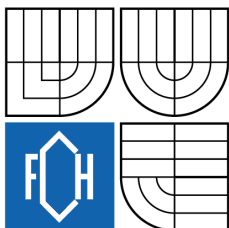


VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA CHEMICKÁ
ÚSTAV CHEMIE POTRAVIN A BIOTECHNOLOGIÍ
FACULTY OF CHEMISTRY
INSTITUTE OF FOOD SCIENCE AND BIOTECHNOLOGY

MOLEKULÁRNÍ TYPIZACE BAKTERIÍ MLÉČNÉHO KVAŠENÍ

MOLECULAR TYPISATION OF LACTIC ACID BACTERIA

BAKALÁŘSKÁ PRÁCE
BACHELOR'S THESIS

AUTOR PRÁCE
AUTHOR

Mgr. ZDENĚK JELÍNEK

VEDOUCÍ PRÁCE
SUPERVISOR

doc. Ing. BOHUSLAV RITTICH, CSc.



Vysoké učení technické v Brně
Fakulta chemická
Purkyňova 464/118, 61200 Brno 12

Zadání bakalářské práce

Číslo bakalářské práce:	FCH-BAK0419/2009	Akademický rok: 2009/2010
Ústav:	Ústav chemie potravin a biotechnologií	
Student(ka):	Mgr. Zdeněk Jelínek	
Studijní program:	Chemie a technologie potravin (B2901)	
Studijní obor:	Biotechnologie (2810R001)	
Vedoucí práce	doc. Ing. Bohuslav Rittich, CSc.	
Konzultanti:		

Název bakalářské práce:

Molekulární typizace bakterií mléčného kvašení

Zadání bakalářské práce:

1. Vypracujte literární přehled k dané problematice
2. Popište použité experimentální metody
3. Vyhodnoťte získané výsledky

Termín odevzdání bakalářské práce: 28.5.2010

Bakalářská práce se odevzdává ve třech exemplářích na sekretariát ústavu a v elektronické formě vedoucímu bakalářské práce. Toto zadání je přílohou bakalářské práce.

Mgr. Zdeněk Jelínek
Student(ka)

doc. Ing. Bohuslav Rittich, CSc.
Vedoucí práce

doc. Ing. Jiřina Omelková, CSc.
Ředitel ústavu

V Brně, dne 1.12.2009

prof. Ing. Jaromír Havlica, DrSc.
Děkan fakulty

Abstrakt

Bakterie mléčného kvašení nacházejí uplatnění nejen v potravinářském průmyslu, ale také ve zdravotnictví. Jsou součástí mnoha fermentovaných mléčných výrobků i běžné střevní mikroflóry.

Pro identifikaci bakterií byla vyvinuta celá řada metod, v této práci se však zaměříme na Rep-PCR a RAPD. Cílem bude zhodnotit vliv různých korelačních koeficientů a algoritmů shlukové analýzy dat na taxonomické zařazení bakterií.

Summary

Lactic acid bacteria play important role both in food industry and health care. They are part of many fermented dairy products and gastrointestinal microflora.

There is a whole variety of methods for bacteria identification, but we will focus on Rep-PCR and RAPD only. The aim of this work will be to discuss the impact of different correlation coefficients and cluster analysis algorithms on the results of taxonomic analysis.

Klíčová slova

Bakterie mléčného kvašení, *Lactobacillus*, taxon, taxonomická jednotka, numerická taxonomie, podobnost znaků, PCR, rep-PCR, RAPD.

Keywords

Lactic Acid bacteria, *Lactobacillus*, taxon, operational taxonomic unit, numerical taxonomy, characters resemblance, PCR, rep-PCR, RAPD.

JELÍNEK, Z. *Molekulární typizace bakterií mléčného kvašení*. Brno: Vysoké učení technické v Brně, Fakulta chemická, 2010. 55 s. Vedoucí doc. Ing. Bohuslav Rittich, CSc.

Prohlašuji, že jsem diplomovou práci vypracoval samostatně a že všechny použité literární zdroje jsem správně a úplně citoval. Diplomová práce je z hlediska obsahu majetkem Fakulty chemické VUT v Brně a může být využita ke komerčním účelům jen se souhlasem vedoucího diplomové práce a děkana FCH VUT.

Mgr. Zdeněk Jelínek

Děkuji doc. Ing. Bohuslavu Rittichovi, CSc., doc. RNDr. Aleně Španové, CSc. a Mgr. Kristýně Turkové za odborné vedení, cenné rady a podněty při zpracování této bakalářské práce.

Mgr. Zdeněk Jelínek

Obsah

1 Úvod	3
2 Teoretická část	4
2.1 Identifikace a taxonomie bakterií	4
2.1.1 Základní pojmy	4
2.1.2 Metody identifikace bakterií	5
2.1.3 Taxonomická analýza	6
2.2 Numerická taxonomie	7
2.2.1 Taxonomický znak	7
2.2.2 Stanovení podobnosti	8
2.2.3 Kódování znaků	8
2.2.4 Míra podobnosti pro binární znaky	8
2.2.5 Míra podobnosti pro ordinální znaky	9
2.2.6 Míra podobnosti pro kvantitativní data	10
2.2.7 Korelační metriky	11
2.3 Shluková analýza dat	11
2.3.1 Hierarchické shlukovací metody	11
2.3.2 Nehierarchické shlukovací metody	13
2.4 Polymerázová řetězová reakce	14
2.4.1 Interrepetitivní polymerázové řetězové reakce (rep-PCR)	15
2.4.2 Náhodná polymerázová řetězová reakce (RAPD)	16
3 Experimentální část	17
3.1 PCR – použité metody	17
3.1.1 Příprava DNA	17
3.1.2 Rep-PCR s DNA zástupců rodu <i>Lactobacillus</i> s využitím primeru (GTG) ₅	17
3.1.3 RAPD-PCR s DNA zástupců rodu <i>Lactobacillus</i> s využitím primeru M13	18
3.1.4 Gelová elektroforéza DNA	19
3.2 Vyhodnocení PCR	19
4 Výsledky	21
4.1 Shrnutí výsledků druhově specifické PCR	21
4.2 Porovnání algoritmů pro 1. sadu fingerprintů	21
4.2.1 Metoda UPGMA	22
4.2.2 Wardova metoda	24
4.2.3 Metoda Neighbor Joining	26
4.2.4 Metoda Single linkage	28
4.2.5 Metoda Complete linkage	30
4.2.6 Zhodnocení výsledků pro 1. sady fingerprintů	31
4.3 Porovnání algoritmů pro 2. sadu fingerprintů	32
4.3.1 Metoda UPGMA	32
4.3.2 Metoda Ward	34

4.3.3	Metoda Neighbor Joining	36
4.3.4	Zhodnocení výsledků shlukové analýzy 2. sady fingerprintů	37
4.4	Porovnání koeficientů míry podobnosti	38
4.4.1	Pearsonův korelační koeficient	38
4.4.2	Korelační koeficient Cosine	40
4.4.3	Koeficient Jaccard	42
4.4.4	Koeficient Dice	44
4.4.5	Jeffreyho koeficient	46
4.4.6	Koeficient Ochiai	48
4.4.7	Zhodnocení koeficientů	49
5	Diskuse	50
6	Závěr	51
	Seznam použitých zdrojů	52
	Seznam použitých zkratk a symbolů	55

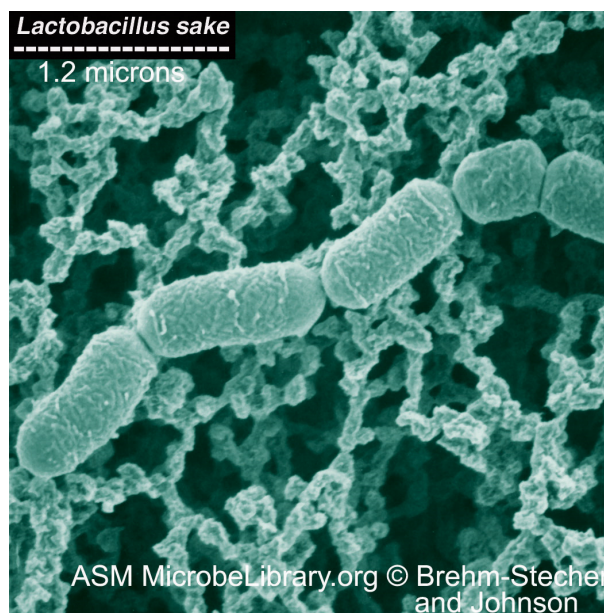
1. Úvod

Bakterie mléčného kvašení jsou nesporulující, grampozitivní koky nebo tyčinky, které spojují podobné morfologické, metabolické a fyziologické charakteristiky. Řadíme sem bakterie rodu *Lactobacillus*, *Leuconostoc*, *Pediococcus* a *Streptococcus*, ale také bakterie rodu *Aerococcus*, *Carnobacterium*, *Enterococcus*, *Lactococcus*, *Oenococcus*, *Tetragenococcus*, *Vagococcus* a *Weissella*. Z historických důvodů bývají k bakteriím mléčného kvašení řazeny také bakterie rodu *Bifidobacterium*. Ty sice s bakteriemi mléčného kvašení sdílí celou řadu vlastností, ovšem nejsou fylogeneticky příbuzné a liší se metabolismem cukrů [19].

Z průmyslového hlediska je pravděpodobně nejvýznamnější rod *Lactobacillus*. Ten obsahuje celou řadu druhů, z nichž většina má zcela nezastupitelnou úlohu nejen při fermentaci potravin a krmiv, ale současně působí probioticky. Ačkoliv se některé druhy bakterií rodu *Lactobacillus* účastní jistých infekčních onemocnění, např. endokarditidy u imunitně oslabených jedinců, nelze je označit za patogeny. Naopak, u zdravých lidí jsou laktobacily přítomny v ústní dutině, v tlustém střevě a jsou dominantními mikroorganismy v ženské vagině. Nezpochybnitelná je jejich schopnost stimulovat imunitu a bránit průjemovým onemocněním. Zvažuje se také jejich použití jako nosičů při vakcinaci [1, 2].

Genetické podstatě bakteriálního rodu *Lactobacillus* je věnována velká pozornost. V genomech jednotlivých druhů se předpokládají rozsáhlé změny genetické informace jak v podobě delecí, tak v důsledku horizontálního přenosu genů. Tyto změny jsou pak zodpovědné za adaptaci jednotlivých druhů na různé prostředí.

Pro přesné určení druhu, stanovení taxonomických závislostí a výzkum genetické podstaty bakterií rodu *Lactobacillus* se používají fenotypové (fyziologické a biochemické) a molekulárně-biologické metody. Ačkoliv jsou fyziologické a biochemické metody přínosné pro výzkum praktické aplikace bakterií, neumožňují jednoznačnou identifikaci blízké příbuzných druhů. Proto se ke stanovení příbuznosti používají nové molekulárně-biologické metody [7].



Obrázek 1.1: Bakterie *Lactobacillus sake* [11]

2. Teoretická část

2.1. Identifikace a taxonomie bakterií

2.1.1. Základní pojmy

Molekulární diagnostika se často zabývá identifikací izolovaných mikroorganismů s jejich následným zařazením do některého existujícího bakteriálního taxonu (druhu). Aby nedošlo k záměně, je na tomto místě vhodné objasnit význam dvou pojmů – detekce a identifikace. Detekcí se rozumí kvalitativní zjištění bakterie v analyzované matrici. Identifikací se rozumí srovnání určované bakterie se známou bakterií, např. se sbírkovým nebo typovým kmenem. Jedná se o proces, při kterém se zjišťuje, ke kterému známému taxonu lze izolát přiřadit (v případě negativního výsledku se může jednat o nový taxon). Při analýze většího počtu izolátů se vedle zařazení do druhu studuje i jejich příbuznost. V mikrobiologii existuje úzká souvislost mezi identifikací a bakteriální taxonomií.

Taxonomická jednotka (OTU – operational taxonomic unit) je skupina konkrétních organismů, které mají společné určité znaky, a tím se odlišují od ostatních taxonů. Taxon je tedy založen na dostatečné míře vzájemné podobnosti jeho členů a zároveň na dostatečné míře odlišnosti těchto členů od členů patřících do jiného taxonu.

Základní jednotkou mikrobiální taxonomie je druh (species, sp.). Podle současné fylogenetické definice druhu je tato jednotka u prokaryot definována [9] jako:

- jasně vymezená skupina navzájem příbuzných kmenů zahrnujících typový kmen,
- sdílející 70% a vyšší homologii DNA (metoda DNA reasociace),
- vykazující, až na výjimky, shodné fenotypové znaky,
- současně mající některé znaky odlišné od jiných skupin,
- rozdíl v T_m (bodu tání) menší než 5 °C.

Ke stanovení fylogenetického postavení jakéhokoliv prokaryotního organismu lze využít sekvenční analýzu 16S rDNA (nebo jiných konzervativních genů). Hodnota podobnosti 97 % a vyšší je při sekvenční analýze udávána jako hraniční pro zástupce stejného druhu. Geneticky příbuzné organismy, které mají v rámci daného druhu vysokou DNA homologii, neodlišitelnou DNA-DNA hybridizací, které se však odlišují genotypem od dalších kmenů v rámci druhu, se označují jako poddruh (subspecies – subsp., ssp.). Druhy jako základní jednotky jsou seskupeny do rodů (genus). Rod se skládá z jednoho či více druhů, či poddruhů.

Základní taxonomické kategorie jsou:

- doména (domain)
- kmen (phyllum)
- třída (classis)
- řád (ordo)

- čeleď (familia)
- rod (genus)
- druh (species)

Jak již bylo uvedeno, základní taxonomickou jednotkou je druh. Jeho jméno se skládá z jména rodového a druhového. Další značení označuje určitý kmen. Taxonomickou jednotkou nižší než druh je poddruh. Pro každý druh je mezinárodní komisí stanoven typový kmen. Tento kmen, jehož znaky by měly být identické s původním popisem v odborné literatuře, je vedle čísla označen písmenem T. Je udržován v čisté kultuře a deponován v některé mezinárodně uznávané sbírce organismů. Vlastnosti typového kmene nemusí být typické pro daný druh. Soubory nebo populace bakteriálních buněk, které rostou na laboratorních živných půdách (tekutých nebo pevných) se označují jako bakteriální kultury. Podle výsledku studia bakteriálních kultur se usuzuje na vlastnosti jednotlivých buněk.

2.1.2. Metody identifikace bakterií

K identifikaci bakterií se používá řada mikrobiologických, biochemických, analytických a fyzikálně-chemických metod. Identifikace mikroorganismů původně spočívala na stanovení fenotypových vlastností. Mezi základní charakteristiky patří:

1. morfologie buněk (tvar, velikost, seskupení, barvení podle Grama, barvení spor apod.)
2. kultivační vlastnosti (tvar a pigmentace kolonií na pevném médiu, růst v tekutém médiu)
3. fyziologické vlastnosti (vztah ke kyslíku, teplotní rozmezí pro růst, citlivost ke žlučovým solím, k nízkému pH, citlivost na antibiotika, typ metabolismu – např. fermentativní, aerobně nebo anaerobně respirační)
4. biochemické testy (využívání sacharidů, redukce dusičnanů, hydrolýza škrobu, želatiny, skulinu, důkaz tvorby indolu, sirovodíku, dekarboxyláz lyzinu nebo ornitinu, důkaz produkce některých enzymů jako např. katalázy, fosfatasy, arginin dehydrogenasy, ureasy, oxidasy, koagulasy, beta-galaktosidasy apod.)
5. další doplňující testy jako zjištění sérologických vlastností, citlivosti k bakteriofágům a další. Je známo, že fyziologické, případně další vlastnosti lze ovlivnit podmínkami kultivace [4]. Proto se k identifikaci bakterií používají ve stále větší míře molekulárně biologické metody
6. molekulárně biologické metody jsou založeny na fyzikálně-chemické charakterizaci molekul (mastné kyseliny, proteiny, sacharidy, nukleové kyseliny), které jsou produkovány mikroorganismy a tudíž jsou univerzálně použitelné

Molekulárně biologické metody lze rozdělit do tří velkých skupin na základě typu makromolekul používaných k analýze:

- polysacharidy a lipidy

- proteiny buněčné stěny a cytoplasmatické membrány
- nukleové kyseliny (DNA a RNA)

V diagnostice mikroorganismů se ze širokého spektra molekulárně biologických metod běžně používají pouze některé. V současné době je dáována přednost metodám založeným na analýze DNA, protože tu obsahují všechny buňky. V dalším textu budou zmíněny pouze vybrané metody, které budou použity při statistickém zpracování výsledků. Jedná se o metody:

- RAPD-PCR (náhodná polymerázová řetězová reakce; RAPD), která se používá k detekci polymorfizmů DNA. Je to typ PCR, při které jsou segmenty DNA amplifikovány v náhodných, předem nedefinovaných místech, ve kterých hybridizuje primer o náhodné sekvenci. Používají se krátké primery (8–12 nukleotidů), přičemž znalost DNA sekvence amplifikovaného fragmentu není vyžadována [18, 20].
- Rep-PCR (polymerázová řetězová interrepetitivní reakce) je typizační metodou, při které se amplifikují sekvence oddělující v genomu různé typy repeticí, tj. v haploidním genomu se opakující DNA sekvence (např. REP element, ERIC, BOX element, IS elementy) [27].
- Restrikční analýza amplifikované ribosomální DNA (Amplified Ribosomal DNA Restriction Analysis - ARDRA), která je založena na restrikční analýze amplikonů (PCR produktů) [10, 26]. Předpokladem použití metody ARDRA je přítomnost vhodného cílového místa pro určitou restriktasu v amplikonu. Jejich výskyt v sekvenci lze zjistit pomocí analýzy *in silico*. Celkové nebo částečné sekvence jednotlivých bakteriálních druhů lze získat z databáze GenBank (<http://www.ncbi.nlm.nih.gov>), která je součástí Národního centra pro biotechnologické informace (NCBI) v USA.

Analýza *in silico* znamená provádět analýzu (experiment) na počítači nebo pomocí počítačové simulace. V této souvislosti je vhodné zmínit význam následujících termínů: *in vivo* – provádět experiment na živém organismu; *in vitro* – provádět experiment v umělých, např. laboratorních podmínkách mimo organismus; ve „zkumavce“ nebo „Petriho misce“; *ex vitro* – znamená provádět experiment s organismem (tkání) mimo organismus za definovaných podmínek.

2.1.3. Taxonomická analýza

Na úrovni jednotlivých taxonů, včetně bakteriálních druhů a poddruhů, existuje genetická diverzita, která umožňuje jejich rozlišení. Ke stanovení diverzity se v řadě případů používají separační metody (gelová elektroforéza, chromatografické metody), obvykle jako způsob vyhodnocení výsledků molekulárně-diagnostických metod. Výsledkem je záznam, který se nazývá fingerprint (otisk). V případě analýzy DNA, mohou mít otisky podobu spekter restrikčních fragmentů nebo produktů PCR rozdělených gelovou elektroforézou nebo restrikčních fragmentů hybridizujících se specifickou sondou. V případě analýzy proteinů mohou mít otisky charakter elektroforeogramu jednotlivých proteinů, hmotnostních spekter peptidů získaných pomocí hmotnostní spektrometrie – např. MALDI-TOF (Matrix-assisted laser desorption/ionization – time-of-flight mass spectrometer). Na základě fingerprintů lze porovnávat jednotlivé mikroorganismy mezi sebou a určovat tak

jejich genetickou příbuznost (typizace). Analogickými pojmy k DNA typizaci jsou DNA profilování a genetický fingerprinting (DNA profiling, DNA typing, genetic fingerprinting).

DNA fingerprinting je tedy technika, pomocí níž lze zobrazit profil DNA konkrétního organismu. Když jsou s neznámými kmeny současně analyzovány jednoznačně identifikované typové a sbírkové kmeny, výsledkem může být zařazení kmenů do určitého taxonu (využití shlukové analýzy).

Mezi základní kritéria, jimiž jsou typizační techniky vzájemně porovnávány, patří typovatelnost, tj. míra schopnosti poskytnout jednoznačný výsledek pro každý testovaný organismus (např. druh), reprodukovatelnost, tj. pravděpodobnost dosažení téhož výsledku při opakování testu, a rozlišovací účinnost, která určuje schopnost metody spolehlivě od sebe odlišit testované mikroorganismy (druhy, kmeny). Typizační metody založené na principu amplifikace DNA nabízejí jednoduché, rychlé a levné postupy analýzy mikroorganismů.

2.2. Numerická taxonomie

Numerická taxonomie je dynamicky se rozvíjející obor, který využívá moderních technologií pro sběr a analýzu taxonomických dat. Často se používá také ekvivalentní název, *matematická taxonomie*.

Filosofie numerické taxonomie je založena na myšlenkách, které vyslovil Michel Adanson již v roce 1763. Adanson tehdy odmítl názor, že některé znaky mohou mít větší důležitost, a naopak vyslovil myšlenku, že by taxonomie měla být založena na analýze podobnosti a že je tedy nutné brát v úvahu všechny znaky. Principy moderní numerické taxonomie stanovili v roce 1973 Sneath a Sokal, když navázali právě na práci Adansona [22]:

1. Klasifikace dat je tím přesnější, čím větší je soubor dat a počet analyzovaných znaků.
2. Každý znak má stejnou váhu.
3. Podobnost dvou entit je funkcí podobností jejich jednotlivých znaků.
4. Každé dva taxony je možné odlišit, protože každé dva různé organismy obsahují některé různé znaky.
5. Z taxonomických struktur a korelací jednotlivých znaků je možné získat jisté znalosti o mechanismech vývoje a evolučních cestách.
6. Taxonomie je empirická věda.
7. Klasifikace je založena na podobnosti.

2.2.1. Taxonomický znak

Pro odlišení jednotlivých taxonů je zaveden pojem taxonomický znak. Je to vlastnost, která jednoznačně odlišuje jeden taxon od jiného. Příkladem může být produkce pigmentu. Obecně platí, že taxonomický znak může nabývat dvou nebo více stavů. Důležité je, aby se současně jednalo o jednotkový znak, tedy znak dále nedělitelný.

Dále je pro numerickou analýzu důležité, aby analyzovaný soubor dat obsahoval alespoň 60 znaků (doporučeno je více než 80) [22].

2.2.2. Stanovení podobnosti

Pro vyhodnocení taxonomických závislostí je nutné stanovit *podobnost* jednotlivých taxonomických jednotek, tedy stanovit příbuznost analyzovaných dat. Zjištěná míra podobnosti bude záviset na porovnávání rysech a na způsobu, jakým byla získaná informace kódována. V některých případech bývá výhodnější pracovat s pojmem opačným, *odlišností* [5].

2.2.3. Kódování znaků

Vhledem k množství analyzovaných dat se stává nezbytným použití výpočetní techniky. Nejen z tohoto důvodu je proto důležitá správná volba znaků. Pro počítačovou analýzu jsou výhodné zejména znaky dvoustavové. Některé znaky však často není snadné na dvoustavové převést. Jedná se kvalitativní vícestavové znaky, např. barva květů (červená, žlutá, bílá. . .), či o kvantitativní vícestavové znaky, např. počet květů. Tyto znaky sice můžeme převést na dvoustavové, např. určením, že květ je bílý nebo barevný, je jeden nebo je jich více, ovšem takovým převodem přidělíme některým stavům větší váhu než stavům ostatním. Tato odchylka se ovšem eliminuje použitím dostatečně velkého souboru znaků [22].

2.2.4. Míra podobnosti pro binární znaky

Obecně je možné kódované znaky zapsat do matice o rozměrech $t \times n$, kde t (počet řádků) označuje počet taxonomických jednotek a n (počet sloupců) počet znaků. V nejběžnějším případě ovšem pracujeme pouze s dvoustavovým kódováním znaků. Výsledný tvar je uveden v Tabulce 2.1.

Tabulka 2.1: Míra podobnosti

	1	0	
1	a	b	$a + b$
0	c	d	$c + d$
	$a + c$	$b + d$	$a + b + c + d$

Symbol a v tabulce označuje počet znaků, pro které nabývají obě taxonomické jednotky hodnoty 1. Symbol b naopak představuje počet znaků, ve kterých mají obě taxonomické jednotky hodnotu 0. Počet znaků, pro které nabývá taxonomická jednotka i hodnoty 1 a taxonomická jednotka j hodnoty 0, označuje symbol c . Obdobně pro hodnotu d . Počet shod je tedy $a + d$, počet neshod $b + c$ [5, 22].

Výpočtem míry podobnosti se zabývá celá řada vzorců, z nichž uvedeme jen ty nejčastěji používané [3, 5, 17]:

Jednoduchý koeficient shody

V roce 1958 zavedli Sokal a Michener tzv. jednoduchý koeficient shody (Simple Matching Coefficient). Tento koeficient představuje poměr počtu shod (tedy počtu znaků, pro které nabývají obě taxonomické jednotky současně hodnoty 1 nebo současně hodnoty 0) a celkového počtu všech znaků.

$$s_{ij} = \frac{a + d}{a + b + c + d} \quad (2.1)$$

Jednoduchý koeficient shody je výhodný, pokud hodnota 0 pokrývá stejné množství informace jako hodnota 1. V opačném případě je výhodnější použít jiné koeficienty.

Jaccardův koeficient

Na rozdíl od jednoduchého koeficientu shody vylučuje Jaccardův koeficient „negativní“ shodu:

$$s_{ij} = \frac{a}{a + b + c} \quad (2.2)$$

Jedná se tedy o poměr pozitivních shod k celkovému počtu znaků s výjimkou všech negativních shod. Výsledná hodnota Jaccardova koeficientu je závislá na směru kódování; zaměníme-li tedy pozitivní a negativní shody, získáme odlišný výsledek.

Diceův koeficient

Třetím nejpoužívanějším koeficientem podobnosti je Diceův koeficient (známý též jako Czekanowskiho koeficient, Sorensenův koeficient či koeficient Nei and Li). Podobně jako Jaccard vylučuje i Dice negativní shodu. Přitom ale dává dvojnásobnou váhu pozitivní shodě:

$$s_{ij} = \frac{2a}{2a + b + c} \quad (2.3)$$

Výsledná hodnota Diceova koeficientu je závislá na směru kódování.

Ochiaiho koeficient

Nejvyšší hodnotou koeficientu je 1. Tato hodnota značí dokonalou podobnost a koeficient ji nabývá, pokud $b = c = 0$, tedy pokud jsou zahrnuty pouze shodné znaky. Nulovou hodnotu nabývá koeficient pro $a = 0$.

$$s_{ij} = \frac{a}{\sqrt{(a + b)(a + c)}} \quad (2.4)$$

Jeffreyho koeficient

Jeffreyho koeficient (Jeffrey's X coefficient) je vyjádřen jako součet poměrů počtu pozitivních shod ku počtu znaků jednotlivých taxonomických jednotek:

$$s_{ij} = \frac{a}{a + b} + \frac{a}{a + c} \quad (2.5)$$

2.2.5. Míra podobnosti pro ordinální znaky

Míru podobnosti pro ordinární znaky lze vyjádřit pomocí Gowerova koeficientu podobnosti.

Gowerův koeficient podobnosti

Někdy bývají taxonomické jednotky popsány binárními znaky, kvalitativními znaky nabývajících více než dvou stavů i kvantitativními znaky. V těchto případech je výhodné použít Gowerův koeficient podobnosti:

$$s_{ij} = \frac{\sum_{k=1}^p s_{ijk}}{\sum_{k=1}^p w_{ijk}} \quad (2.6)$$

V závislosti na tom, jestli je srovnání taxonomických jednotek i a j shledáno pro znak k validním, může váha w_{ijk} nabývat hodnot 0 nebo 1. Hodnoty 0 přitom nabývá, pokud není znak k pro jednu nebo obě taxonomické jednotky známý. Jestliže platí, že $w_{ijk} = 0$, potom platí také, že $s_{ijk} = 0$.

1. *Binární znaky.* Hodnoty a váhy pro binární znaky uvádí následující tabulka:

OTU i	+	+	−	−
OTU j	+	−	+	−
Hodnota s_{ijk}	1	0	0	0
Váha w_{ijk}	1	1	1	0

Pokud by byly taxonomické jednotky popsány pouze binárními znaky, odpovídala by výsledná hodnota Gowerova koeficientu hodnotě Jaccardova koeficientu.

2. *Kvalitativní znaky s více než dvěma stavy.* Hodnota $s_{ijk} = 1$, jestliže je znak k shodný pro obě taxonomické jednotky, a $s_{ijk} = 0$, jestliže se obě jednotky ve znaku k odlišují.

Pokud by byla podobnost popsána pouze kvalitativními znaky, odpovídala by výsledná hodnota Gowerova koeficientu hodnotě jednoduchého koeficientu shody.

3. *Kvantitativní data.* Výpočet hodnoty s_{ijk} popisuje rovnice 2.7:

$$s_{ijk} = \frac{1 - |x_{ik} - x_{jk}|}{R_k} \quad (2.7)$$

x_{ik} resp. x_{jk} je hodnota znaku k pro taxonomickou jednotku i resp. j , R_k je rozsah hodnot jednotlivých znaků v odpovídající populaci nebo rozsah znaků v souboru sledovaných taxonomických jednotek.

2.2.6. Míra podobnosti pro kvantitativní data

Míru podobnosti pro kvantitativní data lze využít pro stanovení taxonomické vzdálenosti.

Taxonomická vzdálenost

Vzdálenost mezi dvěma taxonomickými jednotkami lze vypočítat velmi snadno jako 100 mínus podobnost uváděná v procentech.

Přímý výpočet nabízí vzorec *Euklidovské vzdálenosti* [5, 22]:

$$d_{ij} = \sqrt{\sum_{k=1}^n (x_{ik} - x_{jk})^2} \quad (2.8)$$

Průměrná taxonomická vzdálenost je potom definována jako [5, 22]:

$$\overline{d_{ij}} = \sqrt{\frac{d_{ij}^2}{n}} \quad (2.9)$$

2.2.7. Korelační metriky

Hodnoty korelačních metrik leží vždy v rozmezí -1 až $+1$. $+1$ označuje úplnou korelaci, zatímco -1 úplnou antikorelaci. Žádnou korelaci pak značí 0 [12].

Pearsonova korelace

Nejběžnější ze všech korelačních metrik je Pearsonova korelace (Pearson Product Moment Correlation). Pearsonova korelace popisuje míru lineárního vztahu mezi dvěma proměnnými. Přitom jde o pozitivní vztah – nabývá-li funkce na ose x vysokých hodnot, nabývá vysokých hodnot také na ose y [12]:

$$r_p = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\left[\sum_{i=1}^n (x_i - \bar{x})^2\right] \left[\sum_{i=1}^n (y_i - \bar{y})^2\right]}} \quad (2.10)$$

Korelace Cosine

Velmi blízká Pearsonově korelaci je korelace Cosine:

$$r_c = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}} \quad (2.11)$$

I tento korelační koeficient měří blízkost k přímce lineární regrese. Na rozdíl od Pearsonovy korelace ovšem leží počátek grafu korelace Cosine v bodě $(0, 0)$ [12].

2.3. Shluková analýza dat

Shluková analýza dat zahrnuje celou řadu výpočetních postupů, které umožňují rozdělit daný soubor dat na několik homogenních podmnožin takovým způsobem, aby si byly objekty uvnitř jednotlivých shluků co nejvíce podobné. Objekty z různých shluků by si naopak měly být podobny co nejméně.

Kromě biologie a chemie se shluková analýza uplatňuje i v celé řadě dalších oborů, např. ve statistice, psychologii, při analýze dat nebo strojovém učení.

Metody shlukové analýzy lze rozdělit na *hierarchické* a *nehierarchické*.

2.3.1. Hierarchické shlukovací metody

Shlukovací proces hierarchických metod má charakter hierarchické posloupnosti rozkladů množiny zadaných dat. Podle způsobu tvorby shluků se přitom tyto metody dělí na *aglomerativní* a *divizivní*. Při vytváření shluků se u aglomerativních hierarchických metod tvoří shluky postupně od nejmenších po největší. Použijeme-li k vytváření shluků divizivní metody, budeme postupovat opačně – všechny objekty se postupně rozdělí na skupiny, které se rekurzivně dělí dále [12].

Na vstupu se algoritmu vždy předává množina prvků a *matice nepodobnosti*, která obsahuje vzdálenosti mezi každými dvěma prvky. Výstupem algoritmu je strom hierarchie taxonomických závislostí, tzv. *dendrogram*.

Obecný tvar algoritmu lze shrnout do těchto kroků [24]:

1. V prvním kroku tvoří každý objekt z množiny zadaných dat samostatný shluk.
2. Najdou se dva nejbližší shluky a ty se propojí.
3. Vytvoří se nová matice podobnosti.
4. Pokud neexistuje pouze jeden shluk, pokračuje se 2. krokem.

Metoda nejbližšího souseda (Single Linkage)

Metoda nejbližšího souseda definuje podobnost dvou shluků jako nejmenší vzdálenost mezi libovolnými dvěma objekty uvnitř dvou různých shluků.

Matematicky je vzdálenost mezi dvěma shluky definována jako [24]:

$$d_{r,s} = \min (D(x_{ri}, x_{sj})) \quad (2.12)$$

Metoda nejvzdálenějšího souseda (Complete Linkage)

Metoda nejvzdálenějšího souseda definuje podobnost dvou shluků jako největší vzdálenost (nejmenší podobnost) mezi libovolnými dvěma objekty uvnitř dvou shluků. Vzniká tak mnohem diferencovanější diagram než v případě metody nejbližšího souseda. Tato metoda je současně méně náchylná k šumu [24].

Matematicky se vzdálenost mezi dvěma shluky definuje takto:

$$d_{r,s} = \max (D(x_{ri}, x_{sj})) \quad (2.13)$$

Metoda Neighbor Joining

Podobně jako ostatní výše uvedené metody, sjednocuje i metoda Neighbor Joining shluky po dvojicích tak dlouho, dokud nezůstane pouze jediný shluk. Kritériem sjednocení jsou ovšem tentokrát vzdálenost mezi dvojicí shluků a průměrná vzdálenost k ostatním uzlům v grafu [21].

Dvojice shluků (i, j) se volí tak, aby byla hodnota $Q(i, j)$ minimální:

$$Q(i, j) = D(i, j) - u(i) - u(j), \quad (2.14)$$

kde

$$u(l) = \sum_{k=0}^{r-1} \frac{D(l, k)}{r-2}, \quad (2.15)$$

a r je počet zbývajících shluků.

Poté, co je nalezena minimální Q -hodnota, dojde k přepočtu matice vzdáleností tak, že se odstraní sloupce a řádky i a j . Současně je přidán nový řádek a sloupec se vzdálenostmi k novému shluku [21]:

$$D(a, k) = \frac{D(i, k) + D(j, k) - D(i, j)}{2} \quad (2.16)$$

Wardova metoda

Při výběru shluků volí Wardova metoda takové dvojice, aby součet čtverců vzdáleností mezi každým bodem a těžištěm shluku byl co nejmenší. Začíná se od shluky, které tvoří jediný bod, v každé iteraci se sloučí dva shluky. Již sloučené shluky se znovu nedělí.

Odchylka sumy čtverců pro shluk k je definována jako:

$$E_k = \sum_{i=1}^{n_k} \|z_{ik} - \bar{z}_k\|^2, \quad (2.17)$$

kde z_{ik} je vektor proměnných ve shluku k , jehož velikost je n_k a střední hodnota $\bar{z}_k$.

Celková odchylka sumy čtverců pro g shluků je poté:

$$E = \sum_{k=1}^g E_k \quad (2.18)$$

Protože Wardova metoda minimalizuje sumu čtverců odchylek, maximalizuje homogenitu uvnitř shluků a současně minimalizuje mezishlukovou homogenitu [16].

UPGMA (Metoda průměrné vzdálenosti)

UPGMA (Unweighted Pair Grouping Method with Arithmetic-Mean) patří mezi nejčastěji používané metody shlukové analýzy dat. Kritérium vzniku shluků definuje UPGMA jako průměrnou vzdálenost všech objektů v jednom shluku k průměrné vzdálenosti objektů ve druhém shluku. UPGMA tedy při řešení problému analýzy dat spojuje přístupy metody nejbližšího i nejvzdálenějšího souseda.

Výpočet podobnosti mezi shluky r a s o velikosti n_r resp. n_s lze popsat rovnicí:

$$d_{r,s} = \frac{1}{n_r n_s} \sum_{i=1}^{n_r} \sum_{j=1}^{n_s} (D(x_{ri}, x_{sj})) \quad (2.19)$$

Metodu UPGMA je třeba používat s jistou obezřetností, protože dává všem objektům v souboru dat stejnou váhu. Proto je vždy nutné pracovat s reprezentativním vzorkem dat [14].

WPGMA

Variantou algoritmu UPGMA je algoritmus WPGMA (Weighted Pair Grouping Method with Arithmetic-Mean). Jedinou odlišností je přitom možnost přidělit různým objektům v souboru dat různou váhu. Tímto způsobem umožňuje WPGMA pokrýt situaci, kdy porovnáváme data s rozdílnými počty objektů [14].

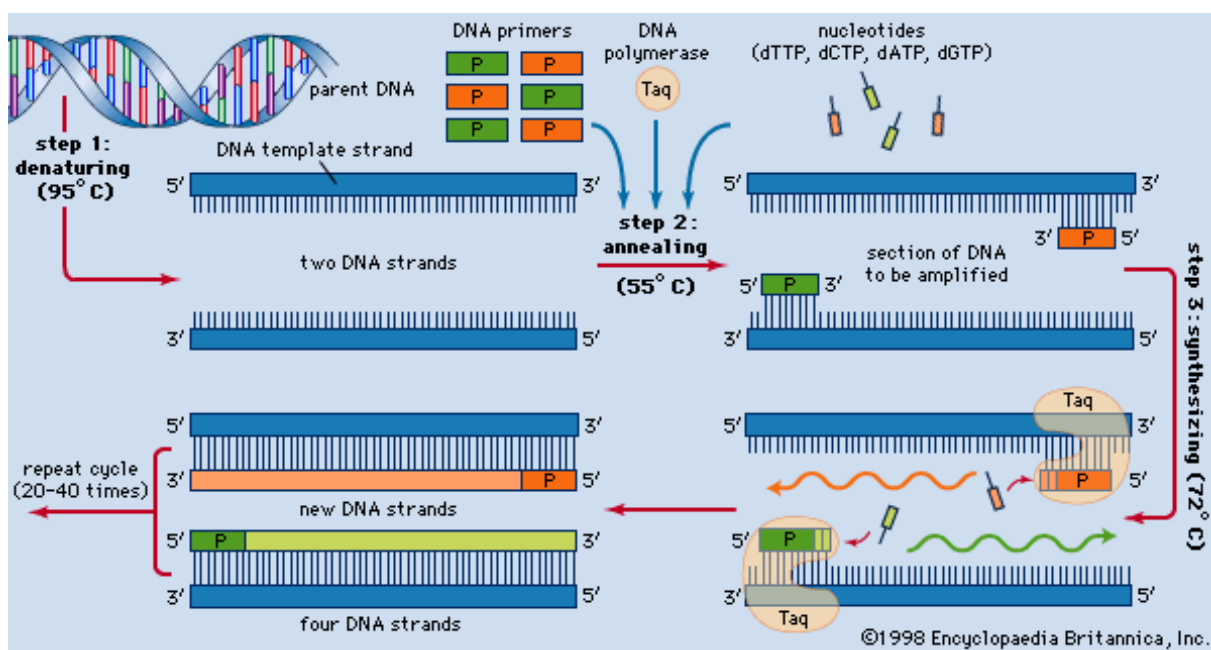
2.3.2. Nehierarchické shlukovací metody

Nehierarchické shlukovací metody naopak nevytváří hierarchickou strukturu, ale rozkládají výchozí množinu do podmnožin tak, aby bylo splněno určité kritérium [12].

2.4. Polymerázová řetězová reakce

Polymerázová řetězová reakce (PCR) a postupy na ní založené patří mezi nejvíce používané molekulárně-diagnostické metody. Podstatou PCR je cyklicky se opakující syntéza určitého úseku DNA. K vybraným úsekům komplementárních vláken denaturované DNA se hybridizují (připojují) krátké syntetické oligonukleotidy (primery), od nichž probíhá syntéza nového řetězce DNA. Syntéza je katalyzována DNA dependentní DNA-polymerasou. PCR je proces, při němž se v závislosti na teplotě reakční směsi pravidelně střídají tři kroky, během nichž probíhají tři odlišné děje [23]:

- denaturace dvouřetězcových molekul DNA (94°C),
- připojení primerů k odděleným řetězcům DNA ($50\text{--}65^{\circ}\text{C}$),
- syntéza nových řetězců DNA katalyzovaná DNA-polymerasou ($65\text{--}75^{\circ}\text{C}$).



Obrázek 2.1: Polymerázová řetězová reakce [8]

Postupným opakováním tohoto procesu se exponenciálně (2^n , kde n představuje počet cyklů) syntetizuje až 10^9 kopií vybraného úseku DNA. Optimální počet cyklů je závislý na výchozí koncentraci templátové DNA a zpravidla se pohybuje v rozmezí od 25–35 cyklů. Příliš vysoký počet cyklů významně snižuje účinnost DNA polymerasy. Výsledným produktem PCR jsou amplikony, což jsou úseky DNA definované délkou o velikosti obvykle desítek až tisíc párů bází (bp). Jejich přítomnost se v reakční směsi prokazuje elektroforézou na agarosovém gelu nebo kvantitativním měřením množství produktu v reálném čase (PCR v reálném čase).

Reakční směs (obvykle v rozmezí 25–100 μl) se skládá z následujících složek:

- *DNA templát*: slouží jako matrice pro syntézu nových řetězců DNA. Obsahuje cílová místa pro primery. Typická množství bakteriální DNA, přidávané do reakce, jsou 10 a 1 ng a 10 pg.

2.4. POLYMERÁZOVÁ ŘETĚZOVÁ REAKCE

- *Oligonukleotidové primery*: bývají synteticky připravené a jsou komplementární k DNA templátu, který má být amplifikován. Typické primery mají 18–30 nukleotidů a obsahují 40–60 % GC bází. Teplota tání primerů bývá v rozmezí 55–80 °C. Primery by neměly být mezi sebou komplementární, zejména ne na 3'-konci, kde překrytí dvou nebo tří bází může vést ke vzniku dimeru primerů, zejména při nadbytku primerů. Potřebná koncentrace každého primeru pro jednu reakci je 0,1–0,5 $\mu\text{mol}\cdot\text{l}^{-1}$.
- *DNA-polymerasa*: syntetizuje novou DNA ve směru 5' $\rightarrow$ 3' podle sekvence nukleotidů v komplementárním řetězci DNA od místa navázání primeru. Ke katalýze se používají termostabilní polymerasy, např. Taq DNA-polymerasa izolovaná z mikroorganismu *Thermus aquaticus*.
- *2'-deoxynukleosid-5'-trifosfáty (dNTP)*: (dATP, dCTP, dGTP, dTTP) – stavební kameny pro syntézu nové DNA. Optimální koncentrace je 200 $\mu\text{mol}\cdot\text{l}^{-1}$, toleranční rozpětí je 20–400 $\mu\text{mol}\cdot\text{l}^{-1}$. Vysoké koncentrace dNTP (od 4 mM a výše) působí inhibičně, protože vyvazují hořčnaté ionty (Mg^{2+}).
- *Mg^{2+} ionty*: jsou nezbytné pro aktivitu DNA-polymerasy. Koncentrace Mg^{2+} musí být optimalizována pro každou kombinaci primerů a DNA templátu. Obvykle se používá koncentrace 1,5 mM Mg^{2+} . Toleranční rozpětí je 0,5–8,0 mM. Vyšší koncentrace iontů snižuje specifitu PCR. U primerů bohatých na G a C báze je ale použití vyšší koncentrace Mg^{2+} vhodnější.
- *PCR pufr*: vytváří optimální prostředí pro DNA polymerasu. Standardní reakční pufr obsahuje 10 mM Tris-HCl (pH 8,3–8,8), 50 mM KCl, 1,5 mM MgCl_2 . Případně může ještě obsahovat acetamid, albumin, želatinu nebo Tween 20.
- *Voda pro PCR*: používá se na doplnění PCR směsi na požadovaný objem. Nejvhodnější je voda o odporu 18 m Ω nebo voda pro injekce ČSL 4.

Při separaci DNA se jako nosiče používají agarosa (pro separaci molekul o velikosti 200–50 000 bp) a polyakrylamid (pro separaci molekul o velikosti 10–1 000 bp).

2.4.1. Interrepetitivní polymerázové řetězové reakce (rep-PCR)

Metoda interrepetitivní (repetitivní) polymerázové řetězové reakce (repeat sequence primed, interrepeat polymerase chain reaction, rep-PCR) je typizační metoda, při které se amplifikují sekvence v mezerníkových oblastech DNA, které oddělují v genomu různé typy repeticí: např. repetitivní mimogenový (mezerníkový) palindrom REP, konvenční enterobakteriální intergenová repetice ERIC, dlouhé roztroušené repetitivní elementy RepMP3, mozaikové repetitivní BOX-elementy a tandemově opakované polynukleotidové sekvence $(\text{GTG})_n$. Uvedené repetice jsou typické pro genomy mnoha bakteriálních druhů a v řadě případů i pro jednotlivé kmeny. K amplifikaci dojde, když se sekvence nacházejí v genomu v obrácené orientaci a v amplifikovatelné vzdálenosti. Výsledkem amplifikace je fingerprint obsahující více ampikonů různé velikosti (délky) [23].

2.4.2. Náhodná polymerázová řetězová reakce (RAPD)

Metoda náhodné polymerázové řetězové reakce (náhodná amplifikace polymorfní DNA, Randomly Amplified Polymorphic DNA – RAPD) se používá k detekci polymorfizmů DNA. Při amplifikaci se používá jeden primer o délce 8–12 nukleotidů s náhodnou sekvencí, za málo přísných podmínek hybridizace primeru. Dochází k nasednutí primeru na více místech DNA, která nejsou od sebe příliš vzdálená a umožňují hybridizaci primerů na protilehlé řetězce DNA s 3'-konci směřujícími proti sobě. Výsledkem amplifikace je vznik více amplikonů o různé délce. Výhodou této metody je skutečnost, že nemusí být známa sekvence DNA analyzovaného mikroorganismu [18].

3. Experimentální část

3.1. PCR – použité metody

Cílem této bakalářské práce bylo zhodnotit vliv korelačních koeficientů a algoritmů shlukové analýzy dat na taxonomické zařazení bakterií rodu *Lactobacillus*.

Zhodnocení algoritmů bylo prováděno s daty získanými z polymerázových řetězových reakcí rep-PCR s využitím primeru (GTG)₅ a RAPD s využitím primeru M13. Všechna data byla převzata z práce Turkové [25], proto je uvedeno pouze krátké shrnutí postupů a metod použitých pro přípravu a vyhodnocení PCR.

3.1.1. Příprava DNA

Bakteriální DNA byla izolována z bakteriálních buněk rodu *Lactobacillus* kultivovaných v médiu MRS. Po lyzi buněk byla provedena fenolová extrakce. DNA byla následně přečištěna vysrážením ethanolem.

3.1.2. Rep-PCR s DNA zástupců rodu *Lactobacillus* s využitím primeru (GTG)₅

PCR směs navržená pro typizaci bakterií rodu *Lactobacillus* byla připravena z komponent uvedených v tabulce. Pro analýzu DNA s využitím interrepetitivních sekvencí bylo možné použít primery (GTG)₅ [9]. Primer (GTG)₅ ve své sekvenci obsahuje pětinasobnou repetici sekvence GTG:

5' GTG GTG GTG GTG GTG 3'

PCR směs byla namíchána z níže uvedených komponent (Tabulka 3.1):

Tabulka 3.1: Komponenty PCR směsi

Komponenta	Objem (μl)
Voda pro PCR	8,0
10× koncentrovaný pufr bez MgCl ₂	2,5
MgCl ₂ (25 mM)	5,0
DMSO	0,5
dNTP směs (10 mM)	1,0
Primer (GTG) ₅ (10 pmol·μl ⁻¹)	2,0
<i>Taq</i> DNA polymeráza (1 U·μl ⁻¹)	1,0
DNA matrice	5,0
Celkem	25,0

Všechny složky PCR směsi byly důkladně promíchány, krátce centrifugovány a mikroz-kumavky byly umístěny do termocykleru. Jako negativní kontrola je možné použít PCR směs bez matricové DNA. Jako pozitivní kontrola lze použít DNA sbírkového nebo typového kmene o koncentraci 10 ng·μl⁻¹. Amplifikace probíhala za následujících podmínek (Tabulka 3.2):

Tabulka 3.2: Podmínky amplifikace s primerem (GTG)₅

Krok	Program	Podmínky	
		Teplota (°C)	Čas (min)
1	Denaturace	94	1
2	Hybridizace primeru	50	1
3	Syntéza řetězce DNA	68	1

PCR směs byla před prvním cyklem zahřáta na teplotu 94 °C po dobu 5 minut. V posledním cyklu byla doba syntézy řetězce DNA (teplota 68 °C) prodloužena na 10 minut. Amplifikace byla prováděna ve 33 cyklech. Po ukončení PCR byly vzorky analyzovány pomocí gelové elektroforézy na agarosovém gelu o hustotě 1,4 %. Na gel bylo naneseno 7 µl směsi, která vznikla smícháním 25 µl PCR produktu a 5 µl nanášecího pufru. Elektroforéza probíhala při napětí 30 V po dobu 16 hodin. Při detekci PCR produktu byl použit velikostní standard DNA (fragменты v rozmezí 100–1 500 bp; žebříček) v množství 5 µl.

3.1.3. RAPD-PCR s DNA zástupců rodu *Lactobacillus* s využitím primeru M13

Pro typizaci kmenů rodu *Lactobacillus* lze použít primer M13 navržený Rossettim et al.[18]:

5' GTA AAA CGA CGG CCA GT 3'

PCR směs navržená pro typizaci bakterií rodu *Lactobacillus* byla připravena z komponent uvedených v Tabulce 3.3 [13, 18]:

Tabulka 3.3: Komponenty PCR směsi

Komponenta	Objem (µl)
Voda pro PCR	7,5
10× koncentrovaný pufr bez MgCl ₂	2,5
MgCl ₂ (25 mM)	5,0
DMSO	0,5
dNTP směs (10 mM)	1,0
Primer M13 (10 pmol·µl ⁻¹)	2,0
<i>Taq</i> DNA polymerasa (1 U·µl ⁻¹)	1,5
DNA matrice	5,0
Celkem	25,0

Amplifikace probíhala za následujících podmínek (Tabulka 3.4):

Tabulka 3.4: Podmínky amplifikace s primerem M13

Krok	Program	Podmínky	
		Teplota (°C)	Čas (min)
1	Denaturace	94	1
2	Hybridizace primeru	45	1
3	Syntéza řetězce DNA	72	1

PCR směs byla před prvním cyklem zahřáta na teplotu 94 °C po dobu 5 minut. V posledním cyklu byla doba syntézy řetězce DNA (teplota 72 °C) prodloužena na 10 minut. Amplifikace byla provedena ve 40 cyklech. Po ukončení PCR byly vzorky analyzovány pomocí gelové elektroforézy na agarosovém gelu o hustotě 1,4 %. Pro gelovou elektroforézu byl použit velikostní standard s fragmenty DNA o definované velikosti (fragmenty DNA o velikosti 100–1 500 párů bazí, bp; DNA žebříček).

3.1.4. Gelová elektroforéza DNA

1. Byl připraven agarosový gel o požadované koncentraci v $0,5 \times$ TBE pufru; suspenze byla pečlivě rozvařena v mikrovlnné troubě, roztok se nechal vychladnout na teplotu asi 60 °C a bylo přidáno 100 μ l ethidiumbromidu. Roztok byl následně nalit do elektroforetické vaničky s hřebínkem a poté ponechán 30 min až 1 hodinu tuhnout. Před nanášením vzorků na gel byl hřebínek opatrně vyjmut.
2. Na podložce (proužek polyethylenu) nebo v mikrozkuhavce bylo smícháno 5 μ l DNA s 1,5 μ l nanášecího pufru ($6 \times$ koncentrovaný). Směs byla nanášena do komůrky gelu (ve směru zprava doleva).
3. Vedle DNA byl na gel nanesen také velikostní DNA standard. Gel byl vložen do elektroforetické vaničky v takové orientaci, aby záporně nabitá DNA migrovala k anodě. Vanička s gelem byla opatrně převrstvena $0,5 \times$ TBE pufrem do výšky 2–3 mm nad gel a poté byl zapnut elektrický zdroj ($1,5 \text{ V} \cdot \text{cm}^{-1}$ přes noc).
4. Separace byla ukončena v okamžiku, kdy nanášecí pufr doputoval do 2/3 délky agarosového gelu.
5. Po skončení elektroforézy byl gel po dobu nejméně 1 hod. obarvován ethidium bromidem ($0,5 \mu\text{g} \cdot \text{ml}^{-1}$).
6. Gel byl opláchnut v destilované vodě, umístěn na transiluminátor a vyhodnocen v UV světle o vlnové délce 305 nm.
7. Byla provedena fotografická dokumentace gelu pomocí digitálního fotoaparátu.

3.2. Vyhodnocení PCR

Fingerprinty získané pomocí rep-PCR s primerem (GTG)₅ a RAPD s primerem M13 byly porovnány pomocí programu BioNumerics (Applied Maths NV, Sint-Martens-Latem, Bel-

gie). Analyzovány byly fingerprinty druhů rodu *Lactobacillus*, konkrétně *L. casei*, *L. paracasei*, *L. zae*, *L. rhamnosus*, *L. plantarum*, *L. salivarium*, *L. fermentum*, *L. acidophilus* a *L. gasseri*. Tyto mikroorganismy byly získány z České sbírky mikroorganismů (CCM) (Brno, ČR) resp. Sbírký mlékařských mikroorganismů Laktoflora (CCDM) (Tábor, ČR). Kromě uvedených druhů byly analyzovány také fingerprinty laktobacilů izolovaných ze stolice plně kojených dětí, které byly získány od prof. Ing. V. Rady, CSc. (Česká zemědělská univerzita v Praze) a jsou uloženy ve sbírce Laktoflora. Uvedené kmeny jsou označeny prefixem RL.

Vyhodnocení bylo provedeno pro Jaccardův, Jeffreyho, Diceův a Ochiaiho koeficient, pro korelační koeficienty Pearson a Cosine a pro algoritmy shlukové analýzy dat UPGMA, Wardovu metodu, metodu Single Linkage a metodu Complete Linkage.

4. Výsledky

4.1. Shrnutí výsledků druhově specifické PCR

Při analýze algoritmů jsme vzali v úvahu výsledky druhově specifické PCR z diplomové práce Turkové [25]. Nejprve jsme prověřili základní algoritmy shlukové analýzy pro průměr z fingerprintů rep-PCR a RAPD. Ověřili jsme, zdali druhovou specifitu potvrzuje také shluková analýza pomocí UPGMA s Pearsonovým korelačním koeficientem, která bývá obvykle doporučována pro výpočet dendrogramů s druhy rodu *Lactobacillus*, a provedli jsme srovnání výsledků s výsledky vypočítanými pomocí různých koeficientů.

Zařazení mikroorganismů pomocí druhově specifické PCR popisuje Tabulka 4.1:

Tabulka 4.1: Druhové zařazení kmenů RL zjištěné pomocí druhově specifické PCR [25]

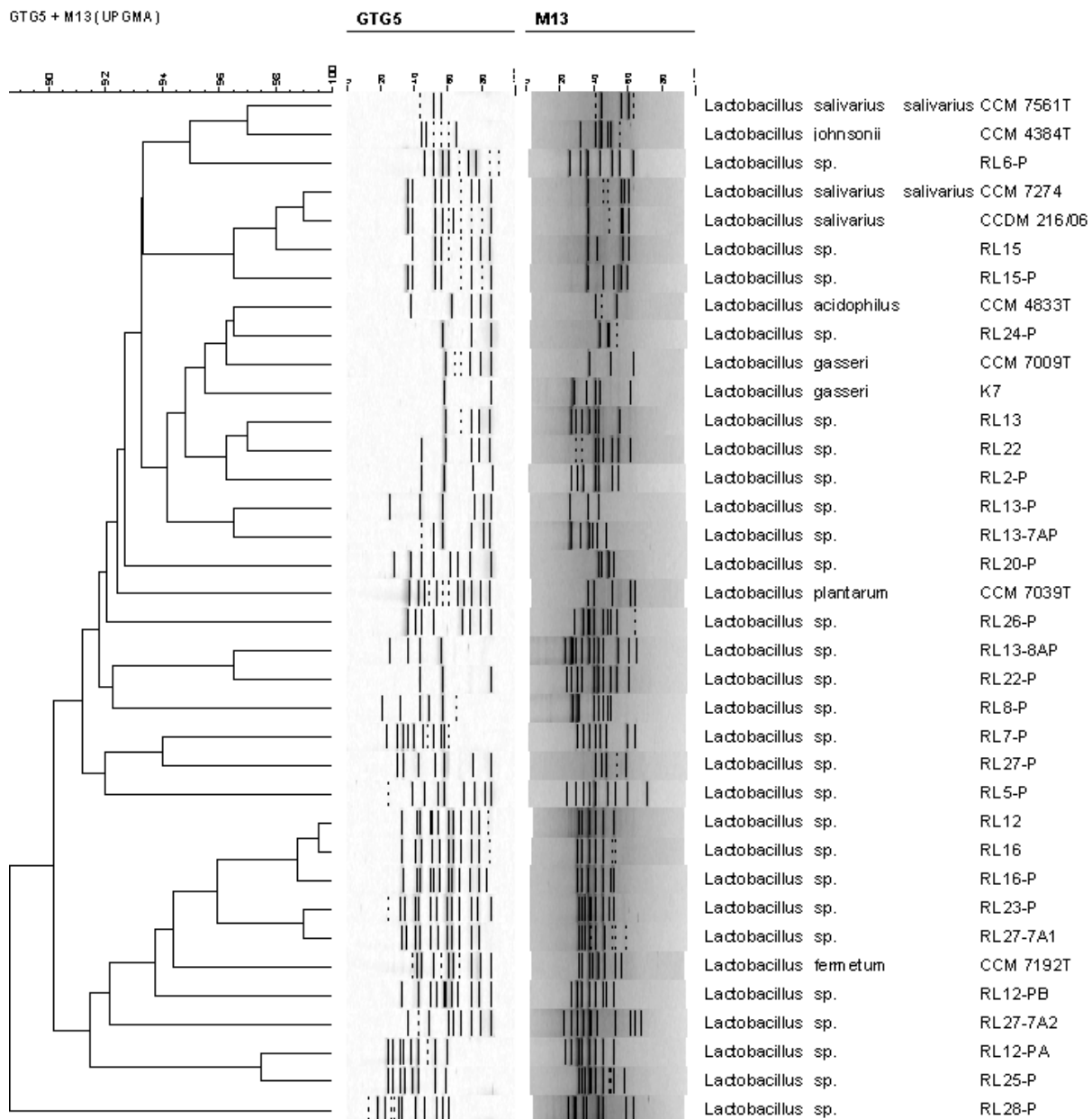
<i>Lactobacillus</i>					
<i>fermentum</i>	<i>johnsonii</i>	<i>paracasei</i>	<i>rhamnosus</i>	<i>salivarius</i>	<i>sp.</i>
RL12	RL2-P	RL4-P	RL1-P	RL15	RL6-P
RL12-PA	RL5-P	RL4-5A	RL3	RL15-P	RL7-P
RL12-PB	RL8-P	RL4-5B	RL3-P	RL16-P	RL20-P
RL16	RL13	RL10	RL9-P		RL26-P
RL16-P	RL13-P	RL10-P	RL11		RL27-P
RL23-P	RL13-7AP	RL10-5A	RL11-P		RL27-7A2
RL25-P	RL13-8AP	RL10-5B	RL14		RL28-P
RL27-7A1	RL22	RL19-P	RL14-P		
	RL22-P	RL21-P	RL17		
	RL24-P		RL17-P		
			RL18		
			RL18-P		
			RL29-P		
			RL30-P		

4.2. Porovnání algoritmů pro 1. sadu fingerprintů

Pomocí dendrogramu sestrojeného na základě shlukové analýzy dat UPGMA (Obrázek 4.1, Tabulka 4.2) s průměrem z fingerprintů pro rep-PCR a RAPD jsme ověřili úspěšnost algoritmu při analýze zadaných dat. Obdobné ověření bylo provedeno také pro Wardovu metodu (Obrázek 4.2, Tabulka 4.3) i metody Neighbor Joining (Obrázek 4.3, Tabulka 4.4), Single linkage (Obrázek 4.4, Tabulka 4.5) a Complete linkage (Obrázek 4.5, Tabulka 4.6).

4.2.1. Metoda UPGMA

Výsledky shlukové analýzy dat pomocí metody UPGMA jsou uvedeny na Obr. 4.1 a v Tabulce 4.2.



Obrázek 4.1: Dendrogram sestavený na základě shlukové analýzy UPGMA s průměrem z fingerprintů pro rep-PCR a RAPD

4.2. POROVNÁNÍ ALGORITMŮ PRO 1. SADU FINGERPRINTŮ

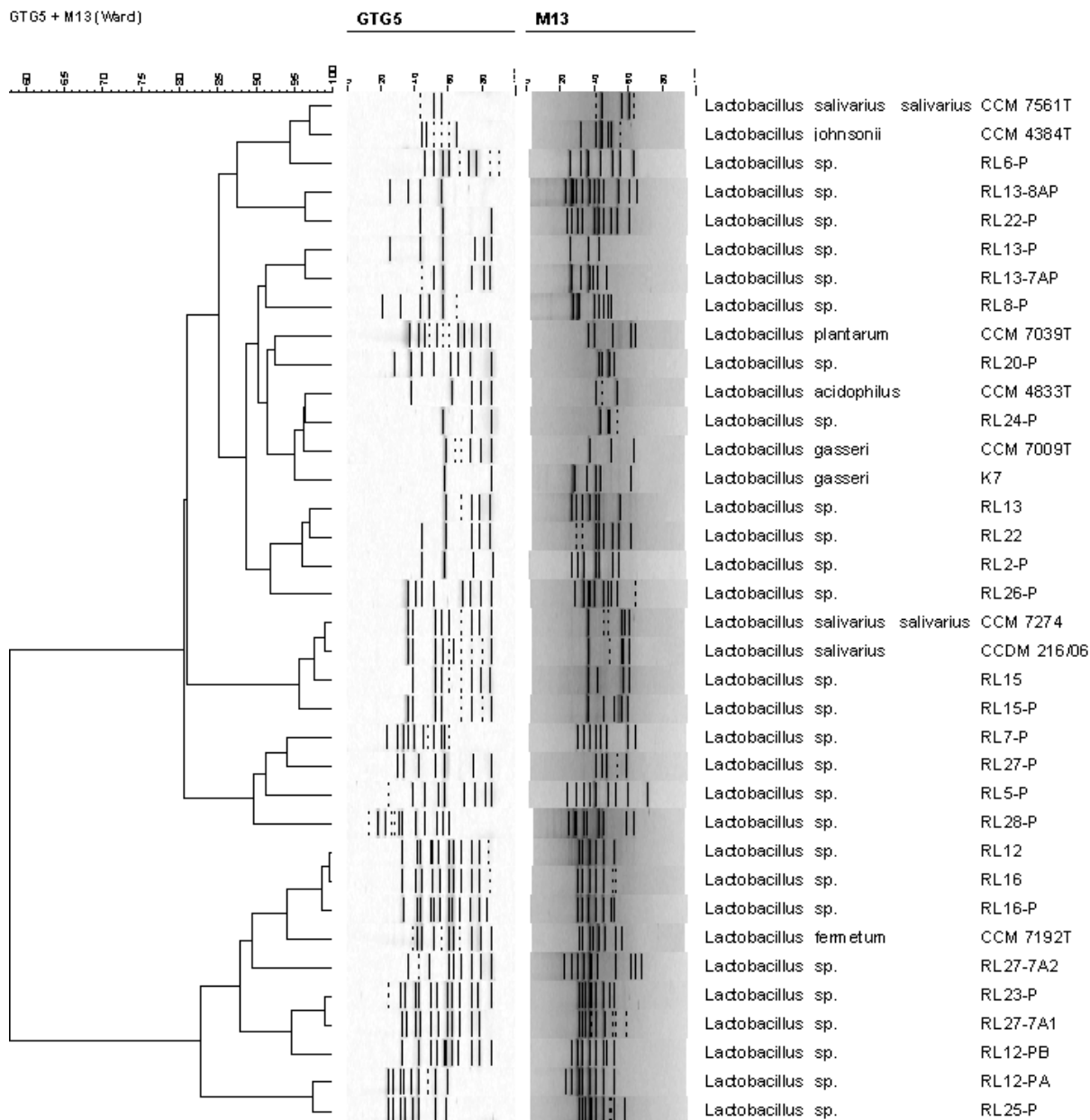
Tabulka 4.2: Rozlišení kmenů analyzovaných pomocí algoritmu UPGMA nad průměrem z fingerprintů pro rep-PCR a M13. Hladina podobnosti **91 %** a **93 %**.

Skup. (91 %)	Podsk. (93 %)	Typový/sbírkový kmen <i>Lactobacillus</i>	RL kmen	Zařazení <i>Lactobacillus</i>
I	I	<i>salivarius</i> ssp. <i>salivarius</i> CCM 7561 ^T <i>salivarius</i> ssp. <i>salivarius</i> CCM 7274 <i>salivarius</i> CCDM 216/06	RL15 RL15-P	<i>salivarius</i>
		<i>johnsonii</i> CCM 4384 ^T	RL2-P RL13 RL13-P RL13-7AP RL22 RL24-P	<i>johnsonii</i>
		<i>acidophilus</i> CCM 4833 ^T <i>gasseri</i> CCM 7009 ^T <i>gasseri</i> K7	RL6-P	sp.
	II		RL20-P	sp.
	III	<i>plantarum</i> CCM 7039 ^T		
	IV		RL26-P	sp.
	V		RL13-8AP RL22-P	<i>johnsonii</i>
	VI		RL8-P	<i>johnsonii</i>
	VII		RL7-P RL27-P	sp.
	VIII		RL5-P	<i>johnsonii</i>
II	IX	<i>fermentum</i> CCM 7192 ^T	RL12 RL12-PB RL16 RL16-P RL23-P RL27-7A1	<i>fermentum</i>
			RL27-7A2	sp.
	XI		RL12-PA RL25-P	<i>fermentum</i>
III	XII		RL28-P	sp.

Z Tabulky 4.2 vyplývá, že se nad hladinou podobnosti 91 % vytvořily pomocí UPGMA 3 skupiny a nad hladinou podobnosti 93 % 12 skupin. První skupina zahrnuje kmeny druhu *L. johnsonii*, *L. salivarius*, *L. gasseri* a *L. plantarum*, druhá skupina potom druh *L. fermentum*. Rozdělení s nižší hladinou významnosti bylo použito ve snaze získat podrobnější zařazení jednotlivých kmenů do druhů – zvláště rozdělení první skupiny na jednotlivé kmeny. Přesto nebyly některé z RL kmenů zařazeny do skupiny s odpovídajícím typovým/sbírkovým kmenem.

4.2.2. Wardova metoda

Výsledky shlukové analýzy dat pomocí Wardovy metody jsou uvedeny na Obr. 4.2 a v Tabulce 4.3.



Obrázek 4.2: Dendrogram sestavený na základě shlukové analýzy Ward s průměrem z fingerprintů pro rep-PCR a RAPD

4.2. POROVNÁNÍ ALGORITMŮ PRO 1. SADU FINGERPRINTŮ

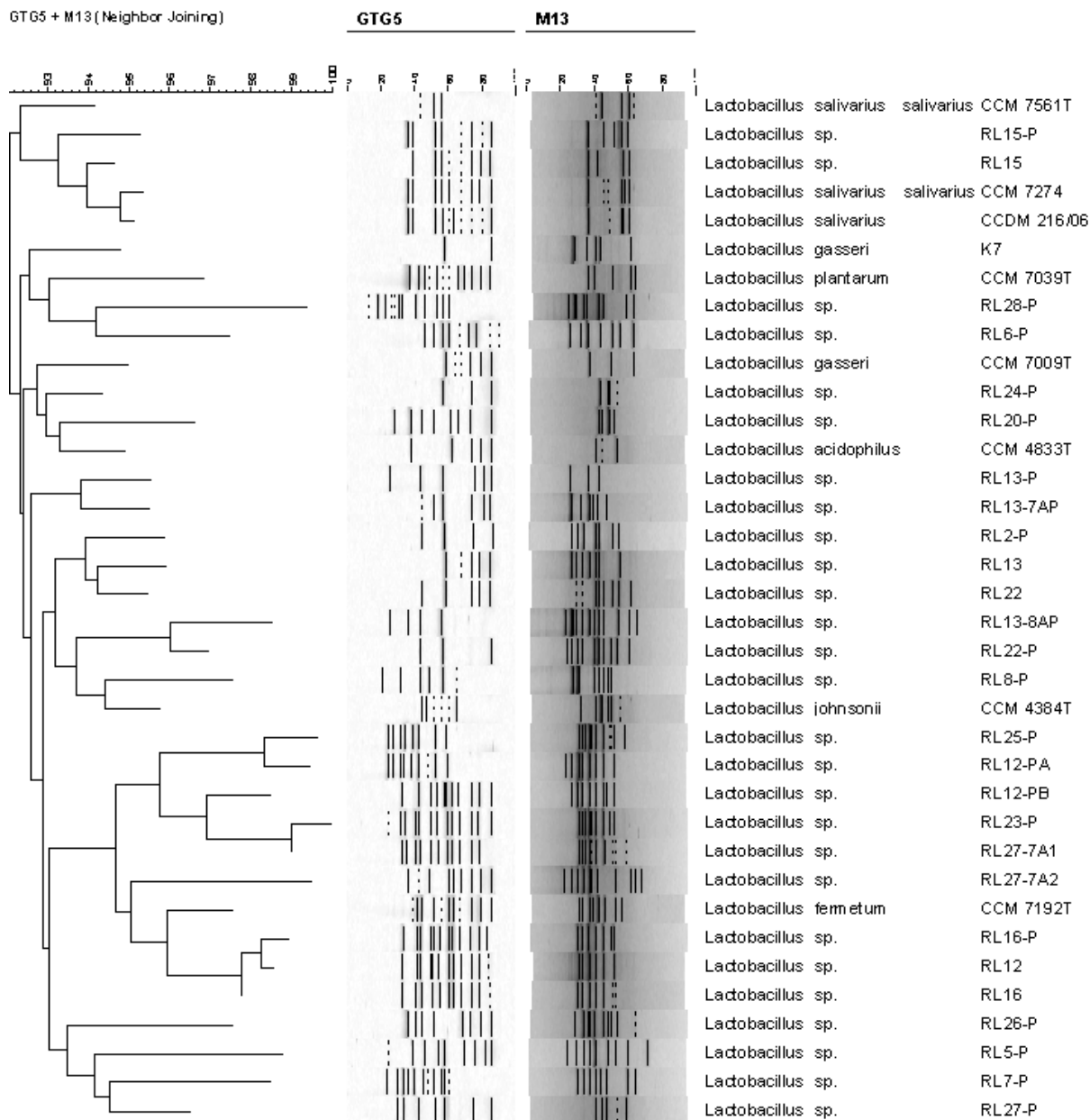
Tabulka 4.3: Rozlišení kmenů analyzovaných pomocí algoritmu Ward nad průměrem z fingerprintů pro rep-PCR a M13. Hladina podobnosti 80 % a 82 %.

Skup. (80 %)	Podsk. (82 %)	Typový/sbírkový kmen <i>Lactobacillus</i>	RL kmen	Zařazení <i>Lactobacillus</i>
I	I	<i>salivarius</i> ssp. <i>salivarius</i> CCM 7561 ^T		
		<i>johnsonii</i> CCM 4384 ^T	RL2-P RL8-P RL13 RL13-P RL13-7AP RL13-8AP RL22 RL22-P RL24-P	<i>johnsonii</i>
		<i>plantarum</i> CCM 7039 ^T <i>acidophilus</i> CCM 4833 ^T <i>gasseri</i> CCM 7009 ^T <i>gasseri</i> K7	RL6-P RL20-P RL26-P	sp.
	II	<i>salivarius</i> ssp. <i>salivarius</i> CCM 7274 <i>salivarius</i> CCDM 216/06	RL15 RL15-P	
	III		RL5-P	<i>johnsonii</i>
			RL7-P RL27-P RL28-P	sp.
II	IV	<i>fermentum</i> CCM 7192 ^T	RL12 RL12-PA RL12-PB RL16 RL16-P RL23-P RL25-P RL27-7A1	<i>fermentum</i>
			RL27-7A2	sp.

Pomocí algoritmu Ward se na hladině podobnosti 80 % vytvořily 2 skupiny a na hladině podobnosti 82 % 4 skupiny. V podrobnějším dělení nebyl jeden z RL kmenů (RL-5P) zařazen do skupiny s odpovídajícím typovým/sbírkovým kmenem (*L. johnsonii*), ale do skupiny bez typového/sbírkového kmene. Přesto lze říci, že Wardův algoritmus provedl rozdělení do podskupin o něco lépe než algoritmus UPGMA – ve stejné skupině jsou téměř všechny RL kmeny *L. johnsonii* a všechny RL kmeny *L. fermentum*.

4.2.3. Metoda Neighbor Joining

Výsledky shlukové analýzy dat pomocí metody Neighbor Joining jsou uvedeny na Obr. 4.3 a v Tabulce 4.4.



Obrázek 4.3: Dendrogram sestrojený na základě shlukové analýzy Neighbor Joining s průměrem z fingerprintů pro rep-PCR a RAPD

4.2. POROVNÁNÍ ALGORITMŮ PRO 1. SADU FINGERPRINTŮ

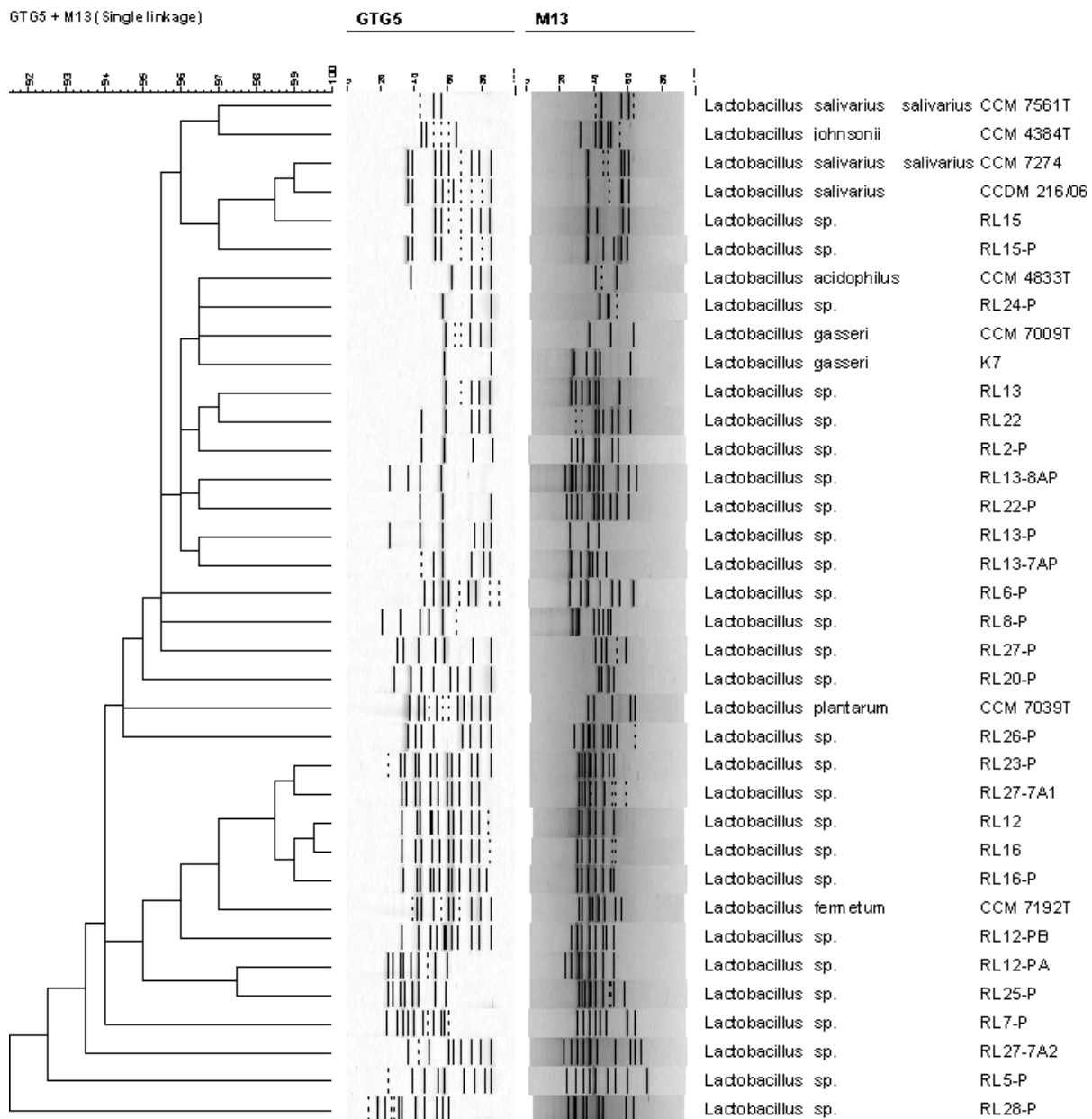
Tabulka 4.4: Rozlišení kmenů analyzovaných pomocí algoritmu Neighbor Joining nad průměrem z fingerprintů pro rep-PCR a M13. Hladina podobnosti **93 %**.

Skupina (93 %)	Typový/sbírkový kmen <i>Lactobacillus</i>	RL kmen	Zařazení <i>Lactobacillus</i>
I	<i>salivarius</i> ssp. <i>salivarius</i> CCM 7561 ^T		
II	<i>salivarius</i> ssp. <i>salivarius</i> CCM 7274 <i>salivarius</i> CCDM 216/06	RL15 RL15-P	<i>salivarius</i>
III	<i>gasseri</i> K7		
IV	<i>plantarum</i> CCM 7039 ^T	RL6-P RL28-P	sp.
V	<i>gasseri</i> CCM 7009 ^T		
VI	<i>acidophilus</i> CCM 4833 ^T	RL20-P	sp.
		RL24-P	<i>johnsonii</i>
VII		RL13-P RL13-7AP	<i>johnsonii</i>
VIII	<i>johnsonii</i> CCM 4384 ^T	RL2-P RL8-P RL13 RL13-8AP RL22 RL22-P	<i>johnsonii</i>
IX	<i>fermentum</i> CCM 7192 ^T	RL5-P RL12 RL12-PA RL12-PB RL16 RL16-P RL23-P RL25-P RL27-7A1	<i>fermentum</i>
		RL7-P RL26-P RL27-P RL27-7A2	sp.

Pomocí algoritmu Neighbor Joining se nad hladinou podobnosti 93 % vytvořilo 9 skupin (Tabulka 4.4). Několik RL kmenů (RL24-P, RL13-P a RL13-7AP) přitom nebylo zařazeno do skupiny s odpovídajícím typovým/sbírkovým kmenem (*L. johnsonii*). Pokud bychom použili nižší hladinu podobnosti, byly by do jedné skupiny zařazeny kmeny druhu *L. fermentum* i *L. johnsonii*.

4.2.4. Metoda Single linkage

Výsledky shlukové analýzy dat pomocí metody Single linkage jsou uvedeny na Obr. 4.4 a v Tabulce 4.5.



Obrázek 4.4: Dendrogram sestrojený na základě shlukové analýzy Single linkage s průměrem z fingerprintů pro rep-PCR a RAPD

4.2. POROVNÁNÍ ALGORITMŮ PRO 1. SADU FINGERPRINTŮ

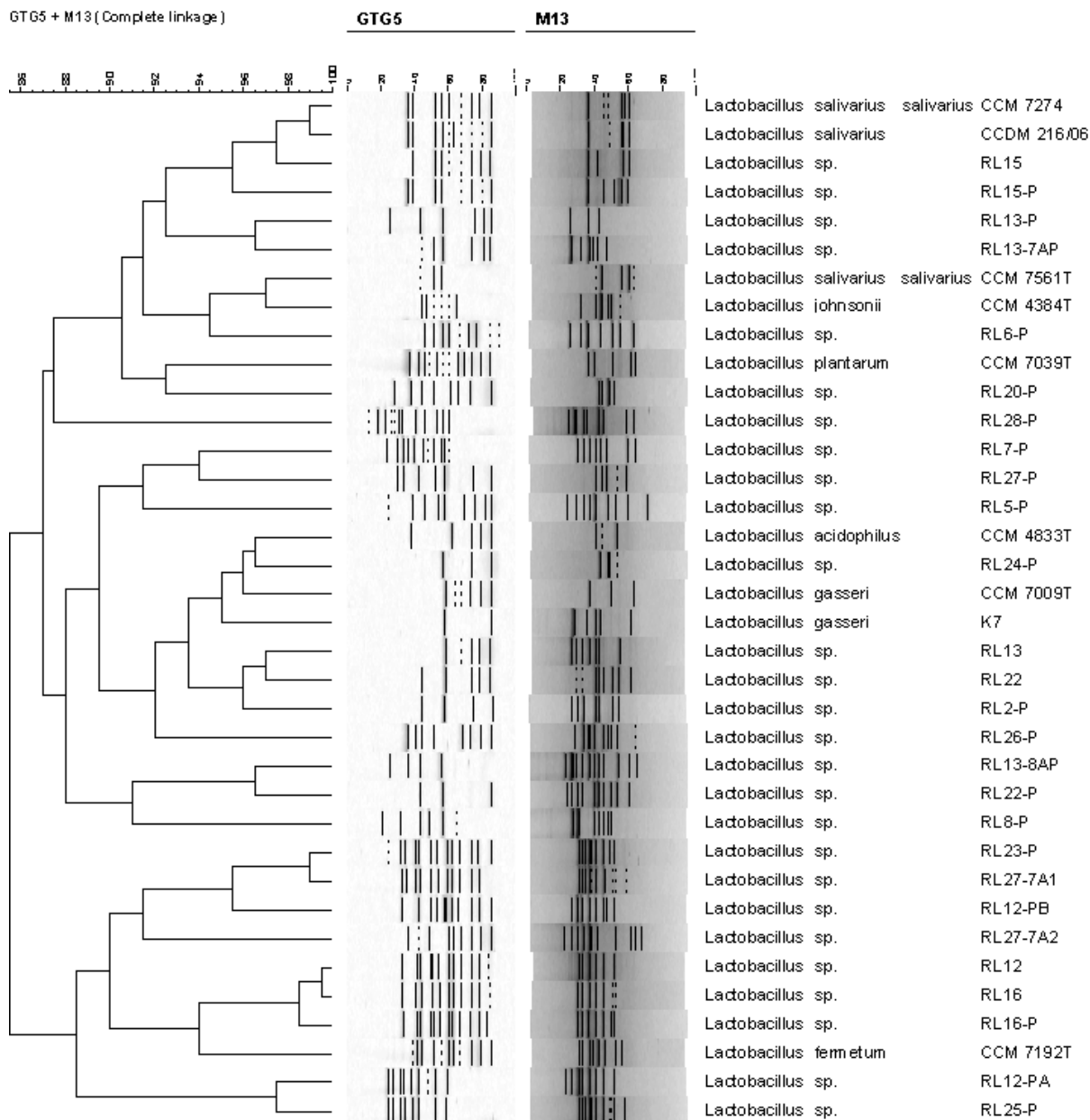
Tabulka 4.5: Rozlišení kmenů analyzovaných pomocí algoritmu Single linkage nad průměrem z fingerprintů pro rep-PCR a M13. Hladina podobnosti **94 %** a **95,4 %**.

Skup. (94 %)	Podsk. (95 %)	Typový/sbírkový kmen <i>Lactobacillus</i>	RL kmen	Zařazení <i>Lactobacillus</i>
I	I	<i>salivarius</i> ssp. <i>salivarius</i> CCM 7561 ^T <i>salivarius</i> ssp. <i>salivarius</i> CCM 7274 <i>salivarius</i> CCDM 216/06	RL15 RL15-P	<i>salivarius</i>
		<i>johnsonii</i> CCM 4384 ^T	RL2-P RL8-P RL13 RL13-P RL13-7AP RL13-8AP RL22 RL22-P RL24-P	<i>johnsonii</i>
		<i>acidophilus</i> CCM 4833 ^T <i>gasseri</i> K7 <i>gasseri</i> CCM 7009 ^T	RL6-P RL27-P	sp.
	II		RL20-P	sp.
	III	<i>plantarum</i> CCM 7039 ^T		
	IV		RL26-P	sp.
II	V	<i>fermentum</i> CCM 7192 ^T	RL12 RL12-PB RL16 RL16-P RL23-P RL27-7A1	<i>fermentum</i>
	VI		RL12-PA RL25-P	<i>fermentum</i>
III	VII		RL7-P	sp.
IV	VIII		RL27-7A2	sp.
V	IX		RL5-P	<i>johnsonii</i>
VI	X		RL28-P	sp.

Pomocí algoritmu Single linkage se na hladině podobnosti 94 % vytvořilo 6 skupin, na hladině podobnosti 95,4 % potom 10 skupin (Tabulka 4.5). Při podrobnějším dělení nebyly RL kmeny RL5-P, RL12-PA a RL25-P zařazeny do skupiny s odpovídajícím typovým/sbírkovým kmenem (*L. johnsonii* resp. *L. fermentum*). Ani při hrubším dělení nebyl RL kmen RL5-P zařazen do skupiny s typovým kmenem *L. johnsonii*. Přesto RL kmeny *L. salivarius* a *L. johnsonii* spadají do stejné skupiny.

4.2.5. Metoda Complete linkage

Výsledky shlukové analýzy dat pomocí metody Complete linkage jsou uvedeny na Obr. 4.5 a v Tabulce 4.6.



Obrázek 4.5: Dendrogram sestavený na základě shlukové analýzy Complete linkage s průměrem z fingerprintů pro rep-PCR a RAPD

4.2. POROVNÁNÍ ALGORITMŮ PRO 1. SADU FINGERPRINTŮ

Tabulka 4.6: Rozlišení kmenů analyzovaných pomocí algoritmu Complete linkage nad průměrem z fingerprintů pro rep-PCR a M13. Hladina podobnosti **87 %**.

Skupina (87 %)	Typový/sbírkový kmen <i>Lactobacillus</i>	RL kmen	Zařazení <i>Lactobacillus</i>
I	<i>salivarius</i> ssp. <i>salivarius</i> CCM 7561 ^T <i>salivarius</i> ssp. <i>salivarius</i> CCM 7274 <i>salivarius</i> CCDM 216/06	RL15 RL15-P	<i>salivarius</i>
	<i>johnsonii</i> CCM 4384 ^T	RL2-P RL5-P RL8-P RL13 RL13-P RL13-7AP RL13-8AP RL22 RL22-P RL24-P	<i>johnsonii</i>
	<i>plantarum</i> CCM 7039 ^T <i>acidophilus</i> CCM 4833 ^T <i>gasseri</i> K7 <i>gasseri</i> CCM 7009 ^T	RL6-P RL7-P RL20-P RL26-P RL27-P RL28-P	sp.
II	<i>fermentum</i> CCM 7192 ^T	RL12 RL12-PA RL12-PB RL16 RL16-P RL23-P RL25-P RL27-7A1	<i>fermentum</i>
		RL27-7A2	sp.

Pomocí algoritmu Complete linkage se nad hladinou podobnosti 87 % vytvořily pouze 2 skupiny (Tabulka 4.6). RL kmeny *L. salivarius* a *L. johnsonii* spadají do stejné skupiny, RL kmeny *L. fermentum* tvoří samostatnou skupinu.

4.2.6. Zhodnocení výsledků pro 1. sady fingerprintů

Pro zadaná data byly nejlepší výsledky vypočteny algoritmem Neighbor Joining, ačkoliv ani ten nevrací ideální výstup. Podobné taxonomické zařazení bylo vypočteno také pomocí algoritmů UPGMA a Ward. Naopak nejhorší výsledky byly vypočteny algoritmem Single linkage, což potvrzuje i v dendrogramu na první pohled patrný efekt zřetězení.

Zajímavé je zařazení typového kmene *L. salivarius* ssp. *salivarius* 7561^T. Tento kmen by kromě algoritmu Neighbor Joining zařadily do stejné skupiny jako kmen *L. johnsonii*

4.3. POROVNÁNÍ ALGORITMŮ PRO 2. SADU FINGERPRINTŮ

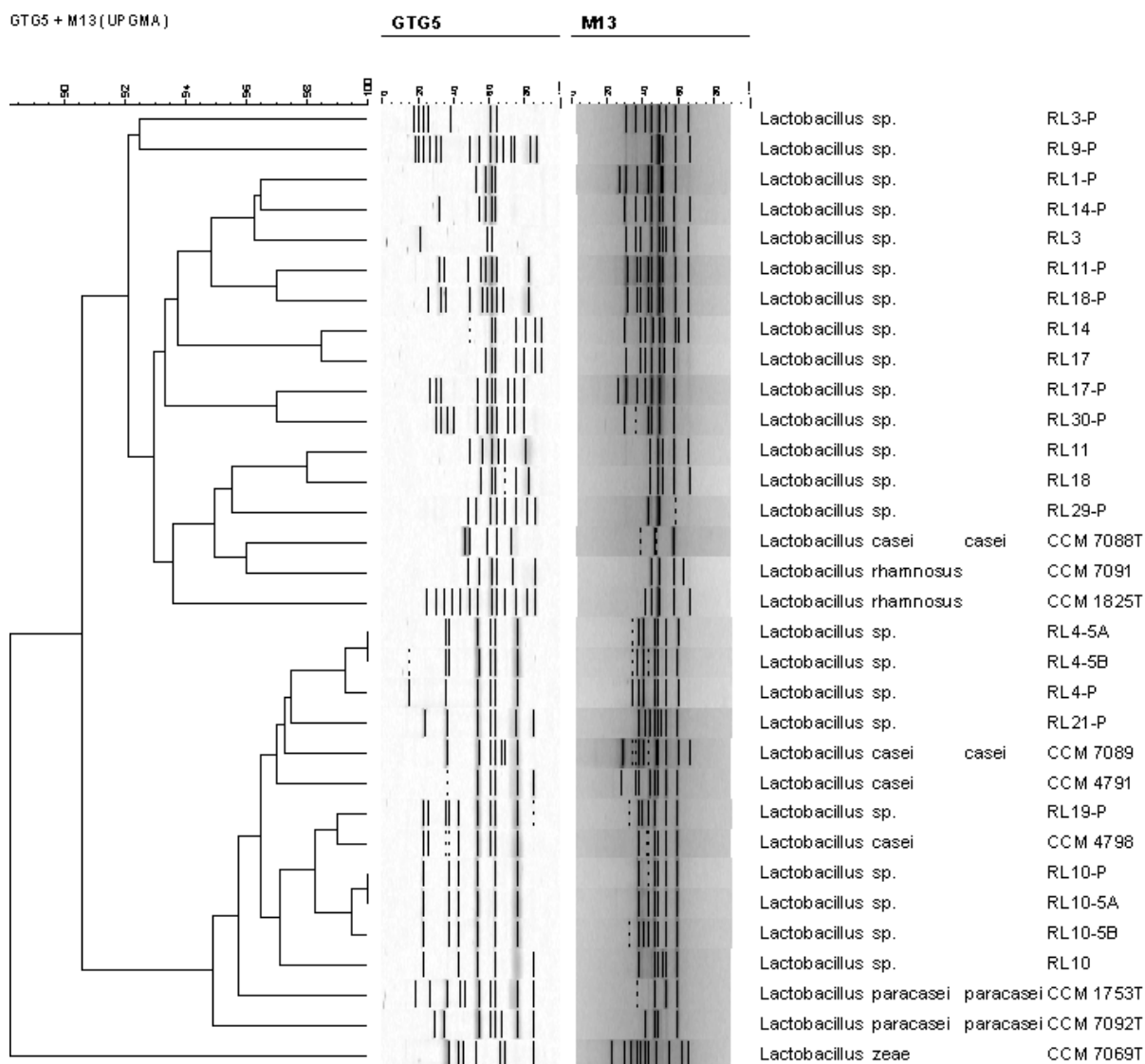
CCM 4384^T všechny použité algoritmy. Naopak algoritmus Neighbor Joining by jako jediný z hodnocených algoritmů zařadil všechny kmeny druhu *L. salivarius* do jedné skupiny (na hladině podobnosti 92,1 %).

4.3. Porovnání algoritmů pro 2. sadu fingerprintů

Protože na základě získaných výsledků nemůžeme rozhodnout, který algoritmus shlukové analýzy je nejlepší, podrobíme nejúspěšnější algoritmy analýze nad jinými daty.

4.3.1. Metoda UPGMA

Výsledky shlukové analýzy dat pomocí metody UPGMA jsou uvedeny na Obr. 4.6 a v Tabulce 4.7.



Obrázek 4.6: Dendrogram sestavený na základě shlukové analýzy UPGMA s průměrem z fingerprintů pro rep-PCR a RAPD (druhý gel)

4.3. POROVNÁNÍ ALGORITMŮ PRO 2. SADU FINGERPRINTŮ

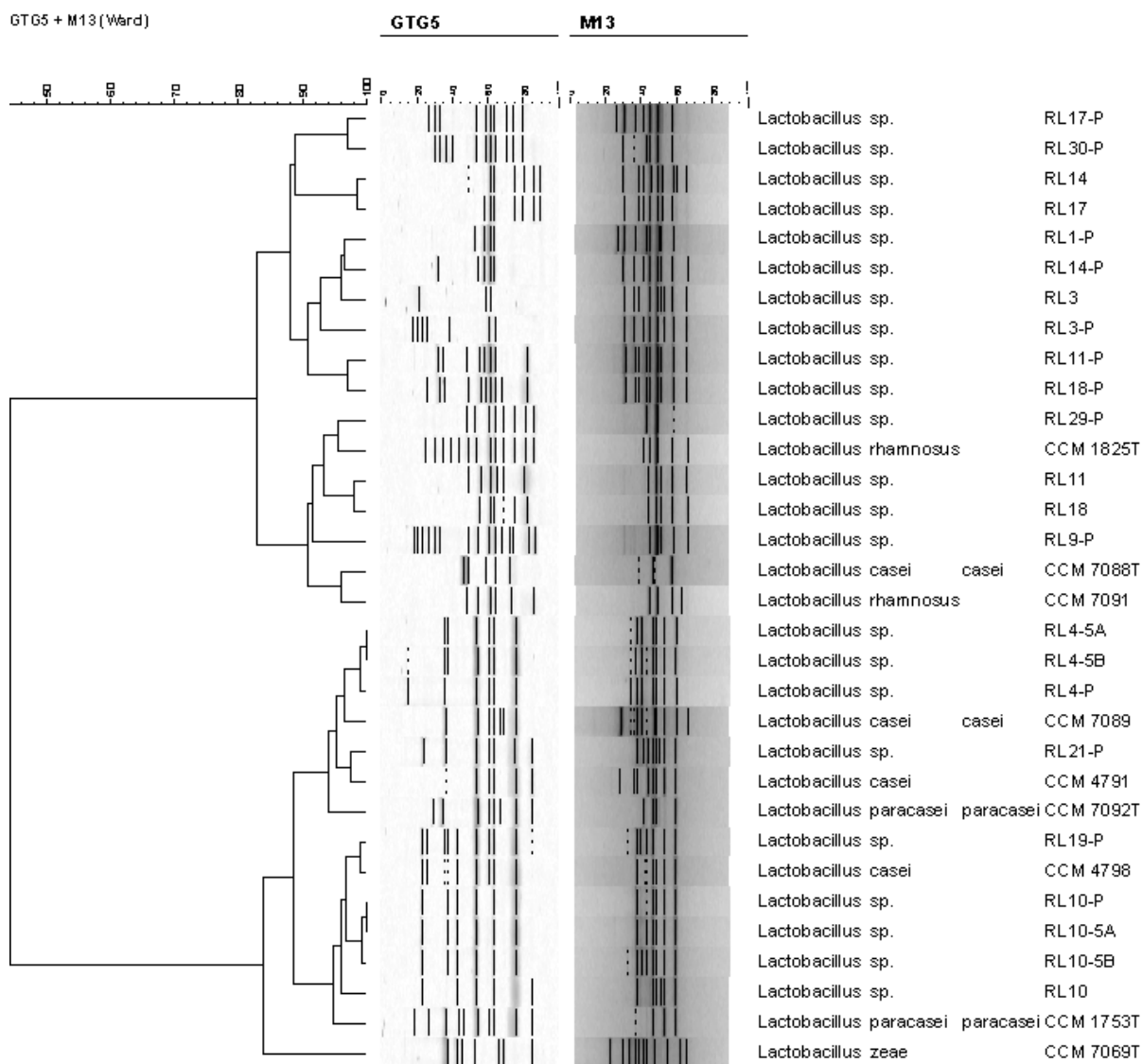
Tabulka 4.7: Rozlišení kmenů analyzovaných pomocí algoritmu UPGMA nad průměrem z fingerprintů pro rep-PCR a M13 (druhý gel). Hladina podobnosti **92 %**.

Skupina (92 %)	Typový/sbírkový kmen <i>Lactobacillus</i>	RL kmen	Zařazení <i>Lactobacillus</i>
I	<i>rhamnosus</i> CCM 7091 <i>rhamnosus</i> CCM 1825 ^T	RL1-P RL3 RL3-P RL9-P RL11 RL11-P RL14 RL14-P RL17 RL17-P RL18 RL18-P RL29-P RL30-P	<i>rhamnosus</i>
	<i>casei</i> ssp. <i>casei</i> CCM 7088 ^T		
II	<i>paracasei</i> ssp. <i>paracasei</i> CCM 1753 ^T <i>paracasei</i> ssp. <i>paracasei</i> CCM 7092 ^T	RL4-P RL4-5A RL4-5B RL10 RL10-P RL10-5A RL10-5B RL19-P RL21-P	<i>paracasei</i>
	<i>casei</i> ssp. <i>casei</i> CCM 7089 <i>casei</i> CCM 4791 <i>casei</i> CCM 4798		
III	<i>zeae</i> CCM 7069 ^T		

Algoritmus UPGMA rozdělil kmeny do 3 skupin (hladina pravděpodobnosti 92 %, Tabulka 4.7). První skupina zahrnuje kmeny druhu *L. rhamnosus* a kmen *L. casei* ssp. *casei* CCM 7088^T. Druhá skupina zahrnuje kmeny druhu *L. paracasei* a *L. casei*. Třetí skupinu tvoří pouze jediný kmen *L. zeae*.

4.3.2. Metoda Ward

Výsledky shlukové analýzy dat pomocí metody Ward jsou uvedeny na Obr. 4.7 a v Tabulce 4.8.



Obrázek 4.7: Dendrogram sestrojený na základě shlukové analýzy metodou Ward s průměrem z fingerprintů pro rep-PCR a RAPD (druhý gel)

4.3. POROVNÁNÍ ALGORITMŮ PRO 2. SADU FINGERPRINTŮ

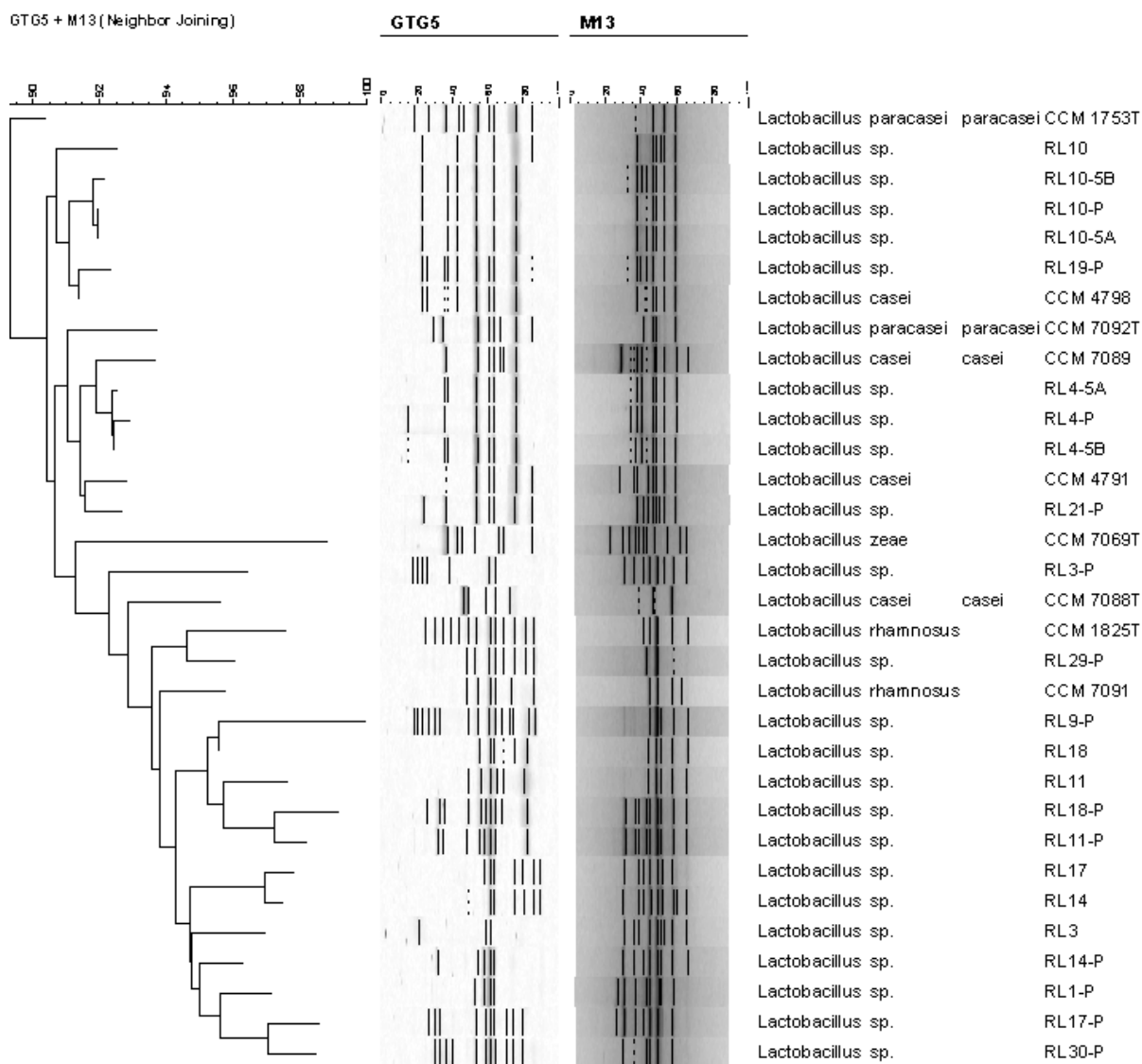
Tabulka 4.8: Rozlišení kmenů analyzovaných pomocí algoritmu Ward nad průměrem z fingerprintů pro rep-PCR a M13 (druhý gel). Hladina podobnosti **82 %**.

Skupina (82 %)	Typový/sbírkový kmen <i>Lactobacillus</i>	RL kmen	Zařazení <i>Lactobacillus</i>
I	<i>rhamnosus</i> CCM 7091 <i>rhamnosus</i> CCM 1825 ^T	RL1-P RL3 RL3-P RL9-P RL11 RL11-P RL14 RL14-P RL17 RL17-P RL18 RL18-P RL29-P RL30-P	<i>rhamnosus</i>
	<i>casei</i> ssp. <i>casei</i> CCM 7088 ^T		
II	<i>paracasei</i> ssp. <i>paracasei</i> CCM 1753 ^T <i>paracasei</i> ssp. <i>paracasei</i> CCM 7092 ^T	RL4-P RL4-5A RL4-5B RL10 RL10-P RL10-5A RL10-5B RL19-P RL21-P	<i>paracasei</i>
	<i>casei</i> ssp. <i>casei</i> CCM 7089 <i>casei</i> CCM 4791 <i>casei</i> CCM 4798 <i>zeae</i> CCM 7069 ^T		

K velmi podobným výsledkům jako pomocí algoritmu UPGMA jsme došli také pomocí algoritmu Ward. Algoritmus Ward rozdělil kmeny do dvou skupin, přitom první skupinu tvoří kmeny druhu *L. rhamnosus* a kmen *L. casei* ssp. *casei* CCM 7088^T, druhou skupinu tvoří kromě kmenů rodu *L. paracasei* také kmeny *L. casei* ssp. *casei* CCM 7089, *L. casei* CCM 4791, *L. casei* CCM 4798 a *L. zeae* CCM 7069^T.

4.3.3. Metoda Neighbor Joining

Výsledky shlukové analýzy dat pomocí metody Neighbor Joining jsou uvedeny na Obr. 4.8 a v Tabulce 4.9.



Obrázek 4.8: Dendrogram sestrojený na základě shlukové analýzy metodou Neighbor Joining s průměrem z fingerprintů pro rep-PCR a RAPD (druhý gel)

4.3. POROVNÁNÍ ALGORITMŮ PRO 2. SADU FINGERPRINTŮ

Tabulka 4.9: Rozlišení kmenů analyzovaných pomocí algoritmu Neighbor Joining nad průměrem z fingerprintů pro rep-PCR a M13 (druhý gel). Hladina podobnosti **91 %**.

Skupina (91 %)	Typový/sbírkový kmen <i>Lactobacillus</i>	RL kmen	Zařazení <i>Lactobacillus</i>
I	<i>paracasei</i> ssp. <i>paracasei</i> CCM 1753 ^T		
II		RL10	<i>paracasei</i>
III	<i>casei</i> CCM 4798	RL10-P	
		RL10-5A RL10-5B RL19-P	<i>paracasei</i>
IV	<i>paracasei</i> ssp. <i>paracasei</i> CCM 7092 ^T	RL4-P RL4-5A RL4-5B RL21-P	<i>paracasei</i>
	<i>casei</i> ssp. <i>casei</i> CCM 7089 <i>casei</i> CCM 4791		
V	<i>rhamnosus</i> CCM 1825 ^T <i>rhamnosus</i> CCM 7091	RL1-P RL3 RL3-P RL9-P	<i>rhamnosus</i>
	<i>zeae</i> CCM 7069 ^T <i>casei</i> ssp. <i>casei</i> CCM 7088 ^T		
VI		RL11 RL11-P RL14 RL14-P RL17 RL17-P RL18 RL18-P RL29-P RL30-P	<i>rhamnosus</i>

Algoritmus Neighbor Joining rozdělil kmeny do 6 skupin. První skupinu tvoří pouze kmen *L. paracasei* ssp. *paracasei* CCM 1753^T, druhou skupinu kmen druhu *L. paracasei*, třetí a čtvrtou skupinu tvoří kmeny druhu *L. casei* a *L. paracasei*. Pátá skupina zahrnuje kmeny druhu *L. rhamnosus*, *L. zeae* a *L. casei*, zatímco šestá skupina pouze kmeny druhu *L. rhamnosus*. I přes velké množství skupin tedy algoritmus zahrnul do stejné skupiny některé kmeny druhu *L. rhamnosus*, *L. zeae* i *L. casei*.

4.3.4. Zhodnocení výsledků shlukové analýzy 2. sady fingerprintů

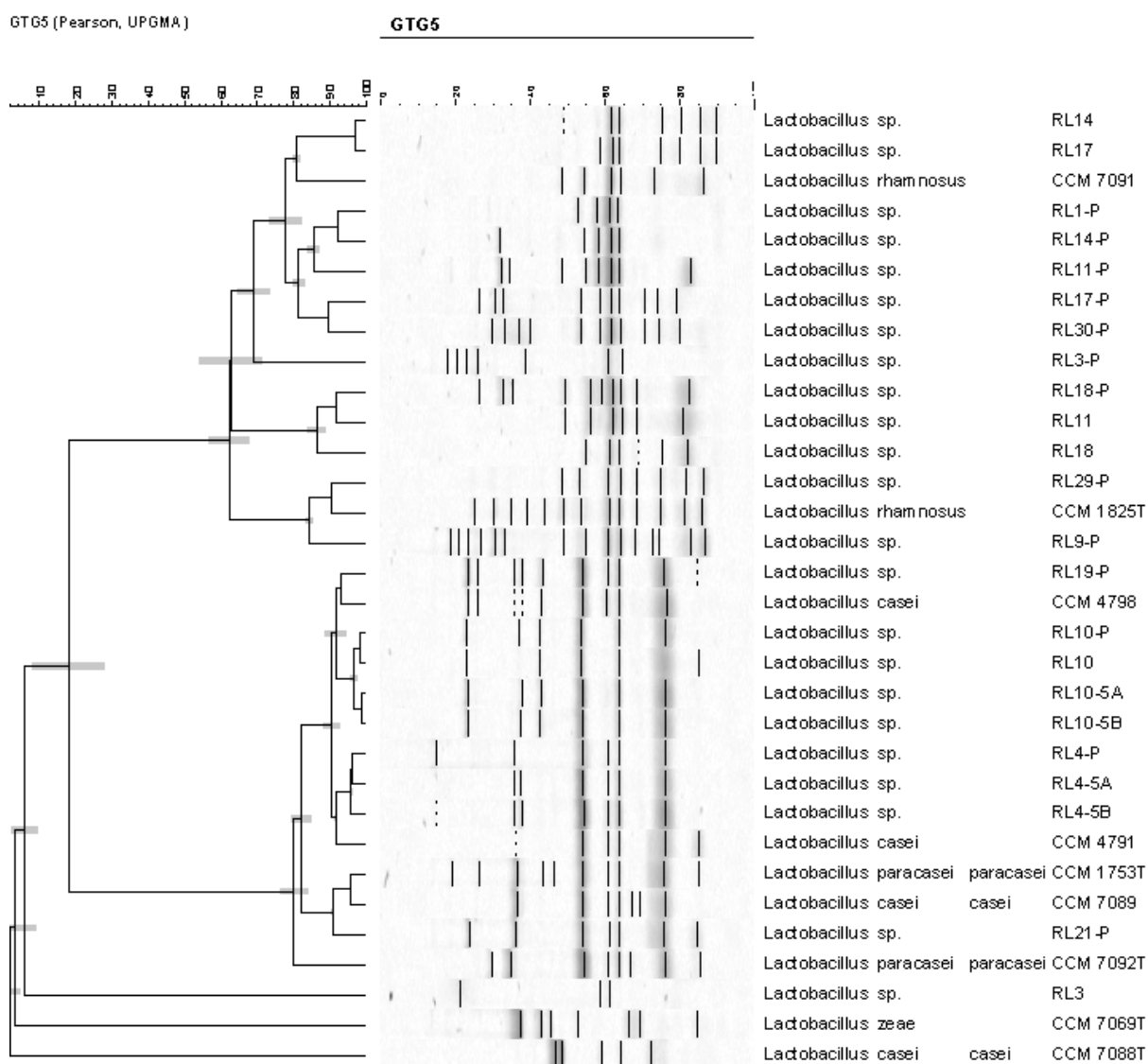
Nejlépeší výsledky vypočítaly algoritmy UPGMA a Ward, naopak algoritmus Neighbor Joining se pro shlukovou analýzu množiny zadaných dat ukázal jako málo vhodný.

4.4. Porovnání koeficientů míry podobnosti

Byla provedena analýza dendrogramů vygenerovaných pomocí algoritmu shlukové analýzy UPGMA pro různé koeficienty. Na základě zjištěných dat byly potom jednotlivé koeficienty porovnány a bylo vyhodnoceno, který koeficient je pro danou aplikaci nejvhodnější. Srovnání bylo provedeno nad druhou sadou kmenů. Použity byly fingerprinty získané pomocí rep-PCR.

4.4.1. Pearsonův korelační koeficient

Výsledky porovnání koeficientů míry podobnosti pomocí Pearsonova korelačního koeficientu jsou uvedeny na Obr. 4.9 a v Tabulce 4.10.



Obrázek 4.9: Dendrogram sestrojený na základě shlukové analýzy metodou UPGMA s Pearsonovým koeficientem

4.4. POROVNÁNÍ KOEFICIENTŮ MÍRY PODOBNOSTI

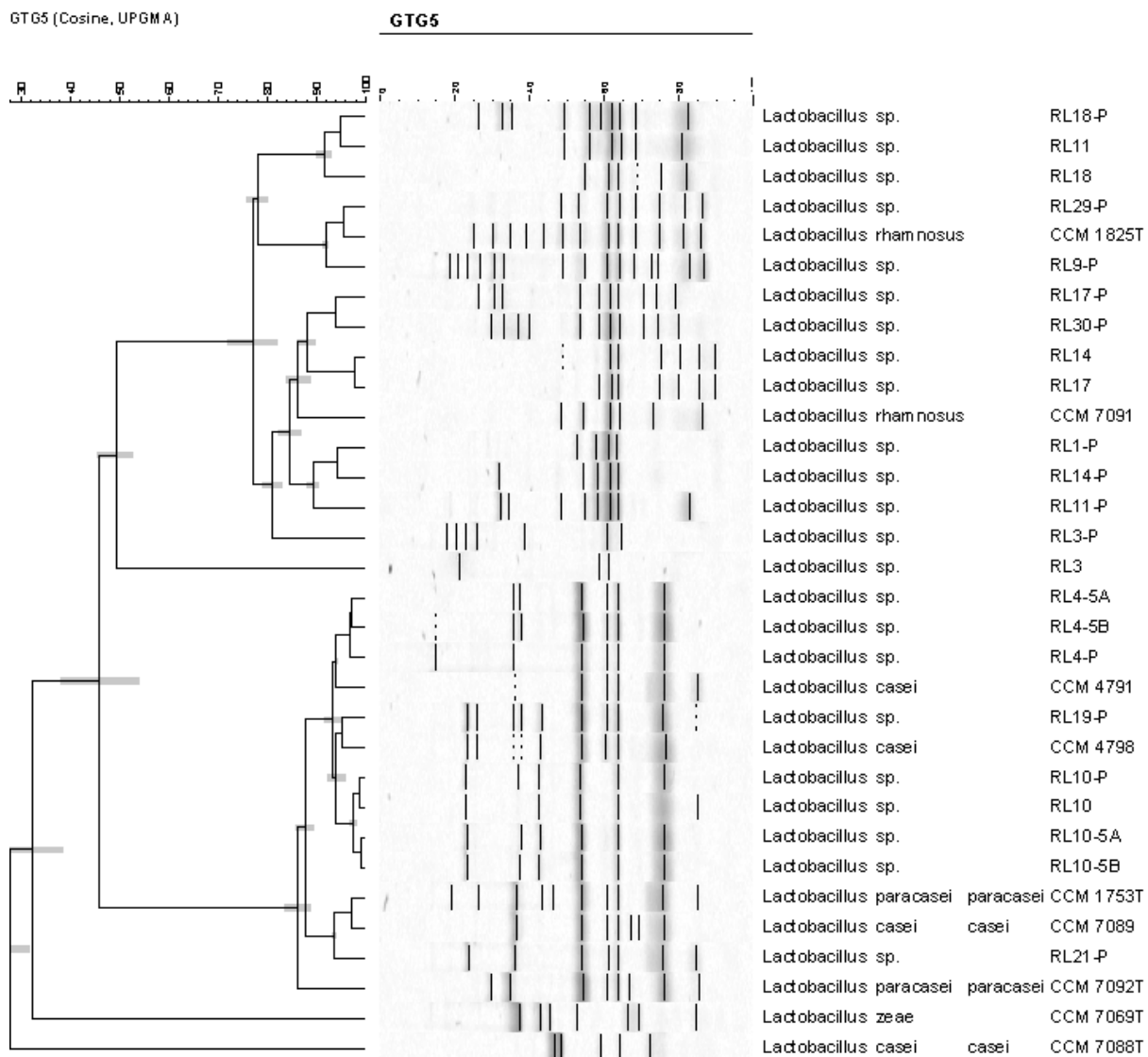
Tabulka 4.10: Rozlišení kmenů analyzovaných pomocí algoritmu UPGMA s Pearsonovým koeficientem. Hladina podobnosti **60 %**.

Skupina (60 %)	Typový/sbírkový kmen <i>Lactobacillus</i>	RL kmen	Zařazení <i>Lactobacillus</i>
I	<i>rhamnosus</i> CCM 7091 <i>rhamnosus</i> CCM 1825 ^T	RL1-P RL3-P RL9-P RL11 RL11-P RL14 RL14-P RL17 RL17-P RL18 RL18-P RL29-P RL30-P	<i>rhamnosus</i>
II	<i>paracasei</i> ssp. <i>paracasei</i> CCM 1753 ^T <i>paracasei</i> ssp. <i>paracasei</i> CCM 7092 ^T <i>casei</i> CCM 4798 <i>casei</i> CCM 4791 <i>casei</i> ssp. <i>casei</i> CCM 7089	RL4-P RL4-5A RL4-5B RL10 RL10-P RL10-5A RL10-5B RL19-P RL21-P	<i>paracasei</i>
III		RL3	<i>rhamnosus</i>
IV	<i>zeae</i> CCM 7069 ^T		
V	<i>casei</i> ssp. <i>casei</i> CCM 7088 ^T		

Pomocí Pearsonova koeficientu rozdělil algoritmus UPGMA kmeny do 5 skupin. První skupinu tvoří kmen druhu *L. rhamnosus*, druhou skupinu tvoří kmeny druhu *L. paracasei* a *L. casei*, třetí skupinu kmen druhu *L. rhamnosus*, čtvrtou skupinu *L. zeae* CCM 7069^T a pátou skupinu kmen *L. casei* ssp. *casei* CCM 7088^T. Algoritmus UPGMA s Pearsonovým koeficientem tedy nezařadil všechny kmeny druhu *L. rhamnosus* do stejné skupiny.

4.4.2. Korelační koeficient Cosine

Výsledky porovnání koeficientů míry podobnosti pomocí korelačního koeficientu Cosine jsou uvedeny na Obr. 4.10 a v Tabulce 4.11.



Obrázek 4.10: Dendrogram sestrojený na základě shlukové analýzy metodou UPGMA s koeficientem Cosine

4.4. POROVNÁNÍ KOEFICIENTŮ MÍRY PODOBNOSTI

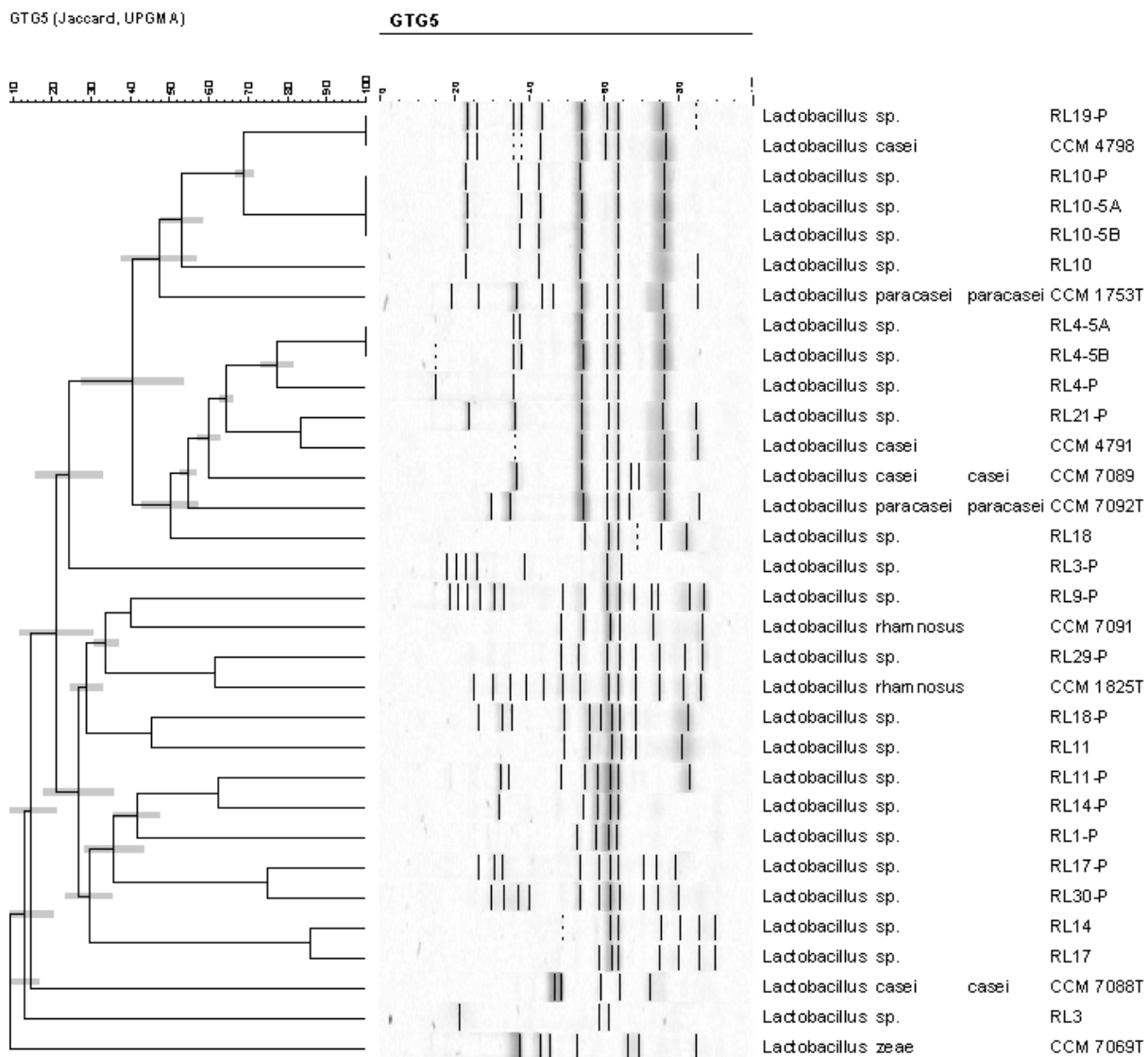
Tabulka 4.11: Rozlišení kmenů analyzovaných pomocí algoritmu UPGMA s koeficientem Cosine. Hladina podobnosti **48 %**.

Skupina (48 %)	Typový/sbírkový kmen <i>Lactobacillus</i>	RL kmen	Zařazení <i>Lactobacillus</i>
I	<i>rhamnosus</i> CCM 7091 <i>rhamnosus</i> CCM 1825 ^T	RL1-P RL3 RL3-P RL9-P RL11 RL11-P RL14 RL14-P RL17 RL17-P RL18 RL18-P RL29-P RL30-P	<i>rhamnosus</i>
II	<i>paracasei</i> ssp. <i>paracasei</i> CCM 1753 ^T <i>paracasei</i> ssp. <i>paracasei</i> CCM 7092 ^T <i>casei</i> CCM 4798 <i>casei</i> CCM 4791 <i>casei</i> ssp. <i>casei</i> CCM 7089	RL4-P RL4-5A RL4-5B RL10 RL10-P RL10-5A RL10-5B RL19-P RL21-P	<i>paracasei</i>
III	<i>zeae</i> CCM 7069 ^T		
IV	<i>casei</i> ssp. <i>casei</i> CCM 7088 ^T		

Pomocí koeficientu Cosine rozdělil algoritmus UPGMA kmeny do 4 skupin. První skupinu tvoří kmeny druhu *L. rhamnosus*, druhou skupinu kmeny druhu *L. paracasei* a *L. casei*, třetí skupinu kmen *L. zeae* 7069^T. Poslední skupinu tvoří kmen *L. casei* ssp. *casei* CCM 7088^T.

4.4.3. Koeficient Jaccard

Výsledky porovnání koeficientů míry podobnosti pomocí Jaccardova koeficientu jsou uvedeny na Obr. 4.11 a v Tabulce 4.12.



Obrázek 4.11: Dendrogram sestavený na základě shlukové analýzy metodou UPGMA s koeficientem Jaccard

4.4. POROVNÁNÍ KOEFICIENTŮ MÍRY PODOBNOSTI

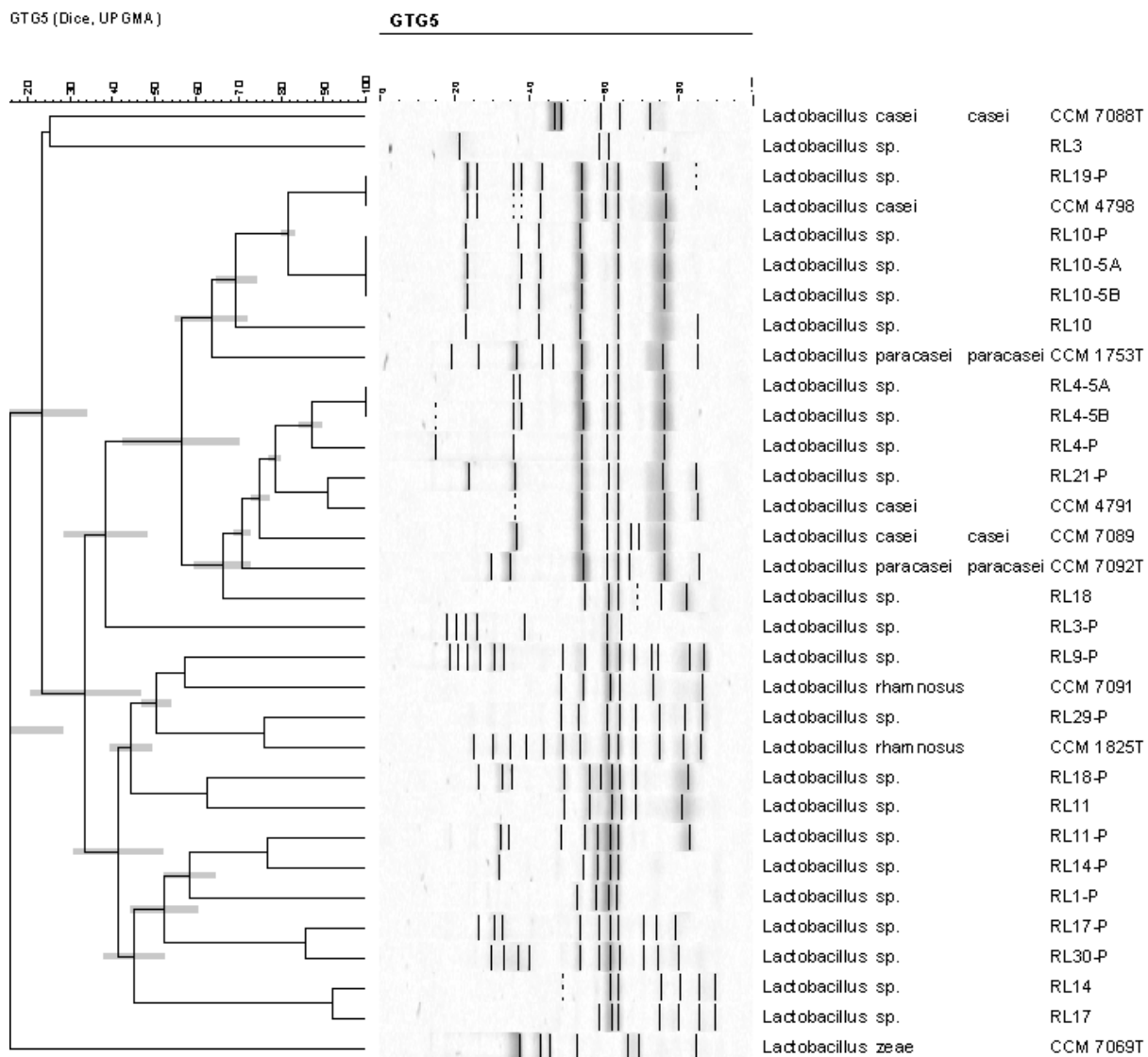
Tabulka 4.12: Rozlišení kmenů analyzovaných pomocí algoritmu UPGMA s koeficientem Jaccard. Hladina podobnosti **26 %**.

Skupina (26 %)	Typový/sbírkový kmen <i>Lactobacillus</i>	RL kmen	Zařazení <i>Lactobacillus</i>
I	<i>paracasei</i> ssp. <i>paracasei</i> CCM 1753 ^T <i>paracasei</i> ssp. <i>paracasei</i> CCM 7092 ^T <i>casei</i> CCM 4798 <i>casei</i> CCM 4791 <i>casei</i> ssp. <i>casei</i> CCM 7089	RL4-P RL4-5A RL4-5B RL10 RL10-P RL10-5A RL10-5B RL18 RL19-P RL21-P	<i>paracasei</i>
II		RL3-P	<i>rhamnosus</i>
III	<i>rhamnosus</i> CCM 7091 <i>rhamnosus</i> CCM 1825 ^T	RL1-P RL3-P RL9-P RL11 RL11-P RL14 RL14-P RL17 RL17-P RL18-P RL29-P RL30-P	<i>rhamnosus</i>
IV	<i>casei</i> ssp. <i>casei</i> CCM 7088 ^T		
V		RL3	<i>rhamnosus</i>
VI	<i>zeae</i> CCM 7069 ^T		

Pomocí Jaccardova koeficientu rozdělil algoritmus UPGMA kmeny do 6 skupin. První skupinu tvoří kmeny druhu *L. paracasei* a *L. casei*, druhou skupinu tvoří kmen druhu *L. rhamnosus*, třetí skupinu kmeny druhu *L. rhamnosus*, čtvrtou skupinu kmen *L. casei* ssp. *casei* CCM 7088^T, pátou skupinu opět kmen druhu *L. rhamnosus* a šestou skupinu kmen *L. zeae* 7069^T. Algoritmus UPGMA s Jaccardovým koeficientem rozdělil kmeny druhu *L. rhamnosus* do tří různých skupin.

4.4.4. Koeficient Dice

Výsledky porovnání koeficientů míry podobnosti pomocí korelačního koeficientu Dice jsou uvedeny na Obr. 4.12 a v Tabulce 4.13.



Obrázek 4.12: Dendrogram sestavený na základě shlukové analýzy metodou UPGMA s koeficientem Dice

4.4. POROVNÁNÍ KOEFICIENTŮ MÍRY PODOBNOSTI

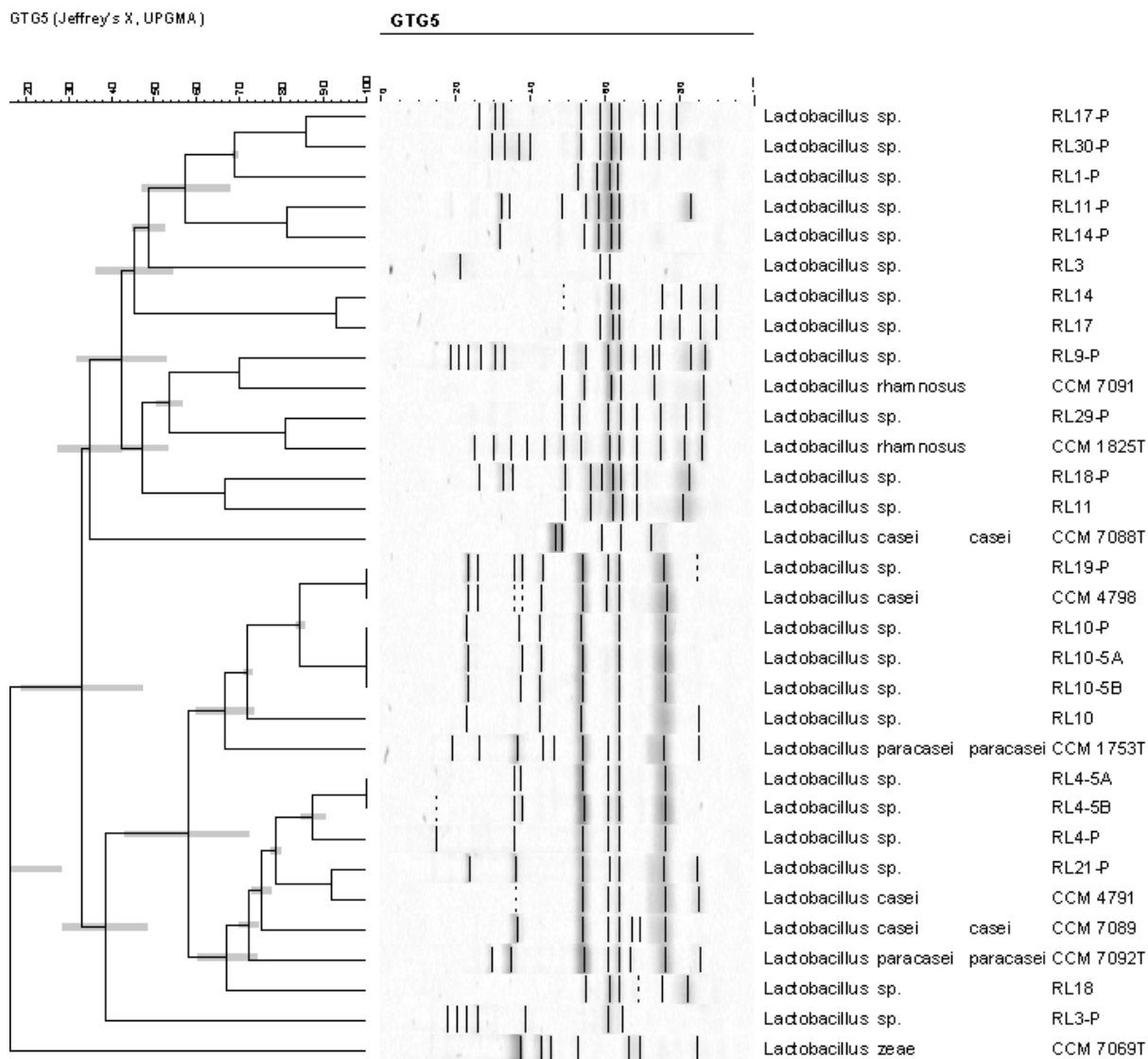
Tabulka 4.13: Rozlišení kmenů analyzovaných pomocí algoritmu UPGMA s koeficientem Dice. Hladina podobnosti **38 %**.

Skupina (38 %)	Typový/sbírkový kmen <i>Lactobacillus</i>	RL kmen	Zařazení <i>Lactobacillus</i>
I	<i>casei</i> ssp. <i>casei</i> CCM 7088 ^T	RL3	<i>rhamnosus</i>
II	<i>paracasei</i> ssp. <i>paracasei</i> CCM 1753 ^T <i>paracasei</i> ssp. <i>paracasei</i> CCM 7092 ^T <i>casei</i> CCM 4798 <i>casei</i> CCM 4791 <i>casei</i> ssp. <i>casei</i> CCM 7089	RL4-P RL4-5A RL4-5B RL10 RL10-P RL10-5A RL10-5B RL19-P RL21-P	<i>paracasei</i>
		RL3-P RL18	<i>rhamnosus</i>
III	<i>rhamnosus</i> CCM 7091 <i>rhamnosus</i> CCM 1825 ^T	RL1-P RL9-P RL11 RL11-P RL14 RL14-P RL17 RL17-P RL18-P RL29-P RL30-P	
IV	<i>zeae</i> CCM 7069 ^T		

Algoritmus UPGMA rozdělil pomocí koeficientu Dice kmeny do 4 skupin. První skupinu tvoří kmen *L. casei* ssp. *casei* CCM 7088^T a RL3 kmen druhu *L. rhamnosus*, druhou skupinu tvoří kmeny druhu *L. paracasei* a *L. casei* a třetí skupinu kmeny druhu *L. rhamnosus*. Čtvrtou skupinu tvoří kmen *L. zeae* 7069^T. Algoritmus UPGMA s koeficientem Dice rozdělil kmeny druhu *L. rhamnosus* do tří různých skupin, přičemž v jedné z těchto skupin jsou současně i kmeny druhu *L. paracasei* a *L. casei*.

4.4.5. Jeffreyho koeficient

Výsledky porovnání koeficientů míry podobnosti pomocí Pearsonova korelačního koeficientu jsou uvedeny na Obr. 4.13 a v Tabulce 4.14.



Obrázek 4.13: Dendrogram sestrojený na základě shlukové analýzy metodou UPGMA s Jeffreyho koeficientem

4.4. POROVNÁNÍ KOEFICIENTŮ MÍRY PODOBNOSTI

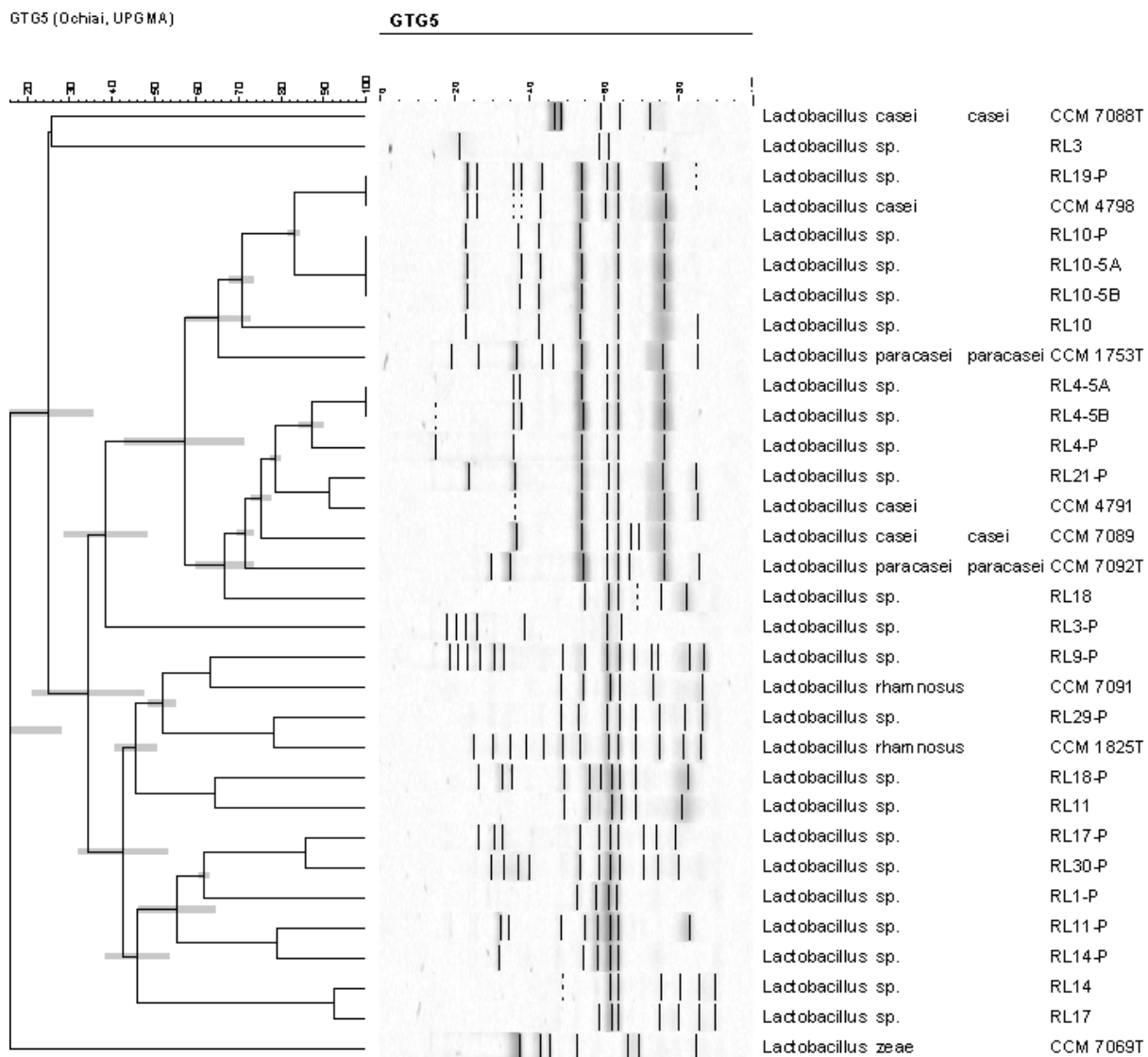
Tabulka 4.14: Rozlišení kmenů analyzovaných pomocí algoritmu UPGMA s Jeffreyho koeficientem. Hladina podobnosti **34 %**.

Skupina (34 %)	Typový/sbírkový kmen <i>Lactobacillus</i>	RL kmen	Zařazení <i>Lactobacillus</i>
I	<i>rhamnosus</i> CCM 7091 <i>rhamnosus</i> CCM 1825 ^T	RL1-P RL3 RL9-P RL11 RL11-P RL14 RL14-P RL17 RL17-P RL18-P RL29-P RL30-P	<i>rhamnosus</i>
	<i>casei</i> ssp. <i>casei</i> CCM 7088 ^T		
II	<i>paracasei</i> ssp. <i>paracasei</i> CCM 1753 ^T <i>paracasei</i> ssp. <i>paracasei</i> CCM 7092 ^T <i>casei</i> CCM 4798 <i>casei</i> CCM 4791 <i>casei</i> ssp. <i>casei</i> CCM 7089	RL4-P RL4-5A RL4-5B RL10 RL10-P RL10-5A RL10-5B RL19-P RL21-P	<i>paracasei</i>
		RL3-P RL18	<i>rhamnosus</i>
III	<i>zeae</i> CCM 7069 ^T		

Algoritmus UPGMA s Jeffreyho koeficientem rozdělil kmeny do 3 skupin. První skupinu tvoří kmeny druhu *L. rhamnosus* a kmen *L. casei* ssp. *casei* CCM 7088^T. Druhou skupinu tvoří kmeny druhu *L. paracasei*, *L. casei* i *L. rhamnosus*, třetí skupinu potom tvoří samostatně kmen *L. zeae* 7069^T. Algoritmus UPGMA s Jeffreyho koeficientem tedy rozdělil kmeny druhu *L. rhamnosus* do dvou různých skupin, přičemž ve druhé skupině jsou současně i kmenu druhu *L. paracasei*.

4.4.6. Koeficient Ochiai

Výsledky porovnání koeficientů míry podobnosti pomocí korelačního koeficientu Ochiai jsou uvedeny na Obr. 4.14 a v Tabulce 4.15.



Obrázek 4.14: Dendrogram sestavený na základě shlukové analýzy metodou UPGMA s koeficient Ochiai

4.4. POROVNÁNÍ KOEFICIENTŮ MÍRY PODOBNOSTI

Tabulka 4.15: Rozlišení kmenů analyzovaných pomocí algoritmu UPGMA s koeficientem Ochiai. Hladina podobnosti **32 %**.

Skupina (32 %)	Typový/sbírkový kmen <i>Lactobacillus</i>	RL kmen	Zařazení <i>Lactobacillus</i>
I	<i>casei</i> ssp. <i>casei</i> CCM 7088 ^T		
II		RL3	<i>rhamnosus</i>
III		RL3-P RL18 RL4-P RL4-5A RL4-5B RL10 RL10-P RL10-5A RL10-5B	<i>rhamnosus</i>
	<i>paracasei</i> ssp. <i>paracasei</i> CCM 1753 ^T <i>paracasei</i> ssp. <i>paracasei</i> CCM 7092 ^T <i>casei</i> CCM 4798 <i>casei</i> CCM 4791 <i>casei</i> ssp. <i>casei</i> CCM 7089	RL19-P RL21-P	<i>paracasei</i>
IV	<i>rhamnosus</i> CCM 7091 <i>rhamnosus</i> CCM 1825 ^T	RL1-P RL9-P RL11 RL11-P RL14 RL14-P RL17 RL17-P RL18-P RL29-P RL30-P	<i>rhamnosus</i>

Algoritmus UPGMA s koeficientem Ochiai rozdělil kmeny do 4 skupin. První skupinu tvoří kmen *L. casei* ssp. *casei* CCM 7088^T, druhou skupinu kmen RL3, který patří do druhu *L. rhamnosus*. Třetí skupinu tvoří současně kmeny druhu *L. paracasei*, *L. casei* i *L. rhamnosus* a čtvrtou skupinu kmeny druhu *L. rhamnosus*. Algoritmus UPGMA s koeficientem Ochiai tedy rozdělil kmeny druhu *L. rhamnosus* do tří různých skupin, přičemž třetí skupina obsahuje současně kmeny druhu *L. rhamnosus*, *L. paracasei* a *L. casei*.

4.4.7. Zhodnocení koeficientů

Zadaná data byla nejlépe vyhodnocena použitím koeficientu Cosine. Pomocí uvedeného koeficientu byly kmeny rozděleny do 4 skupin. Tento koeficient jako jediný zařadil všechny kmeny druhu *L. rhamnosus* do jedné skupiny. Správně zařadil také všechny kmeny skupiny *L. casei/paracasei* s výjimkou kmene *L. casei* ssp. *casei* CCM 7088.

5. Diskuse

Z uvedených výsledků vyplývá, že každá z metod poskytuje značně odlišné výsledky. Otázku, která z metod je tedy pro danou aplikaci nejvhodnější, tak nelze bezchybně zodpovědět. Jednoznačná odpověď totiž neexistuje. Analýza shluků nepatří mezi jednoduché výpočetní metody, ale metody založené na provádění celé řady kroků, které mohou být (a velmi často také jsou) závislé na provedení kroků předchozích. Proto je obecně nemožné předpokládat, která kombinace koeficientů nebo vzdáleností a shlukovacích metod poskytne nejlepší nebo nejspolehlivější výsledky [15]. První soubor dat analyzoval nejlépe algoritmus Neighbor Joining, který ovšem data druhého souboru analyzoval z porovnávaných algoritmů nejhůře. Průměrně podával nejlepší výsledky běžně doporučovaný algoritmus shlukové analýzy dat UPGMA a algoritmus Ward. Z porovnávaných koeficientů se umístil nejlépe korelační koeficient Cosine, který porazil i doporučovaný Pearsonův koeficient.

V knize Everitta a Dunna [5] nalezneme několik studií, které provádějí detailní srovnání jednotlivých metod. I z těchto výstupů je patrné, že použití různých shlukovacích metod poskytne i pro stejné koeficienty, resp. vzdálenosti, významně odlišné výsledky. Proto se často doporučuje provádět analýzu s několika shlukovacími metodami a jako správné označit pouze ty shluky, které jsou společné pro více metod.

Obecně platí, že žádná ze shlukovacích metod není nejlepší pro všechna data. Přitom metoda Single Linkage je obecně nejhorší (to jsme také prokázali při našem měření) – kvůli tendenci k efektu zřetězení. Obecně lze jako nejvhodnější naopak označit metodu UPGMA a Wardovu metodu [15], což jsme prokázali i v naší analýze. Co se týká koeficientů podobnosti, nejlepší výsledky prokázal koeficient Cosine, ovšem ani doporučovaný Pearsonův koeficient se neumístil špatně. Proto můžeme podpořit doporučení tvůrců programu BioNumerics a doporučujeme používat metodu UPGMA s Pearsonovým korelačním koeficientem.

Kromě výběru shlukovacích metod a binárních či korelačních koeficientů nebo koeficientů vzdálenosti je nutné vzít v úvahu také kvalitu dat a chyby zanesené během zpracování dat. Program BioNumerics sice zahrnuje funkci automatické detekce bandů, ale tato metoda je (zvláště pro zašuměné či neostře snímky) dosti nespolehlivá a často je tedy nutné jednotlivé bandy označit ručně. To s sebou ovšem nese velkou míru subjektivity – uživatel často označí ty bandy, které by někdo jiný považoval za šum, a naopak neoznačí ty, které skutečně existují, ale na fotografii gelu splývají s pozadím.

6. Závěr

Pro analýzu fingerprintů analyzovaného souboru různých druhů rodu *Lactobacillus* lze doporučit algoritmus shlukové analýzy dat UPGMA s Pearsonovým korelačním koeficientem. Přesnější závěry bude možné učinit po analýze širšího souboru dat zahrnujících více bakteriálních rodů a druhů.

Seznam použitých zdrojů

- [1] BERGER, B., PRIDMORE, R. D., BARRETTO, C., DELMAS-JULIEN, F., SCHREIBER, K., ARIGONI, F., BRUSSOW, H. Similarity and differences in the *Lactobacillus acidophilus* group identified by polyphasic analysis and comparative genomics. *J. Bacteriol.*, 2007, vol. 189, no. 4, s. 1311–1321.
- [2] BERNARDEAU, M., GUGÉUEN, M., VERNOUX, J. P. Beneficial lactobacilli in food and feed: long-term use, biodiversity and proposals for specific and realistic safety assessments. *FEMS Microbiol. Rev.*, 2006, vol. 30, s. 487–513.
- [3] CARRICO, J. A., PINTO, F. R., SIMAS, C., NUNES, S., SOUSA, N. G., FRAZAO, N., DE LENCASTRE, H., ALMEIDA, J. S. Assessment of Band-Based Similarity Coefficients for Automatic Type and Subtype Classification of Microbial Isolates Analyzed by Pulsed-Field Gel Electrophoresis. *J. Clin. Microbiol.*, 2005, vol. 43, no. 11, s. 5483–5490.
- [4] CODÓN, A. C., BENÍTEZ, T. Variability of the physiological features and of the nuclear and mitochondrial genomes of bakers' yeasts. *Systematic and Applied Microbiology*, 1995, vol. 18, s. 343–352.
- [5] DUNN, G., EVERITT, B. *An introduction to mathematical taxonomy*. 1st illustrated edition. CUP Archive, 1982. 152 s. ISBN 0-521-28388-4.
- [6] DUNN, G., EVERITT, B. *An introduction to mathematical taxonomy*. Courier Dover Publications, 2004. 152 s. ISBN 0-486-43587-3.
- [7] EHRMANN, M. A., VOGEL, R. F. Molecular taxonomy and genetics of sourdough lactic acid bacteria. *Trends Food Sci Technol*, 2005, vol. 16, s. 31–42.
- [8] *Polymerase chain reaction* [online]. 2010 [cit. 2010-05-16]. Encyclopædia Britannica. Dostupné z WWW: <<http://www.britannica.com/EBchecked/topic/468736/polymerase-chain-reaction>>
- [9] GEVERS, D., COHAN, F. M., LAWRENCE, J. G., SPRATT, B. G., COENYE, T., FEIL, E. J., STACKEBRANDT, E., VAN DE PEER, Y., VANDAMME, P., THOMPSON, F. L., SWINGS, J. Opinion: Re-evaluating prokaryotic species. *Nat. Rev. Microbiol.*, 2005, vol. 3, s. 733–739.
- [10] HEYNDRIKX, M., COHAN, F. M., LAWRENCE, J. G., SPRATT, B. G., COENYE, T., FEIL, E. J., STACKEBRANDT, E., VAN DE PEER, Y., VANDAMME, P., THOMPSON, F. L., SWINGS, J. Applicability of combined amplified ribosomal DNA restriction analysis (ARDRA) patterns in bacterial phyloeny and taxonomy. *J. Microbiol. Methods*, 1996, vol. 26, s. 247–259.
- [11] JOHNSON, E., BREHM-STECHER, B. *Electron Micrograph of the Meat Spoilage Bacterium Lactobacillus sake* [online]. 2003-09-22 [cit. 2010-05-16]. MicrobeLibrary.org. Dostupné z WWW: <<http://www.microbelibrary.org/Bacteria/details.asp?id=1471>>.

- [12] KIMMEL, A. R., OLIVER B. DNA microarrays, Part B: Databases and Statistics. In *Methods in enzymology*, vol. 411. Academic Press, 2006. 469 s. ISBN 0-12-182816-6.
- [13] KORDAČOVÁ, L. *Molekulární diagnostika bakterií mléčného kvašení rodu Lactobacillus*. Brno, 2007. 83 s. Diplomová práce. Masarykova univerzita.
- [14] LEGENDRE, P., LEGENDRE, L. *Numerical ecology*. 2nd illustrated edition. Elsevier, 1998. 853 s. ISBN 0-444-89250-8.
- [15] MARHOLD, K., SUDA, J. *Analýza multivariačních dat v taxonomii (fenetické metody)* [online]. Skriptum PřF UK. Praha : [s.n.], 2001, 2001 [cit. 2010-05-16]. Dostupné z WWW: <<http://botany.upol.cz/prezentace/duch/analyza.pdf>>.
- [16] OLEA, R. A. *Geostatistics for engineers and earth scientists*. Ilustrované vydání. Springer, 1999. 303 s. ISBN 0-7923-8523-3.
- [17] ROMESBURG, C. *Cluster Analysis for Researchers*. North Carolina: Lulu Press, 2004. 344 s. ISBN 1-4116-0617-5.
- [18] ROSSETTI, L., GIRAFFA, G. Rapid identification of dairy lactic bacteria by M13-generated, RAPD-PCR fingerprint databases. *J. Microbiol. Methods*, 2005, vol. 63, s. 135–144.
- [19] SALMINEN, S., VON WRIGHT, A., OUWEHAND, A. Lactic acid bacteria: Microbiological and functional aspects. 3rd revised and expanded edition. Marcel Dekker, Inc., 2004. 633 s. ISBN 0-8247-5332-1.
- [20] SAMARŽIJA, D., SIKORA, S., REDZEPOVIC, S., ANTUNAC, N., HAVRANEK, J. Application of RAPD analysis for identification of *Lactococcus lactis* subsp. *cremoris* strains isolated from artisanal cultures. *Microbiol. Res.*, 2002, vol. 157, s. 13–17.
- [21] SIMONSEN, M., MAILUND, T., Pedersen, N. S. Rapid Neighbour-Joining. In *Lecture Notes in Bioinformatics*, vol. 5251. Proceedings of the 8th international workshop on Algorithms in Bioinformatics. Springer-Verlag, 2008. s. 113–122.
- [22] SINGH, G. *Plant systematics: An integrated approach*. 2nd illustrated edition. Science Publishers, Inc., 2004. 561 s. ISBN 1-57808-351-6.
- [23] ŠMARDA, J., DOŠKAŘ, J., PANTŮČEK, R., RŮŽIČKOVÁ, V., KOPTÍKOVÁ, J. *Metody molekulární biologie*. 1. vydání. Masarykova univerzita, 2008. 188 s. ISBN 978-80-210-3841-7.
- [24] TAN, P. N., STEINBACH, M., KUMAR, V. *Introduction to Data Mining*. Addison-Wesley, 2006. 769 s. ISBN 0-321-32136-7.
- [25] TURKOVÁ, K. *Využití metod amplifikace DNA při identifikaci bakterií rodu Lactobacillus*. Brno, 2009. 106 s. Diplomová práce. Masarykova univerzita.
- [26] VENTURA, M., ELLI, M., RENIERO, R., ZINK, R. Molecular microbial analysis of *Bifidobacterium* isolates from different environments by the species-specific amplified ribosomal DNA restriction analysis (ARDRA). *FEMS Microbiol. Ecol.*, 2001, vol. 36, s. 113–121.

- [27] VERSALOVIC, J., KOEUTH, T., LUPSKI, J. R. Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes. *Nucl. Acids Res.*, 1991, vol. 19, no. 24, s. 6823–6831.

Seznam použitých zkratk a symbolů

BOX	mozaikové repetitivní BOX-elementy druhu <i>Streptococcus pneumoniae</i>
dNTP	2'-deoxynukleotid-5'-trifosfát
EDTA	kyselina ethylendiamin-N,N,N',N'-tetraethanová (octová)
ERIC	enterobakteriální opakující se intergenová shodná sekvence (Enterobacterial Repetitive Intergenic Consensus)
BMK	bakterie mléčného kvašení
OTU	taxonomická jednotka (Operational Taxonomic Unit)
PCR	polymerázová řetězová reakce (Polymerase Chain Reaction)
RAPD	náhodná polymerázová řetězové reakce, náhodná amplifikace polymorfní DNA (Random Amplified Polymorphic DNA)
rep-PCR	repetitivní polymerázová řetězová reakce (Repeat Sequence Primed Polymerase Chain Reaction)