VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ
FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

ÚSTAV BIOMEDICÍNSKÉHO INŽENÝRSTVÍ
DEPARTMENT OF BIOMEDICAL ENGINEERING

AUTOMATICKÁ GENOTYPIZACE BAKTERIÍ METODOU REP-PCR
AUTOMATIC GENOTYPING OF BACTERIA BY REP-PCR

DIPLOMOVÁ PRÁCE
MASTER'S THESIS

AUTOR PRÁCE
AUTHOR
Bc. Veronika Pelikánová

VEDOUCÍ PRÁCE
SUPERVISOR
Ing. Helena Škutková, Ph.D.

BRNO 2018

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ

Diplomová práce

magisterský navazující studijní obor **Biomedicínské inženýrství a bioinformatika**
Ústav biomedicínského inženýrství

Studentka: Bc. Veronika Pelikánová ***ID:*** 164991
Ročník: 2 ***Akademický rok:*** 2017/18

NÁZEV TÉMATU:

Automatická genotypizace bakterií metodou rep-PCR

POKYNY PRO VYPRACOVÁNÍ:

1) Seznamte se s metodou rep-PCR a jejím využitím pro stanovení příbuznosti bakterií. 2) Vypracujte literární rešerši metod pro elektroforetickou separaci DNA. Zaměřte se zejména na moderní metody kapilární elektroforézy na čipu, jejich princip, možné příčiny zkreslení a vzniku nepřesností. 3) Navrhněte algoritmus pro automatickou klasifikaci vzorků rep-PCR z čipové elektroforézy. 4) Realizujte dílčí části navrženého algoritmu zajišťující korekci pozice DNA fragmentů pro správné vyhodnocení podobnosti vzorků. 5) V programovém prostředí Matlab vytvořte program s GUI pro automatickou klasifikaci vzorků z rep-PCR formou dendrogramu. 6) Statisticky vyhodnoťte kvalitu realizovaného algoritmu na základě standardizovaných měření DNA markerů i reálných záznamech vybraných bakteriálních kmenů.

DOPORUČENÁ LITERATURA:

[1] OLIVE, Michael D. a Pamela BEAN. Principles and Applications of Methods for DNA-Based Typing of Microbial Organisms. Journal of Clinical Microbiology. 1999, 37(6), 1661-1669.

[2] VERSALOVIC, James, Thearith KOEUTH a James R. LUPSKI. Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes. Nucleic Acids Research. 1991, 19(24), 6823-6831.

Termín zadání: 5.2.2018 ***Termín odevzdání:*** 18.5.2018

Vedoucí práce: Ing. Helena Škutková, Ph.D.
Konzultant:

prof. Ing. Ivo Provazník, Ph.D.
předseda oborové rady

Abstrakt:

Diplomová práce se zabývá automatickou genotypizací bakterií metodou rep-PCR. V teoretické části jsou představeny různé metody typizace DNA, základní informace o elektroforéze a uvedeny moderní elektroforetické přístupy i s jejich problémy, zavádějící zkreslení dat. Za účelem automatické typizace je navržen program pro fylogenetickou klasifikaci vzorků dat z rep-PCR vhodný i obecně pro data z čipové kapilární elektroforézy. Program se skládá ze tří základních částí: digitalizace, úpravy pozic bandů a samotné fylogenetické klasifikace. Výsledkem programu je dendrogram znázorňující podobnost vzorků. K sestavenému programu je vytvořeno uživatelské rozhraní pro možné zavedení do praxe v Dětské nemocnici, na jejichž popud byl program sestavován. Na závěr je vyhodnocena úspěšnost programu na standardizovaných a reálných datech poskytnutých z Dětské nemocnice.

Klíčová slova: genotypizace, rep-PCR, čipová kapilární elektroforéza

Abstract:

This thesis deals with automatic bacteria genotyping by rep-PCR method. Its theoretical part presents various methods of DNA typing, basic information on electrophoresis and modern electrophoretic approaches, including their problems, misleading data distortion. In order to automate typing, there has been introduced a program for phylogenetic sample classification from rep-PCR, also applicable for data from chip capillary electrophoresis. The program consists of three main parts: digitization, bandmatching and clustering apparatus to bacterial type classification. The result of the algorithm is a phylogenetic tree, which indicates the cluster of sapmles according to bacterial type. The program has a graphical user interface for possible use in the Children's Hospital. Finally, the program is tested with data from the Children´s Hospital.

Keywords: genotyping, rep-PCR, chip capillary electrophoresis

PELIKÁNOVÁ, V. *Automatická genotypizace bakterií metodou rep-PCR.* Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2018. 60 s. Vedoucí diplomové práce Ing. Helena Škutková, Ph.D.

Prohlášení autora o původnosti díla:

Prohlašuji, že svou diplomovou práci na téma Automatická genotypizace bakterií metodou rep-PCR jsem vypracovala samostatně pod vedením vedoucí diplomové práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu použité literatury.

Jako autorka uvedené diplomové práce dále prohlašuji, že v souvislosti s jejím vytvořením jsem neporušila autorská práva třetích osob, zejména jsem nezasáhla nedovoleným způsobem do cizích autorských práv osobnostních a/nebo majetkových. Jsem si plně vědoma následků porušení ustanovení § 11 a následujícího zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů, ve znění pozdějších předpisů, včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

V Brně dne ……

Podpis autora

Poděkování:

Chtěla bych poděkovat Ing. Heleně Škutkové Ph.D. za vedení mé diplomové práce, praktické rady, odborný dohled, cenné komentáře, vstřícné jednání a čas, který mi věnovala. Bylo mi ctí s Vámi spolupracovat. Neméně velký dík patří celé rodině, která mě podporovala nejen při psaní diplomové práce, ale v průběhu celého studia. Díky za Vaši trpělivost, shovívavost, zázemí a podporu. Tato práce by bez toho všeho nemohla vzniknout.

Obsah

Úvod

V dnešní době, 90 let po objevení penicilinu, znovu narůstá bakteriálních infekcí, které velmi často komplikují původní onemocnění pacientů. Proto je třeba charakterizace bakteriálních kmenů. Nejčastějšími původci onemocnění jsou *Klebsiella pneumoniae*, *Escherichia coli* a *Psudomonas aeruginosa*. Tyto multirezistentní kmeny bakterií způsobují zejména zápal plic či zánět močových cest. Jsou velmi častým zdrojem nozokomiální infekce. Ohrožení jsou zejména pacienti se sníženou imunitou. U nich může být infekce velmi vážná až smrtelná. V Dětské nemocnici v Brně, na jejíž popud je tato diplomová práce zpracována, mají pacienty po chemoterapii či transplantaci krvetvorných buněk, kteří jsou infekcí ohroženi. V nemocnici předpokládají, že zisk metody porovnání profilů kmenů (pro screening těchto pacientů) by měl velký význam při preventivních opatřeních proti kolonizujícím kmenům bakterií. Program pro automatickou genotypizaci bakterií metodou rep-PCR je založen na typizaci bakteriálních kmenů a porovnání jejich profilů.

Metody typizace musí splňovat vysokou diskriminační sílu (s vysokou přesností rozlišit různé kmeny bakterií, shodně označit stejné kmeny bakterií) a reprodukovatelnost (tvorba databází a možnosti klasifikace). Reprodukovatelnost může být problémem zejména pro virulentní patogeny, které jsou schopny přizpůsobit se antibiotikům.

Ve většině případů jsou metody typizace založeny na modifikaci PCR a následném hodnocení pomocí elektroforetických metod. Přestože velké množství laboratoří využívá primárně klasické „planární“ gelové elektroforézy, stále více pracovišť přechází na kapilární elektroforézu, která slibuje lepší výsledky. Moderní čipová kapilární elektroforéza umožňuje získat ještě kvalitnější signál a je maximálně automatizována. Přesto dochází k problémům (příčinám zkreslení) a následnému zavádění chyb do hodnocení. Navíc protože je celý proces vyhodnocení často zcela automatizovaný, jsou tato zkreslení těžko zjistitelná a tím i problematická jejich kompenzace.

Jednou z metod typizace je rep-PCR. Při této metodě se využívá repetitivních úseků v genomu bakterií, které se v různých kmenech liší délkou. Úseky jsou namnoženy a pomocí elektroforézy jsou separovány fragmenty stejné délky. Získaným signálem jsou velikosti úseků v jednotlivých vzorcích, které je nutné porovnat a určit, zda jsou vzorky stejného či jiného kmene, fylogeneticky je klasifikovat. Samotné fylogenetické klasifikaci předchází ohodnocení jednotlivých vzorků, jemuž musí být předřazen tzv. bandmatching, tj. úprava pozic bandů. Data mají totiž velké rozptyly u hodnot, které by měly být totožné. Úkolem diplomové práce je najít vhodné předzpracování dat a sestavení klasifikačního aparátu k vytvoření dendrogramu, který reprezentuje fylogenetickou klasifikaci. Pro komfortnější použití algoritmu, by měl být navržený program opatřen uživatelským rozhraním.

1 Typizace DNA

Molekulární typizace DNA slouží k určení bakteriálního kmene nebo poddruhu podle genetické variability. Existuje velké množství metod zaměřujících se na odlišné aspekty v genetických datech. Některé jsou pro jisté druhy vhodnější, jiné se pro typizaci daného druhu hodí méně. Je třeba vždy vybrat nejspecifičtější metodu. V následující části budou některé základní metody popsány.

1.1 Metody typizace

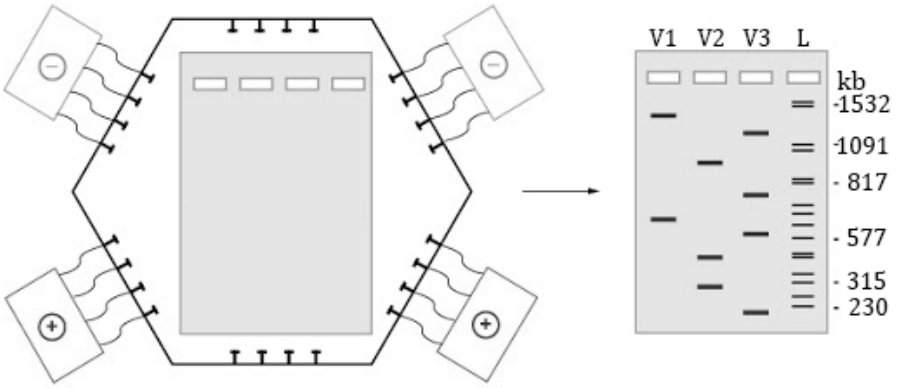
Většina typizačních metod je založena na amplifikaci DNA spojenou s elektroforetickým vyhodnocením. Často se liší způsobem přípravy k amplifikaci DNA. Ve spoustě metod je využívána PCR ([2]) k namnožení žádaných úseků DNA.

Aby bylo metodu možné používat ke genotypizaci druhu či poddruhu, musí splňovat některé podmínky. Metodou by měly být typizovatelné všechny organismy druhu. Některé poddruhy se určují fenotypově (např. reakcí na protilátku). V genomu nemusí být tato reakce odlišitelná. Pro takové kmeny se genotypizace nehodí. Další nutné podmínky pro metody patří vysoká schopnost diferenciace, reprodukovatelnost, jasná, co možná nejsnadnější interpretace měření. Dalšími klady pro metodu jsou rychlost, cenová dostupnost, automatičnost a jednoduchost provedení metody. [1]

PFGE

Pulzní gelová elektroforéza byla poprvé popsána Schwartzem a Cantorem roku 1984. Metoda je často označována za zlatý standard molekulární typizace. Před samotným použitím specifické elektroforézy se používá restrikční štěpení specifickými enzymy. Vzorky jsou poté naneseny na gel a vloženy do speciální vany hexagonálního tvaru. Po spuštění procesu je gel vystaven pulsujícímu elektrickému poli, které mění orientaci. Vzorky mění konformaci a jsou separovány úseky DNA. Vanu s vloženým gelem můžeme vidět na Obr. 1. V pravé části obrázku je naznačený gel po působení elektrického proudu ze tří směrů po obarvení a zobrazení pod UV. [1]

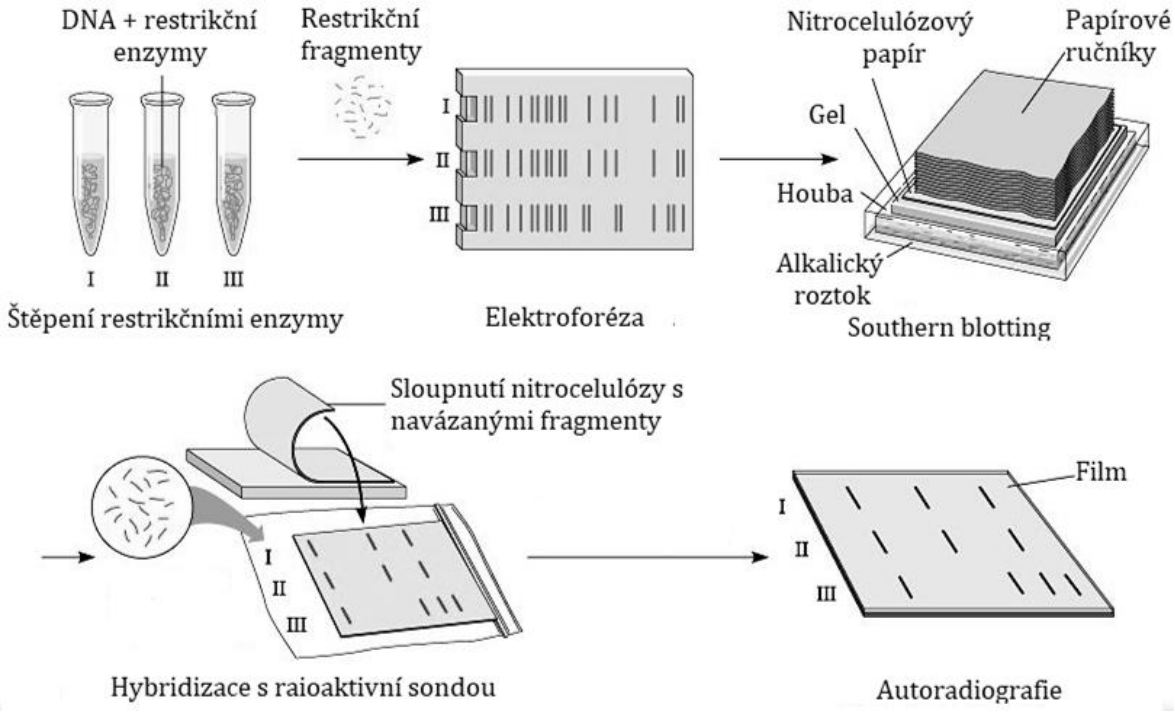
Metoda se hodí zejména pro delší úseky genomu (10-800kb). Po dokončení běhu elektroforézy je DNA obarvena pomocí barviva ethidium bromid k dalšímu zpracování. To probíhá v expertních softwarech, které pracují se snímkem gelu. Velkou nevýhodou metody PFGE je doba zpracování, která je v rozmezí 2 až 3 dní. Vyhodnocení PFGE má velikou diskriminační účinnost. Je snaha vytvořit knihovnu výsledků z PFGE, která by obsáhla většinu živočichů, tím pádem by bylo možné porovnávat vzorky a snadněji je klasifikovat.[1][3]



Obr. 1: Hexagonální vana pro PFGE a následně zobrazený gel po obarvení ethidium bromidem (viditelný pod UV) (převzato z [4])

Southern blotting a RFLP

Southern blotting je metoda pro separaci úseků DNA, získaných štěpením restrikční endonukleázou. RFLP (Restriction fragment length polymorphism, polymorfismus délky restrikčních fragmentů) se používá k separaci cílových sekvencí DNA. Princip je znázorněn na Obr. 2. Fragmenty rozdělené pomocí elektroforézy, jsou následně denaturovány a přeneseny vzlínáním (metodou southern blotting) pomocí pufru na hybridizační membránu (nitrocelulosová nebo nylonová membrána), která je stabilnější než gelový nosič. Po uchycení na membránu jsou vzorky fixovány a radioaktivně značeny. Poloha fragmentů je vyhodnocena autoradiograficky. Metoda RFLP-southern blotting byla používána k diferenciaci bakteriální RNA různých podtypů. Technika je ale překonána. [1][4]



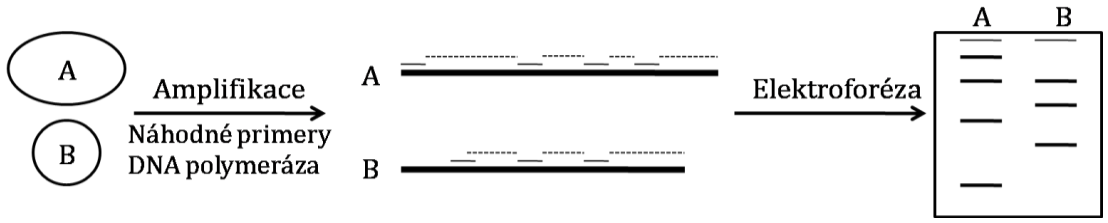
Obr. 2: Princip metody southern blotting s RFLP (převzato z [4])

PCR a Locus-Specific RFLP

Amplifikace pomocí PCR spojená s využitím konkrétných restrikčních enzymů k separaci žádaných lokusů DNA, tedy PCR v kombinaci s RFLP, byla jedna z metod, které nahradily RFLP-souther blotting. Mnohonásobné namnožení zájmových úseků pomocí specifických primerů umožnilo sledovat např. rezistenci bakterií na antibiotika. Nastřihané a amplifikované úseky jsou při této technice elektroforeticky odděleny a dále vizualizovány. [1]

RAPD

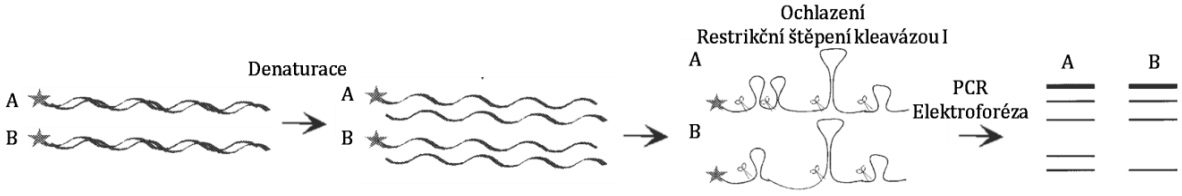
Základem metody RAPD (The random amplified polymorphic DNA, náhodná amplifikace polymorfní DNA) je získání dat pomocí náhodně volených primerů (délka 9 až 10 bází). Náhodností primerů jsou určeny amplifikované úseky (pomocí PCR), které mají pro daný druh stejnou délku, u různých druhů se ale délky amplifikovaných úseků liší. Získané namnožené úseky se hodnotí elektroforeticky. Tím se odlišují velikosti úseků DNA jednotlivých vzorků. Graficky znázorněný postup je na Obr. 3. [1][7]



Obr. 3: Princip průběhu RAPD (upraveno podle [8])

CFLP

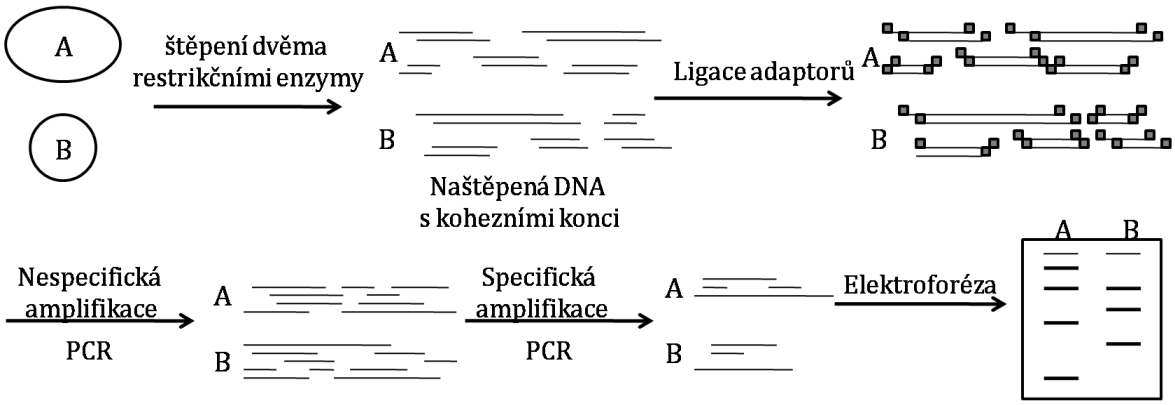
U CFLP (Cleavase fragment length polymorphism, polymorfizmus délek fragmentů vytvořených kleavázou) se užívá k naštěpení vzorku restrikční termostabilní endonukleázou kleavázou I. Tato endonukleáza se váže na struktury, které vzniknou po denaturaci vyšší teplotou. Po ní je reakční směs zchlazena a pro kleavázu I jsou již sestavené smyčky odlišné pro různé druhy. Kleaváza I se na ně váže, pomocí PCR jsou úseky namnoženy a jako u ostatních následuje elektroforetické vyhodnocení. Postup při CFLP je znázorněn na Obr. 4. [1] [9]



Obr. 4: Princip CFLP (převzato z [9])

AFLP

Metoda AFLP (Amplified fragment length polymorfism, polymorfismus délky amplifikovaných fragmentů), zveřejněna Vosem a kol. [10], je založena na štěpení DNA dvěma odlišnými restrikčními enzymy. Princip metody je patrný z Obr. 5. Po štěpení vznikají kohezní konce. Následuje ligace adaptorů (krátkých, známých úseků DNA) na DNA k umožnění nasednutí primerů při nespecifické amplifikaci (první běh PCR). Amplifikací vznikne velké množství ještě nespecifických lokusů. Druhý běh PCR, specifická amplifikace, je dán dvojicí primerů. Tyto primery jsou delší, specifičtější, a snižují počet amplifikovaných úseků. Po specifické PCR je možné data hodnotit opět pomocí elektroforetické separace. [1][11]



Obr. 5: Princip metody AFLP (upraveno podle [8])

DNA Sekvenace

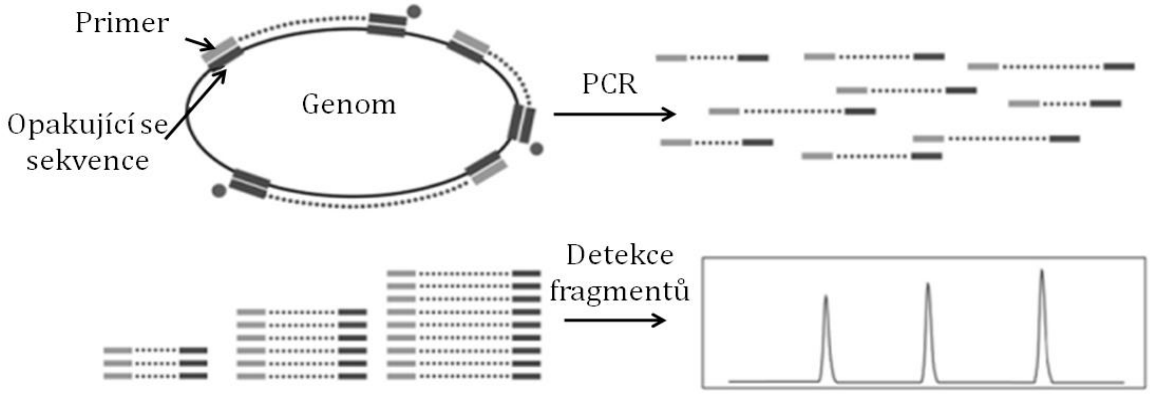
Sekvenováním DNA nám jde o zisk pořadí bází v řetězci DNA. Sekvenačních metod je v dnešní době již celá řada a stále se vyskytují nové a nové přístupy (Max-Gilbert sekvenace, Roche 454, Illuminia, Pacific Biosence, Oxford NAnopore, aj.). Protože v této práci není pracováno s primární strukturou nukleotidů, nebude tato skupina metod představována. O jejím využití pro typizaci nemůže být pochyb. V konkrétním řetězci DNA lze nalézt nejdetailnější rozdíly mezi genomy. [1]

rep-PCR

Jako poslední typizační metodu uvádím rep-PCR (The repetitive extragenic palindromic PCR, interrepetitivní PCR). Data, která jsou v rámci praktické části vyhodnocována, jsou získané právě touto metodou. Roku 1991 ji uvedl tým pana Versaloviče [12]. Metoda je založena na předpokladu existenci hojně se opakujících repetitivních sekvencí v intergenních oblastech genomu. Tyto repetititvní oblasti jsou vhodné pro nasedání primerů při PCR. Amplifikované jsou části mezi těmito úseky. K hodnocení se využívá rozlišná délka oblastí repetitivních úseků jednotlivých kmenů bakterií. [1][12]

Prvním krokem je zisk DNA, při přípravě PCR je nutné zvolit vhodný primer, který bude komplementární k repetitivním oblastem. Za tímto účelem se využívá několik typů

speciálních primerů (např. repetitivní intergenomový palindrom- REP, mozaikové repetitivní BOX elementy, enterobakteriální intergenomové repetice ERIC, dlouhé roztroušené repetitivní elementy RepMP3, tandemově opakované polynukleotidové sekvence $(GTG)_n$) pro PCR. Získané amplifikované části jsou vyhodnoceny pomocí elektroforézy. Rozdílnost kmenů bakterií je v délkách amplifikovaných úseků. [1][12]



Obr. 6: Princip rep-PCR (převzato z [13])

1.2 Porovnání metod

Téměř všechny dříve jmenované metody se zakládají na namnožení úseků DNA a jeho následném elektroforetickém hodnocení. Liší se zejména přístupem PCR, různými primery, opakováním PCR, použitím restrikčních enzymů, aj. Důležitým parametrem je v dnešní době mino jiné rychlost a cena metody, jejich porovnání je uvedeno v Tab. 1.

Tab. 1: Porovnání typizačních metod [14]

	Obtížnost metody	Obtížnost interpretace	Diskriminační síla	Čas provedení [dny]	Cena jednoho testu
PFGE	střední	snadná	vysoká	3	střední
Southern blotting RFLP	snadná	snadná	střední	1	nízká
Lokus specific RFLP	snadná	snadná	střední	1	nízká
RAPD	snadná	snadná	vysoká	1	nízká
CFLP	střední	střední	střední	2	vysoká
AFLP	střední	snadná	vysoká	2	střední
DNA sekvenace	obtížná	střední	vysoká	2	vysoká
rep-PCR	snadná	snadná	vysoká	1	nízká

Z dříve jmenovaných metod má nejvyšší diskriminační sílu PFGE, RAPD, sekvenace DNA a rep-PCR. Nejpřesnější výsledek bude dávat sekvenace DNA, tato metoda je ale velice náročná z mnoha hledisek (finanční, časová, interpretační, dovednostní). Co se týče rychlosti, uplatňují se nejlépe metody RAPD, rep-PCR, Locus-specific RFLP a Souther-Blotting s RFLP. [1][14]

Metodu na určitý experiment je vždy nutné volit nejen z těchto hledisek, ale i s ohledem na úroveň laboratoře, možností zdrojů laboratoře, druh vzorků k typizaci.

Pokud bychom chtěli tvořit referenční databázi, je vhodné volit PFGE (zlatý standard), přesto že je časově náročná, nebo AFLP, vhodná např. pro bakteriální poddruhy. Zpracování virových databází by bylo zase nejvhodnější typizovat pomocí sekvenace DNA, kde jsou nejlépe patrné genetické změny, mutace. [1][14]

Tab. 2: Oblast pro interpretaci jednotlivých typizačních metoda [1]

PFGE	celý genom
Southern blotting RFLP	celý genom
Lokus specific RFLP	specifické lokusy genomu
RAPD	rozdílně dlouhé úseky vzniklé štěpením náhodných primerů
CFLP	fragmenty po štěpení kleavázou
AFLP	specifické úseky značené adaptory
DNA sekvenace	celý genom
rep-PCR	opakující se úseky v genomu

2 Elektroforetická separace DNA

Elektroforéza je separační metoda, při níž lze rozdělit látku na jednotlivé fragmenty podle vlastností jednotlivých částí v elektrickém poli. Jako vlastnost pro separaci se využívá náboj, velikost, členitost, aj. V předchozí kapitole jsme si mohli všimnout, že při typizaci je elektroforéza naprosto zásadní. Je posledním krokem při získávání dat před samotným zpracováním.

Metoda byla poprvé sestavena Arnem Tiseliem roku 1937 [17]. Za tento převratný objev získal Tiselius Nobelovu cenu. Elektroforéza je velmi častá metoda. Používá se její původní podoba i modifikované postupy. Trend vývoje je stejný jako v ostatních oborech zejména miniaturizace, zrychlení, automatizace. Výsledky elektroforézy nejsou za normálních okolností ve většině případů okem viditelné, proto jsou také důležité techniky pro detekci výsledků, zviditelnění koncového stavu elektroforetického běhu.

2.1 Základní princip

Základní princip elektroforézy bude vysvětlen na klasické gelové elektroforéze, od níž jsou všechny další typy odvozeny.

Nejdůležitější částí je při každém pokusu příprava vzorku. Přestože se ve své práci této části nevěnuji, je nutné uvědomovat si, že ani sebelepší metoda nebude v případě špatně získaného či upraveného vzorku dobře vycházet.

Gel je připravován s ohledem na vzorek. Zda separace bude probíhat na základě velikosti, přítomnosti antigenů, náboje, aj. Komerčně se nejčastěji používá polyakrylamidový či agarózový gel. Gel není nutnou podmínkou elektroforézy, jako nosič je možné použít např. i papír.

Další důležitou součástí je pufr. Ten unáší fragmenty vzorku po nosiči. Jeho složení je naprosto zásadní z důvodu ovlivnění vlastností separovaného vzorku. Náboj vzorku v některých případech závisí na tom, v jakém prostředí se vyskytuje. Při nevhodném pH pufru by elektroforetická separace mohla být scestná.

Po přípravě výše jmenovaných částí je gel umístěn do elektroforetické vany naplněné pufrem. Na gel jsou naneseny vzorky, nastavíme napětí a vlivem stejnosměrného elektrického pole, které ve vaně vzniká, je vzorek separován na fragmenty podle připraveného schématu. Nejčastější separace probíhá na základě velikosti. Nejkratší fragmenty se nejsnáze dostávají skrze zesíťované gelové částice, tudíž se při běhu elektroforézy dostanou nejdále.

Výsledkem je rozložení fragmentů na gelu (2D obraz). Pro vyhodnocení je zásadní zviditelnění oblastí a aplikace ladderu na gel, jako „kalibrační prvek“ při kvantifikaci vzorku. Pro zobrazení jsou používána různá barviva, UV světlo, fluorescence, radiační značení, aj. Laddery jsou v dnešní době součástí kitů. Obsahují fragmenty o známých velikostech. Po separaci je možné porovnání vzorku s tímto ladderem.

Při elektroforéze se uplatňují zejména dvě síly: elektrická (urychlující) a odporová. Rychlost separace je dána elektroforetickou pohyblivostí μ_{ef} [$cm^2\ s^{-1}\ V^{-1}$], pro její definici je důležitý předpoklad rovnosti urychlující a odporové síly ve stacionárním stavu, po dodržení této podmínky byl odvozen vzorec (1).

$$\mu_{ef} = \frac{v}{E} = \frac{Q}{6 \cdot \pi \cdot \eta \cdot r} \tag{1}$$

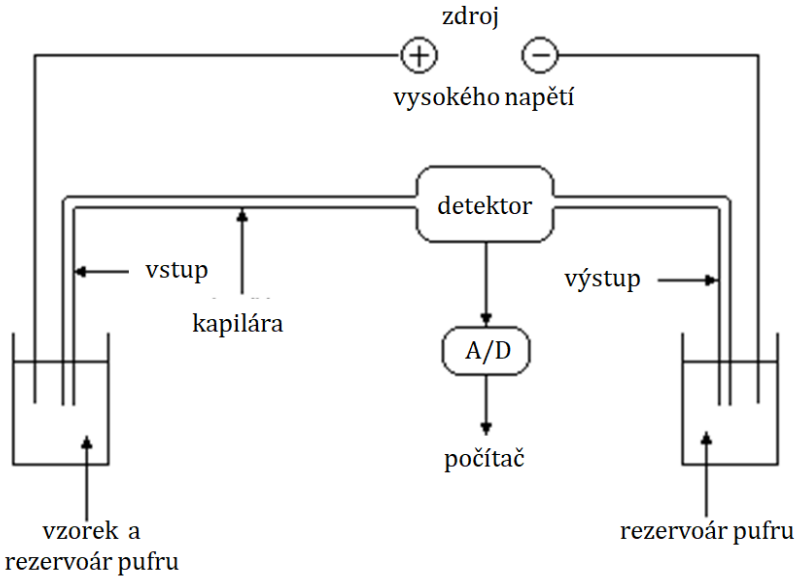
kde v je rychlost [$m \cdot s^{-1}$], E intenzita elektrického pole [$V \cdot m^{-1}$], Q náboj molekuly [C], η viskozita prostředí [$cm^2 \cdot s^{-1}$], r velikost částice [cm]. [15][18]

2.2 Typy elektroforézy

Elektroforetické principy mají v dnešní době širokou škálu technik. Využívají se gelové elektroforézy, izoelektrická fokusace, volná či zónová elektroforéza na různých typech nosičů, denaturační SDS gelová, dvojrozměrná, vysokorozlišovací nebo kapilární elektroforéza. Data pro praktickou část této práce byla získána pomocí čipové kapilární elektroforézy. V další části se zaměřím na kapilární elektroforézu, z níž čipová kapilární elektroforéza vychází.

Kapilární elektroforéza (Capillar electrophoresis, CE)

Počátek kapilární elektroforézy je označován v roce 1967 Stellanem Hjerténem [19]. Používal otevřené milimetrové trubice. Menší kapiláry ještě nebyly k dispozici. Postupem času docházelo k miniaturizaci. Virtanem, Mikkers a Lukacs použili k elektroforéze kapiláry o vnitřním průměru 200 μm z teflonu a skla. V 80. letech 19. století bylo použito křemenné sklo a průměr se zmenšil již na 75 μ. Jorgenson tenkrát popsal teorii kapilární elektroforézy s jejími principy. [20]



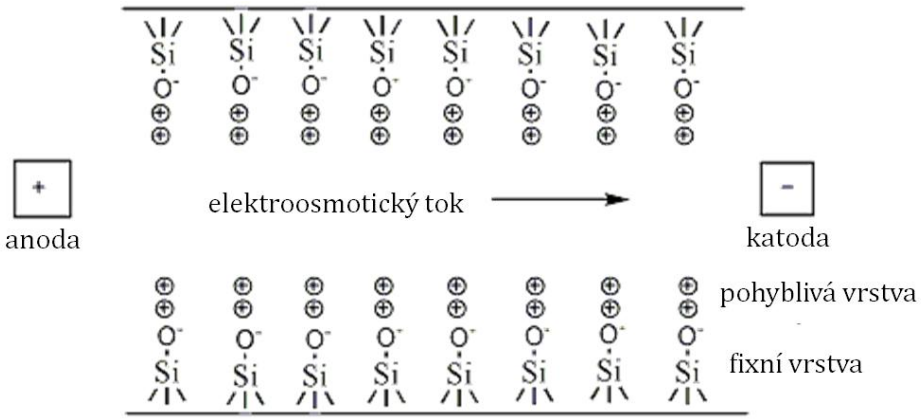
Obr. 7: Schéma kapilární elektroforézy (převzato z [21])

Ke kapilární elektroforéze již není potřeba „vana“ ale přístroj, který se skládá z kazety s kapilárou, nádobkou pro vzorek a pro elektrolyt, zdroj napětí, elektrody, detektor. Často je nezbytný také počítač, kterým se celý děj ovládá.

Na Obr. 7 je znázorněno schéma sestavení kapilární elektroforézy. Dvě nádobky s pufrem jsou propojeny kapilárou, která je stejným pufrem vyplněna. Do jedné z nádobek je přidán vzorek a vlivem přiloženého vysokého napětí je v průběhu kapilárou rozdělen na fragmenty, které jsou zaznamenávány pomocí detektoru. Ziskem není 2D obraz (jako u gelové elektroforézy), ale 1D signál, nejčastěji intenzita fluorescence závislá na čase.

Kapiláry z tavného křemene mají vnitřní průměry 20-200 μm, dlouhé 5-100 cm, jsou zcela naplněny elektrolytem nebo gelem. Díky malému objemu v kapiláře a velkému povrchu stěny kapiláry lze použít mnohem větší napětí než u gelové elektroforézy. Při vyšším napětí vzniká vyšší teplo (Joulovo teplo), které je nežádoucí. Díky povrchu kapiláry je možné stěnu i její obsah chladit. Použití kapilár umožňuje nastavit vyšší napětí a separaci provést za mnohem kratší čas. Nevýhodou kapiláry je interakce se vzorkem. Stěna kapiláry se upravuje, aby neovlivňovala běh separace. K detekci bývá nejčastěji použita absorpce UV, VIS, fluorescence, případně vodivost. [15][18]

$$Si - OH \rightarrow Si - O^- + H^+ \qquad (2)$$



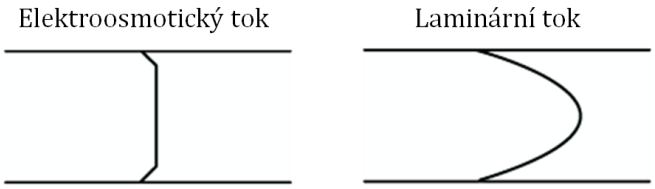
Obr. 8: Elektroosmotický tok (převzato z [22])

Kapilární elektroforéza podléhá elektroforetickému pohybu, jak je popsán pomocí vzorce (1) a často také elektroosmotickému toku, vznikajícímu vlivem adsorpce (disociace) iontů na stěnu kapiláry vlivem stejnosměrného elektrického pole. Na vnitřním povrchu kapiláry vzniká elektrická dvojvrstva, kterou lze popsat rovnicí (2). V kapiláře se vytvoří difuzní vrstva kationtů H^+ (viz Obr. 8), mezi dvojvrstvou a difuzní vrstvou působí zeta potenciál. Ten stoupá společně s pH uvnitř kapiláry. Vrstva kationtů se po přiložení napětí začne silně pohybovat ke katodě, společně s ní se pohybuje veškerá kapalina uvnitř kapiláry. Díky tomuto jevu není prodění uvnitř laminární (ve středu vyšší rychlost proudění), ale proudění o konstantní rychlosti v celém průřezu kapiláry, až u stěny rychlost klesá k nule

(viz Obr. 9). Velikost elektroosmotického toku se dá matematicky popsat pomocí zeta potenciálu, vlastností elektrolytu a rozměru kapiláry, jak je patrné ve vzorci (3).

$$\mu_{eo} = \frac{\varepsilon \cdot \zeta}{\eta \cdot r} \tag{3}$$

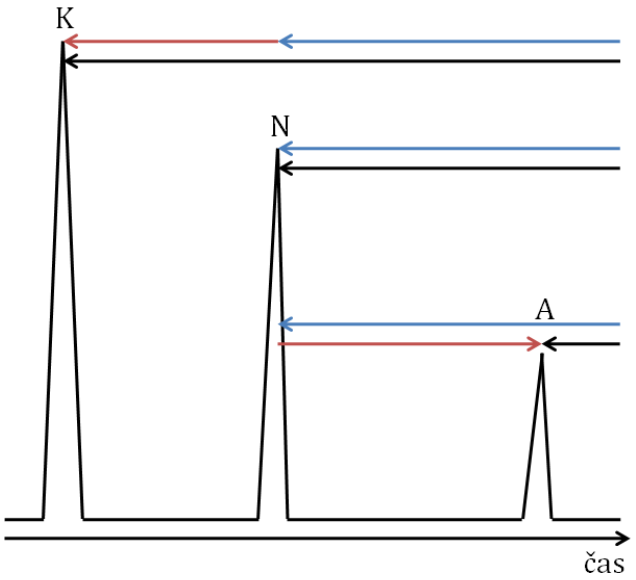
kde je ε permitivita elektrolytu [$C^2 \cdot N^{-1} \cdot cm^{-2}$], ζ zeta potenciál [V], η viskozita [$cm^2 \cdot s^{-1}$] a r poloměr kapiláry [cm]. Velkou výhodou tohoto toku je, že s sebou unáší ke katodě částice, které nemají náboj a zároveň i záporně nabité částice. Díky spojení obou principů jde separovat podle velikosti všechny typy iontů, i neutrální částice. [15][18]



Obr. 9: Porovnání elektroosmotického a laminárního proudění (převzato z [15])

Kapilární zónová elektroforéza (CZE)

CZE je nejjednodušší kapilární elektroforézou. K separaci dochází na základě náboje a hmotnosti fragmentů. Pohyblivost fragmentu v kapiláře je dán součtem μ_{ef} a μ_{eo}. Na Obr. 10 je znázorněn princip vzniku výsledného signálu pro různě nabité částice. Elektroosmotická pohyblivost může být zvyšována spolu s napětím, použitím pufru o vyšším pH, snížením iontové síly (respektive zvýšením potenciálu ζ), zvýšením teploty, aj. [15][18]



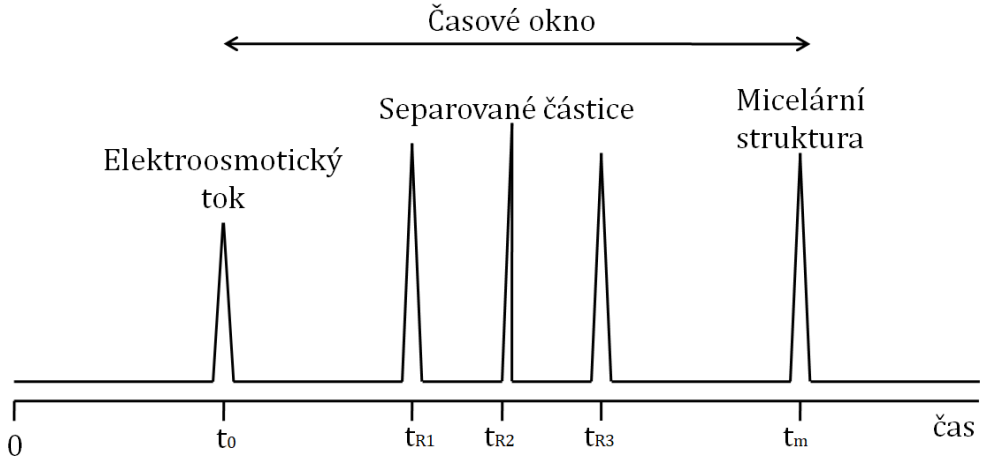
Obr. 10: Princip vzniku výsledného signálu pro kationty (K), neutrální molekuly (N) a anionty (A) černá šipka- výsledná pohyblivost, modrá šipka - elektroforetická pohyblivost, červená šipka elektroforetická pohyblivost (pro kationty kladná, pro neutrální molekuly nulová, pro anionty záporná) (upraveno podle [15])

Kapilární gelová elektroforéza (CGE)

U kapilární gelové elektroforézy je kapilára plněná gelem (polymerem). Jako v klasické elektroforéze se jedná o zesíťovaný polymer, který je připraven s ohledem na analyzovanou látku. Podle ní je volena velikost pórů v částech gelu a hustota. Při GCE je vzorek dělen na základě relativní molekulové hmotnosti. S rostoucí relativní molekulovou hmotností klesá výsledná elektroforetická pohyblivost. Větší molekuly hůře prochází hustou sítí gelu a póry. Plnění kapilár gelem (agarózovým, polyakrylamidovým) je náročné. Proto se používají fyzikální gely, lineární polymery (lineární polyakrylamid, derivát celulosy, dextran, aj.). Ty vytváří síťový efekt a kapiláru lze jejím roztokem snadněji plnit. Často se při CGE používají kapiláry, které jsou ošetřené tak, že v nich nevzniká elektroosmotická pohyblivost. [15][18]

Micelární elektrokinetická chromatografie (MEKC)

MEKC je metoda původně vytvořená k separaci neutrálních molekul. Toho se dociluje přidáním surfaktantu do elektrolytu ve vyšší míře než je kritická molekulární koncentrace (např. 8-9mM SDS). Neutrální molekuly se váží na micelární struktury, které mají na svém povrchu náboj (v případě SDS je náboj záporný). Ta se dále pohybuje již klasickým způsobem podle svého náboje, buď stejným, nebo opačným směrem jako je elektroosmotická pohyblivost. V závislosti na velikosti interakce s micelou je změněna doba průchodu částice kapilárou. Jak je vidět na Obr. 11 jako první je zachycena elektroosmotická pohyblivost, následují separované části a jako poslední se k detektoru dostává micelární struktura, která je největší. MEKC se používá v potravinářské, environmentální, klinické, farmaceutické analýze např. ke stanovení účinné látky v léčivu. [15][18]



Obr. 11: Průběh detekce (upraveno podle [15])

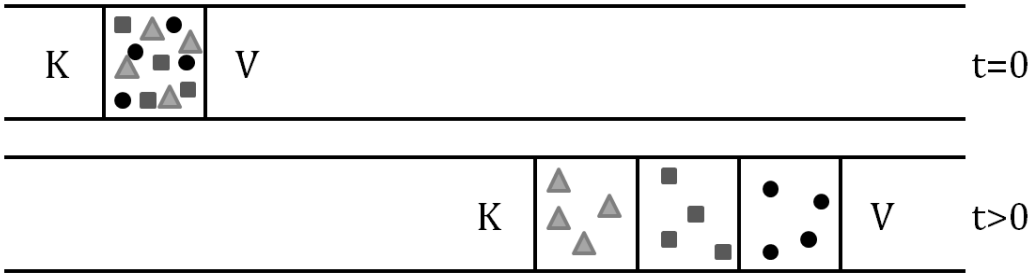
Kapilární izoelektrická fokusace (IEF)

Jak vyplývá z názvu, kapilární izoelektrická fokusace separuje amfolytické látky (např. bílkoviny a peptidy) díky existenci izoelektrického bodu, který je pro danou částici charakteristický. Pro tyto účely se většinou používají kapiláry s potlačeným

elektroosmotickým tokem. V kapiláře je vytvořen elektrickým polem gradient pH (katoda - bazické prostředí, anoda- kyselé prostředí). Látka v kapiláře migruje, dokud není v oblasti pH, kde se chová neutrálně (izoelektrický bod). Po ustálení je potřeba detekovat, kde se jednotlivé části zastavily. Dopravení vzorků k detektoru se uskutečňuje pomocí hydrodynamického toku vyvolaný přetlakem či pod tlakem na jednom z konců kapiláry. [15][18]

Kapilární izotachoforéza (CITP)

Základním principem izotachoforézy je separace na základě elektroforetické mobility. Využívá se k tomu dvou pufrů. Jeden je vedoucí a druhý koncový. Mezi ně se umisťuje vzorek. Je třeba pufry vhodně zvolit, aby platilo, že vedoucí pufr má nejvyšší elektroforetickou mobilitu, vyšší než látky ve vzorku a koncový pufr zase nejnižší, nižší než elektroforetická mobilita vzorku. Při rovnovážném stavu dojde k separaci látek podle jejich elektroforetické mobility (viz Obr. 12), do dosažení tohoto stavu je elektroforetická mobilita různá. Postupně poté putují k detektoru (již se stejnou elektroforetickou mobilitou). Výsledky této metody mají velice ostré hranice mezi vzorky. Při jednom běhu CITP mohou být separovány pouze shodně nabité ionty. [15][18]



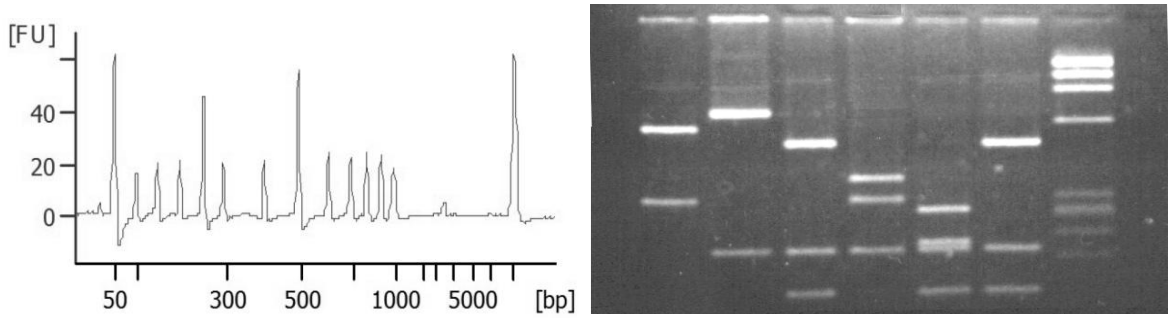
Obr. 12: Separace pomocí CITP (V vedoucí pufr, K koncový pufr) (upreveno podle [15])

Kapilární elektrochromatografie (CEC)

Kapilární elektrochromatografie kombinuje kapalinovou chromatografii a kapilární elektroforézu. Kapilára je naplněna stacionární fází (silikagel, obdoba kolony v chromatografii), která je vystavena elektrickému proudu, díky němu vzniká elektroosmotická mobilita (princip z kapilární elektroforézy). Mobilní fáze je pufr. Elektroosmotický děj probíhá i na povrchu molekul náplně (stacionární fáze). Díky tomu je dosaženo rovného profilu při průchodu kapilárou. Výhodou oproti chromatografii je zisk přesnějších hranic vzorků (menší rozptyl zón). Metoda je nejvhodnější pro neutrální částice. [15][18]

Čipová kapilární elektroforéza

Všechny výše zmíněné typy kapilární elektroforézy je snaha miniaturizovat, automatizovat a přenášet na čip. Poprvé čipovou kapilární elektroforézu zveřejnil roku 1992 Harrison a jeho kolektiv [23]. Oproti kapilární elektroforéze je na čipu možné lépe chladit kapiláry, respektive leptané kanály s rozměry 10-100 μm. Tím pádem je opět možné zvýšit intenzitu elektrického pole, dosáhnout zrychlení celého procesu. Za jeden běh elektroforézy je možné separovat více vzorků. Na čip je potřeba velmi malé množství reaktantů a vzorků. Pohybujeme se v nepatrných rozměrech a objemech. S čipovou kapilární elektroforézou je zisk výsledků analýzy rychlejší (v řádech sekund), za nižší náklady. Výsledky jsou citlivější a selektivnější. V Tab. 4 jsou porovnání klasické CE a čipové CE. [24] [27]



Obr. 13: Ukázka pro porovnání výsledků z čipové (vlevo) a gelové (vpravo) elektroforézy

Čipy se vyrábí ze skla, křemene litografickým leptáním. Dalším a v dnešní době častějším materiálem jsou polymery (polydimethylsiloxan, polymethylmetrhakrylát), které jsou levnější, lehčeji se vyrábí, mají dobré optické vlastnosti pro detekci skrz stěny. Detekce bývá fluorescenční, spektrofotometrická, vodivostní, aj.[24] [25]

Pro čipovou CE je třeba mít celý automatizovaný systém. Na výrobu plně automatizovaných systémů se specializuje mnoho firem, např. Capiler Life Sciences, Agilent Technologies, Bio-Rad, GE Healthcare, Shimadzu Biotech, aj. [28]

K zisku dat byla použita sestava 2100 Bioanalyzer od firmy Agilent (na Obr. 14). Parametry přístroje udávané výrobcem jsou vypsány v Tab. 3.

Tab. 3: Technické parametry přístroje 2100 Bioanalyzer udávané výrobcem [29]

Napájecí napětí	100-240 VAC
Napájecí frekvence	50-60 Hz
Spotřeba	60 VA
Pracovní teplota	15-27 °C

Obr. 14: Sestava 2100 Bioanalyzer (Agilent) použitá k zisku dat (převzato z [29])

Tab. 4: Porovnání CE a čipové CE [27]

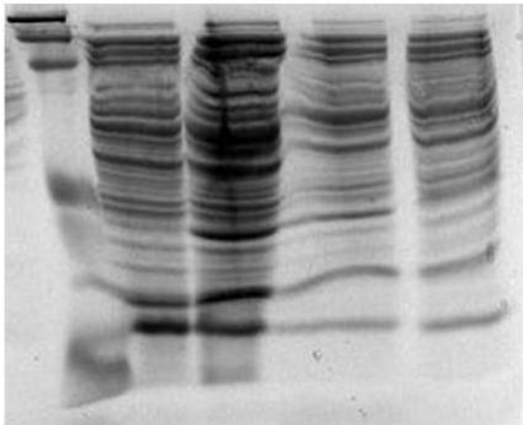
	Kapilární elektroforéza	Čipová kapilární elektroforéza
Vstřikování vzorku	Hydrodynamická, elektrokinetická	Elektrokinetická
Detekce	Ultrafialová, Laserem indukovaná fluorescence	Laserem indukovaná fluorescence
Materiál kapilár	Oxid křemičitý	Sklo nebo polymer
Separační médium	Pufr, síťované polymery, mikročástice	Pufr, síťované polymery, mikročástice
Rychlost analýzy	Rychlá (minuty)	Velmi rychlá (vteřiny)
Kapacita píků	Více píků (delší kapilára)	Méně píků (krátká kapilára)
Integrace	Obtížná (lze špatně propojit bez mrtvého prostoru)	Snadná (např. s PCR)
Automatizace	Vysoce automatizované	Vysoce automatizované (komerční systémy)
Velikost vzorku	Velmi malá (nl až µl)	Velmi malá (nl až µl)
Použití reagencií	Velmi malé (µl až ml za den)	Velmi malé (µl až ml za den)

2.3 Zkreslení a chyby

Ke zkreslení a chybám při elektroforéze může dojít z mnoha důvodů, důsledek bývá často stejný a to, že dojde k rozmytí zóny při detekci, případně přerytí dvou zón. Pro detekci je zásadní, k dosažení tíženého výsledku, udržet dostatečný odstup zón a ponechat pouze drobný rozptyl v rámci jedné scény. Tohoto požadavku není snadné dosáhnout, zejména když je měření vystaveno velkému množství mechanismů, které to ovlivňují. Největší vliv na zkreslení výsledků je špatná úprava vzorku, vnesení nečistot, bubliny v kapiláře, nevhodně zvolený nosič, kapilára, aj. Tyto problémy jsou zásadní a vedou spíše k naprosto špatným, nebo žádným výsledkům. Jsou to chyby lidského faktoru a dále nebudou rozebírány. Dalšími důležitými parametry jsou molekulární difuze, Joulovo teplo, doba vstřiku vzorku, velikost vzorku, povrch kapilár, náplň kapilár aj. [15]

Joulovo teplo

Elektroforéza, jako metoda využívající elektrický proud, je zatížena joulovým teplem, které se při průchodu proudu vzorkem vytváří. Při zahřívání je problémem místní změna gradientu. Ačkoliv při průběhu elektroforézy dochází k chlazení, je právě s teplotním gradientem problém. Ten může být odlišný u stěn a uprostřed kapiláry. Teplota ovlivňuje zejména viskozitu (pufru, gelu, vzorku), její změnou může dojít k rozmývání zón vlivem změny mobility. Změna teploty o jeden stupeň může vyvolat změnu o 2-3 % viskozity. K regulaci joulova tepla (teplotního gradientu) se používají základní čtyři metody: změna velikost elektrického pole, změna vnitřního kapilárního průřezu, změna koncentrace pufru a aktivní kontrola teploty (např. termostatem). [15]



Obr. 15: Elektroforetický gel ovlivněný joulovým teplem (převzato z [16])

Vstřikování vzorku

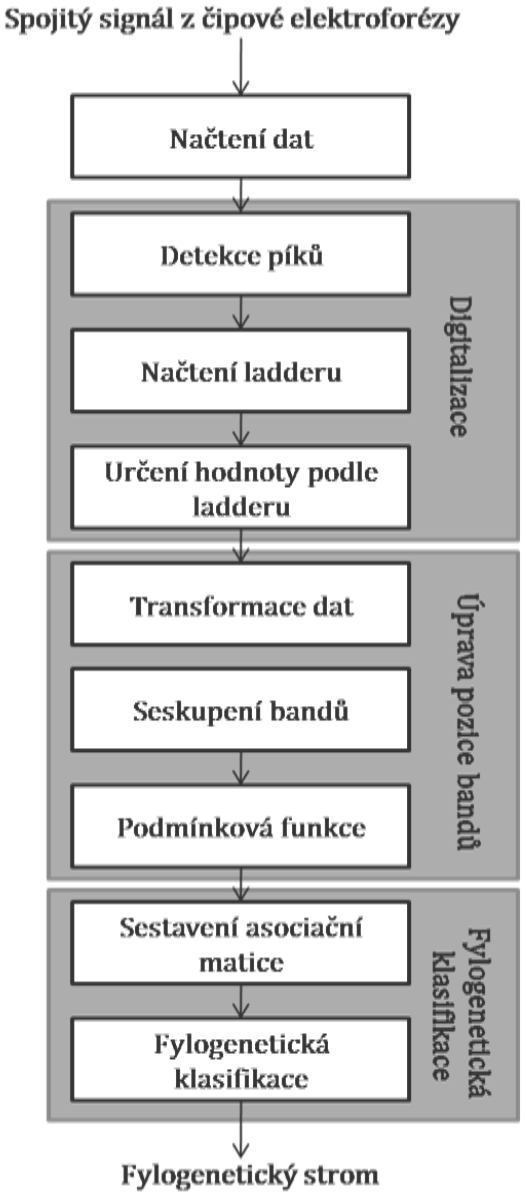
Při vstřikování vzorku je důležité, dbát na velikost otvoru pro vstřikování, době vstřikování, velikosti objemu vzorku. S rostoucím otvorem pro vstřikování klesá rozlišení, velikost otvoru by se měla volit také s ohledem na difuzní koeficient vzorku. Objem vstřikovaného vzorku je třeba určit podle velikosti kapiláry, vhodné je 1 až 2 % z celkového objemu. [15]

Vzorek a povrch kapiláry

Základem kapilární elektroforézy je dříve zmíněný elektroosmotický tok, který je založen na interakci povrchu kapiláry a pufru. Je zřejmé, že zásadní vliv na celkovou účinnost metody bude mít i adsorpce vzorku ke kapiláře. Se zvýšením povrchu kapiláry je možné lépe snižovat joulovo teplo, jak bylo řečeno dříve, ale zároveň je zvýšená pravděpodobnost adsorpce vzorku ke stěně, což je nežádoucí. K interakci mezi kapilárou a vzorkem dochází zejména v podobě hydrofilních a hydrofobních interakcí. [15]

3 Praktická část

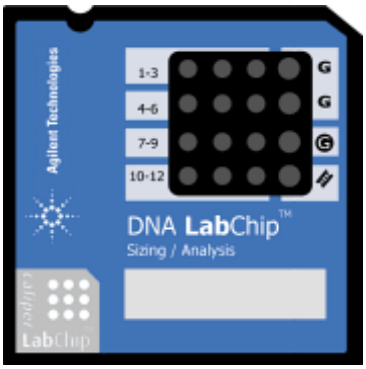
Cílem praktické části diplomové práce je sestavení programu s uživatelským rozhraním pro automatickou klasifikaci vzorků z rep-PCR formou dendrogramu. Vstupními daty do programu jsou hodnoty získané z čipové kapilární elektroforézy (té předchází příprava vzorků metodou rep-PCR). Výstupem algoritmu je dendrogram (resp. fylogenetický strom), umožňující rozhodnutí, zda jsou si vzorky podobné a patří do stejného kmene. Navržený program pro analýzu byl pojmenován GenTyBa, Genová typizace bakteriálních kmenů. Blokové schéma programu je znázorněno na Obr. 16.



Obr. 16: Blokové schéma programu GenTyBa

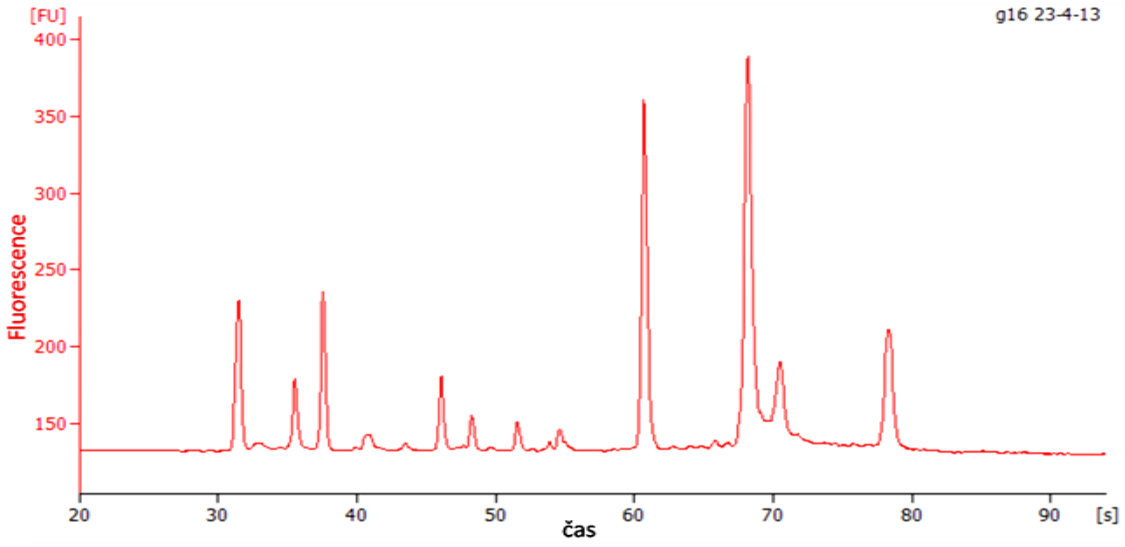
3.1 Vstupní data

V Dětské nemocnici byla naměřena data k sestavení algoritmu. Izolace DNA ze vzorků probíhala pomocí komerčního kitu UltraClean Microbial DNA isolation kit (MO Bio Laboratories). Amplifikace DNA a zároveň štěpení na unikátní fragmenty pro daný kmen probíhalo metodou rep-PCR za použití rep primerů (REP 1R-I (5´-IIIICGICGICATCIGGC-3´), REP 2-I (5´-ICGICTTATCIGGCCTAC-3´)). PCR produkty podléhaly další analýze na DNA čipu (Obr. 17) přístroje Bioanalyzer (Obr. 14) fungujícího na principu kapilární elektroforézy. Výstupem je spojitá závislost fluorescence na čase (viz Obr. 18). Získaná data jsou vstupem pro program k automatické klasifikaci.



Obr. 17: DNA chip firmy Agilent použitý k zisku dat [29]

Data pro sestavení algoritmu nebylo nutné amplifikovat. Standardizované DNA markery mají přesně sestavené profily o různých velikostních fragmentech a byly pouze analyzovány pomocí čipové kapilární elektroforézy. Použité markery měly různé rozpětí a velikosti DNA fragmentů. Jejich charakterizace je v Tab. 6. Každý marker byl proměřen vícekrát. Celkově bylo použito pro sestavování algoritmu přes 100 měření s více než 3000 bandy.



Obr. 18: Vstupní data programu. Výstup z kapilární čipové elektroforézy Bioanalyzer

Dále bylo extrahováno 72 vzorků od pacientů. Vzorky patří do 13 kmenů bakterie *Klebsiella pneumoniae*. V Tab. 5 jsou uvedena pracovní označení a počty vzorků kmenů bakterií získaných ze vzorků pacientů. Tyto vzorky podléhaly celému procesu (rep-PCR a čipové kapilární elektroforéze).

Tab. 5: Kmeny reálných vzorků

Kmen	G	H	I	J	K	L	M	N	O	P	Q	R	S
Počet vzorků	5	6	10	7	6	4	10	5	5	5	5	2	2

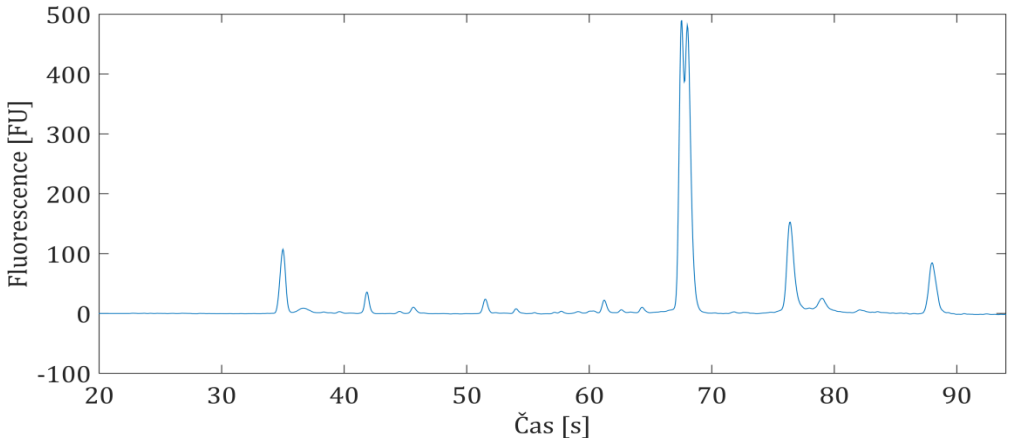
Tab. 6: DNA markery

GeneRuler100 bp [bp] A	GeneRuler50 bp [bp B	O'GeneRuler 1 kb [bp] C	GeneRuler 1 kb [bp] D
3000	1000	10000	10000
2000	900	8000	8000
1500	800	5000	6000
1200	700	4000	4000
1000	600	3500	3500
900	500	3000	3000
800	400	2500	2500
700	300	2000	2000
600	250	1500	1500
500	200	1000	1000
400	150	750	750
300	100	500	500
200	50	250	250
100			

Se spojitým signálem se dá pracovat jako s celkem. Jedna z možností byla, porovnávat celé signály, neřešit jednotlivé píky, bandy, nevytvářet gel. Realizace tohoto postupu se neosvědčila. Po drobné filtraci signálu byla sestavena asociační matice porovnání korelací dvou signálů, jejich diferencí, kovariance. Žádný z pokusů nevedl k uspokojivým výsledkům. Proto byl zvolen postup vytvoření umělého gelu. Současně se jedná o tradiční metodu, která je mikrobiologům dobře známa.

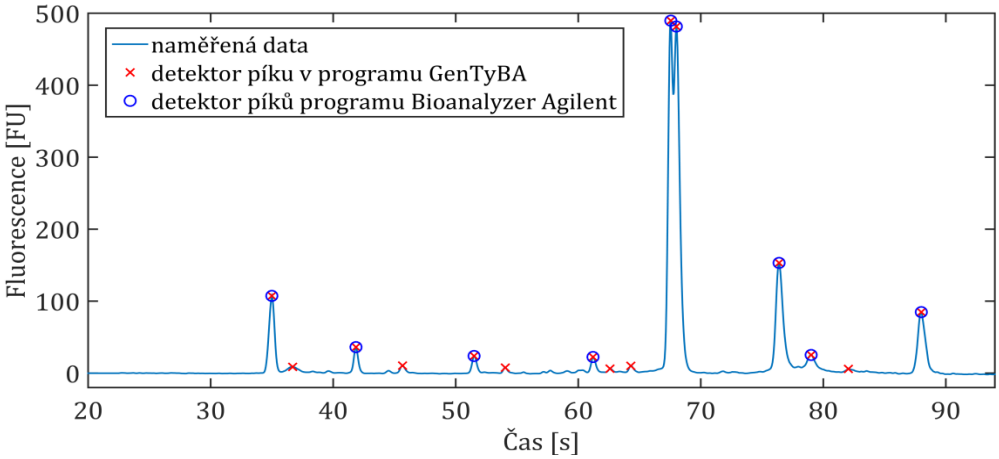
3.2 Digitalizace

Získaný spojitý časově závislý signál je potřeba převést na diskrétní signál závislý na velikosti fragmentů, vytvořit tzv. umělý gel. Na Obr. 19 je zobrazený signál (konkrétně vzorek g18 patřící do kmenu K). Z této podoby signálu je třeba získat časové okamžiky píků a převést je na velikosti fragmentů. Tento postup umí již program Bioanalyzer. Avšak data, obsahující pouze píky, nebyla vhodná pro algoritmus. Nastavení detektoru nebylo dostatečně citlivé, drobné píky nebyly detekovány. Proto program GenTyBa pracuje se surovými daty (Obr. 19) a detektor píků má vlastní.



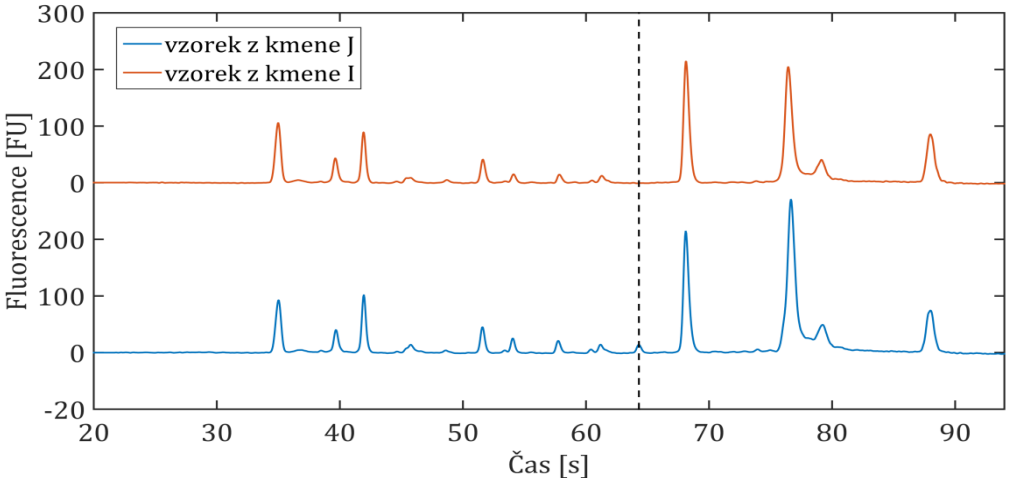
Obr. 19: Spojitý časově signál, vstupní data programu GenTyBa

Na Obr. 20 můžeme vidět porovnání detekce píků v programu GenTyBa a softwaru patřícímu k čipové elektroforéze Bioanalyzer. Detektor Bioanalyzer detekuje 9 píků,oproti tomu GenTyBa označil 15 píků, je citlivější i k méně výraznému signálu. To se projeví i na vytvoření umělého gelu po převodu na velikostní fragmenty, jež je provedeno pomocí zadaných hodnot ladderu, které jsou přiřazeny k časovým okamžikům. Mezi jednotlivými známými píky ladderu dochází k proložení oblasti kubickou závislostí. Časový okamžik vzorku je porovnán s časovými okamžiky fragmentů ladderu a následně pomocí proložených hodnot je určena velikost daného fragmentu.



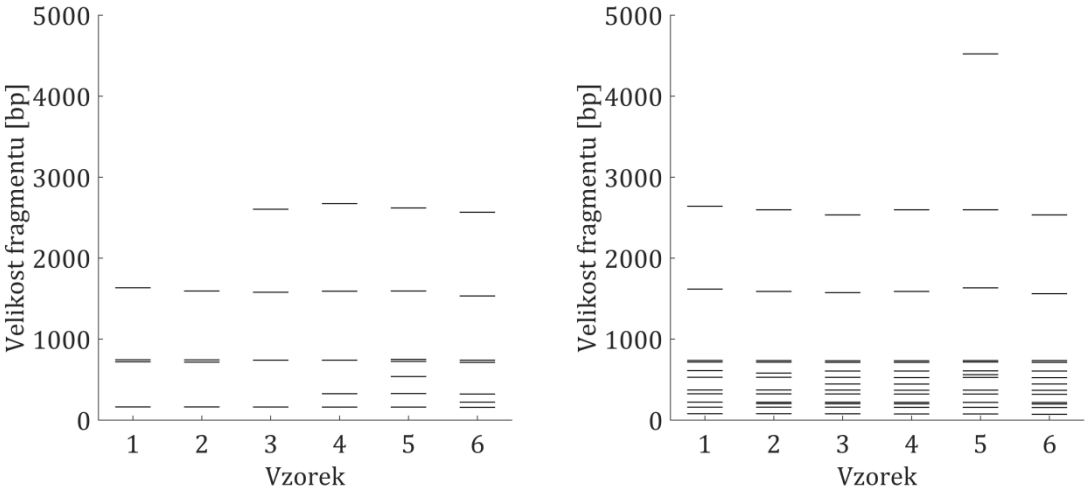
Obr. 20: Porovnání detekce píků u signálu g18 (kmen K); červený křížek detektor programu GenTyBa; modré kolečko detekce programu Bioanalyzer firmy Agilent

Detektor v programu GenTyBa využívá funkce findpeaks. Musí splňovat podmínku velikosti píku min. 4,25 FU (Fluorescent unit) a prominenci min. 4,5. Tyto hodnoty byly získány rozsáhlým testováním a jejich nastavení se ukázalo jako dostatečně citlivé pro rozlišení i malých rozdílů v signálech. Dva kmeny bakterií (I a J) mají téměř shodný profil (viz Obr. 21), rozdíl je v přítomnosti jediného píku navíc (černá čárkovaná čára v Obr. 21). Pík není nijak zvlášť výrazný, ale program jej dokáže detekovat a přiřadit.



Obr. 21: Původní data vzorku z kmene I (nahoře) a z kmene J (dole)

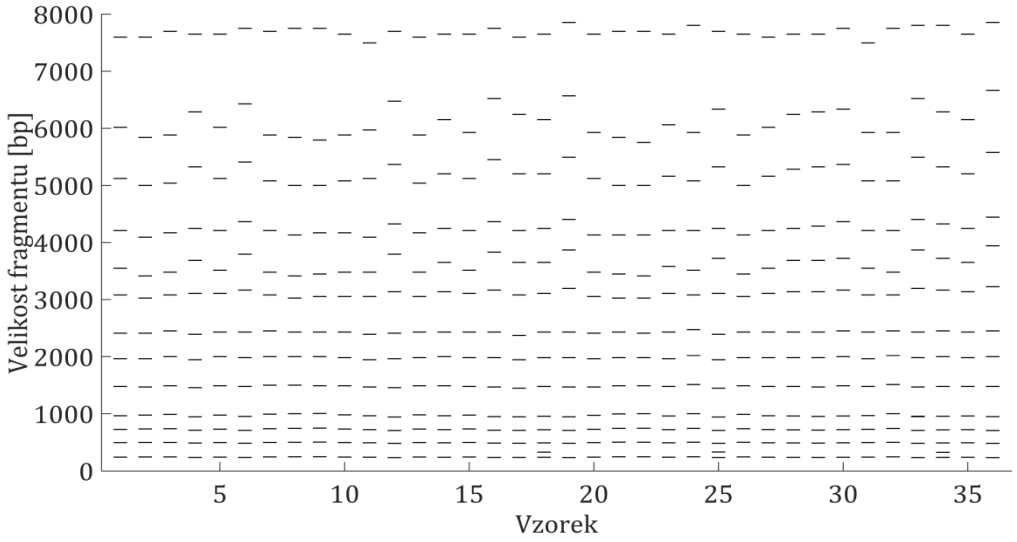
Na Obr. 22 je na první pohled vidět, že porovnání gelu z programu Bioanalyzer (vlevo) není směrodatné. 1. a 2. vzorek obsahuje pouze 4 bandy, oproti tomu vzorek 5 má 7 bandů. Porovnání takového gelu nemůže vést ke správné klasifikaci. V pravé části je sestavený gel programu GenTyBa, profily jsou si na první pohled výrazně podobnější, než v předchozím případě. Je patrné, že sestavení vlastního detektoru bandů byl dobře zvolený krok.



Obr. 22: Uměle vytvořené gely kmenu K po detekci píků; vlevo gel detektoru Bioanalyzer; vpravo detektor v programu GenTyBa

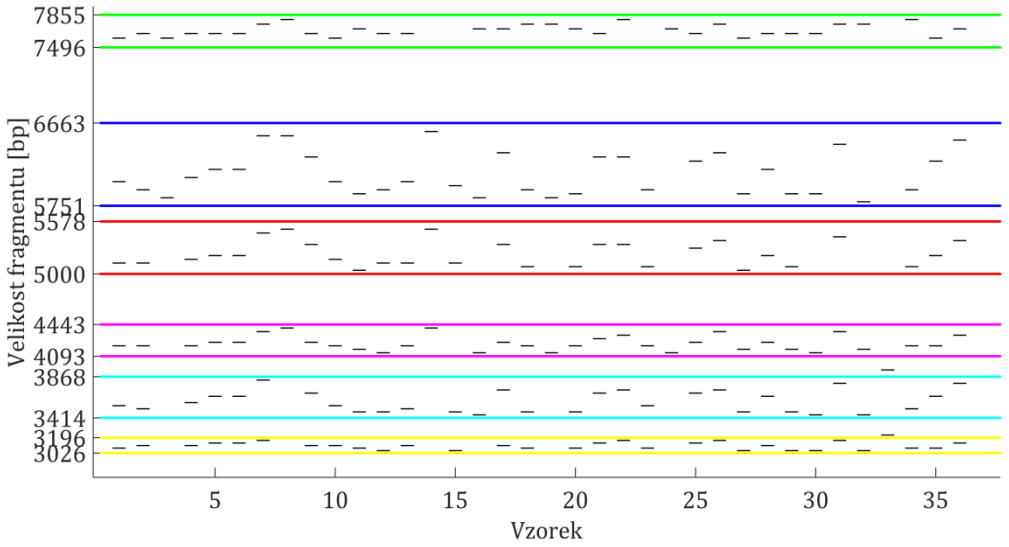
3.3 Úprava pozic bandů

Při zisku dat dochází k nepřesnostem. Data, která by měla být stejná, nemají vždy tutéž hodnotu. Před samotným ohodnocením vzorků a fylogenetickou klasifikací je nutné zajistit korekci DNA fragmentů. Na Obr. 23 je vykreslen originální gel, hodnoty, které byly získány digitalizací.



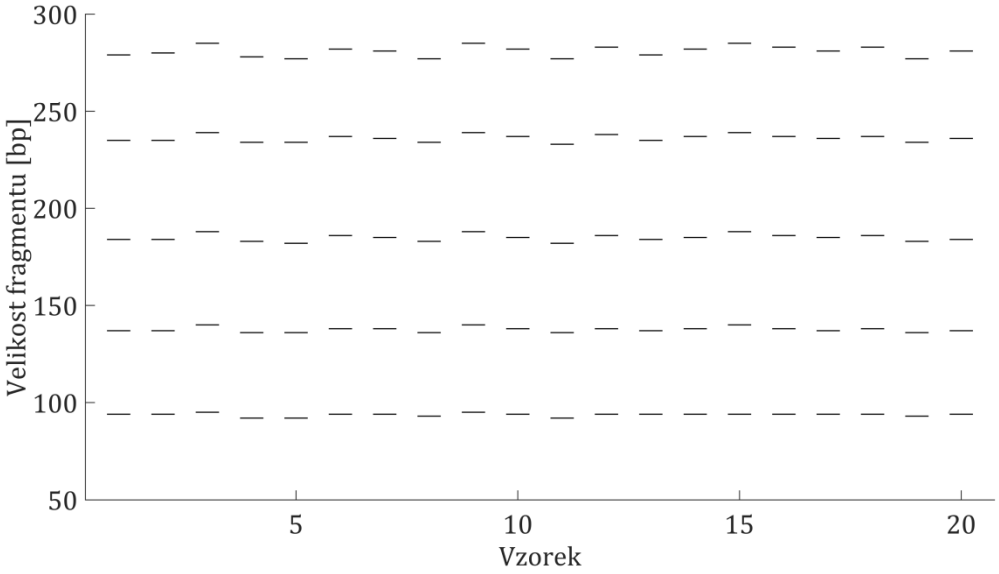
Obr. 23: Originální gel

Na Obr. 24 je ještě více patrné, že bandy, které by měly mít tutéž hodnotu, mají velký rozsah, někdy je rozsah v rámci jednoho velikostního fragmentu větší než samotná mezera mezi nimi. Mezi barevnými značkami (vodorovné čáry) se nachází bandy ze stejné velikostní skupiny fragmentů. Tmavě modrá skupina má největší rozsah (912 bp), druhý největší (578 bp) patří červené skupině. Mezera mezi těmito skupinami (173 bp) je menší než jejich rozsahy. Dále zelená skupina má rozsah 395 bp, očekávali bychom, že v případě větších hodnot bude i rozptyl větší. Úloha úpravy pozic je proto velice ztížena.



Obr. 24: Znázornění rozsahů u ladderu v rámci velikostních fragmentů, mezi shodně barevnými značkami se jedná o stejný velikostní fragment

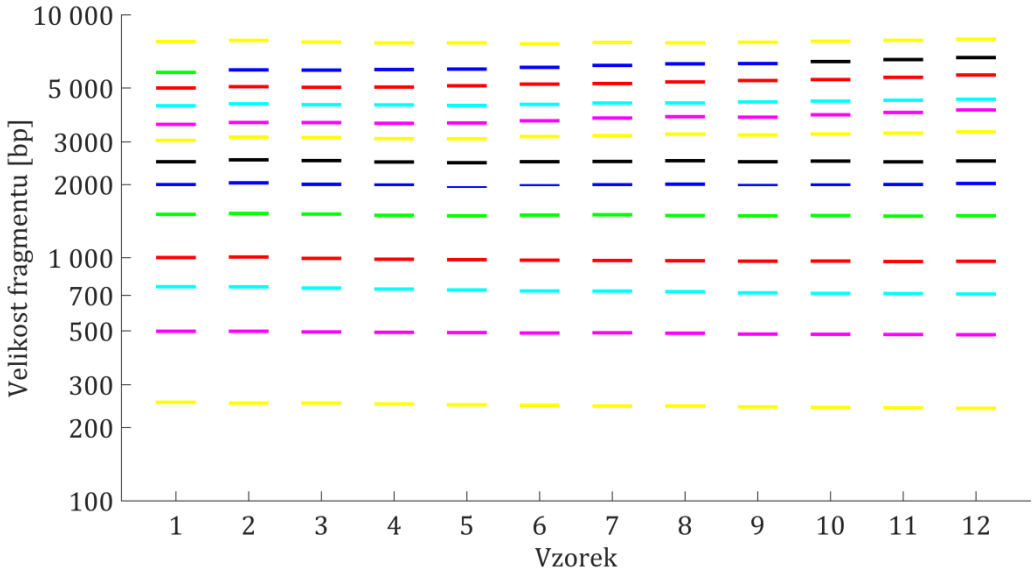
Hlavním úkolem je najít transformaci dat, která pomůže rozptyly hodnot vyrovnat vhodným způsobem. Za tímto účelem bylo využito naměřených dat hmotnostních markerů k provedení analýzy rozptylu dat. Bylo zjištěno, že rozptyly v rámci jedné velikostní skupiny nejsou lineárně závislé na velikosti fragmentů. To dokládá i ukázka na Obr. 24. Nejdelší fragmenty okolo 8000 bp (rozdíl mezi nejkratším a nejdelším fragmentem této skupiny je 359 bp) nemají tak velký rozptyl jako hodnoty okolo 6000 bp (rozdíl mezi nejkratším a nejdelším fragmentem této skupiny je 912 bp). Rozdíl délek je až 15%. Dále na Obr. 25 je vidět, že krátké fragmenty se také odlišují. Bandy okolo 280 bp mají nejméně 277 bp a nejvíce 285 bp, rozpětí 8 bp (rozdíl cca 3% z rozsahu).



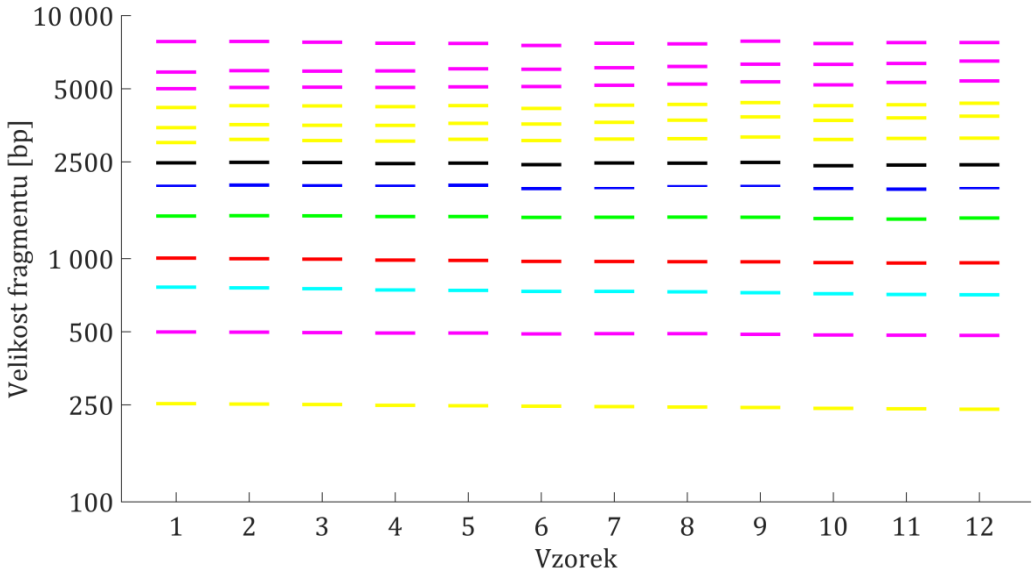
Obr. 25: Část naměřených ladderů, krátké fragmenty

Při úpravě pozic bandů, respektive označení podobně velkých bandů jako jedna skupina, docházelo ke dvěma zásadním problémům. Prvním z nich je rozdělení jedné velikostní skupiny na více. Na Obr. 26 můžeme vidět rozdělení dvou ladderů na tři (cca 6000 bp, předposlední skupina, barvy zelená modrá a černá). V ideálním případě by k tomuto rozdělení nemělo dojít a hodnoty by měly být označeny jednou barvou.

Dalším možným problémem je velká tolerance, kdy jsou dva a více fragmentů zařazeny do jedné skupiny (viz Obr. 27, bandy žluté barvy (cca 2500-5000 bp) a růžové barvy (cca 5000-10000 bp). Velkou tolerancí se může stát, že v jednom vzorku bude označeno více bandů za fragment jedné délky.

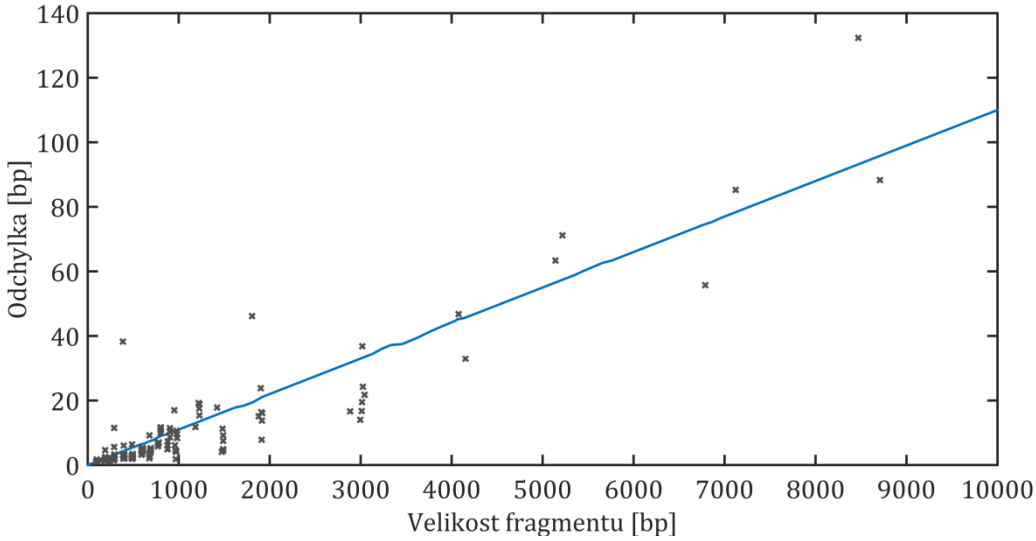


Obr. 26: Problémová situace - rozdělení jednoho fragmentu



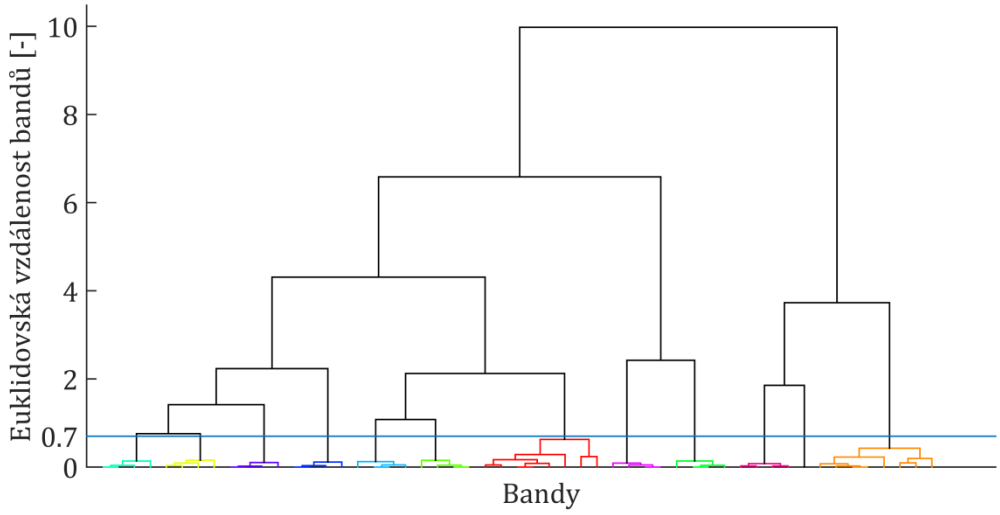
Obr. 27: Problémová situace - sloučení dvou rozdílných fragmentů

Díky analýze rozptylu byla sestavena po částech lineární transformační funkce (viz Obr. 28), která se snaží odstranit rozdíly v odchylkách fragmentů. Funkce je dána směrnicí, která se 125 krát mění (minimální směrnice je -0,0018 a maximální 0,0198). Díky úpravě transformační funkcí je možné použít konstantní práh při shlukování bandů, respektive jejich euklidovských vzdáleností, pomocí metody WPGMA. Další nedílnou částí je podmínková funkce, která je zařazena kvůli výše zmíněnému druhému problému. V jednom vzorku nemohou být dva bandy zařazeny do stejné velikostní skupiny. Základem podmínkové funkce je shlukování k-means.

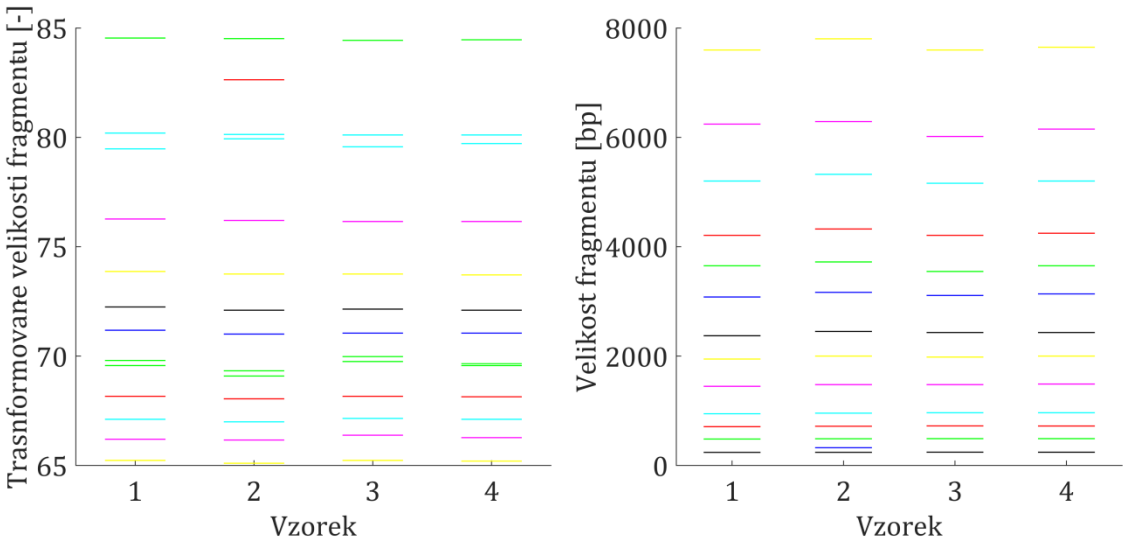


Obr. 28: Transformační funkce

Obr. 29 ukazuje dendrogram po shlukové analýze metodou WPGMA euklidovských vzdáleností transformovaných pozic bandů (transformované pozice bandů viz Obr. 30 vlevo) s prahem 0,7 pro odlišení shluků (na obrázku barevné odlišení). Je vidět, že ve 4 vzorkách jsou nalezeny vždy 4 bandy. Ve dvou případech je ale do shluku zařazeno 8 bandů. Na Obr. 30 je vidět, že bandy okolo hodnoty 80 (světle modré) a okolo 70 (zelené) jsou si velice blízké. Euklidovská vzdálenost ve stromu těchto bandů je menší než 0.7. Tyto bandy jsou zařazeny do jedné skupiny velikosti fragmentů. Na Obr. 30 vpravo je zřejmé, že podmínková funkce, zařazená na konci bloku úpravy pozic bandů, tuto chybu zachytila a všechny bandy jsou zařazeny do správné skupiny velikosti fragmentů.

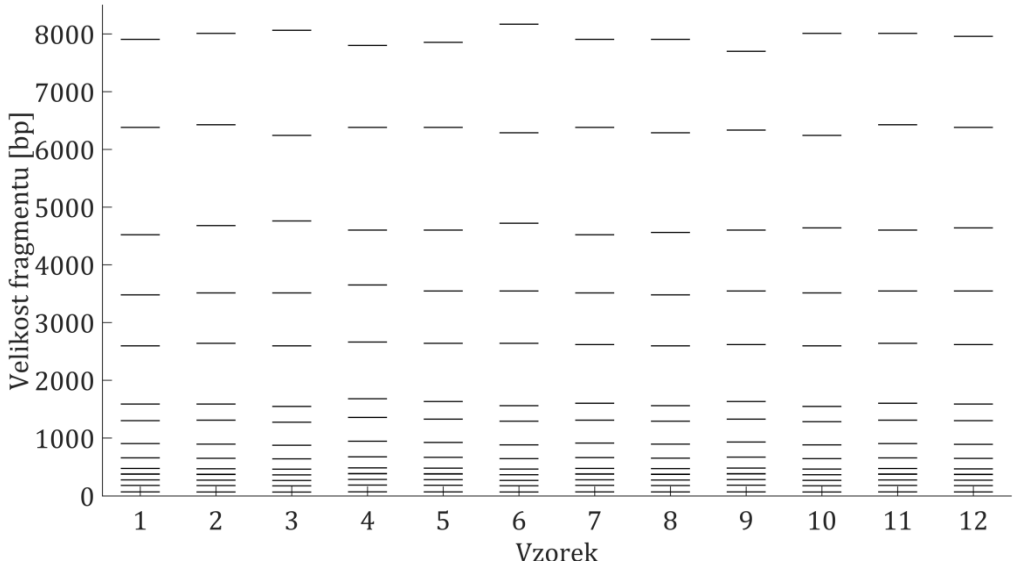


Obr. 29: Přiblížený dendrogram shlukové analýzy WPGMA euklidovských vzdáleností transformovaných pozic bandů

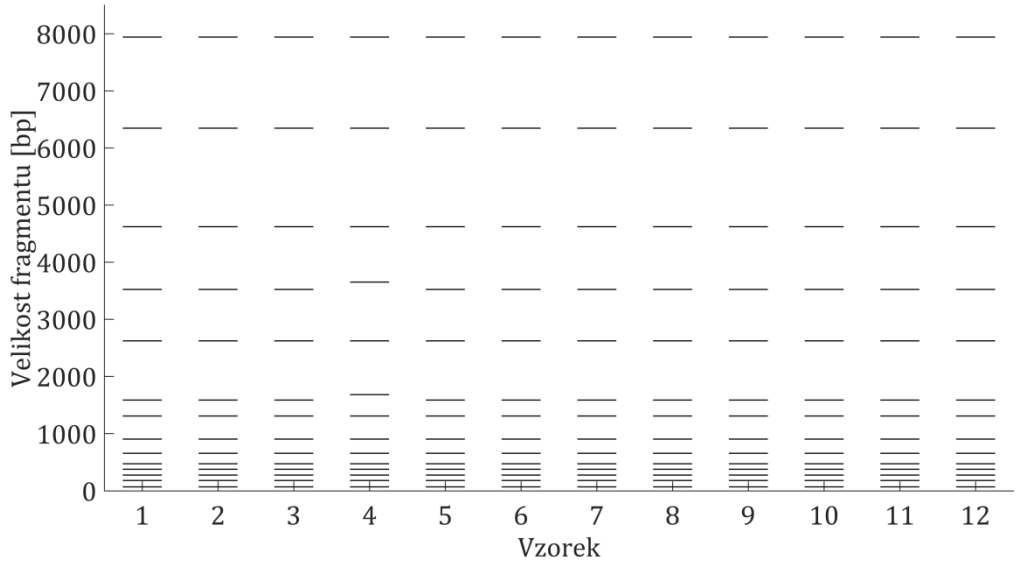


Obr. 30: Ukázky z průběhu úpravy pozic bandů, vlevo vzorky po transformaci dat; vpravo vzorky po celé úpravě pozic bandů

Na Obr. 32 jsou ukázány různé pozice bandů. Je vidět, že bandy nemají tutéž velikost, ačkoli by měly mít. Na Obr. 32 jsou pozice upraveny a velikosti odpovídajících si bandů jsou shodné. Díky transformaci dat, shlukování bandů a podmínkové funkce bylo docíleno tíženého výsledku.



Obr. 31: Uměle vytvořený gel před úpravou pozic bandů

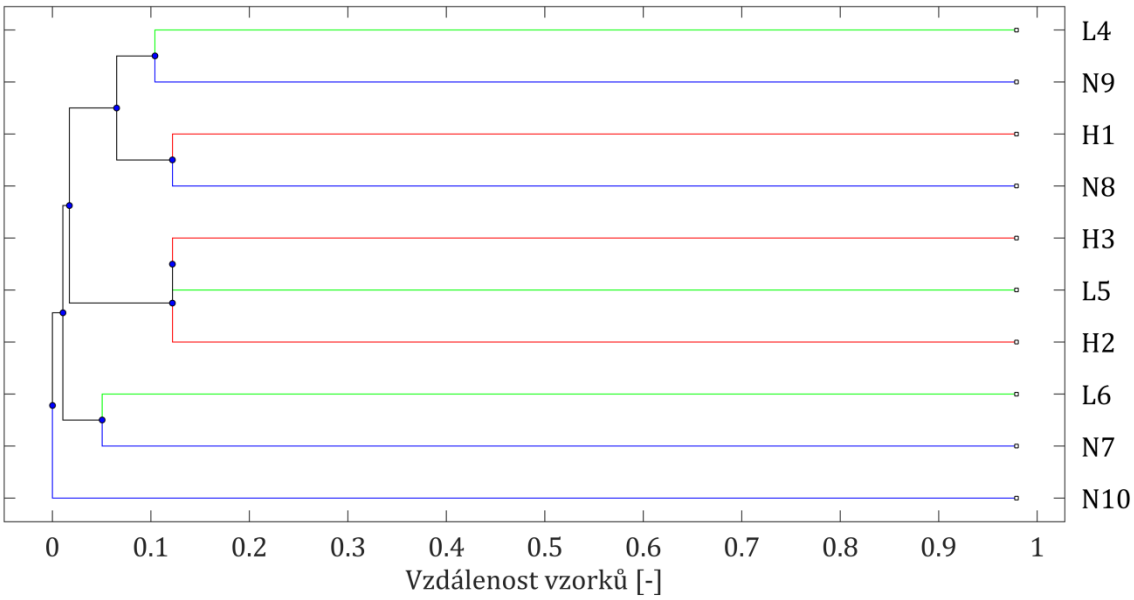


Obr. 32: Uměle vytvořený gel po úpravě pozic bandů

3.4 Fylogenetická klasifikace

K fylogenetické klasifikaci je možné přistoupit až po úpravě pozic bandů. Při použití původních dat, bez úprav, je výsledek shlukování špatný. Na Obr. 33 je ukázka klasifikace bez úpravy pozic bandů.

Ke klasifikaci je použita shluková analýza UPGMA. Vstupem do ní je asociační matice sestavená podle pozic bandů, které jsou převedeny do binární podoby. Pozice v asociační matici odpovídá neshodám (0,1) ve dvou vzorcích podělené neshodami (0,1) a pozitivními shodami (1,1).



Obr. 33: Klasifikace vzorků bez použití úpravy pozic bandů

Ukázka výpočtu asociační matice M:	Velikost bandů [bp]	V1	V2	V3
V1 a V2: 01/10=3 11=3 $M_{12}=M_{21}=3/(3+3)=0,5$ V1 a V3: 01/10=3 11=3 $M_{13}=M_{31}=3/(3+3)=0,5$ V2 a V3: 01/10=2 11=3 $M_{12}=M_{21}=2/(2+3)=0,4$	100	1	1	1
	250	0	1	0
	255	0	0	1
	400	1	0	0
	750	1	1	1
	1000	1	0	0
	3000	1	1	1

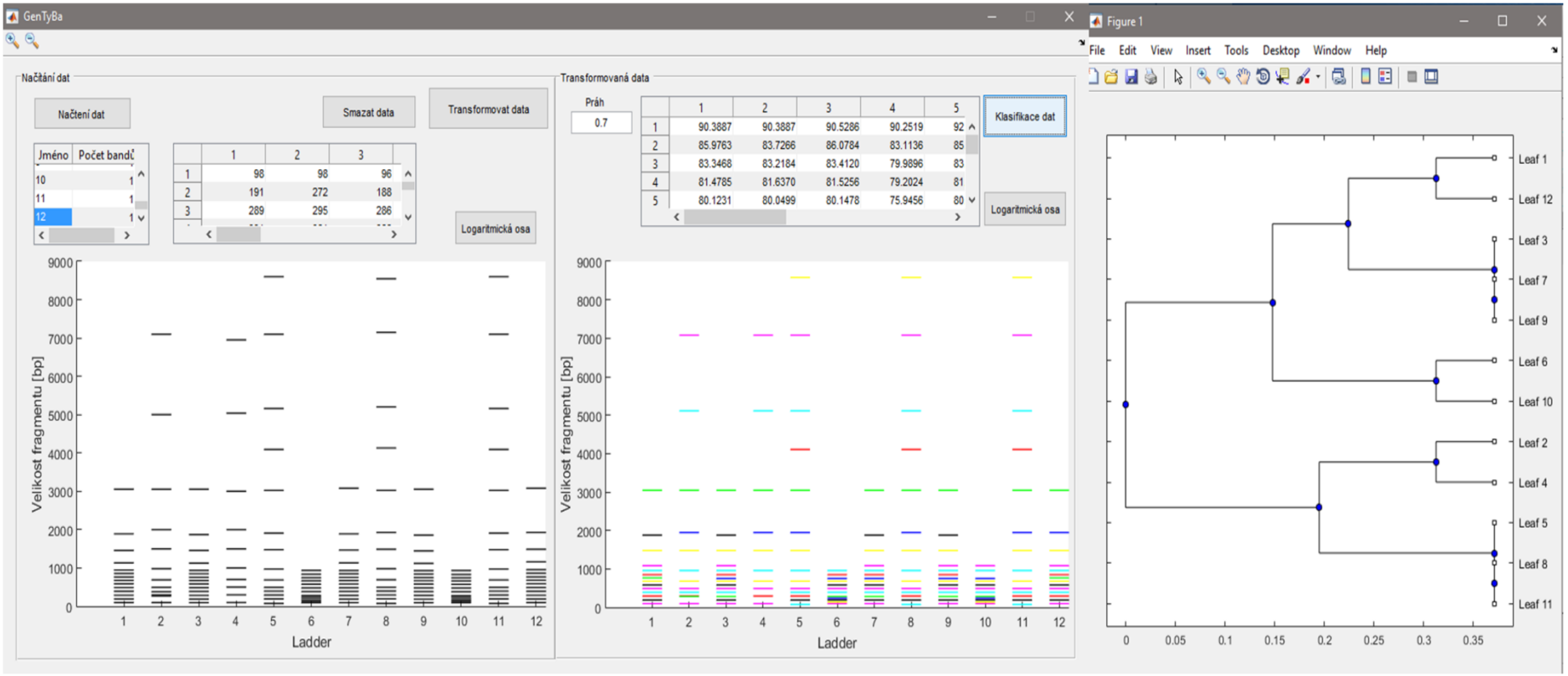
3.5 Uživatelské rozhraní

Pro lepší orientaci a manipulaci s algoritmem bylo vytvořeno uživatelské rozhraní, které je možné nainstalovat jako samostatnou aplikaci do PC. Logo programu je na Obr. 35. Vstupní data je možné vybrat ze složky a jsou požadována v souboru .csv. Podoba tohoto souboru je dána výstupem z programu Bioanalyzer, který je nutný k zisku dat z čipové kapilární elektroforézy. Výstup ze programu lze uložit a to formou .m souboru, který lze v programovém prostředí Matlab opět otevřít.

Zásadní pro fungování programu je nastavení použitého ladderu při získávání dat. Po spuštění programu je nutné zvolit soubor, který obsahuje časový průběh ladderu. Po volbě souboru jsou ihned automaticky nedetekovány píky a je nutné k nim dopsat standardizované velikosti fragmentů, které ladder má. Je nutné jej zadat k časovým okamžikům ve vzrůstajícím pořadí. Aby bylo možné pokračovat dále, nesmí být v žádném poli velikosti fragmentu nulová hodnota a charakter hodnot musí být vzrůstající. Až když je toto splněno zobrazí se tlačítko pro pokračování dále.

Na Obr. 34 je zobrazeno, jak uživatelské rozhraní v dalším kroku vypadá již vyplněné. V levé části jsou vstupní data. Po zvolení jednoho či více souborů jsou do tabulek nahrána data. V malé tabulce úplně vlevo je jméno souboru a počet nalezených bandů. Jméno je editovatelné, jakmile je změněno, upraví se název i v druhé tabulce a na ose pod umělým gelem. V druhé tabulce v části načítání dat je možné měnit a upravit pozice bandů. V prostřední části jsou již upravená data, ty se do tabulek a grafu přepočítají po stisku tlačítka „Transformovat data“. Lze zde také upravit práh pro shlukování při úpravě pozic bandů. V pravé části je poté znázorněn fylogenetický strom odpovídajícím datům. Fylogenetický strom se otevírá v novém okně po stisku tlačítka „Klasifikace dat“.

Na konci práce je přiložena uživatelská příručka pro lepší manipulaci s programem.



Obr. 34: Uživatelské rozhraní programu GenTyBa

3.6 Návrh a testování programového rozhraní

Program GenTyBa detekuje v signálu píky, které mají výšku větší než 4,25 FU a prominenci min. 4,5. Dále uchovává časové okamžiky těchto píků a následně je převádí na velikost fragmentů podle uživatelem zadaného ladderu. Mezi dvěma známými hodnotami ladderu, jsou hodnoty proloženy kubickou závislostí. V proložené oblasti je nalezena tížená hodnota pro naše bandy, resp. vzorky. V tuto chvíli lze již vykreslit umělý gel. Při řešení této části algoritmu nebylo moc inspirace v odborné literatuře. Většina článků o typizaci je doposud vyhodnocována pomocí klasického gelu, proto se zabývá úpravou obrazu, ne spojitou časovou závislostí. Někteří autoři převádí gel na spojitý signál díky intenzitě odpovídající proužku vzorku (např. [30] a [31]), práce se signálem se ale liší.

Uměle vytvořený gel je nezarovnaný a při fylogenetickém shlukování v tento okamžik by docházelo ke zkreslování neshodností pozicí bandů. Vzorky jsou upraveny transformační funkcí podle vzorce (4).

$$H2 = \frac{H}{f(H)^{1.075}} \tag{4}$$

kde *H2* je upravená hodnota bandu, *H* je původní hodnota bandu a f(*H*) je hodnota transformační funkce pro pozici *H.* Úprava pozic pokračuje výpočtem euklidovských vzdáleností (viz vzorec (5)) mezi bandy a následnou shlukovou analýzou metodou WPGMA. Tento krok byl volen podle subjektivního testování více metod pro shlukování. Následně byla zařazena podmínková funkce kontrolující příslušnost max. jednoho bandu ze vzorku k jedné velikostní skupině. Získáme upravený gel.

Výpočet euklidovské vzdálenosti probíhá následovně

$$e = \sqrt{(x_1 - x_2)(x_1 - x_2)} \tag{5}$$

kde *e* je euklidovská vzdálenost, *x1* je velikost prvního bandu a *x2* druhého.

Z upraveného gelu je možné sestavit asociační matici pro shlukování. Nabízelo se velké množství koeficientů, které bylo možno použít. Zvolen byl takový koeficient, aby jím byla vyjádřena vzdálenost. Proto je asociační matice počítána podle následujícího vzorce

$$m_{ij} = \frac{b + c}{a + b + c} \tag{6}$$

kde m_{ij} je odpovídající pozice v asociační matici, *b* (*c*) je neshoda přítomnosti bandu 01 (10) a člen *a* je shoda přítomnosti bandu 11. Je naprosto vynechán člen 00. Ten je součástí problému double zero, v tomto případě je vhodné jej vynechat. 00 značí nepřítomnost bandu v obou vzorcích, je to shoda. Nepřítomnost bandu pro nás ale není směrodatným ukazatelem, zbytečně by snižoval vzdálenost dvou nepodobných vzorků.

Fylogenetická klasifikace pomocí sestavené asociační matice je provedena metodou UPGMA. Volba typu shlukové analýzy proběhla inspirací v článcích [30], [32] a testováním.

Výsledkem je již dendrogram, podle nějž je možné určit podobnost vzorků a jejich zařazení do podtypů kmenů bakterií.

Na Obr. 36, Obr. 37 jsou zobrazeny fylogenetické stromy vybraných reálných vzorků. Stromy byly sestaveny různými způsoby, v jejich nadpisu je zkrácená specifikace změny oproti programu GenTyBa. První fylogenetický strom, nazvaný Originál je sestaven přímo programem. Při podrobném porovnání všech stromů je nejlépe sestaven.

Strom z Obr. 36 s titulkem „Detektor Bioanalyzer" je sestaven bez vlastní detekce píků programu. Vstupní data byla ve formě diskrétních velikostí fragmentů, definované detektorem z programu Bioanalyzer. Ve stromu jsou pouze kmeny L, P a S zařazeny stoprocentně správně. Kmeny I a J nejsou rozlišeny, odlišnost těchto kmenů je velice nepatrná (kmen J má pouze o jeden pík navíc, jak bylo zmíněno dříve na Obr. 21 a bude poukázáno i dále viz Obr. 41). U kmene H a N je úspěšnost klasifikace nulová. Naopak, a to je jediný pozitivní úkaz stromu, kmen S má u obou vzorků naprostou shodu, nejsou od sebe vzdáleny.

Další strom s titulem „Detekce pouze výšky píků" je ukázkou kvality detektoru. Strom při detekci pouze výškou bez dalších podmínek je pro kmen J, P, L a H stoprocentní. Nulová je klasifikace kmenu S a kmenu M má 50% úspěšnost (2 ze 4 vzorků jsou zařazeny do shluku). Testování a příprava detektoru probíhaly s ohledem na velké množství parametrů, aby byla nalezena optimální hodnota. Je ukázáno, že použití detekce s ohledem pouze na velikost píku je nevhodné.

Strom s názvem „Bez trans. f., zvýšený práh 20" napovídá, že při jeho sestavování byla z programu vyjmuta transformační funkce pro úpravu dat. Práh pro shlukování bandů je při jejím použití nastaven na 0,7. V tomto případě nebyla transformační funkce použita a zmíněný práh byl velmi navýšen (na hodnotu 20). 100% vzorků bylo správně zařazeno u kmenů M, I, J, S a P. Bez barevného označení vzorků by kmeny I a J byly brány spíše jako jeden shluk. Vzorky z kmenu L jsou od sebe velice vzdáleny. Transformační funkce zlepšuje klasifikaci a díky ní jsou zvýrazněny rozdíly mezi kmeny.

Jak bylo zmíněno dříve, jedním z možných postupů bylo porovnání celých signálů. Pro ukázku je zde zařazen strom, který je sestaven na základě asociační matice převrácených hodnot korelací signálů (Obr. 37, strom „Asociační matice korelací"). Až na kmeny I, J a M jsou všechny shlukovány správně. Bez správného označení vzorků, by shluky nebyly rozděleny (zejména kvůli délce větví stromu). Kmeny L, K a S by jistě byly zařazeny do společného shluku. Kmen M je vmíchán mezi I a J. Největší nevýhodou stromu shledávám ve velkých délkách větví, neumožňujících stanovení přibližného prahu pro oddělení shluků.

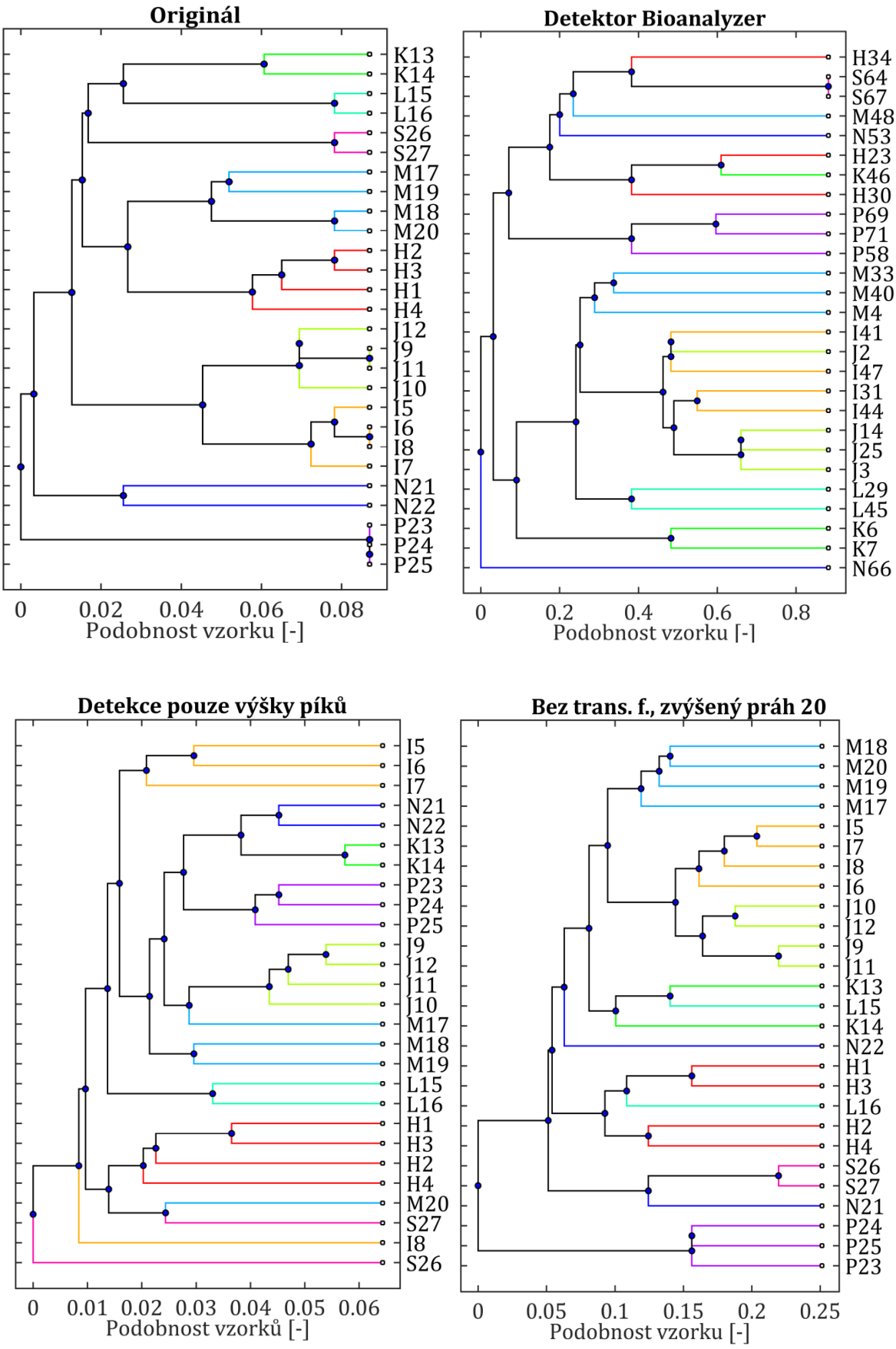
Dalším stromem je „Shluk bandů complete linkage". Jde o strom, na němž chci demonstrovat testování použitých metod pro shlukovou analýzu při úpravě pozic bandů. Metoda nejvzdálenějších sousedů nevhodně shlukovala bandy. Asociační matice je tím pádem zkreslená a celková fylogenetická klasifikace scestná. Těžko bychom hledali jednotný práh pro volbu odlišných shluků kmenů.

Předposlední fylogenetický strom „Asociační matice Dice koeficient“ se liší od programu GenTyBa použitým koeficientem pro sestavení asociační matice. Pro tento případ se skládá z převrácených Dice koeficientů (použitý k porovnávání vzorků v [30]). Dice koeficient se liší od koeficientu použitého v algoritmu GenTyBa jmenovatelem. Dává důraz na shody, proto je použita jeho převrácená hodnota, aby byl dodržen koncept fylogenetické klasifikace probíhající na základě vzdáleností. Zařazení vzorků do shluků více méně odpovídá realitě. Pouze kmen M je rozdělen do dvou samostatných shluků a kmen N je neseskupený. Opět je zde ale problém ve volbě prahů pro rozdělení shluků. Ukazuje se, že postavit hodnocení na shodě, není ideální.

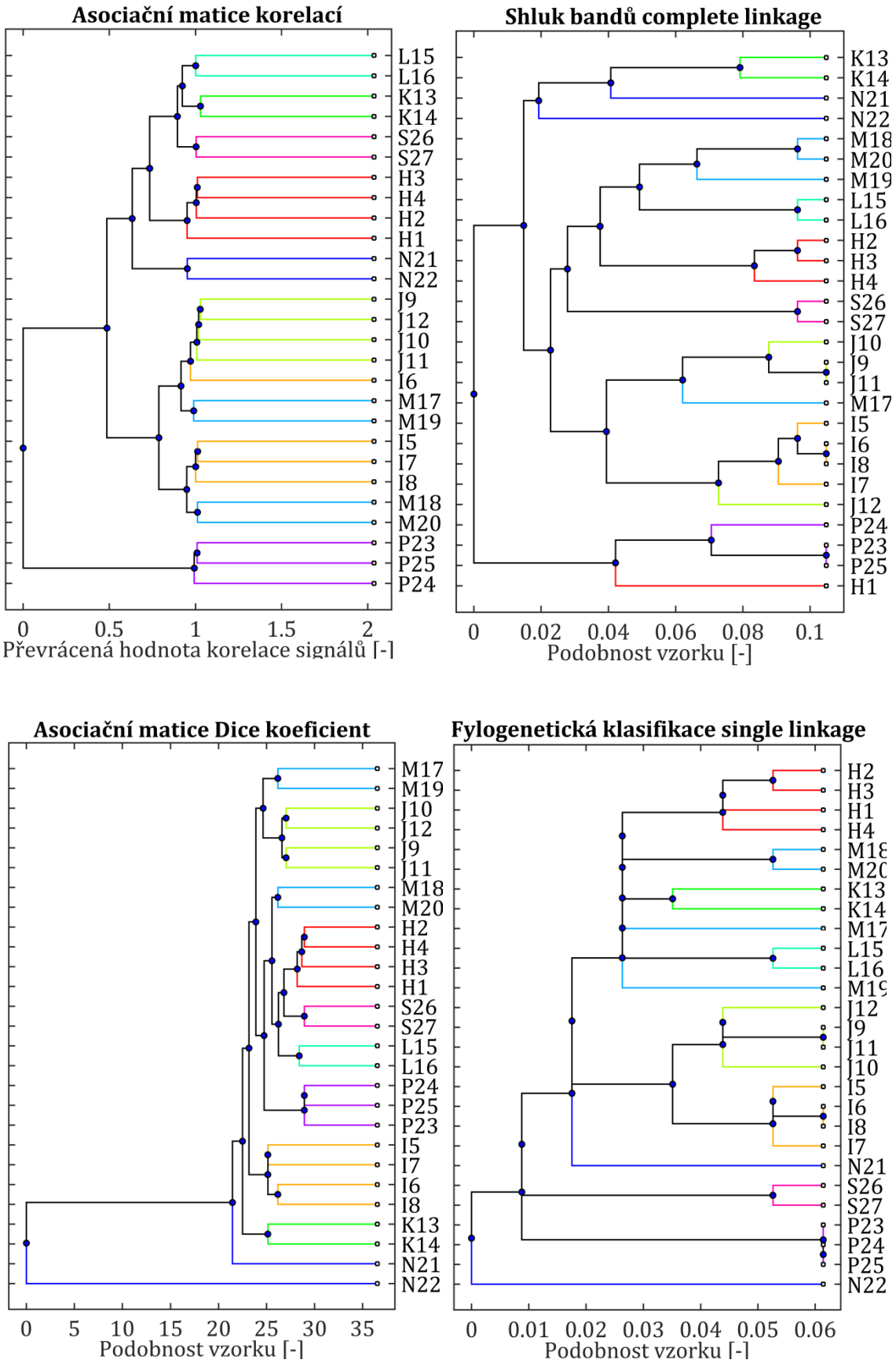
Poslední ukázka „Fylogenetická klasifikace single linkage“ ukazuje změnu v samotném závěru algoritmu. UPGMA je nahrazeno metodou nejbližšího souseda. Kmeny I, J, S a P jsou dostatečně odlišitelné a 100% správně klasifikované. Kmen M a N se nepodařilo vhodně zařadit. Při porovnání stromu s „Originálem“ je vidět, že všechny kmeny kromě M a N mají shluky stejné, s odlišností délky větví. Metoda se opět ukazuje pro klasifikaci méně vhodná než UPGMA.

Na těchto příkladech byl demonstrován zlomek testovaných možností. Je hodně parametrů (použitá vstupní data, způsob nastavení detektoru, shlukovací metriky, koeficienty, transformační funkce, a další), které celkový proces automatické getonypizace bakterií metodou rep-PCR ovlivní. Všechny sestavené stromy na Obr. 36 a Obr. 37 se od originálního programu lišily pouze v jednom parametru. Při kombinaci více nevhodných parametrů zvýrazňuje chyby a dochází k ještě k horším výsledkům. Program GenTyBa byl sestaven z nejvhodnější kombinace parametrů, které z testování vyplynuly.

Obr. 35: Logo programu GenTyBa



Obr. 36: Ukázky fylogenetických stromů sestavených odlišnými způsoby



Obr. 37: Ukázky fylogenetických stromů sestavených odlišným způsobem 2

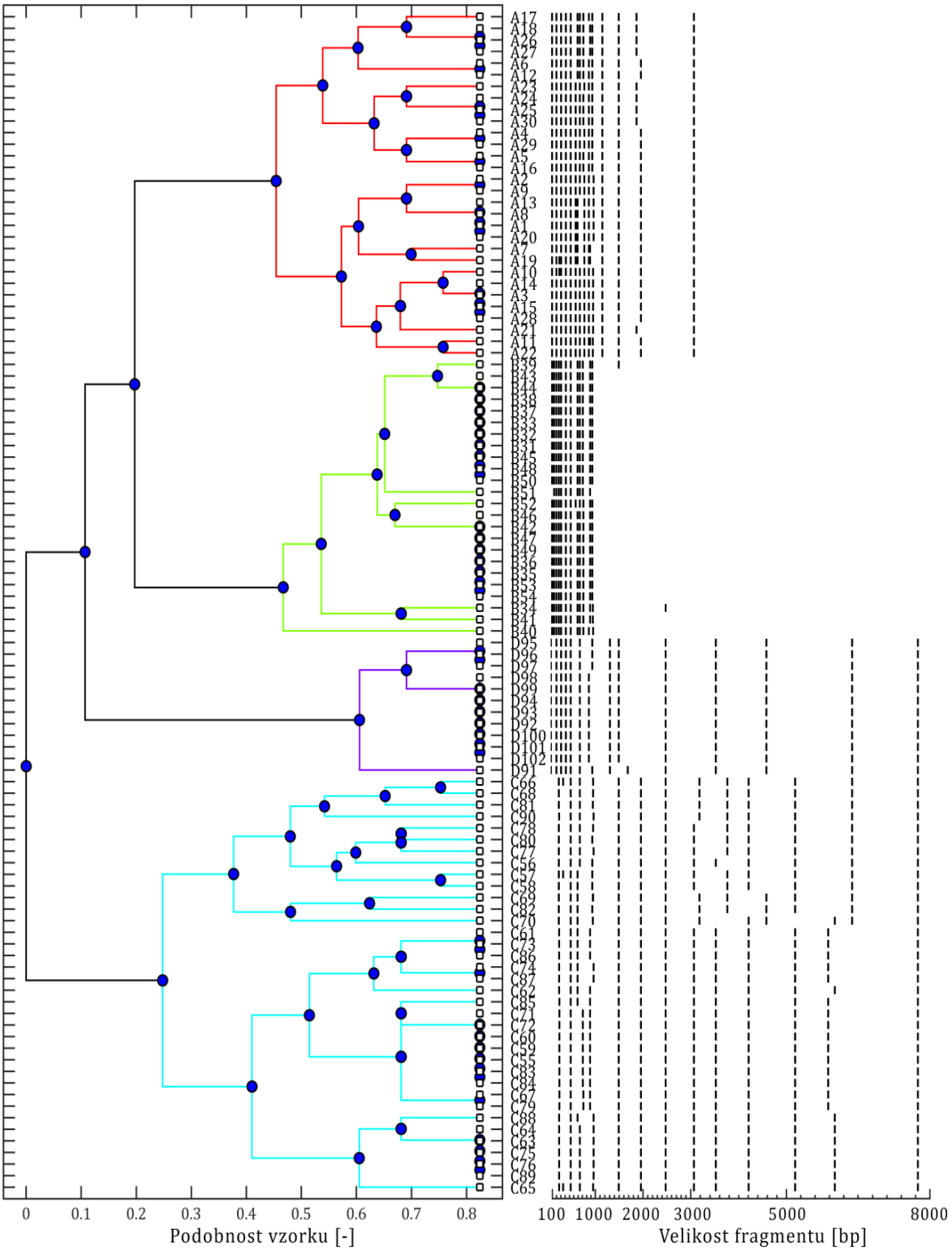
3.7 Výsledky a diskuze

Program byl sestaven na standardizovaných měřeních DNA markerů. Laddery měly různé rozsahy velikostí fragmentů. U dvou ladderů se rozsahy shodovaly, další dva obsahovaly výrazně kratší délky fragmentů. Konkrétní délky fragmentů byly uvedeny dříve v Tab. 6. Klasifikace těchto ladderů je bezchybná. Na Obr. 38 je vidět celý dendrogram s porovnávaným umělým gelem. Je vidět, že profily ladderů se dost výrazně odlišují a jejich klasifikace je správná. Ve fylogenetickém stromu se vytvoří 4 klastry.

Při nahlédnutí o krok zpět před klasifikaci do úpravy pozic bandů je accurency nižší než by se očekávalo. Můžeme si všimnout i velké délky větví ve shlucích, zejména pak u skupiny ladderů C (Obr. 38). Předzpracování není stoprocentní, přesto je vidět, že zařazení malého množství bandů do jiné skupiny velikostí fragmentů neovlivní výslednou klasifikaci.

Tab. 7: Výsledky fylogenetické klasifikace standardizovaných DNA markerů

	Jméno hmotnostního markeru	Velikost fragmentů [bp]	Počet měřených vzorků	ACC (úprava pozic bandů)	Procentuální úspěšnost fylogenetické klasifikace [%]
A	GeneRuler 100 bp DNA Ladder	100-3000	30	0,975	100
B	GeneRuler 50 bp DNA Ladder	50-1000	36	0,932	100
C	O'GeneRuler 1 kb DNA Ladder	250-10000	12	0,979	100
D	GeneRuler 1 kb DNA Ladder	250-10000	24	0,993	100



Obr. 38: Klasifikace ladderů, v levé části fylogenotický strom, v pravé části umělý gel

Z Dětské nemocnice byla poskytnuta data 72 vzorků a výsledky fylogenetické klasifikace (programu BioNumerics [33], který v Dětské nemocnici použili) velké části těchto dat. Při porovnání s jejich klasifikací byl program GenTyBa úspěšnější. Celkově došlo ke správné klasifikaci v 79,16% (přísnější hodnocení, které bere jako správně zařazen pouze kmen, který má všechny vzorky správně klasifikovány). Výsledek je to velice dobrý s ohledem na naměřená data.

Na Obr. 39 je znázorněna klasifikace několika vzorků. Klasifikace kmenů proběhla správně až na jeden vzorek N9. Při prohlédnutí umělého gelu je ale patrné, že profil vzorku se od ostatních vzorků daného kmene odlišuje. Při zkoumání důvodu této špatné klasifikace bylo zjištěno, že špatná jsou již vstupní data. Na Obr. 40 je patrné, že vzorek N9 má velmi odlišný průběh než vzorek N8. Proto je vlastně zařazení vzorku N9 mimo kmen N správné. Při celkovém hodnocení (tj. 79,16% úspěšnost klasifikace) nejsou chyby měření vstupních dat brána v potaz, i proto hodnotím tento výsledek jako velice dobrý.

Tab. 8: Výsledky fylogenetické klasifikace reálných dat

Kmen *Klebsiella pneumoniae*	Počet vzorků	Procentuální úspěšnost klasifikace [%]	Procentuální úspěšnost klasifikace z Dětské nemocnice [%]
G	5	100	40
H	6	100	83
I	10	100	50
J	7	100	43
K	6	100	100
L	4	100	100
M	10	100	90
N	5	80	Neklasifikováno
O	5	40	100 (4 ze 4)
P	5	100	100 (3 ze 3)
Q	5	40	0 (0 ze 3)
R	2	100	Neklasifikováno
S	2	100	Neklasifikováno

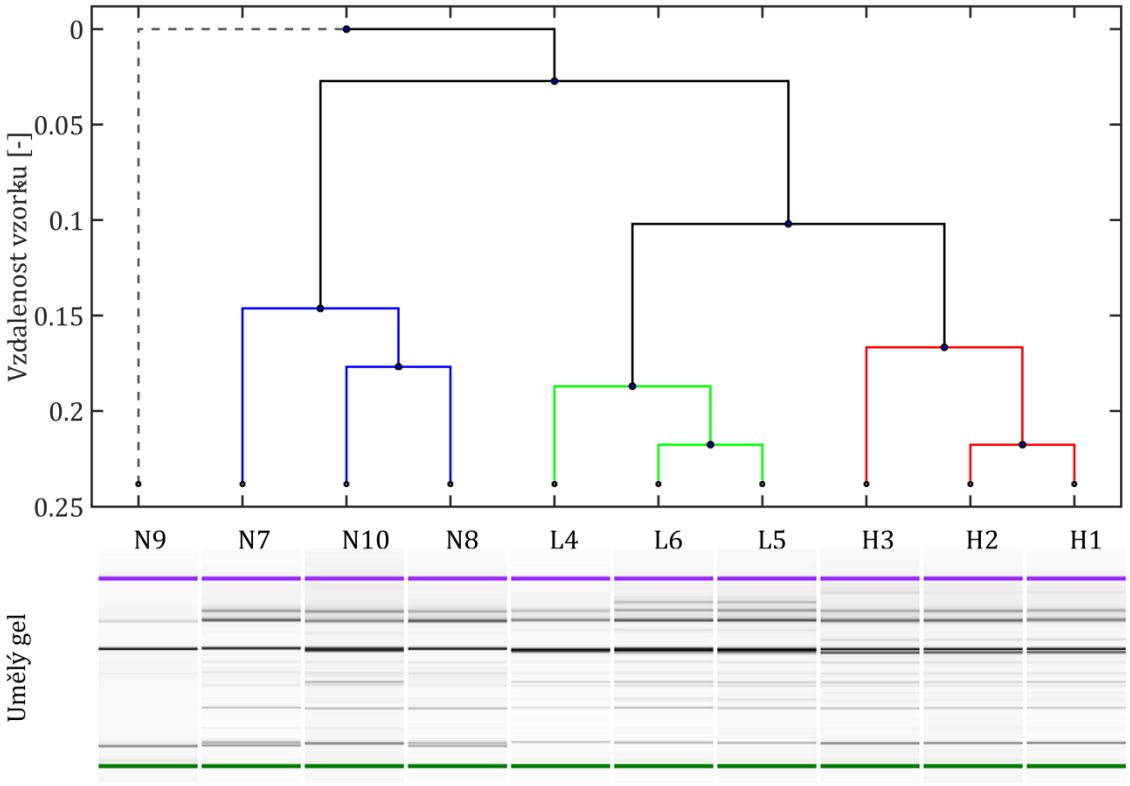
Celková klasifikace je znázorněna na Obr. 42. Umělý gel reálných dat není tak jednoznačně odlišitelný, jako v případě klasifikace ladderů. Přesto jsou patrné odlišnosti v datech a klasifikace proběhne ve většině případů úspěšně.

V Tab. 9 jsou uvedeny úspěšnosti klasifikace pro jednotlivé reálné vzorky. Statistické vyhodnocení specificity pro reálná data vyšlo na 0,999 a sensitivita na 0,903.

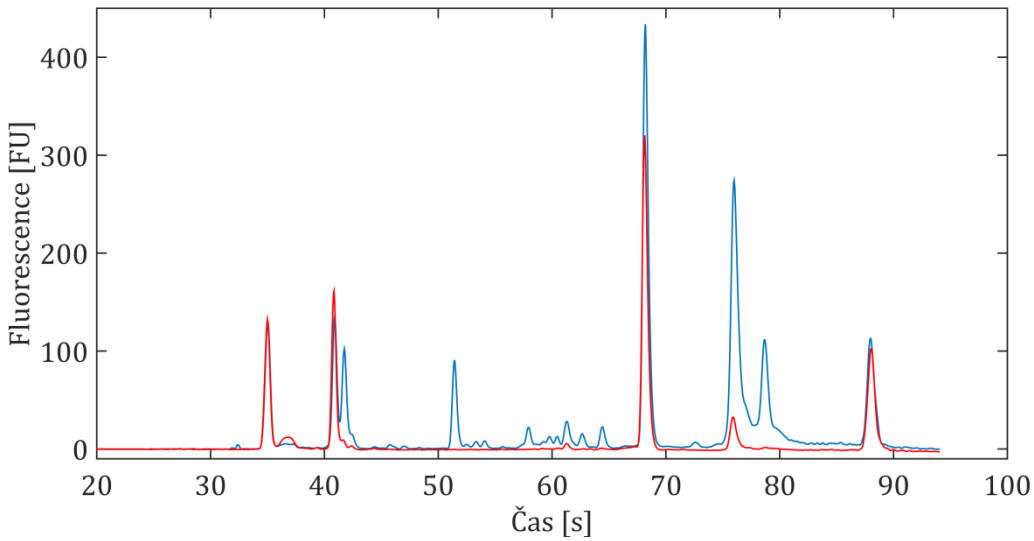
Tab. 9: Statistické hodnocení správnosti klasifikace vzorku

(TP správně pozitivní; TN správně negativní; TF špatně negativní; FP špatně pozitivní)

	G	H	I	J	K	L	M	N	O	P	Q	R	S	Součet
TP	5	6	10	7	6	4	10	4	2	5	2	2	2	65
TN	67	66	62	65	66	48	62	67	67	67	66	70	70	796
FP	0	0	0	0	0	0	0	0	0	0	1	0	0	1
FN	0	0	0	0	0	0	0	1	3	0	3	0	0	7

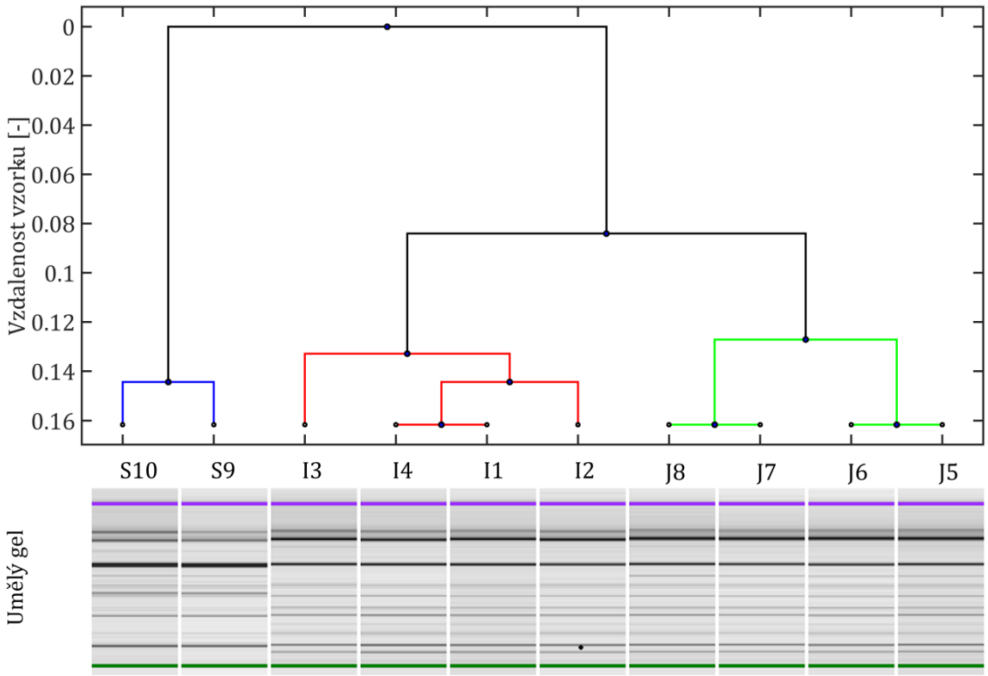


Obr. 39: Ukázka klasifikace s špatně naměřenými vstupními daty

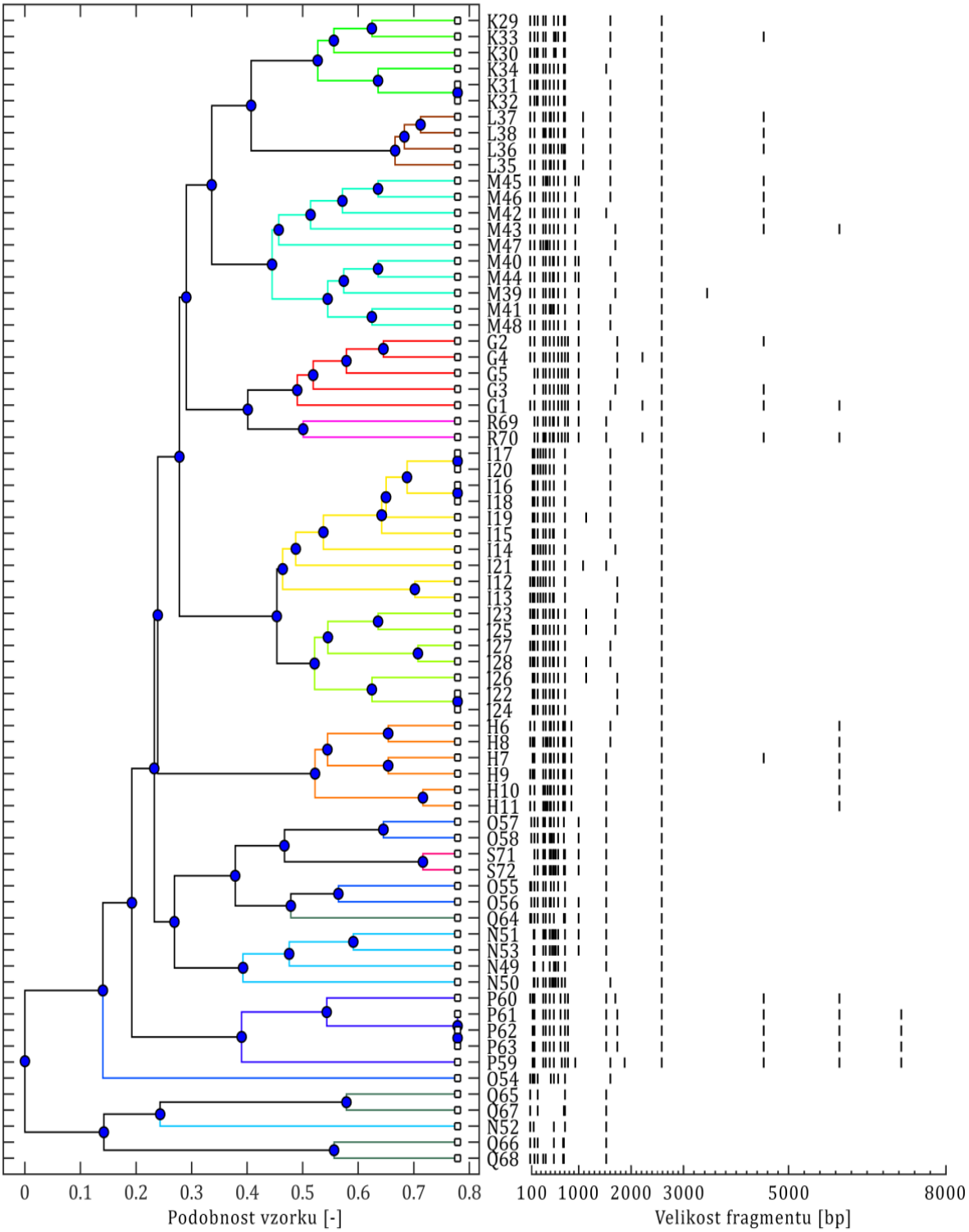


Obr. 40: Ukázka vstupních dat vzorků N8 a N9

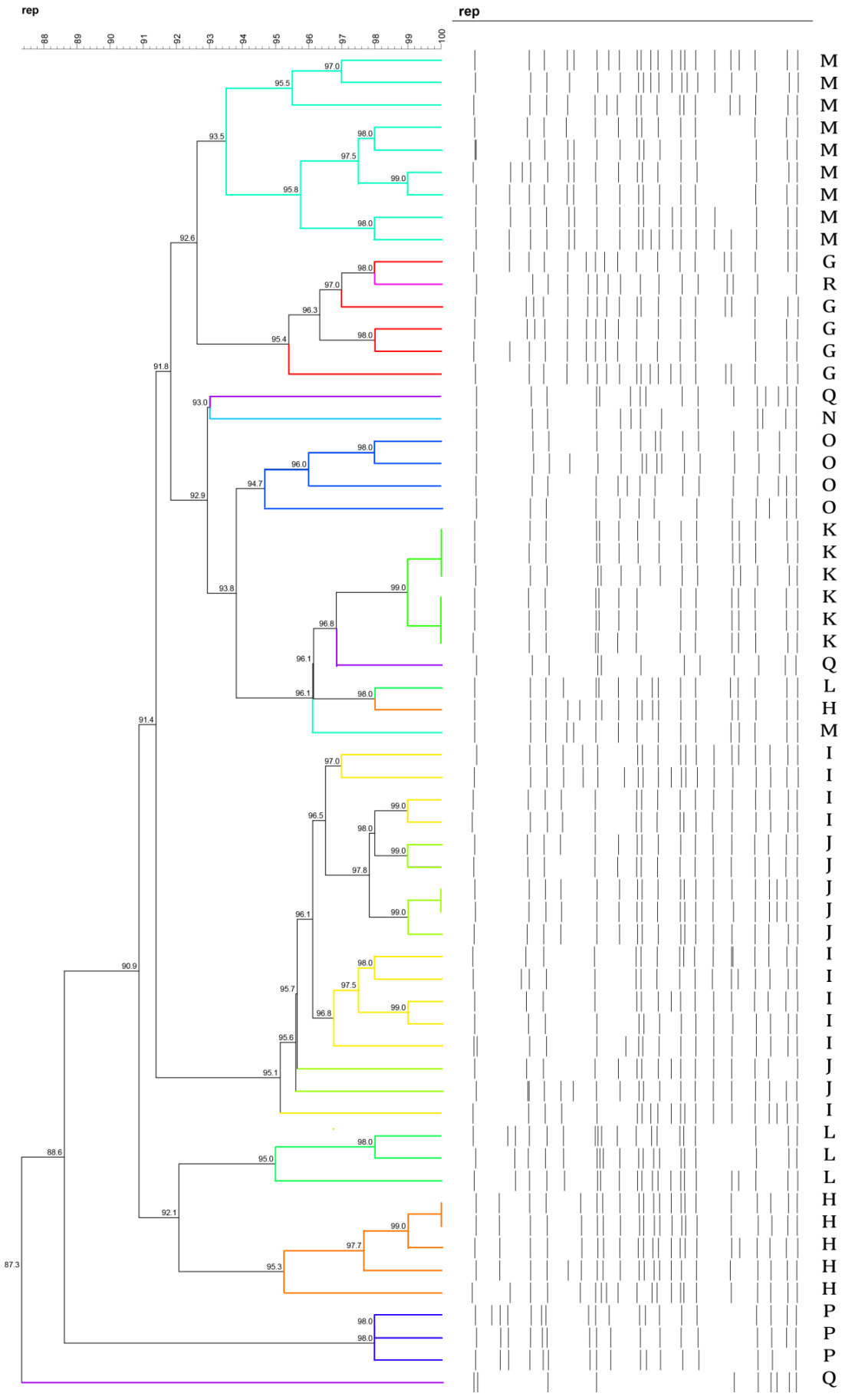
Další příklad (Obr. 41) znázorňuje úspěšnost klasifikace velmi podobných profilů kmenů bakterií. Kmen S má od prvního pohledu jiný profil, je zařazen do samostatné skupiny s velkou vzdáleností vzorků od kmenů I a J. Kmeny I a J mají na první pohled shodné umělé gely. Při podrobnějším prozkoumání je ale u kmenu J výraznější band navíc. Program GenTyBa dokázal tuto odlišnost zachytit a kmeny klasifikoval správně.



Obr. 41: Ukázka klasifikace velmi podobných profilů bakterií



Obr. 42: Fylogenetická klasifikace reálných dat, vlevé části fylogenetický storm, v pravé části umělý gel



Obr. 43: Analýza dat v programu BioNumerics – výsledek poskytnutý z Dětské nemocnice

4 Závěr

V rámci diplomové práce byl sestaven nový algoritmus pro automatickou genotypizaci bakterií metodou rep-PCR, vytvořený pro účely Dětské nemocnice v Brně. Program, nazvaný GenTyBa (Genová typizace bakteriálních kmenů), je opatřen uživatelským rozhraním pro jednodušší užívání v klinické praxi. Lze jej nainstalovat samostatně do PC, aniž by měl uživatel software Matlab. Díky programu je možné využít výsledky z rep-PCR a tím zrychlit a zjednodušit fylogenetickou klasifikaci vzorků oproti PFGE, která je zlatým standardem typizace. Při porovnání je PFGE technicky i časově náročnější (trvá cca 32 hodin čistého času) než rep-PCR (zhruba 11 hodin čistého času). Rep-PCR je díky programu GenTyBa možné použít pro rutinní srovnávání bakteriálních profilů.

Předpokládaná jsou vstupní data v rawe formátu v podobě spojitého signálu závislosti fluorescence na čase, jak je sestaví program Bioanalyzer, příslušející k čipové kapilární elektroforéze Agilent Technologies. Program je sestaven ze tří zásadních bloků – digitalizace dat, úpravy pozic bandů a fylogenetické klasifikace. Všechny kroky byly testovány a optimalizovány, aby bylo dosaženo co nejlepších výsledků fylogenetické klasifikace bakteriálních kmenů.

Digitalizace dat spočívá v detekci píků a převodu časových okamžiků na velikostní fragmenty podle standardizovaného hmotnostního markeru. Detektor píků, sestavený speciálně pro program GenTyBa, je pro klasifikaci vhodnější a citlivější než detektor obsažený v softwaru Bioanalyzer patřící k čipové kapilární elektroforéze.

Unikátnost celého navrženého programového rozhraní je v části úpravy pozic bandů. Ta je, na rozdíl od všech ostatních technologií sloužících k podobnému účelu, založena na transformaci dat pomocí transformační funkce, která byla sestavena na 108 standardizovaných měření DNA hmotnostních markerů. Díky nim bylo možné přesně zjistit profil rozptylů velikostí bandů na konkrétních hladinách a rozptyly kompenzovat. Na reálných datech (72 vzorcích patřících do 13 kmenů) byl testován vliv této korekce v aplikaci fylogenetické klasifikace bakteriálních vzorků. Všechna měření byla provedena certifikovanou laboratoří v Dětské nemocnici. Díky použití transformační funkce, shlukové analýzy WPGMA a podmínkové funkce založené na k-means shlukování, jsou hodnoty bandů upraveny a získáváme zarovnaný umělý gel.

Upravený umělý gel může sloužit jako vstup do bloku fylogenetické klasifikace dat. Podle gelu je sestavena asociační matice, která porovnává vzorky podle shod a neshod přítomnosti bandů, vynechává shodu typu 00, čímž je řešen problém double zero, který by mohl uměle zvyšovat shodu vzorků. Shluková analýza UPGMA vytvořenou matici použije k fylogenetické klasifikaci. Výsledkem je dendrogram (fylogenetický strom) znázorňující podobnost vzorků.

Program byl sestaven na základě dat hmotnostních standardů. Při klasifikaci těchto standardů jsou všechny vzorky klasifikovány do správné skupiny, úspěšnost je 100 %.

Otestování programu pomocí reálných dat dopadlo velmi dobře. Celková úspěšnost klasifikace je 79 %. Při porovnání s programem BioNumerics, patřící k čipové elektroforéze, bylo dosaženo lepších výsledků. Program GenTyBa dokázal na rozdíl od softwaru BioNumerics rozlišit dva téměř stejné kmeny lišící se v jednom bandu. Pouze u 3 kmenů došlo ke 100% správné klasifikaci u obou programů. Program GenTyBa klasifikoval 6 kmenů výrazně lépe (5 kmenů 100 % vzorků, 1 kmen 40 % vzorků) a pouze v jednom případě došlo k horší klasifikaci oproti programu BioNumerics.

Specificita programu GenTyBa dosahuje 0,999 a senzitivita 0,903. S ohledem na chyby měření, které se v souboru dat vyskytují (odlišné signály patřící do stejného kmene), je dosažený výsledek velice dobrý. Chybná vstupní data byla ve statistickém hodnocení ponechána, v důsledku toho dochází k umělému snižování i tak dobré úspěšnosti. Při rutinním použití bych doporučila opakovat měření, kvůli eliminaci těchto primárních chyb.

Program je sestaven k možnému použití v klinické praxi. V Dětské nemocnici by mohl urychlit porovnávání bakteriálních profilů, pomoci ke zjištění četnosti infekcí způsobených kolonizujícími kmeny, mapovat jejich šíření po nemocnici a tím zajistit pro pacienty lepší péči.

Literatura

[1] OLIVE, Michael D. a Pamela BEAN. Principles and Applications of Methods for DNA-Based Typing of Microbial Organisms. Journal of Clinical Microbiology. 1999, 37(6), 1661-1669.

[2] MCDOWELL, D. 2006. The polymerase chain reaction patents: going, going,... still going. *Journal of the Royal Society of Medicine*. 99(2), 62-64.

[3] SCHWARTZ, David C. a Charles R. CANTOR, 1984. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. Cell. 37(1), 67-75. DOI: 10.1016/0092-8674(84)90301-5. ISSN 00928674. Dostupné také z: http://linkinghub.elsevier.com/retrieve/pii/0092867484903015

[4] Pulsed Field Gel Electrophoresis [online]. UVM Genetics & Genomics Wiki is a FANDOM Lifestyle Community. Available at: http://uvmgg.wikia.com/wiki/Pulsed_Field_Gel_Electrophoresis

[5] GOODWIN, Milo. RFLP (Restriction Fragment Length Polymorphism) [online]. Available at: http://slideplayer.com/slide/9282493/28

[6] BOTSTEIN, DAVID, RAYMOND L. WHITE, MARK SKOLNICK a RONALD W. DAVIS, 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics*. 32(3), 314-331. PMC1686077. Dostupné také z: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1686077/?page=2

[7] Genetic analysis using random amplified polymorphic DNA markers, 1993. WILLIAMS, John G.K., Michael K. HANAFEY, J. Antoni RAFALSKI a Scott V. TINGEY. *Methods in enzymology*. San Diego, Calif: Academic Press, s. 704-740. ISBN 9780121821197.

[8] CULTIVAR IDENTIFICATION AND VARIETAL TRACEABILITY IN PROCESSED FOODS: A MOLECULAR APPROACH. 2013. *Cultivars: chemical properties, antioxidant activities and health benefits*. Hauppauge, N.Y.: Nova Biomedical, p. 83-105.

[9] HEISLER, Laura and Chao-Hung LEE. 2002. Cleavase® Fragment Length Polymorphism Analysis for Genotyping and Mutation Detection. *PCR Mutation Detection Protocols*. New Jersey: Humana Press, , 165-178.

[10] Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Kuiper, M. (1995). AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research, 23(21), 4407–4414.

[11] BLEARS, M J, S A DE GRANDIS, H LEE a J T TREVORS, 1998. Amplified fragment length polymorphism (AFLP): a review of the procedure and its applications. *Journal of Industrial Microbiology and Biotechnology*. 21(3), 99–114. DOI: 10.1038/sj.jim.2900537. ISBN 10.1038/sj.jim.2900537. Dostupné také z: http://link.springer.com/10.1038/sj.jim.2900537

[12] VERSALOVIC, James, Thearith KOEUTH a R. LUPSKI, 1991. Distribution of repetitive DNA sequences in eubacteria and application to finerpriting of bacterial enomes. *Nucleic Acids Research*. 19(24), 6823-6831. DOI: 10.1093/nar/19.24.6823. ISSN 0305-1048. Dostupné také z: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/19.24.6823

[13] DiversiLab strain typing. *Biomérieux* [online]. Lyon: bioMérieux. Available at: http://microtyping.nl/content/2012/11/DiversiLab-strain-typing

[14] TENOVER, Fred C., Robert D. ARBEIT a Richard V. GOERING, 1997. How to Select and Interpret Molecular Strain Typing Methods for Epidemiological Studies of Bacterial Infections: A Review for Healthcare Epidemiologists. *Infection Control and Hospital Epidemiology*. 18(6), 426-439. DOI: 10.2307/30141252. ISSN 0899823x. Dostupné také z: http://www.jstor.org/stable/info/10.2307/30141252

[15] LAUER, Henk H. a Gerard P. ROZING, 2014. High Performance Capillary Electrophoresis. 2014. Germany: Agilent Technologies, 174 s. 5990-3777EN. Dostupné také z: https://www.agilent.com/cs/library/primers/public/5990_3777EN.pdf

[16] DICKERSON, Bernard. Supercritical Fluid Chromatography Introduction [online]. Available at: http://slideplayer.com/slide/7903082/

[17] TISELIUS, A., 1937. Electrophoresis of serum globulin: Electrophoretic analysis of normal and immune sera. Biochemical Journal. 31(9), 1464-1477. Dostupné také z: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1267100/

[18] VÁCLAV, Václav, 1997. Teoretické základy a separační principy kapilárních elektromigračních metod. Chemické listy. 91(5), 320-329. ISSN 0009-2770. Dostupné také z: http://ww.w.chemicke-listy.cz/docs/full/1997_05_320-329.pdf8

[19] HJERTÉN, Stellan, 1967. Free zone electrophoresis. Chromatographic Reviews. 9(2), 122-143. DOI: 10.1016/0009-5907(67)80003-6. ISSN 00095907. Dostupné také z: http://linkinghub.elsevier.com/retrieve/pii/0009590767800036

[20] Capillary Zone Electrophoresis, 1994. POMERANZ, Yeshajahu a Clifton E. MELOAN. *Food Analysis: Theory and Practice*. Boston, MA: Springer, s. 228-242. ISBN 978-1-4615-7000-4.

[21] Elektroforéza. 2004. *Biochemie* [online]. Rowan Design. Available at: http://biochemie.sweb.cz/x/metody/elektroforeza.htm

[22] Capillary electrophoresis. *Wikipediea* [online]. Wikimedia Foundation. Available at: https://en.wikipedia.org/wiki/Capillary_electrophoresis

[23] HARRISON, D. Jed., Andreas. MANZ, Zhonghui. FAN, LUEDI a WIDMER, 2002. Capillary electrophoresis and sample injection systems integrated on a planar glass chip. Analytical Chemistry. 64(17), 1926-1932. DOI: 10.1021/ac00041a030. ISSN 0003-2700. Dostupné také z: http://pubs.acs.org/doi/abs/10.1021/ac00041a030

[24] DOLNÍK, V., S. LIU a S. JOVANOVICH, 2000. Capillary electrophoresis on microchip. *ELECTROPHORESIS*. 21(1), 41–54. DOI: 10.1002/(SICI)1522-2683(20000101)21:1<41::AID-ELPS41>3.0.CO;2-7.

[25] VANDAVEER, Walter R., Stephanie A. PASAS-FARMER, David J. FISCHER, Celeste N. FRANKENFELD a Susan M. LUNTE, 2004. Recent developments in electrochemical detection for microchip capillary electrophoresis. *ELECTROPHORESIS*. 25(21-22), 3528-3549. DOI: 10.1002/elps.200406115. ISSN 0173-0835. Dostupné také z: http://doi.wiley.com/10.1002/elps.200406115

[26] HENRY, Charles S., 2006. Microchip Capillary Electrophoresis: An Introduction. *Microchip Capillary Electrophoresis*. New Jersey: Humana Press, 339(1), 1-10. DOI: 10.1385/1-59745-076-6:1. ISBN 1-59745-076-6. Dostupné také z: http://link.springer.com/10.1385/1-59745-076-6:1

[27] LI, S. F.Y., 2006. Clinical Analysis by Microchip Capillary Electrophoresis. *Clinical Chemistry*. 52(1), 37-45. DOI: 10.1373/clinchem.2005.059600. ISSN 0009-9147. Dostupné také z: http://www.clinchem.org/cgi/doi/10.1373/clinchem.2005.059600

[28] ZÍTKA, Ondřej, Soňa KŘÍŽKOVÁ, Vojtěch ADAM, Aleš HORNA, Jiří KUKAČKA, Richard PRŮŠA, Věra ŽIŽKOVÁ a René KIZEK, 2010. Použití automatizované elektroforézy na čipu pro studium laktoferinu a matrixových metaloproteinas. *Chemické Listy*. 104(3), 197-201. Dostupné také z: http://www.chemicke-listy.cz/docs/full/2010_03_197-201.pdf

[29] *Agilent* [online]. 2018. Santa Clara: Agilent Technologies.

[30] PAVEL, Ana Brândușa and Cristian Ioan VASILE. 2012. PyElph - a software tool for gel images analysis and phylogenetics. *BMC Bioinformatics*. 13(1), 9-.

[31] INTARAPANICH, Apichart, Saowaluck KAEWKAMNERD, Philip J SHAW, Kittipat UKOSAKIT, Somvong TRAGOONRUNG and Sissades TONGSIMA. 2015. Automatic DNA Diagnosis for 1D Gel Electrophoresis Images using Bio-image Processing Technique. *BMC Genomics*. 16(Suppl 12), S15-.

[32] KHAKABIMAMAGHANI, Sahand, Ali NAJAFI, Reza RANJBAR and Monireh RAAM. 2013. GelClust: A software tool for gel electrophoresis images analysis and dendrogram generation. *Comput Methods Programs Biomed*. 111(2), 512-518.

[33] *Applied Maths* [online]. c2018. Austin: Applied Maths NV. Available at: http://www.applied-maths.com/bionumerics

Seznam obrázků

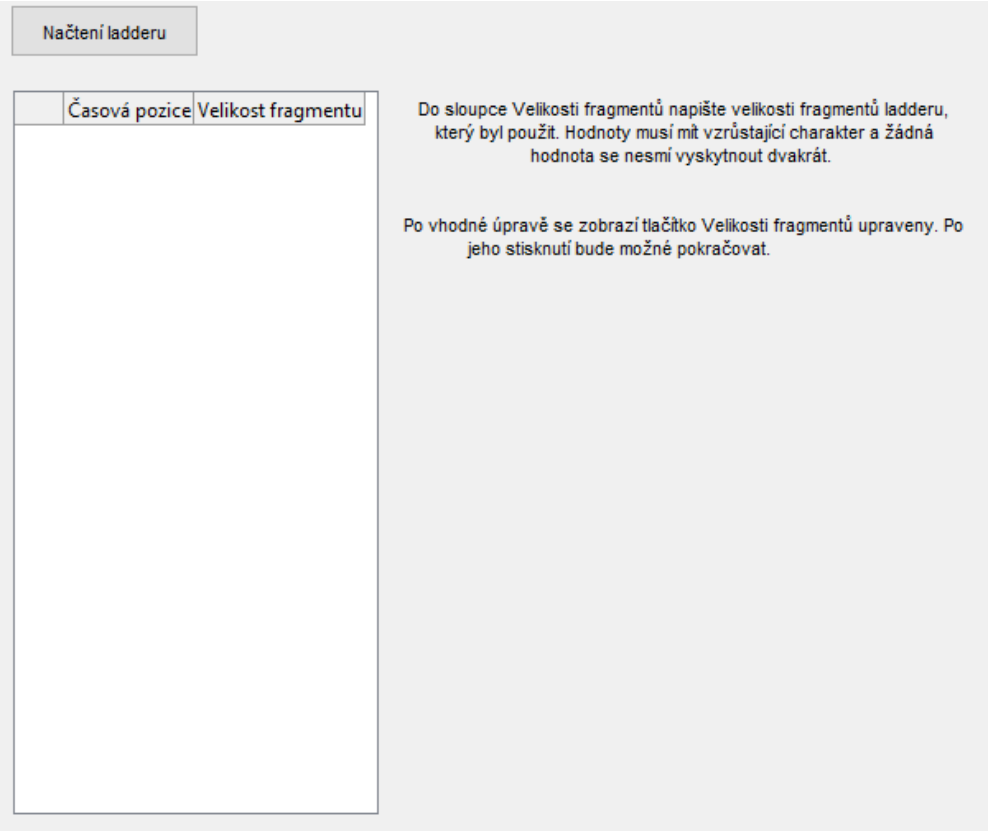
Seznam tabulek

Uživatelská příručka programu GenTyBa

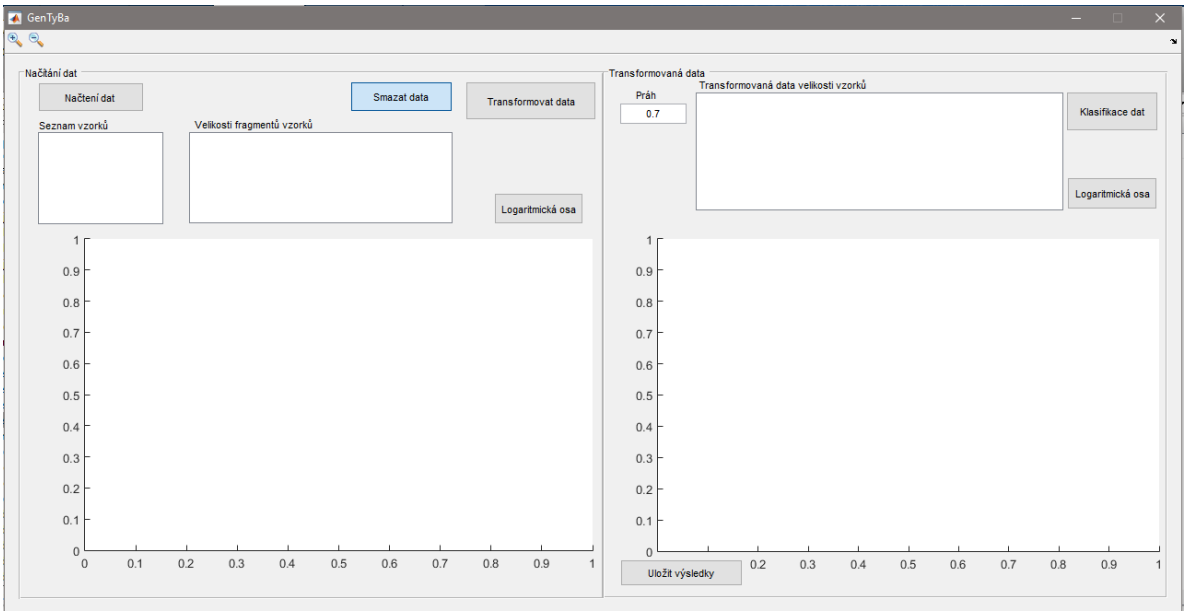
Autor programu: Veronika Pelikánová

GenTyBa je program určený pro automatickou klasifikaci bakterií metodou rep PCR. Nainstalovat si jej můžete pomocí instalačního souboru GenTyBa_instalace, který je umístěný ve složce GenTyBa\for_redistribution. Pro instalaci je nutný přístup k internetu. Vstupními daty pro program jsou soubory ve formátu .asv z programu Bioanalyzer, který slouží k obsluze čipové kapilární elektroforézy. Ukázková data jsou uložena ve složce GenTyBa\Data. Předložený ladder, použitý k zisku dat, má standardizované hodnoty [50, 100, 300, 500, 700, 1000, 1500, 2000, 3000, 5000, 7000, 10380].

Po spuštění programu naběhne prostředí znázorněné na Obrázek 1. V levém horním rohu je tlačítko „Načtení ladderu“. Po jeho stisknutí vyběhne možnost výběru souboru .csv obsahující soubor časového průběhu ladderu z čipové kapilární elektroforézy. Po načtení se vyplní tabulka. Dále je nutné doplnit do pravé části tabulky velikosti fragmentů. Ke každému časovému okamžiku je třeba doplnit vzestupně velikost standardizované hodnoty udávané výrobcem ladderu, který byl použit při běhu elektroforézy, z nějž budou další data analyzována. Po správném zadání se zobrazí v pravé části tlačítko „Velikosti fragmentů upraveny“. Po zadání hodnot a stisknutí tohoto tlačítka program zobrazí další část.

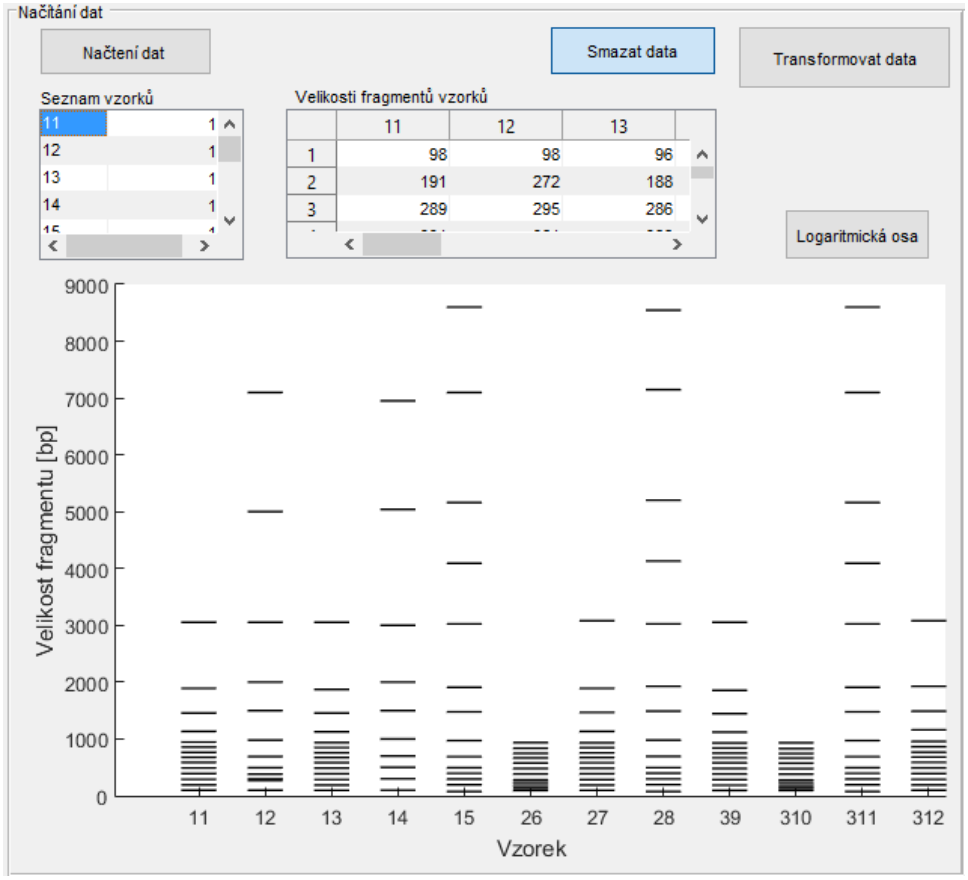


Obrázek 1: Úvodní obrazovka programu (načtení ladderu)



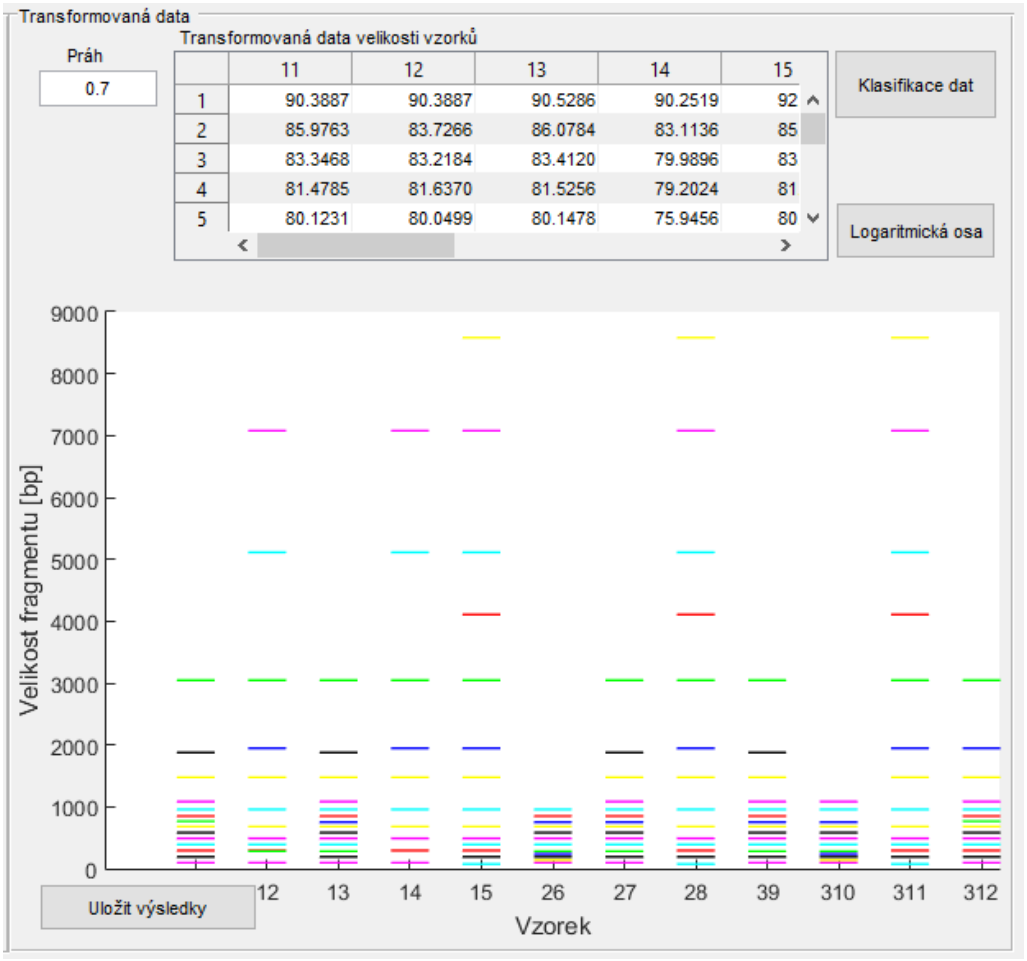
Obrázek 2: Základní obrazovka programu

Další prostředí programu je znázorněné na Obrázek 2. V levém horním rohu je tlačítko „Načtení dat“. Po jeho stisknutí vyběhne možnost výběru souboru .csv, v níž jsou nahrané soubory časových průběhů z čipové kapilární elektroforézy. Je možné zvolit jeden či více průběhů.



Obrázek 3: Část programu pro načtení dat

Po načtení dat se vyplní tabulky v sekci „Načtení dat“ (viz Obrázek 3). V seznamu vzorků lze upravovat jména vzorků. Po jejich úpravě se automaticky změní název v grafu i v tabulce s velikostmi fragmentů. Tabulka „Velikosti fragmentů vzorků“ je také editovatelná. Personál může zasáhnout do naměřených vzorků. Může umazat nesmyslnou hodnotu apod. Po úpravě se hodnota ihned automaticky přepíše i v grafu. V grafu je možné změnit měřítko z lineárního na logaritmické a naopak. Po kontrole dat je možné přistoupit k úpravě pozic bandů, ta se provede stiskem tlačítka „Transformovat data“ a vyplní se další část programu (viz Obrázek 4). Transformovaná data se zapíší do tabulky „Transformovaná data velikosti vzorků“. Tuto tabulku je možné opět editovat a zasáhnout tím do fylogenetické klasifikace pracovníkem dle subjektivního zvážení. Změna hodnoty se opět upraví v grafu a to případným přehodnocením příslušnosti k dané skupině velikosti fragmentů.

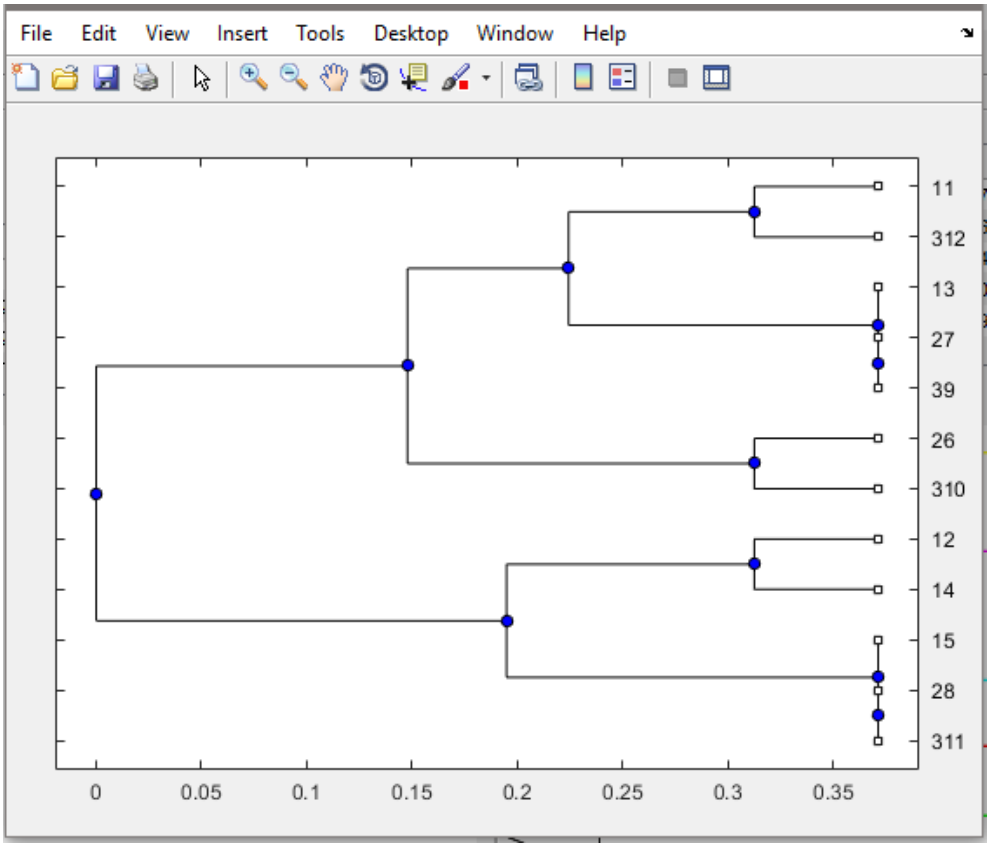


Obrázek 4: Transformovaná data

Po transformaci dat probíhá automaticky úprava pozic bandů. Pro tu je důležitý práh shlukování bandů shodných velikostních fragmentů. Ten je nastaven na 0,7, je ale možné jej upravit. Po jeho změně je nutné znovu nechat transformovat data, a tím následně upravit pozici bandů podle změněné hodnoty, aktuálního v poli „Práh“. V grafu je barevně znázorněno, které bandy byly vyhodnoceny jako stejně veliké. Graf opět můžeme mít v lineárním nebo logaritmickém měřítku.

Výsledná fylogenetická klasifikace je vyhodnocena po stisknutí tlačítka „Klasifikace dat“ v pravém horním rohu. Po jeho stisknutí se zobrazí fylogenetický strom vzorků (ukázka na Obrázek 5).

Data je možné uložit v souboru .mat. Po jeho opětovném spuštění se zobrazí uživatelské rozhraní ve stavu, v jakém byla data uložena (tlačítko „Uložit výsledky“).



Obrázek 5: Výslená klasifikace vzorků