Available online at www.sciencedirect.com

SCIENCE DIRECT®

Gene 344 (2005) 203–211

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

www.elsevier.com/locate/gene

Short tandem repeat (STR) replacements in UTRs and introns suggest an important role for certain STRs in gene expression and disease

Donald E. Riley*, John N. Krieger

Department of Urology, University of Washington, Seattle, WA 98195, USA
Department of Research, VA Puget Sound Health Care System, Room 517, Building 1, 1660 South Columbian Way, Seattle, WA 98012–1532, USA

Received 1 June 2004; received in revised form 25 August 2004; accepted 28 September 2004
Available online 7 December 2004
Received by S.M. Mirkin

Abstract

Some untranslated sequence (UTR)-localized, short tandem repeats (STRs) exhibit evidence of selection pressure, including STR-coupling preferences, STR conservation, interspecies STR–STR replacements, and STR variants implicated in certain diseases. We wished to determine if STR replacements occurred near disease-related genes, including previously unstudied STRs as well as some STRs already implicated in disease. Among nine strong-candidate prostate cancer (CaP)-predisposing genes, three [steroid 5-α-reductase 2 (*Srd5A-2*), macrophage scavenger receptor-1 (*MSR-1*), and tumor necrosis factor receptor-21 (*Tnfr*-21)] exhibited striking STR replacements ($P<0.001$). The glomerular disease-related gene, *CD2AP*, exhibited an STR replacement flanked by well-conserved sequences, suggesting an STR-focused process. Another glomerular disease-related gene, *rabphilin 3A*, exhibited at least two STR replacements at the same UTR position comparing *Drosophila melanogaster*, *Mus musculus*, and *Homo sapiens*. Two genes implicated in blood-clotting disorders, von Willebrand factor (*vWA*) and fibrinogen α (*FGA*), exhibited multiple-intron STR replacements among mammals, extending STR replacement phenomena to introns. Among primates, a tyrosine hydroxylase (*THO1*) intron STR, previously implicated in both schizophrenia and drug withdrawal delirium, exhibited frequent replacements. Some STR replacements were early events in gene divergence. When STR sequences of closely related species were available, STR replacement was observed to be nearly as rapid as speciation. STR replacements expand the list of STR sequences that may contribute to genetic activity and to disease processes.

Keywords: Short tandem repeats; Prostate cancer; Chronic prostatitis; Hemophilia; Tyrosine hydroxylase

1. Introduction

Studies of genetic components of diseases such as prostate cancer (CaP) contain a bias toward investigating coding rather than noncoding sequences. Many short tandem repeats (STRs) are transcribed as 5′ and 3′ untranslated sequences (UTRs) (Riley and Krieger, 2004a). If such STRs influence gene function, then these STRs may be among the first noncoding sequences whose functions can be determined since they represent some of the most recognizable noncoding sequence landmarks.

We define an STR replacement as dissimilar STR sequences identically located near orthologous genes comparing two or more species. We previously demon-

Abbreviations: AQP, Aquaporin; CP/CPPS, Chronic prostatitis/chronic pelvic pain syndrome; CAP, Prostate cancer; DDBJ, DNA Databank of Japan; EMBL, European Molecular Biology Laboratory; *FGA*, Fibrinogen α; *MSR-1*, Macrophage scavenger receptor-1; n, Number of repetitions; STR, Short tandem repeat; *Srd5A-2*, Steroid 5-α-reductase 2; *THO1*, Tyrosine hydroxylase; *Tnfr*, Tumor necrosis factor receptor; *vWA*, von Willebrand factor.

* Corresponding author. Department of Research, VA Puget Sound Health Care System, Room 517, Building 1, 1660 South Columbian Way, Seattle, WA 98012–1532, USA. Tel.: +1 206 277 3608; fax: +1 206 764 2164.
E-mail address: dri@u.washington.edu (D.E. Riley).

0378-1119/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2004.09.034

strated STR replacements in the 3′ UTRs of six (11.5%) mRNAs among 52 independent, orthologous pairs of mRNAs that had STRs (Riley and Krieger, 2004b). Such replacements were highly significant ($P<0.001$). The STRs studied occurred in only 0.13% of all mRNAs, so it was extremely unlikely that the observed replacements reflected a random process. We concluded that there was at least one broad STR-focused process important since it is influenced by natural selection.

In all, we reported three independent lines of evidence supporting the hypothesis that some STRs (or the process that forms them) were under selection pressure. (1) Weakly folding STRs, such as (GATA)n, (AC)n, and (GT)n, were preferentially associated with mRNAs encoding membrane-targeted functions such as plasma membrane receptors, proteins that bind membrane proteins (peripheral membrane proteins), and secreted proteins that negotiate the plasma membrane (Riley and Krieger, 2002, 2004a; all STRs will be written as they appear relative to the sense-strand of the associated gene). Some transcription factors that had membrane-protein-like folds were also encoded by mRNAs that had STRs. (2) STRs such as (AC)n, (GT)n, (AG)n, (AT)n, and (GC)n often occurred in paired couples with strong preferences in polarity (order of STRs) and register (sequence at the STR-couple's joint; Riley and Krieger, 2003). For example, the STR-couple (GT)n(GC)n was strongly preferred over the reverse order, (GC)n(GT)n. (3) STRs such as (AC)n, (GT)n, (AG)n, and (CTTT)n sometimes replace one another in evolution (Riley and Krieger, 2004b). Some STR replacements occurred in the context of well-conserved flanking sequences, supporting a previously unrecognized, STR-focused process. Among the 52 rodent and *Homo sapiens* STR-containing mRNA pairs previously studied, STR–STR replacement (11.5%) nearly rivaled STR conservation (17%). Although we predicted STR replacements, they occurred even more often than anticipated. Another 11% of STRs were excised precisely, a finding that also reflected an STR-focused process.

Since there were limited STR-containing interspecies mRNA-sets available, advancing these studies required the inclusion of genomic sequences. We focused on STR-containing, disease-related genes where STR and flanking sequence information was available for multiple species. First, we searched for STR-replacements among genes suspected as having roles in prostate disease. Second, among the six STR replacements reported previously were three kidney-disease-related genes that we examined in more detail. Finally, we searched for STR replacements among three disease-related genes for which substantial STR polymorphism data were available.

2. Methods

Methods for finding orthologous STR-containing cDNA sequences were previously described (Riley and Krieger, 2004b). For the present study, when reported cDNAs in a given species lacked a reported orthologous STR-containing cDNA in alternative species, genomic sequences were then searched to identify the orthologous sequences. Reported *H. sapiens* or *Mus musculus* cDNA sequences were used to search Genbank, EMBL (European Molecular Biology Laboratory), and DDBJ (DNA Databank of Japan; 2,202,239 sequences), using the Blastn (2.2.8, 1-5-04; available from,http://www.ncbi.nlm.nih.gov/BLAST/) search engine without complexity filtering (Altschul et al., 1997). Selectable parameters used were Expect value=10 and Word value=11. Nine prostate cancer (CaP) predisposing genes were selected based on one or more of the following criteria: LOD scores>3.0 in genomic scans, loss of function mutations documented in families with familial CaP, or effective therapies that targeted the gene product (Table 1). For each CaP-predisposing gene included in this study, a finding of strong CaP-predisposing candidacy had been reported by at least three independent laboratories. To include intron-localized STRs of disease-related genes, we studied three clinically interesting loci widely used by forensic laboratories. These STRs are associated with data from a range of populations (Budowle et al., 1999).

The multiple alignment program Clustal W (MegAlign, LaserGene, DNASTAR, Madison, WI; Thompson et al., 1994) was used to align sequences and to determine phylogenetic trees. For alignments, gap-penalty was set to 15.0, and gap-length penalty was set to 6.7. A gap of 110 bases was inserted manually into the *H. sapiens* macrophage scavenger receptor-1 (*MSR-1*) sequence to align the STR with the *M. musculus* sequence, but no other manual changes to alignments were used in this study. After pairwise or multiple sequence alignment, MegAlign reports

Table 1
Criteria for candidacy as a CaP-predisposing gene or locus

Gene/locus	Primary basis for candidacy	Therapy
AR	Linkage, loss-of-function mutations, familial CaP (Giovannucci et al., 1997)	Androgen ablation shrinks tumors
BRCA2	Loss-of-function in familial CaP, RR=4.6 (Simard et al., 2003)	
ELAC2	LOD=4.5, loss-of-function mutations (Simard et al., 2003)	
MSR-1	Loss-of-function mutations in familial CaP (Xu et al., 2003)	
RNASEL	LOD=5.43, loss-of-function mutations, familial CAP (Simard et al., 2003)	
SRD5A	Increased activity variants show increased risk, RR=3.6–7.7 (Simard et al., 2003)	Androgen ablation shrinks tumors
Tnfr-1b	Differential expression assays (Ashkenazi, 2002)	Target, potent antitumor drugs
Tnfr-21	Differential expression assays (Ashkenazi, 2002)	Target of potent antitumor drugs
VDR	LOD=3.0–4.6 (Ingles et al., 1997), aggressive CaP association, Japanese men	