

JIHOČESKÁ UNIVERZITA V ČESKÝCH BUDĚJOVICÍCH

Fakulta zemědělská a technologická

Bioinformatika pro molekulární diagnostiku: integrace pokročilých
výpočetních metod v diagnostice patogenů

Habilitační práce

Prohlašuji, že jsem habilitační práci na téma „Bioinformatika pro molekulární diagnostiku: integrace pokročilých výpočetních metod v diagnostice patogenů“ zpracoval samostatně a s využitím vlastních publikovaných výsledků, které jsou nedílnou součástí této práce. Veškeré další zdroje a literatura, ze kterých jsem při psaní čerpal, jsou řádně citovány a uvedeny v seznamu použité literatury.

Dále prohlašuji, že nástroje na bázi velkých jazykových modelů byly při přípravě textu využity výhradně jako asistence pro dílčí jazykové a stylistické úpravy. Nesloužily k tvorbě samotného odborného obsahu práce, analýze dat ani jako zdroj informací.

V Českých Budějovicích dne 23. 9. 2026

Ing. Pavel Beran, Ph.D.

Na tomto místě bych rád poděkoval své manželce Kristině za její neúnavnou podporu v osobním i profesním životě. Velký vděk patří také mým rodičům, Janě a Slavomírovi, a sestře Martině, kteří mě vždy povzbuzovali a motivovali k dosažení co nejvyššího vzdělání. V neposlední řadě děkuji všem kolegům, a to jak z domovské katedry, tak i mimo ni, za spolupráci a za inspirativní atmosféru, kterou na pracovišti vytvářejí.

Abstrakt

Tato habilitační práce řeší aktuální výzvu moderní molekulární diagnostiky a hledá efektivní způsoby zpracování stovek dostupných genomů pro spolehlivé nalezení jedinečných diagnostických znaků patogenů. Tradiční výpočetní metody při takových objemech dat narážejí na své limity, a proto se text zaměřuje na integraci pokročilé bioinformatiky přímo do laboratorní a polní praxe. Výpočetní základ tvoří autorské algoritmy KEC a genCRC32, které nevyžadují zarovnávání sekvencí a místo toho využívají matematickou exkluzi a bezkolizní hashování k rychlé analýze rozsáhlých pangenomů. Zmíněné nástroje se staly základem uceleného pracovního postupu „Hundred-Genomes Pipeline“, pomocí kterého byly navrženy vysoce citlivé testy pro metody PCR a izotermální amplifikaci LAMP. Praktická funkčnost těchto postupů je následně demonstrována na rychlé a přesné identifikaci taxonomicky složitých fytopatogenních bakterií. Prezentované molekulární principy navíc prokazují značnou univerzálnost a nacházejí uplatnění i mimo obor fytopatologie, typicky při kontrole falšování potravin nebo ve šlechtění hospodářských zvířat. Celý proces od počítačového návrhu až po reálnou detekci se rovněž přímo promítá do univerzitní výuky a vedení závěrečných prací.

Abstract

This habilitation thesis addresses a current challenge in modern molecular diagnostics and seeks efficient ways to process hundreds of available genomes to reliably find unique diagnostic markers for pathogens. Traditional computational methods reach their limits with such massive data volumes, and therefore the text focuses on integrating advanced bioinformatics directly into laboratory and field practice. The computational foundation consists of the novel algorithms KEC and genCRC32, which do not require sequence alignment and instead utilize mathematical exclusion and collision-free hashing for the rapid analysis of extensive pangenomes. These tools became the basis for the comprehensive "Hundred-Genomes Pipeline" workflow, which is used to design highly sensitive assays for PCR and loop-mediated isothermal amplification (LAMP). The practical functionality of these procedures is subsequently demonstrated by the rapid and precise identification of taxonomically complex phytopathogenic bacteria. Furthermore, the presented molecular principles prove their universality and find applications beyond phytopathology, typically in food fraud detection and livestock breeding. The entire process, from computational design to real-world detection, is also directly reflected in university teaching and the supervision of student theses.

Obsah

Seznam použitých zkratk.....	7
Úvod	8
Výzvy moderní fytopatologie	10
Globální rizika a ekonomické dopady chorob.....	10
Taxonomie a plasticita genomu	11
Mezinárodní obchod a asymptomatický přenos	13
Diagnostické překážky a cíle.....	14
Evoluce molekulární detekce.....	16
Polymerázová řetězová reakce (PCR).....	16
Izotermální amplifikace LAMP	17
Cílené vyhledávání CRISPR-Cas	18
Bioinformatika a strategie <i>in silico</i> designu	20
Principy návrhu molekulárních testů	20
Bioinformatické výzvy.....	22
Algoritmická řešení a jejich integrace do diagnostické praxe	25
KEC: exkluze k-merů	25
genCRC32: Akcelerace analýzy pomocí hashování.....	26
Syntéza: Hundred-Genomes Pipeline	27
Budoucí vědecké směřování	28
Evoluce nástroje KEC.....	28
Dynamická diagnostika v reálném čase.....	28
Adaptace pro CRISPR-Cas12b.....	29
Strojové učení: Hledání skrytých vzorů v datech	29
Rozbor publikovaných prací.....	32
Vývoj autorských softwarových nástrojů.....	32
Aplikace molekulárních metod v diagnostice bakteriálních patogenů.....	45
Univerzálnost metod a jejich adaptace mimo obor fytopatologie	95
Aplikace poznatků ve výuce a pedagogické činnosti	161
Publikace s přímou vazbou na habilitační práci	163
Seznam použité literatury.....	164

Seznam použitých zkratk

- ABIF** – Applied Biosystems DNA sequencing format
ASCII – American Standard Code for Information Interchange
BIP – Backward Inner Primer
BLAST – Basic Local Alignment Search Tool
CRC32 – Cyclic Redundancy Check
CRISPR – Clustered Regularly Interspaced Short Palindromic Repeats
DNA – Deoxyribonucleic acid / Deoxyribonukleová kyselina
ELISA – Enzyme-Linked Immunosorbent Assay
FIP – Forward Inner Primer
FM-index – Ferragina-Manzini index
GUI – Graphical User Interface / Grafické uživatelské rozhraní
HLB – Huanglongbing / Zelenání citrusů
ITS – Internal Transcribed Spacer / Vnitřní transkribovaný mezerík
KEC – K-mer Exclusion by Crossreference
LAMP – Loop-mediated isothermal amplification / Izotermální amplifikace zprostředkovaná smyčkami
LFD – Lateral Flow Device
MSA – Multiple Sequence Alignment / Násobné zarovnání sekvencí
MUSCLE – Multiple Sequence Comparison by Log-Expectation
NCBI – National Center for Biotechnology Information
NGS – Next-Generation Sequencing / Sekvenování nové generace
NP-úplný – Nondeterministic Polynomial-time complete
PAM – Protospacer Adjacent Motif
PCR – Polymerase Chain Reaction / Polymerázová řetězová reakce
RAM – Random Access Memory / Operační paměť
RNA – Ribonucleic acid / Ribonukleová kyselina
SNP – Single Nucleotide Polymorphism / Jednonukleotidový polymorfismus
SSR – Simple Sequence Repeat / Jednoduchá sekvenční repetice (mikrosatelit)
ToBRFV – Tomato Brown Rugose Fruit Virus / Virus hnědé vrásčitosti plodů rajčete
ToLCNDV – Tomato Leaf Curl New Delhi Virus

Úvod

Rostlinná patologie prochází zásadní proměnou. Zatímco v minulosti byla diagnostika původců onemocnění rostlin omezena na pozorování vnějších příznaků či zdlouhavou kultivaci v laboratořích, dnešní možnosti sahají přímo k podstatě genetické informace. Díky pokroku v technologiích sekvenování nové generace (NGS) máme k dispozici obrovské množství dat. Pro většinu hospodářsky významných patogenů již neexistuje pouze jeden referenční genom, ale databáze obsahující stovky genomických sekvencí popisujících jejich globální variabilitu.

Tento informační nadbytek však přináší nový typ problému. Limitujícím faktorem diagnostiky přestává být schopnost data získat, ale stává se jím schopnost data efektivně zpracovat. Tradiční bioinformatické postupy, které byly vyvinuty pro srovnávání několika málo genů, selhávají při pokusu analyzovat stovky kompletních genomů současně. Výpočetní náročnost klasických algoritmů pro zarovnávání sekvencí (alignment) roste s množstvím dat natolik, že se pro moderní pangenomické studie stávají nepoužitelnými.

Předkládaná habilitační práce reaguje na tuto technologickou výzvu. Je zacílena na demonstraci možností, které poskytují nové, výpočetně efektivní postupy založené na principu exaktní matematické exkluze, jež v procesu návrhu diagnostických markerů nahrazují metody zastarávající. Práce se nezabývá pouze teorií. Jejím cílem je překlenout propast mezi bioinformatikou a reálnou zemědělskou praxí. Výsledkem tohoto snažení nejsou jen algoritmy, ale především funkční a robustní nástroje pro detekci patogenů.

Text je strukturován do logických celků, které kopírují proces vývoje diagnostického testu. První část definuje biologické a taxonomické výzvy, kterým čelíme – od kryptických druhů bakterií, jež v předkládaných pracích zastupuje primárně rod *Xanthomonas*, až po nutnost odlišit nebezpečné fytopatogeny od jejich neškodných příbuzných v mikrobiomu rostliny. Následně se práce věnuje samotnému řešení v rovině *in silico*. Představuje vývoj autorských nástrojů pro alignment-free analýzu, konkrétně algoritmus KEC pro vyhledávání unikátních

sekvencí a hashovací nástroj genCRC32, který umožňuje zefektivnění práce s masivními genomickými daty na běžném hardware.

Následující pasáže demonstrují aplikaci těchto metod v praxi. Ukazují, jak lze pomocí „Hundred-Genomes Pipeline“ navrhnout vysoce citlivé diagnostické testy pro metodu LAMP či PCR, které obstojí i v náročných podmínkách s přítomností inhibitorů. Tato část práce tak předkládá ucelený pohled na moderní diagnostiku: od prvotního stažení genomických dat, přes jejich algoritmické zpracování, až po finální detekci patogena na poli. Text však nezůstává pouze u technického řešení a v dalších částech zasazuje dosažené výsledky i do širšího kontextu budoucího výzkumu a pedagogické praxe.

Kapitola Budoucí vědecké směřování nastiňuje plánovaný vývoj prezentovaných nástrojů, zejména s ohledem na jejich automatizaci, adaptaci pro technologie CRISPR a integraci metod strojového učení. Následující část poskytuje rozbor publikovaných prací, který fakticky shrnuje přiložené publikace, dokumentující vývojovou linii od prvotních softwarových nástrojů a algoritmů až po jejich praktickou aplikaci v diagnostice bakterií rodu *Xanthomonas*. Celou práci uzavírá popis implementace bioinformatických postupů do studijních předmětů na Jihočeské univerzitě a jejich využití při vedení bakalářských a diplomových prací. Habilitační práce podává komplexní pohled na moderní diagnostiku a ukazuje, jak nezbytné je dnes propojení teoretické bioinformatiky s reálnou laboratorní praxí.

Výzvy moderní fytopatologie

Molekulární diagnostika stojí na univerzálním principu, který nerozlišuje mezi člověkem, mikroorganismem nebo rostlinou. Úkol je vždy stejný: přesně najít a rozpoznat konkrétní úsek genetické informace v záplavě ostatní DNA. Přestože se dosavadní výzkum autora této práce orientoval na různá témata, klíčová dovednost – tedy návrh citlivých a spolehlivých molekulárních markerů – zůstává neměnná. Tato práce se zabývá postupy precizního návrhu oligonukleotidů v rámci rostlinné patologie. Právě zde totiž narážíme na akutní nedostatek nástrojů, které by byly dostatečně robustní pro použití nejen v laboratoři, ale přímo v proměnlivých podmínkách zemědělské praxe.

Globální rizika a ekonomické dopady chorob

Stabilita světové produkce potravin je dnes pod obrovským tlakem a rostlinné patogeny zde hrají zásadní roli (Savary et al., 2019). Rostoucí mezinárodní obchod (a s tím související introdukce patogenů) v kombinaci s vývojem nových, agresivnějších kmenů patogenů způsobuje roční ekonomické ztráty kolem 220 miliard dolarů (Ristaino et al., 2021). Situaci navíc výrazně zhoršuje změna klimatu. Ta nejen otevírá cestu patogenům do nových oblastí, ale zároveň oslabuje přirozenou obranyschopnost rostlin (Hossain et al., 2024; Jagathjothi et al., 2024). U hlavních plodin, jako jsou obiloviny nebo rýže, přicházíme až o 40 % úrody, což přímo ohrožuje potravinovou bezpečnost v mnoha regionech světa (Savary et al., 2019).

Pohled do historie, například na Velký irský hladomor způsobený epidemií plísně bramboru, nám připomíná, že choroby rostlin mohou vyvolat rozsáhlé společenské krize (Ristaino et al., 2021). Dnešní hrozby mají sice spíše ekonomický rozměr, jejich dopad je ale drtivý. Ukázkovým příkladem je bakterie *Candidatus Liberibacter* způsobující chorobu HLB (Huanglongbing neboli zelenání citrusů), která zdecimovala produkci citrusů na Floridě o více než 90 % a způsobila ztráty v řádu miliard dolarů (USDA, 2024; Ellis et al., 2025). Takové

epidemie jasně ukazují, jak zranitelné jsou moderní monokulturní systémy, pokud nezasáhneme včas (CIRAD, 2024).

Rozhodujícím faktorem v boji proti těmto hrozbám je čas. Epidemiologické modely potvrzují, že zachytit infekci v počáteční fázi je finančně mnohem výhodnější než následná likvidace napadených porostů (Fielder et al., 2025; Wang et al., 2025). Například u nebezpečné bakterie *Xylella fastidiosa* by včasné nasazení detekčních metod mohlo v jižní Evropě ušetřit miliardy eur, jak naznačují aktuální predikce (Calderoni et al., 2025; Luchi et al., 2023). Výzkum se proto nyní soustředí na rychlé diagnostické platformy, jako je izotermální amplifikace LAMP (Loop mediated isothermal amplification) nebo systémy CRISPR. Tyto technologie umožňují vysoce přesnou detekci přímo na místě výskytu a překonávají tak propast mezi laboratorní přesností a potřebami terénní praxe (Serena et al., 2025; Velasco-Amo et al., 2026).

Taxonomie a plasticita genomu

Základním předpokladem úspěšné diagnostiky je přesně vědět, co je předmětem identifikace. V mikrosvětě rostlinných patogenů se však tento cíl v posledních letech pohybuje a mění rychleji než kdy dříve. Statická definice druhu, jak jsme ji znali z učebnic minulého století, se pod náporom moderních sekvenačních metod komplikuje. To, co se jevílo jako jediný monolitický patogen, se často ukazuje být komplexem mnoha blízce příbuzných taxonů, tzv. kryptických druhů. Tento fenomén "skryté diverzity" dnes komplikuje boj s chorobami napříč celým spektrem patogenů (Fielder et al., 2025).

Ukázkovým příkladem takové situace, která je v této práci patrná na více místech, je původce bakteriální skvrnitosti rajčete a papriky. Ještě poměrně nedávno byla za původce choroby označována bakterie zvaná *Xanthomonas campestris* pv. *vesicatoria*. Dnes je toto tvrzení taxonomicky nepřesné a diagnosticky nedostatečné. Fylogenetické analýzy tento taxon aktuálně dělí na čtyři samostatné druhy: *X. euvesicatoria* pv. *euvesicatoria*, *X. euvesicatoria* pv. *perforans*, *X.*

vesicatoria a *X. hortorum* pv. *gardneri* (Constantin et al., 2016; Timilsina et al., 2020).

Záludnost této situace spočívá v tom, že zatímco genetici vidí čtyři odlišné organismy, pěstitel na rostlině vidí stále tytéž příznaky. Fenotypová diagnóza zde naráží na své limity - na základě symptomatologie jsou tyto druhy prakticky nerozeznatelné (Pečenka et al., 2020). Přesto na přesném určení záleží. Jednotlivé druhy se liší agresivitou, hostitelským spektrem a co je nejdůležitější – schopností překonat mechanismy rezistence odrůd hostitelských rostlin. Například odrůda rajčete vyšlechtěná k odolnosti proti *X. euvesicatoria* pv. *euvesicatoria* může být velmi náchylná vůči *X. hortorum* pv. *gardneri*.

Další komplikací je dynamika samotného bakteriálního genomu. Bakterie nejsou geneticky neměnné. Naopak jsou to vysoce adaptabilní organismy schopné rychlých genomických přestaveb. Nedávné studie popisují vysokou plasticitu tzv. postradatelného (dispensable) genomu, kde dochází k zisku či ztrátě genů virulence (Navasca et al., 2025). V roce 2023 byla dokonce navržena reklasifikace, která přesouvá desítky paterů do nových druhových skupin, což do jisté míry činí starší diagnostické protokoly ze dne na den zastaralými (Harrison et al., 2023). Pokud používáme metodu cílenou na gen, který patogen v rámci evoluce ztratil nebo změnil, dostáváme falešně negativní výsledek a s ním falešný pocit bezpečí (Wang et al., 2025).

Právě zde narážíme na limity klasických metod a otevírá se prostor pro pokročilou bioinformatiku. Abychom v tomto taxonomickém labyrintu našli pevný bod, nemůžeme se již spoléhat na jeden či dva geny. Potřebujeme analyzovat celou genetickou informaci, tedy kompletní genom. V této práci byl proto uplatněn přístup komparativní genomiky, kdy srovnáním stovek genomů byly hledány ty vzácné úseky DNA, které jsou stabilní a unikátní pro daný taxon, bez ohledu na taxonomické zařazení. Tento problém "hledání jehly v kupce sena" pomocí algoritmů umožnil vyvinout robustní nástroje pro detekci *X. hortorum* pv. *gardneri* či rozlišení druhů v komplexu *X. euvesicatoria* a *X. vesicatoria* (Beran et al., 2022; Stehlíková et al., 2020).

Mezinárodní obchod a asymptomatický přenos

Zatímco taxonomická nestabilita komplikuje definici patogena, asymptomatický průběh infekce zásadně ztěžuje jeho fyzický záchyt. V kontextu globalizovaného zemědělství, kdy se rostlinný materiál a osivo přesouvají mezi kontinenty ve velkých objemech, vizuální fytosanitární kontrola poskytuje pouze nedostatečný záchyt. Řada závažných fytopatogenů je schopna přežívat v pletivech hostitele v latentní fázi, aniž by vyvolala vnější symptomy, přičemž rostlina zůstává infekční a stává se rezervoárem pro další šíření (Fielder et al., 2025).

Kritickým vektorem tohoto skrytého šíření je infikované osivo. Například virové patogeny, jako je Tomato brown rugose fruit virus (ToBRFV), demonstrují schopnost dlouhodobého přežívání na osemeni i v endospermu. Infikované sazenice se mohou vyvíjet bez viditelných patologických změn až do pozdních fází vegetace, což umožňuje viru nekontrolované šíření v produkčních sklenících ještě před první detekcí (Salem et al., 2023). Podobný mechanismus byl popsán u Tomato leaf curl New Delhi virus (ToLCNDV), kde latentní infekce v plevlech a přenos přes asymptomatické hostitele komplikují eradikační snahy v oblastech Středomoří (Vo et al., 2022; Vo et al., 2025).

U dřevin a trvalých kultur představuje dlouhá inkubační doba ještě závažnější epidemiologické riziko. V případě bakterie *Candidatus Liberibacter asiaticus*, původce choroby HLB (Huanglongbing) u citrusů, může doba mezi infekcí a projevem symptomů trvat měsíce až roky. Modelové studie ukazují, že v době, kdy jsou v sadu pozorovány první příznaky, je míra zamoření v regionu již na úrovni, která činí prosté kácení symptomatických stromů neúčinným opatřením (Ellis et al., 2025). Analogická situace nastává u bakterie *Xylella fastidiosa* u mandloní a olivovníků, kde spolehlivost detekce u bezpříznakových stromů dramaticky klesá v závislosti na použité metodě a typu vzorkovaného pletiva (Velasco-Amo et al., 2026).

Diagnostickým problémem u asymptomatických rostlin je extrémně nízký titr patogena, který se často pohybuje pod detekčním limitem standardních sérologických metod (ELISA, LFD). U latentních infekcí houbovými patogeny,

jako jsou zástupci rodu *Alternaria* v zrně pšenice, je navíc patogen často lokalizován nerovnoměrně, což zvyšuje riziko falešně negativních výsledků při nevhodném vzorkování (Shaizadinova et al., 2025).

Současný výzkum proto směřuje k implementaci vysoce citlivých molekulárních metod schopných amplifikovat i stopová množství cílové DNA přímo v polních podmínkách. Pokročilé techniky, jako je CRISPR-Cas12a nebo optimalizované izotermální amplifikace, vykazují výrazně vyšší senzitivitu než konvenční PCR a umožňují zachytit infekci v presymptomatické fázi (Lagner et al., 2025; Sahu et al., 2025). Úspěšnost těchto metod však přímo závisí na kvalitě bioinformatického návrhu příslušných oligonukleotidů, které musí být schopny cílit na konzervované oblasti genomu patogena s absolutní specificitou, aby nedocházelo k nespecifickým reakcím s genomem nebo mikrobiomem rostliny.

Diagnosticke překážky a cíle

Tradiční diagnostické postupy, opírající se o kultivaci a fenotypovou analýzu, přestávají v konfrontaci s moderními epidemiologickými hrozbami stačit. Hlavním limitem je faktor času a specificity. Izolace bakterií je proces na dny až týdny, což je v případě agresivních patogenů šířících se mezinárodním obchodem neakceptovatelná prodleva (Wang et al., 2025). Sérologické metody, sice nabízejí rychlost, ale často selhávají v rozlišení blízce příbuzných kmenů, které se mohou zásadně lišit virulencí či rezistencí. V éře kryptických druhů, jak bylo popsáno výše, je spoléhání se na fenotyp riskantní (Fielder et al., 2025).

Molekulární diagnostika přináší nezbytnou přesnost tím, že cílí přímo na genetickou informaci. Metody jako PCR a real-time PCR se staly zlatým standardem, avšak jejich závislost na centralizovaných laboratořích vytváří logistické úzké hrdlo. Současný trend proto směřuje k decentralizaci, k technologiím "Point-of-Care", tedy diagnostice přímo na místě, jako je izotermální amplifikace LAMP nebo systémy CRISPR-Cas, které přenášejí laboratorní přesnost přímo do terénu (Luchi et al., 2023; Sahu et al., 2025).

Úspěch těchto metod však nestojí pouze na chemickém principu reakce, ale primárně na kvalitě cílové sekvence - molekulárního markeru. V genomech hostitele, přirozeného mikrobiomu a příbuzných druhů patogenů, je nalezení úseku unikátního pro detekci jednoho velice konkrétního organismu výzvou, kterou již nelze řešit manuálně. Těžiště vývoje se tak přesouvá do roviny *in silico*. Precizní bioinformatický návrh, využívající komparativní analýzu celých genomů, se stává prerekvizitou pro vytvoření diagnostického nástroje, který je dostatečně citlivý, aby odhalil latentní infekci, a zároveň natolik specifický, aby generoval validní výsledky i v složitých podmínkách praxe (Beran et al., 2022).

Evoluce molekulární detekce

V této výzkumné práci nebyl kladen důraz na hledání jednoho univerzálního klíče, ale na strategické využití celého spektra molekulárních metod. Od robustního zlatého standardu PCR, přes terénní flexibilitu izotermální amplifikace LAMP, až po využití dat celogenomového sekvenování. Každá z těchto platform nabízí unikátní výhody, ale také klade specifické nároky na bioinformatický design, který je společným jmenovatelem jejich úspěchu (Bonham-Carter et al., 2014; Sahu et al., 2025).

Polymerázová řetězová reakce (PCR)

I v éře nastupujících nanotechnologií a sekvenování nové generace zůstává polymerázová řetězová reakce (PCR) referenčním bodem, vůči němuž jsou validovány všechny nové přístupy (Kanapiya et al., 2024). Její neotřesitelná pozice v laboratorní praxi vychází z geniálně jednoduchého principu cyklického řízení termodynamiky. Reakce využívá schopnosti termostabilní DNA polymerázy (původně *Taq*) syntetizovat nové vlákno DNA podle templátu, přičemž celý proces je řízen střídáním teplot ve třech fázích: denaturace (rozpletení dvouvláknové DNA), annealing (nasednutí primerů) a elongace (syntéza) (Kralík & Ricchi, 2017).

Právě krok nasednutí primerů (annealing) je zásadním momentem z hlediska specifity. Díky možnosti přesně nastavit teplotu reakce (T_a) můžeme s vysokou přesností diskriminovat i drobné sekvenční rozdíly. To se ukazuje jako klíčové při řešení taxonomických složitostí, které byly popsány v předchozích kapitolách – například při odlišení *Xanthomonas euvesicatoria* pv. *euvesicatoria* od blízce příbuzných patovarů, kde i rozdíl několika bází rozhoduje o správnosti diagnózy (Pečenka et al., 2020; Timilsina et al., 2020). Moderní diagnostika se však posunula od kvalitativní detekce na agarózovém gelu k přesné kvantifikaci pomocí real-time PCR (qPCR). Tato metoda umožňuje sledovat nárůst fluorescence v reálném čase, buď pomocí interkalačních barviv, obvykle SYBR Green, nebo vysoce specifických hydrolyzačních sond, např. TaqMan (Bustin a Nolan, 2017).

Výsledná hodnota prahového cyklu (C_t nebo C_q), tedy momentu, kdy fluorescenční signál překročí šum pozadí, nám dává neocenitelnou informaci o množství patogena ve vzorku. To je zásadní nejen pro diagnostiku, ale i pro hodnocení účinnosti ochranných opatření nebo šlechtění na rezistenci (Mirmajlessi et al., 2015; Yang et al., 2026). Navíc, zavedení multiplexní qPCR, kde v jedné reakci detekujeme více cílů současně (například patogen a interní kontrolu rostliny), se stalo standardem pro vyloučení falešně negativních výsledků způsobených selháním extrakce (Peňázová et al., 2020; Stehlíková et al., 2025). Přes všechny tyto výhody však PCR naráží na fyzikální limit: závislost na precizním cyklování teplot vyžaduje sofistikovanou instrumentaci, což metodu stále poutá k laboratornímu zázemí (Luchi et al., 2023).

Izotermální amplifikace LAMP

Odpovědí na potřebu decentralizace diagnostiky - tedy dostat test z laboratoře přímo do místa diagnostiky - se stala technologie LAMP (Loop mediated isothermal amplification). Tato metoda představuje ideální nástroj pro screeningové testování přímo v provozu. Na rozdíl od PCR eliminuje potřebu teplotních cyklů. Využívá polymerázu, která disponuje aktivitou vytěsňování vláken (strand displacement, např. polymerázy *Bst* nebo *GspSSD*). To umožňuje, aby reakce běžela kontinuálně při konstantní teplotě okolo 60–65 °C (Gomez-Gutierrez & Goodwin, 2022; Notomi et al., 2000).

Princip LAMP je fascinující svým důmyslným mechanismem. Využívá sadu 4 až 6 primerů, které rozpoznávají 6 až 8 různých oblastí na cílové DNA. Vnitřní primery iniciují tvorbu smyčkových struktur, které slouží jako startovací struktury pro masivní autocyklickou syntézu DNA. Výsledkem jsou dlouhé konkatemery DNA a obrovské množství vedlejšího produktu - pyrofosfátu hořečnatého (Serena et al., 2025). Z praktického hlediska je pro fytopatologa nejcennější vlastností LAMP její robustnost. Polymeráza *Bst* je výrazně tolerantnější k inhibitorům (jako jsou rostlinné polyfenoly či polysacharidy) než enzymy používané v PCR. To umožnilo v našich studiích, například při detekci *Xanthomonas hortorum* pv. *gardneri*, provádět detekci bez nutnosti složité

purifikace DNA, což radikálně zkracuje čas analýzy (Mermigka et al., 2026; Sheershika & Ram, 2024; Stehlíková et al., 2020).

Vizualizace výsledku LAMP může být navíc prostá jakékoli přístrojové techniky. Masivní syntéza DNA uvolňuje protony, což vede k acidifikaci reakční směsi. Přidáním pH indikátoru tak lze výsledek odečíst pouhým okem jako změnu barvy, např. z růžové na žlutou. Toto zpřístupnění diagnostiky je pro fytosanitární inspekci zásadním krokem (Amoia et al., 2023; Gudkov et al., 2023). Daní za tuto jednoduchost je však extrémní nárok na návrh primerů. Najít v genomu úsek, kde se vedle sebe složí šest primerů se specifickými termodynamickými vlastnostmi, je bez pokročilé bioinformatiky a práce s celogenomovými daty prakticky nemožné (Beran et al., 2022; de Andrade et al., 2024).

Cílené vyhledávání CRISPR-Cas

Vývoj v diagnostice směřuje k technologiím schopným obejít limity amplifikace: systémy odvozené od bakteriální imunity CRISPR-Cas. Zatímco PCR a LAMP primárně amplifikují DNA, systémy využívající efektory Cas12a, Cas12b pro DNA, nebo Cas13 pro RNA fungují jako vysoce specifické molekulární vyhledávače (Du et al., 2025).

Mechanismus detekce je odlišný od genového inženýrství, kde je metoda CRISPR-Cas již velmi rozšířena. Ribonukleoproteinový komplex (protein Cas + vodící crRNA) skenuje vzorek a hledá sekvenci komplementární k crRNA. Jakmile dojde k vazbě na cíl, enzym se aktivuje a spustí tzv. trans-cleavage (kolaterální) aktivitu. V tomto stavu začne enzym nespecificky štěpit jakoukoli jednovláknovou nukleovou kyselinu ve svém okolí. Pokud do reakce přidáme reportérové molekuly (krátké oligonukleotidy značené fluoroforem a zhasádkem, podobně jako TaqMan pro qPCR), aktivovaný enzym je rozstříhá, čímž uvolní fluorescenční signál (Gootenberg et al., 2017; Ren et al., 2026).

Aktuální data z let 2025 a 2026 naznačují, že CRISPR diagnostika dosahuje attomolární citlivosti (schopnost detekovat jednotlivé molekuly) a dokáže operovat při pokojové teplotě. Unikátní vlastností je schopnost rozlišit

jednobodové změny (SNP) s přesností, které LAMP ani PCR nedosahují (Wei et al., 2026), což otevírá dveře k rychlé identifikaci rezistentních kmenů patogenů (Besati et al., 2025; Lagner et al., 2025). Metoda se také ukazuje jako vysoce adaptabilní pro detekci RNA virů, jako je Tomato brown rugose fruit virus (ToBRFV), bez nutnosti reverzní transkripce v případě použití Cas13 (Chai et al., 2025; Shaizadinova et al., 2025).

Bioinformatika a strategie *in silico* designu

Citlivé platformy jako qPCR, LAMP nebo CRISPR jsou velmi výkonné nástroje. Aby však fungovaly spolehlivě a nedocházelo k falešně pozitivním / negativním výsledkům, je klíčový precizní návrh celého testu. Ať už navrhujeme dvojici primerů pro PCR, sadu šesti oligonukleotidů pro LAMP nebo vodící RNA pro Cas nukleázu, vždy řešíme fundamentální bioinformatický problém: jak v miliardách písmen genetického kódu nalézt unikátní řetězec, který jednoznačně definuje náš cíl a nic jiného (Kamran et al., 2026; Ogbuji, & Agogbua, 2025).

Principy návrhu molekulárních testů

Proces návrhu diagnostického markeru není jen o prostém vyhledání textového řetězce, ale představuje multidimenzionální optimalizační úlohu, která musí v jediném kroku vybalancovat tři často protichůdné požadavky: inkluzivitu, exkluzivitu a termodynamickou stabilitu (Bonham-Carter et al., 2014; Wang et al., 2022).

Prvním kritickým krokem je zajištění inkluzivity, tedy schopnosti detekovat cílový patogen v celé jeho šíři. V ideálním případě by měl být každý druh definován unikátní a stabilní genetickou sekvencí. V reálném biologickém prostředí, formovaném horizontálním přenosem genů a evoluční dynamikou, je však situace složitější. Historický přístup, kdy se návrh zakládal na jednom referenčním genomu staženém z databáze, je dnes považován za rizikový a nedostatečný. Riskujeme totiž, že v terénu narazíme na kmen, který má například v místě vazby primeru bodovou mutaci, což může vést k selhání reakce a falešně negativnímu výsledku. Moderní design proto musí vycházet z konceptu pangenomu, tedy analýzy sekvenčních dat stovek izolátů, jejímž cílem je identifikovat tzv. core genome – evolučně konzervované jádro sdílené celou populací, které je stabilní v čase a prostoru (Beran et al., 2022; Harrison et al., 2023; Timilsina et al., 2020).

Neméně důležitou, a často mnohem náročnější výzvou, je zajištění exkluzivity neboli specifity. Marker nesmí reagovat s žádnou jinou DNA přítomnou ve vzorku. V kontextu rostlinné patologie to znamená, že se musíme vyhnout nejen

křížové reakci s genomem hostitele (např. rajčete či pšenice), ale především s obrovským množstvím mikroorganismů tvořících přirozený mikrobiom rostliny a půdy. Situaci komplikuje fakt, že fytopatogenní bakterie (např. *Xanthomonas* nebo *Pseudomonas*) mají často velmi blízké nepatogenní příbuzné, kteří s nimi sdílejí biotop i většinu genomu. Pokud diagnostický marker nerozliší mezi nebezpečným patogenem a jeho neškodným příbuzným organismem, vede to k falešným poplachům, zbytečné aplikaci pesticidů a ekonomickým ztrátám (Fielder et al., 2025; Osdaghi et al., 2021; Pečenka et al., 2020). Nalezení sekvence, která je přítomna u všech patogenních kmenů (inkluzivita), ale zároveň zcela chybí u tisíců potenciálních environmentálních kontaminantů (exkluzivita), je úkolem srovnatelným s hledáním jehly v kupce sena, kde se stébla neustále mění (Landa et al., 2020; Peňázová et al., 2020).

I když nalezneme sekvenci, která je teoreticky unikátní, nemusí fungovat v reálné reakci, protože diagnostika je ve své podstatě aplikovaná fyzikální chemie. Primery a sondy musí splňovat přísná termodynamická kritéria, jako je optimální teplota tání (T_m) a volná Gibbsova energie (ΔG). Pokud navrhujeme primery s rozdílnou teplotou tání, jeden se naváže dříve než druhý, což vede k asymetrické amplifikaci a snížení citlivosti. Ještě zákeřnější je tvorba sekundárních struktur. Molekuly DNA mají tendenci se navázat samy do sebe a tvořit vlásenky nebo se vázat na sebe navzájem za vzniku dimerů. U metody LAMP je optimalizace podmínek mimořádně náročná, protože je nutné sladit vlastnosti šesti primerů v rámci jediné reakce, přičemž pravděpodobnost nežádoucích interakcí roste exponenciálně (Jasieniecka, & Domingues, 2025; Yang et al., 2023). U systémů CRISPR zase hraje klíčovou roli sekundární struktura vodící RNA (crRNA), která musí zůstat přístupná pro navázání Cas enzymu, jinak nedojde k aktivaci detekčního komplexu (Du et al., 2025; Ren et al., 2026).

Splnění takto přísných nároků však naráží na limity tradičních, heuristických bioinformatických nástrojů, které pro takto náročnou selekci postrádají potřebnou přesnost i výpočetní kapacitu. Moderní design markerů proto opouští pohled zaměřený na jednotlivé geny a přesouvá se k celogenomovým přístupům. Unikátní diagnostické znaky se totiž často nenacházejí v dobře popsanych genech,

ale v intergenových úsecích nebo v oblastech, které standardní anotace přehlíží. Nezbytné jsou nástroje, které dokáží pracovat s genomy jako s masivními soubory dat a hledat unikátní sekvence ne na základě biologické homologie, ale na základě matematické exkluzivity, například pomocí analýzy k-merů nebo hashovacích algoritmů. Právě tato potřeba zpracovat "Big Data" bakteriálních genomů a najít v nich unikátní sekvence, které unikají tradičním alignmentům, vedla k vývoji nových algoritmů a softwarových řešení, jež tvoří jádro předkládané práce a umožňují design robustních detekčních testů i pro taxonomicky nejsložitější skupiny patogenů (Beran et al., 2021; Li et al., 2008; Stehlíková et al., 2025; Yang et al., 2026).

Bioinformatické výzvy

Vstupujeme do éry, kdy limitujícím faktorem diagnostiky přestává být laboratorní technika a stává se jím zpracování dat. Exponenciální pokles ceny sekvenování nové generace vedl k tomu, že pro většinu rostlinných patogenů dnes nemáme k dispozici pouze jeden referenční genom, ale stovky či tisíce kompletních genomových sestav reprezentujících globální diverzitu populací (Ogbuji & Agogbua, 2025). Rychlé přibývání dat je pro fytopatologů velkou výhodou, zároveň však klade vysoké nároky na jejich zpracování a interpretaci (PS & Chintala, 2020). Na jedné straně pangenomické databáze nabízejí kompletní obraz o variabilitě cílového organismu, na straně druhé staví před bioinformatiku výpočetní bariéru, kterou nelze překonat pouhým navýšením hardwarového výkonu (Kamran et al., 2026).

Jádrem problému je samotný princip, na kterém byla bioinformatika po desetiletí stavěna: zarovnávání sekvencí neboli alignment. Tradiční algoritmy pro násobné zarovnání (MSA), jako jsou Clustal či MUSCLE, byly optimalizovány pro srovnávání evolučně příbuzných genů o délce tisíců bází. Jejich úkolem je seřadit sekvence pod sebe tak, aby homologní pozice odpovídaly, což umožňuje identifikovat konzervované úseky a mutace (Chao et al., 2022). Aplikace těchto algoritmů na celogenomová data v měřítku stovek izolátů však vede k extrémnímu nárůstu výpočetní náročnosti. Problém nalezení optimálního zarovnání je z

hlediska teoretické informatiky klasifikován jako NP-úplný, což znamená, že s každou přidanou sekvencí roste časová náročnost řešení exponenciálně. Zarovnat padesát genů 16S rRNA je otázkou sekund; zarovnat pět set kompletních genomů bakterie *Xanthomonas* je úkol, který přesahuje možnosti běžných výpočetních klastrů. Tento přístup navíc obvykle ani nedává biologický smysl kvůli rozsáhlým přestavbám genomu (Bao & Weigel, 2025; Zielezinski et al., 2019).

Druhým pilířem tradiční bioinformatiky jsou heuristické vyhledávací nástroje, jejichž synonymem je BLAST (Basic Local Alignment Search Tool). Tyto algoritmy obětují absolutní přesnost ve prospěch rychlosti. Fungují na principu hledání krátkých shod ("seeds") a jejich následném rozšiřování, přičemž výsledek je hodnocen statisticky (E-value). Pro fylogenetické studie nebo hrubou identifikaci je tento přístup dostačující. V molekulární diagnostice však operujeme v binárním režimu "pozitivní/negativní". Jediná neshoda na 3' konci primeru nebo v oblasti PAM sekvence u CRISPR může zcela zablokovat reakci a vést k falešně negativnímu výsledku. Heuristický přístup, který nezahrne drobnou neshodu v zájmu statistické významnosti, je proto pro návrh vysoce specifických testů rizikový. Navíc, při kontrole specifity proti databázím obsahujícím terabajty environmentálních sekvencí se i rychlý BLAST stává úzkým hrdlem (Gabsi et al., 2010; Wang et al. 2022).

Tváří v tvář těmto limitům se moderní výpočetní biologie obrací k tzv. alignment-free přístupům. Místo složité rekonstrukce celé sekvence genomu pracují tyto metody s daty jako s množinou fragmentů (Ren et al., 2018; Zielezinski et al., 2019). Základní jednotkou se stává k-mer - sekvence o pevné délce k (například 21 písmen). Rozložením genomu na k-mery převádíme složitý biologický problém zarovnávání na exaktní matematický problém množinových operací (Bussi et al., 2021; Wu et al., 2025).

V tomto paradigmatu je hledání diagnostického markeru definováno jako operace rozdílů množin. Hledáme takové k-mery, které jsou prvkem sjednocení všech genomů patogena (inkluzivita) a zároveň nejsou prvkem sjednocení všech genomů necílových organismů (exkluzivita). Tento přístup nabízí dvě zásadní

výhody. Zaprvé, je výpočetně efektivní a lineárně škálovatelný - porovnání množin je řádově rychlejší než zarovnávání sekvencí. Zadruhé, je absolutní - k-mer v množině buď je, nebo není (Roy & Mukhopadhyay, 2024; Song et al., 2024).

Implementace tohoto přístupu však přináší nové technické výzvy, především v oblasti správy paměti. Počet unikátních k-merů v souboru stovek bakteriálních genomů dosahuje miliard. Ukládat a prohledávat takové množství textových řetězců v operační paměti (RAM) vyžaduje optimalizované datové struktury. Zde bioinformatika čerpá inspiraci z počítačové vědy a kryptografie, konkrétně z využití hashovacích funkcí a Bloomových filtrů, které umožňují reprezentovat sekvence DNA v komprimované číselné podobě (Beran et al., 2021; Beran et al., 2025; Shi & Wang, 2019).

Právě na pomezí biologie a informatiky, kde biologická potřeba specifity naráží na limity výpočetní složitosti, vznikl prostor pro vývoj autorských nástrojů. Ty byly navrženy s cílem překlenout propast mezi teoretickou přesností alignment-free metod a praktickou použitelností na běžném hardwaru, a umožnit tak návrh robustních molekulárních testů i v éře pangenomových dat (Beran et al., 2022).

Algoritmická řešení a jejich integrace do diagnostické praxe

Zatímco předchozí kapitoly definovaly problém - nutnost najít v masivních pangenomických datech unikátní a stabilní sekvence - tato kapitola předkládá řešení. Vzhledem k limitům standardních algoritmů bylo přistoupeno k vývoji nových bioinformatických nástrojů. Ty staví na teorii množin a principech tzv. alignment-free analýzy. Místo dalšího čistě akademického programu bylo cílem vyvinout funkční nástroj, který otevře možnosti analýz i lidem bez hlubšího bioinformatického vzdělání. Tento přístup byl uplatněn již v dřívějších pracích zaměřených na vizualizaci a kontrolu kvality Sangerova sekvenování pomocí nástroje ABIF Manager (Beran a Mráz, 2015). Stejná filozofie uživatelské přístupnosti byla přenesena i do oblasti masivních genomických dat.

KEC: exkluze k-merů

Jádrem takového přístupu aplikovaným v této práci je algoritmus KEC (K-mer Exclusion by Crossreference). Byl navržen jako alternativa k nástroji BLAST, kterému při vyhledávání markerových sekvencí v rozsáhlých genomech chybí potřebná rychlost i přesnost. Zatímco BLAST hledá shodu (inkluzi) a následně ji filtruje, KEC přistupuje k problému opačně: hledá unikátnost pomocí exkluze (Beran et al., 2021).

Algoritmus pracuje s genomem jako s množinou podřetězců o pevné délce k (k-mery). Proces detekce unikátních sekvencí lze pak popsat jako sérii množinových operací:

- Cílový soubor (T): Množina všech k-merů přítomných v genomech cílového patogena. Abychom zajistili inkluzivitu, provádíme průnik ($\cap$) k-merů ze všech dostupných izolátů cíle. Tím získáme konzervované jádro (core k-mers).
- Necílový soubor (N): Množina všech k-merů z genomů, které chceme vyloučit (příbuzné druhy, hostitel, environmentální pozadí). Zde provádíme sjednocení ($\cup$) všech těchto k-merů.

- Kandidátní soubor (C): Výsledná množina kandidátních markerů je rozdílem těchto množin: $C = T \setminus N$.

Tento přístup, ač matematicky triviální, je biologicky velmi silný. Umožňuje totiž nalézt sekvence, které jsou pro daný druh absolutně unikátní, aniž bychom museli provádět výpočetně náročné zarovnávání. Nedávné srovnávací studie potvrzují, že k-merové přístupy typu KEC dosahují při analýze bakteriálních pangenomů řádově vyšší rychlosti než metody založené na sufixových stromech nebo FM-indexech, a to při zachování 100% specificity (Andreson et al., 2025; Ghezzi et al., 2024; Matsuda et al., 2025).

genCRC32: Akcelerace analýzy pomocí hashování

Při práci s genomy rostlinných patogenů, jako je *Xanthomonas*, narážíme na limit paměti. Ukládat miliardy textových řetězců (k-merů) ve formátu ASCII je neefektivní. Standardní řešení v informatice představuje hashování - převod textu na číselnou hodnotu (hash) o fixní délce. V bioinformatice se však potýkáme s problémem kolizí, kdy dvě různé sekvence DNA dostanou stejné číslo. V diagnostice, kde kolize může znamenat falešnou pozitivitu, může ale být kolize fatální (Shi & Wang, 2019).

Jako řešení byl vyvinut algoritmus genCRC32, který implementuje kolizně odolnou variantu hashovací funkce CRC32 optimalizovanou specificky pro čtyřpísmennou abecedu DNA. Nástroj mapuje sekvence DNA na 32bitová celá čísla, což radikálně snižuje paměťové nároky analýzy. Klíčovou inovací je ošetření kolizí pro k-mery biologicky relevantních délek, což z genCRC32 činí nástroj vhodný pro exaktní vyhledávání. Tento krok výrazně snižuje nároky na hardware. Analýza stovek genomů, včetně složitějších eukaryotních organismů, je díky tomu proveditelná i na standardním počítači, a to během pár desítek minut až několika hodin (Beran et al., 2025).

Syntéza: Hundred-Genomes Pipeline

Jednotlivé algoritmy (KEC, genCRC32) jsou samy o sobě jen nástroji. Jejich skutečná síla se projeví až v integraci do uceleného pracovního postupu, který byl nazván "Hundred-Genomes Pipeline". Tento systém byl úspěšně aplikován při vývoji duplexního LAMP testu pro detekci *Xanthomonas euvesicatoria* pv. *euvesicatoria* a *X. vesicatoria* (Beran et al., 2022).

Architektura postupu kopíruje logický tok diagnostického návrhu:

- Sběr dat: Stažení všech dostupných relevantních genomů z databáze NCBI.
- Filtrace a extrakce: Identifikace unikátních k-merů, které jsou přítomny ve 100 % cílových genomů a v 0 % necílových genomů (KEC).
- Validace kontextu: Kandidátní k-mery jsou mapovány zpět na referenční genom, aby se ověřilo, zda tvoří repetitivní oblasti nevhodné pro návrh primerů.
- Termodynamický design: Vybrané oblasti jsou předloženy softwaru Primer3, který v nich hledá optimální kombinace primerů (např. pro LAMP sadu F3/B3/FIP/BIP) splňující parametry ΔG a T_m .

Výstupem postupu není finální produkt, ale soubor vysoce pravděpodobných kandidátů. Žádný algoritmus nedokáže plně simulovat komplexitu biochemické reakce v rostlinném extraktu. Proto je nezbytnou součástí procesu empirická validace. V našich pracích se osvědčil dvoustupňový model:

- PCR screening: Rychlé ověření specifity vnějších primerů (F3/B3) na panelu referenčních kmenů.
- LAMP optimalizace: Testování kompletní sady na reálných vzorcích obsahujících inhibitory.

Ten je následován stanovením specifity a senzitivity pomocí laboratorního testování.

Právě spojení masivního bioinformatického výkonu s precizním laboratorním ověřením se v této práci osvědčilo jako efektivní způsob, jak vyvíjet diagnostické metody nové generace.

Budoucí vědecké směřování

Habilitační práce shrnuje etapu předchozího výzkumu, která byla definována přechodem od empirického designu markerů k exaktním algoritmickým přístupům. Vědecké poznání však nestojí na místě. Nástroje, které dnes považujeme za pokročilé, se v krátké době stanou standardem. Ve svém dalším směřování se proto hodlám soustředit na to, jak námi vyvinuté principy (k-merová exkluze, hashování) posunout dál - směrem k uživatelské přívětivosti, širší funkcionalitě a integraci strojového učení.

Evoluce nástroje KEC

Nástroj KEC byl původně navržen s jediným cílem: najít unikátní sekvence pro návrh primerů. Jeho potenciál je však mnohem širší. Cílem je transformovat KEC z jednoúčelové aplikace v komplexní bioinformatickou platformu.

Prvním krokem je automatizace a zpřístupnění pomocí grafického uživatelského rozhraní (GUI). Bioinformatika velmi často trpí bariérou použitelnosti - mnoho biologů neumí ovládat nástroje přes příkazovou řádku. Plánuji vytvořit intuitivní grafické rozhraní, které umožní široké komunitě využívat naše algoritmy bez nutnosti psaní příkazů (Ma et al., 2024). Druhým krokem je rozšíření funkcionality. Využití algoritmu KEC se totiž neomezuje pouze na hledání markerů. Lze jej efektivně adaptovat například pro kontrolu kvality sekvenčních dat, kde na základě vyloučení očekávaných k-merů rychle identifikuje případnou kontaminaci. Další významnou aplikací je genotypizace; analýza k-merového spektra umožňuje rozlišit jednotlivé kmeny a dosáhnout tak mnohem jemnějšího taxonomického zařazení, než jaké nabízí standardní 16S analýza.

Dynamická diagnostika v reálném čase

Dosavadní aplikace našich algoritmů směřovaly k vytvoření konkrétní sady primerů pro konkrétní patogen (např. *Xanthomonas*). Budoucnost však vidím v širší aplikaci. Použitím našich algoritmů na globální pangenomické databáze

můžeme vytvořit "živé" databáze diagnostických cílů pro stovky organismů současně.

Místo toho, aby výzkumník musel pokaždé spouštět celou analýzu znovu, bude systém kontinuálně na pozadí aktualizovat sady unikátních k-merů s každým nově osekvenovaným izolátem, který přibude do veřejných databází (NCBI). Tímto přejdeme od ad-hoc návrhu k systému, který prediktivně nabízí validované markery pro celé spektrum hospodářsky významných patogenů, od virů po houby (Fielder et al., 2025; Kamran et al., 2026).

Adaptace pro CRISPR-Cas12b

Ačkoliv se má dosavadní práce opírala primárně o amplifikační metody (PCR, LAMP), nelze ignorovat nástup programovatelných nukleáz. Technologie CRISPR, konkrétně termostabilní varianta Cas12b, představuje další logický krok v evoluci polní diagnostiky. Cas12b umožňuje provádět amplifikaci (např. LAMP) i detekci v jedné reakci při stejné teplotě (tzv. "one-pot assay"), což radikálně zjednodušuje instrumentaci a omezuje možnost kontaminace (Du et al., 2025).

Z hlediska bioinformatiky je to nová, dosud neprobádaná výzva. Zatímco LAMP vyžaduje 6 primerů, CRISPR vyžaduje specifickou vodicí RNA (sgRNA) s odlišnými pravidly návrhu (PAM sekvence, sekundární struktura scaffoldu). Plánuji rozšířit algoritmus KEC o moduly, které budou schopny v pangenomech vyhledávat nejen unikátní oblasti pro primery, ale zároveň i optimální cíle pro Cas12b komplexy. Tím bude využito know-how z práce s velkými daty a bude aplikováno na technologii, která s velkou pravděpodobností bude definovat příští směřování diagnostiky (Jasieniecka & Domingues, 2025).

Strojové učení: Hledání skrytých vzorů v datech

Poslední oblastí, která v tuto chvíli představuje největší koncepční skok od deterministických přístupů k pravděpodobnostním modelům, je integrace strojového učení. Dosavadní návrh diagnostiky stál pevně na termodynamických

základech, kdy jsme úspěšnost reakce predikovali na základě vypočtené teploty tání a volné energie. Každodenní praxe v laboratoři však ukazuje, že teoretické modely mají své limity. I oligonukleotidy, které podle všech parametrů vypadají dokonale, totiž v reálných podmínkách často z neznámých příčin selhávají. Byla by proto zásadní chyba ignorovat revoluci, kterou do genomiky přináší strojové učení. Zde se otevírá fascinující prostor například pro aplikaci algoritmů typu Random Forest, které vynikají schopností nacházet nelineární vztahy v tabulkových datech a, což je pro vědu klíčové, nabízejí i vhled do toho, který parametr představoval pro rozhodnutí modelu jakou váhu.

Další obrovský a dosud zcela nevyužitý potenciál představuje vytěžování takzvaných negativních dat. Vědecká komunita i naše laboratoře generují tisíce experimentálních výsledků, kde reakce nefungovaly, a tato data končí v zapomnění. Pro natrénování prediktivního modelu jsou však tato selhání cennějším zdrojem informací než úspěchy. Využitím těchto dat lze naučit algoritmy rozpoznávat krátké sekvenční motivy, specifické uspořádání bází nebo kontextuální vlivy okolní DNA, které systematicky vedou k selhání polymerázy či CRISPR komplexu, aniž by byly zachyceny současnými biofyzikálními vzorci. Stejný princip lze aplikovat na samotný algoritmus KEC, který by se mohl z prostého matematického nástroje pro exkluzi vyvinout v inteligentní systém schopný kandidátní unikátní sekvence nejen najít, ale i seřadit podle pravděpodobnosti jejich biologické stability a dostupnosti pro vazbu v reálném buněčném lyzátu.

Další rovinou je analýza komplexních interakcí v multiplexních testech a metodách jako LAMP, kde v jedné zkumavce interagují čtyři a více oligonukleotidů. Počet možných kombinací zde roste velmi rychle a přesahuje možnosti klasického modelování, avšak pro strojové učení představuje ideální trénovací sadu pro predikci nežádoucího pozadí a falešné pozitivitu. V neposlední řadě se nabízí využití těchto metod při zpracování surových dat z přenosných sekvenátorů, kde by modely natrénované na šumu specifickém pro danou technologii mohly v reálném čase oddělovat signál patogena od hostitele s přesností, která by umožnila diagnostiku přímo na poli bez nutnosti přístupu k

cloudu. Přejchod od designu řízeného rovnicemi k designu řízenému daty představuje cestu, jak zvýšit spolehlivost diagnostických nástrojů na úroveň blížící se jistotě a proměnit bioinformatiku z nástroje analytického v nástroj skutečně prediktivní (Ghedia et al., 2026; Wang et al., 2022; Wu et al., 2021).

Rozbor publikovaných prací

Tato kapitola shrnuje vybrané publikované práce autora. Vzhledem k tematickému zaměření jsou práce rozděleny do tří skupin, které ukazují, jak se moderní diagnostika vyvíjí od počítačového kódu až po praktické využití. Každá skupina článků je doplněna vlastním komentářem, který zasazuje jednotlivé práce do širších souvislostí.

První část se věnuje softwarovým nástrojům, které byly vyvinuty, aby bylo možné v obrovském množství genomických dat najít podstatné informace. Druhá část ukazuje využití těchto postupů při řešení konkrétních problémů v rostlinolékařství, zejména u bakteriálních patogenů rostlin. Třetí skupina pak dokládá, že stejné molekulární a bioinformatické metody fungují univerzálně a pomáhají i v jiných oborech, od šlechtění zvířat, až po kontrolu potravin.

Vývoj autorských softwarových nástrojů

Moderní diagnostika rostlinných patogenů vyžaduje robustní softwarovou podporu, která dokáže reagovat na specifické potřeby laboratorní praxe. Tato část práce předkládá shrnutí tří bioinformatických nástrojů, které vznikly s cílem vyřešit konkrétní limitace při zpracování sekvenačních dat. Tyto aplikace pokrývají spektrum od základní kontroly kvality Sangerova sekvenování až po výpočetně náročné analýzy celých genomů.

Prvním nástrojem je ABIF Manager, publikovaný v roce 2015 v časopise Bioinformatics. I v době masivního nástupu nových technologií zůstává Sangerovo sekvenování referenční metodou pro validaci výsledků. Práce s výstupy sekvenátorů však byla často komplikována proprietárními formáty souborů, které neumožňovaly snadnou editaci bez nákladného komerčního software. ABIF Manager tento problém vyřešil poskytnutím rozhraní pro čtení a úpravu ABIF souborů. Umožnil uživatelům provádět detailní inspekci primárních dat a vizuálně ověřit správnost automatického čtení bází (base calling). Tímto způsobem lze v diagnostice efektivně odhalit potenciální heterozygotní pozice či artefakty, které by při pouhém převzetí textové sekvence zůstaly skryty.

S přechodem k sekvenování nové generace (NGS) a analýze pangenomů vyvstala nutnost zpracovávat řádově větší objemy dat, kde již vizuální kontrola není možná. Pro vyhledávání unikátních diagnostických markerů v masivních databázích genomů byl vyvinut autorský algoritmus KEC (K-mer Exclusion by Crossreference), představený v roce 2021. Tento nástroj reagoval na nedostatečnou rychlost tradičních zarovnávacích algoritmů (alignment) a omezenou přesnost heuristických metod, jako je BLAST. Algoritmus KEC využívá principu matematické exkluze k-merů. Místo hledání podobnosti identifikuje sekvence, které jsou přítomny v cílové skupině a zároveň zcela chybí v kontrolní skupině necílových organismů. Tento exaktní přístup umožňuje navrhnout vysoce specifické oligonukleotidy bez rizika křížových reakcí.

Zpracování stovek genomů současně však naráží na hardwarové limity, především na kapacitu operační paměti při ukládání miliard textových řetězců. Odpovědí na tento technický problém je algoritmus genCRC32, který implementuje metodu hashování. Algoritmus převádí sekvence DNA na číselné hodnoty pomocí optimalizované funkce CRC32. Klíčovým přínosem této implementace je zajištění bezkolizního průběhu pro biologicky relevantní délky oligonukleotidů, což je pro diagnostickou přesnost nezbytné. Použití genCRC32 vede k výrazné redukci paměťových nároků a umožňuje provádět komplexní analýzy i na běžném výpočetním hardware.

Představený soubor aplikací pokrývá různé aspekty zpracování genetických dat. Zatímco ABIF Manager cílí na kontrolu kvality u jednotlivých sekvencí získaných Sangerovou metodou, nástroje KEC a genCRC32 řeší algoritmické a paměťové výzvy spojené s analýzou tzv. velkých dat v éře genomiky. Každý z těchto nástrojů tak přispívá k zefektivnění a zpřesnění návrhu molekulárních diagnostických metod.

Související publikace

Beran, P., & Mraz, I. (2015). ABIF Manager: simple reading and editing of ABIF formatted files. *Bioinformatics*, 31(24), 4012-4013.

- Beran, P., Rost, M., Beranová, K., Kváč, M., Stehlíková, D., Udoh, O. E., Jozová, E., & Čurn, V. (2025). genCRC32: Collision-free CRC32 based hashing of DNA sequences. *Bioinformatics Advances*, vbaf315
- Beran, P., Stehlíková, D., Cohen, S. P., & Čurn, V. (2021). KEC: unique sequence search by k-mer exclusion. *Bioinformatics*, 37(19), 3349-3350.

Na tomto místě se v plné verzi habilitační práce nachází text příslušné publikace. Ten však není součástí této veřejně přístupné verze z důvodu licenčních podmínek vydavatele. Níže je k dispozici bibliografická citace a abstrakt článku. Plný text lze nalézt v placeném režimu na webových stránkách časopisu.

Beran, P., & Mraz, I. (2015). ABIF Manager: simple reading and editing of ABIF formatted files. *Bioinformatics*, 31(24), 4012-4013.

Abstrakt

Summary: ABIF Manager is simple application for low-level access to ABIF formatted files. Those are produced by Data Collection software based on the data, collected from capillary sequencing machine. The application shows all the data as readable text, which can be easily edited.

Sequence analysis

genCRC32: collision-free CRC32-based hashing of DNA sequences

Pavel Beran^{1,*}, Michael Rost¹, Kristina Beranová¹, Martin Kváč^{1,2}, Dagmar Stehlíková¹, Obiora Emmanuel Udoh¹, Eva Jozová¹, Vladislav Čurn¹

¹Faculty of Agriculture and Technology, University of South Bohemia in České Budějovice, České Budějovice 370 05, Czech Republic

²Institute of Parasitology, Biology Centre of the Academy of Sciences of the Czech Republic, České Budějovice 370 05, Czech Republic

*Corresponding author. Faculty of Agriculture and Technology, University of South Bohemia in České Budějovice, Studentská 1668, České Budějovice 37005, Czech Republic. E-mail: pavel.beran@centrum.cz.

Associate Editor: Michael DeGiorgio

Abstract

Motivation: Efficient and collision-free hashing of DNA sequences is essential for accuracy and performance in bioinformatics applications such as genome assembly, sequence alignment, and metagenomic classification. Traditional hashing methods often result in collisions, impacting the precision and/or performance of downstream analyses. Thus, it is highly advantageous to have hashing functions that guarantee collision-free mappings for DNA sequences, particularly for k-mers up to length 16, where practical limits for 32-bit hashing are reached. In this study, we evaluate genCRC32 as a hashing primitive, reporting collision behavior, bucket balance, sensitivity to single-base changes, and speed to inform its potential use in downstream tools. Evaluation within specific software tools is outside the scope of this paper and is planned as future work.

Results: We present genCRC32, an innovative hashing method that integrates a straightforward preprocessing step (gen32) with CRC32 hashing, specifically identifying eight CRC32 polynomials that ensure collision-free hashing for all DNA k-mers up to 16 nucleotides in length. Through extensive empirical evaluations, genCRC32 demonstrated zero collisions for these k-mers, achieving a one-to-one mapping without auxiliary data structures. Benchmark tests confirmed minimal computational overhead introduced by preprocessing, maintaining hashing performance comparable to established methods such as MurmurHash3 and xxHash32.

Availability and implementation: The source code for genCRC32 is publicly available at: <https://github.com/berybox/genCRC32>. The implementation is provided in Go (version 1.24) and leverages only standard libraries, ensuring portability and ease of integration into existing bioinformatics workflows.

1 Introduction

Efficient management and rapid querying of large-scale genomic data are fundamental challenges in contemporary bioinformatics. Hash functions are algorithms that map variable-length inputs to fixed-size outputs. They are extensively employed to accelerate key tasks such as sequence alignment, genome assembly, and database searches by enabling constant-time lookups and compact data representations (Rivest 1992, Menezes *et al.* 1996, Ramakrishna and Zobel 1997, Marchet *et al.* 2021, Pibiri 2022). However, conventional hash functions inherently permit collisions—instances where distinct inputs yield identical hash values—with potentially detrimental effects on accuracy and performance in applications requiring unambiguous sequence identification (Dietzfelbinger *et al.* 1994, Majewski *et al.* 1996).

DNA k-mer hashing is particularly sensitive to collisions. Although a 32-bit hash space can theoretically encode up to 416 (≈ 232) distinct 16-mers without collision, in practice collisions occur far more frequently due to uneven mapping and distribution properties of typical hash functions. These collisions can lead to misassemblies in de Bruijn graph construction, erroneous variant calls, and decreased sensitivity in

metagenomic classification (Altschul *et al.* 1990, Phillippy *et al.* 2008). To address collisions, one might consider minimum perfect hash functions. Minimum perfect hash functions (MPHF) guarantee collision-free $O(1)$ lookups on static key sets, and modern constructions are compact and scalable—combining near-optimal space with fast queries and practical builds (Esposito *et al.* 2020, Pibiri and Trani 2021, Lehmann *et al.* 2025, Beling and Sanders 2025). However, being set-specific, they are less suited to our goal of a general-purpose 32-bit hash, which motivates the CRC32-based approach.

Cyclic redundancy check (CRC) codes—ubiquitous in digital communications for error detection—offer an attractive balance of simplicity, hardware acceleration, and strong bit-diffusion properties (Peterson and Brown 1961, Koopman and Chakravarty 2004). In particular, CRC32 applies a 32-bit polynomial divisor to produce checksums that detect common data corruptions with minimal computational overhead (Castagnoli *et al.* 1993, Koopman 2002). Raw CRC32, however, is not designed for collision-free hashing of genomic k-mers: although the theoretical capacity matches the number of possible 16-mers, collisions remain common for $k \geq 5$ in real datasets due to non-ideal polynomial choices and the

Received: June 26, 2025; Revised: November 3, 2025; Accepted: November 7, 2025

© The Author(s) 2025. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

absence of input-specific preprocessing (Ondov et al. 2016, Wood et al. 2019).

Here, we introduce genCRC32, a collision-free hashing scheme for DNA k-mers of length up to 16 bases. Our method employs a lightweight bit-level preprocessing step (“gen32”) that encodes each nucleotide into four bits and packs adjacent bases into bytes, transforming the original k-mer into a $\lceil k/2 \rceil$ -byte intermediate. By aligning the 416 possible 16-mers with the 232 CRC32 output space and empirically identifying CRC32 polynomials that yield zero collisions, genCRC32 achieves a one-to-one mapping without auxiliary lookup tables. Importantly, genCRC32 is non-locality preserving—similar sequences do not necessarily produce similar hashes—which mitigates clustering of related k-mers and can improve discrimination in applications such as variant detection and metagenomic classification (Hosseini et al. 2019, Jafari et al. 2021).

The simplicity of genCRC32’s preprocessing and the ubiquity of hardware-accelerated CRC32 implementations ensure that our approach can be seamlessly integrated into existing bioinformatics pipelines. In the following sections, we detail the gen32 preprocessing algorithm, the selection of collision-free CRC32 polynomials, and performance benchmarks comparing genCRC32 to standard CRC32, MurmurHash3, ntHash, PHast, and xxHash32 on exhaustive k-mer sets. We evaluate genCRC32 as a hashing primitive on these k-mer sets and report collision behavior, bucket balance, sensitivity to single-base changes, and speed. Evaluation within specific software tools is planned for future work.

2 Methods

2.1 Algorithm implementation

We developed genCRC32 in Go (version 1.24) using only the standard library to ensure portability and simplicity. The pipeline comprises a bit-level preprocessing step (gen32) that packs nucleotide string into a byte-level representation, followed by a CRC32 checksum computed via Go’s hash/crc32 package. The code is compiled with the default go build command (no additional flags are necessary).

2.2 Preprocessing (gen32)

The gen32 function converts a DNA sequence of length k into a byte array of $\lceil k/2 \rceil$ bytes. No explicit case normalization is needed: masking each character with 0x0E inherently yields identical codes for uppercase and lowercase nucleotides by isolating bits 1–3 of the ASCII byte, which differ uniquely among A/a, C/c, G/g, and T/t. Two resulting 4-bit codes are concatenated into one byte. The pseudocode for the gen32 function is:

```
function gen32_preprocessing(sequence):
    k ← length(sequence)
    size ← (k ÷ 2) + (k mod 2)
    result ← make([], size) # all bytes zero
    for i from 0 to k − 1:
        # Extract 4-bit nucleotide code via ASCII masking:
        # 'A'/'a' → 0x00, 'C'/'c' → 0x02, 'G'/'g' → 0x06, 'T'/'t'
        # → 0x04
        code ← sequence[i] AND 0x0E
        # Shift existing nibble to high 4 bits
        result[i ÷ 2] ← result[i ÷ 2] << 4
        # Insert new code in low 4 bits
        result[i ÷ 2] ← result[i ÷ 2] OR code
```

return result

By packing two bases per byte, this step encodes all 4^k possible k-mers into $\lceil k/2 \rceil$ bytes—i.e. approximately half the original byte length (e.g. a 16-mer → 8 bytes = 64 bits).

2.3 CRC32 hashing and polynomial selection

We evaluated twenty 32-bit CRC polynomials—commonly used in hardware and software CRC32 implementations (Castagnoli et al. 1993, Koopman 2002, Koopman and Chakravarty 2004)—including: 0xEB31D82E, 0xEDB88320, 0x82F63B78, 0xAD0424F3, 0x945D045D, 0x9D9947FD, 0xA3000000, 0x87496166, 0x8E2371EF, 0x8C746ED4, 0x8EFD4BCD, 0x80000057, 0x8741C726, 0x82608EDB, 0x8EE5368F, 0xE47B4C57, 0xD2C0EF07, 0x80FCB077, 0x93B39B1B, 0xB338ADD6.

Through exhaustive testing, we identified eight polynomials—0x8741C726, 0x87496166, 0x8E2371EF, 0x8EE5368F, 0x8EFD4BCD, 0x945D045D, 0x9D9947FD, and 0xEB31D82E—that yield zero collisions for all gen32 preprocessed k-mers up to length 16.

2.4 Collision testing and benchmarking

For each $k \in \{4 \dots 16\}$, we generated the complete set of 4^k k-mers and applied eight different hashing protocols:

- Raw CRC32 (all 20 polynomials)
- gen32 + CRC32 (same polynomials)
- Raw MurmurHash3 (via github.com/spaolacci/murmur3)
- gen32 + MurmurHash3
- Raw xxHash32 (via <https://github.com/OneOfOne/xxhash>)
- gen32 + xxHash32
- ntHash v2.3.0 (<https://github.com/bcgsc/ntHash>)
- PHast v0.10.0 (<https://docs.rs/crate/ph>)

Every hash value was inserted into an in-memory Go map (dictionary/set) to detect collisions. Each configuration (hash function × polynomial) was run in randomized order and repeated 10× times. Because collisions were identical across replicates, we report the mean collision count (which equals the exact count per run) for each configuration.

Throughput was computed as the total number of k-mers hashed divided by the elapsed wall time for the exhaustive collision-testing runs (for each k), and we report it as hashes per second (H/s or MH/s).

2.5 Uniformity across buckets

We evaluated how well each hashing method distributes outputs by measuring bucket uniformity on three input regimes representative of common k-mer workloads. Each regime comprised 1 000 000 k-mers: (i) Random, uniformly sampled k-mers; (ii) Similar, near-neighbor variants generated by single-base edit of previous k-mer; and (iii) Window, overlapping k-mers obtained with a sliding window from the *Escherichia coli* K-12 MG1655 reference (NC_000913.3). The identical k-mer lists were hashed by all methods to isolate hashing behavior from input composition.

For each method and dataset we mapped hash values to m buckets via $\text{bucket} = h \bmod m$, using a panel of bucket counts spanning typical hash-table sizes (powers of two and nearby odd primes). We tallied bucket occupancies and computed the Pearson χ^2 statistic against the ideal count n/m , converting to P values ($df = m - 1$). To ensure the validity of χ^2 approximations, we restricted values to bucketings with expected counts > 5 .

2.6 Avalanche effect

We quantified output diffusion using the Random k -mer set (same as above). For each original k -mer we generated three independent single-nucleotide substitutions (one base changed per variant), yielding three origin–mutation pairs per k -mer. Each method hashed all originals and all mutated counterparts. For every pair, we computed the XOR of the two hash values and measured the fraction of flipped output bits (Hamming distance divided by the hash width b , with $b = 32$ for 32-bit methods). Our summary metric is mean flip, the average of this fraction across all pairs. An ideal avalanche property corresponds to mean flip = 0.5.

2.7 Performance measurement and environment

We measured elapsed real time for each full k -mer hashing run using Go's time package. All benchmarks were executed on a single workstation (Intel Xeon E5-1630 v3 CPU, 64 GB RAM, Crucial MX500 500 GB SSD, Ubuntu 20.04 LTS).

3 Results

3.1 Collision-free performance of genCRC32

The genCRC32 algorithm combines a novel preprocessing method (gen32) with the CRC32 hashing algorithm. We systematically tested its collision-free capability across a range of k -mer lengths ($k = 4$ –16). Using gen32 preprocessing, eight CRC32 polynomials (0x8741C726, 0x87496166, 0x8E2371EF, 0x8EE5368F, 0x8EFD4BCD, 0x945D045D, 0x9D9947FD, and 0xEB31D82E) demonstrated complete collision-free performance for the entire tested range ($k = 4$ –16). ntHash and PHast were included as comparators. Being collision-free by design, they did not require gen32 preprocessing in our evaluations. By contrast, all other tested hash configurations, including raw CRC32, MurmurHash3 and xxHash32 with and without gen32 preprocessing, exhibited collisions for one or more of the k -mer lengths evaluated (Table 1).

It is notable that $k = 16$ represents the maximum theoretical capacity of a 32-bit hash space, thus making collision-free performance at this length particularly significant.

3.2 Execution performance and speed

Execution speed benchmarks revealed that the gen32 preprocessing step introduces only minimal computational overhead compared to direct hashing without preprocessing. Specifically, the average hashing speed across all tested configurations ($k = 4$ –16) was 23 245 622 hashes per second with preprocessing, compared to 25 031 005 hashes per second without preprocessing (Supplementary Table 1). This indicates that gen32 preprocessing imposes negligible overhead, maintaining efficient hashing even at the largest tested k -mer length ($k = 16$).

Detailed analysis of execution speed at $k = 16$ showed mean speeds of 31 124 560 hashes per second (with gen32) and 31 622 151 hashes per second (without preprocessing). This negligible difference in speed is crucial since $k = 16$ is the critical limit for collision-free 32-bit hashing, underscoring the efficiency of gen32 preprocessing at maximum capacity. Table 2 reports throughput at $k = 16$, computed as the total number of k -mers hashed divided by the elapsed wall time for the exhaustive runs. Across CRC32 variants, MurmurHash3, and xxHash32, observed throughputs were of the same order of magnitude, and adding gen32 preprocessing incurred only

Table 1. Collisions observed for different hash configurations and k .

Hash configuration	Collision-free k value range	k value range with collisions
gen32 + MurmurHash3	4–8	9–16
Raw MurmurHash3		
Raw xxHash	4–9	10–16
Raw CRC32 (D2C0EF07)		
gen32 + xxHash	4–10	11–16
Raw CRC32 (80FCB077, 8741C726, 8C746ED4, 8EE5368F, 945D045D)		
Gen32 + CRC32 (82f63b78)	4–11	12–16
Raw CRC32 (80000057, 82608EDB, 82F63B78, 87496166, 8E2371EF, 8EFD4BCD, 93B39B1B, 9D9947FD, A3000000, AD0424F3, B338ADD6, E47B4C57, EB31D82E, EDB88320)		
gen32 + CRC32 (93B39B1B)	4–13	14–16
gen32 + CRC32 (B338ADD6)	4–14	15–16
gen32 + CRC32 (80000057, 80FCB077, 82608EDB, 8C746ED4, A3000000, AD0424F3, D2C0EF07, E47B4C57, EDB88320)	4–15	16
gen32 + CRC32 (8741C726, 87496166, 8E2371EF, 8EE5368F, 8EFD4BCD, 945D045D, 9D9947FD, EB31D82E)	4–16	–
ntHash		
PHast		

marginal overhead relative to the corresponding raw variant. In our measurements, ntHash attained the highest throughput, consistent with its rolling formulation optimized for adjacent k -mers, whereas PHast was the slowest among the tested implementations. We caution that absolute speed is implementation-sensitive (language, vectorization, memory layout, engineering choices), so the figures should be read as descriptive under our setup; the key observation is that gen32 does not materially affect throughput within our pipeline.

3.3 Scaling with k -mer length

Analyzing execution speed across different k -mer lengths ($k = 4$ –16; Fig. 1), we observed consistent linear scaling of hashing speed with increasing k -mer length. The observed lower speeds at smaller k -mer lengths ($k \leq 10$) are likely due to general computational overhead and external system processes influencing short-duration tasks. Both preprocessed and raw hashing configurations followed similar trends, with preprocessing consistently showing slightly lower yet comparable speeds. Differences in execution speeds diminished substantially as k -mer length increased ($k \geq 10$), confirming minimal preprocessing overhead for larger datasets.

3.4 Consistency and robustness

Throughout our experiments, collision counts remained constant across multiple replicates, demonstrating the robust and deterministic behavior of all tested hashing configurations. This consistency confirms the stability of genCRC32's performance, rendering it highly suitable for reliable integration into bioinformatics pipelines.

3.5 Bucket-occupancy uniformity and avalanche effect

Table 2 summarizes three properties: (i) speed described above (ii) bucket-occupancy uniformity, reported as dataset-level P

Table 2. Hashing metrics and throughput at $k = 16$: bucket-occupancy uniformity (χ^2 P values for Random/Similar/Window), avalanche effect (mean flip; ideal = 0.5), and throughput (M hash/s).

Algorithm	Preprocessing	Bucket-occupancy uniformity			Avalanche effect	
		P value (random)	P value (similar)	P value (window)	Mean flip	Speed MH/s
CRC32 (8741C726)	gen32	.53	.86	.13	0.48	33.24
CRC32 (8741C726)	None	.41	.51	.13	0.50	30.51
CRC32 (87496166)	gen32	.32	.87	.14	0.51	33.28
CRC32 (87496166)	None	.41	.60	.20	0.51	30.24
CRC32 (8E2371EF)	gen32	.46	.80	.16	0.51	33.31
CRC32 (8E2371EF)	None	.59	.70	.06	0.51	30.60
CRC32 (8EE5368F)	gen32	.50	.88	.14	0.51	33.35
CRC32 (8EE5368F)	None	.32	.65	.12	0.50	30.50
CRC32 (8EFD4BCD)	gen32	.55	.80	.09	0.45	33.27
CRC32 (8EFD4BCD)	None	.46	.67	.25	0.49	30.45
CRC32 (945D045D)	gen32	.65	.96	.11	0.48	33.25
CRC32 (945D045D)	None	.43	.57	.20	0.48	30.39
CRC32 (9D9947FD)	gen32	.40	.83	.22	0.50	33.30
CRC32 (9D9947FD)	None	.40	.73	.22	0.51	30.72
CRC32 (EB31D82E)	gen32	.42	.86	.11	0.49	33.70
CRC32 (EB31D82E)	None	.46	.70	.10	0.50	30.45
MurmurHash3	gen32	.43	.57	.04	0.50	29.84
MurmurHash3	None	.37	.45	.11	0.50	35.73
xxHash	gen32	.33	.25	.09	0.50	29.75
xxHash	None	.50	.52	.13	0.50	36.67
ntHash	None	.45	.42	.07	0.49	91.51
PHast	None	1.00	1.00	1.00	0.37	1.14

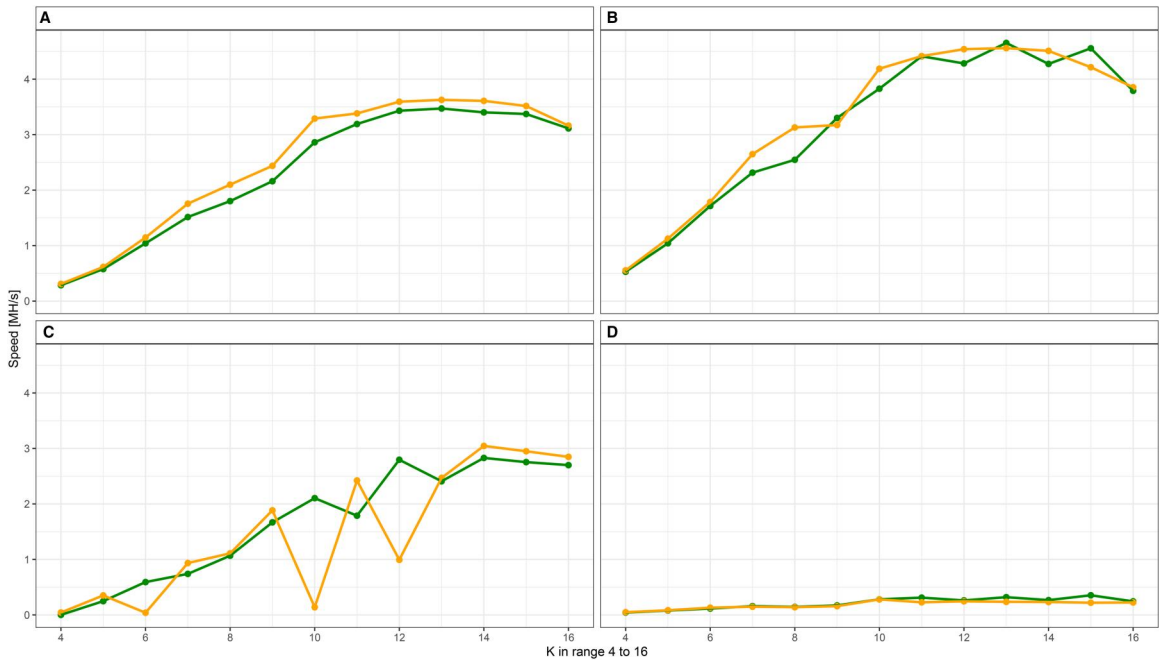


Figure 1. Four-panel comparison of hashing performance across k values (4–16) on X-axis for configurations with gen32 preprocessing (orange line with markers) versus no preprocessing (green line with markers). (A) Average speed, (B) maximal speed, (C) minimal speed, and (D) standard deviation of speed.

values for the Random, Similar, and Window inputs (Pearson’s χ^2 goodness-of-fit against the uniform expectation), and (iii) the avalanche effect, reported as mean flip. These metrics reflect, respectively, how evenly hash values populate buckets after modulo reduction and how sensitively (and uniformly) output bits respond to small input perturbations.

In our measurements, gen32 + CRC32 tended to yield higher uniformity P values than the corresponding raw CRC32

configurations on the Similar dataset—consistent with more balanced bucket occupancies under localized edits—while behavior on Random was broadly comparable and Window appeared most challenging across methods. For avalanche effect, most general-purpose hashes (CRC32 variants, MurmurHash3, xxHash32, ntHash) exhibited mean flip values close to 0.5. By contrast, PHast showed lower mean flip despite very even bucket counts, which aligns with the design goals of

minimal/perfect hashing—injective placement of a fixed key set into a fixed address range (e.g. bucket-then-seed schemes)—prioritizing collision-free, highly regular placement rather than per-bit diffusion.

3.6 Practical utility in hash-based pipelines

Many k -mer-centric pipelines (e.g. exact-match indices, hashtable-based k -mer stores, and sketching/seeding components) require more than a simple symbol-to-bit transformation and may benefit from bucketization, diffusion under small edits, and predictable throughput. In this setting, genCRC32 provides a deterministic, collision-free mapping for all 4^k inputs with $k \leq 16$ under our CRC32 configurations, exhibits balanced bucket occupancies across representative bucketings, shows near-ideal diffusion under single-base edits (mean flip ≈ 0.5), and attains throughput comparable to standard CPU-friendly hashes with only marginal preprocessing overhead. These properties address the practical factors that govern load balance, robustness, and runtime cost in hash-based genomics tooling, complementing compact 2-bit encodings (storage-oriented) and key-set-specific perfect hashing (placement-oriented) by characterized hashing behavior directly.

3.7 Summary of importance

The collision-free performance, combined with minimal computational overhead, confirms the practical utility of genCRC32 in bioinformatics. Its robustness, consistent speed, and capability to achieve theoretical maximum collision-free performance at $k = 16$ make it highly advantageous for efficient and precise management of genomic datasets in various computational genomics applications.

4 Discussion

In this study, we introduced genCRC32, an innovative hashing approach tailored specifically for bioinformatics applications that require precise, collision-free hashing of DNA sequences. By integrating a straightforward preprocessing step (gen32) with carefully selected CRC32 polynomials, we successfully achieved collision-free hashing performance for all possible k -mers up to the theoretical maximum length of $k = 16$ for a 32-bit hash space.

Our identification of eight CRC32 polynomials that provide collision-free hashing was strictly empirical. Currently, there is no established theoretical explanation for why these specific polynomials perform this way, especially considering the preprocessing step transforms nucleotide sequences into intermediate binary representations. It is probable that other (non-tested) CRC32 polynomials could also provide collision-free hashing performance. However, there is currently no basis to believe they would offer any advantage over the eight polynomials identified in this study. Future research could benefit from mathematical or computational studies designed explicitly to explore polynomial performance in hashing tasks.

Collision-free hashing performance at $k \leq 16$ is particularly valuable in genome assembly and alignment applications, where collision-induced ambiguities can significantly impair accuracy or performance of downstream analyses (Phillippy *et al.* 2008, Bankevich *et al.* 2012). Because assemblers and k -mer classifiers use the hash as an internal primitive, the key determinants are the hash's collision rate, bucket balance, diffusion to single-base edits, and throughput; we therefore evaluate on exhaustive k -mer sets and representative workloads. For instance, de Bruijn

graph-based genome assemblers such as SPAdes and Velvet rely on precise k -mer identification, directly affecting graph complexity and assembly accuracy (Zerbino and Birney 2008, Bankevich *et al.* 2012). Similarly, taxonomic classifiers like Kraken2 depend heavily on precise hashing to accurately assign metagenomic sequences, especially when handling complex and extensive reference databases (Wood *et al.* 2019, Lu and Salzberg 2020). In comparison, commonly used hash functions like MurmurHash3 and xxHash32, while computationally efficient, do not guarantee collision-free hashing, highlighting the practical advantage of genCRC32 in scenarios demanding high accuracy and speed.

Hashing is also foundational to minimizer-based sampling of k -mers, where a (pseudo)random hash induces an order within each window and the minimal element is retained as the seed. Minimizers and recent variants (e.g. mod-minimizers) are widely used in seeding, indexing, and sketching (Koerkamp and Pibiri 2024). An interesting direction for future work is to examine how the distributional properties of candidate hash functions (e.g. near-uniformity and independence across k -mers) influence the induced order and sampling density in minimizer schemes.

Nevertheless, it is crucial to recognize genCRC32's limitations when considering k -mer lengths beyond $k = 16$. Given the empirical nature of our polynomial discovery, extending collision-free hashing to larger k -mers would require larger hash spaces and/or alternative hashing methodologies. However, for typical bioinformatics applications, including genome assembly, variant detection, and rapid indexing in metagenomics, collision-free hashing at $k \leq 16$ is sufficiently robust and practical (Ondov *et al.* 2016, Marçais *et al.* 2018, Lu and Salzberg 2020). For $k \leq 16$, the four-letter alphabet admits an injective identifier via direct 2-bit encoding (a 2k-bit state). This representation would support $O(1)$ rolling updates (left-shift by two bits, retain the last 2k bits, then append the incoming base) and thus offers a collision-free alternative orthogonal to our CRC32 design. We did not implement this variant here; we note it as a feasible option for applications that specifically require rolling updates or guaranteed injectivity at small k .

A significant strength of genCRC32 lies in its simplicity and ease of integration. Its straightforward preprocessing step, combined with widely supported and optimized CRC32 hashing implementations, facilitates seamless integration into existing bioinformatics workflows, minimizing disruption and overhead. This practical advantage is particularly beneficial for tools and pipelines already operational in research and production environments.

In conclusion, genCRC32 represents a practical and effective advancement in bioinformatics hashing methodologies, achieving collision-free hashing performance at the theoretical maximum limit for 32-bit hashing. While further theoretical exploration into polynomial selection would be valuable, the current implementation already delivers substantial practical benefits. genCRC32 provides fast, robust, reliable, and collision-free hashing performance suitable for precise and high-throughput bioinformatics applications, highlighting its immediate and practical applicability in modern computational genomics.

Acknowledgements

We acknowledge the use of OpenAI's ChatGPT as an aid in refining the manuscript in accordance with the ISCB

acceptable use policy for large language models. ChatGPT was employed solely for copy-editing (grammar and style correction) and consistency checking. All scientific content, analyses, and interpretations were developed and verified by the authors.

Supplementary data

Supplementary data are available at *Bioinformatics Advances* online.

Conflict of interest

None declared.

Funding

This work has been supported by The Grant Agency of the University of South Bohemia in České Budějovice under grant number GAJU 083/2025/Z.

Data availability

Source code, documentation and executable binary files are available at the GitHub repository (<https://github.com/berybox/genCRC32>).

References

- Altschul SF, Gish W, Miller W *et al.* Basic local alignment search tool. *J Mol Biol* 1990;215:403–10.
- Bankevich A, Nurk S, Antipov D *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;19:455–77.
- Beling P, Sanders P. PHast-Perfect Hashing made fast. In: *Proceedings of the Twenty-Eighth Workshop on Algorithm Engineering and Experiments (ALENEX 2026)*, Vancouver, BC, Canada. Philadelphia, PA: SIAM, 2026 (To appear).
- Castagnoli G, Brauer S, Herrmann M. Optimization of cyclic redundancy-check codes with 24 and 32 parity bits. *IEEE Trans Commun* 1993;41:883–92.
- Dietzfelbinger M, Karlin A, Mehlhorn K *et al.* Dynamic perfect hashing: upper and lower bounds. *SIAM J Comput* 1994;23:738–61.
- Esposito E, Mueller Graf T, Vigna S. RecSplit: minimal perfect hashing via recursive splitting. In: *Proceedings of the Twenty-Second Workshop on Algorithm Engineering and Experiments (ALENEX 2020)*, Salt Lake City, UT, USA. Philadelphia, PA: SIAM, 2020, 175–85.
- Hosseini M, Pratas D, Pinho AJ. A survey on data compression methods for biological sequences. *Information* 2019;10:390.
- Jafari O, Maurya P, Nagarkar P *et al.* A survey on locality sensitive hashing algorithms and their applications. arXiv, <https://doi.org/2102.08942>, 2021, preprint: not peer reviewed.
- Koerkamp GR, Pibiri GE. The mod-minimizer: a simple and efficient sampling algorithm for long k-mers. In: Pissis SP, Sung W-K (eds.), *24th International Workshop on Algorithms in Bioinformatics (WABI 2024)*, Egham, UK. *Leibniz International Proceedings in Informatics (LIPIcs)*, Article No. 11. Wadern, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2024, 11:1–11:23.
- Koopman P. 32-bit cyclic redundancy codes for internet applications. In: Fabre J-C, Jahanian F (eds.), *Dependable Systems and Networks (DSN 2002)*, Bethesda, MD, USA. Los Alamitos, CA: IEEE, 2002, 459–68.
- Koopman P, Chakravarty T. Cyclic redundancy code (CRC) polynomial selection for embedded networks. In: *Int. Conf. Dependable Systems and Networks, Florence, Italy*. Los Alamitos, CA: IEEE, 2004, 145–54.
- Lehmann H-P, Mueller T, Pagh R *et al.* Modern minimal perfect hashing: a survey. arXiv, <https://doi.org/2506.06536>, 2025, preprint: not peer reviewed.
- Lu J, Salzberg SL. Ultrafast and accurate 16S rRNA microbial community analysis using kraken 2. *Microbiome* 2020;8:124.
- Majewski BS, Wormald NC, Havas G *et al.* A family of perfect hashing methods. *Comput J* 1996;39:547–54.
- Marchet C, Kerbiriou M, Limasset A. BLight: efficient exact associative structure for k-mers. *Bioinformatics* 2021;37:2858–65.
- Marçais G, Delcher AL, Phillippy AM *et al.* MUMmer4: a fast and versatile genome alignment system. *PLoS Comput Biol* 2018;14:e1005944.
- Menezes AJ, Van Oorschot PC, Vanstone SA. *Handbook of Applied Cryptography*. Boca Raton, FL: CRC Press, 1996.
- Ondov BD, Treangen TJ, Melsted P *et al.* Mash: fast genome and meta-genome distance estimation using MinHash. *Genome Biol* 2016;17:132.
- Peterson WW, Brown DT. Cyclic codes for error detection. *Proc IRE* 1961;49:228–35.
- Phillippy AM, Schatz MC, Pop M. Genome assembly forensics: finding the elusive mis-assembly. *Genome Biol* 2008;9:R55.
- Pibiri GE. Sparse and skew hashing of k-mers. *Bioinformatics* 2022;38:i185–94.
- Pibiri GE, Trani R. PTHash: revisiting FCH minimal perfect hashing. In: *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '21) Virtual Event*, Canada. New York, NY, USA: ACM, 2021, 1339–48.
- Ramakrishna MV, Zobel J. Performance in practice of string hashing functions. In: *Database Systems for Advanced Applications. Melbourne, Australia*. Singapore: Springer World Scientific, 1997, 215–24.
- Rivest R. RFC 1321: *The MD5 Message-Digest Algorithm*. Fremont, CA: Internet Engineering Task Force, 1992.
- Wood DE, Lu J, Langmead B. Improved metagenomic analysis with kraken 2. *Genome Biol* 2019;20:257.
- Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 2008;18:821–9.

Na tomto místě se v plné verzi habilitační práce nachází text příslušné publikace. Ten však není součástí této veřejně přístupné verze z důvodu licenčních podmínek vydavatele. Níže je k dispozici bibliografická citace a abstrakt článku. Plný text lze nalézt v placeném režimu na webových stránkách časopisu.

Beran, P., Stehlíková, D., Cohen, S. P., & Čurn, V. (2021). KEC: unique sequence search by k-mer exclusion. *Bioinformatics*, 37(19), 3349-3350

Abstrakt

Searching for amino acid or nucleic acid sequences unique to one organism may be challenging depending on size of the available datasets. K-mer elimination by cross-reference (KEC) allows users to quickly and easily find unique sequences by providing target and non-target sequences. Due to its speed, it can be used for datasets of genomic size and can be run on desktop or laptop computers with modest specifications.

Aplikace molekulárních metod v diagnostice bakteriálních patogenů

Vývoj diagnostických metod pro rostlinolékařskou praxi prošel v uplynulých patnácti letech zásadní proměnou, kterou lze sledovat na sérii předložených publikací. Tyto práce mapují cestu od prvotních snah o zachycení genetické variability patogenů, až po vývoj vysoce specifických nástrojů schopných rozlišit fylogeneticky blízce příbuzné druhy přímo v terénu. Ústředním tématem tohoto souboru prací je diagnostika hospodářsky významných bakteriálních patogenů rodů *Pseudomonas* a *Xanthomonas*, u nichž tradiční fenotypové metody často selhávají.

Počátky uvedeného výzkumu byly definovány snahou o pochopení genetické diverzity uvnitř komplexních bakteriálních skupin. Práce zaměřená na *Pseudomonas syringae* využila technologii DNA čipů (microarray) pro screening variability v oblasti vnitřního transkribovaného spaceru (ITS) 16S-23S rRNA. Tato studie demonstrovala, že i v rámci jednoho druhu existuje značná heterogenita, kterou je nutné při návrhu diagnostických nástrojů zohlednit.

V kontextu diagnostiky rodu *Xanthomonas* se výzkum soustředil na taxonomicky složitý komplex druhů, dříve souhrnně označovaný jako *Xanthomonas campestris* pv. *vesicatoria*. Série publikací popisuje postupný vývoj specifických PCR primerů, které reagovaly na potřebu přesného rozlišení jednotlivých pátovarů. Tyto nástroje nebyly vyvíjeny pouze teoreticky, ale byly přímo validovány při monitoringu výskytu choroby v České republice, což umožnilo mapovat šíření patogena v reálných podmínkách. S postupujícím poznáním taxonomie a rozpadem původního druhového komplexu na samostatné druhy (*X. euvesicatoria*, *X. vesicatoria*, *X. gardneri* a *X. perforans*) vyvstala nutnost aktualizace detekčních metod. Byla vyvinuta sada primerů schopná specificky detekovat *Xanthomonas euvesicatoria*, což je úkol komplikovaný vysokou sekvenční podobností s příbuznými druhy.

Vrchol této vývojové linie představuje přechod k metodám, které kombinují rychlost s robustností, a jsou vhodné pro tzv. polní (Point-of-Care) diagnostiku. Pozornost se zaměřila na izotermální amplifikaci metodou Loop-mediated isothermal amplification (LAMP). Pro detekci *Xanthomonas gardneri* byl vyvinut test, který umožňuje vizualizaci výsledku pouhým okem na základě změny barvy reakční směsi, čímž eliminuje potřebu složitého laboratorního vybavení.

Nejnovější práce pak integruje bioinformatický přístup s laboratorní aplikací. S využitím nového výpočetního postupu „Hundred-Genomes Pipeline“ byl navržen duplexní LAMP test, který dokáže v jedné reakci simultánně detekovat a rozlišit *Xanthomonas euvesicatoria* a *X. vesicatoria*. Tento přístup, využívající analýzu stovek genomů pro nalezení unikátních markerů, překonává limity tradičního návrhu primerů a představuje modelový příklad moderní diagnostiky, která je rychlá, přesná a opřená o masivní genomická data.

Související publikace

Beran, P., Stehlikova, D., Cohen, S. P., Rost, M., Beranova, K., & Curn, V. (2023).

Utilization of a new hundred-genomes pipeline to design a rapid duplex LAMP detection assay for *Xanthomonas euvesicatoria* and *X. vesicatoria* in tomato. *Plant Disease*, 107(6), 1822-1828.

Beran, P., Mraz, I., Kokoskova, B., & Bohata, A. (2015). Monitoring the occurrence of bacterial spot of tomato and pepper in the Czech Republic and development of new PCR primers for detection of *Xanthomonas vesicatoria*. *European journal of plant pathology*, 141(3), 617-621.

Beran, P., & Mráz, I. (2013). Species-specific PCR primers for detection of *Xanthomonas vesicatoria*. *Crop Protection*, 43, 213-215.

Lenz, O., Beran, P., Fousek, J., & Mráz, I. (2010). A microarray for screening the variability of 16S–23S rRNA internal transcribed spacer in *Pseudomonas syringae*. *Journal of microbiological methods*, 82(1), 90-94.

- Pecenka, J., Kocanova, M., Baranek, M., Gazdik, F., Ragasova, L., Penazova, E., Čechová, J., Beran, P., & Eichmeier, A. (2020). Species-specific PCR primers for the detection of poorly distinguishable *Xanthomonas euvesicatoria*. *Crop protection*, 127, art. no. 104978.
- Stehliková, D., Beran, P., Cohen, S. P., & Čurn, V. (2020). Development of real-time and colorimetric loop mediated isothermal amplification assay for detection of *Xanthomonas gardneri*. *Microorganisms*, 8(9), 1301.
- Stehlikova, D., Beran, P., Rost, M., & Curn, V. (2026). Genomes-based loop-mediated isothermal amplification assay for rapid and specific detection of walnut pathogen *Xanthomonas arboricola* pv. *juglandis*. *Journal of Plant Pathology*, 1-10.

Na tomto místě se v plné verzi habilitační práce nachází text příslušné publikace. Ten však není součástí této veřejně přístupné verze z důvodu licenčních podmínek vydavatele. Níže je k dispozici bibliografická citace a abstrakt článku. Plný text lze nalézt v placeném režimu na webových stránkách časopisu.

Beran, P., Stehlikova, D., Cohen, S. P., Rost, M., Beranova, K., & Curn, V. (2023). Utilization of a new hundred-genomes pipeline to design a rapid duplex LAMP detection assay for *Xanthomonas euvesicatoria* and *X. vesicatoria* in tomato. *Plant Disease*, 107(6), 1822-1828.

Abstrakt

Xanthomonas euvesicatoria and *X. vesicatoria* are two economically important causal agents of bacterial spot (BS) of tomato and pepper. Management of BS in the field requires rapid and accurate detection. Therefore, this work aimed to develop a pipeline to design a simple, fast, and reliable assay for the detection of *X. euvesicatoria* and *X. vesicatoria* by loop-mediated isothermal amplification. In total, 109 publicly available whole genomic sequences of 24 different species of bacterial pathogens were used to design primers that would amplify the DNA of the two target species. Laboratory testing of the assay was performed on pure bacterial cultures and artificially infected plants, and amplification was conducted with both a sophisticated laboratory instrument and a simple mobile platform. The testing of the assay confirmed its specificity with a sensitivity reaching 1 pg μl^{-1} for both pathogens with an assay duration of 40 min on a mobile detection platform. Our diagnostics development pipeline enables the easy and fast design of a reliable detection assay in the genomics age. By validating the pipeline with *X. euvesicatoria* and *X. vesicatoria* pathogens, we have simultaneously developed an assay with high specificity, sensitivity, and speed, which will allow it to be deployed, contributing to successful management of BS.

Na tomto místě se v plné verzi habilitační práce nachází text příslušné publikace. Ten však není součástí této veřejně přístupné verze z důvodu licenčních podmínek vydavatele. Níže je k dispozici bibliografická citace a abstrakt článku. Plný text lze nalézt v placeném režimu na webových stránkách časopisu.

Beran, P., Mraz, I., Kokoskova, B., & Bohata, A. (2015). Monitoring the occurrence of bacterial spot of tomato and pepper in the Czech Republic and development of new PCR primers for detection of *Xanthomonas vesicatoria*. European journal of plant pathology, 141(3), 617-621.

Abstrakt

We tested 212 bacterial isolates isolated from tomato and pepper plants suspected of being infected by *Xanthomonas vesicatoria* (causing bacterial spot). Samples from plants were collected in the years 2007–2010 from fields located in central Bohemia and central and south Moravia. For detection of the pathogens we used PCR with two previously published primer pairs (RST65, RST69 and XCVF, XCVR) and one primer pair (Xv1, Xv2) designed for specific detection of *Xanthomonas vesicatoria*. In total, 121 of 212 (57 %) samples were positive in PCR. In this article we provide data for additional verification of the newly designed (Xv1, Xv2) primers. Our results confirm the need of precise detection methods, because several tomato and pepper pathogens can share similar visual symptoms. We also suggest, that new PCR primers for detection of bacterial spot of tomato and pepper pathogens should not be overlooked, because they can improve currently used PCR assays for detection of these pathogens.

Na tomto místě se v plné verzi habilitační práce nachází text příslušné publikace. Ten však není součástí této veřejně přístupné verze z důvodu licenčních podmínek vydavatele. Níže je k dispozici bibliografická citace a abstrakt článku. Plný text lze nalézt v placeném režimu na webových stránkách časopisu.

Beran, P., & Mráz, I. (2013). Species-specific PCR primers for detection of *Xanthomonas vesicatoria*. *Crop Protection*, 43, 213-215.

Abstrakt

This study presents the development of a PCR method for specific detection of *Xanthomonas vesicatoria*, causal agent of bacterial spot of tomato and pepper. Primer pair XV1F and XV1R was designed based on partial DNA sequence of the catalytic subunit of the ATP synthase gene *atpD* and tested using a collection of various bacterial pathogens mainly for tomato and pepper. After optimization of the PCR conditions, the 365 bp long fragment was only produced in *X. vesicatoria* samples. Therefore, the primers and protocol allow rapid and species-specific detection of *X. vesicatoria*.

Na tomto místě se v plné verzi habilitační práce nachází text příslušné publikace. Ten však není součástí této veřejně přístupné verze z důvodu licenčních podmínek vydavatele. Níže je k dispozici bibliografická citace a abstrakt článku. Plný text lze nalézt v placeném režimu na webových stránkách časopisu.

Lenz, O., Beran, P., Fousek, J., & Mráz, I. (2010). A microarray for screening the variability of 16S–23S rRNA internal transcribed spacer in *Pseudomonas syringae*. *Journal of microbiological methods*, 82(1), 90-94.

Abstrakt

The 16S–23S ribosomal internal transcribed spacer (ITS1) is often used as a subspecies or strain-specific molecular marker for various kinds of bacteria. However, the presence of different copies of ITS1 within a single genome has been reported. Such mosaicism may influence correct typing of many bacteria and therefore knowledge about exact configuration of this region in a particular genome is essential. In order to screen the variability of ITS1 among and within *Pseudomonas syringae* genomes, an oligonucleotide microarray targeting different configurations of ITS1 was developed. The microarray revealed seven distinct variants in 13 pathovars tested and detected mosaicism within the genomes of *P. syringae* pv. *coronafaciens*, *pisi*, *syringae* and *tabaci*. In addition, the findings presented here challenge the using of rRNA analysis for pathovar and strain determination.

Na tomto místě se v plné verzi habilitační práce nachází text příslušné publikace. Ten však není součástí této veřejně přístupné verze z důvodu licenčních podmínek vydavatele. Níže je k dispozici bibliografická citace a abstrakt článku. Plný text lze nalézt v placeném režimu na webových stránkách časopisu.

Pecenka, J., Kocanova, M., Baranek, M., Gazdik, F., Ragasova, L., Penazova, E., Čechová, J., Beran, P., & Eichmeier, A. (2020). Species-specific PCR primers for the detection of poorly distinguishable *Xanthomonas euvesicatoria*. Crop protection, 127, art. no. 104978.

Abstrakt

New species-specific PCR primers for detection of bacterium *Xanthomonas euvesicatoria*, the causal agent of bacterial spot of tomato and pepper, were designed in this work. Developed PCR primer pair ZnDoF/ZnDoR partially targets the coding sequence of Zn-dependent oxidoreductase gene. The primer pair was tested on various bacterial strains including bacteria from the bacterial spot-causing xanthomonads (BSX) group. Designed primers showed to be species-specific for the samples containing DNA of *X. euvesicatoria*. Specificity of the designed primers was also compared with the primer pairs already published. Subsequently, sensitivity of ZnDoF/ZnDoR primers was determined by ten-fold dilution of standardized sample. Detection capability of developed primers was also investigated on artificially infected pepper plants and seeds collected from bacterial spot-like peppers. Primers proved to be species-specific for in planta and in seed detection as well as to distinguish *X. euvesicatoria* from other bacterial strains used in this assay.



Article

Development of Real-Time and Colorimetric Loop Mediated Isothermal Amplification Assay for Detection of *Xanthomonas gardneri*

Dagmar Stehlíková¹, Pavel Beran^{1,*}, Stephen P. Cohen² and Vladislav Čurn¹¹ Biotechnological Centre, Faculty of Agriculture, University of South Bohemia, Na Sadkach 1780, 37005 Ceske Budejovice, Czech Republic; stehld00@zf.jcu.cz (D.S.); curn@zf.jcu.cz (V.Č.)² Department of Plant Pathology, The Ohio State University, Columbus, OH 43210, USA; cohen.1107@osu.edu

* Correspondence: beranp02@zf.jcu.cz; Tel.: +420-723-289-647

Received: 22 July 2020; Accepted: 24 August 2020; Published: 26 August 2020



Abstract: *Xanthomonas gardneri* is one of the causal agents of bacterial spot (BS), an economically important bacterial disease of tomato and pepper. Field-deployable and portable loop-mediated isothermal amplification (LAMP)-based instruments provide rapid and sensitive detection of plant pathogens. In order to rapidly and accurately identify and differentiate *X. gardneri* from other BS-causing *Xanthomonas* spp., we optimized a new real-time monitoring LAMP-based method targeting the *X. gardneri*-specific *hrpB* gene. Specificity and sensitivity of real-time and colorimetric LAMP assays were tested on the complex of bacterial strains pathogenic to tomato and pepper and on plants infected by the pathogen. The assay detection limit was 1 pg/μL of genomic DNA with an assay duration of only 30 min. The use of portable and handheld instruments allows for fast analysis, reducing the diagnosis time, and can contribute to proper disease management and control of *X. gardneri*. Due to the high efficiency of this method, we suggest its use as a standard diagnostic tool during phytosanitary controls.

Keywords: *Xanthomonas* spp.; bacterial spot; LAMP

1. Introduction

Xanthomonas gardneri is one of the species of the genus *Xanthomonas* that are currently considered to cause bacterial spot (BS) on tomato and pepper. It was first isolated in former Yugoslavia and named *Pseudomonas gardneri* [1]. Later, this name was suggested to be a synonym for *X. vesicatoria* [2]. However, DNA–DNA hybridization indicated that these organisms were genetically distinct [3]. Four distinct groups of *Xanthomonas* were differentiated based on various physiological and molecular tests [4]. Group A was named *X. euvesicatoria*; group B mostly included *X. vesicatoria*, group C *X. perforans* and group D *X. gardneri* [5]. Additionally, based on amplified fragment length polymorphism and multilocus sequence analysis (MLSA) of genes *atpD*, *dnaK*, *efp* and *gyrB* partial sequences, the four bacterial strains pathogenic for tomato and pepper were identified to the species level in *Xanthomonas* spp. [6]. The whole genomic sequence of *X. euvesicatoria* showed a 99.7% relationship with *X. perforans* [7]. It was suggested that the two species should be combined into one and subdivided into two pathovars: *X. euvesicatoria* pv. *euvesicatoria* and *X. euvesicatoria* pv. *perforans* [8]. Based on phylogenetic analysis and comparison of partial gyrase B gene sequences, *X. gardneri* and *X. cynarae* (an artichoke pathogen) were recognized as species closely related to *X. hortorum* [9]. New data from MLSA using four housekeeping genes (*dnaK*, *fyuA*, *gyrB* and *rpoD*) indicated that *X. gardneri* appears to be identical to *X. cynarae* but further investigation of the classification of *Xanthomonas* is still needed [10].

Various molecular, biochemical and physiological assays have been used to characterize *Xanthomonas* species. For *Xanthomonas* spp. causing BS, PCR-based molecular detection assays

were used [11,12]. A region of *hrpB* operon as a potential source target for primers and probes for real-time TaqMan PCR assay for *X. euvesicatoria*, *X. vesicatoria*, *X. gardneri* and *X. perforans* was also evaluated [13]. PCR primer sets with TaqMan probes were developed to distinguish between *X. gardneri* and other *Xanthomonas* groups causing bacterial spot of tomato [14]. Currently, DNA-based techniques are the most used methods for molecular diagnostic of this group of pathogenic bacteria.

Nucleic acid amplification is one of the most valued methods in science but still requires sophisticated instruments for amplification and detection. A majority of the techniques for nucleic acid analysis utilize the polymerase chain reaction (PCR) amplification method, which requires repeated cycles of three temperature-dependent steps during the amplification of the target nucleic acid sequence [15,16]. Isothermal amplification techniques such as nucleic acid sequence-based amplification, helicase-dependent amplification, rolling circle amplification, multiple displacement amplification, recombinase polymerase amplification and loop-mediated isothermal amplification (LAMP) were developed to facilitate DNA amplification in simpler ways [17]. LAMP is a novel isothermal amplification method for amplifying a limited amount of DNA copies into millions of copies in less than an hour. It is a molecular technique of nucleic acid amplification where a set of four to six different primers binds to six to eight different regions on the target gene, providing high specificity [18]. A basic LAMP primer set consists of two outer primers (F3 and B3) and two inner primers (forward inner primer—FIP and backward inner primer—BIP). Loop primers may be used to accelerate the reaction [19]. LAMP methods utilize *Bacillus smithii*, *Bacillus stearothermophilus* or *Geobacillus* sp. DNA polymerases with high strand displacement activity at optimal working temperature ranged between 60 and 65 °C [18,20,21]. This method can be easily performed at the point-of-care thanks to its isothermal nature. It has been successfully used for rapid and specific detection of plant bacteria from infected plant tissues and soil. The objective of this study was to develop a real-time and colorimetric LAMP protocol for specific detection of *X. gardneri* that would be superior to the current PCR-based methods in its speed, ease of use and possibility of point-of-care application.

2. Materials and Methods

2.1. Bacterial Strains and Culturing

Bacterial strains (Table 1) were obtained from Belgian Co-ordinated Collections of Microorganisms—Bacteria Collection, Gent (BCCM/LMG); French Collection of Plant associated bacteria, Beaucauze Cedex (CFBP); Czech Collection of Microorganisms, Brno (CCM); Crop Research Institute Collection, Czech Republic, Prague—Ruzyně (CRI); Deutsche Sammlung Microorganismen und Zellkulturen GmbH, Germany (DSMZ); National Collection of Plant Pathogenic Bacteria, UK, York (NCPBPB); National Collection of Agricultural and Industrial Microorganisms—(NCAIM); and Horticulture Research International, Wellesbourne, UK, (HRI-W). All *Clavibacter* spp. were grown on nutrient broth yeast extract (NBY) [22] at 27 °C for 3 to 7 days, depending on the subspecies. *Pantoea agglomerans* was grown on King's B medium [23] at 25 °C for 24 to 48 h. Other bacteria were cultured on MPAg medium (meat-peptone agar with glucose: 20 g of nutrient agar no. 2, 2.6 g of yeast extract, 5 g of glucose, 10 g of agarose—added to 500 mL of distilled H₂O, pH adjusted to 7.2 with 1 M NaOH and solidified) at 28 °C for 24–48 h, or 3–7 days (only *X. fragariae*).

2.2. DNA Isolation

Freshly grown bacteria (5–10 mg) were taken from plates. Total DNA was extracted using NucleoSpin® Microbial DNA kit (Macherey-Nagel, Dylan, Germany) according to the manufacturer's protocol. For agitation, a Retsch® Mixer Mill MM400 was used for 4 min at maximal frequency (30 Hz). Concentration of extracted DNA was measured with a BioSpec-nano spectrophotometer (Shimadzu, Kyoto, Japan).

Table 1. ¹ Bacteria listed in collection; ² Type; ³ Pathotype; ⁴ not specified (N.S.) by collection; ⁵ (+) for positive detection, (−) for negative detection; BCCM/LMG—Belgian Co-ordinated Collections of Microorganismse Bacteria Collection; CFBP—French Collection of Plant Pathogenic Bacteria; CCM—Czech Collection of Microorganisms; CRI—Crop Research Institute Collection; NCPPB—National Collection of Plant Pathogenic Bacteria; DSMZ—Deutsche Sammlung von Mikroorganismen und Zellkulturen; NCAIM—National Collection of Agricultural and Industrial Microorganisms; HRI-W—Horticulture Research International, Wellesbourne, UK.

Name ¹	Collection	Number in Collection	Geographic Origin	LAMP Result ⁵
<i>Xanthomonas gardneri</i>				
<i>X. gardneri</i> ²	DSMZ	19127	Yugoslavia	+
<i>X. gardneri</i>	CFBP	8588	France (Réunion)	+
<i>X. gardneri</i>	CFBP	7992	France (Réunion)	+
<i>X. gardneri</i>	CFBP	8120	Costa Rica	+
Other (non- <i>gardneri</i>) <i>Xanthomonas</i>				
<i>X. alfalfae</i> subsp. <i>alfalfae</i>	CFBP	3836	Sudan	−
<i>X. arboricola</i> pv. <i>pruni</i>	BCCM/LMG	854	New Zealand	−
<i>X. axonopodis</i> pv. <i>allii</i>	CFBP	6369	Not specified (N.S.) ⁴	−
<i>X. axonopodis</i> pv. <i>carotae</i>	NCPPB	3440	Brazil	−
<i>X. axonopodis</i> pv. <i>vesicatoria</i>	CRI	1013	Czech Republic	−
<i>X. campestris</i> pv. <i>incanae</i>	HRI-W	6377	UK	−
<i>X. campestris</i> pv. <i>phaseoli</i>	NCAIM	B.01695	Hungary	−
<i>X. campestris</i> pv. <i>pisi</i>	NCAIM	B.01393	N.S. ⁴	−
<i>X. campestris</i> pv. <i>raphani</i>	HRI-W	8305	UK	−
<i>X. campestris</i> pv. <i>vesicatoria</i>	BCCM/LMG	934	Brazil	−
<i>X. campestris</i> pv. <i>vesicatoria</i>	BCCM/LMG	921	USA (Long Island)	−
<i>X. euvesicatoria</i>	BCCM/LMG	918	India	−
<i>X. euvesicatoria</i>	BCCM/LMG	922	USA (Florida)	−
<i>X. euvesicatoria</i>	BCCM/LMG	921	USA (Long Island)	−
<i>X. fragariae</i> ²	CFBP	6766	USA	−
<i>X. oryzae</i> pv. <i>Oryzicola</i> ³	CFBP	2286	N.S. ⁴	−
<i>X. perforans</i> 9.2 ²	CFBP	7293	USA (Florida)	−
<i>X. perforans</i> 9.2	CFBP	8122	Thailand	−
<i>X. perforans</i> ²	DSMZ	18975	USA	−
<i>X. vesicatoria</i>	BCCM/LMG	925	Hungary	−
<i>X. vesicatoria</i> ²	CFBP	2537	New Zealand	−
<i>X. vesicatoria</i>	BCCM/LMG	920	Italy	−
Other species				
<i>Agrobacterium tumefaciens</i>	CCM	2835	Czech Republic	−
<i>Burkholderia glumae</i>	BCCM/LMG	20138	Philippines (province Jalajala Riza)	−
<i>B. glumae</i> ²	BCCM/LMG	2196	Japan (Ehime)	−
<i>Clavibacter michiganensis</i> subsp. <i>michiganensis</i>	CFBP	1460	France	−
<i>C. michiganensis</i> subsp. <i>sepedonicus</i>	NCPPB	3467	Poland	−
<i>Erwinia amylovora</i>	CRI	Ea10/96	Czech Republic	−
<i>Pantoea agglomerans</i> ²	CFBP	3845	N.S. ⁴	−
<i>Pseudomonas syringae</i> pv. <i>phaseolicola</i>	CRI	186/2	Czech Republic	−
<i>P. syringae</i> pv. <i>pisi</i>	NCPPB	3496	USA	−
<i>P. syringae</i> pv. <i>syringae</i>	NCPPB	2306	Italy	−
<i>P. syringae</i> pv. <i>tomato</i>	CRI	8119	Czech Republic	−
<i>Ralstonia pseudosolanacearum</i>	CFBP	3936	China (Guangdong)	−
<i>R. solanacearum</i>	NCPPB	2505	Sweden	−
<i>Stenotrophomonas</i> sp.	NCPPB	2859	Turkey	−

2.3. Primer Design

Partial sequence of the *hrpB* gene (GenBank ID: KX437681.1) was used for *X. gardneri* primer design [13]. One set of five primers (external primers F3 and B3; internal primers FIP and BIP and one loop primer LoopB) selected from total of three prospective primer sets was designed in PrimerExplorer software (Eiken Chemical Company, Tokyo, Japan; <https://primerexplorer.jp/e/>) and synthesized by Macrogen (Republic of Korea). Oligonucleotide sequences of the best primer set used for further testing are listed in Table 2. A sixth primer (LoopF) could not be designed due to the configuration of the template binding sites.

Table 2. Loop-mediated isothermal amplification (LAMP) primers.

Primer Name	Primer Length (nt)	T _m (°C)	Sequence (5′–3′)
F3	16	61.60	CGGGGTGCAGGTCAGC
B3	15	61.13	ACCGGCACCGCCAAG
FIP	37	-	CCACCTCGGCACGTTGCAGGCGAGGTATGCGAGTTGC
BIP	35	-	GCCGCCATCTCGCCTTGCGCCCCGATCCGATCACG
LB	17	61.26	CGAGCTGGTGGGCTTGT

2.4. Real-Time LAMP

The LAMP reactions were performed in a QuantStudio™ 6 Flex Real-Time PCR System (ThermoFisher, Waltham, MA, USA) and BioRanger™ (Diagenetix, Honolulu, HI, USA). The reaction mixture contained 12.5 µL of Isothermal master mix (Optigene, Horsham, UK), 0.2 µM of each of outer primers (F3 and B3), 1.6 µM of each of inner primers (FIP and BIP) and 0.4 µM of loop primer (only LoopB). Lastly, 3 µL of template genomic DNA (10 ng/µL) was added and the reaction was brought to a final volume of 25 µL with nuclease-free H₂O. The LAMP reaction mixtures were incubated for 60 min at 65 °C, followed by heating at 98 °C for 2 min to terminate the reactions. Melting analysis followed at temperature range from 70 °C to 99 °C with increments of 0.1 °C per second that allow for the generation of derivative melting curves. All LAMP assays were replicated at least three times, and all experiments included negative (no-template) controls.

2.5. Electrophoresis of LAMP

LAMP reaction products (5 and 10 µL, respectively) were analyzed via electrophoresis on a 1.5% agarose gel made of 1× TBE buffer (Tris/Borate/EDTA: 89 mM Tris, 89 mM boric acid, 2 mM EDTA, pH 8.3) stained with ethidium bromide (SigmaAldrich, St. Louis, MO, USA) at 90 V (4 V/cm) for 1 h. DirectLoad™ Wide Range DNA Marker was used for LAMP samples as a molecular standard for comparing molecular weight. To avoid contamination among samples, different laboratories were used for DNA extraction and for reaction mixture preparation, PCR Clean™ (Minerva Biolabs, Berlin, Germany, DE) was used for surface cleaning and only filtered pipette tips were used. After electrophoresis, the gel was visualized under UV illumination on GeneSys (Syngene, Cambridge, UK).

2.6. Colorimetric LAMP

Colorimetric LAMP reactions were carried out in a 25 µL volume containing 1.6 µM of each of inner primers FIP and BIP, 0.2 µM of each of outer primers F3 and B3, 0.4 µM of LoopB primer, 12.5 µL of 2× Colorimetric LAMP Master Mix (Cat. No. M1800, New England Biolabs) and 5 µL of DNA template (10 ng/µL). Reactions were incubated at 65 °C for 15 and 30 min in a heat block before results were recorded by naked eye.

2.7. Sensitivity and Specificity (Real-Time and Colorimetric LAMP)

Serial 10-fold dilutions of 10 ng/µL *X. gardneri* DNA ranging from 10 ng to 10 fg were used as a template for sensitivity test of DNA amplification assays. Specificity of assays was tested using *Xanthomonas* spp. strains and other related bacteria pathogenic for tomato and pepper (Table 1). No-template control (NC; water) was included in each LAMP run.

2.8. LAMP Assay on Plant Tissues

For testing of the LAMP assay with an infected plant tissue, 50 tomato and pepper plants were grown in growth chamber (Sanyo MLR-351H, Osaka, Japan) in temperature of 24 °C and 72% humidity for 4 weeks. Inoculation was done according to ISTA (2015) [24] methodology. Two youngest fully developed leaves were pierced in the area surrounding the major veins by six needles dipped in the suspension of *X. gardneri* type strain DSMZ 19127. First symptoms were observed after 3–4 weeks.

Samples for subsequent testing were taken from the boundary of healthy and infected tissue and transferred to 50 µL of TE buffer. DNA extraction was performed as described above.

3. Results

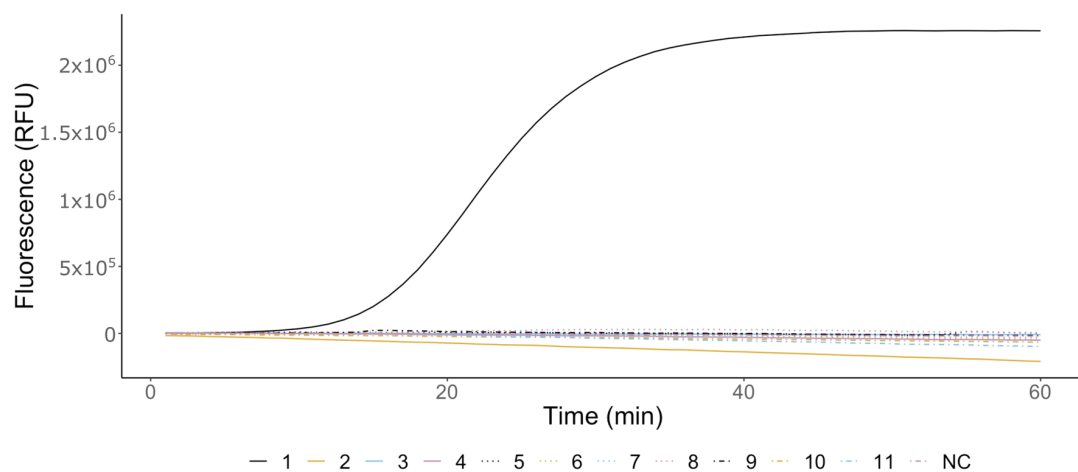
In this study, we developed a LAMP-based real-time and colorimetric assay for specific, sensitive, reliable and robust detection and differentiation of *X. gardneri*. For primer design, we selected the *hrpB* gene due to its high species specificity.

LAMP assays distinctly amplified *X. gardneri* DNA, but none of the nontarget *Xanthomonas* species, including strains closely related to the *X. gardneri*, produced amplicons within 60 min (Figure 1a, Table 1). All the real-time LAMP assays were performed using the Isothermal master mix in both a QuantStudio™ 6 Flex Real-Time PCR System and a BioRanger™ (Diagenetix, USA). Amplification of *X. gardneri* was first observable after 15 min of isothermal amplification. All reactions were subsequently analyzed on electrophoresis gel after 60 min of amplification, to confirm the results (Figure 1b). The *X. gardneri* reaction product displayed the typical ladder-like appearance of LAMP products, while no amplicons were detectable from the other reactions. Colorimetric LAMP assays were performed using 2× Colorimetric LAMP Master Mix in a heat block. No amplification was observed for any sample after 15 min. After 30 min, amplification was only visible in test tubes with *X. gardneri* DNA (Figure 1c).

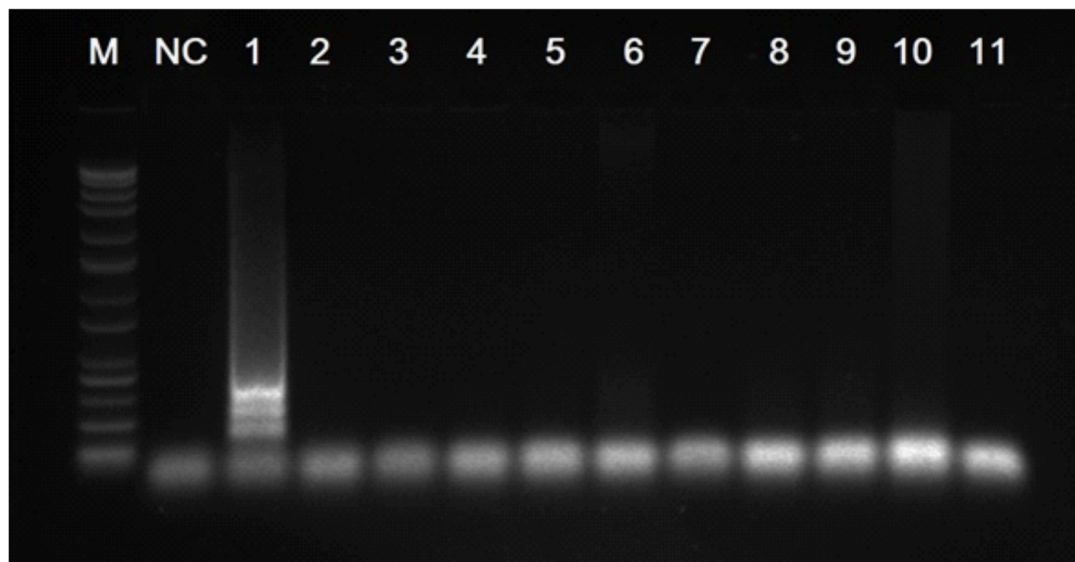
To determine the sensitivity of real-time and colorimetric assays, genomic DNA diluted to concentrations ranging from 10 ng to 10 fg was used. A no-template control (water) was included in each experimental replication. The lowest concentration limit of detection was 1 pg/µL for *X. gardneri* (Figure 2a). However, very weak amplification was also observable via electrophoresis with concentrations of 100 fg and 10 fg (Figure 2b). Considering all the variants of the assay (real-time, electrophoresis, colorimetry), we defined the lowest amount of bacterial genomic DNA required to reliably detect the bacterium to 1 pg/µL after 30 min of amplification (Figure 2a–c) as an aggregate detection limit.

All *X. gardneri* infected tomato and pepper plant samples were positive for *X. gardneri*. No positive amplification was observed when LAMP primers were tested with healthy tomato plants (Figure 3a,b). The melting curves obtained for all *X. gardneri* isolates and from infected tomato plant showed the same melting temperature (T_m) 94.27 °C, indicating similar sequences, and hence similar amplicons (Figure 3c). The possibility of using the LAMP assay in the field was tested on the BioRanger™ platform using tomato plants inoculated by *X. gardneri* (Figure 3b). Amplification was detected by 30 min (45 min on BioRanger™), suggesting that with our primers, this platform is a convenient and reliable method for point-of-care detection of *X. gardneri* if technological limitations of mobile devices are carefully considered.

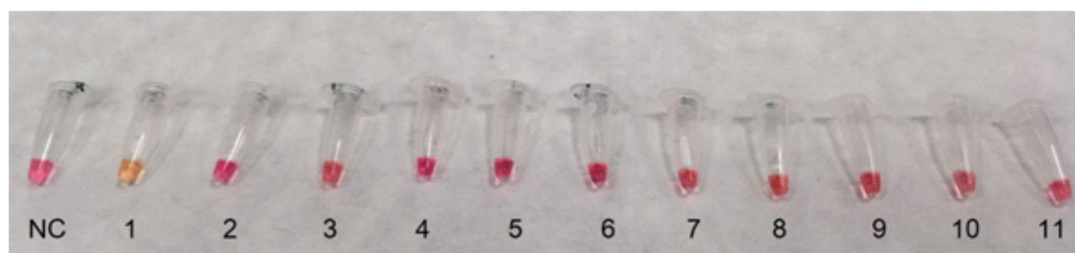
Although 30 min was enough to reliably detect the pathogen in most cases, we recommend 60 min of amplification due to possible variations in detection setup in different laboratories. The real-time nature of the LAMP assay may suggest its use for quantification of *X. gardneri* as well. However, our assay was not designed with quantification in mind, and all data resulting from following our recommendations should only be treated as detection, or not, of *X. gardneri*.



(a)

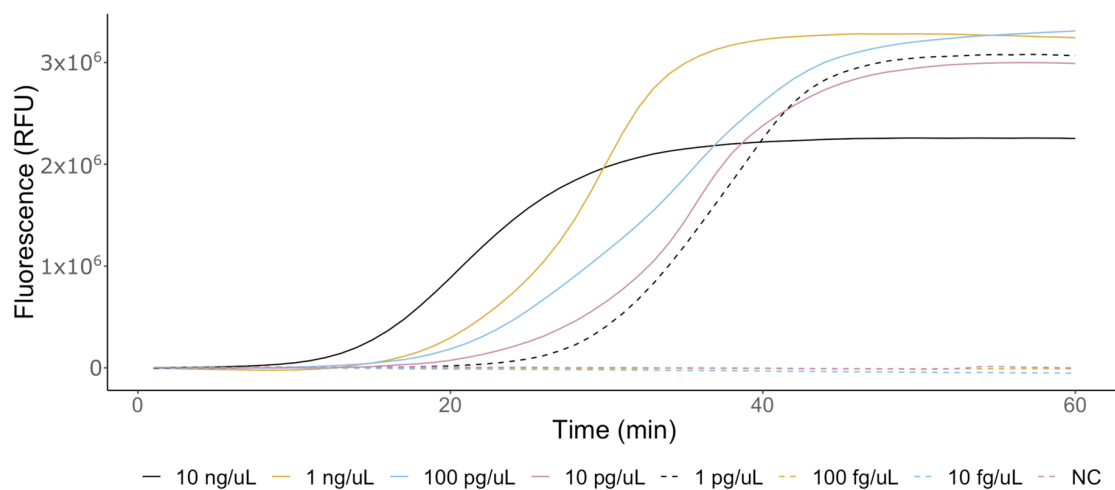


(b)

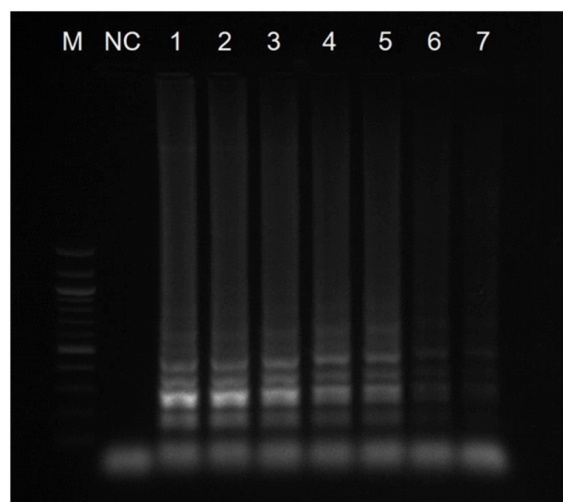


(c)

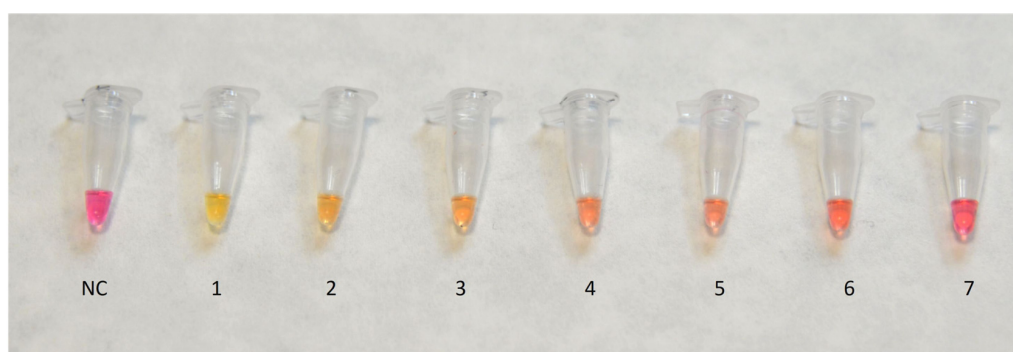
Figure 1. Specificity of selected bacterial species: (a) real-time LAMP assay (QuantStudio™ 6 Flex Real-Time PCR System); (b) gel electrophoresis of LAMP products confirms results of real-time LAMP; (c) Colorimetric Master Mix assay after 30 min reaction time; M—ladder 100 bp (NEB, Hitchin, UK), NC—negative control water, 1—*Xanthomonas gardneri* DSMZ 19127, 2—*Xanthomonas axonopodis* pv. *vesicatoria* CRI 1013, 3—*Xanthomonas perforans* DSMZ 18975, 4—*Xanthomonas vesicatoria* BCCM/LMG 920, 5—*Erwinia amylovora* CRI Ea10/96, 6—*Burkholderia glumae* BCCM/LMG 20138, 7—*Clavibacter michiganensis* susp. *michiganensis* CFBP 1460, 8—*Clavibacter michiganensis* subsp. *sepedonicus* NCPPB 3467, 9—*Pseudomonas syringae* pv. *syringae* NCPPB 2306, 10—*Pseudomonas syringae* pv. *tomato* CRI 8119, 11—*Ralstonia solanacearum* NCPPB 2505.



(a)



(b)



(c)

Figure 2. Serial 10-fold dilutions: (a) real-time LAMP assay (QuantStudio™ 6 Flex Real-Time PCR System); (b) gel electrophoresis of LAMP products confirms results of real-time LAMP; (c) Colorimetric Master Mix assay after 30 min reaction time; M—ladder 100 bp (NEB, UK), NC—negative control water, 1—10 ng/μL, 2—1 ng/μL, 3—100 pg/μL, 4—10 pg/μL, 5—1 pg/μL, 6—100 fg/μL, 7—10 fg/μL.

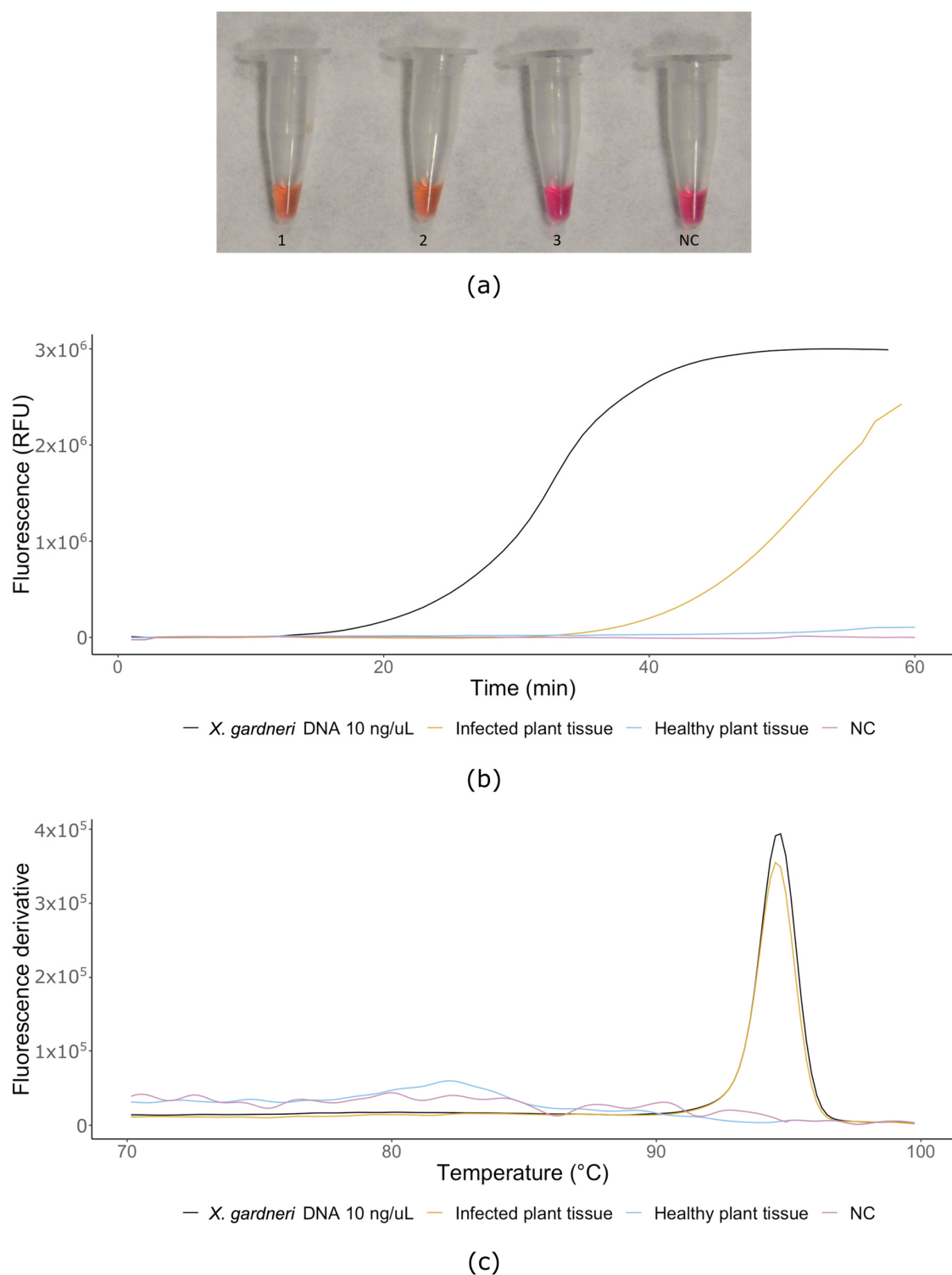


Figure 3. Naturally infected plants: (a) colorimetric LAMP assay, 1—*X. gardneri* bacterial culture DNA (10 ng/ μ L, strain DSMZ 19127), 2—infected plant sample, 3—healthy plant tissue, NC—non-template control (water); (b) real-time LAMP in BioRanger™ device; (c) melting curves of LAMP products.

4. Discussion

Here, we developed a LAMP assay to detect *X. gardneri*, one of the causal agents of BS of tomato. LAMP assays are available for some other *Xanthomonas* spp. like *X. arboricola* pv. *pruni*, the causal agent of Stone Fruit Bacterial Spot; *X. oryzae* pv. *oryzae* the causal agent of Bacterial Blight disease;

X. oryzae pv. *oryzicola*, the causal agent of Bacterial Leaf Streak disease; *X. campestris* pv. *musacearum*, the causal agent of Banana *Xanthomonas* Wilt; *X. citri* subsp. *citri* and *X. fuscans* subsp. *aurantifolii*, the causal agents of Citrus Bacterial Canker; and *X. translucens*, causal agent of cereal leaf streak [25–29]. LAMP assay for detection of *X. euvesicatoria*, the causal agent of BS of tomato, was also developed [30]. Our *X. gardneri*-specific assay substantially improves the detection of the bacteria associated with BS of tomato detection.

Different assays for detection of all the BS of tomato associated *Xanthomonas* were developed before. Most notably, many PCR based assays are available for single pathogens, and several multiplex-PCR assays for detection of all four species has also been developed [12–14,31,32]. In general, specificity of PCR based assays compared to LAMP seems to be similar. Sensitivity varies, and sometimes is not directly comparable (in some tests of sensitivity assessment bacterial cultures are diluted before DNA extraction, other tests use diluted DNA). With a detection limit 1 pg/ μ L of target DNA, our LAMP assay performs considerably better than the most sensitive *X. gardneri* molecular detection assay by Araújo et al. [31] with detection limit of 50 pg/ μ L or 5×10^4 CFU/mL.

The time to obtain results with LAMP is about 30 min, compared to one or more hours using real-time PCR or even more time-consuming end-point PCR. Although single LAMP reactions are usually slightly more expensive than PCR-based alternatives due to additional primers and more expensive polymerase, LAMP reactions can be performed on considerably cheaper and simpler devices such as the BioRanger™ or Genie® II (Optigene, Horsham, UK). This allows the use of LAMP assays in field conditions, which also makes LAMP more suitable for environments where high initial investment is not possible or desirable. Unlike PCR, LAMP does not directly provide information about size of amplified DNA, which is usually determined by performing melting analysis. For routine pathogen detection, this is not an issue, if the assay is well designed and tested. LAMP assays are, however, more difficult to design compared to PCR-based methods due to the higher complexity of the reaction and lower availability of software for LAMP assay design. Additionally, quantification is less accurate than real-time PCR [33]. Lower precision amplification devices (such as mobile devices like BioRanger™), DNA extraction method, in-field contaminants and other factors may further decrease the precision of quantification, so we suggest only qualitative interpretation of results resulting from our assay. When preparing the assay, end users should carefully consider prolonging the time of amplification with mobile devices and for less efficient DNA extraction methods with possible contamination of sample DNA. For evaluation of the data, we suggest using the multi-operator validation test [30], which provides results without further statistical analysis and is therefore suitable for point-of-care detection.

The assay developed in this work is based on the gene sequence of *hrpB* gene. The *hrp* gene cluster was successfully used before as a target for molecular detection of phytopathogenic bacteria [13,34,35]. Other common targets include genes for ribosomal RNA, *atpD*, *gumD*, *gyrB*, *rpfB* and many others, or other non-annotated sequences obtained by whole genome analysis [28,36–38]. As expected from the literature, our own DNA alignment and BLAST search, the *hrpB* gene sequence of *X. gardneri* is conserved within the BS of tomato related *Xanthomonas* (with some variable regions within the group, allowing distinction of the four species), while being distinct enough from other related or tomato specific bacteria. This allowed us to design a LAMP primer set, which is specific only for *X. gardneri*, enabling improvement of current BS-related *Xanthomonas* detection and distinction between *X. gardneri* and other species. To date, no other LAMP assay for detection of this bacteria was published.

Author Contributions: All authors participated in the conception and design of the experiments and analyzed the resulting data; D.S. and P.B. performed the experiments; D.S., P.B., S.P.C. and V.Č. wrote the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research was financially supported by the Ministry of Education, Youth and Sports of the Czech Republic within INTER-EXCELLENCE/INTER-COST programme, grant: LTC19014 (MSMT-15739/2019-8). This article is based on work from European Cooperation in Science and Technology (COST) action CA16107 EuroXanth, supported by COST. S.P.C. is supported by the AFRI Education and Workforce Development

Postdoctoral Fellowship (grant no. 2018-08122) from the U.S. Department of Agriculture, National Institute of Food and Agriculture.

Acknowledgments: The authors would like to thank Eliška Peňázová and the mentioned bacterial strain collections for providing the strains for testing.

Conflicts of Interest: The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

- Šutic, D.D. *Bakterioze crvenog patlidžana = Bacteriosis of tomatoes*; Institut za zaštitu bilja: Belgrade, Serbia, 1957; pp. 1–65.
- Dye, D.W. Cultural and biochemical reactions of additional *Xanthomonas* spp. *N. Z. J. Sci.* **1966**, *9*, 913.
- De Ley, J.; Segers, P.; Gillis, M. Intra- and Intergeneric Similarities of *Chromobacterium* and *Janthinobacterium* Ribosomal Ribonucleic Acid Cistrons. *Int. J. Syst. Bacteriol.* **1987**, *28*, 154–168. [\[CrossRef\]](#)
- Jones, J.B.; Bouzar, H.; Stall, R.E.; Almira, E.C.; Roberts, P.D.; Bowen, B.W.; Sudberry, J.; Strickler, P.M.; Chun, J. Systematic analysis of xanthomonads (*Xanthomonas* spp.) associated with pepper and tomato lesions. *Int. J. Syst. Evol. Microbiol.* **2000**, *50*, 1211–1219. [\[CrossRef\]](#) [\[PubMed\]](#)
- Jones, J.B.; Lacy, G.H.; Bouzar, H.; Stall, R.E.; Schaad, N.W. Reclassification of the Xanthomonads Associated with Bacterial Spot Disease of Tomato and Pepper. *Syst. Appl. Microbiol.* **2004**, *27*, 755–762. [\[CrossRef\]](#)
- Hamza, A.A.; Robène-Soustrade, I.; Jouen, E.; Gagnevin, L.; Lefeuvre, P.; Chiroleu, F.; Pruvost, O. Genetic and Pathological Diversity Among *Xanthomonas* Strains Responsible for Bacterial Spot on Tomato and Pepper in the Southwest Indian Ocean Region. *Plant Dis.* **2010**, *94*, 993–999. [\[CrossRef\]](#)
- Barak, J.D.; Vancheva, T.; Lefeuvre, P.; Jones, J.B.; Timilsina, S.; Minsavage, G.V.; Koebnik, R. Whole-genome sequences of *Xanthomonas euvesicatoria* strains clarify taxonomy and reveal a stepwise erosion of type 3 effectors. *Front. Plant Sci.* **2016**, *7*, 1805. [\[CrossRef\]](#)
- Constantin, E.C.; Cleenwerck, I.; Maes, M.; Baeyen, S.; Van Malderghem, C.; De Vos, P.; Cottyn, B. Genetic characterization of strains named as *Xanthomonas axonopodis* pv. *dieffenbachiae* leads to a taxonomic revision of the *X. axonopodis* species complex. *Plant Pathol.* **2016**, *65*, 792–806. [\[CrossRef\]](#)
- Parkinson, N.; Aritua, V.; Heeney, J.; Cowie, C.; Bew, J.; Stead, D. Phylogenetic analysis of *Xanthomonas* species by comparison of partial *gyrase B* gene sequences. *Int. J. Syst. Evol. Microbiol.* **2007**, *57*, 2881–2887. [\[CrossRef\]](#)
- Young, J.M.; Park, D.C.; Shearman, H.M.; Fargier, E. A multilocus sequence analysis of the genus *Xanthomonas*. *Syst. Appl. Microbiol.* **2008**, *31*, 366–377. [\[CrossRef\]](#)
- Araújo, E.R.; Costa, J.R.; Ferreira, M.A.S.V.; Quezado-Duval, A.M. Widespread distribution of *Xanthomonas perforans* and limited presence of *X. gardneri* in Brazil. *Plant Pathol.* **2017**, *66*, 159–168. [\[CrossRef\]](#)
- Koenraad, H.; Van Betteray, B.; Germain, R.; Hiddink, G.; Jones, J.B.; Oosterhof, J. Development of specific primers for the molecular detection of bacterial spot of pepper and tomato. *Acta Hortic.* **2009**, *808*, 99–102. [\[CrossRef\]](#)
- Strayer, A.L.; Jeyaprakash, A.; Minsavage, G.V.; Timilsina, S.; Vallad, G.E.; Jones, J.B.; Paret, M.L. A Multiplex Real-Time PCR Assay Differentiates Four *Xanthomonas* Species Associated with Bacterial Spot of Tomato. *Plant Dis.* **2016**, *100*, 1660–1668. [\[CrossRef\]](#) [\[PubMed\]](#)
- Cuppels, D.A.; Louws, F.J.; Ainsworth, T. Development and Evaluation of PCR-Based Diagnostic Assays for the Bacterial Speck and Bacterial Spot Pathogens of Tomato. *Plant Dis.* **2006**, *90*, 451–458. [\[CrossRef\]](#) [\[PubMed\]](#)
- Fang, X.; Li, J.; Chen, Q. One new method of nucleic acid amplification—Loop-mediated isothermal amplification of DNA. *Virol. Sin.* **2008**, *23*, 167–172. [\[CrossRef\]](#)
- Parida, M.; Sannarangaiah, S.; Dash, P.K.; Rao, P.V.L.; Morita, K. Loop mediated isothermal amplification (LAMP): A new generation of innovative gene amplification technique; perspectives in clinical diagnosis of infectious diseases. *Rev. Med. Virol.* **2008**, *18*, 407–421. [\[CrossRef\]](#) [\[PubMed\]](#)
- Zanoli, L.M.; Spoto, G. Isothermal Amplification Methods for the Detection of Nucleic Acids in Microfluidic Devices. *Biosensors* **2013**, *3*, 18–43. [\[CrossRef\]](#) [\[PubMed\]](#)
- Notomi, T.; Okayama, H.; Masubuchi, H.; Yonekawa, T.; Watanabe, K.; Amino, N.; Hase, T. Loop-mediated isothermal amplification of DNA. *Nucleic Acids Res.* **2000**, *28*, e63. [\[CrossRef\]](#)

19. Nagamine, K.; Hase, T.; Notomi, T. Accelerated reaction by loop-mediated isothermal amplification using loop primers. *Mol. Cell. Probes* **2002**, *16*, 223–229. [[CrossRef](#)]
20. Aliotta, J.M.; Pelletier, J.J.; Ware, J.L.; Moran, L.S.; Benner, J.S.; Kong, H. Thermostable Bst DNA polymerase I lacks a 3' → 5' proofreading exonuclease activity. *Genet. Anal. Biomol. Eng.* **1996**, *12*, 185–195. [[CrossRef](#)]
21. Hawwa, R.; Aikens, J.; Turner, R.J.; Santarsiero, B.D.; Mesecar, A.D. Structural basis for thermostability revealed through the identification and characterization of a highly thermostable phosphotriesterase-like lactonase from *Geobacillus stearothermophilus*. *Arch. Biochem. Biophys.* **2009**, *488*, 109–120. [[CrossRef](#)]
22. Schaad, N.W.; Jones, J.B.; Chun, W. *Laboratory Guide for the Identification of Plant. Pathogenic Bacteria*, 3rd ed.; American Phytopathological Society, APS Press: St. Paul, MN, USA, 2001.
23. King, E.O.; Ward, M.K.; Raney, D.E. Two simple media for the demonstration of pyocyanin and fluorescein. *J. Lab. Clin. Med.* **1954**, *44*, 301–307. [[CrossRef](#)] [[PubMed](#)]
24. Roberts, S.J.; Koenraadt, H. 7-019a: *Detection of Xanthomonas Campestris pv. Campestris on Brassica spp.*, 4th ed.; International Rules for Seed Testing, Chapter 7: Validated Seed Health Testing Methods; International Seed Testing Association: Bassersdorf, Switzerland, 2015.
25. Bühlmann, A.; Pothier, J.F.; Tomlinson, J.A.; Frey, J.E.; Boonham, N.; Smits, T.H.M.; Duffy, B. Genomics-informed design of loop-mediated isothermal amplification for detection of phytopathogenic *Xanthomonas arboricola* pv. *pruni* at the intraspecific level. *Plant Pathol.* **2013**, *62*, 475–484. [[CrossRef](#)]
26. Hodgetts, J.; Hall, J.; Karamura, G.; Grant, M.; Studholme, D.; Boonham, N.; Karamura, E.; Smith, J.J. Rapid, specific, simple, in-field detection of *Xanthomonas campestris* pathovar *musacearum* by loop-mediated isothermal amplification. *J. Appl. Microbiol.* **2015**, *119*, 1651–1658. [[CrossRef](#)] [[PubMed](#)]
27. Lang, J.M.; Langlois, P.; Nguyen, M.H.R.; Triplett, L.R.; Purdie, L.; Holton, T.A.; Djikeng, A.; Vera Cruz, C.M.; Verdier, V.; Leach, J.E. Sensitive Detection of *Xanthomonas oryzae* Pathovars *oryzae* and *oryzicola* by Loop-Mediated Isothermal Amplification. *Appl. Environ. Microbiol.* **2014**, *80*, 4519–4530. [[CrossRef](#)] [[PubMed](#)]
28. Langlois, P.A.; Snelling, J.; Hamilton, J.P.; Bragard, J.P.; Koebnik, R.; Verdier, V.; Triplett, L.R.; Blom, J.; Tisserat, N.A.; Leach, J.E. Characterization of the *Xanthomonas translucens* Complex Using Draft Genomes, Comparative Genomics, Phylogenetic Analysis, and Diagnostic LAMP Assays. *Phytopathology* **2017**, *107*, 519–527. [[CrossRef](#)]
29. Rigano, L.A.; Marano, M.R.; Castagnaro, A.P.; Do Amaral, A.M.; Vojnov, A.A. Rapid and sensitive detection of Citrus Bacterial Canker by loop-mediated isothermal amplification combined with simple visual evaluation methods. *BMC Microbiol.* **2010**, *10*, 176. [[CrossRef](#)]
30. Larrea-Sarmiento, A.; Dhakal, U.; Boluk, G.; Fatdal, L.; Alvarez, A.; Strayer-Scherer, A.; Paret, M.; Jones, J.; Jenkins, D.; Arif, M. Development of a genome-informed loop-mediated isothermal amplification assay for rapid and specific detection of *Xanthomonas euvesicatoria*. *Sci. Rep.* **2018**, *8*, 1–11. [[CrossRef](#)]
31. Araújo, E.R.; Costa, J.R.; Ferreira, M.A.S.V.; Quezado-Duval, A.M. Simultaneous detection and identification of the *Xanthomonas* species complex associated with tomato bacterial spot using species-specific primers and multiplex PCR. *J. Appl. Microbiol.* **2012**, *113*, 1479–1490. [[CrossRef](#)]
32. Obradovic, A.; Mavridis, A.; Rudolph, K.; Janse, J.D.; Arsenijevic, M.; Jones, J.B.; Minsavage, G.V.; Wang, J.F. Characterization and PCR-based typing of *Xanthomonas campestris* pv. *vesicatoria* from peppers and tomatoes in Serbia. *Eur. J. Plant Pathol.* **2004**, *110*, 285–292. [[CrossRef](#)]
33. Wang, H.; Turechek, W.W. A loop-mediated isothermal amplification assay and sample preparation procedure for sensitive detection of *Xanthomonas fragariae* in strawberry. *PLoS ONE* **2016**, *11*, e0147122. [[CrossRef](#)]
34. Kebede, M.; Timilsina, S.; Ayalew, A.; Admassu, B.; Potnis, N.; Minsavage, G.V.; Jones, J.B. Molecular characterization of *Xanthomonas* strains responsible for bacterial spot of tomato in Ethiopia. *Eur. J. Plant Pathol.* **2014**, *140*, 677–688. [[CrossRef](#)]
35. Leite, R.P.; Minsavage, G.V.; Bonas, U.; Stall, R.E. Detection and identification of phytopathogenic *Xanthomonas* strains by amplification of DNA sequences related to the *hrp* genes of *Xanthomonas campestris* pv. *vesicatoria*. *Appl. Environ. Microbiol.* **1994**, *60*, 1068–1077. [[CrossRef](#)] [[PubMed](#)]
36. Keremane, M.L.; Ramadugu, C.; Rodriguez, E.; Kubota, R.; Shibata, S.; Hall, R.F. A rapid field detection system for citrus huanglongbing associated 'Candidatus Liberibacter asiaticus' from the psyllid vector, *Diaphorina citri* Kuwayama and its implications in disease management. *Crop. Prot.* **2015**, *68*, 41–48. [[CrossRef](#)]

37. Golmohammadi, M.; Llop, P.; Scuderi, G.; Gell, I.; Graham, J.H.; Cubero, J. mRNA from selected genes is useful for specific detection and quantification of viable *Xanthomonas citri* subsp. *citri*. *Plant Pathol.* **2012**, *61*, 479–488. [[CrossRef](#)]
38. Simões, T.H.; Gonçalves, E.R.; Rosato, Y.B.; Mehta, A. Differentiation of *Xanthomonas* species by PCR-RFLP of *rpfB* and *atpD* genes. *FEMS Microbiol. Lett.* **2007**, *271*, 33–39. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Na tomto místě se v plné verzi habilitační práce nachází text příslušné publikace. Ten však není součástí této veřejně přístupné verze z důvodu licenčních podmínek vydavatele. Níže je k dispozici bibliografická citace a abstrakt článku. Plný text lze nalézt v placeném režimu na webových stránkách časopisu.

Stehlikova, D., Beran, P., Rost, M., & Curn, V. (2026). Genomes-based loop-mediated isothermal amplification assay for rapid and specific detection of walnut pathogen *Xanthomonas arboricola* pv. *juglandis*. *Journal of Plant Pathology*, 1-10.

Abstrakt

Xanthomonas arboricola pv. *juglandis* (*Xaj*) causes walnut bacterial blight (WBB), significantly impacting walnut (*Juglans regia* L.) production. This work aimed to develop a loop-mediated isothermal amplification (LAMP) assay for the rapid, sensitive, and specific detection of *Xaj*. Utilizing next-generation sequencing data, we identified a unique 524 nt region in *Xaj* and designed three primer sets. The *Xaj*_K13041 primer set demonstrated 100% specificity, detecting *Xaj* DNA at concentrations as low as 0.1 pg/μL (2 copies per tube) within 40 min, outperforming existing PCR methods. The LAMP assay was validated on DNA from infected walnut fruits, showing good performance compared to established methods and effective detection for future field conditions. The assay's reliability, speed and low cost make it suitable for phytosanitary services. This work presents the first LAMP-based assay for *Xaj* detection, providing a valuable tool for early and accurate identification, essential for managing WBB and ensuring stable walnut crop production.

Univerzálnost metod a jejich adaptace mimo obor fytopatologie

Ačkoliv jádro předkládané práce spočívá v diagnostice rostlinných patogenů, principy molekulární biologie a bioinformatiky jsou univerzální. Schopnost číst a interpretovat genetickou informaci není omezena taxonomickými hranicemi jedné skupiny organismů. Uvedený soubor publikací demonstruje tuto mezidisciplinární přenositelnost, kdy metodiky primárně využívané ve fytopatologii mohou být aplikovány v tak rozmanitých oblastech, jako je šlechtění drůbeže, parazitologie či kontrola pravosti potravin. Společným jmenovatelem těchto prací je snaha o přesnou identifikaci, ať už se jedná o genotyp, patogen nebo biologickou příměs.

Významnou aplikační rovinou je využití molekulárních markerů ve šlechtění a studiu genetické diverzity. V živočišné výrobě se tyto nástroje ukázaly jako klíčové při šlechtění nosnic s unikátní barvou skořápky tzv. Greenshell. Zde molekulární analýza umožnila zpřesnit fenotypovou selekci, zefektivnit šlechtitelský proces. Analogický přístup byl uplatněn v rostlinné genetice při studiu hořčice bílé (*Sinapis alba*). Využití mikrosatelitních markerů (SSR) umožnilo detailně popsat genetickou strukturu populací této plodiny, což je nezbytný krok pro zachování genetických zdrojů i budoucí šlechtění. Do této oblasti spadá i hodnocení proteinových profilů u nigerijských luštěnin, realizované v rámci konzultační činnosti pro doktorský výzkum, které přispívá k poznání nutričního potenciálu méně využívaných plodin.

Bioinformatický návrh detekčních postupů našel uplatnění také v oblasti bezpečnosti a detekce falšování potravin. V souvislosti se zaváděním hmyzího proteinu do potravinářského průmyslu vyvstala potřeba rozlišit povolené a nepovolené druhy hmyzu, které jsou ve zpracovaném stavu nerozeznatelné. Námi vyvinutá duplexní PCR metoda pro detekci potměníka moučného (*Tenebrio molitor*) a potměníka brazilského (*Zophobas morio*) v potravinách ukazuje, jak lze diagnostické principy přenést do potravinářské praxe. Podobně i v parazitologii, při studiu *Cryptosporidium felis* u koček ve střední Evropě, hraje

klíčovou roli molekulární typizace (analýza genu gp60), která odhaluje nové subtypy a pomáhá mapovat epidemiologické riziko.

Samostatnou kapitolu tvoří výzkum fyzikálních metod ochrany rostlin, konkrétně využití nízkoteplotního plazmatu. Zde se práce přesouvá od detekce k přímé intervenci. Studie prokázaly vliv plazmatu na inhibici růstu fytopatogenních bakterií a zkoumaly synergický efekt plazmatu s entomopatogenními a mykoparazitickými houbami pro zvýšení výnosu plodin.

Tato pestrá mozaika témat potvrzuje, že pokročilé výpočetní a molekulární metody netvoří uzavřenou disciplínu. Naopak, ukazují univerzálnost využití moderní biologie, která nám umožňuje řešit komplexní problémy bez ohledu na to, zda je předmětem zájmu bakterie na listu rajčete, parazit u domácího mazlíčka nebo genetický profil hospodářského zvířete.

Související publikace

Beranová, K., Jozová, E., Anderle, V., Rost, M., Zita, L., Beran, P., Tyllero-
Landová, H., Čurn, V., & Tyllér, M. (2024). Molecular tool for efficient
breeding of DOMINANT Greenshell laying hens and significant refinement
of phenotypic selection focused on eggshell color. *Poultry Science*, 103(12),
104425.

Jozová, E., Rost, M., Rychlá, A., Stehlíková, D., Pudhuvai, B., Hejna, O., Beran,
P., Čurn, V., & Klíma, M. (2023). Microsatellite Markers: A Tool to Assess
the Genetic Diversity of Yellow Mustard (*Sinapis alba* L.). *Plants* (Basel),
12(23), 4026.

Mráz, I., Beran, P., Šerá, B., Gavril, B., & Hnatiuc, E. (2014). Effect of low-
temperature plasma treatment on the growth and reproduction rate of some
plant pathogenic bacteria. *Journal of Plant Pathology*, 96(1), 63–67.

Stehlíková, D., Müllerová, V., Adámková, A., Beran, P., Adámek, M., Čurn, V.,
Škrovánková, S., & Mlček, J. (2025). Duplex PCR Detection and

Differentiation of Insect DNA *Tenebrio molitor* and *Zophobas morio* in Various Types of Food. *Insects*, 16(9), 983.

Strejckova, M., Bohata, A., Olsan, P., Havelka, Z., Kriz, P., Beran, P., Bartos, P., Curn, V., & Spatenka, P. (2018). Enhancement of the Yield of Crops by Plasma and Using of Entomopathogenic and Mycoparasitic Fungi: From Laboratory to Large-Field Experiments. *Journal of Biomaterials and Tissue Engineering*, 8(6), 829–836.

Udoh, O. E., Abu, N. E., Egonu, S. N., & Beran, P. (in press). Evaluation of protein profile of ten Nigerian underutilized legumes using SDS–PAGE. *African Journal of Agricultural Research*.

Zikmundová, V., Höflerová, E., Holubová, N., Kicia, M., Rajský, M., Beran, P., Sak, B. & Kváč, M. (2025). Novel gp60 subtypes and genetic diversity of *Cryptosporidium felis* in domestic and stray cats from Central Europe. *Parasites & Vectors*, 18(1), 479.



Molecular tool for efficient breeding of DOMINANT Greenshell laying hens and significant refinement of phenotypic selection focused on eggshell color

Kristina Beranova^{a,1}, Eva Jozova^a, Vojtech Anderle^b, Michael Rost^a, Lukas Zita^c, Pavel Beran^a, Helena Tyllerovalandova^b, Vladislav Curn^a, Milan Tyller^b

^a University of South Bohemia, Faculty of Agriculture and Technology, Department of Genetics and Biotechnology, České Budějovice, Czech Republic

^b Dominant Genetika s.r.o., Voleč, Czech Republic

^c Czech University of Life Science Prague, Faculty of Agrobiological Sciences, Department of Animal Science, Prague, Czech Republic

ARTICLE INFO

Key words:

Czech blue-shelled breed
SLCO1B3
L*a*b* color space
PCR marker
CART methodology

ABSTRACT

In recent decades, interest in non-traditionally colored eggs has increased. For breeders, this market interest means breeding lines of laying hens that lay eggs of varied colors, such as the blue-green eggshells (Dominant Greenshell) in this study. This study presents the results of genotyping the polymorphism of the *O* locus responsible for shell coloration and photometric measurement of eggshell color based on the CIELab system, which was carried out on the unique Czech breeding population Dominant Greenshell. The aim was to use a combination of phenotyping using the CIELab System method and genotyping of the *O* locus using the end-point PCR approach with the main focus on the accuracy of distinguishing shell color genotypes, streamlining the selection of dominant homozygotes in the *O* locus, optimizing this technology for the most efficient and cost-effective selection procedure in practical hen breeding. The optometric method was able to reliably distinguish only dominant and recessive phenotypes and eliminate from the population only undesirable recessive homozygotes with a white colored shell. The parameter *a** (redness/greenness) from the CIELab color space turned out to be absolutely key for distinguishing dominant and recessive phenotypes. Using the CART methodology, a classification tree built on discriminating optometric characteristics *a*-blunt was obtained, however, for the group of desirable *O/O* homozygotes, the selection approach would result in incorrect genotyping of 31% of individuals. Therefore, a combined approach based on rapid and simple elimination of recessive homozygotes using phenotyping (CIELab photometric measurement) and molecular identification of the EAV-HP insertion in the *SLCO1B3* gene in dominant phenotypes, regardless of color intensity affected by laying time/order, and allowing reliable elimination, has proven to be the most effective method to distinguish heterozygotes from the breeding population. The combination of optometric and molecular selection methods then leads to more efficient selection, reduction of overall selection costs. This process led to the stabilization of the breeding population within one generation and the achievement of a pure homozygous line with regard to eggshell color.

INTRODUCTION

Eggshell quality is one of the main breeding goals for laying hen breeders. In many countries, eggshell color is an important indicator of egg quality, with great importance in the sales and marketing strategies of retailers (Dalirfatef et al., 2015). Although no direct relationship has been demonstrated between eggshell color and nutritional values of eggs (Scott and Silversides, 2000), consumers in different countries around

the world prefer a certain coloration, even despite the higher purchase price (Li et al., 2006). In the last few decades, hen breeds laying eggs with a less traditional blue to light green color, originating in Chile, have begun to appear on the markets of Western Europe, North America, and Japan. These breeds include Araucana, Ameraucana, Dongxiang, Lushi and many others (Lukanov et al., 2015). Consequently, lines of laying hens and dual-purpose hybrids have been bred that lay blue-green colored eggs.

Scientific section: Genetics and Genomics

1 Corresponding author: University of South Bohemia, Faculty of Agriculture and Technology, Department of Genetics and Biotechnology, Na Sádkách 1780, 370 05 České Budějovice, Czech Republic. Phone Number: +420 725 595 209

E-mail address: kri.beranova@gmail.com (K. Beranova).

<https://doi.org/10.1016/j.psj.2024.104425>

Received 5 August 2024; Accepted 11 October 2024

Available online 15 October 2024

0032-5791/© 2024 The Authors. Published by Elsevier Inc. on behalf of Poultry Science Association Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

The blue eggshell color in hens is caused by the accumulation of the bile pigments biliverdin and protoporphyrin IX in the surface layer of the eggshell and is determined by an autosomal dominant trait traditionally called oocyan (O) (Punnett, 1933; Zhao et al., 2006; Wragg et al., 2013; Chen et al., 2020). For a long time, the genetic determination of eggshell coloration was not being known and until recently Wang et al. (2013) reported that the blue eggshell is caused by an ~4.2 kb EAV-HP insertion in the 5'-flanking region of the *SLCO1B3* gene that promotes the expression of *SLCO1B3* gene in shell gland of the oviduct in hens. Based on these studies it was possible to analyze breeding populations and to predict blue eggshell phenotype at the molecular level. *SLCO1B3* gene encodes the membrane transporter OATP1B3, which is responsible for the transport of amphipathic organic compounds including bile salts. This mutation in the promotor region activates the expression of the *SLCO1B3* gene in the shell gland of the oviduct and can enhance transportation of biliverdin to eggshell (Wang et al. 2013; Chen et al. 2020). Chen and Wang (2022) found positive correlation between *SLCO1B3* gene expression in shell glands and biliverdin concentration in eggshell which supported role of OATP1B3 in pigment transport and formation of blue-green colored eggs.

Following this discovery, new studies were carried out, for example, Darwish et al. (2019) conducted a genome-wide association study (GWAS) on the Chinese Dongxiang chicken breed, investigating four traits determining hue and color intensity. In total, 10 SNPs were identified that were demonstrably associated with this trait. Additionally, Chen et al. (2020) confirmed the effect of EAV-HP insertion on *SLCO1B3* gene expression. It is anticipated that with the rapid development of high throughput DNA techniques, further and more detailed studies will be forthcoming.

In the breeding process, continuous selection is carried out to obtain eggs not only of a certain color, but also its intensity and uniformity are essential (Lukanov, 2014). High color uniformity also indicates consistent production performance and higher breeding degree (Liao et al., 2021). Until the early 1960s, studies that addressed the possibilities of eggshell color classification were focused on subjective evaluation, e.g., Farnsworth and Nordskog (1955) and Godfrey (1949). Subjective evaluations are not entirely satisfactory due to variables such as lighting, observer's color velocity, limited number of categories, and changing standards (Wei and Bitgood, 1990). A decade later, the use of photoelectric equipment for objective color measurement was published (Hunton, 1962). The development of specific spectrophotometers that measure the intensity and shade of the eggshell has brought significant improvements in selection for layer breeders. Today, the L*a*b* color system (also known as the CIELab System) is the most used for this purpose and is one of the most widely used systems in many areas of color measurement to date (Fairchild, 2013). The measured values can thus be compared at different ages for the same group of hens. From a breeder's point of view, it is essential that the variation within a breeding stock can be analyzed as a normally distributed quantitative trait and can be used for the purpose of systematic selection in the direction of consumer's preferences (Cavero et al., 2012). As far as is known, this system of selection is routinely used in breeding of white and brown-shelled layers (Cavero et al., 2012; Hunton, 1962), to our knowledge there is considerably less data on blue-shelled breeds. Ongoing studies try to identify the most accurate selection procedure for uniformity of blue shell color using the L*a*b* color space also in combination with molecular analysis and statistical models of different difficulty level.

The color variation in blue eggs was studied by Guo et al. (2020). Based on L*a*b* color space measurements, the results attribute this phenomenon to the strong influence of dominant genetic effects. They also suggested the use of their findings in a planned mating system for commercial egg production. Another study conducted by Zeng et al. (2022) used the L*a*b* color space to predict the amount of deposited eggshell pigment as a tool for rapid eggshell color selection in the breeding system. Further, Wang et al. (2024) successfully mapped

functional variants associated with hen eggshell color in *SLCO1B3*, *CPOX*, *ALAS1*, and *ABCG2* genes with the L*a*b* values. Correlation between eggshell colors and eggshell pigments in different subtypes of blue-shelled eggs was studied by Wang et al. (2023). The study revealed correlation between eggshell colors and eggshell pigment content.

Dominant Genetika Ltd., hens leading breeding company in the Czech Republic, has highly developed breeding programs focused on producing non-traditionally colored eggs (so called Easter Eggers). The selection is currently being performed using traditional breeding methods and photometric approaches. Recently the CIELab System has been introduced for selection for eggshell color intensity and PCR-based approaches to identify individuals with EAV-HP insertion in *SLCO1B3* gene.

Specifically, we address the following questions. (1) Will photometric measurements provide outputs on the basis of which individual genotypes can be distinguished? Does egg pigmentation decrease during one clutch and is the photometric analysis and determination of the genotype of individual hens affected? (2) Is it possible to reliably differentiate genotypes on the basis of molecular analysis, and will this analysis enable routine detection of individual alleles? (3) Is it possible to appropriately combine phenotyping and genotyping in a breeding population? As is known, selection using phenotypic approaches is laborious and time-consuming, including some degree of inaccuracy, since it usually takes several generations to achieve the desired goal of eliminating the recessive allele from the population. This paper summarizes the results of the first study focusing on the combination of photometric and PCR-based approaches for the time- and cost-effective selection of desirable individuals/genotypes and the design of an appropriate technology for the successful selection of dominant homozygous individuals in Greenshell lines.

MATERIALS AND METHODS

Ethic Statement

The animals were kept in a private breeding facility that meets the requirements of the Guide for the Care and Use of Agricultural Animals in Research and Teaching. All the material for analyses were collected by a veterinarian specialized in poultry in accordance with the Regulation No. 176/2023 on the protection and rights of animals in force in the Czech Republic.

Animals and Sample Collection

Two groups (119 and 2745 individuals) of the original Czech, blue-shelled DOMINANT Greenshell line were tested. This line was developed by breeding the original South American Araucana breed with the high-performance Dominant lines while maintaining and improving the color of the shell. It differs from the original breed mainly by higher performance parameters (higher laying rate, higher egg weight, more stable shell color intensity until the end of the laying cycle, earlier onset of laying), but also by calmer temperament. The Greenshell line is characterized by autosexing, where it is already possible to distinguish the sex of day-old chicks on the basis of the color of the primary plumage. Overall, the coloration of adult animals is quite varied. The main plumage color is greyish brindle, with brownish brindle on the breast and a distinct silver color with a noticeable transverse brindle on the neck. The male appears lighter overall, otherwise similar in color to the hen. The color of the auricle is white, the crest is relatively large, often connate due to the large scape. The color of the treads is yellow to white. The live weight of the hen is 1.8 to 2.5 kg, that of the male 2.0 to 2.8 kg. The color of the shell is turquoise blue to greenish. The average weight of the eggs is 59 g. In contrast to the Araucana breed, the Dominant Greenshell line does not have the absence of the last links of the tailless tail and does not have the ear tufts, which is the plumage on the chin forming a kind of beard. In contrast, the Greenshell line has a

distinctive crest on its head. All tested birds were kept under the same conditions in individual boxes in the provider's private facilities (Dominant Genetika Ltd, Czech Republic). A group of 119 laying hens was tested by photometric measurement of eggshell color in combination with molecular markers. A total of 357 fresh eggs from 119 DOMINANT Greenshell laying hens at 40-41 weeks of age were photometrically evaluated. Eggs were collected during different stages of the laying cycle on three consecutive days to cover the natural loss of pigment in the shell during the laying cycle. Three eggs were collected from each hen and described by an identifier for the hen and the day of lay. The eggs were stored at room temperature and standard humidity (18-20°C, 45% RH) until the completion of the experiment. The eggs were processed within one week to prevent distortion of results due to gradual pigment degradation. Blood and feather samples from all 119 hens were used for DNA extraction and an appropriate method of molecular genotyping was proposed. The optimized PCR genotyping method was used for screening of large breeding population. This breeding population included a total of 2745 individuals. Of this number, 178 were males and 2567 were females.

Photometric Analysis

The eggshell color was evaluated by an objective photometric method using a QCR reflectometer (TSS, York, UK), operating on the principle of determining the percentage of light reflected by the eggshell surface within the interval from 0 to 100 %. Higher values correspond to lighter eggshell shades. Reflectivity was measured at three points on the eggshell: the sharp, middle and blunt ends. Instrumental color measurements were recorded in the CIELab color space for L* (lightness – 0 = black, 100 = white), a* (redness/greenness - positive values = red, negative values = green), and b* (yellowness/blueness - positive values = yellow, negative values = blue) using a spectrophotometer (CM-700d, Konica Minolta, Osaka, Japan). The instrument has been used in accordance with the manufacturer's instructions. The following parameters were measured: "ref_blunt", "ref_middle", "ref_sharp", "L_blunt", "L_middle", "L_sharp", "a_blunt", "a_middle", "a_sharp", "b_blunt", "b_middle", "b_sharp".

DNA Extraction and Genotyping Analysis

Two types of input material were used to optimize DNA extraction. The first one was peripheral blood collected professionally in 3 ml VACUETTE® haematology collection tubes with K3 EDTA (Greiner Bio-one GmbH, Austria). The second type of material was freshly collected feathers. The material was stored at 4°C until processing. DNA extraction from peripheral blood was performed using a commercial kit NucleoSpin Blood (Machery-Nagel, Germany) according to the manufacturer's instructions. DNA from feathers was extracted according to the in-house protocol using a minor modified CTAB PVP method (Doyle, 1991).

Genomic DNA was amplified by end-point PCR in two different variants. The first variant was adopted from the protocol published by Wang et al. (2013) using three primers test-nor-up 5'TTTGACCAGCGTAGATAA 3', test-nor-down 5' ATGTTAGCAGTG-TAGTTG 3' a test-eav 5' TAGGTTCCGAACGCGATGT 3'. The second variant was performed according to the protocol of Wragg et al. (2013) using primers Multiplex PCR F 5'-GCATTTTCAAAACGGGTGTA, Multiplex PCR R1 5'-CAAACACACAAGGTAATGTTCA-3 and Multiplex PCR R2 5'-CCCAGCAGTAAGCCCTACAT-3. The total volume of the PCR reaction was set up in a volume of 10 µl containing 1x reaction buffer (150 mM Tris-HCl, pH 8.8 (at 25°C), 40 mM (NH₄)₂SO₄, 0.02% Tween 20, 5 mM MgCl₂, 400 µM dATP, 400 µM dCTP, 400 µM dGTP, 400 µM dTTP, 100 U/ml Taq DNA polymerase, dye, stabilisers and additives (PPP Master Mix, Top-Bio, CZ), 20 pmol from each primer and 50 ng of template DNA. The microtubes were placed in a thermocycler (TPro-professional Basic Gradient, Analytik Jena, Germany) under the following

conditions: initial denaturation at 94°C for 5 min, 36 cycles 30 sec at 94°C, 30 sec 53.5°C, 20 sec 72°C, final extension 5 min 72°C for primers test-nor-up, test-nor-down and test-nor-eav. For primers Multiplex PCR F, Multiplex PCR R1 and Multiplex PCR R2 the temperature profile is as follows: initial denaturation at 94°C for 5 min, 30 cycles 10 sec at 94°C, 30 sec 56°C, 30 sec 72°C, final extension 5 min 72°C. Products were separated on a 2% agarose gel stained with ethidium bromide and then visualized using an InGenious3 device (Syngene, UK).

Data Analysis

To evaluate the effect of different genotypes on egg color intensity, a set of measurements obtained using a CM-700d spectrophotometer (Konica Minolta, Osaka, Japan) and a QCR reflectometer (TSS England) were analyzed. Subsequently, basic descriptive statistics were determined along with 95% asymptotic confidence intervals. Differences in the studied traits with respect to O/O, O/o, o/o genotype were assessed using a mixed linear model that accounted for the hierarchical (nested) structure of the data. The restricted maximum likelihood (REML) algorithm implemented in the "lmer" library was used for numerical estimation (Bates et al., 2015). Genotype was considered as a fixed factor, while individual hens and their clutches were considered as random factors nested within genotype, formally expressed as:

$$Y_{ijk} = \beta_0 + \beta_i \text{GENOTYPE} + u_{ji} + \epsilon_{ijk}$$

Y_{ijk} : The value of k -th observation of a particular variable (ref_blunt, etc.) for the i -th level of genotype and the j -th level of hen factor.

β_0 : Overall mean (intercept).

β_i : Genotype fixed effect. Where coefficient β_i indicates the average difference in the value of a particular variable for a i -th genotype from the overall mean/reference.

u_{ji} : Indicates the random effect for the j -th hen within the i -th genotype. The vertical bar "|" emphasizes that the j -th hen is nested within the i -th genotype. Where $u_{ji} \sim N(0; \sigma_u^2)$.

ϵ_{ijk} : The error term (residual), which captures all remaining variability not explained by either fixed or random effects. Where $\epsilon_{ijk} \sim N(0; \sigma^2)$.

Statistical significance was defined as a p -value less than 0.05. To illustrate differences in measured values for different genotypes, the principal components method (Jolliffe, 2002) was used with dimension reduction to two latent variables. At the same time, simple classification rules for approximate prediction of genotype based on egg color measurements were obtained using the Classification Regression Tree Method (CART) by Breiman et al. (2017). All necessary numerical calculations and graphs were performed using the R 4.4.0 programming environment (R Core Team 2024) with the appropriate libraries loaded.

RESULTS

Genotyping of Pilot Population

The analyses were first performed on a test set of 119 laying hens from which a total of 357 eggs were collected. These eggs were collected at different three consecutive laying days, to cover the possible change in pigmentation intensity of the shell during laying. Two types of input material - peripheral blood and freshly collected feathers - were tested for genomic DNA extraction. The genotype of all individuals was successfully determined by PCR analysis regardless of the origin of the input DNA sample and both DNA extraction methods gave identical results. Two different PCR procedures according to Wang et al. (2013) and Wragg et al. (2013) have been optimized to evaluate genotype using molecular markers. Figs. 1b and 1c show a comparison of results obtained by both of these protocols and it can be seen that both procedures lead to the same results providing identical genotyping. Based on the size of the amplified products, three genotypes were then evaluated: The

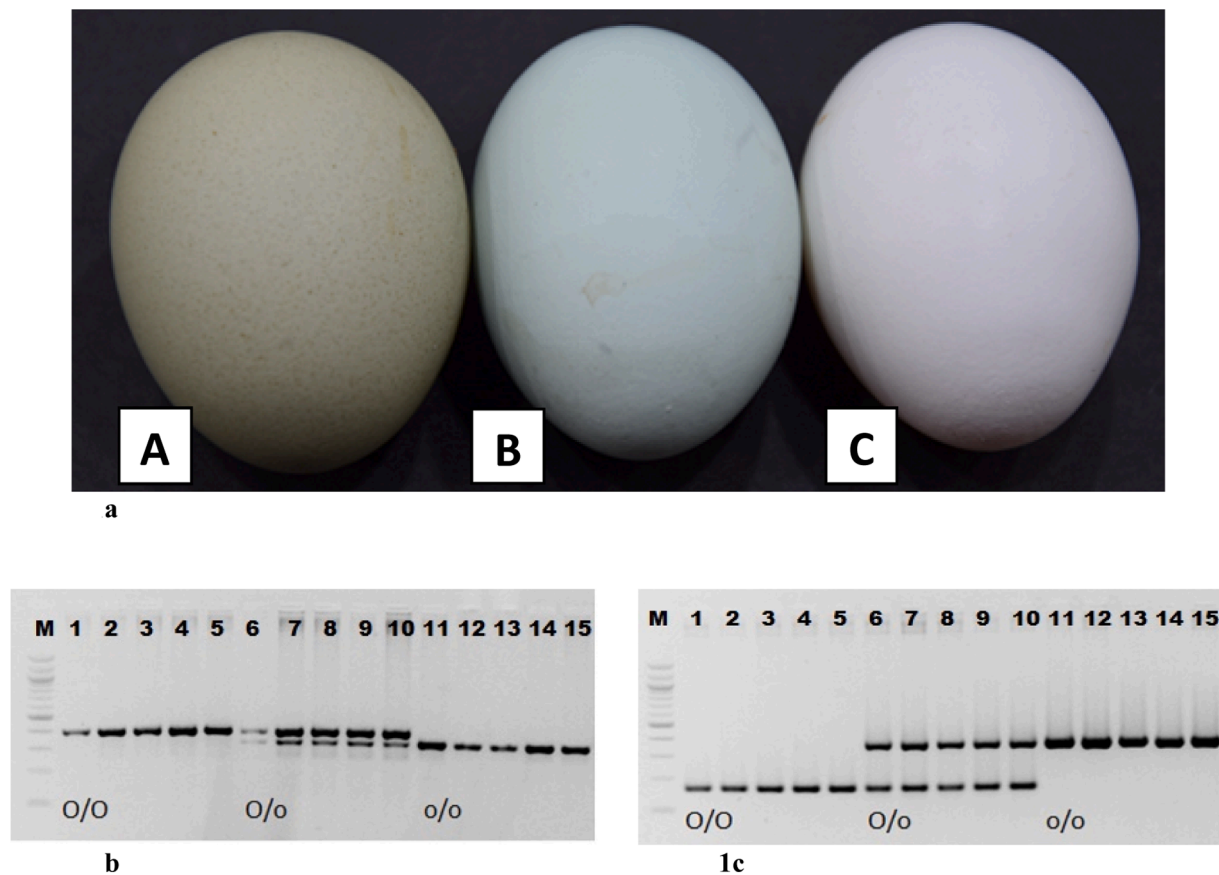


Figure 1. Results of PCR genotyping of blue-shelled line DOMINANT Greenshell. **Fig. 1a:** Eggs A and B have typical coloration for blue egg lines for O/O and O/o genotyped layers. Egg C is pure white in color typical for an o/o genotype. **Fig. 1b:** Results of PCR analyses according to Wang et al. (2013), where in O/O genotype was amplified 425 bp allele (samples 1-5) and in o/o genotype 340 bp allele (samples 11-15), both alleles are present in heterozygote O/o (samples 6-10). **Fig. 1c:** Results of PCR analyses according to Wragg et al. (2013), where in O/O genotype was amplified 167 bp allele (samples 1-5) and in o/o genotype 364 bp allele (samples 11-15), both alleles are present in heterozygote O/o (samples 6-10). Real IDs of the specific genotypes are given in Table S2, supplementary material.

O/O , which is the dominant homozygote genotype determining the green or blue color of the shell, known as green shell, the heterozygous genotype O/o and the recessive homozygote o/o , which is manifested by the white color of the shell. A representative photo of eggs of each genotype is shown in Fig. 1a. In a test set of 119 laying hens, 59 dominant homozygous O/O , 48 heterozygous O/o and 12 recessive homozygous o/o genotypes were detected by molecular genotyping.

Assessment of The Outputs of Photometric and Molecular Analyses

Reflectivity and color characteristics in the CIELAB color space were recorded for all 357 analyzed eggs from 119 laying hens from pilot test population. Descriptive characteristics of individual variables recorded by optometric methods within different genotypes are given in Table 1. For detailed information on laying hens, analyzed eggs, results of photometric analyses and genotyping data see Table S1 in supplementary data. A linear mixed model was chosen for the statistical evaluation of the results of the photometric analyses, where the random factor “hen” is nested within the levels of the fixed factor “genotype”. Table 1 presents the p -values and σ_u , σ values. The σ_u values for all measured parameters indicate that there is considerable variability in the mean value of the corresponding photometric index among individual combinations genotype:hen. The estimated value of σ represents the residual variability. Based on this linear mixed model, we can conclude that both the fixed factor “genotype” and the random factor “hen” have a significant effect on the value of the photometric indices. The variability among different combinations of genotype:hen is relatively high. This suggests that the interaction among genotype and hen plays a significant

role in determining the value of the photometric indices. The ICC (Intraclass Correlation Coefficient) values for all measured indices besides the L^* parameter are close to 1, indicating that most of the variability in the data is due to differences among individual hens with respect to genotype. The results of the molecular genotyping were compared with photometric data using the PCA analysis and displayed through a biplot (Fig. 2). The significance of differences in group means for each variable with respect to genotypes was examined through a one-factor fixed effects analysis of variance. The biplot captures quite clearly the distinctiveness of the group of recessively homozygous individuals – Group A (o/o , marked red in Fig. 2) from a group of individuals with dominant phenotype, a green-colored shell, Group B. These individuals with O/O (marked blue in Fig. 2) and O/o (marked green in Fig. 2) genotypes are not so obviously separated anymore. In addition to the genotypic characteristics, biplot represents the weight of the optometric characteristics (reflectance and color characteristics) and there is visible a high correlation of the variables a_{middle} , a_{blunt} and a_{sharp} with the second main component, in which the separation of the group of individuals with o/o genotype is apparent.

Subsequently, the discriminatory power of individual variables, respectively their suitability for screening purposes, was estimated using logistic regression. Given the identified presence of “complete separation” and the resulting Hauck-Donner phenomenon, i.e. the inability to finitely estimate regression coefficients using the maximum likelihood method, the more appropriate CART methodology was subsequently used. The resulting classification tree (Suppl. Figure S1), built on discriminating optometric characteristics a_{blunt} , can clearly distinguish individuals with a recessive phenotype. The first group includes

Table 1
Descriptive characteristics of individual variables within different genotypes

Variable	Genotype	¹ Mean ± sd	² C.I.	³ Obs	⁴ p- value	⁵ σ_u ⁶ σ ⁷ ICC
Ref_blunt	O/O	64.7 ±7.82	[63.5; 65.8]	177	<0.001	6.86; 2.63
	O/o	64.6 ± 6.15	[63.6; 65.7]	144		0.87
	o/o	54.8 ± 8.73	[51.8; 57.7]	36		
Ref_middle	O/O	66.3 ± 7.50	[65.2; 67.4]	177	<0.001	6.66; 2.55
	O/o	66.1 ± 6.24	[65.1; 67.2]	144		0.87
	o/o	58.7 ± 8.09	[55.9; 61.4]	36		
Ref_sharp	O/O	64.7 ± 7.43	[63.6; 65.8]	177	<0.001	6.42; 3.14
	O/o	64.3 ± 6.46	[63.2; 65.3]	144		0.81
	o/o	56.5 ± 7.83	[53.9; 59.2]	36		
L_blunt	O/O	80.9 ± 5.67	[80.1; 81.8]	177	<0.001	4.57; 3.52
	O/o	78.5 ± 5.77	[77.6; 79.5]	144		0.63
	o/o	74.2 ± 5.97	[72.2; 76.2]	36		
L_middle	O/O	82.8 ± 4.63	[82.1; 83.5]	177	<0.001	3.54; 3.01
	O/o	82.8 ± 4.38	[82.1; 83.6]	144		0.58
	o/o	76.1 ± 5.49	[74.3; 78.0]	36		
L_sharp	O/O	79.0 ± 6.35	[78.1;80.0]	177	=0.039	4.14; 4.89
	O/o	78.9 ± 6.36	[77.8;79.9]	144		0.42
	o/o	75.0 ± 6.64	[72.8;77.3]	36		
a_blunt	O/O	-5.08 ± 1.26	[-5.26; -4.89]	177	<0.001	1.39; 0.66
	O/o	-3.81 ± 1.26	[-4.02; -3.60]	144		0.81
	o/o	5.61 ± 3.03	[4.59;6.64]	36		
a_middle	O/O	-5.27 ± 1.25	[-5.46; -5.09]	177	<0.001	1.42; 0.65
	O/o	-4.20 ± 1.36	[-4.43; -3.98]	144		0.83
	o/o	5.36 ± 3.01	[4.35; 6.38]	36		
a_sharp	O/O	-5.00 ± 1.57	[-5.23; -4.77]	177	<0.001	1.40; 0.90
	O/o	-3.81 ± 1.36	[-4.03; -3.59]	144		0.71
	o/o	5.26 ± 2.79	[4.32;6.21]	36		
b_blunt	O/O	9.59 ± 3.74	[9.03; 10.1]	177	<0.001	3.34; 1.52
	O/o	10.7 ± 3.52	[10.1; 11.3]	144		0.83
	o/o	19.1 ± 3.61	[17.9; 20.3]	36		
b_middle	O/O	9.18 ± 3.72	[8.63; 9.73]	177	<0.001	3.35; 1.48
	O/o	10.2 ± 3.56	[9.65;10.8]	144		0.84
	o/o	18.6 ± 3.57	[17.4;19.8]	36		
b_sharp	O/O	8.83 ± 3.77	[8.27;9.39]	177	<0.001	3.36; 1.54
	O/o	9.82 ± 3.58	[9.23;10.4]	144		0.83
	o/o	18.2 ± 3.56	[17.0;19.4]	36		

¹ Mean and standard deviation,

² 95 % confidence interval for mean;

³ Number of observations;

⁴ p-value obtained from a test of significance of a fixed effect (e.g. genotype) from the LMM set using REML;

⁵ σ_u variation between different genotype-hen groups (hen nested within genotype) is expressed in standard deviation;

⁶ σ residual variability expressed in standard deviation;

⁷ ICC = Intraclass Correlation Coefficient.

individuals with dominant homozygous genotype (150 individuals) and heterozygotes (66 individuals), and the second group includes heterozygotes (78 individuals), but also 27 dominant homozygotes. Classification rules for the resulting tree are described in the Table 2. In the second step, two groups can be distinguished with some degree of uncertainty/error rate. Concerning the group of desirable O/O homozygotes, the selection approach would lead to an incorrect genotype determination in 31% of the individuals (see error matrix in Table 3).

Analysis of Breeding Population

Since phenotypic data fully correlated with genotyping results within the experimental group (see genotype table and laying hens ID in Table S1), the protocol of molecular analysis was subsequently applied to the entire breeding population of the DOMINANT Greenshell line covering 2745 individuals. This population consisted of 178 males and 2567 hens. In the case of males, 146 individuals carrying the O/O genotype, 26 individuals of O/o and 6 individuals of o/o genotype were detected. Undesirable genotypes (O/o and o/o) were detected in a male's group at 18%. A total of 2030 O/O genotypes, 488 O/o genotypes and 49 o/o genotypes were detected in hens. Thus, 20.9% of undesirable (O/o and o/o) genotypes were detected in hen group, 2.9% more than in males. Overall, the analyzed breeding population included 2176 dominant homozygotes with the genotype O/O (79.3%), 514 heterozygotes with the O/o genotype (18.7%) and 55 recessive homozygotes with the o/o genotype (2.0%). Thus, 20.7% of undesirable genotypes were present across the analyzed population.

DISCUSSION

The genetic background of the individual hen and breed/variety is a determining factor for eggshell color (Li et al., 2006). In the same way, the uniformity and intensity of coloring is often an indicator of egg quality for consumers, i.e. also a breeding goal of the producer. However, the color of the shell may also vary depending on some other factors, such as the order of the egg laid (Bi et al., 2018), the genotype of the individuals (Wang et al., 2013) and the level of expression of the *SLCO1B3* gene encoding the membrane transporter OATP1B3 (Chen et al. et al., 2020; Chen and Wang, 2022). The effects of nutrition, health, and age are also reported; however, these factors were not found to be significant (Hargitai et al., 2018; Samiullah et al., 2015).

In our study with regard to the methods and approaches used, three hypotheses have been defined. The first hypothesis assumed that the photometric method would provide data on which it will be possible to determine and distinguish all genotypes. Based on our photometric analysis, it would not be possible to determine the genotype of individual hens and the photometric values of all hens would permeate. Results of PCA analysis (Fig. 2) showed two separate groups: Group A - containing only individuals with genotype o/o and Group B - containing individuals with genotypes O/O and O/o. Thus, individuals carrying only recessive o alleles could only be truly distinguished by photometric measurement without the use of a PCR test. The results showed that a* value is the most sensitive index in evaluating shell color change in the CIELab system / color space and thus supported the results reported by Chen and Wang (2022). On the basis of optometric measurements, there are demonstrable differences between these two groups of genotypes – genotypes with the presence of EAV-HP insertion and without this

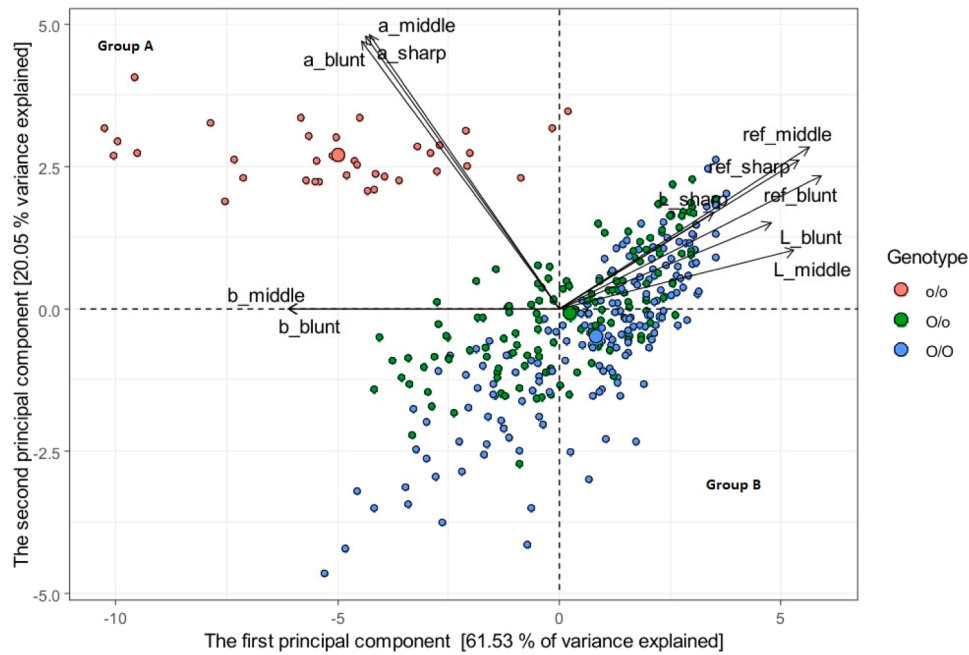


Figure 2. The biplot quite clearly captures the distinctiveness of the group of recessively homozygous *o/o* individuals (Group A) when compared to the group of *O/o* and *O/O* individuals, which are not so obviously separated anymore (Group B). Also evident is the high correlation of the variables *a_middle*, *a_blunt* and *a_sharp* with the second principal component, in the direction of which the separation of the group of individuals with the *o/o* genotype is evident.

Table 2

Classification rules for the resulting tree based on photometric analysis values for parameter *a_blunt*

Genotype ¹	Rules ²
<i>o/o</i> cover 36 (10%)	When <i>a_blunt</i> $\geq$ 0.445
<i>O/O</i> cover 216 (61%)	When <i>a_blunt</i> is -3.925 to 0.445
<i>O/o</i> cover 105 (29%)	When <i>a_blunt</i> < -3.925

¹ The genotype was initially determined by PCR analysis. In combination with the results of photometric measurements, a CART analysis was performed, the outputs of which served as rules for genotyping based on photometry.

² Rules: *o/o* was classified at values $\geq$ 0.445, covering 36 samples. Within the range of values -3.925 to 0.445, genotype *O/O* was classified, covering 216 samples. *O/o* genotype was classified at values < -3.925 covering 105 samples.

Table 3

The error rate matrix

Genotype	<i>o/o</i>	<i>O/o</i>	<i>O/O</i>
<i>o/o</i>	1.00	0.00	0.00
<i>O/o</i>	0.00	0.74	0.26
<i>O/O</i>	0.00	0.31	0.69

If the *o/o* genotype is detected, 0% of individuals are misclassified by photometry. For the *O/o* genotype is the error rate 31% and for the *O/O* genotype it is 26%.

insertion. This finding provides important support for the consistency between the distribution of *a** values alone and the genotypes. These results suggest that this parameter may be of high interest for selection. Li et al. (2006) observed the variation of eggshell color in different-colored chicken lines by measuring the *L*a*b** parameters. As a representative lines for blue-shelled eggs, two hen lines were chosen, one of which was cross of White Leghorn and blue-shelled native breed carrying the *O/o* genotype and the native breed being homozygous *O/O*. They observed different values among parameters measured between those two genotypes, which could be explained by the different genotype and cumulative effect of genes. However, in our study, individuals with genotype *O/O* and *O/o*, showed a partial overlap and based on

measurements of photometric parameters alone, these genotypes cannot be distinguished. Results observed by Wang et al. (2023) during evaluation of *L*a*b** parameters and its correlation with pigment content in blue-green egg subtypes, support our findings. Individual photometric parameters (reflectance and color measurement) do not have the same weight (Fig. 2). A detailed assessment of individual photometric parameters and their ability to distinguish groups of genotypes is given in Fig. 3. It appears that clear separation of recessive homozygotes (Group A) is only possible when evaluating *a** parameter and regardless of the measurement position on the egg. The *a** parameter allows to distinguish the shell color (white and blue) as was also observed by Li et al. (2006) and recently by (Chen and Wang, 2022). This result also corresponds to the PCA analysis results reported in Fig. 2. In theory, therefore, the value of *a** parameter that provided the greatest variability could be used to determine the genotype by measuring the shell color. In contrast, the values *L** and *b** proved to be of little evidence in the color assessment, also concur with the observation of Li et al. (2006). Similarly, Wang et al. (2024), evaluated the *a** parameter as strongly associated with the *SLCO1B3* functional gene variant, affecting the amount of pigment deposited in the blue eggs. The difference in pigment saturation between the eggs of individual laying hens was not sufficiently significant and the values in the PCA analysis appeared to be relatively stable (Fig. 2). This finding may be explained by the effect of hen-to-hen variability, which will have a greater influence on the variation in shade intensity than the order of the egg laid (Zeng et al., 2022).

The second hypothesis tested the level of success of the diagnostic PCR test and possibility of its introduction into routine testing of breeding population. A causal mutation for blue egg color was published by Wang et al. (2013). Wragg et al. (2013) published a study focused on the sequencing analysis of causative mutation – ~4.2 kb retroviral insertion in the *SLCO1B3* gene in the Mapuche fowl and in European breeds of hens. His study revealed a 100% sequential coincidence of both inserts. In addition, they found that EAV-HP insertion sites are different in Asian, South American, and European “oocyan” breeds of hens. DOMINANT Greenshell line was bred on the basis of private breeding know-how using unique donor genotypes. Although the current literature does not mention this, there was a distinct possibility that

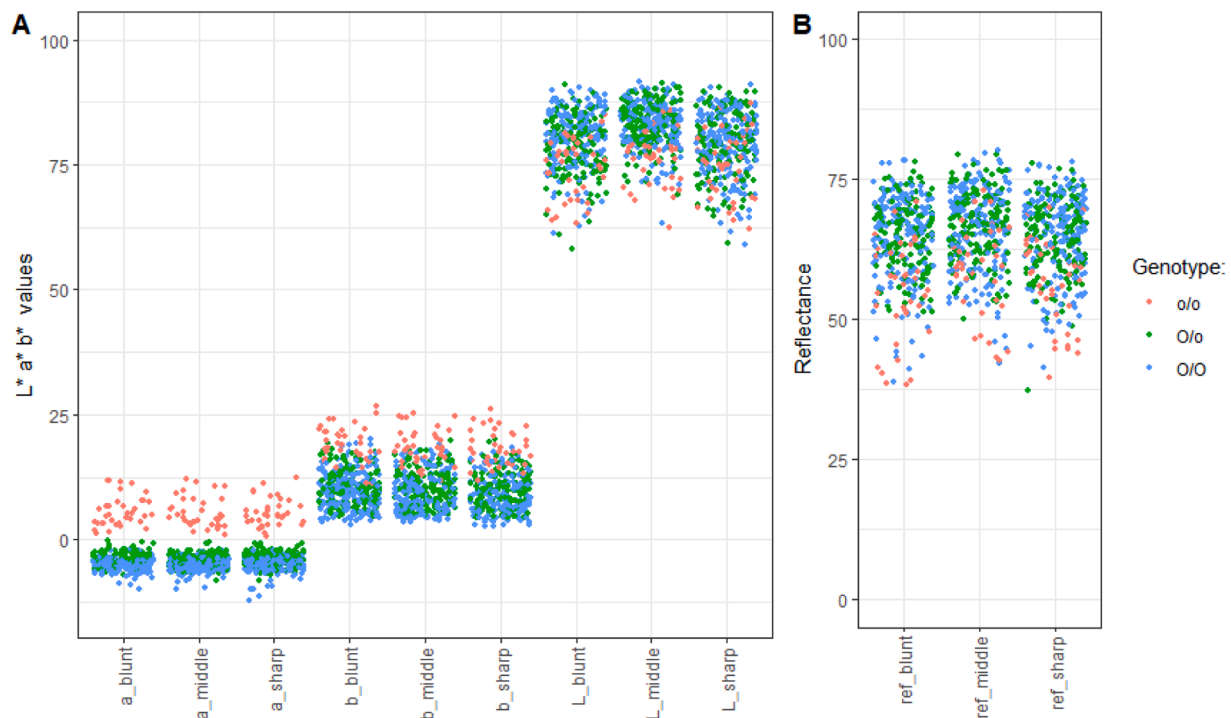


Figure 3. Graph showing complete separation of the *o/o* genotype (marked in red) from the *O/o* and *O/O* genotypes based on the analysis of photometric data for parameter *a**. Fig. 3A shows the separation of genotypes based on the measured individual *L*a*b** values and Fig. 3B shows the separation of genotypes based on reflectance values.

published PCR diagnostic tests would not be 100 percent effective in this specific population. To verify the reliability of the diagnostic test, we tested various input samples (blood and feathers), and we used two DNA isolation methods and two protocols for molecular analysis and identification of retroviral insertion - detection of *O* and *o* alleles. Primers previously published by Wang et al. (2013) and Wragg et al. (2013) were used. To obtain reliable PCR analysis results the conditions of PCR analysis have been modified. The genotype of all individuals was successfully determined by PCR analysis. Regardless of the origin of the input DNA sample, both DNA extraction methods gave identical results (see Table S1 for detailed results). PCR products obtained after analyses of DNA extracted from feathers appeared brighter, with no non-specific products.

The third hypothesis posed the question of whether it is possible to combine phenotyping and genotyping in a suitable way, i.e. optometric measurement of eggs and analysis of the presence of the EAV-HP insertion in the *SLCO1B3* gene. From the results and conclusions given in the first hypothesis, it is evident that the optometric method was able to reliably distinguish only dominant and recessive phenotypes and eliminate only undesirable recessive homozygotes with a white colored eggshell from the population. Even in the case of using the CART methodology and classification trees, it is not possible to reliably distinguish between dominantly homozygous and heterozygous individuals, and the error rate is relatively high. In the group of desirable *O/O* homozygotes, 31% of heterozygotes, carriers of an undesirable recessive allele, was present. For a fast, reliable and cost-effective analysis of the breeding population, a technology based on a combination of a fast optometric method and the measurement of only the key *a** parameter (or only the *a*.blunt value) can be proposed for the elimination of recessive homozygotes. Genotyping and precise determination and selection of dominant homozygotes can then be performed in the remaining population. When screening a large breeding population, feathers collection as an input material for subsequent PCR analyses appears to be more convenient choice to accelerate the selection and also to prevent the animals from increased manipulation, compared to

blood collection. Combination of optometric and molecular selection methods then leads to more efficient selection, reduction of overall selection costs, and stabilization of the breeding population within one generation and the achievement of a pure homozygous line with regard to eggshell color.

Breeding of laying hens for eggshell color was based on methods of subjective shade and color evaluation for about 60 years. For this subjective evaluation own color scales or standardized color palettes/fans (color fan grading) were used (Zeng et al., 2022). Likewise, most of the studies investigating shell color were based on a subjective comparison (Wei and Bitgood, 1990). Since the invention and introduction of CIE-LAB system devices (from roughly the second half of the 1970s in leaders in the poultry sector) into standard breeding practices, measurements have been substantially refined and digitized (Cavero et al., 2012) and became an auxiliary means in breeding programs.

Based on the results of our experiment, in the case of selection/elimination of a recessive allele at the *O* locus, one cannot solely rely only on photometric measurements, because there is a certain intermingling of the measured values of heterozygous and dominant homozygous individuals for this trait. Despite the relatively low cost of PCR analyses, photometric analyses using a simple portable device are less complicated. But only the end-point PCR is reliable for accurate and precise genotype detection. However, a combination of the two methods can be used, where individuals carrying the recessive homozygous genotype *o/o* are selected in first step (simple and fast photometric analysis) and the more expensive and time-consuming PCR test is then applied to a remaining part of the population. In addition, results from analyses of photometric measurements show that this process can be further simplified and rationalized to only a small number of parameters. In case of this study even one photometric parameter in one egg point can be measured, which is consistent with the results of Li et al. (2006). This will allow very rapid assessment and primary screening to eliminate recessive homozygotes. Choosing this combined approach will reduce the negative effect of both methods - the higher inaccuracies of photometric measurement and time losses and financial burden in the

case of PCR. Laboratory analysis also requires collaboration with an external specialized laboratory and increased handling of animals during blood collection. For breeders, the combination of the two methods mentioned above represents not only a considerable timesaving, but also a financial saving.

CONCLUSION

In the previously described approach, we proposed to accelerate and streamline the selection of desired dominantly homozygous individuals by combining photometric measurement of eggshell color and molecular selection. The selection process has thus been significantly shortened by several generations compared to conventional breeding methods, and for the breeder it represents an acceleration of the process by several years. Molecular selection, or the combination of photometric and molecular selection has not yet been used within the breeding company and breeding program of laying hens DOMINANT CZ. In laying hens, 19.38% of heterozygotes with an undesirable recessive allele were detected in the group of dominant phenotypes determined by optometric evaluation. A total of 17.98% of unwanted individuals, recessive homozygotes and heterozygotes were detected in the group of males. In this case, the advantage of the PCR approach was shown in the possibility of applying the method to males, whose genetic structure/nature is determined only by classical approaches during sexual maturity, or to the tests of their sisters. The elimination of undesirable genotypes, carriers of the recessive allele, from the breeding population enabled a rapid stabilization of the shell color in this breeding line. According to the breeding company's internal testing, no recessive allele was observed in subsequent generations of the DOMINANT Greenshell line. Thanks to early detection, this trait was fixed in a clean line guaranteeing the laying of only eggs with a blue-green shell color.

Declaration of competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

ACKNOWLEDGEMENTS

This research was funded by the Grant Agency of the University of South Bohemia in České Budějovice, grant number GAJU 080/2022/Z. Special thanks to Dominat Genetika Ltd., Voleč, Czech Republic, for providing the material for analyses and cooperation.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.psj.2024.104425](https://doi.org/10.1016/j.psj.2024.104425).

REFERENCES

- Bates, D., Maechler, M., Bolker, B., Walker, S., 2015. Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Soft.* 67, 1–48.
- Bi, H., Liu, Z., Sun, C., Li, G., Wu, G., Shi, F., Liu, A., Yang, N., 2018. Brown eggshell fading with layer ageing: dynamic change in the content of protoporphyrin IX. *Poult. Sci.* 97, 1948–1953.
- Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 2017. Classification and regression trees. In *Routledge eBooks*.
- Cavero, D., Schmutz, M., Icken, W., Preisinger, R., 2012. Attractive eggshell color as a breeding goal. *Lohmann Inform* 47, 16–21.
- Chen, J., Dalirfatsat, S.B., Han, D., Dong, X., Hua, G., Zheng, X., Xia, T., Shao, T., Deng, X., Wu, C., 2020. An EAV-HP insertion in the 5' flanking region of *SLCO1B3* is associated with its tissue-expression profile in blue-eggshell Yimeng chickens (*Gallus gallus*). *Poult. Sci.* 99, 6371–6377.
- Chen, Q., Wang, Z., 2022. A new molecular mechanism supports that blue-greenish egg color evolved independently across chicken breeds. *Poult. Sci.* 101, 102223.
- Dalirfatsat, S.B., Dong, X., Deng, X., 2015. Molecular phylogenetic analysis of Chinese indigenous, blue-shelled chickens inferred from whole genomic region of the *SLCO1B3* gene. *Poult. Sci.* 94, 1776–1786.
- Darwish, H.Y.A., Dalirfatsat, S.B., Dong, X., Hua, G., Chen, J., Zhang, Y., Li, J., Xu, J., Li, J., Deng, X., Wu, C., 2019. Genome-wide association study and a post replication analysis revealed a promising genomic region and candidate genes for chicken eggshell blueness. *PLoS One* 14, e0209181.
- Doyle, J., 1991. DNA protocols for plants. Springer eBooks, pp. 283–293.
- Fairchild, M.D., 2013. Color appearance models. Pages 273–286 in *The CIE Color Appearance Model* (1997). Copyright © John Wiley & Sons, Ltd.
- Farnsworth, G.M., Nordskog, A.W., 1955. Breeding for Egg Quality1: 3. Genetic Differences in Shell Characteristics and Other Egg Quality Factors. *Poult. Sci.* 34, 16–26.
- Godfrey, G.F., 1949. Shell Color as a Measure of Egg Shell Quality1. *Poult. Sci.* 28, 150–151.
- Guo, J., Wang, K., Qu, L., Dou, T., Ma, M., Shen, M., Hu, Y., 2020. Genetic evaluation of eggshell color based on additive and dominance models in laying hens. *Asian-Australasian Journal of Animal Sciences* 33, 1217–1223.
- Hargitai, R., Boross, N., Nyiri, Z., Eke, Z., 2018. Effects of food limitation on the intensity of blue-green and brown eggshell coloration: an experimental study with the canary. *Journal of Avian Biol* 49, e01486.
- Hunton, P., 1962. Genetics of eggshell color in a light Sussex flock. *Br. Poult. Sci.* 3, 189–193.
- Jolliffe, I., 2002. Principal component analysis. Springer Science & Business Media.
- Li, X.Y., Xu, G.Y., Hou, Z.C., Zhao, R., Yang, N., 2006. Variation of eggshell color in different egg-type chickens. *Arch. Geflügelk* 70, 278–282.
- Liao, B., Liu, Y., Cao, J., Ning, Z., 2021. Analysis and research on pigment precipitation of eggshell based on green environment-friendly breeding technology. *IOP Conf. Series. Earth and Environmental Sci.* 632, 052046.
- Lukanov, H., 2014. Agricultural Science and Technology 6 (1), 3–10.
- Lukanov, H., Genchev, A., Pavlov, A., 2015. Color traits of chicken eggs with different eggshell pigmentation. *Trakia Journal of Sciences. Series Biomedical Sci.* 13, 149–158.
- Punnett, R.C., 1933. Genetic study in poultry—IX. The blue egg. *Genetics* 27, 465–470.
- R Core Team, 2024. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Accessed May 2024. <https://www.R-project.org/>.
- Samiullah, S., Roberts, J., Chousalkar, K., 2015. Eggshell color in brown-egg laying hens — a review. *Poult. Sci.* 94, 2566–2575.
- Scott, T., Silversides, F., 2000. The effect of storage and strain of hen on egg quality. *Poult. Sci.* 79, 1725–1729.
- Wang, H., Ge, Y., Zhang, L., Wei, Y., Li, Q., Zhang, X., Pan, Y., 2023. The pigments in eggshell with different color and the pigment regulatory gene expression in corresponding chicken's shell gland. *Animal* 17, 100776.
- Wang, Z., Qu, L., Yao, J., Yang, X., Li, G., Zhang, Y., Li, J., Wang, X., Bai, J., Xu, G., Deng, X., Yang, N., Wu, C., 2013. An EAV-HP insertion in 5' flanking region of *SLCO1B3* causes blue eggshell in the chicken. *PLOS Genet* 9, e1003183.
- Wang, Z., Chen, Q., Wang, Y., Wang, Y., Liu, R., 2024. Refine localizations of functional variants affecting eggshell color of Lueyang black-boned chicken in the *SLCO1B3*. *Poult. Sci.* 103, 103212.
- Wei, R., Bitgood, J.J., 1990. A new objective measurement of eggshell color. *Poult. Sci.* 69, 1775–1780.
- Wragg, D., Mwacharo, J.M., Alcalde, J.A., Wang, C., Han, J., Gongora, J., Gourichon, D., Tixier-Boichard, M., Hanotte, O., 2013. Endogenous retrovirus EAV-HP linked to blue egg phenotype in Mapuche fowl. *PLoS One* 8, e71393.
- Zeng, L., Xu, G., Jiang, C., Li, J., Zheng, J., 2022. Research Note: L*a*b* color space for prediction of eggshell pigment content in differently colored eggs. *Poult. Sci.* 101, 101942.
- Zhao, R., Xu, G., Liu, Z., Li, J., Yang, N., 2006. A study on eggshell pigmentation: biliverdin in blue-shelled chickens. *Poult. Sci.* 85, 546–549.

Microsatellite Markers: A Tool to Assess the Genetic Diversity of Yellow Mustard (*Sinapis alba* L.)

Eva Jozová ^{1,*} , Michael Rost ¹ , Andrea Rychlá ², Dagmar Stehlíková ¹ , Baveesh Pudhuvai ¹, Ondřej Hejna ¹, Pavel Beran ¹ , Vladislav Čurn ¹  and Miroslav Klíma ³

¹ Department of Genetics and Biotechnology, Faculty of Agriculture and Technology, University of South Bohemia in Ceske Budejovice, 370 05 Ceske Budejovice, Czech Republic; rost@fzt.jcu.cz (M.R.); stehld@fzt.jcu.cz (D.S.); bpodhuvai@fzt.jcu.cz (B.P.); hejna@fzt.jcu.cz (O.H.); beranp@fzt.jcu.cz (P.B.); curn@fzt.jcu.cz (V.Č.)

² OSEVA Development and Research, 746 01 Opava, Czech Republic; rychla@oseva.cz

³ Crop Research Institute, 161 06 Prague, Czech Republic; klima@vurv.cz

* Correspondence: jozova@fzt.jcu.cz

Abstract: Microsatellite markers were used for the assessment of genetic diversity and genetic structure in a germplasm collection of yellow mustard, *Sinapis alba* L. The comprehensive collection of genetic resources represented 187 registered varieties, landraces, and breeding materials. Microsatellites generated 44 polymorphic alleles in 15 loci. Eleven of them were medium to highly polymorphic, and the high levels of observed heterozygosity (0.12–0.83) and Nei's gene diversity index (0.11–0.68) indicated a high level of polymorphism. Based on PCoA and neighbor joining analyses, the genetic resources were divided into two groups. The range of genetic dissimilarity in the analysed collection was in the range of 0.00–1.00. The high level of dissimilarity between the accessions was documented by the high WAM value (33.82%). Bayesian clustering algorithms were performed in the STRUCTURE 2.3.4 software. The number of clusters was estimated at K = 2. The accessions were classified according to Q1/Q2 values. The low average values of the parameters Fst_1 (0.3482), Fst_2 (0.1916), and parameter alpha (0.0602) indicated substantial mating barriers between varieties and reproductive isolation due to the limited exchange of genetic resources between breeders. These results demonstrated the importance of extensive collections of genetic resources for the maintenance of genetic diversity and indicated considerable genetic differentiation among accessions.

Keywords: molecular markers; SSR; breeding; genetic structure; induced mutagenesis



Citation: Jozová, E.; Rost, M.; Rychlá, A.; Stehlíková, D.; Pudhuvai, B.; Hejna, O.; Beran, P.; Čurn, V.; Klíma, M. Microsatellite Markers: A Tool to Assess the Genetic Diversity of Yellow Mustard (*Sinapis alba* L.). *Plants* **2023**, *12*, 4026. <https://doi.org/10.3390/plants12234026>

Academic Editor: Shri Mohan Jain

Received: 19 October 2023

Revised: 23 November 2023

Accepted: 28 November 2023

Published: 29 November 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Yellow mustard (*Sinapis alba* L.) is a vital crop for Czech agriculture's food and condiment sectors. It originated in the Middle East and Mediterranean regions. Still, this crop has been produced worldwide for millennia as a condiment crop. It is extensively planted as a prominent speciality crop [1], and the seeds and volatile oils are used in traditional Chinese medicine [2]. This crop is drought-tolerant and resistant to major diseases and insect pests [3,4]. Yellow mustard also has phytosanitary properties and can potentially reduce harmful nematodes in soils, and the soil disinfestation effects of biofumigant compounds released from yellow mustard green manure have been reported [5]. Complete data from the literature and original data on *S. alba* exposed to elements such as metals from the thallium, arsenic, and platinum groups were collected, ensuring crop application for phytoremediation in contaminated areas [6]. The Czech Republic is at the forefront of this crop production, due to ideal climatic conditions and a long heritage in cultivation. Together with Canada and Ukraine, the Czech Republic is one of the world's largest producers and exporters of this crop. Yellow mustard is cultivated mainly for seeds and non-production purposes, such as an ameliorating catch crop for green manuring or as a significant component of biobelts. Mustard is also an important bee nectar and pollen-producing plant [7].

The range of genetic diversity, changes in genetic structure over the timespan of breeding in landraces, cultivated varieties, and genetic resources are still little understood, even though yellow mustard, like rapeseed, is a vital crop [4,8]. There is a widespread perception that the genetic variability is declining, due to breeders' emphasis on production and product/seed quality, compared to other intensively cultivated crops [9]. Morphological traits, previously the only tool for selection and cultivar identification, are no longer sufficient to identify emerging varieties and define their phenotypic variance, due to the influences of many environmental factors. This is also reflected in the often problematic use of DUS (Distinctness, Uniformity and Stability) tests for the description and legal protection of varieties [10].

The unremitting reduction in the number of cultivated species and varieties and their widespread loss has led to genetic erosion, the partial or total loss of landrace varieties and a narrowing of genetic diversity. The effects of a lack of genetic diversity can be seen in the loss of important genes and genetic vulnerability, as well as in the reduction in genetic gain in quantitative traits, which is difficult to overcome [9–11]. The breeders then have then sought to expand existing genetic diversity, e.g., by using landrace varieties, by distant hybridisation with taxonomically more distant species, and by applying in vitro techniques and targeted mutagenesis [12]. In practical breeding, the use of classical mutagenesis is limited, and only a few new varieties have been bred based on this approach. Because of the difficult detection of mutations and the not-quite-successful use of induced mutagenesis in plant breeding, a molecular or reverse genetics approach has been adopted since the beginning of the millennium through the TILLING (targeting induced local lesions in genome) method [13]. A more modern mutant detection procedure involves amplifying genes of interest, sequencing amplicons, and identifying point mutations [14].

Plant genetic resources should be characterised by employing molecular markers to reveal genotypic variations and unambiguously identify crop varieties. Molecular methods have been essential components of most research on genetic diversity over the last few decades, and molecular markers recognise DNA polymorphism between varieties and aid in gaining a better understanding of the genetic hierarchy of the parent components [15]. Among the most used types of markers to assess genetic diversity or seed purity are, for example, RAPDs (Random Amplified Polymorphic DNA) [16], which are no longer widely used due to poor repeatability of results, AFLP (Amplified Fragment Length Polymorphism) [17], ISSR (Inter Simple Sequence Repeats) [18], SSR (Simple Sequence Repeats) [4,19,20], or SNP (Single Nucleotide Polymorphism) [21]. However, it is essential to recognise that different markers have distinct attributes/properties that reflect various components of genetic variation and, therefore, can produce different results [22]. Although yellow mustard is an important crop, the number of studies devoted to this species is quite limited. A detailed study of yellow mustard accessions, maintained at Plant Gene Resources of Canada, was performed using AFLP markers [17]. The same collection of genetic resources was recently analysed using a modern approach to genotyping by sequencing, and new sets of SNPs were discovered and evaluated [21]. More attention was focused on coloured mustard, *B. juncea*, *B. carinata*, and *B. nigra*. An SSR marker was used in *B. juncea* [23] and *B. nigra* [24]. In *B. juncea*, the influence of natural outcrossing and cross-hybridization for the development of analysed varieties was demonstrated, based on the analysis of the genetic structure. Unique SSR alleles were used for specific DNA fingerprinting for the identification and legal protection of *B. juncea* varieties. Genome-specific SSR markers in *B. nigra* were used for assessment of the correct taxonomic classification of accessions. Based on multidimensional scaling analyses, genotypes were associated with country or region of origin. In *B. carinata* [25], the genetic diversity and population structure of genotypes of different origins was analysed using high-throughput SNP markers. The analysed genotypes were divided into two clusters that formed landraces/ecotypes, respectively breeding lines and varieties. This indicated a change in genetic structure due to breeding. The most comprehensive marker system used for the assessment of genetic diversity are microsatellites (SSRs, simple sequence repeats), as they have many desirable

properties, such as abundant genome distribution, the ability to form multiplexes, and codominance, i.e., the ability to distinguish between homozygotes and heterozygotes [26]. Molecular approaches improve germplasm conservation efforts and increase the use of plant genetic resources [27]. Furthermore, combining morphological and molecular markers is an excellent approach for selecting optimal genotypes for breeding, while maintaining adequate genetic diversity [28], and knowledge of genetic diversity is the basis for the effective management and utilisation of plant germplasms.

This study aimed to investigate the genetic structure of the yellow mustard core collection and to evaluate the genetic diversity among the yellow mustard accessions. A thorough investigation of the molecular properties of this diversity yields a group of the most diverse genotypes for use in breeding programmes. Marker-assisted selection using polymorphic markers among parents helps with the finding and choosing of suitable parents for hybridisation and backcross breeding for future use.

2. Results and Discussion

2.1. Microsatellite Polymorphism and Genetic Diversity of Genetic Resources

The wide collection of genetic resources for yellow mustard (*S. alba*) was represented by 198 unique accessions (registered varieties, landraces, and breeding materials). Microsatellites generated 113 alleles and, of the total alleles, 44 were polymorphic and detected in 15 loci, with an average of 2.93 alleles per locus and a maximum of six alleles in D3_A, P7, and P35_A loci. Five loci (BoREM1b_B, BoREM1b_C, BolAB19TF, P30_C, and P35_B) were classified as uninformative loci with MAF < 0.01. Only one allele was detected in these loci of one to three individuals/accessions. BoREM1b_B was successfully amplified in genotype 9 (Morocka, landrace) and BoREM1b_C in genotypes 51 (SIN 10/79, breeding material), 167, and 177 (Asta and Raduga, new registered varieties). BolAB19TF was amplified in genotype 130 (Sabon, new registered variety), P30_C in genotypes 169 and 186 (Rumba and Olga, new registered variety), and P35_B in genotype 174 (Bamberka, new registered variety). This study is part of a project focused on yellow and coloured mustard, and only the markers successfully amplified in yellow mustard are presented in Table 1. Samples with the occurrence of uninformative loci were also included in these results, because these genotypes with the presence of even this one allele differ from the set of other varieties. These unique/less common alleles are typical for a particular variety and lead to higher values of genetic diversity (see Figure S1 and WAM value). It is also noticeable that these less frequent alleles, with the exception of the Marocka landrace, occur in newly bred varieties and breeding materials. The size of the amplified alleles was in the range of 125–314 bp. Eleven loci were medium to highly polymorphic, with PIC values varying from 0.19 to 0.65. The highest PIC values were observed in loci D3_A (0.65) and P35_A (0.52), and also in loci BoPC34 (0.51) and P381 (0.50), characterised by a low number of alleles per locus. The level of observed heterozygosity was in the range of 0.12 to 0.83, with an average of 0.21. The highest observed heterozygosity was also recorded in loci with a high polymorphic index, namely D3_A (0.83), BoP34 (0.73), and P35_A (0.53). Nei's gene diversity index ranged between 0.11 and 0.68, with an average of 0.25. Such a high level of polymorphism, compared to other oil seed Brassicas, may be explained by a higher level of cross-pollination in yellow mustard [29]. Detailed results of the microsatellite analysis are given in Table 1.

Table 1. Detailed results of microsatellite analyses.

Microsatellite Marker	Number of Individuals	PIC	Na (Number of Alleles)	Number of Alleles in Population	Length of Alleles in Bp	Median and Max. Frequency of Alleles	Ho (Observed Heterozygosity)	He (Nei, 1973; Expected Heterozygosity)	Fis (Inbreeding Coefficient)	1-D (Simpson Index)	Hexp (Nei, 1978; Gene Diversity)	Evenness
BoREM1b_A	198	0.19	2	209	173–175	104.5; 187	0.12	0.11	−0.06	0.11	0.11	0.50
BoREM1b_B	198	0.00	1	3	181	3; 3	0	0	-	-	-	-
BoREM1b_C	198	0.00	1	3	187	3; 3	0	0	-	-	-	-
BoABI91TF	198	0.00	1	1	314	1; 1	0	0	-	-	-	-
BoPC34	198	0.51	3	323	149–151	148; 172	0.73	0.50	−0.45	0.50	0.50	0.94
P381	198	0.50	2	21	213–229	105; 11	0.05	0.50	0.90	0.50	0.51	1.00
D3_A	198	0.65	6	342	155–163	33.5; 149	0.83	0.67	−0.24	0.67	0.67	0.85
D3_B	198	0.44	5	101	167–171	10; 73	0.20	0.39	0.50	0.38	0.68	0.54
P7	198	0.27	6	119	141–156	3.5; 101	0.11	0.21	0.49	0.21	0.21	0.39
P9_B	198	0.38	3	184	125–138	32; 141	0.28	0.29	0.05	0.27	0.27	0.54
P9_C	198	0.35	4	231	145–160	25; 184	0.25	0.23	−0.06	0.23	0.24	0.49
P30_B	198	0.32	2	5	166–174	2.5; 4	0	0.32	1.00	0.32	0.36	0.72
P30_C	198	0.00	1	2	192	2; 2	0	0	-	-	-	-
P35_A	198	0.52	6	109	149–161	3.5; 68	0.53	0.47	−0.12	0.46	0.47	0.63
P35_B	198	0.00	1	1	173	1; 1	0	0	-	-	-	-

Allele sizes calculated from CE electropherograms were identical for the internal controls and stable in all repetitions. This uniformity and reproducibility for all microsatellites were indicators of correctly set protocols of molecular analyses. The microsatellites used in this study were originally designed for *Brassica oleracea* [30], *B. rapa* ssp. *chinensis*, *B. juncea*, and *B. nigra* genetic resources [31]. Our results showed very good cross-transferability and possible utilisation of these markers in yellow mustard. Successful amplification of specific fragments from the more distantly related *Brassica* species, including *S. alba* using primers designed for *B. napus*, were also reported [19]. Still, they did not analyse the diversity of microsatellites in yellow mustard genetic resources in more detail. Therefore, this study presents the first results of the analysis of microsatellites in a large set of genetic sources of yellow mustard and shows relatively high variability in the genetic resources, as well as the cross-transferability of markers and the occurrence of unique alleles.

The results of the AMOVA (Analysis of molecular variance) showed that the highest proportion of diversity was found within the analysed set of genetic resources (87.13%) and that the proportion of diversity and differentiation between types of genetic resources (4.6%) or between genotypes within types of genetic resources (8.27%), although significant, was always much lower (Table 2). Likewise, the null hypotheses were rejected (for parameters Φ_{CT} , Φ_{SC} — $p < 0.001$, and Φ_{ST} — $p < 0.005$). Thus, there are genetic differences between types of genetic resources, as well as between individual accessions of yellow mustard within the analysed set of genetic resources.

Table 2. Analysis of molecular variance (AMOVA) results for yellow mustard genetic resources (between and within types of genetic resource) with df, sums of squares (SS), mean squares (MS), genetic variance estimates (Sigma), proportion of variance (%) attributed to the different levels in the spatial hierarchy, Phi (Φ) statistics that are analogous to Wright's F-statistic, and p values (9999 permutations).

	df	SS	MS	Genetic Variance Estimates (Sigma)	Proportion of Variance (%)	Phi (Φ)	p -Values
Between types of genetic resources	2	19.41	9.705	0.143	4.602	$\Phi_{CT} = 0.0460$	<0.0001
Between genotypes within types of genetic resources	31	123.07	3.970	0.257	8.272	$\Phi_{SC} = 0.0867$	<0.0001
Within genetic resources	164	414.12	2.707	2.707	87.126	$\Phi_{ST} = 0.1287$	0.0028
Total	197	556.60	2.992	3.107	100		

Type of genetic resource = landrace/variety/breeding material.

The genetic diversity of the yellow mustard germplasms was calculated using DARwin 6 software. The dendrogram in Figure 1a shows the results of the weighted neighbor joining analysis, and, in Figure 1b, the results of the principal coordinates analysis. The extent of genetic variation in the analysed collection was in the range of 0.00–1.00 (Figure S1). The furthest distances were between accessions ‘47-635-YE 13’ and ‘Braco’. Additionally, the seven other most different/distant pairs of accessions always included the Chinese ‘47-635-YE 13’ genotype. Both the genetic diversity and dissimilarity range were much higher than the values reported by Fu et al. [17], given the more considerable extent of the analysed collection and the marker system used. The high level of dissimilarity between the accessions analysed was documented by the WAM value, which was 33.82%. A higher WAM value corresponded with the pattern of genotype distribution after factorial (PCoA) analysis (Figure 1b). In our previous study, focused on *Brassica napus* [32], we observed WAM values that were significantly lower (22.50% for SSR markers). These results showed the importance of extensive collections of genetic resources for preserving genetic diversity. The results of other studies conducted on mustard showed significantly lower values of genetic diversity [33], probably due to the low diversity of the collection and the small number of samples, which was incomparably lower than in this study.

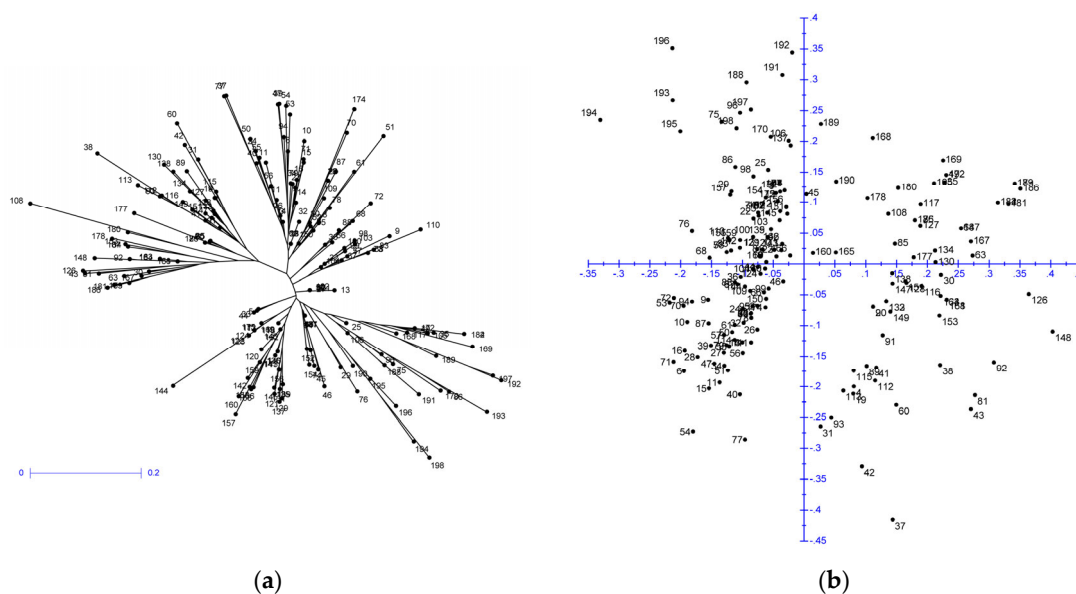


Figure 1. Outputs of statistical analyses of SSR markers in 198 *S. alba* accessions: (a) results of cluster analysis (weighted N J analysis); (b) results of principal coordinates analysis (PCoA).

Based on microsatellite analysis, it was possible to identify a large proportion of the analysed accessions unambiguously, but even when a relatively high number of markers were evaluated, several accessions of different origins had precisely the same patterns of microsatellite markers. There were two main groups of accessions with the same pattern of microsatellites. The first group included genotypes from Germany, Denmark, Netherlands, and Poland, but also genotypes from Australia and an old landrace from the Czech Republic/former Czechoslovakia (‘ABA’, ‘Borowska’, ‘BRSCHW 55705’, ‘Česká krajová’, ‘Perovska’, ‘Sito’, and ‘King’). The second block of genotypes with the same pattern of microsatellites mainly included Czech varieties (‘Andromeda’ and ‘Severka’), as well as varieties from Germany, Austria, and the Netherlands (‘Ascot’, ‘Emergo’, and ‘Mirly’). There were also other groups or pairs of indistinguishable accessions, such as ‘OP-2408’ and ‘OP-2415’, or ‘Olga’ and ‘OP-2291’, originating in one breeding station, but also cases have been reported where landraces have the same microsatellite profile as modern varieties, e.g., ‘Přerovská bílá’ vs. ‘Ludique’ and ‘Budakalászi sárga’, or ‘Dr. Schneider’ vs. ‘Concerta’ and ‘Esprit’. A relatively narrow genetic base should explain these results, and unrelated genotypes with an identical pattern of microsatellites are used in different breed-

ing programmes. These findings also reveal the need for better management of breeding populations based on genetic analysis, and further improvement is also increasingly dependent on the evaluation of genetic diversity resources and breeding populations. Breeding materials from the TILLING population showed significantly higher variability and distinct differences from other genetic resources. These materials were created precisely with the goal of extending genetic variability and their further utilization in breeding programmes (accessions 188–198 in the lower left part of the dendrogram in Figure 1a). Fu et al. [17] reported a limited range of genetic resources used in mustard breeding, and the importance of germplasm protection and narrowing genetic variation has been discussed for other cruciferous oilseeds, including mustards [8,33,34]. For this reason, it is appropriate to use molecular markers in selection programmes. Genetic resources (paternal components used for crossing) should be selected not only according to their phenotypic values but also in terms of genetic diversity or distance, and microsatellites can be used as effective practical tools to provide molecular data and evaluate the genetic relationship [35–37].

2.2. Genetic Structure of Yellow Mustard Varieties

Bayesian clustering algorithms were performed in the STRUCTURE 2.3.4 software to describe the genetic structure better and to explain the results from previous statistical analyses. Although this approach has been highly frequented in ecological studies [38] in the last period, it is also used for agricultural crops including oil seed brassicas, such as *B. napus* [32], *B. juncea* [23], and *B. carinata* [25]. Through the STRUCTURE software, we determined the likelihood of the data for several different K, with K ranging from one to seven, to eventually see the plateau pattern in the diagnostic plot. During these computations, we set the burnin period to 100,000, and 150,000 MCMC repeats after burnin. Such a setting seemed to be sufficient, as we attained good mixing and a stable process in equilibrium, with no excessive variations in the $\ln(\alpha)$ parameter estimates. The parameter alpha (the quantification of relative levels of admixture between populations) was allowed to vary for each population to obtain a more accurate estimate of the ancestry. The starting value for the Dirichlet parameter alpha was set to one with a uniform prior. We applied the model with admixture and correlated allele frequencies to attain greater power, in order to detect potential distinct populations that could be closely related. The distributional parameter lambda was computed in the initial pre-run (with lambda set to one), and the inferred value was used in sequential runs (the quantification of independence between markers in terms of allelic frequency distribution). Other hyperparameters were left at their default value. We did not use the LOCPRIOR parameter in our simulations. Consequently, for optimal K = two, we estimated the individual Q-matrix [38–40]. The detailed outputs of cluster number estimations are given in Table S2 and Figure 2. The results and outputs of the STRUCTURE software (Q1/Q2 values) are given in Figure 3 and Table S1 (Supplementary Data).

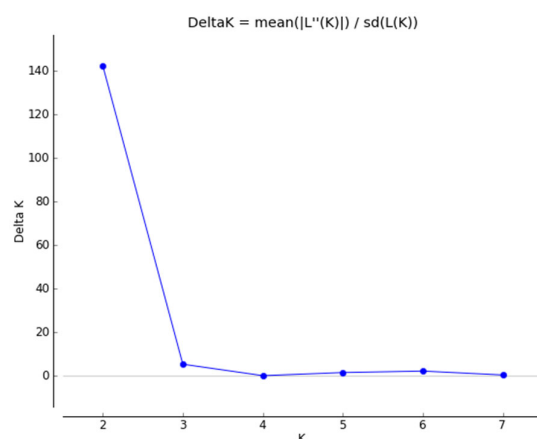


Figure 2. The magnitude of Delta K as a function of K.

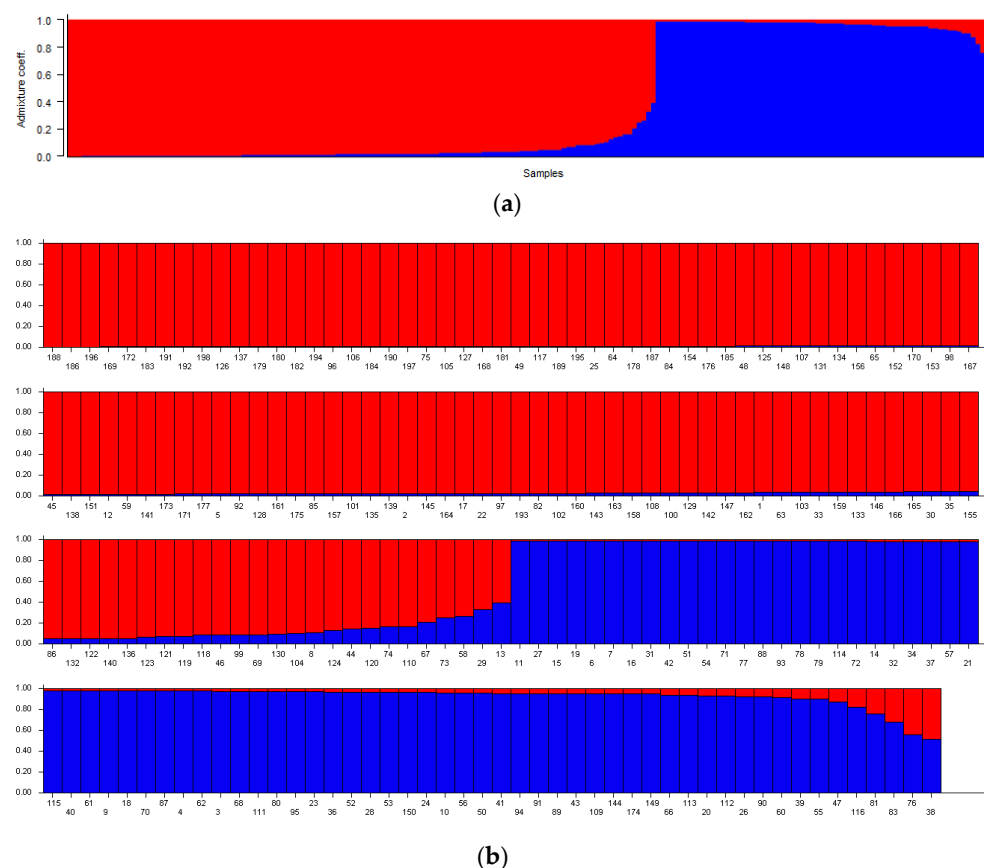


Figure 3. Detailed classification of accessions according to Q1 (blue)/Q2 (red) values: overall overview of the distribution of Q1/Q2 values in the analysed set of genetic resources (a); detailed analysis of genetic structure with marking of individual accessions (b).

The accessions were classified into particular groups, according to Q1/Q2 values (Figure 3a). The mean alpha value was estimated at 0.0602, indicating substantial mating barriers between breeding populations (varieties) and reproductive isolation, due to the limited exchange of genetic resources between individual breeders. The mean values of F_{st_1} and F_{st_2} were 0.3482 and 0.1916, respectively. Also, these values indicated considerable genetic differentiation among the accessions, as was reported previously [41].

The detailed analysis of the raw data and outputs from the STRUCTURE software indicated the presence of a transition zone between the two main clusters estimated, according to parameter Delta K = two. Therefore, detailed classification of accessions according to Q1/Q2 values (Figure 3b) led to the formation of three blocks: a/the first block included accessions with Q1 values in the range of 0.008–0.095; b/the second block formed a transition zone with Q1 values in the range of 0.107–0.396; c/the third block included accessions with Q1 values in the range of 0.515–0.991. Most accessions (113; 57.1%) were assigned to the first block; the third block comprised 73 accessions (36.8%) and 12 accessions (6.1%) formed the transition group. In the first most diverse groups, there were both earlier varieties and landraces, as well as modern varieties and breeding materials. In the second and third groups were landraces and bred varieties, but not new breeding materials. These results from the analysis of the genetic structure were supported by the outputs from the AMOVA, where, although genetic differentiation between types of genetic resources and between individual accessions of yellow mustard within the collection of genetic resources was statistically significant, the proportion of variance between types of genetic resources (landrace, variety, and breeding material) was very low (4.6%). On the contrary, the proportion of variance between genotypes within the analysed set of genetic resources was the highest (87.13%). A more detailed analysis of genetic structure results shows

somewhat more interesting outputs than in the case of the NJ analyses. It was evident that varieties originating from some countries or breeding programmes have similar or the same genetic structures. However, it is more common that, according to the genetic structure, the varieties from one country of origin are classified into two different clusters, and the year of origin of the variety is significantly reflected, because the main factor that distinguishes these clusters is the parameter “year of origin/registration”. It also often reflects changes in the breeding programme, the use of different genetic resources, or the focus on the breeding of new traits (Figure 4).

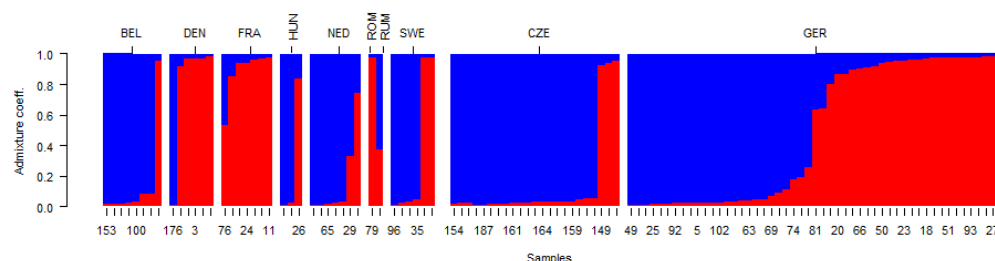


Figure 4. Classification of accessions from selected countries of origin, according to Q1 (red)/Q2 (blue) values and the year of registration.

This distribution of varieties is evident, for example, in Hungarian varieties, where the most recent and older varieties differ significantly in genetic structure. A similar characteristic can be seen in varieties from the Netherlands, Belgium, Denmark, France, Sweden, and Romania. Different characters and explanations may be given for grouping German and Czech varieties. Four German varieties (‘Seco’, ‘Semper’, ‘Setoria’, and ‘Sigri’) belonged to the same group with closely similar Q1 values. They were bred at one breeding station, and their genetic structure reflects the significant similarity of the genetic resources used in breeding programmes. In the case of Czech varieties, both landraces (‘Přerovská bílá’, ‘Česká krajová’), the older variety ‘Zlata’, and modern varieties from the breeding stations Opava and Krukanice belong to the same group. The second cluster comprises current modern varieties from breeding programmes in Selton and BOR (‘Agent’ and ‘Veronika’). In these cases, there is a strong influence of a limited set of genetic resources, and molecular marker methods and the evaluation of genetic diversity can help to appropriate the management of genetic resources and to spread genetic variability and shifts in genetic structures. In this study, some landraces and older varieties were not significantly different from modern varieties. However, it is still evident that the origin of the variety, country of origin, or the breeding programme significantly influences the genetic structure. Similar results were presented in oil seed brassicas [32] or in *B. juncea* [42] and *B. carinata* [34] genetic resources. In contrast, in *B. rapa*, no significant loss of genetic diversity was recorded, although most total variation was attributed to within-cultivar variation in cross-pollinated cultivars [43]. The large geographical variation of genetic resources is a significant source of genetic variation, and high variation among subpopulations of genetic sources has been observed in *B. juncea* [44].

An essential problem in current breeding is the insufficient range of genetic diversity and the related limited availability of interesting genetic resources, such as donors of required traits and characteristics [45]. The depletion and narrowing of genetic diversity occurred in the context of the development of modern breeding in the early 20th century, due to intensive breeding focused on a limited range of traits and the influence of the cultivation of high-yielding and high-quality, but often relatively closely related, varieties of agricultural crops [46,47]. One possibility to expand the range of genetic variability is TILLING technology [48]. This modern approach, combining classical mutagenesis with NGS techniques and the detection of genes carrying mutations, seems to be a very promising breeding method. Plants carrying a change/mutation in the gene of interest are potential donors of a new characteristic and are incorporated into breeding programmes [49]. The

results of the microsatellite analysis show, in this study, a significant spread of genetic variability in the mutant population. All plants from the mutant population were incorporated into the yellow mustard breeding programme and, in the following years, qualitative traits and their possible changes due to mutation will be assessed.

3. Materials and Methods

Plant material and DNA extraction. For the evaluation of genetic diversity, a pilot set of 198 high-value accessions was chosen from the collection of yellow mustard kept at the OSEVA Development and Research in Opava, Czech Republic. These accessions were selected based on their value for current breeding programmes, including modern and old registered varieties, landraces, breeding material, and accessions from the TILLING population. The accessions came from various countries, representing a wide range of spring-type cultivars with a wide range of seed colours. The list of genetic resources, their accession numbers in the gene bank, and their country of origin are given in Table S1, in the Supplementary Data. The seeds were seeded on a horticultural peat substrate and grown under laboratory conditions at a temperature of 21 °C and 16/8 h for five days. The bulk of 36 cotyledons was sampled and gently dried in silica gel. The dried material was homogenised using glass beads in a Retsch® Mixer Mill MM400 (Haan, Germany) for 30 s at a maximal frequency (30 Hz). The DNA was extracted by a modified CTAB PVP method [50], which is more time-consuming, but the obtained DNA is very pure and highly concentrated.

Microsatellite amplification. A total of 15 microsatellite markers were used: BoREM1b_A, BoREM1b_B, BoREM1b_C, BoLAB19TF, BoPC34 [51], P381, D3_A, D3_B, P7, P9_B, P9_C, P30_B, P30_C, P35_A, and P35_B [31]. Three primers were used in each PCR reaction, one universal FAM-labelled primer (Ba02 6-FAM or Ba03 6-FAM), a forward primer extended by the sequence of the labelled primer, and a reverse primer [52]. The PCR reaction was carried out in a 10 µL, in 1× reaction buffer (75 mM Tris—HCl, pH = 8.8, 20 mM (NH₄)₂ SO₄, 0.01% Tween 20, 2.5 mM MgCl₂, 200 µM dNTPs), 10 pmol of each primer, 0.5 U of Taq Purple DNA polymerase PPP Master Mix (Top-Bio, Vestec, Czech Republic) or 2× Master Mix with standard buffer (NEB, Ipswich, MA, USA), 2× BSA, and 50 ng of template DNA. The amplification was carried out on a Biometra TProfessional thermocycler (Göttingen, Germany) in the following temperature profile: initial denaturation of 5 min at 94 °C, 31 cycles of 60 s at 95 °C, 60 s at 52–59 °C (depending on the primer used), and 60 s at 72 °C/68 °C, with a final elongation of 15 min at 72 °C/68 °C.

Data Analysis. The results of the PCR analysis were evaluated by fragment analysis on an ABI 3500 genetic analyser (Applied Biosystems, Waltham, MA, USA). Molecular data were analysed using GeneMapper version 3.0 software. The weighted arithmetic mean of genetic distances (WAM) was calculated based on frequency values in the appropriate class of genetic diversity [32]. Diversity measures were calculated in adegenet for the R software version 4.2.2 [53,54]. They consisted of the number of alleles (NA), major allele frequencies (MAF), observed heterozygosity (Ho), expected heterozygosity (He), and the polymorphism information content (PIC). To determine genetic diversity among and within genetic resources, AMOVAs [55] were performed with the package “pegas” in R [56]. The null hypotheses associated with differentiation parameters (Φ) were as follows: there is no genetic difference between types of genetic resources ($\Phi_{CT} = 0$), between genotypes within types of genetic resources ($\Phi_{SC} = 0$), or between genotypes within collection of genetic resources ($\Phi_{ST} = 0$) [55]. The data were statistically evaluated using traditional genetic diversity metrics in MVSP 3 (Kovach Computing Services, Pentraeth, UK) and DARwin 6 (CIRAD, F) software. Intending to estimate the population/cluster number, we used the Bayesian clustering approach, involving MCMC estimation, and followed Evanno et al. [40] and Kolář et al. [38]. STRUCTURE software was used to assess the genetic structure of analysed germplasms [39].

4. Conclusions

The issue of genetic diversity, its conservation, protection, and exploitation are the key points in plant breeding. In this study, we focused on the detailed description and molecular analysis of the genetic variability of a large collection of yellow mustard genotypes. High values of observed heterozygosity, diversity index, and WAM index indicated a high level of polymorphism between individual accessions. Also, outputs from AMOVA supported these findings and showed a low proportion of variance (4.6%) between types of genetic resources (landrace, variety, and breeding material) and a high proportion of variance between genotypes within the analysed set of genetic resources (87.13%). In addition to the standard descriptive characteristics, two other procedures were applied, for an overall assessment of genetic variability and the degree of diversity of the genetic resources: PCoA analysis and Bayesian clustering algorithms. The results show a more valuable interpretation and highlight another way of processing the genetic diversity data and indicated mating barriers between varieties and reproductive isolation, due to the limited exchange of genetic resources between breeders. The range of genetic diversity was much higher than reported in other studies, concerning the large and variable collection of genetic resources. These results showed the importance of the proper management of genetic resources for the preservation of genetic diversity. Analysis of genetic structure may lead (as shown in our example) to a clarification of the grouping of accessions (genetic resources and varieties) and recommendations for expanding the genetic background in breeding programmes. The choice of parental components based on phenotypic characteristics and genetic diversity facilitates the selection of genetic resources for breeding programmes and can contribute to the success of breeding.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/plants12234026/s1>, Table S1. List of analysed genetic resources of yellow mustard and outputs of the STRUCTURE software (Q1/Q2 values). Table S2. Detailed output of cluster number estimation and statistics for different values of K. Figure S1. The graphic representation of the degree of dissimilarity between genetic resources—output from DARwin 6.0 software.

Author Contributions: Conceptualization, V.Č., M.K. and E.J.; methodology, E.J.; software, M.R.; validation, E.J. and O.H.; formal analysis, E.J. and O.H.; investigation, V.Č.; resources, A.R.; data curation, V.Č.; writing—original draft preparation, V.Č.; writing—review and editing, P.B., D.S., B.P., E.J. and M.K.; visualisation, M.R.; supervision, V.Č.; project administration, M.K.; funding acquisition, M.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Ministry of Agriculture of the Czech Republic, grant number QK1910225, and the Grant Agency of the University of South Bohemia in České Budějovice, grant number GAJU 080/2022/Z.

Data Availability Statement: Additional data files and project outputs are available at <http://biocentrum.zf.jcu.cz/laborator.php?sub=metodiky> (accessed on 20 November 2023).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hemingway, J.S. The mustard species: Condiment and food ingredient use and potential as oilseed crops. In *Brassica Oilseeds Production and Utilization*; Kimber, D., McGregor, D.I., Eds.; CAB International: Wallingford, UK, 1995; pp. 373–383.
2. Ruan, S.F.; Wang, Z.X.; Xiang, S.J.; Chen, H.J.; Shen, Q.; Liu, L.; Wu, W.F.; Cao, S.W.; Wang, Z.W.; Yang, Z.J.; et al. Mechanisms of white mustard seed (*Sinapis alba* L.) volatile oils as transdermal penetration enhancers. *Fitoterapia* **2019**, *138*, 104195. [[CrossRef](#)] [[PubMed](#)]
3. Bodnaryk, R.P.; Lamb, R.J. Mechanisms of resistance to the flea beetle, *Phyllotreta cruciferae* (Goeze), in mustard seedlings, *Sinapis alba* L. *Can. J. Plant Sci.* **1991**, *71*, 13–20. [[CrossRef](#)]
4. Kumari, P.; Singh, K.P.; Rai, P.K. Draft genome of multiple resistance donor plant *Sinapis alba*: An insight into SSRs, annotations and phylogenetics. *PLoS ONE* **2020**, *15*, e0231002. [[CrossRef](#)] [[PubMed](#)]
5. Valdes, Y.; Viaene, N.; Moens, M. Effects of yellow mustard amendments on the soil nematode community in a potato field with focus on *Globodera rostochiensis*. *Appl. Soil Ecol.* **2012**, *59*, 39–47. [[CrossRef](#)]

6. Krasnodębska-Ostrega, B.; Sadowska, M.; Biaduń, E.; Mazur, R.; Kowalska, J. *Sinapis alba* as a useful plant in bioremediation—Studies of defense mechanisms and accumulation of As, Tl and PGEs. *Int. J. Phytoremediat.* **2022**, *24*, 1475–1490. [[CrossRef](#)] [[PubMed](#)]
7. Masierowska, M. Floral nectaries and nectar production in brown mustard (*Brassica juncea*) and white mustard (*Sinapis alba*) (Brassicaceae). *Plant Syst. Evol.* **2003**, *238*, 97–107. [[CrossRef](#)]
8. Katepa-Mupondwa, F.; Gugel, R.K.; Raney, J.P. Genetic diversity for agronomic, morphological and seed quality traits in *Sinapis alba* L. (yellow mustard). *Can. J. Plant Sci.* **2006**, *86*, 1015–1025. [[CrossRef](#)]
9. Khoury, C.K.; Brush, S.; Costich, D.E.; Curry, H.A.; de Haan, S.; Engels, J.M.M.; Guarino, L.; Hoban, S.; Mercer, K.L.; Miller, A.J.; et al. Crop genetic erosion: Understanding and responding to loss of crop diversity. *New Phytol.* **2022**, *233*, 84–118. [[CrossRef](#)]
10. Yu, J.K.; Chung, Y.S. Plant Variety Protection: Current Practices and Insights. *Genes* **2021**, *12*, 1127. [[CrossRef](#)]
11. Govindaraj, M.; Vetriventhan, M.; Srinivasan, M. Importance of genetic diversity assessment in crop plants and its recent advances: An overview of its analytical perspectives. *Genet. Res. Int.* **2015**, *2015*, 431487. [[CrossRef](#)]
12. Garland, S.; Curry, H.A. Turning promise into practice: Crop biotechnology for increasing genetic diversity and climate resilience. *PLoS Biol.* **2022**, *20*, e3001716. [[CrossRef](#)] [[PubMed](#)]
13. McCallum, C.M.; Comai, L.; Greene, E.A.; Henikoff, S. Targeted screening for induced mutations. *Nat. Biotechnol.* **2000**, *18*, 455–457. [[CrossRef](#)] [[PubMed](#)]
14. Tsai, H.; Howell, T.; Nitcher, R.; Missirian, V.; Watson, B.; Ngo, K.J.; Lieberman, M.; Fass, J.; Uauy, C.; Tran, R.K.; et al. Discovery of Rare Mutations in Populations: TILLING by Sequencing. *Plant Physiol.* **2011**, *156*, 1257–1268. [[CrossRef](#)] [[PubMed](#)]
15. El-Esawi, M.A. Genetic diversity and evolution of Brassica genetic resources: From morphology to novel genomic technologies—A review. *Plant Genet. Resour.* **2017**, *15*, 388–399. [[CrossRef](#)]
16. Granot, D.; Shabelsky, E.; Schaffermann, D.; Yaniv, Z. Analysis of genetic variability between populations of *Sinapis alba* and the effect of cultivation on the variability. *Acta Hort.* **1996**, *407*, 67–74. [[CrossRef](#)]
17. Fu, Y.B.; Gugel, R.K.; Katepa-Mupondwa, F. Genetic diversity of *Sinapis alba* germplasm as revealed by AFLP markers. *Plant Genet. Resour.* **2006**, *4*, 87–95. [[CrossRef](#)]
18. Yadav, M.; Rana, J.S. ISSR markers assisted studies on genetic diversity in *Brassica juncea*. *Int. J. Pharma Bio Sci.* **2012**, *3*, 402–411.
19. Plieske, J.; Struss, D. Microsatellite markers for genome analysis in *Brassica*. I. Development in *Brassica napus* and abundance in Brassicaceae species. *Theor. Appl. Genet.* **2001**, *102*, 689–694. [[CrossRef](#)]
20. Javidfar, F.; Cheng, B. Construction of a genetic linkage map and QTL analysis of erucic acid content and glucosinolate components in yellow mustard (*Sinapis alba* L.). *BMC Plant Biol.* **2013**, *13*, 142. [[CrossRef](#)]
21. Fu, Y.B.; Cheng, B.; Peterson, G.W. Genetic diversity analysis of yellow mustard (*Sinapis alba* L.) germplasm based on genotyping by sequencing. *Genet. Resour. Crop Evol.* **2013**, *61*, 579–594. [[CrossRef](#)]
22. Karp, A.; Edwards, K.J. *Molecular Techniques in the Analysis of the Extent and Distribution of Genetic Diversity*; International Plant Genetic Resources Institute: Rome, Italy, 1997; pp. 11–38.
23. Singh, K.H.; Singh, L.; Parmar, N.; Kumar, S.; Nanjundan, J.; Singh, G.; Thakur, A.K. Molecular characterization and genetic diversity analysis in Indian mustard (*Brassica juncea* L. Czern & Coss.) varieties using SSR markers. *PLoS ONE* **2022**, *17*, e0272914. [[CrossRef](#)]
24. Pradhan, A.; Nelson, M.N.; Plummer, J.A.; Cowling, W.A.; Yan, G. Characterization of *Brassica nigra* collections using simple sequence repeat markers reveals distinct groups associated with geographical location, and frequent mislabelling of species identity. *Genome* **2011**, *54*, 50–63. [[CrossRef](#)] [[PubMed](#)]
25. Tesfaye, M.; Feyissa, T.; Hailesilassie, T.; Kanagarajan, S.; Zhu, L.-H. Genetic diversity and population structure in ethiopian mustard (*Brassica carinata* A. Braun) as revealed by single nucleotide polymorphism markers. *Genes* **2023**, *14*, 1757. [[CrossRef](#)] [[PubMed](#)]
26. Guichoux, E.; Lagashe, L.; Wagner, S.; Chaumeil, P.; Léger, P.; Lepais, O.; Lepoitteiven, C.; Malausa, T.; Revardel, E.; Salin, F.; et al. Current trends in microsatellite genotyping. *Mol. Ecol. Resour.* **2011**, *11*, 591–611. [[CrossRef](#)] [[PubMed](#)]
27. Mondini, L.; Noorani, A.; Pagnotta, M. Assessing plant genetic diversity by molecular tools. *Diversity* **2009**, *1*, 19–35. [[CrossRef](#)]
28. Varshney, R.K. Designing future crops: Genomics-assisted breeding comes of age. *Trends Plant. Sci.* **2021**, *26*, 631–649. [[CrossRef](#)]
29. Olsson, G. Self-incompatibility and outcrossing in rape and white mustard. *Hereditas* **1960**, *46*, 241–252. [[CrossRef](#)]
30. Treccarichi, S.; Ben Ammar, H.; Amari, M.; Cali, R.; Tribulato, A.; Branca, F. Molecular markers for detecting inflorescence size of *Brassica oleracea* L. crops and *B. oleracea* complex species (n = 9) useful for breeding of broccoli (*B. oleracea* var. italica) and cauliflower (*B. oleracea* var. botrytis). *Plants* **2023**, *12*, 407. [[CrossRef](#)]
31. Cui, X.; Dong, Y.; Hou, X.; Cheng, Y.; Zhang, J.; Jin, M. Development and characterization of microsatellite markers in *Brassica rapa* ssp. chinensis and transferability among related species. *Agric. Sci. China* **2008**, *7*, 19–31. [[CrossRef](#)]
32. Havlickova, L.; Jozova, E.; Rychla, A.; Klima, M.; Kucera, V.; Curn, V. Genetic diversity assessment in winter oilseed rape (*Brassica napus* L.) collection using AFLP, ISSR and SSR markers. *Czech J. Genet. Plant Breed.* **2014**, *50*, 216–225. [[CrossRef](#)]
33. Vazhakkannadi, V.; Singh, N.; Vasudev, S.; Yadava, D.; Kumar, S.; Naresh, S.; Bhat, S.; Prabhu, K. Assessment of genetic diversity in *Brassica juncea* (Brassicaceae) genotypes using phenotypic differences and SSR markers. *Rev. Biol. Trop.* **2013**, *61*, 1919–1934.
34. Khedikar, Y.; Clarke, W.; Chen, L.; Higgins, E.; Kagale, S.; Koh, C.; Bennett, R.; Parkin, I. Narrow genetic base shapes population structure and linkage disequilibrium in an industrial oilseed crop, *Brassica carinata* A. Braun. *Sci. Rep.* **2020**, *10*, 12629. [[CrossRef](#)] [[PubMed](#)]

35. Sobotka, R.; Dolanská, L.; Čurn, V.; Ovesná, J. Fluorescence-based AFLPs occur as the most suitable marker system for oilseed rape cultivar identification. *J. Appl. Genet.* **2004**, *45*, 161–173. [PubMed]
36. Cowling, W.A. Genetic diversity in Australian canola and implications for crop breeding for changing future environments. *Field Crops Res.* **2007**, *104*, 103–111. [CrossRef]
37. Chikkaputtaiah, C.; Debbarma, J.; Baruah, I.; Havlickova, L.; Deka Boruah, H.P.; Curn, V. Molecular genetics and functional genomics of abiotic stress-responsive genes in oilseed rape (*Brassica napus* L.): A review of recent advances and future. *Plant Biotechnol. Rep.* **2017**, *11*, 365–384. [CrossRef]
38. Kolář, F.; Fér, T.; Štěch, M.; Trávníček, P.; Dušková, E.; Schönschwetter, P.; Suda, J. Bringing together evolution on serpentine and polyploidy: Spatiotemporal history of the diploid-tetraploid complex of *Knautia arvensis* (Dipsacaceae). *PLoS ONE* **2012**, *7*, e39988. [CrossRef] [PubMed]
39. Pritchard, J.K.; Stehens, M.; Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **2000**, *155*, 945–959. [CrossRef]
40. Evanno, G.; Regnaut, S.; Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.* **2005**, *14*, 2611–2620. [CrossRef]
41. Bird, C.; Karl, S.A.; Smouse, P.E.; Toonen, R.J. Detecting and measuring genetic differentiation. In *Phylogeography and Population Genetics in Crustacea*; Koenemann, S., Held, C., Schubart, C., Eds.; CRC Press: Boca Raton, FL, USA, 2011; pp. 31–55. [CrossRef]
42. Yao, Q.L.; Chen, F.B.; Fang, P.; Zhou, G.F.; Fan, Y.H.; Zhang, Z.R. Genetic diversity of Chinese vegetable mustard (*Brassica juncea* Coss) landraces based on SSR data. *Biochem. Syst. Ecol.* **2012**, *45*, 41–48. [CrossRef]
43. Ofori, A.; Becker, H.C.; Kopisch-Obuch, F.J. Effect of crop improvement on genetic diversity in oilseed *Brassica rapa* (turnip-rape) cultivars, detected by SSR markers. *J. Appl. Genet.* **2008**, *49*, 207–212. [CrossRef]
44. Abdel-haleem, H.; Luo, Z.; Szczepanek, A. Genetic diversity and population structure of the USDA collection of *Brassica juncea* L. *Ind. Crops Prod.* **2022**, *187*, 115379. [CrossRef]
45. Sanchez, D.; Sadoun, S.B.; Mary-Huard, T.; Allier, A.; Moreau, L.; Charcosset, A. Improving the use of plant genetic resources to sustain breeding programs' efficiency. *Proc. Natl. Acad. Sci. USA* **2023**, *120*, e2205780119. [CrossRef] [PubMed]
46. Bhandari, H.; Bhanu, A.; Srivastava, K.; Singh, M.; Shreya; Hemantaranjan, A. Assessment of genetic diversity in crop plants: An overview. *Adv. Plants Agric. Res.* **2017**, *7*, 279–286. [CrossRef]
47. Begna, T. Role and economic importance of crop genetic diversity in food security. *Int. J. Agric. Sci. Food Technol.* **2021**, *7*, 164–169. [CrossRef]
48. Kurowska, M.; Daszkowska-Golec, A.; Gruszka, D.; Marzec, M.; Szurman, M.; Szarejko, I.; Maluszynski, M. TILLING: A shortcut in functional genomics. *J. Appl. Genet.* **2011**, *52*, 371–390. [CrossRef] [PubMed]
49. Mba, C. Induced mutations unleash the potentials of plant genetic resources for food and agriculture. *Agronomy* **2013**, *3*, 200–231. [CrossRef]
50. Doyle, J. DNA Protocols for Plants. In *Molecular Techniques in Taxonomy*; Hewitt, G.M., Johnston, A.W.B., Young, J.P.W., Eds.; Springer: Berlin/Heidelberg, Germany, 1991; pp. 283–293. [CrossRef]
51. Louarn, S.; Torp, A.; Holme, I.; Andersen, S.B.; Jensen, B. Database derived microsatellite markers (SSRs) for cultivar differentiation in *Brassica oleracea*. *Genet. Resour. Crop Evol.* **2007**, *54*, 1717–1725. [CrossRef]
52. Jozová, E.; Čurn, V. Microsatellite genotyping with 4-color fluorescent detection using multiple-tailed primers in oilseed rape. *Úroda* **2014**, *12*, 191–194.
53. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020; Available online: <https://www.R-project.org/> (accessed on 30 October 2022).
54. Jombart, T.; Ahmed, I. Adegnet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics* **2011**, *27*, 3070–3071. [CrossRef]
55. Excoffier, L.; Smouse, P.E.; Quattro, J.M. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* **1992**, *131*, 479–491. [CrossRef]
56. Paradis, E. *pegas*: An R package for population genetics with an integrated-modular approach. *Bioinformatics* **2010**, *26*, 419–420. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Na tomto místě se v plné verzi habilitační práce nachází text příslušné publikace. Ten však není součástí této veřejně přístupné verze z důvodu licenčních podmínek vydavatele. Níže je k dispozici bibliografická citace a abstrakt článku. Plný text lze nalézt v placeném režimu na webových stránkách časopisu.

Mráz, I., Beran, P., Šerá, B., Gavril, B., & Hnatiuc, E. (2014). Effect of low-temperature plasma treatment on the growth and reproduction rate of some plant pathogenic bacteria. *Journal of Plant Pathology*, 96(1), 63–67.

Abstrakt

Atmospheric pressure plasmas have received much attention lately as a promising physical tool for biological decontamination and sterilization. Two phytopathogenic bacterial species, *Clavibacter michiganensis* subsp. *michiganensis* and *Erwinia amylovora*, and one cosmopolitan bacterium, *Escherichia coli* (control), were employed to study the bacterial growth mechanism after exposure to the atmospheric pressure GlidArc plasma. The results have shown that low-temperature plasma treatment affects all three bacterial species by slowing down growth and reproduction rate. After plasma treatment, both Gram-negative bacteria (*E. coli* and *E. amylovora*) reached their maximum growth sooner than the Gram-positive *C. michiganensis* subsp. *michiganensis*. This is however caused by a longer cultivation period of *Clavibacter* bacterial genus. The application of GlidArc plasma is a suitable treatment that can be used in agriculture for plant protection.

Article

Duplex PCR Detection and Differentiation of Insect DNA *Tenebrio molitor* and *Zophobas morio* in Various Types of Food

Dagmar Stehlíková ¹, Veronika Müllerová ¹, Anna Adámková ², Pavel Beran ¹, Martin Adámek ³, Vladislav Čurn ¹, Soňa Škrovánková ² and Jiří Mlček ^{2,*}

¹ Department of Genetics and Biotechnology, Faculty of Agriculture and Technology, University of South Bohemia in Ceske Budejovice, Na Sadkach 1780, 370 05 Ceske Budejovice, Czech Republic; stehld@fzt.jcu.cz (D.S.); mullev05@fzt.jcu.cz (V.M.); beranp@fzt.jcu.cz (P.B.); curn@fzt.jcu.cz (V.Č.)

² Department of Food Analysis and Chemistry, Faculty of Technology, Tomas Bata University in Zlin, Vavreckova 5669, 760 01 Zlin, Czech Republic; aadamkova@utb.cz (A.A.); skrovankova@utb.cz (S.Š.)

³ Department of Automation and Control Engineering, Faculty of Applied Informatics, Tomas Bata University in Zlin, Nad Stranemi 4511, 760 05 Zlin, Czech Republic; m2adamek@utb.cz

* Correspondence: mlcek@utb.cz

Simple Summary

In this study, two types of edible insects were examined: mealworms (*Tenebrio molitor*) and superworms (*Zophobas morio*). While mealworms are approved for use in food in the EU, superworms are not. Therefore, only approved insect species can be used in food products. Duplex PCR was developed to detect whether a food product contains mealworms and to ensure that superworms are not included. This method was designed to be highly accurate, even when applied to complex foods such as bread, chocolate, or porridge. Insect DNA was successfully detected, even in very small amounts. Various food products were tested. All of them were labeled as containing mealworms, and this was confirmed through testing. No traces of superworms were found. Even after cooking or baking, the insects could still be identified using the method. Using this approach, food fraud can be prevented, consumer protection can be improved, and EU food regulations can be followed more effectively. The control of insect ingredients in food products, therefore, will be easier.



Received: 11 August 2025

Revised: 2 September 2025

Accepted: 19 September 2025

Published: 21 September 2025

Citation: Stehlíková, D.; Müllerová, V.; Adámková, A.; Beran, P.; Adámek, M.; Čurn, V.; Škrovánková, S.; Mlček, J. Duplex PCR Detection and Differentiation of Insect DNA *Tenebrio molitor* and *Zophobas morio* in Various Types of Food. *Insects* **2025**, *16*, 983. <https://doi.org/10.3390/insects16090983>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract

Edible insects, particularly *Tenebrio molitor* (Linnaeus) (mealworm) and *Zophobas morio* (Fabricius) (superworm), have drawn increasing attention as alternative protein sources. This study aims to develop an accurate molecular detection method for *T. molitor*, an EU-approved food species, and to differentiate it from *Z. morio*, which remains unapproved for human consumption in the EU. The process enables precise and sensitive identification methods by optimizing singleplex and duplex PCR techniques targeting 16S rRNA and COI gene regions. The DNA of *T. molitor* was detected in various food matrices, including pastries, chocolate, and porridge, while avoiding cross-reactivity with *Z. morio*, *Gryllus assimilis*, and *Locusta migratoria*. The detection limit for both singleplex and duplex PCR was 10 pg of DNA, ensuring robustness against inhibitory effects from complex food matrices. The developed approach ensures reliable detection and compliance with EU regulations regarding insect-based foods, providing a critical tool for food authentication and preventing adulteration. The key advancements of this approach lie in its improved specificity and sensitivity, allowing for the ability to detect complex food matrices. An applied perspective was evaluated using real commercial food products.

Keywords: edible insects; molecular detection; novel food

1. Introduction

The demand for food is driven not only by the growing human population, but also by rising living standards. Therefore, the production of proteins from sustainable systems is a highly debated topic. Proteins obtained from non-conventional sources must be safe, nutritionally valuable, produced in sufficient and stable quantities, and be accepted by consumers. Edible insects, one of the non-conventional commodities, have high potential as an alternative protein source, and could become an important food and feed raw material [1]. Additionally, insects may also offer a more efficient way of producing animal protein with respect to environmental sustainability and the use of cost-effective raw materials [2–4].

As a reaction to rising interest in using insects as food, some species from orders Coleoptera and Orthoptera were added to the European Union's novel food list. One of them, *T. molitor*, or the yellow mealworm, used to be solely viewed as a storage pest and pet feed. In 2021, *T. molitor* larvae were marked as safe for consumption by the EU [5]. Recently, foodstuffs containing *T. molitor* larvae have entered the European food market, with the aim of increasing the protein content in products while utilizing a promising alternative protein source. Dried larvae are sold as whole in different flavors, as well as in the form of meal incorporated in pastry, chocolate, porridge, or snack bars. Raising *T. molitor* as a protein source is attractive due to lower space requirements, easy maintenance, high yield, and relatively low price [6].

For safe consumption, control measures need to be established in order to monitor the correct use of product packaging and to detect fraudulent activity. For confirming *T. molitor* content in insect-based foods, a method is needed to differentiate other similar species, which are not on the novel foods list. Confusion in detection could be caused by *Z. morio*, also called superworm. *Z. morio* is also a storage pest belonging to the same family Tenebrionidae. The larger size of *Z. morio* larvae and their lower maintenance costs could lead to potential implementation in insect-based foods. However, *Z. morio* larvae are currently not approved for consumption in the EU, and remain used only as animal feed. Using a modified microscopic method incorporating a double-sedimentation protocol, insect particles were detected in poultry feed containing 1% insect material, including *T. molitor*. While fragments of muscle tissue and tracheoles confirmed the presence of insect-derived material, they did not allow for species-level identification. Therefore, a method to verify the content of approved insect species in insect-based food is needed [7]. There are no validated immunoassays with antibodies specific for *T. molitor* yet. However, there is a more general pilot use of standard ELISA kits (e.g., for crustacean tropomyosin or chitosan detection) [8,9].

Molecular methods pose a reliable approach to insect detection and identification in foodstuffs. Polymerase chain reaction stands out as a reference method due to its sensitivity and accuracy [10]. Nonetheless, the accuracy of this method could be influenced by input material, as certain food ingredients could inhibit the PCR reaction and cause inaccurate outcomes. Thus, optimization of the method is necessary. This study aims to propose a singleplex and multiplex PCR method to verify *T. molitor* content in insect-based foods sold on the Czech food market, such as pastry, chocolate, and porridge. Specificity testing is performed against *Z. morio* and two other novel food insect species, *G. assimilis* and *L. migratoria*. This study addresses the critical need for the reliable identification of approved insect species in food products, focusing on *T. molitor*. Accurate detection is essential for consumer safety and regulatory compliance. This is the first optimized molecular method for the verification of *T. molitor* in processed foods on the Czech market, making this work a new and significant contribution to the field.

2. Materials and Methods

2.1. Insect Sample Preparation

Prior to the analysis, larvae of *T. molitor*, *Z. morio* in the last and second-to-last instars of development (full body length just before pupation), and imagos of *L. migratoria* and *G. assimilis* were collected in the insectarium of UTB Zlín and transferred to the laboratory. The samples, both larvae and imagos, were starved for 48 h, killed with boiling water (100 °C), and dried at 105 °C. Subsequently, the samples were milled in a bead-beater, homogenized, and stored in a refrigerator at 4–7 °C until analysis.

2.2. Pastry Samples

Pastry bread containing different percentages of edible insect larvae meal were prepared for the detection of edible insect species in foodstuffs. The bread dough was composed of 55 mL water, 0.93 g sugar, 0.75 g salt, 0.9 g yeast, and 50 g mix of white flour and different amounts of larvae meal (1%, 2.5%, 10%). Finally, the doughs were baked in a hot-air oven for 35 min at 165 °C.

2.3. Insect-Based Food

In addition to pastry, different insect-based foods were used for analysis. All products are sold on the Czech food market and contain *T. molitor* larvae. Selection of products included samples of dried larvae, “Křupaví červíci” (WormUP, Blansko, Czech Republic), with garlic and chili flavors and dark low-carb chocolate (Farmerio, Svatoňovice, Czech Republic) containing cocoa beans, cocoa butter, and *T. molitor* larvae. The last analyzed product was instant rice porridge (Farmerio, Svatoňovice, Czech Republic), containing husked rice (84%), hominy (2.5%), *T. molitor* larvae meal (11.5%), and lyophilized raspberries (2%).

2.4. DNA Extraction

In total, 30 mg of material was homogenized using 3 mm metal beads in a Retsch® Mixer Mill MM400 (Haan, Germany) for 30 s at a maximal frequency (30 Hz). The DNA was extracted by a modified CTAB PVP method [11]. Finally, the DNA pellet was dissolved in 1× TE buffer and stored in −20 °C. The quality and quantity of the extracted DNA were assessed spectrophotometrically using a BioSpec-nano spectrophotometer (Shimadzu, Long Beach, CA, USA) by measuring the absorbance at 260 nm (A260) and 280 nm (A280). DNA purity was determined by calculating the A260/A280 ratio.

2.5. PCR Analysis

The tested primers were designed to target specific regions of the 16S rRNA and COI genes of *T. molitor* and *Z. morio*, respectively [11]. We tested their coamplification in a novel multiplex PCR by modifying the PCR protocol with primers TM_F (5′ CAAAGGTAGCATAATCATTAGT 3′), TM_R (5′ AGTTAAATAAATTTTCTAACCG 3′), ZM_F (5′ GGGCATCAGTCGATCTCGCA 3′), and ZM_R (5′ CGATCAAAAGTTATTCCTTGTTG 3′) [12]. PCR amplification was performed using a Biometra TProfessional thermocycler (Biometra GmbH, Göttingen, Germany). To optimize annealing temperature, a gradient PCR was conducted (52–60 °C) for each primer set. The optimal annealing temperature was determined to be 55 °C for both sets of primers. Each 25 µL PCR reaction contained 12.5 µL One Taq® Hot Start Quick-Load® Master Mix (New England Biolabs, Ipswich, MA, USA), 9.5 µL PCR Ultra H₂O Top-Bio, 1 µL forward primer, 1 µL reverse primer, and 1 µL template DNA. PCR conditions were as follows: initial denaturation at 94 °C for 5 min, followed by 35 cycles of denaturation at 94 °C for 60 s, annealing at 55 °C for 30 s, and extension at 68 °C for 60 s. A final extension step was performed at 68 °C for 5 min. PCR products were analyzed by electrophoresis on a 2% agarose gel in 1× TBE buffer. A 100 bp

DNA ladder (Neb, Hitchin, UK) was used as a size marker. Electrophoresis was carried out at 120 V for 40 min, and the gel was visualized using InGenius 3 (SynGene, Cambridge, UK).

2.6. Specificity and Sensitivity Testing

The specificity of the PCR detection was verified using DNA from four insect species. Ten nanograms of DNA were used in each PCR reaction, and each DNA extract was tested in duplicate. To confirm the specificity of the primers to their target species, PCR products were extracted from gel using NucleoSpin Gel and PCR Clean-up (Macherey-Nagel, Düren, Germany), sequenced by an external service provider SEQme s.r.o. (Dobris, Czech Republic), and analyzed using BLAST (version 2.15.0; NCBI, Bethesda, MD, USA) [13]. To determine the limit of detection (LOD), 10-fold serial dilutions of known amounts (100 ng/μL) of *T. molitor* and *Z. morio*. DNA were tested using both singleplex and multiplex PCR. The LOD was determined as the lowest concentration of DNA that could be reliably detected.

3. Results

3.1. Samples

Positive controls were prepared milled of dried larvae and imagos of *T. molitor*, *Z. morio* (Figure 1a). Pastry breads were baked with different amounts of larvae meal (1%, 2.5%, 10%) (Figure 1b).



Figure 1. Fresh samples and pastry samples: (a) larvae and imagos of *T. molitor* (left green line) *Z. morio* (right blue line) before drying; (b) baked bread with different amounts (1%, 2.5%, 10%) of larvae meal (example with *T. molitor*).

3.2. DNA Extraction

DNA concentrations ranged from tens to hundreds of ng/ μ L. Purity levels were influenced by the type of input material used. At 1% insect meal content, the total nucleic acid concentration was the lowest among the three tested levels. In commercial insect-containing products, the detection efficacy was the lowest; however, no significant PCR inhibition was observed. Among the tested matrices, chocolate exhibited the highest inhibitory effect.

3.3. PCR

Singleplex and duplex PCR approaches were used to detect the presence of edible insects in food. In the case of *T. molitor*, a 240 bp fragment of the 16S rRNA gene was specifically amplified, while in the case of *Z. morio*, a 120 bp fragment of the COI gene was amplified. Both primer sets were tested at six annealing temperatures. Temperatures of 52 and 55 °C were suitable for the detection of *T. molitor*, while *Z. morio* were detected at all six tested temperatures.

To determine the sensitivity of the singleplex and duplex PCR assays, DNA isolated from crushed *T. molitor* and *Z. morio* samples was used. Singleplex and duplex PCR were performed using 100 ng–1 pg of DNA as a template and the *T. molitor* and *Z. morio* primers. A DNA concentration of 100 pg was determined as the limit of detection for both singleplex and duplex PCR (Figure 2).

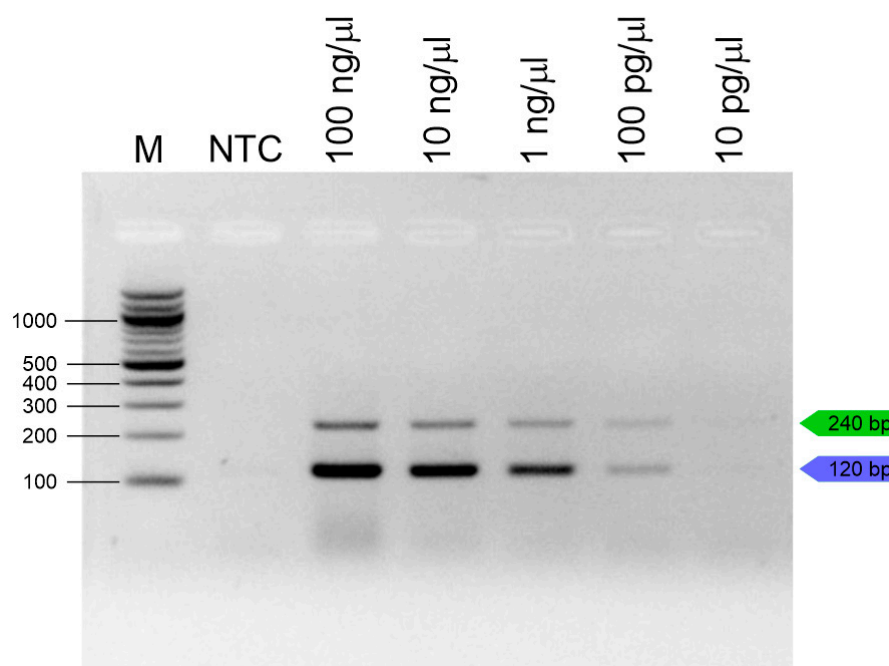


Figure 2. Sensitivity of duplex PCR (10-fold serial dilution); green lines (*T. molitor*); blue lines (*Z. morio*), M—100 bp ladder (NEB, UK); NTC—no-template control.

In duplex PCR, no cross-reactions were observed, and sequencing PCR products showed 100% similarity to *T. molitor* and *Z. morio* sequences in the GenBank database and confirmed the specificity of the primers. Singleplex and duplex PCR accurately and specifically detected insect DNA from the target species. In accordance with EU Regulation 2015/2283, *G. assimilis*, *L. migratoria*, *T. molitor*, and *Z. morio* were selected for specificity testing. Singleplex PCR resulted in a single 240 bp amplicon for *T. molitor* and a single 120 bp amplicon for *Z. morio* (Figure 3).

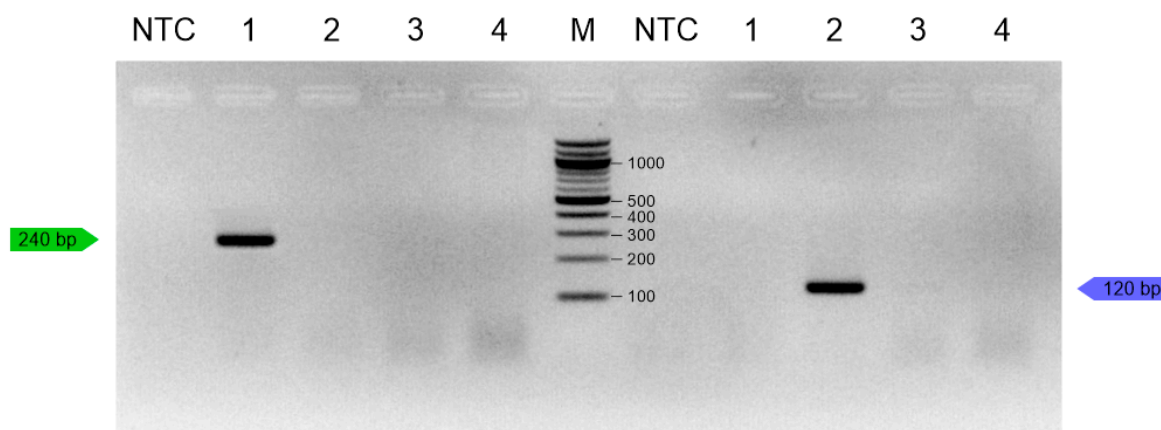


Figure 3. Singleplex specificity in test with 1—*T. molitor* (green line), 2—*Z. morio* (blue line), 3—*G. assimilis*, 4—*L. migratoria*; M—100 bp ladder (NEB, UK); NTC—no-template control.

Duplex PCR was performed using different amounts of DNA from various insect species. In all of the tested combinations, both *Z. morio* and *T. molitor* could be specifically detected in a single reaction (Figure 4).

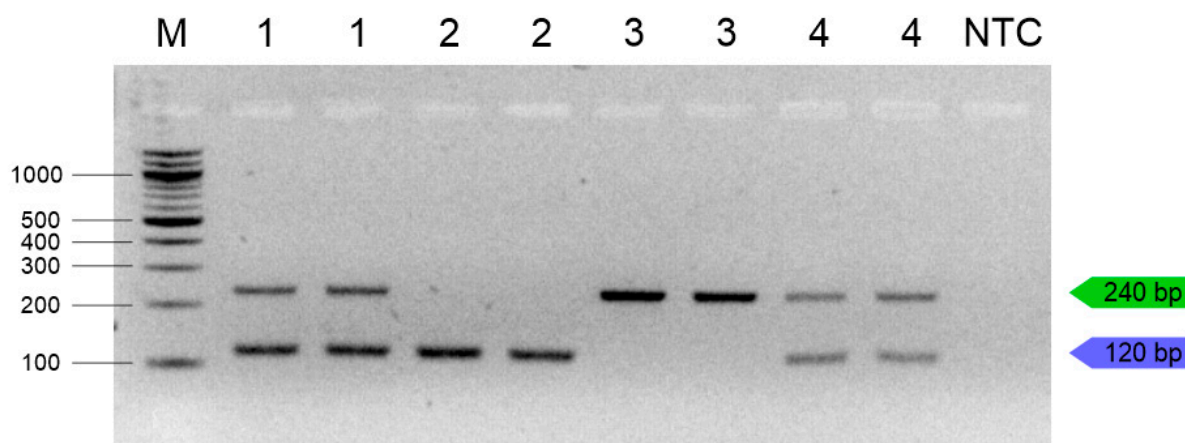


Figure 4. Specificity in multiplex PCR with different combinations DNA samples. 1—0.5 µL DNA *T. molitor* and 0.5 µL *Z. morio*; 2—1 µL *Z. morio*; 3—1 µL *T. molitor*; 4—0.25 µL *Z. morio*, 0.25 µL *T. molitor*, 0.25 µL *G. assimilis*, and 0.25 µL *L. migratoria*; M—100 bp ladder (NEB, UK); NTC—no-template control; *T. molitor* (green line), *Z. morio* (blue line).

3.4. Verification

For specific confirmation, insect-based foods (declared as containing only *T. molitor*) from the Czech food market were used. All products were tested separately for *T. molitor* and *Z. morio* using PCR. Positive detection of *T. molitor* and negative detection of *Z. morio* were observed in all products (Figure 5). This confirms that companies accurately declare the insect species present in their foods.

In pastry breads containing 1–10% insects, detection of the insect content is possible. The baking process did not affect the detection compared to raw insects at these percentages. In baked bread containing a combination of *T. molitor* and *Z. morio*, individual insect species can be specifically detected at levels as low as 1% insect content. This represents a very promising method for identifying counterfeit products (Figure 6).

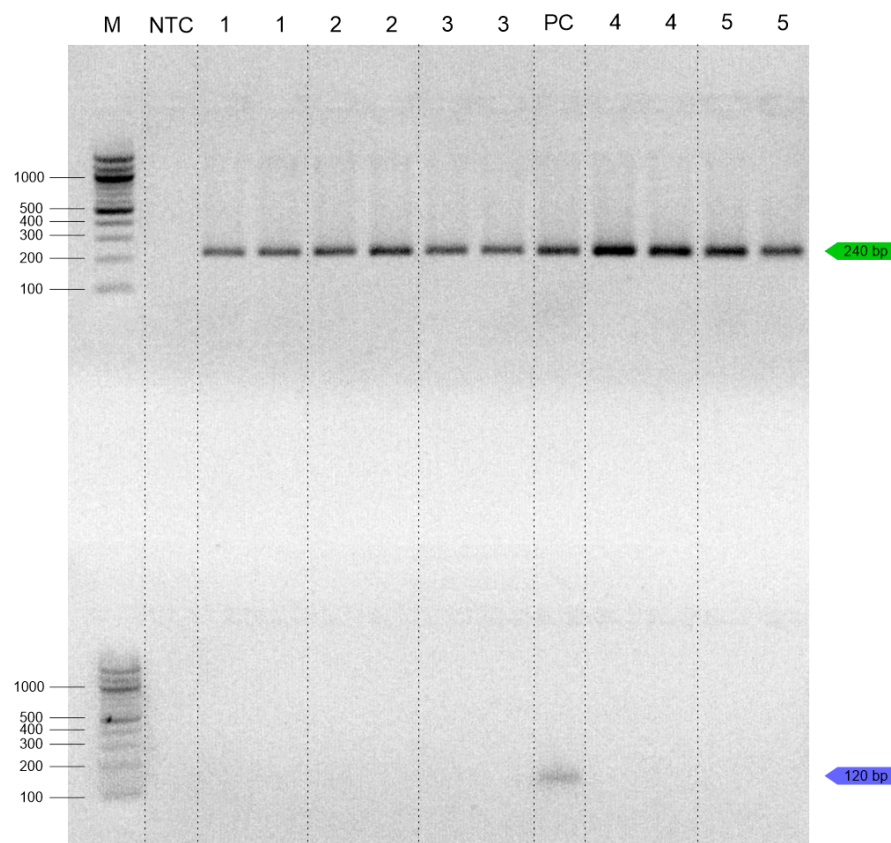


Figure 5. Products with *T. molitor*. 1—prepared boiled instant rice porridge (Farmerio, Svatoňovice, Czech Republic); 2—raw instant rice porridge (Farmerio, Svatoňovice, Czech Republic); 3—low-carb chocolate (Farmerio, Svatoňovice, Czech Republic); 4—dried larvae, “Křupaví červíci” (WormUP, Blansko, Czech Republic) with garlic; 5—dried larvae, “Křupaví červíci” (WormUP, Blansko, Czech Republic), with chili; M—100 bp ladder (NEB, UK); PC—10 ng/μL *T. molitor* (up) and *Z. morio* (down); NTC—no-template control; up (green)—PCR for detection *T. molitor*; down (blue)—PCR for detection *Z. morio*.

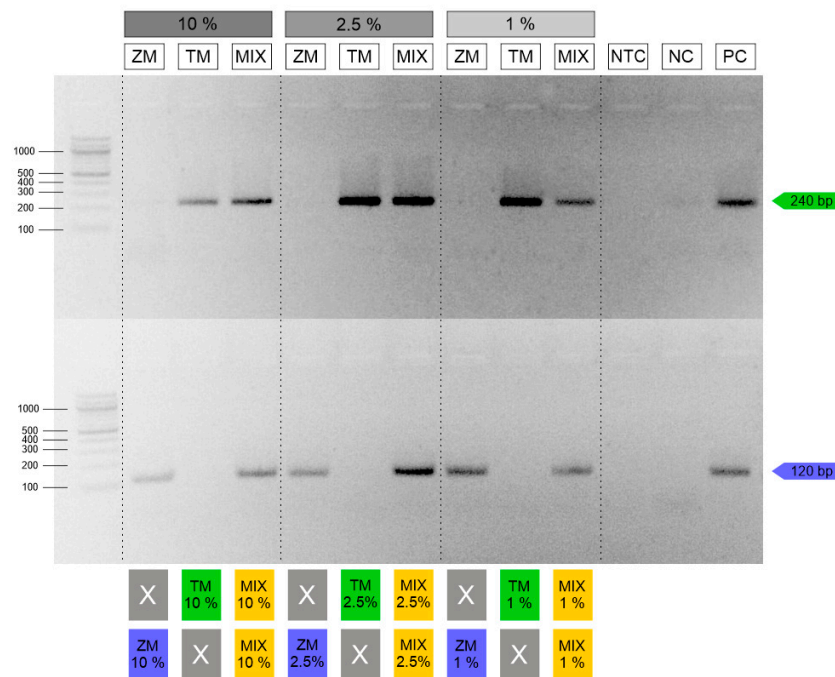


Figure 6. Pastry samples. Up (green) TM—PCR for detection *T. molitor*; down (blue) ZM—PCR for detection *Z. morio*; MIX (yellow)—*T. molitor* and *Z. morio*; PC—10 ng/μL *T. molitor* (up) and *Z. morio* (down); NTC—no-template control; X—not detected.

4. Discussion

Edible insects have recently been used as snacks in Western Europe and as food additives to improve product texture. Fortifying dough with insect meal can alter both its nutritional and rheological properties. It was demonstrated that the substitution of starch with cricket meal in gluten-free baked goods reduced bread hardness and improved dough consistency by enhancing water binding [14]. Similarly, it was found that adding 5% cricket meal to wheat bread dough resulted in higher bread volume and lower crumb stiffness [15].

The rheological properties of gluten-free doughs are generally more complex than those of gluten-containing doughs. Gluten provides the desired baking qualities, such as viscosity and elasticity [16]. However, these properties are suppressed in gluten-free flours due to the absence of gluten, and efforts are made to use improving agents to achieve higher quality and more consumer-acceptable products.

Improving agents can serve various purposes, including speeding up proofing, increasing the volume of products, enhancing elasticity and workability, prolonging shelf life, and reducing production time [17,18]. Edible insect meal can potentially improve the rheological properties of dough, especially gluten-free dough, but the properties may vary depending on the insect species used. Differences in species may also affect price and quality, potentially leading to the substitution (adulteration) of higher-quality insects with cheaper, lower-quality ones. In recent years, instances of fraud have been identified within the edible insect supply chain [5,19].

The legislative process is relatively simpler in Asian countries, while in the EU, only four insect species can be used for human consumption: yellow mealworm (*T. molitor*), migratory locust/grasshopper (*L. migratoria*), grain mould beetle (*Alphitobius diaperinus*), and house cricket (*Acheta domesticus*). These insects are undergoing risk and food safety assessments for the protection of public health by EFSA. One of the safety controls is the assessment of the product's nutritional value, particularly protein content [20]. Different insects can trigger the development of clinical symptoms typical of allergenic reactions in sensitized individuals. This is the next reason why specific detection methods are necessary [21].

Mitochondrial genome analysis was one of the earliest molecular biological methods applied to insect studies. The ribosomal RNA gene (16s and 12s rRNA), NADH5 and NADH4 dehydrogenase gene, and cytochrome c oxidase gene were used for designing specific insect primers [22,23]. Afterward, these genes were used for sequenced insects and new population studies [24,25].

While several modified PCR methods have been described for *T. molitor* detection, cross-amplification remains a challenge. For example, primers targeting the Cadherin gene have been shown to produce false positives for *Z. morio*, *Bombyx mori*, and *Gammarus* sp. Primers targeting the Wingless gene offer improved specificity, but can still cross-react with plant DNA, particularly wheat, a common ingredient in insect-based food and feed products [10,26].

However, existing studies demonstrate the feasibility of PCR for the detection of several edible insect species, but did not address the critical issue of differentiating *T. molitor* from closely related species. Our work provides several novel and practically relevant contributions. First, the duplex PCR developed in this study achieved a detection limit of 100 pg of DNA, confirming high analytical sensitivity suitable for complex food matrices. Second, specificity was rigorously validated not only against *Z. morio*, but also against other insect species relevant to the European novel food context, namely *G. assimilis* and *L. migratoria*. This broader specificity testing minimizes the risk of cross-reactivity and strengthens the robustness of the method. Most importantly, from an applied perspective, the assay was evaluated on real commercial food products, including chocolate,

porridge, and flavored dried larvae. This step bridges the gap between laboratory validation and application, underscoring the method's utility for food control authorities [12,27,28]. According to EU Regulation 2015/2283, specific tests are needed to distinguish between *Z. morio* and *T. molitor* in insect-based food. In this study, we used primers that were previously shown to be highly sensitive and specific in separate PCR detections. This study introduces a multiplex detection method capable of identifying low-level targets, aiming to enhance food product safety and quality control.

5. Conclusions

The multiplex PCR method described in this study is robust and sensitive, making it suitable for its intended purpose: the specific detection of edible insects in insect-based food. The two primer pairs enabled the discrimination of *T. molitor* and *Z. morio*. The sensitivity and specificity of the method were confirmed by the robust detection of target DNA. While inhibitors present in the food or insect matrix can potentially affect DNA extraction, the developed methods were robust and consistently detected *T. molitor* and *Z. morio* target DNA. Further testing on processed foods under more extreme conditions is necessary to support routine food control applications. For these reasons, the method is suitable for detecting the presence/absence of *T. molitor* and *Z. morio* DNA in food, even in small amounts, as well as in insect-based food products.

Author Contributions: Conceptualization, D.S. and V.M.; methodology, D.S. and A.A.; validation, A.A., V.M. and D.S.; formal analysis, M.A.; investigation, J.M.; resources, P.B.; data curation, P.B.; writing—original draft preparation, D.S.; writing—review and editing, J.M., S.Š. and V.Č.; visualization, D.S. and P.B.; supervision, J.M.; project administration, J.M.; funding acquisition, J.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Internal research project of Tomas Bata University in Zlin (No. IGA/FT/2025/003 and RVO/CEBIA/2024/001).

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy restrictions.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Mlček, J.; Adámková, A.; Adámek, M.; Borkovcová, M.; Bednářová, M.; Kouřimská, L.; Hlobilová, V. Selected aspects of edible insect rearing and consumption—A review. *Czech J. Food Sci.* **2021**, *39*, 149–159. [\[CrossRef\]](#)
2. Premalatha, M.; Abbasi, T.; Abbasi, T.; Abbasi, S.A. Energy-efficient food production to reduce global warming and ecodegradation: The use of edible insects. *Renew. Sustain. Energy Rev.* **2011**, *15*, 4357–4360. [\[CrossRef\]](#)
3. Pimentel, D.; Pimentel, M. Sustainability of meat-based and plant-based diets and the environment. *Am. J. Clin. Nutr.* **2003**, *78*, 660S–663S. [\[CrossRef\]](#)
4. Oonincx, D.G.; De Boer, I.J. Environmental impact of the production of mealworms as a protein source for humans—a life cycle assessment. *PLoS ONE* **2012**, *7*, e51145. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Traynor, A.; Burns, D.T.; Wu, D.; Karoonuthaisiri, N.; Petchkongkaew, A.; Elliott, C.T. An analysis of emerging food safety and fraud risks of novel insect proteins within complex supply chains. *npj Sci. Food* **2024**, *8*, 7. [\[CrossRef\]](#)
6. Siemianowska, E.; Kosewska, A.; Aljewicz, M.; Skibniewska, K.A.; Polak-Juszczak, L.; Jarocki, A.; Jedras, M. Larvae of mealworm (*Tenebrio molitor* L.) as European novel food. *Agric. Sci.* **2013**, *4*, 287–291. [\[CrossRef\]](#)
7. Weiner, A.; Kwiatek, K. Laboratory experience with the microscopic method for the detection of insects in poultry feeds. *J. Vet. Res.* **2002**, *66*, 403. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Delgado Calvo-Flores, L.; Garino, C.; Moreno, F.J.; Broll, H. Insects in food and their relevance regarding allergenicity assessment. *EFSA J.* **2022**, *20*, e200909. [\[CrossRef\]](#)

9. Pospiech, M.; Pečová, M.; Bartlová, M.; Javůrková, Z.; Kopecká, A.; Šebelová, K.; Hajšlová, J. Development of Indirect Sandwich ELLA for Detection of Insects in Food. *Appl. Sci.* **2024**, *14*, 10794. [\[CrossRef\]](#)
10. Debode, F.; Marien, A.; Gérard, A.; Francis, F.; Fumière, O.; Berben, G. Development of real-time PCR tests for the detection of *Tenebrio molitor* in food and feed. *Food Addit. Contam.* **2017**, *34*, 1421–1426. [\[CrossRef\]](#)
11. Doyle, J. DNA protocols for plants. In *Molecular Techniques in Taxonomy*; NATO ASI Series H: Cell biology; Hewitt, G.M., Johnston, A.W., Young, J.P.W., Eds.; Springer: Berlin/Heidelberg, Germany, 1991; Volume 57, pp. 283–293.
12. Tramuta, C.; Gallina, S.; Bellio, A.; Bianchi, D.M.; Chiesa, F.; Rubiola, S.; Romano, A.; Decastelli, L. A set of multiplex polymerase chain reactions for genomic detection of nine edible insect species in foods. *J. Insect Sci.* **2018**, *18*, 3. [\[CrossRef\]](#)
13. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [\[CrossRef\]](#)
14. Kowalczewski, P.L.; Walkowiak, K.; Masewicz, Ł.; Bartzak, O.; Lewandowicz, J.; Kubiak, P.; Baranowska, H.M. Gluten-free bread with cricket powder—Mechanical properties and molecular water dynamics in dough and ready product. *Foods* **2019**, *8*, 240. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Roncolini, A.; Milanović, V.; Cardinali, F.; Osimani, A.; Garofalo, C.; Sabbatini, R.; Clementi, F.; Pasquini, M.; Mozzon, M.; Foligni, R.; et al. Protein fortification with mealworm (*Tenebrio molitor* L.) powder: Effect on textural, microbiological, nutritional and sensory features of bread. *PLoS ONE* **2019**, *14*, e0211747. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Cappelli, A.; Oliva, N.; Cini, E. A systematic review of gluten-free dough and bread: Dough rheology, bread characteristics, and improvement strategies. *Appl. Sci.* **2020**, *10*, 6559. [\[CrossRef\]](#)
17. Mondal, A.; Datta, A.K. Bread baking—A review. *J. Food Eng.* **2008**, *86*, 465–474. [\[CrossRef\]](#)
18. Burešová, I.; Kráčmar, S.; Dvořáková, P.; Středa, T. The relationship between rheological characteristics of gluten-free dough and the quality of biologically leavened bread. *J. Cereal Sci.* **2014**, *60*, 271–275. [\[CrossRef\]](#)
19. Moore, J.C.; Spink, J.; Lipp, M. Development and application of a database of food ingredient fraud and economically motivated adulteration from 1980 to 2010. *J. Food Sci.* **2012**, *77*, R118–R126. [\[CrossRef\]](#)
20. EFSA NDA Panel. Safety of dried yellow mealworm (*Tenebrio molitor* larva) as a novel food pursuant to Regulation (EU) 2015/2283. *EFSA J.* **2021**, *19*, 6343. [\[CrossRef\]](#)
21. German Federal Institute for Risk Assessment (BfR); National Reference Laboratory for Animal protein in Feed; NRL-AP; Garino, C.; Zagon, J.; Braeuning, A. Insects in food and feed—allergenicity risk assessment and analytical detection. *EFSA J.* **2019**, *17*, e170907. [\[CrossRef\]](#)
22. Kambhampati, S.; Smith, P.T. PCR primers for the amplification of four insect mitochondrial gene fragments. *Insect Mol. Biol.* **1995**, *4*, 233–236. [\[CrossRef\]](#)
23. Ji, Y.J.; Zhang, D.X.; He, L.J. Evolutionary conservation and versatility of a new set of primers for amplifying the ribosomal internal transcribed spacer regions in insects and other invertebrates. *Mol. Ecol. Notes* **2003**, *3*, 581–585. [\[CrossRef\]](#)
24. Pons, J.; Barraclough, T.G.; Gomez-Zurita, J.; Cardoso, A.; Duran, D.P.; Hazell, S.; Kamoun, S.; Sumlin, W.D.; Vogler, A.P. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Syst. Biol.* **2006**, *55*, 595–609. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Hillinger, S.; Saeckler, J.; Domig, K.J.; Dobrovolny, S.; Hocheegger, R. Development of a DNA Metabarcoding Method for the Identification of Insects in Food. *Foods* **2023**, *12*, 1086. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Natonek-Wiśniewska, M.; Krzyścin, P.; Koseniuk, A. Qualitative and Quantitative Detection of Mealworm DNA in Raw and Commercial Food Products Using Real-Time PCR. *Genes* **2022**, *13*, 1400. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Köppel, R.; Schum, R.; Habermacher, M.; Sester, C.; Piller, L.E.; Meissner, S.; Pietsch, K. Multiplex real-time PCR for the detection of insect DNA and determination of contents of *Tenebrio molitor*, *Locusta migratoria* and *Achaeta domestica* in food. *Eur. Food Res. Technol.* **2019**, *245*, 559–567. [\[CrossRef\]](#)
28. Villa, C.; Moura, M.B.; Teixeira, C.S.; Costa, J.; Mafra, I. Monitoring Yellow Mealworm (*Tenebrio molitor*) as a Potential Novel Allergenic Food: Effect of Food Processing and Matrix. *Nutrients* **2023**, *15*, 482. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Na tomto místě se v plné verzi habilitační práce nachází text příslušné publikace. Ten však není součástí této veřejně přístupné verze z důvodu licenčních podmínek vydavatele. Níže je k dispozici bibliografická citace a abstrakt článku. Plný text lze nalézt v placeném režimu na webových stránkách časopisu.

Strejckova, M., Bohata, A., Olsan, P., Havelka, Z., Kriz, P., Beran, P., Bartos, P., Curn, V., & Spatenka, P. (2018). Enhancement of the Yield of Crops by Plasma and Using of Entomopathogenic and Mycoparasitic Fungi: From Laboratory to Large-Field Experiments. *Journal of Biomaterials and Tissue Engineering*, 8(6), 829–836.

Abstrakt

Recent trends in crop production tend to eliminate the use of pesticides. Instead, environmentalfriendly approaches based on biological and physical methods become more popular in plant protection. In this study we investigate the benefits of such an approach based on seed treatment by plasma discharge in together with *Metarhizium anisopliae* and *Trichoderma virens* coating as an alternative to the chemical dressing of seeds. The presented study covers both laboratory and field experiments, in which various seed treatments of spring barley, winter oilseed rape and spring poppy are compared. The results indicate, that plasma treatment and combined plasma and biological treatment of seeds have a high potential in seed production technology as added-value processes alternative to the standard chemical dressing of seeds achieved by a stimulation of seed germination, early growth of seedlings, canopy development, and they also positively influence the yield.

Na tomto místě se v plné verzi habilitační práce nachází text příslušné publikace. Ten však není součástí této veřejně přístupné verze z důvodu licenčních podmínek vydavatele. Níže je k dispozici bibliografická citace a abstrakt článku. Plný text lze nalézt v placeném režimu na webových stránkách časopisu.

Udoh, O. E., Abu, N. E., Egonu, S. N., & Beran, P. (in press). Evaluation of protein profile of ten Nigerian underutilized legumes using SDS–PAGE. African Journal of Agricultural Research.

Abstrakt

Legumes belong to the family *Fabaceae* and most of them produce seeds within a pod. This study aimed to reassess the taxonomic status of ten selected legumes, examine their genetic similarity and illustrate a potential evolutionary pathway. In this study, protein profiles for seed storage in ten underutilized legumes were analysed by comparing their banding patterns using Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis technique. Protein was extracted from these selected seed samples and crude protein content were evaluated. Upon resolution on 10% SDS polyacrylamide gels, the total seed storage protein of the investigated plants displayed distinct banding patterns. The result showed that some of the seed samples contain higher protein content than the others. Each band contained in each of the seed samples represent a particular protein. A comparison between species was conducted using hierarchical cluster analysis and calculating the genetic resemblance. It was observed that *Vigna subterranea* and *Vigna unguiculata* are genetically very close compared to other species studied and this justifies their placement in the same subfamily. Given the variance in genetic diversity among the plant species under study, it is possible to infer that SDS-PAGE is still a vital tool for genetic investigation.

RESEARCH

Open Access



Novel *gp60* subtypes and genetic diversity of *Cryptosporidium felis* in domestic and stray cats from Central Europe

Veronika Zikmundová^{1,2}, Eliška Höflerová², Nikola Holubová¹, Marta Kicia³, Matúš Rajský^{4,5}, Pavel Beran², Bohumil Sak¹ and Martin Kváč^{1,2*}

Abstract

Background *Cryptosporidium felis*, a host-specific protozoan with possible zoonotic potential, is a common parasite of cats (*Felis catus*). To date, there have been few studies on the molecular subtyping of *C. felis* worldwide, and no have been conducted in Central Europe. The aim of this study was to analyse the prevalence and genetic variability of *C. felis* in domestic and stray cats in Central Europe, particularly in the Czech Republic, Poland and Slovakia.

Methods Faecal samples were collected from domestic and stray cats. The presence of *Cryptosporidium* spp. was analysed by light microscopy with aniline-carbol-methyl violet staining and polymerase chain reaction (PCR)/sequencing of the small subunit rRNA gene (18S rDNA). All PCR-positive samples were further subtyped using PCR and sequencing of the 60 kDa glycoprotein gene (*gp60*). A chi-square test and odds ratio (OR) analysis were used to compare infection rates and assess the risk of *C. felis* infection between domestic and stray cats.

Results A total of 711 faecal samples were collected – 350 from domestic cats and 361 from stray cats. The overall infection rate of *Cryptosporidium* spp. was 4.5% (32/711), with stray cats being significantly more frequently infected (6.7%) than domestic cats (2.3%). Oocysts of *Cryptosporidium* spp. were not detected microscopically in any of the samples. There were no significant differences between the infection rates in the three countries. All isolates were identified as *C. felis*, and analysis of the *gp60* gene revealed five different subtypes, all belonging to the XIXa subtype family of *C. felis*. These subtypes formed five well-supported phylogenetic clusters, none of which had been previously reported worldwide. Only one subtype was found in domestic cats, whereas all five subtypes were found in stray cats. The subtypes identified in stray cats showed a clear geographical distribution in the study region.

Conclusions The results of this study extend our knowledge of the genetic variability of *C. felis* and indicate a possible geographical distribution of the detected subtypes. The significance of the observed genetic variability in terms of geographical distribution, host specificity and zoonotic potential remains unclear and requires further investigation.

Keywords *gp60*, PCR, Cat, *Cryptosporidium*, Subtyping, Geographic distribution

*Correspondence:

Martin Kváč

kvac@paru.cas.cz

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Cryptosporidium (Apicomplexa) is a genus of obligate parasitic protists that predominantly infect the gastrointestinal and respiratory tracts of vertebrates such as mammals, birds, reptiles, fish and amphibians [1–3]. A clinical manifestation of cryptosporidial infection, known as cryptosporidiosis, is characterised by watery diarrhoea, abdominal pain and nausea. The severity of infection depends on the *Cryptosporidium* species and the host species, including its age and immune status. In general, infection is more severe and sometimes fatal in juveniles, non-specific hosts and immunocompromised individuals [4–6].

The domestic cat (*Felis catus*) is one of the most widespread pet species and a favourite companion animal in many households worldwide. According to statistics, 350–370 million cats are kept as pets worldwide and the number of stray cats is estimated at 480 million [7]. Keeping cats has many benefits for humans – they improve mental well-being, reduce stress, promote socialisation and can have a positive effect on cardiovascular health [8]. However, the potential health risks must also be considered. Besides the usual injuries to the owner from scratching or biting and allergic reactions [9], cats can be carriers of various zoonotic parasite infections such as toxoplasmosis, giardiasis, or cryptosporidiosis [10–12].

Cats are primarily parasitised by the host-specific species *Cryptosporidium felis*, but other species and genotypes of *Cryptosporidium* have also been detected in cats, namely *Cryptosporidium muris*, *Cryptosporidium parvum*, *Cryptosporidium baileyi*, *Cryptosporidium ryanae*,

Cryptosporidium hominis, *Cryptosporidium* sp. rat genotype III and IV [10, 13–16]. *Cryptosporidium felis* is classified as a species with zoonotic potential, with low to moderate prevalence in immunocompetent individuals or in populations with limited access to sanitation and clean water [17].

In 2020, Rojas-Lopez et al. [18] developed a molecular method for subtyping *C. felis* on the basis of the analysis of the gene encoding the 60 kDa glycoprotein (*gp60*). Phylogenetic analyses of this gene revealed the existence of five subtypes of *C. felis* (designated XIXa–XIXe), which differ both in their host specificity and geographical distribution [18–20]. Nevertheless, the amount of available data is still quite limited. Especially in the Europe, the genetic variation of *C. felis* is poorly documented, with few published data [10, 18–21].

The aim of the present study is to cover the knowledge gaps on the geographical distribution and genetic variability of *C. felis* in Central Europe by molecular subtyping based on the *gp60* gene.

Methods

Area and specimens studied

For this work, faecal samples were collected from domestic cats kept indoors, and from stray cats in animal shelters or from cats near farm buildings, especially cattle and pig stables. The samples were collected in the Czech Republic, Slovakia and the southern part of Poland (Fig. 1). To avoid re-sampling the same stray cat, samples were collected from the ground after the animal had defecated, within 1 day/location. No domestic cat was

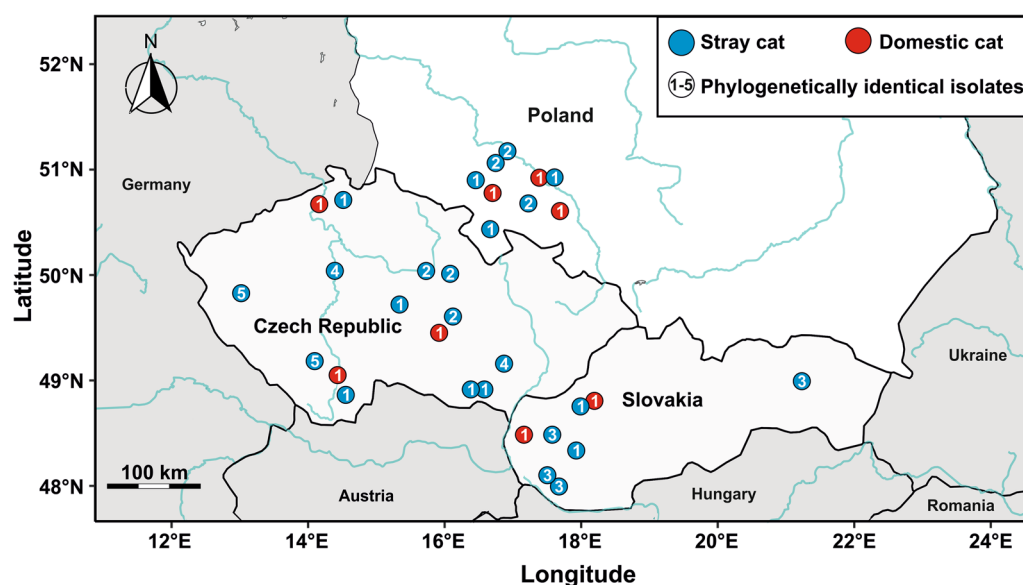


Fig. 1 Map of the geographical distribution of phylogenetically identical *Cryptosporidium felis* subtypes (cluster) in domestic and stray cats in the Czech Republic, Slovakia and Poland, based on genotyping of the gene encoding the 60 kDa glycoprotein (*gp60*)

sampled twice. Each faecal sample was placed in a sterile plastic vial labelled with the sample ID and stored at 4–8 °C without fixative until laboratory processing. A smear was taken from each faecal sample, stained with aniline-carbol-methyl violet and examined using a light microscope to detect *Cryptosporidium* spp. oocysts [22].

Molecular analyses

Total genomic DNA (gDNA) was extracted from 200 mg of faeces by bead disruption at 5.5 m/s for 60 s, using 0.5 mm glass beads in a FastPrepR24 instrument (MP Bio-medicals, CA, USA), followed by DNA isolation using a commercial kit according to the manufacturer's instructions (Exgene™ Stool DNA mini, Genially Biotechnology Co. Ltd., Seoul, Korea). The purified DNA was stored at –20°C before being used for PCR. A nested PCR approach was used to amplify a partial region of the small subunit rRNA gene (18S rDNA) [23] and the *gp60* gene [18]. The PCR reaction was carried out in a total volume of 20 µl. The reaction mixture contained 2 µl of gDNA as template, or 2 µl of the primary PCR product in case of a secondary reaction, 10 µl of 2×AmpONE™ HS-Taq premix (GeneAll), 200 nM of each primer (forward and reverse) and molecular grade water to the total volume. *Cryptosporidium serpentis* (for 18S rDNA) and *C. felis* (for *gp60*) DNA and molecular grade water were used as positive and negative controls, respectively. The secondary PCR products were detected by agarose gel electrophoresis, stained with ethidium bromide and isolated using the GenElute Gel Extraction Kit (Sigma, St. Louis, MO, USA). The purified secondary products were sequenced bidirectionally at a commercial company (SeqMe s.r.o, Dobříš, Czech Republic) using Sanger sequencing. Each positive sample was independently sequenced twice.

Phylogenetic analyses

The nucleotide sequences obtained for each gene in this study were processed and edited using ChromasPro v2.4.1 (Technelysium Pty Ltd., South Brisbane, Australia). Sequence alignments were performed both between samples and with reference sequences from GenBank, using the MAFFT online platform (version 7) with automatic alignment mode selection (<http://mafft.cbrc.jp/alignment/software/>). Phylogenetic analyses were performed using the MEGA X software [24, 25], which was also used to determine the best-fitting DNA/protein evolutionary models based on the Bayesian information criterion (BIC). The maximum likelihood (ML) method was applied to generate phylogenetic tree. The General Time Reversible model [26] was used for *gp60* tree. Branch support was evaluated using 1,000 bootstrap repetitions. The resulting phylograms were generated in

MEGAX and then manually refined in CorelDraw X7. All newly generated sequences of *gp60* genes were deposited in GenBank under the following accession numbers [PV817847–PV817854].

Statistical analysis

Statistical analysis was performed using chi-square tests with Yates correction in EpiInfo™ 7.2.7 (CDC, Atlanta, USA), to compare the frequency of *C. felis* occurrence between domestic and stray cats, and the distribution in each country (Czech Republic, Slovakia and Poland). A *P*-value < 0.05 was considered statistically significant.

Results

Occurrence and prevalence of *Cryptosporidium* spp. in cats

A total of 711 cat samples were collected – 350 from domestic cats and 361 from stray cats (Table 1). Of the 711 samples analysed, no oocysts of *Cryptosporidium* spp. were detected microscopically in any sample. However, the results of the molecular analyses for amplification of 18S rDNA showed the presence of specific *Cryptosporidium* spp. DNA in 32 samples (Table 1).

All PCR products were successfully sequenced, and the results confirmed the presence of *C. felis* in all samples. The overall prevalence of *C. felis* in the analysed regions was 4.5%. The highest prevalence was found in the Czech Republic (5.1%), followed by Slovakia (4.2%) and Poland (4.0%). Of the 350 domestic and 361 stray cats, 8 (2.3%) and 24 animals (6.7%), respectively, were positive for *C. felis*. Domestic cats were 3.04 times less likely to be parasitised by *C. felis* than stray cats ($P = 0.0086$, $\chi^2 = 6.8863$, $df = 1$, $OR = 3.04$).

Table 1 Numbers of tested and positive domestic and stray cats (*Felis catus*) for the presence of *Cryptosporidium felis* in the Czech Republic, Slovakia and Poland

Country	Cat type	Examined	Positivity	
			Microscopy	PCR
Czech Republic	Domestic	141	0	3
	Stray	156	0	12
	Subtotal	297	0	15
Slovakia	Domestic	89	0	2
	Stray	102	0	6
	Subtotal	191	0	8
Poland	Domestic	120	0	3
	Stray	103	0	6
	Subtotal	223	0	9
Total		711	0	32

Molecular subtyping of *Cryptosporidium felis*

All 32 PCR-positive samples were successfully genotyped for the *gp60* locus. Analysis of the *gp60* gene identified five subtypes of *C. felis*, which were assigned to subtype family XIXa (Fig. 2). When analysed phylogenetically, the subtypes formed five separate clusters with high bootstrap support. None of the subtypes detected in this study were identical to subtypes previously identified in cats or humans worldwide (Fig. 2). All isolates from domestic cats were identical to each other and belonged exclusively to cluster no. 1 (Fig. 2), which also included isolates from stray cats (Fig. 2). These isolates were phylogenetically related to isolate *C. felis* 7378 (GenBank MT458675), which originated from a cat in Slovakia (the type of cat is unknown), and to six other isolates obtained in this study from stray cats from Poland and the Czech Republic, which formed a separate cluster (no. 2). The analysis of the occurrence of the isolates in the study area indicates a geographical distribution pattern (Fig. 1). While isolates belonging to cluster no. 1 were detected in all three countries, the Czech Republic, Slovakia and Poland, isolates belonging to cluster no. 2 were found only in the Czech Republic and Poland. Cluster no. 3 was found exclusively in Slovakia, and clusters no. 4 and 5 only in the Czech Republic (Fig. 1 and 2).

Sequence analysis of the individual isolates revealed differences in sequence length, mainly due to the presence of repetitive sequences (Table 2). The sequence designated R1, which contains a 33-bp repetitive stretch of 5' CCA CCT AGT GGC GGT GGC GTG TCC CCT GCT 3' located approximately between nucleotides 450–530 in the sequence alignment, was present in only one copy in the isolates belonging to cluster no. 1 (Table 2; Fig. 2). A repetitive sequence designated R2, containing a 39-bp stretch of the tandem repeat 5' AGC ACA ACT GCG GCT ACA GCG AGC ACT GCG AGT TCG ACA 3' at a position between nucleotides 770–910, was detected in the sequences of eight isolates in repeats 1–3. Isolates with one tandem repeat belonged exclusively to cluster no. 3, those with two repeats to cluster no. 4 and those with three repeats to cluster no. 5 (Fig. 2; Table 2). The number of GTT triplet repeats at the end of the sequence, between nucleotide positions 1143–1154, was between two and four in the isolates analysed. Isolates belonging to clusters nos. 1 and 2 contained two GTT repeats, whereas isolates from clusters nos. 3–5 always had four repeats (Fig. 2; Table 2).

Discussion

This study deals with the occurrence of *C. felis* in cats in the Czech Republic, Slovakia and Poland, focusing on the intraspecific variation within the *gp60* gene.

Cryptosporidium spp. infections in cats are relatively common, and their prevalence in the population varies between different studies and countries, ranging from 0.6% to 40.8% [16]. In the present study, 4.5% of the 711 cats examined were classified as positive. This prevalence represents a relatively low incidence. Previous studies conducted in the Czech Republic, Slovakia and Poland found a similar prevalence of 4.4% [27]. Comparable prevalence rates were found in other European countries, with the exception of Germany, where an overall prevalence of 21.2% was observed [16].

The phylogenetic analyses showed that the cats included in the study were infected only with *C. felis*. These results are in agreement with those of other authors, who also either exclusively detected this species in their studies or dominated the occurrence of *C. felis* [10, 14, 19, 21, 28–33]. Sequence analysis of all 32 samples positive for *Cryptosporidium*-specific DNA showed no evidence of mixed infections, and all phylograms were clean. In a previous study conducted in the same area, no other *Cryptosporidium* species were detected in either stray or domestic cats [27]. The species *C. muris*, *C. parvum*, *C. ryanae*, *Cryptosporidium* sp. rat genotype III and *Cryptosporidium* sp. rat genotype IV, which have been detected together with *C. felis* in previous studies, mostly represent a group of species that are not host-specific to cats [2, 34]. The species and genotypes such as *C. muris*, *Cryptosporidium* sp. rat genotype III and *Cryptosporidium* sp. rat genotype IV are specific to mice and rats, which are a natural food source for cats [35, 36]. It is therefore not surprising that the specific DNA of these *Cryptosporidium* species is occasionally detected in cat faeces using sensitive molecular methods. Similarly, *C. muris* and *C. tyzzeri* have been detected in slurry and pig faeces; however, it was later experimentally confirmed that pigs are not susceptible to these *Cryptosporidium* species [37, 38]. The sporadic occurrence of *C. ryanae* in cats can be explained in a similar way. This *Cryptosporidium* species frequently infects cattle [39], it is likely that cats living on or near farms may become contaminated with cattle faeces.

Our findings, along with other studies, show that stray cats are more likely to be infected with *Cryptosporidium* spp. than domestic cats. In our dataset, the prevalence of *C. felis* infection was higher in stray cats (6.7%) than in domestic cats (2.3%). This is consistent with our previous study, in which 7.4% of stray cats were *C. felis*-positive, compared with only 0.8% in domestic cats [27]. Other studies investigating the influence of cat origin on the prevalence of *Cryptosporidium* spp. have reported similar trends. Yang et al. [15] showed that cats from refuge centres (13.4%) were more frequently infected than privately owned cats (7.1%). Similarly, Xu et al. [40] reported

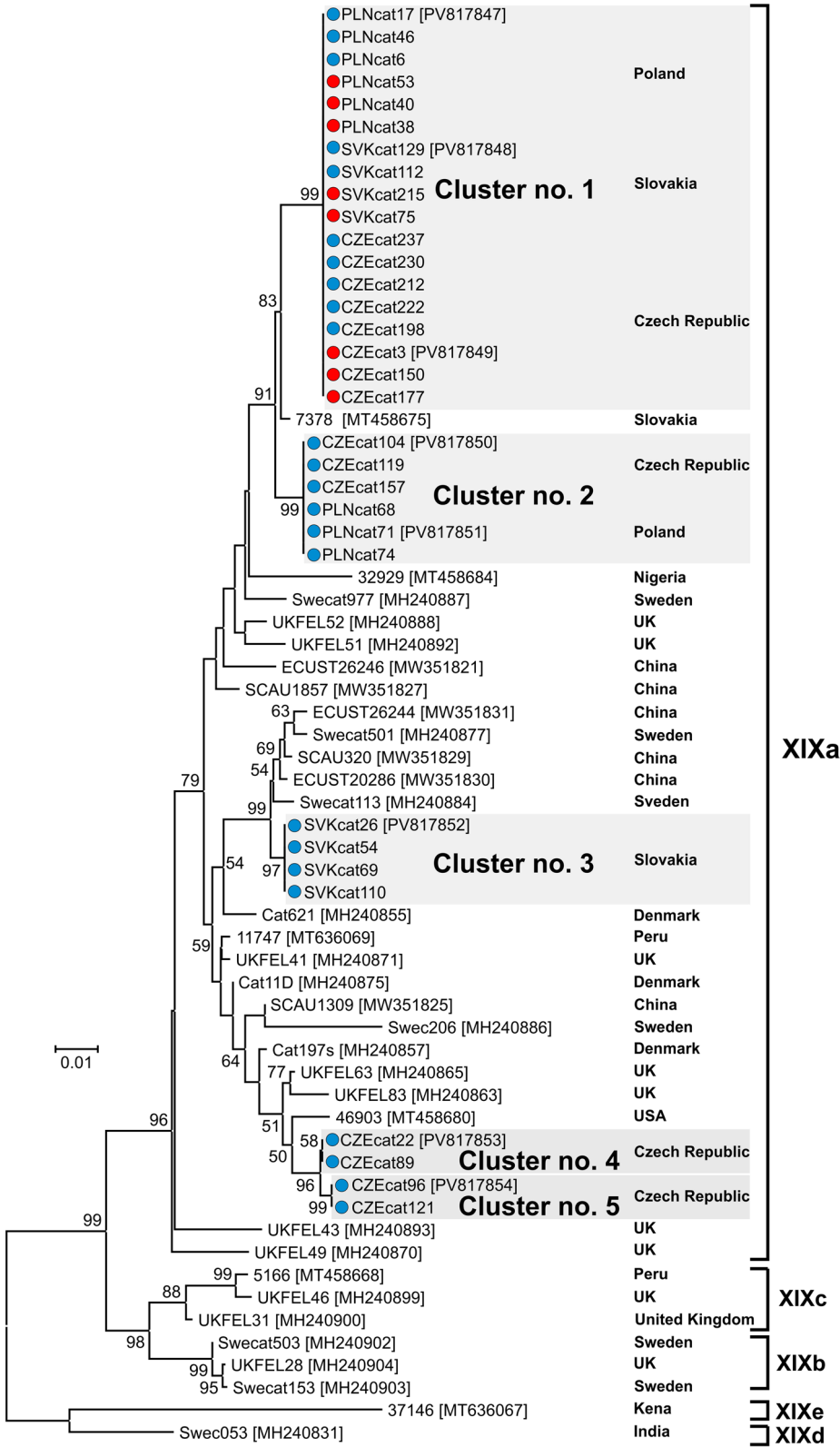


Fig. 2 Phylogenetic relationships between *Cryptosporidium felis* isolates detected in this study and other *C. felis* subtypes, based on the partial sequence of the 60 kDa glycoprotein gene (*gp60*). Maximum likelihood analysis was performed using the General Time Reversible model. Numbers in the nodes represent bootstrap values for nodes that received more than 50% support in the bootstrap test after 1000 replicates. The scale bar is included in the tree. Isolates obtained in this study are highlighted in grey; red dots represent domestic and blue dots represent stray cats

Table 2 The *Cryptosporidium felis* isolates detected in this study, their subtype identity and the copy numbers of major tandem repeats in the *gp60* gene

Country	Cat type	Isolate ID	Genotyping at <i>gp60</i>				Cluster
			Subtype	R1	R2	GGT	
Czech Republic	Domestic	CZEcat3	XIXa	1	–	2	1
	Domestic	CZEcat150	XIXa	1	–	2	1
	Domestic	CZEcat177	XIXa	1	–	2	1
	Stray	CZEcat22	XIXa	–	2	4	4
	Stray	CZEcat89	XIXa	–	2	4	4
	Stray	CZEcat96	XIXa	–	3	4	5
	Stray	CZEcat104	XIXa	–	–	2	2
	Stray	CZEcat119	XIXa	–	–	2	2
	Stray	CZEcat121	XIXa	–	3	4	5
	Stray	CZEcat157	XIXa	–	–	2	2
	Stray	CZEcat198	XIXa	1	–	2	1
	Stray	CZEcat212	XIXa	1	–	2	1
	Stray	CZEcat222	XIXa	1	–	2	1
	Stray	CZEcat230	XIXa	1	–	2	1
	Stray	CZEcat327	XIXa	1	–	2	1
Slovakia	Domestic	SVKcat75	XIXa	1	–	2	1
	Domestic	SVKcat215	XIXa	1	–	2	1
	Stray	SVKcat26	XIXa	–	1	4	3
	Stray	SVKcat54	XIXa	–	1	4	3
	Stray	SVKcat69	XIXa	–	1	4	3
	Stray	SVKcat110	XIXa	–	1	4	3
	Stray	SVKcat112	XIXa	1	–	2	1
	Stray	SVKcat129	XIXa	1	–	2	1
Poland	Domestic	PLNcat38	XIXa	1	–	2	1
	Domestic	PLNcat40	XIXa	1	–	2	1
	Domestic	PLNcat53	XIXa	1	–	2	1
	Stray	PLNcat6	XIXa	1	–	2	1
	Stray	PLNcat17	XIXa	1	–	2	1
	Stray	PLNcat46	XIXa	1	–	2	1
	Stray	PLNcat68	XIXa	–	–	2	2
	Stray	PLNcat71	XIXa	–	–	2	2
	Stray	PLNcat74	XIXa	–	–	2	2

Numbers of the clusters correspond to the clusters in Figs. 1 and 2

a higher prevalence in cats from animal shelters (7.5%) compared with cats from pet shops (3.8%).

Genotyping of *Cryptosporidium* spp. using the gene encoding *gp60* is frequently used to study intraspecific variation, host specificity and pathogenicity of subtypes within individual taxa. The *gp60* gene exhibits a high degree of polymorphism at both inter- and intraspecific levels, which enables the classification of individual isolates into specific families and subtypes within these families [18, 41–43].

Similar to other species, in particular *C. parvum*, *C. hominis* or *Cryptosporidium meleagridis*, a variation of

the *gp60* gene has been described in *C. felis*, comprising five allelic families with dozens of subtypes [19]. In our study, we detected only isolates belonging to allele family XIXa, the most widespread allele family of *C. felis* worldwide. Previous studies suggest a possible geographical splitting of certain *C. felis* subtypes based on the occurrence of genetically identical clusters in geographical areas [19].

In line with these findings, our study not only detected new subtypes not previously reported, but also identified three clusters that appear to be geographically specific. However, it should be emphasised that these findings are

based on limited data, in contrast to the well-studied species *C. parvum* or *C. hominis*. Further studies with larger samples size and in other countries are needed to draw more robust conclusions.

The significance of the diversity of *C. felis* subtypes in relation to pathogenicity – as described for *C. parvum* or *C. hominis*, for example – remains unclear [44–47]. To date, no studies have demonstrated an influence of a specific subtype on the clinical course of *C. felis* infection in cats or humans. In our study, we observed no clinical signs of cryptosporidiosis in any of the positive cats. However, it should be noted that all cats involved in the study were adults, which is consistent with previous findings that adult and healthy cats shed only small amounts of oocysts into the environment after infection [15, 48–51]. This is also consistent with our microscopically negative results, as the low level of oocyst excretion is below the detection limit of conventional staining methods, whereas PCR analyses detected specific *C. felis* DNA in certain samples.

Previous studies have shown that the subtype families XIXb, XIXd and XIXe have been detected exclusively in humans, whereas some subtypes from the families XIXa and XIXc have zoonotic potential [18–21]. Similar differences in the host specificity among subtype families and their subtypes have also been described for *C. parvum* or *Cryptosporidium equi*. For example, the subtypes of subtype family IIc of *C. parvum* are mainly found in humans, with no known natural hosts other than hedgehogs [52, 53]. A similar situation was observed with *C. equi*. While the subtypes of subtype family VIa occur exclusively in horses, donkeys and other equids, infections with subtypes from families VIIb and VIc are almost exclusively associated with humans [54]. Since new subtypes of subtype family XIXa were detected in the countries investigated, and no cases of *C. felis* infections in humans have been reported from these areas, the zoonotic potential of these subtypes of family XIXa cannot yet be assessed.

Conclusions

This study expands the current knowledge on the intraspecific diversity and distribution of *C. felis* in Central Europe. The detection of new subtypes of subtype family XIXa and evidence of a possible geographical clustering emphasise the need for further studies to clarify the zoonotic potential and biological significance. The higher prevalence in stray cats compared with domestic cats emphasises the role of lifestyle and environmental exposure in transmission dynamics. Finally, the discrepancy between microscopically negative and PCR-positive results emphasises the importance of molecular methods for reliable detection.

Abbreviations

BIC	Bayesian information criterion
DNA	Deoxyribonucleic acid
gDNA	Genomic DNA
gp60	60 KDa glycoprotein gene
ML	Maximum likelihood
PCR	Polymerase chain reaction
rRNA	Ribosomal Ribonucleic Acid
OR	Odds ratio
18S rRNA	Small subunit rRNA
18S rDNA	Small subunit rDNA

Author contributions

M.Kv. carried out study conceptualization; M.R., V.Z., M.Ki. and E.H. carried out data curation; M.Ki., P.B. and B.S. carried out formal analysis; M.Ki. and M.Kv. carried out funding acquisition; V.Z., E.H. B.S. M.Ki. and M.R. carried out investigation; M.Ki., B.S. and M.Kv. carried out methodology; M.Kv. carried out project administration; M.Ki., M.R. and M.Kv. acquired resources; B.S. and P.B. carried out software; M.Kv. and M.Ki. carried out supervision; N.H., E.H., M.Ki. and M.Kv. carried out validation; M.Ki. and P.B. carried out visualization; V.Z. and N.H. carried out writing – original draft; B.S., M.Ki. and M.Kv. carried out writing – review and editing. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by Grant Agency of the Czech Republic, grant number GACR 21-23773S and by the Polish Ministry of Health subvention according to number SUBZ.A060.24.016 from the IT Simple system of Wrocław Medical University.

Data availability

All DNA material and datasets on which the conclusions of the manuscript rely are stored at the Institute of Parasitology, Biology Centre, Czech Academy of Sciences, České Budějovice, Czech Republic. Representative nucleotide sequences generated in this study were submitted to the GenBank database under the accession numbers [PV817847–PV817854].

Declarations

Ethics approval and consent to participate

Ethical approval and consent to participate were not required, as the study did not involve the handling of animals in a manner covered by legislation on the protection of animals against cruelty.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Institute of Parasitology, Biology Centre of the Academy of Sciences of the Czech Republic, V.V.I, Branišovská 31, 370 05 České Budějovice, Czech Republic. ²Faculty of Agriculture and Technology, University of South Bohemia in České Budějovice, České Budějovice, Czech Republic. ³Department of Biology and Medical Parasitology, Wrocław Medical University, Wrocław, Poland. ⁴National Agricultural and Food Centre, Lužianky, Slovakia. ⁵Faculty of Forestry, Technical University in Zvolen, Zvolen, Slovakia.

Received: 14 July 2025 Accepted: 23 September 2025
Published online: 21 November 2025

References

- Graczyk TK, Fayer R, Cranfield MR. *Cryptosporidium parvum* is not transmissible to fish, amphibians, or reptiles. *J Parasitol*. 1996;82:748–51.
- Kváč M, McEvoy J, Stenger B, Clark M. Cryptosporidiosis in other vertebrates. In: Cacciò SM, Widmer G, editors. *Cryptosporidium: parasite and disease*. 1st ed. Wien: Springer; 2014.

3. Nader JL, Mathers TC, Ward BJ, Pachebat JA, Swain MT, Robinson G, et al. Evolutionary genomics of anthroponosis in *Cryptosporidium*. *Nat Microbiol*. 2019;4:826–36. <https://doi.org/10.1038/s41564-019-0377-x>.
4. Tůmová L, Ježková J, Prediger J, Holubová N, Sak B, Konečný R, et al. *Cryptosporidium mortiferum* n. sp. (Apicomplexa: Cryptosporidiidae), the species causing lethal cryptosporidiosis in Eurasian red squirrels (*Sciurus vulgaris*). *Parasit Vectors*. 2023;16:235. <https://doi.org/10.1186/s13071-023-05844-8>.
5. Stripen B. Parasitic infections: time to tackle cryptosporidiosis. *Nature*. 2013;503:189–91.
6. Chappell CL, Okhuysen PC. Cryptosporidiosis. *Curr Opin Microbiol*. 2002;15:523–7.
7. World population review: cat population by Country 2025. https://worldpopulationreview.com/country-rankings/cat-population-by-country?utm_source=chatgpt.com. 2025.
8. Ein N, Li L, Vickers K. The effect of pet therapy on the physiological and subjective stress response: a meta-analysis. *Stress Health*. 2018;34:477–89. <https://doi.org/10.1002/smi.2812>.
9. WHO: animal bites. https://www.who.int/news-room/fact-sheets/detail/animal-bites?utm_source=chatgpt.com. 2024.
10. Taghipour A, Khazaei S, Ghodisian S, Shajrizadeh M, Olfatfar M, Foroutan M, et al. Global prevalence of *Cryptosporidium* spp. in cats: a systematic review and meta-analysis. *Res Vet Sci*. 2021;137:77–85. <https://doi.org/10.1016/j.rvsc.2021.04.015>.
11. Montazeri M, Mikaelli G, Galeb T, Moosazadeh M, Sarvi S, Dodangeh S, Javidnia J, et al. The global serological prevalence of *Toxoplasma gondii* in felids during the last five decades (1967–2017): a systematic review and meta-analysis. *Parasit Vectors*. 2020;13:82. <https://doi.org/10.1186/s13071-020-3954-1>.
12. Bouzid M, Halai K, Jeffreys D, Hunter PR. The prevalence of *Giardia* infection in dogs and cats, a systematic review and meta-analysis of prevalence studies from stool samples. *Vet Parasitol*. 2015;207:181–202. <https://doi.org/10.1016/j.vetpar.2014.12.011>.
13. Pavlásek I, Ryan U. The first finding of a natural infection of *Cryptosporidium muris* in a cat. *Vet Parasitol*. 2007;144:349–52. <https://doi.org/10.1016/j.vetpar.2006.10.005>.
14. Scorza V, Tangtrongsup S. Update on the diagnosis and management of *Cryptosporidium* spp infections in dogs and cats. *Top Companion Anim Med*. 2010;25:163–9. <https://doi.org/10.1053/j.tcam.2010.07.007>.
15. Yang RC, Ying J, Monis P, Ryan U. Molecular characterisation of *Cryptosporidium* and *Giardia* in cats (*Felis catus*) in Western Australia. *Exp Parasitol*. 2015;155:13–8. <https://doi.org/10.1016/j.exppara.2015.05.001>.
16. Meng XZ, Li MY, Liu C, Qin YF, Zhao ZY, Yang XB, et al. The global prevalence and risk factors of *Cryptosporidium* infection among cats during 1988–2021: a systematic review and meta-analysis. *Microb Pathog*. 2021;158:105096. <https://doi.org/10.1016/j.micpath.2021.105096>.
17. Ryan U, Zahedi A, Feng Y, Xiao L. An update on zoonotic *Cryptosporidium* species and genotypes in Humans. *Animals (Basel)*. 2021;11:3307. <https://doi.org/10.3390/ani11113307>.
18. Rojas-Lopez L, Elwin K, Chalmers RM, Enemark HL, Beser J, Troell K. Development of a gp60-subtyping method for *Cryptosporidium felis*. *Parasit Vectors*. 2020;13:39. <https://doi.org/10.1186/s13071-020-3906-9>.
19. Li J, Yang F, Liang R, Guo S, Guo Y, Li N, et al. Subtype characterization and zoonotic potential of *Cryptosporidium felis* in cats in Guangdong and Shanghai, China. *Pathogens*. 2021;10:89. <https://doi.org/10.3390/pathogens10020089>.
20. Jiang W, Roellig DM, Lebbad M, Beser J, Troell K, Guo Y, et al. Subtype distribution of zoonotic pathogen *Cryptosporidium felis* in humans and animals in several countries. *Emerg Microbes Infect*. 2020;9:2446–54. <https://doi.org/10.1080/22221751.2020.1840312>.
21. Yun CS, Moon BY, Lee K, Kang SM, Ku BK, Hwang MH. The detection and phylogenetic characterization of *Cryptosporidium*, *Cystoisospora*, and *Giardia duodenalis* of cats in South Korea. *Front Cell Infect Microbiol*. 2023;13:1296118. <https://doi.org/10.3389/fcimb.2023.1296118>.
22. Miláček P, Vitovec J. Differential staining of cryptosporidia by aniline-carbol-methyl violet and tartrazine in smears from feces and scrapings of intestinal mucosa. *Folia Parasitol*. 1985;32:50.
23. Xiao LH, Escalante L, Yang CF, Sulaiman I, Escalante AA, Montali RJ, et al. Phylogenetic analysis of *Cryptosporidium* parasites based on the small-subunit rRNA gene locus. *Appl Environ Microbiol*. 1999;65:1578–83. <https://doi.org/10.1128/AEM.65.4.1578-1583.1999>.
24. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol*. 2003;52:696–704. <https://doi.org/10.1080/10635150390235520>.
25. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*. 2018;35:1547–9. <https://doi.org/10.1093/molbev/msy096>.
26. Tavaré S. Some probabilistic and statistical problems in the analysis of DNA sequences. In: Miura RM, editor. *Some mathematical questions in biology: DNA sequence analysis (Lectures on mathematics in the life sciences)*. New York: American Mathematical Society; 1986.
27. Kváč M, Hofmannová L, Ortega Y, Holubová N, Horčíková M, Kicia M, et al. Stray cats are more frequently infected with zoonotic protists than pet cats. *Folia Parasitol*. 2017;64:1. <https://doi.org/10.14411/Fp.2017.034>.
28. Hsu CH, Liang C, Chi SC, Lee KJ, Chou CH, Lin CS, et al. An epidemiological assessment of *Cryptosporidium* and *Giardia* spp. infection in pet animals from Taiwan. *Animals*. 2023;13:3373. <https://doi.org/10.3390/ani13213373>.
29. Gil H, Cano L, de Lucio A, Bailo B, de Mingo MH, Cardona GA, et al. Detection and molecular diversity of *Giardia duodenalis* and *Cryptosporidium* spp. in sheltered dogs and cats in Northern Spain. *Infect Genet Evol*. 2017;50:62–9. <https://doi.org/10.1016/j.meegid.2017.02.013>.
30. Hinney B, Ederer C, Stengl C, Wilding K, Strkolcova G, Harl J, et al. Enteric protozoa of cats and their zoonotic potential—a field study from Austria. *Parasitol Res*. 2015;114:2003–6. <https://doi.org/10.1007/s00436-015-4408-0>.
31. Piekara-Stepinska A, Piekarska J, Gorczykowski M. *Cryptosporidium* spp. in dogs and cats in Poland. *Ann Agric Environ Med*. 2021;28:345–7. <https://doi.org/10.26444/aaem/120467>.
32. Kostopoulou D, Claerebout E, Arvanitis D, Ligda P, Voutzourakis N, Casaert S, et al. Abundance, zoonotic potential and risk factors of intestinal parasitism amongst dog and cat populations: the scenario of Crete, Greece. *Parasit Vectors*. 2017;10:43. <https://doi.org/10.1186/s13071-017-1989-8>.
33. Scorza V, Willmott A, Gunn-Moore D, Lappin MR. *Cryptosporidium felis* in faeces from cats in the UK. *Vet Rec*. 2014;174:609. <https://doi.org/10.1136/vr.102205>.
34. Robertson LJ, Björkman C, Axén C, Fayer R. Cryptosporidiosis in Farmed Animals. In: Cacciò SM, Widmer G, editors. *Cryptosporidium: parasite and disease*. London: Springer; 2014.
35. Ng-Hublin JS, Singleton GR, Ryan U. Molecular characterization of *Cryptosporidium* spp. from wild rats and mice from rural communities in the Philippines. *Infect Genet Evol*. 2013;16:5–12. <https://doi.org/10.1016/j.meegid.2013.01.011>.
36. Lv C, Zhang L, Wang R, Jian F, Zhang S, Ning C, et al. *Cryptosporidium* spp. in wild, laboratory, and pet rodents in China: prevalence and molecular characterization. *Appl Environ Microbiol*. 2009;75:7692–9. <https://doi.org/10.1128/AEM.01386-09>.
37. Kváč M, Hanzlíková D, Sak B, Květoňová D. Prevalence and age-related infection of *Cryptosporidium suis*, *C. muris* and *Cryptosporidium* pig genotype II in pigs on a farm complex in the Czech Republic. *Vet Parasitol*. 2009;160:319–22. <https://doi.org/10.1016/j.vetpar.2008.11.007>.
38. Kváč M, Kestřánová M, Květoňová D, Kotková M, Ortega Y, McEvoy J, et al. *Cryptosporidium tyzzeri* and *Cryptosporidium muris* originated from wild West-European house mice (*Mus musculus domesticus*) and East-European house mice (*Mus musculus musculus*) are non-infectious for pigs. *Exp Parasitol*. 2012;131:107–10. <https://doi.org/10.1016/j.exppara.2012.03.016>.
39. Lindsay DS, Upton SJ, Owens DS, Morgan UM, Mead JR, Blagburn BL. *Cryptosporidium andersoni* n. sp. (Apicomplexa: Cryptosporidiidae) from cattle, *Bos taurus*. *J Eukaryot Microbiol*. 2000;47:91–5. <https://doi.org/10.1111/j.1550-7408.2000.tb00016.x>.
40. Xu H, Jin Y, Wu W, Li P, Wang L, Li N, et al. Genotypes of *Cryptosporidium* spp., *Enterocytozoon bienersi* and *Giardia duodenalis* in dogs and cats in Shanghai, China. *Parasit Vectors*. 2016;9:121. <https://doi.org/10.1186/s13071-016-1409-5>.
41. Stensvold CR, Beser J, Axen C, Lebbad M. High applicability of a novel method for gp60-based subtyping of *Cryptosporidium meleagridis*. *J Clin Microbiol*. 2014;52:2311–9. <https://doi.org/10.1128/JCM.00598-14>.
42. Stensvold CR, Elwin K, Winiecka-Krusnell J, Chalmers RM, Xiao L, Lebbad M. Development and application of a gp60-based typing assay for *Cryptosporidium viatorum*. *J Clin Microbiol*. 2015;53:1891–7. <https://doi.org/10.1128/JCM.00313-15>.

43. Li N, Xiao L, Alderisio K, Elwin K, Cebelinski E, Chalmers R, et al. Subtyping *Cryptosporidium ubiquitum*, a zoonotic pathogen emerging in humans. *Emerg Infect Dis*. 2014;20:217–24. <https://doi.org/10.3201/eid2002.121797>.
44. Cama VA, Bern C, Roberts J, Cabrera L, Sterling CR, Ortega Y, et al. *Cryptosporidium* species and subtypes and clinical manifestations in children, Peru. *Emerg Infect Dis*. 2008;14:1567–74. <https://doi.org/10.3201/eid1410.071273>.
45. Cama VA, Ross JM, Crawford S, Kawai V, Chavez-Valdez R, Vargas D, et al. Differences in clinical manifestations among *Cryptosporidium* species and subtypes in HIV-infected persons. *J Infect Dis*. 2007;196:684–91. <https://doi.org/10.1086/519842>.
46. Feng Y, Ryan UM, Xiao L. Genetic diversity and population structure of *Cryptosporidium*. *Trends Parasitol*. 2018;34:997–1011. <https://doi.org/10.1016/j.pt.2018.07.009>.
47. Li F, Li J, Tang Y, He W, Fan Y, Huang N, et al. Involvement of a variant secretory protein in virulence of emerging *Cryptosporidium parvum* subtypes. *Virulence*. 2025;16:2514077. <https://doi.org/10.1080/21505594.2025.2514077>.
48. Monticello TM, Levy MG, Bunch SE, Fairley RA. Cryptosporidiosis in a feline leukemia virus-positive cat. *J Am Vet Med Assoc*. 1987;191:705–6.
49. Mtambo MM, Nash AS, Blewett DA, Smith HV, Wright S. *Cryptosporidium* infection in cats: prevalence of infection in domestic and feral cats in the Glasgow area. *Vet Rec*. 1991;129:502–4.
50. Thompson RC, Olson ME, Zhu G, Enomoto S, Abrahamsen MS, Hijjawi NS. *Cryptosporidium* and cryptosporidiosis. *Adv Parasitol*. 2005;59:77–158. [https://doi.org/10.1016/S0065-308X\(05\)59002-X](https://doi.org/10.1016/S0065-308X(05)59002-X).
51. Tangtrongsup S, Scorza AV, Reif JS, Ballweber LR, Lappin MR, Salman MD. Seasonal distributions and other risk factors for *Giardia duodenalis* and *Cryptosporidium* spp. infections in dogs and cats in Chiang Mai, Thailand. *Prev Vet Med*. 2020;174:104820. <https://doi.org/10.1016/j.prevetmed.2019.104820>.
52. Hofmannová L, Hauptman K, Huclová K, Květonová D, Sak B, Kváč M. *Cryptosporidium erinacei* and *C. parvum* in a group of overwintering hedgehogs. *Eur J Protistol*. 2016;56:15–20. <https://doi.org/10.1016/j.ejop.2016.05.002>.
53. Sone B, Ambe LA, Ampama MN, Ajohkoh C, Che D, Nguinkal JA, et al. Prevalence and molecular characterization of *Cryptosporidium* species in diarrheic children in Cameroon. *Pathogens*. 2025;14:287. <https://doi.org/10.3390/pathogens14030287>.
54. Zajackowska Z, Brutovska AB, Akutko K, McEvoy J, Sak B, Hendrich AB, et al. Horse-specific *Cryptosporidium* genotype in human with Crohn's disease and arthritis. *Emerg Infect Dis*. 2022;28:1289–91. <https://doi.org/10.3201/eid2806.220064>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Aplikace poznatků ve výuce a pedagogické činnosti

Nedílnou součástí vědecké práce je přenos získaného know-how na nastupující generaci studentů a výzkumníků. Metodiky a algoritmické principy popisované v této habilitační práci nezůstaly uzavřeny pouze v publikovaných článcích, ale staly se organickou součástí výuky na Jihočeské univerzitě v Českých Budějovicích. Integrace pokročilé bioinformatiky do standardních biologických kurzů se ukazuje jako nezbytný krok pro přípravu absolventů, kteří budou schopni obstat v éře masivních genomických dat.

Transfer poznatků se realizuje ve dvou rovinách: formou přímé výuky v přednáškách a cvičeních, a formou individuálního vedení závěrečných prací. Principy moderního návrhu diagnostických testů se promítají do sylabů klíčových předmětů, jako jsou Bioinformatika, Metody genového inženýrství, Molekulární biologie genu a buňky či Bakteriologie.

V rámci předmětu Bioinformatika nejsou studenti seznamováni pouze s klasickými nástroji typu BLAST, ale přímo konfrontováni s limity těchto metod při analýze celogenomových dat. Výuka zahrnuje praktické ukázky alignment-free přístupů a práce s pangenomy, což studentům umožňuje pochopit rozdíl mezi teoretickou homologií a diagnostickou unikátností. Algoritmy pro exkluzi k-merů, které tvoří jádro této práce, slouží jako modelové příklady pro vysvětlení problematiky tzv. velkých dat v biologii.

V předmětech zaměřených na laboratorní techniky, jako jsou Metody genového inženýrství, se těžiště přesouvá k praktické aplikaci. Studenti se učí, že precizní *in silico* návrh je prerekvizitou úspěšného experimentu. Poznatky z výzkumu termodynamické stability primerů a optimalizace reakčních podmínek pro PCR a LAMP jsou přímo využívány při návrhu laboratorních úloh. Studenti tak získávají unikátní možnost projít celým procesem - od získání sekvenčních dat, přes bioinformatickou filtraci kandidátních markerů, až po jejich empirické ověření v laboratoři.

Nejhmataatelnějším důkazem funkčnosti tohoto propojení vědy a výuky je úspěšné vedení 15 bakalářských a diplomových prací. Témata těchto prací byla úzce provázána s problematikou molekulární detekce a návrhem nových testů. Pod mým vedením studenti aplikovali postupy "Hundred-Genomes Pipeline" na reálné problémy molekulární diagnostiky. Řešili úkoly spojené s detekcí specifických patogenů, přičemž se museli vypořádat s reálnými taxonomickými komplikacemi a variabilitou cílových organismů.

Tato pedagogická praxe potvrzuje, že prezentované metody nejsou jen akademickými koncepty, ale funkčními nástroji, které si dokáží osvojit i studenti bakalářského a magisterského studia. Absolventi, kteří prošli tímto tréninkem, si odnášejí schopnost kritického myšlení v souvislostech genomiky a molekulární diagnostiky. Nejsou pouhými uživateli komerčních kitů, ale rozumí principům, které stojí za jejich fungováním, a jsou schopni samostatně navrhovat řešení pro nově se objevující biologické hrozby.

Publikace s přímou vazbou na habilitační práci

- Beran, P., & Mraz, I. (2015). ABIF Manager: simple reading and editing of ABIF formatted files. *Bioinformatics*, 31(24), 4012-4013.
- Beran, P., Rost, M., Beranová, K., Kváč, M., Stehlíková, D., Udoh, O. E., Jozová, E., & Čurn, V. (2025). genCRC32: Collision-free CRC32 based hashing of DNA sequences. *Bioinformatics Advances*, vbaf315
- Beran, P., Stehlíková, D., Cohen, S. P., & Čurn, V. (2021). KEC: unique sequence search by k-mer exclusion. *Bioinformatics*, 37(19), 3349-3350.
- Beran, P., Stehlikova, D., Cohen, S. P., Rost, M., Beranova, K., & Curn, V. (2023). Utilization of a new hundred-genomes pipeline to design a rapid duplex LAMP detection assay for *Xanthomonas euvesicatoria* and *X. vesicatoria* in tomato. *Plant Disease*, 107(6), 1822-1828.
- Pecenka, J., Kocanova, M., Baranek, M., Gazdik, F., Ragasova, L., Penazova, E., Čechová, J., Beran, P., & Eichmeier, A. (2020). Species-specific PCR primers for the detection of poorly distinguishable *Xanthomonas euvesicatoria*. *Crop protection*, 127, art. no. 104978.
- Stehlíková, D., Beran, P., Cohen, S. P., & Čurn, V. (2020). Development of real-time and colorimetric loop mediated isothermal amplification assay for detection of *Xanthomonas gardneri*. *Microorganisms*, 8(9), 1301.

Seznam použité literatury

- Andreson, R., Brauer, A., Kaplinski, L., Külaots, M., Saumaa, S., Kurg, A., & Remm, M. (2025). K-mer based method for finding sequence type-specific PCR primers for *Listeria monocytogenes*. medRxiv, 2025-12.
- Amoia, S. S., Loconsole, G., Ligorio, A., Pantazis, A. K., Papadakis, G., Gizeli, E., & Minafra, A. (2023). A colorimetric LAMP detection of *Xylella fastidiosa* in crude alkaline sap of olive trees in apulia as a field-based tool for disease containment. Agriculture, 13(2), 448.
- Bao, Z., & Weigel, D. (2025). Complexity welcome: Pangenome graphs for comprehensive population genomics. Quantitative Plant Biology, 1-34.
- Beran, P., & Mraz, I. (2015). ABIF Manager: simple reading and editing of ABIF formatted files. Bioinformatics, 31(24), 4012-4013.
- Beran, P., Mraz, I., Kokoskova, B., & Bohata, A. (2015). Monitoring the occurrence of bacterial spot of tomato and pepper in the Czech Republic and development of new PCR primers for detection of *Xanthomonas vesicatoria*. European journal of plant pathology, 141(3), 617-621.
- Beran, P., Stehlíková, D., Cohen, S. P., & Čurn, V. (2021). KEC: unique sequence search by k-mer exclusion. Bioinformatics, 37(19), 3349-3350.
- Beran, P., Stehlikova, D., Cohen, S. P., Rost, M., Beranova, K., & Curn, V. (2023). Utilization of a new hundred-genomes pipeline to design a rapid duplex LAMP detection assay for *Xanthomonas euvesicatoria* and *X. vesicatoria* in tomato. Plant Disease, 107(6), 1822-1828.
- Beran, P., Rost, M., Beranová, K., Kváč, M., Stehlíková, D., Udoh, O. E., Jozová, E., & Čurn, V. (2025). genCRC32: Collision-free CRC32 based hashing of DNA sequences. Bioinformatics Advances, vbaf315.
- Besati, M., Safarnejad, M. R., Aliahmadi, A., Farzaneh, M., Ruiz, R., Montagud-Martínez, R., Rodrigo, G. & Rafati, H. (2025). Detection of tomato brown

- rugose fruit virus through CRISPR-Cas12a and CRISPR-Cas9 systems. *Scientific Reports*, 15(1), 25638.
- Bonham-Carter, O., Steele, J., & Bastola, D. (2014). Alignment-free genetic sequence comparisons: a review of recent approaches by word analysis. *Briefings in bioinformatics*, 15(6), 890-905.
- Bussi, Y., Kapon, R., & Reich, Z. (2021). Large-scale k-mer-based analysis of the informational properties of genomes, comparative genomics and taxonomy. *PloS one*, 16(10), e0258693.
- Bustin, S., & Nolan, T. (2017). Talking the talk, but not walking the walk: RT-qPCR as a paradigm for the lack of reproducibility in molecular research. *European Journal of Clinical Investigation*, 47(10), 756-774.
- Calderoni, F., Petrontino, A., Frem, M., Fucilli, V., & Bozzo, F. (2025). Economic and social impacts of olive quick decline syndrome: Analysing data from the Italian farm accountancy network. *Plant Pathology*, 74(4), 1010-1023.
- CIRAD. (2024, 12. srpna). Plant disease could spell apocalypse for citrus fruits. CIRAD News.
- Constantin, E. C., Cleenwerck, I., Maes, M., Baeyen, S., Van Malderghem, C., De Vos, P., & Cottyn, B. (2016). Genetic characterization of strains named as *Xanthomonas axonopodis* sp. *dieffenbachiae* leads to a taxonomic revision of the *X. axonopodis* species complex. *Plant Pathology*, 65(5), 792-806.
- de Andrade, A. V., Pereira, F. S., da Silva, F. N., da Silva, G. F., & Magalhães, M. D. L. B. (2024). Validation and optimization of the loop-mediated isothermal amplification (LAMP) technique for rapid detection of wheat stripe mosaic virus, a wheat-infecting pathogen. *Journal of Genetic Engineering and Biotechnology*, 22(2), 100373.
- Du, K., Zeng, Q., Jiang, M., Hu, Z., Zhou, M., & Xia, K. (2025). CRISPR/Cas12a-based biosensing: advances in mechanisms and applications for nucleic acid detection. *Biosensors*, 15(6), 360.

- Ellis, J., et al., (2025). Developing epidemiological preparedness for a plant disease invasion: Modelling citrus huánglóngbìng in the European Union. *Plants, People, Planet*, 7(5), 1403-1423.
- Fielder, H., Cunliffe, N. J., Beale, T., Oliver, G., Szyniszewska, A. M., Taylor, P., & Jeger, M. J. (2025). A synoptic review of plant disease epidemics and outbreaks published in 2023 with synthesis of trends between 2021 and 2024. *CABI Agriculture and Bioscience*, 6(1), 0072.
- Gabsi, N., Clérot, F., & Hébrail, G. (2010, June). Efficient trade-off between speed processing and accuracy in summarizing data streams. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining* (pp. 343-353). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Ghedia, B., Dhokia, G., Dave, P., & Rana, P. (2026). The Role of Artificial Intelligence in the Evolution of Clinical and Diagnostic Microbiology: A Systematic Review. *J Pure Appl Microbiol*, 20(1), 131-138.
- Ghezzi, H., Fan, Y. M., Ng, K. M., Burckhardt, J. C., Pepin, D. M., Lin, X., Ziels, R. M., & Tropini, C. (2024). PUPpy: a primer design pipeline for substrain-level microbial detection and absolute quantification. *Msphere*, 9(7), e00360-24.
- Gomez-Gutierrez, S. V., & Goodwin, S. B. (2022). Loop-mediated isothermal amplification for detection of plant pathogens in wheat (*Triticum aestivum*). *Frontiers in Plant Science*, 13, 857673.
- Gootenberg, J. S., et al. (2017). Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science*, 356(6336), 438-442.
- Gudkov, S. V., Matveeva, T. A., Sarimov, R. M., Simakin, A. V., Stepanova, E. V., Moskovskiy, M. N., ... & Izmailov, A. Y. (2023). Optical methods for the detection of plant pathogens and diseases. *AgriEngineering*, 5(4), 1789-1812.
- Harrison, J., Hussain, R. M., Aspin, A., Grant, M. R., Vicente, J. G., & Studholme, D. J. (2023). Phylogenomic analysis supports the transfer of 20

- pathovars from *Xanthomonas campestris* into *Xanthomonas euvesicatoria*. Taxonomy, 3(1), 29-45.
- Hossain, M. M. et al. (2024). Plant disease dynamics in a changing climate: impacts, molecular mechanisms, and climate-informed strategies for sustainable management. Discover Agriculture, 2(1), 132.
- Chai, L., Cui, X., Liu, H., Zhang, Y., Pan, Y., Li, C., & Le, T. (2025). CRISPR/Cas Toolbox: Unveiling the secret weapon for SNP detection. TrAC Trends in Analytical Chemistry, 192, 118321.
- Chao, J., Tang, F., & Xu, L. (2022). Developments in algorithms for sequence alignment: a review. Biomolecules, 12(4), 546.
- Jagathjothi, N., et al. (2024). Plant pathogen mitigation and adaptation to climate change. In Plant quarantine challenges under climate change anxiety (pp. 53-78). Cham: Springer Nature Switzerland.
- Jasieniecka, A., & Domingues, I. (2025). CRISPR-Cas9 and its bioinformatics tools: A systematic review. Current Issues in Molecular Biology, 47(5), 307.
- Kamran, M., et al. (2026). Integrating genomic technologies with artificial intelligence, CRISPR editing, and high-throughput phenotyping for engineering disease-resistant crops. Plant Stress, 101317.
- Kanapiya, A., Amanbayeva, U., Tulegenova, Z., Abash, A., Zhangazin, S., Dyussebayev, K., & Mukiyanova, G. (2024). Recent advances and challenges in plant viral diagnostics. Frontiers in Plant Science, 15, 1451790.
- Kralik, P., & Ricchi, M. (2017). A basic guide to real time PCR in microbial diagnostics: definitions, parameters, and everything. Frontiers in microbiology, 8, 108.
- Lagner, J. R., Newberry, E. A., Rivera, Y., Zhang, L., Vakulskas, C. A., & Qi, Y. (2025). Amplification-free detection of plant pathogens by improved CRISPR-Cas12a systems: a case study on phytoplasma. Frontiers in Plant Science, 16, 1544513.

- Landa, B. B., et al. (2020). Emergence of a plant pathogen in Europe associated with multiple intercontinental introductions. *Applied and Environmental Microbiology*, 86(3), e01521-19.
- Li, K., et al. (2008). Novel computational methods for increasing PCR primer design effectiveness in directed sequencing. *BMC bioinformatics*, 9(1), 191.
- Luchi, N., Migliorini, D., Pecori, F., & Santini, A. (2023). Real-time portable LAMP assay for a rapid detection of *Xylella fastidiosa* in-field. In *Plant-Pathogen Interactions* (pp. 51-60). New York, NY: Springer US.
- Ma, X. K., et al. (2024). Bioinformatics software development: Principles and future directions. *The Innovation Life*, 2(3), 100083-1.
- Matsuda, T., Sakamoto, H., Kayukawa, T., Kitashima, Y., Kozaki, T., & Gotoh, T. (2025). A PCR primer design method for identifying spider mite species using k-mer counting. *Plos one*, 20(6), e0321199.
- Mermigka, G., Megariti, M., Malliarakis, D., Pagoulatou, M. G., Gizeli, E., & Goumas, D. E. (2026). Development and Validation of a Field-Based Colorimetric LAMP Assay for the Detection of *Clavibacter michiganensis* in Tomato Plants. *Plants*, 15(3), 372.
- Mirmajlessi, S. M., Loit, E., Maend, M., & Mansouripour, S. M. (2015). Real-time PCR applied to study on plant pathogens: potential applications in diagnosis-a review. *Plant Protection Science*, 51(4), 177-190.
- Navasca, A., Singh, J., Rivera-Varas, V., Gill, U., Secor, G., & Baldwin, T. (2025). Dispensable genome and segmental duplications drive the genome plasticity in *Fusarium solani*. *Frontiers in Fungal Biology*, 6, 1432339.
- Notomi, T., Okayama, H., Masubuchi, H., Yonekawa, T., Watanabe, K., Amino, N., & Hase, T. (2000). Loop-mediated isothermal amplification of DNA. *Nucleic acids research*, 28(12), e63-e63.
- Ogbuji, N. G., & Agogbua, J. U. (2025). Genomics in plant pathogen identification and control. *Frontiers in Plant Science*, 16, 1661432.

- Osdaghi, E., et al. (2021). A centenary for bacterial spot of tomato and pepper. *Molecular Plant Pathology*, 22(12), 1500.
- Pecenka, J., et al. (2020). Species-specific PCR primers for the detection of poorly distinguishable *Xanthomonas euvesicatoria*. *Crop protection*, 127.
- Peňázová, E., Dvořák, M., Ragasová, L., Kiss, T., Pečenka, J., Čechová, J., & Eichmeier, A. (2020). Multiplex real-time PCR for the detection of *Clavibacter michiganensis* subsp. *michiganensis*, *Pseudomonas syringae* pv. tomato and pathogenic *Xanthomonas* species on tomato plants. *PloS one*, 15(1), e0227559.
- PS, M. G., & Chintala, B. R. (2020). Big data challenges and opportunities in agriculture. *International Journal of Agricultural and Environmental Information Systems (IJAEIS)*, 11(1), 48-66.
- Ren, J., Bai, X., Lu, Y. Y., Tang, K., Wang, Y., Reinert, G., & Sun, F. (2018). Alignment-free sequence analysis and applications. *Annual review of biomedical data science*, 1(1), 93-114.
- Ren, H., Luo, S., Yu, Y., Chen, P., & Wang, K. (2026). Advances in the Application of CRISPR/Cas Systems in Molecular Diagnostics. *Synthetic Biology and Engineering*, 4(2), 10005.
- Ristaino, J. B., et al. (2021). The persistent threat of emerging plant disease pandemics to global food security. *Proceedings of the National Academy of Sciences*, 118(23), e2022239118.
- Roy, S., & Mukhopadhyay, A. (2024). A randomized optimal k-mer indexing approach for efficient parallel genome sequence compression. *Gene*, 907, 148235.
- Sahu, S., Boukherroub, R., Ritzenthaler, C., & Szunerits, S. (2025). Emerging technologies for in-field plant virus detection: innovations and future directions. *Journal of General Virology*, 106(11), 002182.

- Salem, N. M., Jewehan, A., Aranda, M. A., & Fox, A. (2023). Tomato brown rugose fruit virus pandemic. *Annual review of phytopathology*, 61(1), 137-164.
- Savary, S., Willocquet, L., Pethybridge, S. J., Esker, P., McRoberts, N., & Nelson, A. (2019). The global burden of pathogens and pests on major food crops. *Nature ecology & evolution*, 3(3), 430-439.
- Serena, A. S., et al. (2025). Naked-Eye Molecular Testing for the Detection of *Xylella fastidiosa* in Mallorca (Balearic Island) Almond Orchards by Colorimetric LAMP. *Applied Sciences*, 15(2), 739.
- Shaizadinova, A., Amanzholova, M., Rukavitsina, I., Abeldenov, S., & Zhumakayev, A. R. (2025). CRISPR/Cas12a-based method coupled with isothermal amplification to identify *Alternaria* spp. isolated from wheat grain samples. *Frontiers in Microbiology*, 15, 1468336.
- Sheershika, S., & Ram, M. (2024). Advances in DNA extraction techniques: a comprehensive review of methods and applications. *Plant Cell Biotechnology and Molecular Biology*, 25(5-6), 30-42.
- Shi, L., & Wang, Z. (2019). Computational strategies for scalable genomics analysis. *Genes*, 10(12), 1017.
- Song, B., Buckler, E. S., & Stitzer, M. C. (2024). New whole-genome alignment tools are needed for tapping into plant diversity. *Trends in Plant Science*, 29(3), 355-369.
- Stehlíková, D., Beran, P., Cohen, S. P., & Čurn, V. (2020). Development of real-time and colorimetric loop mediated isothermal amplification assay for detection of *Xanthomonas gardneri*. *Microorganisms*, 8(9), 1301.
- Stehlíková, D., et al. (2025). Duplex PCR Detection and Differentiation of Insect DNA *Tenebrio molitor* and *Zophobas morio* in Various Types of Food. *Insects*, 16(9), 983.

- Timilsina, S., et al. (2020). *Xanthomonas* diversity, virulence and plant–pathogen interactions. *Nature Reviews Microbiology*, 18(8), 415-427.
- USDA ARS. (2024, 28. února). New Discovery Speeds Scientists' Push for HLB-Tolerant Citrus. *Research News*.
- Velasco-Amo, M. P., Olivares-García, C., Román-Écija, M., Marco-Noales, E., Navas-Cortes, J. A., & Landa, B. B. (2026). Modeling the Accuracy of *Xylella fastidiosa* Molecular Diagnostic Tests in Naturally Infected Almond Tree Samples. *Plant Disease*, PDIS-12.
- Vo, T. T. B., Troiano, E., Lal, A., Hoang, P. T., Kil, E. J., Lee, S., & Parrella, G. (2022). ToLCNDV-ES infection in tomato is enhanced by TYLCV: Evidence from field survey and agroinoculation. *Frontiers in Microbiology*, 13, 954460.
- Vo, T. T. B., et al. (2025). The insidious threat: Assessing the dangers and spread of tomato leaf curl New Delhi virus. *The Plant Pathology Journal*, 41(1), 1.
- Wang, H., Tsinda, E. K., Dunn, A. J., Chikweto, F., Ahmed, N., Pelosi, E., & Zemkoho, A. B. (2022). Deep learning forward and reverse primer design to detect SARS-CoV-2 emerging variants. *arXiv preprint arXiv:2209.13591*.
- Wang, X., Liang, S., Gan, Q., Cai, B., & Liu, C. (2025). Current status and future perspectives of the diagnostic of plant bacterial pathogens. *Frontiers in Plant Science*, 16, 1547974.
- Wei, W., Yang, Y., & Shih, J. (2026). Rapid and Sensitive Detection of Phytoplasma Diseases Using a CRISPR/Cas12a DETECTR Assay Combined with Isothermal Recombinase Polymerase Amplification. In *Phytoplasma: Methods and Protocols* (pp. 63-74). New York, NY: Springer US.
- Wu, C., et al. (2021). Application of artificial intelligence ensemble learning model in early prediction of atrial fibrillation. *BMC bioinformatics*, 22(Suppl 5), 93.

- Wu, H., Blanca, A., & Medvedev, P. (2025). A k-mer-Based Estimator of the Substitution Rate Between Repetitive Sequences. 25th International Conference on Algorithms for Bioinformatics (WABI 2025), 20, 1-20.
- Yang, H., Ledesma-Amaro, R., Gao, H., Ren, Y., & Deng, R. (2023). CRISPR-based biosensors for pathogenic biosafety. *Biosensors and Bioelectronics*, 228, 115189.
- Yang, Y., Yang, R., Luo, C., Wang, X., & Ma, F. (2026). Loop-Mediated Isothermal Amplification (LAMP): Current Progress and Future Directions in Rapid Nucleic Acid Detection. *Sensors and Actuators Reports*, 100440.
- Zielezinski, A., et al. (2019). Benchmarking of alignment-free sequence comparison methods. *Genome biology*, 20(1), 144.