Bayesian networks for DNA-based kinship analysis
Functionality and validation of GENis missing person identification module

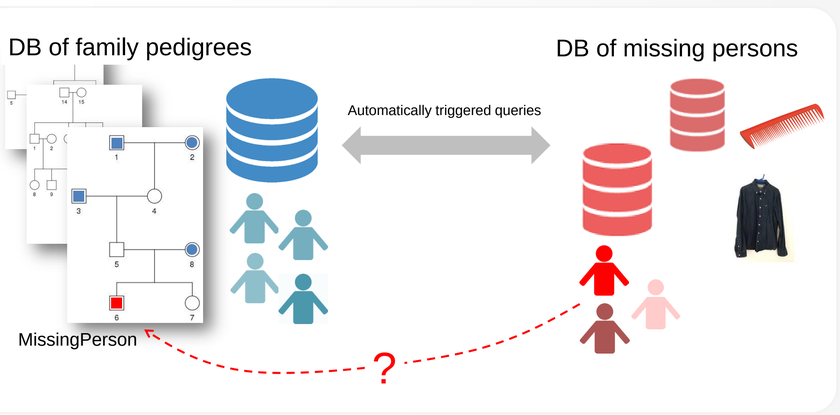
Ariel Chernomoretz*[1,2],F.Marsico[3], J.Iserte[1], M.Herrera Pinero[4], M.S.Escobar[5], M.Balparda[5], G.Sibilla[5]

* ariel@df.uba.ar , achernomoretz@leloir.org.ar / [1]Phys. Department, School of Sciences, University of Buenos Aires/IFIBA CONICET Argentina, [1]Leloir Institute Foundation, [3]UNPAZ, [4]Banco Nacional de Datos Geneticos,[5]Fundación Sadosky

Abstract

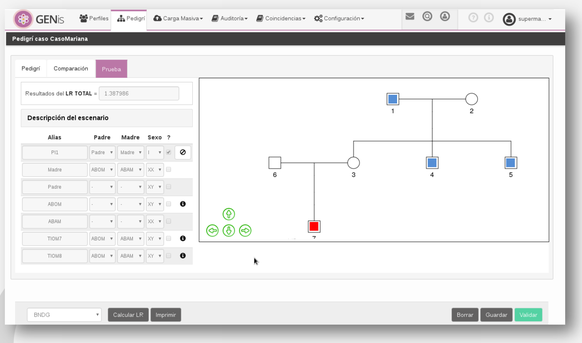
GENis is a recently published open-source multi-tier information system developed to run forensic DNA databases. It relies on a Bayesian Networks framework and it is particularly well suited to efficiently perform large-size queries against databases of missing individuals. In this contribution we present a validation of the missing person identification capabilities of **GENis**. To that end we introduce **fbnet**, a free-software package written in the R statistical language that implements the complete **GENis** functionality to perform kinship analysis based on DNA profiles. With the aid of **fbnet**, we could validate likelihood ratios against estimations draw with **Familias** and **forrel** (two well-recognized R packages for kinship quantification) for complex pedigrees provided by the Argentinian reference databank (Banco Nacional de Datos Geneticos, BNDG).

GENis & fbnet

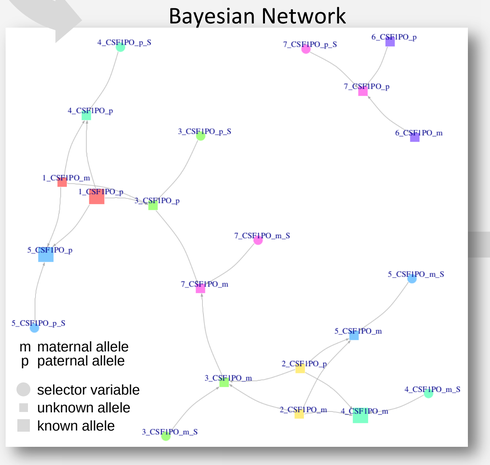


GENis is a highly customizable system composed of three different modules related to: (a) person identification and analysis of forensic evidence, (b) missing person identification, and (c) disaster victim identification[1]. The missing person identification (MPI) module was specifically developed to run automatic queries on family and missing person databases.

For each family, the pedigree structure and available genotypes are integrated into a bayesian network[2] used to infer genotype probability tables for the queried missing person (MP). The availability of these probability tables allows for a rapid calculation of likelihoods for large MP databases.



Bayesian Network



We implemented the complete GENis functionality to perform kinship analysis in **fbnet**[3], an open-source package written in R statistical language.

```
> library(fbnet)
> pbn  <- initBN(ped)
> bnet <- buildBN(pbn,MP=7)
> bnet  <- buildCPTs(bnet)
> resQ <- velim.bn(bnet)
> lProbGeno <- genotypeProbTable(resQ)
```

Get data from *paramlink* pedigree

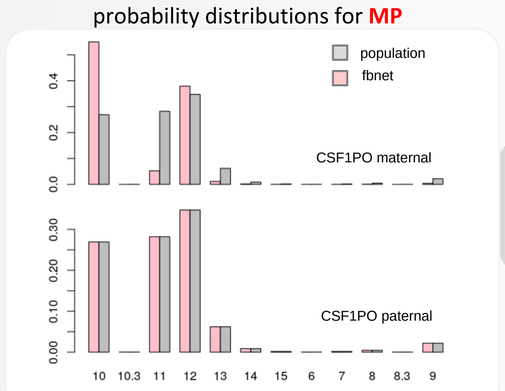
Build the bayesian network associated to pedigree

Compute Conditional Probability Tables

Compute MP allelic probabilities conditioned by evidence

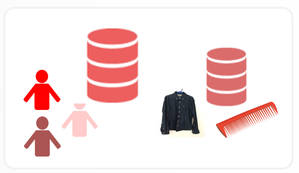
MP genotype probabilities P(MP|ev,π)

Paternal and maternal allelic probability distributions for **MP**



LR estimation

```
> resLR<-reportLR(bn1,resQ,geno)
```



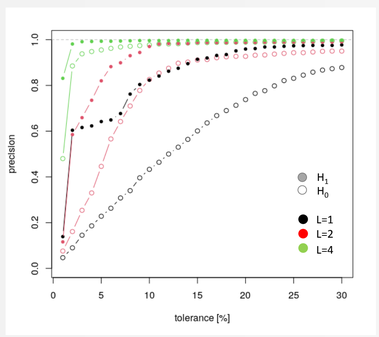
		L_0	L_1	LR
CSF1P0		1.42e-13	4.66e-14	0.327
D1		L_0	L_1	LR
…	CSF1P0	3.12e-13	4.56e-12	14.6
	D16S539	1.21e-12	2.10e-12	1.73
⋱	…	…	…	…

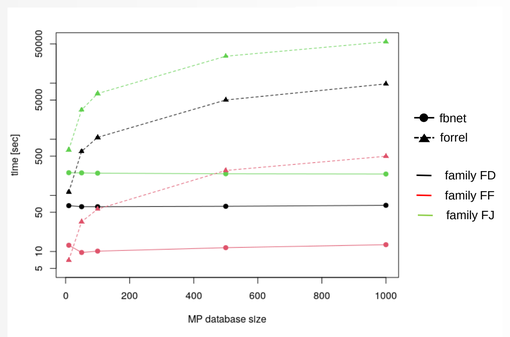
Validation Exercise

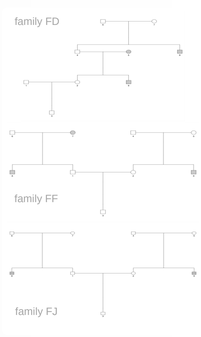
To speed-up calculations **GENis** and **fbnet** implement a truncated version of the canonical stepwise model where probabilities for rare mutations, i.e. more than a given number (*L*) of steps away from the original allelic value (i.e. the diagonal term), are neglected.

We estimated with **fbnet** LR values for 1000 unidentified persons (UPs) simulated profiles (15 markers) considering H1: UP is MP, and 1000 UPs considering H0: UP is not MP. We used L=1, 2 and 4 models. In Panel A we show the precision of these estimations, at a given tolerance level, compared against Familia's LR values for familyFF (Panel C).

LR estimations for large databases can be done very efficiently as the MP genotype probabilities, conditioned by the available evidence and pedigree relationships, are estimated only once for each family. Computational time for **fbnet** LR estimations for the ensemble of simulated profiles are compared against **forrel** running times in Panel B (cpu: ryzen 5 2600x 3.6Ghz, 32 gb ram)





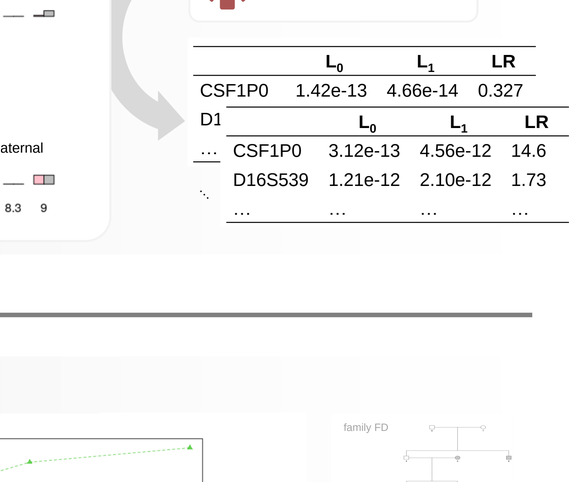


Systematic comparisons of LR values

We considered 24 familial pedigrees from BNDG and randomly generated the genotype (23 molecular markers) of available family contributors. For each pedigree we conditionally simulated 1000 MP genotypes and compared the obtained LR values against **forrel** estimations for both, equal and stepwise mutational models (TSMM with L=4).

	#members	#genotyped	Corr Log[LR]	mean error [%]	max error [%]	# dev>1%	#dev>2%	#dev>5%
F1	12	5	0.99999269	0.17	74.50	10	2	1
F1*	9	3	0.99999794	0.08	14.42	5	3	2
F2	16	9	0.99999994	0.06	2.61	5	1	0
F2*	12	6	0.99999991	0.06	1.86	5	0	0
F3	9	5	0.99999988	0.09	6.42	6	4	1
F3*	9	5	0.99999932	0.07	13.18	6	2	1
F4*	6	3	0.99999981	0.04	5.43	2	2	1
F5	9	5	0.99999998	0.04	1.79	3	0	0
F5*	9	5	0.9999999	0.05	2.33	4	1	0
F6	9	3	0.99999963	0.11	9.80	14	5	2
F6*	6	2	0.99999983	0.07	2.38	6	2	0
F7	14	8	0.99999983	0.10	4.68	7	4	0
F7*	11	6	0.99999974	0.15	2.03	4	1	0
F8	6	2	0.99999991	0.06	1.51	4	0	0
F9	7	5	1	0.02	0.04	0	0	0
F10	11	5	0.99999959	0.11	10.38	5	2	2
F10*	8	3	0.99999859	0.09	20.73	7	3	2
F11	13	6	0.99999989	0.12	5.95	2	2	1
F11*	8	3	0.99999996	0.07	1.42	3	0	0
F12	5	1	0.99999951	0.08	5.82	8	5	2
F13	11	5	0.99999983	0.09	4.69	5	4	0
F13*	6	2	0.99999983	0.05	4.49	2	2	0
F14	16	8	0.99999997	0.07	1.80	2	0	0
F15*	17	8	0.99999968	0.11	9.04	4	1	1

*paternal branch excluded



Conclusions

We found an excellent agreement between our LR estimations and the corresponding reference values. Moreover, the bayesian network approach to the kinship analysis problem provided an extremely efficient methodology to evaluate LRs for large MP databases. Noticeably, the functionality implemented in the open-source package **fbnet** allows us to share with the community the functionality and main design principles behind GENis MPI module

[1] "GENis, an open-source multi-tier forensic DNA information system", Chernomoretz *et al*, Forensic Science International:Reports (2020) http://doi.org/10.1016/j.fsir.2020.100132
[2] "Modeling and reasoning with bayesian networks", Darwiche A., Cambridge University Press (2009).
[3] fbnet R package can be downloaded from : https://CRAN.R-project.org/package=fbnet