

Introduction

OSIRIS (Open Source, Independent Review & Interpretation System) is a public domain quality assurance software package that facilitates the assessment of multiplex STR DNA profiles based on laboratory-specified protocols. OSIRIS evaluates the raw electrophoresis data contained in any .fsa file using a new mathematical approach that assures its independence from other microsatellite genotype analysis software.

The algorithm iteratively fits expected parametric data signatures to the observed data to identify peaks, usually achieving matches with correlations in excess of 0.999. Parametric peak locations are determined with sub-second accuracy and transformed to base pair coordinates. OSIRIS departs from sizing approaches that traditionally rely on either the local or global Southern methods (1-3) to interpolate the ILS into base pair estimates. Instead, OSIRIS uses the correspondence between a sample's ILS and the ILS in the associated allelic ladder to map the time scale of the ladder into that of the sample. This allows effective integration of the ladder with the sample for a straightforward and accurate comparison (typically within 0.1 of a base pair) of sample peaks with ladder locus peaks. Thus, in addition to extremely sensitive peak analysis, OSIRIS offers two new peak quality measures fit level and sizing residual. These new measures can enhance quality metrics currently available to assess STR DNA profiles.

OSIRIS accommodates laboratory-specific signatures, including adjusted sensitivity to background noise, customized naming conventions and internal laboratory controls. When used in complement with other analysis methods, OSIRIS provides an independent review to assure data concordance. With appropriate forensic validation and NDIS approval, OSIRIS may alleviate the need for visual review of passing profiles.

The National Center for Biotechnology Information (NCBI) developed OSIRIS in collaboration with state, local, and federal forensic laboratories and the National Institute of Standards and Technology (NIST). This freely available, object-oriented software is written in C++ to facilitate the development of add-on applications by the private and public sectors. NCBI performs internal quality assurance on its programs and maintains OSIRIS at http://www.ncbi.nlm.nih.gov/projects/SNP/osiris as part of its extensive public domain toolkit for exploring and managing genetic data.

Background

OSIRIS was initiated in response to recommendations of a multidisciplinary advisory group (the Kinship and DNA Analysis Panel, KADAP). KADAP was empanelled to assist the New York City Office of the Chief Medical Examiner (NYC-OCME) and New York State Police (NYSPD) DNA laboratories in the difficult and unprecedented legal and humanitarian challenges faced in the World Trade Center victim identifications.⁴ In its after-action "Lessons Learned" report⁵ the panel focused significant attention on the critical need for quality systems in all facets of mass fatality victim identifications, noting that such systems must be readily available components of the nation's emergency preparedness plans. The group stressed that creating processes with intentional redundancy, such as using multiple test and software systems, assures the validity of results in situations where protocols may be rapidly evolving and identifications are being made in an emotionally charged environment under keen media scrutiny.

As new mass fatalities brought DNA victim identification to the forefront, the need for openly available quality assurance software became even more evident. In the aftermath of Hurricane Katrina, with tens of thousands feared dead, the Louisiana State Police victim identification program initiated careful collections and analyses of known DNA standards from relatives of those reported missing. As is the case with all high throughput DNA profiling, human review became a limiting factor. A pre-alpha version of OSIRIS was tested by the Katrina Victim Identification team to assist analysts with review of controls associated with the files and help expedite sample pipelines for review and identification of files needing re-analysis with appropriate controls.

While the impetus for developing OSIRIS was a direct response to issues in mass fatality victim identification, it also dovetailed with the needs of the federally funded DNA Backlog Reduction program. Convicted offender backlog reduction, identified as a top priority by the National Commission on the Future of DNA Evidence at its first meeting on March 18, 1998, received its initial appropriation in FY2000. The program has funded the analysis of millions of convicted offender profiles, most of which now reside in NDIS which has an Investigations Aided rate of more than 74,000 cases.

Technique maxi

- ratio; good qua residuals
- Signal s(t) and totally uncorrel - Therefore, s(information If n(t) is not " exhibits po Cause for OSIRIS loo artifact-rel

What Kind of

- Based upon sig called "matche - Maximizes s signatures o
- Computationa
- OSIRIS variant signal-to-noise variable profile - Computationa

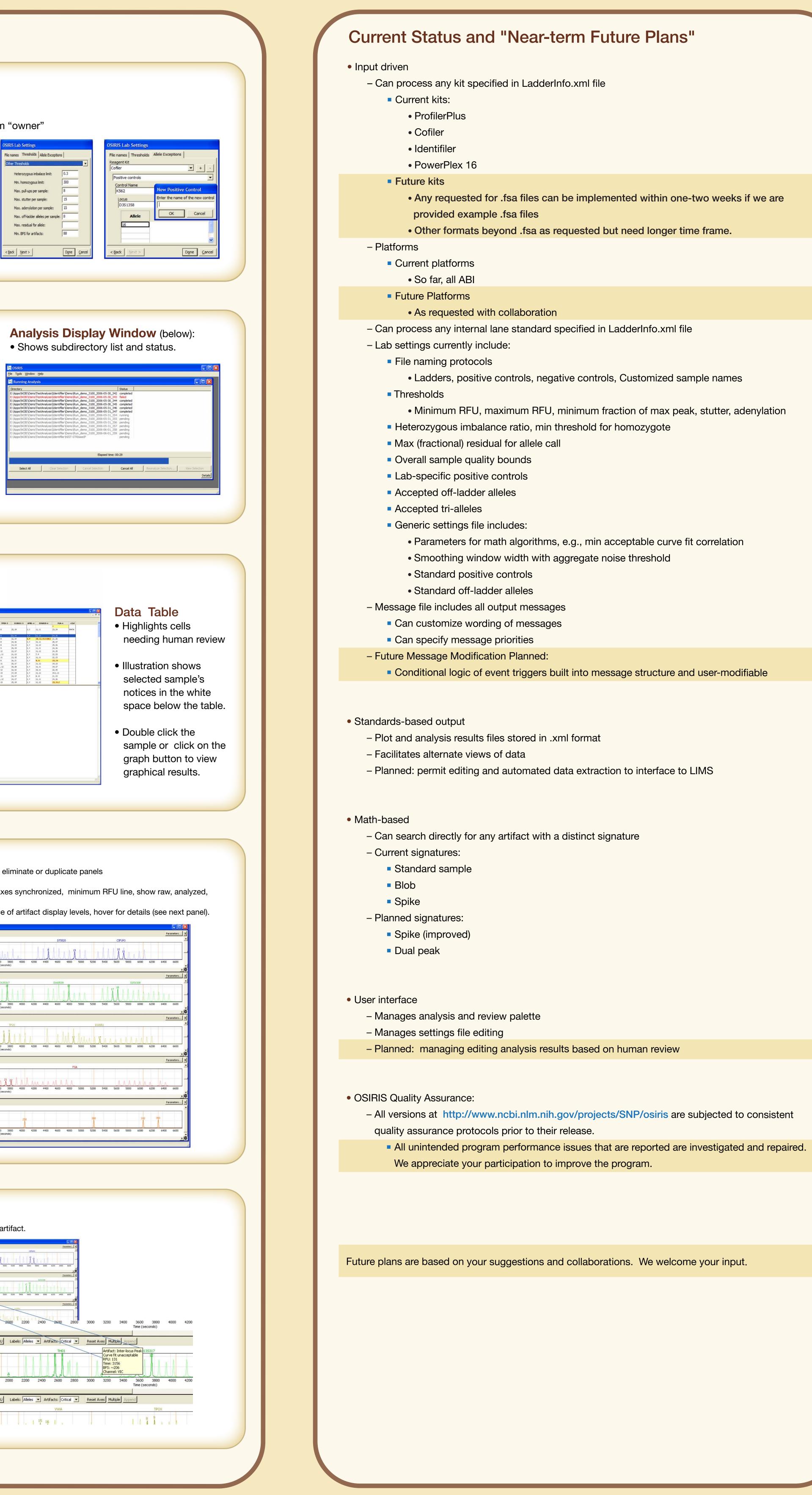
Convert Time

- Conventional: – Uses known t corresponde time into base Assumes app global)
- **OSIRIS**: – Using cubic time into lado – Uses map to sample time f

OSIRIS OUALITY ASSURANCE SOFTWARE

Authors: Lisa Forman Neall (NCBI), Stephen Sherry (NCBI), Robert Goor (NCBI), John Butler (NIST), Margaret Kline (NIST), David Duewer (NIST), Chris Carney (FDLE), Michael Coble (AFDIL), Angelo Della Manna (ADFS), Janos Murvai (SE), Mechthild Prinz (OCME), Tammy Pruit Northrup (SE), Gregory Risch (ADFS), G. Sue Rogers (ADFS), George Riley (AABB), Taylor Scott (ISP), Amanda Sozer (SNA), Mark Squibb (MVRCL), Kerry Zbicz (NCBI), and James Ostell (NCBI)

Nethodological Foundation		Graphic Interface
 OSIRIS Plan Math-based algorithms Independent approaches offer alternatives Reduced reliance on heuristics Transferable to other problem domains C++ / Object-oriented design Transparent code Extend to other genomic methods (e.g., SNP's) Isa input data /.xml input configuration and output files Standards – based Results transformable for use in other tools (e.g. LIMS) 	<section-header><text><text><equation-block><list-item><list-item><list-item><list-item></list-item></list-item></list-item></list-item></equation-block></text></text></section-header>	<text><list-item><list-item></list-item></list-item></text>
<section-header><section-header><list-item><list-item><list-item><list-item><list-item><list-item><list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></section-header></section-header>	<section-header><section-header><text><text><text></text></text></text></section-header></section-header>	OSIRIS Analysis Screen (below): • "Input Directory" data files and/or subdirectories containing .fsa files to be analyzed • "Output Directory" Contains a directory structure identical to input directories for easy identification. • Try Directory [Pogram FleeVCBI/Dairs(VEst/AnalysisUdentifie]]
 Presidual Noise Technique maximizes signal-to-noise ratio; good quality data have small residuals Signal s(t) and residual noise n(t) are totally uncorrelated Therefore, s(t) represents all usable information extracted from data If n(t) is not "small enough", the peak exhibits poor morphology Cause for an alert OSIRIS looks for specific artifact-related signatures 	 Dositioning the Signature Step 1: Look for measurement "density" Aggregate measurements within a moving window of fixed width (e.g., 10 sec) Select intervals in which aggregate peaks above a threshold achieve local minima at endpoints Smoothes noise Step 2: Find best fit "Gaussian" (mean and variance) by maximizing correlation between signature and data Uses numerical iteration 	
5 What Kind of Algorithm Is This? • Based upon signal processing technique called "matched filtering" which: • Maximizes signal-to-noise ratio for signatures of fixed profile • Computationally expensive	<section-header></section-header>	<section-header></section-header>
 OSIRIS variant accomplishes same signal-to-noise maxima for signatures of variable profile (variance) but: Computationally frugal 	<text><figure><figure><text></text></figure></figure></text>	
<section-header><section-header><list-item><list-item><list-item><list-item></list-item></list-item></list-item></list-item></section-header></section-header>	<text><text></text></text>	<section-header></section-header>
9	10	



Summary

sence of tangible remains for families. State and Federal oversight requirements for data generated in these situations compel laboratories to find new resources for data review and reinforce the need for broadly available quality management tools. The increasing capacity challenges on forensic identification systems now provide multiple opportunities to deploy the next generation of management tools for quality assessment.

There are multiple uses for open-source, public-domain quality assurance software for genetic data. As the lessons of 9/11 taught us, the intersection between public health and public safety unites us in the need to assure quality tools are available to any and all who might be responsible for victim identification in mass fatalities. This was the impetus for OSIRIS' original development. However, as those roads diverge, critical needs remain for both public health and public safety to rapidly and efficiently evaluate samples for their value to be maximized. Automating quality assurance with an independent verification through OSIRIS permits valuable human resources to focus on higher level concerns. OSIRIS provides signal-processing algorithms that can be used to assess data in any number of platforms from multiplex STRs to GWAS SNP chips. Similarly, OSIRIS open availability allows it to be used in teaching, training, and management activities beyond the quality concerns of a high throughput environment.

Although OSIRIS' graphic interface was purposely modeled on a look and feel that most analysts' would find familiar and comfortable, the algorithms upon which OSIRIS is based represent a novel approach within signal processing. OSIRIS can be used as a stand-alone quality assessment tool but its operational independence also provides opportunities to test or substantiate results from other sources. Finally, its transparent public domain architecture benefits the legal system as well as encourages opportunities for customization and development of add-on applications.

References

- Elder, JK & Southern, EM (1983) Analytical Biochemistry 128:227-231
- . Hartzell, B., Graham, K. & McCord, B (2003) *Forensic Science International* 133:228-234
- 3. Klein, S.B. Wallin, JM, & Buoncristiani, MR (2003) *Forensic Science Communications*,
- 5 at: http://www.fbi.gov/hq/lab/fsc/backissu/jan2003/klein.htm. . Biesecker, L. et al. (2005) *Science* 310:1122-1123
- 5. http://www.ncjrs.gov/pdffiles1/nij/214781.pdf

Acknowledgements

This work began as a one year Memorandum of Understanding between NCBI and the National Institute of Justice in 2003. The project is supported by the Intramural Research Program of the National Institutes of Health, through the Director of the NCBI, Dr. David Lipman. Jonathan Baker, NIST, contributed extensively to the early development of OSIRIS. David Coffman (FDLE), Cecelia Crouse (PBCSO), and Inspector Mark Dale (NYSPD, retired) have generously shared ideas, critiques and insights throughout OSIRIS' development. Jeff Ban and Susan Greenspoon (VA-DFS) offered helpful perspectives and discussions for considering new platforms. Viviana Van Deerlin (U. Penn Med School) kindly provides OSIRIS with a clinical perspective. We thank Alec Rezanka for his inspired naming of the program. Don Preuss and his team facilitated our demonstration hardware and Phi Ngo, Jalinda Hull and Latisha Wilson provided essential administrative support.

Author Affiliations

NCBI = National Center for Biotechnology Information, Information Engineering Branch, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894

NIST = National Institute of Standards and Technology, Biotechnology Division, 100 Bureau Drive MS8311, Gaithersburg, MD 20899-8311

FDLE = Florida Department of Law Enforcement, DNA Investigative Support Database, 2729 Fort Knox Boulevard, Building 2; Suite 1100, Tallahassee, FL 32308

AFDIL = Armed Forces DNA Identification Laboratory, 1413 Research Blvd. Bldg 101, 2nd floor, Rockville, MD 20850

SE = Self Employed; for Tammy Pruit Northrup: Attorney At Law, LLC, Post Office Box 2213, St. Francisville, LA, 70775

SNA = Sozer, Niezgoda and Associates, LLC, 407 Crown View Drive, Alexandria, VA 22314

ADFS = Alabama Department of Forensic Sciences, 2026 Valleydale Rd., Hoover, AL 35244-2028

OCME-NYC = Office of the Chief Medical Examiner, New York City, 421 E. 26th St., NY, NY 10016 **ISP** = Taylor Scott, Illinois State Police, DNA Indexing Laboratory, 3710 East Lake Shore Drive,

Springfield, IL 62712

MVRCL = Miami Valley Regional Crime Lab, 361 W. Third St., Dayton, OH 45402





