

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 D: 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 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. GenElute PCR Clean-Up Kit (Sigma) nebo PCR Extraction Kit (Geneaid)). 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. Nedochází 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 450-850 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 3130xl Avant (v sekvenační laboratoři PřF UK) jsou produkty cyklického sekvenování elektroforeticky rozděleny podle délky.

D1. 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 μ l.

- sterilní Milli-Q voda..... 15.60 μ l
- 10 \times PCR buffer (Sigma) 2.00 μ l
- dNTP (10 mM)..... 0.40 μ l
- *forward* primer (25 pmol/ μ l)..... 0.25 μ l
- *reverse* primer (25 pmol/ μ l) 0.25 μ l
- (JumpStart) RedTaq DNA Polymerase (Sigma) (1U/ μ l) 0.50 μ l
- + studovaná DNA (5 ng/ μ l)..... 1.00 μ 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 μ l) a přidáme 1 μ 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 μ l na agarosový gel (1%) v 1 \times TAE bufferu. Do první jamky na gelu nanese 1.5 μ l žebříčku (100 bp DNA Ladder Plus). 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 D2). V opačném případě bude třeba PCR reakci optimalizovat (zvýšit teplotu annealingu, zvýšit koncentraci $MgCl_2$, přidat tzv. aditiva... viz obecné poznámky k PCR).

D2. protokol na přečištění PCR produktu pomocí GenElute PCR Clean-Up Kit (Sigma)

PCR produkt přečistíme pomocí kitu GenElute PCR Clean-Up Kitu (Sigma), který využívá tzv. „*spin column technique*“, tj. kolonky s membránou na kterou se váže DNA. Alternativou je PCR Extraction Kit (Geneaid), případně můžeme přečistit fragment vyřízlý z gelu pomocí MinElute Gel Extraction Kit (Qiagen).

- ke zbylým 15 μ l PCR produktu přidáme 5 \times násobek Binding solution a důkladně promícháme na vortexu
- vložíme kolonku do označené (!) 2 ml eppendorfky, 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 μ l Wash Solution (s přidaným etanolem)
- 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 2 minuty při maximálních otáčkách (pro odstranění zbytkového etanolu)
- vložíme kolonku do nové (a označené!) 1.5 ml eppendorfky a přímo do středu membrány napipetujeme 20 μ l 1 \times TE bufferu předehřátého na 65°C (např. v eppendorfce na termobloku)
- necháme stát 15-30 minut
- centrifugujeme 2 minuty při maximálních otáčkách

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

Změříme koncentraci DNA pomocí Nanodropu.

D3. 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 bufferem, 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	

D4. 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 7, 2.suterén), produkt v této formě vydrží 1-2 týdny při -20 °C

protokol E: 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 nebo čtyřikrát, vždy s různou kombinací selektivních primerů a různým fluorescenčním značením. Automatický sekvenátor umí najednou rozeznat čtyři až pět fluorescenčních barev, tou čtvrtou, resp. pátou je standard (viz následující bod).
5. *smíchání selektivních PCR produktů a přesrážení produktu s NaOAc* – každý vzorek vznikne smícháním 1 μ l od každého ze 3-4 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 3130xl (v sekvenační laboratoři PřF UK) jsou AFLP produkty elektroforeticky rozděleny podle délky.

E1. restrikce DNA

Pro restrikci a ligaci používáme AFLP[®] Core Reagent Kit I od firmy Invitrogen (Life Technologies). 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 vracíme do -20 °C.

- 5× Reaction Buffer 1.0 µl
- *EcoRI/MseI*..... 0.4 µl
- sterilní H₂O..... 1.1 µl

Směs promícháme na vortexu, krátce zcentrifugujeme a 2.5 µl rozpipetujeme do 0.2 ml PCR zkumavek. Přidáme 2.5 µl DNA (20-50 ng/µl) a inkubujeme 2-6 hodin (podle optimalizace pro daný druh) při 37 °C v termocykleru.

E2. ligace adaptorů

Pro ligaci si v 1.5 ml eppendorfce připravíme ligační směs pro příslušný počet vzorků (také s využitím AFLP[®] 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 mikrozkušavce, při práci stále sledujeme správné napipetování a vypuštění oboje, PCR zkumavky při práci držíme v ruce!). DNA s ligační směsí inkubujeme při 37 °C 2-12 hodin 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 (po smíchání 1 µl 6×loading dye). Na gelu by měl být vidět rozmazaný pruh rozštěpené DNA, především mezi 50-1000 bp.

E3. preselektivní amplifikace (preamplifikace)

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

- PA mix..... 4.0 µl
- 10× Buffer for JumpStart RedTaq..... 0.5 µl
- JumpStart 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	0:30	
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 2 µl neředěného (!) produktu preamplifikační PCR reakce smíchaného s 1 µl 6× loading dye). Po elektroforéze bychom měli vidět opět rozmazaný pruh fragmentů, některé z nich by měly být výraznější.

Preamplifikační PCR reakci naředíme 10× sterilní H₂O (pokud jsou proužky na kontrolním gelu slabší, možno ředit méně), promícháme na vortexu a krátce zcentrifugujeme. Zamrazíme v -20 °C pokud hned nepokračujeme dalším krokem.

E4. selektivní amplifikace

Selektivní amplifikace se provádí třikrát nebo i čtyřikrát pro každý vzorek, vždy s jinou kombinací primerů a jiným fluorescenčně označeným primerem. Na sekvenátoru lze totiž využít dvou typů tzv. matic neboli toho, jak přístroj rozeznává fluorescenčně značené fragmenty (čtyřbarevný dye set D-30 s barvami 6-FAM, HEX, NED a standardem ROX a pětibarevný dye set D-33 s barvami 6-FAM, VIC, NED, PET a standardem LIZ). Je možné

kombinovat pouze barvy náležející do jednoho setu! Do tří 1.5 ml eppendorfek si tedy připravíme tři (nebo čtyři) PCR směsi, které se budou lišit pouze primery:

- sterilní H₂O..... 5.1 µl
- 10× Buffer for JumpStart 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 0.2 ml PCR zkumavku, případně můžeme pracovat se stripy nebo destičkami. Do každé zkumavky napipetujeme (nejlépe multipipetou) 7.5 µl směsi a přidáme 2.3 µl DNA po preselektivní PCR (pipetováním promícháme). Zkumavky, 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	0:30	} 8 ×	
64 °C	0:30 (v každém kroku -1 °C)		
72 °C	2:00		
94 °C	0:30	} 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.

Správně proběhlou selektivní reakci můžeme opět otestovat na 1.8% agarozovém gelu v 1× TBE bufferu při 60V. Na gel nanese 3 µl produktu selektivní PCR reakce (barvička je již v bufferu polymerázy).

E5. přesrážení produktu s acetátem sodným a přidání standardu (ROX, LIZ)

- do čistých PCR zkumavek (stripů) přepipetujeme po 3 μ l od každého selektivního PCR produktu (případně „vyladěný“ poměr jednotlivých produktů), uzavřeme, promícháme na vortexu, krátce zcentrifugujeme
- na stěnu 1.5 ml eppendorfky napipetujeme dávkovací pipetou 2 μ l NaOAc (acetátu sodného)
- přidáme 3 μ l směsi tří selektivních PCR produktů
- přidáme 25 μ l 96% ethanolu (z mrazáku) pomocí dispensoru
- promícháme na vortexu, krátce zcentrifugujeme
- necháme 20 minut při -20°C
- centrifugujeme v předchlazené centrifuze 30 minut při 4°C
- opatrně slijeme supernatant, otočíme a položíme na buničinu a stočíme v centrifuze při 40 rcf (vyteče ethanol, precipitát zůstane na stěně jamky – není vidět)
- přidáme 100 μ l 70% ethanolu (z mrazáku) pomocí dispensoru
- centrifugujeme v předchlazené centrifuze 5 minut při 4°C , supernatant slijeme a opět stočíme dnem vzhůru při 40 rcf
- otevřené zkumavky vysušíme 5-10 min při 65°C (na termobloku)
- vzorky můžeme skladovat 1-2 týdny při -20°C

Těsně před odnesením vzorků na sekvenátor pokračujeme následujícím způsobem:

- v 1.5 ml eppendorfce si připravíme směs pro příslušný počet vzorků
 - deionizovaný formamid..... 10.00 μ l
 - GeneScan-ROX nebo LIZ (podle dye setu!)..... 0.25 μ l
- směs promícháme na vortexu a krátce zcentrifugujeme
- přidáme 10 μ l směsi ke každému vzorku
- zdenaturujeme na termomixeru nebo v cycleru 3 minuty při 95°C
- vzorky v daném pořadí přepipetujeme na platíčko do sekvenátoru (optical plate) a přelepíme folií

protokol F: Analýza mikrosatelitů (SSRs)

Jako mikrosatelity („*simple sequence repeats*“) 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 podle zvoleného dye setu, viz pozn. u AFLP). 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> (Primery - Mikrosatelitové primery - Seznam publikovaných mikrosatelitových primerů).

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 párů 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 párů můžeme udělat např.

v programu FastPCR (<http://primerdigital.com/fastpcr.html>, freeware). Příklad kombinace primerů do multiplex PCR je v Tabulce 4. Při pipetování multiplexové reakce lze s úspěchem využít Multiplex PCR Kit (Qiagen).

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 3130xl (v sekvenační laboratoři PřF UK) jsou mikrosatelitové alely elektroforeticky rozděleny podle délky.

F1. PCR amplifikace

Do 1.5 ml eppendorfký 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 H₂O..... 13.60 µl
- 10× Buffer for (JumpStart) 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 _a °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).

F2. přesrážení a přidání standardu

Přesrážení produktu a přidání standardu (GeneScan-ROX nebo LIZ) je obdobné jako v případě AFLP (protokol E5).

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/Products/finchtv.shtml>). 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/t-seqmanpro.aspx>, 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*
- rozbalíme 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 např. v programu Geneious (<http://www.geneious.com>), ručně nebo s využitím webové verze programu CAP3 (<http://doua.prabi.fr/software/cap3>).


3. Vyhodnocování AFLP dat v programu GeneMarker (sepsal F. Kolář)

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




Import a kontrola vstupních dat

- data naimportujeme 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* (okno s touto volbou vyskočí také automaticky při importu vzorků), jako *Template* zvolíme ABI, jako *Size Standard* „GS500_1“ (akceptuje všechny peaky standardu GS-500 ROX kromě peaku o délce 250 bazí), jako *Analysis type* „AFLP“ → v dalším okně můžeme nastavit nejruznější parametry automatické analýzy, to bude důležité až při analýze používající panel (viz dále), nyní je vhodné pouze v *Allele call / Auto Range (bps)* nastavit rozsah fragmentů které budou skórovány, tj. většinou 100-500 bp → v dalším okně zatím nic neměníme → *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. V pravé části nahoře je virtuální gel zobrazující všechny vzorky, pod ním je otevřen jeden (několik) vzorků zobrazených v podobě peaků.
- virtuální gel si můžeme přiblížit klepnutím levým tlačítkem myši a tažením doprava a dolů až vznikne obdélník požadovaného rozsahu. Zpět celý gel si zobrazíme pokud




klepneme levým tlačítkem myši a táhneme nahoru a doleva. V přiblíženém gelu se můžeme pohybovat zleva doprava tažením při zmáčknutí pravého tlačítka myši. Zcela analogicky můžeme pracovat i s vzorky zobrazenými jako histogramy (s peaky).

- nyní si zkontrolujeme zda jsou správně srovnány standardy: opakovaným klikáním na tlačítko  se dostaneme na barvu standardu (zpravidla červená) a na virtuálním gelu (popř. i v histogramech) zkontrolujeme, zda jsou peaky příslušné délky seřazeny u všech vzorků pod sebou.
- dále je vhodné zkontrolovat, jak vypadají jednotlivé vzorky – v případě, že je chceme vyřadit (např. příliš slabý signál) stačí kliknout Ctrl+Del a vzorek je *Disabled* (opětovným kliknutím Ctrl+Del vzorek opět aktivujeme)

Vytvoření panelu

- nyní si vytvoříme tzv. panel, tedy jakousi kostru, v níž proběhne (polo)automatické skórování našich vzorků. V praxi je vhodné přípravu finální verze panelu nechat až na okamžik, kdy budeme mít k dispozici většinu vzorků. Veškeré manuální úpravy, které bychom provedli v datech po prvním zanalyzování panelem, by nám totiž po opakované analýze pomocí panelu zmizely.
- Voláme *Tools* → *Panel Editor*, pomocí tlačítka  si zvolíme barvu, pro kterou chceme panel vytvořit, kliknutím na tlačítko  vyvoláme okno pro tvorbu nového panelu; zde necháme přednastavené podmínky (automatické vytvoření panelu s užitím všech vzorků) → *OK*
- takovouto hrubou verzi panelu bude potřeba ručně dopracovat. Klepnutím na  můžeme přepínat mezi zobrazením histogramu jednoho vzorku, histogramů všech vzorků přeložených přes sebe a virtuálního gelu, v nichž vždy vidíme automaticky vyhledané tzv. *bins*, tedy „příhrádky“ pro jednotlivé AFLP lokusy, v nichž bude následně automaticky skórována přítomnost nebo absence fragmentů (peaků). Histogramy, resp. gel můžeme přibližovat a oddalovat analogicky jako v případě vzorků výše. Počty, polohu a rozsah *bins* je možné manuálně měnit (nejlépe ve virtuálním gelu) – patřičný *bin* klepnutím myši vybereme, můžeme ho smazat (klik pravým tlačítkem myši – *Delete Allele*), posouvat (Shift+levé tlačítko myši a táhnout) resp. rozšiřovat (najat na okraj binu, pak Shift+levé tlačítko myši a táhnout). Nový *bin* vytvoříme tak, že zmáčkneme Shift a na patřičné místo klepneme. Po skončení úprav uložíme změny a zavřeme okno Panel Editoru.
- vytvořený panel je možné exportovat jako soubor s příponou *.xml a zálohovat pro případné analyzování dalších vzorků

Analýza pomocí panelu a následné úpravy


- vytvořený panel dále využijeme k automatickému oskórování našich dat (provedeme tzv. *allele call*). Spustíme jakoby opakovanou analýzu pomocí  → v prvním okně v poli Panel vybereme náš uložený panel (ostatní nastavení by mělo být shodné s nastavením první analýzy) → v dalším okně nastavíme parametry automatické analýzy: např. zkontrolujeme rozsah skórovaných fragmentů (*Auto Range (bps)*) na zpravidla 100-500 bp), můžeme snížit *Intensity* u *Peak Detection Threshold* na 50 (bude skórovat i nižší peaky) a odkliknutím vypnout *Stutter peak filter* a *Plus-A Filter* (jejich podmínky bývají při vyhodnocování AFLP dat příliš tvrdé) → ve třetím okně můžeme nastavit při jakém skóre „spolehlivosti“ (lehce mystická, automaticky spočtená hodnota) bude dotyčný peak přijat / označen jako nejistý / odmítnut: vhodné může být snížení hodnoty mezi *Check* a *Pass* na cca 3 → OK → políčko u *Call size again* necháme prázdné a zmáčkeme rovnou *Apply to All* (nebudeme opět provádět kalibraci velikosti standardů). Pozn: Při finální analýze je vhodné zkusit data zanalyzovat s různým nastavení těchto parametrů a pak vybrat optimální řešení.
- po doběhnutí analýzy a odklepnutí okna máme automaticky oskórovány fragmenty (= peaky) v rámci jednotlivých lokusů (*bins*): vidíme peaky přijaté (šedé), nejisté (žluté) a odmítnuté (červené).
- i když si s nastavením parametrů důkladně pohrajeme, bude potřeba data následně pečlivě projít a provést ruční úpravy (časově nejnáročnější část celé práce). Žlutě, resp. červeně označené peaky lze potvrdit (kliknout a Ctrl+M), šedé lze smazat (kliknout a Del) a můžeme i přidat zcela neoznačené píky (kliknout pravým tlačítkem myši a *Insert Allele*). Pro snazší prohlížení je vhodné vypnout zobrazení gelu tlačítkem  a otevřít si víc vzorků naráz poklepáním na příslušné vzorky v seznamu vlevo (max. počet vzorků zobrazených naráz a další vlastnosti jde změnit ve *View – Preferences – Display Settings*). Mezi následujícími vzorky jde jednoduše vertikálně přepínat pomocí PgUp a PgDn. Pozor! Jakékoliv úpravy, které nyní provedeme, budou opět zrušeny, pokud bychom data opět zanalyzovali pomocí panelu.
- export dat (možný jen v plné verzi) provedeme jednoduše uložením tabulky vpravo do *.txt. nebo *.xls. Pokud chceme získat jen jednoduchou matici „0-1-?“ , zmáčkeme  a odškrtnutím vypneme *Show Peak Area*.

Přidávání dalších vzorků

- k již rozpracovanému projektu můžeme v průběhu analyzování libovolně přidávat další vzorky přes *Project – Add Samples to Project*. Vzorky se automaticky zanalyzují použitým panelem a můžeme je dále upravovat.

4. 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).

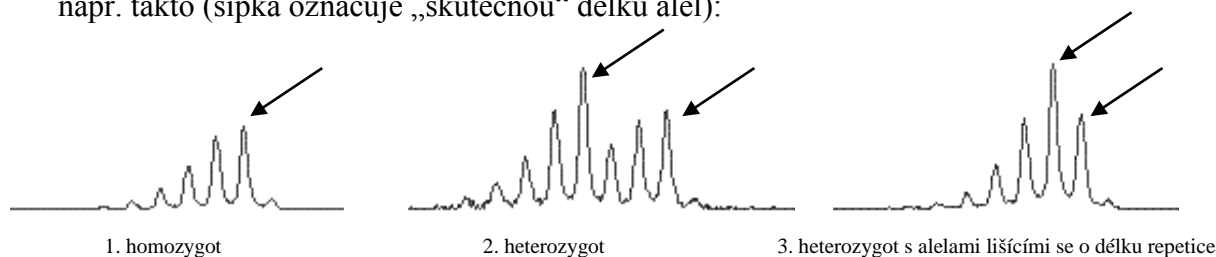
- data naimportujeme pomocí *File → Open Data*, v dalším okně potom *Add...* (více souborů najednou můžeme vybrat při stisknutí Ctrl nebo Shift, potom *Open* a *OK*)
- analýzu dat (kalibraci podle vnitřního standardu) spustíme pomocí *Project → Run*, jako *Template* zvolíme *ABI*, jako *Size Standard* „GS500_1“ (akceptuje všechny peaky standardu GS-500 ROX kromě peaku o délce 250 bází), v dalším okně můžeme nastavit nejružnější 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. Několikrát 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 ploidí je problém analogicky ještě komplikovanější.

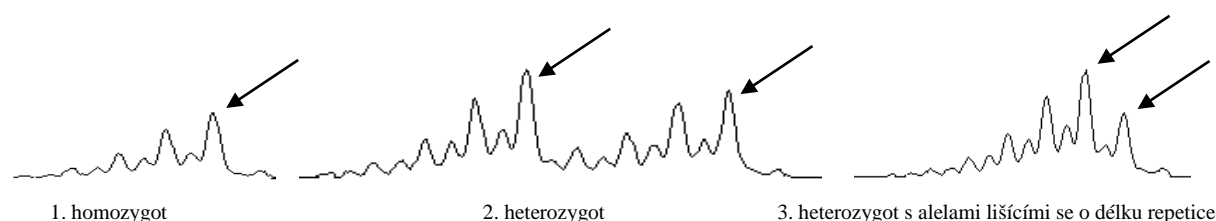
- Pro některé lokusy je příznačný výskyt tzv. „stutter bands“, tj. peaků, které jsou o dvou-, tří-, čtyř- atd. násobek repetice kratší než je skutečná délka alely. Vznikají díky sklouznutí DNA polymerázy při replikaci (tzv. *in vitro* DNA slippage). Typická pattern vypadají např. takto (šipka označuje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

- 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 4.05 (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 flanking region			
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	

jedno- nebo dvousloupcový formát

outbrední (d) nebo inbrední (h) individuum

číslo skupiny populací

chybějící data

název populace

Analyzovaný soubor uložíme do stejného adresáře jako je MSAnalyserMr.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) nebo STRUCTURE (vytvoří vždy). 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).

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

V programu Arlequin v 3.5 (<http://cmpg.unibe.ch/software/arlequin35>, 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.

3. Analýza hlavních koordinát (PCoA) a vytvoření stromů v programu PAST

V program PAST (<http://folk.uio.no/ohammer/past>; freeware, spustitelný *.exe soubor) lze jednoduše a rychle udělat (mimo jiné) analýzu hlavních koordinát (PCoA) a vytvořit stromy (neighbour-joining, UPGMA). Přes schránku vložíme upravenou distanční matici z MSA (je třeba ji upravit např. v Excelu pomocí *Data → Text do sloupců* a jako oddělovač zadat „Mezera“). Matice bude vypadat např. takto:

	vz1	vz2	vz3
vz1	0	0.925	0.85
vz2	0.925	0	0.93
vz3	0.85	0.93	0

Předtím, než ji pomocí schránky (Ctrl+C, Ctrl+V) vložíme, je nutné nahoře zaškrtnout *Edit labels*. Teď je možné definovat skupiny (budou tak obarveny ve všech následujících výstupech): Shift + myši označíme příslušné jedince (řádky) z dané skupiny, klikneme na *Edit → Row color/symbol* a zvolíme barvu a symbol. Nyní je možné soubor uložit (*File → Save*) a kódování barev se uloží i s daty.

- PCoA – označíme všechny jedince (klik do levého horního rohu), zvolíme *Multivar → Principal coordinates* a vybereme distanci jako *User distance*. Zobrazí se tabulka % variability vysvětlených jednotlivými osami (tlačítkem  zkopírujeme do schránky). Kliknutím na *View scatter* se zobrazí PCoA diagram. Kliknutím do grafu ho lze upravovat. Pomocí *View numbers* se zobrazí ordinační skóre jedinců na osách.
- NJ strom – *Multivar → Neighbour joining*, vybereme *User distance* (samozřejmě nelze bootstrapovat). Strom uložíme pomocí *Save as Nexus* a otevřeme např. v programu FigTree (<http://tree.bio.ed.ac.uk/software/figtree>).

D. 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://www.clustal.org/clustal2>).

- 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

Automaticky vytvořený alignment můžeme upravovat buď přímo v textovém editoru např. v poznámkovém bloku nebo využít výhod programu BioEdit (<http://www.mbio.ncsu.edu/BioEdit/bioedit.html>) nebo PhyDE (<http://www.phyde.de>). Pro potřeby druhého editoru je třeba mít alignment ve formátu FASTA nebo NEXUS sequential. Alternativně (a je to vhodnější obzvláště v případě více diverzifikovaných sekvencí) je vhodnější k alignmentu použít např. iterativní algoritmus MAFFT (<http://mafft.cbrc.jp/alignment/server>).

2. Konverze mezi různými formáty zápisu sekvencí

Různé formáty sekvencí můžeme konvertovat např. pomocí programu FORCON (<http://bioinformatics.psb.ugent.be/webtools/ForCon>), který je součástí programu TREECON (<http://bioinformatics.psb.ugent.be/software/details/3>).

- ForCon → Enter
- *Input format*: CLUSTAL, *Output format*: FASTA, OK
- změnit typ souboru na „všechny“, otevřít soubor alignmentu (s koncovkou *.aln)
- *Cut-off after how many characters ?* – napsat více než je pozic alignmentu
- uložit s koncovkou *.fas
- Select All, OK

Další možností je např. program PGDSpider, který umí formátovat pro řadu formátů, např. FASTA, NEXUS, MSA, Arlequin atd. (<http://www.cmpg.unibe.ch/software/PGDSpider>).

Lze také použít on-line službu na adrese

<http://sequenceconversion.bugaco.com/converter/biology/sequences/index.html>

3. Automatické kodování indelů v programu SeqState

Pokud alignment obsahuje mezery (*gaps*), tedy inserce/delece (souhrnně indely), které mohou být fylogeneticky informativní, je potřeba je okódovat jako zvláštní znaky (nejčastěji 0/1), které jsou přidány na konec alignmentu. To lze automaticky provést např. v programu SeqState (<http://bioinfweb.info/Software/SeqState>). Podrobnosti ohledně způsobu kódování indelů lze nalézt např. v článku Müller (2006): Incorporating information from length-mutational events into phylogenetic analysis. *Molec. Phyl. Evol.*, 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

4. FaBox – manipulace se sekvencemi ve formátu FASTA

(<http://users-birc.au.dk/biopv/php/fabox>)

- pomocí tohoto webu lze manipulovat se sekvencemi ve FASTA formátu (pracovat s hlavičkami, ořezávat a spojovat alignmenty, formátovat pro další programy atd.)
- volba „*Create TCS input file from fasta (fasta2tcs)*“ vytvoří správně formátovaný soubor pro analýzu v programu TCS

5. 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
- pro běh programu je třeba Javu (Java Runtime Environment, <http://www.oracle.com/technetwork/java/javase/downloads/index.html>; stahujte JRE!)
- spustíme TCS1.21.jar, 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 sekvenační primery pro nekodující úseky cpDNA a ITS region nrDNA (s vyznačením umístění primerů v genomu).

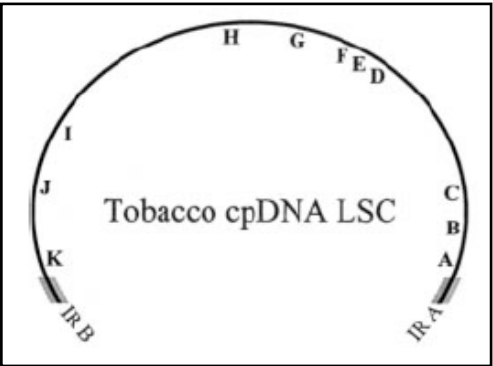
Shaw J. et al. (2005): The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. Am. J. Bot. 92, 142-166.

Primery pro sekvenování variabilních nekodujících úseků cpDNA

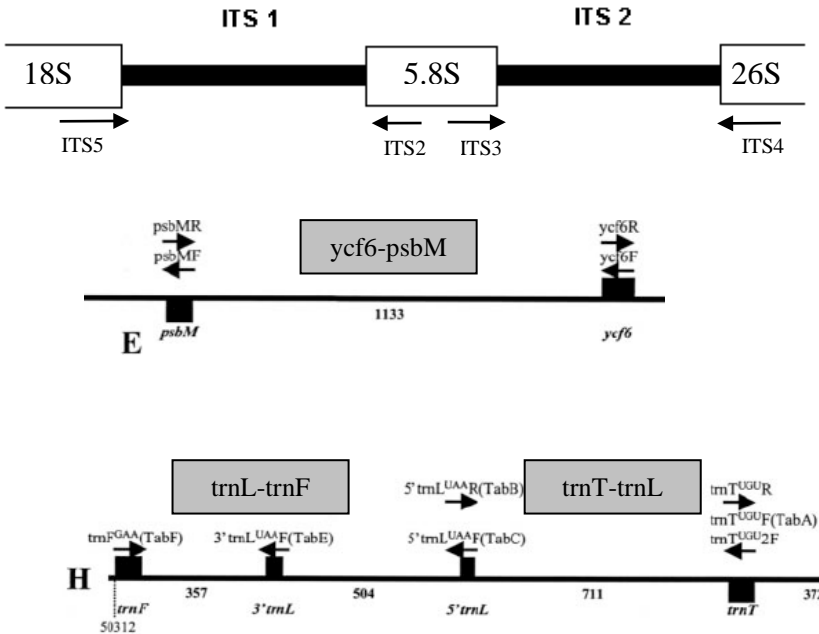
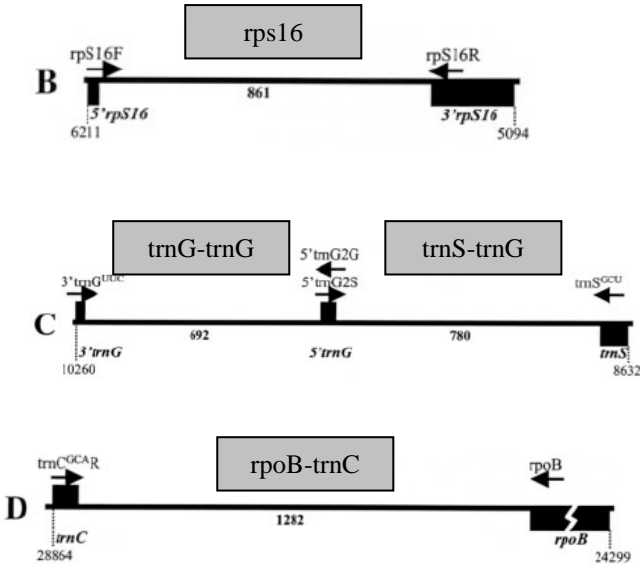
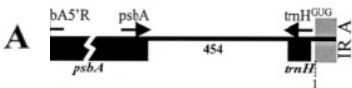
Jméno	Forward primer	Sekvence	Reverse primer	Sekvence	Velikost (bp)	T _a	Reference
<i>trnT-trnL</i>	<i>trnT2F</i>	CAAAATGCGATGCTCTAACCT	<i>trnL-b</i>	TCTACCGATTTCGCCATATC	527–1023	50-55	Taberlet et al. 1991, Cronn et al. 2002
<i>trnL-trnF</i>	<i>trnL-c</i>	CGAAATCGGTAGACGCTACG	<i>trnF-f</i>	ATTTGAACTGGTGACACGAG	602–1076	50	Taberlet et al. 1991
<i>trnH-psbA</i>	<i>trnH</i>	CGCGCATGGATTCACAATCC	<i>psbA</i>	GTTATGCATGAACGTAATGCTC	198–1077	50-56	Tate & Simpson 2003, Sang et al. 1997
<i>trnS-trnG</i>	<i>trnS</i>	AGATAGGGATTCTGAACCCTCGGT	<i>trnG2S</i>	TTTTACCACTAAACTATACCCGC	619–1035	50	Shaw et al. 2005
<i>trnG-trnG</i>	<i>trnG2G</i>	GCGGGTATAGTTTGTGGTAAAA	<i>trnG</i>	GTAGCGGGAATCGGACCCGCATC	697–1008	50	Ohsako & Ohnishi 2000, Shaw et al. 2005
<i>rpoB-trnC</i>	<i>rpoB</i>	CKACAAAAYCCYTCRAATTG	<i>trnC</i>	CACCCRGATTYGAAC TG GGG	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>	ATGGATATAGTAAGTCTYGCTTG GGC	<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	TCCTCCGCTTATTGATATGC	180-240		White et al. 1990



trnH-psbA



Tabulka 2: Kodování kombinací AFLP primerů (se třemi selektivními nukleotidy) a fluorescenční označení *EcoRI* primerů (dye set DS-30 s markerem ROX).

				<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 (dye set DS-30 s markerem ROX).

fluoresc. barva		kód	pozice (bp)	lokalizace	repetice u tabáku	sekvence primerů	T _a		Velikost alel u tabáku (bp)
							Wal	Cpr	
G	HEX	ccmp1	3801	<i>trnK</i> intron	(T) ₁₀	CAGGTAAACTTCTCAACGGA	58	53.3	139
						CCGAAGTCAAAAGAGCGATT	58	57.3	
Y	NED	ccmp2	8609	5' to <i>trnS</i>	(A) ₁₁	GATCCCGGACGTAATCCTG	60	58.0	189
						ATCGTACCGAGGGTTCGAAT	60	58.6	
B	6-FAM	ccmp3	10075	<i>trnG</i> intron	(T) ₁₁	CAGACCAAAGCTGACATAG	58	51.3	112
						GTTTCATTTCGGCTCCTTTAT	56	53.9	
G	HEX	ccmp4	12872	<i>atpF</i> intron	(T) ₁₃	AATGCTGAATCGAYGACCTA	60	55.3	126
						CCAAAATATTBGGAGGACTCT	58	56.1	
Y	NED	ccmp5	16950	3' to <i>rps2</i>	(C) ₇ (T) ₁₀	TGTTCCAATATCTTCTTGTCATTT	62	55.0	121
			16977		(T) ₅ C(A) ₁₁	AGGTTCCATCGGAACAATTAT	58	55.4	
B	6-FAM	ccmp6	45119	ORF 77-ORF 82	(T) ₅ C(T) ₁₇	CGATGCATATGTAGAAAGCC	58	52.7	103
						CATTACGTGCGACTATCTCC	60	52.5	
G	HEX	ccmp7	57339	<i>atpB-rbcL</i>	(A) ₁₃	CAACATATAACCACTGTCAAG	56	45.1	133
						ACATCATTATTGTATACTCTTTC	58	44.5	
Y	NED	ccmp8	71563	<i>rpl20-rps12</i>	(T) ₆ C(T) ₁₄	TTGGCTACTCTAACCTTCCC	60	52.9	77
						TTCTTTCTTATTTCGCAGTGAA	58	53.9	
B	6-FAM	ccmp9	74060	ORF 74b- <i>psbB</i>	(T) ₁₁	GGATTTGTACATATAGGACA	54	45.1	98
						CTCAACTCTAAGAAATACTTG	56	44.2	
G	HEX	ccmp10	86694	<i>rpl2-rps19</i>	(T) ₁₄	TTTTTTTTTAGTGAACGTGTCA	56	53.3	103
						TTCGTCGDCGTAGTAAATAG	58	53.7	

Weising K & Gardner RC (1999) A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. *Genome* 42, 9-19.

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