

Gesamt-Abschlussbericht

Zuwendungsempfänger:

Dr. Jochen Kumlehn (IPK Gatersleben)

Prof. Dr. Daguang Cai (CAU Kiel)

Förderkennzeichen: 2820HS001 & 2820HS012

Bundesministerium
für Ernährung
und Landwirtschaft

ptble

Projekträger Bundesanstalt
für Landwirtschaft und Ernährung

Thema des Forschungsvorhabens:

Machbarkeitsstudie zu Nachweis- und Identifizierungsverfahren für Genom-editierte Pflanzen und pflanzliche Produkte

Laufzeit des Vorhabens:

01.01.2021 – 30.11.2023

Berichtszeitraum:

01.01.2021 – 30.11.2023

Wissenschaftliche Projektmitarbeiter

Dr. habil. Dirk Schenke (CAU Kiel)



Dr. Iris Hoffie (IPK Gatersleben)

Ehemaliger Mitarbeiter: Dr. Martin Becker (IPK Gatersleben)

**Projektpartner**

Dr. Martin Mascher (IPK Gatersleben)

Dr. Marina Pupke Marone (IPK Gatersleben)

Unterauftragnehmer

Prof. Dr. Michael Kleine (Planton GmbH, Kiel)

Dr. Martin Weigel (Planton GmbH, Kiel)



Dr. Olav Grundmann (Impetus GmbH & Co. Bioscience KG, Bremen)

**Zusammenarbeit mit anderen Stellen**

Dr. Lutz Grohmann (BVL, Berlin)



Bundesamt für
Verbraucherschutz und
Lebensmittelsicherheit

Inhaltsverzeichnis

1. Ziele und Aufgabenstellung des Vorhabens	4
1.1. Planung und Ablauf des Vorhabens	5
1.2. Wissenschaftlicher und technischer Stand, an den angeknüpft wurde.....	6
2. Material und Methoden.....	8
3. Ergebnisse	14
3.1. Genomsequenzierung und bioinformatische Datenauswertung	14
3.1.1. <i>Whole-Genome Sequencing</i> , Hi-C Sequenzierung und Assemblierung der Genomdaten in Gerste	14
3.1.2. Ermittlung von Mutations-gekoppelten Sequenz-Polymorphismen in Gerste	16
3.1.3. Ermittlung von Transgenrückständen und <i>Off-Targets</i> in Gerste	17
3.1.4. <i>Whole-Genome Sequencing</i> , Hi-C Sequenzierung und Assemblierung der Genomdaten in Raps	17
3.1.5. Ermittlung von Mutations-gekoppelten Sequenz-Polymorphismen in Raps.....	20
3.1.6. Ermittlung von Transgenrückständen und <i>Off-Targets</i> in Raps.....	22
3.2. PCR-basierte Etablierung der Detektion von Einzelnukleotid-Veränderungen	23
3.2.1. qPCR in Gerste	23
3.2.2. ddPCR in Gerste	26
3.2.3. ddPCR in Raps.....	27
3.3. Charakterisierung der Zielregionen mittels Tiefensequenzierung von Amplikons.....	28
3.3.1. Amplikonsequenzierung in Gerste	28
3.1.2. Amplikonsequenzierung in Raps	30
3.4. Versuch der Etablierung eines Identifikations-Assays in Raps.....	34
3.5. Etablierung einer Analysepipeline für die Tiefensequenzierung von Amplikons	35
4. Diskussion und zusammenfassende Bewertung.....	36
5. Literaturverzeichnis	38

Abkürzungsverzeichnis

CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
Cas	CRISPR-associated
CAU	Christian-Albrechts Universität (AG Cai)
ddPCR	Digital droplet PCR
DNA	Deoxyribonucleic acid (DNS, Desoxyribonukleinsäure)
ENGL	European Network of GMO Laboratories
gDNA	genomische DNA
GVO	Gentechnisch veränderter Organismus (aka GMO, LMO)
Hi-C	Chromosome Conformation Capture Sequenzierung
InDels	Insertionen/Deletionen
IPK	Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung (AGs Kumlehn/Mascher)
LNA	Locked Nucleic Acids
LOD	Limit of Detection (Nachweisgrenze)
MGB	Minor Groove Binder
MRPL	Minimum Required Performance Limit
MT	Mutante/ Pflanze mit Cas9-induzierter Mutation
Nt	Nukleotid
PAM	Protospacer-Adjacent Motif (bei Cas9: NGG)
PCR	Polymerase Kettenreaktion
pDNA	Plasmid-DNA
qPCR	quantitative PCR
RNA	Ribonucleic acid (RNS, Ribonukleinsäure)
(s)gRNA	(single) guide RNA
SNP/SNV	Einzelnukleotid-Polymorphismen/Veränderungen/Variationen
SOP	Standard Operating Procedure
TALENs	Transcription Activator-Like Effector Nucleases
WGS	Whole Genome Sequencing (Komplett-Sequenzierung eines Genoms)
WT	Wildtyp (eine Pflanze mit dem ursprünglichen genetischen Hintergrund)

1. Ziele und Aufgabenstellung des Vorhabens

Der Anbau und Vertrieb von gentechnisch veränderten Organismen (GVO) ist in der europäischen Union stark reglementiert (Richtlinie 2001/18/EG). Genom-editierte Pflanzen fallen unter diese Regelung, wie in der Rechtssache C-528/16 durch den Europäischen Gerichtshof im Jahr 2018 entschieden wurde. Für Pflanzen, die aus artifiziell induzierter Mutagenese resultieren, wurde in diesem Urteil generell klargestellt, dass sie GVO sind. Während jedoch solche Pflanzen mit chemisch oder durch ionisierende Strahlung induzierten Mutationen von der Regulierung ausgenommen sind, unterliegen (Genom-editierte) Organismen mit Zielsequenz-spezifisch hervorgerufenen Mutationen den gesetzlichen Regelungen, die ursprünglich für transgene Pflanzen entwickelt worden sind. Zielsequenz-spezifische Mutationen sind erst seit einigen Jahren mit praxisrelevanter Effizienz möglich. Diese Technologie beruht auf Endonukleasen, die an zu verändernde Ziel motive pflanzlicher Genome angepasst werden können. Die Zielsequenzspezifität solcher Endonukleasen kann Protein-vermittelt (Meganukleasen, Zinkfingernukleasen, *Transcription Activator-like Effector Nucleases* (TALENs)) oder RNA-vermittelt sein (*Clustered Regularly Interspaced Short Palindromic Repeats* (CRISPR)-assoziierte (Cas) Endonukleasen). Cas-Endonukleasen stellen derzeit die für die biotechnologische Praxis geeignetsten molekularen Werkzeuge dieser Art dar. Pflanzen können unter Verwendung dieser Endonukleasen mutiert werden, indem entweder *guide* RNA (gRNA)- und Cas-Endonuklease-codierende DNA oder vorproduziertes Cas-Protein sowie gRNA in Pflanzenzellen übertragen werden. Das dabei adressierte Zielmotiv im pflanzlichen Genom setzt sich aus einem typischerweise 20 Nukleotide (nt) umfassenden Bereich, zu dem das 5'-Ende der dafür synthetisierten gRNA identisch ist, und einem davon stromabwärts gelegenen Triplett aus einem beliebigen Nukleotid und zwei Guanin-Nukleotiden - dem so genannten *Protospacer-Adjacent Motif* (PAM) zusammen. Das Grundprinzip der gezielten Mutagenese beruht darauf, dass die Endonuklease den DNA-Doppelstrang im Zielmotiv durchtrennt und bei der daraufhin aktivierten zelleigenen DNA-Reparatur Fehler auftreten, die sich im Genom als Mutationen manifestieren. Am häufigsten kommen dabei der Verlust oder das Zufügen eines einzelnen Nukleotids vor, es können jedoch auch bis zu mehrere hundert Nukleotide aus der Zielregion verloren gehen oder Nukleotide durch andere ersetzt werden. Die ersten Nutzpflanzen mit Zielsequenz-spezifischen Mutationen werden außerhalb Europas bereits vermarktet und angebaut (wie z. B. in einigen süd- und nordamerikanischen sowie asiatischen Ländern) und es ist aufgrund des züchterischen Potenzials der Genom-Editierung davon auszugehen, dass es in absehbarer Zeit zahlreiche weitere Zulassungen geben wird. Zudem werden solche Pflanzen in den meisten Regionen jenseits von Europa rechtlich nicht wie transgene Pflanzen reguliert und gelangen über vergleichsweise einfache Sortenzulassungen auf den Markt. Eine Voraussetzung für das Inverkehrbringen von GVO und den aus ihnen hergestellten Produkten in der EU ist die Verfügbarkeit von Nachweis- und Identifizierungsverfahren, um eine Marktkontrolle innerhalb der EU zu gewährleisten. Zudem muss kontrolliert werden können, ob nicht zugelassene GVO in Verkehr gebracht werden und ob die Kennzeichnungspflicht für zugelassene GVO eingehalten wird. Aus diesen Gründen ist es erforderlich, zuverlässige Nachweis- und Identifizierungsverfahren für Genom-editierte Pflanzen zu entwickeln.

1.1. Planung und Ablauf des Vorhabens

Laut Bekanntmachung Nr. 02/20/32 des BMEL vom 30. April 2020 wurde vom Projekt die Etablierung von Nachweis- und Identifizierungsverfahren für Genom-editiertes Pflanzenmaterial und die Validierung dieser Verfahren bezüglich ihrer Eignung für GVO-Genehmigungsanträge, amtliche Kontrollen und Eigenkontrollen der Wirtschaft erwartet. Im Wesentlichen ergab sich daraus folgendes Konzept:

1. Genaue und umfassende Sequenzierung des Gesamtgenoms einer durch die Antragsteller vorab bereitzustellenden Genom-editierten Linie im Vergleich zu ihrer Donorlinie
2. Entwicklung von Analyseverfahren für die Identifizierung und den Nachweis von Zielmotivspezifisch induzierten Sequenzveränderungen. Dafür wurden in Vorbereitung des Projektes folgende Methoden für die Detektion und Charakterisierung mutierter genomischer Regionen ausgewählt:
 - quantitative (q)PCR
 - *digital droplet* (dd)PCR
 - Tiefensequenzierung von PCR-Amplifikaten und Etablierung einer dafür geeigneten Analysepipeline

Es sollte in diesem Zusammenhang geprüft werden, ob ein analytisches Verfahren einen Genom-editierten Organismus eindeutig identifizieren und von Organismen unterscheiden kann, bei denen eine identisch modifizierte Nucleotidsequenz auf andere Weise entstanden ist.

Durch ein initiales *Whole Genome Sequencing* (WGS) einer Gerstenlinie und zweier Rapslinien mit marktrelevanten Zielsequenz-spezifischen Veränderungen sowie deren nicht-editierten Ausgangslinien sollte dieses Referenzmaterial zunächst molekular genau charakterisiert werden.

Auf Grundlage dieser umfassenden Sequenzdaten sollten molekularbiologische Methoden und Technologien genutzt werden, um die gezielt induzierten Mutationen in diesem Referenzmaterial nachzuweisen. Geeignete Methoden sollten dann zu routinemäßig einsetzbaren Detektionsverfahren weiterentwickelt werden. Als zu testende Technologien waren die *quantitative real-time* PCR (qPCR), die *digital droplet* PCR (ddPCR) und die Tiefensequenzierung von PCR-Amplikons vorgesehen. Für die Auswertung der Amplikonsequenzen sollte des Weiteren ein bioinformatisches Verfahren (eine Analyse-Pipeline) entwickelt werden. Die Aufteilung sah dabei vor, dass die Experimente mit den Rapslinien hauptsächlich vom Projektpartner CAU Kiel und die Experimente mit den Gerstenlinien vom Projektpartner IPK Gatersleben und den jeweiligen Unterauftragnehmern durchgeführt werden (Abbildung 1). Dennoch gab es auch Überschneidungen. So sollten die WGS und Hi-C Sequenzierungen beider Arten federführend vom Projektpartner IPK durchgeführt werden. Des Weiteren wurden vom IPK und dessen Unterauftragnehmer Impetus Bioscience die ddPCR-Assay Etablierungen für Gerste und Raps durchgeführt.

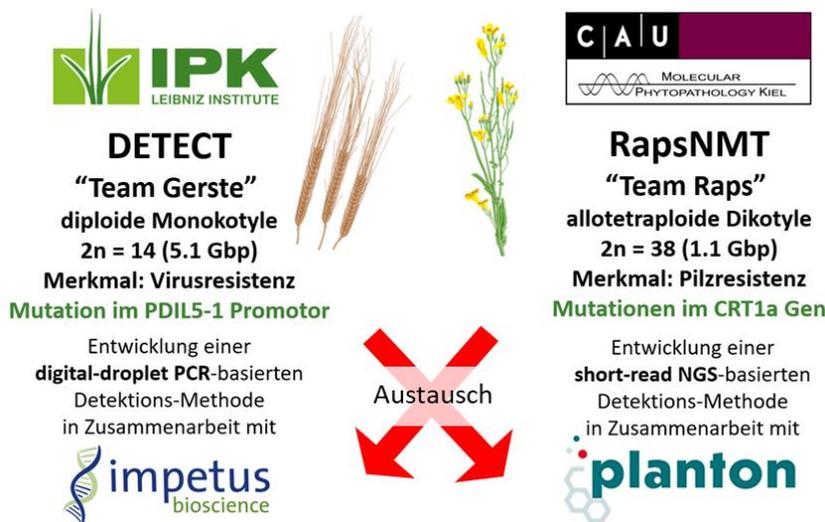


Abbildung 1: Übersicht über das Verbundprojekt beruhend auf Material zweier unterschiedlicher Kulturarten mit durch Genom-Editierung induzierter Resistenz gegenüber Pflanzenschädlingen, d. h. einer Virus-resistenten Gerste „Golden Promise“ (Hoffie *et al.*, 2022) und einem Pilz-resistenten Sommer-Raps „Mozart“ (Pröbsting *et al.*, 2020).

Mit einem geeigneten Detektionsverfahren sollte ein minimal erforderliches Leistungslimit (*minimum required performance limit* (MRPL)) von 0,9 % bzw. 0,1 % Mengenanteilen von GVO-Beimischungen zuverlässig in homogenen und heterogenen (Schüttgut) Proben nachgewiesen werden können. Abschließend war vorgesehen, die als geeignet eingeschätzten Detektionsverfahren anhand optimierter Protokolle in den DIN EN ISO/IEC 17025-akkreditierten Labors der Unterauftragnehmer mit der dort bereits zur Verfügung stehenden Ausstattung hinsichtlich ihrer Mach- und Verwendbarkeit zu testen und weiter zu optimieren.

Mindestens eins der getesteten und optimierten Detektionsverfahren sollte spezifisch und sensitiv genug sein, um Probenmaterial mit gezielt induzierten Mutationen zuverlässig und wiederholbar nachweisen zu können. Im Zuge dieses Projektes sollte zudem beurteilt werden, inwiefern die Unterscheidung einer Genom-editierten Linie von einer entsprechenden spontan entstandenen Mutante möglich wäre.

1.2. Wissenschaftlicher und technischer Stand, an den angeknüpft wurde

Das Forschungsvorhaben dient dem Nachweis der Erfüllbarkeit des Erfordernisses für GVO-Genehmigungsanträge und amtliche Kontrollen sowie Eigenkontrollen der Wirtschaft, aussagekräftige Nachweis- und Identifizierungsverfahren für in Verkehr zu bringende oder bereits in Verkehr gebrachte Genom-editierte Lebens- und Futtermittel zur Verfügung stellen zu können. Es sollte also die Möglichkeit eruiert werden, ob und unter welchen Bedingungen die Etablierung geeigneter Nachweis- und Identifizierungsverfahren möglich ist. Bereits verfügbare theoretische Erkenntnisse sollten anhand aussagekräftiger experimenteller Daten validiert und ergänzt werden. Erwartetes Ergebnis des Forschungsvorhabens war die Etablierung mindestens einer Detektionsmethode, mit der vorab bekannte, durch Zielsequenz-spezifische Nukleasen erzeugte Mutationen detektiert werden können. Die zu detektierenden Veränderungen sind vorrangig Insertionen und Deletionen (Indels) in Cas9-adressierten DNA-Motiven Zielsequenz-spezifisch mutierter Pflanzen.

Nach logischer Betrachtung und aktuellem Wissenstand ist ein Nachweis von Genom-editierten Pflanzen ohne vorherige Informationen über die veränderten DNA-Sequenzen nicht möglich. Für

die Identifizierung Genom-editierter Pflanzen mit bekannter Veränderung wäre ein zweiter bekannter DNA-Polymorphismus, der im Kontext der betreffenden Genom-Editierung auf spontane Weise entstanden ist und sich in unmittelbarer Nähe der gezielt induzierten Mutation befindet ein solides Indiz (Pecoraro *et al.*, 2019). Ansonsten sind Genom-editierte Pflanzen nicht von Pflanzen zu unterscheiden, die durch konventionelle Züchtung oder ungerichtete Mutagenese-Techniken hergestellt wurden. In der EU zugelassene, durch „klassische“ Gentechnik (Transgenese) hergestellte Pflanzen sind mit etablierten PCR-basierten Methoden nachweisbar, identifizierbar und quantifizierbar, da stabil integrierte Transgen-Sequenzen erwartungsgemäß nicht spontan vorkommen und solche Pflanzen somit einzigartig sind (Grohmann *et al.*, 2019; Pallarz *et al.*, 2023). Dagegen enthalten Genom-editierte Pflanzen in der Regel keine rekombinante DNA, sondern weisen vornehmlich kleine Veränderungen auf Nukleotidebene auf. In der GVO-Analytik wurden in der Vergangenheit hauptsächlich PCR-basierte Methoden (zumeist qPCR) eingesetzt, die auch für den Nachweis von kleinen Veränderungen, wie z. B. einzelne Nukleotidvarianten aus technischer Sicht geeignet sein dürften (Grohmann *et al.*, 2019). Da eine sichere Detektion und Quantifizierung von Genom-editierten Pflanzen aber eine hohe Präzision und Auflösung erfordern, sind für die GVO-Analytik ddPCR- und Tiefensequenzierungs-basierte Nachweis-Methoden geeignete Ergänzungen zu den häufig verwendeten qPCR-Assays (Fraiture *et al.*, 2017; Pecoraro *et al.*, 2019; Fraiture *et al.*, 2022; Pallarz *et al.*, 2023). Die größten Vorteile der ddPCR bestehen darin, dass die Methode eine besonders gute Unterscheidbarkeit sehr ähnlicher Sequenzen und eine absolute Quantifizierung der Zielsequenz ermöglicht, ohne von optimalen Referenzmaterialien und Standardkurven abhängig zu sein (Pecoraro *et al.*, 2019). Dies ist besonders interessant, wenn kein Referenzmaterial zur Verfügung steht, wie zum Beispiel bei unbekanntem GVOs (Fraiture *et al.*, 2022). PCR-basierte Methoden sollten jedoch nicht mehr als einzige Technologie für den Nachweis und die Identifizierung von Genom-editierten Pflanzen eingesetzt werden, da Tiefensequenzierungs-basierte Verfahren noch solidere Informationen liefern können (Pallarz *et al.*, 2023; Fraiture *et al.*, 2023; Grohmann *et al.*, 2019). Zum Zeitpunkt des Beginns der Projektlaufzeit waren sequenzierungsbasierte Nachweismethoden, wie z.B. die Amplikon-Tiefensequenzierung, noch nicht in der GVO-Analytik verwendet worden, das Potenzial dieser Methoden war aber bereits in ersten Studien aufgezeigt (Fraiture *et al.*, 2017; Arulandhu *et al.*, 2018; Pecoraro *et al.*, 2019). Qualitätskriterien für die Bewertung von Sequenzierungsdaten und die Validierung sowohl von ddPCR- als auch sequenzierungsbasierten Methoden waren zu diesem Zeitpunkt jedoch noch nicht evaluiert (Pecoraro *et al.*, 2019).

2. Material und Methoden

2.1. Pflanzenmaterial

Im Teilprojekt DETECT des Partners IPK wurde vorrangig mit einer Transgen-freien Gerstenmutante (16-12) gearbeitet, die eine Cas9-induzierte homozygote 1 bp-Deletion im 5'-UTR des *Protein Disulfide Isomerase Like 5-1 (PDIL5-1)*-Gens aufweist (Abbildung 2; Hoffie *et al.*, 2022).

Wildtyp Golden Promise cccagagaaatacatcatcgacccaccccgactggcgatccgccttct
 pdil5-1 Mutante cccagagaaatacatcatcgacccaccc-gactggcgatccgccttct

Abbildung 2: Vergleich der genomischen Zielregionen der Donorlinie 'Golden Promise' und der Gerstenmutante 16-12. Das Zielmotiv inkl. PAM ist unterstrichen

Der Verlust der Funktion dieses Anfälligkeitgens führt zu Resistenz der Gerste gegen Bymoviren, was bereits vor Projektbeginn anhand der hier untersuchten Mutante bestätigt worden war. Für die experimentellen Arbeiten wurden Blattmaterial und Körner der dritten Generation (M₃) der Genom-editierten Transgen-freien Gerstenlinie verwendet. Die Gerstenlinie 'Golden Promise' DH1-6 war als Donormaterial für die Transformation mit einem *PDIL5-1*-gRNA/Cas9-Konstrukt und damit für die Herstellung dieser Mutante verwendet worden. Zudem wurde diese Linie bei allen Experimenten als Wildtyp-Kontrolle verwendet.

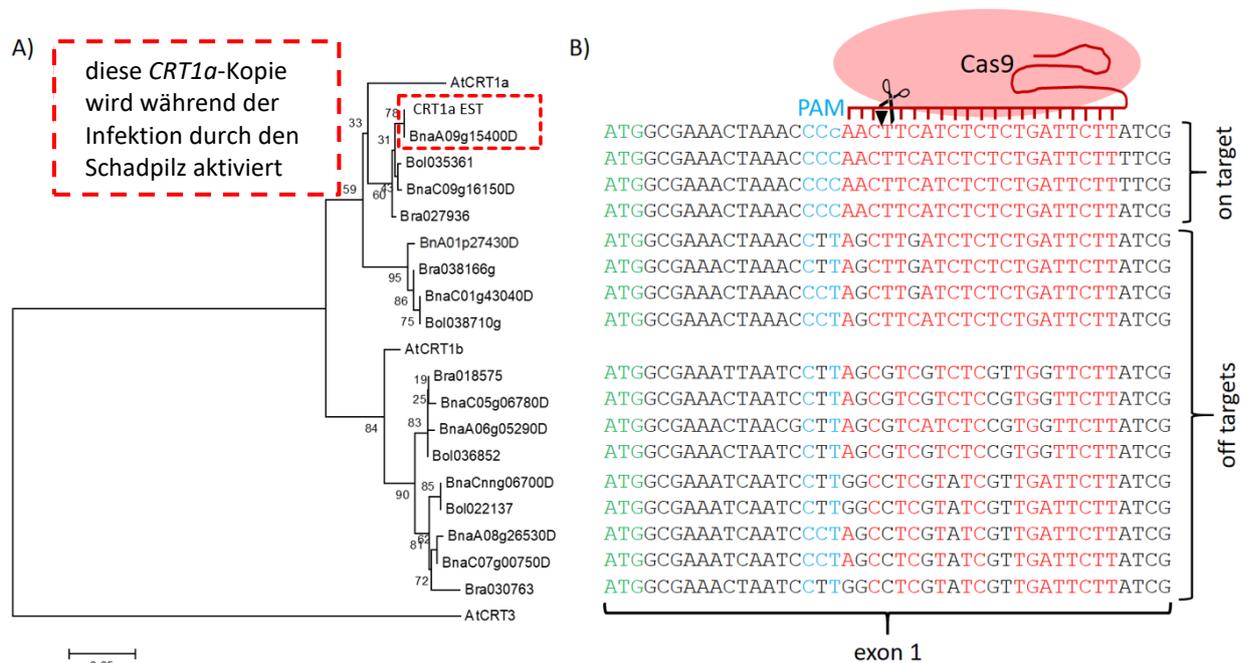


Abbildung 3: Das Genom vom Raps (*Brassica napus*) besteht aus den beiden Subgenomen der „Raps-Eltern“ *Brassica oleracea* (Bol) und *Brassica rapa* (Bra). A) Es gibt insgesamt 12 *Calreticulin*-Loci im Raps, von denen hier nur die 4 *CRT1a*- und die 4 *CRT1b*-Loci gezeigt werden. Der Fokus lag auf dem *CRT1a* A09 Locus, da diese Kopie während der Infektion durch das Pathogen *Verticillium longisporum* aktiviert wird. Von *CRT1a* gibt es somit insgesamt 8 Allele und es ist daher unwahrscheinlich, dass zufällig ein Raps entsteht, bei dem alle diese Kopien spontan mutiert sind und zu einem neuen Phänotyp führen. Dieser auf *Neighbor Joining* basierende phylogenetische Stammbaum wurde mit cDNA Sequenzen des Winter-Raps Referenzgenoms „Darmor“ erstellt. B) Durch Genom-Editierung konnten spezifisch nur die *On-Target* Loci A09 und C09 mutiert werden, sodass die *Off-Target* Loci A01 und C01 intakt blieben.

Im Teilprojekt des Partners CAU wurde vorrangig mit zwei Rapslinien gearbeitet, die Mutationen im Zielgen *CRT1a* in den AA- und CC-Subgenomen aufweisen. Die Rapslinie 'Mozart' diente als Donormaterial für die Transformation und damit für die Herstellung dieser Mutanten. Zudem wurde diese Linie bei allen Experimenten als Wildtyp-Kontrolle verwendet. Durch einen Cas9-induzierten KO von *CRT1a* kommt es zu einer erhöhten Resistenz gegen den bodenbürtigen Schadpilz *Verticillium longisporum* (Pröbsting *et al.*, 2020). Um Resistenz zu generieren, war vor allem die A09-Kopie von Interesse, aber aufgrund der hohen Ähnlichkeit mit der C09-Kopie sind beide Loci *On-Targets*. Dahingehend weisen die beiden A01 und C01 Loci wichtige Unterschiede im Bereich der PAM und der gRNA auf, sodass diese als *Off-Targets* definiert werden können (Abbildung 3).

Die beiden verwendeten T2-Mutanten C3E3 und C3E4 wurden durch Sanger-Sequenzierung hinsichtlich der Mutationen in den *CRT1a* Loci überprüft und entsprachen den in Pröbsting *et al.* (2020) beschriebenen Mutationen. Die C3E3 Mutante war homozygot im A09 Locus mutiert, jedoch nicht in den Allelen des C09 Locus. Die C3E4 Mutante war wie die T0-Ausgangspflanze in beiden Allelen des A09 Locus heterozygot und auch in einem Allel des C09 Locus mutiert (Abbildung 4). In den *Off-Target* Loci A01 und C01 wurden durch Sanger-Sequenzierung hingegen keine Mutationen gefunden.

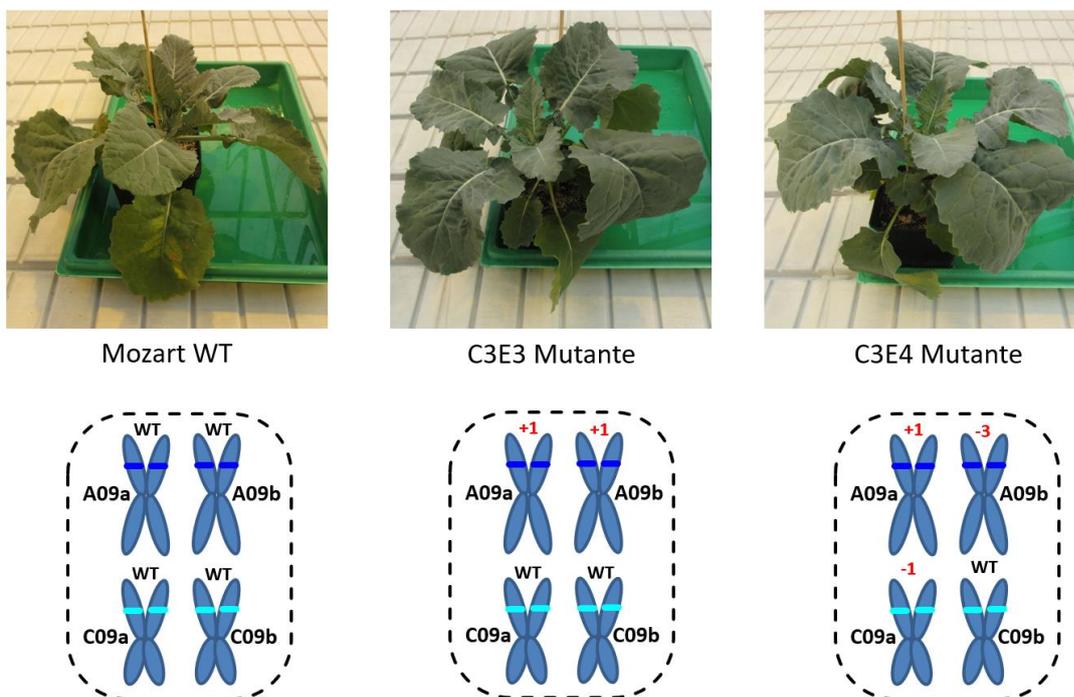


Abbildung 4: Die für das Projekt verwendeten Raps-Pflanzen. Neben der Ausgangslinie „Mozart“ wurden die gRNA/Cas9-transgene C3E3 Mutante und die zu Projektbeginn für Transgen-frei gehaltene C3E4 Mutante ausgewählt. Für die WGS des WT wurden Blätter von 8 Geschwisterpflanzen gepoolt, während die WGS der Mutanten auf Einzelpflanzen beruht. Das Saatgut dieser beiden T2-Mutanten wurde auch für die darauf folgenden Untersuchungen verwendet. Unter den Pflanzen ist die jeweilige genetische Konstitution für die *CRT1a* A09 und C09 Loci dargestellt. Bei der +1 bp Mutation handelt es sich um eine Insertion von einem Nukleotid („T“), durch die -3 bp Deletion ging die Aminosäure Phenylalanin an Position 8 des Genprodukts verloren und bei der -1 bp Deletion eines Nukleotids („T“) kam es, wie auch bei der +1 bp Mutation, zu einem Knockout des Gens durch Verschiebung des translationalen Leserahmens.

2.2. Whole-Genome Sequencing von Gerste und Raps

Die *PacBio HiFi Longread*-Sequenzierung wurde von der Firma Genewiz/Azenta Life Sciences (Chelmsford, MA, USA) durchgeführt und die Sequenzierungsergebnisse wurden wie in den folgenden Abschnitten beschrieben analysiert. Im Zuge des Antrages für einen kostenneutralen Technologiewechsel von der *Illumina Shortread* Sequenzierung zur *PacBio HiFi Longread*-Sequenzierung wurde sich in Anlehnung an frühere Ergebnisse für Gerste MorexV3 (Mascher 2021) für eine Coverage von 20x entschieden, da eine hohe Genauigkeit (> 99 %) bereits ab 15-facher Coverage besteht (Eid *et al.*, 2009).

Die jeweils 6 *HiFi Read Libraries* pro Gerstenlinie (Wildtyp und Mutante) und jeweils 2 *HiFi Read Libraries* pro Rapslinie (Wildtyp und die beiden Mutanten) wurden mittels des *de novo* Genom-Assemblierungs-Tools *Hifiasm* v.0.16.1 (Konfiguration für homozygote Genome, Parameter '-l 0') zu Genom-repräsentierenden *Contigs* assembliert (Cheng *et al.*, 2021). Die GFA-Assemblierung wurde mit *gfatools* v. 0.4-r179 (<https://github.com/lh3/gfatools>) in FASTA konvertiert, und die *Contigs* wurden in sichelförmiger numerischer Reihenfolge mit *seqtk* v. 1.3-r116 mit dem Modul "rename" (new names "contig_X") umbenannt (<https://github.com/lh3/seqtk>).

2.3. Chromosome Conformation Capture von Gerste und Raps

Für diese Sequenzierung wurden für Gerste Geschwisterpflanzen der Genom-sequenzierten Donor-Linie 'Golden Promise' DH1-6 und der *pdil5-1*-Mutante 16-12 verwendet. Für Raps wurden Geschwisterpflanzen der Donor-Linie 'Mozart' und der C3E4 Mutante verwendet. Die Hi-C-Sequenzierung wurde im IPK-Sequenzierlabor (AG GGR) durchgeführt und die Auswertung der Daten erfolgte wie in 2.3.1. und 2.3.2. beschrieben.

2.3.1. Bioinformatische Auswertung der Hi-C-Sequenzierung von Gerste

Die *TRITEX-Pipeline* (Marone *et al.* 2022) wurde verwendet, um das *Scaffolding* der *Assemblies* durchzuführen. Diese Pipeline verwendet eine *Contig*-Assemblierung, Hi-C-Reads und einen *Guide* (eine *Guide Map* oder *Single-Copy*-Regionen eines Referenzgenoms), um eine Assemblierung auf Chromosomen-Ebene zu erzeugen. Um die Anzahl der Sequenzvarianten in Form von Einzelnukleotid-Polymorphismen (SNPs) und Insertionen/Deletionen (Indels) zu ermitteln, wurde eine *Variant Calling*-Analyse durchgeführt.

2.3.2. Bioinformatische Auswertung der Hi-C-Sequenzierung von Raps

Die Assemblierung von Pseudomolekülen und die Suche nach strukturellen Variationen wurde wie in 2.3.1. beschrieben durchgeführt. Für die Rapsanalyse wurden die zwei HiFi-Read-Libraries der Mutante C3E4 sowie zwei Libraries der Donor-Linie verwendet und die zuvor beschriebenen Methoden angewendet. Die *TRITEX*-Analyse wurde mit dem Genom der cv. Darmor-bzh-Genom (Rousseau-Guetin *et al.*, 2020) als Referenz für die Donor-Linie durchgeführt, während die Mutante an den Pseudomolekülen der Donor-Linie ausgerichtet wurde.

2.4. *Off-Target*-Analyse und Suche nach Vektorfragmenten in Gerste

CCTop v.1.0.1 (Stemmer *et al.*, 2015) wurde verwendet, um nach potenziellen *Off-Target*-Regionen in den Pseudomolekülen der Donor-Linie 'Golden Promise' DH1-6 zu suchen. Dazu wurde der Genomindex mit Bowtie v.1.2.2 (Langmead *et al.*, 2009) erstellt und die Analyse in CCTop mit Standardparametern durchgeführt, wobei die Zielsequenz einschließlich der PAM-Sequenz als Eingabe verwendet wurde. Die Positionen möglicher *Off-Targets* wurden anschließend in den ungefilterten Sequenzvarianten-Daten der *pdil5-1*-Mutante 16-12 untersucht (um keine potenzielle *Off-Target*-Mutation zu übersehen), wobei auf Veränderungen 1.000 bp vor und nach dem Zielmotiv geprüft wurde.

Um nach Vektorfragmenten zu suchen, die möglicherweise in das Genom integriert worden sein könnten, wurde eine *k-mer*-basierte Suche mit den *Contig*-Assemblierungen durchgeführt. Aus der Vektorsequenz wurden *22-mers* ermittelt und mit Hilfe von BWA-MEM v. 0.7.17-r1188 (Li and Durbin, 2009) mit Standardparametern sowohl auf die Donor-Linie als auch auf die mutierte *Contig*-Assemblierung kartiert. Die ausgegebene SAM-Datei wurde dann nach alignierten *k-mers* durchsucht.

2.5. *Off-Target*-Analyse und Suche nach Vektorfragmenten in Raps

Zunächst wurden alle *CRT1a* Kopien auf den Chromosomen A09, C09, A01 und C01 untersucht. Mit keiner der angewendeten Methoden (Sanger, Illumina oder PacBio-Sequenzierung) wurden hier Mutationen in den beiden als *Off-Targets* definierten Kopien A01 oder C01 entdeckt. Mittels der Online-Vorhersagetools CCTop (Stemmer *et al.*, 2015; <https://cctop.cos.uni-heidelberg.de/>), CRISPR RGEN Tools (<http://www.rgenome.net/cas-offfinder/>), CRISPR-P (<http://cbi.hzau.edu.cn/CRISPR2/>) und CRISPRdirect (<https://crispr.dbcls.jp/>) wurden 18 weitere potenzielle *Off-Targets* identifiziert. Die Positionen möglicher *Off-Targets* wurden anschließend in den Genom-Daten der beiden Rapsmutanten C3E3 und C3E4 untersucht.

Um nach Transgenrückständen zu suchen, wurden die Genomdaten mit Hilfe eines Kartierungs-Ansatzes nach Vektorsequenzen untersucht.

2.6. Etablierung der Detektion von Einzelnukleotid-Veränderungen mittels qPCR in Gerste

Die qPCR-Analysen wurden am IPK mit einem LightCycler® 480 (Roche Life Science, Basel, Schweiz) und einem Bio-Rad CFX384 Touch (Bio-Rad Laboratories, Hercules, CA, USA) durchgeführt. Bei Impetus Bioscience wurde ein QuantStudio Real-Time-PCR-System (Thermo Fisher Scientific, Waltham, MA, USA) verwendet. Nach einem vorläufigen Test verschiedener Mastermixe, wurde der Kapa Probe Force qPCR-Mastermix (Roche, Mannheim, Deutschland) für alle weiteren qPCR-Experimente nach Herstellerangaben verwendet. Die Primer und Sonden der getesteten Systeme sind in Tabelle 1 aufgeführt. Zur Auswertung der Ergebnisse wurde die LightCycler® 480 Software (Roche Life Science) und die CFX Maestro Software (Bio-Rad) verwendet.

Tabelle 1: Verwendete Oligonukleotide für qPCR- und ddPCR-Assays in Gerste

Name	Sequenz (5'→ 3')	Konzentrationen (µM)	
		qPCR	ddPCR
System 1			
272_PDIL_qRT_fw	GACCGGCCAGAGAAATACATC	0,8	0,9
272_PDIL_qRT_rv	GAGACGAGGAGAAGGCCGATA	0,8	0,9
273_PDIL_Probe_SNV	FAM/GACCCACCCGACT/MGB/BMN-Q535	0,2	0,2
274_PDIL_Probe_WT	HEX/ACCCACCCGACT/MGB/BMN-Q535	0,2	0,2
System 2			
272_PDIL_qRT_fw	GACCGGCCAGAGAAATACATC	0,8	0,9
272_PDIL_qRT_rv	GAGACGAGGAGAAGGCCGATA	0,8	0,9
273 WT-Gerste-NEU	HEX/GACCCACCCGACTGGC/BMN-Q535	0,2	0,2
274 MUT-Gerste-NEU	FAM/CGACCCACCCGACTGGC/BMN-Q535	0,2	0,2
System 3			
IK435_HvPDIL5-1_fw	GTGGCCCATGTCCAGTTGTA	0,8	0,8
IK436_HvPDIL5-1_rv	AGACGAGGAGAAGGCCGATA	0,8	0,8
IK437_HvPDIL_Reference_Probe	HEX/GGACAGATGGCGACGGTGAATCTT/BMN-Q535, 2x	0,2	0,2
IK438_HvPDIL_SNV_Probe	CY5/CCAGTC+G+GGTGGTCGAT/BMN-Q620	0,2	0,2
System 4			
IK435_HvPDIL5-1_fw	GTGGCCCATGTCCAGTTGTA	0,9	
IK441_rev_LNA	CGCCAGT+C+GG+G+T	0,9	
IK437_HvPDIL_Reference_Probe	HEX/GGACAGATGGCGACGGTGAATCTT/BMN-Q535, 2x	0,4	
System 5			
IK435_HvPDIL5-1_fw	GTGGCCCATGTCCAGTTGTA	0,9	
IK442_rev_LNA2	ACGCCAGT+CGGG+T	0,9	
IK437_HvPDIL_Reference_Probe	HEX/GGACAGATGGCGACGGTGAATCTT/BMN-Q535, 2x	0,4	
System 6			
IK424_HvHor3_fw	AGACAAGGCGTGCAGATCG	0,9	0,8
IK425_HvHor3_rev	GACCCTGGACGAGCACACAT	0,9	0,8
IK426_HvHor3_Probe	HEX/CCTCAGCCGCAACAGGTGGGTC/BHQ-1	0,4	0,15

2.7. Etablierung der Detektion von Einzelnukleotid-Veränderungen mittels ddPCR in Gerste und Raps

Am IPK wurde für die ddPCR-Assayentwicklung das naica® System (Stilla, Villejuif, Frankreich) in Kombination mit den Sapphire Chips und dem naica® multiplex PCR Mix (Stilla, Villejuif, Frankreich) nach Herstellerangaben verwendet. Es wurden zunächst die Primer-/Sonden-Systeme 1-3 (Tabelle 1) getestet und dann entschieden, für die weitere Etablierung und Validierung System 2 zu verwenden. Die in Raps verwendeten Primer sind in Tabelle 2 aufgeführt. Die Ergebnisse wurden mit der CrystalMiner Software (Stilla) analysiert. Bei Impetus Bioscience wurde das System von Bio-Rad in Kombination mit dem ddPCR Supermix for Probes (Bio-Rad Laboratories, Hercules, CA, USA) nach Herstellerangaben verwendet. Die Durchführung der ddPCR-Assays war in Raps und Gerste gleich.

Tabelle 2: Verwendete Oligonukleotide für ddPCR-Assays in Raps

Name	Sequenz (5'→ 3')	Konzentrationen (µM)
CRT1 A09_fw	AGGAAGGTGCGATCTGAGTTT+C	0,9
CRT1 A09_rev	CGCGACGAGGCCGA+T	0,9
CRT1 A09_WT-Sonde	HEX/ CAACTTCATCTCTC /MGB/BMN-Q535	0,2
CRT1 A09_MT-Sonde	FAM/ CAACTTCATCTCTC /MGB/BMN-Q535	0,2

2.8. Charakterisierung der Zielregionen mittels Tiefensequenzierung von Amplikons in Gerste und Raps

Die *short-read* Amplikon-Tiefensequenzierungen wurden vom Unterauftragnehmer Planton anhand der MiSeq- und NextSeq-Systeme (Illumina, San Diego, CA, USA) durchgeführt und in separaten SOP-Entwürfen beschrieben. Die Zielregion wurde mit den in Tabelle 3 angegebenen Primern amplifiziert. Die Sequenzierungsergebnisse wurden mit Hilfe der in diesem Projekt entwickelten Analysepipeline analysiert (s. 2.9.). Beim Auftragnehmer Planton erfolgte die Auswertung mit einem von diesem Unterauftragnehmer entwickelten, proprietären Protokoll.

Tabelle 3: Für die Amplikon-Tiefensequenzierung verwendete Oligonukleotide

Name	Sequenz (5'→3')	Beschreibung
NGS_F3	CGGACCGGCCAGAGAAAT	NGS-Assay Gerste
NGS_R3	GAAGACGGGAGACGAGGAG	NGS-Assay Gerste
NGS1_fw	TGACAACACTAGATATGACGTGTA	NGS-Assay Raps
NGS1_rev	ATAACACTAGCAGAGGCGATC	NGS-Assay Raps
NGS2_fw	TGACAACACTAGATATGACGTGTA	NGS-Assay Raps
NGS2_rev	TCCTCGAATATAACACTAGCAGA	NGS-Assay Raps
NGS3a_fw	AGCGCTACTTCACTTGCGGCTT	NGS-Assay Raps
NGS3a_rev	ATAACACTAGCAGAGGCGATC	NGS-Assay Raps
NGS3b_fw	GCGTCTACTTCAGTTGCAGCTT	NGS-Assay Raps
NGS3b_rev	ATAACACTAGCAGAGGCGATC	NGS-Assay Raps
NGS4a_fw	AGCGCTACTTCACTTGCGGCTT	NGS-Assay Raps
NGS4a_rev	TCCTCGAATATAACACTAGCAGA	NGS-Assay Raps
NGS4b_fw	GCGTCTACTTCAGTTGCAGCTT	NGS-Assay Raps
NGS4b_rev	TCCTCGAATATAACACTAGCAGA	NGS-Assay Raps
NGS5_fw	TTCACTTGCGGCTTCTCTCTG	NGS-Assay Raps
NGS5_rev	TACCTTCGAATTTTTCCTCG	NGS-Assay Raps
NGS6_fw	TTCACTTGCGGCTTCTCTCTG	NGS-Assay Raps
NGS6_rev	CGATGCGCCAGCCGAATGAATC	NGS-Assay Raps

2.9. Etablierung einer Analysepipeline für die Tiefensequenzierung von Amplikons

Zur Detektion und Charakterisierung von Einzelnukleotid-Veränderungen wurde eine prototypische bioinformatische Pipeline unter Verwendung frei verfügbarer Sequenzanalyse-Software entwickelt. Eingabedaten sind Sequenzdaten der Illumina-Novaseq-Sequenzierung im FastQ-Format, eine Tabelle mit Barcode der einzelnen Proben und die erwartete Zielsequenz. Zur Dekonvolution der FastQ-Dateien eines Amplicon-Pools mithilfe der Barcode-Sequenzen wird Cutadapt verwendet (Martin 2011). Anschließend werden mit BBMerge (Bushnell *et al.* 2017) die beiden Einzel-*Reads* anhand ihrer überlappenden Endsequenz zusammengefügt. Identische *Reads* werden gruppiert und das Vorkommen jeder unikalen Sequenz gezählt. Sequenzen, deren Häufigkeit geringer als ein vom Nutzer gesetzter Schwellwert ist, werden nicht weiter berücksichtigt. Abschließend wird ein multiples Sequenz-Alignment aller unikalen Sequenzen und der erwarteten Zielsequenz mit dem Tool MAFFT (Kato and Toh, 2008) durchgeführt. Die alignierten Sequenzen werden dem Benutzer im FASTA-Format ausgegeben. Diese können mit dem Programm Unipro UGENE (Unipro, Novosibirsk, Russland) ausgewertet werden.

3. Ergebnisse

3.1. Genomsequenzierung und bioinformatische Datenauswertung

3.1.1. Whole-Genome Sequencing, Hi-C Sequenzierung und Assemblierung der Genomdaten in Gerste

Ziel dieses experimentellen Ansatzes war es, eine mittels Cas9-Endonuklease im Zielgen veränderte Gerstenlinie hinsichtlich ihres gesamten Genoms zu charakterisieren sowie auf Grundlage der vergleichenden Sequenzierung einer entsprechenden nicht-mutierten Linie genomische Referenzdaten für weitere Arbeitspakete des Projektes bereitzustellen. Dazu wurde eine Cas9-induzierte Mutante (16-12) der Gerstensorte 'Golden Promise' DH 1-6 ausgewählt, die im *PDIL5-1*-Gen eine 1 Basenpaar-Deletion in homozygotem Zustand aufweist.

Für die Donor-Linie 'Golden Promise' DH1-6 und die Mutante wurde mittels PacBio *Longread*-Sequenzierung eine 35-fache bzw. 21-fache Genomabdeckung erzielt, womit die diesbezügliche Vorgabe erfüllt ist. Die Gesamtlängen der Genomassemblierungen in Form von *Contigs* entsprechen der erwarteten Größenordnung (Tabelle 4).

Tabelle 4: Assemblierung der Sequenzier-Reads von Wildtyp und *pdil5-1*-Mutante in *Contigs*

Parameter	Golden Promise DH1-6	Mutante
Gesamtlänge (nt)	4.267.761.466	4.244.548.292
Anzahl der <i>Contigs</i>	1.697	1.288
Länge des <i>Contigs</i> N50 (nt)	15.234.071	13.243.152
Länge des <i>Contigs</i> N90 (nt)	3.703.820	3.360.967
Durchschnittliche <i>Contig</i> -Größe (nt)	2.514.885	3.295.456
Maximale <i>Contig</i> -Größe (nt)	53.394.372	78.942.127

Die Assemblierung anhand von *HiFi-Reads* umfasst jedoch noch kein so genanntes *Scaffolding*, d.h. eine verlässliche lokale Zuordnung der *Contigs* zu chromosomalen Pseudomolekülen (*in silico* Repräsentanten ganzer Chromosomen). Für dieses *Scaffolding* war eine zusätzliche Hi-C-Sequenzierung (*chromosome conformation capture sequencing*) erforderlich, dessen Durchführung von uns als essentiell für die Erreichung der Projektziele angesehen wurde, jedoch im Vergleich zur Projektplanung zusätzlichen Zeit- und Kosten-Aufwand erforderte.

Für die Erstellung der Pseudomoleküle, die eine Chromosomensequenz repräsentieren, wurde eine bislang unveröffentlichte pangenomische 'Golden Promise'-Variante als Referenzgenom verwendet, um das *Scaffolding* durchzuführen. Die Gesamtlänge der generierten Pseudomoleküle ohne Berücksichtigung der nicht zugeordneten *Contigs* beträgt 4.267,9 Mb für die Donor-Linie und 4.244,7 Mb für die Mutante. Nur 76 Mb der Donor-Linie und 61 Mb der Mutante konnten keinem der Pseudomoleküle zugeordnet werden. Insgesamt wurden 486 bzw. 534 *Contigs* für das *Scaffolding* und den Aufbau der Pseudomoleküle verwendet. Der Größenunterschied zwischen den beiden Assemblierungen wird auf technische Limitationen bei der Zuordnung von repetitiven Sequenzen z.B. von Telomeren, Subtelomeren, Zentromeren und ribosomaler DNA zurückgeführt (Navratilova *et al.* 2022). Die 1 bp-Deletion in der 5'-UTR des *PDIL5-1*-Gens der Mutante konnte bestätigt werden, während diese Mutation im Wildtyp - ebenfalls erwartungsgemäß - nicht gefunden wurde.

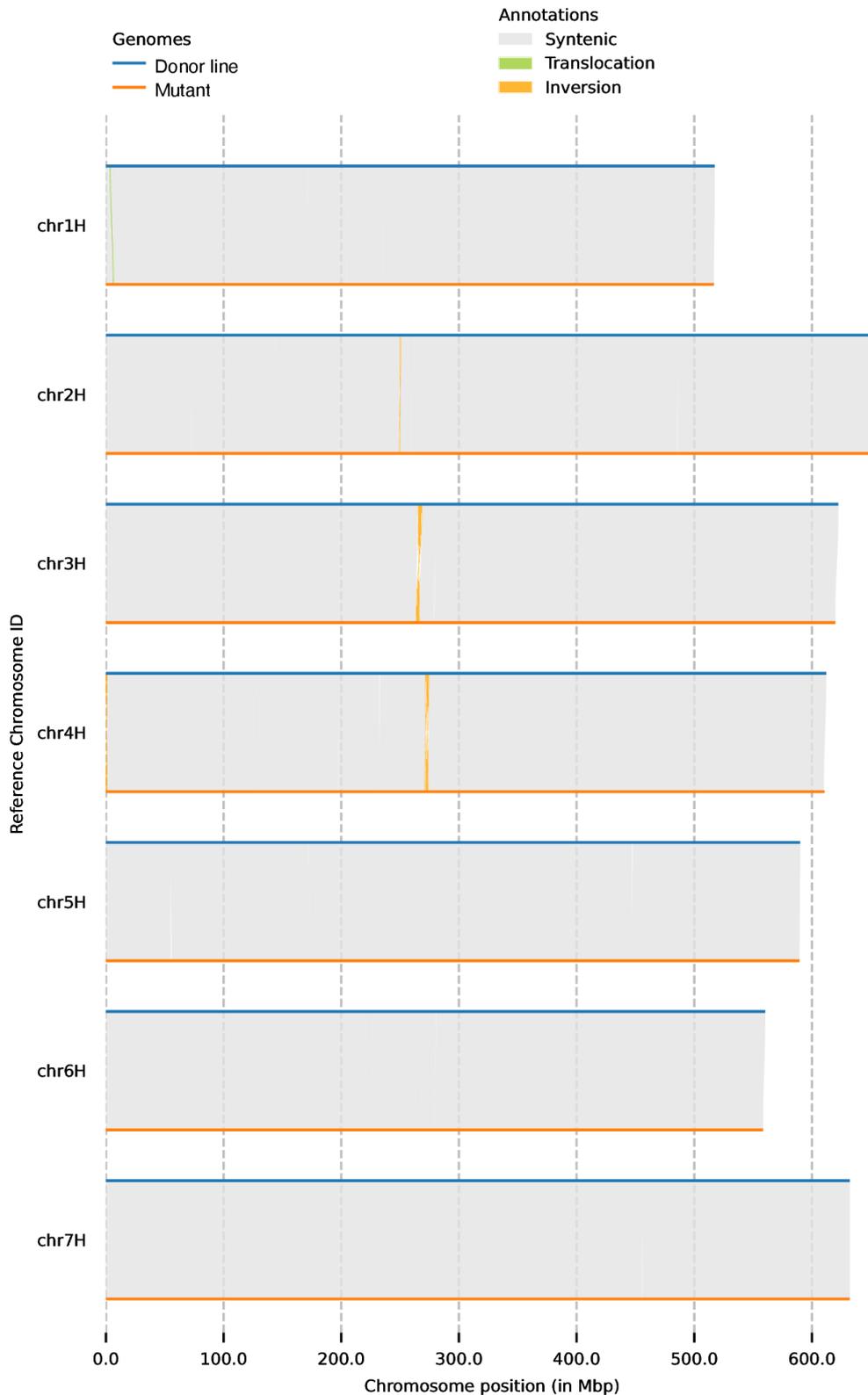


Abbildung 5: Synteny-Plot zwischen der Donor-Linie und der Mutante. Es werden nur Strukturveränderungen gezeigt, die länger als 300 kbp sind.

Die Ermittlung von Sequenzvarianten nach der Assemblierung zu Pseudomolekülen ergab, dass insgesamt 48 homozygote und 80 heterozygote SNPs und 358 homozygote und 1.730 heterozygote InDels in der Mutante im Vergleich zur Donorlinie identifiziert wurden. Daraus ergibt

sich eine Gesamtzahl von 2.088 chromosomal zugeordneten Sequenz-polymorphismen zwischen Mutante und Donorlinie.

Die Assemblierung auf Chromosomenebene ermöglichte zudem eine zuverlässige Suche nach strukturellen Variationen und chromosomalen Neuordnungen, da die Assemblierungen auf *Contig*-Ebene keine Identifizierung von Lücken zwischen verschiedenen *Contigs* ermöglichen (Jayakodi *et al.* 2021). Zur Überprüfung auf strukturelle Variation und Syntenie wurde die Pipeline SyRI (Goel *et al.* 2019) verwendet, die alle Arten von Chromosomen-Neuordnungen identifizieren kann. Die beiden Genome zeigten eine vergleichsweise gute Syntenie (Abbildung 5) mit einigen kleinen Lücken, einer Translokation von 636.105 bp in der telomerischen Region des kurzen Arms von Chromosom 1H und einige Inversionen länger als 700 kbp in den Zentromer-Regionen der Chromosomen 2H, 3H und 4H sowie in der telomerischen Region von 4H. Diese Regionen enthalten repetitive Sequenzen, die sich nur schwer zusammensetzen lassen, worauf wohl auch die wesentlichen Unterschiede zwischen den Assemblierungen beider Genome beruhen. Bislang gibt es keine Genomassemblierung von Gerste, in der solche telomeren und zentromeren Sequenzen verlässlich aufgelöst sind. Selbst die MorexV3-Zusammenstellung (Gersten-Referenz) weist in diesen Regionen Lücken auf (Navratilova *et al.* 2022). Es gibt auch einige Inversionen, die kleiner als 1 kbp sind und sich in der Nähe telomerischer und zentromerischer Regionen befinden.

Abgesehen von diesen Unterschieden, die wahrscheinlich auf technische Limitationen bei der Assemblierung repetitiver Regionen zurückzuführen sind, gibt es keine anderen erkennbaren strukturellen Variationen zwischen den beiden Genomen, was darauf hindeutet, dass die Cas9-vermittelte, gezielte Mutagenese keine chromosomalen Neuordnungen verursacht hat.

3.1.2. Ermittlung von Mutations-gekoppelten Sequenz-Polymorphismen in Gerste

Es wurde nach Sequenz-Varianten in der Nähe der Mutationsstelle im assemblierten Genom der Donor-Linie gesucht, die als Marker verwendet werden könnten. In Tabelle 5 sind die nächstgelegenen SNPs und InDels mit einer maximalen Entfernung von 1,4 Mb von der Mutationsstelle gelistet, wobei der nächstgelegene SNP 522.609 bp strangaufwärts liegt (Nummer 1 in Tabelle 5). Die erzeugte 1 bp-Deletion ist in der Tabelle als Nummer 4 erfasst. Der nächstgelegene InDel liegt 1,4 Mb strangabwärts (Tabelle 5). Das *PDIL5-1*-Gen befindet sich auf Chromosom 4HL in einer distalen Region, mit einer meiotischen Rekombinationsrate von etwa 1 cM/Mb (Mascher *et al.* 2017). Der nächstgelegene SNP ist 522,6 kb von der Mutationsstelle entfernt, was etwa 0,5 cM und in etwa einer Rekombination pro 200 Gameten entspricht. Eine wissenschaftlich fundierte Identifikation des für eine Mutation verantwortlichen (technischen oder spontan verlaufenden) Prozesses ist nach heutigem Verständnis nur denkbar, wenn mit einer gezielt vorgenommenen Mutation eine weitere, ausreichend eng gekoppelte, ungerichtete genetische Veränderung einhergegangen ist. Im Fall der hier verwendeten *pdil5-1*-Mutante ist die Etablierung eines Identifikations-Assays demnach nicht möglich, da der nächstgelegene SNP zu weit von der Mutationsstelle entfernt ist. Zudem kann aus unseren Analysen geschlossen werden, dass eine solche Identifikation Genom-edierter Pflanzen nicht generell möglich ist.

Tabelle 5: SNPs und Indels in der Nähe der Mutationsstelle im Chromosom 4H. Es wurde in einem Intervall von 10 Mb vor und nach der Position des Zielmotivs nach Sequenzvariationen gesucht. Die in der Spalte Genotyp mit * gekennzeichneten Varianten entsprachen nicht dem Coverage-Filter (sie wiesen eine geringere Coverage als den verwendeten Schwellenwert auf). Hom = homozygot; Het = heterozygot.

#	Position	WT-Allel	MT-Allel	Genotyp	Entfernung [bp]
SNPs					
1	454494874	C	T	Hom *	-522.609
2	456466000	C	T	Hom *	1.448.517
Indels					
3	453540821	TGGG	TGG	Het	-1.476.662
4	455017483	ACCCC	ACCC	Hom	0
5	456443647	CAAAAAA	CAAAAAAA	Het	1.426.164

3.1.3. Ermittlung von Transgenrückständen und *Off-Targets* in Gerste

Mit einem *k-mer* Kartierungsansatz wurde untersucht, ob Fragmente des Transformationsvektors unbeabsichtigt in das Genom der *pdil5-1*-Mutante 16-12 integriert worden sind. Für diesen Ansatz wurde eine Kollektion von Fragmenten mit je 22 nt Länge *in silico* generiert, die mit jeder einzelnen Base des Vektors beginnend, überlappend den gesamten verwendeten Transformationsvektor abdecken (22-Mere). Diese Suche ergab, dass keine rekombinante DNA, also keine Transgenrückstände in der Mutante detektierbar sind. Für die *pdil5-1*-Mutante 16-12 ist somit neben der experimentellen Ermittlung mittels PCR zusätzlich durch die WGS belegt, dass sie frei von rekombinanter DNA ist. Im Umkehrschluss bedeutet dies zudem, dass für die Identifizierung dieser mittels Genom-Editierung erzeugten Mutante keine Transgenrückstände herangezogen werden können.

Des Weiteren wurde mittels CCTop nach möglichen *Off-Targets* gesucht. Das Zielmotiv des Cas9/gRNA-Komplexes besteht aus 19 bp mit Homologie zum 5'-Ende der gRNA und der 3 bp umfassenden, von Cas9 erkannten PAM. Bei dieser *Off-Target*-Analyse, die bis zu vier Fehlpaarungen innerhalb des Zielmotivs erlaubte, wurden 74 potenzielle *Off-Targets* gefunden. Um die Koordinaten der Referenz (Donor-Linie) direkt mit denen der Mutante zu vergleichen, wurden diese potenziellen *Off-Targets* in der VCF-Datei überprüft. Dabei wurde nach Sequenzvarianten (sowohl SNPs als auch Indels) im 22 bp umfassenden Zielmotiv jedes möglichen *Off-Targets* gesucht. Es wurden selbst ohne Anwendung der Coverage-Filter keine Sequenzvarianten gefunden. Darüber hinaus wurden auch die Bereiche von 100 bp vor und nach den potenziellen *Off-Target*-Stellen überprüft, wobei ebenfalls keine Sequenzvarianten gefunden wurden.

3.1.4. *Whole-Genome Sequencing*, Hi-C Sequenzierung und Assemblierung der Genomdaten in Raps

Die WGS-*Contig*-Assemblierungen der beiden Mutanten C3E3 und C3E4 und der vom BVL zur Verfügung gestellten Donor-Linie 'Mozart' haben in etwa die für Raps erwartete Größe von ca. 1 GB, wenngleich die Assemblierungen sehr fragmentiert waren (Tabelle 6). Um Assemblierungen im Chromosomenmaßstab zu erhalten und zu überprüfen, ob große strukturelle Veränderungen vorliegen, wurde zusätzlich eine Hi-C-Sequenzierung durchgeführt.

Tabelle 6: Assemblierung der WGS-Reads von der Donor-Linie 'Mozart' und der beiden Mutanten C3E3 und C3E4 in *Contigs*. Diese Assemblierung wurde mit hifiasm für homozygote Genome generiert.

Parameter	Donor-Linie Mozart	Mutante C3E3	Mutante C3E4
Gesamtlänge (nt)	1.324.785.064	1.128.909.168	1.072.337.571
Anzahl der <i>Contigs</i>	6.424	3.304	1.382
N50 (nt)	8.528.487	16.799.602	17.241.869
N90 (nt)	46.077	93.406	2.581.911
Durchschnittliche <i>Contig</i> -Größe (nt)	206.224	341.679	775.931
Maximale <i>Contig</i> -Größe (nt)	31.264.218	60.016.470	49.027.424

Die Hi-C-Sequenzierung wurde mit der Donor-Linie 'Mozart' und Geschwisterpflanzen der Mutante C3E4 durchgeführt, da Projektmittel für die Sequenzierung von nur zwei Linien zur Verfügung standen. Für den *Guide* wurde die Assemblierung von Raps cv. Darmor-bzh (Rousseau-Guetin *et al.*, 2020) verwendet. Die Assemblierungen wiesen einen Unterschied von 300 Mb in der endgültigen Länge auf, der auch in der *Contig*-Assemblierung beobachtet worden war. Die Donor-Linie hat eine kleinere N50, was sich in mehr nicht zugeordneten *Contigs* und mehr *Contigs* pro Pseudomolekül widerspiegelte. Die Größen der Pseudomoleküle A01 und A02 variierten erheblich zwischen den Assemblierungen der Donor-Linie und der Mutante (Abbildung 6). Um nach strukturellen Variationen und chromosomalen Neuordnungen zu suchen, wurden die Pseudomoleküle der Donor-Linie 'Mozart' und der Mutante C3E4 zueinander ausgerichtet. Es wurde SyRI verwendet, um die Syntenie zwischen den beiden Assemblierungen zu überprüfen. In den meisten Regionen weisen beide Assemblierungen eine gute Syntenie auf. Einige zentromerische und telomerische Regionen weisen jedoch Duplikationen, Translokationen und Inversionen auf (Abbildung 6). Diese Unterschiede könnten darauf zurückzuführen sein, dass diese Regionen schwieriger zu assemblieren sind, aber auch darauf, dass Raps polyploid ist und ein gewisses Maß an Heterozygotie aufweist. Aus diesen Daten kann nicht geschlossen werden, ob die beobachteten Variationen auf das gezielte Mutationsereignis oder auf das Assemblierungsverfahren zurückzuführen sind, das nicht speziell für heterozygote Genome konzipiert wurde.

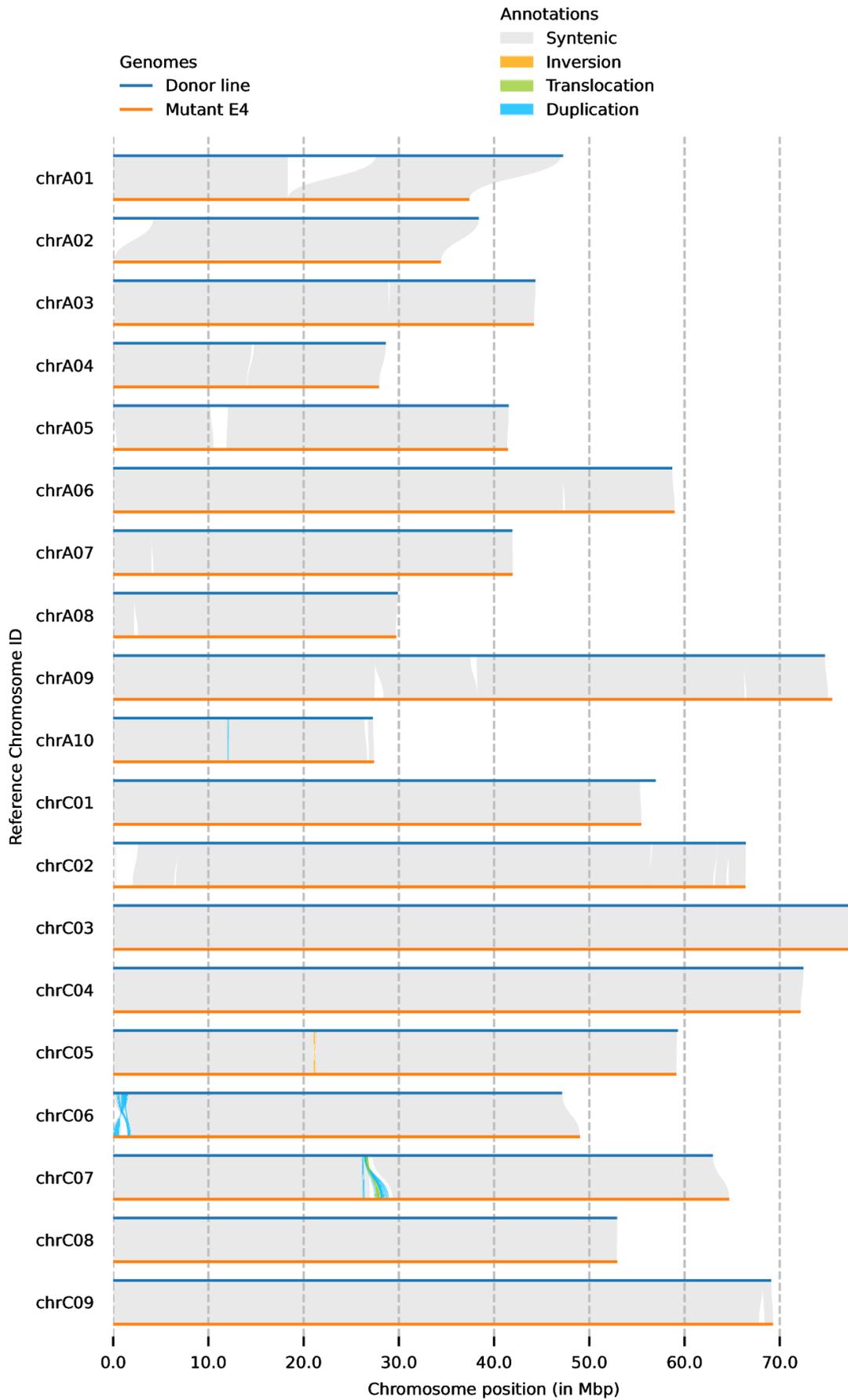


Abbildung 6: Syntenie-Plot zwischen der Donor-Linie und der Mutante C3E4. Es werden nur strukturelle Veränderungen gezeigt, die länger als 30 kbp sind.

3.1.5. Ermittlung von Mutations-gekoppelten Sequenz-Polymorphismen in Raps

Bei der Assemblierung von Haplotypen wurde versucht, beide A09-Schwesterchromosomen (auf 2n-Basis) bis zur maximal möglichen Länge zu assemblieren, um Ereignis-spezifische Polymorphismen zu identifizieren, die entweder mit der relevanten +1 bp (A09a) oder der -3 bp (A09b) Mutation gekoppelt sind. Während in der C3E3 Mutante beide A09 Chromosomen identisch sind (die +1 bp Mutation also homozygot vorliegt), sollten in der C3E4 Mutante beide Versionen vorliegen und anhand des Mutationsorts zu unterscheiden sein. Allerdings wurde in dieser Mutante nur die A09b Version assembliert. Damit stehen aber beide Versionen für einen Vergleich mit dem WT Mozart zur Verfügung. Mit einer Länge von ca. 13.600.000 bp entsprechen diese *Contigs* in etwa 25% der Gesamtlänge des Raps A09 Chromosoms. Ein Vergleich zeigte, dass die meisten Unterschiede zwischen dem Wildtyp und den beiden Mutanten bestehen, welche untereinander sehr ähnlich sind (beide gehen auf das gleiche Ausgangs-Event C3 zurück). Viele gefundene Polymorphismen sind 1 bp InDels und es ist sehr wahrscheinlich, dass es sich dabei um Artefakte handelt, die bei der Sequenzierung in Bereichen mit vielen Wiederholungen einzelner Nukleotide auftreten können. Für den hier untersuchten Bereich konnten zwischen den A09a (aus C3E3) und A09b (aus C3E4) Chromosomen 10 potenzielle Marker identifiziert werden, von denen der nächste 3.370 bp neben der -3 bp Mutation (A09b) lag (Tabelle 7, Nr. 4; Abbildung 7). Diese 3.370 bp stellen eine verhältnismäßig große Entfernung dar und Polymorphismen mit einer solchen Distanz würden normalerweise nicht als eng gekoppelt angesehen werden. Um eine Aussage treffen zu können, in welchem Schwesterchromosom der Unterschied aufgetreten ist, wurden die betreffenden Stellen im Mozart WT und 3 weiteren Raps Genomen gesucht und falls dort gefunden als das natürliche Allel definiert. Das andere ist damit automatisch die Mutation (rot markiert). Keine dieser Mutationen liegt nahe einer dem Zielmotiv der verwendeten gRNA ähnlichen Sequenz und somit handelt es sich höchstwahrscheinlich um spontan aufgetretene Mutationen. Für die Berechnung der Rekombinationsrate wurden publizierte Vergleichswerte der Genomgröße in cM herangezogen, welcher durch die Genomgröße in bp dividiert und mit dem Abstand zwischen Mutation und Marker in bp multipliziert wird (Tabelle 7). Bei einer Rekombinationsrate von 0,01 % ist es zwar recht unwahrscheinlich, dass die Kopplung durch ein Crossing-over gebrochen wird, aber eben doch nicht auszuschließen. Statistisch gesehen sollte in einer von 10.000 Pflanzen ein solcher Kopplungsbruch erfolgen, d.h. man könnte auf einem 1 ha großen Rapsfeld ca. 5 solche Crossing-over-Ereignisse erwarten.

Tabelle 7: Gefundene Event-spezifische Polymorphismen (in rot) in den die Cas9- induzierten Mutationen enthaltenen A09 *Contigs* mit ca. 13.600.000 bp Länge

#	≈ Lage	Typ	A09a (+1) C3E3	A09b (-3) C3E4	p REK (%)
1	-5007703	SNP	T	C	17.85
2	-1258389	InDel	-1 bp		4.48
3	-1047828	SNP	C	T	3.73
4	3370	SNP	A	T	0.01
5	762678	SNP	G	T	2.72
6	771933	InDel		-18 bp	2.75
7	782720	InDel		-13 bp	2.79
8	974374	SNP	T	G	3.47
9	1243103	SNP	T	A	4.43
10	2493529	SNP	C	T	8.89

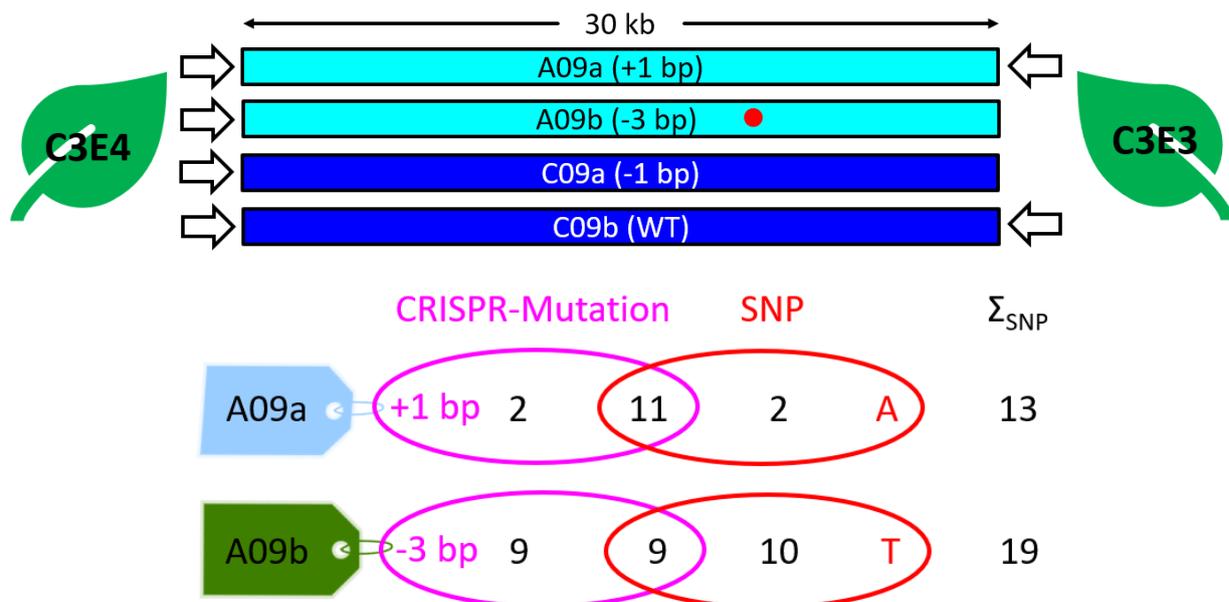


Abbildung 7: Der mit der Cas-induzierten -3 bp Mutation assoziierte SNP (● T statt A) ist eindeutig mit dieser gekoppelt, da 9 physische PacBio-*Reads* beide Marker enthalten. Für beide Versionen gab es in der C3E4 Mutante ähnlich viele *Reads*, sodass es sich aufgrund des beobachteten 50:50 Verhältnisses nicht um Sequenzierfehler handeln dürfte.

Ein Vergleich dieser SNP-Region über das PacBio Raps Pan-Genom (plus der beiden „Raps-Eltern“ *Brassica oleracea* und *Brassica rapa*) zeigte bereits, dass es sich hier wahrscheinlich um einen *Event*-spezifischen SNP handelt, da dieser auch nicht im Mozart Wildtyp Genom zu finden war. Dieser gekoppelte SNP könnte somit für die Etablierung eines Identifikations-Assays durch die Kombination zweier *short-read* Illumina-Assays (einem zum Nachweis der -3 bp Mutation und einem zum Nachweis des SNPs) verwendet werden.

3.1.6. Ermittlung von Transgenrückständen und *Off-Targets* in Raps

Zu Projektbeginn wurden zwei Cas-induzierte Mutanten aus der zweiten Generation (T2) für das Projekt mittels PCR selektiert, wobei die C3E3 Mutante eindeutig transgen (aber nur im A09 Locus homozygot mutiert) war, während die C3E4 Mutante bereits Transgen-frei zu sein schien. Durch die Analyse einzelner Reads aus dem PacBio-WGS, sowie ersten *Assembly*-Versuchen an der CAU konnten in beiden Mutanten allerdings einige T-DNA Integrationsorte bestimmt werden. Überraschenderweise handelte es sich zumeist um T-DNA Fragmente, während das komplette Konstrukt nur einmal in der C3E3 Mutante gefunden wurde. Insgesamt konnten 6 Loci identifiziert werden, welche unvollständige T-DNA enthalten (Tabelle A7). Zwei der T-DNA Fragmente sind mit größeren Chromosomen-Aberrationen assoziiert und ein Genort (A02) beherbergt neben der kompletten T-DNA noch 4 Bruchstücke, von denen zwei invertiert vorliegen. Die erfolgreiche PCR-Amplifikation aller getesteten 5 Integrationsstellen in 10 T1-Pflanzen deutet daraufhin, dass die Fragmentierung und multiple Integration bereits in der Gewebekultur / T0-Ursprungspflanze stattgefunden hat. Nach Selbstung kann rekombinante DNA durch Segregation rausgekreuzt werden und es ist wahrscheinlich, dass in Folgegenerationen mutierte, nicht-transgene Pflanzen selektiert werden können.

Um mögliche *Off-Target* Effekte durch die Genom-Editierung zu ermitteln, wurden zunächst alle *CRT1a* Kopien der beiden Mutanten auf den Chromosomen A09, C09, A01 und C01 untersucht und mit keiner der angewendeten Methoden (Sanger, Illumina oder PacBio) wurden hier Mutationen in den beiden als *Off-Targets* definierten Kopien A01 oder C01 entdeckt. Auch für 18 weitere putative *Off-Targets* (Abbildung 8) konnte nur das WT-Allel in den beiden Mutanten gefunden werden.

	nearby locus	location
1	BnaC05g09240D	intronic
2	BnaC04g27090D	intergenic
3	BnaC01g25880D	intergenic
4	BnaCnng35850D	intergenic
5	BnaA05g04520D	intergenic
6	BnaA10g08750D	intergenic
7	chrAnn_random	intergenic
8	BnaC06g21580D	exonic
9	BnaA07g21220D	exonic
10	BnaA10g11140D	intergenic
11	BnaC04g13430D	intergenic
12	BnaC08g05190D	intergenic
13	BnaA04g22450D	intergenic
14	BnaCnng29210D	intergenic
15	BnaC07g32040D	intergenic
16	BnaC09g28620D	intergenic
17	C04:-14410065	intergenic
18	A05:+12183708	intergenic
	WT in C3E3	WT in C3E4

Abbildung 8: Putative *Off-Targets* in Raps. Die Sequenzen von 1 und 2, 6 und 7, sowie 8 und 9 sind identisch und finden sich zweimal im Genom

3.2. PCR-basierte Etablierung der Detektion von Einzelnukleotid-Veränderungen

Einzelnukleotid-Veränderungen können in einer PCR-Reaktion nachgewiesen werden, sofern die verwendeten Oligonukleotide spezifisch genug sind. In diesem Projekt wurden ausschließlich fluoreszenzmarkierte Hybridisierungs sonden verwendet. Durch die Verwendung von *Locked Nucleic Acids* (LNAs) oder *Minor Groove Binder* (MGB) können die Hybridisierungseigenschaften verbessert werden und auch Oligonukleotide von nur 13-20 bp Länge ausreichend gut funktionieren. Für die GVO-Analytik in Referenzlaboren werden häufig kompetitive duplex Assays verwendet (Olav Grundmann, Auftragnehmer Impetus Bioscience; Fraiture *et al.*, 2022). Dafür werden zwei Sonden, die an die mutierte oder die Wildtyp-Sequenz binden, für die gleichzeitige Quantifizierung von Wildtyp- und mutierten Sequenzkopien aus demselben PCR-Amplikon verwendet (ENGL 2019). Ähnlich funktioniert das ebenfalls oft verwendete *Drop-off Assay*, bei dem die Referenz-Sonde und die *Target-* oder SNV-Sonde im selben PCR-Amplikon binden (Mock *et al.*, 2016; Falabella *et al.*, 2017). Die Referenz-Sonde bindet außerhalb der veränderten Region und somit an die mutierte und die Wildtyp-Sequenz, während die SNV-Sonde entweder nur an die mutierte oder die Wildtyp-Sequenz bindet.

3.2.1. qPCR in Gerste

Es wurde zunächst ein kompetitives Assay (Abbildung 9A, Olav Grundmann, Auftragnehmer Impetus Bioscience) und ein *Drop-off Assay* (Abbildung 9B, Martin Becker, IPK/ Stilla) entwickelt und in qPCR-Experimenten getestet. Dabei wurden sowohl normale Sonden als auch MGB- und LNA-Sonden verwendet.

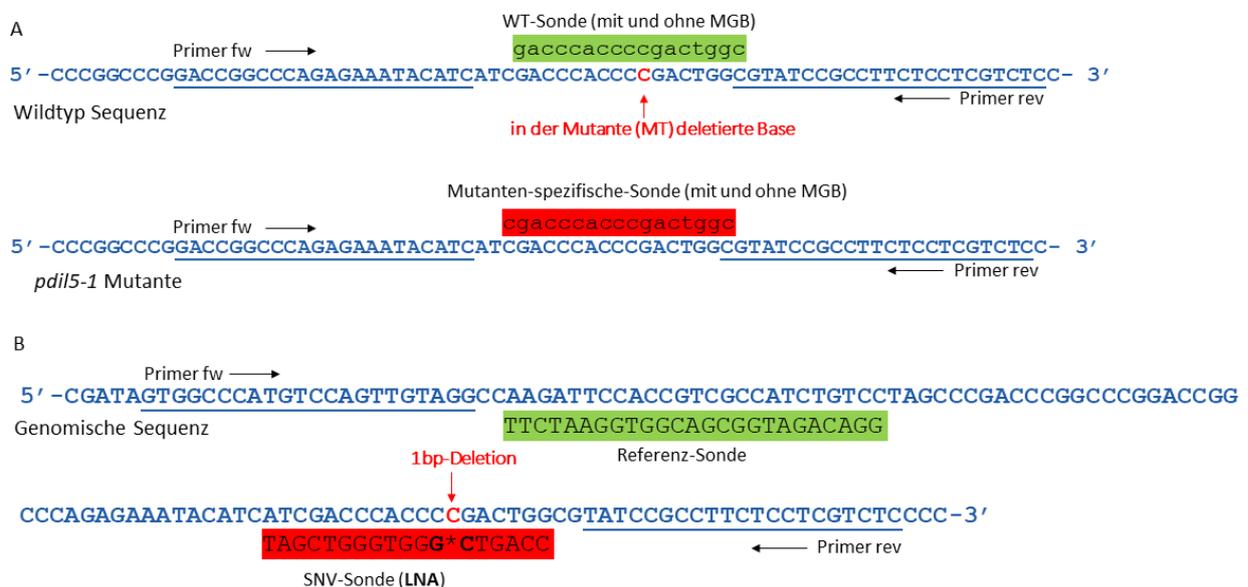


Abbildung 9: Design des kompetitiven (A; System 1 und 2) und des *Drop-off Assays* (B; System 3)

In den qPCR-Assays kam es zu Kreuzreaktionen der Sonden aller Systeme (Abbildung 9). Lediglich die WT-MGB-Sonde (System 1) und die normale WT-Sonde (System 2) haben relativ spezifisch nur in der WT-Probe gebunden (Tabelle 8). Alle anderen theoretisch spezifischen Sonden wurden sowohl in den Wildtyp- als auch in der Mutanten-Probe detektiert. Die repräsentativen Ergebnisse der Systeme 1-3 unter Verwendung von 100 ng genomischer DNA

und dem Kapa Probe Force Master Mix sind in Tabelle 8 dargestellt. Bei diesen starken Kreuzreaktionen ist eine eindeutige Berechnung des Anteils an Mutanten-DNA nicht möglich.

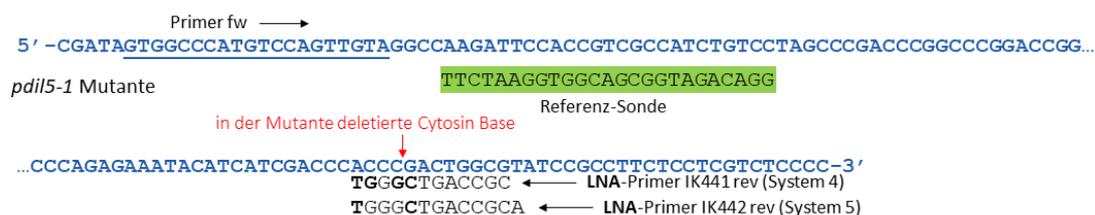
Tabelle 8: Repräsentative qPCR Ergebnisse der Systeme 1 bis 3 unter Verwendung von 100 ng gDNA pro Reaktion und dem Kapa Probe Force Master Mix. n=3, ND=nichts detektiert, CV= Variationskoeffizient der gemittelten Cq-Werte.

Detektion	System 1				System 2				System 3			
	WT-Sonde		MT-Sonde		WT-Sonde		MT-Sonde		Referenz-Sonde		MT-Sonde	
Probe	Cq	CV [%]	Cq	CV [%]	Cq	CV [%]						
Wildtyp	26,24	0,25	30,44	1,44	24,55	0,70	28,25	0,05	26,41	0,71	24,56	0,03
Mutante	38,67	1,22	25,35	1,80	ND	ND	24,90	0,14	25,43	0,23	23,71	1,08
H ₂ O	ND	ND	ND	ND	41,64	3,87	ND	ND	40,95	0,01	ND	ND

Diese Kreuzreaktionen treten in geringerem Maß auch bei der digitalen PCR auf, jedoch lassen sich die falsch-positiven *Droplets* von den korrekt-positiven *Droplets* durch die Höhe und die Separation der Cluster unterscheiden. Aufgrund anderer Prinzipien der Datenerhebung und -auswertung ist eine solche Differenzierung mittels qPCR nicht möglich.

Des Weiteren wurden zwei Assays mit zwei verschiedenen reversen LNA-Primern, die auf der Mutation liegen, getestet (System 4 und 5, Abbildung 10). Neben den Assays für die Zielsequenz-spezifische Mutation wurde ein Gersten-Referenzsystem basierend auf der Publikation von Hernandez *et al.* (2005) verwendet (System 6). Das ursprünglich für ein qPCR-Assay entwickelte Referenzsystem amplifiziert ein 73 bp langes Fragment aus dem γ -*Hordein* (*Hor3*)-Gen und wurde anhand von 10 repräsentativen Gerstensorten getestet (Hernandez *et al.*, 2005).

A



B

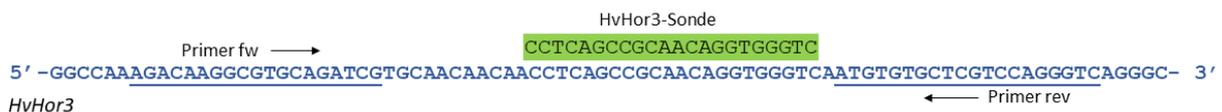


Abbildung 10: Design der Systeme 4 und 5 (A) und des HvHor3 Referenz-Assays (B; System 6)

Parallel zu den Systemen 4 und 5 wurde immer das HvHor3-Referenzassay verwendet, um den Anteil an WT- und Mutanten-DNA abschätzen zu können. Die repräsentativen Ergebnisse dieser Experimente sind in Tabelle 9 dargestellt. Der erste LNA-Primer (IK441) enthält 4 LNAs, die zum einen um die Deletion und zum anderen am 3'-Ende des Oligonukleotids platziert wurden. Mit dem ersten LNA-Primer (System 4) konnte weder mit Plasmid-DNA noch mit gDNA eine

Amplifikation erzielt werden. Auch Optimierungen der PCR-Konditionen führten zu keiner Amplifikation unter Verwendung dieses Primers. Eine Analyse des Primers zeigte keine bzw. nur geringfügige Selbst-Dimer- bzw. Hetero-Dimer-Bildung mit den anderen in dem Assay verwendeten Oligonukleotiden.

Tabelle 9: Repräsentative qPCR Ergebnisse der Systeme 4 bis 6 unter Verwendung von 5 ng pDNA oder 100 ng gDNA pro Reaktion und dem Kapa Probe Force Master Mix. n=3, ND=nichts detektiert, CV= Variationskoeffizient der gemittelten Cq-Werte, - nicht getestet, da Primer/Sonden im Plasmid nicht binden

Detektion	System 4		System 5		System 6	
	Cq	CV [%]	Cq	CV [%]	Cq	CV [%]
Plasmid WT	ND	ND	30,34	0,73	-	-
Plasmid MT	ND	ND	13,61	1,15	-	-
gDNA WT	ND	ND	37,31	0,19	22,26	0,41
gDNA MT	ND	ND	27,12	0,14	22,04	0,08
H ₂ O	ND	ND	42,36	3,02	ND	ND

Ein direkter Vergleich mit System 3 und System 6 im selben qPCR-Durchlauf zeigte, dass es weder zu Problemen mit dem generellen qPCR-Experiment noch mit der DNA gekommen war. Der zweite LNA-Primer (IK442) hatte eine LNA-Base direkt vor der Deletion (5'-3') und eine am 3'-Ende des Oligonukleotids. Mit diesem System 5 konnte mit der Plasmid-DNA als Positivkontrolle eine „normale“ Amplifikation erzielt werden. Allerdings kam es zu einer geringen Kreuzreaktion und WT-Amplifikate wurden ebenfalls erzeugt (Tabelle 9, System 5), obwohl der LNA-Primer nur in der Mutante binden sollte. Im Vergleich mit dem Referenzsystem 6 zeigte sich, dass das System 5 bei der Verwendung von gDNA weniger effizient war. Dennoch war zu sehen, dass es mit dem System 5 auch bei der Verwendung von gDNA zu Kreuzreaktionen kommt.

Eine Besonderheit des Zielmotivs im *PDIL5-1* Gen ist das Vorhandensein von vier Cytosin-Nukleotiden an der Schnittstelle der Cas9. Durch den Cas9-induzierten Doppelstrangbruch an dieser Stelle und einer fehlerhaften Reparatur kam es im Kontext der gezielten Mutagenese zur Deletion eines Cytosins. Jedoch war die entstandene Mutante durch das Vorhandensein der drei verbleibenden Cytosin-Nukleotide im Zielmotiv nur schwer von dem Wildtyp zu unterscheiden. Durch Einsatz von LNA-Sonden ist keine hinreichende Diskriminierung zwischen Wildtyp und Mutation gelungen. LNAs wurden bereits mehrfach für die Diskriminierung von SNPs (Substitution von Nukleobasen) verwendet (u. a. Chhalliyil *et al.*, 2020), allerdings zeigten sich dabei auch Schwierigkeiten bezüglich der Genauigkeit des Nachweises. Die in der GVO-Analytik eingesetzten Methoden müssen zuverlässig zwischen der mutierten Sequenz und dem Wildtyp differenzieren können. Eine besondere Herausforderung ist, dass es sich bei Cas9-induzierten Mutationen in der Regel um InDel-Mutationen handelt. Wurden die LNAs, wie bei den Oligonukleotiden IK438 und IK442, um die Mutationsstelle platziert, kam es zu einer stabilen Bindung dieser Oligonukleotide sowohl mit der Mutanten- als auch der Wildtypsequenz, wobei das Fehlen bzw. die Insertion einer Base diese Bindung offenbar nicht behindert hat. Zumindest haben bei der hier verwendeten Mutante mit einer 1 bp-Deletion in den qPCR-Assays weder die jeweiligen LNA- noch MGB-Sonden bzw. Primer spezifisch nur an die Mutantensequenz gebunden. Auf Grund dessen wurde der Ansatz mit diesen Primer/Sonden-Systemen in einem

qPCR-Assay nicht weiterverfolgt. Angesichts unserer Ergebnisse stellen wir in Frage, dass ein anderes Primer/Sonden-System geeigneter ist, diese spezifische SNV in einem qPCR-Assay zuverlässig detektieren bzw. vom Wildtyp differenzieren zu können. Daraus resultiert auch die verallgemeinerte Einschätzung, dass es mit einem qPCR-Assay nicht generell möglich sein wird, bestimmte SNVs verlässlich genug zu detektieren, für die es aufgrund ihrer Entstehung durch gezielte Mutagenese keine alternativen Polymorphismen gibt. Dennoch sollte jede Mutation und jedes Assay individuell beurteilt werden und die Möglichkeit einer erfolgreichen Detektion ist in Einzelfällen nicht auszuschließen. Für den vorliegenden Fall der *pdil5-1*-Mutante ist es aus objektivem Grund nicht gelungen, ein zuverlässiges qPCR-Assay zu etablieren.

3.2.2. ddPCR in Gerste

Alle drei Sondentypen (normal, MGB, LNA) wurden auch in ddPCR-Assays getestet. Nur mit dem System 2 konnten valide Signale detektiert werden. Dabei gab es eine klare Separierung der positiven von den negativen *Droplets*, jedoch kam es mit diesem System in der ddPCR ebenso wie in der qPCR zu einer Kreuzreaktion der Sonden. Allerdings kann in der digital PCR aufgrund der Separierung der *Droplet*-Cluster klar zwischen den Clustern unterschieden werden, sodass eine konklusive Auswertung möglich ist (Abbildung 11).

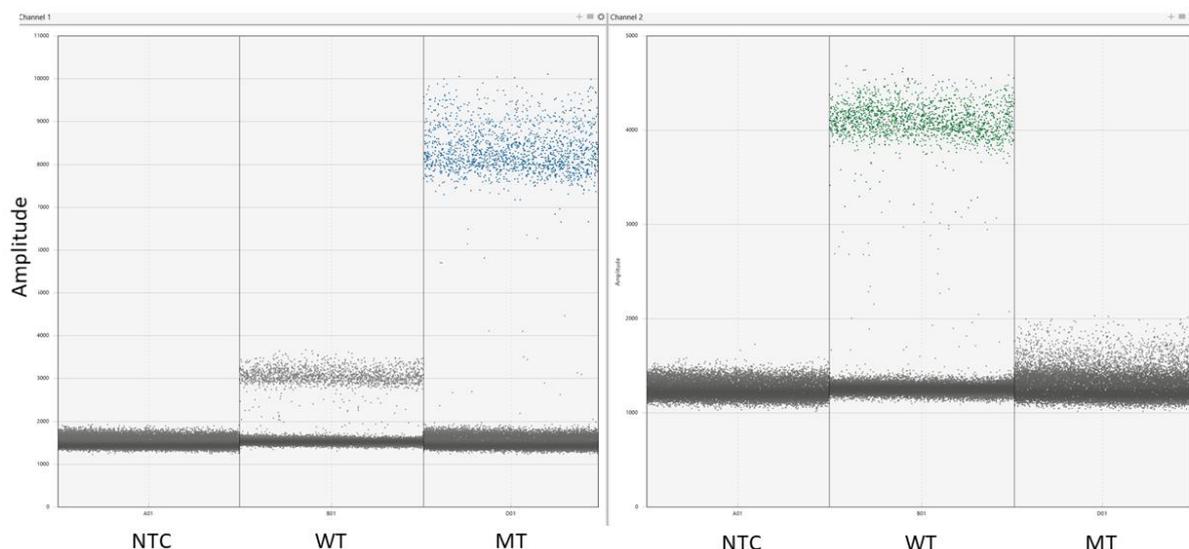


Abbildung 11: Repräsentative Tröpfchenpopulation in der ddPCR unter Verwendung des Systems 2 (Primer 272 fw und rev und Sonden 273_NEU und 274_NEU) und gDNA von 'Golden Promise' DH1-6 (WT), der -1 bp-Mutante 16-12 (MT) und der nicht behandelten Kontrolle (NTC).

Im Vergleich mit dem System 2 waren die Kreuzreaktionen der Sonden von System 3 jedoch zu stark, um eine Unterscheidung zu ermöglichen. Somit war das System 2 das einzige Assay, mit dem eine zuverlässige Diskriminierung zwischen Wildtyp und Mutante gelungen ist. Das System 2 wurde folglich auch für die weitere Etablierung und Validierung des Assays durch Impetus Bioscience ausgewählt. Wie in der Vorhabensbeschreibung aufgeführt, sollen mit dem entwickelten Assay mindestens 20 Kopien der Zielsequenz nachgewiesen werden können. Die qPCR-Assays wurden aufgrund der in 3.1.7. beschriebenen Limitationen nicht weiter getestet. Mit dem ddPCR-Assay (System 2) konnte anhand von 10 Replikaten gezeigt werden, dass

minimal 5 Kopien der Zielsequenz sicher nachweisbar sind (Tabelle 10). Dieses ddPCR-Assay läuft zuverlässig sowohl auf dem Bio-Rad Gerät von Impetus Bioscience als auch auf dem am IPK verfügbaren Stilla Gerät. Sowohl bei Impetus als auch am IPK konnte gezeigt werden, dass mittels des ddPCR-Assays zuverlässig 0,1 % an Beimischung der Mutante in Wildtyp-Hintergrund-DNA detektiert werden kann (Abbildung 12). Dies wurde sowohl anhand von DNA-Mischungen als auch anhand von Saatgutmischungen gezeigt. Die detaillierten Ergebnisse der umfassenden Validierung sind im separaten SOP-Entwurf für ein ddPCR-Assay in Gerste zu finden.

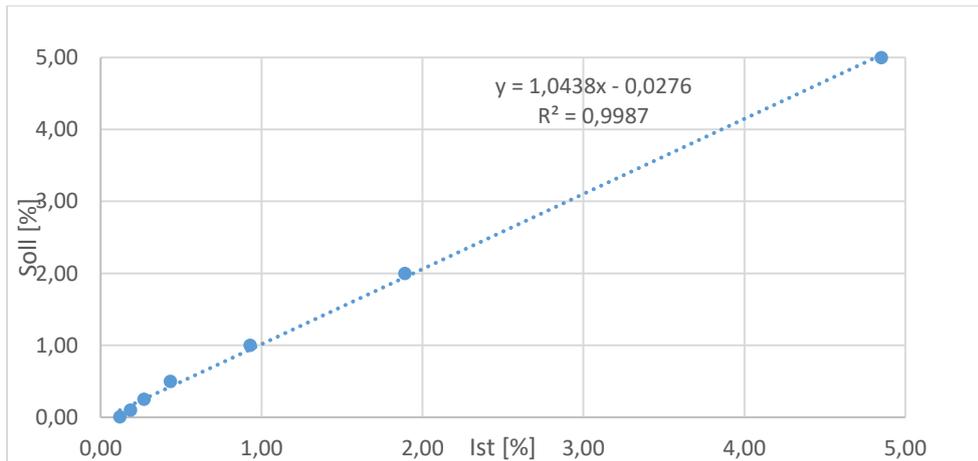


Abbildung 12: Durch ddPCR ermittelte Anteile an Mutanten-DNA in DNA-Mischproben.

Tabelle 10: Sensitivität und Nachweisgrenze (LOD) des ddPCR-Assays in Gerste. 20.000, 10.000, 1.000 und 100 Kopien DNA-Template/PCR-Ansatz wurden in Triplikaten und 20, 10 bzw. 5 Kopien in 10er-Replikaten in 100-200 ng Hintergrund-DNA getestet.

Spezies/GVO Kopien/PCR-Ansatz	WT-Probe			MT-Probe		
	WT		MT	WT	MT	
	Kopien	STABW	Kopien	Kopien	Kopien	STABW
20.000 K	24.457	419,7	nd	nd	20.017	583,6
10.000 K	11.169	1398,2	nd	nd	11.188	912,9
1.000 K	1.007	57,3	nd	nd	1.074	11,2
100 K	79	5,7	nd	nd	84	12,2
20 K	16	6,6	nd	nd	20	4,8
10 K	10	3,0	nd	nd	11	3,4
5 K	5	2,3	nd	nd	4	2,9

3.2.3. ddPCR in Raps

In Analogie zu dem anhand von Gerste entwickelten ddPCR-basierten Detektionsassay wurde auch ein Assay zur Detektion der Cas9-induzierten 1 bp-Insertion im A09 Locus von homozygoten Nachkommen der C3E3 Rapsmutante entwickelt. Das Design des Assays ist in Abbildung 13 dargestellt. Zwei LNA-Primer sollen zwischen dem A09 und dem C09 Locus unterscheiden können. In ersten Experimenten zeigte sich bereits, dass die MBG-Sonden spezifisch nur in den jeweiligen Proben binden und zuverlässig auch geringe Konzentrationen von 0,1 % der Zielsequenz nachgewiesen werden können.

Ähnlich wie für Gerste erfolgte die weitere Etablierung und Validierung des Raps-Assays durch Impetus Bioscience. Die Leistung der entwickelten ddPCR-basierten Methode für Raps wurde in Bezug auf Spezifität (Inklusivität und Exklusivität), Sensitivität (Tabelle 11) Selektivität und Robustheit bewertet. Mit diesem ddPCR-Assay konnte gezeigt werden, dass die Nachweisgrenze bei 5 Kopien der Zielsequenz liegt (Tabelle 11). Die detaillierten Ergebnisse der Validierung sind im separaten SOP-Entwurf für ein ddPCR-Assay in Raps zu finden.

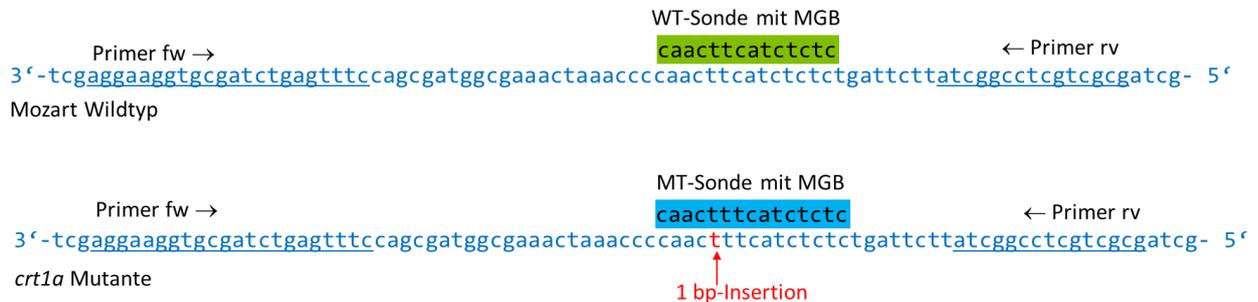


Abbildung 13: Design des ddPCR-Assays für Raps

Tabelle 11: Sensitivität und Nachweisgrenze (LOD) des ddPCR-Assays in Raps. 20.000, 10.000, 1.000 und 100 Kopien DNA-Template/PCR-Ansatz wurden in Triplikaten und 20, 10 bzw. 5 Kopien in 10er-Replikaten in 100-200 ng Hintergrund-DNA getestet.

Spezies/GVO Kopien/PCR-Ansatz	WT-Probe			MT-Probe		
	WT		MT	WT	MT	
	Kopien	STABW	Kopien	Kopien	Kopien	STABW
20.000 K	23.432	1431,5	nd	nd	22.194	360,8
10.000 K	10.686	1355,4	nd	nd	11.174	93,9
1.000 K	1.219	55,4	nd	nd	1.044	55,6
100 K	144	16,4	nd	nd	112	5,8
20 K	32	7,7	nd	nd	19	3,0
10 K	19	8,9	nd	nd	9	2,9
5 K	10	6,0	nd	nd	5	2,8

3.3. Charakterisierung der Zielregionen mittels Tiefensequenzierung von Amplikons

3.3.1. Amplikonsequenzierung in Gerste

Für die Amplikon-Tiefensequenzierung wurde ein Primerpaar (NGS_F3 und NGS_R3, Abbildung 14) ausgewählt, mit dem ein 75 bp großes Fragment des *PDIL5-1* Gens der Gerste amplifiziert wird. Zunächst wurde ein Experiment durchgeführt, um die Sensitivität dieser Methode anhand von DNA-Mischproben zu bestimmen. Es konnte zuverlässig eine Sensitivität von 0,01 % erreicht werden (Tabelle 12; Abbildung 15).

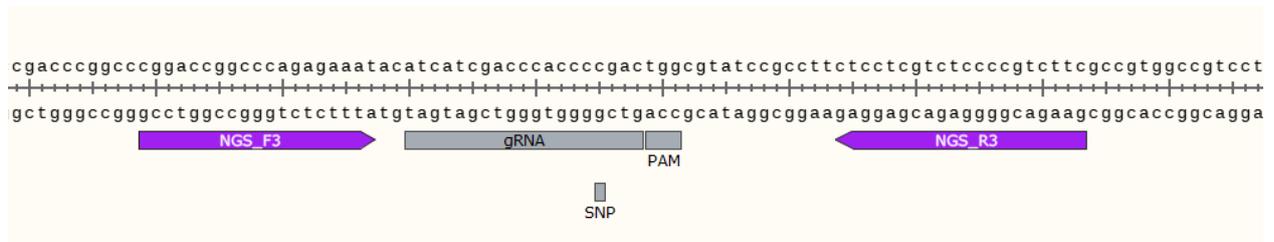


Abbildung 14: Ausgewählte Primer für die Amplikon-Tiefensequenzierung des *PDIL5-1* Zielmotivs

Tabelle 12: Sensitivitätstest anhand von DNA-Mischproben, n = 3 bei 10-1 %, n = 4 bei 0,75-0,25 %, n = 6 bei 0,1–0,01 %

Soll-Anteil von MT-DNA [%]	Ist-Anteil von MT-DNA [%]	STABW
0 (nur WT)	0	0
0,01	0,010	0,011
0,025	0,033	0,006
0,05	0,052	0,001
0,075	0,076	0,001
0,1	0,104	0,003
0,25	0,270	0,014
0,5	0,575	0,053
0,75	0,840	0,064
1	1,097	0,068
5	5,767	0,542
10	11,233	0,872

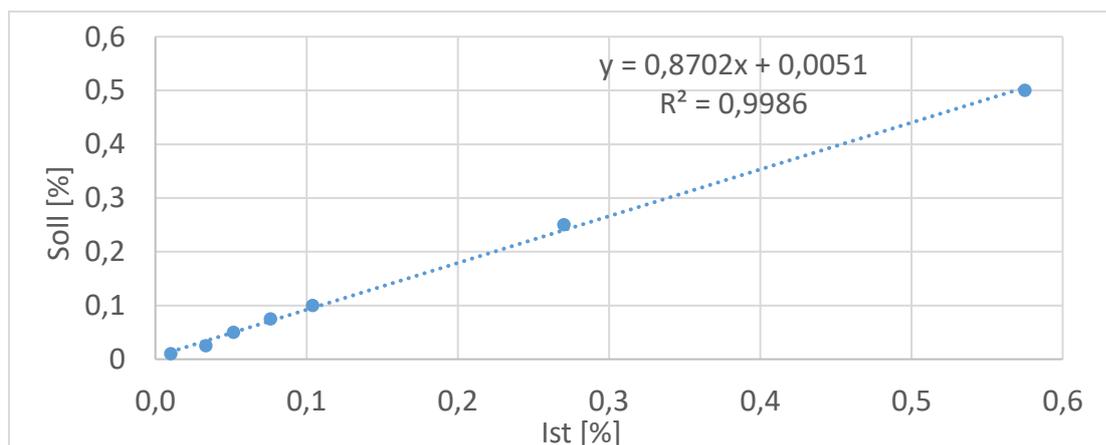


Abbildung 15: Anteil an Mutanten-DNA in DNA-Mischproben ermittelt durch Amplikon-Tiefensequenzierung und dem Primerpaar NGS_F3/NGS_R3

Desweiteren wurde die Spezifität dieses Systems anhand von DNA-Mischproben getestet. Dazu wurden 5 % der Mutanten-DNA mit 95 % der Wildtyp-DNA gemischt und davon dann 10 % in 90 % Hintergrund-DNA von verschiedenen Spezies getestet. Es wurden keine wesentlichen Kreuzreaktionen mit anderen Nutzpflanzen (Raps, Soja, Reis, Mais und Weizen) detektiert. Die Ergebnisse dieses Tests sind in Tabelle 13 dargestellt.

Tabelle 13: Spezifitätstest anhand von DNA-Mischproben; n=2

Hintergrund-DNA	MT-Allel		WT-Allel	
	Anteil [%]	STABW	Anteil [%]	STABW
Raps	5,58	0,05	94,43	0,05
Soja	5,93	0,11	94,07	0,11
Reis	5,79	0,01	94,22	0,01
Mais	5,83	0,24	94,17	0,24
Weizen	6,04	0,05	93,97	0,05
Zuckerrübe	5,83	0,04	94,17	0,04

Für ein weiteres Experiment wurden 0,1-prozentige Saatgutmischungen (2.997 Körner WT + 3 Körner MT) verwendet. Von den 9 Saatgutmischungen wurden jeweils 3-5 DNA-Extraktionen durchgeführt und dann mit jeder Probe 8 Amplikon-Tiefensequenzierungsdurchläufe durchgeführt. Insgesamt wurden auf diese Weise 345 Reaktionen ausgewertet. Die gemittelten Werte der 9 Saatgutmischungen sind in Tabelle 14 aufgeführt. Es konnte gezeigt werden, dass mit dieser Methode ein Anteil von 0,1 % DNA mit der Zielsequenz sicher nachweisbar ist.

Tabelle 14: Sensitivitätstest anhand von Saatgut-Mischproben, n=9

Mischung	MT-Allel		WT-Allel	
	Anteil [%]	STABW	Anteil [%]	STABW
1	0,12	0,01	99,60	0,04
2	0,11	0,03	99,62	0,03
3	0,11	0,03	99,61	0,03
4	0,10	0,03	99,63	0,03
5	0,10	0,02	99,63	0,02
6	0,10	0,02	99,62	0,03
7	0,13	0,03	99,61	0,03
8	0,11	0,02	99,62	0,03
9	0,12	0,02	99,61	0,02

3.1.2. Amplikonsequenzierung in Raps

Da Raps im Vergleich zu Gerste eine allotetraploide Pflanzenart ist, war die Entwicklung von spezifischen Assays ungleich schwerer. Aufgrund der Anzahl von insgesamt 4 *CRT1a*-Loci in Raps und der Unklarheit über die Vollständigkeit der zuvor erhobenen Daten aus der Sanger-Sequenzierung wurden zunächst verschiedene Primer entwickelt und in unterschiedlichen Kombinationen getestet, mit dem Ziel alle 4 Gen-Loci in gleich guter Weise zu amplifizieren. Die Ergebnisse sind in Tabelle 15 dargestellt.

Tabelle 15: Prozentuale Anteile der erzeugten *Reads* mit den bekannten Mutationen bzw. WT Sequenzen (C3E4)

MiSeq Run	% A09a +1	% A09b -3	% C09a -1	% C09b WT	% A01 WT	% C01 WT	% Verluste	# Reads
Assay #1	5,17	4,96	5,85	5,80	25,42	29,45	23,34	18537
Assay #2	2,43	2,68	8,26	8,32	25,84	32,68	19,78	19625
Assay #3	9,77	10,74	12,16	11,84	16,59	14,32	24,58	23430
Assay #4	5,50	5,63	18,80	18,28	17,40	13,57	20,81	16222
Erwartung	12,50%	12,50%	12,50%	12,50%	25,00%	25,00%	Chimären	

Mit keinem der vier zunächst getesteten Assays wurde eine gleichmäßige Amplifikation aller vier Loci erzielt. Zudem kam es zu Verlusten von 20 bis 25 %, welche auf Chimärenbildung während der PCR zurückzuführen sind. Daher wurde ein neuer Assay (#5) entworfen, in welchem nur noch die *On-Targets* amplifiziert werden. Mit diesem Assay wurden verschiedene Rapslinien und die beiden Mutanten C3E3 und C3E4 analysiert (Tabelle 16). Beide Mutanten wiesen in etwa die erwarteten Verhältnisse für die mutierten *CRT1a* Allele auf: für C3E3 waren 50 % *Reads* mit der +1 bp Mutation (A09) zu erwarten und 50 % WT *Reads* für den C09 Locus, während die bei der C3E4 Mutante jedes Allel zu 25 % auftreten sollte. Zu Beginn wurde mittels Sanger-Sequenzierung der Mutationsstatus der beiden verwendeten Mutanten überprüft und es wurde eine neue -2 bp Deletion im C09 Locus der C3E3 Mutante entdeckt. Durch diese *short-read* Analyse mittels Illumina zeigte sich jedoch, dass die neue Deletion im C09 Locus der C3E3 nicht wiedergefunden werden konnte und es sich hiermit höchstwahrscheinlich um eine somatische Mutation im für die Sanger-Sequenzierung verwendeten Blatt handelte. Überraschenderweise wurden in einigen WT-Linien wiederholt kleine Mengen an *Reads* falsch-positiver Signale (0,01%) für einige Mutationen detektiert. Das bedeutet, diese falsch-positiven Signale treten willkürlich auf und sind nicht vorherzusagen. Eine mögliche Erklärung für ihr Auftreten in Amplikon-Sequenzierungen mit Multiplexing durch individuelle Indizes (durch *Barcoding* ermöglichte Zuordnung von *Reads* zu einer bestimmten Pflanze) wäre sogenanntes „Index-Hopping“, ein technisch nicht zu vermeidendes Problem bei solchen Populationsanalysen, oder durch einen PCR-Fehler (wenn z. B. statt TT nur ein T eingebaut wird). Ein möglicher Hintergrund an falsch-positiven Signalen liegt damit rund eine Zehnerpotenz unter der gesetzlich vorgeschriebenen LOD von 0,1 % die sicher bestimmt werden muss. Da aber alle *On-Target* Loci gleich effizient amplifiziert werden, wurde dieser Assay für alle weiteren Analysen verwendet.

Tabelle 16: Ergebnis der Amplikon-Sequenzierung #5 (in rot: Anteil an falsch-Positiven Reads)

Linie	Reads	+1 bp (A09)	-3 bp (A09)	- 1 bp (C09)	WT (C09)
C3E3	44479	47,17 %	0	0	52,83 %
C3E4	330490	22,54 %	24,23 %	26,17 %	27,06 %
Bristol	330351	0	0	0	100 %
Columbus	50605	0	0	0,01 %	99,99 %
Darmor	335676	0	0	0	100 %
Drakkar	332892	0	0	0	100 %
Express	384676	0	0	0	100 %
Forte	315482	0	0	0	100 %
Mozart	44440	0	0	0	100 %
Palma	356439	0	0	0	100 %
Perfect	389337	0	0	0	100 %
Westar	56820	0	0	0,01 %	99,99 %

Um die Spezifität dieses entwickelten Assays #5 zu ermitteln, wurden DNA-Mischproben (5 % der homozygoten C3E3 Mutante in 95 % Raps-WT DNA) in Hintergrund-DNA anderer Spezies (Tomate, Soja, Reis, Mais, Weizen und Zuckerrübe) getestet (Tabelle 17).

Tabelle 17: Spezifitätstest des entwickelten Raps-Assays; n=2

Hintergrund-DNA	MT-Allel		WT-Allel	
	Anteil [%]	STABW	Anteil [%]	STABW
Tomate	4,16	0,00	95,83	0,00
Soja	4,10	0,03	95,90	0,02
Reis	4,31	0,13	95,69	0,12
Mais	4,14	0,26	95,86	0,26
Weizen	4,01	0,00	95,98	0,00
Zuckerrübe	3,96	0,26	96,04	0,26

Es wurden keine wesentlichen Kreuzreaktionen mit anderen Nutzpflanzen detektiert. Die beobachteten Schwankungen liegen innerhalb des zulässigen Bereichs und sind wahrscheinlich auf die Einstellung der DNA-Konzentration mittels qPCR zurückzuführen.

In einem weiteren Experiment wurden DNA-Mischungen verwendet, um die Sensitivität und Konzentrationsabhängigkeit des Assays #5 zu bestimmen. Der Fokus lag wieder auf der homozygoten Mutante C3E3 mit der +1 bp Mutation. Für diese Grundvalidierung wurden insgesamt 12 Verdünnungsstufen (mit mindestens drei und maximal 6 Messungen) eingesetzt. Im relevanten Bereich von 0 % bis 1 % wird deutlich, dass die wichtige 0,1 % LOD im linearen Bereich liegt (Abbildung 16).

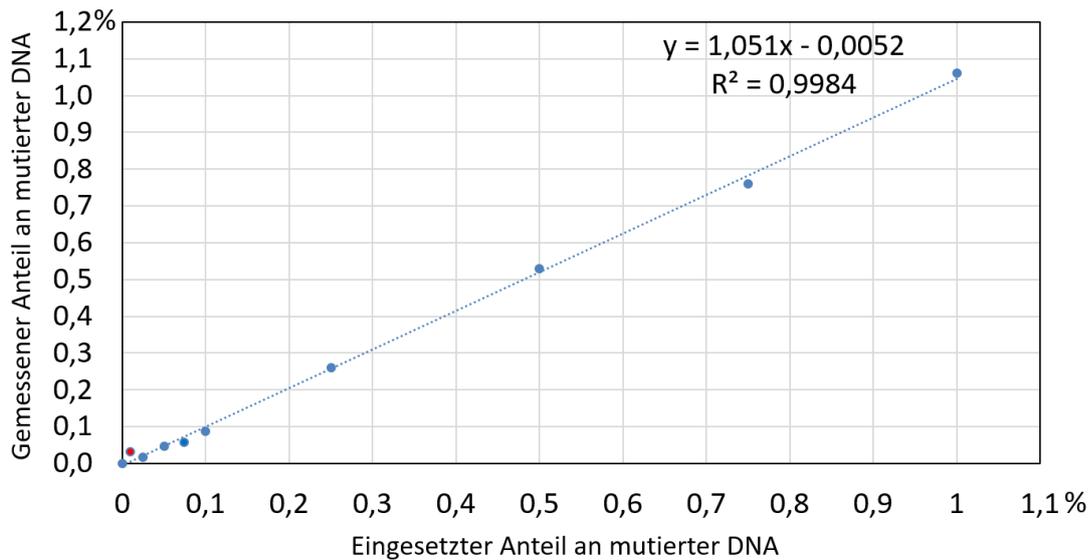


Abbildung 16: Anteil an Mutanten-DNA in DNA-Mischproben in Raps. Die Abbildung stellt den relevanten Bereich <1 % im Detail dar. Der rot hervorgehobene Messwert bei 0,01 % zeigt wie der bei 0,075 eine stochastische Abweichung von der Geraden, allerdings unterhalb der angestrebten 0,1 % LOD.

Die angestrebte LOD von 0,1 % kann somit sicher detektiert werden und damit war dieses Assay generell geeignet für weitere Validierungen mit Saatgut-Mischungen. Mit den Saatgutmischungen sollte erneut die Detektierbarkeit einer LOD von 0,1 % getestet werden (entspricht normalerweise 3 Körner der Mutante in 2997 Wildtyp-Körnern). Da die Körner der Mutanten kleiner sind als die der kommerziellen Sorte Avatar, welche für die Saatgutmischungen verwendet wurden, musste die Einstellung des Mischverhältnisses für 0,1 % Mutantenanteil über das Gewicht erfolgen (Masse/Masse). In einem ersten Versuch wurden 5 Saatgutmischungen erstellt (M1), die DNA isoliert und daraus jeweils 5 Amplikon-Sequenzierungen mit *short-read* Illumina Technik durchgeführt (Abbildung 17). Bei diesen Analysen ist zu bedenken, dass beide *On-Target* Loci (A09 und C09) zu gleichen Anteilen amplifiziert werden, sodass z. B. Werte für die Analyse des A09 Locus theoretisch mit zwei multipliziert werden müssten, oder die Werte der beiden Loci zu addieren sind, um abschätzen zu können wie hoch die Wiederfindungsrate ist. Besonders interessant war die fünfte Mischung mit heterozygotem Saatgut, denn hier liegen die Anteile der einzelnen Allele erwartungsgemäß unter denen der reinen Allele, aber deutlich über dem falsch-positiven Hintergrund (gelb), was zeigt, dass auch kleinste Dosiseffekte nachgewiesen werden können. Um den Nachweis jedoch nicht unnötig zu erschweren, wurde beschlossen, sich bei der Hauptanalyse auf die doppel-homozygote Mutante (#2) zu konzentrieren. Für den zweiten Versuch (M2) wurden 10 Saatgutmischungen zu je 0,1 % mit der doppelt-homozygoten Mutante #2 (A09: +1/+1; C09: -1/-1) erstellt, sowie eine Mischung zu 0,5 % und eine zu 2 %. Nach der Homogenisierung erfolgten aus jeder Probe 4 DNA-Extraktionen und von jeder Extraktion wurden 5 oder 10 Illumina-Reaktionen durchgeführt. Insgesamt konnten 273 Reaktionen ausgewertet werden - mit dem Ergebnis, dass bei einem zu erwartenden Verhältnis von 0,05 % +1 bp (A09) zu 0,05 % -1 bp (C09) nur geringe Abweichungen beobachtet werden konnten: 0,04 % +1 bp (A09) zu 0,06 % -1 bp (C09). Somit konnte auch hier die LOD von 0,1 % erreicht werden.

Verwendeter Genotyp	Suchstring 1			Suchstring 2			Suchstring 3			Suchstring 4		
	WT			+T			-T			-3bp		
	Cluster	Seq	[%] Bs	Cluster	Seq	[%] Bs	Cluster	Seq	[%] Bs	Cluster	Seq	[%] Bs
	CTT?ATCTCTCT?ATT			CTTT?ATCTCTCT?ATT			CT?ATCTCTCT?ATT			A?CATCTCTCT?ATT		
+1/+1	239	120652	99,96	4	34	0,03	2	15	0,01	0	0	<LOD
w/w	280	133138	99,98	4	28	0,02	1	5	<LOD	0	0	<LOD
	243	121794	99,95	4	59	0,05	1	4	<LOD	0	0	<LOD
	262	138045	99,96	5	42	0,03	1	9	0,01	0	0	<LOD
	255	130207	99,96	3	33	0,03	2	15	0,01	0	0	<LOD
+1/+1	256	132113	99,92	3	32	0,02	3	77	0,06	0	0	<LOD
-1/-1	231	120727	99,95	5	37	0,03	2	28	0,02	0	0	<LOD
	256	138447	99,94	3	31	0,02	4	57	0,04	0	0	<LOD
	232	130573	99,94	4	31	0,02	4	50	0,04	0	0	<LOD
	263	134831	99,94	2	17	0,01	3	61	0,05	0	0	<LOD
-3/-3	216	110287	99,93	0	0	<LOD	3	36	0,03	5	36	0,03
-1/-1	281	139826	99,94	0	0	<LOD	3	43	0,03	5	39	0,03
	231	128137	99,93	0	0	<LOD	5	74	0,06	3	22	0,02
	246	128803	99,89	0	0	<LOD	4	89	0,07	5	52	0,04
	235	119712	99,94	0	0	<LOD	2	39	0,03	3	27	0,02
-3/-3	299	150223	99,95	0	0	<LOD	2	8	0,01	7	62	0,04
w/w	230	119403	99,96	0	0	<LOD	2	13	0,01	4	40	0,03
	232	128008	99,96	0	0	<LOD	0	0	<LOD	6	47	0,04
	189	105245	99,96	0	0	<LOD	0	0	<LOD	3	38	0,04
	230	124911	99,96	0	0	<LOD	2	11	0,01	3	45	0,04
+1/-3	167	83372	99,98	2	11	0,01	2	9	0,01	0	0	<LOD
w/-1	321	164822	99,96	2	13	0,01	3	29	0,02	2	16	0,01
	292	145939	99,95	4	38	0,03	2	19	0,01	3	15	0,01
	346	168987	99,97	4	28	0,02	1	10	0,01	2	14	0,01
	297	148225	99,93	2	30	0,02	3	42	0,03	5	28	0,02
M2 \emptyset - Wert	99	75888	99,45	3	34	0,04	2,5	43	0,06	0	0	<LOD

Abbildung 17: Analyse der Saatgutmischung M1 mit homo- und heterozygotem T3-Saatgut. Da Wildtyp-Reads der Mutanten in der großen Menge derer aus der Sorte Avatar untergehen, kann für dieses WT-Allel kein Prozentwert bestimmt werden (blau). In zwei Fällen (gelb) wurden wieder falsch-positive 1 bp Deletionen detektiert, wiederum in einer Größenordnung von 0,01 % (entspricht dem zuvor beobachteten „Hintergrund“ in DNA-Mischungen aus Blättern). Ansonsten entsprechen die Ergebnisse den Erwartungen (grün). Für eine zweite Mischung (M2) wurde nur Saatgut der rot umkreisten Mutante (#2) verwendet und das Durchschnittsergebnis von 273 Reaktionen gezeigt.

3.4. Versuch der Etablierung eines Identifikations-Assays in Raps

Für die Etablierung des Identifikations-Assays wurde ein NextSeq Gerät der Firma Illumina verwendet, da dieses Gerät eine höhere Kapazität hat und parallel zwei Assays mit einer höheren Anzahl an Individuen durchgeführt werden können. Das Assay #5 wurde zum Nachweis der induzierten Mutation in Kombination mit einem zweiten Assay zur Detektion des gekoppelten SNPs verwendet (Abbildung 18). Der Wechsel von der MiSeq zur NextSeq-Plattform und die parallele Analyse eines zweiten Assays erfolgten ohne Probleme, was für die Robustheit des etablierten Assays spricht. Bei Assay #6 zum SNP-Nachweis konnte eine Reaktion nicht ausgewertet werden, aber bei den verbliebenen 99 Reaktion korrelierte die *Read*-Anzahl des SNPs immer mit der Anzahl an *Reads* der -3 bp Mutante, womit die Kopplung im Prinzip bestätigt werden konnte. Somit kann dieses Konzept prinzipiell für die Identifikation Genom-editierter Pflanzen verwendet werden. Im vorliegenden Fall ist die genetische Distanz zwischen Genom-*Edit* und der assoziierten spontanen Mutation jedoch zu groß, um unter landwirtschaftlichen Anbaudimensionen meiotische Entkopplungsereignisse ausschließen zu können (siehe Abschnitt 3.1.5.).

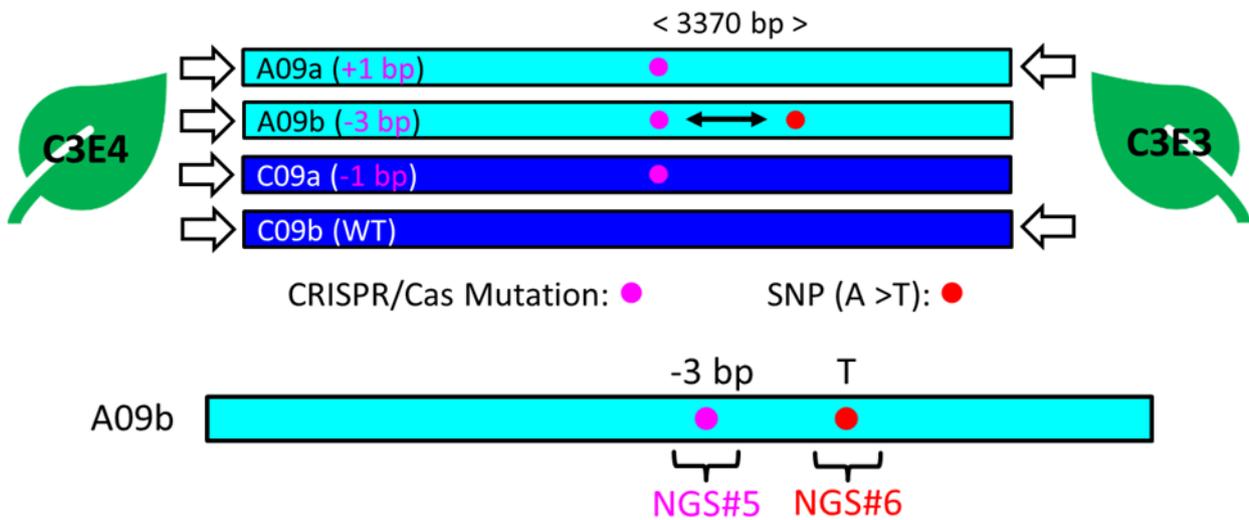


Abbildung 18: Konzept eines Identifikations-Assays durch Kombination zweier *short-read* Amplikon-Tiefensequenzierungen mittels Illumina-Technologie. Oben sind alle 4 möglichen Allele der *On-Target* Loci A09 und C09 gezeigt, von denen das A09b-Allel eine Kopplung zwischen Mutation und Marker (SNP) aufweist. Dieses Allel ist nochmal separat hervorgehoben, um die Sequenzierstrategie durch 2 unabhängige Assays zu verdeutlichen. Die Entfernung zwischen beiden Punkten ist mit 3.370 bp zu groß, um durch Illumina-Technologie überbrückt werden zu können.

3.5. Etablierung einer Analysepipeline für die Tiefensequenzierung von Amplikons

Zur Detektion und Charakterisierung von Einzelnukleotid-Veränderungen wurde eine prototypische bioinformatische Pipeline unter Verwendung frei verfügbarer Sequenzanalyse-Software entwickelt. Dazu wurden die Sequenzdaten der Illumina MiSeq- und NextSeq-Sequenzierung im FASTQ-Format gefiltert, getrimmt und als alignierte Sequenzen im FASTA-Format ausgegeben. Diese prototypische Pipeline wurde anhand von verschiedenen Amplikondatensätzen der AG Kumlehn (PRB, IPK) getestet. Die Pipeline wurde hinsichtlich allgemeiner Verwendbarkeit und Nutzerfreundlichkeit weiter verbessert, indem sie als *Workflow* mit graphischer Oberfläche im IPK-Galaxy-System integriert wurde, sodass eine Nutzung nicht ausschließlich auf Ebene der UNIX-Kommandozeile möglich ist.

Eine vergleichende Auswertung der ersten Experimente der Amplikon-Tiefensequenzierung zeigte, dass die herausgegebenen Ergebnisse der Analysepipelines von Planton (proprietary) und vom IPK sehr ähnlich sind. Zum Beispiel gab es bei der Auswertung des zweiten Experiments zwischen den beiden *Pipelines* nur eine 0,01 %ige Abweichung der Anteile von Wildtypsequenzen in der Mutanten-Probe und keine Abweichungen bei den Anteilen an Mutanten-DNA in der Wildtyp-Probe.

4. Diskussion und zusammenfassende Bewertung

Gemäß der Vorhabensbeschreibung wurden alle relevanten Ziele erreicht. Das WGS und die Hi-C-Sequenzierung der Gersten-Donor-Linie 'Golden Promise' DH1-6 und der *pdil5-1*-Mutante mit einer Zielsequenz-spezifischen Veränderung in Gerste bestätigte, dass die Mutante die erwartete 1 bp-Deletion im *PDIL5-1*-Gen aufweist. Alle im Gerstengenom identifizierten *Off-targets*, also dem adressierten Zielmotiv ähnliche Sequenzen, erwiesen sich hingegen als frei von genetischen Veränderungen. Zudem wurde gezeigt, dass keinerlei Rückstände eines Transgens in der *pdil5-1*-Mutante vorhanden sind und dass es keine mit der Genom-Editierung assoziierten strukturellen Veränderungen gab.

Auch die Genomsequenzierung der beiden Rapsmutanten C3E3 und C3E4 zeigte, dass nur die *On-targets* gezielt mutiert wurden und alle identifizierten *Off-targets* frei von genetischen Veränderungen waren. Jedoch fanden sich in beiden Rapsmutanten noch Rückstände der agrobakteriellen T-DNA wieder. Diese Rückstände sind jedoch aller Wahrscheinlichkeit nach durch Selbstung auskrenzbar, so dass bei Bedarf in den Folgegenerationen transgenfreie Mutanten selektiert werden können. In Raps konnte zudem ein genetisch gekoppelter SNP in 3.370 bp Distanz zu der induzierten Mutation im A09b Locus identifiziert werden. Anhand dieses SNPs wurde versucht, ein Identifikations-Assay für Raps basierend auf der Amplikon-Tiefensequenzierung zu entwickeln.

Mit *long-read* PacBio WGS können viele hochwertige und sehr lange Sequenzen erzeugt werden, die für die Assemblierung von Genomen besser geeignet sind als kurze Illumina-*Reads*. Zudem lassen sich Rückstände eines Transgens zuverlässig ermitteln, während bei den oft verwendeten PCR-Tests kleinere Fragmentierungen übersehen werden können. Außerdem ist eine umfassende Sequenzierung der benachbarten Regionen der induzierten Mutation wichtig, damit womöglich gekoppelte SNPs ermittelt und darauf basierend Identifikations-Assays etabliert werden können. Die WGS und Hi-C-Sequenzierungen haben gezeigt, dass mit den neuen Züchtungstechniken ganz gezielt Mutationen ausschließlich an vorgegebener Stelle ausgelöst werden können. Zudem wurde gezeigt, dass durch die Genom-Editierung mittels Cas-Endonukleasen üblicherweise keine mit der Methode assoziierbaren Rückstände oder sogenannte „Footprints“ im Genom hinterlassen werden, die als Identifikationsmerkmal für die Nutzung dieser neuen Züchtungstechniken bei der Herstellung von Genom-editierten Pflanzen herangezogen werden könnten. Hierfür kommen ausschließlich in unmittelbarer Nähe der gezielt induzierten Mutation spontan und simultan entstandene Sequenzveränderungen in Betracht. Solche Fälle treten jedoch mit zu geringer Wahrscheinlichkeit auf, um dieses Konzept über bestenfalls wenige Ausnahmefälle hinaus für die Identifizierung Genom-editierter Pflanzen in Anwendung zu bringen zu können.

Es konnte kein spezifisches qPCR-Assay für Gerste entwickelt werden, da mit allen 5 dafür getesteten Systemen Kreuzreaktionen der Sonden auftraten und keine eindeutige Differenzierung zwischen dem Wildtyp 'Golden Promise' DH1-6 und der *pdil5-1*-Mutante möglich war. Es konnte hingegen ein ddPCR-basiertes Verfahren zur Detektion der *pdil5-1*-Mutante etabliert werden. Ebenso wurde ein ddPCR-Assay zur Detektion der Cas9-induzierten 1 bp-Insertion im A09 Locus von homozygoten Nachkommen der Rapsmutante C3E3 entwickelt. Mit beiden Assays können auf verlässliche Weise sowohl 0,9 % als auch 0,1 % Mengenteile der jeweiligen Mutante in Mischproben nachgewiesen werden. In Übereinstimmung mit den „*Minimum Performance Requirements for Analytical Methods of GMO Testing*“ des *European*

Network of GMO Laboratories (ENGL, 2015) wurden die ddPCR-Assays von dem akkreditierten Referenzlabor Impetus Bioscience erfolgreich auf Spezifität, Selektivität und Anwendbarkeit getestet. Im Allgemeinen werden qPCR-basierte Methoden an ihre Grenzen stoßen, wenn es darum geht, bestimmte SNVs zuverlässig nachzuweisen, die durch gezielte Mutagenese entstanden sind. Im Gegensatz dazu kann die ddPCR für diese Zielstellung als zuverlässigere Methode angesehen werden. Im Rahmen dieses Projekts ist es uns gelungen, zuverlässige ddPCR-Assays für die getesteten Gersten- und Rapsmutanten zu etablieren. Die Chancen, eine ddPCR-basierte Nachweismethode für diverse andere Mutationen zu etablieren, werden allgemein als recht gut eingeschätzt, aber auch hier gilt, dass jedes Assay und jede Mutation individuell validiert werden muss.

Darüber hinaus wurden jeweils für Gerste und Raps Detektionsverfahren etabliert, die auf Amplikon-Tiefensequenzierung beruhen. Auch mit dieser Methode wurde gezeigt, dass 0,9 % und 0,1 % Mengenanteile der Gerstenmutante und der Rapsmutanten in Mischproben zuverlässig nachgewiesen werden können. Die Amplikon-Tiefensequenzierungen wurden von dem akkreditierten Referenzlabor Planton durchgeführt und die entwickelten Methoden erfolgreich auf Spezifität, Selektivität und Anwendbarkeit getestet. Der Vorteil eines einmal etablierten Amplikon-Sequenzierungs-Assays liegt vor allem in seiner Aussagekraft und Flexibilität unterschiedliche Mutationen in der gleichen genomischen Region detektieren zu können, was nur durch die Erzeugung von Sequenzdaten möglich ist. Aufgrund des hohen Durchsatzes bei einem Illumina-Sequenzierungslauf können alle relevanten Allele auf einmal in ausreichender Abdeckung untersucht werden. Selbst wenn es bei allopolyploiden Pflanzen wie Raps (oder auch Baumwolle, Tabak, Erdnuss, Quinoa, Cassava, Triticale, Weizen etc.) notwendig werden sollte, die Mutationen einem der Subgenome zuzuordnen, sind die Sequenzdaten informativer als andere PCR-basierte Methoden, da auch natürliche SNPs und InDels im Amplikon nachgewiesen werden und anhand dieser eine Differenzierung zwischen den verschiedenen Loci möglich ist.

Am IPK wurde mit Hilfe der am Projekt beteiligten Arbeitsgruppe Domestikationsgenomik eine bioinformatische Analysepipeline für tiefensequenzierte Amplikons entwickelt. Diese Pipeline wurde auf dem Galaxy Server in eine benutzerfreundliche Oberfläche eingebunden. Die Analysepipeline wurde anhand mehrerer Datensätze getestet und eine vergleichende Auswertung anhand der von Planton gelieferten Daten zeigte, dass sie zuverlässige und mit anderen Protokollen vergleichbare Resultate liefert, im Gegensatz zu ihren Alternativen jedoch vorgesehen ist, im Kontext einer öffentlich zugänglichen Plattform (Galaxy) publiziert und freigegeben zu werden.

5. Literaturverzeichnis

- Arulandhu, A.J., van Dijk, J.P., Staats, M., Hagelaar, R., Voorhuijzen, M.M., Molenaar, B., Van hoof, R., Li, R., Yang, L., Shi, J., Scholtens, I.M., Kok, E.J. (2018). NGS-based amplicon sequencing approach; towards a new era in GMO screening and detection. *Food Control*.
- Beier, S., T. Thiel, T. Munch, U. Scholz, and M. Mascher. (2017). MISA-web: a web server for microsatellite prediction. *Bioinformatics*, 33: 2583-85.
- Bushnell, B., Rood, J. and Singer, E. (2017). BBMerge - Accurate paired shotgun read merging via overlap. *PLoS One* 12, e0185056.
- Chhalliyil, P.; Ilves, H.; Kazakov, S.A.; Howard, S.J.; Johnston, B.H.; Fagan, J. (2020). A Real-Time Quantitative PCR Method Specific for Detection and Quantification of the First Commercialized Genome-Edited Plant. *Foods*. 9:1245
- Chen, S., Y. Zhou, Y. Chen, and J. Gu. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, 34: i884-i90.
- Cheng H, Concepcion GT, Feng X, Zhang H, Li H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods* 18: 170–175.
- Danecek, P., J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A. Whitwham, T. Keane, S. A. McCarthy, R. M. Davies, and H. Li. (2021). Twelve years of SAMtools and BCFtools. *Gigascience*, 10.
- Dowle, Matt, and Arun Srinivasan. (2022). Data.table: Extension of 'data.frame'.
- Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., Bibillo, A., Bjornson, K., Chaudhuri, B., Christians, F., Cicero, R., Clark, S., Dalal, R., Dewinter, A., Dixon, J., Foquet, M., ... Turner, S. (2009). Real-time DNA sequencing from single polymerase molecules. *Science (New York, N.Y.)*, 323(5910), 133–138.
- ENGL, European Network of GMO Laboratories (2015). Definition of Minimum Performance Requirements for Analytical Methods of GMO Testing. JRC 95544
- ENGL, European Network of GMO Laboratories (2019). Detection of food and feed plant products obtained by new mutagenesis techniques. JRC116289
- Falabella, M., Sun, L., Barr, J., Pena, A.Z., Kershaw, E.E., Gingras, S., Goncharova, E.A., and Kaufman, B.A. (2017). Single-Step qPCR and dPCR Detection of Diverse CRISPR-Cas9 Gene Editing Events in Vivo. *G3: Genes|Genomes|Genetics*, 7, 3533 - 3542.
- Fraiture, M. A., D'aes, J., Guiderdoni, E., Meunier, A. C., Delcourt, T., Hoffman, S., ... and Roosens, N. H. (2023). Targeted high-throughput sequencing enables the detection of single nucleotide variations in CRISPR/Cas9 gene-edited organisms. *Foods*, 12(3), 455
- Fraiture, M. A., Guiderdoni, E., Meunier, A., Papazova, N., and Roosens, N.H. (2022). ddPCR strategy to detect a gene-edited plant carrying a single variation point: Technical feasibility and interpretation issues. *Food Control*, 137(16):108904
- Fraiture, M.A., Herman, P., Papazova, N., De Loose, M., Deforce, D., Ruttink, T., & Roosens, N.H. (2017). An integrated strategy combining DNA walking and NGS to detect GMOs. *Food chemistry*, 232, 351-358.
- Goel, M., H. Q. Sun, W. B. Jiao, and K. Schneeberger. 2019. 'SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies', *Genome Biology*, 20.
- Grohmann, L., Keilwagen, J., Duensing, N., Dagand, E., Hartung, F., Wilhelm, R., ... and Sprink, T. (2019). Detection and identification of genome editing in plants: challenges and opportunities. *Frontiers in Plant Science*, 10, 236.
- Hernandez M, Esteve T, Pla M. (2005). Real-Time Polymerase Chain Reaction Based Assays for Quantitative Detection of Barley, Rice, Sunflower, and Wheat. *J Agric Food Chem* 53:7003–7009
- Hoffie, R.E., Perovic, D., Habekuß, A., Ordon, F. and Kumlehn, J. (2022). Novel resistance to the Bymovirus BaMMV established by targeted mutagenesis of the PDIL5-1 susceptibility gene in barley. *Plant Biotechnol J*.
- Holst-Jensen, A., Bertheau, Y., De Loose, M., Grohmann, L., Hamels, S., Hougs, L., ... and Wulff, D. (2012). Detecting un-authorized genetically modified organisms (GMOs) and derived materials. *Biotechnology advances*, 30(6), 1318-1335.
- Jayakodi, M., M. Schreiber, N. Stein, and M. Mascher. 2021. 'Building pan-genome infrastructures for crop plants and their use in association genetics', *DNA Res*, 28.
- Katoh, K. and Toh, H. (2008). Recent developments in the MAFFT multiple sequence alignment program. *Briefings in Bioinformatics* 9, 286-298.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology*, 10(3), R25.
- Li, G., Wang, L., Yang, J., He, H., Jin, H., Li, X., Ren, T., Ren, Z., Li, F., Han, X., Zhao, X., Dong, L., Li, Y., Song, Z., Yan, Z., Zheng, N., Shi, C., Wang, Z., Yang, S., Xiong, Z., ...and Wang, D. (2021). A high-quality genome assembly highlights rye genomic characteristics and agronomically important genes. *Nature genetics*, 53(4), 574–584.
- Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 34: 3094-100.
- Li, H. and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, 25(14), 1754–1760.
- Liu, H., and J. Zhang. (2022). Is the Mutation Rate Lower in Genomic Regions of Stronger Selective Constraints?. *Mol Biol Evol*, 39.

- Marone, M. P., H. C. Singh, C. J. Pozniak, and M. Mascher (2022). 'A technical guide to TRITEX, a computational pipeline for chromosome-scale sequence assembly of plant genomes', *Plant Methods*, 18.
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *MBnet. Journal* 17, pp. 10-12.
- Mascher, M., H. Gundlach, A. Himmelbach, ... and N. Stein. (2017). A chromosome conformation capture ordered sequence of the barley genome. *Nature*, 544: 427-33.
- Mascher, M., Wicker, T., Jenkins, J., Plott, C., Lux, T., Koh, C. S., Ens, J., Gundlach, H., Boston, L. B., Tulpová, Z., Holden, S., Hernández-Pinzón, I., Scholz, U., Mayer, K. F. X., Spannagl, M., Pozniak, C. J., Sharpe, A. G., Šimková, H., Moscou, M. J., Grimwood, J., ...and Stein, N. (2021). Long-read sequence assembly: a technical evaluation in barley. *The Plant cell*, 33(6), 1888–1906.
- Milner, S. G., M. Jost, S. Taketa, ... and N. Stein. (2019). Genebank genomics highlights the diversity of a global barley collection. *Nat Genet*, 51: 319-26.
- Mock, U., Hauber, I. and Fehse, B. (2016) Digital PCR to assess gene-editing frequencies (GEF-dPCR) mediated by designer nucleases. *Nat Protoc* 11, 598–615
- Monroe, J. G., T. Srikant, P. Carbonell-Bejerano, ... and D. Weigel. 2022. 'Mutation bias reflects natural selection in *Arabidopsis thaliana*', *Nature*, 602: 101-05.
- Navratilova, P., H. Toegelova, Z. Tulpova, Y. T. Kuo, N. Stein, J. Dolezel, A. Houben, H. Simkova, and M. Mascher. 2022. 'Prospects of telomere-to-telomere assembly in barley: Analysis of sequence gaps in the MorexV3 reference genome', *Plant Biotechnol J*, 20: 1373-86.
- Pallarz, S., Fiedler, S., Wahler, D., Lämke, J., and Grohmann, L. (2023). Reproducibility of next-generation-sequencing-based analysis of a CRISPR/Cas9 genome edited oil seed rape. *Food Chemistry: Molecular Sciences*, 100182.
- Pallotta, M.A., Graham, R.D., Langridge, P., Sparrow, D.H.B., and Barker, S.J. (2000). RFLP mapping of manganese efficiency in barley. *Theoretical and Applied Genetics* 101, 1100–1108.
- Pecoraro S., Berben G., Burns M., Corbisier P., De Giacomo M.,... Spilberg B. (2019). Overview and recommendations for the application of digital PCR. EUR 29673 EN, Publications Office of the European Union, JRC 115736.
- Pucker, B., Kleinbölting, N., and Weisshaar, B. (2021). Large scale genomic rearrangements in selected *Arabidopsis thaliana* T-DNA lines are caused by T-DNA insertion mutagenesis. *BMC genomics*, 22(1), 599.
- Rousseau-Gueutin, M., Belser, C., Da Silva, C., Richard, G., Istace, B., Cruaud, C., Falentin, C., Boideau, F., Boute, J., Delourme, R., Deniot, G., Engelen, S., de Carvalho, J. F., Lemainque, A., Maillet, L., Morice, J., Wincker, P., Denoeud, F., Chèvre, A. M., and Aury, J. M. (2020). Long-read assembly of the *Brassica napus* reference genome Darmor-bzh. *GigaScience*, 9(12), gaa137.
- Sato, K., Abe, F., Mascher, M., Haberer, G., Gundlach, H., Spannagl, M., Shirasawa, K., & Isobe, S. (2021). Chromosome-scale genome assembly of the transformation-amenable common wheat cultivar 'Fielder'. *DNA research : an international journal for rapid publication of reports on genes and genomes*, 28(3), dsab008
- Stemmer, M., Thumberger, T., Del Sol Keyer, M., Wittbrodt, J., and Mateo, J. L. (2015). CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool. *PloS one*, 10(4), e0124633.
- Team, R Core. (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing.
- Thorvaldsdottir, H., J. T. Robinson, and J. P. Mesirov. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*, 14: 178-92.
- Wang, L., Ji, Y., Hu, Y., Hu, H., Jia, X., Jiang, M., Zhang, X., Zhao, L., Zhang, Y., Jia, Y., Qin, C., Yu, L., Huang, J., Yang, S., Hurst, L. D., and Tian, D. (2019). The architecture of intra-organism mutation rate variation in plants. *PLoS biology*, 17(4), e3000191.

30.11.2023




 Datum und Unterschrift der Projektverantwortlichen