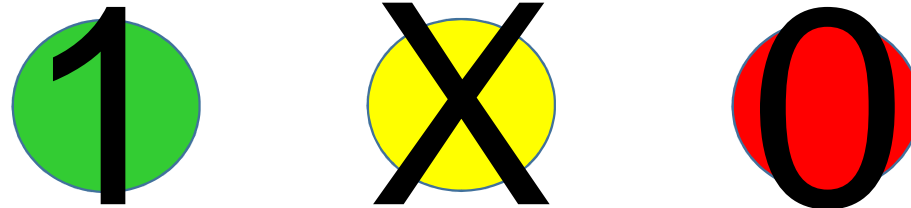


Dominantní molekulární markery

**Bi6589 Laboratorní a bioinformatické metody
rostlinné biosystematiky**

Dominantní molekulární markery

- Nejsme schopni odlišit heterozygoty od dominantních homozygotů
- Bialelická povaha lokusu (fragmentu)
 - > přítomnost (dominantní alela/heterozygot) = 1
 - > nepřítomnost (recesivní alela) = 0
- Binární povaha dat



Metodické přístupy: AFLP, RAPD, ISSR, ...

Výhody dominantních dat

- Obvykle vysoká variabilita
- Mnoho nezávislých lokusů
- Pokrytí „celého“ genomu
- Dostupný a prověřený statistický aparát k analýze dat

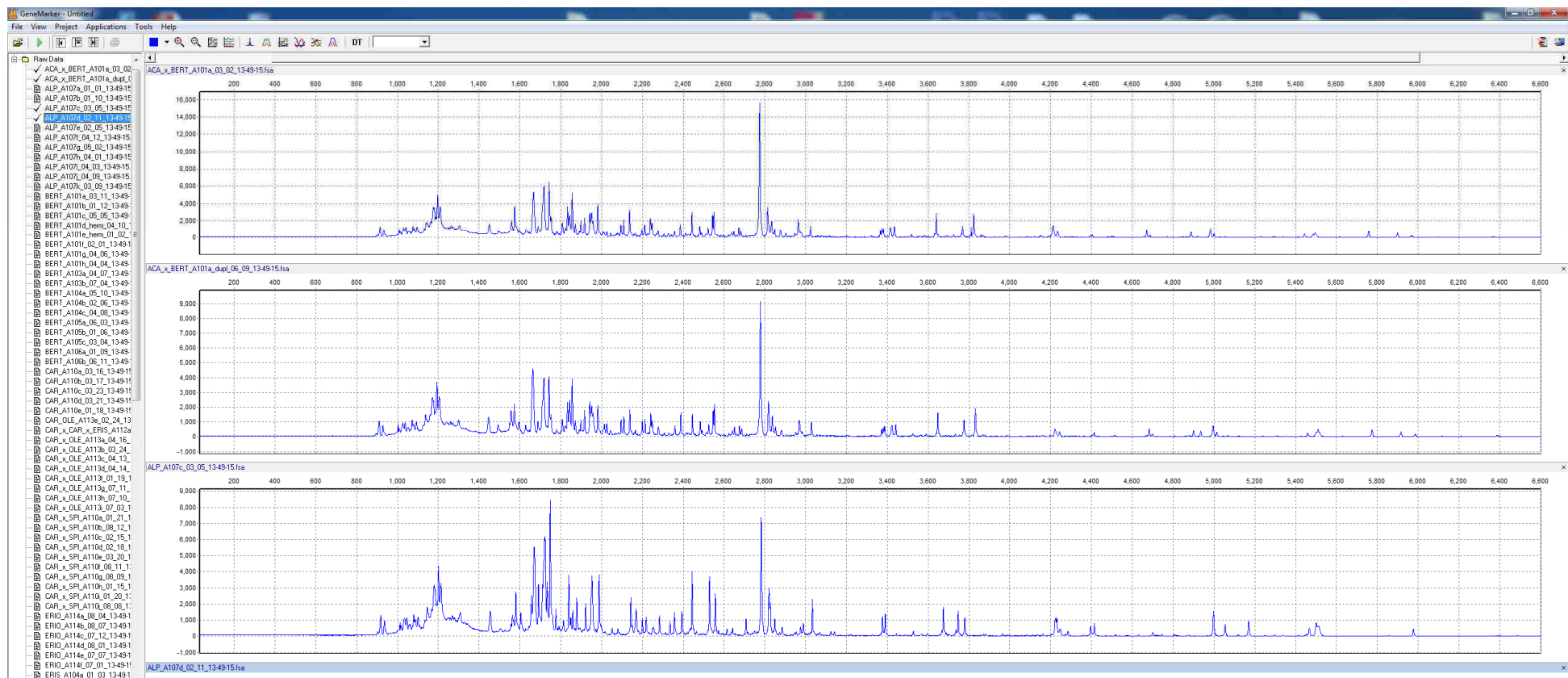
Nevýhody dominantních dat

- Anonymní marker
- Nemožnost odlišit homozygoty od heterozygotů
- Subjektivita hodnocení
- Neznámá rychlost akumulace mutací (nemožnost použít molekulární hodiny)
- Různá rychlost generování „znaků“
- N_e = více důvodů
- Problematické přidávání nových/dalších vzorků
- Homoplazie

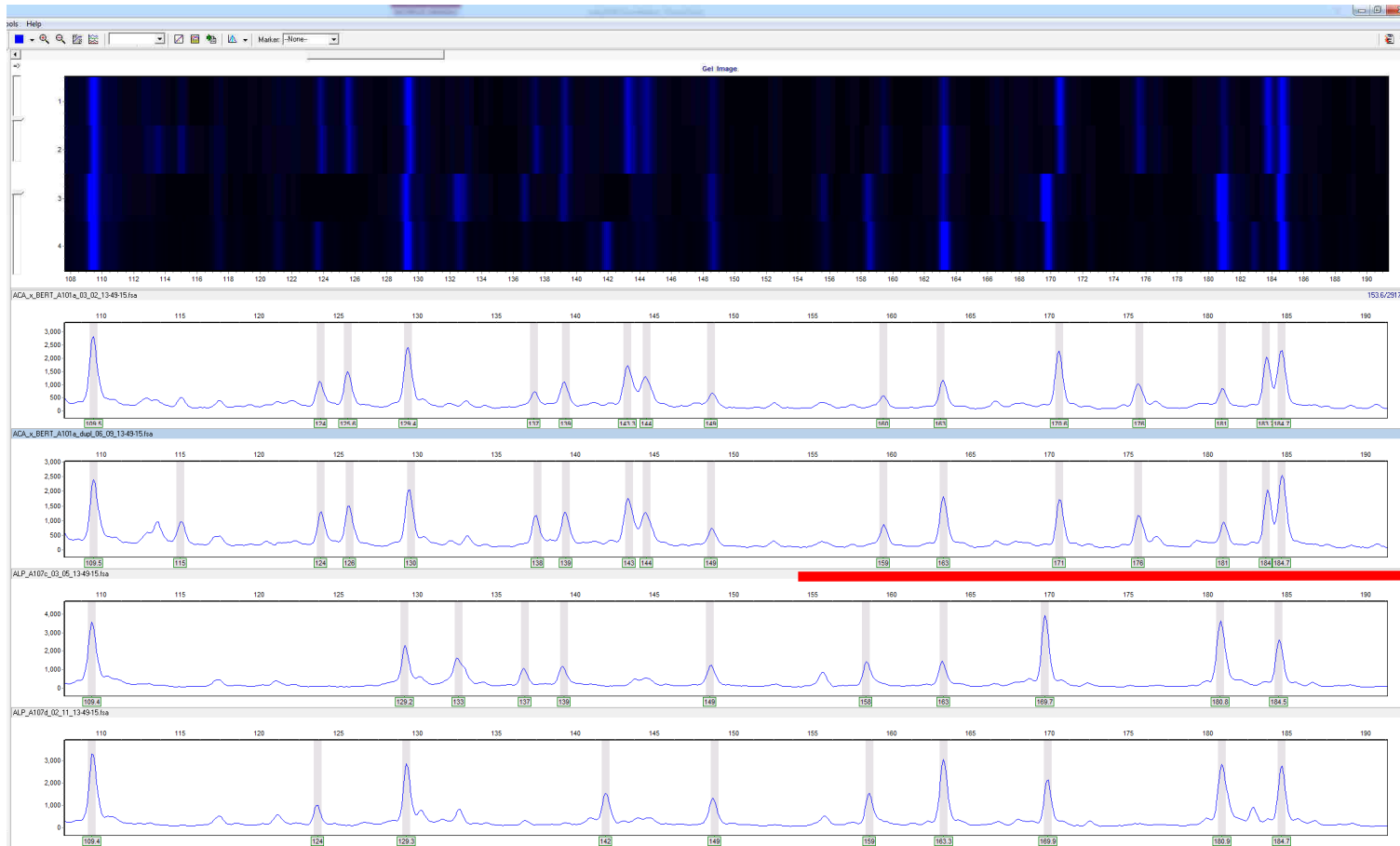
Využití metod

- Populační genetika
- Klonalita (identifikace klonů; error rate)
- Hybridizace (identifikace rodičů; míra introgrese/genetická eroze)
- Fylogeografie (refugia; migrační cesty; kontaktní zóny genových linií)
- Fylogeneze (vs. sekvenční data)
- Polyploidizace (autopolyploidi vs. allopolyploidi)

AFLP - primární data



Hodnocení primárních dat (GeneMarker)



Binární matice

110	115	124	126	129	130
1	0	1	1	0	1
1	1	1	1	0	1
1	0	0	0	1	0
1	0	1	0	1	0

Možnosti analýzy

Vztahy mezi jedinci

- Matice podobností
- Stromy, sítě
- Mnohorozměrné analýzy (PCoA)
- Bayesovské clusterování

Populačně-genetické parametry

- Diverzita (% polymorfních lokusů, index diverzity)
- Divergence (% unikátních lokusů, DW – index)
- F-statistika

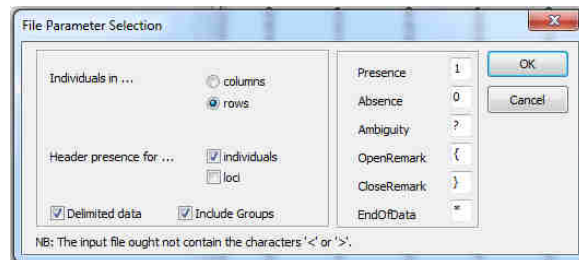
Testování a zjišťování prostorové struktury

- AMOVA
- Bayesovské odhady
- Mantelovy testy
- Prostorová autokorelace

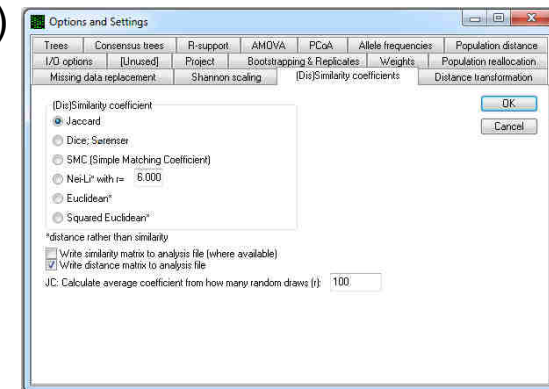
Fylogenetická analýzy

FAMD (Fingerprint Analysis with Missing Data)

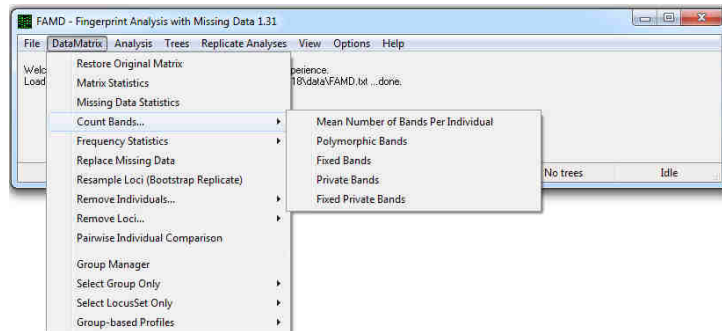
Load data



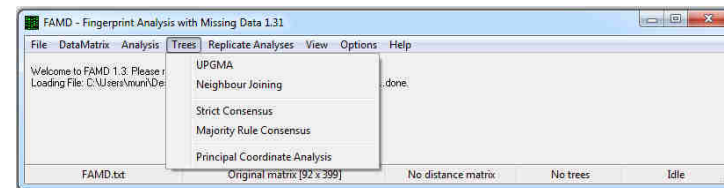
Koeficienty podobnosti (Jaccard, Dice, SMC, Euclidean, ...)



Statistika znaků



Stromy + PCoA zobrazení



Výhody: umí pracovat s chybějícími daty; exporty dat

Vztahy mezi jedinci: koeficienty podobnosti

Prezence bandu		Jedinec B	
		1	0
Jedinec A	1	a	b
	0	c	d

	Jedinec A	Jedinec B
a	—————	—————
b		—————
c	—————	
d		

- Jaccardův koeficient (Jaccard 1908)

$$\frac{a}{a + b + c}$$

- Dice koeficient (Dice 1945) = Nei & Li 1979, Sørensen 1948

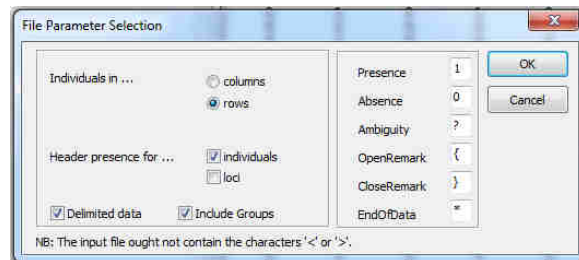
$$\frac{2a}{2a + b + c}$$

- „simple-matching“ koeficient (Sokal & Michener 1958)

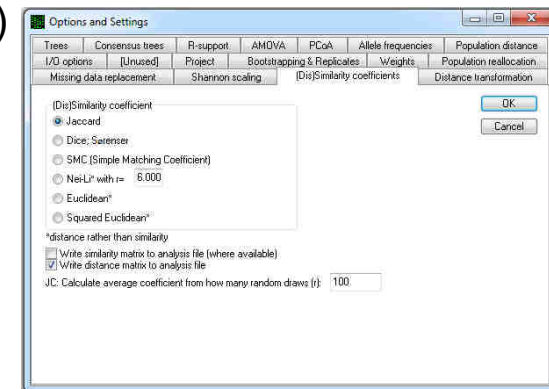
$$\frac{a + d}{a + b + c + d}$$

FAMD (Fingerprint Analysis with Missing Data)

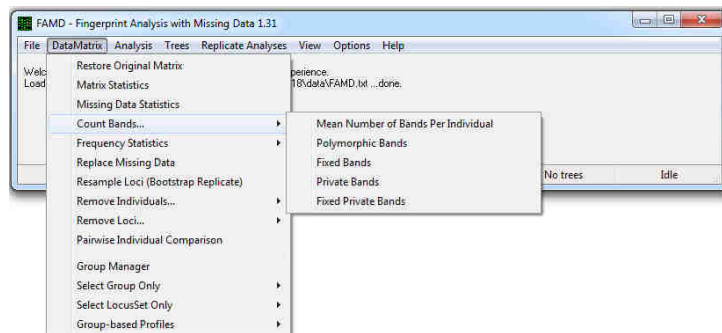
Load data



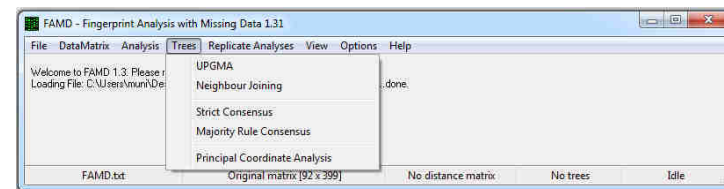
Koeficienty podobnosti (Jaccard, Dice, SMC, Euclidean, ...)



Statistika bandů



Stromy + PCoA zobrazení



Výhody: umí pracovat s chybějícími daty; exporty dat

Principal coordinate(s) analysis - PCoA

- Pro analýzu binárních znaků
- Počet znaků může být větší než počet objektů
- Nehierarchická vizualizace struktury dat
- Výpočet matice vzdáleností mezi objekty (Jaccard)
- Ordinační diagram – převedení objektů do nového prostoru na základě vzdálenosti mezi nimi a maximalizace vysvětlení variability

Software: FAMD

Tvorba stromů

Výpočet matice vzdálenosti – index podobnosti

- Jaccard
- Nei & Li
- Simple matching

Algoritmus na tvorbu stromů

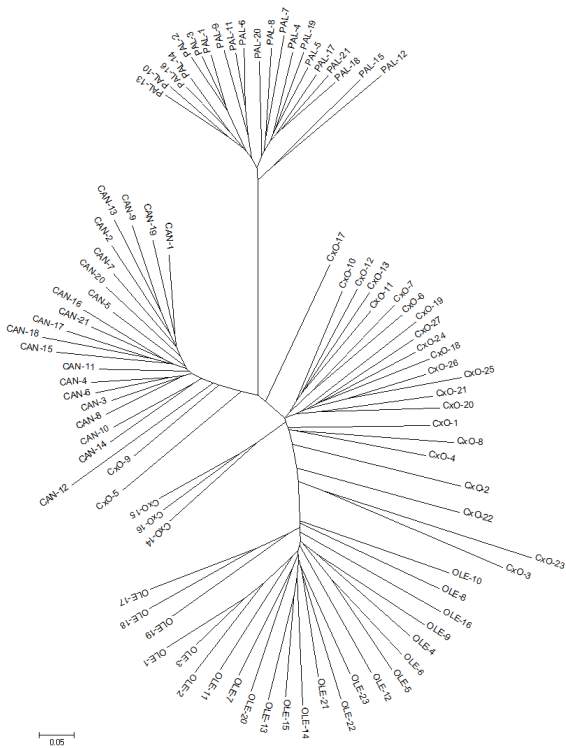
- Shlukování – UPGMA, single link, complete link
- Neighbour joining (NJ) – minimalizace délky
- Maximální parsimonie (MP) – kladistika
- Neighbour - network

Software

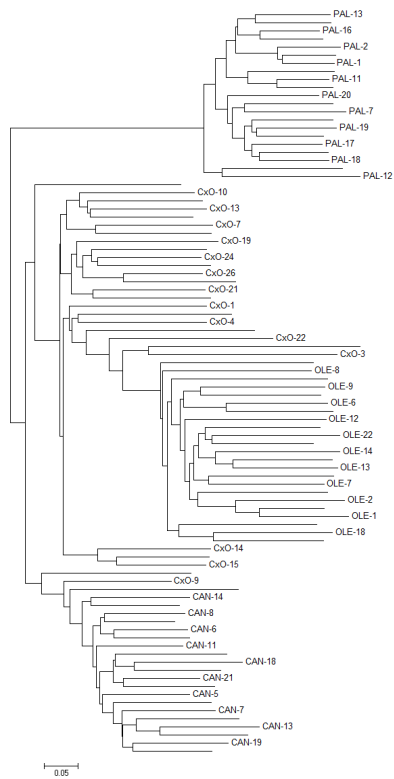
- Tvorba stromů: FAMD
- Prohlížení stromů: MEGA

Typy stromů

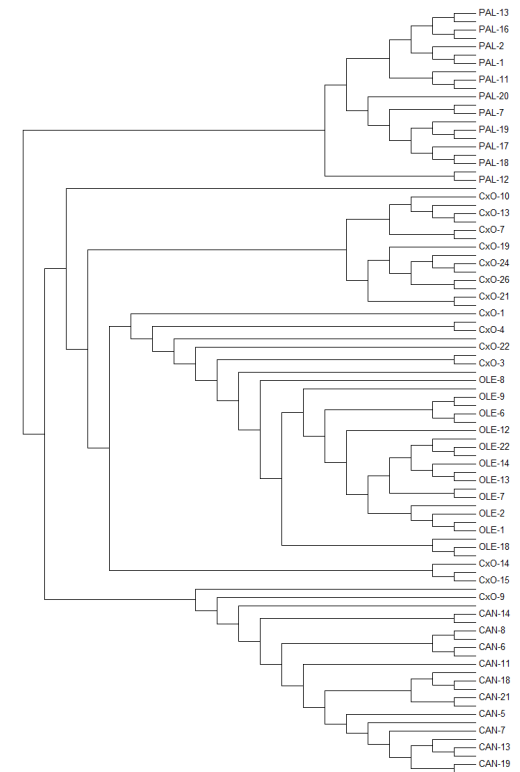
Nezakořeněný



Fylogram

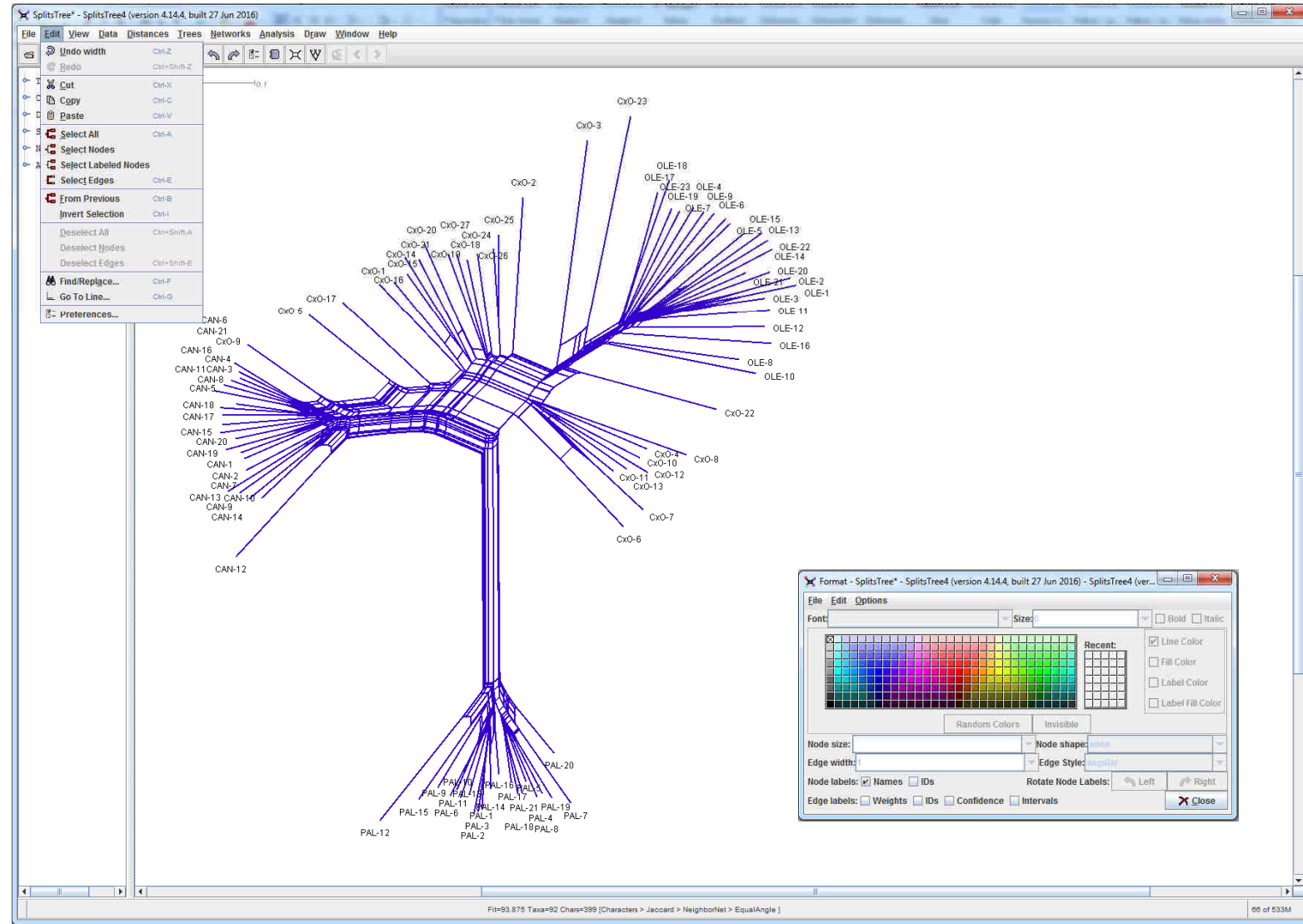


Kladogram



SplitsTree

- Nexus file
- Distance:
nastaveno
uncorrected P



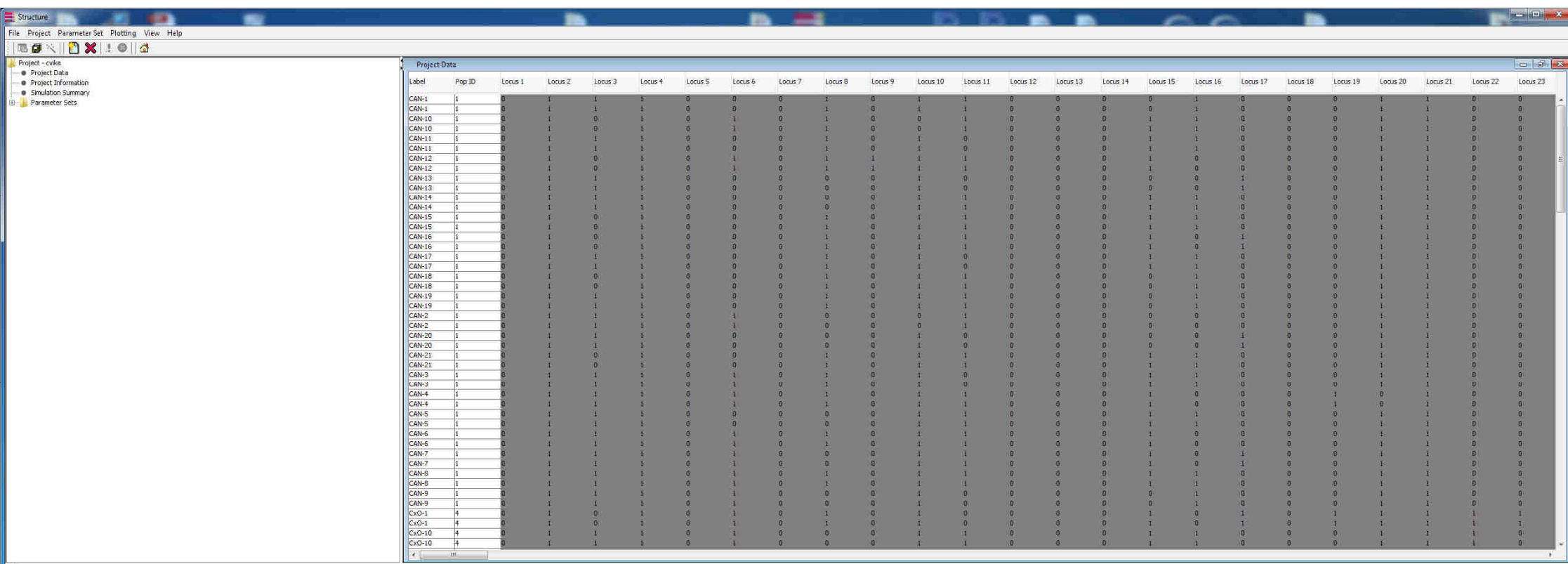
Byaesovské clusterování

- Hledání takového rozdělení jedinců do K clusterů, které je nejvíce pravděpodobné (resp. má maximální záporný logaritmus marginální pravděpodobnosti)
- Výsledkem je rozřazení jedinců do populací (clusterů)
- V rámci populací jsou lokusy v Hardy-Weinbergově rovnováze

Problémy: dominantní data; různé ploidní úrovně

Software: STRUCTURE

STRUCTURE

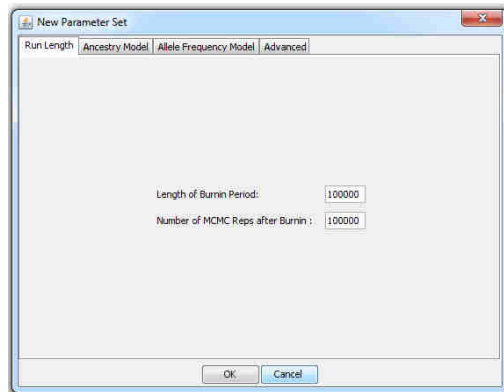


The screenshot displays the STRUCTURE software interface. The main window shows a 'Project Data' table with the following columns: Label, Pop ID, Locus 1, Locus 2, Locus 3, Locus 4, Locus 5, Locus 6, Locus 7, Locus 8, Locus 9, Locus 10, Locus 11, Locus 12, Locus 13, Locus 14, Locus 15, Locus 16, Locus 17, Locus 18, Locus 19, Locus 20, Locus 21, Locus 22, and Locus 23. The rows represent different population labels, including CAN-1 through CAN-19, CAN-2 through CAN-21, CAN-3, LANN-3 through LANN-5, CAN-4 through CAN-9, and CxO-1 through CxO-10. Each cell in the table contains a numerical value, typically 0 or 1, representing the assignment of an individual to a population at a specific locus.

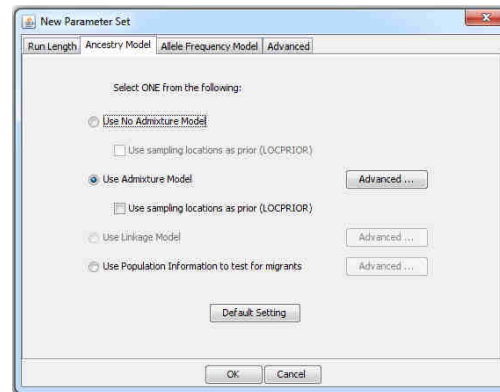
Label	Pop ID	Locus 1	Locus 2	Locus 3	Locus 4	Locus 5	Locus 6	Locus 7	Locus 8	Locus 9	Locus 10	Locus 11	Locus 12	Locus 13	Locus 14	Locus 15	Locus 16	Locus 17	Locus 18	Locus 19	Locus 20	Locus 21	Locus 22	Locus 23
CAN-1	1	0	1	1	1	0	0	0	1	0	1	1	0	0	0	1	0	0	0	1	1	0	0	
CAN-1	1	0	1	1	1	0	0	0	1	0	1	1	0	0	0	1	0	0	0	1	1	0	0	
CAN-10	1	0	1	0	1	0	1	0	1	0	0	1	0	0	1	1	0	0	0	1	1	0	0	
CAN-10	1	0	1	0	1	0	1	0	1	0	0	1	0	0	1	1	0	0	0	1	1	0	0	
CAN-11	1	0	1	1	1	0	0	0	1	0	1	0	0	0	1	1	0	0	0	1	1	0	0	
CAN-11	1	0	1	1	1	0	0	0	1	0	1	0	0	0	1	1	0	0	0	1	1	0	0	
CAN-12	1	0	1	0	1	0	1	0	1	1	1	1	0	0	1	0	0	0	0	1	1	0	0	
CAN-12	1	0	1	0	1	0	1	0	1	1	1	1	0	0	1	0	0	0	0	1	1	0	0	
CAN-13	1	0	1	1	1	0	0	0	1	0	0	0	0	0	0	0	1	0	0	1	1	0	0	
CAN-13	1	0	1	1	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1	1	0	0	
LANN-14	1	0	1	1	1	0	0	0	1	0	0	0	0	0	1	1	0	0	0	1	1	0	0	
LANN-14	1	0	1	1	1	0	0	0	1	0	1	1	0	0	1	1	0	0	0	1	1	0	0	
CAN-14	1	0	1	1	1	0	0	0	0	0	1	1	0	0	1	1	0	0	0	1	1	0	0	
CAN-15	1	0	1	0	1	0	0	0	1	0	1	1	0	0	1	1	0	0	0	1	1	0	0	
CAN-15	1	0	1	0	1	0	0	0	1	0	1	1	0	0	1	1	0	0	0	1	1	0	0	
CAN-16	1	0	1	0	1	0	0	0	1	0	1	1	0	0	1	0	1	0	0	1	1	0	0	
CAN-16	1	0	1	0	1	0	0	0	1	0	1	1	0	0	1	0	1	0	0	1	1	0	0	
CAN-17	1	0	1	1	1	0	0	0	1	0	1	0	0	0	1	1	0	0	0	1	1	0	0	
CAN-17	1	0	1	1	1	0	0	0	1	0	1	0	0	0	1	1	0	0	0	1	1	0	0	
CAN-18	1	0	1	0	1	0	0	0	1	0	1	1	0	0	0	1	0	0	0	1	1	0	0	
CAN-18	1	0	1	0	1	0	0	0	1	0	1	1	0	0	0	1	0	0	0	1	1	0	0	
CAN-19	1	0	1	1	1	0	0	0	1	0	1	1	0	0	0	1	0	0	0	1	1	0	0	
CAN-19	1	0	1	1	1	0	0	0	1	0	1	1	0	0	0	1	0	0	0	1	1	0	0	
CAN-2	1	0	1	1	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	1	1	0	0	
CAN-2	1	0	1	1	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	1	1	0	0	
CAN-20	1	0	1	1	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1	1	0	0	
CAN-20	1	0	1	1	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1	1	0	0	
CAN-21	1	0	1	0	1	0	0	0	1	0	1	1	0	0	1	1	0	0	0	1	1	0	0	
CAN-21	1	0	1	0	1	0	0	0	1	0	1	1	0	0	1	1	0	0	0	1	1	0	0	
CAN-3	1	0	1	1	1	0	1	0	1	0	1	0	0	0	1	1	0	0	0	1	1	0	0	
LANN-3	1	0	1	1	1	0	1	0	1	0	1	0	0	0	1	1	0	0	0	1	1	0	0	
LANN-3	1	0	1	1	1	0	1	0	1	0	1	0	0	0	1	1	0	0	0	1	1	0	0	
CAN-4	1	0	1	1	1	0	1	0	1	0	1	0	0	0	1	0	0	0	1	0	1	0	0	
CAN-4	1	0	1	1	1	0	1	0	1	0	1	0	0	0	1	0	0	0	1	0	1	0	0	
CAN-4	1	0	1	1	1	0	1	0	1	0	1	0	0	0	1	0	0	0	1	0	1	0	0	
CAN-5	1	0	1	1	1	0	0	0	0	0	1	1	0	0	1	1	0	0	0	1	1	0	0	
CAN-5	1	0	1	1	1	0	0	0	0	0	1	1	0	0	1	1	0	0	0	1	1	0	0	
CAN-6	1	0	1	1	1	0	1	0	1	0	1	1	0	0	1	0	0	0	0	1	1	0	0	
CAN-6	1	0	1	1	1	0	1	0	1	0	1	1	0	0	1	0	0	0	0	1	1	0	0	
CAN-7	1	0	1	1	1	0	1	0	0	0	1	1	0	0	1	0	1	0	0	1	1	0	0	
CAN-7	1	0	1	1	1	0	1	0	0	0	1	1	0	0	1	0	1	0	0	1	1	0	0	
CAN-8	1	0	1	1	1	0	1	0	1	0	1	1	0	0	1	1	0	0	0	1	1	0	0	
CAN-8	1	0	1	1	1	0	1	0	1	0	1	1	0	0	1	1	0	0	0	1	1	0	0	
CAN-9	1	0	1	1	1	0	1	0	1	0	1	0	0	0	0	1	0	0	0	1	1	0	0	
CAN-9	1	0	1	1	1	0	1	0	0	0	1	0	0	0	0	1	0	0	0	1	1	0	0	
CxO-1	4	0	1	0	1	0	1	0	1	0	1	0	0	0	1	0	1	0	1	1	1	1	1	
CxO-1	4	0	1	0	1	0	1	0	1	0	1	0	0	0	1	0	1	0	1	1	1	1	1	
CxO-10	4	0	1	1	1	0	1	0	0	0	1	1	0	0	0	1	1	0	0	1	1	1	0	
CxO-10	4	0	1	1	1	0	1	0	0	0	1	1	0	0	0	1	1	0	0	1	1	1	0	

STRUCTURE - nastavení

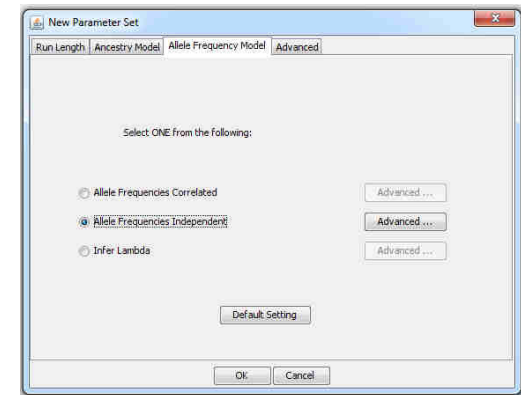
Burn: 100 000; run: 100 000 (1 000 000)



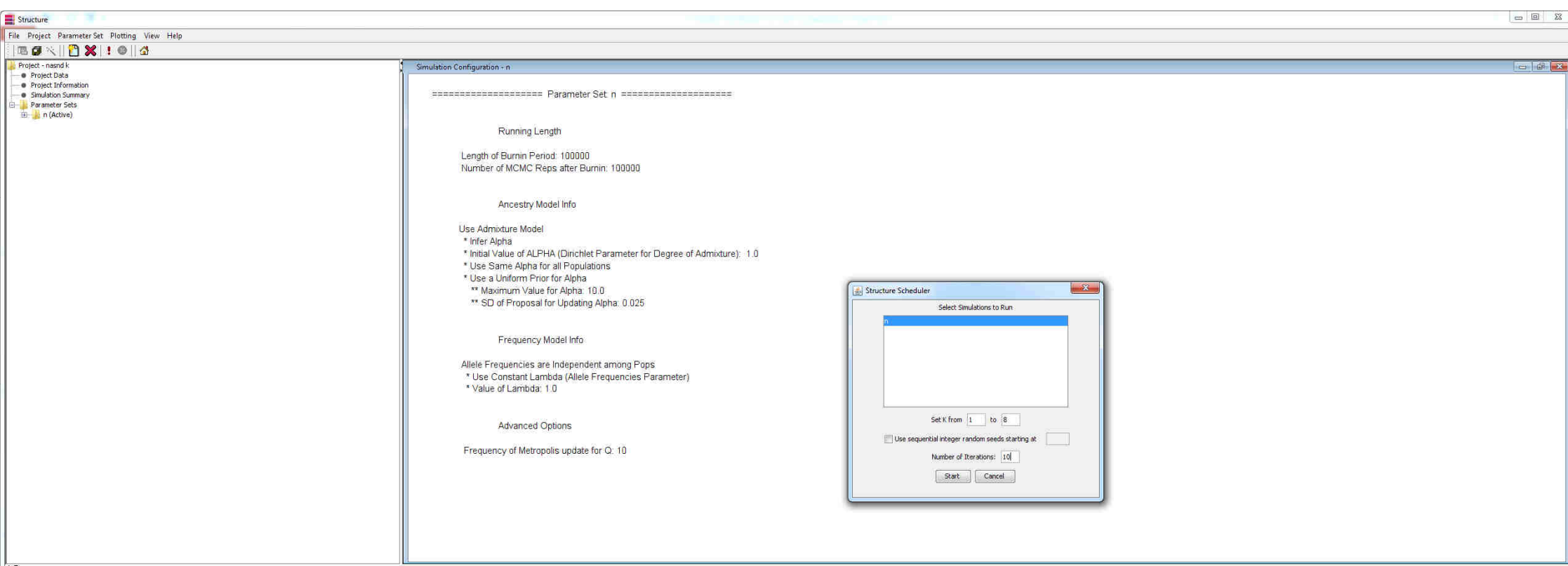
Admixture model



Independent allele frequencies



STRUCTURE - nastavení

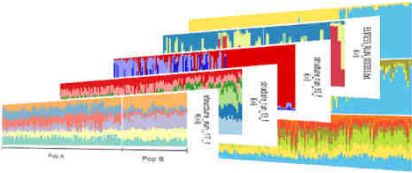


Analýza a zobrazení výsledků: Pophelper

Pophelper: analýza a zobrazení výsledků

← → ↻ ⓘ pophelper.com

Pophelper - web apps
Structure Analyses
Analyse & visualise Structure, fastStructure, Tess & Admixture runs.



Upload Data Evanno Plot Guide Versions

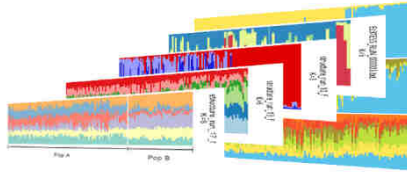
Upload file(s):
Zvolit soubory | Počet souborů: 80 [Upload complete](#)

Upload one or more STRUCTURE, TESS, fastSTRUCTURE or ADMIXTURE run files.
Do not change too many controls at the same time. Adjust an option, wait for the change and then adjust another option.
Not recommended for large datasets. Max upload limits: 200 KBs per file, max 25 files. See guide for more info.
Each session closes after 10 mins of inactivity. Data is not stored/saved on the server. Please download tables and plots that you wish to keep.
Check out the 'Guide' tab for more detailed instructions and sample files to download.

	filename	format	size_kb
1	ovika_run_1.f	STRUCTURE	35.5
2	ovika_run_2.f	STRUCTURE	35.5
3	ovika_run_3.f	STRUCTURE	35.5
4	ovika_run_4.f	STRUCTURE	35.5
5	ovika_run_5.f	STRUCTURE	35.5
6	ovika_run_6.f	STRUCTURE	35.5
7	ovika_run_7.f	STRUCTURE	35.5
8	ovika_run_8.f	STRUCTURE	35.5
9	ovika_run_9.f	STRUCTURE	35.5
10	ovika_run_10.f	STRUCTURE	35.5
11	ovika_run_11.f	STRUCTURE	40.8
12	ovika_run_12.f	STRUCTURE	40.8
13	ovika_run_13.f	STRUCTURE	40.8
14	ovika_run_14.f	STRUCTURE	40.8
15	ovika_run_15.f	STRUCTURE	40.8
16	ovika_run_16.f	STRUCTURE	40.8
17	ovika_run_17.f	STRUCTURE	40.8
18	ovika_run_18.f	STRUCTURE	40.8
19	ovika_run_19.f	STRUCTURE	40.8
20	ovika_run_20.f	STRUCTURE	40.8
21	ovika_run_21.f	STRUCTURE	46.0
22	ovika_run_22.f	STRUCTURE	46.0
23	ovika_run_23.f	STRUCTURE	46.0
24	ovika_run_24.f	STRUCTURE	46.0
25	ovika_run_25.f	STRUCTURE	46.0
26	ovika_run_26.f	STRUCTURE	46.0

Structure Analyses

Analyse & visualise Structure, fastStructure, Tess & Admixture runs.



Upload Data Evanno Plot Guide Versions

Input Output

Evanno results data

	loci	ind	k	runs	elpdmean	elpdsd	elpdmin	elpdmax	lnk1	lnk1max	lnk1min	lnk2	lnk2max	lnk2min	deltaK
1	399	92	1	10	-29913.90	0.45	-29914.6	-29912.9							
2	399	92	2	10	-23486.57	274.98	-24230.9	-23318.3	6427.33	6701.86	6152.80	1721.88	1799.65	1644.11	6.262
3	399	92	3	10	-18781.12	78.22	-18941.9	-18703.5	4705.45	4902.21	4508.69	3789.38	3842.70	3736.06	48.445
4	399	92	4	10	-17865.05	328.30	-18781.3	-17671.2	916.07	1166.15	665.99	804.77	816.01	793.53	2.451
5	399	92	5	10	-17753.75	66.98	-17872.4	-17681.2	111.30	372.42	-150.02	129.49	375.53	-116.55	-1.933
6	399	92	6	10	-17771.94	51.70	-17870.9	-17703.8	-18.19	-2.91	-35.47	6.92	17.12	-3.28	0.134
7	399	92	7	10	-17797.05	56.78	-17866.7	-17715.7	-25.11	-20.03	-30.19	65.28	65.73	64.83	1.150
8	399	92	8	10	-17758.88	62.31	-17880.6	-17660.7	40.17	45.70	34.64				

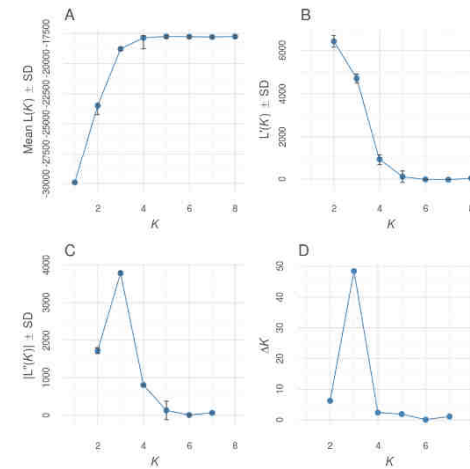
Download Evanno data

File format

tab-delimited

Download

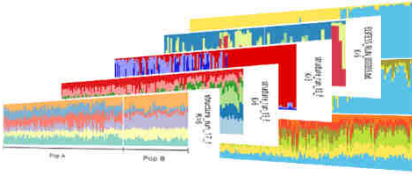
Evanno results plot



Pophelper: hodnocení STRUCTURE výsledků

← → ↻ pophelper.com

Pophelper · web apps
Structure Analyses
Analyse & visualise Structure, fastStructure, Tess & Admixture runs.



Upload Data Evanno Plot Guide Versions

> Plot options

Select file(s) to plot:

file	k
cvilka_run_43_f	5
cvilka_run_44_f	5
cvilka_run_45_f	5
cvilka_run_46_f	5

Order individuals: None

Colour scheme: Viridis

Plot pop labels

Standard Plot Interactive Plot

> Interactive options

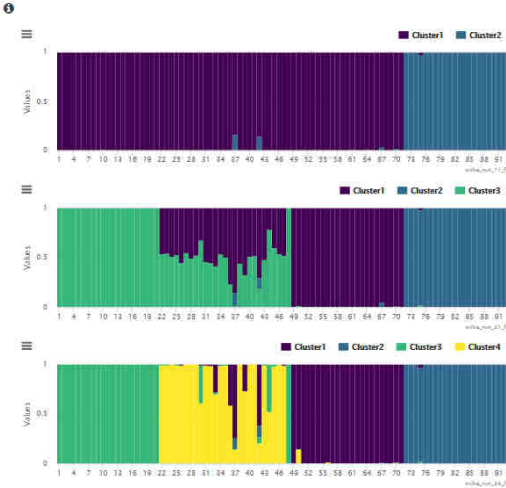
Plot height (px): 200

Plot width (px): 200

Bar border size: 0

Bar gap: 0

Show legend
 Show small filename
 Show large filename
 Fix y-axis to 1



POPHelper Structure Web App v1.0.10, Last Updated: 29-May-2016
© 2012-2016 François Bonhomme

Populačně genetické parametry

Genetická diverzita (jak velké jsou rozdíly mezi jedinci/populacemi)

- % polymorfních fragmentů
- Průměrná genová diverzita (N_e)

-> odráží reprodukční systém, recentní procesy (genový tok, velikost populace,...)

Genetická divergence – vzácnost (jak moc se populace odlišuje od ostatních; jak moc je unikátní -> vznik odlišností kumulací mutací)

- Počet (%) vzácných fragmentů
- Počet unikátních (privátních) fragmentů
- DW – index (*frequency down-weighted marker values*) (= kumulace vzácných fragmentů v populaci)

-> odráží historické procesy (izolace populace, ...)

Software: AFLPdat, Arlequin, ...

Rozložení genetické diverzity (F-statistika)

- celková diverzita: H_T
- vnitropopulační diverzita: H_S
- interpopulační diverzita: $H_T - H_S$

G_{ST} = koeficient genetické diferenciace (míra diferenciace na subpopulace)

0 = žádná genetická struktura populací (všechny populace mají stejné frekvence alel)



$$G_{ST} = \frac{H_T - H_S}{H_T}$$

1 = maximální genetická struktura (každá populace fixovala jinou alelu)



G_{ST} ; podobné: R_{ST} , G'_{ST} , Φ_{ST} , D , G''_{ST}

Testování a zjišťování prostorové struktury

- AMOVA: Analýza molekulární variance
- Mantel testy (genetická x geografická matice; permutace = signifikance; distanční třídy)
- prostorová autokorelace (kinship koeficienty mezi páry jedinců = pravděpodobnost, že dva geny náhodně vytažené z každého jedince jsou si identické vs. geografické vzdálenostmi; permutace)

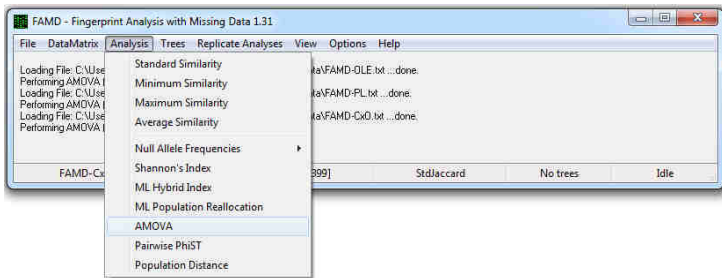
Analýza molekulární variance (AMOVA)

- testování populačně genetické struktury
- rozdělení celkové variance na komponenty (skupiny, populace, jedince)
- založeno na výpočtu F-statistiky

Software: Arlequin, FAMD

FAMD: analýza AMOVA

Nastavení



Výsledky

AMOVA results

Coefficient: Standard Jaccard. Distance Transformation: d=1-s.

SSD(AG)= 10.984597160, df(AG)= 3

SSD(APWG)= 1.676973499, df(APWG)= 12

SSD(WP)= 5.666238567, df(WP)= 76

SSD(T)= 18.327809226, df(T)= 91

Va= 0.153186591, %Var= 63.927824961 (Among Groups)

Vb= 0.011881919, %Var= 4.958562308 (Among Populations/Within Groups)

Vc= 0.074555771, %Var= 31.113612731 (Within Populations)

Vt= 0.239624280

PhiST= 0.688863873

PhiRT= 0.639278250

PhiSR= 0.137462249

AMOVA

