

UNIVERZITA KARLOVA V PRAZE

PŘÍRODOVĚDECKÁ FAKULTA

Studijní program: Genetika, molekulární biologie a virologie



Bc. Kalifa Samaké

## **Polymorfismus transkripčního faktoru NF- $\kappa$ B a Toll-like receptoru 2 u produkční populace skotu (*Bos taurus* L.)**

Polymorphism of the transcription factor NF- $\kappa$ B and Toll-like receptor 2 in a  
production population of cattle (*Bos taurus* L.)

DIPLOMOVÁ PRÁCE

Vedoucí práce: prom. biol. Karel Novák, CSc.

Praha 2018

### **PROHLÁŠENÍ**

Prohlašuji, že jsem tuto diplomovou práci zpracoval samostatně a že jsem uvedl všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze dne .....

.....  
Kalifa Samaké

## **PODĚKOVÁNÍ**

Na tomto místě bych rád poděkoval své školiteli prom. biol. Karlu Novákovi, CSc. za orientaci laboratoře, mnoho cenných rad a pomoc při sepisování této práce.

Děkuji také všem profesorům přírodovědecké fakulty Univerzity Karlovy za trpělivost, rady a mnoho konzultací, které jsem potřeboval během studia.

Děkuji všem kolegům z Výzkumného ústavu živočišné výroby za vytvoření příjemného pracovního prostředí a ochotu kdykoliv poradit či pomoci.

Děkuji také svým rodičům, UNESCO a České vládě za podporu během celé doby studia.

## **ABSTRAKT**

Širší účel práce je najít a interpretovat polymorfismus v genech přirozené imunity skotu využitelný pro zvýšení odolnosti proti infekčním chorobám. Byla aplikována metoda NGS na platformě PacBio pro efektivní resekvenování genu pro klíčový receptor přirozené imunity TLR2 a dvou genů kódujících složky navazujícího transkripčního faktoru NF- $\kappa$ B. V reprezentativní populaci 149 býků plemene ČESTR bylo nalezeno 22 polymorfismů v genu *NFKB1*, z toho 5 nových, zatímco v genu *NFKB2* bylo nalezeno 13 SNP, z toho 10 nových. V genu *TLR2* bylo nalezeno 21 SNP, z toho pouze 3 nové. Z 56 nalezených polymorfismů je 6 nesynonymních SNP, z toho jeden vede k záměně R474G v produktu NF- $\kappa$ B1 a pět k záměnám E63D, R152Q, I211V, R563H, H665Q v proteinu TLR2. Popis haplotypů v genech *NFKB1* a *NFKB2* usnadnil vývoj individuálních genotypovacích reakcí typu extenze primeru. U *TLR2* byl zjištěn vysoký počet haplotypů, a to jak ze sekvencí přečtených technologií PacBio, tak i nezávisle statistickou rekonstrukcí. Rovněž byly u *TLR2* určeny dva shluky haplotypů, což lze vysvětlit diverzifikující selekcí či genovou introgresí. Znalost diverzity v populaci ČESTR dovoluje provést asociační studie s fenotypovými údaji o zdravotním stavu. Lokalizace ve funkčních doménách dovoluje vymezit SNP s potenciálním účinkem na infekční rezistenci, což je zejména H665Q v TIR oblasti TLR2. Znalost diverzity otvírá možnost přenosu údajů o plemenné hodnotě jednotlivých variant z jiných studií.

## **KLÍČOVÁ SLOVA**

imunogenetika; NGS; skot; bioinformatika; genotypizace; zdravotní znaky

## **ABSTRACT**

The broader purpose of the work is to find and interpret polymorphism in the genes of natural immunity of cattle to be used to improve disease resistance. The NGS method on the PacBio platform was applied for the resequencing of the gene for the key receptor of innate immunity TLR2 and two genes coding for the components of the downstream transcriptional factor NF- $\kappa$ B. In the population of 149 bulls of the Czech Simmental breed, 22 polymorphisms were found in the gene *NFKB1* (5 new), while in the *NFKB2* gene 13 SNP were found (10 new). 21 SNP were found in the *TLR2* gene (3 new). Of the 56 found polymorphisms, 6 SNPs were nonsynonymous. One SNP leads to a change R474G in the *NFKB1* product and five to changes E63D, R152Q, I211V, R563H and H665Q in the protein TLR2. Knowledge of the haplotypes facilitated the development of individual genotyping reactions. In *TLR2*, a high number of haplotypes was detected, both from the PacBio reads and the statistical reconstruction. In addition, two clusters of haplotypes were distinguished in *TLR2*, possibly due to diversifying selection or introgression. The knowledge of genetic diversity in the population allows for the planned association studies with health data. Localization in functional domains allow to define the change with the greatest effect, in particular H665Q in the TLR2 TIR domain. The knowledge of diversity also allows to use the data on the value of individual variants from other studies.

## **KEY WORDS**

imunogenetika; NGS; skot; bioinformatika; genotypizace; zdravotní znaky

## OBSAH

1	ÚVOD.....	- 10 -
2	LITERÁRNÍ PŘEHLED .....	- 13 -
2.1	Bos taurus.....	- 13 -
2.2	Přirozená imunita .....	- 14 -
2.2.1	Úloha přirozené imunity .....	- 14 -
2.2.2	Receptory .....	- 14 -
2.2.3	Toll-like receptory .....	- 14 -
2.2.4	Regulace TLR .....	- 16 -
2.2.5	Toll signální dráha .....	- 17 -
2.2.6	Transkripční faktor NF- $\kappa$ B .....	- 20 -
2.2.7	Přirozená diverzita TLR a členů signální dráhy .....	- 22 -
2.3	NGS.....	- 23 -
3	CÍLE PRÁCE .....	- 26 -
4	MATERIÁL A METODIKA .....	- 27 -
4.1	Izolace DNA.....	- 27 -
4.2	Kontrola výsledku izolace DNA .....	- 28 -
4.3	Standardizace preparátů DNA.....	- 30 -
4.4	Amplifikace vybraných genů pomocí PCR a návrh amplifikačních primerů .....	- 30 -
4.5	Příprava vzorku pro NGS.....	- 39 -
4.6	Zpracování výsledků NGS .....	- 40 -
4.7	Přímé určení haplotypů .....	- 40 -
4.8	Návrh genotypovacích reakcí.....	- 41 -
4.9	Funkční charakterizace mutací a modelování produktu.....	- 44 -
5	VÝSLEDKY.....	- 45 -
5.1	Získání knihovny genomické DNA pro populaci ČESTR.....	- 45 -
5.2	Struktura bovinních genů <i>NFKB1</i> a <i>NFKB2</i> .....	- 45 -
5.3	Amplifikace exonů bovinních genů <i>NFKB1</i> a <i>NFKB2</i> .....	- 48 -
5.3.1	Optimalizace amplifikace <i>NFKB1</i> .....	- 48 -
5.3.2	Optimalizace amplifikace <i>NFKB2</i> .....	- 54 -
5.4	Preparativní amplifikace panelu TLR genů a genů <i>NFKB1</i> a <i>NFKB2</i> pro NGS.....	- 57 -
5.5	Zpracování výsledků NGS z technologie PacBio .....	- 59 -

5.6	Návrh a validace genotypovacích reakcí pro SNP v NFkB1 a NFkB2.....	- 68 -
5.6.1	Haplotypová struktura genu NFkB1 v populaci ČESTR.....	- 68 -
5.6.2	Návrh genotypovacích reakcí pro NFkB1 .....	- 69 -
5.6.3	Haplotypová struktura genu NFkB2 v populaci ČESTR.....	- 72 -
5.6.4	Návrh genotypovacích reakcí pro NFkB2 .....	- 72 -
5.6.5	Validace genotypovacích reakcí pro NFkB1 a NFkB2.....	- 75 -
5.6.6	Haplotypová struktura genu TLR2 v populaci ČESTR .....	- 77 -
5.7	Vyhodnocení populačních frekvencí.....	- 80 -
5.8	Kategorizace nalezených polymorfismů a predikované funkční dopady .....	- 86 -
6	DISKUZE.....	- 90 -
6.1	Úloha zvolených genů.....	- 90 -
6.2	Možnost přenosu metod .....	- 90 -
6.3	Kontext nalezeného polymorfismu .....	- 91 -
6.4	Haplotypová diverzita .....	- 92 -
6.5	Funkční dopady diverzity ve studovaných genech .....	- 93 -
6.6	Zdravotní dopady diverzity ve studovaných genech.....	- 94 -
7	ZÁVĚR.....	- 96 -
8	SEZNAM POUŽITÉ LITERATURY .....	- 97 -

## **SEZNAM ZKRATEK**

<b>BSE</b>	Bovinní spongiformní encefalopatie
<b>CDS</b>	Coding DNA Sequence (DNA kodující oblast)
<b>ČESTR</b>	Český strakatý skot (Czech Simmental or czech red cattle)
<b>DC</b>	Dendritic cell (dendritických buněk)
<b>DNA</b>	deoxyribonucleic acid
<b>DTT</b>	dithiothreitol
<b>EBI</b>	European Bioinformatics Institute, Cambridge, VB
<b>EDTA</b>	ethylenediaminetetraacetic acid (kyselinu ethylendiamintetraoctovou)
<b>FAO/OIE</b>	Food and Agriculture Organization/ Organisation for Animal Health
<b>gDNA</b>	genomic deoxyribonucleic acid
<b>IKK</b>	I $\kappa$ B [inhibitor of NF- $\kappa$ B (nuclear factor $\kappa$ B)] kinase
<b>IL-1R</b>	interleukine-1 receptor
<b>IRAKs</b>	IL-1R-associated kinases
<b>LPS</b>	lipopolysaccharide (lipopolysacharid)
<b>LRR</b>	leucine-rich repeats
<b>MyD88</b>	Myeloid differentiation primary response gene 88 (gen 88 primární odezvy myeloidní diferenciaci)
<b>NCBI</b>	National Center for Biotechnology Information
<b>NF-<math>\kappa</math>B</b>	jaderný faktor $\kappa$ B (nuclear factor $\kappa$ B)
<b>NGS</b>	Next-generation sequencing
<b>NO</b>	Nitric oxide (Oxid dusnatý)
<b>PAMPs</b>	pathogen-associated molecular patterns
<b>PCR</b>	Polymerase chain reaction (polymerázová řetězová reakce)
<b>PRR</b>	pattern recognition receptors
<b>RNA</b>	ribonucleic acid
<b>SNP</b>	Single nucleotide polymorphism
<b>TAB</b>	TAK-1-binding proteins
<b>TAK1</b>	TGF (transforming growth factor)- $\beta$ -activated kinase 1
<b>TBE</b>	Tris Borate EDTA
<b>TICAM1/TRIF</b>	TIR-domain-containing adaptor molecule 1/TIR-domain-containing adaptor inducing interferon $\beta$



<b>TIR</b>	Toll/IL-1R/R gene region
<b>TIRAP/Mal</b>	TIR-domain-containing adapter/MyD88 adaptor-like
<b>TLR</b>	Toll-like receptors
<b>TNF</b>	Tumor necrosis factor
<b>TRAF6</b>	TNF receptor associated factor 6
<b>TRAM</b>	TRIF-related adaptor molecule

## 1 ÚVOD

Preventivní povaha rezistenčního šlechtění více odpovídá současným trendům v chovu hospodářských zvířat než dodatečná terapie.

Svět zažívá rychlý růst výroby a spotřeby živočišných produktů a očekává se, že tento trend bude pokračovat. Tradiční systémy živočišné výroby rozhodujícím způsobem přispívají k živobytí 70 procent chudých venkovských obyvatel. Na druhé straně dochází k tomu, že velké podniky se sofistikovanými a mezinárodně provozovanými technologiemi reagují na rychlý růst poptávky na trhu s masem, mlékem a vejci. Živočišná výroba nyní zabírá třetinu orné půdy, která je nyní určena k produkci krmiv, a soutěží s jinými odvětvími o další půdní a vodní zdroje, energii a práci. Vývoj je navíc ovlivněn změnami klimatu a socioekonomickými tlaky.

V důsledku globalizace a změny klimatu dochází k bezprecedentnímu nárůstu chorob zvířat a zoonóz (chorob zvířat přenosných na člověka), nástupu nových chorob a jejich dalšímu šíření. Nemoci postihující hospodářská zvířata mohou mít devastující účinky na objem zemědělské produkce, produktivitu a na obchod s živými zvířaty, masem a jinými živočišnými produkty. Jsou možné i důsledky pro lidské zdraví, přičemž dopad může být celosvětový.

Mezi nedávnými krizemi způsobenými chorobami zvířat stačí připomenout epizootii BSE (bovinní spongiformní encefalopatie) u přežvýkavců (Donnelly et al. 1997) nebo rozšíření slintavky a kulhavky na britských ostrovech v r. 2001 (Woolhouse and Donaldson 2001). Preventivní opatření si v posledním případě vyžádaly utracení více než 10 milionů kusů domácích zvířat na britských ostrovech, především skotu. Jako ekonomicky významné hrozby je třeba uvést reprodukční a respirační syndrom prasat - PRRS (Miller et al., 2009) či aktuální novou verzi ptačí chřipky v r. 2017 (Adlhoch et al. 2018).

Dopady chorob hospodářských zvířat pro mezinárodní společenství iniciují uplatňování vhodných opatření a programů v oblasti zdraví zvířat s cílem chránit veřejné zdraví a bezpečnost potravin ([http://www.fao.org/ag/againfo/themes/fr/animal\\_health.html](http://www.fao.org/ag/againfo/themes/fr/animal_health.html)). Za významný úspěch je třeba považovat vymýcení virového původce dobytčího moru, které je výsledkem úsilí FAO z r. 2011 (FAO/OIE, 2011). Po viru černých neštovic se jedná teprve o druhý případ globální a trvalé eradikace patogena z přírodních ohnisek.

Předpokladem pro další účinný postup proti infekčním chorobám hospodářských druhů je poznání mechanismů patogenese a účinku jednotlivých genů, které kontrolují rezistenci k

těmto chorobám. „Chceme-li dobře porozumět funkci nějakého složitého systému, musíme samozřejmě poznat pokud možno detailně všechny jeho součásti a vztahy mezi nimi. Chceme-li opravit porouchané auto či televizi, musíme vědět, která z mnoha součástí mohla poruchu způsobit. Ve složitých biologických systémech prozatím většinu molekulárních komponent a jejich vzájemných funkčních vztahů neznáme, ale přesto se s větším či menším úspěchem pokoušíme jim porozumět.“ (Hořejší, 2004).

Studium imunitního systému hospodářských zvířat je neoddelitelné od hlavního proudu imunologie a imunogenetiky, který se zabývá imunitním systémem člověka, laboratorních modelů či evolučními aspekty. Řada poznatků získaných u modelových druhů nebo u člověka je přenositelná do imunologie veterinární a pomáhá pochopit, jak se odolnost hospodářských druhů formuje a jak se lze pokusit o kontrolu infekcí těchto zvířat.

Kromě adaptivní imunity hraje důležitou roli i vrozená imunita, která tvoří, společně s fyzikálními a chemickými bariérami, první linii obrany organismu. Na rozdíl od adaptivní imunity není v takové míře formována vývojem jedince, ale je více determinována genetickými faktory. Geny, které kódují jednotlivé části systému přirozené imunity, vykazují přirozenou variabilitu, stejně jako ostatní části genomu. Jednotlivé genové varianty mohou mít odlišné funkční vlastnosti. Předpokládá se, že zárukou funkce přirozené imunity na populační úrovni je její dostatečná přirozená variabilita. Proto polymorfismus genetických faktorů spojených s vrozenou (přirozenou) imunitou poskytuje reálnou možnost pro zlepšení zdraví zvířat.

Šlechtění hospodářských druhů zvířat zaměřené na infekční odolnost by mělo brát ohled na diverzitu těchto genů a cíleně využívat jejich účinnější varianty. Přirozená (vrozená) imunita využívá mj. receptory, které se podílejí na rozpoznávání bakteriálních, virových, houbových a parazitárních struktur.

Nečekaný a z pohledu imunologie pozdní objev Toll-like receptorů (TLR), představuje významný průlom v pochopení molekulárních interakcí mezi hostitelem a patogenem. Tato rodina receptorů, původně nalezená u hmyzu v r. 1985 (Stein et al., 1991), je známa prakticky u všech živočišných skupin a homologní geny jsou známy i u rostlin (Zipfel et al., 2004). TLR jsou transmembránové receptory exprimované vrozenými imunitními buňkami, jako jsou monocyty-makrofágy nebo dendritické buňky, spojené se signální dráhou, která vede k aktivaci nukleárního transkripčního faktoru NF- $\kappa$ B. TLR mohou rozpoznat vysoce konzervované molekulární struktury spojené s patogeny a pak iniciovat imunitní odpověď a systémovou zánětlivou odezvu (Huet et al., 2004).

Vrozená imunitní odpověď umožňuje hostiteli rozeznávat patogeny a spustit rychlou zánětlivou odezvu, včetně tvorby cytokinů a chemokinů, uvolnění efektorových molekul, jako je NO. Probíhá také interakce s adaptivní imunitní odpovědí. Porozumění molekulárním mechanismům vrozené imunity se zrychlilo v polovině devadesátých let 20. století, kdy bylo prokázáno, že protein Toll z banánových mušek *Drosophila* je kritický pro obranu much proti houbovým infekcím (Lemaitre et al., 1996). Toto pozorování otevřelo cestu pro následný popis podobných proteinů, nazývaných Toll-like (TLR), v buňkách savců i prakticky všech dalších živočišných taxonů (Du et al., 2000).

V tomto kontextu byla práce zaměřena na variabilitu genů pro tzv. Toll-like receptory a dvou genů kódujících klíčovou složku Toll signální dráhy, transkripční faktor NF- $\kappa$ B, u jednoho z hlavních plemen skotu (*Bos taurus* L).

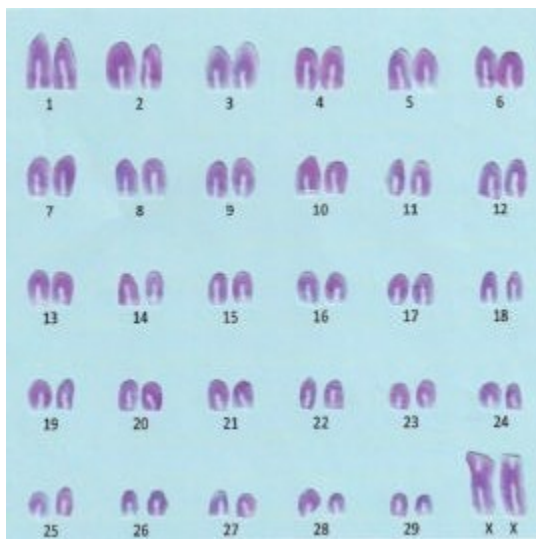
Pro dokumentaci variability u genů vybraných pro diplomovou práci byla použita sekvenační metoda Next Generation Sequencing (NGS) pro detekování polymorfismu, zejména jednonukleotidových záměn (SNP), a následně specifické genotypovací metody.

## 2 LITERÁRNÍ PŘEHLED

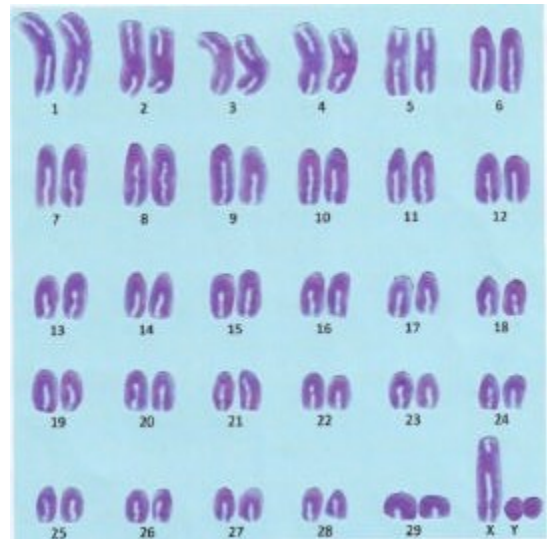
### 2.1 *Bos taurus*

*Bos taurus* (tur domácí, skot) je zemědělsky významným zvířetem. V chudších oblastech převážně agrárních zemí, například Africe, majetek jednotlivců anebo rodiny se odhaduje podle toho, kolik kusů skotu vlastní, protože tato zvířata hrají centrální roli v ekonomické aktivitě. Na druhé straně je faktem, že výroba hovězího masa a mléka skotu je překvapivě největším zpracovatelským oborem i v průmyslově rozvinutých zemích. Skot je i důležitým modelovým organismem pro zdravotní výzkum v oblasti obezity, reprodukčních chorob a infekčních nemocí. Skot se také používá ve studiích v oborech endokrinologie a fyziologie.

Genom skotu je užitečný pro srovnávací genomiku jako typický zástupce placentálních savců, kteří nepatří mezi primáty či hlodavce. Velikost genomu skotu se odhaduje na 3000 Mb, které jsou uspořádané do 29 párů autosomů a dvou pohlavních chromosomů. Genom obsahuje cca 30.000 genů, z toho asi 1000 genů má vztah k infekční odolnosti. Chromosomy skotu jsou akrocentrické (obr. 1).



Karyotyp samice skotu



Karyotyp samce skotu

Obr. 1. Karyotyp skotu podle Ashari et al. (2012)

## **2.2 Přirozená imunita**

### **2.2.1 Úloha přirozené imunity**

Obratlovci jsou neustále ohroženi invazí mikroorganismů a vyvinuli systémy imunitní obrany k odstranění infekčních patogenů v těle. Imunitní systém savců se skládá ze dvou větví: vrozené a získané imunity. Vrozená imunitní odpověď není zcela nespecifická, jak se původně myslelo, ale je schopna rozlišovat mezi vlastními a nevlastními molekulami. Vrozený imunitní systém je první linie hostitelské obrany proti patogenům a je zprostředkována fagocyty včetně makrofágů a dendritických buněk (DC).

### **2.2.2 Receptory**

Účinnost přirozené imunity záleží ve velké míře na patřičné iniciální rekognici patogenů a klíčová úloha receptorů rozpoznávajících struktury (pattern recognition receptors, PRRs) je dobře známá (Palsson-McDermott and O'Neill 2007). Vrozený imunitní systém rozpozná mikroorganismy prostřednictvím omezeného počtu receptorů. Představují základní rozdíl oproti velkému repertoáru rekombinantních receptorů, které využívá systém adaptivní imunity.

PRR mají některé společné charakteristiky.

Za prvé, PRR rozpoznávají mikrobiální složky označované jako molekulární vzory asociované s patogeny (PAMPs, "pathogen-associated molecular patterns"). Tyto struktury bývají nezbytné pro přežití mikroorganismu a patogenitu. Z tohoto důvodu je pro mikroorganismy obtížné je změnit a vyhnout se tak rozpoznání imunitním systémem.

Za druhé, PRRs jsou exprimovány v hostiteli a detekují patogeny bez ohledu na jejich fázi životního cyklu.

Za třetí, PRRs jsou kódovány již v zárodečné linii. Nejsou vázány na určité klony, jsou exprimovány ve všech buňkách daného typu a nezávisle na imunologické paměti. Různé PRR reagují s konkrétními PAMP, vykazují charakteristické vzorce exprese, aktivují specifické signální dráhy a vedou k různým protipatogenním odpovědím. Základní mechanismy vrozeného imunitního rozpoznávání jsou vysoce konzervované mezi skupinami eukaryot, od rostlin přes hmyz k savcům.

### **2.2.3 Toll-like receptory**

Důležitou skupinou PRRs jsou Toll-like receptory. Z celé skupiny PRR, Toll-like receptory (TLR) jsou nejvíce studovanou skupinou molekul, která je zodpovědná za detekci

molekulárních vzorů asociovaných s mikroorganismy (Akira et al., 2006). Tyto receptory se účastní jak spouštění přirozené imunity, tak i regulace imunity adaptivní (Iwasaki a Medzhitov, 2010).

Jednotlivé TLR rozpoznávají a reagují na různé mikrobiální molekuly a umožňují tak přirozenému imunitnímu systému rozlišovat mezi skupinami patogenů a indukovat vhodnou kaskádu efektorových odpovědí. TLR jsou klasifikovány jako členové IL-1 receptor (IL-1R) superrodiny na základě společné cytoplazmatické oblasti, která je známá jako Toll/IL-1R/R gene region (TIR) doména. Zkratka domény zahrnuje Toll-like receptory, receptor pro interleukin 1 (IL-1R) a rostlinné rezistenční (R) geny (Zipfel et al., 2004).

TLR jsou evolučně konzervovány od červa *Caenorhabditis elegans* až po savce (Kawai a Akira, 2010; Beutler, 2004). Gen Toll, zakládající člen rodiny TLR, byl původně identifikován jako genový produkt nezbytný pro vývoj embryonální dorsoventrální polarity u drozofily. Později bylo také prokázáno, že hraje důležitou roli v protiplísňové reakci much (Lemaitre et al., 1996).

Rodina Toll-like receptorů (TLR) je základní rozpoznávací a signalizační složkou hostitele v infekční obraně savců. Nejméně 12 TLR, které byly klonovány u savců, rozeznávají molekulární produkty získané ze všech hlavních tříd patogenů.

Rodina lidských TLR se skládá z deseti receptorů a příslušných genů, které jsou kriticky důležité pro vrozenou imunitu. Jednotlivé TLR rozpoznávají odlišný, ale omezený repertoár konzervovaných mikrobiálních produktů; například dobře charakterizované páry receptor-ligand zahrnují TLR4 a lipopolysacharid (LPS), TLR5 a flagelin, TLR1, TLR2 a TLR6 v kombinaci s lipoproteiny a TLR3, TLR7, TLR8 a TLR9 s různými motivy nukleových kyselin. Souhrnně, kompletní rodina TLR umožňuje hostiteli detekovat infekce způsobené většinou typů (ne-li všemi) mikrobiálních patogenů (Bell et al., 2003).

Zatímco v genomu drozofily bylo identifikováno devět členů genové rodiny proteinu Toll (Valanne et al., 2011), počet genů u obratlovců se pohybuje v rozmezí od 10 do 17. Konkrétně rodina *TLR* obsahuje *TLR1-10* skotu a *TLR1-9* a *11-13* u myši (Kaisho a Akira, 2006; Kawai a Akira, 2010). Pozice genů *TLR* byla mapována v genomu skotu pomocí radiačního hybridního mapování. McGuire a spoluautoři (2005) lokalizovali všech deset bovinních *TLR* genů na sedmi chromozomech. Pouze *TLR1*, *TLR6* a *TLR10* tvoří cluster na chromozómu 6 (Opsal et al., 2006), zatímco *TLR7* a *TLR8* jsou umístěny blízko sebe na pohlavním chromozómu X společně s pseudogenem *TLR*. Podobně byly zmapovány geny *TLR* prasat a ovcí (Jann et al., 2009).

TLR jsou integrální membránové glykoproteiny typu I. Mají charakteristické extracelulární domény obsahující různý počet opakování bohatých na leucin (leucine-rich repeats: LRR). Cytoplazmatické signalizační domény, které jsou homologní k interleukin 1 receptoru (Il-1R), byly pojmenovány Toll/Il-1R oblasti (TIR) (Bowie a O'Neill, 2000). LRR domény jsou složeny z 19 – 25 tandemových LRR motivů, z nichž každý obsahuje 24 – 29 aminokyselin. Ty tvoří motiv XLXXLXLXX, stejně jako ostatní konzervované aminokyselinové zbytky tvoří XØXXØXXXXFXLX kde Ø je hydrofobní zbytek. Každý LRR se skládá z b vlákna a šroubovice spojených smyčkami. Předpokládá se, že LRR doména TLR tvoří strukturu podkovy s vazbou ligandu na konkávním povrchu. Nicméně trojrozměrná struktura LRR motivů lidského TLR3 nabízí odlišný model. Je pravděpodobnější, že negativně nabitá dsRNA se váže na vnější (konvexní) povrch extracelulární části proteinu TLR3 (Choe et al., 2005).

TLR2 tvoří heterodimery s TLR1 nebo TLR6. Tato schopnost má za následek vznik samostatných fixačních kapes, které usnadňují rozpoznávání diacetyl- a triacetylipopeptidů z bakteriálních buněčných stěn. Vazba mezi TLR2 a lipopeptidy je silná, i když je relativně nespecifická. Tato interakce probíhá na úrovni LRR9 až LRR12, kde jsou hydrofobní rezidua, estery mastných kyselin a vnitřní kapsa TLR2. Tato korespondence je dostačující pro indukci intracelulárního signálu fixovanými diacetylipopeptidy. Rozpoznávání triacetylipopeptidu molekulou TLR2 a heterodimerizace vyžadují interakci mezi C-koncem acetylipopeptidu a hydrofobním kanálem PBA z TLR1. U TLR6 tento hydrofobní kanál blokují dva fenylalaniny, které zabraňují připojení třetího acetylovaného lipopeptidu (Schenk et al., 2009).

#### **2.2.4 Regulace TLR**

TLRs jsou exprimovány na různých imunitních buňkách včetně makrofágů, dendritických buněk (DC), B buněk, na specifických typech T buněk a dokonce i na neimunitních buňkách, jako jsou fibroblasty a epiteliální buňky. Exprese TLR není stacionární, ale spíše je modulovaná v rychlé reakci na patogeny různými cytokiny a vlivy vnějšího prostředí.

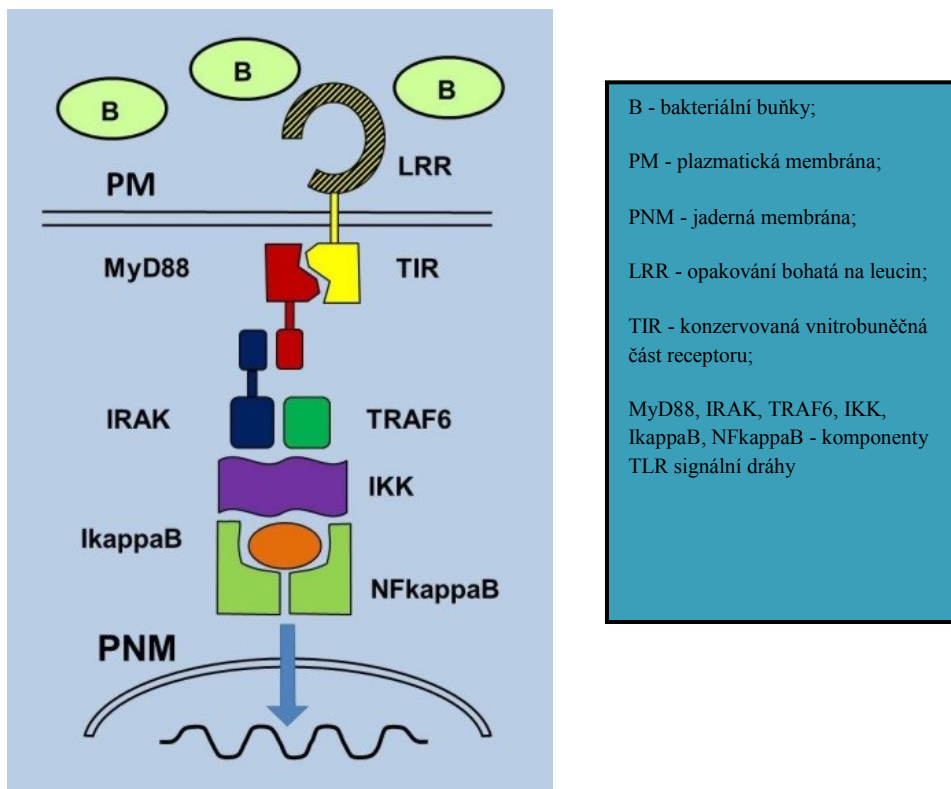
Kromě toho, TLR mohou být exprimovány extra-nebo intracelulárně. Zatímco některé TLR (TLR1, 2, 4, 5 a 6) jsou lokalizovány na povrchu buněk, jiné (TLR3, 7, 8 a 9) se nacházejí téměř výhradně v intracelulárních odděleních, jako jsou endosomy. Jejich ligandy, hlavně nukleové kyseliny, proto vyžadují transport do endosomů před signalizací.



Membrány a membránové proximální regiony jsou důležité pro buněčnou kompartmentalizaci těchto receptorů.

### 2.2.5 Toll signální dráha

Je třeba uvést, že proteiny NF- $\kappa$ B1 a NF- $\kappa$ B2 jsou zastoupen delší a kratší formou, která se může odštěpit při translaci. V případě NF- $\kappa$ B1 to jsou p50 a p105, v případě NF- $\kappa$ B2 p52 a p100. Zatímco těžší proteiny vykonávají úlohu specifických inhibitorů transkripce, menší proteiny se spojují s proteinem c-Rel, případně dalšími Rel proteiny a tvoří tak samotné transkripční faktory. Když heterodimery TLR2/TLR1 nebo TLR2/TLR6 rozeznají své agonisty, aktivují dvě signalizační dráhy, což vyústí v aktivaci heterodimeru P50/c-rel z jaderného faktoru  $\kappa$ B (NF- $\kappa$ B) (Kawai a Akira, 2010). První signalizační dráha od Toll receptorů je MyD88 závislá, zatímco druhá dráha vyžaduje povinně přítomnost Rac1, PI3-K a AKT (Arbibe et al., 2000).



Obr. 2. Zjednodušené schéma Toll signální dráhy. B - bakteriální buňka, PM - plazmatická membrána, PNM - jaderná membrána, LRR - opakování bohatá na leucin, TIR - konzervovaná vnitrobuněčná část receptoru, MyD88, IRAK, TRAF6, IKK, IκB - komponenty signální dráhy.

Variabilní extracelulární části TLR obsahují různé počty opakování bohatých na leucin (Kaisho a Akira, 2006). Po setkání s mikroorganismem TLR spouštějí komplexní kaskádu událostí, které vedou k indukci souboru genů zánětlivé odpovědi. Vazba ligandu vede k připojení ("recruitment") několika molekul k receptorovému komplexu. Tyto molekuly zahrnují adaptorové molekuly obsahující TIR doménu, jako jsou MyD88 (myeloid differentiation primary response gene 88), TIR-domain-containing adapter/MyD88 adaptor-like (TIRAP/Mal), TIR-domain-containing adaptor molecule 1/TIR-domain-containing adaptor-inducing interferon  $\beta$  (TICAM1/TRIF) a TRIF-related adaptor molecule (TRAM) (obr. 2, 3). Další zapojované molekuly zahrnují IRAK ("IL-1R-associated kinases IRAK1, 2, 3 (M) a 4") a také TRAF6 ("Tumor necrosis factor (TNF) receptor-associated factor 6").

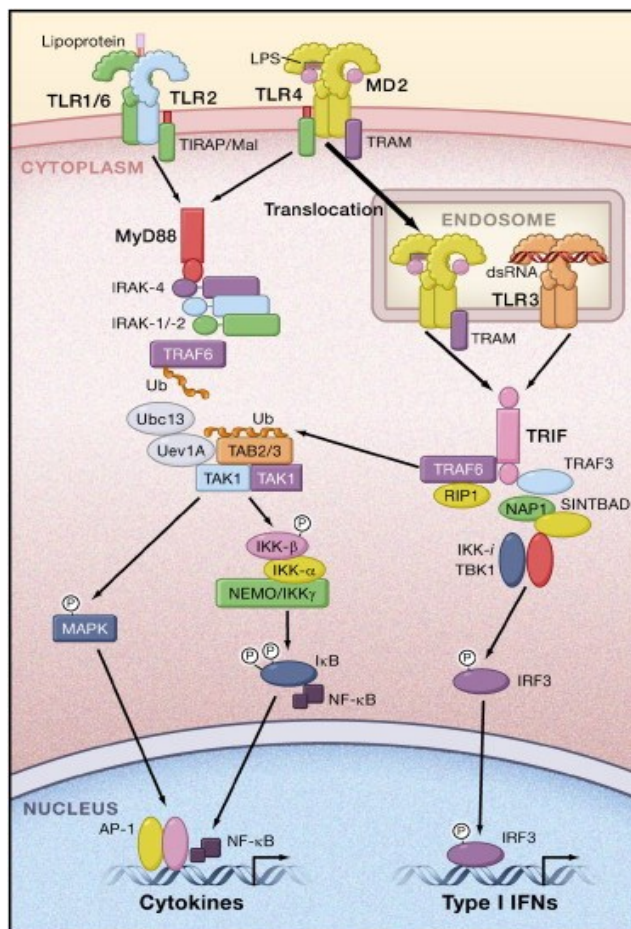
IRAK1 a TRAF6 poté disociují a vážou jiný komplex (obr. 3), který sestává z TAK1 ("TGF - transforming growth factor- $\beta$ -activated kinase 1") a TAB1, 2 a 3 ("TAK-1-binding proteins 1, 2 and 3"). TAK1 potom aktivuje I $\kappa$ B ("inhibitor of NF- $\kappa$ B (nuclear factor  $\kappa$ B) kinase"). Aktivita tohoto komplexu je regulována pomocí IKK $\gamma$  (rovněž známého jako NEMO podle "NF- $\kappa$ B essential modulator"). Fosforylace I $\kappa$ B zprostředkovaná IKK vede k jeho degradaci, která dovoluje NF- $\kappa$ B přemístit se do jádra a podnítit transkripci souboru zánětlivých genů včetně TNF, IL-1 $\beta$  a IL-6 (Misch a Hawn, 2008).

Toll signalizace směřující k transkripčnímu faktoru NF- $\kappa$ B využívá konzervovanou TIR doménu. Ta zprostředkuje interakci s TIR doménou-v molekule adaptéru MyD88 (myeloidní diferenciacní faktor 88). Tuto molekulu jako klíčový adaptér používají všechny TLRs kromě TLR3, který pro přenos signálu využívá adaptorovou molekulu TRIF (Werling a Jungi 2003). Vazba MyD88 na proximální TIR doménu aktivovaných TLRs umožňuje další interakci s příslušníky skupiny Il-1R asociovaných kináz (IRAK) a následnou aktivaci TNF ("tumour necrosis factor") receptoru 6. Tyto události následně vedou k aktivaci NF- $\kappa$ B prostřednictvím kinázového komplexu I $\kappa$ B  $\alpha - \beta - \gamma$  (složky jsou rovněž označovány jako IKK1, IKK2 a NEMO), což je klíčový regulátor celé dráhy.

TLR3 a TLR4 se zřejmě natolik evolučně odchýlily od dalších TLR, že mohou aktivovat genovou expresi a vyvolat antivirovou odpověď v mechanismu, který zahrnuje koordinovanou aktivaci jak NF- $\kappa$ B, tak i IRF-3. Příznakem MyD88 nezávislé signalizace je indukce dráhy zrání dendritických buněk a indukce interferonu typu 1 (IFN- $\beta$ ) (Takeuchi et Akira, 2010).

Další molekulou, která interaguje s adaptérem MyD88, je bílkovina Mal, označovaná také jako TIRAP (obr. 3). Je to bílkovina příbuzná MyD88, která byla objevena na základě její schopnosti zprostředkovat TLR4 signalizaci. Mal/TIRAP ovlivňoval LPS indukovanou

tvorbu IFN- $\beta$  in vitro, nicméně studie na myších s inaktivovaným Mal/TIRAP genem ukazuje, že Mal/TIRAP se podílí na MyD88-dependentní NF- $\kappa$ B aktivaci po stimulaci LPS a zapojení TLR2 v závislosti na ligandu.



Obr. 3. Signalizační dráhy TLR2, TLR3 a TLR4 (Takeuchi a Akira, 2010)

Třetí adaptérový protein (kromě MyD88 a TIRAP) pro Toll-like receptory je TRIF, tj. TIR doménu obsahující adaptér indukující IFN- $\beta$ . Alternativně je označován jako TICAM-1 (TIR obsahující adaptérová molekula. TRIF komunikuje s TLR3 a zprostředkovává TLR3-závislou indukci IFN- $\beta$  přes společnou NF- $\kappa$ B a IRF-3 aktivaci.

Exprese IRF-3 vede k indukci IFN- $\beta$  a ISG-54/56. IRF-3 je aktivován po fosforylaci na shluku specifických COOH-terminálních serinových zbytků, což usnadňuje jeho dimerizaci a interakci s koaktivátory CBP a P300. Aktivovaný komplex IRF-3 se pak přemístí do jádra, kde reguluje přepis cílových genů.

IRF-7 je související regulátor transkripce, který je exprimován převážně v lymfoidních buňkách a je nezbytný pro IFN- $\alpha$  genové exprese. Transkripce IRF-7 je vyvolána IFN a

postranslačně aktivována fosforylací serinových zbytků na COOH konci. Některé z těchto zbytků jsou konzervovány a shodné s IRF-3. IKK $\epsilon$  a TBK-1 (TANK vázící kináza) jsou klíčovými regulátory IRF-3 a IRF-7 při aktivaci drah v těch buňkách, které byly vystaveny některým virům a/nebo aktivovány volnou dsRNA prostřednictvím TLR3. IKK $\epsilon$  a TBK1 jsou také potřebné komponenty TRIF signalizační dráhy, která vede k aktivaci IRF-3.

Úloha IRF3 v LPS-indukované genové expresi IFN- $\beta$  a v endotoxinovém šoku byla demonstrována ve studii na myších s knockoutovaným genem. Nicméně molekulární mechanismy, které regulují odezvu na LPS nezávislou na MyD88, a to přes TLR4 dráhu k IRF-3 a k NF- $\kappa$ B aktivaci nejsou známy dostatečně.

Čtvrtá adaptérová molekula, rovněž obsahující TIR doménu, byla pojmenována TRAM (TRIF-související s adaptérem molekuly). TRAM, stejně jako všechny adaptérové molekuly obsahující TIR domény, aktivuje NF- $\kappa$ B. Kromě toho TRAM, stejně jako TRIF, aktivuje IRF-3 a IRF-7. Na rozdíl od dalších známých adaptérových molekul obsahujících TIR doménu se TRAM zdá být omezen na dráhu TLR4, která signalizuje podnět LPS. Na druhou stranu TRIF hraje roli v obou drahách, jak TLR3, tak i TLR4, které vedou k IRF řízené genové expresi. Tyto vlastnosti naznačují, že adaptér TRAM se mohl vyvinout za účelem zprostředkování TLR4-specifických signálů. Ty potom vedou k profilu genové exprese, který je odlišný od profilu indukovaného TLR3.

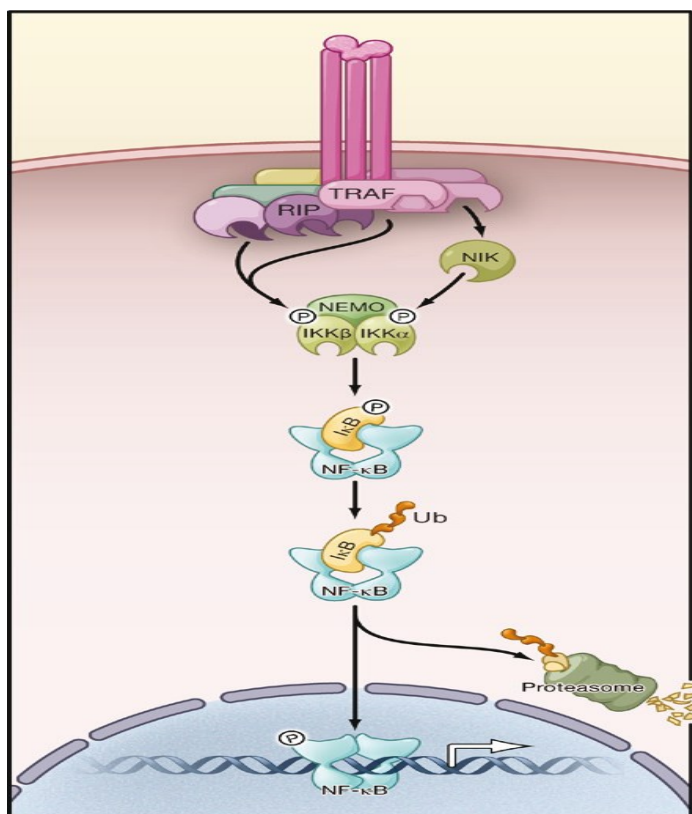
### **2.2.6 Transkripční faktor NF- $\kappa$ B**

Transkripční faktor NF- $\kappa$ B slouží jako typický příklad indukovatelných transkripčních faktorů více než 20 let. Početné podněty, které aktivují NF- $\kappa$ B, a velký počet jím regulovaných genů (400 podle [www.nf-kB.org](http://www.nf-kB.org)) jsou příčinou, že tento transkripční faktor je stále předmětem intenzivního výzkumu (Hayden a Ghosh, 2008).

Signální dráha NF- $\kappa$ B hraje důležitou úlohu v imunitních zánětlivých odpovědích. Nedávno bylo demonstrováno, že se také účastní antiapoptické odpovědi, a to jak v průběhu normální diferenciaci (zejména v hematopoetickém systému), tak i v kancerogenezi.

Proteiny NF- $\kappa$ B jsou transkripční aktivátory, které jsou přítomné v buňce ve formě dimerů. U savců je známo pět členů této proteinové rodiny: p50, p52, c-rel, p65 (rovněž označovaný relA) a relB. Komplexy transaktivátorů, které se vyskytují nejčastěji, jsou heterodimery p50/p65 (které odpovídají původně popsané aktivitě NF- $\kappa$ B), ale existují i početné další kombinace. Podjednotky p50 a p512 jsou syntetizované ve formě cytoplazmatických prekurzorů, a to p105 a p100.

Signalizace k NF- $\kappa$ B sestává ze sady pozitivních a negativních regulačních prvků (obr. 4). Ve své nejzákladnější formě, signální dráha se skládá z receptorů a z adaptorových molekul: komplex IKK s proteiny I $\kappa$ B a dimery NF- $\kappa$ B. Indukce podnětem spouští aktivaci IKK, zahrnující rovněž fosforylaci, ubikvitinylaci a degradaci proteinů I $\kappa$ B (obr. 5). Uvolněné dimery NF- $\kappa$ B jsou potom aktivovány prostřednictvím postranlačních modifikací a jsou přeneseny do jádra, kde se vážou na specifické sekvence a indukují transkripci cílových genů.



Obr. 4. NF- $\kappa$ B signální dráha. Jsou znázorněny kroky po vazbě ligandu na receptor, zapojení proximálních adaptorových proteinů receptoru, a větvení signalizace na IKK a TRAF/RIP komplexy či přes TRAFs a NIK (podle Hayden a Ghosh, 2008).

V nepřítomnosti specifických extracelulárních signálů jsou dimery NF- $\kappa$ B zadržovány v cytoplazmě v neaktivní formě díky jejich interakci s molekulami z rodiny I $\kappa$ B. Tato rodina proteinů zahrnuje pět členů: I $\kappa$ B $\alpha$ , I $\kappa$ B $\beta$ , I $\kappa$ B $\epsilon$ , a prekurzory p105 a p100, které tudíž mají dvojitou funkci. Následkem různých podnětů (jako cytokiny včetně TNF $\alpha$ , interleukin-1, bakteriální či virové proteiny a různé stresové signály) jsou molekuly I $\kappa$ B fosforylovány kinázovým komplexem. Ten se skládá ze tří jednotek: dvou katalytických podjednotek s aktivitou serin/threoninové kinázy (IKK $\alpha$  a IKK $\beta$ ) a z jedné regulační podjednotky NEMO/IKK $\gamma$ . Tyto kroky fosforylace vedou k ubikvitinylaci I $\kappa$ B a následně k jejich

degradaci proteasomem 26S. V důsledku toho je umožněn transport dimerů NF- $\kappa$ B do jádra, kde se váží na promotory svých cílových genů.

Existuje významné množství strukturních informací o NF- $\kappa$ B jak v neaktivní formě vázané na I $\kappa$ B, tak i v aktivním stavu vazby s DNA. Struktury NF- $\kappa$ B ve vazbě na místa  $\kappa$ B dovolily pochopit, jak Ig podobné domény, které zahrnují doménu RHD, interagují s DNA. N-terminální část Ig podobných domén vnáší selektivitu pro určité typy míst  $\kappa$ B, zatímco hydrofobní zbytky přítomné na C. terminální doméně poskytují rozhraní pro dimerizaci subjednotek NF- $\kappa$ B.

V neaktivním stavu jsou dimery NF- $\kappa$ B spojeny s jedním ze tří proteinů I $\kappa$ B: I $\kappa$ B $\alpha$  (NF-KBIA), I $\kappa$ B $\beta$  (NF-KBIB) nebo I $\kappa$ B $\epsilon$  (NF-KBIE), nebo s prekurzory proteinů p:100 (NF-KB2) či p105 (NF-KB1). Tyto proteiny I $\kappa$ Bs udržují dimery NF- $\kappa$ B v cytoplazmě a mají klíčovou roli v indukcí odpovědi. Jsou známy dva indukovatelné proteiny, I $\kappa$ B, Bcl-3 (BCL3) a I $\kappa$ B $\zeta$  (NF-KBZ), jejichž funkce spočívá v inaktivaci NF- $\kappa$ B. Všechny proteiny I $\kappa$ B se vyznačují přítomností mnoha ankyrinových domén.

Nejvíce studovaným členem rodiny je I $\kappa$ B $\alpha$ . Je rychle degradován během aktivace signální dráhy NF- $\kappa$ B, což vede k uvolnění početných dimerů NF- $\kappa$ B. Navíc je heterodimer p65/p50 jedním z prvních cílů I $\kappa$ B $\alpha$ . Aktuální model spočívá na tom, že I $\kappa$ B $\alpha$  udržuje dimery NF- $\kappa$ B v cytoplazmě a zabraňuje tak jejich translokaci do jádra a vazbě na DNA.

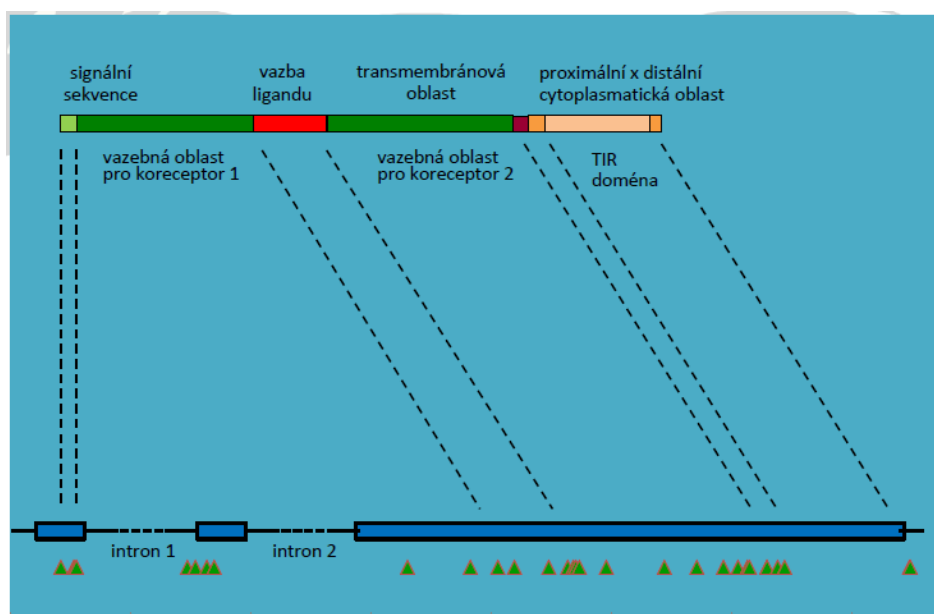
Degradace I $\kappa$ B indukuje rychle kroky signalizace, které jsou iniciované specifickou fosforylací těchto molekul aktivovaným IKK. Komplex IKK obsahuje dvě subjednotky silně homologických kináz, IKK $\alpha$ /IKK1 (CHUK) a IKK $\beta$ /IKK2 (IKBKB), a také regulační podjednotku NEMO (NF- $\kappa$ B essential modulator)/IKK  $\gamma$  (IKNKG). Obecně se nacházejí ve formě heterodimerního kinázového komplexu (Hayden a Ghosh, 2008).

### **2.2.7 Přirozená diverzita TLR a členů signální dráhy**

V případech těch domestikovaných druhů, kde byla struktura TLR alespoň částečně charakterizována, byla využita výhoda plynoucí z dobré homologie k myším a lidským genům a z přítomnosti pouze jednoho paralogu. To usnadňuje návrh in silico specifických oligonukleotidových primerů pro polymerázovou řetězovou reakci (PCR) a identifikaci produktu. Například u skotu byly pro sérii receptorů rozpoznávajících bakterie všechny exony sledovány pomocí 30 PCR produktů (White et al., 2003, Seabury et al., 2007, 2010) a přibližně stejný počet byl potřeba pro tzv. antivirovou sérii (Cargill a Womack, 2007). Charakteristická distribuce polymorfismů se zpravidla vyhýbá vysoce konzervované TIR

oblasti (obr. 5). Navíc se jedná o oblast interagující s několika adaptérovými molekulami, takže případné strukturální odchylky jsou rychle eliminovány.

Řada prací se zabývala diverzitou bovinního *TLR2*, zejména ve vztahu k infekční imunitě (např. Bhaladhare et al., 2016; Bilgen et al., 2015; Iann et al., 2008; Juste et al., 2018; Mucha et al., 2009). Na druhou stranu pro *NFKB1* a *NFKB2* prakticky neexistují, kromě databázových údajů z plošných resekvenčních projektů, publikace, které by se jejich diverzitou zabývaly. V práci Chen et al. (2015) byly charakterizovány dva SNP v *NFKB1* a jeden v *NFKB2*.



Obr. 5. Gen *TLR4* a jeho produkt s vyznačenou distribucí mutací (podle Novák, 2014)

### 2.3 NGS

Experimentální postup se v rámci diplomové práce opíral o moderní a efektivní metodu sekvenování nové generace (NGS). Protože se jedná o metodu, která je klíčová nejenom v rámci diplomové práce, ale i v celé moderní biologii, je její charakterizace zařazena do literárního přehledu.

Od prvního sekvenování v roce 1970 (Ray Wu), Sangerovo sekvanování (1974), Allan Maxam Walter Gilbert (1977) bylo vynaloženo obrovské lidské, finanční a technologické úsilí, aby bylo možné získat nástroje pro dnešní účinnější a automatizované sekvenační metody označované jako NGS ("next generation sequencing"). Časné pokusy o sekvenaci DNA byly těžkopádné. V roce 1968 Wu oznámil použití primerů extenzní metody k určení 12 bazí kohezních konců bakteriofága lambda. V roce 1973 Gilbert a Maxam oznámili 24

báze vazebného místa laktózového represoru. Sekvenování trvalo dva roky rychlostí jedna báze měsíčně.

Obor byl změněn v roce 1976 vývojem dvou metod, které mohly dekodovat stovky bazí za několik hodin. Obě metody: řetězový terminační postup vyvinutý Sangerem a Coulsonem, a postup chemického štěpení vyvinutý Maxamem a Gilbertem použité vzdálenosti podél molekuly DNA od radioaktivní značky na pozici obsazenou příslušnou bazí k určení pořadí nukleotidů. Sangerova metoda zahrnovala čtyři extenze značeného primeru DNA polymerázou, každou se stopovým množstvím terminačního nukleotidu ukončujícího růst řetězce, k tvorbě fragmentů různých délek. Gilbertova metoda vzala terminálně značené DNA fragmenty a ve čtyřech různých reakcích použila chemikálie k báze-specifickému částečnému štěpení. U obou metod byly velikosti fragmentů přítomných v jednotlivých specifických reakcích měřeny elektroforézou na polyakrylamidovém gelu, která umožňovala separaci fragmentů DNA podle velikosti s rozlišením jedné báze. Gely autoradiograficky tvořily obraz žebříčku, ze kterého mohla být sekvence okamžitě přečtena.

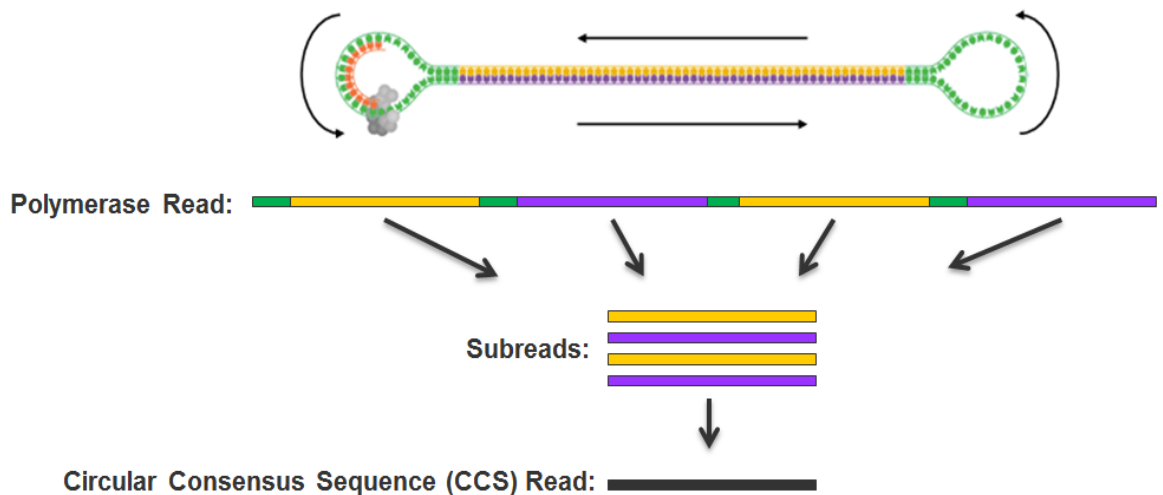
Tyto metody se dostaly do okamžitého použití. Náhodné ("shotgun") sekvenování klonů a následné složení celkové sekvence na základě překrývání bylo navrženo Stadenem v r. 1979. Klonovací vektory na bázi bakteriofága M13 kolem r. 1980 pomohly skládat genomy de novo, jako genom bakteriofága lambda již v 1982. Do roku 1987 již automatizované a na fluorescenci-založené sekvenační přístroje podle Sangera, které vyvinuli Smith a Hood a aplikoval Applied Biosystems, mohly generovat kolem 1.000 bazí za den. Sekvenační data rostla exponenciálně, přibližně podle Moorova zákona. Motivující bylo vytvoření centrálních datových úložišť jako je GenBank, kde je možné hledat prostřednictvím vyhledávacích nástrojů jako je BLAST. To vedlo ke zvýšení hodnoty každé sekvence a vyvolalo ducha sdílení údajů. V roce 1982 bylo v GenBank uloženo více než půl milionu bazí; v roce 1986 téměř 10 000 000 bazí (statistiky GenBank a WGS; Shendure et al., 2017).

Poptávka po rychlých, levných a přesných datech sekvenování DNA vedla k zrození a dominanci nové generace technologií sekvenování. Takzvaná technologie sekvenování "příští generace" umožňuje rychlou generaci dat tím, že souběžně používá různorodé metodiky, které překonávají omezení metod sekvenování podle Sangera používaných k sekvenování prvních lidských genomů. Rozvoj NGS vedl k otevření hranic pro další výzkum genomiky. Přinesl také zásadní odklon od Sangerova sekvenování. Rychlý vývoj sekvenačních technologií navíc zaplavil trh komerčně dostupnými sekvenačními



platformami, jejichž unikátní chemie a rozmanité aplikace stojí jako další překážka omezující potenciál NGS.

V práci byla jako základní metoda sekvenování použita metoda CCS ("circular consensus sequencing") firmy Pacific Biosciences. (Menlo Park, CA, USA). Technologie a možnosti jsou popsány v práci Rhoads a Au (2015) a znázorněny na obr. 6.



Obr. 6. Metoda sekvenování CCS firmy Pacific Biosciences (<https://www.pacb.com/>)

Postup zahrnuje přípravu knihovny typu SMRTbell ze vzorku DNA obohaceného pro sledované sekvence, sekvenování na systému RSII či novějším Sequel, a následně bioinformatické zpracování, tj. filtrování sekvencí (readů) podle kvality, mapování na referenční sekvenci, hledání strukturních variant, anotace variant a identifikace v databázi nukleových kyselin.

Metoda CCS byla použita na čtení produktů PCR v rámci postupu cíleného resekvenování ("targeted sequencing") vybraných genů s významem pro přirozenou imunitu.

Základní výhodou technologie PacBio jsou dlouhé přečtené sekvence. V poslední verzi se jedná až o 20 kb, nicméně praktický limit je 5 kb. Dlouhé sekvence přečtené z jednoho vlákna dovolují určovat přímo haplotypy v rozmezí jednotlivých fragmentů. Na druhou stranu nevýhodou metody PacBio je vysoká chybovost a nízké pokrytí. V práci padla volba na technologii PacBio, neboť znalost haplotypů, v nichž jsou mutace v populaci uspořádány, je důležitá pro šlechtitele. Navíc znalost haplotypů v řadě případů umožňuje snížit počet nezbytných genotypovacích reakcí při určování genotypu jednotlivých zvířat.

### 3 CÍLE PRÁCE

Vytvořit knihovnu genomické DNA ze sledované populace skotu jako východisko pro další postup.

Připravit postup pro resekvenování vybraných genů přirozené imunity skotu (*NFKB1*, *NFKB2* a *TLR2*).

Dokumentovat úplnou diverzitu ve vybraných imunitních genech pro sledovanou populaci skotu pomocí resekvenování.

Interpretovat nalezené genové varianty a nalézt jejich uspořádání do haplotypů na základě výsledků NGS.

Navrhnout a validovat genotypovací reakce pro klíčové varianty.

Nezávisle zjistit distribuci genových variant ve sledované populaci pomocí individuálních genotypovacích reakcí.

Interpretovat výsledky ve vztahu k infekční odolnosti skotu.

## 4 MATERIÁL A METODIKA

### 4.1 Izolace DNA

Použité vzorky DNA byly připraveny ze zmražených inseminačních dávek, případně z konzervovaných krevních vzorků. Pro komerční použití a kryokonzervaci je sperma ředěno tzv. ředidlem tvořícím izotonické a kryoprotektivní prostředí. Inseminační dávka v kapiláře (pejetě) byla rychle rozmrazena při 37 a dále práce probíhala na ledu. Kapilára byla orientována vertikálně, zatavený konec byl odstřižen a obsah byl přenesen do mikrozkušavky 1,5 mL. Alikvóta o objemu 50  $\mu$ L byl centrifugována ve zkumavce 1,5 mL při 8000 ot/min ve stolní centrifuze při 4 po dobu 5 minut. Sediment byl 1x promyt 100  $\mu$ L fosfátového fyziologického roztoku (PBS) s opakovanou centrifugací. Lýze byla prováděna po přidání 225  $\mu$ L lyzačního pufru ze soupravy MagSep (Eppendorf) pro tkáň. Použití většího výchozího objemu inseminačních dávek (až 100  $\mu$ L) vedlo k vysoké viskozitě konečného preparátu DNA, což komplikovalo automatické pipetování. Protože při programovaném pipetování na stanici EpMotion chybí vizuální kontrola, změna vlastností preparátu vedla k chybám v přenášených objemech a zvyšovala riziko kontaminace sousedních vzorků (podle rozmístění na stojánku).

K výchozímu objemu vzorku byla také přidána proteináza K ze soupravy MagSep v objemu 25  $\mu$ L a 22  $\mu$ L 1 M dithiothreitolu (DTT) pro konečnou koncentraci 80 mM. Po počátečním resuspendování lyzační směsi vortexováním po dobu 1 min byla směs inkubována při 56°C po dobu nejméně 4 hodiny. Optimální lýze bylo dosaženo při inkubaci přes noc (12 h), případně 16 h s občasným vortexováním po dobu několika vteřin.

Lyzát byl centrifugován při 11.000 g a při pokojové teplotě po dobu 5 min. 225  $\mu$ L supernatantu bylo přeneseno do 2 mL LoBind mikrozkušavek s kulatým dnem (Eppendorf) a použito jako výchozí vzorek pro izolační program MagSep - Tissue na stanici EpMotion. Podstatou metody je, po lýzi buněk v lyzačním pufru v přítomnosti proteázy K, oddělení DNA z roztoku na afinitních feromagnetických partikulích. Po třikrát opakovaném promývání DNA vázané na částicích je získán čistý preparát pomocí 30 - 100  $\mu$ L elučního pufru. Jednotlivé pipetovací, homogenizační a separační kroky na přístroji jsou plně automatizované a tudíž vysoce reprodukovatelné. Výtěžek izolace se pohyboval kolem 20  $\mu$ g v celkovém objemu eluátu 100  $\mu$ L.

Zásobní 1 M roztok dithiothreitolu (DTT) byl připraven rozpuštěním 1,545 g DTT v 10 mL H<sub>2</sub>O. Roztok byl zfiltrován přes předběžně navlhčený sterilní nitrocelulóзовý filtr s póry 0,2  $\mu$ M pomocí injekční stříkačky a skladován v alikvótách při -20°C.

Při izolaci DNA z krve se postup v prvních krocích lišil. Krev z terénních odběrů o objemu přibližně 1,0 mL byla skladována v mikrozkuvkách o objemu 1,5 mL při -20°C. V řadě případů byla krev sražená, proto byl zvolen postup, který dovoluje získat dostatečné množství použitelné DNA i z takto znehodnocené krve. Vzorky byly zcentrifugovány při 2500 rpm po dobu 2 – 3 minut za účelem odstranění kondenzátu a stop krve z víčka před otevřením a jako prevence vzájemné kontaminace vzorků. Následně bylo do vzorku přidáno 20 µl Na<sub>2</sub>EDTA o koncentraci 10 % jako antikoagulans.

Dále byl vzorek homogenizován vortexováním s inertním materiálem. Pro homogenizaci sražené býčí krve byla testována skleněná balotina (jak jemná, tak i hrubá) a křemičitý písek (Lachema, Brno). Přidávané materiály byly autoklávované, aby nedošlo ke kontaminaci cizí DNA. Jako abrazivum se nejlépe osvědčil křemičitý písek.

Vzorek s EDTA a pískem byl rozmražen na vodní lázni a dále práce probíhala na ledu. Vzorky byly homogenizovány na vortexu nejdříve zběžně a potom důkladně po dobu nejméně 1 minuty. Písek a nečistoty byly odstraněny centrifugací při otáčkách 1500 – 2000 rpm po dobu 2 minut při 4°C. Supernatant byl odebrán a 200 µl bylo použito pro izolaci DNA metodou MagSep - Blood na automatické pipetovací stanici EpMotion. Výtěžek izolace se v tomto případě pohyboval kolem 1 µg v celkovém objemu 50 µL eluátu. I když se jedná asi o pětinový výtěžek v porovnání s výtěžkem z nesražené krve, uvedené množství postačuje pro provedení cca 50 následných genotypovacích reakcí.

Preparáty DNA byly dlouhodobě skladovány při -20°C, krátkodobě (po dobu několika týdnů) při 4°C.

#### **4.2 Kontrola výsledku izolace DNA**

Pro ověření výtěžku a kvality preparátů DNA byla použita metoda měření optické absorbance v UV oblasti, fluorescenční kvantifikace a gelová elektroforéza.

Pro určení koncentrace a čistoty preparátu DNA byl vzorek DNA naředěn 50 - 100 x vodou do objemu nad 50 µL, přenesen do mikrokvyvety a absorbance byla stanovena při třech vlnových délkách v UV oblasti na přístroji Biophotometer (Eppendorf, Hamburk, SRN). Koncentrace DNA v ng/µL byla určena vynásobením absorbance při 260 nm faktorem 50, kontaminace proteiny byla kontrolována podle absorbance při 280 nm a příměs rozptylujících nečistot při 230 nm. Poměr OD260/OD280 se optimálně pohyboval nad 1,8.

Výhodou fluorescenční metody pro stanovení koncentrace DNA je vyšší citlivost a detekce pouze samotné DNA, očištěné od interference volných nukleotidů a proteinů. Preparát byl ředěn 100 x roztokem fluorescenčního barviva SybrGreen (Biotium, Fremont, USA), které bylo přidáno v poměru 1 : 2000 z dodávaného zásobního roztoku do 10 mM Tris pufru, pH = 8,3, s 1 mM EDTA. Kalibrační řada obsahovala ředění komerčního standardu molekulových vah DNA fragmentů GeneRuler 100 bp plus (Fermentas, Vilnius, Litva). Po promíchání a inkubaci 15 min v mikrotitrační destičce byl exponován snímek fluorescence v jednotlivých jamkách s použitím excitace horním modrým světlem s maximem při  $\lambda = 470$  nm a emisního filtru určeného pro fluorescenci ethidium bromidu. Po vyrovnání hodnot pixelů v oblasti jamek destičky v programu ImageJ 1.45s (<http://imagej.nih.gov/ij>) byla hodnota pixelů ve středu jamky se vzorkem odečtena a přepočítána na koncentraci DNA na základě kalibrační křivky. Metoda umožňuje určovat koncentrace DNA od 0,1 ng/uL.

Degradace izolované genomické DNA s hmotností fragmentů kolem 20 kb byla testována pomocí elektroforézy v agarózovém gelu. Na 100 mL 0,6 % agarózového gelu (agaróza pro molekulární biologii Sigma-Aldrich, St. Louis, MO, USA) v pufru TBE bylo přidáno 10  $\mu$ L dodávaného zásobního roztoku fluorescenčního barviva GelRed (Biotium, Fremont, CA, USA). Na přípravu jednoho gelu pro vysokorychlostní horizontální elektroforézu Liberty (Biokeystone Company, El Monte, CA, USA) bylo použito cca 120 mL agarózy. Po solidifikaci gelu byly elektrodové komory zaplněny předchlazeným pracovním pufrům TBE, prostor nad gelem byl zaplněn vodou o teplotě 4°C a tloušťce vrstvy 3 cm. Uspořádání omezuje přehřívání a dovoluje vyšší gradient napětí (celkově 200 V po dobu 30 min). 2  $\mu$ L alikvóty DNA byly aplikovány v 6  $\mu$ L 5 x ředěného nanášecího pufru (Fermentas). Jako standard byla použita směs restričních fragmentů fága lambda po štěpení restriktázou HindIII s fragmenty 23130, 9416, 6557, 4361, 2322, 2027, 564 a 125 bp anebo žebříček standardů GeneRuler 100bp+ (Fermentas). Gely byly dokumentovány fluorescenční kamerou Gbox Chemi XR Bio (Syngene, Frederick, US). Barvivo GelRed dovoluje použít emisní filtr pro ethidium bromid, excitační vlnová délka je variabilní od UV po zelenou oblast.

### **4.3 Standardizace preparátů DNA**

Protože vzorky DNA byly na výstupu z automatu EpMotion uloženy v mikrocentrifugačních zkumavkách, bylo nezbytné provést přeformátování na 96 jamkové mikrotitrační destičky, které umožňují další vysokokapacitní zpracování. Koncentrace byla v knihovně genomické DNA nastavena na 20 ng/ $\mu$ L; to zajišťuje reprodukovatelnost reakcí.

Spektrofotometricky určené koncentrace DNA byly použity pro výpočet objemu vody a DNA vzorku pro ředění. Data byla importována do přístroje EpMotion ve formátu CSV. Byl vytvořen původní program, který zajistil a) převedení kolekce vzorků z formátu mikrozkuvek na mikrotitrační destičku o 96 jamkách, b) normalizaci koncentrací, c) současné vytvoření záložní repliky destičky s knihovnou gDNA.

### **4.4 Amplifikace vybraných genů pomocí PCR a návrh amplifikačních primerů**

Amplifikační primery pro PCR (tab. 1) byly v případě TLR genů převzaty z literatury (White et al., 2003; Seabury et al., 2007). Jejich použití bylo dodatečně optimalizováno (Novák et al., 2017, Novák a Czerneková, 2018a, b). Použité amplifikační reakce plně pokrývají většinu kódujících oblastí genů TLR.

Tab. 1. Amplifikační primery pro všechny bovinní TLR geny Konec fragmentu

Gen	Referenční sekvence podle GenBank	Označení fragmentu	Začátek fragmentu	Konec fragmentu	Délka produktu (párů bází)	Přímý primer - označení	Přímý primer - sekvence 5'→3'	Zpětný primer- označení	Zpětný primer – sekvence 5'→3'	Vazebné teploty v PCR (°C)
Antibakteriální série										
<i>TLR1</i>	FJ147090	1_1	196	1361	1166	1_1F	ATGCCTGACATCCTCTCAC T	1_1R	AGAACCTTGATCTGAGGAG GT	62/60
<i>TLR1</i>	FJ147090	1_2	992	2186	1195	1_2F	TGACCCAGGAAATGAAGTC T	1_2R	CCGTGTTAATGTATTTCTG CTG	62/60
<i>TLR2</i>	EU746465	2_1	1	816	816	2_1F	TCCTGCTCCATATTCCTAC G	2_1R	TGACTGTGTTTGACATCAT GG	62/60
<i>TLR2</i>	EU746465	2_2	556	1223	668	2_2F	CTCATTCATTTATGGCTGG C	2_2R	GACCTGAACCAGGAGGAT G	62/60
<i>TLR2</i>	EU746465	2_3	911	1726	816	2_3F	CGGAAGGAGCCTCTGACCA GGCT	2_3R	CATGGGTACAGTCATCAAA CTC	62/60

<i>TLR2</i>	EU746465	2_4	2206	2935	730	2_4F	AGTTTAACCCAGTGCCTTC C	2_4R	TGGAGTCAATGATGTTGTC G	62/60	
<i>TLR2</i>	EU746465	2_5	2813	3248	436	2_5F	CCTACTGGGTGGAGAACCT C	2_5R	ACCACCAGACCAAGACTG AC	62/60	
<i>TLR4</i>	AC000135	.1	4_1	-3	657	661	4_1F	CCAGGGTATTTTGTTATGG CTGGAACAT	4_1R	TGTTTGCAAATGAACCTAA CCA	62/60
<i>TLR4</i>	AC000135	.1	4_2	4999	5382	384	4_2F	TCTTTGCTCGTCCCAGTAG C	4_2R	AAGTGAATGAAAAGGAGA CCTCA	62/60
<i>TLR4</i>	AC000135	.1	4_3	7941	9154	1214	4_3F	GGAGACCTAGATGACTGG GTTG	4_3R	AAGACAATGCGGATGTTGG T	62/60
<i>TLR4</i>	AC000135	.1	4_4	8924	9596	673	4_4F	TTTCAAGGGGTGCTGTTCT C	4_4R	TGCACACATCATTGCTCA G	64/62
<i>TLR4</i>	AC000135	.1	4_5	9684	#####	737	4_5F	GTCACTGTGCTCCTGGTGT C	4_5R	CTATAGGGCTCGCGTACCA C	64/62
<i>TLR5</i>	EU006635	5_1	-3	638	642	5_1F	TTTGGGAAACGGAGGATA AG	5_1R	GCACCTTTGAGGCTGTGA GCACCTTTGAGGCTGTGA	62/60	
<i>TLR5</i>	EU006635	5_2	553	1241	689	5_2F	GCCTGCTTTTGATACTTTG G	5_2R	AGGTGTCCGCTATGTTCTC A	62/60	
<i>TLR5</i>	EU006635	5_3	1495	2036	542	5_3F	GGCAGATTAGAGGGGAAA GA	5_3R	CCATCAAAGAAGCAGGAA GA	58/56	
<i>TLR5</i>	EU004663	5	5_4	3623	4195	573	5_4F	GAAACCAGCTCCTCTCTCC T	5_4R	ATCTTTCTGCTGCTCCACA C	62/60
<i>TLR5</i>	EU006635	5_5	4536	5299	764	5_5F	GGAGCAGTTTCCACTTATC G	5_5R	ATTCTCATGCCGGTTTCTTT ATTCTCATGCCGGTTTCTTT	58/56	



<i>TLR6</i>	AJ618974	6_1	774	1573	800	6_1F	ACTACCCATTGCTCACTTG C	6_2R	CTATACTCCCAACCCAAGA GC	62/60
<i>TLR6</i>	AJ618974	6_2	1300	2144	845	6_2F	GACACACGCTTTATACACA TGC	6_3R	CACTGACACACCATCCTGA G	62/60
<i>Antivirová série</i>										
<i>TLR3</i>	AJ812027	T3/1	576	1165	590	T3/1F	ATTAGGGACTGAGGGTTTT T	T3/1R	CTCTGTTCTCCTTGTTGAAT G	66/64
<i>TLR3</i>	AJ812027	T3/2	1106	1785	680	T3/2F	GGTCTTTATTTGATCTGGT GCT	T3/2R	GTTTGGGACATGGCGTCT	66/64
<i>TLR3</i>	AJ812027	T3/3	1668		650	T3/3F	TGTTAGTGTGTGTCTGCGT AAA	T3/3R	GGGTGAGTGAAAAAGGAA AA	58/56
<i>TLR3</i>	AJ812027	T3/4	from30 53		600	T3/4F	AGCCATCAGGATCTCTTTC C	T3/4R	CTCAGCGACTTCACTTTC C	58/56
<i>TLR3</i>	AJ812027	T3/5			250	T3/5F	GTCGCCATTTCTTCTCC	T3/5R	CCACAAACTCTCCCCTTCC	58/56
<i>TLR3</i>	AJ812027	T3/6		2591	700	T3/6F	ATTGGAGGCAGGTTCTTCA C	T3/6R	ATCTCATTGTGTTGGAGGT TC	66/64
<i>TLR3</i>	AJ812027	T3/7	2543		700	T3/7F	GTCAAAGTCTCCCTTGGTT G	T3/7R	GCTGACAAGAAAAAGGTG GT	66/64
<i>TLR3</i>	AJ812027	T3/8			700	T3/8F	GGGATGAAAAAGTGTCGA GT	T3/8R	GTCTGTGCTTTGGGATGTT T	66/64
<i>TLR3</i>	AJ812027	T3/9		3587	600	T3/9F	TCAAAAGTAGCACGAAAT GG	T3/9R	AGTCTTGGCATCAAAAATG G	66/64
<i>TLR3</i>	AJ812027	T3/10	3518	4044	527	T3/10F	TGCTATTTTGCTGTCCAGTT	T3/10R	GGACCCTCCACTTCTTTTGT	66/64
<i>TLR3</i>	AJ812027	T3/11	3933		600	T3/11F	CCTTCACACATACTGCTTT	T3/11R	TCCCGATACTCTTCTTCTTG	64/62

							GG		G	
<i>TLR3</i>	AJ812027	T3/12			650	T3/12F	CCTATAACGGAGTAAACCT AACCT	T3/12R	CTGTGTAAAACCACGATAA GCA	66/64
<i>TLR3</i>	AJ812027	T3/13			600	T3/13F	AGTTTCAGGTGATTAGCAA AGG	T3/13R	CTCAATCTTTCCAGCATC A	58/56
<i>TLR3</i>	AJ812027	T3/14			600	T3/14F	CTGTGCTGTATTGCTTCTCT G	T3/14R	GTTCCATCTGTCTTCTGCT CT	66/64
<i>TLR3</i>	AJ812027	T3/15		4382	620	T3/15F	TTCCTTCTCTCCTGCCTTCT	T3/15R	CACTACTCAGCACCCACAT C	66/64
<i>TLR3</i>	AJ812027	T3/16		5014	750	T3/16F	GTCTGGGAGATCAGGGAA G	T3/16R	CCACTGAAAGGAAAAATC GT	72/70
<i>TLR3</i>	AJ812027	T3/17	4902	5537	636	T3/17F	GCTCTTTTTATGGGCTTTCC	T3/17R	CCTTCAGCAACTCGTCATT T	64/62
<i>TLR3</i>	AJ812027	T3/18	5428	6004	577	T3/18F	CCTGAAAAATGTGGACTGC T	T3/18R	GTATTGGGGCGGAGTGTT	66/64
<i>TLR3</i>	AJ812027	T3/19	5947	6536	590	T3/19F	CCACACCAACATCTCTGAA C	T3/19R	AAACTGGACACAGCCAAA TC	66/64
<i>TLR3</i>	AJ812027	T3/20	6433		550	T3/20F	CAAAAGGTAGGTGAACAC TATGAC	T3/20R	ATATGGGACGGGCAGTTT	66/64
<i>TLR3</i>	AJ812027	T3/21		7149	600	T3/21F	GTAGCCATTCCCTTCTCCA	T3/21R	CAGCCCAACGCTCTAAAAT C	58/56
<i>TLR3</i>	AJ812027	T3/22	7082	7590	509	T3/22F	ACCTGGGTTTTAGTGACAA G		GCCTCAGTAACCTGTGTGT TC	66/64
<i>TLR3</i>	AJ812027	T3/23	7404	7673	400	T3/23F	GTCCAGAAATTCAGCACAT	T3/23R	AGGTGTACGTTTTACCCTT	64/62

							T		TCA	
<i>TLR7</i>	DQ333225	T7/1		275	700	T7/1F	CCCAATGTGTAGGGAAAAT G	T7/1R	CACAGGGCAGAGTTTTAGG A	64/62
<i>TLR7</i>	DQ333225	T7/2	164	1014	851	T7/2F	TTTCAGGTGTTTCCAATGT G	T7/2R	GGATCAATCTGTAGGGGAG AAT	66/64
<i>TLR7</i>	DQ333225	T7/3	939	1830	892	T7/3F	GAAATTGCCCTCGTTGTT	T7/3R	AGCCGATTGTTAGAGAAGT CC	66/64
<i>TLR7</i>	DQ333225	T7/4	1743	2564	822	T7/4F	CAGGAAATAGCATTAGCCA GA	T7/4R	TAACCCACCAGACAAACCA C	66/64
<i>TLR7</i>	DQ333225	T7/5	2471	3341	871	T7/5F	CCAGAAAACGTCCTCAACA A	T7/5R	AGTCACATTCGGCAAAGAA G	66/64
<i>TLR7</i>	DQ333225	T7/6	3217		750	T7/6F	AACAAACCCACAGGCTCAC	T7/6R	CAGGAGAGAAAGAGCAAG GA	66/64
<i>TLR8</i>	GU936183	T8/1		193	194	T8/1F	GCGTTTCCTTGAGTTATGC T	T8/1R	CTTCCGTCACATCTTTGTCC	66/64
<i>TLR8</i>	GU936183	T8/2	130	788	659	T8/2F	GCAGAATGTAATGGTCGTC G	T8/2R	CAAGGTACACAGGGAAAT GG	66/64
<i>TLR8</i>	GU936183	T8/3	739	1339	601	T8/3F	AGTGGAAACTGCCCGAGA	T8/3R	GCTTCAGGATGTGACTTTG G	66/64
<i>TLR8</i>	GU936183	T8/4	1263	1854	592	T8/4F	CAGAATATCACCTTGGTC AG	T8/4R	AGCATTCCACAGAAGGTCA A	66/64
<i>TLR8</i>	GU936183	T8/5	1697	2321	625	T8/5F	TAACGCACCGTCTAGGATT T	T8/5R	TCTCCGAAGTCACAGGTAC AG	66/64
<i>TLR8</i>	GU936183	T8/6	2271	2796	526	T8/6F	TGTTTTGGAAGTACTAGGGGGT	T8/6R	CTTGCTTTGGTTGATGCTCT	58/56

							AA			
<i>TLR8</i>	GU936183	T8/7	2688		650	T8/7F	CCTGGAAGAGAGTGAGGA CA	T8/7R	GGCTCTGAAGTGGATGCTA A	66/64
<i>TLR9</i>	AJ509824	T9/1		120	1100	T9/1F	GTTTGTGCTCTGATGGTGC T	T9/1R	CCCCTCCTCTTTCTACTCC	66/64
<i>TLR9</i>	AJ509824	T9/2	12	425	414	T9/2F	CTTCACCTCTCCCCAGACT T	T9/2R	CCGTGTTTCTCTCCATCACT	58/56
<i>TLR9</i>	AJ509824	T9/3	237	802	566	T9/3F	TTCTCACTTCTCTGATCTC T	T9/3R	TGCCTAGCTCTTTCATGCTC	58/56
<i>TLR9</i>	AJ509824	T9/4	731	1236	506	T9/4F	ATTCGTTCTGACCCACAGC A	T9/4R	TTCTCTCTCCAGTGCCCATC	66/64
<i>TLR9</i>	AJ509824	T9/5	1121	1619	499	T9/5F	AGATTGCAGGTCTCAGGAT G	T9/5R	ACAGGTGGACGAAGTCAG AG	64/62
<i>TLR9</i>	AJ509824	T9/6	1557	2274	718	T9/6F	CCAGCCTCTCCTTAATCTC C	T9/6R	CGGAACCAATCTTTCTCTA GTT	58/56
<i>TLR9</i>	AJ509824	T9/7	2189	2839	651	T9/7F	CCTGACACCTTCAGTCACC T	T9/7R	GCGGGTAAACATCTCTTGC T	64/62
<i>TLR9</i>	AJ509824	T9/8	2740	3285	546	T9/8F	CGTCAGCTCAAAGGACTTC A	T9/8R	AGGGTGTGCAGATGGTTCT C	66/64
<i>TLR9</i>	AJ509824	T9/9	3203	3632	430	T9/9F	GGGAGACCTCTATCTCTGC TTT	T9/9R	CGCTCACGTCTAGGATTTT C	64/62
<i>TLR9</i>	AJ509824	T9/10	3582	4419	838	T9/10F	CCTCCTGGTTCGGTTCCTTA	T9/10R	CGGTTATAGAAGTGACGGT TG	66/64
<i>TLRI</i>	EU006653	T10/1		787	813	T10/1F	CTGAGGTGAACCAGTGATA	T10/1R	ATCGTCCCAGGATAAGTCA	66/64

0							AAA		A	
TLRI	EU006653	T10/2	564	1331	768	T10/2F	TGCCCATCTTAAACACAAC	T10/2R	ACCCAAAAACAGAATCAG	64/62
0							A		CA	
TLRI	EU006653	T10/3	1194	1816	623	T10/3F	CCAGCAACACATCCCTGA	T10/3R	AAAGTGGAGGCAGCAGAA	66/64
0									G	
TLRI	EU006653	T10/4	1760	2414	655	T10/4F	ATTGTGGTTGTCATGCTCG	T10/4R	AACCTCCAAACCCTTCATT	66/64
0							T		C	
TLRI	EU006653	T10/5	2347	3094	748	T10/5F	TTTATTAGACACCAGAGGG	T10/5R	GCGGATTCTTTGTGATTGA	66/64
0							ACA		G	
TLRI	EU006653	T10/6	3035		868	T10/6F	TATTGTTGGCTGCACTGAG	T10/6R	AGACGTGTGTTCTGGGAAA	66/64
0							A		G	

Původní primery byly navrženy pro geny *NFKB1* a *NFKB2*. Návrhu předcházelo určení hranic exonů na základě porovnání sekvencí gDNA a mRNA pomocí algoritmu CLUSTALW v programu BioEdit (Larkin et al., 2007). Algoritmus spolehlivě určoval rozsah intronů při penalizaci 3 pro mezeru a extenzi mezery.

Primery byly navrhovány tak, aby PCR fragmenty odpovídaly jednotlivým exonům. Kompatibilita dvojic primerů byla zjišťována on line výpočtem na serveru Primer3web (primer3.ut.ee) a OligoAnalyzer 3.1 (www.idtdna.com/calc/analyze). Primery byly testovány na blízkost % GC u dvojic primerů a na shodu teploty denaturace ( $T_m$ ). Byla prověřována možnost tvorby vlásenkových struktur, tvorby vlastních dimerů a možnost hybridizace s reversním primerem.

Odsolené a lyofilizované primery syntetizované na zakázku (Eurofins Genomics, Ebersberg, SRN) byly podle výtěžku syntézy rozpuštěny na výchozí koncentraci 100  $\mu\text{M}$  v  $\text{H}_2\text{O}$ , následně ředěny 10 x na pracovní koncentraci 10  $\mu\text{M}$  a skladovány ve formátu 96 jamkových mikrotitračních destiček.

Pro amplifikaci fragmentů určených ke genotypování a pro kapilární sekvenování byla použita standardní termorezistentní Taq polymeráza od výrobce Fermentas. Pro přípravu fragmentů určených pro vysokokapacitní sekvenování nové generace byla použita nízkochybová polymeráza pod obchodním označením Phusion od výrobce Finnzymes, součásti Thermo Scientific (Waltham, MA, USA) či od New England Biolabs (Ipswich, MA, USA).

Standardní amplifikační reakce o objemu 10  $\mu\text{L}$  obsahovala 0,03 jednotek/ $\mu\text{L}$  Taq polymerázy v 10 mM Tris pufru, pH 8,8, obsahujícím 50 mM KCl, 0,08% Nonidet, 2 mM  $\text{MgCl}_2$ , čtyři deoxynukleotidtrifosfáty (dNTP) v koncentraci 0,2 mM, oba primery jednotlivě v koncentraci 0,4  $\mu\text{M}$  a templátovou gDNA v koncentraci 0,5 ng/ $\mu\text{L}$ . Sekvence amplifikačních primerů jsou uvedeny v tab. 1.

Preparativní reakce pro NGS v objemu 20  $\mu\text{L}$  a více obsahovala Phusion "high-fidelity" DNA polymerázu v koncentraci 0,1 jednotky/ $\mu\text{L}$  v prostředí firemního HF pufru, který obsahoval 1,5 mM  $\text{MgCl}_2$ , 0,2 mM dNTP, každý primer v koncentraci 0,5  $\mu\text{M}$  (tab. 1) a 0,1 ng/ $\mu\text{L}$  templátové DNA. Amplifikační reakce připravené v jamkách mikrotitračních destiček byly vzhledem k nízkým objemům překryty minerálním olejem (Sigma-Aldrich) a následně uzavřeny PCR fólií.

Amplifikační program byl adaptován podle White et al. (2003) a obsahoval počáteční denaturační krok při 95 v trvání 1 minuty, 5 cyklů při 95 po dobu 30 s, anelační krok při

teplotě  $T_a$  podle tab. 1 po dobu 30 s a 72°C po dobu 2 minut. Dále program pokračoval 43 identickými cykly s anelační teplotou sníženou o 2°C. Závěrečný extenzní krok 6 minut při 72°C zajišťuje dokončení syntézy komplementárních vláken DNA.

Produkty PCR byly skladovány při 4°C. Elektroforetická kontrola očekávané hmotnosti produktů byla prováděna elektroforeticky v 1,6% agarózovém gelu v 1  $\mu$ L vzorku.

#### 4.5 Příprava vzorku pro NGS

Vzorky gDNA býků z normalizované knihovny o koncentraci 20 ng/ $\mu$ L byly spojeny do směsného vzorku populace v ekvimolárním množství. Amplikony studovaných genů byly připraveny s DNA polymerázou Phusion v preparativních reakcích. Dále byla určena koncentrace produktů jednotlivých PCR reakcí podle intenzity fluorescence produktu v gelové elektroforéze. Po kontrole a kvantifikaci PCR produktů byl připraven směsný vzorek druhé úrovně z produktů pro jednotlivé amplikony. Ve vzorku byly zastoupeny v ekvimolární koncentraci, neboť objem byl určen podle výtěžku jednotlivých reakcí a hmotnosti fragmentu.

Takto připravený směsný vzorek byl purifikován na kolonkách NucleoSpin (Macherey-Nagel, Düren, SRN), a to na kolonkách typu PCR Clean-up Gel Extraction. 540  $\mu$ L vzorku (včetně minerálního oleje) bylo smícháno se dvěma objemy pufru NTI a směs byla nanesena na silikátovou membránu kolonky. Kolonka byla centrifugována 30 s při 11 000 g, poté následovalo promytí 700  $\mu$ L pufru NT3 s opakováním centrifugace. Opakované promývání pufrům NT3 minimalizuje kontaminaci chaotropními solemi, která často interferuje s přípravou knihovny pro PacBio sekvenování. Afinitní membrána v kolonce byla zbavena stop ethanolu opakovanou centrifugací po dobu 1 minuty při 11 000 g při umístění na nové zkumavce. DNA byla eluována 30 - 50  $\mu$ L pufru NE po inkubaci při pokojové teplotě po dobu 1 minuty. Eluát byl uvolněn centrifugací při 11 000 g po dobu 1 minuty.

Úspěšnost přečištění vzorku byla určena podle absorbance v UV oblasti. Vzorek, který splňoval kritéria servisní laboratoře, což je množství nejméně 1200 ng, poměr  $OD_{260}/OD_{280} > 1,90$  a  $OD_{260}/OD_{230} > 1,70$ , byl odeslán ve zmrazeném stavu při -20°C k dalšímu postupu v servisní laboratoři GATC (Konstanz, SRN). Příprava knihoven pro NGS byla provedena týmem servisní laboratoře. Vzorky byly přečteny metodou CCS ("consensus circular sequencing") na přístroji PacBio RSII (Pacific Biosciences, Menlo Park, CA, USA).

#### 4.6 Zpracování výsledků NGS

Výsledky sekvenace pomocí technologie PacBio byly poskytnuty ze servisní laboratoře GATC jako výsledky tří běhů přístroje. Protože pracoviště nemá technickou možnost použít software, který zpracovává primární sekvenační data z platformy Pacific Biosciences ve formátu h5, navazující práce vycházela ze souborů ve formátu FastQ a FASTA, které jsou rovněž součástí poskytovaného výstupu. Mapování jednotlivých přečtených úseků, tzv. readů, bylo prováděno na příslušných referenčních sekvencích pro jednotlivé geny. Byla použita místní instalace programu Geneious v. 11 (Biomatters, Auckland, Nový Zéland). Používán byl zpravidla vlastní porovnávací algoritmus Geneious.

Strukturní varianty byly vyhledávány pomocí implementovaného algoritmu v programu Geneious. Varianty byly filtrovány podle frekvence výskytu v populaci  $f \geq 1\%$  a podle pravděpodobností chybného určení  $P \leq 0,01$ . Nalezené strukturní varianty byly exportovány ve formátu CSV. Seznam nalezených odchylek byl porovnán se seznamem již známých SNP v dostupných databázích dbSNP v National Center for Biotechnology Information, Bethesda, MD, USA (NCBI) a Variant Browser v EBI (European Bioinformatics Institute, Cambridge, VB), případně Ensembl (Wellcome Trust Sanger Institute, Hinxton, VB) a s údaji v publikacích. Pro další vyhodnocení a návrh genotypovacích reakcí byly použity v první řadě polymorfismy, jejichž existence byla potvrzena nezávisle ze dvou zdrojů.

Polymorfismy získané vlastním sekvenováním byly považovány za reálné až po validaci pomocí jiné technologické platformy NGS nebo specifickou genotypovací reakcí, případně na základě shody s dříve publikovanými polymorfismy nebo s údaji v databázích.

#### 4.7 Přímé určení haplotypů

Současný výskyt více strukturních variant na jednom vlákně v cis konfiguraci a v těsné genetické vazbě, označované jako haplotypy, je možný rekonstruovat statisticky z populačních dat o výskytu polymorfismů u jednotlivých zvířat. Na druhou stranu technologie Pacific Biosciences dovoluje sekvenovat dlouhé úseky DNA, teoreticky až 20.000 bazí, takže lze výsledky získané pro amplikony s více polymorfismy použít pro přímé určení haplotypů. Amplikony použité v práci byly navrhovány s ohledem na možnost využití pro kapilární sekvenování, proto obsahují od 500 do 1200 párů bazí. Přesto se v řadě případů polymorfismy vyskytují společně v rámci jednoho amplikonu a je tak možné přímo určit jejich uspořádání v haplotypech.

Pro určení haplotypů v populaci byla vytvořena matice polymorfismů, v praxi převážně SNP. Soubor s výsledky mapování readů na referenční sekvenci byl z formátu BAM



exportován do formátu FASTA. Úseky mezi nalezenými SNP byly deletovány, takže zbyla matice nukleotidů, jejíž sloupce odpovídaly nukleotidům v pozici jednotlivých polymorfismů a řádky přiřazeným přečteným sekvencím (readům). Instrukce pro abecední seřazení řádků v programu BioEdit dovolila číst přímo jednotlivé haplotypy.

Na základě znalosti haplotypů bylo provedeno určení diagnostických SNP (tzv. tagging SNP či tagSNP) v rozsahu jednotlivých amplikonů. Genotypování samotných diagnostických SNP dovoluje určit i přítomnost všech SNP, které jsou s nimi v jednom haplotypu. Tímto způsobem lze snížit počet nezbytných genotypovacích reakcí a finanční náklady pro určení genotypu jednotlivých zvířat.

#### **4.8 Návrh genotypovacích reakcí**

Soubory nalezených SNP byly částečně redukovány na základě určení tagSNP. Pro redukováné soubory SNP byly navrženy genotypovací reakce typu extenze primeru (Hoogendoorn et al., 1999). Byla použita komerční varianta označovaná jako reakce SNaPshot (divize Life Technologies v Thermo Scientific). Reakční směs obsahuje oligonukleotid (SNaPshot primer), který 3' koncem přiléhá k variabilnímu místu na templátové DNA, směs fluorescenčně značených dideoxynukleotidtrifosfátů a termostabilní DNA polymerázu. Komplementární značený dideoxynukleotid se inkorporuje na 3' konec primeru. Protože působí jako terminátor polymerázové reakce, k inkorporaci dalších značených dideoxynukleotidů již nedochází. Fluorescenčně značené produkty reakce jsou následně určeny na kapilárním sekvenátoru s pětikanálovou fluorescenční detekcí.

Navrhované SNaPshot primery obsahovaly 21 - 27 nukleotidů převzatých podle referenční sekvence. Za optimální byl považován 60% obsah GC, ale odchylky byly kompenzovány počtem bazí. Podle obsahu GC či interference s dalšími polymorfními místy byla vybrána přímá nebo reversní sekvence. Při volbě reversní sekvence, která interaguje s opačným vláknem, byl v návrhu použit reversní komplement.

Navržené primery pro reakci extenze primeru byly dále sdruženy do skupin – multiplexů, které umožňují provedení reakcí v jednom reakčním objemu. Lze tak dosáhnout snížení pracnosti a nákladů. Produkty jednotlivých reakcí byly odlišeny podle hmotnosti při následné kapilární elektroforéze, neboť hmotnost primerů byla odstupňována přidáním extenze polyT na 5' konec. Hmotnost primerů s extenzí byla odstupňována tak, že hmotnosti se v jednom multiplexu lišily mezi sebou vždy nejméně o 5 nukleotidů.

Jako substrát byly pro reakci SNaPshot připraveny amplikony, v nichž se nacházejí sledované SNP. Fragменты byly získány pomocí PCR s gDNA z jednotlivých zvířat. Postupy pro geny *NFKB1* a *NFKB2*, které jsou výsledkem vlastního vývoje, jsou uvedeny v tab. 5 a 6 v části Výsledky. Pro gen *TLR2* byl používán postup amplifikace podle Nováka a Czerneková (2018a), který je uveden v tab. 1. Primery použité pro vlastní genotypování v reakcích SNaPshot odpovídají rovněž zdroji Novák a Czerneková (2018a) a jsou uvedeny v tab. 2.

Tab. 2. Genotypovací reakce SNaPShot pro bovinní gen *TLR2*

PCR fragment	Poloha a typ SNP <sup>1</sup>	Označení extenzního primeru <sup>2</sup>	Extenzní primer - sekvence 5'→3'	Délka primeru	Objem na reakci 10 µL <sup>3</sup>	Součást multiplexu reakcí	Vizualizované produkty reakce <sup>4</sup>
2.1	77T>G	T077R	(T5)AAAAAATGAA CTCCCTAACA CCTCC	30	0,43	A2	A>C
2.1	115T>C	T115R	(T10)GAAATAACCA AGAGGGAAAT GGAAT	35	0,21	A3	A>G
2.1	423G>A	T423R	(T15)GGATGGACTG GTTGGATCTC CTTGC	40	0,16	A4	C>T
2.1	513C>T	T513F	(T21)CCAGTCCTTG GGGTCACAAA GAGT	45	0,14	A5	C>T
2.2	1009A>G	T009R	ACACACCTCT GCAGGTCTCT GTTGC	25	0,21	B1	T>C
2.2	1060G>A	T060R	(T5)CACACCTCTG CAGGTCTCTG TTGC	29	0,10	B2	C>T
2.2	1044T>C	T044F	(T9)GTGAAAAGCC TTGACCTGTC CAACAA	35	0,40	B3	T>C
2.2	1047T>G	T047R	(T12)GCAGGTCTCT GTTGTCYGACA TAGGTGAT	40	0,11	B4	A>C

2.3	1313G>A	T313R	(T18)CTGTTACTAT TTCCTACTTT TAGGGTC	45	0,18	B5	C>T
2.4	2546G>A	T546F	(T12)GGTCGACTGG CCCGATGACT ACC	35	0,27	C3	G>A
2.4	2565T>C	T565F	(T16)CTACCRCTGT GACTCTCCCT CCCA	40	0,30	C4	T>C
2.4	2883T>C	T883R	(T18)CCACTTGCCA GGAATGAAGT CTCGCTT	45	0,10	C5	A>G
2.5	3206G>A	T206F	TTTGGTTAAA TTTGAGAGCT GCAATAA	27	0,33	G1	G>A

<sup>1</sup> Podle referenční sekvence GenBank EU746465.

<sup>2</sup> Označení primerů určuje, zde se jedná o přímý (F) nebo reverzní (R) primer.

<sup>3</sup> Objem přidávaný z 0,2 µM zásobního roztoku do genotypovací reakce o celkovém objemu 10 µL.

<sup>4</sup> U reverzního primeru se pozoruje komplementární vlákno.

Pro genotypování pomocí extenze primeru byly postačující fragmenty připravené pomocí standardní DNA polymerázy (Fermentas). PCR fragmenty získané v reakčním objemu 10 µL byly čištěny odstraněním nespoteřovaných jednovláknových primerů a deoxyribonukleotid trifosfátů přidáním 1 U/µL exonukleázy I a 0.3 U/µL alkalické fosfatázy FastAP (Fermentas). Směs je zpravidla označována jako EXOSAP. Po 1 h inkubaci při 37 byly enzymy inaktivovány při 85 po dobu 15 minut.

Přečištěné PCR fragmenty byly použity jako templát pro následnou reakci SNaPshot. Reakční směs v objemu 10 µL obsahovala 5 µL reakční směsi dodávané Life Technologies (SNaPshot Ready Reaction Mix), 3 µL přečištěných PCR fragmentů, 1 µL H<sub>2</sub>O a 1 µL extensních primerů, každý ve výchozí koncentraci 0,02 µM. Reakce byla prováděna na PCR destičkách uzavřených fólií na termocykléru Mastercycler (Eppendorf). Byl použit program doporučený výrobcem pro soupravu SNaPshot. Program obsahoval 25cyklů s kroky 96°C po dobu 10 sec, 50°C 5 sec a 60°C 30 sec s následným zchlazením na 4°C.

Nespoteřované fluorescenčně značené dideoxynukleotid trifosfáty byly rozloženy působením 0,3 U/µL alkalické fosfatázy FastAP (Fermentas) po dobu 1 h při 37°C. Enzym byl potom inaktivován při 85°C po dobu 15 minut.

Fluorescenční produkty v objemu 1,5  $\mu\text{L}$  byly přeneseny do formamidu o objemu 8  $\mu\text{L}$  a doplněny 0,5  $\mu\text{L}$  fluorescenčně značených standardů GeneScan 120 LIZ (Thermo Fisher Scientific). Komplexy značených primerů s templátovou DNA byly denaturovány zahřátím na 5 min při 95°C a rychlým zchlazením na ledu. Vzorčky byly děleny pomocí kapilární elektroforézy na automatickém sekvenátoru ABI 3130 (Applied Biosystems; Foster City, CA, USA) na kolonkách o délce 36 cm naplněných polymerem POP7. Pět použitých fluoroforů bylo rozlišeno sadou filtrů E5. Kalibrace byla provedena standardem DS-02 (Applied Biosystems), který zahrnuje použité fluorescenční značky dR110, dRGG, dTAMRA, dROX a LIZ. Primární elektroferogramy produktů genotypovacích reakcí byly vyhodnoceny programem PeakScanner (Thermo Scientific) a exportované polohy a výšky peaků byly zpracovány v programu Excel.

#### **4.9 Funkční charakterizace mutací a modelování produktu**

Primární funkční charakterizace nalezených genových variant byla založena na jejich lokalizaci v genu. Ta byla prováděna zejména na základě nalezených hranic exonů/intronů. Případné změny v proteinové sekvenci byly určeny po překladu od pravděpodobného iniciačního kodonu, resp. iniciačního kodonu převzatého z databáze.

Referenční a mutantní proteinové sekvence byly odeslány na server SWISS-MODEL Basilejské univerzity (<https://swissmodel.expasy.org/>) pro homologní modelování sekundární a terciární struktury. Vyhledávání templátových molekul bylo prováděno programem a byl volen templát s nejvyšší homologií.

Vrácené souřadnicové soubory ve formátu pdb byly následně vizualizovány v grafickém programu Yasara (Yasara Biosciences, Vídeň, Rakousko). Z lokalizace variabilních aminokyselin byla odvozena jejich úloha v doménách a možné funkční dopady nesynonymních záměn.

## 5 VÝSLEDKY

### 5.1 Získání knihovny genomické DNA pro populaci ČESTR

V rámci diplomové práce byla částečně připravena a standardizována knihovna genomické DNA pro klíčovou populaci 150 býků českého strakatého plemene skotu. Vzhledem k nevyrovnané koncentraci izolované DNA bylo nezbytné před přípravou směsného vzorku provést ředění na pracovní standardizovanou koncentraci 20 ng/μL.

Výsledek ředění na standardní koncentraci při přípravě normalizované knihovny byl ověřen přímo na mikrodestičce pomocí intenzity fluorescence přidaného fluorescenčního barviva SybrGreen a rovněž pomocí kontrolních PCR reakcí s fragmentem 16 genu *TLR4*. I přes ojedinělé odchylky v nastavené koncentraci byla kontrolní PCR reakce v souboru vysoce reprodukovatelná.

### 5.2 Struktura bovinních genů *NFKB1* a *NFKB2*

Genomické sekvence genů pro obě komponenty transkripčního faktoru NF-κB byly získány z databáze NCBI (National Center for Biotechnology Information, Bethesda, USA) jako sekvence AC\_000163.1:c23670155-23557313 (chromosom 6) a AC\_000183.1:22890250-22898466 (chromosom 26). Zatímco délka referenční sekvence genu *NFKB2* je 8217 bp, případě *NFKB1* bylo nutné pracovat s úsekem 112843 bp. To již omezovalo možnosti použití algoritmů pro porovnání sekvencí jako je implementovaný algoritmus programu BioEdit (Hall 1999) a algoritmus ClustalW (Larkin et al., 2007). I když jak databáze NCBI, tak i kompletní verze genomu skotu podle Department of Bioinformatics, University of Maryland (UMD\_3.1.1, kopie z NCBI) obsahují anotaci těchto genů, jedná se o strojově generovanou anotaci, která nemusí být spolehlivá. Proto bylo provedeno porovnání s příslušnými sekvencemi pro mRNA (3804 nt pro *NFKB1* a 3116 nt pro *NFKB2* jako BC151467.1) pomocí algoritmu ClustalW při nastavení penalizace 3 za vznik delece a 1 za každý další chybějící nukleotid, a to jak při párovém, tak i celkovém přiřazení. Byly tak určeny hranice exonů, kterých bylo nalezeno 23 u *NFKB1* a 22 u *NFKB2* (tab. 3, 4). Hranice intronů a exonů byly upraveny podle pravidla GT-AG pro konce intronů. Pouze intron 19 u *NFKB2* mohl být typu GC-AG (Szafranski et al., 2007).

Tab. 3. Nalezené hranice exonů pro gen *NFKB1*

Exon č.	Začátek <sup>1</sup>	Konec <sup>1</sup>	Délka (nt)	Vzdálenost sousedních amplikonů (včetně)
1	16836	16881	47	
2	20829	20904	76	4069
3	25350	25390	41	4562
4	29484	29582	99	4233
5	61681	61829	149	32346
6	73683	73846	164	12166
7	75838	75996	159	2314
8	77345	77449	105	1612
9	79879	79970	92	2626
10	81934	82072	139	2194
11	88014	88157	144	6224
12	89828	89917	90	1904
13	91900	92085	186	2258
14	93569	93710	142	1811
15	97042	97156	115	3588
16	102939	103120	182	6079
17	103732	103901	170	963
18	104223	104325	103	594
19	106149	106275	127	2053
20	107767	107833	67	1685
21	108161	108333	173	567
22	109117	109288	172	1128
23	111871	112843	973	3727

<sup>1</sup> Podle referenční sekvence č. AC\_000163.1:c23670155-23557313.

Tab. 4. Nalezené hranice exonů pro gen *NFKB2*

Exon č.	Začátek <sup>1</sup>	Konec <sup>1</sup>	Délka (nt)	Vzdálenost dvou amplikonů (včetně)
1	1143	1427	284	
2	1709	1790	81	647
3	1925	1965	40	256
4	2225	2323	98	398
5	2403	2554	151	329
6	2819	2925	106	522
7	3014	3172	158	353
8	3432	3536	104	522
9	3638	3723	85	291
10	3802	3940	138	302
11	4176	4304	128	502
12	4432	4641	209	465
13	4721	4862	141	430
14	5215	5329	114	608
15	5421	5634	213	419
16	5788	5957	169	536
17	6073	6797	724	1009
18	6856	7007	151	934
19	7138	7207	69	351
20	7438	7610	172	472
21	7724	7835	111	397
22	7937	8212	275	488

<sup>1</sup> Podle referenční sekvence č. AC\_000183.1:22890250-22898466.

Jak je vidět v tab. 3 a 4, celková délka mRNA zůstala u obou homologních genů v průběhu divergence genů srovnatelná, ale délka intronových úseků se u *NFKB1* zvětšila cca 10x.

Srovnatelná zůstala i délka kódující sekvence a proteinového produktu. Po translaci ve třech fázích byla určena délka CDS odpovídající referenční sekvenci jako 2910 nt pro *NFKB1* (ve shodě s NCBI) a 2127 nt pro *NFKB2*.

### **5.3 Amplifikace exonů bovinních genů *NFKB1* a *NFKB2***

#### **5.3.1 Optimalizace amplifikace *NFKB1***

Porovnání sekvence mRNA s gDNA dovolilo určit hranice exonů v genu *NFKB1* (tab. 3, 4). Protože jsou relativně krátké exony rozloženy celkem rovnoměrně podél genomové sekvence o délce 112843 nt, bylo nezbytné navrhovat velký počet PCR reakcí pro amplifikaci jednotlivých exonů. Vzdálenost mezi sousedními exony se pohybovala v rozmezí od 566 do 32068 nt.

Znalost hranic exonů byla využita pro návrh amplikonů, které by pokrývaly celou mRNA genu *NFKB1* (Tab. 5). Navržené primery byly prověřovány na hodnotu % GC, shodu párů v tomto ukazateli, shodu v  $T_a$ , sklon k tvorbě vlásenkových struktur a možnost tvorby autodimerů či heterodimerů. Za přípustnou byla považována volná energie tvorby heterodimeru do 6 kcal/mol.



Tab. 5. Amplikony a navržené primery pro bovinní gen *NFKB1*

Amplikon	Zahrnuje exon	Označení primeru	Začátek <sup>1</sup>	Konec <sup>1</sup>	Délka amplikonu	Sekvence 5'→3'	Délka (nt)	T <sub>m</sub> <sup>2</sup> (°C)	T <sub>a</sub> <sup>3</sup> (°C)
1	1	K1_1F	16723	16744	333	GGGAGAGGTTATATTTGGTCTT	22		62/60
		K1_1R	17034	17055		TGTTTTACATGGCATGTCCTCT	22		62/60
2	2	K1_2F	20624	20644	488	CAACCATAGCCCAACCCCATG	21		62/60
		K1_2R	21091	21111		TGCCTCCTCCAATAACAATGGG	21		62/60
3	3	K1_3F	25293	25314	304	CAAATTTGAGAAGCCTCACAG	22		62/60
		K1_3R	25576	25596		ACTGAGGTTCAGGACTGTCTG	21		62/60
4	4	K1_4F	29329	29351	316	AAGGTTATGGGGAGAAGTAGTTC	23	53.9 °C	62/60
		K1_4R	29625	29644		CTCTTTTGCTCCGAGGACAC	20	55.7 °C	62/60
5	5	K1_5F	61505	61528	407	CACACCCGAAACAAGTACCACTCG	24	59.8 °C	62/60
		K1_5R	61891	61911		AAAAGGGCCTGTCCAGTCCCA	21	61.2 °C	62/60
6	6	K1_6F	73575	73598	375	AGTGTGACTGAGGGTCTGTGCCTG	24	62.6 °C	62/60
		K1_6R	73929	73949		AGCAGCCTGGGGCACAGACTG	21	64.6 °C	62/60

7	7	K1_7F	75633	75653	454	GCAGGCCTTCTTCTGATGCCT	21	59.9 °C	62/60
		K1_7R	76063	76086		CCAGACCTAGTCAGTAACTGCCAG	24	58.5 °C	62/60
8	8	K1_8F	77270	77290	278	AAGGTCCTCACAGGAGGTTA	21	57.3 °C	62/60
		K1_8R	77527	77547		TGGCAGAGGAGAGATGTTACT	21	55 °C	62/60
9	9	K1_9F	79772	79796	334	GGTGGAGTACATAGTAGGTAGACTC	25	55.2 °C	62/60
		K1_9R	80086	80105		CACCTCCAAGACACAGTGCA.	20	57.5 °C	62/60
10	10	K1_10F	81829	81852	337	AGACATTGTGTCTAAGCATTGGGC	24	57.7 °C	62/60
		K1_10R	82146	82165		AGGAGGCTGGTCAAAGGAGG.	20	59.1 °C	62/60
11	11	K1_11F	87875	87894	366	TGGCATCTCTTCCTCCTTGG.	20	56.7 °C	62/60
		K1_11R	88216	88240		ACA ACTACTCAATGCATAGACAAGG	25	54.9 °C	62/60
12	12	K1_12F	89732	89751	320	TGTCAGACACACCGAAGCCT.	20	58.7	62/60

								°C	
		K1_12R	90026	90051		GAATAGAATGTCTACTCTCCACAGGA	26	55.3 °C	62/60
13	13	K1_13F	91793	91813	377	TGCCTCTACCTTTGGCTGACC	21	59.3 °C	62/60
		K1_13R	92149	92169		CTGAAACAGGTTCTCCACCCA	21	57 °C	62/60
14	14	K1_14F	93459	93479	337	TGTGTCGCAGGGTCTCAGAGT	21	60.3 °C	62/60
		K1_14R	93772	93795		TCATCTCATCTCTAGCCAGCACGT	24	58.5 °C	62/60
15	15	K1_15F	96960	96980	354	TAATCCTATCGTGGGAACCTG	21	53.7 °C	62/60
		K1_15R	97291	97313		AAGAACATTCCTATGTAATGGGC	23	52.9 °C	62/60
16	16 + 17	K1_16F	102811	102831	1187	ATGGCCCCTTAGTATGATTCC	21	54.1 °C	62/60
		K1_16R	103977	103997		TAGGGAGATGAGCTGTTAGCG	21	56 °C	62/60
17	18	K1_17F	104147	104166	180	CATGCAGACAGGGGAGGTGG	20	57.2 °C	62/60

		K1_17R	104407	104409		TCAGGAGGTGGCCCGGCTGCTCT	23	67.8 °C	62/60
18	19	K1_18F	106027	106052	338	GCAGGTTTCTGGATATTTTCAGTGGAA	26	58.4 °C	62/60
		K1_18R	106344	106364		AAATACCAGGTGCTGGCTTCG	21	59.1 °C	62/60
19	20 + 21	K1_19F	108031	108051	431	TGCTGACTCTCCAGTGCAAGA	21	65.3 °C	62/60
		K1_19R	108440	108461		AGTTAAGGGAGAGGGGAAGGGA	22	65 °C	62/60
20	22	K1_20F	109025	109045	376	GGGGGACGGGGTGTGGTGTG	21		62/60
		K1_20R	109380	109400		AGAGGGACACAGACGCGGGGA	21		62/60
21	23	K1_21F	111697	111717	891	TGCTCGGTGTTTTCTGCCCA	21		62/60
		K1_21R	112567	112587		GCCACCTGGCTTGCAGGATCT	21		62/60

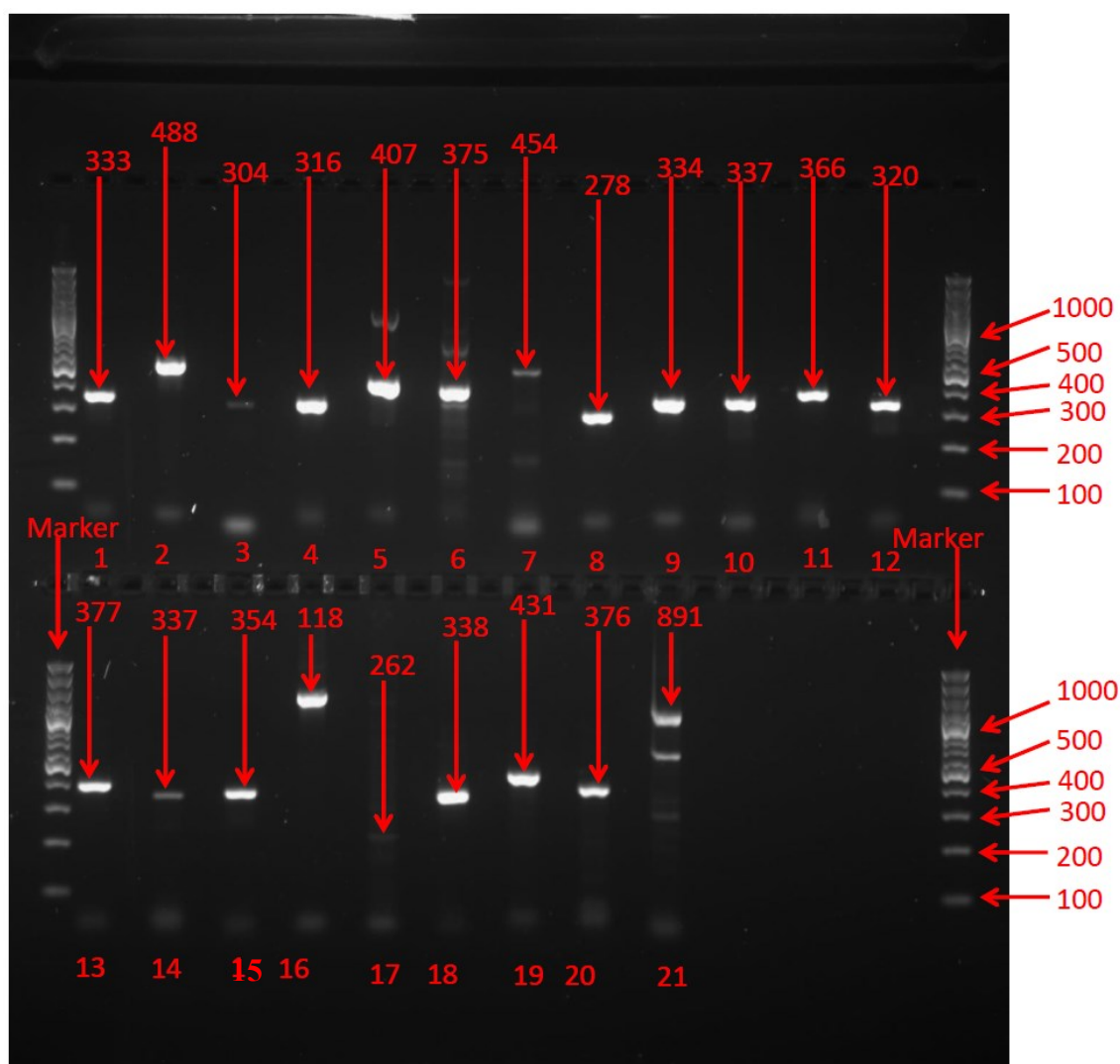
<sup>1</sup> Podle referenční sekvence č. AC\_000163.1:c23670155-23557313

<sup>2</sup> Melting point – vypočítaná teplota denaturace

<sup>3</sup> Teplota pro přisedání v amplifikačním programu při použití základní Taq polymerázy od výrobce Fermentas

Optimalizace vazebné teploty  $T_a$  pro amplifikační program probíhala s použitím základní termostabilní (Taq) polymerázy od výrobce Fermentas a s kontrolou očekávaných produktů elektroforézou v agarózovém gelu (obr. 7). Výtěžek jednotlivých reakcí bylo možné porovnat vzhledem ke stálému objemu nanášené reakční směsi (2  $\mu$ L) a směsi standardů GeneRuler 100 bp. Kontrolní PCR proběhla úspěšně pro většinu párů primerů z 21 amplikonů pro *NFKB1*. pouze u párů primerů 3, 7, 14 a 17 byl pozorován nižší výtěžek. Navíc se u párů primerů 5, 6, 7 a 21 tvořily nespecifické produkty (obr. 7). Přesto nebylo nutné návrh primerů opakovat a sadu primerů doplňovat.

Empiricky potvrzené hodnoty  $T_a$  jsou stejně jako vypočítané  $T_m$  uvedeny v tab. 5. Pro použití s polymerázou Phusion se ukázalo jako postačující zvýšit teplotu  $T_a$  optimalizovanou pro Taq polymerázu Fermentas o 2  $^{\circ}$ C.



Obr. 7. Kontrola PCR fragmentů generovaných navrženými primery pro amplifikaci *NFKB1* a optimalizace  $T_a$

### 5.3.2 Optimalizace amplifikace *NFKB2*

Obdobný postup jako u genu *NFKB1* pro identifikaci exonů, návrh a kontrolu ampliconů byl uplatněn u genu *NFKB2* (tab. 6, obr. 8).

I když je genomická sekvence *NFKB2* daleko kratší než *NFKB1* (8217 nt), tj. 7,28%), stejně se nepodařilo sdružovat více exonů do jednotlivých ampliconů. Proto výsledný počet navržených ampliconů (22) se neliší od počtu nalezených exonů..

Tab. 6. Amplicony a navržené primery pro bovinní gen *NFKB2*

Amplicon	Zahrnuje exon	Primer	Začátek <sup>1</sup>	Konec <sup>1</sup>	Délka ampliconu (nt)	Sekvence 5'→3'	Délka primeru (nt)	Tm <sup>2</sup> (°C)	Ta <sup>3</sup> (°C)
1	1	K2_1F	1063	1083	446	GCGTGGCCAGTGGCGTCATTT	21	64,4	64/62
		K2_1R	1489	1509		AAGGGACTAGGACAGGCAGGC	21	62,2	64/62
2	2	K2_2F	1639	1659	240	AGAGCTGAGAGGTGGGGGGTT	21	63.6	64/62
		K2_2R	1858	1879		CCCAAGACAGAGGGCACACACC	22	63.3	64/62
3	3	K2_3F	1856	1876	188	TTGTCTTGAGGGAGGGGGAG	21	61.7	64/62
		K2_3R	2024	2044		CTGTGGCCAGCCAGCTACAA	21	61.2	64/62
4	4	K2_4F	2140	2160	236	TTATGGTCACTGCTGGCTGGG	21	60.6	64/62
		K2_4R	2356	2376		TGCCCATGGACACACTGTCCA	21	62.3	64/62
5	5	K2_5F	2325	2345	292	TGAGCCAGAGTGGTGCTGGAG	21	62.4	64/62
		K2_5R	2597	2617		GGA CTGGCAAGCGATGAAGGG	21	61.7	64/62

6	6	K2_6F	2726	2746	266	AGTAGAGCAGGGCCAAGCCTT	21	61.9	64/62
		K2_6R	2972	2992		GACCCCTGGACTCCTCCTCTA	21	60.4	64/62
7	7	K2_7F	2937	2957	296	ATGGAGGGAGTCATGGGAGGT	21	60.9	64/62
		K2_7R	3213	3233		GCTGCCCTCCACACTACCTGT	21	62.5	64/62
8	8	K2_8F	3358	3378	256	ACAAAGCACCTGGCTCCCTAG	21	60.3	64/62
		K2_8R	3594	3614		GGGCCATTTGTCCTGAAGCTG	21	59.8	64/62
9	9	K2_9F	3558	3578	242	GGAGCTAGCGATGGGTATCAA	21	57.3	64/62
		K2_9R	3780	3800		TAGGGGGAAGGAGGACATAGG	21	58.3	64/62
10	10	K2_10F	3740	3760	287	GCTGGGGCTGGGCTCAAATTA	21	61.2	64/62
		K2_10R	4007	4027		CACTCATGGGTCCCCTGATCT	21	59.5	64/62
11	11	K2_11F	4066	4086	359	TCCCCACAGTCCTGCCTGTAT	21	61.5	64/62
		K2_11R	4405	4425		GGCAGGACTGTGGGGAGAGGG	21	61.2	64/62
12	12	K2_12F	4318	4338	388	GGTGGCAGGAAGGGCAATGGA	21	63.5	64/62
		K2_12R	4686	4706		GGGCGTGGGTGTGGACTAGGT	21	64.8	64/62
13	13	K2_13F	4643	4663	298	TATGACCTCGGGAATCCCAGC	21	59.7	64/62
		K2_13R	4921	4941		TTCCTTTCCTGCCCCACAGCC	21	63.4	64/62
14	14	K2_14F	5081	5101	305	CTTACCGCAAAGGTGCGCGG	21	62.8	64/62
		K2_14R	5366	5386		CCTTGCCTCACCCACTGACT	21	62.6	64/62
15	15	K2_15F	5338	5358	377	CCTCTATTGGTGAGGGGGAGG	21	59.8	64/62
		K2_15R	5695	5715		CACGTCCACTGTAGGCACTCC	21	60.7	64/62
16	16	K2_16F	5717	5737	297	GGTGTGGAGGTCAGCTGGTCA	21	62.4	64/62

		K2_16R	5994	6014		GACACCTTCACCATCCCCAGC	21	61.3	64/62
17	17	K2_17F	6001	6021	284	ATGGTGAAGGTGTCGGGGGAG	21	62.4	64/62
		K2_17R	6819	6839		GTCACAGGGAGGCTCCCTTGG	21	62.5	64/62
18	18	K2_18F	6745	6765	351	GAACAGGGCTGAGGGGGGAAA	21	62.8	64/62
		K2_18R	7076	7096		CCTTGACCCCCAGTACCAGCT	21	62.2	64/62
19	19	K2_19F	7024	7044	291	ACTAGCAGTACCAGGGGATGG	21	59.2	64/62
		K2_19R	7295	7315		GGAGGACGGGGATGGAGAAAG	21	60.3	64/62
20	20	K2_20F	7332	7352	355	GCCCCGCCTGCACTATCAATC	21	61.8	64/62
		K2_20R	7667	7687		GTTAGGGTGTGGGATCCCGGA	21	61.9	64/62
21	21	K2_21F	7630	7650	298	ATGCTCCATGCCCCAGTCTCA	21	62	64/62
		K2_21R	7908	7928		GGAGTGGACACGGGAAAGGGA	21	62.2	64/62
22	22	K2_22F	7852	7872	365	TGAGAGGTGGATCTGGGCTTG	21	59.8	64/62
		K2_22R	8197	8217		TGCTGTAAGTTGGGGAGGAGG	21	59.8	64/62

1  
Pod  
le  
refe  
ren  
ční  
sek  
ven  
ce  
č.  
AC  
\_00

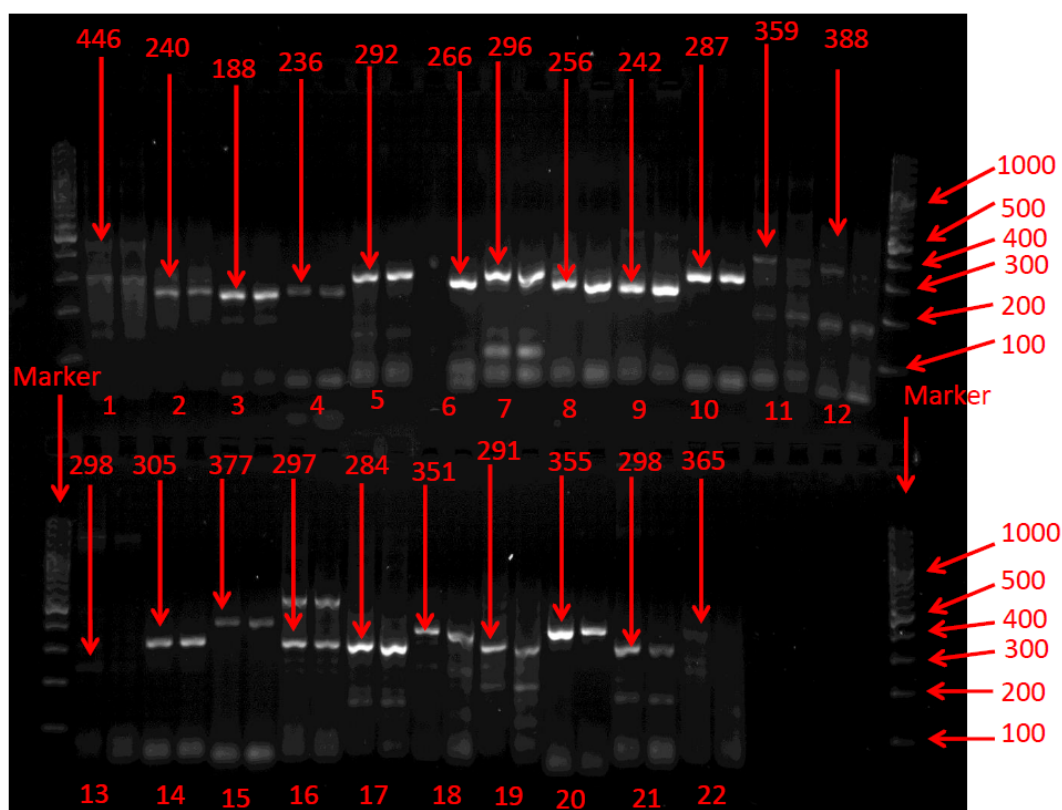
0183.1:22890250-22898466.

<sup>2</sup> Melting point – vypočítaná teplota denaturace

<sup>3</sup> Teplota pro přisedání v amplifikačním programu při použití základní Taq polymerázy od výrobce Fermentas



Výtěžek testovaných amplifikačních PCR reakcí je vizualizován na obr. 8. Ověření návrhu primerů *NFKB2* potvrdilo, že většina fragmentů byla generována úspěšně. Byla nalezena shoda s očekávanou hmotností, absence výrazných vedlejších produktů PCR a také s dostatečný výtěžek jednotlivých reakcí (s možnou výjimkou fragmentů 13 a 18). Přesto nebylo nutné, stejně jako u *NFKB1*, návrh primerů opakovat a sadu primerů doplňovat.



Obr. 8. Kontrola PCR fragmentů generovaných navrženými primery pro amplifikaci *NFKB2* a optimalizace  $T_a$  (aplikovaná  $T_a$  64/62 °C)

U obou *NFKB* genů nebyla identita produktu ověřována kapilárním sekvenováním, neboť jako zpětná kontrola sloužily sekvence získané při NGS směšného vzorku. Přítomnost všech očekávaných readů v dostatečném zastoupení nepřímo potvrdila identitu fragmentů.

#### 5.4 Preparativní amplifikace panelu TLR genů a genů *NFKB1* a *NFKB2* pro NGS

Populační směšný vzorek genomické DNA sledované skupiny býků byl získán spojením alikvót ze 149 standardizovaných vzorků DNA o koncentraci 20 ng/ $\mu$ L. Proto byl

genom každého ze zvířat zastoupen v populaci stejným dílem. Směsný vzorek byl použit jako vstup pro preparativní amplifikaci fragmentů pokrývajících studované geny.

Amplifikace 73 fragmentů pokrývajících exony, resp. kódující sekvence 10 bovinních TLR genů (tab. 1) byla provedena podle publikované metodiky (Novák et al., 2017; Novák a Czerneková, 2018a, b). Pro preparativní účely byl zvýšen objem reakce na 20  $\mu$ L a byla použita nízkochybová Taq polymerázy Phusion (New England Biolabs). Výtěžek byl ověřen pomocí kontrolní elektroforézy (obr. 9). Koncentrace produktu byla kvantifikována porovnáním intenzity fluorescence příslušného pruhu v agarózovém gelu s intenzitou standardu a údaj byl použit pro přípravu směsného vzorku všech fragmentů. Rozdílné koncentrace PCR produktů byly kompenzovány různými pipetovanými objemy. Kromě inverzní závislosti na hmotnostní koncentraci byla zohledněna i molekulová hmotnost fragmentů, čímž bylo dosaženo ekvimolárního zastoupení.

Podobný postup byl použit i pro amplifikaci a přípravu směsného vzorku PCR fragmentů získaných z populačního vzorku gDNA v případě genů *NFKB1* a *NFKB2*.

Pro přípravu vzorku na sekvenování technologií PacBio byl připraven konečný směsný vzorek obsahující fragmenty převážně antivirových TLR (TLR3, -7, -8, -9 a -10) *NFKB1* a *NFKB2*. Fragmenty antibakteriálních TLR (*TLR1*, -2, -4, -5 a -6) byly součástí dalšího směsného vzorku pro jiný běh přístroje.

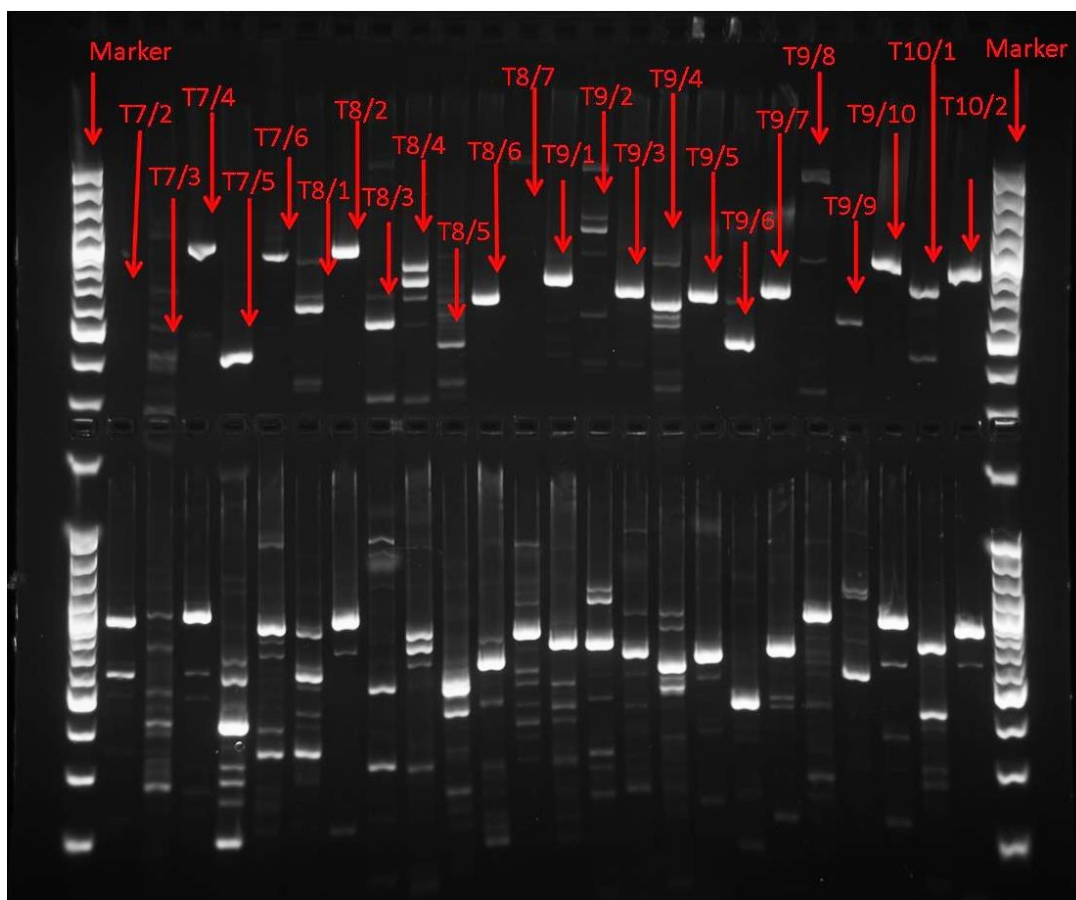
Vzhledem k homologii genů *NFKB1* a *NFKB2* byla ověřována možnost spolehlivého rozlišení readů obou genů v případě jejich společného čtení v jednom směsném vzorku. Protože se homologní amplicony lišily nikoliv jednotkami, ale stovkami nukleotidů, nebyla tato možnost pravděpodobná.

Sekvenační knihovny jako vstup pro technologii cirkulárního konsensního sekvenování (CCS, "circular consensus sequencing") PacBio byly připraveny v rámci servisní laboratoře NGS.

V případě antibakteriálních TLR genů byla převzata a validována metoda jejich sériové amplifikace publikovaná jako užitečný vzor (tab. 1; Novák a Czerneková 2018a) a úspěšně aplikována na směsný vzorek DNA ze sledované populace skotu. Pro sekvenování následujícího směsného vzorku ampliconů metodou PacBio byly generovány amplicony kódujících oblastí *TLR1*, -2, -4, -5 a -6. Ekvimolární koncentrace gDNA byly nastaveny podle spektrofotometrických údajů. Směsný vzorek antibakteriálních ampliconů byl před NGS doplněn amplicony *NFKB1*.

Obdobně byl připraven směsný vzorek ampliconů z antivirových TLR genů (obr. 9), částečně podle Novák a Czerneková (2018b). I když výsledky sekvenování těchto genů

nejsou v rámci diplomové práce interpretovány, vzorek byl použit jako prostředí pro sekvenování ampliconů *NFKB2* technologií PacBio.



Obr. 9. Preparativní amplifikace antivirových TLR genů pomocí Phusion polymerázy před NGS

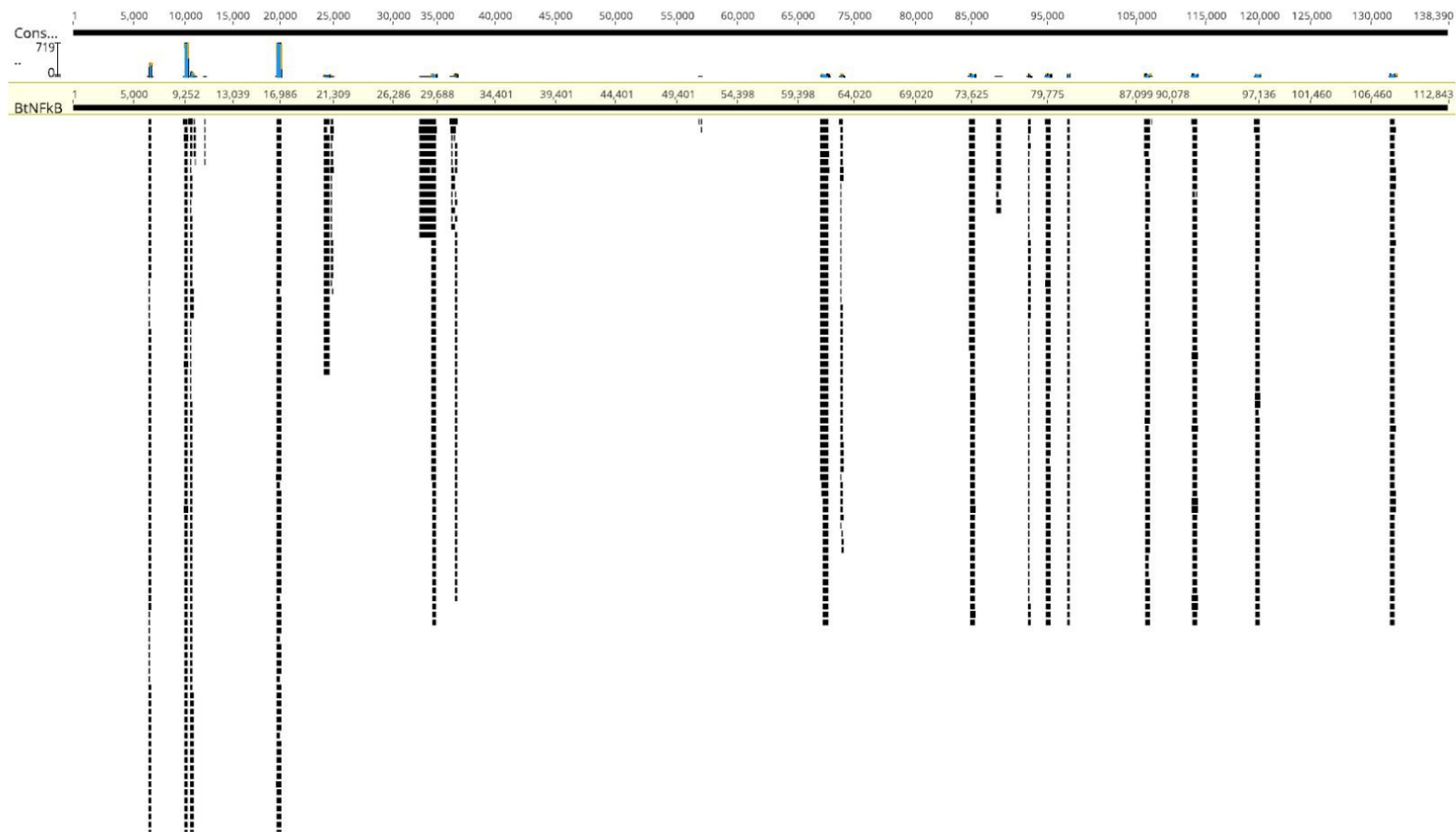
### 5.5 Zpracování výsledků NGS z technologie PacBio

Výpočetní technika VÚŽV nedovolila zpracovávat původní výstupní formát z technologie PacBio ve firemním formátu h5, zejména pro nekompatibilní operační systém LINUX CentOS, Proto zpracování začalo až soubory FastQ , které jsou rovněž poskytovány ze servisní laboratoře.

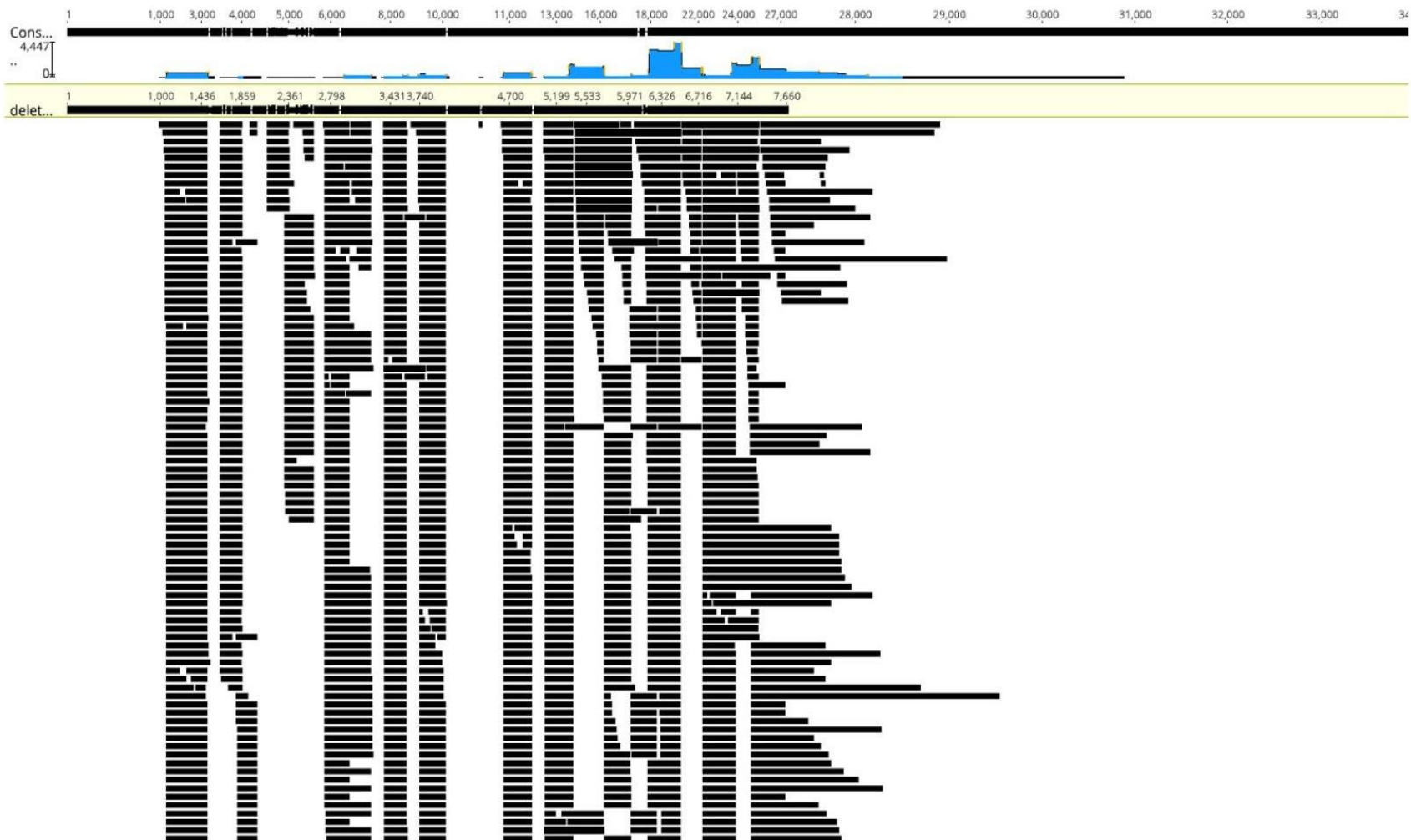
U obou vzorků byly provedeny 3 běhy přístroje v režimu Circular Consensus Sequencing. Tři výstupní soubory, které jim odpovídaly, byly spojeny do jednoho pro snazší zpracování.

Mapování na referenční sekvenci příslušného genu probíhalo pomocí implementovaného algoritmu programu Geneious (obr. 10, 11, 12). Na obr. 10 je zřejmý

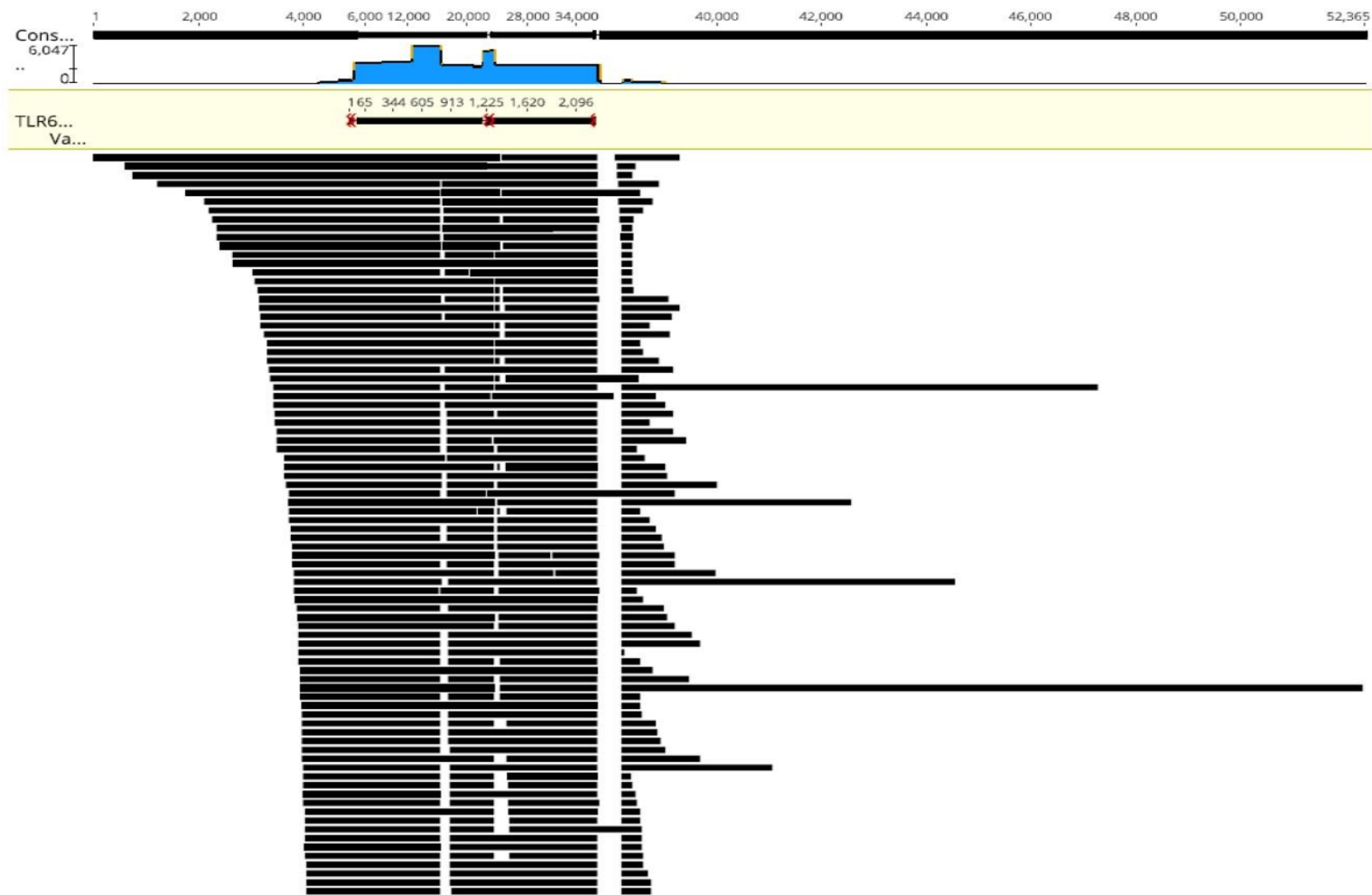
nepoměr mezi rozsahem intronů a exonů u tohoto genu. Expanze intronů po divergenci *NFKB1* a *NFKB2* je vysvětlitelná insercemi retrotransposonů (Nilsson et al., 2012), nicméně tento předpoklad by bylo nezbytné dokázat. Celkem bylo sekvenováním PacBio získáno 84499 readů pro směsný vzorek *NFKB1*, *NFKB2* a antivirové TLR, zatímco pro antibakteriální TLR včetně *TLR2* bylo získáno 45408 readů. Průměrné pokrytí pro všechny amplikony bylo více než 1000 x, což při počtu 149 vzorků (přibližně odpovídá počtu zvířat) bylo téměř 8 readů na vzorek/zvíře. Toto pokrytí dovoluje na populační úrovni detekovat i varianty s nízkou frekvencí pod 5%.



Obr. 10. Distribuce readů *NFKB1* ze sekvenování populačního vzorku ČESTR pomocí PacBio



Obr. 11. Distribuce readů *NFKB2* ze sekvenování populačního vzorku ČESTR pomocí PacBio



Obr. 12. Distribuce readů *TLR2* ze sekvenování populačního vzorku ČESTR pomocí PacBio

Strukturní varianty byly hledány pomocí implementovaného algoritmu programu Geneious. Pro detekci se osvědčil spodní limit frekvence 0,01 a hladina významnosti  $P < 0,01$ . Vizuální kontrolou byly odstraněny polymorfismy, které byly detekovány jako důsledek chybného přiřazení readů s nízkou homologií. Nalezené polymorfismy byly exportovány ve formě tabulky ve formátu csv a dále zpracovány v programu Excel. Polymorfismy v genech *NFKB1*, *NFKB2* a *TLR2* jsou shrnuty v tab. 7, 8 a 9.

Tab. 7. Přehled polymorfismů nalezených resekvenováním exonů *NFKB1* populačního vzorku ČESTR technologií PacBio

Polymorfismus <sup>1</sup>	Typ	V databázi EBI	Frekvence varianty (%)	Signifikance varianty (P)	Lokalizace
61548C>T	ss1088476390, intron	61547G> A	46. 8	0.0	amplikon 5
61871T>C	ss1088476389, intron	61870A> G	46. 1	0.0	amplikon 5
75712A>G	ss1088476319, intron	75711T> C	39. 1	1.1E-56	amplikon 7
75795C>T	SNV, intron, Chen et al., 2015	75795T> C	1.8	0.0045	amplikon 7
79846T>G	ss1088476298, intron	79854A> C	19. 2	0.00008 6	amplikon 7
81892A>G	ss1088476288, intron	81889C> T	2.0	0.0039	amplikon 10
82126G>C	ss1088476285, intron	82125C> G	40. 5	2.5E-57	amplikon 10
91885(T)20>(T) 19	delece, intron	Chybí	90. 1	3.1E- 162	amplikon 10



92022G>A	SNV, G474R	92021C> T	45. 3	0.0	amplikon 13
92029G>C	SNV, synonymní	92028C> G	48. 2	0.0	amplikon 13
96996(T)11>(T) 10	ss1088476191, delece, intron	96995A>	33. 9	1.5E-20	amplikon 15
106293C>T	ss1088476140, intron	106289G> A	35. 5	6.7E-38	amplikon 18
108030C>T	SNV, intron	Chybí	61. 1	1.1E-24	primer amplikonu 19F
108121 del GAATC	delece, intron	Chybí			amplikon 19
108162A>G	2. nt exonu 21	Chybí	3.7	0.0021	amplikon 19
108171A>G	SNV, synonymní	Chybí	3.4	0.0023	amplikon 19
108174A>G	ss1088476120, synonymní	108173T> C	3.3	0.0024	amplikon 19
108282T>C	ss1088476118, synonymní	108281A> G	27. 5	2.9E-34	amplikon 19
108338C>T	ss1088476116, intron	108339A> G	4.1	0.00000 74	amplikon 19
108443C>T	ss1088476114, intron	108442G> C	7.8	9.7E-8	primer 19R
112059 del GTTT	delece, 3'-nepřekládaná oblast	Chybí	72. 2	1.3E-36	amplikon 21
112477CTT>TA G	ss1088476077, 3'- nepřekládaná oblast	112478A> C	2.6	0.0078	amplikon 21

<sup>1</sup> Podle referenční sekvence č. AC\_000163.1:c23670155-23557313.

Tab. 8. Přehled polymorfismů nalezených resekvenováním exonů *NFKB2* populačního vzorku ČESTR technologií PacBio

Polymorfismus <sup>1</sup>	Typ	Databáze EBI	Frekvence varianty (%)	Signifikance varianty (P)	Lokalizace
1831- GGGA	delece, intron	Nový	18.9% > 20.6%	0.0000077	amplikon 2
1843G>C	SNV, intron	Nový	7.3%	0.0000026	amplikon 2
1847G>T	SNV, intron	Nový	7.5%	1.6E-7	amplikon 2
2164A>G	SNV, ss1103392227, intron	Ano	15.4%	2.3E-9	amplikon 4
2299C>A	SNV, synonymní	Ano	13.5%	1.4E-12	amplikon 4
2652G>C	SNV, intron	Nový	5.6%	0.000010	amplikon 5
4586C>T	SNV, synonymní	Ano	30.6%	1.6E-45	amplikon 12
6847T>C	SNV, intron	Nový	19.6%	8.7E-27	amplikon 18
7545C>T	3'- nepřekládaná oblast	Nový	5.7%	0.0000037	ampl. 20
7574A>C	3'- nepřekládaná oblast	Nový	10.3%	3.3E-12	ampl. 20
7930C>A	3'- nepřekládaná oblast	Nový	9.8%	4.0E-12	ampl. 22
8060C>T	3'- nepřekládaná oblast	Nový	4.3%	0.0000065	ampl. 22

<sup>1</sup> Podle referenční sekvence č. AC\_000183.1:22890250-22898466.

Tab. 9. Nalezené polymorfismy v genu *TLR2* pomocí resekvenování v populaci ČESTR technologií PacBio

Polymorfismus <sup>1</sup>	Poloha na chr. 17	Typ	V databázi EBI	Frekvence varianty (%)	Signifikance varianty (P)
115T>C	3953930	5'-nepřekládaná oblast	rs68343162	16.3%	0,001>
222A>T	3953823	5'-nepřekládaná oblast	rs68268242	16.8%	0,001>
412G>A	3953633	5'-nepřekládaná oblast	rs68343164	15.9%	0,001>
423G>A	3953622	5'-nepřekládaná oblast	rs378996667	17.6%	0,001>
459T>C	3953586	5'-nepřekládaná oblast	rs68268244	15.0%	0,001>
528G>A	3953517	5'-nepřekládaná oblast	rs68343166	17.3%	0,001>
538A>C	3953507	5'-nepřekládaná oblast	rs68268246	17.6%	0,001>
624G>T	3953421	5'-nepřekládaná oblast	rs68268247	19.0%	0,001>
657C>T	3953388	5'-nepřekládaná oblast	rs68268248	16.8%	0,001>
817T > A	3953228	5'-nepřekládaná oblast	nový	31.9%	0,001>
825A > G	3953220	5'-nepřekládaná oblast	nový	15.3%	0,001>
1044T > C	3953001	synonymní varianta	rs68268249	11.4%	0,001>
1047G > T	3952998	missense,	rs55617172	25.7%	0,001>

		E63D			
1313G>A	3952732	missense, R152Q	rs43706434	17.7%	0,001>
1489A>G	3952556	missense, I211V	rs43706433	37.4%	0,001>
2546G>A	3951499	missense, R563H	rs68268260	9.5%	0,001>
2565T>C	3951480	synonymní varianta	rs41830058	20.3%	0,001>
2853C>G	3951192	missense, H665Q	rs68268263	13.2%	0,001>
2883T > C	3951162	synonymní varianta	rs68343171	9.9%	0,001>
3153C > T	3950892	synonymní varianta	rs68268267	13.1%	0,001>
3247G > T	3950986	3'-nepřekládaná oblast	nový	18.1%	0,001>

<sup>1</sup> Podle referenční sekvence č. EU746465.

## 5.6 Návrh a validace genotypovacích reakcí pro SNP v NFKB1 a NFKB2

V tab. 10, 12 a 14 je rovněž uvedeno sdružení nalezených SNP do haplotypů v tom rozsahu, v jakém je bylo možné určit na základě dlouhých čtených úseků technologie PacBio.

### 5.6.1 Haplotypová struktura genu NFKB1 v populaci ČESTR

Znalost haplotypové struktury byla vzata v úvahu při výběru diagnostických SNP (tzv. tagSNP), které umožňují určovat přítomnost celých skupin SNP ve vazbě. Výběr diagnostických SNP umožnil určitou redukci genotypovacích reakcí (o cca 20 %), nicméně úspora byla menší, než se očekávalo. Je to důsledek vysoké diverzity v obou genech NFKB a existence velkého počtu rekombinovaných alel přítomných v populaci ČESTR. Nízký

počet alel by i při vysokém počtu SNP mohl snížit počet genotypovacích reakcí pro příslušný gen, teoreticky až na jednu reakci.

Tab. 10. Haplotypy určené ze sekvencí přečtených PacBio technologií u genu *NFKB1* v populaci ČESTR

Haplotyp číslo	Amplikon číslo					
	amplikon 5		amplikon 13		amplikon 19	
	SNP v haplotypu	Frekvence haplotypu (%)	SNP v haplotypu	Frekvence haplotypu (%)	SNP v haplotypu	Frekvence haplotypu (%)
	61548C>T, 61871T>C		92022G>A, 92029G>C		108121ΔA>G, 108162A>G, 108171A>G, 108174A>G, 108282T>C, 108338C>T, 108443C>T	
1	C_C	36,4	A_G	45,6	A_A_A_A_C_C_C_C	11,4
2	C_T	34,5	A_G	1,8	A_A_A_A_T_C_C_C	75,0
3	T_C	18,2	A_C	17,5	A_A_A_A_T_C_T	4,5
4	T_T	10,9	G_G	35,1	A_A_A_A_T_T_C	2,3
5					A_A_A_G_T_C_C	2,3
6					A_A_G_A_T_C_C	4,5
7					G_A_A_A_T_C_C	0,0

### 5.6.2 Návrh genotypovacích reakcí pro *NFKB1*

Přehled genotypovacích reakcí typu extenze primeru navržených pro *NFKB1* je uveden v tab. 11. Celkem bylo navrženo 8 reakcí pro *NFKB1* a 11 reakcí pro *NFKB2* (tab. 13). Pro praktické účely byly sdruženy do jedné série multiplexních reakcí, označených sada A, B a

C po 6, 6 a 7 reakcích., U reakcí v tab. 11 je uveden seznam výchozích PCR fragmentů, které slouží jako substrát. Příslušné PCR fragmenty byly připraveny jednotlivě, výtěžek byl zkontrolován elektroforeticky a pro příslušné multiplexy genotypovacích reakcí byly použity ve směsi.

Spektrum fragmentů použitých pro genotypování je menší než plný rozsah fragmentů, které byly výchozí pro hledání polymorfismů pomocí NGS. I když je citlivost metody SNaPshot dostatečná pro genotypování v původní neredukované směsi PCR fragmentů, z praktických důvodů byly připravovány konkrétní fragmenty pro konkrétní multiplexy reakcí. Vyloučení zbytečných fragmentů sloužilo jako pojistka proti výskytu nespecifických genotypovacích produktů, jejichž přítomnost mohla ztížit vyhodnocení multiplexních reakcí.

Tab. 11. Genotypovací reakce typu extenze primeru navržené pro gen *NFKB1*

Označení primeru <sup>1</sup>	SNV	Sekvence 5'>3'	Délka komplementární části	GC % komplementární části	Délka s extenzí 5'-polyT	Multiplex reakcí SNaPshot	Část multiplexu	Vizualizované produkty reakce <sup>2</sup>
1N54 8F	61548C >T	CACTCGTGTTTTATTTCCAT CCCAT	25	40,0	25,0	A	A1	C > T
1N87 1F	61871T >C	TTTTTTTTTCTGCAGAAGGG AGGCGCAGAT	21	61,9	30,0	A	A2	T > C
1N02 2F	92022G >A	TTTTTTTTTTTTTAACAAGAT GGAGGATCCTGCAAG	23	47,8	35,0	A	A3	G > A
1N02 9R	92029G >C	TTTTTTTTTTTTTTTTTCGGA GCACGTCAGGTTAGCTGCA	23	60,9	40,0	A	A4	C > G

1N12 1F	108121 A>G	TTTTTTTTTTTTTTTTTTTT TCCGGTCCCTGTGGAGAGT CTCTG	23	65,2	45,0	A	A5	A > G
1N16 2R	108162 A>G	TTTTTTTTTTTTTTTTTTTT TTTTCATCTTCAGTCAGTTG TTTCATGTC	25	40,0	50,0	A	A6	T > C
1N28 2R	108282 T>C	AGCGGGACTCAGCCGGAAG GC	21	71,4	21,0	B	B1	A > G
1N44 3R	108443 C>T	TTTTTTGCTCAGAGTTAAGG GAGAGGGGAA	24	54,2	30,0	B	B2	G > A

<sup>1</sup> Označení primerů vyjadřuje 1N = *NFKBI* + poslední tři číslice z lokalice na původně použité referenční sekvenci + F nebo R pro přímý nebo reverzní primer

<sup>2</sup> U reverzního primeru se pozoruje komplementární vlákno

### 5.6.3 Haplotypová struktura genu *NFKB2* v populaci ČESTR

Stejně jako u *NFKB1* byla vyhodnocena haplotypová struktura pro gen *NFKB2* na základě současné přítomnosti nukleotidových variant v readech získaných technologií PacBio (tab. 12). Haplotypovou strukturu bylo možné určit jen ve třech amplikonech (č. 2, 4 a 20). Vzhledem k výskytu rekombinantních haplotypů v amplikonu 2 a 20 nebyl diagnostický soubor zjednodušen. Na druhé straně absolutní a silná vazba markerů v amplikonech 4 a 20 umožnila jednotlivé reakce vynechat.

Tab. 12. Haplotypy určené ze sekvencí přečtených PacBio technologií u genu *NFKB2*

Haplotyp číslo	Amplikon číslo					
	amplikon 2		amplikon 4		ampl. 20	
	SNP v haplotypu	Frekvence haplotypu	SNP v haplotypu	Frekvence haplotypu	SNP v haplotypu	Frekvence haplotypu
	1831G>T, 1843G>C, 1847G>T		2164A>G, 2299C>A		7545C>T, 7574A>C, 7930C>A	
1	GGG	82,7	AC	100,0	CAC	90,5
2	GGT	1,0			CAA	1,6
3	GCG	1,0			CCC	4,8
4	GCT	1,9			CCA	1,6
5	TGG	12,5			TCA	1,6
6	TGT	1,0				

### 5.6.4 Návrh genotypovacích reakcí pro *NFKB2*

Přehled genotypovacích reakcí typu extenze primeru navržených pro *NFKB1* je uveden v tab. 13. Celkem bylo navrženo 11 reakcí pro *NFKB2*. Pro praktické účely byly sdruženy s reakcemi pro *NFKB1* do jedné série multiplexních reakcí, označených jako sada A - C.



Tab. 13. Genotypovací reakce navržené pro gen *NFKB2*

Označení primeru <sup>1</sup>	SNV	Sekvence 5'>3'	Délka komplementární části	GC % komplementární části	Délka s extenzí 5'-polyT	Multiplex reakcí SNaPshot	Část multiplexu	reakce **
2N001R	1TG > GA	TTTTTTTTTTTTTTCCTCTCCCACCACCAGGACTGCA	23	65,2	35	B	B3	A >C
2N831F	1831- GGGA	TTTTTTTTTTTTTTTTTTCCTGCCCTTGTCTTGGAGGGAGG	23	65,2	40	B	B4	G>T
2N843F	1843G > C	TTTTTTTTTTTTTTTTTTTTTCTTGGAGGGAGGGGGAGGGAGGAG	24	66,7	45	B	B5	G > C
2N847F	1847G > T	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGGAGGGAGGGGGAGGGGA GGAG	23	69,6	50	B	B6	G > T
2N164R	2164A > G	GAAACATCCCCCCTCACCAGCCA	23	60,9	23	C	C1	T > C
2N652R	2652G > C	TTTTTCCAACCAGAGTCTTCCCCTTGGT	24	54,2	30	C	C2	C > G

2N586F	4586C > T	TTTTTTTTTTTTGGGGAGGAAGCAGCAGAGCCGAG	24	66,7	35	C	C3	C > T
2N290F	6847T > C	TTTTTTTTTTTTTTTTGGAGCCTCCCTGTGACTATCTCC	23	60,9	40	C	C4	T > C
2N988F	7545C > T	TTTTTTTTTTTTTTTTTTTTTTAGCTGGCAGAACGGCTGGGGTTG	23	65,2	45	C	C5	C > T
2N017F	7574A > C	TTTTTTTTTTTTTTTTTTTTTTTTTTCTGGTGGACACGTATCGAA GGAC	23	56,5	50	C	C6	A > C
2N503F	8070C > T	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGCCCCAGCCTCAGGT GCACTGAA	23	65,2	55	C	C7	C > T

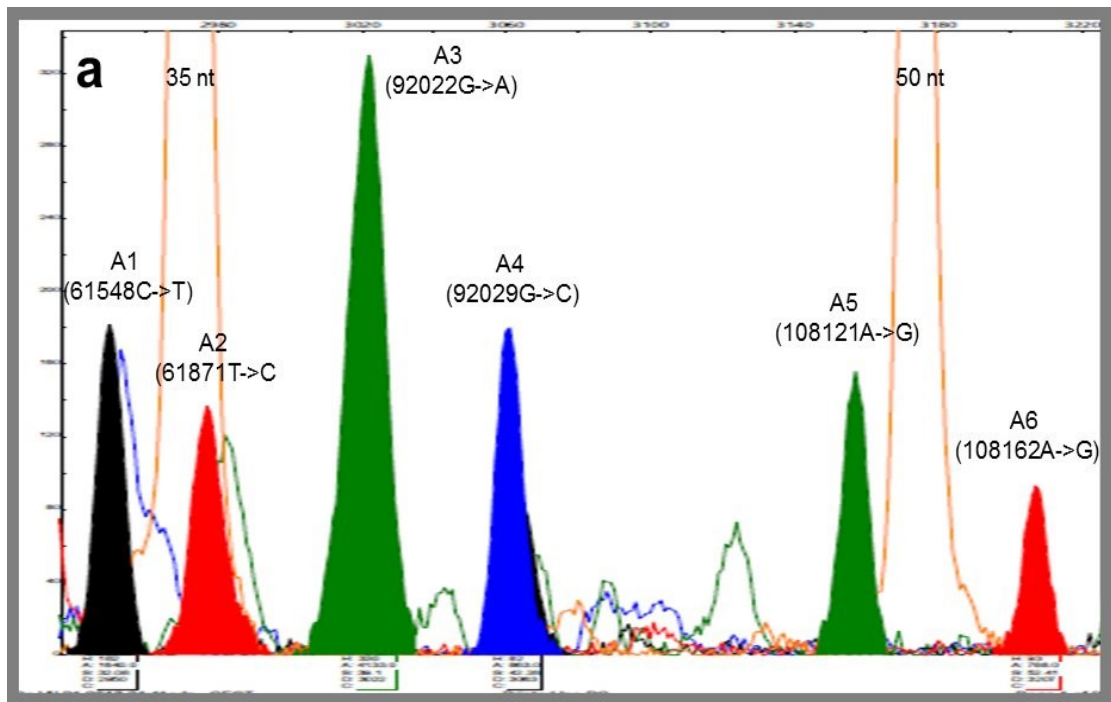
<sup>1</sup> Označení primerů vyjadřuje 1N = *NFKBI* + poslední tři číslice z lokalizace na referenční sekvenci + F nebo R pro přímý nebo reverzní primer

<sup>2</sup> U reverzního primeru se pozoruje komplementární vlákno

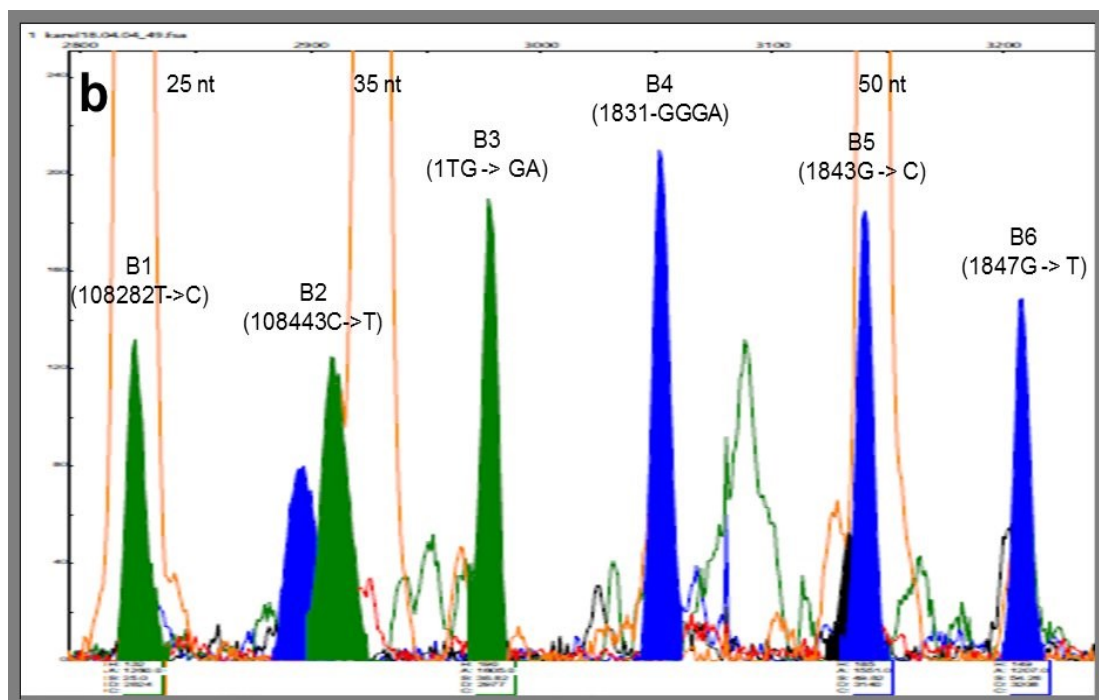
### **5.6.5 Validace genotypovacích reakcí pro NFkB1 a NFkB2**

Pro testování a optimalizaci reakcí SNaPshot byla použita DNA tří náhodně vybraných býků ve směsi, aby byla zvýšena heterozygotnost a byly potlačeny případné vlivy nízké kvality některého z jednotlivých preparátů. Vyšší heterozygotnost zvyšuje pravděpodobnost detekce peaků minoritních alel při optimalizaci a pojišťuje tak jejich spolehlivou identifikaci u jednotlivých vzorků.

Potom, co proběhlo testování genotypovacích primerů jednotlivě, byly primery testovány po skupinách, které odpovídaly předběžně určeným multiplexům. Příklady elektroferogramů fluorescenčně značených produktů reakce extenze primeru jsou uvedeny na obr. 13a, b.



Obr. 13a. Příklad multiplexu reakcí SNaPshot pro *NFKB1* a *NFKB2* (multiplex A)



Obr. 13b. Příklad multiplexu reakcí SNaPshot pro *NFKB1* a *NFKB2* (multiplex B)

### 5.6.6 Haplotypová struktura genu *TLR2* v populaci ČESTR

Haplotypová struktura v populaci byla určena i pro gen *TLR2* na základě dlouhých ampliconů PacBio. V tomto případě se podařilo určit poměrně rozsáhlé haplotypové skupiny ve čtyřech ampliconech (tab. 14). I když se objevuje řada minoritních variant, které by bylo možné přisoudit chybě v sekvenování, v ampliconu 1 je patrné zjevné rozdělení na dvě skupiny. Jedna skupina haplotypů v čteně samotné referenční sekvence se od referenční sekvence liší minimálně, ne více než jedním nukleotidem. Druhá skupina je sdružena kolem haplotypu č. 15, který obsahuje všechny opačné varianty referenční sekvence, tento trend ale již nepokračuje v dalších ampliconech. V ampliconech 2 a 4 převládá jeden majoritní haplotyp, který u ampliconu 2 ani neodpovídá referenční sekvenci.

I když se amplicony 1 a 2 a amplicon 2 a 3 překrývají dvojicemi SNP 624C>T - 657C>T - a 1044T > C - 1047G > T, návaznost frekvencí jednotlivých tříd je pozorována jen mezi haplotypy 1 a 2. Koncová dvojice 1044 a 1047 tvoří v ampliconech 2 a 3 třídy s odlišnou frekvencí. Toto pozorování je obtížné interpretovat a není vyloučena metodická chyba při zpracování souborů s multisekvencemi BAM a FASTA.

Praktickým důsledkem omezené shody frekvencí překrývajících se markerů je nemožnost odvodit haplotypy, které by pokrývaly více ampliconů, i když by to znamenalo významný úspěch metody přímého čtení

Tab. 14. Haplotypy určené ze sekvencí přečtených PacBio technologií u genu *TLR2* v populaci ČESTR. n - počet rozdílů proti referenční sekvenci, f - frekvence haplotypu.

Haplotyp	Amplikon číslo											
	amplikon 1			amplikon 2			amplikon 3			amplikon 4		
	SNP v haplotypu	n	f	SNP v haplotypu	n	f	SNP v haplotypu	n	f	SNP v haplotypu	n	f
	115T>C, 222A>T, 412G>A, 423G>A, 459T>C, 528G>A, 538A>C, 624C>T, 657C>T			624G > T, 657C > T, 817T > A, 825A > G, 825A > G, 1044T > C, 1047G > T			1044T > C, 1047G > T, 1313G > A, 1489A > G			2546G > A, 2565T > C, 2853C > G, 2883T > C		
1	T_A_G_G_T_G_A_G_C	0	68,9	G_C_T_A_T_G	0	4,8	T_G_G_A	0	15,2	G_T_C_T	0	75,3
2	C_A_G_G_T_G_A_G_C	1	1,9	G_C_A_A_T_G	1	4,8	C_G_G_A	1	10,9	A_T_C_T	1	1,1
3	T_T_G_G_T_G_A_G_C	1	1,2	G_C_T_G_T_G	1	4,8	T_T_G_A	1	10,9	G_C_C_T	1	5,9
4	T_A_A_G_T_G_A_G_C	1	2,1	G_C_T_A_C_G	1	4,8	T_G_G_G	1	17,4	G_T_G_T	1	1,5
5	T_A_G_G_G_G_A_G_C	1	0,7	G_C_T_A_T_T	1	47,6	C_T_G_A	2	2,2	G_T_C_G	1	1,5
6	T_A_G_G_T_A_A_G_C	1	1,9	G_T_T_A_T_T	2	4,8	C_G_G_G	2	6,5	A_C_C_T	2	2,1
7	T_A_G_G_T_G_G_G_C	1	1,5	G_C_A_A_T_T	2	9,5	T_T_G_G	2	17,4	G_C_G_T	2	2,4
8	T_A_G_G_T_G_A_T_C	1	5,2	G_C_T_G_T_T	2	4,8	T_G_A_A	2	4,3	G_C_C_C	2	1,1
9	T_A_G_G_T_G_A_G_T	1	2,4	T_T_T_A_C_G	3	4,8	C_T_A_A	3	2,2	A_C_G_T	3	3,0



## 5.7 Vyhodnocení populačních frekvencí

I když určení polymorfismu v genech *NFKB1* a *NFKB2* pomocí NGS bylo úspěšné a proběhl vývoj genotypovacích reakcí pro nalezené polymorfismy, nedošlo v rámci diplomové práce k aplikaci genotypovacích reakcí na populaci ČESTR v důsledku celkového harmonogramu genotypování. Nicméně byla ukončena práce na genotypování polymorfismů v genu *TLR2*, protože bylo možné využít již z větší části dokončené genotypovací metody, které byly založeny na částečných údajích o polymorfismu *TLR2* v populaci mléčného skotu.

Genotypování variant v genu *TLR2* bylo provedeno u 96 býků ze sledované populace ČESTR. Nalezené genotypy jsou shrnuty v tab. 15.

Tab. 15. Distribuce variant genu *TLR2* u býků populace ČESTR na základě individuálních genotypovacích reakcí

Býk číslo	115 T>C	513 C>T	1009 A>G	1060 G>A	1044 T>C	1047 T>G	1313 G>A	2546 G>A	2565 T>C	2883 T>C	3206 G>A
1			AG	GG	TC	GG	GA				GG
2											
3	CT	TT	AG	GA	TC	TG	GG	GA	TC	TC	GA
4			AG	GG	TC	TG	GA				
5	CT	TT	AG	GA	TC	TG	GG	GA	TC	TC	GA
6	CT	TT	AG	GA	TC	TG	GG	GA	TC	TC	AA
7	CT	TT	AG	GA	TC	TG	GA	GA	CC	TC	
8	CT	TT	AG	GG	TC	TT	GG	GA	TT	TT	AA
9	TT	TT	GG	GA	TT	TT	GG	GA	TT	TT	GA
10	TT	TT	GG	GG	TC	TG	GG	GA	TC	TT	GA
11	TT	TT	GG	GG	TT	TG	GA	GA	TC	TT	GG
12			GG	GG	TC	TG	GA	GA	TT	TC	GA
13			AG	GG	TT	TT	GA	GA	TC	TC	
14			AG	GA	TC	TG	GG	GA	TC	TC	
15								GA	TT	TC	GG
16	TT	TT	GG	GG	TT	TG	GG	GA	TT	TC	





50											
51								GA	CC	TC	
52			AG	GA	TC	TG	GA	GA	TT	TT	GG
53					TT			GA	TT	TC	
54	TT	TT			TT			GA	TT	TT	GA
55								GA	TT	TT	
56	TT	TT	AG		TT	TT					
57	CC	TT	AA	AA	TT			GA	TC	TC	GG
58	TT	TT	GG	GG	TC	TG	GA	GA	TC	TC	GG
59	CT	TT	AA	AA	TT						
60	CT	TT	AG	GA	TC	TG	GG	GA	TT	TT	GG
61											
62								GA	TT	TC	
63								GA	TC	TT	AA
64								GA	TT	TT	
65			AA		TT			GA	TT	TT	GG
66								GA	TT	TC	GG
67	TT	TT	GG	GG	TT	TT	GG	GA	TC	TT	GG
68			GG	GG	TT	TG	GA	GA	TC	TC	
69	CT	TT	AG	GA	TC	TG	GG	GA	TT	TC	GG
70	TT	TT						GA	CC	TC	GG
71	CT	TT	AG	GA	TC	TG	GA	GG	TT	TT	GG
72			AG	GG	TT	TT	GA				GG
73											
74								GA	TT	TC	
75								GA	TT	TC	GG
76								GA	TT	TT	
77	TT	TT	AG	GG	TT	TG	GA	GA	TC	TC	GG
78			AG	GG	TT	TG	GA	GA	TT	TC	GG
79	TT	TT	AG	GG	TC	TG	GA	GA	TT	TT	GG
80	TT	TT	AG	GG	TT	TT		GA	TT	TT	GG
81	TT	TT						GA	TT	TC	GG
82	TT	TT						GA	TT	TC	GG

83	TT	TT	AG	GG	TT	TT	GA	GA	TC	TC	GG
84			GG	GG	TT	TG	GA				GG
85								GA	TC	TC	
86											
87								GA	TT	TT	
88								GA	TT	TC	
89			AG	GG	TC	TG	GG	GA	CC	TT	
90			AG	GG	TC	TG	GA	GA	TC	TC	
91			AG	GA	TC	TG	GA				
92											
93	TT	TT						GA	TT	TT	
94	TT	TT									
95			AG	GA	TC	TG	GA	GA	TC	TC	AA
96	TT	TT									

Údaje o genotypch pro gen *TLR2* byly použity pro výpočet základních populačních parametrů pomocí programu PowerMarker (tab. 16).

U řady polymorfismů (115T>C, 1060G>A, 1313G>A, 2565T>C) se distribuce odchyluje od rovnováhy Hardyho - Weinberga, jak by bylo třeba očekávat v případě neutrálních polymorfismů.

Pozorovaná heterozygotnost je u většiny markerů vyšší, než očekávaná. Nejvyšší rozdíl je u markeru 2565T>C. Pokud neuvažujeme možnost metodických chyb při genotypování (ke zdánlivé vysoké heterozygotnosti povedou falešné pozitivy), je možným důvodem i šlechtitelská strategie zaměřená proti riziku inbreedingu.

Nepřímo lze o chybných genotypovacích reakcích usuzovat na základě shody frekvencí nalezených pomocí individuálního genotypování a na základě zastoupení readů NGS (tab. 16).v těch případech, kdy je SNP přítomen v genotypovací sadě (v amplikonech PacBio a v genotypovací sadě se překrývalo pouze sedm SNP: 115T>C, 1044T>C, 1047T>G, 1313G>A, 2546G>A, 2565G>A, 2883T>C). Vyšší hodnoty frekvencí varianty u SNP 2546 a 2883 opět mohou indikovat chyby v genotypovací metodě, které zvýší frekvenci pozitivů. Stejně se ale může jednat o artefakt při provádění NGS, který vede k nižšímu zastoupení jedné z alel ve směsném vzorku. Přesto je potřeba považovat základní shodu ve frekvencích

získaných oběma navzájem nezávislými postupy pro jednu populaci za zjištění, které oba použité přístupy validuje.

Tab. 16. Populační parametry genu *TLR2* v produkční populaci mléčného plemene ČESTR

Polymorfismus	Frekvence varianty(%)	P pro test X <sup>2</sup> H-W	Heterozygotnost očekávaná	Heterozygotnost pozorovaná	PIC (informační obsah polymorfismu)	SNP určen v ampliconu PacBio	Frekvence varianty podle PacBio (% tab. 12)
115T>C	18,3	0,495	0,299	0,327	0,254	1,0	16,3
513C>T	100,0	-	0,000	0,000	0,000	-	-
1009A>G	61,3	0,001	0,475	0,677	0,362	-	-
1060G>A	26,3	0,471	0,387	0,424	0,312	-	-
1044T>C	27,3	0,003	0,397	0,547	0,318	-	-
1047T>G	37,3	0,000	0,468	0,712	0,358	-	-
1313G>A	25,5	0,068	0,380	0,473	0,307	3,0	17,7
2546G>A	48,0	0,000	0,499	0,959	0,375	4,0	9,5
2565T>C	30,7	0,290	0,425	0,373	0,335	4,0	20,3
2883T>C	27,7	0,001	0,401	0,554	0,320	4,0	9,9
3206G>A	16,1	0,017	0,270	0,186	0,234	-	-
Průměr	-	-	0,364	0,476	0,289	-	-

Individuální genotypovací data pro gen *TLR2* v produkční populaci ČESTR rovněž dovolila získat přehled o haplotypech, které jsou v populaci přítomny, pomocí statistického odvození nejpravděpodobnějších kombinací. Tab. 17 obsahuje výstup z programu PHASE, který obsahuje frekvence 39 předpokládaných haplotypů.

Tab. 17. Haplotypy *TLR2* odvozené podle distribuce SNP v populaci ČESTR (SNP v pořadí 115T>C, 513C>T, 1009A>G, 1044T>C, 1047T>G, 1060G>A, 1313G>A, 2546G>A, 2565T>C, 2883T>C, 3206G>A)

Číslo haplotypu	SNP	Frekvence (%)	Číslo haplotypu	SNP	Frekvence (%)
1	C_T_A_T_T_A_G_G_T_T_G	1,0	21	T_T_G_T_G_G_G_G_C_T_G	1,0
2	C_T_A_T_T_A_G_A_C_C_G	0,5	22	T_T_G_T_G_G_A_G_T_T_G	3,1
3	C_T_A_C_G_A_G_G_T_T_G	3,1	23	T_T_G_T_T_G_G_G_T_T_G	26,6
4	C_T_A_C_G_A_G_G_T_T_A	0,5	24	T_T_G_T_T_G_G_G_T_T_A	3,1
5	C_T_A_C_G_A_G_A_C_C_G	1,6	25	T_T_G_T_T_G_G_G_C_T_G	3,6
6	C_T_A_C_G_A_G_A_C_C_A	1,6	26	T_T_G_T_T_G_G_A_T_T_G	10,9
7	C_T_A_C_G_A_A_G_T_T_G	2,6	27	T_T_G_T_T_G_G_A_T_T_A	1,6
8	C_T_A_C_G_A_A_A_C_C_G	0,5	28	T_T_G_T_T_G_G_A_T_C_G	6,8
9	C_T_A_C_G_A_A_A_C_C_A	0,5	29	T_T_G_T_T_G_G_A_T_C_A	0,5
10	C_T_A_C_T_G_G_A_T_T_A	0,5	30	T_T_G_T_T_G_G_A_C_T_G	3,6
11	T_T_A_T_T_G_G_G_T_T_G	2,1	31	T_T_G_T_T_G_G_A_C_T_A	1,0
12	T_T_A_T_T_G_G_A_T_T_G	1,0	32	T_T_G_T_T_G_G_A_C_C_G	7,3
13	T_T_A_T_T_G_G_A_T_C_G	0,5	33	T_T_G_T_T_G_A_A_C_C_G	0,5
14	T_T_A_T_T_G_G_A_C_C_G	0,5	34	T_T_G_T_T_A_G_G_T_T_G	1,0
15	T_T_A_T_T_G_A_G_T_T_G	2,1	35	T_T_G_C_G_G_G_A_C_T_G	1,0
16	T_T_A_T_T_A_G_G_T_T_G	2,1	36	T_T_G_C_G_G_A_G_T_T_G	0,5
17	T_T_A_C_G_G_G_A_C_T_G	0,5	37	T_T_G_C_G_G_A_A_T_C_G	0,5
18	T_T_A_C_G_G_A_G_T_T_G	3,1	38	T_T_G_C_G_A_G_G_T_T_G	0,5
19	T_T_A_C_G_G_A_A_T_T_G	0,5	39	T_T_G_C_G_A_A_G_T_T_A	0,5
20	T_T_G_T_G_G_G_G_T_T_G	1,0			

Bohužel haplotypy, které bylo možné přečíst z NGS dat, a haplotypy rekonstruované na základě individuálního genotypování se překrývaly jen ve dvou dvojicích SNP: 1044 - 1047 a 2565 - 2883. Přesto je porovnání výsledků dvou přístupů uvedeno v tab. 18. V obou případech byla pozorována dobrá shoda výsledků dvou odlišných metod. V případě markerů

1044 a 1047 byly frekvence určeny podle amplikonu 2 genu *TLR2*; pokus o porovnání s frekvencemi získanými přímým odečítáním z amplikonu 3 vedly ke značným rozdílům.

Tab. 18. Shoda vybraných haplotypů v *TLR2* nalezených metodou NGS a statistickým odvozením

1044_1047 <sup>1</sup>			2565_2883		
Haplotyp	Přímo (v PacBio readech)	Fázování podle genotypů	Haplotyp	Přímo (v PacBio readech)	Fázování podle genotypů
T_G	14,3	5,2	T_T	79,1	67,7
T_T	76,2	76,6	T_C	0	8,3
C_G	9,5	17,7	C_T	13,7	10,9
C_T	0	0,5	C_C	7,2	13,0

<sup>1</sup> Frekvence byla stanovena podle amplikonu *TLR2* č. 2.

Zatímco frekvence T\_T, T\_C, C\_T a C\_C byla 67,7, 8,3, 10,9 a 13,0% podle genotypování v populaci, ve směsném vzorku DNA byly detekovány pouze haplotypy T\_T, C\_T a C\_C s frekvencemi 79,1, 13,7 a 7,2%. Stejně jako v případě rozdílů v nalezených frekvencích SNP, příčinou může být menší spolehlivost genotypovacích reakcí typu SNaPshot a výskyt falešných pozitivů při genotypování. Spolehlivost genotypovacích reakcí je ověřována průběžně a teprve po jejich validaci na větších počtech vzorků je lze považovat za dostatečně spolehlivé.

### 5.8 Kategorizace nalezených polymorfismů a predikované funkční dopady

Zatímco v genu *NFKB1* (tab. 7) bylo 16 polymorfismů v intronech nebo v 3'-nepřekládané oblasti a 5 záměn bylo synonymních, tzn. nevedly ke změně aminokyseliny v bílkovinném produktu, pouze jedna mutace vedla k záměně aminokyseliny G474R. Nicméně jedna ze synonymních mutací je lokalizovaná ve 2. nukleotidu exonu 21, takže lze předpokládat vliv na splicing mRNA. Nicméně nebylo v možnostech diplomové práce tento efekt prověřit.

Většina pozorovaných polymorfismů, 13 z celkového počtu 22, je již zřejmě zastoupena v databázi variant Variant Browser v EBI, nicméně tvrzení je pouze přibližné. Data projektů v databázi zřejmě nebyla mapována na aktuální verzi bovinního genomu UMD\_3.1.1, proto

se pozice lišily o 1 - 3 nt. Navíc se SNP objevují jako reverzní komplement. Shodu bylo třeba v řadě případů předpokládat, i když s velkou pravděpodobností.

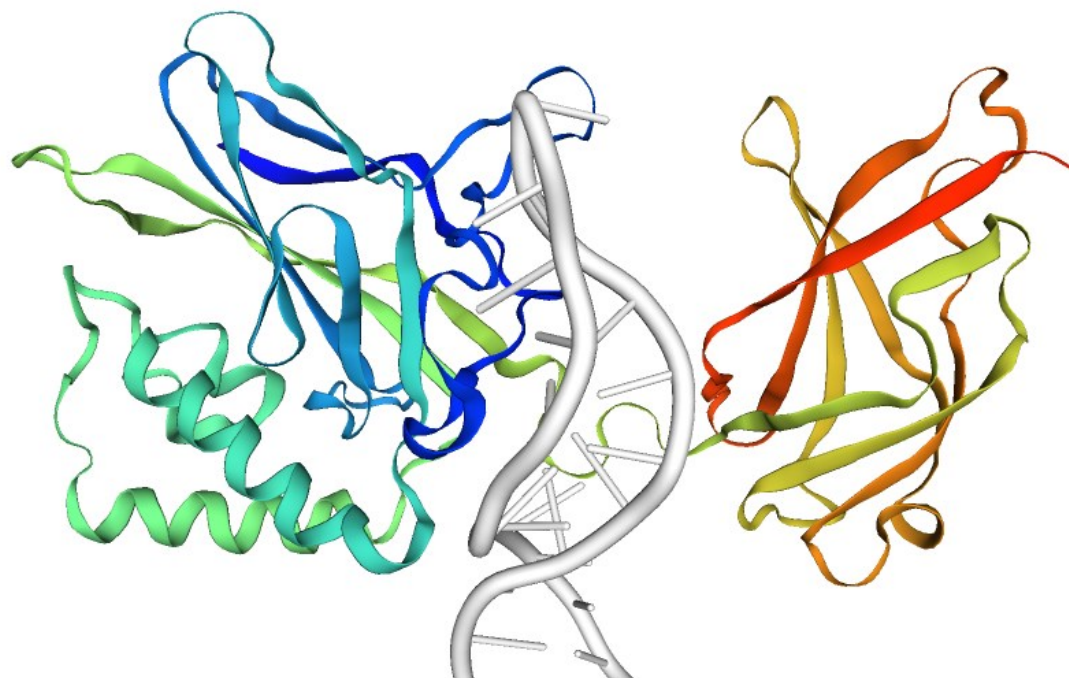
V genu *NFKB2* (tab. 8) bylo nalezeno 10 polymorfismů v intronech nebo v 3'-nepřekládané oblasti a zbývající dvě záměny byly synonymní. Na druhou stranu byly pouze 3 varianty v databázi EBI a 9 mutací je třeba považovat za nové.

V genu *TLR2* (tab. 9) bylo 11 polymorfismů v 5'-nepřekládané oblasti, 1 záměna v 3'-nepřekládané oblasti a 4 záměny byly synonymní. Nicméně 5 mutací vedlo k záměnám aminokyselin: E63ED, R152Q, I211V, R563H a H665Q. Pouze 3 polymorfismy z 21 byly nové.

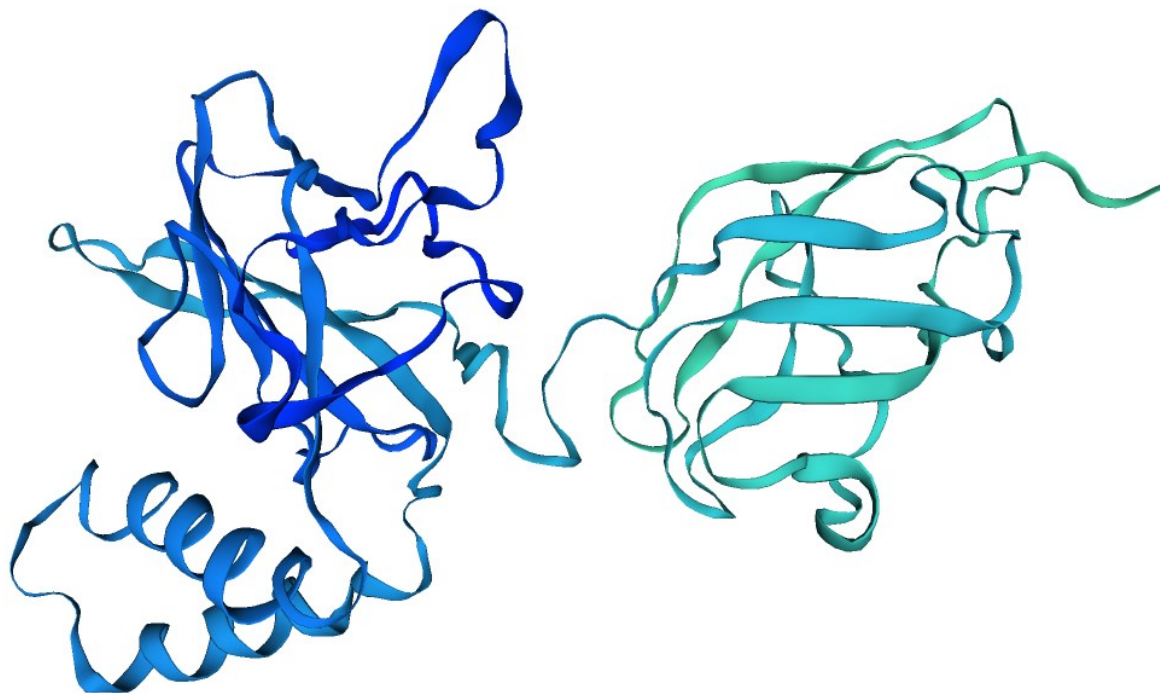
Ve snaze vizualizovat polohu mutací v produktech tří studovaných genů byly připraveny sterické modely (obr. 14, 15). Bohužel, rozsah modelů byl zejména u NF- $\kappa$ B1 a NF- $\kappa$ B2 omezen rozsahem templátů, které byly v databázi SwissProt k dispozici. Proto nebylo možné záměnu G474R v molekule NF- $\kappa$ B1 lokalizovat. Přesto byly pro obě subjednotky transkripčního faktoru získány modely s charakteristickou strukturou bílkoviny interagující s DNA (obr. 14a, b).

V rozsahu modelu extracelulární části bílkoviny TLR2 se podařilo lokalizovat tři záměny aminokyselin v důsledku mutací: R152Q, I211V a R563H (obr. 15b). Charakteristická struktura Toll-like receptoru je pozorovatelná na modelu produktu referenční sekvence *TLR2* (obr. 15a).

Lze konstatovat, že zobrazené tři záměny nepostihují konkávní část bílkoviny, která je esenciální pro rozpoznávání patogenů. Nelze očekávat dramatické změny ve funkci receptoru i z toho důvodu, že ve třech případech popsaných mutací dochází záměně v rámci stejné skupiny: nepolární isoleucin 211 na valin, kyselina glutamová na asparagovou na pozici 63, a bazický arginin 563 na histidin. Funkční změny nejsou vyloučené při záměně bazického argininu 152 a rovněž bazického histidinu 665 na glutamin.

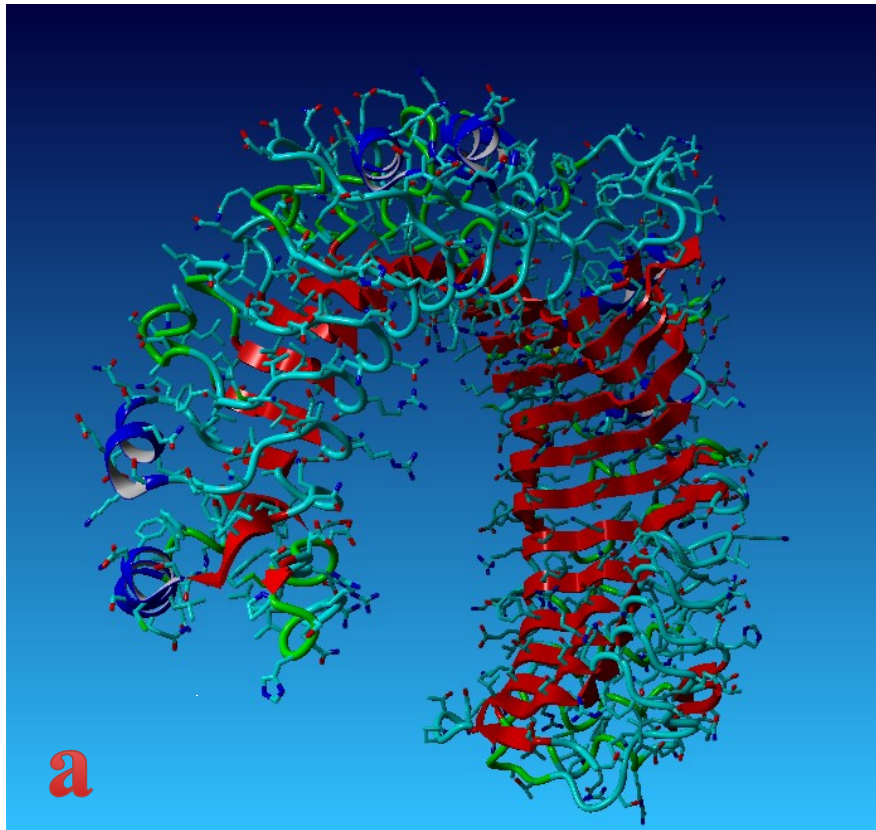


Obr. 14a. Modelovaná část subjednotky 1 transkripčního faktoru NF- $\kappa$ B

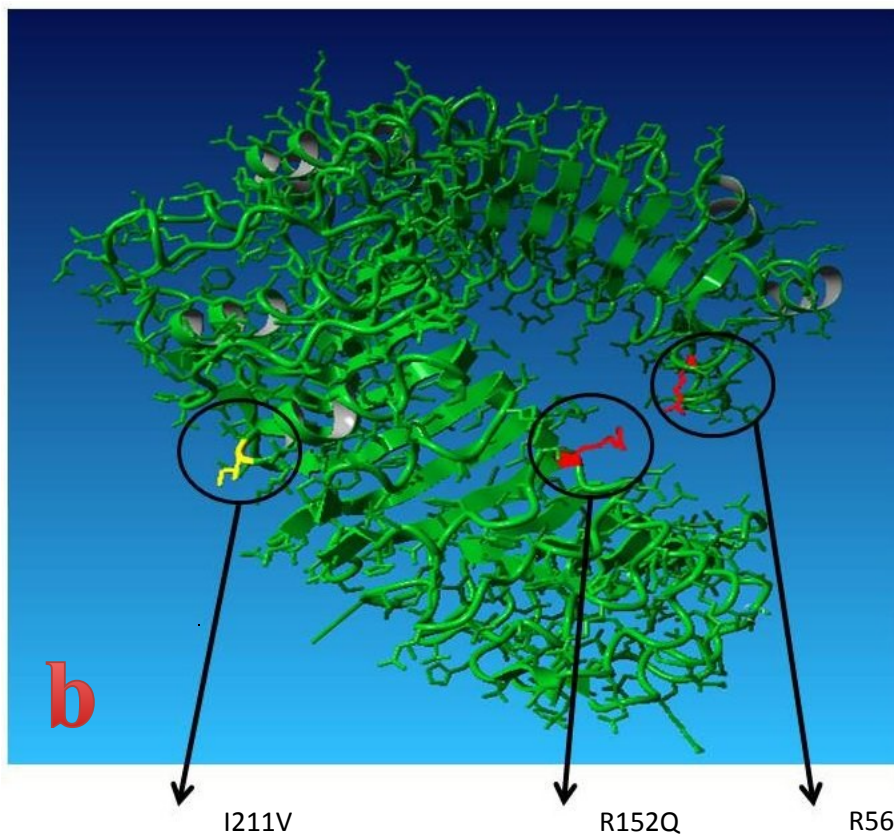


Obr. 14b. Modelovaná část subjednotky 2 transkripčního faktoru NF- $\kappa$ B





Obr. 15a. Modelovaná extracelulární část TLR2 - referenční sekvence. Červeně - opakování bohatá na leucin (LRR)



Obr. 15b. Extracelulární část TLR2 s vyznačenými nalezenými mutacemi

## 6 DISKUZE

### 6.1 Úloha zvolených genů

V rámci diplomové práce byl navržen a použit systém pro selektivní amplifikaci dvou klíčových genů skotu pro složku přirozené imunity NF- $\kappa$ B1 a -2, které kódují dvě subjednotky transkripčního faktoru NF- $\kappa$ B, a genu pro Toll-like receptor 2. Byl získán přehled o polymorfismu těchto genů v populaci českého strakatého skotu a byly zhodnoceny možnosti pro využití tohoto polymorfismu pro účely zvýšení rezistence k infekcím.

Výběr studovaných genů vycházel z klíčové úlohy TLR genů a jejich signální dráhy. TLR jsou transmembranové receptory, které díky extracelulární LRR části rozeznávají PAMPs. Přitom výsledky na laboratorních modelech přesvědčivě prokázaly, že genetické zásahy do těchto receptorů signální dráhy, v krajních případech nulové mutace, vedou ke ztrátě infekční rezistence (Chapes et al., 2001). Bylo demonstrováno, že na buněčné úrovni (leukocytární frakce) nesynonymní mutace v ovčích genech *TLR1* (Ser150Gly a Val220Met) a *TLR2* (Phe670Leu) mohou mít jako důsledek neschopnost vázat ligandy z patogenní bakterie *Mycobacterium avium* (Bhide et al., 2009).

### 6.2 Možnost přenosu metod

Předpokladem pro úspěšnou realizaci práce byl vývoj systému na resekvenování zvolených imunitních genů. Návrh sad amplifikačních primerů byl stoprocentně efektivní díky dodržení zásad pro individuální design:

Primery byly testovány na autokomplementaritu a na komplementaritu s dalšími primery v reakční směsi.

3' konec nesl pokud možno C nebo A,

délka amplifikačních primerů byla mezi 18- 25 nukleotidy, G/C% ideálně mezi 40-60%,  $T_m$  mezi 55 a 65°C.

I když všechny navržené primery fungovaly, v některých případech byla pozorována nízká specificita, například u amplikonů číslo 5, 6, 7 a 21 na obr. 7 (*NFKB1*). Podíl nespecifických produktů se podařilo snížit zvýšením hybridizační teploty v PCR. Proto lze považovat vyvinuté systémy amplikonů za komplexní řešení použitelné v obdobných projektech zaměřených na studium diverzity faktoru NF- $\kappa$ B u skotu.

Práce nezávisle validovala technologii sekvenování nové generace Pacific Biosciences pro resekvenování amplikonů. I přes zvýšenou chybovost metody, která je někdy považována za překážku pro její úspěšné použití pro hledání variant, bylo možné efektivně

hledat polymorfismus v poměrně a rozsáhlé populaci a ve vzorcích, které sdružovaly několik desítek genů. I v tomto případě lze předpokládat převzetí vyvinutých postupů analogickými projekty.

Za snadno aplikovatelné řešení lze považovat i vyvinuté a odzkoušené sady genotypovacích reakcí pro obě podjednotky transkripčního faktoru.

### 6.3 Kontext nalezeného polymorfismu

Je třeba uvést, že nalezená diverzita v genech *NFKB1* a *NFKB2* je spíše nižší, než lze očekávat pro geny skotu. Protože počet polymorfismů známých pro světovou populaci se odhaduje na 100 mln, je teoreticky nejvyšší hustota polymorfismů 1 varianta na 30 nukleotidů. Při celkové délce amplifikovaných a sekvenovaných úseků 8783 nt pro *NFKB1* a 7293 nt pro *NFKB2* lze očekávat při celkové světové diverzitě až 290 a 240 polymorfismů pro tyto geny. Nicméně údaje pro konkrétní mléčná plemena budou výrazně nižší. K úplným údajům v rámci mezinárodního projektu 1000 Bulls aktuálně není přístup.

Jediná dostupná publikace o polymorfismu členů NF- $\kappa$ B dráhy u skotu je článek autorů Chen et al. (2015). V práci byl sledován vliv dvou SNP v intronu 6 a v intronu 9 *NFKB1* a SNP v exonu dvacet *NFKB2* na počet somatických buněk v mléku holštýnského skotu.

Většina mutací v *NFKB1* a *NFKB2* byla v nepřekládaných úsecích genů, i když ty byly navrženým amplifikačním systémem pokryty pouze okrajově. Malý výskyt mutací v kódujících oblastech svědčí o vysoké míře konzervace a stabilizujícím selekčním tlaku na oba dva geny. Rovněž v práci autorů Vinkler et al. (2015) byla nalezená diverzita v genech *TLR4*, -5 a -7, která byla nižší než očekávaná pro volně žijící druhy, spojována se silnou negativní selekcí působící na *TLR* geny.

Skutečnost, že u *NFKB1* byly nalezeny nové polymorfismy u standardního mléčného plemene, ukazuje na dosud neúplnou znalost diverzity tohoto genu a odpovídá i minimálnímu počtu publikací o tomto genu u skotu. Převaha nových polymorfismů při resekvenování bovinního *NFKB2* je neobvyklá a odpovídá absenci publikací o tomto genu.

Na druhou stranu všechny mutace nalezené v genu *TLR2* již byly detekovány resekvenováním v rámci jiných projektů a zařazeny do databází polymorfismu. Je to v souladu se skutečností, že se jedná o tradičně intenzivně studovaný gen v rámci souboru deseti genů pro Toll-like receptory u skotu (Bhaladhare A. et al., 2016; Jann et al., 2008; Juste et al., 2018; Mucha et al., 2009 aj.). K velkému počtu publikací, které se zabývají

diverzitou bovinního *TLR2*, určitě přispívá i skutečnost, že je součástí heterodimerů s *TLR1* a *TLR6* a tak nezbytný pro jejich plnou aktivitu (Schenk et al., 2009).

Zvýšená heterozygotnost pro gen *TLR2* ve sledované populaci zasluhuje vysvětlení. Možným důvodem je snaha plemenářů o maximální heterozygotnost v populaci, což se promítá do přípařovacích plánů. I když plemenářské programy, které mohou zvyšovat heterozygotnost, existují (např. CRV - FITbull aj.) a jako vstupní data používají polymorfismus mikrosatelitů či obecný polymorfismus na 54K microarrayich Illumina, není známo, že by byl tento přístup plošně používán pro plemeno ČESTR. Tento předpoklad bude nicméně nezbytné prověřit jak na základě dat pro neutrální geny, tak i dotazy na plemenářských pracovištích.

#### **6.4 Haplotypová diverzita**

V rámci práce byly rovněž získány poznatky o sdružení přítomných polymorfismů do haplotypů. Pro interpretaci nalezených polymorfismů se jedná o důležitý údaj, neboť dovoluje rozhodnout, v jaké míře diverzita jednotlivých polymorfismů odpovídá diverzitě alel. I značný počet polymorfismů může být projevem přítomnosti pouhých dvou haplotypů, což by mohl být případ haplotypů 1 a 15 v amplikonu 1 genu *TLR2* – oba haplotypy obsahují alternativní nukleotidy ve všech pozicích. Pro šlechtitelské účely znalost haplotypů dovoluje určit, zda lze dvě mutace v jednom genu disociovat či zkombinovat, zjednodušuje genotypování pro účely asociačních studií nebo pro účely introgrese žádoucích variant (MAP, „marker assisted selection“).

Poměrně nepřesná shoda haplotypů přečtených přímo z variant readů NGS směsného populačního vzorku gDNA a haplotypů odvozených z distribuce SNP *TLR2* je zřejmě vysvětlitelná malým rozsahem souboru zvířat genotypovaných individuálně (96 býků) či případnými artefakty v metodě NGS či nepřesnostmi v genotypovací metodě SNaPshot. Na druhou stranu pozorovaná míra shody jak ve frekvencích SNP, tak i haplotypů odvozených oběma způsoby ukazuje, že získané výsledky jsou využitelné pro asociační studie s fenotypovými znaky a pro šlechtitelské aplikace.

Poměrně vysoký počet haplotypů pozorovaný v genu *TLR2* v této práci (15 při přímém čtení NGS v samotném amplikonu 1, 39 odvozených z individuálních dat pro 96 býků) sice překvapuje, ale zdá se být ve shodě s již publikovanými studiemi.

Haplotypová struktura genu *TLR2* byla označena za variabilnější než v případě dalších genů, jmenovitě *TLR4* a *TLR6*, i ve studii čtyř místních plemen Turecka a holštýnského

skotu (Bilgen et al. 2016). Z exonových SNP bylo odvozeno 36 haplotypů, z nich ale pouze 4 byly přítomny u holštýnského skotu. Shodně s mými výsledky, v případě genu *TLR2* se vyskytovaly dva hlavní haplotypy. Autoři tuto vysokou diverzitu haplotypů *TLR2* u historických plemen spojují se selekčním tlakem způsobeným místním prostředím, přičemž tento účinek nebyl překryt intenzivní selekcí na užitkovost a inbreedingem jako v případě vysoce produktivního holštýnského plemen. Nicméně plnohodnotná studie, která by tento předpoklad potvrdila, nebyla zatím provedena.

Rovněž dvě kontrastní skupiny haplotypů nalezené v amplikonu 1 u *NFKB1* mohou odrážet introgresi alternativního haplotypu při vzdálené hybridizaci v minulosti či ztrátu přechodových haplotypů při snížení početnosti populace v minulosti (Booker et al., 2017). Rozlišení obou případů je určitě námětem pro navazující studii.

Vysokou diverzitu haplotypů v *TLR2* dokládají i další práce. Ruiz-Llaranaga et al., (2011) našli v *TLR2* u holštýnského skotu celkem devět haplotypů. Pět různých haplotypů bylo rovněž určeno na základě sedmi SNP soustředěných v TIR oblasti *TLR2* u poměrně malé populace pěti plemen (Mucha et al., 2008).

Pro porovnání lze vzít počet haplotypů nalezených v nejvíce studovaném genu z *TLR* série, genu *TLR4*, kde byl počet haplotypů 20 ve světové populaci (White et al., 2003) a 9 v subpopulaci českého strakatého skotu (Novák et al., 2017).

## 6.5 Funkční dopady diverzity ve studovaných genech

Funkční dopady jednotlivých variant *TLR2* se pokusili předpovědět Jann et al. (2008) na základě lokalizace a typu aminokyselinových záměn. Základem bylo vymezení oblastí vykazujících signály pozitivní selekce. Pomocí koeficientu  $\omega$ , což je poměr nesynonymních a synonymních nukleotidových substitucí pro danou pozici v mezidruhovém porovnání, vymezili oblasti *TLR2*, v nichž u skotu (a *Bovidae* obecně) probíhala pozitivní selekce a oblasti neutrální. Aminokyselinovým záměnám v oblastech pozitivní selekce byly tak přisuzovány větší funkční následky. Tímto postupem byly identifikovány mezi 20 nesynonymními mutacemi identifikovány tři aminokyselinové záměny, konkrétně Leu227Phe, His305Pro a His326Gln, v oblasti s pozitivní selekcí. Tato oblast odpovídala v doménám účastnících se rekognice (cca aa 260 – 360). Z nich je His326Gln považován potenciálně za neúčinnější záměnu, neboť dochází ke změně náboje a navíc v pozici, která je klíčová pro vazbu ligandu u lidského *TLR2*. Kupodivu nižší koeficient  $\omega$  byl v doménách zodpovědných za dimerizaci heterodimerů (cca aa 325 – 400).

Funkční dopady aminokyselinových záměn v TLR2 se snažili predikovat i Bilgen et al., (2016) na základě přesné lokalizace v proteinových modelech. Byly tak lokalizovány i záměny, které byly nalezeny v populaci ČESTR v rámci diplomové práce. E63D (rs55617172) byla umístěna do LRR1, R152Q (rs43706434) do LRR5, I211V (rs43706433) do LRR7, R563H (rs68268260) do LRR20 a H665Q (rs68268263) do TIR oblasti. Umístění všech pozorovaných polymorfismů v oblastech TLR2 kritických pro rekognici a pro přenos signálu zvyšuje pravděpodobnost, že budou prokázány efekty na zaky infekční odolnosti v probíhajících populačních studiích.

### 6.6 Zdravotní dopady diverzity ve studovaných genech

V jediné publikaci o vlivu polymorfismu v *NFKB1* a *NFKB2* na zdravotní znaky skotu autorů Chen et al. (2015) byl sledován vliv dvou SNP v intronu 6 a v intronu 9 *NFKB1* a SNP v exonu dvacet *NFKB2* na počet somatických buněk v mléku holštýnského skotu. Počet somatických buněk je standardně sledován jako ukazatel zánětlivých procesů vemene. Mutace v intronu 9 *NFKB1* a exonová mutace *NFKB2* vykazovaly vliv na počet somatických buněk, navíc obě mutace v intronech *NFKB1* ovlivňovaly ukazatel obsahu tuku. Proto autoři doporučili použít tyto SNP jako markery v selekci na infekční odolnost. Rovněž byl zaznamenán pozitivní vliv varianty v Rel A (p65) jako interaktoru NF-κB1 a -2. Na druhou stranu nebyl zaznamenán efekt mutací ve čtyřech inhibičních faktorech IκB, rovněž interaktorech NF-κB1 a -2.

Protože SNP v intronech 6 a 9 *NFKB1* odpovídají po přepočtu referenční sekvence SNP 75795C>T a 79846T>G detekovaným v populaci ČESTR, je tento poznatek přenositelný do šlechtitelské praxe. Nicméně validace bude provedena v rámci plánované asociační studie s populačními zdravotními daty.

Shodu publikované mutace v *NFKB2* s některým ze SNP nalezených v populaci ČESTR se nepodařilo potvrdit, nicméně není vyloučeno shodu s pozorovanou variantou 7545C>T, pokud bude určena příčina posunu referenčních sekvencí.

Zdravotní efekt jednotlivých variant v *TLR2* na populační úrovni byl již zkoumán ve větší míře, i když nesouvisle.

Ruiz-Llaranaga et al. (2011) našli v *TLR2* u holštýnského skotu celkem 9 haplotypů, ale jen jeden vzácný haplotyp měl častější výskyt u zvířat infikovaných paratuberkulózou. Přitom ale pro jednotlivé alely či genotypy nebyla asociace prokázána.

Ve vřbec první práci na toto téma byla potvrzena asociace záměny Ile680Val v TLR2 s náchylností k paratuberkulóze skotu (Mucha et al., 2008). Další testované záměny v TIR oblasti TLR2, jmenovitě His697Arg, Asn706Ser, Lys709Arg, Ile745Val a Lys759Arg, statisticky významný efekt nevykázaly.

Bhaladhare et al. (2016) sledovali vliv tří polymorfismů v *TLR2* (rs55617172 = E63D, rs43706434 = R152Q a rs68268253) a čtyř polymorfismů v *TLR4* na odolnost k tuberkulóze skotu podle tuberkulinového testu. Jediný efektivní polymorfismus byl rs55617172 spojený se záměnou E63D. Změna alely G na T způsobovala vyšší citlivost na bovinní tuberkulózu. Je překvapivé, že stejný polymorfismus rs55617172 ovlivňoval náchylnost k mastitidě u mléčného skotu v dřívější práci autorů Zhang et al. (2009). Tento efekt lze považovat za reprodukovatelný, neboť byl opakovaně prokázán v nezávislé práci Prebavathy et al. (2015). Protože tento SNP byl nalezen i ve sledované populaci ČESTR, lze uvažovat o přenosu či validaci publikovaných poznatků.

V recentní publikaci (Juste et al., 2018) byl v populaci holštýnského skotu sledován vliv 6 genů spojených s imunitou na výskyt paratuberkulózy vyvolané *Mycobacterium avium*. Studie rozlišila i vliv kombinací variant v těchto šesti genech. Gen *TLR2* byl zastoupen variantami rs110491977, rs68268259, rs41830060, rs109971269, rs41830058, rs43706434 a rs43706433. Z nich jsou podle diplomové práce u ČESTR přítomny polymorfismy rs41830058 (2565T>C, synonymní), rs43706434 (1313G>A = R152Q) a rs43706433 (1489A>G = I211V). Nicméně v nejpříznivější kombinaci šesti genů byla synonymní varianta rs41830058. I synonymní mutace může být potenciálně vázána v haplotypu na další mutace s výraznějším efektem. Vzhledem k projevu této varianty v kombinaci variant šesti genů je šlechtitelská aplikace obtížná a z praktických důvodů málo pravděpodobná.

Obecně znalost diverzity imunitních genů ve studované populaci mléčného skotu otvírá možnost přenosu údajů o plemenné hodnotě jednotlivých variant z jiných publikovaných studií. Zde je ovšem třeba zachovat opatrnost, neboť hodnoty pro jednotlivé varianty se mohou výrazně měnit v závislosti na podmínkách chovů či infekční situaci.

## 7 ZÁVĚR

Prevenici nemocí předchází pochopení mechanismu fungování imunitního systému. Chceme-li pochopit imunitní systém, je nutné znát a pochopit jeho součásti. TLR jsou klíčové v imunitním systému, neboť na začátku Toll dráhy rozeznávají PAMPs a indukují imunitní odpověď odesláním signálu přes tuto dráhu až na NF- $\kappa$ B. Uvolněný transkripční faktor potom vstupuje do jádra a iniciuje mj. transkripci cytokinů, které indukují adaptivní imunitní odpověď. Možnost ovlivnit tento proces prostřednictvím genetické variability poskytuje nástroj pro zlepšení infekční rezistence u hospodářských zvířat obecně.

Nalezení nesynonymních mutací v populaci českého strakatého skotu dovoluje uvažovat o jejich efektu na infekční rezistenci. Rovněž se stává smysluplným i testování na možný efekt na expresi zdravotních znaků v populaci. Protože pro populaci českého strakatého skotu probíhá sběr zdravotních dat od r. 2016, budou údaje o nesynonymních mutacích v genech *NFKB1* a *TLR2* použity pro asociační studie. Nicméně i populační data pro zdánlivě neutrální mutace, pokud budou k dispozici, by měla být testována na asociaci s fenotypovými projevy. V současné době jsou k dispozici populační data pro *TLR2*, ale v budoucnu nesporně proběhne aplikace vyvinutých genotypovacích reakcí na reprezentativní sadu vzorků ČESTR, resp. dalších mléčných plemen skotu. To by mělo umožnit poměrně komplexní asociační studii polymorfismu v transkripčním faktoru NF- $\kappa$ B.



## 8 SEZNAM POUŽITÉ LITERATURY

- Adlhoch, C., Brouwer, A., Kuiken, T., Mulatti, P., Smietanka, K., Staubach, C., Guajardo, I.M., Verdonck, F., Amato, L., Baldinelli, F. (2018) Avian influenza overview February - May 2018. EFSA Journal 16.
- Akira, S., Uematsu, S., Takeuchi, O. (2006): Pathogen recognition and innate immunity. Cell. 124: 783-801.
- Arbibe, L., Mira, J.P., Teusch, N., Kline, L., Guha, M., Mackman, N., Godowski, P.J., Ulevitch, R.J., Knaus, U.G. (2000): Toll-like receptor 2-mediated NF-kappa B activation requires a Rac1-dependent pathway. Nature Immunology 1: 533-540.
- Ashari, M., Busono, W., Nuryadi, Nurgartiningasih, A. (2012): Analysis of chromosome and karyotype in Bali cattle and Simmental-Bali (Simbal) crossbreed cattle. Pakistan Journal of Biological Sciences 15: 736- 741.
- Bell, J.K., Mullen, G.E., Leifer, C.A., Mazzoni, A., Davies, D.R., Segal, D.M. (2003): Leucine-rich repeats and pathogen recognition in Toll-like receptors. Trends in Immunology 24: 528–533.
- Beutler, B. (2004): Innate immunity: an overview. Molecular Immunology 40: 845-859.
- Bhaladhare, A., Sharma, D., Kumar, A., Sonwane, A. (2016): Single nucleotide polymorphisms in toll-like receptor genes and case-control association studies with bovine tuberculosis. Veterinary World 9: 458–464.
- Bhide, M.R., Mucha, R., Mikula, I., Jr., Kisova, L. (2009): Novel mutations in TLR genes cause hyporesponsiveness to *Mycobacterium avium* subsp. *paratuberculosis* infection. BMC Genetics 10: 21.
- Bilgen, N., Kul, B.C., Offord, V., Werling, D., Ertugrul, O. (2016): Determination of genetic variations of Toll-like receptor (TLR) 2, 4, and 6 with next-generation sequencing in native cattle breeds of Anatolia and Holstein Friesian. Diversity 8: 23.
- Booker, T.R., Jackson, B.C., Keightley, P.D. (2017): Detecting positive selection in the genome. BMC Biology 15: 98.
- Bowie, A., O'Neill, L.A.J. (2000): The interleukin-1 receptor/Toll-like receptor superfamily: signal generators for pro-inflammatory interleukins and microbial products. Journal of Leukocyte Biology 67: 508-514.
- Cargill, E.J., Womack, J.E. (2007): Detection of polymorphisms in bovine toll-like receptors 3, 7, 8, and 9. Genomics 89: 745-755.

- Donnelly, C. A., Ghani, A. C., Ferguson, N. M., Anderson R. M. (1997): Recent trends in the BSE epidemic. *Nature* 389: 903.
- Du, X., Poltorak, A., Wei, Y., Beutler, B., (2000): Three novel mammalian toll-like receptors: gene structure, expression, and evolution. *European Cytokine Network* 11, 362–371.
- FAO/OIE (2011): Final Report of Joint FAO/OIE Committee on Global Rinderpest Eradication. FAO/OIE, Rome/Paris.
- Hall, T.A. (1999): BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*. 41: 95-98.
- Hayden, M.S., Ghosh, S. (2008): Shared principles in NF- $\kappa$ B signaling. *Cell*. 132: 344-362.
- Hoogendoorn, B., Owen, M. J., Oefner, P. J., Williams, N., Austin, J., O'Donovan, M.C. (1999): Genotyping single nucleotide polymorphisms by primer extension and high performance liquid chromatography. *Human Genetics*. 104: 89–93.
- Hořejší V. (2004): Receptory TLR - klíčové molekuly imunitního systému. *Vesmír* 83: 134.
- Huet, O., Choukroun, G., Mira, J.P. (2004): Récepteurs de type Toll, réponse inflammatoire et sepsis. *Elsevier SAS* 13: 167-175.
- Chapes, S.K., Mosier, D.A., Wright, A.D., Hart, M.L. (2001): MHCII, Tlr4 and Nramp1 genes control host pulmonary resistance against the opportunistic bacterium *Pasteurella pneumotropica*. *Journal of Leukocyte Biology* 69: 381-386.
- Chen, L., Han, Y., Chen, Y., Li, Z.X.; Wang, H.L., Liu, Y.; He, H.; Chen, S.; Liu, X.L. (2015): Eight SNVs in NF-kappa B pathway genes and their different performances between subclinical mastitis and mixed Chinese Holstein cows. *Gene* 555: 242-249.
- Choe, J., Kelker, M. S., Wilson, I. A. (2005): Crystal structure of human toll-like receptor 3 (TLR3) ectodomain. *Science* 309: 581-585.
- Choe, J., Kelker, M.S., Wilson, I.A. (2015): Crystal structure of human Toll-like receptor 3 (TLR3) ectodomain. *Science* 309: 581-585.
- Iwasaki, A., Medzhitov, R. (2010): Regulation of adaptive immunity by the innate immune system. *Science* 327: 291-295.
- Jann, O.C., King, A., Corrales, N.L., Anderson, S.I., Jensen, K., Ait-ali, T., Tang, H., Wu, C., Cockett, N.E., Archibald, A.L., Glass, E.J. (2009): Comparative genomics of Toll-like receptor signalling in five species. *BMC Genomics* 10: 1471-2164.
- Jann, O.C., Werling, D., Chang, J.S., Haig, D., Glass, E.J. (2008): Molecular evolution of bovine Toll-like receptor 2 suggests substitutions of functional relevance. *BMC Evolutionary Biology* 8: 288.

- Juste, R.A., Vazquez, P., Ruiz-Larrañaga, O., Manzano, M.I.C. (2018): Association between combinations of genetic polymorphisms and epidemiopathogenic forms of bovine paratuberculosis. *Heliyon* 4: e00535.
- Kaisho, T., Akira, S., (2006): Toll-like receptor function and signalling. *Journal of Allergy and Clinical Immunology* 117: 979-987.
- Kawai, T., Akira, S. (2010): The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. *Nature Immunology* 11: 373-384.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G. (2007): Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947-2948.
- Lemaitre, B., Nicolas, E., Michaut, L., Reichhart, J.M., Hoffman, J.A. (1996): The dorsoventral regulatory gene cassette *spatzle/Toll/cactus* controls the potent antifungal response in *Drosophila* adults. *Cell* 86: 973–983.
- McGuire, K., Jones, M., Werling, D., Williams, J.L., Glass, E.J., Jann, O. (2005): Radiation hybrid mapping of all 10 characterized bovine Toll-like receptors. *Animal Genetics* 37: 47–50.
- Miller, L.C., Lager, K.M., Kehrli, M.E. (2009): Role of Toll-Like Receptors in activation of porcine alveolar macrophages by porcine reproductive and respiratory syndrome virus. *Clinical and Vaccine Immunology* 16: 360-365.
- Misch, E.A., Hawn, T.R. (2008): Toll-like receptor polymorphisms and susceptibility to human disease. *Clinical Science* 114: 347-360.
- Mucha, R. Bhide, M.R., Chakurkar, E.B., Novak, M., Mikula, I., Sr. (2009): Toll-like receptors TLR1, TLR2 and TLR4 gene mutations and natural resistance to *Mycobacterium avium* subsp. *paratuberculosis* infection in cattle. *Veterinary Immunology and Immunopathology* 128: 381-388.
- Nilsson, M.A., Klassert, D., Bertelsen, M.F., Hallstrom, B.M., Janke, A. (2012): Activity of ancient RTE retroposons during the evolution of cows, spiral-horned antelopes, and nilgais (*Bovinae*). *Molecular Biology and Evolution* 29: 2885-2888.
- Novák, K. (2014): Functional polymorphisms in Toll-like receptor genes for innate immunity in farm animals. *Veterinary Immunology and Immunopathology* 2: 1-11.
- Novák, K., Czerneková, V. (2018a): Sada pro diagnostiku variant antibakteriálních *TLR* genů skotu. Užitečný vzor CZ 31484 U1. Majitel Výzkumný ústav živočišné výroby, v.v.i.

- Novák, K., Czerneková, V. (2018b): Sada pro diagnostiku variant antivirového genu TLR8 skotu. Užité vzor CZ 31486 U1. 2018-02-20. Majitel Výzkumný ústav živočišné výroby, v.v.i.
- Novák, K., Pikousová, J., Czerneková, V., Mátlová, V. (2017): Diversity of the TLR4 immunity receptor in Czech native cattle breeds revealed using the Pacific Biosciences sequencing platform. *Animal Biotechnology* 28: 228-236.
- Opsal, M.A., Vage, D.I., Hays, B., Berget, I., Lien, S. (2006): Genomic organization and transcript profiling of the bovine toll-like receptor gene cluster *TLR6-TLR1-TLR10*. *Gene* 384: 45–50.
- Pålsson-McDermott, E.M., O'Neill, L.A.J. (2007): Building an immune system from nine domains. *Biochemical Society Transactions* 35: 1437-1444.
- Prebavathy, T., Thanislass, J., Dhanamma, L., Ganesan, R., Mukhopadhyay, H.K. (2015): Association between SNPs in TLR2 gene segment corresponding to LRR functional domain of TLR2 receptor and bovine mastitis. *Asian Journal of Animal Sciences* 9: 45-56.
- Rhoads, A., Au, K.F (2015): PacBio sequencing and its applications. *Genomics, Proteomics and Bioinformatics* 13: 278–289.
- Ruiz-Larrañaga, O., Manzano, C., Iriando, M., Garrido, J.M., Molina, E., Vazquez, P., Juste, R.A., Estonba, A., 2011. Genetic variation of toll-like receptor genes and infection by *Mycobacterium avium* ssp. *paratuberculosis* in Holstein-Friesian cattle. *Journal of Dairy Science* 94: 3635–3641.
- Seabury, C.M., Cargill, E.J., Womack, J.E. (2007): Sequence variability and protein domain architectures for bovine Toll-like receptors 1, 5, and 10. *Genomics* 90: 502-515.
- Seabury, C.M., Seabury, P.M., Decker, J.E., Schnabel, R.D., Taylor, J.F., Womack, J.E. (2010): Diversity and evolution of 11 innate immune genes in *Bos taurus taurus* and *Bos taurus indicus* cattle. *Proceedings of the National Academy of Sciences USA* 107: 151–156.
- Shendure, J., Balasubramanian, S., Church, G.M., Gilbert, W., Rogers, J., Schloss, J.A., Waterston, R.H. (2017): DNA sequencing at 40: past, present and future. *Nature* 550: 345-353.
- Schenk, M., Belisle, J.T., Modlin, R.L. (2009): TLR2 looks at lipoproteins. *Immunity* 31: 847-849.

- Stein, D., Roth, S., Vogelsang, E., Nusslein-Volhard, C. (1991): The polarity of the dorsoventral axis in the *Drosophila* embryo is defined by an extracellular signal. *Cell* 65: 725-735.
- Szafranski, K., Schindler, S., Taudien, S., Hiller, M., Huse, K., Jahn, N., Schreiber, S., Backofen, R., Platzer, M. (2007): Violating the splicing rules: TG dinucleotides function as alternative 3' splice sites in U2-dependent introns. *Genome Biology* 8:154.
- Takeuchi, O.; Akira, S. (2010): Pattern recognition receptors and inflammation. *Cell* 140: 805-820.
- Valanne, S., Wang, J.H., Ramet, M. (2011): The *Drosophila* Toll signaling pathway. *Journal of Immunology* 186: 649–656.
- Vinkler, M., Bainová, H., Bryjová, A. (2015): Characterisation of Toll-like receptors 4, 5 and 7 and their genetic variation in the grey partridge. *Genetica* 143: 101–112.
- Werling, D., Jungi, T.W. (2003): Toll-like receptors linking innate and adaptive immune response. *Veterinary Immunology and Immunopathology* 91: 1–12.
- White, S.N., Taylor, K.H., Abbey, C.A., (2003): Haplotype variation in bovine Toll-like receptor 4 and computational prediction of a positively selected ligand-binding domain. *Proceedings of the National Academy of Sciences USA* 100: 10364–10369.
- Woolhouse, M., Donaldson, A. (2001): Managing foot-and-mouth. The science of controlling disease outbreaks. *Nature* 410: 515-516.
- Zhang, L.P., Gan, Q.F., Ma, T.H., Li, H.D., Wang, X.P., Li, J.Y., Gao, X., Chen, J.B., Ren, H.Y., Xu, S.Z. (2009): Toll-like receptor 2 gene polymorphism and its relationship with SCS in dairy cattle. *Animal Biotechnology* 20: 87-95.
- Zipfel, C., Robatzek, S., Navarro, L., Oakeley, E. J., Jones, J.D.G., Felix, G., Boller, T. (2004): Bacterial disease resistance in *Arabidopsis* through flagellin perception. *Nature* 428: 764–767.