

Česká zemědělská univerzita v Praze

Fakulta lesnická a dřevařská

Katedra genetiky a fyziologie lesních dřevin



Diplomová práce

**Metody sekvenování nové generace (NGS) u smrku
ztepilého – Modifikace analýzy mikrosatelitových
amplikonů**

Autorka: Bc. Andrea Moravcová

Vedoucí práce: Ing. Jiří Korecký, Ph.D.

© 2022 ČZU v Praze

ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA V PRAZE

Fakulta lesnická a dřevařská

ZADÁNÍ DIPLOMOVÉ PRÁCE

Bc. Andrea Moravcová

Lesní inženýrství

Lesní inženýrství

Název práce

Metody sekvenování nové generace (NGS) u smrku ztepilého – Modifikace analýzy mikrosatelitových amplikonů

Název anglicky

New generation sequencing methods (NGS) for Norway spruce – Modification of microsatellite amplicon analysis

Cíle práce

Cílem práce je vytvořit funkční metodu analýzy mikrosatelitových amplikonů pro smrk ztepilý na bázi NGS sekvenování. Protokol využívá mikrosatelitové markery s inovativním přístupem k analýze PCR amplikonů. Běžně používaná detekce a separace amplikonů fragmentační analýzou v kapiláře je nahrazena sekvenací amplikonů pokročilými sekvenačními metodami.

Funkčnost vytvořeného protokolu bude ověřena na testovacím souboru DNA vzorků smrku ztepilého.

Metodika

Součástí diplomové práce bude rešeršní část, kde se diplomantka zaměří na principy analýzy genetické informace s využitím fragmentačních technik a sekvenačních postupů. Budou rozebrány doposud publikované postupy generování molekulárních dat u smrku ztepilého.

Jako vstup pro praktickou část práce budou připraveny vzorky DNA smrku ztepilého vyhovující vstupními parametry požadavkům sekvenační laboratoře.

Ve spolupráci se sekvenační laboratoří budou nastaveny selekční parametry pro potenciálně vhodné mikrosatelitové primery. S využitím vybraných primerů budou připraveny reakční směsi a provedena amplifikace pomocí PCR. Amplikonové produkty budou sekvenovány v laboratoři.

Získaná data budou zpracována a vyhodnocena, očekává se průběžná adjustace experimentu na základě charakteru průběžných výstupů.

Funkčnost finálně navrženého genotypizačního protokolu bude ověřena na testovacím souboru vzorků smrku ztepilého. Nový postup bude porovnán s technikou klasické fragmentační separace pro smrk ztepilý (Bínová et al. 2020).

Harmonogram práce: Do konce prosince 2020 bude vypracována literární rešerše a detailně zpracována metodika diplomové práce. Vyhodnocení NGS dat bude provedeno do konce února 2021.



Doporučený rozsah práce

min. 50 stran

Klíčová slova

NGS sekvenování, smrk ztepilý, mikrosatelitové markery

Doporučené zdroje informací

- Bínová, Z., Korecký, J., Dvořák, J., Bílý, J., Zádřapová, D., Jansa, V., & Lstibůrek, M. (2020). Genetic Structure of Norway Spruce Ecotypes Studied by SSR Markers. *Forests*, 11(1), 110.
- De Barba, M., Miquel, C., Lobréaux, S., Quenette, P. Y., Swenson, J. E., & Taberlet, P. (2017). High-throughput microsatellite genotyping in ecology: improved accuracy, efficiency, standardization and success with low-quantity and degraded DNA. *Mol Ecol Res*, 17(3), 492-507.
- Guichoux, E., Lagache, L., Wagner, S., Chaumeil, P., Léger, P., Lepais, O., ... & Petit, R. J. (2011). Current trends in microsatellite genotyping. *Mol Ecol Res*, 11(4), 591-611.
- Jarne, P., & Lagoda, P. J. (1996). Microsatellites, from molecules to populations and back. *Trends Ecol Evol*, 11(10), 424-429.
- Lesk, A. M. (2017). Introduction to genomics. Oxford University Press.
- Metzker, M. L. (2010). Sequencing technologies—the next generation. *Nat Rev Genet*, 11(1), 31-46.
- Nystedt, B., Street, N. R., Wetterbom, A., Zuccolo, A., Lin, Y. C., Scofield, D. G., ... & Vicedomini, R. (2013). The Norway spruce genome sequence and conifer genome evolution. *Nature*, 497(7451), 579-584.
- White, T. L., Adams, W. T., & Neale, D. B. (2007). *Forest Genetics*. CABI, Wallingford.

Předběžný termín obhajoby

2020/21 LS – FLD

Vedoucí práce

Ing. Jiří Korecký, Ph.D.

Garantující pracoviště

Katedra genetiky a fyziologie lesních dřevin

Elektronicky schváleno dne 25. 10. 2020

prof. Ing. Milan Lstibůrek, MSc., Ph.D.

Vedoucí katedry

Elektronicky schváleno dne 25. 10. 2020

prof. Ing. Róbert Marušák, PhD.

Děkan

V Praze dne 08. 12. 2020

Čestné prohlášení

Prohlašuji, že jsem diplomovou práci na téma "Metody sekvenování nové generace (NGS) u smrku ztepilého – Modifikace analýzy mikrosatelitových amplikonů" vypracovala samostatně pod vedením Ing. Jiřího Koreckého, Ph.D. a použila jen prameny, které uvádím v seznamu použitých zdrojů.

Jsem si vědoma, že zveřejněním diplomové práce souhlasím s jejím zveřejněním dle zákona č. 111/1998 Sb. o vysokých školách v platném znění, a to bez ohledu na výsledek její obhajoby.

V Praze dne 9. 4. 2022

Poděkování

Ráda bych touto cestou poděkovala vedoucímu své diplomové práce Ing. Jiřímu Koreckému, PhD., který mě vždy uměl správně nasměrovat, když jsem se v tématu práce začala ztrácet. Dále bych chtěla poděkovat Fakultě lesnické a dřevařské České zemědělské univerzity v Praze, za možnost studovat obor, který jsem si zvolila a zejména bych chtěla poděkovat všem na Katedře genetiky a fyziologie lesních dřevin za krásných několik let spolupráce ať už na mé bakalářské práci, nebo na jiných projektech v rámci mé pomocné vědecké činnosti na katedře. Největší dík však patří mé rodině, která mě ve vzdělání vždy podporovala a stála při mně.

Abstrakt

Diplomová práce řeší možnosti implementace metod sekvenování nové generace (NGS) do analýzy mikrosatelitových markerů. Tyto moderní metody sekvenování genetické informace jsou založeny na sofistikovanějších a přesnějších postupech, se zaměřením na sekvenování konkrétních SSRs (*Simple Sequence Repeats*) / STRs (*Short Tandem Repeats*) fragmentů DNA s využitím platformy Illumina (Solexa).

V rámci práce byla navržena a otestována metodika modifikované analýzy mikrosatelitových amplikonů smrku ztepilého (*Picea abies* (L.) H. Karst.). Po několika etapovém navrhování primerů a optimalizování multiplexů byly sestaveny sekvenační knihovny. Vybrané knihovny byly sekvenovány a před vlastní analýzou byla ověřena kvalita výstupních dat. Funkčnost navržené metody byla otestována na souboru dvanácti vzorků smrku ztepilého (*Picea abies* (L.) H. Karst.) pocházejících z různých oblastí České republiky.

Pro ověření funkčnosti a přesnosti nově navržené metodiky byly výsledky sekvenace porovnány s výsledky konvenční metody analýzy amplikonů, provedené pomocí kapilární elektroforézy, která se pro práci s krátkými opakujícími se repetitivy nukleotidových motivů (mikrosatelity) hojně používá. Diplomová práce odhaluje silné, ale i slabé stránky spojené s aplikačním využitím metod sekvenování nové generace pro analýzu mikrosatelitových lokusů a nabízí nové pohledy na tuto problematiku.

Klíčová slova: NGS sekvenování, smrk ztepilý, mikrosatelitové markery

Abstract

The diploma thesis addresses the possibilities of implementing new generation sequencing (NGS) methods into the analysis of microsatellite markers. These methods of sequencing genetic information are based on more sophisticated and accurate approaches, focusing on sequencing specific SSRs (Simple Sequence Repeats) / STRs (Short Tandem Repeats) DNA fragments using the Illumina platform (Solexa).

Within the work, the methodology of modified analysis of microsatellite amplicons of Norway spruce (*Picea abies* (L.) H. Karst.) was designed and tested. After several steps of primer design and multiplex optimization, sequencing libraries were assembled. Selected libraries were sequenced, and the quality of the output data was verified before the actual analysis. The functionality of the proposed method was tested on a set of twelve samples of Norway spruce (*Picea abies* (L.) H. Karst.) originating from different areas of the Czech Republic.

To verify the functionality and accuracy of the newly designed methodology, the results of sequencing were compared with the results of the conventional method of amplicon analysis performed by capillary electrophoresis, which is widely used to work with short repetitive repeats of nucleotide motifs (microsatellites). The diploma thesis reveals the strengths and weaknesses associated with the application of next-generation sequencing methods for the analysis of microsatellite loci and offers new perspectives on this issue.

Keywords: NGS sequencing, Norway spruce, microsatellite markers

Obsah

1 Úvod	15
2 Cíle práce	17
3 Literární rešerše	18
3.1 Rod <i>Picea</i>	18
3.1.1 Smrk ztepilý (<i>Picea abies</i> (L.) H. Karst.)	18
3.1.2 Genetická variabilita smrku ztepilého (<i>Picea abies</i> (L.) H. Karst.)	21
3.2 Molekulárně genetické nástroje	22
3.2.1 Mikrosatelitové molekulární markery	23
3.2.1.1 SSRs (Simple Sequence Repeats) / STRs (Short Tandem Repeats) .	23
3.3 Sekvenování první generace	25
3.3.1 Sangerova metoda	26
3.3.2 Analýza fragmentů DNA	28
3.4 Sekvenování nové generace – <i>Next Generation Sequencing</i> (NGS)	30
3.4.1 Platforma Illumina (Solexa).....	32
3.4.2 Roche/454 FLX.....	39
3.4.3 Ion Torrent	41
3.4.4 Pacific Biosciences instruments – <i>Single molecule real time</i> (SMRT™) <i>sequencer</i>	42
3.5 Možnosti využití metod Sekvenování nové generace	43
3.6 Přístupy k datové analýze.....	45
3.6.1 Kontrola kvality dat	45
3.6.1.1 FastQC	45
3.6.1.2 Kvalita bazí.....	46
3.6.1.3 Trimming	46
3.6.1.4 Obsah GC	46
3.6.2 Zarovnání readů dle reference	47
4 Metodika	48
4.1 Návrh metodiky a metodické postupy analýzy genomové DNA.....	48
4.2 Fáze 0	49
4.2.1 Zájmová populace a sběr vzorků	49
4.2.2 Příprava vzorků a homogenizace	51
4.2.3 Extrakce a purifikace nDNA.....	51
4.2.4 Stanovení kvality a koncentrace DNA.....	52
4.2.5 Kontrola kvality DNA	53
4.3 Fáze 1	54
4.3.1 Navržení a vytvoření primerů	54

4.3.2	Vytvoření a optimalizování multiplexů	55
4.3.3	Analýza dat a vytvoření sekvenačních knihoven.....	56
4.4	Fáze 2	57
4.4.1	Příprava knihoven na sekvenování	57
4.4.2	Sekvenování a kontrola kvality dat.....	61
4.5	Fáze 3	62
4.5.1	Rozšíření poolu/knihovny č. 3	62
4.6	Fáze 4	63
4.6.1	Výběr SSR lokusů.....	63
4.6.2	Příprava knihoven	64
4.6.3	Paralelní sekvenování	66
4.6.1	Kontrola kvality QC a analýza sekvenovaných amplikonů.....	66
5	Výsledky	67
5.1	Fáze 0	67
5.1.1	Kvalita a koncentrace DNA	67
5.2	Fáze 1	69
5.2.1	Vytvořené primery	69
5.2.2	Multiplexní PCR a vytvoření sekvenačních knihoven	73
5.3	Fáze 2	76
5.3.1	Sekvenování.....	76
5.3.2	Kontrola kvality QC.....	78
5.4	Fáze 3	79
5.5	Fáze 4	80
5.5.1	Příprava knihoven	80
5.5.2	Analýza sekvenovaných amplikonů	81
5.5.3	Konvenční Fragmentační analýza.....	86
5.6	Porovnání výsledků metod sekvenování	87
6	Diskuse	90
7	Závěr.....	93
8	Seznam literatury a použitých zdrojů	95
9	Seznam příloh	105
10	Přílohy	106

Seznam obrázků

Obrázek 1 - Rozšíření Smrku ztepilého v České republice	19
Obrázek 2 - Chemická struktura ddNTPs a dNTPs	26
Obrázek 3 - Horizontální elektroforéza	28
Obrázek 4 - Typy sekvenačních přístrojů nabízených společnostmi Illumina	33
Obrázek 5 - Grafické znázornění pracovního postupu Illumina.....	33
Obrázek 6 - Typy sekvenační chemie - Illumina.....	37
Obrázek 7 - Příprava DNA knihovny	39
Obrázek 8 - Lokality sběru vzorků na mapě ČR	49
Obrázek 9 – Soustava vybavená pro horizontální gelovou elektroforézu	53
Obrázek 10 - Výsledek elektroforetické separace DNA na agarózovém gelu	68
Obrázek 11 - Lokus EATC1D02A, vzorek A1 a neočekávaná repetice TCC.....	82
Obrázek 12 - Lokus PAAC3 vzorek A1, neočekávaná repetice GT	83

Seznam tabulek

Tabulka 1 - Taxonomické zařazení druhu (<i>Picea abies</i> (L.) H. Karst.)	18
Tabulka 2 - Typy sekvenátorů a na nich závislé typy chemie sekvenování	36
Tabulka 3 - Porovnání současných NGS platforem	44
Tabulka 4 - Značení vzorků a jejich rozdělení dle lokalit sběru.....	50
Tabulka 5 – Složení PCR reakce	54
Tabulka 6 – Program pro průběh PCR reakce	54
Tabulka 7 - Složení 1. PCR reakce	57
Tabulka 8 - Program průběhu 1. PCR.....	57
Tabulka 9 – Složení 2. PCR reakce	58
Tabulka 10 - Program průběhu 2. PCR.....	58
Tabulka 11 - Skladba primerů 1. knihovny	59
Tabulka 12 - Skladba primerů 2. knihovny	59
Tabulka 13 - Skladba primerů 3. knihovny	60
Tabulka 14 - Skladba primerů 4. knihovny	60
Tabulka 15 - Skladba primerů 5. knihovny	60
Tabulka 16 - Kontrola kvality a velikost knihoven (bp).....	61
Tabulka 17 - Výsledky měření koncentrace DNA a čistoty vzorků.....	67
Tabulka 18 - Účinnost amplifikace u jednotlivých primerů	70

Tabulka 19 - Vypočítané jednotlivé zastoupení amplikonů obsahujících primerové sekvence (forward a reverse readů).	71
Tabulka 20 - Vypočítané jednotlivé zastoupení amplikonů obsahujících primerové sekvence (oba ready spojeny).	73
Tabulka 21 - Repetice a výskyt lokusů - pool 1 (6x a11)	76
Tabulka 22 - Výsledky sekvenace a kontroly kvality knihoven (R1, R2)	78
Tabulka 23 – Výsledky sekvenace a kontroly kvality knihoven	78
Tabulka 24 – Jednotlivé lokusy a jejich stav po sekvenování	81
Tabulka 25 - Počet readů a délky lokusů dle NGS	84
Tabulka 26 - Výsledky genotypizování osmi vzorků po fragmentační analýze	86
Tabulka 27 - Porovnání délek lokusů u NGS a konvenční metody sekvenování	88

Seznam grafů

Graf 1 – Spektrální křivka DNA a proteinu	52
---	----

Seznam rovnic

Rovnice 1 – Phred skóre	46
Rovnice 2 - Phred skóre příklad	46

Seznam použitých zkratk, symbolů a výrazů

A – adenin

ATP – adenosintrifosfát (nukleotid)

bp – páry bází (*Base pair*)

C – cytosin

d_dH_2O – ultra čistá voda

DNA – deoxyribonukleová kyselina (*Deoxyribonucleic acid*)

FLD – Fakulta lesnická a dřevařská

G – guanin

GTP – guanosintrifosfát (nukleotid)

IR – Inverted repeats

kbp – kilo páru bází (odvozená jednotka) $1\text{kbp}=1*10^3\text{ bp}$

LSC – *Large single copy*

LVS – lesní vegetační stupeň

Mbp – mega páru bází (odvozená jednotka) $1\text{Mbp}=1*10^6\text{ bp}$

mg – miligram (odvozená jednotka) $1\text{mg}=1*10^{-3}\text{ g}$

mismatch – neshoda, nevhodné spojení

ml –mililitr (odvozená jednotka) $1\text{ml}=1*10^{-3}\text{ l}$

mRNA – mediátorová ribonukleová kyselina (*Messenger Ribonucleic acid*)

nDNA – jaderná DNA (*Nuclear DNA*)

NGS – Sekvenování nové generace (*Next Generation Sequencing*)

nS – počet jednotek Svedberga (odvozená jednotka) $1\text{S}=10^{-13}\text{ sekundy}$

PCR – polymerázová řetězová reakce

read, ready– čtení, DNA sekvence z daného fragmentu

RNA – ribonukleová kyselina

rRNA – ribozomální ribonukleová kyselina

Rubisco (RuBisCo) – ribulosa – 1,5 – bisfosfát – karboxylasa/oxygenasa (enzym)

S-fáze – fáze buněčného cyklu mezi G₁ fází a G₂ fází

SBS – *Sequencing by Synthesis*

SNP – *Single nucleotide polymorphism*

SSC – *Small single copy*

STR – *Short Tandem Repeats*

SSRs – *Simple sequence repeats*

T – thymin

tRNA – transferová ribonukleová kyselina (*Transfer Ribonucleic acid*)

U – uracil

μm – mikrometr (odvozená jednotka) 1 μm = 1 * 10⁻⁶ m

μl – mikrolitr (odvozená jednotka) 1 μl = 1 * 10⁻⁶ l

1 Úvod

V posledních letech, více nežli dříve, cítíme potřebu hledat nové možnosti a cesty, jak přistupovat k lesnímu hospodářství. S gradující změnou klimatu přichází nové výzvy a problémy, které je potřeba překonávat. Kůrovcová kalamita je jeden z mnoha příznaků toho, že smrkové monokultury špatně odolávají klimatické změně. Smrk ztepilý (*Picea abies* (L.) H. Karst.) jakožto naše hlavní hospodářská dřevina má a vždy bude mít v českých lesích své místo. Je zcela zřejmé, že bychom v lesnictví měli usilovat především o zvyšování udržitelnosti hospodaření, a to jak z hlediska ekologického, tak i z hlediska ekonomického a sociálně-politického. S tím souvisí i management genových zdrojů, který je v našich podmínkách často poměrně opomíjený. Možnost odhalit a správně interpretovat informace obsažené v genomech lesních dřevin nám v lesnictví přináší řadu nových možností. Můžeme například efektivněji realizovat šlechtitelské programy, identifikovat od sebe jednotlivé ekotypy druhů lesních dřevin anebo zakládat a udržovat porosty stabilnější a odolnější vůči škůdcům a klimatickým změnám.

Existuje mnoho přístupů a nástrojů pro analýzu genetické informace, některé jsou robustnější a jiné více sofistikované. Mikrosatelitové markery jsou nástrojem, kterým lze odhalit míru příbuzenských vazeb mezi zájmovými jedinci. Běžně se využívá konvenční přístup analýzy získaných mikrosatelitových amplikonů, a tím je délková separace fragmentů založená na odlišné rychlosti jejich prostupu příslušným prostředím, kterým může být například agarózový gel či tenká kapilára přístroje pro fragmentační analýzu. Tento přístup s sebou ovšem nese i určité nevýhody a limity. Informace o délce fragmentů, a tedy i genotypizační výsledky, mohou být nekompatibilní mezi jednotlivými pracovišti, a to i v případě striktního dodržování genotypizačního protokolu. Důvodem může být například odlišná šarže chemikálií pro sekvenační přístroj či jiná laboratorní teplota. Mnohdy však nejsou příčiny vůbec zřejmé. Výsledky mohou být také zkresleny subjektivním ovlivněním hodnotitele při genotypizování. Analýza délkovou separací také neumožňuje odhalovat bodové mutace či další změny v genetické informaci.

Moderní sekvenační platformy jsou již natolik pokročilé a postupně se snižuje jejich cenová náročnost, že se jejich využití neomezuje pouze na humánní genetiku, ale mohou být rutinně využívány i v zemědělství a lesnictví. Primárním impulzem pro práci na metodě analýzy mikrosatelitů pomocí NGS byly zásadní potenciální přínosy této metody. Analýza není omezena pouze na informaci o délce fragmentu, ale je znám

charakter jednotlivých nukleotidů, což umožňuje odhalovat i mutační změny nejenom v repetitivních oblastech, ale také v ohraničujících sekvencích. Je také eliminováno riziko subjektivního vlivu hodnotitele a efekt různých laboratorních podmínek.

2 Cíle práce

Cílem této diplomové práce je vytvořit metodu analýzy mikrosatelitových amplikonů na bázi NGS sekvenování pro smrk ztepilý (*Picea abies* (L.) H. Karst.). Protokol by měl využívat mikrosatelitové DNA fragmenty SSRs (*Simple Sequence Repeats*) / STRs (*Short Tandem Repeats*) s inovativním přístupem k analýze PCR amplikonů. Jde o nahrazení konvenčního přístupu separace amplikonů pomocí kapilární elektroforézy novým přístupem, a to sekvenací amplikonů pokročilými metodami sekvenování. Funkčnost vytvořeného protokolu bude ověřena na testovacím souboru vzorků smrku ztepilého (*Picea abies* (L.) H. Karst.). Metodická část bude provedena v kooperaci s externí společností SEQme (<https://www.seqme.eu>). Nový postup a jeho funkčnost bude porovnán s technikou klasické fragmentační separace pro smrk ztepilý (Bínová et. al 2020).

3 Literární rešerše

3.1 Rod *Picea*

Rod *Picea*, neboli smrk, je oproti jiným druhům dřevin poměrně morfologicky variabilní. Jeho vysoká morfologická variabilita uvnitř rodu znamená zhoršené rozlišování a určování jednotlivých druhů (Musil a Hámerník 2003). Z tohoto důvodu je pro spolehlivé určení daného druhu, hybridu, popř. ekotypu nutné provádět genetické analýzy.

Rod *Picea* je přirozeně rozšířený zejména v jehličnatých lesích boreální až temperátní květenné oblasti severní polokoule. V Evropě se vyskytují autochtonně pouze 3 druhy již zmíněného rodu, z toho v České republice jen 1 druh, a to smrk ztepilý (*Picea abies* (L.) H. Karst.).

3.1.1 Smrk ztepilý (*Picea abies* (L.) H. Karst.)

Smrk ztepilý (*Picea abies* (L.) H. Karst.) je nejpolymorfnějším taxonem z rodu *Picea*, a to jak do interpopulační, tak do intrapopulační variability. U smrku ztepilého lze pozorovat širokou morfologickou variabilitu, zejména u tvaru koruny, šišek a větvení (Musil & Hámerník, 2003).

Tabulka č. 1 znázorňuje taxonomické zařazení zájmového druhu dle biologické systematiky.

Tabulka 1 - Taxonomické zařazení druhu (*Picea abies* (L.) H. Karst.)

Základní taxonomické kategorie		
Taxony	Latinský název	Český název
Říše	<i>Plantae</i>	Rostliny
Kmen	<i>Gymnospermae</i>	Nahosemenné
Oddělení	<i>Pinophyta</i>	Jehličnany
Třída	<i>Pinopsida</i>	Jehličnaté
Řád	<i>Pinales</i>	Borovicotvaré
Čeleď	<i>Pinaceae</i>	Borovicovité
Rod	<i>Picea</i>	Smrk
Druh	<i>P. abies</i>	Smrk ztepilý

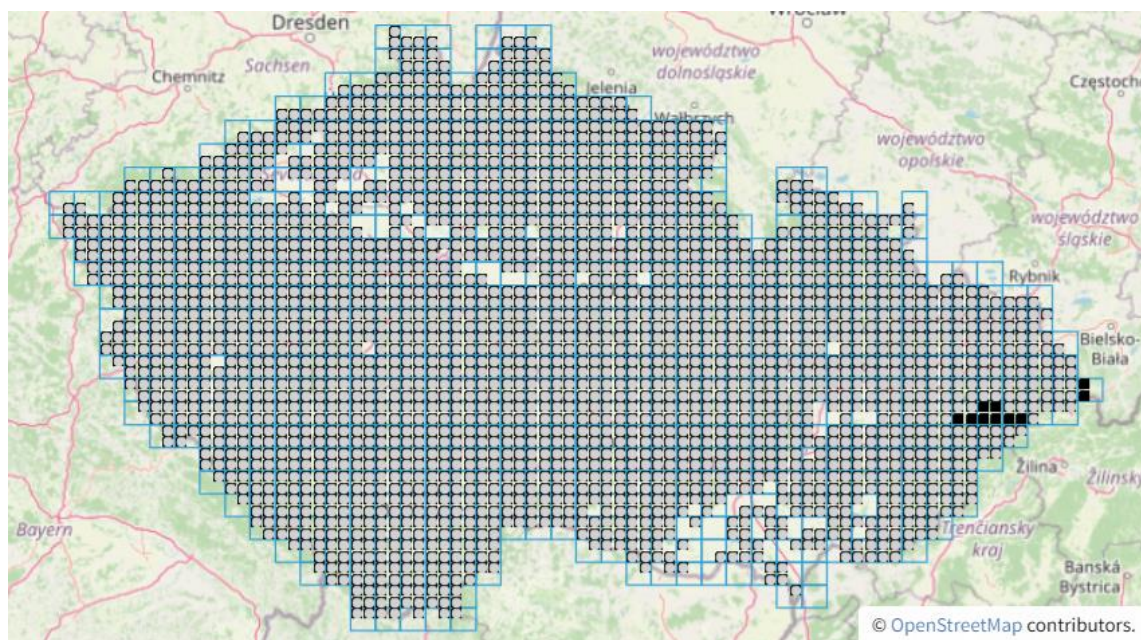
Smrk ztepilý je druhem boreálních hor s širokou a roztroušenou oblastí rozšíření. Přirozeně se vyskytuje v horských lesích. Avšak je vysoce tolerantní k různosti podmínek stanovišť, ve kterých se jeho populace vyskytují. Lze ho tedy najít i jako příměs na stanovištích, kde nikdy nemůže dosahovat své klimaxu – chladnější podmáčená stanoviště, luhy, stinná údolí, rokliny a okraje rašelinišť. Klimaxové smrčiny jsou v České republice v oblasti oreofytika. Nad horní hranici lesa (1300-1400 m) se

smrk ztepilý dostává jednotlivě do pásma kosodřevin, kde vytváří horní hranici stromovou. V našich podmínkách smrk ztepilý roste nejvýše ve výškách kolem 1550 m. n. m. na Sněžce. Naopak nejnižše ho můžeme najít sestoupený až na 140 m. n. m. v klimaticky inverzních soutěskách. Kde roste díky dostatečně koncentrovanému, vlhkému a studenému vzduchu v kombinaci s půdní vlhkostí. V takto nižších polohách by se smrk vyskytoval přirozeně jen ojediněle, a to jako příměs zejména na rašelinách, v chladnějších roklinách nebo na podmáčených půdách.

Hojně bychom ho přirozeně se vyskytující našli také v 700–1000 m v příměsi s bukem lesním (*Fagus sylvatica* L.), jedlí bělokorou (*Abies alba* Mill.) a javorem klenem (*Acer pseudoplatanus* L.).

Je citlivý na extrémně kontinentální klima, jež se vyznačuje velkými rozdíly teploty vzduchu, nízkými úhrny srážek a malou oblačností. Je také citlivý na pozdní jarní přízemní mrazíky. Jednotlivé klimatické ekotypy smrku se vyznačují různými náchylnostmi k rozdílným klimatickým podmínkám (Obidowicz et al. 2004).

Jedná se o druh světlomilný, který však v mládí snese mírný zástin. Jeho požadavky na kvalitu půdy jsou mírné, ale je velmi náročný na půdní vlhkost. Obecně lze říct, že dosahuje svého optima na mírně vlhkých, středně úrodných, a ne příliš kyselých půdách. Aktuální rozšíření tohoto druhu je znázorněno na obrázku č. 1.



Obrázek 1 - Rozšíření Smrku ztepilého (*Picea abies* (L.) H. Karst.) v České republice, zdroj: [Plant Diversity Analysis and Synthesis Centre](#)

Produkční optimum smrku je v rozmezí 600–1000 m. n. m. Navzdory společenským trendům posledních let je smrk ztepilý stále naší nejdůležitější

hospodářskou dřevinou s širokou škálou upotřebení (stavební dříví, vláknina, rezonanční dřevo, pryskyřice, borka) (Slávik 2004).

Je však důležitý i z hlediska mimoprodukčních funkcí lesa a funkcí ekologických (půdoochranná, vodoochranná, klimatická, zdravotní). K mimoprodukčním funkcím lesa nelze opomenout přiřadit i funkci ekologicko-stabilizační a funkci společenskou. (Holuša a Holušová 2020).

V lesích České republiky by se měl smrk vyskytovat přirozeně pouze na 11,2 %, nyní však zaujímá 48,8 % plochy porostní půdy, doporučená skladba je 36,5 %. I když je na našem území více smrku, než by bylo přirozeně, trend jeho výskytu v posledních letech klesá. Pro srovnání od roku 2000 plocha porostní půdy jež zaujímá smrk ztepilý klesla z 1 397 012 ha na 1 274 241 ha (rok 2020), což je pokles o cca 4,7 %. Tento pokles je dlouhodobě zapříčiněn globální změnou klimatu konkrétně suchem a následnou kůrovcovou kalamitou (Ministerstvo zemědělství České republiky 2021).

V roce 2019 došlo opět k odchýlení od dlouhodobého teplotního normálu (z let 1981-2010), a to o +1,6 °C. Ve stejném roce bylo v Česku evidováno 20,7 mil. m³ vytěženého kůrovcového dříví. (Ministerstvo zemědělství České republiky 2020).

V roce 2020 byl teplotní normál překonán o +1,2 °C, z tohoto pohledu byl rok 2020 méně extrémnější než rok 2019. Nejvíce teplotně nadprůměrný byl měsíc únor (+4,6 °C), naopak nejvíce podprůměrný byl květen (-2,1 °C). Rok 2020 byl zlomový z hlediska srážkových úhrnů, které dosáhly za poslední desetiletí nevyšší, a to na 764 mm. Bylo dosaženo 112 % dlouhodobého normálu srážkové bilance. V roce 2020 bylo evidováno cca 21,9 mil. m³ vytěženého kůrovcového dříví, což je rekordní hodnota. Oproti minulému roku to znamená nárůst o 1,2 mil. m³. Na svědomí jej má zejména lýkožrout smrkový (*Ips typographus*), lýkožrout lesklý (*Pityogenes chalcographus*) a lýkožrout severský (*Ips dupliatus*) (Ministerstvo zemědělství České republiky 2021).

3.1.2 Genetická variabilita smrku ztepilého (*Picea abies* (L.) H. Karst.)

Ekotyp je populace s určitou dědičnou genetickou strukturou, kterou dlouhodobě utvářelo prostředí. Morfologie jednotlivých ekotypů se mezi sebou mírně liší, ať už ve tvaru koruny, vzhledu šišek nebo větvení. Na území České republiky rozeznáváme 3 různé ekotypy smrku ztepilého – ekotyp chlumní (*forma acuminata*), horský (*forma europea*) a vysokohorský (*forma obovata*). Při nakládání s ekotypy platí, že je ekologicky i ekonomicky výhodné a důležité zachovávat genofondy původních ekotypů přizpůsobených danému prostředí (Cvrčková et al. 2011).

Proto by se informace o původu osiva, popř. semenáčků jednotlivých dřevin neměly brát na lehkou váhu. Zejména při následné obnově či zakládání nových porostů je důležité brát v potaz, z jakého prostředí rodičovské generace pocházely a v jakém prostředí se tyto generace vyvíjely.

3.2 Molekulárně genetické nástroje

Deoxyribonukleová kyselina – DNA nese dědičnou informaci daného jedince. Analýzou molekul DNA je možné získat informace o genomu unikátním pro každého jedince. Kvůli časové a finanční nákladnosti se ve výzkumech v oboru molekulární biologie se zaměřením na lesní dřeviny běžně neanalyzuje kompletní genom zájmových jedinců, ale spíše se využívá různých markerů. Protože genetický marker je jakýkoliv viditelný znak či jinak testovatelný fenotyp, pro který se alely na lokusech segregují dle pravidel, která popsal J. G. Mendel, bylo zprvu jejich použití u dřevin složité. Dřeviny nemají tak velké množství viditelných Mendelovských znaků, což omezilo lesnický výzkum. Ke změně došlo na počátku 70. let 20. století, kdy byly vyvinuty první biochemické markery vhodné pro dřeviny, a to terpenové a allozymové markery. Ty však nebyly vhodné pro zkoumání genů napříč genomem, kvůli malému množství lokusů, proto byly rychle překonány molekulárně genetickými markery (White et al. 2007).

Molekulárně genetické markery jsou předem vybrané různě dlouhé úseky DNA. Molekulární marker je tedy pro daný experiment vhodně zvolená jednotka genetické informace, která byla získána analýzou DNA. Může se jednat o fragment DNA o určité délce, známou sekvenci, jeden opakující se nukleotid apod. Markery nejsou jen záležitostí jaderné nDNA, ale také například mimojaderné (organelární) DNA. U smrku nejčastěji mitochondriální mtDNA a chloroplastové cpDNA. Ve svých pracích se tímto tématem zabývali například Jaramillo-Correa et al. 2003, kteří tvrdí, že mimojaderná DNA je užitečná jako zdroj haploidních genetických markerů, vhodných ke zkoumání populační genetiky a biogeografie.

Některé vybrané typy analýz molekulárně genetických markerů jsou vypsány níže.

- a) Metoda RFLP (*Restriction Fragment Length Polymorphism*)
- b) Analýza fragmentů DNA s použitím PCR reakce
 - Analýza celého genomu
 - RAPD (*Random Amplified Polymorphic DNA*)
 - AFLP (*Amplified Fragment Length Polymorphism*)
 - ISSRs (*Inter Simple Sequence Repeats*)
 - Analýza konkrétních částí genomu
 - ESTPs (*Expressed Sequence Tagged Polymorphisms*)
 - SSRs (*Simple Sequence Repeats*) / STRs (*Short Tandem Repeats*)
 - SNPs (*Single Nucleotide Polymorphisms*)

3.2.1 Mikrosatelitové molekulární markery

Mikrosatelity (1-10 nukleotidů) a minisatelity (>10 nukleotidů) jsou podkategoriemi tandemových repetic, které spolu s převládajícími rozptýlenými repeticemi tvoří repetitivní oblasti genomu (Vieira et al. 2016).

Mikrosatelitové markery mají široké využití, v nedávné době byl na jejich základě sestaven protokol genotypizace lokusů úspěšně aplikovaný na 3 genotypy smrku ztepilého v České republice. Jedná se tedy mimo jiné o vhodný nástroj k vyhodnocení informací ohledně zdroje a původu semen či sazenic (Bínová et al. 2020).

3.2.1.1 SSRs (Simple Sequence Repeats) / STRs (Short Tandem Repeats)

Tyto mikrosatelitové markery byly vyvinuty původně pro genetické mapování lidského genomu (Weber a May 1989; Litt a Luty 1989). Jedná se o krátké v tandemu se opakující sekvence dvou, tří či čtyř nukleotidů. Počet tandemových repetic na daném lokusu se může lišit, a proto je tento typ markerů často velmi polymorfní.

SSRs markery bývaly nejrozšířenějším typem markerů mezi lety 1996-2016, protože jsou vysoce informativní, kodominantní, multialelové a zároveň jsou experimentálně reprodukovatelné a přenositelné mezi příbuznými druhy (Mason 2015).

Při použití těchto markerů je nejprve potřeba vytvořit a otestovat genomovou knihovnu obsahující mikrosatelitové sekvence. Testují se klony SSR sekvencí, které musí být přesně určeny, aby mohli být navrženy PCR primery pro amplifikaci SSR sekvencí z jednotlivých vzorků. Primery jsou navrženy tak, aby byly specifické pro daný lokus a druh. Délka primerů by měla být 15-25 nukleotidů, je to optimální počet pro efektivní nasedání primerů. Pro většinu druhů již primery existují a jsou dostupné v databázích, vědeckých publikacích a genových bankách (Selkoe a Toonen 2006).

Nedávné studie ukázaly, že nejjednodušší způsob identifikace SSR lokusů je za pomoci NGS (sekvenace genomu či transkriptomu) (Zalapa et al. 2012). Ještě donedávna se používala Sangerova metoda sekvenování s použitím genových knihoven obohacených o sekvence mikrosatelitů. V dnešní době se tato metoda spíše nahrazuje metodou NGS.

DNA SSR markery byly již vytvořeny a otestovány pro mnoho druhů dřevin, a to jak jaderné, tak organelární markery. Pro smrk ztepilý jich bylo v různých studiích použito nespočet. Jen mezi lety 2010-2015 v článku od Vieira et al. z roku 2016 bylo

v databázi Web of Science (Web of Science™ Core Collection) nalezeno pod klíčovými slovy „*SSR markers, microsatellites*“ 933 unikátních záznamů o rostlinných SSR markerech. Poté jejich počet klesal, pravděpodobně nástupem SNP markerů.

3.3 Sekvenování první generace

Již počátkem 70. let 20. století biochemici Angličan Frederick Sanger a Američané Walter Gilbert a Paul Berg přišli s různými metodami sekvenování DNA. Metody sekvenace jsou postupy, díky kterým je možné v určitém úseku DNA analyzovat pořadí jednotlivých nukleotidů (adeninu, cytosinu, thyminu, guaninu). Některé ze sekvenačních metod jsou vhodné i pro nepřímou analýzu sekvencí RNA, která se nejdříve reverzní transkripcí přepíše na DNA. Vezmeme-li v úvahu, že sekundární struktura jaderné DNA (tvar pravotočivé dvoušroubovice složený ze dvou komplementárních polymerních řetězců) byla popsána teprve v roce 1953 Britem Francisem Crickem, Američanem Jamesem Watsonem a dnes méně známým Angličanem Mauricem Wilkinsem, můžeme Sangera, Gilberta a Berga pokládat za opravdové průkopníky sekvenačních metod své doby. Tito biochemici dokonce za svůj objev sekvenačních metod získali v roce 1980 Nobelovu cenu (Sturtevant a Morgan, 2001).

Díky rychle se rozvíjejícím sekvenačním metodám je tedy možné zaměřit se na zkoumání primární struktury DNA či RNA, tedy délky jednotlivých repetitivních úseků či v případě sekvenačních metod nové generace i na konkrétní pořadí jednotlivých nukleotidů. Znalost pořadí jednotlivých nukleotidů v sekvenci DNA nebo RNA se stala neodmyslitelnou nejen pro biologický výzkum, ale i pro mnoho praktických oborů, jako například biotechnologii, virologii, forenzní biologii, diagnostiku v medicíně atd.

Použití sekvenačních metod může být na více úrovních. Lze sekvenovat celé genomy jednotlivých organismů, nebo se zaměřit na dané chromozomy, klastry (větší genové oblasti) či jednotlivé geny (Pareek et al. 2011).

3.3.1 Sangerova metoda

Zatímco Gilbertova metoda je založena na chemických procesech, které zapříčiní rozpad DNA na specifické části jednotlivých nukleotidů, Sangerova metoda využívá ve svůj prospěch proces syntézy DNA. Zejména proces modifikované replikace DNA neboli samozdvojení. Jedná se o děj, který se standartně odehrává při mitóze buněk. Účelem replikace je, aby se dědičná informace, uložená v jádře buněk, dostala při buněčném dělení do všech buněk dceřiných (Koolman a Röhm 2012).

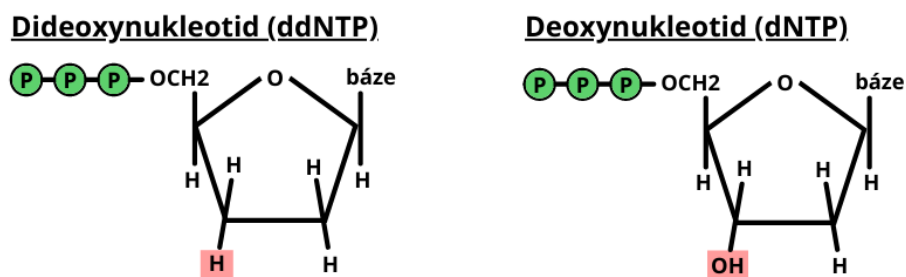
Reakční směs pro modifikovanou replikaci dle Sangera obsahuje:

- templátovou DNA,
- DNA polymerázu,
- specifické sekvenační primery,
- deoxyribonukleotidtrifosfáty – dNTPs (jednotlivé nukleotidy – dA, dC, dT, dG)
- dideoxyribonukleotidfosfáty – ddNTPs (fluorescenčně značené – ddA, ddC, ddT, ddG).

Dále směs obsahuje pufrovací roztoky (ionty, stabilizátory). Tato reakční směs bývá označována jako tzv. mastermix.

Vylepšený a modifikovaný proces dle Sangera začíná tím, že na templátové (zájmové) DNA jsou pomocí primerů ohraničena iniciační místa, která jsou díky DNA polymeráze, dNTPs a ddNTPs replikována. Nukleotidy jsou začleňovány do komplementárního řetězce k ohraničeným úsekům templátové DNA. Jelikož ddNTPs mají jinou chemickou strukturu než dNTPs viz obrázek č. 2, tak se náhodně při replikaci začleňují a ukončují reakci za vzniku fragmentů o různé délce (Wang 2016).

Obrázek 2 - Chemická struktura ddNTPs a dNTPs, převzato z: [OpenStax College™](#), upraveno



Pomocí kapilární elektroforézy je možné seřadit jednotlivé fragmenty dle délky. Protože ddNTPs jsou fluorescenčně značené, je možné za pomoci laseru odečíst konečnou sekvenci. Data jsou pak vyhodnocována počítačovým softwarem např. GeneMarker od společnosti SOFTGENETICS.

Jelikož Sangerova metoda byla objevena v 70. letech 20. století, bioinformatika a biochemie nebyla ještě tolik vyvinutá. Původní Sangerův protokol tedy zkoumal ddNTPs odděleně ve čtyřech zkumavkách, kdy každá ze zkumavek byla pouze na jeden nukleotid. Tím pádem v každé ze 4 reakcí vznikaly fragmenty končící vždy pouze jednou bází. Jednotlivé fragmenty vzniklé ve všech čtyřech zkumavkách se pak seřadili vedle sebe na gelové elektroforéze, aby mohla být odečtena konečná sekvence.

Sangerovo sekvenování má však své limity. Za horní limit je považováno vlákno kolem délky 1000 páru bází. Jak postupují sekvenační reakce, pravděpodobnost ukončení řetězce stoupá. S prodlužující se sekvencí řetězce, nukleotidy lišící se o jeden pár bází se hůře oddělují podle velikosti. Aby bylo možné přečíst dlouhou sekvenci DNA, je nutné ji číst po sekcích, což znamená velmi časově náročnou práci při čtení sérií překrývajících se sekvencí. Na prvních 15-40 bazích není dobrá kvalita sekvence, protože se tam váže primer. Kvalita zároveň klesá mezi 700-900 páry bází. Také může způsobit biologické zkreslení, pokud byl sekvenovaný fragment DNA dříve klonován, část sekvence klonovacího vektoru se může dostat i do konečné sekvence. Další nevýhodou použití této metody může být omezená schopnost manipulovat a analyzovat alelové frekvence. Často pouhé objevení heterozygotního SNP v PCR produktu je těžké, natož jakýchkoliv jiných bází, které nejsou zastoupeny v poměru 1:1. Ovšem největší nevýhodou Sangerovy metody jsou vysoké náklady. Průměrné výzkumné laboratoře nemohou pokrýt náklady na sekvenování projektů, které přesahují megabázové měřítko. Proto jsou vzorky často posílány do velkých genových center (Men et al. 2008).

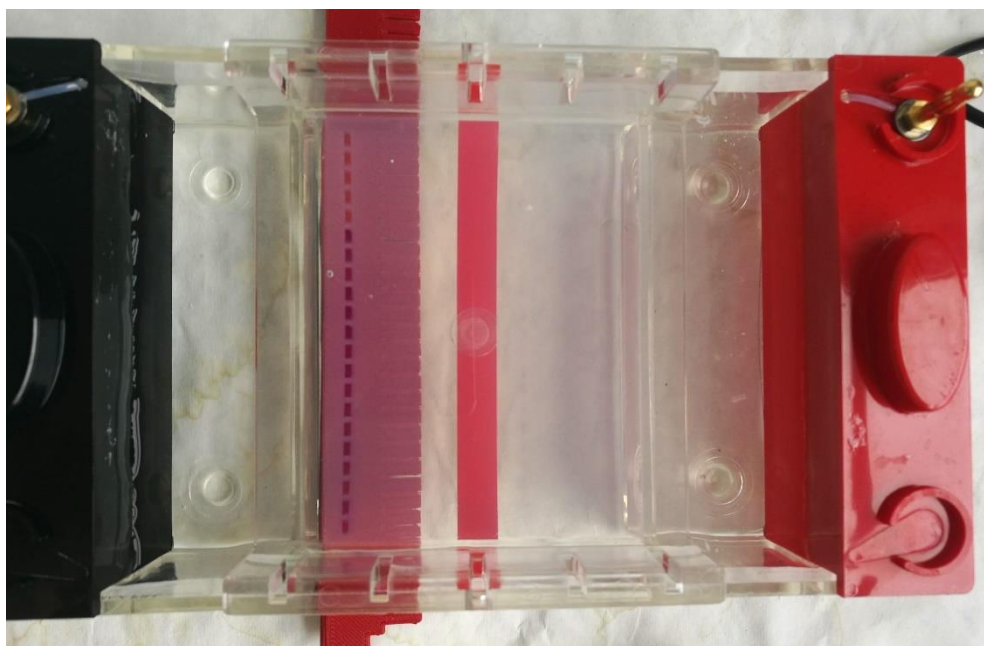
Proto je mimo jiné celkem slibné řešení spojit Sangerovu metodu se sekvenováním nové generace. Samotné Sangerovo sekvenování bude však vždy skvělým řešením pro projekty, kde bude stačit přečíst jen pár krátkých sekcí DNA pro konkrétní projekt.

3.3.2 Analýza fragmentů DNA

Analýza fragmentů genetické informace může být prováděna různými způsoby. Jedná se o konvenční přístup k analýze DNA. Nejčastěji se fragmenty analyzují poté, co dojde k namnožení zájmových úseků pomocí PCR. Fragmenty jsou často fluorescenčně značeny, a provede se tzv. elektroforetická separace.

Elektroforetická separace je metoda, při které dochází k separaci látek na základě jejich odlišné pohyblivosti (rychlosti migrace) ve stejnosměrném elektrickém poli. Protože jsou nukleové kyseliny nabitý záporným nábojem (nesou fosfátovou skupinu), tak se pohybují ke kladně nabitě elektrodě – anodě. Ve většině odborných článků se setkáváme s použitím gelové nebo kapilární elektroforetické separace. Při použití gelové elektroforézy je jako médium použit některý z gelů (agarózový, polyakrylamidový). Dlouhé fragmenty DNA se pohybují k opačně nabitě elektrodě pomaleji, naopak krátké fragmenty se pohybují rychleji a za stejný čas urazí delší vzdálenost. Na obrázku č. 3 je zobrazena vana s gelem horizontální elektroforézy.

Gelová elektroforéza se většinou používá k určení toho, zda se mikrosatelitové primery amplifikují či nikoliv. Dále lze díky této metodě určit, zda jsou primery polymorfní, případně je možné tuto metodu použít ke zjištění kvality DNA (Koolman a Röhm, 2012).



Obrázek 3 - Horizontální elektroforéza

Více sofistikovanou metodou je kapilární elektroforéza neboli separace jednotlivých fragmentů dle Sangerova. Při ní se nepoužívá pouze DNA, ale před samotnou separací je nutné zájmové úseky genetické informace amplifikovat pomocí PCR. Tato metoda funguje na principu pohybu elektricky nabitých látek ve vodivém médiu. Kapiláry jsou naplněny elektrolytem či gelem. Existuje více metod kapilární elektroforézy v závislosti na použitém přístroji, médiu a separovaných částicích (Wallingford a Ewing 1989).

Tato metoda se také často používá pro Sangerovo sekvenování. výstupem pak mohou být délky jednotlivých fragmentů DNA v párech bazí. K vyhodnocení dat (genotypizaci) po sekvenování DNA, se používají nejrůznější softwary např. GeneMarker® (SOFTGENETICS) nebo GeneMapper® (ThermoFisher SCIENTIFIC).

3.4 Sekvenování nové generace – *Next Generation Sequencing* (NGS)

Mezi lety 2000 a 2010 došlo k rychlému vývoji biochemických a bioinformačních technologií. Tento fenomén znamenal rychlý růst v DNA sekvenování. Práce, která dříve se Sangerovým sekvenováním trvala roky, teď může být hotová během několika týdnů. Výhodou těchto platforem je určení sekvenačních dat z amplifikovaných jednovlákných DNA fragmentů bez použití klonování. Další výhodou je redukce sekvenačních chyb. Naopak jednou z výzev, které nové platformy přináší, může být potřeba vývoje nových softwarů a více účinných algoritmů (Ansorge 2009).

V následujících podkapitolách popisují některé hojně využívané platformy pro sekvenování DNA. Zejména se zaměřuji na platformu Illumina, kterou jsem využila i v metodice této diplomové práce.

Platformy pro masivně paralelní sekvenování DNA:

- Roche/454 FLX
- Illumina/Solexa Genome Analyzer
- Applied Biosystems SOLiD™ System
- Helicos Heliscope™
- Pacific Biosciences instruments SMRT
- Ion Torrent
- Oxford Nanopore

Každá z platforem je svým způsobem unikátní, ať už použitím jiného typu přístroje či díky jinému principu sekvenace, nicméně při sekvenování nové generace je používán typický experimentální postup při použití NGS viz níže dle knihy od Wang (2016).

1. Zkoumaná sekvence DNA/RNA

Nejprve je důležité mít zájmový vzorek molekul DNA či RNA, který je potřeba vyextrahovat ze zájmových organismů, což se ve většině případů provádí izolací DNA nebo RNA.

2. Fragmentace

Většina přístrojů pro NGS neumí přímo pracovat s tak velikými vzorky molekul, a proto je potřeba molekuly rozdělit na menší fragmenty. Toho lze docílit pomocí

různých technik, nejčastěji je využívána nebulizace, sonikace, akustický stříh nebo enzymatické ošetření.

3. Selekcce dle velikostí

Z fragmentů se poté vybírají úseky o určitém velikostním rozmezí, velikost daného rozmezí je určena konkrétním experimentem.

4. Navázání adapterů

Klíčovým krokem při přípravě DNA/RNA knihovny je navázání adapterů ke koncům fragmentů. U většiny experimentů s RNA, je reverzní transkripce RNA přepsána na cDNA, a to ještě před navázáním adapterů. Adaptéry jsou uměle vytvořené sekvence obsahující mnoho komponentů, které jsou důležité pro určité části sekvenačních procesů (univerzální primerové sekvence, PCR primery, kotevní sekvence, indexové sekvence). Pro každou z NGS platforem existují specifické adaptéry.

5. Imobilizace sekvencí na pevné podpoře

K upevnění adapterů a sekvencí DNA/RNA se používaná podpora liší v závislosti na použitém přístroji. Někdy se používají agarózové kuličky, jindy např. flowcely neboli reakční komůrky. Tento krok je unikátní pro jednotlivé NGS technologie.

6. Sekvenování za použití NGS technologií

K samotné sekvenaci DNA knihoven různé platformy využívají různé přístupy a získávají sekvenační signály různého původu. Využívá se kupříkladu optických signálů, které jsou běžně založeny na přímé světelné emisi (454/Roche platforma), nebo fluorescenci vytvořenou z modifikovaných nukleotidů, které na sobě mají fluorescenční značení (Illumina a Pacific Biosciences systems). Dále se používají fyzikálně-chemické signály, které se měří z fyzikální či chemické aktivity spojené se sekvenačním procesem, jako je uvolnění H⁺ a kodominantní pH změna při použití platformy Ion Torrent systému, nebo narušení elektrického pole jako při použití Nanoporového sekvenování.

7. Čtení sekvencí

Zpracování takového objemu optických či fyzikálně-chemických signálů vede k dedukci jaká je tedy sekvence DNA fragmentů. Pro datový výstup a uložení jsou syrová data ve formě signálů obvykle uložena ve specifických formátech pro dané platformy. Přečtené sekvence jsou obvykle převedeny do více univerzálních formátů.

Běžně může být sekvence DNA čtena pouze z jednoho konce DNA vzoru (*single-read sequencing*), ale DNA lze také číst z obou konců DNA fragmentu (*pair-end*

sequencing). Mimo zdvojnásobení konečného počtu přečtených sekvencí je čtení z obou párových konců DNA (párové ready) výhodnější v usnadnění následného přiřazení k referenčnímu genomu nebo sestavení genomu.

8. Uspořádání přečtených sekvencí

Díky tomu, že DNA fragmenty jsou obvykle vybrány dle velikosti (bp), je jejich přibližná délka známá, výsledná párová čtení a jejich vzájemná vzdálenost poskytují další informace o tom, jak zarovnat čtení s referenčním genomem nebo jak je sestavit do nového genomu.

Většina současných platform NGS umožňuje párové sekvenování. Z výše uvedeného obecného experimentálního pracovního postupu NGS je zřejmé, že kromě vynalézavosti ve vývoji nových sekvenačních schémat je úspěch NGS technologií při dosahování extrémně vysokého výkonu při současné detekci milionů molekul DNA, způsoben také moderním inženýrstvím a výpočetní technikou. Vývoj moderní optiky, mikroskopie a zobrazovacích technologií umožňuje sledování sekvenačních reakcí ve vysokém rozlišení, vysoké spolehlivosti a taktéž i rychlosti. Některé platformy NGS také spoléhají na desetiletí pokroku v polovodičovém průmyslu nebo na novější, ale rychlý vývoj v technologii nanopórů (jako jsou platformy Ion Torrent a Nanopore). Vysoce výkonná výpočetní technika umožňuje zpracovávat přívaly signálů zaznamenané z milionů těchto procesů.

3.4.1 Platforma Illumina (Solexa)

Společnost Solexa vznikla v roce 1998. Díky sérii kreativních diskuzí došlo k podnícení nápadů týkajících se použití klonálních polí a masivně paralelního sekvenování krátkých čtení za použití pevné fáze a sekvenování pomocí reverzibilních terminátorů – dnes označované jako sekvenování za pomoci syntézy (SBS). Toto se stalo základem nového přístupu k sekvenování DNA. Společnost se velmi rychle vyvíjela a v roce 2006 představila genomový analyzátor, který zvládl osekvenovat 1 Gb dat najednou. Společnost Solexa byla koupena společností Illumina na začátku roku 2007 (Illumina 2021).

Jedná se o jednu z nejrozšířenějších platform na světě. Hlavní výhodou Illuminy je cena za sekvenaci 1 milionu bází (2–15 centů USD) a velký zisk dat za jeden běh (> 3 miliardy bp). I přes veškeré výhody použití této platformy, má Illumina jako jiné platformy založené na detekci fluorescence, i řadu nevýhod, které snižují její přesnost,

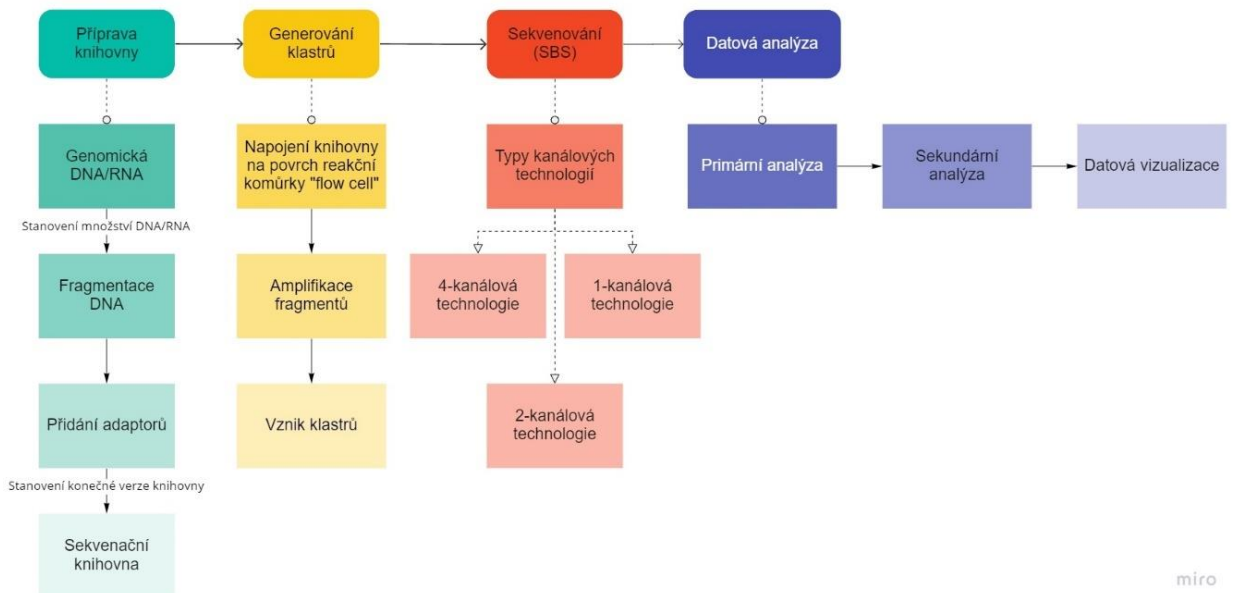
zejména nedostatečnou excitaci fluorescenčního barviva, výrazné spektrální překrytí nebo neefektivní sběr emisních signálů (Metzker 2005).

Délka čtení (50-300 bp) závisí na použitém přístroji a zvolené délce čtení, stejně tak jako čas běhu. V současné době Illumina nabízí 5 typů sekvenátoru viditelných na obrázku č. 4, lišící se svým zaměřením.



Obrázek 4 - Typy sekvenačních přístrojů nabízených společností Illumina, zdroj: [Illumina](#)

Kompletní pracovní postup při použití této platformy je poměrně komplexní a složitý. Proto jsem vytvořila jednoduchou grafickou vizualizaci viz obrázek č. 5.



Obrázek 5 - Grafické znázornění pracovního postupu Illumina, zdroj: [Illumina](#), upraveno

Níže se pokusím vysvětlit jednotlivé kroky pracovního postupu dle společnosti Illumina.

1. Příprava knihovny

Nejprve je nutné připravit knihovnu, ve které jsou informace, které chceme analyzovat. Může se jednat o DNA, RNA apod. Pro přípravu knihovny je tedy nutné získat izolovanou záměrovou jednotku informace v případě této práce nDNA. Tato izolovaná DNA musí být v dostatečné kvalitě a o dostatečné

koncentraci. Poté se může přistoupit i k namnožení zájmového úseku pomocí PCR. Dále se pomocí předem vybraného kitu vytvoří a optimalizuje knihovna. Vhodný kit je vybrán dle několika parametrů. Důležitá je oblast výzkumu, použití dat, zájmový druh, metoda analýzy a konkrétní typ přístroje pro sekvenování. Pro uvedení příkladu v této práci nám jde o oblast lesnické genetiky, použití pro populační studie, metoda za použití mikrosatelitových ampliconů, uplatněná na rostlinném druhu smrku ztepilém a typem přístroje je Miseq.

Vytvoření knihovny je závislé na použitém kitu, a proto se jednotlivé knihovny od sebe mohou lišit. Ve většině případů je nejprve zájmová zhodnocená DNA fragmentována na menší části. Poté jsou k ní přidány adaptéry. Přidání adaptérů je jedním z klíčových částí celého procesu sekvenování, protože na adaptérech jsou závislé následné kroky. Různé typy adaptérů se přidávají k oběma koncům fragmentů DNA. Na úplných koncích jednotlivých fragmentů jsou připojeny adaptéry P7 a P5, které interagují s oligonukleotidy na povrchu reakční komůrky (flow cell), což je zásadní pro vytváření klastrů. Dalšími adaptéry jsou indexy 1 a 2, které jsou důležité pro multiplexování. Indexy jsou unikátní kódy, díky kterým je možné smísit více vzorků dohromady a stále být schopný určit původ sekvence. Posledními adaptéry jsou Rd2 SP a Rd1 SP, tyto adaptéry umožňují navázání primerů tak, aby bylo možné čtení z obou konců fragmentů. Posledním krokem po připojení všech adaptérů je sestavení kompletní knihovny (Bronner a Quail 2019).

2. Generování klastrů

Generování klastrů probíhá už na samotné reakční komůrce – flowcell. Reakční komůrka je skleněná destička s fluidními kanálky nebo žlábků, ve kterých se odehrávají veškeré sekvenační chemické procesy. Každý typ přístroje pro sekvenaci používá unikátní reakční komůrku. Existují varianty komůrek, které mají na povrchu vzorování, díky němuž vznikají předem definované velikosti klastrů. Další variantou jsou komůrky bez vzorování, na nichž vznikají náhodně rozmístěné klastry.

Samotný proces generování klastrů znamená klonování každého fragmentu z knihovny do tisíců identických kopií. Celý proces vzniku klastrů lze rozdělit na následující kroky:

a) Knihovny hybridizují k reakční komůrce

Reakční komůrka je pokryta oligonukleotidy, na které se díky adaptérům připojí fragmenty DNA.

b) Amplifikace

Každý fragment je následně amplifikován neboli rozmnožen. Jednovláčenné molekuly se ohýbají k povrchu komůrky a díky oligonukleotidům a adaptérům dojde k připojení i druhého konce fragmentu. Tímto se vytvoří tzv. most. Enzym polymeráza syntetizuje reverzní vlákno, čímž vznikne kopie fragmentu. Tento amplifikační proces se opakuje několikrát.

c) Vznik klastrů

Jakmile je generování klastrů dokončeno, jsou na povrchu reakční komůrky tisíce identických kopií fragmentů. Optimalizované klastrování určuje kvalitu a konečný výtěžek dat. Aby bylo docíleno optimální klastrové hustoty, je nutné provést správné zhodnocení knihovny už před samotným klastrováním.

3. Sekvenování

Illumina sekvenování je známé pod pojmem sekvenování dle syntézy (SBS). Během tohoto sekvenování jsou k DNA vláknu přidány reagenty. Mimo jiné se přidávají i fluorescenčně značené nukleotidy. Poté co se udělají snímky reakčních komůrek, každá z bazí DNA emituje světlo o jiné intenzitě a vlnové délce. Díky tomu lze identifikovat jednotlivé nukleotidy a jejich pořadí. Typ chemických procesů je závislý na použitém sekvenačním přístroji. Existují tři typy (technologie) chemických procesů sekvenování: 4-kanálové, 2-kanálové, 1 kanálové. Níže v tabulce č. 2 je vidět, který sekvenační přístroj využívá, jaký typ procesu při sekvenování.

Tabulka 2 - Typy sekvenátorů a na nich závislé typy chemie sekvenování, zdroj: Illumina, upraveno

Sekvenátor	4-kanálový typ	2-kanálový typ	1-kanálový typ
iSeq 100			x
MiniSeq		x	
MiSeq Series	x		
NextSeq Series		x	
HiSeq Series	x		
HiSeq X	x		
NovaSeq		x	

a) 4-kanálový chemický typ sekvenování

Tento typ sekvenování používá unikátní fluorescenční značení pro každý jednotlivý nukleotid. Adenin, cytosin, guanin a tymin jsou nukleotidy, které jsou u tohoto typu sekvenování každý značen svojí unikátní barvou. Během každého sekvenačního cyklu je nukleotidová base přidána ke každému vláknu DNA z knihovny na reakční komůrce. Sekvenační přístroj poté udělá čtyři snímky, kde zachytí v každém intenzitu signálu a vlnovou délku. Každý z fluorescenčně značených nukleotidů je emitován pouze v určitém vlnovém rozmezí. Proto každý z klastrů se může objevit jen v jednom ze čtyř snímků.

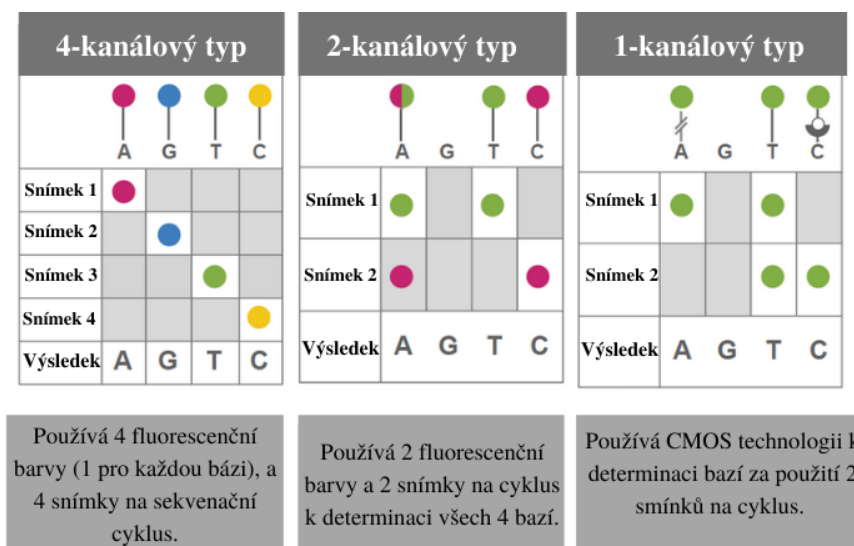
b) 2-kanálový chemický typ sekvenování

Tento typ sekvenování používá pouze dvě barvy na fluorescenční značení – červenou a zelenou. Při tomto typu sekvenování každý z nukleotidů emituje jinou intenzitu barev. Adenin emituje z 50 % červenou a z 50 % zelenou barvu, cytosin emituje ze 100 % červenou barvu, guanin je tmavý a neemituje žádnou intenzitu a tymin emituje ze 100 % zelenou barvu. V každém cyklu se dělají pouze 2 snímky pro rozeznání jednotlivých nukleotidů. Když pak porovnáme červený a zelený snímek, adenin je viditelný na obou snímcích, cytosin je pouze na červeném, tymin je pouze na zeleném a guanin na žádném snímku viditelný není. Intenzita každého nukleotidu je vynesena do korelačního diagramu, kde na ose y je intenzita zelené barvy a na ose x je intenzita

červené barvy. Na základě diagramu lze následně určit, o který ze čtyř nukleotidů se jedná.

c) 1-kanálový chemický typ sekvenování

Poslední typ sekvenování využívá CMOS technologii (komplementární metal-oxid semikonduktor) k určení jednotlivých nukleotidů. Stejně jako 2-kanálový typ používá 2 snímky na jeden cyklus sekvenování. Rozdíl je však v tom, že zároveň využívá 2 chemické procesy. Thymin je značen permanentním barvivem, které zůstává v obou chemických procesech i na obou snímcích. Adenin je značen tak, že je vidět jen na prvním snímku, během chemických procesů se jeho zbarvení změní a na druhém snímku tak není viditelný. Cytosin není vidět na prvním snímku, ale díky změnám během chemických procesů je viditelný na druhém snímku. Guanin není značen, a proto je vždy neviditelný. Tento typ sekvenování je použit v kombinaci s CMOS technologií, což je technologie umožňující zachytit intenzitu jako elektrický signál. Díky analýze signálů na obou snímcích mohou být nukleotidy správně identifikovány viz obrázek č. 6.



Obrázek 6 - Typy sekvenační chemie – Illumina, zdroj: [Illumina](#), upraveno

4. Datová analýza

a) primární analýza

Primární analýza je analýza prováděná v reálném čase, už v průběhu sekvenování, díky softwaru obsaženém v sekvenačním přístroji. Automaticky jsou generována pořadí nukleotidů a skóre kvality těchto dat. Automaticky jsou generována pořadí nukleotidů a skóre kvality těchto dat. Určování DNA sekvencí tzv. *base-calling* je provedeno už v tomto kroku datové analýzy. Výsledky tohoto určování jsou nejčastěji uloženy do standardního FASTQ formátu. Každý FASTQ formát obsahuje masivní počet readů, což jsou hodnoty sekvencí fragmentů DNA získané ze sekvenační knihovny.

b) sekundární analýza

Data o pořadí nukleotidů z primární analýzy jsou použity během sekundární analýzy, kdy jsou nově objevené sekvence přiřazeny k referenčnímu genomu. V této části analýzy dochází také ke kontrole kvality dat. Kontrola kvality dat (QC) zahrnuje zkoumání řady metrik kvality čtení sekvence. Na základě výsledků QC dat jsou sekvenační soubory předzpracovány tak, aby byla odfiltrována čtení s nízkou kvalitou a také aby byly odstraněny sekvence adaptérů nebo jiné uměle přidané sekvence (PCR primery). Následuje mapování, kdy je možné přiřadit sekvence k referenčnímu genomu. Existuje celá řada softwarů, které mohou být použity např. BaseSpace Sequence Hub, HiSeq Analysis software, CASAVA, MiSeq Reporter.

c) Terciální analýza

Při vizualizaci dat, dochází ke změně formátů dat, tak aby mohla být dále více zkoumána a interpretována.

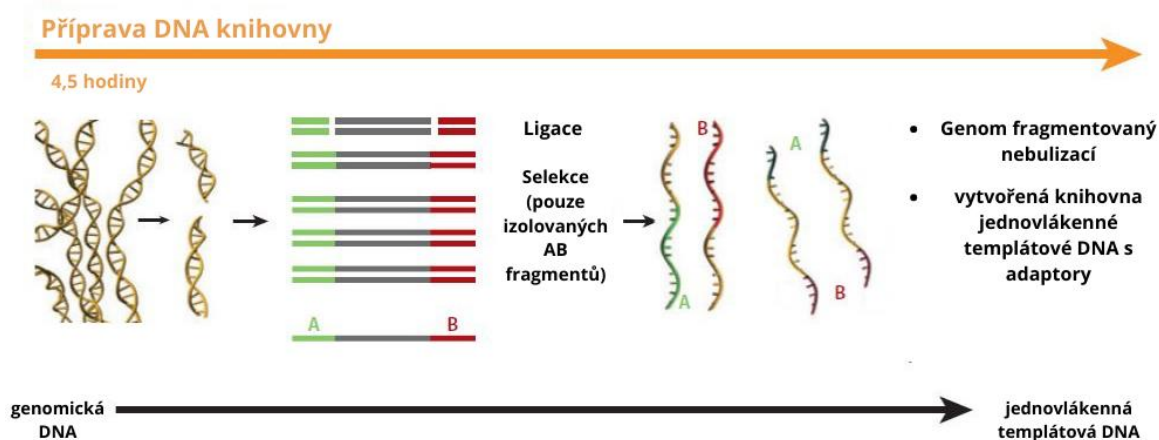
3.4.2 Roche/454 FLX

Tento sekvenátor byl poprvé komerčně představen v roce 2005. Jedná se o první systém na úrovni sekvenování nové generace. Použitá technologie při práci s tímto přístrojem se nazývá pyrosekvenování. Tento název vznikl díky pyrofosfátu, který se uvolňuje při replikaci DNA v okamžik, kdy polymeráza přidá nukleotid k původnímu vlákně DNA. Pyrosekvenování funguje na principu detekce DNA ve formě světla. Každý ze 4 typů nukleotidů tedy při syntéze nového vlákna uvolní pyrofosfát, což způsobuje chemickou reakci s enzymem zvaným luciferáza (Ansorge 2009).

Tento typ sekvenování můžeme rozdělit na 4 kroky:

- Příprava knihovny
- Emulzní PCR
- Načítání
- Pyrosekvenační reakce

Příprava knihovny znamená vytvoření velkého množství fragmentů zájmové DNA viz obrázek č. 7.



Obrázek 7 - Příprava DNA knihovny, převzato z (Mardis 2008), upraveno

Nejprve je tedy potřeba z DNA udělat fragmenty. Toho lze docílit sonikací za pomoci vystavení DNA vysokým zvukovým frekvencím, nebo pomocí nebulizace, což je protlačení DNA malými otvory. Dále se tyto fragmenty DNA spojí s adaptéry, což jsou krátké oligonukleotidy. Tyto adaptéry jsou přidány na 5' a 3' konce fragmentů DNA. Abychom denaturovali dvoušroubovici DNA, je potřeba ji vystavit vyšším teplotám.

Nakonec se hybridní molekuly DNA, amplifikují pomocí polymerázové řetězové reakce (PCR).

Emulzní PCR je jedním z typů PCR, které mohou být použité při sekvenování nové generace k replikaci DNA sekvencí. Fragmenty DNA jsou inkubovány s mikroskopickými agarózovými kuličkami, které jsou po svém obvodu pokryty oligonukleotidy, jež jsou komplementární, k již přidaným adaptérům (v případě technologie Roche/454 se jedná o 454-specifické adaptérové sekvence). Díky tomu každý z fragmentů DNA je spojen s jednou kuličkou (Mardis 2008).

Směs se naředí, tak aby bylo jisté, že na každé kuličce je pouze jedno vlákno DNA. Poté je do směsi přidán olej. Protože tyto 2 tekutiny nejsou smísitelné, vzniká emulze. Emulze vytváří micely, které jsou tak malé, že udrží pouze jednu agarózovou kuličku. Uvnitř těchto micel není pouze jedna kulička s jedním vláknem DNA, ale zároveň veškeré PCR komponenty (polymeráza, dNTPs, primery a buffer). Každá z micel je jako mikroreaktor, kde se odehrává PCR. Díky tomu vznikají tisíce fragmentů DNA amplifikovaných ke kuličkám. Nová vlákna jsou syntetizována včetně adaptérů, takže v každé micelle díky adaptérům na kuličce se znovu a znovu napojují další nově syntetizované fragmenty DNA. Reakce pokračuje v několika desítkách cyklů, dokud na jedné kuličce nevznikne díky polymeráze několik tisíců vláken stejné DNA. A toto se děje na všech agarózových kuličkách najednou. Namnožení DNA je nutné provést zejména proto, abychom mohli kamerami vidět v dostatečné kvalitě emitované světlo (Metzker 2010).

Poté je nutné rozbít emulzi a tím uvolnit kuličky, které se umístí do sekvenačního čipu s velmi malými jamkami, které jsou tak malé, že se do každé vejde pouze jedna kulička. Důležité je DNA na těchto kuličkách imobilizovat, protože reagentie protékající dovnitř a vně z jednotlivých jamek mohou s sebou jednoduše odplavit DNA, čemuž je potřeba se vyhnout. Aby mohlo dojít k samotnému pyrosekvenování používají se tyto reagentie: enzymy (sulfuryláza, luciferáza, apyráza, adenosine fosfosulfát (APS), polymezára) a primery. Počítač začne pumpovat do jamek jednotlivé modifikované nukleotidy (dATP, dGTP, dCTP, dTTP). Tyto nukleotidy pumpuje postupně, vždy pouze jeden typ nukleotidu v daném čase, přičemž po každém nukleotidu dojde k vymytí.

Mezitím se uvnitř jamek polymeráza a primer připojí k jednovlákněmu DNA fragmentu, který je stále připojen ke kuličce a začne syntéza komplementárního vlákna za použití modifikovaných nukleotidů, které dovnitř pumpuje počítač. Jakmile počítač napumpuje nukleotid komplementární k fragmentu, polymeráza ho napojí a uvolní se pyrofosfát. Když dojde k napojení správného nukleotidu, počítač detekuje světlo, protože uvolněný pyrofosfát se v kombinaci s adenosin fosfosulfátem přemění

na molekulu adenosintrifosfátu (ATP). Dále díky enzymu luciferáza a ATP dojde k emitování paprsku světla. Pokud jde o nekomplementární nukleotid, žádné světlo není detekováno, ale je nutné se tohoto nukleotidu zbavit, aby později nezkrasloval výsledek. To se dělá za použití apyrázy, což je enzym, který dokáže degradovat nukleotidy, a tak dokáže odstranit právě nezapojené nukleotidy. Počítač čte mřížkovým způsobem a dotváří si konečnou sekvenci dle intenzity detekovaných světelných paprsků a tím, načítá pořadí jednotlivých nukleotidů (Zheng et al. 2011).

3.4.3 Ion Torrent

V roce 2010 společnost DNA Electronics Ltd. vyvinula a uvedla na trh Ion Torrent Systems Inc. V té době se jednalo o první stolní sekvenátor a produkt, který využíval technologii polovodičového sekvenování, jež bylo ve své době velmi rychlé, masivně škálovatelné, jednoduché a cenově dostupné. S přibývajícím lety se tato technologie ještě více vyvíjela. Oproti konkurenčním sekvenátorům funguje Ion Torrent na úplně jiném principu. Využívá tzv. CMOS technologii (komplementární metal-oxid semikonduktor) k výrobě čipů. Obdobnou technologii využívá i Illumina u 1 kanálového typu sekvenování k určení jednotlivých nukleotidů.

Ion Torrent (nyní divize Life Technologies) vytvořila CMOS senzorový čip, který dokáže provádět sběr dat nezbytných pro sekvenační chemii. Takové čipy mohou monitorovat miliony a potenciálně i miliardy simultánních sekvenačních reakcí, pouze tím, že jsou součástí počítači podobného nástroje. Na rozdíl od všech ostatních sekvenačních nástrojů, které jsou závislé na rozsáhlé instrumentaci při sběru dat, nejčastěji prostřednictvím speciálních vědeckých CCD kamer, je Ion Torrent více automatizovaný.

Polovodičový čip je pokrytý milióny jamek, které zachycují chemické informace z DNA sekvenování a překládají je do digitálních informací nebo pořadí nukleotidů. Celý proces začíná tím, že je DNA rozbita na milióny fragmentů. Každý fragment se napojí na svou vlastní kuličku a je kopírován, dokud kompletně kuličku nepokryje. Tento proces zahrnuje milióny kuliček s milióny kopií fragmentů DNA. Každá z kuliček poté dosedne do jedné z jamek na čipu. Čip je zaplaven jedním ze čtyř DNA nukleotidů. Jakmile se k jednovlákněmu DNA připojí komplementární nukleotid, uvolní se vodíkový iont. Uvolnění H^+ mění pH roztoku v jamce. Na dně jamky je vrstva citlivá na ionty, která snímá změnu pH a přeměňuje ji na napětí. Na základě napětí lze určit který z nukleotidů

byl k vláknu připojen. Každých pár sekund se proces opakuje pokaždé s jiným nukleotidem. (Merriman et al. 2012).

3.4.4 Pacific Biosciences instruments – *Single molecule real time (SMRT™) sequencer*

SMRT sekvenování bylo vyvinuté společností Pacific Biosciences. Největší výhodou této platformy je možnost opravdu dlouhého čtení genetické informace s vysokou přesností. Jedná se o zcela odlišný koncept, kdy délka čtení dosahuje až k 15 kb za velmi krátký čas (Rhoads a Au 2015).

SMRT sekvenátor funguje na principu sekvenování jednotlivých nukleotidů v reálném čase s využitím metody sekvenování syntézou. Na rozdíl od předešlých metody nevyužívá žádný typ amplifikace. Sekvenování se odehrává na sekvenačním čipu, který obsahuje tisíce malých (s objemem v řádech zeptolitřů - 10^{-21} l) zespoda nasvícených jamek, ve kterých jsou vlnovody nultého módu (zero-mode waveguides – ZMWs). Tyto vlnovody jsou tvořené z paladia a mají průměr 100 nm. Nejsou ani schopné zachytit fotony viditelného spektra. Zachycují pouze elektromagnetický vliv. Po vytvoření DNA knihovny se do vzorku přidají primery a polymeráza. Sekvence DNA fragmentů probíhá právě díky molekule DNA polymerázy, která je imobilizována na povrchu každého vlnovodu a syntetizuje komplementární vlákno DNA. Během sekvenační reakce je DNA fragment polymerázou prodloužen o fluorescenčně značené dNTPs (stabilní nukleotidy). Každý ze 4 nukleotidů je značen jinou barvou. Čtení a určení sekvence je možné díky emitovanému světelnému poli na základě fluorescenční detekce nukleotidů. Emitované světelné signály jsou změřeny v reálném čase. Fluorescenční pulz se zastaví po vytvoření fosfodiesterové vazby, která způsobuje uvolnění fluoroforu, který difunduje pryč z ZMW. Určení sekvence je možné, díky následné inkorporaci značených nukleotidů a jejich detekci (Pareek et al. 2011).

Tento typ sekvenování má 2 možné módy. Pro vysoce přesné čtení dlouhých fragmentů (HiFi reads) se používá kruhové konsensuální sekvenování. Dalším módem je nepřetržitě dlouhé čtení, vhodné pro čtení co nejdelších fragmentů/vzorků (Eid et al., 2009).

3.5 Možnosti využití metod Sekvenování nové generace

Sekvenování nové generace nebo také *Next-generation sequencing* (NGS) je masivně paralelní sekvenování, které umožňuje ultra-vysokou průchodnost, škálovatelnost a rychlost. Technologie NGS se používají k určení pořadí nukleotidů v celých genomech nebo v cílených oblastech DNA (A, G, T, C) nebo RNA (A, G, C, U). Rozvoj NGS technologií znamenal revoluci v biologii a genetice, protože umožnil provádět širokou škálu biologických studií a aplikací na takové úrovni, která dříve nebyla možná. Produkce tak velkého množství čtení za relativně nízké náklady dělá z NGS platformou výborný nástroj s širokým spektrem využití.

Možností využití vyspělých NGS technologií je celá řada, níže uvádím pro příklad některé z nich (Metzker 2010).

- studium lidského mikrobiomu
- rychlé sekvenování celého genomu
- hloubkové sekvenování dané cílené oblasti genomu
- identifikace nových patogenů
- sekvenování nádorových vzorků pro studium vzácných somatických variant, tumorových subklonů a dalších
- analýza epigenetických faktorů (např. metylace DNA v celém genomu nebo interakce DNA-protein)
- objevování nových variant RNA, kvantifikace mRNA pro analýzu genové exprese

Použití je závislé na konkrétních vlastnostech platformy. Illumina je vhodná zejména pro resekvenování lidských genomů, protože za jeden běh sekvenování je možné získat obrovský objem vysoce kvalitních dat. Pacific Biosciences je zase vhodný pro sekvenování dlouhých čtení za krátký čas (Quail et al. 2012). Porovnání nejznámějších platform je také v tabulce č. 3.

Tabulka 3 - Porovnání současných NGS platform, převzato z: (Wang 2016) a (Metzker 2010), upraveno

Platforma/ přístroj	HiSeq 2500 (Illumina)	MiSeq (Illumina)	Ion Torrent PGM	Ion Proton	PacBio RSII
Princip	reversibilní oddělení (terminace)	reversibilní oddělení (terminace)	uvolnění protonů a změna pH	uvolnění protonů a změna pH	vlnovod s nulovým videm a sekvenování jedné molekuly
Detekce	fluorescence	fluorescence	změna pH	změna pH	fluorescence
Délka čtení [bp]	125-250	300	400	200	8500
Výstup dat na 1 run [Gb]	1000	15	nad 2	10	0,366
Náklady na Gb	průměrné	mírně drahé	drahé	velmi drahé	nejdražší
Chybovost	10^{-3}	10^{-3}	10^{-2}	10^{-2}	10^{-1}
Běžný typ chyb	záměna jednotlivého nukleotidu	záměna jednotlivého nukleotidu	inzerce a delece	inzerce a delece	inzerce a delece
Sekvenování z obou konců	ano	ano	ano	ano	ne
Požadované množství vzorku DNA [ng]	50-1000	50-1000	100-1000	100-1000	1000

3.6 Přístupy k datové analýze

3.6.1 Kontrola kvality dat

Kontrola kvality dat je při sekvenování nové generace zásadní krok, který by neměl být podceňován. Vzorky mohou selhat z mnoha důvodů. Když v molekulární biologii dojde k tomu, že nebiologické faktory způsobí změny v datech produkovaných z experimentu, může to vést k nepřesným závěrům. Proto je potřeba provádět kontrolu kvality dat, aby byly případné chyby odhaleny a napraveny co nejdříve. Existuje celá řada počítačových nástrojů určených pro kontrolu kvality dat jako například FastQC, které bylo použito i v této práci. Díky FastQC je možné si vytvořit poměrně přesný obrázek o tom, do jaké míry odpovídají získaná data obecným kvalitativním standardům (Wang 2016).

Bohužel datové výstupy ze sekvenování jsou častokrát rozptýleny do více souborů, které nejsou vždy navzájem kompatibilní a porovnatelné. To znamená, že po primární analýze jsou často výsledkem stovky či tisíce souborů, které je potřeba dále upravovat a získat z nich informace. Proto existuje nástroj zvaný MultiQC, který dokáže skenovat veškeré adresáře a analyzovat všechny nalezené protokoly. Tento nástroj vytváří jednotný report shrnující všechny informace na jednom místě. Díky tomu, že jsou vzorky v reportu vedle sebe, je možné si snáze všimnout nepatrných rozdílů u vyčnívajících vzorků.

3.6.1.1 FastQC

Ve zkratce je FastQC textový formát, obsahující sekvence každého readu spolu se skórem spolehlivosti určení pro každou z bazí. FastQC je dostupný jako samostatný Java program s grafickým uživatelským rozhraním, ale je ho také možno snadno použít v příkazovém řádku. FastQC je relativně rychlé a v rámci minut dokáže analyzovat desítky miliónů readů.

Tento program, kromě informací o počtu readů a jejich kvalitě, umí vizualizovat informace o kvalitě a obsahu jednotlivých bazí v readech, délky readů, obsah k-mer a také přítomnost nejednoznačně určených bazí, nadměrně zastoupených sekvencí a duplicit (Andrews 2010).

3.6.1.2 Kvalita bazí

Kvalita bazí udává míru věrohodnosti výsledku čtení bazí. Kvalita bazí je vyjádřena pomocí Phred skóre (Q skóre), které udává pravděpodobnost, že daná báze je určena sekvenátorem nesprávně. Matematicky je vyjádřena:

Rovnice 1 – Phred skóre

$$Q = -10 * \log_{10}(P)$$

$$P = 10_{10}^{-Q}$$

P je pravděpodobnost, že je daná báze nesprávně určena

Q Phred skóre

Pokud existuje šance 1 ze 100, že je báze špatně určena, tak hodnota kvality se rovná:

Rovnice 2 - Phred skóre příklad

$$q = -10 * \log_{10}(0,01) = 20$$

Kvalita se standardně pohybuje od 0 do 40. Typicky hodnota kvality bazí klesá v pozdějších cyklech sekvenování (Eija Korpelainen et al. 2015).

3.6.1.3 Trimming

Do češtiny by se tento proces dal přeložit jako ořezávání nebo ostříhování. Pokud je kvalita bazí na koncích readů nízká, nejjednodušším řešením je oříznout konce na danou délku nebo odříznout z každého konce určitý počet bazí. Bohužel tímto přístupem také odebereme sekvence o dobré kvalitě. Nezaleží, zda oříznutí začne ze 3' nebo 5' konce, báze je odříznuta, pokud je její kvalita pod prahem, který si definujeme (Eija Korpelainen et al. 2015).

3.6.1.4 Obsah GC

Obsah guaninu (G) a cytosinu (C) v readech by měl kopírovat normální rozdělení a soustředit se ve stejném obsahu GC jako u zdrojového genomu organismu. Neobvyklý tvar rozdělení nebo velký posun od obsahu GC ve zdrojovém genomu může značit kontaminaci knihovny.

FastQC umí porovnávat obsah GC v závislosti na pozici bazí, což by mělo být v ideální případě vyjádřeno jako vyrovnaná linie v úrovni obsahu GC zdrojového organismu. Odlišný obsah GC na určité pozici bazí znamená přítomnost nadměrně zastoupených sekvencí v knihovně.

3.6.2 Zarovnání readů dle reference

Poté, co jsou data očištěna, je dalším krokem zmapování a přiřazení readů k referenčnímu genomu. Pokud referenční genom není k dispozici, lze provést tzv. *de novo* sestavení. Většina NGS aplikací vyžaduje mapování readů k referenčnímu genomu, ještě před provedením dalších analýz (Wang 2016).

Před rozmachem NGS metod se hojně používaly algoritmy typu BLAST (*Basic Local Alignment Search Tool*) (Altschul et al. 1990), který porovnává sekvence mezi sebou a identifikuje k nim podobné sekvence z databáze. Bohužel tento typ algoritmů není vhodný pro simultánní mapování miliónů readů u NGS metod. Problémem je, že jakýkoliv genom, ze kterého NGS čte, se na mnoha místech odchyluje od referenčního genomu, kvůli polymorfismu a mutacím. Právě proto musí jakýkoliv algoritmus vytvořený pro práci s NGS daty, reagovat i na tyto odchylky. Další komplikací je to, že sekvenační chyby jsou často nerozeznatelné od sekvenačních odchylek, a tak mohou být snadno zaměněny (Wang 2016).

Zmapované výsledky vytvořené algoritmy jsou obvykle uloženy v textových formátech SAM (*Sequence Alignment Map*) (Li et al. 2009) nebo BAM (*Binary Alignment Map*). BAM ukládá stejná data jako SAM, akorát v komprimovaném binárním zobrazení.

4 Metodika

4.1 Návrh metodiky a metodické postupy analýzy genomové DNA

Koncepčně tato práce vychází z vědecké studie De Barba a kolektivu, která byla publikována v roce 2017 v časopise *Molecular Ecology*. Autoři vyvinuli metodu ke genotypizaci DNA medvěda hnědého, přičemž se jim převodem fragmentační analýzy na vysokokapacitní sekvenování (HTS) podařilo zvýšit přesnost genotypizace a snížit míru genotypizačního neúspěchu.

Cílem této práce bylo vytvořit obdobný genotypizační protokol pro smrk ztepilý a vyvinout tak novou metodiku cílicí přímo na primární sekvenční strukturu ampliconů s potenciálem eliminace dalších nevýhod fragmentační kapilární analýzy. Data získaná uvedenou inovativní technologií jsou porovnávána s výstupy získané konvenční metodou fragmentační analýzy.

Sekvenace a částečně i vývoj genotypizačního protokolu byl prováděn ve spolupráci s externí genomickou společností SEQme (www.seqme.eu).

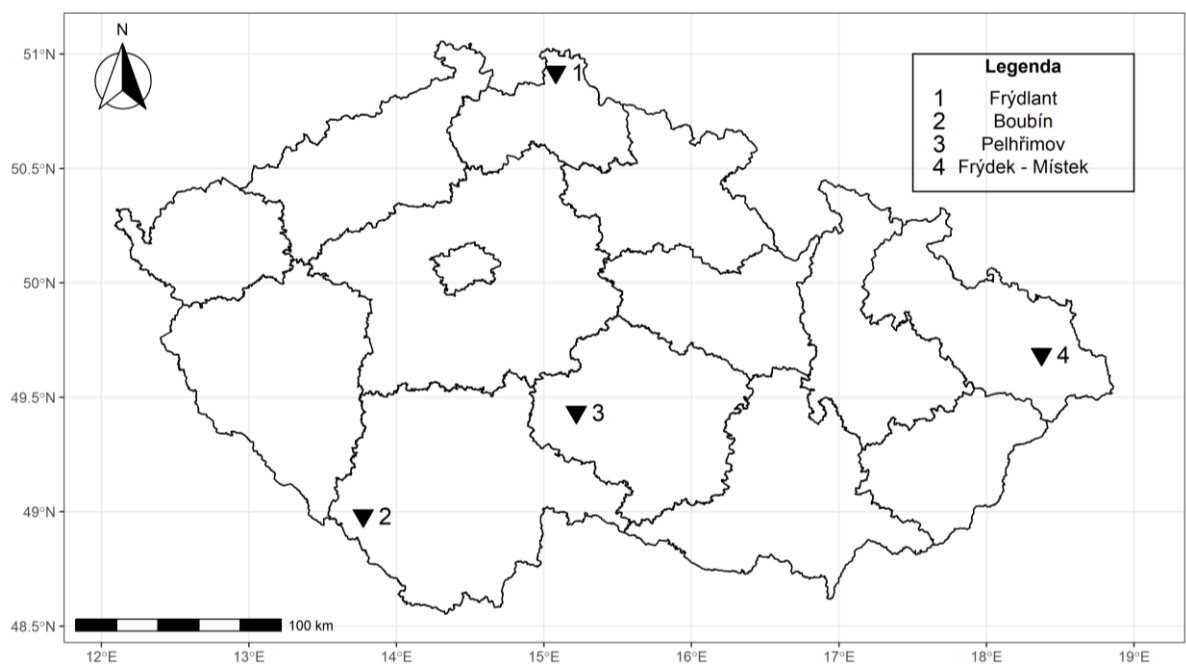
Aktivity lze rozčlenit do 5 fází:

- **Fáze 0:** Sběr a příprava vzorků, izolace DNA, ověření její kvality a koncentrace
- **Fáze 1:** Konverze mikrosatelitní analýzy pomocí fragmentační analýzy na platformu NGS – vytvoření primerů a příprava na vytvoření sekvenační knihovny
- **Fáze 2:** Konverze mikrosatelitní analýzy pomocí fragmentační analýzy na platformu NGS – příprava knihoven a sekvenování
- **Fáze 3:** Design a syntéza 36 oligonukleotidů pro PCR, STR profiling pro 24 primerů
- **Fáze 4:** Optimalizování knihoven a analýza STR lokusů

4.2 Fáze 0

4.2.1 Zájmová populace a sběr vzorků

Pro tuto diplomovou práci byla použita izolovaná DNA smrku ztepilého (*Picea abies* (L.) H. Karst.). Materiál, který byl použit k izolaci genetické informace, pocházel z odebraných jehlic 12 jedinců smrku ztepilého rostoucího na 4 geograficky vzdálených lokalitách – Frýdlant, Boubín, Frýdek-Místek a Pelhřimov. Každou lokalitu reprezentují 3 vzorky. Na obrázku č. 8 jsou zobrazeny jednotlivé lokality na mapě České republiky.



Obrázek 8 - Lokality sběru vzorků na mapě ČR

Počet vzorků byl volen především s ohledem na finanční náročnost vývoje metodiky, přičemž profil 12 jedinců již má potenciál poskytnout robustní informaci o genetické variabilitě. V tabulce č. 4 je vidět z jakých lokalit pochází jednotlivé vzorky.

Tabulka 4 - Značení vzorků a jejich rozdělení dle lokalit sběru

Lokalita	Identifikační označení vzorku FLD	Identifikace vzorku SEQme
Frýdlant	sm 3	A1
	sm 5	A2
	sm 9	A3
Boubín	sm 32	A4
	sm 33	A5
	sm 40	A6
Frýdek-Místek	sm 86	A7
	sm 88	A8
	sm 89	A9
Pelhřimov	sm 97	A10
	sm 99	A11
	sm 100	A12

4.2.2 Příprava vzorků a homogenizace

Po sběru byly vzorky převezeny v uzavíratelných sáčkách se silikagelem a zamraženy na teplotu $-80\text{ }^{\circ}\text{C}$. Silikagel pohltí případnou vlhkost a udržel tak vzorky v konstantním prostředí.

Jednotlivé vzorky jehličí v celkovém množství 80-100 mg byly vyndány z mrazícího boxu a skalpelem nakrájeny na drobné fragmenty. Před krájením každého nového vzorku byly veškeré použité nástroje vyčištěny pomocí ethanolu, aby se předešlo kontaminaci jednotlivých vzorků. Fragmenty byly vloženy do nadepsaných mikrozkušavek o objemu 2 ml. Do každé zkumavky byly přidány 2 kovové drtící kuličky ze slitiny karbidu wolframu a oceli.

Připravené vzorky byly vloženy do konzole oscilačního mlýnu MM400 (Retsch, Haan, Germany). Celou takto připravenou konzoli bylo nutné ponořit do vany s kapalným dusíkem, který vzorky šokově zmrazil na teplotu cca $-196\text{ }^{\circ}\text{C}$, tento krok zajistil zkřehnutí rostlinného materiálu, který se následně mohl lépe nadrtit. Konzole se vzorky byla umístěna zpět do oscilačního mlýnu a následně byl spuštěn 3 minutový program s frekvencí 30 kmitů za 1 sekundu. Díky drtícím kuličkám a mechanickému působení došlo k rozdrčení fragmentů jehličí na jemný prášek. Po skončení drtícího programu byla u všech vzorků okulárně zkontrolována kvalita drcení. Pokud vzorek neobsahoval hrubé části, byla homogenizace úspěšná, v opačném případě bylo nutné přistoupit k opakování procesu drcení.

4.2.3 Extrakce a purifikace nDNA

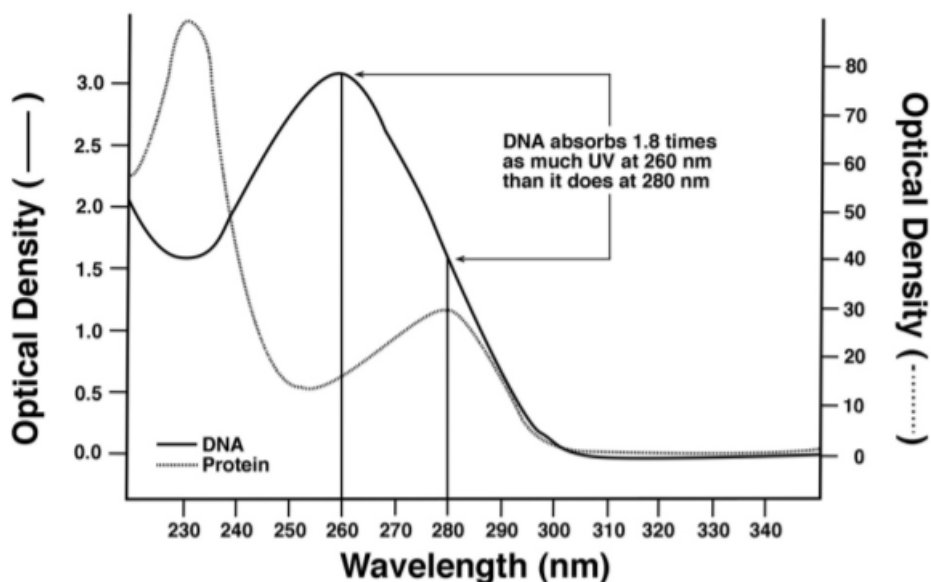
Pro izolaci nDNA by použit izolační kit DNeasy Plant Mini Kit (Qiagen, Hilden, Germany). Při izolaci byly dodrženy instrukce popsané v protokolu od výrobce s jednou malou modifikací, a to s prodloužením inkubace vzorků při lyzační fázi z 10 na 60 minut.

4.2.4 Stanovení kvality a koncentrace DNA

Za účelem ověření kvality pro další sekvenační analýzy, byla změřena koncentrace a čistota DNA. Pro společnost SEQme, se kterou jsme na analýze spolupracovali byly požadované vstupní parametry vzorků následovné:

- minimální koncentrace: 25 ng/μl
- minimální čistota vzorku: 1,8-2,0

Měření a vyhodnocení bylo provedeno na přístroji NanoDrop (Thermo Fisher SCIENTIFIC). Tento přístroj funguje na principu spektrometrického vyhodnocení z 1 μl kapky roztoku DNA/vzorek. Jako referenční vzorek byl použit eluční buffer AE z izolačního kitu od společnosti QIAGEN. Vzorek je považován za čistý, pokud poměrový koeficient absorbance světla při ozáření o vlnových délkách 260 nm a 280 nm je v rozmezí od 1,7 do 1,9. V takovém případě můžeme mluvit o ideální stavu, kdy koncentrace izolované DNA není znečištěna jinými rezidui (proteiny, RNA apod.). Čím více se stanovená hodnota poměrového koeficientu liší od hodnoty 1,8, tím více je pravděpodobné znečištění vzorku. Dále byla sledována spektrální křivka a její průběh. V grafu č. 1 je vidět optimální průběh této křivky u DNA a u proteinu.



Graf 1 – Spektrální křivka DNA a proteinu (osa x znázorňuje vlnovou délku, osa y znázorňuje optickou hustotu), převzato z: (Rothman 2010)

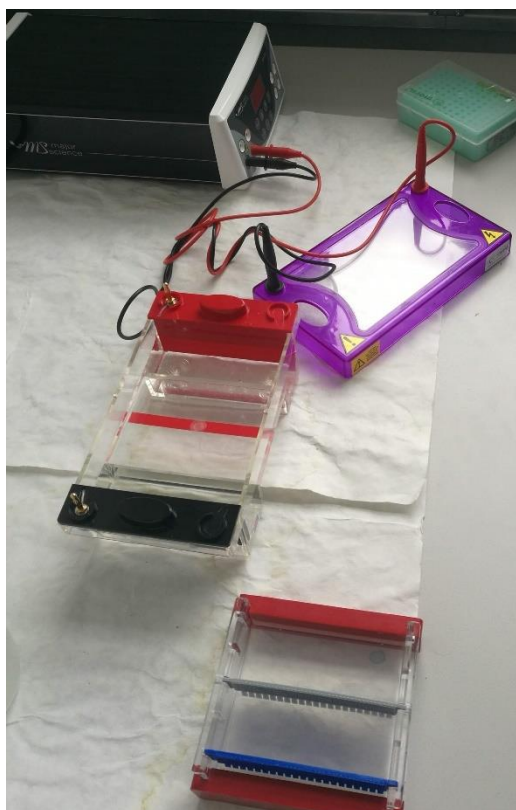
Výsledky měření jsou uvedeny v kapitole výsledky v tabulce č. 17. Po změření byly vzorky uchovány v mrazícím boxu při teplotě -80 °C.

4.2.5 Kontrola kvality DNA

Pro kontrolu kvality DNA, jež byla změřena spektrofotometricky, byla použita metoda horizontální gelové elektroforézy. Princip této metody spočívá v tom, že látky putují různou rychlostí v gelu ovlivněném elektrickým polem. Jako médium byl použit agarózový gel o koncentraci 0,73 %. Izolovaná DNA byla naředěna ultra čistou vodou na koncentraci 10 ng/μl.

Zdroj elektroforetické vany byl nastaven na napětí 70-110 V, proud 250 mA a 35minutový program. Na obrázku č. 9 je zobrazen použitý přístroj. Do krajních jamek byl napipetován velikostní standard.

Snímek gelu je k nahlédnutí kapitole Výsledky.



Obrázek 9 – Kompletní soustava pro horizontální gelovou elektroforézu

4.3 Fáze 1

V první fázi spolupráce se společností SEQme proběhlo hledání repetitivních mikrosatelitových oblastí v referenčním genomu smrku dostupném přes webové rozhraní congenie.org (Nystedt et al. 2013). Celkem bylo navrženo, syntetizováno a otestováno 93 primerových párů pro amplifikaci mikrosatelitových lokusů. Primery nesly názvy PA_1 až PA_93. Každý primerový pár byl vytvořen se specifickou značkou (tzv. adaptérem) umožňující hromadnou PCR amplifikaci (tzv. multiplexing).

Amplifikace perspektivních primerových sekvencí byla ověřována na dodaném vzorku sm 40.

4.3.1 Navržení a vytvoření primerů

Teplota fáze annealingu (nasedání oligonukleotidů) pro PCR – polymerázovou řetězovou reakci (50°C, 55°C, 60°C) a koncentrace Mg²⁺ iontů byla optimalizována v singleplexové reakci u vzorku sm 40. Jako optimální byla zvolena annealingová teplota 55°C a 2,5 mM koncentrace Mg²⁺ iontů. Složení PCR reakce a teplotní profil jsou uvedeny v tabulkách č. 5 a 6.

Tabulka 5 – Složení PCR reakce

Reakce	μl
PCR buffer w/o MgCl ₂ (10x)	2,5
MgCl ₂ (25 mM)	2,5
dNTP's (10 mM)	0,5
DMSO	1,3
BSA 20 mg/ml	0,7
FastStart Taq Roche (5 U/μl)	0,4
DNA (10 ng/μl)	4
primery (konečná koncentrace)	á 300 nM
H ₂ O	do celkového objemu 25 μl

Tabulka 6 – Program pro průběh PCR reakce

Program	Doba trvání	Teplota	Počet opakování
1. iniciální denaturace	5 min	95 °C	1x
2. denaturace	30 s	95 °C	35x
3. annealing	30 s	55 °C	35x
4. extenze	45 s	72 °C	35x
5. finální extenze	7 min	72 °C	1x
6. uchování	∞	4 °C	1x

PCR reakce byly provedeny pro každý z primerů zvlášť. Jednalo se tedy o 93 single reakcí, provedených na referenčním vzorku sm 40. Produkty jednotlivých amplifikací byly následně vizualizovány na agarózovém gelu. Jelikož se amplifikace nezdařila u 27 primerových párů, bylo do dalších fází zařazeno pouze 66 primerových párů, jejichž sekvence jsou uvedeny v příloze č. 1.

4.3.2 Vytvoření a optimalizování multiplexů

Primery úspěšné v prvotním singleplexovém testování byly dále kompletovány do směsi na základě 3 navržených strategií poolingů neboli smíchání primerů do multiplexů.

1. **Ekvimolární pooling** nezohledňující amplifikační účinnost.

Multiplex byl vytvořen na základě směsi stejného látkového množství od každého z 66 amplifikujících primerů – sekvence těchto primerů jsou uvedeny v příloze 1.

2. **Pooling zohledňující amplifikační účinnost** (primery s nejnižší účinností byly v poolu zastoupeny s 3x větší koncentrací, primery se střední účinností s 2x větší koncentrací než primery s největší účinností).

3. **Ekvimolární pooling primerových párů vždy v rámci skupiny s podobnou amplifikační účinností** (pro jeden vzorek tedy použity 3 primerové pooly – low, medium, high).

Při použití vysoce multiplexní PCR je vždy výzvou celková optimalizace reakčních podmínek, zejména koncentrace primerů v dané směsi. Proto byly testovány PCR reakce se 3 výše uvedenými strategiemi poolingů v různých koncentracích pro danou reakci. Jako výchozí stav byla zvolena stejná koncentrace primerů jako pro singleplexovou reakci – 300 nM zastoupení každého jednotlivého primeru z 66 primerových párů. Tato koncentrace bohužel vykazovala velmi silné inhibiční účinky pro samotnou PCR s převahou dimerů vznikajících mezi primery.

4.3.3 Analýza dat a vytvoření sekvenačních knihoven

Následně byly testovány pooly s finální koncentrací jednotlivých primerů 100 nM respektive 60 nM. Jako nejlépe amplifikovatelná byla reakce s koncentrací 100 nM. S touto koncentrací bylo poté sekvenováno na přístroji Miseq (Illumina) v uspořádání 250 bází čtení z obou konců, celkem vzniklo 5 knihoven – ekvimolární pooling, pooling s korekcí primerů a primery, jež byly rozděleny do 3 poolů dle amplifikační účinnosti (low, medium, high).

Kontrola kvality dat, jež je při sekvenování nové generace zásadním krokem, byla provedena pomocí nástrojů FastQC (verze 0.11.5) (Andrews 2010) a MultiQC (verze 1.6) (Ewels 2018). Hrubá analýza sekvencí byla provedena pomocí dvou nezávislých postupů.

- a) Ready R1 (forward) a R2 (reverse) byly demultiplexovány pomocí programu `fastx_barcode_splitter.pl` (Gordon 2009) pomocí sekvencí primerů jakožto barcodů. Poté bylo vypočítáno jednotlivé zastoupení jednotlivých primerů z celkového počtu readů v knihovně. Výsledky jsou uvedeny v tabulce č. 19 v kapitole Výsledky.
- b) Ready R1 (forward) a R2 (reverse) byly spojeny pomocí programu `fastq-join` a následně rozděleny na jednotlivé soubory dle primerů s využitím nástroje `grep`. Výsledky jsou uvedeny v tabulce č. 20 v kapitole Výsledky.

4.4 Fáze 2

V další fázi práce bylo hlavním cílem rozdělit amplifikující primery z fáze 1 do poolů (knihoven) tak, aby byla amplifikace mikrosatelitových fragmentů a jejich následná sekvenace efektivnější a rovnoměrnější.

4.4.1 Příprava knihoven na sekvenování

Dle počtu readů osekvenovaných v rámci ekvimolárního poolu (knihovny) ve fázi 1 (od primerových párů s nejvyšším počtem readů), bylo vybráno 5 primerových poolů. Z těchto poolů bylo za pomoci dvoustupňové PCR vytvořeno 5 knihoven. Pro oba kroky PCR byly pro všechny knihovny zvoleny stejné podmínky vycházející z protokolu pro Type-it Microsatellite PCR Kit (Qiagen) viz tabulky č. 7 až 8. První PCR byla u každé z knihoven provedena zejména za účelem namnožení zájmových úseků DNA. Produkt 1. PCR byl pak použit pro druhou PCR.

Tabulka 7 - Složení 1. PCR reakce

1. PCR	
Reakce:	μl
2x Type-it Multiplex PCR Master Mix	12,5
Q-Solution, 5x	2,5
DNA (10 ng/μl)	4
H2O	3,5
10x primer mix	2,5
Dohromady	25,0

Tabulka 8 - Program průběhu 1. PCR

1. PCR			
Program	Doba trvání	Teplota	Počet opakování
1. iniciální denaturace	5 min	95 °C	1x
2. denaturace	30 s	95 °C	25x
3. annealing	90 s	57 °C	25x
4. extenze	30 s	72 °C	25x
5. finální extenze	90 min	60 °C	1x
6. uchování	∞	4 °C	1x

Při druhé PCR byly přidány tzv. indexy, které umožňují smíchání jednotlivých primerů dohromady. Díky indexům je možné jednotlivé primery odlišit i v rámci multiplexu. Složení 2 PCR reakce a program jejího průběhu je uvedený v tabulce č. 9 a 10.

Tabulka 9 – Složení 2. PCR reakce

2. PCR	
Reakce:	μl
2x Type-it Multiplex PCR Master Mix	12,5
Q-Solution, 5x	2,5
Index 7 Fw primer	3,3
Index 7 Rv primer	3,3
templát 1. PCR	3,3
Dohromady	25,0

Tabulka 10 - Program průběhu 2. PCR

2. PCR			
Program	Doba trvání	Teplota	Počet opakování
1. iniciální denaturace	5 min	95 °C	1x
2. denaturace	30 s	95 °C	8x
3. annealing	30 s	55 °C	8x
4. extenze	45 s	45 °C	8x
5. finální extenze	5 min	72 °C	1x
6. uchování	∞	4 °C	1x

Knihovny byly sestaveny na základě více strategií o různých počtech primerů:

- a) **dle amplifikační účinnosti** – knihovna č. 4 (25 primerů), knihovna č. 5 (66 primerů)
- b) **ekvimolárně** – knihovna č. 2 (25 primerů), knihovna č. 3 (66 primerů)
- c) **ve dvou fázích** – knihovna č. 1 nejprve v single reakci (11 primerů po 6 poolech – **dle amplifikační účinnosti**), poté smíchání (66 primerů – **ekvimolárně**)

V tabulce č. 11 je vidět složení první knihovny. V prvním kroku bylo provedeno 6 nezávislých (pools A-F) PCR reakcí s 11 páry primerů (výsledná koncentrace každého primeru v reakci byla 200 nM). Druhý krok byl proveden tak, že pools A-F byly smíchány ekvimolárně na základě koncentrace DNA a byla provedena PCR. Použité indexy při druhé PCR byly N710/S521.

**Tabulka 11 - Skladba primerů 1. knihovny
1. knihovna 6x_a11**

pool A	pool B	pool C	pool D	pool E	pool F
PA_2	PA_45	PA_10	PA_1	PA_24	PA_31
PA_25	PA_47	PA_19	PA_18	PA_28	PA_37
PA_27	PA_48	PA_38	PA_22	PA_30	PA_43
PA_3	PA_53	PA_58	PA_29	PA_32	PA_5
PA_64	PA_60	PA_59	PA_34	PA_33	PA_62
PA_65	PA_61	PA_66	PA_41	PA_35	PA_77
PA_67	PA_71	PA_85	PA_42	PA_36	PA_83
PA_69	PA_72	PA_90	PA_44	PA_51	PA_89
PA_75	PA_79	PA_91	PA_70	PA_54	PA_9
PA_82	PA_84	PA_95	PA_73	PA_74	PA_92
PA_88	PA_87	PA_97	PA_96	PA_94	PA_93

V tabulce č. 12 je zobrazeno složení 2. knihovny, kde byla první PCR spuštěna s každým z předem vybraných potenciálně nejlepších 25 primerů o koncentraci 200 nM. V druhé PCR byly primery značeny indexy N711/S522.

**Tabulka 12 - Skladba primerů 2. knihovny
2. knihovna 25x_equi**

PA_10	PA_64	PA_88
PA_2	PA_65	PA_91
PA_25	PA_67	PA_95
PA_27	PA_69	
PA_3	PA_71	
PA_45	PA_72	
PA_47	PA_75	
PA_48	PA_79	
PA_53	PA_82	
PA_60	PA_84	
PA_61	PA_87	

V tabulce č. 13 je složení knihovny č. 3, kdy proběhla PCR opět s výslednou koncentrací každého z 66 primerů v reakci 200 nM. A v druhé PCR byly primery značeny indexy N712/UDI6.

**Tabulka 13 - Skladba primerů 3. knihovny
3. knihovna 66_equi**

PA_1	PA_3	PA_42	PA_59	PA_72	PA_88
PA_10	PA_30	PA_43	PA_60	PA_73	PA_89
PA_18	PA_31	PA_44	PA_61	PA_74	PA_9
PA_19	PA_32	PA_45	PA_62	PA_75	PA_90
PA_2	PA_33	PA_47	PA_64	PA_77	PA_91
PA_22	PA_34	PA_48	PA_65	PA_79	PA_92
PA_24	PA_35	PA_5	PA_66	PA_82	PA_93
PA_25	PA_36	PA_51	PA_67	PA_83	PA_94
PA_27	PA_37	PA_53	PA_69	PA_84	PA_95
PA_28	PA_38	PA_54	PA_70	PA_85	PA_96
PA_29	PA_41	PA_58	PA_71	PA_87	PA_97

Při první PCR byla u knihovny č. 4 použita výsledná koncentrace každého primeru v reakci v rozsahu 100-300 nM dle amplifikační účinnosti. Vybrané primery jsou v tabulce č. 14. V druhé PCR byly primery značeny indexy N714/UDI7.

**Tabulka 14 - Skladba primerů 4. knihovny
4. knihovna 25x_corect**

PA_10	PA_64	PA_88
PA_2	PA_65	PA_91
PA_25	PA_67	PA_95
PA_27	PA_69	
PA_3	PA_71	
PA_45	PA_72	
PA_47	PA_75	
PA_48	PA_79	
PA_53	PA_82	
PA_60	PA_84	
PA_61	PA_87	

V tabulce č. 15 je knihovna, kdy byla pro první PCR použita výsledná koncentrace každého primeru v rozsahu 100-300 nM dle amplifikační účinnosti. Pro druhou PCR byly primery značeny indexy N715/UDI10.

**Tabulka 15 - Skladba primerů 5. knihovny
5. knihovna 66x_corect**

PA_1	PA_3	PA_42	PA_59	PA_72	PA_88
PA_10	PA_30	PA_43	PA_60	PA_73	PA_89
PA_18	PA_31	PA_44	PA_61	PA_74	PA_9
PA_19	PA_32	PA_45	PA_62	PA_75	PA_90
PA_2	PA_33	PA_47	PA_64	PA_77	PA_91
PA_22	PA_34	PA_48	PA_65	PA_79	PA_92
PA_24	PA_35	PA_5	PA_66	PA_82	PA_93
PA_25	PA_36	PA_51	PA_67	PA_83	PA_94
PA_27	PA_37	PA_53	PA_69	PA_84	PA_95
PA_28	PA_38	PA_54	PA_70	PA_85	PA_96
PA_29	PA_41	PA_58	PA_71	PA_87	PA_97

4.4.2 Sekvenování a kontrola kvality dat

Celkově bylo osekvenováno 3 093 174 párových readů za účelem multiplexní amplifikace STR za použití specifických primerů. Sekvenování proběhlo na přístroji MiSeq (Illumina) v nastavení 2x250 bp. Čip byl sdílený. Hrubá datová analýza byla provedena pomocí Basespace cloudového rozhraní (Illumina), které je součástí sekvenátoru.

U knihoven byla provedena kontrola kvality pomocí nástrojů FastQC (verze 0.11.8) (Andrews 2010) a MultiQC (verze 1.6) (Ewels 2018). V tabulce č. 16 je vidět velikost každé z knihoven v párech bazí.

Tabulka 16 - Kontrola kvality a velikost knihoven (bp)

Číslo knihovny	Název vzorku	Kvalita vzorku	Velikost knihovny (bp)	Kvalita knihovny
1	6x_a11	prošel	356	prošel
2	25x_equi	prošel	334	prošel
3	66_equi	prošel	312	prošel
4	25x_corect	prošel	329	prošel
5	66x_corect	prošel	334	prošel

Následně byly knihovny zbaveny adaptérů za pomoci programu Trim Galore (verze 0.5.0) (Krueger 2018). Ke spojení readů v jednotlivých knihovnách byl použit program fastq-join (verze 1.3.1) (Aronesty 2013). Spojené ready pro každou knihovnu byly rozděleny dle primerů pomocí nástroje Agrep (Wu a Mamber 1992) s možností jedné chyby/jeden mismatch, a to jak pro forward (F), tak pro reverse (R) primer. Tyto ready byly uloženy do textových souborů dle jednotlivých primerových párů. Dále se pracovalo pouze s ready obsahujícími sekvence F i R primerů – počty readů je možné vidět v kapitole Výsledky.

4.5 Fáze 3

4.5.1 Rozšíření poolu/knihovny č. 3

Na základě naměřených parametrů a vize toho, jak by měla metodika dále postupovat, byla vybrána knihovna č. 3, která byla vytvořena jako ekvimolární multiplex 66 nejlepších primer setů. Za účelem rozšíření funkcionality knihovny i na další vzorky než 1 (SM 40), na kterém jsme knihovny testovali, bylo do knihovny přidáno dalších 18 primerů, které byly smíchány ekvimolárně. Nově přidané primery byly doposud úspěšně používány pro analýzu fragmentů pomocí kapilární elektroforézy, a to i na smrku ztepilém. Bylo tedy již ověřeno, že tyto přidané primery vykazují vysokou variabilitu. Díky přidání těchto konkrétních primerů bylo možné propojit výstupy konvenční a nové metody sekvenování.

Příloha č. 2 obsahuje 18 nově přidaných primerů a jejich vlastnosti.

4.6 Fáze 4

4.6.1 Výběr SSR lokusů

Poté, co se ve fázi 3 nepodařilo s novým multiplexem dojít až do fáze detailního vyhodnocení, vyvstalo několik možností, jak v analýze pokračovat. Při rozhodování byla brána v potaz zejména finanční stránka celého procesu a očekávaný výsledek – analýza SSR lokusů. Dominantní primer Pt71936, který ve třetí fázi obsadil přibližně 93 % sekvenační kapacity, byl i s některými dalšími primery z dalších fází vyřazen.

V rámci fáze 4 byly vytvořeny dva nové soubory primerových párů označovaných jako varianta A a varianta B.

Varianta A obsahovala primerové sekvence, které byly využívány v rámci konvenční metody fragmentační analýzy a zároveň se ukázaly jako perspektivní ve třetí fázi. Tato varianta byla postavena na 14 lokusech viz příloha č. 4.

Pro variantu B byly vybrány lokusy, které byly navrženy v 1. fázi této práce a testovány ve 2. fázi této práce. Konkrétně z knihovny č. 1, která byla sestavena ve 2 etapách – nejprve v single reakci (11 primerů po 6 poolech – dle amplifikační účinnosti), poté smíchání (66 primerů – ekvimolárně). Z této knihovny o šesti poolech (A, B, C, D, E, F) bylo vybráno 25 lokusů. Lokusy byly vybrány pouze z poolů B, C, D a F. Tato varianta byla postavena na 25 lokusech viz příloha č. 5.

Oba tyto soubory primerů byly otestovány na 12 vzorcích smrku ztepilého (A1-A12).

4.6.2 Příprava knihoven

K přípravě knihoven bylo přistoupeno tak, že pro varianty A i B byla příprava provedena v genetické laboratoři FLD, následné analýzy byly provedeny externí genetickou laboratoří SEQme.

V laboratoři FLD došlo nejprve ke dvojfázové PCR u 12ti zájmových vzorků smrku ztepilého za účelem namnožení zájmových lokusů. První PCR reakce tedy obsahovala i vybrané primery. Reakce probíhala zvlášť pro multiplex 1, multiplex 2 a pro primery z knihovny č. 1. Poté byly produkty 1. PCR použity ve 2. PCR, kdy se přidaly do směsi také indexy a adaptory.

Následně byla provedena kvantifikace pomocí fluorometru Quibit s využitím protokolu a kitu dsDNA BR Assay Kit (Invitrogen™).

Měření koncentrace pomocí fluorometru funguje na principu detekce emitovaného světelného signálu z fluorescenčního barviva, které se přidává do PCR produktu (DNA nebo RNA). Fluorescenční barvivo emituje světlo pouze v případě, že dojde k navázání barviva na cílenou molekulu (DNA či RNA). Po zhodnocení výsledků měření koncentrace DNA, byly vzorky přečištěny.

Přečištění bylo provedeno na paramagnetických kuličkách pomocí AMPure XP pro PCR purifikaci (Becman Coulter, Inc). K přečištění paramagnetickým kuličkami je zapotřebí speciální magnetický stojan na zkumavky, 70% ethanol a voda bez obsahu štěpících enzymů. Nejprve bylo přidáno 1.8 µl AMPure XP na 1.0 µl vzorku. DNA fragmenty se naváží na paramagnetické kuličky. Poté dojde díky magnetickému stojanu k separaci fragmentů s kuličkami od zbytku roztoku (kontaminanty). Dále je nutné pomocí ethanolu vymýt kontaminanty ze vzorků. Na závěr dojde k vyplavení DNA fragmentů z magnetických kuliček pomocí elučního bufferu. Eluce proběhla do objemu 40 µl.

Amplikony z jednotlivých variant ze stejného vzorku byly smíchány do jednoho poolu. V tomto případě se jednalo o 2x12 vzorků (varianta A a varianta B). Vzorky varianty A obsahovaly pool smíchaný z multiplexu 1 a multiplexu 2, neboli 14 amplikonů. Vzorky varianty B obsahovaly pool smíchaný z 25 amplikonů z původní knihovny č 1.

Vzorky byly odeslány do externí genetické laboratoře SEQme, kde byla jejich koncentrace opět přeměřena pomocí fluorometru, nicméně za použití jiného kitu Qubit 1X dsDNA high-Sensitivity Assay Kit (Invitrogen™).

Dle výsledků měření koncentrací obou variant nebylo výhodné pokračovat s variantou B, která vykazovala nevyhovující výsledky. Dále se pokračovalo pouze s variantou A, která vykazovala lepší výsledky koncentrace amplikonů. Koncentrace jsou uvedeny v příloze č. 6.

4.6.3 Paralelní sekvenování

Sekvenování z obou konců proběhlo na přístroji Miseq (Illumina) v nastavení 2x250 bp se sdíleným čipem. Hrubá datová analýza byla provedena pomocí Basespace cloudového rozhraní (Illumina), které je součástí sekvenátoru. Další analýzy proběhly ve spolupráci s bioinformatiky ze společnosti SEQme.

V prvním kroku byla ověřována přítomnost repetitivních oblastí pomocí skriptovacího jazyka Bash v příkazovém řádku (Fox 2020). Sekvenační data pro každý vzorek byla rozdělena do jednotlivých souborů na základě přítomnosti sekvencí daných primerů (povoleno jeden mismatch). V jednotlivých souborech byl následně hledán motiv repetyce. Tato data byla použita pouze pro zjištění přítomnosti/nepřítomnosti readů pro daný primer pár, následnou vizuální kontrolu přítomnosti/nepřítomnosti motivu repetyce a nalezení vhodných kotevních „*anchor*“ - sekvencí nutných pro vytvoření konfiguračního souboru pro spuštění programu STRait Razor verze 3.0 (Woerner et al. 2017).

Druhá fáze analýzy byla provedena na trimovaných datech (program Trim Galore, verze 1.18, $Q > 25$) (Krueger 2018). Vzhledem k tomu, že nalezené SSR oblasti jsou u většiny lokusů prosekvenované v celé délce v prvních readech, další analýzy byly pro zjednodušení prováděny jen na R1 souborech. Detekce SSR lokusů, odečet délek a variant byly provedeny programem STRait Razor (Woerner, King a Budowle, 2017). Extrakce dat pro jednotlivé lokusy byla provedena pomocí skriptovacího jazyka Bash (Fox 2020) a následně zaznamenána do tabulky č. 25 v kapitole Výsledky.

4.6.1 Kontrola kvality QC a analýza sekvenovaných amplikonů

U knihoven byla provedena kontrola kvality pomocí nástrojů FastQC (verze 0.11.8) (Andrews 2010) a MultiQC (verze 1.6) (Ewels 2018).

U všech 12 vzorků smrku ztepilého byla provedena bioinformatická analýza amplikonů a po ověření zastoupených sekvencí převedena na údaj o celkové délce amplikonu. Tato forma vyhodnocení umožní pro odpovídající vzorky porovnat získané výstupy nové genotypizační platformy s výstupy klasické metody fragmentační analýzy, která byla realizována dle metodiky použité ve vědecké práci z roku 2020 od Bínové a kolektivu (Bínová et al. 2020).

5 Výsledky

Kapitola Výsledky je pro zvýšení přehlednosti strukturována obdobně jako Metodika, tedy rozčleněna na 5 fází postupu. Obsáhlejší tabulky jsou vloženy do příloh na konci této práce.

5.1 Fáze 0

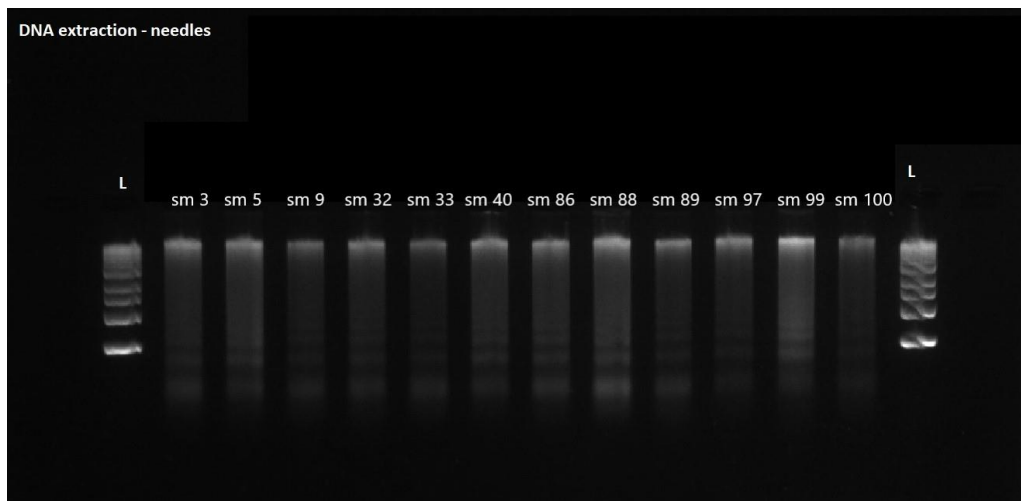
5.1.1 Kvalita a koncentrace DNA

V tabulce č. 17 jsou uvedeny výsledky měření na přístroji NanoDrop (Thermo Fisher SCIENTIFIC). Minimální vstupní hodnoty pro sekvenování splnily všechny vzorky, a to koncentraci minimálně 25 μl a poměrový koeficient 260/280 v rozmezí 1,8-2,0 pro čistotu.

Tabulka 17 - Výsledky měření koncentrace DNA a čistoty vzorků, NanoDrop (Thermo Fisher SCIENTIFIC)

Lokalita	ID vzorku FLD	ID vzorku SEQme	C (ng/ μl)	260/280	260/230
Frýdlant	sm 3	A1	63	1.87	1.94
	sm 5	A2	64.4	1.84	1.86
	sm 9	A3	70.4	1.84	1.86
Boubín	sm 32	A4	46.9	1.86	2.28
	sm 33	A5	42.4	1.85	2.08
	sm 40	A6	52.3	1.83	2.86
Frýdek – Místek	sm 86	A7	39.7	1.89	1.83
	sm 88	A8	36.7	1.87	1.98
	sm 89	A9	49.1	1.87	1.94
Pelhřimov	sm 97	A10	37	1.88	2.41
	sm 99	A11	49.8	1.84	1.94
	sm 100	A12	48.4	1.88	2.13

Za účelem ověření výsledků ze spektrofotometrického měření na přístroji NanoDrop byla provedena elektroforetická separace na agarózovém gelu. Na gelu byla detekována přítomnost DNA u všech 12ti vzorků, které na snímku mají podobnou intenzitu UV světelného odrazu, což potvrzuje správnost předešlého měření koncentrace. Snímek byl pořízen přístrojem s kamerou pro snímání fluorescenčních DNA gelů Azure c200 (Azure Biosystems, Inc.). K nahlédnutí je jako obrázek č. 10.



Obrázek 10 - Výsledek elektroforetické separace DNA na agarózovém gelu

5.2 Fáze 1

5.2.1 Vytvořené primery

Z navržených 93 primerových párů byl referenční vzorek úspěšně a silně amplifikován u 39 párů primerů, u 9 primerů proběhla amplifikace s nižší účinností. Slabě se amplifikovalo 13 párů primerů a velmi slabě se amplifikovalo 5 párů primerů. Z analýzy bylo vyřazeno 27 primerových párů, u kterých nebyl viditelný produkt amplifikace na agarózovém gelu. V tabulce č. 18 je znázorněná míra amplifikace u jednotlivých primerů. Primerové páry, které byly pro další analýzy vyřazeny jsou podbarveny bílou barvou. Se zbylými 66 primerovými páry (zvýrazněny v tabulce č. 18) se pokračovalo v navazujících analýzách.

Tabulka 18 - Účinnost amplifikace u jednotlivých prumerů

Primery	Intenzita amplifikace	Primery	Intenzita amplifikace	Primery	Intenzita amplifikace
PA_1	+++	PA_35	+++	PA_67	+++
PA_2	+++	PA_36	?	PA_68	
PA_3	+	PA_37	+	PA_69	+++
PA_4		PA_38	++	PA_70	+++
PA_5	+	PA_39		PA_71	+++
PA_6		PA_40		PA_72	+++
PA_8		PA_41	++	PA_73	+++
PA_9	++	PA_42	+	PA_74	+++
PA_10	++	PA_43	++	PA_75	+++
PA_11		PA_44	++	PA_76	
PA_12		PA_45	+++	PA_77	+++
PA_13		PA_46		PA_78	
PA_15		PA_47	+++	PA_79	+++
PA_17		PA_48	+	PA_80	
PA_18	+++	PA_49		PA_81	
PA_19	+	PA_50		PA_82	+++
PA_20		PA_51	+++	PA_83	+++
PA_21		PA_52		PA_84	+++
PA_22	+++	PA_53	+++	PA_85	+
PA_23		PA_54	++	PA_86	
PA_24	?	PA_56		PA_87	+
PA_25	+++	PA_57		PA_88	+++
PA_26		PA_58	+++	PA_89	+
PA_27	+++	PA_59	+++	PA_90	+++
PA_28	+	PA_60	+	PA_91	+++
PA_29	+++	PA_61	+++	PA_92	+
PA_30	?	PA_62	?	PA_93	+++
PA_31	++	PA_63		PA_94	+++
PA_32	+	PA_64	?	PA_95	+++
PA_33	+++	PA_65	+++	PA_96	+++
PA_34	++	PA_66	+++	PA_97	+++
+++ vysoká účinnost (39 prumerů), ++ střední účinnost (9 prumerů), + malá účinnost (13 prumerů), ? velmi malá účinnost (5 prumerů)					

Navržené primer sety byly roztrženy na skupiny dle amplifikační účinnosti.

- nefunkční:** 27 primerových párů
- velmi slabě amplifikující:** 5 primerových párů
- slabě amplifikující:** 13 primerových párů
- středně amplifikující:** 9 primerových párů
- silně amplifikující:** 39 primerových párů

Na mikrosatelitní oblasti byl rozptyl v počtu sekvencí při amplifikaci multiplexem sestaveným z všech amplifikujících primerů ze skupin b) - e) (66 primerových párů) výrazně vyšší, než se očekávalo. Tento velký rozptyl je pravděpodobně způsobený výskytem jiných než cílových sekvencí v daném amplikonu např. sekvence neobsahující repetice, sekvence výrazně kratší, než by měly být nebo sekvence s jinými repeticemi. Sekvence s repeticemi, tak jak bychom je očekávali, se ve větší míře vyskytují mezi sekvencemi s délkou, jež odpovídá teoretické délce navrženého amplikonu. Nejlépe z hlediska rozptylu sekvencí vyšel multiplex o 9ti středně amplifikujících primerech ze skupiny d) středně amplifikujících. Tato úspěšnost pravděpodobně vychází z nižšího počtu primerů v multiplexu. V tabulce č. 19 je možné jednotlivé rozptyly v počtu sekvencí pro daný primer vidět. Barvy podbarvení jsou stejné jako u předešlé tabulky: zelená – vysoká účinnost (39 primerů), modrá – střední účinnost (9 primerů), červená malá účinnost (13 primerů) a velmi malá účinnost (5 primerů).

Tabulka 19 - Vypočítané jednotlivé zastoupení amplikonů obsahujících primerové sekvence (forward a reverse readů) z celkového počtu readů v poolech.

ready R1 (forward)/ready R2 (reverse)						
pooly	1	2	3	4	5	
	correction	equimolar	high	medium	low	
celkem	596 651	524 237	243 345	48 538	97 715	
nepárováno	30 915	34 735	18 658	1 903	5 224	
primer	počet	počet	počet	počet	počet	motiv
PA_1	15/29	690/638	610/648	0/0	0/0	(ATAC) ₁₀
PA_2	69/36	9572/7635	4297/2937	0/0	0/0	(ATAC) ₁₃
PA_18	1/13	73/118	73/151	0/0	0/0	(TATC) ₁₁
PA_22	3/72	345/390	185/193	0/0	0/0	(ATAC) ₁₁
PA_25	1137/97	14532/13954	9517/8631	0/0	0/0	(TCAC) ₁₀
PA_27	3/91	124/26134	63/10052	0/1	0/1	(ATAC) ₁₄
PA_29	8/7	166/389	190/245	0/0	0/0	(TATC) ₁₀
PA_33	5/0	82/18	32/15	0/0	0/0	(TATC) ₁₀
PA_35	1/18	4/63	19/47	0/0	0/0	(ATAC) ₁₀
PA_45	3/8519	23/7270	69/3651	0/0	0/0	(TACA) ₁₂
PA_47	87/4	5068/29	3043/30	0/0	0/0	(ATAC) ₁₀
PA_51	4/9	93/49	198/75	0/0	0/0	(ATAC) ₁₀
PA_53	78/73	4594/4593	3108/3221	0/0	0/0	(TATC) ₁₃
PA_58	28/74	1122/1133	1302/1267	0/0	0/0	(ATAC) ₁₄
PA_59	3/5	1640/1320	262/198	0/0	0/0	(TGTA) ₁₂
PA_61	222/1094	1438/3422	3689/3957	0/0	0/0	(TATC) ₁₁
PA_65	725/741	13730/12585	19697/17980	0/0	0/0	(CATA) ₁₁
PA_66	13/304	697/1033	638/675	0/0	0/0	(TGTA) ₁₀
PA_67	465/130	48235/19630	16650/7588	0/0	0/0	(ATAC) ₁₀
PA_69	104/149	9636/9200	2568/1534	0/0	0/0	(TATC) ₁₃

PA_70	14/17	12/99	3/78	0/0	0/0	(ATAC) ₁₁
PA_71	1/12	103/2213	128/1716	0/0	0/0	(CATA) ₁₁
PA_72	227/17	7149/578	4978/1106	0/0	0/0	(ATAC) ₁₀
PA_73	10/3	81/50	61/121	0/0	0/0	(CATA) ₁₁
PA_74	10/4	13/65	4/13	0/0	0/0	(CATA) ₁₃
PA_75	181/196	41702/39661	22863/21931	0/0	0/0	(TACA) ₁₀
PA_77	3/2	33/11	112/13	0/0	0/0	(ATAC) ₁₅
PA_79	1/94	36/4786	56/3548	0/0	0/0	(ATAC) ₁₁
PA_82	1044/1599	161577/166797	105286/106310	0/0	1/2	(TACA) ₁₁
PA_83	0/0	20/0	21/0	0/0	0/0	(ATAC) ₁₃
PA_84	21/31	2060/1531	1681/853	0/0	0/0	(ATAC) ₁₄
PA_88	460/458	24425/25007	19487/19672	0/0	0/1	(TTTA) ₁₁
PA_90	3/4	1245/1249	1225/1203	0/0	0/0	(GATA) ₁₀
PA_91	28/23	1481/1460	780/779	0/0	0/0	(TATC) ₁₀
PA_93	1/2	48/15	13/6	0/0	0/0	(TATC) ₁₀
PA_94	1/0	63/15	89/5	0/0	0/0	(TAGA) ₁₁
PA_95	4/4	1154/1105	1612/1518	0/0	0/0	(CATA) ₁₀
PA_96	1/3	20/252	14/774	0/0	0/0	(TAAA) ₁₀
PA_97	132/5	1200/29	59/58	0/0	0/0	(ATAC) ₁₂
PA_9	23/2	13/1	0/0	1771/1228	0/0	(AATA) ₁₀
PA_10	0/2323	1/1756	0/0	2857/6295	0/0	(TACA) ₁₂
PA_31	15/2	15/1	0/0	4758/4695	0/0	(TATC) ₁₃
PA_34	22/42	20/136	0/0	4703/13907	0/0	(ATAC) ₁₀
PA_38	0/159	2/1170	0/0	9/648	0/0	(CATA) ₁₁
PA_41	85/45	87/92	0/0	15211/13075	0/0	(CATA) ₁₄
PA_43	4/6	0/9	0/1	3445/3273	0/1	(CATA) ₁₁
PA_44	401/4	337/2	0/0	13816/917	0/0	(TACA) ₁₅
PA_54	2/77	13/78	0/0	60/74	0/0	(TATC) ₁₂
PA_3	320308/0	120298/0	3/0	2/0	70454/0	(CATA) ₁₀
PA_5	12/25	4/1	0/0	0/0	17/11	(CATA) ₁₀
PA_19	763/437	518/857	0/0	0/0	84/24	(CATA) ₁₃
PA_24	3971/4132	39/41	0/0	0/0	685/776	(ATAC) ₁₂
PA_28	168/39	47/0	0/0	0/0	41/9	(ATAC) ₁₀
PA_30	381/11	63/9	0/0	0/0	28/3	(GATA) ₁₀
PA_32	90/51	38/24	0/0	0/0	27/12	(ATAC) ₁₄
PA_36	79/23	44/3	0/0	0/0	7/3	(ATAC) ₁₀
PA_37	98/607	0/26	0/0	0/0	34/135	(TACA) ₁₀
PA_42	160/1840	417/175	0/0	0/0	4/556	(TATC) ₁₀
PA_48	1678/1392	261/2470	0/1	0/0	357/268	(ATAC) ₁₆
PA_60	194/181349	22/6926	0/1	0/1	49/15141	(ATAC) ₁₀
PA_62	95/2	39/0	0/0	0/0	6/1	(TATC) ₁₀
PA_64	26399/28221 4	2240/105555	0/4	1/4	3217/61986	(TCTA) ₁₁
PA_85	5614/1388	1115/35	0/0	0/0	549/290	(ATAC) ₁₁
PA_87	199558/72	9568/17	2/0	2/0	16843/16	(ATAC) ₁₀
PA_89	28/20	3/1	0/0	0/0	3/5	(ATAC) ₁₀
PA_92	472/13	12/3	0/0	0/0	85/8	(TACA) ₁₄

5.2.2 Multiplexní PCR a vytvoření sekvenačních knihoven

Po optimalizaci PCR podmínek byla provedena multiplexní PCR s následným vytvořením 5 sekvenačních knihoven. Každá z knihoven byla pomocí přístroje Miseq (Illumina) úspěšně osekvenována přibližně 500 000 ready na jeden vzorek. Primární datová analýza byla zaměřena na vyhledání předpokládaných repetitivních oblastí, než se předpokládalo.

Tabulka č. 20 znázorňuje vypočítané zastoupení ampliconů obsahující primerové sekvence v 5 variantách poolů pro oba ready spojené pomocí fastq-join (Aronesty, 2013).

Tabulka 20 - Vypočítané jednotlivé zastoupení ampliconů obsahujících primerové sekvence z celkového počtu readů v poolech. Oba ready spojeny.

R1 + R2 ready spojené (fastq-join)					
pooly	1	2	3	4	5
	correction	equimolar	high	medium	low
primer	počet	počet	počet	počet	počet
PA_1	34	586	556	0	0
PA_2	54	9 119	3 641	0	0
PA_18	7	102	141	0	0
PA_22	46	287	168	0	0
PA_25	882	10 837	6 830	0	0
PA_27	55	19 037	7 067	0	0
PA_29	10	315	194	0	0
PA_33	3	60	26	0	0
PA_35	15	48	42	0	0
PA_45	6 465	5 462	2 639	0	0
PA_47	60	3 483	1 974	0	0
PA_51	4	77	140	0	0
PA_53	45	3 263	2 074	0	0
PA_58	23	943	1 090	0	0
PA_59	5	904	170	0	0
PA_61	808	2 535	2 218	0	0
PA_65	577	10 786	14 930	0	0
PA_66	230	920	621	0	0
PA_67	312	35 060	11 576	0	0
PA_69	133	8 037	2 137	0	0
PA_70	25	97	64	0	0
PA_71	12	1 637	1 413	0	0
PA_72	104	3 872	2 446	0	0
PA_73	10	83	109	0	0
PA_74	11	51	10	0	0

PA_75	137	30 100	17 250	0	0
PA_77	5	29	72	0	0
PA_79	43	2 314	1 680	0	0
PA_82	1 342	149 006	93 830	0	1
PA_83	0	13	15	0	0
PA_84	17	1 441	991	0	0
PA_88	130	10 457	7 035	0	0
PA_90	4	856	802	0	0
PA_91	21	1 293	686	0	0
PA_93	2	29	15	0	0
PA_94	1	60	63	0	0
PA_95	5	1 032	1 370	0	0
PA_96	2	167	510	0	0
PA_97	93	883	54	0	0
PA_9	18	10	0	1 501	0
PA_10	1 058	1 307	0	5 344	0
PA_31	13	8	0	3 692	0
PA_34	36	93	0	8 631	0
PA_38	124	894	0	503	0
PA_41	72	118	0	11 964	0
PA_43	1	2	0	1 069	0
PA_44	225	172	0	7 672	0
PA_54	57	62	0	46	0
PA_3	221 851	92 266	1	1	50 926
PA_5	21	2	0	0	18
PA_19	569	719	0	0	52
PA_24	3 453	40	0	0	653
PA_28	120	31	0	0	25
PA_30	265	58	0	0	20
PA_32	113	47	0	0	24
PA_36	50	30	0	0	5
PA_37	425	22	0	0	94
PA_42	1 526	443	0	0	454
PA_48	2 066	2 042	0	0	465
PA_60	130 122	5 312	1	0	11 326
PA_62	67	23	0	0	5
PA_64	223 911	86 492	1	3	49 274
PA_85	3 988	838	0	0	487
PA_87	138 381	7 243	1	1	12 134
PA_89	36	4	0	0	5
PA_92	131	8	0	0	33

Sekvenční analýza ukázala, že hloubkové sekvenování metodou NGS může do značné míry kompenzovat optimalizaci multiplexní PCR, nicméně ne do takové míry, aby multiplex nemusel být optimalizován vůbec.

Pro vyladění multiplexu je zapotřebí hlubší analýza dat a určit, či porovnat:

- procenta mapujících a nemapujících sekvencí a jejich délkovou distribuci
- procenta sekvencí s očekávanou repeticí v souborech mapujících se sekvencích a nemapujících se sekvencích
- korelace mezi počtem sekvencí na oblast, procenty mapujících se sekvencí, sekvencí s očekávanou repeticí, délkovými distribucemi jednotlivých skupin a úspěšností amplifikace

Na základě těchto výstupů by bylo možné vysledovat trendy, které by pomohly lépe sestavit multiplex, popř. multiplexy o nižší plexitě, než je celkový počet „použitelných“ primer setů.

5.3 Fáze 2

5.3.1 Sekvenování

U čtyř párů primerů (PA_3, PA_44, PA_64 a PA_87) se nepodařilo detekovat žádné nebo dostatečné množství readů obsahující odpovídající kombinace F a R primerů (v tabulce č. 21 jsou tyto primery přeškrtnuté). U čtyř primerů byly zjištěny interakce s jinými R primery. U všech variant poolování (knihoven) byla ověřena přítomnost požadovaného repetitivního motivu a následně přítomnost samotné repetice. Lokusy byly rozděleny dle obsahujícího motivu a repetice na 4 podskupiny. Pokud obsahovali obojí (Y/Y) jsou v tabulce č. 21 zvýrazněny zeleně, pokud obsahovali pouze motiv bez repetice (Y/N) jsou zvýrazněny modře a lokusy neobsahující motiv ani repetici (N/N) jsou označeny světle červeně. Rudě jsou označeny lokusy, kde nebyly přítomny žádné ready. Tabulka č. 21 znázorňuje přítomnost požadovaného repetitivního motivu a přítomnost případné repetice pro pool č. 1 - 6x a11 (knihovna č. 1). Fialově jsou podbarveny domnělé artefakty.

Tabulka 21 - Repetice a výskyt lokusů – pool 1 (6x a11)

Lokus	Požadovaný motiv	F+R	Repetice	Výskyt	Lokus	Požadovaný motiv	F+R	Repetice	Výskyt
PA_10	TACA	57	10	20	PA_74	CATA	19 319	6	3479
			11	7				7	5610
			12	29				8	5565
PA_25	TCAC	433	3	24	PA_77	ATAC	4 169	10	4140
			4	179	PA_79	ATAC	399	6	69
			5	21	PA_85	ATAC	41	12	13
			6	57	PA_89	ATAC	525	17	25
7	98	8	523						
PA_27	ATAC	29	5	23	PA_9	AATA	1 459	4	1026
PA_28	ATAC	1874	6	5				5	429
			11	10	PA_90	GATA	280	4	48
			12	1858				7	230
PA_31	TATC	6 017	5	4686	PA_96	TAAA	1 342	4	852
			6	45				5	450
			7	1058	PA_97	ATAC	448	5	446
PA_33	TATC	1 805	5	484	PA_1	ATAC	5491		
			7	92	PA_19	CATA	1		
			10	21	PA_2	ATAC	860		
			11	22	PA_29	TATC	7888		
PA_36	ATAC	927	8	237	PA_34	ATAC	436		
			9	676	PA_35	ATAC	649		
PA_37	TACA	3 289	7	39	PA_41	CATA	2028		

			9	517	PA_65	CATA	23045		
PA_38	CATA	66	8	64	PA_83	ATAC	4352		
PA_42	TATC	738	4	632	PA_84	ATAC	626		
			9	106	PA_88	TTTA	1969		
PA_43	CATA	2 817	5	732	PA_95	CATA	83844		
			9	773	PA_18	TATC	21785		
			9	1292	PA_22	ATAC	3908		
PA_44	TACA	738	7	439	PA_24	ATAC	2		
			14	37	PA_3	CATA	4		
PA_45	TACA	368	6	177	PA_30	GATA	1		
			11	17	PA_32	ATAC	61		
			12	168	PA_51	ATAC	4285		
PA_47	ATAC	245	7	244	PA_53	TATC	7569		
PA_48	ATAC	72	11	4	PA_54	TATC	163		
			12	68	PA_59	TGTA	292		
PA_5	CATA	266	5	187	PA_61	TATC	16065		
			11	79	PA_66	TGTA	613		
PA_60	TACA	81	10	41	PA_69	TATC	326		
			13	38	PA_75	TACA	9716		
PA_70	ATAC	1 801	12	21	PA_82	TACA	10594		
			13	1774	PA_91	TATC	647		
PA_71	CATA	155	13	71	PA_92	TACA	51		
PA_72	ATAC	318	10	10	PA_93	TATC	18		
			11	168	PA_94	TAGA	1545		
PA_73	CATA	325	7	173	PA_62				
			9	151	PA_64				
					PA_87				

5.3.2 Kontrola kvality QC

Z tabulky č. 22 je možno vidět, že obsah GC se pohybuje mezi 30-43 %, což je považováno jako nižší hodnota. Ideální je obsah guaninu a cytosinu mezi 40-50 %, kdy mají tyto dvě nukleotidové báze silnější vodíkovou vazbu a pomáhají se stabilitou primerů. V tabulkách č. 23 (podrobněji v tabulce č. 22) je také možné vidět průměrnou délku sekvence, celkový počet readů, případně procentuální vyjádření podílu duplicitních readů a modulů, které sekvenováním neprošly.

Tabulka 22 - Výsledky sekvenace a kontroly kvality knihoven (R1, R2)

Název vzorku	Duplicitní ready (%)	GC (%)	Délka (bp)	neprošlo %	Počet readů
6x-a11_R1	91,0%	34%	199	36%	318 184
6x-a11_R2	86,2%	34%	198	45%	318 184
2-25x-equi_R1	92,6%	38%	171	45%	315 768
2-25x-equi_R2	88,5%	37%	169	45%	315 768
3-66x-equi_R1	92,3%	40%	160	45%	247 265
3-66x-equi_R2	85,4%	40%	160	45%	247 265
4-25x-corect_R1	92,8%	32%	174	45%	289 783
4-25x-corect_R2	87,6%	31%	171	45%	289 783
5-66x-corect_R1	94,8%	43%	171	45%	375 587
5-66x-corect_R2	91,1%	43%	171	45%	375 587

Knihovna č. 1 (rozdělení do 6 skupin po 11 párech primerů) se ukázala jako nejlepší varianta. Kontrola počtů repetitivních oblastí byla u každého primerového páru provedena pomocí nástroje agrep (Wu a Mamber 1992). Pro lepší orientaci byly jednotlivé soubory pro knihovnu 1 rozřizeny dle délek spojených readů viz tabulka č. 23. Dále bylo provedeno mapování readů. Mapované ready byly dále zpracovány pomocí softwaru Samtools (verze 0.1.19-44428cd) (Wysoker et al. 2009).

Tabulka 23 – Výsledky sekvenace a kontroly kvality knihoven

Číslo knihovny	Název vzorku	GC obsah %	Průměrná délka (bp)	Celkový počet readů	Párové ready
1	6x_a11	34	199	636 368	2x 318 184
2	25x_equi	38	170	631 536	2x 315 768
3	66_equi	40	160	494 530	2x 247 265
4	25x_corect	32	173	579 566	2x 289 783
5	66x_corect	43	171	751 174	2x 375 587

5.4 Fáze 3

Ačkoliv se knihovna č. 1 (rozdělení do 6 skupin po 11 párech primerů) ukázala jako nejlepší varianta po analýzách ve 2. fázi metodiky, pro další analýzy byla nahrazena knihovnou č. 3. Tím, že v této fázi bylo potřeba rozšířit multiplex o 18 primer setů, které byly poolovány ekvimolárně, přistoupilo se k použití knihovny č. 3, tedy k ekvimolárnímu poolování. Výsledky kontroly kvality a sekvenování u knihovny č. 1 a knihovny č. 3 nebyly výrazně odlišné, proto byla zvolena knihovna č. 3.

Bohužel, jeden z nově přidaných primerů (*Pt71936*) se ukázal jako vysoce dominantní primer a ~93 % veškerých získaných dat se vztahuje k tomuto primeru. Díky tomu nebylo dosaženo dostatečného sekvenačního výstupu pro provedení pilotní populační analýzy. V příloze č. 3 jsou uvedeny výsledky sekvenování, kde jsou vidět u primeru *Pt71936* vysoké počty readů na dané vzorky (označeno červeně). Žlutě jsou podbarveny nově přidané primery. Podbarvení zelenou, modrou a oranžovou barvou kopíruje označení primerů dle toho, zda u nich v 2. fázi byly či nebyly nalezeny repetitivní úseky případně v jakém počtu.

V principu se opět potvrdilo, že metoda NGS je velice informativní pro studium mikrosatelitů, repeticí a SNP v přesahujících oblastech jsou velice přesně detekovatelné, bohužel však v tomto případě analýza ztroskotává na vyladění multiplexu PCR. Mikrosatelity v jednotlivých vzorcích a pro jednotlivé lokusy jsou detekovatelné, ale ne s pokrytím, dostatečným pro populační analýzu.

5.5 Fáze 4

5.5.1 Příprava knihoven

Výsledky z měření koncentrace přečištěných amplikonů všech vzorků pomocí fluorometru jsou uvedeny v příloze č. 6. Zeleně je podbarvený pool obsahující produkty 2 reakčních multiplexů, a to multiplexu 1 (primerové sekvence pro lokusy *EATC1E03*, *SpAG2*, *paGB3*, *WS00716.F13*) a multiplexu 2 (s primerovými sekvencemi *PA33*, *PA_28*, *PA_56*, *WS0019.F22*, *WS0073.H08*, *PAAC23*, *PAAC3*, *WS0092.A19*, *WS0023.B03*, *EATC1D02A*), které po poolování obsahují primersety k variantě A, naopak modře jsou podbarveny pooly, které byly použity ve variantě B. Z výsledků vyplývá, že koncentrace u varianty A je celkově vyšší oproti variantě B. To znamená, že amplifikace u varianty B, neproběhla tak, jak by pro sekvenování bylo potřeba, a tak nemělo smysl ve variantě B dále pokračovat.

Dle výsledků kontroly kvality je spolehlivost určení báze >99,7%.

5.5.2 Analýza sekvenovaných amplikonů

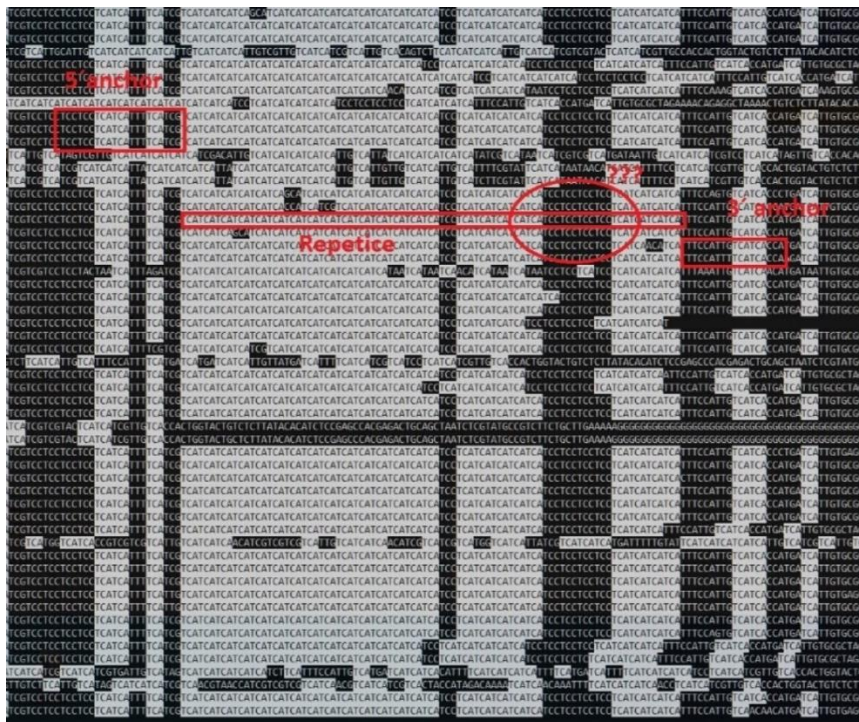
V první fázi analýzy SSR lokusů byla ověřena přítomnost očekávaných repetitivních oblastí za pomoci skriptovacího jazyka Bash (Fox 2020). Pro pět lokusů (*EATC1E03*, *WS0073.H08*, *PAAC23*, *WS0023.B03*, *WS0019.F22*) nebyla detekována žádná, nebo téměř žádná, data, a to u všech dvanácti vzorků. Tyto lokusy byly z následujících analýz vyřazeny. Dále vyhodnocení pokračovalo s devíti zbývajících lokusy. V tabulce č. 24 je seznam lokusů a jejich stav po sekvenování, zda byl lokus vyřazen, případně zda byla pro daný lokus přítomná repetice.

Tabulka 24 – Jednotlivé lokusy a jejich stav po sekvenování

Jednotka repetice	Název lokusu	Sekvenační data	Repetice přítomná
AT	paGB3	OK	ano
TC	SpAG2	OK	ano
GA	WS00716.F13	OK	ano
TCA	EATC1D02A	OK	ano
TCG	Pa_28	OK	ano
AC	WS0092.A19	OK	ano
CGG	Pa_33	OK	ano
AGGTG	PA_56	OK	ano
vyřazeno	PAAC3	OK	ne
vyřazeno	EATC1E03	-	-
vyřazeno	WS0073.H08	-	-
vyřazeno	PAAC23	-	-
vyřazeno	WS0023.B03	-	-
vyřazeno	WS0019.F22	-	-

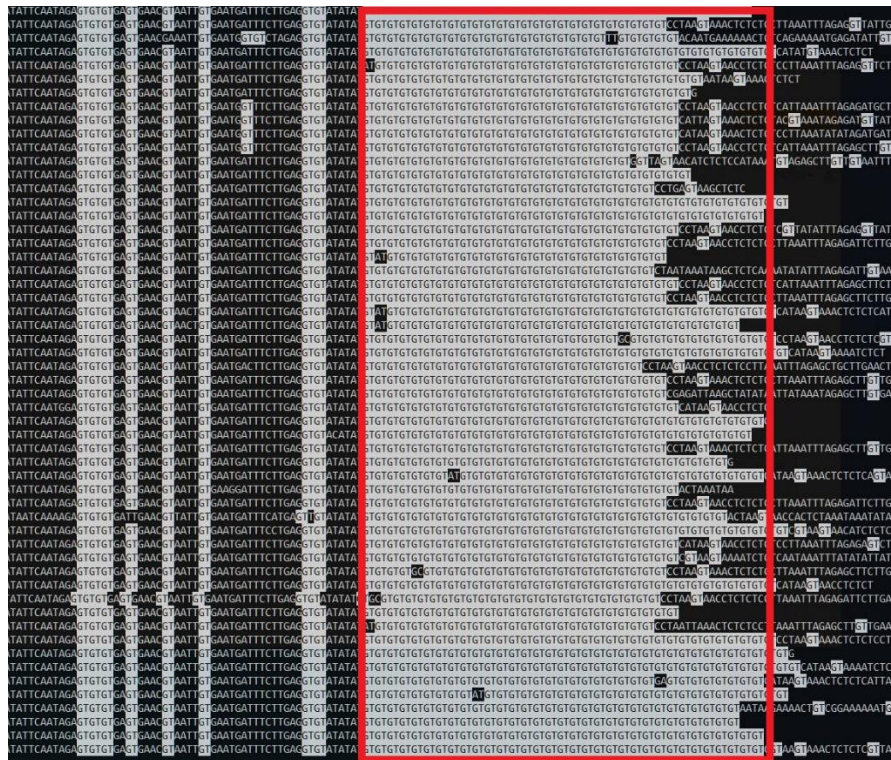
Odečet délek lokusů byl proveden programem STR8rzt (Woerner et al. 2017), kdy pro stanovení délek je potřeba najít tzv. ohraničující sekvence (*anchor*) na 5' i 3' koncích.

U 2 lokusů byl problém tyto oblasti jasně určit (*EATC1D02A*, *PAAC3*). U lokusu *EATC1D02* byla hledaná repetice TCA, ale v datech byly přítomné i SNP TCC, a to i repetitivně viz obrázek č. 11. Neočekávaná repetice TCC byla nakonec zahrnuta do délky repetice tohoto lokusu.



Obrázek 11 - Lokus EATC1D02A, vzorek A1 a neočekávaná repetice TCC

U lokusu *PAAC3* nebyl nalezen očekávaný motiv CA, nicméně u vzorků A1, A2 a A10 byla nalezena repetice jiného motivu, a to GT viz obrázek č. 12. Jelikož nebyla nalezena očekávaná repetice, byl tento lokus z dalších analýz vyřazen.



Obrázek 12 - Lokus PAAC3 vzorek A1, neočekávaná repetice GT

Po vyřazení lokusů, u kterých neproběhla amplifikace nebo nebyla nalezena repetice, pokračovalo vyhodnocení zbývajících osmi lokusů.

Další fáze analýzy byla provedena na trimovaných datech v programu Trim Galore verze 1.18 (Krueger 2018) opět byly hledány SSR lokusy a zaznamenány jejich délky pomocí programu STRait Razor (Woerner et al. 2017). Vzhledem k tomu, že jsou SSR oblasti u většiny lokusů prosekvenované v celé délce v R1 readech, byly další analýzy pro zjednodušení prováděny pouze na R1 readech.

Délky fragmentů a počty readů pro každý z 12 vzorků, byly zaneseny do tabulky č. 25 níže. Žlutě jsou podbarveny homozygotní lokusy a rudě jsou podbarvené lokusy a vzorky, kde neexistují data z NGS kvůli špatné amplifikaci.

Tabulka 25 - Počet readů a délky lokusů dle NGS

ID SEQme	ID FLD	Lokusy		Počet readů	Lokusy		Počet readů
		EAT2A_a	EAT2A_b		paGB3_a	paGB3_b	
A1	sm 3	87	87	2386	12	12	230
A2	sm 5	87	87	4766	14	20	139/28
A3	sm 9	108	108	1538	18	18	122
A4	sm 32	81	81	4670	14	14	27
A5	sm 33	81	81	49	12	12	612
A6	sm 40	81	84	4205/3359	12	18	460/56
A7	sm 86	75	78	3927/5196	12	12	781
A8	sm 88	75	81	1850/2445	12	12	173
A9	sm 89	66	90	2584/1659	20	20	39
A10	sm 97	108	108	3871	14	14	240
A11	sm 99	72	105	3018/1453	16	16	111
A12	sm 100	78	93	2598/2081	14	18	76/28
ID SEQme	ID FLD	Lokusy		Počet readů	Lokusy		Počet readů
		SpAG2_a	SpAG2_b		WSF13_a	WSF13_b	
A1	sm 3	37	37	15	38	38	112
A2	sm 5	27	31	56/25	32	46	5927/1917
A3	sm 9	25	33	74/52	38	38	58
A4	sm 32	29	39	19/16	38	38	3045
A5	sm 33	19	33	71/76	26	38	9721/4680
A6	sm 40	29	31	69/48	40	40	2296
A7	sm 86	25	29	13/9	26	30	14227/10908
A8	sm 88	29	29	14	26	26	4454
A9	sm 89	31	33	19/15	26	26	45
A10	sm 97	29	35	58/36	30	36	9366/5581
A11	sm 99	17	29	67/64	44	46	332/185
A12	sm 100	31	33	25/11	36	40	15/46
ID SEQme	ID FLD	Lokusy		Počet readů	Lokusy		Počet readů
		PA33_a	PA33_b		PA28_a	PA28_b	
A1	sm 3	18	18	7453	18	21	14/14
A2	sm 5	18	18	6616	18	21	2/14
A3	sm 9	18	18	6198	12	12	14
A4	sm 32	18	18	3284	18	18	5
A5	sm 33	18	18	4399	18	21	12/4
A6	sm 40	18	21	2466/2703	0	0	0
A7	sm 86	18	18	2055	18	21	7/1
A8	sm 88	18	18	4430	18	21	1/7
A9	sm 89	18	18	3858	21	21	19
A10	sm 97	18	18	3116	0	0	0
A11	sm 99	18	18	6288	21	21	16
A12	sm 100	18	18	2526	0	0	0

ID SEQme	ID FLD	Lokusy		Počet readů	Lokusy		Počet readů
		WSA19_a	WSA19_b		PA_56_a	PA_56_b	
A1	sm 3	24	38	25/557	15	15	287
A2	sm 5	24	30	1022/925	15	20	20897/19111
A3	sm 9	26	30	590/641	15	20	12546/12339
A4	sm 32	24	30	15/480	15	20	10727/12000
A5	sm 33	24	24	1402	15	20	14273/14184
A6	sm 40	24	24	1061	15	15	15724
A7	sm 86	22	30	417/660	15	15	10112
A8	sm 88	24	24	1681	15	15	272
A9	sm 89	24	30	534/513	15	20	17338/14038
A10	sm 97	24	24	87	15	20	162/24357
A11	sm 99	24	36	717/408	15	20	93/15513
A12	sm 100	24	36	25/616	15	20	106/14580

5.5.3 Konvenční Fragmentační analýza

U stejných vzorků a primerů byla provedena i kapilární elektroforetická separace neboli fragmentační analýza. Toto sekvenování bylo provedeno v rámci jiného výzkumu, avšak výsledky byly pro tuto práci zapůjčeny za účelem porovnání s výsledky NGS. Genotypizace byla provedena nezávislou osobou v software GeneMarker® (SOFTGENETICS). Výsledky jsou uvedeny v tabulce č. 26 níže.

Žlutě jsou podbarveni homozygotní jedinci, bíle jsou podbarveni heterozygotní jedinci a rudě lokusy, které nebyly amplifikovány (0) nebo k nim neexistují data k porovnání z NGS (NA).

Tabulka 26 - Výsledky genotypizování osmi vzorků po fragmentační analýze

Předpokládaná délka jednotlivých alel (a, b) daných lokusů (bp)									
ID SEQme	ID FLD	EAT2A_a	EAT2A_b	paGB3_a	paGB3_b	SpAG2_a	SpAG2_b	WSF13_a	WSF13_b
A1	sm 3	87	87	12	12	37	37	38	38
A2	sm 5	87	87	14	20	27	31	32	46
A3	sm 9	108	108	18	18	25	33	38	38
A4	sm 32	81	81	0	0	29	39	38	38
A5	sm 33	81	81	12	12	19	33	26	38
A6	sm 40	81	84	12	18	29	31	40	40
A7	sm 86	75	78	12	12	25	29	26	30
A8	sm 88	0	0	0	0	0	0	0	0
A9	sm 89	66	90	20	20	31	33	26	26
A10	sm 97	108	108	14	14	29	35	30	36
A11	sm 99	72	105	16	16	17	29	44	46
A12	sm 100	78	93	14	18	31	33	36	40
ID SEQme	ID FLD	PA33_a	PA33_b	PA28_a	PA28_b	WSA19_a	WSA19_b	PA_56_a	PA_56_b
A1	sm 3	18	18	18	21	24	38	15	15
A2	sm 5	18	18	18	21	24	30	15	20
A3	sm 9	18	18	12	12	26	30	15	20
A4	sm 32	18	18	18	18	24	30	15	20
A5	sm 33	18	18	18	21	24	24	15	20
A6	sm 40	18	21	NA	NA	24	24	15	15
A7	sm 86	18	18	18	21	22	30	15	15
A8	sm 88	18	18	18	21	24	24	15	15
A9	sm 89	18	18	21	21	24	30	15	20
A10	sm 97	18	18	NA	NA	24	24	15	20
A11	sm 99	18	18	21	21	24	36	15	20
A12	sm 100	18	18	NA	NA	24	36	15	20

5.6 Porovnání výsledků metod sekvenování

U obou metod došlo na některých lokusech a vzorcích ke špatné amplifikaci, proto některá data nemohou být relevantně porovnána. U NGS se to týká zejména lokusu *PA28*, kde chybí data pro vzorky: A6, A10 a A12. Obdobný problém nastal i u dat z konvenční fragmentační analýzy, kdy díky špatné amplifikaci chybí data pro lokusy *paGB3* (vzorek A4), *PA28* (vzorky A6 a A10). U vzorku A8 nedošlo u lokusů *EAT2A*, *paGB3*, *SpAG2* a *WSF13* k amplifikaci. V tabulce č. 27 jsou tyto případy označeny rudým podbarvením. Jedná se o tzv. nulové alely. U ostatních lokusů a vzorků proběhla amplifikace v pořádku.

V tabulce č. 27 jsou výsledky sekvenování pro obě metody pro každý z lokusů na každém vzorku. Co se týká určení genotypu, tak obě metody u všech lokusů určily shodně zda se jedná o homozygotního či heterozygotního jedince. Locus s nejvíce homozygoty se ukázal jako *PA33*, kde pouze vzorek A6 se dle obou metod jeví jako heterozygot. Jedná se tedy o jeden z nejméně polymorfních markerů. Naopak locus *spAG2* se jeví jako homozygot pouze u 2 vzorků, a to A1 a A8. U vzorku A8 pro tento konkrétní locus neexistují data z konvenční metody sekvenování, ale data mohou být právě díky NGS doplněna.

Délky lokusů se u jednotlivých metod neliší, ale v rámci vyrovnanosti délek napříč vzorky pro daný locus, se jeví jako nejlepší lokusy *PA28* a *PA56*.

U zbylých lokusů to vypadá, že došlo k expanzi mikrosatelitních repeticí tím, že DNA polymeráza tzv. sklouzávala při replikaci repetitivních sekvencí. Jde o poměrně častý jev při použití konvenční metody fragmentační analýzy. Bohužel k této odchylce došlo i u NGS metody, což je pravděpodobně způsobeno chybou při přečišťování amplikonů pomocí magnetických kuliček.

Minimální odchylky v rámci jednotek párů bazí mohou být přičteny genotypizační nepřesnosti při hodnocení dat z fragmentační analýzy. Z tohoto hlediska dopadly nejlépe lokusy (*SpAG2*, *WS00.F13*, *PA33*, *paGB3* a *WS00.A19*) naproti tomu hůře dopadly lokusy (*EAT2A*, *PA_28* a *PA_56*).

Tabulka 27 - Porovnání délek lokusů u NGS a konvenční metody sekvenování

ID SEQme	Konvenční metoda		NGS		Konvenční metoda		NGS	
	EAT2A_a	EAT2A_b	EAT2A_a	EAT2A_b	paGB3_a	paGB3_b	paGB3_a	paGB3_b
A1	87	87	87	87	12	12	12	12
A2	87	87	87	87	14	20	14	20
A3	108	108	108	108	18	18	18	18
A4	81	81	81	81	0	0	14	14
A5	81	81	81	81	12	12	12	12
A6	81	84	81	84	12	18	12	18
A7	75	78	75	78	12	12	12	12
A8	0	0	75	81	0	0	12	12
A9	66	90	66	90	20	20	20	20
A10	108	108	108	108	14	14	14	14
A11	72	105	72	105	16	16	16	16
A12	78	93	78	93	14	18	14	18
ID SEQme	Konvenční metoda		NGS		Konvenční metoda		NGS	
	SpAG2_a	SpAG2_b	SpAG2_a	SpAG2_b	WSF13_a	WSF13_b	WSF13_a	WSF13_b
A1	37	37	37	37	38	38	38	38
A2	27	31	27	31	32	46	32	46
A3	25	33	25	33	38	38	38	38
A4	29	39	29	39	38	38	38	38
A5	19	33	19	33	26	38	26	38
A6	29	31	29	31	40	40	40	40
A7	25	29	25	29	26	30	26	30
A8	0	0	29	29	0	0	26	26
A9	31	33	31	33	26	26	26	26
A10	29	35	29	35	30	36	30	36
A11	17	29	17	29	44	46	44	46
A12	31	33	31	33	36	40	36	40
ID SEQme	Konvenční metoda		NGS		Konvenční metoda		NGS	
	PA33_a	PA33_b	PA33_a	PA33_b	PA28_a	PA28_b	PA28_a	PA28_b
A1	18	18	18	18	18	21	18	21
A2	18	18	18	18	18	21	18	21
A3	18	18	18	18	12	12	12	12
A4	18	18	18	18	18	18	18	18
A5	18	18	18	18	18	21	18	21
A6	18	21	18	21	NA	NA	0	0
A7	18	18	18	18	18	21	18	21
A8	18	18	18	18	18	21	18	21
A9	18	18	18	18	21	21	21	21
A10	18	18	18	18	NA	NA	0	0
A11	18	18	18	18	21	21	21	21
A12	18	18	18	18	NA	NA	0	0

ID SEQme	Konvenční metoda		NGS		Konvenční metoda		NGS	
	WSA19_a	WSA19_b	WSA19_a	WSA19_b	PA_56_a	PA_56_b	PA_56_a	PA_56_b
A1	24	38	24	38	15	15	15	15
A2	24	30	24	30	15	20	15	20
A3	26	30	26	30	15	20	15	20
A4	24	30	24	30	15	20	15	20
A5	24	24	24	24	15	20	15	20
A6	24	24	24	24	15	15	15	15
A7	22	30	22	30	15	15	15	15
A8	24	24	24	24	15	15	15	15
A9	24	30	24	30	15	20	15	20
A10	24	24	24	24	15	20	15	20
A11	24	36	24	36	15	20	15	20
A12	24	36	24	36	15	20	15	20

6 Diskuse

Navržení primerových párů bylo prvním z mnoha kroků provedených v rámci vytváření protokolu v této diplomové práci. V první fázi se podařilo navrhnout, syntetizovat a otestovat 93 primerových párů pro amplifikaci mikrosatelitových lokusů. Z nich se amplifikovalo v dostatečné míře 66 primerových párů. Nedostatečná amplifikace mohla být způsobena neoptimálními podmínkami PCR reakce u některých z primerů jako v publikacích od autorů Hodgetts et al. (2001) nebo Rajora et al. (2001), nebo také rozdíly v nekódujících úsecích referenčního genomu popsáno v publikaci od Nystedt et al. (2013). Se stejným problémem rozdílů v úsecích se setkali i Bernhardsson et al. (2020). V ideálním případě by se měly amplifikovat všechny primery jako u výzkumu od Besnard et al. (2003).

Dalším krokem bylo primery smísit v multiplex. Na základě kombinace parametrů molarity a míry amplifikace, bylo vytvořeno 5 variant poolů, které byly pomocí přístroje Miseq (Illumina) úspěšně sekvenovány. Analýza byla zaměřena na vyhledání předpokládaných repetice ve vzorcích. Bohužel bylo ve vzorcích nalezeno výrazně méně repetitivních oblastí, než se předpokládalo. Nejlépe z hlediska zastoupení amplikonů obsahujících primerové sekvence z celkového počtu readů dopadl pool 2 - zohledňující amplifikační účinnost. Strategie smísení na základě ekvimolarity se ukázala vhodná spíše pro běhy zahrnující více druhů, jako například u Muller et al. (2019), kteří zkoumali strategie poolingů pro NGS u 13 druhů bakterií.

V další fázi bylo potřeba multiplex optimalizovat tak, aby byl vyrovnaný a mohl být použit na širší spektrum vzorků. Proto byly sestaveny sekvenační knihovny na základě více strategií (molarity a amplifikační účinnost) o různých počtech primerů. Jako nejlepší varianta se ukázala knihovna č. 1 (rozdělení do 6 skupin po 11 párech primerů), která měla největší celkový počet readů s průměrnou délkou 199 bp.

Ve 3 fázi bylo potřeba knihovnu rozšířit o markery, které jsou u zájmové druhu široce používané a použili je i Bínová et al. 2020, ve svém výzkumu SSR markerů na různých ekotypech smrku ztepilého. Nově přidaných 18 primerů bylo poolováno ekvimolárně, a proto musela být použita knihovna č. 3, která byla vytvořena jako ekvimolární multiplex 66 nejlepších primerů. Díky přidání těchto konkrétních primerů bylo možné propojit výstupy konvenční metody separace a nové metody sekvenování.

Ve 4 fázi práce byl znovu optimalizován multiplex a ve finální knihovně tedy bylo sekvenováno 14 mikrosatelitových amplikonů. Do finálních analýz se dostalo

8 mikrosatelitových amplikonů. Sekvenováním nebylo možno detekovat téměř 43 % zájmových lokusů, oproti tomu se kolektivu Yang et al. (2019) na jiném zájmovém druhu nepodařilo detekovat téměř 20 % zájmových lokusů.

Výsledky sekvenování souboru primerů ze 4. fáze této práce byly porovnány s výsledky konvenční separace fragmentů pomocí kapilární elektroforézy.

Problémů, které mohou zkreslit vyhodnocení dat, může být celá řada. Například chyby při přípravě knihovny nebo nesprávné zarovnání readu k referenčnímu genomu, jak píše ve své knize Wang (2016).

Dále je vhodné používat algoritmy a nástroje, které umí zlepšit citlivost a specifčnost detekce SNP a určení genotypu z NGS dat jako v případě této práce STRait Razor (Woerner et al. 2017), Bash (Fox 2020), fastQC (Andrews 2010), Trim Galore (Krueger 2018), SAMtools (Li et al. 2009) a další.

Jedním z problémů, který ovlivnil výsledky genotypování u NGS je výskyt nulových alel, o kterém píše i Tibihika et al. (2019) nebo Vartia et al. (2016). Nulové alely mohou vzniknout jako výsledek skluzu artefaktu při PCR reakci, tím pádem vznikají u všech metod závislých na PCR produktech tzn. jak u pokročilých sekvenačních metod, tak u konvenčních metod separace.

Velkým problémem konvenčních metod separace, založených na kapilární elektroforéze je, že genotypizační data jsou přímo závislá na dané platformě. Tomuto se lze elegantně vyhnout pomocí sekvenační metody, která poskytuje přístup přímo k mikrosatelitovým úsekům. Genotypy jsou v tomto případě determinovány pomocí diskrétních dat, jež nepodléhají změnám v závislosti na specifických podmínkách platforme. De Barba et al. (2017) ve své práci zmiňují, že genotypy vzniklé v různých laboratořích na různých platformách, mohou být přímo porovnávány, a to bez nutnosti kalibrace. Tato skutečnost usnadňuje výměnu dat i pro zahraniční programy genetického monitoringu a vytvoření společné genetické databáze ve velkém geografickém měřítku.

Během posledních let došlo k vývoji SSR markerů a genotypizace včetně bioinformatiky. Mikrosatelity však mají stále své limity, například nejsou vhodné pro nemodelové druhy, pro které neexistují sekvenační data, jak ve svém výzkumu poukazují Guichoux et al. (2011). Použití metod sekvenování nové generace k získání sekvenačních dat a identifikaci SSR markerů teprve začíná a jeví se mimořádně slibně. Existuje řada studií zabývajících se technickými a analytickými aspekty SSR sekvenování, přičemž bylo vyvinuto několik bioinformatických přístupů od autorů Hoogenboom et al. (2017), Suez et al. (2016), Zhan et al. (2017) a Barbian et al. (2018), které byly testovány různými

laboratorními protokoly a společně prozkoumaly četné problémy technických a analytických výhod SSR sekvenování oproti konvenčním metodám. Kromě toho se laboratorní metody pro amplifikaci mikrosatelitních markerů a také analytické přístupy k převodu nezpracovaných sekvencí na kodominantní genotypy mezi studii výrazně liší. Z tohoto důvodu je pro potenciální uživatele poměrně obtížné identifikovat jasnou cestu k nastavení nového přístupu SSR sekvenování pro nové zájmové druhy.

Suez et al. (2016) popsali metodu pro genotypování mikrosatelitů dat z NGS. Tato metoda stojí na teoretickém parametrickém modelu pro genotypy a klade si za cíl nalézt optimalizované parametry. Bohužel metoda má některé nevýhody spojené s parametrickým modelem a jeho schopností správně simulovat režim dat. Metoda může vyvolat zkreslení a je velmi náročná na výpočetní náklady. Oproti tomu Zhan et al. (2017) vyvinuli software MEGASAT, jež umožňuje rychlou konverzi DNA sekvenačních dat z vysoce multiplexovaných a poolovaných mikrosatelitových amplikonů na multilokusové genotypy.

Jelikož jsou metody sekvenování nové generace mnohem přesnější a pokročilejší než konvenčně používané metody, lze předpokládat, že odchylky při porovnávání výsledků metod jsou způsobené i chybou při genotypizaci. Pokud bychom se bavili o ideálním počtu variant repetice, tak musíme brát v potaz ploidii zkoumaného organismu, v případě smrku se jedná o diploidní organismus, který má maximálně dvě varianty repetice v případě heterozygota a jednu variantu v případě homozygota. Tato polyploidie byla pravděpodobně způsobena expanzí mikrosatelitních repetic. Volbou nové metody sekvenování bylo předpokládáno, že tento problém nebude nastávat, ale ukázalo se, že při použití NGS právě proces genotypizace může výsledky zkreslit.

7 Závěr

Využití metod sekvenování nové generace není v biologickém a lesnickém odborném prostředí nijak neobvyklé. Dokonce výzkumy s použitím těchto metod byly mnohokrát zaměřeny na rod *Picea*.

Ačkoliv existují studie o míře genetické diverzity u smrku ztepilého (*Picea abies* (L.) H. Karst.) (Maghuly *et al.*, 2006; Bernhardsson *et al.*, 2020; Bínová *et al.*, 2020), stále postrádáme detailní informace o úrovních polymorfismu nukleotidů a nerovnováze vazeb v tzv. nongenic oblastech. Nicméně díky tomu, že je dostupný referenční genom smrku (Nystedt *et al.*, 2013), je možné provádět celou řadu resekvenování genomů jehličnanů, a to i smrku ztepilého (*Picea abies* (L.) H. Karst.). Je známo, že genomy jehličnanů jsou velké 20-40 Gb a mají vysoký obsah repetitivních oblastí, a proto je seskupení genomu často rozděleno do mnoha relativně krátkých sekcí. Navíc z očekávané velikosti genomu chybí v referenčních genomech velké fragmenty (Bernhardsson *et al.* 2020).

V posledních desetiletích se neobyčejně rychle vyvinuly metody sekvenování nové generace (NGS), což vedlo k poklesu nákladů na megabazi (Mb). Tento rozvoj měl vliv na nárůst zájmu o sekvenování genomů. Tyto nové technologie a metody usnadnily, zrychlily a vylepšily mnoho výzkumů z oblasti biologie jako například transkriptomiku, genovou anotaci, identifikaci RNA stříhu a další.

Sekvenace a genotypizace jednoho vzorku pomocí NGS platform, včetně primární analýzy dat, se cenově pohybuje víceméně na srovnatelné úrovni jako při použití fragmentační analýzy. Metoda NGS byla vybrána, protože bylo předpokládáno, že oproti konvenčním separačním metodám, může přinést nepoměrně více informací a odstranit subjektivní chyby, jež vznikají při genotypizaci.

Nicméně s NGS platformami je spjata vyšší chybovost, protože jsou citlivější na kvalitu vstupních parametrů. Délky čtení jsou ve srovnání s konvenčními metodami kratší, a proto jsou výsledky mnohem náročnější na vyhodnocení. Smysluplná analýza NGS dat se přímo spoléhá na přesné určení SNP a genotypů (Nielsen *et al.* 2011).

I přes několik etap navrhování, optimalizování a testování primerů a multiplexů, se v poslední fázi této diplomové práce podařilo vytvořit funkční mikrosatelitový multiplex. Díky přidání primerů používaných již dříve ve výzkumech využívajících konvenčních metod (Bínová *et al.* 2020), se podařilo získat výstup z NGS, který byl porovnatelný s výsledky separace za pomoci fragmentační analýzy.

Ukázalo se, že NGS metody jsou velice přesné a dokáží přinést lepší výsledky než konvenční metody, zejména v množství dat. To však s sebou přináší jistou náročnost na práci s těmito daty, která je naprosto stěžejní a vyžaduje nespočet algoritmů, nástrojů a programů, které data zpracují do interpretovatelné formy. S tím je spojena i potřeba mít určité znalosti v oboru bioinformatiky, případně investovat nemalé prostředky do služeb expertů.

Proto se oproti původnímu plánu analýza nukleotidových sekvencí zredukovala na délkovou analýzu amplikonů. Avšak veškerá potřebná data jsou nyní dostupná, tudíž je možné v budoucnu data použít k přesnějším analýzám konkrétního pořadí nukleotidů. Nicméně k tomu bude potřeba bioinformatická expertíza.

Ukázalo se tedy, že modifikace analýzy mikrosatelitových markerů na platformy NGS je stále perspektivní, avšak je nutno dodat, že podléhá zatížením obdobnými chybami při genotypizaci, jako konvenční metody separace fragmentů.

8 Seznam literatury a použitých zdrojů

ALTSCHUL, Stephen F., Warren GISH, Webb MILLER, Eugene W. MYERS a David J. LIPMAN, 1990. Basic local alignment search tool. *Journal of Molecular Biology* [online]. **215**(3). ISSN 00222836. Dostupné z: doi:10.1016/S0022-2836(05)80360-2

ANDREWS, Simon, 2010. FastQC: a quality control tool for high throughput sequence data [online]. 2010. Dostupné z: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

ANSORGE, Wilhelm J., 2009. *Next-generation DNA sequencing techniques* [online]. 2009. ISSN 18716784. Dostupné z: doi:10.1016/j.nbt.2008.12.009

ARONESTY, Erik, 2013. Comparison of Sequencing Utility Programs. *The Open Bioinformatics Journal* [online]. **7**(1). ISSN 1875-0362. Dostupné z: doi:10.2174/1875036201307010001

BARBIAN, Hannah J., Andrew Jesse CONNELL, Alexa N. AVITTO, Ronnie M. RUSSELL, Andrew G. SMITH, Madhurima S. GUNDLAPALLY, Alexander L. SHAZAD, Yingying LI, Frederic BIBOLLET-RUCHE, Emily E. WROBLEWSKI, Deus MJUNGU, Elizabeth V. LONSDORF, Fiona A. STEWART, Alexander K. PIEL, Anne E. PUSEY, Paul M. SHARP a Beatrice H. HAHN, 2018. CHIIMP: An automated high-throughput microsatellite genotyping platform reveals greater allelic diversity in wild chimpanzees. *International Journal of Business Innovation and Research* [online]. **8**(16). ISSN 20457758. Dostupné z: doi:10.1002/ece3.4302

BERNHARDSSON, Carolina, Xi WANG, Helena EKLÖF a Pär K. INGVARSSON, 2020. Variant Calling Using Whole Genome Resequencing and Sequence Capture for Population and Evolutionary Genomic Inferences in Norway Spruce (*Picea Abies*). In: [online]. Dostupné z: doi:10.1007/978-3-030-21001-4_2

BESNARD, Guillaume, V. ACHÉRÉ, P. Faivre RAMPANT, J. M. FAVRE a S. JEANDROZ, 2003. A set of cross-species amplifying microsatellite markers developed from DNA sequence databanks in *Picea* (Pinaceae). *Molecular Ecology Notes* [online]. ISSN 14718278. Dostupné z: doi:10.1046/j.1471-8286.2003.00456.x

BÍNOVÁ, Zuzana, Jiří KORECKÝ, Jakub DVOŘÁK, Jan BÍLÝ, Dagmar ZÁDRAPOVÁ, Václav JANSKA a Milan LSTIBŮREK, 2020. Genetic structure of Norway spruce ecotypes studied by SSR markers. *Forests* [online]. **11**(1). ISSN 19994907. Dostupné z: doi:10.3390/f11010110

BRONNER, Iraad F. a Michael A. QUAIL, 2019. *Best Practices for Illumina Library Preparation* [online]. 2019. ISSN 19348258. Dostupné z: doi:10.1002/cphg.86

CVRČKOVÁ, Helena, Pavlína MÁCHOVÁ, Jaroslav DOSTÁL a Jana MALÁ, 2011. Sledování genetické proměnlivosti chlumního ekotypu smrku ztepilého pomocí RAPD. *Zprávy Lesnického Výzkumu*. **56**(2), 137–143.

DE BARBA, M., C. MIQUEL, S. LOBRÉAUX, P. Y. QUENETTE, J. E. SWENSON a P. TABERLET, 2017. High-throughput microsatellite genotyping in ecology: improved accuracy, efficiency, standardization and success with low-quantity and degraded DNA. *Molecular Ecology Resources* [online]. **17**(3). ISSN 17550998. Dostupné z: doi:10.1111/1755-0998.12594

EID, John, Adrian FEHR, Jeremy GRAY, Khai LUONG, John LYLE, Geoff OTTO, Paul PELUSO, David RANK, Primo BAYBAYAN, Brad BETTMAN, Arkadiusz BIBILLO, Keith BJORNSON, Bidhan CHAUDHURI, Frederick CHRISTIANS, Ronald CICERO, Sonya CLARK, Ravindra DALAL, Alex DEWINTER, John DIXON, Mathieu FOQUET, Alfred GAERTNER, Paul HARDENBOL, Cheryl HEINER, Kevin HESTER, David HOLDEN, Gregory KEARNS, Xiangxu KONG, Ronald KUSE, Yves LACROIX, Steven LIN, Paul LUNDQUIST, Congcong MA, Patrick MARKS, Mark MAXHAM, Devon MURPHY, Insil PARK, Thang PHAM, Michael PHILLIPS, Joy ROY, Robert SEBRA, Gene SHEN, Jon SORENSON, Austin TOMANEY, Kevin TRAVERS, Mark TRULSON, John VIECELI, Jeffrey WEGENER, Dawn WU, Alicia YANG, Denis ZACCARIN, Peter ZHAO, Frank ZHONG, Jonas KORLACH a Stephen TURNER, 2009. Real-time DNA sequencing from single polymerase molecules. *Science* [online]. **323**(5910), 133–138 [vid. 2022-01-15]. ISSN 00368075. Dostupné z: doi:10.1126/SCIENCE.1162986/SUPPL_FILE/EID.SOM.PDF

EIJA KORPELAINEN, Jarno TUIMALA, Panu SOMERVUO, Mikael HUSS a Garry WONG, 2015. *RNA-Seq Data Analysis: A practical approach*.

EWELS, Phil, 2018. MultiQC: tool that aggregate results from bioinformatics analyses across many samples into a single report [online]. 2018. Dostupné z: <https://multiqc.info/>

FLUCH, Silvia, Agnes BURG, Dieter KOPECKY, Andreas HOMOLKA, Nadine SPIESS a Giovanni G. VENDRAMIN, 2011. Characterization of variable EST SSR markers for Norway spruce (*Picea abies* L.). *BMC Research Notes* [online]. **4**. ISSN 17560500. Dostupné z: doi:10.1186/1756-0500-4-401

FOX, Brian, 2020. Bash software. *GNU Project* [online]. Dostupné z: <http://www.gnu.org/software/bash/>

GORDON, Assaf, 2009. FASTX Barcode Splitter [online]. 2009. Dostupné z: https://github.com/Novartis/yap/blob/master/bin/fastx_barcode_splitter.pl

GUICHOUX, E., L. LAGACHE, S. WAGNER, P. CHAUMEIL, P. LÉGER, O. LEPAIS, C. LEPOITTEVIN, T. MALAUSA, E. REVARDEL, F. SALIN a R. J. PETIT, 2011. Current trends in microsatellite genotyping. *Molecular Ecology Resources* [online]. **11**(4). ISSN 1755098X. Dostupné z: doi:10.1111/j.1755-0998.2011.03014.x

HODGETTS, R. B., M. A. ALEKSIUK, A. BROWN, C. CLARKE, E. MACDONALD, S. NADEEM a D. KHASA, 2001. Development of microsatellite markers for white spruce (*Picea glauca*) and related species. *Theoretical and Applied Genetics* [online]. **102**(8). ISSN 00405752. Dostupné z: doi:10.1007/s00122-001-0546-0

HOLUŠA, Otakar a Kateřina HOLUŠOVÁ, 2020. *Modely návrhu úprav dřevinné skladby porostních směsí s ohledem na ekologickou stabilitu a ekonomiku lesnického hospodaření v přírodních podmínkách Vysočiny.*

HOOGENBOOM, Jerry, Kristiaan J. VAN DER GAAG, Rick H. DE LEEUW, Titia SIJEN, Peter DE KNIJFF a Jeroen F.J. LAROS, 2017. FDSTools: A software package for analysis of massively parallel sequencing data with the ability to recognise and correct STR stutter and other PCR or sequencing noise. *Forensic Science International: Genetics* [online]. **27**. ISSN 18780326. Dostupné z: doi:10.1016/j.fsigen.2016.11.007

ILLUMINA, Inc., 2021. *History of Illumina Sequencing* [online] [vid. 2021-12-20]. Dostupné z: <https://www.illumina.com/technology/next-generation-sequencing/solexa-technology.html>

- JARAMILLO-CORREA, J. P., J. BOUSQUET, J. BEAULIEU, N. ISABEL, M. PERRON a M. BOUILLÉ**, 2003. Cross-species amplification of mitochondrial DNA sequence-tagged-site markers in conifers: the nature of polymorphism and variation within and among species in *Picea*. *Theoretical and Applied Genetics* [online]. **106**(8), 1353–1367. ISSN 0040-5752. Dostupné z: doi:10.1007/s00122-002-1174-z
- KOOLMAN, Jan a Klaus-Heinrich RÖHM**, 2012. *Barevný atlas biochemie*. 4. Praha: Grada. ISBN 978-80-247-2977-0.
- KRUEGER, Felix**, 2018. Trim Galore [online]. 2018. Dostupné z: https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
- LI, Heng, Bob HANDSAKER, Alec WYSOKER, Tim FENNELL, Jue RUAN, Nils HOMER, Gabor MARTH, Goncalo ABECASIS a Richard DURBIN**, 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* [online]. **25**(16). ISSN 13674803. Dostupné z: doi:10.1093/bioinformatics/btp352
- LITT, M. a J. A. LUTY**, 1989. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *American Journal of Human Genetics*. **44**(3). ISSN 00029297.
- MAGHULY, F., W. PINSKER, W. PRAZNIK a S. FLUCH**, 2006. Genetic diversity in managed subpopulations of Norway spruce [*Picea abies* (L.) Karst.]. *Forest Ecology and Management* [online]. **222**(1–3), 266–271. ISSN 03781127. Dostupné z: doi:10.1016/j.foreco.2005.10.025
- MARDIS, Elaine R.**, 2008. *Next-generation DNA sequencing methods* [online]. 2008. ISSN 15278204. Dostupné z: doi:10.1146/annurev.genom.9.081307.164359
- MASON, Annaliese S.**, 2015. Ssr genotyping. *Methods in Molecular Biology* [online]. **1245**. ISSN 10643745. Dostupné z: doi:10.1007/978-1-4939-1966-6_6
- MEN, Artem E., Peter WILSON, Kirby SIEMERING a Susan FORREST**, 2008. Sanger DNA Sequencing. In: *Next Generation Genome Sequencing: Towards Personalized Medicine* [online]. B.m.: Wiley-VCH Verlag GmbH & Co. KGaA, s. 1–11. ISBN 978-3-527-32090-5. Dostupné z: doi:10.1002/9783527625130.CH1

MERRIMAN, Barry, Ion TORRENT a Jonathan M. ROTHBERG, 2012. Progress in Ion Torrent semiconductor chip based sequencing. *ELECTROPHORESIS* [online]. **33**(23), 3397–3417 [vid. 2022-03-03]. ISSN 1522-2683. Dostupné z: doi:10.1002/ELPS.201200424

METZKER, Michael L., 2005. *Emerging technologies in DNA sequencing* [online]. 2005. ISSN 10889051. Dostupné z: doi:10.1101/gr.3770505

METZKER, Michael L., 2010. *Sequencing technologies the next generation* [online]. 2010. ISSN 14710056. Dostupné z: doi:10.1038/nrg2626

MINISTERSTVO ZEMĚDĚLSTVÍ ČESKÉ REPUBLIKY, 2020. *Zpráva o stavu lesa a lesního hospodářství České republiky v roce 2019*.

MINISTERSTVO ZEMĚDĚLSTVÍ ČESKÉ REPUBLIKY, 2021. *Zpráva o stavu lesa a lesního hospodářství České republiky v roce 2020*.

MULLER, B. H., P. MOLLON, E. SANTIAGO-ALLEXANT, F. JAVERLIAT a G. KANEKO, 2019. In-depth comparison of library pooling strategies for multiplexing bacterial species in NGS. *Diagnostic Microbiology and Infectious Disease* [online]. **95**(1). ISSN 18790070. Dostupné z: doi:10.1016/j.diagmicrobio.2019.04.014

MUSIL, Ivan a Jan HÁMERNÍK, 2003. *Lesnická dendrologie 1, Jehličnaté dřeviny*. Praha: Česká zemědělská univerzita v Praze. ISBN 80-213-0992-X.

NIELSEN, Rasmus, Joshua S. PAUL, Anders ALBRECHTSEN a Yun S. SONG, 2011. *Genotype and SNP calling from next-generation sequencing data* [online]. 2011. ISSN 14710056. Dostupné z: doi:10.1038/nrg2986

NYSTEDT, Björn, Nathaniel R. STREET, Anna WETTERBOM, Andrea ZUCCOLO, Yao Cheng LIN, Douglas G. SCOFIELD, Francesco VEZZI, Nicolas DELHOMME, Stefania GIACOMELLO, Andrey ALEXEYENKO, Riccardo VICEDOMINI, Kristoffer SAHLIN, Ellen SHERWOOD, Malin ELFSTRAND, Lydia GRAMZOW, Kristina HOLMBERG, Jimmie HÄLLMAN, Olivier KEECH, Lisa KLASSON, Maxim KORIABINE, Melis KUCUKOGLU, Max KÄLLER, Johannes LUTHMAN, Fredrik LYSHOLM, Totte NIITTYLÄ, Åke OLSON, Nemanja RILAKOVIC, Carol RITLAND, Josep A. ROSSELLÓ, Juliana SENA, Thomas SVENSSON, Carlos TALAVERA-LÓPEZ, Günter THEISSEN, Hannele TUOMINEN, Kevin VANNESTE, Zhi Qiang WU, Bo ZHANG, Philipp ZERBE, Lars ARVESTAD, Rishikesh BHALERAO, Joerg BOHLMANN, Jean BOUSQUET, Rosario GARCIA GIL, Torgeir R. HVIDSTEN, Pieter DE JONG, John MACKAY, Michele MORGANTE, Kermit RITLAND, Björn SUNDBERG, Stacey Lee THOMPSON, Yves VAN DE PEER, Björn ANDERSSON, Ove NILSSON, Pär K. INGVARSSON, Joakim LUNDEBERG a Stefan JANSSON, 2013. The Norway spruce genome sequence and conifer genome evolution. *Nature* [online]. **497(7451). ISSN 00280836. Dostupné z: doi:10.1038/nature12211**

OBIDOWICZ, Andrzej, Magdalena RALSKA-JASIEWICZOWA, Mirosława KUPRYJANOWICZ, Kazimierz SZCZEPANEK, Malgorzata LATALOWA a Dorota NALEPKA, 2004. *Picea abies* (L.) H. Karst. - Spruce. B.m. b.n.

PAREEK, Chandra Shekhar, Rafal SMOCZYNSKI a Andrzej TRETYN, 2011. *Sequencing technologies and genome sequencing* [online]. 2011. ISSN 12341983. Dostupné z: doi:10.1007/s13353-011-0057-x

PFEIFFER, Antonella, Angelo M. OLIVIERI a Michele MORGANTE, 1997. Identification and characterization of microsatellites in Norway spruce (*Picea abies* K.). *Genome* [online]. **40(4), 411–419. ISSN 08312796. Dostupné z: doi:10.1139/g97-055**

QUAIL, Michael A., Miriam SMITH, Paul COUPLAND, Thomas D. OTTO, Simon R. HARRIS, Thomas R. CONNOR, Anna BERTONI, Harold P. SWERDLOW a Yong GU, 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* [online]. **13(1). ISSN 14712164. Dostupné z: doi:10.1186/1471-2164-13-341**

RAJORA, O. P., M. H. RAHMAN, S. DAYANANDAN a A. MOSSELER, 2001. Isolation, characterization, inheritance and linkage of microsatellite DNA markers in white spruce (*Picea glauca*) and their usefulness in other spruce species. *Molecular and General Genetics* [online]. **264**(6). ISSN 00268925. Dostupné z: doi:10.1007/s004380000377

RHOADS, Anthony a Kin Fai AU, 2015. *PacBio Sequencing and Its Applications* [online]. 2015. ISSN 22103244. Dostupné z: doi:10.1016/j.gpb.2015.08.002

ROTHMAN, Robert H., 2010. *Determination of DNA Concentration and Purity by Ultraviolet Spectrophotometry* [online]. [vid. 2019-03-11]. Dostupné z: <https://people.rit.edu/rhrsbi/GEPages/LabManual.html>

RUNGIS, Dainis, Yanik BÉRUBÉ, Jun ZHANG, Steven RALPH, Carol E. RITLAND, Brian E. ELLIS, Carl DOUGLAS, Jörg BOHLMANN a Kermit RITLAND, 2004. Robust simple sequence repeat markers for spruce (*Picea* spp.) from expressed sequence tags. *Theoretical and Applied Genetics* [online]. **109**(6). ISSN 00405752. Dostupné z: doi:10.1007/s00122-004-1742-5

SCOTTI, I., F. MAGNI, R. FINK, W. POWELL, G. BINELLI a P. E. HEDLEY, 2000. Microsatellite repeats are not randomly distributed within Norway spruce (*Picea abies* K.) expressed sequences. *Genome* [online]. **43**(1). ISSN 08312796. Dostupné z: doi:10.1139/g99-095

SCOTTI, I., G. P. PAGLIA, F. MAGNI a Michele MORGANTE, 2002. Efficient development of dinucleotide microsatellite markers in Norway spruce (*Picea abies* Karst.) through dot-blot selection. *Theoretical and Applied Genetics* [online]. **104**(6–7). ISSN 00405752. Dostupné z: doi:10.1007/s00122-001-0843-7

SELKOE, Kimberly A. a Robert J. TOONEN, 2006. *Microsatellites for ecologists: A practical guide to using and evaluating microsatellite markers* [online]. 2006. ISSN 1461023X. Dostupné z: doi:10.1111/j.1461-0248.2006.00889.x

SLÁVIK, Martin, 2004. *Lesnická dendrologie*. B.m.: Česká zemědělská univerzita v Praze.

STURTEVANT, Alfred a Thomas MORGAN, 2001. *A History of Genetics* [online]. New York: Cold Spring Harbor Laboratory Press [vid. 2021-10-28]. ISBN 0-87969-607-9. Dostupné z: <http://www.esp.org/books/sturt/history/index.html>

SUEZ, Marie, Abdelkader BEHDENNA, Sophie BROUILLET, Paula GRAÇA, Dominique HIGUET a Guillaume ACHAZ, 2016. MicNeSs: Genotyping microsatellite loci from a collection of (NGS) reads. *Molecular Ecology Resources* [online]. **16**(2). ISSN 17550998. Dostupné z: [doi:10.1111/1755-0998.12467](https://doi.org/10.1111/1755-0998.12467)

TIBIHIKA, Papius Dias, Manuel CURTO, Eva DORNSTAUDER-SCHRAMMEL, Silvia WINTER, Esayas ALEMAYEHU, Herwig WAIDBACHER a Harald MEIMBERG, 2019. Application of microsatellite genotyping by sequencing (SSR-GBS) to measure genetic diversity of the East African *Oreochromis niloticus*. *Conservation Genetics* [online]. **20**(2). ISSN 15729737. Dostupné z: [doi:10.1007/s10592-018-1136-x](https://doi.org/10.1007/s10592-018-1136-x)

VARTIA, Salla, José L. VILLANUEVA-CAÑAS, John FINARELLI, Edward D. FARRELL, Patrick C. COLLINS, Graham M. HUGHES, Jeanette E.L. CARLSSON, David T. GAUTHIER, Philip MCGINNITY, Thomas F. CROSS, Richard D. FITZGERALD, Luca MIRIMIN, Fiona CRISPIE, Paul D. COTTER a Jens CARLSSON, 2016. A novel method of microsatellite genotyping-by-sequencing using individual combinatorial barcoding. *Royal Society Open Science* [online]. **3**(1). ISSN 20545703. Dostupné z: [doi:10.1098/rsos.150565](https://doi.org/10.1098/rsos.150565)

VENDRAMIN, G. G., L. LELLI, P. ROSSI a M. MORGANTE, 1996. A set of primers for the amplification of 20 chloroplast microsatellites in Pinaceae. *Molecular Ecology* [online]. ISSN 09621083. Dostupné z: [doi:10.1111/j.1365-294X.1996.tb00353.x](https://doi.org/10.1111/j.1365-294X.1996.tb00353.x)

VIEIRA, Maria Lucia Carneiro, Luciane SANTINI, Augusto Lima DINIZ a Carla de Freitas MUNHOZ, 2016. *Microsatellite markers: What they mean and why they are so useful* [online]. 2016. ISSN 16784685. Dostupné z: [doi:10.1590/1678-4685-GMB-2016-0027](https://doi.org/10.1590/1678-4685-GMB-2016-0027)

WALLINGFORD, R. A. a A. G. EWING, 1989. Capillary electrophoresis. *Advances in Chromatography* [online]. **29**, 1–76 [vid. 2022-02-27]. ISSN 0065-2415. Dostupné z: <https://europepmc.org/article/med/2667286>

WANG, Xinkun, 2016. *Next-Generation Sequencing Data Analysis* [online]. B.m.: CRC Press. ISBN 978-1-4822-1788-9. Dostupné z: doi:10.1201/b19532

WEBER, J. L. a P. E. MAY, 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *American Journal of Human Genetics*. **44**(3). ISSN 00029297.

WHITE, Timothy L., W. T. ADAMS a David B. NEALE, 2007. *Forest genetics*. B.m.: CABI. ISBN 9780851990835.

WOERNER, August E., Jonathan L. KING a Bruce BUDOWLE, 2017. Fast STR allele identification with STRait Razor 3.0. *Forensic science international. Genetics* [online]. **30**, 18–23 [vid. 2022-03-19]. ISSN 1878-0326. Dostupné z: doi:10.1016/J.FSIGEN.2017.05.008

WU, S a U MAMBER, 1992. *Agrep - a fast approximate pattern matching tool*. 1992.

WYSOKER, Alec, Andrew WHITWHAM, Bob HANDSAKER, Douglas VOET, James BONFIELD, Jennifer LIDDLE, John MARSHALL, Jue RUAN, Kathleen TIBBETTS, Luke JOSTINS, Petr DANECEK, Rob DAVIES, Tim FENNEL a Valeriu OHAN, 2009. SAM (Sequence Alignment/Map) is a flexible generic format for storing nucleotide sequence alignment. SAM tools provide efficient utilities on manipulating alignments in the SAM format. [online]. 2009. Dostupné z: <https://www.encodeproject.org/software/samtools/>

YANG, Jingjing, Jian ZHANG, Ruixi HAN, Feng ZHANG, Aijun MAO, Jiang LUO, Bobo DONG, Hui LIU, Hao TANG, Jianan ZHANG a Changlong WEN, 2019. Target SSR-seq: A novel SSR genotyping technology associate with perfect SSRS in genetic analysis of cucumber varieties. *Frontiers in Plant Science* [online]. **10**. ISSN 1664462X. Dostupné z: doi:10.3389/fpls.2019.00531

ZALAPA, Juan E., Hugo CUEVAS, Huayu ZHU, Shawn STEFFAN, Douglas SENALIK, Eric ZELDIN, Brent MCCOWN, Rebecca HARBUT a Philipp SIMON, 2012. *Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences* [online]. 2012. ISSN 00029122. Dostupné z: doi:10.3732/ajb.1100394

ZHAN, Luyao, Ian G. PATERSON, Bonnie A. FRASER, Beth WATSON, Ian R. BRADBURY, Praveen NADUKKALAM RAVINDRAN, David REZNICK, Robert G. BEIKO a Paul BENTZEN, 2017. megasat: automated inference of microsatellite genotypes from sequence data. *Molecular Ecology Resources* [online]. **17**(2). ISSN 17550998. Dostupné z: doi:10.1111/1755-0998.12561

ZHENG, Yi, Liangjun ZHAO, Junping GAO a Zhangjun FEI, 2011. IAssembler: A package for de novo assembly of Roche-454/Sanger transcriptome sequences. *BMC Bioinformatics* [online]. **12**. ISSN 14712105. Dostupné z: doi:10.1186/1471-2105-12-453

9 Seznam příloh

Příloha 1 - Nukleotidové sekvence 66 primerových párů z 1. fáze	106
Příloha 2 - Nukleotidové sekvence primerových párů přidaných ve 3. fázi.....	108
Příloha 3 - Počet readů daných primerů a vzorků ve 3. fázi	109
Příloha 4 - Set čtrnácti primerových párů - varianta A, 4. fáze	111
Příloha 5 - Set pětadvaceti primerových párů - varianta B, 4. fáze	112
Příloha 6 - Výsledky měření vstupní koncentrace přečištěných amplikonů před sekvenováním ve 4. fázi.....	114

10 Přílohy

Příloha 1 - Nukleotidové sekvence 66 primerových párů z 1. fáze

Název	Sekvence primerů (5'-3')		Repetice
	Forward primer (5'-3')	Reverse primer (5'-3')	
PA_1	AGGAGTTCCTGTGGAGGAA	TCCTGCACATTACCACTCGG	(ATAC) ₁₀
PA_2	TGGGTAGGTGCACTGGTCTA	ACACCTGGGTAAGTGCACAC	(ATAC) ₁₃
PA_3	CCCCCTCTAGATTTGAACTTGTA	TCCGCACAAAATTGTTTTCTACT	(CATA) ₁₀
PA_5	TCCTTCAAATCCACGATTCTCGA	TGTCAGCATTGGGATTTGAAGA	(CATA) ₁₀
PA_9	TCAAACGAAAGGGAATGACCT	TGGCTCTCTAAAATTGAAGTATGT	(AATA) ₁₀
PA_10	AGAGAATATGAGATGAGATGAAGTGA	TTGAAGGGGATGGGCATGTT	(TACA) ₁₂
PA_18	TGAATCACATGTACATAAAAGCATCAC	GATGTCAAGGGCATGAAGGC	(TATC) ₁₁
PA_19	ACCCAGGCTAGACCATGCTA	TGTGTTACTTTTGCTTCCGCA	(CATA) ₁₃
PA_22	TCACAATAGGGACATTTTGCCA	GGTCGCTTTTGTTGAGCCCT	(ATAC) ₁₁
PA_24	TGTACCTCAAGGAAACACCAGA	ACGTTACGCATCGTTTCCAAG	(ATAC) ₁₂
PA_25	CATGAGCTTCTTGGGGGAGG	CCAAGAGCGCACGAATCAAC	(TCAC) ₁₀
PA_27	TGAGTGCATTACATTTACAAGG	AGTGTTCACGCACACTTGAGT	(ATAC) ₁₄
PA_28	TGGTTTCCACTTGTCTGGCA	GTGATTATGGTCACAATAATGCAAAT	(ATAC) ₁₀
PA_29	GGAATTTCCCTCCCATAGCAGT	AGGGATGATGTTTCCCTTCCATGT	(TATC) ₁₀
PA_30	AGGTTGTGGGTTCCATTCCC	ACCATTAACAAATTATCCACACATGT	(GATA) ₁₀
PA_31	TCACTTTGTTATCAGCCCGT	CCGTTTGGTTTTTCACTTTTTGT	(TATC) ₁₃
PA_32	ACCTTCCACTCCTACTCCT	ACAACATGCTGCAGAAATGTGA	(ATAC) ₁₄
PA_33	TCTTCCAATCAGAAGTTATGCCA	TCAGGAGTTTAAGAAATCCATACCAGA	(TATC) ₁₀
PA_34	CGGGGACACACCACTAATC	TGTGTTTATTGTCCCTTCTGCA	(ATAC) ₁₀
PA_35	ACCCTCCGGAAATCTTGGTT	ACTTTCAAGCCCGTCCAAT	(ATAC) ₁₀
PA_36	GGGGATGTTACACCTTTCAATGC	TGCATATGTATGTATATGTATGCATGT	(ATAC) ₁₀
PA_37	AGCTACGATGGAAAGTAAATGGT	GCTAGCTTAAGACCAAGGCCCT	(TACA) ₁₀
PA_38	TCCATGCTCACATTTGATACTCT	GGTTCTGCAGAAGCTTGAGG	(CATA) ₁₁
PA_41	TCTTGTGACCTCAACACCA	GCATGGGCATATAAAAGAGACCT	(CATA) ₁₄
PA_42	ACAAAATGGATTAACACGTGCCT	TTCCACCCATAAAGTACCAAGA	(TATC) ₁₀
PA_43	ACTTGCATTAGAGTATTTTCAGGAAA	CCAACATCTAGGAAAGTTCTCATGC	(CATA) ₁₁
PA_44	CAGGTTGGGTCATGCCACTT	TGAGAATTGCATGTGCTTTTTGC	(TACA) ₁₅
PA_45	TGTATCATGTGACACCTACAAGG	AAACCTACACATGAGGCCCC	(TACA) ₁₂
PA_47	GGGAGTGAAAGTGATGGGGG	TGAGCATGGAAGGGATTACACA	(ATAC) ₁₀
PA_48	GGGATTGAGCCAGACGATGT	AGCATACACATACGCACGCA	(ATAC) ₁₆
PA_51	CCCAACGATGTATGCCGACT	TGTTCTAGGAGATCTAACCAACT	(ATAC) ₁₀
PA_53	TCGAAGCTTATGTGTCCCATTCT	ACCTCAGTGTGAGCCAAACA	(TATC) ₁₃
PA_54	CTTGAGGGGAGAACCCTGACG	TGGGTGAACGCAGAATCTCT	(TATC) ₁₂
PA_58	AGGAAAGCAATAAAAGGCACAACA	CCTTTCCTCTCTCCAGTGCT	(ATAC) ₁₄
PA_59	GGCCTGTTGGACAACCATAC	GCTTCTTCCATGTGCAACA	(TGTA) ₁₂
PA_60	CGGCATGAAATGAGAGCGGA	CCTCTACCTCGGGTGAGACT	(ATAC) ₁₀
PA_61	AGGGGGTAGAGACTTGACTTCA	AGAACCTCAACATCTCTGGCA	(TATC) ₁₁
PA_62	TTCCAACGTACATGTGGGCT	TCCCACATGTATTACTGGGATGT	(TATC) ₁₀
PA_64	AGAAACGAGGAATCCAACCCT	ATTGTTGGGTGAAGGGGGTC	(TCTA) ₁₁
PA_65	GCTGACCTGCTCGATCTCAA	TGGAACGTTGTTGGATTGGC	(CATA) ₁₁
PA_66	TGCGTGCAATTGATTCGGTT	TCTAGAGCCCAAGGAGGTCC	(TGTA) ₁₀

PA_67	ACATGGATAGGCACACTGGC	GGGTAAGTGCATGTGGTTGC	(ATAC) ₁₀
PA_69	TACAGGATGCGACTGGGGTA	AACTTGTGACCTCTTAGGAGAAA	(TATC) ₁₃
PA_70	CCTCTGGTGGCCGATGATTT	TCAGTCACCAGACCATCGGT	(ATAC) ₁₁
PA_71	ACCGCTTGTCGAGATTAACGA	TGGCTCCCTGGAAACCTAAG	(CATA) ₁₁
PA_72	AGTTTGTAACGTCCCCGACC	TGTGTTATTGTCACCTCCGCA	(ATAC) ₁₀
PA_73	AGAGTGGAAAGGAGGATGCA	CCACATTGTTCACTACTGAGC	(CATA) ₁₁
PA_74	TTGTTTAGACAATACCTTTGGTGTT	TCTTGGAGAAGTAGGGCCCA	(CATA) ₁₃
PA_75	AGGGGTGGAAGGTCTTCTCA	CTTAGCAGGGCTATGGCTCC	(TACA) ₁₀
PA_77	AGGGCTAAGTCCATCTCAACA	TGTCCCTATTGGTTTAGGCCA	(ATAC) ₁₅
PA_79	GCTAGGCAACGTCACAGAGT	TGTGTTATTGTCGCTTCCACA	(ATAC) ₁₁
PA_82	TGGGAATTCGCAAACCTGGGT	AACCCGACTCGTGAAATCCC	(TACA) ₁₁
PA_83	TCTGCCAGTTCGATGCATCT	TGGTACTTGCATACATAGTTATCTCT	(ATAC) ₁₃
PA_84	TGTGACGTACCAAGCCATCC	CACTTCCGCATGAATCTGTGT	(ATAC) ₁₄
PA_85	GCAATGGTTCATGTAAAAAGACACA	CCAGGAGGGTCGATAACGTT	(ATAC) ₁₁
PA_87	GGGACCACTTTGGAGTTCATC	ACTGGGATTACTCTGTATTGTTGA	(ATAC) ₁₀
PA_88	ATTGTGCACCCATCCACGAT	TCACGTTTCTCTTGCAGGGT	(TTTA) ₁₁
PA_89	AGTTGTAGTCCACAGCTGCC	TGAAGACTTGGCACTTACTCAT	(ATAC) ₁₀
PA_90	GGGCTCCCCAAATCCTTGAA	GCTGCATTATGTCTTGCCCA	(GATA) ₁₀
PA_91	TGGGCAGATGTACCAAGTGG	ACCACGCCAACTTTAGCAGT	(TATC) ₁₀
PA_92	GGTGGTTCACTGGCTTACGA	ACCATCAGAAGTCTGTCACAACA	(TACA) ₁₄
PA_93	CGTGCATTGATGATGTCTTCCA	AGACCTCATCAAGCCCATT	(TATC) ₁₀
PA_94	TGGAGCCCTTAGTATGTACCA	TGGATTTTTGTTAAGGATTTCTGTGT	(TAGA) ₁₁
PA_95	ACCAATCAAATGTTTGCAAAATCCT	TGGAGATTTGTCTTATGGTGGAA	(CATA) ₁₀
PA_96	ATCGTGGCTGCGTCTTGTTA	CCACTGGAAGCGGGATGAAT	(TAAA) ₁₀
PA_97	AGCCTAGGTGGTTGAGAGGA	TGTGCACCTAAAGGATATCTAAGGG	(ATAC) ₁₂

Příloha 2 - Nukleotidové sekvence primerových párů přidáných ve 3. fázi

Název	Autor	Sekvence primerů (5'-3')	Repetice	Bp
PAAC23	Scotti et al. 2000	F: TGTGGCCCCACTTACTAATATCAG	GT	300
		R: CGGGCATTGGTTTACAAGAGTTGC		
PAAC3	Scotti et al. 2000	F: CGCTACCTCAGATTTCTCCA	CA	260
		R: AGATATTCCCTCACAAAGTTGG		
paGB3	Besnard et al. 2003	F: AGTGATTAAACTCCTGACCAC	AT	117-153
		R: CACTGAATACACCCATTATCC		
Pt63718	Vendramin et al. 1996	F: CACAAAAGGATTTTTTTTCAGTG	T	92
		R: CGACGTGAGTAAGAATGGTTG		
Pt71936	Vendramin et al. 1996	F: TTCATTGGAAATACACTAGCCC	T	146
		R: AAAACCGTACATGAGATTCCC		
SpAG2	Pfeiffer et al. 1997	F: GCTCTTCACGTGTACTTGATC	TC	96-114
		R: TTCGAAGATCCTCCAAGATAC		
SpAGD1	Pfeiffer et al. 1997	F: GTCAACCAACTTGTAAGCCA	AG	147-164
		R: ACTTGTTTGGCATTTTCCC		
Pa_28	Fluch et al. 2011	R: GGCCGAAAGTGCTACTGCTA	TCG	145-170
		F: TGCTCCAGAAGAACAACACTCACA		
Pa_33	Fluch et al. 2011	R: GGTCGAGGAGGAGGAGGTAG	CGG	90-110
		F: CACCGCTAGTGCAGTCTCTG		
Pa_56	Fluch et al. 2011	F: ATCGTCTGCATTGCATTAC	AGGTG	113-128
		R: CTTCGTTCCCTTCTGATCCA		
WS0019.F22	Rungis et al. 2004	F: AAGCGTTTCTCATTCTTCTGG	AT	350-400
		R: GGGCCCAGAACTAACAATGA		
WS0023.B03	Rungis et al. 2004	F: AGCAGCTGGGGTCAAAGTT	AT	90-105
		R: AAAGAAAGCATGCATATGACTCAG		
WS0022.B15	Rungis et al. 2004	F: TTTGTAGGTGCTGCAGAGATG	AG	169-207
		R: TGGCTTTTTATTCCAGCAAGA		
WS00716.F13	Rungis et al. 2004	F: TCAAGTAATGGACAAACGATACA	GA	281-307
		R: TTTCCAATAGAATGGTGGATTT		
WS0073.H08	Rungis et al. 2004	F: TGCTCTTTATTTCGGGCTTC	GCAG	200-210
		R: AAGAACAAGGCTTCCAATG		
EATC1D02A	Scotti et al. 2002	F: TTGTCATCGTCGTCATTGTC	TCA	200-250
		R: TTTTAGCCTCTGTTTTCTAGCG		
EATC1E03	Scotti et al. 2002	F: CCOCTTATTCCTAAOGTCAA	AC	175
		R: TACCAGTGGTGACAACGATG		
WS0092.A19	Rungis et al. 2004	F: TGTGGTTTTCTGCTTGGAAA	AC	103
		R: CCCATTTTGACTTTGAAATAAGC		

Příloha 3 - Počet readů daných primerů a vzorků ve 3. fázi

primer/vzorek	sm3	sm5	sm9	sm32	sm33	sm40	sm86	sm88	sm89	sm97	sm99	sm100
FR_PA_1	8	8	3	1	17	11	9	0	2	6	7	23
FR_PA_10	0	0	0	0	1	0	0	1	0	0	0	1
FR_PA_18	14	12	12	15	14	11	11	15	2	19	14	5
FR_PA_19	7	5	2	1	0	6	1	2	0	0	5	0
FR_PA_2	19	50	15	9	37	33	16	21	9	15	15	16
FR_PA_22	0	0	0	14	1	0	0	1	0	1	15	0
FR_PA_24	0	0	0	0	0	0	1	0	0	0	0	0
FR_PA_25	3	13	15	2	5	12	7	43	1	6	5	18
FR_PA_27	0	0	0	0	0	0	0	13	1	0	1	0
FR_PA_28	0	0	0	0	0	0	0	0	0	0	0	0
FR_PA_29	1	1	1	0	0	0	2	1	1	2	2	0
FR_PA_3	0	0	0	0	0	0	0	0	0	0	0	0
FR_PA_30	0	0	0	0	0	0	0	0	0	0	0	0
FR_PA_31	0	0	0	2	3	0	2	1	0	1	1	1
FR_PA_32	0	0	0	0	0	0	0	0	0	0	0	0
FR_PA_33	0	0	0	0	2	0	0	0	0	0	0	0
FR_PA_34	0	0	0	0	0	0	0	0	0	0	1	0
FR_PA_35	0	0	0	0	0	0	0	0	0	0	0	0
FR_PA_36	0	0	0	0	1	0	0	1	0	0	0	0
FR_PA_37	0	0	0	0	1	1	0	0	0	1	0	0
FR_PA_38	0	1	0	0	0	0	0	0	0	3	0	1
FR_PA_41	0	0	0	2	3	1	2	0	0	0	2	1
FR_PA_42	0	0	1	0	0	0	0	0	0	0	0	0
FR_PA_43	0	0	0	0	1	0	0	0	0	0	0	0
FR_PA_45	1	2	1	0	1	1	1	1	1	1	0	1
FR_PA_47	0	0	0	1	2	1	0	0	0	0	1	0
FR_PA_48	1	2	0	0	2	0	0	0	0	1	0	1
FR_PA_5	0	0	0	0	0	0	0	0	0	0	0	0
FR_PA_51	0	0	0	1	0	0	1	1	0	0	0	1
FR_PA_53	16	39	28	34	50	33	38	48	16	47	26	22
FR_PA_54	0	0	0	0	0	0	1	0	0	0	0	0
FR_PA_58	20	14	12	13	35	40	18	27	7	21	17	9
FR_PA_59	19	12	8	6	21	14	3	11	13	7	6	15
FR_PA_60	0	0	2	1	1	1	2	1	0	2	0	0
FR_PA_61	57	61	32	67	69	46	90	53	18	56	81	29
FR_PA_62	0	0	0	0	0	0	0	0	0	0	0	0
FR_PA_65	336	366	280	348	519	484	424	442	132	411	276	350
FR_PA_66	18	5	16	26	35	7	2	23	1	5	14	26
FR_PA_67	215	243	207	216	379	255	132	239	147	184	191	62
FR_PA_69	63	58	4	8	13	2	3	14	30	11	13	47
FR_PA_70	0	0	0	1	0	0	0	0	0	0	0	0
FR_PA_71	0	1	0	0	0	2	0	0	0	0	1	0
FR_PA_72	1	3	0	1	3	4	5	11	11	1	4	2
FR_PA_73	0	0	0	1	0	1	0	1	0	0	0	0
FR_PA_74	4	1	2	6	3	4	2	4	1	3	2	6
FR_PA_75	1817	1396	1330	544	2005	1549	461	1596	897	612	1070	810
FR_PA_77	0	0	0	1	0	1	1	0	0	0	1	1
FR_PA_79	0	0	0	0	2	1	5	0	0	0	0	0
FR_PA_82	1417	1222	1356	1417	2633	1805	455	2473	857	974	1090	1544
FR_PA_83	0	1	0	0	0	1	1	0	0	0	0	2
FR_PA_84	15	41	10	16	23	18	6	28	8	5	3	7
FR_PA_85	0	0	0	1	0	1	0	0	0	0	0	0

FR_PA_88	19	8	6	33	5	8	62	36	13	11	32	12
FR_PA_89	0	0	0	0	0	0	0	0	0	0	0	0
FR_PA_9	0	0	0	0	0	0	0	0	0	0	0	0
FR_PA_90	4	8	4	6	5	21	18	11	4	7	22	3
FR_PA_91	9	16	10	11	25	21	21	5	3	9	14	5
FR_PA_92	0	0	0	0	0	0	0	0	0	0	0	0
FR_PA_93	1	6	0	0	0	0	0	0	0	0	0	0
FR_PA_94	0	2	0	0	1	1	0	3	0	1	0	0
FR_PA_95	1201	1875	1149	1610	2524	2200	1108	1911	664	1579	1133	1138
FR_PA_96	0	0	0	0	6	0	0	0	0	1	0	0
FR_PA_97	1	2	1	2	4	2	3	1	1	5	0	1
FR_PAAC23	0	0	0	0	0	0	0	0	0	0	0	0
FR_PAAC3	0	0	0	0	0	1	1	0	0	0	0	0
FR_paGB3	6	2	0	0	9	9	7	13	2	0	1	0
FR_Pt63718	385	445	379	363	501	446	88	368	591	279	255	221
FR_Pt71936	85725	63275	81613	56615	103369	73112	66726	95003	68365	81974	54837	44555
FR_SpAG2	84	191	161	57	329	242	36	217	90	137	148	92
FR_SpAGD1	27	36	0	6	31	0	49	102	40	0	55	11
FR_STR_Pa_28	35	53	48	32	82	63	31	47	22	36	34	40
FR_STR_Pa_33	249	267	234	230	248	250	61	255	96	178	122	79
FR_STR_Pa_56	248	335	292	286	515	458	157	352	171	295	257	267
FR_WS0019.F22	0	0	0	0	0	0	0	0	0	0	0	0
WS0023.B03	0	0	0	0	0	0	0	0	0	0	0	0
FR_WS0022.B15	0	0	0	0	2	1	0	0	0	0	0	0
FR_WS00716.F13	0	0	0	0	0	0	0	0	0	0	0	0
FR_WS0073.H08	0	0	0	0	0	0	0	0	0	0	0	0
FR_EATC1D02A	7	10	1	8	0	12	14	19	3	3	4	6
FR_EATC1E03	54	270	119	64	212	88	31	77	39	151	109	131
FR_WS0092.A19	0	0	0	0	0	0	0	0	0	0	0	0
Celkem	92107	70358	87354	62078	113750	81281	70115	103492	72259	87061	59887	49562

Příloha 4 - Set čtrnácti primerových párů (multiplex 1 a 2) – varianta A, 4. fáze

Název	Sekvence primerů (5'-3')	Repetice
paGB3	F: AGTGATTAAACTCCTGACCAC	(AT) ₁₁
	R: CACTGAATACACCCATTATCC	
SpAG2	F: GCTCTCACGTGTACTTGATC	(TC) ₁₆
	R: TTCGAAGATCCTCCAAGATAC	
EATC1E03	F: CCOCTTATTCCTAAOGTCAAA	(CAT) ₄ CGT
	R: TACCAGTGGTGACAACGATG	(CAT) ₈
WS00716.F13	F: TCAAGTAATGGACAAACGATACA	(GA) ₁₀
	R: TTTCCAATAGAATGGTGGATTT	
EATC1D02A	F: TTGTCATCGTCGTCATTGTC	TCA
	R: TTTTAGCCTCTGTTTTCTAGCG	
WS0073.H08	F: TGCTCTCTTATTTCGGGCTTC	(AT) ₁₄
	R: AAGAACAAGGCTCCCAATG	
Pa_28	F: GGCCGAAAGTGCTACTGCTA	(TCG) _n
	R: TGCTCCAGAAGAACAACACTCACA	
PAAC23	F: TGTGGCCCCACTTACTAATATCAG	(GT) ₁₄
	R: CGGGCATTGGTTTACAAGAGTTGC	
WS0092.A19	F: TGTGGTTTTCTGCTTGGA	(AC) ₉
	R: CCCATTTGACTTTGAAATAAGC	
WS0023.B03	F: AGCAGCTGGGGTCAAAGTT	(AT) ₁₀
	R: AAAGAAAGCATGCATATGACTCAG	
Pa_33	F: GGTCGAGGAGGAGGAGGTAG	(CGG) _n
	R: CACCGCTAGTGCAGTCTCTG	
WS0019.F22	F: AAGCGTTTCTCATTCTTGG	(AT) ₁₃
	R: GGGCCAGAACTAACAATGA	
PA_56	F: ATCGTCTGCATTGCATTAC	(AGGTG) _n
	R: CTTCGTTCTTCTGATCCA	
PAAC3	F: CGCTACCTCAGATTTCTCCA	(CA) ₁₆
	R: AGATATTCCTCACAAAGTTGG	

Příloha 5 - Set pětadvaceti primerových párů (vybrané z poolů B, C, D, F – knihovna č. 1) – varianta B, 4. fáze

Název	Sekvence primerů (5'-3')	Repetice
PA_45	F: TGTATCATGTGACACCTACAAGG	(TACA) ₁₂
	R: AAACCTACACATGAGGCCCC	
PA_47	F: GGGAGTGAAAGTGATGGGGG	(ATAC) ₁₀
	R: TGAGCATGGAAGGGATTACACA	
PA_48	F: GGGATTGAGCCAGACGATGT	(ATAC) ₁₆
	R: AGCATAACACATACGCACGCA	
PA_60	F: CGGCATGAAATGAGAGCGGA	(ATAC) ₁₀
	R: CCTCTACCTCGGGTGAGACT	
PA_71	F: ACCGCTGTGCGAGATTAACGA	(CATA) ₁₁
	R: TGGCTCCCTGGAAACCTAAG	
PA_72	F: AGTTTGTAACGTCCCCGACC	(ATAC) ₁₀
	R: TGTGTTATTGTCACCTCCGCA	
PA_79	F: GCTAGGCAACGTCACAGAGT	(ATAC) ₁₁
	R: TGTGTTATTGTCGCTTCCACA	
PA_10	F: AGAGAATATGAGATGAGATGAAGTGA	(TACA) ₁₂
	R: TTGAAGGGGATGGGCATGTT	
PA_38	F: TCCATGCTCACATTTGATACTCT	(CATA) ₁₁
	R: GGTTCCTGCAGAACTTGAGG	
PA_58	F: AGGAAAGCAATAAAAAGGCACAACA	(ATAC) ₁₄
	R: CCTTTCCTCTCTCCAGTGCT	
PA_85	F: GCAATGGTTCATGTAAAAAGACACA	(ATAC) ₁₁
	R: CCAGGAGGGTCGATAACGTT	
PA_90	F: GGGCTCCCCAAATCCTTGAA	(GATA) ₁₀
	R: GCTGCATTATGTCTTGCCCA	
PA_97	F: AGCCTAGGTGGTTGAGAGGA	(ATAC) ₁₂
	R: TGTGCACCTAAAGGATATCTAAGGG	
PA_42	F: ACAAATGGATTAACACGTGCCT	(TATC) ₁₀
	R: TTTCCACCCATAAAGTACCAAGA	
PA_44	F: CAGGTTGGGTCATGCCACTT	(TACA) ₁₅
	R: TGAGAATTGCATGTGCTTTTTGC	
PA_70	F: CCTCTGGTGGCCGATGATTT	(ATAC) ₁₁
	R: TCAGTCACCAGACCATCGGT	
PA_73	F: AGAGTGGAAAGGAGGATGCA	(CATA) ₁₁
	R: CCACATTGTTCACTTGTGAGC	
PA_96	F: ATCGTGGCTGCGTCTTGTTA	(TAAA) ₁₀
	R: CCACTGGAAGCGGGATGAAT	
PA_31	F: TCACTTTGTTATCAGCCCGT	(TATC) ₁₃
	R: CCGGTTTGGTTTTTCACTTTTTGT	
PA_37	F: AGCTACGATGGAAAGTAAATGGT	(TACA) ₁₀
	R: GCTAGCTTAAGACCAAGGCCT	
PA_43	F: ACTTGCATTAGAGTATTTTCAGGAAA	(CATA) ₁₁
	R: CCAACATCTAGGAAAGTTCTCATGC	
PA_5	F: TCCTCAAATCCACGATTCTCGA	(CATA) ₁₀
	R: TGTCAGCATTGGGATTTGAAGA	
PA_77	F: AGGGCTAAGTCCATCTCAACA	(ATAC) ₁₅
	R: TGTCCCTATTGGTTTAGGCCA	

PA_89	F: AGTTGTAGTCCACAGCTGCC	(ATAC) ₁₀
	R: TGAAGACTGGCACTTACTCAT	
PA_9	F: TCAAACGAAAGGGAATGACCT	(AATA) ₁₀
	R: TGGCTCTCTAAAACCTGAAGTATGT	

Příloha 6 - Výsledky měření vstupní koncentrace přečištěných ampliconů před sekvenováním ve 4. fázi

Lokalita	ID vzorku FLD	ID vzorku SEQme	Pool/multiplex	C (ng/ul)
Frýdlant	sm 3	A1	M1	15
			M2	9.52
			Pool A	2.48
			Pool B	2.38
			Pool C	2.44
			Pool F	4.08
	sm 5	A2	M1	12.7
			M2	7.78
			Pool A	2.26
			Pool B	4.42
			Pool C	3.24
			Pool F	2.08
	sm 9	A3	M1	17.9
			M2	10.3
			Pool A	3.46
Pool B			3.48	
Pool C			3.28	
Pool F			2.28	
Boubín	sm 32	A4	M1	14.8
			M2	7.72
			Pool A	4.58
			Pool B	4.48
			Pool C	3.18
			Pool F	2.42
	sm 33	A5	M1	15.8
			M2	9.68
			Pool A	2.48
			Pool B	6.27
			Pool C	4.14
			Pool F	2.06
	sm 40	A6	M1	17.9
			M2	7.78
			Pool A	4.92
Pool B			2.78	
Pool C			4.06	
Pool F			2.74	
Frýdek – Místek	sm 86	A7	M1	9.08
			M2	4.36
			Pool A	4.42
			Pool B	4.36
			Pool C	3.92
			Pool F	4.63

	sm 88	A8	M1	12
			M2	4.48
			Pool A	7.78
			Pool B	4.58
			Pool C	3.96
			Pool F	2.92
	sm 89	A9	M1	11.9
			M2	6.26
			Pool A	3.48
			Pool B	2.68
Pool C			3.8	
Pelhřimov	sm 97	A10	M1	12.5
			M2	6.46
			Pool A	4.43
			Pool B	9.52
			Pool C	3.94
			Pool F	2.86
	sm 99	A11	M1	13.9
			M2	10.6
			Pool A	4.46
			Pool B	2.38
			Pool C	3.52
			Pool F	2.34
	sm 100	A12	M1	12.8
			M2	4.42
			Pool A	2.98
Pool B			4.78	
Pool C			3.2	
Pool F			3.38	