DNAxs
DNAStatistX

A new software suite for data management and probabilistic interpretation of DNA profiles

NFI Division Biological Traces
History

• NFI experience with development of:
  • Automation solution for the laboratory process
  • Automation solution for the storage of samples
  • Software tools used in DNA case work
    • Bonaparte/Napoleon
    • LOCIM tool
    • LRmix/LRmix Studio
    • SmartRank
    • MixCal
Key projects within NFI strategy

Projects to

- Allow more capacity for more casework and more traces per casework
- Allow a faster workflow and fast answers in the Police investigation process > investigative leads
- Allow a reduction in costs in DNA profiling
- Enhance evidential value by gathering more information from traces by development and implementation of
  - Molecular tools
  - Analytical and Statistical tools
Workflow: Following a case

Reference samples

Client -> Frontdesk -> Planning -> Trace recovery -> DNA lab -> Report by primary RO -> Review by second RO -> Report to client -> Client

R&D

Dutch DNA database

Rework
Support tools in casework interpretation

• Growing number of markers in profiling systems
  • Global STR marker systems available
  • Standard kit the PPF6C kit (27 loci)
• DNA-profile comparison therefor increasingly
  • complex
  • time-consuming
  • error-prone
• Statistical support integrated in casework workflow
What is DNAxs

- NFI developed DNA eXpert System
  - Automatic comparison of sets of DNA-profiles
  - Summary statistics on allele numbers and genotype reproducibility
  - Mixture interpretation
  - Statistical Analysis (March 2019 release)

- In house built (Java)
- Web application (browser)
- Server based
- Validated according to ISO 17025 and NFI standards
- In use since December 2017
- Three releases per year
Functionality of DNAXs

- View profiles
  - Overview of runs and peak heights
  - Bar graphs visualizing alleles/peaks heights/read counts
  - Electropherograms, link to pdf of EPG
- Match profiles
  - Trace vs person
  - Trace vs trace
  - Match matrix
- Derive profiles
  - LoCIM inference of major profile, consensus and composite profile
- Statistics
  - DNAStatistX module
  - Summary statistics (TAC/MAC/type I/II/III loci)
- Supports several NFI profiling workflows (from HVC to complex/severe cases)
- Connectivity to other software LIMS/CODIS/SmartRank/DNAStatistX
- Audit trail
Quality control

- Internal validation according to ISO 17025 standard and internal procedures
  - Validation plans
  - Validation reports
- Q-procedure and software manual
- Internal audit
- External audit
- Integration testing
DNAStatistX

- Based on EuroForMix R code
- As a separate module within DNAxs
- MLE method
- Degradation module included
- Stutter module not included
- Dye Specific detection thresholds
- Tool for number of contributors
DNAStatistX features

- MLE method
- Up to four contributors
- Can handle multiple replicates
- Degradation model
- Model validation

\[ LR = \frac{P(E|Hp)}{P(E|Hd)} = \frac{\sum w(E, Gp|\beta p)P(Gp|Hp)}{\sum w(E, Gd|\beta d)P(Gd|Hd)} \]

- Aim for a maximum run time of 24h for a four-person mixture with three replicates and four unknowns under Hd
From EuroForMix to DNAStatistX: What’s the same?

LR calculation using maximum likelihood estimate (MLE)

User:
- Define hypotheses

Likelihood computation (under Hp and Hd):
1. Estimate parameters (using optimizer, trial and error)
   - Mixture proportions
   - Peak height expectation
   - Peak height variance
   - Degradation slope
2. Determine possible genotype combinations
3. Calculate genotype probabilities (incl. drop-in)
4. Calculate peak height probabilities (incl. drop-in/-out) for each genotype combination
5. Compute profile likelihood

LR calculation:
- Likelihood Hp / likelihood Hd
From EuroForMix to DNAStatistX: What’s different?

- Parts of the EuroForMix code implemented in DNAStatistX
  - Maximum Likelihood Estimate (MLE)

<table>
<thead>
<tr>
<th></th>
<th>EuroForMix</th>
<th>DNAStatistX</th>
</tr>
</thead>
<tbody>
<tr>
<td>Code</td>
<td>R and C++</td>
<td>Java</td>
</tr>
<tr>
<td>Optimizer</td>
<td>nlm</td>
<td>CMA-ES</td>
</tr>
<tr>
<td>Model validation</td>
<td>AdaptIntegrate</td>
<td>Trapezoid Integrator</td>
</tr>
<tr>
<td>Rare allele frequency</td>
<td>Lowest frequency at particular locus</td>
<td>1/(2*size of population)</td>
</tr>
<tr>
<td>Detection threshold</td>
<td>Overall</td>
<td>Dye (locus) specific</td>
</tr>
</tbody>
</table>
Which EuroForMix features in DNAStatistX?

- Degradation model
  - NGM profiles sometimes showed degradation for research samples and often for casework samples
  - All PowerPlex Fusion 6C (PPF6C) profiles showed degradation to some extent

  **Configurable in DNAStatistX, ON by default**

- Stutter model
  - Types of stutter:
    - GeneMapper/GeneMarker etc: -2, -1, -0.5, +0.5, +1 repeat unit
    - EuroForMix: -1 repeat unit
  - Less specific than profile analysis software, very time consuming

  **NOT (YET) in DNAStatistX**
Which EuroForMix features in DNAStatistX?

- Model validation
  - Important quality check: Do observed PHs follow model’s expected PHs

Implementation in DNAStatistX, for every analysis

Example of failed model validation
Developmental validation of DNAStatistX

- **Accuracy:**
  - Comparison to analyses using ground truth parameters
  - Comparison to EuroForMix

- **Precision:**
  - Repeated analyses
  - Optimizer iterations

- **Robustness:**
  - Analyses that should fail

- **Sensitivity:**
  - True positives/ false negatives (Type I errors)

- **Specificity:**
  - True negatives / false positives (Type II errors)

Using a range of case type samples
Collaboration with international partners

- Additional funding for research and development
- Enhance quality of software by incorporating integration testing
- Develop DNAxs in a multi-lab tool for profile comparison/evaluation/interpretation
- Across laboratory validation
- Possibility to disseminate software to other forensic institutes

ISFP-2017-AG-FORENSIC - DNAxs2.0
**Partner choice:** Different languages, different demands
Excellent network (ISFG, EuroForGen, dna.bases)
Casework laboratories of different size
Law enforcement or academic

---

**End-users?**
Forensic DNA caseworkers
High demand for DNAxs-like tools
Partners represent end-users with role in across-Europe validation

---

**Obstacles?**
- **Computer requirements:**
  - Hardware for each partner in project
- **Expert personnel:**
  - Software building expertise within NFI
  - All partners experienced in validations
  - Experienced in organizing workshops
Create guidelines by

- Examining:
  - True positives/negatives - False negatives/positives
  - Effect replicates
  - Effect number of contributors
  - Effect number of drop-outs
  - Etc.

- Defining:
  - Sample types/hypotheses for which LR calculations can be informative
  - LR threshold
  - What to consider when examining results
  - What to do if model validation fails
  - Etc.
Define guidelines for use in forensic DNA casework to:

- Ensure chance of obtaining ‘false-positive’ results is close to zero
- Minimise the number of false-negative results
- Perform LR calculations when regarded useful
- Aim for uniformity among reporting officers
Future functionalities

- MPS data; first module for mtDNA
  - only accessible with mtDNA data
  - Sequential matching
  - Release planned for September 2019
- EMPOP searches
- mtDNA matchbox
- CODIS export
- Followed with STR MPS data
- Under research investigation
  - Stutter model inclusion vs use of laboratory stutter filtered data only
  - Implementation method to estimate number of contributors
  - Deconvolution of all mixed profiles followed by LR computation
## Release history DNAxs

<table>
<thead>
<tr>
<th>Version</th>
<th>Release date</th>
<th>Release theme</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>04 December 2017</td>
<td>Initial release</td>
</tr>
<tr>
<td>1.0.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0.3</td>
<td>24 April 2018</td>
<td>HVC release</td>
</tr>
<tr>
<td>1.0.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1.0</td>
<td>26 July 2018</td>
<td>Tags/notes/Bonaparte export</td>
</tr>
<tr>
<td>1.1.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.2.0</td>
<td>2 April 2019</td>
<td>Initial DNAStatistX release</td>
</tr>
<tr>
<td>1.3.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>