

Univerzita Palackého v Olomouci

Přírodovědecká fakulta

Katedra botaniky

a

Centrum regionu Haná pro biotechnologický a zemědělský výzkum

Ústav experimentální botaniky AV ČR, v.v.i.

pracoviště Olomouc

Pavla Christelová

**Analýza genetické diverzity a evoluce genomu
banánovníku**

Doktorská disertační práce

Studijní obor botanika

Školitel: Mgr. Miroslav Valárik, Ph.D.

Olomouc

2011

Poděkování

Ráda bych poděkovala svému školiteli Mgr. Miroslavu Valárikovi, Ph.D., vedoucímu Laboratoře molekulární cytogenetiky a cytometrie doc. Ing. Jaroslavu Doleželovi, DrSc. a Mgr. Evě Hřibové, Ph.D. za odborné vedení a pomoc v průběhu celého studia, stejně jako všem ostatním členům laboratoře za jejich ochotu a přátelský přístup.

Prohlášení

Prohlašuji, že jsem předkládanou práci vypracovala samostatně, pouze s použitím uvedené literatury.

V Olomouci,

.....

Tato práce byla vytvořena za podpory Ministerstva školství, mládeže a tělovýchovy České republiky a Evropského fondu pro regionální rozvoj – Operační program Výzkum a vývoj pro inovace č. CZ.1.05/2.1.00/01.0007.

BIBLIOGRAPHIC IDENTIFICATION

Author's name: Pavla Christelová (née Němcová)

Title: Analysis of the genetic diversity and evolution of the banana genome

Type of Thesis: Ph.D. thesis

Department: Botany

Supervisor: Mgr. Miroslav Valárik, Ph.D.

The Year of Presentation: 2011

Abstract:

Bananas and plantains (*Musa* spp.) are one of the major world trade commodities with irreplaceable role as a staple food for millions of people, especially in the tropical developing countries. Despite its enormous socio-economic importance, this crop is instantly endangered by various pests and diseases. Moreover, changes in the environmental conditions and farming practices of smallholders cause gradual loss of distinct banana genotypes. Thus, the rich genetic diversity as a potential source of valuable genes for breeding new improved varieties gets narrowed and the need for studying and effectively conserving the genetic diversity becomes ever more urgent. The present study employed several molecular tools with the aim to contribute to the characterization of the genetic diversity and evolution of the *Musaceae* species. Its first goal was to optimize a standardized methodology for efficient genotyping of individual banana species and subspecies using microsatellite markers. The genotyping platform developed in this work proved to be successfully applicable for molecular characterization of individual accessions also on the global level, as a basic methodology used in the *Musa* Genotyping Centre. The second part of the work aimed at reconstructing the phylogenetic relationships within the family *Musaceae* using the to date largest set of nuclear genes. This resulted in a plausible picture of the speciation events leading to the evolution of individual *Musaceae* evolutionary lineages including the first estimates of the divergence dates for clades within the *Musa* genus. In summary, the present study provides a complex picture of the evolution of *Musa* species which can be further utilized in characterizing banana genetic diversity and improving the classification within *Musaceae*.

Keywords: *Musa*, banana, diversity, molecular phylogenetics, evolution

Number of Pages/Apendices: 190 / 11

Language: Czech

BIBLIOGRAFICKÁ IDENTIFIKACE

Jméno: Pavla Christelová (roz. Němcová)

Téma: Analýza genetické diverzity a evoluce genomu banánovníku

Typ práce: Disertační práce

Obor: Botanika

Školitel: Mgr. Miroslav Valárik, Ph.D.

Rok obhajoby: 2011

Abstrakt:

Banánovník (*Musa* spp.) je jednou z nejdůležitějších plodin na světě s nezastupitelnou rolí jako zdroj obživy milionů obyvatel zejména v tropických rozvojových zemích. Navzdory svému velkému významu je však tato plodina neustále ohrožována chorobami a škůdci, a zároveň se díky vlivu prostředí a změnám pěstitelských zvyků některé druhy banánovníku postupně vytrácejí. Tím dochází ke ztrátám bohatých zdrojů genetické diverzity využitelné například pro šlechtění nových odolnějších odrůd. Je tedy žádoucí přistoupit k důkladnému prozkoumání dostupných genových zdrojů a evoluce banánovníku, aby bylo možné efektivně uchovávat a chránit genetickou diverzitu banánovníku. Předkládaná práce si kládla za cíl přispět k charakterizaci genetické diverzity a evoluce banánovníku pomocí molekulárních nástrojů. V první části práce byla optimalizována a standardizována metodika pro genotypování jednotlivých druhů a poddruhů banánovníku pomocí mikrosatelitových markerů. Podařilo se vytvořit fungující genotypovací systém, který je účinný pro charakterizaci jednotlivých položek i na globální úrovni v rámci tzv. Musa Genotyping Centre. V druhé části práce byly pomocí dosud největšího souboru jaderných genových oblastí studovány evoluční vztahy v rámci čeledi Musaceae a vůbec poprvé byly, na základě tohoto širokého souboru dat, poskytnuty první odhady stáří jednotlivých evolučních větví uvnitř rodu *Musa*. Ve svém souhrnu tak předkládaná práce poskytuje komplexní pohled na evoluci banánovníku využitelný pro další studium diverzity genových zdrojů banánovníku a zpřesnění klasifikace jednotlivých druhů.

Klíčová slova: *Musa*, banánovník, diverzita, molekulární fylogenetika, evoluce

Počet stran/příloh: 190 / 11

Jazyk: český

OBSAH

ÚVOD	8
CÍLE DISERTAČNÍ PRÁCE	9
1. CHARAKTERISTIKA BANÁNOVNÍKU	10
1.1 Morfologie banánovníku.....	10
1.2 Taxonomie banánovníku.....	11
1.3 Hospodářský a společenský význam banánovníku.....	13
1.4 Choroby a šlechtění banánovníku.....	14
2. MOLEKULÁRNÍ CHARAKTERIZACE GENETICKÉ DIVERZITY ROSTLIN	15
2.1 Biodiverzita a genetická diverzita.....	15
2.2 Nástroje studia genetické diverzity rostlin.....	15
2.2.1 <i>Biochemické markery</i>	16
2.2.2 <i>Molekulární DNA markery</i>	16
2.3 Genetická diverzita a její charakterizace u banánovníku.....	25
3. MOLEKULÁRNÍ FYLOGENETIKA ROSTLIN	27
3.1 Molekulární data a molekulární fylogenetika.....	27
3.2 Sekvence DNA pro molekulární fylogenetiku.....	29
3.2.1 <i>Chloroplastová DNA ve fylogenetické analýze</i>	30
3.2.2 <i>Jaderná ribozomální DNA ve fylogenetické analýze</i>	32
3.2.3 <i>Geny s nízkým počtem kopií ve fylogenetické analýze</i>	34
3.3 Analýza molekulárně-fylogenetických dat.....	36
3.3.1 <i>Distanční metody</i>	36
3.3.2 <i>Metody založené na kritériu optimality</i>	37
3.3.3 <i>Bayesovské metody</i>	38
3.4 Využití molekulárních dat k datování evolučních událostí.....	38
3.5 Fylogeneze čeledi Musaceae.....	39
PŘEHLED ZÍSKANÝCH VÝSLEDKŮ	55
PUBLIKACE V ODBORNÝCH ČASOPISECH (Přílohy I-VI).....	59
ABSTRAKTA PŘÍSPĚVKŮ NA KONFERENCÍCH (Přílohy VII-XI).....	174
ZÁVĚR	188
SEZNAM POUŽITÝCH ZKRATEK	190

ÚVOD

Banánovník (*Musa* spp., čeleď *Musaceae*) je velká jednoděložná bylina rostoucí v tropických a subtropických oblastech Afriky, jižní Ameriky a jihovýchodní Asie. Plody banánovníku jsou hlavním zdrojem obživy více než 800 milionů obyvatel zejména v rozvojových zemích tropů a jsou také významnou exportní komoditou, čímž se řadí mezi nejvýznamnější plodiny světa. Navzdory velkému společenskému a ekonomickému významu banánovníku však doposud nebyly uspokojivě vysvětleny a popsány fylogenetické vztahy v rámci čeledi *Musaceae* a rovněž zůstává nedořešena otázka klasifikace a charakterizace jednotlivých druhů a poddruhů této čeledi.

Čeleď *Musaceae* náleží do řádu *Zingiberales* a sestává ze tří rodů; *Musa*, *Ensete* a *Musella*. Tradiční morfo-taxonomická klasifikace banánovníku vychází z morfologických znaků jednotlivých druhů a základního počtu chromozomů. S rostoucí dostupností molekulárních nástrojů využitelných pro studium genetické diverzity a fylogeneze rostlinných organizmů se objevily studie, které se pokoušejí o ověření správnosti tradiční taxonomie banánovníku.

Ukázalo se, že výsledky morfo-taxonomických a molekulárně-taxonomických studií nejsou plně v souladu. Toto prakticky znemožňuje výběr optimálních rodičů pro šlechtění nových odrůd a zároveň komplikuje rozumný výběr položek pro konzervaci genetické diverzity v genových bankách. Pro vyřešení tohoto problému bude nutné co nejlépe charakterizovat dostupné genové zdroje a také důkladně prostudovat evoluční vztahy mezi druhy banánovníku. Znalost diverzity je klíčová pro efektivní konzervaci genofondu, protože usnadní racionální výběr položek pro genové banky tak, aby postihly existující diverzitu co nejkomplexněji a zároveň pomůže omezit počet duplikovaných či nesprávně označených položek.

Předkládaná dizertační práce si klade za cíl přispět k charakterizaci genových zdrojů banánovníku, objasnit fylogenetické vztahy uvnitř čeledi *Musaceae* a přispět tak k pochopení evoluční příbuznosti zástupců této čeledi. Výsledky této práce by zároveň měly posloužit ke zpřesnění klasifikace jednotlivých druhů, poddruhů a kultivovaných klonů banánovníku.

CÍLE DISERTAČNÍ PRÁCE

1. Příspěť k charakterizaci genových zdrojů banánovníku

Cílem první části disertační práce bylo zvolit molekulární nástroj vhodný k charakterizaci bohatých zdrojů genetické diverzity banánovníku a zjistit možnosti jeho použití pro rozsáhlé genotypování jednotlivých druhů a poddruhů rodu *Musa*.

2. Optimalizovat vhodný systém pro vysokokapacitní genotypování banánovníku

Dalším cílem předkládané práce bylo optimalizovat zvolený systém molekulární charakterizace pro využití ve vysokokapacitním genotypování banánovníku, ověřit jeho reprodukovatelnost na větším souboru vzorků a jeho využitelnost pro molekulární charakterizaci neznámých položek.

3. Příspěť k objasnění evolučních vztahů v rámci čeledi *Musaceae*

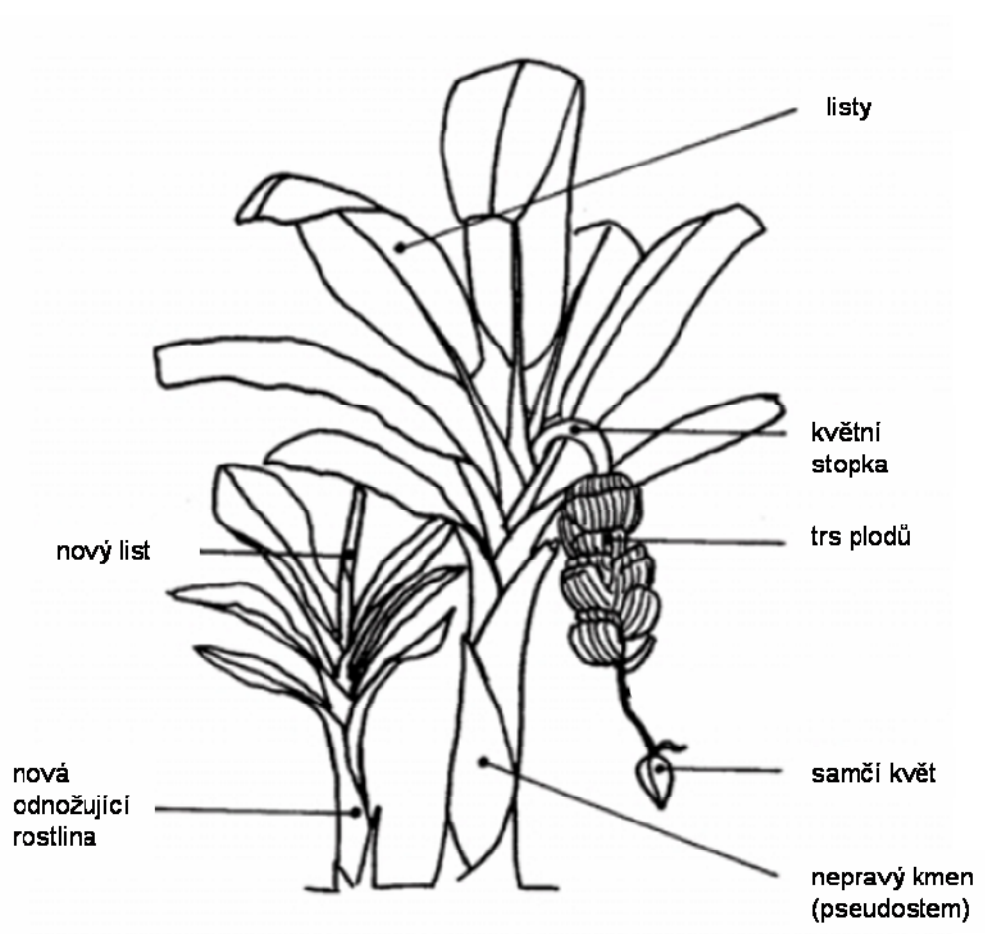
Třetím cílem předkládané disertační práce byla izolace většího souboru genových markerů pro rekonstrukci fylogeneze čeledi *Musaceae* a použití těchto markerů k datování evolučního stáří jednotlivých vývojových linií rodu *Musa*.

LITERÁRNÍ PŘEHLED

1. CHARAKTERISTIKA BANÁNOVNÍKU

1.1 Morfologie banánovníku

Banánovník je tropická rostlina s předpokládaným původem v jihovýchodní Asii, odkud se později rozšířila do tropických a subtropických oblastí Afriky a jižní Ameriky. Přestože kultivované odrůdy běžně narůstají do výšky 2 až 8 m a některé planě rostoucí druhy často dosahují až 15 m, jedná se o jednoděložnou bylinu, která postrádá pro stromy typické dřevnatění a sekundární ztlustění stonku (Karamura & Karamura 1995).



Obr. 1.1: Základní morfologie rostliny banánovníku. Daniells et al. 2001, převzato s úpravami.

Kořenový systém banánovníku je radiální a nazasahuje hluboko do půdy. Ze zakrslého stonku pod úrovní půdy vyrůstá nepravý kmen, tzv. pseudostem, který je tvořen cirkulárně uspořádanými listovými pochvami. Nové listy vyrůstají vždy zevnitř pseudostemu (Obr.1.1). Po vytvoření cca 30-40 listů je prozatím neznámým mechanismem iniciována transformace vegetativního apikálního meristému na generativní a vzniká převislé klasovité květenství. Květy jsou jednopohlavné, přičemž v bazální části květenství vyrůstají květy samičí a v distální části jsou, těsně uzavřeny v samčím pupenu, květy samčí. Plodem banánovníku je bobule, která u kultivovaných jedlých forem neobsahuje semena a tyto druhy banánovníku jsou parthenokarpické, množí se vegetativně pomocí odnoží (Stover & Simmonds 1987). Plané diploidní druhy banánovníku však tvoří plody plné semen a jsou schopny pohlavního rozmnožování.

1.2 Taxonomie banánovníku

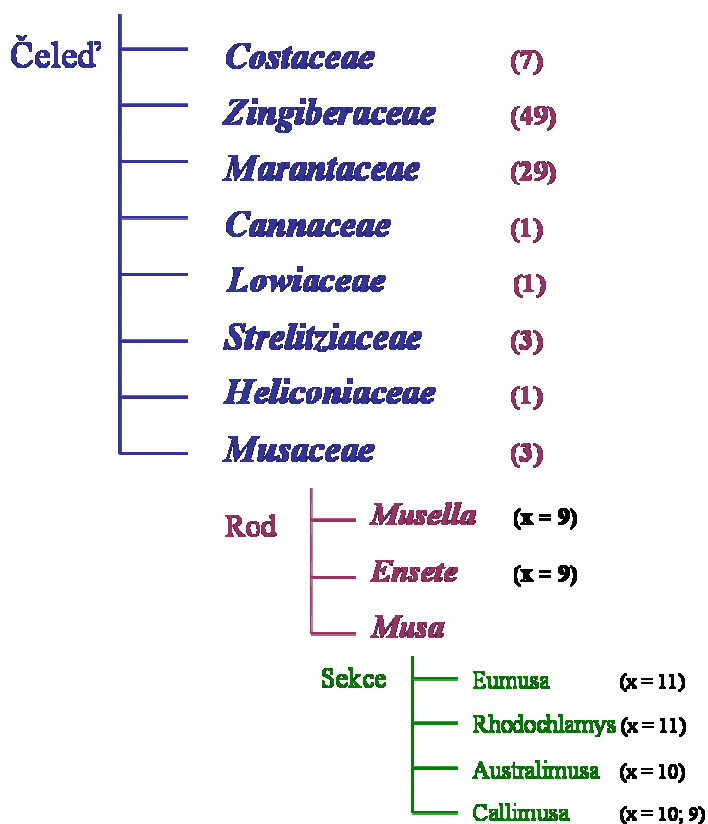
Systematicky je banánovník zařazen do rodu *Musa*, který náleží do čeledi banánovníkovité (*Musaceae*). Tato čeleď společně se sedmi dalšími čeleděmi (Obr.1.2), z nichž některé (např. *Zingiberaceae* – zázvorníkovité) mají podobně velký hospodářský význam, patří do řádu zázvorníkotvaré (*Zingiberales*).

Mimo rod *Musa*, zahrnuje čeleď *Musaceae* ještě další dva samostatné rody, rod *Ensete*, pěstovaný zejména ve východní Africe (tzv. Etiopský banánovník), a monotypický rod *Musella*, jehož rodový status bývá často zpochybňován (Li 1978, Li *et al.* 2010, Liu *et al.* 2010) a bývá proto slučován s rodem *Ensete*.

Tradiční a dodnes používaná klasifikace jednotlivých druhů banánovníku, jichž bylo dosud popsáno více než 40, je založena na základním počtu chromozomů (x) a souboru morfologických deskriptorů (jako např. délka a barva pseudostemu, barva květů, počet a tvar plodů atd.; Cheesman 1947, Simmonds & Shepherd 1955, Simmonds 1962). Podle tohoto systému se rod *Musa* dále rozčleňuje do čtyř sekcí, *Eumusa* [$x = 11$], *Rhodochlamys* [$x = 11$], *Australimusa* [$x = 10$], a *Callimusa* [$x = 9$ a 10]. Přitom sekce *Callimusa* je do jisté míry diskutabilní, neboť druhy do ní náležící se navzájem liší v základním chromozomovém čísle a základem jejich společné klasifikace byla spíše oblast výskytu. Později byla Argenteem (1976) přidána další samostatná sekce, *Ingentimusa* pro druh *Musa ingens* [$x = 7$].

Nejpočetnější a také nejrozšířenější sekcí rodu *Musa* je Eumusa, která zahrnuje dva nejvýznamnější druhy, *Musa acuminata* (A genom) a *M. balbisiana* (B genom) a mimo to také představitele S genomu, druh *M. schizocarpa*. Zástupci T genomu jsou klasifikováni do sekce Australimusa (*M. textilis*). Druhy zahrnuté v sekci Rhodochlamys nevynikají velkou produkcí plodů, jsou však ceněny jako okrasné rostliny pro rozmanitou barevnost květů (Häkkinen & Sharrock 2002).

Řád ZINGIBERALES



Obr. 1.2: Taxonomické zařazení rodu *Musa*. U čeledí jsou v závorce uvedeny počty rodů, u rodů a sekcí je uveden příslušný základní počet chromozomů (x).

Banánovník se vyskytuje v diploidní, triploidní a tetraploidní formě. Nejběžnější jedlé formy banánovníku jsou triploidní vnitrodruhové i mezidruhové hybridní formy AAA, AAB a ABB, odvozené od předchůdců A genomu (*M. acuminata* Colla) a B genomu (*M. balbisiana* Colla). Podle genomového složení rozeznáváme několik

skupin banánů, jsou to Highland bananas a sladké banány tzv. „dessert bananas“ (genom AAA), dále tzv. „plantains“ s genomem AAB a tzv. „cooking bananas“ (genom ABB) používané pro vysoký obsah škrobu spíše jako zelenina (Price 1995). Další významnou skupinu jedlých druhů tvoří tzv. Fe'i banány rostoucí na ostrovech Pacifiku, které jsou typické oranžovou barvou plodů s vysokým obsahem karotenoidů a jsou odvozeny od zástupců sekce Australimusa. Přesný původ těchto dnes pěstovaných forem banánovníku není znám a dnešní podoba těchto klonů je výsledkem dlouholeté selekce původními obyvateli.

1.3 Hospodářský a společenský význam banánovníku

Banánovník představuje jednu z nejdůležitějších hospodářských plodin a v mnoha, zejména rozvojových zemích představuje nenahraditelnou složku v denní výživě obyvatelstva. Celosvětová roční produkce banánů včetně „plantains“ dosahuje téměř 130 milionů tun (FAOstat 2009), přitom ovšem až 90% je zajištěno drobnými pěstiteli a konzumováno v místě produkce. Pouze 10% světové roční produkce je tak určeno pro velkoobchod, ve kterém v současné době dominují kultivary skupiny Cavendish (triploidní AAA genom). Největší spotřebu banánů na obyvatele mají africké země (200 až 250 kg na osobu/rok), kdežto například v Evropě či severní Americe představuje roční konzumace na osobu/rok pouze asi 15 kg. Ve srovnání s Evropou či severní Amerikou, kde převažuje konzumace sladkých, tzv. dezertních banánů, je v Africe také mnohem vyšší obliba tzv. cooking banánů a plantains. Tyto představují v mnoha afrických zemích přímo základní potravinu pro velké skupiny obyvatel.

Banánovník jako rostlina ovšem poskytuje mnohem širší využití než jen jako zdroj plodů. V Indii se banánovník nazývá také „Kalpatharu“, což znamená rostlina se všemožným využitím, jaké si jen lze představit (Roux *et al.* 2008). Listy a pseudostem banánovníku jsou zdrojem vláken, které se zpracovávají pro výrobu látek a lan. Listy se mimo to také využívají jako obalový materiál pro ochranu předmětů při transportu, či jako jednoduché obaly, talíře a misky pro vaření i servírování pokrmů. V některých zemích se také banány používají k výrobě alkoholu či jako krmivo pro hospodářská zvířata. Celá rostlina pak slouží při smíšeném hospodaření ke stínění a vytváření vhodného mikroklimatu pro nižší plodiny, jako např. vanilka, kávovník či pepřovník.

Co do nutriční hodnoty jsou plody banánovníku velmi cenné pro svůj vysoký energetický obsah a vysoký podíl jednoduchých cukrů, při téměř nulovém obsahu tuku a cholesterolu. Banány jsou také zdrojem vlákniny a minerálních prvků, zejména fosforu, vápníku a draslíku (při nízkém obsahu sodíku), ale také zdrojem vitamínu C a A (Chandler 1995).

1.4 Choroby a šlechtění banánovníku

Navzdory každoročně narůstající celosvětové produkci je banánovník neustále ohrožován chorobami a škůdci, což v případě pěstovaných jedlých klonů množících se odnožemi představuje nemalý problém. Dokladem toho je například epidemie tzv. panamské choroby na přelomu 50. a 60. let 20. století, která způsobila obrovské ztráty v celkové produkci do té doby pro komerční trh převážně pěstovaného kultivaru Gros Michel. V důsledku toho se pro účely komerčního pěstování banánovníku začala používat odrůda Cavendish, která je sice vůči panamské chorobě do velké míry odolná, nicméně její klonální množení nadále představuje riziko (Jeger *et al.* 1995). Vedle panamské choroby, způsobované půdním patogenem *Fusarium oxysporum* patří mezi nejčastější choroby i tzv. Black Sigatoka, způsobovaná houbovým patogenem *Mycosphaerella fijiensis* a další choroby virového (tzv. Bunchy top virus) či bakteriálního původu (choroba Moko; *Pseudomonas solanacearum*, Stover & Simmonds 1987).

V současné době není možné chránit se vůči těmto ohrožením jinak než častým chemickým ošetřením pomocí postřiků (až jednou týdně), což ovšem představuje velkou ekologickou zátěž a pro drobné pěstitele je taková ochrana finančně nedostupná. Alternativnímu řešení v podobě vyšlechtění nových odrůd s rezistencí, ale při zachování či zlepšení konzistence a nutričních vlastností plodů prozatím stojí v cestě částečná nebo úplná sterilita a vegetativní množení pěstovaných genotypů. Použití fertálních diploidních, případně tetraploidních druhů zatím nevedlo ke vzniku pro pěstitele přijatelných jedlých odrůd. Klíčem k překlenutí tohoto problému může být přesná charakterizace složení genomů a identifikace druhů které přispěly ke vzniku dnes pěstovaných jedlých odrůd. Pro tuto charakterizaci je nezbytné důkladné prostudování genetické diverzity banánovníku na molekulární úrovni, aby bylo možné plně uplatnit moderní postupy molekulární biologie a biotechnologie pro šlechtění banánovníku.

2. MOLEKULÁRNÍ CHARAKTERIZACE GENETICKÉ DIVERZITY ROSTLIN

2.1 Biodiverzita a genetická diverzita

Biodiverzitu lze chápat jako biologickou rozmanitost na několika úrovních; od DNA a genů, přes druhy až na úroveň populací a ekosystémů. I když je genetická diverzita z hlediska hierarchie na nejnižší úrovni, je základem pro vývoj a přizpůsobivost jedince i populací k nově vzniklým podmínkám prostředí. Z tohoto pohledu se dá genetická diverzita popsat jako různorodost mezi organismy, která je geneticky fixovaná. Studium a ochrana genetické diverzity je důležitá nejenom pro pochopení evoluce dnešních organismů, ale také pro zachování vzácných genových zdrojů využitelných ve šlechtění nových odrůd (Paton 2009, Rands *et al.* 2010). Dokumentace a studium genetické diverzity jsou tedy nepostradatelné pro základní výzkum stejně jako pro jeho aplikace (Miller & Rossman 1995).

V současnosti existují pro mnoho rostlinných druhů, včetně banánovníku tzv. genové banky (<http://www.bioversityinternational.org>), které mají za úkol shromažďovat, dokumentovat a uchovávat jednotlivé položky reprezentující diverzitu daného druhu. Je důležité, aby byly jednotlivé položky, které jsou v genové bance vedeny, ale také ty, které se do ní teprve zavádějí důkladně charakterizovány a předcházelo se tak případným duplikacím či chybám v označení, což jsou typické problémy, se kterými se genové banky potýkají (Zhang *et al.* 2006).

2.2 Nástroje studia genetické diverzity rostlin

Genetickou diverzitu rostlin je možné studovat s použitím mnoha nástrojů, od využití systematické biologie a morfo-taxonomie až po moderní molekulární techniky. Morfologická charakterizace genetické diverzity rostlin sice není náročná na drahé technologie, na druhé straně ovšem vyžaduje značnou zkušenost a je limitována dostupností rostlinných zdrojů, může být výrazně ovlivněna podmínkami okolního prostředí a také její rozlišovací schopnost např. u druhů množících se vegetativně může být omezená (De Langhe *et al.* 2005). U banánovníku je navíc limitující i

potřeba rozsáhlých ploch pro udržování všech hodnocených rostlin v přirozených podmínkách.

Se současným rozvojem a rostoucí dostupností metod molekulární biologie se také rozšířilo použití molekulárních markerů. Molekulární markery popisují diverzitu buď na úrovni primární struktury DNA (DNA markery) nebo na úrovni proteinů a sekundárních metabolitů (biochemické markery).

2.2.1 Biochemické markery

Molekulární markery, které jsou zaměřeny na odhalení polymorfizmu na úrovni proteinů (přítomnost/nepřítomnost proteinu, různě velké formy stejného proteinu, funkčnost/nefunkčnost enzymu), případně sekundárních metabolitů (přítomnost/nepřítomnost metabolitu či jeho rozdílné koncentrace ve stejných pletivech) jsou nazývány biochemické markery. U rostlin je nejčastěji používána analýza isoenzymů, což jsou polymorfnní formy enzymů. Pomocí elektroforetických separačních a imunoanalytických metod lze jednotlivé formy od sebe odlišit a stanovit tak míru genetické diverzity, genetické struktury druhů ale také genetickou strukturu v rámci populace (Mondini *et al.* 2009). Hlavní výhodou isoenzymových markerů je jednoduchost jejich použití, které nevyžaduje předchozí znalost sekvence DNA, primery ani sondy, je poměrně experimentálně a cenově nenáročné a vysoce reprodukovatelné. Naopak mezi jejich slabé stránky patří poměrně nízká úroveň polymorfizmu, a také fakt, že mohou být, stejně jako fenotypové (morfologické) markery ovlivněny vnějšími faktory– např. tím, ze které části rostliny byl izolován materiál pro analýzu (Spooner *et al.* 2005). Proteinové markery jsou většinou kodominantní, čili umožňují rozlišit všechny rodičovské alely, ale zároveň podléhají post-translačním modifikacím a jejich výskyt u rostlin je limitován (Kumar 1999).

2.2.2 Molekulární DNA markery

Jako DNA markery označujeme takové markery, které studují diverzitu přímo na úrovni primární struktury DNA. Narozdíl od morfologických a biochemických markerů jsou DNA markery nezávislé na vnějším prostředí. Jejich použití je založeno na rozpoznání polymorfizmů v sekvenci DNA, neboli na detekci inzercí, delecí, translokací, duplikací a bodových mutací. Tyto změny genetické informace

mohou být studovány v jakékoliv části genomu a na materiálu pocházejícím z kterékoliv části rostliny. Genetická variabilita mezi organizmy existující na úrovni DNA je mnohonásobně vyšší než variabilita projevující se fenotypově, protože velká část, zejména bodových mutací v DNA se nepromítá do sekvence proteinu a tím neovlivní jeho strukturu, expresi ani funkci (Burr 1994). Je tedy zřejmé, že zavedení molekulárních markerů pro charakterizaci, neboli genotypování rostlin značně rozšířilo hranice studování genetické diverzity od evolučně velmi vzdálených až po evolučně velmi příbuzné a fenotypově těžko odlišitelné (Ford-Lloyd 2001, Avise 2004). Zároveň, ve srovnání s časově náročnými postupy klasického šlechtění, fenotypovým hodnocením potomstva a selektováním jedinců s lepšími vlastnostmi, poskytují dnešní molekulární metody velmi rychlý a účinný nástroj (tzv. „marker assisted selection“) pro moderní šlechtění (Winter & Kahl 1995, Langridge & Chalmers 2004).

Ideální DNA marker splňuje několik základních požadavků, mezi něž patří četné a rovnoměrné zastoupení v genomu studovaného organismu, vysoká frekvence polymorfizmu, reprodukovatelnost analýzy a její finanční a experimentální nenáročnost a také, v dnešní době čím dál podstatnější, předpoklad pro vysokou kapacitu (tzv. „high-throughput“) a automatizaci analýzy (Agarwal *et al.* 2008). Navíc je výhodné, pokud pro použití takového DNA markeru není nutná předchozí sekvenční informace o genomu a pokud nevyžaduje izolaci velkého množství DNA. Navzdory poměrně širokému výběru markerů, které se v současné době rutinně používají pro genotypování a analýzu genetické diverzity, žádný z nich nekombinuje všechny uvedené vlastnosti a volba ideálního markeru tak zůstává spojena s konkrétním účelem jeho aplikace. Rozhodování o vhodném typu markerů se většinou řídí poměrem mezi jeho praktičností a přesností, s jakou je schopen detekovat genetickou variabilitu (Sunnucks 2000). Zde je také zásadní rozdíl mezi markery kodominantními (rozlišující všechny alely) a dominantními, které neumožňují rozlišit heterozygoty (Tab. 2.1). Dominantní charakter tedy předurčuje hodnocení DNA fragmentů pouze na přítomnost/nepřítomnost (0/1 systém skórování) a tím limituje možnosti analýzy, jako např. určení frekvence jednotlivých alel apod. (Pérez *et al.* 1998, Hurme & Savolainen 1999). Kodominantní markery jsou sice často jednolokusové a tím méně praktické, protože poskytují menší objem hodnotitelných fragmentů, na druhou stranu však umožňují identifikaci jak recesivních tak dominantních alel, čímž se významně zvyšuje jejich přesnost

(Sunnucks 2000). Navíc, s rozvojem technologií je dnes možné provádět paralelní analýzu mnoha jednolokusových markerů např. na mikročipech a tak postupně mizí rozdíl v praktičnosti z hlediska množství najednou skórovatelných markerů.

Podle principu experimentální metody, na které jsou DNA markery založeny, je lze rozčlenit do tří skupin, a to markery založené na hybridizaci, markery založené na polymerázové řetězové reakci (PCR) a markery založené na sekvenování. I když se pro studium genetické diverzity rostlin uplatňuje široká škála DNA markerů, jsou mezi nimi takové, které se užívají ve větší míře, a pouze těmto bude v následujícím textu věnována pozornost.

Tab. 2.1: Porovnání vlastností vybraných (nejčastěji používaných) typů markerů.

Vlastnosti	RFLP	DArT	RAPD	AFLP	SSR	SNP
Princip	hybridizace	hybridizace	PCR	PCR	PCR	sekvenování
Nároky na kvalitu DNA	vysoká	vysoká	vysoká	střední	střední	vysoká
Nároky na vstupní množství DNA (µg)	10	0.05	0.02	0.5-1.0	0.05	0.05
Četnost výskytu	střední	vysoká	vysoká	vysoká	vysoká	velmi vysoká
Pracnost použití	pracné	jednoduché	jednoduché	jednoduché	jednoduché	jednoduché
Možnost automatizace	ne	ano	ano	ano	ano	ano
Reprodukovatelnost	vysoká	vysoká	nízká	vysoká	vysoká	vysoká
Náročnost vývoje	střední	vysoká	nízká	střední	vysoká	vysoká
Charakter dědičnosti	kodominantní	dominantní	dominantní	dominantní	kodominantní	kodominantní

2.2.2.1 DNA markery založené na hybridizaci

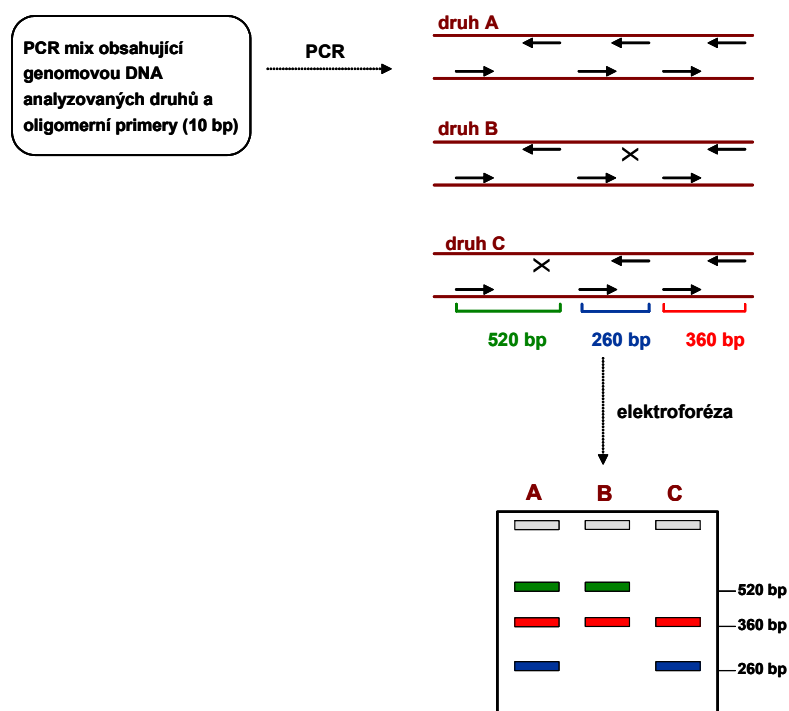
Typickým představitelem DNA markerů založených na hybridizaci a zároveň jedním z prvních v praxi používaných DNA markerů vůbec (Botstein *et al.* 1980) jsou **RFLP** markery (Restriction Fragment Length Polymorphism). Po štěpení genomové DNA specifickou restriční endonukleázou vznikají fragmenty DNA, které se mohou vlivem inzercí, delecí, či bodových mutací mezi jednotlivými jedinci lišit. Po jejich elektroforetickém rozdělení a vizualizaci pomocí hybridizace se značenou homologní sondou lze rozlišit různé restriční profily reprezentující genetickou variabilitu studovaných organismů (Weising *et al.* 2005). Právě RFLP sondy, které jsou tvořeny prostřednictvím knihoven genomové DNA anebo cDNA (komplementární DNA),

představují největší nedostatek této metody. Jejich konstrukce a použití je časově i finančně náročné a není kompenzováno extrémně vysokou mírou detekovatelného polymorfizmu (Agarwal *et al.* 2008). Proto i když byly tyto markery s úspěchem použity pro některé rostlinné druhy (Tanksley *et al.* 1989, Ragot & Hoisington 1993, Takasaki *et al.* 2006;) a jejich reprodukovatelnost je vysoká, jsou postupně nahrazovány modernějšími technikami, jako jsou tzv. DArT markery (Diversity Array Technology).

Technologie **DArT** je v podstatě variací metody RFLP s prvky AFLP (viz níže). Také využívá jako výchozí krok štěpení DNA, restriční endonukleázy jsou zde však dvě, z toho jedna má vzácněji se vyskytující metylsenzitivní rozpoznávací místo štěpení, což znamená produkci méně DNA fragmentů a redukci komplexity genomu. Vzniklé fragmenty se poté ligují do vektoru, který nese adaptéry pouze pro konce vzniklé vzácněji štěpicím enzymem. Tímto způsobem je vytvořena tzv. genomová reprezentace a genom studovaného organismu je zredukován na menší množství fragmentů DNA, obohacených o fragmenty variabilní, tedy polymorfni (Jaccoud *et al.* 2001). Zároveň je díky použití metylsenzitivního enzymu genomová reprezentace obohacena o sekvence s nízkým počtem kopií, často pocházející z genových oblastí. Narozdíl od RFLP se zde polymorfni fragmenty nedetekují hybridizací na mebráně, ale na speciálním čipu, který obsahuje genomovou reprezentaci studovaného druhu. Sondu tvoří také genomová reprezentace, připravená z DNA konkrétní studované rostliny. Použití čipu nabízí obrovskou kapacitu pro paralelní skrining a tato technologie byla již aplikována pro studium genetické diverzity mnoha rostlinných druhů, mezi nimi např. ječmen (Wenzl *et al.* 2004), pšenice (Akbari *et al.* 2006), trávy (Kopecký *et al.* 2009), rýže (Xie *et al.* 2006) a dalších, včetně banánovníku (Risterucci *et al.* 2009). Vývoj čipu pro nový druh je však poměrně finančně náročný, stejně jako analýza v případě, že máme zájem o studium pouze malého počtu vzorků a nenaplní se tak kapacita metody, protože vzrůstá jednotková cena (tzv. datapoint). Jistou nevýhodou metody je také to že, DArT markery jsou dominantní. Celá technologie se poskytuje jako servis bez možnosti provést experimentální část samostatně na svém souboru vzorků, což lze chápat jako výhodu i jako překážku jejího použití pro konkrétní projekt.

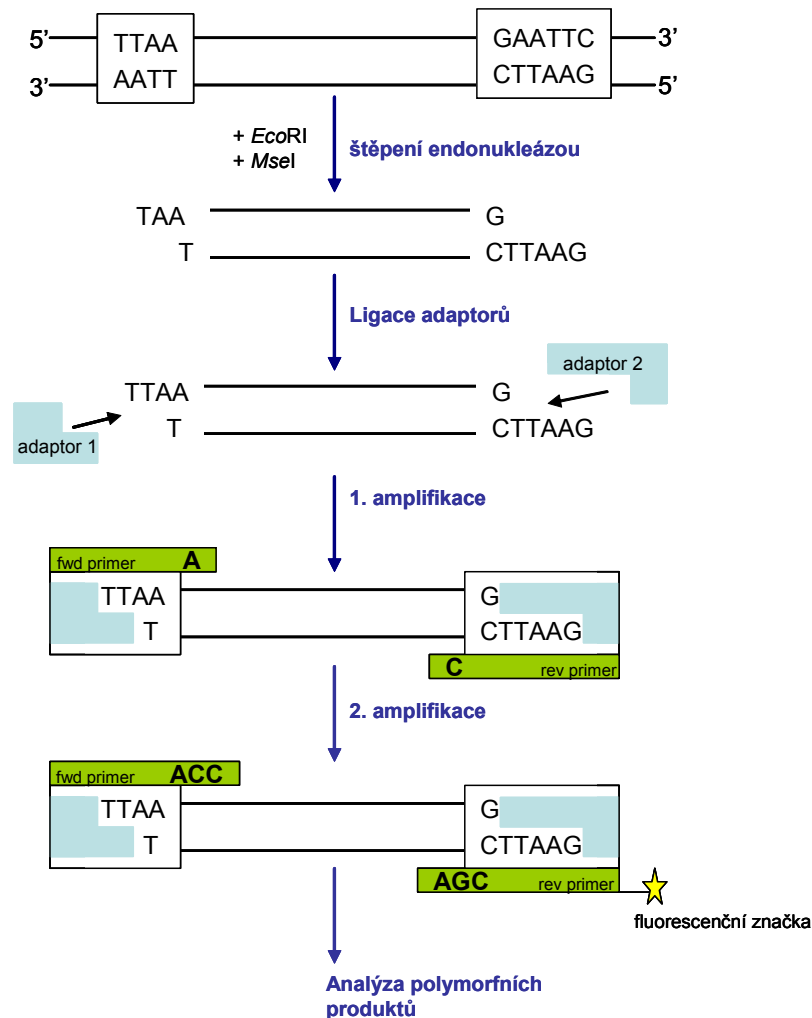
2.2.2.2 DNA markery založené na metodě PCR

S nástupem metody polymerázové řetězové reakce (PCR; Mullis & Faloona 1987) se otevřel prostor pro rozvoj rychlých a experimentálně jednoduchých molekulárních markerů využitelných v genotypování rostlin. Jedním z příkladů jsou **RAPD** markery (Random Amplified Polymorphic DNA; Williams *et al.* 1990), metoda, která využívá nespecifické oligomerní primery pro PCR (délka kolem 10 bp). Tyto primery nasedají na komplementární místa v genomu studované rostliny a z nich se amplifikují náhodné fragmenty DNA, které se mezi jedinci mohou lišit např. v důsledku mutace v místě nasednutí primeru anebo delece/inzerce mezi dvěma takovými místy (Obr. 2.1). Po elektroforetickém rozdělení vzniklých fragmentů tak lze detekovat polymorfizmy dokumentující genetickou diverzitu studovaných jedinců. Vzhledem k tomu, že se používají náhodně generované primery, je u metody RAPD splněna podmínka rovnoměrného pokrytí genomu studované rostliny. Extrémní jednoduchost této metody jí vysloužila poměrně velkou popularitu pro genotypování rostlin (např. Virk *et al.* 1995, Siles *et al.* 2000, Sales *et al.* 2001, Birmeta *et al.* 2002) i přesto, že se jedná o markery dominantní, zatížené poměrně nízkou reprodukovatelností značně závislou na reakčních podmínkách (Ellsworth *et al.* 1993, Muralidharan & Wakeland 1993).



Obr. 2.1: Schéma znázorňující princip detekce polymorfizmu metodou RAPD. Upraveno podle Weising *et al.* 2005.

Později byla vyvinuta metoda s mnohem vyšší spolehlivostí opakování, nazvaná **AFLP** (Amplified Fragment Length Polymorphism; Vos *et al.* 1995). AFLP markery jsou také dominantní, nevyžadují předchozí znalost sekvence DNA pro návržení PCR primerů a byly úspěšně aplikovány v molekulární charakterizaci rostlin (např. Kardolus *et al.* 1998, Koopman *et al.* 2001, Després *et al.* 2003). Princip metody spočívá ve štěpení genomové DNA studovaného organismu dvěma různými enzymy, které zanechávají kohezní konce, následovaném ligací specifických adaptérů o známé sekvenci na konce vzniklých fragmentů. Adaptéry pak tvoří místa pro nasednutí specifických značených primerů a PCR-amplifikaci. Pro redukci příliš velkého množství takto vzniklých fragmentů jsou primery používané k amplifikaci tzv. selektivní, tedy jsou prodlouženy vůči adaptéru o jednu až tři báze (Obr. 2.2).



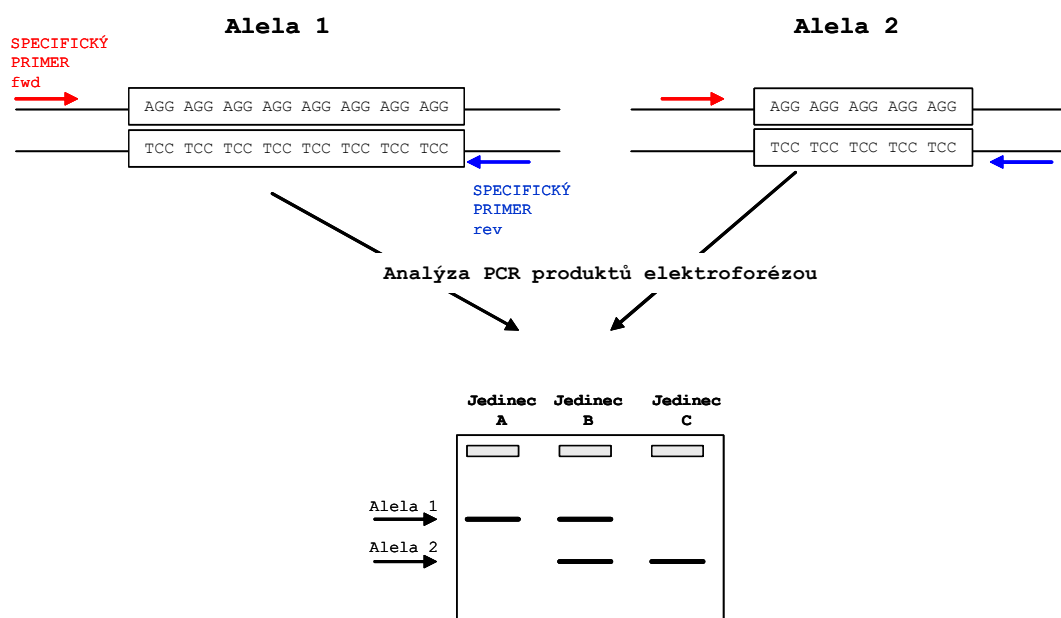
Obr. 2.2: Schematické znázornění principu metody AFLP. Upraveno podle Weising *et al.* 2005.

Podmínky reakce jsou nastaveny tak, aby k amplifikaci došlo pouze tehdy, je-li komplementarita obou primerů k cílovému místu stoprocentní. Statisticky se tak při 1-bázovém prodloužení selektivního primeru sníží množství vzniklých fragmentů 16krát, při 3-bázovém prodloužení až 4096krát, což významně usnadní hodnocení výsledku. Pakliže se navíc použije fluorescenční značení primerů, lze zvýšit kapacitu i přesnost metody použitím kapilární elektroforézy. Přes veškeré výhody, AFLP markery trpí i několika nedostatky, jako jsou vysoké požadavky na kvalitu DNA a její relativně vyšší množství a také ne zcela rovnoměrnou distribuci v genomu v závislosti na použitých restričních enzimech (Young *et al.* 1999).

Zástupcem kodominantních markerů mezi markery založenými na PCR jsou markery odvozené od sekvencí mikrosatelitní DNA, neboli **SSR** (Simple Sequence Repeat) markery. Jedná se o krátké (1-6 bp), tandemově uspořádané repetitivní sekvence, které byly poprvé objeveny v genomu kraba (Skinner *et al.* 1974). Později bylo zjištěno, že jsou tyto sekvence přítomny v genomech většiny živých organismů, rostliny nevyjímaje (Weising *et al.* 1989, Beyermann *et al.* 1992), a že je lze velmi dobře uplatnit pro studium genetické diverzity, či genetické mapování (např. Prasad *et al.* 1999, Gupta & Varshney 2000).

V průběhu evoluce při mnohonásobných cyklech replikace genomové DNA dochází občas v místech přítomnosti SSR motivů k tzv. sklouznutí DNA polymerázy (polymerase slippage), kdy se replikovaný řetězec DNA posune o jednu nebo několik repetitivních jednotek a vznikají tak různě dlouhé varianty stejného lokusu lišící se o několik párů bází. Právě tento jev je považován za nejčastější příčinu vzniku genetické variability, detekovatelné jako polymorfismus SSR markerů (Tautz *et al.* 1986, Schlötterer & Tautz 1992). V praxi se používají specifické primery, které jsou navrženy z oblasti těsně obklopující mikrosatelitový lokus, je tedy nutné znát sekvenci DNA v místě hodnoceného markeru. Z těchto primerů se amplifikují délkově polymorfnní DNA fragmenty reprezentující jednotlivé alely hodnoceného lokusu, které lze separovat na polyakrylamidovém gelu a detekovat barvením (Obr. 2.3). Při využití fluorescenčně značených primerů ve spojení s kapilární elektroforézou lze pak docílit rozlišení jednotlivých alel lišících se o jeden nukleotid (Weising *et al.* 2005). Bylo prokázáno, že ve srovnání s ostatními zde dosud popsanými markery mají mikrosatelity největší informativní hodnotu (měřitelnou jako tzv. polymorphism information content; Powell *et al.* 1996, Nybom 2004). Jejich relativní nevýhodou je velká pracnost a finanční náročnost vývoje. I tak ale

zůstávají SSR markery jedním z nejpoužívanějších nástrojů pro studium a charakterizaci genetické diverzity, genotypování rostlin, populační studie, mapování genomu i pro výběr vhodných jedinců pro šlechtění rostlin (marker assisted selection; Jarne & Lagoda 1996).



Obr. 2.3: Schéma amplifikace mikrosatelitového lokusu. Alely 1 a 2 se vzájemně liší o 3 repetitivní jednotky. Produkty amplifikace se separují elektroforézou a lze rozlišit homozygotní jedince (A a C) od heterozygotního jedince (jedinec B). Upraveno podle Weising et al. 2005.

2.2.2.3 DNA markery založené na sekvenování

Nejvíce přímočarým způsobem jak studovat variabilitu na úrovni DNA je samozřejmě získat její nukleotidovou sekvenci, neboli osekvenovat definované úseky DNA příbuzných organismů, v ideálním případě celé rostlinné genomy, a ty poté vzájemně srovnávat na nukleotidové úrovni. V dnešní době, kdy je sekvenování stále dostupnější a existují databáze s částečnou nebo úplnou sekvenční informací mnoha, nejen rostlinných organismů, je tento přístup ke studiu genetické diverzity dosažitelný a otevírá se prostor pro tzv. markery nové generace, mezi něž se řadí i **SNP** markery (Single Nucleotide Polymorphism). Jedná se o zachycení jednonukleotidových rozdílů v sekvenci DNA různých jedinců. Tyto varianty DNA jsou někdy také označovány jako bodové polymorfizmy. Vyskytují se v genomech

všech živých organismů, přičemž vysoká frekvence jejich výskytu u rostlin (např. u kukuřice každých 60-120 bp; Ching *et al.* 2002, u eukalyptu dokonce každých 16-33bp; Külheim *et al.* 2009) z nich činí nejpočetnější markery vůbec (Gupta *et al.* 2001). I když teoreticky může mít SNP lokus čtyři různé varianty - čtyři různé nukleotidy, ve skutečnosti velmi výrazně převažují bialelické SNP (mající dvě varianty), které vznikají v důsledku tranzice či transverze (Weising *et al.* 2005). Velkou výhodou SNP je vedle jejich četnosti také kodominantní charakter dědičnosti a relativně rovnoměrná distribuce v genomu, i když v souladu s očekáváním lze pozorovat častější výskyt SNP v nekódujících oblastech genomu, jako jsou introny, repetitivní sekvence či pseudogeny (Balasubramanian *et al.* 2002).

Identifikace SNP lokusů v genomu studovaného organismu je možná buď experimentálně, nebo *in silico* s použitím dostupných sekvenčních dat v databázích.

Pro organismy u kterých není známa sekvence genomu se nejčastěji využívá databázi tzv. EST (expressed sequence tag) sekvencí, které pocházejí z exprimované části genomu a jsou tak obohaceny o genové oblasti (Batley *et al.* 2003). V experimentálním přístupu se využívají různé metody, od převedení jiného typu markeru – např. AFLP nebo RFLP na SNP pomocí vytvoření SNP-specifických primerů, přes hybridizaci na mikročipu, až po přímé sekvenování kandidátních oblastí (např. genů). Způsobů genotypování SNP markerů je rovněž několik, mezi nimi např. alelově specifická PCR, kdy se používá speciálně navržených primerů tak, aby docházelo k amplifikaci pouze jedné z variant SNP (Wu *et al.* 1989). Dalším způsobem je tzv. single base extension metoda, neboli prodloužení specifického primeru pouze o jeden nukleotid právě v místě SNP, přičemž nově zabudovaný nukleotid je fluorescenčně značený a podle této značky lze rozlišit konkrétní variantu SNP (Sokolov 1990). Mimo to lze také použít alelově-specifickou hybridizaci oligonukleotidů (často využívaná při vysokokapacitním genotypování na mikročipech) nebo alelově-specifickou ligaci. Všechny přístupy však mají společný rys, kterým je poměrně náročná optimalizace. Pro účely studia genetické diverzity rostlin je vhodné hodnotit paralelně velké množství SNP markerů, což je v současné době dostupné jako servis pro některé druhy, nicméně finanční náročnost vývoje i aplikace těchto čipů je vysoká. S prudkým rozvojem nových vysoce výkonných platform pro sekvenování – tzv. next generation sequencing (454/Roche, Margulies *et al.* 2005; Illumina/Solexa, Shaffer 2007; ABI SOLID, Valouev *et al.* 2008; Helicos Heliscope, Harris *et al.* 2008; Pacific Biosciences SMRT, Eid *et al.* 2009) se

stává sekvenování i celých rostlinných genomů mnohem dostupnější. Je tak zřejmě jen otázkou času, kdy budou k dispozici celé databáze SNP pro mnoho rostlinných druhů a SNP markery se zřejmě stanou nejvyužívanějším nástrojem nejen pro studium genetické diverzity rostlin.

2.3 Genetická diverzita a její charakterizace u banánovníku

V průběhu několika tisíc let domestikace banánovníku vzniklo obrovské množství různých kultivarů, které jsou velmi cenným nositelem genetické diverzity využitelné ve šlechtění nových druhů. Pro ochranu těchto genových zdrojů byl sestaven soubor obecně doporučených pokynů pod názvem *The Global Conservation Strategy for Musa* (INIBAP 2006), který zahrnuje charakterizaci genetické diverzity a racionalizaci národních a mezinárodních genových bank pro její uchování. Mezinárodní kolekce genotypů banánovníku je uchována pod záštitou Bioversity International v International Transit Centre při Katolické Univerzitě v belgické Lovani (<http://bananas.bioversityinternational.org/>). Tato sbírka obsahuje více než 1000 položek, a je neustále rozšiřována o zajímavé druhy banánovníku různého původu. Vedle velmi dobře zavedené morfo-taxonomické klasifikace banánovníku (kapitola 1.2) vyvstává postupně i otázka hledání nejlepšího možného způsobu pro molekulární charakterizaci položek jednak v mezinárodní sbírce, ale také jakýchkoli sporných druhů, které se mohou objevit mezi vědeckou veřejností zabývající se studiem banánovníku. Od 80. let 20. století se objevují práce popisující využití molekulárních markerů pro studium genetické diverzity banánovníku (Tab. 2.2).

Většina těchto prací popisuje relativně úzký soubor genotypů, často se jedná o lokálně významné genotypy. Přitom pro zamezení zkreslení výsledků, je vhodné zahrnout do analýzy co největší počet taxonů a studovat ho pomocí co největšího počtu markerů. V několika případech je zahrnut větší soubor zástupců, který má ambice reprezentovat diverzitu v rámci rodu *Musa* (Ude *et al.* 2002b, Wong *et al.* 2002, Risterucci *et al.* 2009) a některé výsledky ukazují na potřebu reklasifikace – např. Wong *et al.* (2002) ve své studii navrhli sloučení sekce *Eumusa* a *Rhodochlamys*, a sloučení sekce *Australimusa* se sekcí *Callimusa*. Nicméně, dosud nebyl stanoven standard pro způsob genotypování jednotlivých druhů banánovníku, který by sloužil jak širší vědecké veřejnosti, tak pro účely molekulární charakterizace položek genové banky banánovníku.

Tab. 2.2: Přehled vybraných prací zabývajících se použitím molekulárních markerů pro studium genetické diverzity banánovníku.

Studie	Účel studie	Použité markery	Studované druhy
Jarret a Litz, 1986	analýza genetické diverzity	isoenzymy	35 zástupců <i>M. acuminata</i> a <i>M. balbisiana</i>
Gawel <i>et al.</i> 1992	fylogenetická studie	RFLP	19 zástupců rodu <i>Musa</i>
Jarret <i>et al.</i> 1992	fylogenetická studie	RFLP	26 lokálních druhů (Papua Nová Guinea)
Lagoda <i>et al.</i> 1998	vývoj SSR markerů pro čeleď <i>Musaceae</i>	SSR	--
Grapin <i>et al.</i> 1998	analýza genetické diverzity	SSR (STMS)*	59 planých i kultivovaných zástupců <i>M. acuminata</i>
Pillay <i>et al.</i> 2000	identifikace RAPD markerů specifických pro A a B genom banánovníku v hybridech	RAPD	40 triploidních hybridních genotypů
Pillay <i>et al.</i> 2001	analýza genetické diverzity	RAPD	29 lokálních kultivarů tzv. highland banánů (východní Afrika)
Wong <i>et al.</i> 2001	analýza genetické diverzity	AFLP	32 lokálních zástupců <i>M. acuminata</i> (Malajsie)
Ude <i>et al.</i> 2002a	analýza genetické diverzity	AFLP	28 zástupců A a B genomu (diploidní i triploidní hybridní)
Ude <i>et al.</i> 2002b	analýza genetické diverzity a klasifikace	AFLP	39 zástupců ze 4 sekcí rodu <i>Musa</i>
Wong <i>et al.</i> 2002	validace klasifikace rodu <i>Musa</i>	AFLP	21 zástupců ze 4 sekcí rodu <i>Musa</i>
Creste <i>et al.</i> 2003	analýza genetické diverzity	SSR	35 lokálních zástupců rodu <i>Musa</i> (Brazílie)
Noyer <i>et al.</i> 2005	analýza genetické diverzity	SSR, AFLP, MSAP #	30 zástupců plantains
Ge <i>et al.</i> 2005	studie populační struktury	SSR PCR-RFLP	267 lokálních zástupců z 15 populací <i>M. balbisiana</i> (Čína)
Wang <i>et al.</i> 2007	analýza genetické diverzity	AFLP	218 lokálních zástupců z 15 populací <i>M. balbisiana</i> (Čína)
Risterucci <i>et al.</i> 2009	analýza genetické diverzity	DArT	168 zástupců ze sbírek CIRAD a IITA
Till <i>et al.</i> 2010	identifikace DNA polymorfizmu pomocí Ecotillingu	SNP	80 zástupců rodu <i>Musa</i>

* STMS = speciální druh SSR markerů se sekvenční adresou - sequence tagged microsatellite sites

MSAP = methylation sensitive amplification polymorphism

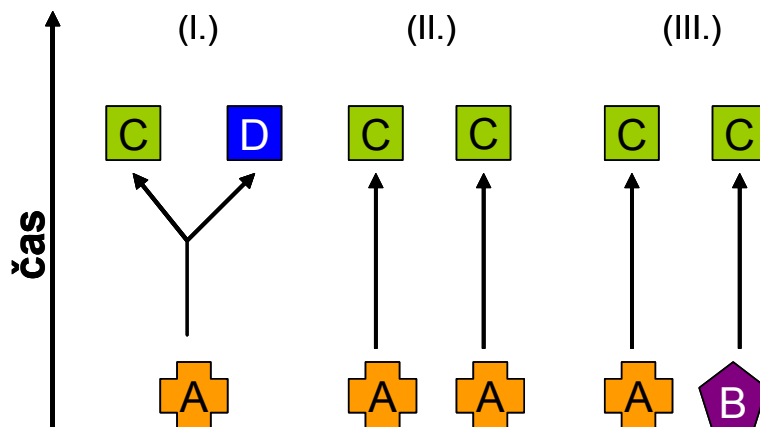
3. MOLEKULÁRNÍ FYLOGENETIKA ROSTLIN

3.1 Molekulární data a molekulární fylogenetika

Data generovaná prostřednictvím molekulárních markerů jsou, jak již bylo řečeno odrazem variability na úrovni DNA. Tato variabilita vzniká mutacemi DNA, které jsou zdrojem evolučních novinek a v průběhu času se akumulují v genomech jednotlivých organismů, čímž zanechávají jakousi molekulární stopu evoluce. Proto lze využít data dokumentující variabilitu DNA (např. generovaná molekulárními markery) nejen pro studium genetické diverzity organismů, ale je možné je také zpětně převést na evoluční vztahy mezi organismy neboli zrekonstruovat obraz evoluce živých organismů – fylogenetický strom.

Molekulární fylogenetika je progresivní mladá vědní disciplína s počátky v 60. letech 20. století, která se zabývá právě popisem evolučních vztahů mezi druhy na základě molekulárně-biologických dat. Vedle molekulárních markerů (viz kapitola 2) se s masivním nárůstem dostupnosti sekvenační technologie zpřístupnil zdroj velmi užitečných sekvenčních informací pro molekulární fylogenetiku. Existuje hned několik důvodů, proč jsou molekulární znaky (zejména pak sekvence DNA) efektivnějším nástrojem pro evoluční biologii, nežli data klasická (morfologická, či fyziologická). Především je na molekulární úrovni k dispozici mnohem více informativních znaků a ve srovnání s klasickými znaky jsou lépe přizpůsobitelné kvantitativnímu hodnocení (Prakash 2007). Dalším důvodem je, že data molekulární lze snadněji uplatnit ve zkoumání vzájemně vzdálenějších druhů a zároveň jsou tyto znaky často selekčně neutrální (bez vlivu na přirozený výběr), tedy jsou obrazem evoluční příbuznosti druhů, nikoli příbuznosti selekčních tlaků, které na ně působily (Flegr 2005). To je důležité zejména z hlediska problému odlišení konvergentní a paralelní evoluce. Při klasické divergentní nebo také darwinovské evoluci, se postupně hromadí rozdíly mezi dvěma druhy pocházejícími ze stejného předka, až dojde ke vzniku nových, odlišných druhů. U konvergentní evoluce však dochází naopak vlivem náhodného působení obdobného selekčního tlaku k vývoji podobných znaků u druhů vzájemně odlišných, které nesdílí společného předka. Navíc existuje mechanismus tzv. paralelní evoluce (Obr. 3.1), kdy se dva podobné druhy nezávisle na sobě vyvíjí podobným způsobem vlivem stabilizujícího vnějšího selekčního tlaku,

aniž by přitom měly společného předka (Niklas 1997). Odlišit tyto rozdílné evoluční pochody pomocí klasických, např. morfologických znaků je velmi obtížné. Ovšem při použití molekulárních dat tento problém není tak zásadní, protože i při působení stejného vnějšího selekčního tlaku nevzniknou na úrovni DNA u nepříbuzných druhů stejné neutrální mutace. I v případě, že dojde k fixaci několika mutací vzniklých konvergencí či paralelní evolucí, nedochází k ovlivnění výsledku fylogenetické analýzy, protože ta většinou pracuje s velkým souborem molekulárních znaků, v němž se statisticky vliv těchto náhodných mutací eliminuje (Flegr 2005).



Obr. 3.1: Schematické znázornění divergentní, paralelní a konvergentní evoluce. U divergentní evoluce (I.) se ze společného předka (A) v průběhu času vytvoří dva druhy s odlišným homologním znakem (C a D). Paralelní evoluce (II.) představuje vznik dvou podobných znaků (C) u podobných předků (A) nezávisle na sobě, působením stejného selekčního tlaku. V případě konvergentní evoluce (III.) se u dvou různých předků (A a B) vyvinou podobné znaky (C). (upraveno podle Flegr 2005)

I v případě molekulární fylogenetiky však platí, že by neměla opomíjet klasická data pocházející z fenotypových znaků. Zejména u prací zabývajících se taxonomií je důležité vzít do úvahy také morfologická data. I když nám totiž molekulární znaky mohou dát odpověď na otázku jaká je kladogeneze (změny vlastností mající za následek odštěpení vývojové linie) studovaných druhů, bez zapojení klasických morfologických znaků nemůžeme rozklíčovat anagenezi (změny vlastností v rámci

vývojové linie) a tudíž nelze spolehlivě rozhodnout, zda odštěpeným větvím přísluší postavení samostatného taxonu (Stuessy 2009).

3.2 Sekvence DNA pro molekulární fylogenetiku

Sekvence DNA, jak bylo zmíněno výše, představuje nejvhodnější a nejpřímější zdroj molekulárních dat pro účely studia fylogeneze živých organismů. V rámci rostlinného genomu, tak jako i u jiných organismů, však existují rozdíly v tom, do jaké míry jsou jednotlivé jeho úseky vhodné či méně vhodné jako zdroj fylogeneticky informativních znaků. Fixace neutrálních mutací dokumentujících evoluci daného druhu není totiž ve všech oblastech genomu stejná. Obecně lze větší pravděpodobnost výskytu selekčně neutrálních mutací očekávat u nekódující DNA (např. v pseudogenech, či intronech; Hartl 2011). Mírou rychlosti molekulární evoluce je tzv. substituční rychlost, která představuje počet fixovaných mutací v dané pozici DNA za časovou jednotku, většinou se vyjadřuje za rok. Substituční rychlost se liší také mezi třemi rostlinnými genomy - mitochondriálním, chloroplastovým a jaderným genomem. Jaderný genom vykazuje u rostlin nejvyšší substituční rychlost, zatímco genom mitochondriální je u rostlin naopak poměrně vysoce konzervovaný (Gaut 1998). Je přitom zajímavé, že u živočichů je situace jiná a mitochondriální genom je naopak pro svou velmi vysokou rychlost molekulární evoluce velmi často využíván v molekulárně-fylogenetických studiích (Brown *et al.* 1979, Harrison 1989, Hyman *et al.* 2007). V chloroplastovém genomu rostlin se co do substituční rychlosti liší oblast variabilnějších sekvencí jednokopiových, včetně genů kódovaných chloroplastovým genomem a oblast tzv. obrácených repetitivních úseků (inverted repeats), která je naopak více konzervovaná (Wolfe *et al.* 1987).

Při výběru vhodné části genomu, resp. vhodného genu pro fylogenetickou analýzu je rovněž důležité posoudit rychlost evoluce ve vztahu k evoluční vzdálenosti studovaných druhů. Pakliže chceme studovat druhy, o kterých víme, že jsou vzájemně poměrně blízce příbuzné, je dobré zvolit takovou oblast genomu, která má vysokou substituční rychlost (např. pseudogeny). Naopak, budeme-li studovat evolučně vzdálené organizmy, vhodnějším zdrojem dat bude ta oblast genomu, která se vyvíjela pomaleji (např. exony).

Dalším faktorem, který by neměl být opomíjen při volbě zdroje molekulárních dat pro fylogenetickou analýzu je počet a velikost/délka analyzovaných lokusů. Bylo

ukázáno, že pro zlepšení přesnosti fylogenetické analýzy je vhodnější použít více vzájemně nezávislých DNA lokusů, které mohou být i relativně kratší, než postavit analýzu na jediném, byť i dlouhém úseku DNA (Rokas *et al.* 2003, Carling & Brumfield 2007). V praxi je ovšem konkrétní projekt často limitován dostupností sekvencí vhodných markerů a finančními zdroji a je proto nutné najít kompromis mezi počtem lokusů a počtem taxonů zahrnutých do analýzy, protože s každým přidaným lokusem cena analýzy exponenciálně narůstá.

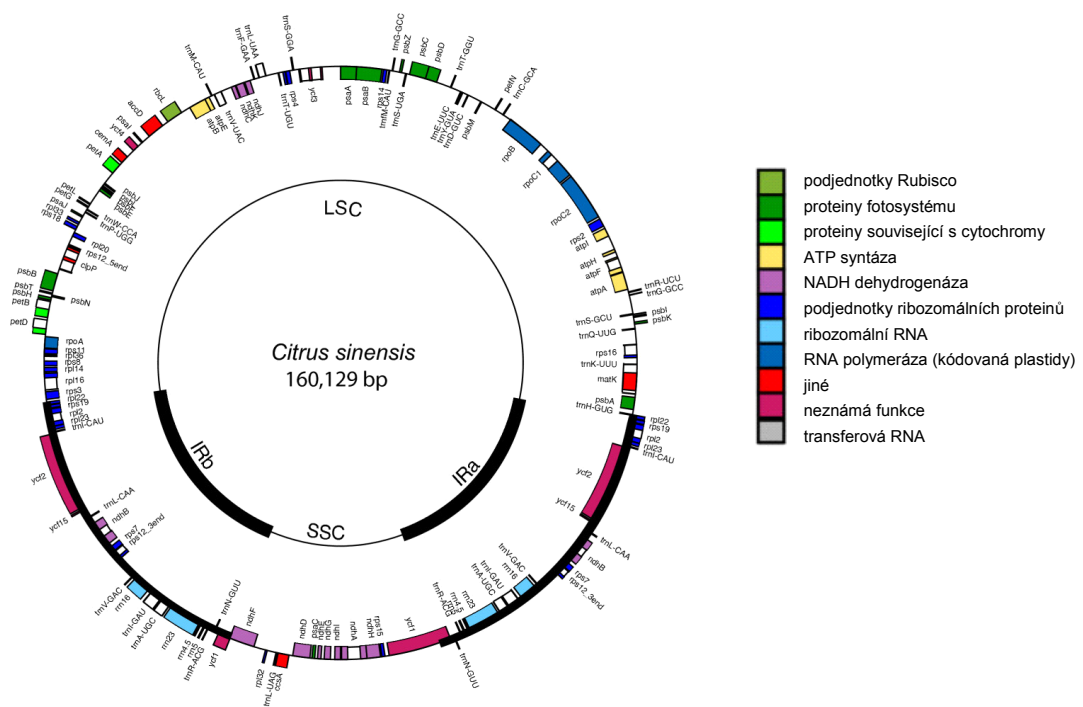
Mezi nejpopulárnější molekulárně-fylogenetické cíle v rámci rostlinného genomu patří chloroplastová DNA a také jaderná ribozomální DNA (nrDNA), především úseky ITS (internal transcribed spacer). V poslední době se také zvyšuje počet studií využívajících k fylogenetické analýze jaderné geny, což úzce souvisí s dostupností částečné nebo úplné genomové sekvence pro narůstající počet rostlinných organismů.

3.2.1 Chloroplastová DNA ve fylogenetické analýze

Rostliny obsahují kromě jaderné ještě mimojadernou dědičnou informaci uloženou v chloroplastovém genomu, který je ve srovnání s jaderným genomem poměrně malý (mezi 37 a 200 kbp). Díky jeho velikosti a malému obsahu repetitivních sekvencí je dnes poměrně snadné získat sekvenci jeho DNA, což dokládá také záznam kompletních sekvencí chloroplastového genomu u 171 zelených rostlin (<http://www.ncbi.nlm.nih.gov/genomes/>, 1.1.2011). Chloroplastový genom má cirkulární strukturu, i když byly pozorovány i jeho linearizované formy, u nichž se ale předpokládá, že jsou pouhými fragmenty kompletní kruhové molekuly (Brown 1999). Počet genů kódovaných chloroplastovým genomem se pohybuje mezi 60 a 200, přičemž jde z větší části o geny pro ribozomální a transferovou RNA, ribozomální proteiny a také proteiny uplatňující se při fotosyntéze (Martin *et al.* 2002). Tyto geny jsou uspořádány ve dvou tzv. jednokopiových oblastech oddělených oblastí tzv. obrácených repetit (IR), která se vyznačuje vysokou konzervovaností (Obr. 3.2, znázorněno na příkladu chloroplastového genomu pomerančovníku).

Mezi výhody, které z těchto genů činí oblíbený nástroj rostlinných fylogenetických studií patří mimo jiné dostupnost univerzálních primerů umožňujících jednoduchou amplifikaci a sekvenování požadovaných úseků DNA. Dalším pozitivem je relativně

vyšší konzervovanost (nižší substituční rychlost) chloroplastových genů ve srovnání s geny jadernými, což je dobře využitelné pro řešení otázek fylogeneze vzdálenějších resp. starších rostlinných druhů (Graham & Olmstead 2000). Na druhé straně ovšem oblasti intronů či tzv. genových mezerníků (intergenic spacers), které jsou také součástí chloroplastové DNA, nejsou pod tak vysokým selekčním tlakem jako geny, vyvíjejí se rychleji a jsou použitelné i pro rozřešení fylogenetických vztahů na nižších taxonomických úrovních mezi vzájemně příbuznějšími druhy (Gielly & Taberlet 1994). V neposlední řadě poskytuje chloroplastový genom proti jadernému výhodu mnohem nižšího obsahu pseudogenů, repetitivních sekvencí a cizorodé DNA.



Obr. 3.2: Genová mapa chloroplastového genomu pomerančovníku (*Citrus sinensis*). Ve vnitřním kruhu jsou znázorněny oblasti jednokopiových sekvencí (velká oblast – LSC, large single copy, a malá oblast – SSC, small single copy) oddělené oblastmi obrácených repetitiv (inverted repeat; IRa a IRb). Bausher et al. 2006, převzato s úpravou.

Všechny tyto vlastnosti se podepsaly pod obrovské množství prací, které využívají chloroplastové DNA pro rekonstrukci fylogeneze rostlinných druhů. Jedna z prvních

velmi komplexních rostlinných fylogenetických studií, která položila základ dnešní představě o fylogenezi krytosemenných rostlin byla provedena právě s použitím sekvence chloroplastového genu *rbcL* (ribulóza-bisfosfát karboxyláza) na souboru 500 taxonů (Chase *et al.* 1993). Tento obrovský soubor taxonů byl později ještě rozšířen na 2538 druhů reprezentujících veškeré fotosyntetizující organizmy (Källersjö *et al.* 1999) a později dokonce až na 13 000 rostlinných taxonů ve studii autorů Smith *et al.* (2009).

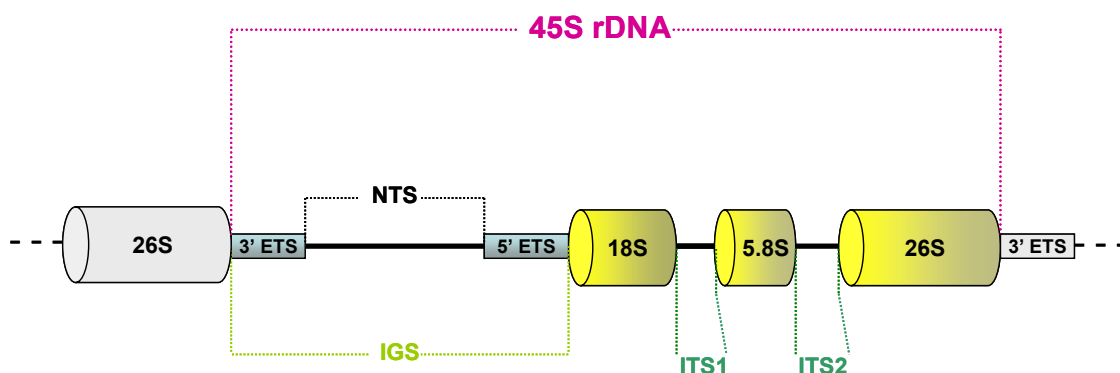
Oblast, kde použití chloroplastové DNA není natolik výhodné, je analýza speciace hybridních druhů, protože chloroplastová DNA se vyznačuje uniparentální dědičností (Birky 1995). V případě většiny krytosemenných rostlin je genom chloroplastu děděn v mateřské linii a dokumentuje tak evoluci hybridních druhů pouze jednostraně.

3.2.2 Jaderná ribozomální DNA ve fylogenetické analýze

Geny kódující ribozomální RNA se v jaderném genomu rostlin, podobně jako u ostatních eukaryotických organizmů, nacházejí ve stovkách až v tisících kopiích, které jsou organizovány jako tandemově se opakující jednotky. Klastry s těmito tandemovými jednotkami jsou v genomu většinou lokalizovány na fyzicky oddělených místech, jedním je lokus obsahující geny pro 5S rRNA, druhým je 45S rDNA lokus, který obsahuje geny pro 18S, 5.8S a 26S rRNA (Obr. 3.3). Tyto geny (18S-5.8S-26S rRNA) jsou v průběhu transkripce DNA přepisovány společně jako jedna jednotka, včetně dvou vnitřních mezerníků, kterými jsou odděleny (ITS1 a ITS2; Internal Transcribed Spacer; Hillis & Dixon 1991) a až následně jsou sestřiženy na samostatné úseky RNA. Jednotlivé kopie 18S-5.8S-26S rRNA genů jsou vzájemně odděleny mezigenovým mezerníkem (IGS; Intergenic Spacer), z něž jsou do mRNA přepisovány pouze jeho konce tvořené tzv. externími přepisovanými mezerníky (ETS; External Transcribed Spacer; Rogers & Bendich 1987).

Právě oblasti ITS1 a ITS2 jsou využívány jako jeden z nejčastějších fylogenetických markerů v rekonstrukci fylogeneze rostlin. Mezi hlavní výhody ITS patří jejich univerzálnost, jsou to totiž oblasti s poměrně vysokou mírou genetické variability, obklopené naopak poměrně konzervovanou strukturou rRNA genů. To umožnilo navržení univerzálních primerů pro jednoduchou amplifikaci ITS mezerníků z rostlinné DNA (White *et al.* 1990). Zároveň průměrná délka kompletního ITS1-5.8S-ITS2 úseku DNA se u většiny rostlin pohybuje pod 700 bp (Feliner & Rosselló

2007), což je ideální délka pro snadné sekvenování této oblasti bez nutnosti použití specifických vnitřních primerů.



Obr. 3.3: Schematické znázornění obecné struktury 45S rDNA lokusu rostlin. Geny pro 18S, 5.8S a 26S rRNA jsou odděleny vnitřními přepisovanými mezerníky (ITS1 a ITS2). Mezi jednotlivými kopiemi 18S-5.8S-26S rDNA je IGS mezerník (intergenic spacer), který se skládá z netranskribované části (NTS- nontranscribed spacer) a přepisovaných 3' a 5' konců (ETS- external transcribed spacer).

Upraveno podle Poczai & Hyvönen 2010.

Bylo prokázáno, že variabilita v rámci ITS sekvencí (zejména ITS1) je dostatečně vysoká aby bylo možné uplatnit tuto oblast DNA pro fylogenetické studie na úrovni rodů až druhů (Baldwin *et al.* 1995). Na druhé straně, relativně nízká variabilita sousedních genů společně s vysokou konzervovaností sekundární struktury sekvence ITS2 zaručují vysokou úspěšnost v identifikaci potenciálních nefunkčních variant – pseudogenů (Bailey *et al.* 2003, Wolf *et al.* 2005), jejichž zahrnutí do fylogenetické analýzy může negativně ovlivnit přesnost výsledku (Mayol & Rosselló 2001).

Jaderná rDNA se také vyznačuje charakteristickým mechanismem evoluce, který bývá označován jako „evoluce v koncertu“ (concerted evolution). Tento mechanismus je zodpovědný za to, že i při velkém počtu kopií rDNA v jaderném genomu rostlin zůstává zachována vysoká uniformita jejich sekvence. Podstata evoluce v koncertu spočívá v homogenizaci genů rDNA, což znamená, že místo aby v průběhu času vznikala různá míra sekvenční variability u každé z mnoha kopií rDNA vlivem akumulace mutací, dochází k hromadění stejných mutací u všech kopií rDNA a ty se tak vyvíjejí jako celek (Zimmer *et al.* 1980, Elder & Turner 1995). I přes rozsáhlou dokumentaci tohoto evolučního mechanismu u rDNA, byly popsány také případy, kdy homogenizace není kompletní a u jednoho druhu tak může být přítomna více než jedna varianta ITS sekvence. Taková situace může nastat zejména

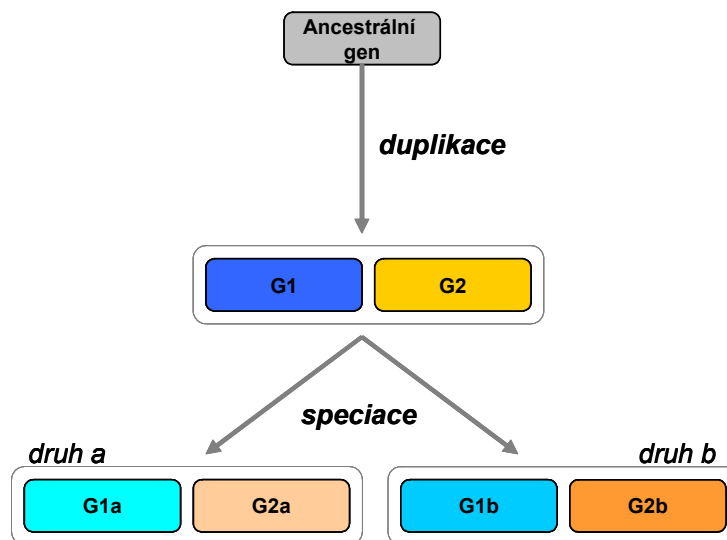
u hybridních druhů, kde v důsledku hybridizace či polyploidizace někdy zůstanou zachovány rodičovské formy rDNA a dále se vyvíjejí nezávisle (např. u knotovky, Popp & Oxelman 2001, nebo pšenice, Zhang *et al.* 2002). V některých případech rodičovské formy rDNA rekombinují za vzniku chimerických ITS sekvencí (Barkman & Simpson 2002). Podobné odchylky od kompletní homogenizace rDNA lokusu mohou být užitečné pro identifikaci rodičovských linií u hybridních druhů (Wendel *et al.* 1995), zároveň však také mohou vnést do fylogenetické analýzy nepřesnost. Aby bylo možné tomu předejít, je vhodné použít při analýze ITS sekvencí vždy raději postup klonování a sekvenování několika klonů pro zachycení všech přítomných variant ITS, nežli sekvenovat ITS oblast přímo z genomové DNA (Álvarez & Wendel 2003).

3.2.3 Geny s nízkým počtem kopií ve fylogenetické analýze

Díky obrovskému množství genů a vysoké evoluční rychlosti zejména v intronech představuje jaderný genom zdaleka nejbohatší zdroj fylogeneticky informativních dat, které mají potenciál vyřešit otázky evoluční příbuznosti i tam, kde rozlišovací schopnost chloroplastové či ribozomální DNA není dostatečná. Jaderné geny vyskytující se v genomu v nízkém počtu kopií, ideálně geny jednokopiové, se stávají čím dál využívanější pro fylogenetickou analýzu na všech taxonomických úrovních (Sang 2002). Při jejich použití odpadá problém uniparentální dědičnosti, který hraje roli u organelové DNA a také není nutné brát v potaz mechanismus evoluce v koncertu, jako je tomu u ribozomální DNA. Navíc se nabízí možnost využít více genů a tedy více vzájemně nezávislých lokusů, což zvyšuje přesnost rekonstruované fylogeneze (Heckman *et al.* 2007).

I při použití jaderných genů je však nutné dbát na několik faktorů, které mohou negativně ovlivnit výsledek analýzy. Jedním z nich je základní předpoklad pro fylogenetické studie, a to že vstupní data pro komparativní analýzu jsou homologní. To znamená, že v případě DNA srovnáváme takové sekvence, jejichž podobnost je důsledkem jejich společného původu (Doyle & Davis 1998). S tím také souvisí problém rozlišení ortologního a paralogního původu genů. V obou případech mají tyto geny společného předka a jsou homologní, ovšem geny ortologní vznikly speciací, kdežto paralogní geny jsou důsledkem duplikace ancestrálního genu (Obr. 3.4, Fitch 1970). Je tedy zřejmé, že nelze spojovat paralogní a ortologní geny v jedné

analýze, protože dokumentují dvě rozdílné evoluční události (Sanderson & Shaffer 2002). Odlišení těchto dvou forem homologních genů není jednoduché, paralogní geny však lze často odhalit pomocí sledování rozdílů v molekulární struktuře, např. v nukleotidovém složení sekvencí (zastoupení jednotlivých bází) nebo ve struktuře kodonů (Cotton 2005).



Obr. 3.4: Znárodnění rozdílů ve vzniku ortologních a paralogních genů. Všechny G geny jsou vzájemně homologní (sdílí společného předka). Geny G1a a G2b jsou vzájemně paralogní (vznikly v důsledku duplikace ancestrálního genu), geny G1a a G1b, resp. G2a a G2b jsou vzájemně ortologní (vznikly v důsledku speciace). Upraveno podle Koonin 2005.

Z praktického hlediska by měly geny, či jejich části zvolené pro fylogenetickou analýzu zohledňovat především její účel. Pokud tedy chceme zkoumat evoluční vztahy na úrovni jednotlivých druhů, je vhodné zahrnout do analýzy také rychle se vyvíjející introny, a naopak pro zkoumání evoluce na vyšších taxonomických úrovních je vhodnější intronové sekvence eliminovat (Sang 2002). Ve snaze vyhnout se paralogním genům je také žádoucí vybírat geny jednokopiové, protože u genů o více kopiích pocházejících z početných genových rodin je vyšší pravděpodobnost, že prošly několika cykly genových duplikací a delecí, což znesnadňuje rozlišení jejich ortologního či paralogního původu (Clegg et al. 1997). Přestože je v jaderném genomu rostlin k dispozici enormní množství fylogeneticky zajímavých cílových oblastí, nejsou dostupné univerzální primery umožňující snadné získání sekvence

DNA z mnoha druhů. Potenciál jaderných genů zodpovědět široké spektrum evolučních otázek v rámci rostlinné říše tak zůstává limitován zejména finanční dostupností vývoje genových fylogenetických markerů.

3.3 Analýza molekulárně-fylogenetických dat

Získaná molekulární data lze převést do podoby fylogenetického stromu odpovídajícího evolučním vztahům studovaných taxonů pomocí několika principiálně odlišných přístupů. První přístup, označovaný jako distanční metody, používá k rekonstrukci kladogeneze matici genetických vzdáleností pro každé dvě sekvence DNA (případně jiná molekulární data). Pomocí výpočetního algoritmu je poté sestaven jeden fylogenetický strom, který nejlépe vystihuje informaci uloženou v distanční matici. Jiným přístupem je hledání nejlepšího fylogenetického stromu z množiny mnoha možných stromů podle zvoleného kritéria optimality. Tento přístup používají např. metody maximální parsimonie a maximální pravděpodobnosti (Vandamme 2003). Mimo tyto dva základní přístupy existuje ještě zvláštní skupina metod pro sestavení fylogenetického stromu, které jsou také metody pravděpodobnostní, ale vycházejí z Bayesova teorému a bývají označovány jako metody bayesovské (Li 1996, Mau 1996, Rannala & Yang 1996).

Žádná z metod analýzy fylogenetických dat není nejsprávnější (Rosenberg & Kumar 2001) a nelze dopředu odhadnout, zda některá poskytne přesnější rekonstrukci kladogeneze na daném souboru, proto se v praxi používá kombinace několika alternativních přístupů. Pokud všechny vyzkoušené metody poskytují stejný výsledek, je velká pravděpodobnost, že odráží skutečné evoluční vztahy.

3.3.1 Distanční metody

Distanční metody pracují s předpokladem, že dva organizmy, které sdílí nedávného společného předka, si budou podobnější, než organizmy, jejichž společný předek je starší. Proto prvním krokem je převedení nukleotidových sekvencí na distanční matici vyjadřující podobnost organismů. Práce s distanční maticí, kde odpadá nutnost zabývat se každou nukleotidovou pozicí v sekvenci DNA zvlášť, je méně náročná pro počítačové zpracování a umožňuje tak poměrně rychle dojít k výsledku i pro velké soubory srovnávaných taxonů, což je hlavní výhodou distančních metod

(Van de Peer 2003). Distanční matice je do formy fylogenetického stromu převedena pomocí výpočetních algoritmů, mezi nimiž nejznámější jsou algoritmy UPGMA (Unweighted Pair Group Method with Arithmetic mean, Michener & Sokal 1957) a Neighbor Joining (NJ, Saitou & Nei 1987). Metoda UPGMA sestavuje výsledný strom tak, že vždy identifikuje nejkratší vzdálenost v matici, spojí příslušné taxony a v dalším kroku je považuje za jednu jednotku, vypočítá novou distanční matici a celý postup se opakuje v potřebném počtu kroků. Výsledkem UPGMA je fenogram, který je obrazem podobnosti organismů, ale jeho topologie se nemusí vždy shodovat s průběhem kladogeneze. Neighbor Joining je distanční metoda, která poskytuje fylogramy, tedy skutečný obraz kladogeneze, a to díky tomu, že umožňuje počítat s nestejnou rychlostí evoluce v různých vývojových větvích stromu, což často mnohem lépe vystihuje skutečnost (Barton *et al.* 2007).

3.3.2 Metody založené na kritériu optimality

Narozdíl od distančních metod, metody založené na kritériu optimality pracují přímo s původní informací uloženou v nukleotidovém složení analyzovaných sekvencí DNA a odráží tedy navíc informaci o tom, do jaké míry ovlivňují jednotlivé nukleotidové pozice délku větví vznikajícího fylogenetického stromu (Page & Holmes 1998). Díky tomu je také možné vnést do analýzy předem stanovený evoluční model (model procesu nukleotidových substitucí). V prvním kroku analýzy jsou u těchto metod vygenerovány všechny možné stromy vyjadřující vztahy mezi analyzovanými druhy podle zvoleného evolučního modelu. Další postup se liší podle zvoleného kritéria optimality.

Metody maximální parsimonie (úspornosti) vycházejí z pravidla tzv. Occamovy břitvy, které říká, že při absenci jiných rozhodovacích faktorů je vždy nejlepším řešením problému to nejjednodušší (Swofford & Sullivan 2003). Z množiny generovaných stromů je tedy podle tohoto kritéria vybírán jako nejlepší ten, k jehož vzniku bylo zapotřebí nejméně kroků - evolučních změn (Camin & Sokal 1965).

U metod, kde je zvoleným kritériem maximální pravděpodobnost (tzv. likelihood metody) se nejlépe vyhovující strom vybírá tak, aby s co největší pravděpodobností odpovídal vstupním datům. Jinými slovy, vybírá se takový strom, jehož topologie, délka větví i parametry evolučního modelu (např. poměr tranzicí a transverzí, rozdíly v evoluční rychlosti na jednotlivých pozicích apod.) je nejpravděpodobnější

vzhledem k pozorovaným vstupním sekvencím DNA (Von Haeseler & Strimmer 2003). Vzhledem k tomu, že obě výše popsané metody vybírají výsledný fylogenetický strom z často obrovského souboru možných stromů, jsou velmi náročné na počítačové zpracování a často je nutné přistoupit ke zjednodušenému, tzv. heuristickému vyhledávání výsledného stromu, kdy se některé skupiny stromů dají předem vyloučit z analýzy.

3.3.3 Bayesovské metody

Bayesovská statistika je podobná metodě maximální pravděpodobnosti (ML; maximum likelihood), ale rozšiřuje ji navíc o tzv. *a priori* pravděpodobnost, neboli bere do úvahy, že většinou máme k dispozici nějakou předchozí znalost vstupních dat. Aplikováno na fylogenetickou analýzu, stejně jako ML metody počítá bayesovská analýza s jasně stanoveným evolučním modelem, ale při výpočtu největší pravděpodobnosti výsledného stromu (zde nazývané *a posteriori* pravděpodobnost) zohledňuje *a priori* pravděpodobnost toho jaké stromy mohou vzniknout vzhledem ke vstupním datům (Huelsenbeck *et al.* 2002, Archibald *et al.* 2003). Díky tomu, že bayesovské metody používají jako výpočetní algoritmus tzv. Markov Chain Monte Carlo simulaci (jejíž podrobnější popis přesahuje zaměření této práce; Metropolis *et al.* 1953), jsou většinou ve srovnání s metodou maximální pravděpodobnosti méně počítačově náročné.

3.4 Využití molekulárních dat k datování evolučních událostí

Rekonstrukce fylogeneze pomocí molekulárně biologických dat poskytuje jednak informaci o příbuznosti studovaných druhů a postupnosti odštěpování jednotlivých vývojových linií, ale zároveň lze její pomocí také získat informaci o časovém sledu těchto dějů. K tomu se využívá konceptu tzv. molekulárních hodin, který je založen na předpokladu, že mutační (substituční) rychlost je v průběhu fylogeneze u všech organismů přibližně konstantní (Zuckerkanndl & Pauling 1962, 1965). Potom tedy množství neutrálních mutací fixovaných v genomech dvou sesterských organismů je úměrné době, před kterou se oddělily od společného předka, čili od okamžiku divergence. Pokud tedy známe substituční rychlost specifickou pro daný druh,

případně gen (tedy průměrný počet mutací fixovaných u daného druhu/genu za časovou jednotku), lze zpětně datovat okamžik divergence podle:

$$T = D / 2k$$

kde T je čas divergence druhů reprezentovaných těmito sekvencemi (v letech), D je počet substitucí mezi dvěma homologními sekvencemi připadající na jednu nukleotidovou pozici a k je absolutní rychlost nukleotidové substituce na pozici za rok (Gaut 1998). V případě, že substituční rychlost není známá, lze molekulární hodiny kalibrovat pomocí jiného druhu, jehož stáří bylo určeno z fosilního paleontologického nálezu. Pokud tedy známe množství fixovaných mutací mezi dvěma studovanými druhy a množství mutací kterými se tyto dva druhy liší od třetího druhu s fosilním záznamem, je možné stanovit okamžik divergence studovaných dvou druhů.

Jak se ovšem později ukázalo, reálná situace v genomech živých organismů se od konceptu striktních molekulárních hodin poněkud liší. Ve skutečnosti totiž substituční rychlost u různých druhů není stejná a také v rámci genomu se může lišit substituční rychlost jednotlivých oblastí (Fitch & Margoliash 1967, Gaut *et al.* 1992). Toto poznání postupně vedlo ke snaze zahrnout do stanovení evolučního stáří také heterogenitu substituční rychlosti mezi jednotlivými vývojovými liniemi. Vznikl tak koncept tzv. relaxovaných molekulárních hodin, který pomocí statistických modelů využívajících např. maximální pravděpodobnosti či Bayesovského modelování umožňuje počítat s nekonstantní substituční rychlostí a zpřesňuje tak odhady evolučního stáří organismů (Sanderson 1997, Thorne *et al.* 1998, Drummond *et al.* 2006).

3.5 Fylogeneze čeledi *Musaceae*

Molekulárně-fylogenetických studií zaměřených na evoluční vztahy v čeledi *Musaceae* není mnoho. První série takto zaměřených studií využívala především různé typy molekulárních markerů, např. RFLP markery (Jarret *et al.* 1992, Gawel *et al.* 1992), nebo RAPD a modifikované SSR markery (Venkatachalam *et al.* 2008).

Teprve v poslední době se objevují práce využívající pro popis fylogenetických vztahů v čeledi *Musaceae* sekvenování DNA, především chloroplastové geny a jejich

introny (Swangpol *et al.* 2007), ale také chloroplastovou DNA v kombinaci s jaderným ITS lokusem (Liu *et al.* 2010, Li *et al.* 2010). V posledních dvou pracích je sice zahrnut poměrně velký soubor taxonů, chybí však podrobné sekvenování jednotlivých klonovaných ITS variant, které by odhalilo případnou heterogenitu ITS lokusu v těchto taxonech a tím podalo přesnější informaci o směru, kterým se ubírala jejich evoluce. Žádná studie dosud nepopisuje fylogenezi *Musaceae* na podkladu většího souboru jaderných genů.

Čeď *Musaceae* byla zahrnuta také do několika molekulárně-fylogenetických prací jako součást jednoděložných rostlin (Bremer 2000, Janssen & Bremer 2004, Anderson & Jansen 2009), nebo jako jedna z čeledí řádu Zingiberales (Kress *et al.* 2001, Kress & Specht 2006). Z těchto prací také plynou první odhady evolučního stáří čeledi *Musaceae*, které bylo stanoveno mezi 61 a 110 miliony let. Nicméně, dosud žádná studie neposkytla informace o evolučním stáří jednotlivých vývojových větví uvnitř čeledi *Musaceae* na nižších taxonomických úrovních.

POUŽITÁ LITERATURA

- Agarwal M, Shrivastava N, Padh H: Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Reports* 27: 617-631, 2008.
- Akbari M, Wenzl P, Vanessa C, Carling J, Xia L, Yang S, Uszynski G, Mohler V, Lehmensiek A, Kuchel H, Hayden MJ, Howes N, Sharp P, Rathmell B, Vaughan P, Huttner E, Kilian A: Diversity Arrays Technology (DArT) for high-throughput profiling of the hexaploid wheat genome. *Theoretical and Applied Genetics* 113: 1409-1420, 2006.
- Álvarez I, Wendel JF: Ribosomal ITS sequence and plant phylogenetic inference. *Molecular Phylogenetics and Evolution* 29: 417-434, 2003.
- Anderson CL, Janssen T: Monocots. V: Hedges SB, Kumar S (eds.): *The timetree of life*. New York: Oxford University Press, 203-212, 2009.
- Archibald JK, Mort ME, Crawford DJ: Bayesian inference of phylogeny: a non-technical primer. *Taxon* 52: 187-191, 2003.
- Argent GCG: The wild bananas of Papua New Guinea. *Notes from the Royal Botanical Garden Edinburgh* 35: 77-114, 1976.
- Avise JC: *Molecular markers, natural history and evolution*. Second edition. Sunderland: Sinauer Associates, 2004.
- Bailey CD, Carr TG, Harris SA, Hughes CE: Characterization of angiosperm nrDNA polymorphism, paralogy, and pseudogenes. *Molecular Phylogenetics and Evolution* 29: 435-455, 2003.
- Balasubramanian S, Harrison P, Hegyi H, Bertone P, Luscombe N, Echols N, McGarvey P, Zhang Z, Gerstein M: SNPs on human chromosomes 21 and 22 - analysis in terms of protein features and pseudogenes. *Pharmacogenomics* 3: 393-402, 2002.
- Baldwin B, Sanderson MJ, Porter JM, Wojciechowski MF, Campbell CS, Donoghue MJ: The ITS region of nuclear ribosomal DNA: A valuable source of evidence on angiosperm phylogeny. *Annals of the Missouri Botanical Garden* 82: 247-277, 1995.
- Barkman TJ, Simpson BB: Hybrid origin and parentage of *Dendrochilum acuiiferum* (*Orchidaceae*) inferred in a phylogenetic context using nuclear and plastid DNA sequence data. *Systematic Botany* 27: 209-220, 2002.
- Barton NH, Briggs DEG, Eisen JA, Goldstein DB, Patel NH: *Evolution*. Cold Spring Harbor: Cold Spring Harbor Laboratory Press, 2007.

- Batley J, Barker G, O'Sullivan H, Edwards KJ, Edwards D: Mining for single nucleotide polymorphisms and insertions/deletions in maize expressed sequence tag data. *Plant Physiology* 132: 84-91, 2003.
- Bausher MG, Singh ND, Lee SB, Jansen RK, Daniell H: The complete chloroplast genome sequence of *Citrus sinensis* (L.) Osbeck var 'Ridge Pineapple': organization and phylogenetic relationships to other angiosperms. *BMC Plant Biology* 6: 21, 2006.
- Beyermann B, Nürnberg P, Weihe A, Meixner M, Eppelen JT, Börner T: Fingerprinting plant genomes with oligonucleotide probes specific for simple repetitive DNA sequences. *Theoretical and Applied Genetics* 83: 691-694, 1992.
- Birky Jr. CW: Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. *Proceedings of the National Academy of Sciences of the USA* 92: 11331-11338, 1995.
- Birmeta G, Nybom H, Bekele E: RAPD analysis of genetic diversity among clones of the Ethiopian crop plant *Ensete ventricosum*. *Euphytica* 124: 315-325, 2002.
- Bremer K: Early Cretaceous lineages of monocot flowering plants. *Proceedings of the National Academy of Sciences of the USA* 97: 4707-4711, 2000.
- Botstein D, White RL, Skolnick M, Davis RW: Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* 32: 314-331, 1980.
- Brown TA: *Genomes*. New York: John Wiley & Sons, 1999.
- Brown WM, George Jr M, Wilson AC: Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences of the USA* 76: 1967-1971, 1979.
- Burr B: Some concepts and new methods for molecular mapping in plants. V: Phillips RL, Vasil IK (eds): *DNA-based markers in plants*. Dordrecht: Kluwer Academic Publishers; 1-7, 1994.
- Camin JH, Sokal RR: A method for deducing branching sequences in phylogeny. *Evolution* 19: 311-326, 1965.
- Carling MD, Brumfield RT: Gene sampling strategies for multi-locus population estimates of genetic diversity (θ). *Plos One* 1: e160, 2007.
- Clegg MT, Cummings MP, Durbin ML: The evolution of plant nuclear genes. *Proceedings of the National Academy of Sciences of the USA* 94: 7791-7798, 1997.
- Cotton JA: Analytical methods for detecting paralogy in molecular datasets. *Methods in Enzymology* 395: 700-724, 2005.

- Creste S, Neto AT, Silva SD, Figueira A: Genetic characterization of banana cultivars (*Musa* spp.) from Brazil using microsatellite markers. *Euphytica* 132: 259-268, 2003.
- Daniells J, Jenny C, Karamura D, Tomekpe K: *Musalogue: A catalogue of Musa germplasm. Diversity in the genus Musa.* (Arnaud E, Sharrock S, compil.). Montpellier: INIBAP, 2001.
- De Langhe E, Pillay M, Tenkouano A, Swennen R: Integrating morphological and molecular taxonomy in *Musa*: the African plantains (*Musa* spp. AAB group). *Plant Systematics and Evolution* 255: 225-236, 2005.
- Després L, Gielly L, Redoutet B, Taberlet P: Using AFLP to resolve phylogenetic relationships in a morphologically diversified plant species complex when nuclear and chloroplast sequences fail to reveal variability. *Molecular Phylogenetics and Evolution* 27: 185-196, 2003.
- Doyle JJ, Davis JJ: Homology in molecular phylogenetics: A parsimony perspective. V: Soltis DE, Soltis PS, Doyle JJ (eds.): *Molecular systematics of plants 2.* Norwell: Kluwer Academic Publishers, 101-131, 1998.
- Drummond A, Ho S, Phillips M, Rambaut A: Relaxed phylogenetics and dating with confidence. *PLoS Biology* 4: e88, 2006.
- Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, deWinter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma C, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D, Zhao P, Zhong F, Korlach J, Turner S: Real-time DNA sequencing from single polymerase molecules. *Science* 323: 133-138, 2009.
- Elder Jr. JF, Turner BJ: Concerted evolution of repetitive DNA sequences in eukaryotes. *The Quarterly Review of Biology* 70: 291-320, 1995.
- Ellsworth DL, Rittenhouse KD, Honeycutt RL: Artifactual variation in randomly amplified polymorphic DNA banding patterns. *BioTechniques* 14: 214-217, 1993.
- Feliner GN, Rosselló JA: Better the devil we know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants. *Molecular Phylogenetics and Evolution* 44: 911-919, 2007.
- FAOstat 2009: <http://faostat.fao.org/>
- Fitch WM, Margoliash E: A method for estimating the number of invariant amino acid coding positions in a gene using cytochrome *c* as a model case. *Biochemical Genetics* 1: 65-71, 1967.

- Fitch WM: Distinguishing homologous from analogous proteins. *Systematic Zoology* 19: 99-113, 1970.
- Flegl J: *Evoluční biologie*. Praha: Academia, 2005.
- Ford-Lloyd BV: Genotyping in plant genetic resources. V: Henry RJ (ed.): *Plant genotyping: the DNA fingerprinting of plants*. Walingford: CABI Publishing; 59-82, 2001.
- Gaut BS, Muse SV, Clark WD, Clegg MT: Relative rates of nucleotide substitution at the *rbcL* locus of monocotyledonous plants. *Journal of Molecular Evolution* 35: 292-303, 1992.
- Gaut BS: Molecular clocks and nucleotide substitution rates in higher plants. V: Hecht MK, MacIntyre RJ, Clegg MT (eds): *Evolutionary biology*. Vol. 30. New York: Plenum Press; 93-120, 1998.
- Gawel NJ, Jarret RL, Whittemore AP: Restriction fragment length polymorphism (RFLP)-based phylogenetic analysis of *Musa*. *Theoretical and Applied Genetics* 84: 286-290, 1992.
- Ge XJ, Liu MH, Wang WK, Schaal BA, Chiang TY: Population structure of wild bananas, *Musa balbisiana*, in China determined by SSR fingerprinting and cpDNA PCR-RFLP. *Molecular Ecology* 14: 933-944, 2005.
- Gielly L, Taberlet P: The use of chloroplast DNA to resolve plant phylogenies: noncoding versus *rbcL* sequences. *Molecular Biology and Evolution* 11: 769-777, 1994.
- Graham SW, Olmstead RG: Utility of 17 chloroplast genes for inferring the phylogeny of the basal angiosperms. *American Journal of Botany* 87: 1712-1730, 2000.
- Grapin A, Noyer JL, Carreel F, Dambier D, Baurens FC, Lanaud C, Lagoda PJJ: Diploid *Musa acuminata* genetic diversity assayed with sequence-tagged microsatellite sites. *Electrophoresis* 19: 1374-1380, 1998.
- Gupta PK, Varshney RK: The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. *Euphytica* 113: 163-185, 2000.
- Gupta PK, Roy JK, Prasad M: Single nucleotide polymorphisms: A new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. *Current Science* 80: 524-535, 2001.
- Harris TD, Buzby PR, Babcock H, Beer E, Bowers J, Braslavsky I, Causey M, Colonell J, Dimeo J, Efcavitch JW, Giladi E, Gill J, Healy J, Jarosz M, Lapen D, Moulton K, Quake SR, Steinmann K, Thayer E, Tyurina A, Ward R, Weiss H, Xie Z: Single-molecule DNA sequencing of a viral genome. *Science* 320: 106-109, 2008.

- Harrison RG: Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. *Trends in Ecology and Evolution* 4: 6-11, 1989.
- Hartl DL: *Essential genetics: A genomics perspective*. 5th edition. Sudbury: Jones and Bartlett Publishers, 2011.
- Häkkinen M, Sharrock S: Diversity in the genus *Musa* – Focus on *Rhodochlamys*. V: INIBAP annual report 2001. Montpellier: INIBAP; 16-23, 2002.
- Heckman KL, Marian CL, Rasoloarison R, Yoder AD: Multiple nuclear loci reveal patterns of incomplete lineage sorting and complex species history within western mouse lemurs (*Microcebus*). *Molecular Phylogenetics and Evolution* 43: 353-367, 2007.
- Hillis DM, Dixon MT: Ribosomal DNA: molecular evolution and phylogenetic inference. *The Quarterly Review of Biology* 66: 411-453, 1991.
- Huelsenbeck JP, Larget B, Miller RE, Ronquist F: Potential application and pitfalls of Bayesian inference of phylogeny. *Systematic Biology* 51: 673-688, 2002.
- <http://bananas.bioversityinternational.org/>
- <http://www.bioversityinternational.org>
- <http://www.ncbi.nlm.nih.gov/genomes/>
- Hurme P, Savolainen O: Comparison of homology and linkage of random amplified polymorphic DNA (RAPD) markers between individual trees of Scots pine (*Pinus sylvestris* L.). *Molecular Ecology* 8: 15-22, 1999.
- Hyman IT, Ho SYW, Jermini LS: Molecular phylogeny of Australian Helicarionidae, Euconulidae and related groups (Gastropoda: Pulmonata: Stylommatophora) based on mitochondrial DNA. *Molecular Phylogenetics and Evolution* 45: 792-812, 2007.
- Chandler S: The nutritional value of bananas. V: Goven S (ed.): *Bananas and Plantains*. London: Chapman and Hall; 468-480, 1995.
- Chase MW, Soltis DE, Olmstead RG, Morgan D, Les DH, Mishler BD, Duvall MR, Price RA, Hills HG, Qiu YL, Kron KA, Rettig JH, Conti E, Palmer JD, Manhart JR, Sytsma KJ, Michaels HJ, Kress WJ, Karol KG, Clark WD, Hedren M, Gaut BS, Jansen RK, Kim KJ, Wimpee CF, Smith JF, Furnier GR, Strauss SH, Xiang QY, Plunkett GM, Soltis PS, Swensen SM, Williams SE, Gadek PA, Quinn CJ, Eguiarte LE, Golenberg E, Learn Jr GH, Graham SW, Barrett SCH, Dayanandan S, Albert VA: Phylogenetics of seed plants: An analysis of nucleotide sequences from the plastid gene *rbcL*. *Annals of the Missouri Botanical Garden* 80: 528-580, 1993.

- Cheesman EE: Classification of the bananas II. The genus *Musa* L. Kew Bulletin 2: 106-117, 1947.
- Ching ADA, Caldwell KS, Jung M, Dolan M, Smith OS, Tingey S, Morgante M, Rafalski A: SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genetics 3: 19, 2002.
- INIBAP: Global conservation strategy for *Musa* (Banana and Plantain). A consultative document prepared by INIBAP with the collaboration of numerous partners in the *Musa* research-and-development community. Montpellier: INIBAP, 2006.
- Jaccoud D, Peng K, Feinstein D, Kilian A: Diversity Arrays: a solid state technology for sequence information independent genotyping. Nucleic Acids Research 29: 4, e25, 2001.
- Janssen T, Bremer K: The age of major monocot groups inferred from 800+ *rbcL* sequences. Botanical Journal of the Linnean Society 146: 385-398, 2004.
- Jarne P, Lagoda PJJ: Microsatellites, from molecules to populations and back. Trends in Genetics 11: 424-429, 1996.
- Jarret RL, Litz RE: Enzyme polymorphism in *Musa acuminata* Colla. Journal of Heredity 77: 183-186, 1986.
- Jarret RL, Gawel NJ, Whittemore A, Sharrock S: RFLP-based phylogeny of *Musa* species in Papua New Guinea. Theoretical and Applied Genetics 84: 579-584, 1992.
- Jeger MJ, Eden-Green S, Thresh JM, Johanson A, Waller JM, Brown AE: Banana diseases. V: Goven S (ed.): Bananas and Plantains. London: Chapman and Hall; 317-381, 1995.
- Karamura EB, Karamura DA: Banana morphology – part II: the aerial shoot. V: Goven S (ed.): Bananas and Plantains. London: Chapman and Hall; 191-205, 1995.
- Kardolus JP, Van Eck HJ, Van den Berg RG: The potential of AFLPs in biosystematics: a first application in *Solanum* taxonomy (*Solanaceae*). Plant Systematics and Evolution 210: 87-103, 1998.
- Källersjö M, Farris JS, Chase MW, Bremer B, Fay MF, Humpries CJ, Petersen G, Seberg O, Bremer K: Simultaneous parsimony jackknife analysis of 2538 *rbcL* DNA sequences reveals support for major clades of green plants, land plants, seed plants and flowering plants. Plant Systematics and Evolution 213: 259-287, 1999.
- Koonin EV: Orthologs, paralogs, and evolutionary genomics. Annual Review of Genetics 39: 309-338, 2005.

- Koopman WJM, Zevenbergen MJ, Van den Berg RG: Species relationships in *Lactuca* s.l. (Lactuceae, Asteraceae) inferred from AFLP fingerprints. *American Journal of Botany* 88: 1881-1887, 2001.
- Kopecký D, Bartoš J, Lukaszewski AJ, Baird JH, Černoch V, Kölliker R, Rognli OA, Blois H, Caig V, Lubberstedt T, Studer B, Shaw P, Doležel J, Kilian A: Development and mapping of DArT markers within the *Festuca* - *Lolium* complex. *BMC Genomics*, 10, 473, 2009.
- Kress WJ, Prince LM, Hahn WJ, Zimmer E: Unraveling the evolutionary radiation of the families of the Zingiberales using morphological and molecular evidence. *Systematic Biology* 50: 926-944, 2001.
- Kress WJ, Specht CD: The evolutionary and biogeographic origin and diversification of the tropical monocot order Zingiberales. *Aliso* 22, 619-630, 2006.
- Kumar LS: DNA markers in plant improvement: An overview. *Biotechnology Advances* 17: 143-182, 1999.
- Külheim C, Yeoh SH, Maintz J, Foley WJ, Moran GF: Comparative SNP diversity among four *Eucalyptus* species for genes from secondary metabolite biosynthetic pathways. *BMC Genomics* 10: 452, 2009.
- Lagoda PJJ, Noyer JL, Dambier D, Baurens FC, Grapin A, Lanaud C: Sequence tagged microsatellite site (STMS) markers in the *Musaceae*. *Molecular Ecology* 7: 657-666, 1998.
- Langridge P, Chalmers K: The principle: Identification and application of molecular markers. V: Lörz H, Wenzel G (eds): *Molecular marker systems in plant breeding and crop improvement*. Berlin: Springer; 3- 22, 2004.
- Li HW: The *Musaceae* of Yunnan. *Acta Phytotaxonomica Sinica* 16: 54-64, 1978.
- Li LF, Häkkinen M, Yuan YM, Hao G, Ge XJ: Molecular phylogeny and systematics of the banana family (*Musaceae*) inferred from multiple nuclear and chloroplast DNA fragments, with a special reference to the genus *Musa*. *Molecular Phylogenetics and Evolution* 57: 1-10, 2010.
- Li S: Phylogenetic tree construction using Markov chain Monte Carlo. *Doktorská disertační práce*. Columbus: Ohio State University, 1996.
- Liu AZ, Kress WJ, Li DZ: Phylogenetic analyses of the banana family (*Musaceae*) based on nuclear ribosomal (ITS) and chloroplast (*trnL-F*) evidence. *Taxon* 59: 20-28, 2010.
- Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP,

- Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM: Genome Sequencing in Open Microfabricated High Density Picoliter Reactors. *Nature* 437: 376-380, 2005.
- Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, Lins T, Leister D, Stoebe B, Hasegawa M, Penny D: Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proceedings of the National Academy of Sciences of the USA* 99: 12246-12251, 2002.
- Mau B: Bayesian phylogenetic inference via Markov chain Monte Carlo methods. *Doktorská disertační práce*. Madison: Wisconsin University, 1996.
- Mayol M, Rosselló JA: Why nuclear ribosomal DNA spacers (ITS) tell us different stories in *Quercus*. *Molecular Phylogenetics and Evolution* 19: 167-176, 2001.
- Metropolis N, Rosenbluth AW, Rosenbluth N, Teller AH, Teller E: Equations of state calculations by fast computing machines. *Journal of Chemical Physics* 21: 1087-1091, 1953.
- Michener CD, Sokal RR: A quantitative approach to a problem of classification. *Evolution* 11: 490-499, 1957.
- Miller DR, Rossman AY: Systematics, Biodiversity, and Agriculture. *BioScience* 45: 680-686, 1995.
- Mondini L, Noorani A, Pagnotta MA: Assessing plant genetic diversity by molecular tools. *Diversity* 1: 19-35, 2009.
- Mullis KB, Faloona F: Specific synthesis of DNA in vitro via polymerase chain reaction. *Methods in Enzymology* 155: 350-355, 1987.
- Muralidharan K, Wakeland EK: Concentration of primer and template qualitatively affects products in random-amplified polymorphic DNA PCR. *BioTechniques* 14: 362-364, 1993.
- Niklas KJ: *The evolutionary biology of plants*. Chicago: The University of Chicago, 1997.
- Noyer JL, Causse S, Tomekpe K, Bouet A, Baurens FC: A new image of plantain diversity assessed by SSR, AFLP and MSAP markers. *Genetica* 124: 61-69, 2005.
- Nybom H: Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Molecular Ecology* 13: 1143-1155, 2004.

- Page RDM, Holmes EC: Molecular evolution: A phylogenetic approach. Oxford: Blackwell Publishing, 1998.
- Paton A: Biodiversity informatics and the plant conservation baseline. Trends in Plant Science 14: 629-637, 2009.
- Pérez T, Albornoz J, Domínguez A: An evaluation of RAPD reproducibility and nature. Molecular Ecology 7: 1347-1357, 1998.
- Pillay M, Nwakanma, DC, Tenkouano A: Identification of RAPD markers linked to A and B genome sequences in *Musa L.* Genome 43: 763-767, 2000.
- Pillay M, Ogundiwin E, Nwakanma DC, Ude G, Tenkouano A: Analysis of genetic diversity and relationships in East African banana germplasm. Theoretical and Applied Genetics 102: 965-970, 2001.
- Poczai P, Hyvönen J: Nuclear ribosomal spacer regions in plant phylogenetics: problems and prospects. Molecular Biology Reports 37: 1897-1912, 2010.
- Popp M, Oxelman B: Inferring the history of the polyploid *Silene aegaea* (*Caryophyllacea*) using plastid and homoeologous nuclear DNA sequences. Molecular Phylogenetics and Evolution 20: 474-481, 2001.
- Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski JA: The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Molecular Breeding 2: 225-238, 1996.
- Prakash M: Molecular genetics. New Delhi: Discovery Publishing House, 2007.
- Prasad M, Varshney RK, Kumar A, Balyan HS, Sharma PC, Edwards KJ, Singh H, Dhaliwal HS, Roy JK, P.K. Gupta PK: A microsatellite marker associated with a QTL for grain protein content on chromosome arm 2DL of bread wheat. Theoretical and Applied Genetics 99: 341-345, 1999.
- Price NS: The origin and development of banana and plantain cultivation. V: Goven S (ed.): Bananas and Plantains. London: Chapman and Hall; 1-13, 1995.
- Ragot M, Hoisington DA: Molecular markers for plant breeding: comparisons of RFLP and RAPD genotyping costs. Theoretical and Applied Genetics 86: 975-984, 1993.
- Rands MRW, Adams WM, Bennun L, Butchart SHM, Clements A, Coomes D, Entwistle A, Hodge I, Kapos V, Schalerman JPW, Sutherland WJ, Vira B: Biodiversity conservation: Challenges beyond 2010. Science 329: 1298-1303, 2010.
- Rannala B, Yang Z: Probability distribution of molecular evolutionary trees: A new method of phylogenetic inference. Journal of Molecular Evolution 43: 304-311, 1996.

- Risterucci AM, Hippolyte I, Perrier X, Xia L, Caig V, Evers M, Huttner E, Kilian A, Glaszmann JC: Development and assessment of Diversity Arrays Technology for high-throughput DNA analyses in *Musa*. *Theoretical and Applied Genetics*, 119: 1093-1103, 2009.
- Rogers SO, Bendich AJ: Ribosomal RNA genes in plants: variability in copy number and in intergenic spacer. *Plant Molecular Biology* 9: 509-520, 1987.
- Rokas A, Williams BL, King N, Carroll SB: Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* 425: 798-804, 2003.
- Rosenberg MS, Kumar S: Traditional phylogenetic reconstruction methods reconstruct shallow and deep evolutionary relationships equally well. *Molecular Biology and Evolution* 18: 1823-1827, 2001.
- Roux N, Baurens FC, Doležel J, Hříbová E, Heslop-Harrison P, Town C, Sasaki T, Matsumoto T, Aert R, Remy S, Souza M, Lagoda P: Genomics of banana and plantain (*Musa* spp.), major staple crops in the tropics. V: Moore PH, Ming R (eds.): *Genomics of tropical crop plants*. New York: Springer, 83-111, 2008.
- Saitou N, Nei M: The Neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4: 406-425, 1987.
- Sales E, Nebauer SG, Mus M, Segura J: Population genetic study in the Balearic endemic plant species *Digitalis minor* (*Scrophulariaceae*) using RAPD markers. *American Journal of Botany* 88: 1750-1759, 2001.
- Sanderson M: A nonparametric approach to estimating divergence times in the absence of rate constancy. *Molecular Biology and Evolution* 14: 1218-1231, 1997.
- Sanderson MJ, Shaffer HB: Troubleshooting molecular phylogenetic analyses. *The Annual Review of Ecology, Evolution and Systematics* 33: 49-72, 2002.
- Sang T: Utility of low-copy nuclear gene sequences in plant phylogenetics. *Critical Reviews in Biochemistry and Molecular Biology* 37: 121-147, 2002.
- Shaffer C: Next-generation sequencing outpaces expectations. *Nature Biotechnology* 25: 149, 2007.
- Schlötterer C, Tautz D: Slippage synthesis of simple sequence DNA. *Nucleic Acids Research* 20:2211-2215, 1992.
- Siles BA, O'Neil KA, Fox MA, Anderson DE, Kuntz AF, Ranganath SC, Morris AC: Genetic fingerprinting of grape plant (*Vitis vinifera*) using random amplified polymorphic DNA (RAPD) analysis and dynamic size-sieving capillary electrophoresis. *Journal of Agricultural and Food Chemistry* 48: 5903-5912, 2000.

- Simmonds NW, Shepherd K: The taxonomy and origins of cultivated bananas. *Botanical Journal of the Linnean Society* 55: 302-312, 1955.
- Simmonds NW: The evolution of the bananas. London: Longman, 1962.
- Skinner DM, Beattie WG, Blattner FR: The repeat sequence of a hermit crab satellite deoxyribonucleic acid is $(-T-A-G-G-)_n \times (-A-T-C-C-)_n$. *Biochemistry* 13: 3930-3937, 1974.
- Smith SA, Beaulieu JM, Donoghue MJ: Mega-phylogeny approach for comparative biology: an alternative to supertree and supermatrix approaches. *BMC Evolutionary Biology* 9: 37, 2009.
- Sokolov BP: Primer extension technique for the detection of single nucleotide in genomic DNA. *Nucleic Acids Research* 18: 3671, 1990.
- Spooner D, Van Treuren R, De Vincente MC: Molecular markers for genebank management. Rome: Bioversity International, 2005.
- Stover RH, Simmonds NW: Bananas. Third edition. Burnt Mill, Harlow: Longman Scientific and Technical, 1987.
- Stuessy TF: Plant taxonomy: The systematic evaluation of comparative data. New York: Columbia University Press, 2009.
- Sunnucks P: Efficient genetic markers for population biology. *Trends in Ecology and Evolution* 15: 199-203, 2000.
- Swangpol S, Volkaert H, Sotto RC, Seelanan T: Utility of selected non-coding chloroplast DNA sequences for lineage assessment of *Musa* interspecific hybrids. *Journal of Biochemistry and Molecular Biology* 40: 577-587, 2007.
- Swofford DL, Sullivan J: Phylogeny inference based on parsimony and other methods using PAUP*. V: Salemi M, Vandamme AM (eds.): The phylogenetic handbook. A practical approach to DNA and protein phylogeny. New York: Cambridge University Press, 160-206, 2003.
- Takasaki T, Moriya Y, Okada K, Yamamoto K, Iwanami H, Bessho H, Nakanishi T: cDNA cloning of nine S alleles and establishment of a PCR-RFLP system for genotyping European pear cultivars. *Theoretical and Applied Genetics* 112: 1543-1552, 2006.
- Tanksley SD, Young ND, Paterson AH, Bonierbale MW: RFLP mapping in plant-breeding - New tools for an old science. *Bio-Technology* 7: 257-264, 1989.
- Tautz D, Trick M, Dover GA: Cryptic simplicity in DNA is a major source of genetic variation. *Nature* 322: 652-656, 1986.
- Thorne J, Kishino H, Painter I: Estimating the rate of evolution of the rate of molecular evolution. *Molecular Biology and Evolution* 15: 1647-1657, 1998.

- Till BJ, Jankowicz-Cieslak J, Sági L, Huynh OA, Utsushi H, Swennen R, Terauchi R, Mba C: Discovery of nucleotide polymorphisms in the *Musa* gene pool by Ecotilling. *Theoretical and Applied Genetics* 121: 1381-1389, 2010.
- Ude G, Pillay M, Nwakanma D, Tenkouano A: Genetic diversity in *Musa acuminata* Colla and *Musa balbisiana* Colla and some of their natural hybrids using AFLP markers. *Theoretical and Applied Genetics* 104: 1246-1252, 2002a.
- Ude G, Pillay M, Nwakanma D, Tenkouano A: Analysis of genetic diversity and sectional relationships in *Musa* using AFLP markers. *Theoretical and Applied Genetics* 104: 1239-1245, 2002b.
- Valouev A, Ichikawa J, Tonthat T, Stuart J, Ranade S, Peckham H, Zeng K, Malek JA, Costa G, McKernan K, Sidow A, Fire A, Johnson SM: A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning. *Genome Research* 18: 1051-1063, 2008.
- Vandamme AM: Basic concepts of molecular evolution. V: Salemi M, Vandamme AM (eds.): *The phylogenetic handbook. A practical approach to DNA and protein phylogeny*. New York: Cambridge University Press, 1-23, 2003.
- Van de Peer Y: Phylogeny inference based on distance methods: Theory. V: Salemi M, Vandamme AM (eds.): *The phylogenetic handbook. A practical approach to DNA and protein phylogeny*. New York: Cambridge University Press, 101-119, 2003.
- Venkatachalam L, Sreedhar RV, Bhagyalakshmi N: The use of genetic markers for detecting DNA polymorphism, genotype identification and phylogenetic relationships among banana cultivars. *Molecular Phylogenetics and Evolution* 47: 974-985, 2008.
- Virk PS, Fordlloyd BV, Jackson MT, Newbury HJ: Use of RAPD for the study of diversity within plant germplasm collections. *Heredity* 74: 170-179, 1995.
- Von Haeseler A, Strimmer K: Maximum likelihood methods with TREE-PUZZLE. V: Salemi M, Vandamme AM (eds.): *The phylogenetic handbook. A practical approach to DNA and protein phylogeny*. New York: Cambridge University Press, 137-159, 2003.
- Vos P, Hogers R, Bleeker M, Reijans M, Van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* 23: 4407-4414, 1995.
- Wang XL, Chiang TY, Roux N, Hao G, Ge XJ: Genetic diversity of wild banana (*Musa balbisiana* Colla) in China as revealed by AFLP markers. *Genetic Resources and Crop Evolution* 54: 1125-1132, 2007.
- Weising K, Weigand F, Driesel A, Kahl G, Zischler H, Epplen JT: Polymorphic simple GATA/GACA repeats in plant genomes. *Nucleic Acids Research* 17: 10128, 1989.

- Weising K, Nybom H, Wollf K, Kahl G: DNA Fingerprinting in Plants. Principles, Methods, and Applications. Second Edition. London: Taylor & Francis Group, 2005.
- Wendel JF, Schnabel A, Seelanan T: Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). Proceedings of the National Academy of Sciences of the USA 92: 280-284, 1995.
- Wenzl P, Carling J, Kudrna D, Jaccoud D, Huttner E, Kleinjohs A, Kilian A: Diversity arrays technology (DArT) for whole-genome profiling of barley. Proceedings of the National Academy of Sciences of the USA 101: 9915-9920, 2004.
- White TJ, Bruns T, Lee S, Taylor J: Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. V: Innis M, Gelfand D, Sninsky J, White T (eds.): PCR Protocols: A guide to methods and applications. San Diego: Academic Press, 315-322, 1990.
- Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV: DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Research 18: 6231-6235, 1990.
- Winter P, Kahl G: Molecular marker technologies for plant improvement. World Journal of Microbiology and Biotechnology 11: 438-448, 1995.
- Wolf M, Achtziger M, Schultz J, Dandekar T, Müller T: Homology modeling revealed more than 20,000 rRNA internal transcribed spacer 2 (ITS2) secondary structures. RNA 11: 1616-1623, 2005.
- Wolfe KH, Li WH, Sharp PM: Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proceedings of the National Academy of Sciences of the USA 84: 9054-9058, 1987.
- Wong C, Kiew R, Loh JP, Gan LH, Set O, Lee SK, Lum S, Gan YY: Genetic diversity of the wild banana *Musa acuminata* Colla in Malaysia as evidenced by AFLP. Annals of Botany 88: 1017-1025, 2001.
- Wong C, Kiew R, Argent G, Set O, Lee SK, Gan YY: Assessment of the validity of the sections in *Musa* (*Musaceae*) using AFLP. Annals of Botany 90: 231-238, 2002.
- Wu DY, Ugozzoli L, Pal BK, Wallace RB: Allele-specific enzymatic amplification of beta-globin genomic DNA for diagnosis of sickle cell anemia. Proceedings of the National Academy of Sciences of the USA 86: 2757-2760, 1989.
- Xie Y, McNally K, Li CY, Leung H, Zhu YY: A high-throughput genomic tool: diversity array technology complementary for rice genotyping. Journal of Integrative Plant Biology 48: 1069-1076, 2006.

- Young WP, Schupp JM, Keim P: DNA methylation and AFLP marker distribution in the soybean genome. *Theoretical and Applied Genetics* 99: 785-790, 1999.
- Zhang D, Mischke S, Goenaga R, Hemeida AA, Saunders JA: Accuracy and reliability of high-throughput microsatellite genotyping for cacao clone identification. *Crop Science* 46: 2084-2092, 2006.
- Zhang W, Qu J, Gu H, Gao W, Liu M, Chen J, Chen Z: Studies on the origin and evolution of tetraploid wheats based on the internal transcribed spacer (ITS) sequence of nuclear ribosomal DNA. *Theoretical and Applied Genetics* 104: 1099-1106, 2002.
- Zimmer EA, Martin SL, Beverley SM, Kan YW, Wilson AC: Rapid duplication and loss of genes coding for the alpha chains of hemoglobin. *Proceedings of the National Academy of Sciences of the USA* 77: 2158-2162, 1980.
- Zuckerlandl E, Pauling L: Molecular disease, evolution, and genetic heterogeneity. V: Bryson B, Vogel HJ (eds.): *Horizons in Biochemistry*. New York: Academic Press, 189-225, 1962.
- Zuckerlandl E, Pauling L: Evolutionary divergence and convergence in proteins. V: Bryson B, Vogel HJ (eds.): *Evolving genes and proteins*. New York: Academic Press, 97-116, 1965.

PŘEHLED ZÍSKANÝCH VÝSLEDKŮ

PUBLIKACE V ODBORNÝCH ČASOPISECH

PŘÍLOHA I

Němcová P, Valárik M, Hřibová E, Doležel J: A complex approach for unravelling *Musaceae* phylogeny at molecular level. *Acta Horticulturae*, (přijato k tisku), 2010.

[podíl doktorandky na práci: sepsání přehledového článku – 95 %]

PŘÍLOHA II

Christelová P, Valárik M, Hřibová E, De Langhe E, Doležel J: A multi gene sequence-based phylogeny of the *Musaceae* (banana) family. *BMC Evolutionary Biology*, 11: 103, 2011.

[podíl doktorandky na práci: odvození, sekvenování fylogenetických markerů, analýza dat-experimentální část – 90 %, sepsání publikace – 85%]

PŘÍLOHA III

Christelová P, Valárik M, Hřibová E, Van den houwe I, Chanellière S, Roux N, Doležel J: A platform for efficient genotyping in *Musa* using microsatellite markers. *AoB Plants*, (odesláno do tisku), 2011.

[podíl doktorandky na práci: optimalizace genotypovacího systému-experimentální část – 90 %, sepsání publikace – 90%]

PŘÍLOHA IV

Hřibová E, Čížková J, Christelová P, Taudien S, De Langhe E, Doležel J: The ITS1-5.8S-ITS2 sequence region in the *Musaceae*: structure, diversity and use in molecular phylogeny. *Plos One* 6: e17863, 2011.

[podíl doktorandky na práci: klonování a sekvenování ITS lokusu u vybraných zástupců-experimentální část – 20 %, sepsání publikace – 20%]

PŘÍLOHA V

Kopecký D, Bartoš J, Christelová P, Černoch V, Kilian A, Doležel J: Genomic constitution of *Festuca* × *Lolium* hybrids revealed by the DArTFest array. Theoretical and Applied Genetics 122: 355-363, 2011.

[podíl doktorandky na práci: sekvenování DArT markerů, základní úprava dat-experimentální část – 40%, sepsání publikace – 20%]

PŘÍLOHA VI

Bartoš J, Sandve SR, Kölliker R, Kopecký D, Christelová P, Stočes S, Østrem L, Larsen A, Kilian A, Rognli OA, Doležel J: Genetic mapping of DArT markers in the *Festuca-Lolium* complex and their use freezing tolerance association analysis. Theoretical and Applied Genetics 122: 1133-1147, 2011.

[podíl doktorandky na práci: sekvenování DArT markerů, základní úprava dat-experimentální část – 40%, sepsání publikace – 20%]

ABSTRAKTA PŘÍSPĚVKŮ NA KONFERENCÍCH

PŘÍLOHA VII

Hřibová E, Čížková J, Němcová P, Doležel J: Phylogenetic relationships in the *Musaceae* based on nucleotide sequence of the ITS1-5.8S-ITS2 region. V: Abstracts of the International Conference “Molecular Mapping & Marker Assisted Selection in Plants”. p. 58. University of Vienna, Vienna, 2008.

(abstrakt posteru)

PŘÍLOHA VIII

Hřibová E, Němcová P, Čížková J, Schillerová L, Doležel J: Phylogenetic relationships in family *Musaceae* based on the nucleotide sequence of ITS1-5.8S-ITS2 region. V: Abstracts of the Annual Meeting for the Society of Molecular Biology and Evolution (SMBE). p. 636. SMBE, Barcelona 2008.

(abstrakt posteru)

PŘÍLOHA IX

Hřibová E, Němcová P, Čížková J, Schillerová L, Kilian A, Doležel J: Nucleotide sequence of the ITS1-5.8S-ITS2 region and DArT markers shed light on phylogenetic relationships within the *Musaceae*. V: Abstracts of the International Conference on “Polyploidy, Hybridization and Biodiversity”. p. 143. INRA, Saint Malo, 2009.

(abstrakt posteru)

PŘÍLOHA X

Němcová P, Hřibová E, Valárik M, Čížková J, Schillerová L, Kilian A, Doležel J: Phylogenetic relationships in the family *Musaceae* based on the genic sequences, sequence of the ITS1-5.8S-ITS2 region and DArT markers. V: Global Perspectives on Asian Challenges. p. 18. International Society for Horticultural Science/ProMusa, Guangzhou, 2009.

(abstrakt přednášky)

PŘÍLOHA XI

Němcová P, Valárik M, Hřibová E, Doležel J: Použití genových sekvencí pro objasnění fylogenetických vztahů v rámci čeledi *Musaceae*. V: Sborník abstrakt z vědecké konference “5. Metodické dny”. p. 77. Bulletin České společnosti experimentální biologie rostlin a Fyziologické sekce Slovenské botanické společnosti, Malenovice, 2009.

(abstrakt posteru)

PŘÍLOHY

PŘÍLOHA I

A complex approach for unravelling *Musaceae* phylogeny at molecular level

Němcová P., Valárik M, Hřibová E, Doležel J.

Acta Horticulturae (přijato k tisku), 2010.

[podíl doktorandky na práci: sepsání přehledového článku – 95 %]

A Complex Approach for Unravelling Musaceae Phylogeny at Molecular Level

P. Němcová, E. Hřibová, M. Valárik and J. Doležel

Laboratory of Molecular Cytogenetics and Cytometry, Institute of Experimental Botany, Academy of Sciences of the Czech Republic

Keywords: DArT, ITS, low-copy genes, molecular phylogenetics, *Musa*.

Abstract

Bananas and plantains (*Musa* species, Zingiberales) are a relatively underresearched plant group in contrast to their great socio-economic importance and their significance as one of the earliest domesticated crops. The classification of species and clones is hampered by the traditional morphology-based taxonomy, with limited knowledge on specific molecular characteristics of individual taxons. The phylogeny of the Musaceae has never been fully resolved, and the accuracy of the current classification system is a common point of debate among the scientific community. Different types of molecular markers have been used to investigate the diversity at various taxonomic levels within *Musa* species and have reached various degrees of success in terms of phylogenetic reconstruction. The phylogenetic history of Musaceae as a part of the order Zingiberales was investigated in previous studies (Kress et al., 2001). However, within the family Musaceae itself, a few papers have addressed this question (Lescot et al., 2008; Boonruangrod et al., 2008) and the evolutionary relationships remain fuzzy. This paper is an overview of molecular-based approaches that have been employed by our research group in an attempt to shed more light on Musaceae diversity and evolutionary relationships, and thereby serve for the classification and development of reliable markers for unambiguous banana species identification.

INTRODUCTION

Modern, currently grown bananas are mainly diploid and triploid intra- and interspecific hybridisations between the two wild *Musa* progenitors, *Musa acuminata* Colla (AA_w genome) and *Musa balbisiana* Colla (BB_w genome). The classification of *Musa* genotypes is traditionally based on morphological descriptors and the basic chromosome number (Cheesman, 1947; Simmonds and Shepherd, 1955). Several studies have investigated the diversity of bananas and plantains, and have eventually proposed changes to the original system. Diversity studies have been carried out using a wide range of molecular markers, e.g. amplified fragment length polymorphism – AFLPs (Ude et al., 2002; Wang et al., 2007), microsatellites – SSRs (Buhariwalla et al., 2005), random amplified polymorphic DNA – RAPD (Pillay et al., 2000; Venkatachalam et al., 2008) or PCR combined with restriction fragment length polymorphism – PCR-RFLP (Ning et al., 2007). Although each of the methods mentioned above allowed for assessment of the diversity and relationships among certain group of cultivars, a clear picture of Musaceae phylogeny is still lacking.

Here, we provide an overview of a set of molecular tools that are being used in our research to address banana phylogenetics at the molecular level.

RIBOSOMAL DNA ITS1-5.8S-ITS2 REGION

In higher plants, the 45S ribosomal RNA gene unit consists of 18S, 5.8S and 26S rRNA coding regions, separated by internal transcribed spacers (ITS1 and ITS2). The internal transcribed spacer is the most popular non-plastid DNA region used in plant phylogenetic studies. Numerous features make the ITS locus an ideal target for molecular phylogenetics. First, it is a highly variable DNA region with a significant level of inter-specific mutations (Baldwin, 1992). At the same time, the highly variable ITS locus is surrounded by conserved sequences, which facilitates the use of a set of universal primers for PCR amplification. Moreover, the average length of the spacers varies around 700 bp, providing perfect material for diversity analysis at the sequence level, as sequencing can be performed without the need for a time-consuming internal primer design. On the other hand, the 18S, 5.8S and 26S rDNA repeat units show homogenisation within species, probably due to unequal crossing-over and gene conversion, mechanisms leading to concerted evolution (Dover, 1982). Concerted evolution causes the different copies of the spacer sequences to be more similar within species than among species (Elder and Turner, 1995). One of the accented arguments against the exclusive use of ITS region for phylogenetic studies is that such an analysis based on a single locus within the genome may not thoroughly reflect the real picture of evolution.

In *Musa*, a restriction-based PCR-RFLP strategy was used to study the ITS region in selected species and was proved to be useful for distinguishing between the A and B genome (Nwakanma et al., 2003). In order to get down to the level of distinguishing species, subspecies and clones, it would be necessary to study the internal transcribed spacer in detail by a comparative sequence analysis of a number of representatives of the genus. Recently, our research group has carried out a project on comparative sequence analysis of the ITS1 and ITS2 region from 88 accessions, including AAA, AAB, ABB, AxS and AxT hybrids (Hřibová et al., 2009).

INTRON-SPANNING LOW-COPY NUCLEAR GENE-DERIVED MARKERS

An alternative target for phylogenomics are low-copy nuclear genes. As compared to the ITS region, they offer the advantage of a lower level of homoplasy, which as a result of the convergent evolution impedes the possibility of building unambiguous alignment of homologous genome regions (Álvarez and Wendel, 2003). At the same time, the phylogenetic informativeness of obtained data is relatively high, especially when non-coding introns are included in the analysis (Bailey and Doyle, 1999). As introns are non coding, they are subject to a less stringent selection pressure in contrast to the coding regions. Thus, intron-spanning single-copy gene-derived markers serve as a rich source of targets that can be used with success to resolve phylogenies, even at lower taxonomic levels (Creer, 2007; Peng et al., 2009).

Zou et al. (2008) used a set of 142 single-copy genes from rice, which proved to have enough informative power to completely resolve the diversification of the rice genus. However, development of a set of low-copy nuclear gene markers for non-model organisms requires large investments in terms of funding and labour, especially when compared to the use of more or less universally amplifiable rDNA intergenic spacers or high-throughput techniques, such as DArT and/or next-generation sequencing methods. Feltus et al. (2006) have surveyed the use of “conserved intron-scanning primers” (CISP) derived from the genomic sequence information of completely sequenced species. Another study (Lohithaswa et al., 2007) has worked this system out for a number of representatives from the orders

Poales, Arecales, Asparagales and Zingiberales, including *Musa*, showing the potential utilisation of this type of CISP in comparative studies. There are several drawbacks that need to be carefully considered when using intron-spanning gene markers for phylogenomics, such as the problem of paralogy, heterozygotic length variants, insertion/deletions treatment, etc. (reviewed in Creer, 2007). On the other hand, when all the potential interfering properties of this type of phylogenetic marker are handled cautiously, targeting low-copy or, ideally, single-copy nuclear genes that are evenly distributed across the genome provides an excellent source of locus-independent sequence information that can address different evolutionary questions (Feliner and Roselló, 2007).

DIVERSITY ARRAY TECHNOLOGY (DArT) MARKERS

In contrast to the comparative sequence-based analysis, Diversity Array Technology (DArT) represents a different approach to estimate species diversity that does not require sequence information. This method has been widely used for genotyping purposes on numerous organisms, including *Musa* species (Amorim et al., 2009). DArT technology is based on solid-state surface (microarray) hybridisation. Polymorphic DNA fragments from a given organism are amplified and spotted onto the microarray, which is subsequently used for hybridisation with a reduced representation of the genomic DNA of the species to be genotyped (Jaccoud et al., 2001). It is a high-throughput technique, where hundreds of polymorphic markers can be amplified in a single PCR, suitable for automation (Luikart et al., 2003).

DArT analysis mainly results in dominant markers, which compared to codominant markers possess lower informative power. However, it can be successfully implemented in phylogenetic studies, if not as the exclusive source of phylogenetic information, then as a complementary approach to assess the diversity within the species of interest.

CONCLUSIONS

This paper describes different approaches that have been used by our research team in the Laboratory of Molecular Cytogenetics and Cytometry for studying phylogeny of the *Musa* species. Our preliminary results confirmed that comparative sequence analysis of single-copy genes is a powerful tool for resolving the evolutionary history of Musaceae and is complementary with the analysis of the ribosomal DNA ITS1-5.8S-ITS2 region and DArT markers analysis. We believe that well-balanced combination of these methods provides a useful toolbox for Musaceae phylogeny analysis and can significantly contribute to reliable identification of *Musa* species, subspecies and cultivated clones.

ACKNOWLEDGEMENTS

This work has been supported by the Grant Agency of the Academy of Sciences of the Czech Republic (grant award no. IAA600380703) and the International Atomic Energy Agency (research agreement no. 13192).

Literature Cited

- Álvarez, I. and Wendel, J.F. 2003. Ribosomal ITS sequences and plant phylogenetic inference. *Mol. Phylogenet. Evol.* 29:417-434.
- Amorim, E.P., Vilarinhos, A.D., Cohen, K.O., Amorim, V.B.O., dos Santos-Serejo, J.A., Oliveira e Silva, S., Pestana, K.N., dos Santos V.J., Paes, N.S., Monte, D.C.

- and dos Reis, R.V. 2009. Genetic diversity of carotenoid-rich bananas evaluated by Diversity Arrays Technology. *Genet. Mol. Biol.* 32(1):96-103.
- Bailey, C.D. and Doyle, J.J. 1999. Potential phylogenetic utility of the low-copy nuclear gene *pistillata* in dicotyledonous plants: Comparison to nrDNA ITS and *trnL* intron in *Sphaerocardamum* and other Brassicaceae. *Mol. Phylogenet. Evol.* 13:20-30.
- Baldwin, B.G. 1992. Phylogenetic utility of the internal transcribed spacer of nuclear ribosomal DNA in plants: An example from the Compositae. *Mol. Phylogenet. Evol.* 1:3-16.
- Boonruangrod, R., Desai, D., Fluch, S., Berenyi, M. and Burg, K. 2008. Identification of cytoplasmic ancestor gene-pools of *Musa acuminata* Colla and *Musa balbisiana* Colla and their hybrids by chloroplast and mitochondrial haplotyping. *Theor. Appl. Genet.* 118:43-55.
- Buhariwalla, H.K., Jarret, R.L., Jayashree, B., Crouch, J.H. and Ortiz, R. 2005. Isolation and characterization of microsatellite markers from *Musa balbisiana*. *Mol. Ecol. Notes.* 5:327-330.
- Cheesman, E.E. 1947. Classification of the bananas. *Kew. Bull.* 2:106-117.
- Creer, S. 2007. Choosing and using introns in molecular phylogenetics. *Evol. Bioinform. Online.* 3:99-108.
- Dover, G. 1982. Molecular drive: a cohesive mode of species evolution. *Nature.* 299:111-117.
- Elder Jr., J.F. and Turner, B.J. 1995. Concerted evolution of repetitive DNA sequences in eukaryotes. *Q. Rev. Biol.* 70:297-320.
- Feliner, G.N. and Roselló, J.A. 2007. Better the devil we know? Guidelines for insightful utilization of nrDNA ITS species-level evolutionary studies in plants. *Mol. Phylogenet. Evol.* 44:911-919.
- Feltus, F.A., Singh, H.P., Lohithaswa, H.C., Schulze, S.R., Silva, T.D. and Paterson, A.H. 2006. A comparative genomics strategy for targeted discovery of single-nucleotide polymorphisms and conserved-noncoding sequences in orphan crops. *Plant Physiol.* 140:1183-1191.
- Hřibová, E., Němcová, P., Čížková, J., Schillerová, L., Kilian, A. and Doležel, J. 2009. Nucleotide sequence of the ITS1-5.8S-ITS2 region and DArT markers shed light on phylogenetic relationships within the *Musaceae*. Program and abstracts: International Conference on Polyploidy, Hybridization and Biodiversity, Saint-Malo, France 17-20 May. p.143.
- Jaccoud, D., Peng, K., Feinstein, D. and Kilian, A. 2001. Diversity arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acids Res.* 29:4.
- Kress, W.J., Prince, L.M., Hahn, W.J., and Zimmer, E.A. 2001. Unraveling the evolutionary radiation of the families of the Zingiberales using morphological and molecular evidence. *Syst. Biol.* 50(6):926-944.
- Lescot, M., Piffanelli, P., Ciampi, A.Y., Ruiz, M., Blanc, G., Leebens-Mack, J., da Silva, F.R., Santos, C.M.R., D'Hont, A., Garsmeur, O., Vilarinhos, A.D., Kanamori, H., Matsumoto, T., Ronning, C.M., Cheung, F., Haas, B.J., Althoff, R., Arbogast, T., Hine, E., Pappas Jr., G.J., Sasaki, T., Souza Jr., M.T., Miller, R.N.G., Glaszmann, J-Ch. and Town, Ch.D. 2008. Insights into the *Musa* genome: Syntenic relationships to rice and between *Musa* species. *BMC Genomics* 9:58.

- Lohithaswa, H.C., Feltus, F.A., Singh, H.P., Bacon, C.D., Bailey, C.D. and Paterson, A.H. 2007. Leveraging the rice genome sequence for monocot comparative and translational genomics. *Theor. Appl. Genet.* 115:237-243.
- Luikart, G., England, P.R., Tallmon, D., Jordan, S. and Taberlet, P. 2003. The power and promise of population genomics: from genotyping to genome typing. *Nat. Rev. Genet.* 4:981-994.
- Ning, S.P., Xu, L.B., Lu, Y., Huang, B.Z. and Ge, X.J. 2007. Genome composition and genetic diversity of *Musa* germplasm from China revealed by PCR-RFLP and SSR markers. *Sci. Hortic.-Amsterdam* 114:281-288.
- Nwakanma, D.C., Pillay, M., Okoli, B.E. and Tenkouano, A. 2003. PCR-RFLP of the ribosomal DNA internal transcribed spacers (ITS) provides markers for the A and B genomes in *Musa* L. *Theor. Appl. Genet.* 108:154-159.
- Peng, Z., Elango, N., Wildman, D.E. and Yi, S.V. 2009. Primate phylogenomics: developing numerous nuclear non-coding, non-repetitive markers for ecological and phylogenetic applications and analysis of evolutionary rate variation. *BMC Genomics* 10:247.
- Pillay, M., Nwakanma, D.C. and Tenkouano, A. 2000. Identification of RAPD markers linked to A and B genome sequences in *Musa* L. *Genome* 43:763-767.
- Simmonds, N.W. and Sheperd K. 1955. The taxonomy and origins of the cultivated bananas. *J. Linn. Soc.* 55:302-312.
- Ude, G., Pillay, M., Nwakanma, D. and Tenkouano, A. 2002. Genetic diversity in *Musa acuminata* Colla and *Musa balbisiana* Colla and some of their natural hybrids using AFLP. *Theor. Appl. Genet.* 104:1246-1252.
- Venkatachalam, L., Sreedhar, R.V. and Bhagyalakshmi. 2008. The use of genetic markers for detecting DNA polymorphism, genotype identification and phylogenetic relationships among banana cultivars. *Mol. Phylogenet. Evol.* 47:974-985.
- Wang, X.L., Chiang, T.Y., Roux, N., Hao, G. and Ge, X.J. 2007. Genetic diversity of wild banana (*Musa balbisiana* Colla) in China as revealed by AFLP markers. *Genet. Resour. Crop Evol.* 54:1125-1132.
- Zou, X.H., Zhang, F.M., Zhang, J.G., Zang, L.L., Tang, L., Wang, J., Sang, T. and Ge, S. 2008. Analysis of 142 genes resolves the rapid diversification of the rice genus. *Genome Biol.* 9:R49.

PŘÍLOHA II

A multi gene sequence-based phylogeny of the *Musaceae* (banana) family

Christelová P, Valárik M, Hřibová E, De Langhe E, Doležel J.

BMC Evolutionary Biology 11: 103, 2011.

*[podíl doktorandky na práci: odvození, sekvenování fylogenetických markerů, analýza dat-
experimentální část – 90 %, sepsání publikace – 85%]*

PŘÍLOHA III

A platform for efficient genotyping in *Musa* using microsatellite markers

Christelová P, Valárik M, Hřibová E, Van den houwe I, Chanellière S, Roux N,
Doležel J.

AoB Plants (odesláno do tisku), 2011.

[podíl doktorandky na práci: optimalizace genotypovacího systému-experimentální
část – 90 %, sepsání publikace – 90%]

A Platform for Efficient Genotyping in *Musa* using Microsatellite Markers

Pavla Christelová¹, Miroslav Valárik¹, Eva Hřibová¹, Ines Van den houwe², Stéphanie Channelière³, Nicolas Roux³ and Jaroslav Doležel^{1,*}

¹Centre of the Region Haná for Biotechnological and Agricultural Research, Institute of Experimental Botany, Sokolovská 6, Olomouc, CZ-77200 Czech Republic.

²Laboratory of Tropical Crop Improvement, Katholieke Universiteit Leuven, Kasteelpark Arenberg 13, B-3001 Leuven, Belgium.

³Bioversity International, Parc Scientifique Agropolis II, Montpellier Cedex 5, 34397, France.

*Corresponding author

Corresponding author's e-mail address: dolezel@ueb.cas.cz

ABSTRACT

Background and aims

Bananas and plantains (*Musa* spp.) are one of the major fruit crops worldwide with acknowledged importance as a staple food for millions of people. The rich genetic diversity of this crop is, however, endangered by diseases, adverse environmental conditions and changed farming practices, and the need for its characterization and preservation is urgent. With the aim to provide a simple and robust approach for molecular characterization of *Musa* species, we developed an optimized genotyping platform using 19 published SSR markers.

Methodology

The genotyping systems is based on 19 microsatellite loci, that are scored using fluorescently labeled primers and high-throughput capillary electrophoresis separation with high resolution. This genotyping platform was tested and optimized on a set of 70 diploid and 38 triploid banana accessions.

Principal results

The marker set used in this study provided enough polymorphism to discriminate between individual species, subspecies and subgroups of all accessions of *Musa*. Likewise, the capability of identifying duplicate samples was confirmed. Based on the results of a blind test, the genotyping system was confirmed to be suitable for characterization of unknown accessions.

Conclusions

Here we report on the first complex and standardized platform for molecular characterization of *Musa* germplasm that is ready-to-use for the wider *Musa*

research and breeders community. We believe that this genotyping system offers a versatile tool that can accommodate all possible requirements for characterizing *Musa* diversity, and is economical for samples ranging from one to many accessions.

INTRODUCTION

The important role of bananas and plantains (*Musa* spp.) as one of the top world trade commodities and as a food security for millions of people especially in humid tropics is unquestionable. However, this crop faces a serious endangerment by numerous pests and diseases. Breeding efforts are hampered by a high degree of banana sterility and a lack of characterized germplasm as potential parents for breeding. Currently grown banana cultivars are mainly triploid clones, which originated as intraspecific hybrids of *M. acuminata* and interspecific hybrids between *M. acuminata* and *M. balbisiana*, with a possible involvement of a few other species within the genus. To set up an efficient strategy for breeding improved banana varieties and support the choice of crossing parents, a solid understanding of the genetic diversity of available resources is needed. Coessentially, the conservation of the existing gene resources is highly eligible, especially when we can observe the continuous loss of banana diversity due to indelicate environmental treatment of the rain forests, as well as changed farming practices of the smallholders. The main objectives and means for *Musa* diversity conservation were formulated in The Global Conservation Strategy for *Musa* (INIBAP 2006) under the scope of GMGC (Global Musa Genomics Consortium). Nevertheless, irrespective of the selected strategy, efficient collecting and preserving the banana diversity highly depends on the unambiguous sample identification. To avoid problems of duplicities within national, regional and Global germplasm collections, an accurate and standardized characterization of newly introduced accessions as well as those already deposited in gene banks would be of a great benefit. This rationalization effort will allow *Musa* accessions to be efficiently conserved.

Traditional classification of *Musa* species is based on morphological characters and chromosome counts (basic chromosome number; x) (Cheesman 1947; Simmonds and Shepherd 1955). Although a morphotaxonomic system allows for a differentiation of specific banana clones (Stover and Simmonds 1987), insufficiencies of this approach start to emerge as the genetic basis of the plants under study gets narrow. Additionally, a small change at the DNA level can cause a large phenotypic manifestation, while sometimes none or minor morphological changes can be observed after extensive genetic changes. Obviously, a classification system that relies exclusively on the phenotypic manifestations of the genome suffers from a limited accuracy (Crouch 2000; De Langhe 2005), but can be made robust if supported by molecular-based characterization.

The enormous increase in the availability of various molecular techniques over the past decades has facilitated the classification of new banana cultivars, as well as reassessment of the traditional taxonomy. Among the broad portfolio of molecular tools, some of the markers have gained special

attention in terms of their use in diversity studies and molecular characterization of banana genotypes. Most recently, the Diversity Arrays Technology (DArT) was used for the assessment of the genetic diversity within *Musa* spp. (Risterruci *et al.* 2009). While holding the advantage of a high-throughput approach suitable for large numbers of genotypes, its use for a limited number of samples in a short turn-around time would rank it within the more demanding methods in terms of funding support. The same applies to the Genotyping by Sequencing approach, which has gained special attention recently (Elshire *et al.* in press). Other molecular markers applied in *Musa* diversity studies were RAPDs (random amplified polymorphic DNA; Pillay *et al.* 2000; Pillay *et al.* 2001; Ruangsuttapha *et al.* 2007; Venkatachalam *et al.* 2008) and AFLPs (amplified fragment length polymorphism; Loh *et al.* 2000; Wong *et al.* 2001a; Ude *et al.* 2002; Wang *et al.* 2007). Both of these markers have relatively high level of polymorphism, but they are dominant and in case of the RAPDs, their reproducibility is a serious limitation (Jones *et al.* 1997). The more advantageous co-dominant markers were also used for *Musa*, such as the RFLPs (restriction fragment length polymorphism; Gawell *et al.* 1992; Nwakanma *et al.* 2003; Ning *et al.* 2007) and SSRs (simple sequence repeats; e.g. Kaemmer *et al.* 1997; Grapin *et al.* 1998; Lagoda *et al.* 1998; Buhariwalla *et al.* 2005). While RFLPs perform well in terms of reproducibility, they suffer from relatively low level of polymorphism and are laborious to use. On the contrary, the SSR markers are outperforming the RFLPs and RAPDs in all the above mentioned aspects.

Microsatellites (SSRs) are stretches of simple 1-6 base pair long repeat motifs arranged tandemly within the genomes of prokaryotic and eukaryotic organisms. Their flanking regions that are usually highly conserved are suitable for designing locus-specific primers. SSRs have been successfully applied in molecular genotyping of many important crops such as rice (Pessoa-Filho *et al.* 2007), cereals (Hayden *et al.* 2007), grapevine (This *et al.* 2004) or cacao (Zhang *et al.* 2006). Moreover, the use of SSR markers opens the possibility of automation and multiplexing which significantly increases the throughput of the technique.

With the aim of developing a standardized protocol to classify *Musa* germplasm we have tested and optimized the use of 22 published SSR markers on a set of banana genotypes. The goal of the present study was to investigate the potential of this marker set to distinguish individual accessions and to develop a standardized procedure for *Musa* genotyping that could serve as a basis for molecular characterization of new samples introduced into the Global *Musa* gene bank (International Transit Centre, ITC, Leuven, Belgium) as well as to the wider *Musa* research and breeding community.

MATERIALS AND METHODS

Plant material and the Reference DNA collection

The Reference DNA collection, comprising a total of 65 accessions (Table 1) was established to represent genetic diversity within the genus *Musa* and

corresponding *in vitro* plantlets be available for distribution from the Bioversity International Transit Center (ITC). The genomic DNA of 61 out of the 65 accessions is stored in the Musa Genome Resources Centre (http://www.musagenomics.org/cetest_firstpage1/genomic_dna.html) and is available for distribution. Out of the 65 accessions, 54 were successfully included in the analysis (Table 1). To extend the diploid representation of the genotype set, 39 additional diploid accessions were included (Table 2), with three of them being duplicate samples to the Reference DNA collection. These duplicates were included intentionally to test the capability of the genotyping platform to identify sample duplicities. All the 39 additional diploid accessions originated from the ITC collection (International Transit Centre, Leuven, Belgium) as *in vitro* rooted plants and were maintained in a heated greenhouse after transfer to soil. DNA of these 39 entries was isolated from young leaf tissue using the Invisorb® Spin Plant Mini kit (Invitex, Berlin, Germany), following the manufacturer's instructions.

PCR amplification and fragment analysis

The 22 SSR loci (Table 3) were amplified using specific primers (Crouch *et al.* 1998; Lagoda *et al.* 1998; Hippolyte *et al.* 2010) that were adjusted by 5'-M13 tails to enable the use of universal fluorescently labeled primer according to Schuelke (2000). Four different fluorophores were used for the primer labeling (6-FAM, VIC, NED, PET), allowing for subsequent multiplexing of the reactions (Table 3). The reaction was performed in a final volume of 20 µl containing 10 ng of template genomic DNA, reaction buffer (consisting of 10mM Tris-HCl (pH 8), KCl 50 mM, 0.1% Triton-X100 and 1.5 mM MgCl₂), 200 µM dNTPs (each), 1U of Taq polymerase, 8 pmol of the M13-tailed locus specific forward primer, 6 pmol of the fluorescently labeled universal M13 forward primer, 10 pmol of the locus specific reverse primer. The cycling conditions were set as follows; initial denaturation step at 94°C for 5 min, followed by 35 cycles of denaturation (94°C/45 s), annealing at temperature corresponding to the locus-specific primer (1 min) and extension (72°C/1 min). The final extension was allowed for 5 min at 72°C. Purification of the PCR products was performed by ethanol/ sodium acetate precipitation. Three independent PCR reactions were performed in order to improve the accuracy of alleles binning.

For automatic capillary electrophoresis, optimized amounts of amplification products were combined with highly deionized formamide and internal standard (GeneScan™-500 LIZ size standard, Applied Biosystems, Foster City, CA, USA). After 5 min denaturation at 95°C, samples were loaded onto the automatic 96-capillary ABI 3730xl DNA Analyzer and electrophoretic separation and signal detection was carried out with default module settings. In order to reduce the cost and increase the capacity of the genotyping platform, samples were multiplexed for the second and third round of electrophoretic separation. Up to 4-fold multiplexing was applied by combining 4 PCR products, labeled with different fluorescent dyes (6-FAM, VIC, NED and PET; Table 3) into a single sample for loading. The level of multiplexing could be further increased by combining products of different expected lengths, labeled with the identical fluorescent dyes.

Fragment sizing and data analysis

The resulting data were analyzed and called for alleles using GeneMarker® v1.75 (Softgenetics, LLC., State College, PA, USA). Automated scoring of the data was followed by a careful manual check and low quality DNA samples were discarded from the analysis. The marker panels were built based on the allele calls of the Reference DNA Collection sample set and later extended by additional diploid accessions allele calls, in order to increase the reference SSR-profiles database. Bins for each allele were set with respect to the allele frequencies and signal strength extracted from the three repeated runs of each sample.

The diploid and triploid accessions were analyzed separately, because in case of polyploid species, the polysomic inheritance brings along the simultaneous occurrence of several alleles of a single SSR. In such a situation, the exact number of copies of individual alleles cannot be determined; therefore the genotypic data are converted into binary (coded by 1-presence/ 0-absence) and analyzed as a dominant marker's record (Weising *et al.* 2005). Both, genotypic and binary data were used to generate a genetic similarity matrices based on Nei's genetic distance coefficient (Nei 1973) in the software PowerMarker v3.25 (Liu and Muse 2005). The unweighted pair-group method with arithmetic mean (UPGMA; Michener and Sokal 1957) was used to assess the relationship between individual genotypes. The results of UPGMA cluster analysis were visualized in a form of a tree using TreeView v1.6.6 (Page 1996). Polymorphism information content (PIC) and heterozygosity of individual markers was estimated in Powermarker v3.25. The overall probability of identity (P_{ID}) of unrelated multilocus genotypes was assessed according to Paetkau *et al.* (1995) as implemented in the IDENTITY program (Wagner and Sefc 1999).

Blind test

In order to verify the reliability of the optimized genotyping platform and its potential as a standardized methodology for molecular characterization of new accessions, a set of anonymous samples was analyzed (Table 4). The genomic DNA was extracted from lyophilized leaf tissue provided by ITC and samples were analyzed following identical experimental procedure as for the Reference DNA collection. Negative and positive controls (5 previously analyzed reference genotypes) were included in the Blind test to assure correct allele sizing and control the consistency of the electrophoretic condition. The unknown samples were coded numerically and their true identity was disclosed by our partners only after the data analysis. As revealed subsequently, the Blind test sample set contained additional four samples that were duplicates to the Reference DNA collection (see Table 1 and 4).

Genotyping errors handling

To eliminate the genotyping errors several precautions were employed in the genotyping pipeline, following the recommendations by Bonin *et al.* (2004). First, to minimize the allelic dropout effect, the multitube approach (Taberlet *et al.* 1996) was used with three independent reactions for each marker/genotype combination. The error prone samples with low quality DNA

were discarded from the analysis. Second, the multilocus genotype was examined and accessions differing at a single locus were carefully inspected and reanalyzed (if needed) to confirm the difference. Third, to decrease the human factor errors, the sample preparation was performed by two different persons for the replicated reactions. Data evaluation was ruled by strictly pre-set parameters to avoid errors such as misinterpretation of stutter peaks.

RESULTS

Twenty two SSR markers were selected by CIRAD as a set enabling to distinguish between individuals in the *Musa* Reference DNA collection (Crouch *et al.* 1998; Lagoda *et al.* 1998; Hippolyte *et al.* 2010; website1; Table 3). After the initial double-repeated primer test screening using our protocol, 19 markers were selected out of the initial 22 markers set, for their clear reproducible amplification pattern. The three markers that were excluded from the analysis produced extensive stuttering of peaks, disabling the reproducible interpretation of the SSR profiles. All further analyses were performed with the selected 19 SSRs. Altogether, the SSR profiles were collected for 70 diploid and 38 triploid banana accessions.

Analysis of diploid accessions

Diploid accessions were underrepresented in the Reference DNA collection; therefore, we decided to include additional diploids into the analysis to increase the number of reference SSR-profiles (Table 2). In the resulting set of 70 diploid accessions (incl. the Blind test entries), a total of 292 alleles were scored from the 19 loci, with the average of 15.4 alleles per locus. The observed heterozygosity (the fraction of all individuals who are heterozygous for the observed locus) ranged between 0.179 and 0.714 (mean 0.450). The polymorphism information content (PIC) of the markers used was relatively high (mean value 0.827), ranging between 0.625 – 0.936 (see Table 5 for details). The P_{ID} probability of identity (combined over all loci), which represents the probability of observing identical genotypes purely by chance, was 9.44×10^{-29} denoting the extremely high resolution power of this marker set.

The UPGMA cluster analysis based on the Nei (1973) genetic distance revealed relatively clear grouping of genotype groups and subgroups (Figure 1). The B-genome representatives, *M. balbisiana* including the diploid hybrid cultivars (AB and BBxT) formed a separate cluster (cluster I). The A-genome representatives, *M. acuminata* species were grouped in several clusters depending upon their subspecies classification. *M. acuminata* ssp. *banksii* entries grouped within the cluster II, *M. acuminata* ssp. *microcarpa* grouped together with *M. schizocarpa* and AS hybrids within the cluster III. The sole representative of *errans* subspecies, cultivar Agutay, was present at the separate clade with relation to the above described *M. acuminata* clusters. Subcluster VI contained the *M. acuminata* ssp. *zebrina* representatives. Subspecies *burmannica*, *burmannicoides* and *siamea* were grouped within cluster VII, sharing their position with several entries from the section *Rhodochlamys*. *M. acuminata* ssp. *malaccensis* subspecies formed a separate cluster labeled VIII (Figure 1). Most of the AA cultivars were

grouped within the cluster IV. The Australimusa section representatives included in the study formed cluster V, together with *M. beccarii* (classified under the Callimusa section). *M. coccinea*, another representative of the Callimusa section, was separated from all the other groups, resembling a behavior of an outgroup species. As mentioned before, Rhodochlamys species were partly present in the cluster VII (specifically the *M. ornata* and *M. mannii* entries). *M. velutina* accessions, another representatives of the Rhodochlamys section formed a separate cluster labeled IX together with a single *M. ornata* accession (ITC 1330).

Blind test with diploid accessions

When the anonymous samples were included into the dataset, the clustering was slightly changed (Figure 2). The position of accession Agutay (*M. acuminata* ssp. *errans*) moved into the cluster II containing mostly the *M. acuminata* ssp. *banksii* entries. Another alteration could be seen in the position of *M. acuminata* ssp. *zebrina* species, which no longer formed a separate subclade (previously labeled with VI), but instead clustered within the cluster IV containing the AA cultivars. Finally, cluster VII, although not changed in the content, showed now a different subclustering pattern, with the *M. acuminata* ssp. *burmannica*, *burmannicoides* and *siamea* species grouped together within one subcluster (VIIa), separated from the Rhodochlamys entries (subcluster VIIb).

Out of the nine anonymous accessions, eight were assessed correctly as the closest related species to the corresponding reference accession (Figure 2). The only exception was blind sample no.4 (*M. acuminata* ssp. *malaccensis* ITC 0250), which did not group together with its reference genotype (the same ITC 0250 accession), but instead clustered together within the *M. acuminata* ssp. *banksii* subgroup (clade II). The multilocus genotypes of the blind sample no.4 (ITC 0250) and the closest related genotype Higa (ITC 0428) differed at a single locus only, suggesting that the blind sample no.4 belonged very likely to the *banksii* subspecies.

In order to further investigate this incongruence in the blind test results, we conducted the ITS locus sequence analysis according to Hřibová *et al.* (in press) in the problematic *malaccensis* accessions. This analysis confirmed that the blind sample no.4 was not identical to the genotype *M. acuminata* ssp. *malaccensis* ITC 0250, which was originally received from the ITC and stored in the local greenhouse [see **ADDITIONAL INFORMATION**]. The results are, however, not conclusive about the identity of the blind sample no.4, as only a single representative of the *banksii* subspecies was used for the ITS analysis in our previous study. Thus, it cannot be explicitly stated, whether the blind sample no.4 is different genotype of *M. acuminata* ssp. *malaccensis* or ssp. *banksii*, or rather a hybrid between *malaccensis* and *banksii* subspecies. Only a more detailed sequence analysis would probably provide a definite answer.

Analysis of triploid accessions

Altogether, 38 triploid accessions were analyzed (incl. the Blind test entries). The 19 microsatellite loci scored a total of 267 alleles, ranging between 8 and

24 per locus, with the mean value of 14 alleles per locus. The average PIC of the SSR markers applied on the triploid accessions was 0.850 (Table 6).

The UPGMA analysis majority rule consensus tree showed two main clusters, cluster A and B (Figure 3). Cluster A contained solely the AAA hybrid accessions, with separated clade bearing the Lujugira/Mutika subgroup representatives, as well as a distinct clade leading to the edible species from the Cavendish and Gros Michel subgroups. Among all the AAA entries included in the analysis, only the accession Pisang Berangan clustered outside the A cluster, sharing a clade (IVa) with the African plantains representatives within the main cluster B. The second main cluster B was split into 4 subclusters/subclades. While subcluster II was formed exclusively by the AAB hybrid entries, subcluster I contained also an ABB genotype Namwa Khom (Pisang Awak subgroup), as a closest relative of the AAB Figue Pomme Géante accession from the Silk subgroup. Two of the ABB hybrids representatives, Kluai Tiparot and Pelipita formed the third subclade within the B cluster (III). The majority of the ABB hybrids was grouped under IVb, together with an AAB accession Popoulou. The African plantains formed a separate clade IVa with a single AAA representative P. Berangan, as mentioned above.

Blind test with triploid accessions

Six encoded triploid samples were included in the Blind test and all of them were assessed correctly as the closest related species to the corresponding reference genotype from identical subgroup, with significant statistical support (Figure 4). The position of some clades was slightly altered after the inclusion of anonymous samples into the analysis (Figure 4). Specifically, the UPGMA cluster analysis has now shown an altered position of the clade previously labeled with III (ABB accessions Pelipita and Kluai Tiparot) and the subclade of the cluster previously labeled with II, bearing the AAB genotypes P. Palembang, P. Rajah and P. Raja Bulu. However, the bootstrap statistical support for the nodes leading to these clades was not significantly strong in either dataset and the position of all the other clades in the consensus tree remained unchanged.

Identification of duplicate accessions

The 100% similarity in multilocus genotypes was seen in 9 pairs of duplicate accessions (Table 7). Some of the duplicities were introduced into the accession set intentionally from the local greenhouse (originally coming from the ITC collection) to assess the capability of our genotyping system of spotting the duplicate accessions. Others were introduced through the Blind test samples (see M&M). All the duplicates were identified (Table 7), with two exceptions. The *M. textilis* Reference collection DNA sample (ref. 50), that was reported to correspond to the ITC accession ITC 1072, was shown to be identical to another *M. textilis* accession (ITC 0539). This suggests that the reference sample (ref. 50) was either mislabeled or its origin was not reported correctly.

Another anticipated duplicate, introduced into the triploid entries through the Blind test, was the accession blind 12 (Pisang Bakar ITC 1064). Its corresponding reference DNA sample was ref. 19; however, their identity

based on the multilocus molecular profile was not approved. Although the two samples differed at 7 out of the 19 scored SSR loci, their closest relationship was revealed after the UPGMA cluster analysis (Figure 4), suggesting that their mutual subgroup classification (subgr. Ambon) may be correct, but the identity of one of the samples was confused.

Moreover, more than one duplicate accession were reported for both, accession ref. 8 (*M. acuminata* ssp. *burmannicoides* 'Calcutta4') and ref. 21 (*M. balbisiana* 'Tani'). The second duplicate for each of the two reference samples were classified under the same species/subspecies (Table 7). This indicates that either the marker set used had not enough resolution power to distinguish these accessions or, more likely based on the low P_{ID} value mentioned above, these accessions were mislabeled.

DISCUSSION

While the use of microsatellite markers to analyze genetic diversity among *Musa* species is well documented (e.g. Kaemmer *et al.* 1997; Grapin *et al.* 1998; Buhariwalla *et al.* 2005; Ning *et al.* 2007; Venkatachalam *et al.* 2008; Wang *et al.* 2010), its application in a form of a standardized platform to serve for genotyping purposes for the wider *Musa* community is still missing. In this study, we attempted to develop an optimized SSR-based system for molecular characterization of *Musa* accessions that could be used as the basis for the foundation of the Musa Genotyping Centre.

Mislabeled accessions and sample duplications are common problems in germplasm collections (e.g. Virk *et al.* 1995; Zhang *et al.* 2006). The resolution of the marker set tested in this study was high enough ($P_{ID} = 9.44 \times 10^{-29}$) to distinguish between different accessions and proved to be powerful enough to identify mislabeled accessions, as documented in case of *M. acuminata* ssp. *malaccensis* accession. Similarly, its potential for identifying duplicates was clearly approved on the present dataset. Nevertheless, we wanted to assure the reproducibility of results and genotyping errors minimization prior to its implementation into practice. When compared to the original data reported by Lagoda *et al.* (1998) for a subset of the markers, the allele size ranges were overlapping, but not identical. Similar problems were described previously and most often they could be attributed to the method used and the conditions of electrophoretic separation (e.g. Testolin *et al.* 2000; Creste *et al.* 2003). Also, the automatic capillary electrophoresis system used in this experiment allows for much higher resolution and run-to-run precision than the previously used gel-based systems. Therefore, the wider range of allele sizes and higher numbers of identified alleles are adding to the resolution power of the marker set, rather than infirming the capability of the platform.

Among the common genotyping errors that are responsible for misidentification of a particular genotype, allele dropout and false allele amplification play an important role. Allele dropout is an accidental failure of PCR to amplify one of the alleles present at the heterozygous locus, which produces false homozygous patterns (Pompanon *et al.* 2005). To deal with

this problem, three options have been proposed. The first relies on systematic replication of the genotyping, i.e. multitube approach, which in most cases would expose the underlying allelic dropouts or allele shifts due to poor amplification (Taberlet *et al.* 1996). Another possibility is to allow for a certain level of mismatch tolerance, providing that enough loci are scored. Then based on the multilocus genotype, the differences generated by genotyping errors can be distinguished by the low number of mismatches from those that are actual differences between two genotypes (McKelvey and Schwartz 2004). The third option combines the two former ones, with replicated genotyping only for the samples where 3 or fewer mismatches at different loci were observed. These multilocus genotypes are reevaluated after the repeated typing to prove that they are different genotypes in reality, but the cost increase by PCR replications is minimized (Zhang *et al.* 2006). In this pilot study, we adopted the multitube approach with three replicates to assure the maximum precision. However, with many more samples coming to be analyzed in the Musa Genotyping Centre, and thereby increasing the reference database of molecular profiles, the third (combined) option appears to be adequate and is currently being tested.

The grouping revealed by the UPGMA cluster analysis was consistent with the characterization based on the morphotaxonomic classification of the accessions (Figure 2 and 4). The Callimusa section, however, did not form a separate cluster, which reflects its controversial position and agrees with its previously reported close relationship to the Australimusa species (Wong *et al.* 2001b; Wong *et al.* 2002; Jarret and Gawel 1995). Also, the close relation between *Rhodochlamys* and *M. acuminata* species (Wong *et al.* 2002; Bartoš *et al.* 2005; Li *et al.* 2010; Liu *et al.* 2010) was confirmed. The marker set enabled distinction to the level of individual subgroup/subspecies. The degree of polymorphism varied between subgroups and subspecies and polymorphic sites were still to be found within the subgroups and subspecies. For example, in contrast to the study of Creste *et al.* (2003) who were not able to find polymorphic loci among the Cavendish subgroup of bananas in their study based on 6 SSR loci, the marker set used in our study did provide polymorphic loci among the three representatives of Cavendish subgroup allowing for their distinction. Obviously, the larger number of loci scored increases the possibility to find enough polymorphic loci. On the other hand, the limitations in the resolution of microsatellite markers become evident when somatic mutants are analyzed; as they share the common origin and the genetic variation that is narrowed through the cycles of vegetative propagation may not be reflected in the SSR molecular profile (Cipriani *et al.* 1994; Creste *et al.* 2003; Esselink *et al.* 2003). As most of the commercial banana cultivars are vegetatively propagated clones, the assessment of their genetic variability through the marker set tested in this study may not be successful and is yet to be confirmed. However, it still presents a very useful platform for molecular characterization of unknown samples and assessment of the genetic integrity of the *Musa* germplasm collections.

Although microsatellites have been used as reliable markers for projects with labor division among laboratories (Bredemeijer *et al.* 2002; Röder *et al.* 2002), several works have shown that there was a significant level of

incongruence between results obtained at different workplaces, thus complicating the transferability and comparability of the data (Jones *et al.* 1997; Weeks *et al.* 2002; This *et al.* 2004; Van Treuren *et al.* 2010). In the light of this, centralization of genotyping activities in *Musa* and its standardization as a service to the research community appears to be a preferable option. In addition to facile quality control, the core facility would enable using other methods to support the genotyping, such as flow cytometric estimation of ploidy level and/or genome size, keeping in mind that the genotyping data treatment differs for the diploid and polyploid accessions. Obviously, sample transfer requirements can be minimized if both types of analysis are performed at single site. Moreover, with every new sample passing through the analysis, the database of reference SSR profiles is enlarged and the probability of identifying the closest relative or exactly matching accession is enhanced.

Based on our results obtained with the SSR markers presented in this work and those of Hřibová *et al.* (in press) obtained with ITS, as well as the long-term experience in DNA flow cytometry (Doležel 1991; Lysák *et al.* 1999; Roux *et al.* 2003; Bartoš *et al.* 2005; Doleželová *et al.* 2005) the *Musa* Genotyping Centre has been established at the Institute of Experimental Botany in Olomouc (Czech Republic) under the umbrella of Bioversity International. The Centre serves the whole *Musa* research and breeders community. Moreover, the genotyping platform has already been included in the pipeline for characterization of newly introduced accessions to the International banana germplasm collection (ITC). Although it is obvious that new high-content high-throughput genotyping approaches will gradually replace marker-based systems, we feel confident that the platform described here offers a well-founded and ready-to-use approach, which can be applied immediately and which offers higher flexibility in scaling the analysis with respect to the sample size, cost-efficiency and results turn-around time.

CONCLUSIONS and FORWARD LOOK

The platform for genotyping of *Musa* germplasm described here provides a robust and reproducible approach to characterize genetic variability of this important crop, support the management of germplasm collections and help to direct genotype selection for breeding improved cultivars. The database of molecular profiles keeps growing with every new sample passing through the analytical pipeline, resulting in stepwise improvement in the grouping, and consequently increasing a chance of finding an exact match for unknown samples. As a part of the future plans, a batch of tetraploid accessions will be included in the analysis to make it more versatile and satisfying all possible requirements for molecular characterization of the diverse *Musa* gene pool.

ADDITIONAL INFORMATION

The following additional information is available in the on-line version of this article:

Additional file 1: The detailed results of the ITS sequence analysis of the blind sample no. 4 and its putative corresponding reference accession – *M. acuminata* ssp. *malaccensis* (ITC 0250).

SOURCES OF FUNDING

This work has been supported by the Bioversity International (LOA CfL 2009/48 and LoA CfL 2010/58), the Internal Grant Agency of Palacký University, Olomouc, Czech Republic (grant award no. Prf-2010-001) and by the Ministry of Education, Youth and Sports of the Czech Republic and the European Regional Development Fund (Operational Programme Research and Development for Innovations No. CZ.1.05/2.1.00/01.0007).

CONTRIBUTIONS BY THE AUTHORS

All authors have contributed to, read and approved the manuscript.

CONFLICTS OF INTEREST

No conflict of interest.

ACKNOWLEDGEMENTS

We thank our colleague Marie Seifertová, MSc. for her excellent technical assistance. The work was performed under the International Atomic Energy Agency (IAEA) Research Agreement No. 13192: "Molecular Cytogenetic Mapping as a Tool to Characterize Genetic Diversity and Induced Mutants in Banana" as a part of the Agency's Co-ordinated Research Project "Molecular tools for quality improvement in vegetatively propagated crops including banana and cassava".

REFERENCES

- Bartoš J, Alkhimova O, Doleželová M, De Langhe E, Doležel J. 2005.** Nuclear genome size and genomic distribution of ribosomal DNA in *Musa* and *Ensete* (Musaceae): taxonomic implications. *Cytogenetic and Genome Research* **109**: 50-57.
- Bonin A, Bellemain E, Bronken Eidesen P, Pompanon F, Brochmann C, Taberlet P. 2004.** How to track and assess genotyping errors in population genetic studies. *Molecular Ecology* **13**: 3261-3273.
- Bredemeijer GMM, Cooke R, Ganai M, Peeters R, Isaac P, Norordijk Y, Rendell S, Jackson J, Röder MS, Wendehake K, Dijcks M, Amelaine M, Wickaert V, Bertrand L, Vosman B. 2002.** Construction and testing of a microsatellite database containing more than 500 tomato varieties. *Theoretical and Applied Genetics* **105**: 1019-1026.
- Buhariwalla HK, Jarret RL, Jayashree B, Crouch JH, Ortiz R. 2005.** Isolation and characterization of microsatellite markers from *Musa balbisiana*. *Molecular Ecology Notes* **5**: 327-330.
- Cheesman EE. 1947.** Classification of the bananas II. The genus *Musa* L. *Kew Bulletin* **2**: 106-117.
- Cipriani G, Frazza G, Peterlunger E, Testolin R. 1994.** Grapevine fingerprinting using microsatellite repeats. *Vitis* **33**: 211-215.
- Creste S, Neto AT, Silva SD, Figueira A. 2003.** Genetic characterization of banana cultivars (*Musa* spp.) from Brazil using microsatellite markers. *Euphytica* **132**: 259-268.
- Crouch HK, Crouch JH, Jarret RL, Cregan PB, Ortiz R. 1998.** Segregation of microsatellite loci in haploid and diploid gametes of *Musa*. *Crop Science* **38**: 211-217.

- Crouch HK, Crouch JH, Madsen S, Vuylsteke DR, Ortiz R. 2000.** Comparative analysis of phenotypic and genotypic diversity among plantain landraces (*Musa* spp., AAB group). *Theoretical and Applied Genetics* **101**: 1056-1065.
- De Langhe E, Pillay M, Tenkouano A, Swennen R. 2005.** Integrating morphological and molecular taxonomy in *Musa*: the African plantains (*Musa* spp. AAB group). *Plant Systematics and Evolution* **255**: 225-236.
- Doležel J. 1991.** Flow cytometric analysis of nuclear DNA content in higher plants. *Phytochemical Analysis* **2**: 143-154.
- Doleželová M, Doležel J, Van den houwe I, Roux N, Swennen R. 2005.** Focus on the *Musa* collection: Ploidy levels revealed. *InfoMusa* **14**: 34-36.
- Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011.** A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *Plos One*, in press.
- Esselink GD, Smulders MJ, Vosman B. 2003.** Identification of cut rose (*Rosa hybrida*) and rootstock varieties using robust sequence tagged microsatellite site markers. *Theoretical and Applied Genetics* **106**: 277-286.
- Gawel NJ, Jarret RL, Whittemore AP. 1992.** Restriction fragment length polymorphism (RFLP)-based phylogenetic analysis of *Musa*. *Theoretical and Applied Genetics* **84**: 286-290.
- Grabin A, Noyer JL, Carreel F, Dambier D, Baurens FC, Lanaud C, Lagoda P. 1998.** Diploid *Musa acuminata* genetic diversity assayed with sequence-tagged microsatellite sites. *Electrophoresis* **19**: 1374-1380.
- Hayden MJ, Nguyen TM, Waterman A, McMichael GL, Chalmers KJ. 2007.** Application of multiplex-ready PCR for fluorescence-based SSR genotyping in barley and wheat. *Molecular Breeding* **21**: 271-281.
- Hippolyte I, Bakry F, Seguin M, Gardes L, Rivallan R, Risterucci AM, Jenny C, Perrier X, Carreel F, Argout X, Piffanelli P, Khan IA, Miller RNG, Pappas GJ, Mbéguié-A-Mbéguié D, Matsumoto T, De Bernardinis V, Huttner E, Kilian A, Baurens FC, D'Hont A, Cote1 F, Courtois B, Glaszmann JC. 2010.** A saturated SSR/DArT linkage map of *Musa acuminata* addressing genome rearrangements among bananas. *BMC Plant Biology* **10**: 65.
- Hřibová E, Čížková J, Christelová P, Taudien S, de Langhe E, Doležel J. 2011.** The ITS1-5.8S-ITS2 sequence region in the Musaceae: structure, diversity and use in molecular phylogeny. *Plos One*, in press.
- INIBAP. 2006.** Global conservation strategy fo *Musa* (Banana and Plantain). A consultative document prepared by INIBAP with the collaboration of numerous partners in the *Musa* research-and-development community. INIBAP, France.
- Jarret RL, Gawel NJ. 1995.** Molecular markers, genetic diversity and systematics. In: Gowen S, ed. *Bananas and plantains*. London: Chapman and Hall, 67-83.
- Jones CJ, Edwards KJ, Castaglione S, Winfield MO, Sala F, Van de Wiel C, Bredemeijer G, Vosman B, Matthes M, Daly A, Brettschneider R, Bettini P, Buiatti M, Maestri E, Malcevski A, Marmioli N, Aert R, Volckaert G, Rueda J, Linacero R, Vazquez A, Karp A. 1997.** Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. *Molecular Breeding* **3**: 381-390.

- Kaemmer D, Fischer D, Jarret RL, Baurens FC, Grapin A, Dambier D, Noyer JL, Lanaud C, Kahl G, Lagoda PJJ. 1997.** Molecular breeding in the genus *Musa*: a strong case for STMS marker technology. *Euphytica* **96**: 49-63.
- Lagoda PJJ, Noyer JL, Dambier D, Baurens FC, Grapin A, Lanaud C. 1998.** Sequence tagged microsatellite site (STMS) markers in the Musaceae. *Molecular Ecology* **7**: 657-666.
- Li LF, Häkkinen M, Yuan YM, Hao G, Ge XJ. 2010.** Molecular phylogeny and systematics of the banana family (Musaceae) inferred from multiple nuclear and chloroplast DNA fragments, with a special reference to the genus *Musa*. *Molecular Phylogenetics and Evolution* **57**: 1-10.
- Liu AZ, Kress WJ, Li DZ. 2010.** Phylogenetic analyses of the banana family (Musaceae) based on nuclear ribosomal (ITS) and chloroplast (*trnL-F*) evidence. *Taxon* **59**: 20-28.
- Liu K, Muse SV. 2005.** PowerMarker: Integrated analysis environment for genetic marker data. *Bioinformatics* **21**: 2128-2129.
- Loh JP, Kiew R, Set O, Gan LH, Gan YY. 2000.** Amplified fragment length polymorphism fingerprinting of 16 banana cultivars (*Musa* cvs.). *Molecular Phylogenetics and Evolution* **17**: 360-366.
- Lysák MA, Doleželová M, Horry JP, Swennen R, Doležel J. 1999.** Flow cytometric analysis of nuclear DNA content in *Musa*. *Theoretical and Applied Genetics* **98**: 1344-1350.
- McKelvey KS, Schwartz MK. 2004.** Genetic errors associated with population estimation using non-invasive molecular tagging: Problems and new solutions. *Journal of Wildlife Management* **68**: 439-448.
- Michener CD, Sokal RR. 1957.** A quantitative approach to a problem of classification. *Evolution* **11**: 490-499.
- Nei M. 1973.** Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences, USA* **70**: 3321-3323.
- Ning SP, Xu LB, Lu Y, Huang BZ, Ge XJ. 2007.** Genome composition and genetic diversity of *Musa* germplasm from China revealed by PCR-RFLP and SSR markers. *Scientia Horticulturae* **114**: 281-288.
- Nwakanma DC, Pillay M, Okoli BE, Tenkouano A. 2003.** PCR-RFLP of the ribosomal DNA internal transcribed spacers (ITS) provides markers for the A and B genomes in *Musa* L. *Theoretical and Applied Genetics* **108**: 154-159.
- Page RDM. 1996.** TREEVIEW: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* **12**: 357-358.
- Paetkau D, Calvert W, Stirling I, Strobeck C. 1995.** Microsatellite analysis of population structure in Canadian polar bears. *Molecular Ecology* **4**: 347-354.
- Pessoa-Filho M, Beló A, Alcochete AAN, Rangel PHN, Ferreira ME. 2007.** A set of multiplex panels of microsatellite markers for rapid molecular characterization of rice accessions. *BMC Plant Biology* **7**: 23.
- Pillay, M, Nwakanma, DC, Tenkouano A. 2000.** Identification of RAPD markers linked to A and B genome sequences in *Musa* L. *Genome* **43**: 763-767.
- Pillay M, Ogundiwin E, Nwakanma DC, Ude G, Tenkouano A. 2001.** Analysis of genetic diversity and relationships in East African banana germplasm. *Theoretical and Applied Genetics* **102**: 965-970.

Pompanon F, Bonin A, Bellemain E, Taberlet P. 2005. Genotyping errors: Causes, consequences and solutions. *Nature Reviews Genetics* **6**: 847-859.

Risterucci AM, Hippolyte I, Perrier X, Xia L, Caig V, Evers M, Huttner E, Kilian A, Glaszmann JC. 2009. Development and assessment of Diversity Arrays Technology for high-throughput DNA analyses in *Musa*. *Theoretical and Applied Genetics* **119**: 1093-1103.

Röder MS, Wendehake K, Korzun V, Bredemeijer G, Laborie D, Bertrand L, Isaac P, Rendell S, Jackson J, Cooke RJ, Vosman B, Ganai M. 2002. Construction and analysis of a microsatellite- based database of European wheat varieties. *Theoretical and Applied Genetics* **106**: 67-73.

Roux N, Toloza A, Radecki Z, Zapata-Arias FJ, Doležel J. 2003. Rapid detection of aneuploidy in *Musa* using flow cytometry. *Plant Cell Reports* **21**: 483-490.

Ruangstutapha S, Eimert K, Schröder MB, Silayoi B, Denduangboripant J, Kanchanapoom K. 2007. Molecular phylogeny of banana cultivars from Thailand based on HAT-RAPD markers. *Genetic Resources and Crop Evolution* **54**: 1565-1572.

Schuelke M. 2000. An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology* **18**: 233-234.

Simmonds NW, Sheperd K. 1955. Taxonomy and origin of cultivated bananas. *Botanical Journal of the Linnean Society* **55**: 302-312.

Stover RH, Simmonds NW. 1987. *Bananas*, 3rd edn. Burnt Mill, Harlow: Longman Scientific and Technical.

Taberlet P, Griffin S, Goossens B, Questiau S, Manceau V, Escaravage N, Waits LP, Bouvet J. 1996. Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research* **24**: 3189-3194.

Testolin R, Marrazzo T, Cipriani G, Quarta R, Verde I, Dettori MT, Pancaldi M, Sansavini S. 2000. Microsatellite DNA in peach (*Prunus persica* L. Batsch) and its use in fingerprinting and testing the genetic origin of cultivars. *Genome* **43**: 512-520.

This P, Jung A, Boccaccii P, Borrego J, Botta R, Constantini L, Crespan M, Dangl GS, Eisenheld C, Ferreira-Monteiro F, Grando S, Ibañez J, Lacombe T, Laucou V, Magalhães R, Meredith CP, Milani N, Peterlung E, Regner F, Zulini L, Maul E. 2004. Development of a standard set of microsatellite reference alleles for identification of grape cultivars. *Theoretical and Applied Genetics* **109**: 1448-1458.

Ude G, Pillay M, Nwakanma D, Tenkouano A. 2002. Genetic diversity in *Musa acuminata* Colla and *Musa balbisiana* Colla and some of their natural hybrids using AFLP markers. *Theoretical and Applied Genetics* **104**: 1246-1252.

Van Treuren R, Kemp H, Ernsting G, Jongejans B, Houtman H, Visser L. 2010. Microsatellite genotyping of apple (*Malus x domestica* Borkh.) genetic resources in the Netherlands: application in collection management and variety identification. *Genetic Resources and Crop Evolution* **57**: 853-865.

Venkatachalam L, Sreedhar RV, Bhagyalakshmi N. 2008. The use of genetic markers for detecting DNA polymorphism, genotype identification and phylogenetic relationships among banana cultivars. *Molecular Phylogenetics and Evolution* **47**: 974-985.

- Virk PS, Newbury HJ, Jackson TM, Ford-Lloyd BV. 1995.** The identification of duplicate accessions within a rice germplasm collection using RAPD analysis. *Theoretical and Applied Genetics* **90**: 1049-1055.
- Wagner HW, Sefc KM. 1999.** IDENTITY 1.0. Centre for Applied Genetics, University of Agricultural Sciences Vienna.
- Wang XL, Chiang TY, Roux N, Hao G, Ge XJ. 2007.** Genetic diversity of wild banana (*Musa balbisiana* Colla) in China as revealed by AFLP markers. *Genetic Resources and Crop Evolution* **54**: 1125-1132.
- Wang JY, Zheng LS, Huang BZ, Liu WL, Wu YT. 2010.** Development, characterization and variability analysis of microsatellites from a commercial cultivar of *Musa acuminata*. *Genetic Resources and Crop Evolution* **57**: 553-563.
- Website 1.** Global Musa Genomics Consortium. http://www.musagenomics.org/cetest_firstpage1/genomic_dna.html. (28 January 2011).
- Weeks DE, Conley YP, Ferrell RE, Mah TS, Gorin MB. 2002.** A tale of two genotypes: Consistency between two high-throughput genotyping centers. *Genome Research* **12**: 430-435.
- Weising K, Nybom H, Wollf K, Kahl G. 2005.** *DNA Fingerprinting in Plants. Principles, Methods, and Applications*, 2nd edn. London: Taylor & Francis Group.
- Wong C, Kiew R, Loh JP, Gan LH, Set O, Lee SK, Lum S, Gan YY. 2001a.** Genetic diversity of the wild banana *Musa acuminata* Colla in Malaysia as evidenced by AFLP. *Annals of Botany* **88**: 1017-1025.
- Wong C, Kiew R, Ohn S, Lamb A, Lee SK, Gan LH, Gan YY. 2001b.** Sectional placement of three Bornean species of *Musa* (Musaceae) based on AFLP. *Gardens' Bulletin Singapore* **53**: 327-341.
- Wong C, Kiew R, Argent G, Set O, Lee SK, Gan YY. 2002.** Assessment of the validity of the sections in *Musa* (Musaceae) using AFLP. *Annals of Botany* **90**: 231-238.
- Zhang D, Mischke S, Goenaga R, Hemeida AA, Saunders JA. 2006.** Accuracy and reliability of high-throughput microsatellite genotyping for cacao clone identification. *Crop Science* **46**: 2084-2092.

FIGURE LEGENDS

Figure 1. Dendrogram showing the results of the UPGMA analysis of diploid accessions dataset. Bootstrap support values higher than 50% are marked below the corresponding branches. The classification of the genotypes into individual sections, species and subspecies of the genus *Musa* is indicated by the colored side bars and legends. Complete list of accessions with their taxonomic details can be found in Tables 1 and 2.

Figure 2. Dendrogram showing the results of the UPGMA analysis of diploid accessions dataset including the Blind test samples. Bootstrap support values higher than 50% are marked below the corresponding branches. The anonymous samples included in the Blind test are highlighted in red. The classification of the genotypes into individual sections, species and subspecies of the genus *Musa* is indicated by the colored side bars and

legends. Complete list of accessions with their taxonomic details can be found in Tables 1, 2 and 4.

Figure 3. Dendrogram showing the results of the UPGMA analysis of triploid accessions dataset. Bootstrap support values higher than 50% are marked below the corresponding branches. Complete list of accessions with their taxonomic details can be found in Tables 1 and 2.

Figure 4. Dendrogram showing the results of the UPGMA analysis of the triploid accessions dataset including the Blind test samples. Bootstrap support values higher than 50% are marked below the corresponding branches. The anonymous samples included in the Blind test are highlighted in red. Complete list of accessions with their taxonomic details can be found in Tables 1, 2 and 4.

Figure 1.

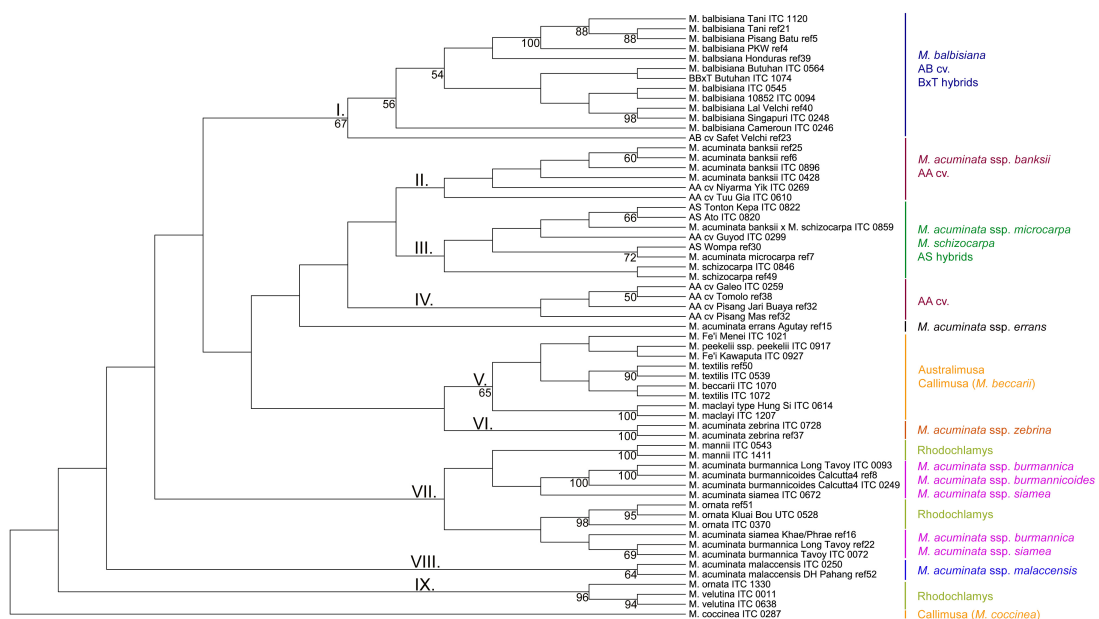


Figure 2.

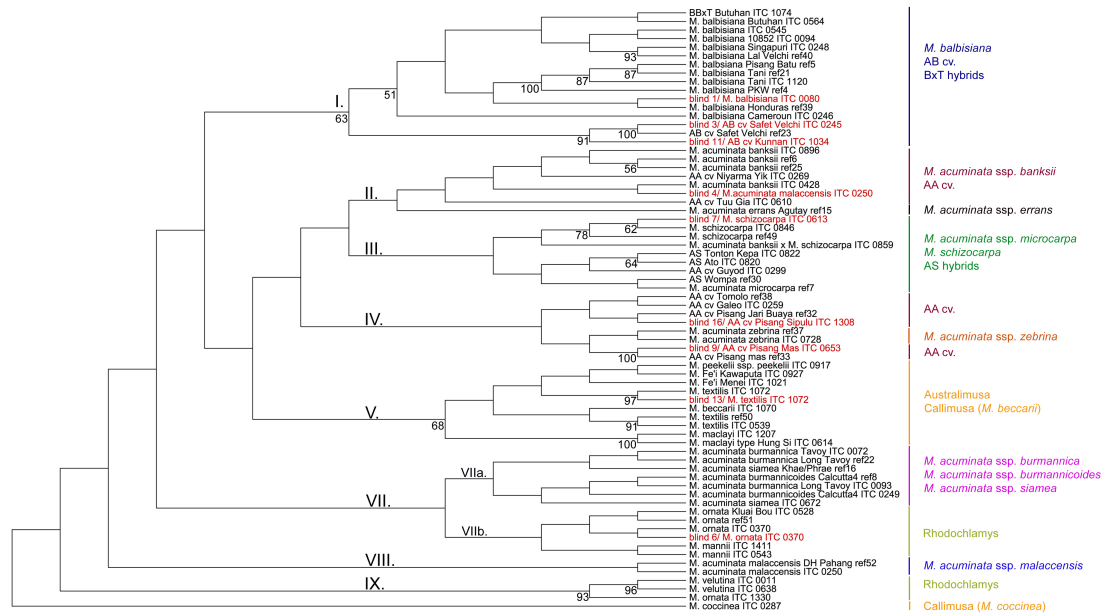


Figure 3.

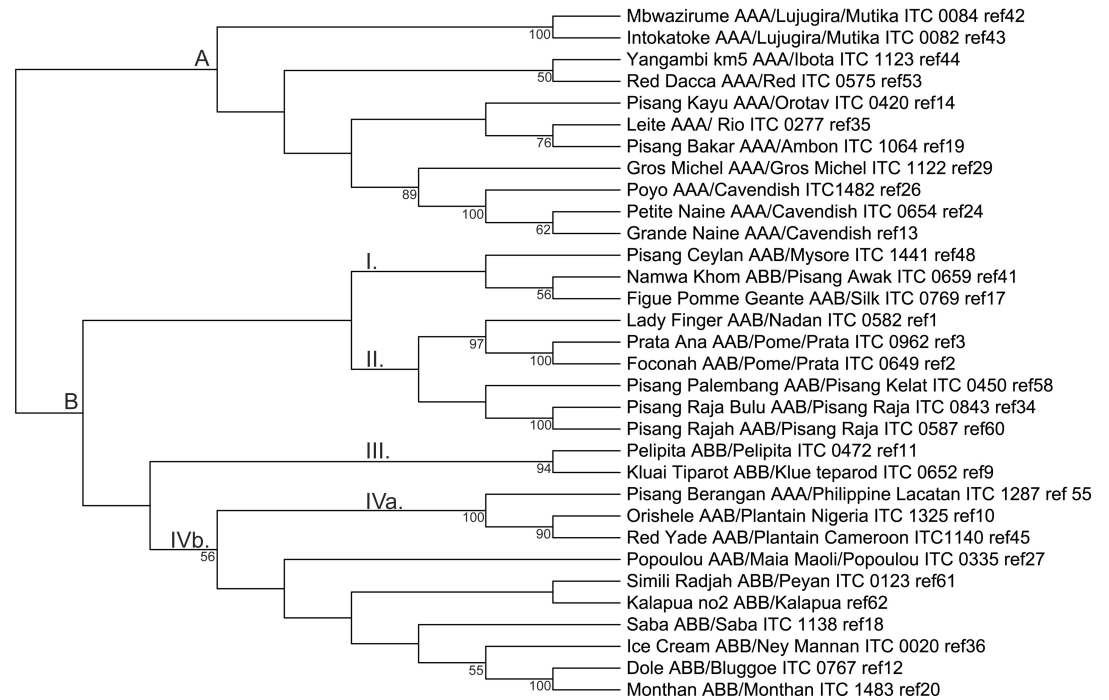


Figure 4.

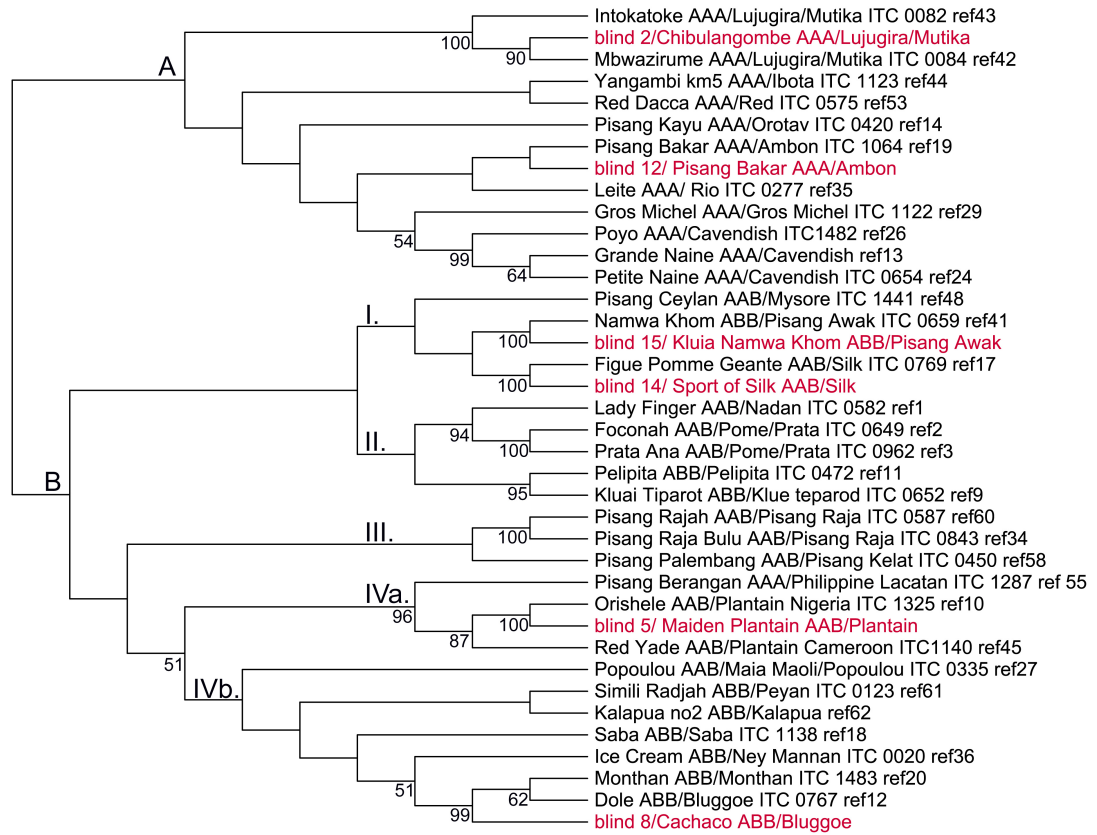


Table 1. Taxonomic details of the Reference DNA Collection accessions.

(http://www.musagenomics.org/cetest_firstpage1/genomic_dna.html)

Ref. sample number	Accession name	Genus	Section	Species (group)	Subspecies (subgroup)	DNA Availability	Corresponding ITC accessions
1	Lady Finger	Musa	Eumusa	AAB	Nadan Pome /	Yes	ITC 0582
2	Foconah	Musa	Eumusa	AAB	Prata Pome /	Yes	ITC 0649
3	Prata Ana Pisang Klutuk	Musa	Eumusa	AAB	Prata	Yes	ITC 0962
4	Wulung	Musa	Eumusa	balbisiana	type 4	Yes	ITC 1063
5	Pisang Batu	Musa	Eumusa	balbisiana	type 4	Yes	ITC 1156
6	Banksii	Musa	Eumusa	acuminata	banksii	Yes	ITC 0623
7	Borneo	Musa	Eumusa	acuminata	microcarpa burmannicoi	Yes	ITC 0253
8	Calcutta 4	Musa	Eumusa	acuminata	des	Yes	ITC 0249
9	Kluai Tiparot	Musa	Eumusa	ABB	Klue teparod	Yes	ITC 0652
10	Orishele	Musa	Eumusa	AAB	Plantain	Yes	ITC 1325
11	Pelipita	Musa	Eumusa	ABB	Pelipita	Yes	ITC 0472
12	Dole Grande	Musa	Eumusa	ABB	Bluggoe	Yes	ITC 0767
13	Naine	Musa	Eumusa	AAA	Cavendish	Yes	--
14	Pisang Kayu	Musa	Eumusa	AAA	Orotav	Yes	ITC 0420
15	Agutay	Musa	Eumusa	acuminata	errans	Yes	ITC 1028
16	Khae (Phrae) Figue Pomme Géante	Musa	Eumusa	acuminata	siamea	Yes	ITC 0660
17	Saba	Musa	Eumusa	AAB	Silk	Yes	ITC 0769
18	Saba	Musa	Eumusa	ABB	Saba	Yes	ITC 1138
19	Pisang bakar	Musa	Eumusa	AAA	Ambon	Yes	ITC 1064
20	Monthan	Musa	Eumusa	ABB	Monthan	Yes	ITC 1483
21	Tani	Musa	Eumusa	balbisiana		Yes	ITC 1120
22	Long Tavoy	Musa	Eumusa	acuminata	burmannica	Yes	ITC 0283
23	Safet Velchi	Musa	Eumusa	AB cv	Ney Poovan	Yes	ITC 0245
24	Petite Naine	Musa	Eumusa	AAA	Cavendish	Yes	ITC 0654
25	Paliama	Musa	Eumusa	acuminata	banksii	Yes	ITC 0766
26	Poyo	Musa	Eumusa	AAA	Cavendish Maia Maoli/	Yes	ITC 1482
27	Popoulou	Musa	Eumusa	AAB	Popoulou	Yes	ITC 0335
28 ^a	Simili Radjah	Musa	Eumusa	ABB	Peyan	Yes	--
29	Gros-Michel	Musa	Eumusa	AAA	Gros Michel	Yes	ITC 1122
30	Wompa	Musa	Eumusa	AS		Yes	ITC 1152
31 ^b	Kunnan Pisang Jari Buaya	Musa	Eumusa	AB cv		Yes	ITC 1034
32	Buaya	Musa	Eumusa	AA cv (18)	jari buaya	Yes	ITC 0312
33	Pisang Mas Pisang Raja	Musa	Eumusa	AA cv (2)	Sucrier	Yes	ITC 0653
34	Bulu	Musa	Eumusa	AAB	Pisang raja	Yes	ITC 0843
35	Leite	Musa	Eumusa	AAA	Rio	Yes	ITC 0277
36	Ice Cream	Musa	Eumusa	ABB	Ney Mannan	Yes	ITC 0020
37	Zebrina	Musa	Eumusa	acuminata	zebrina	Yes	ITC 1177
38	Tomolo	Musa	Eumusa	AA cv	Cooking AA	Yes	ITC 1187
39	Honduras	Musa	Eumusa	balbisiana	type 1	Yes	ITC 0247
40	Lal Velchi Namwa	Musa	Eumusa	balbisiana	type 3	Yes	no
41	Khom	Musa	Eumusa	ABB	Pisang Awak Lujugira/Muti	Yes	ITC 0659
42	Mbwazirume	Musa	Eumusa	AAA	ka	Yes	ITC 0084

Ref. sample number	Accession name	Genus	Section	Species (group)	Subspecies (subgroup)	DNA Availability	Corresponding ITC accessions
43	Intokatoke Yangambi	Musa	Eumusa	AAA	Lujugira/Mutika	Yes	ITC 0082
44	km5	Musa	Eumusa	AAA	Ibota	Yes	ITC 1123
45	Red Yade	Musa	Eumusa	AAB	Plantain Nendra	Yes	ITC 1140
46 ^b	Pisang Rajah	Musa	Eumusa	AAB	padaththi	Yes	ITC 0243
47 ^c	Yawa 2 Pisang	Musa	Eumusa x Australimusa	ABBT	ABBT	Yes	ITC 1238
48	Ceylan	Musa	Eumusa	AAB	Ssp/sgr 501	Yes	ITC 1238
49	Schizocarpa	Musa	Eumusa	schizocarpa	Mysore	Yes	ITC 1441
50	Textilis	Musa	Eumusa	textilis		Yes	ITC 0856
51	Musa ornata CIRAD 930 (DH Pahang)	Musa	Australimusa Rhodochlamys	ornata	textilis	Yes	ITC 1072
52	Red Dacca Khai Thong	Musa	Eumusa	acuminata	ornata	Yes	ITC 0637
53	Ruang	Musa	Eumusa	AAA	malaccensis	Yes	ITC 1511
54 ^b	Pisang Berangan	Musa	Eumusa	AAA	Red	Yes	ITC 0575
55	Obino l'Ewai	Musa	Eumusa	AAB	Ibota Philippine Lacatan/Sgr. 555	Yes	ITC 0662
56 ^b	Ihitisim Pisang Palembang	Musa	Eumusa	AAB	Plantain-French	Yes	ITC 1287
57 ^b	Uzakan	Musa	Eumusa	AAB	Plantain-Horn	Yes	ITC 0109
58	Pisang Rajah	Musa	Eumusa	AAB	Pisang Kelat	Yes	ITC 0121
59 ^d	Simili Radjah	Musa	Eumusa	AAB	Iholena	No	ITC 0450
60	Kalapua No2	Musa	Eumusa	ABB	Pisang raja	Yes	ITC 0825
61	Igitsiri	Musa	Eumusa	AAA	Peyan	Yes	ITC 0587
62	Blue Java	Musa	Eumusa	ABB	Kalapua Lujugira/Mutika	Yes	ITC 0123
63 ^d	Pisang Lilin	Musa	Eumusa	AA		No	--
64 ^d					Ney Mannan	No	ITC 0081
65 ^d					Pisang Lilin	No	ITC 0361

^a replaced with accession no. 61 and eliminated from the set.

^b excluded from the analysis due to low quality DNA.

^c excluded from the analysis-sole tetraploid representative.

^d DNA not available for distribution, not included in the analysis.

Table 2. List of additional diploid accessions from ITC collection (maintained in local greenhouse) included in the analysis.

ITC code	Accession name	Species (group)	Subspecies (subgroup)	Section
ITC 0917	M. peekelii ssp. peekelii	peekelii	peekelii	Australimusa
ITC 1207	Musa maclayi	maclayi		Australimusa
ITC 0614	Musa maclayi type Hung Si	maclayi		Australimusa
ITC 0927	Kawaputa	Fe'i		Australimusa
ITC 1021	Menei	Fe'i		Australimusa
ITC 1072*	Musa textilis	textilis	textilis	Australimusa
ITC 0539	Musa textilis	textilis	textilis	Australimusa
ITC 1070	Musa beccarii	beccarii	beccarii	Callimusa
ITC 0287	Musa coccinea	coccinea	coccinea	Callimusa
ITC 0610	Tuu Gia	AA cv		Eumusa
ITC 0259	Galeo	AA cv		Eumusa
ITC 0299	Guyod	AA cv		Eumusa
ITC 0269	Niyarma Yik	AA cv		Eumusa
ITC 0249*	Calcutta 4	acuminata	burmannicoides	Eumusa
ITC 0072	Tavoy	acuminata	burmannica	Eumusa
ITC 0093	Long Tavoy	acuminata	burmannica	Eumusa
ITC 0250	Malaccensis	acuminata	malaccensis	Eumusa
ITC 0896	M. acuminata ssp. banksii	acuminata	banksii	Eumusa
ITC 0428	Higa	acuminata	banksii	Eumusa
ITC 0728	Maia Oa	acuminata	zebrina	Eumusa
ITC 0672	Pa (Rayong)	acuminata	siamea	Eumusa
ITC 0564	Butuhan, salient apex	balbisiana		Eumusa
ITC 0094	Balbisiana (10852)	balbisiana		Eumusa
ITC 1120*	Tani	balbisiana		Eumusa
ITC 0248	Singapuri	balbisiana		Eumusa
ITC 0545	Balbisiana	balbisiana		Eumusa
ITC 0246	Cameroun	balbisiana		Eumusa
ITC 1074	Butuhan	balbisiana x textilis		Eumusa x Australimusa
ITC 0846	Schizocarpa	schizocarpa		Eumusa
ITC 0822	Tonton Kepa	AS		Eumusa
ITC 0859	M. acuminata ssp banksii x M. schizocarpa	AxS		Eumusa
ITC 0820	Ato	AS		Eumusa
ITC 0528	Kluai Bou	ornata	ornata	Rhodochlamys
ITC 1330	Musa ornata (red fingers)	ornata	ornata	Rhodochlamys
ITC 0370	Musa ornata	ornata	ornata	Rhodochlamys
ITC 0011	Musa velutina	velutina	velutina	Rhodochlamys
ITC 0638	Musa velutina	velutina	velutina	Rhodochlamys
ITC 0543	Musa mannii H. Wendl	sanguinea	sanguinea	Rhodochlamys
ITC 1411	Musa mannii			Rhodochlamys

* accessions included as duplicates to the Reference DNA collection samples

Table 3. Detailed list of the SSR markers used in the study.

Marker	Fluorophore	Motif	Reference	Accession GenBank	Annealing temperature (this study; °C)	Minimum allele (this study; bp)	Maximum allele (this study; bp)
mMaCIR01	6-FAM	(GA)20	Lagoda et al. 1998	X87262	55	241	440
mMaCIR03	6-FAM	(GA)10	Lagoda et al. 1998	X87263	55	111	147
mMaCIR07	NED	(GA)13	Lagoda et al. 1998	X87258	53	136	195
mMaCIR08	VIC	(TC)6N24(TC)7	Lagoda et al. 1998	X87264	55	229	283
mMaCIR13	PET	(GA)16N76(GA)8	Lagoda et al. 1998	X90745	53	268	427
mMaCIR24	PET	(TC)7	Lagoda et al. 1998	Z85972	48	240	291
mMaCIR27*	PET	(GA)9	Lagoda et al. 1998	Z85962	58	232	277
mMaCIR39	VIC	(CA)5GATA(GA)5	Lagoda et al. 1998	Z85970	52	329	390
mMaCIR40	6-FAM	(GA)13	Lagoda et al. 1998	Z85977	54	169	247
mMaCIR45	6-FAM	(TA)4CA(CTCGA)4	Lagoda et al. 1998	Z85968	57	274	318
mMaCIR150	VIC	(CA)10	Hippolyte et al. 2010	AM950440	54	253	376
mMaCIR152	6-FAM	(CTT)18,(CT)17,(CA)6	Hippolyte et al. 2010	AM950442	54	147	195
mMaCIR164	VIC	(AC)14	Hippolyte et al. 2010	AM950454	55	256	458
mMaCIR195*	VIC	(GA)11,(GA)6	Hippolyte et al. 2010	AM950461	54	262	306
mMaCIR196	NED	(TA)4, (TC)17, (TC)3	Hippolyte et al. 2010	AM950462	55	163	201
mMaCIR214	NED	(AC)7	Hippolyte et al. 2010	AM950480	53	115	238
mMaCIR231	NED	(TC)10	Hippolyte et al. 2010	AM950497	55	236	286
mMaCIR260	PET	(TG)8	Hippolyte et al. 2010	AM950515	55	204	264
mMaCIR264	6-FAM	(CT)17	Hippolyte et al. 2010	AM950519	53	234	383
mMaCIR307	NED	(CA)6 (GA)17AA(GA)8AA(GA)2	Hippolyte et al. 2010	AM950533	54	143	173
Ma-1-32*	NED	(GA)2	Crouch et al. 1998	n/a	58	208	251
Ma-3-90	PET	(CT)11	Crouch et al. 1998	n/a	53	147	191

* excluded from the analysis due to unreproducible amplification.

Table 4. List of encoded accessions included in the Blind test.

Anonymous				
Code	ITC code	Accession name	Species (group)	Subspecies (subgroup)
blind 1	ITC 0080	Balbisiana	balbisiana	
blind 2	ITC 0138	Chibulangombe	AAA	Lujugira / Mutika
blind 3	ITC 0245*	Safet Velchi	ABcv	Ney Poovan
blind 4	ITC 0250	Malaccensis	acuminata	malaccensis
blind 5	ITC 0322	Maiden Plantain	AAB	Plantain
blind 6	ITC 0370*	Musa ornata	ornata	
blind 7	ITC 0613	Schizocarpa no. 1	schizocarpa	
blind 8	ITC 0643	Cachaco	ABB	Bluggoe
blind 9	ITC 0653*	Pisang Mas	AAcv	
blind 11	ITC 1034	Kunnan	ABcv	Ney Poovan
blind 12	ITC 1064	Pisang Bakar	AAA	Ambon
blind 13	ITC 1072*	M. textilis	textilis	
blind 14	ITC 1222	Sport of Silk	AAB	Silk
blind 15	ITC 1304	Kluai Namwa Khom	ABB	Pisang Awak
blind 16	ITC 1308	Pisang Sipulu	AAcv	

* duplicates to the Reference DNA collection samples

Table 5. Allele number, frequency of the major allele, unique genotypes observed, heterozygosity and informativeness (PIC) of the 19 microsatellite loci applied on the dataset of 70 diploid *Musa* accessions.

Marker	Major allele frequency	Number of unique genotypes observed	Allele number	Observed heterozygosity	PIC*
mMaCIR01	0.125	39	26	0.531	0.936
mMaCIR03	0.357	13	7	0.400	0.694
mMaCIR07	0.181	33	21	0.551	0.883
mMaCIR08	0.231	22	12	0.646	0.830
mMaCIR13	0.229	28	19	0.543	0.870
mMaCIR24	0.328	19	15	0.344	0.767
mMaCIR39	0.200	39	20	0.714	0.893
mMaCIR40	0.233	29	23	0.534	0.887
mMaCIR45	0.207	16	8	0.357	0.801
mMaCIR150	0.328	20	15	0.522	0.797
mMaCIR152	0.232	19	11	0.250	0.849
mMaCIR164	0.161	28	22	0.322	0.916
mMaCIR196	0.250	23	13	0.453	0.855
mMaCIR214	0.383	12	7	0.313	0.670
mMaCIR231	0.214	27	14	0.540	0.880
mMaCIR260	0.329	20	14	0.357	0.765
mMaCIR264	0.239	35	24	0.522	0.900
mMaCIR307	0.500	10	6	0.179	0.625
Ma-3-90	0.167	31	15	0.474	0.893
Mean	0.258	24.4	15.4	0.450	0.827

* Polymorphism Information Content

Table 6. Major allele frequency, allele number and informativeness (PIC) of the 19 microsatellite loci applied on the dataset of 38 triploid *Musa* accessions.

Marker	Major allele frequency	Allele number	PIC*
mMaCIR01	0.105	24	0.942
mMaCIR03	0.237	12	0.839
mMaCIR07	0.132	17	0.912
mMaCIR08	0.237	14	0.867
mMaCIR13	0.342	12	0.804
mMaCIR24	0.289	12	0.817
mMaCIR39	0.316	18	0.859
mMaCIR40	0.289	9	0.817
mMaCIR45	0.289	12	0.814
mMaCIR150	0.263	8	0.808
mMaCIR152	0.263	12	0.850
mMaCIR164	0.131	18	0.913
mMaCIR196	0.237	15	0.881
mMaCIR214	0.263	8	0.788
mMaCIR231	0.132	16	0.905
mMaCIR260	0.474	13	0.733
mMaCIR264	0.158	18	0.913
mMaCIR307	0.342	8	0.760
Ma-3-90	0.105	21	0.934
Mean	0.242	14.1	0.850

* Polymorphism Information Content

Table 7. List of duplicates identified among the analyzed genotypes.

Accession name	No. in the Reference DNA collection / corresponding ITC code	Accession label in the UPGMA tree	DNA sample origin	Identical multilocus genotype to accessions (duplicate samples)	DNA sample origin of the duplicates
<i>M. acuminata</i> ssp. <i>burmannicoides</i> 'Calcutta4'	ref. 8 / ITC 0249	<i>M. acuminata burmannicoides</i> Calcutta4 ref8	Reference DNA collection (CIRAD, Guadeloupe)	<i>M. acuminata burmannicoides</i> Calcutta4 ITC 0249 <i>M. acuminata burmannica</i> Long Tavoy ITC 0093	ITC collection / local greenhouse on-site ITC collection / local greenhouse on-site
<i>M. balbisiana</i> 'Tani'	ref. 21 / ITC 1120	<i>M. balbisiana</i> Tani ref21	Reference DNA collection (CIRAD, Guadeloupe)	<i>M. balbisiana</i> Tani ITC 1120 <i>M. balbisiana</i> P. Batu ref. 5	ITC collection/local greenhouse on-site Reference DNA collection (CIRAD, Guadeloupe)
AB cv. Safet Velchi	ref. 23 / ITC 0245	AB cv Safet Velchi ref23	Reference DNA collection (CIRAD, Guadeloupe)	blind 3 / AB cv Safet Velchi ITC 0245	ITC collection / lyophilized leaf tissue for the Blind test
AA cv. Pisang Mas	ref. 33 / ITC 0653	AA cv Pisang Mas ref33	Reference DNA collection (CIRAD, Guadeloupe)	blind 9 / AA cv. Pisang Mas ITC 0653	ITC collection / lyophilized leaf tissue for the Blind test
<i>M. textilis</i>	ref. 50 / ITC 1072	<i>M. textilis</i> ref50	Reference DNA collection (CIRAD, Guadeloupe)	<i>M. textilis</i> ITC 0539	ITC collection / local greenhouse on-site
<i>M. textilis</i>	ITC 1072	<i>M. textilis</i> ITC 1072	ITC collection/local greenhouse on-site	blind 13 / <i>M. textilis</i> ITC 1072	ITC collection / lyophilized leaf tissue for the Blind test
<i>M. ornata</i>	ITC 0370	<i>M. ornata</i> ITC 0370	ITC collection/local greenhouse on-site	blind 6 / <i>M. ornata</i> ITC 0370	ITC collection / lyophilized leaf tissue for the Blind test

PŘÍLOHA IV

The ITS1-5.8S-ITS2 sequence region in the *Musaceae*: structure, diversity and use in molecular phylogeny

Hřibová E, Čížková J, Christelová P, Taudien S, De Langhe E, Doležel J.
Plos One 6: e17863, 2011.

*[podíl doktorandky na práci: klonování a sekvenování ITS lokusu u vybraných zástupců-
experimentální část – 20 %, sepsání publikace – 20%]*

PŘÍLOHA V

Genomic constitution of *Festuca* × *Lolium* hybrids revealed by the DArTFest array

Kopecký D, Bartoš J, Christelová P, Černocho V, Kilian A, Doležel J.

Theoretical and Applied Genetics 122: 355-363, 2011.

[podíl doktorandky na práci: sekvenování DArT markerů, základní úprava dat-experimentální část – 40%, sepsání publikace – 20%]

PŘÍLOHA VI

Genetic mapping of DArT markers in the *Festuca-Lolium* complex and their use freezing tolerance association analysis

Bartoš J, Sandve SR, Kölliker R, Kopecký D, Christelová P, Stočes S, Østrem L,
Larsen A, Kilian A, Rognli OA, Doležel J.

Theoretical and Applied Genetics 122: 1133-1147, 2011.

*[podíl doktorandky na práci: sekvenování DArT markerů, základní úprava dat-experimentální část –
40%, sepsání publikace – 20%]*

PŘÍLOHA VII

Phylogenetic relationships in the *Musaceae* based on nucleotide sequence of the ITS1-5.8S-ITS2 region

Hřibová E, Čížková J, Němcová P, Doležel J.

V: Abstracts of the International Conference “Molecular Mapping & Marker Assisted Selection in Plants”. p. 58. University of Vienna, Vienna, 2008.

(abstrakt posteru)

Phylogenetic Relationships in the Musaceae Based on Nucleotide Sequence of the ITS1-5.8S-ITS2 Region

Eva Hřibová, Jana Čížková, Pavla Němcová, Jaroslav Doležel.

The family Musaceae comprises genera *Musa*, *Ensete* and *Musella*. Traditional taxonomy based on the plant phenotype and chromosome number divided the genus *Musa* into four sections: *Eumusa* ($x=11$), *Rhodochlamys* ($x=11$), *Australimusa* ($x=10$) and *Callimusa* ($x=10$). This classification has been questioned; the taxonomy of the genus *Musa* has never been fully resolved and remains a subject of debate. Recently, the internal transcribed spacer (ITS) region of the 18S-5.8S-26S nuclear rDNA has been utilized as a marker in many phylogenetic studies. In this work we have amplified and sequenced the ITS region of different diploid taxa from the four sections of *Musa*, two species of the genus *Ensete* and one *Musella* species. Nucleotide sequences were edited using BioEdit software and a phylogenetic tree was constructed based on the neighbor joining method. The tree rooted on *Zingiber* spp. supports the genus *Musa* as a monophyletic group that is separated from *Ensete* and *Musella*. The genus *Musa* is divided into two distinct clades – a clade of *Callimusa* and *Australimusa* and a clade of *Eumusa* and *Rhodochlamys*, where A genome is strictly separated from the B genome. Our results provide useful information for the taxonomy of Musaceae and we expect that the ITS sequence data will aid in resolving some relationships at lower taxonomic levels within the genus *Musa*. This work was supported by ASCR (award A600380703), IAEA (Research Contract no. 12230 and Research Agreement no. 13192) and Generation Challenge Programme (LOA INIB 2006/15).

Phylogenetic relationships in family Musaceae based on the nucleotide sequence of ITS1-5.8S-ITS2 region

E. Hřibová, J. Čížková, P. Němcová, J. Doležel

Laboratory of Molecular Cytogenetics and Cytometry, Institute of Experimental Botany, Sokolovská 6, 77200 Olomouc, Czech Republic

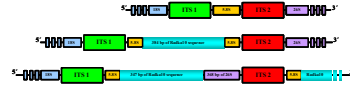
Introduction

- The family Musaceae comprises genera *Musa*, *Ensete* and *Musella*
- Traditional taxonomy based on the plant phenotype and chromosome number has been questioned
- The taxonomy of the genus *Musa* has never been fully resolved
- The 18S-5.8S-26S nuclear rDNA contains the internal transcribed spacers ITS1 and ITS 2

ITS region has been utilized as a marker in phylogenetic studies



Sequence analysis in *M. balbisiana* 'Cameroun' (ITC 0246) revealed variability in the ITS region and insertion of a part of *Radka10* repetitive sequence into the 5.8S rRNA gene



Radka10 contains a sequence homologous to the MUS8 clone carrying repetitive part of cacao swollen shoot badnavirus

Experimental Design

- Genomic DNA isolation ? PCR amplification of the ITS region using specific primers ITS 1 and ITS 4 (Nwakanma *et al.* 2003) ? direct sequencing ? sequence data analysis ? cloning of heterogenous ITS region (TOPO cloning, Invitrogen) ? sequencing of clones ? sequence data analysis (Staden package, BioEdit) ? construction of the phylogenetic tree
- ITS 2 - secondary structure prediction (mfold, Zuker 2003)

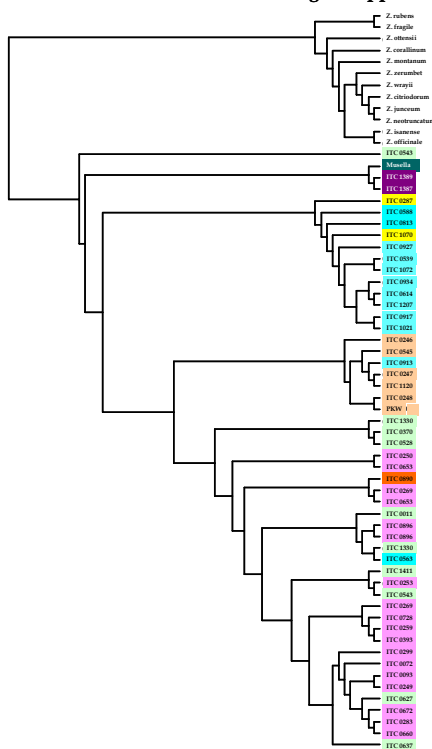
Conclusions

- Phylogenetic tree based on the Neighbour-Joining method was rooted by *Zingiber* (DQ064573-91) and supports the genus *Musa* as monophyletic group that is separated from genus *Ensete* and *Musella*
- Musa* is divided into two distinct clades - clade of *Callimusa* and *Australimusa* and clade of *Eumusa* and *Rhodochlamys*
- The A genome is strictly separated from the B genome
- Sequence analysis revealed that a part of *Radka10* repetitive sequence is inserted into 5.8S rRNA gene of *M. balbisiana* 'Cameroun' (ITC 0246)
- The secondary structures of ITS 2 predicted for different *Eumusa*, *Rhodochlamys* and *Callimusa* species have the constitution characteristic for plants
- The secondary structures of the most of the *Australimusa* species predicted using mfold differ from the other *Musa* species

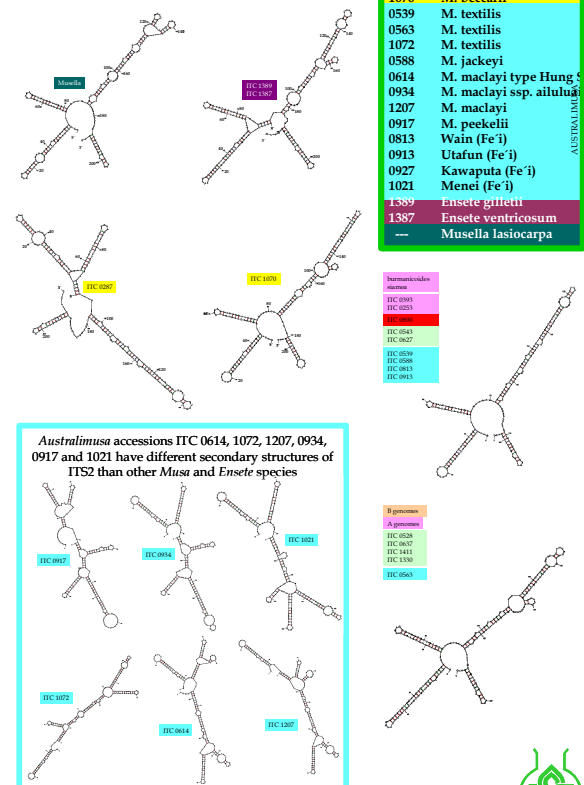
Evaluated accessions

ITC	Accession name
0072	Tavoy
0093	Long Tavoy
0283	Long Tavoy
0249	Calcutta4
0660	Khao (Phrae)
0672	Pa (Rayong)
0250	Malaccensis
0259	Galeo
0253	Borneo
0269	Niyarma Yik
0299	Guyod
0393	Truncata
0653	Pisang Mas
0728	Maia Oa
0896	Banksii
0890	Schtzocarpa
0246	Cameroun
0247	Honduras
0248	Singapuri
0545	M. balbisiana
1120	Tani
---	Pisang Klutug Wulung
0011	M. velutina
0370	M. ornata
0637	M. ornata
1330	M. oranata - red finger
0528	Kluai Bou (ornata)
0543	M. mannii
1411	M. mannii
0627	M. laterita
0287	M. coccinea
1070	M. beccarii
0539	M. textilis
0563	M. textilis
1072	M. textilis
0588	M. jackeyi
0614	M. macleayi type Hung
0934	M. macleayi ssp. ailuana
1207	M. macleayi
0917	M. peskelii
0813	Wain (Fe'i)
0913	Utafun (Fe'i)
0927	Kawaputa (Fe'i)
1021	Menei (Fe'i)
1389	Ensete gittuetii
1387	Ensete ventricosum
---	Musella lasiocarpa

Dendrogram based on Neighbour Joining method of Saitou and Nei, rooted on *Zingiber* spp.



SECONDARY STRUCTURES OF ITS2



PŘÍLOHA VIII

Phylogenetic relationships in family *Musaceae* based on the nucleotide sequence of ITS1-5.8S-ITS2 region

Hřibová E, Němcová P, Čížková J, Schillerová L, Doležel J.

V: Abstracts of the Annual Meeting for the Society of Molecular Biology and Evolution (SMBE). p. 636. SMBE, Barcelona 2008.

(abstrakt posteru)

Phylogenetic relationships in family Musaceae based on the nucleotide sequence of ITS1-5.8S-ITS2 region

Eva Hřibová, Pavla Němcová, Jana Čížková, Lenka Schillerová, Jaroslav Doležel.

Laboratory of Molecular Cytogenetics and Cytometry, Institute of Experimental Botany, Sokolovská 6, 77200 Olomouc, Czech Republic

Traditional taxonomy of banana (*Musa* spp.) is based on the plant phenotype and chromosome number. However, phylogenetic relationships within the genus *Musa* have never been fully resolved. In this study we focused on the ITS1-5.8S-ITS2 region of ribosomal DNA. We have sequenced ITS region of different diploid taxa from four sections of *Musa* genus, two species of genus *Ensete* and one *Musella* species. Phylogenetic tree was constructed based on the neighbor joining method. The tree rooted on *Zingiber* spp. supports the genus *Musa* as monophyletic group that is separated from genus *Ensete* and *Musella*. The genus *Musa* is divided into two distinct clades - clade of *Callimusa* and *Australimusa* and clade of *Eumusa* and *Rhodochlamys*, where the A genome is strictly separated from the B genome. The results based on the ITS sequence data presented in this study provide useful informations on taxonomy of the family *Musaceae*. Moreover, the ITS sequence data are useful to resolve some relationships at lower taxonomic levels within the genus *Musa*. This work was supported by the International Atomic Energy Agency (Contract No. 12230) and the Grant Agency of the Academy of Sciences of the Czech Republic (Contract No. IAA600380703).

Key words: *Musa*, ITS region, molecular phylogenetics

Phylogenetic relationships in family *Musaceae* based on the nucleotide sequence of ITS1-5.8S-ITS2 region

Eva Hřibová, Pavla Němcová, Jana Čížková, Lenka Schillerová, Jaroslav Doležel

Laboratory of Molecular Cytogenetics and Cytometry, Institute of Experimental Botany, Sokolovská 6, 77200 Olomouc, Czech Republic

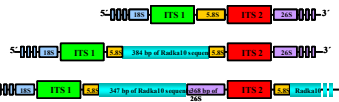
Introduction

- The family *Musaceae* comprises genera *Musa*, *Ensete* and *Musella*
- Traditional taxonomy based on the plant phenotype and chromosome number has been questioned
- The taxonomy of the genus *Musa* has never been fully resolved
- The main problem is the characterization of banana hybrids
- The 18S-5.8S-26S nuclear rDNA contains the internal transcribed spacers ITS1 and ITS 2

ITS region has been utilized as a marker in phylogenetic studies



Sequence analysis in *M. balbisiana* 'Cameroun' (ITC 0246) revealed variability in the ITS region and insertion of a part of *Radka10* repetitive sequence into the 5.8S rRNA gene



Radka10 contains a sequence homologous to the MUSA8 clone carrying repetitive part of cacao swollen shoot badnavirus

Experimental Design

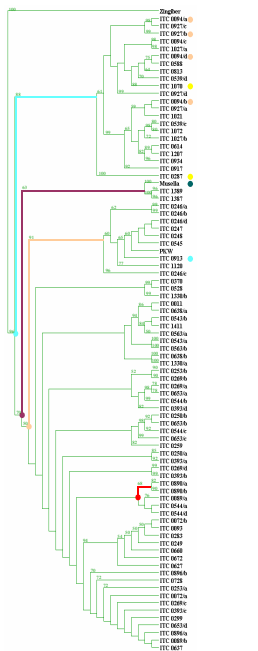
- Genomic DNA isolation ? PCR amplification of the ITS region using specific primers ITS L and ITS 4 (Nwakanma *et al.*, 2003) ? direct sequencing ? sequence data analysis ? cloning of heterogenous ITS region (TOPO cloning, Invitrogen) ? sequencing of clones ? sequence data analysis (ClustalW, BioEdit) ? construction of the phylogenetic tree based on ITS1 and ITS2 spacer (Splits Tree)

Conclusions

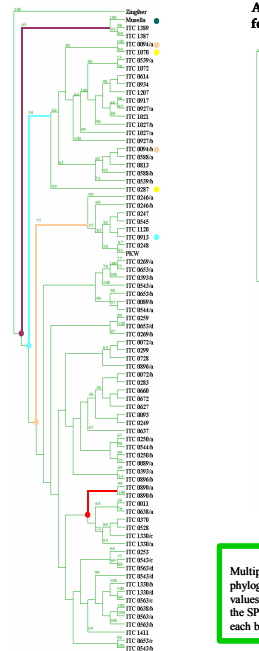
- Phylogenetic tree based on the Neighbour-Joining method was rooted by *Zingiber officinale* (DQ064590)
- Analysis based on ITS2 spacer supports the genus *Musa* as monophyletic group that is separated from genus *Ensete* and *Musella*
- Clade of *Australimusa* contains also *Callimusa* species and clade of *Eumusa* contains *Rhodochlamys* species
- The B genome is strictly separated from the other *Eumusa* and *Rhodochlamys* species
- Sequence analysis revealed that a part of *Radka10* repetitive sequence is inserted into 5.8S rRNA gene of *M. balbisiana* 'Cameroun' (ITC 0246)
- Sequence analysis based on ITS1 and ITS2 spacers is useful for characterization of different genomes in hybrid banana clones

RESULTS

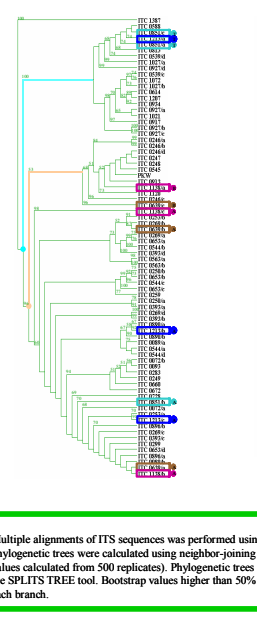
NJ tree of ITS1 rooted by *Zingiber*



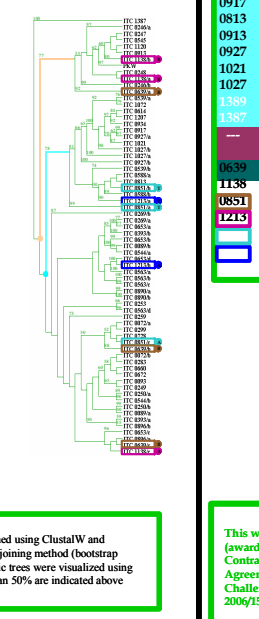
NJ tree of ITS2 rooted by *Zingiber*



NJ tree of ITS1 assembled for selected A, B and T genomes and four hybrid clones



NJ tree of ITS2 assembled for selected A, B and T genomes and four hybrid clones



Multiple alignments of ITS sequences was performed using ClustalW and phylogenetic trees were calculated using neighbor-joining method (bootstrap values calculated from 500 replicates). Phylogenetic trees were visualized using the SPLITS TREE tool. Bootstrap values higher than 50% are indicated above each branch.

Evaluated accessions

ITC	Accession name
0072	Tavoy
0093	Long Tavoy
0283	Long Tavoy
0249	Calcutta4
0660	Khae (Phrae)
0672	Pa (Rayong)
0250	Malaccensis
0259	Galeo
0253	Borneo
0269	Niyarma Yik
0299	Guyod
0393	Truncata
0653	Pisang Mas
0728	Maia Oa
0896	Banksii
0089	M. paradisiaca
0544	M. paradisiaca
0890	Schizocarpa
0094	Balbisiana (10852)
0246	Cameroun
0247	Honduras
0248	Singapuri
0545	M. balbisiana
1120	Tani
—	Pisang Klutug Wulung
0011	M. velutina
0638	M. velutina
0370	M. ornata
0637	M. ornata
1330	M. ornata – red finger
0528	Kluai Bou (ornata)
0543	M. manni
1411	M. manni
0627	M. laterita
0287	M. coccinea
1070	M. beccarii
0539	M. textilis
0563	M. textilis
1072	M. textilis
0588	M. jackeyi
0614	M. maclayi type Hung Si
0934	M. maclayi ssp. ailuluai
1207	M. maclayi
0917	M. peekelii
0813	Wain (Fe' i)
0913	Utafun (Fe' i)
0927	Kawaputa (Fe' i)
1021	Menei (Fe' i)
1027	Asupina (Fe' i)
1389	Ensete gillettii
1387	Ensete ventricosum
—	Musella lasiocarpa
0639	Maritú (AAB)
1138	Saba (ABB)
0851	Karoina (AAxT)
1243	Sar (AAxT)



This work was supported by ASCR (award A60380703), IAEA (Research Contract no. 12230 and Research Agreement no. 13192) and Generation Challenge Programme (LOA INIB 2006/15).

PŘÍLOHA IX

Nucleotide sequence of the ITS1-5.8S-ITS2 region and DArT markers shed light on phylogenetic relationships within the *Musaceae*

Hřibová E, Němcová P, Čížková J, Schillerová L, Kilian A, Doležel J.

V: Abstracts of the International Conference on “Polyploidy, Hybridization and Biodiversity”. p. 143. INRA, Saint Malo, 2009.

(abstrakt posteru)

Nucleotide Sequence of the ITS1-5.8S-ITS2 Region and DArT Markers Shed Light on Phylogenetic Relationships within the *Musaceae*

E. Hřibová¹, P. Němcová¹, J. Čížková¹, L. Schillerová¹, A. Kilian² and J. Doležel¹

¹Laboratory of Molecular Cytogenetics and Cytometry, Institute of Experimental Botany, Olomouc, Czech Republic

²Diversity Arrays Technology, Canberra, Australia

Email: hribova@ueb.cas.cz

Bananas and plantains (*Musa* spp.) provide a staple food for millions of people in sub-Saharan Africa, South and Central America and much of Asia. All cultivars that are grown are seed sterile clones and are thought to have been selected as naturally occurring hybrids in Southeast Asia by the earliest of farmers. In fact, many believe that banana was one of the first crops to be domesticated by man. Almost all of the 300 or more cultivars that are known arose from two seeded, diploid species, *Musa acuminata* Colla and *M. balbisiana* Colla; they are diploid, triploid and tetraploid hybrids among subspecies of *M. acuminata*, and between *M. acuminata* and *M. balbisiana*. Despite the socioeconomic importance of bananas and plantains, the exact mode of their origin and phylogenetic relationships within the family Musaceae remain subjects of debate. In this study we focused on the ITS1-5.8S-ITS2 region of the ribosomal DNA locus. We have sequenced ITS region of different diploid taxa from four sections of *Musa*, two species of genus *Ensete* and one species of *Musella*. Phylogenetic tree was constructed based on the neighbour joining method. The tree, which was rooted on *Zingiber* spp., supports the genus *Musa* as monophyletic group that is separated from genus *Ensete* and *Musella*. The genus *Musa* is divided into two distinct clades – a clade of *Callimusa* and *Australimusa* and a clade of *Eumusa* and *Rhodochlamys*, where the A genome is strictly separated from the B genome. The results based on the ITS sequence data presented in this study provide useful information on taxonomy of the family *Musaceae*. Moreover, the ITS sequence data are useful to resolve some relationships at lower taxonomic levels within the genus *Musa*. Our conclusions derived from the study of ITS nucleotide sequences were supported using a different phylogenetic strategy based on DArT markers to avoid a phylogenetic reconstruction based only on one locus. This work has been supported by the Grant Agency of the Academy of Sciences of the Czech Republic (grant award no. IAA600380703) and the International Atomic Energy Agency (research agreement no. 13192).

Keywords: ITS, ribosomal DNA locus, taxonomy

Nucleotide Sequence of the ITS1-5.8S-ITS1 Region and DArT Markers Shed Light on Phylogenetic Relationships within the *Musaceae*

E. Hřibová¹, P. Němcová¹, J. Čížková¹, L. Schillerová¹, A. Kilian², J. Doležel¹

¹Laboratory of Molecular Cytogenetics and Cytometry, Institute of Experimental Botany, Sokolovská 6, 77200 Olomouc, Czech Republic

²Diversity Arrays Technology, Canberra, Australia

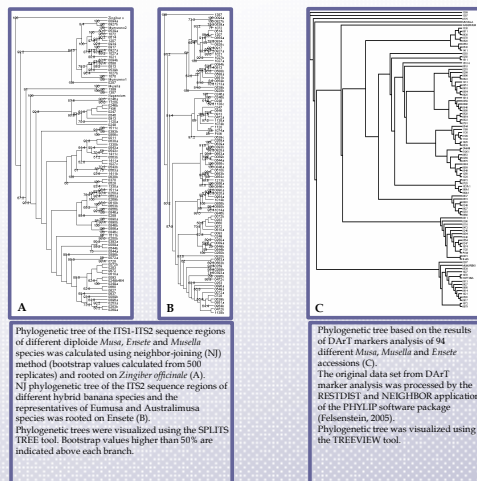
Introduction

- The family *Musaceae* comprises genera *Musa*, *Ensete* and *Musella*
- Traditional taxonomy based on the plant phenotype and chromosome number has been questioned
- The taxonomy of the genus *Musa* has never been fully resolved
- The main problem is the characterization of banana hybrids
- The 18S-5.8S-26S nuclear rDNA contains the internal transcribed spacers ITS1 and ITS2
- ITS region has been utilized as a marker in phylogenetic studies
- Banana DArT chip was prepared from XX banana species and provided by A. Kilian
- DArT markers are derived from methyl-filtered genomic DNA

Experimental Design

- Genomic DNA isolation ? PCR amplification of the ITS region using specific primers ITS L and ITS 4 (Nwakanma *et al.*, 2003) ? direct sequencing ? sequence data analysis ? cloning of heterogenous ITS region (TOPO cloning, Invitrogen) ? sequencing of clones ? sequence data analysis (Staden Package, Clustalw) ? construction of the phylogenetic tree based on ITS1-5.8S-ITS2 region (Phylip, ATV Forester)
- Comparison of the nucleotide sequence of the ITS region of *M. acuminata* 'Calcutta4' obtained using standart protocol (PCR amplification and direct Sanger sequencing), BAC clone sequencing (several ITS region obtained from Sanger sequencing of MA4BAC clone carry 45S rDNA units) and 454 sequencing of *M. acuminata* 'Calcutta4' genomic DNA (based on pyrosequencing, provided by GS FLX system)

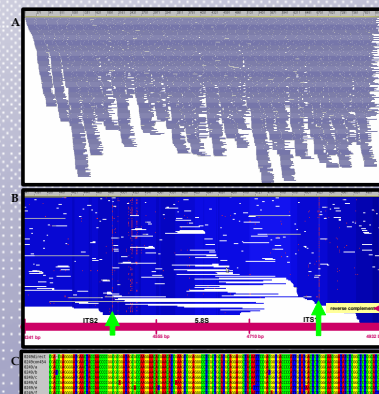
Results



Conclusions

- Phylogenetic tree of the ITS1-ITS2 region based on the Neighbour-joining method was rooted by *Zingiber officinale* (DQ064590) and supports the genus *Musa* as monophyletic group that is separated from genus *Ensete* and *Musella*
- Clade of *Australimusa* contains also *Callimusa* species and clade of *Eumusa* contains *Rhodochlamys* species
- The B genome is strictly separated from the other *Eumusa* and *Rhodochlamys* species
- Sequence analysis based on ITS2 spacer is useful for characterization of genome constitution in hybrid banana clones
- Phylogenetic tree based on the DArT markers is in agreement with the results obtained using ITS nucleotide sequences
- DArT markers are useful for classification of different diploid banana species
- The 454 sequencing of *M. acuminata* 'Calcutta4' nuclear genome provided useful data for analysis of structure and heterogeneity of the rDNA locus
- The 454 data showed that ITS1-5.8S-ITS2 region of heterozygous banana clone *M. acuminata* 'Calcutta4' was highly conserved
- The nucleotide sequence of ITS1-5.8S-ITS2 region of 'Calcutta4' obtained from direct sequencing and 454 sequencing differ only in one mismatch (see below)

Comparison of the nucleotide sequence of the ITS region of *M. acuminata* 'Calcutta4'



- ITS1-5.8S-ITS2 region of *M. acuminata* 'Calcutta4' was obtained by direct sequencing from genomic DNA - products of four PCR reaction were sequenced using dideoxy sequencing and resulted in the same nucleotide sequences
- In addition, six ITS sequences were obtained from BAC clone sequencing
- 3835 sequencing reads provided by 454 sequencing were assembled into 7.553 bp long contig of 45S rRNA gene surrounded by partial IGS with average read depth of 128 reads (figure A)
- The entire ITS1-5.8S-ITS2 region contains only two sites with mismatches in the nucleotide sequence - labeled by green arrows in the figure B
- Spacer ITS1 contains one site with mismatch: A-G, at this position, out of 129 total reads, 82times adenine vs. 47times guanine present.
- Spacer ITS2 contains the second site with mismatch: G-A, at this position, out of the total 124 reads, 68times guanine vs. 56 times adenine present.
- Also 454 data showed that this region is highly conserved in the *M. acuminata* 'Calcutta4'
- The ITS sequence region obtained from direct sequencing and ITS region from 454 data differ only at one position (figure C) in contrast with the ITS regions obtained from BAC clone sequencing
- ITS regions obtained from BAC sequencing have been found more heterogenous (XX nucleotide changes were found between the ITS region obtained from BAC sequencing)

Evaluated accessions

ITC	Accession name	
0072	Tavoy	
0093	Long Tavoy	
0283	Long Tavoy	
0249	Calcutta4	
0660	Khae (Phrae)	
0672	Pa (Rayong)	
0250	Malaccensis	
0259	Galeo	
0253	Borneo	
0269	Niyarma Yik	
0393	Truncata	
0653	Pisang Mas	
0728	Maia Oa	
0610	Tuu Gia	
0896	Banksii	
0089	<i>M. paradisiaca</i>	
0544	<i>M. paradisiaca</i>	
0890	Schizocarpa	
0846	Schizocarpa	
0856	Schizocarpa	
0094	Balbisiana (10852)	
0246	Cameroun	
0247	Honduras	
0248	Singapuri	
0545	<i>M. balbisiana</i>	
1120	Tani	
—	Pisang Klutug Wulung	
0011	<i>M. velutina</i>	
0638	<i>M. velutina</i>	
0370	<i>M. ornata</i>	
0637	<i>M. ornata</i>	
1330	<i>M. ornata</i> - red finger	
0528	Kluai Bou (ornata)	
0543	<i>M. mannii</i>	
1411	<i>M. mannii</i>	
0627	<i>M. laterita</i>	
0287	<i>M. coccinea</i>	
1070	<i>M. beccarii</i>	
0539	<i>M. textilis</i>	
0563	<i>M. textilis</i>	
1072	<i>M. textilis</i>	
0588	<i>M. jackeyi</i>	
0614	<i>M. maclayi</i> type Hung Si	
0934	<i>M. maclayi</i> ssp. ailuluai	
1207	<i>M. maclayi</i>	
0917	<i>M. peekelii</i>	
0813	Wain (Fe' i)	
0913	Utafun (Fe' i)	
0927	Kawaputa (Fe' i)	
1021	Menei (Fe' i)	
1027	Asupina (Fe' i)	
1389	<i>Ensete gillettii</i>	
1387	<i>Ensete ventricosum</i>	
—	<i>Musella lasiocarpa</i>	
0639	Maritu (AAB)	
1138	Saba (ABB)	
0472	Pelipita (ABB)	
0851	Karoina (AAxT)	
1213	Sar (AAxT)	
0854	Umbubu (AAxT)	
0928	Kabulupusa (AAxT)	
1074	Butuhan (BBxT)	
0820	Ato (AS)	
0954	Ungota (AS)	
1014	? AAxSS	
0549	Hochuchu (AAA)	

This work has been supported by the Grant Agency of the Academy of Sciences of the Czech Republic grant award no. IAA600380703 and the International Atomic Energy Agency (research agreement no. 13192).



PŘÍLOHA X

Phylogenetic relationships in the family *Musaceae* based on the genic sequences, sequence of the ITS1-5.8S-ITS2 region and DArT markers

Němcová P, Hřibová E, Valárik M, Čížková J, Schillerová L, Kilian A, Doležel J.

V: Global Perspectives on Asian Challenges. p. 18. International Society for Horticultural Science/ProMusa, Guangzhou, 2009.

(abstrakt přednášky)

Phylogenetic Relationships in Family *Musaceae* Based on the Genic Sequences, Sequence of the ITS1-5.8S-ITS2 Region and DArT Markers

P. Němcová¹, E. Hřibová¹, M. Valárik¹, J. Čížková¹, L. Schillerová¹, A. Kilian² and J. Doležel¹

¹Laboratory of Molecular Cytogenetics and Cytometry, Institute of Experimental Botany, Olomouc, Czech Republic

²Diversity Arrays Technology, Canberra, Australia

Despite the socio-economic importance of banana (*Musa* spp.), phylogenetic relationships within the family *Musaceae* as well as the classification of subspecies and clones remain subject to debate. Different types of molecular markers have been used until now with various degrees of success and a detailed picture is still lacking. An unexplored opportunity is to compare DNA sequences of a large set of genes across the *Musaceae*. In this work, we searched banana ESTs for homology to single-copy rice genes and hybridized candidates with banana genomic DNA. During the first phase of the project, sequences of 24 candidate genes from 14 carefully selected species (ITC collection, Belgium) representing maximal diversity within the *Musaceae* were analyzed. The sequence analysis confirmed previously reported observations on clear divergence of *Musa*, *Ensete* and *Musella* as monophyletic groups. Moreover, within the *Musa* genus clade, section *Eumusa* clustered with *Rhodochlamys*, clearly separated from the cluster of *Australimusa* and *Callimusa*. These findings are in concordance with the previously reported close relation of these sections. The significance of our results was further supported by comparison with the results of nucleotide sequence analysis of ITS1-5.8S-ITS2 region in selected *Musaceae* species as well as analysis with DArT markers. Although each of the methods used is based on a different principle, data that were collected and analyzed supported the same concept of evolutionary relationship within the *Musaceae* species. We expect that a detailed analysis of the sequence data from higher number of genes and extension to other members of the family will clearly reveal phylogenetic relationships within the *Musaceae* and provide markers for unambiguous identification of subspecies and clones. This work has been supported by Grant Agency of the Academy of Sciences of the Czech Republic (grant award no. IAA600380703).

PŘÍLOHA XI

Použití genových sekvencí pro objasnění fylogenetických vztahů v rámci čeledi *Musaceae*

Němcová P, Valárik M, Hřibová E, Doležel J.

V: Sborník abstrakt z vědecké konference “5. Metodické dny”. p. 77. Bulletin České společnosti experimentální biologie rostlin a Fyziologické sekce Slovenské botanické společnosti. Malenovice, 2009.

(abstrakt posteru)

POUŽITÍ GENOVÝCH SEKVENCÍ PRO OBJASNĚNÍ FYLOGENETICKÝCH VZTAHŮ V RÁMCI ČELEDI *MUSACEAE*

PAVLA NĚMCOVÁ¹, MIROSLAV VALÁRIK¹, EVA HŘIBOVÁ¹, JAROSLAV DOLEŽEL¹

¹ Laboratoř molekulární cytogenetiky a cytometrie, Ústav experimentální botaniky AV ČR v.v.i., Sokolovská 6, 77200 Olomouc, ČR.

E-mail: nemcova@ueb.cas.cz, tel.: 585 205 860

Banánovník (*Musa* spp., čeleď *Musaceae*) patří mezi nejvýznamnější plodiny světa, jako důležitý zdroj obživy mnoha obyvatel zejména v rozvojových zemích a také jako významná vývozní komodita. Navzdory svému společenskému a ekonomickému významu však doposud nebyly plně vysvětleny a popsány fylogenetické vztahy v rámci čeledi *Musaceae* a rovněž zůstává nedořešena otázka klasifikace a charakterizace jednotlivých druhů a poddruhů této čeledi. Různí autoři se ve svých studiích s větším či menším úspěchem zabývali použitím rozličných typů molekulárních markerů pro objasnění fylogeneze banánovníku, nicméně celkový a podrobný přehled zatím nebyl získán. Jedna z možností, která dosud u *Musaceae* nebyla použita, je srovnání sekvencí DNA odvozených z genů s jedinou kopií v genomu pro různé druhy banánovníku. Touto cestou je možné velmi detailně porovnat oblasti s homologními úseky mezi jednotlivými genomy, což by při použití sekvencí s vysokým počtem kopií nebylo možné. Vývoj takového typu markerů byl v této práci proveden podle následující strategie. Databáze dostupných sekvencí EST z čeledi *Musaceae* byla srovnána s genomovou sekvencí rýže, která je pro banánovník nejbližší příbuznou rostlinou se sekvenovaným genomem. Byly vybrány ty EST sekvence, které vykazovaly vysokou homologii k jedinému lokusu v genomu rýže a výskyt intronů v rámci tohoto lokusu, pro zvýšení pravděpodobnosti výskytu druhově specifických mutací v odvozených markerech. Jako počáteční záměr bylo stanoveno získat alespoň jeden kandidátní gen pro každé z ramen rýžových chromozómů, tak aby byla zajištěna co nejvyšší míra pravděpodobnosti náhodné distribuce kandidátních genů v genomu banánovníku. Následně bylo v první fázi projektu vybráno 13 zástupců z čeledi *Musaceae*, tak aby co nejlépe reprezentovali diverzitu v rámci čeledi a jeden outgroup, druh *Strelizia nicolai* z příbuzné čeledi *Strelitziaceae*. Soubor kandidátních genů byl po amplifikaci z genomové DNA vybraných druhů sekvenován. Analýza získaných sekvencí a fylogenetická analýza potvrdily monofyletický původ rodu *Musa*. Dále byla jasně potvrzena blízká příbuznost sekce *Eumusa* a *Rhodochlamys*, stejně jako blízký evoluční původ sekcí *Australimusa* a *Callimusa*, což je v souladu s předešlými zjištěními. V rámci sekce *Eumusa* je navíc patrná samostatná evoluční větev tvořena zástupci A genomu banánovníku, jasně separovaná od zástupců B genomu. Dokončení analýzy by mělo vést k datování divergence jednotlivých rodů čeledi *Musaceae* a sekcí rodu *Musa*. Předpokládáme, že další detailní analýza sekvenačních dat z většího množství genů a rozšíření studie o další druhy poskytne podrobný obraz fylogeneze čeledi *Musaceae* a cenný zdroj markerů pro charakterizaci jednotlivých druhů banánovníku.

Tato práce byla podpořena grantem IAA600380703 Grantové Agentury České republiky.

POUŽITÍ GENOVÝCH SEKVENCÍ PRO OBJASNĚNÍ FYLOGENETICKÝCH VZTAHŮ V ČELEDI MUSACEAE



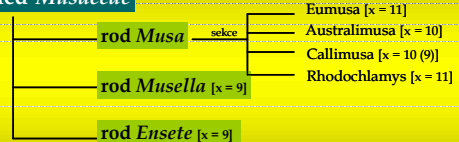
P. Němcová¹, M. Valárik¹, E. Hřibová¹, J. Doležel¹

¹ Laboratoř molekulární cytogenetiky a cytometrie, Ústav experimentální botaniky AV ČR v.v.i., Sokolovská 6, 772 00 Olomouc
e-mail: nencova@ueb.cas.cz, tel.: 585 205 860

Úvod

- Banánovník (čeleď *Musaceae*) patří mezi ekonomicky nejvýznamnější plodiny světa
- Klasifikace jednotlivých druhů je založena na morfologických znacích a základním počtu chromozomů (x)
- Čeleď *Musaceae* obsahuje tři rody: *Musa*, *Ensete* a *Musella*
- Evoluční příbuznost jednotlivých druhů v rámci čeledi nebyla dosud detailně prostudována
- Podrobný popis fylogenetických vztahů v čeledi *Musaceae* napomůže přesnější charakterizaci a určení jednotlivých druhů, popř. poddruhů a kultivovaných klonů banánovníku
- V tomto projektu byla pro studium evoluce čeledi *Musaceae* použita sekvenční analýza genů s nízkým počtem kopií

Čeleď *Musaceae*



Experimentální provedení a výsledky

1. Identifikace a výběr kandidátních genů s nízkým počtem kopií

- EST sekvenční banánovníku jsou srovnány s jednotlivými chromozomy rýže pro identifikaci orthologických genů
- Jako kandidátní gen je zvolen ten, u něž je prokazatelná silná homologie s genem, který se v genomu rýže vyskytuje v jediné kopii
- Kandidátní geny jsou vybrány tak, aby každý z nich byl v genomu rýže lokalizován na jiném chromozomovém raménku. Tímto způsobem se zvýší pravděpodobnost rovnoměrného rozložení lokusů kandidátních genů v genomu banánovníku

2. Design specifických primerů odvozených ze sekvencí EST

- Primery pro amplifikaci kandidátních genů jsou odvozeny ze sekvence EST, jsou tedy komplementární ke kódující oblasti genu
- Na základě předpokládané podobnosti s rýží jsou primery navrženy tak, aby překlenuly oblasti intronů
- Pro amplifikaci kandidátních genů je jako templát použita genomová DNA banánovníku
- Dochází k amplifikaci kandidátního genu společně s přítomnými introny a tím ke zvýšení pravděpodobnosti výskytu většího počtu fylogeneticky informativních molekulárních znaků v sekvenci genu

3. Amplifikace kandidátních genů z vybraných genotypů banánovníku

Rod	Sekce	1TC	Název kultivarů	Druh	Poddruh	Oblast původního výskytu
<i>Musa</i>	<i>Eumusa</i>	0249	<i>capitata</i>	<i>acuminata</i>	<i>burmanica</i>	Myanmar
<i>Musa</i>	<i>Eumusa</i>	0728	<i>Musa</i> sp.	<i>acuminata</i>	Zebona	Havaj
<i>Musa</i>	<i>Eumusa</i>	1100	PKW	<i>babingtonii</i>		Thajsko, Thajland
<i>Musa</i>	<i>Rhodochlamys</i>	0537	<i>Musa ornata</i>	<i>ornata</i>		Thaj, Gámské Myanmar
<i>Musa</i>	<i>Rhodochlamys</i>	0511	<i>Musa velutina</i>	<i>velutina</i>		Japonsko
<i>Musa</i>	<i>Australimusa</i>	0539	<i>Musa textilis</i>	<i>textilis</i>		Filipíny
<i>Musa</i>	<i>Australimusa</i>	0514	<i>M. maculata</i>	<i>maculata</i>		PHG, Malajsko
<i>Musa</i>	<i>Australimusa</i>	1031	<i>Musa</i> sp.	sp.	<i>subcordata</i>	PHG, Malaj, SP
<i>Musa</i>	<i>Callimusa</i>	1079	<i>Musa beccarii</i>	<i>beccarii</i>		Borneo
<i>Musa</i>	<i>Callimusa</i>	0287	<i>Musa coccinea</i>	<i>coccinea</i>		S. Indonésie
<i>Ensete</i>		1387	<i>Ensete ventricosum</i>	<i>ventricosum</i>		Highland E. Africa
<i>Musella</i>			<i>Musella</i>			S. China, B. Burma, Thailand
<i>Strelitzia</i>	Outgroup					
<i>Utricularia</i>						

- V první fázi projektu bylo vybráno 13 druhů tak, aby reprezentovaly diverzitu v rámci čeledi *Musaceae* (viz tabulka)
- Jako outgroup byla použita *Strelitzia nicotai* z příbuzné čeledi *Strelitziaceae*
- Vybrané kandidátní geny byly po ověření počtu kopií v genomu banánovníku (Southernovými hybridizací) amplifikovány ze všech 14 uvedených druhů a sekvenovány
- Získané sekvenční daty byly použity k rekonstrukci fylogeneze čeledi *Musaceae*

4. Sekvenování, analýza sekvenčních dat, rekonstrukce fylogeneze

Fylogenetický strom získaný analýzou sekvenčních dat z 8 kandidátních genů metodou maximální parsimonie (software MEGA4, Tamura et al., 2007). Spolehlivost topologie stromu byla ověřena metodou bootstrapping + počtem replikací 1000 (výsledné hodnoty v % jsou uvedeny u jednotlivých uzlů).

Souhrn

- Pro studium fylogeneze v rámci čeledi *Musaceae* byla použita sekvenční analýza genů s nízkým počtem kopií
- Předběžné výsledky potvrzují blízkou příbuznost A genomu banánovníku (sekce *Eumusa*) se sekci *Rhodochlamys* a příbuznost sekce *Australimusa* se sekci *Callimusa*
- Kompletní analýza souboru 24 kandidátních genů bude použita pro datování divergence jednotlivých větví v rámci čeledi *Musaceae* a přispěje k přesnější a spolehlivější klasifikaci druhů a poddruhů banánovníku

Tato práce byla podpořena grantem IAA600380703 Grantové Agentury ČR.

ZÁVĚR

Předkládaná dizertační práce byla zaměřena na analýzu genetické diverzity a fylogeneze čeledi *Musaceae* s využitím molekulárních nástrojů.

Cílem první části práce bylo vybrat vhodný způsob pro molekulární charakterizaci genových zdrojů čeledi *Musaceae* a optimalizovat jej pro použití ve vysokokapacitním genotypování jednotlivých druhů, poddruhů a kultivovaných odrůd banánovníku. Z dostupných molekulárních nástrojů byl pro tento účel zvolen soubor 22 mikrosatelitových markerů, které byly již dříve vyvinuty pro studium diverzity banánovníku. Za použití moderních metod separace PCR produktů pomocí kapilární elektroforézy ve spojení s fluorescenčním značením primerů byl tento soubor markerů otestován na referenční sbírce genotypů banánovníku. Byla ověřena jeho funkčnost a reprodukovatelnost a také možnosti automatizace a tzv. multiplexingu analýzy vzorků. Bylo zjištěno, že 19 z původních 22 markerů je použitelných jako dostatečně spolehlivý a účinný nástroj pro potřeby molekulární identifikace jednotlivých genotypů banánovníku. Zároveň byla pomocí tzv. slepého testu u tohoto genotypovacího systému ověřena a potvrzena schopnost charakterizovat neznámé vzorky, identifikovat jejich pravděpodobné taxonomické zařazení na úrovni druhů a poddruhů a také identifikovat případné duplikované vzorky. Výsledky této části práce tak přispívají k důkladnému poznání genetické diverzity banánovníku a nabízejí první standardizovanou, vysokokapacitní genotypovací platformu, která se stala základní metodou používanou v činnosti tzv. Musa Genotyping Centre. Toto Centrum působí v rámci Laboratoře molekulární cytogenetiky a cytometrie ÚEB v Olomouci a bylo pověřeno ověřováním molekulární identity položek genové banky banánovníku a molekulární charakterizací zajímavých, sporných, případně nových genotypů banánovníku pro širší vědeckou veřejnost. Optimalizace spolehlivého genotypovacího systému pomocí SSR markerů poprvé umožní charakterizaci diverzity banánovníku na globální úrovni a zajistí vzájemnou porovnatelnost jednotlivých položek, a to i díky tomu, že všechny položky nově zaváděné do genové banky budou procházet molekulární charakterizací v Musa Genotyping Centre. Zde mohou být tyto analýzy doplněny také o sekvenční analýzu oblasti ITS (vhodnou pro případnou bližší charakterizaci hybridních klonů) a také o analýzu pomocí již dříve vypracovaných

metod (např. průtokové cytometrie pro měření velikosti genomu a stupně ploidie; či fluorescenční *in situ* hybridizace).

Druhá část práce byla zaměřena na důkladné prostudování evolučních vztahů v rámci čeledi *Musaceae*. K tomuto účelu bylo použito dvou přístupů. V prvním, na němž jehož experimentální části jsem se částečně podílela, bylo využito fylogenetické analýzy ITS lokusu ribozomální DNA na velmi širokém souboru 87 genotypů banánovníku obsahujícím také hybridní triploidní klony. Ve druhém přístupu byl naopak použit rozsáhlý soubor 19 jaderných genů, resp. jejich intronů, odvozených z dostupných sekvencí DNA banánovníku. Vybrané kandidátní geny byly amplifikovány a následně sekvenovány z genomové DNA 13 diploidních zástupců banánovníku, zvolených tak, aby reprezentovali diverzitu čeledi *Musaceae*. Fylogenetická analýza získaných sekvencí DNA umožnila jednak pro rekonstrukci evolučních vztahů v čeledi *Musaceae* a také datování divergence jednotlivých vývojových linií uvnitř čeledi. V souladu s předchozími pracemi bylo minimální stáří čeledi datováno do období přelomu třetihor a křídly (~ 69 mil. let) a divergence rodu *Musa* do období staršího paleogénu (~ 50 mil. let). Získaná data dále ukazují na počátek oddělování vývojových větví vedoucích k sekcím rodu *Musa* před řádově 28 mil. let. Vývojová linie nejmladší ze sekcí, sekce *Australimusa*, se začala diverzifikovat před asi 5 mil. let. Tato část práce představuje fylogenetickou analýzu čeledi *Musaceae* na dosud nejrozsáhlejším souboru jaderných genů a poskytuje první údaje stáří jednotlivých evolučních větví uvnitř této čeledi, čímž posunuje současné znalosti o evoluci jedné z nejdůležitějších plodin.

Společně s první částí práce tak získané výsledky poskytují komplexní pohled na evoluci banánovníku využitelný pro další studium diverzity genových zdrojů banánovníku a zpřesnění klasifikace jednotlivých druhů. Ve svém souhrnu znamenají získané výsledky zásadní posun ve studiu genetické diverzity a fylogeneze čeledi *Musaceae*.

SEZNAM POUŽITÝCH ZKRATEK

AFLP	polymorfizmus délky amplifikovaných fragmentů (amplified fragment length polymorphism)
DArT	Diversity Array Technology
DNA	deoxyribonukleová kyselina
ETS	vnější přepisovaný mezerník (external transcribed spacer)
IGS	mezigenový mezerník (intergenic spacer)
IR	obrácená repetice (inverted repeat)
ITC	International Transit Centre
ITS	vnitřní přepisovaný mezerník (internal transcribed spacer)
ML	maximální pravděpodobnost (maximum likelihood)
mRNA	mediátorová RNA
NJ	Neighbor-joining
NTS	nepřepisovaný mezerník (non-transcribed spacer)
PCR	polymerázová řetězová reakce (polymerase chain reaction)
RAPD	náhodně amplifikovaná polymorfní DNA (random amplified polymorphic DNA)
rDNA	ribozomální DNA
<i>rbcL</i>	ribulóza-bisfosfát kaboxyláza
RFLP	polymorfizmus délky restričních fragmentů (restriction fragment length polymorphism)
RNA	ribonukleová kyselina
rRNA	ribozomální RNA
SNP	jednonukleotidový polymorfizmus (single nucleotide polymorphism)
SSR	jednoduchá repetitivní sekvence (simple sequence repeat)
UPGMA	unweighted pair group method with arithmetic mean