DNAxs/DNAStatistX:
A software suite for the data management and probabilistic interpretation of DNA profiles

Frequently asked questions (FAQ)

Support

- ### How do I get support?
    - For general information on DNAxs, see the website www.forensicinstitute.nl/research-and-innovation/international-projects/dnaxs .
    - For help with installing DNAxs, see the installation manual.
    - For guidance through the DNAxs software, see the manual and instructional videos under the *Help* button within DNAxs.
    - If none of the above answer your question, please contact the DNAxs team at DNAxs@NFI.nl .

About the software

- ### Are DNAxs and DNAStatistX free software?
    DNAxs 2.0 is provided free of charge on request. For more information read *Is the software available to anyone?* & *How can I get DNAxs 2.0?*

- ### Is DNAStatistX open source software?
    DNAStatistX is freely available, but not open source. The MLE module which is used by DNAStatistX is open-source and will be available via GitHub. This means the Java-code of the software will be available. The link is provided through the DNAxs website.

- ## Is DNAxs open source software?

  DNAxs is not open-source software. If you have specific questions about the software code, please contact us through DNAxs@NFI.nl.

- ## Is the software available to anyone?

  The DNAxs2.0 version was supported with a grant from the EU (European Union's Internal Security Fund — Police (Proposal Number: 820838, Proposal Acronym: DNAxs2.0). For this reason, the 2.0 version is freely available. For sustainability reasons upcoming versions and support for the use of the software is only available through a paid licensing structure.

  DNAxs 2.0 is provided free of charge on request. The only condition is that the requesting party signs a license agreement. See *How can I get DNAxs 2.0?* for more information.

- ## DNAxs support service/Will DNAxs be supported continuously?

  The NFI is a governmental institution and as such any support fees only cover costs for the support itself and for future developments of the software. We are in the process of setting up a structure for paid licenses for upcoming versions of DNAxs. If you are interested in upcoming versions and support please send an email to DNAxs@NFI.nl and we will inform you when we have these licenses available.

- ## How can I get DNAxs?

  DNAxs 2.0 can be provided free of charge by sending an application request to the following e-mail address: DNAxs@NFI.nl. For the supported version of DNAxs, requests can also be sent to this email address.

- ## Are there any restrictions to the DNAxs 2.0 free version?

  The free version of DNAxs2.0 is restricted to 4 users (roles can be changed) and 1 administrator account. Furthermore, the software will become inaccessible after the 31st of December 2021. If you wish to increase the user limit of DNAxs please refer to: *DNAxs support service/Will DNAxs be continuously supported*?

- ## What is the difference between DNAxs and DNAStatistX?

  DNAStatistX is a freely available statistical software tool and implements MLE which is an open-source statistical module based on EuroFormix [http://www.euroformix.com/]. The development of DNAStatistX was partly funded by the European Union's Internal Security Fund — Police (Proposal Number: 820838, Proposal Acronym: DNAxs2.0).

  DNAStatistX can be operated by a command line and graphical user interface to calculate likelihood ratios and deconvolve mixed DNA profiles.

  DNAxs is a forensic software suite, which integrates the use of DNAStatistX. DNAxs can initiate DNAStatistX calculations through its premium user interface. The results are sent back to DNAxs for visualization, further analysis and archival purposes.

  DNAxs is not solely developed for the statistical use of DNAStatistX and integrates other software tools and functionalities such as SmartRank searches and automated CODIS database uptake requests.

- ## How many releases will there be on a yearly basis? Do we receive updated versions?

  NFI is committed to future development of DNAxs and will make new versions available through its licensing structure. Since the launch of DNAxs, 2 to 3 releases per year have been made. For sustainability reasons upcoming versions and support for the use of the software is only available through a paid licensing structure. For more information, please refer to: *DNAxs support service/Will DNAxs be continuously supported?*

- ## Will the newly developed reporting tool be available to users outside the NFI?

  The reporting tool, which is currently being developed, will be modular and will be part of future releases of DNAxs.

Software installation

- ### What are the computer system requirements for the software?

  DNAxs is able to run on any system that is able to run the latest version of Java 8 or higher in combination with one of the following internet browsers: Internet Explorer (version 9 or higher), Edge, Chrome and Firefox. For the optimal experience it is recommended to run DNAxs on Windows 7 or higher with Google Chrome as internet browser.

  DNAStatistX is able to run on any system which supports the latest version of Java 8 or higher.

- ### Where should I install DNAxs on my computer?

  In a folder where you have full access, for example your user home. Preferably not in the 'Program Files'. (For more information, see *I have installed DNAxs, but it does not start after clicking the 'Launch DNAxs suite'*.)

- ### I have installed DNAxs, but it does not start after clicking the 'Launch DNAxs suite'

  This problem usually occurs when DNAxs is installed in 'Program Files'. This could cause a problem as this is usually a location where admin rights are required. Try uninstalling and reinstalling DNAxs in another folder, for example your user home.

- ### Would it be possible to connect DNAxs and/or DNAStatistX to other systems?

  Both DNAxs and DNAStatistX are designed to be modular. As such it is possible to connect them to any software which accepts messages from JSON based API's.

- ### How are the different systems (i.e. DNAxs and LIMS) communicating? Is it direct communication or communication using intermediate files?

  DNAxs communicates between APIs with network URLs. These URLs don't require an internet connection to communicate locally.

- **What kind of integration is possible, REST or SOAP?**
  We use REST connectivity. As DNAxs is built in a modular fashion, technically any application can be connected.

- **How is DNAxs tested before a release?**
  Each release is both manually tested and automatically tested through integration tests. Integrations tests are run every time a new feature is added to DNAxs. New methodologies are validated according to the ISO 17025 standard.

- **Can I install DNAxs on any computer or server?**
  DNAxs can run both locally and on a Windows or Linux server that supports Java 8 or higher. Generally, it is recommended to run DNAxs on a server with a dedicated calculation cluster. DNAxs comes bundled with a custom cluster manager which manages the queuing of multiple calculations requested by multiple users at once.

- **Which database engine is used?**
  DNAStatistX receives its input and sends its output through text files and as such no server or database is used for standalone DNAStatistX.

  DNAxs is compatible with PostgreSQL and SQLite out of the box. For DNAxs 2.0 SQLite is used locally and PostgreSQL is used when DNAxs is run on a server. However, on request any type of database can be easily integrated, for more information, see DNAxs support service.

- **Is it possible to create a back-up when you generate a lot of data?**
  Yes, it is possible, this is something that should be arranged within your own IT infrastructure. DNAxs does not automatically generate back-ups.

- **For what purpose is the kit_properties.xml used?**
  DNAStatistX gets all the kit data for a calculation from the kit_properties.xml file. If you want to change locus-specific parameters, such as detection thresholds and drop-in probabilities or if you want to add a kit or add a new CE setting, you can edit this file. More information can be found in the installation manual.

Type of data

- **What type of data can be loaded into DNAxs?**
  Currently DNAxs can handle CE-STR data.

- **What data formats are accepted by DNAxs?**
  DNAxs can import DNA profiles in CSV format. The file must have the following column structure: Sample Name, Marker, Allele 1, Allele 2, Allele n, Height 1, Height 2, Height n. The column separator is automatically detected. The Allele and Height columns can also be set alternately. The height columns are optional.

  To test whether your data format is suitable for DNAxs, we have created a Profile Read Tester, which will be provided with the software.

- **What data is used by DNAxs, raw or stutter filtered data?**
  Stutter filtered data is used by DNAxs, as it does not include a stutter filter or stutter model.

DNAxs user questions

- **Why would you change the CE/PCR settings of your data in DNAxs?**
  If no information is provided on CE/PCR settings and on values that apply to this data (e.g. detection threshold), default values will be used in e.g. DNAStatistX LR calculations. If other settings apply to your data, they can be changed manually.

- **Why does the NOC estimation tool estimate one person in my sample where there are three alleles on a locus? This is contradictory to the 'Complexity' in the profile info.**

  The NOC tool is based on machine learning. The generic model uses only those markers that are available in most of the commercial STR kits (FGA, TH01, VWA, D1S1656, D2S441, D3S1358, D8S1179, D10S1248, D12S391, D18S51, D21S11, D22S1045). As the generic model was trained on those markers only, other markers (that may exhibit more alleles) are excluded from the estimation. Within the NFI, the generic model was tested on PowerPlex Fusion 6C, GlobalFiler and NGMSelect (Express) data and showed improved results when compared to the maximum allele count approach. For more information please read the following paper: C.C.G. Benschop, J. van der Linden, J. Hoogenboom, R. Ypma, H. Haned. Automated estimation of the number of contributors in autosomal short tandem repeat profiles using a machine learning approach. Forensic Sci. Int. Genet. 43 (2019) 102150.

- **The profile qualities 'complete', 'sufficient', 'insufficient': is this automated or does the user need to fill this in?**

  This is automated. The profile is denoted 'complete' when it contains at least one allele on all loci of its kit. 'Sufficient' is reached when the profile contains at least one allele on all SGM+ loci. The quality of the profile will be 'insufficient' when it does not meet the requirements of 'complete' or 'sufficient'. The settings for quality can be changed manually.

- **Can I combine data from different kits (e.g. sample with runs by different kits)?**

  In Profile Comparison, data from different kits are combined for the overlapping loci. Data of different kits cannot be combined in the LR calculations.

DNAStatistX

- **Is DNAxs required for running DNAStatistX, or can DNAStatistX be run as stand-alone software?**
  DNAStatistX can be used both stand-alone and in combination with DNAxs. The stand-alone version of DNAStatistX can be used with either the command line or a user interface. DNAxs both incorporates and expands upon DNAStatistX. For more information on this subject see: _What is the difference between DNAxs and DNAStatistX?_.

- **Do you have guidelines for using DNAStatistX?**
  Summarized guidelines based on internal validation using PowerPlex Fusion 6C data is provided in FSIG 42 (2019) 81-89. These guidelines are not intended as strict rules but as a helpful tool in the decision-making process. These guidelines may inspire others to create their own internal guidelines.

- **How should I use the degradation model?**
  By default, the degradation model is turned on in DNAStatistX. Validation data showed that most often the data could be better explained with the degradation model turned on. With the degradation model turned off, the degradation model will have no effect on the expected peak heights. To see which model fit best for your data, run both and compare the likelihoods obtained using both methods. The largest likelihood is the best explanation. For guidance on model selection and explanation of the Akaike Information Criterion see also: Ø. Bleka, C.C.G. Benschop, G. Storvik, P. Gill. A comparative study of qualitative and quantitative models used to interpret complex STR DNA profiles. Forensic Sci. Int. Genet. 25 (2016) 85-96.

- **Does DNAStatistX have an option for non-contributor tests?**
  DNAStatistX does not have an option for non-contributor tests.

- **Is it possible to replace unknown under Hd (H2) with a relative of the POI?**
  Not yet, this feature is added to the development list and will be part of a future release.

Validation

- **What types of analyses would you suggest to validate the software for use in my own laboratory?**
  Internal validation can be described as: the accumulation of test data within the laboratory to demonstrate that established parameters, software settings, formulas, algorithms and mathematical functions perform as expected; and that the information/results/data obtained is correct and consistent with expected values. To that end, the user can:
  - Run its own laboratory specific dataset to cover the range of profile types as can be encountered in forensic DNA casework.
  - Run the dataset as provided by the NFI, to get familiarized with the DNAxs software and to compare the DNAStatistX results obtained by other laboratories. This dataset will be made available through publication.
  - For guidelines on (internal) validation of software we suggest reading, *e.g.*:
    - UK Forensic Science Regulator, 31 July 2018, Software validation for DNA mixture interpretation, FSR-G-223, Issue 1,
      https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/730994/G223_Mixture_software_validation_Issue1.pdf.
    - J.A. Bright, I.W. Evett, D. Taylor, J.M. Curran, J. Buckleton, A series of recommended tests when validating probabilistic DNA profile interpretation software, Forensic Sci. Int. Genet. 14 (2015) 125–131.
    - SWGDAM, Guidelines for the Validation of Probabilistic Genotyping Systems, (2015).
    - H. Haned, P. Gill, K. Lohmueller, K. Inman, N. Rudin, Validation of probabilistic genotyping software for use in

forensic DNA casework: definitions and illustrations, Sci. Justice 56 (2016) 104–108.

- M.D. Coble, J. Buckleton, J.M. Butler, T. Egeland, R. Fimmers, P. Gill, L. Gusmão, B. Guttman, M. Krawczak, N. Morling, W. Parson, N. Pinto, P.M. Schneider, S.T. Sherry, S. Willuweit, M. Prinz, DNA Commission of the International Society for Forensic Genetics: recommendations on the validation of software programs performing biostatistical calculations for forensic genetics applications, Forensic Sci. Int. Genet. 25 (2016) 191–197.

- ENFSI (2017) Best Practice Manual for the Internal Validation of Probabilistic Software to Undertake DNA Mixture Interpretation, (2019) QCC-BPM-003 issue 001.

Background information:

- ## Are there publications related to the software and its features?

  Yes, see:

  o C.C.G. Benschop, J. Hoogenboom, P. Hovers, M. Slagter, D. Kruise, R. Parag, K. Steensma, K. Slooten, J.H.A. Nagel, P. Dieltjes, V. van Marion, H. van Paassen, J. de Jong, C. Creeten, T. Sijen, A.L.J. Kneppers. DNAxs/DNAStatistX: Development and validation of a software suite for the data management and probabilistic interpretation of DNA profiles. Forensic Sci. Int. Genet. 42 (2019) 81-89.

  o C.C.G. Benschop, A. Nijveld, F.E. Duijs, T. Sijen, An assessment of the performance of the probabilistic genotyping software EuroForMix: trends in likelihood ratios and analysis of Type I & II errors, Forensic Sci. Int. Genet. 42 (2019) 31–38.

  o C.C.G. Benschop, J. van der Linden, J. Hoogenboom, R. Ypma, H. Haned. Automated estimation of the number of contributors in autosomal short tandem repeat profiles using a machine learning approach. Forensic Sci. Int. Genet. 43 (2019) 102150.

  o C.C.G. Benschop, J. de Jong, L. van de Merwe, H. Haned, Adapting a likelihood ratio model to enable searching DNA databases with complex STR DNA profiles, Proceedings of the 27th ISHI, Promega.com.

  o C.C.G. Benschop, L. van de Merwe, J. de Jong, V. Vanvooren, M. Kempenaers, C. van der Beek, F. Barnie, E. López Reyes, L. Moulin, L. Pene, P. Gill, H. Haned, T. Sijen, SmartRank: a likelihood ratio

software for searching national DNA databases with complex DNA profiles, Forensic Sci. Int. Genet. 29 (2017) 145–153.

o Ø. Bleka, G. Storvik, P. Gill, EuroForMix: an open source software based on a continuous model to evaluate STR DNA profiles from a mixture of contributors with artefacts, Forensic Sci. Int. Genet. 21 (2016) 35–44.

o C.C.G. Benschop, T. Sijen, LoCIM-tool: an expert's assistant for inferring the major contributor's alleles in mixed consensus DNA profiles, Forensic Sci. Int. Genet. 11 (2014) 154–165.