

Interpretace profilu DNA a Baysovská statistika

Člověk, pravděpodobnost a Spravedlnost v Science Café Praha

<https://slideslive.com/38899791/clovek-pravdepodobnost-a-spravedlnost>



Statistika a STR profilování

STR lokus

Mám typický multialelický autozomální STR lokus „A“

- známo je celkem 9 alel (1-9)
- každá z těchto alel má nějakou frekvenci ve studované populaci

1	2	3	4	5	6	7	8	9
0,02	0,08	0,11	0,24	0,09	0,21	0,15	0,09	0,01

Jak ale vlastně zjistím frekvenci alel v populaci?

→ musím udělat populační studii

Populační studie

„Vezmi reprezentativní vzorek dostatečně velký, aby tvá data byla relevantní, a genotypuj ho.“

Co to je „REPREZENTATIVNÍ VZOREK (VÝBĚR)“?

→ jen sladké mámení... nikdy nemáme plně reprezentativní vzorek dané populace, můžeme se mu jen blížit

Jaké budou hlavní požadavky na reprezentativnost?

Co to je DOSTATEČNĚ VELKÝ“ ?

➤ záleží na nás, s jakou přesností chceme získat údaj o alelických frekvencích - čím větší vzorek, tím přesnější data (tzn. menší SD)

Lze dosáhnout 100% přesnosti?

Populační studie

každou osobu z výběru genotypuji – tzn. stanovím sestavu jejích alel
stanovím četnost výskytu každé alely

- POZOR! – heterozygot 2/3 – započítávám 1x alelu 2 a 1x alelu 3
homozygot 3/3 – započítávám 2x alelu 3

pak stanovím frekvence jednotlivých alel – pro alelu X platí:

$$P(X) = n(X)/N$$

1	2	3	4	5	6	7	8	9
8	32	44	48	36	42	60	36	4
400	400	400	400	400	400	400	400	400
0,02	0,08	0,11	0,24	0,09	0,21	0,15	0,09	0,01

Populační studie

...tolik stanovení frekvence alel. Toto mohu udělat u jakéhokoli polymorfismu, mohu to udělat u libovolného genetického markeru, biochemického markeru, morfologického markeru, mohu to udělat i pro polymorfismus barvy auta atd.

Ale to je jen začátek!

Už známe frekvence, ale teď musíme ověřit, jak se alely chovají, abychom mohli dále hodnotit

Co je Svatým Grálem populační genetiky?

HW - equilibrium

HW - equilibrium

Co jsou základní předpoklady HWE?

- nekonečná velikost populace
 - žádné mutace
 - žádná selekce
- náhodný pohlavní výběr

HW - equilibrium

NEKONEČNÁ VELIKOST POPULACE

Nekonečnost populace je nezbytným předpokladem, který eliminuje náhodné změny frekvencí alel

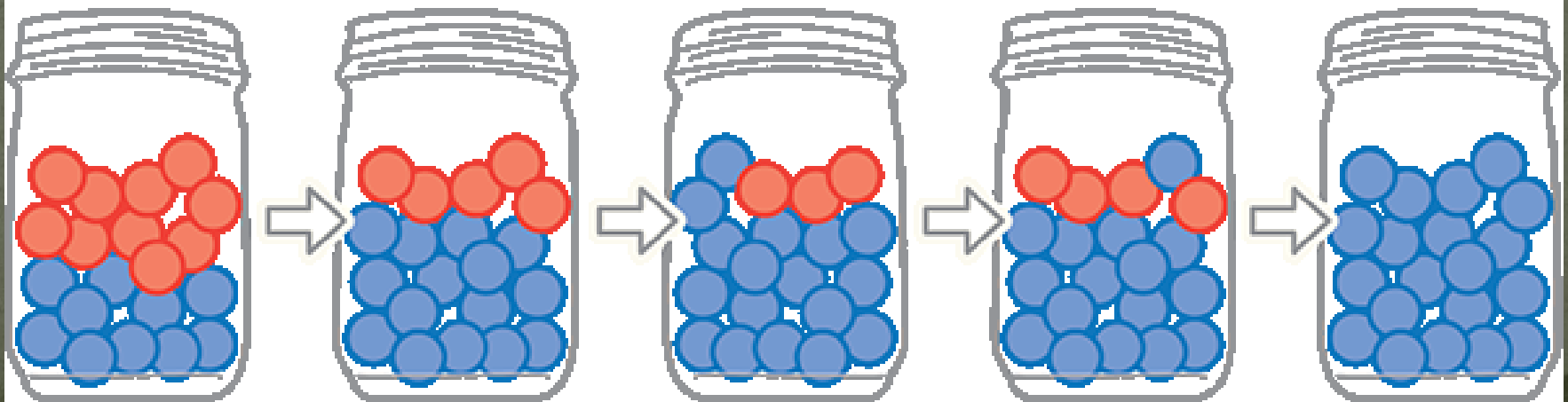
Jak takovým náhodným změnám říkáme?

genetický drift

Každá populace je konečná = je tvořena konečným počtem jedinců, nicméně má-li alespoň rozumnou velikost, je vliv driftu zanedbatelný a HWE poskytuje dobré odhady

Genetický drift

- Proces, kdy dochází k náhodným změnám (posunu = driftu) ve frekvencích alel v dané populaci.
- Není zapříčiněn selekčními tlaky.
- V konečně velkých populacích může genetický drift vést k tzv. fixaci jedné z variantních alel (její frekvence dosáhne 100 %), popř. její eliminaci (frekvence 0 %).



HW - equilibrium

NEEXISTENCE MUTACÍ

Neexistence mutací je dalším nezbytným předpokladem, který eliminuje náhodné změny frekvencí alel

Mohou existovat polymorfní STR lokusy, které nemutují a nikdy nemutovaly?

Opět platí, že dochází-li k mutacím pouze v rozumné míře, je jejich vliv zanedbatelný a HWE poskytuje dobré odhady

HW - equilibrium

NULOVÝ SELEKČNÍ TLAK

Neexistence selekce je nezbytným předpokladem, který eliminuje nenáhodné změny frekvencí alel

Jednotlivé alely samy o sobě nesmějí přinášet nějakou selekční výhodu respektive nevýhodu

Existují STR lokusy, které selekční (ne)výhodu přinášejí?

Pokud samy alely nejsou předmětem selekce, nesmí se tímto předmětem selekce stát ani nepřímo – nesmí existovat genová vazba mezi daným STR lokusem a nějakým funkčním genem, který je selektován

Jak nazýváme tuto „nepřímou selekci“?

genetický draft = selekce na pozadí

Genetický draft

- zvýšení frekvence alely díky genetické vazbě s výhodnou mutací
- selekční vymetení („selective sweep“) = snížení genetické variability v okolí výhodné mutace
- zvyšuje míru vazebné nerovnováhy v populaci

Selekce na pozadí

- snížení frekvence alel díky genetické vazbě s nevýhodnou mutací
- také vede ke snížení genetické variability v populaci, není však tak výrazné, jako u selekčního vymetení.

HW - equilibrium

Multialelický autozomální STR lokus - několik jednoduchých úloh:

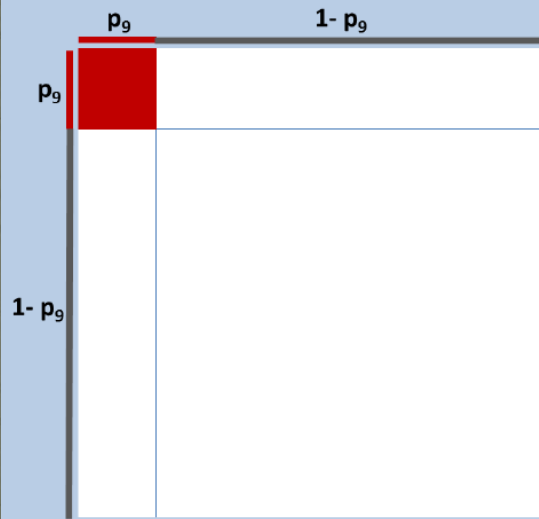
1	2	3	4	5	6	7	8	9
0,02	0,08	0,11	0,24	0,09	0,21	0,15	0,09	0,01

1. kolik % je v populaci osob s genotypem 9/9 ?
2. kolik % osob má alespoň jednu alelu 4 ?
3. kolik % osob má právě jednu alelu 8 ?
4. kolik % spermií v celé populaci nese alelu 7 ?

HW - equilibrium

1. kolik % je v populaci osob s genotypem 9/9 ?

1	2	3	4	5	6	7	8	9
0,02	0,08	0,11	0,24	0,09	0,21	0,15	0,09	0,01



$$P = p_9 \times p_9 = p_9^2$$

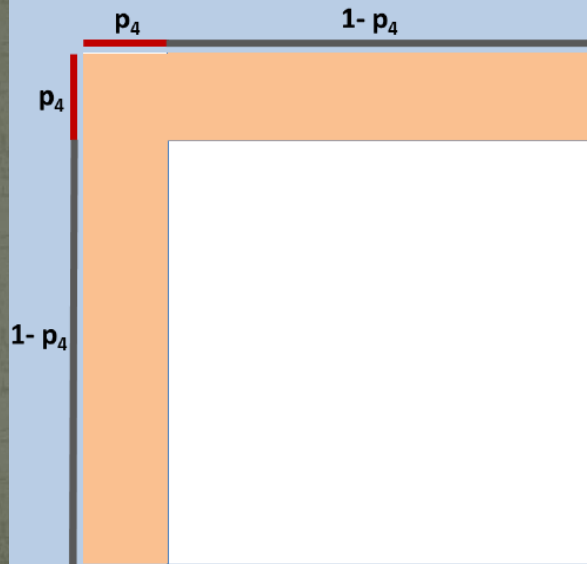
$$P = 0,01 \times 0,01 = \mathbf{0,0001}$$

0,01%

HW - equilibrium

2. kolik % osob má alespoň jednu alelu 4 ?

1	2	3	4	5	6	7	8	9
0,02	0,08	0,11	0,24	0,09	0,21	0,15	0,09	0,01



$$P = (p_4 \times p_4) + 2(p_4 \times 1-p_4)$$

$$P = 2p_4 - (p_4 \times p_4)$$

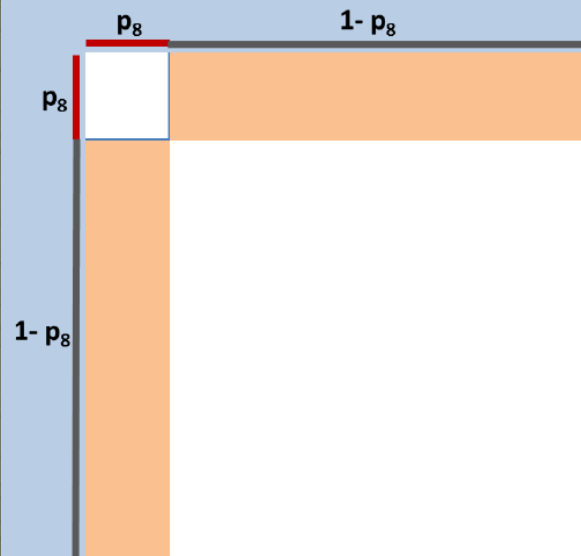
$$P = 1 - (1 - p_4)^2$$

$$P = 0,4224 \quad 42,24\%$$

HW - equilibrium

3. kolik % osob má právě jednu alelu 8 ?

1	2	3	4	5	6	7	8	9
0,02	0,08	0,11	0,24	0,09	0,21	0,15	0,09	0,01



$$P = 2 \times (1-p_8) \times p_8 = 2(p_8 - p_8^2)$$

$$P = 0,1638$$

16,38%

HW - equilibrium

4. kolik % spermií v celé populaci nese alelu 7 ?

1	2	3	4	5	6	7	8	9
0,02	0,08	0,11	0,24	0,09	0,21	0,15	0,09	0,01

$$P = p_7 = 0,15 \quad 15\%$$

HW - equilibrium

OVĚŘENÍ

mám lokus A s alelami **1** ($p_1=0,1$), **2** ($p_2=0,3$) a **3** ($p_3=0,6$)

genotyp	1/1	2/2	3/3	1/2	1/3	2/3
Očekávaná četnost	0,01	0,09	0,36	0,06	0,12	0,36
Pozorovaná četnost	0,02	0,06	0,39	0,06	0,13	0,34
rozdíl	0,01	0,03	0,03	0	0,01	0,02

musím posoudit, zda tyto rozdíly mezi očekávaným a pozorovaným mohou být pouze dílem náhody, či zda jsou signifikantní a ukazují, že HWE neplatí

jak? - χ^2

HW - equilibrium

Důsledky pro forenzní výpočty

pokud zjistím, že pro lokus HWE neplatí, nemohu počítat frekvence jednotlivých genotypů prostou kombinatorikou; mohu

- a) užívat výrazně komplikovanější výpočty s koeficienty**
- b) vyloučit lokus z testování**

pokud zjistím, že pro lokus HWE platí, pracuji při hodnocení důkazů s frekvencemi tak, jak je počítá HWE

Proč používat statistiku ve forenzní genetice ?

- Ve forenzní genetice nestačí pouhé konstatování shody či neshody DNA profilů.
- Potřebujeme zjistit, jak silným důkazem jsou námi zjištěné výsledky (shoda v 6 lokusech x 24 lokusů).
- Proto používáme statistiku, abychom zjistili, jaká je pravděpodobnost, že výsledky svědčí pro jednu z hypotéz pouze shodou náhod.

Základní zákony pravděpodobnosti

- $\Pr(A)$ – pravděpodobnost jevu A
- $\Pr(A|E)$ – pravděpodobnost, že tvrzení A je pravdivé, za předpokladu, že známe E (důkaz – z angl. evidence)
- Obecně je $\Pr(\text{Tvrzení}|\text{Informace})$

Forenzní důkaz

IDEÁLNÍ

- Má znaky, které jsou unikátní pro individuální osobu
- Tyto znaky se nemění v čase
- Znaky jsou jednoznačně určitelné na různých místech různými experty
- Umožňuje potvrdit přítomnost osoby na místě činu; vyskytuje se vždy, když hypotéza platí, a zároveň se nevyskytuje nikdy jindy
- Má jednoduché a levné zjištění hodnoty znaku

TYPICKÝ

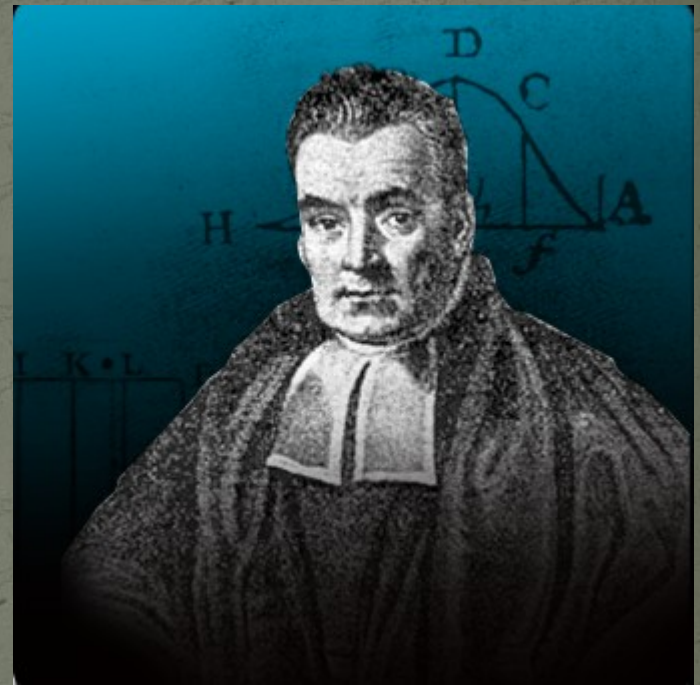
- Znak je přítomen, i když hypotéza není pravdivá (falešná pozitivita testu)
- Znak není přítomen, i když je hypotéza pravdivá (falešná negativita testu)
- Je pravděpodobnější, že se znak vyskytuje, pokud je hypotéza pravdivá

Interferenční logiky

- není doménou vyhrazenou pro znalce, mohou a měli by ji používat všichni účastníci soudního řízení
- znalec nemůže vyvodit závěr (např. krevní skvrnu zanechala určitá osoba, dítě bylo zneužito) na základě jediného důkazu
- vědecký důkaz by měl být zkombinován s dalšími důkazy k případu. Nejlepším způsobem, jak to provést, je použití věrohodnostního poměru, který může být zkombinován s dalšími důkazy prostým vynásobením
- důkaz má váhu jen v takovém kontextu, kdy pomáhá rozlišit mezi hypotézami
- problémy s vědeckými důkazy vznikají častěji při interpretaci než z experimentálních chyb

Bayesovská inference

- usuzování postaveno na baysevorském teorému
 - úprava pravděpodobnostních očekávání
ve světle nových informací
- postavení si dvou hypotéz H_1 a H_2
- H_n – pravděpodobnost důkazu,
- resp. jak bylo pravděpodobné,
když uvidím důkaz, když platí ta či
ta hypotéza



- T. Bayes (1702 - 1761)-
anglický duchovní

Apriorní pravděpodobnost (prior)

- to, co víme nebo si myslíme předem, ještě před získáním dat (např. výsledky předchozích experimentů)
- problém nastává, když předem nevíme nic

Věrohodnost (likelihood)

- pravděpodobnost, s jakou bychom pozorovali právě ta data, která skutečně pozorujeme, kdyby naše hypotéza byla správná
- jde o opak toho, co doopravdy chceme znát

Aposteriorní pravděpodobnost (posterior)

- výsledek celého snažení: pravděpodobnost naší hypotézy v závislosti na předchozích znalostech (prior) a současně nových datech

Mám profil DNA ... a co dál?

profil je x není vhodný ke vložení/porovnání

CODIS – vložení, porovnání s profily DNA uloženými v
databázi

profil DNA jedné osoby

profil DNA dvou osob – majorita

směsné profily dvou osob bez majority-

Popstats

paternity (O M D, R D)

správné nastavení hypotézy, co chci zjistit

Profil jedné osoby

- vložím
- prohledám
- shoda – je x není
- vyřídím písemně oznámení o nalezené shodě
- je uveden prověřovaný, podezřelý, obviněný – porovnáám s ním

Směs dvou osob - majorita

- musí splňovat určitá kritéria daná pokyny PP nebo Ř KÚP
- dle kvality – mezinárodní výměna nebo jen v rámci ČR
- opakovaný profil DNA – dedukovaný major lze vložit

Směs dvou osob bez rozlišení

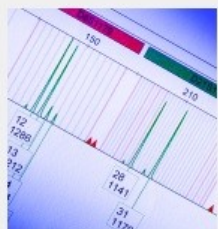
- lze porovnat s předloženou osobou / CODIS a pak propočítat hypotézy
- čím větší číslo, resp. menší, tím lépe – menší pravděpodobnost náhodné shody s další osobou
- na základě frekvence alel v populaci
- vztahuje se k populaci, jejíž alelické frekvence jsou vedené v databázi programu

POPSTATS

- dopočítá RMP hodnotu profilu DNA
- směs dvou osob – zda se podílí nebo nepodílí profil DNA osoby na směsném profilu DNA
- příbuzenské vztahy – R a D, trio
- jiný typ příbuzenství

Dostupný program – LRmix studio

- k dopočítání likelihood ratio pro profil DNA
- umí dopočítat pro 5 osob, resp. 4 a pátou známou osobu
- využívá hypotézy obhajoby a obžaloby a dopočítá váhu té či oné hypotézy
- slouží i jako nástroj k porovnávání profilů DNA



Active	Sample	Source File
<input checked="" type="checkbox"/>	podezřelý	podezřelý.csv

Add profile...

Load from file...

Locus	podezřelý
VWA	18 / 18
D16S539	12 / 12
D2S1338	20 / 26
D8S1179	12 / 15
D21S11	30 / 30.2
D18S51	13 / 14
D19S433	12 / 15
TH01	8 / 9
FGA	23 / 24
D3S1358	13 / 17
D7S820	9 / 12
CSF1PO	10 / 10
D13S317	9 / 11
TPOX	8 / 8
D5S818	12 / 13



Alleles in the **replicate** that are not present in the **reference profiles**

Alleles in the **reference profiles** that are not present in the **replicate**

Matching alleles in the **replicate** and **podezřelý**

- Text Colour
- Background Colour
- Bold**
- Italic*
- Underlined

Highlighted 44 alleles

Print...

Select	Name	Replicate	podezřelý	Distinct Alleles
<input checked="" type="checkbox"/>	stopa			
<input checked="" type="checkbox"/>	VWA	14 16 18	18 18	3
<input checked="" type="checkbox"/>	D16S539	10 12 13	12 12	3
<input checked="" type="checkbox"/>	D2S1338	20 23 26	20 26	3
<input checked="" type="checkbox"/>	D8S1179	12 13 15	12 15	3
<input checked="" type="checkbox"/>	D21S11	28 30 30.2 32.2	30 30.2	4
<input checked="" type="checkbox"/>	D18S51	13 14 15 16	13 14	4
<input type="checkbox"/>	D19S433		12 15	0
<input checked="" type="checkbox"/>	TH01	8 9	8 9	2
<input type="checkbox"/>	FGA		23 24	0
<input checked="" type="checkbox"/>	D3S1358	15 17	13 17	2
<input checked="" type="checkbox"/>	D7S820	9 11 12	9 12	3
<input checked="" type="checkbox"/>	CSF1PO	10 12	10 10	2
<input checked="" type="checkbox"/>	D13S317	9 11 112	9 11	3
<input checked="" type="checkbox"/>	TPOX	8 11	8 8	2
<input checked="" type="checkbox"/>	D5S818	12	12 13	1

Prosecution Hypothesis

Contributor	ID	Dropout Probability
<input checked="" type="checkbox"/>	podezřelý	0,1

Unknown Contributors
 Dropout Probability for unknowns

Defense Hypothesis

Contributor	ID	Dropout Probability
<input type="checkbox"/>	podezřelý	0,1

Unknown Contributors
 Dropout Probability for unknowns
 One of the unknowns is a relative
 Parent/Child of

Parameters

Allele Frequencies
 Rare allele frequency
 Drop-in probability Theta correction
 Max Threads

Results

Locus	LR
VWA	1.1460
D16S539	0.7862
D2S1338	76.6355
D8S1179	6.5818
D21S11	5.2213
D18S51	4.4066
TH01	10.1047
D3S1358	0.3090
D7S820	5.3798
CSF1PO	2.0648
D13S317	4.8540
TPOX	0.9928
D5S818	0.4720

Overall Likelihood Ratio



Vary Dropout Profile

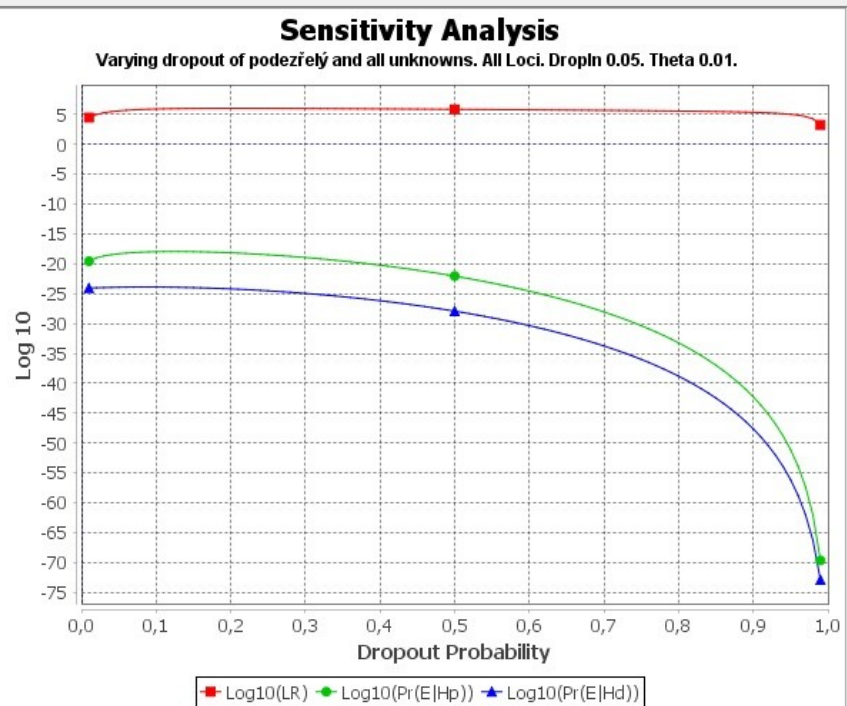
- podezřelý
- Defense Unknown Contributors
- Prosecution Unknown Contributors

Set dropout of selected profiles to:

Sensitivity Analysis Settings Dropout Estimation Settings

Drop-Out variation to in steps at locus

Drop In Theta



Log10(LR)
Log10(Pr(E|Hp))
Log10(Pr(E|Hd))

Prosecution Hypothesis

Profile	Dropout Probability
podezely	0.10
1 Unknown	0.10

Defense Hypothesis

Profile	Dropout Probability
2 Unknown	0.10

RESULTS

Locus	Pr(E Hp)	Pr(E Hd)	LR
VWA	2,92417E-002	2,55146E-002	1,14607E000
D16S539	1,65567E-002	2,10587E-002	7,86214E-001
D2S1338	4,53726E-002	5,92058E-004	7,66355E001
D8S1179	2,31694E-001	3,52022E-002	6,58181E000
D21S11	2,01808E-002	3,86503E-003	5,22139E000
D18S51	3,23177E-002	7,33379E-003	4,40668E000
TH01	1,21467E-001	1,20207E-002	1,01048E001
D3S1358	1,54527E-002	4,99982E-002	3,09065E-001
D7S820	1,41333E-001	2,62708E-002	5,37983E000
CSF1PO	2,80363E-001	1,35777E-001	2,06489E000
D13S317	7,41990E-004	1,52860E-004	4,85405E000
TPOX	3,06138E-001	3,08340E-001	9,92860E-001
D5S818	1,92596E-002	4,08027E-002	4,72019E-001
Overall Likelihood Ratio			8,25317E005



Active	Sample	Source File
<input checked="" type="checkbox"/>	rep1	sample.txt
<input checked="" type="checkbox"/>	rep2	sample.txt
<input checked="" type="checkbox"/>	rep3	sample.txt
<input checked="" type="checkbox"/>	rep4	sample.txt

Case Number Example

Restore session from Log

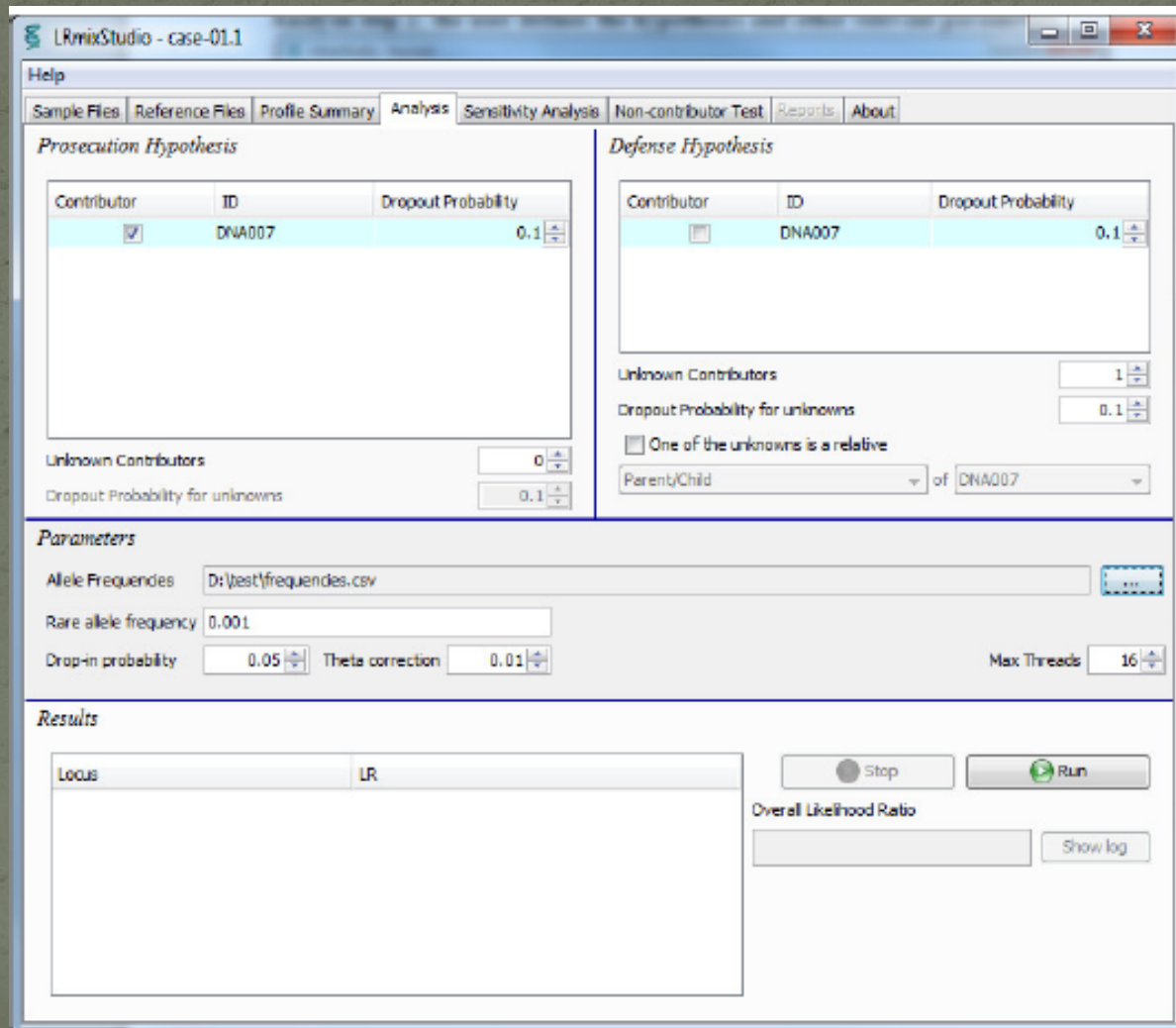
Restart

Add replicate

Load from file...

Locus	rep1	rep2	rep3	rep4
D10S1248	12 15	12 15	12 15	12 15
VWA	14 16	14 16	14 16	14 16
D16S539	9 10	9 10	9 10	10
D2S1338	20 23	20 23	20 23	20 23
AMEL	XY	XY	XY	XY
D8S1179	12 13	12 13	12 13	12 13
D21S11	28 31	28 31	28	
D18S51	12 15	12	12 15	12 15
D22S1045	11 16	16	11 16	11 16
D19S433	14 15	14 15	14 15	14 15
TH01	7 9,3	7 9,3	7 9,3	7 9,3
FGA	24 26	24 26	24 26	24 26
D2S441	14 15	14 15	14 15	14 15
D3S1358	15 16	15 16	15 16	15 16
D151656	13 16	13 16	13 16	13 16
D12S391	18	18	18	18

LRmix studio



LRmixStudio - case-01.1

Help

Sample Files | Reference Files | Profile Summary | Analysis | Sensitivity Analysis | Non-contributor Test | Reports | About

Prosecution Hypothesis

Contributor	ID	Dropout Probability
<input checked="" type="checkbox"/>	DNA007	0.1

Unknown Contributors:

Dropout Probability for unknowns:

Defense Hypothesis

Contributor	ID	Dropout Probability
<input type="checkbox"/>	DNA007	0.1

Unknown Contributors:

Dropout Probability for unknowns:

One of the unknowns is a relative

Parent/Child: of

Parameters

Alele Frequencies: ...

Rare alele frequency:

Drop-in probability: Theta correction: Max Threads:

Results

Locus	LR
D10S1248	55.5290
VWA	22.3325
D16S539	49.9784
D2S1338	31.3001
D8S1179	9.3984
D21S11	26.6025
D18S51	20.5438
D22S1045	9.1672
D19S433	7.2339

Stop

Overall Likelihood Ratio:

Výstup

Chyba obžaloby (Transposed conditional)

- **Správný závěr:** Výsledky DNA analýzy jsou 1000 krát pravděpodobnější, pokud obžalovaný zanechal stopu, než pokud ji zanechal jiný člověk.
- **Nesprávný závěr:** Je 1000 krát větší pravděpodobnost, že podezřelý zanechal stopu, než pokud ji zanechal jiný člověk.