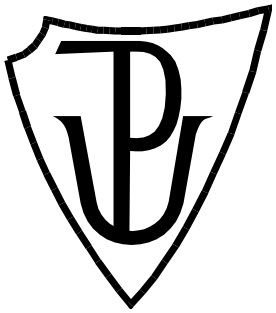
UNIVERZITA PALACKÉHO V OLOMOUCI

Přírodovědecká fakulta

Katedra biochemie



Identifikace a analýza alternativního sestřihu v transkriptomu ječmene

BAKALÁŘSKÁ PRÁCE

Autor:	**Gabriela Majzlíková**
Studijní program:	B1406 Biochemie
Studijní obor:	Bioinformatika
Forma studia:	Prezenční
Vedoucí práce:	**Mgr. Filip Kokáš**
Rok:	2018

Prohlašuji, že jsem bakalářskou práci vypracoval/a samostatně s vyznačením všech použitých pramenů a spoluautorství. Souhlasím se zveřejněním bakalářské práce podle zákona č. 111/1998 Sb., o vysokých školách, ve znění pozdějších předpisů. Byl/a jsem seznámen/a s tím, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorský zákon, ve znění pozdějších předpisů.

V Olomouci dne 9. 1. 2018

Poděkování

Velmi ráda bych touto cestou poděkovala svému vedoucímu bakalářské práce Mgr. Filipovi Kokášovi za odborné vedení, cenné rady, věnovaný čas, vstřícnost, a trpělivost, která mi pomohla k dokončení bakalářské práce.

Bibliografická identifikace

Jméno a příjmení autora	Gabriela Majzlíková
Název práce	Identifikace a analýza alternativního sestřihu v transkriptomu ječmene
Typ práce	Bakalářská
Pracoviště	Katedra biochemie
Vedoucí práce	Mgr. Filip Kokáš
Rok obhajoby práce	2018

Abstrakt

Studium alternativního sestřihu hraje důležitou roli v mnoha výzkumných projektech prováděných na pracovištích molekulární biologie. V rámci bakalářské práce byla vypracována literární rešerše na téma detekce a kvantifikace alternativního sestřihu. Kvalitativní a kvantitativní detekce alternativního sestřihu byl studován v rostlinách transgenního ječmene setého, který obsahoval gen AtCKX1 pod kořenově specifickým promotorem a který byl podroben stresu suchem. Pro analýzu byly použity programy Cufflinks a StringTie. Na základě kvalitativní analýzy izoforem byly zjištěny rozdílné údaje v počtu detekovaných izoforem což bylo způsobeno rozdílným přístupem softwarů k jejich detekci. Pro kvantitativní analýzu izoforem byly vybrány geny CKX (cytokinin oxidáz/reduktáz) u nichž byla následně provedena analýza, zda u nich dochází k odlišnému alternativnímu sestřihu v průběhu vystavení stresu suchem. Výsledky ukazují, že v případě srovnání WT a mutantních rostlin nedochází ve sledovaných časových bodech k výrazným změnám v expresi izoforem sledovaných genů. Zajímavé údaje byly ovšem zjištěny při sledování změn exprese těchto izoforem v průběhu stresu a následném revitalizačním procesu což ukazuje na zapojení studovaných izoforem do procesu adaptace rostlin při vystavení stresu suchem.

Klíčová slova	ječmen, cytokininy, stres suchem, alternativní sestřih, bioinformatika
Počet stran	54
Počet příloh	4
Jazyk	Český

Bibliographical identification

Autor’s first name and surname	Gabriela Majzlíková
Title	Identification and analysis of alternative Splicing in Barley transcriptome
Type of thesis	Bachelor
Department	Department of biochemistry
Supervisor	Mgr. Filip Kokáš
The year of presentation	2018

Abstract

Studies of alternative splicing plays an important role in many research projects carried out at the workplaces of molecular biology. In the framework of the bachelor's thesis was prepared a literature search on the topic detection and quantification of alternative splicing. Qualitative and quantitative alternative splicing detection was studied in transgenic barley plants that contained the AtCKX1 gene under the root-specific promoter and which was subjected to drought stress. For analysis were used the programs Cufflinks and StringTie. Based on qualitative analysis of isoforms have been identified differing data in the number of detected isoforms which was caused by differences in the access software for their detection. CKX genes (cytokinin oxidase / reductase) were selected for quantitative isoform analysis and analyzed whether there was any alternative splicing during dry stress. The results show that, in the case of the comparison of WT and mutant plants, in the monitored time points there were not significant changes in the expression of the isoforms of the monitored genes. Interesting data were found in monitoring changes in the expression of these isoforms during stress and subsequent revitalization process, indicating the involvement of the studied isoforms in the process of plant adaptation when exposed to drought stress.

Keywords	barley, cytokinins, drought stress, alternative splicing, bioinformatics
Number of pages	54
Number of appendices	4
Language	Czech

OBSAH

Cíle práce

1. Vypracování literární rešerše na téma ječmen setý, rostlinný stres a problematika vzniku detekce a kvantifikace alternativního sestřihu.

2. Kvalitativní analýza alternativního sestřihu ve vzorcích ječmene setého divokého typu a transgenního ječmene AtCKX1 vystavených stresu suchem.

3. Kvantitativní analýza alternativního sestřihu vybraných genů CKX (cytokinin oxidáz/reduktáz) v průběhu vystavení stresu suchem.

1 ÚVOD

Ječmen setý (*Hordeum vulgare*) je rostlina řadící se do čeledi lipnicovitých (*Poaceae*). Pomocí genetických modifikací, jsou vytvářeny kultivary ječmene, tolerantní vůči abiotickým a biotickým stresům. Dlouholetá studie genomu ječmene také prokázala, že je o polovinu větší než genom lidský, a to z důvodu velkého obsahu repetitivních sekvencí DNA. Pomocí metod RNA sekvenování je možné ječmen studovat z hlediska kvantifikace genové exprese a rovněž i na úrovni alternativního sestřihu. Alternativní sestřih představuje buněčný proces, během kterého prekurzorová mRNA (pre-mRNA), obsahující kódující a nekódující úseky, které jsou jinak známé jako exony a introny, prochází úpravami za vzniku mRNA. Díky těmto různým sestřihovým variantám vzniká z jednoho genu více bílkovinných produktů. V případě ječmene bylo zjištěno, že téměř 70% genů podléhá alternativnímu sestřihu. Pro zkoumání alternativního sestřihu z kvalitativního i kvantitativního hlediska lze využít řadu softwarových aplikaci.

Bakalářská práce se zaměřuje detekci a následnou kvantifikaci nových isoforem v ječmeni u genů podléhajících alternativnímu sestřihu a je rozdělena na teoretickou a praktickou část. Teoretická část práce je zaměřena na ječmen setý, abiotický a biotický stres rostlin a zejména na popis prekurzorové mRNA a její následnou úpravu alternativním sestřihem.

Praktická část práce je rozdělena do několika podkapitol. První podkapitola se zabývá kontrolou kvality dat poskytnutých v rámci celotranskriptomového sekvenování pomocí k tomuto účelu určenému bioinformatickému nástroji. Druhá a třetí podkapitola praktické části obsahuje popis a zhodnocení procesu optimalizace mapování „readů“ získaných v rámci procesu sekvenování na referenční genom ječmene. Kapitola rovněž dokumentuje následnou kvantifikaci namapovaných „readů“ na genomové úrovni. Poslední část praktické části se zabývá kvalitativní a kvantitativní analýzou alternativního sestřihu pre-mRNA.

2 SOUČASNÝ STAV ŘEŠENÉ PROBLEMATIKY

2.1 Ječmen setý

Rodina jednoděložných rostlin čeledi lipnicovitých zahrnuje velké množství zemědělsky významných rostlin, jako je kukuřice (*Zea mays L.*), pšenice (*Triticum aestivum L.*) nebo rýže (*Oryza sativa L.*). Členem této rodiny je také ječmen setý (*Hordeum vulgare*) (Obr. 1). Ječmen je jednoletá rostlina jarního a ozimého charakteru, jejíž výška se pohybuje v rozmezí 0,8–1,2 m. Je možné jej pěstovat na hnědých, černozemích a lužních půdách s optimálním pH v rozmezí 6,2 – 7,2 (Close *et al.,* 2004). Podle FAO statistics (FAOSTAT: http://www.fao.org/faostat/en/#data/QC) je ječmen čtvrtou nejdůležitější obilovinou, jehož celosvětová produkce přesahuje 130 milionů tun ročně. Podle informací Zemědělského svazu České republiky o postupu sklizně obilovin ke dni 28. 8. 2017 je sklizeno 97 157 ha ječmene ozimého, jeho produkce činí 562,8 tis. tun a průměrný výnos odpovídá 5,79 t/ha. Jarního ječmene je sklizeno 228 100 ha, jeho produkce odpovídá 1146,9 tis. tun a průměrný výnos je 5,03 t/ha (Zemědělský svaz ČR: https://www.zscr.cz/clanek/postup-sklizne-obilovin-a-repky-k-7-8-2017-3178).

Využitím metod genetických modifikací lze vytvořit kultivary ječmene tolerantní k chladu, zasolení, suchu a alkalické půdě. Geneticky modifikovaný ječmen je možné kultivovat v různých prostředích, od oblasti Arktického kruhu až k uměle zavlažovaným polím subsaharské Afriky (Mayer *et al.*, 2011, 2012).

Obr 1. Ječmen setý (z KWS: https://www.kws-uk.com/aw/Company/Submenu-1-Topic-1/KWS-Infinity-Lifts-Two-Row-Yields/~fyuf/; 13. 2. 2017)

Ječmen setý je vysoce přizpůsobivá obilnina, je důležitým zdrojem potravin v mnoha částech světa a je považován za nejvíce tolerantní obilovinu vůči stresu suchem a zasolením (Close *et al.,* 2004). Ječmen představuje rovněž klíčovou složku výroby piva a whisky (Ogle *et al.,* 2006) a v Evropě je nejvíce obdělávanou obilovinou, která se používá na výrobu krmiv pro zvířata (Todorov 1988). Za posledních několik let byla zrna ječmene úspěšně použita ve farmaceutickém průmyslu jako důležitý bioreaktor přizpůsobený k výrobě terapeutických proteinů. V potravinářském průmyslu se z ječmene vyrábí kroupy, kávová náhražka a slad (Mayer *et al.,* 2012). Ječmen je také považován za modelový organismus rodiny lipnicovitých, v důsledku nedávného zlepšení účinnosti transformace, zkrácení doby potřebné pro přípravu stabilních transgenních linií, dostupnosti rozsáhlých zdrojů zárodečné plazmy a mutantních forem ječmene. Výhodou ječmene jako modelového organismu je rovněž diploidní genom a poměrně nízká složitost genomu (Mayer *et al.,* 2011, 2012).

Genom ječmene má diploidní počet chromozomů (2n = 2x = 14) (Wicker *et al.,* 2008) s 10 až 20 % tvořenými tandemově uspořádanými repetitivními sekvencemi neboli tandemovými repeticemi DNA, což jsou opakující se motivy jednoho či více nukleotidů (Gray *et al.,* 1991). S velikostí haploidního genomu přibližně 5,3 Gbp v 7 chromozomech se jedná o jeden z největších diploidních genomů. Aktuální počet genů ječmene je podle databáze Ensembl Plants 39 809 (Ensembl Plants: http://plants.ensembl.org/Hordeum_vulgare/Info/Annotation/).

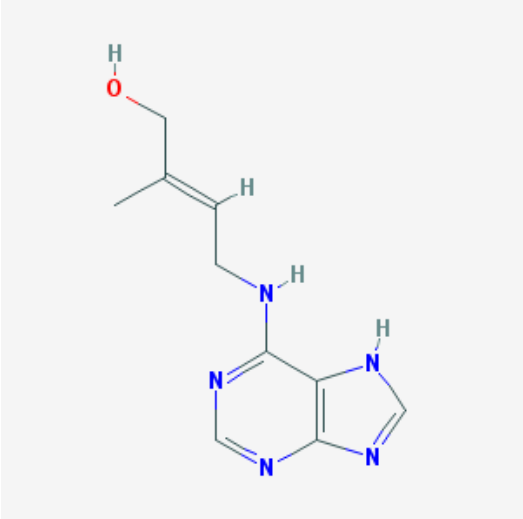
Ke studiu ječmene lze využít mnoho analytických metod. Jednou z nich je RNA sekvenování (RNA-seq), které umožňuje ve spojení s následnou bioinformatickou analýzou kromě kvantifikace genové exprese, také studium strukturálních variant populace RNA, role alternativního sestřihu a průzkum dosud necharakterizovaných transkripčně aktivních oblastí genomu ječmene. Zarovnání RNA-seq dat na predikované geny ječmene s vysokou jistotou ukázalo, že více než 70 % transkriptů může být modifikováno alternativním sestřihem (Close *et al.,* 2009; Haseneyer *et al.,* 2011).

2.2 Cytokininy

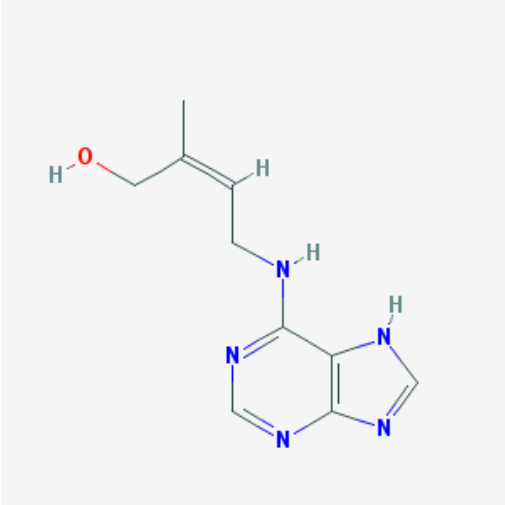
Cytokininy přestavují skupinu rostlinných hormonů, které se podílejí na regulaci morfogeneze rostlin, jako je tvorba výhonků a kořenů nebo apikální dominance. Při koordinovaném působení s jinými hormony, zejména auxiny, podporují buněčné dělení (Wernet et al., 2009 a Spíchal et al., 2012). V biotechnologických aplikacích

jsou cytokininy ceněny za svou schopnost zprostředkovat obranné reakce vůči environmentálním stresům, jako je slanost nebo sucho (Macková et al., 2013). Chemicky jsou cytokininy deriváty adeninu substituované v poloze N6 isoprenoidním nebo aromatickým postranním řetězcem.

Nejrozšířenějšími přirozeně se vyskytujícími rostlinnými cytokininy jsou isopentenyladenin (iP) a jeho hydroxylované deriváty *trans*-zeatin (tZ) (Obr. 2) a *cis*-zeatin (cZ) (Obr. 3).



Obr 2. *Trans*-zeatin (z The PubChem Project: https://pubchem.ncbi.nlm.nih.gov/compound/Zeatin#section=2D-Structure,15. 5. 2017)



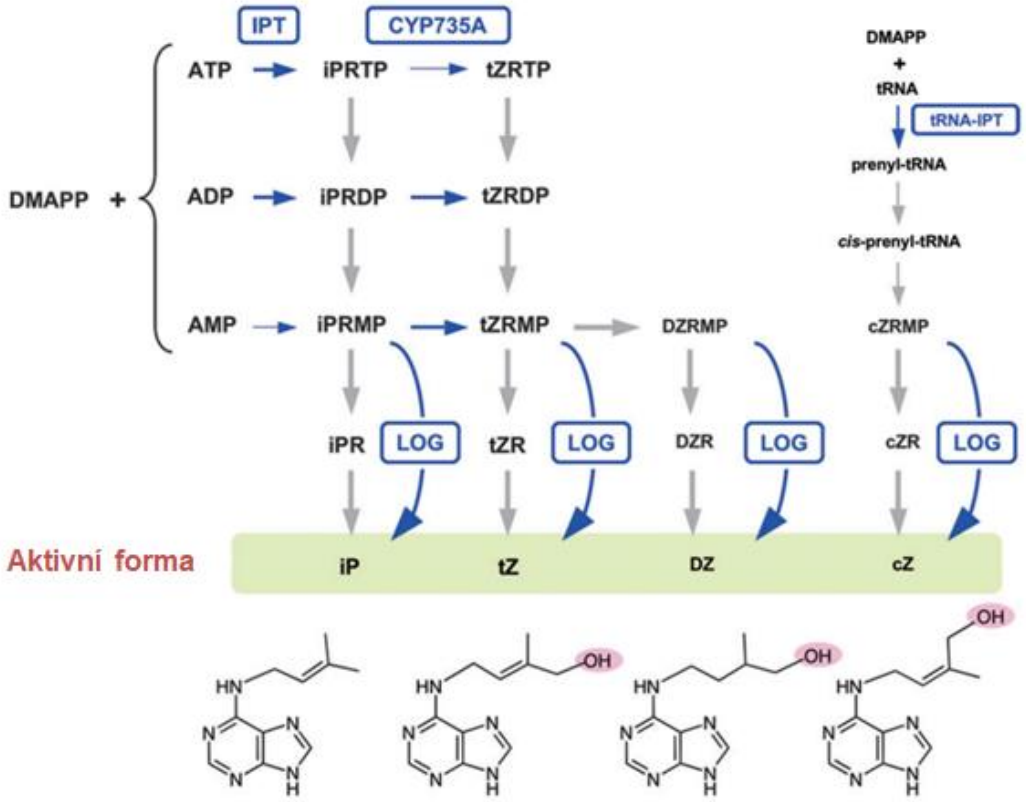
Obr 3. *Cis*-zeatin (z The PubChem Project: https://pubchem.ncbi.nlm.nih.gov/compound/cis-zeatin#section=2D-Structure, 15. 5. 2017)

Aromatické cytokininy, jako jsou benzyladeninové deriváty a kinetin, se používají při rozmnožování rostlin a při různých biotechnologických aplikacích kvůli jejich vyšší stabilitě proti degradaci endogenními enzymy rostlin. Přírodní cytokininy (Mok et al., 2001) se vyskytují ve čtyřech hlavních formách: (1) nukleotidové, které jsou produkovány během *de novo* biosyntézy a poté konvertovány na jiné deriváty; (2) volné báze, které mají nejvyšší aktivitu (Romanov et al., 2006); (3) ribosidy, které představují transportní formu (Hirose et al., 2008) a (4) glukosidy, které jsou skladovacími/inaktivovanými formami (Moore et al., 1989). Strukturní modifikace cytokininů přímo ovlivňují jejich biologickou aktivitu a vazebnou schopnost na receptory (Spíchal et al., 2004).

Přímé zapojení různých forem cytokininů do důležitých fyziologických procesů, jako je zrání semen, reakce na stres a dobrá znalost jejich metabolismu, z nich činí vhodný cíl pro umělou manipulaci s rostlinným fenotypem (Wernet et al., 2001). Nadměrná exprese genu kódujícího cytokinin dehydrogenázu (CKX, EC 1.5.99.12) v tabáku, u *Arabidopsis thaliana* a v ječmenu, způsobuje zvětšení kořenového systému, ale bohužel také významnou redukci nadzemních částí rostlin a snížení nebo ztrátu plodnosti (Mrízová et al., 2013). Přestože rozšířený kořenový systém dává rostlinám výhodu ve spotřebě vody a živin, dobře rozvinuté horní části jsou z agronomického hlediska rovněž důležité. Modifikací degradace cytokininu zvýšenou expresí genu CKX dochází ke zvýšení kořenové biomasy bez ovlivnění vývoje a funkce nadzemních částí (Wernet et al., 2010).

Počátečním krokem v biosyntéze cytokininů (Obr. 4) je konverze dimethylallyldifosfátu (DMAPP) pomocí adenosin-fosfát-isopentenyltransferázy (IPT). Ve vyšších rostlinách je hlavním produktem této reakce iP nukleotid, jako iP ribosid 5'-trifosfát (iPRTP) nebo iP ribosid 5'-difosfát (iPRDP), protože IPT jako substrát převážně používá DMAPP a ATP nebo ADP (Kakimoto, 2001; Sakakibara *et al.,* 2005). V *Arabidopsis* se nukleotidy iP konvertují na nukleotidy cytochromem P450 mono-oxygenázou CYP735A1 a CYP735A2 (Takei *et al.,* 2004b). Aby se staly biologicky aktivní, nukleotidy iP a tZ se převádějí na formy volných bází defosforylací a deribosylací, ale geny kódující nukleotidázu (Chen a Kristopeit, 1981a) a nukleosidázu (Chen a Kristopeit, 1981b) dosud nebyly identifikovány. Biosyntéza cytokininů může rovněž probíhat alternativní biosyntetickou cestou, která přímo uvolňuje aktivní cytokinin z nukleotidu. U této biosyntézy se kromě enzymu tRNA-IPT

účastní také enzym cytokinin nukleosid 5'-monofosfát fosforibohydroláza (LOG; Kurakawa *et al.,* 2007).



Obr 4. Model biosyntézy cytokininů. DMAPP – dimethylallylpyrofosfát, IPT – isopentenyltransferáza, iPRTP – iP ribosid 5'-trifosfát, iPRDP – iP ribosid 5'-difosfát, iPRMP – iP ribosid 5'-monofosfát, tZRTP – tZ ribosid 5'-trifosfát, tZRDP – tZ ribosid 5'-difosfát, tZRMP – tZ ribosid 5'-monofosfát, DZRMP – DZ ribosid 5'-monofosfát, cZRMP – cZ ribosid 5'-monofosfát, iPR – iP ribosid, tZR – tZ ribosid, DZR – DZ ribosid, cZR – cZ ribosid. (Převzato a upraveno z Hirose *et al.,* 2007)

2.3 Biotický a abiotický stres

Stres je souhrnné označení pro stav, ve kterém se rostlina nachází při působení stresových faktorů. Stresové faktory se dělí na biotické a abiotické (Cramer *et al.,* 1990). Na působení stresových faktorů odpovídají rostliny aktivací obranných mechanismů (Chinnusamy *et al.,* 2007).

Jedním z nejdůležitějších abiotických stresů, které ovlivňují rostliny je stres způsobený suchem. Tento stres rostliny zažívají v okamžiku, kdy dochází k omezení přívodu vody do kořenů, nebo pokud stoupne intenzita transpirace. V případě vysoké půdní slanosti a také v jiných podmínkách, jako je zaplavení a nízká teplota půdy, sice

existuje voda v půdním roztoku, kterou však rostliny nemohou přijmout. Tento stav je známý jako tzv. fyziologické sucho (Ha *et al.,* 2014). Schopnost rostlin odolávat tomuto stresu má nesmírný ekonomický význam. Znalost biochemických a molekulárních reakcí na sucho je zásadní pro komplexní vnímání mechanismu rezistence rostlin vůči stresovým podmínkám (Srivastava *et al.,* 2012). Stres suchem v rostlinách snižuje vodní potenciál a turgor rostlinných buněk, což je poté následováno akumulací kyseliny abscisové (ABA) a kompatibilních osmolytů, látek, které jsou schopny regulovat proces proudění vody, jako je například prolin. Dále dochází ke zhoršení stavu nadprodukcí reaktivních druhů kyslíku a tvorbou askorbátu a glutathionu (Lisar *et al.,* 2012). Sucho rovněž ovlivňuje stomatální uzavření, omezuje výměnu plynů, snižuje transpiraci a zastavuje nebo zpomaluje asimilaci uhlíku (fotosyntézu). Negativní účinky sucha na minerální výživu (příjem a transport živin) a metabolismus rostliny, vedou k poklesu listové plochy a ke změnám rozdělení asimilace mezi orgány nebo změně elasticity stěn rostlinných buněk a k narušení homeostázy a distribuce iontů v buňce (Alvarez *et al.,* 2008). Dalším důsledkem je syntéza nových proteinů a mRNA, spojených s reakcí na sucho (Chinnusamy *et al.,* 2007).

Dalším abiotickým stresem je stres teplotní. Vysoká teplota způsobuje rozmanité a často nepříznivé změny v růstu rostliny, v jejím vývoji, ve fyziologických procesech a produktivitě (Hasanuzzaman *et al.,* 2012). Jedním z hlavních důsledků stresu vysokou teplotou je přebytek reaktivních forem kyslíku, což vede k oxidačnímu stresu (Hasanuzzaman *et al.,* 2011). Rostlina je schopna do určité míry tolerovat přehřívání fyzikálními změnami uvnitř rostlinného těla a často také vytvořením signálů pro změnu metabolismu (Valliyodan *et al.,* 2006, Munns *et al.,* 2008). Tepelný stres na molekulární úrovni, způsobuje změny v expresi genů podílejících se na přímé ochraně před stresem vysokou teplotou (Chinnusamy *et al.,* 2007). Tyto geny jsou zodpovědné za expresi detoxikačních enzymů, transportérů a regulačních proteinů (Krasensky *et al.,* 2012). V podmínkách vysoké teploty a ovlivnění genové exprese specifických genů, postupně dochází k vytvoření tolerance na teplo ve formě aklimatizace nebo v ideálním případě přizpůsobení (Deluc *et al.,* 2009).

Teplotní stres může být způsobem i vlivem teploty nízké, která je důležitým faktorem, který omezuje rozšíření kultivačních ploch a tím i produktivitu pěstování několika rostlin (Chinnusamy *et al.,* 2007). Při nízkých teplotách dochází k zásadnímu

poškozování buněčných membrán hlavně v důsledku dehydratace způsobené tvorbou extracelulárního ledu (Steponkus 1984; Steponkus *et al.*, 1993).

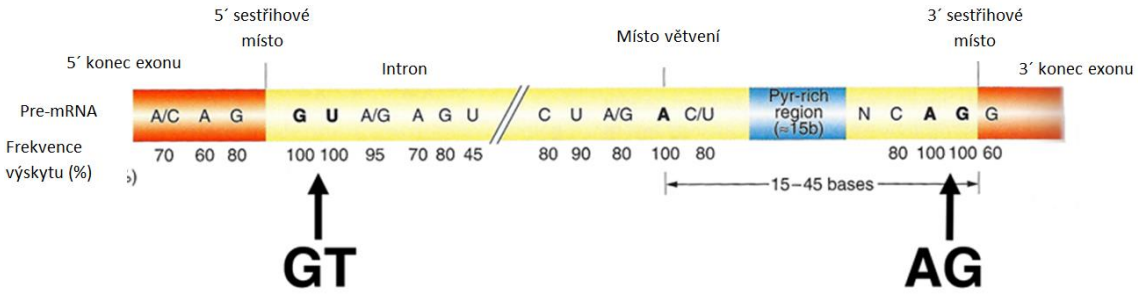
Stres rostlin může být způsoben i zasolením půdy. Salinita je jedním z nejdůležitějších environmentálních parametrů, které určují úspěch nebo selhání při pěstování rostlin (Cramer *et al.*, 1990). Nejvíce rizikovým vývojovým stupněm při stresu zasolením je klíčení. Nárůst koncentrace chloridu sodného snižuje počet kultivarů, délku klasu, počet zrn na klasu a výnos zrna na rostlinu (Ahmad *et al.*, 2003).

Poškození rostlin může být rovněž vyvoláno stresem biotickým, a to pomocí hmyzu, plísní, bakterií a virů. Mikroorganismy způsobují vadnutí rostlin, skvrny na listech, hnilobu kořenů nebo poškození semen a hmyz naopak může za fyzické poškození rostlin, včetně listů, kůry a květů (Christensen *et al.*, 2004).

K opatření vůči abiotickým, ale i biotickým stresům je možné použít zejména agrotechnické metody, které jsou využitelné například u nedostatku půdního vzduchu, na který je ječmen vysoce citlivý či v případě chorob přenosných osivem, například prašná sněť a hnědá skvrnitost (Saritha *et al.*, 2007).

2.4 Pre-mRNA a metody sestřihu

Geny kódující proteiny, jsou během procesu transkripce přepisovány z DNA do prekurzorové mRNA, jinak také pre-mRNA (Obr. 5), která je tvořena exony (sekvencemi kódující protein) a nekódujícími introny. Sekvence intronů jsou během procesu zvaného sestřih odstraněny a exony jsou opět spojeny ligací. K samotnému sestřihu dochází ve specifických sestřihových místech prekurzorové mRNA (Yitzhaki *et al.*, 1996).



Obr 5. *Pre-mRNA* (převzato a upraveno z BIOL 202 Genetics http://www.discoveryandinnovation.com/BIOL202/notes/lecture12.html, 14. 3. 2017)

Intron je ohraničen 5‘ sestřihovým místem a 3‘ sestřihovým místem. Mezi 5‘ a 3‘ sestřihovým místem, ve vzdálenosti 20-50 nukleotidů od 3‘ sestřihového místa, se nachází místo větvení. Toto místo obsahuje adenosin, nukleosid důležitý pro první krok sestřihu (Burge *et al.,* 1999). Je známou skutečností, že téměř všechna místa sestřihu jsou v souladu s konsenzuální sekvencí, na jejímž základě k sestřihu dochází. Tyto konsenzuální sekvence zahrnují téměř neměnné dinukleotidy na každém konci intronu, GT na 5 'konci intronu a AG na 3' konci intronu.

Sestřih pre-mRNA lze rozdělit na konstitutivní sestřih a alternativní sestřih. Výsledkem konstitutivního sestřihu je molekula mRNA poskytující vždy stejnou primární strukturu.

Alternativní sestřih tvoří centrální buněčný proces, během kterého je z jediného genu (prekurzorové mRNA) produkováno několik různých izoforem mRNA (Lander *et al.,* 2001) a ty jsou následně přeloženy do proteinů. Tento pozoruhodný jev tak vede ke vzniku proteinových izoforem s různými funkcemi, což značně rozšiřuje proteomickou rozmanitost u vyšších eukaryot (Ben-Dov *et al.,* 2008). Fenomén alternativního sestřihu byl poprvé objeven v roce 1978, a poté experimentálně ověřen v roce 1987. Dříve byl považován za poměrně neobvyklou formu regulace genů (Romero *et al.,* 2006). Alternativní sestřih je běžně se vyskytující jev eukaryotních buněk, avšak více se vyskytuje u eukaryot vyšších. Nahromaděné Expressed Sequence Tags (EST), mRNA data sety a data z celogenomových studii alternativního sestřihu prokázaly, že až 95% lidských genů, 60% genů *Drosophily melanogaster* a 61% introny obsahujících rostlinných genů podléhá alternativnímu sestřihu (Kampa *et al.,* 2004).

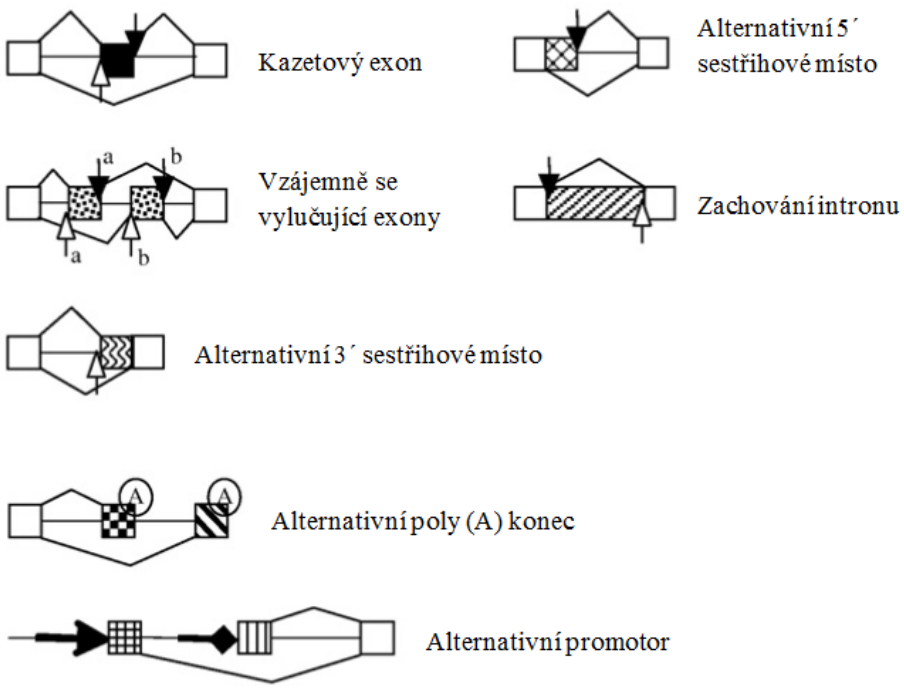
Kvalitativní a kvantitativní identifikace vzniklých izoforem je nezbytná pro pochopení různých rolí alternativně sestřižených genů v buňce (Stamm *et al.,* 2005) mezi které lze zařadit regulaci genové exprese v reakci na environmentální podněty a vývojové změny. Některé izoformy vzniklé alternativním sestřihem mohou být nefunkční nebo rychle degradovány a poskytují buňce další mechanismy k regulaci genové exprese po transkripci a před translací (Keren *et al.,* 2010). Alternativní sestřih se rovněž může uplatňovat v různých tkáních, kde mohou vznikat různé izoformy jednoho genu.

Alternativní sestřih má několik variant. Patří mezi ně kazetový exon (cassette exon nebo také exon skipping), vzájemně se vylučující exony (mutually exclusive exons), zachování intronu (retained intron), alternativní 5´ sestřihové místo (alternative 5´ splice sites), alternativní 3´ sestřihové místo (alternative 3´ splice sites), alternativní

promotorová oblast (alternative promoters) a alternativní poly-A místa (alternative poly-A sites).

Nejběžnější a nejznámější je kazetový exon (Obr. 6). Jedná se o sestřihový mechanismus, ve kterém jsou exony ponechány či odstraněny z konečného transkriptu, což vede ke vzniku prodloužené nebo zkrácené varianty mRNA (Chen, 2011).

Vzájemně se vylučující exony (Obr. 6) představují vzácný podtyp alternativního sestřihu. Tento podtyp se vyznačuje koordinovaným sestřihem exonů. Jak vyplývá z názvu, tak jeden ze dvou exonů (či jedna skupina exonů ze dvou exonových skupin) je zachován, zatímco druhý je vystřižen. V případě menších změn výsledné proteinové sekvence, mohou vzájemně se vylučující exony poskytnout výhodu mnoha typům proteinů, jako například iontovým kanálkům tím, že je zachována jejich prostorová struktura, ale dochází ke změně funkce výsledného proteinu (Sorek *et al.,* 2004b).



Obr 6. Varianty alternativního sestřihu (převzato a upraveno z Tazi *et al.,* 2009)

Zachování intronu (Obr. 6) byla po dlouhou dobu nejméně prostudovanou variantou alternativního sestřihu, neboť se předpokládalo, že tento typ je do značné míry odvozen od nesestřižených či částečně sestřižených pre-mRNA. Zachování intronu přispívá k molekulární rozmanitosti savčích buněk. Tyto introny mohou být rovněž

substrátem pro cytoplazmatický sestřih, a také slouží jako vazebné místo pro spliceosom s cytoplasmatickou aktivitou (Wang *et al.,* 2006).

Alternativní 5‘ a 3‘ sestřihové místo (Obr. 6) představují významný podíl všech variant alternativního sestřihu. V tomto případě sestřihu, dvě či více alternativních 5‘ sestřihových míst soupeří o připojení ke dvěma či více alternativním 3‘ sestřihovým místům (Hastings *et al.,* 2001).

Kromě těchto základních variant alternativního sestřihu, existují další dva mechanismy, díky kterým mohou být získány různé typy mRNA z téhož genu. Jsou jimi alternativní promotory a alternativní poly-A místa (Obr. 6). Spuštěním transkripce na různých místech, mohou být generovány transkripty s odlišnými 5‘ koncovými exony. Na druhé straně, alternativní poly-A místa poskytují transkriptu různé 3‘ koncové části. Oba tyto mechanismy poskytují další možnosti pro zvýšení rozmanitosti výsledné mRNA odvozené z jednoho genu (Kimura *et al.,* 2006).

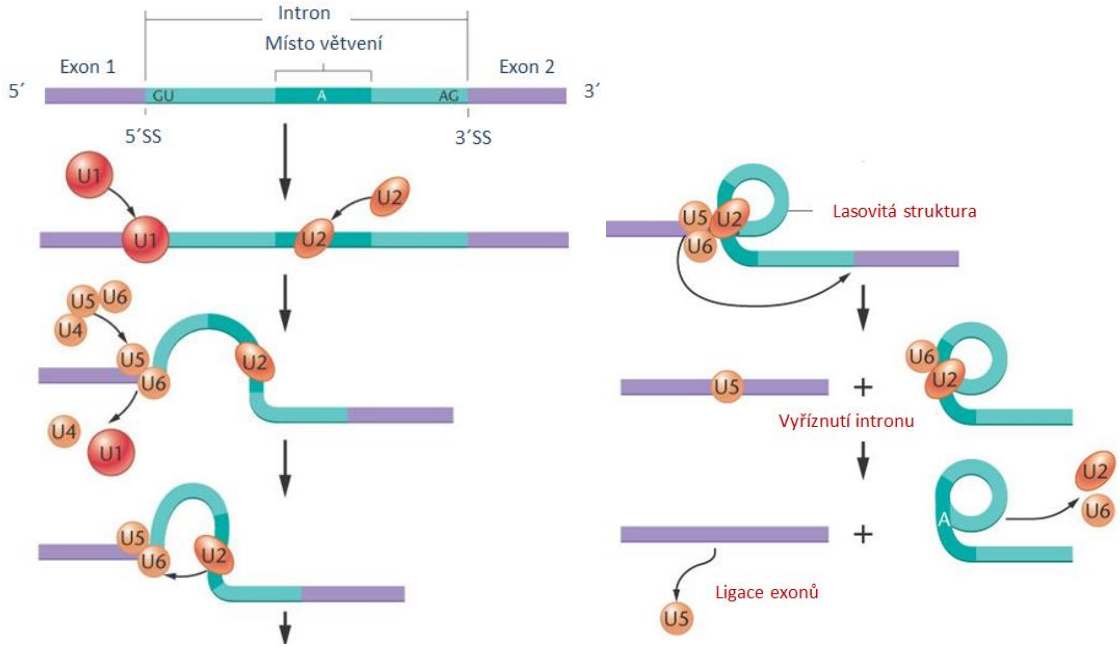
Ze všech sestřihových variant je obecně nejtěžší rozpoznat zachování intronu, neboť je obtížné jej odlišit od experimentálních artefaktů. Například neúplně sestřižené transkripty mohou obsahovat fragmenty intronu, které pak mohou být považovány za variantu zachování intronu. Mnoho genů má více variant alternativního sestřihu se složitými kombinacemi exonů, produkujících různorodé izoformy transkriptů. Například gen Dscam v *Drosophila melanogaster,* může potenciálně produkovat 38 016 odlišných variant mRNA pomocí různých kombinací 95 exonů (Graveley *et al.,* 2004). Živočichové a rostliny se liší v nejběžnějších variantách alternativního sestřihu. Kazetový exon je nejběžnější variantou u člověka (> 40%), ale nejméně obvyklou variantou u rostlin (5%). U rostlin se nejčastěji vyskytuje varianta zachování intronu (~ 40%). Tento rozdíl naznačuje, že rostliny a zvířata mohou rozpoznávat exony a introny odlišným způsobem. Některé důkazy také ukazují, že rostliny a zvířata regulují alternativní sestřih různě (Keren *et al.,* 2010).

Jednou z nejčastějších možností realizace sestřihu pre-mRNA je sestřih pomocí spliceozomu což je komplex struktur RNA a proteinů (Obr. 7). Spliceozomy obsahují malé molekuly RNA označované jako snRNA a dalších přibližně 40 proteinů. Sestřihu pre-mRNA se jako složky spliceozomu účastní pět snRNA které lze označit jako U1, U2, U4, U5 a U6 (Lander *et al.,* 2001). Spliceozom plní dvě základní funkce: rozpoznání vazby mezi exony a introny a následnou katalýzu reakce, během které se odstraňují introny a exony jsou spojovány. (Villa *et al.,* 2002). Sestavení spliceosomu je

velmi dynamický proces, ve kterém dochází ke komplexním interakcím RNA-RNA a RNA-protein. Katalytickou podjednotkou spliceozomu je U6.

V prvním kroku se podjednotka U1 váže na 5‘ sestřihové místo a podjednotka U2 na místo větvení přes RNA-RNA interakci mezi snRNA a pre-mRNA (Obr. 7). Následuje vazba zbylých podjednotek spliceozomu, jmenovitě U4, U5 a U6. Vazba mezi podjednotkami U4 a U6 se naruší vlivem interakce mezi podjednotkami U6 a U2 a dojde ke štěpení 5‘ sestřihového místa pre-mRNA. Následně se 5‘ konec intronu spojuje s adeninem v místě větvení a vytváří lasovitou strukturu s následným uvolněním podjednotek U1 a U4. Následuje štěpení 3‘ sestřihového místa a 5‘ konec druhého exonu se připojuje k 3‘ konci exonu jedna.

V průběhu tohoto kroku rovněž dochází k uvolnění intronu ve tvaru lasa společně se zbylými podjednotkami spliceozomu (Sun a Chasin, 2000). Obě sestřihové reakce (5‘ sestřihová reakce i 3‘ sestřihová reakce) vyžadují energii ve formě ATP (Lander *et al.*, 2001).



Obr 7. Schéma sestřihu genů (převzato a upraveno z Biology 3400 Genetics: http://bio3400.nicerweb.com/Locked/media/ch13/spliceosome.html, 12. 3. 2017)

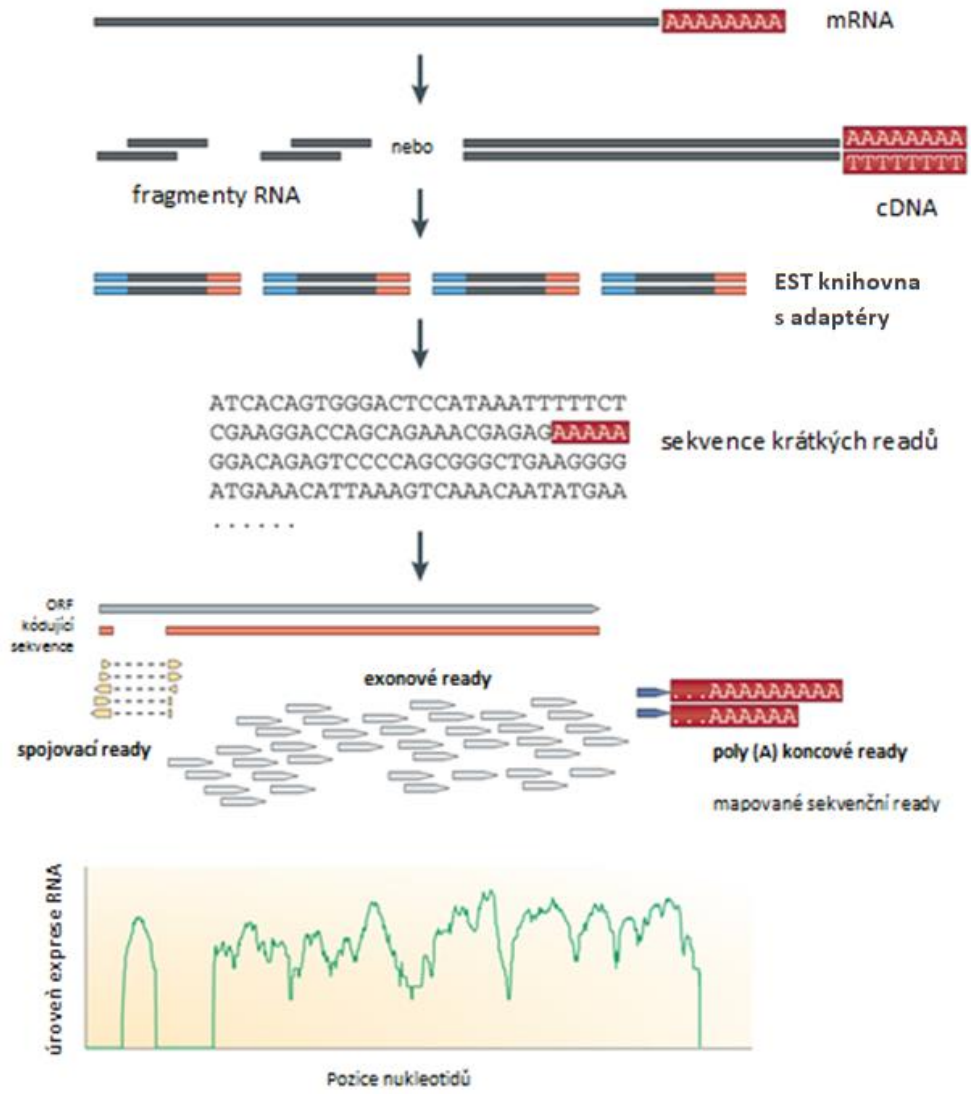
K rozpoznávání exonů spliceozomem slouží pomocné cis-elementy známé jako exonové a intronové sestřihové enhancery (ESE a ISES) a exonové a intronové sestřihové silencery (ESS a ISS) (Lim a Burge, 2001).

2.5 Metody studia alternativního sestřihu

Díky velkému množství genomových sekvencí a technikám s vysokou propustností, je možné alternativní sestřih studovat na genomové úrovni (Orengo *et al.,* 2007). Významným úkolem je také detekce specifických izoforem, vznikajících neobvyklým sestřihem, u kterých je známo, že jsou zodpovědné za nemoci spojené s různými typy rakoviny u člověka. Současné metody detekují alternativní sestřih primárně posuzováním sekvenčních „readů" (krátkých sekvencí produkovaných při RNA sekvenování), zmapovaných do unikátních jednotlivých izoforem nebo sestavením transkriptů a odhadu nejpravděpodobnější izoformy dle namapovaných sekvenčních „readů" (Wang *et al.,* 2008). V současné době, výzkumné metody alternativního sestřihu existují společně s metodami sekvenování nové generace, jinak známé jako Next-Generation Sequencing (Griffith *et al.,* 2010).

Ke studiu alternativního sestřihu se tedy využívají data získaná RNA sekvenováním (RNA-Seq) transkriptomu. RNA-seq poskytuje vysoce citlivý a přesný nástroj pro měření exprese genů a jejich isoforem (Wang *et al.* 2009, Garber *et al.* 2011) s výhodami jako je především široký dynamický rozsah, citlivé a přesné měření genové exprese. Metoda může být aplikována na jakýkoliv druh, a to i za předpokladu, že není k dispozici referenční sekvence genomu pro daný organismus (Shen *et al.,* 2014).

Typický RNA-Seq experiment (Obr. 8) probíhá tak, že molekuly mRNA extrahované z biologického vzorku jsou nejprve převedeny na knihovnu fragmentů cDNA, s následnou fragmentací mRNA a přepisem do cDNA. Sekvenční adaptéry, což jsou krátké oligonukleotidy, poskytují vazebné místo pro primery při amplifikaci a sekvenování. Tyto oligonukleotidy se následně přidávají ke každému fragmentu cDNA a z každé cDNA se získá krátká sekvence („read") s použitím sekvenačních technologií s vysokou propustností. Výsledné sekvence „readů" jsou mapovány na referenční genom nebo transkriptom a mohou být klasifikovány jako exonové, spojovací, nebo poly (A) koncové „ready". Tyto tři typy „readů" se používají pro generování expresního profilu pro každý gen (Wang *et al.,* 2009).

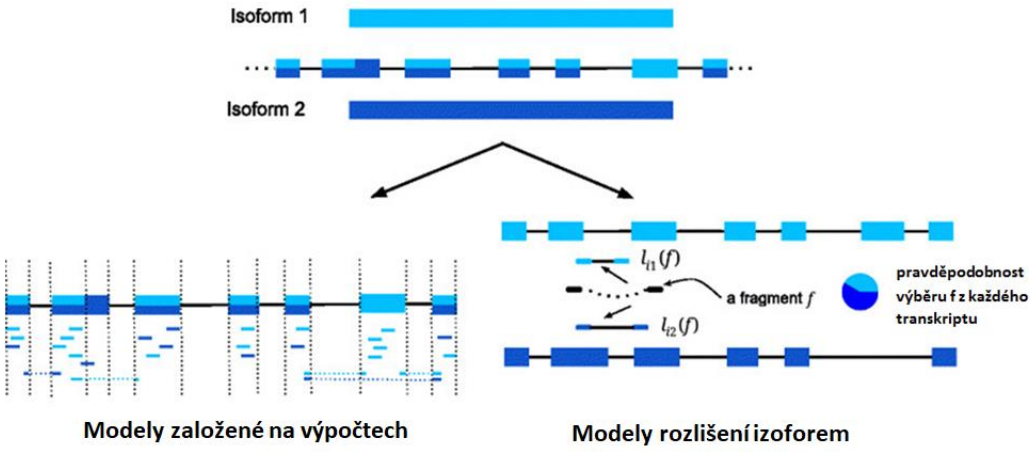


Obr 8. Typický RNA-Seq experiment (převzato a upraveno z Wang *et al.*, 2009)

„Ready“ vzniklé RNA sekvenováním se využívají pro detekci alternativního sestřihu. K tomuto účelu bylo vyvinuto několik bioinformatických softwarů, které lze rozdělit do dvou skupin na základě používané metody. První skupinou jsou modely založené na výpočtech (count-based models) a do druhé skupiny patří modely rozlišení izoforem (isoform resolution models). Obě skupiny modelů slouží ke kvantifikaci izoforem, i pro jejich kvalitativní detekci.

U modelů zakládajících se na výpočtech (Obr. 9) se „ready“ přiřazují výpočetním jednotkám (tj. exonům). Pro každou výpočetní jednotku se testují dva možné výsledky, a to zda došlo k zahrnutí nebo vystřižení jednotky. U modelů rozlišení izoforem (Obr. 9) se dva konce páru „readů“ (na obr. 9 zobrazeny jako tmavé obdélníky spojeny

pomlčkou) zarovnávají před a za 5´ sestřihové místo. $l_{i1}(f)$ je v rámci tohoto modelu délka zarovnání fragmentu f na izoformu i1 a je v případě obrázku 9 kratší než $l_{i2}(f)$. Proto je-li známa distribuce velikosti fragmentů, je možné odvodit, která izoforma pravděpodobně generuje fragment f.



Obr 9. Modely sloužící ke kvantifikaci isoforem (převzato a upraveno z Liu *et al.*, 2014)

2.5.1 Modely založené na výpočtech

Mezi programy využívající tyto metody patří například DEXSeq (Anders *et al.*, 2012), DSGseq (Wang *et al.*, 2013), SplicingCompass (Aschoff *et al.*, 2013) a SeqGSEA (Wang et al., 2014). Některé dokumenty odkazují na tento model, jako na model založený na variantách alternativního sestřihu (Alamancos *et al.*, 2013).

Modely zakládající se na výpočtech, jsou založeny na metodách používaných pro kvantifikaci transkriptů s jednou izoformou. Metoda RPKM (Reads Per Kilobase per Million mapped reads) se využívá pro normalizaci počtu „readů“ namapovaných na daný transkript. Pro rozlišení sestřihu z kvantitativního a kvalitativního hlediska, jsou tyto modely modifikovány k výpočtu „readů“ v malých výpočetních jednotkách (tj. exonech), nežli v celých oblastech transkriptu. Metody se rovněž zaměřují na diferenciální analýzu výpočetních jednotek z hlediska počtu „readů“ namapovaných na tyto výpočetní jednotky. Tato skutečnost umožňuje kvantitativní analýzu isoforem. Modely z této skupiny obvykle konfigurují každý gen do jedinečné reprezentace sestávající se z výpočetních jednotek. Výpočetní jednotky mohou být buď celé exony

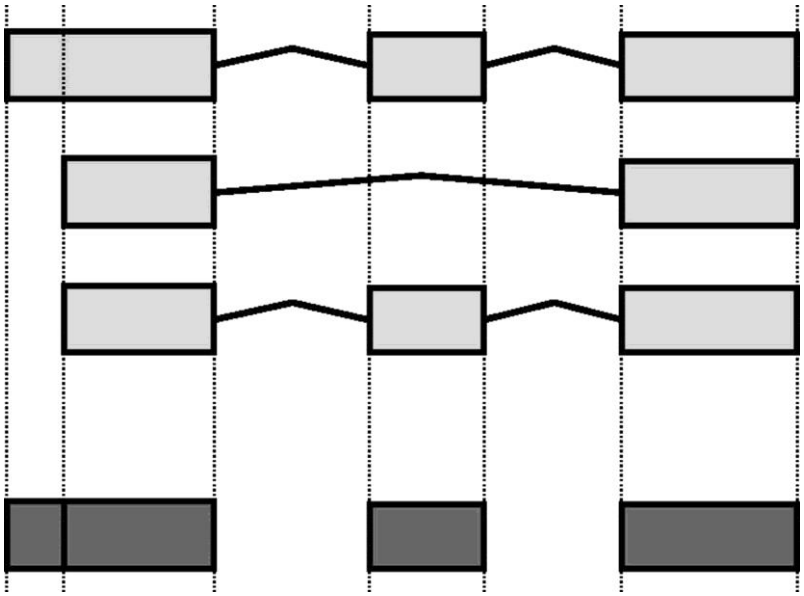
nebo zkrácené exonové oblasti (př. DEXSeq a DSGseq). Model založen na výpočtech se používá zejména pro testování dvou možných sestřihových výsledků, a to zahrnutí a/nebo vyloučení každé výpočetní jednotky do určitého sestřihového výsledku (Alamancos *et al.,* 2013).

Programy využívající modely založené na výpočtech jsou obvykle závislé na již existující anotaci poskytující strukturu genu a používají Poissonovo nebo negativně binomické rozdělení (NB) (Bullard *et al.,* 2010, Marioni *et al.,* 2008).

Programy SeqGSEA a DSGseq jsou si v mnoha věcech podobné. Známou sadu transkriptů v lokusu, oba tyto programy tzv. zplošťují do jednoho transkriptu tvořeného z výpočetních jednotek (tzv. matematické exony u DSGseq a tzv. sub-exony u SeqGSEA). Oba programy modelují počet „readů“, odpovídající výpočetním jednotkám jako NB náhodné proměnné. Modely pro daný gen počítají $\hat{p}_{ij}$ jako očekávaný počet „readů“ výpočetních jednotek i ve skupině jednotek j a odchylku hodnoty $\hat{p}_{ij}$. Obě metody definují genovou statistiku pro měření rozdílu v očekávaném počtu „readů“, a to průměrem všech výpočetních jednotek a úpravou odchylky. SeqGSEA používá přístup založený na permutaci k výpočtu p-hodnoty, zatímco DSGseq pouze uvádí statistiky a nepočítá p-hodnotu. Jak DSGseq tak SeqGSEA hlásí, který gen je alternativně sestřižen. Alternativně sestřižený gen může být předpovězen pouze tehdy, je-li zjištěno, že anotovaný konstitutivní exon je vyloučeným exonem („exon skipping“). DSGseq také určuje, kde může k vyloučení exonu skutečně dojít (Wang et al., 2014).

Stejně jako SeqGSEA a DSGseq, DEXSeq transformuje známé genové modely do setů výpočetních jednotek, pomocí všech možných sestřihových míst (Anders *et al.,* 2012). Počátečním krokem analýzy je zarovnání „readů“ na genom. Ústřední strukturou dat je tabulka, která v nejjednodušším případě obsahuje pro každý exon každého genu počet „readů“ v každém vzorku, které se překrývají s exonem. Hranice exonu však nejsou ve všech transkriptech stejné a v tom případě je exon ve dvou nebo více částech nařízнut (Obr. 10). Využívá se metoda „počítání zásobníků“ k odkazům na exony či části exonů. k_{ijl} je pro tento model počet překrývajících se „readů“ v zásobníku l genu i ve vzorku j. k_{ijl} interpretuje náhodnou proměnou K_{ijl} a počet vzorků je označen jako m ($j = 1, \ldots, m$). μ_{ijl} je očekávaná hodnota koncentrace cDNA fragmentů podílejících se na zásobníku l genu i. $E(K_{ijl})$ je očekávaný počet „readů“ a s_j je faktor velikosti, který odpovídá hloubce, kterou byl vzorek j sekvenován. Platí vztah: $E(k_{ijl}) = s_j\mu_{ijl}$. Hodnota s_j závisí pouze na j, tzn., že rozlišení v sekvenční hloubce způsobí lineární škálování počtu

„readů“. Pro modelování počtu „readů“ se používá zobecněný lineární model. Konkrétně K_{ijl} je vyjádřeno negativním binomickým rozdělením: $K_{ijl} \sim \mathrm{NB}(s_j\mu_{ijl}, \alpha_{il})$, kde α_{il} je rozptyl počítání zásobníku (i,l) (Anders *et al.,* 2012).



Obr 10. Případ variabilní délky exonu. Světle šedá – exony třech transkriptů, tmavě šedá – počítání zásobníků. Exon s variabilní délkou je rozdělen na dva zásobníky (převzato z Anders *et al.,* 2012)

SplicingCompass ve srovnání s ostatními softwary nepoužívá žádný statistický model založený na výpočetních procesech. Nejprve konstruuje vektory „readů“ na exonech a stejně tak na sestřihových spojeních, pro každý gen a vzorek. Dále využívá geometrické úhly mezi dvěma vektory. Nakonec je pro každý gen prováděn jednovýběrový t-test, který porovnává podmínky charakterizující zkonstruované vektory, pro každý gen. Na základě tohoto testu, SplicingCompass hlásí, který gen je alternativně sestřižen (Rogers *et al.,* 2012). Proto pokud se zmíněný test ukáže jako pozitivní, je nalezen nový alternativně sestřižený gen. Tímto přístupem lze ovšem detekovat pouze variantu kazetového exonu (Aschoff *et al.,* 2013).

2.5.2 Modely rozlišení izoforem

Programy Cufflinks (Trapnell *et al.,* 2010), DiffSplice (Hu *et al.,* 2013) a TopHat (Trapnell *et al.,* 2009) patří do modelů rozlišení izoforem. Metody založené na modelech rozlišení izoforem, na rozdíl od metod založených na výpočtech, usilují o přímé vyřešení tohoto problému srovnáním četností izoforem mezi vzorky a/nebo mezi podmínkami. Modely rozlišení izoforem se snaží přiřadit „ready“ nebo fragmenty

do transkriptů, z nichž pocházejí, za cenu dodatečné nejistoty při mapování „readů“, kvůli překryvu mezi izoformami. U modelů založených na výpočtech neexistuje žádná nejasnost při přiřazování „readů“ k výpočetním jednotkám (Rogers *et al.,* 2012).

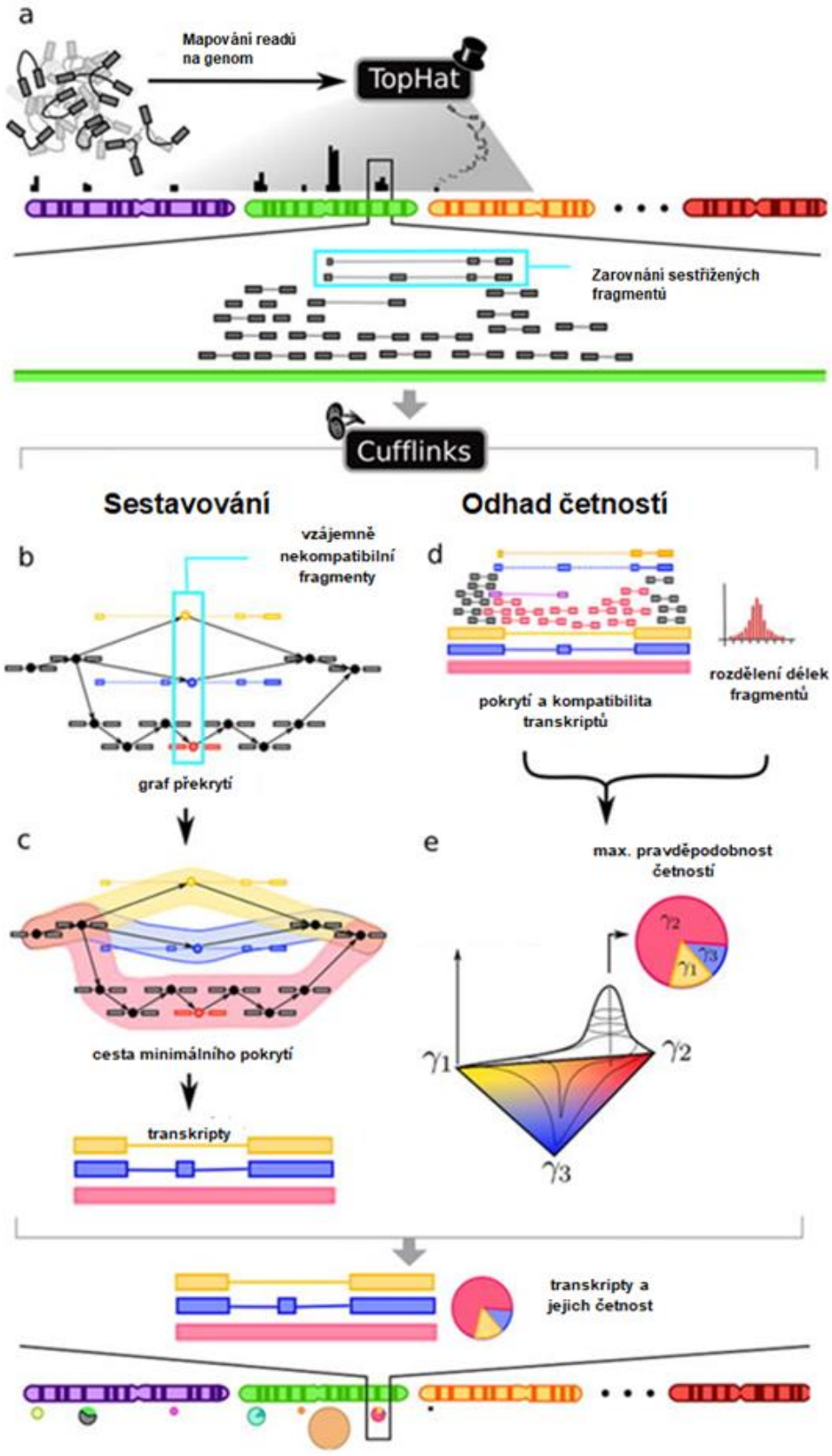
Program Cufflinks sestavuje transkripty a kvantifikuje namapované „ready“. Díky schopnosti sestavování transkriptů, je méně závislý na přesnosti genové anotace (Trapnell *et al.,* 2010).

Algoritmus programu Cufflinks, bere jako vstupní sekvence cDNA fragmenty získané zarovnáním na genom pomocí softwaru TopHat (Obr. 11a). Při použití metody „paired-end“ RNA sekvenování, jsou páry fragmentu „readů“ považovány za jedinou entitu. Algoritmus sestavuje překrývající se síť „uzlů“ zarovnaných fragmentů (na obr. 11 b-c), což snižuje dobu výpočtů a využití paměti, protože každý uzel obsahuje fragmenty z ne více než několika genů. Cufflinks poté odhaduje počet sestavených transkriptů (na obr. 11 d-e).

Prvním krokem při sestavování fragmentů (Obr. 11b) je identifikace dvojice „nekompatibilních“ fragmentů, které musí pocházet z odlišně sestřižených izoforem mRNA (Trapnell *et al.,* 2010). Fragmenty jsou spojeny v tzv. grafu překrytí, pokud jsou kompatibilní a jejich zarovnání se v genomu překrývají. Každý fragment má v grafu jeden uzel a mezi každou dvojici kompatibilních fragmentů je umístěna hrana grafu směřující zleva doprava podél genomu. Následně dojde k sestavení izoforem podle grafu překrytí (Obr. 11c). Cesty v grafu odpovídají množinám vzájemně kompatibilních fragmentů, které lze sloučit do úplných izoforem.

Dilworthova věta říká, že počet vzájemně nekompatibilních „readů“ je stejný jako minimální počet transkriptů potřebných k popisu všech fragmentů. Cufflinks představuje důkaz Dilworthovy věty, produkcí minimální množiny cest, které pokrývají všechny fragmenty v grafu překrytí tím, že najde největší množinu „readů“ s vlastnostmi takovými, že nemohou pocházet ze stejné izoformy.

Dalším krokem ve výpočetním procesu je odhad četnosti transkriptů (Obr. 11d). Odhad relativní četnosti transkriptu je uveden ve formě FPKM (fragmenty na kilobázy na milion mapovaných fragmentů), což představuje metodu normalizace ekvivalentní s RPKM v případě sekvenování single-end „readů“ (Trapnell *et al.,* 2010). Fragmenty odpovídají těm transkriptům, ze kterých mohly vzniknout. Cufflinks odhaduje četnost transkriptu za použití statistického modelu, ve kterém je pravděpodobnost pozorování každého fragmentu lineární funkcí četností transkriptů, ze kterých mohl vzniknout.



Obr 11. Princip výpočtu programem Cufflinks. $\gamma_1,\gamma_2,\gamma_3$ – četnosti izoforem (převzato a upraveno z Trapnell *et al.*, 2010).

Vzhledem k tomu, že jsou sekvenovány pouze konce každého fragmentu, jeho délka může být neznámá. Přiřazení fragmentu k různým izoformám, často znamená, že mají různou délku. Cufflinks může využít rozdělení délky fragmentů k přiřazení fragmentů k izoformám. Program poté číselně maximalizuje funkci, která přiřazuje pravděpodobnost všem možným sadám relativních výskytů izoforem (př. $\gamma_1,\gamma_2,\gamma_3$) a produkuje četnosti, které nejlépe popisují pozorované fragmenty (Trapnell *et al.,* 2010; Obr 11e).

Program DiffSplice nevyžívá model rozlišení izoforem, ale rozlišení alternativních cest. V algoritmu DiffSplice, alternativní cesty představují cesty z alternativního sestřiženého modulu (ASM) v grafech sestřihu a každý ASM má alespoň dvě alternativní cesty. ASM je oblast v sestřihových grafech, kde se izoformy navzájem liší. ASM se snaží minimalizovat nejednoznačnost v rozlišení izoforem tím, že zváží pouze oblasti, které nejsou sdíleny všemi izoformami (Hu *et al.,* 2013). DiffSplice testuje diferenciální sestřih na každém ASM místo na celých transkriptech. Relativní zastoupení alternativních cest je odhadnuto pomocí metody maximální pravděpodobnosti. Rozdíl v kompozicích relativních abundancí se měří metodou Jensen-Shannon Divergence (Hu *et al.,* 2013).

Cufflinks rozšiřují tento model na případ paired-end „readů“, zatímco DiffSplice se omezuje na ASM. Stejně jako metoda Cufflinks, je DiffSplice schopen namapovat zarovnané „ready“ na transkriptom. Proto jsou obě metody schopny rozpoznat varianty AS, které nejsou anotovány (Hu *et al.,* 2013).

Dalším programem, který využívá uvedenou metodu je TopHat. Jedná se o rychlý mapovač sestřihových spojení („splice junction“), tedy spojů mezi exony a introny. TopHat nachází spojení pomocí mapování „readů“ na referenční sekvenci ve dvou fázích. V první fázi mapuje všechny „ready“ na referenční genom pomocí Bowtie. Bowtie je ultrarychlý, paměťově výhodný zarovnávač krátkých „readů“, na dlouhé genomy. Pro dosažení vyšší rychlosti mapování je schopen současně použít několik procesorů. Výstupy Bowtie mohou být ve standartním formátu SAM, což umožňuje softwaru Bowtie spolupracovat s dalšími nástroji podporujícími formát SAM, včetně SAMtools. Všechny „ready“, které nejsou namapovány, jsou označeny jako iniciálně nemapované „ready“ (IUM). Bowtie pro každý „read“ hlásí jedno či více zarovnání, obsahující několik neshod (standardně 2) u bází na 5´ konci „readu“. Zbývající část na 3´ konci může mít dodatečné neshody za předpokladu, že Hammingova vzdálenost, která reprezentuje minimální počet substitucí potřebných pro změnu jednoho řetězce

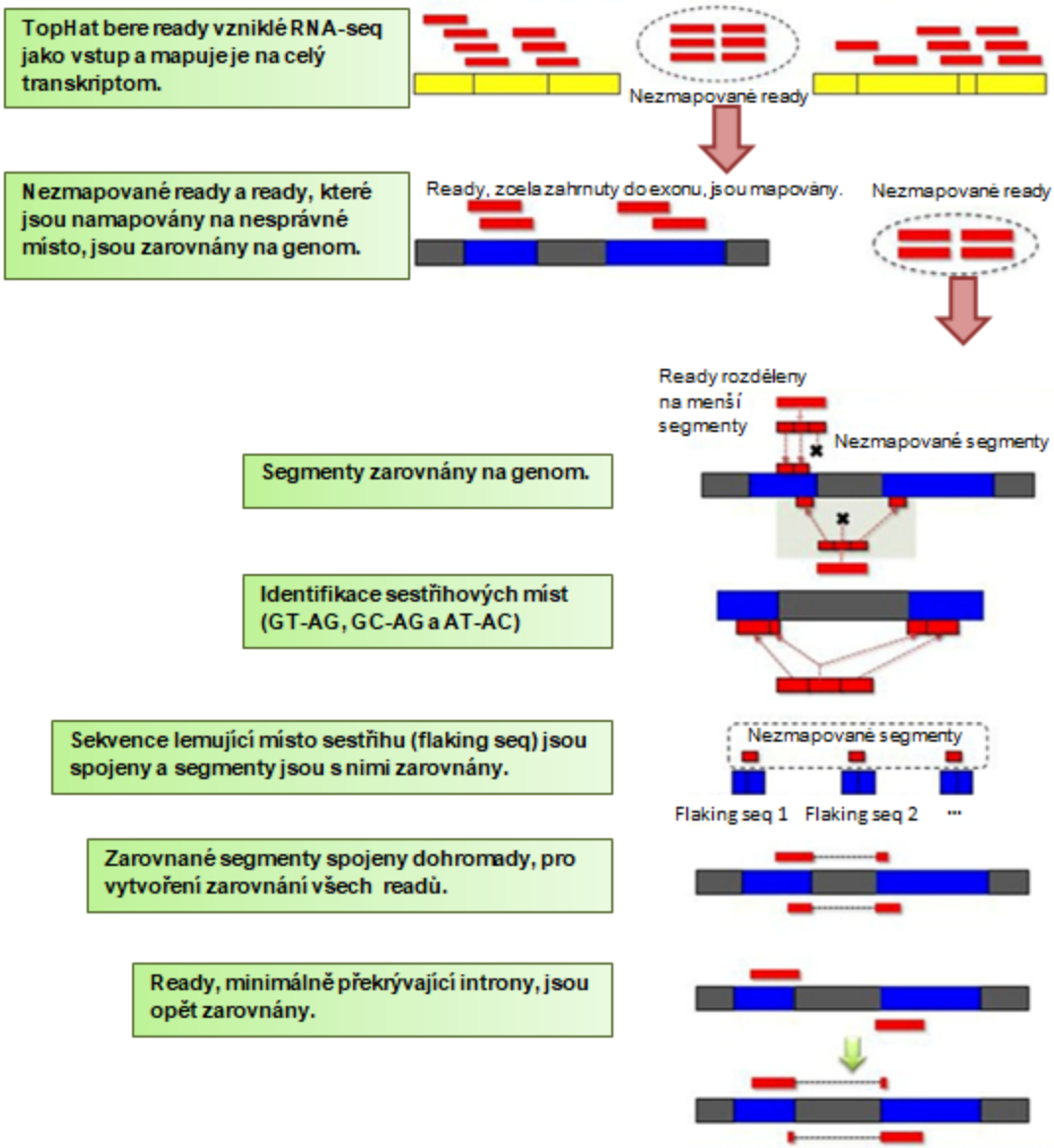
do druhého nebo minimální počet chyb, které by mohly mít, při transformaci řetězce do druhého (Cohen *et al.,* 1997), je nižší než definovaná prahová hodnota (70 ve výchozím nastavení). Tato skutečnost je založena na empirickém zjištění, že 5' konec „readu“ obsahuje méně sekvenčních chyb než 3' konec (Hillier *et al.,* 2008). TopHat umožňuje Bowtie hlásit více než jedno zarovnání pro jeden „read“ (výchozí hodnota 10) a potlačuje všechna zarovnání „readů“ mající větší počet zarovnání. Tato pojistka dovoluje hlásit tzv. vícenásobné „ready“, ale vylučuje mapování na sekvenci s nízkou složitostí (low-complexity sequence), což je oblast s neobvyklým složením nukleotidů (např. AAATAAAAAAAATAAAAAAT), ke které se chybné „ready“ často zarovnávají. „Ready“ s nízkou složitostí, které překrývají oblasti s nízkou složitostí, nejsou zahrnuty do sady IUM „readů“, ale jsou prostě vyřazeny (Li *et al.,* 2008).

TopHat (Obr. 12) nachází místa spojení bez referenční anotace. Při prvním mapování RNA-seq „readů“ na genom, TopHat identifikuje potenciální exony, poněvadž mnoho RNA-seq „readů“ bude souvisle zarovnáno na genom. Pomocí tohoto počátečního mapování vytvoří TopHat databázi možných spojení, a pak mapuje „ready“ proti těmto spojům a potvrdí je. Nástroje na sekvenování krátkých „readů“, mohou v současné době produkovat „ready“ o délce 100bp nebo delší, ale mnoho exonů je kratších než tato délka a v počátečním mapování vyvstává problém s jejich mapováním. TopHat tento problém řeší rozdělením vstupních „readů“ na menší fragmenty a ty poté samostatně mapuje. Zarovnané fragmenty ve finálním kroku tzv. vlepuje dohromady, čímž se vytvoří „paired-end“ zarovnání (Pozzoli *et al.,* 2007). TopHat generuje databázi možných spojů ze tří zdrojů. Prvním zdrojem je párování tzv. ostrovů pokrytí, což jsou odlišné oblasti nahromaděných „readů“ počátečního mapování.

Sousední regiony ostrovů pokrytí jsou často spojeny v transktriptomu často společné, takže TopHat hledá cesty k jejich spojení s intronem. Druhý zdroj se používá pouze tehdy, když je aplikace TopHat spuštěna s párovými koncovými „ready“ (vznikají sekvenováním typu paired-end). Pokud „ready“ v párech pocházejí z exonů v transkriptu, budou obecně mapovány daleko od sebe v genomovém prostoru. V této situaci se TopHat pokouší uzavřít mezeru mezi nimi, hledáním podsekvencí s celkovou délkou, která se přibližně rovná očekávané vzdálenosti mezi sousedy.

Introny v této podsekvenci jsou přidány do databáze. Třetím a nejsilnějším zdrojem důkazů o sestřihových místech je případ, kdy jsou dva segmenty ze stejného „readu“ mapovány daleko od sebe nebo pokud vnitřní segment nelze mapovat. U dlouhých

„readů“ (>=75bp), mohou být introny "GT-AG", "GC-AG" a "AT-AC" nalezeny ab initio, pomocí metod, které se pokoušejí předpovídat geny na základě statistických vlastností dané sekvence. U kratších „readů“, TopHat hlásí pouze introny "GT-AG" (Hillier *et al.*, 2008).



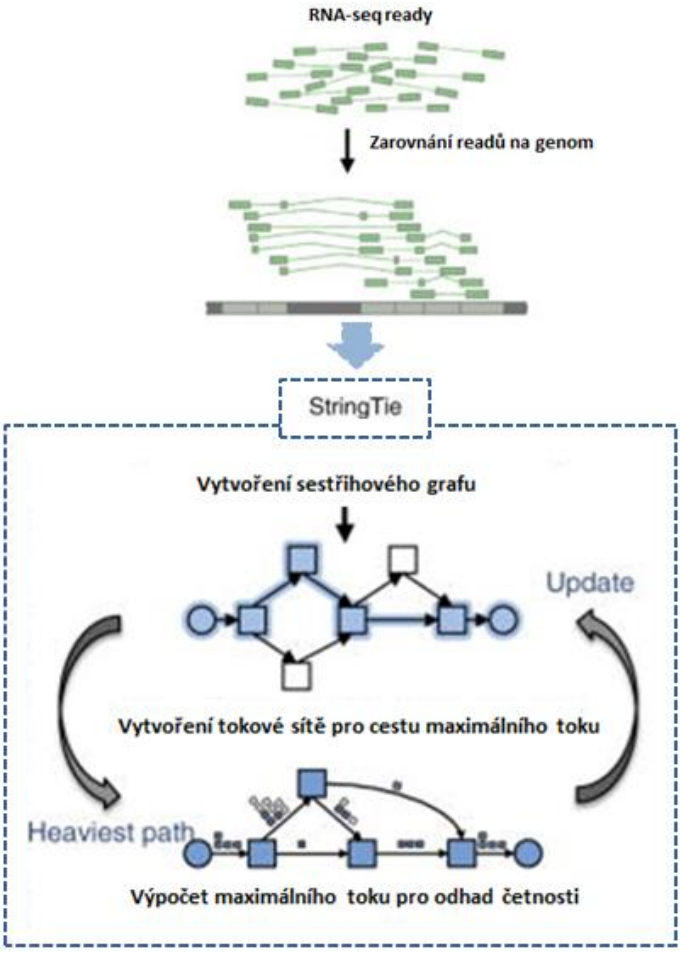
Obr 12. Princip TopHat. Červeně – „ready“, žlutě – exony z anotovaného transkriptu, modře – neanotované transkripty, šedě – intron nebo intergenový region (převzato a upraveno z Kim *et al.*, 2013)

Dalším programem, využívaným k analýze alternativního sestřihu je StringTie. Tento program využívá algoritmus toků v sítích. Při analýze simulovaných a reálných datových souborů dosáhl StringTie lepších výsledků, než jiné „assemblery“, jako je například Cufflinks. StringTie produkuje úplnější a přesnější rekonstrukce genů

a má i lepší odhady úrovně exprese a rovněž ve srovnání s jinými programy StringTie běží rychleji.

StringTie (Obr. 13) v rámci své činnosti sestavuje sestřihový graf, který následně využívá pro hledání izoforem. Vstupy StringTie mohou zahrnovat nejen zarovnání sestřižených „readů“, ale také zarovnání kontigů, které jsou již předsestaveny ze sekvenačních „readů“. Poté StringTie iterativně extrahuje nejlepší cestu ze sestřihového grafu, zkonstruuje tokovou síť, vypočítá maximální tok pro odhad četnosti a následně aktualizuje sestřihové grafy vyřazením „readů“, které byly přiřazeny algoritmem toku. Tento proces se opakuje, dokud nebudou přiřazeny všechny „ready“.

StringTie na rozdíl od Cufflinks, sestaví transkripty a úrovně exprese odhaduje u všech současně. StringTie nejprve seskupí „ready“ do klastrů a pro každý vzniklý klastr vyrobí sestřihový graf, z něhož identifikuje transkripty, a poté pro každý transkript vytvoří tokovou síť pro odhad úrovně exprese použitím algoritmu maximálního toku (Pertea *et al.*, 2015).



Obr 13. Princip StringTie (převzato a upraveno z Pertea *et al.*, 2015)

3 EXPERIMENTÁLNÍ ČÁST

3.1 Biologický materiál a sekvenování RNA

Jako biologický materiál byly vybrány transgenní rostliny s vakuolárním genem AtCKX1 a divoké (WT) rostliny odrůdy jarního ječmene Golden Promise, které byly pěstovány v prostředí s fotoperiodou 15°C/16 hodin ve světle a 12°C/8 hodin ve tmě. Zdroj světla byl kombinací rtuťových wolframových lamp a sodíkových výbojek o intenzitě 300 μmol m^{-2} s^{-1}. Rostliny byly pěstovány v hydroponii v modifikovaném Hoaglandově roztoku.

Před začátkem stresu byly odebrány tři vzorky jako biologické replikáty. Stres suchem byl vyvolán u rostlin starých 4 týdny, vylitím Hoaglandova roztoku z nádoby. Po 24 hodinách aplikace stresu byly rostliny vráceny do nádoby. U vybraných rostlin byl odebrán kořenový systém po 24 hodinovém stresu a 12 hodin a 14 dní po revitalizaci. Revitalizace byla provedena návratem rostlin do živného roztoku.

Izolace celkové RNA byla provedena s použitím RNAqueous Kit (Life Technologies, USA). Izolovaná RNA byla poté ošetřena s použitím soupravy TURBO DNA-free Kit (Life Technologies) a purifikována pomocí magnetických kuliček (Agencourt RNA CLEAN XP, Beckman Coulter, USA).

2,5 μg celkové RNA z každého vzorku, extrahovaného výše popsaným způsobem, byla použita pro přípravu cDNA knihovny pomocí Ilumina TruSeq Stranded mRNA Sample Preparation Kit (Illumina, USA). Koncentrace knihovny byla vyhodnocena s využitím Kapa Library Quantification Kit (Kapa Biosystems, USA) a všechny knihovny byly shromážděny do konečné koncentrace 8 pM pro generování klastru a sekvenování. Klastry byly generovány za použití Illumina1 TruSeql SR Cluster Kit v3 cBot HS a sekvenovány na HiSeq SR Flow Cell v3 s HiSeq 2500 Sequencing Systém.

3.2 Bioinformatická analýza

Pro samotnou bakalářskou práci byla poskytnuta RNA-seq data, a tato vstupní data byla ve formátu FASTQ. Všechny programy použité v experimentální části byly spouštěny na operačním systému Linux systém Ubuntu 12.04.

Pomocí volně dostupného programu FastQC verze 0.11.5 (z FastQC: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), sloužícího ke kontrole kvality sekvenčních dat s vysokou propustností, byly získány informace o celkovém počtu sekvencí a poměru GC a dalších charakteristikách vzorků. Získaná data byla následně zpracována do tabulky.

Pro mapování „readů“ vzniklých RNA sekvenováním byl použit program TopHat2 (z TopHat: http://ccb.jhu.edu/software/tophat/index.shtml). „Ready“ byly mapovány na referenční genom ječmene v32, který byl získán z databáze ENSEMBL.

S pomocí programu TopHat2 byla provedena optimalizace procesu mapování „readů“ s využitím parametru -N představujícího počet neshod mezi mapovanými „ready“ a referenčním genomem. Cílem bylo najít optimální hodnotu tohoto parametru, při níž došlo k namapování největšího množství „readů“ na referenční genom. Program byl použit pro všechny vzorky. Syntaxe příkazu používaného ke spuštění programu TopHat2 byl v následujícím tvaru:

Tophat2 -p 12 -N 2 –o výstupní_soubor reference_soubor vstupní_soubor.fastq

Popis parametrů využitých ke spuštění programu TopHat2:

-p	Počet CPU, využitých při procesu mapování.
-N	„Ready“, které jsou mapovány na referenční genom a mající počet neshodných nukleotidů, než jaký určuje -N, jsou vyřazeny z další analýzy.
-o	Výstupní soubor
vstupní_soubor	Vstupní soubor ve fastq formátu.
reference_soubor	Soubor referenčního genomu ve formátu fasta.

Výstupní soubory získané ze všech analyzovaných vzorků pomocí programu TopHat2, byly použity ke kvantifikaci počtu zarovnaných „readů“ k anotovaným genům, a to využitím programu featureCounts (ze Subread: http://subread.sourceforge.net/). Analyzován byl multimaping a další nežádoucí jevy.

Syntaxe příkazu používaného ke spuštění programu featureCounts byl v následujícím tvaru:

featureCounts -a reference_soubor -o výstupní_soubor -F GTF -t exon -s 2 -T 5 vstupní_soubor

Popis parametrů využitých ke spuštění programu featureCounts:

-a	Anotační soubor - soubor obsahující referenční genom ve formátu gtf.
-o	Výstupní soubor ve formátu txt, obsahující informace o kvantifikaci „readů“.
-F	Specifikace formátu anotačního souboru. Defaultní hodnota je GTF.
-t	Specifikace exonů. Do kvantifikace počtu „readů“ budou zahrnuty pouze ty řádky, které jsou odpovídající pro tuto specifikaci. Defaultní hodnota parametru je exon.
-s	Specifikace knihovny, která je využita pro generování vstupního souboru (0 = unstranded, 1 = single-stranded, 2 = reverse-stranded).
-T	Počet CPU využitých při kvantifikaci.
vstupní_soubor	Vstupní soubor ve formátu BAM (obsahuje výsledky výpočtů programu TopHat2).

Dalším krokem experimentální části byla analýza alternativního sestřihu dvěma k tomu určenými bioinformatickými nástroji. První z vybraných programů byl Cufflinks (z Cufflinks: http://cole-trapnell-lab.github.io/cufflinks/).

Syntaxe příkazu používaného ke spuštění programu Cufflinks byl v následujícím tvaru:

cufflinks -p 6 -g reference_soubor -o výstupní_soubor vstupní_soubor

Popis parametrů využitých ke spuštění programu Cufflinks:

-p	Počet CPU využitých při sestavení transkriptů.
-g	Anotační soubor. Využití referenční sekvence jako průvodce k sestavení transkriptů, hlášení nových genů a izoforem.
-o	Výstupní soubor.
vstupní_soubor	Vstupní soubor ve formátu BAM (obsahuje výsledky výpočtů programu TopHat2).

Druhým programem použitým pro analýzu alternativního sestřihu byl použit rychlý StringTie (z StringTie: https://ccb.jhu.edu/software/stringtie/). Syntaxe příkazu používaného ke spuštění programu StringTie byl v následujícím tvaru:

stringtie vstupní_soubor -p 6 -G reference_soubor -o výstupní_soubor

Popis parametrů využitých ke spuštění programu StringTie:

vstupní_soubor	Vstupní soubor ve formátu BAM (obsahuje výsledky výpočtů programu TopHat2).
-p	Počet CPU využitých při sestavení transkriptů.
-G	Anotační soubor – reference ve formátu gtf.
-o	Výstupní soubor, obsahující informace o sestavených transkriptech.

Výsledky obou programů byly zpracovány do tabulek a ze získaných informací byl vysloven závěr. Rovněž byla u vybraných genů provedena analýza zastoupení jednotlivých izoforem ve zkoumaných skupinách vzorků. Jednalo se zejména o geny zapojené do metabolismu cytokininů, a to jmenovitě o cytokinin-oxidasy-dehydrogenasy.

4 VÝSLEDKY A DISKUZE

4.1 Kontrola kvality sekvenačních dat

Provedením kontroly kvality dat získaných RNA sekvenováním pomocí programu FastQC, byly zjištěny informace o poměru GC a celkovém počtu sekvencí.

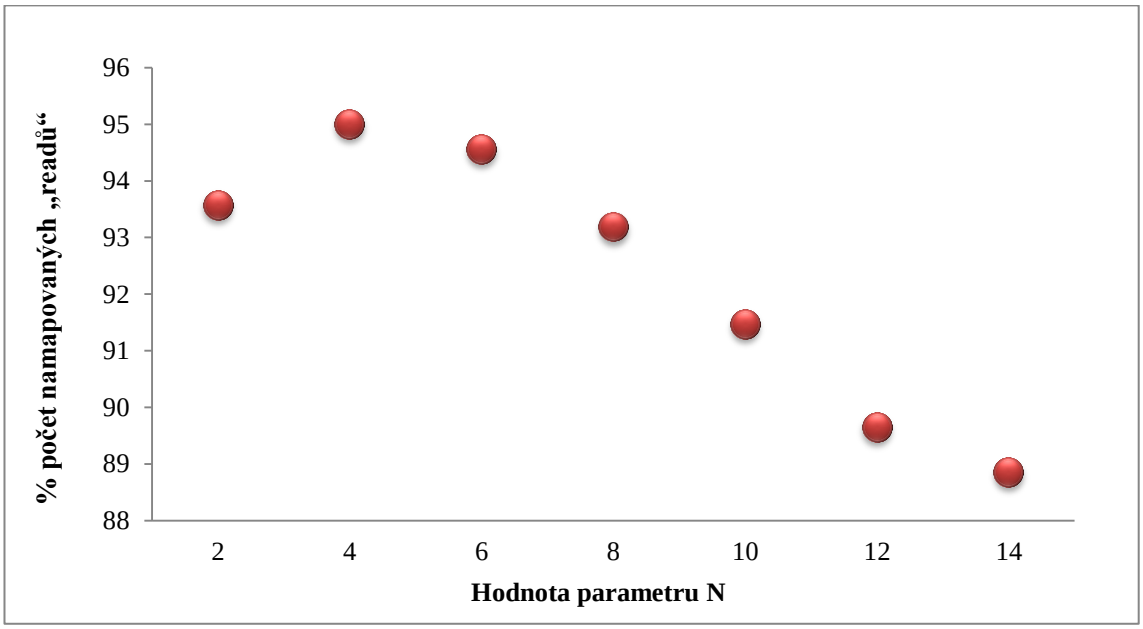
Výsledky poměru GC, jehož očekávaná hodnota je 50 % prokázaly, že sekvenace proběhla úspěšně. Hodnoty poměru GC v analyzovaných vzorcích se pohybovaly v rozmezí 48 – 52 %. Získáním informací o celkovém počtu „readů“ bylo zjištěno, že nejvyšší počet „readů“ má vzorek CTRL_WT_1 a nejnižší počet „readů“ má vzorek S_WT_2 (Tab. 1). U žádného ze zkoumaných vzorků, nebyla nalezena kontaminace krátkými k-ticemi nukleotidů a rovněž nebyly nalezeny kontaminující sekvence z adaptorů využitých v průběhu sekvenování.

Tab. 1: Výsledky z programu FastQC. Zkratka VAK – transgenní rostliny s vAtCKX1 genem a WT – netransgenní rostliny. Zkratka CTRL značí dobu před stresem, S značí dobu během stresu, R12H je 12 hodin po ukončení stresu, R2F je 2 týdny po ukončení stresu.

Název vzorku	**Celkový počet „readů“**	**Poměr GC [%]**
CTRL_VAK_1	75 279 326	50
CTRL_VAK_2	71 354 832	51
CTRL_WT_1	84 517 106	50
CTRL_WT_2	63 270 045	50
R2T_VAK_1	71 287 568	51
R2T_VAK_2	59 187 683	51
R2T_WT_1	68 521 784	49
R2T_WT_2	66 446 892	50
R12H_VAK_1	71 082 657	49
R12H_VAK_2	69 436 337	51
R12H_WT_1	65 959 107	51
R12H_WT_2	51 142 352	52
S_VAK_1	71 985 223	50
S_VAK_2	73 507 253	50
S_WT_1	77 801 469	48
S_WT_2	46 104 470	51

4.2 Optimalizace procesu mapování

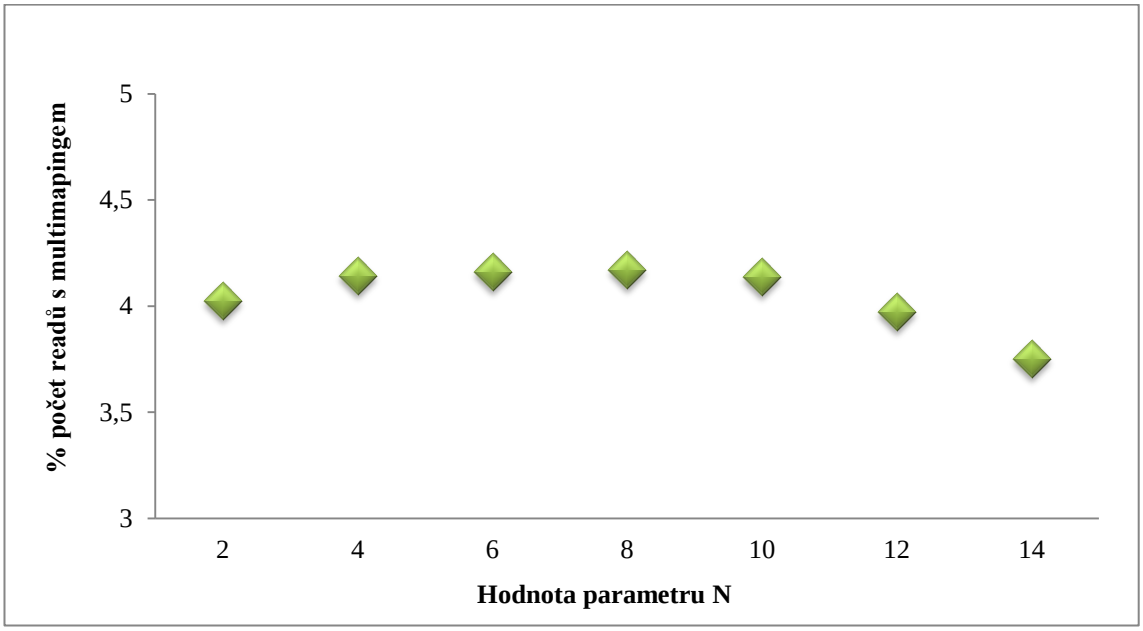
K určení optimální hodnoty parametru -N, se jeho hodnota postupně navyšovala o 2, v rozmezí od 2 do 14. Sledovanými charakteristikami při procesu mapování, byly informace o namapovaných „readech“ na referenční genom a informace o mnohočetném přiřazení „readů“, které jsou uváděny jako multimaping (Příloha 1). Vyhodnocením charakteristik bylo zjištěno, že optimální hodnota, při které dochází k namapování největšího počtu „readů“ na referenční genom, je 4. U analyzovaného vzorku S_WT_1, byla optimální hodnota parametru -N rovna 6. Nicméně pro další analýzu byla i u tohoto vzorku použita hodnota 4. U hodnot vyšších než 4, se množství úspěšně namapovaných „readů“ postupně zmenšovalo. Tento nárůst a postupný pokles procentuálního počtu namapovaných „readů“ byl potvrzen výpočtem průměrného množství z procentuálního počtu namapovaných „readů“, které odpovídají daným analyzovaným vzorkům, a to postupně pro každou hodnotu parametru –N (Obr. 14).



Obr 14. Grafické znázornění závislosti procentuálního počtu namapovaných „readů“ na hodnotě parametru -N.

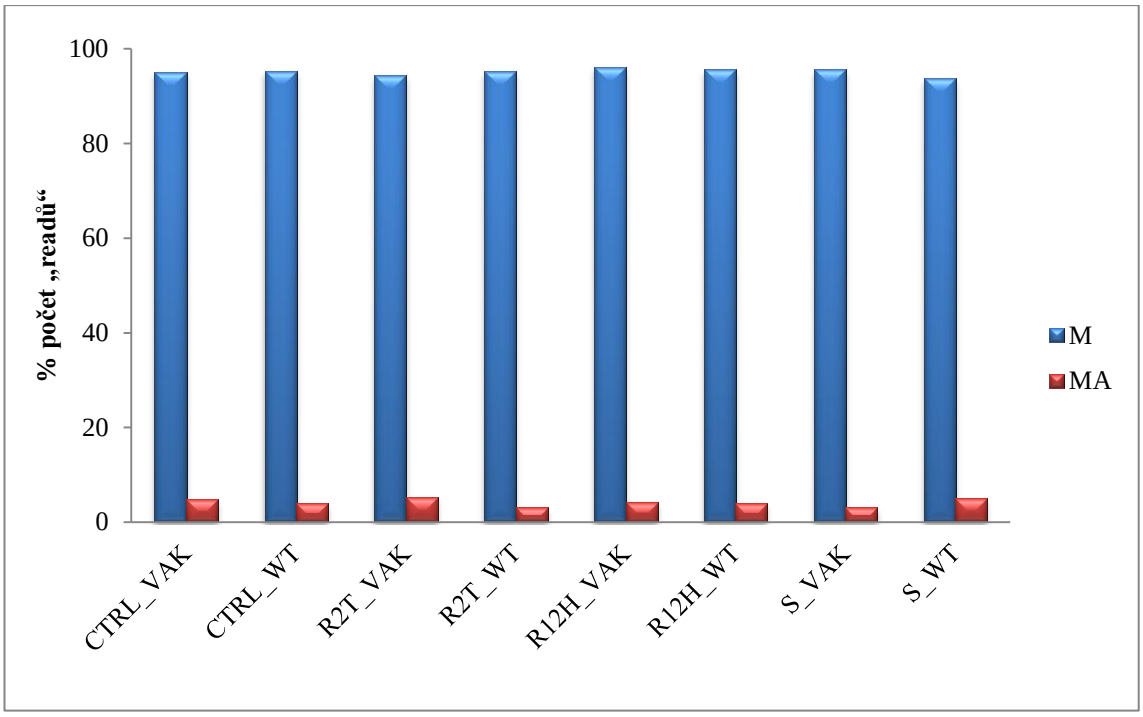
Analýza pomocí programu TopHat2 také poskytla informace o multimapingu, přesněji o procentuálním množství „readů“ s multimapingem, a to z celkového počtu namapovaných „readů“, u každého analyzovaného vzorku. Z procentuálních hodnot počtu „readů“ s multimapingem bylo podobně jako u analýzy namapovaných „readů“ vypočítáno průměrné množství, pro každou hodnotu parametru –N. Výpočtem

průměrných hodnot bylo zjištěno, že se procentuální počet „readů“ s multimapingem pohyboval kolem 4 % u všech analyzovaných vzorků (Obr. 15).



Obr 15. Grafické znázornění závislosti procentuálního počtu „readů“ s multimapingem na hodnotě parametru –N.

Ke zjištění odlišností v rámci transgenních rostlin a WT („wild type“) rostlin, byly vypočteny průměrné hodnoty pro každou fázi experimentu a rovněž pro mutantní a WT rostliny pro optimální hodnotu parametru -N (N = 4). K výpočtu průměru bylo využito procentuální množství „readů“ namapovaných na referenční genom, a to vždy v rámci dvou vzorků stejného typu (např. výpočet průměru u vzorků CTRL_VAK_1 a CTRL_VAK_2, či vzorků CTRL_WT_1 a CTRL_WT_2) a následně byly vypočítány také průměrné hodnoty z procentuálního zastoupení mnohočetného přiřazení. Vypočtené průměrné hodnoty byly následně zpracovány do grafu (Obr. 16), z něhož bylo odvozeno, že k odlišenostem v rámci transgenních a WT rostlin nedošlo a tedy u všech vzorků bylo docíleno přibližně stejných procentuálních hodnot jak namapovaných „readů“ tak i „readů“ podléhajících multimapingu.



Obr 16. Grafické znázornění průměrných hodnot namapovaných „readů“ a mnohočetného přiřazení „readů“ pro typy rostlin (WT a VAK) a jednotlivé fáze experimentu (CTRL – kontrola před stresem, S – průběh stresu, R12H – 12 hodin po revitalizaci, R2T – 2 týdny po revitalizaci). M – namapované „ready“, MA – „ready“ s multimapingem

4.3 Kvantifikace „readů“

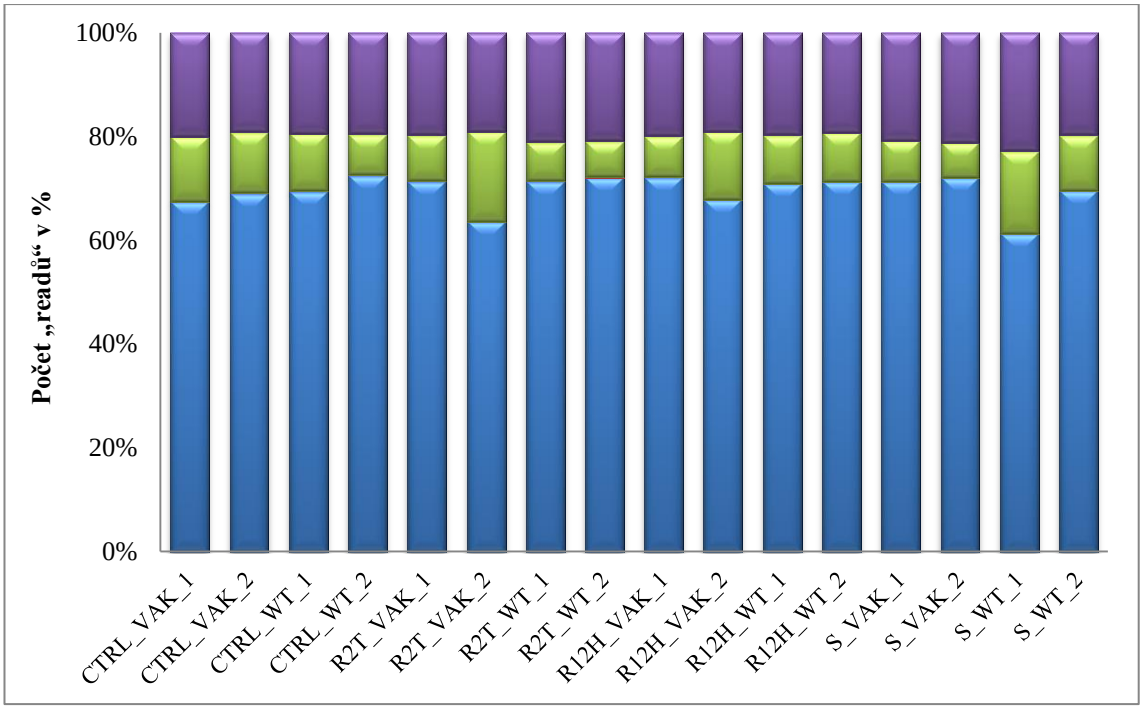
Výstupní soubory získané procesem mapování na referenční genom pomocí programu TopHat2, byly zpracovány programem FeatureCounts, který slouží ke kvantifikaci „readů“ získaných v procesu sekvenování.

Program FeatureCounts poskytl užitečné informace o „readech“ (Příloha 2), a to jmenovitě o počtu „readů“, které byly mapovány do oblastí genů, které jsou již anotovány, tedy podává zprávy o anotované oblasti reference. Nejvyšší počet „readů“ v anotované oblasti měl vzorek CTRL_WT_1 (Příloha 2) a naopak nejnižší počet „readů“ byl nalezen u vzorku S_WT_2 (Příloha 2). Dále se díky tomuto programu získaly informace o „readech“, u kterých nelze přesně určit, ve kterém genu se nachází, neboť zasahují současně do oblastí dvou genů. A program také podal zprávy o multimapingu, kdy se v tomto případě mohou „ready“ mapovat na více míst v referenčním genomu.

Pro analýzu je významná také neanotovaná oblast reference. Jedná se o případ, kdy „ready“ nejsou umístěny v oblasti žádného anotovaného genu. Tato oblast je důležitá z toho důvodu, že může obsahovat oblasti, ve kterých mohou

být anotovány nové izoformy, případně i dosud neanotované geny které je následnou bioinformatickou analýzou možné anotovat a následně také kvantifikovat. Nejvyšší a nejnižší počet „readů“, které se vyskytovaly v této oblasti, měly opět vzorky CTRL_WT_1 a S_WT_2 (Příloha 2).

Analýzou procentuálního rozdělení (Obr. 17) jednotlivých oblastí, do kterých byly „ready“ mapovány, bylo zjištěno, že oblast, obsahující nejvyšší počet „readů“ zahrnuje „ready“ v anotované oblasti sekvence. Tato skutečnost ukazuje na poměrně dobrou kvalitu anotace referenčního genomu, který postihuje většinu unikátních „readů“ namapovaných na referenční genom.



Obr 17. Procentuální zastoupení výskytu „readů“ v oblastech referenčního genomu. Fialová barva – neanotovaná oblast reference, zelená barva – nepřiřazené „ready“ v důsledku mnohočetného přiřazení, modrá barva – anotovaná oblast reference.

U téměř všech vzorků byl nalezen podobný poměr „readů“ namapovaných do oblastí již anotovaných genů. Počet „readů“ téměř všech vzorků v anotované oblasti odpovídal necelým 70 % z celkového počtu namapovaných readů. Oblast multimapingu zahrnovala 10 % z celkového počtu „readů“. Výskyt „readů“ tohoto druhu je zde očekáván, a to především z důvodu výskytu repetitivních sekvencí v referenčním genomu. Výskyt „readů“ v neanotované oblasti odpovídal cca 20 % z celkového počtu namapovaných „readů“ u všech vzorků. Tento podíl „readů“ může být zapříčiněn jednak mírnými neshodami readů získaných v průběhu sekvenování se zvolenou referencí nebo také skutečností, že v referenčním genomu použitém pro tento

experiment stále ještě nejsou exaktně popsány všechny geny, které podléhají transkripci do mRNA.

Počet readů, které nemohly být jednoznačně přiřazeny, například v důsledku přerývajících se genů se pohyboval na velmi nízkých hodnotách v řádech tisíců a tudíž i jeho procentuální podíl na celkovém počtu namapovaných „readů“ dosahoval zanedbatelných hodnot. Z toho důvodu, tyto hodnoty v grafu (Obr. 18) nejsou pro svou velikost fyzicky viditelné.

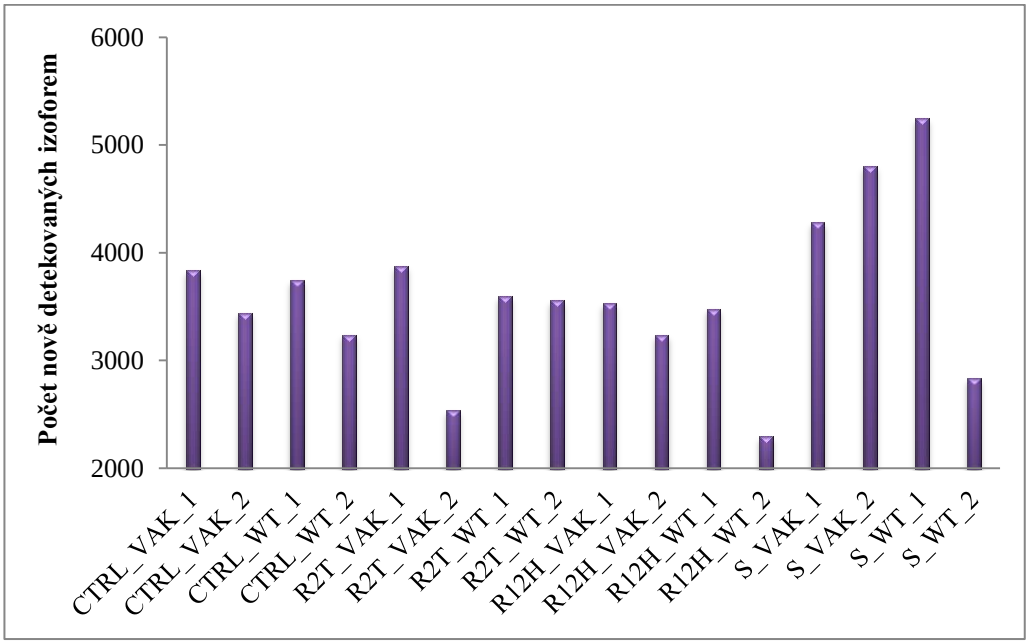
Hodnoty procentuálního poměru se u analyzovaných vzorků R2T_VAK_2 a S_WT_1 oproti hodnotám ostatních vzorků mírně lišily. Důvodem byl vyšší počet „readů“ nacházejících se v oblasti, kde podléhají multimapingu. Této oblasti totiž odpovídalo přibližně 16 % z celkového počtu „readů“ u obou daných analyzovaných vzorků. V anotované oblasti se nacházelo pouze cca 60 % „readů“. Počet „readů“ v neanotované oblasti již správně odpovídal 20 % z celkového počtu namapovaných „readů“ jako to bylo v případě zbylých vzorků a zanedbatelnému množství nejednoznačně namapovaných „readů“ opět odpovídalo 0 %.

4.4 Kvalitativní a kvantitativní analýza alternativního sestřihu

K získání důležitých informací o počtu nově detekovaných izoforem byly použity programy Cufflinks a StringTie. Výstupem programu Cufflinks byly čtyři výstupní soubory, pro analýzu byl vybrán soubor transcripts.gtf obsahující údaje o úspěšně nalezených izoformách. Nejdůležitějšími charakteristikami pro další analýzu byl relativní výskyt izoformy ve fragmentech na kilobázy na milión mapovaných fragmentů (FPKM) a hodnoty určující pokrytí transkriptu mapovanými „ready“.

Každé jednotlivé izoformě v souboru odpovídal jeden řádek ve výstupním souboru. Pro zjištění počtu nově detekovaných izoforem byly vybrány pouze transkripty, jejichž pokrytí přesahovalo 100 (byly tedy minimálně 100x pokryty sekvenovanými „ready“). Ze získaných transkriptů byly následně vybrány pouze ty, které nebyly anotovány v použité referenci. Získané počty nově definovaných izoforem byly zobrazeny ve formě grafu.

Tímto způsobem byly analyzovány všechny vzorky a bylo zjištěno, že největší množství nově detekovaných izoforem bylo nalezeno v analyzovaném vzorku S_WT_1 a naopak nejnižší počet byl nalezen u vzorku R12H_WT_1 (Obr. 18, Příloha 3).



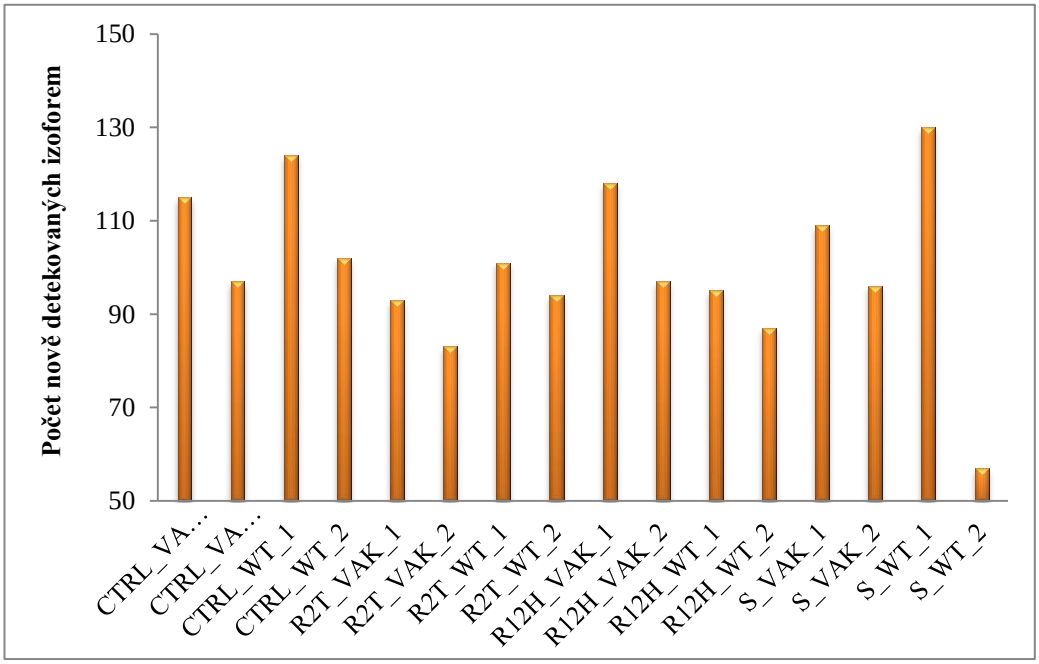
Obr 18. Grafické znázornění množství nově detekovaných izoforem u všech analyzovaných vzorků.

Analýza výsledků získaných programem StringTie byla provedena podobně jako u Cufflinks. Pro zjištění počtu nově detekovaných izoforem byly opět vybrány pouze transkripty jejichž pokrytí přesahovalo 100x a rovněž nebyly anotovány v použité referenci. Počty získaných izoforem byly následně zpracovány do podoby grafu (Obr. 19).

Výsledky analýzy pomocí StringTie poskytly výrazně méně nově detekovaných izoforem ve srovnání s programem Cufflinks (Příloha 3). Při srovnání výstupů z obou programů byl v obou případech nalezen největší počet nově detekovaných izoforem u vzorku S_WT_1. Počet nových izoforem se u vzorků R12H_VAK_1 a CTRL_WT_1 pohyboval kolem 120, u vzorku S_WT_2 bylo nalezeno nejméně nových izoforem, a to pouze 57 (Obr 19, Příloha 4).

V případě programu Cufflinks se ukazuje zvýšený výskyt u vzorků podrobených stresu ve srovnání se všemi ostatními vzorky. Tuto skutečnost si lze vysvětlit, že v průběhu stresu dochází k odlišnému sestřihu transkriptů, a tudíž je možné detekovat rozdílné množství nových izoforem mezi vzorky odebranými v průběhu stresu a vzorky odebranými před a po stresu. Ověření této hypotézy by ovšem

vyžadovalo hlubší porovnání získaných transkriptů především na úrovni sekvencí a rovněž bližší studium diferenciální exprese izoforem, které jsou těmito sekvencemi reprezentovány.



Obr 19. Grafické znázornění množství nově detekovaných izoforem, které byly zjištěny pomocí programu StringTie.

V rámci kvantitativní analýzy izoforem byla zvláštní pozornost věnována vybraným genům CKX (cytokinin oxidáz/dehydrogenáz) které jsou zapojeny do degradace cytokininů. Pro následnou analýzu byly vybrány geny CKX5 a CKX11 jelikož pouze tyto geny dosahovaly dostatečného pokrytí ve studovaných vzorcích, které dosahovalo alespoň desetinásobného pokrytí pro celý gen. Výsledky pro tyto geny byly zpracovány do tabulky, kde jsou uvedeny normalizované FPKM hodnoty pro srovnávané skupiny vzorků (Tab. 2).

Na základě údajů v tabulce bylo zjištěno, že u transkriptu CKX5.1 byly nalezeny rozdíly hodnot FPKM v průběhu stresu a následném revitalizačním procesu ve srovnání s hodnotou zjištěnou před stresem. To ukazuje na změnu exprese této izoformy v průběhu vystavení stresu což koresponduje s předpokladem zapojení cytokininů v adaptaci rostlin na abiotický stres. Mírné změny v genové expresi transkriptu CKX5.2 mohou naznačovat regulaci celkové hladiny exprese genu CKX5 pomocí alternativního sestřihu, nicméně pro exaktní zhodnocení nárůstu této izoformy je třeba ověřit expresi této izoformy dalšími molekulárně biologickými metodami, kde jako příklad lze uvést například vhodnou techniku PCR. Nejvyšší exprese genu CKX5 byla dosažena u vzorků

odebraných 12 hodin po revitalizaci. Následně již hodnota exprese genu CKX5 klesala přibližně na polovinu což bylo pozorováno u vzorků odebraných dva týdny po revitalizaci. Mezi WT a mutantními rostlinami nebyly zaznamenány výrazné rozdíly mezi jednotlivými časovými body ve kterých byly odebírány vzorky.

Tab. 2: Normalizované hodnoty FPKM pro srovnávané skupiny vzorků. VAK – mutantní rostliny s vakuolárním AtCKX1; WT – divoký typ rostlin; CTRL – vzorky před stresem; S – vzorky v průběhu stresu; R12H – vzorky 12 hodin po stresu; R2T – vzorky 2 týdny po stresu.

Analyzované vzorky	**Analyzované geny**			
	CKX5.1	**CKX5.2**	**CKX11.1**	**CKX11.2**
CTRL_VAK	3,55	0,13	2,24	0,21
S_VAK	3,44	0,06	*ND	*ND
R12H_VAK	6,07	0,18	2,52	2,03
R2T_VAK	3,34	0,34	2,47	*ND
CTRL_WT	4,47	*ND	*ND	*ND
S_WT	2,45	0,28	*ND	*ND
R12H_WT	7,42	0,13	2,41	1,78
R2T_WT	2,62	0,1	2,28	1,25

*ND – nenaměřené hodnoty (not detected)

Transkripty genu CKX11 poskytly méně informací ve srovnání s transkripty u genu CKX5. Nejvyšší exprese těchto transkriptů byla opět nalezena u vzorků odebraných 12 hodin po revitalizačním procesu, a to jak u vzorků WT, tak i u vzorků z mutantu. Tato shoda s expresí transkriptů genu CKX5 rovněž podporuje hypotézu, že v průběhu stresu dochází k zapojení genů pro CKX do procesu adaptace rostlin na stres suchem. Ve srovnání s genem CKX5 jsou ovšem u genu CKX11 poskytnuty zajímavější informace o zastoupení izoforem tohoto genu v průběhu stresu. Za zajímavé lze určitě považovat změnu zastoupení transkriptu CKX11.2 v průběhu revitalizačního procesu, ve srovnání se stavem před aplikací stresu u vzorků z WT. Je očividné, že u WT dochází k výraznému nárůstu exprese transkriptů CKX11.2 což ukazují nárůst hodnot FPKM z hodnoty 0,21 na hodnotu 2,03. U mutantů v počáteční fázi experimentu není možné ani jednu z uvedených izoforem detekovat, nicméně v průběhu revitalizačního procesu jsou pozorovány podobné hodnoty jako v případě WT vzorků. Skutečnost, zda je nemožnost detekovat expresi izoforem genu CKX11 způsobena velmi nízkou expresí těchto izoforem v mutantních rostlinách, nebo nedostatečnou hloubkou sekvenování bude pravděpodobně nutné ověřit pomocí vhodné PCR techniky.

U genu CKX11 rovněž nebyly pozorovány výrazné rozdíly v genové expresi mezi WT a mutantem v pozorovaných časových bodech.

Celkově výsledky získané v rámci kvantitativní analýzy genů CKX ukazují na zapojení těchto genů v procesu adaptace rostliny na stres suchem, a to jak u mutantních rostlin, tak i u rostlin WT. Zajímavou skutečností je změna exprese izoformy CKX11.2 u WT rostlin v reakci na revitalizační proces po aplikaci stresu ve srovnání s transkriptem CKX11.1 u kterého jsou pozorované hodnoty pouze mírně odlišné.

5 ZÁVĚR

Bakalářská práce byla zaměřena na analýzu RNA-Seq dat ječmene setého s ohledem na identifikaci nových izoforem vzniklých alternativním sestřihem, využitím bioinformatických programů sloužících k tomuto účelu. U sledovaných vzorků byla rovněž provedena kvantitativní analýza vybraných genů CKX a jejich izoforem ve sledovaných časových bodech před stresem, v průběhu stresu suchem a 12 hodin a dva týdny po zahájení revitalizačního procesu.

Teoretická část bakalářské práce je strukturována do čtyř kapitol. Úvodní kapitola je zaměřena na popis ječmene setého a jeho využití v zemědělství a k výzkumným účelům. Druhá kapitola je věnována rozdělení cytokininů a jejich zapojením do důležitých fyziologických procesů a poslední odstavec kapitoly popisuje jejich biosyntézu. Třetí kapitola teoretické části se zabývá rostlinným stresem, konkrétně jeho rozdělením na abiotický a biotický a jejich působením na rostlinu. Následující kapitola je věnována popisu kódujících a nekódujících oblastí prekurzorové mRNA a také popisem sestřihových míst. Následně se kapitola zabývá alternativním sestřihem, rozdělením variant alternativního sestřihu a také jeho samotným průběhem. Poslední pátá kapitola se zabývá rozdělením a popisem metod a jim příslušným programům sloužících ke kvalitativní a kvantitativní analýze alternativního sestřihu.

Experimentální část byla zahájena kontrolou kvality sekvenačních dat. Následně byla provedena optimalizace procesu mapování na referenční genom pomocí programu TopHat2. Data získaná tímto procesem byla dále zpracována programem na kvantifikaci „readů“ featureCounts. Dalším krokem byla kvalitativní analýza transkriptů a k tomuto účelu byly použity programy Cufflinks a StringTie. Následnou analýzou výstupů z těchto programů, byl zjištěn počet nově detekovaných izoforem a bylo zjištěno, že programy poskytují různé informace což je pravděpodobně způsobeno odlišným algoritmickým přístupem k jejich detekci a klasifikaci.

Posledním úkolem experimentální části byla analýza genů cytokinin oxidáz/reduktáz (CKX). Informace o těchto genech byly zjišťovány na základě výsledků poskytnutých programy Cufflinks a StringTie a bylo zjištěno, že pouze geny CKX5 a CKX11 mají dostatečné pokrytí pro celý gen a tedy pouze tyto geny byly vybrány pro kvantitativní analýzu izoforem. Následnou analýzou bylo

zjištěno, že u těchto genů nedochází k výrazným změnám mezi rostlinami WT a rostlinami mutanta AtCKX1. S pomocí kvantitativní analýzy byly ovšem odhaleny změny v expresu jednotlivých izoforem v průběhu aplikace stresu a v rámci následného revitalizačního procesu, a to jak u rostlin mutantních, tak i u rostlin WT.

6 LITERATURA

Ahmad M., Mohammed F., Maqbool K., Azamand A., Iqbal S. (2003): Genetic variability and traits correlation in wheat. *Sarhad J. Agric* **19**, 347-351.

Alamancos G. P., Agirre E., Eyras E. (2013): Methods to study splicing from high-throughput RNA Sequencing data. *Methods Mol Biol* **1126**, 357–397.

Alvarez V. E. (2008): Autophagy Is Involved in Nutritional Stress Response and Differentiation in Trypanosoma cruzi. *J Biol Chem* **283**, 3454-64.

Anders S., Reyes A., Huber W. (2012): Detecting differential usage of exons fromRNA-seq data. *Genome Res* **22**, 2008–2017.

Aschoff M., Hotz-Wagenblatt A., Glatting K. H., Fischer M., Eils R., Kdonig R. (2013): SplicingCompass: differential splicing detection using RNA-seq data. *Bioinformatics* **29**, 1141–1148.

Ben-Dov C., Hartmann B., Lundgren J., Valcarcel J. (2008): Genome-wide analysis of alternative pre-mRNA splicing. *J. Biol. Chem* **283**, 1229–1233.

Bullard J. H., Purdom E., Hansen K. D., Dudoit S. (2010): Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics* **11**, 94.

Burge C. B., Tuschl T. H., Sharp P. A. (1999): Splicing of precursors to mRNAs by the spliceosomes. *The RNA World* **2**, 525–560.

Close TJ., Wanamaker S. I., Caldo R. A., Turner S. M., Ashlock D. A., Dickerson J. A. (2004): A new resource for cereal genomics: 22 K barley GeneChip comes of age. *Plant Physiol* **134**, 960–968.

Cohen G., Honkala I., Litsyn S., Lobstein A. (1997): Covering Codes. *North-Holland Mathematical Library* **54**, 16–17.

Cramer G. R., Epstein E., Läuchli A. (1990): Effects of sodium, potassium and calcium on salt-stressed barley. I. Growth analysis. *Physiol. Plant* **80,** 83–88.

Deluc L. G., Quilici D. R., Decendit A., Grimplet J., Wheatley M. D., Schlauch K. A., Merillon J. M., Cushman J. C., Cramer G. R. (2009): Water deficit alters differentially metabolic pathways affecting important flavor and quality traits in grape berries of Cabernet Sauvignon and Chardonnay. *BMC Genomics* **10,** 212.

Garber M., Grabherr M. G. (2011): Computational methods for transcriptome annotation and quantification using RNA-seq. *Nat Methods* **8**, 469–77.

Graveley B. R., Kaur A., Gunning D., Zipursky S. L., Rowen L., Clemens J. C. (2004): The organization and evolution of the dipteran and hymenopteran *Down syndrome cell adhesion molecule (Dscam)* genes. *RNA* **10**, 1499–1506.

Gray I. C., Jeffreys A. J. (1991): Evolutionary transience of hypervariable minisatellites in man and the primates. *Proc Biol Sci* **243**, 241–253.

Griffith M., Griffith O. L., Mwenifumbo J., Goya R., Morrissy A. S., Morin R. D., Corbett R., Tang M. J., Hou Y. C., Pugh T. J. (2010): Alternative expression analysis by RNA sequencing. *Nat. Methods* **7**, 843–847.

Ha C. V., Leyva-Gonzalez M. A., Osakabe Y., Tran U. T., Nishiyama R., Watanabe Y. (2014): Positive regulatory role of strigolactone in plant responses to drought and salt stress. *Proc. Natl. Acad. Sci. U.S.A.* **111** 581–856.

Hasanuzzaman M., Hossain M. A., Fujita M. (2011): Nitric oxide modulates antioxidant defense and the methylglyoxal detoxification system and reduces salinity-induced damage of wheat seedlings. *Plant Biotechnology Reports* **5**, 353–365.

Hasanuzzaman M., Hossain M. A., da Silva J. A. T., Fujita M. (2012): Plant responses and tolerance to abiotic oxidative stress: antioxidant defense is a key factor. In: Bandi V, Shanker AK, Shanker C, Mandapaka M., eds. Crop stress and its management: perspectives and strategies. *Germany: Springer*, 261–316.

Haseneyer G., Schmutzer T., Seidel M., Zhou R., Mascher M., Schön C. C., Taudien S., Scholz U., Stein N., Mayer K. F., Bauer E. (2011): From RNA-seq to large-scale genotyping - Genomics resources for rye (Secale cereale L.). *BMC Plant Biol* **11**, 131.

Hastings M. L., Krainer A. R. (2001): Pre-mRNA splicing in the new millennium. *Curr Opin Cell Biol* **13**, 302–309.

Hillier,L.W.,et al. (2008): Whole-genome sequencing and variant discovery in C. elegans. *Nat. Meth* **5**, 183–188.

Hirose N., Takei K., Kuroha T., Kamada-Nobusada T., Nayashi H., Sakakibara H. (2007): Regulation of cytokinin biosynthesis, compartmentalization and translocation. *Journal of Experimental Botany* **59**, 75-83.

Hu Y., Huang Y., Du Y., Orellana C. F., Singh D., Johnson A. R.,Monroy A., Kuan P. F., Hammond S. M., Makowski L., Randell S. H., Chiang D. Y., Hayes D. N., Jones C., Liu Y., Prins J. F., Liu J. (2013): DiffSplice: the genome-wide detection of differential splicing events with RNA-seq. *Nucleic Acids Res* **41**, 39.

Chen I. C., Hill J. K., Ohlemüller R., Roy D. B., Thomas C. D. (2011): Rapid Range Shifts of Species Associated with High Levels of Climate Warming. *Science* **333**, 1024-1026.

Chen M., Manley J. L. (2009): Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat Rev Mol Cell Biol* **10**, 741–754.

Chinnusamy V., Zhu J., Zhu J. K. (2007): Cold stress regulation of gene expression in plants. *Trends in Plant Science* **12**, 444-451.

Christensen A. B., Thordal-Christensen H., Zimmermann G., Gjetting T., Lyngkjaer M. F., Dudler R., (2004): The germinlike protein GLP4 exhibits superoxide dismutase activity and is an important component of quantitative resistance in wheat and barley. *Mol Plant Microbe Interact* **17**, 109–17.

Kampa D., Cheng J., Kapranov P., Yamanaka M., Brubaker S., Cawley S., Drenkow J., Piccolboni A., Bekiranov S., Helt G., Tammana H., Gingeras T. R. (2004): Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. *Genome Res* **14**, 331–342.

Keren H., Lev-Maor G., Ast G. (2010): Alternative splicing and evolution: diversification, exon definition and function. *Nat Rev Genet* **5**, 345-355.

Kim D., Pertea G., Trapnell C., Pimentel H., Kelley R., Salzberg S. L. (2013): TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology* **14**, 1474.

Kimura K., Wakamatsu A., Suzuki Y., Ota T., Nishikawa T. (2006): Diversification of transcriptional modulation: large-scale identification and characterization of putative alternative promoters of human genes. *Genome Res* **16**, 55–65.

Krasensky J., Jonak C. (2012): Drought, salt, and temperature stress-induced metabolic rearrangements and regulatory networks. *J Exp Bot* **63,** 1523–1524.

Lander E. S., Linton L. M., Birren B., Nusbaum C., Zody M. C., Baldwin J., Devon K., Dewar K., Doyle M., FitzHugh W., Funke R., Cage D., Harris K., Heaford A., Howland J., Kann L., Lehoczky J., LeVine R., McEwan P., McKernan K., Meldrim J., Mesirov J. P., Miranda C., Morris W., Naylor J., Raymond C., Rosetti M., Santos R., Sheridan A., Sougnez C., Stange-Thomann N., Stojanovic N., Subramanian A., Wyman D., Rogers J., Sulston J., Ainscough R., Beck S., Bentley D., Burton J., Clee C., Carter N., Coulson A., Deadman R., Deloukas P., Dunham A., Dunham I., Durbin R., French L., Grafham D., Gregory S., Hubbard T., Humphray S., Hunt A., Jones M., Lloyd C., McMurray A., Matthews L., Mercer S., Milne S., Mullikin J. C., Mungall A., Plumb R., Ross M., Shownkeen R., Sims S., Waterston R. H., Wilson R. K., Hillier L. W., McPherson J. D., Marra M. A., Mardis E. R., Fulton L. A., Chinwall A. T., Pepin K. H., Gish W. R., Chissoe S. L., Wendl M. C., Delehaunty K. D., Miner T. L., Delehaunty A., Kramer J. B., Cook L. L., Fulton R. S., Johnson D. L., Minx P. J., Clifton S. W., Hawkins T., Branscomb E., Predki P., Richardson P., Wenning S., Slezak T., Doggett N., Cheng J. F., Olsen A., Lucas S., Elkin C., Uberbacher E., Frazier M. (2001): Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921.

Li H., et al. (2008): Mapping short dna sequencing reads and calling variants using mapping quality scores. *Genome Res* **18**, 1851–1858.

Lim L. P.,Burge C. B. (2001): A computational analysis of sequence features involved in recognition of short introns. *Proc. Natl. Acad. Sci* **98**, 11193–11198.

Liu R., Loraine A. E., Dickerson J. A. (2014): Comparisons of computational methods for differential alternative splicing detection using RNA-seq in plant systems. *BMC Bioinformatics* **15**, 364.

Marioni J. C., Mason C. E., Mane S. M., Stephens M., Gilad Y. (2008): RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res* **18**, 1509–1517.

Mayer K. F. X., Martis M., Hedley P. E., Simkova H., Liu H., Morris J. A. (2011): Unlocking the barely genome by chromosomal and comparative genomics. *Plant Cell* **23**, 1249–1263.

Mayer K. F. X., Waugh R., Langridge P., Close T. J., Wise R. P., Graner A. (2012): A physical, genetic and functional sequence assembly of the barley genome. *Nature* **491**, 711–716.

Munns R., Tester M. (2008): Mechanisms of salinity tolerance. *Annu Rev Plant Biol* **59,** 651-681.

Ogle, Maureen (2006): Ambitious brew : the story of American beer. *Orlando: Harcourt*, 70–72.

Orengo J. P., Cooper T. A. (2007): Alternative splicing in disease. *Adv. Exp. Med. Biol* **623**, 212–223.

Pertea M., Pertea G. M., Antonescu C. M., Chang T-C., Mendell J. T., Salzberg S. L. (2015): StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* **33**, 290 – 295.

Pozzoli U., et al. (2007): Intron size in mammals: complexity comes to terms with economy. *Trends Genet* **23**, 20–24.

Rogers M. F. (2012): SpliceGrapher: detecting patterns of alternative splicing from RNA-Seq data in the context of gene models and EST data. *Genome Biol* **13**.

Romero P. R., Zaidi S., Fang Y., Uversky V. N., Radivojac P., Oldfield C. J., Cortese M. S., Sickmeier M., LeGall T., Obradovic Z., Dunker A. K. (2006): Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms, *Proc. Natl. Acad. Sci. U. S. A.* **103**, 8390–8395.

Saritha P., Aparna C., Himabindu V., Anjaneyulu Y. (2007): Comparison of various advanced oxidation processes for the degradation of 4-chloro-2 nitrophenol. *J. Hazard. Mater* **149**, 609–614.

Lisar S. Y. S., Motafakkerazad R., Hossain M. M., Rahman I. M. M. (2012):Water Stress in Plants: Causes, Effects and Responses, *InTech*. 1–14.

Shen S., Park J. W., Lu Z. X., Lin L., Henry M. D., Wu Y. N., Zhou Q., Xing Y. (2014): rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl. Acad. Sci. USA* **111**, 5593–5601.

Sorek R., Shemesh R., Cohen Y., Basechess O., Ast G., Shamir R. (2004): A non-EST-based method for exon-skipping prediction. *Genome Res* **14**, 1617–1623.

Srivastava S., Pathak A.D., Gupta P. S., Shrivastava A. K., Srivastava, A. K. (2012): Hydrogen peroxide-scavenging enzymes impart tolerance to high temperature induced oxidative stress in sugarcane. *J. Environ. Biol* **33**, 657–661.

Stamm S., Ben-Ari S., Rafalska I., Tang Y., Zhang Z., Toiber D., Thanaraj T. A., Soreq H. (2005): Function of alternative splicing, *Gene* **344**, 1–20.

Steponkus P. L. (1984): Role of the plasma membrane in freezing injury and cold acclimation. *Annu Rev Plant Physiol* **35**, 543-584.

Steponkus P. L., Uemura M., Webb M. S. (1993): A contrast of the cryostability of the plasma membrane of winter rye and spring oat-two species that widely differ in their freezing tolerance and plasma membrane lipid composition. In PL Steponkus, ed, Advances in Low-Temperature Biology. *JAI Press* **2**, 211-312.

Sun H., Chasin L. A. (2000): Multiple splicing defects in an intronic false exon. *Mol. Cell. Biol* **20**, 6414–6425.

Tazi J., Bakkour N., Stamm S. (2009): Alternative splicing and disease. *Biochim Biophys Acta* **1792**, 14–26.

Todorov, N. A. (1988): *Livestock Prod. Sci.* 19:47.

Trapnell C., Pachter L., Salzberg S. L. (2009): TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-1111.

Trapnell C., Williams B. A., Pertea G., Mortazavi A., Kwan G., van Baren M. J., Salzberg S. L., Wold B. J., Pachter L. (2010): Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**, 511–515.

Valliyodan B., Nguyen H. T. (2006): Understanding regulatory networks and engineering for enhanced drought tolerance in plants. *Curr Opin Plant Biol* **9**, 189-195.

Villa T., Pleiss J. A., Guthrie C. (2002): Spliceosomal snRNAs: Mg2+-dependent chemistry at the catalytic core. *Cell* **109**, 149–152.

Wang E. T., Sandberg R., Luo S., Khrebtukova I., Zhang L., Mayr C., Kingsmore S. F., Schroth G. P., Burge C. B. (2008): Alternative isoform regulation in human tissuetranscriptomes. *Nature* **456**, 470–476.

Wang W., Qin Z., Feng Z., Wang X., Zhang X. (2013): Identifying differentially spliced genes from two groups of RNA-seq samples. *Gene* **518**, 164–170.

Wang X., Cairns M. J. (2014): SeqGSEA: a bioconductor package for gene set enrichment analysis of RNA-seq data integrating differential expression and splicing. *Bioinformatics* **30**, 1777–1779.

Wang Z., Gerstein M., Snyder M. (2009): RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10**, 57–63.

Wang Z., Xiao X., Nostrand E. V., Burge C. B. (2006): General and specific functions of exonic splicing silencers in splicing control. *Mol. Cell* **23**, 61–70.

Wicker T., Narechania A., Sabot F., Stein J., Vu G. T., Graner A., Ware D., Stein, N. (2008): Low-pass shotgun sequencing of the barley genome facilitates rapid identification of genes, conserved non-coding sequences and novel repeats. *BMC Genomics* **9**, 518.

Yitzhaki S., Miriami E., Sperling J., Sperling R. (1996): Phosphorylated Ser/Arg-rich proteins: Limiting factors in the assembly of 200S large nuclear ribonucleoprotein particles. *Proc. Natl. Acad. Sci. USA* **93**, 8830–8835.

7 PŘÍLOHY

7.1 Příloha 1 – Tabulka s výsledky programu TopHat2

Tab. 3: Výsledky optimalizace procesu mapování, hodnoty namapovaných „readů“ a mnohočetného přiřazení vyjádřeno v procentech.

Název vzorku	*	N = 2	N = 4	N = 6	N = 8	N = 10	N = 12	N = 14
CTRL_VAK_1	M	93,6	95,1	94,6	93,3	91,7	90,0	89,2
	MA	4,8	4,9	4,9	4,9	4,9	4,8	4,5
CTRL_VAK_2	M	93,5	95,0	94,6	93,3	91,6	89,8	89,0
	MA	4,5	4,6	4,7	4,7	4,7	4,6	4,4
CTRL_WT_1	M	93,5	95,0	94,6	93,2	91,5	89,6	88,8
	MA	4,3	4,5	4,5	4,5	4,5	4,4	4,1
CTRL_WT_2	M	93,7	95,5	95,0	93,5	91,7	89,8	88,9
	MA	3,2	3,4	3,5	3,5	3,4	3,3	3,1
R2T_VAK_2	M	91,8	93,7	93,5	92,5	91,2	89,7	89,1
	MA	6,5	6,7	6,7	6,7	6,7	6,6	6,5
R2T_VAK_1	M	93,5	95,0	94,6	93,1	91,4	89,5	88,7
	MA	3,5	3,7	3,7	3,7	3,7	3,5	3,3
R2T_WT_2	M	94,0	95,2	94,7	93,2	91,4	89,5	88,6
	MA	2,9	3,0	3,0	3,0	2,9	2,7	2,4
R2T_WT_1	M	93,9	95,2	94,6	93,1	91,3	89,3	88,5
	MA	3,0	3,1	3,1	3,1	3,1	2,9	2,6
R12H_VAK_2	M	93,4	94,8	94,3	93,0	91,2	89,4	88,6
	MA	5,0	5,1	5,1	5,2	5,2	5,0	4,9
R12H_VAK_1	M	94,2	95,7	95,0	93,4	91,5	89,5	88,6
	MA	3,2	3,4	3,4	3,43	3,3	3,1	2,9
R12H_WT_2	M	94,1	95,5	95,0	93,6	91,8	90,0	89,1
	MA	3,8	3,9	3,9	3,9	3,9	3,7	3,5
R12H_WT_1	M	94,4	95,6	95,0	93,4	91,6	89,6	88,7
	MA	3,7	3,8	3,8	3,8	3,8	3,6	3,3
S_VAK_1	M	94,5	95,6	95,0	93,5	91,5	89,5	88,7
	MA	3,2	3,2	3,3	3,3	3,2	3,0	2,7
S_VAK_2	M	94,6	95,8	95,2	93,7	91,8	89,9	98,1
	MA	2,8	2,9	2,9	2,9	2,8	2,6	2,3
S_WT_1	M	90,6	92,2	92,6	92,1	91,1	90,0	89,5
	MA	5,9	5,9	5,9	5,9	5,9	5,8	5,7
S_WT_2	M	93,9	95,2	94,7	93,2	91,3	89,3	88,5
	MA	4,1	4,2	4,2	4,2	4,2	4,0	3,8

*M – namapované „ready“, MA – „ready“ podléhající multimapingu

7.2 Příloha 2 – Tabulka s výsledky programu FeatureCounts

Tab. 4: Výsledky kvantifikace „readů“ pro parametr N = 4 s pomocí programu FeatureCounts.

Analyzované vzorky	Anotovaná oblast reference	Nejednoznačně namapované ready	Multimaping	Neanotovaná oblast reference
CTRL_VAK_1	52 532 917	1 864	9 744 434	15 529 778
CTRL_VAK_2	50 665 937	2 060	8 677 799	13 998 367
CTRL_WT_1	60 004 052	1 803	9 627 429	16 725 399
CTRL_WT_2	45 951 406	1 453	5 021 898	12 378 095
R2T_VAK_1	51 131 828	2 922	6 321 471	14 120 604
R2T_VAK_2	39 864 435	1 789	10 880 291	11 873 213
R2T_WT_1	48 775 764	3 309	5 083 590	14 384 051
R2T_WT_2	47 629 117	3 575	4 621 308	13 753 980
R12H_VAK_1	51 516 850	3 764	5 623 142	14 191 055
R12H_VAK_2	48 755 165	6 176	9 415 281	13 673 736
R12H_WT_1	47 506 915	3 280	6 302 958	13 148 715
R12H_WT_2	36 991 600	2 306	4 914 854	9 949 032
S_VAK_1	51 609 116	6 230	5 732 735	14 962 067
S_VAK_2	52 863 509	7 424	4 992 063	15 500 524
S_WT_1	49 213 813	2 408	12 831 949	18 282 881
S_WT_2	32 764 804	5 365	5 015 709	9 301 428

7.3 Příloha 3 – Tabulka s výsledky programu Cufflinks

Tab. 5: Tabulka s počty nově detekovaných izoforem jednotlivých analyzovaných vzorků získaných pomocí programu Cufflinks.

Analyzované vzorky	Počet nově detekovaných izoforem
CTRL_VAK_1	3837
CTRL_VAK_2	3443
CTRL_WT_1	3743
CTRL_WT_2	3238
R2T_VAK_1	3877
R2T_VAK_2	2535
R2T_WT_1	3594
R2T_WT_2	3558
R12H_VAK_1	3529
R12H_VAK_2	3235
R12H_WT_1	3478
R12H_WT_2	2296
S_VAK_1	4284
S_VAK_2	4801
S_WT_1	5248
S_WT_2	2836

7.4 Příloha 3 – Tabulka s výsledky programu StringTie

Tab. 6: Tabulka s počty nově detekovaných izoforem jednotlivých analyzovaných vzorků získaných pomocí programu Cufflinks.

Analyzované vzorky	Počet nově detekovaných izoforem
CTRL_VAK_1	115
CTRL_VAK_2	97
CTRL_WT_1	124
CTRL_WT_2	102
R2T_VAK_1	93
R2T_VAK_2	83
R2T_WT_1	101
R2T_WT_2	94
R12H_VAK_1	118
R12H_VAK_2	97
R12H_WT_1	95
R12H_WT_2	87
S_VAK_1	109
S_VAK_2	96
S_WT_1	130
S_WT_2	57