STRBase: a short tandem repeat DNA database for the human identity testing community

Christian M. Ruitberg, Dennis J. Reeder and John M. Butler*

Biotechnology Division, National Institute of Standards and Technology, 100 Bureau Drive, Mail Stop 8311, Gaithersburg, MD 20899-8311, USA

Received August 31, 2000; Accepted September 17, 2000

ABSTRACT

The National Institute of Standards and Technology (NIST) has compiled and maintained a Short Tandem Repeat DNA Internet Database (http://www.cstl.nist.gov/biotech/strbase/) since 1997 commonly referred to as STRBase. This database is an information resource for the forensic DNA typing community with details on commonly used short tandem repeat (STR) DNA markers. STRBase consolidates and organizes the abundant literature on this subject to facilitate on-going efforts in DNA typing. Observed alleles and annotated sequence for each STR locus are described along with a review of STR analysis technologies. Additionally, commercially available STR multiplex kits are described, published polymerase chain reaction (PCR) primer sequences are reported, and validation studies conducted by a number of forensic laboratories are listed. To supplement the technical information, addresses for scientists and hyperlinks to organizations working in this area are available, along with the comprehensive reference list of over 1300 publications on STRs used for DNA typing purposes.

INTRODUCTION

Tandemly repeated DNA sequences, which are widespread throughout the human genome, are polymorphic in nature, making them important genetic markers for mapping studies, disease diagnosis, and human identity testing (1). Short tandem repeats (STRs) contain repeat units that are 2–6 bp in length and can be readily amplified with the polymerase chain reaction (PCR). STRs have become popular in forensic laboratories because low amounts of DNA, even in a degraded form, can be successfully typed. Sample mixtures can be more readily resolved with STR results than with previously used DNA typing technologies (2).

In the United States, a core set of 13 STR markers are being used to generate a nationwide DNA database called the FBI Combined DNA Index System (CODIS). The CODIS Database and similar DNA databases around the world have been successful at linking DNA profiles from repeat offenders and crime scene evidence. STR typing results are also used to aid hundreds of thousands of paternity testing cases each year.

Forensic DNA testing requires stringent guidelines for DNA sample processing and data analysis. New STR kits are validated by conducting a number of tests to verify that results are reliable and robust. A set of quality assurance standards issued by the DNA Advisory Board must be followed by forensic laboratories in order to submit STR profiles to the national CODIS database (3).

CONTENT OF STRBase

The primary content of STRBase is shown in Figure 1. The information is broken into three sections: general, forensic and supplemental information. An introductory PowerPoint presentation in the ‘STRs 101’ section explains STRs and their use in forensics to help familiarize people with this field.

General information

Information describing each commonly used STR marker forms the core of STRBase in a format we call STR fact sheets (Fig. 2). These fact sheets are composed of four sections: general information, PCR primers used, PCR product sizes and additional information. The general information section describes other names commonly used for the STR locus, its chromosomal location, the sequence of the core STR repeat unit, the GenBank accession number and the number of repeat units in the reference sequence. Underlined words shown in Figure 2 are hyperlinked to further information. The second section of these fact sheets, PCR primers used, lists the sequence for published primers, or the amplification kit(s) in which the primer set is available for commercial primers (sequences for primers in commercial kits have not been released at this time) (4). Each primer set is referenced to the published paper from our reference list or the commercial source of the kit. Each reference is assigned a number when it is entered into the database, and is referred to in the ‘Ref.’ column of Figure 2 (see Supplemental information). The PCR product section gives the length and sequence of an amplicon generated using each primer set listed in the PCR primer section for reported alleles. Each sequence is also referenced to a published paper in our reference listing. Links are provided to annotated GenBank sequence information for commonly used STR loci, which include both repeat and flanking regions.
Commercial multiplex STR kits have become widely used by laboratories worldwide because of their ease of use and high discriminatory power. Charts showing the allele product size ranges of commercially available multiplex STR kits, and hyperlinks to vendors are available on STRBase. Information on published multiplexes not commercially available is also provided. Examples from two kits are shown in Figure 3. The PowerPlex® 16 kit simultaneously amplifies 15 different STR loci, including the two pentanucleotide repeat loci Penta D and Penta E, as well as the sex-typing marker amelogenin. A DNA ladder (ILS-600) labeled with the dye CXR (Promega Corporation, Madison, WI) is used as an internal sizing standard. The Profiler Plus® kit amplifies 9 STRs and amelogenin, and uses ROX (Applied Biosystems, Foster City, CA) to label its internal sizing standard (GS-500).

Verification of microvariant alleles is becoming very important as STR typing expands, since they are being observed on a regular basis as more samples are studied. STRBase provides a publication venue by listing new alleles as they are found. We invite scientists to submit information on these new and rare alleles so they may be recognized.

Forensic interest information

The FBI has selected thirteen core loci for the CODIS database, which will be used for linking serial crimes and unsolved cases with repeat offenders nationwide. A chromosomal map in STRBase shows the location of each core locus, with links to STR fact sheets, which are available for all thirteen CODIS core loci.

Verification of microvariant alleles is becoming very important as STR typing expands, since they are being observed on a regular basis as more samples are studied. STRBase provides a publication venue by listing new alleles as they are found. We invite scientists to submit information on these new and rare alleles so they may be recognized.

Figure 1. Overview of topics covered in STRBase.

Figure 2. An example STR fact sheet for the marker D8S1179.
testing or other cases of mixed male/female DNA. STR fact sheets are available for several Y-chromosome STRs as well as relevant links, references and a PowerPoint presentation explaining the use of Y-chromosome STRs for forensic purposes.

Numerous PCR-based sex-typing assays have been reported in the literature. The most commonly used is amelogenin, which differentiates a 6 bp deletion on the X-chromosome from the Y-chromosome. Primer sequences, PCR product sizes and references for amelogenin and three other sex-typing markers are all listed in STRBase.

Supplemental information

Over 1300 references pertaining to STRs and their application to forensic DNA typing have been gathered from journals, conference proceedings, book chapters and other sources. They come from almost 120 sources, and over 800 are from peer-reviewed journals. The abundance of literature available on the use of STRs for forensic DNA typing shows that it has become an established technology worthy of being used as court evidence.

The rapid pace of developing technologies for DNA analysis can make it difficult to keep track of and understand all methodologies. STRBase includes a brief review of techniques that have been successfully implemented for resolving and detecting STR alleles. Relevant references and hyperlinks to groups working in each area are also provided.

STRBase has more than 60 hyperlinks to organizations involved in DNA typing, commercial sources of instrumentation or DNA testing kits, paternity testing laboratories, electronic journals where STR publications have been found and other useful DNA databases.

Addresses for scientists working with STR markers are listed in STRBase with email links, phone and fax numbers. All scientists working with STR markers are invited to add their information to aid in cooperation of DNA typing laboratories around the world. To have your name listed, send an email message with the above information to john.butler@nist.gov

STRBase ACCESS AND DATA ACQUISITION

The short tandem repeat DNA database is available throughout the world at http://www.cstl.nist.gov/biotech/strbase/. When using information from STRBase, please cite this paper and the date which the information was gathered from STRBase.

The information contained in STRBase is taken from published works on short tandem repeats used for DNA typing purposes. The literature is regularly searched for new publications and updates are periodically made. Comments on the database, suggestions for further improvements or submissions should be sent to the corresponding author, attn.: STRBase, or john.butler@nist.gov

NOTE

Certain commercial equipment, instruments or materials are identified in this report to adequately specify an available source of information. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

REFERENCES