

Abschlussbericht

Zuwendungsempfänger:

IPK Gatersleben

Projektleiter: Dr. Jochen Kumlehn

Förderkennzeichen: 2820HS001

Bundesministerium
für Ernährung
und Landwirtschaft



Thema des Forschungsvorhabens:

Machbarkeitsstudie zu Nachweis- und Identifizierungsverfahren für genomeditierte Pflanzen und pflanzliche Produkte (DETECT)

Laufzeit des Vorhabens:

01.01.2021 – 30.11.2023

Berichtszeitraum:

01.01.2021 – 30.11.2023

Projektmitarbeiter

Dr. Iris Hoffie

Ehemaliger Mitarbeiter: Dr. Martin Becker

Projektpartner

Dr. Martin Mascher

Dr. Marina Pupke Marone



Kooperationspartner

Prof. Dr. Daguang Cai

Dr. habil. Dirk Schenke



Unterauftragnehmer

Prof. Dr. Michael Kleine

Dr. Martin Weigel



Dr. Olav Grundmann



Zusammenarbeit mit anderen Stellen

Dr. Lutz Grohmann



Bundesamt für
Verbraucherschutz und
Lebensmittelsicherheit

Inhaltsverzeichnis

1. Ziele und Aufgabenstellung des Vorhabens	4
1.1. Planung und Ablauf des Vorhabens	4
1.2. Wissenschaftlicher und technischer Stand, an den angeknüpft wurde.....	5
2. Material und Methoden.....	6
3. Ergebnisse	15
3.1. Ausführliche Darstellungen der wichtigsten Ergebnisse.....	15
3.1.1. <i>Whole-Genome Sequencing</i> von Gerste	15
3.1.2. Hi-C Sequenzierung und Assemblierung der Genomdaten zu chromosomalen Pseudomolekülen in Gerste	16
3.1.3. Ermittlung von Transgenrückständen und <i>Off-Targets</i> in Gerste	20
3.1.4. <i>Whole-Genome Sequencing</i> , Hi-C Sequenzierung und Assemblierung der Genomdaten zu chromosomalen Pseudomolekülen in Raps.....	20
3.1.5. Etablierung der Detektion von Einzelnukleotid-Veränderungen mittels qPCR in Gerste	23
3.1.6. Etablierung der Detektion von Einzelnukleotid-Veränderungen mittels ddPCR in Gerste	29
3.1.7. Etablierung der Detektion von Einzelnukleotid-Veränderungen mittels ddPCR in Raps	31
3.1.8. Charakterisierung der Zielregionen mittels Tiefensequenzierung von Amplikons in Gerste	32
3.1.9. Etablierung einer Analysepipeline für die Tiefensequenzierung von Amplikons	34
3.1.10. Verifizierung spontan entstandener Hintergrundmutationen in Gerste	35
3.2. Voraussichtlicher Nutzen und Verwertbarkeit der Ergebnisse.....	37
4. Zusammenfassung	37
4.1. Zusammenfassende Bewertung	38
5. Gegenüberstellung der ursprünglich geplanten zu den tatsächlich erreichten Zielen	39
6. Literaturverzeichnis	40

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 (AG 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 Nucleasen
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 künstlich induzierter Mutagenese resultieren, wurde in diesem Urteil generell klargestellt, dass sie GMO sind. Während jedoch solche Pflanzen mit ungerichtet (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 Zielstrukturen pflanzlicher Genome angepasst werden können. Die Zielsequenzspezifität solcher Endonukleasen kann Protein-vermittelt (Meganukleasen, Zinkfinger-nukleasen, *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 Endonuklease 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 Triplet 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 sind bereits auf nationalen Märkten außerhalb Europas (wie z. B. in zahlreichen süd- und nordamerikanischen und 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 GMO 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 GMO in Verkehr gebracht werden und ob die Kennzeichnungspflicht für zugelassene GMO 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 zielte das DETECT-Projekt auf die Etablierung von Nachweis- und Identifizierungsverfahren für Genom-editiertes Pflanzenmaterial und die Validierung dieser Verfahren bezüglich ihrer Eignung für GMO-Genehmigungsanträge, amtliche Kontrollen und Eigenkontrollen der Wirtschaft. Im Wesentlichen wurde dafür folgende wissenschaftliche Bearbeitung erwartet:

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

Darüber hinaus sollte geprüft werden, ob ein analytisches Verfahren einen Genom-editierten Organismus eindeutig identifizieren und von Organismen unterscheiden kann, bei denen eine identisch modifizierte Nukeotidsequenz auf andere Weise entstanden ist

Durch ein initiales *Whole Genome Sequencing* (WGS) einer Gerstenlinie mit einer marktrelevanten Zielsequenz-spezifischen Veränderung sowie dessen nicht-editierter Ausgangslinie sollte das 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 qPCR, die ddPCR und die Tiefensequenzierung von PCR-Amplikons (*deep sequencing of amplicons*) vorgesehen. Für die Auswertung der Amplikonsequenzen sollte des Weiteren ein bioinformatisches Verfahren (eine Analysepipeline) entwickelt werden.

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 Proben (Schüttgut) 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 Anwendbarkeit 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 DETECT 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 hergestellte Pflanzen sind mit etablierten PCR-basierten Methoden nachweisbar, identifizierbar und quantifizierbar, da stabil integrierte Transgene in der Natur in dieser Kombination nicht vorkommen und sie diese GVO-Pflanzen somit einzigartig machen (Grohmann et al., 2019; Pallarz et al., 2023). Dagegen enthalten Genom-editierte Pflanzen in der Regel keine rekombinante DNA, sondern weisen nur kleine Veränderungen auf Nukleotidebene auf. In der GVO-Analytik wurden in der Vergangenheit hauptsächlich PCR-basierte Methoden (zumeist qPCRs) 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 ein eindeutiger Nachweis und eine Quantifizierung von Genom-editierten Pflanzen aber eine hohe Präzision und Auflösung erfordern, sind für die GVO-Analytik ddPCR- und Amplikon-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 sehr sensitiv ist und eine absolute Quantifizierung der Zielsequenz möglich ist ohne abhängig zu sein von optimalen Referenzmaterialien und Standardkurven (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 sequenzierungs-basierte Verfahren tiefere und eindeutige Informationen liefern können (Pallarz et al., 2023; Fraiture et al., 2023; Grohmann et al., 2019). Zum Zeitpunkt des Beginns der Projektlaufzeit waren sequenzierungs-basierte 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). Jedoch waren die Qualitätskriterien für die Bewertung von Sequenzierungsdaten und die Validierung sowohl von ddPCR- als auch sequenzierungs-basierten Methoden zu diesem Zeitpunkt noch nicht evaluiert (Pecoraro et al., 2019).

2. Material und Methoden

2.1. Pflanzenmaterial

Im Teilprojekt DETECT wurde vorrangig Blattmaterial und Körner von einer Genom-editierten Gerstenlinie und deren Elternlinie verwendet. Die Gerstenmutante 16-12 weist eine Cas9-induzierte homozygote 1 bp-Deletion im 5'-UTR des *Protein Disulfide Isomerase Like 5-1 (PDIL5-1)*-Gens auf (Abbildung 1; Hoffie et al., 2022). 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. Blattmaterial und Körner der dritten Generation (M₃)

der Genom-editierten Transgen-freien Gerstenlinie wurden verwendet. Als Wildtyp-Kontrolle wurde bei allen Experimenten die direkte Donor-/Elternlinie 'Golden Promise' DH1-6 verwendet.

```
Wildtyp Golden Promise cccagagaaatacatcatcgaccaccccgactggcgatatccgccttct
pdil5-1 Mutante       cccagagaaatacatcatcgaccaccccgactggcgatatccgccttct
```

Abbildung 1: Vergleich der DNA-Sequenz des Wildtyps 'Golden Promise' DH1-6 und der *pdil5-1* Mutante 16-12. Das Zielmotiv inkl. PAM ist unterstrichen.

Zudem wurde für die Genomsequenzierung und die Etablierung eines ddPCR-Assays in Raps geeignetes Material von den Projektpartnern aus Kiel bereitgestellt (siehe Berichtsteil RapsNMT).

Sofern nicht anders angegeben, wurde genomische DNA aus Blättern von 1-2 Wochen alten Keimlingen extrahiert. Es wurde das NucleoSpin Plant II Kit (Macherey-Nagel) verwendet. Die Konzentration von genomischer DNA wurde mit einem Nanodrop 2000 Spektralphotometer (Thermo Fisher Scientific, Waltham, MA, USA) oder einem Qubit® Fluorometer (Invitrogen/Thermo Fisher Scientific, Waltham, MA, USA) bestimmt.

2.2. Oligonukleotide

Alle verwendeten Oligonukleotide sind im Anhang in der Tabelle A1 aufgeführt. Die für die Detektions-Assays verwendeten Oligonukleotide sind zusätzlich gesondert in den jeweiligen Abschnitten aufgeführt. Alle Oligonukleotide für Standard-PCRs wurden von der Firma Biologio (Nijmegen, Niederlande) hergestellt. Alle Primer und Sonden für die PCR-basierten Detektionsassays wurden von der Firma Biomers (Ulm, Deutschland) hergestellt.

2.3. 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).

2.3.1. Vorbereitung der WGS-Proben

Um das Gerstengenom mit 5 Gigabasen ausreichend abzudecken, wurden für PacBio inklusive Sicherheitspuffer ca. 100 µg und für Raps ca. 40 µg genomische DNA mit einer durchschnittlichen Fragmentgröße von über 50 Kilobasen (hochmolekulare DNA) benötigt. Es wurde das NucleoBond HMW DNA Kit (Macherey-Nagel, Düren) verwendet, das hohe Ausbeuten an hochmolekularer DNA bei einer relativ geringen Menge an eingesetztem Blattmaterial verspricht. Für Gerste wurden Keimlinge angezogen und über einen Zeitraum von ca. drei Wochen sukzessive das jeweils jüngste Blatt geerntet, um auf eine Gesamt-Blattfrischmasse von 3 g zu kommen. Das Blattmaterial der beiden Rapsmutanten wurde auf Trockeneis vom Projektpartner CAU/ Kiel) bereitgestellt und die Extraktion der hochmolekularen DNA ebenfalls mit dem DNA-Extraktions-Kit von Macherey-Nagel durchgeführt. Sowohl für Gerste als auch Raps erwies sich

das Blattmaterial einer Einzelpflanze als hinreichend, um die für die PacBio-Sequenzierung benötigten Mengen an hochmolekularer DNA bereitzustellen. Die Verteilung der Fragmentgrößen innerhalb einer Probe wurde von der IPK-Arbeitsgruppe GGR (IPK) mittels einer *FEMTO Pulse*-Messung ermittelt (Abbildung 2).

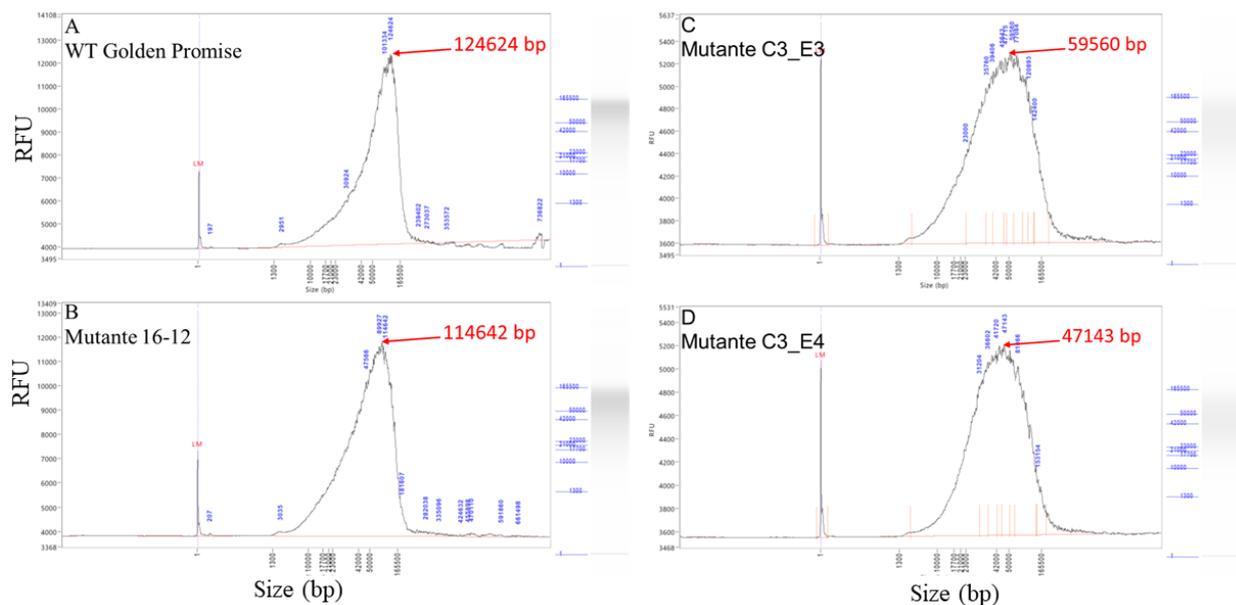


Abbildung 2: Verteilung von Fragmentgrößen in DNA-Proben. (A) Gerste 'Golden Promise' DH1-6. (B) Gerstenmutante 16-12. (C) Rapsmutante C3E3. (D) Rapsmutante C3E4.

2.3.2. Assemblierung der Genome

Die Sequenzierungs-Rohdaten der Gersten- und Rapsmutanten sowie der Gersten-Wildtypie wurden an die am Projekt beteiligten Kollegen der IPK-Arbeitsgruppe Mascher (DG) übergeben. Zudem wurden die Genomdaten der Raps-Donor-Linie 'Mozart' vom BVL zur Verfügung gestellt. 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>).

2.4. Bioinformatische Auswertung der WGS von Gerste

Qualität der Sequenzen

Um die durchschnittliche Qualität der Roh-Daten zu bestimmen, wurde die Software fastp v.0.20.0 (Chen et al. 2018) verwendet und eine Analyse für jede PacBio *HiFi Read Library* durchgeführt. Sowohl für die Mutante als auch für die Donor-Linie erwiesen sich die Lesequalitäten erwartungsgemäß hoch, mit Ausnahme des 3'-Endes eines einzigen *Reads*. Der Prozentsatz der Q30-Basen (Basen mit 99,9 % Erkennungsgenauigkeit) lag bei allen Proben im Durchschnitt bei 91,4 %, was bedeutet, dass nur wenige Basen eine niedrigere Erkennungsgenauigkeit aufwiesen. Der Prozentsatz der Q20-Basen (Basen mit 99 % Erkennungsgenauigkeit) betrug bei allen Proben 95,8 %. Dieses Ergebnis zeigt, dass die Genauigkeit den Erwartungen entsprach.

Ermittlung von Sequenzvarianten

Um die Anzahl der Sequenzvarianten in Form von Einzelnukleotid-Polymorphismen (SNPs) und Insertionen/Deletionen (Indels) zu ermitteln, wurde eine *Variant Calling*-Analyse durchgeführt. Bei solchen Analysen werden die *Reads* der entsprechenden Probe mit einem Referenzgenom verglichen. Wenn eine von der Referenz abweichende Variante vorhanden und homozygot ist, wird dies als *Coverage* von *Reads* angezeigt, die ein alternatives Allel belegen. Wenn die Stelle heterozygot ist, gibt es *Reads*, die sowohl das Referenz- als auch das alternative Allel repräsentieren. Ist die Stelle hingegen identisch mit der Referenz, zeigen die *Reads* ausschließlich das Referenzallel. Bei der Analyse von Varianten aus einer einzigen Probe ist es schwierig, Artefakte zu erkennen, da Abweichungen durch Sequenzierungs- oder *Alignment*-Fehler verursacht sein könnten. Bei der Analyse wurde entsprechend versucht, solche Fehler durch die Anwendung strenger Filter zu minimieren. In diesem Projekt wurde die Donor-Linie 'Golden Promise' DH1-6 als Referenz verwendet. Es wurden die *HiFi-Reads* der Donor-Linie (als Negativkontrolle) und der Mutante genutzt. Die *Reads* jeder Probe wurden mit Hilfe der Software pbmm2 v.1.9.0 (<https://github.com/PacificBiosciences/pbmm2>) und einem *Wrapper* für *Long-Read Assemblies* von minimap2 v2.2 (Li, 2018) unter Verwendung des Parameters „preset – CCS“ mit der gewählten Referenz kartiert.

Die ausgegebenen BAM-Dateien wurden zusammengeführt und in CRAM mit SAMTools v.1.10 (Danecek et al. 2021) konvertiert, um ein *Variant Calling* mittels BCFTools v.1.12 (Danecek et al. 2021) durchführen zu können. Es wurden SNPs und Indels mit einer Qualität von ≥ 40 und einer Abdeckung pro Position von ≥ 20 separat ausgewählt. Auf diese Weise wurden zwei VCF-Dateien erzeugt (eine für SNPs und eine für Indels), die für nachfolgende Filterungsschritte in R v.4.2 (Team 2022) unter Verwendung des Pakets data.table v.1.14.6 (Dowle und Srinivasan 2022) genutzt wurden. Die VCF-Dateien wurden mit einem benutzerdefinierten Skript in ein langes Format konvertiert. Die Varianten wurden gemäß den Kriterien in Tabelle 1 genotypisiert. Diejenigen, die nicht einer dieser Kategorien zugeordnet werden konnten, also mit hoher Wahrscheinlichkeit Sequenzierfehler repräsentieren oder auf kleine Sektoren der Pflanzenprobe begrenzt und daher nicht vererbbar sind, wurden ausgeschlossen. Sowohl für SNPs als auch für Indels wurden folgende Filter verwendet: (1) Entfernen von technischen Artefakten, d. h. von Varianten, die im Referenzgenom als alternative Allele oder Heterozygot bezeichnet werden, und (2) Entfernen von Varianten mit zu geringer oder zu hoher *Coverage*. Die durchschnittliche Genom-*Coverage* der zwei Genome (Donor-Linie und Mutante) betrug 28. Ausgehend davon wurden Grenzwerte für die *Coverage* festgelegt: $14 < Coverage < 56$. Für Indels wurde ein weiterer Filterschritt hinzugefügt, um SSR-Sequenzen (einfache Sequenzwiederholungen) zu entfernen. Für die Identifizierung der Positionen von SSRs wurde MISA v.2.1 (Beier et al. 2017) verwendet und in R wurden sie dann entfernt.

Tabelle 1: Kategorisierung der Sequenz-Varianten

Genotyp	Beschreibung
Homozygotes Referenz Allel	mindestens 90 % der Reads entsprechen dem Referenzallel
Heterozygot	je 30-70 % der Reads werden beiden Allelen zugeordnet
Homozygotes alternatives Allel	mindestens 90 % der Reads entsprechen dem alternativen Allel

2.5. 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 2.5.1. und 2.5.2. beschrieben.

2.5.1. 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. TRITEX wurde bereits erfolgreich für die Assemblierung vieler Pflanzengenome eingesetzt. Zunächst wurde ein *in-silico*-Verdau der *Contig*-Assemblierungen mit DpnII (GATCGATC) durchgeführt, demselben Enzym, das auch bei der Vorbereitung der *Libraries* verwendet wurde (Skript "digest_emboss.zsh" unter <https://bitbucket.org/tritexassembly>). Als Leitfaden diente das 'Golden Promise'-Genom im Rahmen des Gersten-Pangenom-Projekts (Martin Mascher, Nils Stein, unveröffentlichte Ergebnisse). Dazu wurde mit dem Skript "mask_assembly.zsh" unter <https://bitbucket.org/tritexassembly> *Single-Copy*-Regionen > 100 bp erhalten und diese auf die *Contig*-Assemblierungen kartiert (Parameter '-l 20G -K5G -x asm5').

Mit diesen Datensätzen wurde das *Scaffolding* und die manuelle Zuordnung in R (Team, 2022) unter Verwendung der TRITEX-*Pipeline* durchgeführt. Zunächst wurden die Informationen von den *Contigs* und dem *Guide* zusammengeführt, um zu prüfen, ob es sich um chimäre *Contigs* handelt, und um sie auf den Chromosomen zu platzieren. Nur die Mutante wies ein chimäres *Contig* auf (contig_447 auf 1 Mb-1.5Mb). Dieses *Contig* wurde aufgelöst, und dann wurde das *Scaffolding* mit den Hi-C Reads durchgeführt. Hierfür wurden inter- und intrachromosomale Hi-C-Matrizen erstellt und manuell korrigiert. Die erzeugten Pseudomoleküle wurden in FASTA-Sequenzen umgewandelt. Die Pseudomolekül-Assemblierung ermöglicht die Suche nach großen strukturellen Variationen in den Daten. Zunächst wurden die beiden Assemblierungen (Donor-Linie 'Golden Promise' DH1-6 und der *pdil5-1*-Mutante 16-12) mit minimap2 (Li, 2018) mit den Parametern "-c -f 0.005 -x asm5 -eqx" ausgerichtet, um die Kolinearität zu überprüfen. Zur Überprüfung von strukturellen Variationen und Syntenie wurde die Pipeline SyRI (Goel et al., 2019) mit den Parametern "-k -F P" verwendet. Um die Syntenie-Informationen darzustellen, wurde die Funktion "plotsr" von SyRI mit dem Parameter "-s 300000" angewendet. Die erneute Ermittlung von Sequenzvarianten erfolgte wie in 2.4. beschrieben.

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

Die Assemblierung von Pseudomolekülen und die Suche nach strukturellen Variationen erfolgte wie in 2.5.1. beschrieben. 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 des cv. Darmor-bzh (Rousseau-Guetin et al., 2020) als Referenz für die Donor-Linie durchgeführt, während die Mutante an den Pseudomolekülen der Donor-Linie ausgerichtet wurde. Es wurden einige chimäre *Contigs* gefunden, die mit TRITEX aufgelöst wurden: in der Donor-Linie contig_1 (2,5-4Mb) und contig_4 (6-7Mb) sowie in der Mutante contig_20 (30-31Mb). Der Schritt der manuellen Korrektur war schwieriger als bei Gerste. In der Donor-Linie wurde eine Translokation fixiert, die bei der Assemblierung zwischen den Chromosomen A07-C06 entstanden war, und *Contigs* wurden von

A01 nach C01 und A02 nach C02 verschoben. Danach wurden gute Hi-C-Matrizen erhalten und die Pseudomoleküle generiert. Diese wurden dann als Referenz für die Assemblierung der Mutante verwendet.

2.6. *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 potenziellen *Off-Target*-Mutationen zu übersehen), wobei auf Änderungen 1.000 bp vor und nach dem Zielmotiv geprüft wurde. Alle 74 gefundenen potenziellen *Off-Targets* sind in Tabelle A2 im Anhang aufgeführt.

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. 22-*mers* wurden aus der Vektorsequenz 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.

2.7. Verifizierung spontan entstandener Hintergrundmutationen in Gerste

Einige der in der initialen WGS gefundenen SNVs und Sequenzierfehler sollten experimentell differenziert werden. Insgesamt sollten 25 SNVs verteilt über alle *Contigs* untersucht und in 4 Geschwisterpflanzen (M_3) der verwendeten *pdil5-1*-Mutante und der Donorlinie 'Golden Promise' DH1-6 verifiziert werden. Die dafür verwendeten Primer sind im Anhang in der Tabelle A1 aufgeführt. Zunächst wurde die optimale Annealing-Temperatur (T_{opt}) eines jeden Primerpaares in Gradienten-PCRs ermittelt. Daraufhin erfolgte die Amplifikation der jeweiligen Zielregionen mit einer 25 μ L Standardreaktion, die folgende Komponenten enthielt: 5 μ L Go Taq Puffer und 0,25 μ L Go Taq Polymerase (Promega, Madison, WI, USA), 400 nM dNTPs (Thermo Fisher Scientific), jeweils 500 nM Forward- und Reverse-Primer und eine Endkonzentration von ca. 100 ng DNA. Die PCRs wurden in einem Mastercycler® nexus (Eppendorf, Hamburg, Deutschland) durchgeführt und das Standardprogramm war wie folgt: 95 °C/ 3 min; [95 °C/ 15 s; T_{opt} / 30 s; 72 °C/ 30 s] x 35 + 72 °C/ 5 min. Jeweils 12 μ L aufgereinigtes PCR-Produkt + 2 μ L des jeweiligen Forward-Primers wurden zum Sanger-Sequenzieren geschickt (LGC Genomics GmbH, Berlin, Deutschland). Die erhaltenen Sequenzierungsergebnisse wurden mit Hilfe der Programme Clone Manager 9 (Scientific & Educational Software, Morrisville, NC, USA) und Unipro UGENE (Unipro, Novosibirsk, Russland) ausgewertet.

2.8. 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. Es wurden verschiedene Mastermixe (Tabelle 2) getestet und diese nach Herstellerangaben verwendet.

Tabelle 2: Verwendete qPCR-Mastermixe

Bezeichnung	Firma
Takara Probe qPCR Master Mix	Takara Bio Europe SAS (Saint Germain en Laye, Frankreich)
Kapa Probe Fast	Roche (Mannheim, Deutschland)
Kapa Probe Force	Roche (Mannheim, Deutschland)
Fast qPCR Master Mix plus low ROX	Eurogentec (Seraing, Belgien)
GoTaq Probe qPCR Master Mix	Promega (Madison, WI, USA)
TaqMan Fast Advanced Master Mix	Applied Biosystems/ Thermo Fisher Scientific (Waltham, MA, USA)
Inhibitor-Tolerant qPCR Mix	Meridian Bioscience (Cincinnati, OH, USA)

Tabelle 3: Verwendete Oligonukleotide für qPCR- und ddPCR-Assays in Gerste. Die hier angegebenen, in den Reaktionen verwendeten Endkonzentrationen der Oligonukleotide wurden in Optimierungsassays ermittelt.

Name	Sequenz (5'→3')	Konzentrationen (µM)	
		qPCR	ddPCR
System 1			
272_PDIL_qRT_fw	GACCGGCCAGAGAAATACATC	0,8	0,9
272_PDIL_qRT_rv	GAGACGAGGAGAAGGCGGATAC	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	GAGACGAGGAGAAGGCGGATAC	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	AGACGAGGAGAAGGCGGATAC	0,8	0,8
IK437_HvPDIL_Reference_Probe	HEX/GGACAGATGGCGACGGTGAATCTT/BMN-Q535, 2x	0,2	0,2
IK438_HvPDIL_SNV_Probe	CY5/CCAGTC+G+GGTGGTTCGAT/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

Eine 20 µL Standard-qPCR-Reaktion umfasste in der Regel 10 µL des jeweiligen Mastermix (abhängig von den jeweiligen Konzentrationen), die Endkonzentrationen der jeweiligen Primer und Sonden (Tabelle 3) sowie 5 µL DNA [20 ng/µL]. Die Primer und Sonden der jeweiligen getesteten Systeme sind in Tabelle 3 aufgeführt. Das Standard-PCR Programm wurde unter den folgenden Bedingungen durchgeführt: 95 °C/ 3 min; [95 °C/ 10 s; 61 °C/ 30 s] x 45. Zur Auswertung der Ergebnisse wurde die LightCycler® 480 Software (Roche Life Science) und die CFX Maestro Software (Bio-Rad) verwendet.

2.9. Etablierung der Detektion von Einzelnukleotidveränderungen 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 3) getestet und auf dieser Basis entschieden, für die weitere Etablierung und Validierung System 2 zu verwenden. Die Ergebnisse wurden mit der CrystalMiner Software (Stilla) analysiert. Bei Impetus Bioscience wurde das System von Bio-Rad (C1000 Touch) in Kombination mit dem ddPCR Supermix for Probes (Bio-Rad Laboratories, Hercules, CA, USA) nach Herstellerangaben verwendet. Für die ddPCR-Assays in Gerste wurde System 2 (Tabelle 3) verwendet und die in Raps verwendeten Primer sind in Tabelle 4 aufgeführt. Die Durchführung der ddPCR war für beide Systeme gleich. Insgesamt wurden 20 µL einer Standardreaktion (10 µL 2x ddPCR Supermix for Probes, 900 nM von jedem Primer, 200 nM von jeder Sonde und 5 µL DNA [20 ng/µL]) mit 70 µL Droplet Generator Oil in einem QX200™ Droplet Generator (Bio-Rad) gemischt. Mit 40 µL der erzeugten Tröpfchen pro Reaktion wurde eine PCR-Amplifikation auf einem Bio-Rad C1000 Touch-Thermocycler durchgeführt. Das PCR-Programm war wie folgt: 95 °C/ 10 min; [94 °C/ 30 s; 60 °C/ 1 min] x 45 + 98 °C/ 10 min. Für jede ddPCR Reaktion wurden die Ergebnisse im HEX- und FAM-Kanal mit einem QX200-Reader (Bio-Rad) erfasst und mit QuantaSoft (Bio-Rad) analysiert. Der *Threshold* wurde manuell oberhalb des Clusters negativer Partitionen eingestellt. Die Details der ddPCR-Reaktionen sind den angefügten *Standard Operating Procedure* (SOP)-Entwürfen zu entnehmen.

Tabelle 4: 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/ CAACTTTCATCTCTC /MGB/BMN-Q535	0,2

2.10. Charakterisierung der Zielregionen mittels Tiefensequenzierung von Amplikons in Gerste

Die *short-read* Amplikon-Tiefensequenzierung wurde vom Unterauftragnehmer Planton mit einem MiSeq-System (Illumina, San Diego, CA, USA) wie im als Anlage bereitgestellten SOP-Entwurf beschrieben durchgeführt. Kurz zusammengefasst wurde die Zielregion mit den letztendlich verwendeten Primern NILP-IPKv2-F3 –TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-cggaccggcccagagaaat- und NILP-IPKv2-R3 –GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGgaagacggggagacgaggag- amplifiziert. Der 20 µL PCR-Ansatz enthielt 5 µL Template-DNA (40 ng/µL), 10 µL 2x Mastermix Platinum Superfi

II (Invitrogen/ Thermo Fisher Scientific, Waltham, MA, USA), je 0,1 µL Forward- und Reverse-Primer (100 µM), 0,2 µL SybrGreen (1/300 Vorverdünnung) und 4,6 µL HPLC-Wasser. Das PCR-Programm war wie folgt: 98 °C/ 30 s; [98 °C/ 10 s; 60 °C/ 10 s; 72 °C/ 30 s + Fluoreszenzaufnahme] max. 35 Zyklen + 72 °C/ 5 min; 16 °C/ Pause. Die PCR-Amplifikate wurden chemisch mittels ExoI/CIP-System von NEB (New England Biolabs, Ipswich, MA, USA) aufgereinigt, wodurch die Oligos degradiert und dephosphoryliert wurden. Hierfür wurden 6 µL PCR-Produkt und 2 µL ExoI/CIP zusammen für 15 min bei 37 °C inkubiert. Die Deaktivierung der Enzyme erfolgte dann für 10 min bei 90 °C. Die Indexierung der chemisch aufgereinigten Amplifikate erfolgte mittels einer zweiten PCR und dem Nextera XT Index Kit v2 (Illumina). Der 50 µl PCR-Ansatz enthielt 5 µl aufgereinigtes Template, jeweils 5 µL S- und N-Oligos, 25 µL 2x Q5-Mastermix (NEB) und 10 µL HPLC-Wasser. Das zweite PCR-Programm war wie folgt: 95 °C/ 5 min; [95 °C/ 30 s; 55 °C/ 30 s; 72 °C/ 30 s]x 7; + 72 °C/ 5 min; 16 °C/ Pause. Die Amplifikate aus der Indexierung wurden mittels 2-3 % Agarose-Elektrophorese aufgetrennt und über ein Silica-Membran-System aufgereinigt. Die Vorbereitung und Erstellung der *Library* erfolgte mit dem MiSeq Reagent Kit v2 (300, Illumina) nach Herstellerangaben. Nach dem MiSeq-Sequenzierungslauf wurden die ausgegebenen bcl-Dateien manuell ins FastQ-Format umgewandelt.

Die Sequenzierungsergebnisse wurden mit Hilfe der in diesem Projekt entwickelten Analyse-Pipeline analysiert (siehe 2.11.). Bei Planton erfolgte die Auswertung mit einem von diesem Unterauftragnehmer entwickelten, proprietären Protokoll. Die für diese Auswertungs-Methode benötigten Programme sind in Tabelle 5 aufgeführt.

Tabelle 5: Software für das Auswertungs-Protokoll von Planton

Bezeichnung	Version	Link
usearch	(64bit Version)	https://www.drive5.com/usearch/
CutAdapt	ab v4.x	https://cutadapt.readthedocs.io/en/stable/
fastx_ToolKit	ab 0.0.13 (2010)	http://hannonlab.cshl.edu/fastx_toolkit/
Excel ab 2019	64bit Version	https://www.microsoft.com/de-de/microsoft-365
ASAP für Excel	ab v7.9	https://www.asap-utilities.com/

2.11. Etablierung einer Analysepipeline für die Tiefensequenzierung von Amplikons

Zur Detektion und Charakterisierung von Einzelnukeotid-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 mit Hilfe 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 wird gezählt. Seltene 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 (Katoh and Toh, 2008) durchgeführt. Die alignierten Sequenzen werden dem Benutzer im FASTA-Format ausgegeben. Diese konnten wiederum mit dem Programm Unipro UGENE (Unipro, Novosibirsk, Russland) ausgewertet werden.

3. Ergebnisse

3.1. Ausführliche Darstellungen der wichtigsten Ergebnisse

3.1.1. *Whole-Genome Sequencing* von 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.

Zur Zeit der Antragstellung, -begutachtung und Bewilligung des Projektes stellte die Illumina-Technologie den Stand der Technik für Genomsequenzierungen dar, was dem experimentellen Konzept des Projektes auch entsprechend zugrunde gelegt worden ist. Ein Technologiewechsel von der Illumina- zu der PacBio *Hifi Longread*-Sequenzierung wurde im ersten Projektjahr bewilligt. Für die PacBio-basierte Genom-Sequenzierung von Gerste mit hinreichender Sequenziertiefe (*Coverage*) war es uns möglich ohne Aufstockung der für die Genom-Sequenzierung bewilligten Projektmittel zwei Gersten-Linien (Mutante vs. Wildtyp-Linie) zu sequenzieren. Das bewilligte Projekt-Budget reichte gleichermaßen für die Genom-Sequenzierung zweier Raps-Linien (C3E3 und C3E4) mittels PacBio. Die Sequenzdaten einer dritten Raps-Linie (Wildtyp) wurden uns darüber hinaus vom Bundesamt für Verbraucherschutz und Lebensmittelsicherheit (BVL) zur Verfügung gestellt, was über eine Kooperationsvereinbarung zwischen IPK und BVL verbindlich geregelt worden ist. Im Gegenzug werden dem BVL die im DETECT-Projekt generierten Raps-Sequenzen für ein dort bearbeitetes Projekt (mit anderer Zielstellung) zur Verfügung gestellt. Die methodischen Spezifikationen der Genomsequenzierungen wurden zwischen BVL und IPK zur Sicherstellung der Datenkonformität aufeinander abgestimmt. Die Sequenzierung erfolgte für beide Einrichtungen anhand der gleichen Technologieplattform (PacBio Sequel II HiFi) mit einer Tiefe von mindestens 20-facher Genomabdeckung.

Für die Donor-Linie 'Golden Promise' DH1-6 und die Mutante wurde eine 35-fache bzw. 21-fache Genomabdeckung erzielt, womit die Vorgaben erfüllt sind. Die Gesamtlängen der Genomassemblierungen (in Form von *Contigs*) kommen den erwarteten Genomgrößen nahe (Tabelle 6; Anzahl der *Contigs*: 1.697 und 1.288). Die Standardassemblierung 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.

Tabelle 6: 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

Ermittlung von Sequenzvarianten

Insgesamt wurden in der initialen WGS 2.816 homozygote und 2.055 heterozygote SNPs in der Mutante im Vergleich zur Donor-Linie identifiziert (Tabelle 7). Diese Häufigkeiten waren um mehrere Größenordnungen niedriger als für bisher beschriebene SNPs zwischen unabhängig entstandenen Gerstensorten (durchschnittlich 1 sortenspezifischer SNP pro Gb, ~5 pro Jahr; entsprechend der für *B. dystachion* beschriebenen Mutationsrate, die für die am nächsten mit der Gerste verwandten Art verfügbar ist [Wang et al., 2019]), überstiegen jedoch den derzeit angenommenen Umfang der pro Generation spontan entstehenden heritablen Mutationen (zwischen 1 SNP pro 100 Mb und 1 SNP pro 1 Gb). Eine Fehlerquelle dieser Variantenbestimmung konnte daher nicht ausgeschlossen werden. Die von uns für die Genom-Editierung verwendete Donor-Linie 'Golden Promise' DH1-6 ist eine in unserem Labor entstandene doppelhaploide (d.h. vollständig homozygote) Linie. Zudem gilt es in diesem Kontext zu beachten, dass die sequenzierte *pdil5-1*-Mutante bereits in der dritten Generation ist und dass durch die *Whole-Genome* Sequenzierung in jedem Individuum auch nicht-heritable, ausschließlich im Soma vorliegende Mutationen in bislang unbekannter Proportion detektiert werden. Die Bestimmung geeigneter Filter zur Ermittlung von zufälligen Mutationsraten ist ein aktuelles Forschungsfeld und selbst bei dem vergleichsweise kleinen *Arabidopsis*-Genom eine Herausforderung (Monroe et al. 2022; Liu und Zhang 2022). Bei der Durchführung zukünftiger ähnlicher Experimente könnten technische Wiederholungen von Vorteil sein, um die Anzahl der technischen Artefakte zu minimieren.

Tabelle 7: Gesamtzahl der alternativen homozygoten und heterozygoten SNPs und Indels, die in der mit der Donorlinie 'Golden Promise' DH1-6 verglichenen Mutanten-Probe nach der Filterung in der WGS (vor Hi-C Sequenzierung) ermittelt wurde.

Variante	Genotyp	Mutante
SNPs	Homozygot	2.816
	Heterozygot	2.055
	Gesamt	4.871
Indels	Homozygot	302
	Heterozygot	2.090
	Gesamt	2.392

3.1.2. Hi-C Sequenzierung und Assemblierung der Genomdaten zu chromosomalen Pseudomolekülen in Gerste

Für die Erstellung der Pseudomoleküle, die eine Chromosomensequenz repräsentieren, wurde die 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 4267,9 Mb für die Donor-Linie und 4244,7 Mb für die Mutante (Tabelle 8), was mit dem vorherigen Ergebnis der *Contig*-Assemblierung übereinstimmt (Tabelle 6). 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.

Tabelle 8: Assemblierungsdaten des *Scaffoldings*. Die Pseudomoleküle wurden mit der TRITEX-Pipeline assembliert. Einige *Contigs* konnten keinem Pseudomolekül (Chromosom) zugeordnet werden.

Genom	Chromosom	Länge (Mb)	Anzahl der <i>Contigs</i>	N50 (Mb)	Max. Länge (Mb)	Min. Länge (kb)
Donor-Linie GP DH1-6	chr1H	517,4	58	16,2	47,4	502,6
	chr2H	656,4	60	20,2	43,8	504,1
	chr3H	622,5	72	13,8	43,2	502,3
	chr4H	612,1	87	12,1	53,1	500,3
	chr5H	590,0	66	17,8	44,2	600,5
	chr6H	560,3	66	20,3	53,4	630,3
	chr7H	632,2	77	14	31,7	526,1
	1-7	4.267,9	486	15,9	53,4	500,3
	Nicht zugeordnete <i>Contigs</i>	77	1.211	0,1	0,5	1,1
pdil5-1 Mutante	chr1H	516,7	67	10,4	51,9	527,5
	chr2H	656,0	73	16,8	45,1	561,6
	chr3H	619,9	82	14,3	55,7	568,8
	chr4H	610,7	88	12,3	78,9	514,9
	chr5H	589,4	77	12,4	31,2	512,9
	chr6H	558,4	73	16,3	57,0	619,4
	chr7H	632,2	74	15,3	43,6	536,3
	1-7	4.244,7	534	13,3	78,9	512,9
	Nicht zugeordnete <i>Contigs</i>	61,3	756	0,1	0,5	1,4

Eine erneute Ermittlung von Sequenzvarianten nach der Assemblierung zu Pseudomolekülen ergab eine deutlich reduzierte Anzahl an gefundenen SNPs und InDels. Insgesamt wurden 48 homozygote und 80 heterozygote SNPs und 358 homozygote und 1.730 heterozygote InDels in der Mutante im Vergleich zur Donorlinie identifiziert (Tabelle 9).

Tabelle 9: Gesamtzahl der alternativen homozygoten und heterozygoten SNPs und Indels, die in der mit der Donorlinie 'Golden Promise' DH1-6 verglichenen Mutanten-Probe nach der Assemblierung zu Pseudomolekülen ermittelt wurde

Variante	Genotyp	Mutante
SNPs	Homozygot	48
	Heterozygot	80
	Total	128
Indels	Homozygot	358
	Heterozygot	1.730
	Total	2.088

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 10 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 10). Die erzeugte 1 bp-Deletion ist in der Tabelle als Nummer 4 erfasst. Der nächstgelegene InDel liegt 1,4 Mb strangabwärts (Tabelle 10). 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 mit etwa 0,5 cM einer Rekombination pro .200 Gameten entspricht. Eine wissenschaftlich fundierte Identifikation des für eine Mutation verantwortlichen (technischen oder spontan verlaufenen) 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.

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). Daher wurden die Pseudomoleküle der Donor-Linie und der Mutante zueinander ausgerichtet. 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 3) mit einigen kleinen Lücken, einer Translokation von 636.105 bp in der telomerischen Region des kurzen Arms von Chromosom 1H und einigen Inversionen länger als 700 kbp in den Zentromerregionen 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 Genom-assemblierung von Gerste, in der solche telomeren und zentromeren Sequenzen verlässlich aufgelöst sind. Selbst die MorexV3 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.

Tabelle 10: 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 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

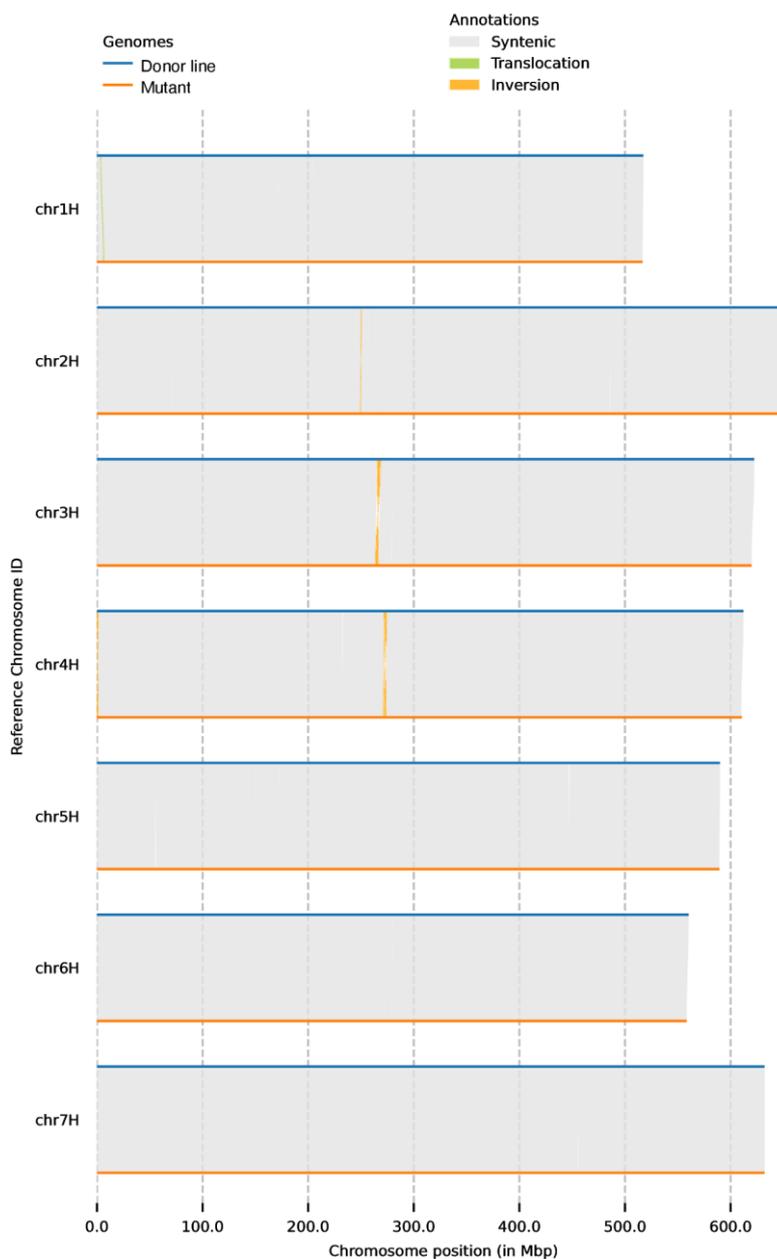


Abbildung 3: Synteny-Plot zwischen der Donor-Linie und der Mutante. Es werden nur Strukturveränderungen gezeigt, die länger als 300 kbp sind.

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.

3.1.3. 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-mers*). Diese Suche ergab, dass keine rekombinante DNA, also keine Transgenrückstände, in der Mutante vorliegen. 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 Detektion 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 dem 3 bp umfassenden, von Cas9 erkannten PAM. Bei der *Off-Target*-Analyse, die bis zu 4 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 74 potenziellen *Off-Targets* in der VCF-Datei überprüft. Dabei wurde nach Sequenzvarianten (sowohl SNPs als auch Indels) im 22 bp umfassenden Zielmotiv des möglichen *Off-Targets* gesucht. Sequenzvarianten wurden selbst ohne Anwendung der *Coverage*-Filter nicht gefunden. Darüber hinaus wurden auch die Bereiche von 100 bp beidseitig der potenziellen *Off-Target*-Stellen überprüft, wobei keine Sequenzvarianten gefunden wurden.

3.1.4. *Whole-Genome Sequencing*, Hi-C Sequenzierung und Assemblierung der Genomdaten zu chromosomalen Pseudomolekülen 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 (ca. 1 GB), obwohl die Assemblierungen relativ stark fragmentiert waren (Tabelle 11). Um Assemblierungen im Chromosomenmaßstab zu erhalten und zudem zu überprüfen, ob große strukturelle Veränderungen vorliegen, wurde zusätzlich eine Hi-C-Sequenzierung durchgeführt.

Tabelle 11: Assemblierung der WGS-*Reads* von der Donor-Linie 'Mozart' und der beiden Mutanten C3E3 and 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 nur für die Sequenzierung von 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 bezüglich ihrer endgültigen Länge auf, der auch in der *Contig*-Assemblierung beobachtet wurde (Tabelle 11, Tabelle 12). Die Donor-Linie hat eine kleinere N50, was sich in mehr nicht zugeordneten *Contigs* und mehr *Contigs* pro Pseudomolekül widerspiegelt. Die Größen der Pseudomoleküle A01 und A02 variierten erheblich zwischen den Assemblierungen (Tabelle 12).

Tabelle 12: Assemblierungsdaten des *Scaffoldings* für die Raps-Donor-Linie 'Mozart' und die Mutante C3E4. Die Pseudomoleküle wurden mit der TRITEX-Pipeline assembliert. Einige *Contigs* konnten keinem Pseudomolekül (Chromosom) zugeordnet werden.

Genom	Chromosom	Länge (Mb)	Anzahl der <i>Contigs</i>	N50 (Mb)	Max. Länge (Mb)	Min. Länge (kb)
Donor-Linie	chrA01	47.2	14	4.3	6.5	838.5
	chrA02	38.4	7	8.5	14.8	1327.1
	chrA03	44.3	5	13.5	16.8	535.9
	chrA04	28.6	3	10.2	14.1	4309.7
	chrA05	41.5	11	7.6	14.9	585.1
	chrA06	58.7	6	31.3	31.3	714.3
	chrA07	41.9	7	10.9	12.1	998.4
	chrA08	29.9	9	7.8	8.2	697.5
	chrA09	74.7	10	10.6	17.8	562.7
	chrA10	27.3	3	24	24	897.2
	chrC01	57	10	11.1	21.2	871.8
	chrC02	66.4	17	6.7	12.3	548.5
	chrC03	79.9	8	16.2	26.9	650.3
	chrC04	72.5	10	11.1	22.9	573.8
	chrC05	59.3	3	20.9	29.1	9296.7
	chrC06	47.1	4	29.9	29.9	656
	chrC07	63	7	26.1	26.3	550.8
	chrC08	52.9	10	12.5	17.4	531.9
	chrC09	69.1	16	5.9	18.4	682.5
	1-09	1325.4	160	12.1	31.3	531.9
	Nicht zugeordnete <i>Contigs</i>	324.9	6268	0.1	1.9	1.3
Mutante C3E4	chrA01	37.4	8	9.9	11.7	839
	chrA02	34.4	5	8.6	15	1286.8
	chrA03	44.2	3	29.4	29.4	1472.2
	chrA04	27.9	2	14.1	14.1	13824
	chrA05	41.4	5	30.1	30.1	787.2
	chrA06	59	5	36.3	36.3	797.2
	chrA07	42	7	8.1	11.4	1134.3
	chrA08	29.7	5	15.3	15.3	1027.9
	chrA09	75.5	7	18.8	27.7	564.3
	chrA10	27.4	2	26.9	26.9	522
	chrC01	55.5	3	23.1	23.5	8828.1
	chrC02	66.4	8	10.8	30.5	1098.1
	chrC03	79.8	4	41	41	5415.3
	chrC04	72.2	4	49	49	2202.4
	chrC05	59.1	3	20.9	29.1	9164.5
	chrC06	49	3	31.8	31.8	5138.7
	chrC07	64.7	8	26.4	31.8	537.6
	chrC08	52.9	5	14.4	21.6	983.1
	chrC09	69.3	10	12.6	31.5	685.8
	1-09	1072.5	97	20.9	49	522
	Nicht zugeordnete <i>Contigs</i>	84.6	1287	0.1	5.1	13.4

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 4). Diese Unterschiede könnten darauf zurückzuführen sein, dass diese Regionen schwieriger zu assemblieren sind, aber auch darauf, dass Raps polyploid ist und ein erhebliches 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.

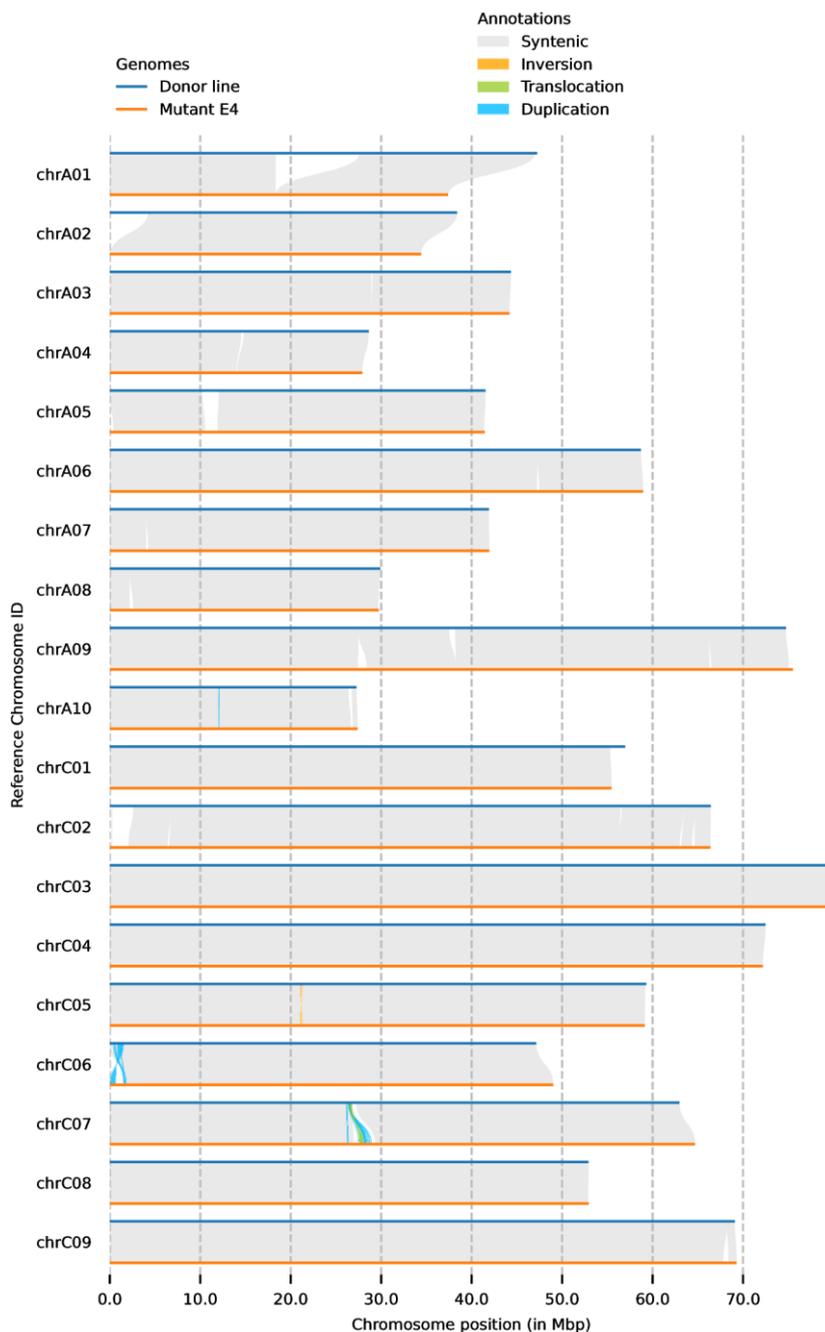


Abbildung 4: Syntenie-Plot zwischen der Donor-Linie und der Mutante C3E4. Es werden nur strukturelle Veränderungen gezeigt, die länger als 30 kbp sind.

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

Einzelnukleotidveränderungen können in einer qPCR-Reaktion nachgewiesen werden, sofern die verwendeten Oligonukleotide spezifisch genug sind. In diesem Projekt wurden ausschließlich fluoreszenzmarkierte Hybridisierungssonden 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 (Dr. Olav Grundmann, 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 Referenzsonde 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. Es wurde zunächst ein kompetitives Assay (Abbildung 5A, Dr. Olav Grundmann, Auftragnehmer Impetus Bioscience) und ein *Drop-off* Assay (Abbildung 5B, Dr. Martin Becker, IPK/ Stilla) entwickelt. Dabei wurden sowohl normale Sonden als auch MGB- und LNA-Sonden getestet.

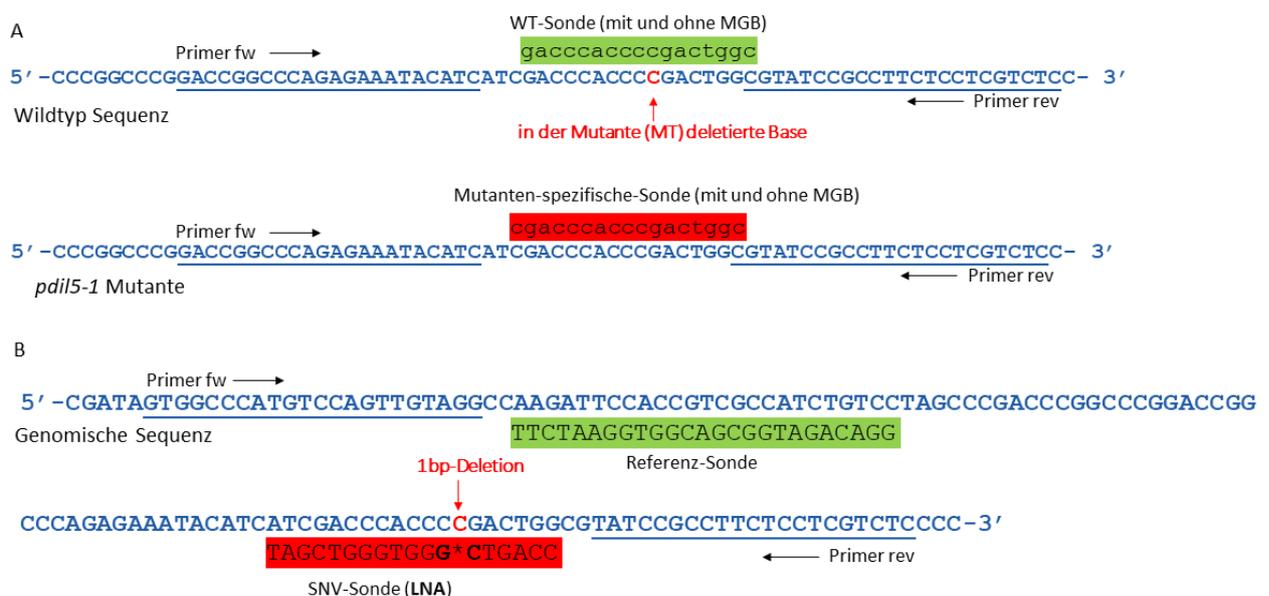


Abbildung 5: Design des kompetitiven (A; System 1 und 2) und des *Drop-off*-Assays (B; System 3)

Tabelle 13: Übersicht der verwendeten PCR-Systeme

Name	Beschreibung
System 1	Kompetitives Assay mit Primern 272 fw/rv und Sonden 273/274
System 2	Kompetitives Assay mit Primern 272 fw/rv und Sonden 273_NEU/ 274_NEU
System 3	Drop-off Assay mit Primern IK435/IK436 und Sonden IK437/IK438
System 4	Assay mit Primern IK435/IK441 und Sonde IK437
System 5	Assay mit Primern IK435/IK442 und Sonde IK437
System 6	<i>Hor3</i> Referenz Assay mit Primern IK424/IK425 und Sonde 426

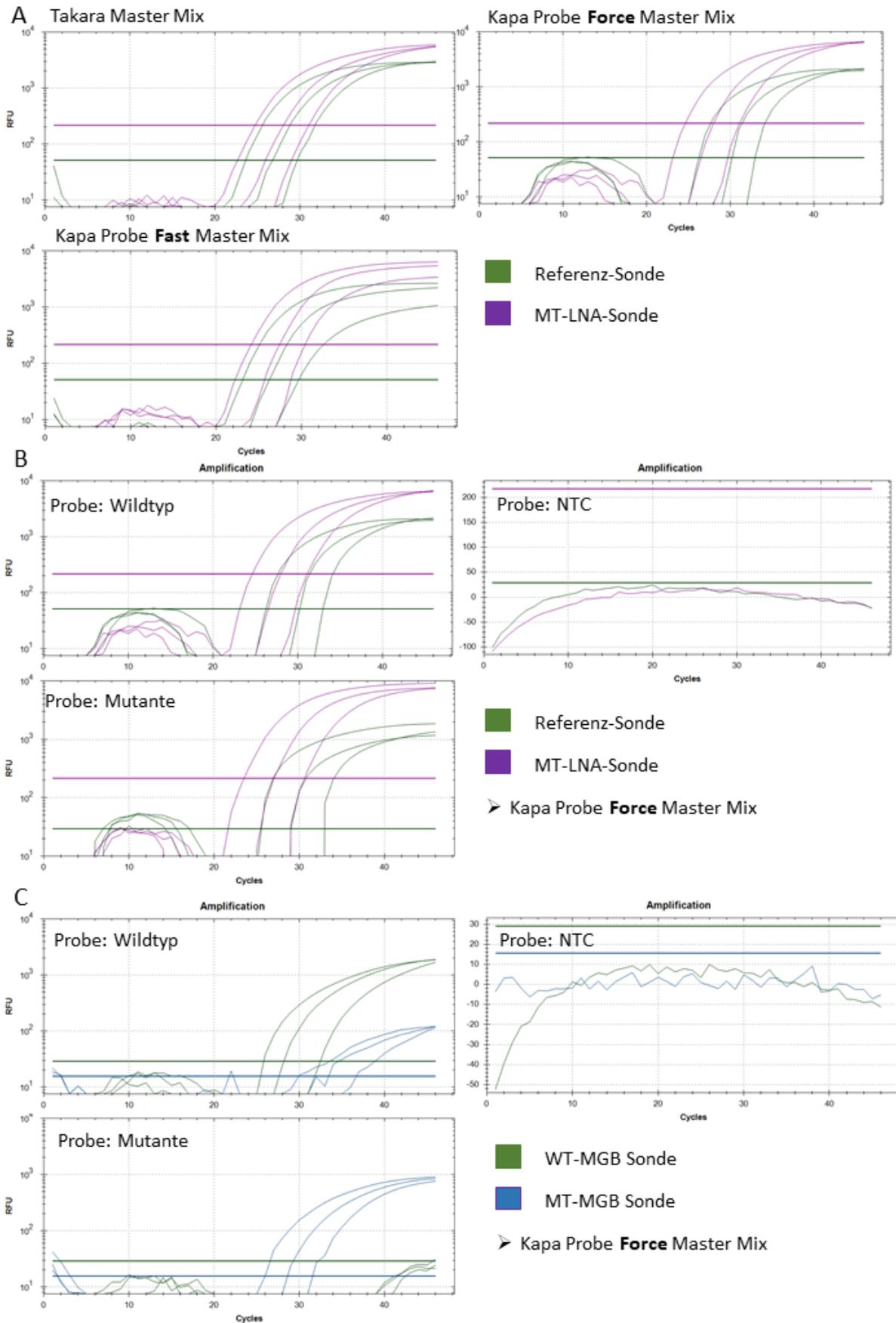


Abbildung 7: Drop-off und kompetitives qPCR-Assay, (A) Vergleich des Takara Probe qPCR Master Mix, Kapa Probe Fast Master Mix und des Kapa Probe Force Master Mix im Drop-off Assay mit jeweils der Referenzsonde und der MT-LNA-Sonde (System 3). Es wurden 2 serielle 10-fach Verdünnungen (1:10, 1:10²) von Wildtyp-gDNA und die Startkonzentration von 100 ng pro Reaktion verwendet. (B) System 3 mit Wildtyp- oder Mutanten-gDNA [2 serielle 10-fach Verdünnungen (1:10, 1:10²) und die Startkonzentration von 100 ng pro Reaktion] und NTC, (C) System 1 (MGB-Sonden) mit Wildtyp- oder Mutanten-gDNA [2 serielle 10-fach Verdünnungen (1:10, 1:10²) und die Startkonzentration von 100 ng pro Reaktion] und NTC.

So wurden Standardmastermixe von Impetus Bioscience, der am IPK verwendete Takara Mastermix und die empfohlenen Kapa Probe Mastermixe eingesetzt. In Abbildung 7A ist beispielhaft ein Vergleich der Mastermixe mit dem Primer/Sonden-System 3 gezeigt. Alle anderen dargestellten qPCR-Experimente wurden dann mit dem vom BVL und LGL empfohlenen Kapa Probe Force Mastermix durchgeführt. Lediglich die WT-MGB-Sonde (System 1; Abbildung 7C) und die normale WT-Sonde (System 2) haben relativ spezifisch nur in der WT-Probe gebunden (Tabelle 14). 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 14 dargestellt. Bei diesen starken Kreuzreaktionen ist eine eindeutige Berechnung des Anteils an Mutanten-DNA nicht möglich.

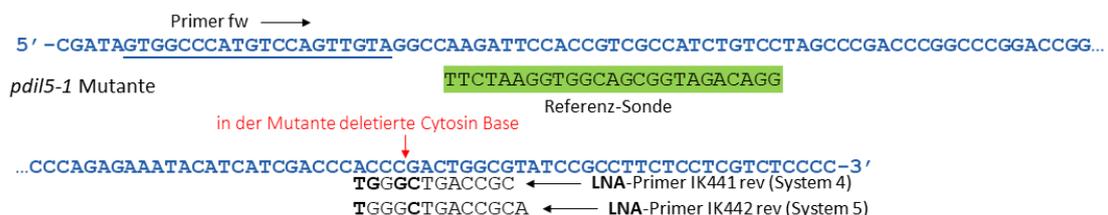
Tabelle 14: Repräsentative qPCR Ergebnisse der Systeme 1 bis 3 unter Verwendung von 100 ng gDNA pro Reaktion und dem Kapa Probe Force Master Mix. n=3, ND=nichts detektiert, CV= Variationskoeffizient der gemittelten Cq-Werte.

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 von den korrekt-positiven *Droplets* durch die Höhe und die Separation der Cluster unterscheiden. In der qPCR ist dies nicht möglich, da sich die Methoden in Bezug auf die Datenerhebung und –auswertung unterscheiden.

Des Weiteren wurden zwei Assays mit zwei verschiedenen reversen LNA-Primern getestet, die auf der Mutation liegen (System 4 und 5, Abbildung 8).

A



B

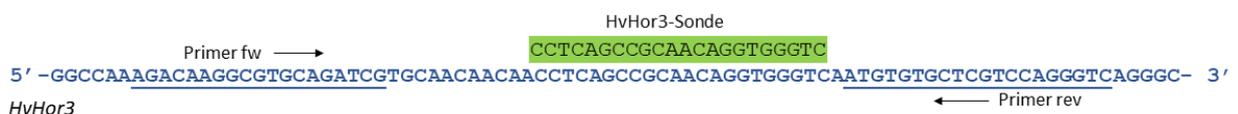


Abbildung 8: Design der Systeme 4 und 5 (A) und des HvHor3 Referenz-Assays (B; System 6)

Neben den Assays für die Zielsequenz-spezifische Mutation wurde ein Gersten-Referenzsystem basierend auf der Publikation von Hernandez et al. (2005) verwendet (System 6). Zum einen erlaubt dies eine vergleichende Betrachtung der Leistungseigenschaften von neu entwickelten mit bereits etablierten Assays und zum anderen dient dies zur Bestimmung der Kopienanzahl eines *Templates*. Normalerweise sollte bei einem neuen Assay gezeigt werden, dass das Zielgen in allen Akzessionen in gleicher Kopienzahl vorliegt, damit es nicht zu Fehlern bei der Bestimmung der Kopienzahl in der GVO-Analytik kommt. Als praktikable Lösung dafür wurde entschieden, ein bereits etabliertes System als Referenz für die Bestimmung der Kopienzahl eines Gens zu verwenden. Das ursprünglich für ein qPCR-Assay entwickelte Referenzsystem amplifiziert ein 73 bp langes Fragment aus dem γ -*Hordein* (*Hor3*)-Gen und war anhand von 10 repräsentativen Gerstensorten getestet worden (Hernandez et al., 2005).

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 Abbildung 9 und Tabelle 15 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. 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 15, 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.

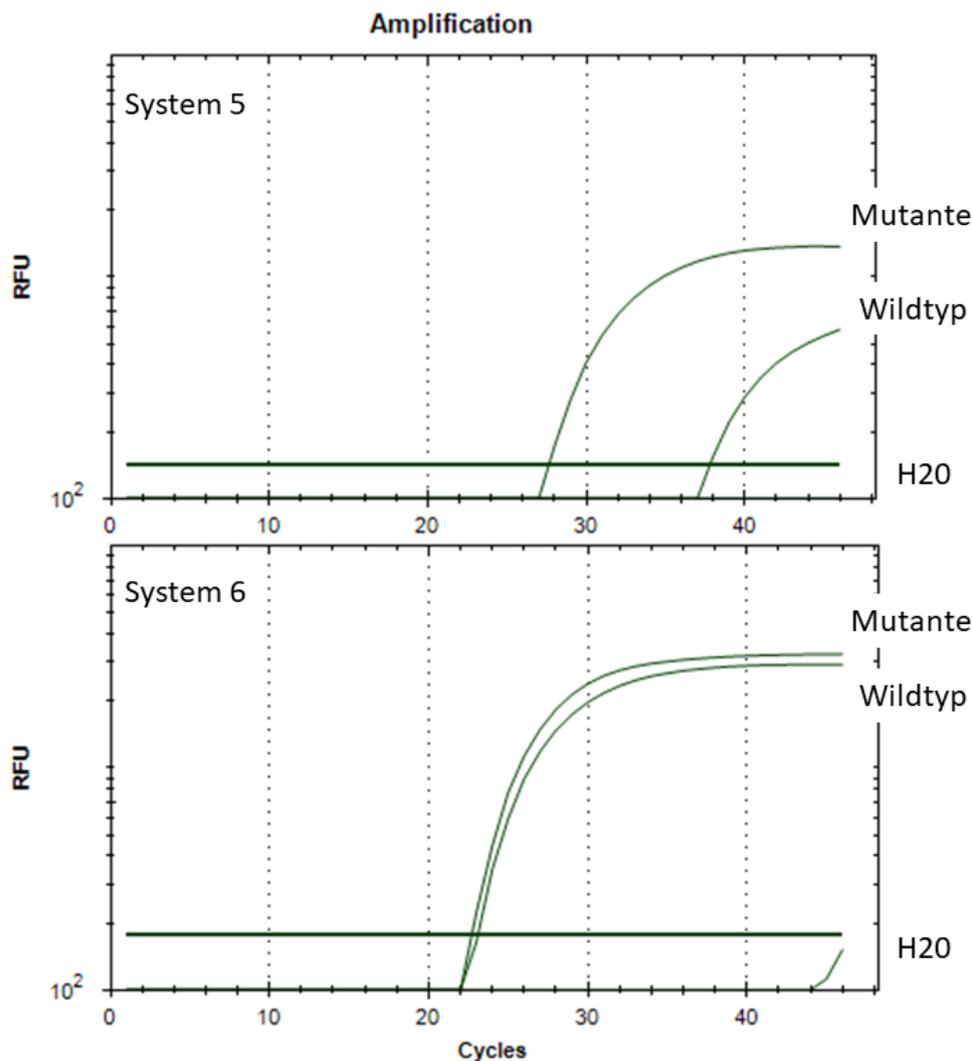


Abbildung 9: Repräsentative Ergebnisse der Systeme 5 und 6 unter Verwendung von 100 ng gDNA pro Reaktion

Tabelle 15: 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

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 Nucleobasen) verwendet (u. A. Chhalliyil et al., 2020), allerdings zeigten sich auch dabei 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.

3.1.6. Etablierung der Detektion von Einzelnukleotid-Veränderungen mittels ddPCR in Gerste

Alle Systeme mit den drei verschiedenen Sondentypen (normal, MGB, LNA) wurden auch in ddPCR-Assays getestet. Die Ergebnisse sind in Tabelle 16 zusammengefasst.

Tabelle 16: Sondentest im ddPCR-Assay

DNA-Probe	Bindung der Sonden im ddPCR Assay					
	System 1		System 2		System 3	
	WT-MGB	MT-MGB	WT	MT	Ref	MT-LNA
Wildtyp	✓	x	✓	*	✓	✓
Mutante	x	x	x	✓	✓	✓

✓ spezifische Bindung; ✓ unspezifische Bindung; x keine Bindung (erwartet); x keine Bindung (unerwartet); *geringe Kreuzreaktion, klare Separierung der Cluster; Ref = Referenz-Sonde

Zunächst wurden die MGB-Sonden (System1) getestet, wobei nur mit der WT-MGB-Sonde in der Wildtyp-Probe ein geringes Signal detektiert werden konnte. Mit den normalen Sonden von System 2 konnten dagegen valide Signale detektiert werden. Es gab 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 10).

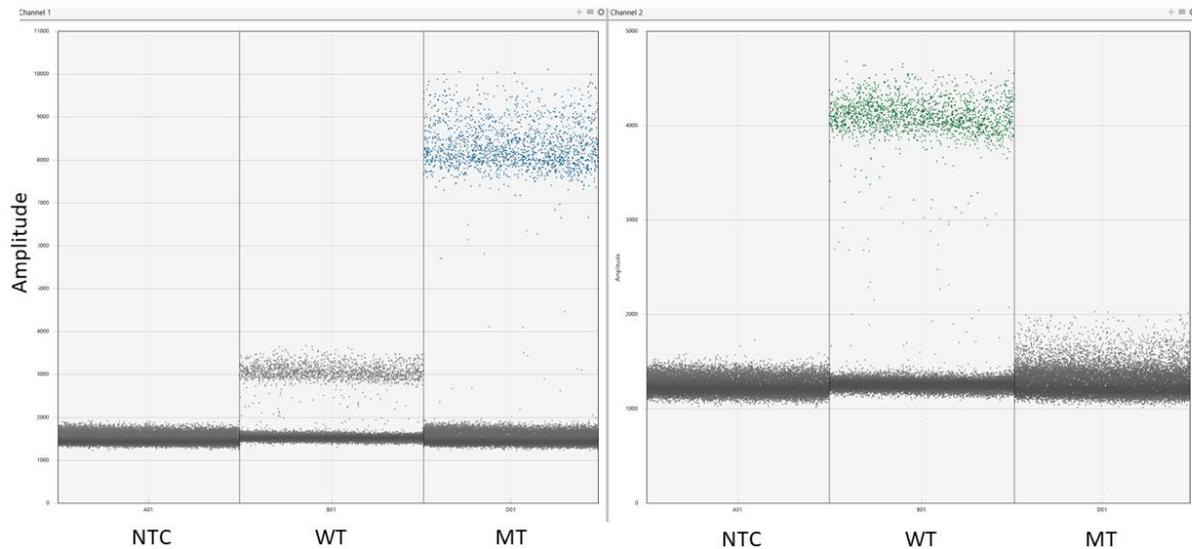


Abbildung 10: Repräsentative Tröpfchenpopulation in der ddPCR unter Verwendung des Systems 2 (Primer 272 fw und rev und Sonden 273_NEU und 274_NEU) und gDNA von 'Golden Promise' DH1-6 (WT), der -1 bp-Mutante 16-12 (MT) und der nicht behandelten Kontrolle (NTC).

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 3.1.5. dargelegten Limitationen nicht weiter getestet. Das in dieser Studie etablierte ddPCR-Assay (System 2) läuft hingegen 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 11). In Übereinstimmung mit den "Mindestanforderungen an die Leistungsfähigkeit von Analysemethoden für GVO-Tests" (ENGL, 2015; ENGL, 2021; Pecoraro et al., 2019) wurde die Leistung der entwickelten ddPCR-basierten Methode (System 2) in Bezug auf Spezifität (Inklusivität und Exklusivität), Sensitivität, Selektivität und Robustheit bewertet (Tabelle 17). Mit diesem ddPCR-Assay konnte gezeigt werden, dass die Nachweisgrenze bei 5 Kopien der Zielsequenz liegt. Dies wurde sowohl anhand von DNA-Mischungen als auch anhand von Saatgutmischungen gezeigt. Die detaillierten Ergebnisse der umfassenden Validierung sind im angefügten SOP-Entwurf für ein ddPCR-Assay in Gerste zu finden.

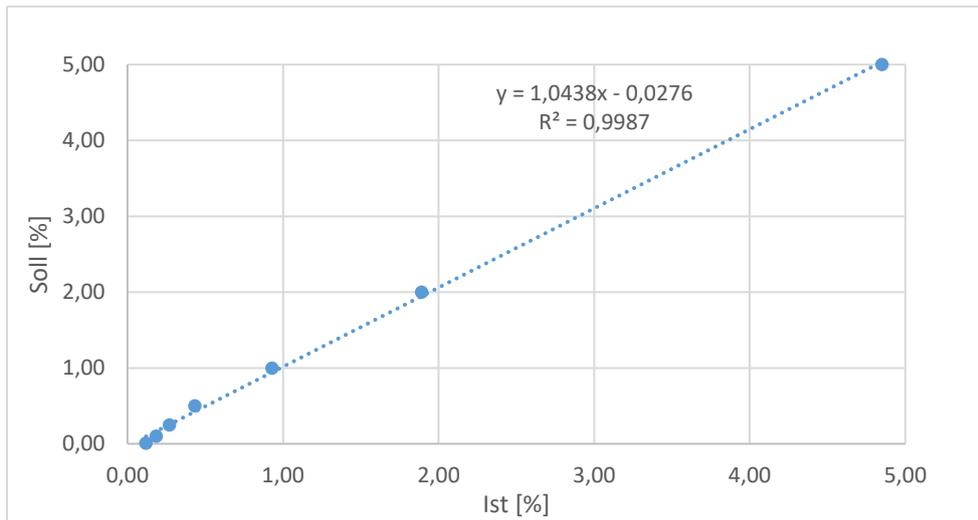


Abbildung 11: Durch ddPCR ermittelte Anteile an Mutanten-DNA in DNA-Mischproben.

Tabelle 17: 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 und 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

3.1.7. Etablierung der Detektion von Einzelnukleotid-Veränderungen mittels 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 12 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.

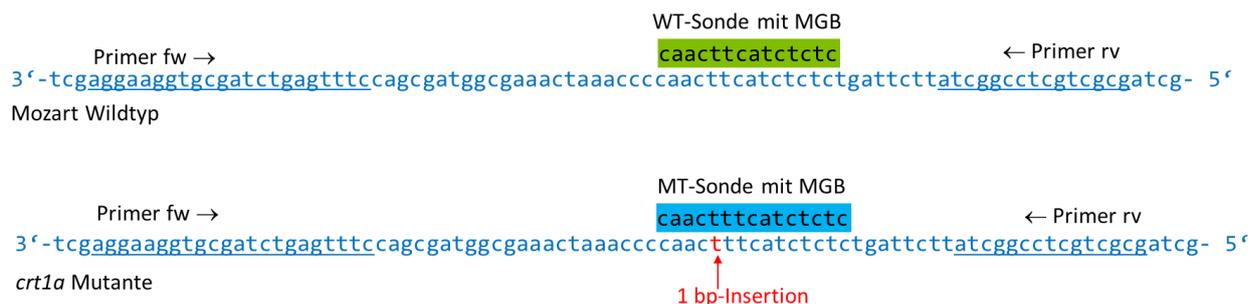


Abbildung 12: Design des ddPCR-Assays für Raps

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, Selektivität und Robustheit bewertet (Tabelle 18). Mit diesem ddPCR-Assay konnte gezeigt werden, dass die Nachweisgrenze bei 5 Kopien der Zielsequenz liegt. Die detaillierten Ergebnisse der Validierung sind im angefügten SOP-Entwurf für ein ddPCR-Assay in Raps zu finden.

Tabelle 18: 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 und 5 Kopien in 10ner-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

3.1.8. Charakterisierung der Zielregionen mittels Tiefensequenzierung von Amplikons in Gerste

Für die Amplikon-Tiefensequenzierung wurde zunächst ein Primerpaar (NGS_F1 und NGS_R1, Abbildung 13) ausgewählt, mit dem ein 189 bp großes Fragment des *PDIL5-1* Gens der Gerste amplifiziert wird.



Abbildung 13: Ausgewählte Primer für die Amplikon-Tiefensequenzierung des *PDIL5-1* Zielmotivs

Allerdings trat mit diesem Primerpaar ein Problem bezüglich der Sequenzierqualität auf. Nur etwa die Hälfte der *Reads* war mit der am IPK entwickelten Pipeline auswertbar. Herr Dr. Weigel (Auftragnehmer Planton) benannte vor allem den reversen Primer als problematisch. Um eine bessere Sequenzierqualität zu erreichen, wurden drei neue Primer getestet (NGS_F2, NGS_F3 und NGS_R3). Das beste Ergebnis lieferte das Primerpaar NGS_F3 und NGS_R3 mit denen ein 75 bp großes Fragment erzeugt wird. Dieses Primerpaar wurde dann für ein erneutes Experiment zur Bestimmung der Sensitivität dieser Methode anhand von DNA-Mischproben verwendet. Neben einer verbesserten Sequenzierqualität konnte mit diesem Ansatz zuverlässig eine Sensitivität von 0,01 % erreicht werden (Tabelle 19, Abbildung 14).

Tabelle 19: Sensitivitätstest anhand von DNA-Mischproben, n = 3 bei 10-1 %, n = 4 bei 0,75-0,25 %, n = 6 bei 0,1–0,01 %

Soil-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

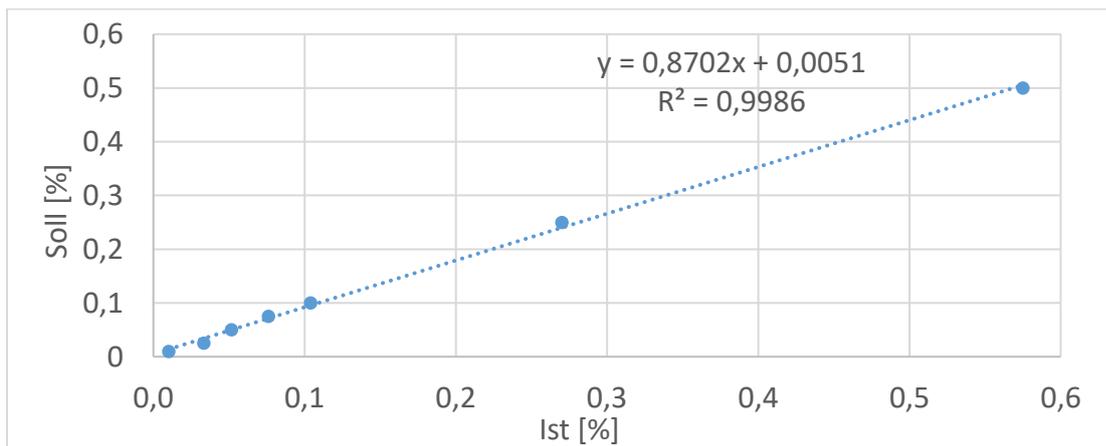


Abbildung 14: Anteil an Mutanten-DNA in DNA-Mischproben ermittelt durch Amplikon-Tiefensequenzierung und dem Primerpaar NGS_F3/NGS_R3

Tabelle 20: 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

Die Spezifität dieses Systems wurde auch 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 20 dargestellt.

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 21 aufgeführt. Zusammengefasst konnte gezeigt werden, dass mit dieser Methode ein Anteil von 0,1 % DNA mit der Zielsequenz sicher nachweisbar ist. Die Details der Amplikon-Tiefensequenzierung und dessen Auswertung sind im entsprechenden SOP-Entwurf für ein Amplikon-Tiefensequenzierungs-Assay in Gerste als Anlage des RapsNMT-Berichtsteils dokumentiert.

Tabelle 21: 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

3.1.9. 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. Zusammengefasst wurden die Sequenzdaten der Illumina MiSeq- und NextSeq-Sequenzierung im FASTQ-Format gefiltert und 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 sie auch ohne Zugriff auf die UNIX-Kommandozeile verwendet werden kann.

Eine vergleichende Auswertung der ersten Experimente der Amplikon-Tiefensequenzierung zeigte, dass die herausgegebenen Ergebnisse der Analysepipelines von Planton und vom IPK sehr ähnlich sind. Zum Beispiel gab es bei der Auswertung des zweiten Experiments zwischen den beiden Protokollen 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.

3.1.10. Verifizierung spontan entstandener Hintergrundmutationen in Gerste

Dieses Arbeitspaket diente der experimentellen Differenzierung von SNVs und Sequenzierfehlern anhand eines repräsentativen Sets von Unterschieden zwischen den Genomdaten (WGS) von der *pdil5-1*-Mutante vs. der Donorlinie 'Golden Promise' DH1-6 mittels Sanger-Sequenzierung entsprechender PCR-Amplifikate von potenziell segregierenden Nachkommen. Dieses Experiment wurde als wichtig erachtet, da zum Zeitpunkt der Bewilligung der Verlängerungsphase keine Hi-C Sequenzierungsdaten vorlagen und die Anzahl der zunächst gefundenen SNPs als recht hoch eingestuft wurde. Zu diesem Zweck hat die AG Mascher (DG, IPK) ein Set von homozygoten und heterozygoten SNVs und Sequenzierfehlern aus den initialen Genomdaten herausgesucht. Dabei wurde darauf geachtet, dass die SNVs quer über das gesamte Genom verteilt waren. Es sollten jeweils 5 homozygote SNVs und 5 potenzielle Sequenzierfehler überprüft werden. Bei den heterozygoten SNVs wurde noch zusätzlich zwischen den Anteilen an *Reads* unterschieden, die entweder das Wildtyp- oder das Mutanten-Allel belegen (Verhältnis von 1:1, 1:4 oder 1:8). Hier sollten auch jeweils 5 SNVs überprüft werden. Insgesamt sollten also 25 PCR-Systeme mit jeweils 4 Geschwisterpflanzen (M_3) der verwendeten *pdil5-1*-Mutante und der Donorlinie 'Golden Promise' DH1-6 etabliert werden und 125 Proben sequenziert werden. Jedoch sind insbesondere bei den PCRs und Sequenzierungen der potenziellen Sequenzierfehler Probleme aufgetreten, sodass zwar für diese Kategorie sogar 10 PCR-Systeme etabliert wurden, aber in der Sanger-Sequenzierung nur für 3 Zielbereiche valide Sequenzen generiert werden konnten. Auch bei den heterozygoten SNVs konnten nicht für alle 3 Kategorien 5 SNVs validiert werden. Für die heterozygoten SNVs mit einem Verhältnis von 1:4 Anteilen an *Reads* konnten 4 SNVs validiert werden, aber für die mit einem Verhältnis von 1:8 Anteilen an *Reads* waren nur 2 SNVs auswertbar. Dies wird als Bestätigung dafür gesehen, dass insbesondere jene Bereiche schwer zu sequenzieren waren in denen in der WGS Sequenzierfehler aufgetreten waren. Die Ergebnisse der SNV-Analyse sind in Tabelle 22 zusammengefasst. Angezeigt werden die Ergebnisse der WGS mit den Anteilen an Sequenzier-*Reads*, die entweder das WT- oder das MT-Allel belegen und die Ergebnisse der Sanger-Sequenzierung der Donorlinie 'Golden Promise' DH1-6 und der 4 Geschwisterpflanzen der *pdil5-1*-Mutante 16-12.

Insgesamt konnte nur ein SNV (ID contig_213:3211765, 3. in Tabelle 22), der in der initialen WGS gefunden wurde in diesem Experiment bestätigt werden. Für die Donorlinie konnte sowohl in der WGS als auch in der Sanger-Sequenzierung ein Guanin an der vorgegebenen Position detektiert werden, während die WGS-sequenzierte *pdil5-1*-Mutante und ihre 4 Geschwisterpflanzen ein Adenin an dieser Position aufwiesen. Dieser SNV kann deshalb mit ziemlicher Sicherheit als eine „echte“ Veränderung angesehen werden. Kein weiterer homozygoter SNV konnte bestätigt werden, was erklärt, weswegen zunächst mehr SNVs als erwartet in der WGS detektiert wurden. Auch die Ergebnisse der Analysen der heterozygoten SNVs und der potenziellen Sequenzierfehler ergaben ein einheitliches Bild zwischen der Donorlinie und den Geschwisterpflanzen der *pdil5-1*-Mutante. Bei 3 SNVs (ID contig_171:6922627, contig_8:2666351 und contig_138:10209410) scheint es bei der WGS der Donorlinie zu einem Fehler gekommen zu sein, da in der Sanger-Sequenzierung nur das alternative Allel detektiert

werden konnte. Zusammenfassend zeigt dieses Experiment, dass - wie erwartet - bei der WGS Sequenzierfehler entstehen und diese durch die angewendeten Filterschritte nur bedingt eliminiert werden können. Außerdem zeigt es, wie wichtig es ist, zusätzlich zu der WGS eine Hi-C Sequenzierung durchzuführen, um verlässliche Daten zu generieren. Aus diesem Experiment geht hervor, dass die Anzahl der in der initialen WGS gefundenen SNVs (sowohl homozygote als auch heterozygote) nach unten korrigiert werden muss. Dies bestätigte sich nach der Assemblierung zu Pseudomolekülen, was durch die Hi-C Sequenzierung ermöglicht wurde. Wie in 3.1.2. beschrieben, wurden damit deutlich weniger SNVs gefunden, wobei deren Häufigkeit nun auch im Rahmen bisheriger Annahmen liegt.

Tabelle 22: SNV-Analyse; DP= Sequenzier-Reads, WT = Donorlinie 'Golden Promise' DH1-6, MT= *pdil5-1*-Mutante 16-12, MT1-4 = Geschwisterpflanzen 1-4 der *pdil5-1*-Mutante 16-12, - = Sequenzierung fehlgeschlagen/ keine Auswertung möglich

SNV ID	WGS			Sanger-Sequenzierung						
	WT Allel	MT Allel	DP	DP WT	DP MT	WT	MT 1	MT 2	MT 3	MT 4
<i>Homozygote SNVs</i>										
contig_20:20323203	C	T	25	0	25	C	C	C	C	C
contig_249:20982977	C	G	21	0	21	C	C	C	C	C
contig_213:3211765	G	A	29	0	29	G	A	A	A	A
contig_354:49243	G	A	19	0	19	G	G	G	G	G
contig_38:9278940	G	C	15	0	15	G	G	G	G	-
<i>Heterozygote SNVs (Verhältnis 1:1)</i>										
contig_37:19813867	G	T	17	9	8	G	G	G	G	G
contig_171:6922627	T	G	13	7	6	G	G	G	G	G
contig_183:12123205	G	A	26	14	12	G	G	G	G	G
contig_61:16566771	G	A	27	14	13	G	G	G	G	G
contig_343:1891353	G	T	27	13	14	G	G	G	G	G
<i>Heterozygote SNVs (Verhältnis 1:4)</i>										
contig_176:922374	G	A	17	4	13	G	G	G	G	G
contig_24:8710835	C	T	12	3	9	C	C	C	C	C
contig_8:2666351	G	A	16	4	12	A	A	A	A	A
contig_326:12129791	G	A	30	22	8	G	G	G	G	G
<i>Heterozygote SNVs (Verhältnis 1:8)</i>										
contig_69:1220665	A	C	30	26	4	A	A	A	A	A
contig_138:10209410	A	C	24	21	3	C	C	C	C	C
<i>Potenzielle Sequenzierfehler</i>										
contig_149:6715167	C	T	2	0	2	C	C	C	C	C
contig_269:8545511	G	C	3	0	3	G	G	G	-	G
contig_203:22930580	C	T	6	0	6	C	C	C	C	C

3.2. Voraussichtlicher Nutzen und Verwertbarkeit der Ergebnisse

Laut Zuwendungsbescheid vom 10.12.2020 ist ein Verwertungsplan nicht erforderlich, da keine Pflicht zur Verwertung besteht. Dennoch sind die in diesem Verbundprojekt generierten Ergebnisse von großem wissenschaftlichem und politischem Interesse. Insbesondere die limitierte Machbarkeit des Nachweises von Genom-edierten Pflanzen trotz Kenntnis der Zielsequenz-spezifischen Veränderung ist für Akteure aus der Politik sowie für die für die Rückverfolgbarkeit zuständigen Behörden relevant. Zudem wird anhand der Projektergebnisse ersichtlich, dass eine eindeutige Identifikation von Genom-edierten Pflanzen nur in sehr bestimmten Fällen (zufällige, gekoppelte Mutationen vorhanden) möglich sein wird. Darüber hinaus sind insbesondere die Ergebnisse der Sequenzierung auch aus wissenschaftlicher Sicht von großem Nutzen, und es ist zu erwarten, dass aus diesem Projekt einige Publikationen hervorgehen werden.

4. Zusammenfassung

Gemäß der Vorhabensbeschreibung wurden alle relevanten Ziele erreicht. Das WGS der Donor-Linie 'Golden Promise' DH1-6 und der *pdil5-1*-Mutante mit einer Zielsequenz-spezifischen Veränderung bestätigte, dass die Mutante die erwartete 1 bp-Deletion im *PDIL5-1*-Gen aufweist. Durch die erfolgte Hi-C Sequenzierung war es möglich, die durch die WGS generierten Genom-ähnlichen *Contigs* zu Pseudomolekülen zu assemblieren, wodurch valide Aussagen über Sequenzunterschiede zwischen der Donor-Linie und der Mutante getroffen werden konnten. Die Mutante wies insgesamt 48 homozygote SNPs und 358 homozygote Indels im Vergleich mit der Donor-Linie auf. Diese Größenordnung liegt im Bereich der Erwartungen bezüglich pro Generation spontan auftretender Mutationen. Diese geringe Anzahl an aufgetretenen Sequenzunterschieden wurde auch durch eine zusätzlich durchgeführte SNP/SNV-Analyse mittels Sanger-Sequenzierung belegt. Zudem war es durch die Hi-C-Sequenzierung möglich, chromosomale Neuordnungen und große strukturelle Veränderungen auszuschließen. Alle im Gerstengenom identifizierten *Off-Targets*, also dem adressierten Zielmotiv ähnliche Sequenzen erwiesen sich als frei von genetischen Veränderungen. Zudem wurde gezeigt, dass keinerlei Rückstände eines Transgens in der *pdil5-1*-Mutante vorhanden sind.

Ebenso wie für Gerste erfolgte eine WGS der beiden Rapsmutanten C3E3 und C3E4. Zusätzlich wurde eine Hi-C Sequenzierung der Raps-Donor-Linie 'Mozart' und der Mutante C3E4 durchgeführt. Dies ermöglichte die Assemblierung der vom BVL zur Verfügung gestellten Genomdaten der Donor-Linie 'Mozart' und der Mutante zu Pseudomolekülen. Im Vergleich zu Gerste erwies sich die Assemblierung bei Raps aufgrund seiner Polyploidie als schwieriger. Dennoch können keine chromosomalen Neuordnungen oder strukturellen Veränderungen auf die gezielte Mutagenese mittels Cas9-System zurückgeführt werden.

Es konnte kein spezifisches qPCR-Assay für Gerste entwickelt werden, da mit allen 5 getesteten Systemen Kreuzreaktionen der Sonden auftraten und keine eindeutige Differenzierung zwischen dem Wildtyp 'Golden Promise' DH1-6 und der *pdil5-1*-Mutante möglich war. Dagegen konnte ein ddPCR-basiertes Detektionsverfahren zum Nachweis 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 vom Projektpartner der CAU Kiel bereitgestellten Rapsmutante C3E3 entwickelt. Mit beiden Assays können auf verlässliche Weise sowohl 0,9 % als auch 0,1 % Mengenanteile der jeweiligen Mutante in Mischproben nachgewiesen werden. Die Übertragbarkeit dieser Assays wurde durch erfolgreiche Anwendung sowohl beim Auftragnehmer Impetus Bioscience als auch am IPK Gatersleben unter Verwendung von Geräten unterschiedlicher Hersteller belegt. 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 vom akkreditierten Referenzlabor Impetus Bioscience erfolgreich auf Spezifität, Selektivität und Anwendbarkeit getestet.

Darüber hinaus wurde ein Detektionsverfahren etabliert, das auf Amplikon-Tiefensequenzierung beruht. Auch mit dieser Methode wurde gezeigt, dass 0,9 % und 0,1 % Mengenanteile der Mutante in Mischproben zuverlässig nachgewiesen werden können. Die entwickelte Methode wurden vom akkreditierten Referenzlabor Planton durchgeführt und ebenfalls erfolgreich auf Spezifität, Selektivität und Anwendbarkeit getestet.

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

4.1. Zusammenfassende Bewertung

Die WGS und Hi-C-Sequenzierungen haben gezeigt, dass mit den neuen Züchtungstechniken ganz gezielt nur Mutationen an vorgegebener Stelle ausgelöst werden können. Die hier untersuchte Gerstenmutante hatte in allen potenziellen *Off-Targets* keinerlei Sequenzveränderungen. Zudem wurde für diesen Fall gezeigt, dass durch die Genom-Editierung mittels Cas-Endonukleasen keine Rückstände oder sogenannte „Footprints“ im Genom hinterlassen werden, die als Identifikationsmerkmal für Genom-editierte Pflanzen herangezogen werden können. Aufgrund der in diesem Projekt durchgeführten Experimente können wir feststellen, dass ddPCR-basierte Nachweismethoden für die Detektion sowohl der in diesem Projekt verwendeten Gersten- als auch der Rapsmutante geeignet sind. Es war jedoch nicht möglich ein zuverlässiges qPCR-Assay in Gerste zu etablieren. Daraus folgt die Einschätzung, dass es nicht generell möglich sein wird, für jede Genom-editierte Pflanze ein zuverlässiges PCR-basiertes Detektionsassay zu etablieren. Dies wird immer von Fall zu Fall beurteilt werden müssen. Zusätzlich zu dem ddPCR-Assay war es uns möglich, ein valides Amplikon-Tiefensequenzierungs-basiertes Detektionsassay zu etablieren. Diese Methode hat den ganz klaren Vorteil einen viel tieferen und genaueren Einblick geben zu können, als es mit ausschließlich PCR-basierten Methoden möglich ist. Zudem kann ein einmal etabliertes Assay für die Detektion von verschiedenen Mutationen/Sequenzveränderungen im gleichen Sequenzbereich eingesetzt werden. Bei der Verwendung von ausschließlich PCR-basierten Methoden ist es hingegen erforderlich, für jede Sequenzveränderung - selbst im gleichen Sequenzabschnitt - ein neues Assay zu entwickeln. Aus diesen Gründen werden Sequenzierungs-basierte Methoden in der GVO-Analytik in Zukunft vermutlich bevorzugt eingesetzt werden. Abschließend lässt sich konstatieren, dass es zwar möglich ist PCR-basierte Methoden für die Detektion von Genom-editierten Pflanzen zu entwickeln, es aber nicht generell möglich sein wird, Genom-editierte Pflanzen zu identifizieren, weil es keine Unterscheidungsmöglichkeit zu einer gleichartigen, spontan entstandenen Mutation gibt. Eine Identifikation kann ausschließlich dann gelingen, wenn in unmittelbarer Nähe der

zielsequenzspezifisch-induzierten Mutation simultan auch eine spontane Sequenzänderung im Genom aufgetreten ist, die für ein Identifikationsassay herangezogen werden kann.

5. Gegenüberstellung der ursprünglich geplanten zu den tatsächlich erreichten Zielen

Objektiv bedingt durch den Wechsel der für die Genomsequenzierung angewendeten Technologie (von ursprünglich geplanten Illumina *Short Reads* zu zeitgemäßerer PacBio *Long Reads*) während der Projektlaufzeit war es zwischenzeitlich zu erheblichen Verzögerungen gekommen. Die Ergebnisse der Genomsequenzierung konnten erst in der zweiten Projekthälfte ausgewertet werden und dadurch verzögerten sich viele der weiteren, auf diesen Ergebnissen aufbauenden Experimente. Die vorgesehenen Arbeitspakete der regulären Laufzeit (AP 1.1. bis 2.4) und der Projektverlängerung (V1 bis V4) sind in Tabelle 23 zusammengestellt. Trotz der Verzögerungen konnten die für die umfassende Validierung des ddPCR-Assays in Gerste erforderlichen Experimente innerhalb der regulären Projektlaufzeit abgeschlossen werden. Dabei wurden die Ergebnisse der Arbeitspakete 2.1, 2.2 und 2.4 in den jeweiligen SOP-Entwürfen dokumentiert. Die Etablierung eines entsprechenden Assays für Raps war während der regulären Projektlaufzeit nicht möglich, wurde jedoch während der bewilligten Projektverlängerung ebenfalls erfolgreich abgeschlossen (Arbeitspaket V1). Trotz umfangreicher Experimente erwies sich die Entwicklung eines spezifischen qPCR-basierten Nachweis-Assays für Gerste (Arbeitspaket 1.2) als nicht möglich.

In der Projektverlängerung wurden die *Chromosome Conformation Capture* (Hi-C) Sequenzierungen von Gerste und Raps durchgeführt. Die bioinformatische Auswertung dieser Sequenzierung erfolgte durch den IPK-internen Projektpartner der AG Mascher (DG). Dabei wurden die vorliegenden Genomdaten zu chromosomalen Pseudomolekülen assembliert. Die Hi-C-Sequenzierung in Verbindung mit der *Long-Read*-Sequenzierung ist unerlässlich, um Assemblierungen im Chromosomenmaßstab (Pseudomoleküle) zu erzeugen und zuverlässig auf strukturelle Variationen zu prüfen, die durch das Mutationsereignis verursacht werden könnten (Pucker et al., 2021). Wie in unseren Ergebnissen zu sehen ist, erhielten wir weniger falsch-positive Varianten nach Assemblierung von Pseudomolekülen. Zudem wurde in der Verlängerungsphase ein Set von spontan entstandenen Hintergrundmutationen validiert, die nach der WGS identifiziert wurden (Arbeitspaket V3). Hierbei zeigte sich, dass von den letztendlich untersuchten 19 SNVs nur ein homozygoter SNV mittels Sanger-Sequenzierung bestätigt werden konnte. Diese Ergebnisse bestätigten die Resultate der Hi-C Sequenzierung, nämlich, dass es deutlich weniger Unterschiede zwischen der Mutante und der Donor-Linie gibt als zunächst nach der WGS angenommen und dass durch die Hi-C Sequenzierung eine höhere Aussagekraft der Genomdaten erreicht wurde.

Da kein mit der gezielt induzierten Mutation einhergegangener und dieser ausreichend nahe gelegener spontan entstandener SNP oder InDel in der *pdil5-1* Mutante gefunden wurde, war die Bearbeitung des Arbeitspakets V4 hinfällig.

Tabelle 23: Darstellung der geplanten Arbeiten

Arbeitspaket	Aufgabe	Ziele erreicht
AP 1.1	WGS zur Charakterisierung von Referenzlinien (Gerste & Raps) und bioinformatische Auswertung von Indels	✓
AP 1.2	Entwicklung eines qPCR-Assays unter Verwendung von TaqMan und LNA-Sonden	nicht möglich
AP 1.3	Etablierung der ddPCR unter Verwendung von TaqMan und LNA-Sonden	✓
AP 1.4	Tiefensequenzierung von Amplikons der Zielregion und möglicher <i>Off-Targets</i> der durch WGS charakterisierten Mutanten und Kontrollen	✓
AP 2.1	Vergleichende Validierung der Detektionstechnologien bezüglich Genauigkeit und Sensitivität	✓
AP 2.2	Schaffung von statistisch abgesicherten Verfahrenskennwerten der jeweiligen Methoden und Entwicklung von Analyse-Richtlinien	✓
AP 2.3	Etablierung einer Analysepipeline für die Amplikon-Sequenzierung	✓
AP 2.4	Validierung der überprüften, optimierten Protokolle der qPCR, ddPCR & Amplikon-Sequenzierung	✓
V1	Etablierung eines Nachweisverfahrens für Genom-editierten Raps mittels ddPCR	✓
V2	Chromosome Conformation Capture (Hi-C) von Gerste und Raps	✓
V3	Verifizierung spontan entstandener Hintergrundmutationen in Gerste	✓
V4	Identifizierung von Genom-editierter Gerste mittels genetisch gekoppelter SNVs	nicht möglich

6. 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.,... and 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., Boutte, 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), g1aa137.
- 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.

Anlagen

Anhang mit Primer- und *Off-Target*-Liste (Tabellen A1, A2)

Entwurf einer *Standard Operating Procedure* (SOP) für ein quantitatives ddPCR-System zum Nachweis einer Gerstenmutante

Entwurf einer *Standard Operating Procedure* (SOP) für ein quantitatives ddPCR-System zum Nachweis einer Rapsmutante

Kurzfassung des DETECT-Abschlussberichtes

Die Rohdaten wurden der BLE zur Verfügung gestellt.

30.11.2023



Datum und Unterschrift des Projektverantwortlichen

Anhang

Tabelle A1: Verwendete Oligonukleotide

Name	Sequenz (5'→3')	Beschreibung
272_PDIL_qRT_fw	GACCGGCCAGAGAAATACATC	Primer-kompetitives Assay
272_PDIL_qRT_rv	GAGACGAGGAGAAGGCGGATAC	Primer-kompetitives Assay
273_PDIL_Probe_SNV	FAM/GACCCACCCGACT/MGB/BMN-Q535	SNV-Sonde mit MGB
274_PDIL_Probe_WT	HEX/ACCCACCCCGACT/MGB/BMN-Q535	Referenz-Sonde mit MGB
273_WT-Gerste-NEU	HEX/GACCCACCCCGACTGGC / BMN-Q535	Normale Referenz-Sonde
274_MUT-Gerste-NEU	FAM/CGACCCACCCGACTGGC/ BMN-Q535	Normale SNV-Sonde
IK435_HvPDIL5-1_fw	GTGGCCCATGTCCAGTTGTA	Primer Drop-off Assay
IK436_HvPDIL5-1_rv	AGACGAGGAGAAGGCGGATA	Primer Drop-off Assay
IK437_HvPDIL_Reference_Probe	HEX/GGACAGATGGCGACGGTGGAAATCTT/BMN-Q535, 2x	Normale Referenz-Sonde
IK438_HvPDIL_SNV_Probe	CY5/CCAGTC+G+GGTGGGTCGAT/BMN-Q620	SNV-Sonde mit LNAs
IK441_rev_LNA	CGCCAGT+C+GG+G+T	Rev LNA Primer
IK442_rev_LNA2	ACGCCAGT+CGGG+T	Rev LNA Primer
IK424_HvHor3_fw	AGACAAGGCGTGCAGATCG	Primer Hor3 Referenz Assay
IK425_HvHor3_rev	GACCTGGACGAGCACACAT	Primer Hor3 Referenz Assay
IK426_HvHor3_Probe	HEX/CCTCAGCCGCAACAGGTGGGTC/BHQ-1	Sonde Hor3 Referenz Assay
CRT1 A09_fw	AGGAAGGTGCGATCTGAGTTT+C	Primer Raps ddPCR Assay
CRT1 A09_rev	CGCGACGAGGCCGA+T	Primer Raps ddPCR Assay
CRT1 A09_WT-Sonde	HEX/ CAATTCATCTCTC /MGB/BMN-Q535	WT-Sonde Raps ddPCR Assay
CRT1 A09_MT-Sonde	FAM/ CAATTTTCATCTCTC /MGB/BMN-Q535	MT-Sonde Raps ddPCR Assay
IK447_SNP_hom2_fw	GAGTGACAATAGCCGGAACC	SNV-Analyse
IK448_SNP_hom2_rv	CTAGCTCCAGTTCGTCATG	SNV-Analyse
IK451_SNP_hom1_fw	CTAGCCAGCCGGCTAGGAAC	SNV-Analyse
IK452_SNP_hom1_rv	CCACGCCAAGTGCAGTAATG	SNV-Analyse
IK453_SNP_hom2_fw	CCATGACCCGTCATGAAC	SNV-Analyse
IK454_SNP_hom2_rv	GCATGCTGCAAAATGGATGT	SNV-Analyse
IK455_SNP_hom3_fw	GGAGAAAGCGAGTACTTAGG	SNV-Analyse
IK456_SNP_hom3_rv	GGGACTTAGCTGCGATCAAG	SNV-Analyse
IK457_SNP_hom8_fw	CTAGTCCCCTCTGGTGCTT	SNV-Analyse
IK458_SNP_hom8_rv	GGCTGTTATTTCACTTCACG	SNV-Analyse
IK459_SNP_het1_fw	CTTGTTACGGAAAGGACTC	SNV-Analyse
IK460_SNP_het1_rv	GCACTTGCTTGAATTCATC	SNV-Analyse
IK461_SNP_het5_fw	GAGAAAGGACCCGTTTCAAG	SNV-Analyse
IK462_SNP_het5_rev	ATAGGGCACCACGTAATTCC	SNV-Analyse
IK463_SNP_het6_fw	ACAAATCCACAATTCATCGG	SNV-Analyse

IK464_SNP_het6_rv	TGGAGGGCTTGTACATCAAC	SNV-Analyse
IK465_SNP_het7_fw	GCAAAGTGACATGGATCTG	SNV-Analyse
IK466_SNP_het7_rv	TCCACACATGTAAATGAGTC	SNV-Analyse
IK467_SNP_het8_fw	AGCGGATGATTGCGTATGTC	SNV-Analyse
IK468_SNP_het8_rv	CGAAAGTGCAACTGGCAAAG	SNV-Analyse
IK469_SNP_het11_fw	CCTTCTCAAGCTCCCAATC	SNV-Analyse
IK470_SNP_het11_rv	GTACATCGTGATCGGAATTA	SNV-Analyse
IK471_SNP_het12_fw	TGACCATGTCGCCATGAAAC	SNV-Analyse
IK472_SNP_het12_rv	CCTGCCTGAAGGAGCAGAC	SNV-Analyse
IK473_SNP_het13_fw	GTCCCGTTCCAACATCATCC	SNV-Analyse
IK474_SNP_het13_rv	TTCTTCGGGCATGTTAGAG	SNV-Analyse
IK475_SNPerr1_fw	TTTGCCCGACAGTGTCCCTC	SNV-Analyse
IK476_SNPerr1_rv	CCGGCTCCTCTCTATGAGT	SNV-Analyse
IK477_SNPerr2_fw	CGAAACATATGCATGACATT	SNV-Analyse
IK478_SNPerr2_rv	CACACACTCATATGTCCC	SNV-Analyse
IK479_SNPerr3_fw	CTCCTTGCCTTAATTTCCC	SNV-Analyse
IK480_SNPerr3_rv	CCCTCCTTCTCTCCTTCTC	SNV-Analyse
IK481_SNPerr4_fw	GATCTTGCGGAGACAAAAGC	SNV-Analyse
IK482_SNPerr4_rv	GAACGGAGTCCAAACGGAAT	SNV-Analyse
IK483_SNPerr5_fw	TGAGGGTTACCTCATGAATC	SNV-Analyse
IK484_SNPerr5_rv	CTACAGGAATGGCCCTCCTC	SNV-Analyse
IK489_SNP_Hom7_fw	CTACGTATCACACTTGGAGC	SNV-Analyse
IK490_SNP_Hom7_rev	GCGCTAAGTATGGACTTGAC	SNV-Analyse
IK491_SNP_Het2_fw	CTTCTCCTCTTGTCCCATT	SNV-Analyse
IK492_SNP_Het2_rv	GCATAGTTTGCTCCGGAATG	SNV-Analyse
IK493_SNP_Het17_fw	GATGAGTTCTCCAGCACGAC	SNV-Analyse
IK494_SNP_Het17_rv	AGGGTTTCCAAGTGGAGAAG	SNV-Analyse
IK495_SNP_Het18_fw	GAAGGAAGGCACTGCAGCTC	SNV-Analyse
IK496_SNP_Het18_rv	CCATGGTACGGGCTCGTTG	SNV-Analyse
IK497_SNP_Het25_fw	GGTCCGGTTATGCGGAACTC	SNV-Analyse

IK498_SNP_Het25_rv	GAGTTCCACGTAACCGGACC	SNV-Analyse
IK499_SNP_Het26_fw	GATCTCGCGAAACGGAAGC	SNV-Analyse
IK500_SNP_Het26_rv	TTGGAATGGAGTCCAAACGG	SNV-Analyse
IK501_SNP_Het28_fw	CCAAACGTGTGATCATATGC	SNV-Analyse
IK502_SNP_Het28_rv	GGAGCTCGAAAAAGGAACC	SNV-Analyse
IK503_SNP_Het29_fw	TCTCTAACTGGGCATCCTG	SNV-Analyse
IK504_SNP_Het29_rv	GATGGGTAAAGCGGGTCTTC	SNV-Analyse
IK505_SNP_Het30_fw	ATGGCTCTTGGGCCAAATTC	SNV-Analyse
IK506_SNP_Het30_rv	GGCGGCTTACGATGACAAGA	SNV-Analyse
IK507_SNPerr6_fw	AATTAGCCCATTTAGCTCCA	SNV-Analyse
IK508_SNPerr6_rv	AGAAGGTCCTTGGCCTTCTC	SNV-Analyse
IK509_SNPerr7_fw	TCATCCTCGCGGACCATCAC	SNV-Analyse
IK510_SNPerr7_rv	CTCGCCTGCTAAGCTGTGAC	SNV-Analyse
IK511_SNPerr9_fw	CTTTGTGATCTAATGCAAGC	SNV-Analyse
IK512_SNPerr9_rv	GTGAGACACCACTGGAGACG	SNV-Analyse
IK513_SNPerr10_fw	ATCGCCTGGCGTGCTAGATG	SNV-Analyse
IK514_SNPerr10_rv	TGGCTTCGTAGTGTGCTGTC	SNV-Analyse

Tabelle A2: Potentielle Off-Targets in Gerste

Chromosome	start	end	strand	MM	target_seq	PAM
chr4H	455017471	455017493	+	0	CATCATCGACCCACCCCGAC	TGG
chr2H	303234283	303234305	+	2	GATCATCGACTCACCCCGAC	AGG
chr3H	555516541	555516563	-	4	GTGCATCGTCCCACCCCGAC	GGG
chr2H	396059136	396059158	+	4	CCACCTCGCCCCACCCCGAC	CGG
chr7H	504603382	504603404	-	4	CCACCTCGCCCCACCCCGAC	CGG
chr5H	318353770	318353792	-	4	CCACCTCGCCCCACCCCGAC	CGG
chr5H	104123778	104123800	-	4	CCACCTCGCCCCACCCCGAC	CGG
chr4H	379124202	379124224	+	4	CACCAAGGATCCACCCCGAC	AGG
chr5H	116492561	116492583	-	4	CACCAAGGATCCACCCCGAC	AGG
chr2H	314095209	314095231	-	4	CACCAAGGATCCACCCCGAC	AGG
chr2H	575165420	575165442	+	4	CCTCGTCAACGCACCCCGAC	CGG
chr4H	367821445	367821467	-	4	CCTCGTCAACGCACCCCGAC	CGG
chr2H	387818512	387818534	+	4	CATGATAGCCTCACCCCGAC	CGG
chr4H	327959727	327959749	+	4	AATGATCTACCCTCCCGAC	AGG
chr1H	344996169	344996191	-	4	CATGAGCGAAGCACCCCGAC	CGG
chr3H	522280334	522280356	+	3	CATGATCGATCCCCCGAC	AGG
chr5H	275293168	275293190	-	3	CATGATCGATCCCCCGAC	AGG
chr5H	257194866	257194888	+	4	AATCATCATCTACCCCGAC	AGG
chr7H	78832550	78832572	-	4	CAAATCGCCCCAGCCCGAC	GGG
chr4H	312631865	312631887	+	4	CATGATTGATCCCCCGAC	AGG
chr6H	519426825	519426847	-	4	CATCGGCGACGCCCGAC	CGG
chr7H	602812813	602812835	-	4	CATCGGCGACGCCCGAC	CGG
chr5H	244310283	244310305	-	4	CCGCCTGACCCACCGCGAC	CGG
chr3H	288026433	288026455	-	4	TAGCATCGACCTAGCCCGAC	TGG
chr1H	348842475	348842497	+	4	CACCATAGATCCACACCGAC	GGG
chr2H	363233449	363233471	-	3	CGTCATCGACCCCTCCGAC	AGG
chr3H	451944222	451944244	-	3	CGTCATCGACCCCTCCGAC	AGG
chr2H	82034004	82034026	+	3	CATGATCGACCCCTCCGAC	AGG

chr4H	501901405	501901427	+	3	CATGATCGACCCCCTCCGAC	AGG
chr3H	340563025	340563047	+	3	CATGATCGACCCCCTCCGAC	AGG
chr7H	343430688	343430710	+	3	CATGATCGACCCCCTCCGAC	AGG
chr2H	65935061	65935083	+	3	CATGATCGACCCCCTCCGAC	AGG
chr1H	355222871	355222893	+	3	CATGATCGACCCCCTCCGAC	GGG
chr7H	509200190	509200212	-	3	CATGATCGACCCGCTCCGAC	AGG
chr5H	507285907	507285929	-	3	CATGATCGACCCCCTCCGAC	AGG
chr5H	139831616	139831638	-	3	CATGATCGACCCCCTCCGAC	AGG
chr1H	438395447	438395469	-	3	CATGATCGACCCCCTCCGAC	AGG
chr1H	363543708	363543730	-	4	TATGATCGACCCCCTCCGAC	GGG
chr7H	262060885	262060907	-	4	TATGATCGACCCCCTCCGAC	AGG
chr5H	271383797	271383819	-	4	TATGATCGACCCCCTCCGAC	AGG
chr2H	211779303	211779325	-	4	TATAATCGACCCCCTCCGAC	AGG
chr3H	281477718	281477740	+	4	CGTGATCGACCCCCTCCGAC	AGG
chr5H	532112853	532112875	+	4	CGTGATCGACCCCCTCCGAC	AGG
chr5H	275698134	275698156	+	4	CGTGATCGACCCCCTCCGAC	AGG
chr6H	275581252	275581274	+	4	CTTGATCGACCCCCTCCGAC	AGG
chr2H	499845815	499845837	-	4	CGTGATCGACCCCCTCCGAC	AGG
chr2H	522253735	522253757	-	4	CGTGATCGACCCCCTCCGAC	AGG
chr5H	345743172	345743194	-	4	CCTGATCGACCCCCTCCGAC	AGG
chr2H	643266014	643266036	+	4	CATGACCGACCCCCTCCGAC	AGG
chr1H	257096319	257096341	+	4	CATGACCGACCCCCTCCGAC	AGG
chr3H	441807954	441807976	+	4	CATGATTGACCCCCTCCGAC	AGG
chr4H	85990801	85990823	+	4	CATGATTGACCCCCTCCGAC	AGG
chr7H	379213276	379213298	+	4	CATGATTGACCCCCTCCGAC	AGG
chr2H	94589049	94589071	+	4	CATAATAGACCCCCTCCGAC	AGG
chr2H	214974839	214974861	-	4	CATGATTGACCCCCTCCGAC	AGG
chr3H	70636718	70636740	-	4	CATGATTGACCCCCTCCGAC	AGG
chr2H	540112902	540112924	-	2	CATCATCGTCCCACCCCAC	GGG
chr3H	335242726	335242748	-	4	CATCAGTGACCCCCTCCGAC	AGG

chr6H	371057304	371057326	+	4	CGTGATCGACCCCTCGAC	AGG
chr1H	60812766	60812788	-	4	CCTCGTCGACCTCCACGAC	GGG
chr3H	460790054	460790076	-	4	CACCATAGACCCATACCGAC	GGG
chr4H	88370215	88370237	+	4	TAGCTTCGACCCACCCGCC	TGG
chr1H	70213833	70213855	-	4	AAGCAGCGACCCACCCGAA	AGG
chr3H	246579070	246579092	+	4	CCTCTTCGACCAACCCGTC	AGG
chr2H	133537565	133537587	-	4	CAACACCCACCCACCCGAG	GGG
chr7H	158532044	158532066	+	4	G TTCATCGACCCACCCACAC	CGG
chr1H	379402276	379402298	-	3	CATCATCCACCTACCCGAA	GGG
chr1H	57707268	57707290	+	4	GATCGTCGACCCAACCCGAA	AGG
chr3H	263600869	263600891	+	4	AATCATCAACCCACCCAGCC	TGG
chr2H	226094647	226094669	+	4	CTTCATGGACCCACCCCCC	AGG
chr7H	564939314	564939336	-	4	CTTCATGGACCCACCCCCC	AGG
chr4H	26066497	26066519	+	4	CATCAGCCACCCACCCAGAG	AGG
chr2H	635355386	635355408	+	4	CATCTTAGACCCACCCAAA	CGG
chr6H	172168291	172168313	+	4	CATCAGCCACCCACCCGGA	GGG
