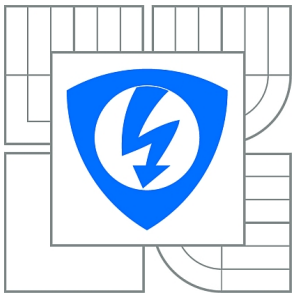
VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ

ÚSTAV BIOMEDICÍNSKÉHO INŽENÝRSTVÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

DEPARTMENT OF BIOMEDICAL ENGINEERING

KLASIFIKACE VZORKŮ 1D GELOVÉ ELEKTROFORÉZY

CLASSIFICATION OF 1D GEL ELECTROPHORESIS SAMPLES

DIPLOMOVÁ PRÁCE

MASTER'S THESIS

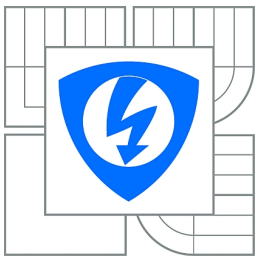
AUTOR PRÁCE — Bc. ONDŘEJ KRUPKA

AUTHOR

VEDOUCÍ PRÁCE — Ing. MARTIN VÍTEK, Ph.D.

SUPERVISOR

BRNO 2015



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

Fakulta elektrotechniky a komunikačních technologií

Ústav biomedicínského inženýrství

Diplomová práce

magisterský navazující studijní obor
Biomedicínské inženýrství a bioinformatika

Student:	Bc. Ondřej Krupka	***ID:***	137253
Ročník:	2	***Akademický rok:***	2014/2015

NÁZEV TÉMATU:

Klasifikace vzorků 1D gelové elektroforézy

POKYNY PRO VYPRACOVÁNÍ:

1) Nastudujte a popište problematiku zpracování obrazu 1D gelové elektroforézy. Zaměřte se na různé přístupy předzpracování obrazu, detekce hranic jednotlivých vzorků a zejména jejich následnou klasifikaci. 2) Vytvořte databázi vlastních referenčních snímků s ohledem na dosažení co nejvyšší výsledné obrazové kvality. 3) Navrhněte a v Matlabu realizujte metodu detekce hranic jednotlivých vzorků a otestujte ji na vytvořené databázi. Dosažené výsledky diskutujte. 4) Navrhněte a v Matlabu realizujte různé metody klasifikace jednotlivých vzorků založené na jejich zarovnání a následné shlukové analýze. Metody otestujte na vytvořené databázi a statisticky vyhodnoťte. 5) Program opatřete vhodným grafickým uživatelským rozhraním.

DOPORUČENÁ LITERATURA:

[1] SKUTKOVA, H., M. VITEK, S. KRIZKOVA, R. KIZEK, et al. Preprocessing and classification of electrophoresis gel images using dynamic time warping. Int. J. Electrochem. Sci., 2013, 8(2).
[2] AUSUBEL, F.M., R. BRENT, R.E. KINGSTON, D.D. MOORE, et al. Current Protocols in Molecular Biology. Edtion ed.: John Wiley & Sons, Incorporated, 2003. ISBN 9780471142720.

Termín zadání:	9.2.2015	***Termín odevzdání:***	22.5.2015

Vedoucí práce: Ing. Martin Vítek, Ph.D.
Konzultanti diplomové práce:

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

UPOZORNĚNÍ:

Autor diplomové práce nesmí při vytváření diplomové práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č.40/2009 Sb.

Abstrakt

Tato diplomová práce se zabývá klasifikací vzorků 1D gelové elektroforézy. Popisuje teoretické informace týkající se gelové elektroforézy, různým druhům rušení, zpracování obrazu a jeho klasifikace pomocí shlukové analýzy. Jeden z hlavních úkolů je vytvoření databáze obrazů s ohledem na jejich co nejvyšší obrazovou kvalitu. Je realizována metoda předzpracování a detekce hranic jednotlivých vzorků v prostředí MATLAB. A konečně, je realizována klasifikace na základě shlukové analýzy vzorků, které je posléze statisticky ohodnocena.

Klíčová slova

Gelová elektroforéza, zpracování elektroforeogramu, shluková analýza, MATLAB, statistická analýza

Abstract

This term project deals with the classification of 1D gel electrophoresis samples. It describes the theoretical information about gel electrophoresis, various types of errors, processing of the image and its classification using the cluster analysis. One of the main goals is creation of images with the highest quality as possible. A realization of pre-processing and detection of the sample borders is made in the MATLAB environment. And finally, classification of samples is done with subsequent statistical analysis.

Keywords

Gel electrophoresis, processing of the elektroforeogram, cluster analysis, MATLAB, statistical analysis

Bibliografická citace díla:

KRUPKA, O. KLASIFIKACE VZORKŮ 1D GELOVÉ ELEKTROFORÉZY. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií. Ústav biomedicínského inženýrství, 2015. 98 s. Diplomová práce. Vedoucí práce: Ing. Martin Vítek, Ph.D.

Prohlášení:

Prohlašuji, že svoji diplomovou práci na téma KLASIFIKACE VZORKŮ 1D GELOVÉ ELEKTROFORÉZY jsem vypracoval samostatně pod vedením vedoucího diplomové práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autor uvedeného diplomové práce dále prohlašuji, že v souvislosti s vytvořením tohoto projektu jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a jsem si plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

V Brně dne 22. 5. 2015

…….

Poděkování:

Děkuji vedoucímu semestrální práce Ing. Martinu Vítkovi, Ph.D. za účinnou metodickou, pedagogickou a odbornou pomoc a další cenné rady při zpracování mé semestrální práce. Také děkuji Doc. Baštincovi za velice vstřícnou odbornou pomoc se zpracováním statistické analýzy výsledků.

Obsah

Seznam obrázků

1 Úvod

Gelová elektroforéza je významná separační metoda, využívající rozdílu pohyblivosti částic, na které je aplikováno elektrické pole. Je to nejrozšířenější způsob, jakým lze separovat biopolymery, například nukleové kyseliny a proteiny.

V této práci je použita elektroforéza na nosiči, kterým je agarový gel. Metoda je jednoduchá a vysoce efektivní pro separaci, identifikaci a purifikaci různě dlouhých DNA fragmentů. [10]

Díky širokým možnostem aplikace a obrovskou komerční využitelností byla gelová elektroforéza v posledních letech aplikována i na mikročipové platformy. [9]

Předmětem diplomové práce je popsání problematiky zpracování obrazu gelové elektroforézy a jeho klasifikace, vytvoření databáze ideálních obrazů, detekce hranic jednotlivých vzorků, návrh a realizace různých metod klasifikace a následná statistická analýza výsledků.

2 Elektroforéza

2.1 Princip

Elektroforéza představuje pohyb elektricky nabitých částic ve stejnosměrném elektrickém poli mezi dvěma elektrodami. Prostředí mezi elektrodami je vytvořeno použitím elektrolytu, ten zajišťuje vodivé spojení celého systému, do kterého je umístěn vzorek. Při připojení stejnosměrného elektrického pole se kationty pohybují směrem k zápornému pólu a anionty ke kladnému pólu. Díky odlišné rychlosti migrace složek vzorku se v průběhu elektroforézy jednotlivé složky oddělují a to na základě *elektroforetické pohyblivosti*.

$$\mu_e = \frac{q}{6\pi\eta r} \tag{1.1}$$

Kde q je náboj částice, r poloměr částice a η viskozita elektrolytu. Ze vztahu plyne, že elektroforetická pohyblivost je přímo úměrná náboji a nepřímo úměrná její velikosti a viskozitě prostředí. Tudíž se nejlépe pohybuje malá částice s velkým nábojem.

Dalším faktorem, ovlivňujícím pohyblivost je stupeň disociace slabých kyselin a zásad – α.

$$\mu_{e(efektivní)} = \alpha * \mu_e \tag{1.2}$$

Stupeň disociace lze měnit volbou pH prostředí a tím ovlivnit separaci těchto látek. [2]

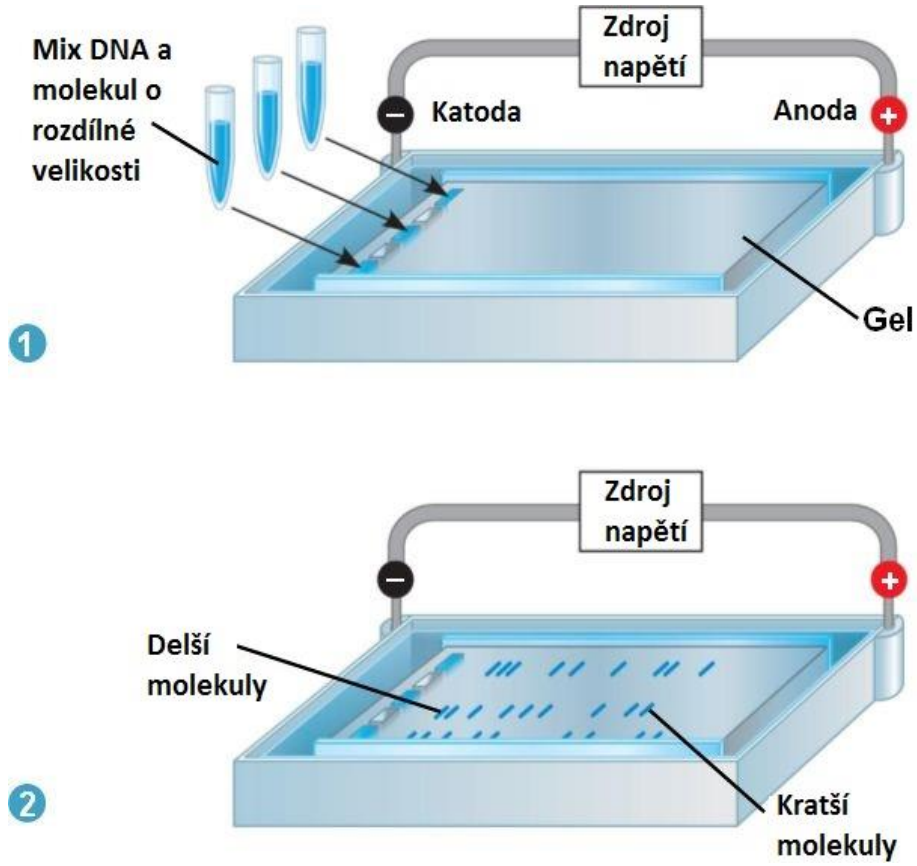
2.2 1D gelová elektroforéza

Pro vykonání elektroforézy je nutné zkoumaný vzorek umístit do vhodného nosiče. Tímto nosičem u 1D gelové elektroforézy je gel, například agarosový, škrobový či polyakrylamidový. Takový gel musí být hydrofilní, nerozpustný ve vodě a mít co nejméně adsorpčních vlastností.

U této metody je rychlost pohybu částic ovlivněna také principem molekulárního síta. Částice vzorku jsou omezovány přítomným gelem, například velikost pórů gelu ovlivňuje rychlost pohybu přímo úměrně. Velké molekuly tedy budou více zbrzďovány póry gelu, než molekuly malé. [3]

Běžně používaným nosičem je polyakrylamidový gel, který je inertní, průhledný, mechanicky pevný a poskytuje možnost různého nastavení vlastností, například hustota zesíťování, či gradient hustoty.

Metodu lze realizovat jako sloupcovou, nebo plošnou. Sloupcová metoda, kde migrace je uskutečněna v tenkých trubičkách, je jednoduchá a levná. Plošná metoda je citlivější, avšak nevýhodou je komplikovanější příprava gelu. Princip plošné elektroforézy je vyobrazen na Obrázek 1. [4]



Obrázek 1: Gelová elektroforéza [6]

Druhým typem gelu je agarosový, metoda je identická pouze s rozdílnými vlastnostmi nosiče. Agarosový gel je méně náročný na přípravu, avšak je také křehčí než polyakrylamidový gel a při větší koncentraci je zakalený. Je také nutné gel zbarvit UV činidlem, např. ethidium bromidem, či fluorescenční nukleovou kyselinou (menší toxicita). Následně je gel prosvícen UV světlem a vyfocen digitální kamerou. [5]

2.3 Druhy rušení v elektroforéze

V elektroforeogramu můžeme očekávat různé druhy rušení, například vnějšími vlivy, nedokonalou přípravou gelu, chybou při postupu vypracování metody, či zvolení špatných parametrů metody. [1]

Druhů rušení je velké množství, v této práci je popsáno několik z nich.

2.3.1 Špatné rozlišení fragmentů DNA

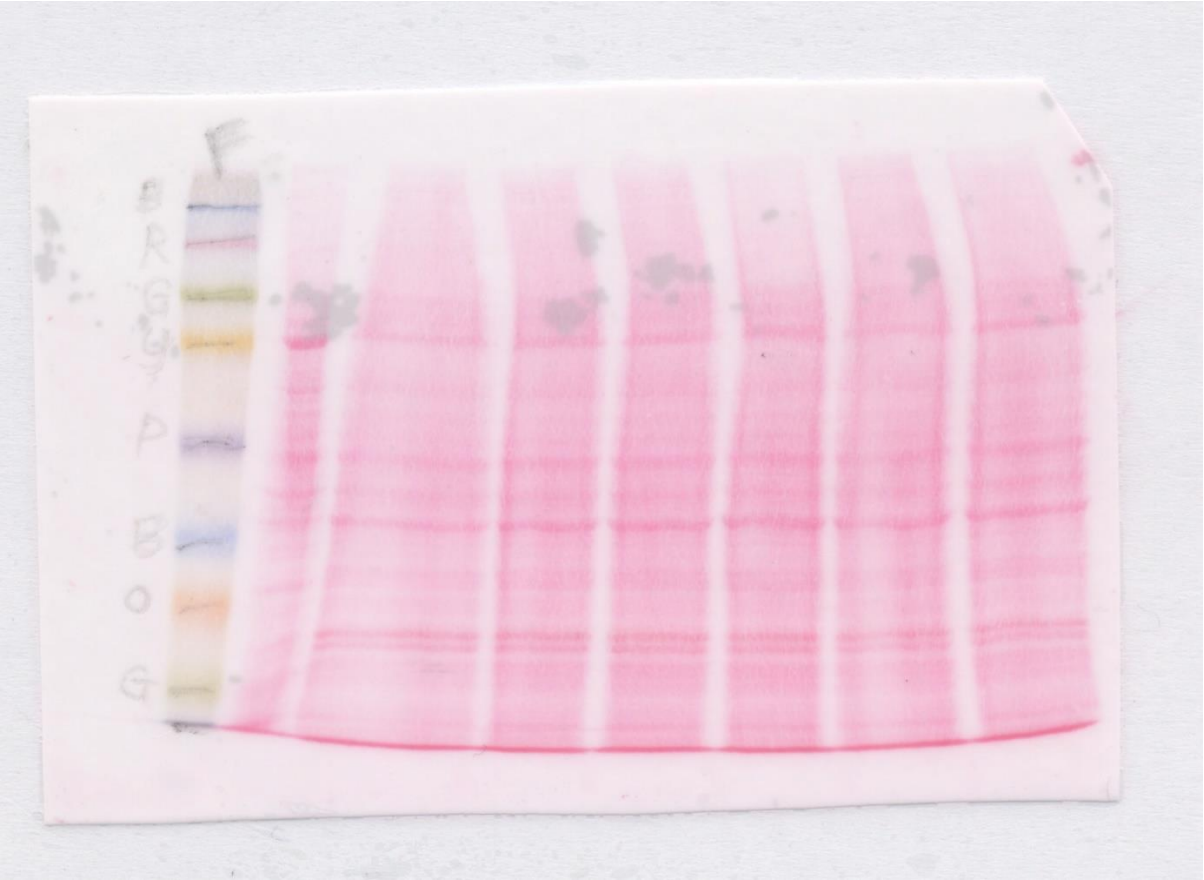
Nejčastější příčinou špatného rozlišení DNA je špatně zvolená hustota agarového gelu. Pro DNA fragmenty s vysokou molekulární hmotností je nutné použít gel méně koncentrovaný a naopak.

Také příliš nízké napětí a tedy i dlouhá doba elektroforézy způsobuje málo výrazné až nejasné proužky a to především u malých fragmentů DNA. [10]

2.3.2 Smile effect

Obvodem elektroforézy procházející stejnosměrný proud způsobuje vznik tepla, který má za následek ohřívání gelu které může způsobit znehodnocení zkoumaných bílkovin, nebo morfologické změny v obraze.

Tyto změny jsou způsobeny nerovnoměrným zahřátím gelu, kdy teplejší částí prochází vzorky snadněji než částí s menší teplotou. Příklad smile effect rušení je na Obrázek 2.



Obrázek 2: Smile effect

Jelikož je míra zahřátí přímo úměrná procházejícímu proudu, je nutné zvolit správné parametry výstupního napětí el. zdroje,. Při nutnosti užití vysokého napětí je žádoucí použít v systému například peristaltickou pumpu, která zaručí cirkulaci pufru v systému a tedy i rovnoměrnou distribuci tepla v gelu. [10]

2.3.3 Další typy rušení

Mezi rušení elektroforézy patří například *rozmazání proužků*, vyskytující se u DNA fragmentů s větší hmotností. Bývá způsobeno přílišným naplněním důlků v gelu, kdy je použito většího objemu vzorků. Také použití příliš vysokého napětí a mechanické poškození důlků v gelu způsobuje rozmazání.

Mimo jiné, příprava menšího množství pufru, nebo vypaření pufru z nádrže elektroforézy způsobuje *roztavení gelu* a tedy i snížení kvality elektroforeogramu. Dle literatury je vhodné pro dlouhé nastavení času použít TBE pufr místo TAE kvůli větší pufrovací kapacitě. Další metodou prevence je použít nádrž s rezervoárem pro doplnění pufru při vypaření. [10]

Jiné typy rušení jsou - nehomogenita osvětlení, zkreslení hran, parazitní gradient atd.

2.4 Předzpracování snímků elektroforézy

Jelikož výsledné obrázky mívají v praxi k dokonalosti relativně daleko, je nutné obraz předzpracovat. Prvním krokem je zlepšení celkové kvality obrazu a to kvůli jednodušší automatizaci zpracování vzorku jako takového a dále kvůli zajištění vysokého kontrastu výsledného obrazu.

Obraz je konvertován na 1D signál pomocí mediánu, tento postup může být považován za nelineární filtraci. Dochází ke zvýraznění užitečných informací vzorku, například jde o zvýšení kontrastu mezi proužky a pozadím obrazu.

Pro zpracování je vyžadováno použít normalizovaný (hodnoty pixelů v intervalu od 0 do 1), šedotónový obraz, kde pozadí je bílé (odpovídá 1) a proužky jsou černé (odpovídají 0).

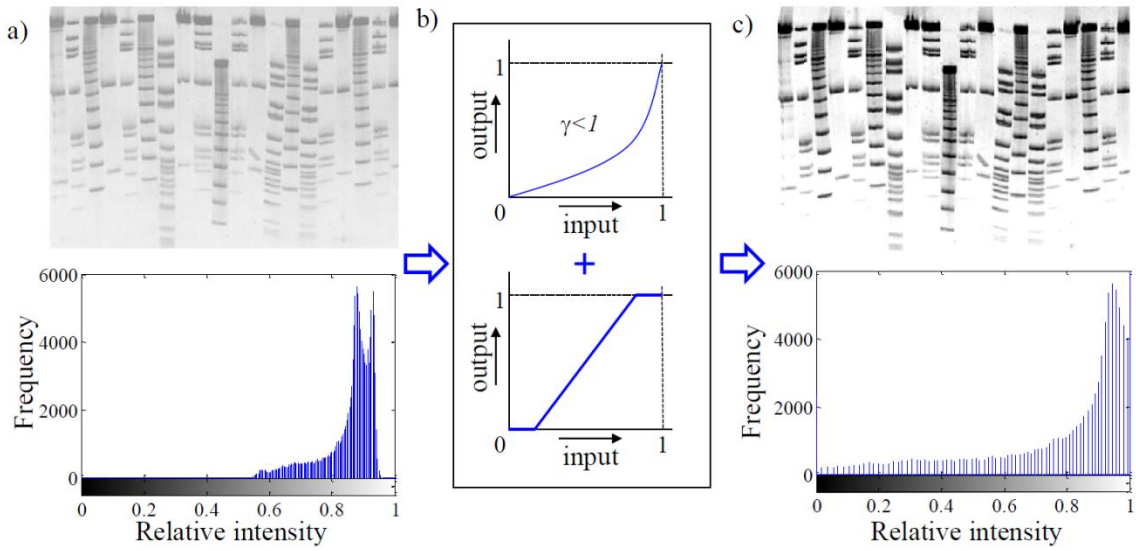
Nejprve je vykreslen histogram znázorňující frekvence jednotlivých odstínů šedi v obrazu. Obvykle je v histogramu patrný výrazný nárůst hodnot od 0,55 do 1, který reprezentuje bílé pozadí. Vzhledem k tomu, že obrazy mohou být přeexponované (většina hodnot se blíží 0), nebo podexponované (většina hodnot se blíží 1), je nutné použít po částech lineární transformační funkci pro zlepšení kvality obrazu. Dojde k roztažení histogramu do větší šířky, znamenající citelné zlepšení kvality obrazu. Pokud jsou všechny odstíny obrazu rozloženy rovnoměrně, jedná se o takzvaný *brilantní obraz*. [1]

Druhou použitou metodou ke zvýraznění obrazu je *gama korekce*, jedná se o nelineární transformaci kontrastu a je popsána rovnicí:

$$g = f^{\gamma^{-1}} \tag{1.3}$$

Výsledek korekce je závislý na hodnotě parametru γ, který se vyskytuje v intervalu hodnot <0,6;2,5>. Hodnoty γ<1 se používají u přeexponovaných snímků, γ>1 je použito u podexponovaných snímků. [1], [7], [8]

Příklad použití po částech lineární transformační funkce a gama korekce je vyobrazen na Obrázek 3.



Obrázek 3: Zlepšení kontrastu snímku za pomocí po částech lineární transformační funkce a gama korekce. a) Původní obraz gelu a jeho histogram. b) Gama korekce (γ=0.65) a po částech lineární transformace obrazu. c) Výsledný zpracovaný obraz a jeho histogram. [1]

2.5 Detekce hranic vzorků

Druhým krokem ve zpracování elektroforeogramů je detekce linií a proužků a tudíž detekce hranic proužků.

2.5.1 Detekce linií

Automatická detekce linií je důležitým krokem k plné automatizaci zpracování elektroforeogramů. Přístupů k tomuto problému je více, některé metody využívají tzv. *sledování linií*, kde je počítáno s prostředkem každé linie. Nevýhodou je ignorování informací vně středu linie. Další metoda sleduje hranice jednotlivých linií a dovoluje kompletní segmentaci linií.

Tato metoda může být rozdělena na dvě části:

- Detekce prvního pixelu každé hranice linií
- Sledování hranice skrz další pixely

Detekce prvního pixelu je provedena pomocí výpočtu průměru (či standardní odchylky) každého sloupce pixelů obrazu. Pixely jsou vybrány pouze z první třetiny obrazu (od shora) a to kvůli absenci zakřivení linií (zakřivení se výrazněji projevuje až dále. Takto je vypočten 1D průběh průměrů sloupců pixelů a ten je vykreslen. Z takto získaného průměru jsou detekována lokální maxima, s určitou minimální vzdáleností, která představují první pixel hledané linie.

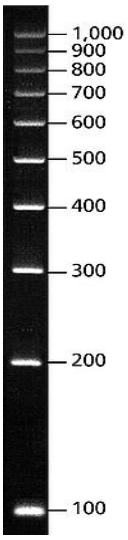
Dále jsou vypočteny další pixely každé detekované linie. V každém kroku jsou porovnány hodnoty intenzit tří pixelů pod aktuálním pixelem. Ten, který má největší hodnotu, je vybrán jako další pixel linie. Pro zajištění nejvíce přímé cesty je nutné vynásobit prostřední pixel určitou konstantou (například 1,04). [1]

2.5.2 Detekce proužků

Detekce proužků může být realizována obdobně jako detekce linií. Obraz je rozdělen na jednotlivé linie, ty jsou otočeny o 90° po směru hodinových ručiček a v nich je vypočten průměr pixelů ve sloupcích. Tento průměr je vynesen do grafu jako 1D signál. Jsou detekována lokální maxima, která značí přítomnost proužku.

2.6 Určení velikosti molekul

Pro kvantitativní měření v elektroforeogramu je nutné použít alespoň jeden standardizovaný vzorek – tzv. *ladder*. Ladder má jasně danou hodnotu jednotlivých proužků ve foreogramu. Jde o komerčně dostupnou směs molekul, která vykazuje stálé stejné výsledky. Příklad 100bp ladderu od výrobce Sigma-Aldrich Co. je na Obrázek 4.



Obrázek 4: 100bp ladder Sigma-Aldrich Co. - uvedené hodnoty představují počet párů bazí pro konkrétní proužek [11]

Při výběru ladderu je nutné zvážit velikost zkoumaných molekul, aby rozsah hodnot ladderu přibližně odpovídal rozsahu hodnot zkoumané látky. Používanými rozsahy jsou 100bp, 1000bp, 10kbp atd. Toto jsou laddery s lineární stupnicí, dále se můžeme setkat také se stupnicí logaritmickou, například 2log. [10], [11]

2.7 Klasifikace

Pro klasifikaci elektroforézy pomocí shlukové analýzy je nejprve vhodné vybrat správné příznaky, podle kterých budou jednotlivé vzorky shlukovány. Na základě příznaků je možno vypočíst tzv. *vzdálenost* jednotlivých vzorků a za pomoci matice vzdáleností tyto vzorky seskupit do jednotlivých skupin (shlukovat). [12], [15]

Metody výpočtu podobností a shlukování byly vybrány z prostředí MATLAB, jmenovitě z funkcí *pdist* a *linkage* obsažených v *Statistics and Machine Learning Toolbox* verze 2015a.

2.7.1 Zarovnání vzorků

Po dohodě s vedoucím práce byla část o zarovnání vzorků vyřazena z této práce. Zarovnání vzorků je žádoucí pouze u některých obrazů a v případě pouze jednoho proužku ve vzorku může zarovnání obraz výrazně zkreslit.

2.7.2 Příznaky vzorků

Volba správných příznaků jednotlivých vzorků je stěžejní pro následnou klasifikaci snímků elektroforézy. Hodnoty těchto příznaků reprezentují určitou vlastnost konkrétního vzorku,

která je pro něj charakteristická. Například pro vzorky 1D gelové elektroforézy je jednou z charakterizujících hodnot počet proužků ve vzorku, či šířka jednotlivých proužků. Těchto příznaků lze vybrat celá řada a výběr závisí na konkrétním využití shlukové analýzy. [15]

Výběr příznaků bude podrobněji popsán v kapitole 5.1.

2.7.3 Standardizace dat

Tímto krokem je matice příznaků konvertována z původních do nových, bezrozměrných hodnot. Tento krok se používá především díky rozdílným jednotkám jednotlivých příznaků. Tyto rozdíly mohou ovlivnit podobnost mezi jednotlivými vzorky. Standardizace dat je však volitelná a v této práci použita není. [15]

2.7.4 Výpočet matice podobností

Před shlukovou analýzou je nutné vypočíst tzv. matici podobností. Použitím hodnot daného páru objektů z matice příznaků je možné vypočítat, jak podobné jsou si tyto objekty. Hodnota v matici může vyjadřovat podobnost, či vzájemnou vzdálenost těchto objektů. Rozdíl těchto hodnot je otázka *směru.* Menší hodnota vzdálenosti vyjadřuje vyšší hodnotu podobnosti a naopak. Matici podobností lze vypočíst několika metodami. [15]

Mezi metody výpočtu patří:

1. *Euklidovská vzdálenost:*

$$d^2{}_{st} = (x_s - x_t)(x_s - x_t)' \tag{2.1}$$

2. *Vzdálenost City block:*

$$d_{st} = \sum_{j=1}^{n} |x_{sj} - x_{tj}| \tag{2.2}$$

3. *Minkovského metrika:*

$$d_{st} = \sqrt[p]{\sum_{j=1}^{n} |x_{sj} - x_{tj}|^p} \tag{2.3}$$

4. *Chebishevova vzdálenost:*

$$d_{st} = max_j\{|x_{sj} - x_{tj}|\} \tag{2.4}$$

5. *Cosinová vzdálenost:*

$$d_{st} = 1 - \frac{x_s x_t{}'}{\sqrt{(x_s x_s{}')(x_t x_t{}')}} \tag{2.5}$$

6. *Korelační vzdálenost:*

$$d_{st} = 1 - \frac{(x_s - \overline{x_s})(x_t - \overline{x_t})'}{\sqrt{(x_s - \overline{x_s})(x_s - \overline{x_s})'}\sqrt{(x_t - \overline{x_t})(x_t - \overline{x_t})'}} \tag{2.6}$$

kde

$$\overline{x_s} = \frac{1}{n}\sum_j x_{sj} \tag{2.7}$$

a

$$\overline{x_t} = \frac{1}{n}\sum_j x_{tj} \tag{2.8}$$

7. *Hammingova vzdálenost:*

$$d_{st} = \left(\frac{\#\left(x_{sj} \neq x_{tj}\right)}{n}\right) \tag{2.9}$$

8. *Jaccardova vzdálenost:*

$$d_{st} = \frac{\#\left[\left(x_{sj} \neq x_{tj}\right) \cap \left(\left(x_{sj} \neq 0\right) \cup \left(x_{tj} \neq 0\right)\right)\right]}{\#\left[\left(x_{sj} \neq 0\right) \cup \left(x_{tj} \neq 0\right)\right]} \tag{2.10}$$

9. *Spearmanova vzdálenost:*

$$d_{st} = 1 - \frac{(r_s - \overline{r_s})(r_t - \overline{r_t})'}{\sqrt{(r_s - \overline{r_s})(r_s - \overline{r_s})'}\sqrt{(r_t - \overline{r_t})(r_t - \overline{r_t})'}} \tag{2.11}$$

kde r_{sj} *je řád* x_{sj} *vypočtený z* x_{1j},.. x_m *a* r_s, r_t *jsou řády vektorů* x_s, x_t *s ohledem na jejich polohu,*

$$\overline{r_s} = \frac{1}{n}\sum_j r_{sj} = \frac{(n+1)}{2} \tag{2.12}$$

a

$$\overline{r_t} = \frac{1}{n}\sum_j r_{tj} = \frac{(n+1)}{2} \tag{2.13}$$

kde x_s a x_t jsou vektory zkoumaných objektů. [19]

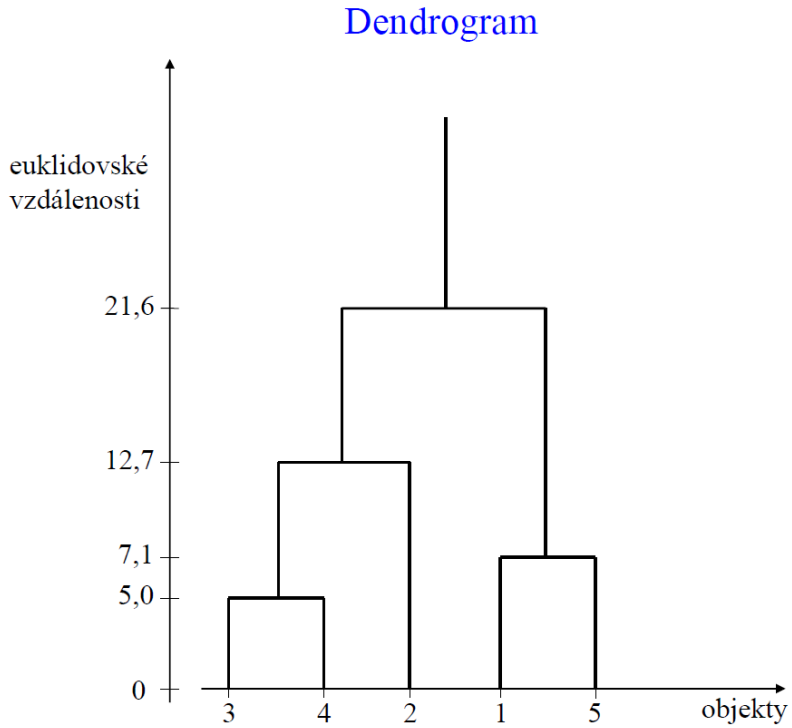
2.7.5 Shluková analýza

Shluková analýza je seskupování podobných objektů, jinými slovy roztřiďuje množinu objektů podle charakteristik (příznaků), jež nejsou nezávislé na individuálním pohledu každého člověka. Snažíme se nalézt shluky objektů, aby si členové shluku byli podobní navzájem, ale nebyli podobní s objekty, které jsou mimo shluk. [13]

V případě klasifikace elektroforézy jde o shlukování tvrdé, hierarchické, aglomerativní – tyto shlukové metody nabízí více alternativních řešení, jsou vyjádřeny dendrogramy a při analýze vycházíme z jednotlivých objektů, které následně spojujeme.

Mezi metody shlukové analýzy paří například metoda UPGMA (z angličtiny *unweighted pair-group method using arithmetic averages*), která využívá neváhované párování pomocí aritmetických průměrů. Pro výpočet metody je nutné nejdříve vytvořit dříve zmíněnou *matici podobnosti*. [14]

U metody UPGMA jde o nalezení nejmenší hodnoty euklidovské vzdálenosti a tyto dva objekty sloučit a za pomoci výpočtu aritmetického průměru dopočítat hodnoty vzdáleností mezi sloučenými a ostatními objekty. Dále vybereme nejmenší vzdálenost a postup opakujeme, přičemž postupně zakreslujeme sloučené objekty s jejich vzdáleností do dendrogramu (stromová struktura) viz Obrázek 5, kde jako první byly sloučeny objekty 3 a 4 a jejich vzdálenost měla hodnotu 5,0.



Obrázek 5: Dendrogram vytvořený pomocí metody UPGMA [14]

Z této metody vychází také UPGMC (*Unweighted Pair-Group Method using Centroids*), která počítá euklidovské vzdálenosti pro geometrické středy (*centroidy*) vypočtených distancí.

Median linkage, neboli WPGMC (*Weighted Pair-Group Method using Centroids*) využívá pro výpočet euklidovských vzdáleností váhované geometrické středy.

Metoda WPGMA (*Weighted Pair-Group Method with Arithmetic mean*) využívá pro výpočet vzdáleností mezi shluky průměrnou hodnotu euklidovské vzdálenosti.

Obdobnou metodou je SLINK (*single linkage clustering method*), která vytváří shluky se stejným postupem, ovšem při přepočítávání vzdáleností nepoužíváme již aritmetický průměr, nýbrž pro novou vzdálenost určíme tu menší ze vzdáleností obou objektů.

Další metodou je CLINK (*complete linkage clustering method*), která je prakticky stejná jako CLINK, jen při výpočtu hodnoty vzdáleností vybereme hodnotu maximální. [14]

Wardova shlukovací metoda je po UPGMA nejpoužívanější metodou. Jako ostatní metody i Wardova metoda používá kroky shlukování, které začínají s určitým počtem shluků, každý z těchto shluků obsahuje jeden objekt a končí s jedním shlukem obsahujícím všechny objekty. Pracuje s hodnotami *E*, které reprezentují *rozptyl* neboli tzv. *index sumy čtverců*. [15]

V prvním kroku spojujeme objekty tak, aby se hodnota E zvýšila co nejméně. To znamená, že musíme vyzkoušet všechny možné spojení objektů a pro ně vypočítat hodnotu E. Poté vybereme ty dva objekty, pro které se hodnota E zvýšila nejméně.

Ve druhém kroku vypočteme vzdálenosti mezi všemi objekty v daném shluku a průměr tohoto shluku. Pro příklad můžeme vzít, že shluk obsahuje tři objekty a každý z nich je popsán pěti atributy. Pro první objekt bychom vypočítali pět vzdáleností objektu od průměru shluku. Tyto vzdálenosti mohou být pozitivní, či negativní. Pro další objekty vypočteme to samé, tudíž bychom měli patnáct vzdáleností.

Ve třetím kroku umocníme každou vzdálenost, která byla vypočtena v kroku předchozím, a přičteme ji ke každému shluku. Tímto získáme hodnotu *sumy čtverců* pro každý shluk.

V posledním kroku vypočteme hodnotu E přičítáním hodnoty sum čtverců každého shluku.

Opět sestavíme dendrogram na základě výběru nejmenší hodnoty daného shlukovacího kroku. [15]

Pro další analýzu je také vhodné vypočíst, jaké zkreslení přineslo vytvoření dendrogramu. Tento výpočet je proveden pomocí *Pearsonova korelačního koeficientu*, kdy

dokonalou shodu vyjadřuje koeficient roven jedné. Pokud je koeficient větší nebo rovno 0,8 je zkreslení považováno za přijatelné. [14]

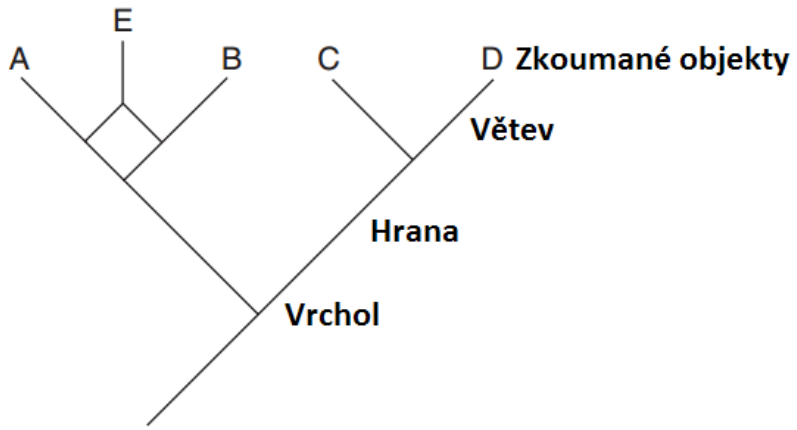
Je také nutné dodat, že některé metody shlukování vyžadují jako vstupní hodnoty vypočtené euklidovské vzdálenosti, jmenovitě jde o UPGMC, WPGMC a Wardovu metodu. [15]

V této práci však budou vyzkoušeny všechny kombinace metod podobností a metod shlukování. Teoreticky by tyto metody měly vykazovat špatné, či zkreslené výsledky shlukování.

2.7.6 Fylogenetická analýza

Fylogenetická analýza je další částí klasifikace elektroforézy. Popisuje posloupnost událostí formující množinu druhů, jinými slovy vývoj druhů organismů v historickém sledu ve smyslu evoluční teorie. [16]

Fylogenetika se zabývá tvorbou evolučních stromů ze sekvencí, které mají podobné vlastnosti, a tedy mají i společného předka. Chceme znát, v jakém pořadí jednotlivé sekvence divergovaly. Tuto divergenci popisuje *fylogenetický strom*, viz Obrázek 6. Větve a hrany reprezentují evoluční vzdálenost jednotlivých zkoumaných objektů, vrchol označuje společného předka.



Obrázek 6:Fylogenetický strom [16]

Pro sestrojení stromu můžeme použít již zmíněnou shlukovou analýzu, kdy dendrogram již tuto evoluční vzdálenost vyjadřuje. Jiné metody sestavení fylogenetického stromu je například metoda spojování sousedů (*neighbor joining*), dále metoda maximální parsimonie (úspornost), nebo maximální pravděpodobnosti, kdy se jedná o znakové metody. [17]

Jiné metody výpočtu evolučních vzdáleností se zakládají na tzv. evolučních modelech, které umožňují zjistit podobnost dvou sekvencí, počet mutací vzniklých vývojem apod. Mezi tyto modely patří například Jukes-Cantorův, Kimurův, Tamurův, či Tamura-Neiův model. [16]

Fylogenetická analýza v této práci realizována není a klasifikace je provedena pomocí shlukové analýzy.

3 Vytvoření referenčních snímků

V rámci této práce byla v laboratoři genomiky a proteomiky na Ústavu biomedicínského inženýrství vytvořena databáze elektroforeogramů. Tato databáze byla vypracována s důrazem na co nejvyšší kvalitu výsledných snímků. Vytvoření probíhalo ve spolupráci s Bc. Tomášem Dvořáčkem.

Metoda může být rozdělena na tři části:

- Příprava gelu v určité koncentraci
- Vložení vzorku DNA do jamek v gelu a aplikace napětí na preparát po určitý čas
- Snímání elektroforeogramu digitální kamerou pod UV světlem

Jednotlivé parametry (koncentrace gelu, volba pufru, objem barviva apod.) budou popsány v kapitole 3.4.

3.1 Příprava gelu

Při přípravě gelu je nutné nejprve namíchat správné množství pufru (TBE, či TAE), abychom mohli vyrobit agarový gel a naplnit celou elektroforézní nádrž. Nejprve byla nachystána vanička určená k tuhnutí gelu a následné elektroforéze. Vanička byla oblepena izolepou, aby mohlo být dosaženo zamýšleného tvaru agarového gelu. Do vaničky byl následně vložen hřeben pro vytvoření důlků pro pipetaci vzorků do gelu.

V kádince byl smíchán pufr s adekvátní množstvím agarosy (dle požadované koncentrace gelu). Tato směs byla rozehřáta na bod varu v mikrovlnné troubě a míchána, dokud nebyla směs naprosto čirá. Následně bylo do této směsi přidáno barvivo GelRed®, které umožňuje zobrazení elektroforeogramu pod UV světlem. Při stálém míchání byla směs ochlazena pod tekoucí vodou na 60°C a byla nalita do předem připravené vaničky. Vrstva nalité směsi bývá obvykle 0,5 – 1cm vysoká.

Tato směs byla ponechána v klidu po určitý čas, aby došlo ke ztuhnutí gelu. Hřeben byl z gelu opatrně vyzvednut a izolepa odlepena. Poté byl gel i s vaničkou vložen do nádrže elektroforézy, která byla zalita pufrem tak, aby došlo k vodivému spojení katody i anody elektroforézy.

3.2 Příprava vzorků

Následně je nutné připravit vzorky DNA, pro elektroforézu. Při vytvoření databáze byly použity pouze referenční vzorky – tzv. *laddery*. Tyto vzorky mají předem známou velikost jednotlivých proužků a používají se jako vztyčné body při klasifikaci, viz kapitola 2.7.

Šlo například o PCR laddery Sigma-Aldrich Co. 100bp, či BioLabs® 2log, které byly namíchané od výrobce – vzorek již obsahoval ladder i barvivo v dané koncentraci. Ladder MO BIO® 100bp bylo nutné namíchat v laboratoři, za pomoci barviva, destilované vody a samotného ladderu. [11]

Dále byly vzorky napipetovány do důlku v gelu, nádrž elektroforézy byla uzavřena a elektrody byly připojeny na zdroj napětí.

3.3 Elektroforéza a zobrazení výsledků

Následuje již samotná metoda, kdy vlivem procházejícího proudu dochází k migraci molekul vzorků směrem od katody k anodě. Je nutné do systému přivést stejnosměrné napětí tak dlouho, aby molekuly dorazily na konec gelu, ale ne příliš dlouho, aby z gelu nevytekly.

Po ukončení procesu je gel vyzvednut z lázně a umístěn pod zdroj UV záření. To umožňuje zobrazení vzorků a následné snímání digitální kamerou.

3.4 Parametry elektroforézy

Při vytváření databáze snímků s ohledem na co největší kvalitu bylo nutné brát v úvahu velké množství parametrů, které mohou výsledný obraz ovlivnit. Je nutné na začátku zmínit, že výběr pufru byl pro všechny obrazy stejný – TBE.

Prvním parametrem byla hustota gelu. Z literatury lze vyčíst, že méně husté gely jsou určené pro molekuly s větší hmotností, zato více husté gely dosahují lepších výsledků pro molekuly s menší hmotností. Viz ***Tabulka 1***.

Tabulka 1: Koncentrace agarosy pro dělení fragmentů DNA [10]

Agarosa (%)	Efektivní rozsah hodnot fragmentů DNA
0,5	30 až 1
0,7	12 až 0,8
1,0	10 až 0,5
1,2	7 až 0,4
1,5	3 až 0,2

Při vypracování bylo ve většině případů použito gelu o koncentraci 1% a to kvůli použití ladderů v rozsahu hodnot 10 až 0,5.

Dále je nutné zvážit, jaké barvivo je vhodné použít. Obvykle je použit ethidium bromid, který má mutagenní a karcinogenní účinky. Vhodnou alternativou je například barvivo GelRed®.

Kritickým parametrem, pro správnou přípravu gelu je také teplota po ochlazení, je žádoucí teplotu nesnížit příliš, jinak dochází k brzkému tuhnutí gelu. To může být problém, pokud s gelem je nutné stále manipulovat. Odstraňování bublinek pomocí pipetovací špičky z gelu při nízké teplotě může způsobit nehomogenitu tvaru připravovaného gelu – důlek po prasklé bublince v gelu zůstane. Správná teplota gelu pro nalévání je 55-60°C. [10]

Doba tuhnutí gelu je závislá na parametrech gelu samotného. Zvyšuje se s jeho hustotou a jeho tloušťkou. Obvyklá doba tuhnutí gelu při přípravě v laboratoři byla 15 minut. Některé vzorky byly nechány ztuhnout po dobu 30 minut, avšak toto nebylo nutné. Již po 15 minutách nedocházelo ke změně tuhosti gelu.

Nejdůležitějšími parametry elektroforézy je však výstupní napětí zdroje a délka trvání elektroforézy. Obecně je napětí nastaveno dle délky gelu a to pro každý cm délky je nutné přičíst 1 až 10 voltů. Délku trvání je nutné zvolit dle přivedeného napětí. Vzorky migrují rychleji pod větším napětím, a tedy je nutné ohlídat, aby nedosáhly konce gelu a tzv. „nevytekly“. V laboratoři byly ve většině případů zvoleny tyto hodnoty: 70V/90min. Pokud vzorky nedosáhly poslední třetiny gelu (vlivem větší hustoty gelu apod.) byl čas dodatečně navýšen. [10]

Ze všech měření a o všech použitých hodnotách byly v laboratoři vedeny příslušné protokoly. Ukázka takového protokolu je na Obrázek 7.

Laboratorní protokol

Pracující: Krupka Dvořáček	ID měření: 11a	Provedeno: Elektroforéza
Datum: 27. 10. 2014	Měření úspěšné (ano/ne/částečně): ano	

Typ, ID a datum výroby pufru:	TBE, 1, 20 a 22.10. 2014, 1x konc
Množství použitého pufru [ml]:	50+250

Typ barviva a ředění:	GelRed, 10x zředěný
Množství barviva [µl]	25

Celkový objem gelu [ml]:	50
Teplota gelu při nalévání [°C]:	60
Doba tuhnutí gelu před pipetováním vzorků:	30

Napipetované vzorky: pouze 100bp!!									
1	2	3	4	5	6	7	8	9	
Mobio	Mobio	Mobio	Sigma	Sigma	Sigma	Biolabs	Biolabs	Biolabs	

Nastavená prvotní délka elektroforézy:	90
Dodatečná délka elektroforézy:	
Nastavené napětí:	70

Neúmyslné chyby měření:

Úmyslné chyby měření:

Názvy výstupních souborů:
elfo_id_11_a

Obrázek 7: Ukázka vyplněného protokolu pro měření 27. 10. 2014

3.5 Zdroje chyb

Při vypracování databáze se bylo možné setkat s různými zdroji chyb, které velmi snižovaly kvalitu výsledných obrazů. Tyto chyby byly po několika měřeních odstraněny a výsledky odpovídají prakticky ideálním obrazům, viz kapitola 3.6.

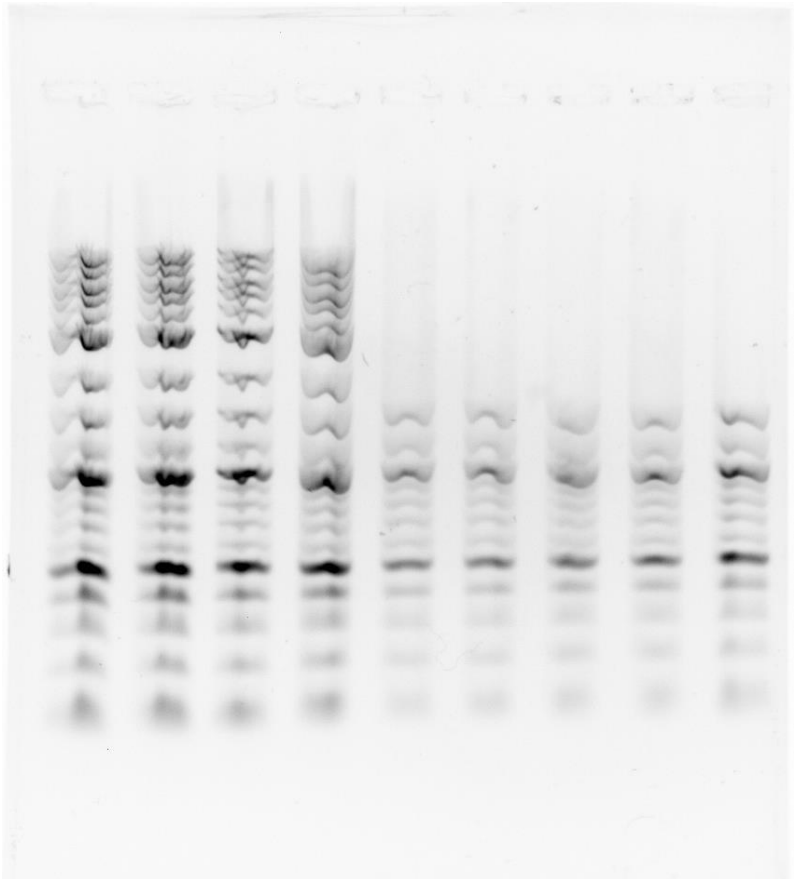
Ukázalo se, že správný postup přípravy gelu je stěžejní pro celou metodu, kdy bublinky, zářezy, důlky apod. poškodily výsledný elektroforeogram. Je také nutné vyvarovat se nahodilých chyb, kdy například při odlepování izolepy z vaničky došlo ke sklouznutí gelu z vaničky a jeho pádu na zem. Takový gel je již nepoužitelný.

Velký důraz je také nutné klást důraz na správné namíchání a napipetování vzorků. Špatně provedená pipetaci – propíchnutí gelu, nerovnoměrná distribuce vzorku v jamce, či použití většího objemu vzorku také vede k chybám viditelným na výsledném obrazu.

3.6 Výsledné obrazy

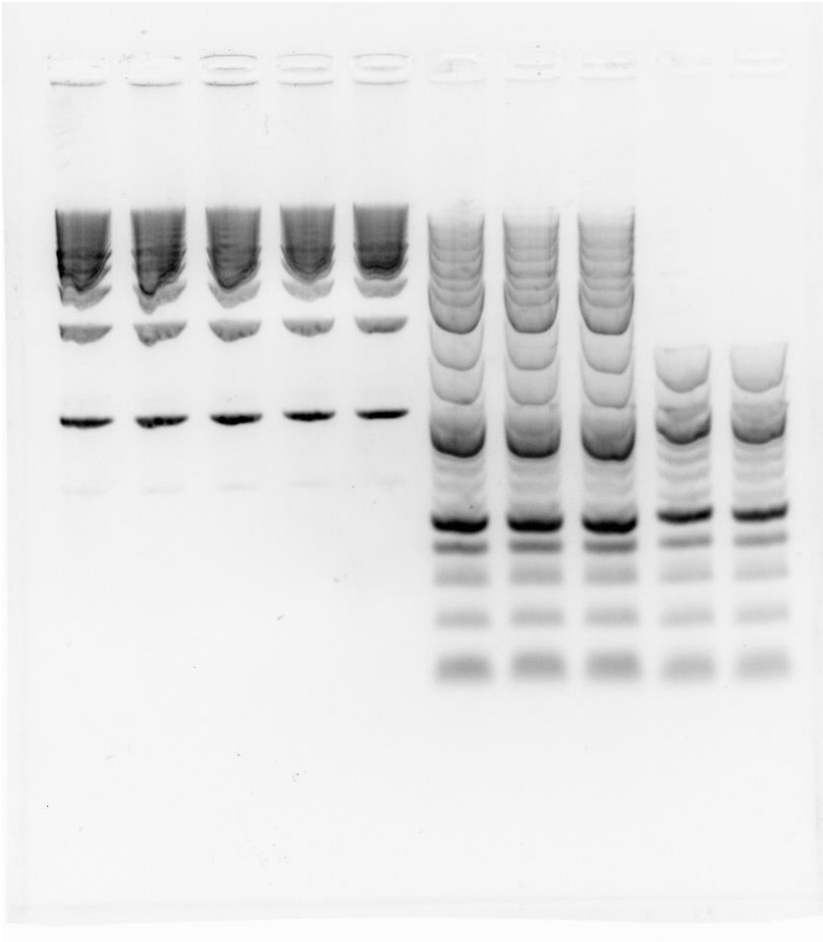
Při vypracování databáze se ukázalo, že první snímky nejsou ani z daleka ideální.

U Obrázek 8 jsou některé proužky velmi zakřiveny, jsou z části výrazné, z části nevýrazné, jiné nejsou výrazné vůbec. Je také vidět sklon linie posledních proužků v obraze. Celkově je obraz nevýrazný a pro použití v ideální databázi nepoužitelný. Zde chyby nastaly především při přípravě a pipetaci vzorků. Při přípravě gelu váženka s agarosou spadla do Erlenmeyerovy baňky a mohla gel kontaminovat. Došlo k propíchnutí gelu hned v několika jamkách a v jiných došlo k vyplavení vzorku z jamky (nevýrazné proužky).



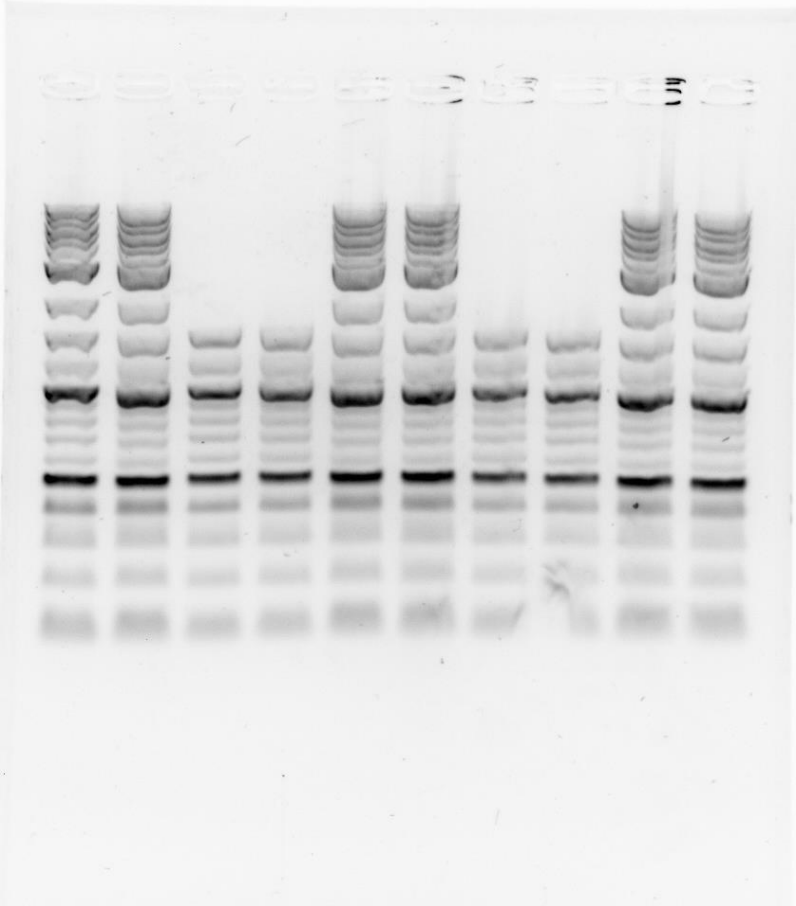
Obrázek 8: První ostré měření - elfo_id_2_a

Další neúspěšný pokus o vytvoření ideálního elektroforeogramu je na Obrázek 9. Zde se vyskytují artefakty u větších fragmentů (1. až 5. linie, délka 1kb), kdy první proužky jsou velmi rozmazané. U druhé poloviny linií je vidět „U“ profil proužku, který reprezentuje pohyb špičky pipety v důlku při pipetování vzorků. Zde lze poukázat na rozdíl mezi prvním snímkem (Obrázek 8) a druhým snímkem (Obrázek 9), je patrné že ze začátku pipetování příliš nešlo a proužky jsou na Obrázek 8 rozvlněné, u Obrázek 9 jsou již téměř rovné a kopírují „U“ profil.



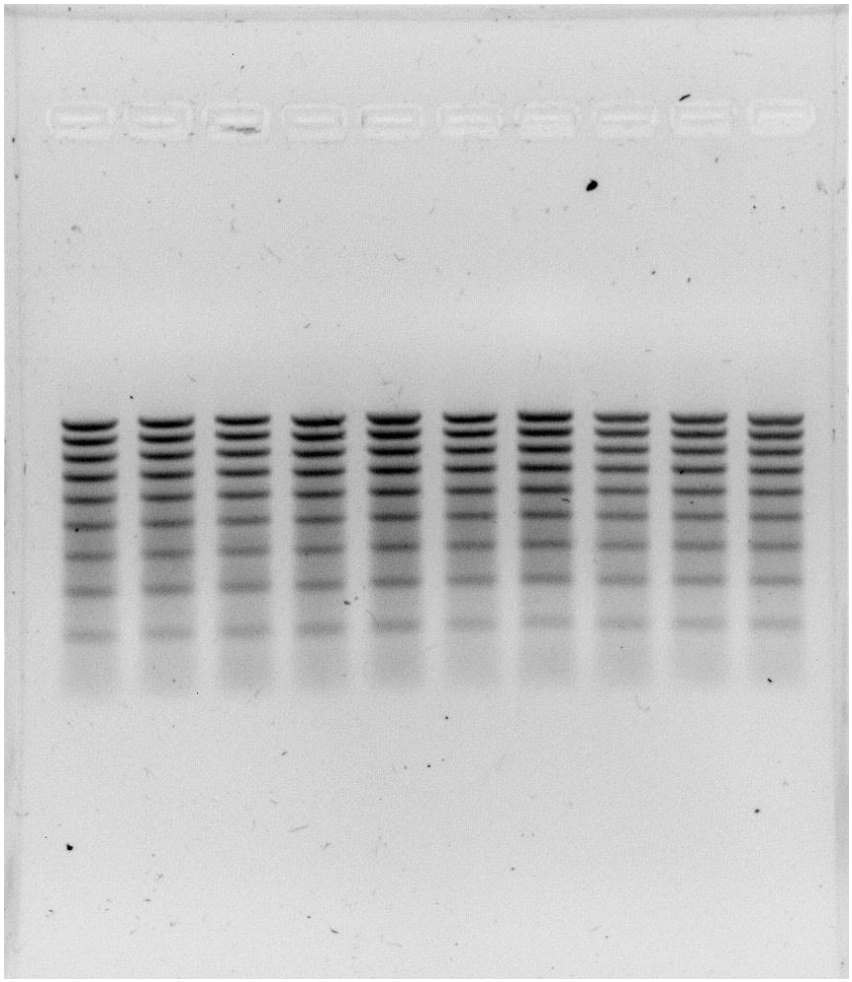
Obrázek 9: Ukázka, čtvrté měření - elfo_id_4_a

V dalším elektroforeogramu (Obrázek 10) nastala chyba při přípravě gelu, který byl mechanicky poškozen a delší molekuly v liniích číslo 7 a 8 nedorazily až na konec a jejich proužky jsou poškozené. Dále došlo k propíchnutí gelu při pipetování – především proužky 6 a 9. Krom těchto problémů je z elektroforeogramu jasný kvalitativní posun. Proužky jsou v první třetině gelu zakřivené („U“), avšak se postupem průchodu gelem narovnávají. Tento elektroforeogram se již blíží ideálnímu obrazu.



Obrázek 10: Ukázka, sedmé měření - elfo_id_7_a

Poslední elektroforeogram (Obrázek 11) lze již označit za takřka ideální, proužky jsou rovné, nedošlo k propíchnutí gelu a konce linií jsou prakticky stejně dlouhé. Jediné dvě chyby jsou - horší kontrast než u minulých snímků a nečistoty v elektroforeogramu. Nečistoty mohly vzniknout při přípravě gelu, nebo skleněná podložka, na kterou se umisťuje gel v kameře, mohla být špinavá. Tyto menší nedostatky lze však odstranit správným předzpracováním obrazu před následnou analýzou, viz kapitola 2.4.



Obrázek 11: Ukázka, jedenácté měření - elfo_id_11_b

Další obrazy i s uvedenými protokoly jsou přiloženy v kapitole 10.

4 Detekce hranic jednotlivých vzorků

V této kapitole bude navrhnut a realizován algoritmus detekce hranic jednotlivých vzorků. Samotný algoritmus může být rozdělen na několik částí:

- Předzpracování
- Detekce linií
- Převedení obrazu na medián
- Detekce proužků

které budou popsány v dalších kapitolách.

4.1 Algoritmus

4.1.1 Předzpracování

Jelikož databáze obrazů není zdaleka ideální, je nutné tyto obrazy předzpracovat. Jak již bylo zmíněno v kapitole 2.4, je vhodné docílit maximální kvality obrazu.

V první části je obraz normalizován, převeden na obraz šedotónový a je zajištěno, aby v obraze měly proužky černou barvu a pozadí barvu bílou. Obraz je také oříznut. V obrazech se vyskytuje především horší kontrast, který znesnadňuje následnou detekci linií a proužků. Ke zlepšení kontrastu je použita *po částech lineární transformace kontrastu* a *gama korekce*.

4.1.2 Detekce linií

Při detekci linií je nejprve vypočítána standardní odchylka obrazu, která vytvoří 1D reprezentaci obrazu ve vertikálním směru. Tento průběh je vyhlazen mediánovým filtrem a jsou v něm detekována lokální maxima. Tato maxima reprezentují hranice linií obrazu.

4.1.3 Převedení obrazu na medián

Polohy detekovaných linií jsou využity k převedení obrazu na medián. V každé linii je vypočítána hodnota mediánu jednotlivých řádků. Tato hodnota je následně roztažena po celém řádku. Takto dochází k vyhlazení obrazu a k vyrovnání proužků. Nevýhodou této metody je mírné roztažení proužků, které jsou rozmazané, či deformované ve vertikálním směru.

4.1.4 Detekce proužků

K detekci proužků přistupujeme analogicky jako k detekci linií. Obraz je rozdělen na jednotlivé linie, ve kterých je provedena detekce. Jako vstup slouží průběh dříve vypočteného mediánu jednotlivých proužků, ten je převeden na 1D signál a z něj jsou detekována lokální maxima.

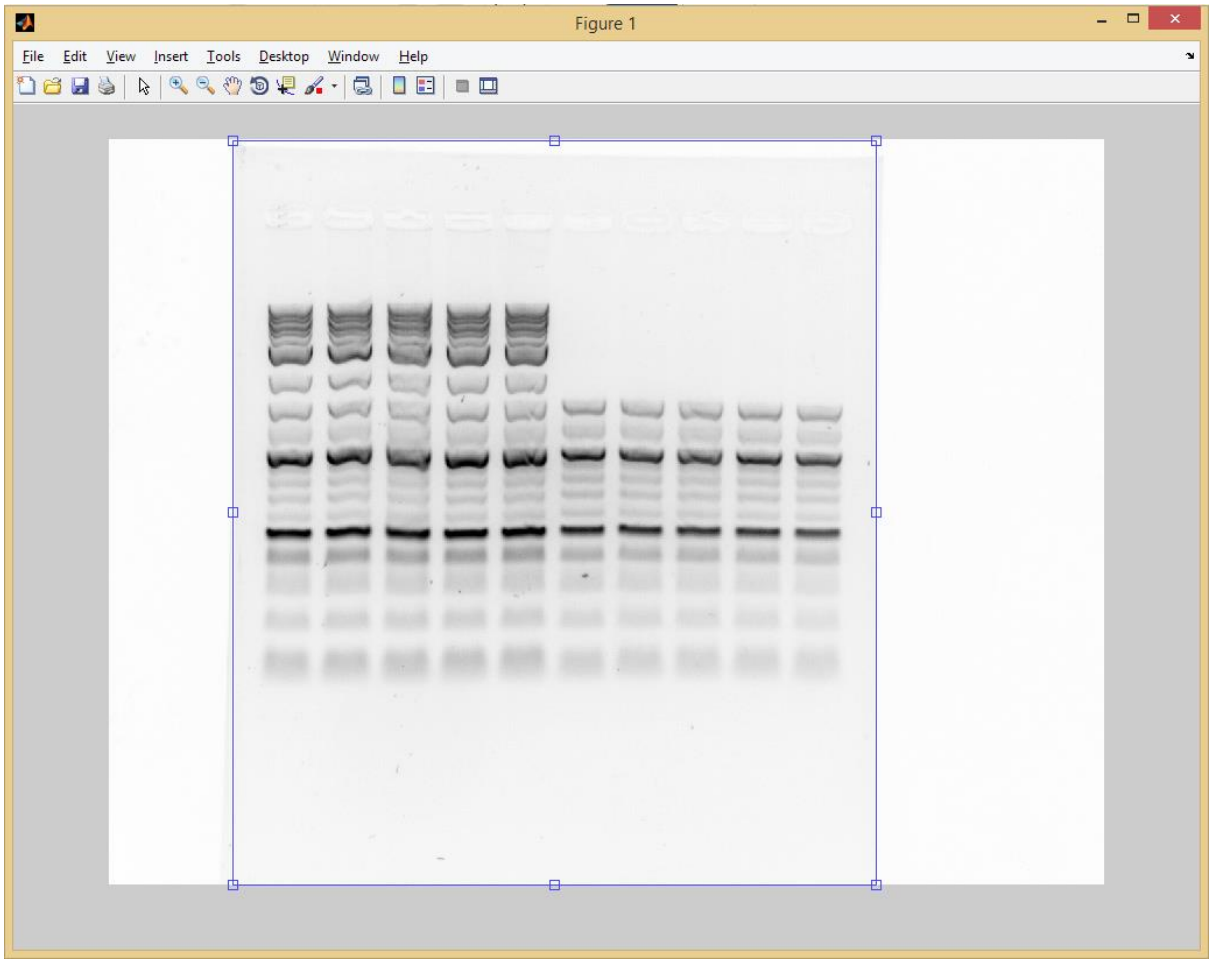
4.2 Realizace

Následuje realizace samotného algoritmu.

4.2.1 Předzpracování

V první části algoritmu je načten příslušný obrázek, v bezztrátovém formátu .tiff, a je provedena jeho normalizace a převedení obrazu do šedotónového. Také je ošetřeno, aby v obraze byly proužky reprezentovány černou barvou a pozadí barvou bílou. Je vypočítána průměrná hodnota obrazu, a pokud tato hodnota je menší, než 0,5 je provedena inverze.

Je také nutný uživatelský vstup, při kterém je zvolen počet linií v obraze a také metoda ořezu obrazu – automatická nebo manuální. Při volbě automatického ořezu je obraz vyhlazen mediánovým filtrem s oknem 5*5px a vysoké hodnoty větší než 0,95 jsou zaokrouhleny na 1. To má za následek vyhlazení okolí gelu a jeho snazší detekci. Poté je provedena detekce hranic gelu porovnáváním hodnot a následný ořez. Manuální ořez je proveden pomocí funkce *imcrop*, viz Obrázek 12. Uživatel zde pomocí myši označí vhodnou oblast a potvrdí ji dvojitým klikem na modré ohraničení. Je žádoucí, aby uživatel označil celý gel, ne pouze výřez, a to kvůli spolehlivosti detekce.

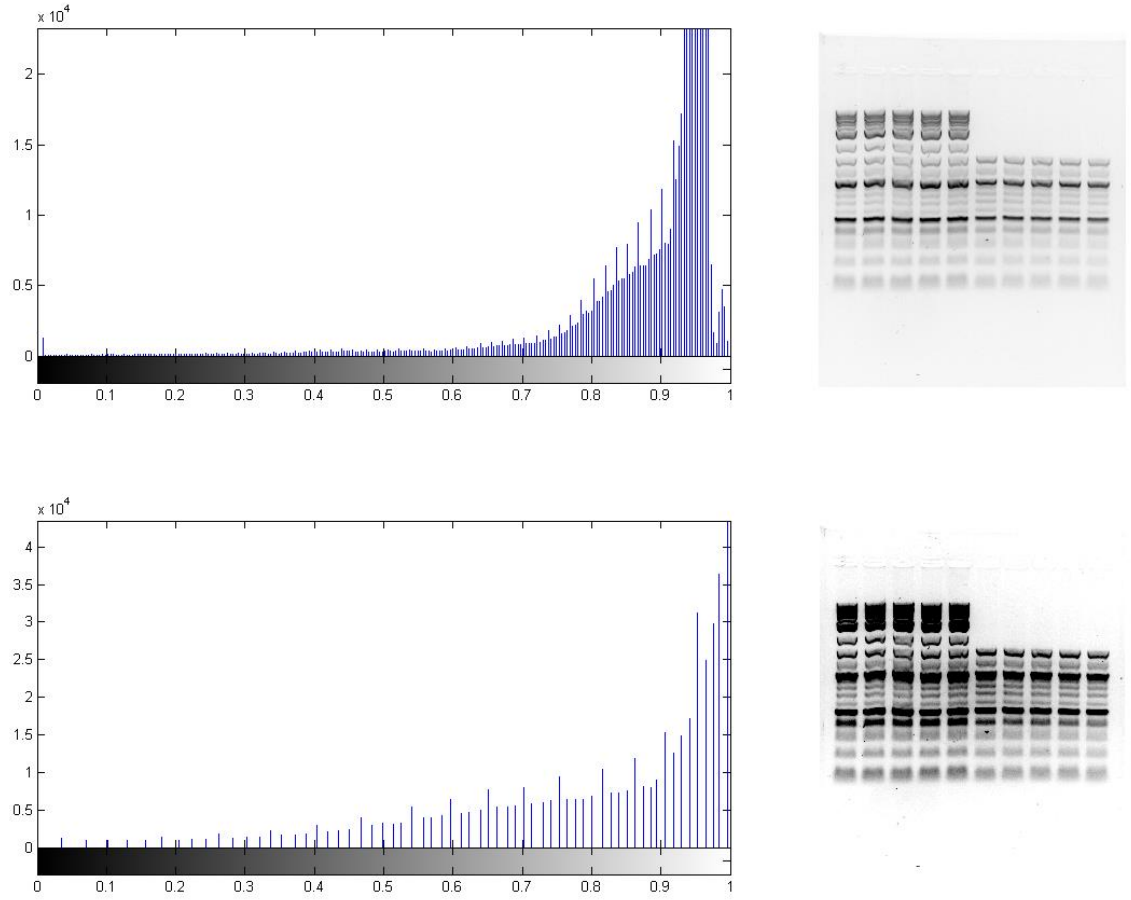


Obrázek 12: Ukázka funkce imcrop

Následuje již samotné předzpracování obrazu, jsou zvoleny parametry začátku a konce intervalu, ve kterém má proběhnout *po částech lineární transformace* obrazu a následně je zvolena hodnota γ pro *gama korekci* – Obrázek 13. Hodnoty *po částech lineární* transformace byly zvoleny na základě prvního histogramu, kde je v tomto intervalu zastoupeno nejvíce složek obrazu, tedy začátek=0,45 a konec 0,92. Hodnota γ byla zvolena menší než 1 – přeexponovaný obraz.

Tyto hodnoty je možné nastavit na pevno v rámci jednoho data setu s předpokladem podobně laděných obrazů. V případě nefungující detekce, je žádoucí parametry přenastavit.

V této práci jsou pro data set vytvořený v laboratoři použity pouze tyto uvedené hodnoty předzpracování.



Obrázek 13: Transformace kontrastu

Nutno podotknout, že takto zpracovaný obraz slouží pouze pro detekci linií a proužků. Pro vykreslení výsledků je opět použit obraz původní.

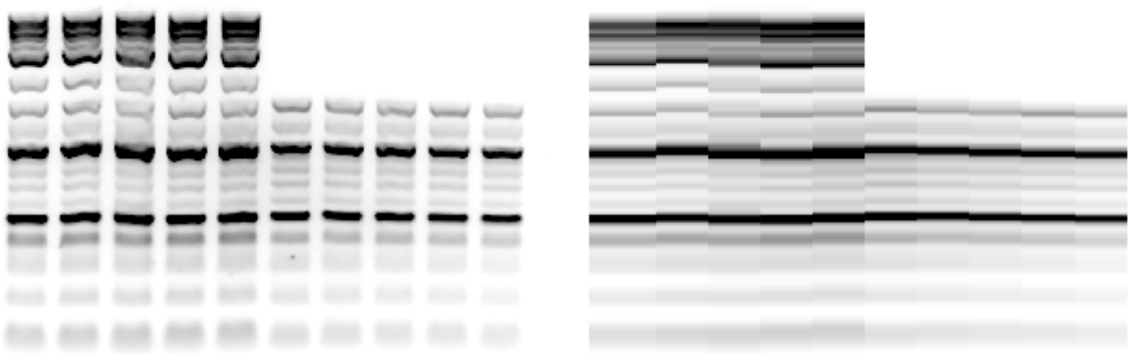
4.2.2 Detekce linií

Pro detekci linií je vypočtena standardní odchylka v první polovině obrazu. Dle literatury je vhodné vybrat pouze pixely z první třetiny obrazu, avšak pro některé laddery, jmenovitě proužky 100bp ladderu začínají až kolem třetiny obrazu, tudíž je polovina vhodnější. [1]

Takto je získán 1D průběh standardní odchylky obrazu. Průběh je dále vyhlazen mediánovým filtrem s oknem 5px a invertován. Následně je vypočtena minimální vzdálenost píku podle šířky ořízнutého snímku a počtu zvolených linií. Od této vzdálenosti je odečtena hodnota 15px, pro případ nestandardní vzdálenosti píků. Jsou nalezena lokální maxima v tomto průběhu a jejich polohy jsou uloženy.

4.2.3 Převedení obrazu na medián

V tomto kroku je obraz rozdělen na jednotlivé linie a v každé z nich algoritmus prochází linii po řádcích. V každém kroku jsou vybrány všechny hodnoty konkrétního řádku. Z vybraných hodnot je vypočten jejich medián, který následně nahrazuje všechny pixely v konkrétním řádku. Ukázka výpočtu mediánu obrazu z obrazu předzpracovaného je na Obrázek 14.



Obrázek 14: Převedení obrazu na medián: Vlevo se nachází předzpracovaný obraz a na pravé straně je zobrazen medián tohoto obrazu

4.2.4 Detekce proužků

Detekce je počítána v cyklu, který se zopakuje tolikrát, kolik je detekováno linií. Každá linie je vypočtena zvlášť. Na začátku se provede výběr linie a otočení obrazu do horizontální roviny – o 90° po směru hodinových ručiček.

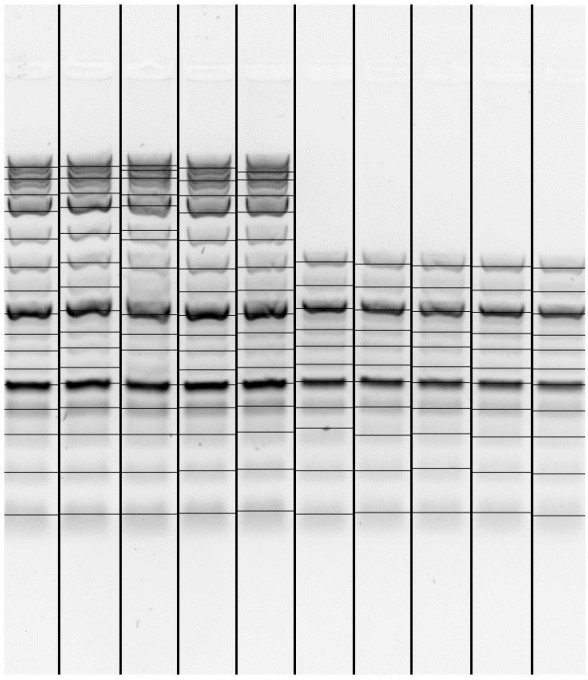
V tomto případě je medián obrazu – z každé linie se vybere pouze řádek pixelů, čímž dostaneme 1D signál odpovídající reálným hodnotám mediánu obrazu. Tento 1d signál je následně invertován.

Je zde nastaven práh detekce, který byl zvolen jako 10% z maximální hodnoty průměru. Tato hodnota byla nastavena heuristicky a nemusí vykazovat kladné výsledky pro jiné, než obrazy blížící se ideálním.

Opět dochází k detekci lokálních maxim a jejich polohy jsou uloženy.

4.3 Výsledky

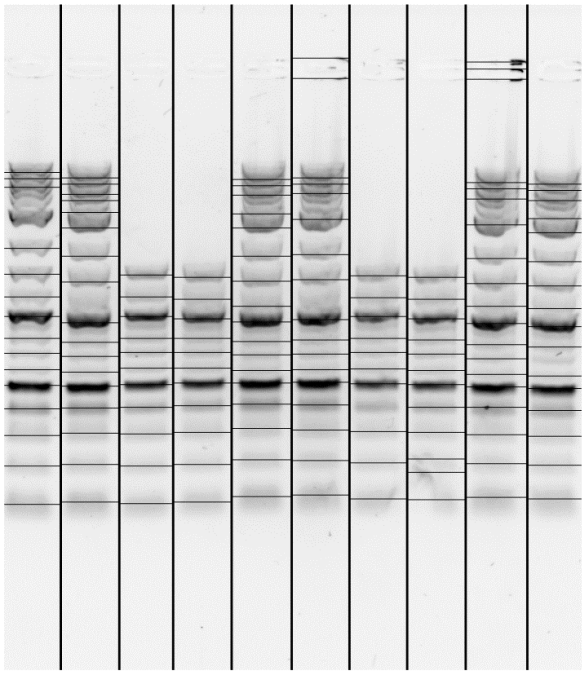
Při detekci hranic jednotlivých vzorků je žádoucí správná detekce všech linií a proužků a vyhnutí se například falešně pozitivních, či falešně negativních detekcí. Příklad detekce v obrazu *elfo_id_3_a.tif* provedené dle výše popsaného algoritmu je na Obrázek 15.



Obrázek 15: Detekce elfo_id_3_a.tif

Z tohoto obrazu lze vidět, že byly detekovány prakticky všechny proužky. A to bez jediné falešně pozitivní detekce.

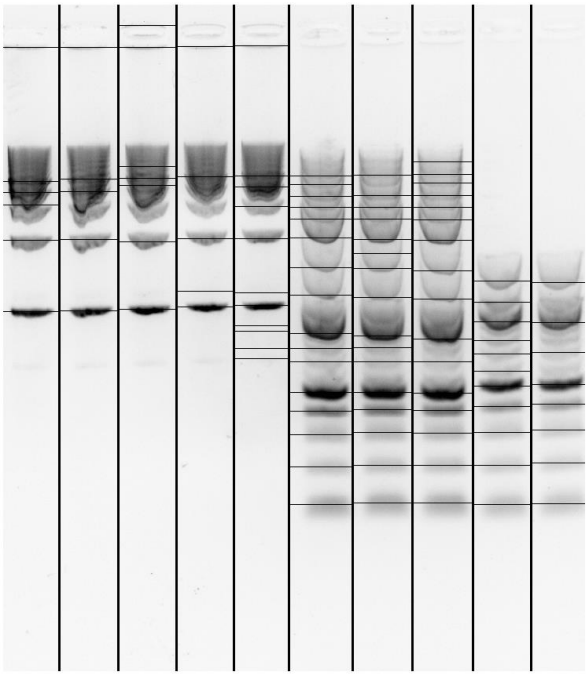
Další příklad detekce je na Obrázek 16.



Obrázek 16: Detekce elfo_id_7_a.tif

V tomto případě došlo k několika falešně negativním detekcím a to například díky poškození proužků v liniích 7 a 8. Také došlo k falešně pozitivním detekcím v místě propíchnutí gelu při pipetování.

Jiná ukázka detekce je na Obrázek 17.



Obrázek 17: Detekce elfo_id_4_a.tif

V tomto případě detekce naráží na hranice kvality analyzovaného obrazu, avšak opět je většina proužků úspěšně detekována.

Hodnoty výsledků jsou zobrazeny v ***Tabulka 2***. Jako testovací obrazy byly použity obrazy *elfo_id_3_a*, *elfo_id_4_a* a *elfo_id_7_a*. Z hodnot je patrné, že detekce je úspěšná z takřka 90%, avšak selhává ve falešně negativní detekci v jednom z deseti případů. Velkou měrou na úspěšnosti má především správné předzpracování obrazu a především nastavení hodnot parametrů předzpracování.

Tabulka 2: Hodnoty úspěšnosti detekcí

	Proužek	Bez proužku
Detekce	89,45%	5,61%
Bez detekce	10,55%	-

Je nutné dodat, že uvedené výsledky byly dosaženy za použití prostředí MATLAB ve verzi 2014a. Při použití starších verzí nemusí být výsledky konzistentní. Konkrétně u verze 2008b vykazovala funkce *findpeaks* diametrálně odlišné detekce, které měly řádově horší výsledky.

5 Klasifikace vzorků

Takto detekované linie a proužky je možné použít jako vstupní data pro následnou klasifikaci vzorků 1D gelové elektroforézy ve smyslu zařazování podobných, či stejných vzorků do jednotlivých shluků.

Takto lze shlukovat reálné vzorky například s ladderem a následně odečíst přesné hodnoty délky sekvencí.

5.1 Volba příznaků

Volba příznaků neboli určitých vlastností vzorků jej charakterizující je stěžejní částí vypracování shlukové analýzy jako takové. Jde o vybrání co nejobjektivnějšího ukazatele podobnosti, který sdílejí všechny vzorky obsažené v gelu. Pokud tento požadavek není splněn, může dojít ke zkreslení výsledků a chybné rozřazení vzorků do shluků.

V této práci byly vybrány dva typy příznaků:

- Medián jako 1D signál
- Parametry detekovaných proužků

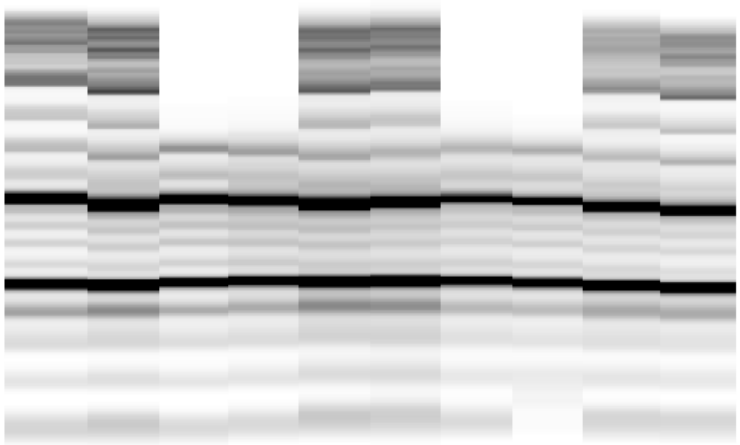
Tyto příznaky byly vybrané po dohodě s vedoucím práce, ovšem ve výsledku výběr záleží na konkrétním použití shlukové analýzy.

5.1.1 Medián linií jako 1D signál

Prvním typem příznaku je medián jako 1D signál.

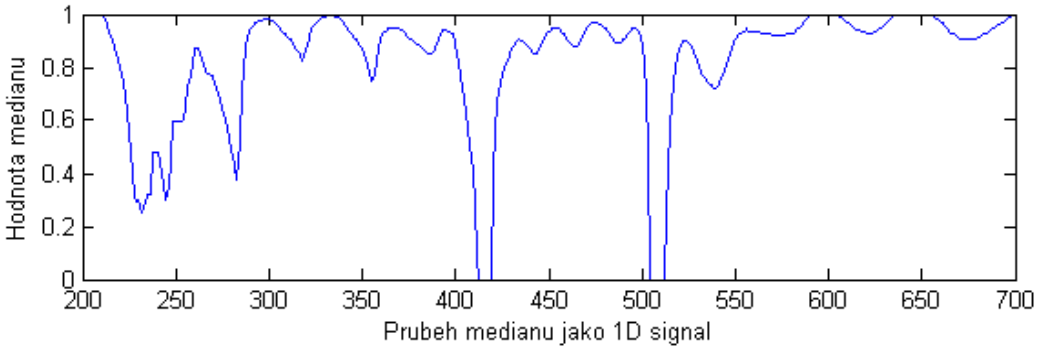
Tento konkrétní typ příznaku se vyznačuje především značnou jednoduchostí získání matice příznaků pro jednotlivé vzorky. V předzpracovaném obrazu jsou detekovány linie hranic jednotlivých vzorků. Obraz je následně převeden na medián, tímto získáme vstupní obraz pro navazující vytvoření matice příznaků - Obrázek 18.

Tímto krokem je eliminována například detekce proužků, které může při zhoršení výsledků detekce zkreslovat matici příznaků.



Obrázek 18: Vstupní obraz provolbu příznaku Medián jako 1D signál

Z každé linie je vybrán pouze jeden sloupec pixelů – tímto jsou získány 1D signály těchto linií – Obrázek 19.



Obrázek 19: Medián jako 1D signál

Průběhy těchto signálů jsou uloženy do matice a slouží jako vstupní hodnoty pro výpočet podobností.

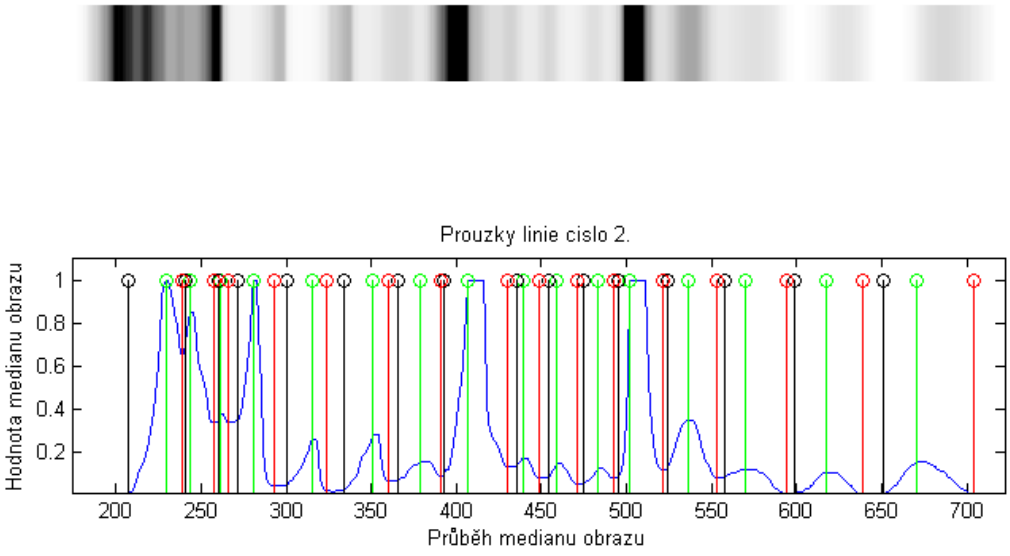
5.1.2 Parametry detekovaných proužků

Jako druhá varianta výběru příznaků byly vybrány Parametry detekovaných proužků. Výběr vychází z tvrzení, že podobné vzorky mají podobné proužky z hlediska jejich parametrů. Použité parametry jsou:

- Počet proužků
- Polohy proužků
- Šířky proužků
- Průměrné hodnoty proužků

Počty a polohy proužků jsou již detekovány v kapitole 4.2.4. Je tedy nutné vypočíst šířku každého proužku a jeho průměrnou hodnotu.

Šířka proužku je vypočtená na základě detekované polohy proužků, kdy je na signál přivedeno testovací okno, které se posouvá po signálu a kontroluje rozdíl první a poslední hodnoty okna. Pokud je rozdíl nulový (první a poslední hodnoty jsou stejné), je detekován konec proužku. Tento postup je vykonán dvakrát, jednou pro začátek proužku a jednou pro konec proužku. Ukázka detekce je na Obrázek 20.



Obrázek 20: Detekce šířky proužků. Nahoře je detekovaný vzorek, dole se nachází průběh hodnot vzorku společně s příslušnými detekcemi. Polohy proužků jsou označeny zelenou barvou. Černou barvou jsou označeny začátky a červenou konce

Pro výpočet hodnot proužků jsou použity detekované polohy proužků, jejich konce a začátky. Z původního obrazu jsou vybrány detekované proužky a tyto hodnoty jsou zprůměrovány.

Komplikace nastává v případě nestejného počtu proužků, kdy chybějící hodnoty jsou doloženy nulami. Dle předpokladu, že počet proužků charakterizuje podobné vzorky, je toto ošetření chybějících hodnot dostačující a neovlivňuje jejich vzájemnou podobnost. Naopak doložení nulami zvyšuje vzdálenost mezi vzorky s nestejným počtem proužků.

Všechny hodnoty zvolených parametrů jsou uloženy do jedné matice, která je vstupem pro výpočet podobností.

5.2 Použité metody

Při výběru metod pro výpočet podobností a shlukování byl zohledněn požadavek pro vyzkoušení co největšího počtu těchto metod. Proto byly zvoleny veškeré dostupné metody ve *Statistics and Machine Learning Toolbox*u prostředí MATLAB, a to i přes teoretickou nekompatibilitu některých metod shlukování s jinou, než euklidovskou vzdáleností příznaků. Tyto kombinace metod by tedy měly teoreticky vykazovat řádově horší výsledky, než kombinace navzájem kompatibilní.

5.2.1 Výpočet vzdáleností

Jelikož jsou v prostředí MATLAB vypočítávány vzdálenosti jednotlivých vzorků namísto jejich podobností, je tento krok i související proměnné v kódu pojmenovány jako výpočet vzdáleností (distancí).

Výběr metod, které byly popsány v kapitole 2.7.4, je následující:

- Euklidovská vzdálenost (euclidean)
- City block (cityblock)
- Minkovského metrika (minkowski)
- Chebyshevova vzdálenost (chebyshev)
- Cosinová vzdálenost (cosine)
- Korelační vzdálenost (correlation)
- Spearmanova vzdálenost (spearman)
- Hammingova vzdálenost (hamming)
- Jaccardova vzdálenost (jaccard)

Kde v závorkách je označení metod v prostředí MATLAB u funkce *pdist*.

5.2.2 Shluková analýza

Obdobně jako výběr metod vzdáleností byly vybrány metody shlukování. Tyto metody jsou:

- UPGMA (average)
- WPGMC (median)
- UPGMC (centroid)
- WPGMA (weighted)
- SLINK (single)
- CLINK (complete)
- Wardova metoda (ward)

Kde v závorkách je označení metod v prostředí MATLAB u funkce *linkage*.

5.3 Realizace metod

V této kapitole budou popsány konkrétní realizace jednotlivých metod v prostředí MATLAB.

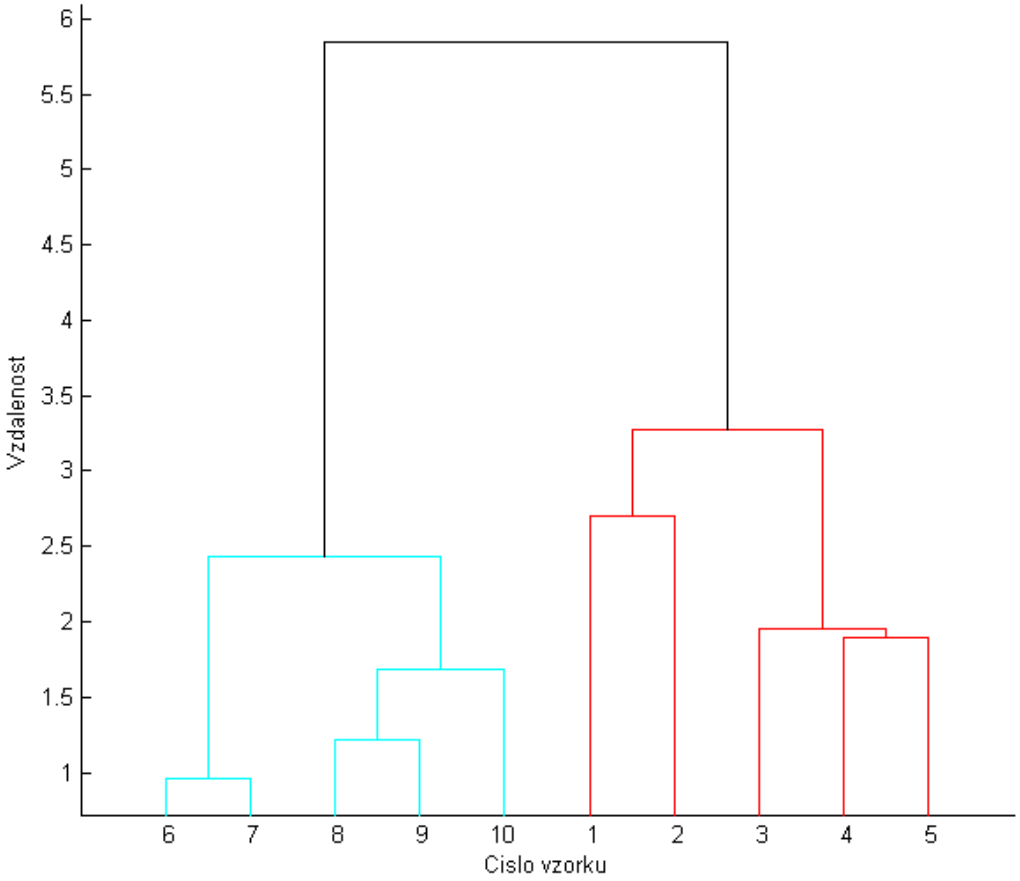
5.3.1 Medián linií jako 1D signál

Tato metoda využívá převedený obraz na medián, popsaný v kapitole 4.2.3.

V každém vzorku je vybrán první sloupec pixelů, tímto je získán 1D signál, zobrazený na Obrázek 19. Tyto 1D signály jsou poté uloženy do matice příznaků, kde každý řádek reprezentuje jednotlivé vzorky. Výsledná matice tedy může mít rozměry například 10x960 vzorků – což reprezentuje deset vzorků a devět set šedesát vzorků 1D signálu.

Pomocí funkce *pdist*, s příslušným nastavením metody výpočtu, je následně z matice příznaků vypočtena matice vzdáleností jednotlivých vzorků.

Poté je funkcí *linkage* vypočten aglomerativní hierarchický shlukovací strom z matice vzdáleností. Tento strom je následně vykreslen funkcí *dendrogram*. Barevné odlišení shluků je provedeno výpočtem maximální hodnoty vzdálenosti shluků a aplikování parametru *colorthreshold* na příkaz *dendrogram*. Příklad takového stromu je na Obrázek 21.



Obrázek 21: Vypočtený dendrogram s použitím příznaku Medián jako 1D signál

5.3.2 Parametry proužků

Tato metoda obdobně využívá obraz převedený na medián a také originální oříznutý obraz jako vstupní hodnoty.

Pro vytvoření matice příznaků je nutné znát nejen polohu proužků ve vzorcích, ale i jejich šířku a průměrnou hodnotu. K výpočtu šířky proužku je využit cyklus již představený v kapitole 4.2.4. V tomto cyklu je označen začátek i konec každého detekovaného proužku.

Algoritmus detekce spočívá v plovoucím okně o velikosti tří vzorků, které se posouvá po signálu. Testování začíná od polohy detekce proužků, okno se posouvá po signálu s jednotkovým krokem a v každém cyklu je vypočtena diference prvního a posledního pixelu v okně. Pokud tato diference bude nula, tedy první a poslední pixel jsou stejné hodnoty, uloží se poloha okna a je detekován začátek, či konec proužku. Tento test je proveden v kladném i

v záporném směru signálu a to pro detekci konců a začátků proužků. Výsledek detekce začátků a konců proužků je na Obrázek 20.

Takto získáme matice začátků a konců proužků. Šířka proužků je vypočtena prostým odečtením zmíněných matic.

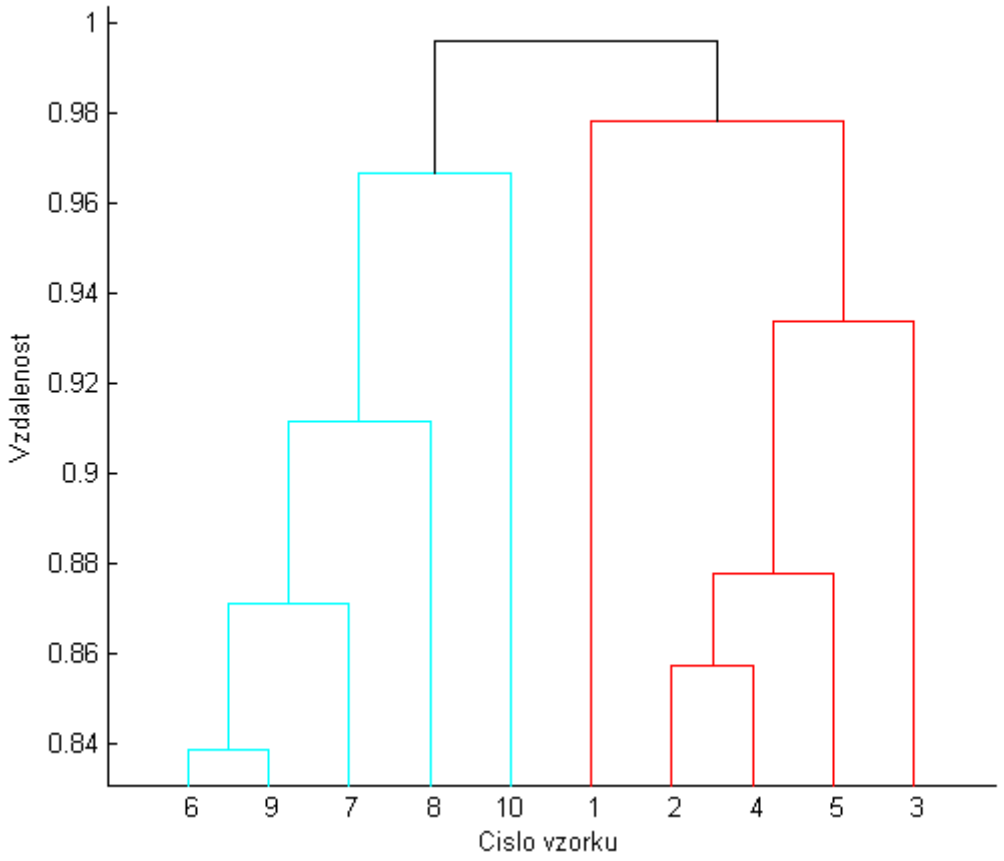
Ze zmíněného cyklu je využita také matice poloh proužků. Pokud je ve vzorcích rozdílný počet proužků, jsou chybějící hodnoty nahrazeny nulami.

Průměrná hodnota proužků je vypočtena průměrováním vyříznutého úseku původního obrazu na základě detekce začátků a konců proužků.

Následně jsou ošetřeny nulové hodnoty chybějících proužků pomocí logického indexování, kdy jsou nuly dosazeny do matice průměrných hodnot na základě nul v matici poloh proužků.

Dále je vytvořena matice příznaků seskupením všech vypočtených parametrů proužků. Na každém řádku jsou jednotlivé vzorky na gelu a ve sloupcích jsou vyneseny hodnoty parametrů proužků. Nejprve je vynesena hodnota počtů proužků, následně jsou přidány hodnoty poloh proužků, šířek proužků a konečně průměrných hodnot proužků. Výsledná matice tedy může mít rozměry například 10x52 vzorků – což reprezentuje deset vzorků, jeden sloupec počtů linií a tři krát sedmnáct parametrů proužků.

Obdobně jako v kapitole 5.3.1 je pro výpočet matice distancí použit příkaz *pdist* a pro shlukování příkaz *linkage*. Je také vypočten práh vykreslení z maximální hodnoty vzdáleností shluků a vykreslen dendrogram – Obrázek 22.



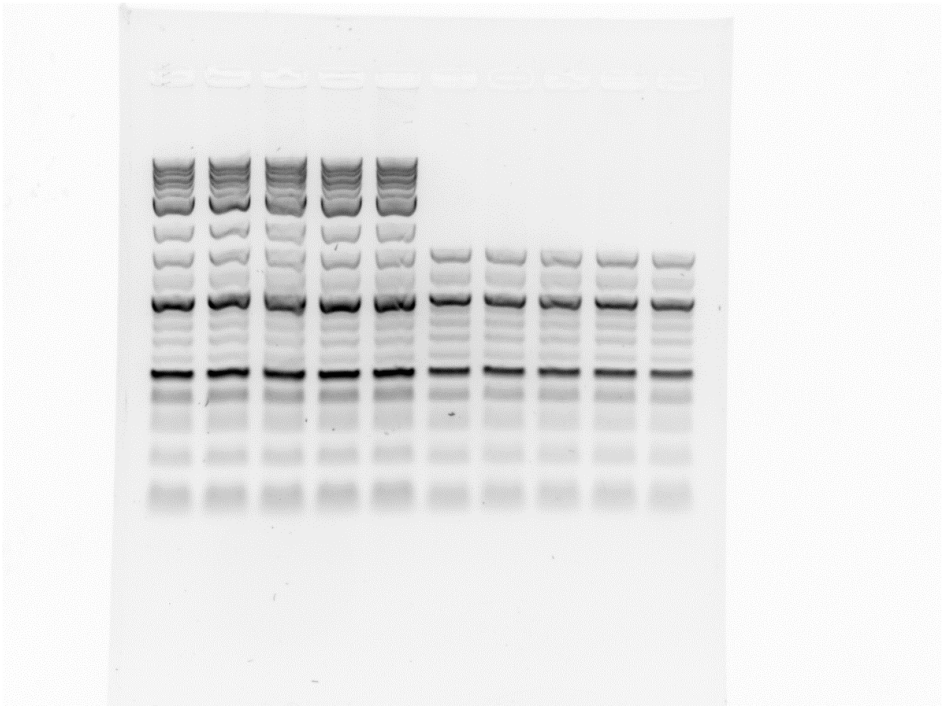
Obrázek 22: Vypočtený dendrogram s použitím příznaku Parametry proužků

5.4 Výsledky klasifikace

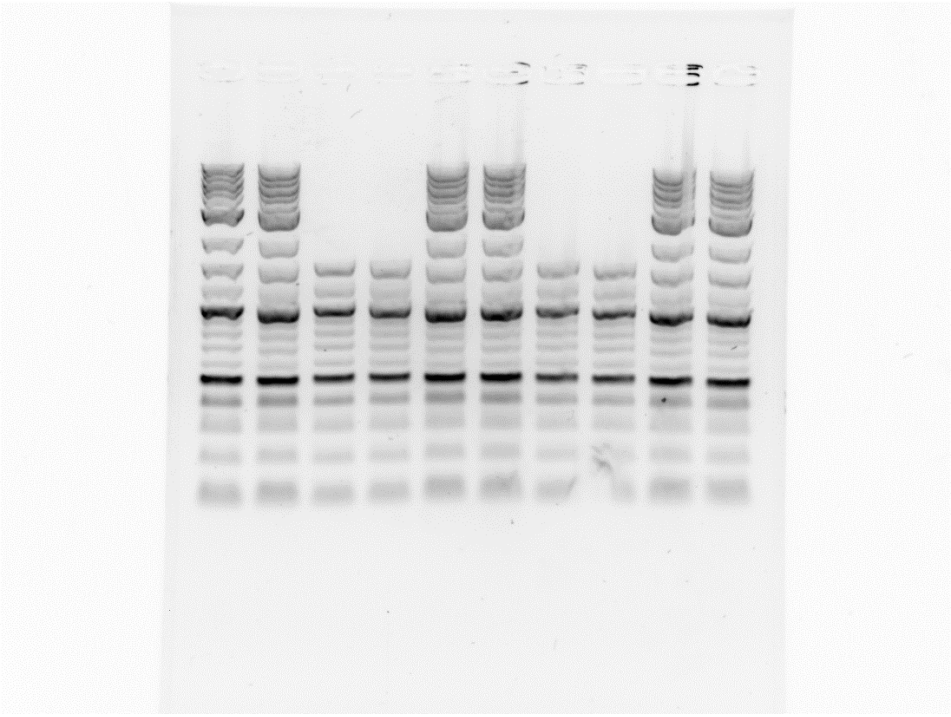
V této kapitole budou představeny výsledky klasifikace obou metod příznaků. Pro porovnání metod bude sloužit výběr euklidovské distance a shlukování pomocí UPGMA. V tomto případě se jedná o demonstraci spolehlivosti shlukování u obou metod příznaků. Statistická analýza výsledků jednotlivých metod a jejich kombinací, za pomoci vybraného kritéria účinnosti, bude popsána v kapitole 6.

Pro zjištění účinnosti shlukování byly vybrány dva obrazy, které obsahují dva druhy ladderu. Při správné funkci shlukování by tyto dva laddery měly od sebe být bezpečně rozlišeny v dendrogramu jeho morfologií a také barvou.

Tyto vstupní obrazy jsou *elfo_id_3_a* (Obrázek 23) a *elfo_id_7_a* (Obrázek 24).



Obrázek 23: elfo_id_3_a jako první testovací obraz



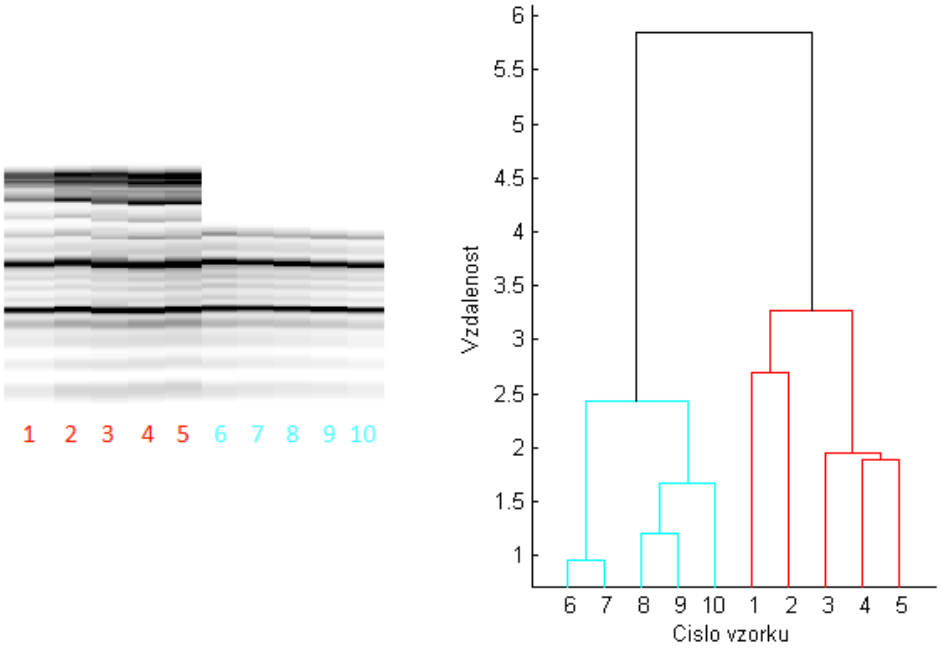
Obrázek 24: elfo_id_7_a jako druhý testovací obraz

5.4.1 Medián linií jako 1D signál

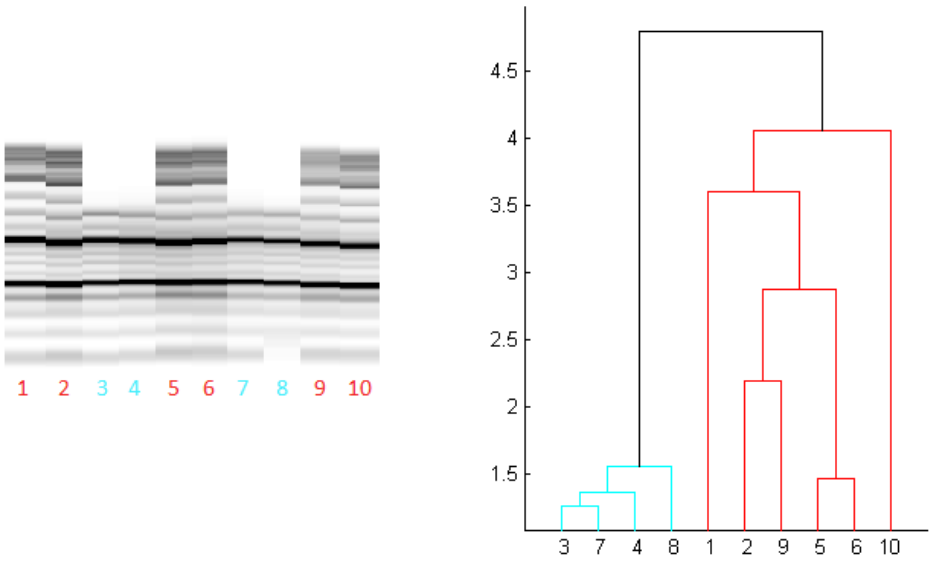
Následují výsledky shlukování pro příznaky typu Medián linií jako 1D signál. Jak již bylo zmíněno, použité metody jsou:

- Euklidovská vzdálenost
- UPGMA

Testování bylo provedeno na obrazech z kapitoly 5.4. Byl zadán příslušný počet linií (10) a byly nastaveny parametry předzpracování z kapitoly 4.2.1.



Obrázek 25: Výsledky shlukování pomocí příznaků Medián jako 1D signál obrazu elfo_id_3_a, kde vlevo je předzpracovaný obraz a vpravo výsledný dendrogram

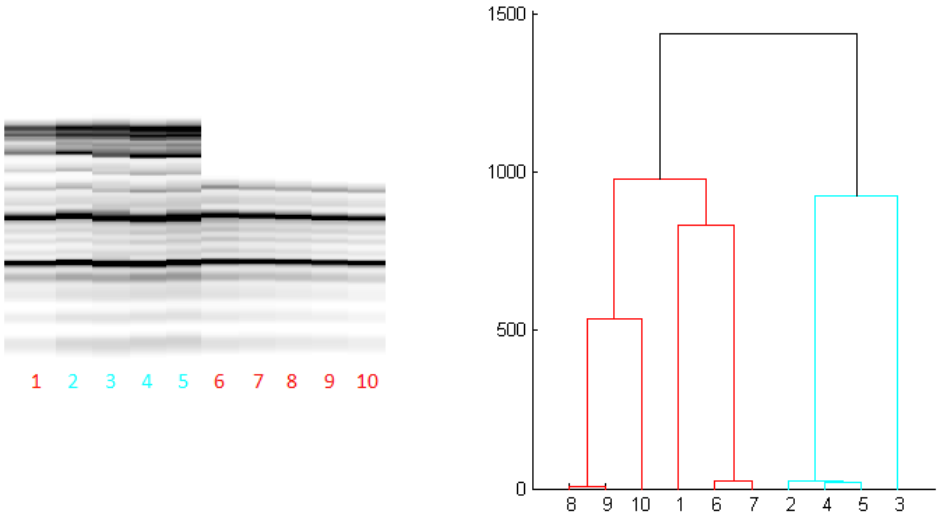


Obrázek 26: Výsledky shlukování pomocí příznaků Medián jako 1D signál obrazu elfo_id_7_a, kde vlevo je předzpracovaný obraz a vpravo výsledný dendrogram

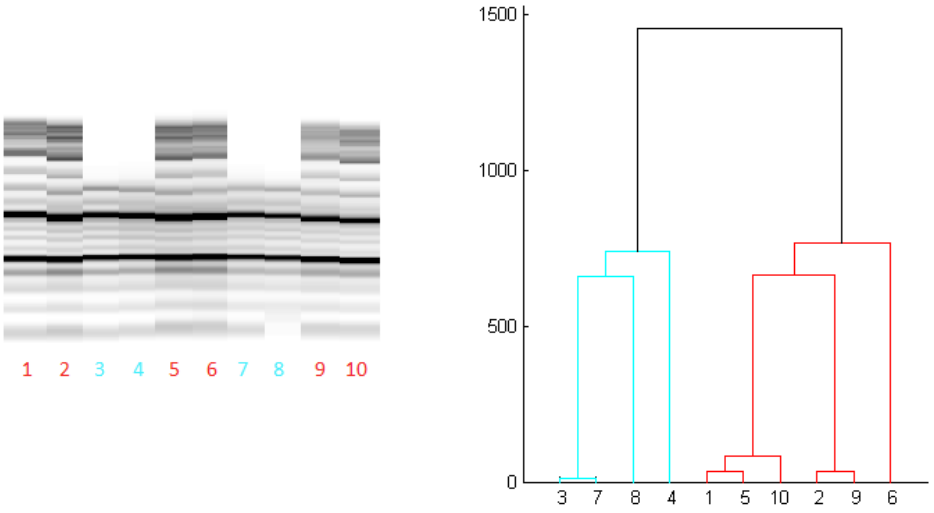
Z uvedených výsledků je patrné, že jednotlivé obrazy byly shlukovány úspěšně. Vypočtené vzdálenosti shluků stejných vzorků jsou způsobeny především vzájemným posunutím jednotlivých vzorků ve vertikálním směru. Hodnoty vzdáleností budou dále rozebrány v kapitole 6.

5.4.2 Parametry proužků

Identicky, jako v předchozí kapitole, jsou nastaveny parametry předzpracování, metody výpočtu vzdáleností, shlukování a vstupní obrazy.



Obrázek 27: Výsledky shlukování pomocí příznaků Parametry proužků obrazu elfo_id_3_a, kde vlevo je předzpracovaný obraz a vpravo výsledný dendrogram

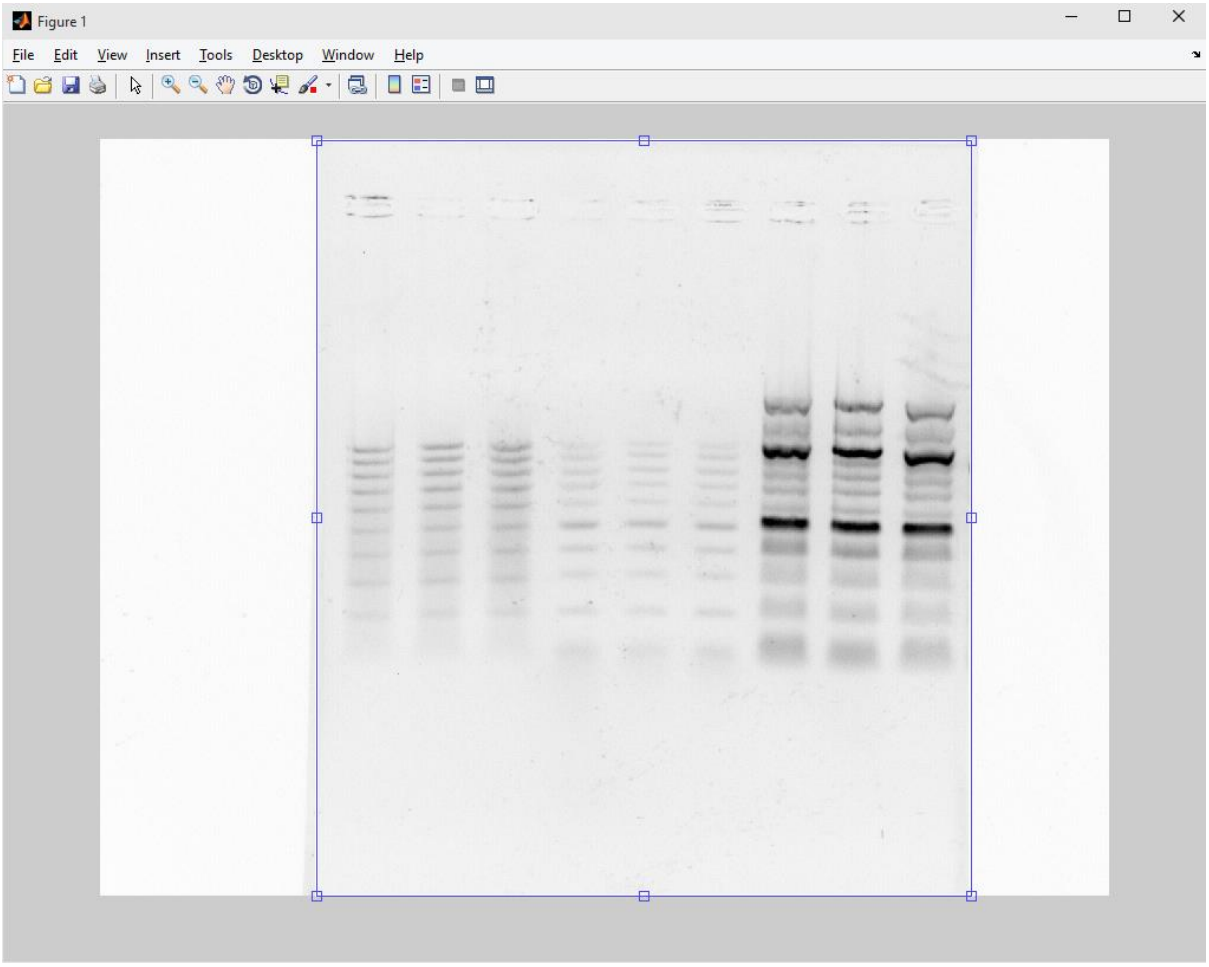


Obrázek 28: Výsledky shlukování pomocí příznaků Parametry proužků obrazu elfo_id_7_a, kde vlevo je předzpracovaný obraz a vpravo výsledný dendrogram

Z Obrázek 27 je patrné, že byla první linie přiřazena do špatného shluku. Toto může být způsobeno výběrem metody výpočtu vzdáleností a metody shlukování. Při volbě metod *Cityblock* a *Ward*, dochází již ke správnému shlukování.

Druhý vzorek již byl klasifikován bez problémů. Vypočtené vzdáleností v rámci jednoho shluku jsou způsobeny především vzájemným posunem linií.

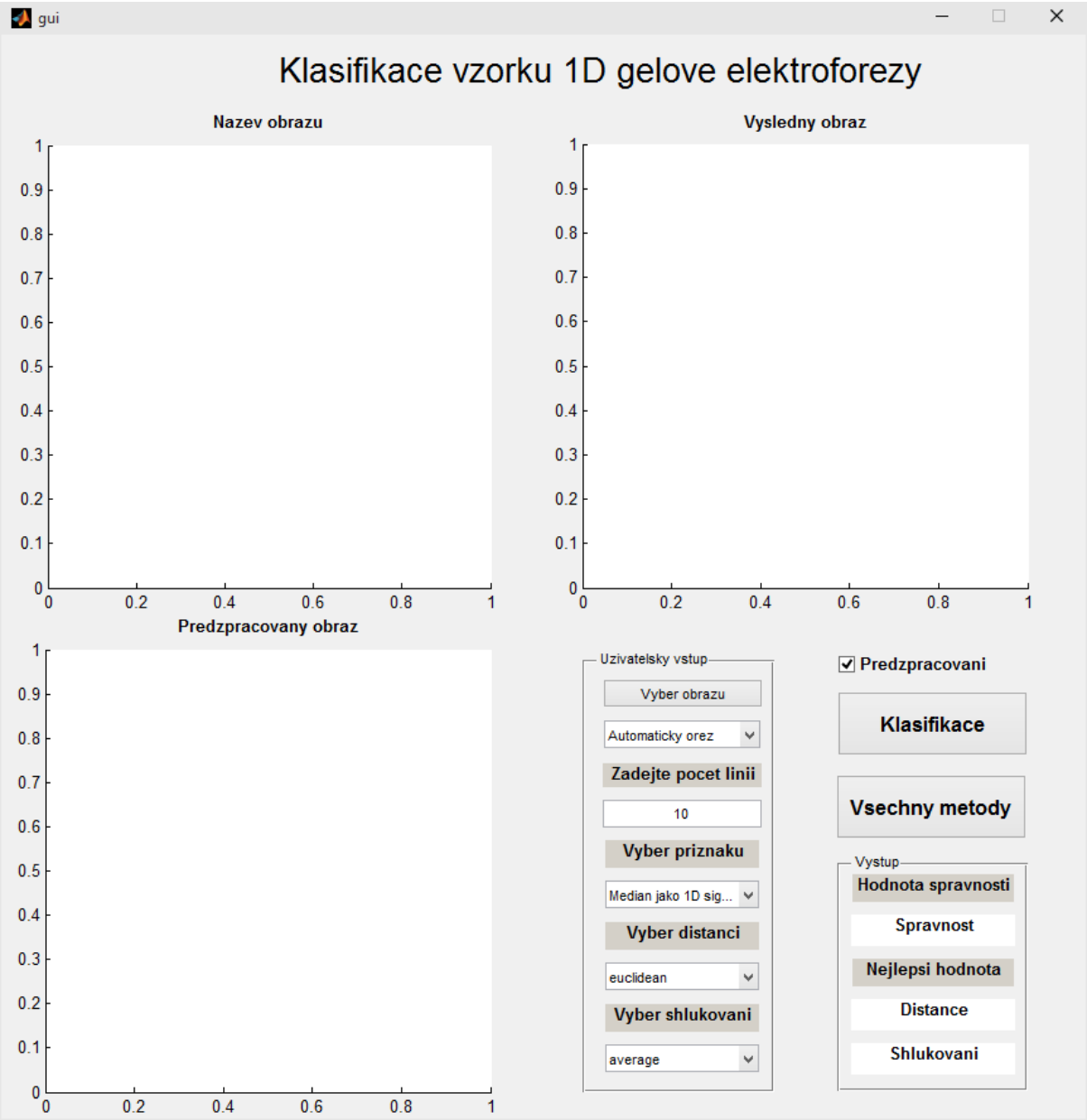
Je však nutné dodat, že pro některé obrazy (především se zhoršeným kontrastem, překrývajícími se proužky a nehomogenním pozadím gelu) selhává detekce šířek proužků. Pro správný chod algoritmu je nutné snímek správně (manuálně) oříznout tak, aby modré pole lícovalo s hranami gelu – Obrázek 29.



Obrázek 29: Doporučené oříznutí gelu

5.5 Grafické uživatelské prostředí

Dle zadání práce bylo vytvořeno grafické uživatelské prostředí. Toto GUI má za cíl co nejvíce usnadnit práci se zde vytvořeným algoritmem a od uživatele se při používání očekává pouze několik málo kroků k dosažení klasifikace snímků. Celý program se spouští pomocí skriptu *gui.m*. Toto prostředí je zobrazeno na Obrázek 30.



Obrázek 30: Grafické uživatelské prostředí

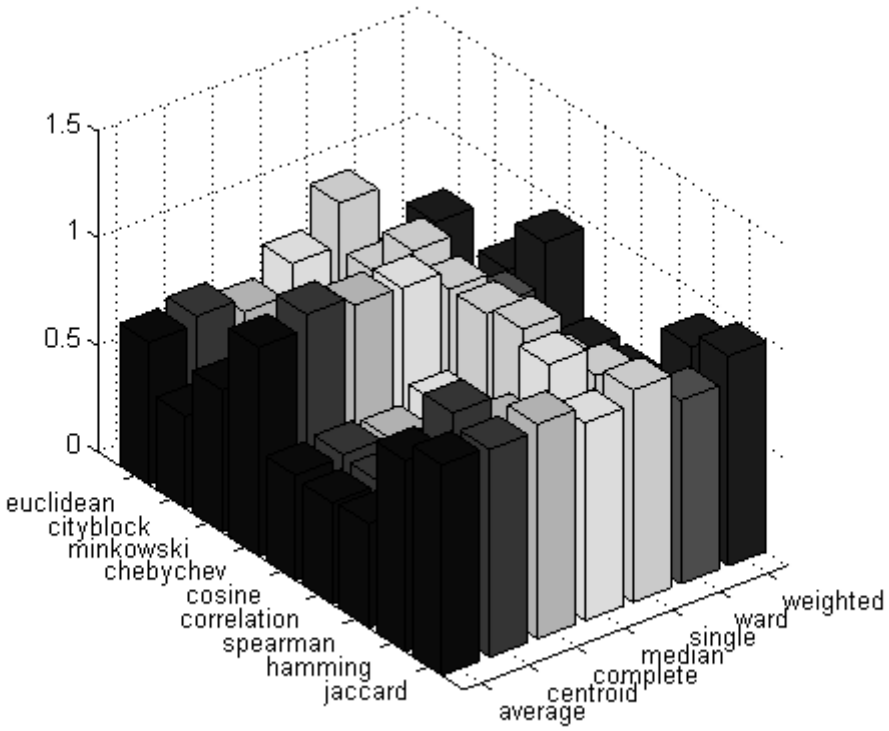
Grafické uživatelské prostředí se skládá ze čtyř částí:

- Levý horní roh – původní obraz
- Levý spodní roh – zpracovaný obraz
- Pravý horní roh – výsledný dendrogram
- Pravý spodní roh – uživatelský vstup a zobrazení výsledných hodnot

Pravý spodní roh se skládá z uživatelského vstupu, který obsahuje tlačítko *Vyber obrazu*. Po stisknutí dojde k vyvolání kontextové nabídky, ve které si uživatel zvolí testovaný obraz. Po dokončení výběru se obraz vykreslí v levém horním rohu. Pod tlačítkem se nachází *dropdown menu* pro vybrání metody ořezu obrazu. Je možné, že automatický ořez selže (z důvodu nehomogenního okolí gelu) a v tomto případě je žádoucí použít ořez manuální. Následuje uživatelský vstup pro zadání počtu linií (pro zlepšení detekce linií gelu). Dále tato část obsahuje výběr metody příznaků, výběr metody distancí a výběr metody shlukování.

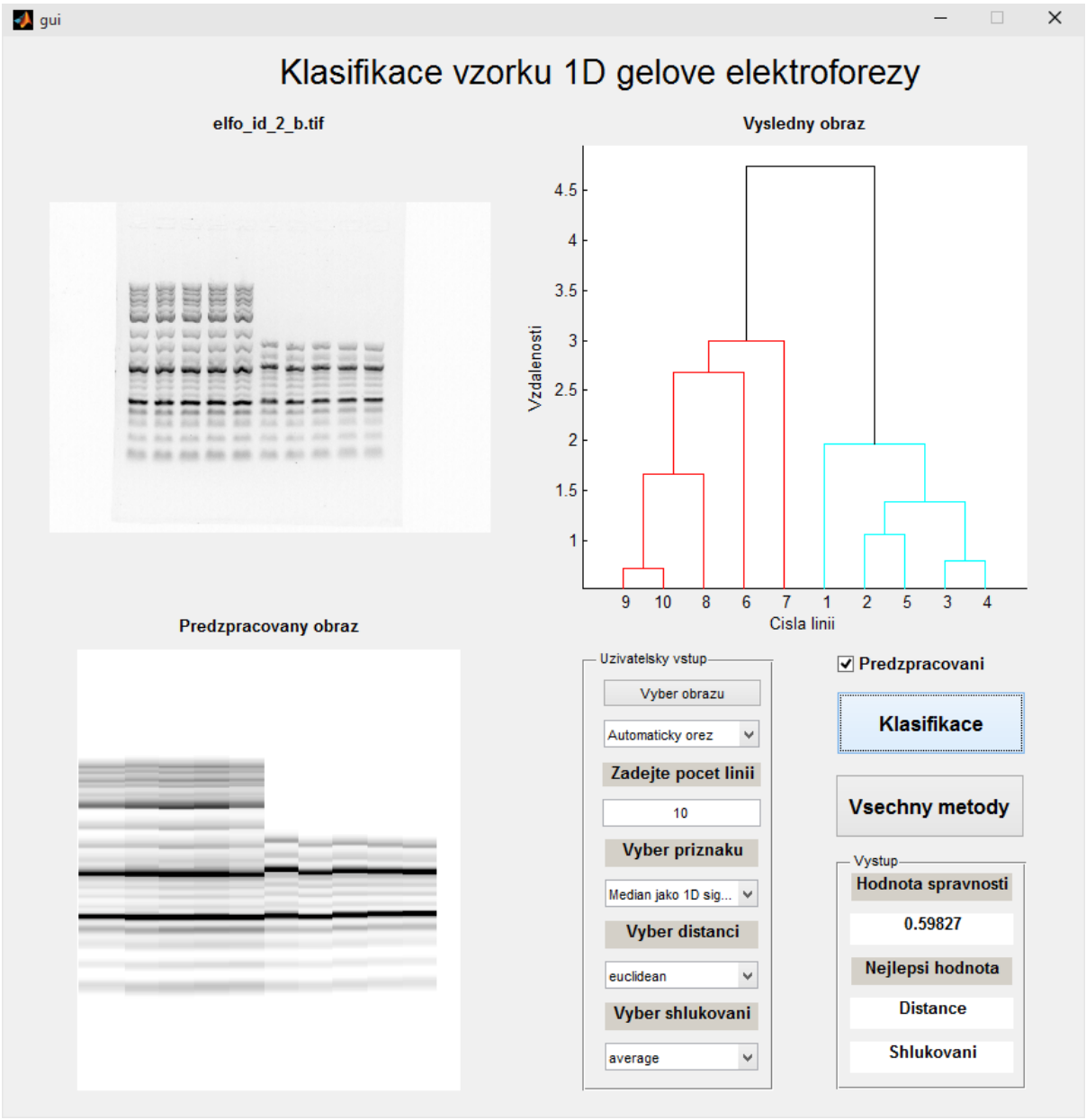
Vedle uživatelského vstupu se nachází zaškrtávací tlačítko, pro zamezení předzpracování obrazu dle kapitoly 4.2.1. Implicitně je tato volba zapnutá. Pokud je předzpracování vypnuté, tak i přesto dojde k oříznutí, detekci linií a k převedení obrazu na medián – to z důvodu samotné metody příznaků Medián linií jako 1D signál. Tato volba je vytvořena pro případ použití již předzpracovaných obrazů.

Následují tlačítka *Klasifikace* a *Všechny metody*. První z nich vykoná shlukování dle nastavených kritérií metody příznaků, metody distancí a metody shlukování. Vykreslí dendrogram obdobný tomu na Obrázek 28. Spočítá se hodnota kritéria účinnosti metod, výpočet této hodnoty bude představen v kapitole 6.2, která je zobrazena v poli *Hodnota spravnosti* v sekci *Vystup*. Při stisku tlačítka *Všechny metody* dojde k výpočtu všech kombinací metod výpočtu vzdáleností a výpočtů shlukování. A vykreslí se 3D sloupcový graf (Obrázek 31), kde na ose *x* jsou metody výpočtu vzdáleností, na ose *y* metody výpočtu shlukování a na ose *z* je vynesena hodnota vypočteného kritéria. V sekci *Vystup* je nyní zobrazena kombinace metod výpočtů s nejlepší hodnotou kritéria, která je obdobně vypsána do políčka *Hodnota spravnosti*.



Obrázek 31: Ukázka zobrazení výsledků pro kombinace všech metod výpočtu

Ukázka spuštěného programu je na Obrázek 32.



Obrázek 32: Spuštěné grafické uživatelské prostředí

6 Statistická analýza

V této kapitole budou vypsány a interpretovány výsledky statistické analýzy účinnosti představených metod. Bude představen vzorek zkoumaný dat a především určené kritérium účinnosti jednotlivých metod.

Pro statistickou analýzu byla připravena zvláštní funkce *testovaci_fce.m*, která výpočte všechny kombinace metod pro oba typy příznaků pro všechny vstupní obrazy. U některých obrazů je nutné zvolit manuální ořez pro správný běh algoritmu.

Jsou tedy postupně načítány obrazy, volány všechny vytvořené funkce a hodnoty výsledků se ukládají do dvou datových balíků (*data_1.m, data_2.m*) . Tyto datové soubory jsou koncipovány tak, že na řádcích jsou vynesené metody vzdáleností, ve sloupcích metody shlukování a ve třetím rozměru jsou hodnoty jednotlivých klasifikací. Pro devět metod vzdáleností, sedm metod shlukování a dvanáct klasifikovaných obrazů je velikost matice 9x7x12.

Tato data byla následně upravena v programu Microsoft Excel, kde byly vytvořeny tabulky pro statistickou analýzu těchto dat. Tento soubor se jmenuje *data.xls*. Statistická analýza je dále provedena v programu Statistica.

6.1 Vstupní data

Jako vstupní data byly zvoleny obrazy, které jsou dostupné v kapitole 10 – Obrazové přílohy. Použity byly i obrazy, které nedosahovaly kvalit ideálních obrazů, a to především z důvodu zvýšení objemu testovaných dat tedy zvýšení robustnosti statistické analýzy. Tyto obrazy jsou:

- elfo_id_2_b.tif
- elfo_id_3_a.tif
- elfo_id_4_a.tif
- elfo_id_5_a.tif
- elfo_id_7_a.tif
- elfo_id_8_a.tif
- elfo_id_8_b.tif
- elfo_id_11_a.tif
- elfo_id_11_b.tif

- elfo_id_12_a.tif
- elfo_id_13_a.tif
- elfo_id_14_b.tif

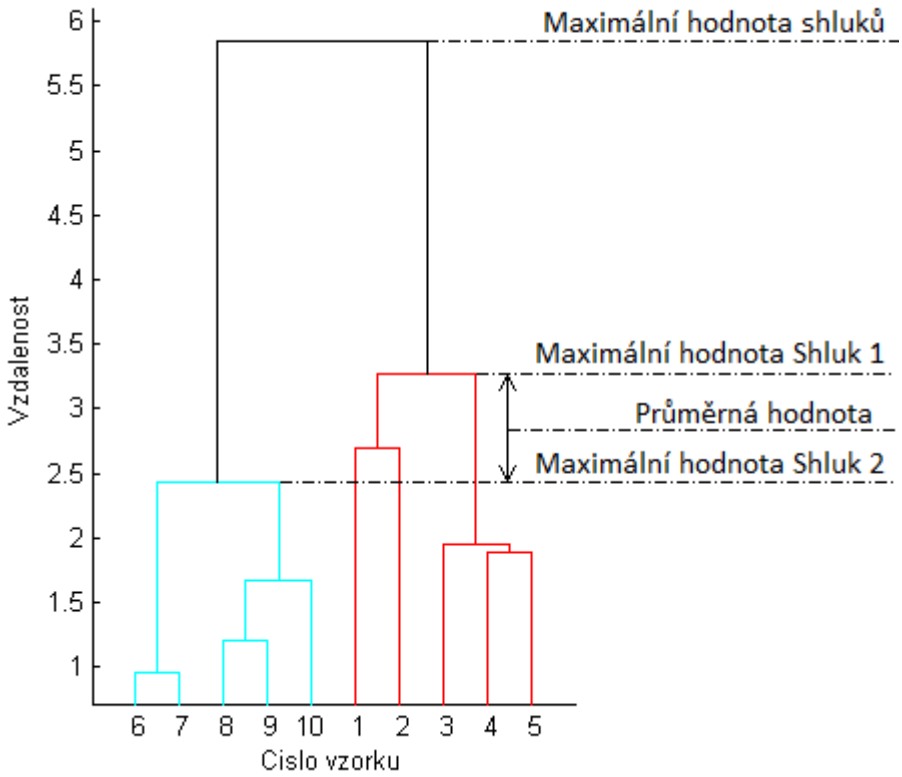
6.2 Zvolené kritérium účinnosti metod

Při výběru objektivního kritéria účinnosti shlukové analýzy byly zohledněny teoretické předpoklady, dle kterých dosahuje kvalitnější výsledek shlukování větších hodnot vzdálenosti jednotlivých shluků. Jinými slovy, stejné vzorky se shluknou hned (se vzdáleností 0) a vzdálenost rozdílných vzorků bude maximální (v ideálním případě nekonečná).

Kritérium účinnosti tedy bylo zvoleno jako:

$$Kritérium\ účinnosti = \frac{Průměr(\text{Maximum}(\text{Shluk}_1), Maximum(\text{Shluk}_2))}{Maximum(Shluk)}$$

kde dělenec vyjadřuje průměrnou hodnotu dvou nejvyšších shluků a dělitel je jejich vzdálenost. Tyto hodnoty jsou zobrazeny na Obrázek 33.



Obrázek 33: Kritérium účinnosti

Pro výslednou hodnotu kritéria tedy platí, že čím nižší je vypočtená hodnota, tím kvalitnějšího výsledku bylo dosaženo.

Je nutné zmínit, že toto kritérium považuje metodu za účinnou v tom případě, když je vzdálenost shluků co nejvyšší. Toto může být problém například při posuzování evoluční vzdálenosti organismů, kdy bude shlukování podobných organismů, tedy těch které jsou si evolučně blízko, označeno jako méně kvalitní.

Také toto kritérium nemusí být směrodatné pro obrazy s pouze jedním typem vzorků. V tomto případě je vhodné spíše vypočíst celkovou vzdálenost, za kterou se vzorky podařilo shlukovat. Vzhledem k tomu, že ve vstupních datech jsou přítomny obrazy s jedním typem vzorků, je možné určité zkreslení očekávat. Těchto obrazů je však menší část a tedy toto zkreslení není příliš podstatné.

6.3 Výsledky statistické analýza

V této kapitole budou představeny vybrané nulové hypotézy, použité statistické testy a především jejich výsledky.

6.3.1 Srovnání účinnosti obou metod příznaků

Jelikož jsou výsledky klasifikace pro oba dva typy příznaků, uvedené v kapitole 5.4, velice podobné, je nutné tyto metody nejdříve navzájem porovnat.

Pro statistickou analýzu je použit list *Všechny shluky* z výše zmíněného datasetu.

Je tedy použita nulová hypotéza H_0, která tvrdí, že obě metody dosahují stejných výsledků. Jestliže je tato hypotéza potvrzena, je možné říci, že u představeného algoritmu nezáleží na výběru příznaků.

Toto tvrzení je nejdříve otestováno Wilcoxovým testem pro párové hodnoty. Tento test určuje rozdíly mezi párovými hodnotami obou metod, těmto rozdílům přiřazuje pořadí dle jejich vzestupné velikosti zvlášť pro kladné rozdíly a zvlášť pro rozdíly záporné. Je vypočten součet pořadí kladných i záporných pořadí. Tyto dva součty jsou stejné za předpokladu, že oba testované soubory pocházejí ze stejného základu, neboli mají stejné rozložení hodnot. Menší ze součtů je označen jako *T*. Je vypočtena kritická hodnota testu *Z*, která má pro více než 25 vzorků přibližně normální rozdělení. [18]

$$Z = \frac{T - \frac{m(m+1)}{4}}{\sqrt{\frac{m(m+1)(m+2)}{24}}} \tag{6.1}$$

kde *m* je počet hodnot rozdílů.

Tabulka 3: Výsledky Wilcoxova testu pro srovnání účinnosti metod příznaků

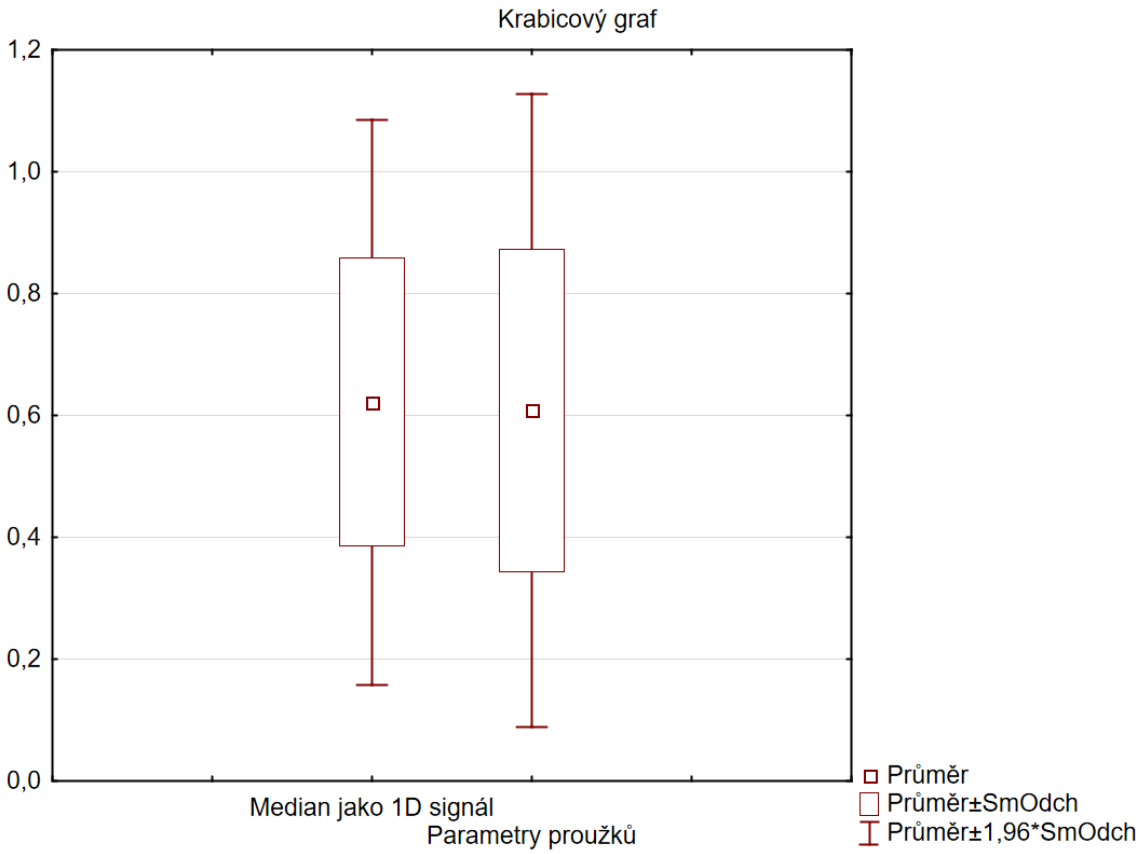
Dvojice proměnných	Wilcoxonův párový test (data) Označené testy jsou významné na hladině p <,05000			
	Počet (platných)	T	Z	p-hodn.
Median jako 1D signál & Parametry proužků	756	136763	1,050526	0,293477

Výsledné hodnoty jsou zobrazeny v **Error! Reference source not found.**. P-hodnota označuje statistickou významnost testu, která překročila limit 0,05, a tudíž zamítáme hypotézu H_0. Výsledky obou metod příznaků nejsou stejné.

V tomto případě je tedy nutné vybrat metodu, která vykazuje lepší výsledky. Pro to použijeme jejich popisné statistiky.

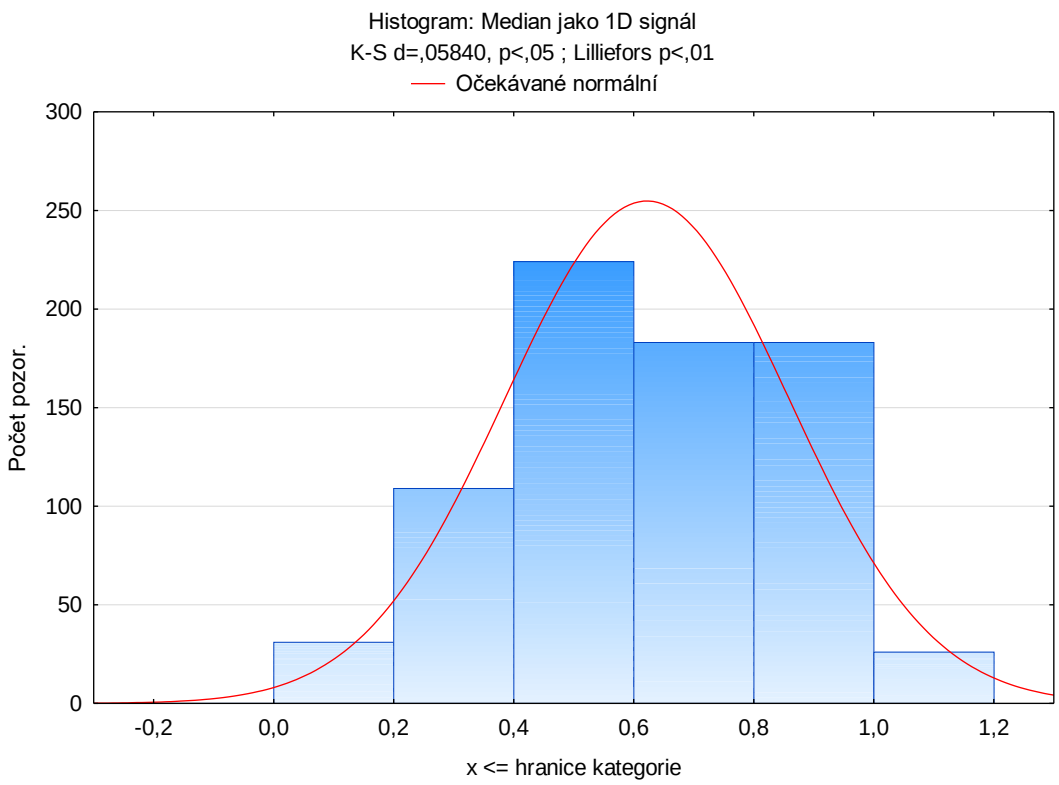
Tabulka 4: Hodnoty popisných statistik pro srovnání účinnosti metod příznaků

Proměnná	Popisné statistiky (data)				
	N platných	Průměr	Minimum	Maximum	Sm.odch.
Median jako 1D signál	756	0,622046	0,067038	1,124411	0,236717
Parametry proužků	756	0,608365	0,06796	1,120483	0,265098

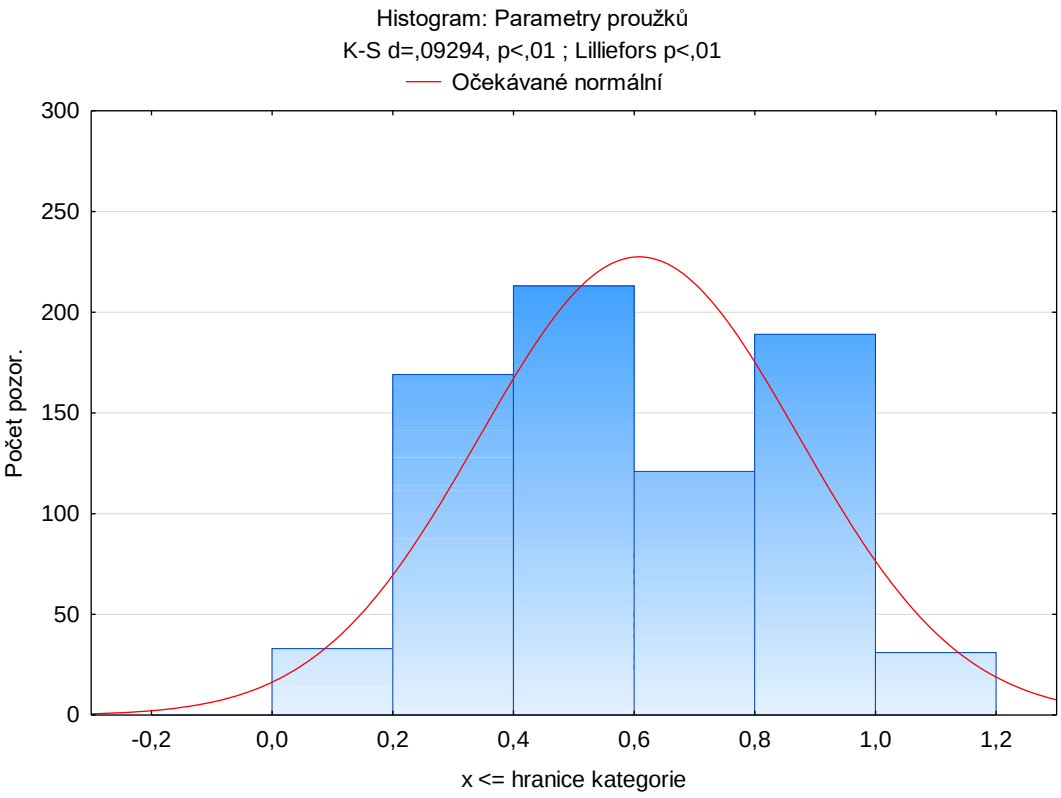


Obrázek 34: Krabicové grafy popisných statistik hodnot obou metod příznaků

Také je vhodné uvést příslušné histogramy těchto hodnot s testem normality rozložení. Test normality je proveden pomocí *Lillieforsova testu*, kde normálně rozložená data vykazují hodnoty větší než 0,05. [18]



Obrázek 35: Histogram hodnot metody příznaku Medián jako 1D signál



Obrázek 36: Histogram hodnot metody příznaku Medián jako 1D signál

Z uvedené tabulky (**Error! Reference source not found.**) a grafů (Obrázek 34) je patrné, že Medián jako 1D signál má větší průměrnou hodnotu, avšak s menší směrodatnou odchylkou (tedy menším rozptylem hodnot). Za to parametry proužků mají menší průměrnou hodnotu s větší směrodatnou odchylkou. Jelikož je výsledek *Lillieforsova testu* menší než 0,05, obě metody nemají normální rozdělení hodnot (Obrázek 35, Obrázek 36).

V tomto případě je na řadě *Test významnosti rozdílu dvou výběrových průměrů pro párové hodnoty* (tzv. t-*test*). Představíme tedy novou nulovou hypotézu H_0, podle které je rozdíl průměrů hodnot obou metod statisticky významný a tedy účinnost obou metod není stejná. [18]

Tento test pracuje s testovacím kritériem:

$$t = \frac{|\bar{d}|}{s_d}\sqrt{n-1} \tag{6.2}$$

kde n reprezentuje počet párů měření, d_i rozdíl jednotlivých měření,

$$\bar{d} = \frac{1}{n}\sum_i d_i, \tag{6.3}$$

a

$$s_d = \sqrt{\frac{1}{n}\sum_i (d_i - \bar{d})^2}. \tag{6.4}$$

Tabulka 5: Výsledné hodnoty Testu významnosti rozdílu dvou výběrových průměrů pro párové hodnoty

Proměnná					
	N	Rozdíl	Sm.odch. (rozdílu)	t	p
Median jako 1D signál					
Parametry proužků	756	0,013681	0,22323	1,68505	0,092392

Dle výsledků (Tabulka 5) je možné nulovou hypotézu zamítnout (hodnota p je větší než 0,05) a představit alternativní hypotézu H, která uvádí, že rozdíly průměrů jednotlivých metod jsou statisticky nevýznamné.

Pomocí statistické analýzy bylo zjištěno, že obě metody vykazují obdobné výsledky se statisticky zanedbatelnými rozdíly. Určit, která metoda příznaků je účinnější tedy nezávisí na výsledných hodnotách, ale spíše na osobní preferenci uživatele.

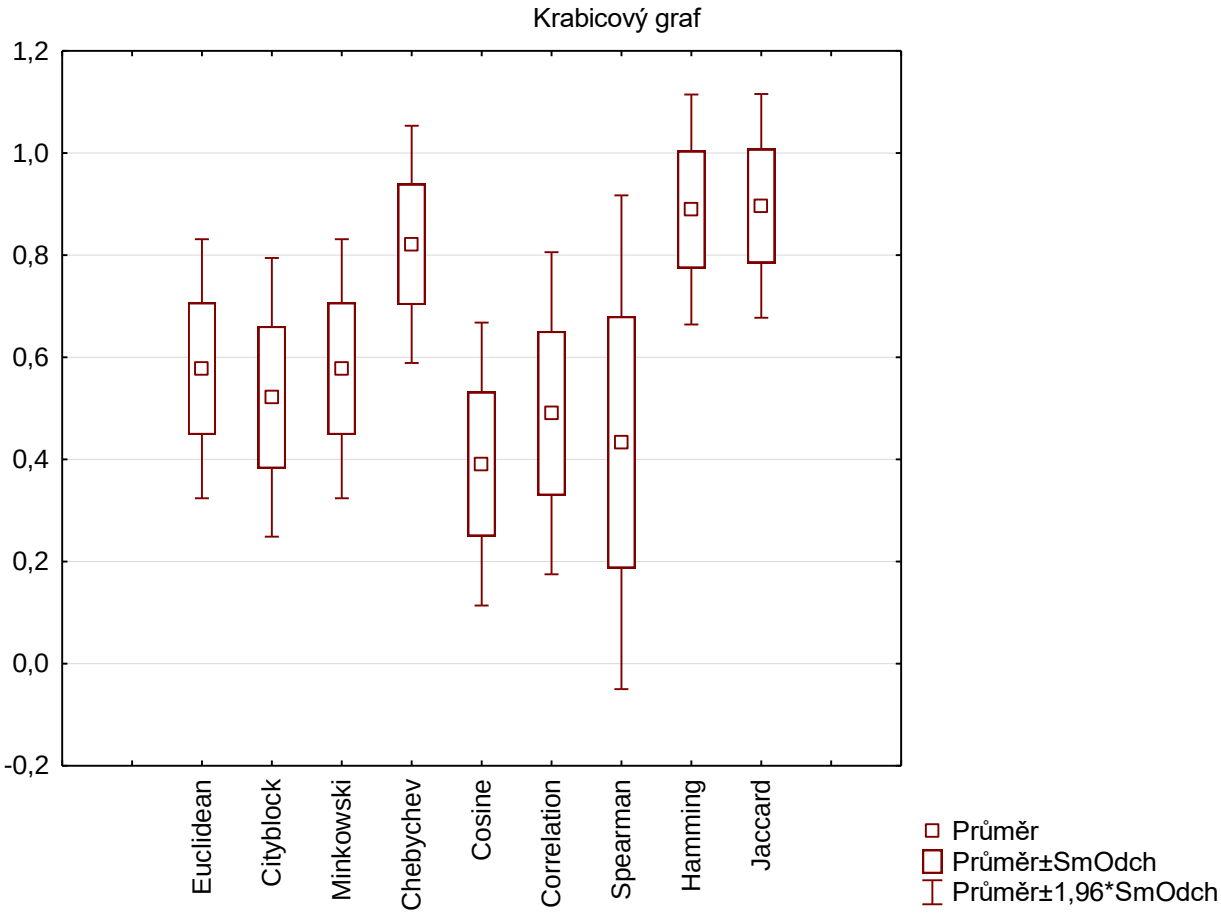
6.3.2 Kombinace metod výpočtu vzdáleností a shluků

Dále je nutné porovnat samotné metody výpočtu vzdáleností a shlukování. K těmto účelům byl opět použit soubor *data.xls*, konkrétně jeho listy *data_1_vzdalenosti*, *data_1_shluky* (pro *Medián linií jako 1D signál*), *data_2_vzdalenosti* a *data_2_shluky* (pro *Parametry proužků*).

Opět je využito popisných statistik dosažených hodnot.

Tabulka 6: Popisné statistiky jednotlivých metod vzdáleností pro Medián linií jako 1D signál

	Popisné statistiky (data)				
Proměnná	N platných	Průměr	Minimum	Maximum	Sm.odch.
Euclidean	84	0,577638	0,285942	0,825671	0,12943
Cityblock	84	0,521504	0,223332	0,829216	0,139271
Minkowski	84	0,577638	0,285942	0,825671	0,12943
Chebychev	84	0,821136	0,47988	1,124411	0,118472
Cosine	84	0,390858	0,176901	0,736816	0,141393
Correlation	84	0,49034	0,178271	0,807126	0,160898
Spearman	84	0,433612	0,067038	0,939434	0,246619
Hamming	84	0,889267	0,519518	1,058969	0,114933
Jaccard	84	0,896417	0,5234	1,075375	0,111792

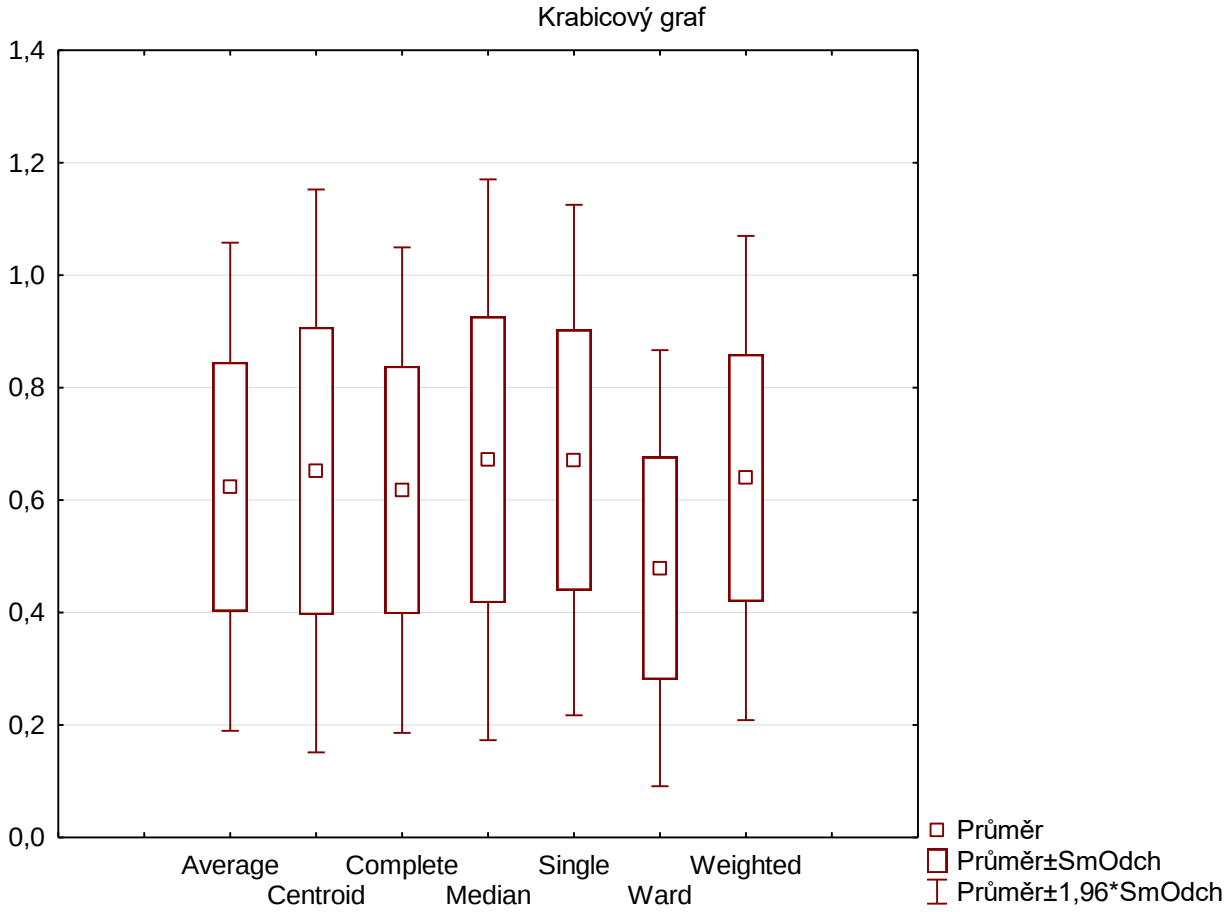


Obrázek 37: Krabicový graf hodnot jednotlivých metod vzdáleností pro Medián linií jako 1D signál

Z uvedených hodnot (Tabulka 6, Obrázek 37) je zřejmé, že nejlepší výsledky vykazuje metoda výpočtu *Cosine* s nejnižší průměrnou hodnotou a relativně nízkou směrodatnou odchylkou (a tedy i rozptylem hodnot). Metoda *Spearman* dosahuje u některých snímků lepších hodnot, nicméně má největší směrodatnou odchylku – právě tato hodnota může být zatížena zkreslením v souvislosti s obrazy s jedním typem příznaků (Kapitola 6.2). Následují metody *Euclidean, Cityblock, Correlation* a *Minkowski,* které jsou srovnatelné. Nejhorší výsledky vykazují metody *Chebychev, Hamming* a *Jaccard.*

Tabulka 7: Popisné statistiky jednotlivých metod shlukování pro Medián linií jako 1D signál

Proměnná	Popisné statistiky (data)				
	N platných	Průměr	Minimum	Maximum	Sm.odch.
Average	108	0,623669	0,099318	0,964433	0,221415
Centroid	108	0,651877	0,09676	1,104328	0,25541
Complete	108	0,617768	0,133151	0,966255	0,220243
Median	108	0,671691	0,099763	1,124411	0,254471
Single	108	0,671162	0,072098	0,991027	0,2316
Ward	108	0,478871	0,067038	0,849976	0,197887
Weighted	108	0,639281	0,102466	0,968414	0,219627

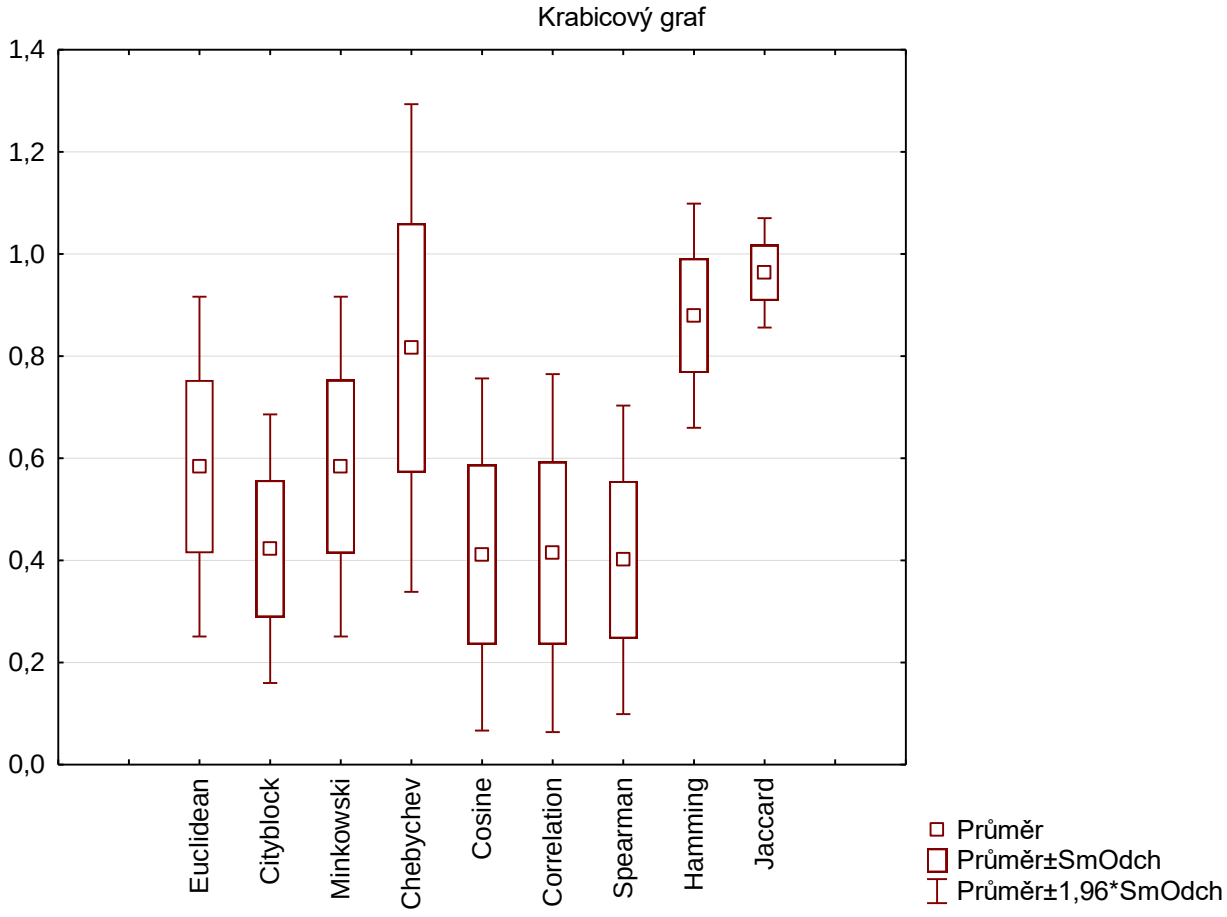


Obrázek 38: Krabicový graf hodnot jednotlivých metod shlukování pro Medián linií jako 1D signál

V tomto případě (Tabulka 7, Obrázek 38) jasně vede metoda shlukování *Ward*. Vykazuje výrazně nižší průměrné výsledky společně s nejnižší směrodatnou odchylkou. Ostatní metody jsou srovnatelné.

Tabulka 8: Popisné statistiky jednotlivých metod vzdáleností pro Parametry proužků

	Popisné statistiky (data)				
Proměnná	N platných	Průměr	Minimum	Maximum	Sm.odch.
Euclidean	84	0,583613	0,180649	0,971119	0,169723
Cityblock	84	0,422906	0,157272	0,86153	0,134221
Minkowski	84	0,583613	0,180649	0,971119	0,169723
Chebychev	84	0,815931	0,167492	1,120483	0,243637
Cosine	84	0,411468	0,071393	0,899696	0,175933
Correlation	84	0,414309	0,06796	0,939333	0,178822
Spearman	84	0,401148	0,088162	0,882776	0,154142
Hamming	84	0,879175	0,561197	1,098333	0,112047
Jaccard	84	0,963123	0,770919	1,04499	0,054654

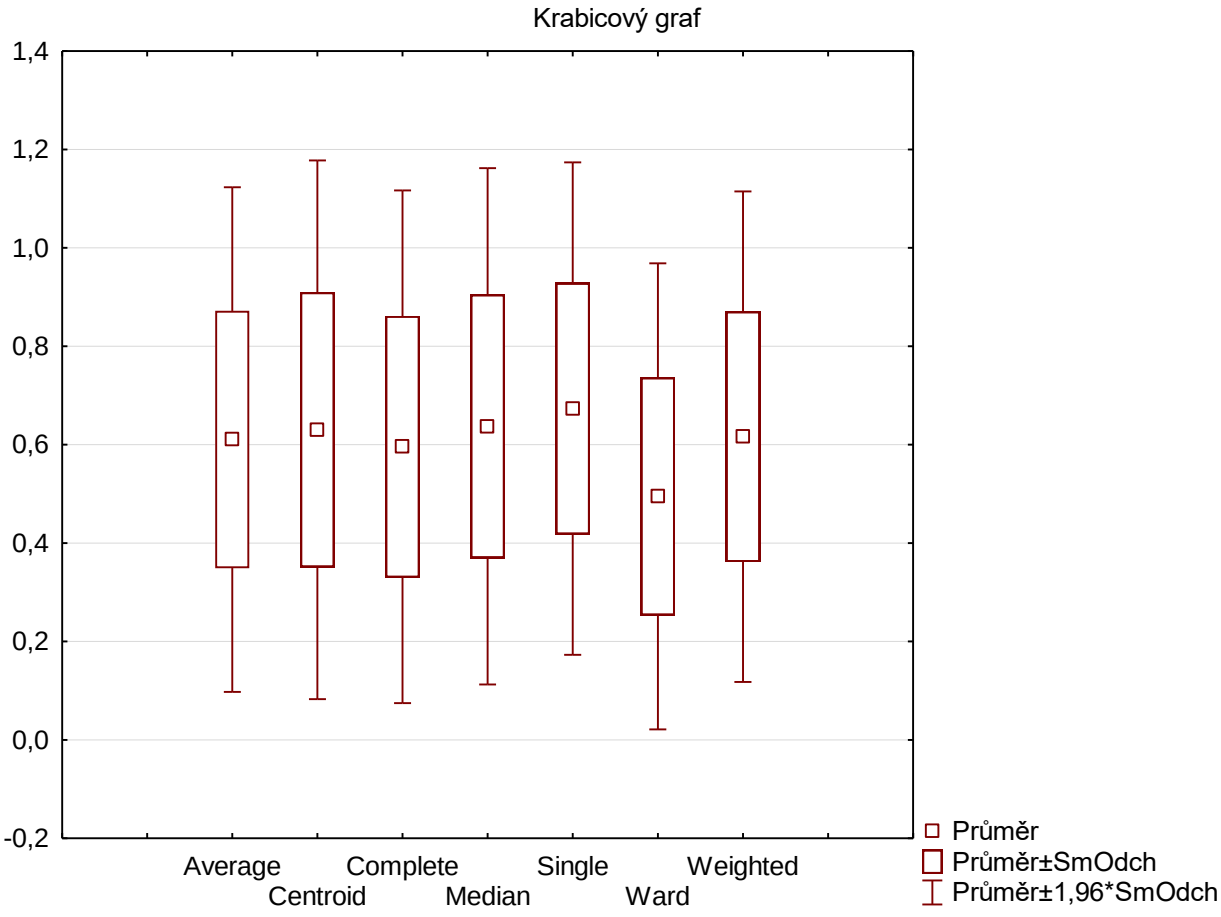


Obrázek 39: Krabicový graf hodnot jednotlivých metod vzdáleností pro Parametry proužků

U příznaku *Parametry* proužků (Tabulka 8, Obrázek 39) vykazuje metoda *Spearman* výrazně lepší výsledky (nižší průměr a především směrodatná odchylka) než v případě *Medián linií jako 1D signál*. Také průměrná hodnota *Correlation* a *Cityblock* je nižší. Tyto metody společně s *Cosine* vykazují obdobné výsledky. *Euclidean* a *Minkowski* dosahují průměrných hodnot. Mezi metody s nižší účinností se opět řadí *Chebychev*, *Hamming* a *Jaccard.*

Tabulka 9: Popisné statistiky jednotlivých metod shlukování pro Parametry proužků

	Popisné statistiky (data)				
Proměnná	N platných	Průměr	Minimum	Maximum	Sm.odch.
Average	108	0,610494	0,113233	0,999467	0,26164
Centroid	108	0,630404	0,113431	1,11781	0,279259
Complete	108	0,595764	0,100336	1	0,265779
Median	108	0,637328	0,107849	1,120483	0,267684
Single	108	0,673464	0,115145	1	0,255347
Ward	108	0,494894	0,06796	0,942108	0,241667
Weighted	108	0,616208	0,107618	0,999429	0,254364



Obrázek 40: Krabicový graf hodnot jednotlivých metod shlukování pro Parametry proužků

V tomto případě (Tabulka 9, Obrázek 40) vychází výsledky velmi podobně jako u metody Medián linií jako 1D signál. Opět platí, že metoda *Ward* vykazuje nejlepší výsledky a ostatní metody jsou srovnatelné.

7 Diskuze

Z uvedených obrazů (Kapitola 3) lze říci, že se podařilo přiblížit ideálním elektroforeogramům. Jejich kvalita závisí především na nastavení parametrů a na zkušenostech vykonavatele. Při vypracování se kvalita obrazů zvyšovala s každým dalším pokusem. Je nutné se však vyvarovat nahodilých chyb, které jakost řádově snižují (propíchnutí gelu pipetou, vytečení vzorků, zašpinění gelu apod.) a v některých případech znemožňují použití gelu pro další klasifikaci. Po určitém počtu provedení metody však není velký problém získat velice kvalitní elektroforeogram.

Z výsledků detekce hranic vzorků (Kapitola 4) je očividné, že při dostatečně dobrém vstupním obrazu není samotná detekce ničím složitým. Je také nutné správně nastavit parametry předzpracování obrazu. Toto předzpracování nekompenzuje jiné typy rušení, jakožto *smile effect*, rozmazání proužků atd., které se ovšem nevyskytuje u ideálních snímků. Při detekci linií je pro uvedený algoritmus vhodné uživatelsky zadat počet linií v testovaném gelu. Tento krok výrazně zlepšuje výsledky detekce. Také správný ořez snímku je pro tuto část důležitý. Pokud není ořez úplný a ve snímku zůstane okolí gelu, detekce nemusí vykazovat správné výsledky. Tudíž při nestandardním okolí gelu, kdy automatický ořez snímku selhává, je nutné použít ořez manuální.

Detekce proužků se odvíjí od správného nalezení hranic linií. Pro zlepšení výsledků detekce je obraz převeden na medián. Také nastavení prahu je důležité pro správnou detekci. Je nutné jej zvolit adekvátně pro zpracovaný snímek. Při příliš nízkém prahu algoritmus vykazuje falešně pozitivní detekce, naopak při příliš vysokém prahu můžeme očekávat falešně negativní detekce.

Pro klasifikaci vzorků (Kapitola 5) je žádoucí vybrat správnou metodu příznaků. První metoda (Medián linií jako 1D signál) využívá již získanou informaci, obsaženou v obrazu převedeném na medián (Kapitola 4.2.3). Druhá metoda (Parametry proužků) vyžaduje dodatečnou detekci šířky proužků, která může do měření přivést chyby – a to především u rozmazaných, či překrývajících se proužků. Při srovnání výsledků shlukování (Kapitola 5.4) však nedochází k velkým rozdílům mezi metodami příznaků.

V kapitole 6 byla představena vstupní data, která byla vytvořena testovací funkcí vykonávající výpočet pro všechny přiložené snímky. Bylo zvoleno kritérium účinnosti pro porovnání jednotlivých metod. Toto kritérium zohledňuje co největší rozdíly ve vzdálenostech shluků, a tedy nemusí být použitelné pro popis evolučních vzdáleností.

Následně byly jednotlivé metody porovnány pomocí statistické analýzy. Ukázalo se, že mezi výsledky metod příznaků je statisticky nevýznamný rozdíl, tedy že obě volby příznaků vykazují obdobné výsledky. Je však vhodné připomenout že výsledky metody *Parametry*

proužků závisí na správné detekci proužků a jejich šířky a proto může pro obrazy horší kvality vykazovat chybné výsledky. Obecně vzato, metoda *Medián linií jako 1D signál* vyniká svojí jednoduchostí a není zatížena chybou detekcí proužků a především jejich šířky.

Při porovnání výsledků jednotlivých metod výpočtu vzdáleností a shlukování bylo zjištěno, že se hodnoty metod výpočtu vzdáleností v rámci obou voleb příznaků od sebe liší. V obou případech se mezi nejlepší metody dostaly metody *Cosine* a *Spearman* i přes to, že dle specifikací jsou některé metody shlukování použitelné pouze pro metodu *Euklidovských vzdáleností*. Hodnoty výsledků metody *Spearman* mohly být zkresleny vstupními obrazy s jedním typem vzorků. U shlukování vykazuje nejlepší hodnoty metoda *Ward* shodně pro oba typy příznaků a pro všechny kombinace výpočtů vzdáleností. Ostatní metody shlukování dosahují srovnatelných, avšak o něco horších, výsledků. [15]

8 Závěr

Cílem diplomové práce bylo nastudovat a popsat problematiku zpracování obrazu 1D gelové elektroforézy. Byly popsány různé přístupy předzpracování obrazu, detekce hranic jednotlivých vzorků a jejich následná klasifikace týkající se shlukové analýzy.

Byla vytvořena databáze dvanácti referenčních snímků s ohledem na co nejvyšší obrazovou kvalitu. Z uvedených snímků je patrné, že obrázky se blíží prakticky k ideálnímu obrazu. Z průběhu vypracování je patrné, že kvalita obrazů následovala vzestupný trend. Při vypracování obrazů je důležité řídit se pokyny pro vypracování elektroforézy, nastavit správné parametry a vyvarovat se nahodilých chyb. Pokud jsou tyto faktory splněny, výsledný obraz by měl být velmi kvalitní.

Byl navrhnut algoritmus pro detekci hranic jednotlivých vzorků a následně byl realizován v prostředí MATLAB. Byly uvedeny výsledky detekce. Ty závisí především na kvalitě obrazu a jeho správného předzpracování. Je nutné správně nastavit parametry předzpracování a správný práh detekce proužků. V případě správného nastavení vykazuje detekce velmi dobré výsledky a to nejen u ideálních snímků.

Byly navrhnuty, realizovány a otestovány kombinace metod klasifikace čítající dvě metody příznaků, devět metod výpočtů vzdáleností a sedm metod shlukové analýzy. Obě metody příznaků jsou dostatečně objektivními ukazateli vlastností jednotlivých vzorků a dokáží tyto vzorky úspěšně shlukovat.

Účinnost metod byla otestována na vytvořených souborech. Takto vzniklý balík dat byl následně statisticky ohodnocen v programu *Statistica*. Bylo zjištěno, že obě metody příznaků vykazují podobné (nikoliv však identické) výsledky a jejich vzájemná rozdílnost je statisticky nevýznamná. Jako nejlepší metody výpočtu vzdáleností se ukázaly *Cosinová* a *Spearmanova vzdálenost*, avšak i *Korelační* a vzdálenost vykazuje podobné, ale o něco horší výsledky. Jako nejúčinnější varianta shlukové analýzy byla shledána *Wardova metoda*, která dosahuje nejlepších výsledků a to i v kombinaci jiných výpočtů vzdáleností, než *Euklidovských*.

Realizace algoritmů byla opatřena uživatelským prostředím, byla vytvořena testovací funkce *testovaci_fce.m* a také dataset změřených hodnot *data.xls* pro vstupní data, upraven pro následnou statistickou analýzu.

9 Reference

[1] SKUTKOVA, Helena, Martin VITEK, Sona KRIZKOVA, Rene KIZEK and Ivo PROVAZNIK. Preprocessing and Classification of Electrophoresis Gel Images Using Dynamic Time Warping. 2013, vol. 8, pp. 1609–1622.

[2] KLOUDA, Pavel. *Moderní analytické metody*. 2., upr. a dopl. vyd. Ostrava: Pavel Klouda, 2003, 132 s. ISBN 80-863-6907-2.

[3] *Bioanalytické metody*. 3., přeprac. vyd. Praha: Vysoká škola chemicko-technologická, 2001, 254 s. ISBN 80-708-0449-1.

[4] KÁŠ, Jan, Milan KODÍČEK a Olga VALENTOVÁ. *Laboratorní techniky biochemie*. 1. vyd. Praha: VŠCHT, 2005, 258 s. ISBN 80-708-0586-2.

[5] LEWIS, Matt. DEPARTMENT OF PATHOLOGY UNIVERSITY OF LIVERPOOL. *Agarose gel electrophoresis (basic method)* [online]. 2001 [cit. 2014-11-16]. Dostupné z: http://www.methodbook.net/dna/agarogel.html

[6] CHEN, Peter. COLLEGE OF DUPAGE. *Electrophoresis* [online]. [cit. 2014-11-16]. Dostupné z: http://bio1151.nicerweb.com/Locked/media/ch20/electrophoresis.html

[7] JAN, Jiří. *Číslicová filtrace, analýza a restaurace signálů*. 2. upr. a rozš. vyd. Brno: VUTIUM, 2002, 427 s. ISBN 80-214-2911-9.

[8] WALEK, Petr, Martin LAMOŠ a Jan JIŘÍ. *Analýza biologických obrazů: Počítačová cvičení* [online]. první. Brno: Vysoké učení technické v Brně, 2013 [cit. 2014-11-17]. ISBN 978-80-214-4792-9. Dostupné z: http://www.dbme.feec.vutbr.cz/sites/default/files/news/fabo.pdf

[9] ED. BY: EDWARD R. DOUGHERTY, Ed. by: Edward R.Ilya Shmulevich. *Genomic signal processing and statistics*. New York, NY [u.a.]: Hindawi Publ. Corporation, 2005. ISBN 97-759-4507-0.

[10] AUSUBEL, Frederick M. *Current protocols in molecular biology*. Media, Pa.: J. Wiley, order fulfillment, c1987-, 2 v. (loose-leaf). ISBN 97804715033782-.

[11] PCR 100 bp Low Ladder. *Sigma-Aldrich Co.* [online]. 2014 [cit. 2014-12-20]. Dostupné z: http://www.sigmaaldrich.com/catalog/product/sigma/p1473?lang=en&region=CZ

[12] MOUNT, David W. *Bioinformatics: sequence and genome analysis*. 2nd ed. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press, c2004, xii, 692 p. ISBN 08-796-9712-1.

[13] ŽÁK, L. Shluková analýza I. Automatizace. 2004, roč. 47, č. 3, s. 3. Dostupné z: http://www.volny.cz/elzet/Libor/Aut_cl_1.pdf

[14] KOZUMPLÍK, Jiří. *Umělá inteligence v medicíně: Shluková analýza* [online]. Brno, 2014 [cit. 2014-12-30]. Přednáška. Vysoké Učení Technické v Brně.

[15] ROMESBURG, H. Charles. *Cluster analysis for researchers*. [Repr.]. Lulu Pr: Lulu Press, 2004. ISBN 978-141-1606-173.

[16] WILEY, E a Bruce S LIEBERMAN. *Phylogenetics: theory and practice of phylogenetics systematics*. Hoboken, N.J.: Wiley-Blackwell, c2011, xvi, 406 p. ISBN 978-047-0905-968.

[17] NEI, Masatoshi a Sudhir KUMAR. *Molecular evolution and phylogenetics*. New York: Oxford University Press, 2000, xiv, 333 p. ISBN 01-951-3585-7.

[18] BAŠTINEC, Jaromír. VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ. *Statistika, stochastické procesy, operační výzkum* [Skripta předmětu DMA1]. Brno, 2014 [cit. 2015-05-16].

[19] THE MATHWORKS, INC. *Statistics and Machine Learning Toolbox* [online]. 2015a. 2015 [cit. 2015-05-16]. Dostupné z: http://www.mathworks.com/help/stats/index.html

10 Přílohy

Laboratorní protokol

Pracující: Krupka Dvořáček	ID měření: 2	Provedeno: Elektroforéza 2
Datum: 3.10.2014	Měření úspěšné (ano/ne/částečně): částečně	

Typ, ID , datum výroby pufru a jeho koncentrace:	TBE, 1, 3.10.2014, 1x konc
Množství použitého pufru [ml]:	100 + [250 +250 (zalití vany)]

Typ barviva:	GelRed
Množství barviva [µl]	100

Celkový objem gelu [ml]:	100
Teplota gelu při nalévání [°C]:	Cca 55
Doba tuhnutí gelu před pipetováním vzorků[min]:	30

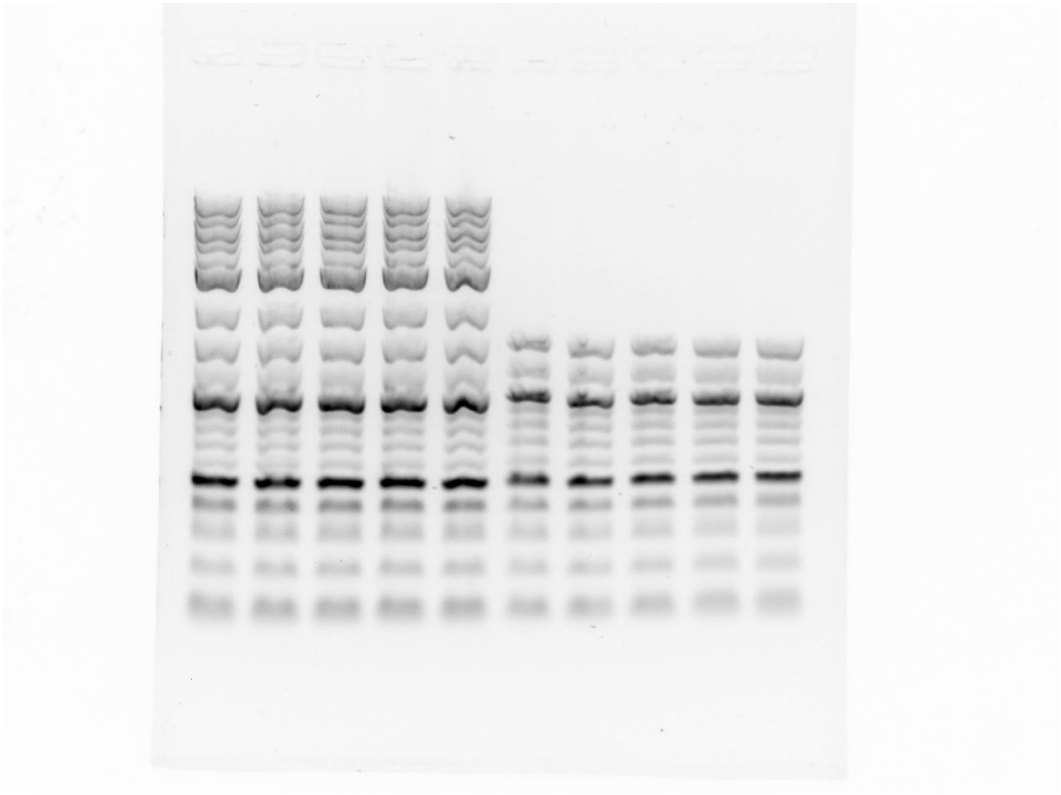
Napipetované vzorky:								
1	2	3	4	5	6	7	8	9
log	log	log	log	log	100	100	100	100x2

Nastavená prvotní délka elektroforézy [min]:	70
Dodatečná délka elektroforézy [min]:	-
Nastavené napětí [V]:	90

Neúmyslné chyby měření:
Nasýpací vanička spadla do Erl. Baňky s gelem, propíchnutí jamek v gelu (označeny červeně)

Úmyslné chyby měření:
Názvy výstupních souborů:
Elfo_ID_2_b

Obrázek 41: Protokol měření 02



Obrázek 42: elfo_id_2_b

Laboratorní protokol

Pracující: Krupka Dvořáček	ID měření: 3	Provedeno: Elektroforéza 1x
Datum: 6.10.2014	Měření úspěšné (ano/ne/částečně): částečně	

Typ, ID a datum výroby pufru:	TBE, 1, 3.10.2014, 1x konc
Množství použitého pufru [ml]:	50 + 250

Typ barviva a ředění:	GelRed, 10x zředěný
Množství barviva [µl]	50

Celkový objem gelu [ml]:	50
Teplota gelu při nalévání [°C]:	61
Doba tuhnutí gelu před pipetováním vzorků:	30

Napipetované vzorky:									
1	2	3	4	5	6	7	8	9	10
Log	Log	Log	Log	Log	100	100	100	100	100

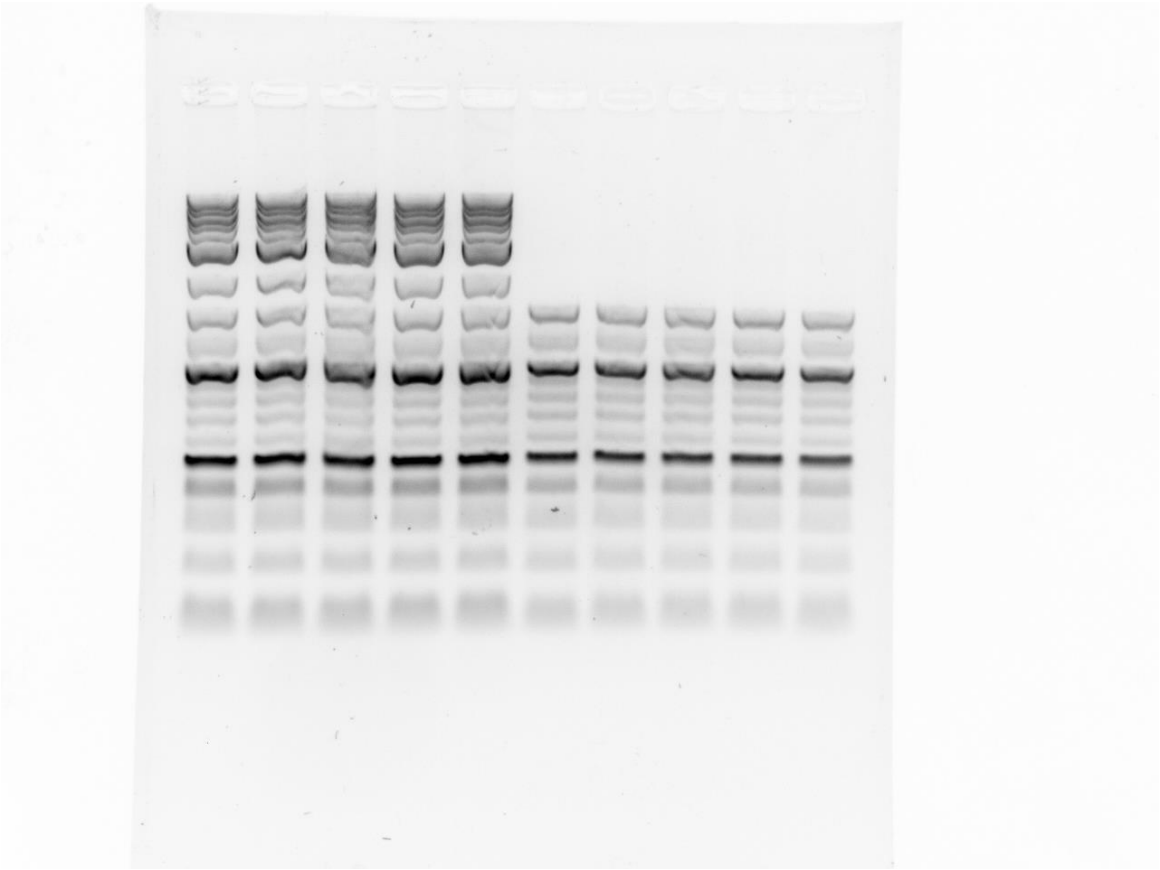
Nastavená prvotní délka elektroforézy:	90
Dodatečná délka elektroforézy:	-
Nastavené napětí:	70

Neúmyslné chyby měření:
Opětovné artefakty větších fragmentů.

Úmyslné chyby měření:

Názvy výstupních souborů:
elfo_id_3_a

Obrázek 43: Protokol měření 03



Obrázek 44: elfo_id_3_a

Laboratorní protokol

Pracující: Krupka Dvořáček	ID měření: 4	Provedeno: Elektroforéza 1x
Datum: 13.10.2014	Měření úspěšné (ano/ne/částečně): částečně	

Typ, ID a datum výroby pufru:	TBE, 1, 3.10.2014, 1x konc
Množství použitého pufru [ml]:	100+250

Typ barviva a ředění:	GelRed, 10x zředěný
Množství barviva [µl]	50

Celkový objem gelu [ml]:	100
Teplota gelu při nalévání [°C]:	60
Doba tuhnutí gelu před pipetováním vzorků:	30

Napipetované vzorky:									
1	2	3	4	5	6	7	8	9	10
1kb	1kb	1kb	1kb	1kb	Log	Log	Log	100	100

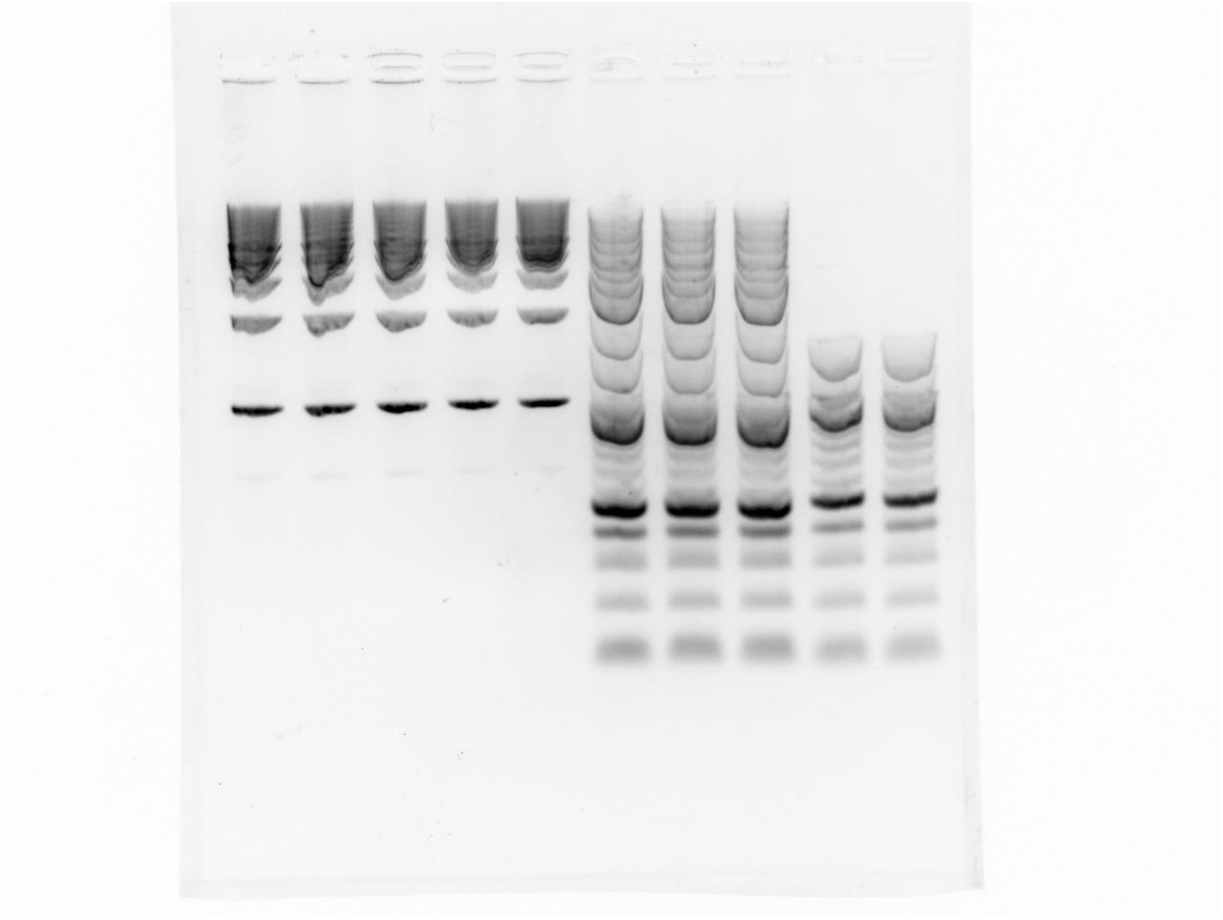
Nastavená prvotní délka elektroforézy:	70
Dodatečná délka elektroforézy:	-
Nastavené napětí:	90

Neúmyslné chyby měření:
Opětovné artefakty větších fragmentů. Jeden z gelů se rozbil před pipetováním. Přílišná koncentrace vzorků 1kb. (??)

Úmyslné chyby měření:

Názvy výstupních souborů:
elfo_id_4_a

Obrázek 45: Protokol měření 04



Obrázek 46: elfo_id_4_a

Laboratorní protokol

Pracující: Krupka Dvořáček	ID měření: 5	Provedeno: Elektroforéza 1x
Datum: 13.10.2014	Měření úspěšné (ano/ne/částečně): částečně	

Typ, ID a datum výroby pufru:	TBE, 1, 13.10.2014, 1x konc
Množství použitého pufru [ml]:	50+250

Typ barviva a ředění:	GelRed, 10x zředěný
Množství barviva [µl]	50

Celkový objem gelu [ml]:	50, 80% gel
Teplota gelu při nalévání [°C]:	60
Doba tuhnutí gelu před pipetováním vzorků:	30

Napipetované vzorky:									
1	2	3	4	5	6	7	8	9	10
2log	2log	2log	2log	2log	100bp	100bp	100bp	100bp	100bp

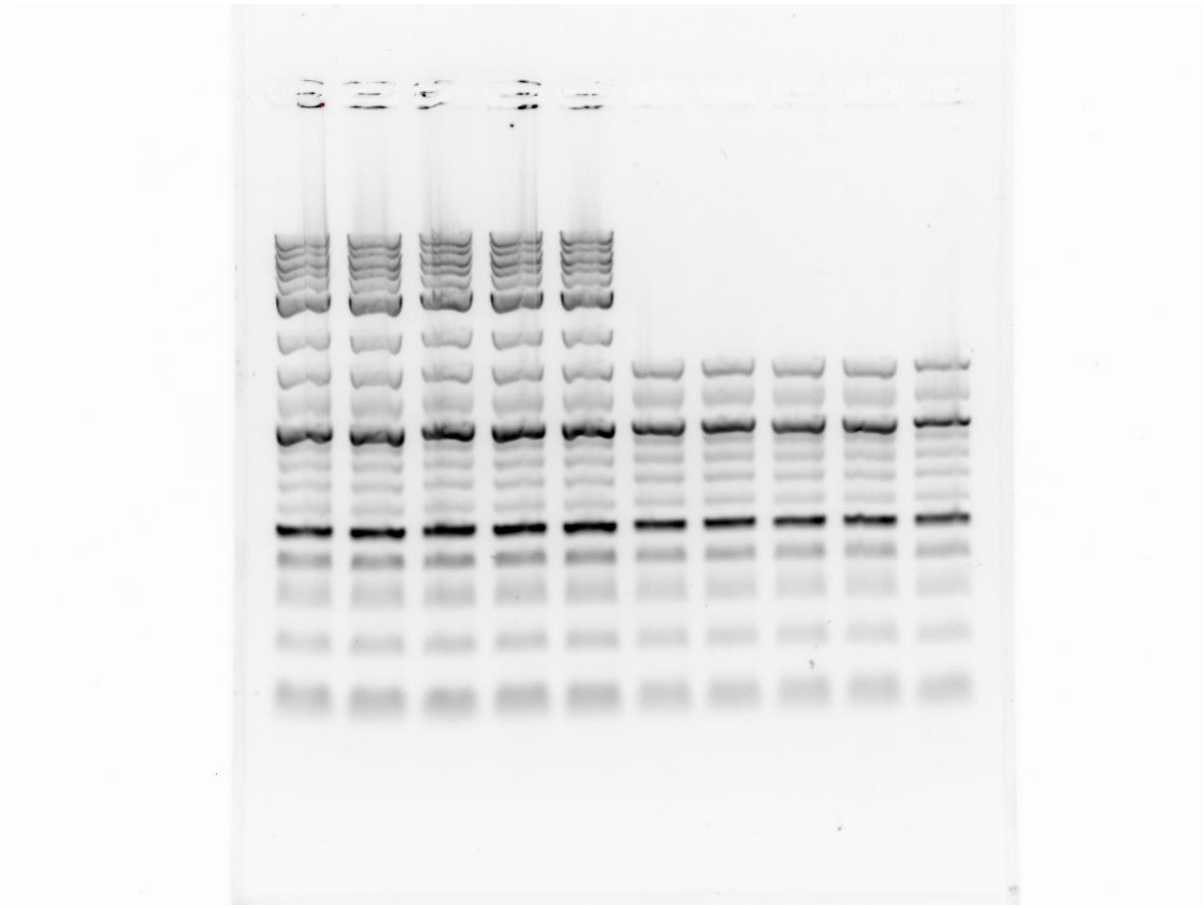
Nastavená prvotní délka elektroforézy:	70
Dodatečná délka elektroforézy:	-
Nastavené napětí:	90

Neúmyslné chyby měření:

Úmyslné chyby měření:

Názvy výstupních souborů:
elfo_id_5_a

Obrázek 47: Protokol měření 05



Obrázek 48: elfo_id_5_a

Laboratorní protokol

Pracující: Krupka Dvořáček	ID měření: 7a	Provedeno: Elektroforéza 1x
Datum: 20.10.2014	Měření úspěšné (ano/ne/částečně):	

Typ, ID a datum výroby pufru:	TBE, 1, 13.10.2014, 1x konc
Množství použitého pufru [ml]:	50+250

Typ barviva a ředění:	GelRed, 10x zředěný
Množství barviva [µl]	50

Celkový objem gelu [ml]:	50
Teplota gelu při nalévání [°C]:	60
Doba tuhnutí gelu před pipetováním vzorků:	30

Napipetované vzorky:									
1	2	3	4	5	6	7	8	9	10
2log	2log	100bp	100bp	2log	2log	100bp	100bp	2log	2log

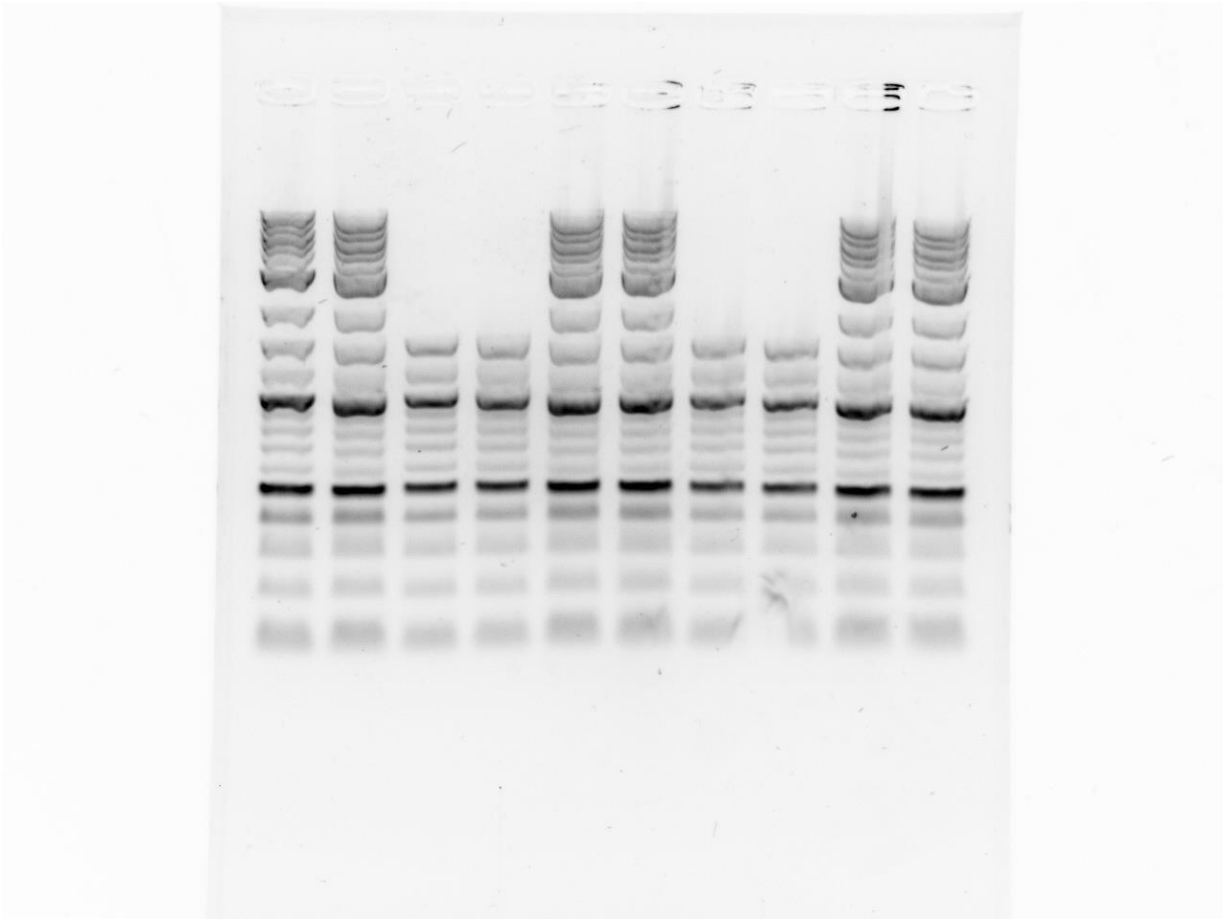
Nastavená prvotní délka elektroforézy:	70
Dodatečná délka elektroforézy:	
Nastavené napětí:	90

Neúmyslné chyby měření:

Úmyslné chyby měření:

Názvy výstupních souborů:
elfo_id_7_a

Obrázek 49: Protokol měření 07



Obrázek 50: elfo_id_7_a

Laboratorní protokol

Pracující: Krupka Dvořáček	ID měření: 8b	Provedeno: Elektroforéza 1x
Datum: 20. 10. 2014	Měření úspěšné (ano/ne/částečně):	

Typ, ID a datum výroby pufru:	TBE, 1, 13. 10. 2014, 1x konc
Množství použitého pufru [ml]:	50+250

Typ barviva a ředění:	GelRed, 10x zředěný
Množství barviva [µl]	50

Celkový objem gelu [ml]:	50
Teplota gelu při nalévání [°C]:	60
Doba tuhnutí gelu před pipetováním vzorků:	30

Napipetované vzorky:									
1	2	3	4	5	6	7	8	9	10
100bp	100bp	100bp	100bp	100bp	100bp	100bp	100bp	100bp	x

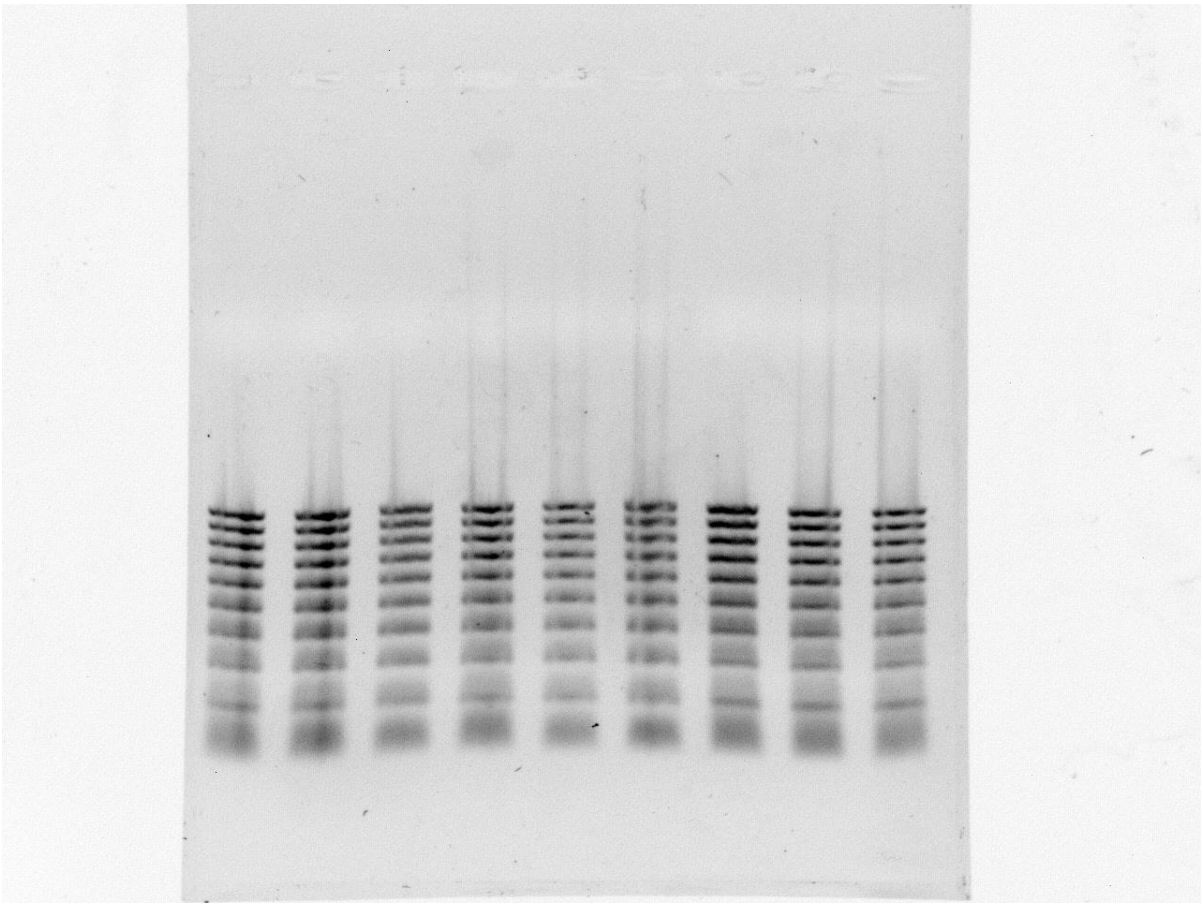
Nastavená prvotní délka elektroforézy:	20 +15 + 5 (o napětí 200V)
Dodatečná délka elektroforézy:	
Nastavené napětí:	150

Neúmyslné chyby měření:

Úmyslné chyby měření:
Pokus o smile effect v důsledku zvýšeného napětí a následného zahřívání gelu procházejícím proudem.

Názvy výstupních souborů:
elfo_id_8_b

Obrázek 51: Protokol měření 8b



Obrázek 52: elfo_id_8_b

Laboratorní protokol

Pracující: Krupka Dvořáček	ID měření: 11a	Provedeno: Elektroforéza
Datum: 27. 10. 2014	Měření úspěšné (ano/ne/částečně): ano	

Typ, ID a datum výroby pufru:	TBE, 1, 20 a 22.10. 2014, 1x konc
Množství použitého pufru [ml]:	50+250

Typ barviva a ředění:	GelRed, 10x zředěný
Množství barviva [µl]	25

Celkový objem gelu [ml]:	50
Teplota gelu při nalévání [°C]:	60
Doba tuhnutí gelu před pipetováním vzorků:	30

Napipetované vzorky: pouze 100bp!!									
1	2	3	4	5	6	7	8	9	
Mobio	Mobio	Mobio	Sigma	Sigma	Sigma	Biolabs	Biolabs	Biolabs	

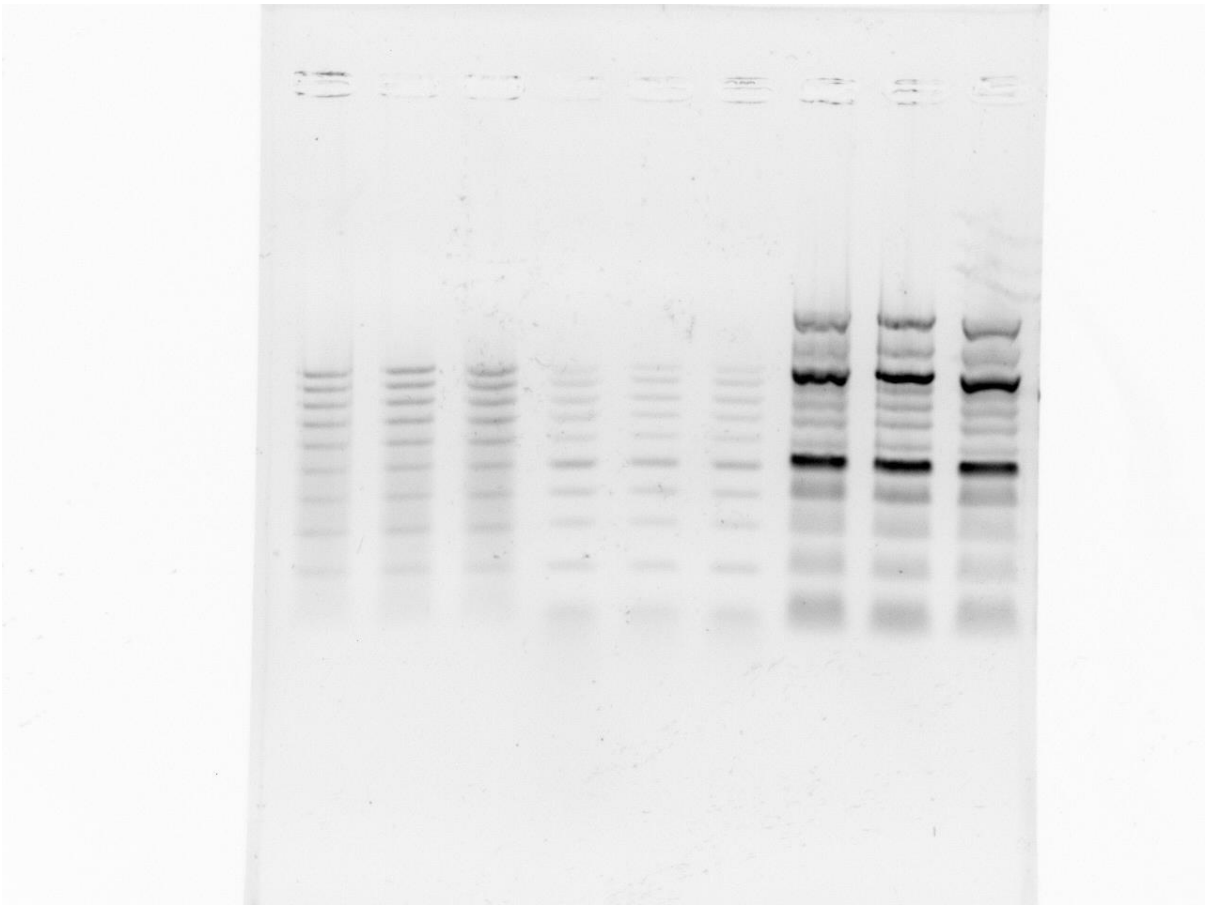
Nastavená prvotní délka elektroforézy:	90
Dodatečná délka elektroforézy:	
Nastavené napětí:	70

Neúmyslné chyby měření:

Úmyslné chyby měření:

Názvy výstupních souborů:
elfo_id_11_a

Obrázek 53: Protokol měření 11a



Obrázek 54: elfo_id_11_a

Laboratorní protokol

Pracující: Krupka Dvořáček	ID měření: 11b	Provedeno: Elektroforéza
Datum: 27. 10. 2014	Měření úspěšné (ano/ne/částečně):	

Typ, ID a datum výroby pufru:	TBE, 1, 20 a 22.10. 2014, 1x konc
Množství použitého pufru [ml]:	50+250

Typ barviva a ředění:	GelRed, 10x zředěný
Množství barviva [µl]	25

Celkový objem gelu [ml]:	50
Teplota gelu při nalévání [°C]:	60
Doba tuhnutí gelu před pipetováním vzorků:	25

Napipetované vzorky:									
1	2	3	4	5	6	7	8	9	10
100bp	100bp	100bp	100bp	100bp	100bp	100bp	100bp	100bp	100bp

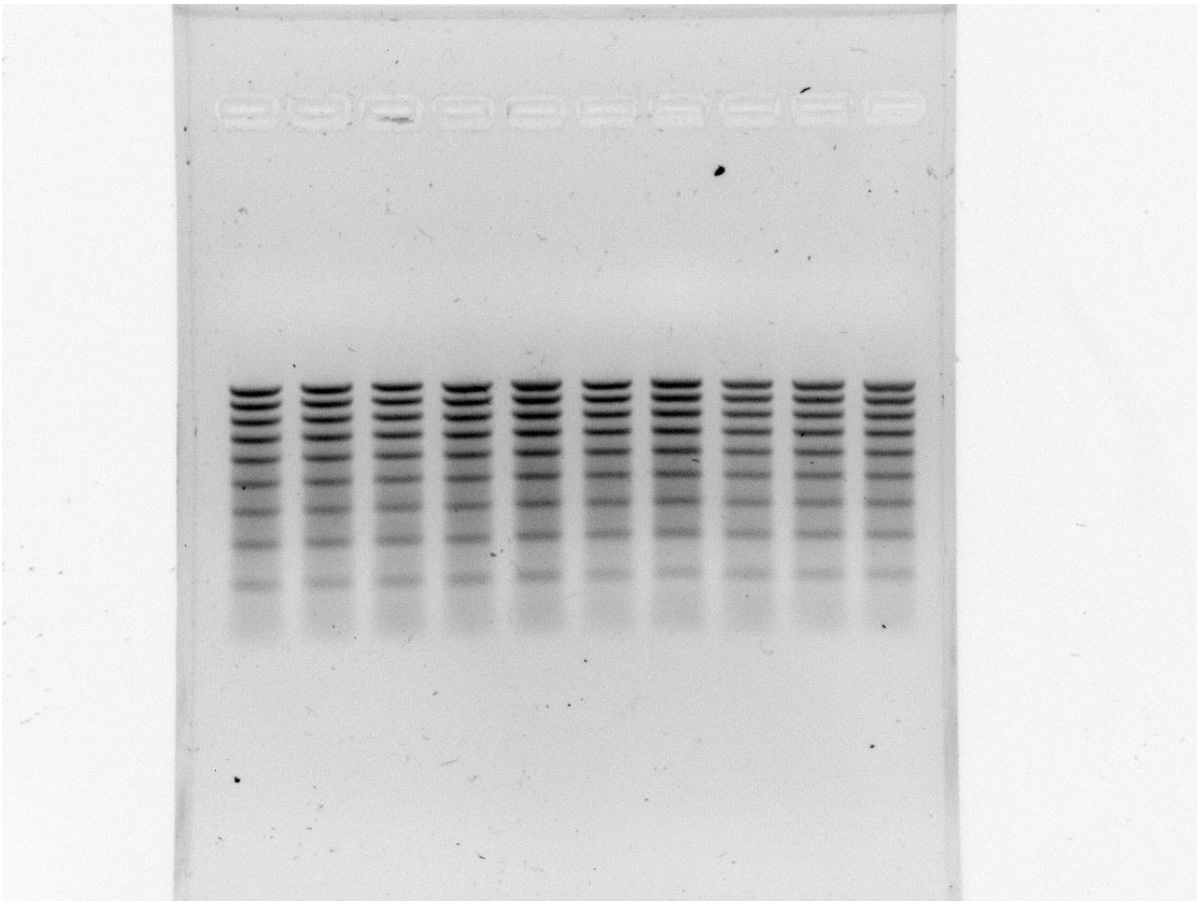
Nastavená prvotní délka elektroforézy:	90
Dodatečná délka elektroforézy:	
Nastavené napětí:	70

Neúmyslné chyby měření:

Úmyslné chyby měření:

Názvy výstupních souborů:
elfo_id_11_b

Obrázek 55: Protokol měření 11b



Obrázek 56: elfo_id_11_b

Laboratorní protokol

Pracující: Krupka Dvořáček	ID měření: 12a	Provedeno: Elektroforéza
Datum: 30. 10. 2014	Měření úspěšné (ano/ne/částečně):	

Typ, ID a datum výroby pufru:	TBE, 1, 20 a 22.10. 2014, 1x konc
Množství použitého pufru [ml]:	50+250

Typ barviva a ředění:	GelRed, 10x zředěný
Množství barviva [µl]	25

Celkový objem gelu [ml]:	50
Teplota gelu při nalévání [°C]:	60
Doba tuhnutí gelu před pipetováním vzorků:	30

Napipetované vzorky: pouze 100bp!!									
1	2	3	4	5	6	7	8	9	
100	100	100	100	100	100	100	100	100	100

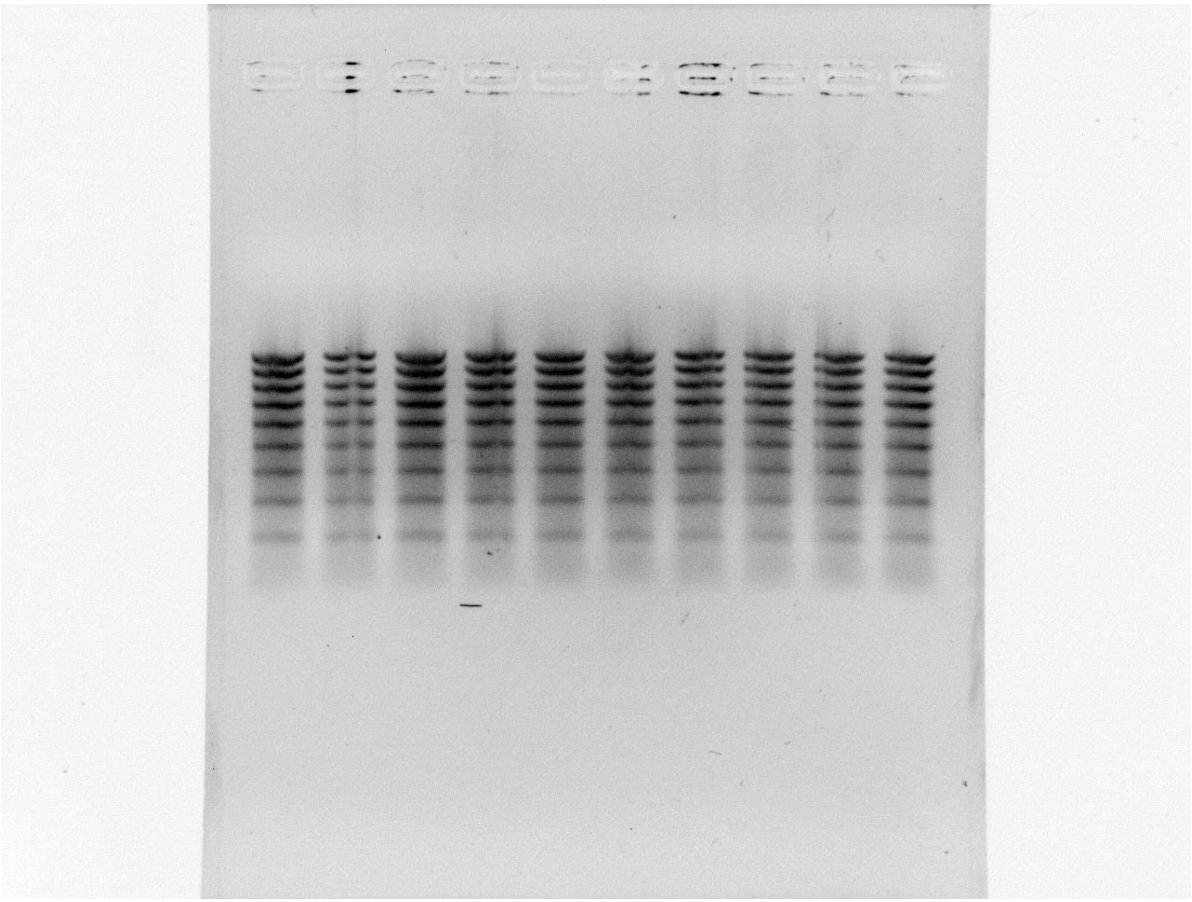
Nastavená prvotní délka elektroforézy:	90
Dodatečná délka elektroforézy:	
Nastavené napětí:	70

Neúmyslné chyby měření:

Úmyslné chyby měření:
Vložený hřebínek byl zleva doprava nadzdvihnut o úhel 3°

Názvy výstupních souborů:
elfo_id_12_a

Obrázek 57: Protokol měření 12a



Obrázek 58: elfo_id_12_a

Laboratorní protokol

Pracující: Krupka Dvořáček	ID měření: 13a	Provedeno: Elektroforéza
Datum: 31. 10. 2014	Měření úspěšné (ano/ne/částečně):	

Typ, ID a datum výroby pufru:	TBE, 1, 20 a 22.10. 2014, 1x konc
Množství použitého pufru [ml]:	50+250 + 0,4g agarozy

Typ barviva a ředění:	GelRed, 10x zředěný
Množství barviva [µl]	50

Celkový objem gelu [ml]:	50
Teplota gelu při nalévání [°C]:	60
Doba tuhnutí gelu před pipetováním vzorků:	30

Napipetované vzorky: pouze 100bp!!									
1	2	3	4	5	6	7	8	9	
100	100	100	100	100	100	100	100	100	

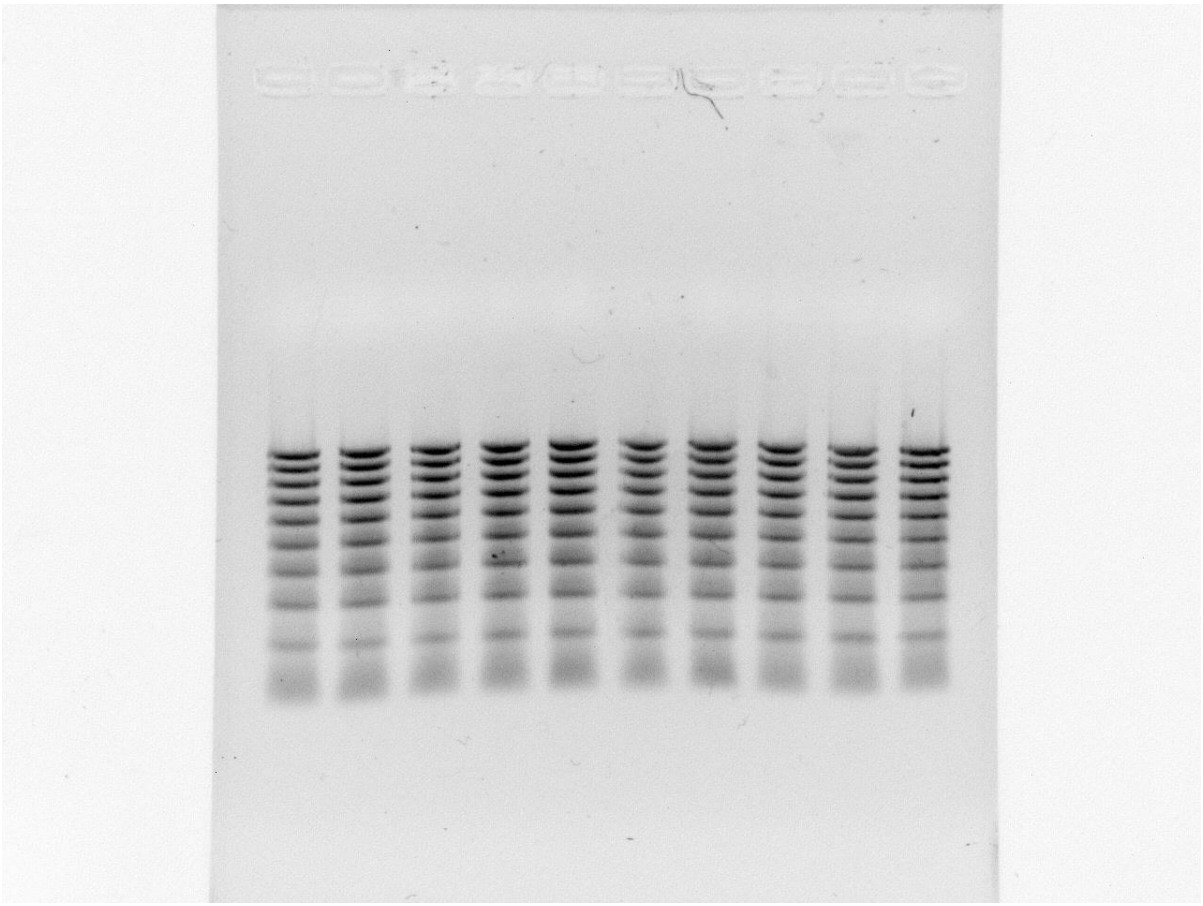
Nastavená prvotní délka elektroforézy:	45
Dodatečná délka elektroforézy:	
Nastavené napětí:	130

Neúmyslné chyby měření:

Úmyslné chyby měření:
Zvýšené napětí pro vytvoření smile efektu (pokus s 0.8 % gelem)

Názvy výstupních souborů:
elfo_id_13_a

Obrázek 59: Protokol měření 13a



Obrázek 60: elfo_id_13_a

Laboratorní protokol

Pracující: Krupka Dvořáček	ID měření: 14b	Provedeno: Elektroforéza
Datum: 7. 11. 2014	Měření úspěšné (ano/ne/částečně):	

Typ, ID a datum výroby pufru:	TBE, 1, 20 a 22.10. 2014, 1x konc
Množství použitého pufru [ml]:	50+250

Typ barviva a ředění:	GelRed, 10x zředěný
Množství barviva [µl]	50

Koncentrace gelu	2%
Celkový objem gelu [ml]:	50
Teplota gelu při nalévání [°C]:	60
Doba tuhnutí gelu před pipetováním vzorků:	30

100bp ladder									
1	2	3	4	5	6	7	8	9	10
Sigma	Sigma	Sigma	Sigma	Sigma	BioLabs	BioLabs	BioLabs	BioLabs	BioLabs

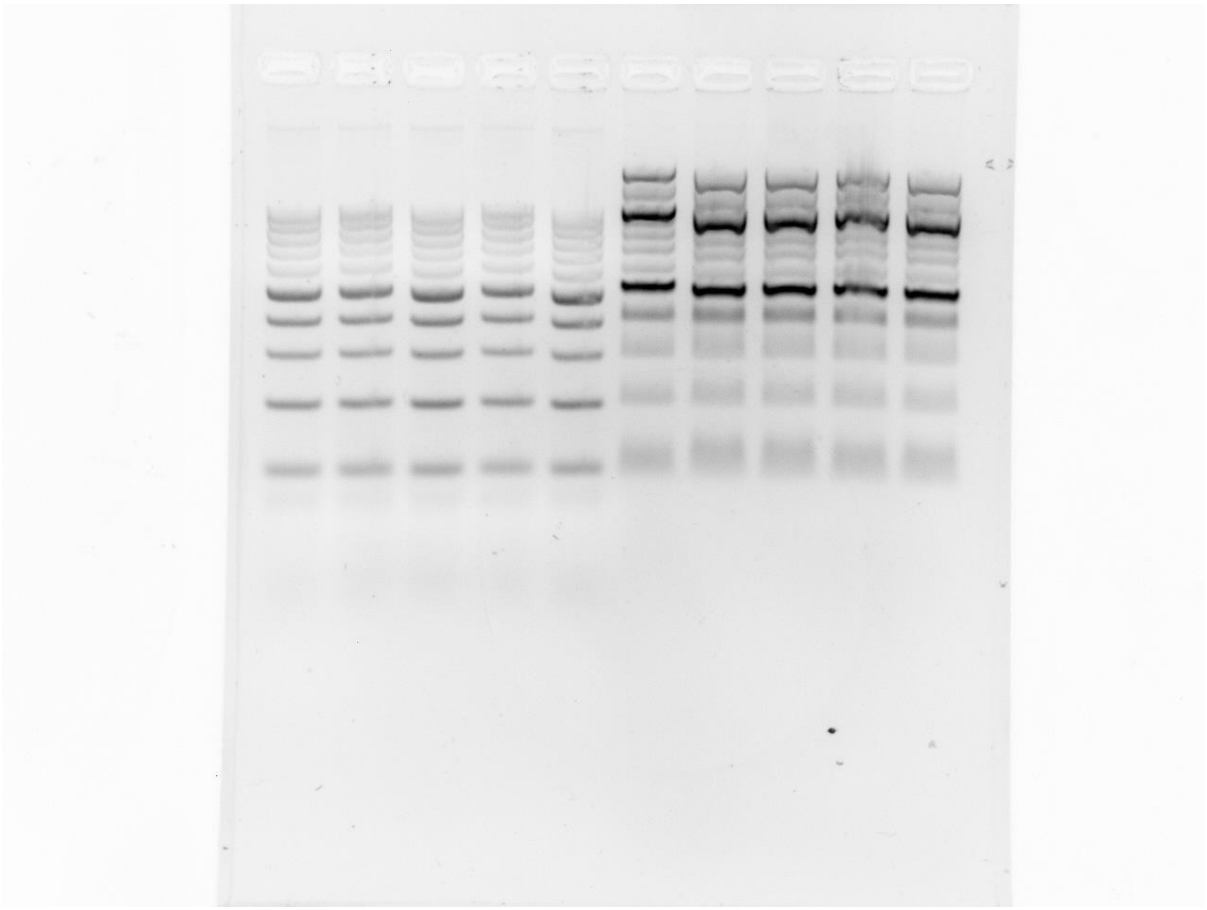
Nastavená prvotní délka elektroforézy:	70
Dodatečná délka elektroforézy:	
Nastavené napětí:	90

Neúmyslné chyby měření:

Úmyslné chyby měření:

Názvy výstupních souborů:
elfo_id_14_b

Obrázek 61: Protokol měření 14b



Obrázek 62: elfo_id_14_b