OSIRIS Version 2.2:
Intelligent Software for Casework and CODIS DNA Analysis

George Riley, PhD, Robert Goor, PhD, Douglas Hoffman, MS, Stephen Sherry, PhD,; Charity Noreuil, MS, Taylor Scott, PhD, Tim Kalafut, PhD, Joel Sutton, MSFS*

Disclaimer

Names of commercial manufacturers or products included are incidental only, and inclusion does not imply endorsement by the authors, NCBI or NIH

We are not endorsing, and have no financial interest in any commercial products that mentioned in this talk.

- Osiris was initiated in response to recommendations of the KADAP advisory panel brought together by the US Dept. of Justice to develop guidelines and recommendations during the World Trade Center victim identification effort. A prerelease version of Osiris was tested in the Hurricane Katrina victim identification, leading to additional development in collaboration with the State and Federal laboratories and NIST.

OSIRIS

- Rapid STR analysis and identification of artifacts
- Validated as an expert system for CODIS
  - Accepted by NDIS http://www.fbi.gov/about-us/lab/biometric-analysis/codis/codis-and-ndis-fact-sheet

- Source Code is publically available
  - GitHub: https://github.com/NCBITools/

- Designed to rapidly analyze STR profiles and identify artifacts and quality issues
- Download from the Osiris homepage
  http://www.ncbi.nlm.nih.gov/projects/SNP/osiris/

Click or cut and paste links into browser:
http://www.fbi.gov/about-us/lab/biometric-analysis/codis/codis-and-ndis-fact-sheet
https://github.com/NCBITools/

Or scan QR code with your cell phone or tablet.

OSIRIS

- Runs on Windows XP, Windows 7, and Macintosh
- Analyzes file formats: .fsa and .hid
- Data produced by 310, 31xx, 3500, and 37xx Genetic Analyzers
- Not limited by the number of dye colors
- Not limited by number of users on network
- Fast - ~.25 second per sample

Dye color number not limited (up to 99)
At a quarter of a second per sample, the typical 96 well plate analyzes in less than 30 seconds.

How does it work?

Mathematical basis
- Fitting curves to peaks
  - Alleles - Double Gaussian
  - Artifacts – spike, blob, split peak (crater) other signatures.
- Controls:
  - ILS, Ladder – time mapped to bp sizes by cubic spline
- Samples:
  - Ladder to sample comparison for best fit
  - Peak identification, artifact identification
- New quality metrics
- Artifact priority - critical, non-critical

- Osiris uses a unique set of mathematical algorithms for analysis
- Other mathematical signatures for spike, blob, split peak (crater)
- For the ILS and the ladder, Osiris maps the time to bp (using a cubic spline) then compares ladder ILS Peak times to the sample ILS Peak times (also using a cubic spline) to ensure that the best ladder is chosen for the analysis of each sample. This reduces the amount of peak 'shifting' that is seen if an arbitrary ladder or a ladder average is used for analysis.
- This gives us three new metrics to measure analysis quality: Peak "Residual" (measure of the sample peak alignment to the ladder peak), curve "Fit" (peak morphology), fit of the sample ILS to the ladder ILS (new in v. 2.3) (goodness of calls)
  - The quality metric for the Ladder to Sample fit is the maximum of the absolute value of the 2nd derivative of the cubic spline used to map the ladder ILS peak size to the sample ILS peak size. When this metric rises above 150, it is indicative of shifting, and peak calling accuracy falls off rapidly, as indicated by the increasing number of excessive peak Residuals and/or off-ladder alleles

Fitting allele curves

- Allele peaks are fitted to a double Gaussian
- "Fit" quality metric (peak morphology) - correlation between raw data and the double Gaussian curve

Fit ≥ 0.9999
Raw
Analyzed
Time (sec)
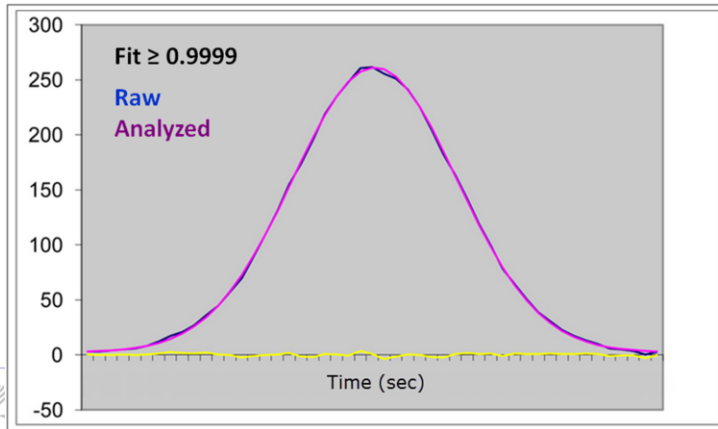
Typical allele curve "fit" is 3 to 4 nines (0.999 - 0.9999). Fit less than 0.98 triggers an artifact notice.

Fitting artifacts

- Signatures for spikes, blobs and split peaks (craters)
- Complex algorithms for
  - Stutter – plus and minus
  - Minus A
  - Pull-up
  - Split peaks (craters)
  - Many other artifacts and conditions to assess quality and make review more efficient

Osiris has mathematical curve signatures for artifacts like spikes, blobs and split peaks and more complex algorithms for plus/minus stutter, -A, and pull-up. Split peak analysis also involves a complex algorithm in addition to mathematical signature. Osiris analyzes many other artifacts and conditions to assess quality and to make review more efficient.

Fitting ILS/ladder/baseline

- Cubic spline fit for
  - Baseline
  - internal marker/ILS peaks
  - Ladder peaks
- Relative spacing of peaks (quality metric)
- Very tolerant of noise and artifacts

- Rather than trying to find the baseline locally, Osiris uses a cubic spline to fit a smooth curve to the baseline
- Cubic spline is a piecewise cubic designed to be smooth connection between points.
- In addition to taking into account the relative spacing of Ladder and ILS peaks, Osiris takes into account the relative peak height differences in ladders and ILS to accurately identify the correct peaks
- This makes Ladder and ILS calling in Osiris very accurate and very tolerant of noise, artifacts and spurious peaks in the ladder and ILS.

Baseline

- Without normalization baseline is taken as the right end of the plot
- Normalization (new) → low analysis threshold
  - Cubic spline fit baseline subtracted from raw data
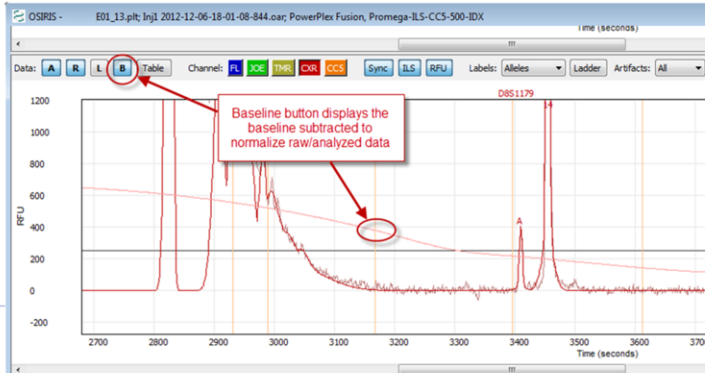
The original raw data baseline that has been fit with a cubic spline (curve that is circled) can be subtracted from the raw data to give a very flat normalized raw data baseline as shown.

Normalization allows accurate allele calling and artifact identification at much lower RFU, important when the lab uses an analysis threshold in the range between 40 and 100 RFU, or any time the raw data baseline rises above the analytical threshold.

Sample peak analysis

- ILS fit
- Peaks fit
- <u>Sample-to-ladder ILS comparison to each ladder to choose best ladder</u>
- Peak assignment
- Artifact identification

First the ILS is fit, then the sample peaks are fit.
The ILS of each ladder in the analysis is compared to the ILS of the sample to choose the best ladder to analyze that specific sample.
Then there is peak assignment and artifact identification.

Rapid identification of artifacts/quality issues

- ILS
- Ladders
- Peak Morphology
- Minus/Plus Stutter, spikes, pull-up/pull-down, split peaks, minus-A
- Verification of: Negative primer peaks, Negative contamination, Kit positive, Custom positive
- Under/over-loaded or -amped
- Homozygote threshold, Het. Peak imbalance

- Osiris tests ILS and ladder peaks' morphology, spacing, height (for choosing correct peaks) and tests for spurious peaks, and the correct number of peaks
- If a peak is misidentified, it causes an artifact notification
- Peak morphology where the Fit (correlation) is less than 0.98 is not acceptable and triggers an artifact notification
- In addition to all the other artifacts that Osiris checks, it verifies:
    - that positive and negative control samples are present in the analysis run,
    - that the negative control sample has primer peaks and is not contaminated, and
    - that kit and lab-specific positive controls have the correct allele values.
- Osiris determines if the sample is under or over-loaded or –amplified
- It checks homozygous loci to determine if the locus could be heterozygous with one allele dropped out
- Checks for heterozygous peak imbalance

Rework prediction

- Depending on failure mode:
  - Dropout
  - Degradation
  - Inhibition
- Suggests:
  - More/Less template,
  - Re-extract/Re-inject
- Help automate re-analysis

Employs user specified parameters to predict reanalysis conditions for various types of failed samples

This data can be fed back into the LIMS system to help automate the flow of the reanalysis process.

User configurable

- Thresholds
  - Analysis, Detection
    - Sample, Ladder, ILS
  - Plus/minus stutter, Minus A
- Parameters for Peak, Sample & Batch metrics
- Multiple process quality options

Osiris is highly configurable.

- Analysis and Detection thresholds can be individually set for samples, ladders and the ILS.
- Plus stutter (e.g., N+4), Minus stutter (e.g., N-4) and Minus A (e.g., N-1) can be individually set for each locus or a default.
- There are also multiple parameters that a lab can set to monitor process quality.

Table view

- Color coding changes with editing
- Graph preview
- Notifications panes

In the Table view, Loci with artifacts are colored red and the color coding changes with editing to give visual cues of completion of the review process

Table view

In the Table view, Loci with artifacts are colored red and the color coding changes with editing to give visual cues of completion of the review process

The center pane is an optional configurable graphic display

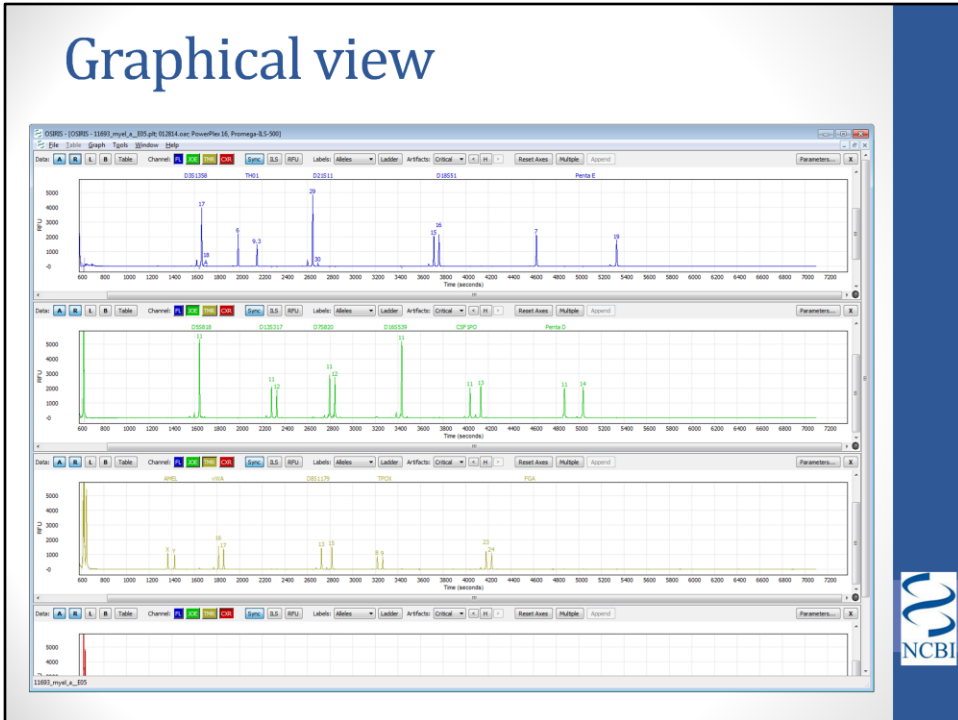The bottom right pane gives information regarding artifact notifications

Graphical View

- Stacked
- Overlayable
- Flexible

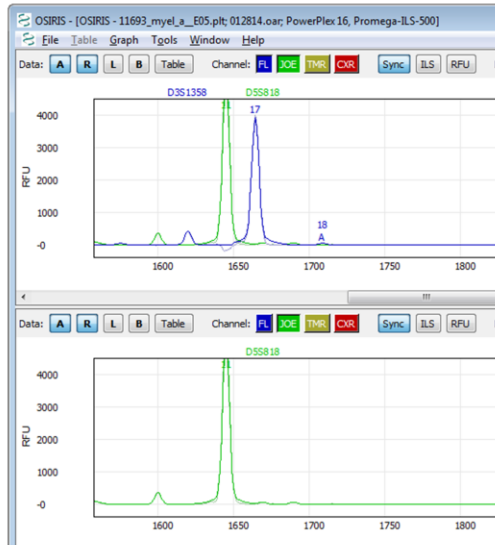The graphical view is a classic stacked electropherogram

Classic Stacked Graph view displays multiple channels

The display is highly configurable.
- User may display Analyzed and/or Raw data,
- Peaks can be labelled with calls, RFU, BP, etc.
- Peak artifacts are labelled with an 'A'.
- Critical or both critical and non-critical artifact labels can be displayed.
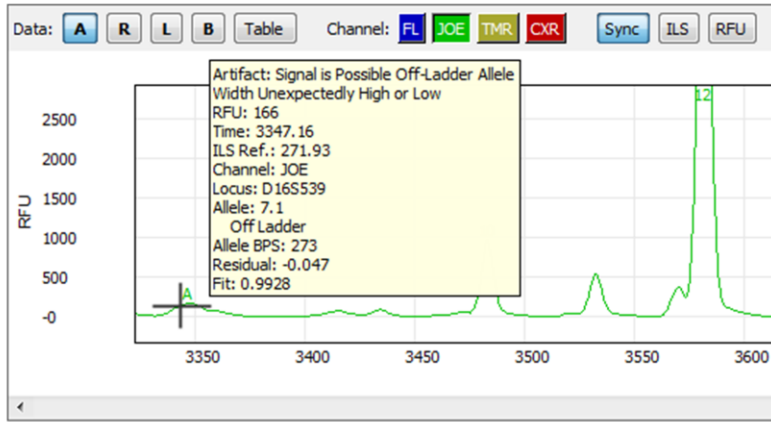
Non-critical artifacts are the artifacts such as stutter, which the user has determined do not require analyst intervention.

You can display multiple colors to verify cross-channel artifacts such as pull-up.

Display peak and artifact info

Hovering the cursor over a peak label to see peak info:
Artifacts, RFU, BP, fit, etc
This peak under the cursor at the bottom left has artifact notifications indicating that it has poor morphology and is off ladder.

Review and Editing

- Flexible review
- 'Accept' artifacts or 'Edit' peaks/artifacts
- Software enforceable

Reviewing and editing allele calls and artifacts is flexible
You can edit calls and artifacts, or simply accept the allele and artifact calls that Osiris made, speeding review
The analysis and review process can be software enforced to prevent departures from the laboratory's SOP.

Analysis and Editing History

- Analysis parameters
- Analysis history
  - Edits
  - Notes
  - Users
  - Can view each set of changes saved

Osiris retains a history of the analysis and review whose changes can be viewed step by step, for each save point.
Including parameters, edits, notes and the users that made them.

Forensic lab use

- Network capable
  - Multi-user, unlimited number of simultaneous users
- Secure analysis parameters to prevent change
- Analysis/Editing/Review history

Large numbers of users running Osiris from a network server have no impact on Osiris performance

Export

- Sample data
- LIMS data
- CODIS data
- Analysis project/Discovery project
- Process quality data
- Multiple formats
  - txt, csv, tsv (tab), XML, *etc.*

Export:
- Allele calls, RFU, and other peak data in various layouts
- LIMS import files
- CODIS upload files
- Analysis projects, including discovery-specific samples
- Data regarding process quality
Data can be in essentially any format (CMF, text, csv, tsv (tab), XML, oer, etc)

Use cases

- CODIS Indexing analysis
  - Accepted by NDIS as expert system
- Casework analysis
  - DNA Profile Analysis
  - DNA Profile QC
- Automating Sample Reanalysis
- Process monitoring
- Training

Osiris can be used in various ways for process monitoring.
For example, Osiris could be used in a high throughput system to automatically check that a batch's controls are good, without analyst intervention.

Take Home

- Fast, flexible STR analysis
- Automatable profile QC
- Artifact identification
- Flexible export scripting
- Validated as an expert system for CODIS samples
- Casework validation

Increase accuracy and reduce the amount of time you spend looking at profiles

Reduce the amount of time you spend looking at profiles and increase accuracy at the same time.

Casework validation talk by Tim Kalafut of USACIL follows this talk at AAFS.

Collaborators on version 2.2

Illinois State Police DNA Indexing Laboratory
Charity Noreuil
Taylor Scott
Rebecca Wheeler
Donald Parker

USACIL
Tim Kalafut
Joel Sutton
Other analysts

NIST
John Butler
Becky Hill

Stanford Bone Marrow Transplant program

Osiris users

Many previous collaborators

In addition to the many people who have contributed ideas, data and time previously, the following were major contributors of ideas in version 2.2 and the upcoming version 2.3.

Thank You!

Download OSIRIS at the homepage:
http://www.ncbi.nlm.nih.gov/projects/SNP/osiris/
(Search the internet for OSIRIS NCBI)

Questions? Suggestions?
forensics@ncbi.nlm.nih.gov

George Riley
George.riley@nih.gov
301-402-8252

We welcome feedback.

Download OSIRIS at:
http://www.ncbi.nlm.nih.gov/projects/SNP/osiris/

Questions? Suggestions?
forensics@ncbi.nlm.nih.gov

george.riley@nih.gov