

Využití molekulárních metod při studiu taxonomie a ekologie hub

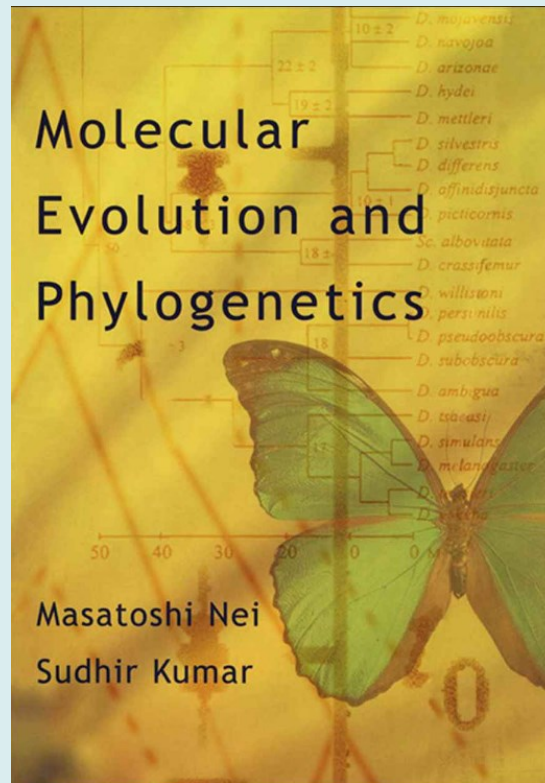
Miroslav Kolařík

mkolarik@biomed.cas.cz

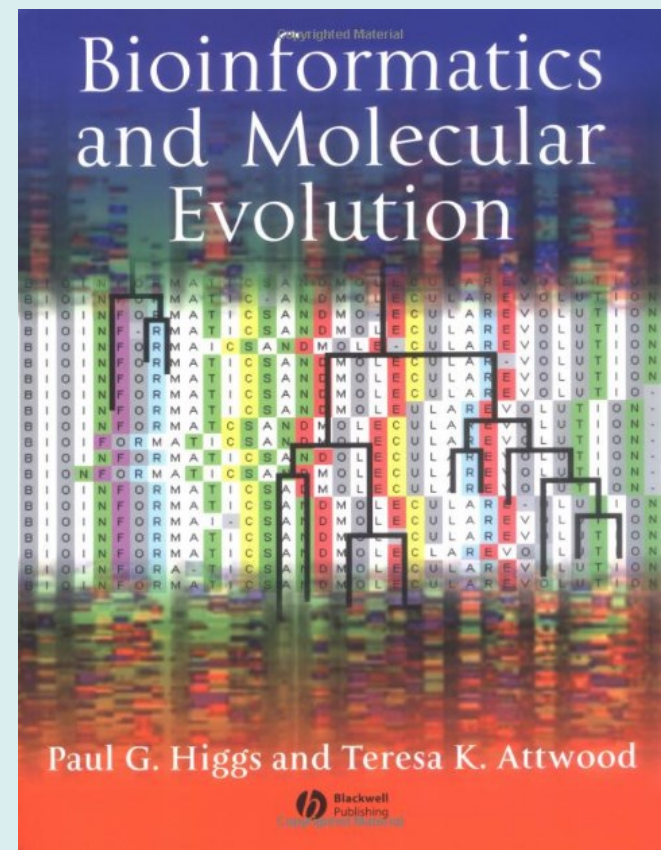
Molekulární genetika na PřF UK

- **Helena Štorchová** - Molekulární přístupy v botanice
- **Tomáš Fér** - Využití molekulárních markerů v systematice a populační biologii rostlin
- **Fatima Cvrčková** - Bioinformatika
- **Vladimír Hampl** - Molekulární taxonomie
- **Jan Zima a kol.** - Genetické metody v zoologii
- **Karol Marhold a kol.** - Fenetika a kladistika
- **Vladimír Vondrejs** - Genové inženýrství
- **Pavel Munclinger, Karel Janko** - Molekulární ekologie
- **Michail Rovatsos** - Practical Course of Evolutionary Genetics and Genomics

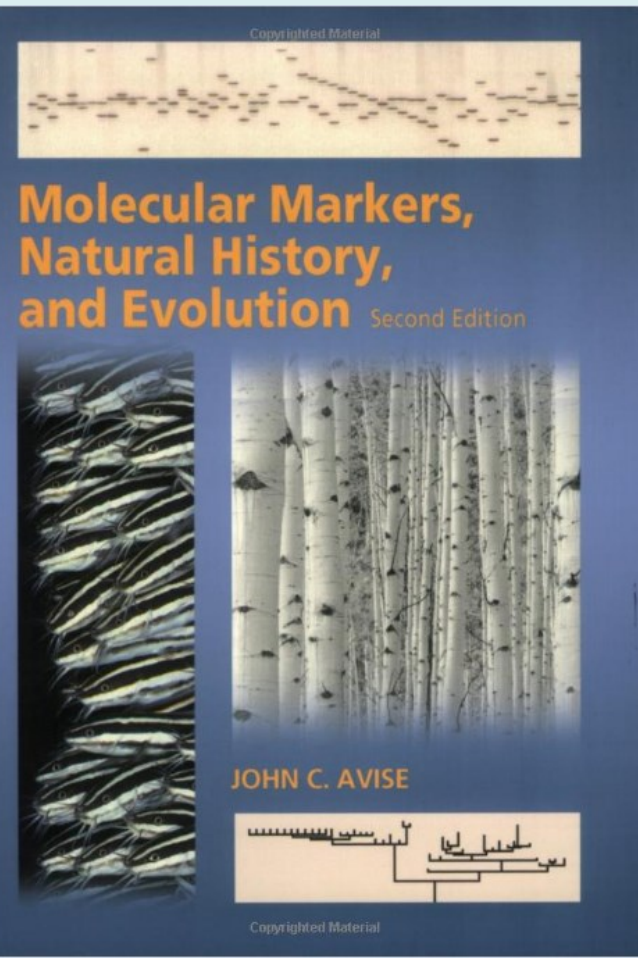
2000

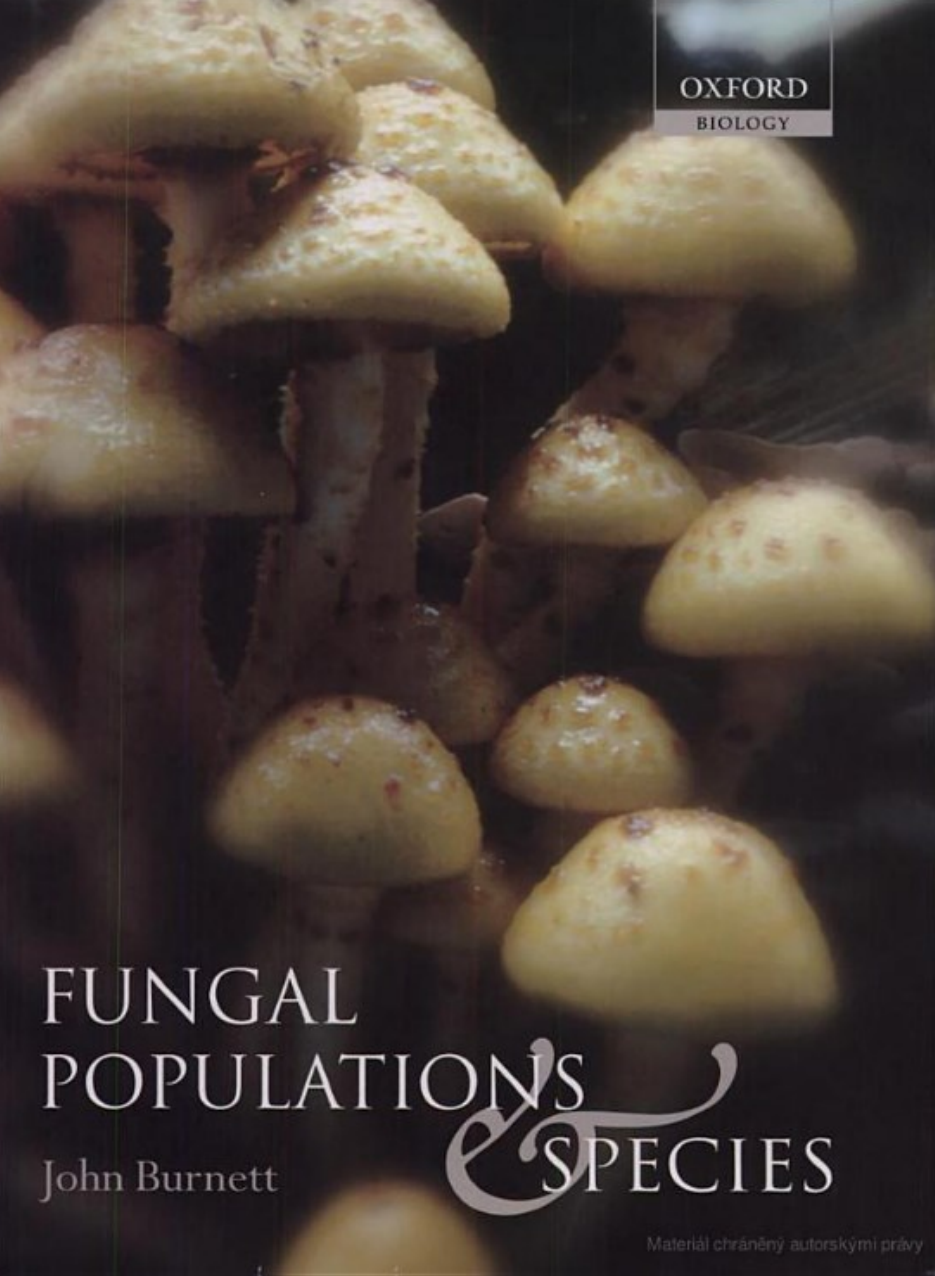


2005

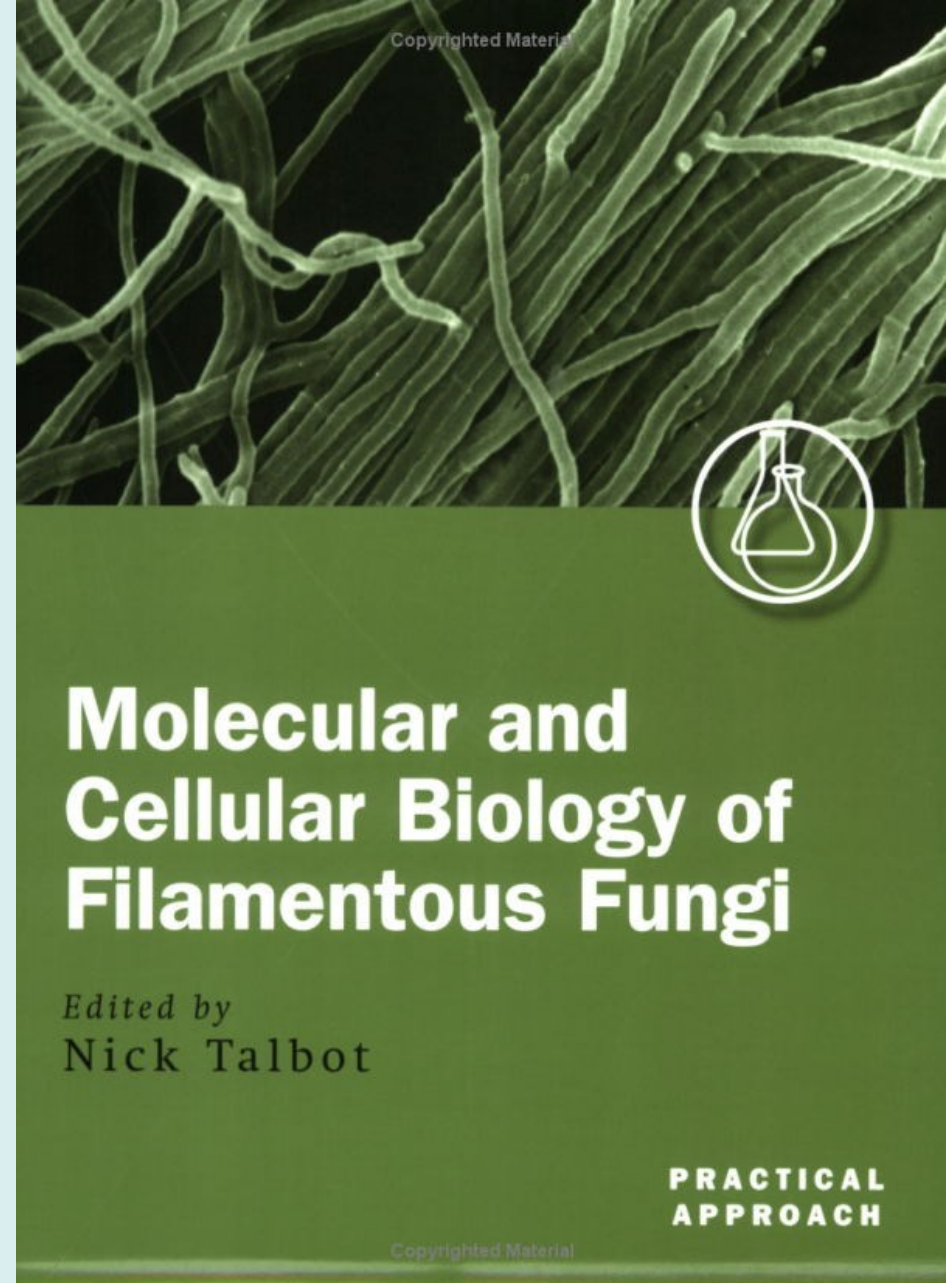


2004

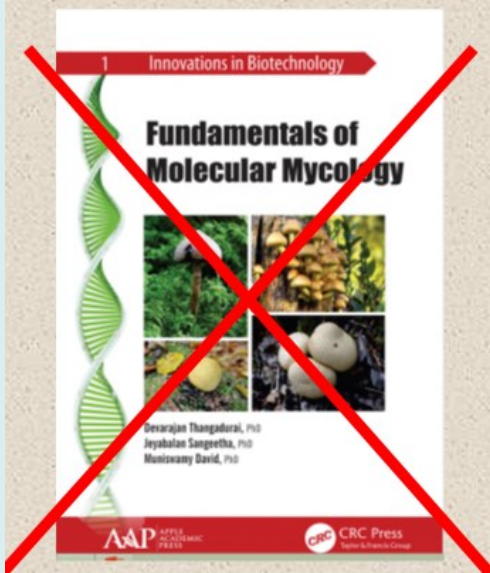




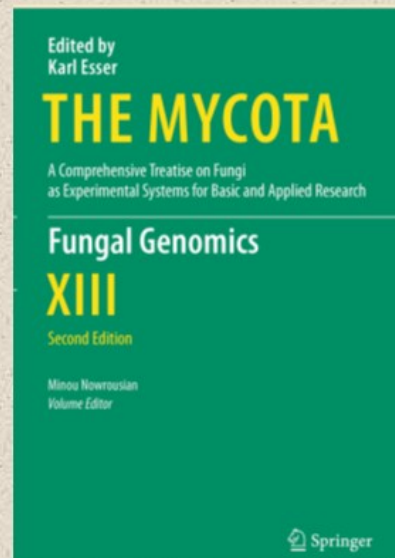
2003



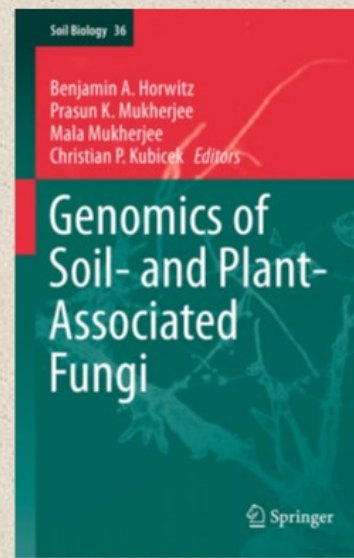
2001



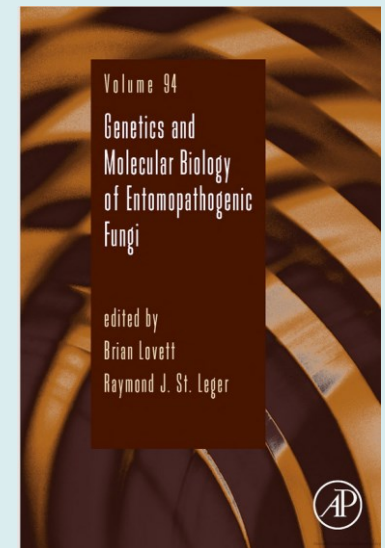
2016



2014



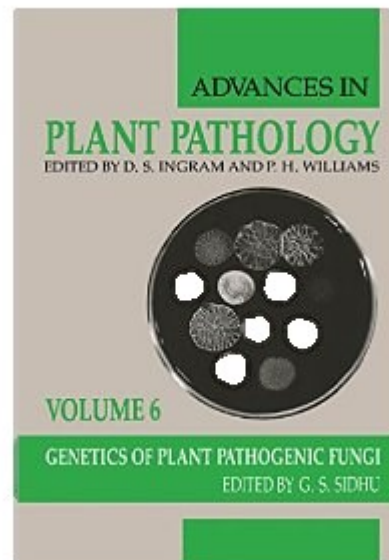
2013



2016



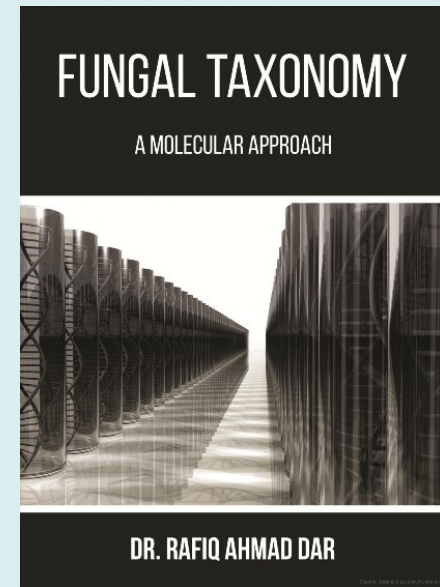
2014



2016



2016



2017

21st Century Guidebook to Fungi, SECOND EDITION
by David Moore, Geoffrey D. Robson and Anthony P. J. Trinci

revised 2020 edition

Now updated to 2020 online ahead of publication

A **COMPLETE** update and revision of this comprehensive text book of general fungal biology

In the form you are using this 'ebook' at the moment it is arranged as a website

Each page has its own **navigation** and **search** tools.

Treat yourself

PDF files of the Chapters in

21st Century Guidebook to Fungi Online

are available at only \$5 each

[You can get the complete 2020 version for only \$30]

[CLICK HERE](#) to view the order page

Choose a page to visit from the [blue](#) hyperlinks in the complete Table of Contents below

http://www.davidmoore.org.uk/21st_Century_Guidebook_to_Fungi_PLATINUM/

the hibbett lab

at Clark University

Navigation

[People](#)
[Projects](#)
[Publications](#)
[Lab News](#)
[Tour](#)
[Links](#)
[Contact Us](#)
[Clark Biology](#)
[Clark University](#)

Our lab studies evolutionary biology and ecology of Fungi, principally Basidiomycota (mushroom-forming fungi and relatives). Much of our work is focused on molecular systematics and the uses of phylogenetic trees for studying the evolution of morphological and ecological characters.

Other projects involve molecular ecology, evolution of lignin-degrading enzymes, and phyloinformatics.

Follow the links on this page to learn more about our current research projects, outreach efforts, and opportunities for graduate study.



Site designed & built by Julie Schaefer



Resources

[Poly PEET Project](#)

taxonomy and evolution of the Polyporales (Basidiomycota, Fungi)

[AFTOL Classification Project](#)

an ordinal-level classification of Fungi

[more automated phylogenetic taxonomy](#)

automated phylogenetic taxonomy of Fungi

[Clark Fungal Database](#)

search DNA & specimen records

[Lab Protocols](#)

includes primer information

[Undergrad courses & research](#)

with syllabi

[Prospective grad students](#)

wanted: talented graduate students

[Teaching the Fungal Tree of Life](#)

for K-12 teachers and students

[Outreach activities](#)

programs for high school students & biology teachers



Photo: Irwin Brodo

[François Lutzoni's page](#)
[Lab Members](#)
[Visitors](#)
[Group Photos](#)
[News Release](#)
[Publications & Data Sets](#)
[Downloadable Programs](#)
[Primers](#)
[Ascomycota Phylogenies](#)
[Peltigera Project](#)
[Teloschistales Project](#)
[AFToL](#)
[Assembling the Fungal Tree of Life](#)
[EnDoBiodiversity.org](#)
[Lichen Collection](#)


Lichenology group at Duke University during Fall 2011. From left to right, back row: François Lutzoni, Emilie Lefevre, Olaf Mueller & Martin Ramirez. Middle Row: Kathleen Miglia, Eimy Rivas Plata, Molly McMullen, Kohsuan Chen, Nicolas Magain & Chioita Culberson. Front row: Jolanta Miadlikowska, Tami McDonald, Ryoko Oono & Kathryn Picard.

Not pictured: Ester Gaya & Daniele Armaleo.



Latest News

August 2012
Welcome, Edgar Medina, new Ph.D. Student

June 2012
Tami McDonald is the recipient of the 2012 Harold Sanford Perry Prize!

April 2012
Welcome, Ko-Hsuan Chen, new Ph.D. Student

November 2011
Congratulations, Dr. Tami McDonald!

November 2011
Congratulations, Kathryn Picard, for successfully passing her prelim exam

November 2011
Welcome, Emily Lefevre, New Postdoctoral Research Associate

October 2011
Welcome, Geir Hestmark, Visiting Researcher

September 2011
Welcome, Nicolas Magain, Visiting Researcher

Useful links

[Assembling the Fungal Tree of Life \(AFToL\)](#)
[Cyanosite](#)
[Deep Hypha](#)
[Duke Lichen Herbarium](#)
[Duke University - Biology Department](#)
[EnDoBiodiversity.org](#)
[Index to American Botanical Literature](#)
[International Association for Lichenology](#)
[Lichen determination keys](#)
[Links to Lichens and Lichenologists](#)
[Mycological Society of America \(MSA\)](#)
[Myconet](#)
[National Center for Biotechnology Information](#)
[National Science Foundation](#)
[North American Lichen Checklist](#)
[North American Lichens and Bryophytes: Sensitive Indicators of Environmental Quality and Change](#)
[Phylogeny Software Programs](#)
[Recent Literature on Lichens](#)
[Sharnoff's photos of lichens](#)
[Systematic Botany and Mycology Fungal Databases](#)
[Tree of Life web project \(ToL\)](#)
[World Data Centre for Microorganisms](#)
[WWW Virtual Library - Mycological Resources](#)





the Bruns lab

321 Koshland Hall

Department of Plant and Microbial Biology
University of California at Berkeley

For an overview of the research currently being
 conducted in the Bruns Lab [click here](#).



Lab Personnel

- Rachel Adams
- Ben Araki
- Sara Branco
- Tom Bruns
- Thea Chesney
- Gerry Cobian
- Sydney Glassman
- Sobia Ilyas
- Don Julio
- Holly Mendonca
- Nhu Nguyen
- Abdul Razaq
- Lisa Rosenthal
- Tim Szaro
- Else Vellinga

Former Post Docs and Grad Students

- Anthony Amend
- Jacqueline Baar
- Gery Baura
- Martín Bidartondo
- Enrico Bonello
- Francisco Camacho
- Ken Cullings
- Matteo Garbelotto
- Monique Gardes
- Roberto Garibay Orijel
- Paul Grogan
- Lisa Grubisha
- Tasha Teutsch Hausmann
- Nils Högberg
- Tom Horton
- Nicole Hynson
- Antonio Izzo
- Peter Kennedy
- Jennifer Kerekes
- Rasmus Kjeller
- Annette Kretzer
- Tonya Lander
- Anna Levin
- Yunan Li
- Erik Lilleskov
- Jane Marks
- Vincent Merckx
- Kabir Peay
- Liz Pine
- Dirk Redecker
- Nick Rosenstock
- Shannon Schechter
- Lee Taylor
- Det Vogler
- Val Wong

Former Undergraduates

- Josephine Agbowo
- Alison Bender
- Kristi Bischel
- Aiona Bones
- Ryan Bowman
- Primrose Boynton
- Danielle Bricker
- Sophie Burda-Wilkin
- Megan Canright
- Chieh-I Chen
- Kari Cheng
- Karen Chiu
- Judy Chung
- Loleta Chung
- Remi Cohen
- Michael Ernandes
- Amanda Gaide
- Alex Grote
- Chrissie Howell
- Julie Kemp
- Pei-ying Liao
- Sermed Naaman
- Diem Thi Nguyen
- Jennifer Nguyen
- Jean Pan
- Natalia Pinzon
- Andricia Ramirez
- Annie Rowe
- Theresa Hendrick Rusca
- Max Schubert
- Rima Shamieh
- Eric Stendell
- Jenny Tan
- Lauren Thompson
- Vicky Valinluck
- Bernadette Lato Verbeke
- Ted Szeto
- Elanore Zhou
- Sara Zlotnik

Other Emeritus Personnel

- Eden Abram
- Glade Dlott
- Haley Dunleavy
- Esther McLaughlin
- Tina Popenuck
- David Read
- Clare Robinson
- Bitty Roy
- Caroline Sidi
- Ian Singleton

www.zealsoft.com

view the Fungi we study

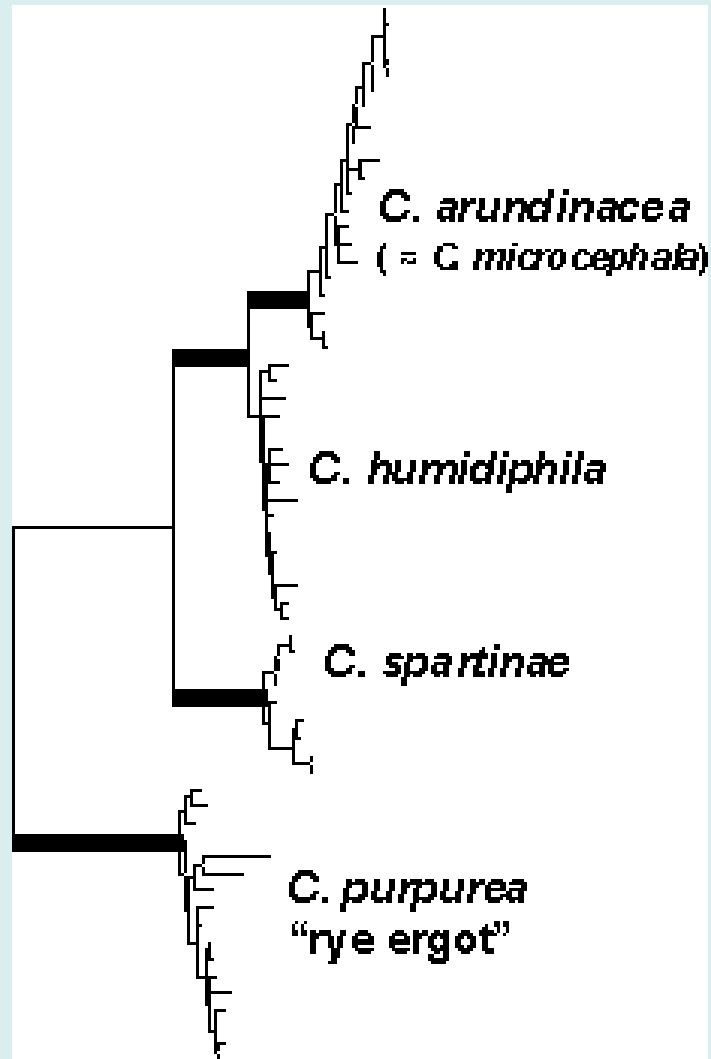
- Mushrooms
- Mycorrhizal Fungi Part I
- Mycorrhizal Fungi Part II
- Rust Fungi
- Post Fire Fungi from Point Reyes

Useful Items

- Primer Sequences
- Primer Maps:
 - Primer Map
 - 18S Primer Map
 - 28S Primer Map
 - ITS Primer Map
 - Mitochondrial Large Primer Map
- Sequence Alignment Files:
 - Amyloporon Paup file
 - atp6+mtLSU.aln Paup file
 - atp6.all_fungi.aln Paup file
 - Cronartium ITS Paup file
 - Glomales 5.8S Paup file
 - ML56 database: the ML56 database used in the Bruns (1998) et. al. paper (152 taxa)
 - ML56 database: an updated version (159 taxa) (January 1998)
 - ML56 database: updated by Tom Horton (175 taxa) (November 2000)
 - nuclear ribosomal 5.8S RNA gene sequence database
 - Suillus ITS Paup file used in Kretzer (1996) et. al. paper: [text only version](#) of the file
 - Gautieria ITS Paup file used in Bidartondo & Bruns (2002) paper.
 - Tricholoma ITS Paup file used in Bidartondo & Bruns (2002) paper.
- Famous Mycologists
- Fungal Environmental and Informatics Network
- Mushroom Cards
 - [readme file](#) [pdf]
 - [Mushroom cards](#) [pdf]
 - [Mushroom cards \(no lichens\)](#) [pdf]
 - [Mushroom cards](#) [xls]
- Mycorrhizal Reading Group
- Point Reyes MycoBlitz
- Point Reyes MycoBlitz fungal collections and photos

Druh jako základní jednotka diverzity

- Biologická diverzita netvoří kontinuum, ale je organizována do diskretních skupin



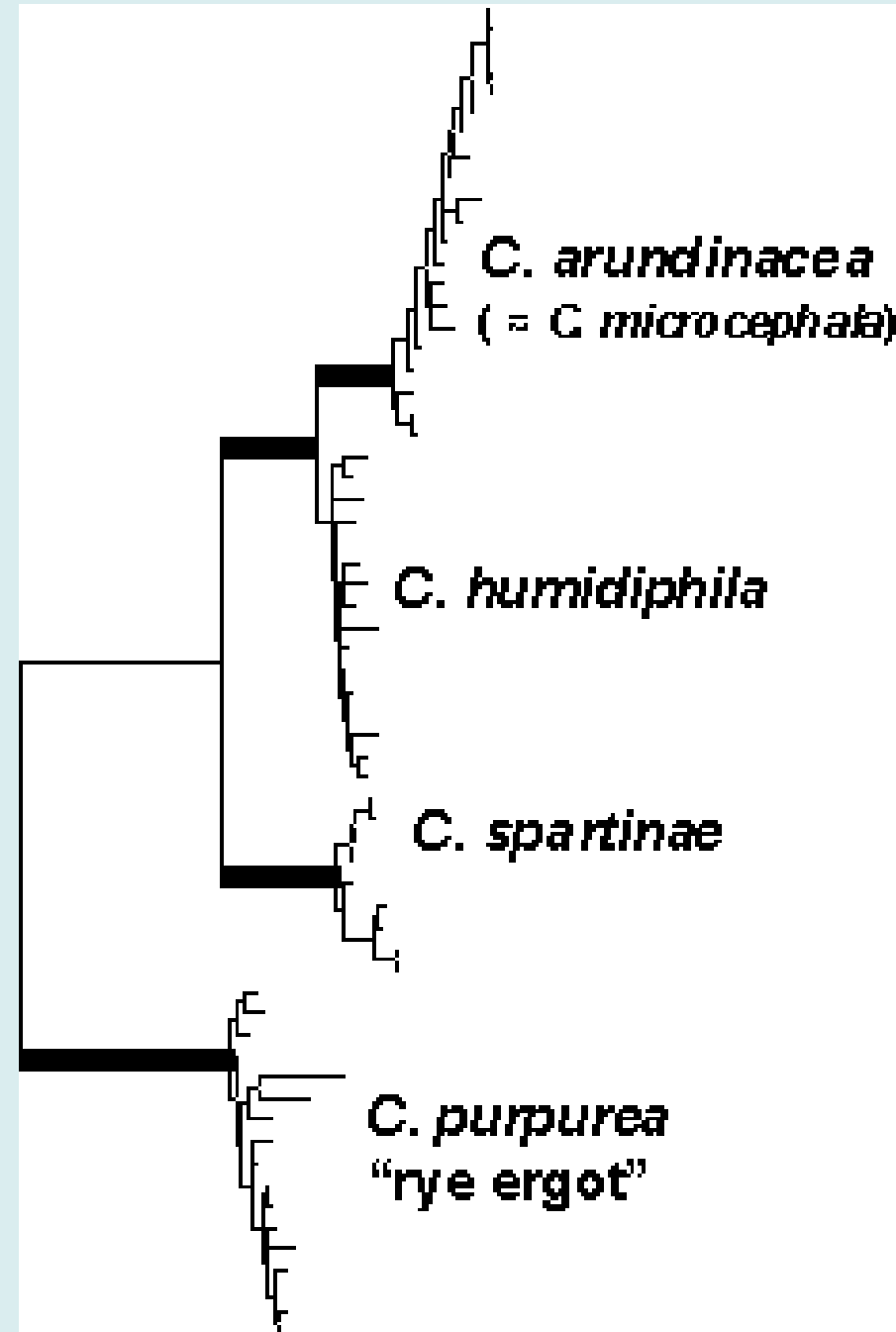
Druh jako základní jednotka diverzity

- **The morphological species concept...**
-
- **The biological species concept.** Species is defined as an interbreeding population that is somehow reproductively isolated from other populations. Not working in asexual fungi (?)
- **Ecological and physiological species concepts.** Differentiates species by their ecological niche and the constraints on their evolution that determine their maintenance and reproduction in that niche. Often in pathogens.
- **Evolutionary/phylogenetic species concept.** Envisages a species as being a monophyletic group of organisms sharing molecular characters that derive from a common ancestor.
- **Consolidated Species Concept (CSC).** Species are distinguished based on a polyphasic approach, combining morphological, ecological and phylogenetic species concepts

- **Proč jsou druhy?**

1) druhy jsou výsledek adaptace k různým ekologickým nikám

2) druhy jsou důsledkem reprodukčních bariér
(biologických i nebiologických).



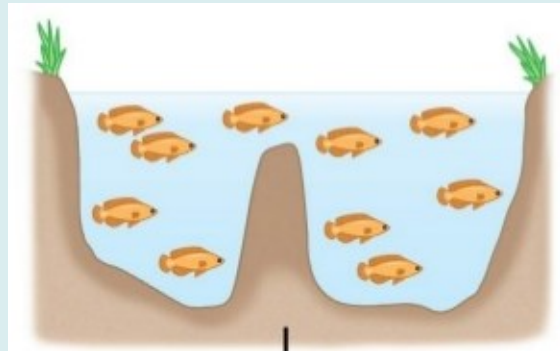
species as a result of the adaptation to existing niches
individuals sharing the same ecological, physiological,
phenotypic and genetic traits

Darwin's finches

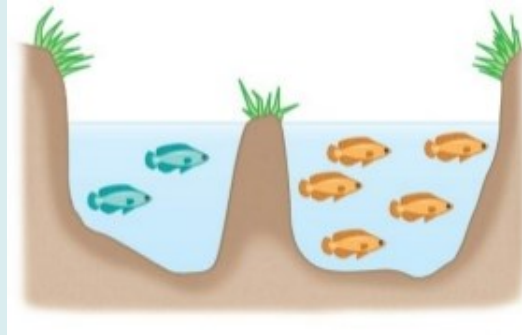




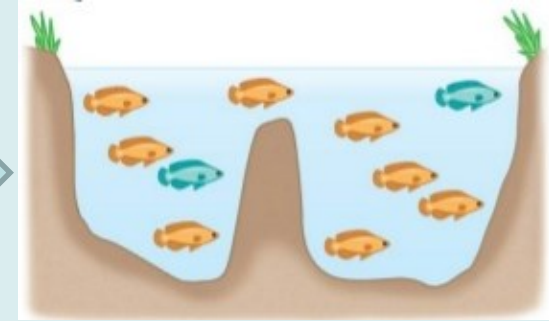
species as a results of the reproductive barriers



ancestral population



Period of isolation



and species meet again...

geographic barrier



X



mating disruption

- Allopatrická speciece následovaná migrací do areálu původního druhu
- možnost křížení s původním druhem, se kterým nebyla nastolena reprodukční bariéra – vznik nových patogenů

Proceedings of the 4th International Workshop on Genetics of Host-Parasite Interactions in Forestry

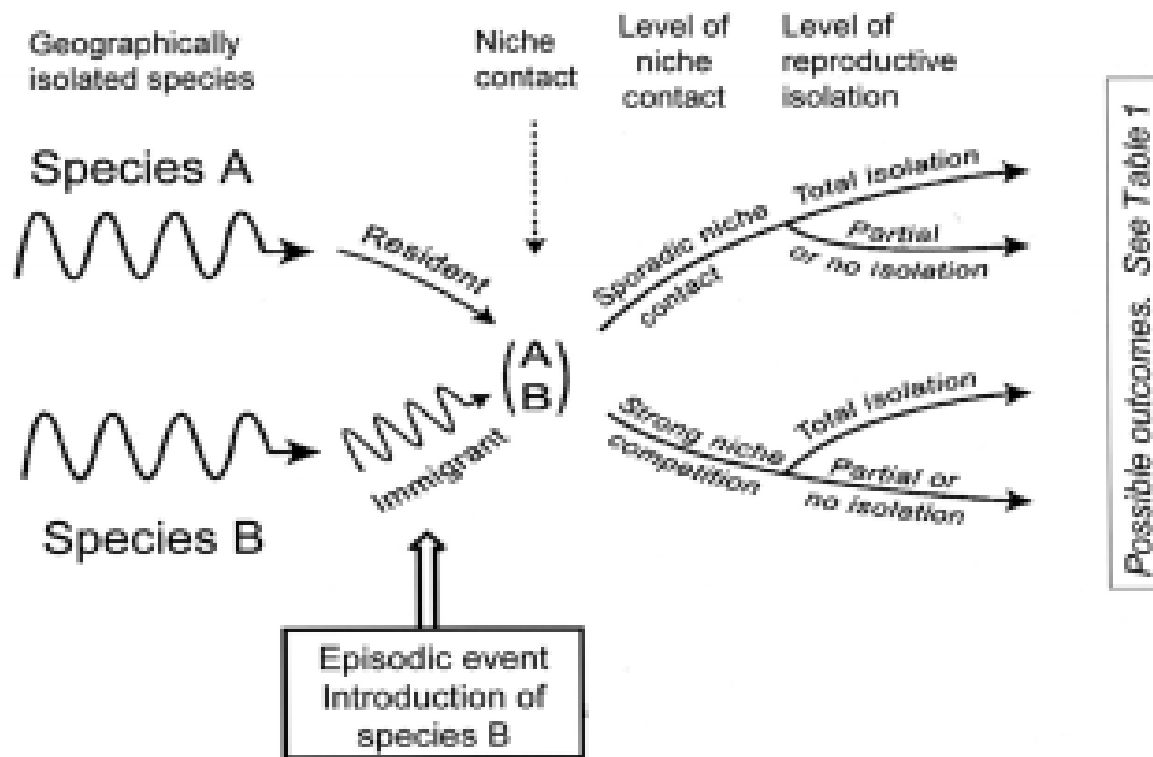


Figure 2—Genetical and environmental constraints on emergence of species hybrids in fungal pathogens. Adapted from Brasier (1995).

Molekulárně taxonomické metody

- klasifikují organismy na podkladě studia jejich **genomu** (DNA, RNA) a jeho translačních produktů (**proteiny**)
- nepatří sem metody využívající jiné molekuly (tzv. **chemical taxonomy**)

CLINICAL MICROBIOLOGY REVIEWS, July 1999, p. 454–500
0893-8512/99/\$04.00+0
Copyright © 1999, American Society for Microbiology. All Rights Reserved.

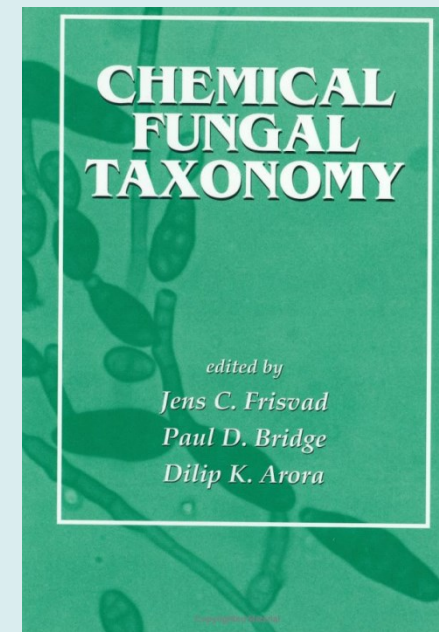
Vol. 12, No. 3

Developments in Fungal Taxonomy

JOSEP GUARRO,* JOSEPA GENÉ, AND ALBERTO M. STCHIGEL

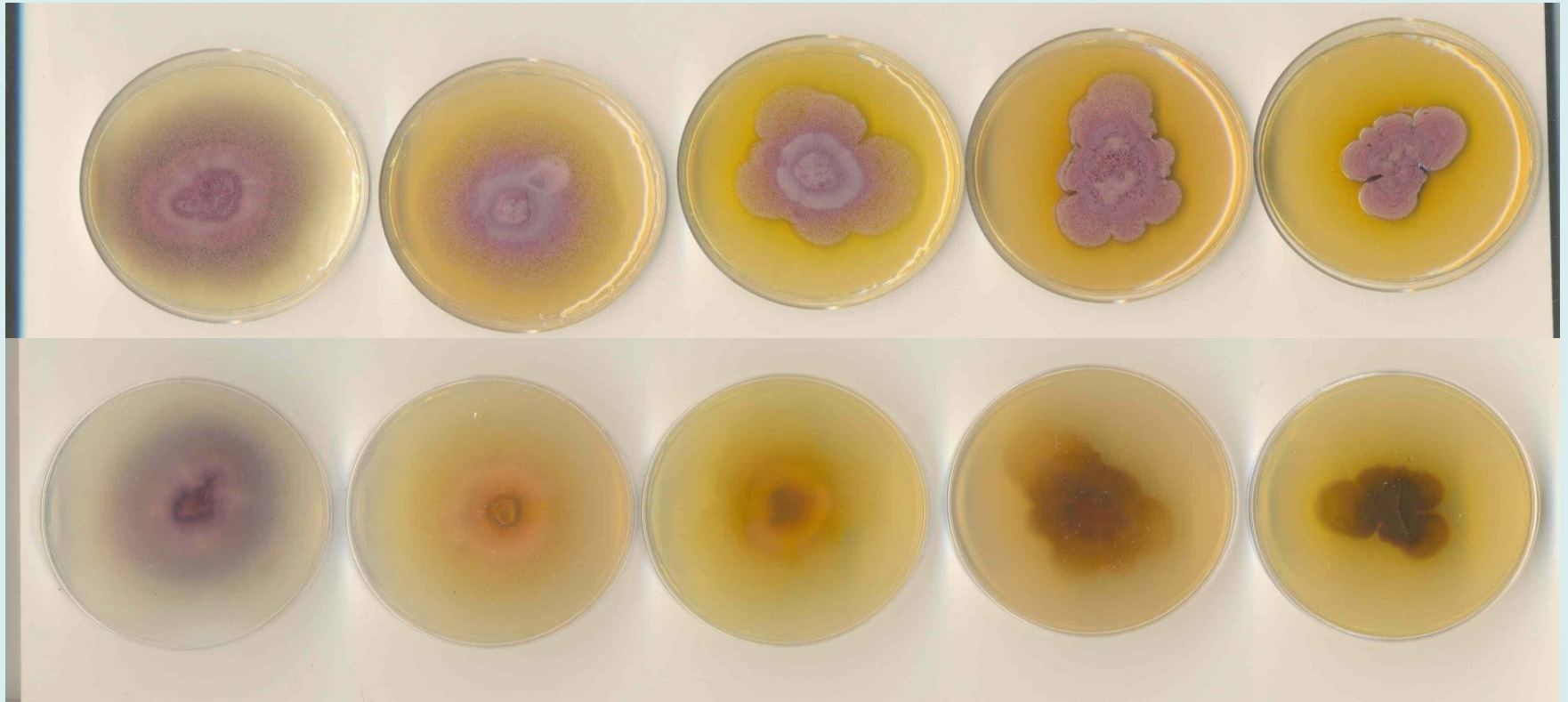
*Unitat de Microbiologia, Departament de Ciències Mèdiques Bàsiques, Facultat de Medicina i
Ciències de la Salut, Universitat Rovira i Virgili, 43201 Reus, Spain*

INTRODUCTION	454
THE CONCEPT OF SPECIES IN FUNGI	455
PHYLOGENY AND EVOLUTION	455
NOMENCLATURE	456
CURRENT MYCOLOGICAL TYPING METHODS	457
Morphology	457
Molecular Techniques	459
Other Techniques	460
Physiological and biochemical techniques	460
Secondary metabolites	460
Ubiquinone systems	461
Fatty acid composition	461
Cell wall composition	461
Protein composition	461



Velká fenotypová plasticita hub

- zpětné posuzování kvality fenotypových znaků



Mutanti jednoho monosporického kmene *G. lavendula*

Biochemické markery - chemotaxonomie

Nejčastější metody

- sekundární metabolity
- Proteinové profily (MALDI...)
- převažující ubichynonový systém
- profil mastných kyselin a sterolů (FAST- profiles)
- profil polysacharidů buněčné stěny
- izozymy, allozymy

Použití

- přídatné markery v taxonomii
- identifikace významných hub (aplikovaná mikrobiologie)
- cenná data, ale problematická interpretace (fenoplasticita)

Biochemické markery – sek. metabolity

FRISVAD at al. 2007: The use of secondary metabolite profiling in chemotaxonomy of filamentous fungi. – Mycol. Res.

- látky které nejsou nezbytné pro život
- důležité markery u řady skupin (lišejníky, *Aspergillus*, *Penicillium*, *Fusarium*, *Xylaria*, *Daldinia*, *Hypoxylon*, tvrdohouby obecně)
- navíc získáme informaci o chemickém složení organismu (toxiny atd.)
- **možnost zpracovat staré herbářové položky**

Postup

extrakce rozpouštědlem (MetOH, Aceton, EtylAcetát, EtOH)



Chromatografická separace (TLC, HPLC, UPLC, případně plynová chromatografie u těkavých látek)



Detekce látek (dle polohy a zbarvení na TLC, dle hmoty, UV spekter, NMR, krystalu)

Biochemické markery – sek. metabolismy

- Ehrlichův test u *Penicillium* a *Aspergillus*
 - detekce indolů přímo na agarové plotně
- na agar (či otisk kolonie na filtračním papíru) se kápne Ehrlichova činidlo (dimethylaminobenzaldehyd, ethanol, HCl)
- pozitivní reakce je modré zbarvení

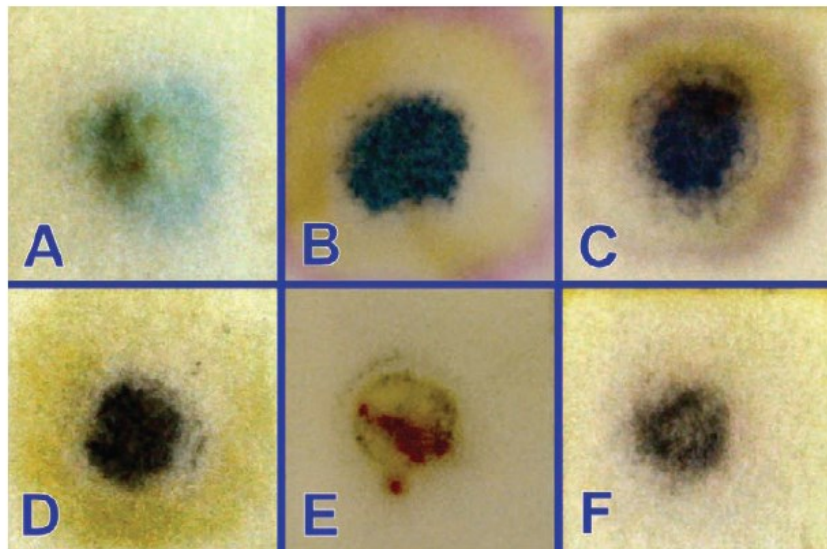


Fig. 8. Ehrlich colour reaction of some *Aspergillus* species in section *Nigri*. (A) blue in *A. costaricensis*, (B) purple ring in *A. heteromorphus*, (C) purple ring in *A. homomorphus*, (D) yellow-green in *A. niger*, (E) purple-red at sclerotial area of *A. sclerotioriger* and (F) no reaction in *A. vadensis*.



Biochemické markery – sek. metabolity

Table I. Secondary products isolated and characterized from taxa of *Pseudocyphellaria*. (Results expressed as % dry weight).

	<i>P. nitida</i>	<i>P. meyenii</i>	<i>P. Exanthem</i>	<i>P. divulsa</i>	<i>P. berberina</i>
Stictic acid 1	0.02		0.26	0.08	
Norstictic acid 2			0.08		
Hyposalazinic acid 3			0.08		
Hypostictic acid 4			0.05		
Constictic acid 5				0.04	
Cryptostictic acid 6				0.06	
Tenuiorin 7	0.10	0.18	0.67	1.92	
Calycin 8					0.04
Pulvinic dilactone 9					0.69
Pulvinic acid 10					0.04
Zeorin 11	0.03		0.23	0.41	
7 β -acetoxy-22-hydroxyhopane 12	0.07	0.45	0.74	4.08	
Leucotylin 13		0.03			
Ergosterol peroxide 14	0.13	0.06	0.03	0.31	

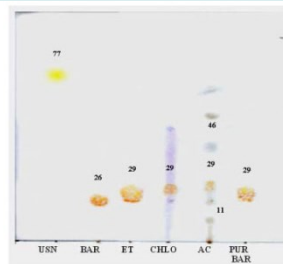
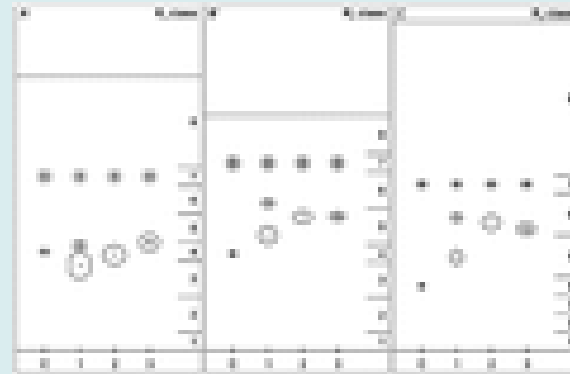


Figure 1 - Thin layer chromatogram of organic extracts of *Cladia aggregata* (Sw.) Nyl. Captions: UNS – standard usnic acid, BAR – standard barbatric acid, ET – ether extract, CHLO – chloroform extract, AC – acetone extract, PUR BAR – purified barbatric acid. Numbers means the spots' R_f values.



TLC sekundárních metabolitů lišejníků (acetonový extrakt)

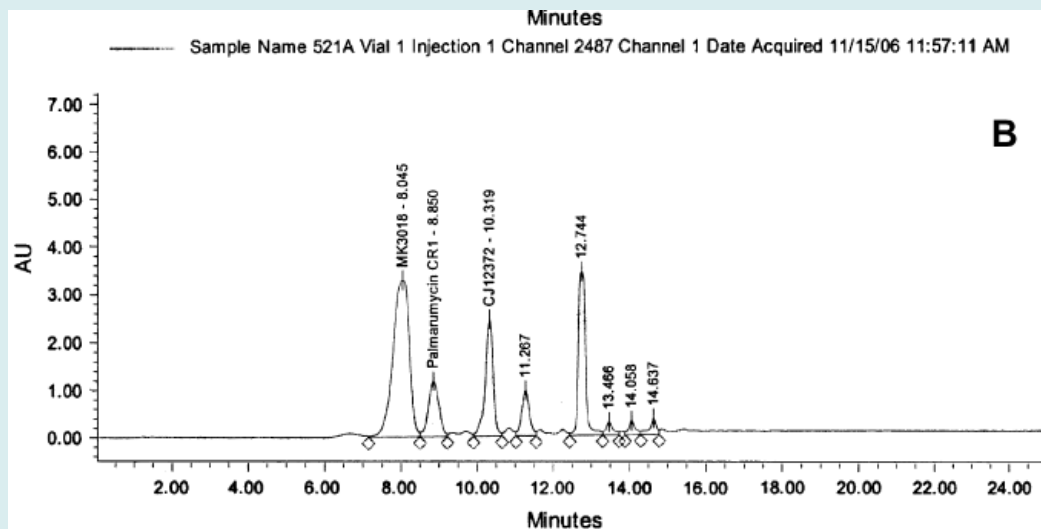
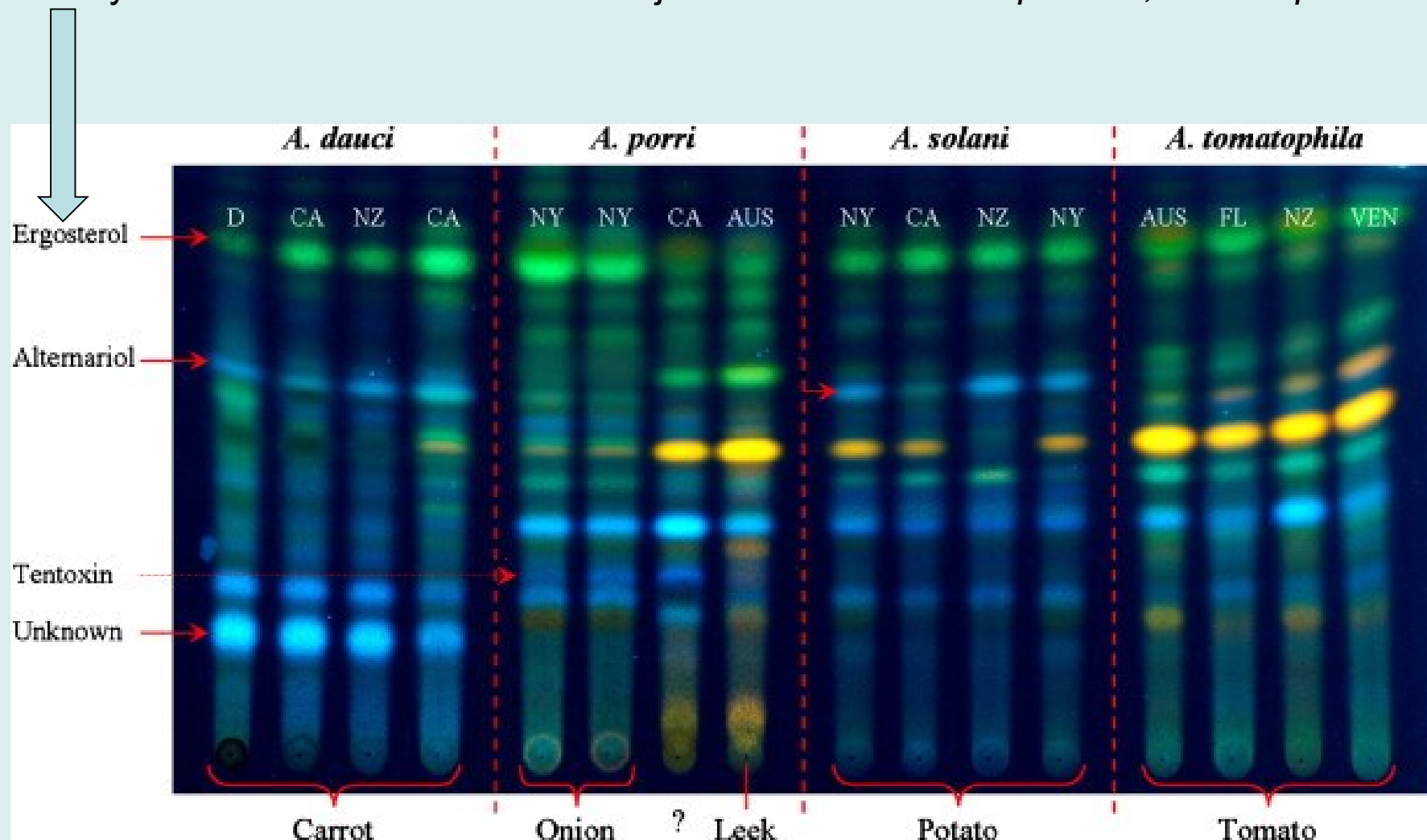


Fig 3 – Hplc traces (220 nm) of crude extracts prepared from liquid cultures of *Rhytidhysterium rufulum* from Costa Rica. The occurrence of the peaks is indicated in the corresponding chromatograms (A) sample 521A-clade I, (B) sample 512A-clade II, (C) sample 525A-clade III, and (D) sample 510A-clade IV.

- často není jisté, že jde o sek. metabolity, protože může jít i o látky s neznámou funkcí, či látky neidentifikované ...adekvátní je termín *natural compounds*, *natural products*



TLC po sprejování $H_2SO_4/MeOH$ a pozorovaná při 366 nm (UV). Převzato z Frisvad et al. (2008)

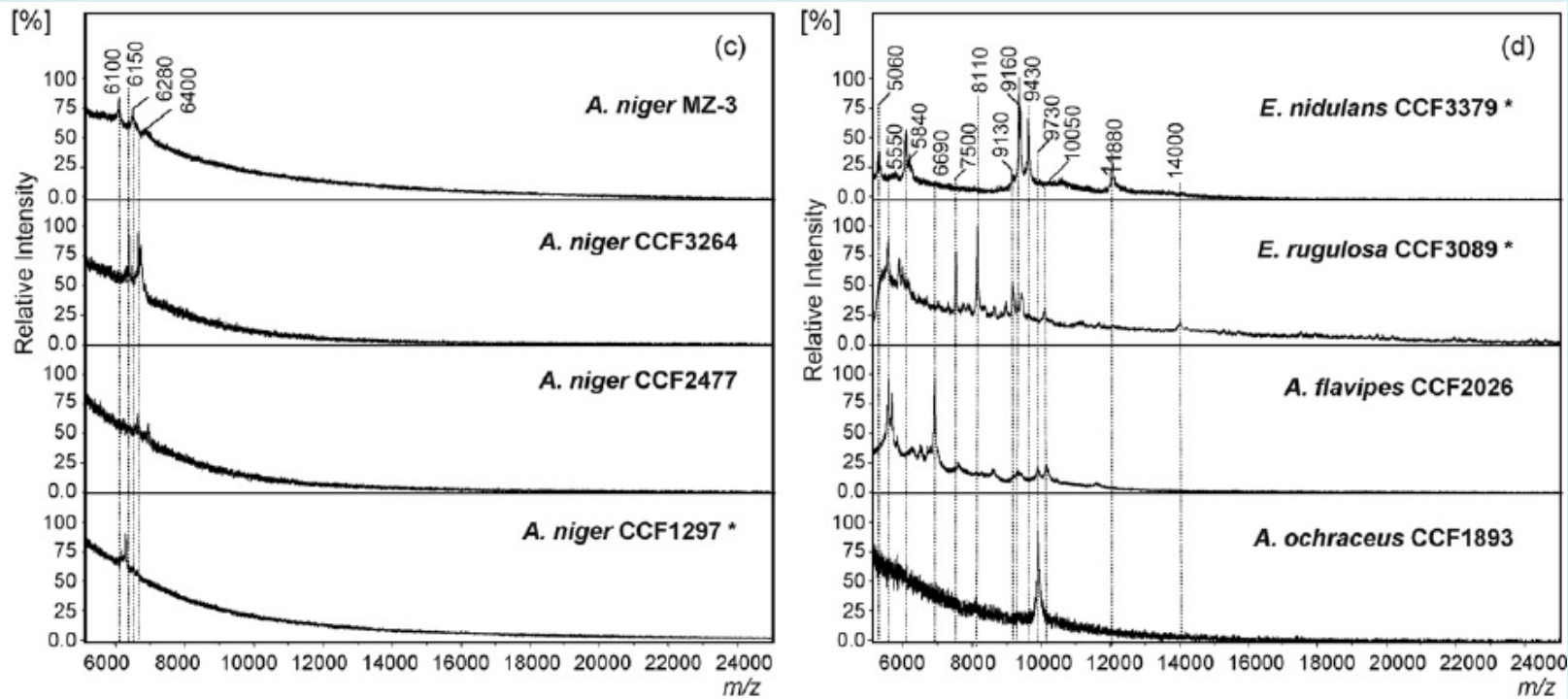
Table 2. Extrolites produced by species assigned to *Aspergillus* section *Fumigati*.

Species	Extrolites produced
<i>Aspergillus brevipes</i>	roquefortine C, meleagrins-like
<i>Aspergillus duricaulis</i>	pseurotin A, fumagillin, asperpentyn, duricaulic acid and asperdurin, phthalides, chromanols, cyclopaldic acid, 3-O-methylcyclopolic acid
<i>Aspergillus fumigatiaffinis</i>	auranthine, cycloechinuline, fumigaclavines, helvolic acid, neosartorin, palitantin, pyripyropenes A, E, O & S, tryptoquivaline, tryptoquivalone
<i>Aspergillus fumigatus</i>	fumagillin, fumitoxins, fumigaclavines A & C, fumitremorgins, gliotoxin, trypacidin, pseurotins, helvolic acid, pyripyropenes, methyl-sulochrin, verruculogen, fumiquinazolines
<i>Aspergillus fumisynnematus</i>	neosartorin, pyripyropenes, fumimycin
<i>Aspergillus lentulus</i>	cyclopiazonic acid, pyripyropenes A, E & O, terrein, auranthine, neosartorin
<i>Aspergillus novofumigatus</i>	aszonalenin, cycloechinuline, fiscalins, helvolic acid, neosartorin, palitantin, terrein, territrem B
<i>Aspergillus turcosus</i>	kotanins and several unique but not yet elucidated secondary metabolites
<i>Aspergillus unilateralis</i>	mycophenolic acid, other unique secondary metabolites
<i>Aspergillus viridinutans</i>	viriditoxin, 13-O-methylviriditin, phomaligin A, variotin, viriditin, wasabidienone B0, B1, viriditin, 4-acetyl-6,8-dihydroxy-5-methyl-2-benzopyran-1-1 A
<i>Neosartorya assulata</i>	indole alkaloids and apolar metabolites
<i>Neosartorya aurata</i>	helvolic acid, yellow unidentified compounds
<i>Neosartorya aureola</i>	fumagillin, tryptoquivaline, tryptoquivalone, pseurotin A and viriditoxin (FRR 2269 also produces helvolic acid)
<i>Neosartorya australensis</i>	wortmannin-like, aszonalenin-like
<i>Neosartorya coreana</i>	aszonalenins
<i>Neosartorya denticulata</i>	gliotoxin, viriditoxin

Biochemické markery – MALDI-TOF profil

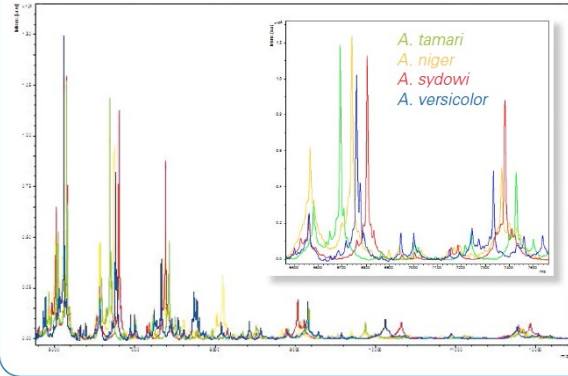
- Vytvoření proteinového profilu (fingerprintu), který je porovnán s databází
- levná a rychlá metoda, široce používaná v klinické a průmyslové mikrobiologii
- ke stanovení molekulových hmotností se používá metoda **hmotnostní spektrometrie používající ionizace laserem za přítomnosti matrice** (MALDI, *matrix assisted laser desorption/ionization*) v kombinaci s detektorem doby letu (TOF, *time-of-flight*)
- Abundantní proteiny či jejich kombinace jsou charakteristické pro druh (či nižší jednotku)

International Journal of Mass Spectrometry 280 (2009) 162–168



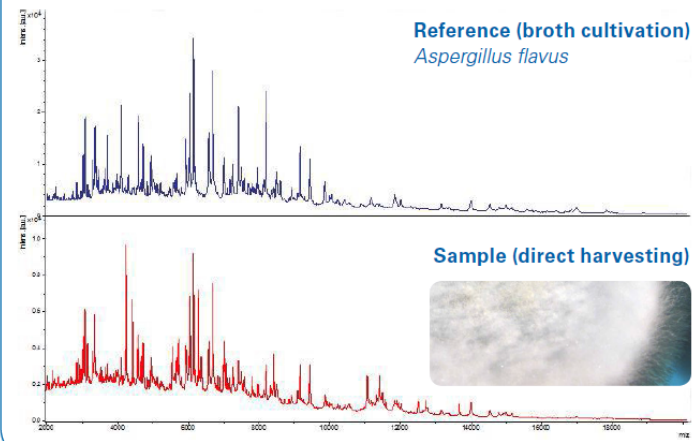
- MALDI Biotyper firmy Bruker

Different species of the genus *Aspergillus*



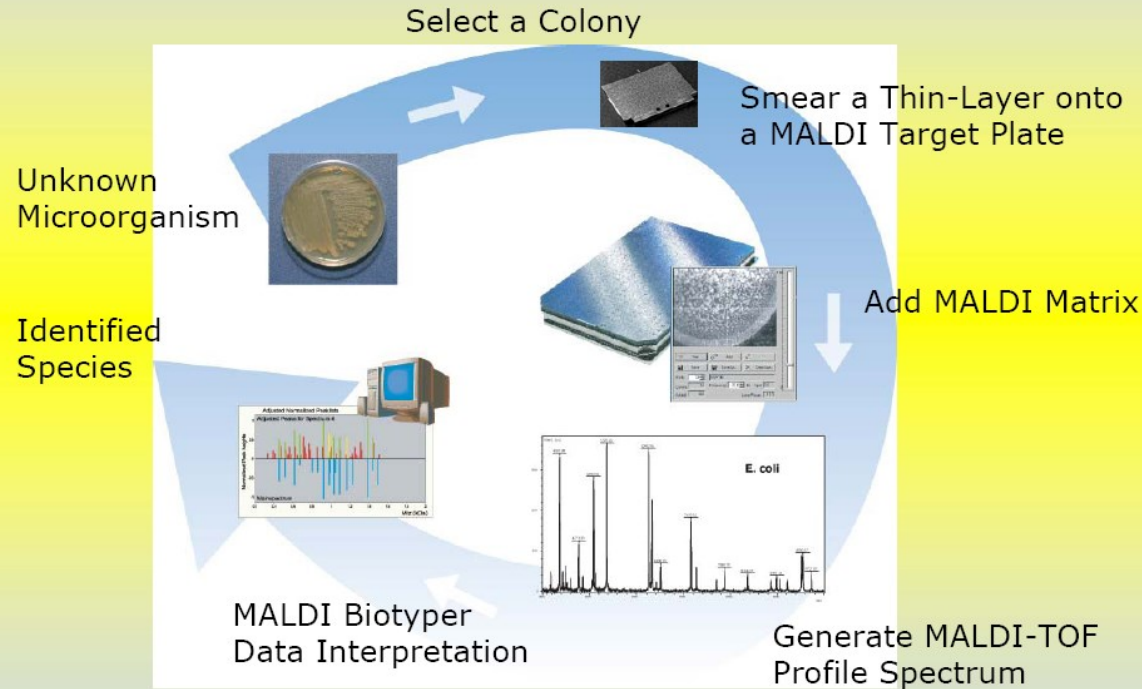
MALDI Biotyper spectra overlay of various species of *Aspergillus*.

Typical MALDI Biotyper mass spectra



Top spectrum is achieved after liquid cultivation of *Aspergillus flavus* and the bottom spectrum is achieved by direct harvesting of *Aspergillus flavus* from agar. Note from the image that the "front mycelium" is both clearly visible and can be easily harvested.

MALDI Biotyper Workflow



127 species entries

<i>Absidia coerulea</i>	<i>Cladosporium</i> sp	<i>Monilinia laxa</i>	<i>Phaeoacremonium</i> sp
<i>Acremonium strictum</i>	<i>Cunninghamella elegans</i>	<i>Mucor circinelloides</i>	<i>Phialemonium</i> sp
<i>Alternaria alternata</i>	<i>Curvularia clavata</i>	<i>Mucor ramosissimus</i>	<i>Phialophora bubakii</i>
<i>Arthrinium phaeospermum</i>	<i>Curvularia lunata</i>	<i>Paecilomyces farinosus</i>	<i>Phoma glomerata</i>
<i>Arthrographis kalrae</i> [ana] (<i>Eremomyces langeronii</i> [teleo])	<i>Curvularia pallescens</i>	<i>Paecilomyces lilacinus</i>	<i>Phoma herbarum</i>
<i>Aspergillus candidus</i>	<i>Curvularia verruculosa</i>	<i>Paecilomyces marquandii</i>	<i>Phoma sorghina</i>
<i>Aspergillus clavatus</i>	<i>Epicoccum nigrum</i>	<i>Paecilomyces variotii</i>	<i>Rhizomucor pusillus</i>
<i>Aspergillus flavus</i>	<i>Epidermophyton floccosum</i>	<i>Penicillium brevicompactum</i>	<i>Rhizopus microspores</i>
<i>Aspergillus fumigatus</i>	<i>Fennellia flavipes</i>	<i>Penicillium chrysogenum</i>	<i>Rhizopus oryzae</i>
<i>Aspergillus glaucus</i>	<i>Fusarium aquaeductuum</i>	<i>Penicillium citreonigrum</i>	<i>Rhizopus stolonifer</i>
<i>Aspergillus niger</i>	<i>Fusarium cerealis</i>	<i>Penicillium citrinum</i>	<i>Scedosporium prolificans</i>
<i>Aspergillus nomius</i>	<i>Fusarium chlamydosporum</i>	<i>Penicillium commune</i>	<i>Scedosporium apiospermum</i> [ana] <i>Pseudallescheria boydii</i> [teleo]
<i>Aspergillus ochraceus</i>	<i>Fusarium culmorum</i>	<i>Penicillium corylophilum</i>	<i>Schizophyllum commune</i>
<i>Aspergillus oryzae</i>	<i>Fusarium dimerum</i>	<i>Penicillium crustosum</i>	<i>Scopulariopsis acremonium</i>
<i>Aspergillus parasiticus</i>	<i>Fusarium equiseti</i>	<i>Penicillium daleae</i>	<i>Scopulariopsis brevicaulis</i>
<i>Aspergillus sclerotiorum</i>	<i>Fusarium incarnatum</i>	<i>Penicillium dierckxii</i>	<i>Scopulariopsis brumptii</i>
<i>Aspergillus sydowi</i>	<i>Fusarium moniliforme</i>	<i>Penicillium digitatum</i>	<i>Scytalidium lignicola</i>
<i>Aspergillus tamarii</i>	<i>Fusarium oxysporum</i>	<i>Penicillium discolor</i>	<i>Sporothrix schenckii</i>
<i>Aspergillus terreus</i>	<i>Fusarium proliferatum</i>	<i>Penicillium expansum</i>	<i>Syncephalastrum racemosum</i>
<i>Aspergillus unguis</i>	<i>Fusarium solani</i>	<i>Penicillium funiculosum</i>	<i>Thanatephorus cucumeris</i>
<i>Aspergillus ustus</i>	<i>Fusarium tabacinum</i>	<i>Penicillium glabrum</i>	<i>Trichoderma koningii</i>
<i>Aspergillus versicolor</i>	<i>Fusarium verticillioides</i>	<i>Penicillium italicum</i>	<i>Trichoderma longibrachiatum</i>
<i>Aspergillus amstelodami</i> [ana] <i>Eurotium amstelodami</i> [teleo]	<i>Geomyces pannorum</i>	<i>Penicillium lanosum</i>	<i>Trichophyton eboreum</i>
<i>Aspergillus nidulans</i> [ana] <i>Emericella nidulans</i> [teleo]	<i>Geosmithia argillaceae</i>	<i>Penicillium olsonii</i>	<i>Trichophyton equinum</i>
<i>Aureobasidium pullulans</i>	<i>Lecythophora hoffmannii</i>	<i>Penicillium pseudostromaticum</i>	<i>Trichophyton interdigitale</i>
<i>Beauveria bassiana</i>	<i>Lichtheimia corymbifera</i>	<i>Penicillium purpurogenum</i>	<i>Trichophyton rubrum</i>
<i>Botrytis cinerea</i>	<i>Microsporium cookie</i>	<i>Penicillium roqueforti</i>	<i>Trichophyton tonsurans</i>
<i>Chaetomium funicola</i>	<i>Microsporium equinum</i>	<i>Penicillium rugulosum</i>	<i>Trichophyton violaceum</i>
<i>Chaetomium globosum</i>	<i>Microsporium fulvum</i>	<i>Penicillium</i> sp	<i>Trichophyton mentagrophytes</i> var. <i>erinacei</i> [ana] <i>Arthroderma benhamiae</i> [teleo]
<i>Chrysosporium keratinophilum</i>	<i>Microsporium gypseum</i>	<i>Penicillium striatisporum</i>	<i>Trichurus</i> sp
<i>Cladosporium cladosporioides</i>	<i>Microsporium persicolor</i>	<i>Penicillium turbatum</i>	
<i>Cladosporium herbarum</i>	<i>Microsporium praecox</i>	<i>Penicillium verrucosum</i>	

Biochemické metody - starší

byly dražší, náročnější, znaky často fenoplastické a s omezeným použitím

- **spektrum polysacharidů ve stěně** (vodo-rozpustné, extrahovatelné v alkáliích)
- **spektrum mastných kyselin v membránách** (převzato z bakterií, kde je větší diverzita těchto látek a tedy lepší rozlišení druhů). Poměrně fenoplastické. Spektrum je druhově specifické. Stále se používají pro odlišení bakteriální vs houbové biomasy. Analýza pomocí plynové chromatografie.
- **variability ubichinonů** (koenzym Q) – přenašeče elektronů v dýchacím řetězci se specificky liší v uspořádání isoprenových jednotek na chinonovém jádře (hlavně taxonomie vyššího řádu)
- **variabilita proteinů (analýza izozymů=izoenzymů)** proteiny o stejné funkci mohou mít mnoho forem s rozdílnou strukturou (tj. mobilitou na elektroforetickém gelu). Hlavně na populačně genetické studie.

Molekulárně taxonomické metody

- **Molekulární, neboli genetický znak (marker)** - úsek DNA (či jeho translačního produktu), používaný pro zjištění genetické příbuznosti porovnávaných organismů - např. sekvence určitého genu, proteinu, délka PCR fragmentu...
- **alternativní markery** – jiný marker může ukázat jiné výsledky, nutnost ověřovat a brát vše s rezervou
- kvalita výsledku dle počtu sledovaných markerů, lokusů - **uni x multilokusové metody**

Využití

vhodná volba markeru a metody

- identifikace (...jedinec, druh, rod)
-
- vymezení druhů, zejména těch kryptických (= *sibling species*, *microspecies*), které se neliší morfologicky
 - spojení anamorfy s teleomorfou
- studium evoluce dané skupiny (fylogenetické vztahy)
- studium diverzity (nekultivovatelné houby)
- populační genetika (zjištění příbuznosti populací, fylogeografie, sledování migrace, odlišení jedince)
- způsob rozmnožování (klonalita x pohlavní křížení)
- studium evoluce velikosti a komplexity genomu při speciaci

Povaha mol. tax. znaků

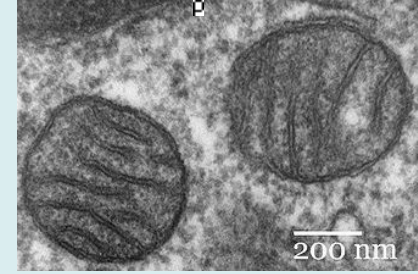
- **kvantita** - omezena jen finančně
- **diskrétní povaha**
- **selekční neutrálnost** – nepodléhá selekčnímu tlaku – ten se může měnit
- **studium skutečného polymorfismu** – u fenoplastických hub důležité
- **polární znaky** – možnost odlišit původní stav znaku
- **znaky jsou nevázané** – jedna evoluční událost neovlivní jiné znaky
- **omezeno riziko konvergence (homoplazie)** (neplatí u hypervariabilních úseků)

Stavba genomu hub

mt DNA, jaderná DNA, plazmidová DNA

mt DNA

- uniparentálně i biparentálně dědičná
- non-Mendelian inheritance, cytoplasmic inheritance
- u některých chytridií chybí (anaerobní)
- 27-174 kbp
- 30-40 genů (tj. málo)
- transpozóny
- **Důležité markery: mt rDNA (LSU, SSU), cytochrom oxidáza subunit I (COI, COX1)**



DNA plazmidová

- mitochondriální i cytoplasmatické plazmidy
- mohou se vmezeřit do *nc* i mtDNA (to se používá při genet. transformaci)
- lineární i cirkulární
- mohou přecházet anastomózami, mohou nést důležitou gen. informaci (rDNA u *O. ulmi*, toxiny, rezistence)
- některé jsou virového původu



Research review

Inheritance and recombination of mitochondrial genomes in plants, fungi and animals

Author for correspondence:

Camille M. Barr

Tel +1 434 982 5218

Fax: +1 434 982 5626

Email: cbarr@virginia.edu

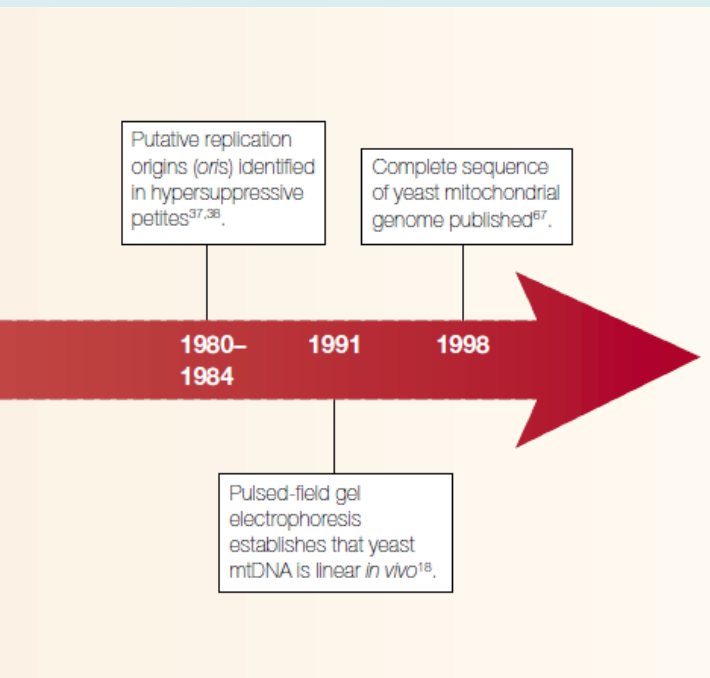
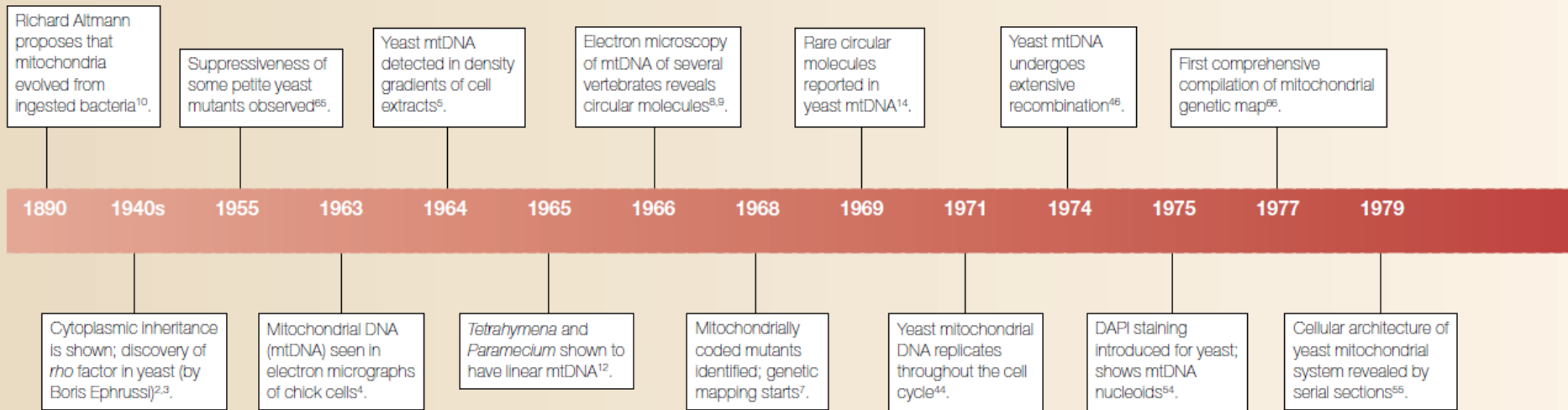
Received: 22 February 2005

Accepted: 3 May 2005

Camille M. Barr, Maurine Neiman and Douglas R. Taylor

Department of Biology, University of Virginia, Charlottesville, VA 22904, USA

Timeline | A history of yeast mitochondrial DNA



- *petite* mutanti u pivovarské kvasinky
- senescenční syndrom – porucha v mtDNA (COX gen)
- pouze fermentace, nikoliv dýchání



Stavba genomu hub - jaderná DNA

- malý genom s vysokou genovou densitou: 37-61 genů na 100 kbp
- **velikost nejčastěji mezi 20-40 Mbp** (0.01-0.1 pg v jedné buňce)
- cca 250 houbových anotovaných genomů (2014), 1250 genomových projektů (2016)

Table 2 Average genome size, and average number of coding genes and exons present in the different phyla/sub-phyla of the Kingdom Fungi

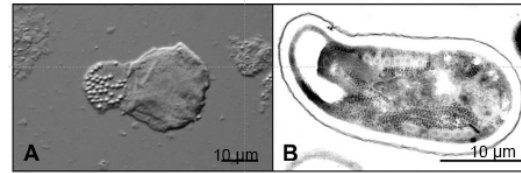
Fungal division	Average Genome Size (Mb)	Average No. Of Genes	Average No. Of Exons
Ascomycota	36.91	11129.45	2.58
Basidiomycota	46.48	15431.51	5.28
Oomycota	74.85	24173.33	2.23
Mucoromycotina	38.777	13306.85	4.25

Fungal genome size

Genome size is an important biodiversity character the study of which has both practical and theoretical uses in biology.

Variation in fungal genome sizes ranges:

from 2.5 Mb in *Encephalitozoon cuniculi*



Parasitological findings from the brain abscess.
Ditrich O et al. J. Clin. Microbiol. 2011;49:2769-2771

to 3050 Mb in *Neottiella* sp.



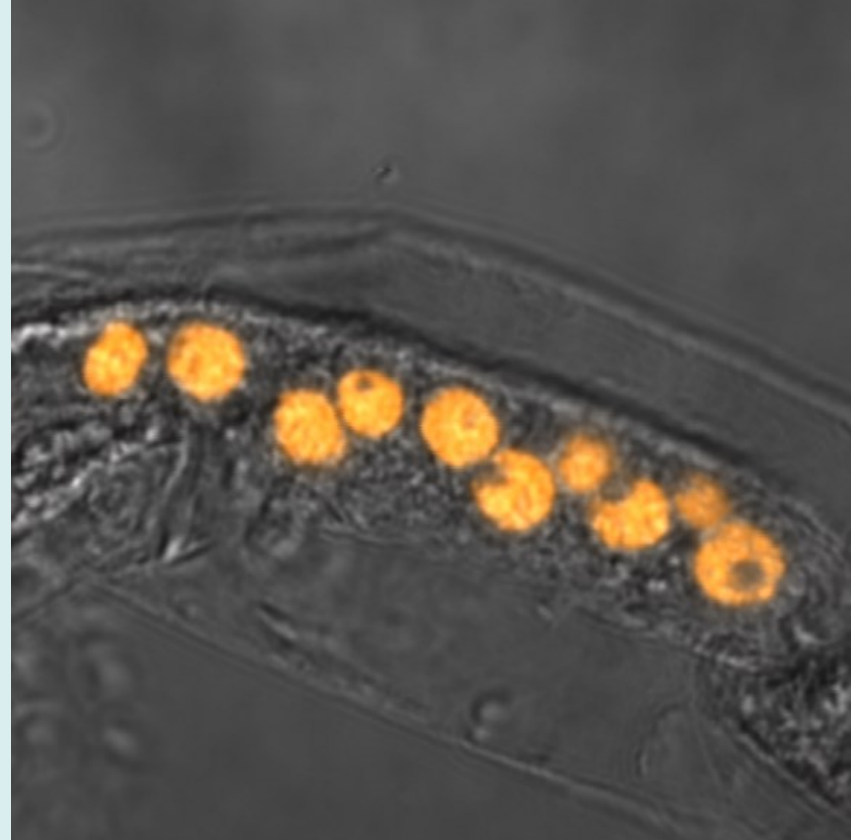
highly polyploid genera of ascomycetes (Kullman 2008)

Estimation of fungal genome size by flow cytometry

Bellis Kullman

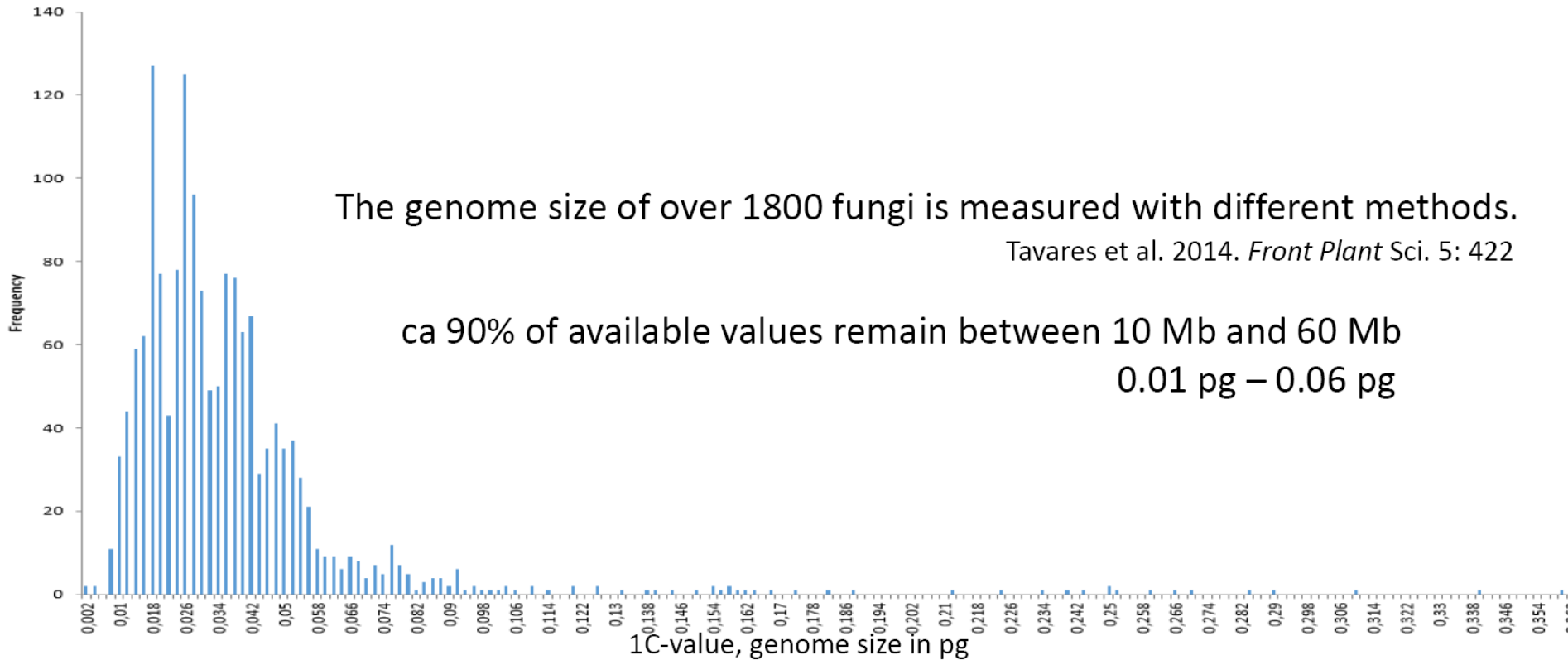
Department of Mycology
Institute of Agricultural and Environmental Sciences,
Estonian University of Life Sciences

Entomophthora muscae – 3.5 Gbp



Fungal Genome Size Database. Kullman et al., 2005

<http://www.zbi.ee/fungal-genomesize>



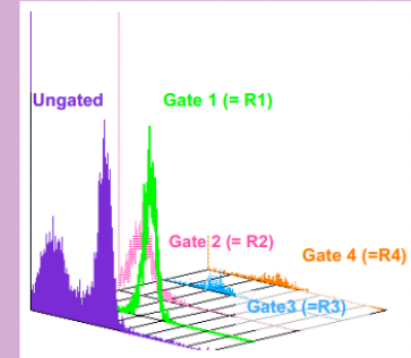
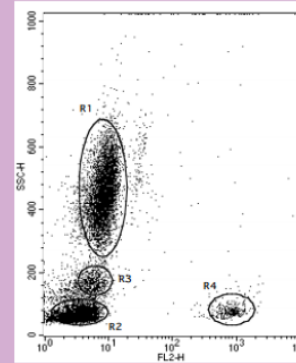
Estimation of fungal genome size by flow cytometry

Bellis Kullman

Department of Mycology
Institute of Agricultural and Environmental Sciences,
Estonian University of Life Sciences

Flow Cytometry (FC or FCM)

Flow Cytometry (cyto=cell) (metry=measurement)
Measurements performed
On individual cells
In a liquid stream.



<http://www.flowcytometri.dk/literature/Leslie-FCBasic.pdf>

The first fluorescence-based flow cytometer was developed in 1968 by Wolfgang Göhde from the University of Münster, and was commercialized one year later by Partec in Germany.

It soon became an essential method in medicine, veterinary science, food microbiology and water analysis.

However, it has remained unknown and unused in mycology.

C value – velikost haploidního (1n) genomu

CV - coefficient of variation - udává chybovost měření (u hub až 10%)

Estimation of fungal genome size by
flow cytometry

Bellis Kullman

Department of Mycology
Institute of Agricultural and Environmental Sciences,
Estonian University of Life Sciences

Application of Flow Cytometry for Genome Size Determination in *Geosmithia* Fungi: A Comparison of Methods

Tereza Veselská,^{1,2*} Jan Svoboda,² Zdenka Růžicková,² Miroslav Kolarík^{1,2}

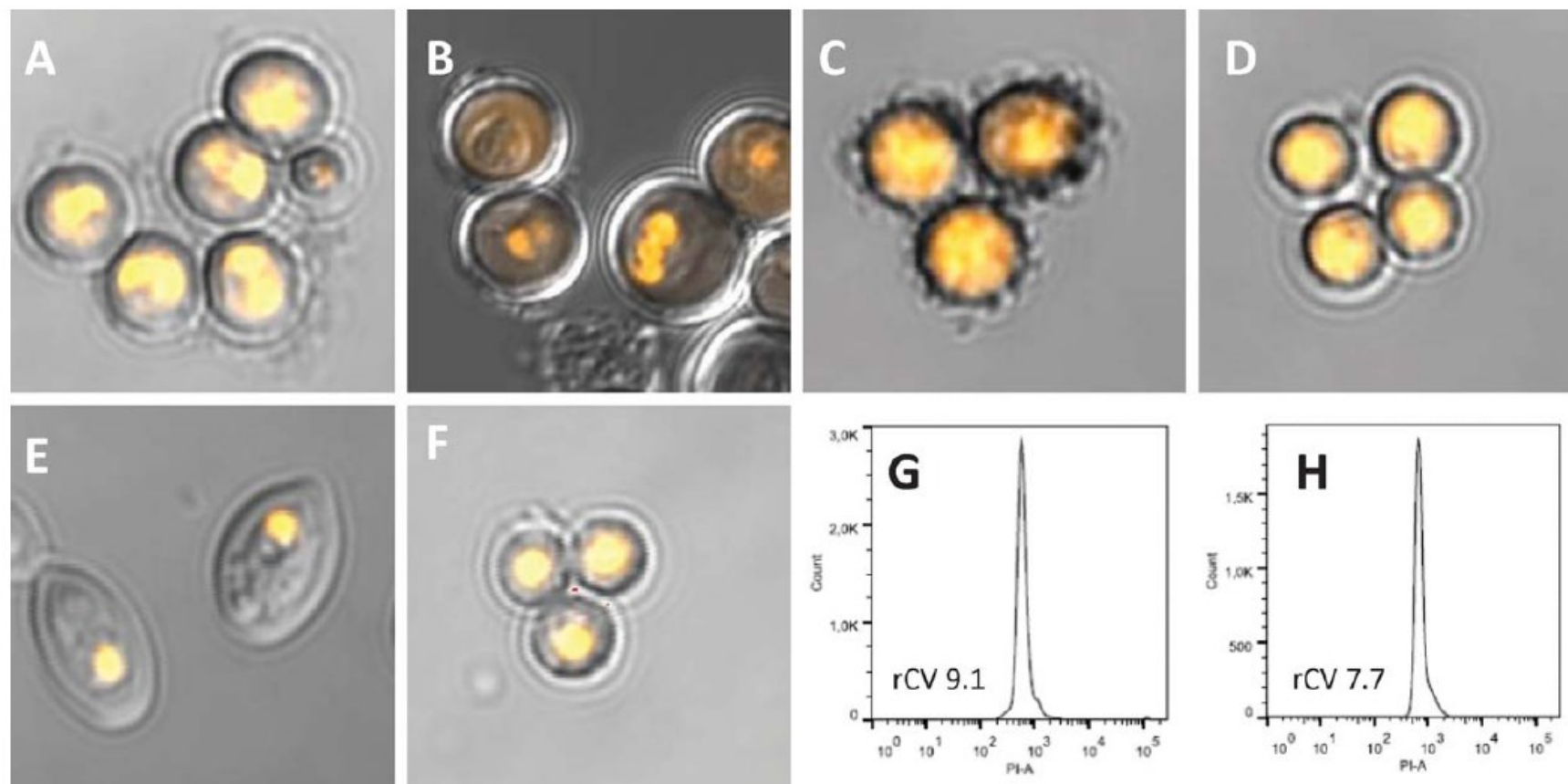


Figure 3. PI staining of standards. **A:** *A. niger* NRRL350, **(B)** *A. niger* CBS513.88, **(C)** *A. aculeatus* CBS172.66, **(D)** *A. fumigatus* Af293, **(E)** *S. cerevisiae* BY4743 α , **(F)** *A. fumigatus* CEA10 (our method), **(G)** FCM histogram of PI-stained *S. cerevisiae* BY4743 α , **(H)** FCM histogram of PI-stained *A. fumigatus* CEA10 (our method). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com]

C value paradox



Application of flow cytometry for exploring the evolution of *Geosmithia* fungi living in association with bark beetles: the role of conidial DNA content

Tereza VESELSKÁ^{a,b}, Miroslav KOLÁŘÍK^{a,b,*}

^aDepartment of Botany, Faculty of Science, Charles University, Benátská 2, 128 01, Czech Republic

^bInstitute of Microbiology of the ASCR, v.v.i., Václavská 108/3, 142 20 Praha 4, Czech Republic

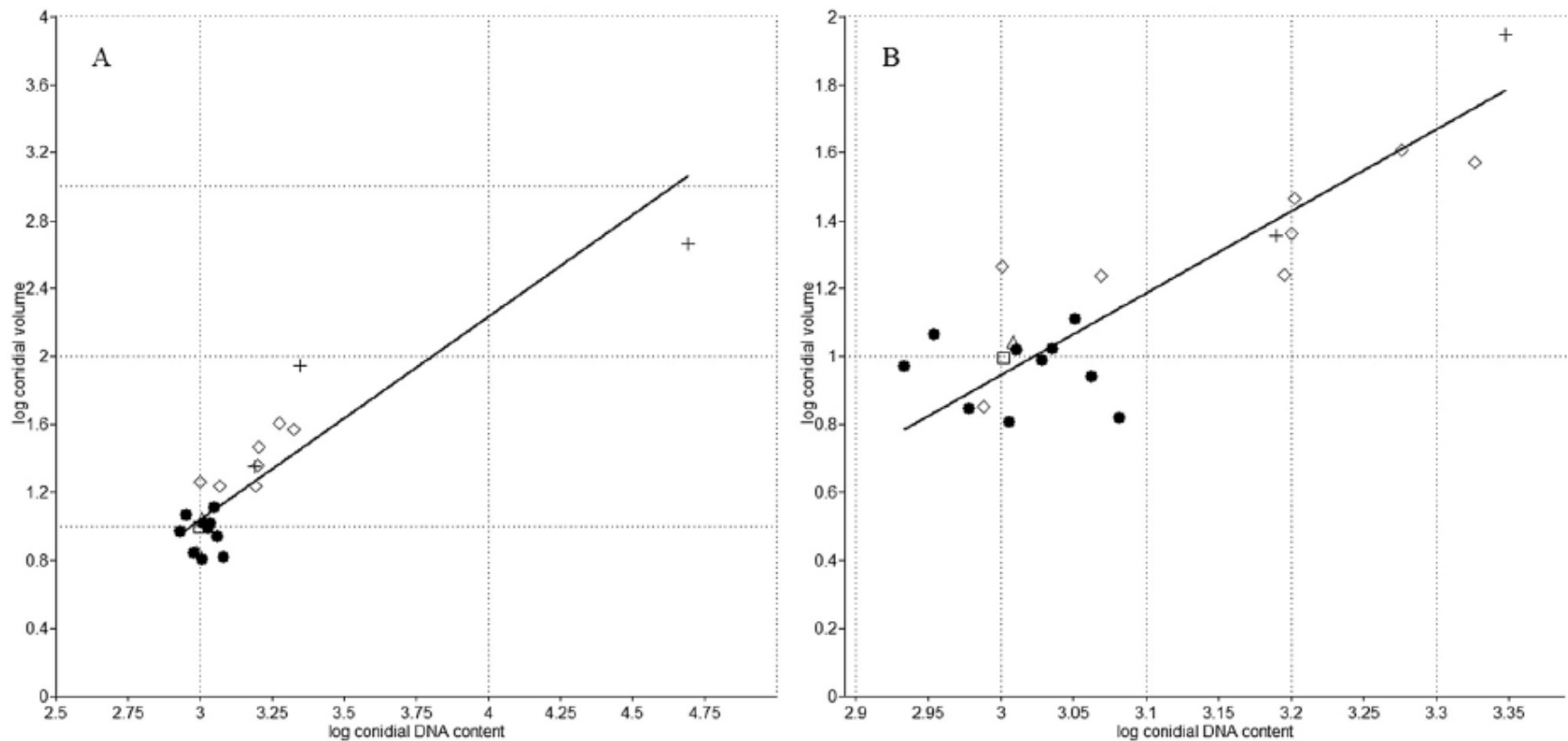


Fig 4 – Conidial DNA content and volume regression. (A) All species are plotted ($R^2 = 0.80$, $p < 0.0001$), (B) all species except *G. eupagioceri* are plotted ($R^2 = 0.76$, $p < 0.0001$). Cross – ambrosia species, diamond – specialists, circle – generalists, square – auxiliary ambrosia species, triangle – pathogen.

available at www.sciencedirect.comjournal homepage: www.elsevier.com/locate/funeco

Relationship between basidiospore size, shape and life history characteristics: a comparison of polypores

Håvard KAUSERUD^{a,*}, Jonathan E. COLMAN^b, Leif RYVARDEN^a

^aMicrobial Evolution Research Group, Department of Biology, University of Oslo, P.O. Box 1066 Blindern, N-0316 Oslo, Norway

^bDepartment of Ecology and Natural Resource Management, Norwegian University of Life Sciences, P.O. Box 5003, N-1432 Ås, Norway

ARTICLE INFO

Article history:

Received 24 July 2007

Received in revised form

19 October 2007

Accepted 6 December 2007

Published online 13 February 2008

Corresponding editor: Lynne Boddy

Keywords:

Basidiospore

Dispersal

Life history characteristics

Offspring size

Polypores

ABSTRACT

Although spore dispersal is the predominant mode of spread in fungi, the functional ecology of offspring (spore) size and shape has received little attention. We investigated the relationship between spore size (volume) and shape and various life history parameters among 303 European polypore species. In an analysis of variance, basidiocarp size, nutritional mode (parasitic vs. saprotrophic) and host (conifer vs. deciduous) accounted for a significant part of the variation in spore size. Species producing large basidiocarps also produce large spores, parasites produce larger spores than saprotrophs, and species colonizing deciduous trees generally produce larger spores than those colonizing conifers. There was a correlation between spore size and shape, with larger spores being more spherical. The most important factors accounting for variation in spore shape were rot type (white rot vs. brown rot) and nutritional mode, with white rot species and parasites having more spherical spores compared to brown rotters and saprotrophs.

© 2008 Elsevier Ltd and The British Mycological Society. All rights reserved.



Number of publications in Web of Science (from All Databases)

- Searched for „flow cytometry“ and **fungi** – ca 460 (of these, „genome size“ 17)
- Searched for „flow cytometry“ and **plant** – ca 6 000
- Searched for „flow cytometry“ and **animal** – ca 65 000

Estimation of fungal genome size by flow cytometry

Bellis Kullman

Department of Mycology
Institute of Agricultural and Environmental Sciences,
Estonian University of Life Sciences

Alternatively you can submit C-value data.

Estimation Methods:

- > GS = Complete genome Sequencing
- > PFGE = Pulsed Field Gel Electrophoresis
- > CHEF = Contour clamped Homogeneous Electric Field gel electrophoresis
- > FC = Flow Cytometry (Flow Microfluorometry, Flow Microfluorimetry, Fluorescence-Activated Cell Sorting, a cytophotometry technique). Technique using an instrument system for making, processing, and displaying one or more measurements on individual cells or nuclei obtained from a suspension. Cells are usually stained with one or more fluorescent dyes specific to cell components of interest, e.g., DNA, and fluorescence of each cell is measured as it rapidly transverse the excitation beam (laser or mercury arc lamp). Fluorescence provides a quantitative measure of various biochemical and biophysical properties of the cell, as well as a basis for cell sorting. Other measurable optical parameters include light absorption and light scattering, the latter being applicable to the measurement of cell size, shape, density, granularity, and stain uptake (a cytophotometry technique).
- > PI-FC = Flow Cytometry, stained with Propidium Iodide
- > DAPI-FC = Flow Cytometry, stained with DAPI
- > PC = Photometric Cytometry (classical cytophotometry technique, microscope with a photometer)
- > Fe-PC = microspectrophotometry, stained with Feulgen, measuring light absorption
- > FI-PC = microfluorometry (cytofluorometry), stained with Fluorochrome, measuring the light intensities
- > IC = Image Cytometry (a cytophotometry technique, an image analysis system which grabs images from the microscope via a digital camera, and calculates intensity or optical density from the grey values of pixels in the nucleus)
- > Fe-IC = Image Cytometry, stained with Feulgen, measuring light absorption (also called optical density, OD)
- > FI-IC = Image Cytometry, stained with Fluorochrome, measuring the light intensities
- > NS = Not Specified

Cell types:

- > FB = Fruit Body
- > PC = Pure Culture
- > S = Spores
- > C = Conidia
- > N = intact Nuclei
- > V = Various
- > NS = Not Specified

Standards and C-value for standards in Mb:

- k-mer frequency analysis z WGS dat
- WGS pro měření 1-n genomu
- FC pro změření ploidie

Estimation of fungal genome size by flow cytometry

Bellis Kullman

Department of Mycology
Institute of Agricultural and Environmental Sciences,
Estonian University of Life Sciences

Fungal Genomics Program

Exploration of fungal diversity



Genomic Encyclopedia of Fungi

Plant Feedstock Health

[Mycorrhizal Symbiosis](#)

[Plant Pathogenicity](#)

[Biocontrol](#)

Biorefinery

[Lignocellulose Degradation](#)

[Sugar Fermentation](#)

[Industrial Organisms](#)

Fungal Diversity

1000 Fungal Genomes project

Announcements

- March 26-28, 2013
JGI User Meeting, Walnut Creek, CA, USA
- March 12-17, 2013
Fungal Genetics Meeting in Asilomar, Pacific Grove, CA, USA
- January 12-16, 2013
Plant and Animal Genomics Meeting, San Diego, CA, USA

Recent Genome Releases

- November 15, 2012
[Aspergillus zonatus v1.0](#)
- November 12, 2012
[Aspergillus glaucus v1.0](#)
- November 4, 2012
[Hebeloma cylindrosporum h7 v2.0](#)
- November 2, 2012
[Didymella exigua CBS 183.55 v1.0](#)
- November 2, 2012
[Trichoderma longibrachiatum ATCC 18648 v1.0](#)
- November 2, 2012
[Trichoderma citrinoviride v1.0](#)
- October 10, 2012
[Aspergillus tubingensis v1.0](#)
- October 10, 2012
[Aspergillus wentii v1.0](#)
- October 5, 2012
[Paxillus rubicundulus Ve08.2h10 v1.0](#)
- October 4, 2012
[Amanita muscaria Koide v1.0](#)
- October 4, 2012
[Zasmidium cellare ATCC 36951 v1.0](#)
- October 4, 2012
[Apiospora montagnei NRRL 25634 v1.0](#)
- September 26, 2012
[Fomitopsis pinicola FP-58527 SS1 v3.0](#)
- September 14, 2012
[Hypoxylon sp. CI-4A v1.0](#)
- September 14, 2012
[Daldinia eschscholzii EC12 v1.0](#)
- September 14, 2012
[Hypoxylon sp. CO27-5 v1.0](#)
- September 14, 2012
[Hypoxylon sp. EC38 v1.0](#)
- August 31, 2012
[Amanita thiersii Skay4041 v1.0](#)
- August 24, 2012
[Tulasnella calospora AL13/4D v1.0](#)

- [Fungal Genomics Program](#)
- projekt 1000+ genomů

MycoCosm portal: gearing up for 1000 fungal genomes

Igor V. Grigoriev*, Roman Nikitin, Sajeet Haridas, Alan Kuo, Robin Ohm, Robert Otilar, Robert Riley, Asaf Salamov, Xueling Zhao, Frank Korzeniewski, Tatyana Smirnova, Henrik Nordberg, Inna Dubchak and Igor Shabalov

US Department of Energy Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

Received October 9, 2013; Revised October 31, 2013; Accepted November 1, 2013

ABSTRACT

MycoCosm is a fungal genomics portal (<http://jgi.doe.gov/fungi>), developed by the US Department of Energy Joint Genome Institute to support integration, analysis and dissemination of fungal genome sequences and other ‘omics’ data by providing interactive web-based tools. MycoCosm also promotes and facilitates user community participation through the nomination of new species of fungi for sequencing, and the annotation and analysis of resulting data. By efficiently filling gaps in the Fungal Tree of Life, MycoCosm will help address important problems associated with energy and the environment, taking advantage of growing fungal genomics resources.

The first JGI fungal portal was created for the white rot fungus *Phanerochaete chrysosporium* in 2004 (2). To enable comparative genome analysis within the same analytical framework, in 2008 JGI integrated data from six genomes of Dothideomycete plant pathogens (three sequenced by JGI and three others provided by the Dothideomycetes Genome Consortium). This served as a prototype for MycoCosm, the first official version of which was released in March 2010 in response to a call for a pan-fungal genomic resource from the community of fungal biologists brought together by the Burroughs Wellcome Fund (Alexandria, VA). By 2011, MycoCosm included over 100 fungal genomes (3), and continues to grow. Here we provide an update on the latest developments of MycoCosm, which includes flexible tools for the analysis of exponentially growing genomics and other ‘omics’ data sets, and extends user community participation.

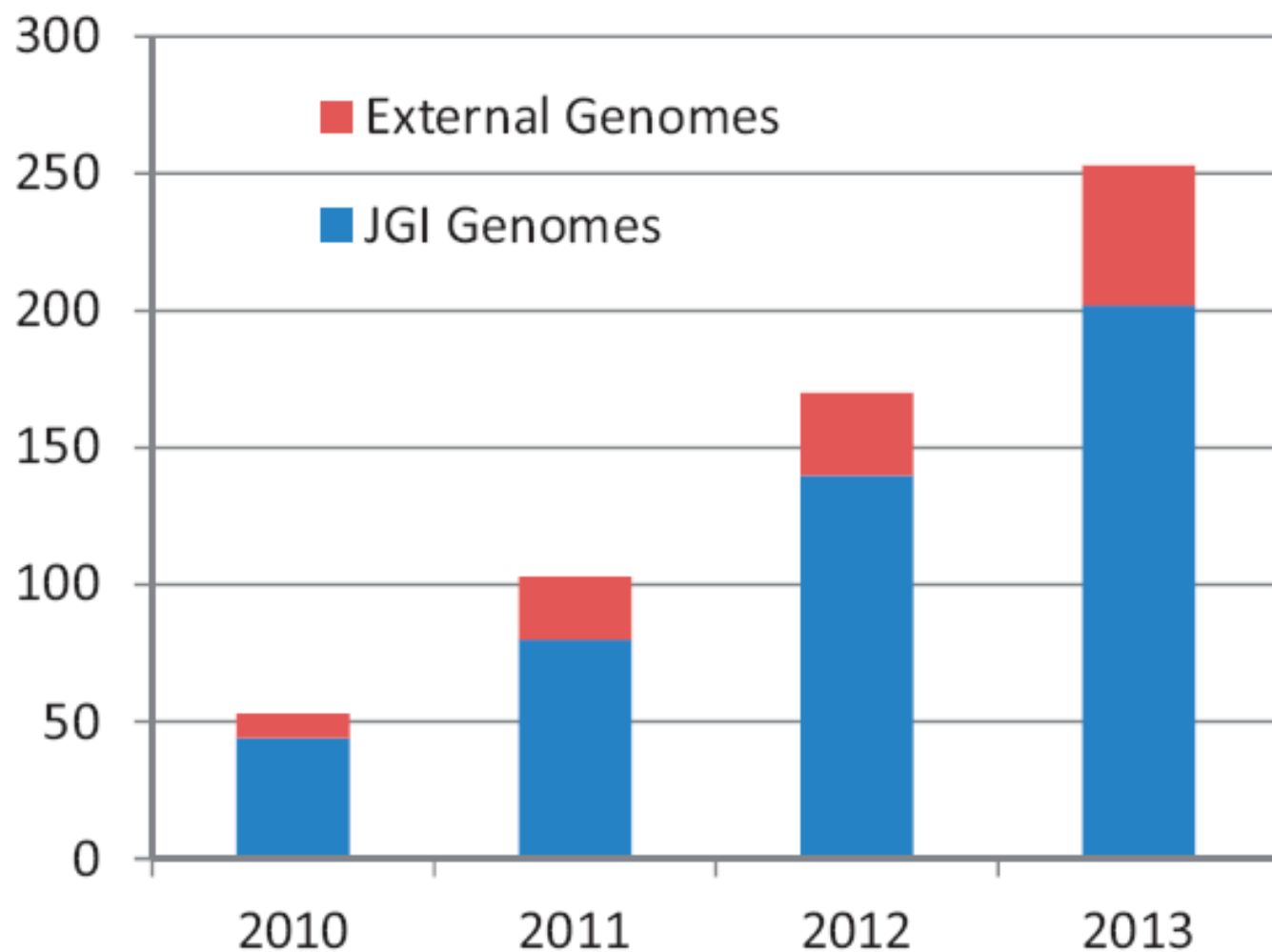
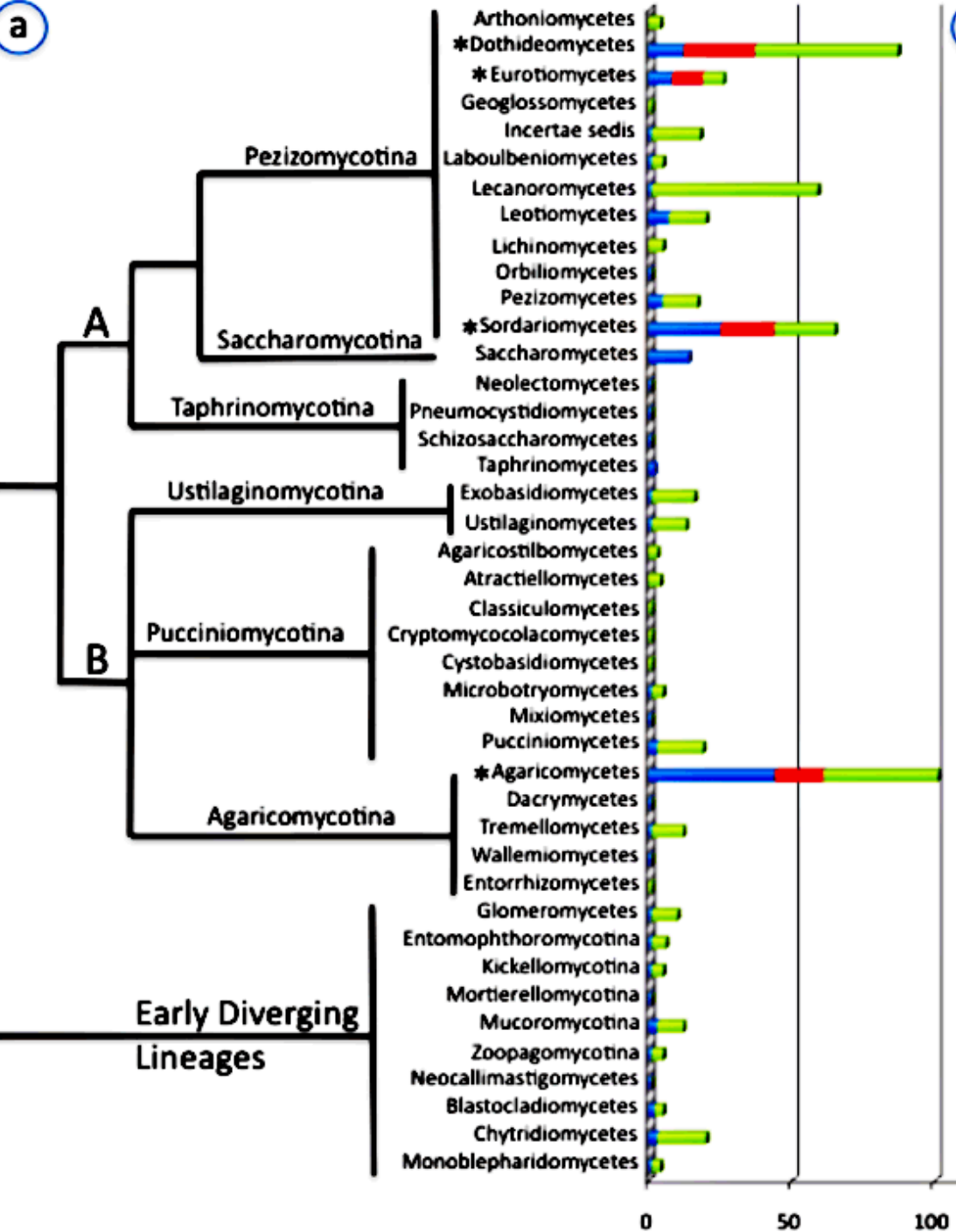


Figure 1. Growth of annotated genomes in MycoCosm. The genomes sequenced by JGI are shown in blue and those sequenced by others are shown in red.



- přehled studovaných genomů
- absolutní počet čeledí ve třídě
 - hotové genomy
 - plánované genomy
 - nesekvované

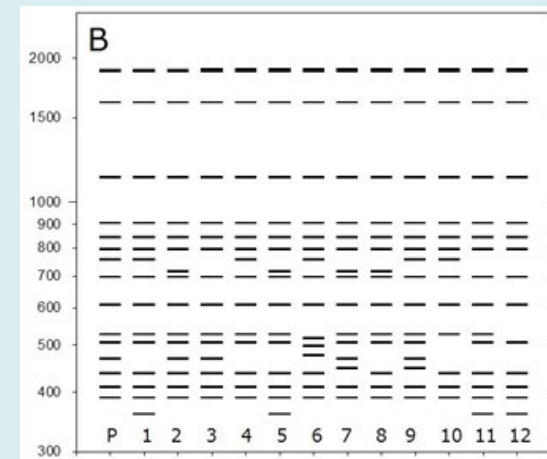
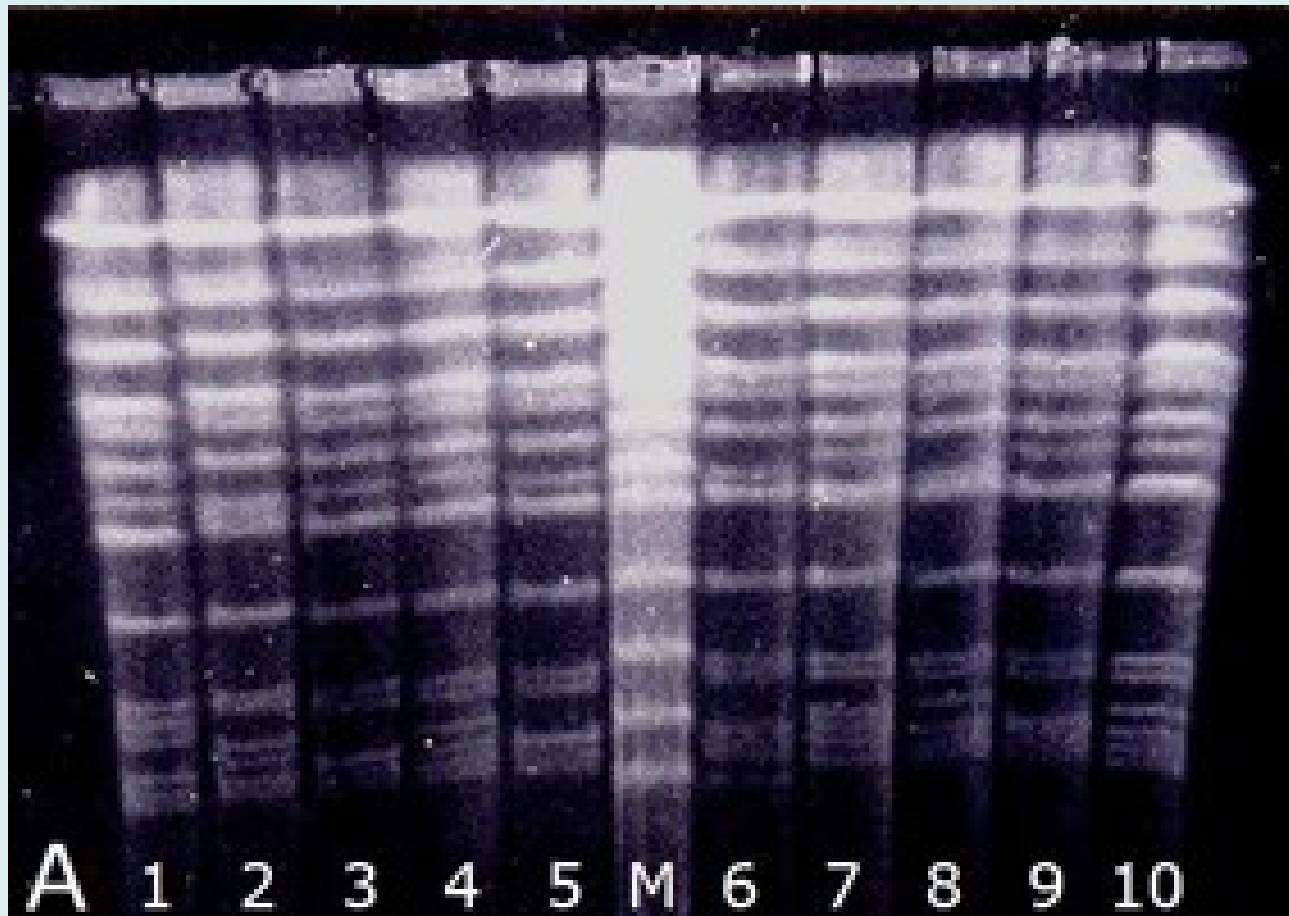
Stavba genomu hub - jaderná DNA

- **velmi variabilní genom, řada specifických genů** – zejména u patogenů a symbiontů (74% rozdíl *N. crassa* vs. *Magnaporthe grisea* = savci vs. ryby)
- **srovnávací genomika** (*Saccharomyces bayanus*, *S. mitake* a *S. paradoxus*, *Aspergillus* spp., *Penicillium* spp.)
 - velmi malá syntenie (shoda v umístění jednotlivých genů). U rodu *Aspergillus* je syntenie 68%
 - 8-14% druhově specifických genů v rámci rodu.
- *Laccaria bicolor* – pouze 70% genů má protějšek jinde
- geny samostatně, daleko méně v klastrech (vázané geny) - důležité sek. metabolity (PKS), geny pro patogenezí

Stavba genomu hub - jaderná DNA

- **3–22 chromozomů** (*Schizosaccharomyces pombe* – *Ustilago hordei*, *Batrachochytrium dendrobatidis*)
- **velmi malé** (detekce PFGE)
- **vnitrodruhová variabilita** v délce a počtu chromozómů (chromozómové zlomy, nehomologní překřížení, chybná segregace v meioze) – větší u nepohlavních hub
- **nadbytečné chromosomy** (minichromosomy, B-chromosomy *dispensable ch.*, *supernumerary ch.*), variabilní mezi jedinci jednoho druhu, obsahují repetice i unikátní sekvence, mohou být důležité v patogenezi

- velká schopnost adaptace díky chromozomálním poruchám (*in vitro* pokusy s *Candida albicans*, *S. cerevisiae* atd.)



Chromosome instability in industrial strains of *Saccharomyces cerevisiae* batch cultivated under laboratory conditions. Genet. Mol. Res. 6 (4): 1072-1084 (2007)

Stavba genomu hub - jaderná DNA

Příklad patogenní houby *Mycosphaerella graminicola* a *Fusarium* spp.

- Variabilita v počtu i struktuře chromosomů (CNP, *chromosome number polymorphism*) v rámci populace
- 8 z 21 chromozómů postradatelné (*dispensable*) chromosomy (nemají vliv na životaschopnost organismu) s velkým množstvím repetitivních sekvencí a s geny, které se pravděpodobně uplatňují při kolonizaci hostitele.
- Genom *Fusarium oxysporum* má o 4 chromosomy více než jeho blízce příbuzné druhy *F. graminearum* a *F. verticillioides*. Tyto chromosomy, nazývané liniově specifická (LS) genomová oblast (*lineage specific genome region*), zabírají $\frac{1}{4}$ jeho genomu a jsou opět bohaté na repetitivní sekvence a geny potřebné pro interakci s hostitelem
- Postradatelné a LS chromozomy jsou zodpovědné za rychlou adaptační schopnost těchto hub

Stavba genomu hub - jaderná DNA

- **úseky kódující** (proteiny, specifické RNA, regulační místa)
 - podléhají selekčnímu tlaku
 - uchová se jenom pozitivní nebo neutrální mutace
 - konzervativní
 - *singlecopy* vs. *multicopy* (jednotlivé kopie se nemusí shodovat)
- **nekódující** (introny, mezerníky, repetitivní DNA)
 - selekčně neutrální
 - každá mutace je uchována
 - méně konzervativní
 - pozice intronů a exonů je konzervativní - možnost použití univerzálních. primerů

Stavba genomu hub - jaderná DNA

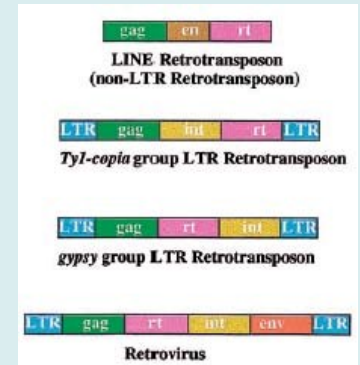
- velmi kondenzovaný genom typický malým zastoupením nekódujících oblastí
- velmi variabilní velikost a počet genů a intronů napříč taxonomickými i ekologickými skupinami.
 - *Ustilago maydis* (1n=20.5 Mb, 6784 genů)
 - *Puccinia graminis* (17 773 genů)
 - *Fusarium graminearum* (11 640 genů)
 - *Blumeria graminis* (5854 genů)
- Velmi variabilní počet intronů
 - *Ustilago maydis* – 0.46 intronů na gen (průměr)
 - *Coprinopsis cinerea* – 4.7 intronů na gen
 - *Phanerochaete chrysosporium* – 2.6 intronů na gen

Stavba genomu hub - jaderná DNA

- **repetitivní DNA – asco a basidiomycota typicky 5-30% celého genomu**
 - **mikrosatelity** (2-10 bp, = SSR - *single sequence repeats*)
 - seřazeny tandemově v jednom lokusu (150 bp)
 - velmi variabilní počet, často násobné rozdíly
 - velmi variabilní úseky mezi jednotlivými lokusy
 - vhodné pro populační studie
 - 0,08-0,67 % celkové DNA (u člověka 3 %, u rostlin daleko více)
 - typicky G-C či A-T bohaté
 - **minisatelity** (15- 30 bp) (např. M13 fág)
 - **transpozóny** (3% z celého genomu *S. cerevisiae*, 17% genomu *M. graminicola*, 0.02% u *Pseudozyma* spp., 30% u *Serpula lacrymans* or *Puccinia graminis*, 50% genomu savců, 80% genomu rostlin)

Stavba genomu hub - transpozóny

- **Class I retrotransposons:** copy-paste (přepis do RNA a následná reverzní transkripce)
- **Class II transposons:** cut-paste
- v rámci rep-DNA tvoří největší část (70% repetitivního genomu *M. graminicola*)



- velký evoluční význam, zdroj variability chromozómové délky

- umístěny zejména v telomeře a subtelomeře

- hlavní repetice (TTAGGG)_n, jako u prim. obratlovců
- *telomere repeat associated genes*
- klastry genů pro interakce s hostitelem, toxiny
- urychlená evoluce
- **zásadní význam v adaptaci na imunitní systém hostitele**

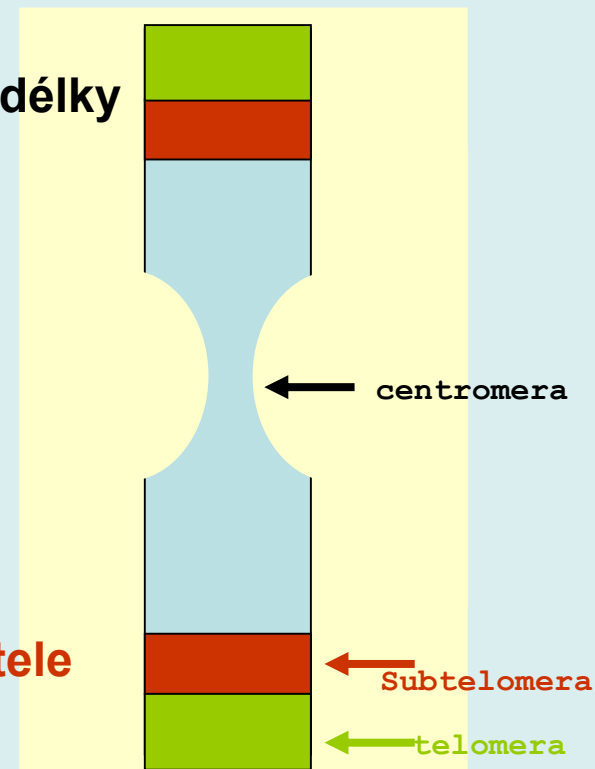


Table 1

Genomes of filamentous plant pathogens discussed in this paper

Species	Host	Genome size (Mbp)	Repeat content (%)	Reference
Oomycetes				
<i>Albugo candida</i>	Crucifers	45	17	[59**]
<i>Hyaloperonospora arabidopsidis</i>	Arabidopsis	100	43	[22]
<i>Phytophthora infestans</i>	Potato, tomato	240	74	[23]
<i>Phytophthora ramorum</i>	Broad range	65	28	[4]
<i>Phytophthora sojae</i>	Soybean	95	39	[4]
Fungi				
<i>Aciculosporium take</i>	Bamboo	60	57	[24,28]
<i>Blumeria graminis</i> f. sp. hordei	Barley	120	64	[14]
<i>Claviceps fusiformis</i>	Pearl millet	52	48	[24,28]
<i>Fusarium graminearum</i>	Wheat, barley	36	44	[19]
<i>Fusarium oxysporum</i>	Broad range	60	28	[19]
<i>Leptosphaeria maculans</i>	Crucifers	45	66	[15,16*]
<i>Magnaporthe oryzae</i>	Rice	40	52	[3]
<i>Melampsora larici-populina</i>	Poplar	101	45	[38]
<i>Melampsora lini</i>	Flax	220	45	[9]
<i>Melanopsichium pennsylvanicum</i>	Smartweed	20	3	[49*]
<i>Puccinia striiformis</i> f. sp. tritici	Wheat	110	49	[40]
<i>Zymoseptoria tritici</i>	Wheat	32–40	18	[27]

Available online at www.sciencedirect.com

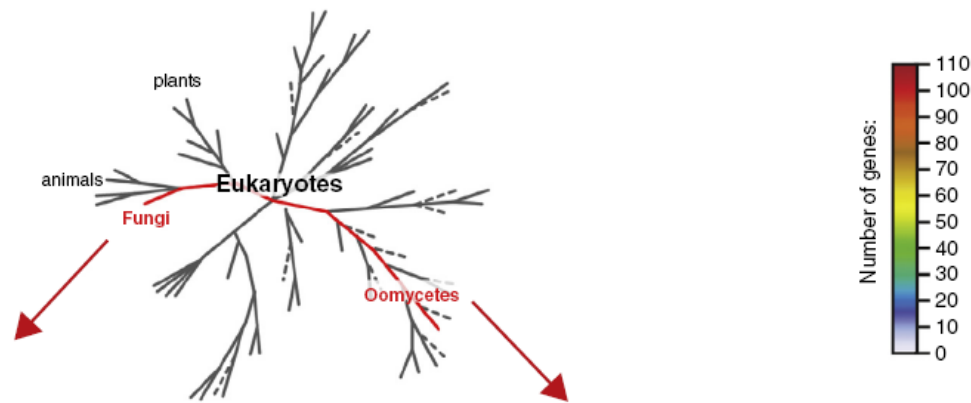
ScienceDirect

Current Opinion in
Genetics
& Development

The two-speed genomes of filamentous pathogens: waltz with plants

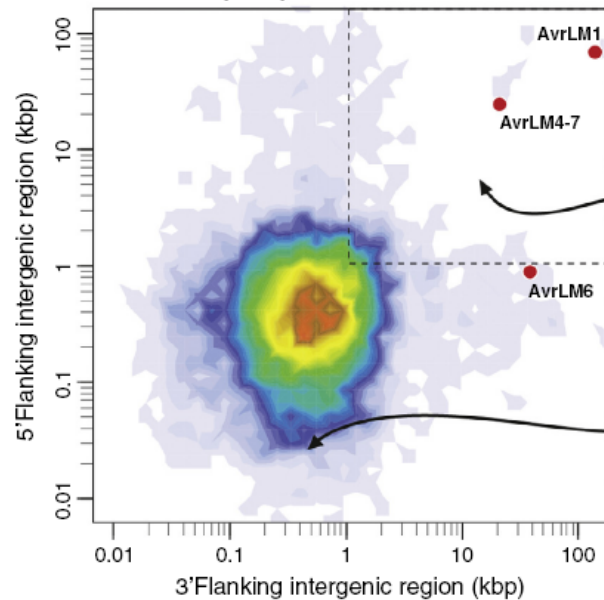
Suomeng Dong¹, Sylvain Raffaele^{2,3} and Sophien Kamoun⁴

Current Opinion in Genetics & Development 2015, 35:57–65

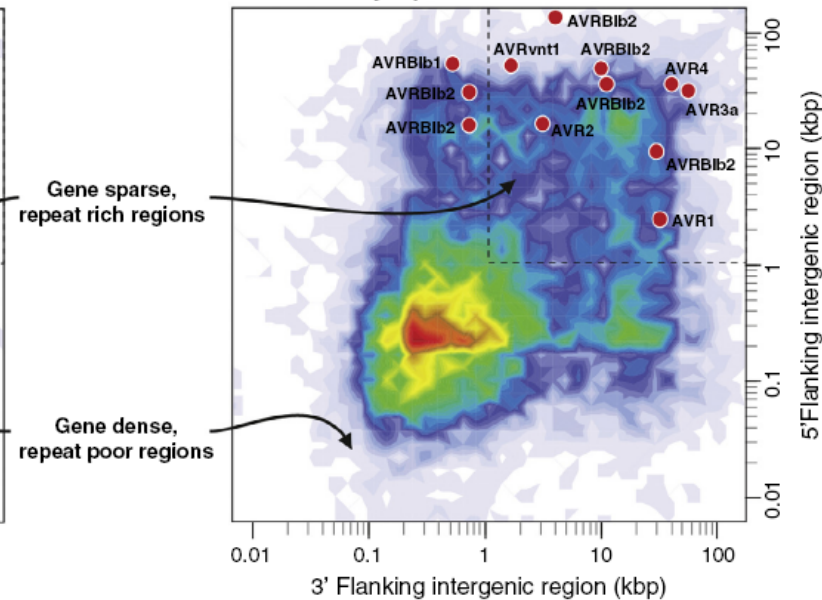


- pomalu a rychle mutující části genomu
- Oblasti s četnými repetitivy – velká mutační rychlost

Leptosphaeria maculans



Phytophthora infestans



Gene sparse,
repeat rich regions

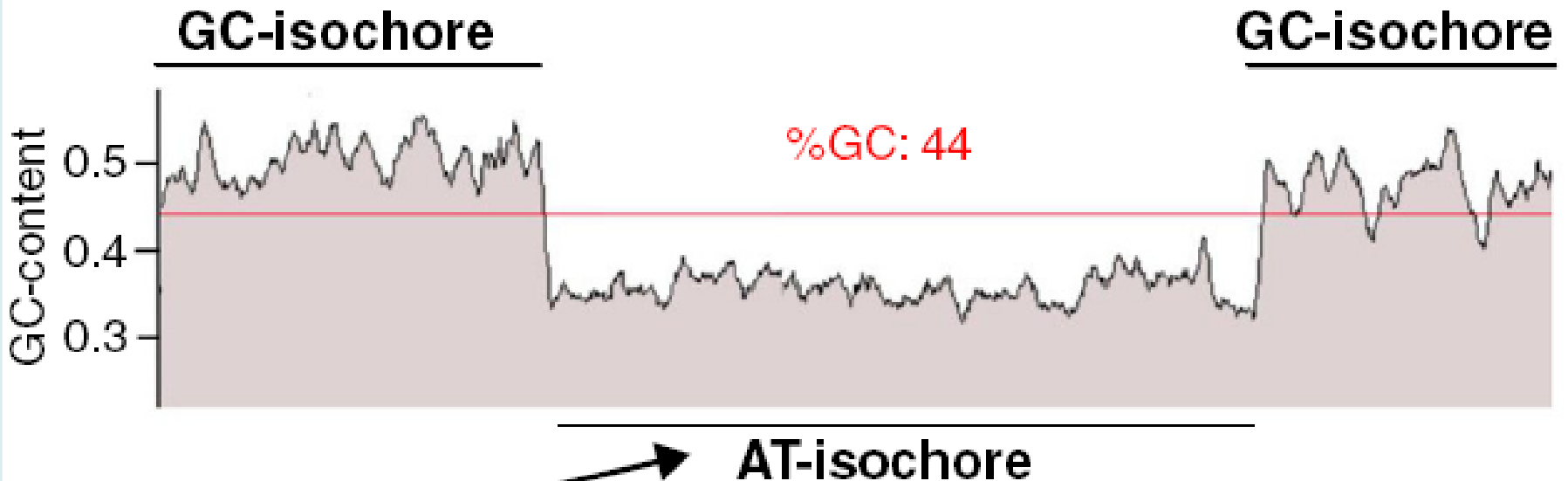
Gene dense,
repeat poor regions

Current Opinion in Genetics & Development

Convergence towards a similar genome architecture in phylogenetically unrelated plant pathogens. The flanking distance between neighboring genes provides a measurement of local gene density and is displayed as a color-coded heat map based on a whole genome analysis of the fungus *Leptosphaeria maculans* and the oomycete *Phytophthora infestans* [23]. In addition, the figure displays the distribution of Avr effector genes of *L. maculans* and *P. infestans* according to the length of their 5' and 3' flanking intergenic regions. Note how in both cases the Avr effector genes primarily occupy the gene sparse regions of the genome.

- rychle mutující části genomu jsou AT bohaté

Location in TE-rich regions

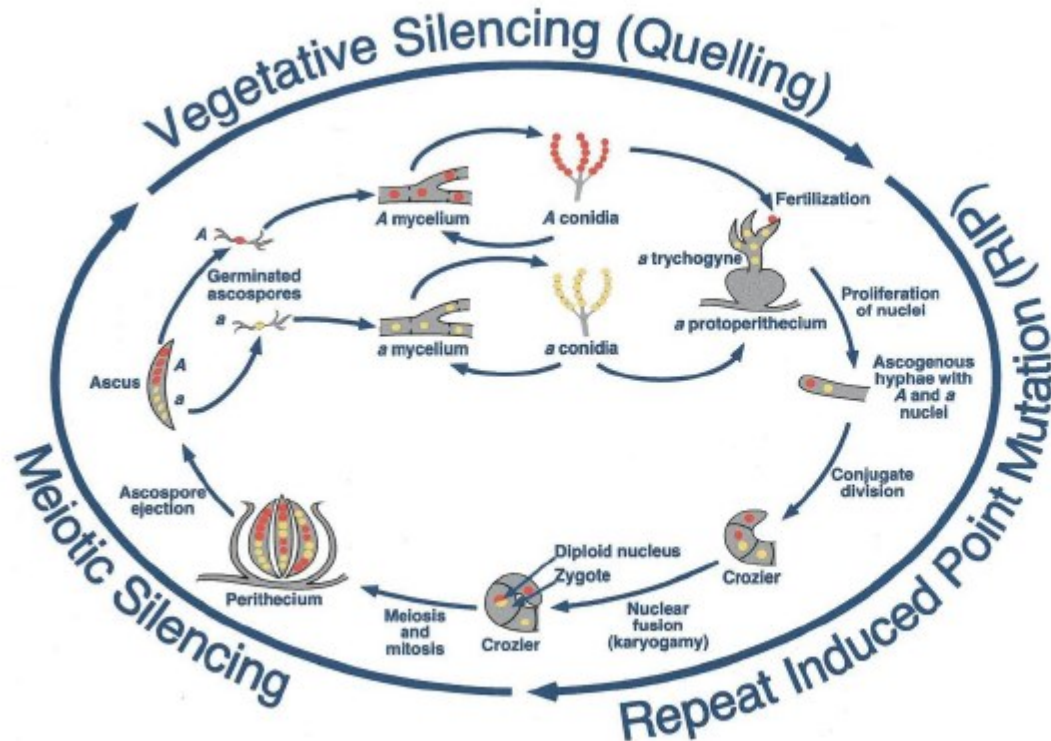


20% of the genes of TE-rich areas
are putative effector genes

Existují tři známé obrany proti transposabilním elementům.

Vegetative/Meiotic Silencing jsou založeny na post-transkripčním utišení genu (PTGS, post-transcriptional gene silencing)

RIP (repeat induced point mutation) je znám jen u hub a je založen na indukovaných mutacích a methylaci. Funguje jen v pohlavní fázi.



Obr. 2. Působení typů obrany proti repetitivním sekvencím během životního cyklu *N. crassa*.

RIP se uplatňuje během sexuální fáze v dikaryotickém myceliu. Po karyogamii nastupuje MSUD, meiotické utišení. Ve vegetativní fázi působí *quelling*. Blíže vysvětleno v textu. Schéma převzato z Borkovich et al. (2004).

Duální fenomén: sektorování u *Fusarium*, *Penicillium*, *Geosmithia* ...



Geosmithia pallida

Primární zdroj variability - mutace

- frekvence se liší dle pozice v genomu
- dříve studováno na úrovni funkčních genů, frekvence 10^{-5} – 10^{-7} /gen (např. na jeden milion dělení připadá na jeden gen jedna mutace)
- nověji pomocí celogenomických dat
- indely, celochromozomální změny, duplikace genomu (kvasinky), variace v počtu kopií (zejména mikrosatelity)
- velká schopnost mutovat je umocněná haploidní povahou většiny hub
- kolonie *Penicillium* 3.5 cm - cca 400 miliónu konidií
- uchovávání v metabolicky aktivním stavu

Precise estimates of mutation rate and spectrum in yeast

Yuan O. Zhu^{a,b,1}, Mark L. Siegal^c, David W. Hall^d, and Dmitri A. Petrov^{b,1}

- 145 diploidních jedinců (mezi každým jedincem cca 2000 dělení). Frekvence mutací u haploidů jsou vyšší cca o řád
- **867 substitučních změn**, = 5×10^{-10} na jednu pozici a generaci
- **26 indels pod 50 bp** (silně převažuje delece nad inzercí) = 5×10^{-12} na jednu pozici a generaci, více časté v nekódující oblasti, často vedle mikrosatelitů
- three copy number variants
- **Mikrosatelity**. 1000 lokusů typu 5-10 repetit, 280 lokusů 11-20 repetit, 7 lokusů 30-46 repetit, mikrosatelitových lokusů v kategorii 4-14 bp. Frekvence mutací 10^{-7} - 10^{-5} dle délky repetice
- **31 změn v počtu chromozomů** (29 x duplikace, 2x ztráta) = 9.7×10^{-5} mutací na jednu generaci. Mutanti se ztrátou chromozomu neměli schopnost sporulovat. Čím menší chromozom, tím větší šance ho ztratit.
- frekvence dominantních letálních mutací 2×10^{-5} na jeden diploidní genom a generaci

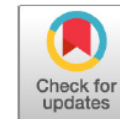
Precise estimates of mutation rate and spectrum in yeast

Yuan O. Zhu^{a,b,1}, Mark L. Siegal^c, David W. Hall^d, and Dmitri A. Petrov^{b,1}

^aDepartment of Genetics, Stanford University, Stanford, CA 94305-5120; ^bDepartment of Biology, Stanford University, Stanford, CA 94305-5020; ^cGenomics and Systems Biology, Department of Biology, New York University, New York, NY 10003; and ^dDepartment of Genetics, University of Georgia, Athens, GA 30602-7223

Edited by Michael Lynch, Indiana University, Bloomington, IN, and approved April 17, 2014 (received for review December 12, 2013)

Mutation is the ultimate source of genetic variation. The most direct and unbiased method of studying spontaneous mutations is via mutation accumulation (MA) lines. Until recently, MA experiments were limited by the cost of sequencing and thus provided us with small numbers of mutational events and therefore imprecise estimates of rates and patterns of mutation. We used whole-genome sequencing to identify nearly 1,000 spontaneous mutation events accumulated over ~311,000 generations in 145 diploid MA lines of the budding yeast *Saccharomyces cerevisiae*. MA experiments are usually assumed to have negligible levels of selection, but even mild selection will remove strongly deleterious events. We take advantage of such patterns of selection and show that mutation classes such as indels and aneuploidies (especially monosomies) are proportionately much more likely to contribute mutations of large effect. We also provide conservative estimates of indel, aneuploidy, environment-dependent dominant lethal, and recessive lethal mutation rates. To our knowledge, for the first time in yeast MA data, we identified a sufficiently large number of single-nucleotide mutations to measure context-dependent mutation rates and were able to (i) confirm strong AT bias of mutation in yeast driven by high rate of mutations from C/G to T/A and (ii) detect a higher rate of mutation at C/G nucleotides in two specific contexts consistent with cytosine methylation in *S. cerevisiae*.



Fungus Causing White-Nose Syndrome in Bats Accumulates Genetic Variability in North America with No Sign of Recombination

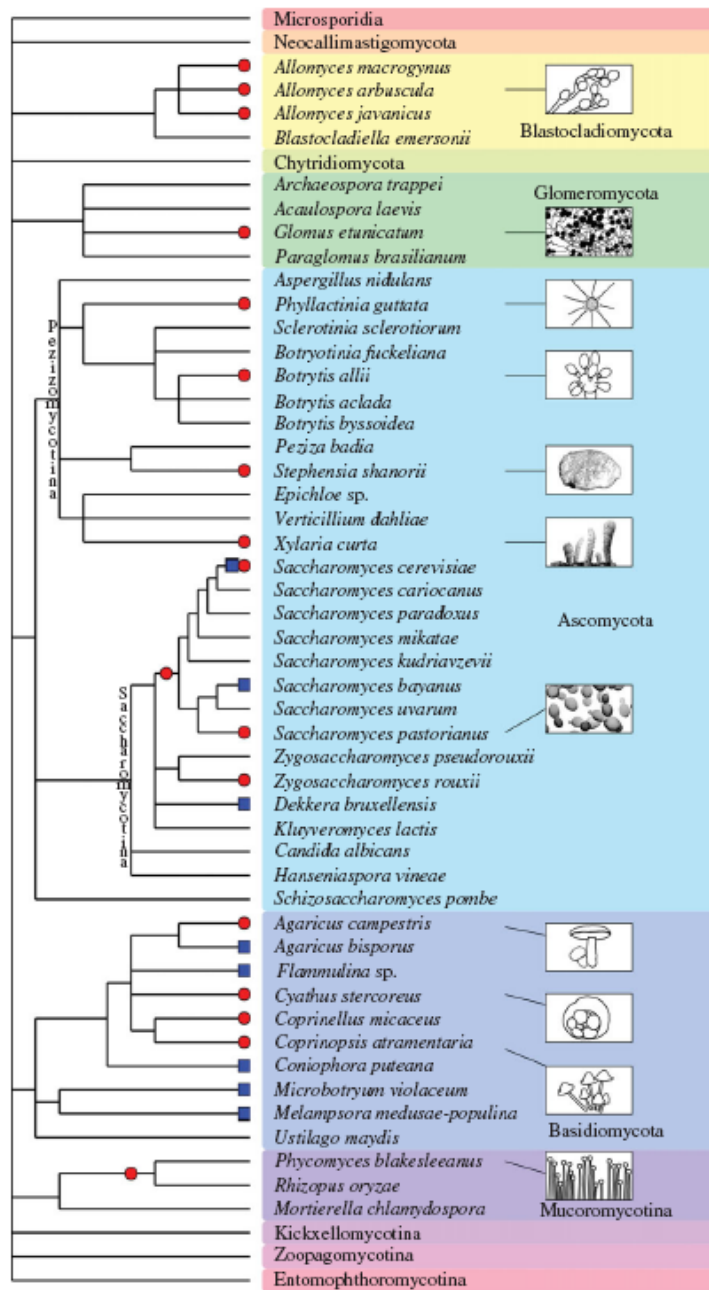
Jigar Trivedi,^a Josianne Lachapelle,^a Karen J. Vanderwolf,^b Vikram Misra,^c Craig K. R. Willis,^d John M. Ratcliffe,^a Rob W. Ness,^a James B. Anderson,^a Linda M. Kohn^a

Department of Biology, University of Toronto, Mississauga, Ontario, Canada^a; New Brunswick Museum, Saint John, New Brunswick, Canada^b; Western College of Veterinary Medicine, University of Saskatchewan, Saskatoon, Saskatchewan, Canada^c; Department of Biology and Centre for Forest Interdisciplinary Research (C-FIR), University of Winnipeg, Winnipeg, Manitoba, Canada^d

ABSTRACT Emerging fungal diseases of wildlife are on the rise worldwide, and the white-nose syndrome (WNS) epidemic in North American bats is a catastrophic example. The causal agent of WNS is a single clone of the fungus *Pseudogymnoascus*

Received 17 June 2017 Accepted 19 June 2017 Published 12 July 2017

- 22 kmenů (od společného předka se odlišili se nejpozději před 10 lety)
- 70 mutací

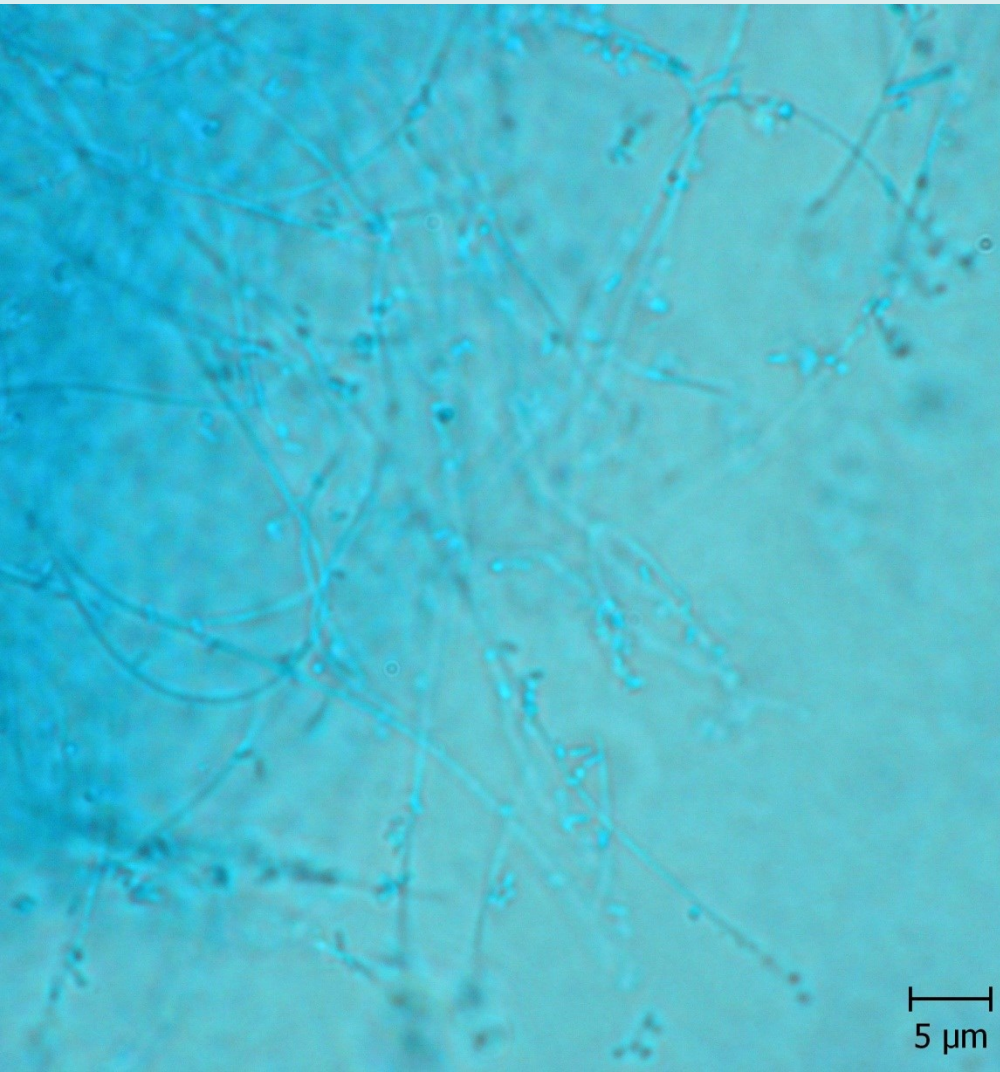


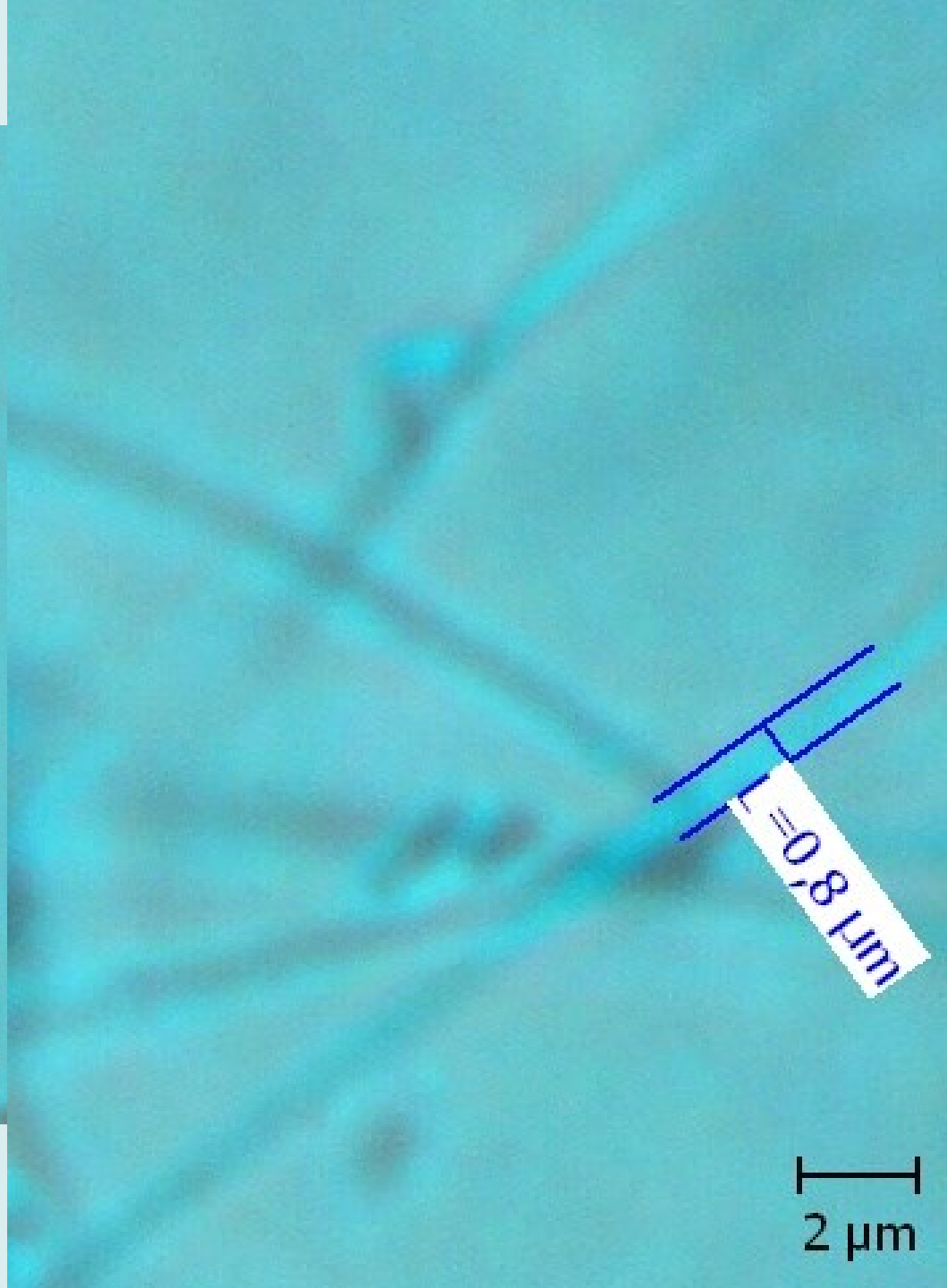
- Polyploidizace, patrně opomíjený mechanismus speciace

- hybridi
- polyploidi

Figure 3. Inferred polyploidy and reticulate evolution in fungi. Schematic phylogeny and classification of the fungi, based on Hibbet *et al.* [19]. Red circles indicate suspected polyploidy, blue squares indicate lineages with individuals having hybrid origin (reticulate evolution). Branch lengths are not proportional to genetic distances. Full references are available in electronic supplementary material, tables S1 and S2.

Zásady práce – vědět s čím pracuji !





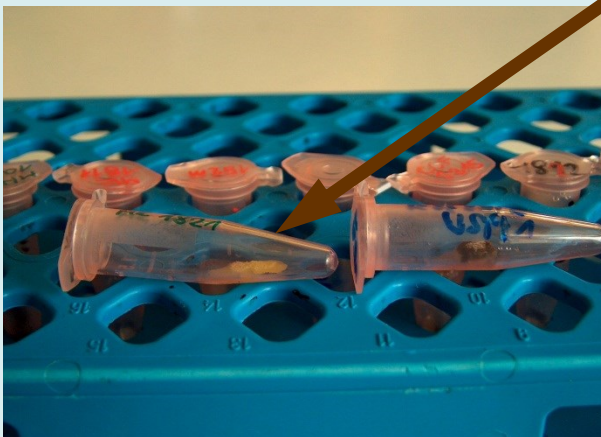
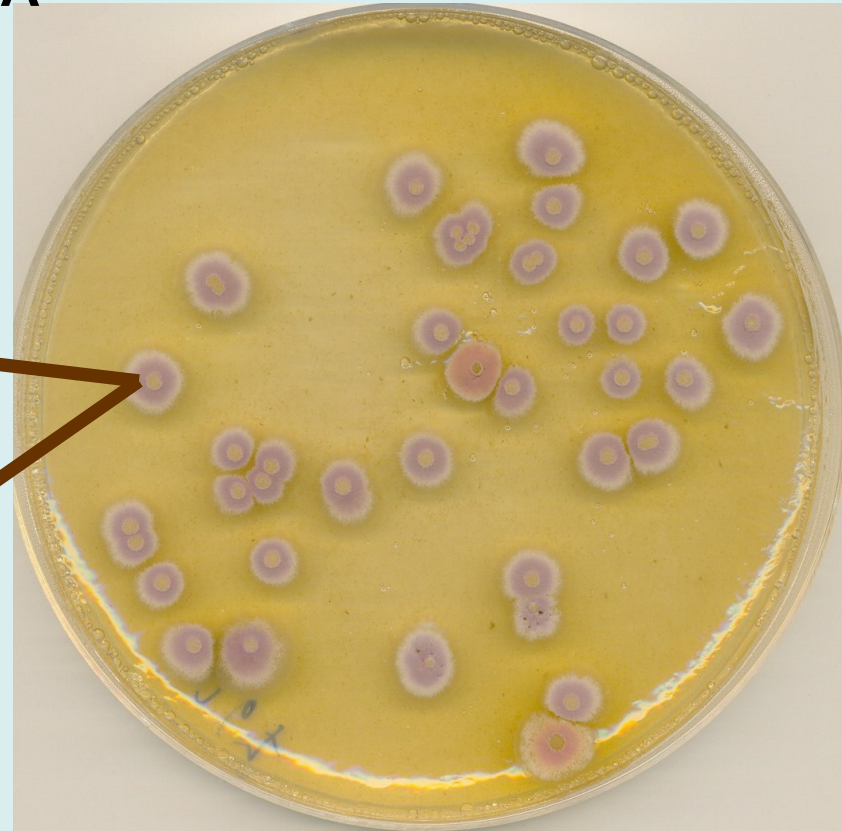
Vstupní materiál – přírodní materiál

- **čerstvý** – mladé plodnice, materiál nekontaminovaný okolím (trama z vnitřku plodnice), kultura
- **fixovaný** - 90% EtOH, zmrazený (ideální), kvalitně sušený (cca 10 let, pak problematické), různé fixáže (nejčastěji hexadecyltrimethylammonium bromid = CTAB, TRIZOL, skladovaný při pokojové teplotě)
- směsný materiál (více druhů hub)



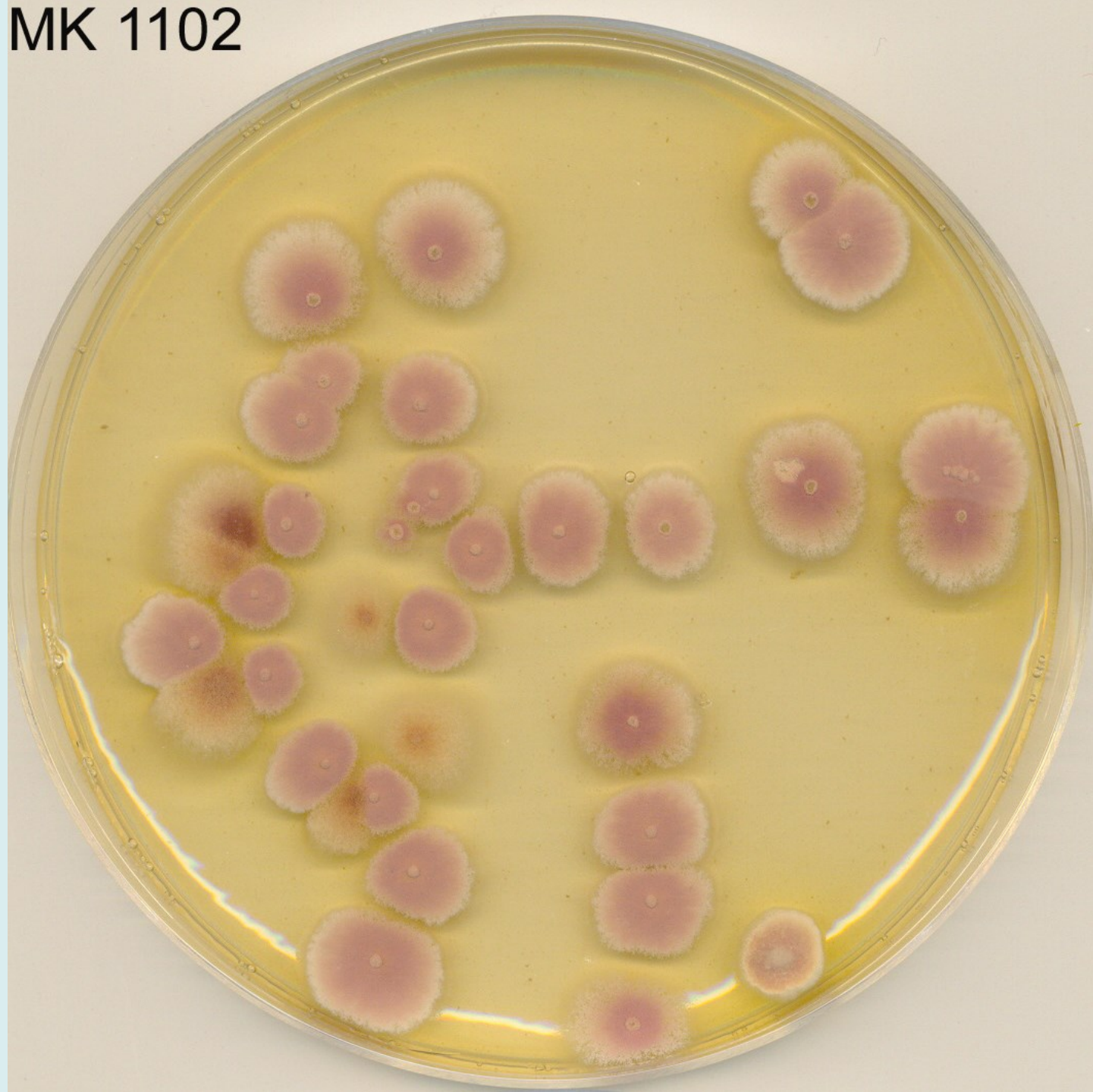
Vstupní materiál – čistá kultura

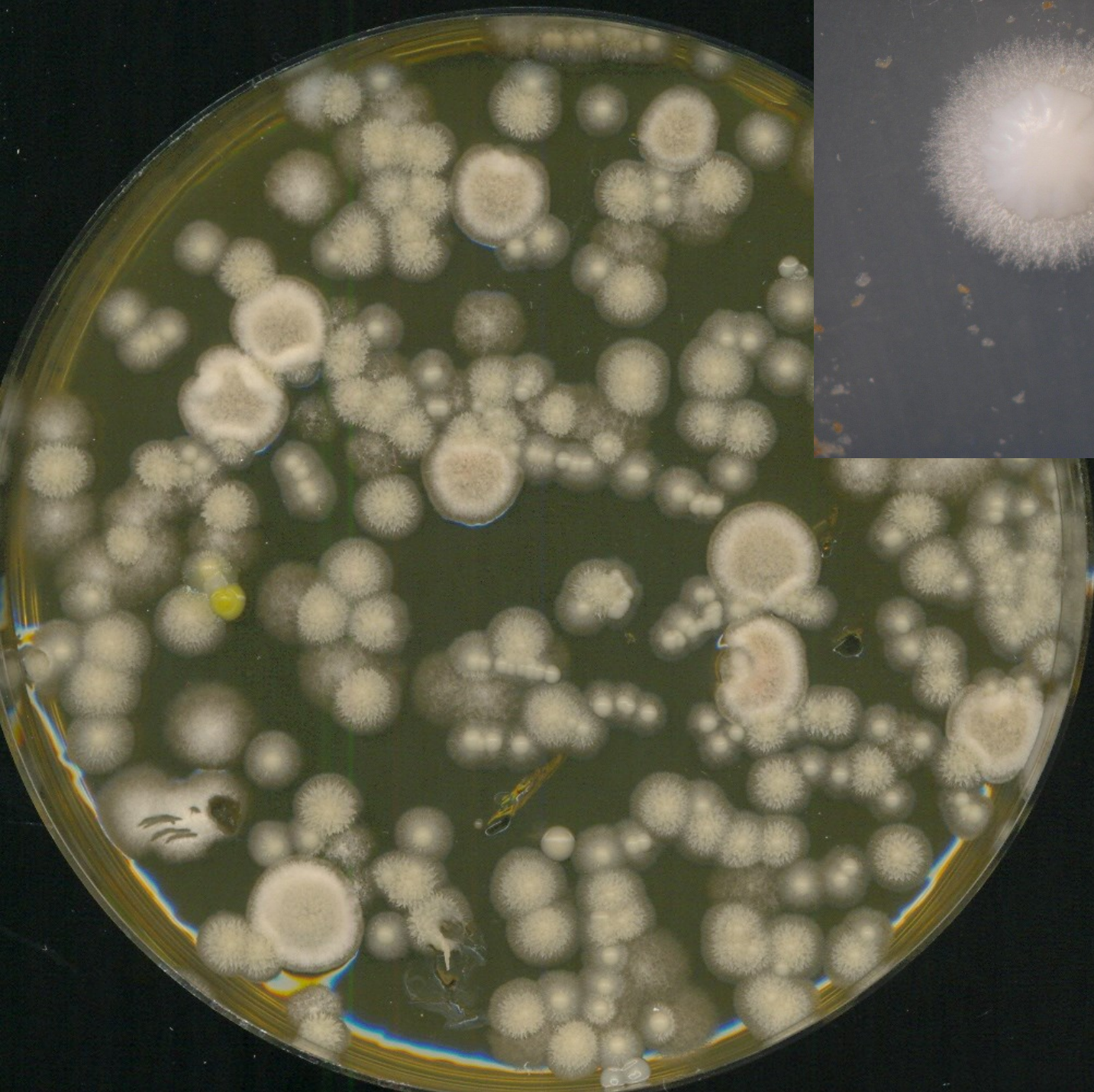
- monosporické izoláty
- část na další kultivace část na DNA





MK 1102





Základní techniky- izolace DNA

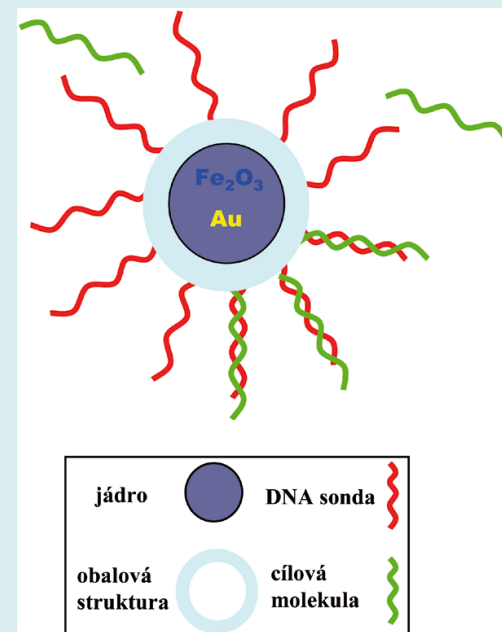
- **Garbage in, garbage out** – základem je kvalitní a čistý materiál. Čím mladší kultura (1-2 d) tím méně polysacharidů a sek. metabolitů.
- **rozdrcení buněčných stěn** – u hub problém (polysacharidy, polyfenoly)
 - drcení v tekutém dusíku, balotino, skleněný písek, proteinkynáza A (proteaza), glukonazy, chitinazy, lyticase (zymolase, bakteriální enzymy s proteasovou i endoglukanazovou aktivitou)
- **extrakční pufr** – detergenty naruší b.s. a zejména membrány, inaktivace enzymů (denaturace). Detergenty: SDS (dodecylsulfát sodný), guanidinium hydrochlorid, sarcosyl, NP-40 (Triton X100)
- **oddělení z nadmolekulárních struktur (odstranění proteinů a polysacharidů)**
 - fenol-chloroform (jeho zbytky inhybují PCR)
 - CTAB (cetrimehyl ammoniumbromid) – vysráží polysacharidy
 - PVP (polyvinylpyrrolidon), aceton
 - sorbenty vázající DNA (viz dále)
- **odstranění RNA (RNAáza)**
- DNA je rozpustná v polárních rozpouštědlech – zůstává ve vodné fázi
- **DNA se sráží pomocí alkoholů** (isopropanol, ethanol, PEG polyethylenglykol) – sraženina oddělena centrifugací. Nutno vyvázat vodu. Nejčastěji octan (sodný, amonný)

Základní techniky- izolace DNA

- **sedimentační metody** – s vysráženou DNA se usadí i zbytek nečistot. Výtěžky jsou velké
- **metody s využitím sorbentu DNA** (sklo, křemelina, silikagel, iontoměniče). Výtěžky menší, DNA čistější
 - **silikagel (SiO_2)**, nejčastější součást komerčních souprav. Silikagel v kolonce váže DNA v míře ovlivněné iontovou silou prostředí. Pro zvýšení iontové síly roztoku se používá guanidine hydrogen sulfát (dříve jodid sodný). Opakovaným promýváním guanidine hydrogen sulfátem se DNA zbaví nečistot. DNA se vymyje roztokem o malé iontové síle (voda, TRIS pufr). Kolonku lze recyklovat vymytím 0.1M HCl a vodou.
- **RESIN (pryskyřice)** - k vázání a následnému uvolňování požadovaného materiálu využívají iontové výměny (iontoměniče). Např. DEAE (diethylaminoethyl) v kitu od firmy Qiagen. Dále Chelex 100. K separaci DNA se použije kolonka, která naváže resin+DNA. Velké výtěžky a velká čistota. Někdy používáno na čištění již hotové DNA. Vhodné na odstranění PCR inhibitorů.

Základní techniky- izolace DNA

- **izolace pomocí magnetických částic.** Dražší metoda, ale je schopna izolovat čistou DNA z velmi komplexních substrátů (půda).
- 5 nm až 100 μm kuličky (maghemit $\gamma\text{-Fe}_2\text{O}_3$, magnetit Fe_3O_4 , Au)
- povrch částic je dále modifikován (SiO_2), aby mohl vázat DNA
- možnost kombinovat se hybridizační DNA sondou
- kuličky s navázanou DNA oddělíme magnetem.



Obr.7: Schéma magnetické částice

Základní techniky- izolace DNA

- rozdílné dle typu vstupního materiálu (izolace z půdy, dřeva...)
- výtěžek sedimentační metody cca 50 µg, z kolonek cca 0,5 µg.
- velmi rychlé metody založené na rozdrcení a vaření v pufru (DNA přežije vaření, mikrovlny) či drcení a rychlé extrakce v NaOH, Chelex 100 atd. Většinou nejde dlouhodobě uchovávat.
- možnost izolace z jedné spory (Glomeromycota)
- archivní DNA lze izolovat, ale záleží na způsobu ošetřování položek (formaldehyd degraduje DNA)
- ideální uchovávání biomasy je mražení. Případně usmrcení EtOH a sušení. Uchovávání DNA se děje v -20 °C či -80 °C. Další možnost je CTAB, TRIZOL roztok (+ řada dalších komerčních roztoků).

Základní techniky – elektroforéza DNA

- nositelem el. náboje u DNA jsou záporně nabitě fosfátové skupiny
- pohyblivost DNA je nepřímo úměrná logaritmu jejich velikosti (nikoliv sekvenci)
- nosič: - **agaróza** (DNA fragmenty o velikosti cca 200 bp - 50 kb),
DNA vizualizována ethidium bromidem
 - **polyakrilamid** (PAGE, velikost 10-1000 bp)
DNA vizualizována autoradiogramem, fluorescentně

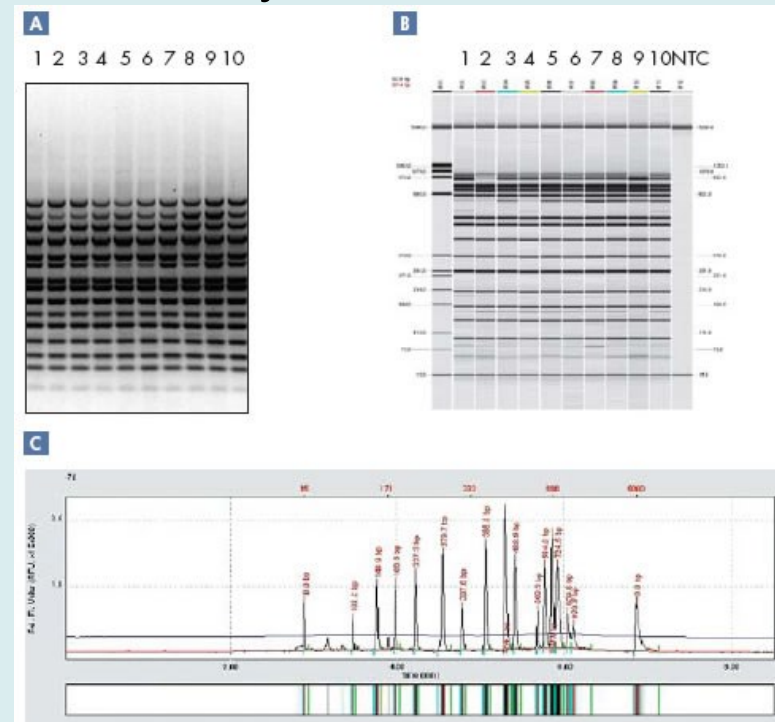
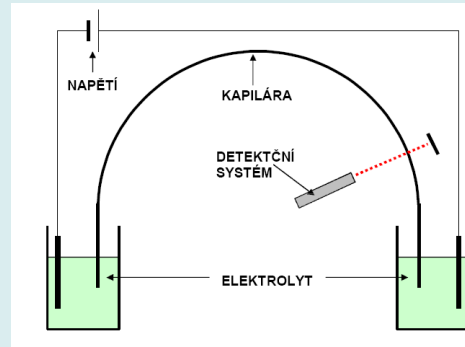
Základní techniky – elektroforéza DNA

- PFGE (Pulse Field Gel Electrophoresis), pro fragmenty > 50 kbp, slouží při karyotypování

problém: fragmenty o stejné hmotnosti nemusí mít stejnou sekvenci

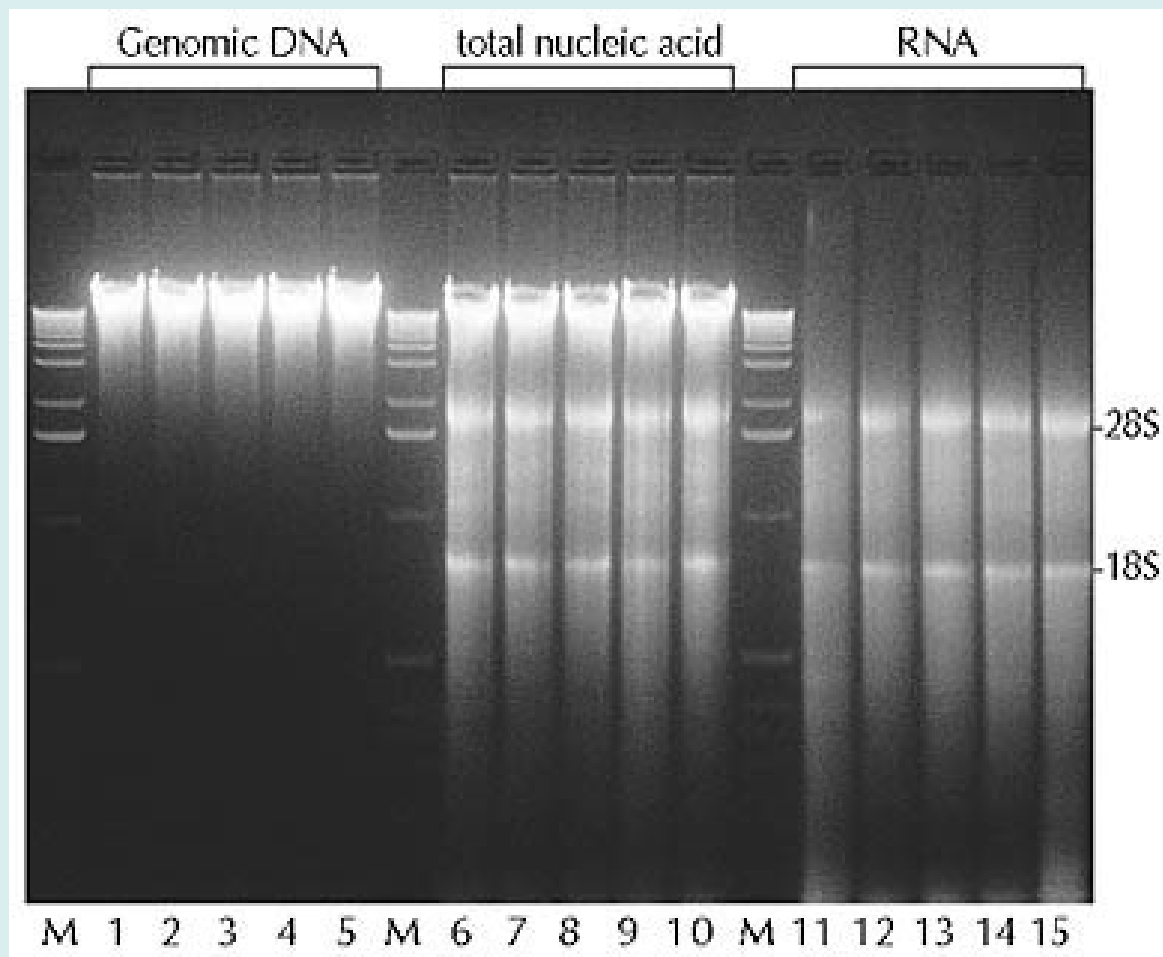
- DGGE (Denaturing Gradient Gel Elektroporesis), TGGE...
- kapilární elektroforéza (indikace provádí fotometricky v UV oblasti spektra)

ABI 3100 capillary systems



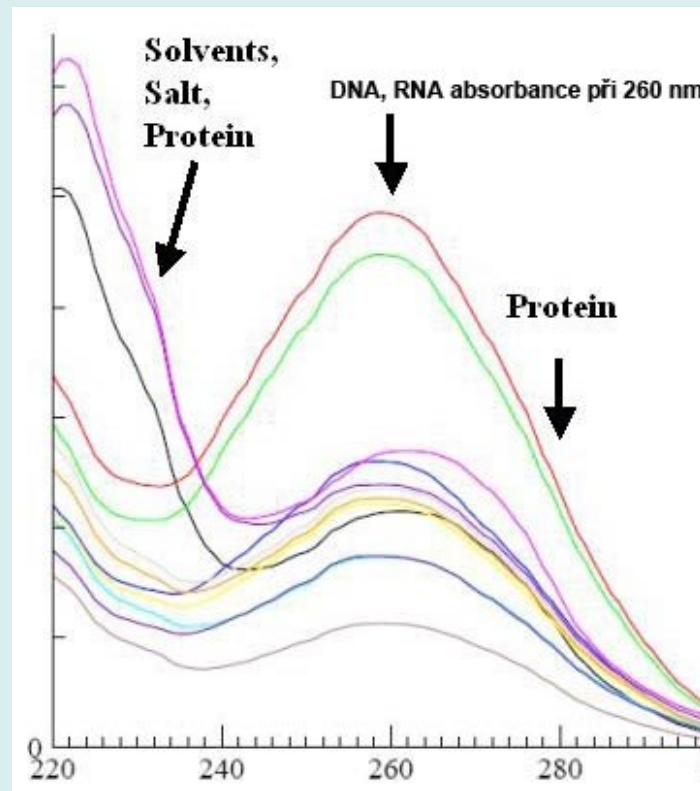
Základní techniky – zjištění koncentrace a kvality DNA

- Porovnáním svítivosti se standardem



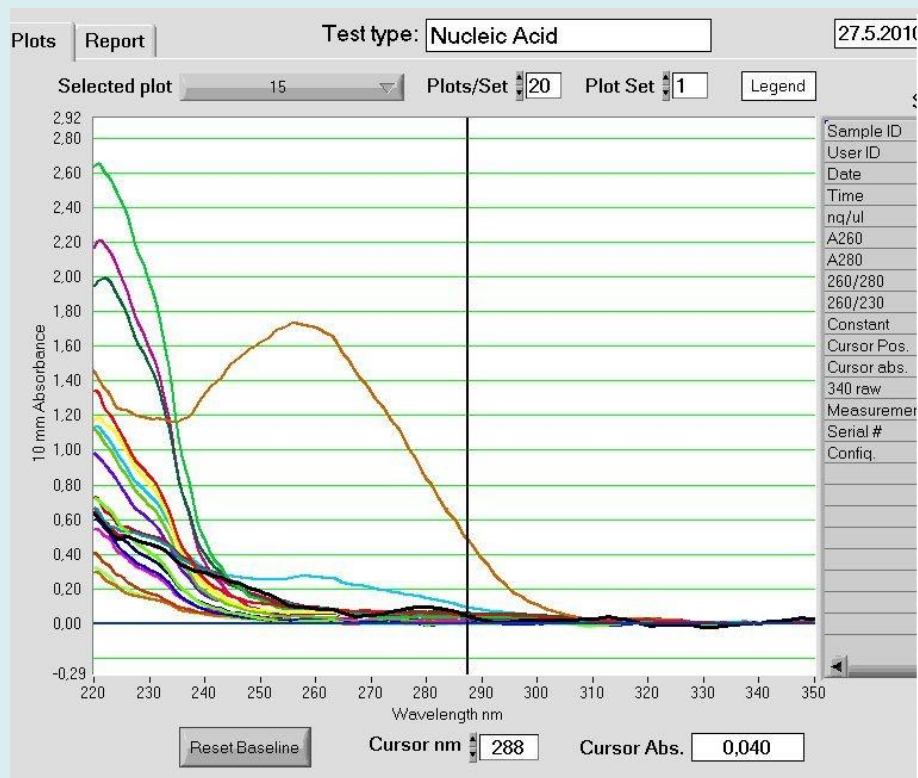
Základní techniky – zjištění koncentrace a kvality DNA

- spektrofotometricky. Firma Nanodrop. 0.5 μ l - 2.0 μ l vzorku.
- Měří koncentraci DNA, RNA (A260) a bílkovin (A280) a čistotu vzorku (poměr 260/280)



Základní techniky – zjištění koncentrace a kvality DNA

- poměr 260/280: **1.8** u čisté DNA
- poměr 260/230: **druhořadý údaj, 2-2,2** u čisté DNA

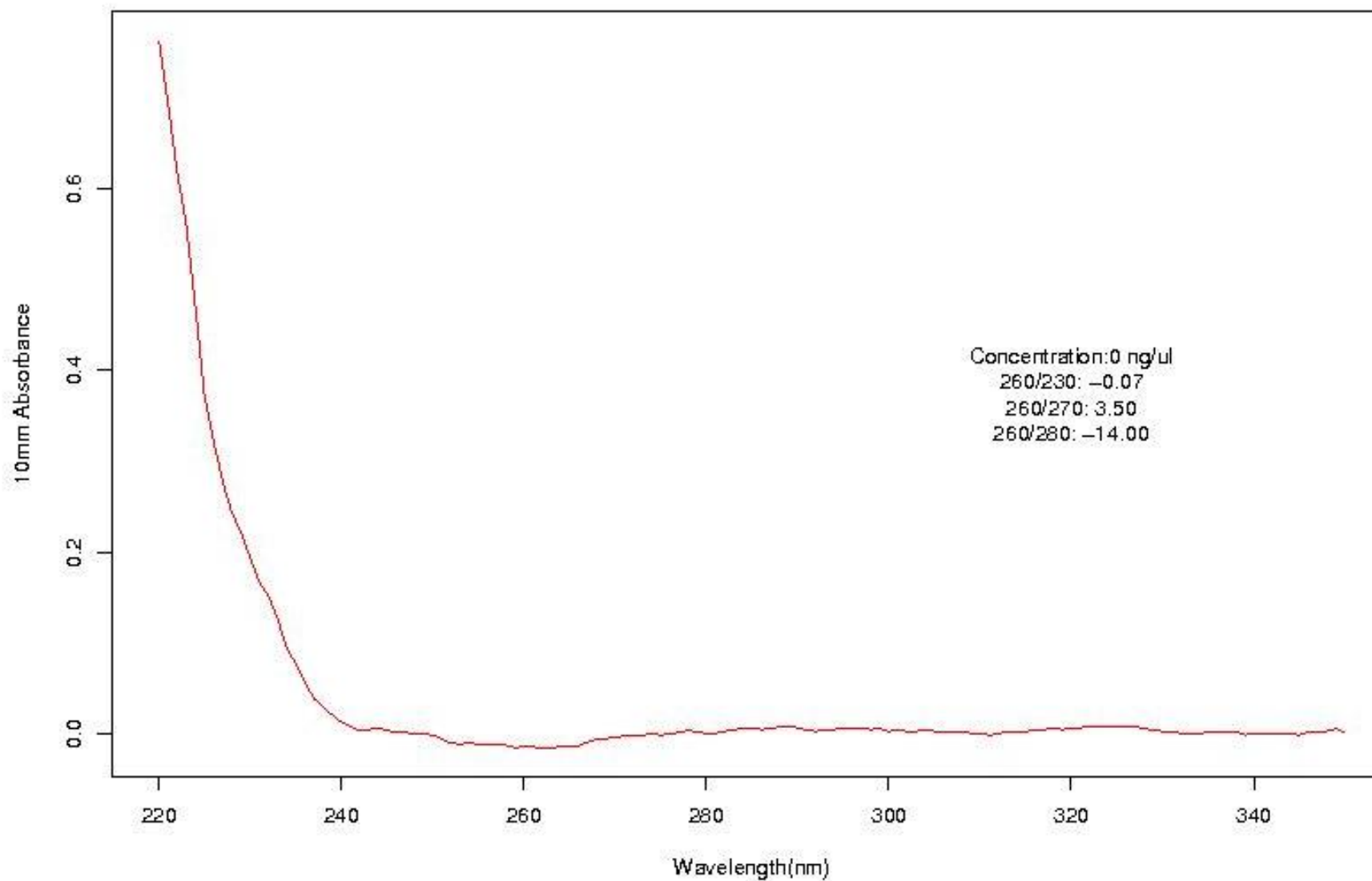


Sample ID	ng/ul	A260	A280	260/280	260/230	340 raw
15	4,30	0,086	0,094	0,91	0,19	0,047
gd9	3,94	0,079	0,056	1,40	0,09	0,070
gd7	3,93	0,079	0,032	2,45	0,04	0,002
sk7	85,33	1,707	0,835	2,04	1,45	0,012
gd22	0,63	0,013	-0,013	-0,98	0,03	-0,005
11	4,33	0,087	0,048	1,80	0,18	0,440
gd11	0,81	0,016	-0,014	-1,18	0,05	-0,003
gd30	0,99	0,020	0,004	4,94	0,07	-0,003
9	4,63	0,093	0,036	2,57	0,18	0,010
gd29	4,37	0,087	0,063	1,40	0,06	0,049
gd21	2,96	0,059	0,055	1,08	0,07	-0,002
gd14	1,76	0,035	0,007	4,94	0,05	-0,002
gd1	1,13	0,023	0,015	1,54	0,13	0,024
gd33	0,88	0,018	-0,007	-2,55	0,13	-0,007
gd32	13,38	0,268	0,143	1,87	0,36	0,001
gd12	2,69	0,054	0,042	1,27	0,09	-0,002
gd34	0,37	0,007	-0,012	-0,61	0,05	-0,009
19	3,86	0,077	0,039	2,00	0,17	-0,001
gd31	0,98	0,020	0,014	1,39	0,05	-0,003
gd19	1,70	0,034	0,034	0,99	0,02	-0,010
6	5,58	0,112	0,063	1,77	0,20	0,049
gd15	1,59	0,032	0,018	1,81	0,02	-0,029
gd23	3,03	0,061	0,010	5,98	0,15	0,316
gd17	3,98	0,080	0,026	3,02	0,17	-0,009
gd35	3,58	0,072	0,022	3,30	0,01	-0,015
sk6	6,83	0,137	0,069	1,99	0,25	0,003
sk6	7,46	0,149	0,096	1,55	0,24	0,162
gd26	1,20	0,024	-0,016	-1,50	0,02	-0,005
8	5,30	0,106	0,046	2,31	0,20	0,034
13	3,45	0,069	0,017	4,14	0,14	0,139
gd20	2,93	0,059	0,030	1,93	0,10	-0,021
ccf3997	8,84	0,177	0,065	2,70	0,21	0,001
ccf3998	9,33	0,187	0,084	2,21	0,54	0,009
ccf3999	6,59	0,132	0,062	2,11	0,40	0,493
11	4,02	0,080	0,018	4,49	0,16	0,011



Častý zdroj nečistot je octan (acetate)

nanodrop – psm_100mM_NaOAc



Základní techniky – zjištění koncentrace a kvality DNA

- Nanodrop měří absorbance DNA + šum (RNA, fenol a další látky s absorbancí při 260nm). Při nízkých koncentracích (pod 5ng/ul) není přesný.
- PicoGreen (případně konkurenční Qubit): měří se fluorescence obarvené DNA. Měří pouze DNA (fluorometrem) i o nízkých koncentracích (10 pg/μl -100 ng/μl)

Table 3. Comparison of PicoGreen and NanoDrop quantitation measurements in DNA samples greater than 20 ng/μl in concentration.

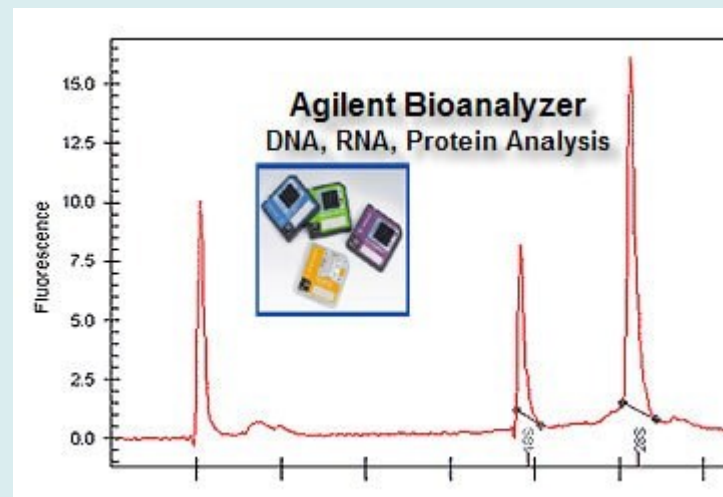
Sample	PicoGreen Average (ng/μl)	NanoDrop Average (ng/μl)	NanoDrop Average A260/280
1	38.4	39.0	1.6
2	37.0	41.2	1.7
3	35.6	34.5	1.6
4	118.1	103.0	1.8
5	102.4	90.0	1.6
6	37.5	35.6	1.7
7	23.9	23.7	1.6
8	35.6	37.2	1.6
9	44.1	44.1	1.6
10	37.9	38.4	1.6
11	33.5	33.4	1.6
12	40.1	41.4	1.6
13	37.1	39.5	1.6
14	46.2	45.9	1.7
15	37.2	40.2	1.6
16	28.3	29.3	1.6

Table 4. Comparison of PicoGreen and NanoDrop quantitation measurements for DNA samples less than 10 ng/μl in concentration.

PicoGreen Average (ng/μl)	NanoDrop Average (ng/μl)
0.37	3.77
1.31	5.70
4.19	7.83
0.74	5.10
1.39	5.67
2.47	5.97
0.61	4.33
2.15	6.30
3.17	7.40

Základní techniky – zjištění koncentrace a kvality DNA

- Kapilarní elektroforeza
- Kvantita i délka fragmentů



Základní techniky- dekontaminace DNA

- velmi odolná molekula, problémy s malými řetězci o nízké koncentraci
- **povrchy:** γ -záření, oheň, SAVO, UVA (10 min, velmi blízko k zářivce)

Table 1. Elimination of surface contamination.

	Air shower	Water	Detergent	DNA away®	UV	CoPA solution	Bleach
% deg	0	50,3	70,9	62,3	95,6	99,5	99,4
SD	45,5	34,5	8,9	16,9	5,6	0,5	1,2

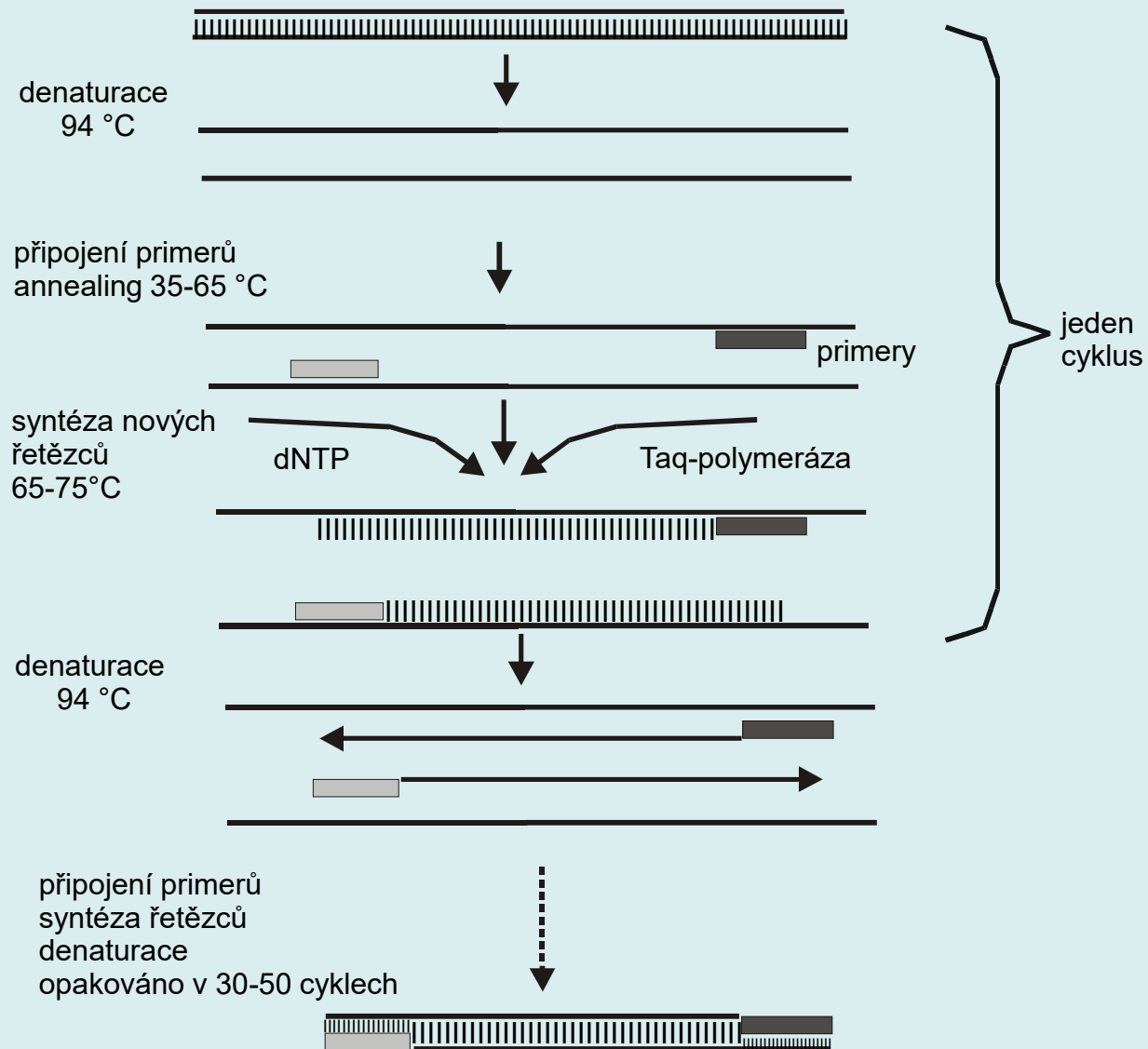
- **roztoky:** UVA záření, γ -záření, DNAáza, hl-dsDNase (heat-labile double strand specific) na čištění roztoku primerů
- autokláv spolehlivě účinný až po 120 minutách
- alkoholy sráží DNA, čímž se může stát více odolnější

Základní techniky – PCR

PCR (*Polymerase chain reaction*, polymerázová řetězová reakce)

- replikace DNA *in vitro*, slouží k namnožení úseku, který je ohraničen primery
- primer: očko, oligonukleotid který je potřeba k iniciaci replikace
- reakční směs obsahuje:
 - volné nukleotidy (dATP, dGTP, dCTP, dTTP)
 - DNA-polymerázu a její katalyzátor Mg^{2+}
 - primery
 - templátová DNA
 - PCR aditiva (BSA, DMSO, betain... omezení vzniků sek. struktur rDNA, vlivu G-C% bohaté DNA)

Princip PCR



Ze dvou řetězců získáme po 30 cyklech teoreticky celkem $2^{30} = 107$ milionů kopií

Základní techniky – PCR

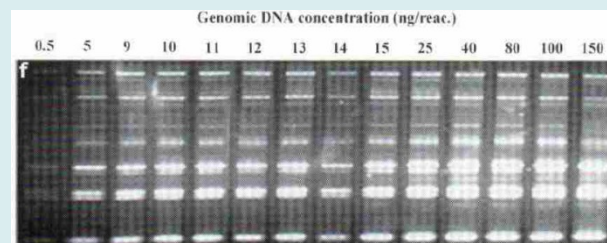
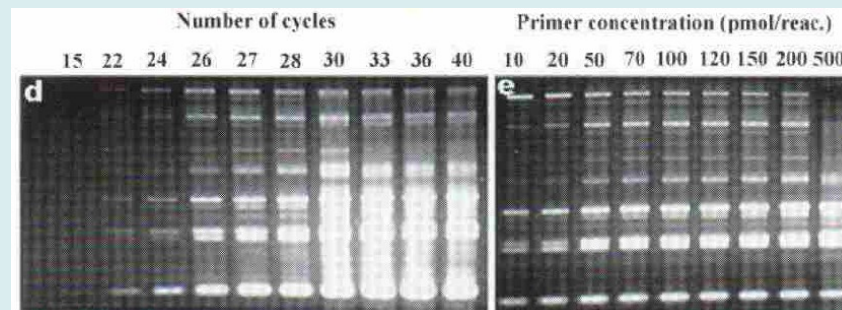
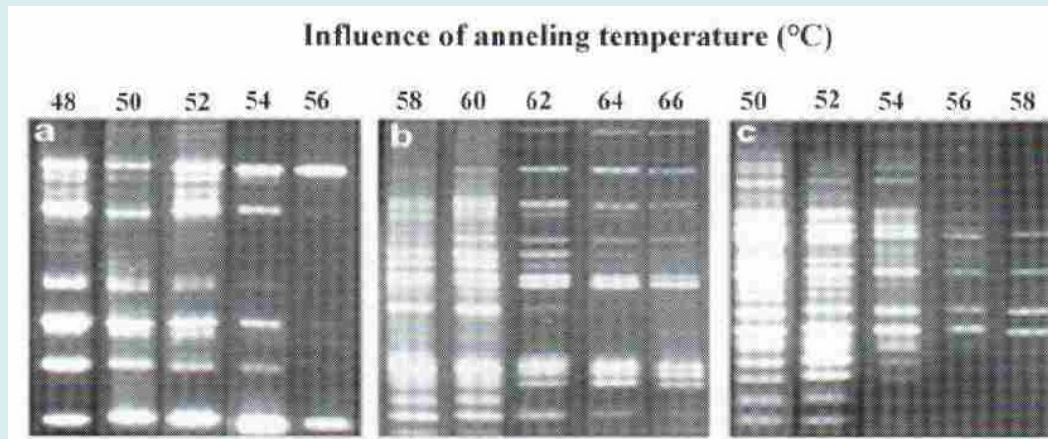
- annealing – teplota zaručuje specifitu nasednutí primerů

35-55°C - nespecifické nasednutí

55-65°C - specifické (komplementární) nasednutí

- u každé PCR metody je nutná optimalizace reakce
 - koncentrace Mg_{2+}
 - koncentrace DNA
 - teplota annealingu (gradient PCR, touchdown PCR)
 - rychlost RAMP (rychlost změn teplot, nejdůležitější je ta z denaturace na annealing)
 - typ polymerázy
 - PCR aditiva
 - PCR fail save kit (různé kombinace enzymu, aditiv...)
 - nested PCR (PCR produkt z 1. PCR se použije jako templát v druhé PCR, kde se použijí vnitřní primery).

Vliv teploty annealingu, počtu cyklů, koncentrace primerů a templátové DNA na příkladu ISSR fingerprintu



Plant Molecular Biology Reporter **19**: 209–215, 2001

© 2001 International Society for Plant Molecular Biology. Printed in Canada

Základní techniky – restriční enzymy

- štěpí nemetylovanou DNA v místech specifických sekvencí
- palyndromatické, 4 až 7 nukleotidové

Restriční místo *EcoRI*



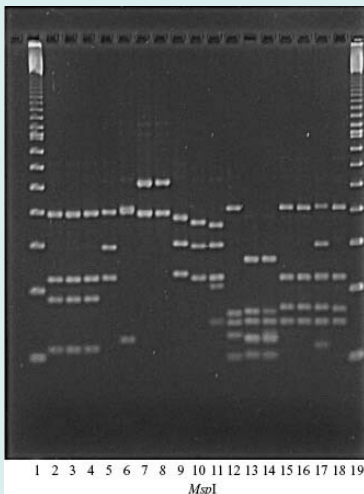
Mol. Tax. metody - RFLP

RFLP (Restriction Fragment Length Polymorphism)

- polymorfismus délky restrikčních fragmentů
- **RFLP-PCR**: štěpení PCR produktu (mtDNA, rDNA, β -tubulin) pomocí restr.

Enzymů

- RFLP detekuje rozdíly v sekvenci PCR produktu, ale oproti sekvenování poskytuje jen malé množství znaků, z tohoto důvodu je v současné době používána hlavně k rutinní detekci patogenů o známé sekvenci



Elektroforetogram restrikčních fragmentů, oblast ITS-rDNA. Štěpeno *MspI*. *Phytophthora* spp.

Mol. Tax. metody - RFLP

- používá se k odlišení izolátů u arbuskulárních hub (*Glomus intraradices*)
- mtLSU je zde velmi variabilní, odlišuje jedince

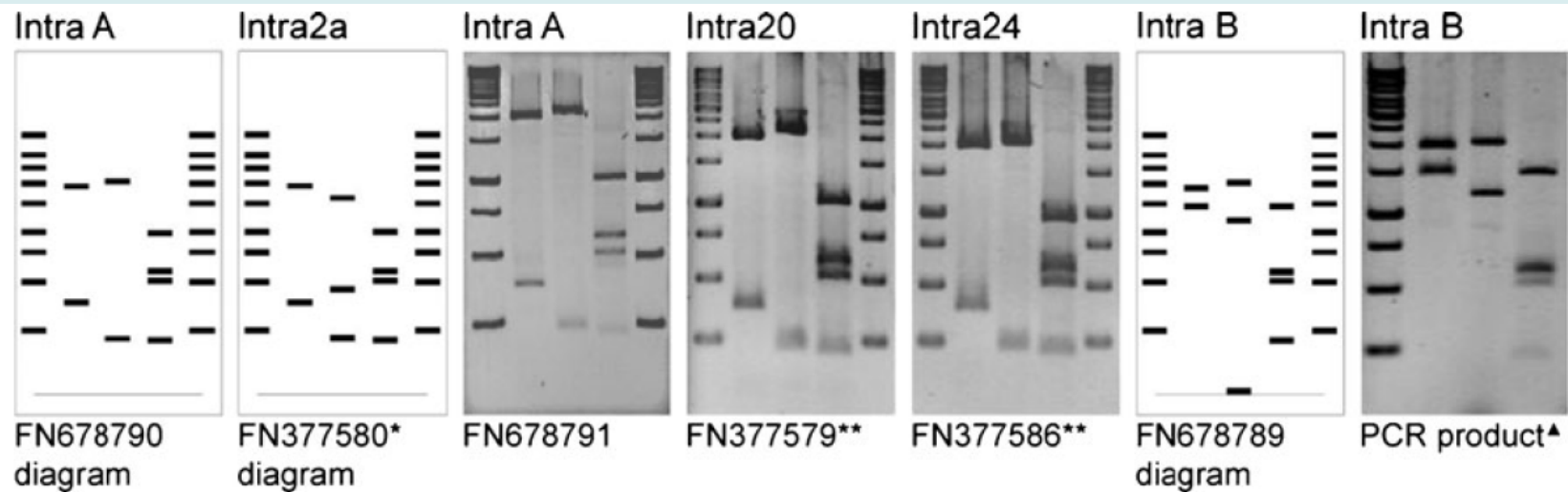


Fig. 1 RFLP banding patterns of haplotypes A and B (Intra A and Intra B) of the mtLSU in comparison with the respective patterns of Intra2a, Intra20, and Intra24. PCR products were amplified using the primer pair RNL-29/RNL-30 and digested using the restriction enzymes *Dra*III, *Bsa*II, and *Hind*III and loaded in the same order per sample onto the gels. DNA ladder in the left-most lane of each gel (bp): 250, 500, 750, 1,000, 1,500, 2,000, 2,500, 3,000, 3,500, 4,000,

5,000, 6,000, 8,000, and 10,000. The size standard in the left-most lane of the diagrams produced in silico (bp): 250, 500, 750, 1,000, 1,500, 2,000, 2,500, 3,000, and 4,000. Accession numbers of the respective clones are shown below the patterns. *Triangle*, original digested PCR product from which the clone FN678789 was derived. *Single asterisk*, from Börstler et al. (2008); *double asterisk*, from Börstler et al. (2010)

specifická PCR – rychlá identifikace

- Specifické primery navržené pro jednotlivé druhy (či populace) využívány rutinně ve fytopatologii a lékařství, kde pracujeme s environmentální DNA.
- cíleno na rDNA, b-tubulin, různé genomické úseky
- Nutnost křížových zkoušek kdy se primery testují na směsi různých druhů hub
- „kvantitativní PCR“ umožňuje změřit množství PCR produktů – kvantifikace infekce, kolonizace...

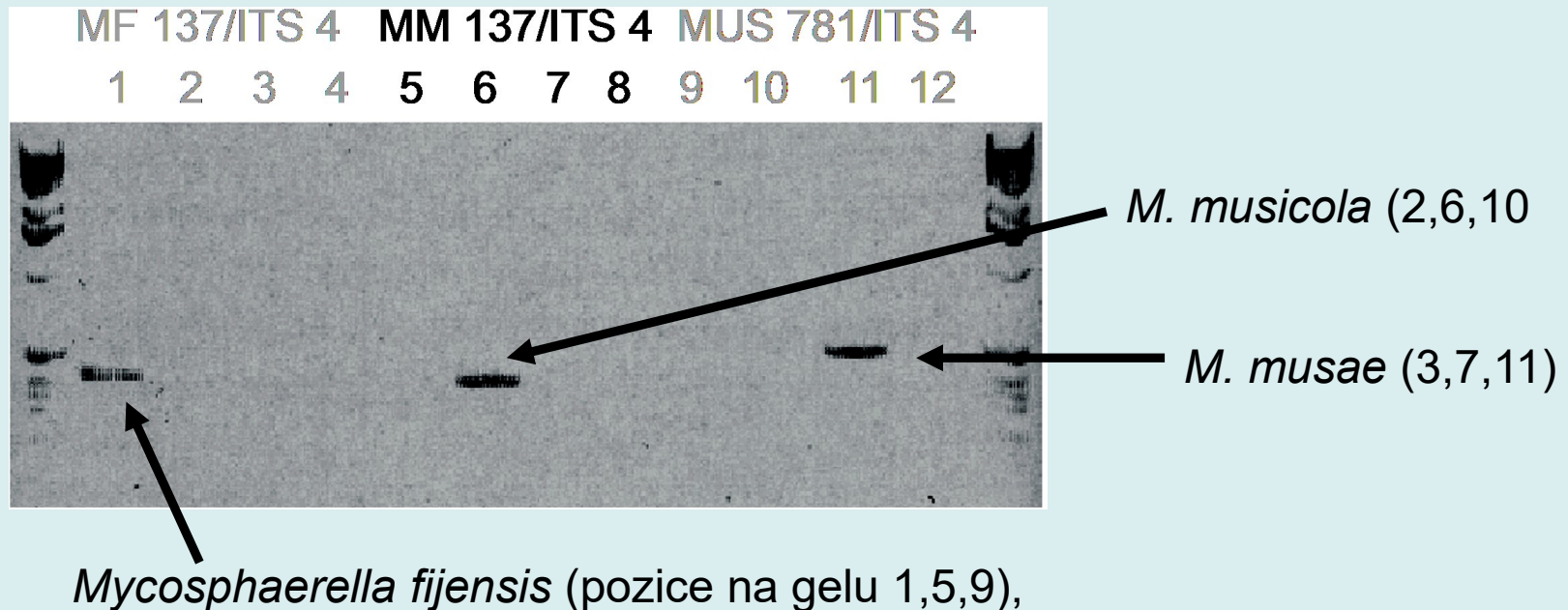


Table 2. *Neofabraea* primer sets and their respective species-specific fragment sizes.

<i>Neofabraea</i> species	Upper primer	Lower primer	Fragment size (bp)
<i>N. alba</i>	Neofab-UpTub-100	Neo_alba-loTub-439	358
<i>N. malicorticis</i>	Neofab-UpTub-100	Neo_mal-loTub-262	178
<i>N. perennans</i>	Neofab-UpTub-100	Neo_per-loTub-382	300
<i>N. populi</i>	Neofab-UpTub-100	Neo_pop-loTub-475	397
<i>N. sp. nov.</i>	Neofab-UpTub-100	Neo_spnov-loTub-319	236

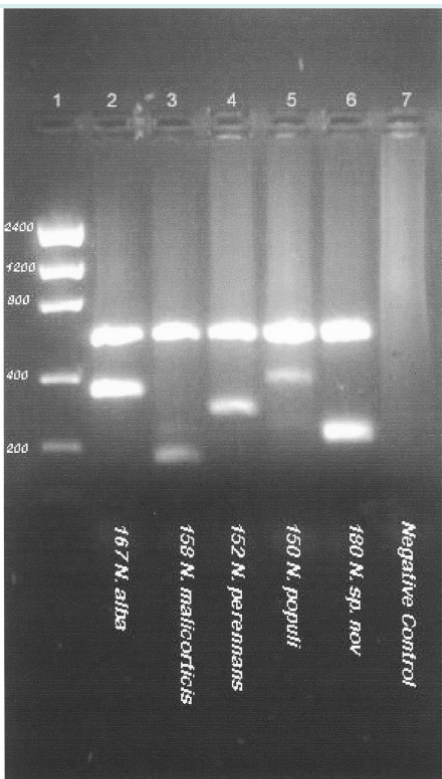


Fig. 6. Amplification of each of the five species of *Neofabraea* using a combination of the primers Neofab-upTub-100 with Neo_alba-loTub-439, Neo_mal-loTub-262, Neo_per-loTub-382, Neo_pop-loTub-475 and Neo_spnov-loTub-319 in a single reaction mixture. The universal fungal primers UN-UP18S-42 and UN-LO28S-22 used as internal positive controls amplified an approximately 600 bp fragment.

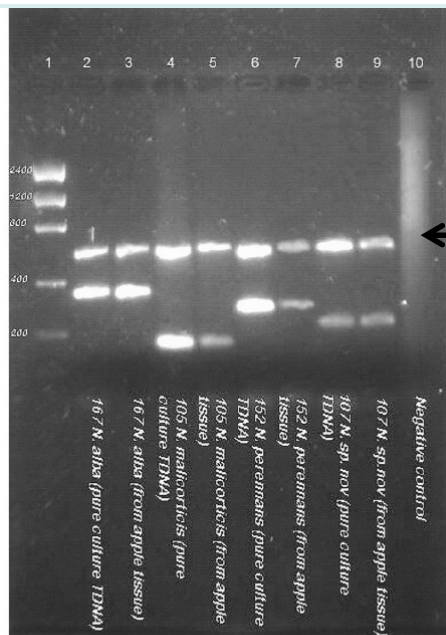


Fig. 7. PCR of *Neofabraea* DNA isolates 167 (*N. alba*), 105 (*N. malicorticis*), 152 (*N. perennans*), and 107 (*Neofabraea* sp. nov.) extracted from infected apple fruits (Lanes 3, 5, 7 and 9, respectively). Each isolate was amplified using the primer combination designed to be specific for that species. Lanes 2, 4, 6 and 8 represent pure culture total DNA (TDNA) of *N. alba*, *N. malicorticis*, *N. perennans*, and *Neofabraea* sp. nov., respectively, amplified with the appropriate species-specific primer set. Universal fungal primers UN-UP18S-42 and UN-LO28S-22 were used in each reaction as internal positive controls, generating an approximately 600 bp band for all isolates.

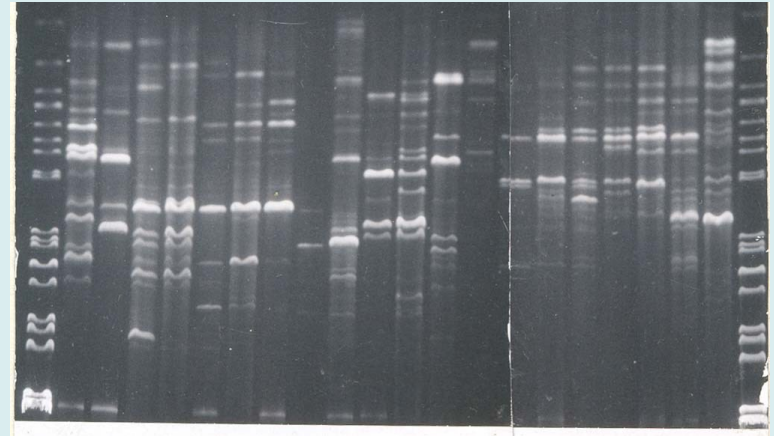
Amplikon získaný pomocí univerzálních houbových primerů pro kontrolu přítomnosti houbové DNA

Species specific identification of the *Neofabraea* pathogen complex associated with pome fruits using PCR and multiplex DNA amplification

Mol. Tax. metody –PCR fingerprinting

- skupina multilokusových metod založených na PCR

- Fingerprít (*fingerprint, banding pattern*)
„otisk prstu“, unikátní kombinace
alel (proužků) charakterizující daný vzorek



- jednotlivé metody se liší typem primeru, a teplotou annealingu
- používají se na „typování“ izolatů morfologicky špatně odlišitelných hub

Lze rozdělit na dvě skupiny

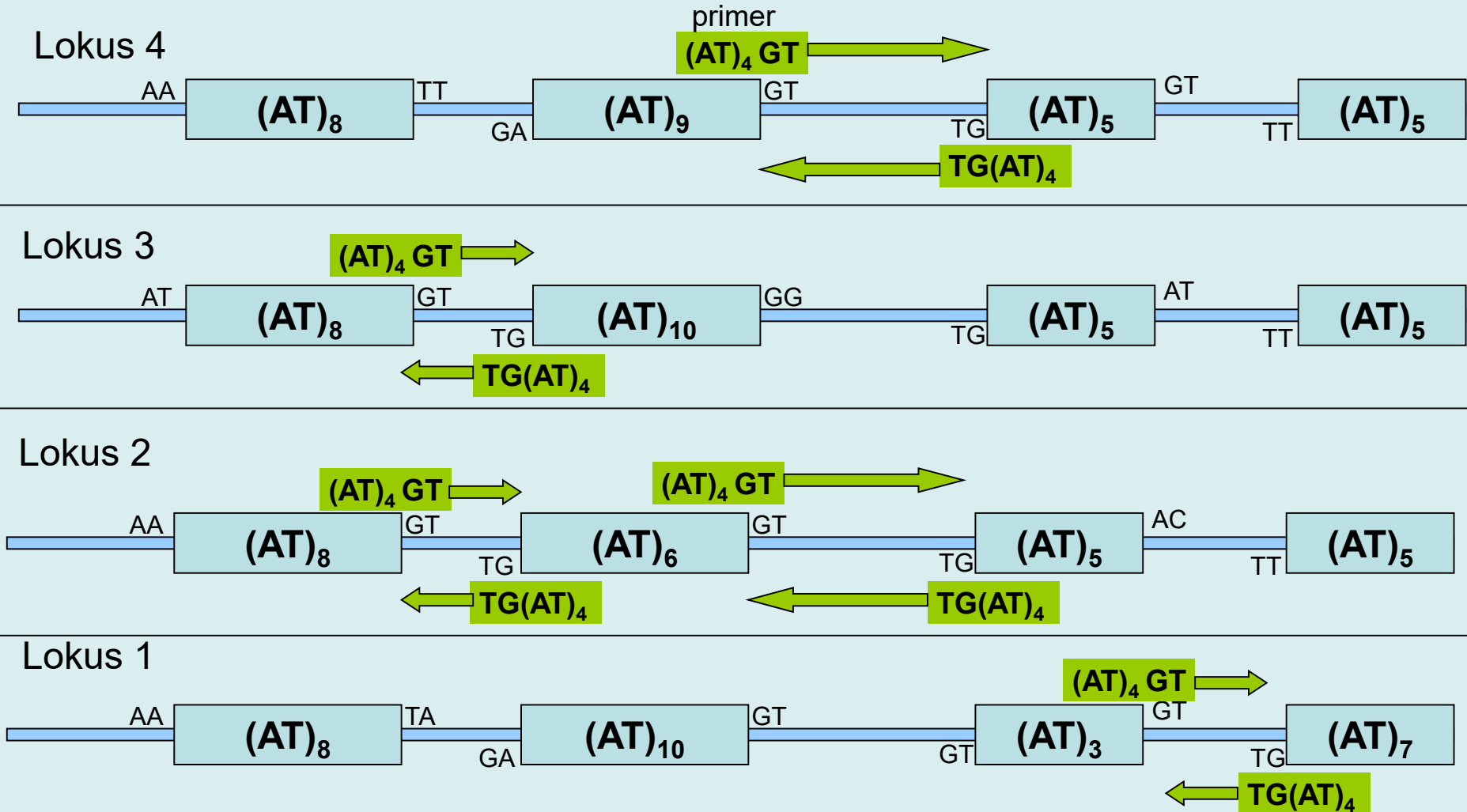
- *High-stringency PCR site-specific polymorphisms*
- *Low-stringency PCR* neboli *MAAP (Multiple Arbitrary Amplicon Profiling)*

Mol. Tax. metody –PCR fingerprinting

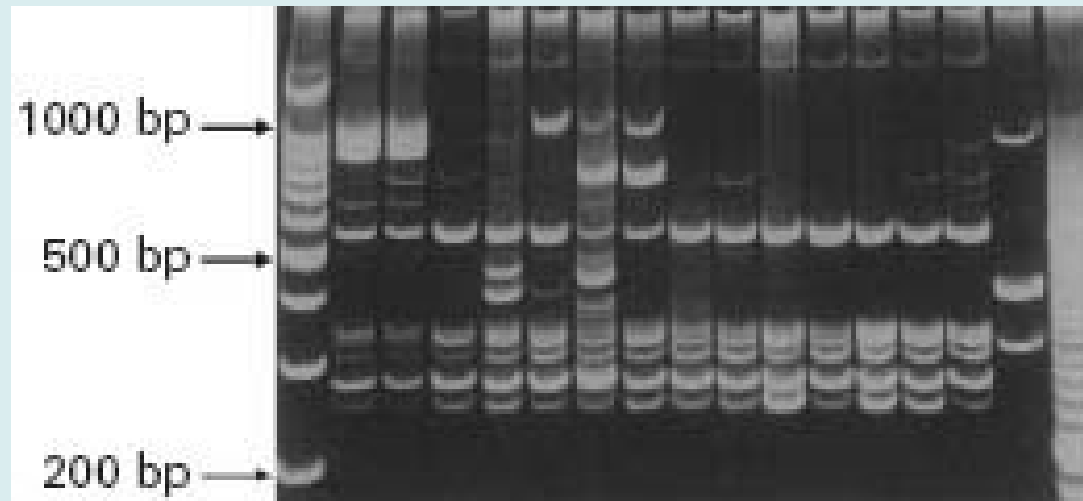
ISSR-PCR (*Inter single sequence repeat PCR*)

- jako primer použita sekvence SSR
- Namnoží se úseky SSR a oblasti oddělující SSR
- díky možnému přeskokování primerů během PCR se můžou namnožit nespecifické proužky – silné pozadí
- řešení je v ukotvení primeru na konec SSR. Verze zvaná **anchored-ISSR**
- primer (AG)₈ GT, nebo směs - degenerované primery (AG)₈ GT, (AG)₈ TG, (AG)₈ AG
- 16-18 bp dlouhý primer
- annealing 45-65 °C
- metoda dobře reprodukovatelná a snadná

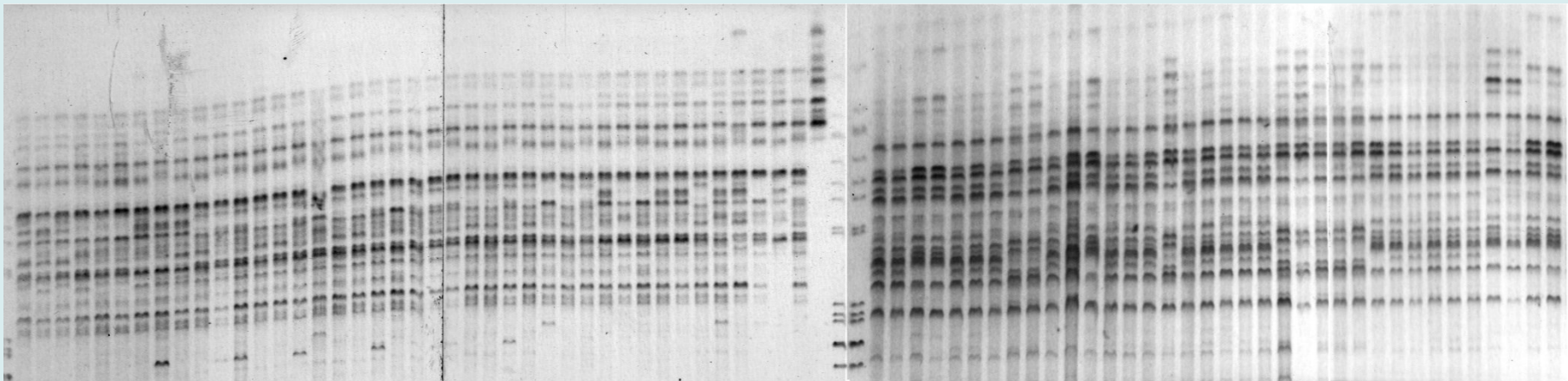
Rozložení mikrosatelitu (AT) v jednotlivých lokusech a zobrazení možného polymorfismu v počtu repetit a délce vmezeřených sekvencí. Dvojice zakotvených primerů naznačuje úsek zmnožený při PCR



ISSR. Ukázka fingerprintu jedinců klouzku sličného (*Suillus grevillei*) - metoda je schopna odlišit jednotlivé klony



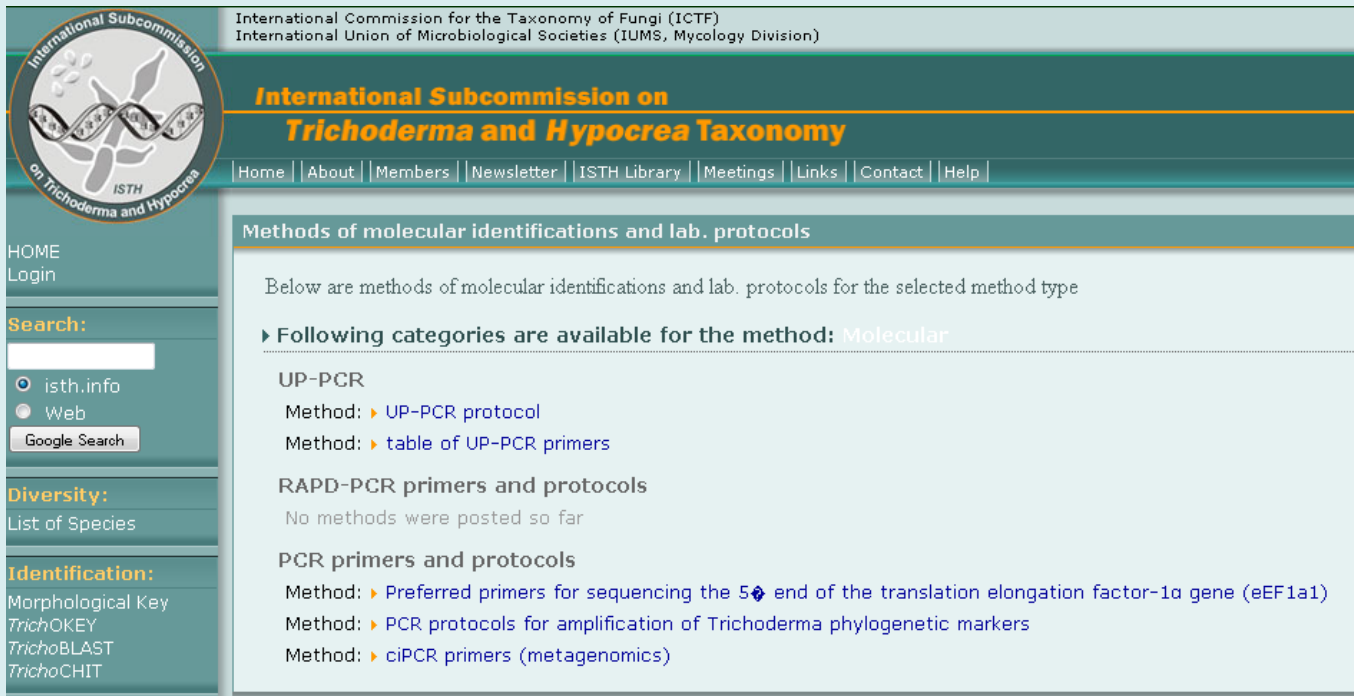
ISSR. Ukázka fingerprintu dvou druhů rodu *Geosmithia* (nepohlavní druhy)



Mol. Tax. metody –PCR fingerprinting

High-stringency PCR

- **UP-PCR** (universal primed PCR) univerzálně nasedající PCR
 - primery 18-20 bp, univerzální motivy ve variabilních úsecích genomu (původně u bakterií)
 - annealing 52-56 °C (dle primeru)
 - rychlá „ramp“ fáze, 4°C/s (rychlost změny teploty, zvyšuje specifitu)



International Commission for the Taxonomy of Fungi (ICTF)
International Union of Microbiological Societies (IUMS, Mycology Division)

**International Subcommittee on
Trichoderma and Hypocrea Taxonomy**

[Home](#) | [About](#) | [Members](#) | [Newsletter](#) | [ISTH Library](#) | [Meetings](#) | [Links](#) | [Contact](#) | [Help](#)

HOME
Login

Search:

☐ isth.info
☐ Web

Diversity:
[List of Species](#)

Identification:
[Morphological Key](#)
[TrichOKEY](#)
[TrichoBLAST](#)
[TrichoCHIT](#)

Methods of molecular identifications and lab. protocols

Below are methods of molecular identifications and lab. protocols for the selected method type

► **Following categories are available for the method: Molecular**

UP-PCR
Method: ► [UP-PCR protocol](#)
Method: ► [table of UP-PCR primers](#)

RAPD-PCR primers and protocols
No methods were posted so far

PCR primers and protocols
Method: ► [Preferred primers for sequencing the 5' end of the translation elongation factor-1 \$\alpha\$ gene \(eEF1a1\)](#)
Method: ► [PCR protocols for amplification of Trichoderma phylogenetic markers](#)
Method: ► [ciPCR primers \(metagenomics\)](#)

- Poměrně reprodukovatelné
- Oblíbené na typování do druhů
- Úroveň druhů a nižší
- Odlišení izolátů

TABLE 2. Primers used in this study

Primer type	Primer	Sequence
UP	3-2	5'-TAAGGGCGGTGCCAGT-3'
	AS4	5'-TGTGGGCGCTCGACAC-3'
	AS15	5'-GGCTAAGCGGTCGTTAC-3'
	AS15inv	5'-CATTGCTGGCGAATCGG-3'
	AA2M2	5'-CTGCGACCCAGAGCGG-3'
	Fok1 ^a	5'-GGATGACCCACCTCCTAC-3'
	L21	5'-GGATCCGAGGGTGGCGGTTCT-3'

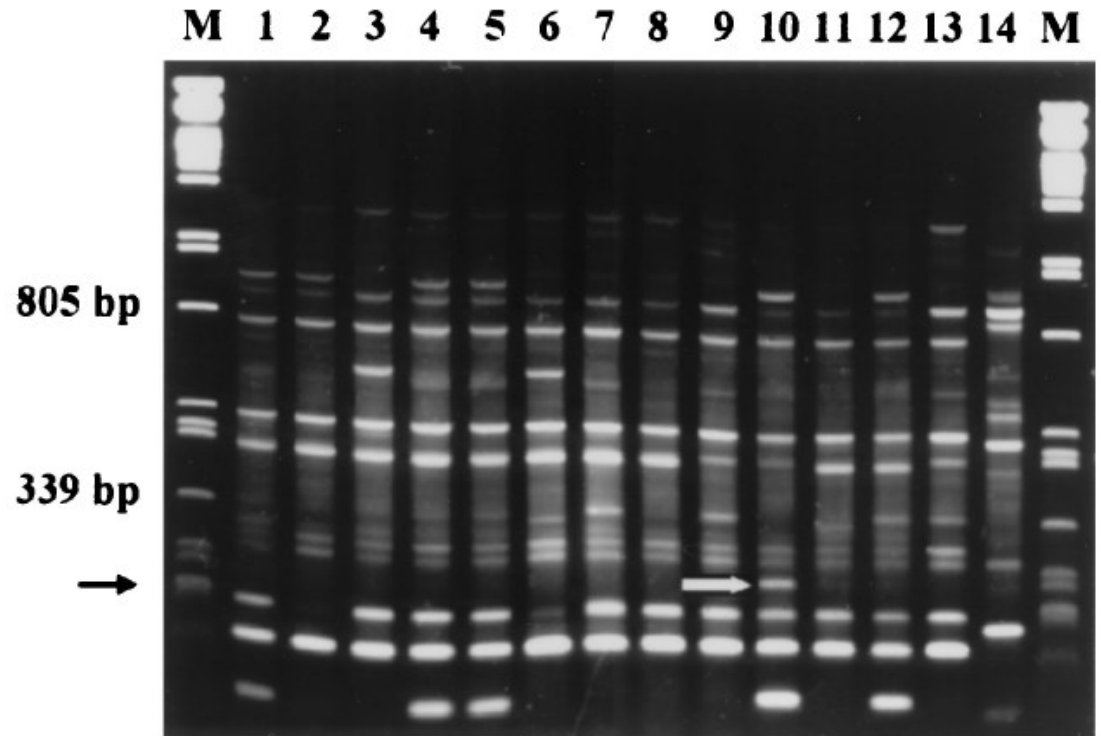
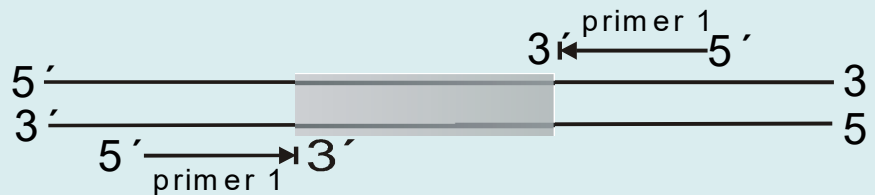


FIG. 1. UP-PCR banding profiles for *C. rosea* strains generated with the AS15inv-AA2M2 UP primer combination. Lanes 1 to 14, strains GR3, IK726, GR6, GR10, GR12, GR11, GR9, GR8, GR4, GR5, GR1, GR13, GR7, and GR2, respectively. Lanes M, molecular size markers (λ phage DNA digested with *Pst*I). Arrows shows the marker of interest for strain GR5.

Mol. Tax. metody –PCR fingerprinting

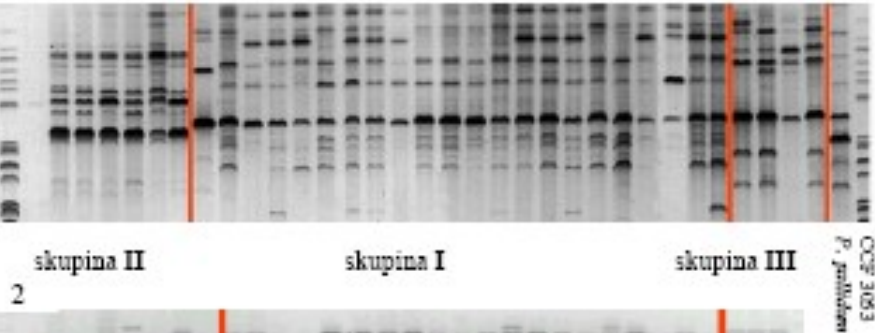
Low-stringency PCR: ap-PCR (arbitrary primed) – náhodné primery

- **RAPD** (Random Amplified Polymorphic DNA) polymorfismus náhodně zmnožených fragmetnů – běžně nazýváno jako „Rapid“ – rychlý

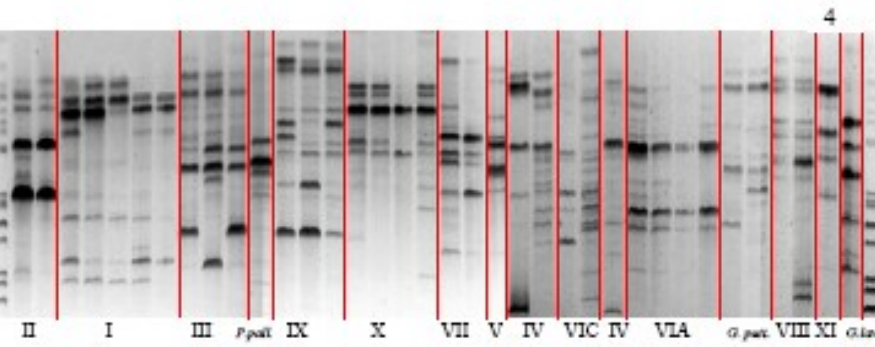
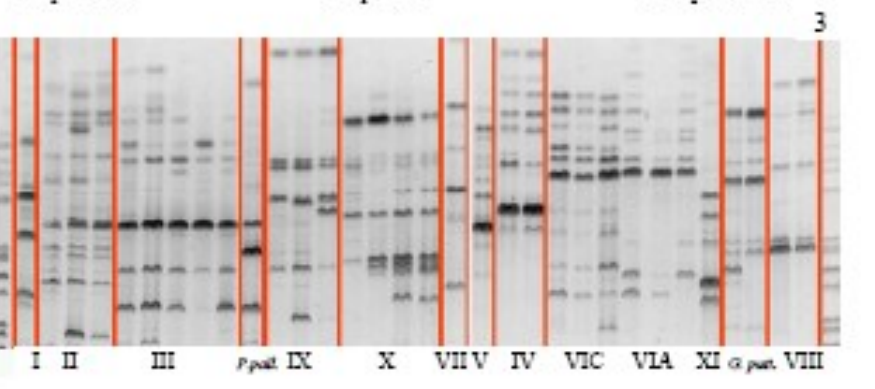
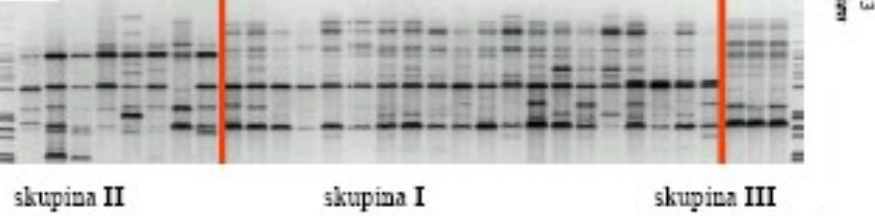


- Nízká teplota annealingu v počátečních cyklech (35 °C) způsobuje, že použité primery nemusí přisednout zcela komplementárně
- jeden primer
- primery nesmyslné a krátké 8-12 bp
- Přidává se $MgCl_2$ který zvyšuje aktivitu polymerázy (nutná optimalizace)
- Při nízkém annealingu a krátkých primerech, primery nasedají nehomologně
vznikne hodně PCR fragmentů

1



2



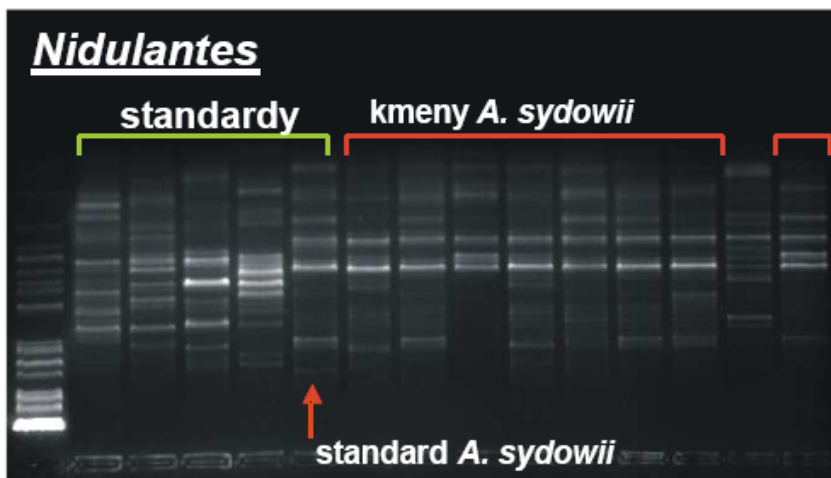
- Velký výběr primerů
- primery se liší taxonomickou citlivostí
- Primery pro zvýraznění rozdílů mezi druhy
- Primery pro studium vnitrodruhové variability

Tabule IV Ukázka RAPD *fingerprints* zástupců rodu *Geosmithia* 1. Skupiny I-III a *P. pallidum*, primer 10R. 2. Zástupci skupin I-III, primer OPA20. 3. Vybraní zástupci skupin rodu *Geosmithia*, primer 10R. 4. Vybraní zástupci skupin rodu *Geosmithia*, primer 8F.

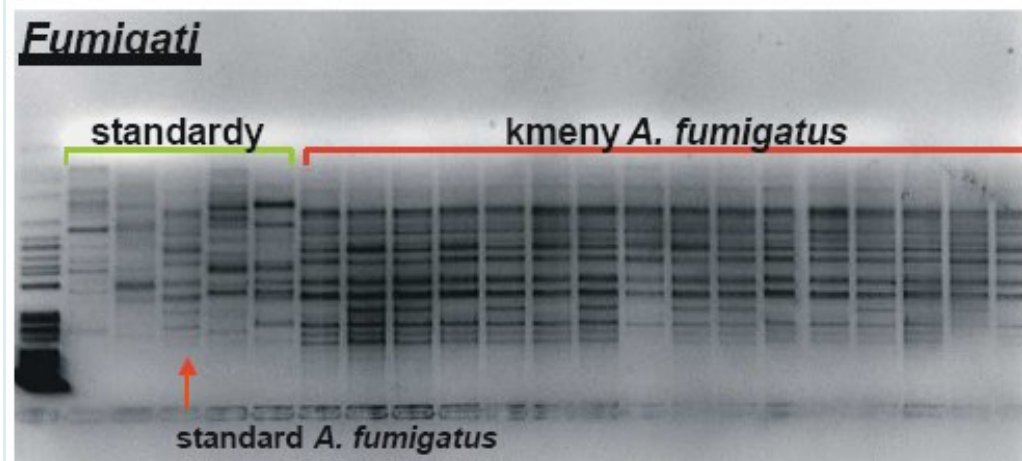
Mol. Tax. metody –PCR fingerprinting

rep-PCR (repetition PCR)

- metody používající primer odpovídající nějaké repetitivní sekvenci
- annealing 40-55°C
- **M13-PCR fingerprinting** - primery M13 a M13core – odvozeny od ubikvitního fága M13 (jde vlastně o minisatelit)



Obr. 3.1 Příklad fingerprintu získaného po amplifikaci s primerem M13 u zástupců sekce *Nidulantes*. Jako standardy byly vybrány dříve osekvenované druhy ze stejné sekce.



Obr. 3.1 Příklad fingerprintu získaného po amplifikaci s primerem M13 u zástupců sekce *Fumigati*. Jako standardy byly vybrány dříve osekvenované druhy ze stejné sekce.

Mol. Tax. metody –PCR fingerprinting

rep-PCR (repetition PCR)

- **ERIC-PCR** (*Enterobacterial repetitive consensus*), primery komplementární ke konzervativním rep. oblastem, poprvé známým u enterobakterií
- primery ERIC1R a ERIC2
- annealing 50-58 °C

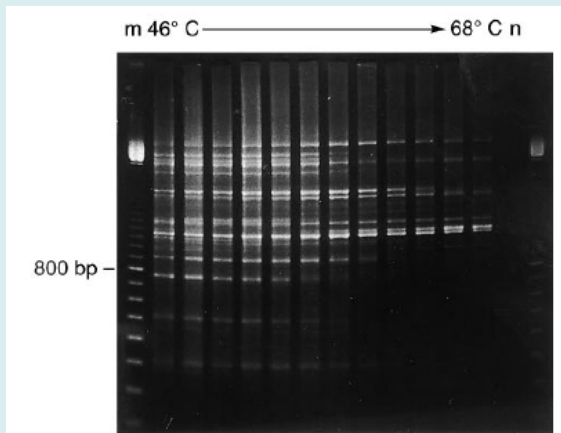


Fig. 1 ERIC-PCR performed on *Escherichia coli* JM101 using a gradient of annealing temperatures. Electrophoresis was on a 2% agarose gel stained with ethidium bromide. Primer annealing temperatures were varied in 2°C increments from 48°C to 68°C (right hand side of gel). Tracks (m) and (n) are 100 base pair ladders (AMRAD Pharmacia) and negative control, respectively

- Pouze při specifickém annealingu jsou amplifikovány výhradně ERIC sekvence

Mol. Tax. metody –PCR fingerprinting

RAPD, rep-PCR, UP-PCR, ISSR, ...

Výhody a využití

- náhodné primery či primery nasedající na univerzální motivy
 - téměř žádná tax. specificita primerů („jeden primer na všechny houby“)
 - velké množství znaků, velmi snadná a levná metoda
- Omezená reprodukovatelnost. PCR fingerprint se liší dle kvality DNA, typu DNA polymerázy, PCR cyklu. Tj. samotný fingerprint nelze reprodukovat mezi laboratořemi (proto nemá publikace obrázků fingerprintů velký smysl). Ale genetické vzdálenosti jím zjištěné jsou reprodukovatelné.
- Použitelné na úrovni druhu a nižší (blízce příbuzné druhy, populace)

Mol. Tax. metody –PCR fingerprinting

RAPD, rep-PCR, UP-PCR, ISSR, ...

Výhody a využití

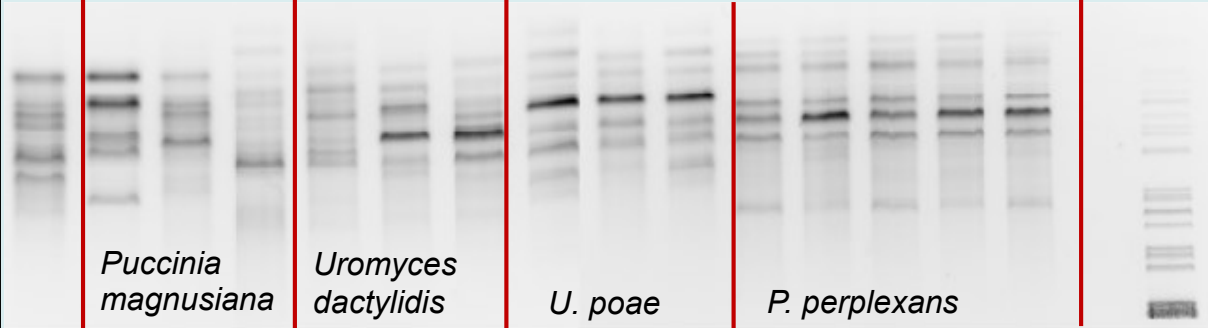
- u porovnávání vzdálenějších taxonů je problém s nepravou homologií proužků (komigrace)

- studuje pouze dominantní alely (tj. zachytíme jen produkt komplementrání k primerům, kdežto absence PCR produktu může být způsobena řadou různých mutací, takže nejde vyhodnotit „nulová alela“)

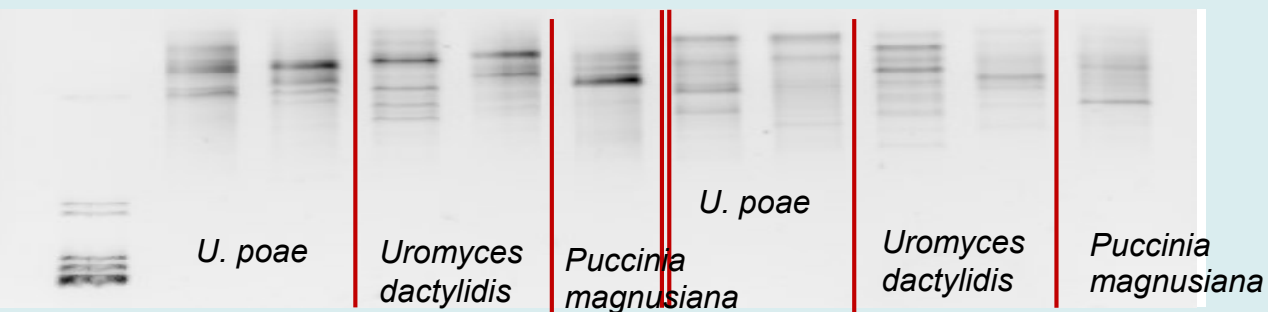
- dříve se používalo pro vyhodnocení genetických vzdáleností a konstrukci kladogramů

- vhodná při „typování“ - redukce identických klonů či druhů hub

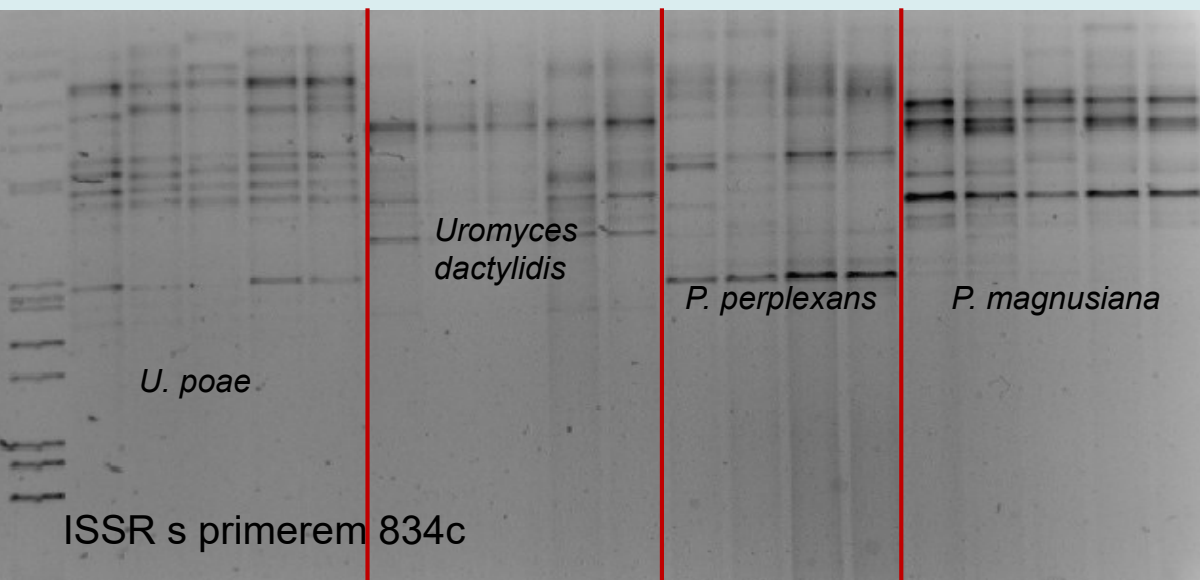
(dereplikace)



Rep-PCR s primerem M13 core (převzato z Hejná 2008)



RAPD s primerem 8F (5 vzorků vlevo) a 10R



ISSR s primerem 834c

- Hledání vhodné fingerprintov metody
- Optimalizace primerů
- Možnost kombinace různých metod

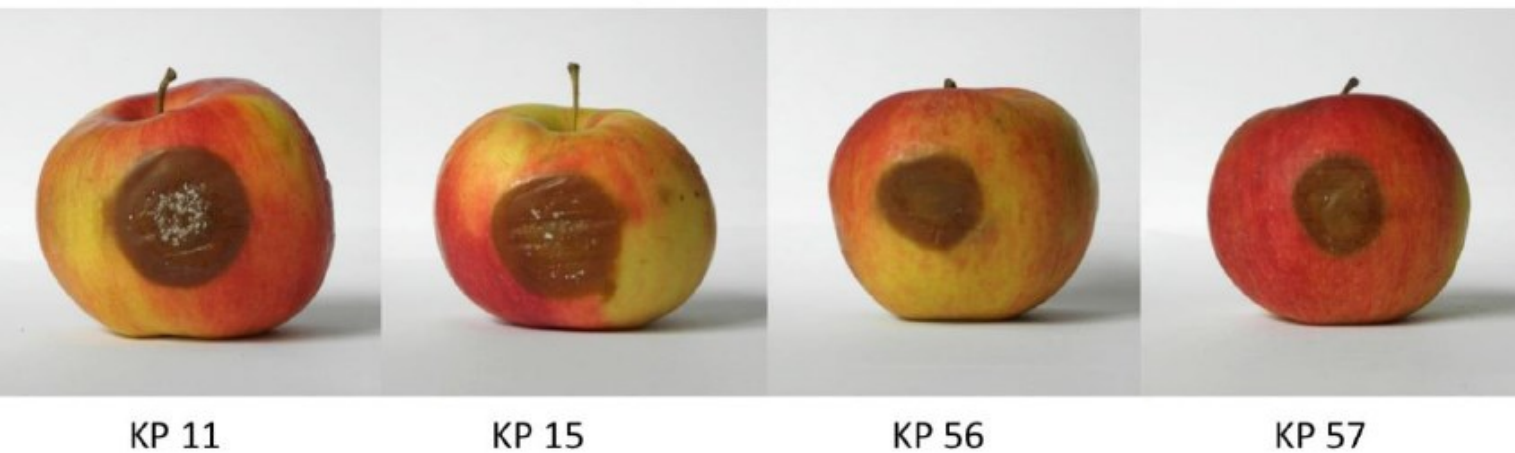


Obr. 7

PCR fingerprinty s primerem M13-core, část 2/2. A: *N. alba* (CPPF 467), K: *C. kienholzii* (CBS 126461), P: *N. perennans* (CPPF 468), M: *N. malicorticis* (CBS 102863), 4–81: KP4–KP81. ■ *N. alba*, ■ *N. perennans*, ■ *N. malicorticis*, ■ *C. kienholzii*, ■ KP4.

Rep-PCR s primerem M13 core hub rodu *Neofabraea* (převzato z Pešicová 2013)

Neofabraea perennans



Ověření identity sbírkových kmenů

Potom



Předtím



Metody PCR- fingerprintingu (RAPD, rep-PCR)

Mol. Tax. metody –mikrosatelity

High-stringency PCR site-specific polymorphisms

- vysoká teplota annealingu, **55-65 °C** zaručuje specifické nasednutí
- patří sem zejména metody **analýzy mikrosatelitů (SSR)**
- SSR se liší ve složení a počtu repetice (10-50). Tyto repetice se nachází v genomu ve větším počtu lokusů (8-17 u *Aspergillus fumigatus*) a v rámci jednoho lokusu v řadě alel

Mononucleotide SSR (**A**)¹¹

AAAAAAAAAAAA

Dinucleotide SSR (**GT**)⁶

GTGTGTGTGTGT

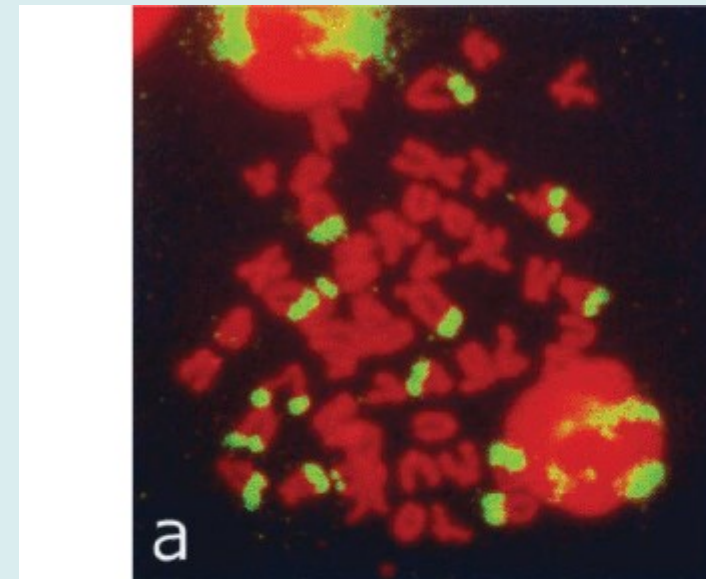
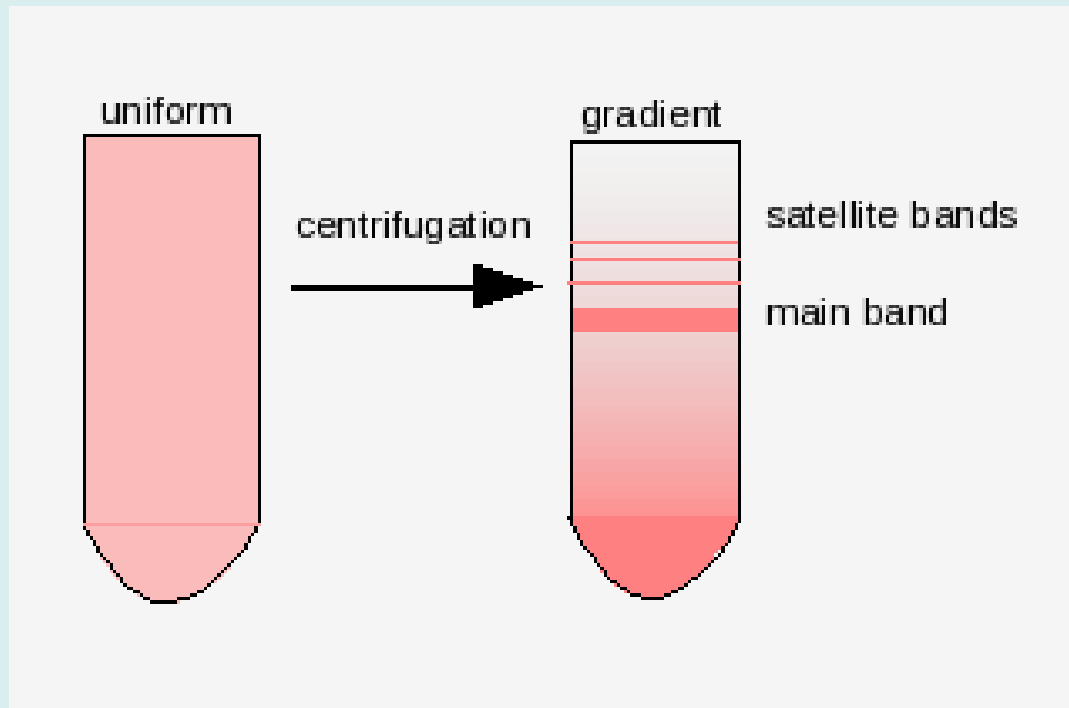
Trinucleotide SSR (**CTG**)⁴

CTGCTGCTGCTG

Tetranucleotide SSR (**ACTC**)⁴

ACTCACTCACTCACTC

- při ultracentrifugaci fragmetované DNA v gradientu CsCl dojde k separaci dle vznášivosti
- **sateletní** frakce mají odlišné složení bazí, typické pro repetitivní úseky
- na chromozomu často v telomeře



Mol. Tax. metody –mikrosatelity

- velmi variabilní i mezi jedinci - základní nástroj populační genetiky

variabilita vzniká díky „**replication slippage**“ a „**Unequal crossing-over during**

meiosis“

(a)

Normal replication



5' CAG CAG CAG CAG CAG CAG n
3' GTC GTC GTC GTC GTC GTC n

(b)

Backward slippage

5' CAG CAG CAG CAG CAG CAG 3' $n+1$

3' GTC GTC GTC GTC GTC GTC 5' n

Second replication

Insertion
mutation

5' CAG CAG CAG CAG CAG CAG n+1 3'

3' GTC GTC GTC GTC GTC GTC n+1 5'

Normal



5' CAG CAG CAG CAG CAG CAG 3'

3' GTC GTC GTC GTC GTC GTC 5'

(c)

Forward slippage

5' CAG CAG CAG CAG CAG $n-1$ 3'

3' GTC GTC GTC GTC GTC n 5'

Second replication

Deletion mutation

5' CAG CAG CAG CAG CAG 3' n-1

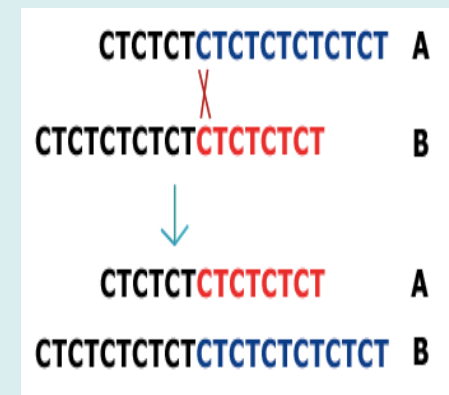
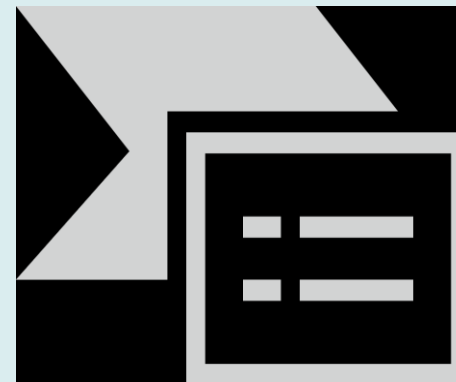
3' GTC GTC GTC GTC GTC 5' n

Normal



5' CAG CAG CAG CAG CAG CAG 3'

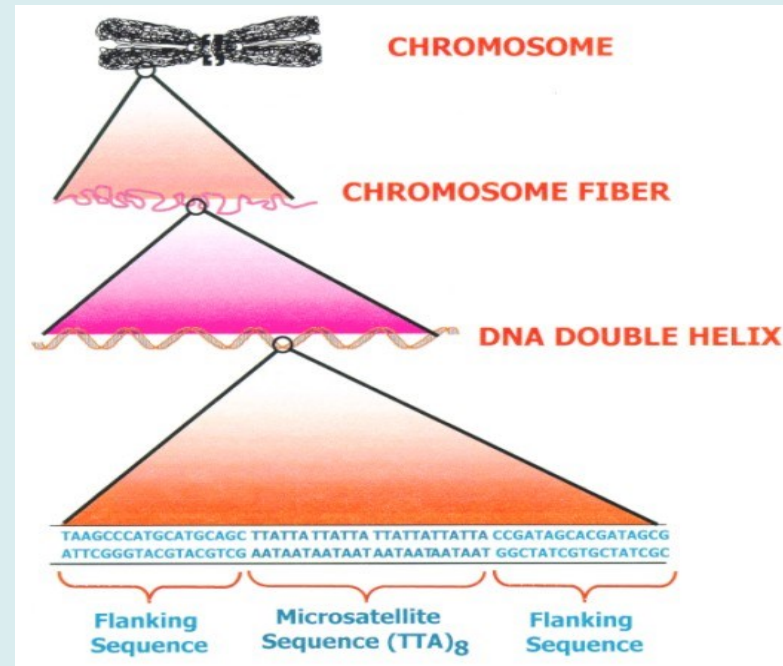
3' GTC GTC GTC GTC GTC GTC 5'



Mol. Tax. metody –mikrosatelity

Pomymorfismus v mikro a minisatelitech je nazýván VNTR (variable number of tandem repeats)

- nejčastěji se používá PCR pomocí specifických primerů navržených z ohraničujících sekvencí
- nutnost znát sekvence ohraničujících SSR (flanking seq.)
- ty můžou být druhově specifické, ale mohou fungovat i na odlišné druhy i rody
- z genomických sekvencí (lze použít i transkriptom)
- sekvenováním ISSR fragmentů atp.

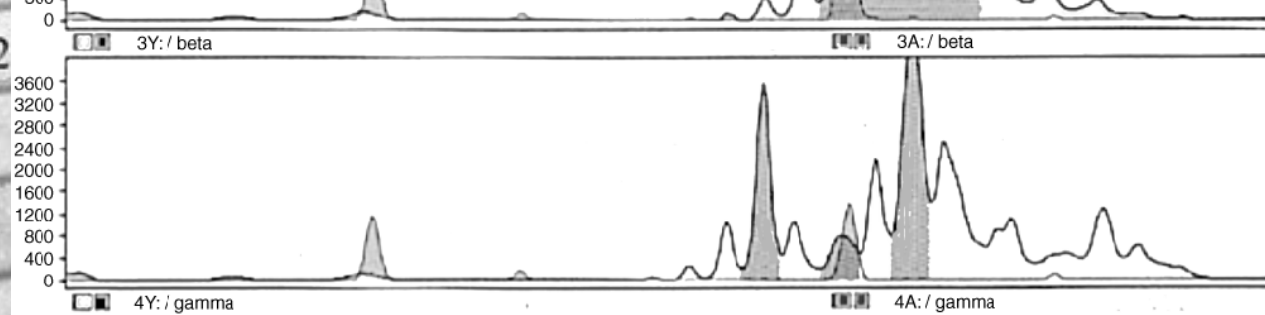
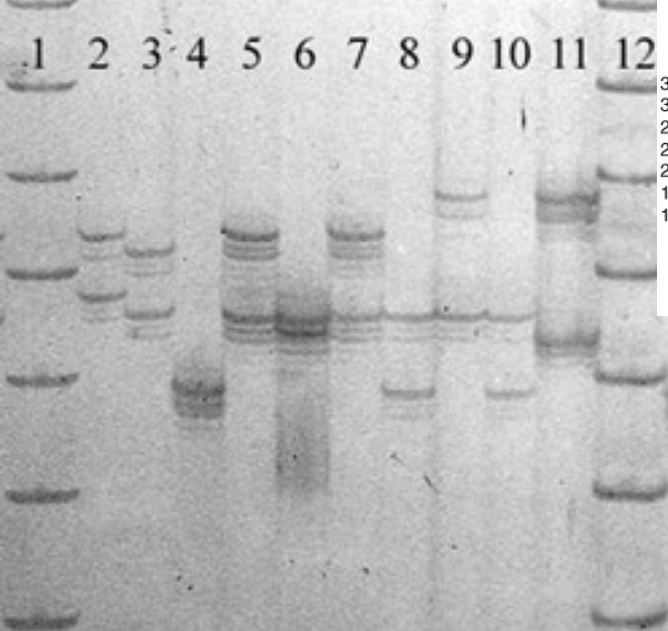


- každý lokus je jeden znak
- hledáme SSR s polymorfními lokusy (SSR se studují vždy u více izolátů)

Table 1 Characterization of sixteen microsatellite loci in *Geosmithia morbida*

Locus ^a	Primer sequence (5'–3')	Repeat	T _a	Allelic class size range (bp)	N _A
GS004	F:TAGACGAATAACAGCGCAAGAA R:CCCGCTACCGTATGAAGGTATT	(TC) ₁₃	61	221–254	5
GS005	F:CGCGTCCCTTATGAATATCTTG R:CCCTAGCAGATTTCAGAGTTT	(CA) ₅ ...(AC) ₆ ...(TG) ₈	60	239–257	3
GS0017	F:AAAATTTCCCTCTTCGTCCTCC R:GTATGGCGTCTATGCAAGTCAG	(CTT) ₅ c(TCT) ₃ (TCCCTCT) ₃ ...(GTCCT) ₅	60	216–240	7
GS0027	F:ACCATTGGCATTGACTCTTCT R:ACCATGACGACACATATCTTGG	(AC) ₁₁	60	220–236	5
GS0036	F:CTAGGGAAAAATGGTCAGCATC R:TCGACATCTAGATCACGGAATG	(AC) ₆	60	125–135	4
GS0037	F:GTTTGCCATCTGCATTACAAAA R:TCCGTCTTATTCTTGGTGTGTG	(CACCCA) ₃ ...(CTCCCA) ₅ ...(CA) ₈ c(CACCGT) ₅ (CA) ₈ t(ACACATAC) ₂	60	213–262	8
GS0041	F:TTTGGCTTCTCCTCTTGTAGC R:GGAATCCCGTGATACGAAATAA	(CA) ₆ ...(CA) ₆	60	161–182	6
GS0048	F:TCCATTTCTCCACCTCTACAC R:TTTCTTACGCAGAAAAGGTAGC	(CT) ₁₀	61	174–229	6
GS0054	F:CCCTTCTCAACGGTATGTTCTC R:GTTAGCGTGCTTTGATCCTTGT	(CT) ₉ (CA) ₁ t(AC) ₆ ...(AC) ₆ g c(AT) ₆	60	145–183	6
GS0057	F:GTGCGGTGTGGTGTATAGAAAA R:AGACATACGACAGTGCAAGAC	(TG) ₆ [CG(TG) ₂] ₂ CG(TG) ₆ CG(TG) ₅ CG	60	180–195	6
GS0060	F:CGAATCCTGATCTTGTCTTTCC R:CTGGACCAATAAGGTGCTGCT	(TGC) ₆ ...(TTG) ₃ (CTGTTG) ₃ (TTG) ₂ (CTG) ₃ t(GTG) ₅ (TTG) ₁₁	61	218–245	8
GS0066	F:TCGCGCTTACCTATTCTTCTTC R:CAGCTCTCGTAGTGCGTTATG	(CGCA) ₂ (CGACA) ₄ (CA) ₃₆	60	216–260	5
GS0078	F:CAACTCCCTCCAGTACACAAC R:AGGTTGACGATAGACCGATGAC	[CAT(CAC) ₂] ₂ CAT(CAC) ₃	61	149–173	4
GSA0051	F:TACCAACCTTGAGTTTGTGTCG R:GTCAGCATATGTGTGGTCTGT	(CGA) ₅	60	195–204	3
GSA0058	F:ATAACGCCGTAGAGATGCAAAT R:CCGTATGCACCTCCCTCTATC	(GT) ₁₀	60	196–204	4
GSA0077	F:CCATCATTGCTGACTAGGAACA R:CTCCCAACCAACTACGACTAC	(TGT) ₈	60	223–235	4

T_a annealing temperature; allelic class size range (bp), N_A number of alleles^a GenBank accession number: GS004–GS0078: JN580435–JN580447. GSA0051–GSA0077: JN627439–JN627441



- Detekce fragmentační analýzou (kapilární elfo) nebo na polyakryamidových gelech
- klasická či kapilárová Elfo
- možné problémy s odečítáním proužků
- *stutter bands* - liší se od sousedního, majoritního, o několik málo bp
 - vznik pomocí *slippage*, a poly-A aktivitě polymerázy

Mol. Tax. metody –mikrosatelity

Využití

- odlišení jedince – jedinec definován např. jako kmen s unikátním setem mikrosatelitových alel
- díky specifickým primerům lze použít na přírodním materiálu, kde je směs více druhů hub
- snadná, reprodukovatelná, mnoho variabilních alel
- stále základ většiny populačně genetických studií

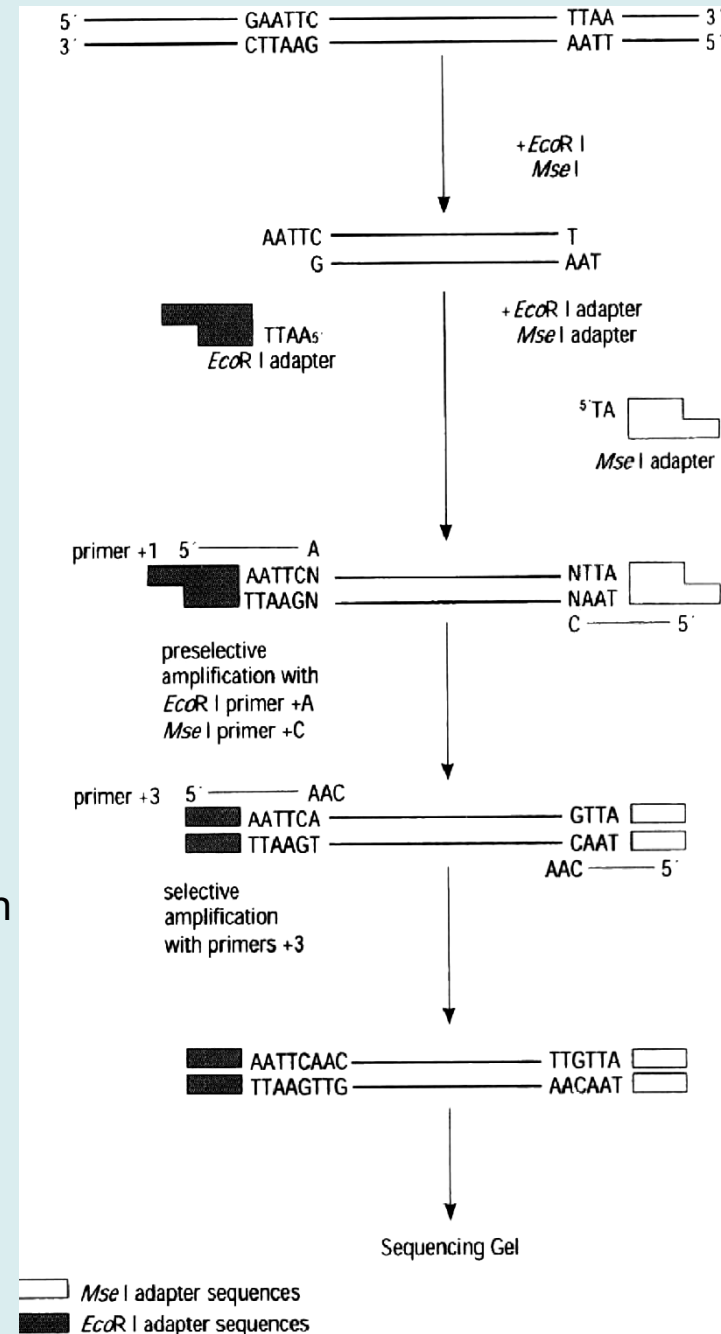
Mol. tax. metody –AFLP

Trocha historie -
mimo zkoušku

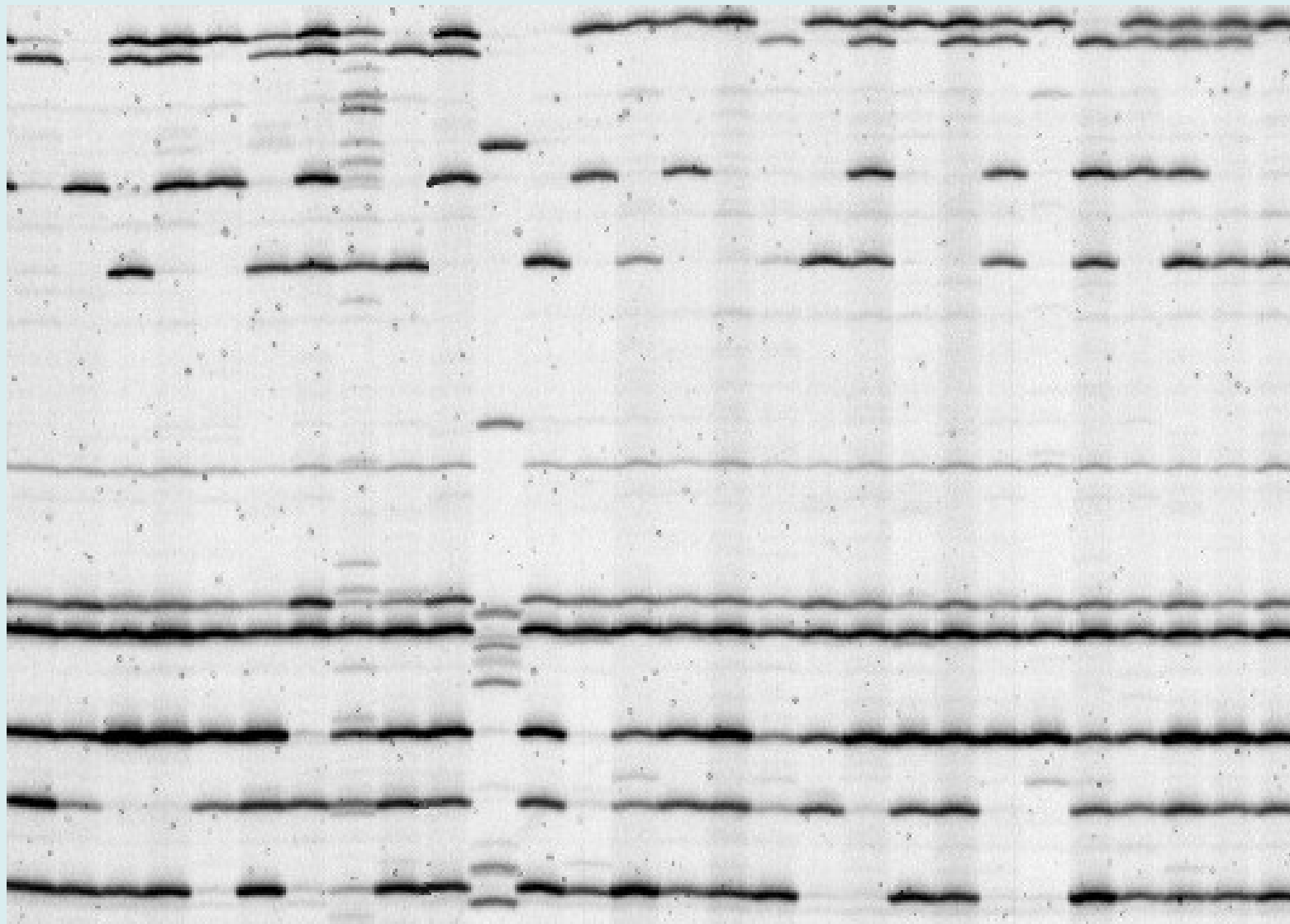
AFLP (*Amplified Fragment Length Polymorphism*)

- Tato metoda je založena na selektivní amplifikaci restrikčních fragmentů získaných štěpením pomocí dvou různých restrikčních enzymů
- Metoda multilokusová
- dobrá reprodukovatelnost
- značně náročná optimalizace
- použití na úrovni druhu a nižší
- s její pomocí bylo učiněno mnoho zásadních objevů
- v současné době se téměř nepoužívá a je nahrazena studiem SNP...

- 1) Štěpení RE pomocí dvou RE (jedna je v nadbytku)
- 2) Navázání adaptérů DNA-ligázou
- 3) Preselektivní amplifikace – primery přesahují do fragmentu
- 4) selektivní amplifikace – primery komplementární s předešlými a opět přesahují do fragmentu
- 6) Výsledkem je zmnožení fragmentů, u kterých je okrajová sekvence komplementární k selektivním primerům
- 7) Jeden z primerů značen. Vizualizace pomocí PAGE či kapilárové elfo (sekvenátor). Výsledně jsou tedy vidět jen fragmenty které jsou štípány jen jedním enzymem (na jedné straně)



AFLP ukázka polyakriamidového gelu



High-resolution DNA melt-curve analysis (HRM)

- jedna z „post-PCR“ metod
- Založeno na křivce tání (*melt curve*) dsDNA v PCR produktu
- roztok DNA se postupně ohřívá (0.1°C za 2s) a měří se úbytek dsDNA
- DNA je značena fluorescenční barvičkou, která se váže pouze na dsDNA
- Měření intenzity fluorescence se děje v PCR cykleru
- křivka tání je dána sekvencí DNA a je tedy druhově specifická
- velmi levné a „*high-throughput*“
- funguje i se směsnou DNA
- ideální na určování taxonů, kde přichází v úvahu jen omezené množství druhů

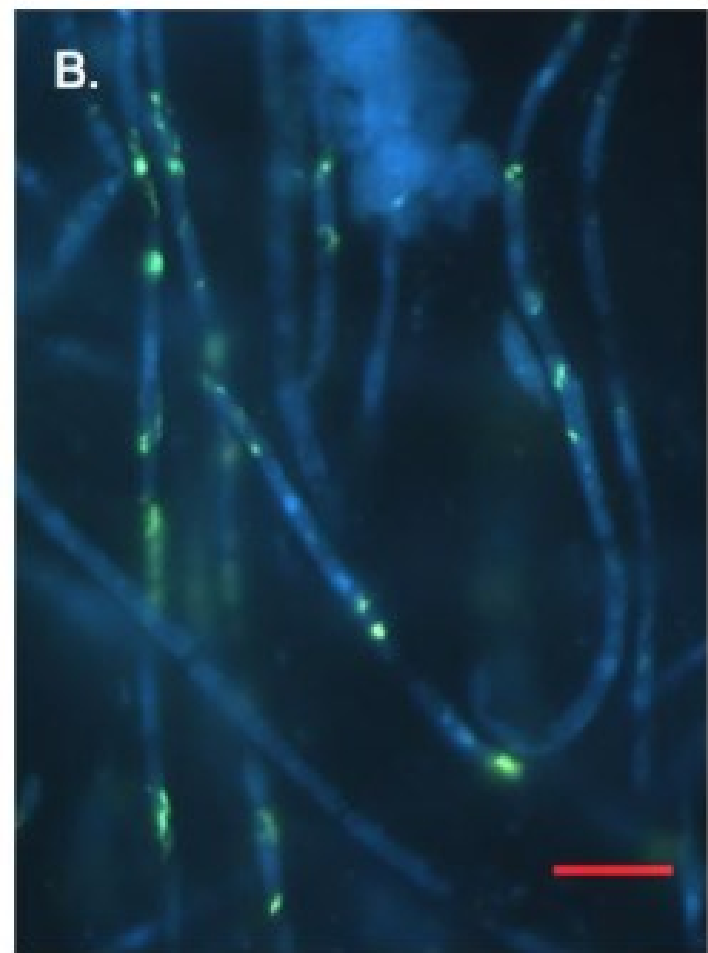
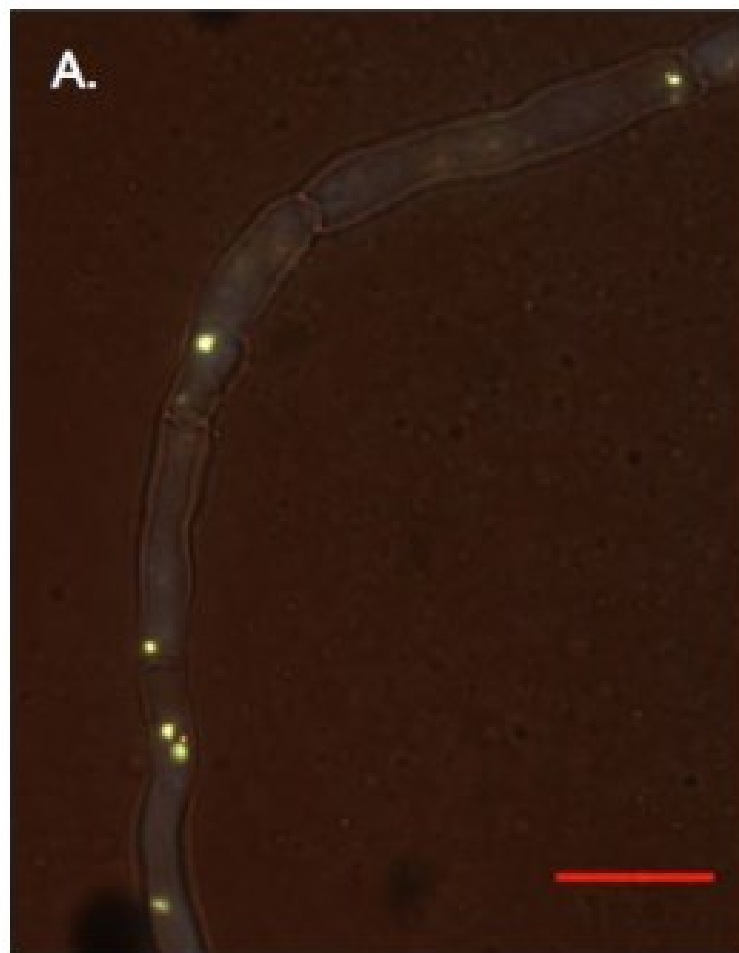


FIG. 1. Fluorescent *in situ* hybridization (FISH) microscopy showing hyphae of two isolates of endophytic fungi harboring endohyphal bacteria. Panel A (isolate 9084b; Dothideomycetes) shows the TAMRA fluorophore at the site of internal structure in hyphal cells. Panel B (isolate 9143; Sordariomycetes) shows the TAMRA fluorophore with the DAPI counterstain (blue), highlighting the fungal nuclear and mitochondrial DNA in addition to bacteria (yellow/green). Scale bar, 10 μm (A) or 25 μm (B).

- Zvolené primery nasadajú i na bakteriální DNA, nutno brát v potaz

Diverse Bacteria Inhabit Living Hyphae of Phylogenetically Diverse Fungal Endophytes^{▽†}

Michele T. Hoffman and A. Elizabeth Arnold*

Division of Plant Pathology and Microbiology, School of Plant Sciences, 1140 E. South Campus Drive,



Rhizopus microsporus
- parazit rýže

Burkholderia rhizoxinica

Rhizoxin

High-resolution DNA melt-curve analysis (HRM)

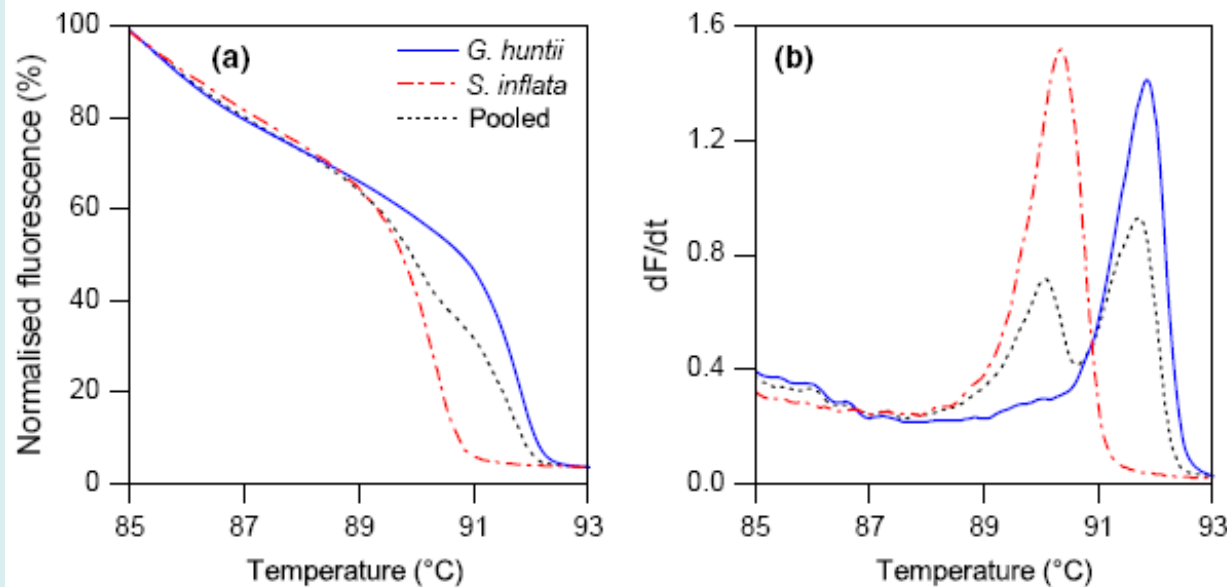


Fig. 1 Methodological validation showing accurate discrimination of mixed DNA samples using (a) aligned HRM curves and (b) derivative HRM curves. Separate pure DNA isolates of *Sporothrix inflata* and *Grosmannia huntii* compared with a single mixed DNA sample. The dF/dt represents the change in fluorescence per unit time as temperature was increased by 0.1 °C every 2 s.

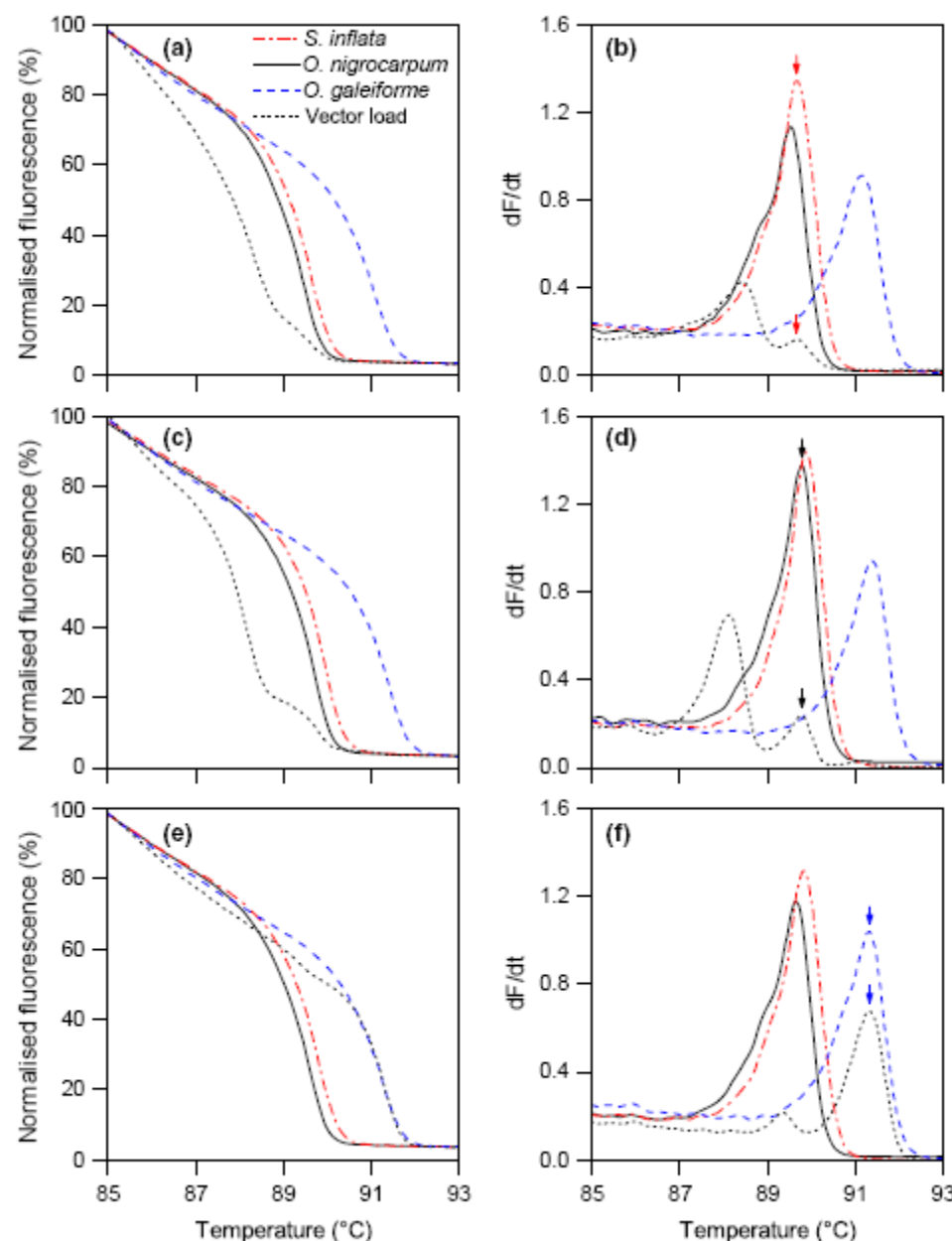
MOLECULAR ECOLOGY RESOURCES

Molecular Ecology Resources (2013) 13, 908–917

doi: 10.1111/1755-0998.12143

High-resolution DNA melt-curve analysis for cost-effective mass screening of pairwise species interactions

JAMES K. MCCARTHY,*+ RAPHAEL K. DIDHAM,*‡§ ECKEHARD G. BROCKERHOFF,+ KATHERINE A. VAN BYSTERVELDT* and ARVIND VARSANI*¶**



Studium SNP...

- *Single nucleotide polymorphism*, „snip“
- sledujeme jednotlivé mutace u velkého počtu lokusů
- genotyping
- data použitelná zejména v populační genetice, kde nahrazují mikrosatelity
- studováno pomocí NGS metod
- Např.. restriction site-associated DNA sequencing (RAD-Seq)

Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers

Nathan A. Baird¹*, Paul D. Etter¹*, Tressa S. Atwood², Mark C. Currey³, Anthony L. Shiver¹, Zachary A. Lewis¹, Eric U. Selker¹, William A. Cresko³, Eric A. Johnson^{1*}

¹ Institute of Molecular Biology, University of Oregon, Eugene, Oregon, United States of America, ² Floragenex, Eugene, Oregon, United States of America, ³ The Center for Ecology and Evolutionary Biology, University of Oregon, Eugene, Oregon, United States of America

Abstract

Single nucleotide polymorphism (SNP) discovery and genotyping are essential to genetic mapping. There remains a need for a simple, inexpensive platform that allows high-density SNP discovery and genotyping in large populations. Here we describe the sequencing of restriction-site associated DNA (RAD) tags, which identified more than 13,000 SNPs, and mapped three traits in two model organisms, using less than half the capacity of one Illumina sequencing run. We demonstrated that different marker densities can be attained by choice of restriction enzyme. Furthermore, we developed a barcoding system for sample multiplexing and fine mapped the genetic basis of lateral plate armor loss in threespine stickleback by identifying recombinant breakpoints in F_2 individuals. Barcoding also facilitated mapping of a second trait, a reduction of pelvic structure, by *in silico* re-sorting of individuals. To further demonstrate the ease of the RAD sequencing approach we identified polymorphic markers and mapped an induced mutation in *Neurospora crassa*. Sequencing of RAD markers is an integrated platform for SNP discovery and genotyping. This approach should be widely applicable to genetic mapping in a variety of organisms.

Mol. Tax. metody – sekvence DNA

Co sekvenovat?

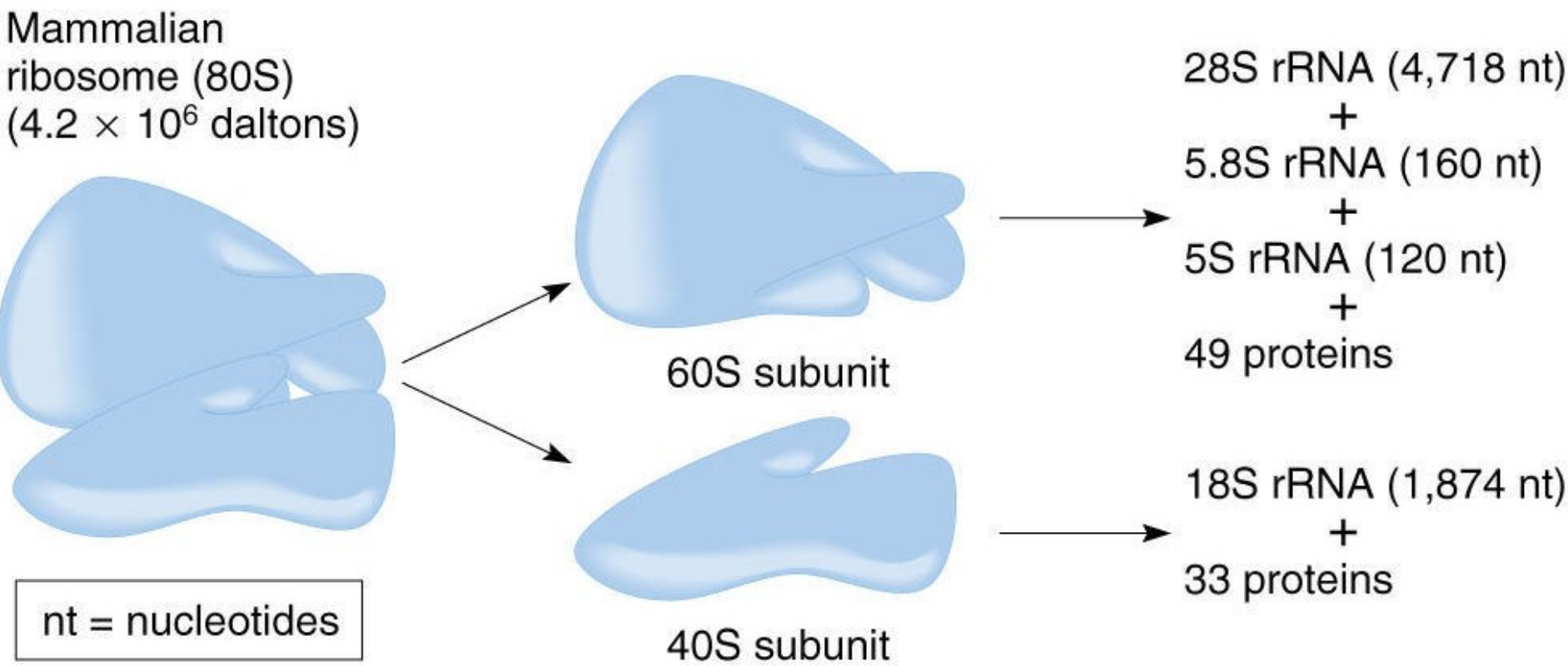
- vybráno několik desítek regionů (nejčastěji genů)
- další geny (i anonymní) lze získat z genomických sekvencí (možnost vybrat geny rovnoměrně rozložené v genomu)
- tyto geny se musí nacházet u všech porovnávaných organismů
- jejich translační produkty nesmí během evoluce měnit svojí funkci (aby nepodléhaly změněnému selekčnímu tlaku)
- „multivopy“ (desítky až stovky kopií) (nebezpečí paralogních genů z jiných lokusů)
- „low copy“ (desítky) či „singlecopy“ geny (v většina proteinů)
- počty kopií (i lokusů) jednotlivých genů se liší mezi druhy (např. ITS, TEF-1 α)
- problémy s amplifikací u různých druhů hub (absence univerzálních primerů)
- známy různě variabilní geny – použití dle taxonomické úrovně

Mol. Tax. metody – sekvence DNA

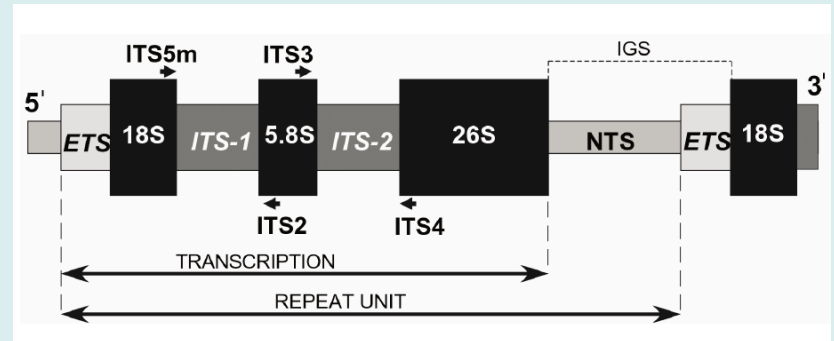
rDNA – multi copy gen

- rDNA sa vyskytuje v genomu repetitívne, 7,7-24 kb, (185 kopií rDNA u *Neurospora crassa*, 45 u *Aspergillus nidulans*), v jednom (*Sach. cer.*) či dvou klastrech (*Schi. pomp.*) Variabilní počet kopií mezi kmeny *Kluyveromyces lactis* – kmeny s malým počtem rostou pomalu.
- skládá se jak z konzervativních, tak z variabilních podjednotek, což umožňuje použití na různých taxonomických úrovních
- nejmenší podjednotka **5S** (délka cca 120 bp) je u hub většinou lokalizovaná mimo rDNA (až na některé bazidiomycety), v současné době se nepoužívá
- konzervativní podjednotky – exony:
 - **5,8S** (cca 160 bp)
 - **16S-18S (SSU rDNA** - small subunit, cca 1800 bp)
 - **25S-28S (LSU rDNA** - large subunit, cca 3400 bp)

Termín LSU označuje jak celou 60S (tj. 28S, 5.8S, 5S rDNA) tak jen 28S rDNA
Zatímco SSU se vztahuje jen na 18S rDNA



Mol. Tax. metody – sekvence DNA



Využití rDNA

16S-18S (SSU rDNA) - porovnávání na úrovni řádu a vyšší

25S-28S (LSU rDNA)

konzervativní gen, obsahuje i tzv. divergentní domény, kódující oblasti RNA (např. smyčky) které nemají takový dopad na funkčnost molekul zejména její doména D2, slouží k fylogenetickým analýzám na úrovních nižších řádu. Sedimentační konstanta 25 či 26S milně udávána u kvasinek a basidiomycetů. Správněji jen jako 28S.

5,8S - úroveň řádu a nižší, studována spolu s ITS

- Nefunkční, nepřepisované sekvence oddělující jednotlivé kopie

IGS (intergenetic spacer) – velmi variabilní, použití v pop. genetice

NTS (nontranscribed spacer)

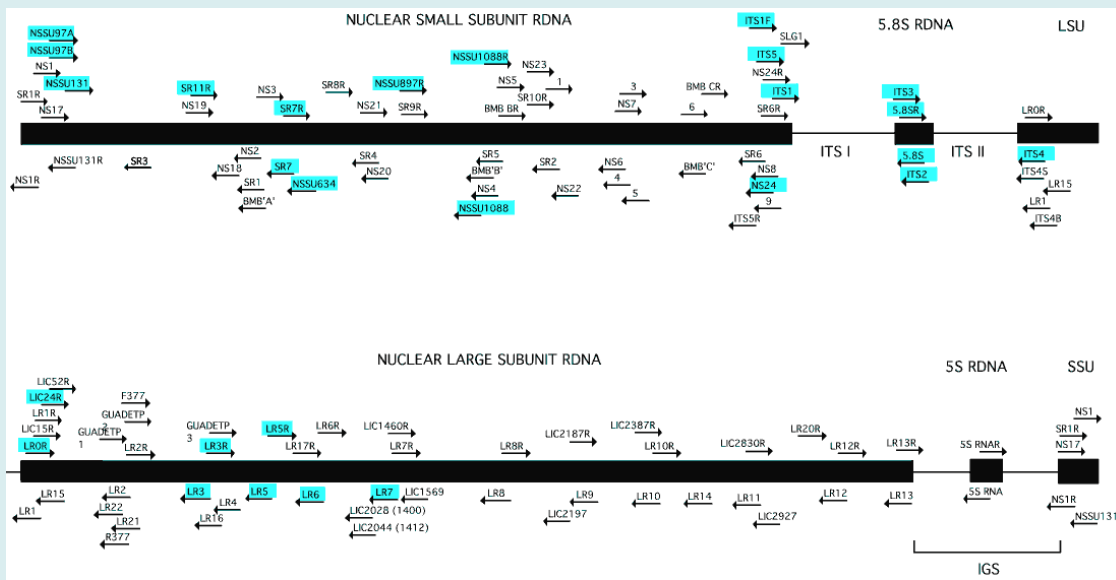
mt rDNA – mtSSU, LSU rDNA velmi variabilní. mtLSU rDNA k odlišení jedinců

Glomus intraradices – sledování efektivity inokulace

Mol. Tax. metody – sekvence DNA

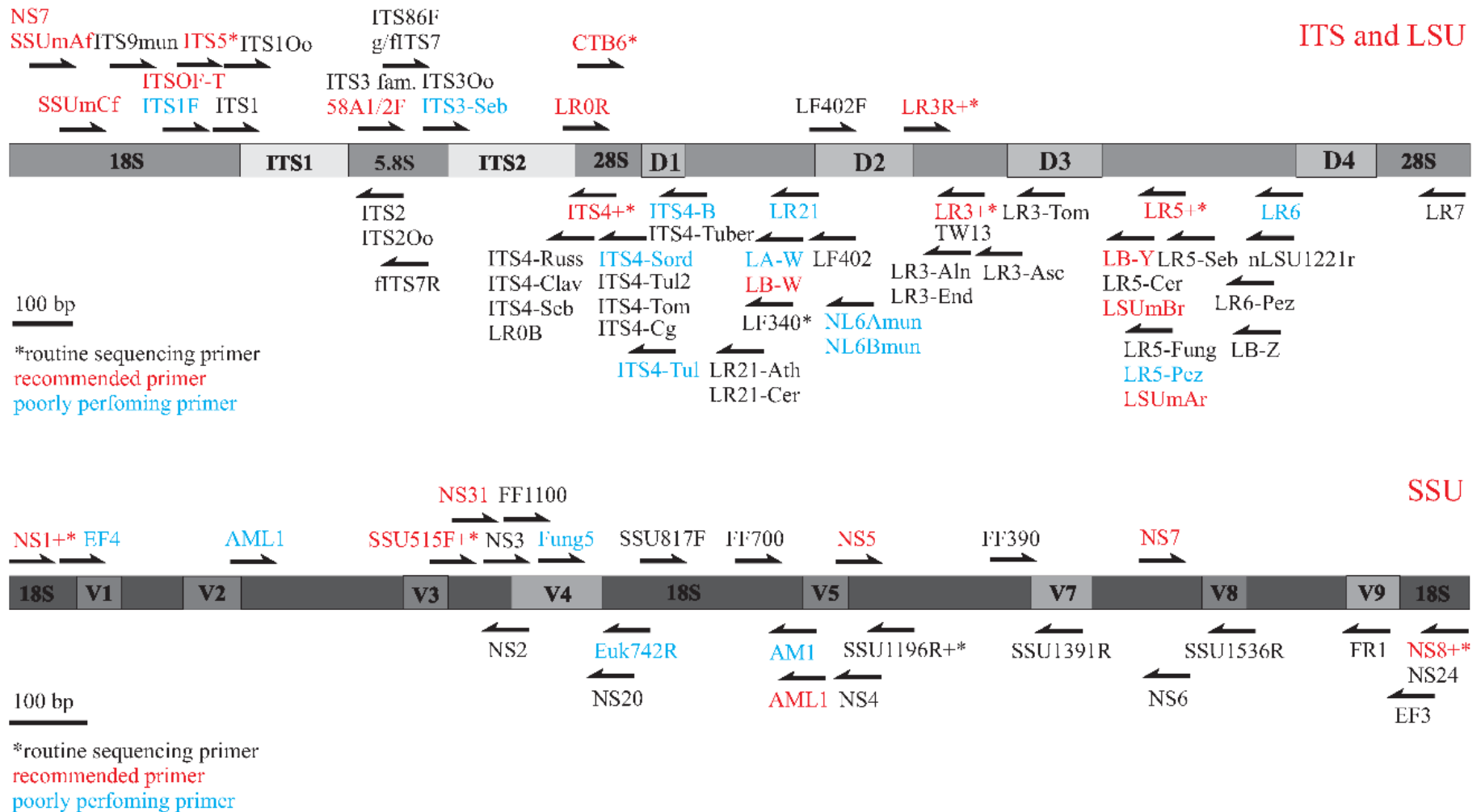
Univerzální houbové primery

- Díky konzervativním úsekům lze najít místa pro primery společná pro všechny houby
- Existují i specifické primery pro různé houbové skupiny
- Specifické primery navržené pro jednotlivé druhy (či populace) využívány v diagnostice

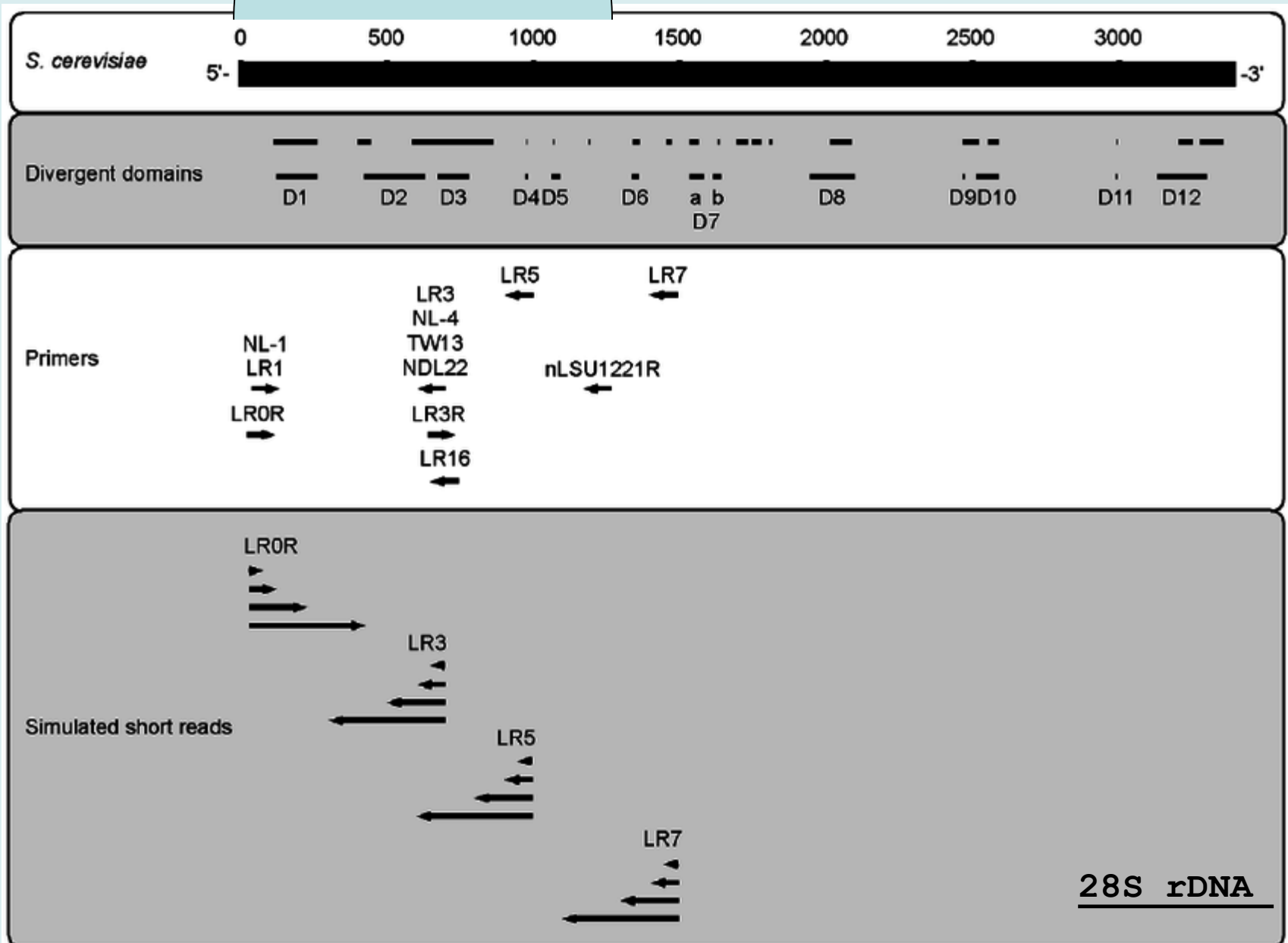


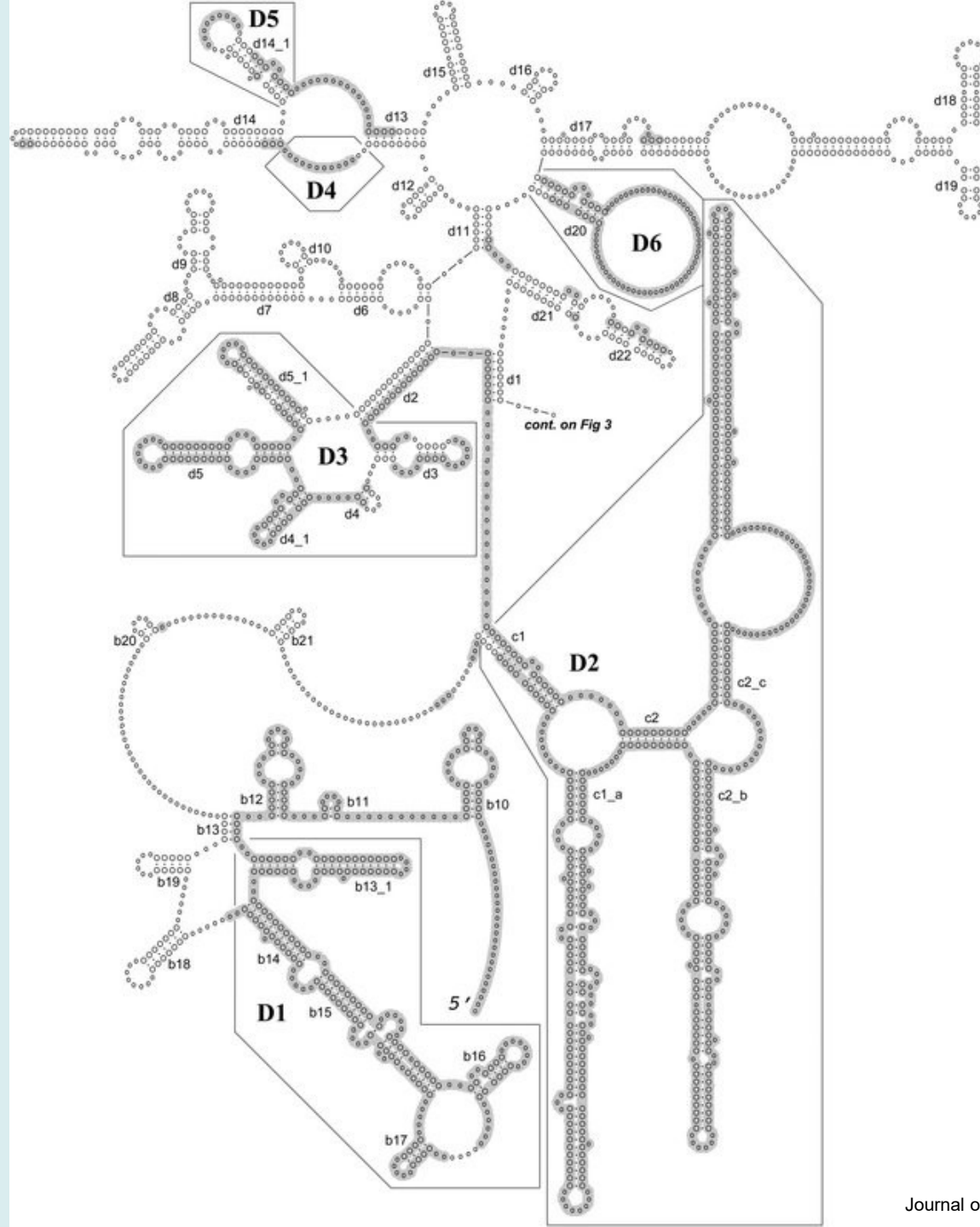
<http://aftol.biology.duke.edu/public/disclaimer>

- patrně nejaktuálnější přehled univerzálních i specifických houbových rDNA primerů je na Unite <https://unite.ut.ee/>
- Primery specifické na určité skupiny (např. basidiomycota)
- Sekvenační vs PCR amplifikační primery
- Amplifikační primery nemusí být vhodné na sekvenování

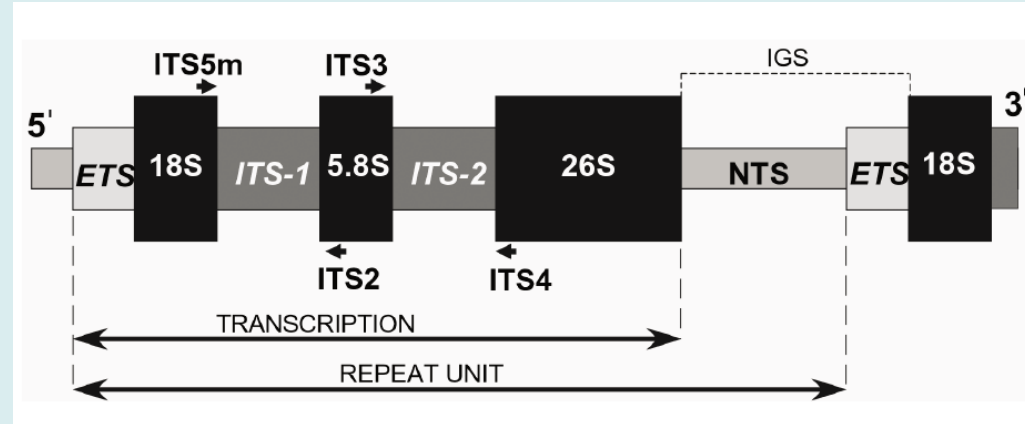


Nejčastěji se studuje D1-D2 doména, kvalitnější studie D1-D5





Mol. Tax. metody – sekvence DNA



rDNA

- Funkční podjednotky oddělují mezerníky - *spacery*

ITS 1 a ITS 2 (internal transcribed spacer)

ETS (external transcribed spacer)

ITS1 + 5,8S + ITS2 nazýván **ITS region či oblast** (nejčastěji 450-550 bp, výjimečně 250-1000 bp)

- ITS 1,2 nejsou zcela nefunkční, krátké oblasti pro vystřížení podjednotek podléhají selekci
- ITS region je nejstudovanější marker – důležitý nástroj na určení neznámé houby (spolu s LSU u basidiomycetů)

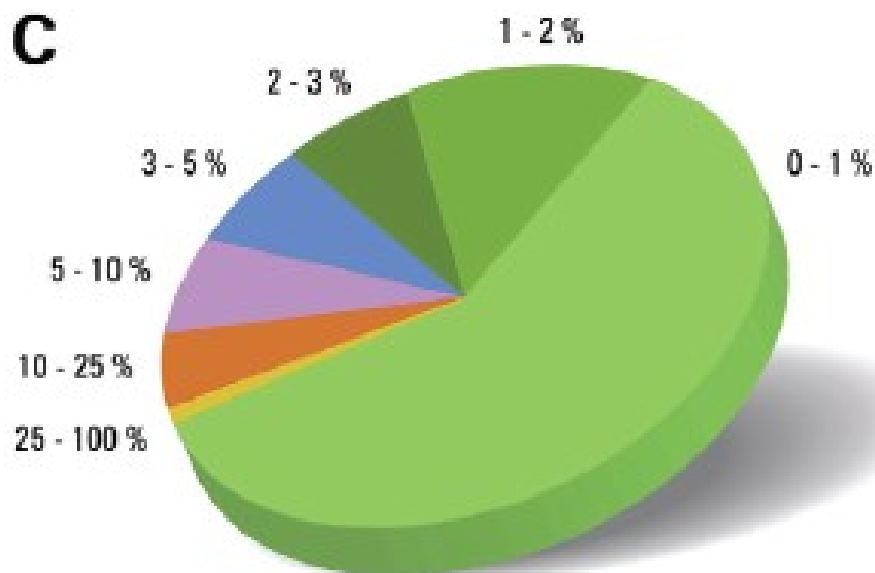
Vnitrodruhová variabilita se liší dle skupiny hub

% SD

Weighted intraspecific *ITS* variability in the kingdom *Fungi*
Weighted intraspecific *ITS* variability for the five conceptual
phyla of the kingdom

<i>Ascomycota</i>	(2509 species)	1.96% [3.73]
<i>Basidiomycota</i>	(1582 species)	3.33% [5.62]
<i>Chytridiomycota</i>	(11 species)	5.63% [10.49]
<i>Glomeromycota</i>	(23 species)	7.46% [4.14]
<i>Zygomycota</i>	(60 species)	3.24% [6.12]

C



- Použito 4200 druhů hub
- 172000 sekvencí v GenBank (56% určeno do druhu, 15500 druhů)

Intraspecific *ITS* Variability in the Kingdom *Fungi* as Expressed in the International Sequence Databases and Its Implications for Molecular Species Identification

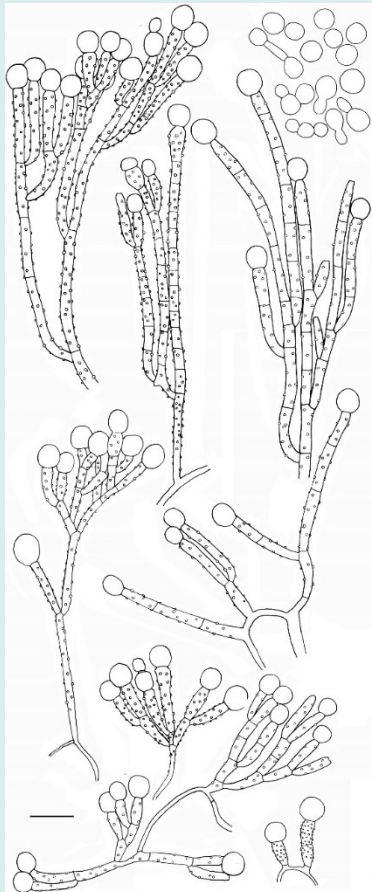
R. Henrik Nilsson¹, Erik Kristiansson², Martin Ryberg¹, Nils Hallenberg¹ and Karl-Henrik Larsson¹

Figure 1. The proportion of fungal species in this study having intraspecific variabilities in the ranges depicted of (A) the *ITS1* region, (B) the *ITS2* region, and (C) the *ITS1*, *5.8S*, and *ITS2* regions combined.

Vnitrodruhová variabilita se liší dle skupiny hub

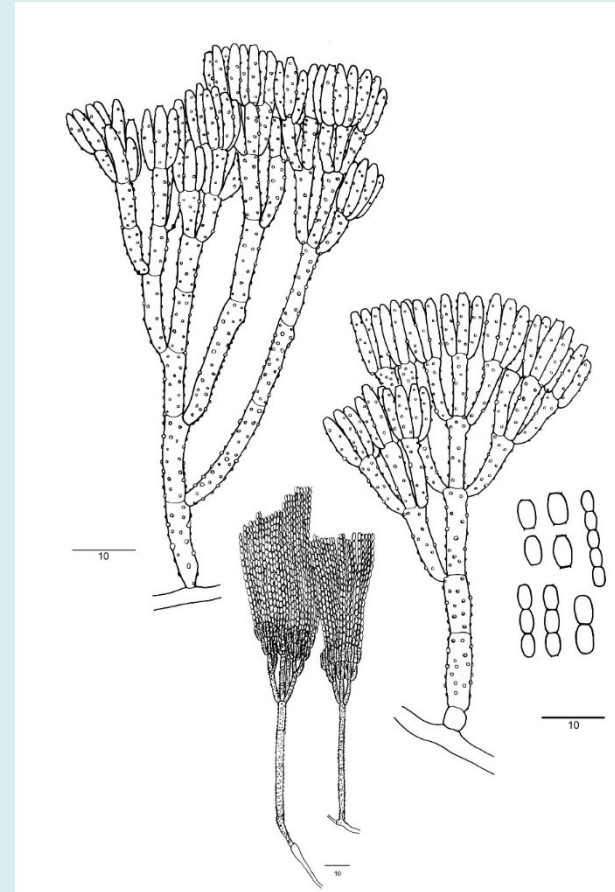
- řada druhů Ascomycota nejde odlišit pomocí ITS rDNA
- *Aspergillus* (např. *A. niger*), *Penicillium* (podrod *Penicillium*), *Phoma*, *Trichoderma*, *Cladosporium*....
- naopak u Basidiomycota ITS rDNA odlišuje druhy spolehlivě (?)

Geosmithia sp. 8, phleom feeding bark beetles



ITS-LSU rDNA	0%
RPB2	0.1 %
β -tubulin	1.5%
TEF1 α	1.5%
IGS rDNA	0.6%

G. microcorthyli
ambrosia fungus



Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for *Fungi*

Conrad L. Schoch^{a,1}, Keith A. Seifert^{b,1}, Sabine Huhndorf^c, Vincent Robert^d, John L. Spouge^a, C. André Levesque^b, Wen Chen^b, and Fungal Barcoding Consortium^{a,2}

^aNational Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20892; ^bBiodiversity (Mycology and Microbiology), Agriculture and Agri-Food Canada, Ottawa, ON, Canada K1A 0C6; ^cDepartment of Botany, The Field Museum, Chicago, IL 60605; and ^dCentraalbureau voor Schimmelcultures Fungal Biodiversity Centre (CBS-KNAW), 3508 AD, Utrecht, The Netherlands

Edited* by Daniel H. Janzen, University of Pennsylvania, Philadelphia, PA, and approved February 24, 2012 (received for review October 18, 2011)

- Cytochrom c oxidáza 1 (univerzální barcode), LSU rDNA, SSU rDNA, RPB1 (rybozomální polymeráza), MCM7 (minichromosome maintenance protein)
- **ITS rDNA**
 - největší rozdíl mezi vnitro a mezidruhovou variabilitou (barcode gap)
 - Univerzální PCR primery – velký úspěšnost PCR

available online at www.studiesinmycology.org

STUDIES IN MYCOLOGY 92: 135–154 (2019).

Large-scale generation and analysis of filamentous fungal DNA barcodes boosts coverage for kingdom fungi and reveals thresholds for fungal species and higher taxon delimitation

D. Vu^{1*}, M. Groenewald¹, M. de Vries¹, T. Gehrman¹, B. Stielow¹, U. Eberhardt², A. Al-Hatmi¹, J.Z. Groenewald¹, G. Cardinali³, J. Houbraken¹, T. Boekhout^{1,4}, P.W. Crous^{1,5,6}, V. Robert¹, and G.J.M. Verkley^{1*}

¹Westerdijk Fungal Biodiversity Institute, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands; ²Staatliches Museum f. Naturkunde Stuttgart, Abt. Botanik, Rosenstein 1, D-70191 Stuttgart, Germany; ³University of Perugia, Dept. of Pharmaceutical Sciences, Via Borgo 20 Giugno 74, I 06121 Perugia, Italy; ⁴Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, The Netherlands; ⁵Wageningen University and Research Centre (WUR), Laboratory of Phytopathology, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands; ⁶Department of Genetics, Biochemistry and Microbiology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria 0028, South Africa

Intraspecies variation

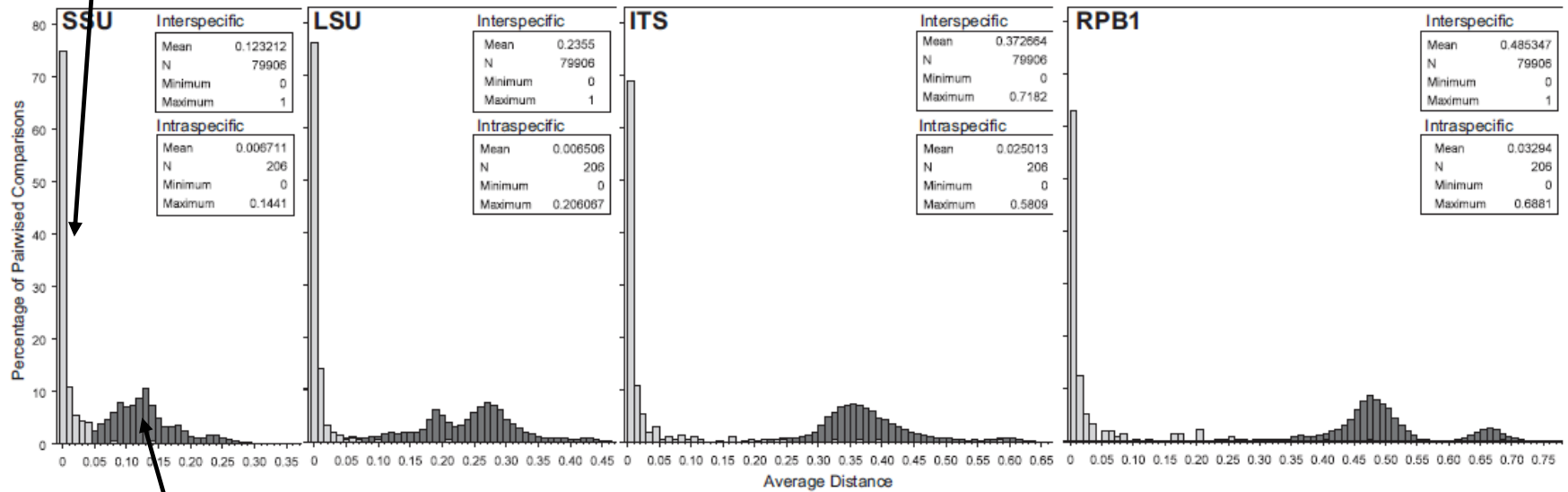
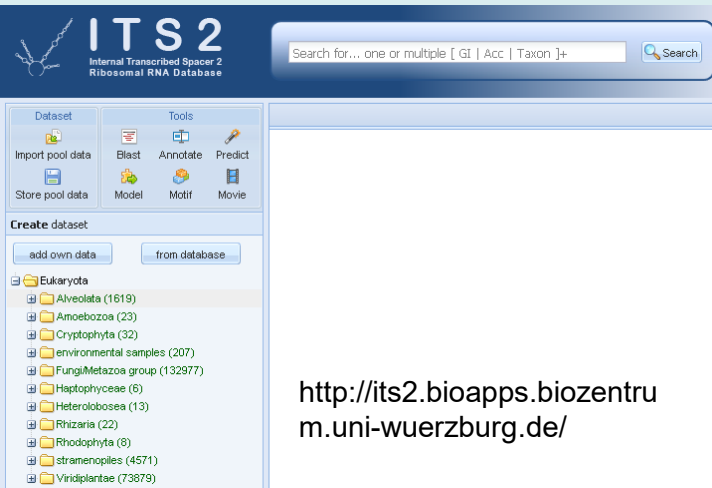


Fig. 3. Barcode gap analyses using distance histograms for each marker. Histograms display intraspecific variation in light gray and interspecific variation in dark gray. Inserts summarize distance data.

Interspecies variation

Využití sekundární struktury rRNA

- složeno ze smyček (*loops*) a stonků (*stems*)
- modelováno hlavně u ITS2 (univerzální 4 stems)
- *compensatory base changes* (CBC) – změny nenarušující stonek



The screenshot shows the ITS2 database website. At the top, there's a search bar with the text "Search for... one or multiple [GI | Acc | Taxon]+" and a "Search" button. Below the search bar, there's a "Dataset" section with icons for "Import pool data", "Store pool data", "Blast", "Model", "Annotate", "Motif", "Predict", and "Movie". To the left, there's a "Create dataset" section with "add own data" and "from database" buttons. Below that, there's a list of taxonomic groups under "Eukaryota": Alveolata (1619), Amoebozoa (23), Cryptophyta (32), environmental samples (207), Fungi/Metazoa group (132977), Haptophyceae (6), Heterokontophyta (13), Rhizaria (22), Rhodophyta (8), stramenopiles (4571), and Viridiplantae (73679). At the bottom, the URL <http://its2.bioapps.biozentrum.uni-wuerzburg.de/> is displayed.

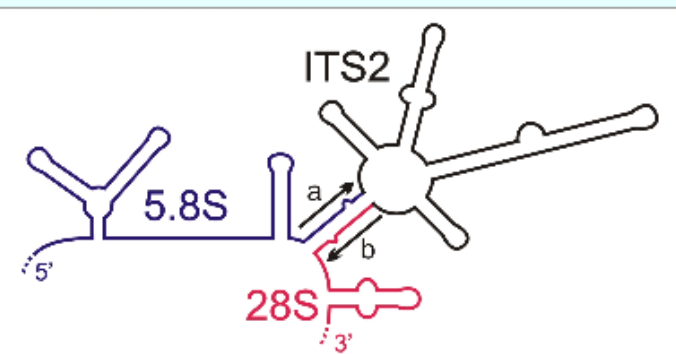
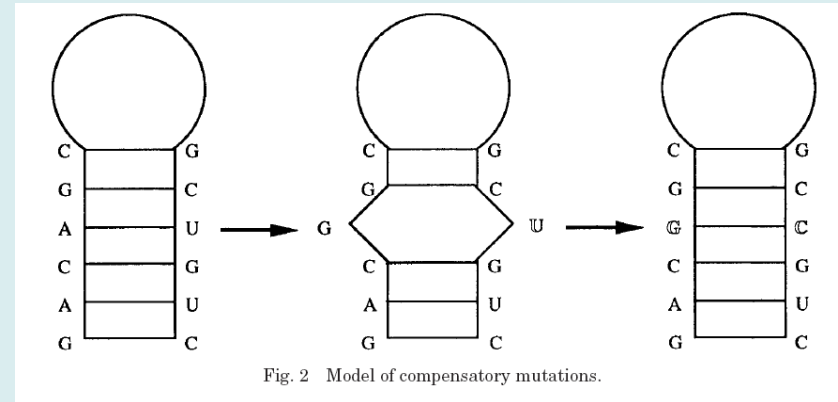


Fig. 1: Secondary structure of the internal transcribed spacer 2 (ITS2) with tailing parts of the 28S and 5.8S rRNA. This annotation tool searches for the end of the 5.8S and start of the 28S indicated by arrows (a and b, respectively). These parts are known to hybridize during ribogenesis before ITS2 cleavage.



CBCAnalyzer - Inferring phylogenies based on compensatory base changes (CBCs)

Matthias Wolf, Joachim Friedrich, Benjamin Ruderisch, Thomas Dandekar, and Tobias Müller

Department of Bioinformatics, Biocenter, University of Würzburg, Germany

Home

Download

Installation

Quickstart

Contact

- jedinci jednoho druhu by se neměly lišit v CBC
- **přítomnost CBC je náznakem biologické izolace druhu** (93% z 1300 druhů)
- nicméně nepřítomnost CBS není informativní (může jít o druhy i nemusí)
- samotné mapování přítomnosti různých CBC mezi druhy lze použít při fylogenezi

Distinguishing species

TOBIAS MÜLLER,¹ NICOLE PHILIPPI,¹ THOMAS DANDEKAR, JÖRG SCHULTZ, and MATTHIAS WOLF

Department of Bioinformatics, Biocenter, University of Würzburg, Am Hubland, D-97074 Würzburg, Germany

ABSTRACT

Given two organisms, how can one distinguish whether they belong to the same species or not? This might be straightforward for two divergent organisms, but can be extremely difficult and laborious for closely related ones. A molecular marker giving a clear distinction would therefore be of immense benefit. The internal transcribed spacer 2 (ITS2) has been widely used for low-level phylogenetic analyses. Case studies revealed that a compensatory base change (CBC) in the helix II or helix III ITS2 secondary structure between two organisms correlated with sexual incompatibility. We analyzed more than 1300 closely related species to test whether this correlation is generally applicable. In 93%, where a CBC was found between organisms classified within the same genus, they belong to different species. Thus, a CBC in an ITS2 sequence-structure alignment is a sufficient condition to distinguish even closely related species.

Počet rDNA kopií u hub (45-200 repetice)

druh	Počet rDNA kopií
Cochliobolus heterostrophus	90
Cochliobolus heterostrophus	130
Melampsora sp.	200
Colletotrichum acutatum	20
Mucor miehei	100
Phaffia rhodozyma	60 (na 3 chromozomech)
Candida spp., Kluyveromyces, Schizosaccharomyces, Torulaspora	10-100 (cca), na 1-2 chromozomech
Aspergillus fumigatus	38 – 91 (dle použitého kmenu)
Kluyveromyces lactis	10
Neurospora crassa	185

Mol. Tax. metody – sekvence DNA

- u všech *multicopy* genů možný polymorfismus v rámci jednoho jádra– jak u diploidů tak haploidů
- homogenita v sekvencích v rámci genomu udržována pomocí *concerted evolution* a genové konverze pomocí homologní rekombinace– známo u ITS rDNA
- u hub dokázáno u *Aspergillus nidulans*
- Možnost sekvenovat jednotlivé kopie (klonování, masivní paralelní sekvenování)

Ganley ARD, Kobayashi T. 2007. Highly efficient concerted evolution in the ribosomal DNA repeats: total rDNA repeat variation revealed by whole-genome shotgun sequence data. Gen Res. 17:184–191.

Intragenomic Variation of Fungal Ribosomal Genes Is Higher than Previously Thought

Uwe K. Simon* and Michael Weiß†

Mol. Biol. Evol. 25(11):2251–2254. 2008

*Lehrstuhl für Pharmazeutische Biologie, Universität Würzburg, Würzburg, Germany; and †Lehrstuhl Spezielle Botanik und Mykologie, Universität Tübingen, Tübingen, Germany

Table 2
Percentage of Polymorphisms (p = numbers of polymorphic sites in relation to fragment length) in the Nuclear Ribosomal DNA (complete SSU, complete ITS1-5.8S-ITS2, D1–D3 region of LSU) of *Davidiella tassiana*, *Mycosphaerella punctiformis*, *Phoma exigua* var. *exigua*, and *Teratosphaeria microspora*

	<i>P. exigua</i> var. <i>exigua</i>	<i>M.</i> <i>punctiformis</i>	<i>T.</i> <i>microspora</i>	<i>D.</i> <i>tassiana</i>	<i>Aspergillus</i> <i>nidulans</i>
SSU	2.9	2.6	2.0	3.7	n.d.
ITS	2.2	3.6	2.2	3.0	n.d.
LSU	3.0	2.8	1.7	3.6	1.7

n.d., not determined.

Table 1
Origin of Fungal Material and NCBI GenBank Accession Numbers of Majority Rule Consensus Sequences Obtained in an Earlier Study (Simon UK, Groenewald J, Crous P, unpublished)

Species	CBS Accession Number	GenBank Accession Number
<i>Davidiella tassiana</i> (C)	CBS 723.79	EU167558
<i>Mycosphaerella punctiformis</i> (C)	CBS 113265	EU167569
<i>Phoma exigua</i> var. <i>exigua</i> (P)	CBS 118.94	EU167567
<i>Teratosphaeria microspora</i> (C)	CBS 101951	EU167572

C, Capnodiales; P, Pleosporales.

- nalezené sekvence z jenoho kmene odpovídají sekvencím z více kmenů z GenBank

TABLE I. Number of ITS sequence variants obtained from 1399 clones derived from 19 isolates of six *Laetiporus* species and two isolates of *Wolfeiporia dilatohypha*

Species	Clade designation from Banik et al. 2010	Isolate	Number of clones sequenced	Number of sequence variants ^a	Percentage of clones within a given percent similarity of the consensus ^b		
					98–100% similar to consensus	95–98% similar to consensus	< 95% similar to consensus
<i>L. cincinnatus</i>	Clade K	DA-37	96	33	86.5% (83)	1% (1)	12.5% (12)
<i>L. cincinnatus</i>	Clade K	DA-37ss-1 ^c	24	15	75% (18)	4% (1)	21% (5)
<i>L. cincinnatus</i>	Clade K	DA-37ss-2 ^c	23	7	95.7% (22)	0%	4.3% (1)
<i>L. cincinnatus</i>	Clade K	DA-37ss-3 ^c	24	11	75% (18)	0%	25% (6)
<i>L. cincinnatus</i>	Clade K	DA-37ss-4 ^c	23	9	100% (23)	0%	0%
<i>L. cincinnatus</i>	Clade K	MAS-1	86	33	79.1% (68)	12.8% (11)	8.1% (7)
<i>L. cincinnatus</i>	Clade K	MO-4	93	39	89.3% (83)	2.4% (2)	8.6% (8)
<i>L. cincinnatus</i>	Clade K	IL-51	94	23	95.7% (90)	0%	4.3% (4)
<i>L. conifericola</i>	Clade B	NV-2	92	11	100% (92)	0%	0%
<i>L. conifericola</i>	Clade B	TR-5	2	1	100% (2)	0%	0%
<i>L. gilbertsonii</i>	Clade F	TJV-2000-101	87	25	100% (87)	0%	0%
<i>L. gilbertsonii</i>	Clade F	CA-7	94	22	100% (94)	0%	0%
<i>L. huroniensis</i>	Clade A1	MI-5	29	3	100% (29)	0%	0%
<i>L. sp. nov.</i>	Clade J	GDL-1	95	17	100% (95)	0%	0%
<i>L. sp. nov.</i>	Clade J	PR6521	91	25	97.8% (89)	0%	2.2% (2)
<i>L. sulphureus</i>	Clade E	CT-1	62	18	100% (62)	0%	0%
<i>L. sulphureus</i>	Clade E	NAMA-2	95	22	100% (95)	0%	0%
<i>L. sulphureus</i>	Clade E	DA-11	92	18	100% (92)	0%	0%
<i>L. sulphureus</i>	Clade E	MAS-2	108	32	100% (108)	0%	0%
<i>W. dilatohypha</i>		CS-63	45	11	100% (45)	0%	0%
<i>W. dilatohypha</i>		FP72162	44	19	98% (43)	2% (1)	0%

^aNumber of variants based on 100% sequence similarity.
^bPercentage is given first, followed by the raw number of sequences in parentheses.
^cSingle-spore isolate derived from the fruiting body of DA-37.

Intragenomic variation in the ITS rDNA region obscures
phylogenetic relationships and inflates estimates of
operational taxonomic units in genus *Laetiporus*

Daniel L. Lindner¹ spacer region, intragenomic variation, molecular
Mark T. Banik drive, sulfur shelf
*US Forest Service, Northern Research Station, Center for
Forest Mycology Research, One Gifford Pinchot Drive,
Madison, Wisconsin 53726*

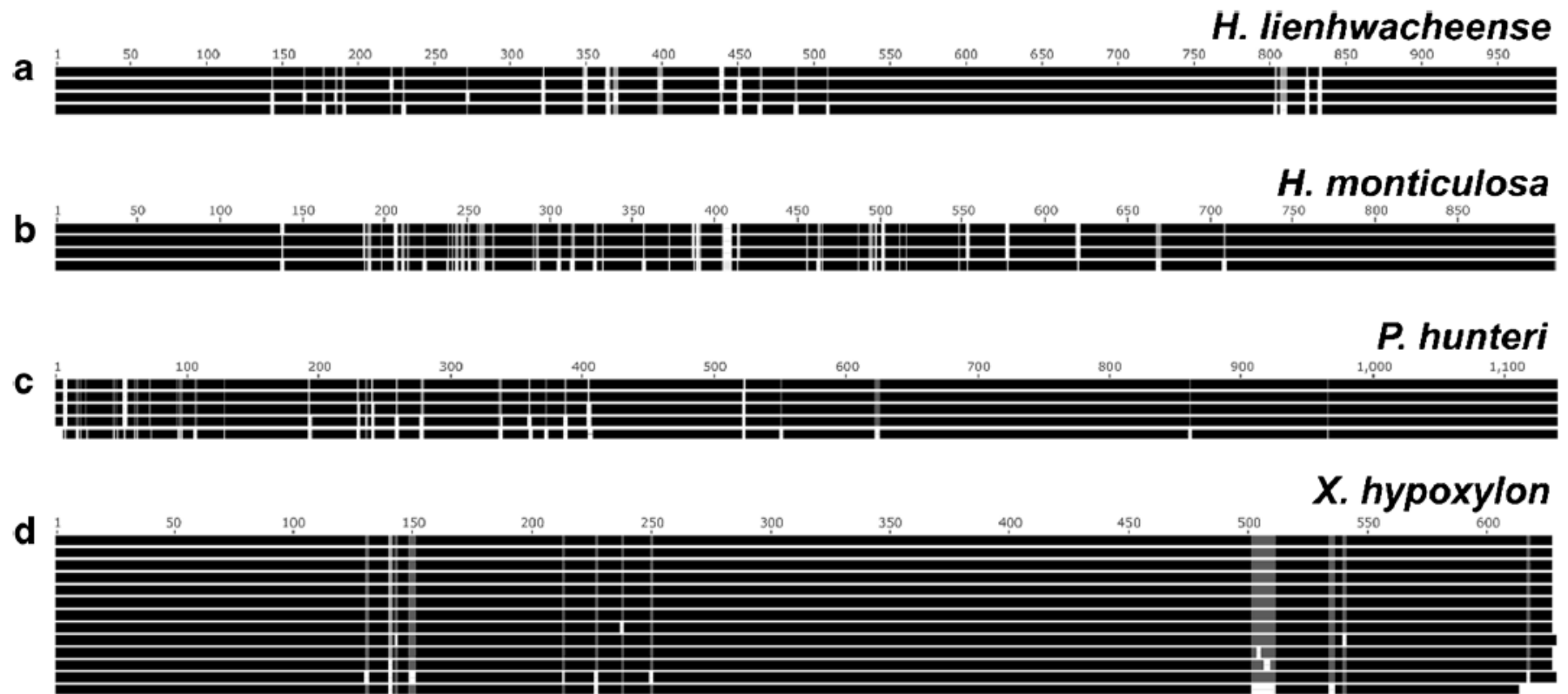


Fig. 2 Intragenomic polymorphism between the ITS regions of the identified and aligned nrDNA copies of different Xylariales species (**a** *Hypoxylon lienhwacheense* MFLUCC 14-1231, **b** *Hypomontagnella monticulosa* MUCL 54604, **c** *Pyrenopolyporus hunteri* MUCL 49339, **d** *Xylaria hypoxylon* CBS 122620). The shown region spans the partial

18S rDNA, ITS1, 5.8S rDNA, ITS2, and partial 28S rDNA. The fragment sizes are indicated by numbers. Identical regions are shown in black, and variable areas are highlighted in dark or light gray. Gaps are depicted in white

- Rozložení mutací není rovnoměrné

Intragenomic polymorphisms in the ITS region of high-quality genomes of the Hypoxylaceae (Xylariales, Ascomycota)

Marc Stadler¹ • Christopher Lambert¹ • Daniel Wibberg² • Jörn Kalinowski² • Russell J. Cox³ • Miroslav Kolařík^{1,2,3,4} • Eric Kuhnert³

Check for updates

Table 1 Details of the genome sequences generated for the selected Xylariales. O/I Oxford nanopore/Illumina, PB – PacBio. ET - ex epitype strain; HT – ex holotype strain.

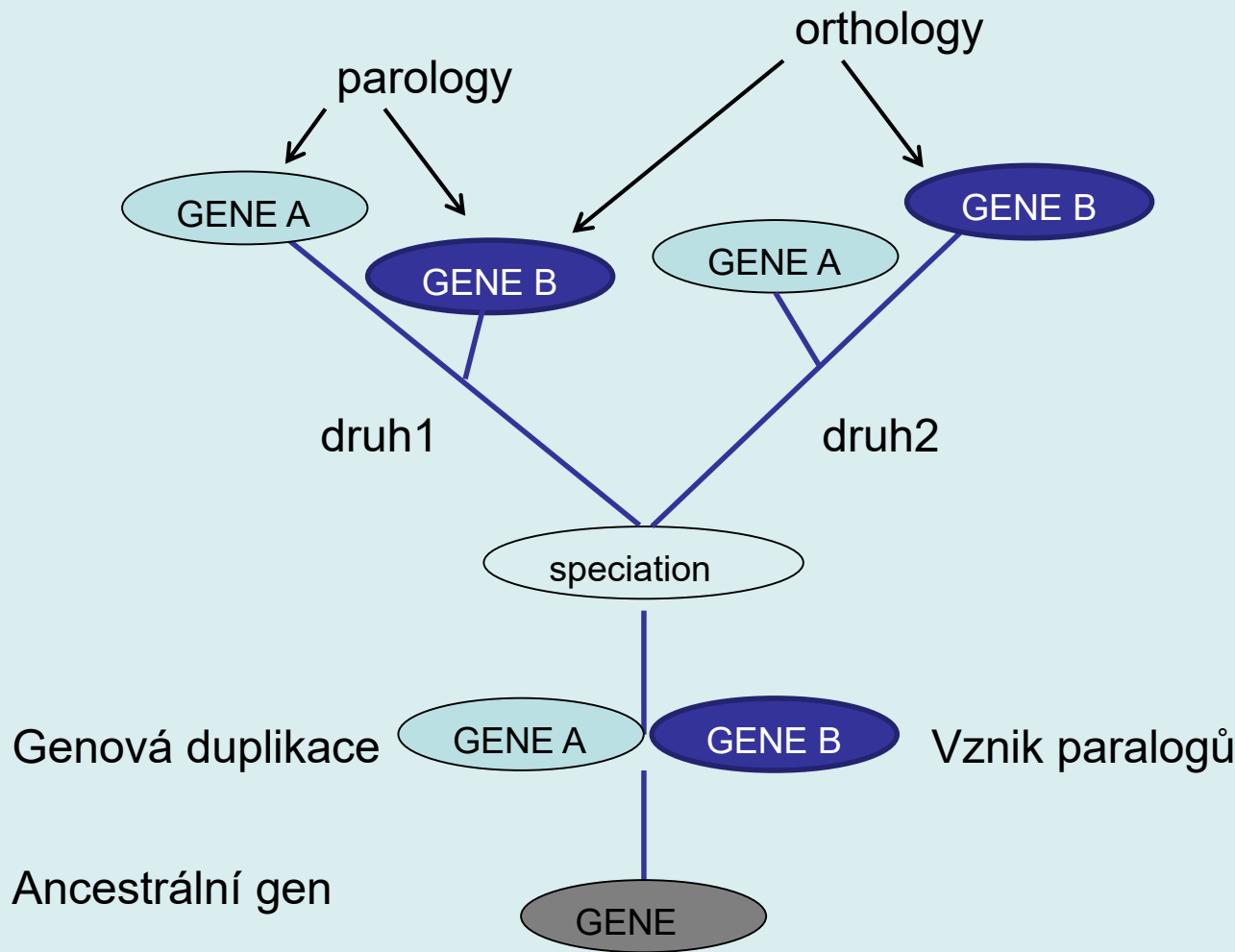
Organism	Strain	Reference	Genome size/ sequencing method [#]	# of ITS copies	Maximum deviation (# of bp)	% homology divergence
<i>Annulohypoxyton truncatum</i> (ET)	CBS 140778	Kuhnert et al. 2017	38,511,861 O/I	3	3	99.5
<i>Daldinia concentrica</i>	CBS 113277	Stadler et al. 2014b	37,605,921 O/I	3	1	99.8
<i>Entonaema liquescens</i> *	ATCC 46302	Triebel et al. 2005; Rogers 1982	39,197,785 O/I	4	4	99.3
<i>Hypomontagnella monticulosa</i>	MUCL 54604	Lambert et al. 2019	42,889,121 O/I	4	64	87
<i>Hypomontagnella spongiphila</i> (sp. nov. ined. *)	CLL-205	Leman-Loubiere et al. 2017	42,173,915 O/I	5	3	99.8
<i>Hypoxyton fragiforme</i> (ET)	MUCL 51264	Wendt et al. 2018	38,198,373 PB	19	4	99.4
<i>Hypoxyton lienhwacheense</i>	MFLUCC 14-1231	Sir et al. 2016; Sir et al. 2019	35,785,595 O/I	4	23	97.7
<i>Hypoxyton pulicicidum</i> (HT)	ATCC 74245	Bills et al. 2012	43,543,700 O/I	4	16	97.4
<i>Hypoxyton rickii</i> (ET)	MUCL 53309	Kuhnert et al. 2014	41,846,710 O/I	4	23	95.8
<i>Jackrogersella multiformis</i> (ET)	CBS 119016	Wendt et al. 2018	38,501,162 O/I	3	1	99.8
<i>Pyrenopolyporus hunteri</i> (ET)	MUCL 49339	Wendt et al. 2018	40,356,773 O/I	5	40	96.5
<i>Xylaria hypoxyton</i> (ET)	CBS 122620	Stadler et al. 2014a	54,341,593	13	19	96.9

* authenticity of the strain is unclear (for detailed discussion about this issue see submitted follow-up work by Wibberg et al. 2020) ; ** originally reported as *Hypoxyton monticulosum* from a marine environment by Leman-Loubière et al. 2017); # either sequenced with Pacbio (PB) or with Oxford Nanopore/Illumina (O/I) as indicated.

Intragenomic polymorphisms in the ITS region of high-quality genomes of the Hypoxylaceae (Xylariales, Ascomycota)

Marc Stadler¹  · Christopher Lambert¹ · Daniel Wibberg² · Jörn Kalinowski² · Russell J. Cox³ · Miroslav Kolářik^{1,2,3,4} · Eric Kuhnert³

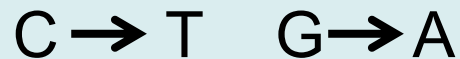
Check for
updates



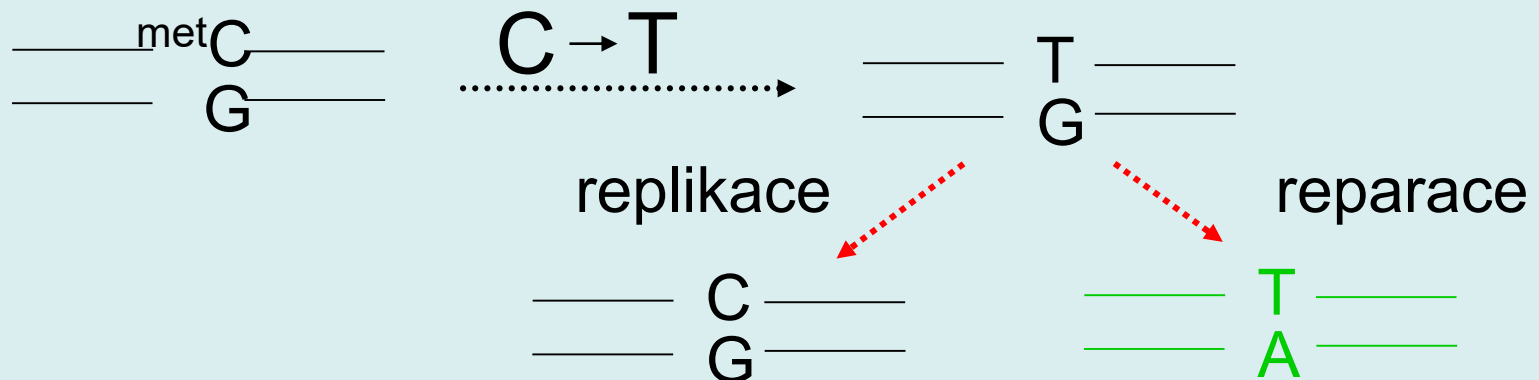
- následkem genové duplikace vznikají **paralogy**, funkce kopií může (ale nemusí) divergovat. Kopie můžou mít různou mutační rychlost (a ta se liší i v rámci rodu)
- **různé kopie ITS rDNA v genomu jsou paralogy**
- ortholog vzniká speciací. Orthologní jsou například Tub2 geny (beta tubulin) u různých druhů

RNA pseudogeny: odhalení

- **Methylation-related substitutions:** pseudogeny jsou GC chudé



- **velká mutační rychlost, náhodně rozložena**
 - velká divergence (dlouhé větve ve fyl. stromech)
 - destabilizace sekundární struktury
 - neobvykle velká variabilita v délce sekvence



RNA pseudogeny: odhalení

TABLE 1

Sequence Characteristics, Thermodynamic Parameters, and Structural Motifs of the ITS Region in Eight Species of *Quercus*

	ITS-1				ITS-2				Pyrimidine pairing at helix II
	Length	G + C	ΔG (37°C) (kcal/mol)	Paired conserved region (bp)	Length	G + C	ΔG (37°C) (kcal/mol)	Four-helix structure	
Sect. Lobatae									
<i>Q. rubra</i> L.	213	66.67	-107.4	Y (21)	211	70.62	-100.0	Y	Y
	226	53.98	-63.8	N (21)	142	50.70	-41.9	N	N
Sect. Quercus									
<i>Q. robur</i> L.	221	65.61	-105.8	Y (21)	209	66.51	-92.7	Y	Y
	200	65.50	-89.5	N (16)	162	67.90	-75.9	N	Y
<i>Q. virginiana</i> Mill.	222	66.22	-110.4	Y (21)	210	68.10	-97.2	Y	Y
	199	63.32	-77.6	N (19)	157	66.88	-51.0	N	N
<i>Q. acutissima</i> Carruth.	221	67.42	-112.2	Y (21)	205	66.83	-104.5	Y	Y
	227	55.95	-68.4	N (21)	132	49.24	-38.4	Y	N
<i>Q. cerris</i> L.	220	66.82	-108.1	Y (21)	209	67.46	-94.3	Y	Y
	199	63.52	-79.8	N (16)	160	66.25	-61.6	N	N
<i>Q. suber</i> L.	214	65.89	-91.0	Y (21)	210	68.57	-100.7	Y	Y
	183	51.91	-53.3	N (21)	139	50.36	-35.5	N	N
<i>Q. coccifera</i> L.	222	66.22	-106.6	Y (21)	212	68.87	-106.6	Y	Y
	205	63.90	-84.9	Y (18)	167	68.26	-73.3	N	Y
<i>Q. ilex</i> L.	221	64.71	-95.4	Y (21)	214	67.29	-97.4	Y	Y
	217	65.44	-96.6	N (16)	169	68.64	-80.2	N	Y

Note. For each species the first row refers to the data of Manos *et al.* (1999) and the second to Samuel *et al.* (1998).

TABLE 1
Comparisons of class I and II alleles

	No. of Alleles	PCR	ITS stability (ΔG) ^a		Relative rate tests ^b		Methylation-induced substitution ^c		Substitution freq ^d	
		DMSO/ non-DMSO	ITS1 + 2 (37°)	ITS1 + 2 (72°)	ITS	5.8S	f(C → T + G → A) (%)	Difference (mC – C)	Liu Sites vs. ITS	5.8S vs. ITS
Gossypium										
Class I	16	0/16	–154	–29	2.22 (63)	0.00 (0)	20	$P = 0.001$	3.9	–
Class II	6	0/6	–139	–23			53		4.5	–
Nicotiana										
Class I	17	16/1	–161	–43	4.79 (100)	7.71 (50)	36	$P = 0.006$	–2.0	–*
Class II	8	0/8	–123	–24			53		16.2	+
Tripsacum										
Class I	8	8/0	–178	–58	2.83 (100)	0.56 (0)	14	$P < 0.001$	–3.9	–
Class II	2	2/0	–133	–28			40		7.4	–*
Winteraceae										
Class I	7	0/7	–149	–31	1.42 (0)	8.62 (0)	5	$P = 0.286$	0.3	–
Class II	3	0/3	–127	–22			17		3.0	+
Zea										
Class I	61	53/8	–176	–51	2.18 (64)	4.61 (31)	22	$P < 0.001$	–1.4	–*
Class II	5	0/5	–152	–39			56		11.1	–

^a Free energy (ΔG) is reported in kcal/mol.

^b Average relative substitution rate ratios of class I and II alleles from same species. Percentage values in parentheses refer to the frequency of significant comparisons ($P < 0.05$).

^c The relative substitution frequencies of C → T and G → A substitutions (corrected for base content as in LI *et al.* 1984); P of a G-test of heterogeneity between the class II alleles' and the class I alleles' relative substitution frequencies; and difference between the average number of C → T at standard methylation sites (CpG and CpNpG) and the average number of C → T at nonmethylation sites.

^d + or – indicates whether there were more or less substitutions at the conserved ITS sites (LIU and SCHARDL 1994) or 5.8S region than the rest of the ITS region. *Comparison was significant at $P < 0.05$.

Table 2 Summary of the length and the secondary structure free energy of *H. monticulosa* MUCL 54604 deviated ITS copy with other strain from this study

Taxon	5.8S rDNA length (bp)	ITS GC content (%)	5.8S secondary structure energy (kcal/mol)
<i>H. monticulosa</i> MUCL 54604 deviating ITS	172	39.9	− 43.6
<i>H. monticulosa</i> MUCL 54604 main ITS	158	43.1	− 44.6
other strains from this study (main ITS copy)	158	43.1–53.1 (avg. 49.4)	− 46.9

Mycological Progress (2020) 19:235–245
<https://doi.org/10.1007/s11557-019-01552-9>

DGfM

ORIGINAL ARTICLE



Intragenomic polymorphisms in the ITS region of high-quality genomes of the Hypoxylaceae (Xylariales, Ascomycota)

Marc Stadler¹ · Christopher Lambert¹ · Daniel Wibberg² · Jörn Kalinowski² · Russell J. Cox³ · Miroslav Kolařík^{1,2,3,4} · Eric Kuhnert³

Použití protein kódujících genů

- nutná znalost pozice intronů
- pomocí alignmentu s příbuznými taxony, či anotovanými sekvencí (Blast)

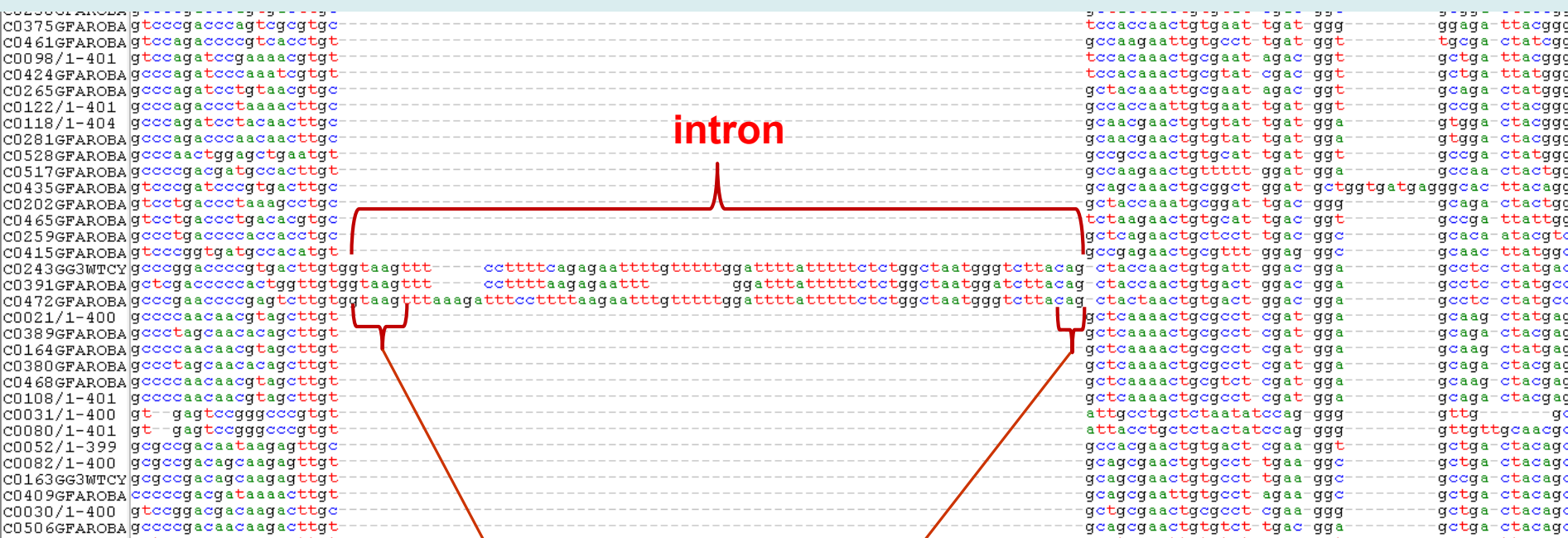


TABLE 2. Cellobiohydrolase (*cbhI*) gene fragments recovered from sporocarp DNAs of six fungal species and best BLAST P matches of putative CBHI protein fragments recovered after removal of introns and translation*

Species (putative gene)	Length of amplicon (bp)	Intron				BLAST P match (% identity)
		Insertion position(s)	Length(s) (bp)	Splice site(s)		
				5'	3'	
<i>Clitocybe nuda cbhI570</i>	570	418	58	GTAAGC	CAG	<i>Agaricus bisporus</i> cellobiohydrolase, CAA90422 (82.4)
<i>Clitocybe nuda cbhI580</i>	580	418	68	GTGAGT	TAG	<i>Volvariella volvacea cbhI-II</i> , AAT64007 (81.2)
<i>Clitocybe nuda cbhI571</i>	571	418	59	GTAAGT	CAG	<i>Volvariella volvacea cbhI-II</i> , AAT64007 (76.5)
<i>Clitocybe gibba cbhI562</i>	562	418	49	GTGAGA	CAG	<i>Volvariella volvacea cbhI-II</i> , AAT64007 (81.2)
<i>Clitocybe gibba cbhI564</i>	564	418	51	GTCAGC	TAG	<i>Volvariella volvacea cbhI-II</i> , AAT64007 (81.2)
<i>Clitocybe gibba cbhI566</i>	566	418	53	GTCAGT	TAG	<i>Volvariella volvacea cbhI-I</i> , AAT64006

Protein kódující geny

- **single- low copy genes:** - β -tubulin
 - jaderný elongační faktor 1 α (EF-1 α , TEF-1 α)
 - podjednotky RNA polymerázy II (RPB1, RPB2)
 - kalmodulin
 - glyceraldehyd-3-fosfát dehydrogenáza
 - aktin
 - cyklophilin
 - MAT faktory (MAT1, 2)...
 - McM7, Tsr1
- **Obsahují jak konzervativní exony tak variabilní introny**
- **Exony mohou být značně variabilní díky tichým mutacím (silent mutations – mutace, které nemají vliv na AMK sekvenci)**
- **Možnost translace a použití AMK sekvence**

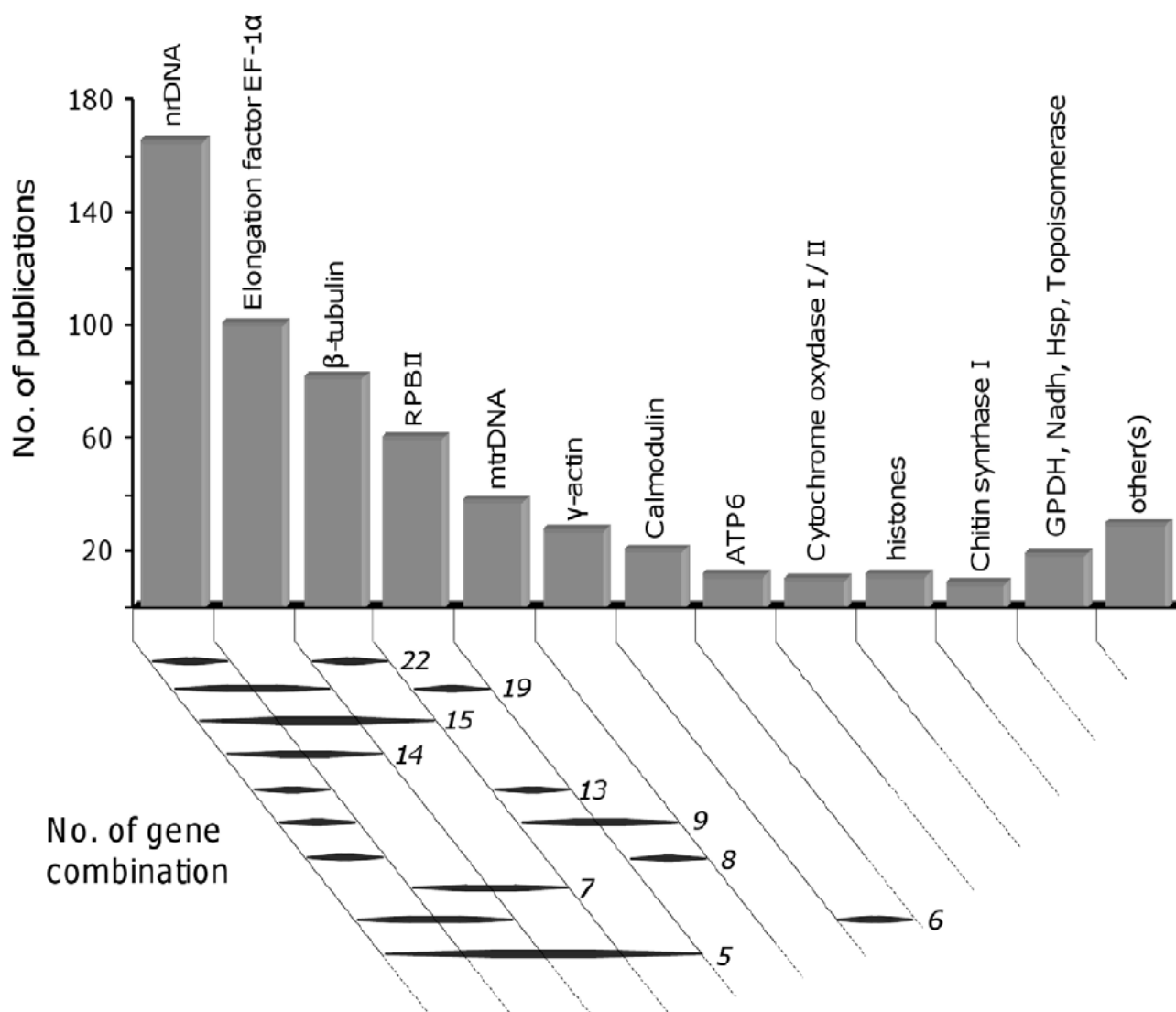


Figure 1. Genes most commonly used in fungal multigene phylogenies. Data were compiled from 196 multigene phylogenies published between 1996 and 2009 retrieved in ISI Thompson Reuters using the following keywords: "fungi" OR "fungal" AND "multigene" or "multilocus" AND "phylogeny". A, distribution of genes/number of publications; B, occurrence of the ten most commonly gene combination used among the 196 studies.

doi:10.1371/journal.pone.0018803.g001

Translační elongační faktor 1 α (EF-1 α , TEF-1 α)

- intronová oblast – dobré rozlišení druhů, horší informace pro vyšší fylogenezi
- exonová oblast – konzervativnější
- různé studie používají různé úseky, které se nemusí překrývat (*Trichoderma*, *Alternaria*) – **nutná dobrá volba sekvenované části**

www.zeallsoft.com

Hypocrea lixii strain zd 57 translation elongation factor 1-alpha (tef1-alpha) gene, large intron

317 bp linear DNA

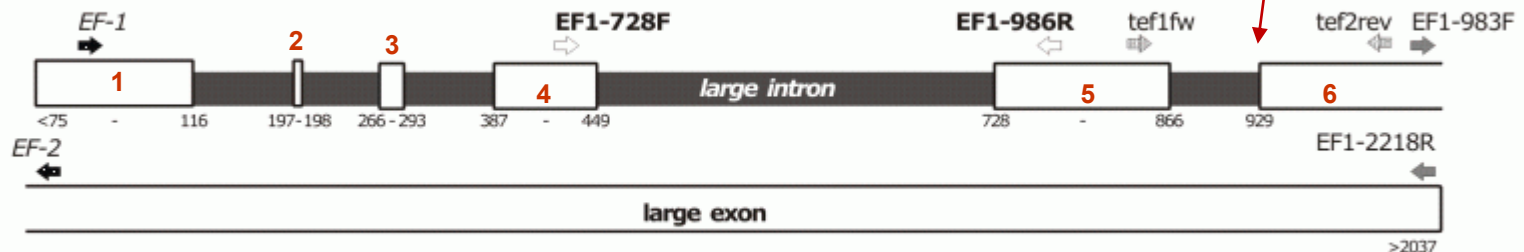
Accession: AY602977.1 GI: 55121930

www.zeallsoft.com

Trichoderma saturnisporum isolate SF21 elongation factor 1a (tef1) gene, exons 5, 6 and partial cds

509 bp linear DNA

Accession: EU595046.1 GI: 192757942



TrichoKey

- různé studie používají různé úseky, které se nemusí překrývat (*Trichoderma*, *Alternaria*)
- nutná dobrá volba sekvenované části

www.zeallsoft.com

[Hypocrea lixii strain zd 57 translation elongation factor 1-alpha \(tef1-alpha\) gene, large intron](#)

317 bp linear DNA

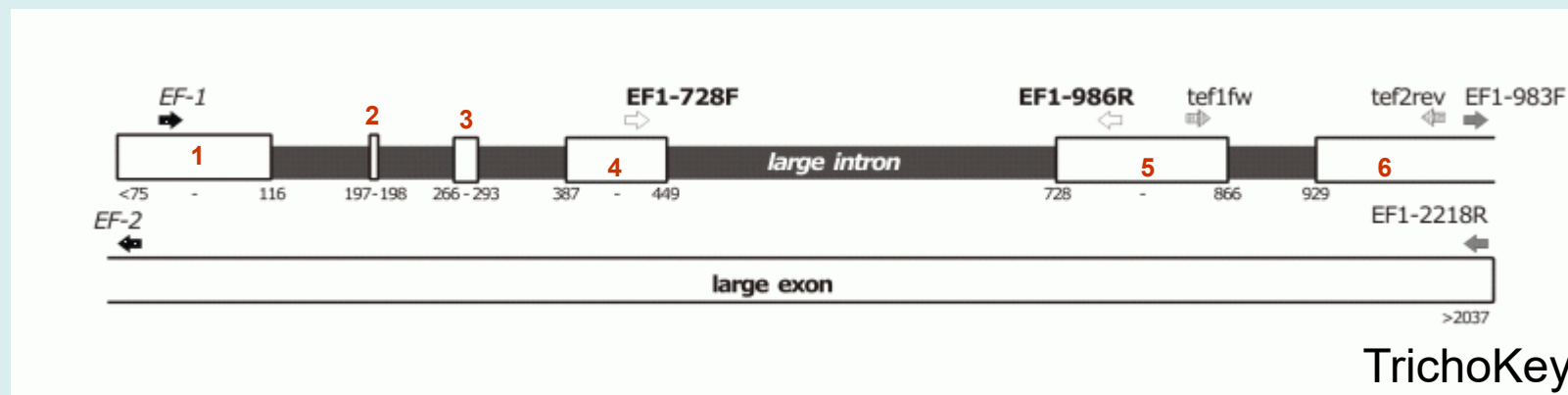
Accession: AY602977.1 GI: 55121930

www.zeallsoft.com

[Trichoderma saturnisporum isolate SF21 elongation factor 1a \(tef1\) gene, exons 5, 6 and partial cds](#)

509 bp linear DNA

Accession: EU595046.1 GI: 192757942

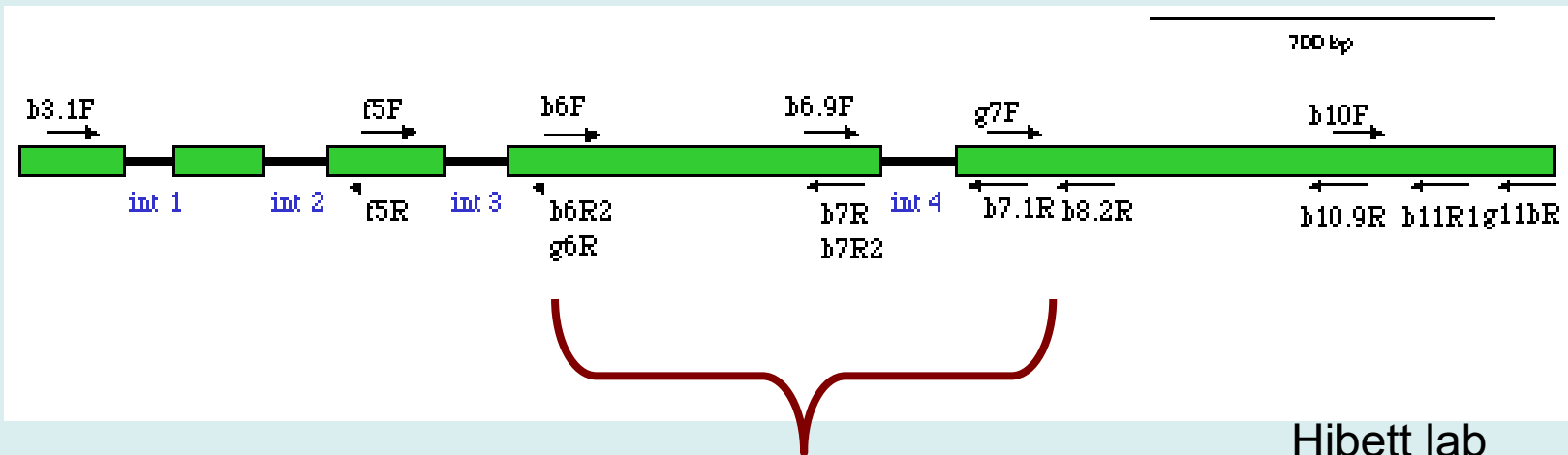


TrichoKey

Mol. tax. metody – sekvence DNA

RNA polymerase II second largest subunit – RPB2

- převážně exony



- Použitím více genů se dostáváme ke realističtějšímu pohledu na příbuznost
- Jednogenové analýzy se mezi sebou mohou lišit, ale většinou ne v koncových nodech

Genome-scale approaches to resolving incongruence in molecular phylogenies

Antonis Rokas*, Barry L. Williams*, Nicole King & Sean B. Carroll

Howard Hughes Medical Institute, Laboratory of Molecular Biology, R. M. Bock Laboratories, University of Wisconsin-Madison, 1525 Linden Drive, Madison, Wisconsin 53706, USA

* These authors contributed equally to this work

One of the most pervasive challenges in molecular phylogenetics is the incongruence between phylogenies obtained using different data sets, such as individual genes. To systematically investigate the degree of incongruence, and potential methods for resolving it, we screened the genome sequences of eight yeast species and selected 106 widely distributed orthologous genes for phylogenetic analyses, singly and by concatenation. Our results suggest that data sets consisting of single or a small number of concatenated genes have a significant probability of supporting conflicting topologies. By contrast, analyses of the entire data set of concatenated genes yielded a single, fully resolved species tree with maximum support. Comparable results were obtained with a concatenation of a minimum of 20 genes; substantially more genes than commonly used but a small fraction of any genome. These results have important implications for resolving branches of the tree of life.



Phylogenomic analysis of 2556 single-copy protein-coding genes resolves most evolutionary relationships for the major clades in the most diverse group of lichen-forming fungi

David Pizarro¹ · Pradeep K. Divakar¹ · Felix Grewe² · Steven D. Leavitt³ · Jen-Pan Huang² · Francesco Dal Grande^{1,4} · Imke Schmitt^{4,5} · Mats Wedin⁶ · Ana Crespo¹ · H. Thorsten Lumbsch²

Assessing the Performance of Single-Copy Genes for Recovering Robust Phylogenies

G. AGUILETA,^{1*} S. MARTHEY,² H. CHIAPELLO,² M.-H. LEBRUN,⁴ F. RODOLPHE,² E. FOURNIER,³
A. GENDRAULT-JACQUEMARD,² AND T. GIRAUD¹

- databáze ortologních rodin single copy genů z 30 genomů
- topologie stromů ze spojených 246-ti genů odlišná od topologie z normálně používaných genů
- nalezeny dva geny (MS465, MS277), které odpovídají 246-ti genové topologii

FUNYBASE
FUNgal phYlogenomic dataBASE

Home
Home Orthologs Advanced search Documentation



Welcome to **FUNYBASE**

FUNYBASE is a database dedicated to the analysis of fungal proteins extracted from complete public fungal genomes, and their classification in clusters of orthologs. The aim of the project is to build a bioinformatic resource useful to answer the following question: how many and which genes are sufficient to build robust fungal phylogenies?

For this purpose, proteins have been extracted from 31 fungal genomes and classified into families using an automated procedure (see [Documentation](#) for details). A subset of 21 genomes has subsequently been defined for which confident gene annotations were available. This dataset have been used to extract single-copy orthologous genes and to build phylogenies using both individual and concatenated alignments.

Using **FUNYBASE** it is possible to access to two sets of results:

1. [the complete protein families dataset](#), including orthologs and paralogs, built from the 31 fungal genomes (use the [advanced search](#) page to access this results).
2. [the subset of 246 families of single-copy orthologs obtained from 21 genomes](#) for which further phylogenetic analyses have been performed. For each of these 246 families, FUNYBASE provides complete information, such as the aminoacid substitution model that best fits the data, the available annotation for the family, the mean percent of identity of the sequences in the family, the number of a variable sites, the aligned proteins, the corresponding phylogenetic tree and its similarity with the tree resulting from the concatenated dataset (go to [orthologs](#) to view this results).

The FUNGI project is funded by [BRG](#) (Bureau des Ressources Génétiques) and started in january 2006.
Research labs associated to this project are :

[Mathématique Informatique et Génome](#)-INRA (Hélène Chiapello, Sylvain Marthey, Annie Gendraul, François Rodolphe)
[Ecologie Systématique et Evolution](#)-CNRS-UPSUD (Tatiana Giraud, Gabriela Aguilera)
[UMR 2847-CNRS-Bayer CropScience](#) (Marc-Henri Lebrun)
[Phytopathologie et Méthodologie de la détection](#)-INRA (Elisabeth Fournier)
[Unité Recherche Génomique Informatique](#)-INRA (Joelle Anselem, Emmanuel Quevillon)

Fungal genomes are issued from [NCBI](#), [MIPS](#), [JGI](#), [Pasteur Institute](#) and [INRA](#).
If you use **FUNYBASE** please cite Aguilera et al, 2007.



Problems? Questions?
Contact [FunYbase Team](#).

**Proteiny pro
procesing pre-rRNA
a pro DNA replikaci**

Tsr1
McM7

Mol. tax. metody – sekvenca DNA

Mcm7 a Tsr1

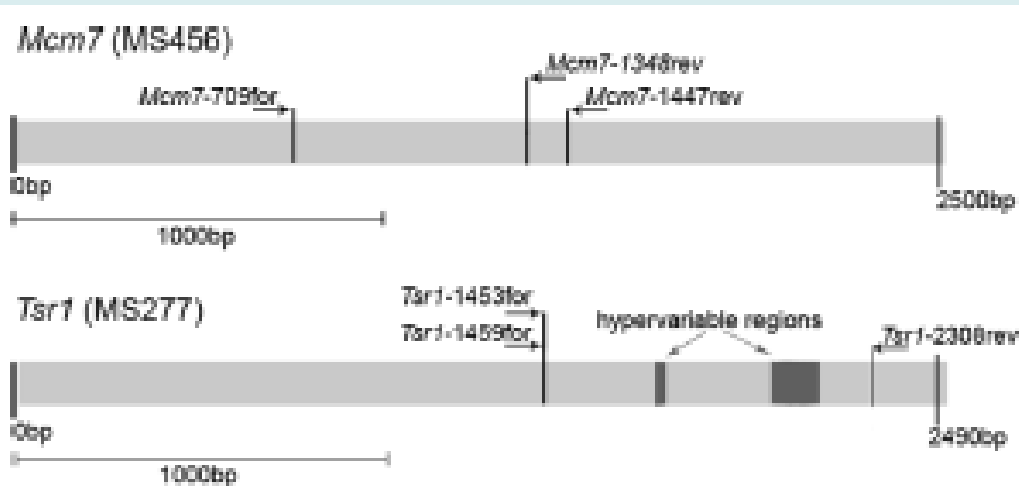


Fig. 1 Locations of the new primers for *Mcm7* and *Tsr1* using *Aspergillus nidulans* mRNA (XM_658504 and XM_658778) as reference sequence. Shaded areas in *Tsr1* indicate regions of high sequence variability.

Table 5 *Mcm7* and *Tsr1* sequence and alignment properties.

	<i>Mcm7</i> (MS456)	<i>Tsr1</i> (MS277)
Introns	None	some (length: 189–272 bp)
Total alignment length (bp)	573	827
Hypervariable (excluded) sites	None	198
Variable sites	357/573 (62.3 %)	489/629 (77.7 %)
Constant sites	216/573 (37.7 %)	140/629 (22.3 %)
Within-genus sequence variation (p-distances) excluding hypervariable sites:		
<i>Malcolmiella</i> (11 OTUs)	0.0055–0.2227	0.0332–0.2193
<i>Aspergillus</i> (7 OTUs)	0.0230–0.2307	0.0357–0.3076
<i>Lecanora</i> (6 OTUs)	0.0377–0.2756	0.0226–0.4148

Table 2 Primers developed in the current study.

Primer Name	Direction	Sequence (5'-3')	Position In <i>A. nidulans</i> mRNA (XM_658504 and XM_658778)	Corresponding amino acid sequence In <i>A. nidulans</i> (AN5992 and AN6266)	Length	Degeneracy
<i>Mcm7</i> -709for	For	ACI MGI GTI TCV GAY GTH AAR CC	709	TRVSDVKP	23 bp	32
<i>Mcm7</i> -1348rev	Rev	GAY TTD GCI ACI CCI GGR TCW CCC AT	1348	MGDPGVAKS	26 bp	16
<i>Mcm7</i> -1447rev	Rev	C ATI ACI GCI GCI GTR AGR CC	1447	GLTAAVM	21 bp	4
<i>Tsr1</i> -1453for	For	GAR TTC CCI GAY GAR ATY GAR CT	1453	EPFDEIEL	23 bp	32
<i>Tsr1</i> -1459for	For	CCI GAY GAR ATY GAR CTI CAY CC	1459	PDEIELHP	23 bp	32
<i>Tsr1</i> -2308rev	Rev	CTT RAA RTA ICC RTG IGT ICC	2308	GTHGYFK	21 bp	8



Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev



New molecular markers for fungal phylogenetics: Two genes for species-level systematics in the Sordariomycetes (Ascomycota)

Donald M. Walker^{a,b,*}, Lisa A. Castlebury^b, Amy Y. Rossman^b, James F. White^a

^aDepartment of Plant Biology and Pathology, Rutgers University, 59 Dudley Rd., New Brunswick, NJ 08901, USA

^bSystematic Mycology and Microbiology Laboratory, USDA Agricultural Research Service, 10300 Baltimore Ave., Bldg. 010A, BARC-West, Beltsville, MD 20705, USA

New molecular markers for fungal phylogenetics: Two genes for species-level systematics in the Sordariomycetes (Ascomycota)

Donald M. Walker^{a,b,*}, Lisa A. Castlebury^b, Amy Y. Rossman^b, James F. White^a

^aDepartment of Plant Biology and Pathology, Rutgers University, 59 Dudley Rd., New Brunswick, NJ 08901, USA

^bSystematic Mycology and Microbiology Laboratory, USDA Agricultural Research Service, 10300 Baltimore Ave., Bldg. 010A, BARC-West, Beltsville, MD 20705, USA

ARTICLE INFO

Article history:

Received 10 June 2011

Revised 5 March 2012

Accepted 4 May 2012

Available online xxxxx

Keywords:

Fungal phylogenetic markers

FG1093

Gnomoniaceae

MS204

Phylogenetic performance

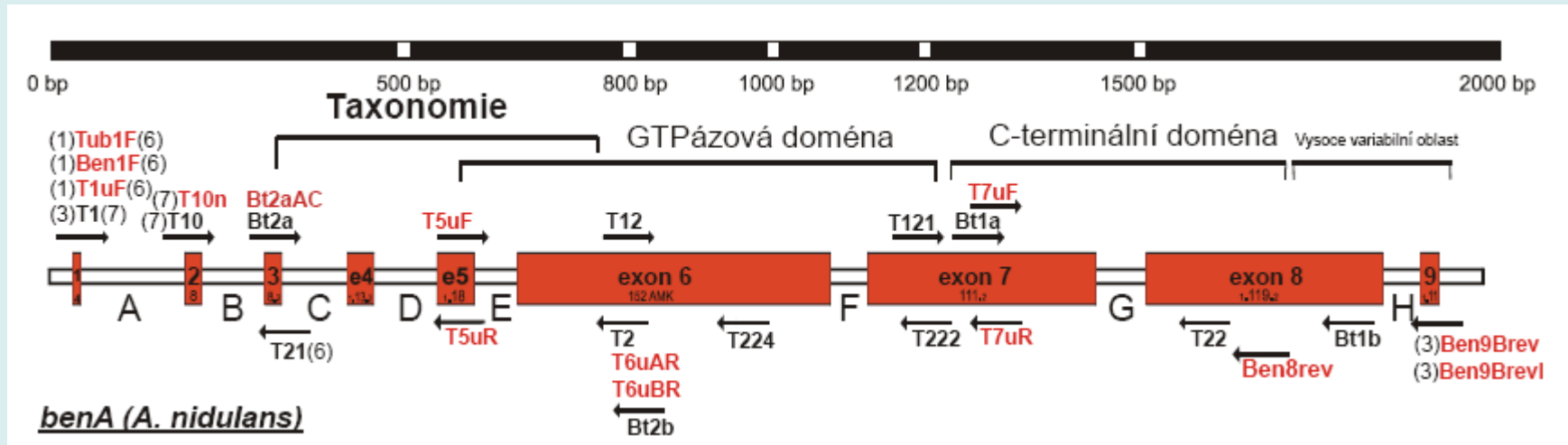
Site rate evolution

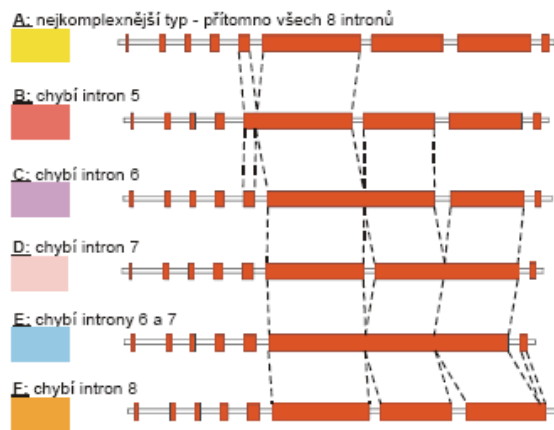
ABSTRACT

Although significant progress has been made resolving deep branches of the fungal tree of life, many fungal systematists are interested in species-level questions to both define species and assess fungal biodiversity. Fungal genome sequences are a useful resource to systematic biologists for developing new phylogenetic markers that better represent the whole genome. Here we report primers for two newly identified single-copy protein-coding genes, FG1093 and MS204, for use with ascomycetes. Although fungi were the focus of this study, this methodological approach could be easily applied to marker development for studies of other organisms. The tests used here to assess phylogenetic informativeness are computationally rapid, require only rudimentary datasets to evaluate existing or newly developed markers, and can be applied to other non-model organisms to assist in experimental design of phylogenetic studies. Phylogenetic utility of the markers was tested in two genera, *Gnomoniopsis* and *Ophiognomonia* (Gnomoniaceae, Diaporthales). The phylogenetic performance of β -tubulin, ITS, and *tef-1 α* was compared with FG1093 and MS204. Phylogenies inferred from FG1093 and MS204 were largely in agreement with β -tubulin, ITS, and *tef-1 α* although some topological conflict was observed. Resolution and support for branches differed based on the combination of markers used for each genus. Based on two independent tests of phylogenetic performance, FG1093 and MS204 were determined to be equal to or better than β -tubulin, ITS, and *tef-1 α* in resolving species relationships. Differences were found in site-specific rate of evolution in all five markers. In addition, isolates from 15 orders and 22 families of Ascomycota were screened using primers for FG1093 and MS204 to demonstrate primer utility across a wide diversity of ascomycetes. The primer sets for the newly identified genes FG1093 and MS204 and methods used to develop them are useful additions to the ascomycete systematists' toolbox.

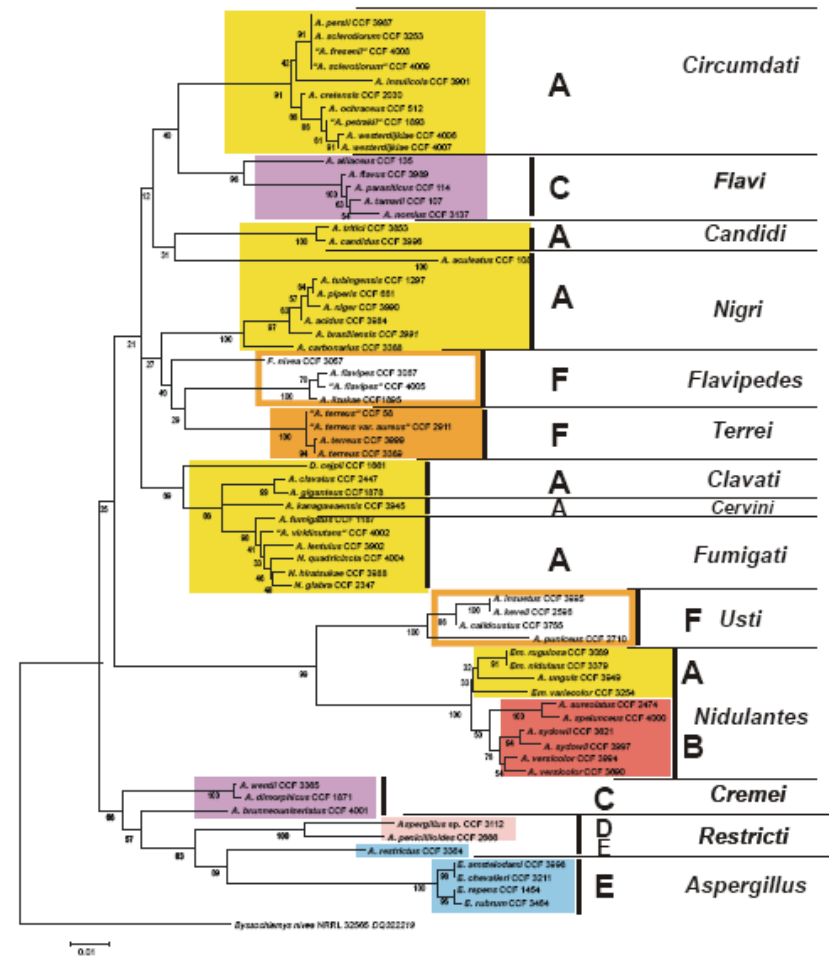
Mol. tax. metody – sekvence DNA

Beta tubulin (tub2)





- počet a pozice intronů má fylogenetický signál



Převzato z Hubka 2011

Obr. 3.18 Architektura exon-intron u genu *benA* napříč sekcemi rodu *Aspergillus*. Každému typu uspořádání je přiřazena barva a písmeno podle legendy a přehledně vyznačena na fylogenetický strom (ITS rDNA; MEGA5, Minimum Evolution, bootstrap 300).

Použití protein kódujících genů

- Paralogie u genu pro beta-tubulin
 - *tub2* = *benA*, hlavní gen
 - *tubC* – přítomen jen někde, není nezbytný, důležitý patrně při sporulaci

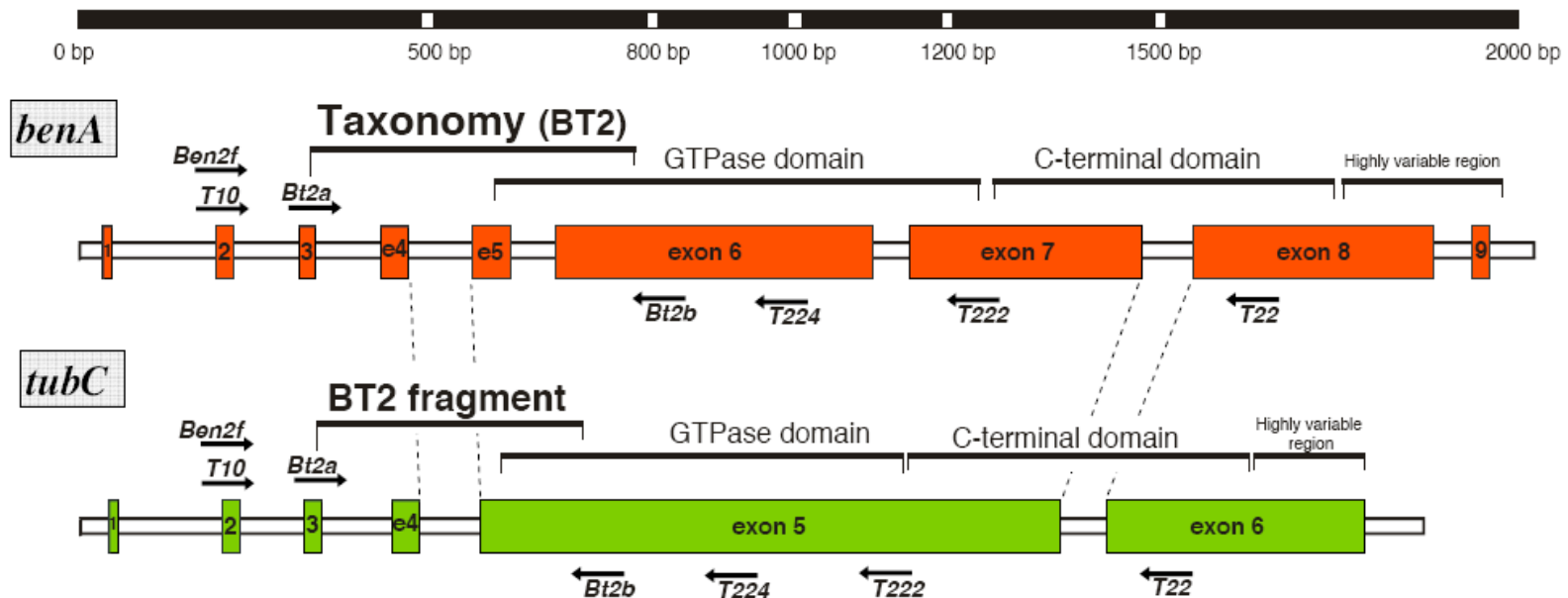
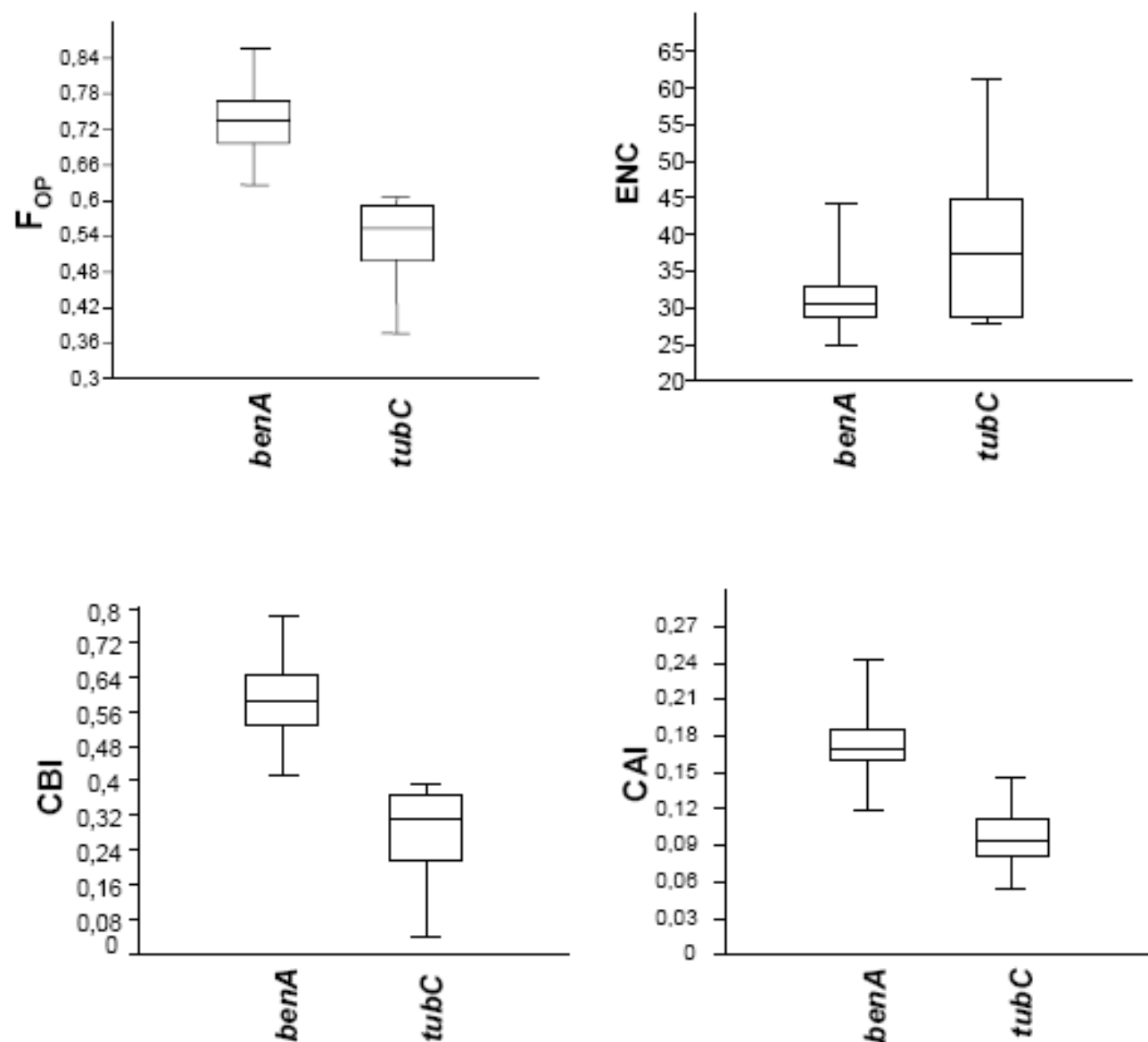


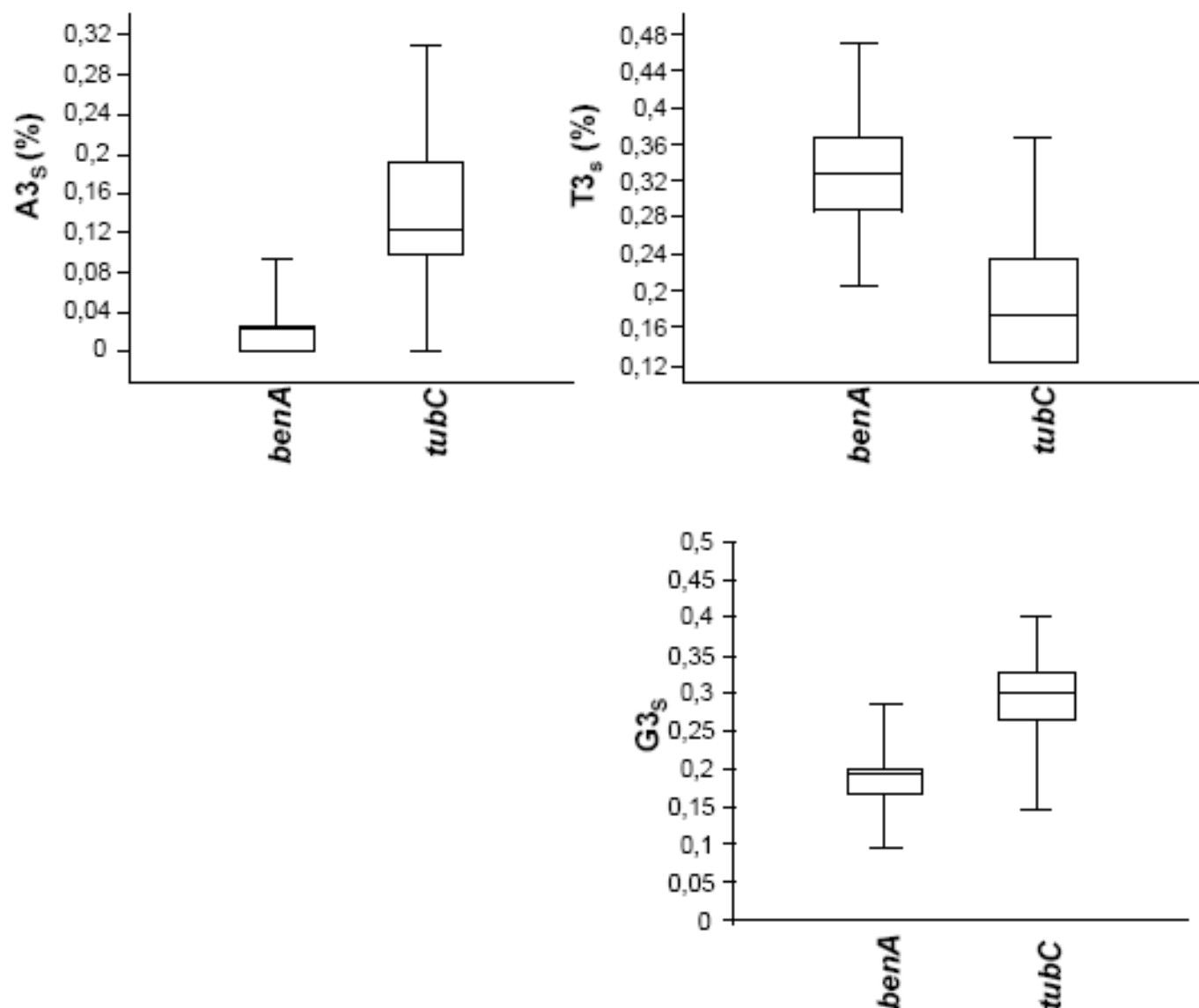
Fig. 1 The exon-intron arrangement of *benA* and *tubC* gene of *A. aculeatus*. The exons are in red (*benA*) and green (*tubC*) colour. The position of primers used for specificity testing (see Table 4) and their orientation is designated by arrows above and below exons.

Odhalení paralogních genů – codon usage bias

- Genetický kód je degenerovaný (1-6 kodonů na 1 AMK) a jednotlivé geny se liší ve spektru použitých kodónů. Obecně lze říci, že více exprimované geny využívají méně kodonů. „Ko-exprimované“ geny využívají podobné spektrum kodonů.
- Několik metod vyhodnocení codon usage bias.
- **Relative synonymous codon usage – RSCU (a podobně i effective number of codons – ENCs)**
frekvence výskytu určitého kodonu dělený náhodným počtem výskytů. Hodnoty RSCU či ENCs jsou nezávislé na AMK složení a jsou zvláště vhodné pro porovnání CU mezi geny, které se liší AMK složením a délkou. RSCU charakterizuje každý gen 59 (~ počet kodonů) proměnnými, což je ale neprakticky velké množství dat.
- **obsah G+C na třetí pozici synonymních kodonů - GC3S.** Tato hodnota vyjadřuje podíl kodonů končících guaninem nebo cytosinem ze skupin synonymních kodonů v daném genu. Podobně může být počítáno zastoupení všech bazí zvlášť pro třetí pozici synonymních kodonů ("silent base composition") - G3s, C3s, A3s a T3s. Více přepisované geny mají často nízké GC3S, protože G≡C vazba je nejpevnější.
- **codon adaptation index - (CAI), Codon bias index (CBI) frequency of optimal codons – Fop.** Tyto metody porovnávají výskyt kodonů s tabulkou reálně používaných kodonů u silně exprimovaných genů *S. cerevisiae* a *A. nidulans*. Některé jsou nezávislé na velikosti a AMK složení a dovedou zohlednit různý obsah G+C.
- více na <http://codonw.sourceforge.net>

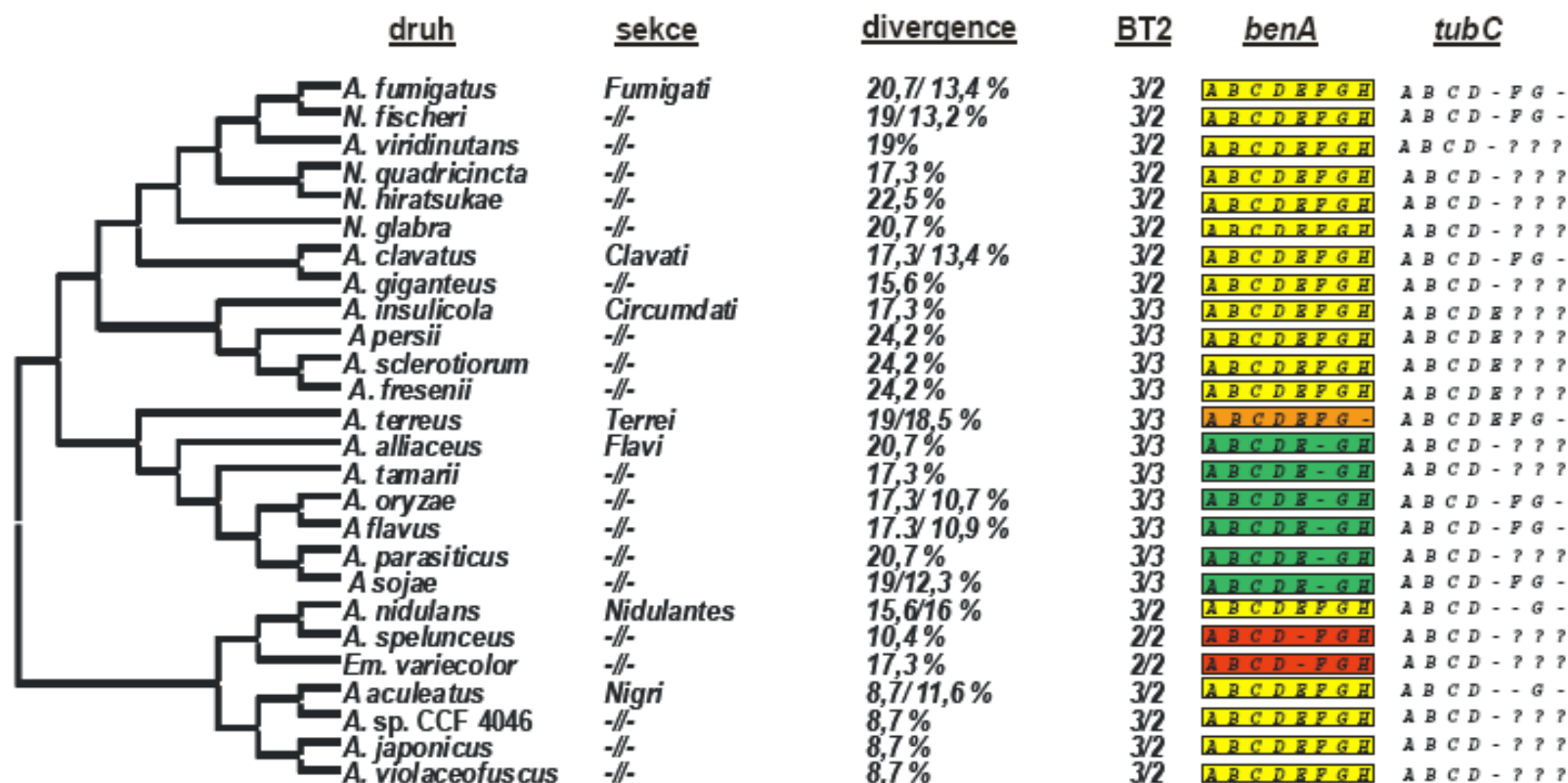


Obr. 3. 26 Charakteristiky pracující se setem "optimálních" kodonů a efektivní počet kodonů (ENC) jako vhodné markery pro odlišení paralogů *benA* a *tubC*. Vlevo nahoře: F_{OP} (MWt $p < 0,0001$); vpravo nahoře: ENC (MWt $p < 0,009$); vlevo dole: CBI (MWt $p < 0,0001$); vpravo dole: CAI (MWt $p < 0,0001$).



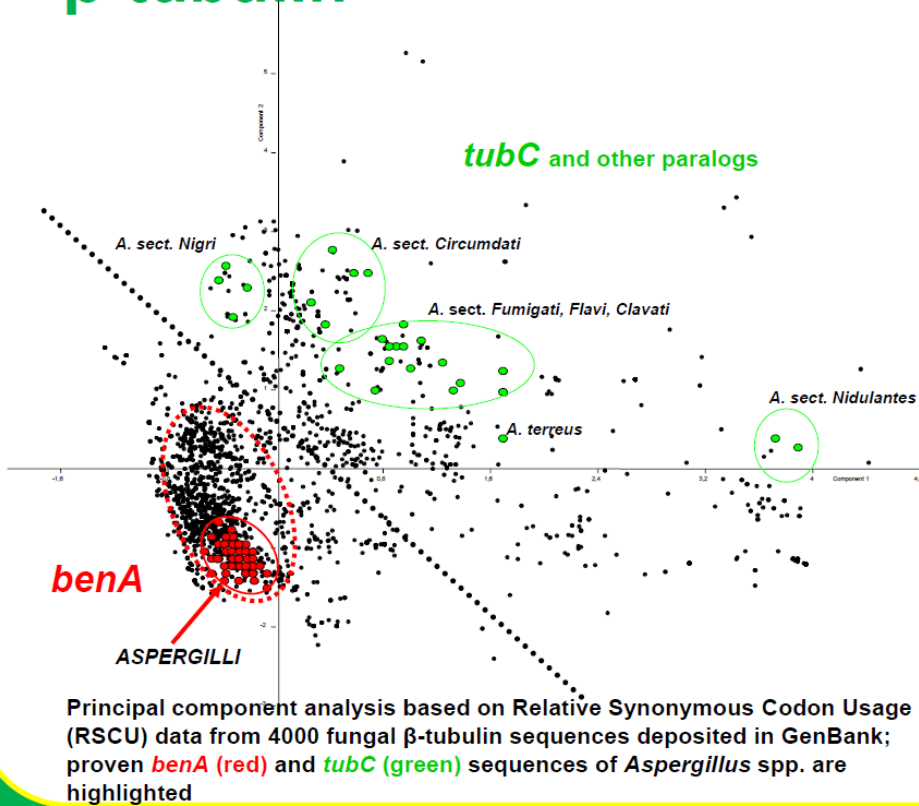
Obr. 3.25 Charakteristiky obsahu bází ve třetí pozici synonymních kodonů, ve kterých se dají nalézt signifikantní rozdíly mezi geny *benA* a *tubC*. Vlevo nahoře: A3_s (MWt $p < 0.0001$); vpravo nahoře: T3_s (MWt $p < 0.0001$); vpravo dole: G3_s (MWt $p < 0.0001$).

- U genů s paralogy může být rozdíl ve funkci, divergenci, codon usage a evoluční rychlosti mezi druhy



Obr. 3.19 Divergence a struktura genů *benA* a *tubC* u druhů, kde byl *tubC* amplifikován. Ve sloupci "divergence" je spočítána rozdílnost aminokyselinové sekvence paralogů *benA* a *tubC* u příslušného druhu. Pokud byly k dispozici sekvence celého genu, je za lomítkem divergence kompletních genů. Ve sloupci "BT2" je porovnán počet intronů v BT2 fragmentu genu *benA* (první číslo) a *tubC* pro daný druh. Uspořádání intronů pro oba geny je zaneseno do posledních dvou sloupců (označení intronů písmeny je podle obrázku 3.15).

β-tubulin



Number of β-tubulin paralogues in selected

Phylum	Number of β-tubulin paralogues
Eurotiales	
<i>Aspergilli</i>	1-2
Hypocreales	
<i>Aciculosporium take</i>	2
<i>Beauveria bassiana</i>	3
<i>Claviceps fusiformis</i> , <i>C. paspali</i>	2
<i>Cordyceps militaris</i>	2
<i>Epichloe</i> spp. (6)	2
<i>Fusarium</i> spp. (3)	2
Pleosporales	
<i>Bipolaris sorokiniana</i>	2
Sordariomycetes	
<i>Glomerella graminicola</i>	3
Capnodiales	
<i>Dothistroma septosporum</i>	2
Glomerellales	
<i>Colletotrichum</i> spp. (3)	2
Pezizales	
<i>Tuber melanosporum</i>	2
Mucorales	
<i>Mucor circinelloides</i>	4
<i>Rhizopus delemar</i>	4
Chytridiomycota	≥2

- misidentification of β-tubulin paralogues in *Aspergillus* sect. *Nigri* and taxonomic consequences were reviewed by Hubka and Kolařík (2012)
- codon usage is particularly useful in differentiating of paralogues *benA* and *tubC*
- β-tubulin gene is especially inconvenient for taxonomy of Zygomycetes due to presence of several hardly distinguishable paralogues
- other group in which paralogues could be mistaken are probably Hypocreales

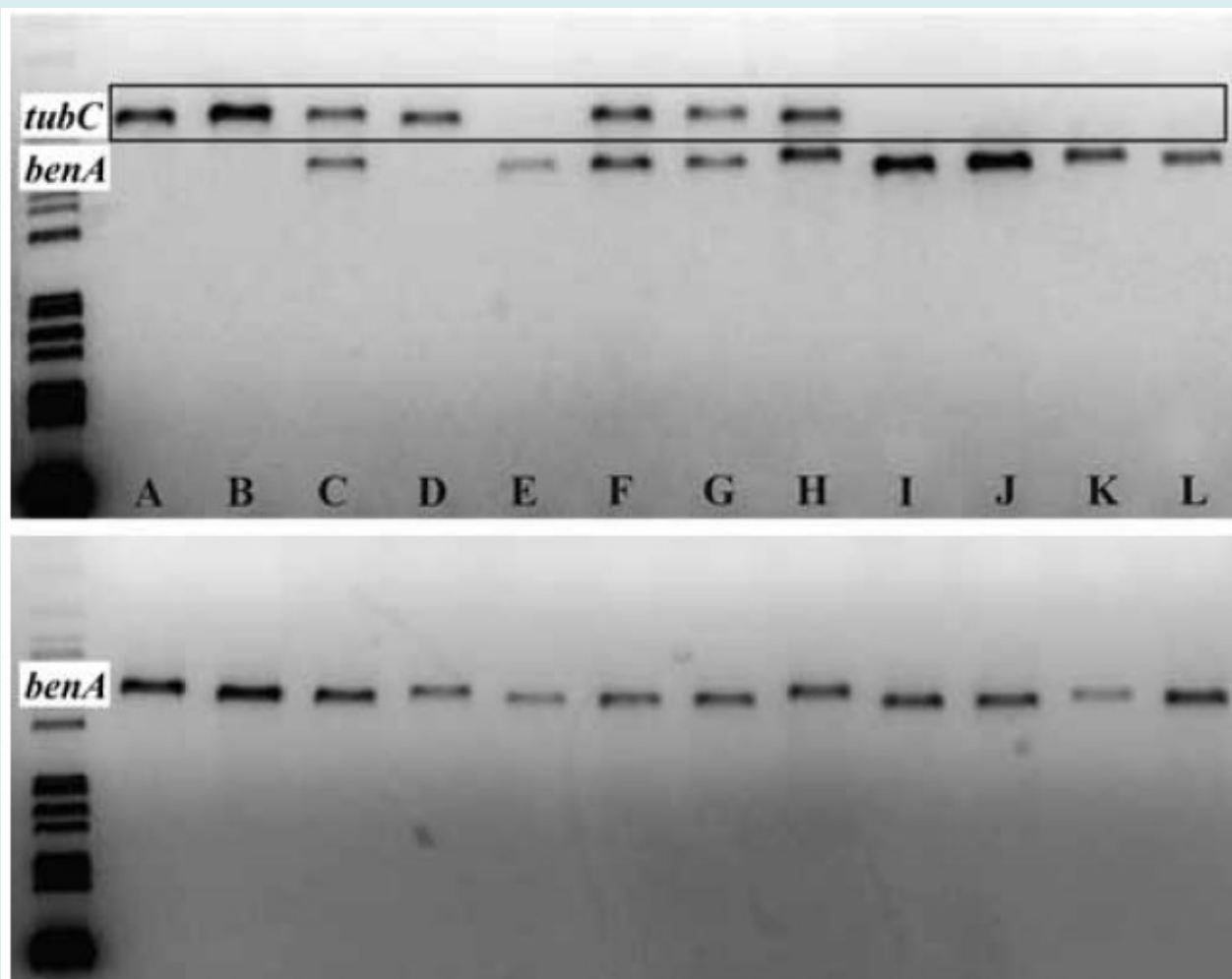


Fig. 5 The electrophoretogram showing amplification products of the β -tubulin gene. The PCR reaction was performed at an annealing temperature of 55 °C. The reaction with primers *Bt2a* and *Bt2b* is shown on the upper part of the image. The lower part shows a reaction with primers *Ben2f* and *Bt2b*. The taxa used for primer testing are listed in Table 1 and designated A-L. The specificity of the *Bt2a* and *Bt2b* primer pair is apparently low in contrast to primer pair *Ben2f* and *Bt2b* that is highly *benA* specific.

Zjištění identity neznámé sekvence

```
>sekvence1  
AGGGTCTCGAGCTGCGTNGTCC.....
```

■ vyhledávání v databázích sekvencí

Možný výsledek

- Zařazení do druhu (**dle podobnosti**)
- Zařazení do vyšší jednotky, následuje fylogenetická analýza (**studium příbuznosti**)

International Nucleotide Sequence Database Collaboration (INSDC)

- **European Molecular Biology Lab. (EMBL)** - European Bioinformatics Institute
- **GenBank** - National Center for Biotechnology Information (NCBI)

(<http://www.ncbi.nlm.nih.gov>)

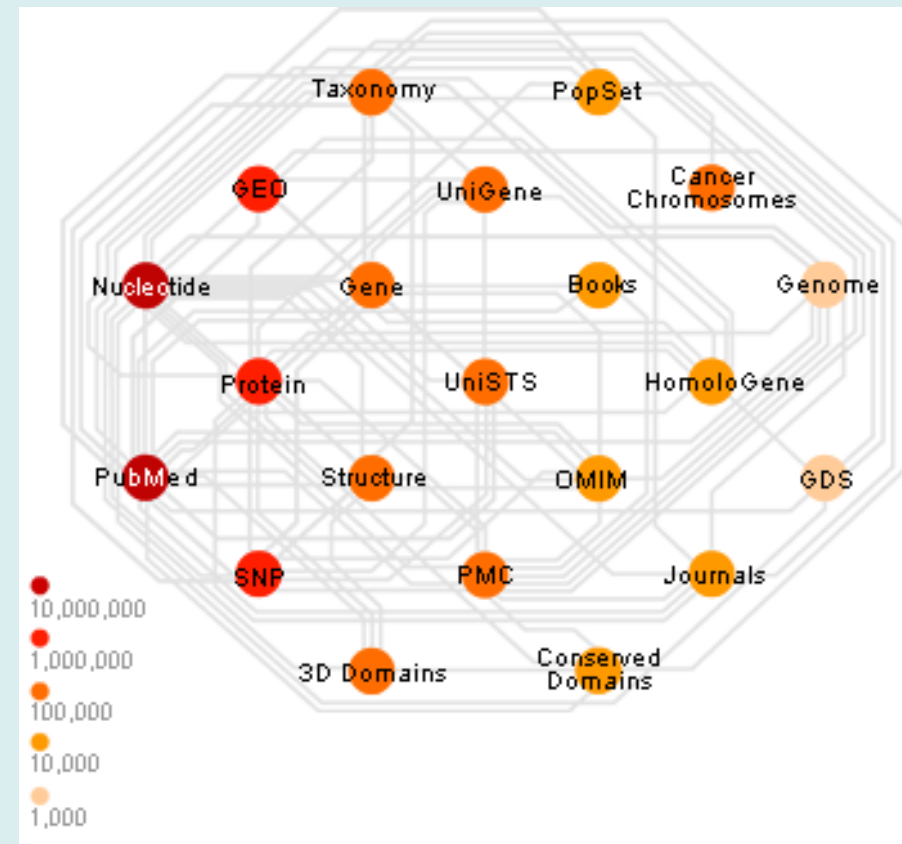
- **DNA Data Bank of Japan (DDBJ)** na National Institute of Genetics of Japan

NCBI - National Center for Biotechnology Information

-(<http://www.ncbi.nlm.nih.gov>)

Entrez - „vstupte“, vyhledávací a stahovací systém pracující s databázemi NCBI (PubMed, Nucleotide and Protein Sequences, Protein Structures, Complete Genomes, Taxonomy....)

<http://www.ncbi.nlm.nih.gov/Database/index.html>



Příklad záznamu v GenBank

Geosmithia sp. CCF3558 partial tub2 gene for beta tubulin, strain CCF3558, exons 1-3

GenBank: HE604180.1

[FASTA](#) [Graphics](#)

[Go to:](#) ☐

LOCUS	HE604180	733 bp	DNA	linear	PLN 09-DEC-2011
DEFINITION	Geosmithia sp. CCF3558 partial tub2 gene for beta tubulin, strain CCF3558, exons 1-3.				
ACCESSION	HE604180				
VERSION	HE604180.1 GI:359802740				
KEYWORDS	.				
SOURCE	Geosmithia sp. CCF3558				
ORGANISM	Geosmithia sp. CCF3558 Eukaryota; Fungi; Dikarya; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; Bionectriaceae; mitosporic Bionectriaceae; Geosmithia.				
REFERENCE	1				
AUTHORS	Kolarik,M. and Jankowiak,R.				
TITLE	Diversity and taxonomy of Geosmithia (Ascomycota: Hypocreales) associated with bark beetles living on Pinaceae in temperate Europe				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 733)				
AUTHORS	Kolarik,M.				
TITLE	Direct Submission				
JOURNAL	Submitted (09-OCT-2011) to the INSDC. Kolarik M., Institut of Microbiology, Czech Academy of Sciences, Videnska 1083, Prague, 142220, CZECH REPUBLIC				

```

FEATURES
    source            1..733
                      /organism="Geosmithia sp. CCF3558"
                      /mol_type="genomic DNA"
                      /strain="CCF3558 = MK860"
                      /isolation_source="host larva on Abies alba"
                      /host="Cryphalus abietis or Pityophthorus pityographus"
                      /db_xref="taxon:367687"
                      /country="Slovakia"
                      /collected_by="M. Kolarik"
                      /identified_by="M. Kolarik"
    gene              <217..>733
                      /gene="tub2"
    mRNA              join(<217..240,383..505,585..>733)
                      /gene="tub2"
    CDS               join(<217..240,383..505,585..>733)
                      /gene="tub2"
                      /codon_start=1
                      /product="beta tubulin"
                      /protein_id="CCE25882.1"
                      /db_xref="GI:359802741"
                      /translation="VHLQTGQCQGNQIGAAFQWTIAGEHGLDSNGVYNGTSELQLERMS
                      VYFNEASGNKYVPRSVLVDLEPGTMDAVRAGPFGQLFRPDNPFVFGASGAGNNWA"
    exon              <217..240
                      /gene="tub2"
                      /number=1
    intron            241..382
                      /gene="tub2"
                      /number=1
    exon              383..505
                      /gene="tub2"
                      /number=2
    intron            506..584
                      /gene="tub2"
                      /number=2
    exon              585..>733
                      /gene="tub2"
                      /number=3

```

ORIGIN

```

    1 tttgacgcgt ttctttctccc ctccctacctc gcccttggt ttttgactat accccgcgcg
   61 cacacaaaaa acgtgtgtct gcacccaagt caacattcca acccagaggg ctccccagag
  121 ggagaggagg agagacagtc tcgcgctggt tctgggttaa atgacccttg gggaagatgg
  181 tgttttgcta acgacgtttc ttttgcctc aaatagggtc acctccagac cggtcagtgc
  241 gtaagtcgat cctccccccc cctccccaga cgaaggacag caggagaagg accattgatg
  301 gcctcctggt ttctcctggt gggggcctgc tacgcgtctt tcttgagca gacgactaac
  361 gtgtttatth tcaatcaaac agggtaacca gattggtgct gctttctggc agaccattgc
  421 tggcgagcat ggctctgaca gcaacggtgt ctacaacggc acctctgagc tccagctcga
  481 gcgcatgagc gtctacttca acgaggtagc ttttagacat ccatgatatg gacatggctg
  541 tgatgggtgac gtttgggtga caggttctcg ctatttgtca ataggcttcc ggcaacaagt
  601 atgttccccg ctccgtcctc gtcgacctcg agcccggtag catggacgcc gtccgcgcgcg
  661 gtccctttgg ccagctcttc cgtcccgaca acttcgtctt cggtgcttcc ggtgccggaa
  721 acaactgggc caa

```

//

GenBank: 35 000 druhů z 5300 rodů

- 70-100 000 druhů popsáno
- 11 500 ve sbírkách kultur hub
- odhad celkem 500 000 - 2 500 000

Ranks:	higher taxa	genus	species	lower taxa	total
Archaea	196	162	620	0	978
Bacteria	1,732	3,237	16,349	871	22,189
Eukaryota	22,530	76,665	359,225	26,121	484,541
Fungi	1,641	5,356	35,199	1,188	43,384
Metazoa	16,066	52,801	182,768	13,122	264,757
Viridiplantae	3,074	15,402	130,658	11,511	160,645
Viruses	693	594	2,174	0	3,461
All taxa	25,184	80,665	378,394	26,992	511,235

Dates: [2008](#) [2009](#) [2010](#) [2011](#) [2012](#) [2013](#) [2014](#) [2015](#) [2016](#) [2017](#) [all dates](#)

Basic BLAST

Choose a BLAST program to run.

<u>nucleotide blast</u>	Search a nucleotide database using a nucleotide query <i>Algorithms: blastn, megablast, discontinuous megablast</i>
<u>protein blast</u>	Search protein database using a protein query <i>Algorithms: blastp, psi-blast, phi-blast</i>
<u>blastx</u>	Search protein database using a translated nucleotide query
<u>tblastn</u>	Search translated nucleotide database using a protein query
<u>tblastx</u>	Search translated nucleotide database using a translated nucleotide query

- Možnost hledat příbuzné sekvence pomocí programu BLAST - Basic Local Alignment Search Tool
- Rozlišovat EMBL accession no., GenBank či DDBJ accesion no.

Nucleotide Sequence Databases – není možnost hledat ve všech najednou

Nr – GenBank + RefSeq Nucleotides (= vybrané anotované a „curated“ sekvence z Genbank). Chybí HTGS, EST, GSS, STS, PAT, WGS.

refseq_rna – referenční RNA database– anotované sekvence proteinů přeložených z cDNA (mRNA)
(nezískáme informace o intronu)

Chromosome, refseq_genomic- anotované genomy z RefSeq

Est – databáze „Expressed Sequence Tags“, krátké transkripty genů (mRNA)

Gss – Genome Survey Sequence, krátké a často náhodné genomické sekvence

Htgs – nedokončené High Throughput Genomic Sequences (dokončené jsou v Nr)

Wgs - whole genome shotgun sequence entries – neanotované genomy

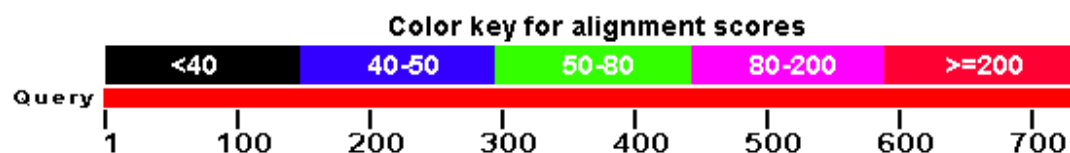
env_nt – sekvence z environmentálních vzorků (řada dalších je v nr)

Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
HE573028.1	Basidiomycota sp. MK-2011-90 genomic DNA containing 18S rRNA gene	503	503	100%	3e-139	100%	
JF691458.1	Uncultured Hymenochaetales clone FM612.1 internal transcribed spacer	119	119	37%	1e-23	87%	
JF691457.1	Uncultured Hymenochaetales clone FM611.2 internal transcribed spacer	119	119	37%	1e-23	87%	
JF691058.1	Uncultured Hymenochaetales clone FM025.1 internal transcribed spacer	119	119	37%	1e-23	87%	

Distribution of 1640 Blast Hits on the Query Sequence

Mouse-over to show define and scores, click to show alignments



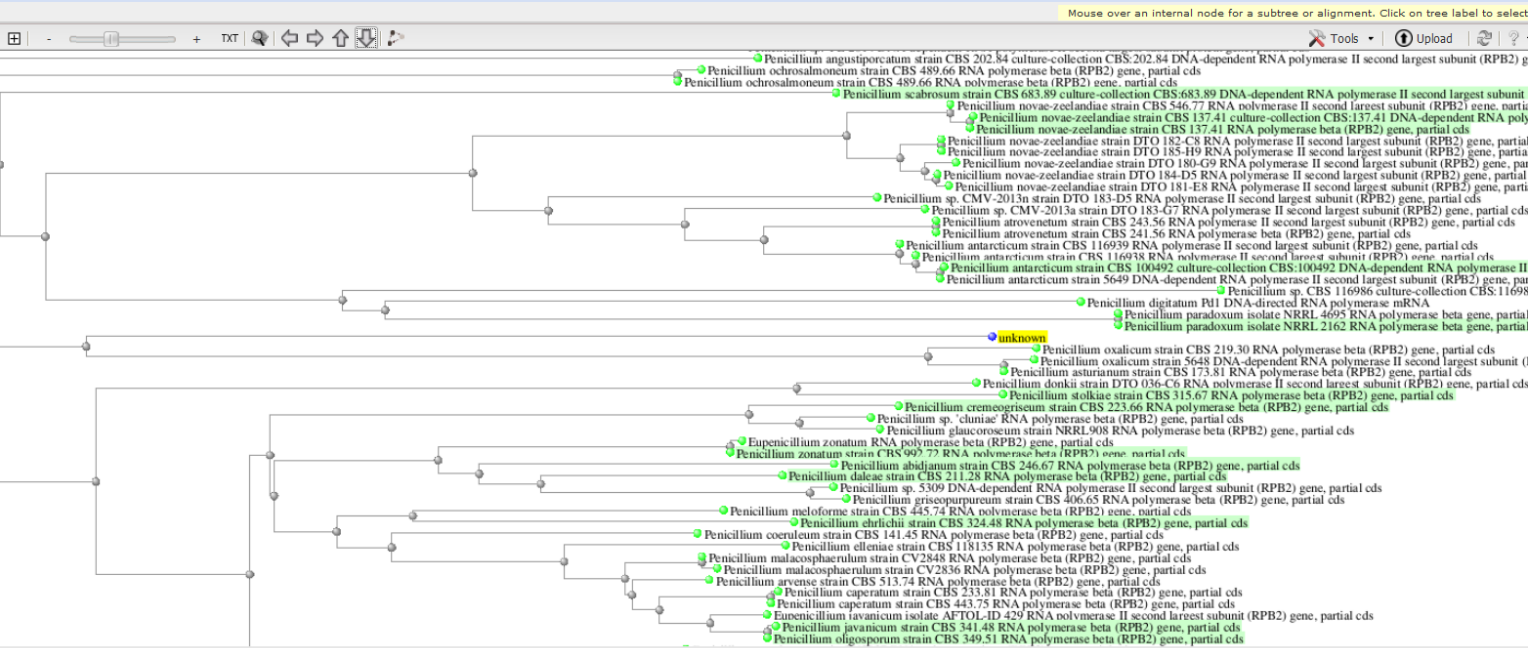
- grafický výstup z Blast
- barevně je zobrazena míra homologie
- v případě protein kódujících genů lze použít na vyhledávání exonů a intronů

Blast- řazení dle podobnosti

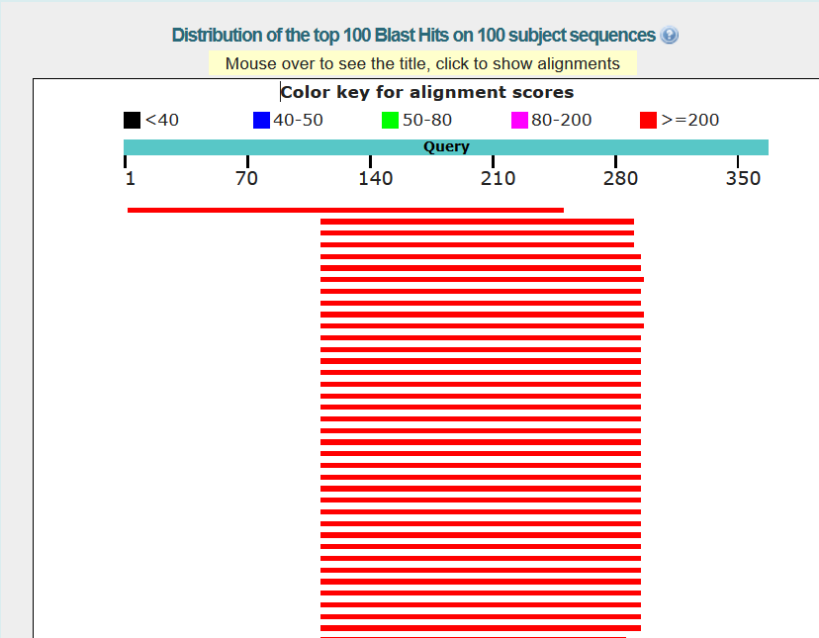
Alignments Download GenBank Graphics Distance tree of results

	Description	Max score	Total score	Query cover	E value	Ident	Accession
☐	Alternaria sp. P11 YNW-2014 internal transcribed spacer 1, partial sequence; 5.8S ribosomal RNA gene and internal transcribed spacer 2, complete sequence; and 28S ribosomal RNA gene, partial sequence	582	1033	78%	5e-162	82%	KP027305.1
☐	Fungal sp. RTL2 internal transcribed spacer 1, partial sequence; 5.8S ribosomal RNA gene and internal transcribed spacer 2, complete sequence; and 28S ribosomal RNA gene, partial sequence	569	807	80%	4e-158	82%	KF910776.1
☐	Alternaria alternata strain DUCC5024 internal transcribed spacer 1, partial sequence; 5.8S ribosomal RNA gene and internal transcribed spacer 2, complete sequence; and 28S ribosomal RNA gene, partial sequence	540	996	78%	3e-149	81%	KJ728679.1
☐	Alternaria alternata 18S ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal transcribed spacer 2, complete sequence; and 28S ribosomal RNA gene, partial sequence	540	1065	78%	3e-149	81%	JQ080319.1
☐	Alternaria sp. 2.5 V1 internal transcribed spacer 1, partial sequence; 5.8S ribosomal RNA gene and internal transcribed spacer 2, complete sequence; and 28S ribosomal RNA gene, partial sequence	494	720	78%	3e-135	80%	KP749178.1
☐	Alternaria alternata strain Lahuf-A alter internal transcribed spacer 1, partial sequence; 5.8S ribosomal RNA gene and internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA gene, partial sequence	486	720	82%	4e-133	79%	MF099863.1
☐	Alternaria sp. C GRMP-59 internal transcribed spacer 1, partial sequence; 5.8S ribosomal RNA gene and internal transcribed spacer 2, complete sequence; and 28S ribosomal RNA gene, partial sequence	483	698	78%	6e-132	80%	KF516006.1
☐	Alternaria alternata strain Lahuf-A al internal transcribed spacer 1, partial sequence; 5.8S ribosomal RNA gene and internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA gene, partial sequence	473	705	81%	3e-129	79%	MF099865.1
☐	Corynespora sp. isolate OLS1 internal transcribed spacer 1, partial sequence; 5.8S ribosomal RNA gene and internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA gene, partial sequence	472	737	78%	1e-128	80%	KU898065.1
☐	Alternaria alternata isolate AA-13 18S ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal transcribed spacer 2, complete sequence; and 28S ribosomal RNA gene, partial sequence	460	685	78%	3e-125	81%	JQ907485.1

Blast- řazení dle příbuznosti



- řazení dle MAX SCORE zohledňuje „coverage“ i podobnost
- Pokud jsou v databázi podobnější, ale výrazně kratší sekvence, tak je Max score upozadí
- Možnost řazení dle „Identity“



Select: [All](#) [None](#) Selected:0

[Alignments](#) [Download](#) [GenBank](#) [Graphics](#) [Distance tree of results](#)

Description	Max score	Total score	Query cover	E value	Ident	Accession
☐ Uncultured fungus clone 2 5.8S ribosomal RNA gene, partial sequence; internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA gene, partial sequence	329	329	67%	3e-86	91%	MF568703.1
☐ Microporellus obovatus voucher Dai 10937 internal transcribed spacer 1, partial sequence; 5.8S ribosomal RNA gene, complete sequence; and internal transcribed spacer 2, partial sequence	305	305	48%	5e-79	98%	KX880613.1
☐ Flabellophora sp. 3 OM-2011 isolate X1277 18S ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal transcribed spacer 2, complete sequer	305	305	48%	5e-79	98%	JN710535.1
☐ Flabellophora sp. 1 OM-2011 isolate X1357 18S ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal transcribed spacer 2, complete sequer	305	305	48%	5e-79	98%	JN710533.1
☐ Entoloma incanosquamulosum voucher MICH 8890 internal transcribed spacer 1, partial sequence; 5.8S ribosomal RNA gene and internal transcribed spacer 2, complete sequence; and large subu	303	303	49%	2e-78	97%	MK169374.1
☐ Nolanea edulis var. concentrica voucher MICH 9131 internal transcribed spacer 1, partial sequence; 5.8S ribosomal RNA gene and internal transcribed spacer 2, complete sequence; and large subu	303	303	49%	2e-78	97%	MK169373.1
☐ Entoloma subcapitatum voucher WTU-F-043799 small subunit ribosomal RNA gene, partial sequence; internal transcribed spacer 1 and 5.8S ribosomal RNA gene, complete sequence; and internal	303	303	50%	2e-78	97%	MK169341.1
☐ Entoloma minutisporum voucher ST258 small subunit ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal transcribed spacer 2, complete se	303	303	49%	2e-78	97%	MK044846.1
☐ Entoloma juncinum strain CBS 384.50 small subunit ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal transcribed spacer 2, complete sec	303	303	49%	2e-78	97%	MH856672.1
☐ Entoloma sp. isolate CM13 224 internal transcribed spacer 1, partial sequence; 5.8S ribosomal RNA gene and internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA g	303	303	50%	2e-78	97%	KY774213.1
☐ Entoloma sp. isolate PGK13 169 small subunit ribosomal RNA gene, partial sequence; internal transcribed spacer 1 and 5.8S ribosomal RNA gene, complete sequence; and internal transcribed sp	303	303	50%	2e-78	97%	KY774192.1
☐ Nolanea sp. voucher FLAS-F-61537 internal transcribed spacer 1, partial sequence; 5.8S ribosomal RNA gene and internal transcribed spacer 2, complete sequence; and large subunit ribosomal R	303	303	49%	2e-78	97%	MH211959.1
☐ Entoloma sp. isolate SA5 internal transcribed spacer 1, partial sequence; 5.8S ribosomal RNA gene and internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA gene, p	303	303	49%	2e-78	97%	KY706196.1
☐ Entoloma sericeum isolate BF8 internal transcribed spacer 1, partial sequence; 5.8S ribosomal RNA gene and internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA g	303	303	49%	2e-78	97%	KY706153.1
☐ Entoloma xanthochroum voucher HMJAU34390 small subunit ribosomal RNA gene, partial sequence; internal transcribed spacer 1 and 5.8S ribosomal RNA gene, complete sequence; and internal t	303	303	49%	2e-78	97%	KY820058.1
☐ Entoloma brunneosericeum voucher 20160813 internal transcribed spacer 1, partial sequence; 5.8S ribosomal RNA gene and internal transcribed spacer 2, complete sequence; and large subunit ri	303	303	49%	2e-78	97%	KY820057.1
☐ Nolanea cf. bicoloripes F12-JS-73 small subunit ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal transcribed spacer 2, complete sequen	303	303	49%	2e-78	97%	KU574738.1
☐ Entoloma sp. R-106 internal transcribed spacer 1, partial sequence; 5.8S ribosomal RNA gene, complete sequence; and internal transcribed spacer 2, partial sequence	303	303	49%	2e-78	97%	KR135338.1
☐ Entoloma sp. isolate MICH-20000 ITS region from T2000 material	303	303	49%	2e-78	97%	NR_120005.1

Table 1. Overview of the five guidelines.

Target of guideline	Way of getting there
1. Establish that the sequences come from the intended gene or marker	Do a multiple alignment of the sequences and verify that they all feature some suitable, conserved sub-region (here the 5.8S gene)
2. Establish that all sequences are given in the correct (5' to 3') orientation	Examine the alignment for any sequences that do not align at all to the others; re-orient these; re-run the alignment step; and examine them again
3. Establish that there are no (bad cases of) chimeras in the dataset	Run the sequences through BLAST in INSD/UNITE and verify that the best match comprises more or less the full length of the query sequences
4. Establish that there are no other major technical errors in the sequences	Examine the BLAST results carefully, particularly the graphical overview and the pairwise alignment, for anomalies
5. Establish that any taxonomic annotations given to the sequences make sense	Examine the BLAST hit list to see that the species names produced make sense

Five simple guidelines for establishing basic authenticity and reliability of newly generated fungal ITS sequences

R. Henrik Nilsson¹, Leho Tedersoo^{2,3}, Kessy Abarenkov², Martin Ryberg⁴, Erik Kristiansson⁵, Martin Hartmann^{6,7}, Conrad L. Schoch⁸, Johan A. A. Nylander⁹, Johannes Bergsten¹⁰, Teresita M. Porter¹¹, Ari Jumpponen¹², Parag Vaishampayan¹³, Otso Ovaskainen¹⁴, Nils Hallenberg¹, Johan Bengtsson-Palme¹⁵, K. Martin Eriksson¹, Karl-Henrik Larsson¹⁶, Ellen Larsson¹, Urmas Kõljalg^{2,3}

- Sekvence ukládány ve směru 5'-3'

Download ▾[GenBank](#) [Graphics](#)

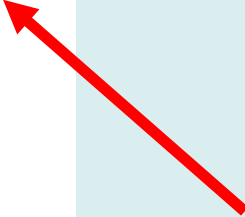
Omphalina sphagnicola small subunit ribosomal RNA gene, partial sequence, internal transcr

Sequence ID: [U66453.1](#) Length: 2069 Number of Matches: 1

Range 1: 11 to 571 [GenBank](#) [Graphics](#)

▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
841 bits(455)	0.0	526/561(94%)	1/561(0%)	Plus/Minus
Query 1	GATTATAAGCGAATCCATGGTTCACTGCCAATGACATAGATAATTATCATGGCAGAACTG	60		
Sbjct 571	GATTATAAGCGAATCCATGTTTCACTGCAAATGACATAGATAATTATCATGCCAGAGCTG	512		
Query 61	ATCCACTTAAAGAGATTTTCGCTAATGCATTTAAAGAGGAGCTGACCTTTAAAACCAGCA	120		
Sbjct 511	ATCCACTTAGAGAGATTTTCGCTAATGCATTTAAAGAGGAGCTGACCTTTAAAGCCAGCA	452		
Query 121	CACTCCAAGGCCAAGCTTAAACAAGGAATAAACCTTGAAAGGTGATAATTTAATGACAC	180		
Sbjct 451	CACTCCAAGTCCAAGCTTAAACAAGGAATAAACCTTGAAAAGTTGATAATTTAATGACAC	392		
Query 181	TCCAAAGCCGATCCCGCTCCCAATACCAAGCCCGCCCAAGCCCGCTTCAAGCATTCGATCA	240		



forma prezentace identifikace pomocí „blasting“

TABLE 1. Database typing of ITS sequences^a

Sum (%) ^b	OTU	Accession no.	Best BLAST hit					% Similarity ^e
			Taxon	Phylum	Class	Score ^c	Accession no. ^d	
10.58	Ms9Mb4 ^f	AJ875353	<i>Cladosporium</i> sp. strain 4/97-17	Ascomycota	Dothideomycetes	1,158	AJ279487	99.8
5.27	Ms4Mb4a ^f	AJ875341	Unidentified isolate 4/97-9	Ascomycota	Sordariomycetes	1,088	AJ279484	100
5.11	Ms6Mb4	AJ875345	<i>Chaetomium globosum</i>	Ascomycota	Sordariomycetes	476	AY429056	80.5
3.70	Ms28Mb4a	AJ875376	<i>Cercophora appalachianensis</i>	Ascomycota	Sordariomycetes	667	AF177155	79.6
2.55	Ms9Mb43a	AJ875357	<i>Stenella araguata</i>	Ascomycota	Dothideomycetes	359	AF362066	81.0
2.45	Ms39Mb28 ^g	AJ875381	<i>Aureobasidium pullulans</i>	Ascomycota	Dothideomycetes	1,197	AF455533	99.7
2.45	Ms51Mb4	AJ875391	<i>Thielavia intermedia</i>	Ascomycota	Sordariomycetes	603	AJ271588	89.9
2.40	Ms9Mb43b	AJ875358	<i>Cladosporium</i> sp.	Ascomycota	Dothideomycetes	581	AJ222807	83.3
2.35	Ms22Mb7	AJ875369	Ericoid mycorrhiza	Ascomycota	Leotiomycetes	928	AY046400	98.0
2.35	Ms25Mb17	AJ875374	<i>Terfezia leptoderma</i>	Ascomycota	Pezizomycetes	509	AF396863	86.4
2.19	Ms7Mb28b ^f	AJ875351	<i>Stagonospora</i> sp. strain 4/99-18	Ascomycota	Dothideomycetes	1,185	AJ496627	100
1.82	Ms5Mb4	AJ875343	<i>Podospora ellisiana</i>	Ascomycota	Sordariomycetes	458	AY515360	79.6
1.77	Ms7Mb4	AJ875347	<i>Podospora decidua</i>	Ascomycota	Sordariomycetes	684	AF443851	90.6
1.51	Ms41Mb20	AJ875382	<i>Hymenogaster griseus</i>	Basidiomycota	Homobasidiomycetes	1,203	AF325636	97.8
1.46	Ms5Mb7 ^g	AJ875344	<i>Cryptococcus carnescens</i>	Basidiomycota	Heterobasidiomycetes	914	AB105438	97.8
1.41	Ms20Mb11 ^f	AJ875368	<i>Cylindrocarpon</i> sp.	Ascomycota	Sordariomycetes	963	AY295332	99.8
1.36	Ms25Mb24	AJ875375	<i>Sebacina</i> sp.	Basidiomycota	Heterobasidiomycetes	452	AF284135	88.4
1.36	Ms4Mb4b	AJ875342	<i>Colletotrichum sublineolum</i>	Ascomycota	Sordariomycetes	325	AJ301978	74.2
1.20	Ms48Mb21	AJ875390	<i>Sporobolomyces roseus</i>	Basidiomycota	Urediniomycetes	1,191	AY015438	99.8
1.09	Ms28Mb4b	AJ875377	<i>Podospora intestinacea</i>	Ascomycota	Sordariomycetes	549	AY515363	87.5
1.09	Ms9Mb28	AJ875356	<i>Dactylaria junci</i>	Ascomycota	Orbiliomycetes	379	AY265321	73.3
1.04	Ms17Mb11 ^f	AJ875366	<i>Cylindrocarpon</i> sp. strain 5/97-12	Ascomycota	Sordariomycetes	1,142	AJ279482	99.7

Zjištění identity neznámé sekvence

- **GenBank** - mnoho nekvalitních dat
 - chyby v sekvencích (chybovost Taq polymerázy, chiméry)
 - špatné určení
- **řešení** - vybírat jen sekvence z taxonomických prací
 - anotované a kontrolované databáze (obecné i specializované databáze)
 - RDB (LSU, ITS), obecná databáze
 - NCBI- RefSeq, RefRNA (řada genů), obecná databáze
 - Unite
 - Phymyco-DB (SSU, TEF1a)
 - Q-bank (řada genů), karanténní fytopatogenní houby
 - ISHAM ITS Database (ITS), patogení živočichů

- do databáze může uložit kdokoliv cokoliv
- snaha o kontrolu administrátory
- Řetězení chybné identifikace

Alignments Download GenBank Graphics Distance tree of results							
	Description	Max score	Total score	Query cover	E value	Ident	Accession
☐	Cladosporium tenellum isolate CLAD185 internal transcribed spacer 1, partial sequence; 5.8S ribosomal RNA gene and internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA gene, partial sequence	889	889	100%	0.0	100%	MK111467.1
☐	Aff. Cladosporium sp. strain B64 small subunit ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA gene, partial sequence	889	889	100%	0.0	100%	MF615037.1
☐	Cladosporium sp. isolate D3 small subunit ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA gene, partial sequence	889	889	100%	0.0	100%	MG548567.1
☐	Cladosporium herbarum strain IGB-KR52 internal transcribed spacer 1, partial sequence; 5.8S ribosomal RNA gene and internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA gene, partial sequence	889	889	100%	0.0	100%	MK012417.1
☐	Cladosporium herbarum strain CBS 128234 small subunit ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA gene, partial sequence	889	889	100%	0.0	100%	MH864860.1
☐	Cladosporium macrocarpum strain CBS 127778 small subunit ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA gene, partial sequence	889	889	100%	0.0	100%	MH864773.1
☐	Cladosporium herbarum strain CBS 127777 small subunit ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA gene, partial sequence	889	889	100%	0.0	100%	MH864772.1
☐	Cladosporium herbarum strain CBS 127961 internal transcribed spacer 1, partial sequence; 5.8S ribosomal RNA gene and internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA gene, partial sequence	889	889	100%	0.0	100%	MH864758.1
☐	Cladosporium cladosporioides strain CBS 127051 small subunit ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA gene, partial sequence	889	889	100%	0.0	100%	MH864397.1
☐	Cladosporium herbarum strain CBS 125831 small subunit ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA gene, partial sequence	889	889	100%	0.0	100%	MH863798.1
☐	Cladosporium variable strain CBS 121636 small subunit ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA gene, partial sequence	889	889	100%	0.0	100%	MH863132.1
☐	Cladosporium variable strain CBS 121635 small subunit ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA gene, partial sequence	889	889	100%	0.0	100%	MH863131.1
☐	Cladosporium tenellum strain CBS 121634 small subunit ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA gene, partial sequence	889	889	100%	0.0	100%	MH863130.1
☐	Cladosporium subinflatum strain CBS 121630 small subunit ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA gene, partial sequence	889	889	100%	0.0	100%	MH863129.1
☐	Cladosporium allicinum strain CBS 121624 small subunit ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA gene, partial sequence	889	889	100%	0.0	100%	MH863126.1
☐	Cladosporium macrocarpum strain CBS 121623 small subunit ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA gene, partial sequence	889	889	100%	0.0	100%	MH863125.1
☐	Cladosporium herbarum strain CBS 121621 small subunit ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA gene, partial sequence	889	889	100%	0.0	100%	MH863124.1
☐	Cladosporium ossifragi strain CBS 842.91 small subunit ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA gene, partial sequence	889	889	100%	0.0	100%	MH862342.1
☐	Cladosporium variable strain CBS 195.54 internal transcribed spacer 1, partial sequence; 5.8S ribosomal RNA gene and internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA gene, partial sequence	889	889	100%	0.0	100%	MH857289.1
☐	Cladosporium allicinum strain CBS 188.54 small subunit ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA gene, partial sequence	889	889	100%	0.0	100%	MH857286.1
☐	Cladosporium phlei strain CBS 307.50 small subunit ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA gene, partial sequence	889	889	100%	0.0	100%	MH856639.1
☐	Cladosporium phlei strain CBS 306.50 small subunit ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA gene, partial sequence	889	889	100%	0.0	100%	MH856638.1



Current version: **8.2**; Last updated: 2020-01-15 ([read more](#))

Number of ITS sequences (UNITE+INSID): **2 480 043**; Number of UNITE fungal Species Hypotheses with DOIs at 1.5% threshold: **102 100** ([more statistics](#))

UNITE is a database and sequence management environment centered on the eukaryotic nuclear ribosomal ITS region. All eukaryotic ITS sequences from the International Nucleotide Sequence Database Collaboration are clustered to approximately the species level (distance between species in steps of 0.5%), and all such species hypotheses are given a DOI to facilitate unambiguous scientific communication and data assembly. The user can interact with these species hypotheses (e.g., **SH1566366.08FU**) in various ways, and third-party sequence and metadata annotation are supported... [Read more](#)

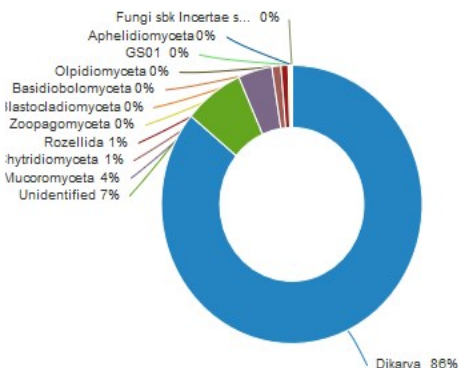
Fungal Species Hypotheses

Threshold	1.5 %	▼	Include	All SH-s	▼	Start typing taxon name here ...	Go	Reset
-----------	-------	---	---------	----------	---	----------------------------------	----	-------



- ▶ Dikarya (97,214)
- ▶ Unidentified (8,380)
- ▶ Mucoromyceta (4,767)
- ▶ Chytridiomyceta (1,251)
- ▶ Rozellida (1,006)
- ▶ Zoopagomyceta (283)
- ▶ Blastocladiomyceta (79)
- ▶ Basidiobolomyceta (53)
- ▶ Olpidiomyceae (35)
- ▶ GS01 (informal name, see [Tedersoo et al., 2017](#)) (30)
- ▶ Aphelidiomyceta (19)
- ▶ Fungi sbk Incertae sedis (1)

SH graph: Fungi



Search by SH code ...

Go

Species Hypotheses

Sequences

... or specify search filters below:

Top 50 most wanted

Filters

Version

8.0

Threshold

1.5 %

Include

All SH-s



Taxon name

Start typing here ...

☐ Include sub-taxa

Accession number

UNITE / INSD accession no.

Go

Reset

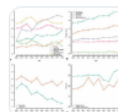
TWEETS BY UNITE



UNITE Community

@unite_sh

When mycologists describe new species, not all relevant information is provided (clearly enough):
mycokeys.pensoft.net/article/56691/



When mycologists describe new specie...
Taxonomic mycology struggles with what seems to be a perpetual shortage of
mycokeys.pensoft.net



Sep 10, 2020

RECENT POSTS

PhD students course "Identifying and publishing HTS/Sanger DNA sequence datasets"
29.10.2018

Call for Citizen Scientists "Submit specimens for sequencing!"
04.05.2017

PhD students course "Metabarcoding Data Management and Open Data"
20.04.2017

View More...

ABOUT US

Contact
How to cite UNITE
FAQ
History
UNITE Board
Acknowledgements

Species Hypothesis - any species-level group of individuals that share a given set of observed characters. In this case it is sequences that share a given level of similarity – that is, operational taxonomic units (OTUs). However, since we'd like to traverse the step between OTUs and species, we'd like to define our OTUs in such a way that they correspond to the species level as currently practiced in mycology. This is a non-trivial exercise and the results will not always be perfect, hence the suffix "hypothesis".

Remark: This is a practical definition of the SH, which is a major building block of the new UNITE system.

UNITE - Firefox

Soubor Úpravy Zobrazit Přejít Záložky Nástroje nápověda

← → ↺ × 🏠 ➕ 🖨️ 🔍 http://unite.ut.ee/sp_desc.php?sp_id=A

Getting Started Latest Headlines

SEZNAM Hledej EN -> CZ

Search NCBI PubMed for

Home Run Analysis Search Pages Not

Tomentella bryophila (Pers.) M.J.Larsen



Tomentella bryophila (Pers.) M. J. Larsen, Mycol. Mem. 4. p. 51. 1974.

Basionym: *Sporotrichum bryophilum* Pers., Mycol. Europ. 1: 78. 1822.

Synonym: *Tomentella ferruginella* Bourdot & Galzin, Bull. Soc. Mycol. France 40: 157. 1924. (PC)

Synonym: *Hypochnus fuscoferruginosus* Bres., Ann. Mycol. 1: 109. 1903. (S)

Tomentella fuscoferruginosa (Bres.) Litsch., Ann. Mycol. 39: 377. 1941.

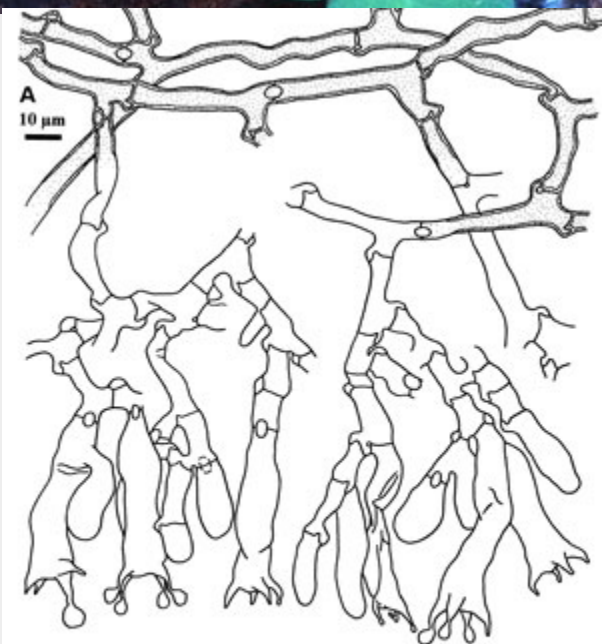
Synonym: *Tomentella spongiosa* (Schwein.: Fr.) Bourdot & Galzin var. *brevispina* Bourdot & Galzin, Bull. Soc. Mycol. France 40: 154. 1924. (PC)

Tomentella brevispina (Bourdot & Galzin) M. J. Larsen, Mycologia 62: 136. 1970.

Synonym: *Hypochnus viridescens* Bres. & Torrend in Torrend, Broteria 11: 85. 1913. (S)

Tomentella viridescens (Bres. & Torrend) Bourdot & Galzin, Hym. France, p. 477. 1928.

Lectotype: In Persoon herbarium. (L)





RefSeq: NCBI Reference Sequence Database

A comprehensive, integrated, non-redundant, well-annotated set of reference sequences including genomic, transcript, and protein.

Using RefSeq

[About RefSeq](#)

[Prokaryotic RefSeq Genomes](#)

[FAQ](#)

[Growth Statistics](#)

[NCBI Handbook](#)

[Factsheet](#)

RefSeq Access

[RefSeq FTP](#)

[RefSeq genomes FTP](#)

[New RefSeq genomic \(last 30 days\)](#)

[New RefSeq transcripts \(last 30 days\)](#)

[New RefSeq proteins \(last 30 days\)](#)

[Searching for RefSeq records \(Queries\)](#)

RefSeq projects

[Consensus CDS \(CCDS\)](#)

[HIV-1, Human Protein Interaction Database](#)

[RefSeqGene](#)

[Targeted Loci](#)

[Virus Variation](#)

[Publications and Citing RefSeq](#)

Announcements

November 7, 2016

RefSeq Release 79 is available for FTP

This release includes:

Proteins: 73,099,060

Transcripts: 16,967,019

Organisms: 64,227

Available at: <ftp://ftp.ncbi.nlm.nih.gov/refseq/release/>

Documentation: [Release Notes](#)

See [previous announcements](#), follow [NCBI on Twitter](#), or subscribe to [NCBI's refseq-announce mail list](#) to receive announcements.

Related Links

[Assembly](#)

[Gene](#)

[Genome](#)

[HomoloGene](#)

[Map Viewer](#)

[UniGene](#)

Feedback & Credits

[Contact RefSeq Help Desk](#)

[Contact CCDS Help Desk](#)

[Submit a GeneRIF](#)

[Collaborators](#)

BARCODING

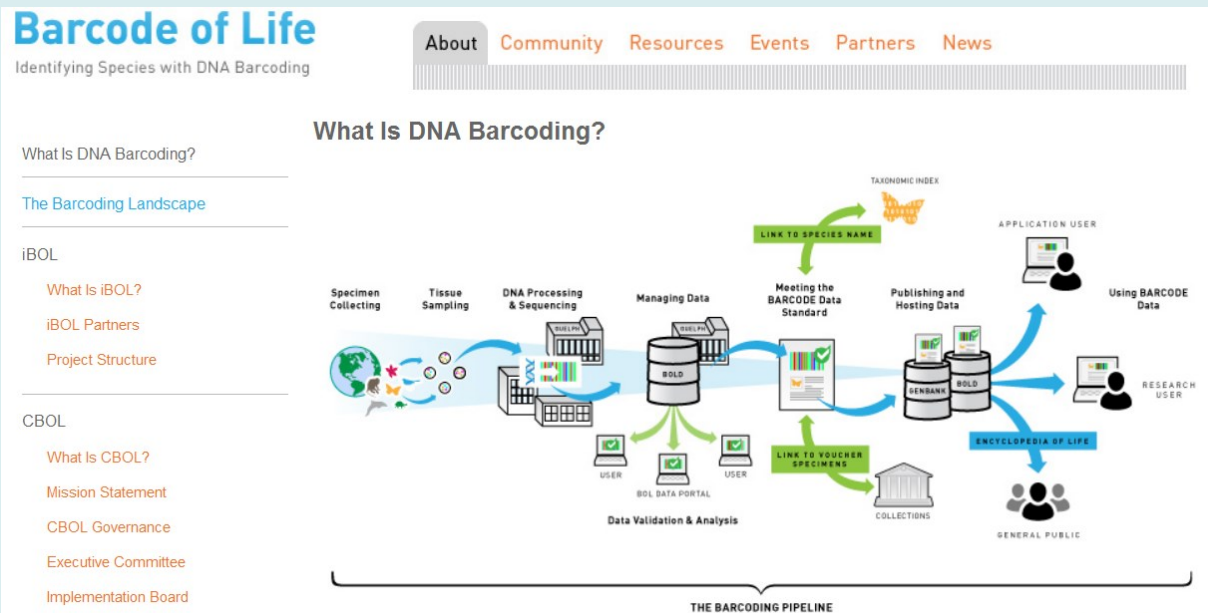
- determinace neznámá sekvence

Identifikace druhů za použití krátké DNA sekvence

- gen s velkou variabilitou mezi druhy a s malou uvnitř druhu
- gen společný pro co největší počet organismů
- COI u většiny Eukaryot, matK a rbcL u rostlin, 16S rRNA u prokaryot
- ITS rDNA u hub (+ RPB1, RPB2, Mcm7, TUB2, LSU, SSU u některých skupin)



<http://www.barcodeoflife.org/content/about/what-dna-barcoding>



Fungal Barcoding

International Fungal Working Group



[Home](#) [Search strain database](#) [Identification](#) [Deposit](#) [Registration](#) [Primers](#) [Publications](#) [Meetings](#) [Links](#) [Help](#) [Unknown user](#)

Pairwise sequence alignment

☐ I have read [the disclaimer](#) and I agree with the conditions and limitations associated with the usage of the software

Pairwise sequence alignment parameters

Minimum similarity to keep results (0-100%):	0	Gap creation penalty (1-100):	12
Minimum overlap to keep results (0-100%):	50	Gap extension penalty (1-10):	2
Minimum overlap for rating reward (0-100%):	85	Word size (1-256):	20
Maximum alignments to display (1-1000):	50	Penalty for a nucleotide mismatch:	-1
Select sorting mode:	Rating	Reward for a nucleotide match:	1

Select the reference(s) file(s) to be used for the alignment:

☐	Name	Location	Status	#Match(s) found
☐	FunBOLD	Local	Reachable	
☐	FUNSTRAI	http://www.q-bank.eu/Fungi/	Reachable	
☐	FUNCBS	http://www.cbs.knaw.nl/Collections/	Reachable	
☐	UNITE	http://www.cbs.knaw.nl/Collections/	Reachable	
☐	Yeast	http://www.cbs.knaw.nl/Collections/	Reachable	
☐	Fusarium	http://www.cbs.knaw.nl/Fusarium/	Reachable	
☐	Calonec	http://www.cbs.knaw.nl/Calonectria/	Reachable	
☐	Indoor.txt	http://www.cbs.knaw.nl/Indoor/	Reachable	
☐	ITSSeq	http://www.mycologylab.org/	Reachable	
☒	PYCC	http://pycc.bio-aware.com/	Reachable	
☐	Non-redundant GenBank	http://blast.ncbi.nlm.nih.gov/Blast.cgi	Reachable	

SEZUAM

Getting Started Latest Headlines

Search NCBI PubMed for

EN -> CZ

Pravopis

E-mail

S-Rank

Nastavení

http://www.cbs.knaw.nl/phaeoacremonium/BioloMICSID.aspx

Přejít

paeoacremonium

BioloMICS

HOME SEARCH IDENTIFICATION DEPOSIT CART GIS MY DATA HELP

Phaeoacremonium -> Phaeoacremonium identifier

Polyphasic identification

Select database:
Phaeoacremonium

Select table:
Phaeoacremonium identifier

First, select the database and the table to be used for the comparisons. Then, **fill in the form below** and **click** on the **"start identification"** button.

start identification

Minimum number of characters to be accounted: 1

Main index (use it to navigate to the wanted set of characters):

1. Morphology

- 1.1. Conidiophores
- 1.2. Phialides
- 1.3. Conidia
- 1.4. Warts

2. Culture characters

3. DNA sequences

1. Morphology

Hotovo

start

D...

8 F

5

M...

Br...

A...

CS

16:47



BARCODING

Mycoscience (2006) 47:55–64
DOI 10.1007/s10267-006-0279-7

© The Mycological Society of Japan and Springer-Verlag Tokyo 2006

REVIEW

Irina S. Druzhinina · Alex G. Kopchinskiy
Christian P. Kubicek

The first 100 *Trichoderma* species characterized by molecular data

IDENTIFICATION:
Morphological Key

TrichoKEY **NEW!**
TrichoBLAST

SEQUENCE LIBRARY:
Multiloci Database
Loci Distribution

BIOLOGY:

Morphology
Phylogeny
Ecology
Biocontrol
Evolution
Holomorphs

METHODS:

Molecular
Phylogenetic
Phenetic
Cultivation
Microscopy
Other

Links

Software **NEW!**
Forum

Analysing sequence: **Non-fasta_sequence**

First anchor (GSH) was found in position **89**
Second anchor (GSH) was found in position **110**
Third anchor (GSH) was found in position **257**
Fourth anchor (GSH) was found in position **415**
Fifth anchor (GSH) was found in position **519**

Found **5** genus-specific hallmarks (Anchors):

Genus Identification: *Hypocrea/Trichoderma*, *Hypocreaceae*, *Hypocreales*, *Ascomycota*

region 1 (8nt) was detected; region 2 (141nt) was detected; region 3 (188nt) was detected; 5.8S RNA gene was removed

Section Identification: **Section Trichoderma**

Section Identification: **XII Rufa Clade**


Species identification:

Hypocrea atroviridis/Trichoderma atroviride

Identification reliability: **high**

barcode was developed on the basis of **46** vouchered sequences which showed **1** ITS 1 and 2 alleles

BLAST N AGAINST *HYPOCREA/TRICHODERMA* DATABASE: >>

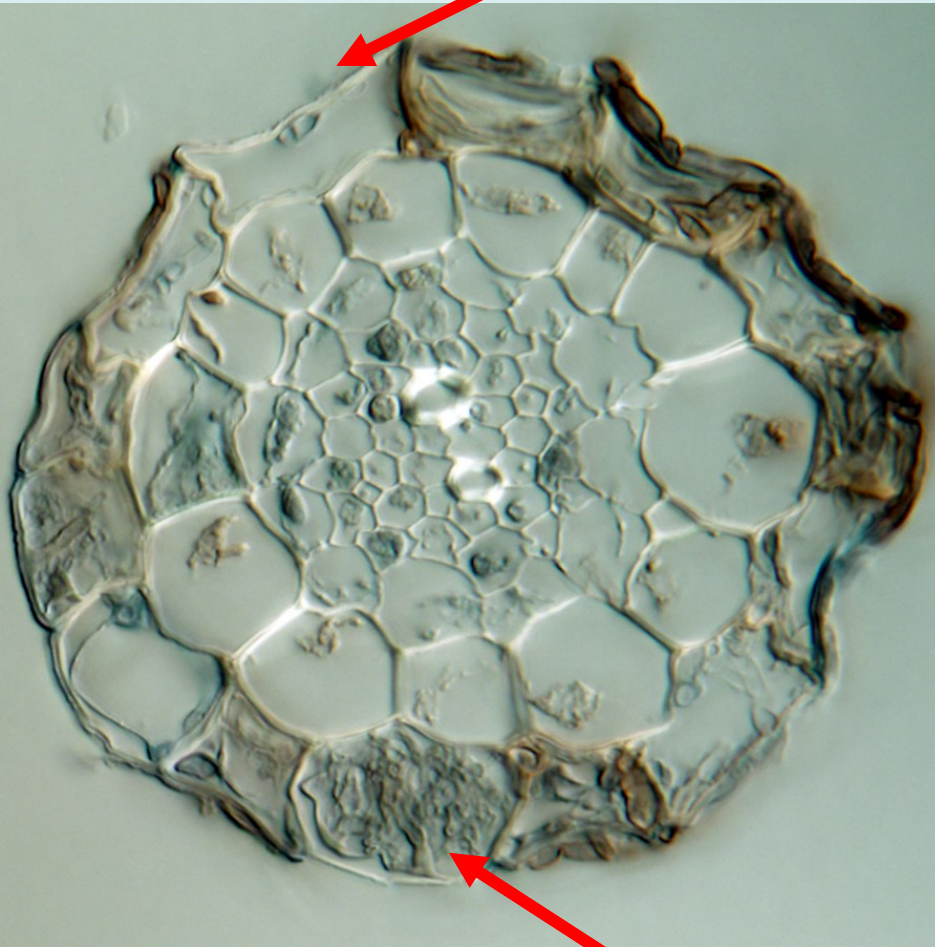
Query sequence: AAGTCGTAACAAGGTCTCCGTTGGTGAACAGCGGAGGGATCATTACCGAGTTTACAACTCCCAAAACCCA
ATGTGAACCATACCAAACTGTTGCCCTCGGCGGG [GTCACG] CCCCCGG [TGCGTCGCG] CCCCCG
AACCAGGCGCCGCGGAGGGACCAACCA [AACTCTTTTCTGTA] GTCCCTCGCGGACGTTATTTCT
TACAGCTCTGAGCAAAAATTCAAAATGAATCAAACTTTCAACAACGGATCTCTTGGTTCTGGCATCGAT
GAAGAACGCGACGAAATGCGATAAGTAATGTGAATTGCAGAAATTCAGTGAAATCATCGAATCTTTGAACGC
ACATTGCGCCCGCCAGTATTCTGGCGGGCATGCCTGTCCGAGCGTCATTTCAACCCCTCGAAACCCCTCCGG
GGGGTCGGCGTTGGG [GACCTCGGGAGC] CCCTAAGACGGGATCCCGGCCCGAAATACAGTGGCGGT
CTCGCCCGAGCCTCTCTGCGCAGTAGTTTGCACAACTCGCACCGGGAGCGCGCGCTCCACGTCCGTA
AAACACCCAACTTCTGAAATGTTGACCTCGGATCAGGTAGGAATACCCGCTGAACCTTAAGCATATCA

Příčiny inkongruence

Podobnosti druhů se mohou lišit v závislosti na použitém genu

- Paralogie
- výskyt pseudogenu
- heterogenita v nukleotidovém složení
- odlišná rychlost evoluce (heterotachie)
- incomplete lineage sorting
- hybridizace

no hyphal sheath



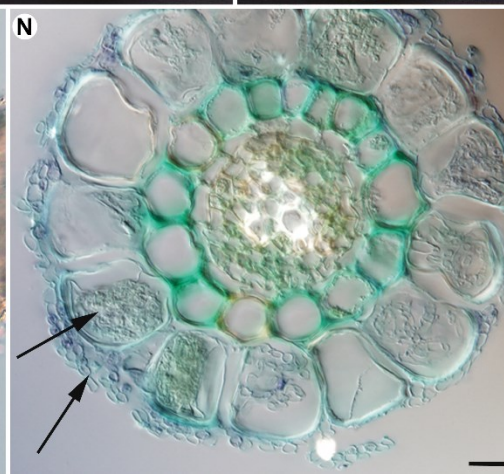
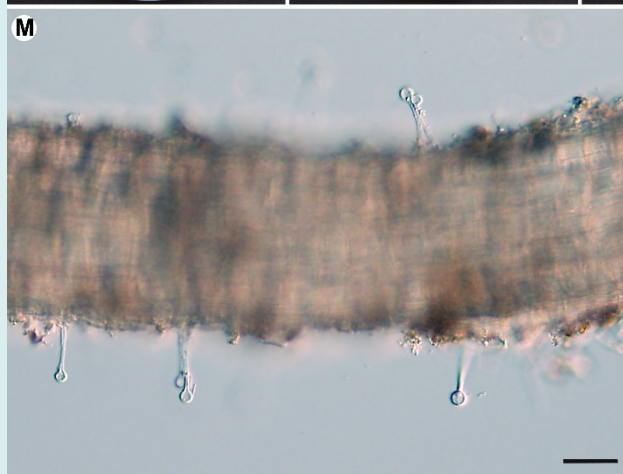
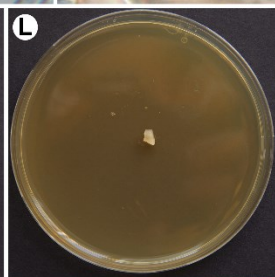
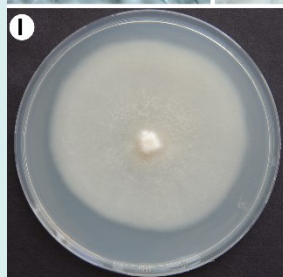
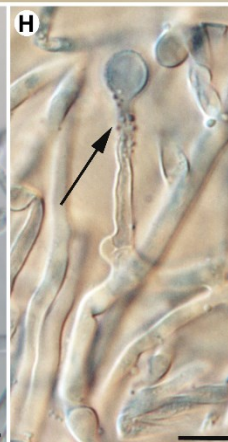
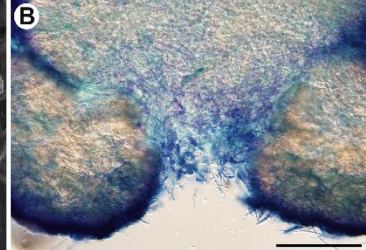
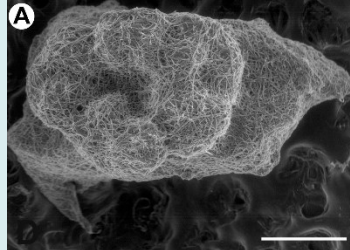
hyphal sheath

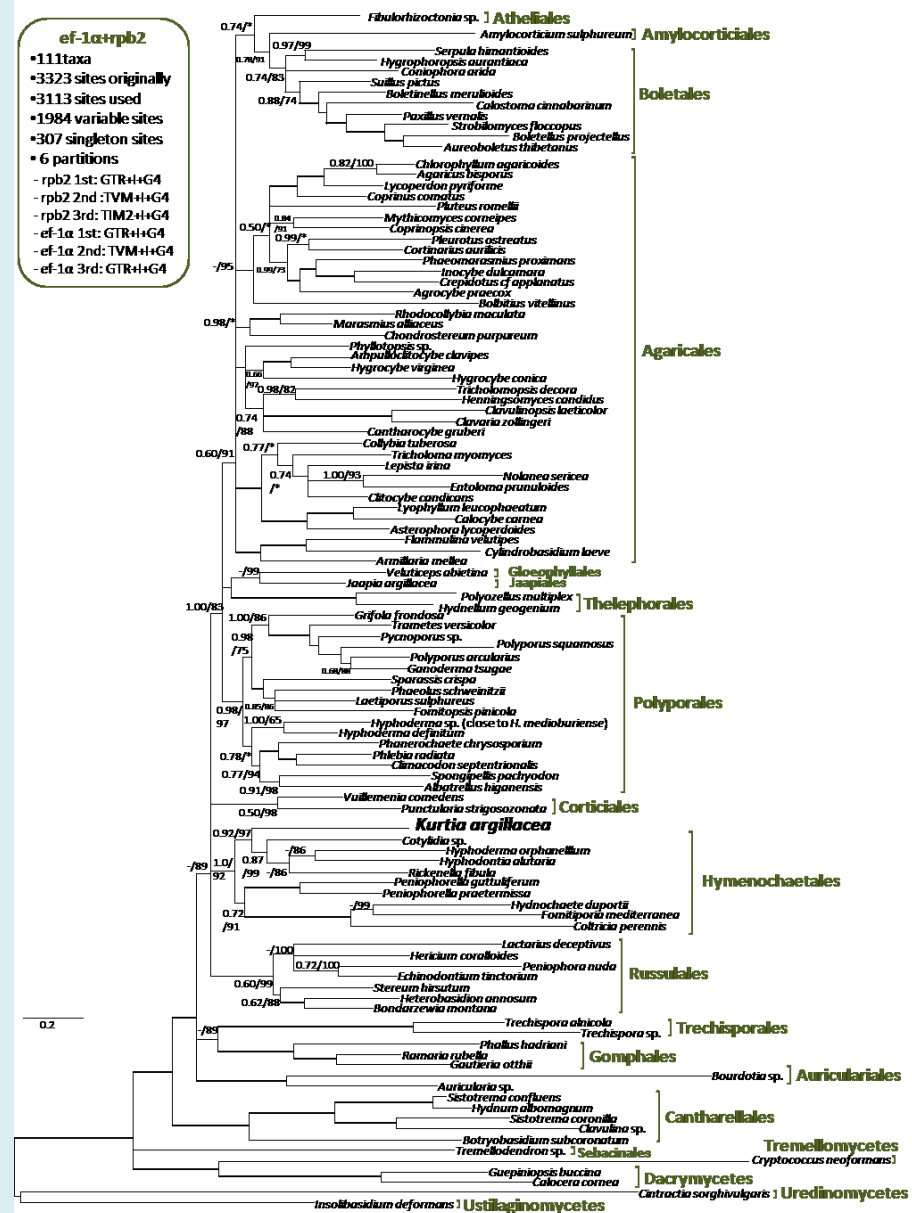
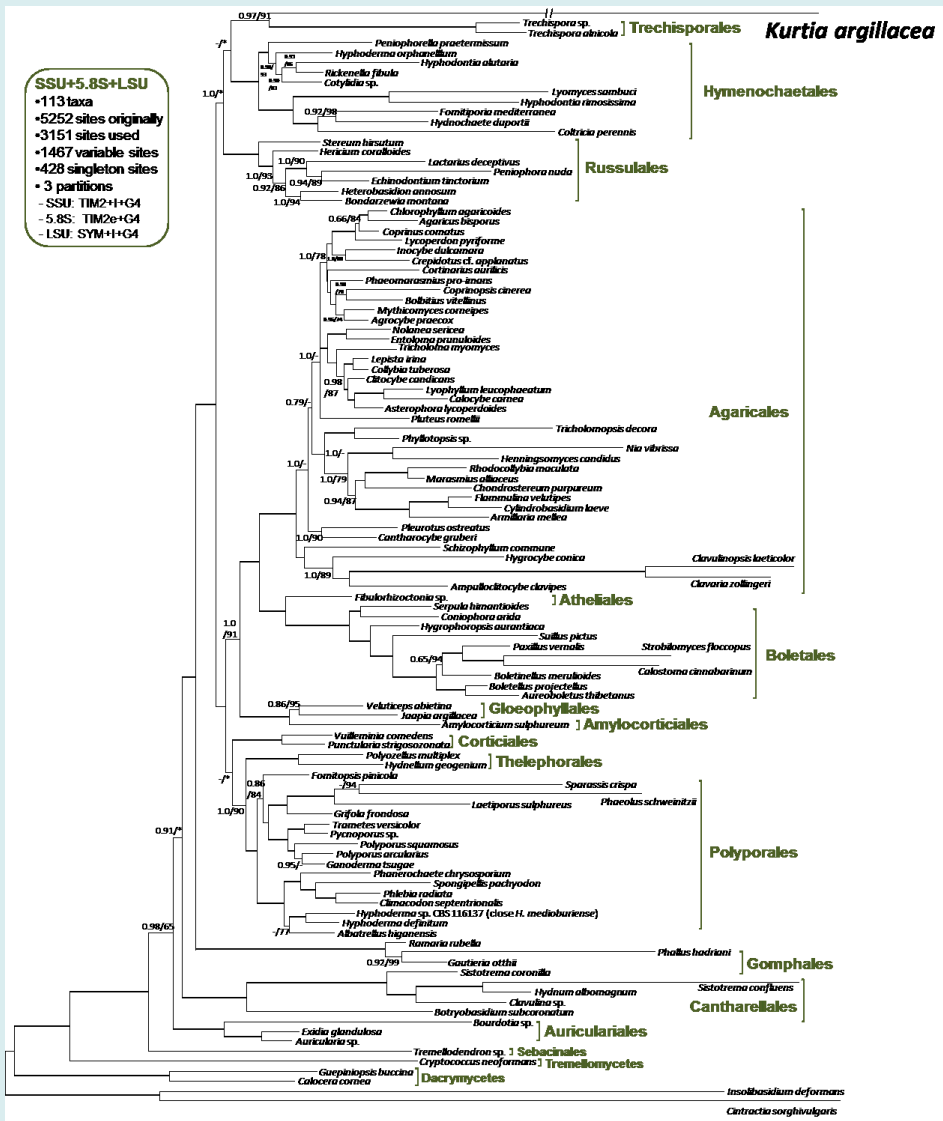


intracellular hyphal coils (no intercellular colonization)

Vaccinium, **typical ericoid mycorrhiza**
(multiple fungi)

Vaccinium, **sheathed ericoid mycorrhiza**

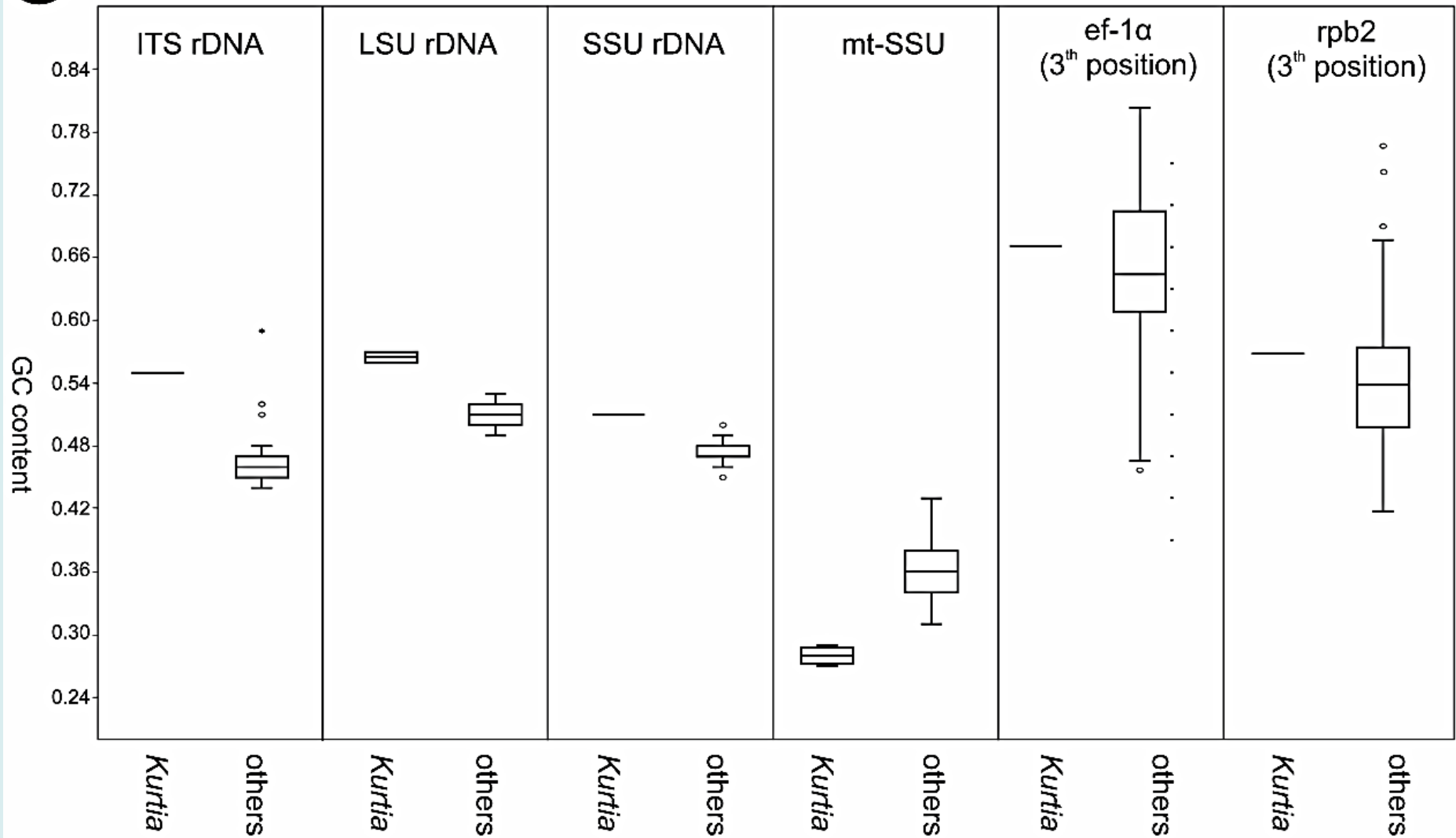




taxon	nuclear ITS region length [bp]			ITS2 secondary structure energy [kcal/mol]	nuc-LSU length [bp]	nuc-SSU length [bp]
	IITS1	5.8S	ITS2			
<i>Kurtia</i>	195	167	226	-42,5	953	1748
Other fungi^a	200(217. 7)242	162(162,3) 163	187(205.7) 247	-8.9(-57.9)- 74.8)	916(932.6)938	1672(17 37)1754

Heterogenita v nukleotidovém složení

A



Heterogenita v nukleotidovém složení

- Každý z „life history-related traits“ má vliv na substituční rychlost, ale i na nukleotidové složení
- Nukleotidové složení jako výsledek neutrálních procesů, bez adaptivní funkce ?
- **gBGC** – GC-biased gene conversion
 - objeveno u kvasinek
 - preferenční fixace GC alel u AT/GC heterozygotů
 - větší v místech s častou rekombinací
 - tj. souvisí s pozicí na chromozomu, generační dobou a pohlavností

Weber et al. 2014. Evidence for GC-biased gene conversion as a driver of between-lineage differences in avian base composition. *Genome Biology* **15**: 549.

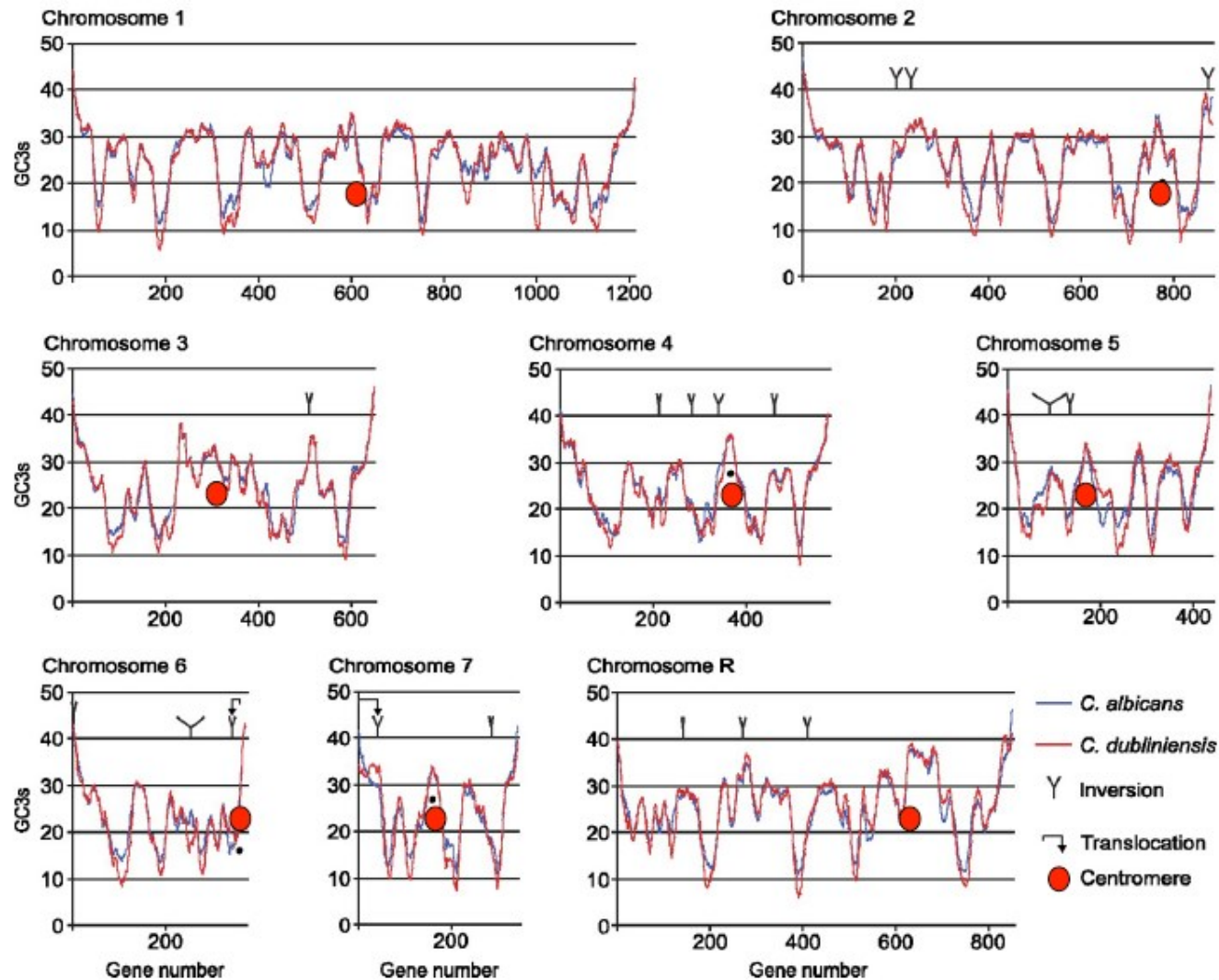


Analytical Biases Associated with GC-Content in Molecular Evolution

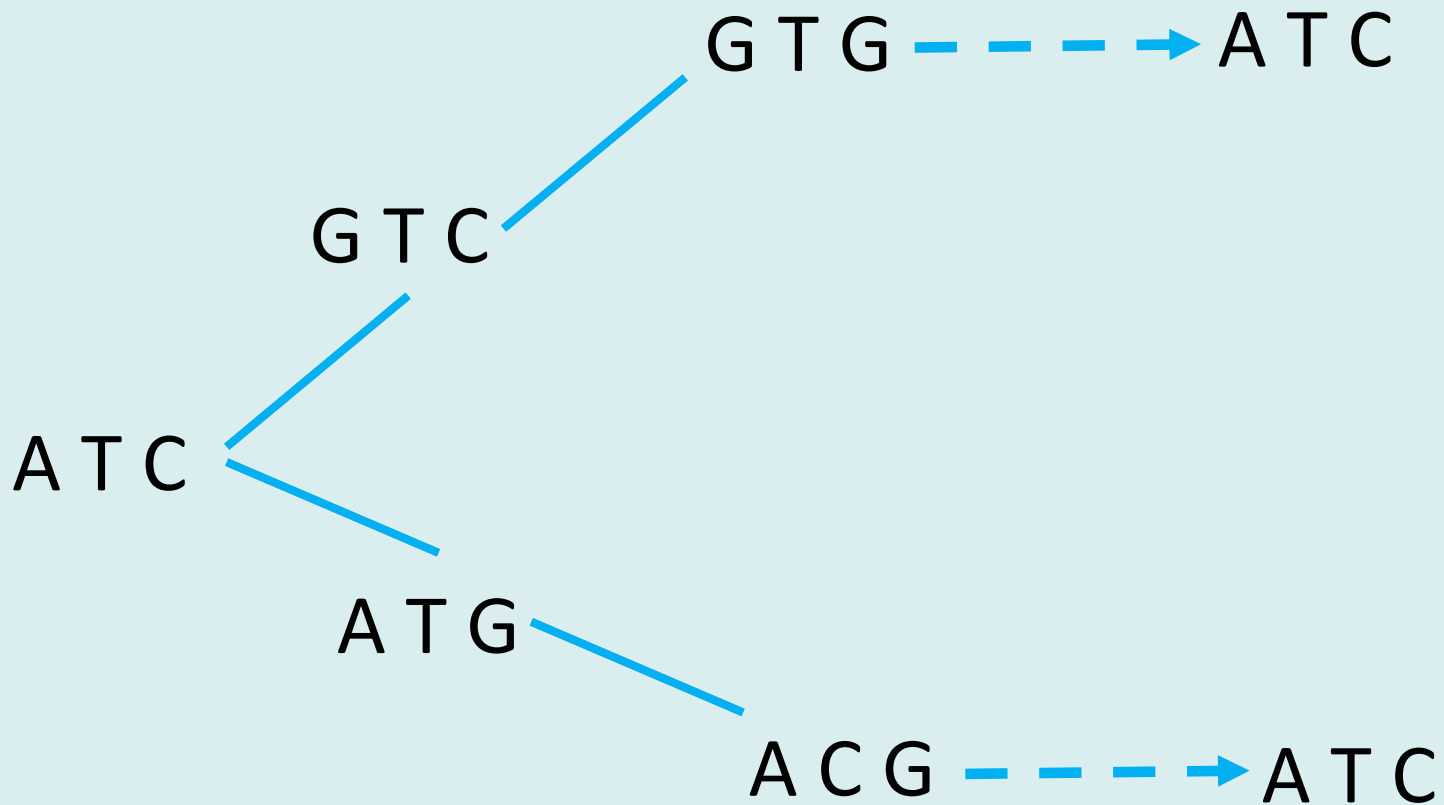
Jonathan Romiguler and Camille Roux*

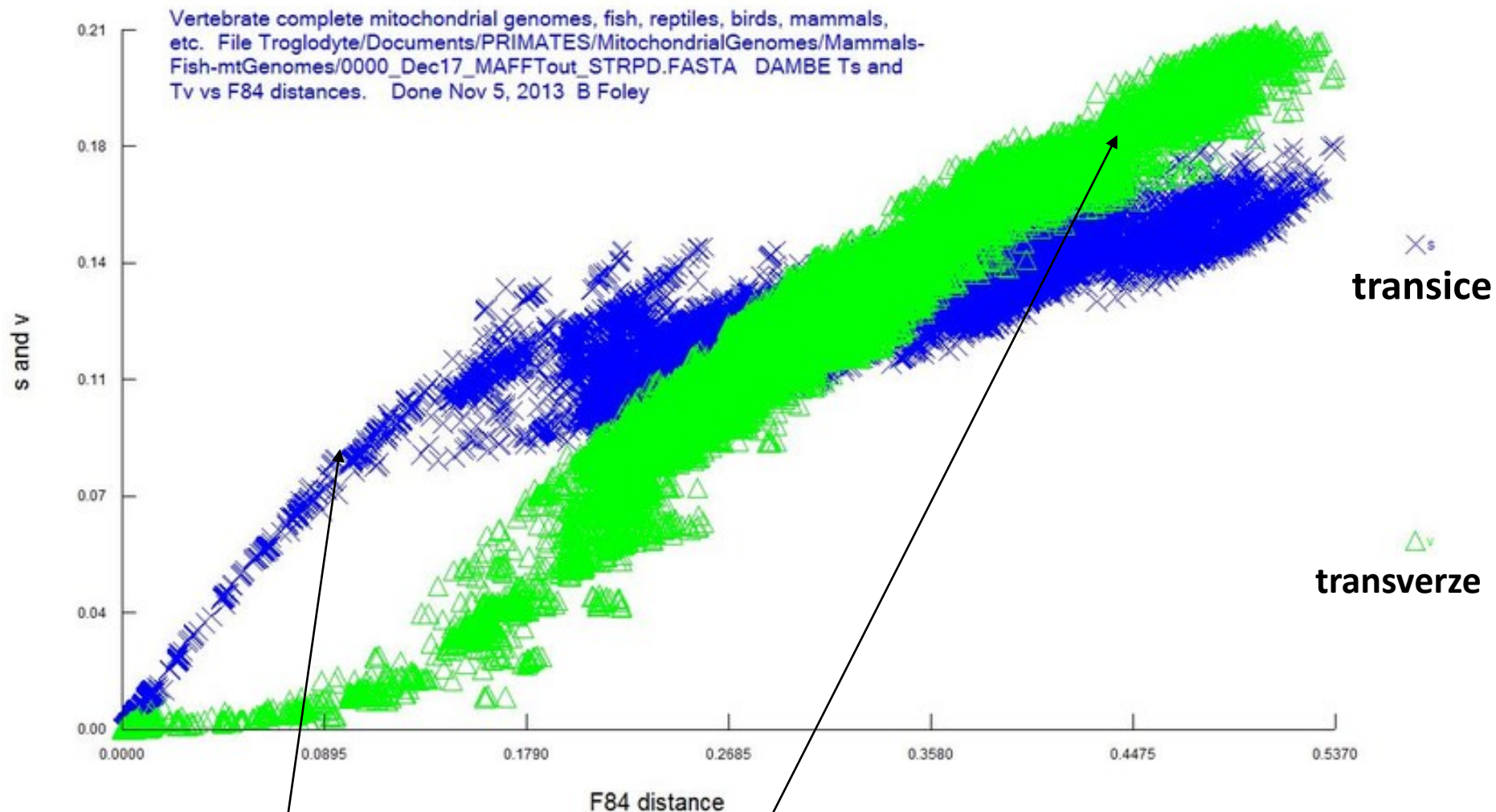
Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland

Chromosomal GC3s profiles in *Candida albicans* and *C. dubliniensis*



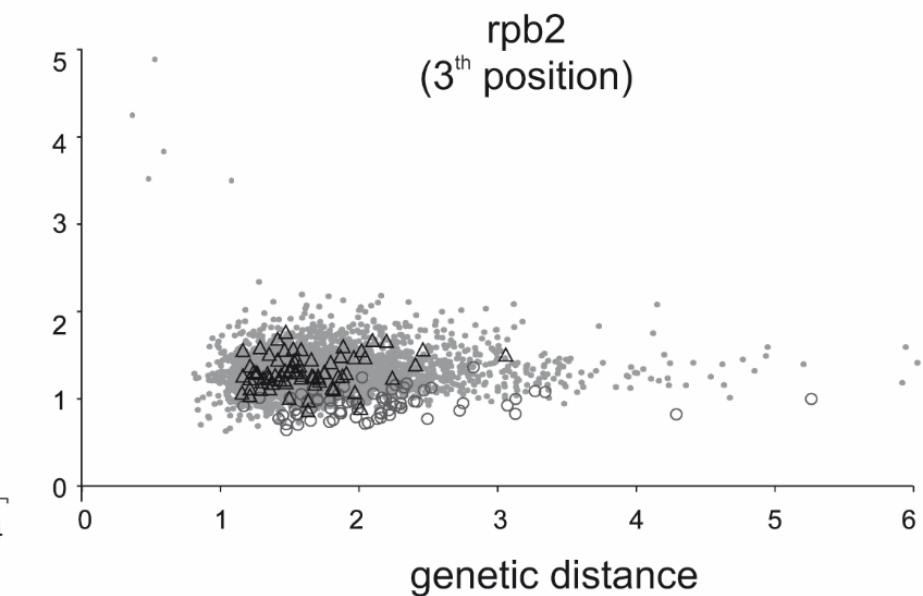
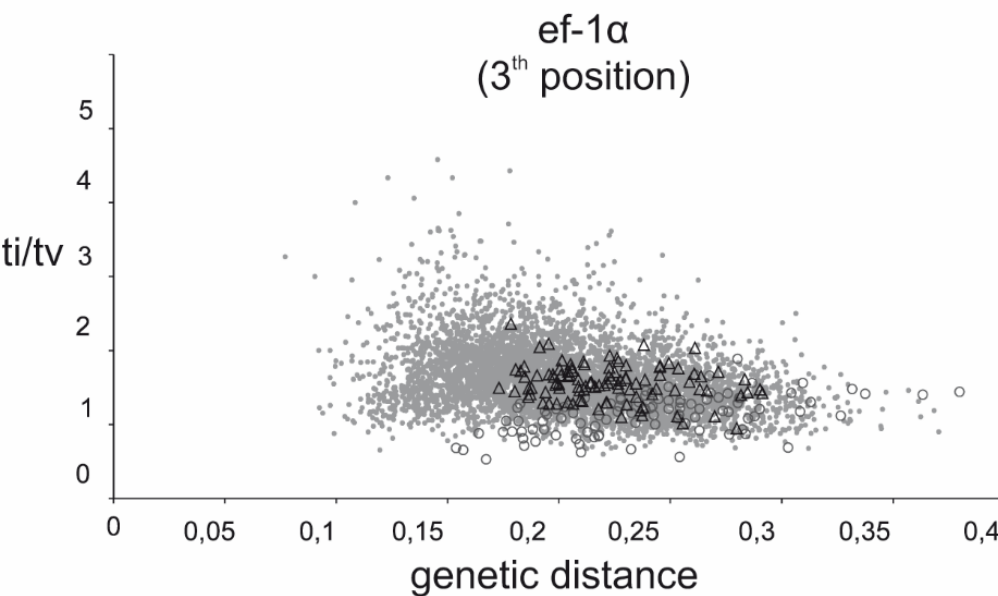
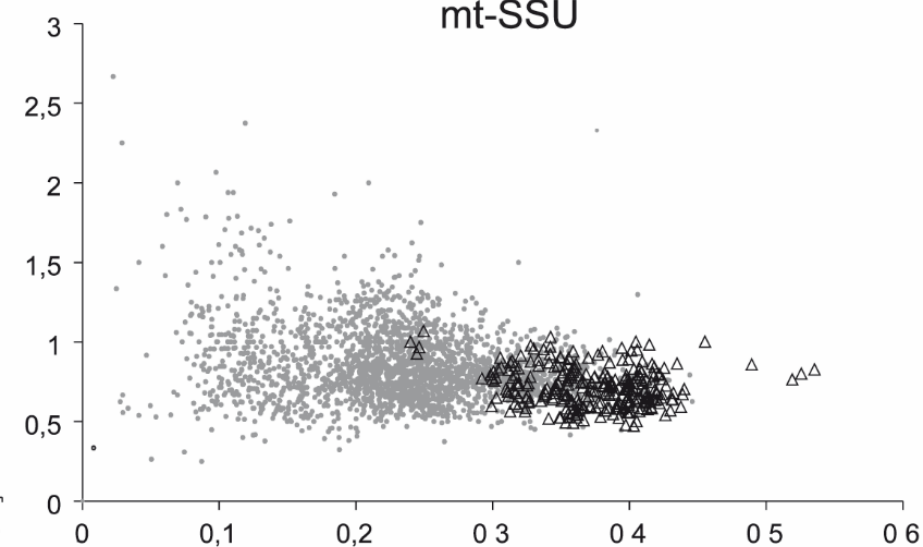
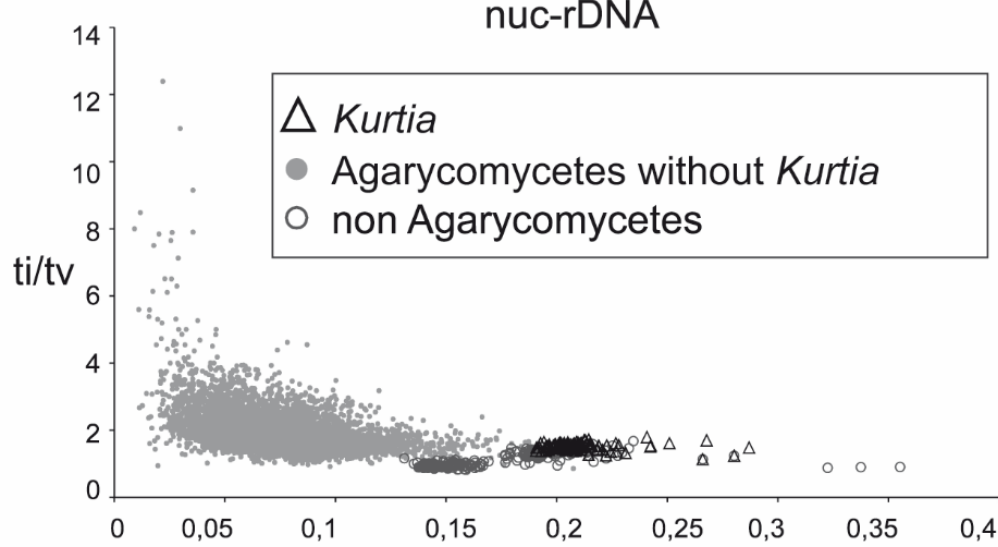
- nukleotidová (genetická) saturace





Transice je více častá než transverze

S narůstajícím evolučním časem (genet. distancí) je větší šance, že dojde k transverzi

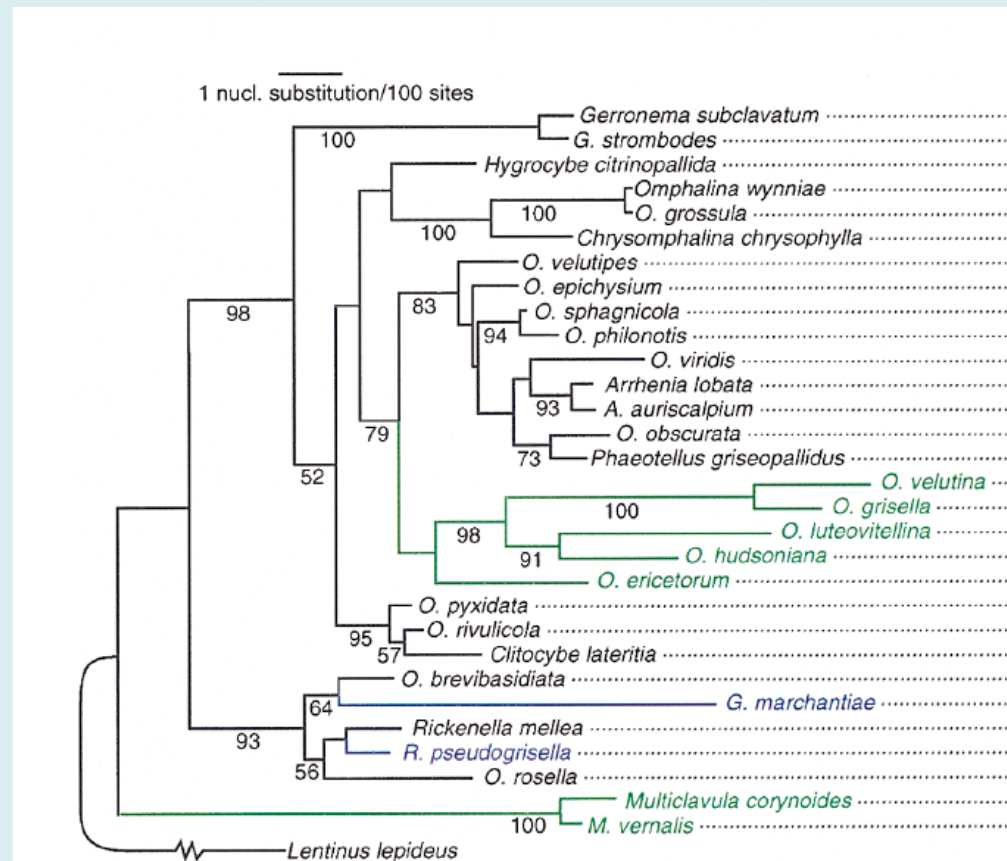


Podíl transicí a transverzí

Odlišná rychlost evoluce a nukleotidová saturace

- Marais GAB, Calteau A, Tenaillon O, 2008. Mutation rate and genome reduction in endosymbiotic and free-living bacteria. *Genetica* 134: 205-210.
- Lutzoni F, Pagel M, 1997. Accelerated evolution as a consequence of transitions to mutualism. *Proceedings of the National Academy of Sciences of the United States* **94**: 11422-11427.

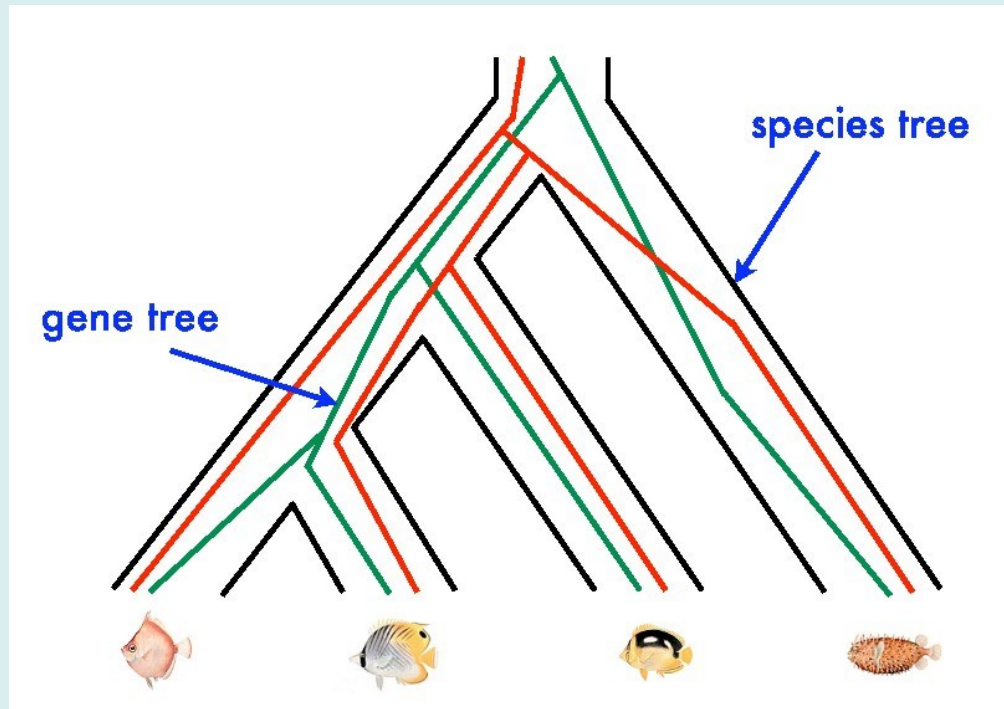
- Redukce genomu u mutualistů
- Ztráta reparačních mechanismů
- bottleneck efekt
- kratší generační čas
- malá populační velikost
- asexuální reprodukce
- volné radikály z hostitele



Konflikty při fylogenetických analýzách

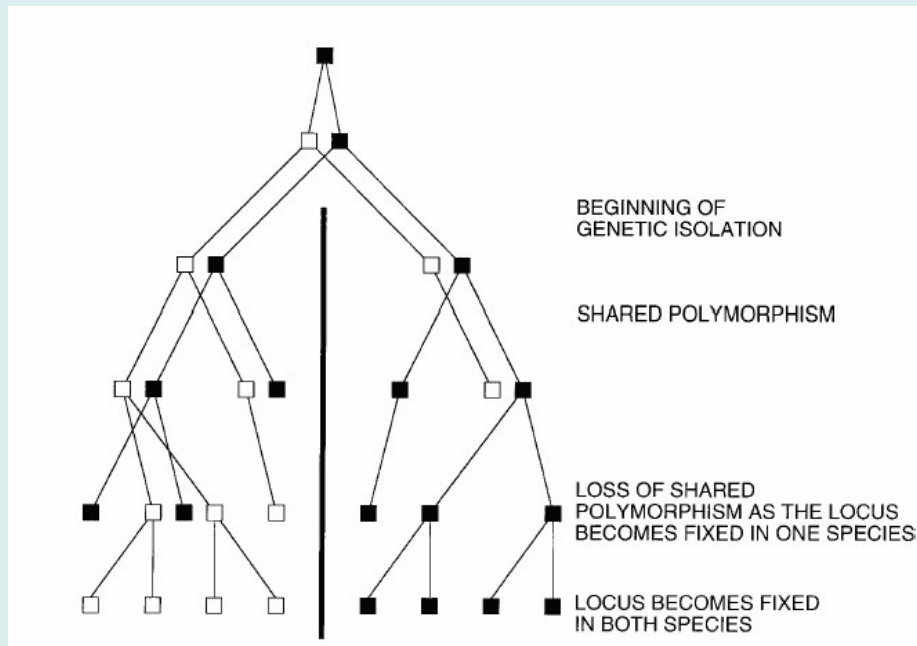
Genový strom x druhový strom

- genealogie jednotlivých genů se mohou lišit
- výsledkem je špatné vyhodnocení vzájemné podobnosti druhů (většinou jen v rámci jednoho rodu)
- řeší mnohagenová analýza



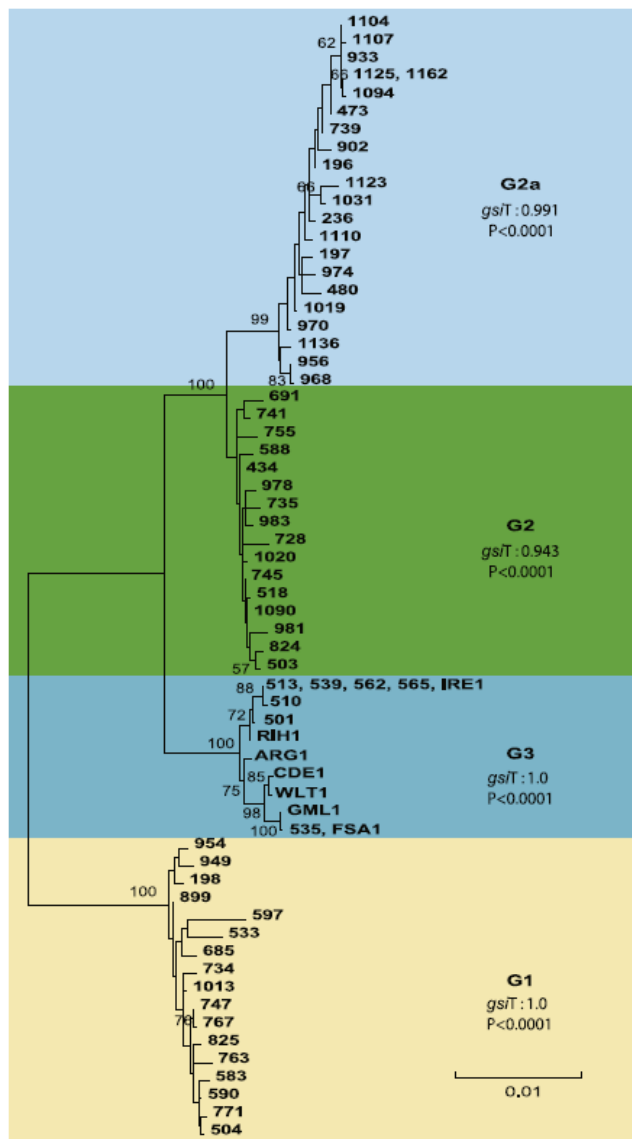
Konflikty při fylogenetických analýzách

- V populaci se vyskytuje mnoho alel jednoho genu (např. jedinci s různou sekvencí ITS)
- Dojde-li zde ke vzniku biologického druhu, tak se důsledná izolace projeví fixací určité alely pro každý druh. Každý druh má pak například specifickou sadu příbuzných ITS sekvencí.
- U mladých druhů není „sortování linií genů“ kompletní - ***incomplete lineage sorting***
- Ve stavu **ancestrálního polymorfismu** sdílí oba nové druhy společné alely, které mohou mít odlišnou genealogii

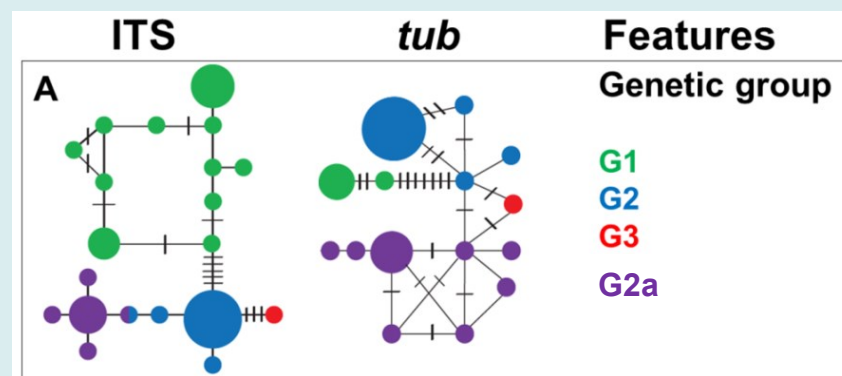




0.1



- *Claviceps purpurea* – 4 kryptické druhy
- ITS je ve stavu ancestrálního polymorfismu u *C. humidiphila* vs *C. arundinis* – některé jedince nelze odlišit pomocí ITS
- Při mnohagenových analýzách je možné takový gen vynechat



Delimitation of cryptic species inside *Claviceps purpurea*

Sylvie PAŽOUTOVÁ^{a,†}, Kamila PEŠICOVÁ^{a,b}, Milada CHUDÍČKOVÁ^a, Petr ŠRŮTKA^c, Miroslav KOLARÍK^{a,*}

- Species pod velkým či nestálým selekčním tlakem mají unikátní genealogii

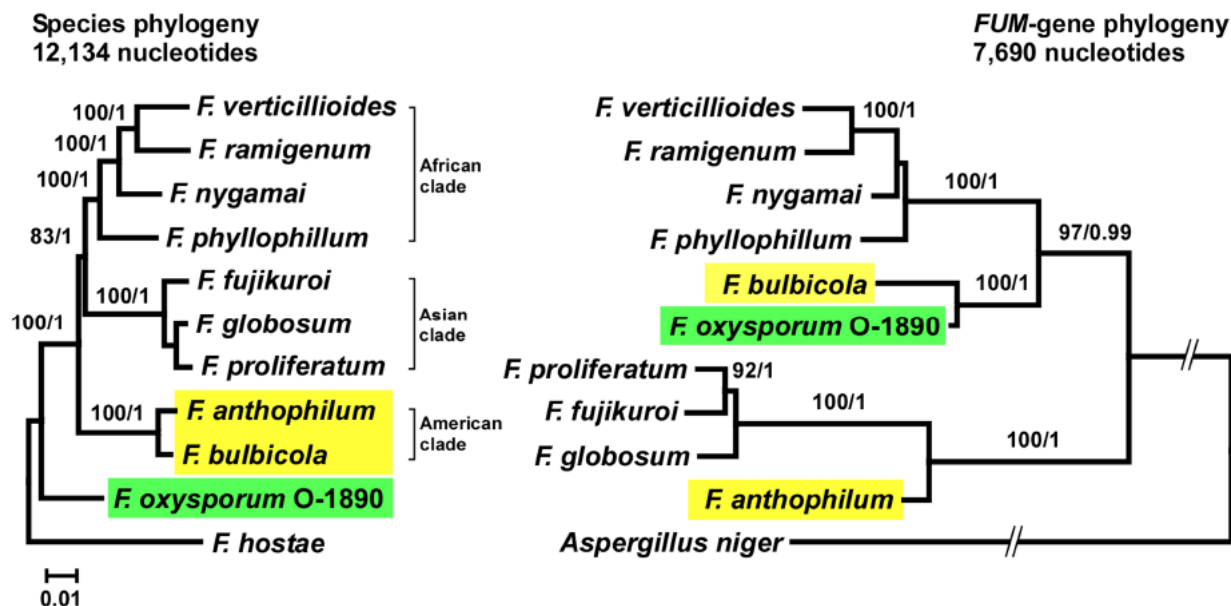


Fig. 2. Species and *FUM* gene phylogenies based on maximum likelihood analysis of concatenated nucleotide sequences of 12 PM and nine *FUM* genes amplified from fumonisin-producing species of *Fusarium*. Numbers near branches indicate bootstrap values based on 1000 pseudoreplicates from maximum likelihood analysis (left of slash mark) and posterior probabilities from Bayesian analysis (right of slash mark). GenBank accession numbers for sequences generated and employed in this analysis are KF416128–KF415191 (*FUM* genes) and KF466326–KF466456 (PM genes).

Birth, death and horizontal transfer of the fumonisin biosynthetic gene cluster during the evolutionary diversification of *Fusarium*

Robert H. Proctor,¹* François Van Hove,² Antonia Susca,³ Gaetano Stea,³ Mark Busman,¹ Theo van der Lee,⁴ Cees Waalwijk,⁴ Antonio Moretti^{3†} and Todd J. Ward¹

¹United States Department of Agriculture[†], Agricultural Research Service, National Center for Agricultural Utilization Research, Peoria, IL, USA.

²Université catholique de Louvain, Earth and Life Institute, Applied Microbiology, Mycology, Mycotoxins de l'Université catholique de Louvain (BCCMTM/MUCL), Louvain-la-Neuve, Belgium.

³National Research Council, Institute of Sciences of Food Production, Bari, Italy.

⁴Wageningen UR, Melle, The Netherlands.

with two competing hypotheses: (i) multiple horizontal transfers of the cluster from unknown donors to FFSC recipients and (ii) cluster duplication and loss (birth and death). Furthermore, low levels of *FUM* gene divergence in *F. bulbicola*, an FFSC species, and *F. oxysporum* provide evidence for horizontal transfer of the cluster from the former, or a closely related species, to the latter. Thus, uniform gene organization within the *FUM* cluster belies a complex evolutionary history that has not always paralleled the evolution of *Fusarium*.

Introduction

The first step in the evolution of *Fusarium* is the acquisition of the

Konflikty při fylogenetických analýzách

- Horizontální transfer vs *Birth and death evolution*

300 R. H. Proctor et al. ■

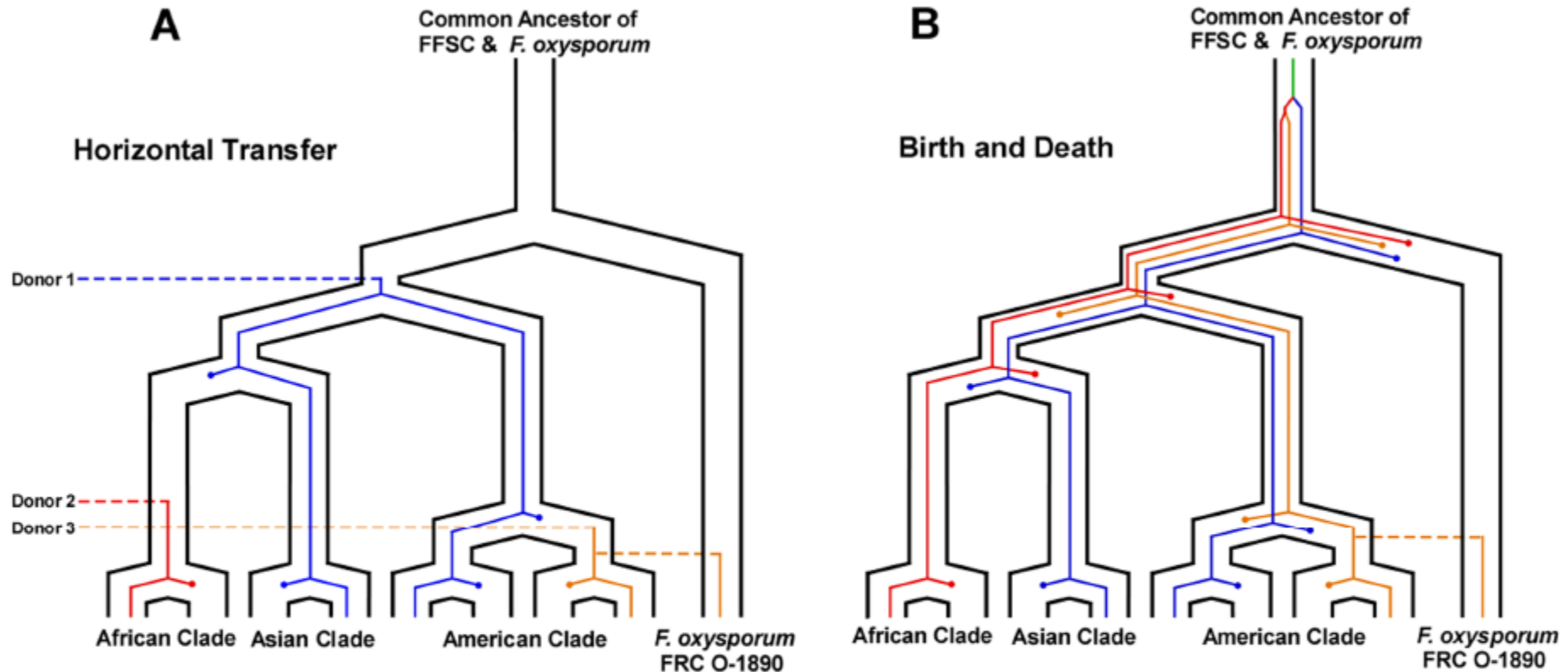


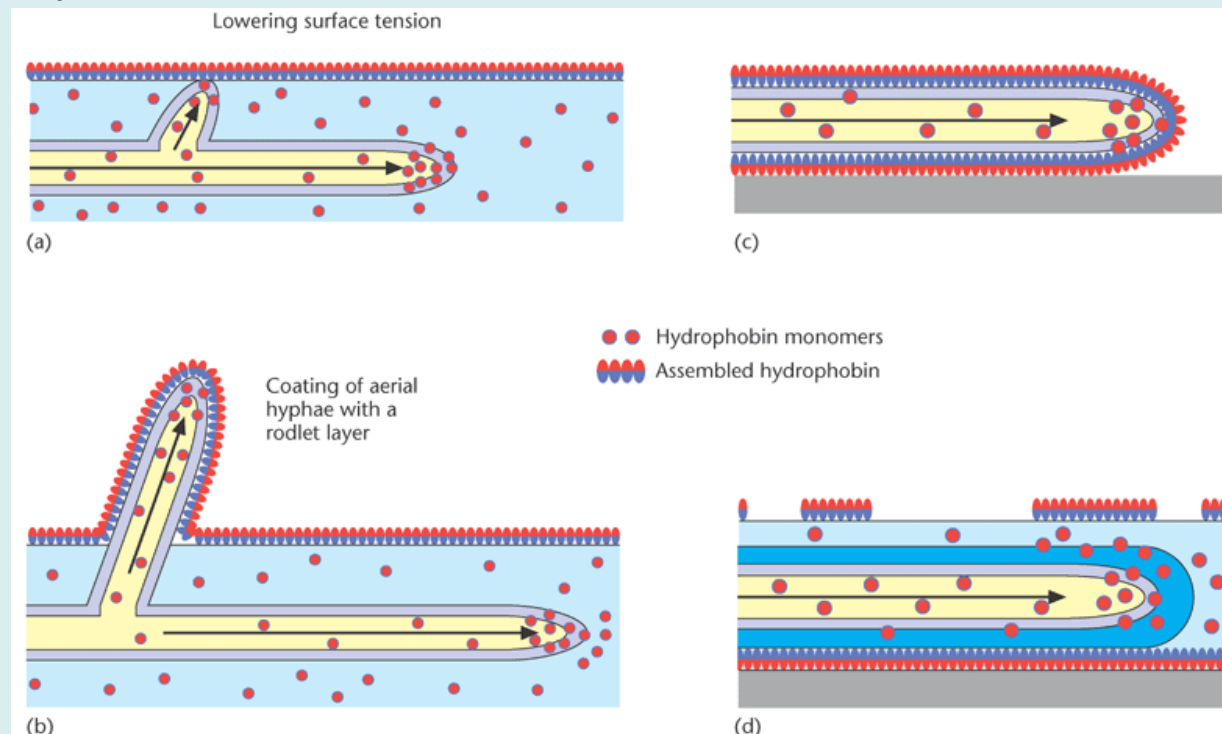
Fig. 5. Models for the evolutionary history of the *FUM* cluster during divergence of the FFSC and *F. oxysporum* O-1890.

A. Model based on independent horizontal transfer of the *FUM* cluster from three unknown but related donors to three lineages within the FFSC.

B. Model based on duplication followed by retention or loss (birth and death) of the *FUM* cluster. In both models, thick black lines depict the divergence of *Fusarium* lineages from the common ancestor of the FFSC and *F. oxysporum*; solid coloured lines represent evolution of the *FUM* cluster during divergence of the *Fusarium* lineages; closed circle indicate loss of a cluster homologue; and the dashed coloured line indicates horizontal duplication. In the birth-and-death model, bifurcations of coloured lines prior to divergence of the FFSC and *F. oxysporum* indicate cluster duplication events.

Konflikty při fylogenetických analýzách

- Proteiny zodpovědné za adaptaci (často povrchové proteiny) podléhají velké selekci
- hydrofobiny hub, geny pro syntézu toxinů...
- assemble spontaneously into amphipathic monolayers at hydrophobic–hydrophilic interfaces
- crucial in the interaction with the environment, masking
- pathogenicity factors



Konflikty při fylogenetických analýzách

- Případem jsou hydrofobiny hub

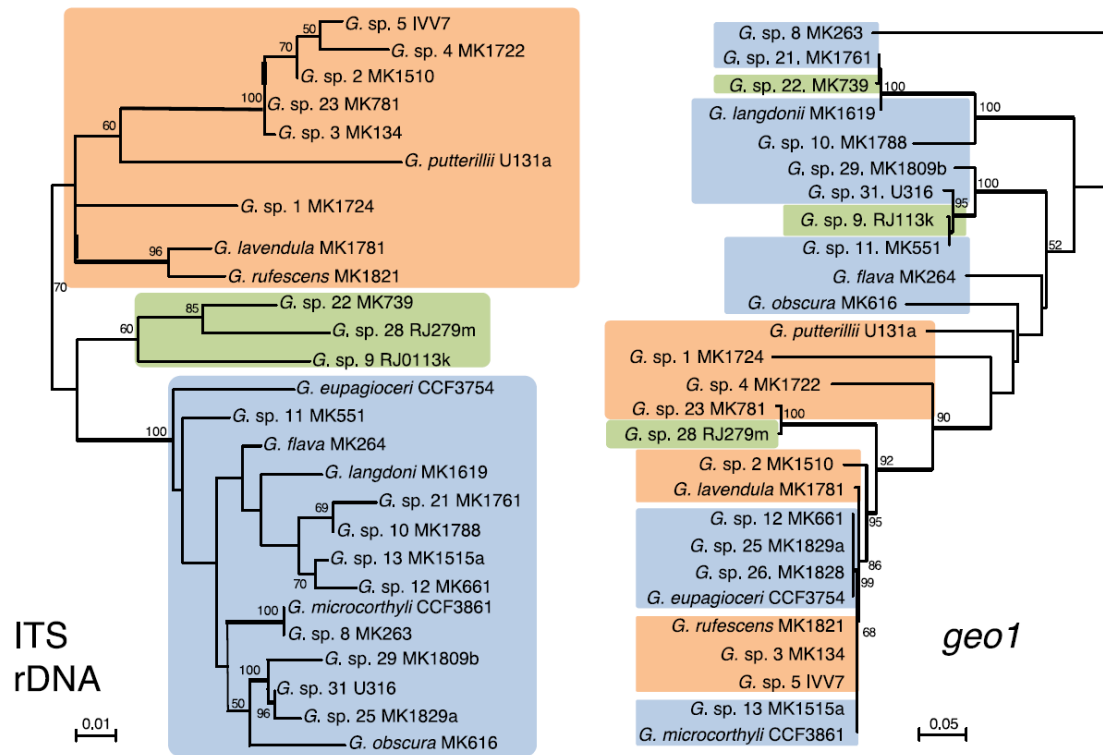


Fig 1 – Maximum likelihood trees of the *Geosmithia* species used in this study based on ITS rDNA gene or on the genomic sequences of the *geo1* gene. Phylogenetic trees were constructed with PhyML 3.0 performing bootstrap analysis with 500 replicates. Thicker lines indicate strongly supported branches (bootstrap values > 90).

Interspecific variability of class II hydrophobin GEO1 in the genus *Geosmithia*

Arcangela FRASCELLA^{a,b}, Priscilla P. BETTINI^{a,*}, Miroslav KOLAŘÍK^c, Cecilia COMPARINI^b, Luigia PAZZAGLI^d, Simone LUTI^d, Felice SCALA^e, Aniello SCALA^b

- geny zodpovědné za adaptaci (často povrchové proteiny) mají unikátní evoluci
- *Intragenic tandem repeats* (ITR) zodpovědné za rychlou generaci variability
- Například u *Saccharomyces cerevisiae* umožňují ITR velmi rychlé změny v adhezi k substrátům.

Geosmithia sp. 8, MK263
Geosmithia sp. 1, MK1724
G. langdonii, MK1619
Geosmithia sp. 21, MK1761
Geosmithia sp. 22, MK739
Geosmithia sp. 10, MK1788
Geosmithia sp. 29, MK1809b
Geosmithia sp. 31, U316
Geosmithia sp. 9, RJ131k
Geosmithia sp. 11, MK551
G. putterillii, U131a
G. flava, MK264
Geosmithia sp. 23, MK781
Geosmithia sp. 28, RJ279m
G. lavendula, MK1781
Geosmithia sp. 12, MK661
Geosmithia sp. 25, MK1829a
Geosmithia sp. 3, MK134
Geosmithia sp. 5, IVV7
Geosmithia sp. 13, MK1515
Geosmithia sp. 26, MK1828
G. eupagioceri
G. microcorthyli
G. rufescens, MK1821
Geosmithia sp. 2, MK1510
G. obscura, MK616
Geosmithia sp. 4, MK1722

Fig 2 – MUSCLE alignment of the deduced GEO1 amino acid sequences from the different *Geosmithia* species. Strain number is indicated after the species name. The eight cysteine residues are highlighted in grey. The sequence of the signal peptide is underlined in the reference species *Geosmithia* sp. 5 strain IVV7. Symbols: (*) indicates invariant amino acid positions, (:) indicates conserved substitutions, (.) indicates semi-conserved substitutions.

Interspecific variability of class II hydrophobin GEO1 in the genus *Geosmithia*

Arcangela FRASCELLA^{a,b}, Priscilla P. BETTINI^{a,*}, Miroslav KOLÁŘÍK^c,
Cecilia COMPARINI^b, Luigia PAZZAGLI^d, Simone LUTTI^d, Felice SCALA^e,
Aniello SCALA^b

Horizontální přenos genů a mezidruhov^á hybridizace

- vektorem může být plazmid, transpozom, virus - přenesení několika genů
- mezidruhov^á hybridizace – přenos větších genomických úseků
- β -laktamy (peniciliny) u *Penicillii* jsou zřejmě bakteriální
- důkaz přenosu genů mezi „rumen fungi“ (anaerobní chytridie) a bakteriemi v bachoru přežvýkavců
- Mezidruhoví hybridi v komplexu *Epichloe* – *Neotyphodium*
- *Phytophthora* sp., *Ophiostoma*, *Melampsora*, *Tilletia*
- velmi dobře popsáno u kvasinek

Qiu et al. *BMC Biology* (2016) 14:41
DOI 10.1186/s12915-016-0264-3

BMC Biology

RESEARCH ARTICLE

Open Access



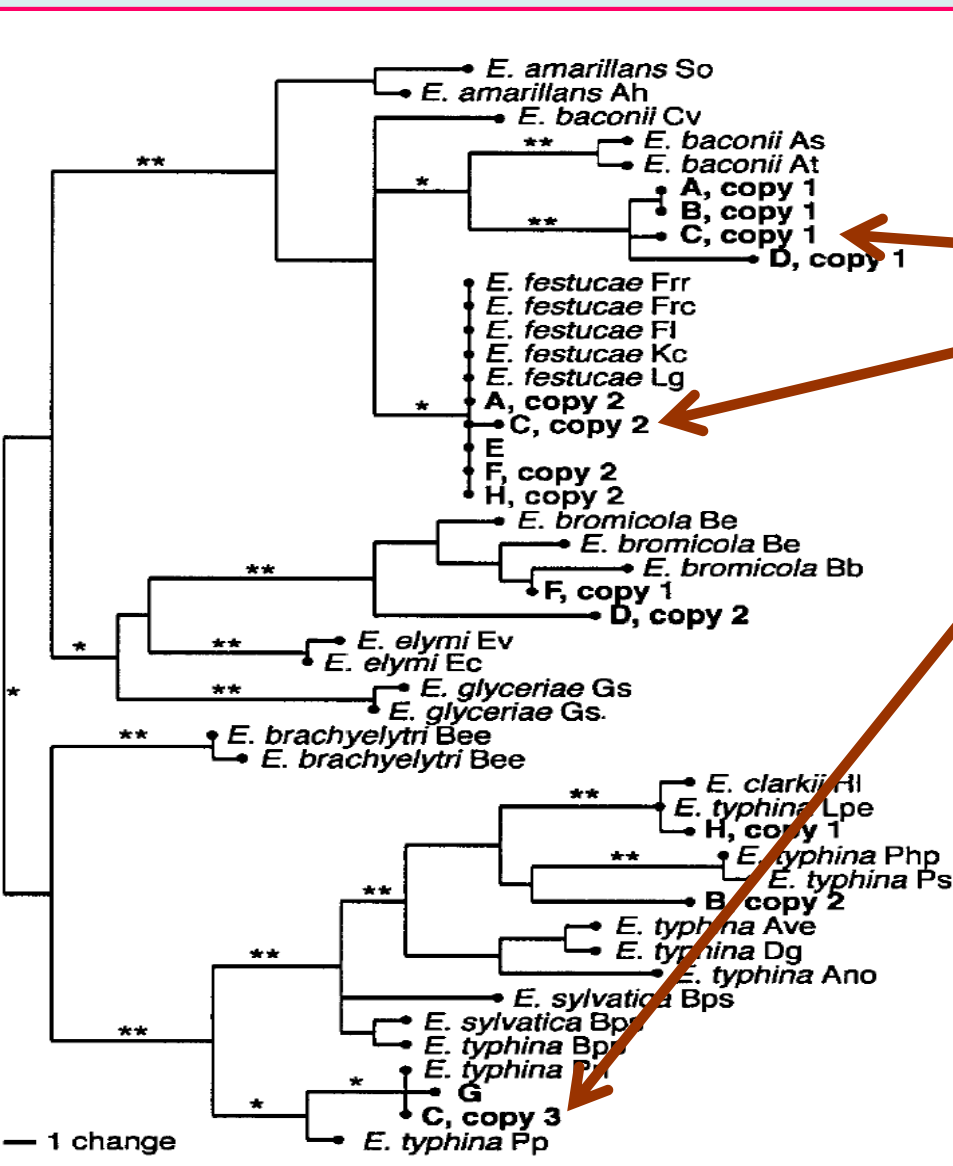
Extensive horizontal gene transfers between plant pathogenic fungi

Huan Qiu^{1*}, Guohong Cai², Jing Luo³, Debashish Bhattacharya¹ and Ning Zhang^{3,4*}

Interspecifická hybridizace

Asexuální endofyty

Beta – tubulin



A - *Neotyphodium* sp.

B - *Neotyphodium* sp.

C - *N. coenophialum*

D - *N. occultans*

E - *N. lolii*

F - *N. siegelii*

G - *N. uncinatum*

H - *Neotyphodium* sp.

Parentální druhy:

E. festucae

E. baconii

E. typhina

Studium diverzity

Výběr kmenů pro analýzu

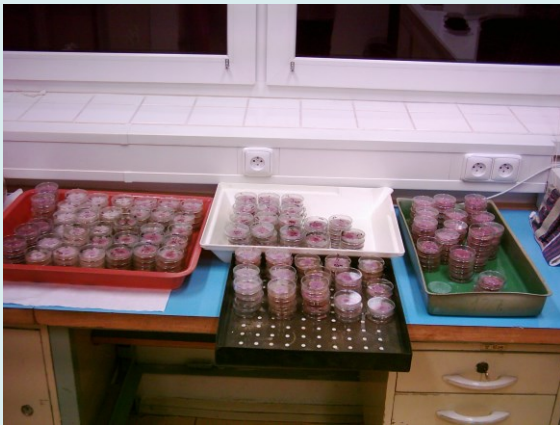
- z každého vzorku se vyberou unikátní kolonie („koloniální morfotypy“)
- okem definované morfotypy souhlasí s mol. genet. vymezením
- funguje i pro *mycelia sterilia*
- 3-5 ks – případné odhalení kryptických druhů



Studium diverzity

Výběr kmenů pro analýzu

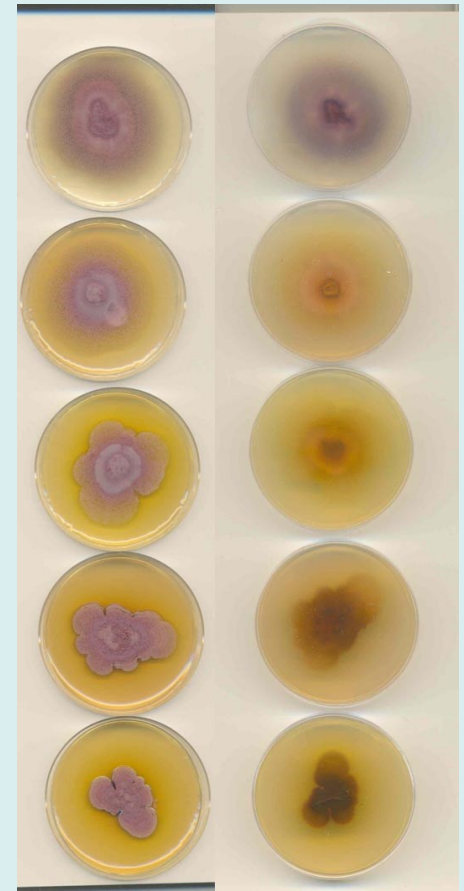
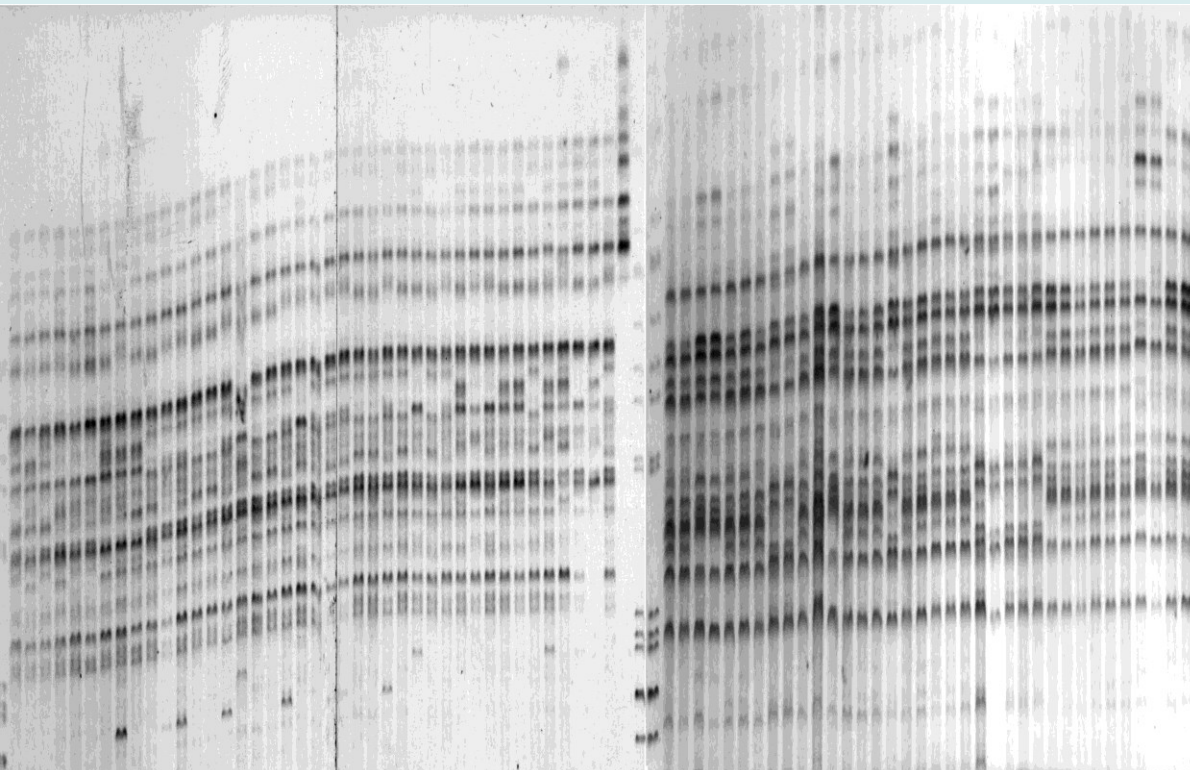
- Zástupci koloniálních morfotypů z více vzorků ze stejného substrátu a lokality jsou sjednoceny do morfotypů dle kolonie na MEA (24C, 37), CYA a CZD



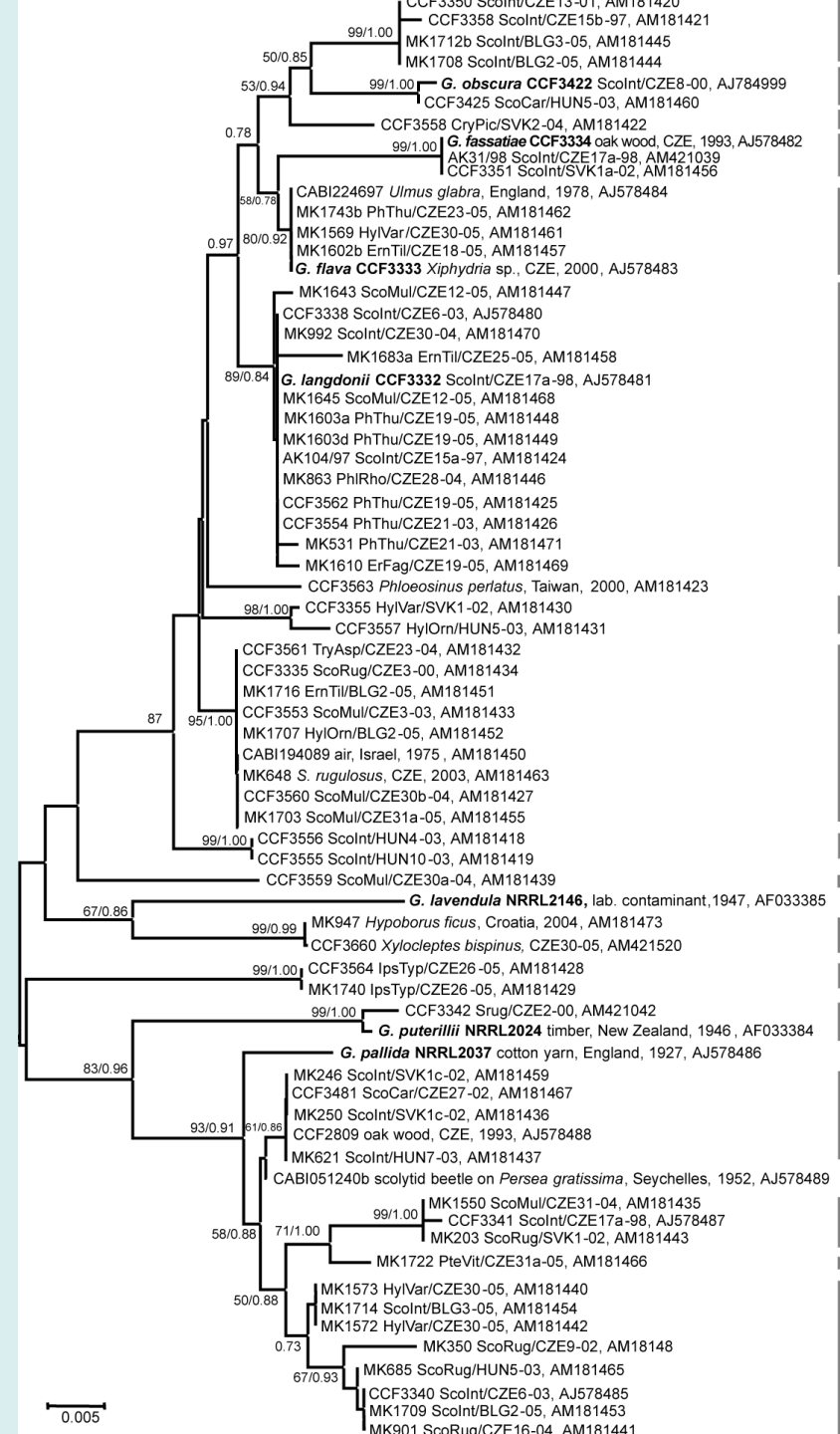
Studium diverzity

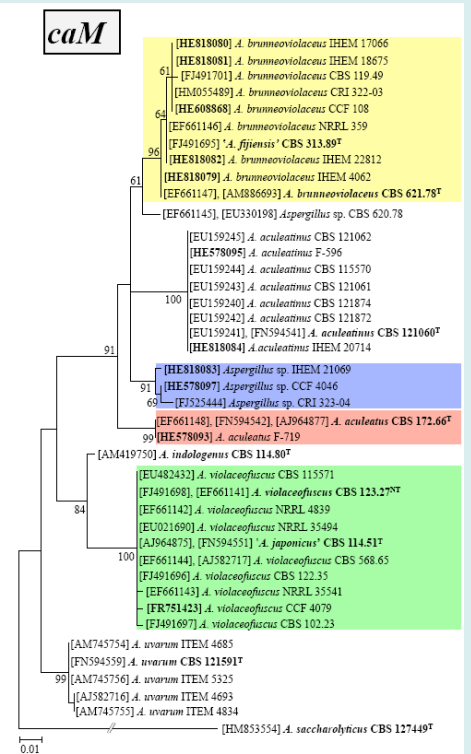
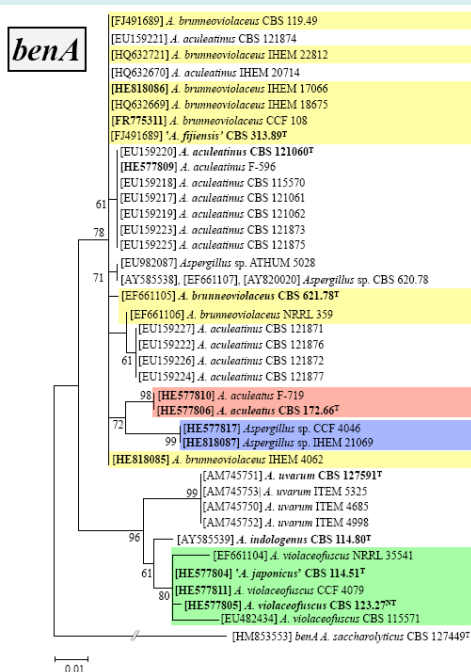
redukce identických kmenů pro sekvenování

- Zástupci na sjednocení kmenů do geneticky unikátních linií (např. RAPD-typ) lze použít PCR fingerprinting (ISSR, ERIC-PCR, RAPD, ITS-RFLP), či DNA barcoding...
- skupiny o 80-100% podobnosti

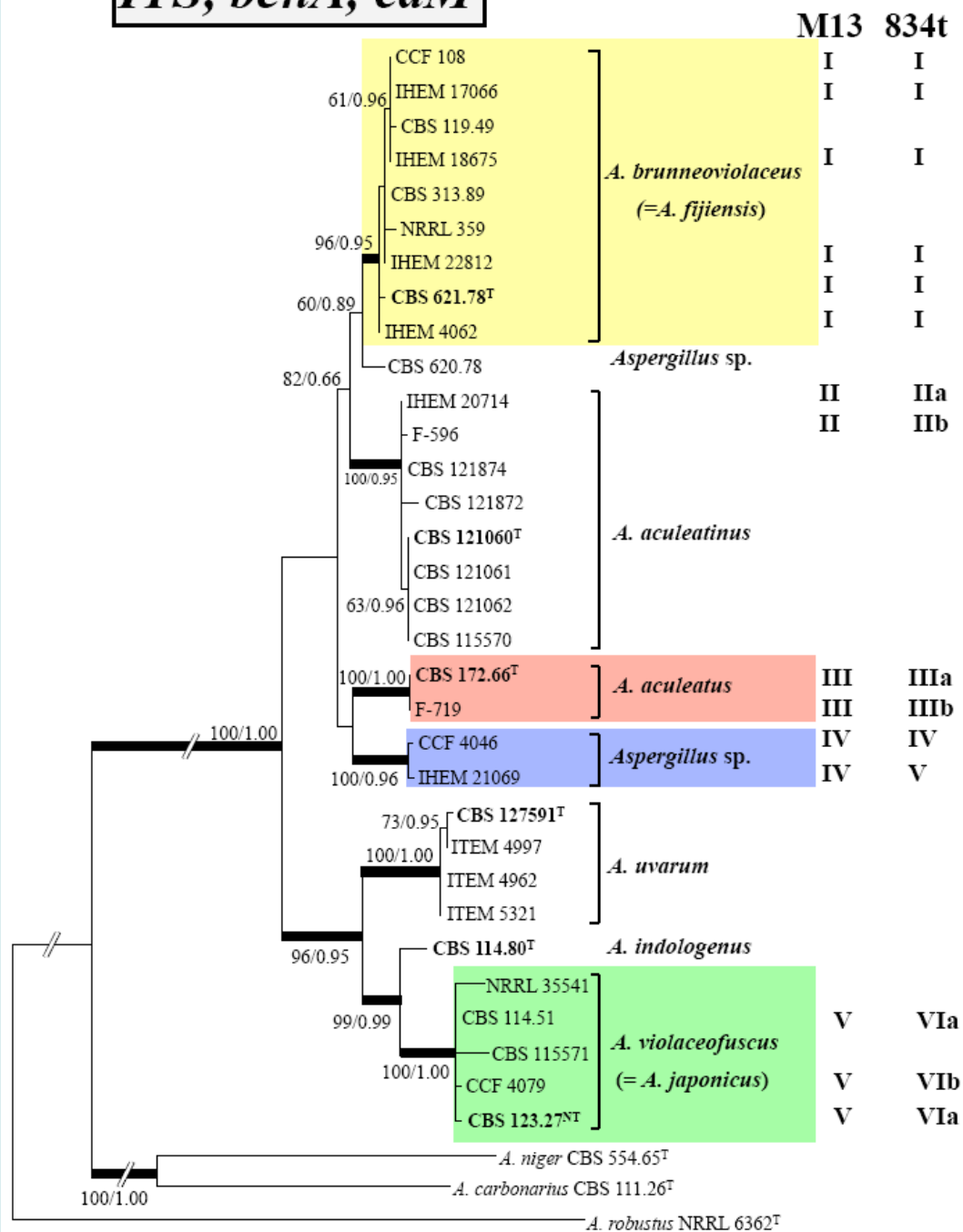


- rozdělení do skupin dle sekvencí
- Případně více genů (MLST – multilocus sequence typing)
- vymezení OTU (operational taxonomic unit), či MOTU (molecular OTU)





ITS, benA, caM



Studium diverzity



- spora 5 μm , 125 μm^3
- objem 1000 cm^3 (1l)
- 6 600 000 000 000 (6×10^{13}) spor v jedné plodnici
- organické hmoty na zemi 10^{13} - 10^{14} kg
- **Kdyby vyklíčily spory z jedné velké pýchavky, tak pozřou vsechnu organickou hmotu na Zemi**



Pýchavka obrovská

Upon being stirred, the spores became evenly scattered in the fluid. With the counting apparatus, as an average of several trials, the 0.1 gram of the fruit-body was calculated to contain 3,245,000,000 spores. The whole fruit-body, therefore, was calculated to contain 7,500,000,000,000 spores. The fruit-body had a small sterile base and a very thin, although imperfect, peridium, but a small part of the gleba had been lost by accident. Taking these factors into account, it seems fairly safe to state that the puff-ball produced about 7,000,000,000,000 spores which would be liberated by about 4000 good-sized probably a large Giant Puff-ball which, it is said, can almost attain the dimensions of a sheep, is the largest organism living on our planet.

RESEARCHES ON FUNGI

AN ACCOUNT OF THE PRODUCTION, LIBERATION,
AND DISPERSION OF THE SPORES OF HYMENOMYCE-
TES TREATED BOTANICALLY
AND PHYSICALLY
ALSO SOME OBSERVATIONS UPON THE DISCHARGE AND
DISPERSION OF THE SPORES OF ASCOMYCETES
AND OF PHLOBOLUS

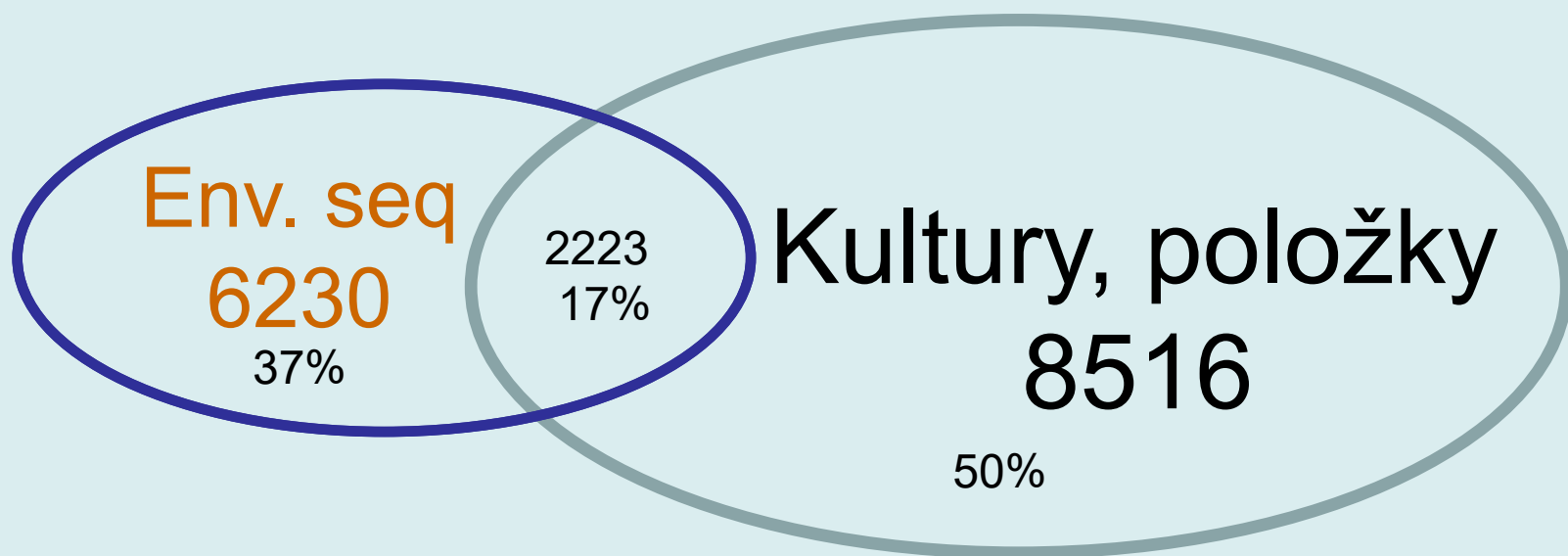
BY

A. H. REGINALD BULLER
B.Sc. (Lond.); D.Sc. (Birm.); Ph.D. (Leip.)
PROFESSOR OF BOTANY AT THE UNIVERSITY OF MANITOBA

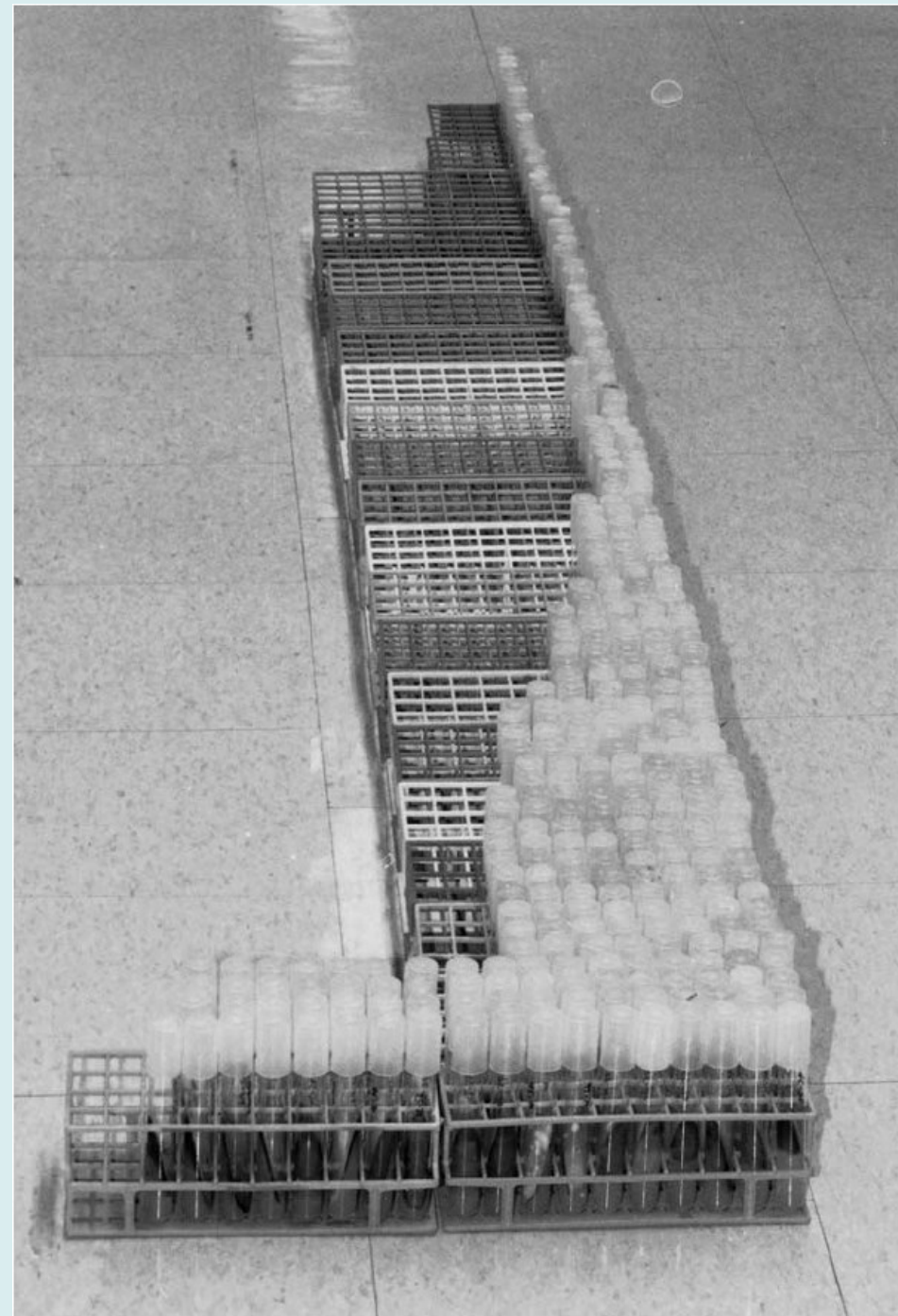
- 1 g půdy – 800 m aktivního mycelia
- 16 km mycelia celkem
- 21 milionů km silnic na celém světě (...2 m³ půdy)



- Analýza sekvencí z Genbank (90 tis.)
- Rozdělení na OTU dle 97% podobnosti sekvencí (hrubě koresponduje s druhem)
- 17000 OTU

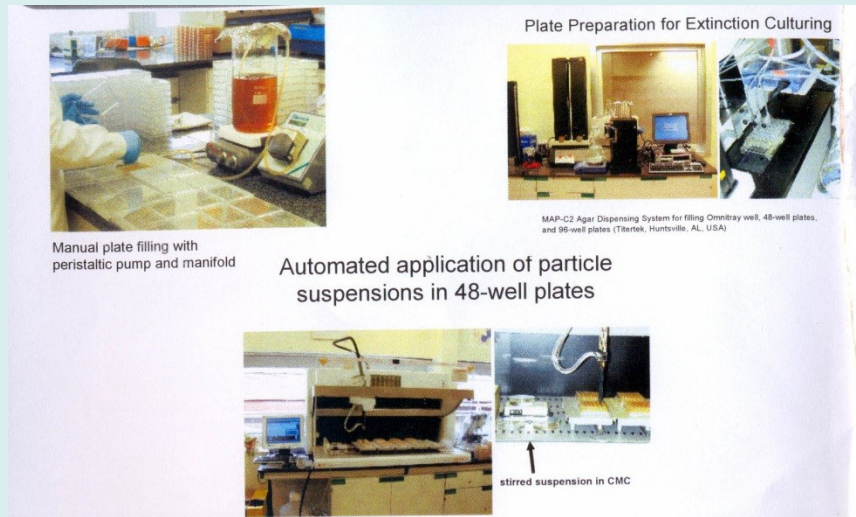


- Logaritmické rozdělení četnosti druhů
- Jsme schopni studovat jen více frekventované druhy
- Omezené zachycení méně frekventovaných druhů
- 20000 – 1 000 000 CFU/g
- dalších 67-81% nelze kultivovat
- 100-1200 OTU na gram půdy



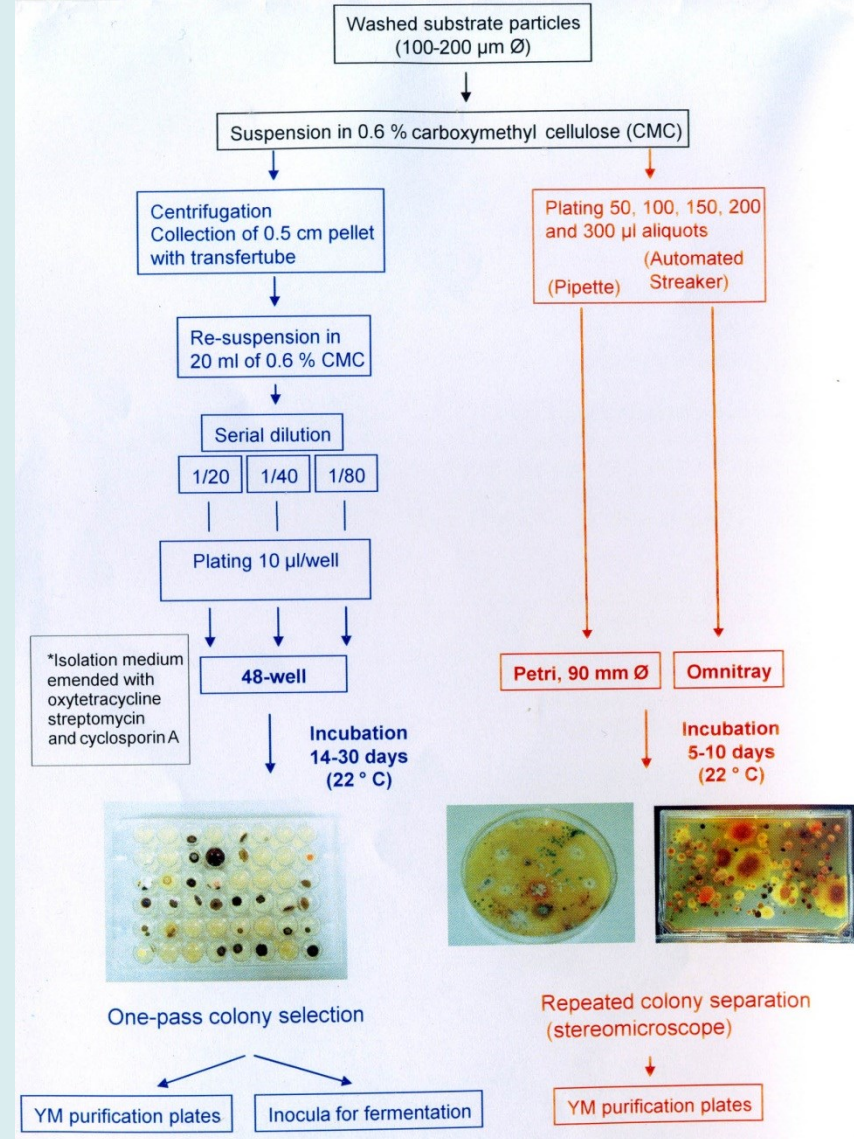
Studium diverzity

- environmentální DNA
- klasické a zdokolané izolační techniky



Extinction culturing

- vyvinuto pro pomalu rostoucí vodní bakterie (objev mnoha nových druhů)
- „pozorovaná diverzita narůstá s klesajícím množstvím inokula“
- málo použito u hub

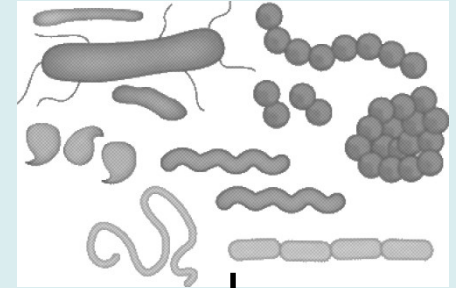
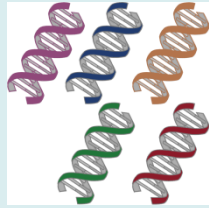


METAGENOMIC APPROACHES

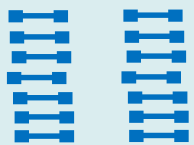
EXTRACTED DNA

ISOLATES

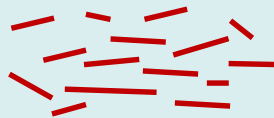
ENVIRONMENTAL
SAMPLE



Targeted Gene
Sequencing



Shotgun Metagenome
Sequencing



binning



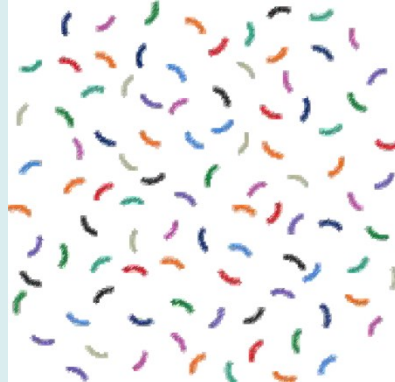
Single Cell Genome
Sequencing



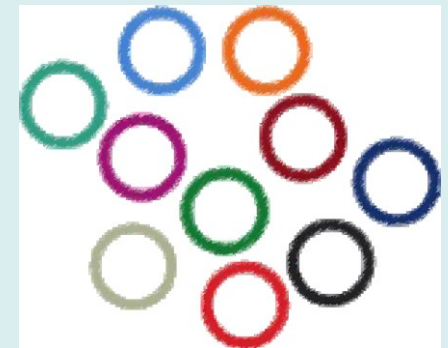
known genes
*identity, diversity,
distribution, dynamics*

"all" genes
*functional potential
metabolic reconstruction*

genomes
*linking phylogeny & function,
genome diversification*



binning to
metagenome
assembled
genomes (MAGs)



Gene-centric

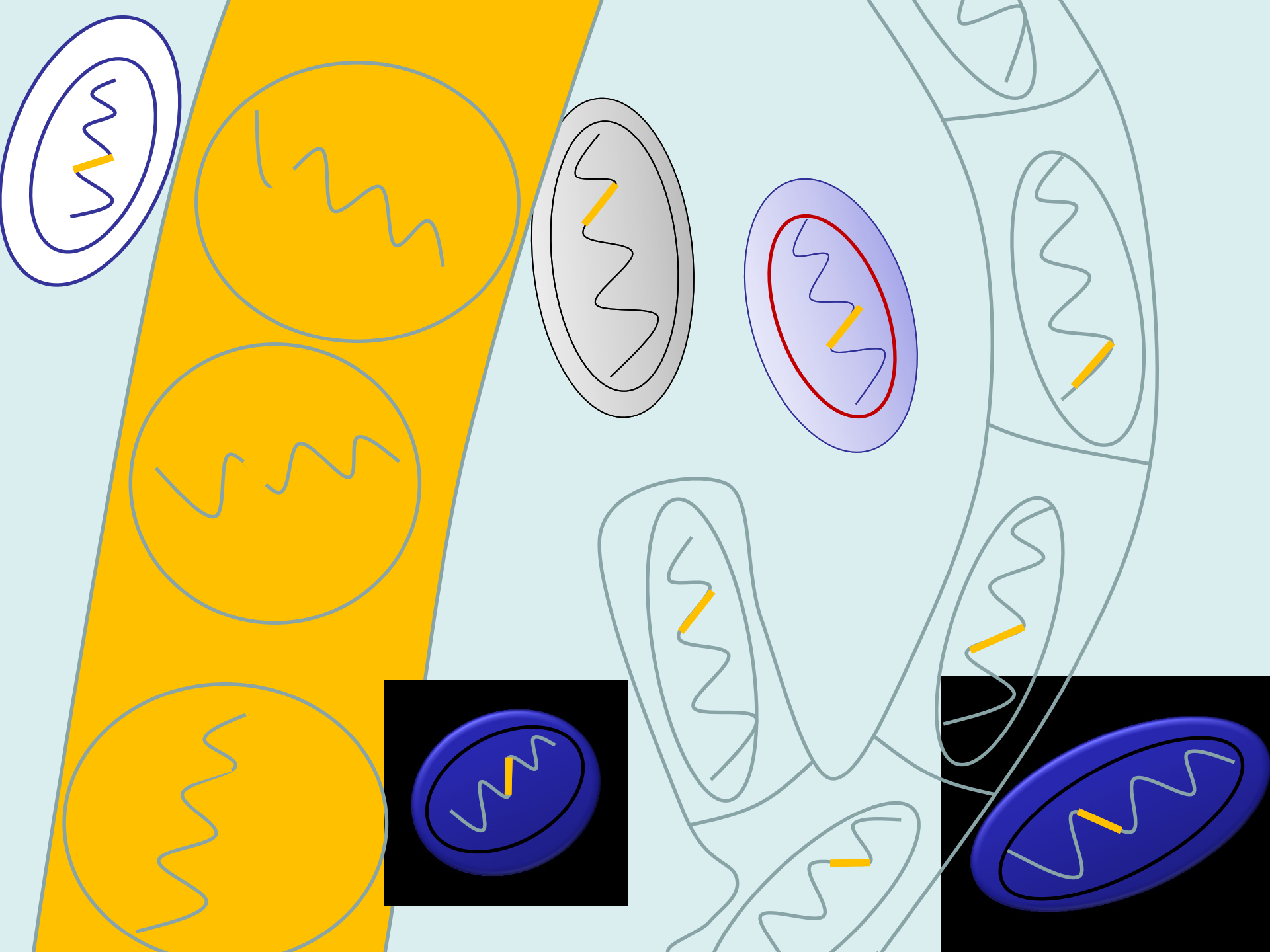
Genome-centric

metabarcoding

metaganomics

Metabarcoding

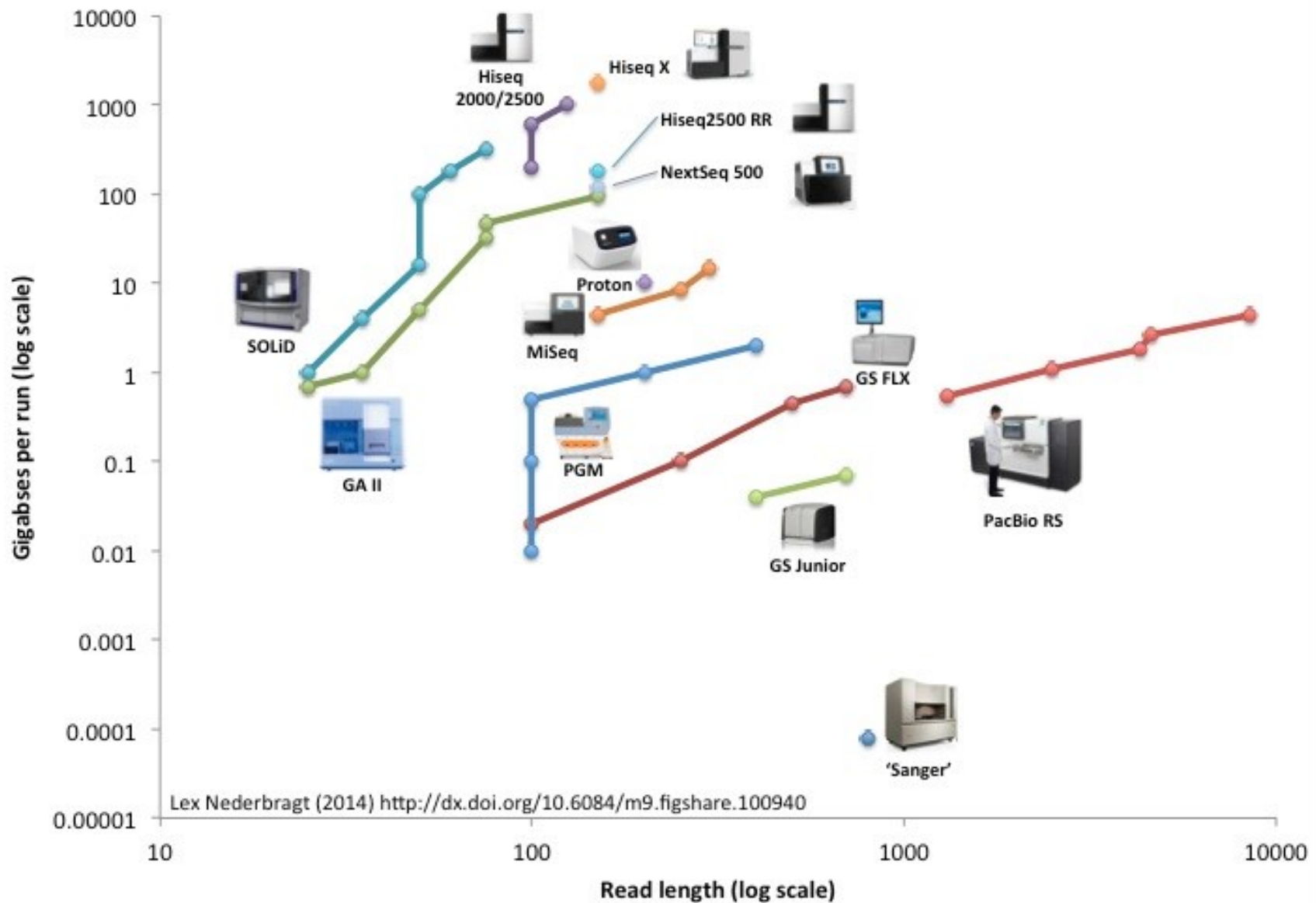
- totální environmentální DNA
- PCR amplifikace barcode genu
- možnost paralelizace pomocí Tagů na primerech, či ligovaných na PCR produkt
- nutnost separace jednotlivých PCR produktů
 - DGGE – rozdělí produkty dle sekvence
 - t-RFLP
 - inzerce do plazmidu *E. coli* („klonování“)
 - masivní paralelní sekvenování



Analýza enviromtálních vzorků

- **high-throughput sequencing = next-generation sequencing, massive parallel sequencing**
 - » 454 pyrosequencing (přerušena podpora)
 - » **Illumina (Solexa) sequencing**
 - » SOLiD sequencing
 - » Ion torrent
 - » Helicos
 - » **Pacific Biosciences**
 - » **Nanopore**
- Sekvenování jednotlivých DNA fragmentů (PCR fragmenty, či fragmenty genomické DNA)
- Jednotlivé metody se liší délkou přečtených fragmentů, chybovostí, cenou...

Developments in High Throughput Sequencing



2020 – Illumina, PacBio, Nanopore



Illumina
(HiSeq 4000)

PacBio
(Sequel)

Oxford Nanopore
(MinION)

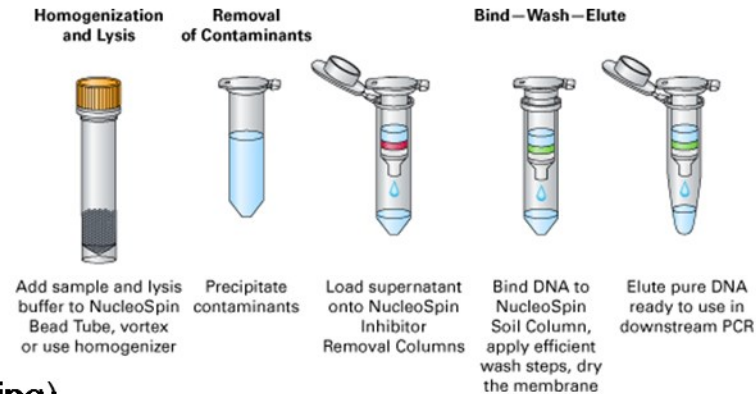
Read length	Up to 150 bp	10-15kb	Up to 900kb
Number of reads	2.5-5 Million	500 K	Up to 1 M
Processing time	<1-3.5 days	Up to 10 hours	~ 6 hours
Error rate	<1%	10-15%	5-15%
Cost per run	~\$3000	~\$850	\$500-\$900
Instrument price	\$900 K	\$350K	\$1K
Advantages	Highly accurate	Sequence long reads	Sequence long reads Portable device

Kontinuální pokrok ve zlepšení chybovosti

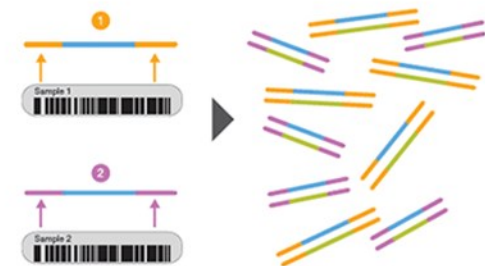
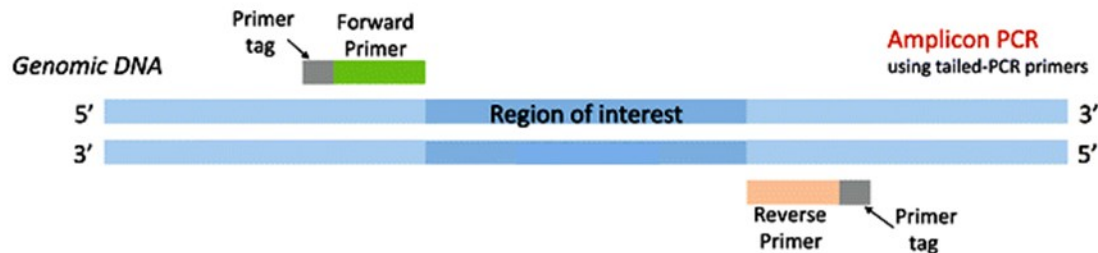


Illumina amplicon sequence library preparation

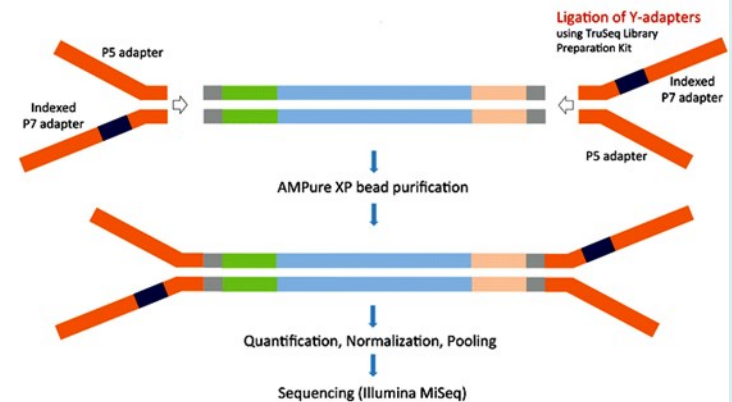
1. DNA isolation



2. PCR with barcoded primers (multiplexing)



3. Ligation of sequencing adapters – to attach short oligonucleotides (60bp) to your DNA fragments, these oligonucleotides are used to attach to the sequencing flow cell and they are also used as barcode of library



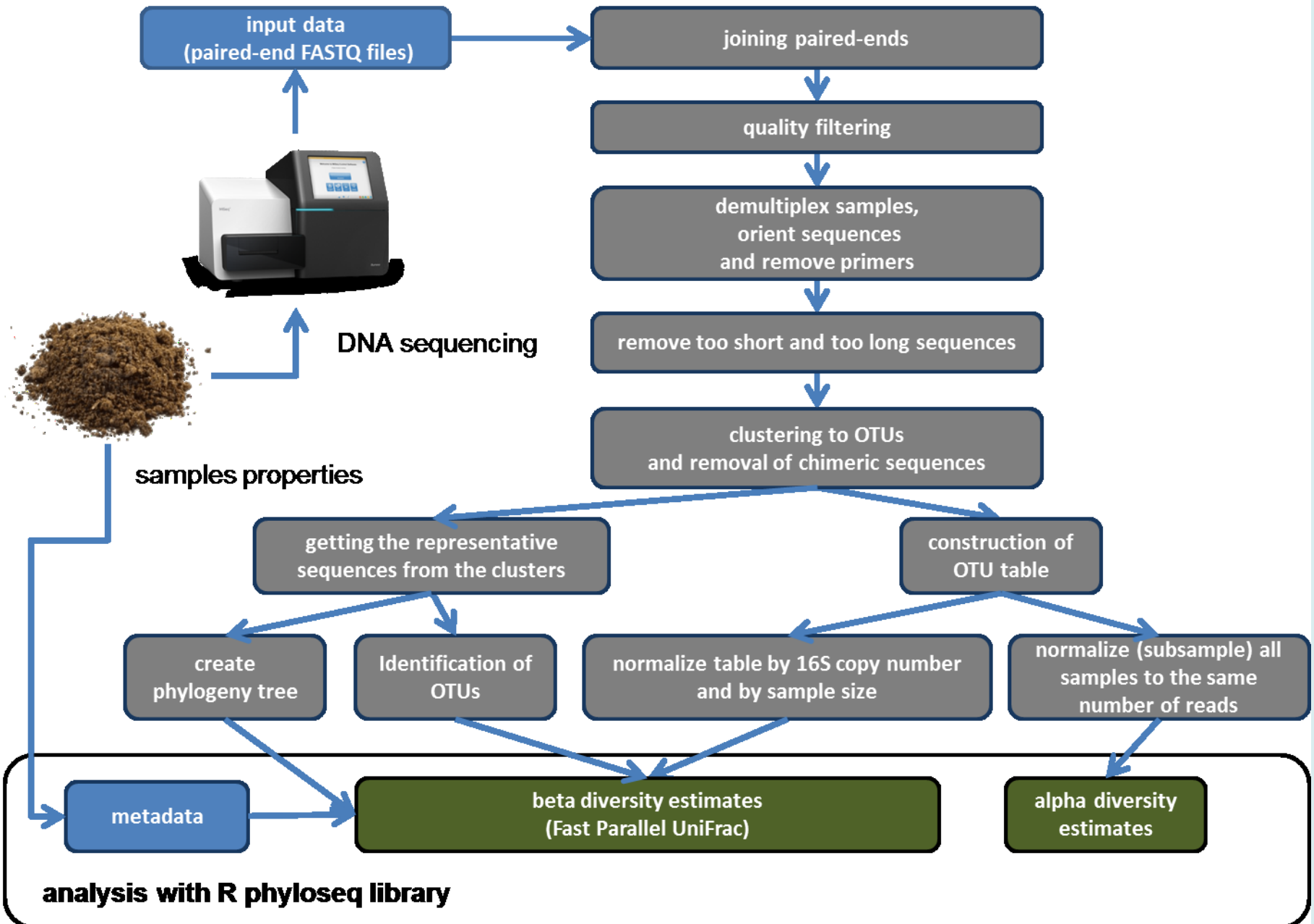
4. Quantification of the library by qPCR – to quantify of the exact amount of ligated fragments

Metabarcoding

- Minimalizace PCR bias a vlivu DNA extrakce
 - více nezávislých DNA izolací a PCR reakcí
 - menší počet PCR cyklů
 - polymeráza s menší chybovostí
- odstranění chybných sekvencí (tzv. denoising) a chimér
- klastrování sekvencí do **OTU (či MOTU)** (dle % podobnosti či jinak), nejčastěji 97%
- Klastrovací metody se značně liší, takže i neklastrovaná data
- přiřazení důležitých MOTU do druhů (Blast)

Amplicons pipeline workflow

T. Větrovský



[Home](#)[Install](#)[Tutorials](#)[Scripts](#)[Help](#)[Resources](#)[File Formats](#)[Workshops](#)[Blog](#)[Developer](#)[Citing QIIME](#)[Forum](#)[Documentation](#)

What is QIIME?

QIIME™ (canonically pronounced *chime*) stands for Quantitative Insights Into Microbial Ecology.

QIIME 2 has succeeded QIIME 1 as of January 1, 2018. QIIME 1 is no longer supported at this time, as development and support effort for QIIME is now focused entirely on [QIIME 2](#). For more information, see our blog post: [QIIME 2 has succeeded QIIME 1](#).

QIIME 1 users should transition from QIIME 1 to [QIIME 2](#). If you're new to QIIME, you should start by learning [QIIME 2](#), not QIIME 1.

This site documents QIIME 1. To learn more about [QIIME 2](#), see <https://qiime2.org>.

QIIME is an open-source bioinformatics pipeline for performing microbiome analysis from raw DNA sequencing data. QIIME is designed to take users from raw sequencing data generated on the Illumina or other platforms through publication quality graphics and statistics. This includes demultiplexing and quality filtering, OTU picking, taxonomic assignment, and phylogenetic reconstruction, and diversity analyses and visualizations. QIIME has been applied to studies based on billions of sequences from tens of thousands of samples.

Getting started with QIIME

Installing: The quickest way to get started using QIIME is with [MacQIIME](#) (if you're running Mac OS X), the [QIIME VirtualBox](#) or the [QIIME Amazon EC2 image](#) (if you're using Windows, Mac OS X, or Linux), or [pip](#) (if you're using Linux or Mac OS X). See the [QIIME install documentation](#) for details.

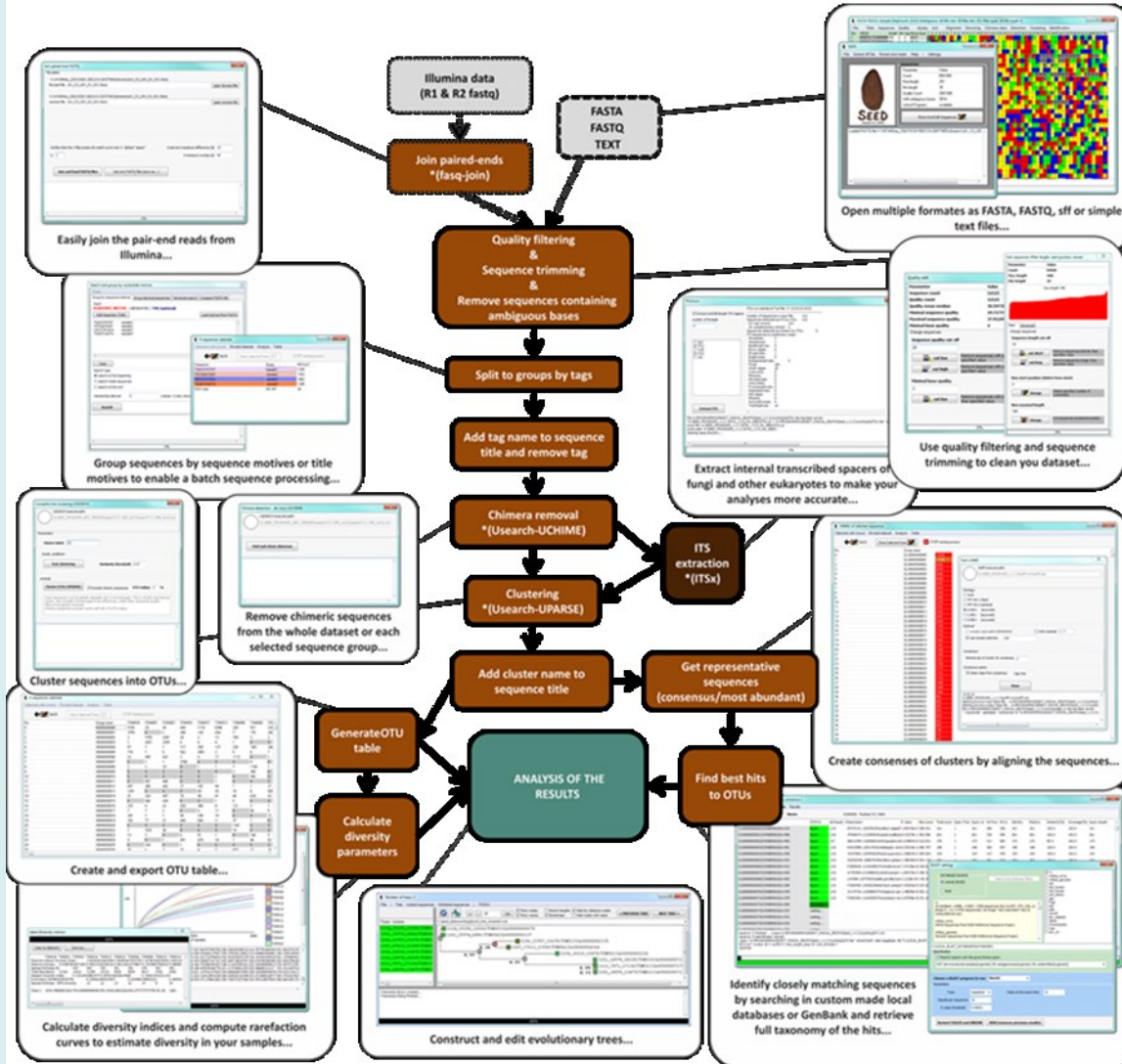
Running: Once you've installed QIIME, move on to the [QIIME Tutorials](#). The [Illumina overview tutorial](#) or the [454 overview tutorial](#) are good first analyses to run. In each of these tutorials you'll download a small data set and work through a series of commands that will introduce you to some of QIIME's commonly used features and analyses.

Getting help: For help with QIIME, see help.qiime.org. For getting started on interacting with the command line, we recommend the [Software Carpentry lessons and workshops](#).

QIIME scripts: [The QIIME script documentation](#) will help you explore and learn QIIME's functionality.

[illegible]

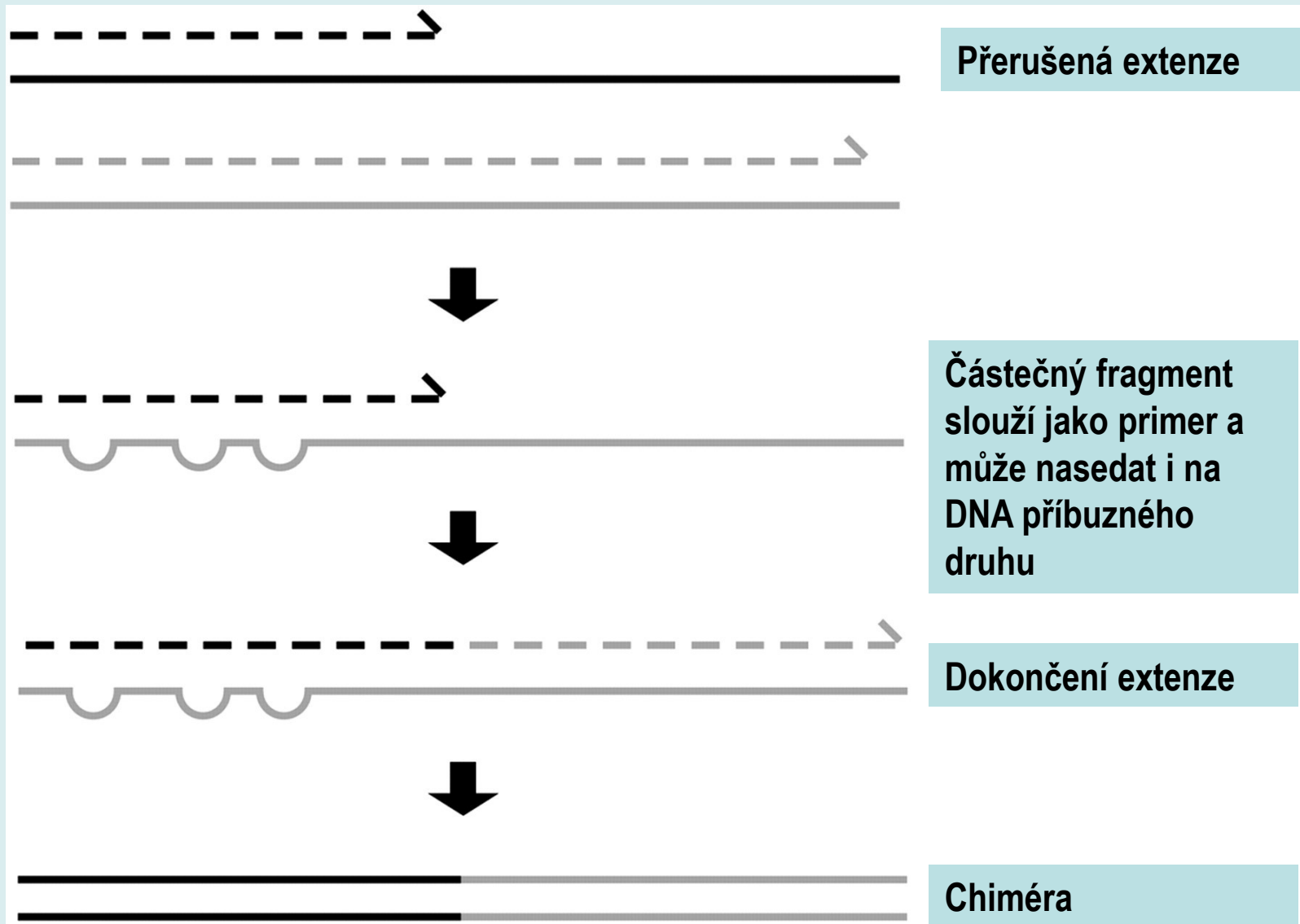
GUI based alternative for Windows (<http://www.biomed.cas.cz/mbu/lbwrf/seed/>)



SEED 2: a user-friendly platform for amplicon high-throughput sequencing data analyses

- editing of sequences and their titles
- sorting
- quality trimming
- pair-end joining
- grouping of sequences based on sequence motifs or sequence titles
- batch processing of sequence groups
- denoising
- chimera removal
- ITS extraction
- sequence alignments and clustering
- OTU table construction
- construction of consensus sequences
- creation of local databases for BLAST
- searching either local databases or the whole NCBI
- retrieval of taxonomical classification from the NCBI
- calculation of diversity parameters
- many more...

- Nedokončený PCR produkt z předešlých PCR cyklů může fungovat jako templát v následujících PCR cyklech. Možný vznik chimér..
- Částečně je lze odhalit pomocí poč. programů (Chimera check)



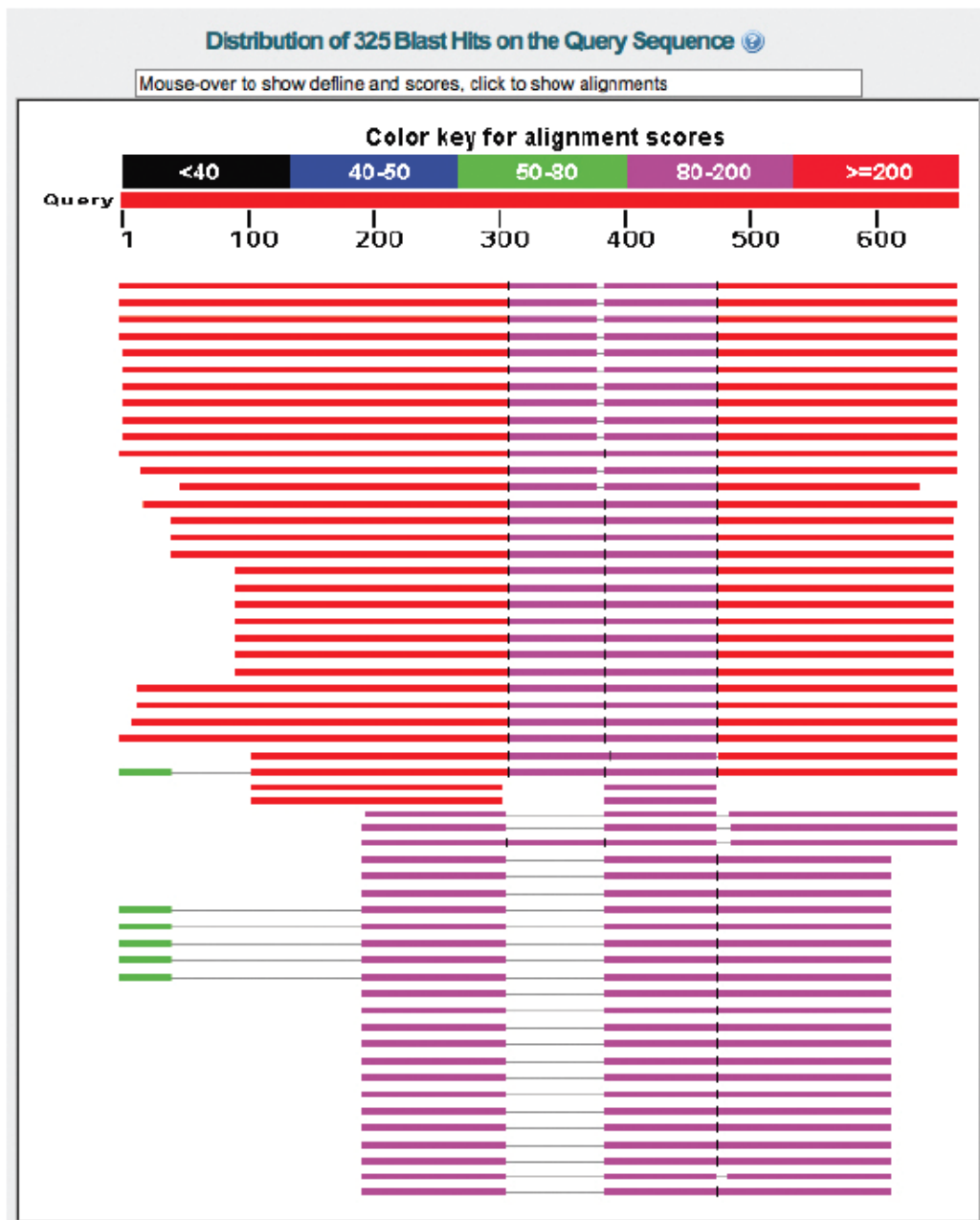


Figure 6. An assembly chimera. The black dashed lines indicate breaks in the BLAST alignment and should always be taken to mean that manual examination is needed.

Clustering to OTUs

OTU (Operational taxonomic unit) group of similar sequences grouped based on some similarity threshold usually 97% similarity (16S, ITS) represents Species

Heuristic

- comparison of each sequence with representative sequence („seed“)
- depends on sequence order

FAST

Hierarchical

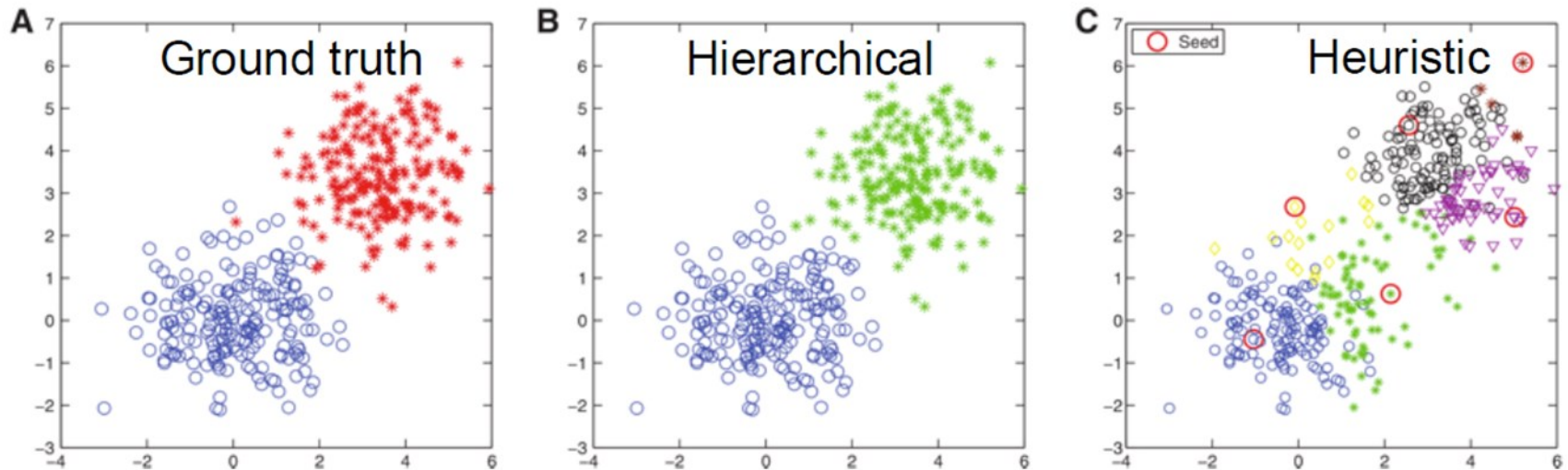
- comparison of each sequence with each other (tree construction)

SLOW

Model based

- probabilistic, iterative
- uses more information than the sequence identity

**VERY
SLOW**



Hierarchical clustering

- is able to identify the real clusters (ideally)
- computationally expensive

x

Heuristic clustering

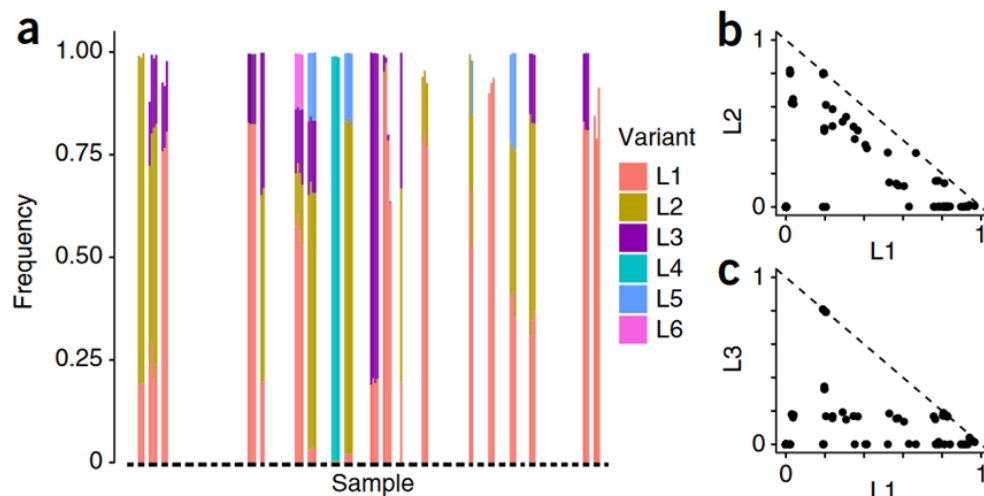
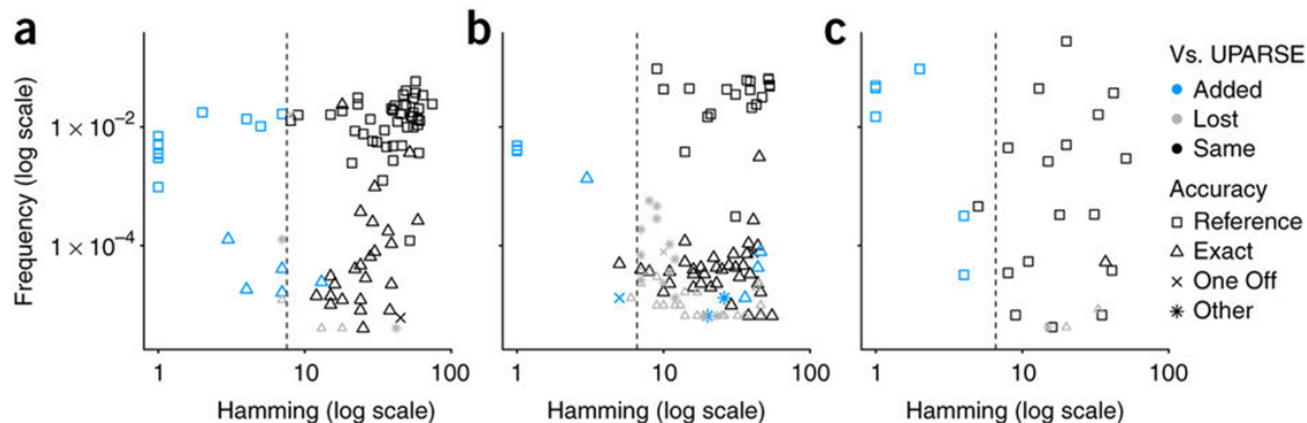
- computationally cheap
- often generates artificial clusters (overestimated diversity)

- Algorithmy
UPARSE
USEARCH
CROP

- Jde to i bez klastrování - amplicon sequence variant (ASV), tedy každá unikátní sekvence je použita ja taxon

Clustering-independent methods (DADA2)

Comparison of sequence variants inferred by DADA2 with OTUs constructed by UPARSE.

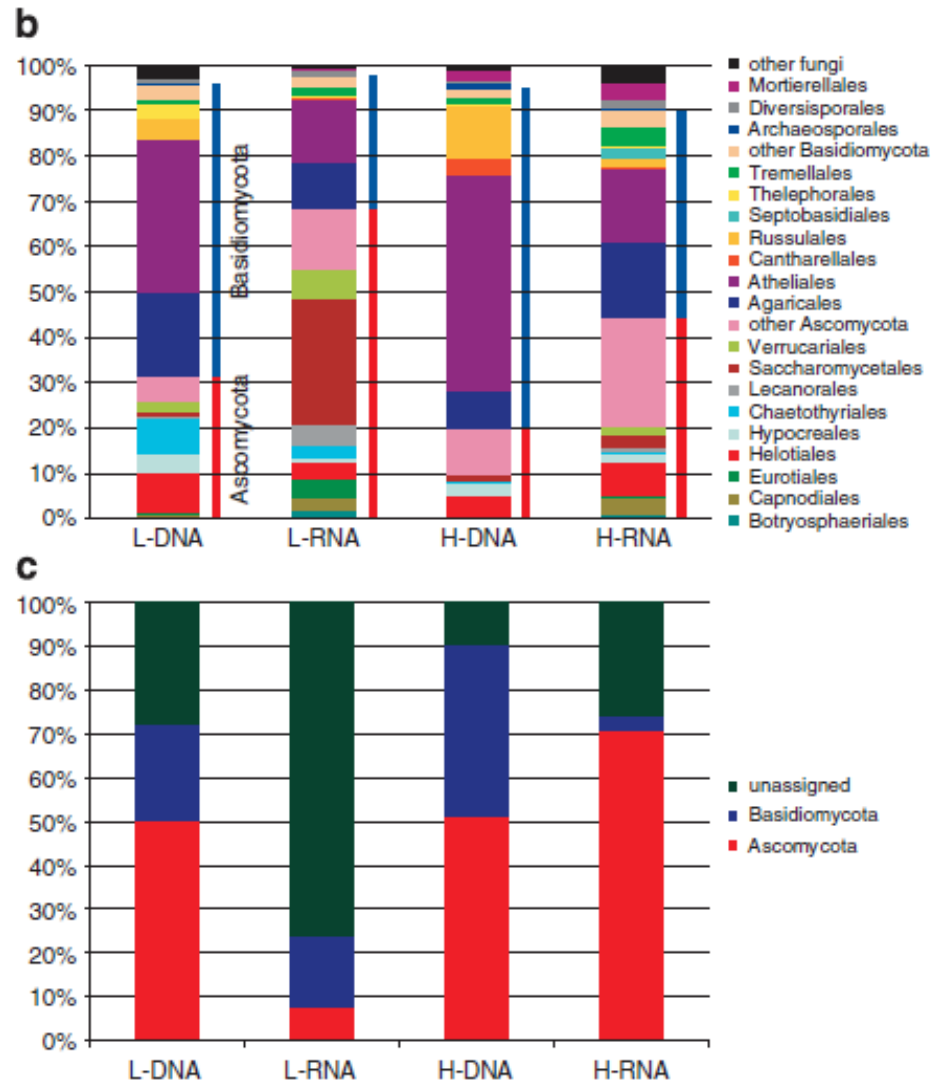


L. crispatus sequence variants in the human vaginal community during pregnancy. DADA2 identified six *L. crispatus* 16S rRNA sequence variants present in multiple samples and a significant fraction of all reads.

(a) The frequency of L1–L6 in each sample. Black bars at the bottom link samples from the same subject. (b,c) The frequency of (b) L1 vs. L2 and (c) L3 vs. L1, by sample. The dashed line indicates a total frequency of 1.

Analýza enviromtálních vzorků

- **metabarcoding**
- Sekvenování barcode markeru z env. vzorků
- ITS region (ITS1, ITS2)
- Mnohokopiový marker, problém s kvantifikací
- Vnitrogenomová variabilita
- Rozdílná délka mezi taxony
- Alternativní markery, RPB2, TEF1a,
 - jednokopiové markery
 - Problém s univerzálními primery
 - Specifika PCR s degenerovanými primery
 - Absence alternativních markerů v databázích
 - Problém s introny



• Rozdíl mezi DNA a RNA

Figure 3 Phylogenetic assignment of bacterial, fungal and *cbhI* sequences from *Picea abies* forest topsoil. The data represent mean values from four study sites. (a) Bacteria, (b) *fur* (c) *cbhI* sequences.

Active and total microbial communities in forest soil are largely different and highly stratified during decomposition

Petr Baldrian¹, Miroslav Kolařík¹, Martina Štursová¹, Jan Kopecký², Vendula Valášková¹, Tomáš Větrovský¹, Lucia Žifčáková¹, Jaroslav Šnajdr¹, Jakub Rídl³, Čestmír Vlček³ and Jana Voříšková¹

ITS rDNA

- Odlišná velikost 500-900 bp
- 10-200 kopií na genom
- vnitrogenomová variabilita
- 3%
- >3%

Sangerovo sekvenování

AGGC TTGCAAGT CAGGC TTGAAA

➡ 1 OTU

NGS

GGGc TTGcAAGTcAGGC TTGAAA
AGGc dA TGC AAGT CAGGC TTGAAA
AGGC TT GGG GGT CAGGC TTGAAA
AGGC TTGCAAGT CAGGC CCc TTGAAA
AGGGGTGT CAGGC TTGAAA

➡ 4 OTU

AGGC TTGCAAGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC TTGAAA

AGGC dA TGC AAGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC TTGAAA

AGGC TT GGG GGT CAGGC TTGAAA

AGGGGTGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC TTGAAA

AGGC TTGCAAGT CAGGC CCc TTGAAA

AGGC TTGCAAGT CAGGC CCc TTGAAA

TABLE II. Number of *Laetiporus* OTUs based on cloned and consensus ITS sequences from cultures of six *Laetiporus* species

Percent similarity used to define OTUs	Number of <i>Laetiporus</i> OTUs observed	
	Consensus ITS sequences	Cloned ITS sequences
90%	7	20
95%	7	20
97%	7	21
98%	7	28
99%	7	35
100%	16	338

**Intragenomic variation in the ITS rDNA region obscures
 phylogenetic relationships and inflates estimates of
 operational taxonomic units in genus *Laetiporus***

Daniel L. Lindner¹
 Mark T. Banik

spacer region, intragenomic variation, molecular
 drive, sulfur shelf

*US Forest Service, Northern Research Station, Center for
 Forest Mycology Research, One Gifford Pinchot Drive,
 Madison, Wisconsin 53726*

INTRODUCTION

data z přímého sekvenování (direct sequencing, Sanger)

nejsou identická s

daty z „amplicon sequencing“ (= metabarcoding, klonování)

LINDNER AND BANIK: VARIATION IN *LAETIPORUS*

735

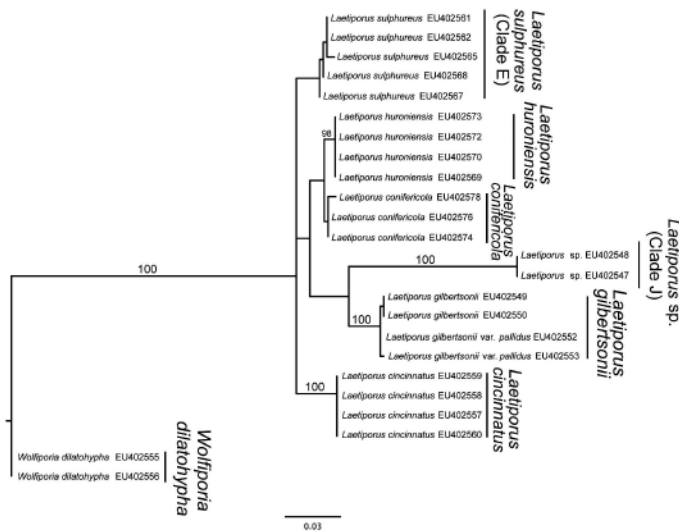


FIG. 1. Maximum likelihood phylogeny of *Laetiporus* isolates based on consensus ITS sequences from 22 cultures of six *Laetiporus* species and two cultures of *Wolfiporia dilatohypha*. Each sequence is labeled with the corresponding GenBank accession number, and species clades are indicated with vertical black lines. Bootstrap values greater than 70% are reported above branches.

736

MYCOLOGIA

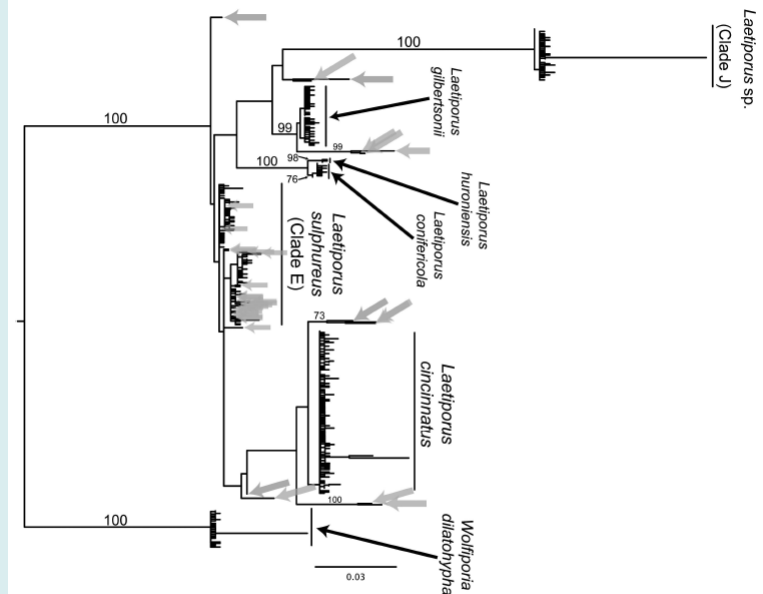
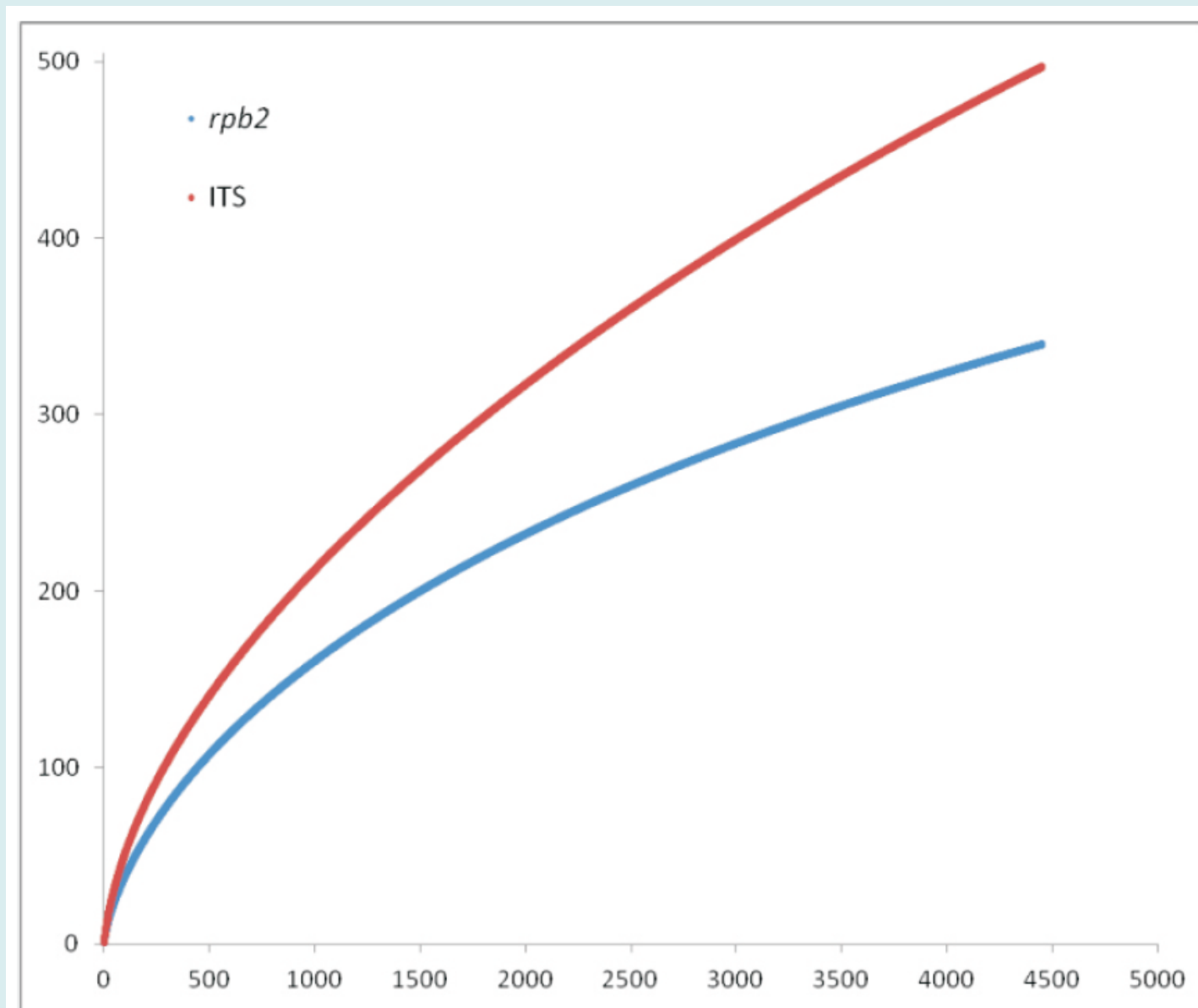


FIG. 2. Maximum likelihood phylogeny of cloned ITS sequences from 1399 clones derived from 19 isolates of six *Laetiporus* species and one *Wolfiporia* species. Species clades based on consensus sequences are indicated with vertical black lines. Gray arrows indicate the position of cloned *L. cincinnatus* ITS sequences that fell outside the *L. cincinnatus* consensus clade. Bootstrap values greater than 70% are reported above branches.



- Rarefakční křivka - ukazatel pokryvu celkové diverzity

Table 2 Identification of the thirty most abundant OTUs in spruce forest topsoil using ITS and *rpb2* as alternative molecular markers. The identity was assigned using BLASTN (for ITS) or tBLASTX (for *rpb2*). The percentage similarity of protein sequence is given in the case of *rpb2*. The taxonomic classification to Ascomycota (A) and Basidiomycota (B) is marked

ITS					<i>rpb2</i>				
Identification	Accession number	Similarity (%)	Coverage (%)	Abundance (%)	Identification	Accession number	Similarity (%)	Coverage (%)	Abundance (%)
<i>Piloderma</i> (B)	JQ711958	97	100	10.6	<i>Lactarius</i> (B)	DQ408128	98	99	22.9
<i>Lachnellula</i> (A)	KC464638	98	100	7.0	<i>Lactarius</i> (B)	DQ408128	98	99	13.3
<i>Mycena</i> (B)	EF093152	100	100	6.2	<i>Ceratomyrium</i> (A)	AY485617	79	99	8.2
<i>Cenococcum</i> (A)	AM087244	99	100	4.4	<i>Rasamsonia</i> (A)	JQ729684	97	99	4.3
<i>Marasmius</i> (B)	FR717227	99	97	4.1	<i>Lactarius</i> (B)	DQ408128	99	99	3.6
<i>Cladophialophora</i> (A)	EF016381	97	100	3.8	<i>Hydropus</i> (B)	DQ472722	71	62	3.5
<i>Russula</i> (B)	HM189931	99	100	2.9	<i>Albotricha</i> sp. (A)	AB481347	98	99	3.2
<i>Hygrophorus</i> (B)	JP908073	100	100	2.5	<i>Piloderma</i> (B)	GU187797	99	99	2.4
<i>Melinisomyces</i> (A)	HQ157837	100	100	2.4	<i>Ceratomyrium</i> (A)	AY485617	92	90	2.3
<i>Tylospora</i> (B)	JN943896	100	100	2.4	<i>Gymnopus</i> (B)	DQ472716	99	97	1.7
<i>Tricholoma</i> (B)	AB036899	99	100	1.8	<i>Cudonia</i> (A)	AY641033	92	99	1.4
Uncultured (<i>Trechispora</i>) (B)	FJ475683	99	100	1.8	<i>Tyrannosorus</i> (A)	DQ470928	74	52	1.4
<i>Cladophialophora</i> (A)	EF016381	96	100	1.5	<i>Veluticeps</i> (B)	HM536125	92	100	≤1
Ascomycete (<i>Scytalidium</i>) (A)	GU067746	99	100	1.5	<i>Umbelopsis</i> *	DQ302787	64	96	≤1
<i>Inocybe</i> (B)	AJ889955	99	99	1.5	<i>Ellisembia</i> (A)	DQ435090	79	94	≤1
<i>Lactarius</i> (B)	JQ712010	99	100	1.3	<i>Microbotryum</i> (B)	DQ789985	90	99	≤1
<i>Rhizocyphus</i> (A)	JQ711796	96	100	1.3	<i>Mortierella</i>	DQ302784	92	100	≤1
<i>Lyophyllum</i> (B)	HE819396	99	45	1.3	<i>Microthyrium</i> (A)	GU371734	77	99	≤1
<i>Lachnellula</i> (A)	AB481245	98	100	1.3	<i>Microbotryum</i> (B)	DQ789985	90	99	≤1
<i>Xerocomus</i> (B)	HQ207696	100	100	1.1	<i>Spiromyces</i> *	DQ302790	54	99	≤1
<i>Cadophora</i> (A)	AB543058	97	87	≤1	<i>Microthyrium</i> (A)	GU371734	75	99	≤1
<i>Rhizocyphus</i> (A)	JQ711893	95	100	≤1	<i>Hydropus</i> (B)	DQ472722	71	62	≤1
<i>Mortierella</i>	JQ272448	99	100	≤1	<i>Mortierella</i>	DQ302784	95	99	≤1
<i>Amamita</i> (B)	EF493271	100	100	≤1	<i>Mortierella</i> *	DQ302784	78	96	≤1
<i>Melinisomyces</i> (A)	HQ157926	99	100	≤1	<i>Ceratomyrium</i> (A)	AY485617	89	99	≤1
<i>Leptodontidium</i> (A)	GU067735	96	100	≤1	<i>Mortierella</i>	DQ302784	92	86	≤1
<i>Exophiala</i> (A)	HE605215	87	100	≤1	<i>Trichopeziza</i> (A)	AB481360	98	98	≤1
<i>Tylospora</i> (B)	JN943896	99	100	≤1	<i>Coprinopsis</i> (B)	XM1829088	100	93	≤1
Uncultured (<i>Leptodontidium</i>)	HM488455	100	100	≤1	<i>Gymnopus</i> (B)	DQ472716	99	99	≤1
Uncultured (<i>Rhodospiridium</i>)	FJ475820	100	62	≤1	<i>Mortierella</i>	DQ302784	92	92	≤1

*OTUs with low similarity to known fungal sequences were placed into basal fungal lineages of uncertain position based on the phylogenetic tree.

Table 3 Comparison of the representation of major fungal groups in the mock community and in the sequence data sets obtained with ITS and *rpb2* as molecular markers

		ITS				<i>rpb2</i>		
		Mock community	Species detected	OTUs		Species detected	OTUs	Reads (%)
Basidiomycota	83 (63.8%)	64 (63.4%)	205 (74.5%)	90.0	68 (70.1%)	133 (75.1%)	62.9	
Ascomycota	42 (32.3%)	32 (31.7%)	62 (22.5%)	8.6	25 (25.8%)	38 (21.5%)	35.3	
Mucoromycotina and Mortierellomycotina	5 (3.9%)	5 (4.9%)	8 (3.0%)	1.4	4 (4.1%)	6 (3.4%)	1.8	

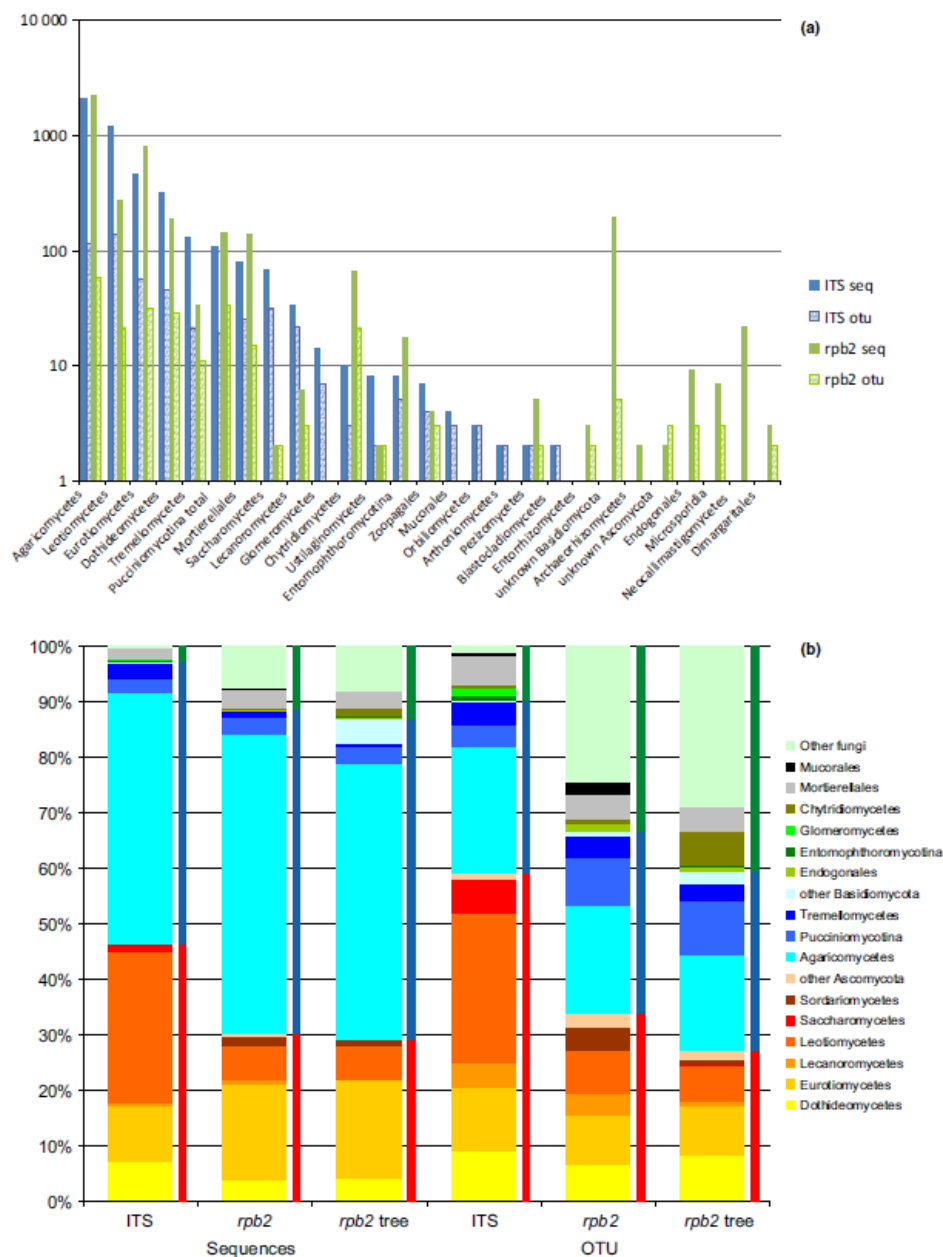
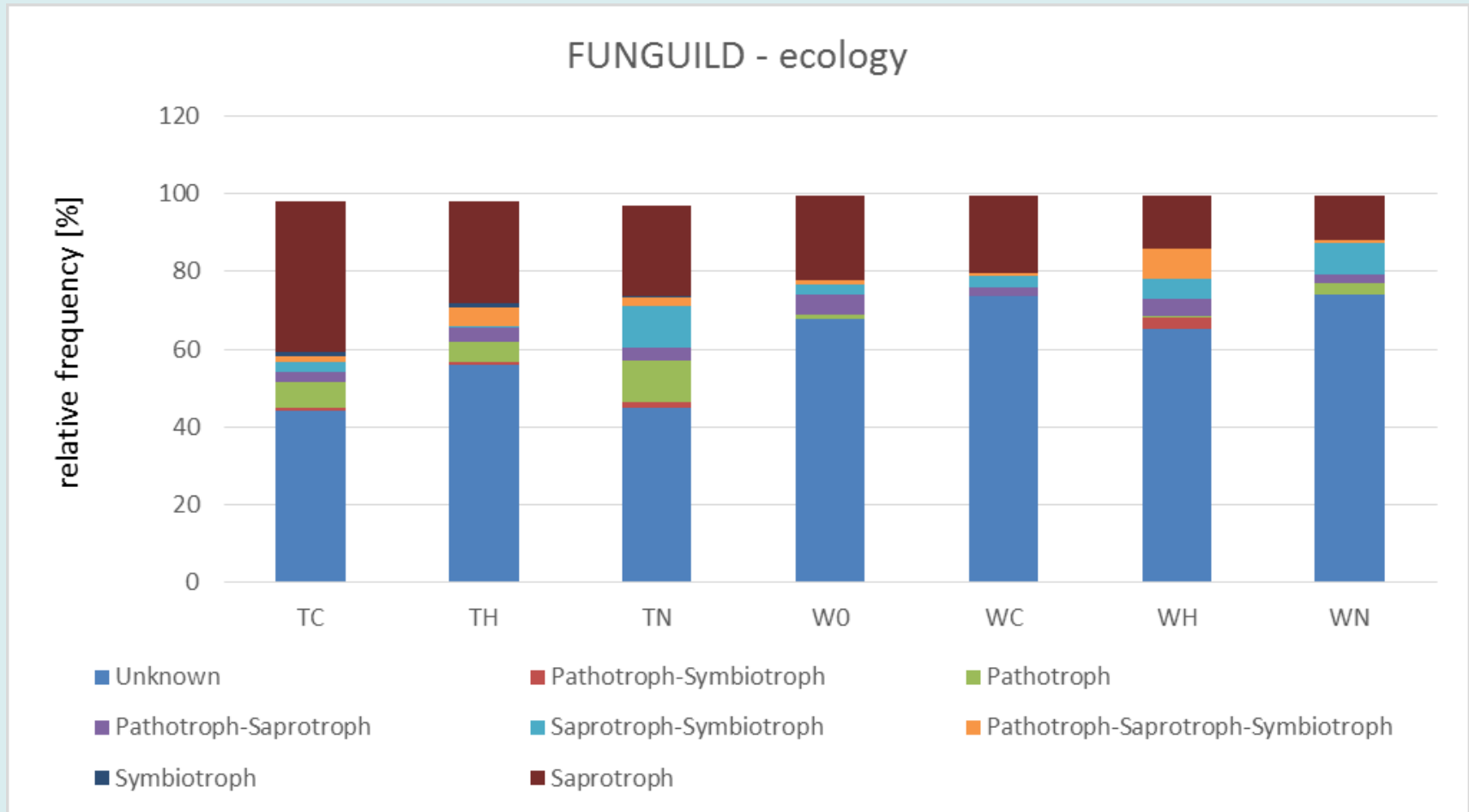


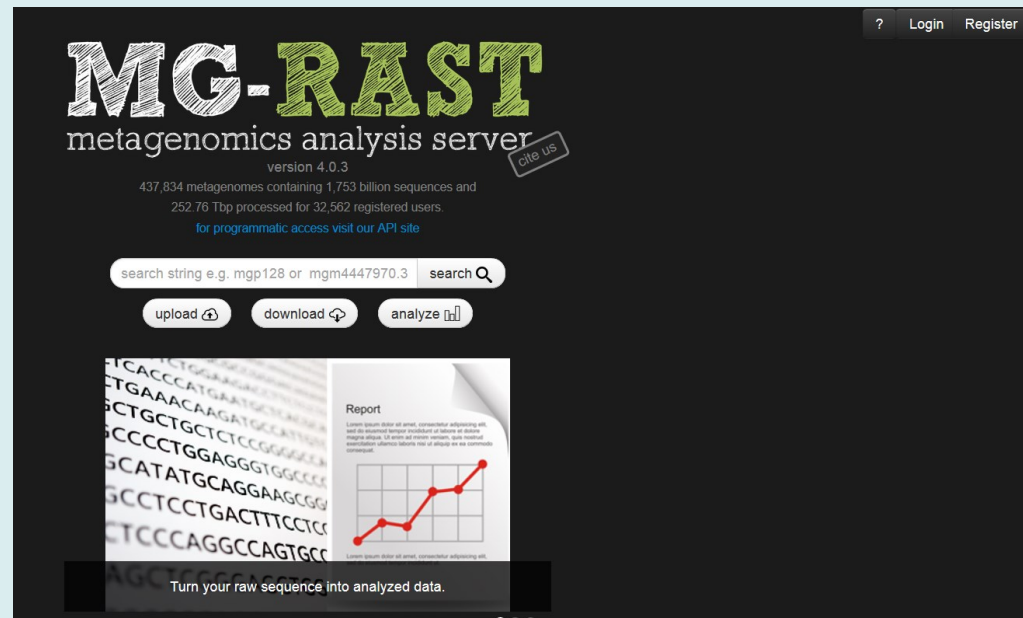
Fig. 1 Abundance of sequences and OTUs at the level of fungal class identified from the ITS and *rpb2* data sets. (a) abundances of fungal classes, (b) relative composition of fungal communities based on sequence and OTU counts. The thin bars in the panel B use red colour to indicate the Ascomycota, blue for the Basidiomycota and green for other fungal lineages.

- Možnosť kategorizovať OTU (či ASV) dle
různych kategorií (ekologie atd)



Využití metabarcoding dat při studiu biogeografie...

- Nejsou jednotná pravidla na deponování NGS dat
- MG-RAST
- NCBI Sequence Read archive (SRA)
- Dryad
- velmi často nikde...
- Lze vyhledávat jen v rámci jednotlivých datasetů



The screenshot shows the MG-RAST metagenomics analysis server homepage. The header features the logo "MG-RAST" in a stylized font, with "metagenomics analysis server" underneath. A "cite us" button is visible. The main content area displays statistics: "437,834 metagenomes containing 1,753 billion sequences and 252.76 Tbp processed for 32,562 registered users." Below this is a search bar with the text "search string e.g. mgp128 or mgm4447970.3" and a "search" button. There are also buttons for "upload", "download", and "analyze". At the bottom, there is a section titled "Report" showing a line graph and a table of data. The footer text reads "Turn your raw sequence into analyzed data."

tool (Letunic and Bork, 2019). Data were deposited in the MG-RAST database under accession number mgp91984 and in NCBI Sequence Read Archive (SRA) under BioProject accession number PRJNA639228. Processed data (extracted ITS2 reads) were depos-

Využití metabarcoding dat při studiu biogeografie...

- <https://globalfungi.com/>
- Actual number of samples in database 20009; Studies 207. Number of ITS sequence variants 119566862; Number of ITS1 sequences 416291533; Number of ITS2 sequences 231278756.

GlobalFungi Database

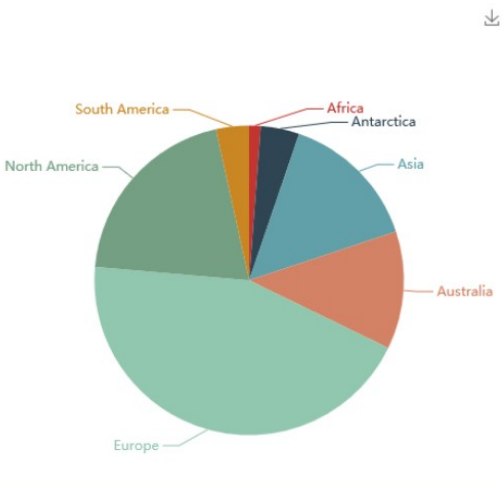
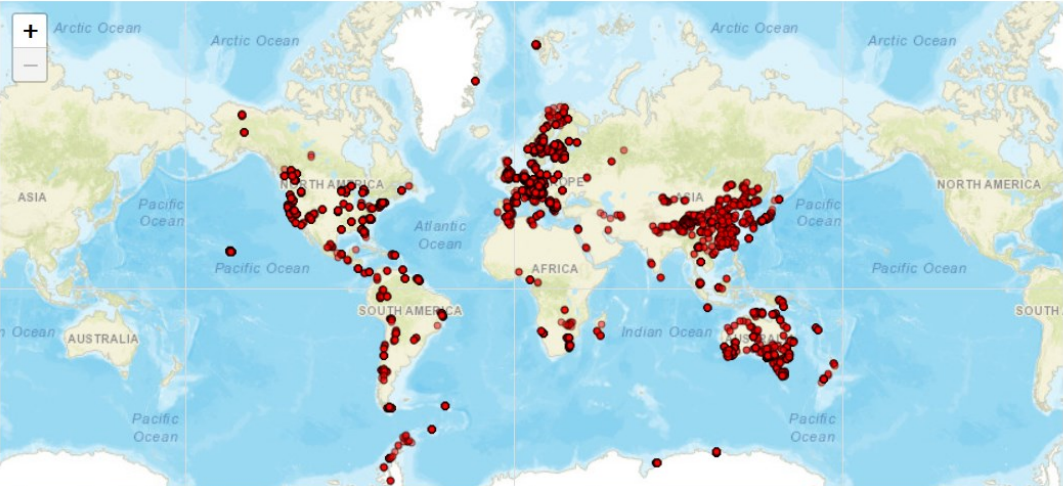
NOTE: optimal usability would be achieved using Firefox or Chrome on PC

Home
Taxon search
Sequence search
Studies
Geosearch
How to cite
About GlobalFungi
Help
Results
Leave a message
Submit your study
Collaborators

Database version v0.9.6; Release version 1.0; Taxonomy based on UNITE version 8.1(2.2.2019); Last update 27.3.2020; First release.
Actual number of samples in database 20009; Studies 207.
Number of ITS sequence variants 119566862; Number of ITS1 sequences 416291533; Number of ITS2 sequences 231278756.

How to use it (tutorial) How to Submit your Study (tutorial)

site design & programming
Tomas Vetrovsky
Daniel Morais
(c) 2020

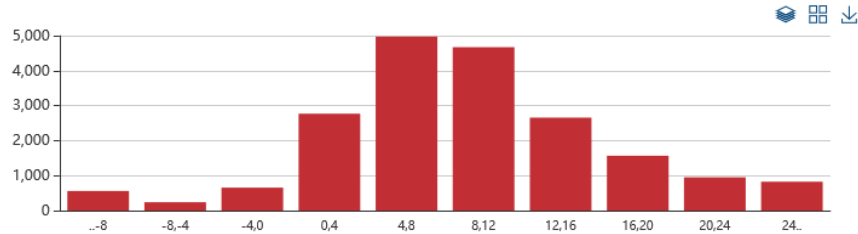


Continent	Approximate Percentage
Europe	35%
Asia	25%
North America	15%
Australia	10%
South America	5%
Africa	3%
Antarctica	1%

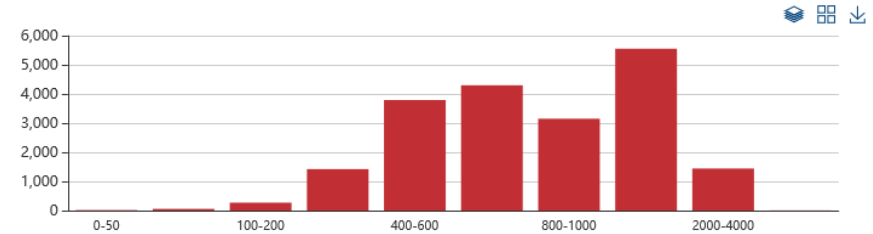
Využití metabarcoding dat při studiu biogeografie...

- <https://globalfungi.com/>

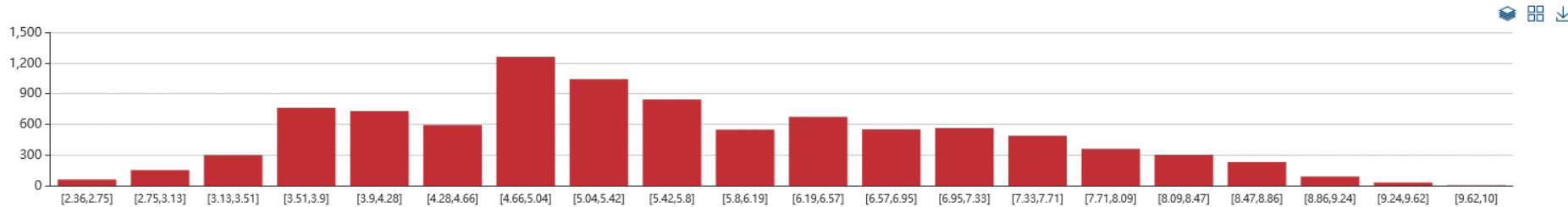
Histogram of MAT of all samples



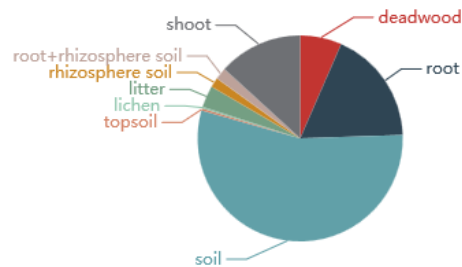
Histogram of MAP of all samples



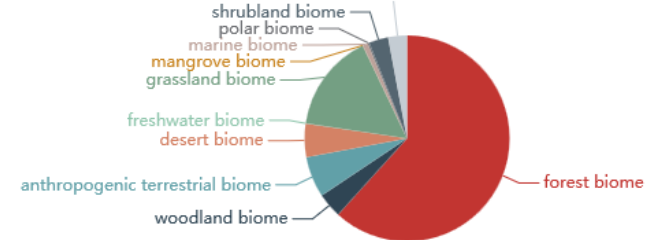
Histogram of pH of all samples (9561 out of 20009 total observations)



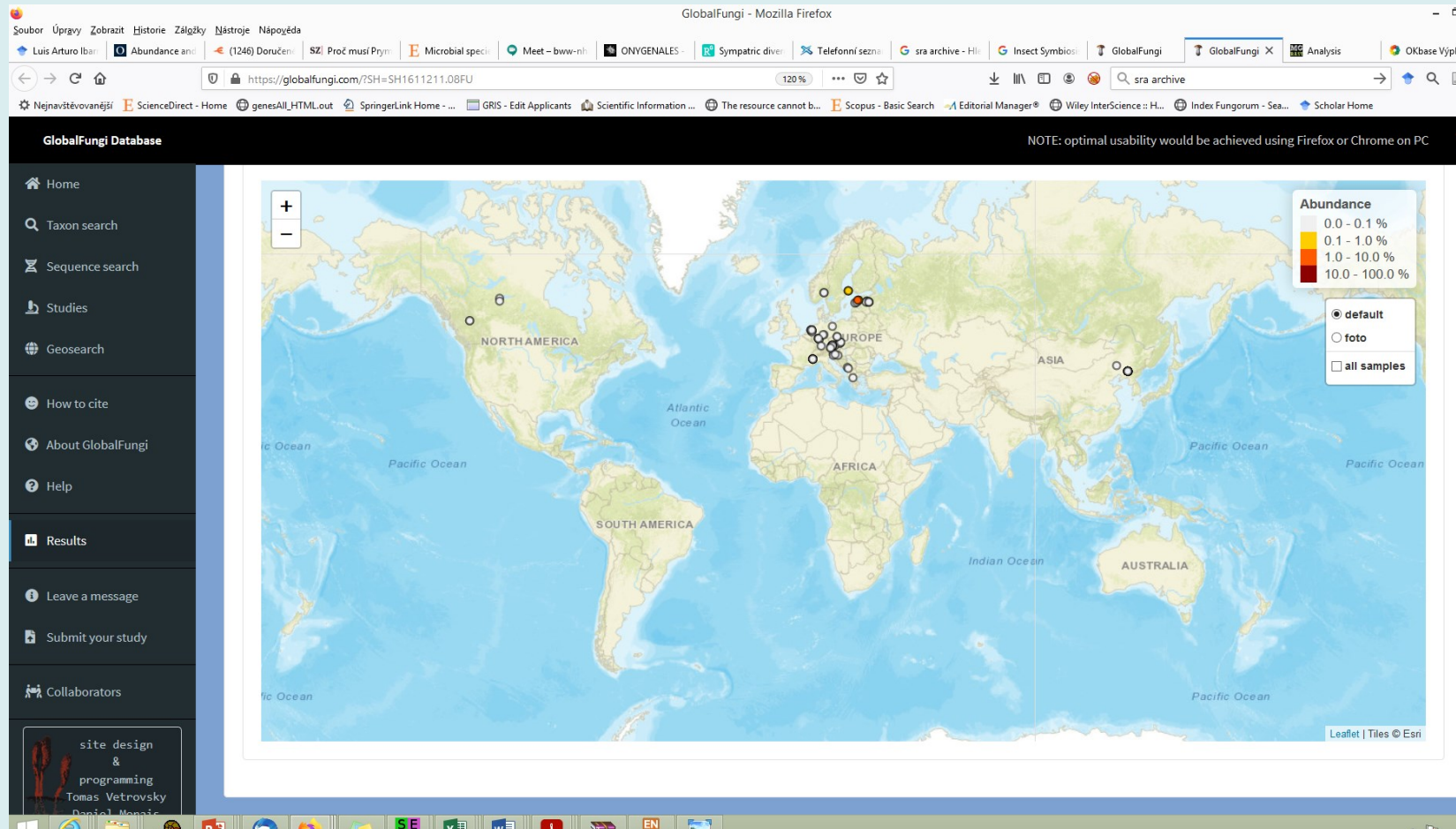
Sample type breakdown of all samples



Sample biome breakdown of all samples



Pholiota squarrosa



Využití metabarcoding dat při studiu biogeografie...

- <https://globalfungi.com/>
- Možnost hledání pomocí algoritmu Blast

Home

Taxon search

Sequence search

Studies

Geosearch

How to cite

About GlobalFungi

Help

Results

Leave a message

Submit your study

Collaborators

site design

 Search by sequence!

Paste your sequence

CCGAAGTACAGGCCCTCTCGTAGGGCTAAACTTCACCCCTTTGTTTATCATACCATGTTGCTTTGGCGAGACGTCTCGGACCACCGGCCCTCGGGCGGGTGCGCGCTCGCCAGAGAAAAATCAAACCCAAACCATTTAGTAGTAGTCTGAAAAAAGTTTCAATTATTA

Choose FASTA file

Browse...

No file selected

Reset/Clear Input

Search type:

☐ Exact hit (input 1-100 sequences; only complete ITS1 or ITS2)

☐ BLAST - best hit (input 1-100 sequences; ITS1 or ITS2)

☒ BLAST - group results (input 1 sequence; ITS1 or ITS2)

Max of BLAST results:

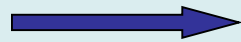
10

Search

Analýza enviromtálních vzorků

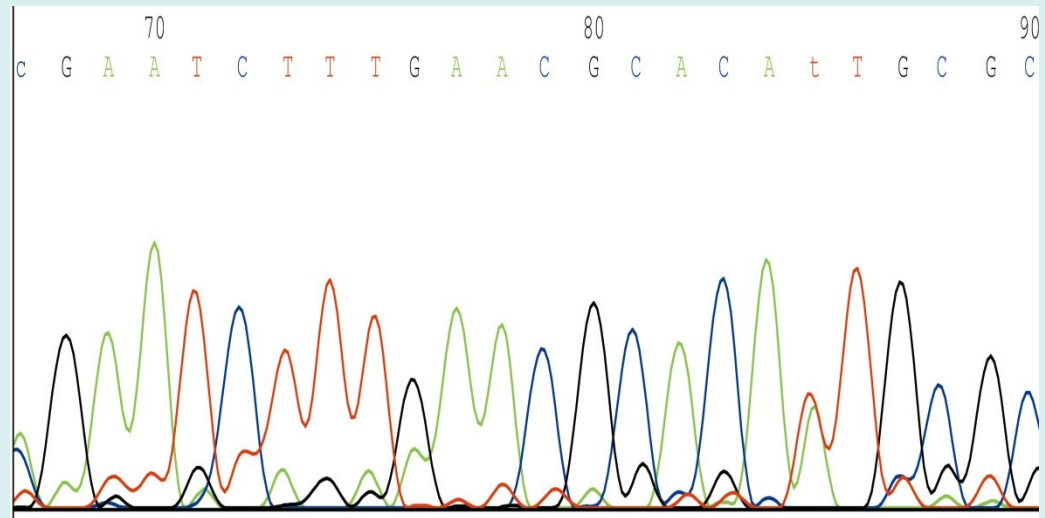
- **Metagenomika, metabolomika, metatranskriptomika („omics“)**
- Sekvenování barcode markeru z env. vzorků
- ITS region (ITS1, ITS2)
- Mnohokopiový marker, problém s kvantifikací
- Vnitrogenomová variabilita
- Rozdílná délka mezi taxony
- Alternativní markery, RPB2, TEF1a,
 - jednokopiové markery
 - Problém s univerzálními primery
 - Specifika PCR s degenerovanými primery
 - Absence alternativních markerů v databázích
 - Problém s introny

Mol. Tax. metody – sekvence DNA



sekvenování

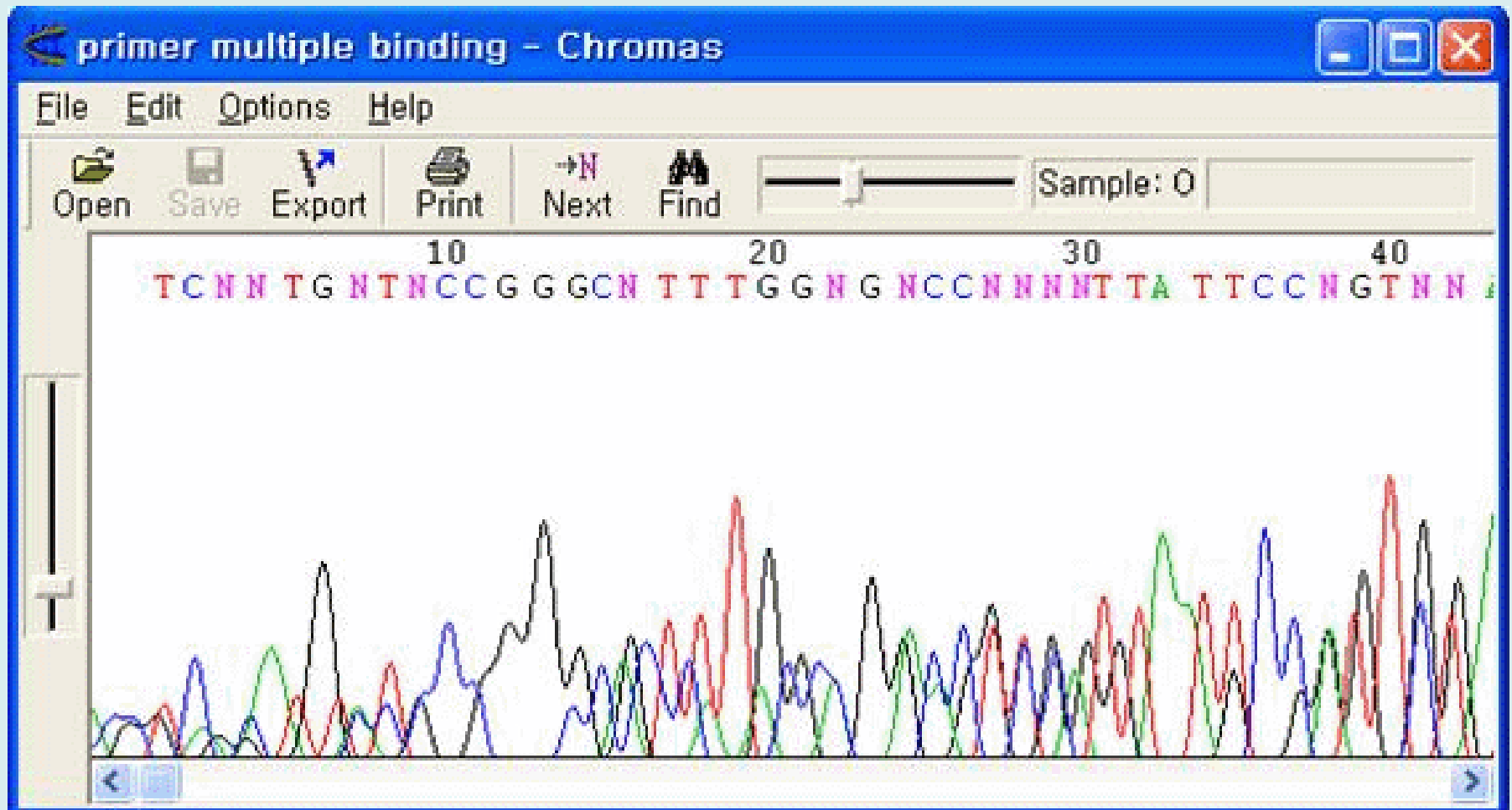
Chromatogram z automatického sekvenátoru



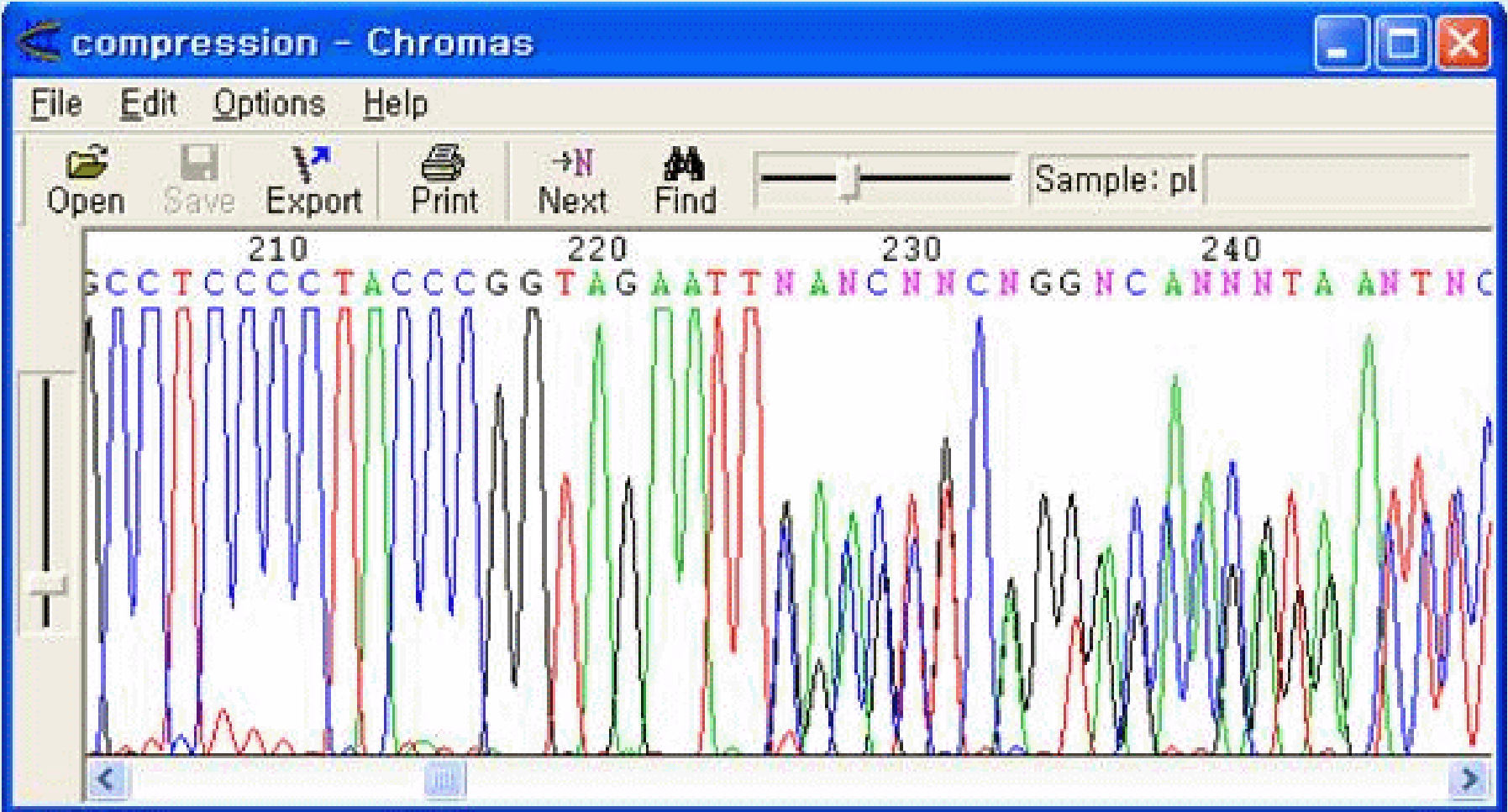
PCR amplifikace

- konečná sekvence je průměrem všech PCR produktů (problém u heterozygotů)
- možnost sekvenovat jednotlivé kopie (klonování, masivní paralelní sekvenování)
- nebezpečí vzniku chimér při PCR ze směsi DNA templátů

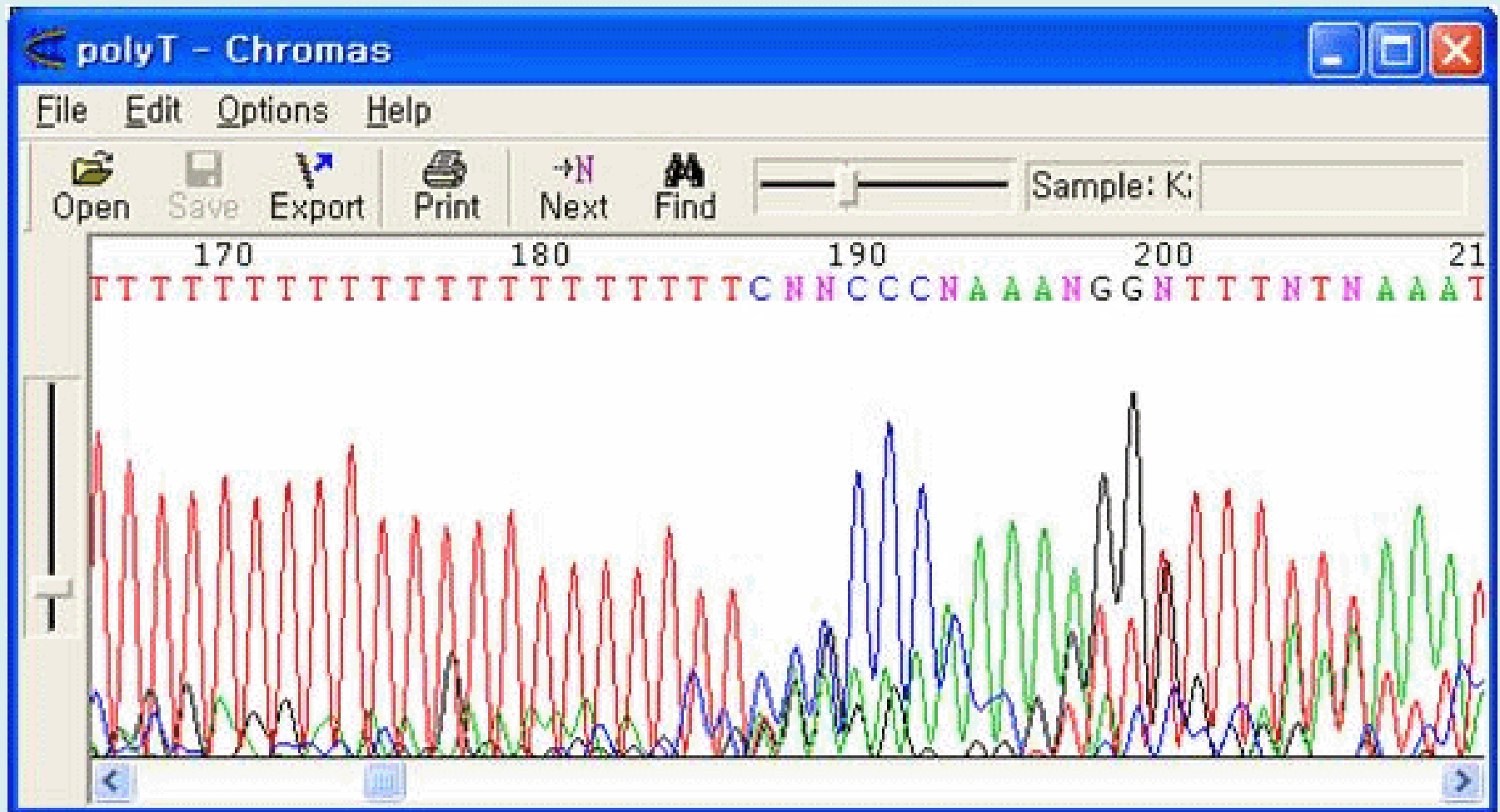
směsný vzorek čí více alel nebo více míst nasednutí primeru



Dvojitý signál od určité pozice. Typicky díky přítomnosti stejného PCR produktu s jinou sekundární strukturou. Často v místech bohatých na G/C či na A/T.

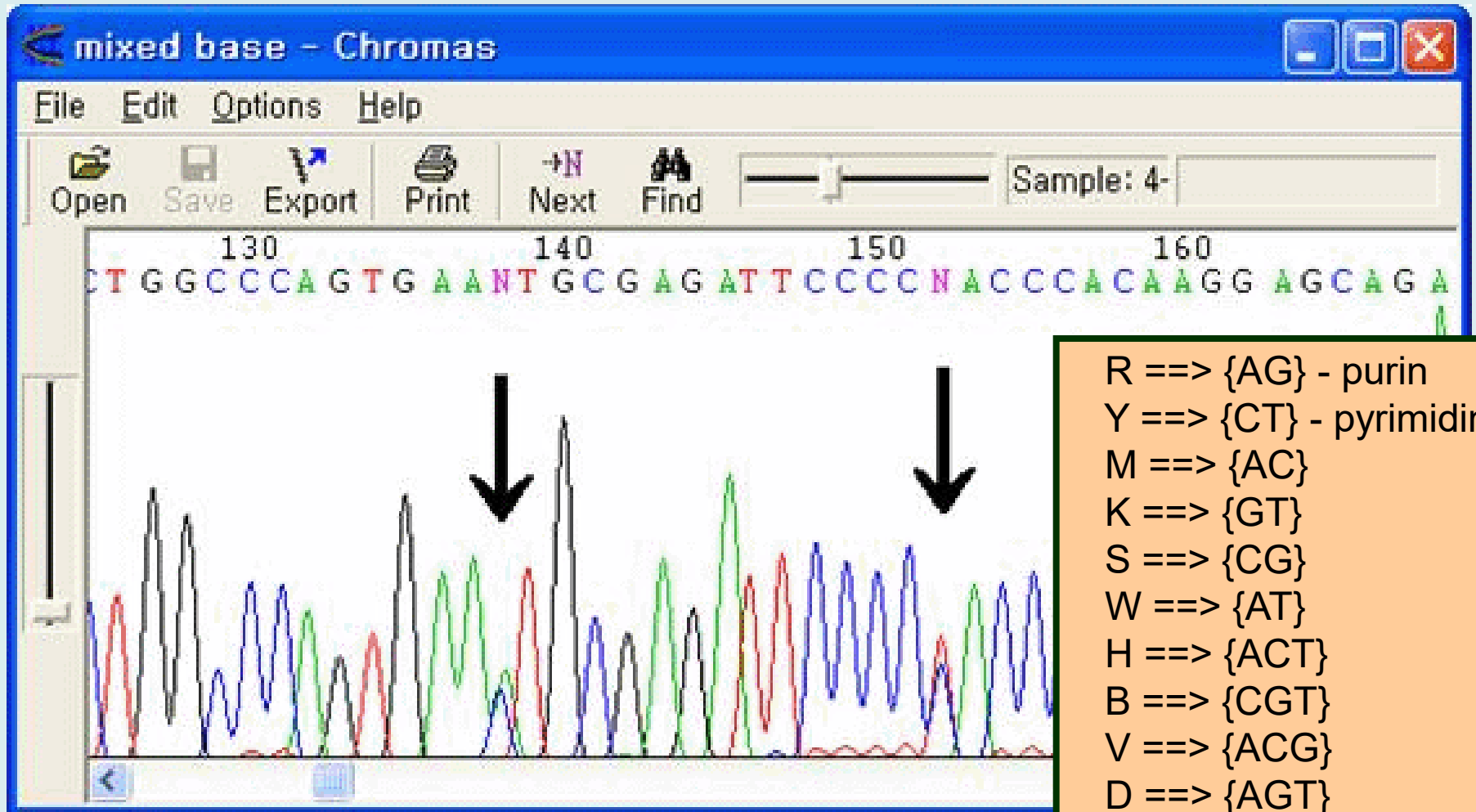


PCR homopolimerních úseků může díky „*slippage*“ – sklouznutí, produkovat variabilní délku těchto úseků.



- typicky u heterozygotů – např. u dikaryotických bazidiomycetů

wobbles (superimposed base calls)

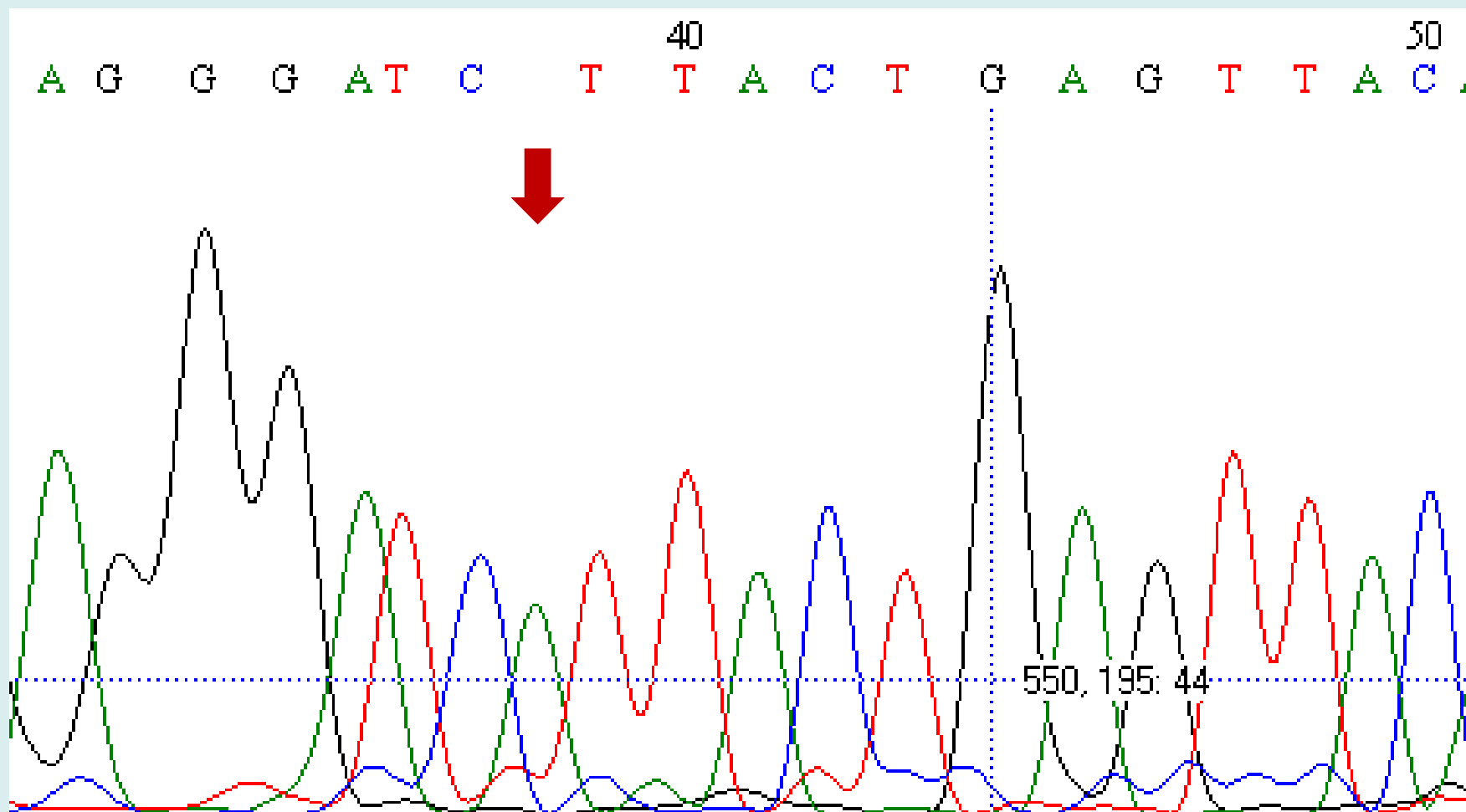


R ==> {AG} - purin
 Y ==> {CT} - pyrimidin
 M ==> {AC}
 K ==> {GT}
 S ==> {CG}
 W ==> {AT}
 H ==> {ACT}
 B ==> {CGT}
 V ==> {ACG}
 D ==> {AGT}
 N ==> {ACGT}

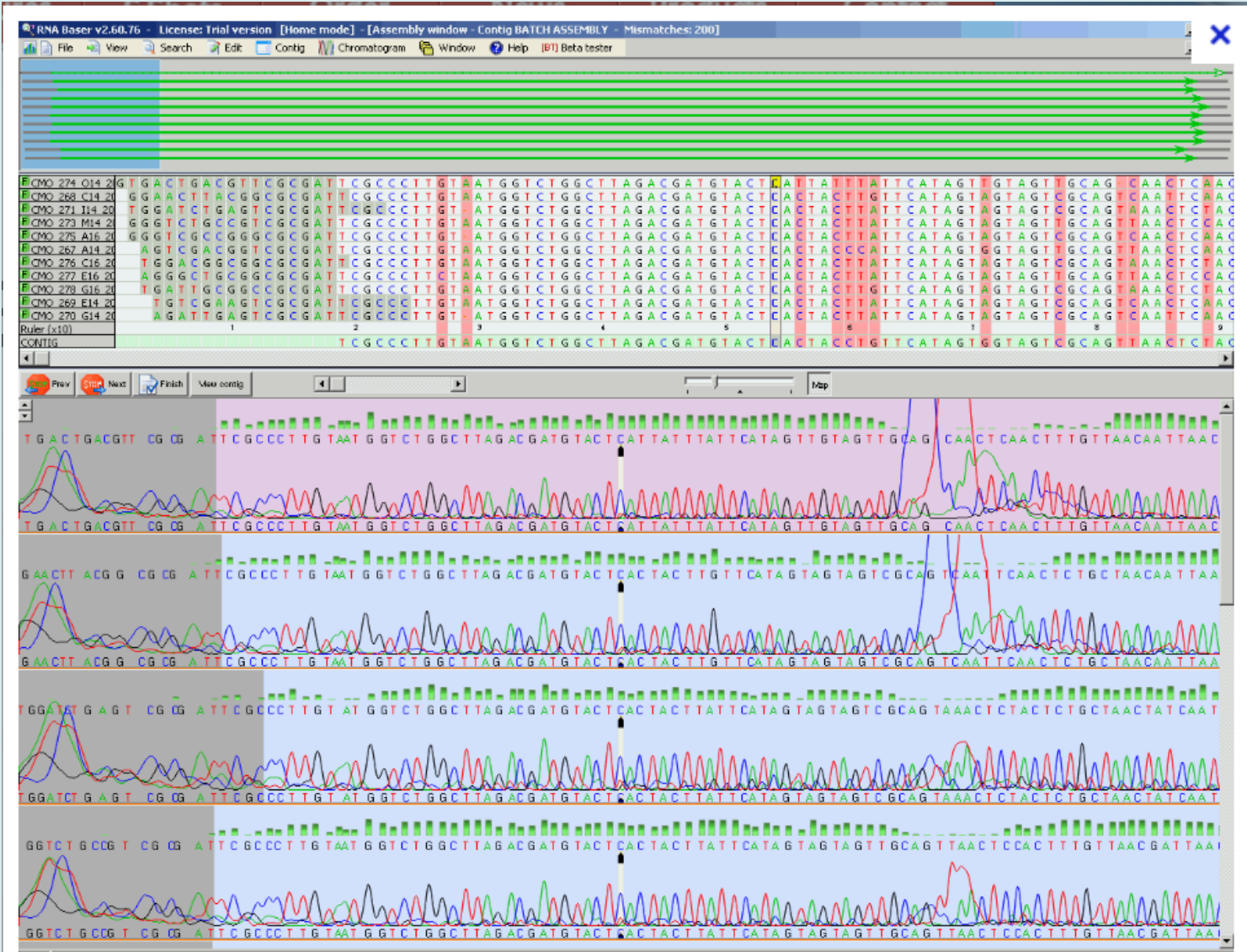
nečistý či degradovaný primer



- Vizuální inspekce chromatogramu
- Sekvenci čteme (sekvenujeme) z obou směrů



- Poslední možnost je hledání podezřelých pozic v alignmentu
- Publikujeme pouze spolehlivé sekvence



```

>Gfla_U67
gcggaggggatcattatcgagttacaaactccaaacctttgtgaacataccac
>Gfla_U334
gcggaggggatcattatcgagttacaaactccaaacctttgtgaacataccac
>Gfla_U134
gcggaggggatcattatcgagttacaaactccaaacctttgtgaacataccac
>Gfla_U212
gcggaggggatcattatcgagttacaaactccaaacctttgtgaacataccac
>Gfla_U302
gcggaggggatcattatcgagttacaaactccaaacctttgtgaacataccac

```

```

#NEXUS

BEGIN DATA;
  DIMENSIONS  NTAX=5  NCHAR=52;
  FORMAT DATATYPE = DNA GAP = - MISSING = ?;
  MATRIX
    'Gfla_U67'  gcggaggggatcattatcgagttacaaactccaaacctttgtgaacataccac

    'Gfla_U334' gcggaggggatcattatcgagttacaaactccaaacctttgtgaacataccac

    'Gfla_U134' gcggaggggatcattatcgagttacaaactccaaacctttgtgaacataccac

    'Gfla_U212' gcggaggggatcattatcgagttacaaactccaaacctttgtgaacataccac

    'Gfla_U302' gcggaggggatcattatcgagttacaaactccaaacctttgtgaacataccac

;

END;

```

```

5 52 I
Gfla_U67      gcggaggggat cattatcgag ttacaaactc caaacctttg tgaacatacc
              ac
Gfla_U334     gcggaggggat cattatcgag ttacaaactc caaacctttg tgaacatacc
              ac
Gfla_U134     gcggaggggat cattatcgag ttacaaactc caaacctttg tgaacatacc
              ac
Gfla_U212     gcggaggggat cattatcgag ttacaaactc caaacctttg tgaacatacc
              ac
Gfla_U302     gcggaggggat cattatcgag ttacaaactc caaacctttg tgaacatacc
              ac

```

formáty sekvencí

- FASTA (*.fas)
- PHYLIP (*.phy)
- NEXUS (*.nex, *.nxs)
- GenBank (*.gen, *.gbn)

editory sekvencí

- BioEdit
- DAMBE
- SeaView
- MEGA
- Geneious...

PHYLIP

Alignment DNA sekvencí

<http://portal.matematickabiologie.cz/index.php?pg=analiza-genomickych-a-proteomickych-dat>

- Přiřazení homologních pozic srovnávaných sekvencí
- Správné přiřazení sekvencí je nutné k odhalení jednotlivých mutací
- delece či inzerce (**=indel**) je označena mezerou (**gap**)

	1	10	20	30	40	50	60	70	80	90	98
Phaeo4	ATCATTACAGAGTTGCAAACTCCCAACCCA-TGCGTACATACC--GTGTCGTTGCTTCGGCGGGCTGCG-----CCTCCCCC--GGGGGGTG										
Oph9	ATCATTACCGAGTTGCAAACTCCCAAAACCACTGCGTACGTACC--ATGTCGTTGCTTCGGCGGGCCGCCG-----CCTCCCCC--GGGGGGGGCG										
Leaf	ATCATTACAGAGTTGCAAACTCCC-AAACCATTTGTGACGTACCC-GTATCGTTTCTTCGGCGGGCGGCTCC-----CCACCCGGGGGCCGCGC										
Thielavia	ATCATTACAGAGTTGCAAACTCCCAAAACCATTTGTGACCTACCCCTAACCGTTGCTTCGGCGGGCGGGCCC-GGGCCTCACCGCCCGGACGCCCC										
raetoniun	ATCATTACAGAGTTGCAAACTCCCTAAACCATTTGTGACGTACCTAACCGTTGCTTCGGCGGGCGGGCCC-GGGGTTTACCCCCGGG--CGCCC										
ialophora	ATCATTACAGAGTTGCAAACTCCCTAAACCATTTGTGACGTACCTAACCGTTGCTTCGGCGGGCGGGCCCCGGGGTTTACCCCCGGG-GCGCCC										
P.nustea	ATCATTACCGAGTTGCAAACTCCCAAAACCAAC-GCGTACCTACC--TACACGTTGCTTCGGCGGTGCCGCCGCGGGG----GGTGCCC-----AGCTG										
Consensus	ATCATTACaGAGTTGCAAACTCCCaAAaCCA.tGcGtAC.TacC.....CGTTgCTTCGgcGgGC.Gcc.c.ggg.....cc.CCC.gg.g.gc.g										
	99	108	118	128	138	148	158	168	178	188	196
Phaeo4	CGGCCCGGGCCGCGAGGGAGCCCCGCCGGAGGT--CTAAA-CTCTGTATCTCAGCGAGCTGTCTTCTGAGTAT-CTTGAAAAATCAAGTAAAACTTT										
Oph9	CGGCCCTGCGCGAGAGGGGGCCCCGCCGGAGGT--CTAAA-CTCTGTATCTCAGCGAGACGTCTTCTGAGCAT-CTTGAAAAATCAAGTAAAACTTT										
Leaf	CTACCCCTCAGGGGTGGCAGCCCCGCCGGAGGCGTCGAAA-CTCTCAGCATCTTATGGCA-TCT-CTGAGTAA-CT-TACAAAATAAGTTAAACTTT										
Thielavia	CCAGGCCCCCTCGCGG-GGC-GCCCCGCCGGAGGTACCCAAA-CTCTTGATATATATGGCC-TCT-CTGAGTCTTCTGTACTGAATCAGTCAAACTTT										
raetoniun	CTGGGCCCCACCGCG-GGC-GCCCCGCCGGAGGTACCCAAA-CTCTTGATATATATGGCC-TCT-CTGAGTCTTCTGTACTGAATCAGTCAAACTTT										
ialophora	CTGGGCCCCACCGCG-GGC-GCCCCGCCGGAGGTACCCAAA-CTCTTGATATATATGGCC-TCT-CTGAGTCTTCTGTACTGAATCAGTCAAACTTT										
P.nustea	CTGGGCCCCCCCCG---GGGCGCCCCGCCGGAGGCCACTAAACTCTTTGTATCTCGCGTAC--CATCTGAG---CGAATGGCGAATAATAAACTTT										
Consensus	CtgggCCcccCcg.g.GGg.GCCCCGCCGGAGGt..CtAAA.CTCTttatatct...ggcc.tCttCTGAGt.t.Ct..a...aa.aAgTaAAACTTT										

Ukázka výstupu z programu BioEdit

- Provedení alignmentu je zatíženo největší subjektivitou ze všech kroků při zpracování sekvenčních dat

SeqA	GARFIELD	THE	LAST	FA-T	CAT
SeqB	GARFIELD	THE	FAST	CA-T	---
SeqC	GARFIELD	THE	VERY	FAST	CAT
SeqD	-----	THE	----	FA-T	CAT

Pairwise alignment - porovnává jen dvě sekvence v jednu chvíli.

Rychlá metoda, která dobře najde podobné sekvence o podobné délce (využívá se při hledání podobných sekvencí v databázi)

Multiple sequence alignment – porovnání více sekvencí najednou, náročné, kvalitní. Různé metody (progresivní, iterativní, strukturální)

local alignment (Smith–Waterman algorithm)

dobře hledá motivy podobných úseků mezi úseky variabilními (pokouší se nalézt podobné úseky v sekvencích, nemusí nezbytně nalézt nejlepší alignment celých sekvencí)

global alignment (Needleman–Wunsch algorithm)

rychlé pro podobné sekvence o podobné délce (snaží se najít nejlepší alignment celých sekvencí)

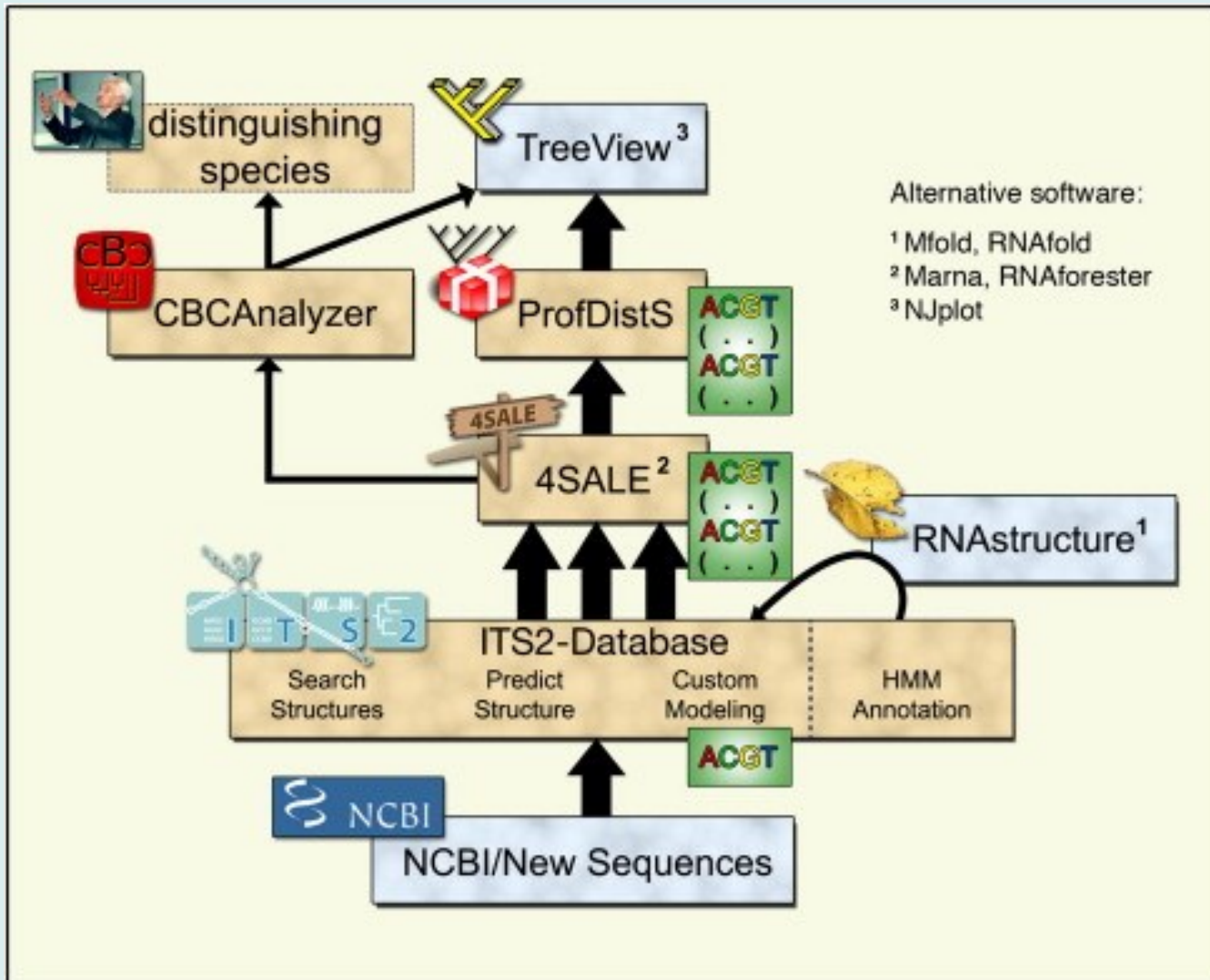
Global	FTFTALILLAVAV
	F--TAL-LLA-AV
Local	FTFTALILL-AVAV
	--FTAL-LLAAV--

-

Gap open penalty, Gap extension penalty – možnost penalizace za včlenění a rozšíření gapu, zvětšujeme když neočekáváme introny, indely

Structural alignment - alignemnt s přihlédnutím na sekundární strukturu

- alignemnt proteinů může zohlednit i jeho terciální strukturu
- rDNA lze alignovat dle sekundární struktury (např. 4SALE)



- alignment protein kódující sekvence proteinů musí respektovat kodony)

Sekv1	AGT	CGG	TAA	AGAT	CGGA
Sekv2	AGT	GGG	ACA	AGATT	TGA
Sekv1	AGT	C ~ ~ ~	CA	AGAT	CGGA

Sekv1	AGT	CGG	TAA	AGAT	CGGA
Sekv2	AGT	GGG	ACA	AGATT	TGA
Sekv1	AGT	— — —	CCA	AGAT	CGGA

Pohodlná volba parametrů v on-line programech

- Řada strategií a programů.
- Nejpoužívanější: MAFFT, MUSCLE, T-Coffe, ClustalW

MAFFT version 7
Multiple alignment program for amino acid or nucleotide sequences

[Download version](#)
[Mac OS X](#)
[Windows](#)
[Linux](#)
[Source](#)

[Online version](#)
[Alignment](#)
[mafft --add](#)
[Merge](#)
[Phylogeny](#)
[Rough tree](#)
[Merits / limitations](#)
[Algorithms](#)
[Tips](#)
[Benchmarks](#)
[Feedback](#)
[Follow](#)

SubmitReset

Advanced settings

Strategy:

☒ Auto (FFT-NS-1, FFT-NS-2, FFT-NS-i or L-INS-i; depends on data size) [Updated](#)

Progressive methods

☐ FFT-NS-1 (Very fast; recommended for >2,000 sequences; progressive method)

☐ FFT-NS-2 (Fast; progressive method)

☐ G-INS-1 (Slow; progressive method with an accurate guide tree)

Iterative refinement methods

☐ FFT-NS-i (Slow; iterative refinement method)

☐ E-INS-i (Very slow; recommended for <200 sequences with multiple conserved domains and long gaps) [Help](#) [Updated](#) (2015/Jun)

☐ L-INS-i (Very slow; recommended for <200 sequences with one conserved domain and long gaps) [Help](#)

☐ G-INS-i (Very slow; recommended for <200 sequences with global homology) [Help](#)

☐ Q-INS-i (Extremely slow; secondary structure of RNA is considered; recommended for a global alignment of highly divergent ncRNAs with <200 sequences $\times$ <1,000 nucleotides; the number of iterative cycles is restricted to two, 2016/May) [Help](#)

Align unrelated segments, too? *in Alpha Testing* (2014/Mar)

If the input data is expected to be globally conserved but locally contaminated by unrelated segments, try 'Unalignlevel>0' and possibly 'Leave gappy regions'.

Unalignlevel:

0.0

0.8

↑ Default

This feature is available only when G-INS-1 or G-INS-i is selected in the **Strategy** section above.

☒ Try to align gappy regions anyway

☐ Leave gappy regions (Not recommended for >~1,000 sequences)

Parameters:

Scoring matrix for amino acid sequences:

BLOSUM62

Scoring matrix for nucleotide sequences:

200PAM / $\kappa=2$

↑ Switch it to '1PAM / $\kappa=2$ ' when aligning closely related DNA sequences.

Gap opening penalty:

1.53

 (1.0 - 5.0)

Offset value:

0.0

 (0.0 - 1.0)

Score of N in nucleotide data: [Example](#)

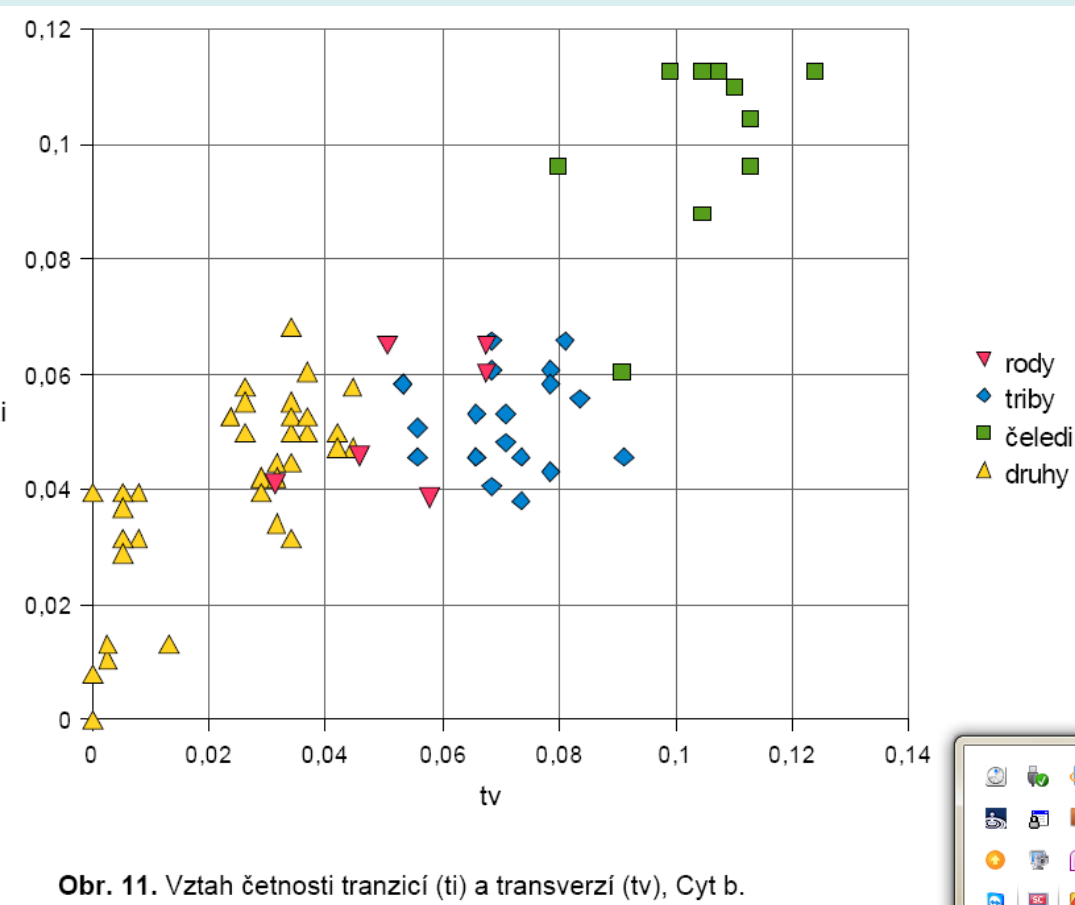
↑ Long stretches of Ns tend to be gapped (excluded from the alignment).

☒ (nzero) N has no effect on the alignment score.

☐ (nwildcard) N is treated like a wildcard. **Experimental option** (2016/Apr/26)

↑ Try this if Ns should be aligned with usual letters.

- **substituční saturace (nucleotide substitution saturation)**– jedna pozice může projít více substitucemi, což nám následně nedovolí odhadnout skutečný počet evolučních změn
- Odhady substituční saturace nukleotidů (např. podíl transverzí a tranzicí)
- tranzice — změna purinu (A,G) na purin nebo pyrimidinu (C,T) na pyrimidin
- transverze — změna purinu na pyrimidin nebo naopak ($A \rightarrow T$, $T \rightarrow A$, $C \rightarrow A$, $A \rightarrow C$, $G \rightarrow T$, $T \rightarrow G$, $G \rightarrow C$, $C \rightarrow G$)



- Reálná frekvence tranzic je větší než frekvence transverzí
- Tj. u saturaovaných pozic bude přibývat transverzí



Obr. 11. Vztah četnosti tranzicí (t_i) a transverzí (t_v), Cyt b.

Gblocks 0.91b Results

Processed file: **nad3.pir**

Number of sequences: **17**

Alignment assumed to be: **Protein**

New number of positions: **56** (selected positions are underlined in blue)

```
      10      20      30      40      50      60
=====+=====+=====+=====+=====+=====+
nad3_parde -----MEYLLQEYLPILVFLGMSALAIVLILAAAVIAVRN--PDPEKVSAYECGFNAF
nad3_acaca -----MTLEYIYIFIFFWGAFFISCLLIFLSYFLVYQE--SDIEKNSAYECGFQPF
nad3_allma -----MTYLVVYIVFTTIVLVGLILVLSYLLSQAQ--PDSEKVSAYECGFSP
nad3_apec -----IFNFLTLFVLSILIFLITTLITFAAHFLPSRN-TD-SEKSSPYECGFDP
nad3_arath -----MMSEFAPISIYLVISLLVSLILLGVPPFFASNS-STYPEKLSAYECGFDP
nad3_balca -----MNSFLIYLLIAITLSFILLSIVGHRIPTRN-MD-CEKLSPYECGFDP
nad3_chocr -----MKLIFTEYSAILIFFAFAISSLSVIFLLSYFLIPQK--PDCEKVSAYECGFNPF
nad3_drome -----MFSIIFIALILLITTIVMFLASILSKKA-LIDREKSSPFECGFDPK
nad3_human -----MN-FALILMINTLLALLLMIITFLLPOLN-GY-MEKSTPYECGFDP
nad3_ktun -----MFFVLSLVLFSTFLLSLVLLSVLSLTTKKK-MMNREKSSPFECGFDPK
nad3_lter -----MILTALSSAIALLVPIIILGAAWVLASRS-TEDREKSSPFECGFDPK
nad3_marpo -----MEFAPIFVYLVISLLLSLILLIGVSFLFASSSSLAYPEKLSAYECGFDP
nad3_metse -----MYTEFYGILVLLIFSVVLSAISGASYILGDKQ--PDREKVSAYECGFDP
nad3_picca MLNYFVYPYGIENDMGMKFYMMLVPMMSMVLMMINYMNTNKS-DNNMNMKTGPECGFDSF
nad3_podan -----MSSMTLFIILFVSIALLFLFINLIFAPHN-SEYQKYSLEECGFHSF
nad3_prowi -----MYEFLGILIIYFFIALALSLLLLGLPFLVSTRK--ADPEKISAYECGFDP
nad3_recam -----MNTMILSEYLSVLIFFIFSFGLSCIIILGLSYVLATQN--ADTEKLSPYECGFNPF
```

```
      70      80      90     100     110     120
=====+=====+=====+=====+=====+=====+
nad3_parde D-DARMKFDVRFYLVLSILFIIFDLEVAFLFPWAVSFASLS-DVAFWGLMVFLAVLTVGFA
nad3_acaca E-DTRSKFNVRYYLIAILFMIFDLEIMYLFPPWSISISTGS-FFGVWAIFLFLIILTVGFI
nad3_allma G-DARQKFDVSRFYLIAILFIIFDLEVVFILFFASVIHNVS-LLGGMITIIFLVILTIGFI
nad3_apec N-SARVPFVSFRFFFLVAIFLFLFDLEIALLFFLPFSVFFH--P--IHTP----LILTVGLI
nad3_arath G-DARSRFDIRFYLVLSILFLLIPDLEVTTFPPWAVPPNKID-LFGFWSMMAFLFILTIGFL
nad3_balca A-SARLPFSLRFFFLVAIFLFLFDLEIALLLPFAALSARDPQLSFTLAFLLILILTIGLI
nad3_chocr D-DARATFDIRFYLVAILFLIFDLEISFLFPWSLVLGEIS-IIGFWSMIVFLVILTIGFI
nad3_drome S-SSELFPFSLRFFLITIIIFLFDVEIALILPMIIMKYSNIMIWTITSIIIFILILLIGLY
nad3_human S-PARVPFSMKRFFLVAITFLLFDLEIALLLPLPWALQTTNLPMLVMSSLLLIILALSLSA
nad3_ktun S-SARLPFSMRFFLITVVFVLFVDVEIVLPLLPLFSSGWSIDVFSLVGSMMLVILIIIGVL
nad3_lter S-TARIPFSTRFFLLAIIFIVFDIEIVLLMPLPTILHTSDVFTTIVTTSVLFMLIIGLI
nad3_marpo D-DARSRFDIRFYLVLSILFIIFDLEVTFLLPWAVALNKIG-LFGFWSMMVFLFILTIGFV
nad3_metse G-TPGRPFSSIRFFLIGILFIIFDLEISFLFPWCVCVNQVF-PFGFVMTMIVFLAVLTLGLV
nad3_picca R-QSRTTYSIRFILLIAILFLFFDLELTSILPYTLSMYNTN-IYGLFILLYFLLPLIIGFI
nad3_podan LGQNRQTQGVKRFIFALVYLLLDLEILLTFPFVAVSEYVNM-IYGLIILLGFITIIIGFV
nad3_prowi D-DARGRFDICFYLVAILFIIFDLEVAFLFPWALTNLNKIG-YFGFWSMMLFLFILTIGFI
nad3_recam D-DARGAFDVRFFYLVAILFIIFDLEVAFLFPWAVALSDVT-IFGFWTMFIIFLLILTIGFI
```

Gblocks Server

Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis

About the Gblocks Server

Version 0.91b, January 2002

Copyright © Jose Castresana



Gblocks eliminates poorly aligned positions and divergent regions of a DNA or protein alignment so that it becomes more suitable for phylogenetic analysis. This server implements the most important features of the Gblocks program to make its use as simple as possible without loosing the functionality that it is necessary in most of the cases. Other options can be changed in the stand-alone program. You can see here an [example output file](#) showing the blocks selected from a protein alignment. Further information can be found in the [online documentation](#). Please see the [Gblocks](#) page for citations.

Gblocks Server

Paste an alignment in NBRF/PIR or FASTA format:

Or upload an alignment file:

Procházet...

Type of sequence:

DNA ☐ || Protein ☒ || Codons ☐

Options for a less stringent selection:

- ☐ Allow smaller final blocks
- ☐ Allow gap positions within the final blocks
- ☐ Allow less strict flanking positions

Options for a more stringent selection:

- ☐ Do not allow many contiguous nonconserved positions

[Get Blocks](#)

[Clear](#)

View

Reformat to GCG, PHYLIP, MSF, NEXUS, uppercase/lowercase, *etc.* with Readseq

GUIDANCE2 computes the residue-wise confidence scores and extracts well-aligned residues.

Refine dataset

Phylogenetic tree *Visualization updated, 2016/Sep*

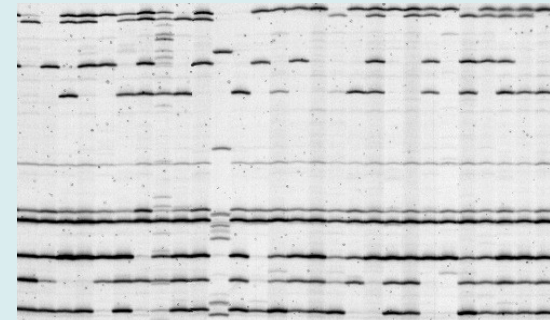
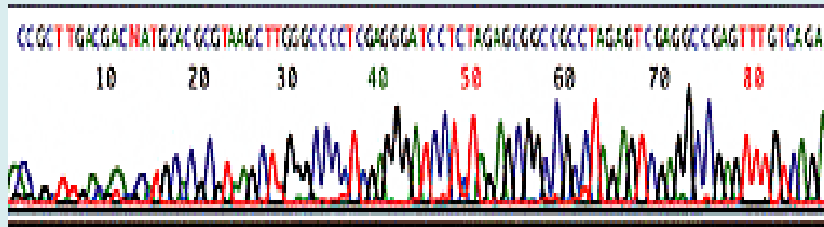
MAFFT-[L-INS-i](#) Result

CLUSTAL format alignment by MAFFT (v7.306b)	
1LYLA	FNDELNRNRREKL---AALRQQGVAFPNDFR-----RDHTSDQLHEEFDAKD
c p2	MP-----
1B8AA	MYR-----THYSSEITEELNGQK
1ASZB	EDTAKDNYGK-----LPLIQSRDSDRTGQKRVKFVDLDEAKDSK
c p1	MRTEYCGQLR-----
1ADJA	TAR-----
j p2	M-----
j p2	LRDNMK-----MLLIH---SDYLEFEAKEKTKIAEETENLK
c p1	MPVITLPDGSQRHYDHAVSPMDVAL--DIGPGLAKACIAGRVNGELVDACDLIENDAQLS
c p1	MRT-----SQYLLSTLKETPADAE

Zpracování mol. genet. dat

Typy dat

- **znaková (*characters*)** - sekvence, kde známe přímo pořadí bazí a každé přísluší hodnota znaku, přítomnost daného PCR fragmentu



- **distanční (*distance*)** - podobnosti při DNA-DNA hybridizaci, či znaková data převedené na distance

	<i>B. sesqui</i>	<i>E. terrico</i>	<i>A.altern</i>	<i>Geo III</i>	<i>P.pallid</i>	<i>Geo II</i>	<i>Geo I</i>	<i>G.putter</i>	<i>G.laven</i>	<i>Geo XI</i>	<i>Geo X</i>	<i>Geo V</i>	<i>Geo VI C</i>	<i>Geo VI A</i>	<i>Geo VII</i>	<i>Geo VIII</i>	<i>Geo IV</i>	<i>Geo IX</i>
<i>B. sesquicillii</i>	0.00000	0.23119	0.35323	0.27937	0.32078	0.28097	0.30348	0.30039	0.28617	0.29055	0.30052	0.30713	0.30713	0.30713	0.31350	0.32717	0.32094	0.30980
<i>E. terricola</i>	0.23119	0.00000	0.24151	0.19817	0.20600	0.21679	0.20423	0.23158	0.22160	0.19414	0.19590	0.20083	0.20514	0.20514	0.21920	0.23604	0.21674	0.22605
<i>A.alternatum</i>	0.35323	0.24151	0.00000	0.14532	0.14645	0.17076	0.15328	0.16361	0.18294	0.16263	0.16273	0.16696	0.16696	0.16696	0.18115	0.19340	0.18027	0.17182
<i>Geo III</i>	0.27937	0.19817	0.14532	0.00000	0.02170	0.01944	0.01144	0.05360	0.06917	0.05797	0.06087	0.06099	0.06099	0.06099	0.07978	0.07945	0.08305	0.07212
" <i>P.pallidum</i> "	0.32078	0.20600	0.14645	0.02170	0.00000	0.02361	0.01340	0.06624	0.07146	0.06052	0.06092	0.05575	0.05813	0.05813	0.06636	0.07351	0.07693	0.06388
<i>Geo II</i>	0.28097	0.21679	0.17076	0.01944	0.02361	0.00000	0.01935	0.06371	0.07951	0.06771	0.07078	0.06539	0.06785	0.06785	0.08174	0.08403	0.08502	0.07410
<i>Geo I</i>	0.30348	0.20423	0.15328	0.01144	0.01340	0.01935	0.00000	0.05548	0.06600	0.05531	0.05814	0.05306	0.05306	0.05306	0.06847	0.07053	0.07387	0.06104
<i>G.putterilli</i>	0.30039	0.23158	0.16361	0.05360	0.06624	0.06371	0.05548	0.00000	0.05852	0.05085	0.05582	0.06049	0.06049	0.06049	0.08189	0.07904	0.08736	0.06870
<i>G.lavendla</i>	0.28617	0.22160	0.18294	0.06917	0.07146	0.07951	0.06600	0.05852	0.00000	0.04130	0.04580	0.04137	0.04361	0.04361	0.05489	0.05515	0.07036	0.05285
<i>Geo XI</i>	0.29055	0.19414	0.16263	0.05797	0.06052	0.06771	0.05531	0.05085	0.04130	0.00000	0.01137	0.01531	0.01729	0.01729	0.03418	0.02984	0.03645	0.02153
<i>Geo X</i>	0.30052	0.19590	0.16273	0.06087	0.06092	0.07078	0.05814	0.05582	0.04580	0.01137	0.00000	0.01134	0.01328	0.01328	0.03431	0.02995	0.03198	0.01730
<i>Geo V</i>	0.30713	0.20083	0.16696	0.06099	0.05575	0.06539	0.05306	0.06049	0.04137	0.01531	0.01134	0.00000	0.00183	0.00183	0.02334	0.01925	0.02738	0.01326
<i>Geo VI C</i>	0.30713	0.20514	0.16696	0.06099	0.05813	0.06785	0.05306	0.06049	0.04361	0.01729	0.01328	0.00183	0.00000	0.00000	0.02538	0.02126	0.02945	0.01521
<i>Geo VI A</i>	0.30713	0.20514	0.16696	0.06099	0.05813	0.06785	0.05306	0.06049	0.04361	0.01729	0.01328	0.00183	0.00000	0.00000	0.02538	0.02126	0.02945	0.01521
<i>Geo VII</i>	0.31350	0.21920	0.18115	0.07978	0.06636	0.08174	0.06847	0.08189	0.05489	0.03418	0.03431	0.02334	0.02538	0.02538	0.00000	0.02314	0.03846	0.02554
<i>Geo VIII</i>	0.32717	0.23604	0.19340	0.07945	0.07351	0.08403	0.07053	0.07904	0.05515	0.02984	0.02995	0.01925	0.02126	0.02126	0.02314	0.00000	0.04266	0.02543
<i>Geo IV</i>	0.32094	0.21674	0.18027	0.08305	0.07693	0.08502	0.07387	0.08736	0.07036	0.03645	0.03198	0.02738	0.02945	0.02945	0.03846	0.04266	0.00000	0.01529
<i>Geo IX</i>	0.30980	0.22605	0.17182	0.07212	0.06388	0.07410	0.06104	0.06870	0.05285	0.02153	0.01730	0.01326	0.01521	0.01521	0.02554	0.02543	0.01529	0.00000

Zpracování mol. genet. dat

- kolik evolučních změn se událo mezi jednotlivými sekvencemi ?
- zjistíme pomocí koeficientů pro odhad genetických vzdáleností

	A	B	C	D
A	-			
B	0.5	-		
C	0.45	0.15	-	
D	0.55	0.4	0.35	-

= modely pro substituce nukleotidů (*nucleotide substitutions models*)

- **Výsledkem je matice genetických vzdáleností (jedno číslo pro každou dvojici)**

- **Nei & Li** koeficient, **Jaccardův** koeficient a **Simple matching** koeficient - pro binární data
- **Koeficient Jukes-Cantor**: stejná frekvence nukleotid. substitucí pro všechny nukleotidy

- **Tamura-Nei**: výměna (transice) mezi puriny (A,G) není stejně častá jako mezi pyrimidiny (T,C)

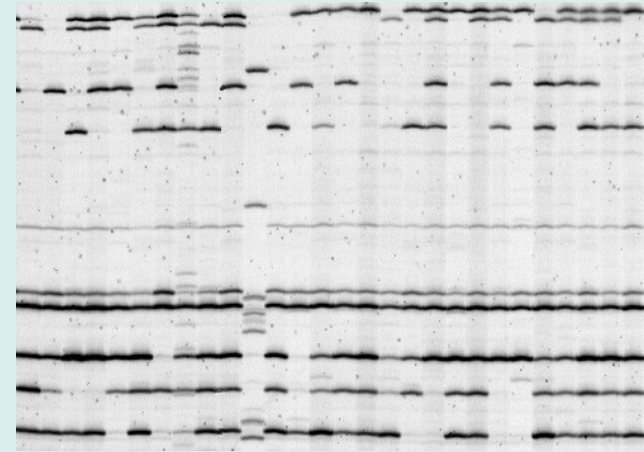
- **HKY, T92, GTR...**

Mutační rychlost se v sekvenci může lišit místo od místa. Toto rozložení je specifikováno koeficientem gama (γ), který je odvozen z rozptylu míry substitucí.

Nejvhodnější model a hodnotu gama může spočítat - **MODELTEST** (PAUP)

V případě analýzy rRNA lze brát v úvahu její sekundární strukturu

Zpracování mol. genet. dat

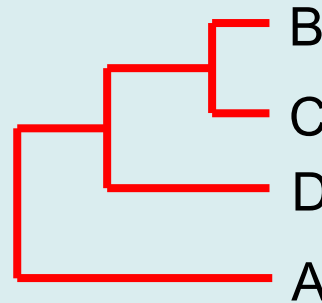


binární data

- fingerprint převeden na binární matici
(proužek přítomen - 1, nepřítomen - 0)
- Z binární matice se provede odhad genetické podobnosti mezi jednotlivými organismy pomocí koeficientů
 - Nei & Li (NL) koeficient
 - Jaccardův koeficient (J)
 - Dice-Sorensenův koeficient
- koeficienty jsou podobnostmi. Při konstrukci fylogenetických stromů se počítá s genetickými vzdálenostmi (distance). Distance = 1 – podobnost
- fylogram či kladogram se počítá některou metodou pro distanční data (**UPGMA, Neighbor joining, mnohorozměrné metody – PCoa, NMDS**)

Konstrukce fylogenetických stromů

A. Distanční metody



	A	B	C	D
A	-			
B	0.5	-		
C	0.45	0.15	-	
D	0.55	0.4	0.35	-

- z genetických vzdáleností sestrojíme fylogenetický strom pomocí algoritmu
- výpočet je rychlejší a poskytuje pouze jeden strom, ale nemusí být přesný (riziko homoplázie)
- u podobných sekvencí funguje dobře

- Shlukovací algoritmus **UPGMA** (*Unweighted Pair Grouping Method Using Arithmetic Average*)

- UPGMA počítá se stejnou rychlostí kumulace mutačních změn v jednotlivých větvích kladogramu

- pouze nezakořeněné kladogramy

- **Neighbor-joining** - počítá s různou mutační změnou v různých větvích a lze jí tvořit zakořeněné kladogramy = **fylogramy**

- **Minimální evoluce** – ze všech vygenerovaných stromů vybere nejkratší

Zpracování mol. genet. dat – analýza sekvencí

Konstrukce fylogenetických stromů – znaková data

- nevychází z matice gen. distancí („% homologie mezi sekvencemi“)
- využívá se přímo alignment, každá pozice alignmentu se porovnává zvlášť
- **hledáme strom**, který s největší pravděpodobností vygeneruje pozorovaná data za platnosti evolučního modelu (opět modely typu K2P, JC, GTR...)
- výpočetně náročné
- metody generují více stromů a vytvoří se jeden strom konsenzuální (*strict consensus, majority rule consensus*)

Metody znakové

- **metoda maximální úspornosti (*maximum parsimony*)**
Tento soubor metod hledá takový fylogenetický strom, který by vyžadoval co nejmenší počet evolučních změn (mutací), nepracuje s evolučním modelem (neparametrická metoda)
- **metoda maximální pravděpodobnosti (*maximum likelihood*)**
Přesnější a náročnější. Umožňuje začlenění modelu a řady dalších parametrů. Odhalí „long branch attraction“.
- **Bayéská metoda (*Bayesian analysis*)** – obdoba maximum likelihood

Zpracování mol. genet. dat – analýza sekvencí

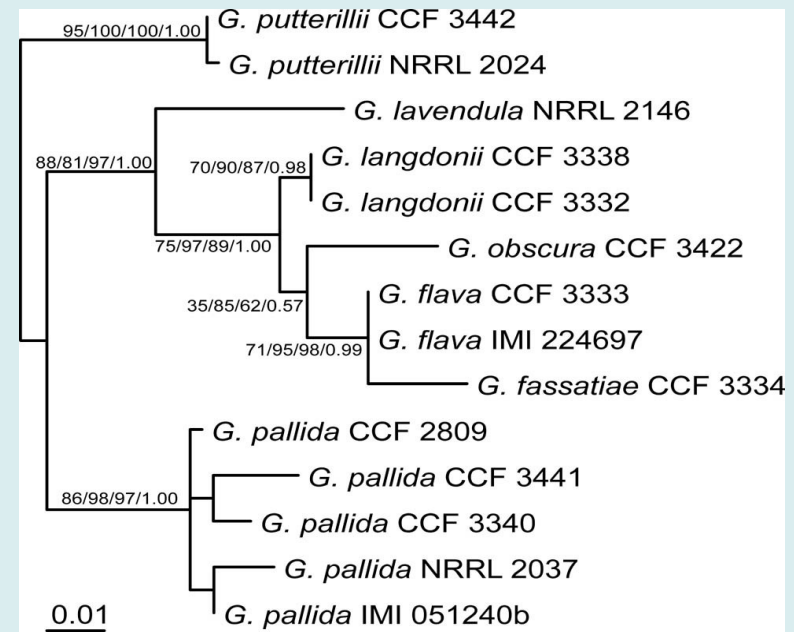
Bootstrap

”důvěryhodnost” získaného fylogramu

- Při výpočtu bootstrapu, provádějí fylogenetické programy tzv. *resampling* ze všech získaných sekvencí se náhodně vybírá velký počet podmnožin (400 - 1000 ×) a pro každou podmnožinu se počítá fylogenetický strom
- Bootstrap o hodnotě 95 udává, že ze všech sestavených stromů se určitý taxon nachází z 95% ve stejném monofyletické skupině

posterior probability

- podobně jako bootstrap
- výsledek Bayesian analysis



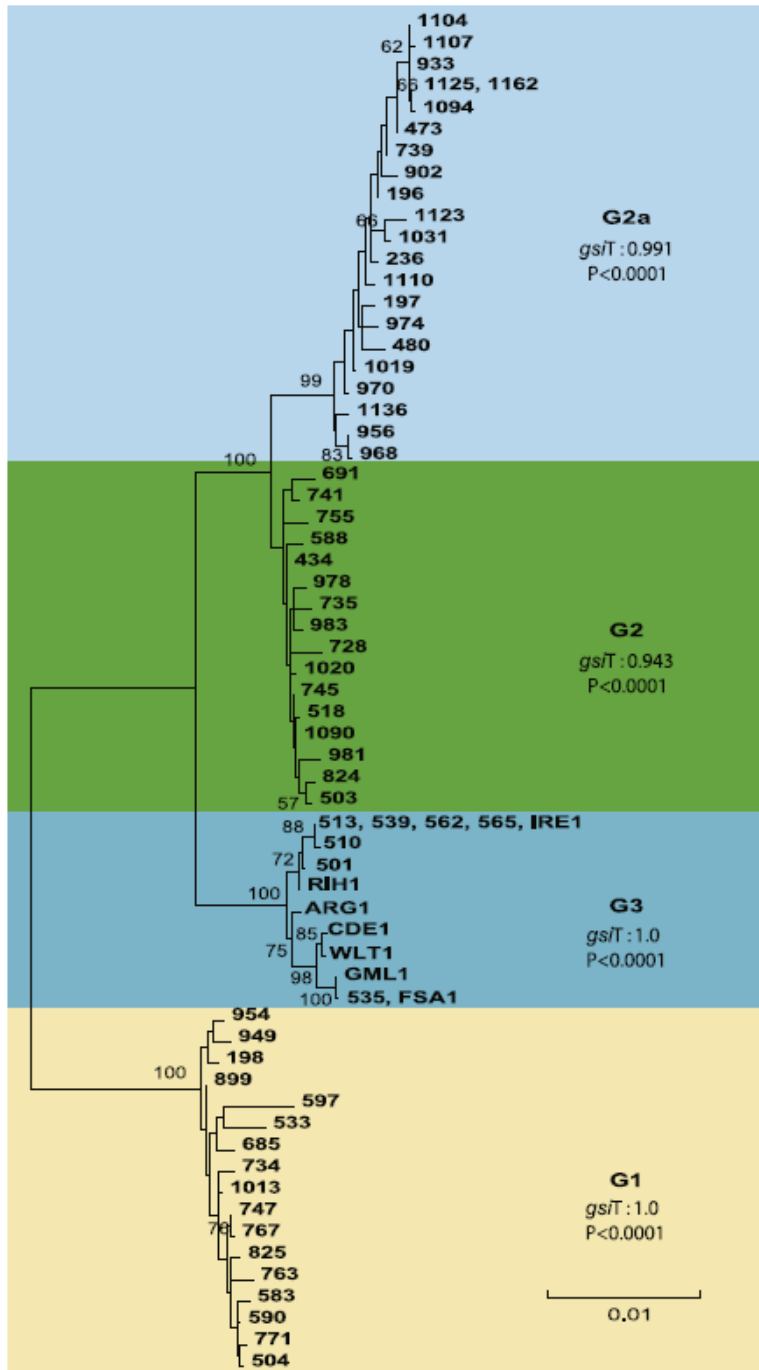


Table 6 – Divergence time estimates from *BEAST analysis.

Species/populations	Time (Ma)	95 % HPD
G1 vs G3, G2, G2a	7.83	4.36–12.4
G3 vs G2, G2a	2.9	1.24–5.14
G2 vs G2a	1.17	0.52–2.07

HPD, highest posterior density interval

- Mladé druhy nelze rozlišit klasickými fylogenetickými metodami
- Aplikace populačně genetických metod

Fig 1 – Maximum likelihood phylogenetic tree of the multigene dataset (*Mcm7*, *RPB2*, *EF-1 α* , *benA*) constructed in Garli 2.0.

Populační genetika u hub

- Jedinci s úspěšnějšími fenotypy jsou zvýhodněni přírodním výběrem či jsou náhodně fixováni
- Poměrné změny v četnostech, ve kterých přispívají jednotlivé fenotypy do další generace, vedou ke změnám v genových (= alelových) frekvencích
- Neustálé mezigenerační změny v genových (genotypových) frekvencích jsou základem evolučních procesů

Populační úroveň

Nastolení genetické izolace (geograficky, geneticky)

- Na populační úrovni se tedy omezení *gene flow* projeví změnou v alelických **frekvencích**. Dojde-li zde ke vzniku biologického druhu, tak se důsledná izolace projeví většinou fixací určité alely pro každý druh
- Každý druh má pak například specifickou ITS sekvenci nebo tvoří vlastní kladus po analýze PCR fingerprintingem
- mezi druhy porovnáváme sekvence alel, ne jejich frekvenci
- fixace určité alely nemusí být důsledná - ancestrální polymorfismus

Druhovú úroveň

- příbuznost velmi podobných sekvencí je zobrazena pomocí haplotypových sítí (*haplotype network*)
- každý úsečka zobrazuje jednu mutační změnu
- koláče zobrazují haplotypy (malé černé tečky značí nenalezené haplotypy)

POPULATION GENETICS OF *XANTHORIA PARIETINA* 1551

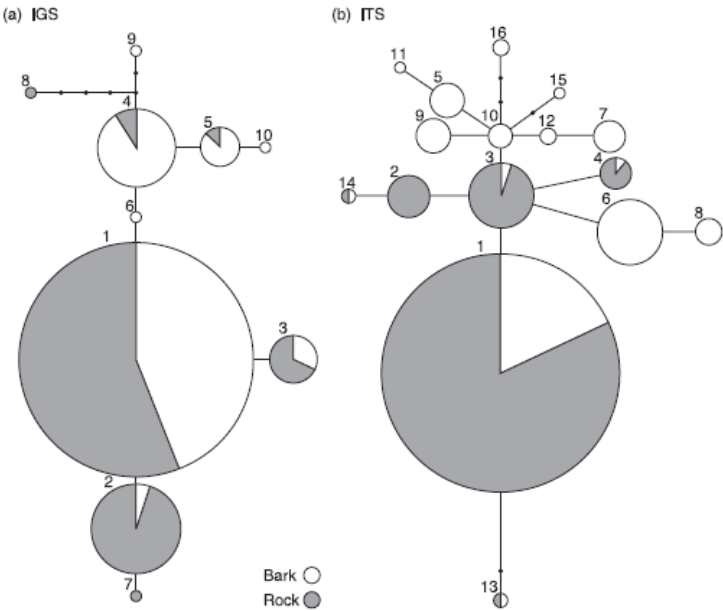


Fig. 2 Haplotype network for *Xanthoria parietina* under the 95% parsimony criterion. The circle area of each haplotype is proportional to the frequency of this haplotype in the total sample. For exact haplotype frequencies and GenBank Accession nos, see Table 3. Grey areas in the circles mark the proportion of individuals growing on rock, white areas on bark. Small filled circles represent haplotypes not present in the sample. One mutational step between haplotypes is represented by a line. (a) Network based on IGS sequences. (b) Network based on ITS sequences.

Table 4 Hierarchical AMOVA of seven *Xanthoria parietina* populations on either bark or rock. Variation was partitioned (a) among populations without habitat grouping; (b) among substrates, viz bark vs. rock grouping

Source of variation	d.f.		Variance component		% Variation		P		Fixation indices	
	IGS	ITS	IGS	ITS	IGS	ITS	IGS	ITS	IGS	ITS
(a) Among populations	6	6	0.15	0.19	19.87	20.34	< 0.0001	< 0.0001		
With Molecular Ecology (2006) 15, 1545–1559			0.59	0.74	80.13	79.66	< 0.0001	< 0.0001	$\Phi_{ST} = 0.199$	0.203
(b) Among groups	1	1	0.21	0.28	25.27	27.30	0.054	0.037	$\Phi_{CT} = 0.253$	0.273
Among populations within groups	5	5	0.03	0.03	3.08	2.90	< 0.0001	< 0.0001	$\Phi_{SC} = 0.041$	0.040
Within populations	206	211	0.59	0.73	71.66	69.81	< 0.0001	< 0.0001	$\Phi_{ST} = 0.283$	0.302

- **Distační metody** – PAUP, MEGA,
- **Parsimonie** – PAUP, MEGA, TNT
- **Maximum likelihood** – PAUP, RAxML, Phyml, IQ-Tree, TreePuzzle, GARLI
- **Bayéská metoda** – MrBayes
Phylobayes

Table 1 | **Functionalities of a few commonly used phylogenetic programs**

Name	Brief description	Link	Refs
Bayesian evolutionary analysis sampling trees (BEAST)	A Bayesian MCMC program for inferring rooted trees under the clock or relaxed-clock models. It can be used to analyse nucleotide and amino acid sequences, as well as morphological data. A suite of programs, such as Tracer and FigTree, are also provided to diagnose, summarize and visualize results	http://beast.bio.ed.ac.uk	135
Genetic algorithm for rapid likelihood inference (GARLI)	A program that uses genetic algorithms to search for maximum likelihood trees. It includes the GTR + Γ model and special cases and can analyse nucleotide, amino acid and codon sequences. A parallel version is also available	http://code.google.com/p/garli	55
Hypothesis testing using phylogenies (HYPHY)	A maximum likelihood program for fitting models of molecular evolution. It implements a high-level language that the user can use to specify models and to set up likelihood ratio tests	http://www.hyphy.org	136
Molecular evolutionary genetic analysis (MEGA)	A Windows-based program with a full graphical user interface that can be run under Mac OSX or Linux using Windows emulators. It includes distance, parsimony and likelihood methods of phylogeny reconstruction, although its strength lies in the distance methods. It incorporates the alignment program ClustalW and can retrieve data from GenBank	http://www.megasoftware.net	37
MrBayes	A Bayesian MCMC program for phylogenetic inference. It includes all of the models of nucleotide, amino acid and codon substitution developed for likelihood analysis	http://mrbayes.net	71
Phylogenetic analysis by maximum likelihood (PAML)	A collection of programs for estimating parameters and testing hypotheses using likelihood. It is mostly used for tests of positive selection, ancestral reconstruction and molecular clock dating. It is not appropriate for tree searches	http://abacus.gene.ucl.ac.uk/software	137
Phylogenetic analysis using parsimony* and other methods (PAUP* 4.0)	PAUP* 4.0 is still a beta version (at the time of writing). It implements parsimony, distance and likelihood methods of phylogeny reconstruction	http://www.sinauer.com/detail.php?id=8060	
PHYLIP	A package of programs for phylogenetic inference by distance, parsimony and likelihood methods	http://evolution.gs.washington.edu/phylip.html	
PhyML	A fast program for searching for the maximum likelihood trees using nucleotide or protein sequence data	http://www.atgc-montpellier.fr/phyml/binaries.php	53
RAxML	A fast program for searching for the maximum likelihood trees under the GTR model using nucleotide or amino acid sequences. The parallel versions are particularly powerful	http://scoih-its.org/exelixis/software.html	54
Tree analysis using new technology (TNT)	A fast parsimony program intended for very large data sets	http://www.zmuc.dk/public/phylogeny/TNT	42

Note: all programs can run on Windows, Mac OSX and Unix or Linux platforms. Except for PAUP*, which charges a nominal fee, all packages are free for download. See Felsenstein's comprehensive list of programs at <http://evolution.genetics.washington.edu/phylip/software.html>. GTR, general time reversible; MCMC, Markov chain Monte Carlo.

Table 2 | **A summary of strengths and weaknesses of different tree reconstruction methods**

Strengths	Weaknesses
<i>Parsimony methods</i>	
• Simplicity and intuitive appeal • The only framework appropriate for some data (such as SINES and LINES)	• Assumptions are implicit and poorly understood • Lack of a model makes it nearly impossible to incorporate our knowledge of sequence evolution • Branch lengths are substantially underestimated when substitution rates are high • Maximum parsimony may suffer from long-branch attraction
<i>Distance methods</i>	
• Fast computational speed • Can be applied to any type of data as long as a genetic distance can be defined • Models for distance calculation can be chosen to fit data	• Most distance methods, such as neighbour joining, do not consider variances of distance estimates • Distance calculation is problematic when sequences are divergent and involve many alignment gaps • Negative branch lengths are not meaningful
<i>Likelihood methods</i>	
• Can use complex substitution models to approach biological reality • Powerful framework for estimating parameters and testing hypotheses	• Maximum likelihood iteration involves heavy computation • The topology is not a parameter so that it is difficult to apply maximum likelihood theory for its estimation. Bootstrap proportions are hard to interpret
<i>Bayesian methods</i>	
• Can use realistic substitution models, as in maximum likelihood • Prior probability allows the incorporation of information or expert knowledge • Posterior probabilities for trees and clades have easy interpretations	• Markov chain Monte Carlo (MCMC) involves heavy computation • In large data sets, MCMC convergence and mixing problems can be hard to identify or rectify • Uninformative prior probabilities may be difficult to specify. Multidimensional priors may have undue influence on the posterior without the investigator's knowledge • Posterior probabilities often appear too high • Model selection involves challenging computation^{138,139}

"A la Carte" Mode

Alignment
MUSCLE



Curation
Gblocks



Phylogeny
PhyML



Tree Rendering
TreeDyn

1. Workflow Setup

Workflow Settings

Name of the analysis (optional):

Choose processing steps to run and select software to use:

☒ Multiple Alignment:

☒ MUSCLE

☐ ProbCons

☐ T-Coffee

☐ 3D-Coffee

☐ ClustalW

☒ Alignment curation:

☒ Gblocks

☐ Remove positions with gaps

☒ Construction of phylogenetic tree:

Maximum Likelihood

☒ PhyML

Parsimony

☐ TNT

Distances

☐ ProtDist/FastDist + BioNJ

☐ ProtDist/FastDist + Neighbor

Bayesian inference

☐ MrBayes (limit: 30 sequences)

☒ Visualisation of phylogenetic tree:

☒ TreeDyn

- online aplikace nabízející propojení různých aplikací

http://www.phylogeny.fr/version2_cgi/alacart

Citation for Study 13673

- 1 Citation title: "Taxonomic revision of *Eurotium* and transfer of species to *Aspergillus*".
- 1 Study name: "Taxonomic revision of *Eurotium* and transfer of species to *Aspergillus*".
- 1 This study is part of submission 13673 (Status: In Progress).

Citation

Hubka V., Kolarik M., Kubatova A., & Peterson S.W. 2013. Taxonomic revision of *Eurotium* and transfer of sp

Authors

- Hubka V. (submitter)  hubka@biomed.cas.cz  +420 739 663 218
- Kolarik M.  mkolarik@biomed.cas.cz
- Kubatova A.  alena.kubatova@natur.cuni.cz
- Peterson S.W.  stephen.peterson@ars.usda.gov  309-681-6384

Abstract

Aspergillus section *Aspergillus* contains economically important, xerophilic fungi that are widely distributed in food, soil, and indoor environments. This species seems to be relatively common and has been often confused with *A. glaucus*. In addition, green ornamentation, and the ability to grow at specific combinations of temperature and water activity, is provided designated for some of the revised taxa.

Keywords

food spoilage, low water activity, multilocus sequence typing, scanning electron microscopy, stored grain, xe

- **Veškeré vstupní datasety (alignments, stromy) je možno uložit v databázy**
- **Někdy podmínkou publikace (Mycologia)**
- **TREEBASE: <http://www.treebase.org>**

Staré slajdy mimo přednášku

allopatrie

peripatrie

parapatrie

sympatrie

Původní populace



První krok speciace



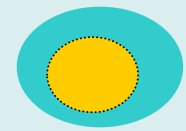
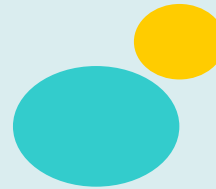
Geogr. bariera

Osídlení nového
habitatu

Osídlení nového
habitatu

Genetická mutace

Evoluce reprodukční
izolace



V geogr. izolaci

V izolovaných
habitatch

Osídlení nového
habitatu

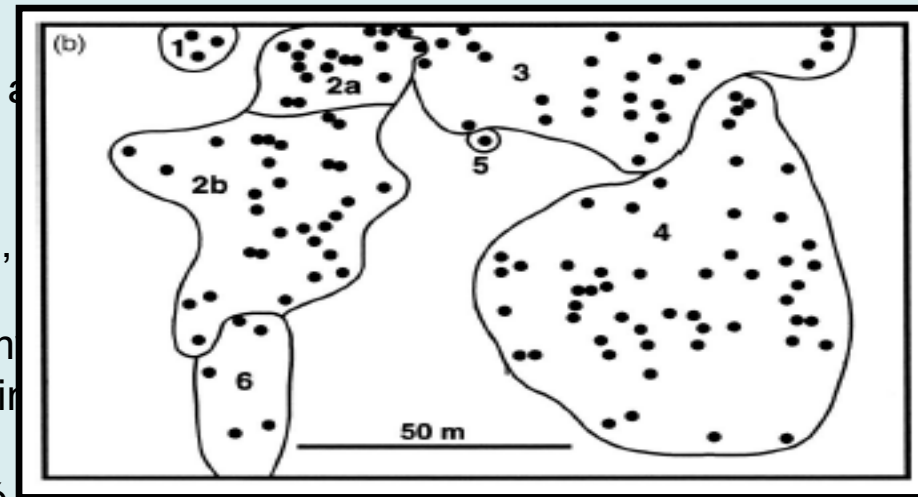
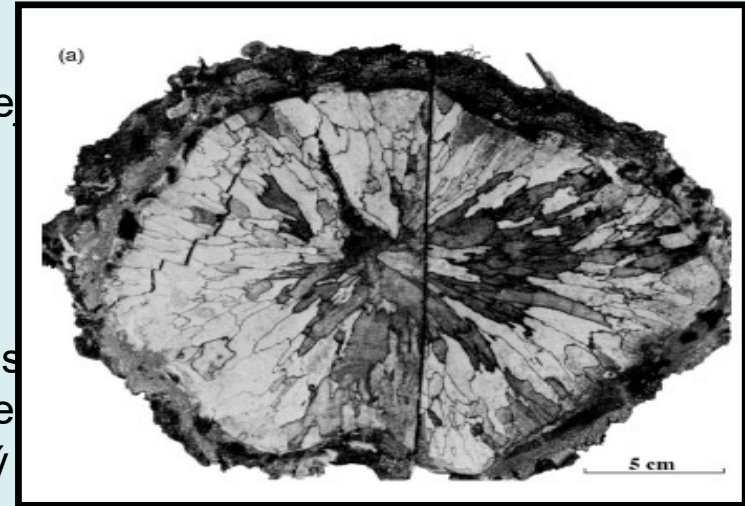
V původní populaci

Nové biologické
druhy po
splynutí areálu



Vymezení jedince u hub, sampling

- Geneticky unikátní jedinec se nazývá **geneta**
- Prakticky ho detekovat jako soubor klonů (ramet) se stěží
- to bývá označováno jako haplotyp
- Geneta - každý produkt jednoho meiotického dělení
 - u Zygomycetů a Ascomycetů jednak jedince, co vyrostlo z jedné bazidiospory, ale i následné mitotické deriváty tohoto jedince (klony této genety)
 - u Bazidiomycetů je jednou genetou i dikaryon vzniklý z jedné bazidiospory a všichni mitotiční potomci
- U hub je jedna geneta prostorově neomezena a může být ve fyzickém kontaktu.
- *Armillaria gallica*, jedna geneta na ploše 15 ha,
- Více plodnic může patřit jednomu mycelium (množství klonů), větvi může růst několik geneticky odlišných jedinců
- Sampling - zamezit „resamplingu“ klonu jedné genety
- Velký počet genet z každé populace (30 ks -několik set) sebraných v průběhu více sezón



Biochemické markery - ubichinony

- koenzymy Q6-10, přenašeče elektronů v dých. řetězci, liší se izoprenovými jednotkami
- detekce pomocí HPLC se standardy Q6-Q10
- převažující ubichinonový systém (90%) může být stabilní pro druh, rod, řád
- využití v taxonomii kvasinek, pouze jako přídatná informace

Table 2. Distribution of the groups formed using ubiquinone system as chemotaxonomic marker in *Cunninghamella* strains.

Group I CoQ 6 main		Group II CoQ 9 main		Group III Major percentage
A Presence / CoQ 10	B Absence / CoQ 10	A Presence / CoQ 10	B Absence / CoQ 10	CoQ 9
<i>C. bertholletiae</i> IFM46987	<i>C. ramosa</i> URM1918	<i>C. elegans</i> URM3172	<i>C. elegans</i> L14	<i>C. echinulata</i> URM2136
<i>C. bertholletiae</i> IFM46110	<i>C. bertholletiae</i> IFM46114		<i>C. elegans</i> L15	
<i>C. bertholletiae</i> IFM46111	<i>C. bertholletiae</i> L13		<i>C. elegans</i> L16	
<i>C. elegans</i> IFM40505			<i>C. elegans</i> L17	
<i>C. elegans</i> IFM47050			<i>C. elegans</i> URM2084	
<i>C. elegans</i> IFM46109			<i>C. blakesleeana</i> URM168	
<i>C. elegans</i> IFM47591				

Table 4. Fast-profiles (for cultivations on MOS only) of 10 isolates with highest average within-race/type mismatch values. Extremely high or low values are printed in bold

Strain	Type	Fatty acids															Sterols							Average mismatch
		TAE	9-			OH- 9,12-			6,9-			x,x,x- 18:3	OH- 22:0	OH- 25:0	OH- 24:0	Ergosta- trienol	STA	Ergos- tenon	STE	Ergos- terol	Ergosta- dienol	STB		
			15:0	16:1	16:0	17:1	16:0	18:2	C18:1	T18:1	18:0												18:2	
CF-86-0007	NA-LTT	2.8	1.1	0.4	7.5	1.2	2.3	35.7	15.5	1.1	0.5	0.8	0.0	0.9	0.3	4.1	1.2	0.0	5.0	1.2	20.5	0.4	0.0	24
CF-86-0035	NA-STT	1.4	1.6	0.3	9.5	1.3	2.2	41.9	16.8	1.1	0.7	1.0	0.2	1.4	0.1	2.8	1.7	0.0	2.3	1.4	13.6	0.0	0.3	20
CF-87-0014	NA-LTT	2.4	0.7	0.4	9.8	1.0	2.0	35.7	20.5	1.2	0.7	0.9	0.0	0.8	0.4	3.3	1.1	0.0	3.8	0.9	16.6	0.4	0.0	20
18-76	NA-STT	8.8	0.2	1.0	9.8	0.5	0.2	36.9	42.5	1.4	1.2	1.0	0.1	0.1	0.1	0.3	0.0	0.0	1.0	0.1	3.6	0.0	0.0	19
18-74	NA-STT	3.2	0.3	2.0	12.9	0.6	0.4	51.5	25.2	1.7	0.9	0.7	0.6	0.3	1.0	0.3	0.0	0.0	1.4	0.0	0.3	0.0	0.0	17
CF-85-0157	NA-STT	2.6	0.8	0.8	12.3	1.1	1.6	51.8	19.0	1.6	0.7	1.0	0.5	0.6	0.7	1.7	0.0	0.3	2.1	0.3	2.1	0.0	1.0	17
CF-87-0025	NA-STT	2.1	1.6	0.3	10.1	1.2	1.6	44.2	21.0	1.5	0.5	1.5	0.6	0.7	0.0	1.8	0.8	0.1	3.3	1.1	7.3	0.0	0.9	16
CF-88-0008	NA-STT	2.8	0.9	0.9	10.3	1.4	1.4	49.8	21.7	1.6	0.6	1.2	0.7	0.7	1.3	1.0	0.2	0.1	2.3	0.4	2.6	0.0	0.8	16
18-51	NA-STT	7.0	0.3	2.0	13.5	0.5	0.3	35.6	38.6	2.5	1.3	1.1	0.1	0.1	0.1	0.4	0.0	0.0	0.7	0.0	2.9	0.0	0.0	16
CF-85-0107	NA-STT	2.4	0.9	0.7	12.2	1.2	1.2	50.7	20.2	1.2	0.8	0.9	0.6	0.6	0.3	1.3	0.2	0.2	2.3	0.4	2.8	0.1	0.9	15

Diversity of *Gremmeniella abietina*

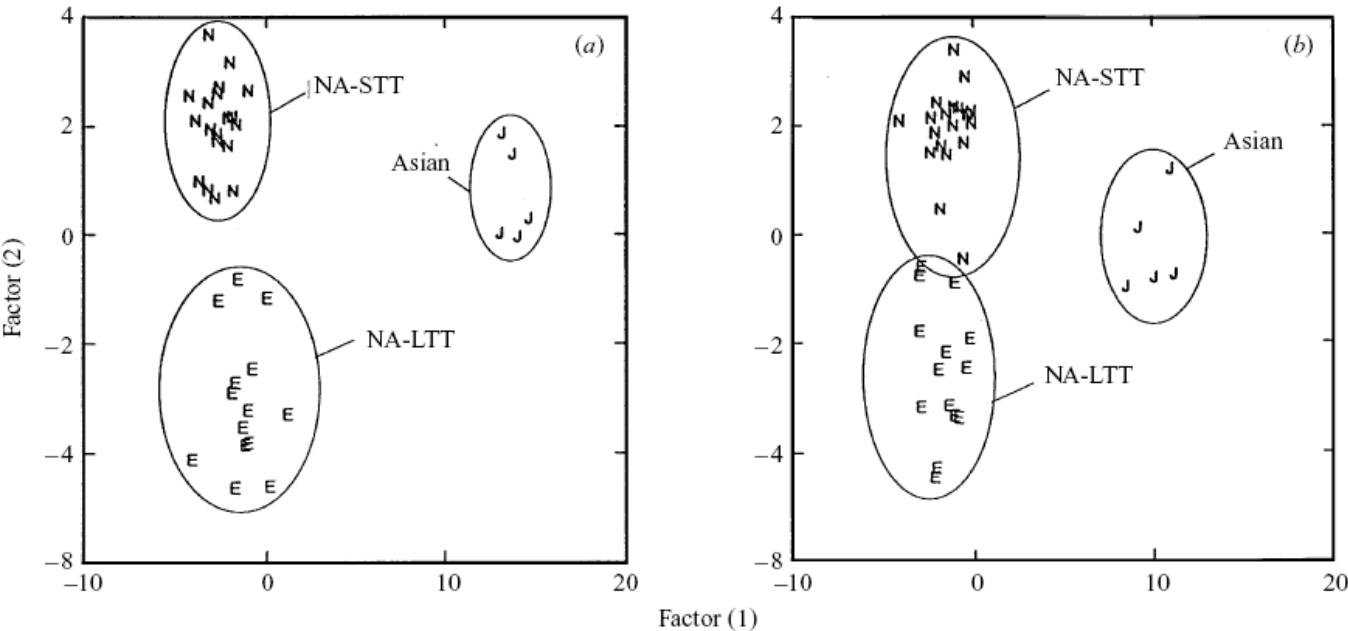


Fig. 1. Separation of *G. abietina* FAST-profiles by a multivariate discriminant analysis into three groups according to their *a priori* grouping: NA-STT, NA-LTT and Asian isolates. The isolates were cultivated on PTS (a) and MOS (b).

Biochemické markery- isozymy

- **isoenzymy (isozymy)** – různé enzymy (v různých lokusech a s jinou strukturou) se stejnou funkcí
- **allozymy** – různé alely jednoho enzymu
- vnitrodruhová i mezidruhová variabilita
- liší v sekvenci a následné struktuře (náboj, mobilita)
- separace na Elfo (detekce enzym. aktivity - obarvený substrát, produkt...)
- esteráza, amyláza, malát dehydrogenáza, peroxidáza, pektináza, phosphatáza

Biochemické markery- isozymy

- při zjištění genetické variability se téměř nepoužívají a byly vytlačeny DNA technikami
 - vliv kultivačních podmínek a ontogenetického stádia
 - malá variabilita a ne všechny změny Elfo zachytí
 - pouze v kódujících sekvencích
 - pouze nesynonymní mutace

Mycologist, Volume 19, Part 4 November 2005

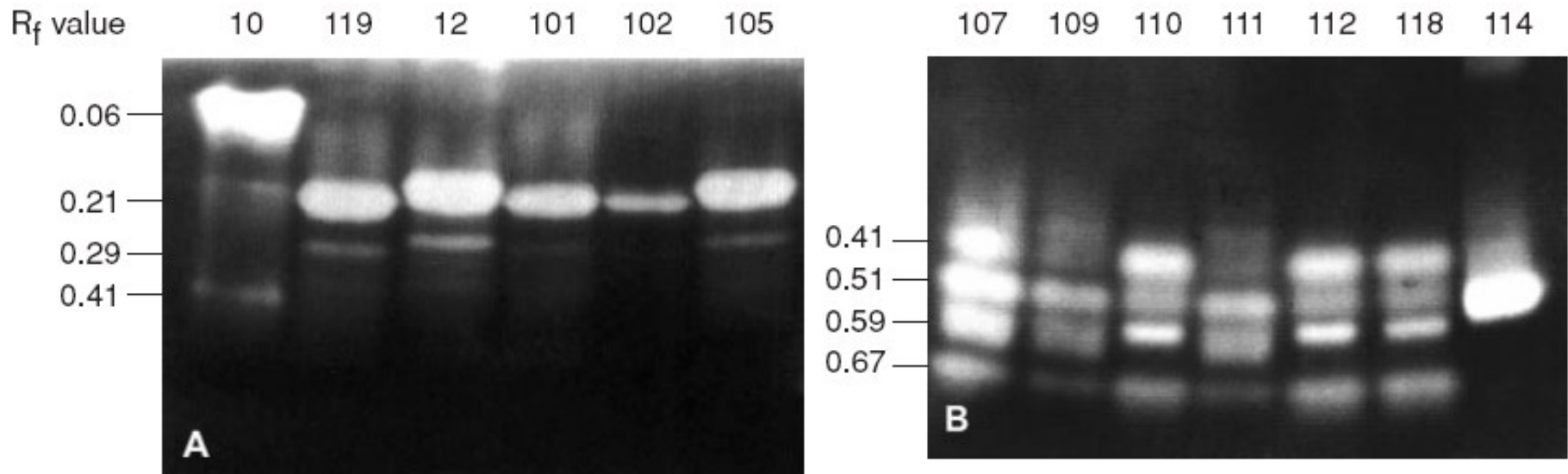


Fig 2 Amylase zymogram of different strains of *Aspergillus niger* (A) and *Aspergillus flavus* (B)

Biochemické markery – mastné kyseliny

- pyrolysis gas chromatography-MS se standardy mastných kyselin
- myristic acid (C14:0), palmitic acid (C16:0), stearic acid (C18:0), oleic acid (C18:1), linolenic acid (C18:3) and γ-linolenic acid (C18:3)
- FAST profiling: mastné kyseliny + steroly, srovnání s databází profilů MK
- druhově specifické profily, vnitrodruhové variabilita
- využití u významných hub

Rev Iberoam Micol 2001; 18: 123-127 **123**

Biochemical markers in taxonomy of the genus *Cunninghamella*

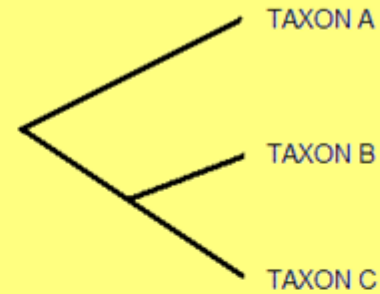
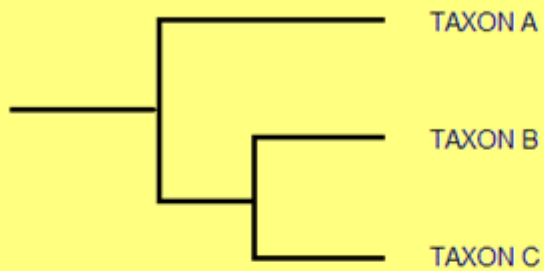
Ricardo Kenji Shiosaki^{1,4,5}, Kaoru Okada^{1,2}, Norma Buarque de Gusmão¹, Pragati Nigam¹, Peter S. Falcão³, Nicácio Henrique da Silva¹, Kazutaka Fukushima¹, Makoto Miyaji¹ & Galba Maria de Campos-Takaki^{1,3,4}

Table 4. Fatty acids chemical composition average percentage in *C. elegans* and *C. bertholletiae*.

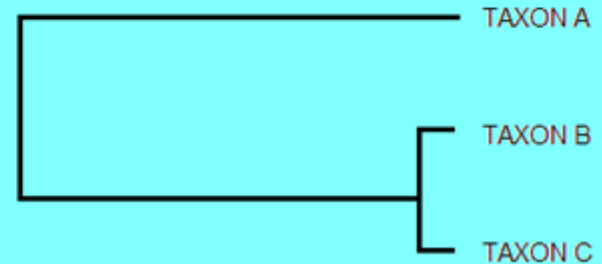
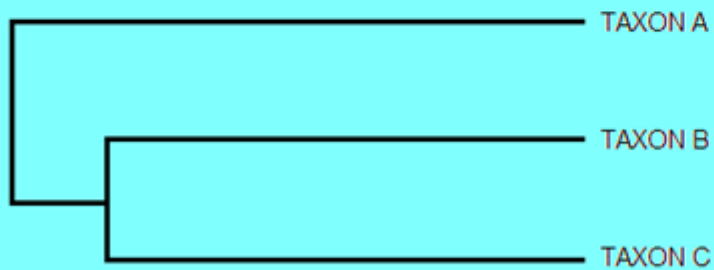
Species	Average percentage (average ± error standard)					
	C14:0	C16:0	C18:0	C18:1	C18:3	C18:3 (γ)
<i>C. elegans</i>	1.37 ± 0.11	23.39 ± 0.84	11.98 ± 1.63	36.63 ± 1.44	13.17 ± 1.55	13.56 ± 1.94
<i>C. bertholletiae</i>	1.63 ± 0.24	24.89 ± 0.88	8.21 ± 0.74	36.47 ± 1.42	18.56 ± 1.05	10.24 ± 1.21

Dendrogramy

Kladogram



Fylogram



termín *clade* použitelný spíše u kladogramu. U fylogenetického stromu je možno použít termín *phylogenetic clade*, *phyloclade* (či jiné alternativy...*lineage*..)

PARYBY RYBY PLAZI SAVCI



=

PARYBY RYBY SAVCI PLAZI

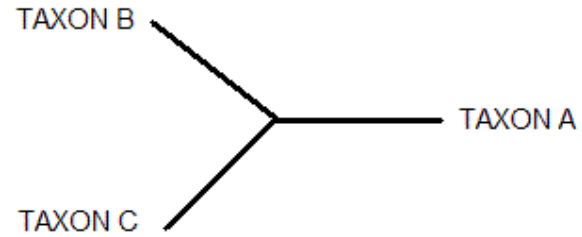
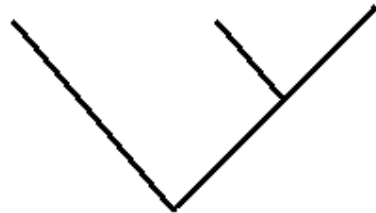


=

PARYBY PLAZI SAVCI RYBY



TAXON A TAXON B TAXON C



Analýza enviromtálných vzorků

- nové linie ascomycot známé pouze z environmentálných vzorků

Schadt, C.W., Martin, A.P., Lipson, D.A., Schmidt, S.K., 2003. Seasonal dynamics of previously unknown fungal lineages in tundra soils. *Science* 301, 1359–1361.

Vandenkoornhuyse, P., Baldauf, S.L., Leyval, C., Straczek, J., Young, J.P.W., 2002. Extensive fungal diversity in plant roots. *Science* 295, 2051.

638

T.M. Porter et al. / *Molecular Phylogenetics and Evolution* 46 (2008) 635–644

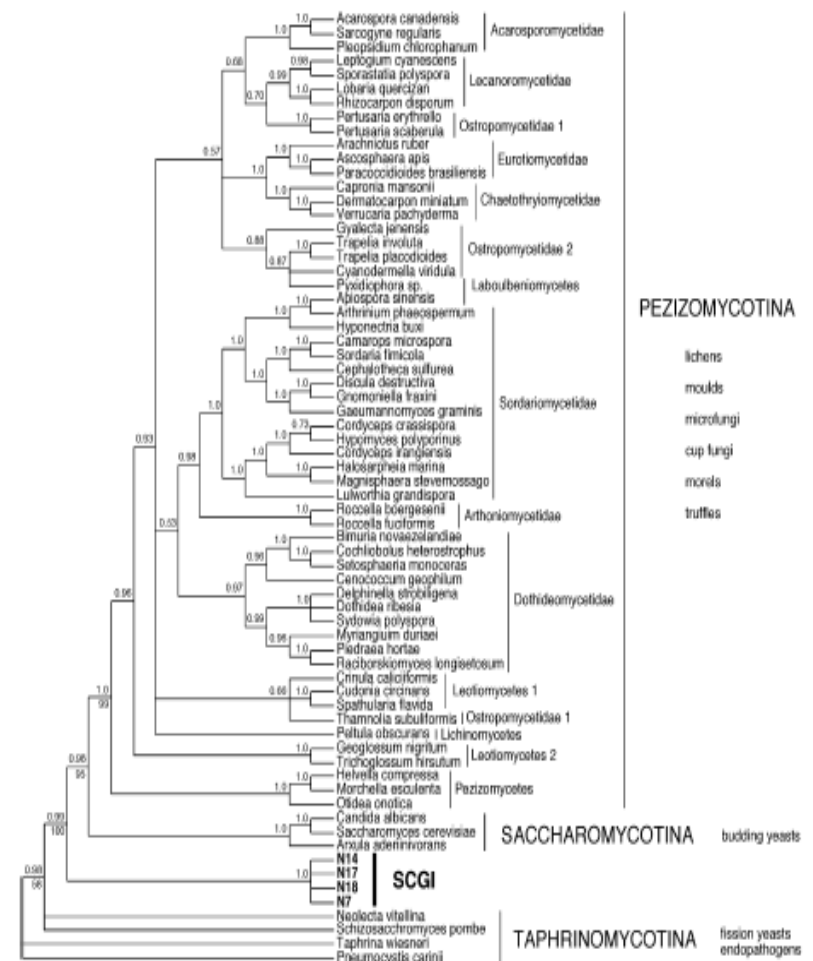
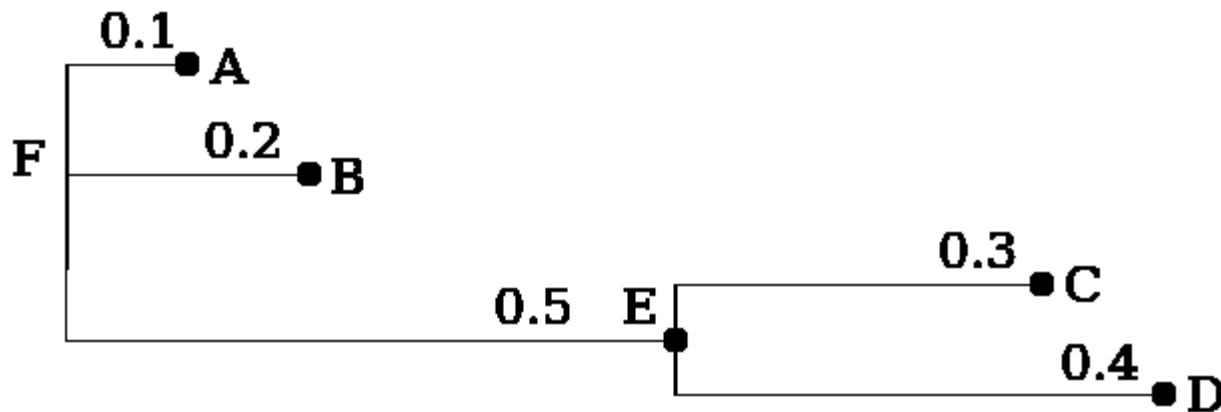


Fig. 1. Bayesian consensus tree for the SSU + LSU combined rDNA analysis. The data set includes 65 taxa and 2,364 SSU and LSU rDNA characters and is rooted with *Pneumocystis carinii*. Class and subclass labels largely follow the convention used by Lutzoni et al. (2004). Bayesian support is shown above the branch and parsimony bootstrap support is shown below branches of interest.

- formáty stromů
 - NEXUS
 - Newick
- editory stromů (FigTree, TreeGraph, MEGA)



```

(,, (,));
(A,B, (C,D));
(A,B, (C,D)E)F;
(:0.1,:0.2, (:0.3,:0.4):0.5);
(:0.1,:0.2, (:0.3,:0.4):0.5):0.0;
(A:0.1,B:0.2, (C:0.3,D:0.4):0.5);
(A:0.1,B:0.2, (C:0.3,D:0.4)E:0.5)F;
((B:0.2, (C:0.3,D:0.4)E:0.5)F:0.1)A;
  
```

no nodes are named
leaf nodes are named
all nodes are named
all but root node have a distance to parent
all have a distance to parent
distances and leaf names (popular)
distances and all names
a tree rooted on a leaf node (rare)

Zpracování mol. genet. dat – analýza sekvencí

- Fylogenetické stromy lze obecně rozdělit na zakořeněné a nezakořeněné
- Zakořeněný fylogenetický strom obsahuje taxon stojící vně sledované skupiny, který s ní sdílí původní znak - *outgroup*, určen dle jiných biologických dat. Nebo jako nejvzdálenější taxon v nezakořeněném stromu.
- Zakořeněné stromy vyjadřují historii jednotlivých taxonů, tedy jejich postupné odštěpování od nejpůvodnějšího taxonu (**topologie**)

Rozpory v pojetí fylogenetického a biologického druhu

Neurospora crassa x *Neurospora intermedia*

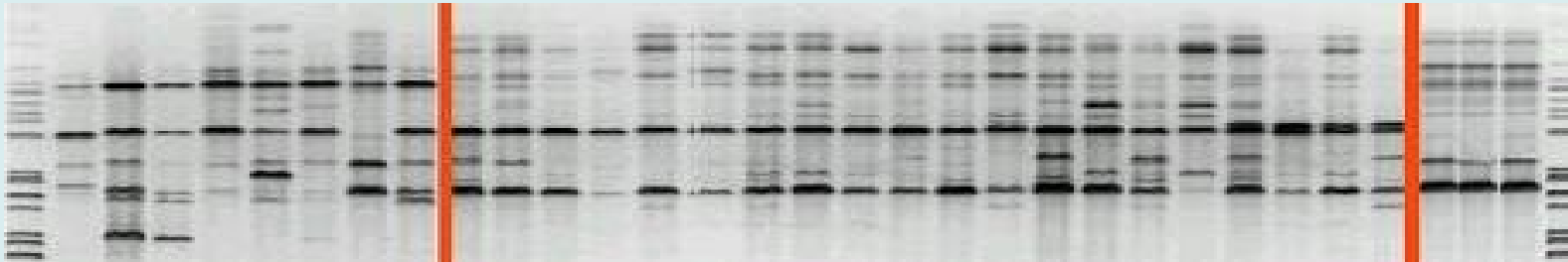
Dva jasné fylogenetické druhy

- Vnitrodruhová variabilita stírá morfologické rozdíly (nelze je rozeznat)
- společný habitat
- Některé izoláty schopny mezidruhového křížení (1-10% fertilních askospor)

Zpracování mol. genet. dat

Problémy při odčítání distančních dat

- nepravá homologie
- náhodné rozhození vzorků
- falešná absence znaku...



Rozpory v pojetí fylogenetického a biologického druhu

Ophiostoma ulmi x *O. novo-ulmi*

Dva morfologické, fylogenetické i biologické druhy

Ophiostoma ulmi: Evropská epidemie 1900

Ophiostoma novo-ulmi: ssp. *novo-ulmi* (Euroasiiská rasa)
pandemie v Evropě 1940- 50

: ssp. *Americana* (Severoamerická rasa)
1940 USA 1960-1980 Evropa

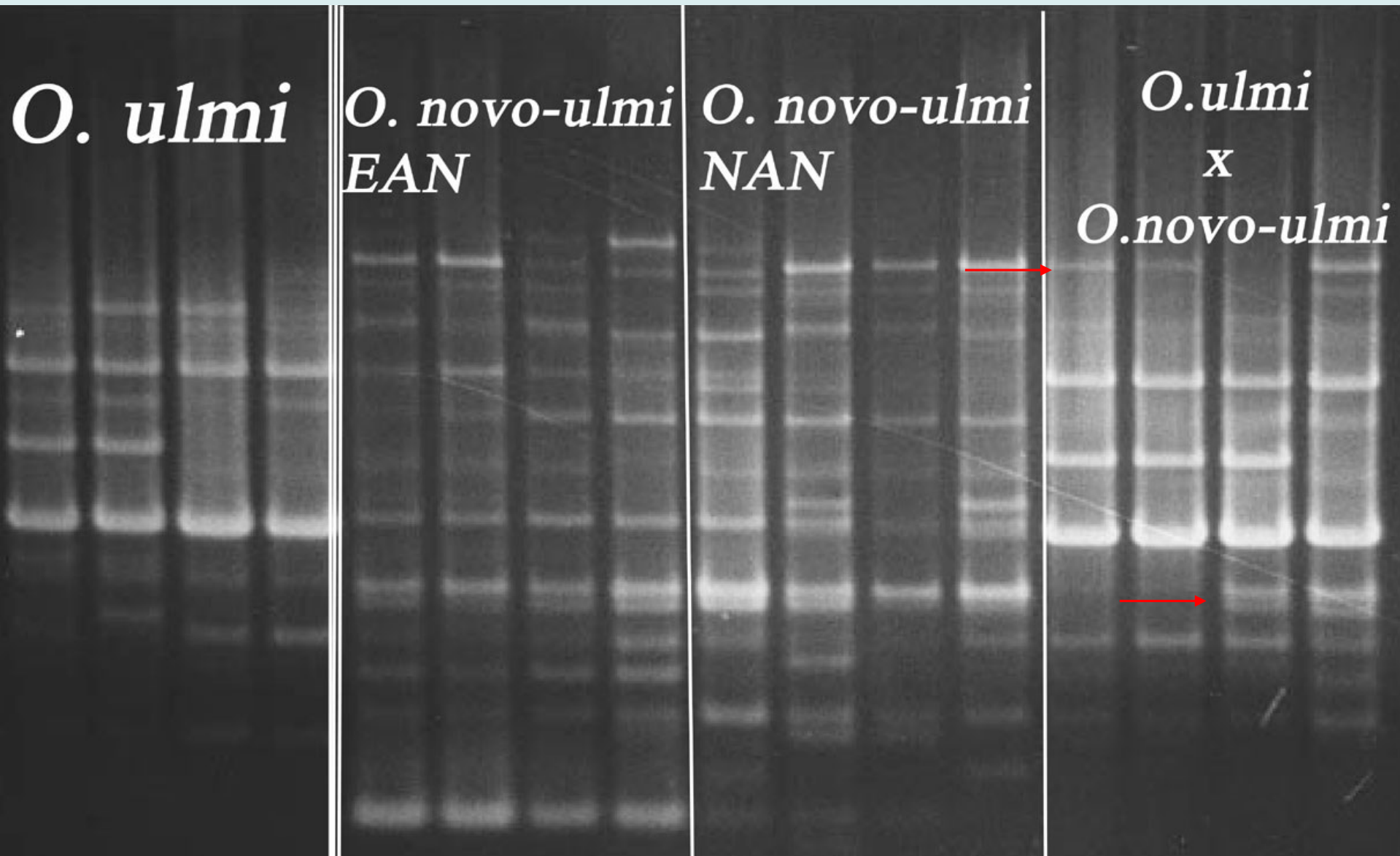
Od roku 1973 zkoumáno 11 000 izolátů...nelezeno 9 hybridů
(malé fitness)

O. ulmi

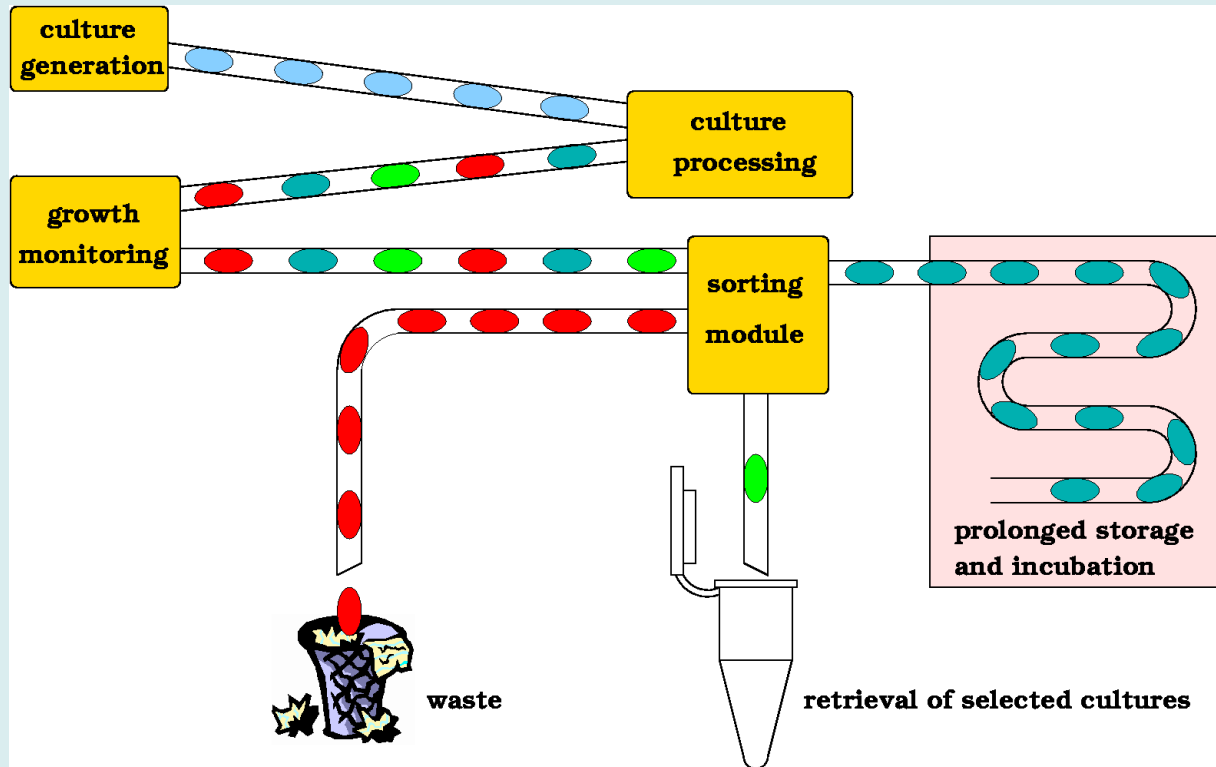
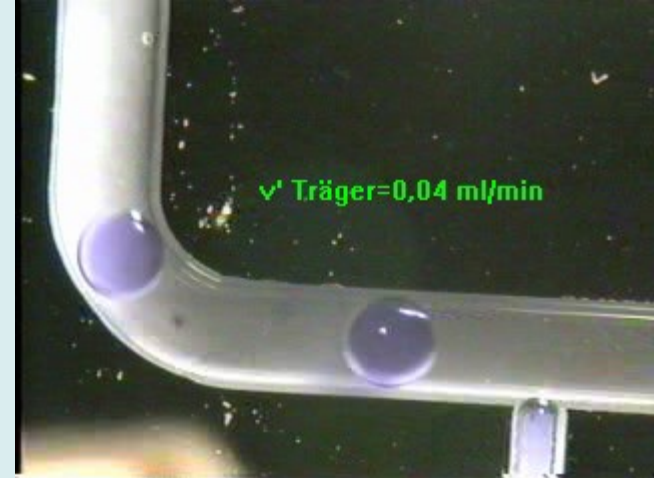
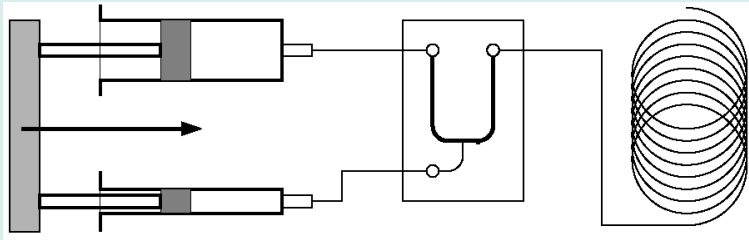
O. novo-ulmi
EAN

O. novo-ulmi
NAN

O. ulmi
x
O. novo-ulmi



- automatická selekce kmenů označených DNA značkou



Mol. Tax. metody – G+C poměr, DNA-DNA reasociace

% G+C (G+C contents, G+C value)

- porovnání organismů na základě relativního množství G+C
- pomocí střední teploty tání homoduplexu T_m . (*inflection point* - teplota, kdy polovina dsDNA je denaturována)
- T_m se zvětšuje se stoupajícím množstvím G+C
- množství ssDNA stanoveno např. spektrofotometricky
- udává se v mol% nebo v %
- částečně nahrazeno celogenomickými daty

Mol. Tax. metody – G+C poměr, DNA-DNA reasociace

DNA-DNA hybridizace (DNA-DNA- hybridization, DNA-DNA reassociation)

- při DNA-DNA hybridizaci se nechávají spojovat ss DNA od dvou organismů v nový heteroduplex, u nějž se zjišťuje „kvalita připojení“, tedy míra homologie studovaných DNA
- obě ssDNA jsou zbaveny velmi variabilních repetitivních sekvencí – ty jsou odstraněni denaturací a rychlou renaturací
- zjišťuje se např. porovnáním T_m jednotlivých homoduplexů a následně jejich heteroduplexu. T_m vzniklého heteroduplexu je přímo úměrná jeho homologii
- Udává se v %

Mol. Tax. metody – G+C poměr, DNA-DNA reasociace

Využití

- multilokusová metoda, dává informace z velké části genomu, k identifikaci a studium příbuznosti na úrovni druhu
- od nástupu PCR postupně mizí. Určitou dobu se používala v taxonomii kvasinek a heterobasidiomycetů. Dnes standardní metoda u bakterií.
- Nevýhody:
 - velká technická náročnost
 - DNA řetězce z jednotlivých vzorků musí porovnávat každý s každým
- Problematická interpretace naměřené homologie, tedy rozhodnutí o druhové hranici – to bylo většinou stanoveno arbitrárně. Např. DNA-DNA reasociační homologie 84% naměřená u kmenů heterobasidiomycetu *Cuniculitrema polymorpha* (teleomorfa) a *Sterigmatosporidium polymorphum* (anamorfa), je dostatečná k prohlášení, že jde o konspecifické druhy

Mol. Tax. metody – SCCP, DGGE

SSCP (*Single Strand Conformation Polymorphism*) a **DGGE** (*Denaturing Gradient Gel Electrophoresis*)

- možnost odlišit fragmenty které se neliší jen velikostí, ale i sekvencí
- SSCP
 - Dvoušroubovice se denaturuje a jednotlivé řetězce (ssDNA) se dělí na polyakrylamidovém gelu
 - ssDNA se sbalí do sekundární struktury - tvar (a tedy i mobilita) se liší podle sekvence

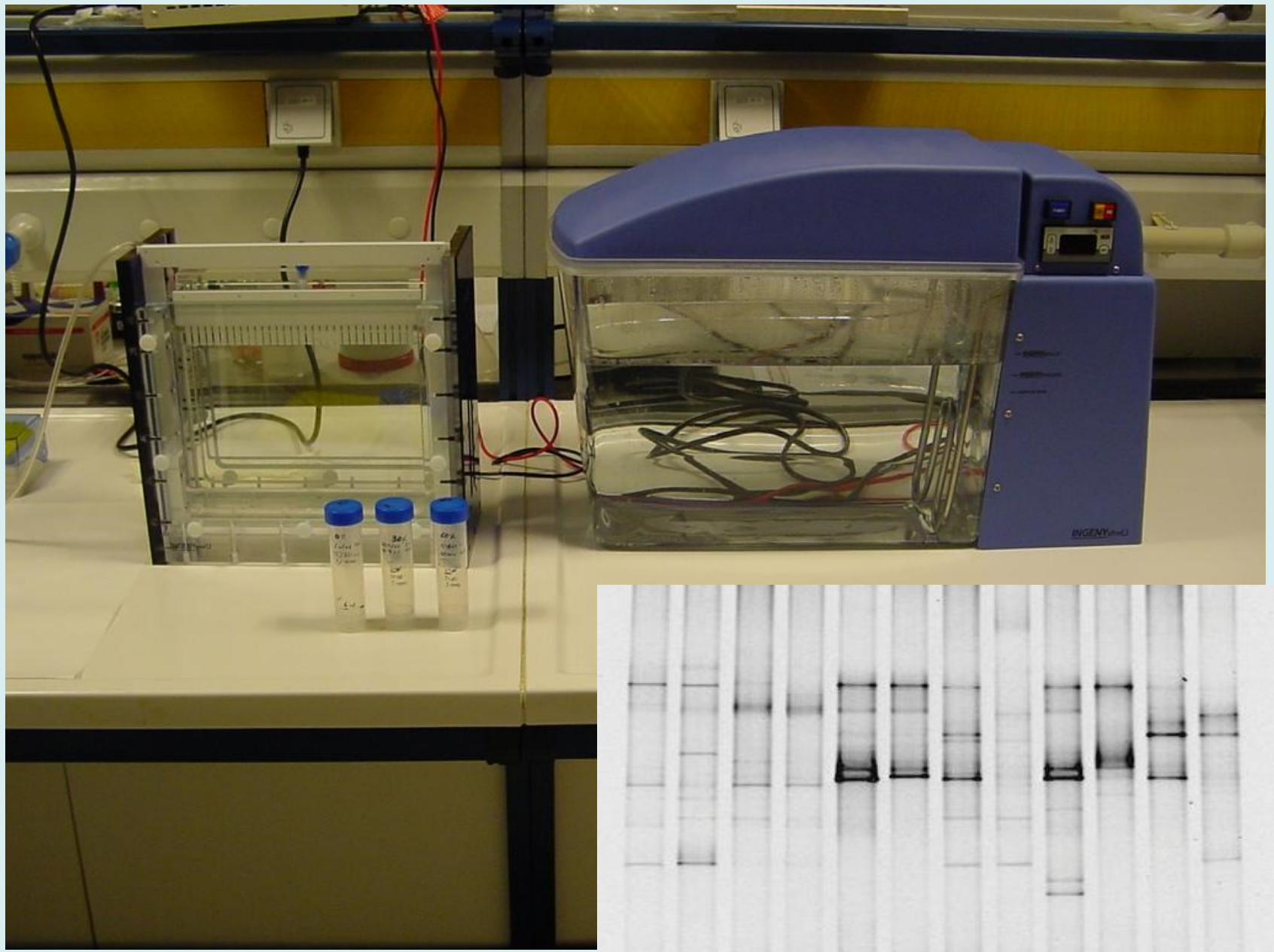
Mol. Tax. metody - SSCP a DGGE

•DGGE

- dsDNA putuje gelem v prostředí se stoupajícím gradientem denaturující látky (močovina, formaldehyd) nebo denaturační teploty (verze zvaná TGGE – *temperature gradient gel electrophoresis*)
- jakmile dsDNA doputuje do bodu kde zdenaturuje (což je ovlivněno právě sekvencí), tak se změní její mobilita

Použití DGGE a SSCP

- Možnost odlišit jednotlivé alely
- Využití v populačně genetických studiích (odhalit polymorfní alelu bez sekvenování)
- Vybrané geny (často z rDNA operonu) se z velkého počtu jedinců se amplifikují pomocí PCR. Metoda SSCP nebo DGGE rozdělí PCR produkty do skupin dle stejných sekvencí a sekvenovat stačí jen zástupce těchto skupin.
- Analýza enviromentálních vzorků
- z větší části nahrazeno metodami NGS



DGGE

SCAR markery (Sequence-characterised amplified region)

- vybereme amplifikační produkt
- zjistíme jeho sekvenci a primery na jeho amplifikaci (zaklonováním do vektoru)

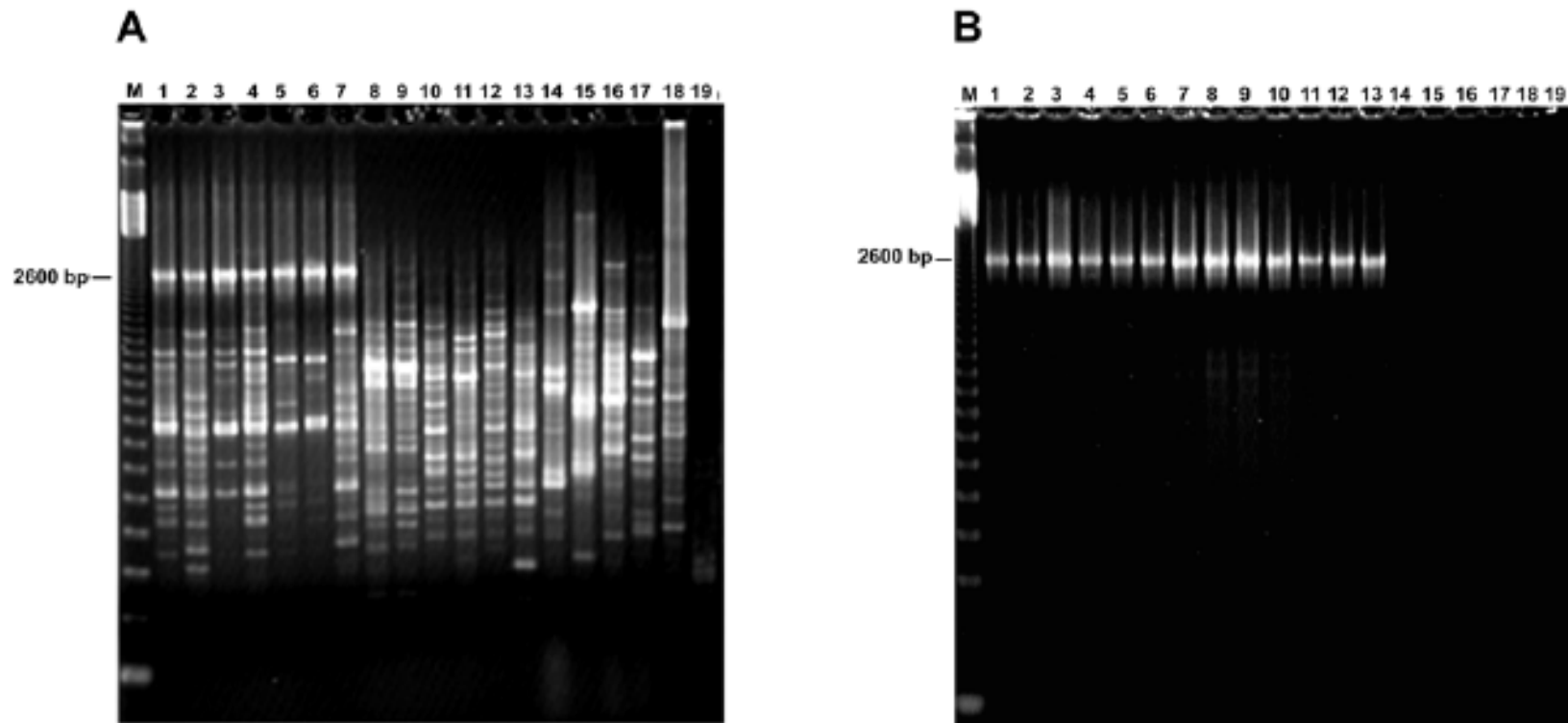


Fig. 1. A, Agarose gel showing amplification products from polymerase chain reaction (PCR) of genomic DNA using the random amplified polymorphic DNA (RAPD) decamer primer OpC15. Marker indicates the 2.6-kb fragment of *Phoma sclerotioides* sequenced to design SCAR primers PSC15₂₆₀₀. B, Amplification products from PCR of genomic DNA of *P. sclerotioides* using the sequence-amplified characterized region (SCAR) primer pair OpC15₂₆₀₀. Lanes 1 to 7: isolates of *P. sclerotioides* from Farson, Lyman, Eden, Boulder, Saratoga, and Laramie regions in Wyoming; and Canada (ATCC), respectively. Lanes 8 and 9: *Phoma medicaginis*, lanes 10 to 13: *Phoma betae*, lane 14: *Aphanomyces euteiches*, lane 15: *Rhizoctonia solani*, lane 16: *Fusarium oxysporum*, lane 17: *Pythium ultimum*, lane 18: *Phytophthora infestans*, lane 19: water control.

A Rapid Method Using PCR-Based SCAR Markers for the Detection and Identification of *Phoma sclerotioides*: The Cause of Brown Root Rot Disease of Alfalfa

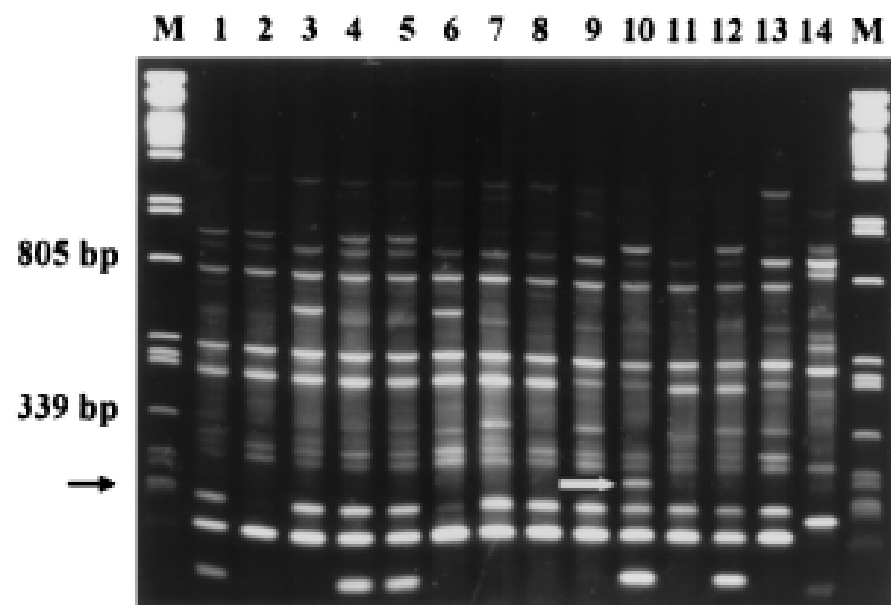


FIG. 1. UP-PCR banding profiles for *C. rosea* strains generated with the AS15inv-AA2M2 UP primer combination. Lanes 1 to 14, strains GR3, IK726, GR6, GR10, GR12, GR11, GR9, GR8, GR4, GR5, GR1, GR13, GR7, and GR2, respectively. Lanes M, molecular size markers (λ phage DNA digested with *Pst*I). Arrows shows the marker of interest for strain GR5.

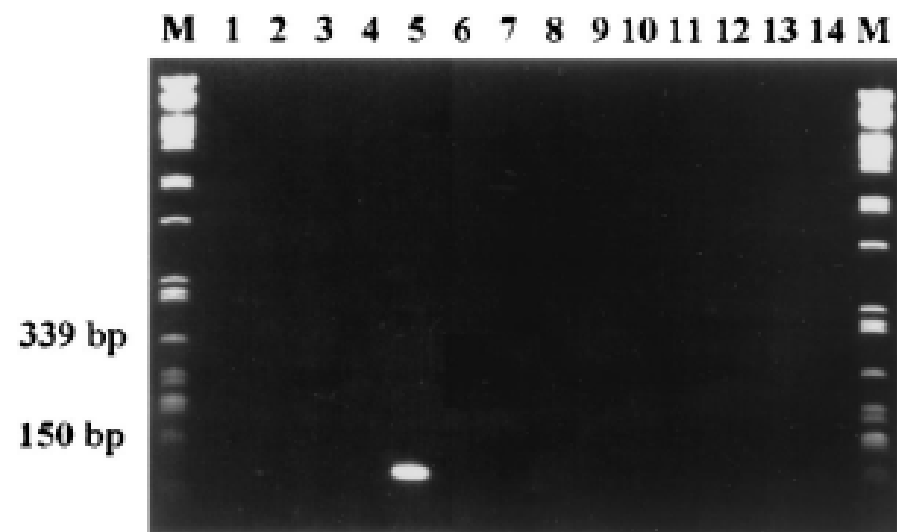


FIG. 3. Testing of *C. rosea* strains for the AS15inv-AA2M2 marker using the SCAR primer set. Lanes 1 to 14, strains IK726, GR12, GR10, GR13, GR5, GR3, GR2, GR4, GR7, GR6, GR11, GR9, GR8, and GR1, respectively. Lanes M, molecular size markers (λ phage DNA digested with *Pst*I).

DISCUSSION

APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Nov. 2000, p. 4758–4763

0099-2240/00/\$04.00+0

Copyright © 2000, American Society for Microbiology. All Rights Reserved.

Vol. 66, No. 11

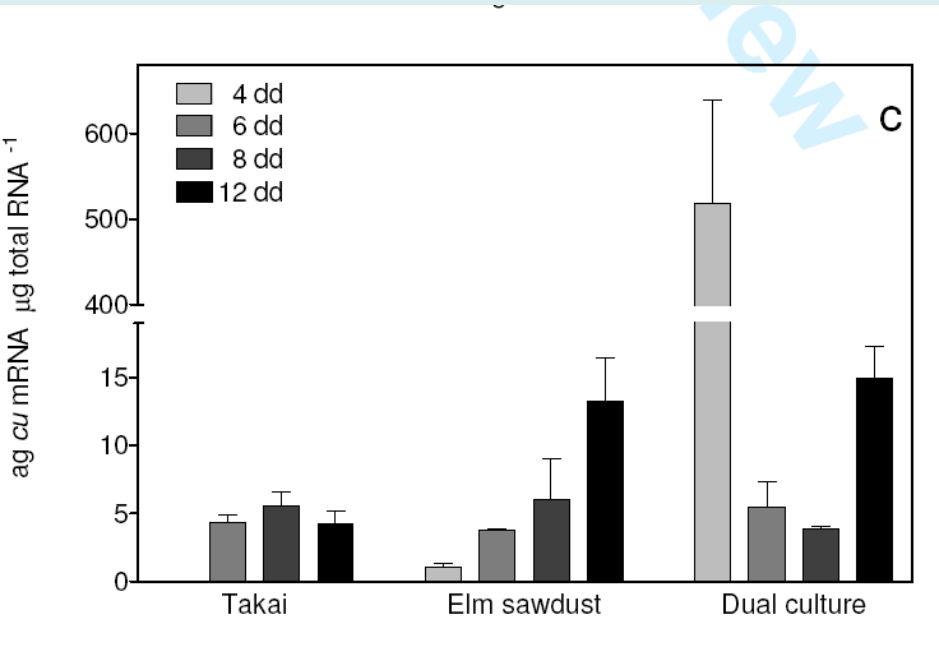
Identification of a Universally Primed-PCR-Derived Sequence-Characterized Amplified Region Marker for an Antagonistic Strain of *Clonostachys rosea* and Development of a Strain-Specific PCR Detection Assay

SERGEY A. BULAT,¹ METTE LÜBECK,^{2*} IRINA A. ALEKHINA,¹ DAN FUNCK JENSEN,² INGE M. B. KNUDSEN,² AND PETER STEPHENSEN LÜBECK^{2†}

 Cerato-ulmin expressed by numerous *Geosmithia* strains, five species – only those from elms

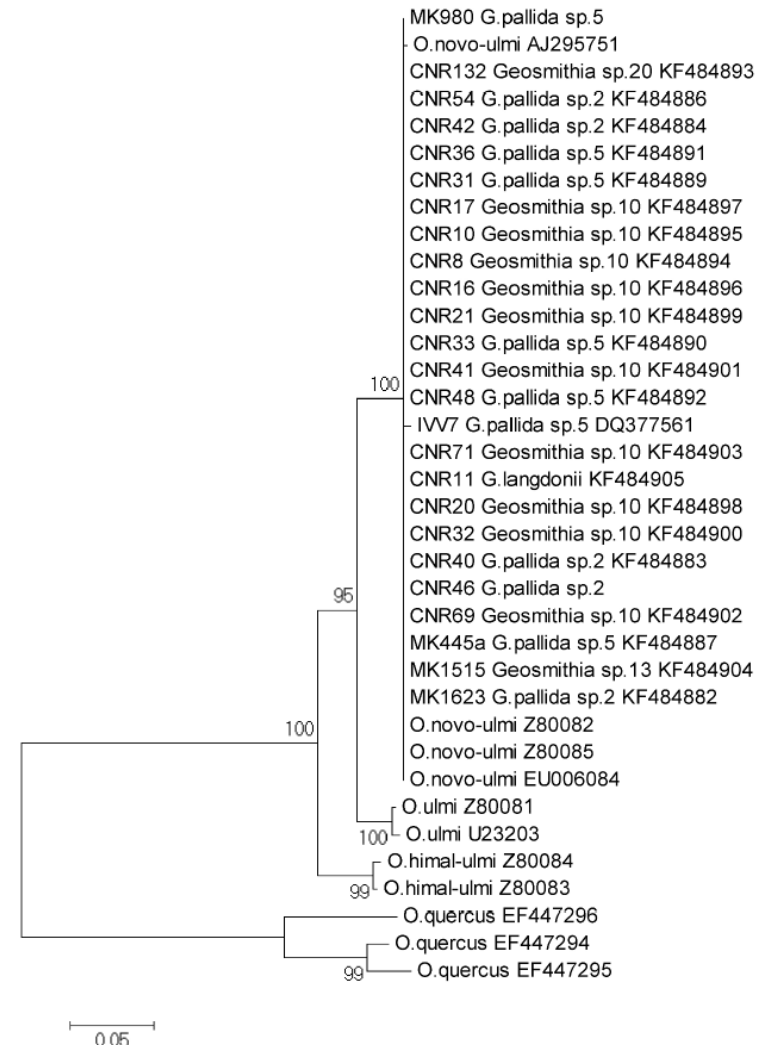
 identical DNA sequence as *Ophiostoma novo-ulmi*

 horizontal transfer affected much more genes



Widespread horizontal transfer of the cerato-ulmin gene between *Ophiostoma novo-ulmi* and *Geosmithia* species

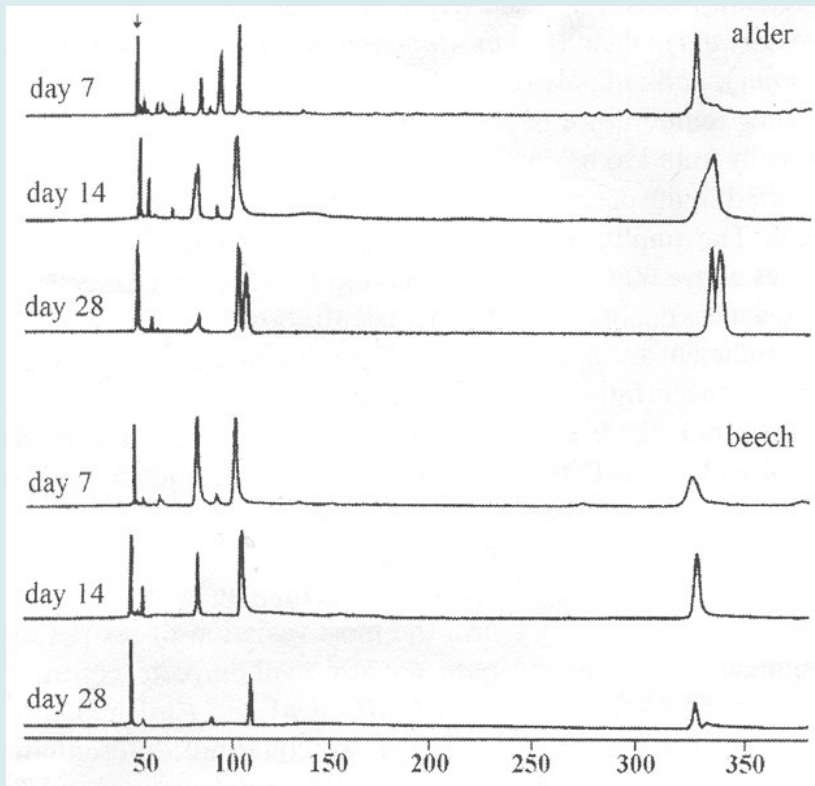
Francesca P. BETTINI^{a,*}, Arcangela FRASCELLA^{a,b}, Miroslav KOLARÍK^{c,d}, Cecilia COMPARINI^b, Alessia L. PEPORI^e, Alberto SANTINI^e, Felice SCALA^f, Aniello SCALA^b



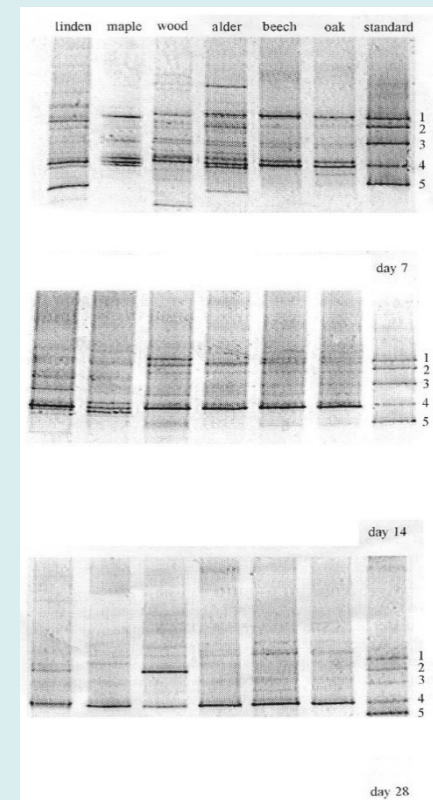
Analýza enviromtálních vzorků

- počet fragmentů = cca počet zjištěných druhů
- klonování či DGGE je citlivější
- nedochází k amplifikaci výhradně houbové DNA
- nedostatečné rozlišení podobných sekvencí
- obtíže s identifikací fragmentů (vhodná srovnávací databáze t-RFLP fragmentů)

t-RFLP



DGGE



Analýza enviromtálních vzorků - klonování

Inserce do vektoru

- vložení PCR produktu do specifického místa vektoru (plazmid)
- ligace (uzavře se plazmid)

Transformace *E.coli* (vložení plasmidu)

- teplotní šok

Kultivace transformovaných buněk

Ověření přítomnosti PCR produktu

- vysetí na agarovou plotnu a selekce kolonií s plazmidem (to lze pomocí reportérových genů – např. modro-bílá selekce), nebo pomocí restrikce
- Izolace DNA
- PCR amplifikace pomocí známých primerů plasmidu
- rozdělení do RFLP typů (3 enzymy), sekvenování