

Abschlussbericht

Zuwendungsempfänger:

Prof. Dr. Daguang Cai (CAU Kiel)
Dr. habil. Dirk Schenke (CAU Kiel)

Förderkennzeichen: 2820HS012



Bundesministerium
für Ernährung
und Landwirtschaft



Thema des Forschungsvorhabens:

Implementation der Next Generation Sequencing (NGS) Technologie zum gezielten Nachweis genomeditierter Rapspflanzen und derer Produkte (RapsNMT)

Laufzeit des Vorhabens:

01.02.2021– 31.07.2023

Berichtszeitraum:

01.02.2021 – 31.07.2023

Zusammenarbeit mit anderen Stellen

Dr. Jochen Kumlehn
Dr. Iris Hoffie
Dr. Martin Mascher
Dr. Marina Pupke Marone



Unterauftragnehmer:

Prof. Dr. Michael Kleine
Dr. Martin Weigel



Dr. Olav Grundmann



Dr. Lutz Grohmann
Dr. Steffen Pallarz



Inhaltsverzeichnis

1. Ziele und Aufgabenstellung des Vorhabens.....	3
1.1 Planung und Ablauf des Vorhabens.....	4
1.2 Wissenschaftlicher und technischer Stand, an den angeknüpft wurde.....	4
2. Material und Methoden.....	5
3. Ergebnisse.....	10
3.1 Ausführliche Darstellung der wichtigsten Ergebnisse.....	10
3.1.1. Erste Ergebnisse aus dem long-read PacBio WGS.....	10
3.1.2. Ergebnisse aus der short-read Illumina Amplikon-Tiefensequenzierung	16
3.2 Voraussichtlicher Nutzen und Verwertbarkeit der Ergebnisse.....	23
4. Zusammenfassung.....	24
5. Gegenüberstellung der ursprünglich geplanten zu den tatsächlich erreichten Zielen.....	25
5.1 Vergleich der Arbeitspakete mit dem Bearbeitungsstand zum Projektende.....	25
5.2 Hinweise auf weiterführende Fragestellungen.....	26
6. Literaturverzeichnis.....	27

Abkürzungsverzeichnis

CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
Cas	CRISPR-associated (die "Genschere")
CAU	Christian-Albrechts Universität (AG Cai)
DNA	Deoxyribonucleic acid (DNS, Desoxyribonukleinsäure)
DSB	DNA Doppelstrangbruch
ENGL	European Network of GMO Laboratories
GE	Genom-Editierung
GVO	Gentechnisch veränderter Organismus (aka GMO, LMO)
HDR	Homology Directed Repair (Reparatur der DSB mittels Reparaturvorlage mit dem Edit)
InDels	Insertionen/Deletionen
IPK	Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung (AG Kümlehn/Mascher)
LOD	Limit Of Detection (Nachweisgrenze)
NGS	Next Generation Sequencing Techniken (z. B. Illumina short-read, PacBio long-read)
NHEJ	Non-Homologous End Joining (Reparaturmechanismus für DSB)
NGT	Neue Gentechnik
NMT	Neue Mutagenese Techniken (aka Neue Züchtungstechnologien)
PAM	Protospacer Adjacent Motif (bei Cas9: NGG)
PCR	Polymerase Kettenreaktion
RNA	Ribonucleic acid (RNS, Ribonukleinsäure)
RNP	Ribonucleoprotein-Komplexe (bestehend aus Genschere plus sgRNA)
SDN	Site-Directed Nuclease (wirken ortsspezifisch an definierter Stelle, z. B. Cas9, TALEN, ...)
SDN-1	Einfachste Methode zur Zufallsmutagenese an der Stelle des DSB durch NHEJ
SDN-2	Gezielte Gen-Korrektur/Modifikation an der Stelle des DSB durch HDR
(s)gRNA	(single) guide RNA (dirigiert die Genschere an eine genau definierte Stelle: „site-directed“)
SNP	Einzelnukleotid-Polymorphismen/Veränderungen/Variationen (aka SNV)
SOP	Standard Operating Procedure
TALEN	Transcription Activator-Like Effector Nuclease
WGS	Whole Genome Sequencing (Komplett-Sequenzierung eines Genoms)
WT	Wildtyp (eine Pflanze mit dem ursprünglichen genetischen Hintergrund, hier Mozart)

1. Ziele und Aufgabenstellung des Vorhabens

Nach der derzeitigen europäischen Gesetzeslage (EuGH Urteil in der Rechtssache C-528/16 vom 25.07.2018) müssen Genveränderungen in der Pflanzenzüchtung, welche durch die Neuen Mutagenese Techniken (NMT) erzeugt wurden, wie GVOs reguliert werden. Selbst wenn in der EU solche Pflanzen nicht angebaut würden, so gelten doch in einigen Drittstaaten (z. B. Argentinien, Japan, USA, Israel) Genom-editierte (GE) Organismen nicht als GVO, weshalb eine diesbezügliche Regulierung, Kennzeichnung und Informationsweitergabe nicht erfolgt. Es ist daher zu erwarten, dass GE Pflanzen bzw. daraus hergestellte Produkte in den nächsten Jahren auch in die EU gelangen können. Zum Beispiel wurde in Japan bereits eine Tomate von Sanatech Seed mit erhöhtem GABA-Gehalt zugelassen (Waltz, 2021), sodass sich die Frage nach geeigneten Nachweismethoden für den Nachweis und die Identifizierung genomeditierter Organismen, bzw. daraus hergestellter Produkte stellt. Mit der Chhalliyil-Publikation (2020) wurde eine erste qPCR-basierte Methode zur Detektion von Mutationen im Cibus-Raps beschrieben. Durch das BVL konnte jedoch gezeigt werden, dass die Leistungsparameter nicht den Vorgaben für Methoden in der amtlichen Überwachung entsprechen. Dies zeigt, wie schwierig die Entwicklung einer spezifischen und robusten quantitativen Polymerasekettenreaktion (qPCR)-Methode zum Nachweis von Punktmutationen ist. Die Detektion einer Mutation aus einer Genom-Editierung (GE) beinhaltet:

- ① den einfachen Nachweis (hierbei handelt es sich lediglich um das Auffinden der Zielsequenz),
- ② die Identifikation (d.h. die nachgewiesene Sequenz kann einem spezifischen Event/Prozess zugeschrieben werden) und
- ③ die Quantifizierung des Anteils an Genom-editierter DNA in einer Probe,

wobei die Nachweisgrenze für registrierte Events bei 0,1% liegt und unterhalb derer ein Produkt noch als unbelastet gelten würde (Grohmann et al., 2019, ENGL, 2019). Bis dato gab es noch keine ausreichende Validierung von Next Generation Sequencing-Ansätzen für die Quantifizierung von Zielsequenzen in komplexen Proben und auch keine Möglichkeit den Ursprung von SNPs oder InDels auf Genom-Editierung, klassische Mutagenese-Techniken oder natürliche Entstehung zurückzuführen (ENGL, 2019). Aus diesem Grund erfolgte 2020 durch die Bundesanstalt für Landwirtschaft und Ernährung (BLE) eine Bekanntmachung, um geeignetere Nachweis und Identifizierungsverfahren für genomeditierte Pflanzen und pflanzliche Produkte zu prüfen (BAnz AT 30.04.2020 B5). In ähnlicher Weise gibt es auch ein durch das BfN gefördertes Forschungs- und Entwicklungsvorhaben zur „Weiterentwicklung der molekularen Methoden für die Risikobewertung klassischer und genomeditierter GVO unter Berücksichtigung der technischen Entwicklung von Analytik und Datenbanken“ (FKZ 3520 84 0200). Zwei Vorschläge wurden für das Verbundprojekt „Machbarkeitsstudie zu Nachweis- und Identifizierungsverfahren für genomeditierte Pflanzen und pflanzliche Produkte“ ausgewählt (Abb. 1), von denen sich das DETECT Teilprojekt mit Gerste beschäftigt und das RapsNMT Teilprojekt Gegenstand dieses Berichts ist. Im Rahmen des Zulassungsverfahrens für GVO und daraus hergestellter Produkte soll ein validiertes Nachweis- und Identifizierungsverfahren und entsprechendes Referenzmaterial zur Rückverfolgbarkeit und Marktkontrolle verfügbar gemacht werden. Für die amtliche Überwachung wird eine gerichtsfeste Nachweismethode benötigt, welche die GE-Pflanzen von klassischen Züchtungsprodukten unterscheiden und auch geringe Beimengungen (0,1 %) ausreichend sicher detektieren kann. Eine wichtige Voraussetzung für die Etablierung jeglicher Nachweismethoden wäre allerdings eine vollständige, internationale und transparente Datenbank, in welcher alle notwendigen Informationen zu jeglichen GE Events vorliegen (Ribartis et al., 2021). Im Anbetracht der weltweit unterschiedlichen Bewertung / Regulierung dieser Technologie scheint allerdings ein Wechsel zu einer produktbasierten Regulierung in der EU pragmatischer (Gould et al., 2022).

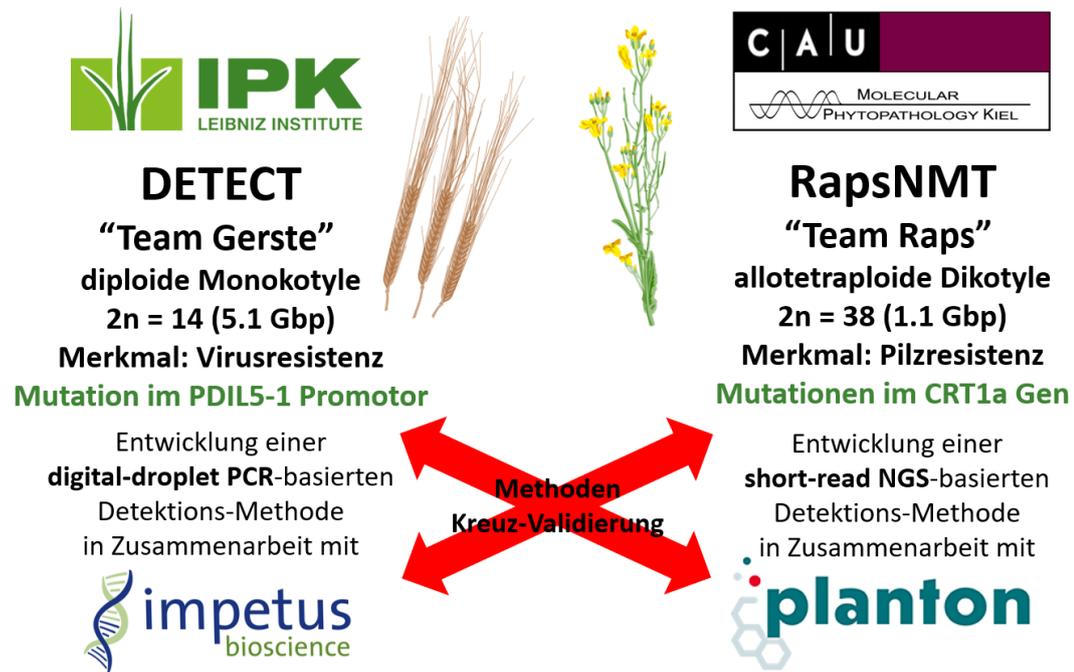


Abb. 1: Übersicht über das Verbundprojekt beruhend auf Material zweier unterschiedlicher Kulturarten mit durch GE induzierter Resistenz gegenüber Pflanzenschädlingen, einer Virus-resistenten Gerste „Golden Promise“ (Hoffie et al., 2022) und einem Pilz-resistenten Sommer-Raps „Mozart“ (Pröbsting et al.,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

1.1 Planung und Ablauf des Vorhabens

In Phase 1 sollen die GE-Pflanze und ihre Elternlinie mittels Whole Genome Sequencing (WGS) genau charakterisiert werden und die erhaltenen Sequenzen mit einer Referenzkarte des bekannten Raps-Pangenoms (Song et al., 2020, Lee et al., 2020) abgeglichen werden.

In einer zweiten Phase sollen geeignete Nachweis- und Identifizierungsverfahren entworfen werden. Diese Assays werden anschließend an homogenen und heterogenen Proben in Hinsicht auf „GVO“-Gehalte von 0,1 % und 0,9 % (Masse/Masse) getestet, wobei die Analysen durch ein ISO 17025 akkreditiertes molekularbiologisches Untersuchungslabor durchzuführen sind.

1.2 Wissenschaftlicher und technischer Stand, an den angeknüpft wurde

Nach allem wissenschaftlichen Dafürhalten ist es nicht möglich einer Mutation anzusehen wie diese entstanden ist – auf natürlichem Wege, durch klassische Mutagenese mittels Strahlung, bzw. chemischer Behandlung oder durch GE (Leopoldina, 2019; ENGL, 2019). Der präzise Nachweis von GE-Pflanzen ist nach jetzigem Wissensstand von bekannten Informationen über die Genomänderung und die Verfügbarkeit von Referenzmaterial abhängig. Nur wenn auch der Hintergrund, in welchem die Mutation erzeugt wurde, ausreichend gut charakterisiert ist und es in der Elternlinie eine nur dort vorkommende spezifische Marker-Sequenz (im Vergleich zum restlichen Pan-Genom) gibt, ist auch eine Identifizierung theoretisch denkbar. Sowohl induzierte Mutation als auch Marker-Sequenz müssten allerdings sehr eng gekoppelt vorliegen, damit es nicht durch ein Crossing-over in den Folgegenerationen zu einem Kopplungsbruch kommt. Da die Neue Gentechnik *in trans* wirkt, besteht die Möglichkeit die evtl. verwendeten Konstrukte mit der Genschere von den induzierten Mutationen durch Segregation oder Kreuzungen unabhängig zu vererben, sodass klassische PCR-Nachweise auf die im Genom der editierten Pflanze verbliebenen Transgen-Resten nicht mehr greifen. Als mögliche besser geeignete Methoden sind die digitale droplet PCR oder die Amplikon-Tiefensequenzierung mittels short-read NGS zu nennen. Allerdings waren beide Ansätze zum Zeitpunkt der Ausschreibungen noch nicht systematisch evaluiert um GE-Pflanzen nachzuweisen, insbesondere auch hinsichtlich der Quantifizierung des GE-Anteils in komplexen Mischungen (ENGL, 2019).

2. Material und Methoden

2.1 Material

Im Teilprojekt RapsNMT untersuchen wir Mutanten, welche aus unserem BMBF-geförderten BS-TALEN Projekt zur Anwendung von CRISPR/Cas in der Resistenzzüchtung hervorgegangen waren. BS-TALEN hatte zum Ziel Raps weniger anfällig gegenüber dem bodenbürtigen Schadpilz *Verticillium longisporum* zu machen (Pröbsting et al., 2020). Wir hatten zuvor unser Zielgen „*CRT1a*“ durch „Funktionale Genomik“ identifiziert, d.h. wir haben festgestellt, dass eine Kopie in der Pflanze-Pilz Interaktion verstärkt exprimiert wird. Durch einen CRISPR/Cas-induzierten KO kam es zu einer erhöhten Resistenz gegen den Pilz, sodass es sich bei *CRT1a* um ein sogenanntes Anfälligkeits/Suszeptibilitäts-Gen handelt (Abb. 2A). Raps (*Brassica napus*) entstand vor ca. 7500 Jahren durch eine Hybridisierung zwischen Rübsen (*Brassica rapa*, AA-Genom) und Gemüsekohl (*Brassica oleracea*, CC-Genom) und ist damit eine allotetraploide Kulturpflanze (Chalhoub et al., 2014). Daher findet sich in der Regel jedes Raps-Gen auf mindestens zwei Genorten (Loci), einem im AA-Subgenom und einem im CC-Subgenom. Im Fall von *CRT1a* gibt es aber durch Genduplikation insgesamt 4 Loci, von denen wir speziell die A09-Kopie ausschalten wollten, aber aufgrund der hohen Ähnlichkeit mit der C09-Kopie sind beide Loci „On-Target“, während die beiden A01 und C01 Loci wichtige Unterschiede im Bereich der PAM und der sgRNA aufweisen, sodass diese als „Off-Target“ definiert werden können (Abb. 2B).

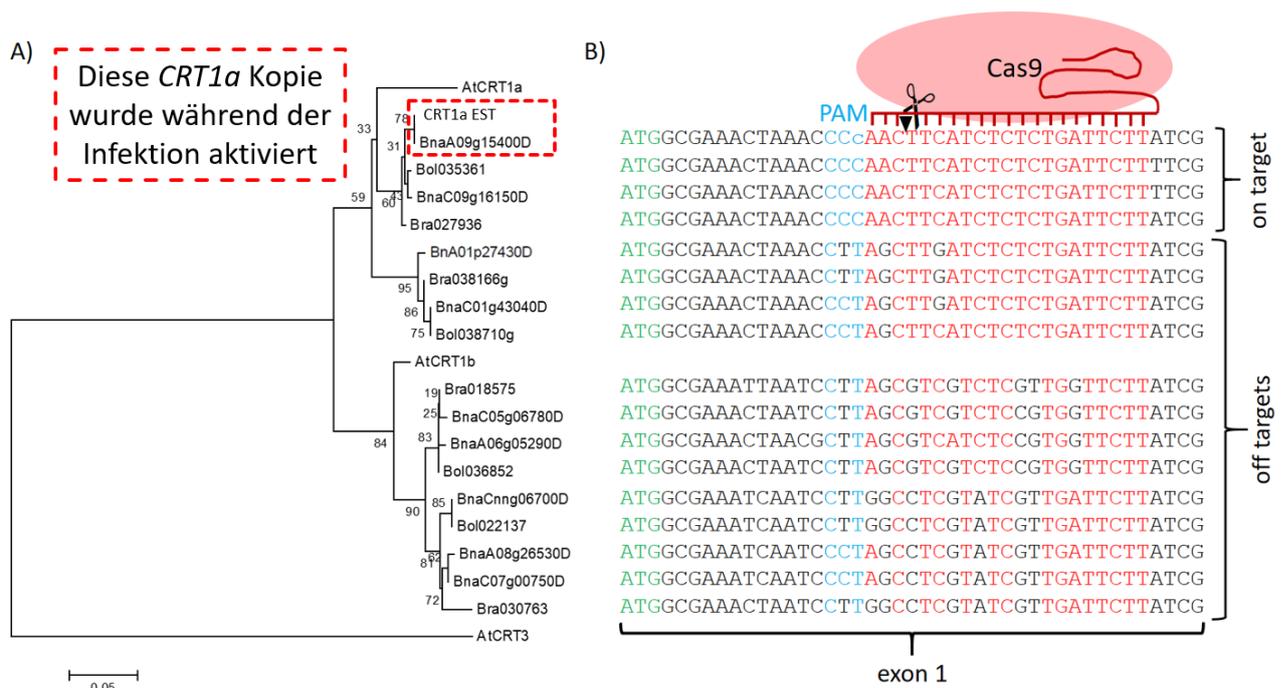


Abb. 2: Das Genom vom Raps (*Brassica napus*) besteht aus den beiden Subgenomen der „Raps-Eltern“ *Brassica oleracea* (Bol) und *Brassica rapa* (Bra). A) Im Raps gibt es insgesamt 12 Calreticulin Loci, von denen hier jedoch nur die 4 *CRT1a* und die 4 *CRT1b* Loci dargestellt sind (kompletter Baum in Abb. S8A). In unserem Fall liegt der Fokus auf dem *CRT1a* A09 Locus, da gerade diese Kopie während der Infektion angesprochen wurde. Von *CRT1a* gibt es somit insgesamt 8 Allele und es ist daher nicht damit zu rechnen durch Zufall einen Raps zu identifizieren, in dem all diese Kopien auf natürliche Weise mutiert und es so zu einem neuen Phänotyp kommt. Dieser auf Neighbor-Joining basierende phylogenetische Stammbaum wurde mit cDNA Sequenzen des Winter-Raps Referenzgenoms „Darmor“ erstellt. B) Durch CRISPR/Cas9 wurde es nun möglich spezifisch nur die On-Target Loci A09 und C09 zu mutieren, sodass die Off-Target Loci A01 und C01 intakt blieben und so weniger Zielkonflikte („trade-offs“) zu erwarten sind.

Raps ist also eine junge Kulturpflanze mit einem aus diesem Grund relativ engen Genpool. Das bedeutet, dass es vergleichsweise wenig genetische Variation in den verfügbaren Rapslinien gibt und es so nur künstlich möglich ist einen neuen Phänotyp hervorzurufen, indem man mittels Mutagenese versucht Unterschiede in allen Genkopien zu induzieren. Die Mutationszüchtung spielt

damit eine ganz wesentliche Rolle bei der Verbesserung der Rapseigenschaften. Unser Fokus lag bei der Mutagenese auf der auf Chromosom A09 liegenden *CRT1a* Kopie, weil diese zu 100% identisch mit dem von uns während der Infektion exprimierten Sequenz (EST) übereinstimmt (Abb. 2A). Weil wir aber beim Einsatz ungerichteter chemischer Mutagenese widersprüchliche Ergebnisse erhalten haben, wurde schließlich CRISPR/Cas9 als Alternative genutzt und ein Konstrukt erzeugt, welches mit Hilfe des *Agrobacterium tumefaciens* Stammes GV31010 in den Sommer-Raps Mozart transformiert wurde (Pröbsting et al. 2020). Die sgRNA für die Steuerung der Genschere zur Zielsequenz wurde so programmiert, dass nur die beiden *CRT1a*-Loci auf dem A09 und C09 Chromosom geschnitten und mutiert werden sollten, was auch der Fall war (Abb. 2C). Das so erhaltene Referenzmaterial sollte zum einen der veröffentlichten Mutante (C3E3) entsprechen, zum anderen aber auch bereits Transgen-frei sein. Ursprünglich war geplant, neben der Raps Elternlinie Mozart und einer Mutante auch Express 617 als Kontrolllinie zu sequenzieren. Da die Genom-Sequenz von Express 617 aber zu Projektbeginn veröffentlicht wurde (Lee et al. 2020), konnte stattdessen eine zweite Mutante hinzugenommen werden. Aus diesem Grund wurde eine geeignete Schwester-Pflanze (C3E4), von der wir annehmen konnten Transgen-frei zu sein, mit in das Projekt aufgenommen (Abb. 3).

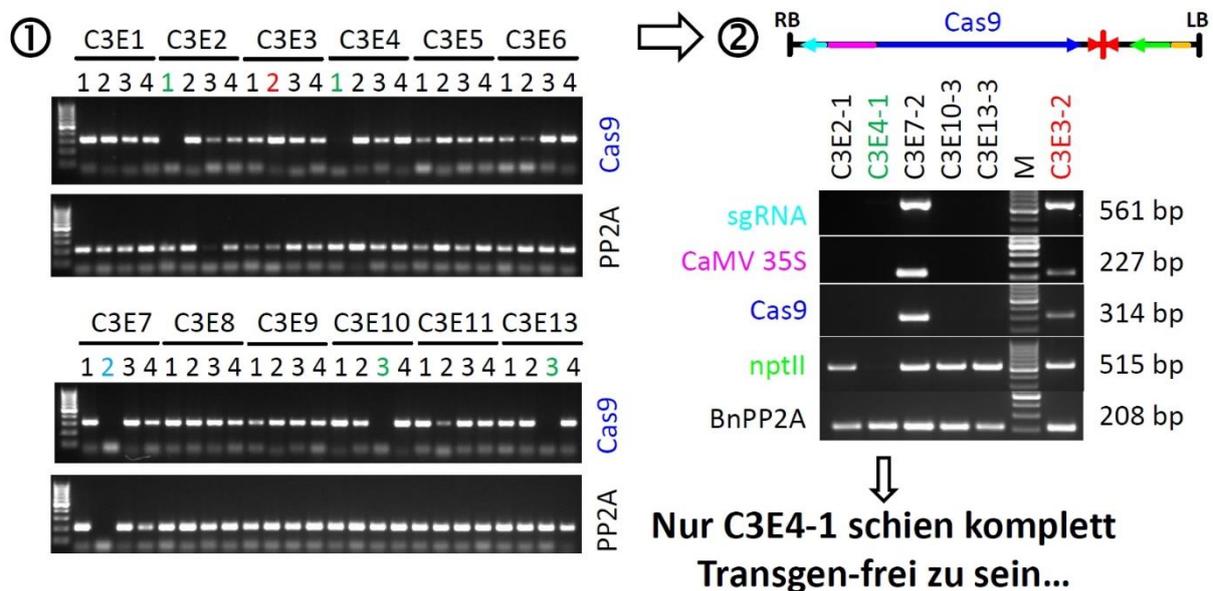


Abb. 3: Ergebnisse der PCR-Analyse. Nach einer Vorauswahl verschiedener T2-Nachkommen aus dem C03-Event (Cas9-Negative Pflanzen ①), wurde in einem zweiten Experiment nach weiteren Konstrukt-spezifischen Sequenzen gesucht (②). Nur die Pflanze C3E4-1 schien komplett Transgen-frei zu sein, während die Pflanze C3E3-2 eindeutig transgen ist. Das Gen *BnPP2A* fungiert jeweils als Kontrolle, dass die isolierte Template-DNA in Ordnung war. Erst durch eine Komplettssequenzierung des Genoms (Whole Genome Sequencing, WGS) mittels long-read PacBio wurde entdeckt, dass auch die C3E4-Mutante noch ein kleines T-DNA Fragment des Nos-Promoters (gelb) enthielt und die als Kontrolle mitgeführte C3E3 insgesamt vier weitere T-DNA Integrationen besitzt (vgl. Tab. 2 und Abb. 6).

Diese beiden T2-Mutanten C3E3 und C3E4 wurden anschließend durch Sanger-Sequenzierung hinsichtlich der Mutationen in den *CRT1a* Loci überprüft und entsprachen den in Pröbsting et al. (2020) beschriebenen induzierten Mutationen, wobei die C3E3 Mutante homozygot im A09 Locus, aber nicht in den Allelen des C09 Locus mutiert war und die C3E4 Mutante wie die T0-Ausgangspflanze in beiden Allelen des A09 Locus heterozygot und auch in einem Allel des C09 Locus mutiert war (Abb. 4). Dies beweist, dass die Genschere beide On-Target Loci gefunden und geschnitten hat, sodass dort Mutationen entstehen konnten. In den Off-Target Loci A01 und C01 fanden wir durch Sanger-Sequenzierung hingegen keine Mutationen.

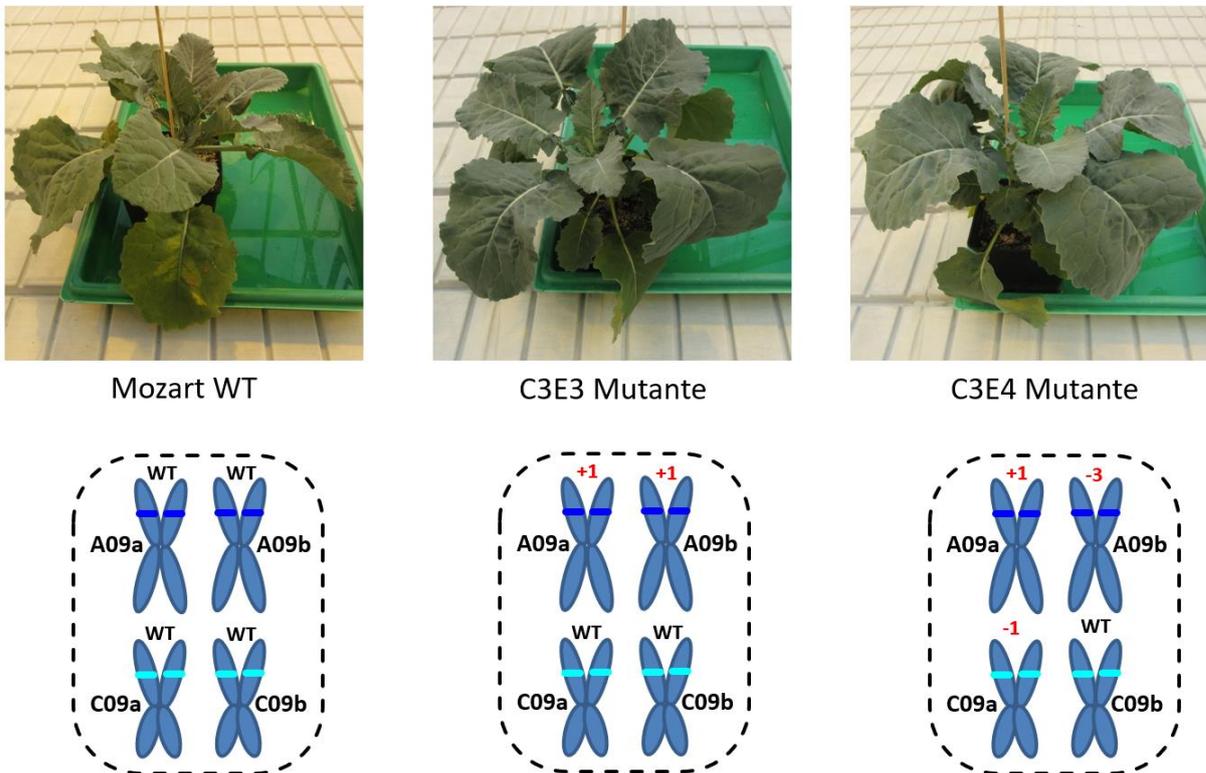


Abb. 4: Die für das Projekt verwendeten Pflanzen. Neben der Ausgangslinie „Mozart“, einem recht gut zu transformierenden und zu regenerierenden Sommer-Raps wurde die eindeutig transgene C3E3 Mutante und die zu Projektbeginn für Transgen-frei gehaltene C3E4 Mutante ausgewählt. Für das WGS des WT wurden allerdings Blätter von 8 verschiedenen Pflanzen gepoolt, während das WGS der Mutanten auf diese Einzelpflanzen zurückging. Das Saatgut dieser beiden T2 Mutanten wurde geerntet und in späteren Versuchen verwendet. Unter den Pflanzen ist die jeweilige genetische Konstitution für die *CRT1a* A98 und C09 Loci dargestellt. Bei der +1 bp Mutation handelt es sich um eine Insertion von einem Nukleotid („T“), durch die -3 bp Deletion ging die Aminosäure Phenylalanin an Position 8 verloren und bei der -1 bp Deletion kann es ebenfalls zu einem Frame-Shift durch den Verlust eines Nukleotids („T“).

Dieses Material ist einzigartig in den Möglichkeiten sowohl einen einzelnen, homozygot mutierten Locus (A09) zu untersuchen (C3E3 Mutante), als auch zwischen verschiedenen Allelen zweier homeologer Loci anhand von 7 SNPs differenzieren zu können (C3E4 Mutante). Gerade aus der C3E4 Mutante können Populationen mit unterschiedlichen Allel-Zusammensetzungen, aber vorhersagbaren Anteilen untersucht werden was eine statistische Analyse erlaubt. Die kann bei allopolyploiden Kulturpflanzen mit komplexen Genomen wie dem Raps (auch Baumwolle, Tabak, Erdnuss, Quinoa, Maniok, Triticale, Emmer oder Hartweizen) eine wichtige Rolle spielen, da es von jedem Gen mindestens zwei – meist sehr ähnliche – Gen-Loci bestehend aus je 2 Allelen gibt. Falls dann die Notwendigkeit zur Differenzierung besteht (wie z. B. bei der C3E3 Mutante, da die den Phänotyp verursachende Mutation ja auf dem A09 Chromosom verortet ist), kann quasi nur anhand von Sequenzdaten eine genaue Zuordnung der Mutation erfolgen. Daher scheint short-read Amplicon-Tiefensequenzierung ideal um solche Daten zu generieren. Auch original WGS-Daten enthalten im Idealfall Informationen zu beiden Allelen (allerdings in der Regel nur von einem Individuum), während in den Datenbanken immer nur der haploide Chromosomensatz dargestellt ist. Das bedeutet, dass es für die Chromosomen A09a/b und C09a/b der C3E3 Mutante keine Rolle spielen würde, da wahrscheinlich beide A09 und beide C09 Chromosomen identisch sind. Aber im Falle der C3E4 Mutante kann man durchaus differenzieren und müsste sich für eines der beiden Chromosomen entscheiden. Wenn man nur A09a statt A09b betrachtet ginge damit potentiell interessante Information verloren wie sich in unserem Raps-Beispiel noch zeigen wird.

Weitere verwendete Rapslinien (alle WT) zum Test der Primer-Spezifität:

Drakkar, Forte, Perfekt, Palma, Westar, Express617, Darmor, Sherlock, Columbus, Bristol, Tapidor, Ningyou7 und Avatar (für die Saatgutmischungen)

2.2 Methoden

2.2.1 DNA-Isolation

Die Isolation genomischer DNA für PCR Amplifikationen erfolgte mit dem GeneJET Plant Genomic DNA Purification Mini Kit (#K0791, Thermo Scientific) aus 100 mg frischem Blattmaterial unter der Verwendung einer Kugelmühle (Precellys® 24 der Firma PEQLAB, 2x mit Programm „Soft Tissue“) und Zircosil Beads (Mühlmeier Mahltechnik). Als Modifikation beim letzten Schritt wurden statt 100 µl nur 70 µl des zuvor auf 80°C erwärmten Elutionspuffers verwendet.

2.2.2 Klonierungen

Die Klonierung von PCR-Fragmenten für die Sanger-Sequenzierung (bei eurofins) erfolgte mit dem pGEM®-T Easy Vector System (Promega) in kompetente *E. coli* DH5α durch Hitzeschock-Transformation und anschließender Selektion auf Ampicillin/IPTG/X-Gal-haltigen Agarplatten). Plasmid-DNA wurde mit dem NucleoSpin Plasmid Mini Kit (Macherey-Nagel) isoliert.

2.2.3 PCR

Die parallele Amplifikation mit unterschiedlichen Primer-Kombinationen erfolgte mittels Touch-Down PCR (Initial-Denaturierung für 3 Min. bei 95°C, 20 Sek. Denaturierung, 20 Sek. Annealing die ersten 10 Zyklen beginnend mit 65°C/-1°C pro Zyklus, dann 25 Zyklen bei 55°C und je nach erwarteter Produktlänge 20 Sek. bis 2 Min. Elongation bei 72°C, sowie einer abschließenden Elongationsphase bei 72°C für 5 Min. und Abkühlen auf 12°C). Als DNA-Polymerasen kamen die DreamTaq (Fermentas) oder die Phusion High Fidelity mit Proof-reading Funktion (Thermo Fisher) zum Einsatz. Die verwendeten Primer finden sich im Anhang (Tab. S1). PCR-Produkte wurden mit dem Nucleo Spin Gel and PCR clean-up Kit (Macherey-Nagel) gemäß Protokoll durchgeführt mit dem einzigen Unterschied, dass für die Elution der Puffer auf 80°C erhitzt wurde.

2.2.4 Standard Operating Procedure (SOP) für den Nachweis

Der Nachweis bekannter Mutationen erfolgte mittels short-read Amplikon-Tiefensequenzierung an einem MiSeq und später auch an einem NextSeq (beides Geräte von Illumina) wie in der separaten SOP-Datei beschrieben.

2.2.5 Bioinformatische Analysen / verwendete Software

2.2.5.1 Read-Count Statistics: EmEditor / Notepad++

Suchstrings	Zielsequenz
@ / >	Anzahl Gesamt-Reads/Sequenzen in Datei
GTAGGCT	Bezugssystem selektiv On-Targets A09 / C09
GGATTAT	Bezugssystem selektiv Off-Targets A01 / C01
CTTTCATCTCTCTGATT	A09a Mutation +1 bp (+T)
CATCTCTCTGATT	A09b Mutation -3 bp (+TTC)
CTCATCTCTCTGATT	C09a Mutation -1 bp (-T)
CTTCATCTCTCTGATT	C09b / On-Target WT Reads
CTTGATCTCTCTGATT	A01 / C01 Off-Target WT Reads
TCATCCA	Bezugssystem selektiv A09 (SNP-Region)
TCAGCCA	Bezugssystem selektiv C09 (SNP-Region)
GAAAGAAGCTGATACATTTTGCAGATG	A09a SNP (A)
GAAAGAAGCTGATTCATTTTGCAGATG	A09b SNP (T)

2.2.5.2 Multisequenz-Alignment (MSA)

Reverse-Komplementierung: https://www.bioinformatics.org/sms/rev_comp.html

Algorithmen: CLUSTALW, MUSCLE, MAFFT

z. B. via <https://www.genome.jp/tools-bin/clustalw>

Visualisierung: GeneDoc, AliView

2.2.5.3 Datenbanken

Allgemein: NCBI > Nucleotide via BLAST: <https://www.ncbi.nlm.nih.gov/>

Raps-Referenzgenom: https://www.genoscope.cns.fr/projet_CCM/cgi-bin/gbrowse/bnapus-bzh/

Raps-Pan-Genom Browser: <http://cbi.hzau.edu.cn/cgi-bin/bnapus/blast>

2.2.5.4 Off-Target Vorhersage



<https://cctop.cos.uni-heidelberg.de/>



<http://www.rgenome.net/cas-offinder/>



<http://cbi.hzau.edu.cn/CRISPR2/>



<https://crispr.dbcls.jp/>

2.2.5.5 Software-Pipeline AmplikonSeq-Workfow v1.2_17.10.2022_PLANTON GmbH (siehe SOP)

Programm	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/

3. Ergebnisse

3.1 Ausführliche Darstellungen der wichtigsten Ergebnisse

Weitere Ergebnisse oder technische Details zu den hier gezeigten Ergebnissen sind in einer EXCEL-Datei im Anhang (Supplement) aufgeführt, auf die jeweils durch den Zusatz „S“ verwiesen wird.

3.1.1 Erste Ergebnisse aus dem long-read PacBio WGS

Der Fokus des Projekts liegt zwar auf der Anwendung der short-read Amplikon-Tiefensequenzierung durch Illumina-Technologie, aber eine Grundvoraussetzung für deren Anwendung ist die Kenntnis der Genom-Sequenz und der Mutationsorte. Daher ist es angezeigt zunächst auf die Ergebnisse des ebenfalls durchgeführten long-read WGS durch PacBio-Technologie einzugehen. Da die Genom-Assemblierung noch nicht abgeschlossen ist, sind die hier vorgestellten Ergebnisse als vorläufig zu betrachten. Die PacBio Daten stammen für die beiden Mutanten von der Firma Genewiz in Leipzig (Azenta Life Sciences) und für die Mozart WT Sequenzierung von der Firma BGI (via Lutz Grohmann vom BVL).

3.1.1.1 WGS-Analyse als Basis für die Identifizierung

Durch ein Haplotypen-Assembly wurde für die Identifizierung versucht beide A09 Schwesterchromosomen (auf 2n-Basis) auf maximal möglicher Länge zu assemblieren, um so Event-spezifische Polymorphismen zu identifizieren, welche entweder mit der relevanten +1 bp (A09a) oder der -3 bp (A09b) Mutation gekoppelt sind. Während in der C3E3 Mutante beide A09 Chromosomen identisch sind (die +1 bp Mutation also homozygot vorliegt), sollten in der C3E4 Mutante beide Versionen vorliegen und anhand des Mutationsorts zu unterscheiden sein. Allerdings wurde in dieser Mutante nur die A09b Version assembliert. Damit stehen aber beide Versionen für einen Vergleich mit dem WT Mozart zur Verfügung. Mit einer Länge von ca. 13600000 bp entsprechen diese Contigs in etwa 25% der Gesamtlänge des Raps A09 Chromosoms. Ein Vergleich zeigte, dass die meisten Unterschiede zwischen dem Wildtyp und den beiden Mutanten bestehen, welche untereinander sehr ähnlich sind (und ja auch beide auf das gleiche Ausgangs-Event C3 zurückgehen). Viele gefundene Polymorphismen sind jedoch 1 bp InDels und es ist sehr wahrscheinlich, dass dies Artefakte sind, welche bei PacBio manchmal an Stellen mit vielen Wiederholungen einzelner Nukleotide auftreten. Die PacBio Daten sind dennoch erstaunlich gut und Fehler können wenn die Reads untereinander verglichen werden relativ leicht erkannt und eine korrekte Konsensussequenz abgeleitet werden. Nur wenn die Polymorphismen zwischen den Reads in etwa dem Verhältnis 50:50 entsprechen handelt es sich wahrscheinlich um einen echten Unterschied. Für den hier untersuchten Bereich konnten zwischen den A09a (aus C3E3) und A09b (aus C3E4) Chromosomen 10 potentielle Marker identifiziert werden, von denen der nächste 3370 bp neben der -3 bp Mutation (A09b) lag (Tab. 1). Um eine Aussage treffen zu können in welchem Schwesterchromosom der Unterschied aufgetreten ist wurden die betreffenden Stellen im Mozart WT und 3 weiteren Raps Genomen gesucht und falls dort gefunden als das natürliche Allel definiert. Das andere ist damit automatisch die Mutation (rot markiert). Keine dieser Mutationen liegt nahe einer der sgRNA ähnlichen Sequenz und somit handelt es sich höchst-wahrscheinlich um spontan aufgetretene Mutationen. Für die Berechnung der Rekombinationsrate wurden publizierte Vergleichswerte der Genomgröße in cM herangezogen, welcher durch die Genomgröße in bp dividiert und mit dem Abstand zwischen Mutation und Marker in bp multipliziert wird (Tab. 1, Tab. S4). Für einen mit der -3 bp Mutation gekoppelten SNP aus der T0 (Abb. S10) ergibt sich die niedrigste Rekombinationsrate womit dieser SNP für die Identifikation am besten geeignet scheint. Bei einer Rekombinationsrate von 0,01% ist es zwar recht unwahrscheinlich, dass die Kopplung durch ein Crossing-over gebrochen wird, aber eben doch nicht auszuschließen. Statistisch gesehen sollte in einer von 10.000 Pflanzen ein solcher Kopplungsbruch erfolgen, d.h. man könnte auf einem 1 ha großen Rapsfeld ca. 5 solche Crossing-over erwarten.

Tab. 1: Gefundene Event-spezifische Polymorphismen (in rot) in dem die CRISPR-Mutation enthaltenen A09 Contigs mit ca. 13.600.000 bp Länge

≈ Lage	Typ	A09a (+1)	A09b (-3)	Mozart	Westar	ZS11	Darmor	p REK (%)
-5007703	SNP	T	C	A09b	A09b	A09b	A09b	17.85
-1258389	InDel	-1 bp		A09b	A09b	A09b	A09b	4.48
-1047828	SNP	C	T	A09a	A09a	A09a	A09a	3.73
3370	SNP	A	T	A09a	A09a	A09a	A09a	0.01
762678	SNP	G	T	A09a	A09a	A09a	A09a	2.72
771933	InDel		-18 bp	A09a	A09a	A09a	A09a	2.75
782720	InDel		-13 bp	A09a	A09a	A09a	A09a	2.79
974374	SNP	T	G	A09b	n.d.	A09b	n.d.	3.47
1243103	SNP	T	A	A09a	A09a	A09a	A09a	4.43
2493529	SNP	C	T	A09a	A09a	A09a	A09a	8.89

Wir haben in einer Korrelationsanalyse 100 Pflanzen untersucht, um zu schauen, ob die -3 bp Mutation auf A09b immer mit dem SNP („T“) zusammenvererbt wird. 99 Sequenzen des SNP-Assays waren auswertbar und sowohl Mutation als auch SNP korrelierten in den erwarteten Verhältnissen (Tab. S3). Allerdings ist es durchaus möglich, dass bei einer höheren Fallzahl mit mehr als 10.000 Individuen doch Kopplungsbrüche detektiert werden. Neben der Gefahr eines Kopplungsbruchs durch ein Crossing-over, welches im Raps auch zwischen den homeologen Chromosomen beider Subgenome auftreten kann (Abb. S8), gilt es zu bedenken, dass die -3 bp Mutation züchterisch überhaupt nicht relevant ist, da der Verlust nur einer Aminosäure (Phe8) keine Auswirkungen auf den Resistenzphänotypen hat (Abb. S4). Dennoch gibt es zu diesem SNP keine wirkliche Alternative, denn nur dieser ist sicher mit dem Mutationsort physisch gekoppelt, da es 9 PacBio Reads gibt welche dies beweisen und es in dieser Studie nicht um den Sinn dieses GE Nachweises, sondern nur um dessen Machbarkeit geht (Abb. 5).

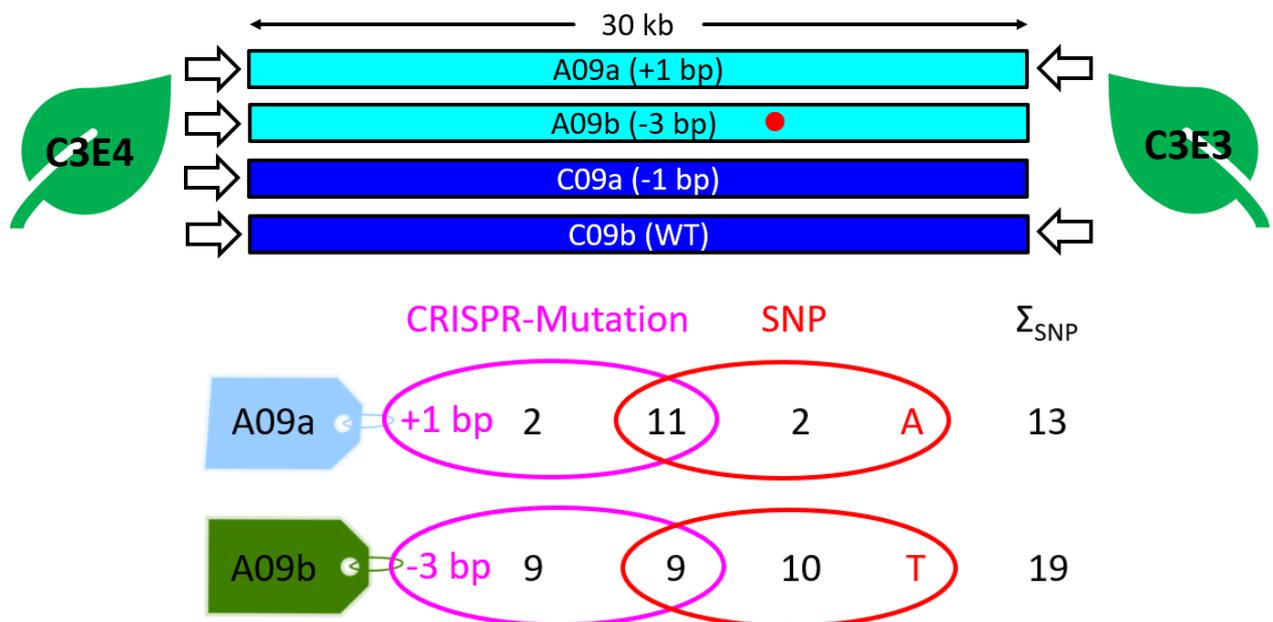


Abb. 5: Der mit der -3 bp CRISPR/Cas-Mutation assoziierte SNP (● T statt A) ist eindeutig mit dieser gekoppelt, da 9 physische PacBio-Reads beide Markern enthalten. Für beide Versionen gab es in der C3E4 Mutante ähnlich viele Reads, sodass man aufgrund des beobachteten 50:50 Verhältnisses davon ausgehen kann, dass es sich nicht um Sequenzierfehler handelt.

Ein Vergleich dieser SNP-Region über das PacBio Raps Pan-Genom (plus der beiden „Raps-Eltern“ *Brassica oleracea* und *Brassica rapa*) deutet bereits an, dass es sich hier um einen **Event-spezifischen SNP** handelt, da dieser auch nicht im Mozart Wildtyp Genom zu finden ist (Tab. S3). Damit steht dem Versuch einer Identifikation durch Kombination zweier short-read Illumina-Assays (einem zum Nachweis der -3 bp Mutation und einem zum Nachweis des SNP) nichts mehr im Wege (siehe Abschnitt 3.1.2.5).

3.1.1.2 Off-Target Analyse

Um mögliche Off-Target Effekte durch das Genom-Editing zu untersuchen wurden zunächst alle *CRT1a* Kopien auf den Chromosomen A09, C09, A01 und C01 untersucht und mit keiner Methode (Sanger, Illumina oder PacBio) wurden hier Mutationen in den beiden als Off-Targets definierten Kopien A01 oder C01 entdeckt. Durch bioinformatische Vorhersagen mit verschiedenen Online-Tools (beruhen auf dem Darmor Referenzgenom) können noch weitere mögliche Off-Target Sequenzen identifiziert werden. Insgesamt wurden 4 Online-Dienste (siehe 2.2.5.4) genutzt und es wurden drei Kriterien zugrunde gelegt:

- ① Die Sequenz sollte zu 80% identisch mit der sgRNA-Bindestelle sein
- ② Die Sequenz muss neben einer PAM für *SpCas9* (NGG) liegen
- ③ Es sollten keine Unterschiede in der Seed-Region (erste 12 bp neben PAM) geben

Nach Abgleich der durch die 4 Online-Tools vorgeschlagenen Kandidaten-Sequenzen wurde eine Liste mit weiteren 18 mehr oder weniger wahrscheinlichen Sequenzen für die Off-Target Analyse erstellt (Tab. S5). Ein Problem ist die Schwierigkeit hier eine vollständige Liste zu erstellen, da Mozart (noch) nicht als Referenz-Genom herangezogen werden kann. Ein anderes Problem ist, dass eine direkte Suche nach einer unbekanntem Mutation im Genom nicht ohne weiteres möglich ist. Daher wurden zunächst nur die WT Sequenzen in den WGS-Reads der C3E3 und C3E4 Mutanten direkt in sense und antisense Orientierung gesucht. Wenn diese gefunden wurde ist diese Sequenz in zumindest einem der beiden Allele nicht mutiert (theoretisch könnte also das zweite Allel noch eine Off-Target Mutation tragen). Es wurden allerdings auch durch die Analysen am IPK keine Off-Target Effekte detektiert, sodass man davon ausgehen kann dass CRISPR/Cas9 erstaunlich genau eingesetzt werden kann.

3.1.1.3 Analyse des Transgen-Status

Prinzipiell liegt der Vorteil der NMT darin, dass bei der Verwendung der Agrobakterium-vermittelten Transformation T-DNA Konstrukt und Mutation nicht zwangsläufig gekoppelt vorliegen und damit durch Segregation bei der pflanzlichen Selbstbestäubung oder durch Kreuzung getrennt vererbt werden können. So können mutierte, aber im Endeffekt Transgen-freie Pflanzen entstehen. Zu Projektbeginn wurden zwei CRISPR/Cas-Mutanten aus der zweiten Generation (T2) für das Projekt mittels PCR selektiert, wobei die C3E3 Mutante eindeutig transgen (aber nur im A09 Locus homozygot mutiert) war, während die C3E4 Mutante bereits Transgen-frei zu sein schien (Abb. 3). Durch die Analyse einzelner Reads aus dem PacBio-WGS, sowie ersten Assembly-Versuchen an der CAU konnten in beiden Mutanten allerdings einige T-DNA Integrationsorte bestimmt werden. Überraschenderweise handelte es sich meist lediglich um T-DNA Fragmente und das komplette Konstrukt konnte nur einmal in der C3E3 Mutante wiedergefunden werden, was erklärt warum das kleine Nos-Promoter Fragment im C04 Chromosom der C3E4 Mutante übersehen werden konnte. Insgesamt konnten 6 Loci identifiziert werden, welche T-DNA (Reste) enthalten (Tab. 2). Als Besonderheiten sind zwei der T-DNA Fragmente mit größeren Chromosomen-Aberrationen assoziiert und ein Genort (A02) beherbergt neben der kompletten T-DNA noch 4 Bruchstücke, von denen zwei invertiert vorliegen (Abb. 6). Prinzipiell stellt sich nun die Frage warum, wie und in welcher Generation die T-DNA fragmentierte. Die erfolgreiche PCR-Amplifikation aller getesteten 5 Integrationsstellen in 10 T1-Pflanzen deutet zumindest daraufhin, dass die Fragmentierung und multiple Integration bereits in der Gewebekultur / T0-Ursprungspflanze stattgefunden hat (Abb. S5). Die einzelnen Signale der Kanamycin-Resistenz in den Pflanzen C3E2-1, C3E10-3 und C3E13-3 (Abb. 2.2) sind damit wahrscheinlich auf das T-DNA Fragment auf Chromosom C01 zurückzuführen. Da laut dieser Analyse die C3E3 Mutante noch zwei intakte Konstrukte mit der Genschere und der zugehörigen sgRNA enthält, kann dies erklären warum bei der Sanger-Sequenzierung eines Blattes dieser Mutante ein Allel im *CRT1a* C09 Locus eine -2 bp Deletion aufwies, während bei der Sequenzierung eines anderen Blattes derselben Pflanze mittels short-read NGS beide C09 Allele noch als WT vorlagen. Mit der hohen Anzahl an Konstrukt-Resten in der C3E3 Mutante geht die wissenschaftliche Sorge um mögliche Positionseffekte einher, welche bei Homozygotie den zu untersuchenden Phänotyp beeinflussen könnten. Das träfe z.B. auf die A01 und A05 Chromosomen zu, aus welchen das T-DNA Fragment nur noch durch Kreuzung entfernt werden kann. Ansonsten kann man sehen, dass auch durch das Selbsten bereits viele in der T0 vorhandene Konstrukt-Reste durch Segregation verloren gehen können – wie anhand der C3E4 Mutante zu sehen ist (Abb. 3, Tab. 2). Da in der C3E4 Mutante das T-DNA Fragment auf Chromosom C04 nur heterozygot vorliegt, sollten in der folgenden T3 Generation Pflanzen zu finden sein, welche tatsächlich nur noch mutiert und nicht mehr transgen sind (Abb. S7).

Tab. 2: In den Mutanten gefundene T-DNA Fragmente

Pflanze	Fragment	Lage	Chromosom / Gen	Arabidopsis	Potentielle Positionseffekte	Status	Kommentar
C3E3	3008 bp	Promotor	A01 p17650_DAR A01 T0079300WE	AT4G26430	COP9 signalosome complex (CSN6B)	homozygot (nur in C3E3)	
C3E3	25269 bp	Promotor	A02 p34910_DAR A02 T0337400WE	AT3G26940	Receptor Like Cytoplasmic Kinase (CDG1)	heterozygot (nur in C3E3)	Komplette T-DNA + 4 Fragmente
C3E3	6395 bp	Intron	A05 p38840_DAR A05 T00105079001WE	AT3G10270	DNA Gyrase B1 (GYRB1)	homozygot (nur in C3E3)	Komplette T-DNA
C3E3	1989 bp	ORF	C01 p00510_DAR 1716T0014800WE	AT1G79620	VASCULAR-RELATED RLK (VRLK1)	heterozygot (nur in C3E3)	
C3E3	315 bp	3'UTR Intron	C04 p71980_DAR C04 p72150_DAR	??? AT2G46370	unknown protein JAR1-like (KO durch Deletion)	heterozygot (nur in C3E3)	Translokation oder Deletion
C3E4	563 bp	Promotor Promotor	A03 T0013100WE A03 T0031800WE	AT5G01090 AT5G06390	Concanavalin A-like lectin P Fascilin-like AGP (FLA17)	heterozygot (nur in C3E4)	Deletion von ca. 920216 bp

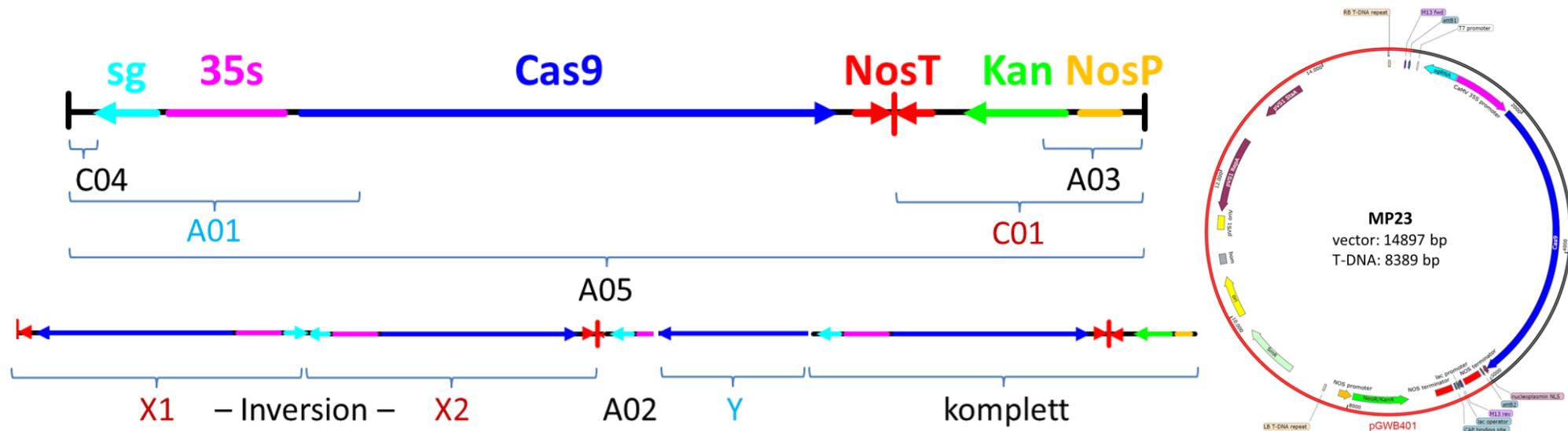


Abb. 6: Übersicht über die gefundenen 6 T-DNA Fragmente in den Chromosomen A01, C01, A02, A03, C04 und A05, sowie den verwendeten Pflanzentransformationsvektor. Nur im A02 und A05 Integrationsort liegt der Vektor noch komplett vor, im A02 zusammen mit 4 weiteren T-DNA Bruchstücken. Wie eine solche Fragmentierung zustande kommen konnte ist ein Rätsel. Ob einige der T-DNA Bruchstücke im A02 Locus aus dem C01 Locus (rotes X) oder A01 Locus (blaues Y) stammen wäre eine interessante Arbeitshypothese.

Für die erneute Suche nach möglicherweise doch Transgen-freien Nachkommen der durch WGS sequenzierten heterozygoten C3E4 Mutante wurden daher Nos-Promotor spezifische Primer entwickelt und für die DNA-Analyse der 48 T3 Pflanzen aus der Populationsanalyse (Tab. S2) eingesetzt. Zehn Nachkommen dieser Pflanzen wiesen wiederholt keine Amplifikation auf und so wurde die T3 Pflanze #18 für eine wiederholte Populationsanalyse (Tab. S3) ausgewählt. Leider fanden sich auch unter den T4 Nachkommen wieder Nos-Promoter Signale in der PCR, was an der Eignung einer konventionellen PCR für den Transgen-Nachweis zweifeln lässt (Abb. S7) und die Notwendigkeit eines WGS für den Beweis der Transgen-Freiheit einer Pflanze unterstreicht.

3.1.1.4 Bewertung der PacBio-Technologie / Notwendigkeit eines WGS

Mittels long-read PacBio WGS können viele qualitativ hochwertige und sehr lange Sequenzen erzeugt werden, welche sich viel besser als kurze Illumina Reads dazu eignen Genome zu assemblieren. Durch das WGS konnten wir nicht nur die Sanger- und Illumina Sequenzdaten an den Mutationsstellen bestätigen, sondern auch zeigen, dass es wahrscheinlich nicht zu Off-Target Mutationen gekommen ist. Darüber hinaus zeigte sich, dass es in der T0 des C3 Ausgangs-Events an mindestens 5 unterschiedlichen Stellen im Genom zur Integration verschiedener T-DNA Fragmente gekommen ist, von denen 4 in der C3E3 Mutante und ein weiteres in der C3E4 Mutante nachgewiesen werden konnten. Ein PCR-Test auf Vektorreste kann lediglich der Vorselektion für ein WGS dienen, welches als einzige Methode zweifelsfrei beweisen kann, dass die Pflanze wirklich nur mutiert und nicht doch noch durch ein verwendetes Konstrukt Transgen ist. Ein WGS wäre auch unabdingbar um zusätzliche Event-spezifische Marker zu detektieren, welche für eine eindeutige Identifizierung des GE Events eng mit der induzierten Mutation gekoppelt sein müssen. Wie stabil ein Identifikations-Assay wäre wird in Abschnitt 3.1.2.7 diskutiert.

3.1.1.5 Bewertung der CRISPR/Cas9-Technologie / Spezifität hinsichtlich möglicher Off-Targets

Die beiden von uns untersuchten *crt1a* Mutanten C3E3 und C3E4 entsprechen der T2 Generation, sodass nicht nur die CRISPR/Cas9-induzierten Mutationen von den T-DNA Resten segregieren konnten, sondern auch evtl. aufgetretene Off-Target Mutationen. Wir haben beide Mutanten zuerst durch short-read Illumina NGS charakterisiert und dabei sowohl die On-Target Loci (A09 und C09), als auch die beiden Off-Target Loci (A01 und C01) amplifiziert. In beiden Fällen konnten keine Mutationen in den *CRT1a* Off-Target Kopien detektiert werden. Diese Beobachtung wurde durch das anschließende long-read PacBio WGS bestätigt, in dem auch keine Mutationen in 18 weiteren untersuchten Off-Target Stellen gefunden wurden (vgl. Bessoltane et al., 2022; Sturme et al., 2022). Damit scheint CRISPR/Cas9 ideal zur Erzeugung von Mutationen an einer gewünschten Zielstelle, auch wenn nicht jede dieser Zufallsmutationen einen Phänotyp hervorrufen wird (z. B. unsere -3 bp Deletion in A09b, welche sich in homozygotem Zustand wie Wildtyp verhält). Eine Diskussion um so wenige putative Off-Target Effekte ist allerdings züchterisch wenig relevant, da Pflanzenzüchter mit den herkömmlichen und von der GVO-Regulation ausgenommenen Mutagenesetechniken (z. B. Strahlung oder chemisch mit EMS) bereits mit wesentlich mehr solch unerwünschter Off-Target Mutationen konfrontiert wurden (Pixley et al., 2022). Diese könnten nun mit den NMT drastisch reduziert werden, wodurch sich zeitintensive Rückkreuzungen vermeiden lassen. Allerdings zeigt sich auch, dass SDN-1 Ansätze mit Hilfe von Agrobakterium-vermittelter Transformation auch Nachteile mit sich bringen können. Das Versprechen schnell Transgen-freies Material zu erhalten, weil die erzeugten Mutationen in der Regel von den verwendeten Konstrukten unabhängig segregieren ist bei einer multiplen Integration in mehrere Chromosomen wie sie hier beobachtet wurde wohl nicht einzuhalten und erfordert dann doch einige Rückkreuzungen. Während eine stabile Transformation der Genschere die Effizienz enorm steigert, entstehen durch die Integration von Fremd-DNA eindeutig GVOs falls es nicht gelingt diese wieder zu entfernen. Daher gibt es Bestrebungen beim GE ganz auf DNA-haltige Konstrukte zu verzichten und die Genschere in Proteinform zusammen mit ihrer sgRNA (sogenannte RNPs) in die Zellen einzuschleusen. Solche Ansätze sind aber noch ineffizient.

3.1.2 Ergebnisse aus der short-read Illumina Amplikon-Tiefensequenzierung

In den meisten Fällen wurde die Sequenzierung an einem Illumina MiSeq Gerät durchgeführt, aber gegen Ende des Projekts liefen einige Runs auf einem Illumina NextSeq, worauf dann gesondert hingewiesen wird. Die Proben wurden für die Homogenisierung, DNA-Isolation und das NGS an die Planton GmbH in Kiel übergeben, die für solche Untersuchungen nach ISO 17025 akkreditiert sind.

3.1.2.1 Etablierung eines Nachweisverfahrens für die *CRT1a* Mutationen

Da es im Raps insgesamt 4 *CRT1a* Loci gibt und Unklarheit über die Vollständigkeit der zuvor erhobenen Daten aus der Sanger-Sequenzierung herrschte wurden zunächst verschiedene Primer entwickelt und in unterschiedlichen Kombinationen getestet, mit dem Ziel alle 4 Gen Loci in gleich guter Weise zu amplifizieren (Abb. S2). Die Ergebnisse sind in Tabelle 3 dargestellt.

Tab. 3: Prozent-Anteile der erzeugten Reads mit den bekannten Mutationen oder WT Sequenzen (C3E4)

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

Von all diesen Assays lagen die Ergebnisse von #3 noch am nächsten am erwarteten Verhältnis, aber man kann schon beobachten, dass in diesem Assay und in Assay #1 alle On-Target Loci (A09 und C09) zu gleichen Anteilen amplifiziert und sequenziert werden konnten. Die Relation zu den Off-Target Loci A01 und C01 stimmt aber oft nicht und wenn man sehr stringent nach der kompletten Sequenz der Amplikons sucht fallen Verluste von 20 bis 25% auf, welche auf Chimärenbildung während der PCR zurückzuführen sind. Dies ist ein interessanter Aspekt wenn komplexe Genome mit sehr ähnlichen Genkopien untersucht werden, wie dies bei allopolyploiden Kulturpflanzen oft der Fall ist. Ähnliche Verlustwerte wurden in den beiden Populationsanalysen festgestellt (Tab. S2 und S3) und die Chimärenbildung bedingt sogar für einen kleinen Teil aller Reads, dass eine Sequenz aus dem A-Subgenom plötzlich mit dem C-Subgenom assoziiert werden kann (Abb. S3). Durch ein stringentes Filtern werden solche PCR-Artefakte aussortiert, aber dadurch stünden auch ca. 25% weniger Reads für die Auswertung zur Verfügung. Weitere Reads gehen bei diesen vier Ansätzen an die Off-Targets verloren, welche für die Detektion der CRISPR/Cas9-Mutationen aber keine Rolle spielen. Daher wurde ein neuer Assay #5 entworfen, in welchem nur noch die On-Targets amplifiziert werden (Tab. S2). Mit diesem Assay wurden verschiedene Rapslinien und die beiden Mutanten C3E3 und C3E4 analysiert (Tab. 4). Beide Mutanten wiesen in etwa die erwarteten Verhältnisse für die mutierten *CRT1a* Allele auf: für C3E3 waren 50% Reads mit der +1 bp Mutation (A09) zu erwarten und 50% WT Reads für den C09 Locus, während die bei der C3E4 Mutante jedes Allel zu 25% auftreten sollte. Zu Beginn wurde mittels Sanger-Sequenzierung der Mutationsstatus der beiden verwendeten Mutanten überprüft und es wurde eine neue -2 bp Deletion im C09 Locus der C3E3 Mutante entdeckt. Durch diese short-read NGS Analyse mittels Illumina zeigte sich jedoch, dass die neue Deletion im C09 Locus der C3E3 nicht wiedergefunden werden konnte und es sich hiermit höchstwahrscheinlich um eine somatische Mutation im für die Sanger-Sequenzierung verwendeten Blatt handelte. Dies zeigt wie wichtig es ist, dass die Konstrukte für das GE wieder entfernt werden wenn Wert auf einen intakten zweiten On-Target Locus aus dem anderen Subgenom gelegt wird. Überraschenderweise zeigten sich in einigen WT-Linien kleine Mengen an Reads falsch-positiver Signale (0,01%) für einige Mutationen (vgl. auch Tab. S2 und S3). Daher wurde dieses Experiment wiederholt und nun traten insgesamt weniger Falsch-Positive auf, nur für Westar gab es keine Verbesserung.

Tab. 4: Ergebnis der NGS-Sequenzierung #5 (in rot Anteil an Falsch-Positiven Reads)

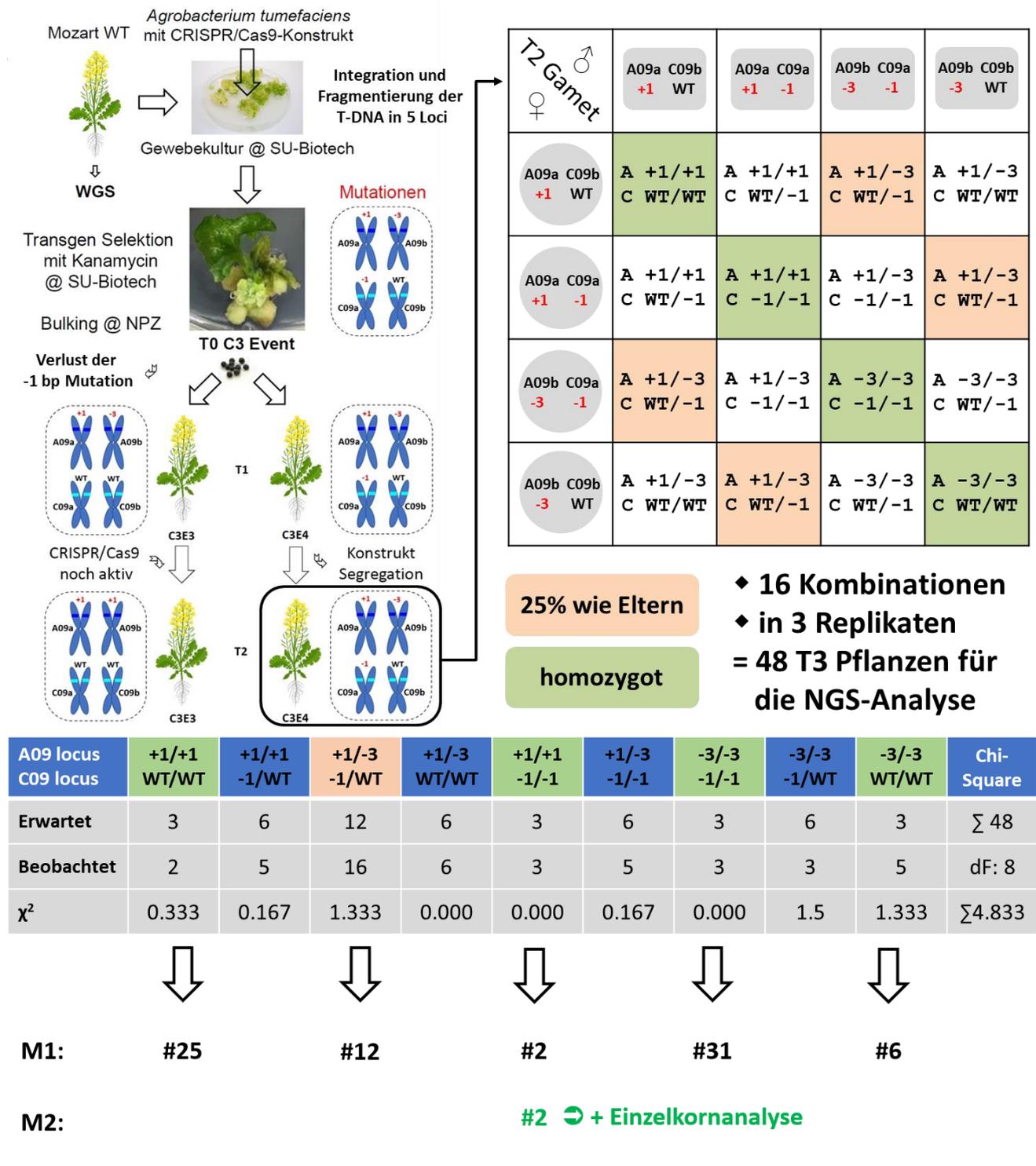
Linie	Reads	+1 bp (A09)	-3 bp (A09)	-1 bp (C09)	WT (C09)
C3E3	44479	47.17%	0	0	52.83%
C3E4	330490	22.54%	24.23%	26.17%	27.06%
Bristol	330351	0	0	0	100%
Columbus	50605	0	0	0.01%	99.99%
Darmor	335676	0	0	0	100%
Drakkar	332892	0	0	0	100%
Express	384676	0	0	0	100%
Forte	315482	0	0	0	100%
Mozart	44440	0	0	0	100%
Palma	356439	0	0	0	100%
Perfect	389337	0	0	0	100%
Westar	56820	0	0	0.01% (2x)	99.99%

Das bedeutet diese Falsch-Positiven Signale treten willkürlich auf und sind nicht vorherzusagen. Eine Mögliche Erklärung für ihr Auftreten in NGS-Läufen mit Multiplexing durch individuelle Indizes (durch „Barcoding“ ermöglichte Zuordnung von Reads zu einer bestimmten Pflanze) wäre sogenanntes „Index-Hopping“, ein technisch nicht zu vermeidendes Problem bei solchen Populationsanalysen, oder durch einen PCR-Fehler (wenn z. B. statt TT nur ein T eingebaut wird). Ein möglicher Hintergrund an falsch-positiven Signalen liegt damit rund eine Zehnerpotenz unter der gesetzlich vorgeschriebenen LOD von 0,1% die sicher bestimmt werden muss. Da aber alle On-Target Loci gleich effizient amplifiziert werden ist dieser Assay für alle weiteren Analysen bestens geeignet.

3.1.2.2 Ergebnisse der ersten Populationsanalyse

In diesem Versuch wurde der Mutationsstatus in einer T3 Population aus 48 Nachkommen der C3E4 Mutante bestimmt. Dabei wurde DNA aus Blattmaterial der Einzelpflanzen verwendet und wiederum stringent gefiltert, wodurch von durchschnittlich 90205 Reads ca. 24% verloren gingen (Tab. S2). Damit beträgt die durchschnittliche Ausbeute an Reads mit dem MiSeq 68415. Wenn es sich nicht um Einzelpflanzen, sondern um eine Mischprobe handeln würde, müssten für das Erreichen der LOD von 0,1% von der verbliebenen Gesamtzahl an Reads 68,4 auf die GE Mutationen entfallen. Liegen in einer Pflanze alle 3 Mutationen vor (wie für 25% aller Pflanzen zu erwarten wäre, siehe Abb. 7), gäbe es anteilig 4 unterschiedliche Allele zu bedenken und man würde ca. 17,1 Reads für jedes mutierte Allel erwarten. Bei einer homozygoten Mutante wie der C3E3 wären es entsprechend doppelt so viele Reads (34,2). Beide Werte liegen jedoch nur geringfügig über der beobachteten Anzahl falsch-positiver Reads, welche Dank der Einzelpflanzen-Analyse klar als solche zu erkennen sind. Bei zwei Pflanzen (#27 und #34) überstieg der prozentuale Anteil Falsch-Positiver sogar die 0,1% LOD, sodass es zu einer Falsch-Interpretation bei der Bewertung dieser Proben käme. Ein Grund könnte hier eine mögliche Kontamination des Blattmaterials mit Pollen einer anderen Pflanze gewesen sein. Ein weiteres Ziel der Populationsanalyse war die Gewinnung von homozygotem Saatgut für die Analyse von Saatgutmischungen und Resistenztests gegenüber dem bodenbürtigen Schadpilz *Verticillium longisporum* (Abb. S4). Für eine statistische

Bewertung der Übereinstimmung von Befund und Erwartung in dieser Populationsanalyse ist die Zahl an untersuchten Individuen mit 48 eigentlich zu klein. Ein Chi-Quadrat Test deutet aber bereits an, dass die Mutationen wie vorhergesagt vererbt werden, auch wenn man für eine verlässliche Aussage mindestens 5 Replikate a 16 Pflanzen (also insgesamt 80 Pflanzen) analysieren müsste (Abb. 7).



#18 ➔ nun wahrscheinlich Transgen-frei...

Abb. 7: Ergebnis der ersten Populationsanalyse und Verwendung des erhaltenen homo- und heterozygoten T3-Saatguts aus der Selbstung der mittels PacBio sequenzierten C3E4-Mutante (Tab. S2). Für die erste Saatgutmischung (M1) wurden 5 verschiedene Mutanten Samen zu 0,1% in 3000 Körner kommerzielles Saatgut beigemischt, für die zweite Saatgutmischung (M2), sowie eine Einzelkornanalyse (siehe Sachstandsbericht im Dez. 2022, Abb. 3) nur noch die homozygot mutierten Samen der Pflanze #2. Das Saatgut der für Transgen-frei gehaltenen Pflanze #18 (Abb. S7) wurde für eine zweite Populationsanalyse mit mehr Individuen und dem Versuch einer Identifikation durch zusätzlichen Nachweis eines gekoppelten Markers verwendet (Tab. S3).

3.1.2.3 Detektion von CRISPR/Cas Mutationen in DNA-Mischproben

Das Hauptziel des Projekts ist die Entwicklung eines Nachweisverfahrens mit einer 0,1%igen Nachweisgrenze der GE-Mutationen in Saatgut- oder Futtermittelmischungen. Dazu wurde zunächst 5% DNA der homozygoten C3E3 Mutante in 95% Raps-WT DNA gemischt und davon 10% in 90% DNA anderer Spezies (Tab. 5).

Tab. 5: Nachweis von 5% mutierter Raps DNA in DNA anderer Spezies

Spezies	Brutto Anzahl Seq.	Bezugs-system Netto-Seq.	Verluste [%]	Nicht analysierbar				Wildtyp			Mutante +T		
				zugeordnet	[%]	nicht zuzuordnen	[%]	Cluster	Seq	[%] Bs	Cluster	Seq	[%] Bs
Tomate	106154	106095	0,06	105220	0,82	101	0,1	228	100828	95,83	19	4381	4,16
Tomate	106154	106095	0,06	105220	0,82	101	0,1	228	100828	95,83	19	4381	4,16
Soja	108530	108460	0,06	107659	0,74	49	0,05	244	103212	95,87	18	4441	4,13
Soja	98358	98271	0,09	97574	0,71	51	0,05	227	93596	95,92	19	3967	4,07
Reis	101942	101860	0,08	101091	0,75	62	0,06	209	96853	95,81	23	4230	4,18
Reis	101358	101252	0,1	100516	0,73	48	0,05	223	96062	95,57	17	4454	4,43
Mais	94970	94891	0,08	94181	0,75	47	0,05	194	90529	96,12	19	3652	3,88
Mais	106003	105938	0,06	105148	0,75	66	0,06	236	100518	95,6	21	4622	4,4
Weizen	108195	108109	0,08	107186	0,85	139	0,13	223	102878	95,98	18	4294	4,01
Weizen	108195	108109	0,08	107186	0,85	139	0,13	223	102878	95,98	18	4294	4,01
Zuckerrübe	109049	108967	0,08	108120	0,78	57	0,05	228	103553	95,78	17	4555	4,21
Zuckerrübe	125167	125057	0,09	124236	0,66	96	0,08	279	119623	96,29	22	4601	3,7

Der Abgleich der DNA-Konzentrationen für die Mischverhältnisse erfolgte *via* qPCR mittels SybrGreen. Durchschnittliche Schwankungen bei der qPCR liegen im Bereich von $\pm 19\% = 0,25$ Ct (laut Planton max. 25% erlaubt) und wurden für dieses Experiment mit durchschnittlich 13% bestimmt, sodass die gemessenen Werte noch um diesen Faktor korrigiert werden müssen. Daraus folgt, dass $\emptyset 4,11\% + 0,51\%$ (13% Korrekturwert) = 4,62% ergeben, welche von 5% eingesetzter Raps DNA wiedergefunden wurden, sodass es kein Problem mit einer eventuellen Kreuzreaktivität gibt. Damit sind die Daten repräsentativ für Futtermittelmischungen und die NGS-Primer #5 amplifizieren nur im spezifischen Rapsanteil der Mischung, denn sonst wären WT-Reads weit über 95% zu erwarten gewesen.

In einem nächsten Schritt wurden Blatt DNA-Mischungen verwendet, um die Linearität und Konzentrationsabhängigkeit des Assays #5 zu bestimmen. Der Fokus liegt wieder auf der homozygoten Mutante C3E3 mit der +1 bp Mutation. Für diese Grundvalidierung wurden insgesamt 12 Verdünnungsstufen (mit mindestens drei und maximal 6 Messungen) eingesetzt. Im relevanten Bereich von 0% bis 1% wird deutlich, dass die wichtige 0,1% LOD im linearen Bereich liegt (Abb. 8).

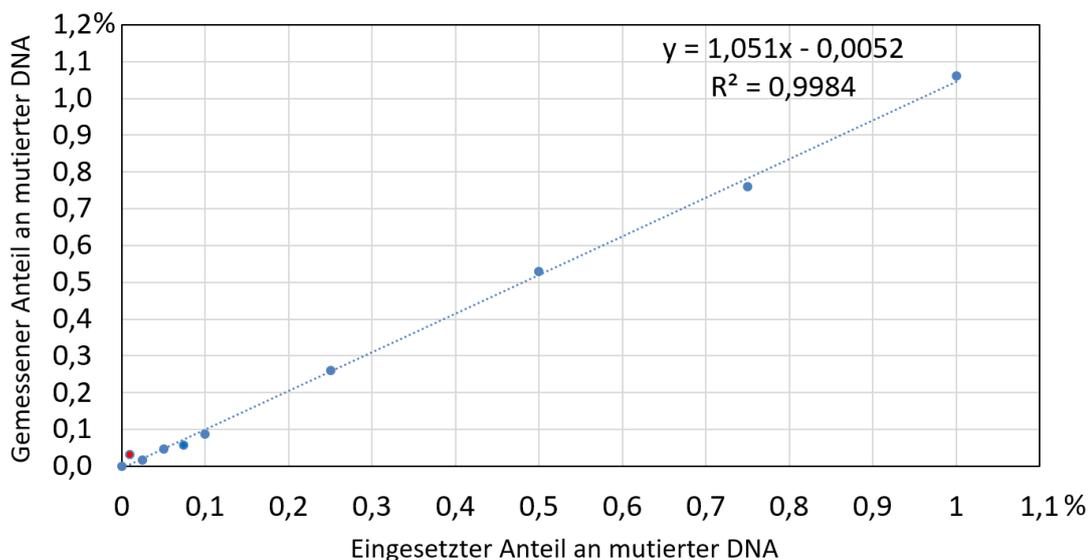


Abb. 8: Soll-Ist Vergleich der prozentualen Beimischungen mutierter DNA zu WT-DNA. Die Abbildung stellt den relevanten Bereich <1% im Detail dar. Der rot hervorgehobene Messwert bei 0,01% zeigt wie der bei 0,075 eine stochastische Abweichung von der Geraden, allerdings unterhalb der angestrebten 0,1% LOD.

Der leicht erhöhte Wert bei 0,01% (rot hervorgehoben) entspricht in etwa der zuvor beobachteten Rate an falsch-positivem Hintergrund des Assays durch technisch bedingtes Index-Hopping. Damit ist der Assay generell geeignet um als nächstes mit der Analyse von Saatgut-Mischungen zu beginnen.

3.1.2.4 Detektion von CRISPR/Cas Mutationen in Raps Saatgut-Mischungen

Bisher wurde mit DNA aus Blattmaterial gearbeitet und nun liegt der Fokus auf Raps-Samen und Saatgut-Mischungen (vergleichbar 3 Mutanten-Körner in 2997 Wildtyp-Körnern der Sorte Avatar, entsprechend der 0,1% LOD). Da die Körner der Mutanten kleiner sind als die der kommerziellen Sorte Avatar, welche wir für die Saatgutmischungen verwendet haben, musste die Einstellung des Mischverhältnisses für 0,1% Mutantenanteil über das Gewicht erfolgen (Masse/Masse). In einem ersten Versuch wurden 5 Saatgutmischungen erstellt (M1), die DNA isoliert und daraus jeweils 5 Amplikon-Sequenzierungen mit short-read Illumina Technik durchgeführt (Abb. 9). Man darf bei diesen Analysen nie vergessen, dass beide On-Target Loci (A09 und C09) zu gleichen Anteilen amplifiziert werden, sodass z. B. Werte für die Analyse des A09 Locus theoretisch mit zwei multipliziert werden müssten, oder die Werte der beiden Loci zu addieren sind um abschätzen zu können wie hoch die Wiederfindungsrate ist.

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

Abb. 9: Analyse der Saatgutmischung M1 mit homo- und heterozygotem T3-Saatgut (vgl. auch Abb. 7). Da Wildtyp-Reads der Mutanten in der großen Menge derer aus der Sorte Avatar untergehen kann für dieses WT-Allel kein Prozentwert bestimmt werden (blau). In zwei Fällen (gelb) wurden wieder falsch-positive 1 bp Deletionen detektiert, wiederum in einer Größenordnung von 0,01% (entspricht dem zuvor beobachteten „Hintergrund“ in DNA-Mischungen aus Blättern). Ansonsten entsprechen die Ergebnisse den Erwartungen (grün). Für eine zweite Mischung (M2) wurde nur Saatgut der rot umkreisten Mutante (#2) verwendet und das Durchschnittsergebnis von 273 Reaktionen gezeigt.

Interessant ist vor allem die fünfte Mischung mit heterozygotem Saatgut, da hier die Anteile der einzelnen Allele so wie erwartet unter denen der reinerbig vorliegenden Allele liegen, aber deutlich über dem falsch-positivem Hintergrund (gelb) was zeigt das auch kleinste Dosiseffekte detektiert werden können. Um aber den Nachweis nicht unnötig zu verkomplizieren wurde entschieden sich

für die Hauptanalyse auf die doppelt-homozygote Mutante (#2) zu fokussieren, deren Saatgut zuvor durch Einzelkornanalysen verifiziert wurde (siehe Sachstandsbericht im Dez. 2022, Abb. 3).

Für einem zweiten Versuch (M2) wurden 10 Saatgutmischungen zu je 0,1% mit der doppelt-homozygoten Mutante #2 (A09: +1/+1; C09: -1/-1) erstellt, sowie eine Mischung zu 0,5% und eine zu 2%. Nach der Homogenisierung erfolgten aus jeder Probe 4 DNA-Extraktionen und von jeder Extraktion wurden 5 oder 10 Illumina NGS Reaktionen durchgeführt. Insgesamt konnten 273 Reaktionen ausgewertet werden mit dem Ergebnis, dass bei einem zu erwartenden Verhältnis von 0,05% +1 bp (A09) zu 0,05% -1 bp (C09) nur geringe Abweichungen beobachtet werden konnten: 0,04% +1 bp (A09) zu 0,06% -1 bp (C09).

Auch dieser Assay zeigte sich linear (Abb. S9) und damit wurde die LOD von 0,1% erreicht.

3.1.2.5 Ergebnisse der zweiten Populationsanalyse zur Identifikation

Die zweite Populationsanalyse mit 100 Individuen erfolgte an einem NextSeq Gerät der Firma Illumina mit dem Assay #5 zum Nachweis der Mutation und einem zweiten Assay, um die Region um den mit der -3 bp gekoppelten SNP (Tab. 1) zu analysieren. Dieses Gerät besitzt eine größere Kapazität und ist in der Lage beide Assays parallel und mit einer höheren Anzahl an Individuen durchzuführen (Abb. 10).

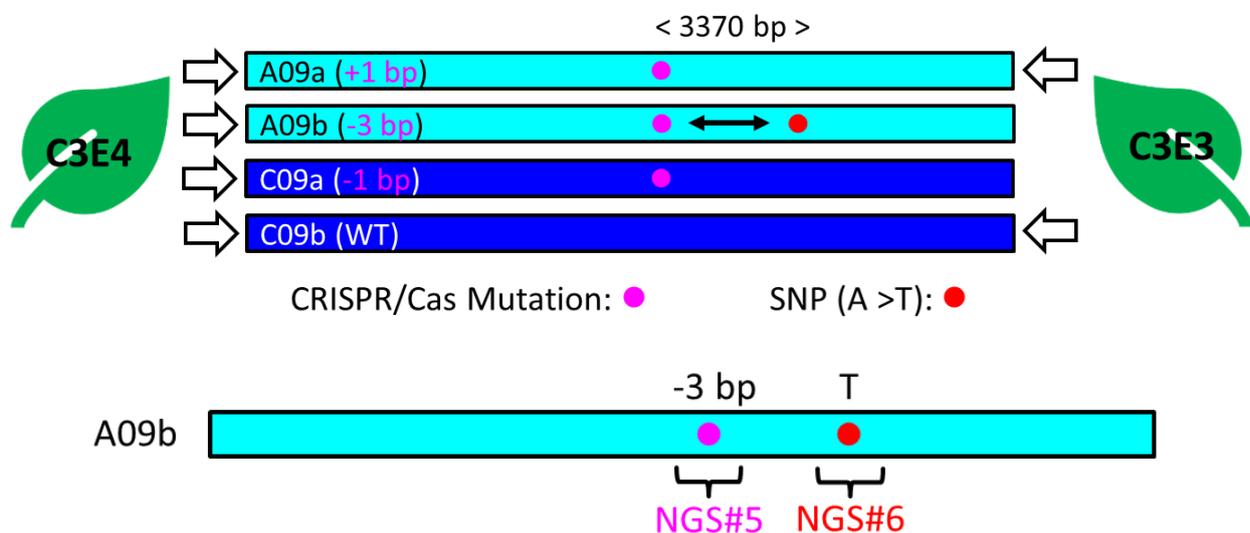


Abb. 10: Versuch einer Identifikation durch Kombination zweier short-read Amplikon-Tiefensequenzierungen mittels Illumina Technologie. Oben sind alle 4 möglichen Allele der On-Target Loci A09 und C09 gezeigt, von denen das A09b Allel eine Kopplung zwischen Mutation und Marker (SNP) aufweist. Dieses Allel ist nochmal separat hervorgehoben, um die Sequenzierstrategie durch 2 unabhängige Assays zu verdeutlichen. Leider ist die Entfernung zwischen beiden Punkten mit 3370 bp zu groß, um durch Illumina Technologie überbrückt werden zu können, weshalb kein Beweis einer physischen Kopplung wie bei long-read PacBio (Abb. 5) erbracht werden kann. Daher kann nur durch die Analyse einer spaltenden Population eine statistische Absicherung der Kopplung versucht werden.

Der Wechsel von der MiSeq zur NextSeq-Plattform und die parallele Analyse eines zweiten Assays erfolgten ohne Probleme, was für die Robustheit des etablierten Assays spricht. Die Ergebnisse beider Assays sind im Detail im Anhang zu finden (Tab. S3). Generell gab es durch die größere Anzahl untersuchter Individuen eine Verbesserung in der statistischen Auswertung mittels eines Chi-Quadrat Tests (vorher: $\chi^2 = 4,833$ bei $N=48$ und 8 dF), welcher durch eine Kombination nun ein geringeres $\chi^2 = 3,695$ bei $N=144$ und 8 dF aufweist (Abb. 11), was belegt dass die beobachteten Abweichungen von den Erwartungswerten zufällig (= nicht signifikant) sind. Bei Assay #6 zum SNP-Nachweis konnte eine Reaktion nicht ausgewertet werden, aber bei den verbliebenen 99 Reaktion korrelierte die Readmenge des SNPs immer mit der Anzahl an Reads der -3 bp Mutante, womit die Kopplung im Prinzip bestätigt werden konnte.

T2 Gamet ♂ ♀	A09a C09a +1 WT	A09a C09b +1 -1	A09b C09b -3 -1	A09b C09a -3 WT
A09a C09a +1 WT	A +1/+1 C WT/WT	A +1/+1 C WT/-1	A +1/-3 C WT/-1	A +1/-3 C WT/WT
A09a C09b +1 -1	A +1/+1 C WT/-1	A +1/+1 C -1/-1	A +1/-3 C -1/-1	A +1/-3 C WT/-1
A09b C09b -3 -1	A +1/-3 C WT/-1	A +1/-3 C -1/-1	A -3/-3 C -1/-1	A -3/-3 C WT/-1
A09b C09a -3 WT	A +1/-3 C WT/WT	A +1/-3 C WT/-1	A -3/-3 C WT/-1	A -3/-3 C WT/WT

25% wie Eltern

homozygot

Exp 1: C3E4 (T2) Nachkommen

Exp 2: C3E4 #18 (T3) Nachkommen

- ♦ 16 Kombinationen möglich
- ♦ Exp 1: 3 „Replikate“ = 48 T3 Pflanzen
- ♦ Exp 2: 6 „Replikate“ = 96 T4 Pflanzen
- ♦ total: 9 „Replikate“ = 144 Pflanzen

A09 locus C09 locus	+1/+1 WT/WT	+1/+1 -1/WT	+1/-3 -1/WT	+1/-3 WT/WT	+1/+1 -1/-1	+1/-3 -1/-1	-3/-3 -1/-1	-3/-3 -1/WT	-3/-3 WT/WT	Chi-Square
Erwartet	9	18	36	18	9	18	9	18	9	∑ 144
Beobachtet	9	19	41	17	5	15	8	19	11	dF: 8
χ ²	0.000	0.056	0.694	0.056	1.778	0.500	0.111	0.056	0.444	∑ 3.695

Abb. 11: Statistische Auswertung der kombinierten Populationsanalysen unter Verwendung eines Chi-Quadrat Tests zum Vergleich von Beobachtung und Erwartung.

3.1.2.6 Bewertung der short-read NGS-Technologie / Nachweis einer bekannten Mutation

Der Vorteil eines einmal etablierten NGS Assays liegt vor allem in seiner Aussagekraft und Flexibilität unterschiedliche Mutationen in der gleichen Genregion detektieren zu können, was nur durch die Erzeugung von Sequenzdaten möglich ist. Durch den Hochdurchsatz in einem Illumina NGS-Lauf werden alle Allele in ausreichender Abdeckung auf einmal untersucht, was bei der Sanger Amplikon-Sequenzierung problematisch ist (da hier bei unbekanntem Mutationen die Anzahl der benötigten Klone für die Sequenzierung nur schwer abzuschätzen ist). Trotzdem bietet die Sanger-Sequenzierung auch einen Vorteil, da Amplikons bis zu 1 kb kein Problem darstellen und man so auch sehr große Insertionen (z. B. +117 bp) und Deletionen (z. B. -67 bp) erkennen kann. Dies wäre bei einem *ab initio* Ansatz mit Illumina Technik schwierig, da die Amplikongröße hier auf ca. 300 bp beschränkt ist. Kleinere InDels und Punktmutationen (SNPs) stellen allerdings kein Problem dar und ein einmal etablierter Assay sollte auch gut geeignet sein ähnliche SDN-1 oder SDN-2 Mutationen in anderen Elite-Zuchtlinien nachzuweisen, wenn diese dort an gleicher Stelle eingeführt wurden. So konnten bei unserer Untersuchung der C3E4 Mutante insgesamt 3 CRISPR/Cas9-induzierte Mutationen plus ein WT Allel, alle anteilig (zu jeweils ca. 25% aller erzeugten Sequenzen) detektiert werden. Das half vor allem bei der Analyse einer spaltenden Population (aus der C3E4 Mutante), wodurch Falsch-Positive Sequenzen eindeutig identifiziert werden konnten – nämlich immer dann, wenn die untersuchten Allele homozygot vorlagen und Signale für die übrigen Allele nicht zu erwarten waren. In unserem Assay auf dem MiSeq-Gerät fanden sich Falsch-Positive Sequenzen zu ca. 0,016% und auf dem NextSeq nur zu ca. 0,009%. Auch wenn es bei allopolyploiden Pflanzen wie dem Raps (auch Baumwolle, Tabak, Erdnuss, Quinoa, Maniok, Triticale, Emmer oder Hartweizen) notwendig werden sollte die Mutationen einem der Subgenome zuzuordnen sind Sequenzdaten aussagekräftiger als andere PCR-basierte Verfahren, da hier auch natürliche SNPs und InDels im Amplikon erkannt werden und anhand derer eine Differenzierung zwischen den unterschiedlichen Loci möglich wird. In einem solchen Fall dürften jedoch nur die größten Cluster in die Auswertung einbezogen werden (ca. 75% aller erzeugten Sequenzen), weil auch bei dieser Methode PCR-Zwischenschritte (Zielgen-spezifische PCR und Barcoding PCR) zum Einsatz kommen, was bei sehr ähnlichen Genkopien immer auch zu Chimärenbildung führen wird (Abb. S3). In unserem Fall durchschnittlich ca. 24,21% Chimären in der Populationsanalyse mit dem MiSeq

(N=48; Tab. S2) und 27,67% Chimären in der Populationsanalyse mit dem NextSeq (N=100; Tab. S3), sodass man gut abwägen muss wie stringent man nach der Mutation sucht damit z. B. wenigstens 40.000 auswertbare Reads für die Bestimmung der 0.1% LOD an Saatgutmischungen zur Verfügung stehen. In beiden Populationsanalysen mit Blatt-DNA von unseren Raps Einzelpflanzen in denen nach der kompletten Amplikon-Sequenz gesucht wurde (und so alle Chimären und PCR-Artefakte herausgefiltert wurden) gab es aber in jedem Fall ausreichend Reads für die Analyse des Mutationsorts (\approx 68415 Reads auf dem MiSeq und \approx 192065 Reads auf dem NextSeq) und ebenfalls wenn die Raps-DNA zuvor mit den DNAs anderer Spezies vermischt wurde. Bei unserem Assay liegt die gesetzlich vorgeschriebene Nachweisgrenze von 0,1% nicht nur im linearen Bereich, sondern auch knapp eine 10er-Potenz über dem Hintergrund möglicher Falsch-Positiver Signale. Da alle parallelen DNA-Analysen aus individuellen Pflanzen ein Barcoding verlangen, könnten hier die Falsch-Positiven Werte auf Index-Hopping zurückzuführen sein. Bei dem NextSeq-Ansatz wurden allerdings auch wesentlich mehr Sequenzen generiert, sodass selbst bei stringenter Filterung die Anzahl Falsch-Positiver Sequenzen prozentual geringer ausfiel als am MiSeq und in 33% aller Samples mit Homozygotie-Anteil gar keine Falsch-Positiven gefunden wurden (nur 1 Sample am MiSeq). Damit ist die short-read Amplikon-Tiefensequenzierung mit Illumina Technologie (NextSeq besser als MiSeq) ideal für die Detektion bekannter Mutationen in einem genau definierten Genabschnitt (vgl. auch Fraiture et al., 2023).

3.1.2.7 Bewertung der Event-Identifizierung / Rückführbarkeit der Mutation auf Genom-Editing

Da SDN-1 oder SDN-2 induzierte Mutationen in gleicher Form auch natürlich auftreten können, ist es notwendig ihren Ursprung auf diese Techniken zurückführen zu können, damit eine Regulation überhaupt stattfinden kann. Wir konnten bereits für die *CRT1a* Loci zeigen, dass es im Raps Pan-Genom natürliche InDels gibt, welche 3 bp neben einer Cas9-typischen PAM (NGG) liegen (Abb. S1, unten), sodass dies keinen hinreichenden Beweis für ein GE Event darstellt – zumal es inzwischen auch viele CRISPR/Cas System gibt, welche ganz ohne PAM funktionieren (Ren et al., 2021). Damit die eindeutige Identifizierung eines Events möglich wird, müsste wenigstens ein eng gekoppelter Marker, welcher sonst im zugehörigen Pan-Genom nicht vorkommt, (durch WGS) identifiziert werden. In unserem Fall gab es (durch großes Glück) einen SNP ca. 3,5 kb neben der A09b -3 bp Mutation. Bei einer Rekombinationsrate von 0,01% ist es zwar recht unwahrscheinlich, dass die Kopplung durch ein Crossing-over gebrochen wird, aber eben doch nicht auszuschließen. Statistisch gesehen sollte in einer von 10.000 Pflanzen ein solcher Kopplungsbruch erfolgen, d. h. man könnte auf einem 1 ha Rapsfeld ca. 5 solche Crossing-over erwarten. Wir haben in einer Korrelationsanalyse 100 Pflanzen untersucht, um zu schauen, ob die -3 bp Mutation auf A09b immer mit dem SNP („T“) zusammenvererbt wird. 99 Sequenzen des SNP-Assays waren auswertbar und sowohl Mutation als auch SNP korrelierten in den erwarteten Verhältnissen. Allerdings ist es durchaus möglich, dass bei einer höheren Fallzahl (> 10.000) doch Kopplungsbrüche auftreten, was einen gerichtsfesten Nachweis erschweren würde.

3.2 Voraussichtlicher Nutzen und Verwertbarkeit der Ergebnisse

Laut Zuwendungsbescheid ist ein Verwertungsplan nicht erforderlich. Da die hier untersuchten Mutationen allerdings in einer ökonomisch völlig uninteressanten Sommer-Rapssorte (Mozart) erzeugt wurden und eine Identifizierung nur mit einer genauso uninteressanten Mutation (die -3 bp auf A09b bewirkt keinen Resistenzeffekt gegen den Schadpilz; Abb. S4) erfolgen kann, ist noch kein unmittelbarer Nutzen für das SOP abzusehen. Dazu müsste zunächst in der A09-Kopie des *CRT1a* Gens einer kommerziell bedeutenden Winter-Rapssorte eine Frame-shift Mutation an ähnlicher Stelle induziert werden (am besten DNA-frei unter Verwendung von RNPs), der Transgen-Status durch WGS bestimmt und GE-Mutationen in der EU schließlich genau wie die Mutationen durch Strahlung oder chemische Behandlung von der Regulierung ausgenommen werden, damit ein erster Resistenz-Test unter Feldbedingungen möglich wird.

4. Zusammenfassung

Ziel des Projekts ist es festzustellen, ob ein im Rahmen der geltenden Zulassungsrichtlinien für GVOs und daraus hergestellter Produkte validiertes Nachweis- und Identifizierungsverfahren für GE-Pflanzen mit entsprechendem Referenzmaterial zur Rückverfolgbarkeit und Marktkontrolle verfügbar gemacht werden kann. In diesem Teil der Machbarkeitsstudie geht es um die Bewertung der **short-read Amplikon-Tiefensequenzierung mittels Illumina-Technologie** hinsichtlich ihrer Eignung bekannte CRISPR/Cas9-induzierte Mutationen im Raps *CRT1a*-Gen nachzuweisen und falls möglich auf die verwendete Neue Mutagenese-Technik rückzuschließen (Event Identifikation). Im Raps gibt es insgesamt 4 *CRT1a* Gen-Loci, zwei auf den Chromosomen A01 und C01 (Off-Target Loci), sowie zwei auf A09 und C09 (On-Target Loci). Für dieses Projekt wurde das Genom der Elternlinie Mozart, sowie die beiden Mutanten C3E3 (A09a: +1 bp/A09b: +1 bp; C09a: WT/C09b: WT) und C3E4 (A09a: +1 bp/A09b: -3 bp; C09a: -1 bp/C09b: WT) sequenziert. Durch eine vorläufige Auswertung des PacBio WGS mit einer durchschnittlichen Read-Länge von ca. 15 kb und 45 GB je erzeugtem Datensatz konnten alle drei bekannten Mutationen in den *CRT1a* Loci gefunden werden. Von diesen liegt nur die züchterisch nicht relevante -3 bp Mutation (auf Chromosom A09b) mit 3070 bp nahe genug an einem für dieses Event spezifischen Marker (der SNP T statt A ist weder im Mozart WT noch im restlichen Raps Pan-Genom zu finden). Nur mit Hilfe solch eines eng gekoppelten Markers kann versucht werden durch Kombination zweier Assays den Ursprung der Mutation auf die NMT zurückzuführen (Abb. 10). Allerdings gilt es zu bedenken, dass es durch Crossing-over leicht zu einem Kopplungsbruch kommen kann, sodass ein gerichtsfester Nachweis nicht ohne weiteres garantiert werden kann.

Durch das WGS konnten keine Off-Target Mutationen gefunden werden, dafür aber auf 6 unterschiedlichen Chromosomen T-DNA Fragmente des Konstrukts mit der Genschere, welches zur Erzeugung der beiden Raps-Mutanten während der Gewebekultur eingesetzt wurde (Abb. S6). Damit sind sowohl die hier analysierte homozygote C3E3-Mutante mit fünf T-DNA Fragmenten, als auch die heterozygote C3E4-Mutante mit nur einem verbliebenen T-DNA Fragment nicht als Transgen-frei einzustufen (Abb. 3, Tab. 2). Sollten GE-Mutationen einmal von der Regulierung ausgenommen werden, scheint es daher angezeigt, dass NMT-Pflanzen welche z.B. mit Hilfe von Agrobakterien erzeugt wurden, vor einem Freilandversuch mittels eines long-read WGS zunächst genauer untersucht werden – zum Beweis, dass diese auch tatsächlich Transgen-frei sind.

Mit Hilfe einer *CRT1a* Referenzkarte (Abb. S1) wurden für das Illumina-NGS zunächst Primer erstellt, welche alle 4 *CRT1a* Loci im Bereich der Mutationsstelle amplifizieren können, womit gezeigt werden konnte, dass es nur in den beiden On-Target Loci zu Mutationen gekommen ist und dass alle bekannten Mutationen auch in den erwarteten Verhältnissen nachgewiesen werden konnten (Abb. S2). Mit neuen, auf die On-Target Zielkopien optimierten Primern wurden dann verschiedene Raps Linien untersucht, Populationsanalysen durchgeführt (Tab. S2+S3), DNA-Verdünnungsreihen und Mischungen getestet und schließlich die LOD an Saatgut-Mischungen überprüft (Abb. 9). Der Assay zeigt keine Kreuzreaktionen mit DNAs anderer Spezies, verhält sich linear und Konzentrationsabhängig und besitzt auch die notwendige Sensitivität, um die gesetzliche Nachweisgrenze von 0,1% einhalten zu können, womit die wichtigsten Parameter für die Erstellung eines SOPs eingehalten wurden.

Damit ist es durch diese Technik möglich eine bekannte Mutation nachzuweisen und durch die Erzeugung von Sequenzdaten auch eindeutig einem Gen-Locus/Subgenom zuzuordnen. Das hier entwickelte SOP erlaubt somit den Nachweis einer durch NMT erzeugten Mutation und stellt einen wichtigen Schritt in Richtung Identifizierung dar. Eine Identifizierung ist allerdings wohl nur in Ausnahmefällen möglich, denn dazu bedarf es eines eng mit der Mutation gekoppelten zweiten Markers.

5. Gegenüberstellung der ursprünglich geplanten zu den tatsächlich erreichten Zielen

Anpassung der WGS-Sequenzierstrategie

Die Sequenzierung der geplanten Kontrolllinie Express 617 wurde durch eine Publikation der Genomsequenz überflüssig (Lee et al., 2020), sodass man darauf verzichten und stattdessen eine zweite Mutante (C3E4) sequenzieren konnte. Der angenommene Vorteil, dass diese zweite Mutante inzwischen Transgen-frei sein sollte (Abb. 3) erwies sich leider als falsch. Damit wäre es theoretisch auch möglich in beiden GE-Mutanten den Transgen-Anteil wie bei klassischen GVOs mittels PCR nachzuweisen. Dafür wurden durch diese Mutante Analysen an spaltenden Populationen möglich, z. B. mit Aussagen zur Differenzierung zwischen den Raps-Subgenomen oder der Korrelation von -3 bp Mutation und Event-spezifischem SNP für die Identifizierung.

Planung: short-read Illumina WGS durch IPK	Anpassung: long-read PacBio WGS
C3E3 Mutante (A-Genom-Kopie vollständig und homozygot mutiert, C-Genom-Kopie WT)	C3E3 (eindeutig transgen) bei Genewiz
Express 617 als Kontrolllinie	C3E4 (heterozygot mutiert) bei Genewiz
Mozart als Elternlinie	8 gepoolte Mozart-Pflanzen über das BVL

5.1 Vergleich der Arbeitspakete mit dem Bearbeitungsstand zum Projektende

Ziele laut Arbeitsplan	Stand
<ul style="list-style-type: none"> Erstellung einer Referenzkarte aller CRT1a Sequenzen aus dem Raps Pan-Genom Multiples Sequenz-Alignment and -Analysen mit Mozart Re-Sequenzierungs-Daten Ausschluss natürlicher Polymorphismen an den Mutationsstellen im Pan-Genom Primer-Design und Validierung für einen Illumina-NGS Assay zum Mutationsnachweis Durchführung der Amplikon-Tiefensequenzierung / Überprüfung der PCR- Amplifikate Assay funktioniert in unterschiedlich verdünnten und gemischten DNA-Proben Durchführung der NGS-Analysen und der bioinformatischen Datenanalysen 	Ziel erreicht Ziel erreicht Ziel erreicht Ziel erreicht Ziel erreicht Ziel erreicht Ziel erreicht
<ul style="list-style-type: none"> Durchführung der CRT1a Mozart Re-Sequenzierung durch Sanger mittels Klonierung Durchführung des PacBio-WGS Vergleich der Sanger-Sequenzen mit den PacBio-WGS Daten und dem Pan-Genom Suche nach GE-Mutationen, Off-Targets und T-DNA Resten in den Mutanten Ergänzendes Chromosome Conformation Capture Sequencing (Hi-C Sequenzierung) Datenbearbeitung/Auswertung des Pac-Bio WGS am IPK. 	Ziel erreicht Ziel erreicht Ziel erreicht Ziel erreicht laufend laufend
<ul style="list-style-type: none"> Detektion Event-spezifischer SNPs/InDels in den WGS-Daten für die Identifizierung Etablierung eines zweiten Assays für die Identifizierung Identifizierung einer GE-Mutation in einer Population aus Einzelpflanzen 	Ziel erreicht Ziel erreicht Ziel erreicht
<ul style="list-style-type: none"> Bestimmung der Spezifität und Sensitivität des NGS-Verfahrens an Saatgutmischungen Erstellung eines Praxis-entsprechenden SOPs zum GE-Nachweis in Raps, bzw. Gerste 	Ziel erreicht Ziel erreicht
<ul style="list-style-type: none"> Erstellung von Manuskripten zur Publikation 	laufend

Das PacBio-WGS ist abgeschlossen und es liegen bereits erste Erkenntnisse vor. Für ein vollständiges Assembly der WGS-Daten muss nun am IPK noch eine Hi-C Sequenzierung durchgeführt werden.

5.2 Hinweise auf weiterführende Fragestellungen

5.2.1 SOP

Für ein amtlich anerkanntes und validiertes SOP müssten noch Ringversuche mit weiteren akkreditierten Laboren erfolgen.

5.2.2 Identifizierung

Für die eindeutige Identifizierung eines Events benötigt man eine Assay-Kombination zum Nachweis der induzierten Mutation und eines eng mit dieser gekoppelten Markers. Die gemeinsame Vererbung konnte für eine kleine Population von 100 Einzelpflanzen durch eine Korrelationsanalyse bestätigt werden (Tab. S3). Allerdings ist dies noch kein Beweis einer physischen Kopplung und auch ein Test des zweiten Assays für die Detektion des Markers in Mischproben mit 0,1% und 0,9% Anteil der Mutante steht noch aus.

5.2.3 Feldversuche

Ohne Feldversuche unter natürlichen Bedingungen mit ausreichend vielen Pflanzen wird das CRT1a Event wohl in kein Zuchtprogramm aufgenommen werden, denn die im Gewächshaus beobachtete Resistenz muss zuerst einmal unter natürlichen Bedingungen reproduziert werden. Für die Resistenz gegen den Schadpilz ist eine Frame-shift Mutation im *CRT1a* A09 Locus (vgl. Abb. S4) erforderlich. Diese müsste jedoch zunächst in einer kommerziell relevanten Winter-Raps Sorte „nachgebaut“ werden, für die es dann aber wahrscheinlich keinen gekoppelten Marker geben wird und damit auch keine Identifikationsmöglichkeit. Das SOP könnte jedoch auch nur für den einfachen Nachweis der Mutation genutzt werden, um z.B. in parallelen Untersuchungen die eventuelle Ausbreitung von GE-Mutationen in angrenzende Streifen mit WT-Raps zu untersuchen.

5.2.4 Transgen-Status

Die Verwendung von Agrobakterien zur stabilen Transformation unserer Kulturpflanzen bedingt immer auch die Integration fremder T-DNA in das pflanzliche Genom. Theoretisch ist ein Herauskreuzen/Segregieren dieser Fremd-DNA zwar möglich, aber durch unser WGS konnten wir zeigen, dass dies sehr schwierig werden kann, wenn es zur Integration an multiplen Stellen im Genom kommt und die T-DNA dabei auch noch fragmentiert / neu arrangiert wird (Abb. S5). Daher sollten GE-Pflanzen durch ein WGS untersucht werden, um zu belegen, dass keine Konstrukt-Reste mehr vorhanden sind. DNA-freie Ansätze mit der Genschere in Proteinform, welche zusammen mit der sgRNA (RNPs) eingeschleust werden sind zum gegenwärtigen Zeitpunkt zwar noch nicht effizient zu nutzen, könnten aber dieses Problem lösen. Eine nähere Untersuchung der T-DNA Fragmentierung scheint allerdings nicht notwendig, da eine Publikation von Pucker et al. (2021) zeigt, dass es auch in Arabidopsis multiple Insertionen von T-DNA Bruchstücken zu beobachten gibt.

6. Literaturverzeichnis

- Bessoltane N., Charlot F., Guyon-Debast A., Charif D., Mara K., Collonnier C., Perroud P.F., Tepfer M. and Nogué F. (2022). Genome-wide specificity of plant genome editing by both CRISPR-Cas9 and TALEN. *Scientific reports*, 12(1), 9330.
- Bundesamt für Verbraucherschutz und Lebensmittelsicherheit, „Stellungnahme der ZKBS zu neuen Techniken für die Pflanzenzüchtung“ (2012).
- Bundesamt für Verbraucherschutz und Lebensmittelsicherheit, „Neue Nachweismethode verspricht spezifische Detektion genomeditierter Rapslinien – was kann das Verfahren tatsächlich leisten?“ (2020).
- Bundesamt für Naturschutz, „Analyse von Nachweismethoden für genomeditierte und klassische GV-Pflanzen“, BfN-Skripten 622 (2022).
- Cai G., Yang Q., Yi B., Fan C., Edwards D., Batley J. and Zhou Y. (2014). A complex recombination pattern in the genome of allotetraploid *Brassica napus* as revealed by a high-density genetic map. *PloS one*, 9(10), e109910.
- Chalhoub B., ... and Wincker P. (2014). Plant genetics. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* (New York, N.Y.), 345(6199), 950–953.
- Chen X., Tong C., Zhang X., Song A., Hu M., Dong W., Chen F., Wang Y., Tu J., Liu S., Tang H. and Zhang L. (2021). A high-quality *Brassica napus* genome reveals expansion of transposable elements, subgenome evolution and disease resistance. *Plant biotechnology journal*, 19(3), 615–630.
- Chhalliyil P., Ilves H., Kazakov S.A., Howard S.J., Johnston B.H. and Fagan J. (2020). A Real-Time Quantitative PCR Method Specific for Detection and Quantification of the First Commercialized Genome-Edited Plant. *Foods* (Basel, Switzerland), 9(9), 1245.
- Delourme R., ... and Pauquet J. (2013). High-density SNP-based genetic map development and linkage disequilibrium assessment in *Brassica napus* L. *BMC genomics*, 14, 120.
- European Network of GMO Laboratories (ENGL), “Detection of food and feed plant products obtained by new mutagenesis techniques”, 26 March 2019 (JRC116289).
- Fraiture M.A., D'aes J., Guiderdoni E., Meunier A.C., Delcourt T., Hoffman S., Vandermassen E., De Keersmaecker S.C.J., Vanneste K. and Roosens N.H.C. (2023). Targeted High-Throughput Sequencing Enables the Detection of Single Nucleotide Variations in CRISPR/Cas9 Gene-Edited Organisms. *Foods* (Basel, Switzerland), 12(3), 455.
- Giraut L., Falque M., Drouaud J., Pereira L., Martin O.C. and Mézard C. (2011). Genome-wide crossover distribution in *Arabidopsis thaliana* meiosis reveals sex-specific patterns along chromosomes. *PLoS genetics*, 7(11), e1002354.
- Gould F., Amasino R.M., Brossard D., Buell C.R., Dixon R.A., Falck-Zepeda J.B., Gallo M.A., Giller K.E., Glenna L.L., Griffin T., Magraw D., Mallory-Smith C., Pixley K.V., Ransom E P., Stelly D.M. and Stewart C. N., Jr (2022). Toward product-based regulation of crops. *Science* (New York, N.Y.), 377(6610), 1051–1053.
- Grohmann L., Keilwagen J., Duensing N., Dagand E., Hartung F., Wilhelm R., Bendiek J. and Sprink T. (2019) Detection and Identification of Genome Editing in Plants: Challenges and Opportunities. *Front. Plant Sci.* 10:236.
- 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 biotechnology journal*, 21(2), 331-341.
- Lee H., Chawla H.S., Obermeier C., Dreyer F., Abbadi A. and Snowdon R. (2020). Chromosome-Scale Assembly of Winter Oilseed Rape *Brassica napus*. *Frontiers in plant science*, 11, 496.
- Leopoldina - Nationale Akademie der Wissenschaften, Deutsche Forschungsgemeinschaft und Union der deutschen Akademien der Wissenschaften (2019) „Wege zu einer wissenschaftlich begründeten, differenzierten Regulierung genomeditierter Pflanzen in der EU“. ISBN: 978-3-8047-4064-8.
- Lloyd A., Blary A., Charif D., Charpentier C., Tran J., Balzergue S., Delannoy E., Rigaiil G. and Jenczewski E. (2018). Homoeologous exchanges cause extensive dosage-dependent gene expression changes in an allopolyploid crop. *The New phytologist*, 217(1), 367–377.
- Nicolas S.D., Le Mignon G., Eber F., Coriton O., Monod H., Clouet V., Huteau V., Lostanlen A., Delourme R., Chalhoub B., Ryder C.D., Chèvre A.M. and Jenczewski E. (2007). Homeologous recombination plays a major role in chromosome rearrangements that occur during meiosis of *Brassica napus* haploids. *Genetics*, 175(2), 487–503.

- Pixley K.V., Falck-Zepeda J.B., Paarlberg R.L., Phillips P.W.B., Slamet-Loedin I.H., Dhugga K.S., Campos H. and Gutterson N. (2022). Genome-edited crops for improved food security of smallholder farmers. *Nature genetics*, 54(4), 364–367.
- Pröbsting M., Schenke D., Hossain R., Häder C., Thureau T., Wighardt L., Schuster A., Zhou Z., Ye W., Rietz S., Leckband G. and Cai D. (2020). Loss of function of CRT1a (calreticulin) reduces plant susceptibility to *Verticillium longisporum* in both *Arabidopsis thaliana* and oilseed rape (*Brassica napus*). *Plant biotechnology journal*, 18(11), 2328–2344.
- 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.
- Ren Q., ... and Zhang Y. (2021) PAM-less plant genome editing using a CRISPR-SpRY toolbox. *Nature plants*, 7(1), 25–33.
- Ribarits A., Eckerstorfer M., Simon S. and Stepanek W. (2021). Genome-Edited Plants: Opportunities and Challenges for an Anticipatory Detection and Identification Framework. *Foods (Basel, Switzerland)*, 10(2), 430.
- Rousseau-Gueutin M., ... and Aury J.M. (2020). Long-read assembly of the *Brassica napus* reference genome Darmor-bzh. *GigaScience*, 9(12), g1aa137.
- Samans B., Chalhoub B. and Snowdon R.J. (2017). Surviving a Genome Collision: Genomic Signatures of Allopolyploidization in the Recent Crop Species *Brassica napus*. *The plant genome*, 10(3), 10.3835/plantgenome2017.02.0013.
- Singer T., Fan Y., Chang H.S., Zhu T., Hazen S.P. and Briggs S.P. (2006). A high-resolution map of *Arabidopsis* recombinant inbred lines by whole-genome exon array hybridization. *PLoS genetics*, 2(9), e144.
- Song J.M., Guan Z., Hu J., Guo C., Yang Z., Wang S., Liu D., Wang B., Lu S., Zhou R., Xie W.Z., Cheng Y., Zhang Y., Liu K., Yang Q.Y., Chen L.L. and Guo L. (2020). Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of *Brassica napus*. *Nature plants*, 6(1), 34–45.
- Sturme M.H.J., van der Berg J.P., Bouwman L.M.S., De Schrijver A., de Maagd R.A., Kleter G.A., and Battaglia-de Wilde E. (2022). Occurrence and Nature of Off-Target Modifications by CRISPR-Cas Genome Editing in Plants. *ACS agricultural science & technology*, 2(2), 192–201.
- Tiley G.P. and Burleigh G. (2015) The relationship of recombination rate, genome structure, and patterns of molecular evolution across angiosperms. *BMC Evolutionary Biology* 15:194.
- Waltz E. (2022). GABA-enriched tomato is first CRISPR-edited food to enter market. *Nature biotechnology*, 40(1), 9–11.
- Wang M., King G.J., Shi L., Li R., Zhang Y., Wang X., Meng J., Tu J. and Zou J. (2023) Genome-wide recombination variation in biparental segregating and reciprocal backcross populations provides information for introgression breeding in *Brassica napus*. *The Crop Journal* 11(1), 208-219.

PROF. DR. DAGUANG CAI
 LEHRSTUHL MOLEKULARE PHYTOPATHOLOGIE
 DER CHRISTIAN-ALBRECHTS-UNIVERSITÄT ZU KIEL
 HERMANN-RODEWALD-STRASSE 9 · 24118 KIEL

(Prof. Daguang Cai)

Kiel, 27. Juli 2023

Datum und Unterschrift des Projektverantwortlichen