**Univerzita Palackého v Olomouci**

# Diplomová práce

**Olomouc 2018** **Bc. Lucie Bílková**

**Univerzita Palackého v Olomouci**

**Přírodovědecká fakulta**

**Katedra buněčné biologie a genetiky**

# Rekonstrukce kompletní chloroplastové DNA u vybraných rostlinných druhů a jejich komparativní analýza

**Diplomová práce**

**Bc. Lucie Bílková**

Studijní program: Biologie

Studijní obor: Molekulární a buněčná biologie

Forma studia: Prezenční

**Olomouc 2018**

**Vedoucí práce: Mgr. Eva Hřibová, Ph.D.**

Prohlašuji, že jsem diplomovou práci vypracovala samostatně pod vedením Mgr. Evy Hřibové, Ph.D. a za použití uvedených literárních zdrojů.

V Olomouci Lucie Bílková

## Souhrn

Předložená diplomová práce se zabývá rekonstrukcí kompletní chloroplastové DNA u vybraných taxonů rodu *Dactylorhiza* a jejich anotací a následnou komparativní analýzou.

Předmětem teoretické části diplomové práce bylo vypracovat literární rešerši, která se zaměřuje na strukturu a organizaci genomu vyšších rostlin, zejména na mimojadernou DNA, konkrétně chloroplastovou DNA, její využití a sekvenační přístupy využívané pro sestavení celogenomové sekvence chloroplastové DNA.

V praktické části byla analyzována Next-Gen data (Illumina sekvence), provedena *de-novo* rekonstrukce chloroplastové DNA s následným *in silico* i experimentálním ověřením. Byla rovněž provedena anotace kompletní chloroplastové DNA a fylogenetická analýza. Podařilo se zrekonstruovat celkový genom chloroplastové DNA u osmi studovaných taxonů rodu *Dactylorhiza*. U druhu *Dactylorhiza fuchsii* subsp. *soóana* se povedlo složit celkový clDNA genom v jednom dlouhém *scaffoldu*. Pouze u jediného druhu, *Dactylorhiza bohemica*, se nepodařilo složit celkový genom clDNA. Na základě provedených analýz bylo zjištěno, že velikost chloroplastové DNA studovaných taxonů rodu *Dactylorhiza* je 154 113-156 724 kb. Proteiny kódující geny, tRNA a rRNA tvořily zhruba 70 % celého genomu chloroplastu. Z toho připadalo zhruba 48 % na proteiny kódující geny, 19 % na geny pro tRNA a 3 % na geny pro rRNA. Zbylých 30 % genomu tvořily inter-genové mezerníky, introny a pseudogeny. V oblasti velké kódující podjednotky bylo přítomno přibližně 47 % genů, v oblasti malé kódující podjednotky přibližně 9 % genů a v oblastech duplikace přibližně celkem 44 % genů.

## Summary

The thesis deals with use of partial illumina sequencing data of total genomic DNA for *de-novo* assembly of a complete chloroplast genome. Paired-end illumina sequences of nine selected representatives of *Dactylorhiza* genus and two different assembly programs using different computing algorithms were used to verify this task.

The theoretical part of the master thesis focuses a genome structure and organization of higher plants, especially extracellular DNA, namely chloroplast DNA; sequencing techniques and bioinformatics tools used for reconstruction of whole chloroplast genome sequence as well as utilization of clDNA sequences in phylogenetic studies.

The practical part of the thesis was focused on analyzes Next-Gen data (Illumina sequencing), *de-novo* assembly of chloroplast DNA and a subsequent *in silico* analysis and experimental verification. Out of nine analyzed representatives, complete chloroplast genome was reconstructed in eight of them. By my approach, I was not able to reconstruct whole genome sequence only in one analyzed dataset of *Dactylorhiza bohemica*. Based on the analyses carried out in this thesis, it was discovered that the size of the chloroplast DNA of the studied *Dactylorhiza* genus taxa is 154 113-156 724 kb. Protein-coding genes, tRNA and rRNA, represented approximately about 70% of the whole chloroplast genome. About 48% of chloroplast genome was specific to protein-coding genes, 19% to tRNA genes, and 3% to rRNA genes. The remaining 30% of the genome was composed of intergenic spacers, introns, and pseudogenes. Approximately about 47% of the genes were present in the large single copy regions, 9% in the small single copy regions, and 44% in the inverted repeats region.

The results obtained in the diploma thesis can be used in future for analysis of evolutionary relationships in *Dactylorhiza* as well as in broader groups of orchids.

Ráda bych poděkovala Mgr. Evě Hřibové, Ph.D. za její odborné rady, čas, který mi věnovala, ale také za ochotu, trpělivost a vstřícný přístup při vypracování mé diplomové práce.
Tato práce vznikla za podpory projektů CERIT Scientific Cloud (LM2015085) a CESNET (LM2015042) financovaných z programu MŠMT Projekty velkých infrastruktur pro VaVaI.

# Obsah

# 1 Úvod

Chloroplastový genom obsahuje geny, které se primárně účastní fotosyntézy, transkripce a translace. Obsah genů a obecná struktura chloroplastových genomů v suchozemských rostlinách jsou značně zachované. Typický plastom je čtyřdílný s dlouhou jednokopiovou oblastí (LSC, ~80 kb) a krátkou jednokopiovou oblastí (SSC, ~20 kb), které jsou oddělené dvěma identickými obrácenými duplikacemi (IR, ~25 kb), jež jsou nezbytná pro stabilizaci struktury genomu plastidu. IR jsou nejvíce zachované prvky plastomu. Ačkoliv jsou plastidové genomy relativně malé, kódující sekvence obsahují pouze 50 % plastomů suchozemských rostlin. Typické plastidové genomy vyšších rostlin kódují všechny typy rRNA (23S, 16S, 5S, 4,5S), 27 až 31 genů tRNA a řadu proteinů (např. ~85 fotosyntetických proteinů). Asi 45 vysoce konzervovaných genů bylo nalezeno v plastomech téměř všech fotosyntetických organismů. Plastidy také obsahují řadu nekódujících RNA, včetně mnoha antimediátorových RNA.

Konzervovaná struktura chloroplastové DNA a absence rekombinace, nízký stupeň mutací společně s dědičností většinou po mateřské linii, činí chloroplastovou DNA častým markerem využívaným ve fylogenetických studiích rostlinných druhů. Vysoký počet kopií chloroplastů na jednu buňku přispívá k tomu, že je clDNA pro sekvenování snazším cílem než nízkokopiové jaderné geny, zejména pak z malých nebo degradovaných vzorků. Přestože se více uplatňují variace DNA v jaderném genomu oproti plastidovým genům, mnohé oblasti výzkumu, jako je fylogenetika a fylogeografie, budou sekvence chloroplastů nadále využívat, ať už z technických nebo biologických důvodů.

Vývoj metod sekvenování nové generace poskytl rychlejší a levnější metody sekvenování genomů. Sekvenování nové generace představují metody tzv. masivně paralelního sekvenování, kdy v jediném sekvenačním cyklu lze získat obrovský objem dat ekonomickým způsobem. Technologie sekvenování DNA nové generace mají v současné době nenahraditelné místo ve výzkumu a přicházejí i do oblasti klinické praxe. Sekvenační přístroje produkují velké množství dat, jejichž analýza metodami bioinformatiky je nezbytná k získání relevantních výsledků. Sekvenování se tak bez pokročilého výpočetního zpracování specializovanými algoritmy neobejde.

# 2 Cíle práce

1. Analýza sekvenačních dat druhé generace (Illumina sekvencí) vybraných rostlinných druhů.
2. *De-novo* rekonstrukce chloroplastové DNA pomocí různých programů ("assemblerů") a porovnání získaných výsledků.
3. Experimentální ověření sestavených oblastí (*kontigů* nebo *skafoldů*) a překlenutí nespojených úseků chloroplastového genomu pomocí PCR a následného Sangerova sekvenování.
4. Anotace a komparativní analýza kompletní chloroplastové DNA.

# 3 Literární přehled

## 3.1 Struktura a organizace genomu vyšších rostlin

Převážná většina dědičné informace u vyšších rostlin, stejně jako u ostatních vyšších eukaryot, je uložena v buněčném jádře. Kromě jaderné DNA, je dědičná informace vyšších rostlin uložena také v semiautonomních organelách – chloroplastech a mitochondriích [Gill a kol. 2008, Heslop-Harrison a Schwarzacher 2011].

### 3.1.1 Organizace a velikost jaderných genomů

Jaderný genom rostlin, složený z DNA a asociovaných proteinů, je organizován do jednotlivých chromozomů. Stejně jako velikost genomu se počet chromozomů v rostlinných druzích značně liší a může se pohybovat v rozmezí od 4 do více než 1000, přičemž pro polyploidní druhy jsou charakteristické genomy s vyššími počty chromozomů. Oproti tomu, počet chromozomů daného druhu, s výjimkou nadpočetných B chromozomů, je obvykle konstantní [Jones a kol. 2008], ale některé taxony, jako např. ty z rodiny *Cruciferae*, však mohou mít velmi proměnlivé počty chromozomů [Jeelani a kol. 2013].

Jednou ze základních charakteristik jaderných genomů je jejich velikost, která je v případě rostlinných druhů velmi rozlišná a pohybuje se v rozmezí od 0,063 Gb do 148,8 Gb, což je 2400násobný rozdíl [Dodsworth a kol. 2015; Kelly a Leitch 2011, Pellicer a kol. 2010]. Dlouho se jako nositel největšího genomu udával řebčík asyrský (*Fritillaria assyriaca*) s 127 Gbp (127,40 pg DNA/1C) [Leitch a kol. 2005]. Dnes je zřejmé, že publikovaný údaj byl nadhodnocený a navíc se nejspíše vztahoval k blízce příbuznému druhu *F. uva-vulpis*. Největší známý genom rostlin byl zjištěn u vraního oka japonského (*Paris japonica*) s 148 852 Mbp (152,20 pg DNA/1C) [Pellicer a kol. 2010]. Naopak nejmenší genom, o velikosti 63 Mbp (0,065 pg DNA/1C), byl objeven u genlisei zlaté (*Genlisea aurea)* [Greilhuber a kol. 2006]. Například genomy nejdůležitějších plodin jsou středně velké a spadají mezi tyto dva extrémy: *Oryza sativa* má velikost genomu 489 Mbp (0,50 pg DNA/1C), *Zea mays* 2 665 Mbp (2,73 pg DNA/1C) a *Triticum aestivum* 16 944 Mbp (17,33 pg DNA/1C) [http://data.kew.org/cvalues/]. V rámci této diplomové práce jsem se zabývala rodem *Dactylorhiz*a, který obsahuje jak diploidní tak polyploidní zástupce. Velikost genomu diploidního druhu *Dactylorhiza fuchsii* je ~2 826 Mbp (2,89 pg DNA/1C) a ~3 467 Mbp (3,55 pg DNA/1C) u diploidního zástupce *Dactylorhiza incarnata.*

Tetraploidní druh *Dactylorhiza lapponica* má velikost genomu ~6 538 Mbp (což odpovídá ~335 pg DNA/1Cx) [Aagaard a kol. 2005].

Množství DNA v organismu je udáváno jako tzv. "C-hodnota" [Swift 1950], která představuje obsah DNA haploidního genomu a vyjadřuje se v párech bází (bp). Následně bylo zjištěno, že neexistuje žádný vztah mezi C-hodnotou DNA a složitostí organismů [Mirsky a Ris 1951]. Nedostatek korelace byl Thomasem (1971) později nazýván paradoxem C-hodnoty.

Jednou z nejčastějších metod využívaných pro stanovení velikosti genomu rostlin je průtoková cytometrie [Doležel a Bartoš 2005]. Pro analýzu obsahu jaderné DNA u rostlin byla vypracována řada postupů. Ty se sestávají z přípravy suspenzí intaktních jader, barvených pomocí fluorochromů specifických pro DNA, a analýzy relativní intenzity fluorescence jader pomocí průtokového cytometru [Doležel a kol. 2007]. Protože průtoková cytometrie analyzuje relativní intenzitu fluorescence a tím relativní obsah DNA, může být velikost genomu neznámého vzorku stanovena až po porovnání s jádry referenčního standardu, jehož velikost genomu je známá [Doležel a Bartoš 2005]. Průtoková cytometrie má oproti předchozím metodám (biochemická extrakce a dvě metody založené na mikroskopii - mikrospektrometrie a kvantitativní cytofluororometrie) výhodu v jednoduchosti přípravy vzorků (posekání rostlinného pletiva žiletkou), vysoké citlivosti, nízké destruktivnosti (pouze malá část rostliny je analyzována), rychlosti analýzy (tisíce buněk (jader) mohou být analyzovány během několika minut) a relativně nízkým nákladům na analýzy [Suda 2005]. Databáze velikostí genomů rostlin (C-hodnot) udržuje Královská botanická zahrada v Kew ve Velké Británii [http://data.kew.org/cvalues/].

### 3.1.2 Rozdíly ve velikosti jaderných genomů

Jak již bylo zmíněno, velikost jaderných genomů vyšších rostlin se velmi liší. Počet genů kódujících proteiny se v genomech vyšších rostlin výrazně neliší, odhady se pohybují přibližně od 30 000 do zhruba 60 000 genů [Ming a kol. 2008, Yu a kol. 2002], v závislosti na druhu. Příčinou tak vysoké variability ve velikostech jaderných genomů je především přítomnost různého množství repetitivních DNA sekvencí [Bennetzen a kol. 2005, Hawkins a kol. 2008, Hřibová a kol. 2010, Kelly a Leitch 2011, Macas a kol. 2007, Renny-Byfield a kol. 2011, Swaminathan a kol. 2007, Vitte a Panaud 2005, Wicker a kol. 2009]. Dalším faktorem, který stojí za vysokou variabilitou ve velikostech genomu vyšších rostlin, je fakt, že evoluce velkého množství rostlinných druhů byla doprovázena četnými vnitro-

nebo mezi-druhovými hybridizacemi, a nebo polyploidizací [Adams a Wendel 2005, Jiao a kol. 2011, Leitch a Leitch 2008, Leitch a kol. 2008, Soltis a kol. 2009, Wendel 2000].

Repetitivní elementy v genomech se skládají z tandemových repetic a rozptýlených mobilních elementů (například transpozonů a retrotranspozonů). U krytosemenných rostlin jsou tyto repetice rozmanité a početné; zaujímají 70-80 % jaderné genomové DNA (gDNA), čímž činí kvetoucí rostliny vhodnou skupinou ke studiu evoluce dynamiky repetitivních elementů [Hansen a Heslop-Harrison 2004, Kelly a kol. 2012, Leitch a Leitch 2008, Wicker a kol. 2007]. Pozorování, že několik rodin [Hawkins a kol. 2006, Piegu a kol. 2006], nebo dokonce jednotlivé rodiny [Neumann a kol. 2006] transponovatelných elementů (TEs) mohou převažovat v rostlinných genomech a zodpovídat za variabilitu ve velikosti genomu mezi blízce příbuznými druhy, vedlo k návrhu, že rozdílná náchylnost k amplifikaci TE hraje primární roli v řízení změn ve velikosti genomu [Grover a Wendel 2010].

Přestože je většina rozmanitosti výsledkem rozdílné expanze a ztráty repetic, v průběhu evoluce se výrazně podílí na výsledné velikosti jaderného genomu proces celogenomové duplikace (WGD), také známé jako polyploidie a paleopolyploidie [Proost a kol. 2011, Soltis a Burleigh 2009]. Polyploidizace je již dlouho považována za významný mechanismus speciace u rostlin, zejména krytosemenných rostlin [Otto a Whitton 2000, Soltis a Soltis 2000, 2009]. Všechny krytosemenné rostliny prošly nejméně jedním kolem (ne-li více) celogenomové duplikace [Vision a kol. 2000, Bowers a kol. 2003, Jaillon a kol. 2007, Jiao a kol. 2011] a proto sdílejí staré WGD, stejně jako všechny semenné rostliny [Jiao a kol. 2011]. Studie ukázaly, že dochází jednak ke zvětšování genomů prostřednictvím akumulace repetitivní DNA [Renny-Byfield a kol. 2011], ale také ke zmenšování genomů ("genome downsizing") ztrátou DNA, která následuje po polyploidizaci [Leitch a Bennett 2004, Leitch a kol. 2008]. Polyploidi, kteří vzniknou zdvojením chromozomů téhož druhu či znásobením celého vlastního genomu, jsou označovaní jako autopolyploidi. Polyploidní jedinci, kteří vznikají křížením mezi různými druhy, se nazývají alopolyploidi. Tento proces může vyvolat rychlé, opakovatelné a směrodatné změny subgenomů předků [Comai a kol. 2003, Feldman a Levy 2009, Chen a Ni 2006, Lim a kol. 2004, Liu a Wendel 2003, Matyasek a kol. 2007]. *Dactylorhiza* spp. také prošla hybridizací a polyploidizací; mnoho euroasijských druhů tvoří polyploidní komplex. Tyto aloploidní druhy se vyvinuly opakovaně hybridizací mezi dvěma široce rozšířenými rodičovskými liniemi: diploidní ($2n = 40$) *D. incarnata* a *D. maculata* (včetně diploidní *D. fuchsii*) [Balao a kol. 2016, Hedrén a kol. 2012, Naczk a kol. 2015, Paun a kol. 2011].

## 3.2 Mimojaderná DNA

V roce 1909 dvě publikace od Corrense a Baura, publikované ve svazku 1 Zeitschrift für induktive Abstammungs- und Vererbungslehre (nyní Molekulární genetika a genomika), se zabývaly nemendelovskou dědičnou deficiencí chlorofylu [Correns 1909, Baur 1909]. Tyto dokumenty uvádějí první příklady mimojaderné dědičnosti a položily základy nové oblasti výzkumu: mimojaderné genetice. Correns pozoroval čistě mateřský typ dědičnosti (u rodu *Mirabilis*), zatímco Baur našel biparentální dědičnost (u rodu *Pelargonium*). Baur následně vyvinul teorii dědičnosti plastidů [Baur 1910, 1911]. V mnoha rodech jsou plastidy přenášeny matkou pouze uniparentálně, zatímco u několika rodů dochází k biparentální dědičnosti plastidů. Obvykle dochází k náhodnému třídění plastidů během ontogenetického vývoje. Renner a Schwemmle, stejně jako genetici z jiných zemí, přidali k této teorii další podrobnosti [Renner 1922, 1924, 1929, 1934, 1936, Schwemmle 1940, 1941, 1943, 1957]. Průkopnické studie se zabývaly mitochondriální dědičností u kvasinek, kdy byly prokázány buňky s deficiencí respirace (petity v kvasinkách, poky u *Neurospora*) díky mitochondriálním mutacím [Ephrussi 1949, Slonimski a Ephrussi 1949]. Elektronová mikroskopie a biochemické studie následně ukázaly, že plastidy a mitochondrie obsahují organelově specifické DNA molekuly [Beale a Knowless 1978, Gibor a Izawa 1963, Granick a Gibor 1967, Hagemann 1968, Kirk 1963, 1986, Kirk a Tilney-Bassett 1967, Leff a kol. 1963, Ris 1961, Ris a Plaut 1962, Sager 1972, Sager a Ishida 1963]. Tato zjištění položila molekulární základ pro mimojadernou dědičnost - plastidovou a mitochondriální genetiku.

### 3.2.1 Mitochondriální DNA

Mitochondriální genomy rostlin se značně se liší ve velikosti, dokonce i mezi velmi blízkými druhy nebo v rámci druhu [Huot a kol. 2014, Lohse a kol. 2013]. Rozdíly ve velikosti rostlinné mitochondriální DNA (mtDNA) jsou 22násobné, s 1,5násobným rozdílem v množství genů [Gualberto a kol. 2014]. Velikost mitochondriální DNA u krytosemenných rostlin se pohybuje obvykle v rozmezí 200-750 kb (průměrně 484 kb) [Lohse a kol. 2013, Richardson a kol. 2013], vyskytují se však v několika rodech výjimky, kde pozorujeme obrovský nárůst ve velikosti [Sugiyama a kol. 2005]. Velikost mitochondriálního genomu semenných rostlin kolísá v rozmezí 200-2900 kb [Alverson a kol. 2011, Gualberto a kol. 2014, Sloan a kol. 2012].

Obrovská diverzita ve velikosti rostlinné mitochondriální DNA není způsobena množstvím genů, ale přítomností značného množství nekódující DNA v genomech (tj. rozptýlených repetic, intronů, inter-genových mezerníků a cizích sekvencí DNA). Důležitou charakteristikou rostlinných mitogenomů je přítomnost repetitivních sekvencí DNA lišících se různou velikostí i počtem opakování [Alverson a kol. 2011]. Tyto repetitivní DNA sekvence jsou často klasifikovány na základě své délky - dlouhé repetice (> 500 párů bází), které mohou být zapojeny do časté homologní rekombinace; střední repetice (50-500 párů bází), které se podílejí na vzácné ektopické homologní rekombinaci; a krátké repetice (<50 párů bází), které mohou podpořit nelegitimní mikrohomologicky zprostředkovanou rekombinaci [Arrieta-Montiel a kol. 2009, Davila a kol. 2011, Gualberto a kol. 2014].

Mitochondriální DNA vyšších rostlin kóduje 60-70 genů [Oudot-Le Secq a kol. 2011]. Počet mitochondriálních genů u krytosemenných rostlin se u různých druhů velmi liší (od 32 do 67), což odráží ztráty a přenos genů do jaderného genomu během evoluce [Kubo a Newton 2008, Paux a kol. 2006, Rice a kol. 2013, Richardson a kol. 2013, Sugiyama a kol. 2005, Yurina a Odintsova 2016]. mtDNA kóduje několik genů mitochondriálního elektronového transportního řetězce. Pro expresi těchto genů má mitochondriální genom svůj vlastní translační systém, který je také částečně kódován mtDNA, včetně rRNA, tRNA a proměnného počtu ribozomálních proteinů, které závisí na druhu organismu [Lohse a kol. 2013, shrnuto v Kubo a Newton 2008]. Několik proteinů podílející se na sestavení funkčních respiračních komplexů může být také kódováno rostlinnou mtDNA. Avšak všechny faktory potřebné pro udržení mtDNA a expresi jejich genů jsou kódovány jádrem a importovány z cytosolu, čímž se replikace, strukturní organizace a exprese genů mtDNA dostává pod kontrolu jádra [shrnuto v Huot a kol. 2014]. Horizontální přenos je zodpovědný za získávání exogenních sekvencí a část rostlinných mitogenomů může být považována za odvozenou z chloroplastové, jaderné nebo virové DNA, ale většina nekódujících sekvencí je neznámého původu [Bergthorsson a kol. 2003].

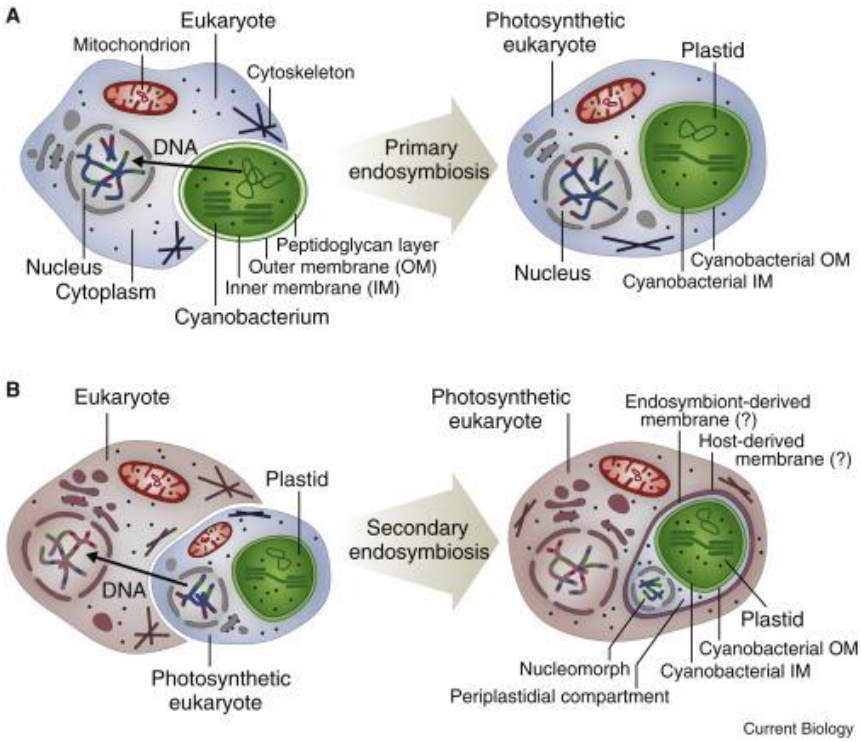
### 3.2.2 Chloroplastová DNA

#### 3.2.2.1 Chloroplasty a jejich původ

Přestože je jako klíčová funkce plastidů často uznávána fotosyntéza, hrají plastidy důležitou roli i v dalších aspektech fyziologie a vývoje rostlin, včetně syntézy aminokyselin, nukleotidů, mastných kyselin, fytohormonů, vitamínů a množství metabolitů a asimilace síry

a dusíku. Metabolity, které jsou syntetizovány v chloroplastech, jsou důležité pro interakce rostlin s jejich prostředím (reakce na teplo, sucho, sůl, světlo apod.) a jejich obranu proti napadení patogeny. Chloroplasty slouží jako centra metabolismu při buněčných reakcích na signály a reagují pomocí retrográdní signalizace [Bobik a Burch-Smith 2015, Daniell a kol. 2016].

Dnes je všeobecně přijímána endosymbiotická teorie původu plastidů v buňkách rostlin a řas (Obr. 1). Plastidy vznikly před více než miliardou let v důsledku symbiózy mezi nefotosyntetickými eukaryotickými buňkami a fotosyntetickými sinicemi [Xu a kol. 2015]. Vznik plastidů a nakonec rostlin a řas měl obrovský vliv na vývoj života na Zemi. Plastidy rostlin a zelených řas (chloroplasty), červených řas (rodoplasty) a glaukofytů (cyanely nebo cyanoplasty) pocházely z primární symbiózy. Rostliny z několika fotosyntetických eukaryotických taxonů obsahují plastidy, které se objevily ze sekundární nebo terciární symbiózy, tj. symbiózy nefotosyntetických eukaryot s volně žijícími fotosyntetickými eukaryoty (např. červené nebo zelené řasy) [Jensen a Leister 2014]. Sekundární plastidy jsou typické pro *Cryptophyceae*, *Chlorarachnea*, *Haptophyta*, *Euglenida*, většinu *Dinoflagellates* a *Diatomae*. *Diatomae* a *Stramenopila* (*Heterokonta*) byly tvořeny symbiózou zelených a červených řas [Green 2011, Howe a kol. 2008, Ruck a kol. 2014]. Terciární plastidy vznikly, když fotosyntetická eukaryota ztrácela sekundární plastidy a nahrazovala je plastidy z jiných symbiontů. Terciární plastidy byly nalezeny u *Dinoflagellata* [Howe a kol. 2008]. Na rozdíl od primárních plastidů, které vznikly symbiózou mezi eukaryoty a prokaryoty a jsou obklopeny dvěma membránami, sekundární a terciární plastidy jsou vázány více než dvěma membránami a obvykle nemají jádro ve fotosyntetických eukaryotech [Barbrook a kol. 2010]. Rostliny z některých taxonů se sekundárními plastidy, jako jsou *Chlorarachnea* a *Cryptophyceae* si uchovávají reliktní eukaryotické jádro – nukleomorfní – nacházející se mezi vnitřními dvěma a vnějšími membránami chloroplastu [Barbrook a kol. 2010, Green 2011].




Obr. 1: Původ plastidů z A) primární a B) sekundární endosymbiosy [Archibald 2015].

### 3.2.2.2 Plastidové nukleoidy

Plastidová DNA je uspořádána do proteinů s vysokou molekulovou hmotností a komplexů obsahujících RNA, které jsou připojeny k vnitřním membránám plastidu. Tyto struktury jsou podobné bakteriálním nukleoidům (odtud pochází jejich název) a někdy se nazývají plastidová jádra [Jensen a Leister 2014, Krupinska a kol. 2013]. Nukleoidy jsou považovány za hlavní formu plastomu v organelách [Golczyk a kol. 2014]. Plastidové nukleoidy vyšších rostlinných plastidů jsou vysoce dynamické struktury: jejich počet, morfologie, strukturní uspořádání a obsah proteinů závisí na podmínkách prostředí a během vývoje chloroplastů se výrazně mění [Pfalz a Pfannschmidt 2015, Woodson a Chory 2012]. Plastidové nukleoidy jsou umístěny v obalové membráně nezralých proplastidů a poté se přemísťují na tylakoidy ve zralých chloroplastech, kde dochází ke snižování jejich velikosti a stávají se kompaktnějšími a hojnějšími. Zralé chloroplasty obsahují mnoho malých nukleoidů připojených k tylakoidům [Yagi a Shiina 2012, 2014]. Ukázalo se, že nukleoidy se mohou v rámci stejného chloroplastu lišit strukturálně a funkčně. Plastidové nukleoidy obsahují v průměru 10–20 kopií plastidového genomu [Melonek a kol.2010, Powikrowska a kol. 2014].

Plastidové nukleoidy mají jedinečné složení a strukturu – připomínají jak prokaryotické nukleoidy, tak eukaryotický chromatin [Jensen a Leister 2014, Krupinska

a kol. 2013]. Základní rozdíl mezi uspořádáním genomu u plastidů a bakterií spočívá v tom, že plastidy mají více nukleoidů s různým počtem kopií genomu, zatímco bakterie mají pouze jediný nukleoid obsahující proměnný počet molekul DNA, který se u různých bakteriálních druhů liší. Kompaktní struktura DNA v plastidových nukleoidech byla porovnávána s chromatinem jádra eukaryotických buněk. Centrální tělo nukleoidu s hustým obalem (jádrem) může odpovídat eukaryotickému heterochromatinu, zatímco více disperzní periferní oblasti (oblasti aktivní transkripce) připomínají eukaryotický euchromatin [Krupinska a kol. 2013, Powikrowska a kol. 2014].

#### 3.2.2.3 Struktura genomu chloroplastu

Velikost plastidového genomu (plastomu) suchozemských rostlin a fotosyntetických řas se pohybuje od 120 do 190 kb [Wicke a kol. 2011, Yu a kol. 2014] (například velikost plastomu u huseníčku rolního (*Arabidopsis thaliana*) je 154 kb [Ortelt a Link 2014]). Chloroplastový genom je složen z mnoha kruhových a pravděpodobně identických dvouřetězcových molekul DNA. Kromě kruhových molekul DNA mohou plastomy obsahovat alternativní formy DNA, jako jsou multimerické kruhy a lineární a rozvětvené molekuly DNA [Oldenburg a Bendich 2015, Ruhlman a Jansen 2014]. Každá kruhová molekula DNA vyšších rostlin obsahuje 100–150 genů (< 5 % typického cyanobakteriálního genomu) [Ortelt a Link 2014]; u suchozemských rostlin a fotosyntetizujících řas se počet genů v jedné kruhové molekule DNA pohybuje od 100 do 200 [Rochaix a Ramundo 2015]. Chloroplastový genom obsahuje geny, které se primárně účastní fotosyntézy, transkripce a translace. Obsah genů a obecná struktura chloroplastových genomů v suchozemských rostlinách jsou značně zachované [Raman a Park 2015]. Plastom tabáku virginského (*Nicotiana tabacum*) (155 939 bp) je považován za etalon starověkého uspořádání plastomu. Umístění genů v tomto plastomu je typické pro krytosemenné rostliny, jejichž plastomy během evoluce neprošly významnými přeskupeními.

Typický plastom je čtyřdílný s dlouhou jednokopiovou oblastí (LSC, ~80 kb) a krátkou jednokopiovou oblastí (SSC, ~20 kb), které jsou oddělené dvěma identickými obrácenými duplikacemi (IR, ~25 kb), jež jsou nezbytná pro stabilizaci struktury genomu plastidu. IR jsou nejvíce zachované prvky plastomu. V suchozemských rostlinách obrácená opakování obvykle obsahují základní skupinu genů pro čtyři rRNA (4,5S, 5S, 16S, 23S) a pět tRNA (trnA-UGC, trnI-GAU, trnN-GUU, trnR-ACG, trnV-GAC). Kromě tohoto jaderného shluku (klastru) rRNA/tRNA obsahují obrácená opakování mnohých

suchozemských rostlin, zejména pak cévnatých rostlin, také řadu dalších genů. U neotropické liány *Tanaecium tetragonolobum* (čeleď trubačovité (*Bignoniaceae*)) je například 10 z 86 genů kódujících proteiny lokalizováno uvnitř oblasti IR, a je tedy v plastomu plně duplikováno [Nazareno a kol. 2015]. Určité rodové linie chloroplastových genomů suchozemských rostlin také vykazují významná strukturální přeskupení, což svědčí o ztrátě oblastí IR nebo celých rodin genů.

Ačkoliv jsou plastidové genomy relativně malé, kódující sekvence obsahují pouze 50 % plastomů suchozemských rostlin. Zbytek plastomu je zastoupen introny, regulačními sekvencemi a inter-genovými mezerníky. Introny jsou stejně jako geny v genomech chloroplastu suchozemských rostlin obecně zachovány, ale ztráta intronů uvnitř genů kódujících proteiny byla popsána u několika druhů rostlin [Jansen a kol. 2007], včetně ječmene (*Hordeum vulgare*) [Saski a kol. 2007], bambusu (*Bambusa* sp.) [Wu a kol. 2009], manioku (*Manihot esculenta*) [Daniell a kol. 2008] a cizrny (*Cicer arietinum*) [Jansen a kol. 2008]. Proteiny kódované geny, u nichž je známá ztráta intronů, mají různé funkce; patří mezi ně ATP syntáza (atpF), Clp proteáza (clpP), RNA polymeráza (rpoC2) a ribosomální proteiny (rpl2, rps12 a rps16) [Jansen a kol. 2007].

Nejvíce kompaktní plastom mezi fotosyntetickými suchozemskými rostlinami (66 % kódujících sekvencí) byl nalezen u velvičie podivné (*Welwitschia mirabilis)*, čeleď welwitschiovité (*Welwitschiaceae*), což je rostlina z genetofytní rodové linie nahosemenných rostlin [McCoy a kol. 2008]. U fotosyntetických řas se obsah kódujících sekvencí v plastomu pohybuje od 50 % (zelená řasa *Chlamydomonas reinhardtii*) do 93,5 % (červená řasa *Cyanidioschyzon merolae*) [Ortelt a Link 2014]. Plastomy jsou také vysoce bohaté na AT (60 – 70 %); celkový obsah GC je obvykle 30 – 40 %, ačkoli v některých oblastech, které nekódují proteiny, obsah AT překračuje 80 %. Podíl GC, který je vyšší v sekvencích kódujících proteiny, se v různých plastomech liší. Například geny kódující fotosyntetické proteiny mají nejvyšší obsah GC, zatímco geny dehydrogenázy NAD(P)H mají obsah nejnižší [Ruhlman a Jansen 2014]. Extrémně vysoký obsah AT byl nalezen u zbytkových plastidových (apikoplastových) genomů parazitárních výtrusovců (*Apicomplexa*), jako je rod zimnička (*Plasmodium*) a kokcidie (*Toxoplasma*) [Wicke a kol. 2013].

#### 3.2.2.4 Chloroplastové geny

clDNA se u většiny druhů dědí po mateřské linii [Zhang a Sodmergen 2010]. Plastidové geny lze rozdělit na tři funkční skupiny: geny kódující komponenty

fotosyntetického aparátu, geny kódující komponenty genetického systému a geny kódující proteiny zapojené do dalších buněčných procesů (biosyntézy aminokyselin, mastných kyselin, pigmentů apod.) [Odintsova a Yurina 2003, Tiller a Bock 2014]. Typické plastidové genomy vyšších rostlin kódují všechny typy rRNA (23S, 16S, 5S, 4,5S), 27 až 31 genů tRNA a řadu proteinů (např. ~85 fotosyntetických proteinů) [Powikrowska a kol. 2014]. Asi 45 vysoce konzervovaných genů bylo nalezeno v plastomech téměř všech fotosyntetických organismů [Barbrook a kol. 2006]. Plastidy také obsahují řadu nekódujících RNA, včetně mnoha antimediátorových RNA [Borner a kol. 2015].

Jedenáct chloroplastových genů kóduje podjednotky ndh, které se účastní fotosyntézy. Proteiny ndh se sestavují do komplexu fotosystému I za účelem zprostředkování transportu cyklických elektronů v chloroplastech [Munekage a kol. 2004, Ueda a kol. 2012] a usnadnění chlororespirace [Peltier a Cournac 2002]. Některé autotrofní rostliny v chloroplastovém genomu nemají funkční geny ndh [Blazier a kol. 2011, Braukmann a kol. 2009, Chang a kol. 2006, McCoy a kol. 2008, Pan a kol. 2012, Sanderson a kol. 2015, Weng a kol. 2014, Wu a kol. 2010, Yang a kol. 2013]. Na rozdíl od dříve popsaných ztrát jednoho genu byla v těchto rostlinách deletována celá rodina genů ndh. Sedm chloroplastových genomů u orchidejí vykazuje alespoň tři nezávislé delece genů ndh [Lin a kol.2015]. Některé fragmenty DNA genů ndh u orchidejí byly identifikovány v mitochondriálním genomu, ale kompletní geny ndh potřebné k translaci domnělých funkčních proteinových komplexů u těchto rostlin chybí [Lin a kol. 2015].

#### 3.2.2.5 Využití chloroplastové DNA

Konzervovaná struktura chloroplastové DNA a absence rekombinace, nízký stupeň mutací společně s dědičností většinou po mateřské linii, činí chloroplastovou DNA častým markerem využívaným ve fylogenetických studiích rostlinných druhů [Shaw a kol. 2005]. Vysoký počet kopií chloroplastů na jednu buňku přispívá k tomu, že je clDNA pro sekvenování snazším cílem než nízkokopiové jaderné geny, zejména pak z malých nebo degradovaných vzorků [Staats a kol. 2013]. Přestože se více uplatňují variace DNA v jaderném genomu oproti plastidovým genům [Hollingsworth a kol. 2011, Lemmon a Lemmon 2013, Mandel a kol. 2014, Weitemier a kol. 2014, Zimmer a Wen 2013], mnohé oblasti výzkumu, jako je fylogenetika a fylogeografie, budou sekvence chloroplastů nadále využívat, ať už z technických nebo biologických důvodů. Obecně platí, že mitochondriální genom rostlin se vyvíjí nejpomaleji, genom chloroplastu mírně rychleji a jaderný genom

nejrychleji [Wang a kol. 2014]. Kódující oblasti se vyvíjejí pomaleji než nekódující oblasti (introny a inter-genový spacer). Z tohoto důvodu se sekvence genu clDNA (např. *rbc*L, *atp*B, *mat*K a *ndh*F) používají značně na úrovni rodiny a výše, zatímco nekódující sekvence jako introny (např. *rp*L16, *rpo*C1, *rp*S16, *trn*L, *trn*K) a inter-genový mezerník (např. *trn*T-*trn*L, *trn*L-*trn*F, *atp*B-*rbc*L, *psb*A-*trn*H) jsou používány častěji na nižších taxonomických úrovních [Bonatelli a kol. 2013, Shaw a kol. 2007]. Inter-genové mezerníky (jako je *trn*L-*trn*F) a třetí pozice kodonu genů kódujících proteiny (jako jsou *rbc*L a *mat*K) představují nejčastěji používané oblasti clDNA při konstrukci fylogeneze [Chase a kol. 2007, Hollingsworth a kol. 2009, Pleines a kol. 2009, Shaw a kol. 2005, 2007]. Jednou z nejvariabilnějších oblastí clDNA krytosemenných rostlin je inter-genový mezerník *psb*A-*trn*H, který je také využívaným nástrojem pro analýzu fylogenetických vztahů [Štorchová a Olson 2007]. Pro fylogenetickou analýzu je také možné použít více genů nebo celé sekvence clDNA, díky čemuž jsme schopni identifikovat například vztahy mezi jednotlivými taxony v příslušné fylogenetické skupině [Hou a kol. 2016].

## 3.3 Sekvenační přístupy využívané pro sestavení celogenomové sekvence chloroplastové DNA

Výzkum zaměřený na zjišťování primární struktury plastomů začal podstatně dříve než podobné studie týkající se jaderných genomů. První osekvenované chloroplastové genomy pocházely z tabáku virginského (*Nicotiana tabaccum* L.) a porostnice mnohotvárné (*Marchantia polymorpha* L.) [Ohyama a kol. 1986, Shinozaki a kol. 1986], zatímco analogická data o eukaryotických jaderných genomech a prokaryotických genomech se objevují až o deset let později. Což bylo způsobeno velkým rozdílem ve velikosti genomu a úrovní pokroku v sekvenačních technikách. Tyto dva faktory jsou hlavním důvodem současného množství znalostí o struktuře a fungování plastomů. Zpočátku mělo čistě kognitivní charakter, ale časem získalo i významný aplikační význam.

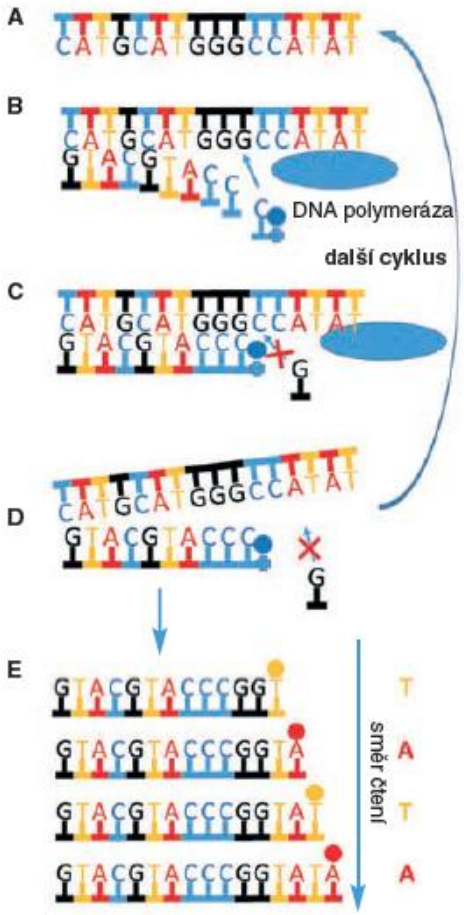
Až donedávna bylo osekvenováno přibližně 500 kompletních chloroplastových genomů a tyto informace jsou veřejně dostupné (http:// www.ncbi.nlm.nih.gov/genome). Většina kompletních chloroplastových genomů byla získána u hospodářsky významných plodin náležících do devíti čeledí: hvězdnicovité (*Asteraceae*), brukvovité (*Brassicaceae*), bobovité (*Fabaceae*), šácholanovité (*Magnoliaceae*), slézovité (*Malvaceae*), myrtovité (*Myrtaceae*), borovicovité (*Pinaceae*), lipnicovité (*Poaceae*) a čajovníkovité (*Theaceae*) [Nazareno a kol. 2015]. Plastomy ze zelených řas, krytosemenných rostlin, semen a suchozemských rostlin byly rozsáhle studovány, zatímco údaje o vlastnostech plastomů

z jiných taxonů jsou vzácnější [Xu a kol. 2015]. Zde lze nicméně nalézt zástupce i jiných taxonů, což poskytuje více příležitostí ke komparativním (srovnávacím) studiím.

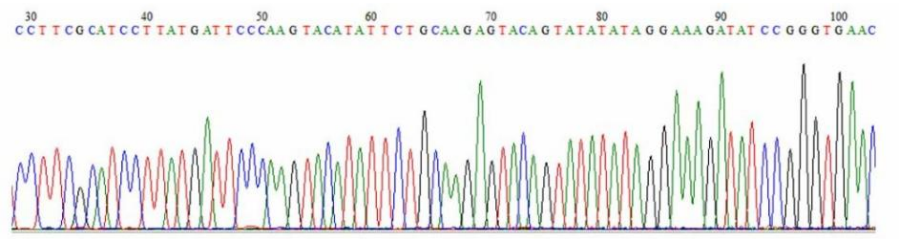
### 3.3.1 Metody sekvenování první generace

První metody využité pro sekvenování DNA byly vyvinuty v roce 1977 [Maxam a Gilbert 1977, Sanger a kol. 1977]. Označují se jako sekvenování první generace (FGS) a patří mezi elektroforetické metody [Slatko a kol. 2011]. Stejně jako Sangerova metoda, i postup dle Maxama a Gilberta vede k produkci množství různě dlouhých kopií výchozí sekvence, které jsou na jednom konci radioaktivně značeny a jsou separovány podle délky gelovou elektroforézou. Na základě vizualizace radioaktivních značek byla podle délky sekvence determinována konkrétní báze v daném místě. Zásadní rozdíl oproti Sangerově metodě však spočívá v tom, že místo syntézy pro vytvoření nového vlákna DNA o délce odpovídající rozmezí od počátku reakce k modifikované, terminační bázi, tato metoda je založena na štěpení molekul analyzovaného fragmentu DNA [Maxam a Gilbert 1977, Sanger a kol. 1977].

Sangerovo sekvenování se vyvinulo do současného automatizovaného sekvenování DNA (Obr. 2, Obr. 3). Tyto sekvenátory používají čtyři různě barevné fluorescenční značky pro odlišení dideoxy-nukleotidů a celý proces tak může probíhat v jedné reakci. Sekvenační produkty jsou následně separovány kapilární elektroforézou a barevné značení je snímáno automaticky a převedeno do podoby grafu (elektroforegramu), ve kterém jsou vyneseny intenzity jednotlivých bází podle barvy na daných pozicích [Dovichi a Zhang 2000, Karger a Guttman 2009, Prober a kol. 1987, Slatko a kol. 2011, Smith a kol. 1986, Swerdlow a Gesteland 1990, Tipu a Shabbir 2015]. Sangerovo sekvenování dominovalo po dobu asi dvou desetiletí a vedlo k mnoha úspěchům, včetně dokončení celogenomové sekvence lidského genomu [Collins a kol. 2003, International human genome sequencing consortium 2004, Lander a kol. 2001, Venter a kol. 2001].




Obr. 2: Sangerova metoda [Kolísko 2017]. Templátová DNA (A); DNA polymeráza přidává nukleotidy k rostoucímu řetězci DNA podle předlohy - templátu (B). Přidáním dideoxynukleotidu, který je označen fluorescenční značkou, dojde k zastavení syntézy nového řetězce DNA (C). Zahřátím se dva řetězce DNA oddělí (D) a proces syntézy nového řetězce polymerázou se může opakovat. Výsledné molekuly se seřadí podle velikosti a podle fluorescenčního značení se odvodí výsledná sekvence (E).



Obr. 3: Elektroforegram [http://labguide.cz]. Výsledný graf automatického vyhodnocení elektroforetického rozdělení fragmentů DNA fluorescenčně značených na 3´ konci podle přítomné báze.

### 3.3.2 Metody sekvenování druhé generace

Vývoj metod sekvenování nové generace (NGS) poskytl rychlejší a levnější metody sekvenování genomů. Moore s kolegy poprvé použili NGS k sestavení sekvencí chloroplastového genomu u rostliny nandina (*Nandina)* a platanu (*Platanus*) [Moore a kol. 2006]. Sekvenování nové generace představují metody tzv. masivně paralelního sekvenování, kdy v jediném sekvenačním cyklu lze získat obrovský objem dat ekonomickým způsobem (Tab. 1). Různé typy NGS platforem se liší v principu sekvenování jednotlivých bází, což vede k rozdílům ve výkonnosti, délce čtení, chybovosti, pokrytí genomu, nákladech a provozní době [Metzker 2009, Scholz a kol. 2012]. Dva kroky, které se týkají všech NGS platforem, zahrnují přípravu templátů a sekvenování. Existují dva základní sekvenační přístupy a to sekvenování syntézou (SBS) a sekvenování hybridizací a ligací (SBL).

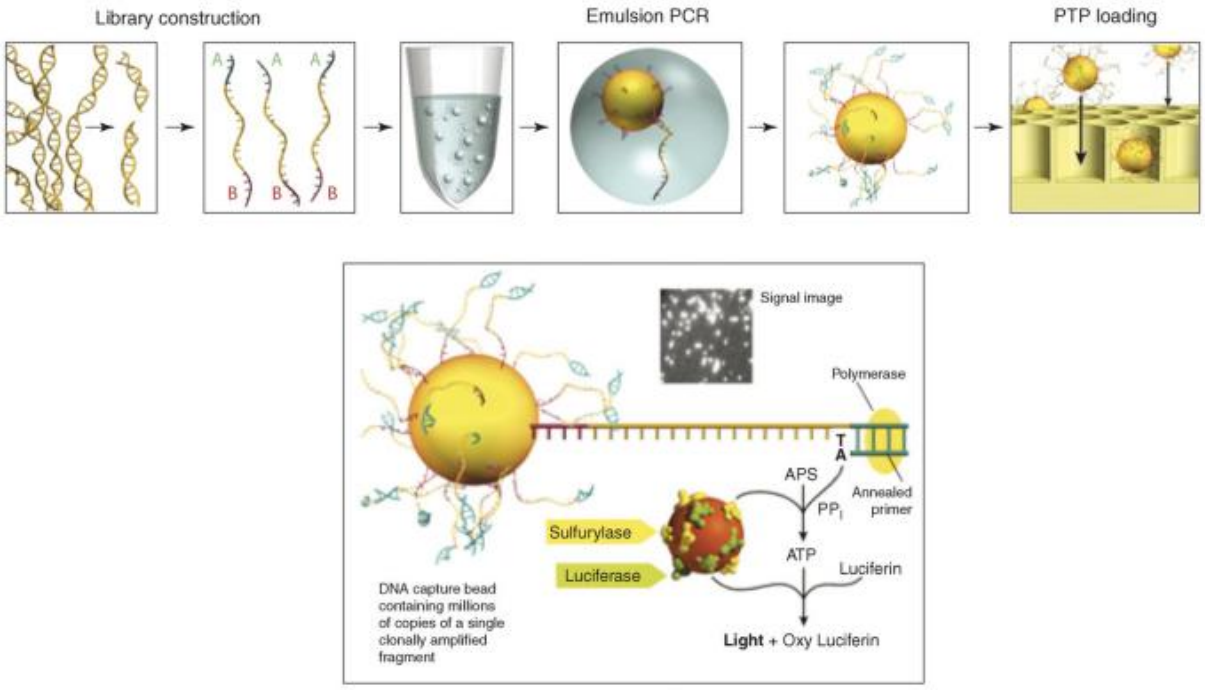
Tab. 1: Porovnání sekvenačních platforem sekvenování druhé a třetí generace [Ambardar a kol. 2016]

| Sekvenační platforma | Chybovost (%) | Délka čtení (nukleotidy) | Počet čtení na 1 běh (v milionech) | Výkonnost (Gb/běh) |
|---|---|---|---|---|
| Pyrosekvenování (Roche/454) | 1 | 500 | 1 | 0,5 |
| Sekvenování sytémem Solexa/Illumina (Illumina HiSeq 2500) | 0,26 | 2 × 100 | 8000 PE | 720-800 |
| Sekvenování sytémem Solexa/Illumina (Illumina HiSeq 2500 Rapid Run) | 0,26 | 2 × 250 | 1200 PE | 150-180 |
| Sekvenování sytémem Solexa/Illumina (Illumina NextSeq) | 0,8 | 2 × 150 | 800 PE | 100-120 |
| Sekvenování sytémem Solexa/Illumina (Illumina MiSeq) | 0,8 | 2 × 300 | 44-50 PE | 13,2-15 |
| Sekvenování sytémem Solexa/Illumina (Illumina MiniSeq) | 0,8 | 2 × 150 | 50 | 6,5-7,5 |
| Sekvenování ligací SOLiD | 0,01 | 35 | 1400 | 155 |
| Sekvenování detekcí vodíkových iontů (Ion Torrent) | 1,78 | 200 | 80 | 10 |
| Sekvenační technologie firmy Pacific Biosciences | 13 | 40 000 | 0,1 | 0,1 |
| Technologie Oxford nanopore (minION) | 38,2 | 2000 | 0,03 | 1 |

#### 3.3.2.1 Pyrosekvenování (Roche/454)

Společnost 454 Life Sciences (později koupená mezinárodním farmaceutickým gigantem Roche) přišla v roce 2005 jako první na trh se sekvenátorem druhé generace umožňujícím masivní paralelní analýzu stovek krátkých úseků DNA během jedné procedury [Margulies a kol. 2005]. Genomový sekvenátor s označením GS (a doprovázeným zkratkou jednotlivé verze přístroje) umožňoval provádět pyrosekvenování v miliónech oddělených mikroreaktorech [Ronaghi a kol. 1996]. Sekvenační reakce probíhala v pikotitračních destičkách, což je skleněná plocha s mikroskopickými jamkami, kdy do každé z jamek pikotitrační destičky je schopna zapadnout jedna sekvenační kulička s namnoženými kopiemi fragmentů DNA. Kuličky jsou výsledkem emulzní PCR [Shao a kol. 2011] a jsou imobilizovány v jamkách, ve kterých v miniaturním objemu probíhají následné cykly pyrosekvenační reakce (Obr. 4).

Pyrosekvenace opět využívá syntézu vlákna komplementárního k jednovláknovému templátu obdobně jako Sangerova metoda, ale přidání každého nukleotidu je monitorováno za pochodu skrze sledování uvolnění molekuly pyrofosfátu bez nutnosti separace vláken gelovou elektroforézou. Pyrofosfát, přirozený produkt inkorporace volného nukleotidu polymerázou na 3´ konci nově syntetizovaného řetězce, vstupuje do enzymatické kaskády, jejímž konečným výsledem je emise světelného signálu, který je detekován CCD kamerou. Kaskáda využívající enzymy přidané na počátku do reakce zahrnuje přeměnu pyrofosfátu na molekulu ATP enzymem sulfurylázou; ATP je dále použit luciferázou k produkci světelného záblesku. Pyrosekvenační reakce běží ve čtyřech opakujících se cyklech, kdy je přidán vždy pouze jeden typ dNTP, zaznamenán signál v mikroreaktorech, kde došlo v dané pozici (podle pořadí cyklu) k inkorporaci nukleotidu a následuje odmytí volných dNTP před přidáním dalšího nukleotidu v následujícím cyklu [Fakruddin a kol. 2012]. Aby v cyklu, kdy je přidáván adenosintrifosfát, nedošlo k jeho přímému využití luciferázou pro vysvícení světelného signálu, je místo běžného dATP používán modifikovaný deoxyadenosinthiotrifosfát (aATPαS), který luciferáza není schopna použít jako substrát pro reakci. Celá destička je snímána kamerou skrze její dno a pozice jamek je automaticky rozpoznána podle pozitivního signálu během prvních cyklů [Huse a kol. 2007, Marzorati a kol. 2013].



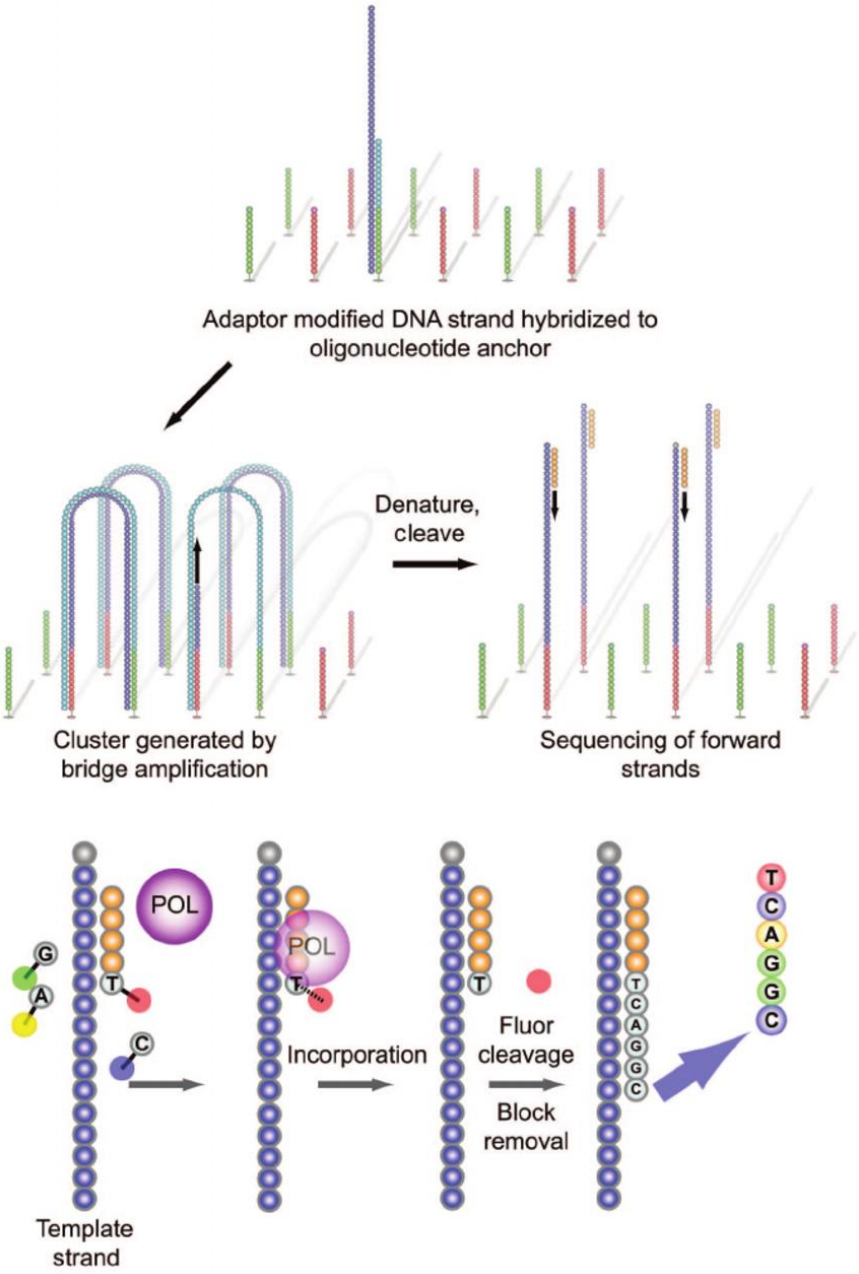

Obr. 4: Pyrosekvenování [Mardis 2008]. DNA se naváže na mikrokuličku, na níž je posléze enzymatickou reakcí namnožena. Mikrokuličky se vloží do komůrek na sekvenační destičce. Vždy je do reakční směsi přidán pouze jeden typ nukleotidu. Jestliže DNA polymeráza zařadí daný nukleotid do nového řetězce, dojde k uvolnění pyrofosfátu (PPI), který je převeden sulfurylázou na adenosintrifosfát (ATP). Luciferáza pak za použití ATP převede luciferin na oxyluciferin, přičemž dojde k vyzáření světla, které zachytíme kamerou.

### 3.3.2.2 Sekvenování sytémem Solexa/Illumina

První sekvenátor založený na principu sekvenace syntézou ve spojení s „můstkovou“ amplifikací uvedla na trh společnost Illumina v roce 2007 (Obr. 5). U této technologie se templátová DNA hybridizuje na opticky transparentní pevný povrch skleněné průtokové komůrky (glass flow cell) a využívá chemicky reverzibilně modifikovaných nukleotidů [https:/www.illumina.com/science/technology/next-generation-sequencing/sequencing-technology.html].

Příprava sekvenační knihovny zahrnuje mechanické štěpení DNA na fragmenty o velikosti 300-800 bází. Na vzniklé tupé konce DNA fragmentů se následně naliguje adeninový přesah na obou koncích fragmentů; specifické adaptory se naligují na oba konce všech fragmentů. Selektují se fragmenty požadované velikosti pomocí gelové extrakce nebo na speciálních kuličkách. Po denaturaci jsou jednotlivé fragmenty hybridizovány ke skleněné průtokové komůrce, jejíž povrch je hustě pokryt komplementárními adaptory k adaptorům

připojeným k DNA fragmentům. Každý fragment je tak svým jedním koncem imobilizován ke skleněnému povrchu průtokové komůrky. Poté je přidána směs reagencií potřebných pro PCR. Adaptory na povrchu průtokové komůrky slouží jako primery pro syntézu DNA. Dvouřetězcová DNA je následně denaturována a původní templát je odmyt. Nově nasyntetizovaný řetězec zůstává kovalentně navázáný na povrch reakční komůrky. Řetězec ssDNA se ohne a nahybridizuje na sousední primer komplementární k druhému konci molekuly. Nahybridizovaný primer je prodloužen DNA polymerázou. Vytvoří se dvouřetězcový můstek (proto je tato amplifikace nazývána „můstková“ amplifikace). Dvouřetězcový můstek je denaturován, tím vzniknou dvě kopie kovalentně navázaného jednořetězcového templátu. Celý proces se cyklicky opakuje až do vytvoření mnohočetných můstků. Dvouřetězcové můstky jsou denaturovány, reverzní řetězce odštěpeny a odmyty. Výsledkem je klastr tvořený pouze forward DNA řetězci, které budou následně sekvenovány. 3´ konce se zablokují reverzním terminátorem, aby se řetězce nemohly prodlužovat. Sekvenační primery jsou nahybridizovány na sekvence adaptérů a do skleněné průtokové komůrky s klastry je nalita směs polymerázy a čtyř rozdílně fluorescenčně značených nukleotidů s chemicky inaktivovanou 3’- OH skupinou. Je tak zaručeno, že v jednom cyklu je inkorporován pouze jeden nukleotid. Jakmile dojde k začlenění nukleotidu do řetězce DNA, pozice a typ nukleotidu jsou zaznamenány díky jeho flourescenční značce pomocí CCD kamery. Terminační skupina na 3’- konci nukleotidu i fluorescenční barvička jsou odstraněny a cyklus je opakován. Sekvence každého klastru je generována speciálním algoritmem, který jednotlivým bázím přiděluje určitou hodnotu, na jejímž základě jsou vyřazeny sekvence nízké kvality [Ambardar a kol. 2016, Bentley a kol. 2008, Goodwin a kol. 2016, Guo a kol. 2008, Heather a Chain 2016, Reuter a kol. 2015, Tipu a Shabbir 2015, Turcatti a kol. 2008, Twyford 2016, Wu a kol. 2012, 2014].

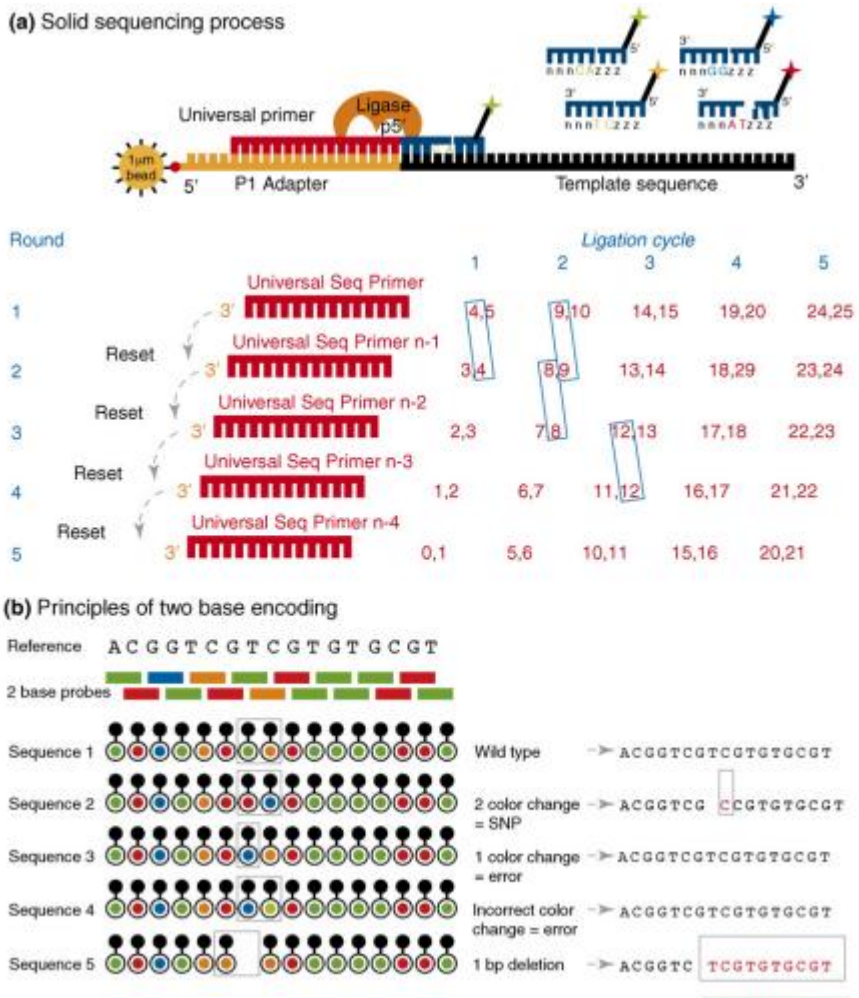



Obr. 5: Sekvenování sytémem Solexa/Illumina [Voelkerding a kol. 2009]. Jednotlivé fragmenty ssDNA jsou hybridizovány na povrch skleněné průtokové komůrky. Adaptory slouží jako primery pro syntézu DNA. Následuje „můstková“ amplifikace. Výsledkem je klastr tvořený pouze forward řetězci, které jsou následně sekvenovány. DNA polymeráza přidá do rostoucího řetězce jeden modifikovaný nukleotid značený fluorescenčním barvivem, jenž zároveň reverzibilně blokuje navázání dalšího nukleotidu. Kamera pak zachytí fluorescenční signál pro každou skupinu DNA molekul. Fluorescenční označení a blokace jsou odbourány a může dojít k připojení dalšího nukleotidu.

#### 3.3.2.3 Sekvenování ligací SOLiD

V roce 2007 představila společnost Applied Biosystems (dnes Life Technologies) platformu SOLiD, která využívá ligace pro postupné prodlužování vlákna DNA komplementárního k jednovláknovému templátu. Do reakční směsi jsou přidávány uměle syntetizované oligonukleotidy o délce osm nukleotidů (oktamery). Používá se čtyř druhů oktamerů, které jsou na 5´ konci označeny čtyřmi různými fluorescenčními značkami. Čtyři typy oktamerů však zde nepředstavují čtyři typy bází, ale ve skutečnosti každá barva odpovídá určité dvojici nukletidů na 3´ konci oligonukleotidů. Oktamery jsou totiž navrženy tak, že dva nukleotidy na 3´ konci jsou známé, následuje šest naprosto náhodných nukleotidů a na 5´ konci fluorescenční značka, která kromě signalizace prvních dvou nukleotidů zabraňuje ligaci dalšího oktameru na 5´ konec během jednoho cyklu sekvenace (Obr. 6).

Na jednovláknový templát je hybridizován sekvenační primer, ohraničující začátek reakce. Za jeho 5´ konec nasedá komplementární oktamer (podle aktuálních dvou nukleotidů komplementárních k prvním dvěma nukleotidům na 3´ konci, a nese odpovídající barevnou značku) a je ligací spojen s vláknem primeru. Poté, co je fluorescenční signál aktuálně naligovaného oktameru přečten, značka společně se třemi posledními nukleotidy oktameru je odštěpena, takže sekvenační primer zůstane prodloužen o pět nukleotidů [McKernan a kol. 2009, Roeh a kol. 2017]. Následují další cykly, které stejným způsobem prodlužují 5´ konec oktameru ligovaného v předchozím cyklu. Barevné značení kóduje dva nukleotidy, ale v každém cyklu je sekvence prodloužena o pět nukleotidů. Abychom dostali signál o dinukleotidech ve všech pozicích, je po 10 cyklech procedura přerušena, nově vzniklý ligační produkt je denaturován a omyt společně s primerem. Reakce je nastavena od začátku, ale tentokrát s novým primerem, který je o jednu bázi kratší. Tyto kroky jsou opakovány celkem s pěti primery [Valouev a kol. 2008, Yegnasubramanian 2013].

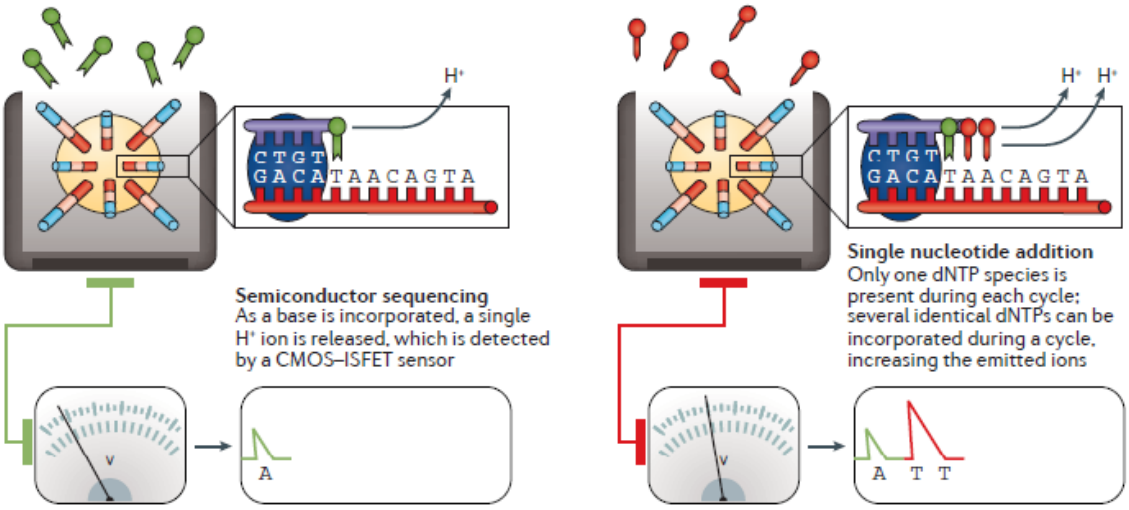



Obr. 6: Sekvenování ligací SOLiD [Mardis 2008]. Znázorněny jsou cykly sekvenačního procesu, kdy je při každém cyklu ligován jeden ze čtyř oligonukleotidů podle prvních dvou bází. Příslušné báze jsou po sekvenační proceduře dekódovány na základě barvy značky podle klíče uvedeného dole.

### 3.3.2.4 Sekvenování detekcí vodíkových iontů (Ion Torrent)

V roce 2010 představila firma Life Technologies přístroj Ion Personal Genome Machine (PGM) založený na nové technologii schopné přímo převádět chemický signál do digitální podoby [Rothberg a kol. 2011]. K přípravě knihovny se opět využívá emulzní PCR. Kuličky s templátem jsou poté deponovány na čip tak, že v každé jamce je pouze jedna molekula DNA. Proces probíhá na polovodičovém čipu hustě pokrytém mikrojamkami, pod nimiž je umístěna na ionty citlivá vrstva. Čip je postupně zaplavován jednotlivými druhy nukleotidů a dochází k syntéze DNA. Místo optického způsobu zaznamenávání jednotlivých nukleotidů se zde využívá detekce vodíkových protonů uvolněných v průběhu syntézy nově

vznikajícího řetězce katalyzovaného DNA polymerázou. Inkorporace nukleotidu způsobí uvolnění H+, čímž dojde ke změně pH, kterou zaznamenává detektor (Obr. 7). Ion Torrent měří pH v rekční směsi; podle intenzity změny pH lze poznat, kolik bází bylo přiřazeno (pH roztoku se mění s každou přidanou bází o 0,02 jednotky). Pokud nukleotid není komplementární, detektor zaznamená nulový signál. Dojde-li k začlenění dvou nukleotidů, je signál dvojnásobný [Ambardar a kol. 2016, Goodwin a kol. 2016, Heather a Chain 2016, Malapelle a kol. 2015, Mascher a kol. 2013].



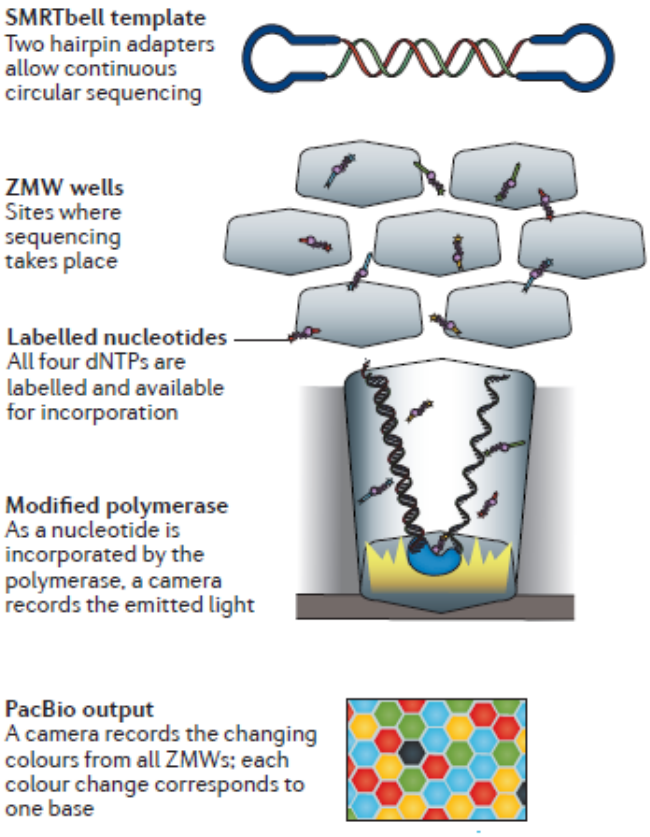

Obr. 7.: Ion Torrent [Goodwin a kol. 2016]. Během každého cyklu je polovodičový čip zaplavován jedním typem dNTP, inkorporací nukleotidu dojde k uvolnění iontu vodíku, což detekuje senzor CMOS-ISFET. Při začlenení více nukleotidů se signál znásobí. Nukleotidy, které se neinkorpovaly, jsou odmyty a v dalším cyklu je přidán jiný druh dNTP.

### 3.3.3 Sekvenování třetí generace

Technologie „single-molecule“ sekvenování, označovaná jako třetí generace sekvenování, nevyužívá amplifikaci před vlastním sekvenováním, což zkracuje dobu přípravy DNA a snižuje chybovost pramenící z amplifikace, jelikož při vytváření kopií může docházet k vnášení chyb. Třetí generace sekvenování je však ve skutečnosti více chybová, než druhá generace, jelikož signál při sekvenování jediné molekuly v reálném čase je slabý a existuje riziko, že bude odečten špatně [Heather a Chain 2016].

#### 3.3.3.1 Sekvenační technologie firmy Pacific Biosciences

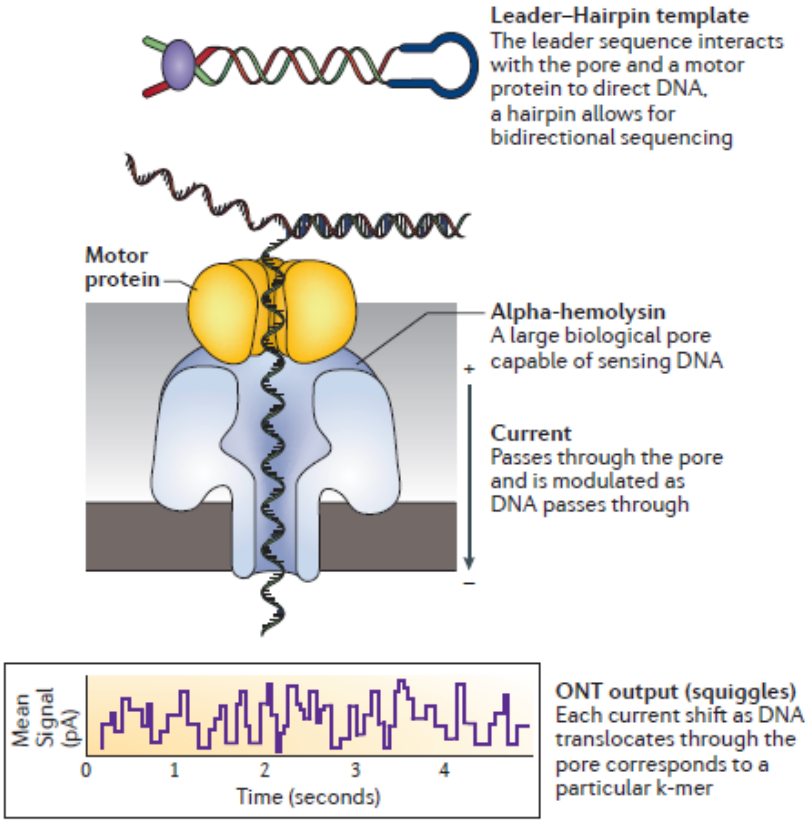
Společnost Pacific Bioscience zveřejnila v roce 2010 sekvenační technologii tzv. SMRT (single-molecule, real- time sequencing) [McCarthy 2010]. Postup zahrnuje mechanickou fragmentaci DNA a úpravu konců (zatupení, přidání dATP), následuje ligace vlásenkových adaptérů, přečištění a velikostní selekce (Obr. 8). Tento systém využívá nanostruktury zvané Zero Mode Waveguide - destičky s deseti tisíci jamkami o průměru 10 nm. Sekvenování probíhá v ZMW komůrkách v pikolitrových objemech. Sekvenuje se syntézou řetězce komplementárního ke kruhovému templátu, a to několikrát dokola pro snížení chybovosti. Během sekvenačního procesu se komplementární vlákno syntetizuje pomocí modifikované DNA polymerázy ukotvené na dně každé jamky. Fluorescence může být detekována pouze u dna komůrky. Fluorescenční značka je umístěna na fosfátové skupině nukleotidu, což má za následek uvolnění záblesku zároveň s jeho inkorporací. Proces inkorporace i uvolnění fluorescence trvá po určitou dobu, čehož se využívá pro určení identity báze. V okamžiku, kdy polymeráza přichytí inkorporovaný nukleotid, je zaznamenán puls fluorescence odpovídající danému nukleotidu [Ambardar a kol. 2016, Goodwin a kol. 2016, Heather a Chain 2016].




Obr. 8.: Sekvenační technologie firmy Pacific Biosciences [Goodwin a kol. 2016]. Obrázek znázorňuje vlásenkové adaptéry, které umožňují kontinuální syntézu řetězce komplementárního ke kruhovému templátu, dále ZMW komůrky, kde probíhá sekvenování pomocí modifikované DNA polymerázy s využitím fluorescenčně značených nukleotidů. Jakmile polymeráza přichytí inkorporovaný nukleotid, je zaznamenán puls fluorescence odpovídající danému nukleotidu.

### 3.3.3.2 Technologie Oxford Nanopore (MinION)

Technologii „nanopore sequencing“ uvedla společnost Oxford Nanopore, která roce 2014 představila „nanopore“ sekvenátor MinION. Metoda je založena na biologických vlastnostech nanopóru (Obr. 9). Nanopóry jsou součástí proteinových kanálků v membránách a dovolují výměnu iontů. Nanopórem protéká konstantní proud. Analyt, kterým je v případě využití pro sekvenování jednořetězcová molekula DNA, prochází nanopórem a dochází k detekci jednotlivých nukleotidů, přičemž pro každý typ nukleotidu je předem určena modulace proudu [Deamer a kol. 2016, Jain a kol. 2016, Lu a kol. 2016, Schneider a Dekker 2012].




Obr. 9: Technologie Oxford Nanopore [Goodwin a kol. 2016]. Jednořetězcová molekula DNA je protahována mikroskopickým pórem na syntetické membráně. Nanopórem protéká konstantní proud. Detektor zaznamenává, jaká báze v daný okamžik membránou prochází.

# 4 Data a metody

K veškerým *in silico* analýzám (kromě anotace clDNA, pro kterou bylo využito webového rozhraní) byl využit IBM server, který disponuje 40 procesory, 1.5Tb RAM paměti a 90Tb diskovým polem a pracuje v prostředí systému Linux Debian. Navíc, některé procesy byly provedeny na strojích Metacentra (http://www.metacentrum.cz/en/index.html), které poskytuje přístup pro všechny registrované členy CESNETu. Metacentrum poskytlo nejen možnost práce na vzdálených serverech, ale také přístup k četným programům využitých v předkládané diplomové práci. Schéma postupu bioinformatických analýz je uvedeno v Příloze 1.

## 4.1 Illumina sekvence analyzovaných druhů *Dactylorhiza* spp.

V předložené práci byla analyzována částečná sekvenační data 9 zástupců druhů *Dactylorhiza* spp. (Tab. 2), která byla získána pomocí párového Illumina sekvenování a to na přístroji MiSeq a nebo HiSeq. Jednotlivé taxony byly sesbírány Mgr. Veronikou Nývltovou v rámci jejího PhD. Studia.

Tab. 2: Seznam analyzovaných druhů

| Vědecký název druhu | Český název druhu |
|---|---|
| *Dactylorhiza bohemica* Businský | Prstnatec český |
| *Dactylorhiza carpatica* (Batoušek & C. A. J. Kreutz) P. Delforge | Prstnatec karpatský |
| *Dactylorhiza fuchsii* (Druce) Soó subsp. f*uchsii* | Prstnatec Fuchsův pravý |
| *Dactylorhiza fuchsii* (Druce) Soó subsp. *soóana* (Borsos) Borsos | Prstnatec Fuchsův Soóův |
| *Dactylorhiza incarnata* (L.) Soó subsp. *incarnata* | Prstnatec pleťový |
| *Dactylorhiza majalis* (Reichenb.) Hunt & Summerh. subsp. *majalis* | Prstnatec májový pravý |
| *Dactylorhiza majalis* (Reichenb.) Hunt & Summerh. subsp. *majalis* * | Prstnatec májový pravý |
| *Dactylorhiza majalis* (Reichenb.) Hunt & Summerh. subsp. t*urfosa* Procházka | Prstnatec májový rašelinný |
| *Dactylorhiza traunsteineri* (Sauter ex Reichenb.) Soó | Prstnatec Traunsteinerův |

* morfologicky odlišný taxon z lokality Kalábová

## 4.2 Analýza kvality získaných Illumina sekvencí a selekce kvalitních sekvenčních čtení (tzv. *trimování* dat)

Kvalita Illumina sekvencí byla analyzována pomocí programu FastQC verze 0.10.1 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Ze sekvenačních dat byly následně odstraněny příliš krátké sekvence, případné sekvenační adaptory a sekvence s nízkou kvalitou pomocí programu Trimmomatic V0.36 [Bolger a kol. 2014] a FASTX-Toolkit verze 0.0.14 [fastq_quality_trimmer -t 20 -l 100] (http://hannonlab.cshl.edu/fastx_toolkit /index.html). Následně jsem identifikovala párová čtení pomocí nástroje Pairfq verze 0.13.1 (https://github.com/sestaton/Pairfq/).

## 4.3 Sestavení Illumina sekvencí – tzv. *assembly* částečných Illumina sekvenačních dat

Pro sestavení dlouhých úseků DNA bylo použito dvou programů: programu Ray verze 2.3.1, využívajícím de Bruijnovi grafy [Boisvert a kol. 2010] a programu MaSuRCA (Maryland Super-Read Celera Assembler) verze 2.1.0, založeném na hybridním přístupu, který využívá výpočetní účinnost de Bruijnových grafů a flexibilitu strategie založenou na identifikaci překrývajících se homologních úseků [Zimin a kol. 2013].

*Assembly* pomocí programu Ray byla nejdříve provedena pro dva taxony (*D. carpatica* a *D. incarnata*) a to za použití různé hodnoty *k-meru* (k = 55; k = 65; k = 75). Na základě získaných výsledků bylo zjištěno, že kvalita *assembly* (hodnota N50) se zvyšovala se vzrůstající hodnotou *k-meru*. Proto pro ostatní taxony byla sestavena *assembly* za použití hodnoty k = 75 [mpiexec -n 50 Ray -k 75].

Pro *assembly* pomocí programu MaSuRCA bylo potřeba vytvořit konfigurační soubor, který musí mimo jiné obsahovat hodnotu tzv. JELLYFISH (JF_SIZE), která je specifická pro daný taxon a je vypočítána jako: velikost genomu (bp) * *coverage* / 2. Velikosti genomů a JF_SIZE jednotlivých taxonů je uvedena v Tab. 3. Příklad konfiguračního souboru je uveden v Příloze 2.

Tab. 3: Hodnoty využité pro vytvoření konfiguračního souboru programu MaSuRCA

| Název taxonu | Velikost genomu 1Cx [Gb] | JF_SIZE |
|---|---|---|
| *Dactylorhiza bohemica* | ~3,58 | 2 310 538 600 |
| *Dactylorhiza carpatica* | ~3,48 | 2 046 378 770 |
| *Dactylorhiza fuchsii* subsp. *fuchsii* | ~3,19 | 3 696 920 000 |
| *Dactylorhiza fuchsii* subsp. *soóana* | ~3,19 | 2 027 916 210 |
| *Dactylorhiza incarnata* | ~3,50 | 2 440 870 170 |
| *Dactylorhiza majalis* subsp. *majalis* | ~3,03 | 9 238 530 000 |
| *Dactylorhiza majalis subsp. majalis* * | ~3,04 | 13 293 540 000 |
| *Dactylorhiza majalis* subsp. t*urfosa* | ~3,81 | 3 653 760 000 |
| *Dactylorhiza traunsteineri* | ~3,35 | 2 557 899 920 |

***** morfologicky odlišný taxon z lokality Kalábová

### 4.4 Rekonstrukce celogenomové clDNA

Prvním krokem při sestavování celogenomové chloroplastové DNA byla identifikace *kontigů* a nekontinuálních sekvencí – tzv. *scaffoldů*, vzniklých po sestavení sekvencí (*assembly*) programy Ray a MaSuRCA. Za tímto účelem bylo využito programu BLASTN [Altschul a kol. 1990], kdy jako referenční DNA sekvence bylo využito již dříve sestavených celkových chloroplastových genomů příbuzných druhů, především druhu *Habenaria pantlingiana* (GenBank kód: KJ524104.1); *Anoectochilus roxburghii* (GenBank kód: KR779936.2); *Masdevallia coccinea* (GenBank kód: KP205432.1) a *Corallorhiza trifida* (GenBank kód: KM390019.1).

Ve druhém kroku byly *kontigy* a *scaffoldy* vykazující vysokou homologii k clDNA příbuzných druhů orchidejí namapovány na nejhomolognější referenci – clDNA druhu *Habenaria pantlingiana*, *Masdevallia coccinea* a nebo *Corallorhiza trifida* pomocí programu MAFFT verze 7.029 [--localpair --maxiterate 1000] [Katoh a kol. 2005] s cílem identifikovat

přesnou pozici a zrekonstruovat tak výsledný genom clDNA studovaných taxonů rodu *Dactylorhiza*. Mapování *scaffoldů* pomocí programu MAFFT a rekonstrukce celkové clDNA byla ověřena také programem Dotter [Sonnhammer a Durbin 1995]. Program Dotter byl využit také pro identifikaci duplikovaných oblastí nacházejících se v chloroplastovém genomu.

## 4.5 *In silico* ověření zrekonstruovaných celogenomových chloroplastových sekvencí

Nově sestavené celogenomové sekvence chloroplastové DNA vybraných druhů rodu *Dactylorhiza* byly uloženy ve formátu fasta a to ve dvojím provedení: jak celková clDNA, tak sekvence obsahující jen první duplikovaný úsek clDNA. Za účelem ověření sestavení celkové chloroplastové DNA, respektive identifikaci problematických oblastí – např. oblastí sestavených na základě velmi nízkého pokrytí genomu Illumina sekvencemi – tzv. *coverage*, byly na výsledné clDNA sekvence mapovány jednotlivé párové Illumina čtení programem BWA verze 0.7.15 [Li a Durbin 2009] [bwa mem -t 5 -M]. Pomocí programu SAMtools verze 1.4 [Li a kol. 2009] byly převedeny výsledné soubory z formátu *SAM* do formátu *BAM* [samtools view -Sb]. SAMtools byl využit také pro třídění namapovaných párových čtení na sekvenci chloroplastové DNA [samtools sort -m 30G], který byl nutný pro následnou analýzu souboru v programu BEDtools verze 2.26.0 [Quinlan a Hall 2010]. Programem BEDtools byla následně stanovena hustota čtení – tzv. *read depth* pro každý úsek dlouhý 100 bp [bedtools coverage]. Výsledný soubor byl použit pro vytvoření grafu, který ukazuje *coverage* v každém 100 bp douhém úseku sestavené celogenomové sekvence chloroplastu vybraných zástupců rodu *Dactylorhiza.* Graf byl vytvořen pomocí programu Excel verze 14.0.4760.1000.

## 4.6 Experimentální ověření sestavených celogenomových chloroplastových sekvencí

### 4.6.1 PCR amplifikace

Specifické primery byly navrženy programem Primer3 verze 0.4.0. PCR reakce byla provedena v celkovém objemu 20 µl a obsahovala 10x naředěný pufr pro Taq polymerázu, biotinem nebo digoxigeninem značený nukleotidový mix o koncentraci 200 µmol.l-1 (poměr biotin-16-dUTP, resp. digoxigenin-11-dUTP : dTTP = 1:3, koncentrace biotin-16-dUTP, resp. digoxigenin-11-dUTP = 12,5 µmol.l-1 ), specifické primery o koncentraci 1 µmol.l-1 , 0,4 U

Taq polymerázy a 30 ng templátové genomové DNA. Objem PCR reakce byl doplněn sterilní redestilovanou vodou. Sekvence primerů jsou uvedeny v Tab. 4.

PCR reakce probíhala za následujících podmínek:

| | | | |
|---|---|---|---|
| Počáteční denaturace | 94 °C | 5 min | |
| Denaturace | 94 °C | 50 s | 35 cyklů |
| Nasedání primerů | 55 °C | 50 s | |
| Extenze | 72 °C | 1 min | |
| Závěrečná extenze | 72 °C | 5 min | |

Tab. 4: Sekvence primerů

| Název primeru | Sekvence primeru (5´-3´) | Překlenovaná oblast |
|---|---|---|
| LD – levý primer | TTGCGAACCAAAAAGAATGA | velká kódující podjednotka – oblast první duplikace |
| LD – pravý primer | GTGAGACATGCGAGAAACGA | velká kódující podjednotka – oblast první duplikace |
| DS – levý primer | CCGTCGCCTATTGTCACTAA | oblast první duplikace – malá kódující podjednotka |
| DS – pravý primer | GGAATTCCTTTTAACGGAGGA | oblast první duplikace – malá kódující podjednotka |
| SD – levý primer | TTCCTCGATATGGTCCGTTT | malá kódující podjednotka – oblast druhé duplikace |
| SD – pravý primer | GGAAGAAGGGGGAGAAAGAA | malá kódující podjednotka – oblast druhé duplikace |
| DL – levý primer | GTGAGACATGCGAGAAACGA | oblast druhé duplikace – velká kódující podjednotka |
| DL – pravý primer | AATATCGTAGCCGCTCATGG | oblast druhé duplikace – velká kódující podjednotka |

Po proběhnutí PCR reakce byla provedena kontrolní elektroforéza v 1,2% agarózovém gelu, kde se jako molekulární marker použilo 50 ng λ DNA štěpené enzymem PstI. Elektroforéza probíhala při 4 V/cm po dobu 90 min, a vizualizována barvením gelu v roztoku ethidium bromidu.

### 4.6.2 Přečištění PCR produktu a sekvenování

Výsledný produkt PCR reakce byl enzymaticky přečištěn pomocí alkalické fosfatázy a exonukleázy I, tj. komerčně dodávané směsi s označením ExoSAP-IT (USB, Cleveland, OH, USA). Reakční směs připravena dle návodu výrobce, byla napipetována do destičky a vložena do termocykléru s následujícím programem: 30 minut při 37 °C, 5 minut při 96 °C. Získaný, přečištěný produkt byl sekvenován Sangerovým přístupem. Sekvenování bylo provedeno s pomocí BigDye Terminator v3.1 Cycle Sequencing kitu (Applied Biosystems, Foster City, USA). Složení reakční směsi je uvedeno v Tab. 5. Reakční směs byla napipetována do ABI destičky a vložena do termocykléru na kterém byl nastaven tento profil: 98 °C/ 5 min (jeden cyklus); 96 °C/10 s, 50 °C/ 5s, 60 °C/ 4 min (60 cyklů). Získané PCR produkty byly přečištěny pomocí paramagnetických kuliček CleanSeq (Beckman Coulter) a vyhodnoceny pomocí 3730xl DNA analyzátoru (Applied Biosystems, Foster City, USA).

Tab. 5: Složení reakční směsi pro sekvenační reakci

| **Položka** | **Množství na 1 test [µl]** |
|---|---|
| Pufr 5x | 1,5 |
| BDX64 | 0,875 |
| BigDye (ředění 64×) | 0,125 |
| Primer (10 µM) | 1 |
| Deionizovaná voda | 3,9 |
| Templátová DNA | 3,4 |
| Celkový objem reakce | 10 |

Získané nukleotidové sekvence byly analyzovány programem DNA Baser (verze 3.5.3). „Sanger“ sekvence byly namapovány na kompletní clDNA prstnatců pomocí programu Muscle a graficky znázorněny v programu Seaview (verze 4.4.2).

## 4.7 Anotace clDNA sekvence a fylogenetická analýza

Anotace kompletní složené clDNA sekvence byla provedena pomocí webového rozhraní programu DOGMA [Wyman a kol. 2004]. Anotované chloroplastové sekvence byly následně graficky znázorněny programem GenomeVX (http://wolfe.ucd.ie/GenomeVx/).

Pro fylogenetickou analýzu celkových chloroplastových sekvencí bylo využito nově sestavených clDNA druhu *Dactylorhiza* a dalších již známých sekvencí skupiny

*Orchidoideae* stažených z databáze GenBank: *Anoectochilus roxburghii* (KR779936.2), *Anoectochilus emeiensis* (NC033895.1), *Goodyera fumata* (NC026773.1), *Goodyera procera* (NC029363.1), *Goodyera velutina* (NC029365.1), *Habenaria pantlingiana* (NC026775.1), *Habenaria radiata* (NC035834.1) a *Ludisia discolor* (NC030540.1). Pro fylogenetickou analýzu bylo využito sekvencí obsahující jen první duplikovanou oblast. V prvním kroku bylo provedeno mnohočetné přiřazení (tzv. multiple alignment) programem MAFFT verze 7.029 [--localpair --maxiterate 1000] [Katoh a kol. 2005] a fylogenetický strom byl zrekonstruován metodou BioNJ [Gascuel 1997] s Bootstrap hodnotou 500. Fylogram byl graficky znázorněn programem FigTree v1.4.0 (http://tree.bio.ed.ac.uk/software/figtree/).

# 5 Použité chemikálie, roztoky a komerční kity

## 5.1 Použité chemikálie

- Agaróza (Sigma)
- Bromfenolová modř (Sigma-Aldrich, Saint Louis, Missouri, USA)
- Deionizovaná voda
- Ethanol - nedenaturovaný (Lach-ner)
- Ethidium bromid (Sigma-Aldrich, Saint Louis, Missouri, USA)
- Ethylendiaminotetraoctová kyselina (EDTA) (Fluka)
- Kyselina boritá ($H_3BO_3$) (Lach-ner)
- Tris base (Sigma)
- Xylenecyanol (Sigma-Aldrich, Saint Louis, Missouri, USA)

## 5.2 Použité roztoky

- 5x TBE pufr
  - ✓ 54 g tris base
  - ✓ 27,5 g kyseliny borité
  - ✓ 20 ml 0,5 M EDTA, pH 8
  - ✓ doplnit redestilovanou vodou do 1 l
- 6x STOP C
  - ✓ 2 ml 0,5mol.$l^{-1}$ EDTA
  - ✓ 1 ml 10% SDS
  - ✓ 4,3 ml 99,9% glycerolu
  - ✓ 5 mg bromfenolové modři
  - ✓ 5 mg xylenecyanolu
  - ✓ doplnit redestilovanou vodou do 10 ml

## 5.3 Použité komerční kity

- BigDye® Terminator v 3.1 Cycle Sequencing Kit, Applied Biosystems, USA
- ExoSAP-IT®, USB®Corporation, Cleveland, Ohio, USA
- Paramagnetické kuličky Agencourt CleanSEQ, BeckmanCoulter, Danvers, USA

# 6 Seznam laboratorních přístrojů

- Centrifuga IEC Micromax RF, ThermoScientific, Waltham, USA
- Horizontální vana a zdroj napětí pro elektroforézu MP-300 V, Major Science, Taiwan
- PCR Termocyklér PTC-200, MJ Research Inc., Massachusettes, USA
- Sekvenátor 3730 xl DNA Analyzer, Aplied Biosystems, Kalifornie, USA
- UV Transluminátor, InGenius Bio Imaging, Syngene, Cambridge, Velká Británie
- Zdroj napětí pro eletroforézu OSP-300V, OWL separation Systems, Porthsmouth, USA

# 7 Výsledky

Cílem mé diplomové práce byla rekonstrukce kompletní chloroplastové DNA u vybraných rostlinných druhů a následné provedení jejich komparativní analýzy. Celkem bylo pro testování vybráno devět druhů *Dactylorhiza* spp., u kterých byla již dříve získána částečná sekvenační data – Illumina párové čtení.

## 7.1 Analýza kvality získaných Illumina sekvencí a selekce kvalitních sekvenčních čtení (tzv. *trimování* dat)

Primární analýza získaných Illumina sekvencí byla provedena programem FastQC, díky kterému byly velmi rychle získány základní informace týkající se množství získaných Illumina sekvenačních dat, délky jednotlivých čtení, kvality čtení, respektive přítomnosti adaptérů využitých při přípravě Illumina sekvenačních knihoven. Na základě výsledků získaných programem FastQC byla sekvenační data selektována tak, aby sekvenační čtení neobsahovaly adaptery, byly dlouhé alespoň 100 bp bází při kvalitě sekvenace q = 20 alespoň pro 90% nukleotidů daného čtení. Tato hodnota kvality sekvenování udává pravděpodobnost chyby 1:100. U všech použitých Illumina sekvenačních dat bylo pomocí výše zmíněné selekce dat – tzv. t*rimování* – odstraněno 5,35 – 12,00 % málo kvalitních dat, které by v následujících krocích analýzy – *assembly* – mohly mít za následek sestavení hybridních *kontigů* nebo *scaffoldů.* Výsledná *trimovaná* data tedy představovala 1,18x – 8,74x *coverage* genomů studovaných taxonů rodu *Dactylorhiza* (Tab. 6).

Tab. 6: Základní informace o sekvenčních datech a jednotlivých taxonech, s cílem stanovení pokrytí genomu - *coverage*

| Název taxonu | Ploidie | Sekvenační platforma | Délka čtení [bp] | Počet získaných čtení [1] | Počet otrimovaných dat [1] | Velikost genomu 2C [Gb] | Velikost genomu 1Cx [Gb] | Coverage otrimovaných dat |
|---|---|---|---|---|---|---|---|---|
| *Dactylorhiza bohemica* | tetraploid | MiSeq | 2 × 300 | 22 811 412 | 20 091 640 | ~14,30 | ~3,58 | 1,29x |
| *Dactylorhiza carpatica* | tetraploid | MiSeq | 2 × 300 | 19 251 068 | 17 794 598 | ~13,92 | ~3,48 | 1,18x |
| *Dactylorhiza fuchsii* subsp. *fuchsii* | diploid | MiSeq | 2 × 300 | 36 130 990 | 32 110 340 | ~6,37 | ~3,19 | 2,32x |
| *Dactylorhiza fuchsii* subsp. *soóana* | diploid | MiSeq | 2 × 300 | 20 039 172 | 17 634 054 | ~6,37 | ~3,19 | 1,27x |
| *Dactylorhiza incarnata* | diploid | MiSeq | 2 × 300 | 23 267 736 | 21 224 958 | ~7,00 | ~3, 50 | 1,39x |
| *Dactylorhiza majalis* subsp. *majalis* | tetraploid | HiSeq | 2 × 250 | 97 751 856 | 92 525 052 | ~12,17 | ~3,04 | 6,09x |
| *Dactylorhiza majalis subsp. majalis* [2] | tetraploid | HiSeq | 2 × 250 | 143 857 468 | 132 471 244 | ~12,13 | ~3,03 | 8,74x |
| *Dactylorhiza majalis* subsp. t*urfosa* | tetraploid | MiSeq | 2 × 300 | 35 919 978 | 31 772 640 | ~15,22 | ~3,81 | 1,92x |
| *Dactylorhiza traunsteineri* | tetraploid | MiSeq | 2 × 300 | 24 859 658 | 22 242 608 | ~13,40 | ~3,35 | 1,53x |

[1] počet párových čtení

[2] morfologicky odlišný taxon z lokality Kalábová

## 7.2 Sestavení Illumina sekvencí – tzv. *assembly* částečných Illumina sekvenačních dat

Pro sestavení dlouhých úseků DNA bylo použito dvou programů: programu MaSuRCA a programu Ray. Celková délka *assembly* vytvořené programem Ray u jednotlivých analyzovaných taxonů představovala 1 642 869 bp u *Dactylorhiza bohemica*, 3 209 250 bp u *Dactylorhiza carpatica*, 11 273 468 bp u *Dactylorhiza fuchsii* subsp. *fuchsii*, 3 976 677 bp u *Dactylorhiza fuchsii* subsp. *soóana*, 4 979 623 bp u *Dactylorhiza incarnata*, 43 325 070 bp u *Dactylorhiza majalis* subsp. *majalis*, 160 417 750 bp u *Dactylorhiza majalis* subsp. *majalis* (morfologicky odlišný taxon z lokality Kalábová), 3 411 374 bp u *Dactylorhiza majalis* subsp. t*urfosa* a 3 459 444 bp u *Dactylorhiza traunsteineri*. Zatímco program MaSuRCA vyprodukoval *assembly*, která představovala podstatně větší část genomu. Avšak oproti *assembly* vytvořené programem Ray, program MaSuRCA sestavil mnohem fragmentovanější *assembly* - větší počet kratších *scaffoldů* a *kontigů* v porovnání s programem Ray. Tato skutečnost se promítla také do hodnoty N50, která udává kvalitu získané *assembly*. U programu Ray byly hodnoty N50 u většiny analyzovaných druhů vyšší. Výsledky společně se základní statistikou *assembly* pomocí programu Ray a MaSuRCA jsou uvedeny v Tab. 7 a 8.

Tab. 7: Výsledky *assembly* pro *scaffoldy* a *kontigy* při použití programu Ray

| | *Scaffoldy* | | | | *Kontigy* | | | |
|---|---|---|---|---|---|---|---|---|
| Název taxonu | Počet *scaffoldů* | Celková délka *assembly* [bp] | Maximální délka *scaffoldu* [bp] | n50 [bp] | Počet *kontigů* | Celková délka *assembly* [bp] | Maximální délka *kontigu* [bp] | n50 [bp] |
| *Dactylorhiza bohemica* | 581 | 1 642 869 | 55 941 | 8 789 | 581 | 1 642 869 | 55 941 | 8 789 |
| *Dactylorhiza carpatica* | 1 565 | 3 209 250 | 84 321 | 4 378 | 1 592 | 3 199 223 | 84 321 | 3 856 |
| *Dactylorhiza fuchsii* subsp. *fuchsii* | 12 373 | 11 273 468 | 136 352 | 780 | 12 419 | 11 255 622 | 136 352 | 779 |
| *Dactylorhiza fuchsii* subsp. *soóana* | 2 421 | 3 976 677 | 110 701 | 3 655 | 2 457 | 3 966 214 | 110 701 | 2 949 |
| *Dactylorhiza incarnata* | 3 793 | 4 979 623 | 130 496 | 1 763 | 3 816 | 4 973 487 | 130 496 | 1 732 |
| *Dactylorhiza majalis* subsp. *majalis* | 55 727 | 43 325 070 | 87 801 | 693 | 55 860 | 43 277 407 | 87 801 | 693 |
| *Dactylorhiza majalis subsp. majalis* * | 207 490 | 160 417 750 | 84 348 | 734 | 207 771 | 160 307 205 | 84 348 | 733 |
| *Dactylorhiza majalis* subsp. t*urfosa* | 3 302 | 3 411 374 | 128 911 | 997 | 3 335 | 3 400 993 | 128 911 | 990 |
| *Dactylorhiza traunsteineri* | 1 623 | 3 459 444 | 110 306 | 5 068 | 1 651 | 3 452 097 | 110 306 | 4 495 |

* morfologicky odlišný taxon z lokality Kalábová

Tab. 8: Výsledky *assembly* pro *scaffoldy* a *kontigy* při použití programu MaSuRCA

| | *Scaffoldy* | | | | *Kontigy* | | | |
|---|---|---|---|---|---|---|---|---|
| Název taxonu | Počet *scaffoldů* | Celková délka *assembly* [bp] | Maximální délka *scaffoldu* [bp] | n50 [bp] | Počet *kontigů* | Celková délka *assembly* [bp] | Maximální délka *kontigu* [bp] | n50 [bp] |
| *Dactylorhiza bohemica* | 5 571 | 4 520 938 | 57 021 | 727 | 5 571 | 4 520 938 | 57 021 | 727 |
| *Dactylorhiza carpatica* | 27 424 | 18 023 404 | 69 724 | 698 | 28 193 | 17 983 368 | 69 724 | 661 |
| *Dactylorhiza fuchsii* subsp. *fuchsii* | 114 479 | 86 483 766 | 63 761 | 795 | 118 236 | 86 285 220 | 63 761 | 770 |
| *Dactylorhiza fuchsii* subsp. *soóana* | 32 949 | 26 945 256 | 81 701 | 853 | 33 851 | 26 890 562 | 81 701 | 829 |
| *Dactylorhiza incarnata* | 1 396 | 1 700 688 | 28 075 | 2 256 | 1 406 | 1 700 434 | 28 075 | 2 228 |
| *Dactylorhiza majalis* subsp. *majalis* | 620 862 | 501 267 961 | 55 803 | 877 | 643 391 | 500 082 371 | 55 803 | 850 |
| *Dactylorhiza majalis subsp. majalis* * | 2 123 538 | 1 991 909 057 | 15 892 | 1 097 | 2 123 538 | 1 991 909 057 | 15 892 | 1 097 |
| *Dactylorhiza majalis* subsp. t*urfosa* | 45 668 | 35 951 360 | 80 044 | 812 | 46 306 | 35 927 474 | 80 044 | 802 |
| *Dactylorhiza traunsteineri* | 26 547 | 23 299 029 | 75 715 | 932 | 26 892 | 23 286 072 | 75 715 | 919 |

* morfologicky odlišný taxon z lokality Kalábová

## 7.3 Rekonstrukce celogenomové clDNA

V prvním kroku byly pomocí programu BLASTN identifikovány *kontigy* a *scaffoldy* homologní k clDNA blízce příbuzných druhů, a to *Habenaria pantlingiana, Anoectochilus roxburghii, Masdevallia coccinea* a *Corallorhiza trifida.* V *assembly* vytvořené programem MaSuRCA nebyly identifikovány dlouhé *kontigy* a *scaffoldy* představující clDNA, na rozdíl od programu Ray, kdy zpravidla ty nejdelší *scaffoldy* a *kontigy* představovaly clDNA sekvence. Proto bylo pro rekonstrukci kompletní chloroplastové DNA dále využito jen *assembly* vytvořené pomocí programu Ray.

*Kontigy* a *scaffoldy* pocházející z *de-novo assembly* vytvořené programem Ray a homologní k clDNA příbuzných druhů byly dlouhé 84000-130000 bp. U druhu *Dactylorhiza fuchsii* subsp. *soóana* se díky *de-novo assembly* programem Ray dokonce podařilo složit celkový clDNA genom v jednom dlouhém *scaffoldu*.

Ve druhém kroku rekonstrukce celogenomové clDNA bylo pracováno s druhy, u kterých byly clDNA sekvence identifikovány ve více *scaffoldech*, respektive *kontizích*. Tyto byly následně mapovány na již známý chloroplastový genom příbuzného druhu s cílem zjistit, zda se jednotlivé *kontigy* a *scaffoldy* na svých koncích překrývají a zrekonstruovat tak kompletní chloroplast studovaných taxonů. Tímto přístupen se podařilo zrekonstruovat výsledný genom clDNA u 7 dalších studovaných taxonů rodu *Dactylorhiza*. Pouze u jediného druhu, *Dactylorhiza bohemica*, se nepodařilo složit celkový genom clDNA (Tab. 9).

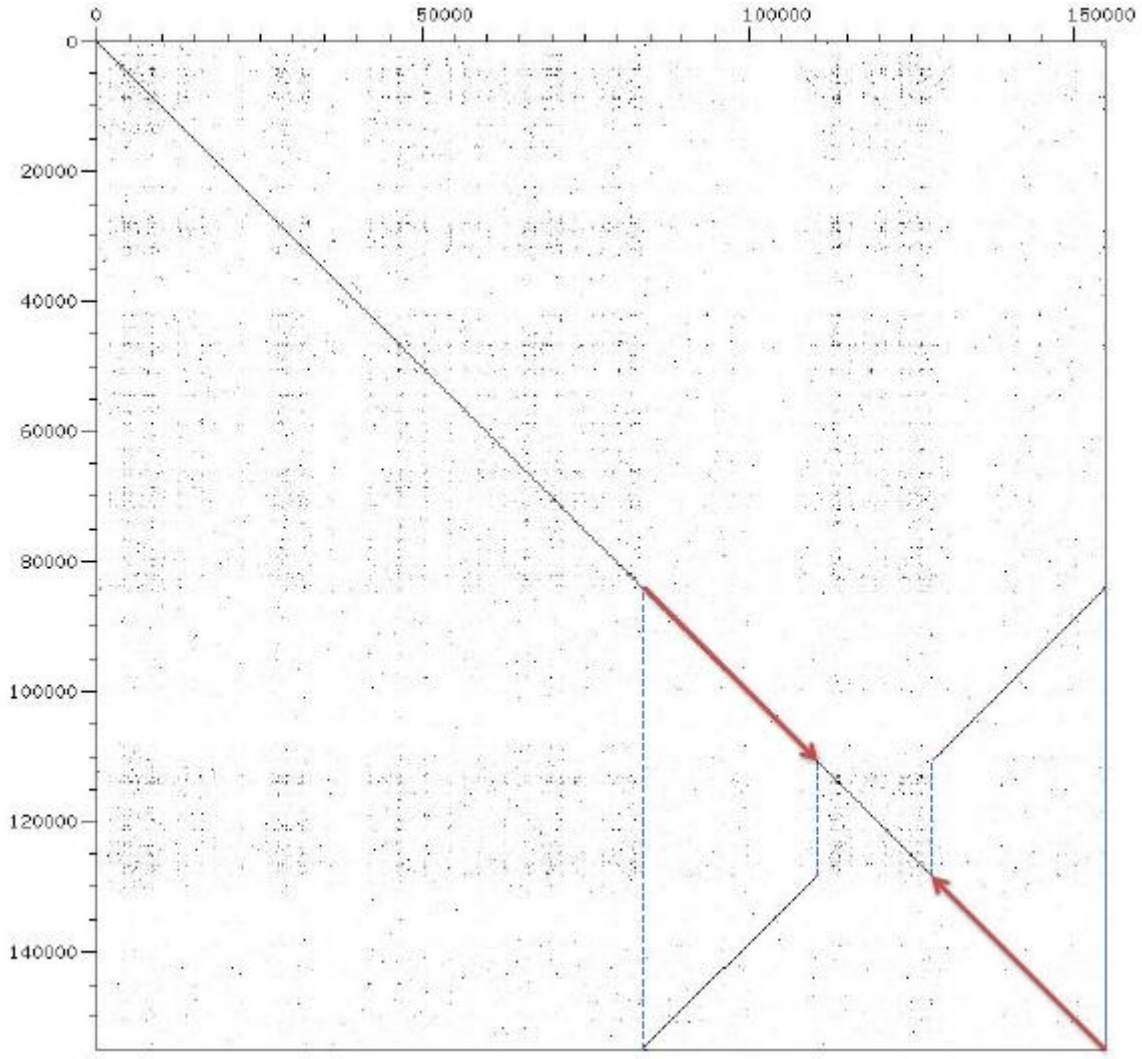
Následně byly pomocí programu Dotter u všech zrekonstruovaných clDNA identifikovány duplikované oblasti nacházející se v chloroplastovém genomu (Obr. 10).

Tab. 9: Výsledky rekonstrukce clDNA

| Název taxonu | Délka celkové clDNA [bp] |
|---|---|
| *Dactylorhiza bohemica* | nepodařilo se složit celkový genom clDNA |
| *Dactylorhiza carpatica* | 154 669 |
| *Dactylorhiza fuchsii* subsp. *fuchsii* | 154 113 |
| *Dactylorhiza fuchsii* subsp. *soóana* | 155 376 |
| *Dactylorhiza incarnata* | 154 314 |
| *Dactylorhiza majalis subsp. majalis* | 154 651 |
| *Dactylorhiza majalis subsp. majalis* * | 154 741 |
| *Dactylorhiza majalis* subsp. t*urfosa* | 154 568 |
| *Dactylorhiza traunsteineri* | 156 724 |

* morfologicky odlišný taxon z lokality Kalábová

Obr. 10: Příklad výsledné identifikace duplikovaných oblastí clDNA u *Dactylorhiza carpatica*. Na obrázku je vidět, že je clDNA kompletně složená, včetně dvou oblastí duplikace. Duplikované oblasti jsou v obrázku zvýrazněny červenými šipkami udávající také jejich orientaci (přímá obrácená duplikace).



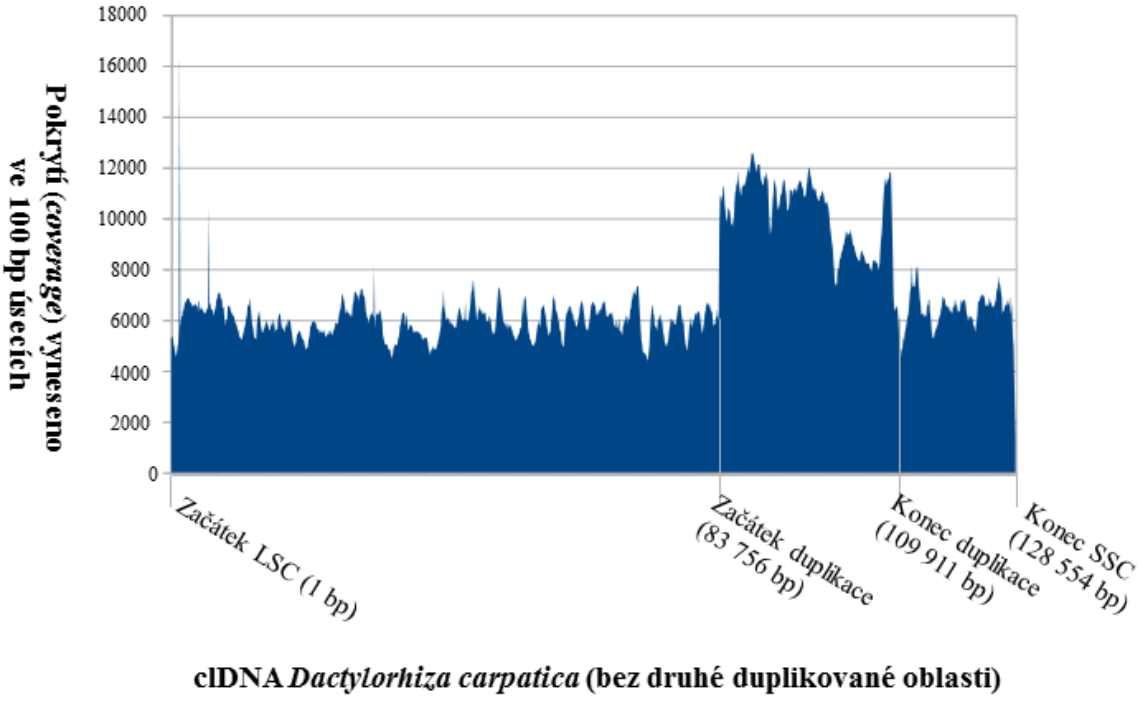

## 7.4 *In silico* ověření zrekonstruovaných celogenomových chloroplastových sekvencí

Pro ověření správnosti sestavení jednotlivých párových Illumina čtení do dlouhých úseků (*scaffoldů*) a následné rekonstrukce celkové chloroplastové DNA bylo v prvním kroku využito *in silico* analýzy. Za tímto účelem byla z celogenomové clDNA odstraněna druhá duplikovaná sekvence a párové sekvenační čtení byly mapovány na takto upravenou sekvenci. Cílem bylo zjistit, zda je zrekonstruovaný chloroplast rovnoměrně pokrytý sekvenačními čteními, nebo zda se v něm vyskytují oblasti, kde je celkové pokrytí (*coverage*) v porovnání s okolím velmi nízké. Tato skutečnost by naznačovala špatně spojené *kontigy* nebo *scaffoldy*,

respektive oblasti, které by byly v rámci *assembly* spojeny s nízkou pravděpodobností (potenciální hybridní úseky nebo úseky nutné pro další experimentální ověření).

Mapování párových sekvenačních čtení na zrekonstruované clDNA prstnatců neodhalilo ani v jednom případě problematické úseky (potenciálně špatně propojené *kontigy* nebo *scaffoldy* nebo úseky sestavené s nízkou pravděpodobností ) - úseky, které by se vyznačovaly velmi nízkým pokrytím párovými Illumina sekvencemi (Obr. 11, Příloha 3). Navíc, vzhledem k tomu, že clDNA obsahovala jen první duplikovaný úsek, který má 100% homologii s druhou duplikovanou oblastí, mapování párových Illumina sekvencí vedlo k namapování vyššího (zhruba dvojnásobného) počtu sekvencí na oblast duplikace u všech analyzovaných druhů a tak poskytlo nezávislé potvrzení přítomnosti duplikovaných oblastí v kompletní zrekonstruované clDNA studovaných druhů (Obr. 11, Příloha 3).

Obr. 11: Grafické znázornění pokrytí sestavené celogenomové sekvence clDNA u *Dactylorhiza carpatica* párovými Illumina čteními




## 7.5 Experimentální ověření sestavených celogenomových chloroplastových sekvencí

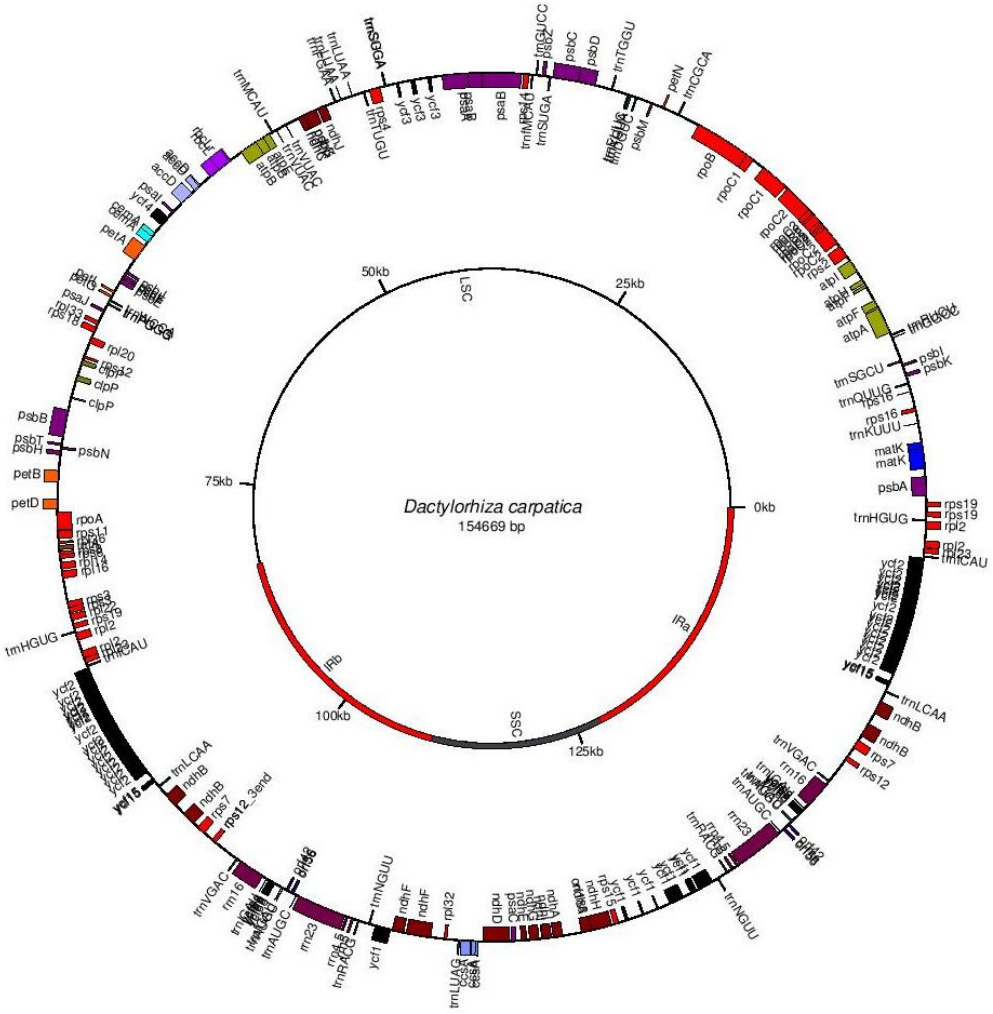
Vzhledem ke skutečnosti, že *in silico* analýza neodhalila přítomnost problematicky složených úseků v kompletně zrekonstruovaných clDNA, byly pomocí PCR ověřeny oblasti přechodu mezi dlouhým kódujícím úsekem a první duplikovanou oblastí, přechodu mezi první duplikovanou oblastí a krátkým kódujícím úsekem, přechodu mezi krátkým kódujícím úsekem a druhou duplikovanou oblastí a přechodu mezi druhou duplikovanou oblastí a dlouhým kódujícím úsekem. PCR amplifikace byla provedena na 23 vybraných zástupcích rodu *Dactylorhiza*, a u převážné většiny poskytla pozitivní PCR produkty odpovídající délce navržené *in silico*. PCR reakce byla provedena v 96 jamkové destičce a všechny DNA sekvence byly tak amplifikovány při stejné teplotě nasedání primerů (55 °C).

PCR produkty byly následně sekvenovány Sangerovým přístupem z obou stran, analyzovány programem DNA Baser a následně použity pro ověření úseků zrekonstruovaných z Illumina sekvenačních dat. Všechny získané „Sanger“ sekvence byly namapovány (přiřazeny *multiple alignment*) na místa v kompletních clDNA, ze kterých byly původně navrženy specifické primery, a zároveň bylo ověřeno, že všechny „Sanger“ DNA sekvence se shodují se sekvencí clDNA sestavenou z párových Illumina dat.

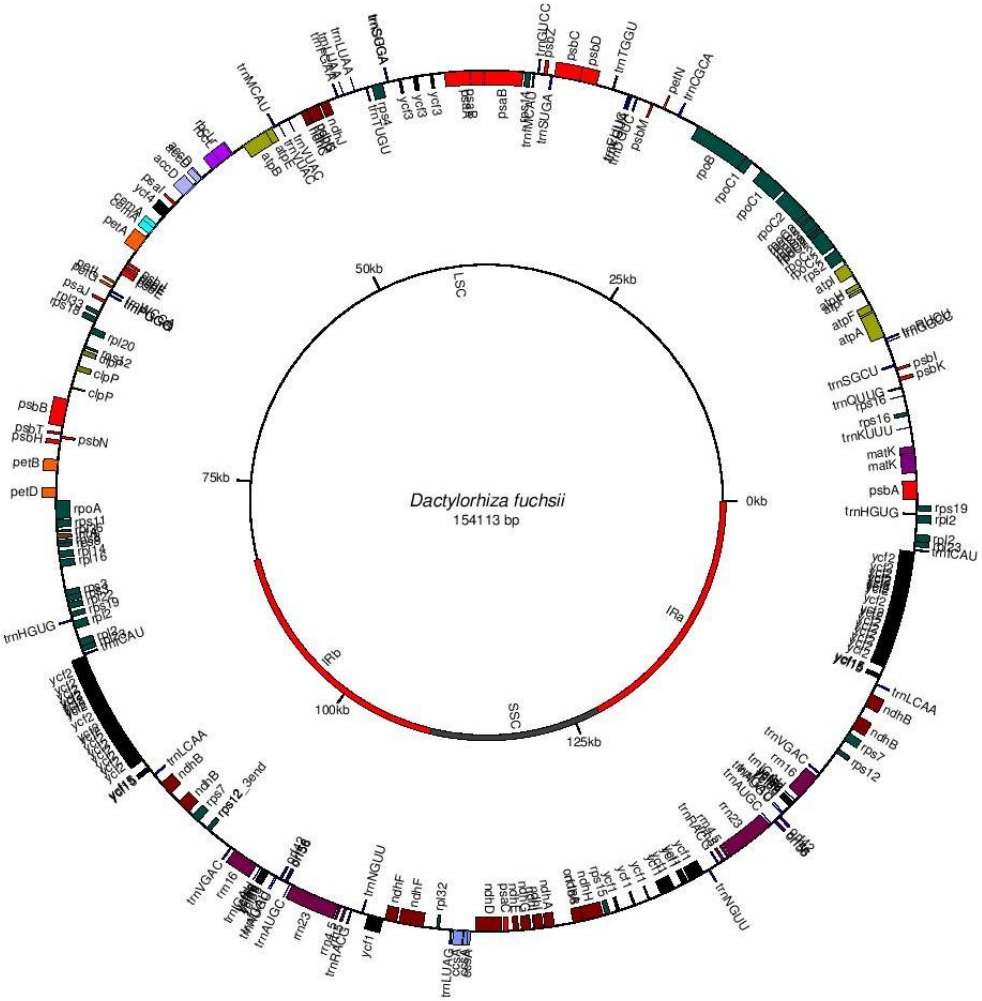
## 7.6 Anotace a komparativní analýza clDNA sekvencí

Anotace kompletní složené clDNA sekvence byla provedena pomocí webového rozhraní programu DOGMA. Počet strukturních genů nalezených u většiny studovaných druhů rodu *Dactylorhiza* byl 113, z toho 15 genů bylo duplikováno v repetitivních oblastech (*ycf*1, *orf56*, *orf42*, *ycf68*, *ycf68*, *rps12*, *rps7*, *ndhB*, *ndhB*, *ycf15*, *ycf2*, *rpl23*, *rpl2*, *rpl2*, *rps19*). Genů pro tRNA jsem detekovala 46, kdy deset z nich bylo přítomno ve dvou kopiích, protože byly duplikovány v IR oblastech (*trnH*-GUG, *trnI*-CAU, *trnL*-CAA, *trn*V-GAC, *trnI*-GAU, *trnI*-GAU, *trnA*-UGC, t*rnA*-UGC, *trnR*-ACG, *trnN*-GUU). Dále byly u studovaných druhů anotovány celkem 4 geny pro rRNA (*rrn16, rrn23, rrn4.5, rrn5*), které byly duplikovány v IR oblastech. Proteiny kódující geny, tRNA a rRNA tvořily u zkoumaných druhů rodu *Dactylorhiza* zhruba 70 % celého genomu chloroplastu. Z toho připadalo zhruba 48 % na proteiny kódující geny, 19 % na geny pro tRNA a 3 % na geny pro rRNA. Zbylých 30 % genomu tvořilo inter-genové mezerníky, introny a pseudogeny. V oblasti velké kódující podjednotky bylo přítomno přibližně 47 % genů, v oblasti malé kódující podjednotky přibližně 9 % genů a v oblastech duplikace přibližně celkem 44 % genů.

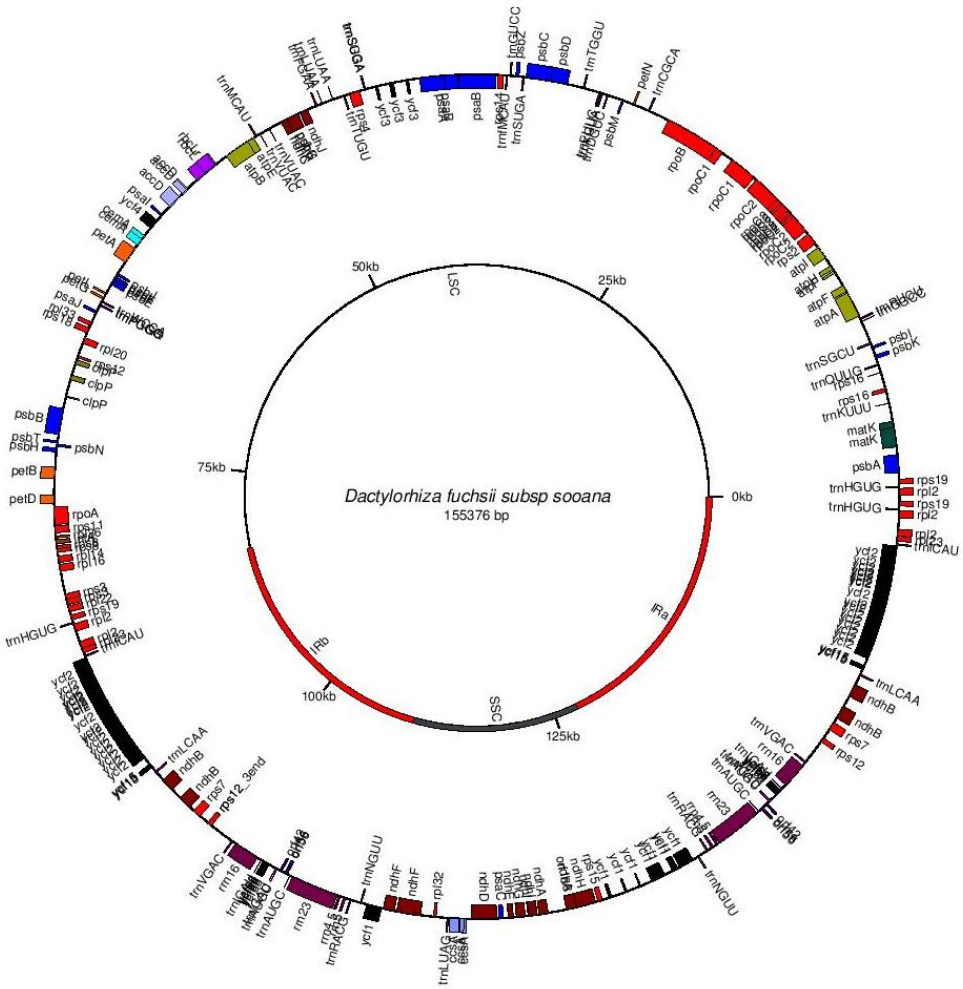
Anotované chloroplastové sekvence byly následně graficky znázorněny a jsou uvedené na Obr. 12 - 19.



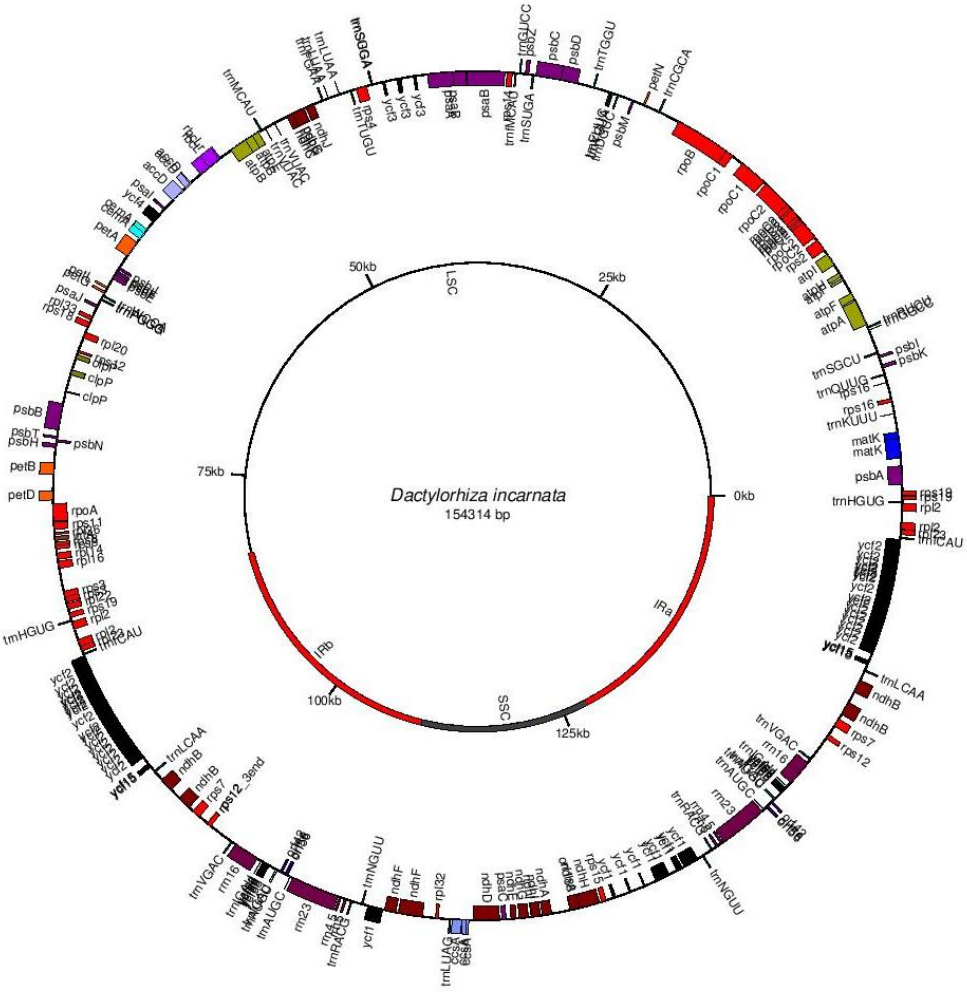

Obr. 12: Cirkulární mapa clDNA druhu *Dactylorhiza carpatica*. Červeně jsou zvýrazněny repetitivní oblasti IRb a IRa, které rozdělují genom na krátkou (SSC) a dlouhou (LSC) jednokopiovou oblast. Geny na vnější straně kruhu leží na 5' - 3' vlákně DNA, geny na vnitřní straně kruhu leží na reverzním vlákně DNA (3'- 5').



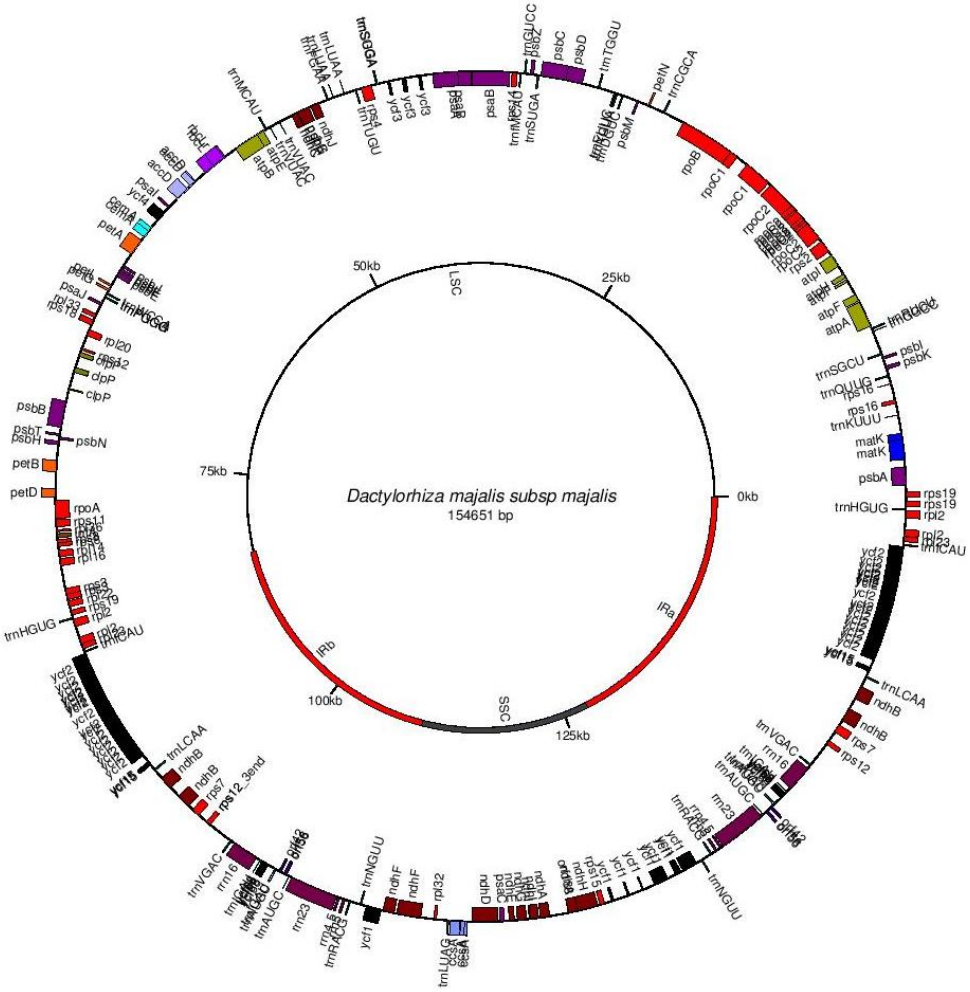

Obr. 13: Cirkulární mapa clDNA druhu *Dactylorhiza fuchsii* subsp. *fuchsii*. Červeně jsou zvýrazněny repetitivní oblasti IRb a IRa, které rozdělují genom na krátkou (SSC) a dlouhou (LSC) jednokopiovou oblast. Geny na vnější straně kruhu leží na 5’ - 3’ vlákně DNA, geny na vnitřní straně kruhu leží na reverzním vlákně DNA (3’- 5’).



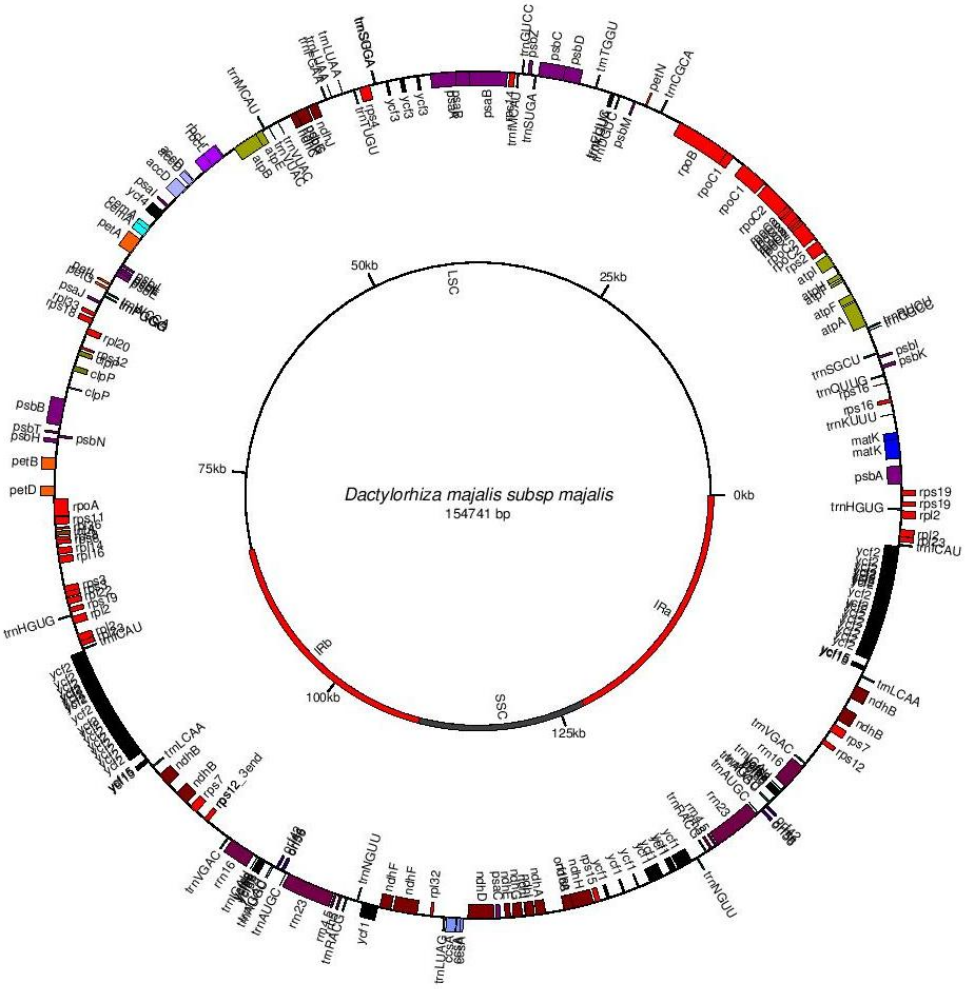

Obr. 14: Cirkulární mapa clDNA druhu *Dactylorhiza fuchsii* subsp. *soóana*. Červeně jsou zvýrazněny repetitivní oblasti IRb a IRa, které rozdělují genom na krátkou (SSC) a dlouhou (LSC) jednokopiovou oblast. Geny na vnější straně kruhu leží na 5’ - 3’ vlákně DNA, geny na vnitřní straně kruhu leží na reverzním vlákně DNA (3’- 5’).



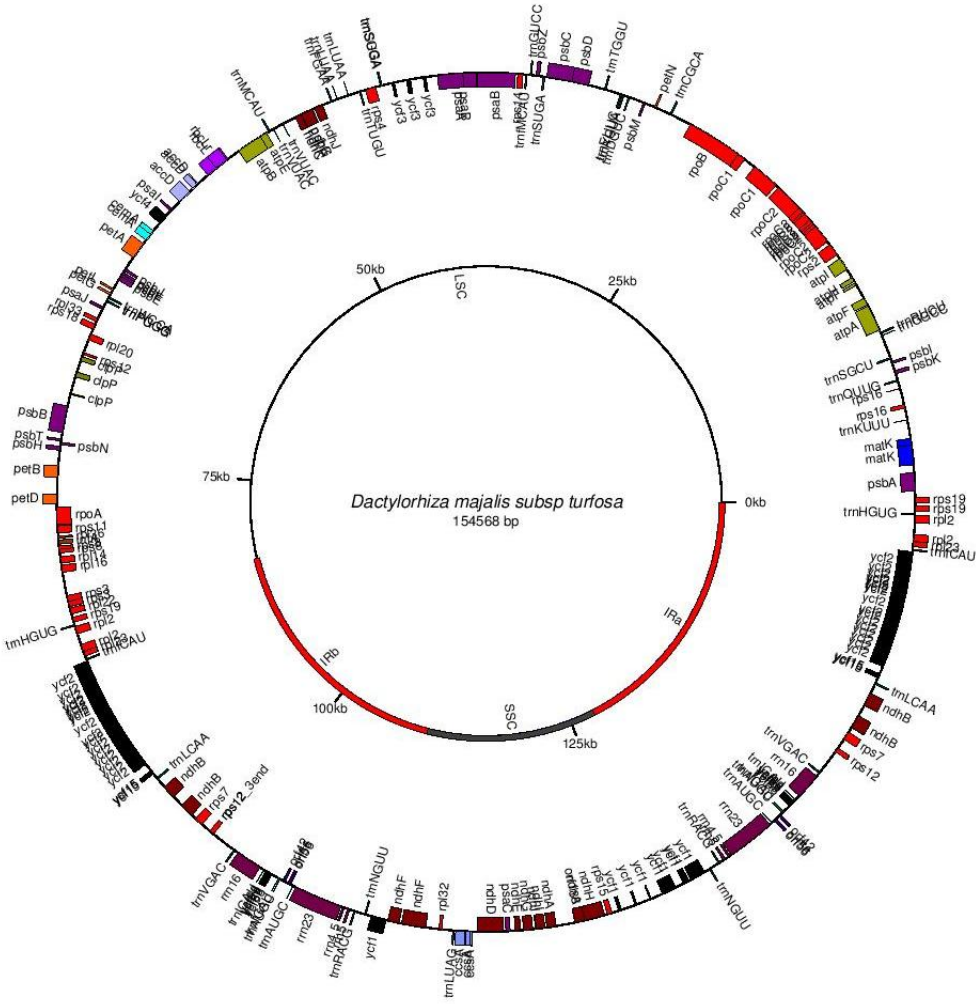

Obr. 15: Cirkulární mapa clDNA druhu *Dactylorhiza incarnata*. Červeně jsou zvýrazněny repetitivní oblasti IRb a IRa, které rozdělují genom na krátkou (SSC) a dlouhou (LSC) jednokopiovou oblast. Geny na vnější straně kruhu leží na 5' - 3' vlákně DNA, geny na vnitřní straně kruhu leží na reverzním vlákně DNA (3'- 5').



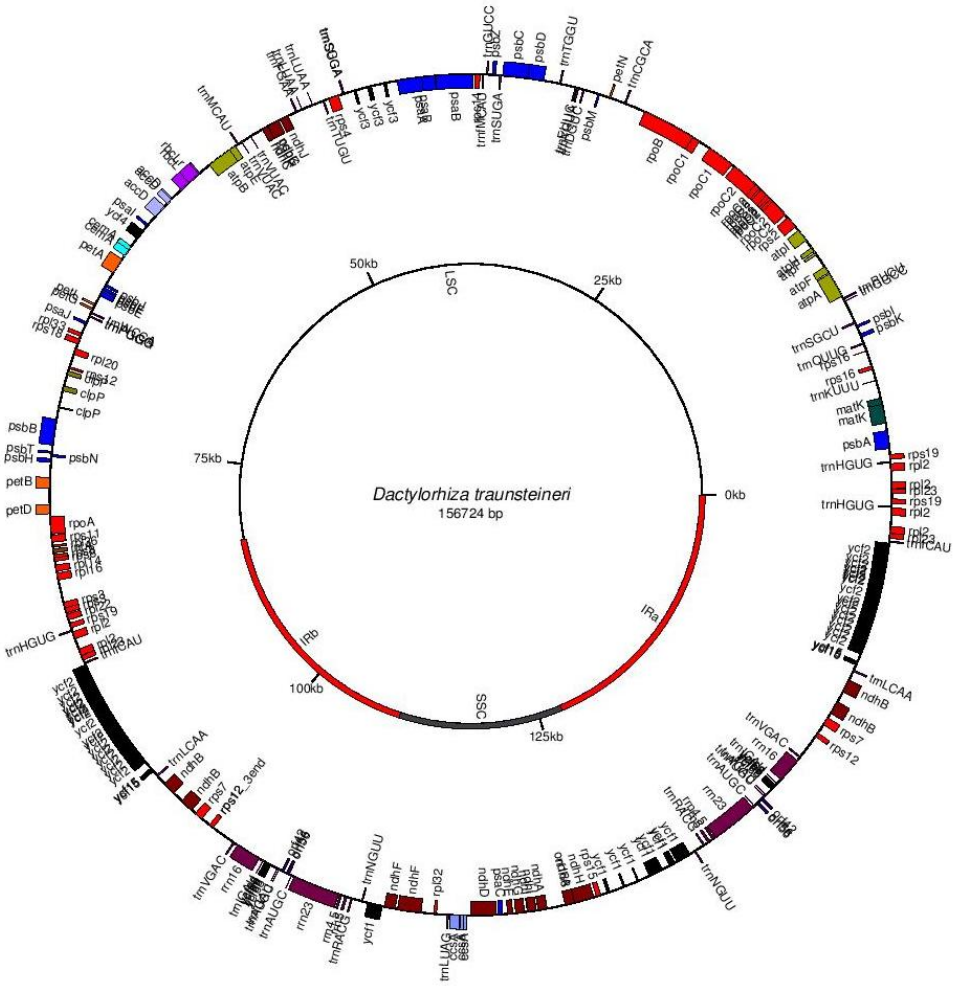

Obr. 16: Cirkulární mapa clDNA druhu *Dactylorhiza majalis* subsp. *majalis*. Červeně jsou zvýrazněny repetitivní oblasti IRb a IRa, které rozdělují genom na krátkou (SSC) a dlouhou (LSC) jednokopiovou oblast. Geny na vnější straně kruhu leží na 5’ - 3’ vlákně DNA, geny na vnitřní straně kruhu leží na reverzním vlákně DNA (3’- 5’).




Obr. 17: Cirkulární mapa clDNA druhu *Dactylorhiza majalis* subsp. *majalis* (morfologicky odlišný taxon z lokality Kalábová). Červeně jsou zvýrazněny repetitivní oblasti IRb a IRa, které rozdělují genom na krátkou (SSC) a dlouhou (LSC) jednokopiovou oblast. Geny na vnější straně kruhu leží na 5' - 3' vlákně DNA, geny na vnitřní straně kruhu leží na reverzním vlákně DNA (3'- 5').




Obr. 18: Cirkulární mapa clDNA druhu *Dactylorhiza majalis* subsp. t*urfosa*. Červeně jsou zvýrazněny repetitivní oblasti IRb a IRa, které rozdělují genom na krátkou (SSC) a dlouhou (LSC) jednokopiovou oblast. Geny na vnější straně kruhu leží na 5’ - 3’ vlákně DNA, geny na vnitřní straně kruhu leží na reverzním vlákně DNA (3’- 5’).

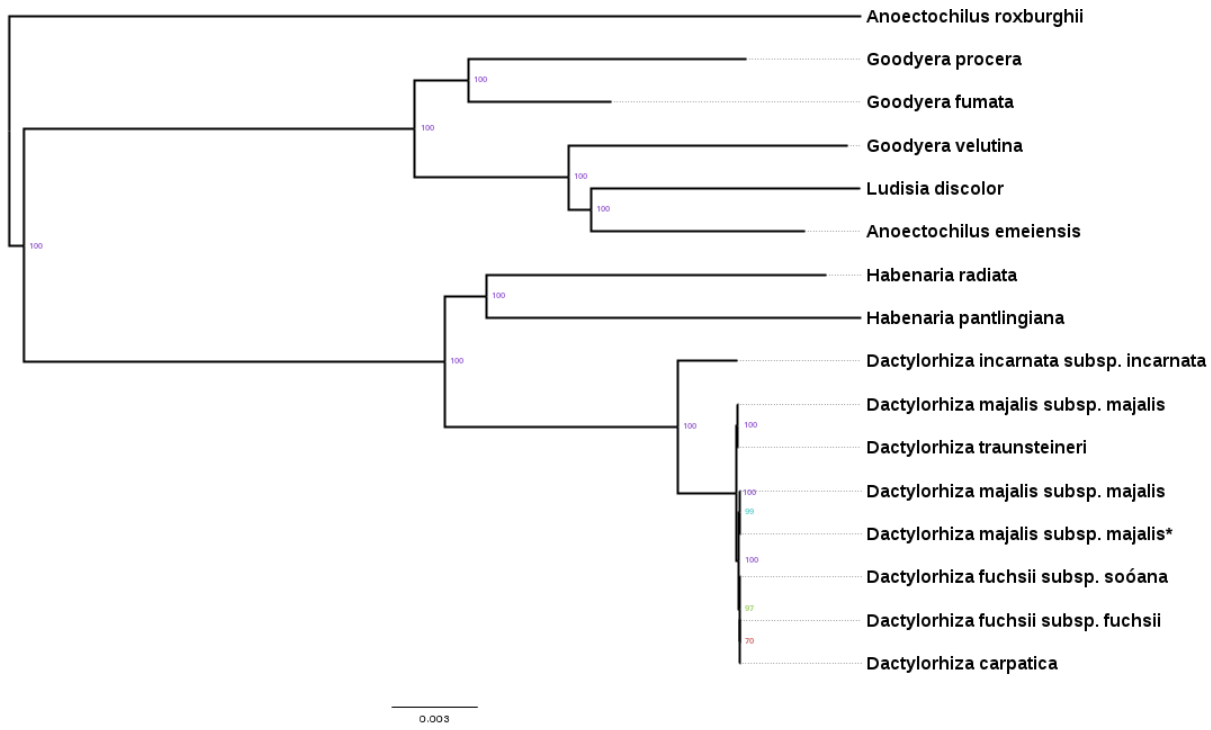



Obr. 19: Cirkulární mapa clDNA druhu *Dactylorhiza traunsteineri*. Červeně jsou zvýrazněny repetitivní oblasti IRb a IRa, které rozdělují genom na krátkou (SSC) a dlouhou (LSC) jednokopiovou oblast. Geny na vnější straně kruhu leží na 5' - 3'vlákně DNA, geny na vnitřní straně kruhu leží na reverzním vlákně DNA (3'- 5').

### 7.7. Využití celkové clDNA pro fylogenetickou analýzu

Mnohočetné přiřazení sekvencí celkové chloroplastové DNA, obsahující pouze první duplikace, bylo provedeno programem MAFFT a výsledný alignment byl graficky znázorněn v programu Seaview. Program Seaview byl využit i pro základní analýzu počtu variabilních míst. Bylo zjištěno, že všechny analyzované druhy rodu *Dactylorhiza*, kromě *D. incarnata,* se vyznačují vysokou homologií. Počet variabilních míst celkové clDNA mezi druhy *Dactylorhiza carpatica, Dactylorhiza fuchsii* subsp. *fuchsii, Dactylorhiza fuchsii* subsp. *soóana*, *Dactylorhiza majalis* subsp. *majalis*, *Dactylorhiza majalis* subsp. *majalis* *, *Dactylorhiza majalis* subsp. *turfosa* a *Dactylorhiza traunsteineri* byl pouze 41 nukleotidů. Všechny tyto změny se vyskytovaly ve formě jednonuleotidových polymorfizmů (tzv. SNP). Dva morfologicky odlišné taxony *Dactylorhiza majalis subsp. majalis,* sebrané na lokalitě Kalábová, mají naprosto totožnou chloroplastovou sekvenci. Některé tyto jednonuleotidové polymorfizmy byly specifické jen pro taxony *Dactylorhiza majalis* subsp. *turfosa* a *Dactylorhiza traunsteineri*. Podstatně variabilní byla celková clDNA druhu *D. incarnata* (662 variabilních míst vůči zbývajícím analyzovaným taxonům rodu *Dactylorhiza*). Kromě jednonukleotidových polymorfismů, se u druhu *D. incarnata* vyskytují také kratší či delší inzerce/delece, které mohou být využity v dalších analýzách pro vytvoření specifických markerů vhodných pro analýzu fylogeneze nebo genetické variability.

Multiple alignment všech analyzovaných druhů skupiny *Orchodoideae* představuje celkem 13858 variabilních míst, z toho 6628 informativních. Fylogenetická analýza skupiny *Orchidoideae* potvrdila, že nejbližším druhem, u kterého byla dosud osekvenována a sestavena celková clDNA je rod *Habenaria* (Obr. 20).




Obr. 20: Fylogenetický strom vytvořený z chloroplastových sekvencí, obsahujících pouze první duplikaci, taxonomické skupiny *Orchidoideae*

# 8 Diskuze

Chloroplastová DNA je typická pro zelené fotosyntetizující organismy a využívá se především v evolučních studiích [Nazareno a kol. 2015, Pan a kol. 2012, Raman a Park 2015, Wu a kol. 2010, Yang a kol. 2013], a to díky vysoce konzervované struktuře na úrovní nukleotidové sekvence a skutečnosti, že se u většiny organismů dědí jen po jedné rodičovské linii – především po mateřské linii [Zhang a Sodmergen 2010]. Znalost inzercí, delecí a tranzicí nebo transverzí v chloroplastovém genomu může být kromě studia evoluce studovaných druhů využita také pro identifikaci a vývoj vhodných molekulárních markerů využitelných např. pro studium genetické diverzity populací orchidejí.

V převážné většině fylogenetických studií se využívá jen krátkých genových a mimogenových úseků clDNA (např. *rbc*L, *atp*B, *mat*K, *ndh*F, *rp*L16, *rpo*C1, *rp*S16, *trn*L, *trn*K, *trn*T-*trn*L, *trn*L-*trn*F, *atp*B-*rbc*L, *psb*A-*trn*H), které jsou naamplifikovány pomocí PCR se specifickými primery a následně sekvenovány klasickým Sangerovým přístupem [Bonatelli a kol. 2013, Hollingsworth a kol. 2009, Hou a kol. 2016, Chase a kol. 2007, Pleines a kol. 2009, Shaw a kol. 2007, Štorchová a Olson 2007]. Tyto krátké oblasti clDNA jsou však tak vysoce konzervované, že je mnohdy nelze úspěšně využít pro evoluční studie na úrovni nižších taxonomických jednotek (druhů nebo poddruhů).

V dřívějších studiích byla celková chloroplastová DNA sestavena použitím klasického přístupu založeném na izolaci celogenomové nebo chloroplastové DNA, konstrukci DNA knihoven, jejich následném skríningu se sondami pro clDNA a sekvenování vybraných DNA klonů nesoucích homologní chloroplastové sekvence [Bortiri a kol. 2008, Saski a kol. 2005]. Tento přístup byl však relativně zdlouhavý a navíc byl zatížen i sekvenováním klonů, které ne vždy pocházely z chloroplastového genomu, a to díky skutečnosti, že se fragmenty clDNA vyskytují inzertované i v jaderném genomu [Chen a kol. 2015]. Tímto přístupem tak může docházet i k sestavení hybridních sekvencí.

Velký rozvoj nastal až s využitím sekvenování druhé generace a bioinformatických nástrojů vhodných pro analýzu a sestavení krátkých sekvenačních čtení (především párových sekvenačních čtení) do dlouhých genomových *kontigů* [Schatz a kol. 2010, Smith 2014]. Dosud bylo osekvenováno a sestaveno celkem 231 celkové chloroplastové DNA u orchidejí, z toho 6 zástupců patřilo do podčeledi *Apostasioideae*, 14 zástupců do podčeledi *Cypripedioideae*, 187 zástupců do podčeledi *Epidendroideae*, 19 zástupců do podčeledi *Orchidoideae* a 5 zástupců do podčeledi *Vanilloideae*.

Dosud nebyla sestavena celková chloroplastová DNA druhu *Dactylorhiza*, což je jeden

z nepočetnějších druhů evropských orchidejí, vyskytující se mimo jiné i na našem území. Proto bylo cílem předkládané diplomové práce sestavit celkovou chloroplastovou DNA devíti zástupců druhů *Dactylorhiza* spp. a zjistit variabilitu a možné využití celkové clDNA pro studium fylogeneze tohoto druhu, respektive pro identifikaci vhodných molekulárních markerů využitelných pro analýzu variability.

Pro sestavení celkové chloroplastové DNA byla využita částečná sekvenační data 9 zástupců druhů *Dactylorhiza* spp., která byla získána pomocí párového Illumina sekvenování a to na přístroji MiSeq a nebo HiSeq. Sekvenační data byla selektována tak, aby sekvenační čtení neobsahovaly adaptery, byly dlouhé alespoň 100 bp bází při kvalitě sekvenace q = 20 alespoň pro 90% nukleotidů daného čtení. Tato hodnota kvality sekvenování udává pravděpodobnost chyby 1:100. U všech použitých Illumina sekvenačních dat bylo pomocí výše zmíněné selekce dat – tzv. t*rimování* – odstraněno 5,35 – 12,00 % málo kvalitních dat, které by v následujících krocích analýzy – *assembly* – mohly mít za následek sestavení hybridních *kontigů* nebo s*caffoldů.* Výsledná *trimovaná* data tedy představovala 1,18x – 8,74x *coverage* genomů studovaných taxonů rodu *Dactylorhiza*. Přítomnost sekvenačních adaptorů v neotrimovaných datech by mohla vést k vytvoření hybridní *assembly*, kdy by se sekvenační čtení mohla spojit do *kontigů* skrze adaptory.

Pro sestavení chloroplastové DNA bylo využito *de-novo assembly*, která oproti tzv. „reference guided assembly“ poskytuje mnohem přesnější informaci o struktuře a organizaci sestavené DNA sekvence. Sestavení celogenomové sekvence přístupem známým jako „reference guided assembly“ využívá již známé celogenomové sekvence (tzv. reference) blízce příbuzného druhu, na kterou jsou namapovány jednotlivé sekvenační čtení a ty jsou poté využity pro vytvoření konsenzuální sekvence studovaného druhu. Tento přístup však může vést k vytvoření hybridních úseků nebo špatně sestavených úseků (tzv. *missasemblies*), a k nemožnosti identifikace potenciálních inzercí nebo delecí [Phillippy a kol. 2008, Schneeberger a kol. 2011].

Díky rozvoji bioinformatických nástrojů využívaných pro sestavení (*assembly*) dlouhých DNA sekvencí nebo dokonce celkových genomů (v mém případě chloroplastového genomu) z krátkých párových sekvenačních čtení se proto v dnešní době upouští od přístupu „reference guided assembly“ a stále častěji se využívá *de-novo assembly*.

Jedním z cílů předkládané diplomové práce bylo zjistit, zda jsou i částečná sekvenační data vhodná pro sestavení celogenomové clDNA přístupem *de-novo assembly*. Za tímto účelem bylo využito dvou programů využívaných pro sestavení genomových sekvencí ze sekvenačních čtení druhé generace – programu Ray a programu MaSuRCA. Program

MaSuRCA byl vyvinut v roce 2013 a je založený na hybridním přístupu, využívá výpočetní účinnost de Bruijnových grafů a flexibilitu strategie založenou na identifikaci překrývajících se homologních úseků. Tento program byl využit v několika studiích, například u studie *Bubassealus bubalis* [Williams a kol. 2017], *Aegilops tauschii* [Zimin a kol. 2017] a *Pinus taeda*, který má jeden z největších dosud sekvenovaných genomů [Zimin a kol. 2017]. Martínez-García a kol. (2016) pro *assembly* genomu *Juglans regia* rovněž využili dvou různých programů, a to SOAPdenovo2 a MaSuRCA, kdy program MaSuRCA poskytl lepší *assembly*, proto byl použit jako základní nástroj pro následné analýzy a anotaci genomu *Juglans regia*.

Oproti tomu, program Ray je založený na de Bruijnově grafu a využívá se především pro analýzu metagenomů. De Bruijnův graf je vygenerován s použitím čtení a jejich překryvů přesně naopak než u metody Overlap/ Layout/Consensus (OLC) – vrcholy (uzly) jsou překryvy a hrany mezi vrcholy reprezentují unikátní sekvenci každého čtení. Hledá se Eulerovské cesta, kdy je každá hrana navštívena jednou. Namísto alignmentu celých sekvencí jsou původní čtení rozdělena do krátkých úseků (*k-mer*) předem definované délky, ty jsou podle shody prefixů (*k-mer* mínus poslední báze) a sufixů (*k-mer* bez první báze) pospojovány do grafu, po kterém se pak algoritmus pohybuje posouváním slov o jednu pozici s navštívením každého spoje jedenkrát [Boisvert a kol. 2010]. Tento program byl úspěšně využit například ve studii clDNA u *Eleusine indica* [Zhang a kol. 2017].

Tyto dva programy byly vybrány jednak s ohledem na jejich rozdíl ve výpočetním algoritmu a také s ohledem na jejich úspěšnost pro sestavení tzv. primárních *assemblies* v několika probíhajících projektech v Centru strukturní a funkční genomiky rostlin, ÚEB AV ČR a na spolupracujících pracovištích. V diplomové práci byla tak ověřena úspěšnost těchto dvou programů pro analýzu částečných sekvenačních dat. Bylo zjištěno, že programem MaSuRCA byla sestavena mnohem fragmentovanější *assembly* - větší počet kratších *scaffoldů* a *kontigů* v porovnání s programem Ray. Tato skutečnost se promítla také do hodnoty N50, která udává kvalitu získané *assembly*. U programu Ray byly hodnoty N50 u většiny analyzovaných druhů vyšší, kdy navíc mezi nejdelší *kontigy* a *scaffoldy* patřily právě chloroplastové DNA sekvence. U druhu *Dactylorhiza fuchsii* subsp. *fuchsii* se dokonce pomocí primární *assembly* programem Ray podařilo zrekonstruovat celkovou clDNA, včetně obou duplikovaných oblastí. Proto bylo pro rekonstrukci kompletní chloroplastové DNA dále využito *assembly* vytvořené pomocí programu Ray. Naopak, více fragmentovaná *assembly* vytvořená programem MaSuRCA může sloužit v budoucnu jako základní *assembly* pro identifikaci druhově specifických molekulárních markerů, např. SSR markerů vhodných

pro studium genetické diverzity [Nývltová a kol., nepublikováno]. Rozdíly ve výsledné *assembly* získané dvěma použitými programy odrážejí právě rozdílný matematický algoritmus, který je využit pro analýzu sekvenačních čtení a jejich sestavení do delších úseků – *kontigů* a *scaffoldů*.

Celogenomovou clDNA se podařilo zrekonstruovat u osmi z devíti analyzovaných zástupců. Ani jedna *assembly* nevedla u druhu *Dactylorhiza bohemica* k sestavení celkové chloroplastové DNA, respektive neumožnila její rekonstrukci. Důvodem mohla být přítomnost menšího počtu sekvenačních čtení specifických pro clDNA v celkových sekvenačních datech tohoto druhu. Za účelem sestavení celkové clDNA *D. bohemica* tak bude nutné tento druh dosekvenovat na vyšší *coverage*.

*In silico* analýza, stejně jako Sangerovo sekvenování specifických úseků clDNA sestavených druhů, navíc potvrdila, že *de-novo assembly* částečných sekvenačních dat umožnila sestavit vysoce kvalitní celogenomovou sekvenci clDNA, ve které nebyly zjištěny žádné hybridní úseky, respektive úseky sestavené s nízkou podporou. Bylo tak možné použít nově sestavené clDNA osmi zástupců pro následné analýzy – především analýzu diverzity clDNA sekvence rodu *Dactylorhiza*, respektive fylogenetickou analýzu.

Vzhledem k malé velikosti genomu clDNA, a četným studiím, které se věnují právě sestavení celkové clDNA a analýzou variability chloroplastové DNA a jejím využitím v evolučních studiích vyšších rostlin, byly v nedávné době vytvořeny také komplexní webové nástroje, které usnadňují analýzu, respektive anotaci clDNA – např. DOGMA [Wyman a kol. 2004] a GenomeVX (http://wolfe.ucd.ie/GenomeVx/). Použití těchto webových nástrojů, specifických právě pro anotaci a analýzu clDNA nebo mtDNA, je výhodné také z hlediska toho, že stejné typy sekvencí jsou u různých druhů analyzované stejnými nástroji, což je výhodné právě pro komparativní analýzy nebo evoluční studie [Chang a kol. 2005].

Obsah genů a obecná struktura chloroplastových genomů u suchozemských rostlin jsou značně zachované, jak dokládají studie provedené na orchidejích, například u *Liparis loeselii* [Krawczyk a kol. 2017], *Cattleya crispata* [da Rocha Perini a kol. 2016], *Dendrobium nobile* [Konhar a kol. 2016], *Oncidium ´Gower Ramsey´* [Wu a kol. 2010], *Anoectochilus roxburghii* [Yu a kol. 2016] a *Cymbidium* spp. [Yang a kol. 2013]. Plastomy čtyř nově osekvenovaných druhů orchidejí (*Dendrobium moniliforme, Goodyera schlechtendaliana, Paphiopedilum armeniaclum* a *Veronica aphylla*) se liší v posunu hranice oblasti duplikace a variabilní ztrátou/retencí ndh genů. Mann-Whitney test naznačil, že ndh geny hrají důležitou roli ve stabilitě spojení oblasti duplikace a malé kódující podjednotky [Niu a kol. 2017].

Na obrázcích genomů všech anotovaných druhů (Obr. 12 - 19) je vidět, že clDNA

u všech analyzovaných zástupců rodu *Dactylorhiza* má strukturu typickou pro krytosemenné rostliny, která obsahuje dvě IR oblasti (oblasti duplikace), IRa (oblast první duplikace) a IRb (oblast druhé duplikace), které jsou od sebe odděleny LSC (velká kódující podjednotka) a SSC (malá kódující podjednotka) oblastmi. clDNA sekvence studovaných druhů *Dactylorhiz* jsou si až na druh *Dactylorhiza incarnata* extrémně podobné, což naznačuje jednak velmi blízkou evoluční příbuznost všech studovaných českých zástupců, a také to, že všechny analyzované druhy, u kterých se podařilo sestavit celkovou chloroplastovou DNA, kromě *D. Incarnata*, vznikly ze stejného, minimálně jednoho mateřského rodiče. Další analýza na celogenomové (jaderné) úrovni nebo použití vhodných molekulárních markerů by mohlo objasnit bližší evoluční vztahy a vznik zástupců rodu *Dactylorhiza* rostoucích na území ČR. Sestavení celogenomové clDNA u dvou morfologicky odlišných zástupců – *Dactylorhiza majalis* subsp. *majalis* sebraných na lokalitě Kalábová, kdy byla získána naprosto totožná clDNA, indikuje jejich stejný genetický základ a zároveň ukazuje na skutečnost, že na vysoké morfologické variabilitě prstnatců se podílí epigenetické změny, což bylo naznačeno již ve studii Balao a kol. 2016.

Závěrem, rekonstrukce celkové chloroplastové DNA osmi zástupců *Dactylorhiza* spp. rostoucích na území ČR poskytla první informaci o struktuře a organizaci clDNA genomu. Byla potvrzena relativně nízká variabilita chloroplastové DNA na úrovni druhů ale zároveň byly identifikovány variabilní oblasti clDNA mezi jednotlivými druhy orchidejí, které mohou být využity v budoucnu - např. pro rekonstrukci evoluce této velmi druhově početné skupiny orchidejí, respektive vývoj specifických markerů vhodných pro charakterizaci jednotlivých druhů nebo analýzu variability populací.

# 9 Závěr

Cílem předkládané práce byla rekonstrukce kompletní chloroplastové DNA u vybraných rostlinných druhů a následné provedení jejich komparativní analýzy. Bylo testováno devět druhů *Dactylorhiza* spp., u kterých byla již dříve získána částečná sekvenační data – Illumina párové čtení. Pro sestavení dlouhých úseků DNA bylo použito dvou programů: programu MaSuRCA a programu Ray. V *assembly* vytvořené programem MaSuRCA nebyly identifikovány dlouhé *kontigy* a *scaffoldy* představující clDNA, na rozdíl od programu Ray, kdy zpravidla ty nejdelší *scaffoldy* a *kontigy* představovaly clDNA sekvence. Proto bylo pro rekonstrukci kompletní chloroplastové DNA dále využito jen *assembly* vytvořené pomocí programu Ray.

Podařilo se zrekonstruovat celkový genom chloroplastové DNA u 8 studovaných taxonů rodu *Dactylorhiza*. U druhu *Dactylorhiza fuchsii* subsp. *soóana* se dokonce podařilo složit celkový clDNA genom v jednom dlouhém *scaffoldu*. Pouze u jediného druhu, *Dactylorhiza bohemica*, se nepodařilo složit celkový genom clDNA. U všech zrekonstruovaných clDNA byly identifikovány duplikované oblasti. *In silico* analýza neodhalila přítomnost problematicky složených úseků v kompletně zrekonstruovaných clDNA. Experimentálně jsem ověřila správnost sestavení jednotlivých párových Illumina čtení do dlouhých úseků (*scaffoldů*) a následnou rekonstrukci celkové chloroplastové DNA.

Velikost chloroplastového genomu studovaných taxonů rodu *Dactylorhiza* se pohybovala od 154 113 kb u *Dactylorhiza fuchsii* subsp. *fuchsii* do 156 724 kb *Dactylorhiza traunsteineri*. Počet strukturních genů nalezených u většiny studovaných druhů byl 113, z toho 15 genů bylo duplikováno v repetitivních oblastech; genů pro tRNA bylo detekováno 46, kdy deset z nich bylo přítomno ve dvou kopiích, protože byly duplikovány v IR oblastech; a 4 geny pro rRNA, které byly duplikovány v IR oblastech. Proteiny kódující geny, tRNA a rRNA tvořily u zkoumaných druhů rodu *Dactylorhiza* zhruba 70 % celého genomu chloroplastu. Z toho připadalo zhruba 48 % na proteiny kódující geny, 19 % na geny pro tRNA a 3 % na geny pro rRNA. Zbylých 30 % genomu tvořily inter-genové mezerníky, introny a pseudogeny. V oblasti velké kódující podjednotky bylo přítomno přibližně 47 % genů, v oblasti malé kódující podjednotky přibližně 9 % genů a v oblastech duplikace přibližně celkem 44 % genů.

# 10 Seznam použité literatury


Aagaard S. M. D., Såstad S. M., Greilhuber J., Moen, A. (2005). A secondary hybrid zone between diploid *Dactylorhiza incarnata* ssp. *cruenta* and allotetraploid *D. lapponica* (*Orchidaceae*). Heredity 94(5): 488.

Adams K. L., Wendel J. F. (2005). Polyploidy and genome evolution in plants. Current Opinion in Plant Biology 8: 135–141.

Altschul S. F., Gish W., Miller W., Myers E. W., Lipman D. J. (1990). Basic local alignment search tool. Journal of molecular biology 215(3): 403-410.

Alverson A. J., Zhuo S., Rice D. W., Sloan D. B., Palmer J. D. (2011). The mitochondrial genome of the legume *Vigna radiata* and the analysis of recombination across short mitochondrial repeats. PLOS ONE 6: e16404.

Ambardar S., Gupta R., Trakroo D., Lal R., Vakhlu J. (2016). High throughput sequencing: an overview of sequencing chemistry. Indian journal of microbiology 56(4): 394-404.

Archibald J. M. (2015). Endosymbiosis and eukaryotic cell evolution. Current Biology 25(19): R911-R921.

Arrieta-Montiel M. P., Shedge V., Davila J., Christensen A. C., Mackenzie S. A. (2009). Diversity of the *Arabidopsis* mitochondrial genome occurs via nuclear-controlled recombination activity. Genetics 183: 1261–68.

Balao F., Tannhäuser M., Lorenzo M. T., Hedrén M., Paun O. (2016). Genetic differentiation and admixture between sibling allopolyploids in the *Dactylorhiza majalis* complex. Heredity 116(4): 351-361.

Balao F., Tannhäuser M., Lorenzo M. T., Hedrén M., Paun O. (2016). Genetic differentiation and admixture between sibling allopolyploids in the *Dactylorhiza majalis* complex. Heredity 116(4): 351.

Barbrook A. C., Howe C. J., Kurniawan D. P., Tarr S. J. (2010). Organization and expression of organellar genomes. Philosophical Transactions of the Royal Society B 365: 785-797.

Barbrook A. C., Howe C. J., Purton S. (2006). Why are plastid genomes retained in nonphotosynthetic organisms? Trends in Plant Science 11: 101-108.

Baur E. (1909). Das Wesen und die Erblichkeitsverhältnisse der "Varietates albomarginatae hort" von *Pelargonium zonale*. Z Indukt Abstamm Vererbungsl 1: 330–351.

Baur E. (1910). Untersuchungen über die Vererbungs von Chromatophorenmerkmalen bei *Melandrium*, *Antirrhinum* und *Aquilegia*. Z Indukt Abstamm Vererbungsl 4: 81–102.

Baur E. (1911). Einführung in die experimentelleVererbungslehre. Gebr. Borntraeger Berlin (2.AuX. 1914, 3. AuX. 1919).

Beale G., Knowles J. (1978). Extranuclear genetics. Edward Arnold, London.

Bennetzen J. L., Ma J., Devos K. M. (2005). Mechanisms of recent genome size variation in flowering plants. Annals of Botany 95: 127–132.

Bentley D. R., Balasubramanian S., Swerdlow H. P., Smith G. P., Milton J., Brown C. G., *et al.* (2008). Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456(7218): 53-59.

Bergthorsson U., Adams K. L., Thomason B., Palmer J. D. (2003). Widespread horizontal transfer of mitochondrial genes in flowering plants. Nature 424: 197–201.

Blazier J., Guisinger M. M., Jansen R. K. (2011). Recent loss of plastid-encoded ndh genes within *Erodium* (*Geraniaceae*). Plant Molecular Biology 76: 263–72.

Bobik K., Burch-Smith T. M. (2015). Chloroplast signaling within, between and beyond cells. Frontiers in Plant Science 6: 781.

Boisvert S., Laviolette F., Corbeil J. (2010). Ray: simultaneous assembly of reads from a mix of high-throughput sequencing technologies. Journal of computational biology 17(11): 1519-1533.

Bolger A. M., Lohse M., Usadel B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. Bioinformatics 30(15): 2114-2120.

Bonatelli I. A. S., Zappi D. C., Taylor N. P., Moraes E. M. (2013). Usefulness of cpDNA markers for phylogenetic and phylogeographic analyses of closely-related cactus species. Genetics and Molecular Research 12(4): 4579-4585.

Borner T., Aleynikova A. Y., Zubo Y. O., Kusnetsov V. V. (2015). Chloroplast RNA polymerases: role in chloroplast biogenesis. Biochimica et Biophysica Acta 1847: 761-769.

Bortiri E., Coleman-Derr D., Lazo G. R., Anderson O. D., Gu Y. Q. (2008). The complete chloroplast genome sequence of *Brachypodium distachyon*: sequence comparison and phylogenetic analysis of eight grass plastomes. BMC Research Notes 1: 61.

Bowers J. E., Chapman B. A., Rong J., Paterson A. H. (2003). Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature 422(6930): 433-438.

Braukmann T. W. A., Kuzmina M., Stefanović S. (2009). Loss of all plastid ndh genes in *Gnetales* and conifers: extent and evolutionary significance for the seed plant phylogeny. Current Genetics 55: 323–37.

Collins F. S., Morgan M., Patrinos A. (2003). The Human Genome Project: lessons from large-scale biology. Science 300(5617): 286-290.

Comai L., Madlung A., Josefsson C., Tyagi A. (2003). Do the different parental 'heteromes' cause genomic shock in newly formed allopolyploids? Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 358: 1149–1155.

Correns C. (1909). Vererbungsversuche mit blass(gelb)grünen und buntblättrigen Sippen bei *Mirabilis jalapa*, *Urtica pilulifera* und *Lunaria annua*. Z Indukt Abstamm Vererbungsl 1: 291–329.

da Rocha Perini V., Leles B., Furtado C., Prosdocimi F. (2016). Complete chloroplast genome of the orchid *Cattleya crispata* (*Orchidaceae*: *Laeliinae*), a Neotropical rupiculous species. Mitochondrial DNA Part A 27(6): 4075-4077.

Daniell H., Chan H. T., Pasoreck E. K. (2016). Vaccination through chloroplast genetics: affordable protein drugs for the prevention and treatment of inherited or infectious human diseases. Annual review of genetics 50: 595-618.

Daniell H., Wurdack K. J., Kanagaraj A., Lee S. B., Saski C., Jansen R. K. (2008). The complete nucleotide sequence of the cassava (*Manihot esculenta*) chloroplast genome and the evolution of atpF in *Malpighiales*: RNA editing and multiple losses of a group II intron. Theoretical and Applied Genetics 116: 723–37.

Davila J. I., Arrieta-Montiel M. P., Wamboldt Y., Cao J., Hagmann J., *et al.* (2011). Double-strand break repair processes drive evolution of the mitochondrial genome in *Arabidopsis*. BMC Biology 9: 64.

Deamer D., Akeson M., Branton D. (2016). Three decades of nanopore sequencing. Nature biotechnology 34: 518–24.

Dodsworth S., Leitch A. R., Leitch I. J. (2015). Genome size diversity in angiosperms and its influence on gene space. Current Opinion in Genetics & Development 35: 73-78.

Doležel J., Bartoš J. A. N. (2005). Plant DNA flow cytometry and estimation of nuclear genome size. Annals of Botany 95(1): 99-110.

Doležel J., Greilhuber J., Suda J. (2007). Estimation of nuclear DNA content in plants using flow cytometry. Nature protocols 2(9): 2233-2244.

Dovichi N. J., Zhang J. (2000). How capillary electrophoresis sequenced the human genome. 39: 4463-8.

Ephrussi B. (1949). Action de l'acriXavine sur les levures. In: Unites biologiques douees de continuite genetique. du Centre Nat Rech Sci, Paris, pp 165–180.

Fakruddin M., Chowdhury A., Hossain M., Mannan K. S. B., Mazumdar R. M. (2012). Pyrosequencing-principles and applications. Life 50: 65.

Feldman M., Levy A. A. (2009). Genome evolution in allopolyploid wheata revolutionary reprogramming followed by gradual changes. Journal of Genetics and Genomics 36: 511–518.

Gascuel O. (1997). BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. Molecular Biology and Evolution 14: 685-695.

Gibor A., Izawa M. (1963). The DNA content of the chloroplasts of *Acetabularia*. USA 50: 1164–1169.

Gill N., Hans C. S., Jackson S. (2008). An overview of plant chromosome structure. Cytogenetic and genome research 120(3-4): 194-201.

Golczyk H., Greiner S., Wanner G., Weihe A., Bock R., Borner T., Herrmann R. G. (2014). Chloroplast DNA in mature and senescing leaves: a reappraisal. Plant Cell 26: 847-854.

Goodwin S., McPherson J. D., McCombie W. R. (2016). Coming of age: ten years of next-generation sequencing technologies. Nature Reviews Genetics 17(6): 333-351.

Granick S., Gibor A. (1967). The DNA of chloroplasts, mitochondria and centrioles. 6: 143–186.

Green B. R. (2011). Chloroplast genomes of photosynthetic eukaryotes. The Plant Journal 66: 34-44.

Greilhuber J., Borsch T., Müller K., Worberg A., Porembski S., Barthlott W. (2006). Smallest angiosperm genomes found in *Lentibulariaceae*, with chromosomes of bacterial size. Plant Biology 8(06): 770-777.

Grover C. E., Wendel J. F. (2010). Recent insights into mechanisms of genome size change in plants. Journal of Botany 2010: article ID 382732.

Gualberto J. M., Mileshina D., Wallet C., Niazi A. K., Weber-Lotfi F., Dietrich A. (2014). The plant mitochondrial genome: dynamics and maintenance. Biochimie 100: 107–20.

Guo J., Xu N., Li Z., Zhang S., Wu J., Kim D. H., Shi S. *el al.* (2008). Four-color DNA sequencing with 3′-O-modified nucleotide reversible terminators and chemically cleavable fluorescent dideoxynucleotides. Proceedings of the National Academy of Sciences 105(27): 9145-9150.

Hagemann R. (1968). Extrachromosomale Vererbung. Fortschr Botanik 30: 225–241.

Hansen C. N., Heslop-Harrison J. S. (2004). Sequences and phylogenies of plant pararetroviruses, viruses and transposable elements. h 41: 165–193.

Hawkins J. S., Grover C. E., Wendel J. F. (2008). Repeated big bangs and the expanding universe: directionality in plant genome size evolution. Plant Science 174: 557–562.

Hawkins J. S., Kim H. R., Nason J. D., Wing R. A., Wendel J. F. (2006). Differential lineage-specific amplification of transposable elements is responsible for genome size variation in Gossypium. Genome Research 16: 1252–1261.

Heather J. M., Chain B. (2016). The sequence of sequencers: the history of sequencing DNA. Genomics 107(1): 1-8.

Hedrén M., Nordström S., Ståhlberg D. (2012). Geographical variation and systematics of the tetraploid marsh orchid *Dactylorhiza majalis* subsp. *sphagnicola* (*Orchidaceae*) and closely related taxa. Botanical Journal of the Linnean Society 168(2): 174-193.

Heslop-Harrison J. S., Schwarzacher T. (2011). Organisation of the plant genome in chromosomes. The Plant Journal 66(1): 18-33.

Hollingsworth M. L., Andra Clark A. L. E. X., Forrest L. L., Richardson J., Pennington R., Long, D. G., *et al.* (2009). Selecting barcoding loci for plants: evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. Molecular Ecology Resources 9(2): 439-457.

Hollingsworth P. M., Graham S. W., Little D. P. (2011). Choosing and using a plant DNA Barcode. PLoS ONE 6: e19254.

Hou C., Wikström N., Strijk J. S., Rydin C. (2016). Resolving phylogenetic relationships and species delimitations in closely related gymnosperms using high-throughput NGS, Sanger sequencing and morphology. Plant Systematics and Evolution 302(9): 1345-1365.

Howe C. J., Barbrook A. C., Nisbet R. E., Lockhart P. J., Larkum A. W. (2008). The origin of plastids. Philosophical Transactions of the Royal Society B 363: 2675-2685.

Hřibová E., Neumann P., Matsumoto T., Roux N., Macas J., Doležel J. (2010). Repetitive part of the banana (*Musa acuminata*) genome investigated by low-depth 454 sequencing. BMC Plant Biology 10: 204.

Huot J. L., Enkler L., Megel C., Karim L., Laporte D., Becker H. D., Duchene A. M., Sissler M., Marechal-Drouard L. (2014) Idiosyncrasies in decoding mitochondrial genomes. Biochimie 100: 95-106.

Huse S. M., Huber J. A., Morrison H. G., Sogin M. L., Welch D. M. (2007). Accuracy and quality of massively parallel DNA pyrosequencing. Genome biology 8(7): R143.

Chang C. C., Lin H. C., Lin I. P., Chow T. Y., Chen H. H., Chen W. H., *et al.* (2006). The chloroplast genome of *Phalaenopsis aphrodite* (*Orchidaceae*): comparative analysis of evolutionary rate with that of grasses and its phylogenetic implications. Molecular Biology and Evolution 23: 279–91.

Chang C. C., Lin H. C., Lin I. P., Chow T. Y., Chen H. H., Chen W. H., Chaw S. M. (2005). The chloroplast genome of *Phalaenopsis aphrodite* (*Orchidaceae*): comparative analysis of evolutionary rate with that of grasses and its phylogenetic implications. Molecular Biology and Evolution 23(2): 279-291.

Chase M. W., Cowan R. S., Hollingsworth P. M., van den Berg C., Madriñán S., Petersen G., *et al.* (2007). A proposal for a standardised protocol to barcode all land plants. Taxon 56(2): 295-299.

Chen H., Yu Y., Chen X., Zhang Z., Gong C., Li J., Wang A. (2015). Plastid DNA insertions in plant nuclear genomes: the sites, abundance and ages, and a predicted promoter analysis. Functional & integrative genomics 15(2): 131-139.

Chen Z. J., Ni Z. F. (2006) Mechanisms of genomic rearrangements and gene expression changes in plant polyploids. Bioessays 28: 240–252.

Illumina. [online] [navštíveno 5.2.2018] Dostupné z https://www.illumina.com/science/technology/next-generationsequencing/sequencing-technology.html.

International human genome sequencing consortium (2004). Finishing the euchromatic sequence of the human genome. Nature: 431: 931-45.

Jaillon O., Aury J. M., Noel B., Policriti A., Clepet C., Casagrande A., *et al.* (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449(7161): 463.

Jain M., Olsen H. E., Paten B., Akeson, M. (2016). The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. Genome biology 17(1): 239.

Jansen R. K., Cai Z., Raubeson L. A., Daniell H., Leebens-Mack J., Müller K. F., *et al.* (2007). Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proceedings of the National Academy of Science of the USA 104: 19369–74.

Jansen R. K., Wojciechowski M. F., Sanniyasi E., Lee S. B., Daniell H. (2008). Complete plastid genome sequence of the chickpea (*Cicer arietinum*) and the phylogenetic distribution of rps12 and clpP intron losses among legumes (*Leguminosae*). Molecular Phylogenetics and Evolution 48: 1204–17.

Jeelani S. M., Rani S., Kumar S., Kumari S., Gupta R. C. (2013). Cytological studies of *Brassicaceae* Burn. (*Cruciferae* Juss.) from Western Himalayas. Cytology and genetics 47(1): 20-28.

Jensen P. E., Leister D. (2014). Chloroplast evolution, structure and functions. F1000Prime Reports 6: 40.

Jiao Y., Wickett N. J., Ayyampalayam S., Chanderbali A. S., Landherr L., *et al.* (2011) Ancestral polyploidy in seed plants and angiosperms. Nature 473: 97–U113.

Jones R. N., Viegas W., Houben A. (2008). A century of b chromosomes in plants: so what? Annals of Botany 101: 767–775.

Karger B. L., Guttman A. (2009). DNA sequencing by capillary electrophoresis. Electrophoresis 30: 196-202.

Katoh K., Kuma K., Toh H., Miyata T. (2005). MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids research 33(2): 511-518.

Kelly L. J., Leitch A. R., Fay M. F., Renny-Byfield S., Pellicer J., Macas J., Leitch I. J. (2012). Why size really matters when sequencing plant genomes. Plant Ecology & Diversity 5: 415–425.

Kelly L. J., Leitch I. J. (2011). Exploring giant plant genomes with next-generation sequencing technology. Chromosome Research 19(7): 939-953.

Kirk J. T. O. (1963). The deoxyribonucleic acid of broad bean chloroplasts. 76: 130–141.

Kirk J. T. O. (1986). The discovery of chloroplast DNA. Bioessays 4: 36–38.

Kirk J. T. O., Tilney-Bassett R. A. E. (1967). The plastids: their chemistry, structure, growth and inheritance. W H Freeman, London.

Kolísko M. (2017). Principy sekvenování DNA vybranými moderními metodami. Živa 3: 120.

Konhar R., Biswal D. K., Debnath M., Parameswaran S., Sundar D., Tandon P. (2016). Complete chloroplast genome sequence of *Dendrobium nobile* from northeastern India. Genome announcements 4(5): e01088-16.

Krawczyk K., Wiland-Szymańska J., Buczkowska-Chmielewska K., Drapikowska M., Maślak M., Myszczyński K., Sawicki, J. (2017). The complete chloroplast genome of a rare orchid species *Liparis loeselii* (L.). Conservation Genetics Resources 1-4.

Krupinska K., Melonek J., Krause K. (2013). New insights into plastid nucleoid structure and functionality. Planta 237: 653-664.

Kubo T., Newton K. J. (2008). Angiosperm mitochondrial genomes and mutations, Mitochondrion 8: 5e14.

LABGuide. Průvodce laboratoří. [online] [navštíveno 5.2.2018] Dostupné z http://labguide.cz.

Lander E. S., Linton L. M., Birren B., Nusbaum C., Zody M. C., Baldwin J., *et al.* (2001). Initial sequencing and analysis of the human genome. Nature 409(6822): 860-921.

Leff J., Mandel M., Epstein H. T., Schiff J. A. (1963). DNA satellites from cells of green and aplastidic algae. 13: 126–130i.

Leitch A. R., Leitch I. J. (2008). Perspective – Genomic plasticity and the diversity of polyploid plants. Science 320: 481–483.

Leitch I. J., Bennett M. D. (2004). Genome downsizing in polyploid plants. Biological journal of the Linnean Society 82(4): 651-663.

Leitch I. J., Hanson L., Lim K. Y., Kovarik A., Chase M. W., Clarkson J. J., Leitch A. R. (2008). The ups and downs of genome size evolution in polyploid species of *Nicotiana* (*Solanaceae*). Annals of Botany 101: 805–814.

Leitch I. J., Soltis D. E., Soltis P. S., Bennett M. D. (2005). Evolution of DNA amounts across land plants (*Embryophyta*). Annals of Botany 95(1): 207-217.

Lemmon E. M., Lemmon A. R. (2013). High-throughput genomic data in systematice and phylogenetics. Annual Review of Ecology, Evolution, and Systematics 44: 99–121.

Li H., Durbin R. (2009). Fast and accurate short read alignment with Burrows-Wheeler Transform. Bioinformatics 25: 1754-60.

Li H., Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25(14): 1754-1760.

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., *et al.* (2009). The Sequence alignment/map (SAM) format and SAMtools. Bioinformatics, 25: 2078-2079.

Lim K. Y., Matyasek R., Kovarik A., Leitch A. R. (2004). Genome evolution in allotetraploid *Nicotiana*. Biological Journal of the Linnean Society 82: 599–606.

Lin C. S., Chen J. J., Huang Y. T., Chan M. T., Daniell H., Chang W. J., *et al.* (2015). The location and translocation of ndh genes of chloroplast origin in the *Orchidaceae* family. Scientifics Reports 5: 9040.

Liu B., Wendel J. F. (2003). Epigenetic phenomena and the evolution of plant allopolyploids. Molecular Phylogenetics and Evolution 29: 365–379.

Lohse M., Drechsel O., Kahlau S., Bock R. (2013). Organellar Genome DRAW – a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. Nucleic Acids Research 41: 575-581.

Lu H., Giordano F., Ning Z. (2016). Oxford Nanopore MinION sequencing and genome assembly. Genomics, proteomics & bioinformatics 14(5): 265-279.

Macas J., Neumann P., Navratilova A. (2007). Repetitive DNA in the pea (*Pisum sativum* L.) genome: comprehensive characterization using 454 sequencing and comparison to soybean and *Medicago truncatula*. BMC Genomics 8: 427.

Malapelle U., Vigliar E., Sgariglia R., Bellevicine C., Colarossi L., Vitale D., *et al.* (2015). Ion Torrent next-generation sequencing for routine identification of clinically relevant mutations in colorectal cancer patients. Journal of clinical pathology 68(1): 64-68.

Mandel J. R., Dikow R. B., Funk V. A. *et al.* (2014). A target enrichment method for gathering phylogenetic information from hundreds of loci: an example from the *Compositae*. Applications in Plant Sciences 2: 1300085.

Mardis E. R. (2008). The impact of next-generation sequencing technology on genetics. Trends in genetics 24(3): 133-141.

Margulies M., Egholm M., Altman W. E., Attiya S., Bader J. S., Bemben L. A., *et al.* (2005). Genome sequencing in open microfabricated high density picoliter reactors. Nature 437: 376-380.

Martínez-García P. J., Crepeau M. W., Puiu D., Gonzalez-Ibeas D., Whalen J., Stevens K. A., Chakraborty S. (2016). The walnut (*Juglans regia*) genome sequence reveals diversity in genes coding for the biosynthesis of non-structural polyphenols. The Plant Journal 87(5): 507-532.

Marzorati M., Maignien L., Verhelst A., Luta G., Sinnott R., Kerckhof F. M., Possemiers S. (2013). Barcoded pyrosequencing analysis of the microbial community in a simulator of the human gastrointestinal tract showed a colon region-specific microbiota modulation for two plant-derived polysaccharide blends. Antonie Van Leeuwenhoek 103: 409–420.

Mascher M., Amand P. S., Stein N., Poland J. (2013). Application of genotyping-by-sequencing on semiconductor sequencing platforms: a comparison of genetic and reference-based marker ordering in barley. PLoS ONE 8: e76925.

Matyasek R., Tate J. A., Lim Y. K., Srubarova H., Koh J., *et al.* (2007). Concerted evolution of rDNA in recently formed *Tragopogon* allotetraploids is typically associated with

an inverse correlation between gene copy number and expression. Genetics 176: 2509–2519.

Maxam A. M., Gilbert W. (1977). A new method for sequencing DNA. Proceedings of the National Academy of Sciences 74: 560–564.

McCarthy A. (2010). Third generation DNA sequencing: pacific biosciences' single molecule real time technology. Chemistry & biology 17(7): 675-676.

McCoy S. R., Kuehl J. V., Boore J. L., Raubeson L. A. (2008). The complete plastid genome sequence of *Welwitschia mirabilis*: an unusually compact plastome with accelerated divergence rates. BMC Evolutionary Biology 8: 130.

McKernan K. J., Peckham H. E., Costa G. L., McLaughlin S. F., Fu Y., Tsung E. F., *et al.* (2009). Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. Genome research 19(9): 1527-1541.

Melonek J., Mulisch M., Schmitz-Linneweber C., Grabowski E., Hensel G., Krupinska K. (2010). Whirly1 in chloroplasts associates with intron containing RNAs and rarely colocalizes with nucleoids. Planta 232: 471-481.

Metzker M. L. (2009). Sequencing technologies: the next generation. Nature Reviews Genetics 11: 31–46.

Ming R., Hou S., Feng Y., *et al.* (2008). The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* L.). Nature 452: 991–996.

Mirsky A. E., Ris H. (1951). The desoxyribonucleic acid content of animal cells and its evolutionary significance. Journal of General Physiology 34: 451–462.

Moore M. J., Dhingra A., Soltis P. S., Shaw R., Farmerie W. G., Folta K. M., Soltis D. E. (2006). Rapid and accurate pyrosequencing of angiosperm plastid genomes. BMC Plant Biology 6: 17.

Munekage Y., Hashimoto M., Miyake C., Tomizawa K. I., Endo T., Tasaka M., Shikanai T. (2004). Cyclic electron flow around photosystem I is essential for photosynthesis. Nature 429: 579-82.

Naczk A. M., Górniak M., Szlachetko D. L., Ziętara M. S. (2015). Plastid DNA haplotype diversity and morphological variation in the *Dactylorhiza incarnata*/*maculata* complex (*Orchidaceae*) in northern Poland. Botanical journal of the Linnean Society 178(1): 121-137.

Nazareno A. G., Carlsen M., Lohmann L. G. (2015). Complete chloroplast genome of *Tanaecium tetragonolobum*: the first *Bignoniaceae* plastome. PLoS One 10: 129930.

Neumann P., Koblizkova A., Navratilova A., Macas J. (2006). Significant expansion of *Vicia pannonica* genome size mediated by amplification of a single type of giant retroelement. Genetics 173: 1047–1056.

Niu Z., Xue Q., Zhu S., Sun J., Liu W., Ding X. (2017). The complete plastome sequences of four orchid species: Insights into the evolution of the *Orchidaceae* and the utility of plastomic mutational hotspots. Frontiers in plant science 8: 715.

Odintsova M. S., Yurina N. P. (2003). Plastidic genome of higher plants and algae: structure and function. Molecular Biology 37: 1-16.

Ohyama K., Fukuzawa H., Kohchi T., Shirai H., Sano T., *et al.* (1986). Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. Nature 322.6079: 572-574.

Oldenburg D. J., Bendich A. J. (2015). DNA maintenance in plastids and mitochondria of plants. Frontiers in Plant Science 6: 883.

Ortelt J., Link G. (2014). Plastid gene transcription: promoters and RNA polymerases. Methods in Moleclar Biology 1132: 47-72.

Otto, S. P., Whitton, J. (2000). Polyploid incidence and evolution. Annual review of genetics 34(1): 401-437.

Oudot-Le Secq M. P., Green B. R. (2011). Complex repeat structures and novel features in the mitochondrial genomes of the diatoms *Phaeodactylum tricornutum* and *Thalassiosira pseudonana*. Gene 476: 20-26.

Pan I. C., Liao D. C., Wu F. H., Daniell H., Singh N. D., Chang C., *et al.* (2012). Complete chloroplast genome sequence of an orchid model plant candidate: *Erycina pusilla* apply in tropical oncidium breeding. PLoS One 7: e34738.

Paun O., Bateman R. M., Fay M. F., Luna J. A., Moat J., Hedrén M., Chase M. W. (2011). Altered gene expression and ecological divergence in sibling allopolyploids of *Dactylorhiza* (*Orchidaceae*). BMC evolutionary biology 11(1): 113.

Paux E., Roger D., Badaeva E., Gay G., Bernard M., Sourdille P., Feuillet C. (2006). Characterizing the composition and evolution of homoeologous genomes in hexaploid wheat through BAC-end sequencing on chromosome 3B. The Plant Journal 48(3): 463-474.

Pellicer J., Fay M. F., Leitch I. J. (2010). The largest eukaryotic genome of them all? Botanical Journal of the Linnean Society 164: 10–15.

Peltier G., Cournac L. (2002). Chlororespiration. Annual Review of Plant Biology 53: 523–50.

Pfalz J., Pfannschmidt T. (2015). Plastid nucleoids: evolutionary reconstruction of a DNA/protein structure with prokaryotic ancestry. Frontiers in Plant Science 6: 220.

Phillippy A. M., Schatz M. C., Pop M. (2008). Genome assembly forensics: finding the elusive mis-assembly. Genome biology 9(3): R55.

Piegu B., Guyot R., Picault N., Roulin A., Saniyal A., Kim H., Collura K., Brar D. S., Jackson S., Wing R. A. *et al.* (2006). Doubling genome size without polyploidization:

dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. Genome Research 16: 1262–1269.

Pleines T., Jakob S. S., Blattner F. R. (2009). Application of non-coding DNA regions in intraspecific analyses. Plant Systematics and Evolution 282(3-4): 281-294.

Powikrowska M., Oetke S., Jensen P. E., Krupinska K. (2014). Dynamic composition, shaping and organization of plastid nucleoids. Frontiers in Plant Science 5: 424.

Prober J. M., Trainor G. L., Dam R. J., Hobbs F. W., Robertson C. W., Zagursky R. J., *et al.* (1987). A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. Science 238: 336-41.

Proost S., Pattyn P., Gerats T., Van de Peer Y. (2011). Journey through the past: 150 million years of plant genome evolution. The Plant Journal 66(1): 58-65.

Quinlan A. R., Hall I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26(6): 841-842.

Raman G., Park S. (2015). Analysis of the complete chloroplast genome of a medicinal plant, *Dianthus superbus* var. *longicalyncinus*, from a comparative genomics perspective. PLoS One 10: e0141329.

Renner O. (1922). Eiplasma und Pollenschlauchplasma bei den Oenotheren. Z Indukt Abstamm Vererbungsl 27: 235–237.

Renner O. (1924). Die Scheckung der Oenotherenbastarde. Biol Zbl 44: 309–336.

Renner O. (1929). Artbastarde bei PXanzen. Handbuch Vererbungswissensch. Bd IIA. Gebr. Borntraeger, Berlin.

Renner O. (1934). Die pXanzlichen Plastiden als selbständige Elemente der genetischen.

Renner O. (1936). Zur Kenntnis der nichtmendelnden Buntheit der Laubblätter. Flora 130: 218–290.

Renny-Byfield S., Chester M., Kovařík A., *et al.* (2011). Next-generation sequencing reveals genome downsizing in allotetraploid *Nicotiana tabacum*, predominantly through the elimination of paternally derived repetitive DNAs. Molecular biology and evolution 28: 2843–2854.

Reuter J. A., Spacek D. V., Snyder M. P. (2015). High-Throughput Sequencing Technologies. Molecular Cell 58: 586–597.

Rice D. W., Alverson A. J., Richardson A. O., Young G. J., Sanchez-Puerta M. V., Munzinger J., Barry K., Boore J. L., Yan Zhang Y., De Pamphilis C. W., Knox E. B., Palmer J. D. (2013). Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm *Amborella*. Science 342: 1468-1473.

Richardson A. O., Rice D. W., Young G. J., Alverson A. J., Palmer, J. D. (2013). The "fossilized" mitochondrial genome of *Liriodendron tulipifera*: ancestral gene content and order, ancestral editing sites, and extraordinarily low mutation rate. BMC Biology 11: 29.

Ris H. (1961). Ultrastructure and molecular organization of genetic systems. 3: 95–120.

Ris H., Plaut W. (1962). Ultrastructure of DNA-containing areas in the chloroplast of *Chlamydomonas.* The Journal of cell biology 13: 383–391.

Roeh S., Weber P., Rex-Haffner M., Deussing J. M., Binder E. B., Jakovcevski M. (2017). Sequencing on the SOLiD 5500xl System–in-depth characterization of the GC bias. Nucleus 1: 11.

Rochaix J. D., Ramundo S. (2015). Conditional repression of essential chloroplast genes: evidence for new plastid signaling pathways. Biochimica et Biophysica Acta 1847: 986-992.

Ronaghi M., Karamohamed S., Pettersson B., Uhlén M., Nyrén P. (1996). Real-time DNA sequencing using detection of pyrophosphate release. Analytical biochemistry 242(1): 84-89.

Rothberg J. M., Hinz W., Rearick T. M., Schultz J., Mileski W., Davey M., *et al.* (2011). An integrated semiconductor device enabling non-optical genome sequencing. Nature 475(7356): 348-352.

Royal Botanic Gardens, Kew. WCSP. World Checklist of Selected Plant Families. [online] [navštíveno 5.2.2018] Dostupné z http://apps.kew.org/wcsp/.

Ruck E. C., Nakov T., Jansen R. K., Theriot E. C., Alverson A. J. (2014). Serial gene losses and foreign DNA underlie size and sequence variation in the plastid genomes of diatoms. Genome Biology Evolution 6: 644-654.

Ruhlman T. A., Jansen R. K. (2014). The plastid genomes of flowering plants. Methods of Molecular Biology 1132: 3-38.

Sager R. (1972). Cytoplasmic genes and organelles. Academic Press, New York.

Sager R., Ishida M. R. (1963). Chloroplast DNA in *Chlamydomonas*. Proceedings of the National Academy of Sciences USA 50: 725–730.

Sanderson M. J., Copetti D., Búrquez A., Bustamante E., Charboneau J. L., Eguiarte L. E., *et al.* (2015). Exceptional reduction of the plastid genome of saguaro cactus (*Carnegiea gigantea*): loss of the ndh gene suite and inverted repeat. American Journal of Botany 102: 1115–27.

Sanger F., Nicklen S., Coulson A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences 74: 5463–5467.

Saski C., Lee S. B., Daniell H., Wood T. C., Tomkins J., Kim H. G., Jansen R. K. (2005). Complete chloroplast genome sequence of *Glycine max* and comparative analyses with other legume genomes. Plant Moleclar Biology 59: 309–22.

Saski C., Lee S. B., Fjellheim S., Guda C., Jansen R. K., Luo H., *et al.* (2007). Complete chloroplast genome sequences of *Hordeum vulgare*, *Sorghum bicolor* and *Agrostis stolonifera*, and comparative analyses with other grass genomes. Theoretical and Applied Genetics 115: 571-90.

Shao K., Ding W., Wang F., Li H., Ma D., Wang H. (2011). Emulsion PCR: a High Efficient Way of PCR Amplification of Random DNA Libraries in Aptamer Selection. PLoS ONE 6: e24910.

Shaw J., Lickey E. B., Beck J. T. *et al.* (2005). The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. American Journal of Botany 92: 142–166.

Shaw J., Lickey E. B., Schilling E. E., Small R. L. (2007). Comparison of whole chloroplast sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. American Journal of Botany 94: 275–288.

Shinozaki K., Ohme M., Tanaka M., Wakasugi T., *et al.* (1986). The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. The EMBO journal 5(9): 2043.

Schatz M. C., Delcher A. L., Salzberg S. L. (2010). Assembly of large genomes using second-generation sequencing. Genome Research 20: 1165–73.

Schneeberger K., Ossowski S., Ott F., Klein J. D., Wang X., Lanz C., *et al.* (2011). Reference-guided assembly of four diverse *Arabidopsis thaliana* genomes. Proceedings of the National Academy of Sciences 108(25): 10249-10254.

Schneider G. F., Dekker C. (2012). DNA sequencing with nanopores. Nature biotechnology 30(4): 326-328.

Scholz M. B., Lo C. C., Chain P. S. G. (2012). Next generation sequencing and bioinformatic bottlenecks: the current state of metagenomic data analysis. Current Opinion in Biotechnology 23: 9–15.

Schwemmle J. (1940). Plastidenmutationen bei Eu-Oenotheren. Z Indukt Abstamm Vererbungsl 79: 171–187.

Schwemmle J. (1941). Weitere Untersuchungen an Eu-Oenotheren über die genetische Bedeutung des Plasmas und der Plastiden. Z Indukt Abstamm Vererbungsl 79: 321–335.

Schwemmle J. (1943). Plastiden und Genmanifestation. Flora 137: 61–72.

Schwemmle J. (1957). Der EinXuss des Plasmas und der Plastiden auf die AYnität zwischen Samenanlagen und Pollenschläuchen. Biol Zbl 76: 529–549.

Slatko B. E., Kieleczawa J., Ju J., Gardner A. F., Hendrickson C. L., Ausubel F. M. (2011). "First generation" automated DNA sequencing technology. Current protocols in molecular biology 7: 2.

Sloan D. B., Alverson A. J., Chuckalovcak J. P., Wu M., McCauley D. E., Palmer J. D, Taylor D. R. (2012). Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. PLoS Biology 10: e1001241.

Slonimski P., Ephrussi B. (1949). Action de l'acriXavine sur les levures, V. Le systeme des cytochromes des mutants 'petite colonie'. Ann Inst Pasteur 77: 47–83.

Smith D. R. (2014). Buying in to bioinformatics: an introduction to commercial sequence analysis software. Briefings in bioinformatics 16(4): 700-709.

Smith L. M., Sanders J. Z., Kaiser R. J., Hughes P., Dodd C., Connell C. R. *et al.* (1986). Fluorescence detection in automated DNA sequence analysis. Nature 321: 674-9.

Soltis D. E., Albert V. A., Leebens-Mack J., Bell C. D., Paterson A. H., *et al.* (2009) Polyploidy and angiosperm diversification. American Journal of Botany 96: 336–348.

Soltis D. E., Burleigh J. G. (2009). Surviving the KT mass extinction: New perspectives of polyploidization in angiosperms. Proceedings of the National Academy of Sciences, 106(14), 5455-5456.

Soltis P. S., Soltis D. E. (2000). The role of genetic and genomic attributes in the success of polyploids. Proceedings of the National Academy of Sciences 97(13): 7051-7057.

Soltis P. S., Soltis D. E. (2009). The role of hybridization in plant speciation. Annual review of plant biology 60: 561-588.

Sonnhammer E. L., Durbin R. (1995). A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. Gene 167(1): GC1-10.

Staats M., Erkens R. H. J., van de Vossenberg B., *et al.* (2013). Genomic treasure troves: complete genome sequencing of herbarium and insect museum specimens. PLoS ONE 8: e69189.

Suda J. (2005). Co se skrývá za rostlinnou průtokovou cytometrií. Živa 1: 46-48.

Sugiyama Y., Watase Y., Nagase M., Makita N., Yagura S., Hirai A., Sugiura M. (2005). The complete nucleotide sequence and multipartite organization of the tobacco mitochondrial genome: comparative analysis of mitochondrial genomes in higher plants. Molecular Genetics and Genomics 272: 603-615.

Swaminathan K., Varala K., Hudson M. E. (2007). Global repeat discovery and estimation of genomic copy number in a large, complex genome using a high-throughput 454 sequence survey. BMC Genomics 8: 132.

Swerdlow H., Gesteland R. (1990). Capillary gel electrophoresis for rapid, high resolution DNA sequencing. Nucleic Acids Research 18(6): 1415-1419.

Swift H. H. (1950). The constancy of desoxysibose nucleic acid in plant nuclei. Proceedings of the National Academy of Sciences 36: 643-654.

Štorchová H., Olson M. S. (2007). The architecture of the chloroplast *psb*A-*trn*H non-coding region in angiosperms. Plant Systematics and Evolution 268(1): 235-256.

Thomas Jr, C. A. (1971). The genetic organization of chromosomes. Annual review of genetics 5(1): 237-256.

Tiller N., Bock R. (2014). The translational apparatus of plastids and its role in plant development. Molecular Plant 7: 1105–20.

Tipu H. N., Shabbir A. (2015). Evolution of DNA sequencing. Journal of College of Physicians and Surgeons Pakistan 25(4): 210-15.

Turcatti G., Romieu A., Fedurco M., Tairi A. P. (2008). A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis. Nucleic acids research 36(4): e25-e25.

Twyford A. D. (2016). Will Benchtop Sequencers Resolve the Sequencing Trade-off in Plant Genetics? Frontiers in plant science 7.

Ueda M., Kuniyoshi T., Yamamoto H., Sugimoto K., Ishizaki K., Kohchi T., *et al.* (2012). Composition and physiological function of the chloroplast NADH dehydrogenase-like complex in Marchantia polymorpha. Plant Journal 72: 683–93.

Valouev A., Ichikawa J., Tonthat T., Stuart J., Ranade S., Peckham H., *et al.* (2008). A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning. Genome research 18(7): 1051-1063.

Venter J. C., Adams M. D., Myers E. W., Li P. W., Mural R. J., Sutton, G. G., *et al.* (2001). The sequence of the human genome. Science 291(5507): 1304-1351.

Vision T. J., Brown D. G., Tanksley S. D. (2000). The origins of genomic duplications in Arabidopsis. Science 290(5499): 2114-2117.

Vitte C., Panaud O. (2005). LTR retrotransposons and flowering plant genome size: emergence of the increase/decrease model. *Cytogenetic and genome research* 110: 91–107.

Voelkerding K. V., Dames S. A., Durtschi J. D. (2009). Next-generation sequencing: from basic research to diagnostics. *Clinical chemistry 55*(4): 641-658.

Wang W., Li H., Chen, Z. (2014). Analysis of plastid and nuclear DNA data in plant phylogenetics--evaluation and improvement. Science China. Life Sciences 57(3): 280.

Weitemier K., Straub S. C. K., Cronn R. C., *et al.* (2014). HYB-SEQ: combining target enrichment and genome skimming for plant phylogenomics. Applications in Plant Sciences 2: 1400042.

Wendel J. F. (2000). Genome evolution in polyploids. Plant Molecular Biology 42: 225–249.

Weng M. L., Blazier J. C., Govindu M., Jansen R. K. (2014). Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats and nucleotide substitution rates. Molecular Biology and Evolution 31: 645–59.

Wicke S., Muller K. F., De Pamphilis C. W., Quandt D., Wickett N. J., Zhang Y., Renner S. S., Schneeweiss G. M. (2013). Mechanisms of functional and physical genome reduction in photosynthetic and nonphotosynthetic parasitic plants of the broomrape family. Plant Cell 25: 3711-3725.

Wicke S., Schneeweiss G. M., De Pamphilis C. W., Muller K. F., Quandt D. (2011). The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Molecular Biology 76: 273-297.

Wicker T., Sabot F., Hua-Van A., Bennetzen J. L., Capy P., Chalhoub B., Flavell A., Leroy P., Morgante M., Panaud O., Paux E., San Miguel P., Schulman A. H. (2007). A unified classification system for eukaryotic transposable elements. Nature Reviews Genetics 8: 973–982.

Wicker T., Taudien S., Houben A., Keller B., Graner A., Platzer M., Stein N. (2009). A whole-genome snapshot of 454 sequences exposes the composition of the barley genome and provides evidence for parallel evolution of genome size in wheat and barley. Plant Journal 59: 712–722.

Williams J. L., Iamartino D., Pruitt K. D., Sonstegard T., Smith T. P., Low W. Y., Coletta A. (2017). Genome assembly and transcriptome resource for river buffalo, *Bubalus bubalis* (2 n= 50). GigaScience 6(10): 1-6.

Woodson J. D., Chory J. (2012). Organelle signaling: how stressed chloroplasts communicate with the nukleus. Current Biology 22: R690-692.

Wu F. H., Chan M. T., Liao D. C., Hsu C. T., Lee Y. W., Daniell H., *et al.* (2010). Complete chloroplast genome of *Oncidium Gower Ramsey* and evaluation of molecular markers for identification and breeding in *Oncidiinae*. BMC Plant Biology 10: 68.

Wu F. H., Chan M. T., Liao D. C., Hsu C. T., Lee Y. W., Daniell H., Lin C. S. (2010). Complete chloroplast genome of *Oncidium Gower Ramsey* and evaluation of molecular markers for identification and breeding in *Oncidiinae*. BMC plant biology 10(1): 68.

Wu F. H., Kan D. P., Lee S. B., Daniell H., Lee Y. W., Lin C. C., *et al.* (2009). Complete nucleotide sequence of *Dendrocalamus latiflorus* and *Bambusa oldhamii* chloroplast genomes. Tree Physiology 29: 847–56.

Wu J., Liu B., Cheng F., Ramchiary N., Choi S. R., Lim Y. P., Wang X. W. (2012). Sequencing of chloroplast genome using whole cellular DNA and Solexa sequencing technology. Frontiers in Plant Science 3: 234.

Wu Z., Gui S., Quan Z., Pan L., Wang S., Ke W., *et al.* (2014). A precise chloroplast genome of *Nelumbo nucifera* (*Nelumbonaceae*) evaluated with Sanger, Illumina MiSeq, and PacBio RS II sequencing platforms: insight into the plastid evolution of basal eudicots. BMC Plant Biology 14: 289.

Wyman S. K., Jansen R. K., Boore J. L. (2004). Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20(17): 3252-3255.

Xu J. H., Liu Q., Hu W., Wang T., Xue Q., Messing J. (2015). Dynamics of chloroplast genomes in green plants. Genomics 106: 221-231.

Yagi Y., Shiina T. (2012). Evolutionary aspects of plastid proteins involved in transcription: the transcription of a tiny genome is mediated by a complicated machinery. Transcription 3: 290-294.

Yagi Y., Shiina T. (2014). Recent advances in the study of chloroplast gene expression and its evolution. Frontiers in Plant Science 5: 61.

Yang J. B., Tang M., Li H. T., Zhang Z. R., Li D. Z. (2013). Complete chloroplast genome of the genus Cymbidium: lights into the species identification, phylogenetic implications and population genetic analyses. BMC Evolutionary Biology 13: 84.

Yang J. B., Tang M., Li H. T., Zhang Z. R., Li D. Z. (2013). Complete chloroplast genome of the genus *Cymbidium*: lights into the species identification, phylogenetic implications and population genetic analyses. BMC evolutionary biology 13(1): 84.

Yegnasubramanian S. (2013). Preparation of fragment libraries for next-generation sequencing on the applied biosystems SOLiD platform. Methods in enzymology 529: 185.

Yu C. W., Lian Q., Wu K. C., Yu S. H., Xie L. Y., Wu Z. J. (2016). The complete chloroplast genome sequence of *Anoectochilus roxburghii*. Mitochondrial DNA Part A 27(4): 2477-2478.

Yu J., Hu S., Wang J., Wong G. K. S., Li S., Liu B., *et al.* (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science 296(5565): 79-92.

Yu Q. B., Huang C., Yang Z. N. (2014). Nuclear encoded factors associated with the chloroplast transcription machinery of higher plants. Frontiers in Plant Science 5: 316.

Yurina N. P., Odintsova M. S. (2016). Mitochondrial genome structure of photosynthetic eukaryotes. Biochemistry 81(2): 101.

Zhang H., Hall N., McElroy J. S., Lowe E. K., Goertzen L. R. (2017). Complete plastid genome sequence of goosegrass (*Eleusine indica*) and comparison with other *Poaceae*. Gene 600: 36-43.

Zhang Q., Sodmergen X. (2010). Why does biparental plastid inheritance revive in angiosperms? Journal of plant research 123(2): 201-206.

Zimin A. V., Marçais G., Puiu D., Roberts M., Salzberg S. L., Yorke J. A. (2013). The MaSuRCA genome assembler. Bioinformatics 29(21): 2669-2677.

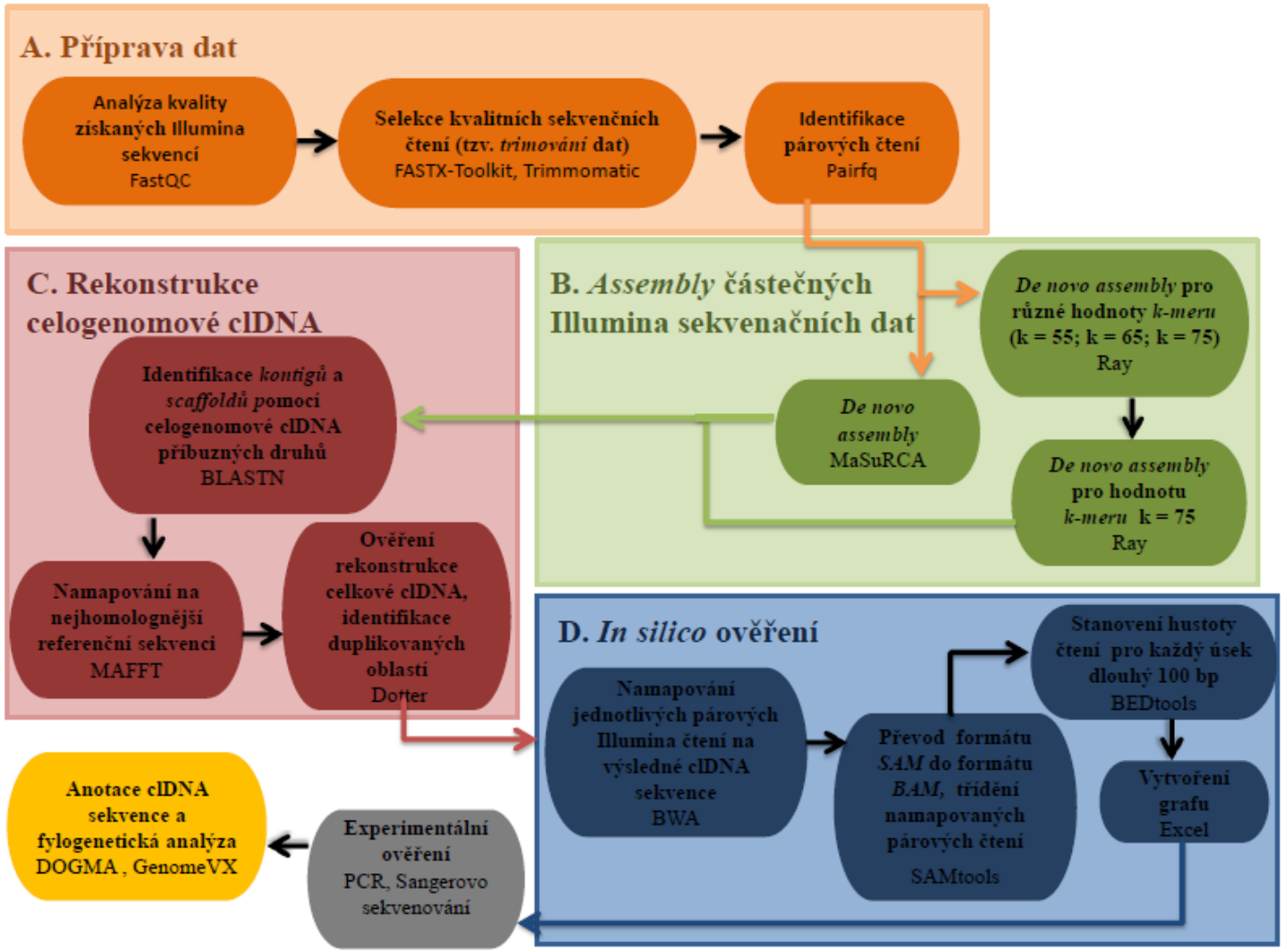
Zimin A. V., Puiu D., Luo M. C., Zhu T., Koren S., Marçais G., Salzberg S. L. (2017a). Hybrid assembly of the large and highly repetitive genome of *Aegilops tauschii*, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm. Genome research 27(5): 787-792.

Zimin A. V., Stevens K. A., Crepeau M. W., Puiu D., Wegrzyn J. L., Yorke J. A., Salzberg, S. L. (2017b). An improved assembly of the loblolly pine mega-genome using long-read single-molecule sequencing. Gigascience 6(1): 1-4.

Zimmer E. A., Wen J. (2013). Reprint of: Using nuclear gene data for plant phylogenetics: Progress and prospects. Molecular phylogenetics and evolution 66(2): 539-550.

# 11 Přílohy

**Příloha 1: Schéma postupu *in silico* analýz**

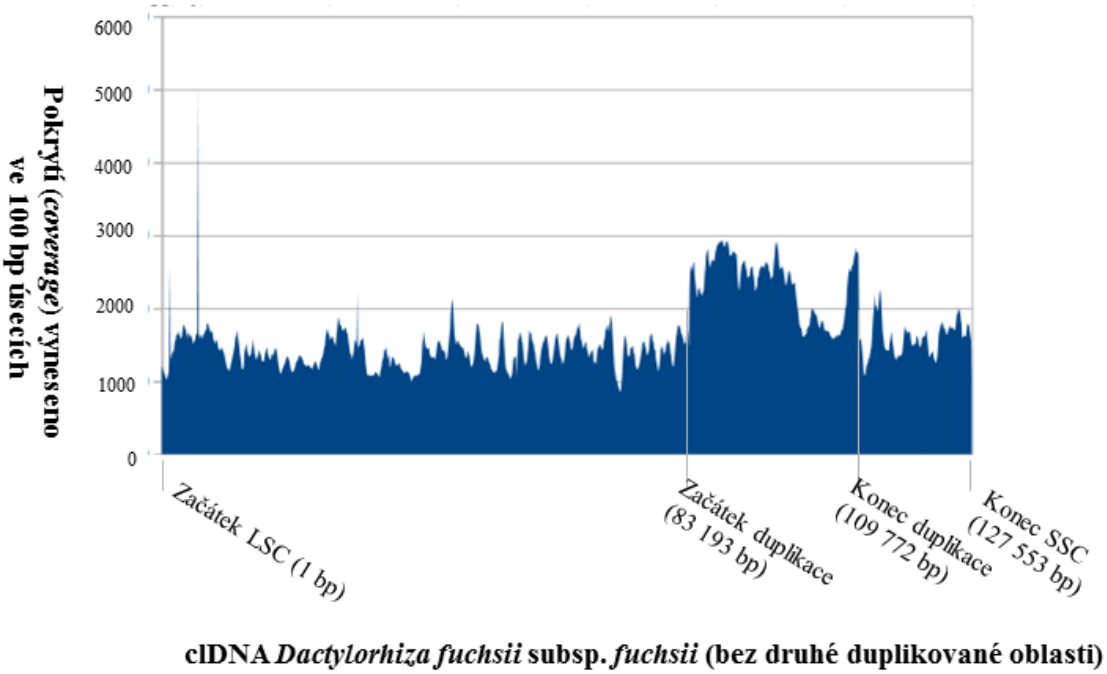


**Příloha 2: Příklad konfiguračního souboru programu MaSuRCA**

```
DATA
PE = pa 600 200 CARP_R1_Qtrim_P.fastq CARP_R2_Qtrim_P.fastq
END
PARAMETERS
GRAPH_KMER_SIZE=auto
USE_LINKING_MATES=1
LIMIT_JUMP_COVERAGE = 300
CA_PARAMETERS = ovlMerSize=30 cgwErrorRate=0.25 ovlStoreMemory=30GB
KMER_COUNT_THRESHOLD = 1
NUM_THREADS= 10
#this is mandatory jellyfish hash size
JF_SIZE=2 046 378 770 #velikost genomu (bp) * coverage / 2
DO_HOMOPOLYMER_TRIM=1
END
```

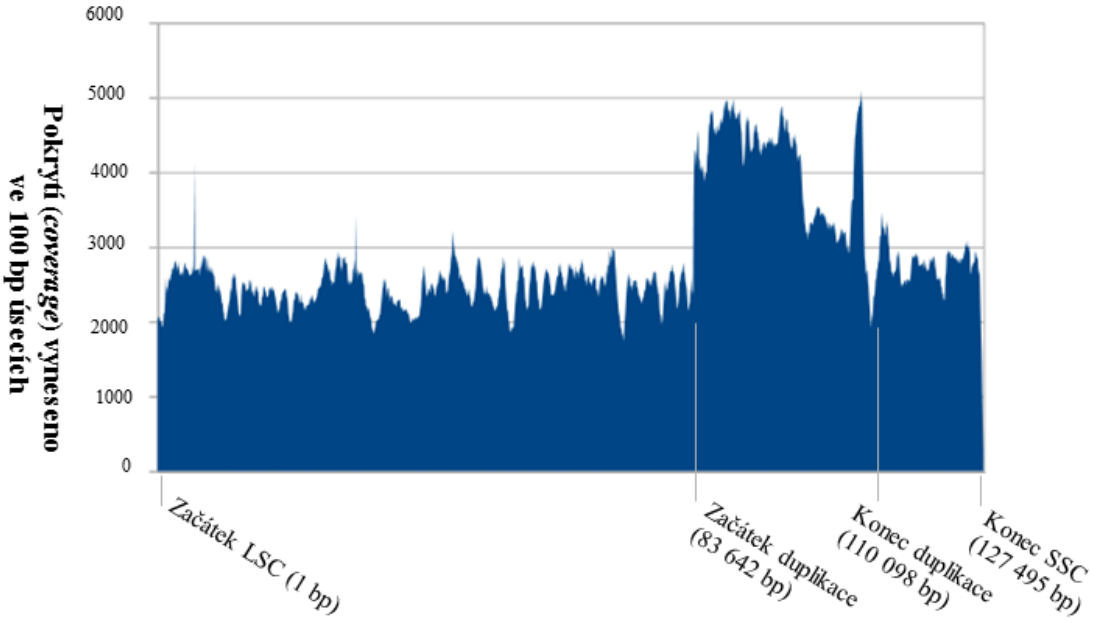
**Příloha 3: Grafické znázornění pokrytí sestavené celogenomové sekvence clDNA u studovaných druhů párovými Illumina čteními (Obr. 1-7)**

Obr. 1: Grafické znázornění pokrytí sestavené celogenomové sekvence clDNA u *Dactylorhiza fuchsii* subsp. *fuchsii* párovými Illumina čteními



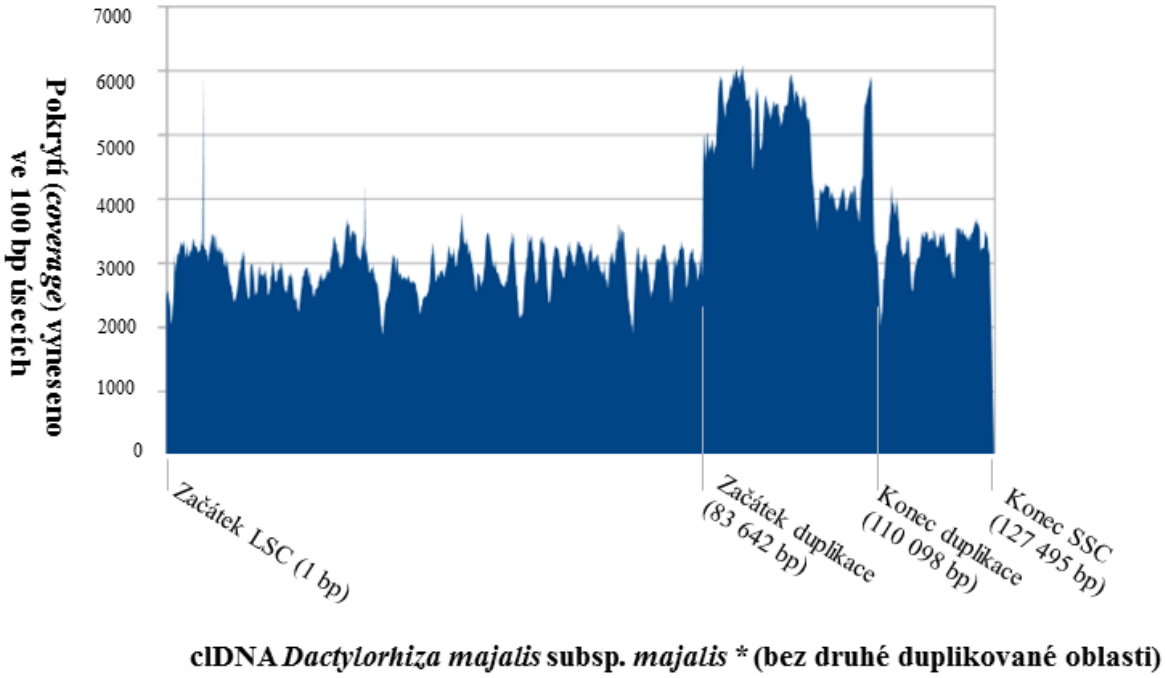

**clDNA *Dactylorhiza fuchsii* subsp. *fuchsii* (bez druhé duplikované oblasti)**

Obr. 2: Grafické znázornění pokrytí sestavené celogenomové sekvence clDNA u *Dactylorhiza majalis* subsp. *majalis* párovými Illumina čteními



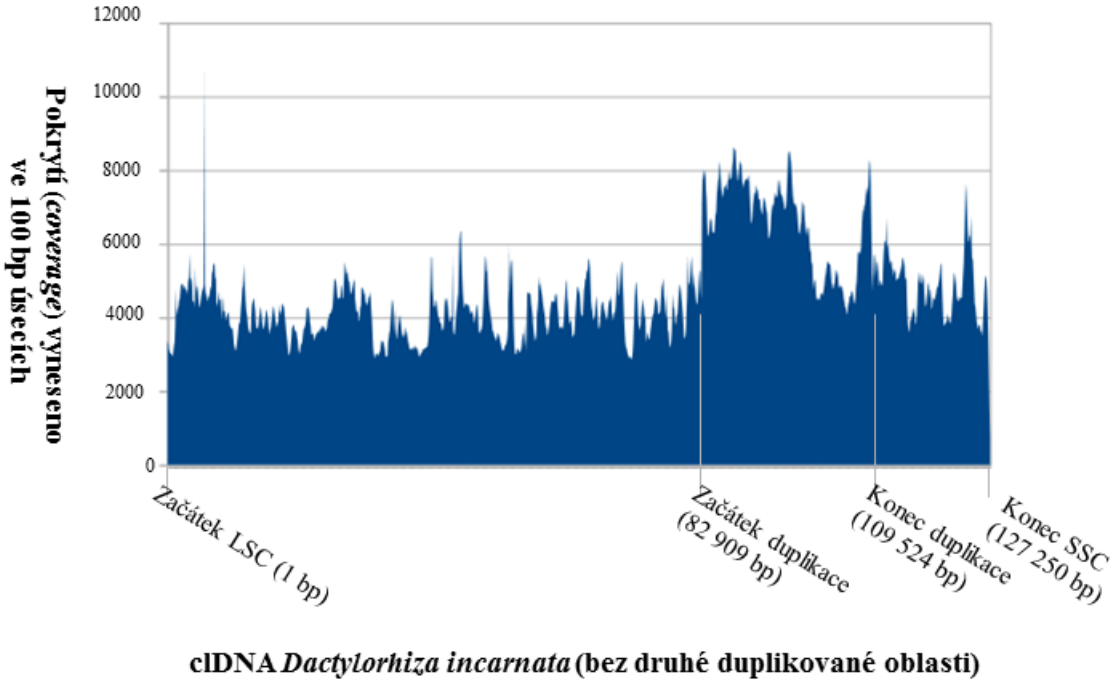

**clDNA *Dactylorhiza majalis* subsp. *majalis* (bez druhé duplikované oblasti)**

Obr. 3: Grafické znázornění pokrytí sestavené celogenomové sekvence clDNA u *Dactylorhiza Dactylorhiza majalis* subsp. *majalis* (morfologicky odlišný taxon z lokality Kalábová) párovými Illumina čteními

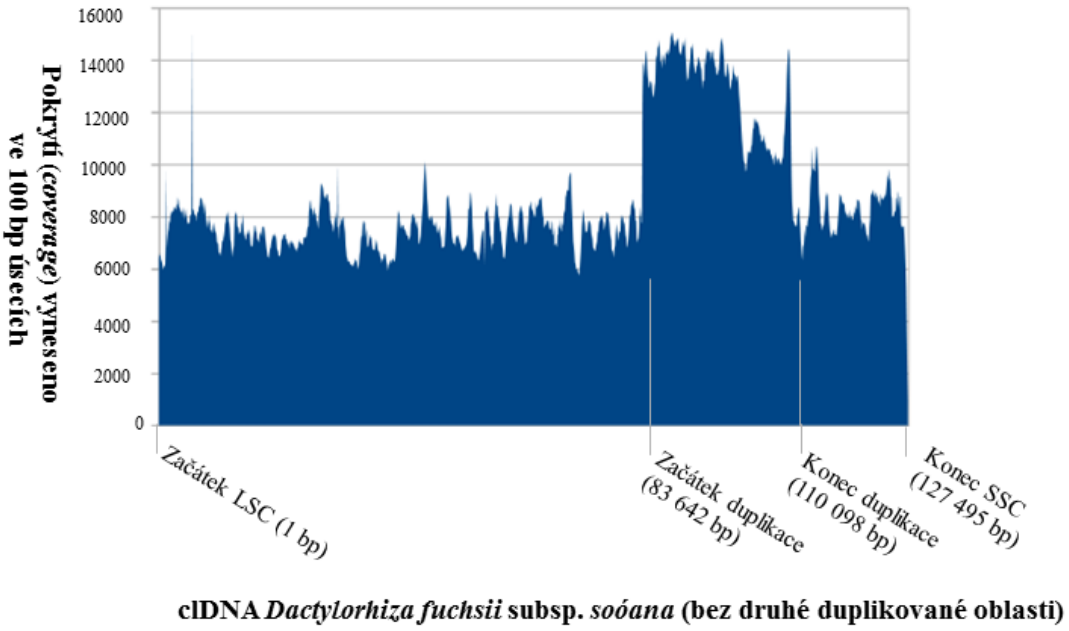



**clDNA *Dactylorhiza majalis* subsp. *majalis* * (bez druhé duplikované oblasti)**

Obr. 4: Grafické znázornění pokrytí sestavené celogenomové sekvence clDNA u *Dactylorhiza incarnata* párovými Illumina čteními

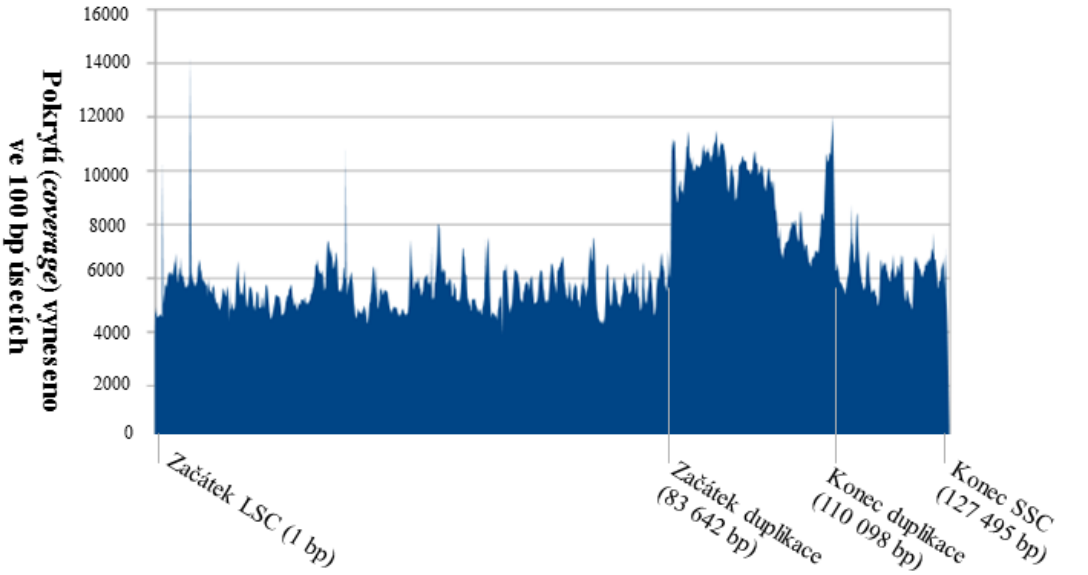


Obr. 5: Grafické znázornění pokrytí sestavené celogenomové sekvence clDNA u *Dactylorhiza fuchsii* subsp. *soóana* párovými Illumina čteními



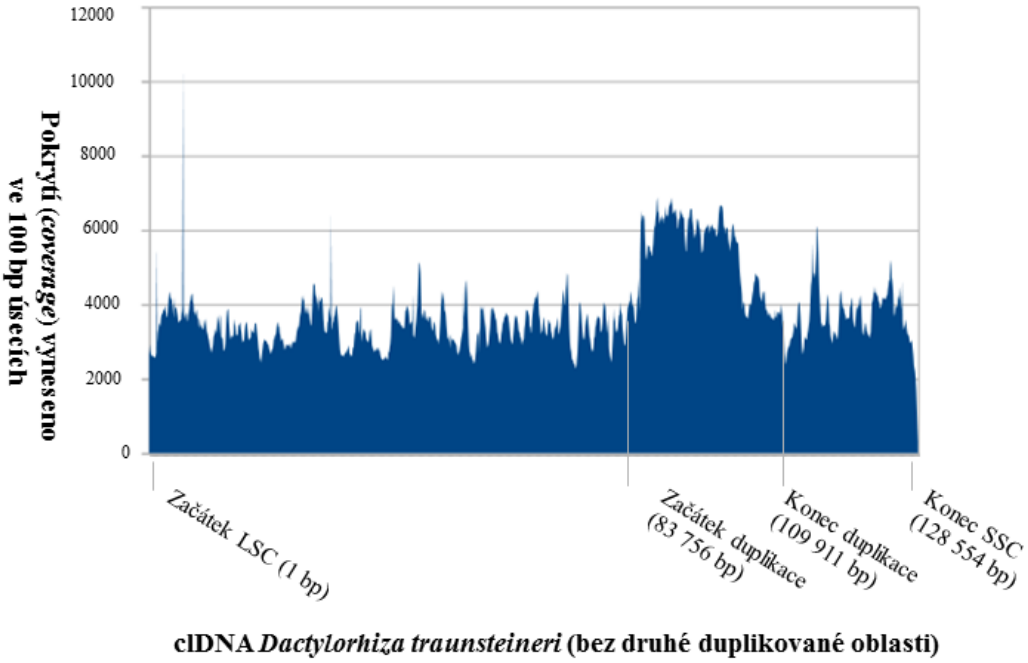

**clDNA *Dactylorhiza fuchsii* subsp. *soóana* (bez druhé duplikované oblasti)**

Obr. 6: Grafické znázornění pokrytí sestavené celogenomové sekvence clDNA u *Dactylorhiza majalis* subsp. t*urfosa* párovými Illumina čteními




**clDNA *Dactylorhiza majalis* subsp. *turfosa* (bez druhé duplikované oblasti)**

Obr. 7: Grafické znázornění pokrytí sestavené celogenomové sekvence clDNA u *Dactylorhiza traunsteineri* párovými Illumina čteními




**clDNA *Dactylorhiza traunsteineri* (bez druhé duplikované oblasti)**