

UNIVERZITA KARLOVA V PRAZE

PŘÍRODOVĚDECKÁ FAKULTA

Studijní program: Speciální chemicko-biologické obory
Studijní obor: Molekulární biologie a biochemie organismů



Studium environmentálních vzorků pomocí pyrosekvenování a vliv intragenomové variability ITS rDNA oblasti na odhady diverzity hub

Pyrosequencing analysis of fungal assemblages and the effect of ITS rDNA intragenomic variation on diversity estimates

BAKALÁŘSKÁ PRÁCE

Autor: Tomáš Zelenka

Školitel: Mgr. Miroslav Kolařík, Ph.D.

Praha, 2013

Prohlášení:

Prohlašuji, že jsem závěrečnou práci zpracoval samostatně a že jsem uvedl všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze dne 17.05.2013

.....

Poděkování

Rád bych poděkoval zejména svému školiteli Mirku Kolaříkovi za jeho bezbřehou vstřícnost, neskutečnou trpělivost při naší kooperaci a za nesčetné cenné rady do studia i do života. Velký DÍK patří osazenstvu našeho bytu (Aničce s Nelou, Míše s Andreou a Katce), Frankovi a samozřejmě také rodině za jejich pomoc a podporu.

Abstrakt

ITS region jaderné ribozomální DNA je dnes jeden z nejčastějších houbových *barcode* markerů. Používá se v environmentálních metagenomických analýzách, při determinaci hub, ale i při studiu jejich fylogeneze. V genomu se rDNA vyskytuje ve formě multikopiových tandemových repetit. Běžným Sangerovým sekvenováním získáme vždy pouze konsenzus repetit, zakrývající většinu možných intragenomových variací. Stále častěji používané 454 pyrosekvenování odhaluje sekvence jednotlivých kopií. Většina mutací se během jevu zvaného *concerted evolution* vytrácí. Částečná variabilita v kopiích však zůstává zachována, včetně možných pseudogenů. Vzhledem k častému používání 454 sekvenování u environmentálních studií nejsme schopni tuto variabilitu odlišit od reálné diverzity. Tím pravděpodobně dochází k nadhodnocování skutečné diverzity a mylnému vyhodnocování dalších studií. Práce shrnuje studie zabývající se zjišťováním diverzity hub pomocí 454 sekvenování a zároveň předkládá rešerši současných poznatků o intragenomové variabilitě rDNA coby potenciálním zdroji nepravé diverzity. Doposud chybí experimentální práce, která by se komplexním propojením těchto problematik zabývala.

Klíčová slova: pyrosekvenování, 454 sekvenování, ITS oblast, vnitřní přepisovaný mezerník, intragenomová variabilita, intradruhovká variabilita, pseudogeny, houby, diverzita hub

Abstract

The ITS region of nuclear ribosomal DNA has recently become a frequently studied region for its use as a barcode marker. It is employed in environmental metagenomic analyses, the determination of fungi and in studies of fungal phylogenetics. In genomes, rDNA occurs in the form of large multicopy tandem arrays. Thus common Sanger sequencing leads to a consensus of many copies which in fact conceals most of the potential intragenomic variation. However 454 pyrosequencing reveals the sequence of single copies. Although it is believed that most of the variation among copies is reduced during a process called concerted evolution, some variation might still be conserved, including variation of non-functional pseudogenes. As a consequence, when using 454 pyrosequencing, we are not able to discriminate between this variation and real diversity. This could have a huge impact on the estimation of real diversity as well as on the correct assessment of pyrosequence studies. This thesis reviews current knowledge of intragenomic variation among fungi and summarizes some papers applying pyrosequencing in the research of fungal diversity. At the same time it indicates intragenomic variation as a potential cause of untrue diversity in diversity studies. Up until now there hasn't existed any experimental survey covering these topics.

Key words: pyrosequencing, 454 sequencing, ITS region, internal transcribed spacer, intragenomic variation, intraspecific variation, pseudogenes, fungi, fungal diversity

OBSAH

Seznam použitých zkratk

1. Úvod.....	1
2. Sekvenování nukleových kyselin	2
2.1. Sangerovo sekvenování	2
2.2. Sekvenační analýza 454.....	3
2.2.1. Příprava DNA knihovny	3
2.2.2. Emulzní PCR	4
2.2.3. Sekvenování DNA	5
2.3. Porovnání metod	6
2.4. Vyhodnocování výsledků sekvenování	7
2.4.1. Purifikace sekvencí.....	7
2.4.2. Klastrování sekvencí	7
2.4.3. Metagenomické analýzy	9
3. Poznatky o ITS oblasti u hub	10
3.1. Uspořádání rDNA u hub	10
3.1.1. Mechanismy boje s mutacemi v rDNA	12
3.1.2. Vliv počtu repetice rDNA na biologii houby	12
3.2. Variabilita ITS oblasti	13
3.2.1. Intradruhovká variabilita	13
3.2.1. Intragenomová variabilita.....	14
3.2.2. Fúze vegetativních hyf	17
4. Rozlišení pseudogenů u hub	17
4.1. Původ pseudogenů	18
4.2. Vyhledávání pseudogenů	19
5. Mykologické práce využívající 454 sekvenování	20
5.1. Metodické chyby	21
5.2. Nalezené práce.....	21
5.3. Diskuze výsledků.....	22
6. Závěr.....	24
7. Použitá literatura.....	25
8. Přílohy	35

Seznam použitých zkratk

BLAST	The Basic Local Alignment Search Tool	Algoritmus na základní srovnávání sekvencí
CBC	Compensatory Base Change	Změny nenarušující stonek
CCD	Charge Coupled Device	Zařízení s vázanými náboji
emPCR	The EMulsion-based clonal amplification of a DNA	Emulzní PCR
FASTA	FAST-All	Základní textový formát nukleotidových a peptidových sekvencí
FISH	Fluorescence In Situ Hybridization	Fluorescenční in situ hybridizace
CHEF	Contour-clamped Homogeneous Electric Field electrophoresis	Technika pulzní gelové elektroforézy
IDV	Intraspecies Variation	IntraDruhová Variabilita
IGS	InterGenic Spacer	Mezigenový mezerník
IGV	IntraGenomic Variation	IntraGenomová Variabilita
INSD	The International Nucleotide Sequence Databases	Mezinárodní databáze sekvencí nukleotidů
ITS	Internal Transcribed Spacer	Vnitřní přepisovaný mezerník
LINE	Long Interspersed Nuclear Elements	Dlouhé rozptýlené jaderné elementy
LSU	Large SubUnit	Velká ribozomální podjednotka
MIENS	Minimum Information about an ENvironmental Sequence	Směrnice o minimálních informacích o environmentálních sekvencích
MOTU	Molecular Operational Taxonomic Unit	Molekulární operačně taxonomická jednotka
mRNA	Messenger RNA	Mediátorová RNA
NGS	Next Generation Sequencing	Sekvenování nové generace
NUMT	NUclear MiTochondrial DNA	Jaderná mitochondriální DNA
OTU	Operational Taxonomic Unit	Operačně taxonomická jednotka
PCR	Polymerase Chain Reaction	Polymerázová řetězová reakce
qPCR	Quantitative real time Polymerase Chain Reaction	Kvantitativní PCR
rDNA	Ribosomal DNA	Ribozomální DNA
rRNA	Ribosomal RNA	Ribozomální RNA
SINE	Short Interspersed Nuclear Elements	Krátké rozptýlené jaderné elementy
SMRT	Single Molecule Real Time sequencing	Sekvenování jedné molekuly v ten daný moment
SOLiD	Sequencing by Oligonucleotide Ligation and Detection	Sekvenování pomocí ligace a detekce oligonukleotidu
SSU	Small SubUnit	Malá ribozomální podjednotka

1. Úvod

V posledních letech je pyrosekvenování velmi populární metodou k získávání sekvencí z environmentálních vzorků pro výzkum diverzity. Jako hlavní *barcode* marker se používá ITS oblast jaderné ribozomální DNA. Ta se v genomu vyskytuje ve formě tandemových repetit v počtu až 200 kopií. I přes homogenizační účinky *concerted evolution* zůstává mezi kopiemi zachována určitá variabilita. Tato intragenomová variabilita může být prezentována ve formě funkčních paralogů, ale také jako nefunkční pseudogeny. Pyrosekvenování, na rozdíl od přímého sekvenování, odhaluje sekvence jednotlivých kopií. Především zmíněné pseudogeny a tzv. „hluboké“ paralogy mohou způsobit nadhodnocení výsledků a tedy mylný odhad reálné diverzity.

V první kapitole shrnuji základní principy 454 sekvenační analýzy v kontextu metod sekvenování nové generace. Významnou část věnuji vyhodnocování sekvencí, protože chyby v této oblasti mohou také zkreslit výsledky. Následují poznatky o ITS regionu u hub, včetně jeho vlivu na celkovou biologii houby. Dále shrnuji charakteristiku a rozšíření variability mezi kopiemi, která vzniká při nedostatečné homogenizaci kopií.

Vzhledem k absenci prací, které by detailně popisovaly vliv intragenomové variability na výsledky pyrosekvenování, jsem vytvořil přehledovou tabulku 60 mykologických prací, které 454 sekvenování použily. U jednotlivých studií jsem porovnal použité metody vyhodnocování a výsledné počty OTU. V části prací jsem navíc porovnal získané OTU s diverzitním ukazatelem Chao1, který odhaduje skutečnou diverzitu daného habitatu. Ze sepsaných hodnot je patrné, že většina studií odkrývá pouze malé procento celkové diverzity (Kap. 5.3.). To zároveň pomáhá uvědomit si nedostatečnou hloubku sekvenování a nutí tak k zamýšlení nad *rare biosphere* a nedostatečným odhalováním reálné diverzity. Hlavním cílem tabulky je podat důkaz o tom, že je potřeba dalšího studia zmíněné problematiky a mimo to také sjednocení metodiky a typu prezentovaných dat.

V současné době se nacházíme na začátku nové éry studia environmentální DNA. Zavedení NGS metod u hub bylo převratnou událostí, protože velká část jich není kultivovatelná. Nekultivovatelné jsou zejména houby parazitické a symbiotické, které však mohou být v některých habitatech dominantní (např. mykorhizní houby v lesní půdě). Pro představu lze uvést práci Hibbett et al. (2011), porovnávající data získaná z GeneBanku. Z celkového počtu klastrů (definovaných na 97% hladině podobnosti) jich bylo 6 230 (37 %) z environmentálních vzorků a pouze 2 223 (13 %) se překrývalo s OTU z ostatních vzorků – plodnic, kultur atp.

První práce využívající 454 sekvenování při studiu hub byla publikována až v roce 2009. To nasvědčuje, že se nacházíme teprve ve stádiu optimalizací a poznávání metodických možností. Předložená rešerše by mohla sloužit jako takový první manuál pro další studenty využívající tento metodický přístup. Díky rychlému technologickému postupu je zřejmé, že zvolená metoda 454

sekvenování může být brzy nahrazena přesnějšími a levnějšími postupy. Proto bylo mou snahou přiblížit obecné principy vyplývající z použití NGS metod.

Kvůli velkému množství cizích termínů, které lze mnohdy jen obtížně přeložit, jsem se rozhodl některé ponechat v původním znění, psané kurzívou. Protože ne všechna popisovaná témata mají dostatek relevantních informací o dané problematice u hub, extrapoloval jsem je též ze studií bakteriologických či rostlinných.

2. Sekvenování nukleových kyselin

Sekvenování je souhrnný název pro metody zjišťující pořadí bází v nukleových kyselinách a řadí se mezi nejdůležitější platformy studia biologických systémů současnosti. Snaha sekvenovat DNA existuje již od dob Watsona a Cricka, kteří strukturu této dvoušroubovice popsali. Mezi nejdůležitější metody patří Sangerovo dideoxysekvenování (Sanger et al., 1977), později upravované, mj. pro použití s fluorescenčními barvivami (Prober et al., 1987). V dnešní době začínají převažovat tzv. metody sekvenování nové generace (dle anglického Next Generation Sequencing = NGS), které se vyznačují především miniaturizací a vysokou paralelitou prováděných reakcí. Kromě snahy snížit cenu celé procedury hrají důležitou roli také její rychlost či velikost sekvenovaných fragmentů. NGS metod je poměrně značné množství (Liu et al., 2012), ale mezi nejpoužívanější stále patří SOLiD, Illumina a 454 sekvenování. Srovnání několika NGS metod se Sangerovým sekvenováním ukazuje tab. 1.

Při tzv. *metabarcodingu* se k analýze environmentálních vzorků využívají markerové sekvence získané z prostředí (Taberlet et al., 2012). K sekvenaci těchto markerů je velmi vhodné 454 pyrosekvenování, které umožňuje přečíst i dostatečně dlouhé sekvence (Tab. 1). Proto bych rád nastínil podrobnější principy jeho fungování nutné k pochopení problematiky vyhodnocování pyrosekvenačních studií. Pro potřeby porovnání 454 a Sangerova sekvenování nastíním i základní princip této metody.

2.1. Sangerovo sekvenování

V dnešní době se již používá řada různých modifikací této metody, základní princip však zůstává stále zachován. Při běžné replikaci DNA se ke stávajícímu řetězci připojují jednotlivé nukleotidy podle antiparalelní matrice. Vzniká fosfodiesterová vazba mezi OH skupinou na 3' uhlíku cukerné složky posledního nukleotidu v řetězci a proximálním fosfátem nukleotidu přicházejícího. Aby reakce mohla pokračovat, musí nově zařazený nukleotid také obsahovat 3' OH skupinu. Právě toho využívá Sangerova metoda: do reakční směsi jsou ve vhodném poměru přidány dideoxynukleotidy, které mají 3' OH skupinu nahrazenou vodíkem. Řetězec pak není schopen další elongace (Sanger et al., 1977). Výsledkem je velké množství různě dlouhých fragmentů, které jsou vzápětí rozděleny

podle délky např. kapilární gelovou elektroforézou. Následně je zdetekováno, jakou bázi úsek končí, nejčastěji pomocí fluorescenčních barviv (Dolnik, 1999). Ze všech fragmentů je pak vytvořena výsledná konsenzuální sekvence.

Platforma	Sanger 3730xl	454 GS FLX	SOLiDv4	Illumina HiSeq 2000	Ion Torrent PGM	PacBio RS (SMRT)
Délka čtení	400 ~ 900 bp	700 bp	50 bp	150 bp	~ 200 bp	1500
Data získaná za jeden cyklus	1.9 ~ 84 kb	0.7 Gb	120 Gb	600 Gb	1 Gb (318 chip)	100 Mb
Délka trvání cyklu	20 min ~ 3 hod	24 hod	7 dní pro SE	11 dní	2 hod	2 hod
Cena přístroje	95 000 \$	500 000 \$ (108 000 \$ GS Junior)	495 000 \$	690 000 \$	80 000 \$	695 000 \$
Cena za Mb	2400 \$	10 \$	0,13 \$	0,07 \$	1 \$ (318 chip)	2 \$
Dosahovaná přesnost	Q 50	> Q 20	Q 40	> Q 30	Q 20	< Q 10

Tab. 1: Ukázka NGS metod a základní popis jejich vlastností [převzato a upraveno z Liu et al. (2012); Quail et al. (2012)].

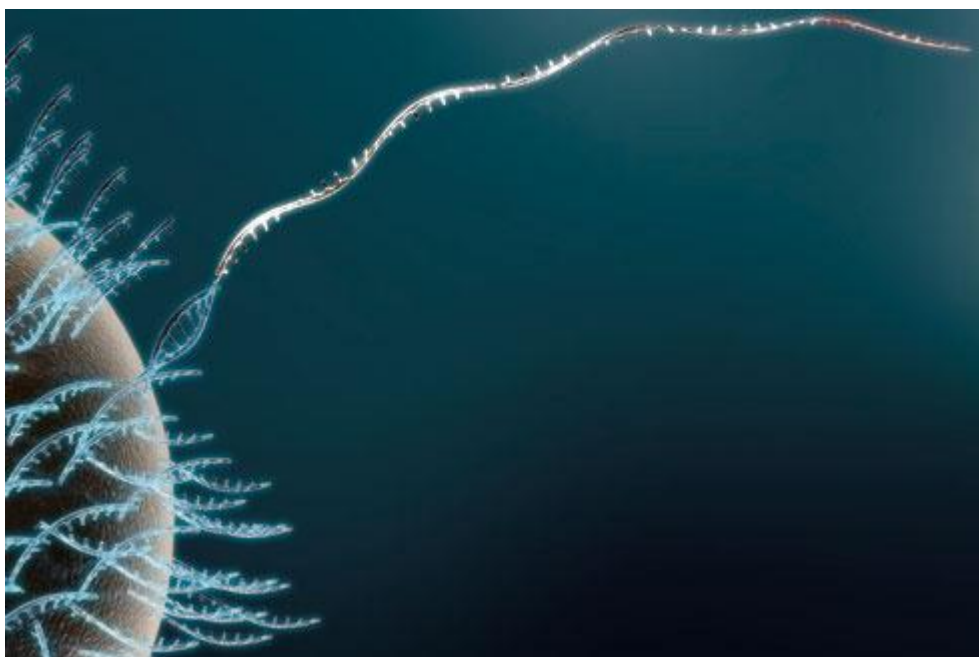
2.2. Sekvenační analýza 454

S metodou 454 sekvenování přišla na trh firma Life Sciences (nyní Roche Diagnostics; <http://www.454.com>). Mezi důležité klady patří paralelita zpracování vzorků a celková miniaturizace procesu. Pro představu: skříčko, na kterém probíhá sekvenování, obsahuje přibližně 1,6 milionu jamek o objemu 75 pl. Do nich se po jedné usazují kuličky s cca 10 miliony navázaných fragmentů DNA. Ve výsledku je generováno i přes 25 milionů bází se skórem kvality Phred 20 a vyšším (Margulies et al., 2005). Toto skóre udává pravděpodobnost chybného přečtení sekvenované báze, podle vzorce $Q = -10 * \log_{10}(P)$. Při skóre 20 je tedy pravděpodobnost chybného vyhodnocení báze 99 %, což odpovídá jedné chybné bázi ze sta přečtených (Richterich, 1998).

2.2.1. Příprava DNA knihovny

Vstupní materiál je buď směs PCR ampliconů (délky 200-800 bp) či celá genomická DNA, ze které je třeba vytvořit DNA knihovnu. Její příprava začíná fragmentací DNA na krátké úseky, přibližně 300 až 800 párů bází dlouhé, případně sestavením kolekce DNA této délky. V dalším kroku, po zatupení fragmentů, jsou na jejich konce ligovány adaptéry. Tyto se klasicky označují A a B a obsahují primerovou sekvenci důležitou pro další amplifikaci a sekvenování

(<http://www.gsjunior.com/instrument-workflow.php>). Často se používají také ke značení jednotlivých fragmentů tím, že obsahují specifickou sekvenci. Dále obsahují kupř. restrikční místo, které pomáhá k odfiltrování nesprávně nalogovaných adaptérů (Meyer et al., 2008). B adaptér obsahuje na 5' konci biotin. Ten je afinitní látkou ke streptavidinu, který obaluje malé kuličky na jejichž povrch se DNA váže – vždy jeden fragment na kuličku (Obr. 1). Toho je docíleno patřičným ředěním. Dalšími kroky, jako např. působením zásady, se lze zbavit nebiotinylovaných řetězců a dále odstranit nepřipojené či špatně připojené fragmenty (Margulies et al., 2005). Tímto je DNA knihovna připravena k dalšímu kroku.



Obr. 1: Reakční kulička s navázanými B adaptéry (modře) a jedním fragmentem, který bude amplifikován a sekvenován (bíle). Na jeho druhém konci je ligovaný načervenalý adaptér A. Převzato z <http://www.454.com>.

2.2.2. Emulzní PCR

Z připravené knihovny je vytvořena emulze reakčních kapiček vody v oleji, přičemž v každé kapce se nachází právě jedna kulička. Následuje PCR, která se proto označuje jako emulzní, a jejím výsledkem jsou kuličky obalené uniformními kopiemi původně navázaného fragmentu (Margulies et al., 2005).

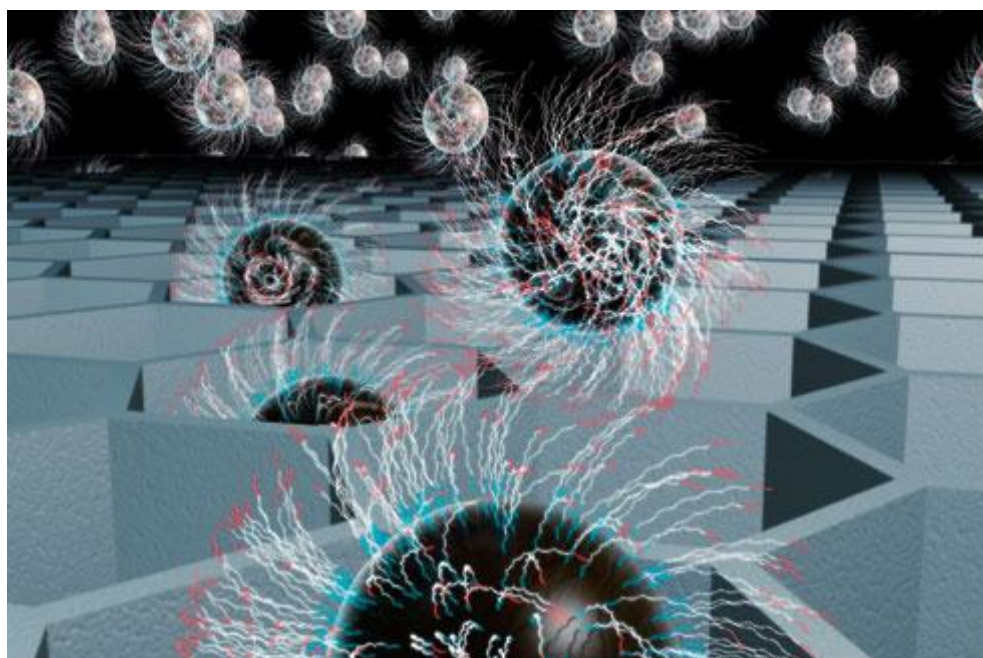
Jelikož pravděpodobně vzniká mnohem větší podíl chyb než při samotném sekvenování, měla by se jejímu designu věnovat zvýšená pozornost. Proto se mj. doporučuje minimalizovat počet PCR cyklů (Patin et al., 2012). V úvahu je potřeba brát také fakt, že existuje vztah mezi množstvím původní DNA a počtem cyklů PCR. Jejich nevhodná synchronizace pak může vést k nad- nebo podhodnocení výsledného počtu fragmentů (Avis et al., 2010).

Významně lze reakci ovlivnit i zvolením pufru či odstraněním nežádoucích inhibitorů a nečistot. Velmi důležitá je také volba polymerázy. Osvědčilo se používání směsí polymeráz s různými vlastnostmi (Arezi et al., 2003). Jako příklad lze uvést procesivní mutantu *Taq* polymerázy s malým přídavkem archeální polymerázy *Pfu*. Ta je sice procesivní méně, nicméně obsahuje 3'–5' exonukleázovou, *proofreading*, aktivitu, která významně zvyšuje přesnost celé amplifikace. Tato kombinace dokáže amplifikovat mnohem delší úseky DNA než jednotlivé polymerázy samostatně (Barnes, 1994).

Archeální polymerázy navíc disponují tzv. *read-ahead* funkcí, která rozpoznává uracil v templátovém řetězci, a spolu s *proofreadingem* zvyšuje jejich účinnost. Pozoruhodným zjištěním bylo, že *hot start* varianty polymeráz celkovou účinnost amplifikace nijak výrazně nezvyšují (Arezi et al., 2003).

2.2.3. Sekvenování DNA

Po emPCR se kuličky s dalšími potřebnými agens nanesou do pikolitrových jamek v reakční destičce – vždy jedna kulička do jedné jamky (Obr. 2). V jamkách probíhá dosyntetizování druhého řetězce fragmentů podle primeru, který je komplementární k primerové části adapteru A. Destičky jsou postupně promývány nukleotidy tak, aby bylo možno rozpoznat, kdy se jaký nukleotid připojil (Margulies et al., 2005).

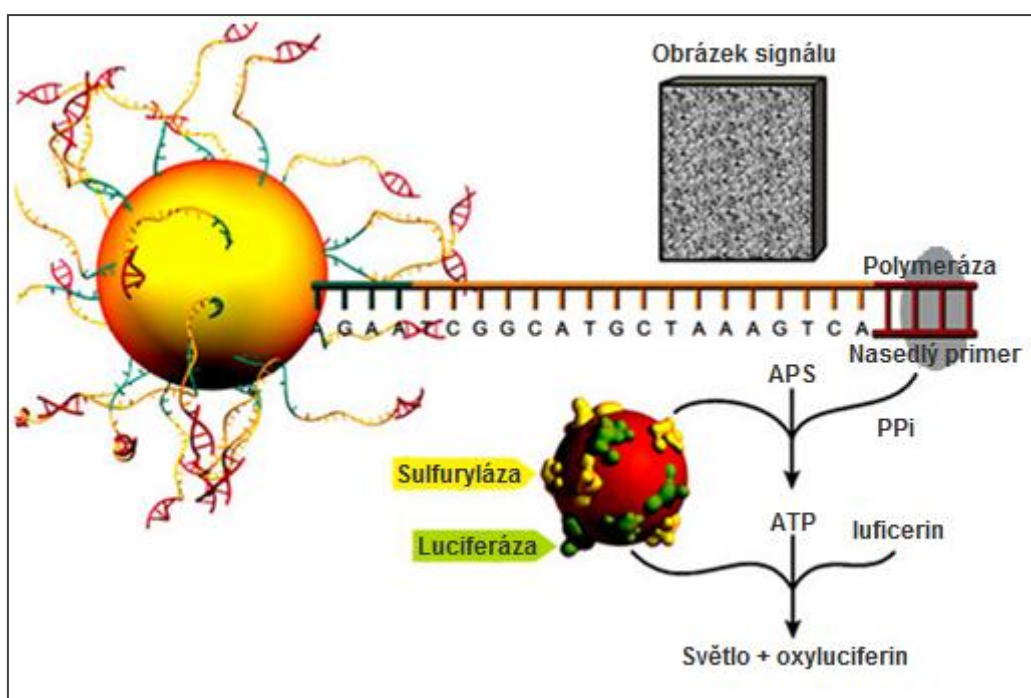


Obr. 2: Ukázka nanášení kuliček obalených uniformními fragmenty do pikolitrových jamek. Převzato z <http://www.454.com>.

Po připojení nukleotidu se uvolňuje pyrofosfát, proto se tato metoda označuje jako pyrosekvenování. Pyrofosfát je konvertován ATP sulfurylázou na ATP (Obr. 3), které zapříčiní

změnu luciferinu na oxyluciferin pomocí enzymu luciferázy (Ronaghi et al., 1996). Při této reakci se uvolňuje světlo, jehož množství koresponduje s počtem zařazených nukleotidů. Intenzitou tedy odpovídá délce právě sekvenovaného homopolymeru. Světlo je detekováno CCD čipy a počítačově převedeno do výsledného *flowgramu*. Z tohoto principu vyplývá nejzávažnější zdroj chyb, které jsou způsobené právě čtením délky homopolymerů. Díky nedokonalé kalibraci detekce světla se tak může například zaměnit polymer TTT za TTTT a naopak (Ronaghi et al., 1998).

Zbylé nukleotidy jsou odstraňovány apyrásou, aby mohlo proběhnout další kolo promývání a nedocházelo tak k falešným výsledkům (Ronaghi et al., 1998).



Obr. 3: Znárodnění principu vzniku světelného záblesku při pyrosekvenování. Po inkorporaci nukleotidu se uvolní pyrofosfát, který je enzymem sulfurylázou konvertován na ATP. To způsobí změnu luciferinu na oxyluciferin za pomoci enzymu luciferázy. Při této reakci je uvolňováno také světlo, jež můžeme detekovat. Převezato a upraveno z <http://www.454.com>.

2.3. Porovnání metod

454 i Sangerovo sekvenování se v různých ohledech významně liší. V rámci této práce bych opomenul rozdíly v délce sekvenovaných fragmentů apod. Za účelem pochopení dalších částí textu je důležité uvědomit si základní podstaty obou typů.

Obecně při metodách přímého sekvenování (Sangerovo) je výsledkem konsenzus sekvencí. Máme-li multikopiový úsek DNA, je tak v podstatě irrelevantní, na které z nich primer nasedne. Nakonec je vždy vidět výsledný konsenzus sekvencí jednotlivých klonů. Tím se efektivně zastře většina PCR chyb, ale i malá přirozená variabilita daného PCR produktu.

Naopak meritem „nepřímého“ sekvenování (454, klonování) je výsledek ve formě jednotlivých osekvenovaných kopií. Ty jsou (v případě 454 metody) během emPCR amplifikovány a posléze

sekvenovány. Jakékoliv mutace, byť jen v jediné kopii, pak mohou vést k chybnému závěru, že se například jedná o novou OTU. Protože se polymorfismus v říši hub vyskytuje poměrně hojně (Nilsson et al., 2008; Simon et Weiss, 2008; Smith et al., 2007), jsou obavy z nadhodnocených počtů OTU namístě.

2.4. Vyhodnocování výsledků sekvenování

Získání přesných dat použitím jakékoliv metody je vždy poměrně náročný proces. Původní celková přesnost 454 sekvenování byla pro genomické knihovny vypočítána na 96 % (Margulies et al., 2005). Vzniklé chyby se ale dají částečně eliminovat vytvořením konsenzu z velkého počtu přečtených sekvencí a docílit tak přesnosti stejné, či dokonce lepší, než je tomu u klasických metod. (Huse et al., 2007).

2.4.1. Purifikace sekvencí

V současné době existuje mnoho algoritmů a metod na detekci a eliminaci chyb. Základní postup při jejich odstraňování je však stále podobný. Během tzv. filtrování se odstraní sekvence, které jsou s vysokou pravděpodobností špatné. Odfiltrují se kupř. extrémně krátké a dlouhé sekvence anebo ty, které se vyskytují pouze v malém počtu kopií. Také se kontroluje, zda sekvence obsahují primerovou část (Quince et al., 2011).

Dále se provádí tzv. *denoising*, při kterém se odfiltrovávají nepřesné signály a chyby zavedené při sekvenování. Tyto nepřesné signály způsobují především různou výslednou délku sekvencí u homopolymerních úseků (Quince et al., 2011). Přehled několika významných *denoising* algoritmů, resp. programů, zachycuje tab. 2.

Důležitou součástí purifikace je rozpoznání a odstranění chimér – DNA fragmentů vzniklých při PCR, složených z několika částí z různých mateřských DNA (Quince et al., 2011). Četnost chimér se dá částečně ovlivnit designem PCR – kombinací vhodného počtu cyklů, počáteční koncentrací vzorku a samozřejmě i volbou polymerázy (Lahr et Katz, 2009).

Název softwaru	PyroNoise	Pyrotagger	DeNoiser	AmpliconNoise
Reference	Quince et al., (2009)	Kunin et al., (2010)	Reeder et Knight, (2010)	Quince et al., (2011)

Tab. 2: Ukázka několika programů používaných při *denoisingu*.

2.4.2. Klastrování sekvencí

Dalším krokem je klastrování do molekulárních, operačně-taxonomických, jednotek (MOTU), často zvaných pouze OTU. Pod tímto pojmem se skrývá terminální taxonomická skupina pomocí níž

jsme schopni určovat diverzitu. ITS region ribozomální DNA se pro rozlišení OTU standardně klastruje na základě konvenční hladiny 97% podobnosti sekvencí. To umožňuje vyhnout se dosud neodstraněným chybám a v nejlepším případě také vlivu vnitrodruhové variability (Nilsson et al., 2008).

Výsledky práce Nilsson et al. (2008) ale ukazují, že hodnoty intradruhové variability (IDV) se mezi různými skupinami hub významně liší. Pro klastrování velké části hub ze skupiny Ascomycota by bylo často vhodnější používat 98% či dokonce 99% hladinu podobnosti. Tato výjimka může být pravděpodobně způsobena vyhodnocováním teleomorfních i anamorfních forem od jednoho druhu zároveň (Blaalid et al., 2013). Opačný případ vhodnější konzervativní hranice 95 % nalezneme u výzkumu endofytů (Arnold et Lutzoni, 2007). U většiny hub je ale 97% hladina optimální, což dokazují výsledky mnoha studií (Bjorbaekmo et al., 2010; Blaalid et al., 2013; Kunin et al., 2010; Tedersoo et al., 2010) a pro potřeby environmentálních studií je patrně nejlepším řešením.

Problematika klastrování není žádnou novinkou, nicméně dřívější algoritmy využívaly srovnávání všech sekvencí se všemi. To začalo být s příchodem obrovského množství dat z NGS, vzhledem k vysokým technickým nárokům, obtížně proveditelné (Li et al., 2012). Proto vznikaly nové způsoby klastrování založené na důmyslnějších algoritmech.

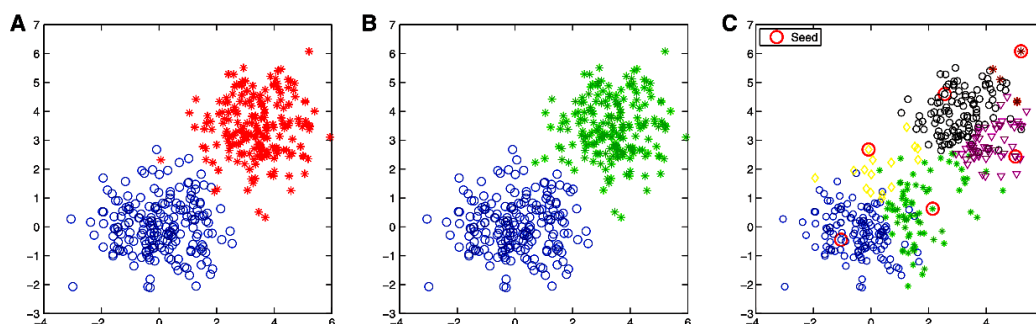
Jedním z nejpoužívanějších klastrovacích programů je CD-HIT, který se vyskytuje v mnoha modifikacích. Pro NGS data by bylo vhodné použít verzi 4.6 (Fu et al., 2012). CD-HIT-454 je pak uzpůsobený k vyhledávání umělé duplikáty sekvencí, které je velmi těžké odlišit od těch přirozeně se vyskytujících (Niu et al., 2010). CD-HIT-OTU dokáže dokonce během několika kroků odstranit chyby sekvenování, a to za podobných výsledků jako *denoising* programy typu AmpliconNoise (Li et al., 2012).

CD-HIT používá ke zrychlení svých výpočtů následující heuristický princip. Nejprve se srovnají sekvence podle velikosti. Nejdelší sekvence slouží jako představitel prvního klastru (tzv. *seed*). Následující sekvence se srovnávají se sekvencemi z již zavedených klastrů. Klastry jsou tvořeny podle toho, kolika procentní nastavíme hladinu podobnosti pro uskupení sekvencí (Li et Godzik, 2006).

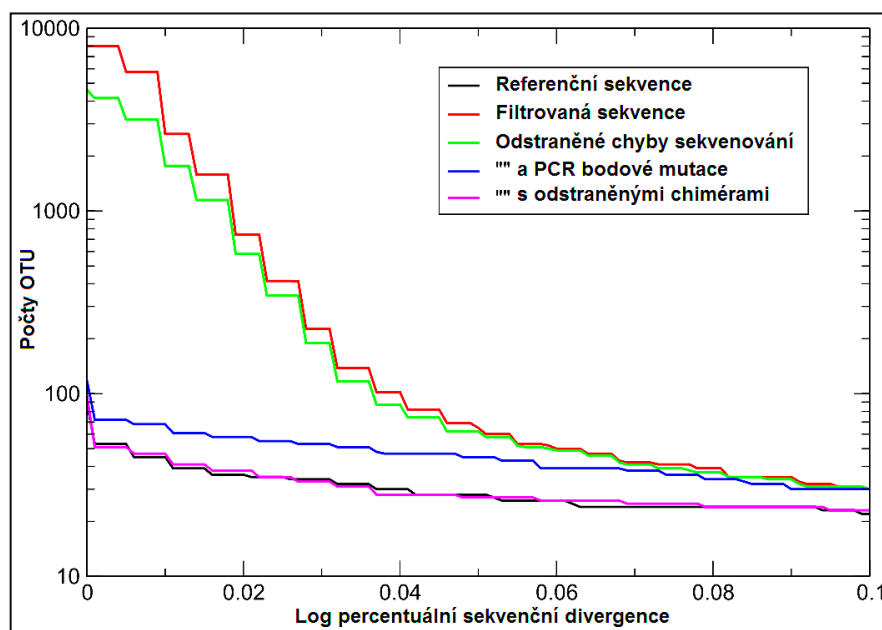
Na podobném principu jako CD-HIT pracují ještě např. UCLAST (Edgar, 2010), DNACLUST (Ghodsi et al., 2011) nebo SEED (Bao et al., 2011). Všechny ale patří mezi tzv. hladové algoritmy (*greedy incremental approach*) stejně jako CD-HIT.

Jiná studie však ukazuje, že i klasické hierarchické klastrování je po určitých úpravách použitelné na velké objemy dat. Příkladem je program ESPRIT-Tree, který klastruje na základě hierarchického *average linkage-based* algoritmu (Cai et Sun, 2011). Rozdíl ve výkonnosti obou typů klastrování je vidět na obr. 4.

Důležitost všech uvedených metod zpracovávání sekvencí jasně vyplývá z obr. 5, který porovnává počty OTU při postupném zavádění metod. Schéma ideálního postupu při purifikaci sekvencí uvádějí ve své práci Větrovský et Baldrian (2013).



Obr. 4: Příklad ukazující chování algoritmů využívajících přístupy *greedy incremental* a klasický hierarchický. (A) Data tvořená dvěma rozdílnými Gaussovými distribucemi. (B) Použití klasického hierarchického klastrování, které odkrývá pravou strukturu – v porovnání s obr. 4A. (C) Hladový algoritmus, jehož výsledek odpovídá původním datům pouze slabě. Červené kroužky představují tzv. *seedy*, tedy jakási mateřská data pro jednotlivé klastry. Vidíme, že se oba přístupy při použití stejných hladin sekvenční divergence chovají odlišně. Převzato a upraveno z Sun et al. (2012).



Obr. 5: Ukázka semilogaritmického vynesení závislosti počtu OTU na procentuálním rozdílu sekvencí během klastrování při použití různých metod odstraňování chyb. Převzato a upraveno z Quince et al. (2011).

2.4.3. Metagenomické analýzy

Jelikož se 454 sekvenování používá velmi hojně k získávání environmentálních metagenomických dat, nalezneme i dostatek programů na taxonomickou či funkční klasifikaci těchto sekvencí. Podle principu je můžeme rozdělit na *supervised* a *unsupervised*. Druhý jmenovaný slouží k odstranění sekvencí ze vzorků. *Supervised* porovnává neznámé sekvence s databázovými a lze je rozdělit do tří skupin dle použité metody. Jedná se o metody porovnávání sekvenčních podobností, složení sekvencí a metody fylogenetické (Bazin et Cummings, 2012). Jako příklad by se dal použít

program MEGAN, který na základě sekvenční podobnosti porovnává data s databází známých sekvencí pomocí BLAST a následně je klasifikuje podle NCBI taxonomie pomocí *lowest common ancestor* algoritmu (Huson et al., 2007).

Právě kvůli značnému množství algoritmů vznikají softwarové balíky sdružující několik z nich pod společné uživatelské rozhraní. Příkladem je program MOTHUR, který dokáže upravovat sekvence, srovnávat je, počítat jejich vzdálenosti a přiřazovat sekvence k jednotlivým OTU. Proto může být užitečným nástrojem pro popisování a porovnávání celých mikrobiálních společenstev (Schloss et al., 2009). Dalším balíkem je FastGroupII (Yu et al., 2006), který, na rozdíl od programu MOTHUR, používá algoritmy pracující na bázi *the nearest neighbour* a vykazuje tak vyšší počet OTU (Krüger et al., 2012). Nově je dostupný program SEED, který lze doplnit o další externí rozšíření a získat tak komplexní nástroj pro vyhodnocování sekvencí (Větrovský et al., 2013).

3. Poznatky o ITS oblasti u hub

Název ITS je zkratkou z anglického *Internal Transcribed Spacer*. Je to mezerák mezi geny pro malou (SSU) a velkou (LSU) ribozomální podjednotku v jaderné DNA (rDNA). ITS region je rozdělen genem pro 5,8S rRNA na ITS1 a ITS2 oblasti. Jednotlivé kopie jsou od sebe odděleny netranskribovaným, tzv. intergenovým, mezerákem IGS. Gen pro 5S rRNA bývá často obsažen právě v IGS. Mapa zmíněných genů je znázorněna na obr. 6.

3.1. Uspořádání rDNA u hub

Ribozomální DNA se vyskytuje v genomu ve formě tzv. *head-to-tail* tandemových repetice v různých počtech zhruba 30–200 kopií (Tab. 3). Již tento fakt nasvědčuje, proč používat právě ITS region jako genetický marker. Stačí pouze malé množství DNA a vzhledem k abundanci repetice ITS pravděpodobně podaří namnožit. Tato skutečnost velmi napomáhá např. při zpracovávání starých herbářových položek.

Délka ITS regionu, včetně 5,8S genu, se pohybuje mezi 270–972 bp. To je značný rozdíl vůči délce „konkurenční“ *barcode* oblasti genu *COXI*, která je v rozmezí 642 bp až 12,3 kbp, kvůli variabilní přítomnosti intronů (Seifert, 2009). Další nevýhodou *COXI*, stejně jako dalších uvažovaných genů (*Tub2*, *RPB1*, *RPB2*, *TEF1α*), je absence univerzálních primerů, které naopak pro ITS oblast existují. Navíc ITS obvykle obsahuje velkou mezeru mezi inter- a intradruhovou proměnlivostí a celkově je na průřezu říší hub dostatečně variabilní. Všechny tyto aspekty spolu se snadným vyhledáváním z již vytvořených, obsáhlých databází ITS sekvencí ho předurčují k úloze vhodného *barcode* markeru mezi houbami (Schoch et al., 2012).

Název organismu	Počet kopií rDNA	Použitá metoda	Reference
<i>Cochliobolus heterostrophus</i>	130	Southern blot	Garber et al. (1988)
<i>Phaffia rhodozyma</i>	61 (na 3 chromozomech)	CHEF	Wery et al. (1997)
<i>Mucor miehei</i>	100	Southern blot	Maicas et al. (2000)
<i>Cochliobolus heterostrophus</i>	90	FISH	Tsuchiya et Taga (2001)
<i>Melampsora sp.</i>	±200	qPCR	Boyle et al. (2005)
<i>Colletotrichum acutatum</i>	20	qPCR	Debode et al. (2009)
<i>Aspergillus fumigatus</i>	38–91 (dle kmene)	qPCR	Herrera et al. (2009)

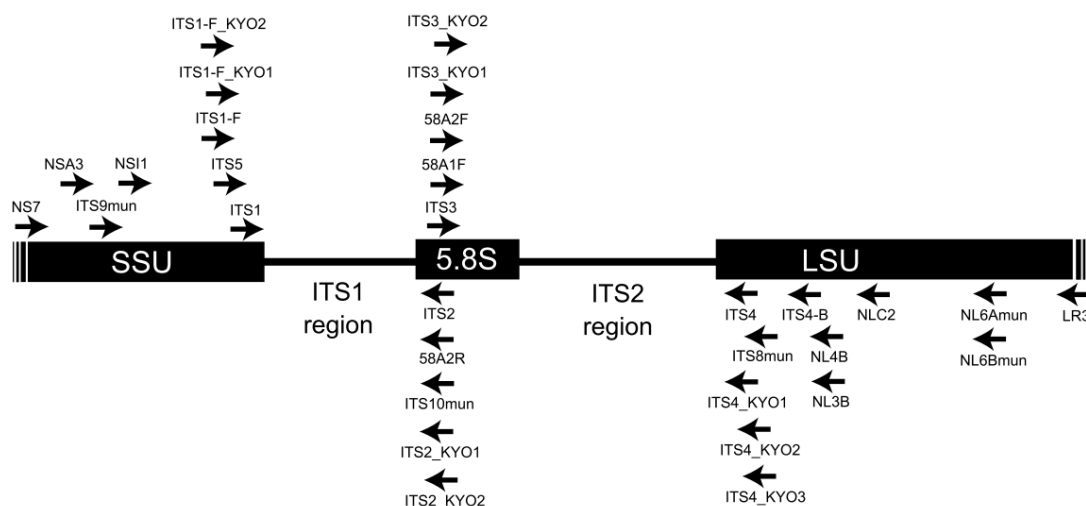
Tab. 3: Počty kopií rDNA u vybraných druhů. Informace jsou z prací využívajících Southernův přenos, značené sondy (FISH), qPCR (porovnáním s PCR křivkou *single-copy* genu) a metodu porovnání velikosti kopie rDNA s výsledkem CHEF elektroforézy.

Na druhou stranu je třeba dávat pozor i na jeho omezení – od intragenomové variability až po rozdílnost ve výsledcích odlišných typů sekvenování (Kiss, 2012; Lindner et Banik, 2011). Navíc je pravděpodobnost správné identifikace hub pomocí ITS oblasti pouze asi 73 % (Schoch et al., 2012). Právě vzhledem ke všem neduhům ITS se navrhuje, minimálně u problematických skupin, používat spolu s ním nějaký sekundární marker jako kontrolu (Seifert, 2009).

Pro amplifikaci ITS regionu z genomické DNA lze vybírat z mnoha možných primerů (Obr. 6). Jejich výběru je nutno věnovat velkou pozornost, protože rozdíly mezi použitelností jednotlivých primerů jsou markantní (Toju et al., 2012).

U arbuskulárních mykorhizních hub (Glomeromycota) nelze ITS oblast nejspíše vůbec používat – vzhledem k velké, až > 23%, intradruhovému a intrasporální variabilitě (Stockinger et al., 2009). Jako případnou možnost se doporučuje používat celý, 1500 bp dlouhý, fragment SSUmCf-LSUmBr. Obsahuje část SSU, celou oblast ITS a část LSU zahrnující D1 a D2 oblasti. U pyrosekvenačních studií generujících přibližně 400 bp dlouhé úseky jsou pak nejvhodnější oblasti ITS2 a LSU-D2 (Stockinger et al., 2010).

Některé skupiny jako Zygomycota jsou zastoupeny v ITS knihovnách v počtu 1,5 % a oddělení Chytridiomycota pak dokonce vůbec. Naopak v SSU knihovnách jsou zastoupeny obě po 4 %, což značí rozdíly v primerové specifitě (O'Brien et al., 2005). Volba primerů u různých skupin by se tedy měla zohledňovat také podle těchto informací.



Obr. 6: Schéma ribozomálních genů pro malou (SSU) a velkou (LSU) ribozomální podjednotku a genu pro 5,8S rRNA. Mezi podjednotkami jsou vidět regiony ITS1 a ITS2, často spolu s genem pro 5,8S rRNA nazývají pouze oblast ITS. Region IGS, který se nachází jako mezerák mezi jednotlivými klastry rDNA, zde není zakreslen. Šipkami jsou znázorněny různé typy primerů, které se při amplifikaci ITS regionu používají. Převzato a upraveno z Toju et al. (2012).

3.1.1. Mechanismy boje s mutacemi v rDNA

Genetická informace neustále podléhá mutacím. Integrita multigenních úseků DNA, jako je rDNA, zůstává zachována zásluhou hned několika mechanismů.

Jedním je tzv. *birth-and-death evolution*. Principem této formy evoluce je genová duplikace, přičemž vzniklé geny mohou perzistovat i velmi dlouhou dobu. Ostatní geny naopak mohou být inaktivovány nebo dokonce vymazány. Tento mechanismus v rámci rDNA patrně ovlivňuje především geny pro 5S rRNA, které bývají dispergovány po genomu (Rooney et Ward, 2005).

Velmi důležitá je homogenizace kopií u jevu zvaného *concerted evolution*. Základními rekombinačními principy této „evoluce v koncertu“ jsou genová konverze a nerovnoměrný *crossing over* (Ganley et Kobayashi, 2007). Za významnou hnací sílu *concerted evolution* se dají určitě považovat delece rozsáhlých úseků rDNA. Bylo zjištěno, že takto může být odstraněno až 75 % repetice (Ganley et Kobayashi, 2011).

Z hlediska dlouhodobého mají však buňky i velmi účinné amplifikační mechanismy – aby si udržely víceméně stabilní počet repetice. Rekombinací snížený počet kopií se pak může opět zvýšit opravou dvouvláknových zlomů DNA pomocí nerovnoměrné rekombinace sesterských chromatid. V podstatě se jedná o opakovanou replikaci některých repetice (Kobayashi et Ganley, 2005).

3.1.2. Vliv počtu repetice rDNA na biologii houby

Počet kopií rDNA v rámci jednotlivých druhů se může lišit. Již dlouho se ví, že kvantita rDNA koreluje s růstovou rychlostí daného kmene (Maleszka et Clark-Walker, 1990). Kmeny kvasinky *Saccharomyces cerevisiae* s 20 a 110 kopiemi rDNA produkují stejné celkové množství RNA. Ty s nižším počtem kopií jsou však mnohem náchylnější k poškození DNA (Ide et al., 2010).

Z uvedených informací lze vyvodit několik závěrů. U kmenů s malým počtem kopií se vyskytuje vyšší senzitivita k mutačnímu poškození. Výsledkem je, že jsou celkově adaptabilnější, tedy schopné snáze odolávat změnám v rychle se měnícím prostředí. Tento větší evoluční potenciál pak způsobuje vyšší variabilitu mezi jedinci. Ti s více kopiemi mají naopak mohutnější proteosyntetický aparát a tedy i rychlejší růst (r-stratégové). Počet kopií ribozomální DNA tak silně souvisí nejen s životní strategií, ale i s celkovou biologií a *fitness* samotné houby (Kobayashi, 2011).

3.2. Variabilita ITS oblasti

Artefaktům v sekvenčních datech se můžeme snažit nejrozličnějšími způsoby předcházet anebo je *ex post* eliminovat (viz kap. 2.1.4.). Přirozeně se vyskytující variabilitu však ovlivnit nemůžeme. Dokonce i přes všechny zmíněné opravné mechanismy zůstává její část zachována. Existence variability sekvencí mezi druhy se dá považovat téměř za samozřejmou. Na výsledky studií mají mnohem větší vliv variabilita intradruhov (IDV; vnitrodruhov) a intragenomová (IGV; vnitrogenomová). Oblasti genů pro velkou a malou ribozomální podjednotku jsou v rámci nižších taxonomických skupin poměrně konzervativní. Proto se budu dále zabývat především variabilitou ITS oblasti, která hraje v determinaci druhů největší roli. Procentuální hodnoty variability vyjadřují poměr polymorfních míst k celkovému počtu bází.

3.2.1. Intradruhov (IDV) variabilita

V rámci jednotlivých druhů hub se vyskytuje mnoho rozdílů. Četné studie poukazují na lišící se fyziologii uvnitř druhu – např. citlivost na toxické kovy, využití organického či anorganického dusíku aj. (Cairney, 1999). Již dlouho se věnuje určitá pozornost i problematice variability ITS regionu (Kárén et al., 1997; O'Donnell, 1992; O'Donnell et al., 1997).

Pionýrská studie z roku 1997 porovnává intra- a interdruhov (IDV) variabilitu na souboru 44 druhů ektomycorhizních hub pomocí délkového polymorfismu restrikčních fragmentů. Výsledky ukazují IDV jako poměrně nízkou a pouze 16 % druhů vykazovalo polymorfní RFLP vzorec. Naopak byla potvrzena velká mezidruhov (IGV) variabilita (Kárén et al., 1997).

Podobné výsledky variability byly zaznamenány i v několika dalších studiích, které shrnuje tab. 4. Většina z nich byla prováděna na severní polokouli, což také může mít na výsledky určitý vliv. Jak známo, jižní polokoule, obzvláště Austrálie a oblast bývalé Gondwany, se profiluje mnohem výraznější biologickou diverzitou. S tím bude pravděpodobně souviset i masivnější zastoupení polymorfismů a větší genetická variabilita vůbec (Glen et al., 2001).

Doposud zmíněné studie využívaly převážně metodu RFLP. Tato metoda ovšem skýtá mnohá negativa. Například mohou vznikat artefakty neúplnou digestí pomocí restrikčních enzymů. Další možnostmi jsou různé bodové mutace, při kterých substituovaná báze změní restrikční místo enzymu,

který ho poté nerozpozná. Stejně tak může dojít i k inzerci či delecí, kteréžto pozmění restriční místo nebo minimálně změní délku fragmentu (Avis et al., 2006).

Zkoumané houby	Metoda	Počet zkoumaných druhů	Počet variabilních druhů		Percentuální podíl		Reference
Ektomykorhizní	RFLP	44	7		16 %		Kårén et al. (1997)
Ektomykorhizní	RFLP	42	12		29 %		Glen et al. (2001)
Ektomykorhizní	sekvenování	68	27	7 Asc.	39,50 %	10 % Asc.	Smith et al. (2007)
				20 Bas.		29,5 % Bas.	
Houby z databáze INSD	–	4185	3264		78 %		Nilsson et al. (2008)
<i>Erysiphales</i>	sekvenování	6	4		67 %		Kovacs et al. (2011a)

Tab. 4: Výsledky několika náhodně vybraných studií IDV ukazujících počty variabilních druhů včetně jejich percentuálních zastoupení v celkovém počtu studovaných druhů. Z tabulky jsou vidět velké rozdíly mezi výsledky studií využívajících RFLP a sekvenování.

IDV variuje přibližně mezi 0,16–2,85 %. Na druhou stranu byla prokázána 5,7–12,5% variabilita u druhových komplexů, které se posléze ukázaly jako kryptické (Smith et al., 2007). V jiné práci je vážený průměr IDV počítaný ze souboru 4185 druhů z INSD a činí 2,51 % (Nilsson et al., 2008). Průměrné hodnoty IDV pro jednotlivá oddělení hub zachycuje graf 1. Jak bylo zmíněno dříve (Kap. 2.4.2.), hodnoty vnitrodruhové variability by se měly zohledňovat při klastrování sekvencí do OTU.

Na stejném souboru se ukázala také nejednotná variabilita v sekvencích ITS1 a ITS2. ITS2 oblast byla variabilnější pouze v 34 % případů a výsledek tedy potvrzuje převažující názor, že je variabilnější ITS1 (Nilsson et al., 2008).

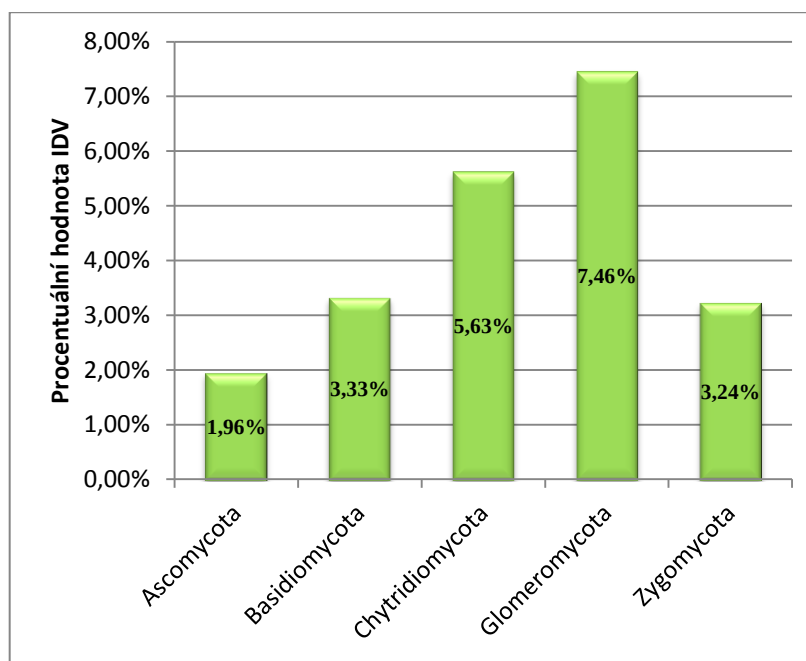
Na výsledky vypočítané z těchto databází nelze úplně spoléhat. Jak známo, nepřesnosti v databázích jsou ohromné. Některé sekvence jsou vyloženě pochybné, jiné chimérické nebo nesprávně identifikované, a to buď kvůli špatné izolaci DNA nebo nesprávné determinaci houby podle vnějších znaků (Bridge et al., 2003).

3.2.1. Intragenomová variabilita

Intragenomová variabilita je považována za nejvíce znepokojující zdroj chyb v popisech taxonů založených na sekvencích DNA (Hibbett et al., 2011).

Až do příchodu celogenomových dat jí však nikdo nepřikládal přílišnou pozornost. Postupně se objevilo několik studií, které IGV pouze zmínily (Fell et al., 2007; Kim et al., 2006; Lachance et al., 2003; Pawlowska et Taylor, 2004).

Při studiu procesů *concerted evolution* a *birth-and-death evolution* ovlivňujících homogenitu repetitivních úseků DNA si autoři všimli většího významu IGV. Právě úroveň IGV je totiž klíčovým rozdílem mezi těmito dvěma formami evoluce (Ganley et Kobayashi, 2007). Rooney et Ward (2005) objevili v genomech několik typů genů pro 5S rRNA podléhajících pravděpodobně *birth-and-death evolution*.



Graf 1: Graf znázorňuje procentuální zastoupení IDV pro jednotlivá oddělení hub. Data jsou spočítaná ze souboru 4185 druhů z databáze INSD. Vytvořeno dle Nilsson et al. (2008).

Ganley et Kobayashi (2007) zjistili, že IGV je u kvasinek dispergována relativně rovnoměrně po celé rDNA. To kontruje logickému předpokladu zvýšeného výskytu IGV v oblastech s nižším selekčním tlakem. Zároveň bylo zastoupení polymorfismů v rámci rDNA poměrně nízké, což by potvrdzovalo homogenizační účinek *concerted evolution* (Ganley et Kobayashi, 2007).

IGV může vznikat např. jako možný důsledek relaxované *concerted evolution* nebo rekombinací během křížení (Kageyama et al., 2007). V říši hub bylo odhaleno poměrně vysoké zastoupení IGV (Alper et al., 2011; James et al., 2009; Kovacs et al., 2011b; Lindner et Banik, 2011; Simon et Weiss, 2008). Potenciální formy výskytu IGV souvisí také s životním cyklem houby. Měly by se proto brát v potaz možné rozdíly v IGV způsobené množstvím genetické informace – haplo- versus diploidní buňky (kvasinky), mono- versus dikaryotické buňky (Ascomycota versus Basidiomycota) nebo mnohoaderné buňky u skupiny Glomeromycota (Simon et Weiss, 2008).

U arbuskulárních mykorhizních hub (Glomeromycota) se obecně vyskytuje velká variabilita (Stockinger et al., 2009). Hyfy těchto hub totiž obsahují stovky (až 5000) jader a navíc má každé z nich kolem 90 kopií rDNA (Lanfranco et al., 1999). Takový počet jader patrně představuje pufrační systém zabráňující přílišné akumulaci mutací. Na druhou stranu u asexuálních jedinců, za

jaké se arbuskulární houby považují, umožňuje neúplná homogenizace jader perzistenci určitého evolučního potenciálu (Pawlowska et Taylor, 2004). Výzkum intraizolátové variability LSU arbuskulárních hub ukazuje, že variabilita přetrvává i do transkripční úrovně – tedy není během transkripce nijak odstraněna (Boon et al., 2010).

Ke zvýšené IGV může přispívat i existence extrachromozomální rDNA ve formě lineárních či cirkulárních plazmidů. Vyskytuje se minimálně u některých kvasinek, a přestože bývá množství plazmidů malé, neměla by se jejich role podceňovat (Simon et Weiss, 2008).

Při studiu kvasinky *Geotrichum candidum* si Alper et al. (2011) všimli dvojitých vrcholů chromatogramu, které by mohly značit IGV. Tuto hypotézu si následně ověřili klonováním, kdy u čtyř kmenů zjistili z 68 klonů celkem 32 variant ITS.

Studie na rodu *Laetiporus* (Basidiomycota) ukázala významnou IGV ≥ 5 % (Lindner et Banik, 2011). Mezi čtyřmi druhy ze skupiny Ascomycota se našla IGV v rozmezí 2,2–3,6 % (Simon et Weiss, 2008). Velké množství polymorfismů se vyskytuje pod úrovní 1 % rozdílnosti sekvencí, které však bývají odstraněny během klastrování (viz kap. 2.4.2.). Několik náhodně vybraných výsledků shrnuje tab. 5.

Skupina hub	Celkový počet	Polymorfní počet	Procentuální podíl	Hodnota IGV	Reference
Více skupin	68 druhů	5 druhů (Basidiomycota)	7 %	0,16–0,55 %	Smith et al. (2007)
Ascomycota, 4 druhy	234 klonů	48 klonů	21 %	2,2–3,6 %	Simon et Weiss (2008)
Ascomycota, 4 druhy	110 jedinců	10 jedinců	9 %	0,15–1,81 %	Kovacs et al. (2011b)

Tab. 5: Výsledky náhodně vybraných studií IGV ukazující procentuální zastoupení IGV s jejími hodnotami v souborech testovaných entit.

Co se týká charakteru polymorfismů, ve všech studiích se vyskytovaly nejen substituční mutace typu tranzice a transverze, ale i mutace inzerční a deleční. Tranzice byly nejčastější, s výskytem 71 % (James et al., 2009) až 90,8 % (Simon et Weiss, 2008) ze všech mutačních událostí. Konkrétně se jednalo hlavně o tranzice typu A-G, které mají malý vliv na sekundární strukturu DNA (Lindner et Banik, 2011). Tedersoo et al. (2010) varují před *singletony* (sekvencemi, které se vyskytují v datasetu pouze jednou), jelikož mohou představovat umělé sekvence. Nalezli pouze 24,4 % *singletonů*, které měly kvalitní sekvenci, ale tyto neodpovídaly žádnému druhu odhalenému pomocí Sangerova sekvenování. Většina *singletonů* dávala vznik domnělým umělým OTU a autoři je proto odstranili. Tato práce byla provedena na ektomykorhizních kořenových špičkách, tedy na substrátu u kterého se nepředpokládá velká diverzita.

Na druhou stranu nám 454 sekvenování umožňuje odhalit tzv. *rare biosphere*. Jedná se o společenství, ve kterých se vedle několika dominantních populací vyskytují stovky až tisíce dalších s velmi málo zastoupenými organismy (Sogin et al., 2006). Ve studiích takovýchto

hyperdiverzitních environmentálních vzorků je užitečné *singletony* nebo málo-kopiové sekvence neodstraňovat, protože pomáhají odhalit pravou diverzitu společenstva. Míra důvěryhodnosti těchto sekvencí stoupá spolu s hloubkou sekvenování. Čím více se získá z pyrosekvenování kopií, tím je větší pravděpodobnost, že se *rare biosphere* hyperdiverzitních substrátů podaří odhalit (Huse et al., 2010).

3.2.2. Fúze vegetativních hyf

Poměrně zajímavá je problematika fúze hyf jako potenciální tvůrce variability. Existují studie negativně nahlízející na fúzi vegetativních hyf u skupiny Glomeromycota (Giovannetti et al., 2003; Pawlowska et Taylor, 2004, 2005) poukazující na možnou vegetativní inkompatibilitu bránící této fúzi. Přesto však někteří autoři tento jev podporují (Bever et Wang, 2005; Croll et al., 2009). V jedné práci vyústilo 46–51 % kontaktů hyf v bezchybné anastomózy. Minimálně jedna jedno-sporová linie, s jádry z geneticky odlišných rodičů, pak dávala vznik života-schopnému potomstvu (Croll et al., 2009).

Ve skupině vláknitých hub je fenomén fúze vegetativních hyf bráný jako všeobecně rozšířený (Glass, 2004). Proběhne-li následně fúze jader, můžeme získat dokonce hyfy alopolyloidní – obsahující genetické sady z více druhů, či tzv. amfihaploidní – také s genetickou informací z parentálních jedinců různého druhu, avšak v haploidním stavu (Clewes et Barbara, 2008; Selosse et Schardl, 2007).

Heteroploidní hybridy již byli potvrzeni v mnoha endofytech. V hybridech dochází k významnému zachytu pozitivních mutací rozšiřujících arzenál toxinů interagujících s mutualistickým či parazitickým vztahem k hostiteli. Pravděpodobně je možná i interchromozomální rekombinace při opětovné haploidizaci. Těmito událostmi se otevírá mnoho možností, jak obměňovat genetickou informaci. S tím pak souvisí i rozšiřování okruhu hostitelů a také udržení si pozitivních změn celkově zvyšujících *fitness* jedince (Selosse et Schardl, 2007).

Nenalezl jsem však žádnou studii, která by kalkulovala s možností alopolyloidních hub při výzkumech variability. Tento fenomén by měl být zohledňován, zejména pokud by se ukázalo jeho masivnější rozšíření.

4. Rozlišení pseudogenů u hub

Vnitrogenomová variabilita může být způsobena přítomností divergentních paralogů. Především u rostlin jsou rozdílné paralogy ITS hojné a jsou tak studovány poměrně často, jelikož mohou představovat vážný problém při rekonstrukci fylogenetických stromů (Buckler et al., 1997). Velmi problematické mohou být především tzv. „hluboké“ paralogy, které se duplikovaly dříve, než došlo

ke speciaci. Vyskytují se tak u několika druhů zároveň a je obtížné je odlišit od ortologů vzniklých právě v průběhu speciace (Bailey, 2003).

Některé tyto paralogy jsou nefunkční a označují se jako pseudogeny. Důkaz jejich významnosti demonstrovali Mayol et Rosselló (2001) porovnáním dvou systematických studií na dubech. Výsledkem obou studií byly rozdílné fylogenetické stromy, což bylo pravděpodobně způsobeno právě neodhalením pseudogenových sekvencí v jedné z nich.

U hub byly popsány divergentní paralogy pouze v několika případech. V roce 1997 byl zjištěn výskyt dvou typů ITS2 oblasti u komplexu *Gibberella fujikuroi* (Ascomycota) (O'Donnell et Cigelnik, 1997). Tři, nejspíše funkční, paralogy byly nalezeny u *Trichaptum abietinum* (Basidiomycota) (Ko et Jung, 2002) a několik typů ITS objevili i Kovacs et al. (2011a) u řádu *Erysiphales* (Ascomycota).

Dále existují studie (Boon et al., 2010; Corradi et al., 2007; Lafontaine et Dujon, 2010; Lindner et Banik, 2011; Rajashekar et al., 2007; Rydholm et al., 2007), které domnělou existenci pseudogenů u hub zmiňují, ale více se jí nezabývají.

Práce, která by provedla důkladnou analýzu vztahu ITS oblasti, pseudogenů a snad i intragenomové variability u hub ale zřejmě zatím chybí. Byly však prozkoumány některé jiné lokusy. Na kvasinkách (Ascomycota) byly systematicky vyhledávány všechny potenciální pseudogeny. Mezi osmi zkoumanými druhy bylo nalezeno celkem 871 pseudogenů v rozmezí 38–230 pseudogenů na genom. Proporčně to odpovídá 0,7–3,6 % udávaných kódujících sekvencí. Celkem 62 pseudogenů bylo nalezeno u více kmenů zároveň, což podporuje hypotézu, že k jejich vzniku došlo již u společného předka (Lafontaine et Dujon, 2010). Orientační zastoupení pseudogenů u hub vzhledem k počtu funkčních genů ukazuje tab. 6.

4.1. Původ pseudogenů

Předpokládá se, že většina pseudogenů vzniká genovou duplikací. Když dojde v paralogu k fatální mutaci znemožňující kódování proteinu, vzniká pseudogen (Rooney et Ward, 2005). Další možností je, že *concerted evolution* nefunguje stoprocentně a ribozomální lokus začne kumulovat mutace až také vznikne pseudogen (Feliner et Rosselló, 2007). Pseudogeny mohou vznikat i ze *single-copy* genů – opět mutací způsobující ztrátu funkčnosti. Velmi zajímavá je teorie duplikace pomocí retrotranspozice (Lafontaine et Dujon, 2010; Rooney et Ward, 2005).

Kromě toho, že 5S geny (*Aspergillus*, *Fusarium*, *Magnaporthe*, *Neurospora*) a SINE mají stejný typ vnitřního promotoru RNA polymerázy III, je navíc jeden typ SINE pravděpodobně odvozený od produktu fúze LINE a 5S rDNA (Rooney et Ward, 2005). Lidské LINE retrotranspozony umí využít vlastní enzymatický aparát k transpozici transkriptu cizích genů a jako pseudogeny je integrovat do jiného místa genomu (Esnault et al., 2000). To nabízí možnost, že by i tyto houbové entity mohly interagovat – 5S geny spolu s LINE či jinými „retroelementy“ –

a využívat tak jejich enzymatický aparát (Rooney et Ward, 2005). Transpozice na jiné místo pak způsobí to, že gen již není ovlivňován homogenizačními účinky *concerted evolution*, a mohou se v něm hromadit mutace (Glass et al., 2013). U druhu *Magnaporthe grisea* (Ascomycota) byly dokonce objeveny trojitě DNA chiméry, nejspíše vzniklé právě pomocí reverzní transkripce řízené LINE (Gogvadze et al., 2007).

Gen	Druh	Celkový počet kopií	Počet pseudogenů	Percentuální podíl	Reference
5S	<i>Aspergillus nidulans</i>	52	16	31 %	Rooney et Ward (2005)
5S	<i>Magnaporthe grisea</i>	71	33	46 %	Rooney et Ward (2005)
5S	<i>Fusarium graminearum</i>	66	7	11 %	Rooney et Ward (2005)
5S	<i>Neurospora crassa</i>	92	14	15 %	Rooney et Ward (2005)
hyd	<i>Paxillus involutus</i>	8	1	13 %	Rajashekar et al. (2007)

Tab. 6: Zastoupení kopií určitých genů a jejich paralogních pseudogenů u vybraných druhů hub. Gen *hyd* představuje gen pro hydrofobin, ostatní jsou 5S geny rDNA.

Výsledky ukazují, že tvorba pseudogenů souvisí s rychlostí celkového genového *turnoveru*, tedy s rychlostí, jakou se v buňkách obměňuje genetická informace (Rooney et Ward, 2005). To dokazuje také fakt, že se jich hodně nachází v subtelomerických oblastech. Toto místo je známé post-transkripčním umlčováním genů, redukcí aktivních genů a obecně odkládáním genových reliktů (Fairhead et Dujon, 2006; Lafontaine et Dujon, 2010). Pseudogeny v těchto oblastech 3–16krát převyšují počty aktivních genů (Lafontaine et Dujon, 2010).

V některých případech se potvrdily pseudogeny s určitou funkcí, jako je především regulace genové exprese a tvorba genetické diverzity (Lafontaine et Dujon, 2010). Transkript pseudogenu, jenž už se nemůže dále translatovat, někdy dokáže alespoň ovlivnit stabilitu mRNA svého kódujícího paralogu a tím ho regulovat (Hirotsume et al., 2003).

4.2. Vyhledávání pseudogenů

Pseudogeny se kvůli množství naakumulovaných mutací profilují ve fylogenetických stromech jako dlouhé větve. To znesnadňuje správnou interpretaci fylogramu (Xiao et al., 2010). Obtížně odlišitelné a tím pádem znepokojující mohou být i NUMT – transponovatelné geny mitochondriálního původu komplementární k těm jaderným. Stejně potíže mohou činit i pseudogeny s malým počtem mutací, tedy vzniklé v nedávné historii. Velmi podobné paralogy by však neměly mít na vyhodnocování velký vliv – alespoň ne při rozlišování vyšších taxonů (Glass et al., 2013). Ve fylogenezi naopak mohou dobře posloužit pseudogeny s vysokou divergencí, a to především při odlišování velmi blízkých druhů, kde ostatní funkční paralogy selhávají (Xiao et al.,

2010). Problém je v tom, že spolu mohou jednotlivé paralogy rekombinovat (Buckler et al., 1997), a to dokonce i funkční gen spolu s pseudogenem (Alvarez et Wendel, 2003).

Pseudogeny se dají determinovat např. pomocí délkové variability, protože jsou často zkrácené. Kratší délka má zprostředkovaně vliv i na sekundární strukturu (Rooney et Ward, 2005). Sekundární struktura ITS nutná ke správnému utvoření vyzrálé rRNA je evolučně velmi omezená (Feliner et Rosselló, 2007). ITS1 a ITS2 mají velkou roli ve zpracovávání rRNA při ribozomální biogenezi. Dá se usuzovat, že toto funkční omezení vytváří silný evoluční tlak na zachování interakcí mezi bázemi a konzervace určitých domén (Mullineux et Hausner, 2009). Po analýze těchto oblastí jsme tedy schopni přibližně říct, zda bude daná sekvence kódovat funkční RNA či nikoliv (Glass et al., 2013).

Nosným pilířem stability sekundární struktury jsou GC páry. Pseudogeny a chiméry by měly mít, kvůli absenci *concerted evolution*, jejich obsahy v ITS1 a ITS2 divergentní. To je podstatný rozdíl oproti funkčním genům pod selekčním tlakem (Glass et al., 2013). Udržení GC bohatých oblastí umožňují především *compensatory base changes* (= CBC; G–C na A–U a naopak), neboli změny, které nenarušují stonek. Dalšími pomocnými událostmi jsou hemi-CBC (G–C na G–U a naopak), kompenzační inserce či delece a potenciálně i *slippage* (čili sklouznutí) RNA řetězce. Poměr GC párů se ukazuje mnohonásobně vyšším v oblastech ITS1 a ITS2, na rozdíl od 5,8S – což také ukazuje důležitost jejich sekundární struktury (Mullineux et Hausner, 2009). Vzhledem k četným mutacím a tedy i CBC v ITS mohou vznikat homoplazie vedoucí k synapomorfii a ty opět až k mylné interpretaci výsledků (Alvarez et Wendel, 2003). Pseudogeny můžeme rozeznávat i pomocí primární struktury – především kvůli velkému zastoupení substitucí spojených s metylací. Vzhledem k velké mutabilitě metylovaných dinukleotidových ostrůvků CpG, nalezneme v pseudogenech hojně jejich deaminované protějšky thymin a adenin (Buckler et al., 1997).

5. Mykologické práce využívající 454 sekvenování

V této kapitole bych rád přiblížil mykologické práce, které využívaly 454 sekvenování a nastínil jejich případné nedostatky.

S tím, jak jsou v posledních letech produkována kvanta NGS dat, stoupá pochopitelně i množství chyb v databázích. Může za to především nedisciplinovanost autorů, kteří nejsou schopni dodržovat několik základních pravidel. Protože veřejná databáze INSD závisí právě na datech nahraných autory, bylo potřeba ke zlepšení kvality databáze zavést určité standardy. Směrnice MIENS obsahuje základní informace o markerových genech, které by se měly u environmentálních vzorků udávat (Glöckner et al., 2010). Velká část dat z INSD je identifikována špatně nebo latinské binomické pojmenování postrádá úplně. Většině položek navíc chybí důležitá metadata, jež jsou v MIENS zmíněna. Přitom se jedná o základní údaje, které je logické ke vzorku přidávat. Jako příklady mohou být uvedeny: detailní informace o původu vzorku (stát a lokalita,

včetně zeměpisných souřadnic), ovlivňovaný organismus (u patogenů či mykorhizních jedinců) anebo samotný charakter vzorku – zda se jednalo o plodnici, hyfy apod. Chybět by pochopitelně neměly ani údaje o autorovi, datu sběru atp. (Glöckner et al., 2010; Ryberg et al., 2009; Tedersoo et al., 2011)

5.1. Metodické chyby

Jak bylo uvedeno dříve (Kap. 3.2.), rozšíření IDV a IGV mezi houbami je poměrně velké. Přítomnost variability může mít mnoho důsledků pro fylogenetiku, mikrobiální ekologii či diagnostiku patogenů pomocí metod využívajících DNA. Navíc hrozí nezanedbatelné ovlivnění výsledků klastrování OTU a tím pádem i determinace druhů nebo určování diverzity (Smith et al., 2007). S tímto problémem souvisí také rozdíly mezi přímým sekvenováním a metodami „nepřímými“ (Kap. 2.3.).

Studie rodu *Laetiporus* např. ukazuje, že při použití výsledků ze sekvenování klonů může dojít k až trojnásobnému nadhodnocení údajů proti přímému konsenzuálnímu sekvenování (Lindner et Banik, 2011). Je důležité si uvědomit, že konsenzus je pouhý teoretický konstrukt. V nejextrémnějším případě, kdy by se všechny klony vyznačovaly malými a rovnoměrně rozprostřenými polymorfismy, by žádný z reálných klonů nemusel tomuto konsenzu vůbec odpovídat (Lindner et Banik, 2011).

Všechny zmíněné aspekty je třeba zvážit již při experimentálním designu – aby se vybraly ty nejvhodnější metody.

5.2. Nalezené práce

Nejdříve bych rád nastínil několik základních rad pro budoucí autory, díky nimž je snazší plně porozumět provedeným pokusům a navíc je to umožní znovu zopakovat v budoucnu. Kolektiv významných autorů, Nilsson et al. (2011), popsal vhodný postup pro výzkum hub pomocí NGS metod následovně:

Aby bylo možné prezentované výsledky *ex post* ověřovat, je vhodné důležitá data deponovat do centrálního archivu. Kromě toho by měla být k dispozici i doplňková data, kupř. na autorových stránkách, spolu s odkazem na původní článek. K dispozici by měla být původní, hrubá, 454 data a dále pak informace o jednotlivých krocích – počínaje filtrováním a ořezáváním sekvencí, až po *denoising* a údaje o počtu vyřazených sekvencí spolu s důvody. Vše, i pro jednotlivé vzorky, by mělo být dostupné v doplňkových datech ve formátu FASTA. Dále je potřeba připojit podrobné informace o zvoleném genetickém markeru, typu klastrování, použitých programech a o případném dalším nakládání s danými taxonomickými jednotkami (Nilsson et al., 2011).

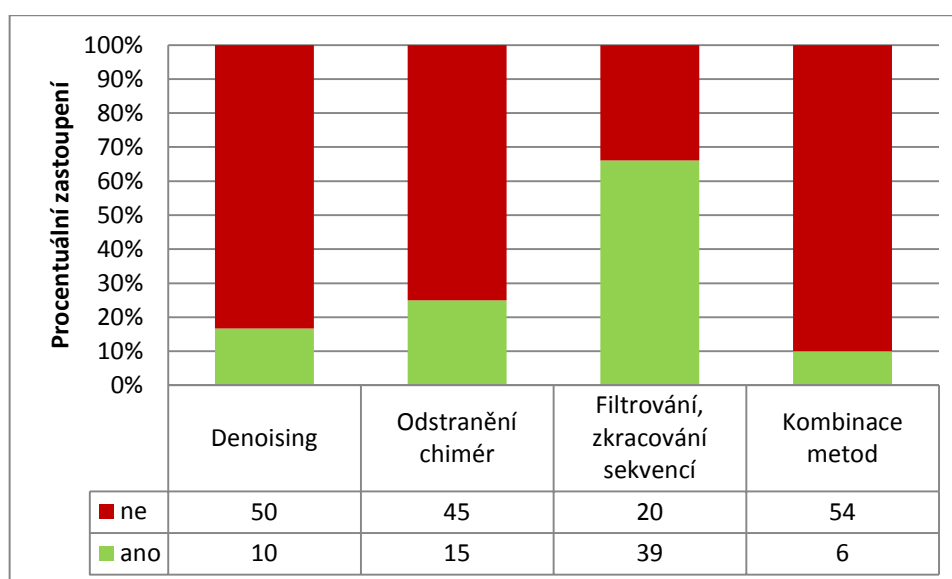
Vytvořil jsem přehledovou tabulku (Kap. 8.) z celkem 60 prací, které používaly 454 sekvenování ke studiu hub. Tyto práce jsou dohledatelné pouze do roku 2009, ale v jejich počtu je

zřejmá vzrůstající tendence. Při výběru jsem se inspiroval tabulkou vytvořenou týmem Větrovský et Baldrian (2013), ze které jsem některá data převzal a dále je rozšířil. Snažil jsem se vybírat práce z hyperdiverzitních substrátů, protože by mohly být zatíženy největší chybou. Z celkového počtu se osm prací nezmiňovalo o rozlišování diverzity, zatímco 52 ji nějakým způsobem studovalo. Z těchto prací jsem vybral 19 takových, které uváděly ukazatel diverzity Chao1 (popř. Chao2 nebo ACE) a z nich jsem pro větší přehlednost vytvořil samostatnou tabulku 2 (Kap. 8.). Chao1 je při porovnávání s počty získaných OTU velmi vypovídající a umožňuje tak lépe práce komplexně analyzovat. U informací o *denoisingu*, odstraňování chimér a taxonomickém zařazování jsem se pokusil přiblížit i použitý algoritmus či program dle toho, co autoři zrovna uvedli.

5.3. Diskuze výsledků

Vzhledem k silné heterogenitě prací, je velmi obtížné vyvozovat ucelené závěry. Mnoho uváděných informací je navíc neúplných, či nejednoznačně formulovaných a závěry zde uváděné jsou tak jenom orientační.

Pouze jedna jediná práce udává, že použili filtrování a zkracování sekvencí, *denoising*, odstranění chimér a odstranění *singletonů*. 10 % prací pak aplikovalo současně všechny tři kategorie purifikace získaných sekvencí – jak je vidět v grafu 2.



Graf 2: Ukázka zastoupení jednotlivých metod upravujících získané sekvence v prozkoumaných pracích.

Porovnávat celkové počty získaných OTU mezi sebou je obtížné kvůli rozdílným vztažným soustavám. Některé práce udávají informace vztažené na celý dataset, jiné pouze na vzorek. To ale nebrání možnosti porovnávat výsledky jednotlivých studií spolu se zmíněným ukazatelem Chao1. Tento ukazatel se snaží odhadnout reálnou diverzitu v daném souboru. Porovnáním se získaným

počtem OTU tak lze konstatovat, nakolik daná studie dokázala odhalit diverzitu. I tento odhad však může být zatížený chybou a pro přesnější závěry je třeba zohlednit více ukazatelů diverzity zároveň.

Některé studie udávají navíc informace o počtu OTU, ze kterých byly vynechány *singletons* nebo málo-kopiové sekvence. Ve většině případů vedou tato čísla k věrohodnějším výsledkům, nejspíše vzhledem k všeobecné absenci metod purifikace sekvencí. Zda by se *singletons* měly odstraňovat či nikoliv je složitá otázka. Některé studie dokládají, že se může jednat o umělé sekvence (Tedesoo et al., 2010). Jindy by však mohlo po jejich odstranění dojít právě k neúplnému odhalení skutečné diverzity (Kap. 3.2.2.). Je možné, že dataset obsahuje *singletons* pouze z důvodu malé hloubky sekvenování. Kdyby se celkově získalo více sekvencí, možná by se již nejednalo o *singletons* a vyskytovaly by se i tyto sekvence čteněji. Někdy sice autoři udávají počty OTU bez *singletonů*, nicméně nezmiňují, že by je ze závěrečných výsledků odstranili. Takto se patrně zachovali především ti, jejichž výsledky se zahrnutými *singletons* dosahovaly malého podílu odhadnuté diverzity pomocí Chao1. To by totiž mohlo znamenat právě skutečnost, že *singletons* představují sekvence málo abundantních druhů.

Na základě výsledků lze říci, že ve většině případů byl získán neuspokojivý počet sekvencí pro odhalení celkové diverzity. Vzorky nebyly sekvenovány dostatečně do hloubky a nebyla tak odhalena *rare biosphere* (Kap. 3.2.2.). Toto stanovisko dokládá u některých studií ukazatel *rarefaction* (zde neuvedeno). Porovnáním tohoto ukazatele s počty získaných sekvencí je možno stanovit, zda bylo dosaženo saturace – tedy *plató* fáze *rarefaction* křivky.

Do tab. 7 jsem vybral čtyři studie umožňující vzájemné porovnání nalezené diverzity. Všechny tyto práce ukazují počty OTU z celého datasetu a studovaly půdní houby. Bohužel byla množství odebíraných vzorků rozdílná a čísla není možné přesně vztáhnout na stejný objem půdy z důvodu nelineární funkce přírůstku počtu OTU vůči množství vzorku. Nicméně lze pozorovat velkou diverzitu půdních hub, dosahující řádově stovek až tisíců OTU na gram půdy. Dále jsem porovnal počty získaných OTU s *rarefaction* křivkami (data neuvedena). Žádná z těchto studií neodhalila skutečnou diverzitu. Je to způsobeno zřejmě tím, že nebylo provedeno dostatečně hluboké sekvenování, aby mohla být celková diverzita odhalena. Z tabulky je patrná korelace mezi počty sekvencí a OTU, která naznačuje, že by se při větším množství sekvencí mělo dosáhnout i většího množství OTU. Pro tyto odhady jsou ale určitě vhodnější zmíněné *rarefaction* křivky, jejichž tvar detailně vypovídá o saturaci diverzity při daném počtu sekvencí.

Některé studie potvrzují divergenci v počtech OTU při použití různých primerů, jak bylo nastíněno dříve (Kap. 3.1.). Převažující hladina pro klastrování sekvencí do OTU byla 97 % – v celkovém počtu 37 prací, tedy přibližně v 63 %. Nezanedbatelným faktorem ovlivňujícím výsledky je také délka sekvencí po ořezání – např. Monard et al. (2013) prezentovali lišící se výsledky pro dvě různé minimální délky sekvencí.

Ze všeho výše uvedeného je jasné, že by se autoři měli pokoušet o získání co největšího počtu sekvencí, protože pouze to umožní přiblížit se odhadům reálné diverzity. Dále by si měli nechat více

záležet na dodržování metod, které zpřesňují dosažené výsledky a odstraňují výsledky chybné. Aby nedocházelo ke zbytečným misinterpretacím výsledků, bylo by užitečné věnovat se srozumitelnému popisu pracovního postupu a uvádění všech podrobných informací.

Množství vzorku	Počet sekvencí *	Počet OTU*	Odhadnutá diverzita*	Hladina	Reference
250 mg SV	205 131	2 847	12 000	98 %	Lentendu et al. (2011)
500 mg	ITS1F: 4 767 ; ITS4: 5 551	119 ; 115	155 ; 138	98 % Asco. / 97 % Bas.	Arfi et al. (2012a)
1 000 mg SV	8 945	456 (262 NS)	689–784	97 %	Menkis et al. (2012)
1 000 mg	15 408	1 655	3 310	97 %	Wubet et al. (2012)

Tab. 7: Demonstrace závislosti počtů OTU na množství pořízených sekvencí a na odhadu diverzity pomocí Chao1.

* U práce Arfi et al. (2012a) jsou uvedeny sekvenační primery. Bližší informace o primerech jsou v kap. 8. Pořadí počtů OTU v dalších sloupcích zůstává zachováno. Vysvětlivky: SV: suchá váha; NS: vynechány *singletony*

6. Závěr

Cílem této práce bylo objasnit problematiku intragenomové variability ITS oblasti u hub a osvětlit její důsledky pro výzkum diverzity hub. Pro lepší uvedení do problematiky jsem postupně shrnul základní principy 454 pyrosekvenování včetně jeho vyhodnocování, poznatky o ITS oblasti u hub a informace o intradruhovém a intragenomovém variabilitě rDNA.

V závěrečné části jsem představil přehledovou tabulku mykologických prací, které použily 454 sekvenování. Při prozkoumání dat z tabulky je patrná diskutabilní váha výsledků u některých studií. Autoři nevyužili všechny dostupné možnosti ke zlepšení přesnosti výsledků a navíc mnohdy udali neúplné či matoucí informace. Vezmou-li se v potaz sepsané informace o intragenomové variabilitě a pyrosekvenování, může mít tato nedůslednost ve zpracovávání sekvencí ohromný vliv na odhady diverzity. Kromě toho je zarážející absence celkového cytogenetického pohledu na IGV. Při výzkumech IGV by se měly zohledňovat i informace na pozadí genetické variability jako je množství jader, ploidie, či lokace genů na chromozomech. Nezanedbatelný vliv při získávání sekvencí z environmentálních vzorků mají také rozdílné počty repetitivní rDNA mezi jednotlivými skupinami (Kap. 3.1.). Všechny tyto aspekty jsou studovány hojně u rostlin, u hub však nikoliv.

Všeobecně v odhadech diverzity stojí naproti sobě dva odlišné problémy. Prvním je nadhodnocování počtu OTU z důvodu zavedených chyb nebo přítomnosti intragenomové variability. Na druhé straně jsou podhodnocené počty z důvodu nevhodně designovaného

klastrování či nedostatečné hloubky sekvenování. Obě stinné stránky jsou ve studiích reálně přítomné a hrají tak v odhadech diverzity roli. Jelikož o globálním vlivu IGV mezi houbami se toho zatím příliš neví, rád bych navázal na tuto řešerši svojí diplomovou prací. Plánuji v ní porovnat vlastní výsledky přímého sekvenování a pyrosekvenování na známém datasetu. Bude třeba zvolit dostatečně rozsáhlý okruh houbových skupin, aby byl obsažen co možná nejširší rozsah potenciální variability, která by se v environmentálních vzorcích mohla vyskytovat. Z těchto výsledků pak vyhodnotím její rozšíření mezi jednotlivými skupinami a odvodím chybu v odvození diverzity jako její důsledek.

7. Použitá literatura

- Alper, I., Frenette, M., et Labrie, S. (2011). Ribosomal DNA polymorphisms in the yeast *Geotrichum candidum*. *Fungal Biology* 115, 1259–1269.
- Alvarez, I., et Wendel, J.F. (2003). Ribosomal ITS sequences and plant phylogenetic inference. *Mol. Phylogenet. Evol.* 29, 417–434.
- Amend, A.S., Seifert, K.A., Samson, R., et Bruns, T.D. (2010a). Indoor fungal composition is geographically patterned and more diverse in temperate zones than in the tropics. *PNAS*.
- Amend, A.S., Seifert, K.A., et Bruns, T.D. (2010b). Quantifying microbial communities with 454 pyrosequencing: does read abundance count? *Molecular Ecology* 19, 5555–5565.
- Amend, A.S., Barshis, D.J., et Oliver, T.A. (2012). Coral-associated marine fungi form novel lineages and heterogeneous assemblages. *ISME J* 6, 1291–1301.
- Arezi, B., Xing, W., Sorge, J.A., et Hogrefe, H.H. (2003). Amplification efficiency of thermostable DNA polymerases. *Anal. Biochem.* 321, 226–235.
- Arfi, Y., Marchand, C., Wartel, M., et Record, E. (2012a). Fungal diversity in anoxic-sulfidic sediments in a mangrove soil. *Fungal Ecology* 5, 282–285.
- Arfi, Y., Buée, M., Marchand, C., Levasseur, A., et Record, E. (2012b). Multiple markers pyrosequencing reveals highly diverse and host-specific fungal communities on the mangrove trees *Avicennia marina* and *Rhizophora stylosa*. *FEMS Microbiology Ecology* 79, 433–444.
- Arnold, A.E., et Lutzoni, F. (2007). Diversity and host range of foliar fungal endophytes: are tropical leaves biodiversity hotspots? *Ecology* 88, 541–549.
- Avis, P.G., Dickie, I.A., et Mueller, G.M. (2006). A “dirty” business: testing the limitations of terminal restriction fragment length polymorphism (TRFLP) analysis of soil fungi. *Molecular Ecology* 15, 873–882.
- Avis, P.G., Branco, S., Tang, Y., et Mueller, G.M. (2010). Pooled samples bias fungal community descriptions. *Molecular Ecology Resources* 10, 135–141.
- Bailey, C. (2003). Characterization of angiosperm nrDNA polymorphism, paralogy, and pseudogenes. *Molecular Phylogenetics and Evolution* 29, 435–455.
- Baldrian, P., Kolařík, M., Štursová, M., Kopecký, J., Valášková, V., Větrovský, T., Žifčáková, L., Šnajdr, J., Řídl, J., Vlček, Č., et al. (2012). Active and total microbial communities in forest soil are largely different and highly stratified during decomposition. *ISME J* 6, 248–258.
- Bao, E., Jiang, T., Kaloshian, I., et Girke, T. (2011). SEED: efficient clustering of next-generation sequences. *Bioinformatics* 27, 2502–2509.

- Barnes, W.M. (1994). PCR amplification of up to 35-kb DNA with high fidelity and high yield from lambda bacteriophage templates. *Proc Natl Acad Sci U S A* *91*, 2216–2220.
- Bazinet, A.L., et Cummings, M.P. (2012). A comparative evaluation of sequence classification programs. *BMC Bioinformatics* *13*, 1–13.
- Bazzicalupo, A.L., Bálint, M., et Schmitt, I. (2013). Comparison of ITS1 and ITS2 rDNA in 454 sequencing of hyperdiverse fungal communities. *Fungal Ecology* *6*, 102–109.
- Becklin, K.M., Hertweck, K.L., et Jumpponen, A. (2012). Host identity impacts rhizosphere fungal communities associated with three alpine plant species. *Microb. Ecol.* *63*, 682–693.
- Bever, J.D., et Wang, M. (2005). Arbuscular mycorrhizal fungi: Hyphal fusion and multigenomic structure. *Nature* *433*, E3–E4.
- Bik, H.M., Sung, W., De Ley, P., Baldwin, J.G., Sharma, J., Rocha-Olivares, A., et Thomas, W.K. (2012). Metagenetic community analysis of microbial eukaryotes illuminates biogeographic patterns in deep-sea and shallow water sediments. *Mol Ecol* *21*, 1048–1059.
- Bjorbækmo, M.F., Carlsen, T., Brysting, A., Vrålstad, T., Høiland, K., Ugland, K.I., Geml, J., Schumacher, T., et Kauserud, H. (2010). High diversity of root associated fungi in both alpine and arctic *Dryas octopetala*. *BMC Plant Biology* *10*, 244.
- Blaalid, R., Carlsen, T., Kumar, S., Halvorsen, R., Ugland, K.I., Fontana, G., et Kauserud, H. (2012). Changes in the root associated fungal communities along a primary succession gradient analysed by 454 pyrosequencing. *Molecular Ecology* *21*, 1897–1908.
- Blaalid, R., Kumar, S., Nilsson, R.H., Abarenkov, K., Kirk, P.M., et Kauserud, H. (2013). ITS1 versus ITS2 as DNA metabarcodes for fungi. *Mol. Ecol. Resour.* *13*, 218–224.
- Boon, E., Zimmerman, E., Lang, B.F., et Hijri, M. (2010). Intra-isolate genome variation in arbuscular mycorrhizal fungi persists in the transcriptome. *Journal of Evolutionary Biology* *23*, 1519–1527.
- Boyle, B., Hamelin, R.C., et Seguin, A. (2005). In vivo monitoring of obligate biotrophic pathogen growth by kinetic PCR. *Appl Environ Microbiol* *71*, 1546–1552.
- Bridge, P.D., Roberts, P.J., Spooner, B.M., et Panchal, G. (2003). On the unreliability of published DNA sequences. *New Phytologist* *160*, 43–48.
- Buckler, E.S., Ippolito, A., et Holtsford, T.P. (1997). The evolution of ribosomal DNA divergent paralogues and phylogenetic implications. *Genetics* *145*, 821–832.
- Buée, M., Reich, M., Murat, C., Morin, E., Nilsson, R.H., Uroz, S., et Martin, F. (2009). 454 Pyrosequencing analyses of forest soils reveal an unexpectedly high fungal diversity. *New Phytologist* *184*, 449–456.
- Cai, Y., et Sun, Y. (2011). ESPRIT-Tree: hierarchical clustering analysis of millions of 16S rRNA pyrosequences in quasilinear computational time. *Nucl. Acids Res.*
- Cairney, J.W.G. (1999). Intraspecific physiological variation: implications for understanding functional diversity in ectomycorrhizal fungi. *Mycorrhiza* *9*, 125–135.
- Clewes, E., et Barbara, D.J. (2008). Two allopolyploid ascomycete fungal plant pathogens were not rescued by vertical transmission. *New Phytol.* *177*, 583–585.
- Cordier, T., Robin, C., Capdevielle, X., Desprez-Loustau, M.-L., et Vacher, C. (2012). Spatial variability of phyllosphere fungal assemblages: genetic distance predominates over geographic distance in a European beech stand (*Fagus sylvatica*). *Fungal Ecology* *5*, 509–520.
- Corradi, N., Croll, D., Colard, A., Kuhn, G., Ehinger, M., et Sanders, I.R. (2007). Gene copy number polymorphisms in an arbuscular mycorrhizal fungal population. *Appl. Environ. Microbiol.* *73*, 366–369.
- Croll, D., Giovannetti, M., Koch, A.M., Sbrana, C., Ehinger, M., Lammers, P.J., et Sanders, I.R. (2009). Nonspecific vegetative fusion and genetic exchange in the arbuscular mycorrhizal fungus *Glomus intraradices*. *New Phytologist* *181*, 924–937.

- Daghino, S., Murat, C., Sizzano, E., Girlanda, M., et Perotto, S. (2012). Fungal diversity is not determined by mineral and chemical differences in serpentine substrates. *PLoS ONE* 7, e44233.
- Danielsen, L., Thürmer, A., Meinicke, P., Buée, M., Morin, E., Martin, F., Pilate, G., Daniel, R., Polle, A., et Reich, M. (2012). Fungal soil communities in a young transgenic poplar plantation form a rich reservoir for fungal root communities. *Ecology and Evolution* 2, 1935–1948.
- Davey, M.L., Heegaard, E., Halvorsen, R., Ohlson, M., et Kauserud, H. (2012). Seasonal trends in the biomass and structure of bryophyte-associated fungal communities explored by 454 pyrosequencing. *New Phytologist* 195, 844–856.
- Davey, M.L., Heegaard, E., Halvorsen, R., Kauserud, H., et Ohlson, M. (2013). Amplicon-pyrosequencing-based detection of compositional shifts in bryophyte-associated fungal communities along an elevation gradient. *Molecular Ecology* 22, 368–383.
- Davison, J., Öpik, M., Zobel, M., Vasar, M., Metsis, M., et Moora, M. (2012). Communities of arbuscular mycorrhizal fungi detected in forest soil are spatially heterogeneous but do not vary throughout the growing season. *PLoS ONE* 7, e41938.
- Debode, J., Van Hemelrijck, W., Baeyen, S., Creemers, P., Heungens, K., et Maes, M. (2009). Quantitative detection and monitoring of *Colletotrichum acutatum* in strawberry leaves using real-time PCR. *Plant Pathology* 58, 504–514.
- Dolnik, V. (1999). DNA sequencing by capillary electrophoresis (review). *J. Biochem. Biophys. Methods* 41, 103–119.
- Drell, T., Lillsaar, T., Tummeleht, L., Simm, J., Aaspõllu, A., Väin, E., Saarma, I., Salumets, A., Donders, G.G.G., et Metsis, M. (2013). Characterization of the vaginal micro- and mycobiome in asymptomatic reproductive-age Estonian women. *PLoS ONE* 8, e54379.
- La Duc, M.T., Vaishampayan, P., Nilsson, H.R., Torok, T., et Venkateswaran, K. (2012). Pyrosequencing-derived bacterial, archaeal, and fungal diversity of spacecraft hardware destined for Mars. *Appl Environ Microbiol* 78, 5912–5922.
- Dumbrell, A.J., Ashton, P.D., Aziz, N., Feng, G., Nelson, M., Dytham, C., Fitter, A.H., et Helgason, T. (2011). Distinct seasonal assemblages of arbuscular mycorrhizal fungi revealed by massively parallel pyrosequencing. *New Phytologist* 190, 794–804.
- Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461.
- Esnault, C., Maestre, J., et Heidmann, T. (2000). Human LINE retrotransposons generate processed pseudogenes. *Nat Genet* 24, 363–367.
- Fairhead, C., et Dujon, B. (2006). Structure of *Kluyveromyces lactis* subtelomeres: duplications and gene content. *FEMS Yeast Research* 6, 428–441.
- Feliner, G.N., et Rosselló, J.A. (2007). Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants. *Molecular Phylogenetics and Evolution* 44, 911–919.
- Fell, J.W., Scorzetti, G., Stätzell-Tallman, A., et Boundy-Mills, K. (2007). Molecular diversity and intragenomic variability in the yeast genus *Xanthophyllomyces*: the origin of *Phaffia rhodozyma*? *FEMS Yeast Research* 7, 1399–1408.
- Fu, L., Niu, B., Zhu, Z., Wu, S., et Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28, 3150–3152.
- Ganley, A.R.D., et Kobayashi, T. (2007). Highly efficient concerted evolution in the ribosomal DNA repeats: Total rDNA repeat variation revealed by whole-genome shotgun sequence data. *Genome Res* 17, 184–191.
- Ganley, A.R.D., et Kobayashi, T. (2011). Monitoring the rate and dynamics of concerted evolution in the ribosomal DNA repeats of *Saccharomyces cerevisiae* using experimental evolution. *Mol Biol Evol* 28, 2883–

Garber, R.C., Turgeon, B.G., Selker, E.U., et Yoder, O.C. (1988). Organization of ribosomal RNA genes in the fungus *Cochliobolus heterostrophus*. *Curr. Genet.* 14, 573–582.

Geml, J., Laursen, G.A., Timling, I., Mcfarland, J.M., Booth, M.G., Lennon, N., Nusbaum, C., et Taylor, D.L. (2009). Molecular phylogenetic biodiversity assessment of arctic and boreal ectomycorrhizal *Lactarius Pers.* (*Russulales*; Basidiomycota) in Alaska, based on soil and sporocarp DNA. *Molecular Ecology* 18, 2213–2227.

Ghannoum, M.A., Jurevic, R.J., Mukherjee, P.K., Cui, F., Sikaroodi, M., Naqvi, A., et Gillevet, P.M. (2010). Characterization of the oral fungal microbiome (mycobiome) in healthy individuals. *PLoS Pathog* 6, e1000713.

Ghodsi, M., Liu, B., et Pop, M. (2011). DNACLUSt: accurate and efficient clustering of phylogenetic marker genes. *BMC Bioinformatics* 12, 271.

Giovannetti, M., Sbrana, C., Strani, P., Agnolucci, M., Rinaudo, V., et Avio, L. (2003). Genetic diversity of isolates of *Glomus mosseae* from different geographic areas detected by vegetative compatibility testing and biochemical and molecular analysis. *Appl. Environ. Microbiol.* 69, 616–624.

Glass, N. (2004). Hyphal homing, fusion and mycelial interconnectedness. *Trends in Microbiology* 12, 135–141.

Glass, D.J., Takebayashi, N., Olson, L.E., et Taylor, D.L. (2013). Evaluation of the authenticity of a highly novel environmental sequence from boreal forest soil using ribosomal RNA secondary structure modeling. *Molecular Phylogenetics and Evolution* 67, 234–245.

Glen, M., Tommerup, I.C., Bougher, N.L., et O'Brien, P.A. (2001). Interspecific and intraspecific variation of ectomycorrhizal fungi associated with *Eucalyptus* ecosystems as revealed by ribosomal DNA PCR-RFLP. *Mycological Research* 105, 843–858.

Glöckner, F.O., Yilmaz, P., Kottmann, R., Field, D., Knight, R., Cole, J., Amaral-Zettler, L., Gilbert, J., Karsch-Mizrachi, I., Johnston, A., et al. (2010). The “Minimum Information about an ENvironmental Sequence” (MIENS) specification. *Nature Precedings*.

Gogvadze, E., Barbisan, C., Lebrun, M.-H., et Buzdin, A. (2007). Tripartite chimeric pseudogene from the genome of rice blast fungus *Magnaporthe grisea* suggests double template jumps during long interspersed nuclear element (LINE) reverse transcription. *BMC Genomics* 8, 360.

Gottel, N.R., Castro, H.F., Kerley, M., Yang, Z., Pelletier, D.A., Podar, M., Karpinets, T., Uberbacher, E., Tuskan, G.A., Vilgalys, R., et al. (2011). Distinct microbial communities within the endosphere and rhizosphere of *Populus deltoides* roots across contrasting soil Types. *Appl. Environ. Microbiol.* 77, 5934–5944.

Hartmann, M., Howes, C.G., VanInsberghe, D., Yu, H., Bachar, D., Christen, R., Henrik Nilsson, R., Hallam, S.J., et Mohn, W.W. (2012). Significant and persistent impact of timber harvesting on soil microbial communities in Northern coniferous forests. *ISME J* 6, 2320–2320.

Herrera, M.L., Vallor, A.C., Gelfond, J.A., Patterson, T.F., et Wickes, B.L. (2009). Strain-dependent variation in 18S ribosomal DNA copy numbers in *Aspergillus fumigatus*. *J. Clin. Microbiol.* 47, 1325–1332.

Hibbett, D.S., Ohman, A., Glotzer, D., Nuhn, M., Kirk, P., et Nilsson, R.H. (2011). Progress in molecular and morphological taxon discovery in fungi and options for formal classification of environmental sequences. *Fungal Biology Reviews* 25, 38–47.

Hirotsune, S., Yoshida, N., Chen, A., Garrett, L., Sugiyama, F., Takahashi, S., Yagami, K., Wynshaw-Boris, A., et Yoshiki, A. (2003). An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene. *Nature* 423, 91–96.

Hui, N., Jumpponen, A., Niskanen, T., Liimatainen, K., Jones, K.L., Koivula, T., Romantschuk, M., et Strömmer, R. (2011). EcM fungal community structure, but not diversity, altered in a Pb-contaminated shooting range in a boreal coniferous forest site in Southern Finland. *FEMS Microbiology Ecology* 76, 121–132.

- Huse, S.M., Huber, J.A., Morrison, H.G., Sogin, M.L., et Welch, D.M. (2007). Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biology* 8, R143.
- Huse, S.M., Welch, D.M., Morrison, H.G., et Sogin, M.L. (2010). Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environmental Microbiology* 12, 1889–1898.
- Huson, D.H., Auch, A.F., Qi, J., et Schuster, S.C. (2007). MEGAN analysis of metagenomic data. *Genome Res* 17, 377–386.
- Ide, S., Miyazaki, T., Maki, H., et Kobayashi, T. (2010). Abundance of ribosomal RNA gene copies maintains genome integrity. *Science* 327, 693–696.
- Ihrmark, K., Bodeker, I.T.M., Cruz-Martinez, K., Friberg, H., Kubartova, A., Schenck, J., Strid, Y., Stenlid, J., Brandström-Durling, M., Clemmensen, K.E., et al. (2012). New primers to amplify the fungal ITS2 region – evaluation by 454-sequencing of artificial and natural communities. *FEMS Microbiology Ecology* 82, 666–677.
- Illegheems, K., De Vuyst, L., Papalexandratou, Z., et Weckx, S. (2012). Phylogenetic analysis of a spontaneous cocoa bean fermentation metagenome reveals new insights into its bacterial and fungal community diversity. *PLoS ONE* 7, e38040.
- James, S.A., O’Kelly, M.J.T., Carter, D.M., Davey, R.P., Oudenaarden, A. van, et Roberts, I.N. (2009). Repetitive sequence variation and dynamics in the ribosomal DNA array of *Saccharomyces cerevisiae* as revealed by whole-genome resequencing. *Genome Res.* 19, 626–635.
- Jumpponen, A., et Jones, K.L. (2009). Massively parallel 454 sequencing indicates hyperdiverse fungal communities in temperate *Quercus macrocarpa* phyllosphere. *New Phytologist* 184, 438–448.
- Jumpponen, A., et Jones, K.L. (2010). Seasonally dynamic fungal communities in the *Quercus macrocarpa* phyllosphere differ between urban and nonurban environments. *New Phytologist* 186, 496–513.
- Jumpponen, A., Jones, K.L., et Blair, J. (2010a). Vertical distribution of fungal communities in tallgrass prairie soil. *Mycologia* 102, 1027–1041.
- Jumpponen, A., Jones, K.L., David Mattox, J., et Yaege, C. (2010b). Massively parallel 454-sequencing of fungal communities in *Quercus spp.* ectomycorrhizas indicates seasonal dynamics in urban and rural sites. *Molecular Ecology* 19, 41–53.
- Kageyama, K., Senda, M., Asano, T., Suga, H., et Ishiguro, K. (2007). Intra-isolate heterogeneity of the ITS region of rDNA in *Pythium helicoides*. *Mycological Research* 111, 416–423.
- Kårén, O., Högborg, N., Dahlberg, A., Jonsson, L., et Nylund, J.-E. (1997). Inter- and intraspecific variation in the ITS region of rDNA of ectomycorrhizal fungi in Fennoscandia as detected by endonuclease analysis. *New Phytologist* 136, 313–325.
- Kausrud, H., Kumar, S., Brysting, A.K., Nordén, J., et Carlsen, T. (2012). High consistency between replicate 454 pyrosequencing analyses of ectomycorrhizal plant root samples. *Mycorrhiza* 22, 309–315.
- Kim, M.S., Klopfenstein, N.B., Hanna, J.W., et McDonald, G.I. (2006). Characterization of North American *Armillaria* species: genetic relationships determined by ribosomal DNA sequences and AFLP markers. *Forest Pathol.* 36, 145–164.
- Kiss, L. (2012). Limits of nuclear ribosomal DNA internal transcribed spacer (ITS) sequences as species barcodes for fungi. *PNAS* 109, E1811–E1811.
- Ko, K.S., et Jung, H.S. (2002). Three nonorthologous ITS1 types are present in a polypore fungus *Trichaptum abietinum*. *Molecular Phylogenetics and Evolution* 23, 112–122.
- Kobayashi, T. (2011). Regulation of ribosomal RNA gene copy number and its role in modulating genome integrity and evolutionary adaptability in yeast. *Cell. Mol. Life Sci.* 68, 1395–1403.
- Kobayashi, T., et Ganley, A.R.D. (2005). Recombination regulation by transcription-induced cohesin dissociation in rDNA repeats. *Science* 309, 1581–1584.

- Kovacs, G.M., Jankovics, T., et Kiss, L. (2011a). Variation in the nrDNA ITS sequences of some powdery mildew species: do routine molecular identification procedures hide valuable information? *European Journal of Plant Pathology* *131*, 135–141.
- Kovacs, G.M., Balazs, T.K., Calonge, F.D., et Martin, M.P. (2011b). The diversity of *Terfezia* desert truffles: new species and a highly variable species complex with intrasporocarpic nrDNA ITS heterogeneity. *Mycologia* *103*, 841–853.
- Krüger, D., Kapturska, D., Fischer, C., Daniel, R., et Wubet, T. (2012). Diversity measures in environmental sequences are highly dependent on alignment quality — data from ITS and new LSU primers targeting *Basidiomycetes*. *PLoS ONE* *7*, e32139.
- Kubartová, A., Ottosson, E., Dahlberg, A., et Stenlid, J. (2012). Patterns of fungal communities among and within decaying logs, revealed by 454 sequencing. *Molecular Ecology* *21*, 4514–4532.
- Kunin, V., Engelbrektson, A., Ochman, H., et Hugenholtz, P. (2010). Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environmental Microbiology* *12*, 118–123.
- Lafontaine, I., et Dujon, B. (2010). Origin and fate of pseudogenes in *Hemiascomycetes*: a comparative analysis. *BMC Genomics* *11*, 260.
- Lahr, D.J.G., et Katz, L.A. (2009). Reducing the impact of PCR-mediated recombination in molecular evolution and environmental studies using a new-generation high-fidelity DNA polymerase. *BioTechniques* *47*, 857–866.
- Lachance, M. a., Daniel, H. m., Meyer, W., Prasad, G. s., Gautam, S. p., et Boundy-Mills, K. (2003). The D1/D2 domain of the large-subunit rDNA of the yeast species *Clavispora lusitaniae* is unusually polymorphic. *FEMS Yeast Research* *4*, 253–258.
- Lanfranco, L., Delpero, M., et Bonfante, P. (1999). Intrasporal variability of ribosomal sequences in the endomycorrhizal fungus *Gigaspora margarita*. *Molecular Ecology* *8*, 37–45.
- Lekberg, Y., Schnoor, T., Kjølner, R., Gibbons, S.M., Hansen, L.H., Al-Soud, W.A., Sørensen, S.J., et Rosendahl, S. (2012). 454-sequencing reveals stochastic local reassembly and high disturbance tolerance within arbuscular mycorrhizal fungal communities. *Journal of Ecology* *100*, 151–160.
- Lentendu, G., Zinger, L., Manel, S., Coissac, E., Choler, P., Geremia, R.A., et Melodelima, C. (2011). Assessment of soil fungal diversity in different alpine tundra habitats by means of pyrosequencing. *Fungal Diversity* *49*, 113–123.
- Li, W., et Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* *22*, 1658–1659.
- Li, W., Fu, L., Niu, B., Wu, S., et Wooley, J. (2012). Ultrafast clustering algorithms for metagenomic sequence analysis. *Brief Bioinform.*
- Liggenstoffer, A.S., Youssef, N.H., Couger, M.B., et Elshahed, M.S. (2010). Phylogenetic diversity and community structure of anaerobic gut fungi (phylum Neocallimastigomycota) in ruminant and non-ruminant herbivores. *ISME J.* *4*, 1225–1235.
- Lin, X., Feng, Y., Zhang, H., Chen, R., Wang, J., Zhang, J., et Chu, H. (2012). Long-term balanced fertilization decreases arbuscular mycorrhizal fungal diversity in an arable soil in North China Revealed by 454 Pyrosequencing. *Environ. Sci. Technol.* *46*, 5764–5771.
- Lindner, D.L., et Banik, M.T. (2011). Intragenomic variation in the ITS rDNA region obscures phylogenetic relationships and inflates estimates of operational taxonomic units in genus *Laetiporus*. *Mycologia* *103*, 731–740.
- Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., et Law, M. (2012). Comparison of next-generation sequencing systems. *J Biomed Biotechnol* *2012*.
- Lumini, E., Orgiazzi, A., Borriello, R., Bonfante, P., et Bianciotto, V. (2010). Disclosing arbuscular

mycorrhizal fungal biodiversity in soil through a land-use gradient using a pyrosequencing approach. *Environmental Microbiology* 12, 2165–2179.

Maicas, S., Adam, A.C., et Polaina, J. (2000). The ribosomal DNA of the Zygomycete *Mucor miehei*. *Curr Genet* 37, 412–419.

Maleszka, R., et Clark-Walker, G.D. (1990). Magnification of the rDNA cluster in *Kluyveromyces lactis*. *Molec. Gen. Genet.* 223, 342–344.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.-J., Chen, Z., et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature*.

Mayol, M., et Rosselló, J.A. (2001). Why nuclear ribosomal DNA spacers (ITS) tell different stories in *Quercus*. *Molecular Phylogenetics and Evolution* 19, 167–176.

McGuire, K.L., Fierer, N., Bateman, C., Treseder, K.K., et Turner, B.L. (2012). Fungal community composition in neotropical rain forests: the influence of tree diversity and precipitation. *Microb. Ecol.* 63, 804–812.

Mello, A., Napoli, C., Murat, C., Morin, E., Marceddu, G., et Bonfante, P. (2011). ITS-1 versus ITS-2 pyrosequencing: a comparison of fungal populations in truffle grounds. *Mycologia* 103, 1184–1193.

Menkis, A., Burokienė, D., Gaitnieks, T., Uotila, A., Johannesson, H., Rosling, A., Finlay, R.D., Stenlid, J., et Vasaitis, R. (2012). Occurrence and impact of the root-rot biocontrol agent *Phlebiopsis gigantea* on soil fungal communities in *Picea abies* forests of northern Europe. *FEMS Microbiology Ecology* 81, 438–445.

Meyer, M., Stenzel, U., et Hofreiter, M. (2008). Parallel tagged sequencing on the 454 platform. *Nature Protocols* 3, 267–278.

Monard, C., Gantner, S., et Stenlid, J. (2013). Utilizing ITS1 and ITS2 to study environmental fungal diversity using pyrosequencing. *FEMS Microbiology Ecology* 84, 165–175.

Monchy, S., Sancier, G., Jobard, M., Rasconi, S., Gerphagnon, M., Chabé, M., Cian, A., Meloni, D., Niquil, N., Christaki, U., et al. (2011). Exploring and quantifying fungal diversity in freshwater lake ecosystems using rDNA cloning/sequencing and SSU tag pyrosequencing. *Environmental Microbiology* 13, 1433–1453.

Moora, M., Berger, S., Davison, J., Öpik, M., Bommarco, R., Bruehlheide, H., Kühn, I., Kunin, W.E., Metsis, M., Rortais, A., et al. (2011). Alien plants associate with widespread generalist arbuscular mycorrhizal fungal taxa: evidence from a continental-scale study using massively parallel 454 sequencing. *Journal of Biogeography* 38, 1305–1317.

Mullineux, T., et Hausner, G. (2009). Evolution of rDNA ITS1 and ITS2 sequences and RNA secondary structures within members of the fungal genera *Grosmannia* and *Leptographium*. *Fungal Genetics and Biology* 46, 855–867.

Nilsson, H.R., Tedersoo, L., Lindahl, B.D., Kjoller, R., Carlsen, T., Quince, C., Abarenkov, K., Pennanen, T., Stenlid, J., Bruns, T., et al. (2011). Towards standardization of the description and publication of next-generation sequencing datasets of fungal communities. *New Phytologist* 191, 314–318.

Nilsson, R.H., Kristiansson, E., Ryberg, M., Hallenberg, N., et Larsson, K.-H. (2008). Intraspecific ITS variability in the kingdom fungi as expressed in the International Sequence Databases and its implications for molecular species identification. *Evol Bioinform Online* 4, 193–201.

Niu, B., Fu, L., Sun, S., et Li, W. (2010). Artificial and natural duplicates in pyrosequencing reads of metagenomic data. *BMC Bioinformatics* 11, 187.

O'Brien, H.E., Parrent, J.L., Jackson, J.A., Moncalvo, J.-M., et Vilgalys, R. (2005). Fungal community analysis by large-scale sequencing of environmental samples. *Appl. Environ. Microbiol.* 71, 5544–5550.

O'Donnell, K. (1992). Ribosomal DNA internal transcribed spacers are highly divergent in the phytopathogenic ascomycete *Fusarium sambucinum* (*Gibberella pulicaris*). *Curr. Genet.* 22, 213–220.

- O'Donnell, K., et Cigelnik, E. (1997). Two divergent intragenomic rDNA ITS2 types within a monophyletic lineage of the fungus *Fusarium* are nonorthologous. *Molecular Phylogenetics and Evolution* 7, 103–116.
- Opik, M., Zobel, M., Cantero, J.J., Davison, J., Facelli, J.M., Hiiesalu, I., Jairus, T., Kalwij, J.M., Koorem, K., Leal, M.E., et al. (2013). Global sampling of plant roots expands the described molecular diversity of arbuscular mycorrhizal fungi. *Mycorrhiza*.
- Öpik, M., Metsis, M., Daniell, T.J., Zobel, M., et Moora, M. (2009). Large-scale parallel 454 sequencing reveals host ecological group specificity of arbuscular mycorrhizal fungi in a boreonemoral forest. *New Phytologist* 184, 424–437.
- Orgiazzi, A., Bianciotto, V., Bonfante, P., Daghino, S., Ghignone, S., Lazzari, A., Lumini, E., Mello, A., Napoli, C., Perotto, S., et al. (2013). 454 Pyrosequencing analysis of fungal assemblages from geographically distant, disparate soils reveals spatial patterning and a core mycobiome. *Diversity* 5, 73–98.
- Ovaskainen, O., Nokso-Koivisto, J., Hottola, J., Rajala, T., Pennanen, T., Ali-Kovero, H., Miettinen, O., Oinonen, P., Auvinen, P., Paulin, L., et al. (2010). Identifying wood-inhabiting fungi with 454 sequencing – what is the probability that BLAST gives the correct species? *Fungal Ecology* 3, 274–283.
- Patin, N.V., Kunin, V., Lidström, U., et Ashby, M.N. (2012). Effects of OTU clustering and PCR artifacts on microbial diversity estimates. *Microb. Ecol.*
- Pawlowska, T.E., et Taylor, J.W. (2004). Organization of genetic variation in individuals of arbuscular mycorrhizal fungi. *Nature* 427, 733–737.
- Pawlowska, T.E., et Taylor, J.W. (2005). Arbuscular mycorrhizal fungi: Hyphal fusion and multigenomic structure (reply). *Nature* 433, E4–E4.
- Prober, J.M., Trainor, G.L., Dam, R.J., Hobbs, F.W., Robertson, C.W., Zagursky, R.J., Cocuzza, A.J., Jensen, M.A., et Baumeister, K. (1987). A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. *Science* 238, 336–341.
- Quail, M.A., Smith, M., Coupland, P., Otto, T.D., Harris, S.R., Connor, T.R., Bertoni, A., Swerdlow, H.P., et Gu, Y. (2012). A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* 13, 341.
- Quince, C., Lanzén, A., Curtis, T.P., Davenport, R.J., Hall, N., Head, I.M., Read, L.F., et Sloan, W.T. (2009). Accurate determination of microbial diversity from 454 pyrosequencing data. *Nature Methods* 6, 639–641.
- Quince, C., Lanzen, A., Davenport, R.J., et Turnbaugh, P.J. (2011). Removing noise from pyrosequenced amplicons. *BMC Bioinformatics* 12, 38.
- Rajashekar, B., Samson, P., Johansson, T., et Tunlid, A. (2007). Evolution of nucleotide sequences and expression patterns of hydrophobin genes in the ectomycorrhizal fungus *Paxillus involutus*. *New Phytologist* 174, 399–411.
- Reeder, J., et Knight, R. (2010). Rapidly denoising pyrosequencing amplicon reads by exploiting rank-abundance distributions. *Nature Methods* 7, 668–669.
- Richterich, P. (1998). Estimation of errors in “raw” DNA sequences: A validation study. *Genome Res* 8, 251–259.
- Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlén, M., et Nyrén, P. (1996). Real-time DNA sequencing using detection of pyrophosphate release. *Analytical Biochemistry* 242, 84–89.
- Ronaghi, M., Uhlén, M., et Nyrén, P. (1998). A sequencing method based on real-time pyrophosphate. *Science* 281, 363–365.
- Rooney, A.P., et Ward, T.J. (2005). Evolution of a large ribosomal RNA multigene family in filamentous fungi: Birth and death of a concerted evolution paradigm. *Proc Natl Acad Sci U S A* 102, 5084–5089.
- Rousk, J., Bååth, E., Brookes, P.C., Lauber, C.L., Lozupone, C., Caporaso, J.G., Knight, R., et Fierer, N. (2010). Soil bacterial and fungal communities across a pH gradient in an arable soil. *ISME J* 4, 1340–1351.

- Ryberg, M., Kristiansson, E., Sjökvist, E., et Nilsson, R.H. (2009). An outlook on the fungal internal transcribed spacer sequences in GenBank and the introduction of a web-based tool for the exploration of fungal diversity. *New Phytol.* *181*, 471–477.
- Rydholm, C., Dyer, P.S., et Lutzoni, F. (2007). DNA sequence characterization and molecular evolution of MAT1 and MAT2 mating-type loci of the self-compatible Ascomycete mold *Neosartorya fischeri*. *Eukaryotic Cell* *6*, 868–874.
- Sanger, F., Nicklen, S., et Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* *74*, 5463–5467.
- Seifert, K.A. (2009). Progress towards DNA barcoding of fungi. *Molecular Ecology Resources* *9*, 83–89.
- Selosse, M.-A., et Schardl, C.L. (2007). Fungal endophytes of grasses: hybrids rescued by vertical transmission? An evolutionary perspective. *New Phytologist* *173*, 452–458.
- Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., et al. (2009). Introducing MOTHUR: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* *75*, 7537–7541.
- Schoch, C.L., Seifert, K.A., Huhndorf, S., Robert, V., Spouge, J.L., Levesque, C.A., Chen, W., Bolchacova, E., Voigt, K., Crous, P.W., et al. (2012). Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for fungi. *Proc Natl Acad Sci U S A* *109*, 6241–6246.
- Simon, U.K., et Weiss, M. (2008). Intragenomic variation of fungal ribosomal genes is higher than previously thought. *Molecular Biology and Evolution* *25*, 2251–2254.
- Smith, M.E., Douhan, G.W., et Rizzo, D.M. (2007). Intra-specific and intra-sporocarp ITS variation of ectomycorrhizal fungi as assessed by rDNA sequencing of sporocarps and pooled ectomycorrhizal roots from a *Quercus* woodland. *Mycorrhiza* *18*, 15–22.
- Sogin, M.L., Morrison, H.G., Huber, J.A., Welch, D.M., Huse, S.M., Neal, P.R., Arrieta, J.M., et Herndl, G.J. (2006). Microbial diversity in the deep sea and the underexplored “rare biosphere.” *PNAS* *103*, 12115–12120.
- Stockinger, H., Walker, C., et Schüßler, A. (2009). “*Glomus intraradices* DAOM197198”, a model fungus in arbuscular mycorrhiza research, is not *Glomus intraradices*. *New Phytologist* *183*, 1176–1187.
- Stockinger, H., Krüger, M., et Schüßler, A. (2010). DNA barcoding of arbuscular mycorrhizal fungi. *New Phytologist* *187*, 461–474.
- Sun, Y., Cai, Y., Huse, S.M., Knight, R., Farmerie, W.G., Wang, X., et Mai, V. (2012). A large-scale benchmark study of existing algorithms for taxonomy-independent microbial community analysis. *Brief Bioinform* *13*, 107–121.
- Štursová, M., Žifčáková, L., Leigh, M.B., Burgess, R., et Baldrian, P. (2012). Cellulose utilization in forest litter and soil: identification of bacterial and fungal decomposers. *FEMS Microbiology Ecology* *80*, 735–746.
- Taberlet, P., Coissac, E., Hajibabaei, M., et Rieseberg, L.H. (2012). Environmental DNA. *Molecular Ecology* *21*, 1789–1793.
- Tedersoo, L., Nilsson, R.H., Abarenkov, K., Jairus, T., Sadam, A., Saar, I., Bahram, M., Bechem, E., Chuyong, G., et Kõljalg, U. (2010). 454 Pyrosequencing and Sanger sequencing of tropical mycorrhizal fungi provide similar results but reveal substantial methodological biases. *New Phytologist* *188*, 291–301.
- Tedersoo, L., Abarenkov, K., Nilsson, R.H., Schüssler, A., Grelet, G.-A., Kohout, P., Oja, J., Bonito, G.M., Veldre, V., Jairus, T., et al. (2011). Tidying up International Nucleotide Sequence Databases: ecological, geographical and sequence quality annotation of ITS sequences of mycorrhizal fungi. *PLoS ONE* *6*, e24940.
- Toju, H., Tanabe, A.S., Yamamoto, S., et Sato, H. (2012). High-coverage ITS primers for the DNA-based identification of *Ascomycetes* and *Basidiomycetes* in environmental samples. *PLoS ONE* *7*, e40863.
- Tsuchiya, D., et Taga, M. (2001). Application of fibre-FISH (fluorescence in situ hybridization) to filamentous

- fungi: visualization of the rRNA gene cluster of the ascomycete *Cochliobolus heterostrophus*. *Microbiology* 147, 1183–1187.
- Větrovský, T., et Baldrian, P. (2013). Analysis of soil fungal communities by amplicon pyrosequencing: current approaches to data analysis and the introduction of the pipeline SEED. *Biology and Fertility of Soils*.
- Voříšková, J., et Baldrian, P. (2013). Fungal community on decomposing leaf litter undergoes rapid successional changes. *ISME J* 7, 477–486.
- Wallander, H., Johansson, U., Sterkenburg, E., Brandström Durling, M., et Lindahl, B.D. (2010). Production of ectomycorrhizal mycelium peaks during canopy closure in Norway spruce forests. *New Phytologist* 187, 1124–1134.
- Wery, J., Gutker, D., Renniers, A.C.H.M., Verdoes, J.C., et van Ooyen, A.J.J. (1997). High copy number integration into the ribosomal DNA of the yeast *Phaffia rhodozyma*. *Gene* 184, 89–97.
- Wubet, T., Christ, S., Schöning, I., Boch, S., Gawlich, M., Schnabel, B., Fischer, M., et Buscot, F. (2012). Differences in soil fungal communities between European beech (*Fagus sylvatica* L.) dominated forests are related to soil and understory vegetation. *PLoS ONE* 7, e47500.
- Xiao, L.-Q., Möller, M., et Zhu, H. (2010). High nrDNA ITS polymorphism in the ancient extant seed plant *Cycas*: Incomplete concerted evolution and the origin of pseudogenes. *Molecular Phylogenetics and Evolution* 55, 168–177.
- Xu, L., Ravnskov, S., Larsen, J., et Nicolaisen, M. (2011). Influence of DNA extraction and PCR amplification on studies of soil fungal communities based on amplicon sequencing. *Can. J. Microbiol.* 57, 1062–1066.
- Xu, L., Ravnskov, S., Larsen, J., Nilsson, R.H., et Nicolaisen, M. (2012). Soil fungal community structure along a soil health gradient in pea fields examined using deep amplicon sequencing. *Soil Biol. Biochem.* 46, 26–32.
- Yu, L., Nicolaisen, M., Larsen, J., et Ravnskov, S. (2012a). Molecular characterization of root-associated fungal communities in relation to health status of *Pisum sativum* using barcoded pyrosequencing. *Plant Soil* 357, 395–405.
- Yu, L., Nicolaisen, M., Larsen, J., et Ravnskov, S. (2012b). Succession of root-associated fungi in *Pisum sativum* during a plant growth cycle as examined by 454 pyrosequencing. *Plant Soil* 358, 225–233.
- Yu, Y., Breitbart, M., McNairnie, P., et Rohwer, F. (2006). FastGroupII: A web-based bioinformatics platform for analyses of large 16S rDNA libraries. *BMC Bioinformatics* 7, 57.
- Zimmerman, N.B., et Vitousek, P.M. (2012). Fungal endophyte communities reflect environmental structuring across a Hawaiian landscape. *Proc Natl Acad Sci U S A* 109, 13022–13027.

Elektronické zdroje:

- 454 Life Sciences [online]. [cit. 2013-05-10]. Dostupný z: <http://www.454.com>.
- 454 Sequencing - GS Junior- Instrument and Workflow [online]. [cit. 2013-05-10]. Dostupný z: <http://www.gsjunior.com/instrument-workflow.php>.

8. Přílohy

Tabulka 1 (str. 1/3): Přehled mykologických prací využívajících 454 sekvenování. V této části jsou shrnuty práce bez bližší specifikace odhadování diverzity.

Typ vzorku	Množství	Marker	Primery	Počet sekvencí	Počet OTU	Odhad diverzity	Délka	Filtrování, zkracování	Hladina klastrování	Klastrovací algoritmus	<i>Denoising</i>	Odstranění chimér	Odstranění singletonů	Taxonomické zařazení	Reference
půda	1 000 – 5 000 mg	ITS; 28S	ITS1F/TW13	918	24; 38	✓			95; 97 %	CAP3				UniFrac	Geml et al., (2009)
sediment	100 mg	ITS; 28S	ITS1F/ITS4; LRORF/LR5F	ITS: 97 557 ; LSU: 187 668	ITS: 4 473	✓	> 300	RDP; CLC Gen. W.	97 %	CD-HIT (ITS); MAFFT (LSU)				MEGAN	Amend et al., (2010a)
sediment	100 mg	ITS	ITS1F/ITS4	ITS1F: 3 596 ; ITS4: 4 379	ITS1F: 19 ; ITS4: 6	✓	> 300	CLC Gen. W.	97 %	CD-HIT				MEGAN	Amend et al., (2010b)
kořeny	40 – 50 mg	ITS	ITS1F/ITS4	576	137				97 %	Sequencher		ručně		BLASTN	Bjorbækmo et al., (2010)
ústní výplach	15 ml	ITS	ITS1F/ITS2	34 049	85		> 100	vlastní skript	98 %	CD-HIT; KALIGN				BLASTN	Ghannoum et al., (2010)
půda	500 mg	ITS	ITS5/ITS4	14 578	1 151 (597 NS)	✓	> (200)		95 %	CAP3			✓	BLASTN	Jumpponen et al., (2010a)
exkrementy	500 mg	ITS	MN100/MNGM 2	267 287	1 037	✓	> 130		95 %	MOTHUR				BLASTN	Liggenstoffer et al., (2010)
piliny	150 – 250 mg MV	ITS	ITS1F/ITS2	62 214			> 150							BLASTN	Ovaskainen et al., (2010)
půda	500 mg	18S	5,8S/ITS1F	4 700 (85 PNV)	21–36 PNV	✓		QIIME	97 %	CD-HIT				BLAST	Rousk et al., (2010)
kořeny	5 mg MV	18S; ITS	ITS1F/ITS2; ITS5/ITS2	44 411	312 (225 NS)	✓	> 140	Newbler	97 %	TGICL		ručně BLASTN	✓	BLASTN	Tedersoo et al., (2010)
kořeny (G)	50 mg	18S	NS31/AM1; WANDA/AM1	108 245	70	✓	> 100		97 %	NW; BLASTCLUST				BLASTN	Dumbrell et al., (2011)
půda, kořeny	2,0 ml půdy; 50 mg kořinek	28S	LROR/LR3	322 100 (316 360 NS)	72	✓	150–300	MOTHUR	95 %	MOTHUR; Cluster			✓	BLASTN	Gottel et al., (2011)
kořeny, plodnice	1–3 kořinky; x mg SV lupenů	ITS	ITS1F/ITS2	8 194	161 (79 NS)	✓	> (200)	PREGAP; GAP	95 %	CAP3			✓	BLASTN	Hui et al., (2011)

Tabulka 1 (str. 2/3): pokračování

Typ vzorku	Množství	Marker	Primery	Počet sekvencí	Počet OTU	Odhad diverzity	Délka	Filtrování, zkracování	Hladina klastrování	Klastrovací algoritmus	<i>Denoising</i>	Odstranění chimér	Odstranění <i>singletonů</i>	Taxonomické zařazení	Reference
filtrační sediment	150; 300 ml sladké vody	18S	18S-82F/Ek-516r; nu-SSU-0817-5'/nu-SSU-1196-3'	53 856; 49 579		✓								BLASTN; MEGAN	Monchy et al., (2011)
koráli	150–200 mg	28S	LROR/LR5	26 022	5410 (840 NS)	✓	> 250	MOTHUR	97 %	MOTHUR		MOTHUR	✓	BLASTN; MEGAN	Amend et al., (2012)
kořeny	cca 15 cm	18S	nu-SSU-0817-5'/nu-ssu-1536-3'	3 921	44 (G); 186 (nonG) (= 230)	✓	> 195		97 %	CAP3			✓	BLASTN; ClustalX	Becklin et al., (2012)
sediment	0,2 ml	18S	SSUF04/SSUR22; NF1/18Sr2b	969 343	986 (36 houby)	✓	> 200	OCTUPUS; QIIME	95; 99 %	MegaBLAST; UCLUST	AmpliconNoise; Perseus	OCTUPUS; QIIME		BLASTN	Bik et al., (2012)
kořeny	5 – 315 mg	ITS	ITS1F/ITS4; ITS5/ITS2	101 870	1663 (470 NS)	✓	> 150	CLOTU	98.5 %	BLASTCLUST			✓	CLOTU	Blaalid et al., (2012)
listy	2 cm ² / list	ITS	ITS1F/ITS2	96 130	1 604	✓	> 100	RDP	97 %	UCLUST				BLASTN	Cordier et al., (2012)
půda	500 mg	ITS	ITS1F/ITS2; ITS3/ITS4	ITS1: 4 952; ITS2: 2 707	ITS1: 791 (386 NS); ITS2: 576 (266 NS)	✓	> 200		97 %	CD-HIT				BLASTN; MEGAN	Daghino et al., (2012)
listy, kořeny, půda	100 mg	ITS	ITS1F/ITS2	686 053	75 – 800 NS	✓	> 150		97 %	USEARCH			✓	BLASTN	Danielsen et al., (2012)
mech	rostlinka	ITS	ITS1F/ITS4; ITS3/ITS4	327 273	9 367 (2 676 NS)	✓	> 200	QIIME	97 %	UCLUST	QIIME		✓	BLASTN	Davey et al., (2012)
půda (G)	250 mg SV	18S	NS31/AML2	13 320	37 (G)	✓	> 170	Python; Java skripty	97 %	BLASTN		UCHIME	✓	BLASTN	Davison et al., (2012)
půda	500 mg	ITS	ITS3/ITS4	2 224 307	19 353	✓		MOTHUR	97 %	CRUNCHCLUST (NW)	CRUNCH-CLUST			MOTHUR	Hartmann et al., (2012)
půda, dřevo, kořeny, seno		ITS	ITS1F/ITS4; fITS7/ITS4; gITS7/ITS4; fITS9/ITS4	612 279	487–492	✓	146–274	SCATA	98,5 %	SCATA			✓	PlutoF	Ihrmark et al., (2012)
kakaový kvas	20 000 mg	ITS		755	17	✓		ShortRead; CD-HIT						BLASTN; MEGAN	Illegghems et al., (2012)

Tabulka 1 (str. 3/3): pokračování

Typ vzorku	Množství	Marker	Primery	Počet sekvencí	Počet OTU	Odhad diverzity	Délka	Filtrování, zkracování	Hladina klastrování	Klastrovací algoritmus	<i>Denoising</i>	Odstranění chimér	Odstranění <i>singletonů</i>	Taxonomické zařazení	Reference
kořeny	90 mg	ITS	ITS1F/ITS4; ITS5/ITS2	10 430	57 (35 NS)		> 150	CLOTU	97 %	BLASTCLUST				BLASTN	Kauserud et al., (2012)
přiliny		ITS	ITS1F/ITS4	231 337	1 914		> 200	SCATA	98,5 %	SCATA			✓	SCATA; BLASTN	Kubartová et al., (2012)
sediment prach	500 µl směsi	ITS	ITS1F/ITS4	143 557	456		> 200	MOTHUR		CRUNCHCLUST				Greengenes	La Duc et al., (2012)
kořeny (G)	25 mg SV	28S	0061/LR5; glo454/NDL22	213 323	32	✓	> 150	RDP	97 %	MAFFT			✓	BLASTN	Lekberg et al., (2012)
půda (G)	500 mg MV	18S	AMV4.5NF/AM DGR	124 998	493 (231 houby; 37 PNV G)	✓	> 200	QIIME	97 %					BLASTN	Lin et al., (2012)
půda	250 mg	18S	SSU817f/SSU1 196r	57 560	9 733			QIIME	97 %					BLASTN	McGuire et al., (2012)
půda	500 mg	ITS	ITS1/ITS4	7 075	413		> 300	MOTHUR	97 %	CD-HIT	Denoiser	UCHIME		PlutoF	Štursová et al., (2012)
kořeny	420 mg PNV kořinky; 1600 mg výhony	ITS	ITS1F/ITS2	60 071	661 (142 NS)	✓	> 150	CLOTU	97 %	BLASTCLUST			✓	BLASTN	Yu et al., (2012a)
listy	1 cm ² / list	ITS	ITS1F/ITS4	ITS1: 52 825; ITS2: 49 010	ITS1: 169; ITS2: 201	✓	300–450	Sffinfo	97 %		Amplicon-Noise	Amplicon-Noise		BLASTN; MEGAN	Bazzicalupo et al., (2013)
mech	výhonek	ITS	ITS1F/ITS5; ITS3/ITS4	326 018	3 329 (1 695 NS)	✓	> 250	QIIME	97 %	UCLUST	Denoiser	Perseus	✓	BLASTN	Davey et al., (2013)
vaginální stěr		ITS	ITS1F/ITS2	276 336	196	✓	> 170		97 %	CD-HIT				BLASTN	Drell et al., (2013)
půda	500 mg	ITS	ITS1F/ITS4	ITS1: 190 102/67 246; ITS2: 194 711/39 124	ITS1: 4 542/1 563; ITS2: 4 671/1 239	✓	> 100; > 250	SCATA	98,5 %	SLC			✓	BLASTN	Monard et al., (2013)
kořeny (G)	30 – 100 mg SV	18S	NS31/AML2	22 391	173	✓	> 170		97 %	Neighbor-joining		UCHIME	✓	BLASTN	Opik et al., (2013)
půda	500 mg	ITS	ITS1F/ITS2; ITS3/ITS4	ITS1: 16 758; ITS2: 11 234	ITS1: 1 660; ITS2: 1 393	✓	> 200	QIIME	97 %	USEARCH	QIIME	QIIME		BLASTN	Orgiazzi et al., (2013)
listy, opad	300 mg	ITS	ITS1/ITS4	23 760	1 874 (681 NS)	✓	200–380		97 %	CD-HIT	Denoiser	UCHIME		PlutoF	Voříšková et Baldrian, (2013)

Tabulka 2 (str. 1/2): Přehled mykologických prací využívajících 454 sekvenování. V této části jsou shrnuty práce, které odhadovaly diverzitu pomocí ukazatele Chao1, příp. Chao2 a ACE. Dále jsou přidány sloupce ukazující počty odhadnutých OTU a vztaznou soustavu tohoto odhadu.

Typ vzorku	Množství	Marker	Primery	Počet sekvencí	Počet OTU	Odhadnutá diverzita	Vztazeno k	Estimátor	Délka	Filtrování, zkracování	Hladina klastrování	Klastrovací algoritmus	<i>Denoising</i>	Odstranění chimér	Odstranění singletonů	Taxonomické zařazení	Reference
půda	500 mg	ITS	ITS1F/ITS2	166 350	823 PNV	2 240	4 g	Chao1	> 100	EMBOSS	97 %	BLAST-CLUST				MEGAN	Buée et al., (2009)
listy	3,14 cm ² / list	ITS	ITS1F/ITS2	18 020; 1 005 PNV	689 (360 NS); PNV: 58	203	PNV	Chao1	> 200		95 %	CAP3			✓	BLASTN	Jumpponen et Jones, (2009)
kořeny (G)	20 cm	18S	NS31/AM1	158 358	47 (G)	10 – 38	PNV	Chao1 = Chao2	> 160		97 %					BLASTN	Öpik et al., (2009)
listy	2,4 cm ² / list	ITS	ITS1F/ITS4	84 956	1 232 (744 NS)	345; 607; 317	PNV	Chao1; Chao2; ACE	> 200		95 %	CAP3			✓	BLASTN	Jumpponen et Jones, (2010)
kořeny	cca 80–400 ks PNV	ITS	ITS1F/ITS2	29 910; 618 PNV	19 PNV	119	PNV	Chao1	> 200		95 %	CAP3				BLASTN	Jumpponen et al., (2010b)
půda (G)	2500 mg	18S	AMV4.5NF/AM DGR; NS31/Ammix	3 189; 1 003	117 (37 NS); 28 (9 NS)	156 a 161; 37 a 35	dataset	Chao1 a ACE	> 230; > 250		97 %	DOTUR					Lumini et al., (2010)
mycelium	50 – 100 mg směsi	ITS	ITS1F/ITS4	> 18 000	248	65	dataset	Chao2 (95%)	> 200	SCATA	98.5 %	SCATA				BLASTN	Wallander et al., (2010)
půda	250 mg SV	ITS	ITS5/ITS2	205 131	2 847	12 000; 5 100	dataset	Chao1; ACE	196–310	nespec.	98 %	Borneo, MCL				ecoTag (OBITools)	Lentendu et al., (2011)
půda	500 mg	ITS	ITS1F/ITS2; ITS3/ITS4	13 646	552 PNV	801; 846	PNV	Chao1; ACE	> 60	Sequencher	97 %	RapidOTU				BLASTN	Mello et al., (2011)
kořeny (G)	20 cm	18S	NS31/AM1	109 884	73 (G) (21 PNV)	23	PNV	Chao1	> 160							BLASTN	Moora et al., (2011)
půda	3 000 mg	ITS	ITS1/58A2R	6 490	363 (180 NS)	83	dataset	Chao1	> 200	CLOTU	97 %	BLAST-CLUST			✓	BLASTN	Xu et al., (2011)
půda	500 mg	ITS	ITS1F/ITS4	10 318 (ITS1F: 4 767 ; ITS4: 5 551)	ITS1F: 119 ; ITS4: 115	ITS1F: 155 ; ITS4: 138	dataset	Chao1	> 200		98 % Asco. / 97 % Bas.	CD-HIT		Chimera-test tool	✓	BLASTN	Arfi et al., (2012a)
listy, kmen, kůra	500 mg	ITS; 18S	ITS1F/ITS4; nu-SSU-0817-5/nu-ssu-1536-3'	209 544 (ITS1F: 62 485 ; ITS4: 82 097 ; nu-ssu-0817: 33 746 ; nu-ssu-1536: 31 216)	ITS1F: 2 877 ; ITS4: 3 506 ; nu-ssu-0817: 1 683 ; nu-ssu-1536: 1 455 (vše NS)	ITS1F: 835 ; ITS4: 974 ; nu-ssu-0817: 707 ; nu-ssu-1536: 505	PNV	Chao1	> 200	PANGEA	98 %	CD-HIT		Chimera-test tool	✓	MEGAN	Arfi et al., (2012b)

Tabulka 2 (str. 2/2): pokračování

Typ vzorku	Množství	Marker	Primery	Počet sekvencí	Počet OTU	Odhadnutá diverzita	Vztaženo k	Estimátor	Délka	Filtrování, zkracování	Hladina klastrování	Klastrovací algoritmus	<i>Denoising</i>	Odstranění chimér	Odstranění <i>singletonů</i>	Taxonomické zařazení	Reference
půda	3 000 mg	ITS	ITS1/ITS4	1 000 PNV	181	315	PNV	Chao1	200–380	MOTHUR	97 %	CD-HIT	Denoiser	UCHIME		PlutoF	Baldrian et al., (2012)
kořeny půda humus	15 mg SV kořínků, 1 000 mg SV půdy, 100 mg SV humusu	ITS; 28S	ITS9/ITS4	8 945	456 (262 NS)	689–784	dataset	Chao1	> 200	SEQMAN PRO	97 %					BLASTN	Menkis et al., (2012)
kořeny	500 mg	ITS	ITS1F/ITS4	67 925	569 (121 NS)	72	PNV	Chao1 (95%)	> 150	CLOTU	97 %	BLAST-CLUST			✓	BLASTN	Yu et al., (2012b)
půda	1 000 mg	ITS	ITS1F/ITS4	15 408	1 655	3 310	dataset	Chao1 = ACE	200–450	QIIME; MOTHUR	97 %	CD-HIT; MAFFT		UCHIME		CANGS	Wubet et al., (2012)
půda	3 000 mg	ITS	ITS1/58A2R	55 460	434 (145–200 PNV)	172–212	PNV	Chao1 = ACE	> 50		97 %	CAP3			✓	BLASTN	Xu et al., (2012)
listy	50 mg SV	ITS	ITS1F/ITS2	665 155	4 253 (2 552 NS)	6 812	dataset	Chao1	> 100	HMM; MUSCLE; Geneious	95 %	NW; ESPRIT	SLP		✓	BLASTN; MEGAN	Zimmerman et Vitousek, (2012)

Vysvětlivky:

HMM – *hidden Markov model*; SLC – *single-linkage clustering*; SLP – *single-linkage preclustering*; (G) – zkoumány houby ze skupiny Glomeromycota; NV – na vzorek; PNV – průměr na vzorek; SV – suchá váha;

MV – mokrá váha; NW – *Needleman-Wunsch*; RDP – *the Ribosomal Database Project*; CLCgenW – *CLC Genomics Workbench 2*; F – *forward*; R – *reverse*

Byly vybrány pouze práce využívající část rDNA jako marker. Slovo půda zahrnuje také opad. Čísla jsou zaokrouhlená, udaná bez směrodatných odchylek.

Pozn. k práci Mello 2011: Byly zkoumány vypálená a nevypálená oblast. Uvnitř byly počty OTU pro ITS1: 514; ITS2: 591 a vně ITS1: 599; ITS2: 503 a celkové hodnoty zahrnující průnik OTU z obou oblastí byly ITS1: 900 a ITS2: 885 OTU.