

**Protokoly a návody pro praktikum**

**Molekulární markery v systematice a populační  
biologii rostlin**

(B120C44)

**část II. Sekvenování DNA, AFLP a mikrosatelity**

**v DNA laboratoři Katedry botaniky PŘF UK, Benátská 2, Praha**



# A. PROTOKOLY

## protokol E: sekvenování DNA

Sekvenováním zjistíme pořadí bazí (A, C, T, G) v nějakém konkrétním úseku DNA. Postup můžeme rozdělit do několika kroků:

1. *PCR amplifikace* konkrétního úseku DNA pomocí dvojice specifických primerů. Primery jsou specifické pro jedno konkrétní místo v jaderné nebo chloroplastové DNA. V jaderné neboli nukleární DNA (nDNA) se nejjednodušeji a nejčastěji se sekvenují oblasti tzv. internal transcribed spaceru (ITS), které jsou přítomny až v desítkách tisíc kopií. Úseky chloroplastového genomu se také velmi jednoduše amplifikují, protože kruhových molekul cpDNA je v izolátu zpravidla velké množství. Často se sekvenují nekódující úseky (introny nebo spacers) jako např. *trnL-trnF*, *psbA-trnH* ad. (viz Tabulka 1)
2. *přečištění PCR produktu* – provede se nejčastěji pomocí komerčních kitů (např. QIAGEN nebo Genomed JETQUICK). Přečištěním se produkt zbaví zbylých primerů a neinkorporovaných dNTP. Toto je velmi důležité, protože při následné sekvenační reakci je použit pouze jeden primer a také koncentrace dNTP musí být vybalancována vzhledem k ostatním složkám, především ddNTP (viz další krok). Aby sekvenační reakce proběhla optimálně, je třeba zajistit i správný poměr koncentrace primer:PCR produkt. Proto je dobré změřit koncentraci PCR produktu např. fotometricky.
3. *cyklické sekvenování* (cycle sequencing) je v podstatě modifikovaná PCR reakce, při které použijeme pouze jeden sekvenační primer a PCR produkt jako templát. Nedochozí tedy k exponenciální, ale pouze k lineární amplifikaci templátu. Kromě klasických dNTP jsou v sekvenační směsi přítomny i ddNTP (2',3'-dideoxynukleotidtrisfosfáty). Ty jsou fluorescenčně značeny (každá база jinou barvou) a navíc při inkorporaci do vznikajícího řetězce DNA zamezí jeho dalšímu prodlužování. Výsledkem sekvenační reakce je tak směs různě dlouhých řetězců DNA a každý je fluorescenčně označen barvou podle toho, kterou bazí končí. Tyto produkty jsou pak elektroforeticky rozděleny na automatickém sekvenátoru a je možno přečíst posloupnost jednotlivých bazí. Sekvenuje se vždy jen s jedním primerem (umožní číst cca 450-550 bazí, záleží na kvalitě PCR produktu a na sekvenci bazí). Pokud chceme sekvenovat delší úsek než cca 500 bazí, je třeba provést dvě sekvenační reakce s různými primery (tzv. *forward* a *reverse*), PCR produkt bude osekvenován z obou stran, které se potom pomocí specifického software spojí dohromady.

4. *přesrážení (precipitace) produktů sekvenační reakce* – před analýzou na automatickém sekvenátoru je třeba produkty cyklického sekvenování přechistit a přesrážet s NaOAc (acetát sodný) a ethanolem.
5. *sekvenační analýza na automatickém sekvenátoru* – v kapilárním sekvenátoru ABI 3100 Avant (v sekvenační laboratoři PřF UK) jsou produkty cyklického sekvenování elektroforeticky rozděleny podle délky.

## E1. protokol na PCR amplifikaci

Do 1.5 ml eppendorfky si připravíme PCR směs pro příslušný počet vzorků. Množství polymerázy i studované DNA je možné měnit, v tom případě upravíme objem vody tak, aby celkové množství bylo 20  $\mu$ l.

- sterilní Milli-Q voda..... 15.60  $\mu$ l
- 10 $\times$  PCR buffer (Sigma) ..... 2.00  $\mu$ l
- dNTP (10 mM) ..... 0.40  $\mu$ l
- *forward* primer (25 pmol/ $\mu$ l)..... 0.25  $\mu$ l
- *reverse* primer (25 pmol/ $\mu$ l) ..... 0.25  $\mu$ l
- (Jump Start) Red Taq DNA Polymerase (Sigma) (1U/ $\mu$ l).. 0.50  $\mu$ l
- + studovaná DNA (5 ng/ $\mu$ l)..... 1.00  $\mu$ l

Se všemi chemikáliemi pracujeme na ledu. Polymerázu vyndáme z mrazáku až těsně před pipetováním a pak ji ihned vracíme zpátky. Jako primery použijeme *forward* a *reverse* primery podle toho, jaký úsek chceme sekvenovat (Tabulka 1). Rozpipetujeme do jednotlivých zkumavek příslušný objem směsi (v tomto případě 19  $\mu$ l) a přidáme 1  $\mu$ l DNA.

Zkumavky vložíme do termocykleru s následujícím programem:

94 °C	1:00	} 25-35 $\times$
94 °C	0:45	
52-62 °C	0:45	
72 °C	1:00-2:00	
72 °C	10:00	
10 °C	hold	

Teplotu annealingu a dobu elongace nastavíme podle Tabulky 1. Téměř vždy je dobré provést gradientový test optimální teploty annealingu před sekvenováním většího množství vzorků.

Po proběhnutí reakci otestujeme amplifikovaný produkt nanesením cca 5  $\mu$ l na agarosový gel (1%) v TAE bufferu. Do první jamky na gelu nanese 3  $\mu$ l žebříčku (100 bp Gene Ruler). Na gelu by měl být vidět jeden proužek v délce zhruba odpovídající očekávané dle tabulky.

Jeho délku odhadneme pomocí softwaru KODAK 1D Image Analysis Software porovnáním s proužky žebříčku.

Pokud je na gelu viditelný jeden jasný proužek v délce zhruba odpovídající délce fragmentu dle tabulky, povedlo se nám namnožit studovaný úsek a můžeme přistoupit k přečištění PCR fragmentu (protokol E2). V opačném případě bude třeba PCR reakci optimalizovat (zvýšit teplotu annealingu, zvýšit koncentraci  $MgCl_2$ , přidat tzv. aditiva... viz obecné poznámky k PCR).

## **E2. protokol na přečištění PCR produktu pomocí GENOMED Jetquick kitu**

PCR produkt přečistíme pomocí kitu GENOMED Jetquick PCR Purification Spin Kit, který využívá tzv. „*spin column technique*“, tj. kolonky s membránou na kterou se váže DNA. Alternativou je QIAquick PCR Purification Kit (Qiagen), případně můžeme přečistit fragment vyřízlý z gelu pomocí MinElute Gel Extraction Kit (Qiagen).

- ke zbylým 15  $\mu$ l PCR produktu přidáme 60  $\mu$ l „H1 solution“ a důkladně promícháme na vortexu
- vložíme kolonku do označené (!) 2 ml eppendorfky bez víčka (z kitu), přepipetujeme do ní směs z předchozího bodu
- centrifugujeme 1 minutu při maximálních otáčkách
- vylijeme to, co proteklo kolonkou (DNA zůstala navázána na membráně)
- vložíme kolonku zpátky do 2 ml eppendorfky a přidáme 500  $\mu$ l „H2 solution“
- centrifugujeme 1 minutu při maximálních otáčkách
- vylijeme to, co proteklo kolonkou (DNA navázaná na membráně se promyla)
- vložíme kolonku zpátky do 2 ml eppendorfky a centrifugujeme 1 minutu při maximálních otáčkách
- vložíme kolonku do nové (a označené!) 1.5 ml eppendorfky a přímo do středu membrány napipetujeme 30  $\mu$ l 1  $\times$  TE bufferu předehřátého na 60 °C (např. v eppendorfce na termobloku)
- necháme stát 5-10 minut
- centrifugujeme 2 minuty při maximálních otáčkách
- DNA dále skladujeme při -20 °C

Pro podrobnější návod a další poznámky k přečištění PCR fragmentů viz návod ke kitu.

Změříme koncentraci DNA pomocí fotometru.

### E3. cyklické sekvenování

Sekvenační reakci provádíme pomocí kitu BigDye Terminator v3.1 Cycle Sequencing Kit od firmy ABI. Napipetujeme sekvenační reakci dle tabulky (viz poznámky pod tabulkou):

	celá	1/2	1/4
Ready Reaction Mix	8 µl	4 µl	2 µl
primer	3.2 pmol	3.2 pmol	3.2 pmol
templátová DNA	....	....	....
Sequencing Buffer 5×	0	2 µl	4 µl
sterilní voda	....	....	....
celkový objem	20 µl	20 µl	20 µl

- celkový objem sekvenační reakce je 20 µl, z toho 7 µl tvoří Ready Reaction Mix, Buffer a primer, vody a PCR produktu je zbylých 13 µl
- v případě 1/2 a 1/4 reakce šetříme sekvenační premix a část jí nahrazujeme sekvenačním pufrům, použijeme v případě sekvenování bezproblémového templátu
- primer již je naředěn na 3.2 pmol/µl, tj. přidáváme 1 µl
- množství templátové DNA je podle následující tabulky (nejběžněji sekvenujeme fragmenty okolo 1000 bp a používáme 15 ng DNA), množství PCR produktu v µl tedy závisí na naměřené koncentraci:

templát (PCR produkt)	množství
100-200 bp	1-3 ng
200-500 bp	3-10 ng
500-1000 bp	5-30 ng
1000-2000 bp	10-40 ng

- pro každý vzorek podle koncentrace PCR produktu spočítáme kolik µl je potřeba, aby v reakci bylo přítomno např. 15 ng DNA a podle toho upravíme množství vody pro příslušný vzorek
- připravíme směs z Ready Reaction Mix, Bufferu a primeru pro příslušný počet vzorků (pracujeme na ledu, obsahuje polymerázu), promícháme na vortexu, krátce stočíme

- rozpipetujeme směs do PCR zkumavek nebo do stripů a ke každému vzorku přidáme spočítané množství vody a PCR produktu
- promícháme na vortexu, lehce stočíme v minicentrifuze a vložíme do PCR cycleru s následujícím programem:

96 °C	0:10	} 25 ×
50 °C	0:05	
60 °C	4:00	
10 °C	hold	

#### E4. precipitace produktu

- produkt sekvenační reakce přepipetujeme do popsaných 0.5 ml eppendorfek
- ke 20 µl produktu přidáme 2 µl 3M NaOAc (acetát sodný) a 50 µl 96% ethanolu (z mrazáku)
- promícháme a necháme stát při laboratorní teplotě 10-15 minut
- centrifugujeme v chlazené centrifuze 30 minut při 13 200 rpm a 22 °C
- opatrně(!) odpipetujeme supernatant (precipitát není vidět), pipetujeme po druhé straně eppendorfky než má být pellet
- přidáme 250 µl 70% ethanolu (z mrazáku), promícháme a centrifugujeme v chlazené 15 minut při 13 200 rpm a 22 °C
- zopakujeme předchozí dva body
- otevřené zkumavky s odpipetovaným supernatantem necháme 5 minut stát při laboratorní teplotě v zapnutém laminárním boxu
- vysušíme 5-10 minut při 65 °C (na termobloku), ubrouskem otřeme zbytky ethanolu ze svrchní části zkumavky
- ve vysušené formě je produkt připraven k odnesení do Laboratoře sekvenace DNA (Viničná 5, 2.suterén), produkt v této formě vydrží 1-2 týdny při -20 °C

## protokol F: AFLP (Amplified Fragment Length Polymorphism)

Metoda AFLP je určena ke zjištění polymorfismu v celkové DNA. Je založena na restrikci DNA pomocí dvou restrikčních endonukleáz a následné selektivní amplifikaci pouze určité části fragmentů. Metodiku můžeme rozdělit do několika částí:

1. *restrikce* celkové DNA – dvojicí restrikčních endonukleáz (typicky *EcoRI* a *MseI*) je DNA rozštěpána na mnoho fragmentů, které na svých koncích mají *EcoRI* nebo *MseI* restrikční místo. Restrikce je při dodržení standardních podmínek (teplota, pH) velmi specifická a reprodukovatelná.
2. *ligace* – na konce fragmentů jsou pomocí enzymu T4 ligázy připojeny tzv. adaptory, tj. krátké úseky dvouřetězcové DNA o známé sekvenci, které mají jeden konec komplementární k fragmentům rozštěpeným restrikčními endonukleázami. Výsledkem jsou tedy fragmenty, jejichž počáteční a koncová sekvence je známa.
3. *preselektivní amplifikace* (preamplifikace) – pomocí dvou primerů, jejichž sekvence je komplementární k sekvenci adaptorů + 1 tzv. selektivní baze, je zajištěna PCR amplifikace pouze 1/16 restrikčních fragmentů, které mají na jednom konci *EcoRI* a na druhém *MseI* adaptor.
4. *selektivní amplifikace* – PCR reakce s tzv. selektivními primery (tj. primery se dvěma až čtyřmi selektivními bazemi), dochází k další redukci počtu fragmentů na optimální počet, který je možné vyhodnocovat. Nejčastěji se používají kombinace primerů se třemi selektivními bazemi (Tabulka 2). Zpravidla se selektivní amplifikace pro každý vzorek dělá třikrát, vždy s různou kombinací selektivních primerů a různým fluorescenčním značením. Automatický sekvenátor umí najednou rozeznat čtyři fluorescenční barvy, tou čtvrtou je standard (viz následující bod).
5. *smíchání selektivních PCR produktů a přesrážení produktu s NaOAc* – každý vzorek vznikne smícháním 1  $\mu$ l od každého ze 3 selektivních PCR produktů s primerem značeným vždy jinou fluorescenční barvou. Tento poměr bývá zpravidla nutné „vyladit“ pro každý druh a primerový pár. Na intenzitu má vliv jak použitá fluorescenční barva, tak zvolená kombinace *EcoRI* a *MseI* primerů v kombinaci s příslušnou DNA. Přesrážení produktu ethanolem v přítomnosti solí (acetát sodný) slouží k přečištění PCR produktů. Poté je k produktu přidán fluorescenčně značený žebříček (např. GeneScan-ROX-500). Ten se přidává jako vnitřní standard ke každému vzorku, aby bylo později možné identifikovat délku každého AFLP fragmentu a jednoznačně ji srovnávat s délkou fragmentů v jakémkoliv jiném běhu.



6. *fragmentační analýza na automatickém sekvenátoru* – v kapilárním sekvenátoru ABI 3100 Avant (v sekvenační laboratoři PřF UK) jsou AFLP produkty elektroforeticky rozděleny podle délky.

## F1. restrikce DNA

Pro restrikci a ligaci používáme AFLP<sup>®</sup> Core Reagent Kit I od firmy Invitrogen. Přesné složení jednotlivých částí kitu lze nalézt v bookletu přiloženému ke kitu. Do 1.5 ml eppendorfky si napipetujeme restrikční směs pro příslušný počet vzorků podle návodu. Stále pracujeme na ledu, směs restrikčních enzymů vyndáme z mrazáku až těsně před pipetováním a okamžitě poté opět vrátíme do -20 °C.

- 5× Reaction Buffer ..... 1.0 µl
- *EcoRI/MseI*..... 0.4 µl
- sterilní H<sub>2</sub>O..... 1.1 µl

Směs promícháme na vortexu, krátce zcentrifugujeme a 2.5 µl rozpipetujeme do PCR eppendorfek. Přidáme 2.5 µl DNA (50-100 ng/µl) a inkubujeme 2-3 hodiny při 37 °C v termocykleru nebo v inkubátoru (potom ale musíme do každé zkumavky přidat kapku PCR oleje, aby nedocházelo k odpařování!).

## F2. ligace adaptorů

Pro ligaci si v 1.5 ml eppendorfke připravíme ligační směs pro příslušný počet vzorků (také s využitím AFLP<sup>®</sup> Core Reagent Kit I. Opět pracujeme na ledu, ligační roztok i T4 ligáza jsou velmi náchylné k vyšším teplotám.

- Adapter/Ligation Solution..... 4.8 µl
- T4 DNA Ligase ..... 0.2 µl

Po 5 µl přidáme ke vzorkům po restrikci (pracujeme velmi opatrně, pipetujeme bílou špičkou těsně pod hladinu kapaliny v mikrozkuhavce, při práci stále sledujeme správné napipetování a vypuštění objemu, PCR zkumavky při práci držíme v ruce!). DNA s ligační směsí inkubujeme při 37 °C 2-3 hodiny nebo přes noc (opět nejlépe v termocykleru). Poté vzorky zamrazíme v -20 °C až do další práce.

Zda restrikce proběhla správně otestujeme na 1.8% agarosovém gelu v 1×TBE bufferu při 60V. Nanášíme 3-5 µl produktu po restrikci/ligaci. Na gelu by měl být vidět rozmazaný pruh rozštěpené DNA, především mezi 50-1000 bp.

### F3. preselektivní amplifikace (preamplifikace)

Pro preamplifikaci používáme AFLP® Pre-amp Primer Mix I od firmy Invitrogen. Do 1.5 ml eppendorfky smícháme směs podle následujícího návodu. Opět pracujeme na ledu.

- PA mix..... 4.0 µl
- 10× Buffer for RedTaq..... 0.5 µl
- RedTaq Polymerase (Sigma)..... 0.1 µl

Popíšeme si nové PCR zkumavky a rozpipetujeme do nich 4.6 µl směsi. Přidáme 0.5 µl DNA po restrikci a ligaci z minulého kroku (pracujeme velmi pečlivě s 1 µl pipetou, vizuálně kontrolujeme napipetování a vypuštění malých objemů, PCR zkumavky při práci držíme v ruce!). Zkumavky pečlivě uzavřeme a vložíme do termocykleru s následujícím programem:

72 °C	2:00	} 20 ×
94 °C	0:01	
56 °C	0:30	
72 °C	2:00	
60 °C	30:00	
10 °C	hold	

Správně proběhlou preamplifikaci můžeme opět otestovat na 1.8% agarozovém gelu v 1× TBE bufferu při 60V. Na gel nanese 3 µl neředěného (!) produktu preamplifikační PCR reakce smíchaného se 2-3 µl nanášecí barvičky (loading dye). Po elektroforéze bychom měli vidět opět rozmazaný pruh fragmentů, některé z nich by měly být výraznější.

Preamplifikační PCR reakci naředíme 10 × sterilní H<sub>2</sub>O, promícháme na vortexu a krátce zcentrifugujeme. Zamrazíme v -20 °C pokud hned nepokračujeme dalším krokem.

### F4. selektivní amplifikace

Selektivní amplifikace se provádí třikrát pro každý vzorek, vždy s jinou kombinací primerů a jiným fluorescenčně označeným primerem (6-FAM, HEX, NED). Do tří 1.5 ml eppendorfek si tedy připravíme tři PCR směsi, které se budou lišit pouze primery:

- sterilní H<sub>2</sub>O..... 5.1 µl
- 10× Buffer for RedTaq..... 1.0 µl
- dNTP ..... 0.2 µl
- *EcoRI* primer (1 pmol/µl)..... 0.5 µl
- *MseI* primer (5 pmol/µl)..... 0.5 µl
- (JumpStart) RedTaq DNA Polymerase (1U/µl)..... 0.2 µl

Směs promícháme na vortexu a krátce zcentrifugujeme. Pro kombinaci každého vzorku a barvy si popíšeme PCR zkušavku, případně můžeme pracovat se stripy nebo destičkami. Do každé zkušavky napipetujeme (nejlépe multipipetou) 7.5 µl směsi a přidáme 2.3 µl DNA po preselektivní PCR (pipetováním promícháme). Zkušavky, stripy nebo destičky krátce zcentrifugujeme a vložíme do termocykleru s následujícím programem:

94 °C	2:00		
65 °C	0:30		
72 °C	2:00		
94 °C	0:01	}	8 ×
64 °C	0:30 (v každém kroku -1 °C)		
72 °C	2:00		
94 °C	0:01	}	23 ×
56 °C	0:30		
72 °C	2:00		
60 °C	30:00		
10 °C	hold		

Pro selektivní PCR je využito tzv. touch-down protokolu, který umožňuje zvýšit specifitu amplifikace bez výrazného poklesu výtěžku. Při vyšší teplotě annealingu v prvních devíti cyklech (65-57 °C) se amplifikují pouze specifické produkty (sice s nižším výtěžkem, ale s mnohem větší pravděpodobností než nespecifické). Nižší teplota annealingu v dalších 23 cyklech zvyšuje výtěžek a protože koncentrace specifických produktů je zvýšena díky předchozí fázi reakce, jsou tyto „správné“ fragmenty účinně namnoženy.

#### **F5. přesrážení produktu s acetátem sodným a přidání standardu (ROX)**

- do čistých PCR zkušavek (stripů) přepipetujeme po 3 µl od každého selektivního PCR produktu (případně „vyladěný“ poměr jednotlivých produktů), uzavřeme, promícháme na vortexu, krátce zcentrifugujeme
- na stěnu 1.5 ml eppendorfky napipetujeme 2 µl NaOAc (acetátu sodného)
- přidáme 3 µl směsi tří selektivních PCR produktů

- přidáme 25  $\mu$ l 96% ethanolu (z mrazáku)
- promícháme na vortexu, krátce zcentrifugujeme
- necháme 20 minut při  $-20\text{ }^{\circ}\text{C}$
- centrifugujeme v předchlazené centrifuze 30 minut při  $4\text{ }^{\circ}\text{C}$
- opatrně slijeme supernatant
- přidáme 100  $\mu$ l 70% ethanolu (z mrazáku)
- centrifugujeme v předchlazené centrifuze 5 minut při  $4\text{ }^{\circ}\text{C}$ , supernatant slijeme
- otevřené zkumavky necháme 5 minut stát při laboratorní teplotě v zapnutém laminárním boxu
- vysušíme 5-10 min při  $65\text{ }^{\circ}\text{C}$  (na termobloku)
- vzorky můžeme skladovat 1-2 týdny při  $-20\text{ }^{\circ}\text{C}$

Těsně před odnesením vzorků na sekvenátor pokračujeme následujícím způsobem:

- v 1.5 ml eppendorfci si připravíme směs pro příslušný počet vzorků
  - deionizovaný formamid..... 10.00  $\mu$ l
  - GeneScan-ROX-500..... 0.25  $\mu$ l
- směs promícháme na vortexu a krátce zcentrifugujeme
- přidáme 10  $\mu$ l směsi ke vzorku
- promícháváme na termomixeru 3 minuty při  $95\text{ }^{\circ}\text{C}$  a 300 rpm
- vzorky v daném pořadí přepipetujeme na platíčko do sekvenátoru (optical plate)

## protokol G: Analýza mikrosatelitů

Jako mikrosatelity se označují repetitivní DNA sekvence, které jsou nejčastěji tvořeny opakováním mono-, di-, tri- nebo tetranukleotidů, např.  $(A)_n$ ,  $(AT)_n$ ,  $(ATA)_n$  apod. Počet opakování jednotky (repetice) v konkrétním místě DNA (lokusu) definuje alelu. U diploidních jedinců tak můžeme odhalit dvě alely, u tetraploidních až čtyři atd. Délku alel(y) zjistíme PCR amplifikací daného lokusu pomocí primerů kompatibilních k okolním sekvencím. Délku PCR fragmentů pak zjistíme např. kapilární elektroforézou v automatickém sekvenátoru. Pro analýzu mikrosatelitů tedy potřebujeme znát okolní sekvence.

Mikrosatelity se vyskytují jak v jaderné, tak v chloroplastové DNA. *Jaderné mikrosatelity* jsou často velmi druhově specifické, musíme tedy pracovat s druhem, pro nějž jsou primery již publikované. Jsou velmi variabilní, používají se pro na vnitrodruhové úrovni. Nejčastěji se analyzují lokusy s di- a trinukleotidovými repeticemi. *Chloroplastové mikrosatelity* se naopak vyskytují v místech mezi konzervovanými sekvencemi, tj. na základě tzv. konsenzuálních primerů můžeme určité lokusy amplifikovat pro druhy z mnoha různých čeledí. Nejčastěji se jedná o jednobázové repetice. Jejich variabilita odpovídá variabilitě v cpDNA, jsou využitelné pro hodnocení variability např. mezi druhy.

Protokol můžeme rozdělit do dvou částí:

1. *PCR s dvojicí specifických primerů* – pro amplifikaci konkrétního úseku DNA, který obsahuje repetitivní mikrosatelitovou sekvenci jsou použity dva primery, jeden z primerů je fluorescenčně značený (používáme barvy HEX, NED, 6-FAM). Pro analýzu cpDNA mikrosatelitů zvolíme primery na základě Tabulky 3. V případě jaderných mikrosatelitů je nejlépe použít protokol uvedený v originální publikaci. Seznam rostlinných druhů, pro něž jsou mikrosatelitové primery publikovány je na adrese <http://botany.natur.cuni.cz/dna/primers/publikSSRs.htm>.

Pokud je třeba analyzovat více mikrosatelitových lokusů (často se analyzuje 8-10), je výhodné pokusit se sestavit tzv. multiplex PCR. Je to v podstatě běžná PCR reakce, do které přidáme více primerových páru najednou a amplifikujeme tak více lokusů (až třeba pět) v jedné PCR reakci. Můžeme tak výrazně snížit finanční i časové náklady na analýzu. Pro úspěšnou a dostatečnou amplifikaci všech lokusů musíme zaručit, aby (1) primery měly přibližně stejnou teplotu annealingu, (2) primery v jedné PCR reakci nevytvářely dimery, (3) lokusy s překrývajícím se rozsahem předpokládaných délek fragmentů byly označeny jinou fluorescenční barvou. Otestovat vytváření primerových dimerů a vhodnost kombinace primerových páru můžeme udělat např.

v programu FastPCR (<http://www.biocenter.helsinki.fi/bi/Programs/fastpcr.htm>, freeware). Příklad kombinace primerů do multiplex PCR je v Tabulce 4.

2. *přesrážení produktu s NaOAc a přidání fluorescenčně značeného standardu* – slouží k přečištění PCR produktů. Poté je k produktu přidán fluorescenčně značený žebříček (např. GeneScan-ROX-500). Ten se přidává jako vnitřní standard ke každému vzorku, aby bylo později možné identifikovat délku každého fragmentu (alely) a jednoznačně ji srovnávat s délkou fragmentů v jakémkoliv jiném běhu.
3. *fragmentační analýza na automatickém sekvenátoru* – v kapilárním sekvenátoru ABI 3100 Avant (v sekvenační laboratoři PřF UK) jsou mikrosatelitové alely elektroforeticky rozděleny podle délky.

## G1. PCR amplifikace

Do 1.5 ml eppendorfky si napipetujeme směs pro příslušný počet vzorků podle návodu. Stále pracujeme na ledu, DNA polymerázu vyndáme z mrazáku až těsně před pipetováním a okamžitě poté opět vracíme do -20 °C. Návod je obecný příklad, v případě multiplex PCR se přidají forward a reverse primery pro příslušné kompatibilní lokusy, objem vody se dopočítá do 20 µl (včetně templátové DNA!).

- sterilní H<sub>2</sub>O..... 13.60 µl
- 10× Buffer for RedTaq..... 2.00 µl
- dNTP ..... 0.40 µl
- forward primer (25 pmol/µl) ..... 0.25 µl
- reverse primer (25 pmol/µl) ..... 0.25 µl
- (JumpStart) RedTaq DNA Polymerase (1U/µl)..... 0.50 µl

Promícháme na vortexu, krátce stočíme v centrifuze a po 17 µl rozpipetujeme do PCR zkumavek (případně stripů). Přidáme 3 µl naředěné DNA (5 ng/µl), promícháme na vortexu a vložíme do termocykleru s následujícím programem:

94 °C	1:00	} 35 ×
94 °C	0:30	
T <sub>a</sub> °C	0:30	
72 °C	0:35	
72 °C	15:00	
10 °C	hold	

Po proběhnutí reakce otestujeme úspěšnost amplifikace na 1% agarosovém gelu v TAE bufferu. Na gelu by měl být vidět jeden proužek (nebo i více v případě multiplex PCR) odpovídající předpokládané velikosti alel (např. podle Tabulky 4 nebo originální práce). Podle intenzity fragmentu naředíme PCR produkt 1:2 až třeba 1:20 (ředění je třeba vyzkoušet a podle intenzity peaků ze sekvenátoru případně upravit).

## **G2. přesrážení a přidání standardu**

Přesrážení produktu a přidání standardu (GeneScan-ROX-500) je obdobné jako v případě AFLP (protokol F5).

## B. Analýza primárních dat

### 1. Prohlížení sekvencí pomocí programu FinchTV

Sekvenační data (ve formátu \*.ab1) uložena jsou na serveru (<ftp://botany.natur.cuni.cz/>, user: sequences, přístup např. přes Total Commander) v adresáři „sekvence“ a dále podle data analýzy (ve formátu rr-mm-dd). Jednotlivé stažené sekvence si lze prohlédnout např. v programu FinchTV (freeware, <http://www.geospiza.com/finchtv>). Sekvence by měla být čitelná od cca 30-50 bází do cca 550-650 bází. Pokud nevidíme jednoznačné peaky pro jednotlivé báze, je něco v nepořádku (pravděpodobně se sekvenační reakcí). Pouze kvalitní sekvence můžeme dále analyzovat. V programu FinchTV můžeme sekvence editovat, tj. měnit „base calling“, mazat i vkládat báze. Program dále umožňuje vyhledávat v sekvencích, export do formátu FASTA, tisk chromatogramu a také prohlížení primárních dat (*raw data*).

### 2. Vytvoření contigů v programu SeqMan (DNASTAR)

Tzv. contig vznikne spojením *forward* a *reverse-complement* (tj. komplementární báze v opačném pořadí) sekvence ze stejného úseku (PCR fragmentu). To můžeme automaticky udělat např. v programu SeqMan (součást komerčního balíku DNASTAR Lasergene, <http://www.dnastar.com/web/r13.php>, je možné si on-line vyžádat zkušební verzi programu.). Zde také můžeme editovat evidentně špatný „base calling“ (překlad peaků na jednotlivé báze).

- sekvence vložíme *File* → *New* → *Add Sequences*
- vložíme *forward* i *reverse* sekvence od jednoho vzorku (dvojklikem nebo označíme + „Add“)
- klikneme na „*Assemble*“, potom dvojklik na *Contig 1* (pokud jsou sekvence v pořádku a dostatečně se překrývají, měl by se vytvořit pouze jeden contig)
- zkontrolujeme, zda je *forward* sekvence zleva doprava, pokud není, tak zvolíme *Contig* → *Complement Contig*
- rozbálíme všechny sekvence kliknutím na trojúhelníček vlevo od názvu
- zobrazení začátku i konce každé sekvence můžeme měnit posunováním konců myši
- překontrolujeme „base calling“ v řádku *Translate* (consensus sekvence), případné jasné chyby editujeme (přepíšeme)



- editovaný a upravený contig uložíme ve formátu programu SeqMan (\*.sqd), může se hodit pro pozdější kontrolu
- označíme myší a zkopírujeme (Ctrl+C) horní řádek (*Translate*), případně jeho část
- vložíme do textového souboru (např. do poznámkového bloku), název sekvence uvedeme ve formátu FASTA, např.

```
>sekvence1
ATTGCGTATGCCTGTCAT
```

- toto provedeme pro všechny sekvence a textový soubor uložíme s koncovkou \*.fas

Alternativně lze contig vytvořit ručně nebo s využitím webové verze programu CAP3 (<http://pbil.univ-lyon1.fr/cap3.php>).

### **3. Analýza AFLP vzorků ze sekvenátoru pomocí programu GeneScan Analysis v3.7.1**

Analýzou dat z fragmentační analýzy se rozumí identifikace jednotlivých peaků markeru (GS-500 ROX) a následné přiřazení délky (v bazích) všem identifikovaným peakům v jednotlivých barvách. Teprve zanalyzovaná data je možno dále načíst do programu Genographer.

Data ve formátu \*.fsa jsou uložena na serveru (<ftp://botany.natur.cuni.cz/>, user: sequences) v adresáři „fragmentace“ a dále podle data analýzy (ve formátu rr-mm-dd) a jsou automaticky analyzována již na počítači sekvenátoru. Výsledky analýzy jsou shrnuty v \*.log souboru. Abychom se mohli přesvědčit o správnosti analýzy a případně zpracovat vzorky, které nebyly automaticky zanalyzovány, je třeba použít program GeneScan Analysis. Následující text shrnuje nejdůležitější kroky při analýze, podrobnější údaje lze nalézt v manuálu k tomuto programu.

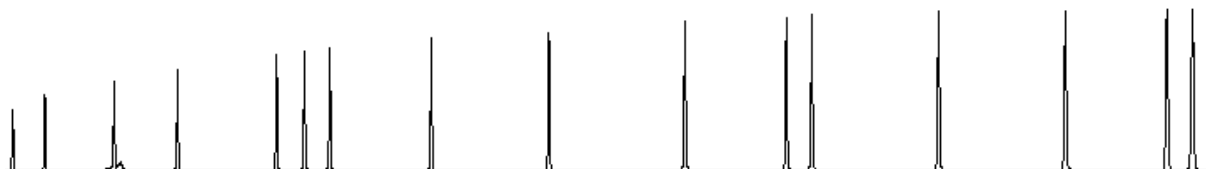
#### *a) Vytvoření projektu, analýza*

- *File* → *New* → *Project*
- *Project* → *Add Sample Files* → nalistovat příslušné vzorky, označit → *Add* (případně *Add All*) → *Finish*
- zobrazí okno „*Analysis Control*“ se seznamem vzorků s příslušnými „*Size Standard*“ a „*Parameters*“ – typicky se používá GS500-250.szs a GS500Analysis.gsp (GS500.szs použije všechny peaky standardu GS-500 ROX, GS500-250.szs znamená vynechání peaku o délce 250 bazí, který údajně nemigruje konzistentně v kapilárních sekvenátorech)

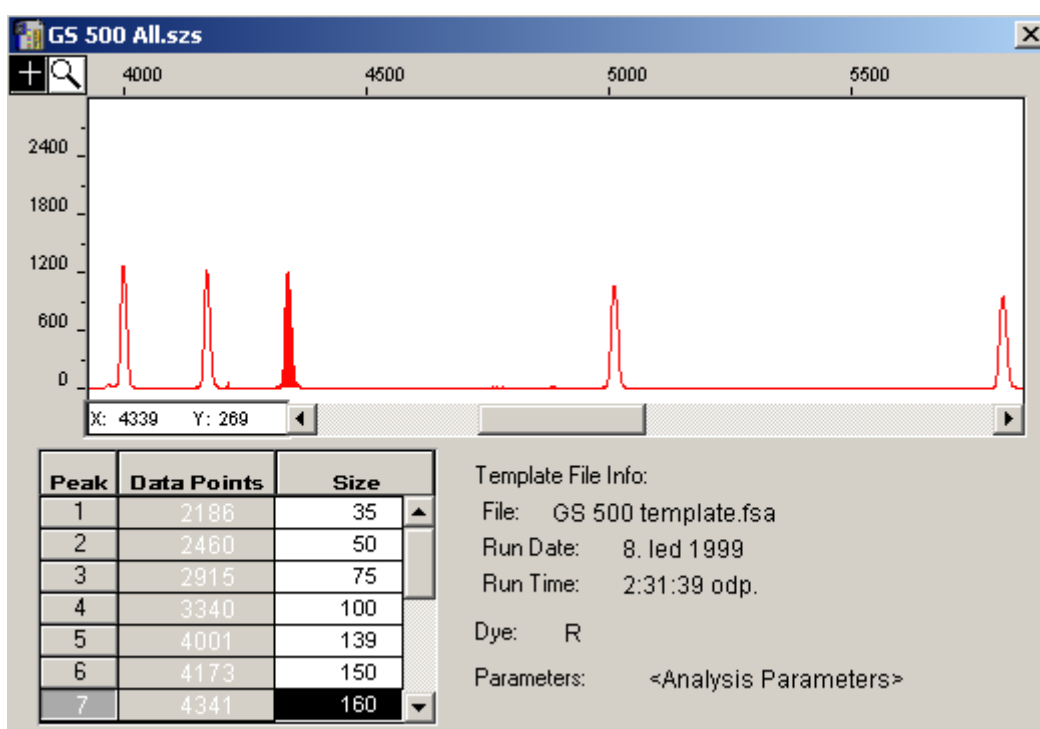
- pro všechny vzorky najednou můžeme určit standard nebo parametry vybráním příslušného typu po kliknutí na šipku v záhlaví nebo můžeme pro každý vzorek určit standard a parametry zvlášť
- na primární data se můžeme podívat přes *Sample* → *Raw Data*
- vzorky, které chceme analyzovat označíme kliknutím na jejich číslo (ve sloupci úplně vlevo), u vzorku se vybarví sloupce B,G,Y,R. Všechny vzorky najednou můžeme označit kliknutím na záhlaví sloupce s čísly
- klikneme na *Analyze* – program začne analyzovat jednotlivé vzorky, úspěšná analýza je indikována odbarvením políček
- po skončení analýzy vyskočí okno s informací o jejím průběhu (*Analysis Log*) – je zde uvedeno, u kterých vzorků analýza proběhla (ne)úspěšně a v jakém rozsahu se podařilo identifikovat marker a přiřadit peakům jejich správnou délku
- vzorky, které se nepodařilo analyzovat předvoleným standardem a parametry bude třeba zanalyzovat s jiným nastavením, tj. definovat jiný „*Size Standard*“ a „*Parameters*“ (nejlépe přímo podle vzorku, který chceme analyzovat – viz dále část „b“ a „c“)
- po úspěšné analýze všech vzorků je možno se přesvědčit, jestli analýza proběhla správně, tj. jestli byly peaky standardů správně přiřazeny (viz dále – část „d“)
- po kontrole analýzy uložíme celý projekt (*File* → *Save Project As...*) a zavřeme
- do všech vzorků je zanesena informace o délkách jednotlivých peaků ve všech barvách a jsou připraveny k dalšímu hodnocení např. v programu Genographer

#### b) *Definování nového standardu („Size Standard“)*

- kliknutím na šipku ve sloupci „*Size Standard*“ u příslušného vzorku vybereme volbu „*Define New*“
- otevře se okno, kde jsou zobrazeny peaky standardu (červená barva) v příslušném vzorku a dole v tabulce potom jejich definované délky
- všechny peaky standardu mají typický tvar („Gaussovská křivka“) a zhruba stejnou intenzitu (výšku), ostatní peaky (šum, nespecifity...) jsou většinou „ostřejší“ a jsou vyšší nebo nižší než správné peaky
- peaky standardu GS-500 ROX jsou (zleva doprava): 35, 50, 75, 100, 139, 150, 160, 200, 250, 300, 340, 350, 400, 450, 490, 500 bází



- kliknutím na ikonku lupy a posléze do obrázku je možno horizontálně zoomovat
- kliknutím na ikonku kříže a poté na jednotlivé peaky (vyplní se barvou) přiřadíme všem peakům standardu správnou délku, případným ostatním peakům dáme „0“, stejně tak peaku o velikosti 250 bazí (viz poznámka výše)



- nově definovaný standard uložíme (*File* → *Save As...*) do adresáře „\AppliedBio\Shared\Analysis\Sizecaller\SizeStandards“
- nově definovaný standard je nyní k dispozici při výběru

c) *Definování nových parametrů analýzy („Analysis Parameter“)*

- kliknutím na šipku ve sloupci „Parameters“ u příslušného vzorku vybereme volbu „Define New“
- otevře se okno, kde je možno definovat různé parametry analýzy, zde jsou zmíněny jen některé (blíže viz manuál k programu)

- *Analysis Range* – možno specifikovat jen určitou část vzorku (např. s vynecháním úvodní části s nespecifickými peaky), ve které budou hledány peaky standardu; zadává se v tzv. Data Points, požadovaný rozsah je možné zjistit při zobrazení RawData
- *Peak Amplitude Thresholds* – pokud jsou peaky standardu příliš malé, je možné snížit práh jejich detekce uvedením menšího čísla u R:
- *Min. Peak Half Width* – zvýšením tohoto čísla eliminujeme detekci příliš úzkých peaků
- nově definovaný soubor parametrů uložíme (*File* → *Save As...*) do adresáře „\AppliedBio\Shared\Analysis\Sizecaller\Params“
- nově definovaný parametr je nyní k dispozici při výběru

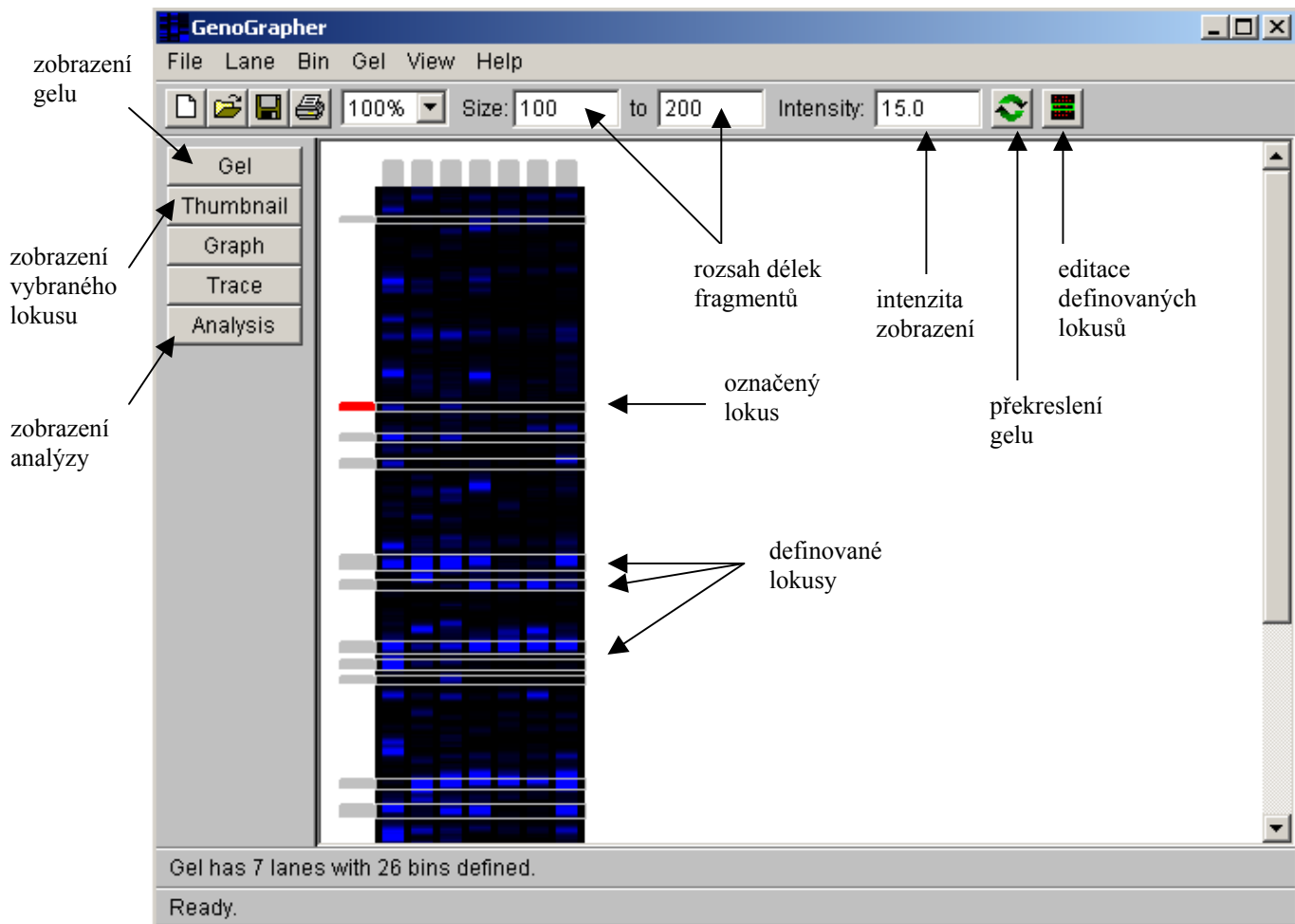
#### d) *Kontrola správnosti analýzy*

- kontrolou analýzy se rozumí ověřit si, zda byly peakům standardu přiřazeny správné délky a jestli se tak stalo stejně u všech analyzovaných vzorků
- pokud si tedy zobrazíme peaky standardu z několika vzorků do jednoho grafu, měly by se překrývat
- to lze udělat pomocí *Windows* → *Results Control*
- zobrazí se okno, kde je v levé části seznam vzorků a v pravé jejich rozdělení do tzv. panelů, jejich počet lze zvolit (max. 8)
- nejjednodušeji a nejrychleji zobrazíme peaky standardu z různých vzorků přes sebe zaškrtnutím volby „*Quick Tile*“ na „*On*“ a kliknutím na písmeno R v červeném poli – program automaticky rozdělí všechny vzorky pravidelně do zvoleného počtu panelů
- klikneme na „*Display*“, zobrazí se několik řádků peaků standardu (v každém řádku je několik vzorků, podle toho, kolik jsme zvolili panelů)
- *View* → *Align By Size* – zarovná peaky podle jejich délky v bazích
- vizuálně zkontrolujeme, jestli se všechny peaky standardů ze všech vzorků překrývají (alespoň v úseku 75-500 bazí)
- pokud je peakům u některého vzorku přiřazena evidentně špatná délka (evidentně správný peak se nepřekrývá s peaky téže délky jiných vzorků), můžeme vzorek identifikovat kliknutím na peak – v tabulce (ve spodní části okna) jsou vysvíceny údaje o něm, včetně jména vzorku; tento vzorek bude pravděpodobně nutné znovu zanalyzovat
- pokud byly všem peakům přiřazeny správné délky, tj. jednotlivé vzorky se překrývají, je možné projekt uložit, zavřít a pokračovat v analýze vzorků např. v programu Genographer

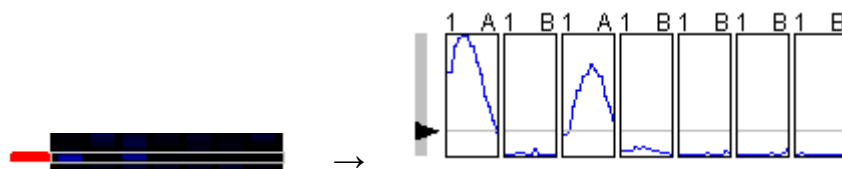
#### 4. Vyhodnocování AFLP dat v programu Genographer

Data ze sekvenátoru, která byla analyzována (automaticky na počítači sekvenátoru nebo manuálně v programu GeneScan), můžeme pro jednoduché vyhodnocování importovat do programu Genographer. Program je freeware (<http://hordeum.oscs.montana.edu/genographer>) a k jeho běhu je potřeba mít nainstalováno JRE (Java Runtime Environment, <http://java.sun.com/j2se/1.4.2/download.html>; stahujte J2SE JRE, nikoliv SDK!). Program Genographer rozbalíme do složky C:\genographer. Na ploše vytvoříme nového zástupce, který bude odkazovat na soubor javaw.exe. Ten příkazem -cp spustí Genographer. „Cíl“ zástupce tedy bude např. "C:\Program Files\Java\j2re1.4.2\_05\bin\javaw.exe -mx64m -cp C:\genographer genographer", Volba „Spustit v“ bude "C:\genographer". Nyní by se spuštěním zástupce měl spustit Genographer. (Příkazem -mx64m vyhradíme programu 64 MB paměti pro práci, hodnotu je možné změnit, u mnoha vzorků při větším zvětšení je to dokonce nutné.)

Data (soubory \*.fsa) naimportujeme do programu volbou *File* → *Import*. Příslušné soubory vybereme dvojklikem, případně můžeme importovat celý obsah adresáře volbou „*Add All*“. Files of Type vybereme jako ABI 3100/3700. V dalším okně zvolíme barvu, kterou chceme importovat a nastavíme *Size Standard* na „GS-TAMRA-36A-2400“. Soubory se naimportují a dále s nimi můžeme pracovat podle následujících voleb:

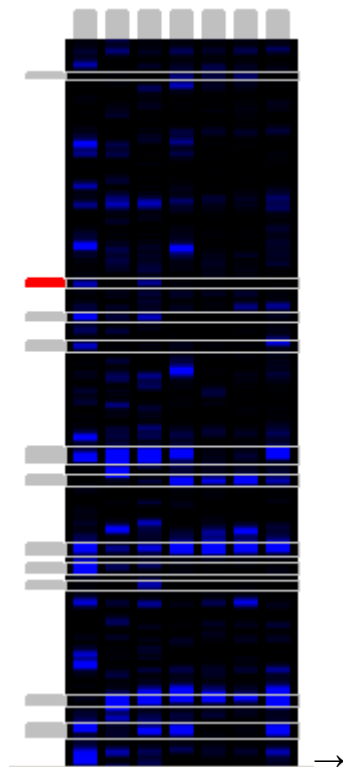


Jednotlivé lokusy, které považujeme za spolehlivé (můžeme tedy jednoznačně určit, zda fragment je či není přítomen), skorujeme nezávisle na jejich intenzitě. Lokus vybereme myší přímo v gelu, můžeme upravovat šířku a polohu fragmentu. O správném skorování se můžeme přesvědčit vybráním fragmentu vlevo od gelu (šedé záhlaví zčervená) a kliknutím na „Thumbnail“. Zobrazí se peaky pro jednotlivé vzorky, posunováním černého trojúhelníčku můžeme definovat hranici pro skorování (A jako přítomnost, B nepřítomnost).



Takto definujeme veškeré lokusy, je výhodné si celý rozsah zobrazovat např. po 30 bp a ve větším rozlišení (např. 200%). Skorujeme zpravidla fragmenty v rozsahu 100-500 bp, fragmenty kratší mají mnohem větší pravděpodobnost být nemohologní, fragmenty delší již

nemají správně přiřazenou délku. Poté si můžeme výsledek analýzy zobrazit kliknutím na „Analysis“, kdy se ukáže výsledek skorování jednotlivých lokusů.




```
131.0 +/- 0.64  
Mean: 130.96 StdDev: 0.15 #A: 4 #B: 3 from 4/4 peaks  
ABAABBA  
  
133.92 +/- 0.44  
Mean: 134.0 StdDev: 0.28 #A: 6 #B: 1 from 6/6 peaks  
BAAAAA  
  
145.38 +/- 0.39  
Mean: 145.59 StdDev: 9.223372036854776E16 #A: 1 #B: 6 from 1/1 peaks  
BBABBBB  
  
147.22 +/- 0.43  
Mean: 147.29 StdDev: 0.09 #A: 3 #B: 4 from 3/3 peaks  
AAABBBB  
  
149.09 +/- 0.58  
Mean: 149.19 StdDev: 0.05 #A: 7 #B: 0 from 7/7 peaks  
AAAAAAA  
  
155.87 +/- 0.46  
Mean: 155.89 StdDev: 0.03 #A: 5 #B: 2 from 5/5 peaks
```

Data je možné zkopírovat např. do Wordu, kde jednoduše nahradíme „A“ za „1“ a „B“ za „0“ . V Excelu potom datovou matici transponujeme, abychom měli v řádcích druhy a ve sloupcích jednotlivé lokusy.

Soubor ve formátu Genographer (včetně definovaných lokusů) je možné si uložit příkazem *File* → *Save As...* (pokud zadáme pouze „*Save*“ program přepíše předchozí uložený soubor, protože neobsahuje dodatečný dotaz na přepsání!).

## 5. Vyhodnocování mikrosatelitových dat v programu GeneMarker

Mikrosatelitová data ze sekvenátoru můžeme analyzovat a velmi názorně vizualizovat v programu GeneMarker (<http://www.softgenetics.com/genemarker.html>, komerční program, z webu je možné stáhnout zkušební verzi nebo demoverzi, která neumožňuje ukládat).

- data nainportujeme pomocí *File* → *Open Data*, v dalším okně potom *Add...* (více souborů najednou můžeme vybrat při stisknutí Ctrl nebo Shift, potom *Open* a *OK*)
- analýzu dat (kalibraci podle vnitřního standardu) spustíme pomocí *Project* → *Run*, jako *Template* zvolíme *ABI*, jako *Size Standard* „GS500\_1“ (akceptuje všechny peaky standardu GS-500 ROX kromě peaku o délce 250 bází), v dalším okně můžeme nastavit nejrůznější parametry automatické analýzy, můžeme nechat default beze změn, v dalším okně nic neměníme a odklikneme *OK*
- proběhne analýza všech vzorků, v okně si můžeme zkontrolovat, zda se u všech vzorků objevilo „*Completed*“. Pokud je zde „*Failed*“ vzorek pravděpodobně nebude možno použít. Okno se zprávou o analýze zavřeme.
- v levé části obrazovky je seznam vzorků, kompletně analyzované vzorky jsou zeleně, problémové vzorky žlutě, neanalyzované vzorky červeně
- nejlepším způsobem, jak vyhodnotit mikrosatelitová data, je zobrazit si jednotlivé vzorky v příslušných lokusech pod sebou a odečíst jednotlivé alely. Klikneme pravým tlačítkem na první vzorek a vybereme „*Select Max*“. Až 32 prvních vzorků se zobrazí pod sebou ve spodním prostředním okně. Několikerým kliknutím na tlačítko  vybereme jednu ze tří barev, které chceme vyhodnocovat. Klikáním na ikonu lupy a pomocí posuvníků zobrazíme v dostatečném zvětšení příslušný lokus.
- s přihlédnutím k následujícím poznámkám můžeme odečíst alelickou kompozici u jednotlivých vzorků pro daný lokus a zapsat si ji např. do excelové tabulky

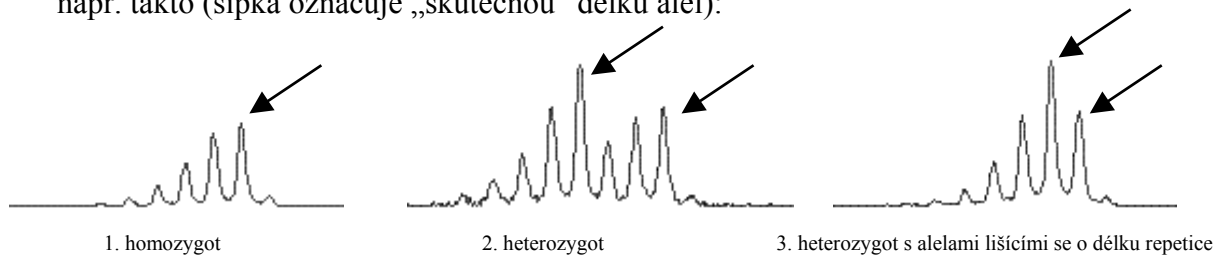
### *Poznámky k vyhodnocování mikrosatelitových pattern*

1. Diploidi mohou mít v jednom lokusu max. 2 alely, pokud skorujeme pouze jednu, jedná se o homozygota. Tetraploidi mohou mít v jednom lokusu až 4 alely, pokud skorujeme



čtyři nebo jednu, lze jednoznačně určit genotyp. Pokud ale skorujeme např. tři alely, jedna z nich je přítomna dvakrát, ale nelze jednoznačně říci, která to je. Jsme tedy schopni pouze konstatovat prezenci alely, ale už nikoliv genotyp. U vyšších ploidii je problém analogicky ještě komplikovanější.

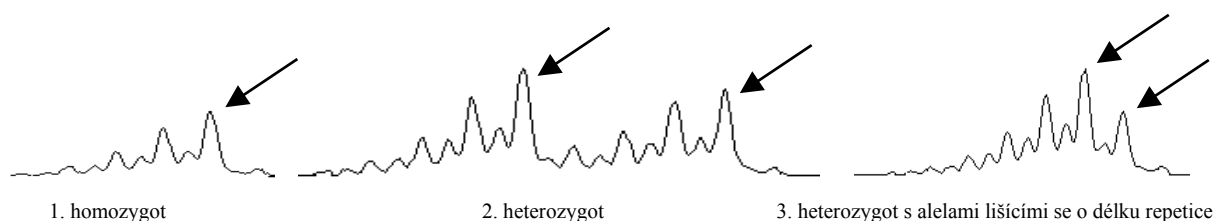
2. Pro některé lokusy je příznačný výskyt tzv. „stutter bands“, tj. peaků, které jsou o dvou-, tří-, čtyř- atd. násobek repetice kratší než je skutečná délka alely. Vznikají díky sklouznutí DNA polymerázy při replikaci (tzv. *in vitro DNA slippage*). Typická pattern vypadají např. takto (šipka označuje „skutečnou“ délku alel):



Pokud je jedinec homozygot (1), je skorování jednoduché – konzistentně skorujeme peak nejvíce vpravo (nejvyšší). V případě heterozygotů, jejichž alely se od sebe liší alespoň o dvě délky repetice (2) je situace také jednoduchá – skorujeme dva nejvyšší peaky, za nimiž vždy následuje peak o hodně nižší. Pokud se ale alely liší pouze o délku repetice (3), skorujeme dvě alely nejvíce vpravo. Výška druhého peaku zprava je dána i stutter produktem alely o repetici delší.

Stutter produkty neznemožňují čtení mikrosatelitového pattern, naopak mohou někdy pomoci identifikovat skutečné alely od nespecifit či pozadí (skutečné alely budou mít vždy podobné stutter bands pattern).

3. Dalším fenoménem, který komplikuje interpretaci mikrosatelitových pattern jsou tzv. „-A“ produkty. *Taq* polymeráza totiž přidává ke každému produktu adenin (tzv. terminal transferase activity) a prodlužuje tak jeho délku o 1 bazi. Proto se závěrečná doba elongace prodlužuje na 10-15 minut, aby byl adenin přidán ke všem produktům a výsledky byly konzistentní. Přesto pattern některých lokusů vykazuje „-A“ produkty. Většinou to nebývá na závadu, jen je třeba konzistentně skorovat vždy pravou, tj. „+A“ alelu. Příklady pattern s „-A“ produkty (šipka označuje „skutečnou“ délku alel):



# C. Vyhodnocování kodominantních dat

## 1. Analýza mikrosatelitových dat pomocí programu MSA

Pomocí programu MSA ([http://i122server.vu-wien.ac.at/MSA/MSA\\_download.html](http://i122server.vu-wien.ac.at/MSA/MSA_download.html), freeware) můžeme (pro diploidy!) počítat mimo jiné následující populačně-genetické parametry:

- popisné statistiky pro populaci a lokus (počet alel,  $H_O$ ,  $H_E$ ,  $H_{Sh}$ ,  $F_{IS}$ ...)
- vzdálenosti mezi jedinci a populacemi ( $D$  – standard genetic distance,  $(\delta\mu)^2$ ,  $D_{ps}$  – proporce sdílených alel,  $D_{kf}$  – kinship koeficient...)
- $F$ -statistika ( $F_{ST}$ ,  $F_{IS}$ ,  $F_{IT}$  – globálně i párově mezi populacemi)

Kromě toho program ohlásí varování, pokud by v datech mohla být chyba (alely neodpovídající násobku repetice, příliš velké vzdálenosti mezi alelami, odlehlé alely...)

Vstupní soubor pro MSA je nejlépe zformátovat např. v Excelu a uložit jako TAB-delimited text. Vstupní soubor může vypadat třeba takto (pro diploidní druh):

		název lokusu		délka repetice				délka repetitivní oblasti				
2		2	2	2	2	2	2	2	2	2	2	
		64	74	46	46	24						
		NLGA1	NLGA2	NLGA3	NLGA4	NLGA5						
A	d	1	160	160	86	96	142	142	198	198	100	100
A	d	1	166	166	86	86	152	152	198	198	100	100
A	d	1	166	166	86	86	152	152	198	198	100	100
A	d	1	166	166	86	86	152	152	198	198	100	100
A	d	1	166	166	86	86	152	152	198	198	100	100
A	d	1	166	166	86	86	152	152	198	198	100	100
A	d	1	160	166	86	86	152	152	198	198	100	100
B	d	1	166	166	86	96	150	150	196	198	100	100
B	d	1	160	166	84	84	150	150	196	198	100	104
B	d	1	160	166	92	92	150	152	196	198	100	100
B	d	1	160	166	92	92	150	152	196	198	100	100
B	d	1	166	166	90	92	150	150	198	198	100	100
B	d	1	166	166	82	82	150	152	200	200	100	100
B	d	1	160	162	86	96	nd	.	198	198	94	100
D	d	2	152	160	86	96	152	152	198	198	100	100
D	d	2	152	162	92	96	152	152	198	198	94	100
D	d	2	160	160	-1	-1	150	150			100	100

Annotations:

- ↑ jedno- nebo dvousloupcový formát (pointing to the first two columns)
- ↑ název lokusu (pointing to the column headers NLGA1-5)
- ↑ délka repetice (pointing to the first two columns of the data rows)
- ↑ délka repetitivní oblasti (pointing to the last two columns of the data rows)
- ↑ outbrední (d) nebo inbrední (h) individuuum (pointing to the first two columns of the data rows)
- ↑ číslo skupiny populací (pointing to the third column of the data rows)
- ↑ chybějící data (pointing to the 'nd' and '.' values in the data rows)
- ↑ název populace (pointing to the first column of the data rows)

Analyzovaný soubor uložíme do stejného adresáře jako je MSAnalyzerMr.exe, který potom spustíme. Otevře se okno pro zadání vstupního souboru a parametrů analýzy. Vložením (i) zadáme vstupní soubor (včetně koncovky!). Pod (d) a následně (p) můžeme nastavit nejružnější genetické vzdálenosti (1)-(9) a také nastavit výpočet vzdáleností mezi populacemi (c) a jedinci (i), případně zapnout bootstrapové testování (n). Napsáním (b) se dostaneme zpátky do „distance menu“. Pod (s) lze nastavit parametry  $F$ -statistik, pomocí (c) zapneme výpočet  $F$ -statistik, pomocí (g) zvolíme, zda budou počítány globálně, párově nebo obojí. Pomocí (m) se vrátíme do hlavního menu, kde pomocí (c) lze nastavit např. vytvoření vstupního souboru pro program Arlequin (3). Analýzu lze kdykoliv spustit pomocí (!). Pokud nebyl nikde zadán bootstrap, proběhne analýza velmi rychle. Na základě nastavených parametrů program vytvoří řadu Excelových tabulek a textových souborů a uloží je do přehledné adresářové struktury:

- Allelecount – počty a frekvence alel pro jednotlivé lokusy a populace
- Distance\_data – textové soubory s maticemi vzdáleností mezi jedinci
- Formats&Data – vstupní soubory pro Arlequin a další software
- F-Statistic –  $F$ -statistika globální i párová
- Group\_data – informace o parametrech podle zadaných skupin populací
- Single\_data – informace o parametrech pro jednotlivé populace

Pro další informace o parametrech, metodách výpočtu a interpretaci výstupních souborů viz manuál ([http://i122server.vu-wien.ac.at/MSA/info.html/MSA\\_info.html](http://i122server.vu-wien.ac.at/MSA/info.html/MSA_info.html)).

## **2. Analýza molekulární variance (AMOVA) mikrosatelitových dat v programu Arlequin**

V programu Arlequin v 3.1 (<http://cmpg.unibe.ch/software/arlequin3>, freeware) lze vypočítat např. analýzu molekulární variance (AMOVA). Jako vstupní data použijeme soubor vygenerovaný programem MSA (viz předchozí část). Soubor s koncovkou \*.arp otevřeme pomocí „Open project“, na kartě „Project“ jsou zobrazené jednotlivé populace (*Samples*) a skupiny (*Groups*). Na kartě „Settings“ postupně zaškrtneme, co všechno chceme spočítat, pro AMOVA je to *Genetic structure* → AMOVA. Můžeme zvolit metodu výpočtu matice vzdáleností mezi jedinci: *number of different alleles* (počet rozdílných alel, tj.  $F_{ST}$  analog) nebo *sum of squared differences* (součet čtverců rozdílů mezi délkami alel, tj.  $R_{ST}$  analog). Analýzu spustíme kliknutím na tlačítko „Start“. Program vygeneruje webovou stránku s výsledky, ve kterých se dá přehledně orientovat pomocí menu v levé části (v IE je třeba

povolit spuštění aktivního obsahu!). Jednotlivé analýzy jsou v adresáři podle data a hodiny spuštění, na začátku každé analýzy je vypsáno nastavení, včetně geografické struktury populací a metody výpočtu vzdáleností. Interpretace výsledků AMOVA je podobná jako u analýzy dominantních dat.

# D. Vyhodnocování dominantních dat

## 1. Tvorba stromů pomocí programu TREECON

Program se používá pro konstrukci stromů na základě indexů podobnosti clusterovými metodami (např. UPGMA) nebo metodou neighbor-joining (NJ). Program je freeware [http://bioinformatics.psb.ugent.be/psb/Userman/treecon\\_userman.html](http://bioinformatics.psb.ugent.be/psb/Userman/treecon_userman.html)

Vstupní data je třeba zformátovat následujícím způsobem (a uložit nejlépe s koncovkou .seq).

```
[počet lokusů (proužků) ]
sample1
1011101...
sample2
1011101...
...
```



Po spuštění programu se zobrazí okno s několika tlačítky. Nejprve je třeba spočítat koeficienty podobnosti pro všechny páry vzorků (*Distance estimation*). Klikneme na toto tlačítko a posléze na *Start distance estimation*. Vybereme požadovaný soubor, zaškrtneme *Sequence Type* jako *RFLP/AFLP/RAPD data* a *Sequence format* jako *TREECON*. Nyní máme možnost vybrat vzorky pro analýzu, případně vybereme všechny (*Select all*). Jako koeficient podobnosti vybereme *Nei and Li (1979)*, dále můžeme definovat znaky, na základě kterých budeme index počítat (standardně *Columns* → *All*) nebo počítat *Bootstrap*. Dalším krokem je výpočet stromu (*Infer tree topology*). Máme možnost vybrat shlukovací metodu (*Neighbor-joining* nebo *Clustering*, jehož kritérium v dalším kroku upřesníme – nejčastěji *UPGMA*). Před zobrazením je třeba strom zakořenit (*Root unrooted trees*), např. uprostřed nejdelší vzdálenosti (*midpoint rooting*) nebo zvolením tzv. outgroup (*Single sequence*). Strom nakreslíme po stisku *Draw phylogenetic tree* a kliknutím na ikonu bílého papíru vlevo nahoře (*Load new tree*). V tomto grafickém editoru je možno strom všelijak upravovat. Pokud chceme znát i bootstrapovou podporu jednotlivých větví stromu, provedeme všechny kroky ještě jednou a v každém zaškrtneme *Bootstrap* → *Yes*. Více viz nápověda k programu a skriptu „Analýza multivariačních dat v taxonomii (Marhold & Suda)“.

## 2. PCoA pomocí programu SYNTAX 2000

Analýza hlavních koordinát (*principal coordinate analysis* – *PCoA*) slouží k nalezení směrů největší variability v datech na základě koeficientů podobnosti mezi jednotlivými jedinci. Pro


analýzu pomocí programu SYNTAX je třeba si připravit dva datové soubory: (1) soubor s daty, (2) soubor s popisky (labels)

1. datový soubor (nejlépe s koncovkou .dat) by měl vypadat následovně:

```
[jméno]
[počet vzorků] [počet lokusů]
1 0 1 1 1 0 1 0 1 0 0 1 ... (na konci mezera)
1 1 1 0 1 0 1 1 1 0 1 1 ... (na konci mezera)
...
```

2. soubor s popisky dat (nejlépe s koncovkou .lab)

```
[počet vzorků] [jméno]
vzorek1
vzorek2
...
```

Spustíme ordinační modul programu SYNTAX (např. ikonou ). Pomocí *File* → *Open Raw Data* otevřeme náš datový soubor, v okně se objeví naše matice dat. Pomocí *File* → *Open Labelfile for rows* otevřeme soubor s popisky. Řádkům jsou přiřazeny příslušné popisky. Vybereme metodu (*Method*), tedy *PCoA* a koeficient podobnosti (*Coefficient* → *For binary data* → *Jaccard*). Dále je třeba vybrat ordinaci řádků (*Ordination of objects as Rows*) a zvolit příslušný počet požadovaných os (*No. of axes*), např. 3. Nyní je všechno připraveno pro stisk tlačítka *Analyze*. Pokud byl formát vstupních dat v pořádku, proběhne výpočet a výstup analýzy je zobrazen do nového okna. V hlavním okně SYNTAXu můžeme kliknout na *Scattergram for objects* pro grafický výstup z analýzy. Pokud program při pokusu o analýzu zahlásí chybu, je třeba zkontrolovat formát vstupních dat (pozor na mezery, přebytečné odklepy apod.). Problémem může být i příliš složitá adresářová cesta pro přístup k datům.

Více viz skripta „Analýza multivariačních dat v taxonomii (Marhold & Suda)“ a manuál k programu SYNTAX.

### 3. Výpočet populačně-genetických charakteristik pomocí programu POPGEN

Program POPGEN (Population Genetic Analysis) je freeware stažitelný např. z adresy <http://www.ualberta.ca/~fyeh/>. Program umožňuje analýzu dominantních i kodominantních markerů a výpočet nejrůznějších indexů a statistik na populační úrovni (více viz manuál k programu). Vstupní data je třeba zformátovat následujícím způsobem (vlastní čísla a data se doplní místo údajů označených kurzívou):

```
/* jméno */
Number of populations = číslo
Number of loci = číslo
Locus name :
```

```


lokus1 lokus2 ...

name = jméno populace1
1011101...
1011101...

name = jméno populace2
1001...
...

```



Otevřeme program  a datový soubor načteme pomocí *File* → *Load Data* → *Dominant Marker Data*. RAPD nebo AFLP data budeme analyzovat pomocí *Dominant* → *Diploid Data*. V novém okně vybereme *Data Format* jako *Variable as column*, zvolíme zadání *Hierarchical structure (Single Populations* – parametry pro každou populaci zvlášť, *Multiple Populations* – parametry přes všechny populace, *Group* – možnost zadání skupin populací) a vybereme parametry, které chceme odhadnout (*Estimation*). Po stisku *OK* máme možnost vyloučit některé lokusy nebo populace z analýzy. Pokud bylo zaškrtnuto *Group*, následuje zadání počtu skupin populací a možnost jejich definice. Výstup z analýzy je zobrazen v novém okně a obsahuje odhady zadaných parametrů pro každou populaci (*Single Populations Descriptive Statistics*) nebo přes všechny populace (*Multiple Populations Descriptive Statistics*). Více viz manuál k programu a v něm obsažené citace na originální práce k jednotlivým odhadovaným parametrům.

#### 4. AMOVA (zjištění rozdělení genetické variability pomocí programu Arlequin)

Analýza molekulární variance (AMOVA) je používána ke zjištění, jaká část z celkové variability je uvnitř populací a jaká mezi populacemi, případně mezi skupinami populací. To je možné vypočítat např. pomocí programu Arlequin (freeware), verze 3 je stažitelná z adresy <http://cmpg.unibe.ch/software/arlequin3>. Datový soubor může vypadat třeba takto (místo textu napsaného *kurzívou* je třeba doplnit vlastní údaje, „...“ znamenají pokračování datového souboru). V části *[Data]* jsou údaje o jedincích (*[[HaplotypeDefinition]]*), populacích (*[[Samples]]*) a skupinách populací (*[[Structure]]*). Soubor je třeba uložit s koncovkou \*.arp.

```


[Profile]
Title="jméno"
NbSamples="počet populací"
DataType=RFLP
GenotypicData=0
LocusSeparator=NONE
CompDistMatrix=1
[Data]
[[HaplotypeDefinition]]
HaplListName="jméno"
HaplList= {

```

```

sample1      101110101...
sample2      101110101...
sample3      100110011...
...
}
[[Samples]]
SampleName="jméno populace1"
SampleSize="počet vzorků v populaci1"
SampleData= {
sample1  "počet"
sample2  "počet"
...
}
SampleName="jméno populace2"
SampleSize="počet vzorků v populaci2"
SampleData= {
sample3  "počet"
sample4  "počet"
}
[[Structure]]
StructureName="jméno"
NbGroups="počet skupin"
#"jméno skupiny1"
Group={
"jméno populace1"
"jméno populace2"
...
}
#"jméno skupiny2"
Group={
"jméno populace3"
"jméno populace4"
}

```

Po spuštění programu  je třeba načíst data (*Open Project*). V levé části okna se zobrazí seznam vzorků (*Samples*) a struktura datového souboru (*Groups*). Na listě *Settings* klikneme na *Genetic structure* a vybereme *AMOVA*. Zaškrtneme *AMOVA computations* a můžeme kliknout na *Start*. Po provedení výpočtů se zobrazí výsledky v html formátu, AMOVA výsledky najdeme pod *Run* → *Genetic structure* → *AMOVA*. Tabulka ukazuje rozdělení celkové variability na jednotlivé úrovně (*Source of variation*, *Percentage of variation*). Více viz manuál k programu Arlequin.

## 5. Práce s programem FAMD

Program FAMD (*Fingerprinting Analysis with Missing Data*) umožňuje jednoduché analýzy binárních dat jako je tvorba matic podobností, stromů pomocí UPGMA a NJ, PCoA, AMOVA ad. Výhodou je, že data nemusí být nijak speciálně formátována, stačí 0-1 matici přepokopírovat z tabulkového procesoru (např. Excelu) do textového souboru, příp. pouze přenést přes schránku. Vstupní soubor může vypadat např. takto:



vz1	0	1	1	0	1	1	1	1	1
vz2	0	1	1	0	1	1	1	0	1
vz3	0	1	0	1	1	0	0	0	1
vz4	1	1	0	1	1	0	0	0	1
vz5	0	0	1	0	1	1	1	1	0
vz6	0	0	1	0	1	1	1	0	0
vz7	1	0	0	0	1	0	0	0	0
vz8	1	0	0	0	1	0	0	0	0

\*

[Groups]

AllData/skup1= vz1, vz2, vz3, vz4;

AllData/skup2= vz5, vz6, vz7, vz8;

AllData/skup1/pop1= vz1, vz2;


AllData/skup1/pop2= vz3, vz4;

AllData/skup2/pop3= vz5, vz6;

AllData/skup2/pop4= vz7, vz8;

\*

V sekci [Groups] jsou definovány čtyři populace (pop1-pop4) a dvě skupiny populací (skup1,

skup2). Po spuštění programu  otevřeme data pomocí *File* → *Load*, zvolíme *Individuals in rows* a *header presence for individuals*, zaškrtneme *Delimited data* a *Include Groups*. Před vytvořením stromu je třeba vytvořit matici podobností (*Analysis* → *Standard Similarity*). Program ji zapíše do souboru *analysis.txt* (do stejného adresáře odkud jsme načetli data). Ještě před tím je možno provést výběr koeficientu podobnosti pomocí *Options* → *Similarity Coefficient Selection*. Strom vytvoříme pomocí *Trees* → *Neighbour Joining*. Strom bude zapsán do souboru *outtree.ph*. Strom zobrazíme pomocí *View* → *Tree File* (je potřeba mít nainstalovaný program *TreeView*). PCoA provedeme jako *Trees* → *Principal Coordinate Analysis*. Zobrazí se další okno s grafickým výstupem, který je možno modifikovat (výsledky analýzy jsou opět zapsány do souboru *analysis.txt*). Pokud jsme ve vstupním souboru definovali skupiny, je možné si různé skupiny v koordinačním diagramu zobrazit různou barvou nebo symbolem (pod *Group appearance* zvolíme skupinu a přiřadíme jí barvu a symbol).

Program dále umožňuje bootstrap stromů (*Replicate Analyses* → *Bootstrap Std. Tree*), konsenzuální strom lze spočítat pomocí *Tree* → *Strict Consensus* a zobrazit pomocí *View* → *Consensus Tree File*.

Pokud v *Options* → *Similarity Coefficient Selection* zvolíme Euklidovskou vzdálenost, je možno počítat i AMOVA (*Analysis* → *AMOVA*). V tom případě je nutné mít ve vstupním souboru definovanou příslušnost každého jedince do právě jedné populace, případně skupiny.

## E. Práce se sekvencemi

### 1. Vytvoření alignmentu pomocí programu ClustalX

Alignment znamená seřazení homologních bází pod sebe (positional homology). Automaticky to lze udělat např. v programu ClustalX (<http://bips.ustrasbg.fr/fr/Documentation/ClustalX>).

- soubor se sekvencemi ve FASTA formátu otevřeme pomocí *File* → *Load Sequences*
- zvolíme *Alignment* → *Do Complete Alignment*
- vytvořený alignment je automaticky uložen do souboru s koncovkou \*.aln

### 2. Automatické kodování indelů v programu SeqState

Pokud alignment obsahuje mezery (*gaps*), tedy inserce/delece (souhrnně indely), které mohou být fylogeneticky informativní, je potřeba je okódotovat jako zvláštní znaky (nejčastěji 0/1), které jsou přidány na konec alignmentu. To lze automaticky provést např. v programu SeqState (<http://www.nees.unibonn.de/downloads/SeqState>). Podrobnosti ohledně způsobu kódování indelů lze nalézt např. v článku Müller (2006): Incorporating information from length-mutational events into phylogenetic analysis. *Molecular Phylogenetics and Evolution*, 38, 667-676.

- *File* → *Load NEXUS file* (nebo *Load FASTA file*)
- vybrat soubor s koncovkou \*.nex (nesmí být v interleaved, ale v sequential formátu, tj. vytvořený např. přeformátovaný pomocí FORCONu), resp. \*.fas
- *IndelCoder* → *modified complex coding* (bližší informace o způsobu kódování viz zmiňovaný článek)
- vygeneruje NEXUS soubor (k názvu původního souboru přidá \_mcic.nex) s původní maticí a přidaným kódováním indelů, který lze přímo použít pro analýzu např. v programu PAUP

### 3. FaBox – manipulace se sekvencemi ve formátu FASTA

(<http://www.birc.au.dk/fabox/>)

- pomocí tohoto webu lze manipulovat se sekvencemi ve FASTA formátu (pracovat s hlavičkami, ořezávat a spojovat alignmenty, formátovat pro další programy)

- volba „*Create TCS input file from fasta (fasta2tcs)*“ vytvoří správně formátovaný soubor pro analýzu v programu TCS

#### 4. Analýza sekvencí v programu TCS

- program TCS 1.21 (<http://darwin.uvigo.es/software/tcs.html>) zkolabuje sekvence na jednotlivé haplotypy a pomocí tzv. statistické parsimonie vytvoří síť mezi haplotypy
- spustíme TCS1.21.jar (pro běh programu je třeba mít nainstalované JRE (Java Runtime Environment, <http://java.sun.com/javase/downloads>); stahujte JRE, nikoliv SDK!), zvolíme „*Start New TCS Analysis*“.
- *File* → *Select NEXUS/PHYLIP Sequence file*
- vše necháme jako default, pouze změníme „*Gaps = missing*“, *RUN*
- program nakreslí haplotypovou síť, kterou je možno editovat (uložena v souboru \*.graph), případně exportovat (jako PostScript nebo PICT file)
- informace o analýze program uloží do souboru \*.graph.log

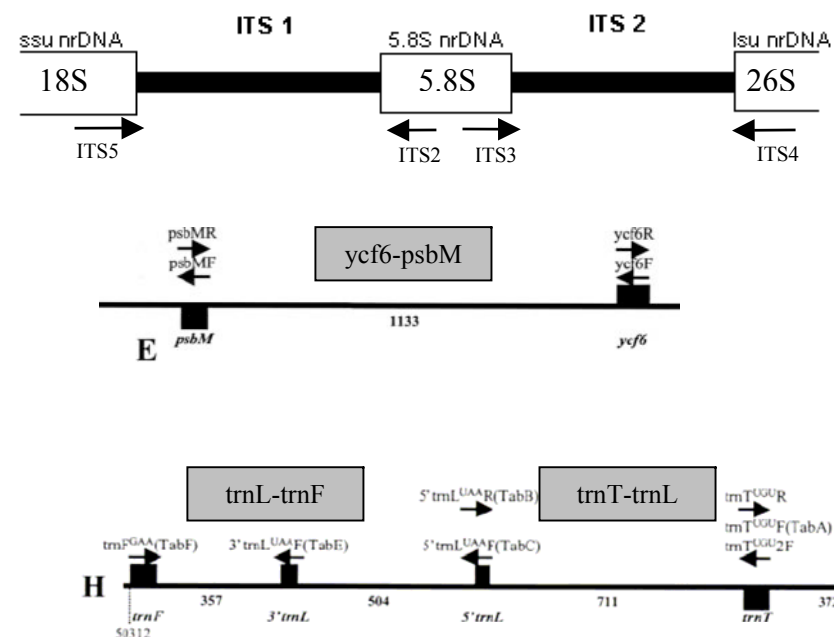
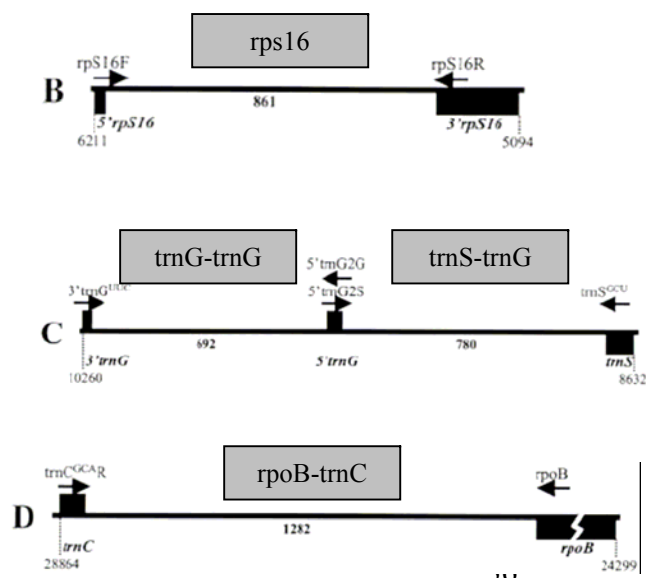
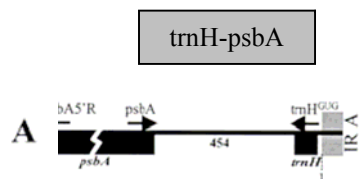
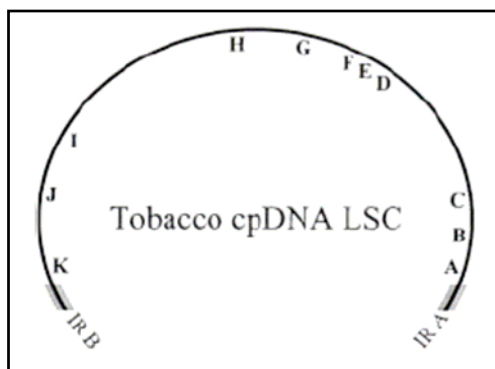
**Tabulka 1:** PCR a sekvenční primery pro nekodující úseky cpDNA a ITS region nrDNA (s vyznačením umístění primerů v genomu).

**Primery pro sekvenování variabilních nekodujících úseků cpDNA**

Jméno	Forward primer	Sekvence	Reverse primer	Sekvence	Velikost (bp)	T <sub>a</sub>	Reference
<i>trnT-trnL</i>	<i>trnT2F</i>	CAAAATGCGATGCTCTAACCT	<i>trnL-b</i>	TCTACCGATTTTCGCCATATC	527–1023	50-55	Taberlet et al. 1991, Cronn et al. 2002
<i>trnL-trnF</i>	<i>trnL-c</i>	CGAAATCGGTAGACGCTACG	<i>trnF-f</i>	ATTTGAACTGGTGACACGAG	602–1076	50	Taberlet et al. 1991
<i>trnH-psbA</i>	<i>trnH</i>	CGCGCATGGATTCCACAATCC	<i>psbA</i>	GTTATGCATGAACGTAATGCTC	198–1077	50-56	Tate & Simpson 2003, Sang et al. 1997
<i>trnS-trnG</i>	<i>trnS</i>	AGATAGGGATTTCGAACCCCTCGGT	<i>trnG2S</i>	TTTTACCACTAAACTATAACCCGC	619–1035	50	Shaw et al. 2005
<i>trnG-trnG</i>	<i>trnG2G</i>	GCGGGTATAGTTTAGTGGTAAAA	<i>trnG</i>	GTAGCGGAATCGGACCCGCATC	697–1008	50	Ohsako & Ohnishi 2000, Shaw et al. 2005
<i>rpoB-trnC</i>	<i>rpoB</i>	CKACAAAAYCCYTCRAATTG	<i>trnC</i>	CACCCRGATTYGAACCTGGGG	914–1309	50-57	Shaw et al. 2005
<i>rps16</i>	<i>rpS16F</i>	AAACGATGTGGTARAAAGCAAC	<i>rpS16R</i>	AACATCWATTGCAASGATTTCGATA	784–946	50-55	Shaw et al. 2005
<i>ycf6-psbM</i>	<i>ycf6F</i>	ATGGATATAGTAAGTCTYGTGGGC	<i>psbMR</i>	ATGGAAGTAAATATTCTYGCATTTATTGCT	406-1283	50-55	Shaw et al. 2005

**Primery pro sekvenování ITS úseku rDNA**

ITS1	ITS5	GGAAGTAAAAGTCGTAACAAGG	ITS2	GCTGCGTTCTTCATCGATGC	200-300	White et al. 1990
ITS2	ITS3	GCATCGATGAAGAACGCAGC	ITS4	TCCTCCGTTATTGATATGC	180-240	White et al. 1990



**Tabulka 2:** Kodování kombinací AFLP primerů (se třemi selektivními nukleotidy) a fluorescenční označení *EcoRI* primerů.

				<i>MseI</i> primer							
				1	2	3	4	5	6	7	8
				CAA	CAC	CAG	CAT	CTA	CTC	CTG	CTT
<i>EcoRI</i> primer	Y	NED	AAC	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8
	G	HEX	AAG	G1	G2	G3	G4	G5	G6	G7	G8
	B	6-FAM	ACA	B1	B2	B3	B4	B5	B6	B7	B8
	Y	NED	ACC	Y9	Y10	Y11	Y12	Y13	Y14	Y15	Y16
	G	HEX	ACG	G9	G10	G11	G12	G13	G14	G15	G16
	B	6-FAM	ACT	B9	B10	B11	B12	B13	B14	B15	B16
	Y	NED	AGC	Y17	Y18	Y19	Y20	Y21	Y22	Y23	Y24
	G	HEX	AGG	G17	G18	G19	G20	G21	G22	G23	G24
	B	6-FAM	ATC	B17	B18	B19	B20	B21	B22	B23	B24

***MseI* primer**

5' - GATGAGTCCTGAGTAAxxx -3'

***EcoRI* primer**

5' - GACTGCGTACCAATTCxxx -3'

**Tabulka 3:** Seznam primerů pro amplifikaci cpDNA mikrosatelitových lokusů včetně fluorescenčního značení.

fluoresc. barva		kód	pozice (bp)	lokalizace	repetice u tabáku	sekvence primerů	T <sub>a</sub>		Velikost alel u tabáku (bp)
							Wal	Cpr	
G	HEX	ccmp1	3801	<i>trnK</i> intron	(T) <sub>10</sub>	CAGGTAAACTTCTCAACGGA	58	53.3	139
						CCGAAGTCAAAAGAGCGATT	58	57.3	
Y	NED	ccmp2	8609	5' to <i>trnS</i>	(A) <sub>11</sub>	GATCCCGGACGTAATCCTG	60	58.0	189
						ATCGTACCGAGGGTTTCAAT	60	58.6	
B	6-FAM	ccmp3	10075	<i>trnG</i> intron	(T) <sub>11</sub>	CAGACCAAAAGCTGACATAG	58	51.3	112
						GTTTCATTTCGGCTCCTTTAT	56	53.9	
G	HEX	ccmp4	12872	<i>atpF</i> intron	(T) <sub>13</sub>	AATGCTGAATCGAYGACCTA	60	55.3	126
						CCAAAATATTBGGAGGACTCT	58	56.1	
Y	NED	ccmp5	16950	3' to <i>rps2</i>	(C) <sub>7</sub> (T) <sub>10</sub>	TGTTCCAATATCTTCTTGTCATTT	62	55.0	121
			16977			(T) <sub>5</sub> C(A) <sub>11</sub>	AGGTTCCATCGGAACAATTAT	58	
B	6-FAM	ccmp6	45119	ORF 77-ORF 82	(T) <sub>5</sub> C(T) <sub>17</sub>	CGATGCATATGTAGAAAGCC	58	52.7	103
						CATTACGTGCGACTATCTCC	60	52.5	
G	HEX	ccmp7	57339	<i>atpB-rbcL</i>	(A) <sub>13</sub>	CAACATATACCACTGTCAAG	56	45.1	133
						ACATCATTATTGTATACTCTTTC	58	44.5	
Y	NED	ccmp8	71563	<i>rpl20-rps12</i>	(T) <sub>6</sub> C(T) <sub>14</sub>	TTGGCTACTCTAACCTTCCC	60	52.9	77
						TTCTTTCTTATTTTCGCAGTGAA	58	53.9	
B	6-FAM	ccmp9	74060	ORF 74b- <i>psbB</i>	(T) <sub>11</sub>	GGATTTGTACATATAGGACA	54	45.1	98
						CTCAACTCTAAGAAATACTTG	56	44.2	
G	HEX	ccmp10	86694	<i>rpl2-rps19</i>	(T) <sub>14</sub>	TTTTTTTTTAGTGAACGTGTCA	56	53.3	103
						TTCGTGDCGTAGTAAATAG	58	53.7	

**Tabulka 4:** Příklad designu multiplex PCR pro 10 mikrosatelitových lokusů u *Nymphoides peltata*.

rozsah alel	bp	152-162	147-175	149-173	176-204	160-174	220-234	147-169	143-167	135-137	127-129
bp	lokus	Np152	Np207	Np274	Np306	Np382	Np641	Np694	Np730	Np274-2	Np382-2
152-162	Np152		y	y		y		y	y		
147-175	Np207			y	x,y	y		x,y	y		
149-173	Np274				y	y		y	x,y		
176-204	Np306					y	x	y	y		
160-174	Np382							y	y		x
220-234	Np641										
147-169	Np694								y	y	
143-167	Np730									x,y	
135-137	Np274-2										
127-129	Np382-2										

x = tvorba primerových dimerů

y = délkový překryv alel

**Multiplex PCR:**

multiplex I	rozsah alel
Np152	152-162
Np207	147-175
Np382	160-174
Np274-2	135-137
Np641	220-234

multiplex II	rozsah alel
Np274	149-173
Np382-2	127-129
Np694	147-169
Np306	176-204

multiplex III	rozsah alel
Np730	143-167