Published in final edited form as:

Gene. 2009 January 15; 429(1-2): 80–86. doi:10.1016/j.gene.2008.09.030.

UTR dinucleotide simple sequence repeat evolution exhibits recurring patterns including regulatory sequence motif replacements

Donald E. Riley^{a,b,*} and John N. Krieger^{a,b}

^a Department, Urology, University of Washington, Seattle, WA 98195

^b Department, Research, VA Puget Sound Health Care System, Room 517, Building 1, 1660 South Columbian Way, Seattle, WA 98012

Abstract

New genome sequence information was used to study evolution of 22 dinucleotide simple sequence repeat (diSSR) sites whose upstream flanking sequences were shown to be conserved comparing *Homo sapiens* with the marsupial, *Monodelphis domestica*. Among mammals, most of these diSSR sites were conserved both upstream and downstream of the diSSR. However, individual diSSRs were frequently replaced by alternative repeats. Conserved among mammals examined, the *Vsn11* gene's 3' UTR-localized (AC)_n repeat replaced an A-rich tract in non-mammalian vertebrates examined. The *Sema6D* gene's (GT)_n was also well conserved among mammals examined. Such conservation provides evidence of a functional role. The UTR-localized diSSRs of other genes evolved by replacing alternative diSSRs, by replacing mononucleotide-rich tracts and, in fewer cases, by expansion from short repeating sequences. Extension of the study to less conserved diSSR sites revealed that some diSSRs replaced post-transcriptional regulatory motifs, such as AU-rich elements (AREs) and C-rich tracts. The *Mtap2* gene's UTR-localized (AC)_n was located within a known dendritic targeting element. These evolutionary replacements suggest that some diSSRs belong to a broader group of weak-folding repetitive sequences with potential regulatory roles.

Keywords

SSR, simple sequence repeat; HCE, Highly conserved element; STR, short tandem repeats; microsatellites; nervous system; development; post-transcriptional regulation; ARE, AU-rich elements

Introduction

Stringent selection of UTR-localized dinucleotide simple sequence repeats (diSSRs) flanked upstream by highly conserved elements (HCEs) resulted in selection of 22 genes, including 18 (95%) of 19 with known functions proving important in nervous system development or

Address for correspondence: Donald Riley, Puget Sound Health Care System (VA) Room 517, Building 1; 1660 South Columbian Way, Seattle, WA 98108 email, dri@u.washington.edu, Tel. 206-277-3608; Fax: (206) 764-2164.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

function (please see companion article). Newly available genome sequence data provided the opportunity to examine the evolution of these diSSR sites over a broad range of mammals.

An important observation supporting function of diSSRs was the one-for-one replacement of diSSRs during evolution (Riley and Krieger, 2004; Riley and Krieger, 2005; Riley et al., 2007). For example, *Homo sapiens* aquaporin 3 (*AQP3*) gene UTR has (GT)_n while the orthologous *Rattus rattus* UTR has (AG)_n in the same position. In most cases, such replacements were documented in mouse-human comparisons making it difficult or impossible to discern when replacements occurred. It was also difficult to discern which diSSR was ancestral. In theory, recently available genome sequences would allow us to examine a broader range of mammals to help narrow down the approximate lineages involved in diSSR replacements. Identifying the lineages involved is likely prerequisite to understanding the mechanisms and potential functions of diSSR replacement. For purposes of this study, we relied on published molecular-phylogenetic relationships.

Studying diSSR replacement evolution by genome comparisons should improve our understanding of how diSSRs may be related to other UTR sequences. The current study also examines diSSR sites among nervous system genes with less conserved UTRs because restrictions in the original diSSR selection strategy led to a limited group of diSSR sites. Study of less conserved sites was performed to include additional data that helped expand our understanding of the repertoire of sequence replacements that have occurred at diSSR sites. Our findings on the evolution of diSSRs suggest potential regulatory roles for these sequences.

Methods

2.1. Genome searches

We selected UTR-localized diSSR sites with the most conserved flanking sequences (CFS) available in mammals (please see companion article). The diSSR-upstream flanking sequences of the 22 CFS strategy UTRs were employed in BLASTn searches (<http://www.ensembl.org/Multi/blastview>) of all available vertebrate genomes including: *Mus musculus*; *Rattus norvegicus* (rat); *Oryctolagus cuniculus* (rabbit); *H. sapiens*; *Pan troglodytes* (chimpanzee); *Macaca mulatta* (Rhesus monkey); *Bos Taurus* (cow); *Canis familiaris* (dog); *Felis catus* (cat); *Dasyurus novemcinctus* (armadillo); *Loxodonta africana* (elephant); *Echinops telfairi* (hedgehog tenrec); *Monodelphis domestica* (opossum); *Gallus gallus* (chicken); *Xenopus tropicalis* (African tree frog); *Danio rerio* (zebra fish, “Fish 1”); *Takifugu rubripes* (Japanese puffer fish, “Fish 2”); and *Tetraodon nigroviridis* (green spotted pufferfish, “Fish 3”). Search sensitivity was set to pre-defined parameters described as, “near exact matches”, with search parameters defined in the “help” file of the website. Then orthologous sequences were downloaded, including the diSSR plus 1,000 bases of upstream and downstream sequence.

2.2. DNA sequence alignments

The multiple alignment program, *Clustal W* (*MegAlign*, 6.1, LaserGene; DNASTar™) (Thompson et al., 1994) was used to align diSSR-flanking sequences with the Slow-Accurate algorithm, gap-penalty set to 15.0, gap-length penalty set to 6.7 and a DNA transition weight of 0.5. (Flanking sequences are shown as unedited alignment consensus plots produced by *MegAlign*.) Consensus strength graphics were copied from *MegAlign*’s consensus report output, configured to show minimal extra space between residues. For each gene, approximately 2000 bases of sequence were aligned. Consensus strength plots represent portions of the alignments with the repeat regions near the center unless coding sequences were nearby, in which case the repeat region is off-center. Pre-aligned genomes at

<http://genome.ucsc.edu/cgi-bin/hgBlat?command=start&org=Human&db=hg18&hgsid=-79462048> (Kent, 2002) were also examined to verify *Clustal W* alignments.

The diSSRs and other repetitive sequences were arranged manually since *Clustal W* is inappropriate for certain repeats (e.g., when diSSR-diSSR replacements have occurred). In many cases, the alignment algorithm introduced gaps within repeated sequences, making them difficult to visualize. For this reason, diSSRs and other repeats were arranged for viewing ease with the repeated sequences manually grouped together with as few gaps as possible. This was done with the sequence sliding tools of *MegAlign*. No bases were inserted, removed or rearranged in sequence. For this study, diSSR site remodeling is defined as emergence of a diSSR by any of the following mechanisms: expansion from a very short precursor such as (AC)₃ (Messier et al., 1996) or replacement of another repeat or mononucleotide-rich tract conserved in other mammals. Species are listed in consistent order to assist the reader in identifying which species either lacked an orthologous upstream flanking sequence or, in some cases, had an absence of sequence data where the repeat region would be.

2.3. Phylogeny

Phylogenetic relationships were redrawn from Murphy et al. (Murphy et al., 2001), a phylogenetic study of mammals based on 15 nuclear genes, including coding sequences and UTR sequences. The phylogeny also relied upon 3 mitochondrial genes. Another recent study involved 218 genes and was validated with highly conserved elements (HCEs) taken from genomic sequences (Nikolaev et al., 2007). For the species examined in the current study, the two topologies proved indistinguishable.

Results

3.1. Conservation of upstream and downstream sequences

The CFS strategy identified diSSR sites with, in most cases, both upstream and downstream flanking sequences conserved over most mammalian species whose genome sequences were available (Fig. 1 and supplementary figures). Alignments for genes not shown can be recovered using gene names and genome searches. The *Pou4f2* gene UTR was an exception with the diSSR downstream flanking sequences not well conserved. Conservation of most downstream sequences was unanticipated because only upstream sequences were used to select CFS strategy UTRs. There was no sequence conserved in human and opossum genomes that was not conserved in a wide range of other mammals, although occasionally one or more mammalian species failed to be represented. Potentially, this reflected information gaps in some genome sequences, although other explanations are possible. In every case, diSSR-flanking sequences proved unique, single-copy sequences in their respective genomes.

3.2. Genes with deeply conserved 3'UTR-localized diSSR sites

The *Vsn11* 3' UTR-localized (AC)_n repeat and its flanking sequences were highly conserved among mammals (Fig. 2a). This diSSR replaced an A-rich tract present in fish and *Xenopus* genomes. Among mammals, rare, in-phase substitutions were observed that preserved dinucleotide spacing. The *Sema6D* (GT)_n repeat was conserved, albeit with in-phase substitutions, for the mammals examined (Fig. 2b). Sequences flanking this repeat lacked orthologous sequences among the non-mammalian species examined. The *Pou4f2* gene's 3'-UTR localized (AC)_n replaced a (CT)_n repeat in opossum (Fig. 2c). Alternatively, an ancestral (AC)_n may have been replaced by (CT)_n in the lineage leading to opossum.

3.3. Genes with 3' diSSR sites conserved after the marsupial divergence

We will describe the evolution of additional CFS strategy diSSR sites because some common patterns emerged. Among mammals, the *Nrxn1* (AC)_n repeat was broadly conserved except in opossum, representing the earliest branching lineage examined (Supplementary material). For the *Ches1* gene, opossum exhibits a short A-rich tract; dog exhibits (AG)_n; and mouse exhibits (GT)_n; while most other mammals exhibit (AC)_n (Fig. 3). Relative to other CFS strategy UTRs, this represented a surprising number of replacements. The *Igsf9* 3' diSSR exhibited (GT)_n among mammals representing post-marsupial branches except that the repeat site acquired (CT)_n in the primate lineage (not shown).

3.4. diSSR sites remodeled in the primate- or primate/rodent lineages

Among the 22 CFS repeat sites identified, 7 (32%) underwent remodeling localized to the primate lineage as represented by the species examined (Fig. 4). For the CFS group of genes, replacements proved more common than “new-born” diSSRs (Messier et al., 1996) but we can't rule out the possibility that many diSSRs emerged by expansion at some point in evolution.

Among primates examined, *Camk2n1* and *Zfx1a* genes' 3' UTRs both had well organized (AC)_n tracts (likely due in part to our asymmetric search that required perfect repeats in humans) while other mammals only had A-rich tracts at the corresponding positions for these genes. The *Kif1b* gene's UTR exhibited (AC)_n among primates while other mammals had short, pre-expansion (AC)_n or A-rich tracts (not shown). For *Cln3*, primates examined had lengthy (CT)_n tracts as did rabbit, perhaps representing an independent expansion. In contrast, the rodents examined exhibited poly C at the same position.

The *Cd2ap* gene is critical for kidney podocyte development (Eppig et al., 2005). Flanking the *Cd2ap* 3' diSSR are sequences that, like the majority of CFS diSSR sites, form a conserved island within the UTR. Slight departures from flanking sequence identity (orange regions in Fig. 1) reflect primarily the rodent lineage represented by mouse and rat. Depending on the species examined, three repeats were evident, (AC)_n, (GT)_n and (AT)_n (Fig. 5).

3.5. Investigation of neuronal gene UTRs not recovered by the original CFS strategy

Among 252 human genes identified as having diSSRs, there were 21 nervous system genes not recovered by the CFS strategy because the UTRs were either not conserved or were conserved over more limited ranges of mammals. Among these 21 nervous system genes with diSSRs but not identified by the CFS strategy, we observed three independent replacements of T-rich tracts with diSSRs in *Camk2a*, *Olf628* (Fig. 6) and *Clic6* 3' UTRs. The *Mtap2* UTR-localized diSSR site underwent replacements involving (AC)_n, (GT)_n and (AT)_n and the downstream flanking sequences were less conserved than the upstream flanking sequences. Involvement of (AT)_n in such replacements proved rare. Function studies indicated the *Mtap2* gene UTR-localized (AC)_n was located within a known dendritic targeting element (Fig. 7). Other sites not identified by the CFS strategy represented divergent sequences among mammals, limiting the usefulness of such comparisons.

Discussion

4.1. Conservation of regions broader than the selected diSSR upstream flanking sequences

When selected for their conserved upstream flanking sequences some diSSRs proved remarkably well-conserved across a broad range of mammals. In most cases, the downstream flanking sequences were well-conserved also. These findings were unexpected since we only selected for upstream sequence conservation and diSSRs are sometimes thought of as unstable sequences. Even though nestled among HCEs, some diSSRs are polymorphic (e.g., *Vsn11* and

Sema6D gene UTRs) among individuals within a species as described in the companion manuscript. Taken together, these observations strongly support the previously proposed hypothesis that some diSSRs function. Providing genetic variation is likely to represent part of that function (Kashi and King, 2006).

Some diSSRs exhibited replacement histories involving alternative diSSRs or mononucleotide tracts. Alternative phylogenetic models to the one used here (Fig. 4) might lead to some slight alterations in estimates of when diSSR site remodeling occurred. But, the phylogenetic relationships presented are supported by two major studies (Methods). To our knowledge, the major branch points are uncontroversial. In any case, slightly altered phylogenies would not impact our conclusions.

4.2. A pattern of dinucleotide repeats replacing mononucleotide repeats

Among the 22 genes identified by the CFS strategy, 4 had (AC)_n sequences that replaced A-rich tracts present in representatives of earlier branching lineages. The approximate lineages involved in the replacements and their direction (A-rich to (AC)_n) are fairly unambiguous. These genes included *Vsn11* (Fig. 2), *Zfx1a*, *Camk2n1* (both in supplementary material) and *Kif1b* (not shown). The *Ublcp1* 3' UTR has an (AC)_n repeat in human. For the same gene, other species were surprisingly varied, exhibiting either a T-rich tract, an A-rich tract or a C-rich tract at the diSSR position (Fig. S6). We also observed replacements involving (CT)_n and poly-T tracts (Fig. 6). For *Cln3*, (CT)_n was present among mammals examined except mouse and rat that had C-rich tracts instead (not shown). Due to its presence in a wide range of mammals the *Cln3* (CT)_n appears more ancient, an exception because mononucleotide tracts appear to be more ancestral among the majority of sites examined.

4.3. Weak-folding sequences and potential regulatory sequences

The weak-folding hypothesis is that sequences at diSSR sites are selected based on a relative lack of potential to form higher order structures. The hypothesis has provided a fairly consistent explanation for the wide variety of sequence replacements observed.

As described in the companion article, the CFS strategy led to a focus on neuronal genes. The 3' UTRs of neuronal genes contain multiple elements important for regulating mRNA stability and transport (Snee et al., 2002; Shan et al., 2003; Smith, 2004; Hirokawa and Takemura, 2005). Such post-transcriptional regulatory elements include AU-rich elements (AREs) and poly C tracts that both modulate mRNA stability, and poly T tracts involved in mRNA transport (Blichenberg et al., 1999; Mori et al., 2000; Chen et al., 2002; Erlitzki et al., 2002; Snee et al., 2002; Sommer et al., 2005). Like the weak-folding diSSRs, AREs, C-rich and T-rich tracts are weak-folding sequences. Because the original strategy was somewhat arbitrary, we performed additional genome comparisons for 21 human neuronal genes initially identified as having diSSRs but not recovered by the initial CFS selection strategy. Taken together with the genome comparisons of the 22 original UTRs, and the function studies, our data support several potentially interesting findings.

For example, *Camk2a* mRNA is transported to dendritic extensions (Mori et al., 2000). The T-rich tracts in the 3' UTR-localized HCE of the *Camk2a* gene encode U-rich tracts that interact with RNA-binding proteins that regulate multiple genes post-transcriptionally (Chung et al., 1997; Aranda-Abreu et al., 1999). In the rodent lineage, the T-rich tracts were replaced by (CT)_n. Thus, a diSSR replaced a suspected post-transcriptional regulatory element.

For *Olfir628*, the dog UTR has an ARE motif. Rabbit exhibited multiple ARE motifs and a poly T tract in this position (Fig. 6b). Mouse and rat had (CT)_n in the same position while hedgehog tenrec had (AC)_n. These replacements again suggest potential regulatory roles for the diSSRs.

The mouse myristoylated alanine-rich C kinase substrate (*Marcks*) gene 3' UTR-localized HCE contains an ARE shown to be critical for posttranscriptional *Marcks* gene regulation (Wein et al., 2003). Most other species have T-rich tracts in this position.

Deletion of a 92 bp T-rich tract in the rat *Tau* protein's mRNA 3' UTR sequence prevents anterograde axonal transport of the mRNA, indicating that the T-rich tract is an mRNA transport motif (Aranda-Abreu et al., 1999; Poorkaj et al., 2001). The T-rich tract for *Tau* protein was not well conserved among mammals. For this reason, more conserved repeat sites such as the diSSR sites studied here may prove useful for determining if regulatory sequences like T-rich tracts can be replaced by alternative sequences that function similarly.

The diSSR (AC)_n of *Mtap2* (Fig. 7) is physically located within a known dendritic targeting element (Blichenberg et al., 1999). C-rich tracts were replaced by the diSSRs (AC)_n and (CT)_n in the 3' UTRs of the *Ublcp1* (Supplementary Fig. S6) and *Igsf9* genes respectively. The C-rich sequence in the 3' UTR of the α -globin gene is crucial to mRNA stability and is mediated by an RNA-binding protein, α CP (Wang et al., 1995; Ji et al., 2007). These observations suggest that some 3' localized diSSRs might function as targeting or stability elements because during evolution, some diSSRs replaced regulatory and targeting motifs occurring in the same positions among otherwise well conserved sequences. The sequence comparisons in our study are suggestive but further verification is needed.

Other investigators reported that HCEs are enriched in predicted secondary structure (Siepel et al., 2005). Maintaining loop forming sequences, such as weak-folding diSSRs, within some HCEs may provide access to otherwise tightly folded sequences (Pedersen et al., 2006). Such access might be needed for RNA-binding and RNA-unfolding activities. Based on lack of complementary bases, many repeating sequences (e.g., poly A, poly C, poly T (AC)_n, (AG)_n, (GT)_n, (CT)_n and most AREs) could in theory replace one another if function only depends on preservation of loop-forming potential. We cannot exclude the possibility that primary sequence changes provide currently unsuspected, selective advantages.

To identify and test putative mRNA trafficking and stability elements, it may prove helpful to further expand the repertoire of known diSSR replacements as we have started to do in this study. This approach might prove especially helpful for understanding human trafficking and stability elements, since the mouse model may not work in cases where such motifs have been replaced by alternative sequences.

4.4. The role of (AT)_n

Potential roles of the “strong-folding” diSSR (AT)_n remain uncertain. (AT)_n is only slightly less common among UTRs than the weak-folding diSSRs. Replacements involving (AT)_n occur (Riley et al., 2007) but proved less common than weak-folding diSSR replacements. The current studies did not uncover (AT)_n replacements that involved functional motifs. However the finding of some (AT)_n sequences within highly conserved flanking sequence environments supports the hypothesis that (AT)_n may have an important function.

4.5. Evolution of diSSRs and developmental differences

Previous studies indicate that SSRs modulate gene activity to provide an important source of genetic variation (Li et al., 2004; Kashi and King, 2006). Our findings also suggest regulatory roles based on replacements of regulatory motifs. Consistent with function, some diSSRs exhibited long periods of conservation or maintenance by replacement with other weak-folding sequences including alternative diSSRs. Future studies are needed to elucidate potential selective forces leading to diSSR site remodeling. Our study begins this process by narrowing the range of potential lineages involved in diSSR site remodeling events (Fig. 4).

Conclusions

Strong conservation of some UTR-localized diSSR sites supports the hypothesis that these sequences function. Evolution of diSSRs shows recurring patterns frequently involving diSSR-diSSR replacement and the replacement of mononucleotide rich tracts. In several cases, post-transcriptional regulatory sequence motifs were replaced by diSSRs suggesting a regulatory role for the latter. Our observations are consistent with the weak-folding hypothesis, where repeating sequences lacking canonical base pairs tend to replace one another during evolution. The diSSR (AT)_n is an exception. Single base substitutions within diSSRs usually preserve dinucleotide phasing consistent with some selection at the primary sequence level. Further genome comparisons of diSSR sites and other SSR sites may help improve understanding of these relatively common, non-coding sequences.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Abbreviations

ARE	AU-rich element
CPS	coding potential score
diSSR	dinucleotide simple sequence repeat
HCE	highly conserved element
UTR	untranslated region

References

- Aranda-Abreu GE, Behar L, Chung S, Furneaux H, Ginzburg I. Embryonic lethal abnormal vision-like RNA-binding proteins regulate neurite outgrowth and tau expression in PC12 cells. *J Neurosci* 1999;19:6907–17. [PubMed: 10436048]
- Blichenberg A, Schwanke B, Rehbein M, Garner CC, Richter D, Kindler S. Identification of a cis-acting dendritic targeting element in MAP2 mRNAs. *J Neurosci* 1999;19:8818–29. [PubMed: 10516301]
- Chen CY, Xu N, Shyu AB. Highly selective actions of HuR in antagonizing AU-rich element-mediated mRNA destabilization. *Mol Cell Biol* 2002;22:7268–78. [PubMed: 12242302]
- Chung S, Eckrich M, Perrone-Bizzozero N, Kohn DT, Furneaux H. The Elav-like proteins bind to a conserved regulatory element in the 3'-untranslated region of GAP-43 mRNA. *J Biol Chem* 1997;272:6593–8. [PubMed: 9045688]
- Eppig JT, et al. The Mouse Genome Database (MGD): from genes to mice--a community resource for mouse biology. *Nucleic Acids Res* 2005;33:D471–5. [PubMed: 15608240]
- Erlitzki R, Long JC, Theil EC. Multiple, conserved iron-responsive elements in the 3'-untranslated region of transferrin receptor mRNA enhance binding of iron regulatory protein 2. *J Biol Chem* 2002;277:42579–87. [PubMed: 12200453]
- Hirokawa N, Takemura R. Molecular motors and mechanisms of directional transport in neurons. *Nat Rev Neurosci* 2005;6:201–14. [PubMed: 15711600]

- Ji X, Kong J, Carstens RP, Liebhaber SA. The 3' untranslated region complex involved in stabilization of human alpha-globin mRNA assembles in the nucleus and serves an independent role as a splice enhancer. *Mol Cell Biol* 2007;27:3290–302. [PubMed: 17325043]
- Kashi Y, King DG. Simple sequence repeats as advantageous mutators in evolution. *Trends Genet* 2006;22:253–9. [PubMed: 16567018]
- Kent WJ. BLAT--the BLAST-like alignment tool. *Genome Res* 2002;12:656–64. [PubMed: 11932250]
- Li YC, Korol AB, Fahima T, Nevo E. Microsatellites within genes: structure, function, and evolution. *Mol Biol Evol* 2004;21:991–1007. [PubMed: 14963101]
- Messier W, Li SH, Stewart CB. The birth of microsatellites. *Nature* 1996;381:483. [PubMed: 8632820]
- Mori Y, Imaizumi K, Katayama T, Yoneda T, Tohyama M. Two cis-acting elements in the 3' untranslated region of alpha-CaMKII regulate its dendritic targeting. *Nat Neurosci* 2000;3:1079–84. [PubMed: 11036263]
- Murphy WJ, Eizirik E, Johnson WE, Zhang YP, Ryder OA, O'Brien SJ. Molecular phylogenetics and the origins of placental mammals. *Nature* 2001;409:614–8. [PubMed: 11214319]
- Nikolaev S, Montoya-Burgos JI, Margulies EH, Rougemont J, Nyffeler B, Antonarakis SE. Early history of mammals is elucidated with the ENCODE multiple species sequencing data. *PLoS Genet* 2007;3:e2. [PubMed: 17206863]
- Pedersen JS, Bejerano G, Siepel A, Rosenbloom K, Lindblad-Toh K, Lander ES, Kent J, Miller W, Haussler D. Identification and classification of conserved RNA secondary structures in the human genome. *PLoS Comput Biol* 2006;2:e33. [PubMed: 16628248]
- Poorkaj P, Kas A, D'Souza I, Zhou Y, Pham Q, Stone M, Olson MV, Schellenberg GD. A genomic sequence analysis of the mouse and human microtubule-associated protein tau. *Mamm Genome* 2001;12:700–12. [PubMed: 11641718]
- Riley DE, Krieger JN. Simple repeat replacements support similar functions of distinct repeats in inter-species mRNA homologs. *Gene* 2004;328:17–24. [PubMed: 15019980]
- Riley DE, Krieger JN. Short tandem repeat (STR) replacements in UTRs and introns suggest an important role for certain STRs in gene expression and disease. *Gene* 2005;344:203–211. [PubMed: 15656986]
- Riley DE, Jeon JS, Krieger JN. Simple repeat evolution includes dramatic primary sequence changes that conserve folding potential. *Biochem Biophys Res Commun* 2007;355:619–25. [PubMed: 17320822]
- Shan J, Munro TP, Barbarese E, Carson JH, Smith R. A molecular mechanism for mRNA trafficking in neuronal dendrites. *J Neurosci* 2003;23:8859–66. [PubMed: 14523087]
- Siepel A, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* 2005;15:1034–50. [PubMed: 16024819]
- Smith R. Moving molecules: mRNA trafficking in Mammalian oligodendrocytes and neurons. *Neuroscientist* 2004;10:495–500. [PubMed: 15534035]
- Snee M, Kidd GJ, Munro TP, Smith R. RNA trafficking and stabilization elements associate with multiple brain proteins. *J Cell Sci* 2002;115:4661–9. [PubMed: 12415010]
- Sommer S, Cui Y, Brewer G, Fuqua SA. The c-Yes 3'-UTR contains adenine/uridine-rich elements that bind AUF1 and HuR involved in mRNA decay in breast cancer cells. *J Steroid Biochem Mol Biol* 2005;97:219–29. [PubMed: 16289864]
- Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994;22:4673–4680. [PubMed: 7984417]
- Wang X, Kiledjian M, Weiss IM, Liebhaber SA. Detection and characterization of a 3' untranslated region ribonucleoprotein complex associated with human alpha-globin mRNA stability. *Mol Cell Biol* 1995;15:1769–77. [PubMed: 7862166]
- Wein G, Rossler M, Klug R, Herget T. The 3'-UTR of the mRNA coding for the major protein kinase C substrate MARCKS contains a novel CU-rich element interacting with the mRNA stabilizing factors HuD and HuR. *Eur J Biochem* 2003;270:350–65. [PubMed: 12605686]

Figure 1.

Consensus strength plots of diSSR regions from Megalign (*Clustal W*) alignments. Red, taller vertical bars represent identical bases in all species; Orange, one or two mismatches; Green, 50% identity; White, minimal identity. The number of species whose sequences were available is shown in parentheses next to gene names. Bracket, diSSR region. *, *Camk2α*, was conserved in a limited range of mammals.

Sema6d

Pou4f2

Gene. Author manuscript; available in PMC 2010 January 15.

Gene. Author manuscript; available in PMC 2010 January 15.

Figure 4.

Phylogeny of mammals with available genomes showing approximate branch points involved in CFS gene diSSR remodeling events. The phylogenetic tree was redrawn from Murphy et al. (Murphy et al., 2001). Horizontal branch lengths represent relative time since divergence based on DNA sequence data.

Upstream 3' UTR-diSSR region

Dendrite targeting element

200 bp

Figure 7. Consensus strength plot and diSSR sequences at the *Mtap2* 3' UTR-localized upstream diSSR (one of two diSSRs in the UTR). Red, taller vertical bars represent identical bases in all rodent and primate species examined; Orange, one or two mismatches; Green, 50% identity; White, minimal identity. Underlined bases, in-phase substitutions.