

Research Articles

New resources for the forensic genetics community available on the NIST STRBase website[☆]

John M. Butler^{*}

*National Institute of Standards and Technology, Biochemical Science Division, 100 Bureau Dr.,
M/S 8311, Gaithersburg, MD 20899-8311, USA*

Received 16 August 2007; accepted 7 October 2007

Abstract

For more than a decade, the U.S. National Institute of Standards and Technology (NIST) has maintained the short tandem repeat DNA Internet database (STRBase), which is located at <http://www.cstl.nist.gov/biotech/strbase/>. The purpose of STRBase has been and continues to be an attempt to bring together the abundant literature and information in the forensic genetics field in a cohesive fashion to make current and future work easier. New materials are regularly added to expand the valuable information contained on the STRBase website.

© 2008 Elsevier Ireland Ltd. All rights reserved.

Keywords: STRBase; Website; Short tandem repeats; STRs; NIST

1. Introduction

In late 1996 and early 1997, while a postdoctoral researcher working with Dennis Reeder at the National Institute of Standards and Technology (NIST), I compiled the initial information for and launched the short tandem repeat DNA Internet database (STRBase). Many of the core materials came from my Ph.D. dissertation on STR typing by capillary electrophoresis, which was written in the summer of 1995. A comprehensive examination of more than 500 articles from the scientific literature on STRs was conducted to organize useful information from the rapidly growing field of forensic genetics. The initial information contained in STRBase was described previously [1–3]. Over the past decade, aggressive collection of

articles on STRs from forensic journals and other sources has resulted in over 2900 publications, which are listed in the STRBase reference section (http://www.cstl.nist.gov/biotech/strbase/str_ref.htm). This comprehensive collection of publications on the subject has led to several reviews on forensic DNA analysis [4,5].

2. Materials and methods

Information on the STRBase website is contained in hypertext markup language (HTML) files that were created primarily using Microsoft FrontPage Software (Richmond, WA). Over 1000 files now exist containing 10,000 plus printed pages of information that are connected with approximately 4600 hyperlinks. An additional 2300 hyperlinks connects various information on STRBase to other Internet websites including 342 direct links to various organizations, journals, academic and forensic institutes, commercial sites, genetic genealogy labs, parentage testing labs, and legal sites dealing with forensic DNA (see <http://www.cstl.nist.gov/biotech/strbase/weblink.htm>). Thus, STRBase is not a true searchable “database” but rather a collection of information with interconnected files. Since July 2006, resources that are added to STRBase are now tracked on a “recent updates” page: <http://www.cstl.nist.gov/biotech/strbase/updates.htm>.

[☆] Contribution of the U.S. National Institute of Standards and Technology. Not subject to copyright. Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the US Department of Justice. Certain commercial equipment, instruments and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by the National Institute of Standards and Technology nor does it imply that any of the materials, instruments or equipment identified are necessarily the best available for the purpose.

* Tel.: +1 301 975 4049; fax: +1 301 975 8505.

E-mail address: john.butler@nist.gov.

Table 1
Software tools available for download from STRBase (<http://www.cstl.nist.gov/biotech/strbase/software.htm>)

Program name	Author	Purpose
AutoDimer	Peter Vallone (NIST)	Aids multiplex assay design through comparing batches of primers to one another
mixSTR	David Duewer (NIST)	Compares reference profiles to mixtures for rapid allele inclusion/exclusion
Multiplex_QA	David Duewer (NIST)	Enables comparison of quality metrics for ABI 310/3100 electropherogram data
STR_ConvertFormats	David Duewer (NIST)	Transforms data from columns into rows
STR_MatchSamples	David Duewer (NIST)	Compares STR types from multiple data sets to aid concordance checking
OmniPop	Brian Burritt (San Diego PD)	Enables profile match probability calculations across up to 202 populations http://www.cstl.nist.gov/biotech/strbase/populationdata.htm http://www.cstl.nist.gov/biotech/strbase/population/OmniPop200.1.xls

3. Results and discussion

3.1. STR fact sheets, multiplexes, and variant alleles

STR fact sheets are the centerpiece of STRBase and list information regarding genomic location, GenBank accession, repeat structure, reported PCR primer sets, observed allele sizes and sequence structure, commercially available allelic ladders, common multiplexes, and mutation rates. A total of 53 STR fact sheets including all 18 core or common loci used in current commercial STR kits [6] are available at http://www.cstl.nist.gov/biotech/strbase/str_fact.htm. In recent months, new fact sheets for SE33, D2S1338, D19S433, D1S1656, and DYS635 have been created and information on other loci brought fully up-to-date to enable coverage of all commercial STR kits.

Multiplex assay and kits are visually summarized by locus-specific allele size ranges and dye colors with each locus name hyperlinked to the appropriate STR fact sheet. Labs worldwide continue to contribute to knowledge regarding rare alleles such that we now have catalogued 364 variant alleles and 153 tri-allelic patterns.

3.2. Training materials and NIST project team output

Numerous PowerPoint slides, NIST publications and presentations, software programs, and other useful information are available for download and use by the forensic genetics community. Materials from over a dozen recent workshops consisting of thousands of slides covering capillary electrophoresis, low-copy number DNA testing, mixture interpretation, qPCR DNA quantitation, Y-chromosome and mtDNA analysis, and validation are available for use at <http://www.cstl.nist.gov/biotech/strbase/training.htm>.

3.3. Information and tools to aid PCR primer set concordance studies

A new section of STRBase was recently created to track null alleles detected through DNA testing with different primer sets (<http://www.cstl.nist.gov/biotech/strbase/NullAlleles.htm>). Thus far, 38 entries covering 12 different STR loci are listed from various published studies including a recent concordance study of the new MiniFiler kit with other commercial STR kits [7].

A new software program running in Microsoft Excel and entitled STR_MatchSamples was created by David Duewer

(NIST Chemical Sciences and Technology Laboratory) that enables concordance checking with multiple data sets. This program is listed in Table 1 along with others that can be downloaded from STRBase.

4. Conclusions

STRBase has been well received and widely used by the forensic genetics community and we continue to create additional resources for our website. In 2005, NIST adopted STRBase as an official Standard Reference Database (SRD) giving further credence to the value of the information contained in our website.

Acknowledgments

The initial efforts of Christian Ruitberg and on-going support of Jan Redman with maintaining updates to STRBase and the Reference Manager database of our literature collection are much appreciated as was the initial encouragement of Dennis Reeder in starting this project over 10 years ago.

This work was funded in part by the National Institute of Justice through interagency agreement 2003-IJ-R-029 with the NIST Office of Law Enforcement Standards.

Funding source

U.S. National Institute of Justice - but without involvement in the development of this paper.

Conflict of interest

None.

References

- [1] C.M. Ruitberg, D.J. Reeder, J.M. Butler, STRBase: a short tandem repeat DNA database for the human identity testing community, *Nucleic Acids Res.* 29 (2001) 320–322.
- [2] J.M. Butler, D.J. Reeder, STRBase: a short tandem repeat DNA database for the DNA typing community, in: *Profiles in DNA*, vol. 1, no. 2, Promega Corporation, 1997, p. 10, http://www.promega.com/profiles/102/Profile-sinDNA_102_10.pdf.
- [3] J.M. Butler, C.M. Ruitberg, D.J. Reeder, STRBase: a short tandem repeat DNA Internet-accessible database, in: *Proceedings of the Eighth International*

- Symposium on Human Identification 1997, Promega Corporation, 1998, pp. 38–47, <http://www.promega.com/geneticidproc/ussymp8proc/13.html>.
- [4] T.A. Brettell, J.M. Butler, R. Saferstein, *Forensic science, Anal. Chem.* 77 (2005) 3839–3860.
- [5] T.A. Brettell, J.M. Butler, J.R. Almirall, *Forensic science, Anal. Chem.* 79 (2007) 4365–4384.
- [6] J.M. Butler, Genetics and genomics of core STR loci used in human identity testing, *J. Forensic Sci.* 51 (2) (2006) 253–265.
- [7] C.R. Hill, M.C. Kline, J.J. Mulero, R.E. Lagace, C.-W. Chang, L.K. Hennessy, J.M. Butler, Concordance study between the AmpFISTR Mini-Filer PCR amplification kit and conventional STR typing kits, *J. Forensic Sci.* 52 (4) (2007) 870–873.