

Gefördert durch:



aufgrund eines Beschlusses
des Deutschen Bundestages

Schlussbericht zum Thema

“Züchtung und Agronomie neuartiger, Vicin-armer Ackerbohnen und Einsatz als einheimisches Eiweißfutter“

FKZ: 2815EPS066

Projektnehmer: RLP AgroScience GmbH

Gefördert durch das Bundesministerium für Ernährung und Landwirtschaft auf Grund eines Beschlusses des Deutschen Bundestages im Rahmen der BMEL Eiweißpflanzenstrategie.

Abschlußbericht

Gefördert durch:



aufgrund eines Beschlusses
des Deutschen Bundestages

Inhaltsverzeichnis

Planung und Ablauf des Projektes	2
Wissenschaftlicher und technischer Stand, an den angeknüpft wurde	3
Material und Methoden	4
Darstellung der wichtigsten Ergebnisse und Diskussion	6
I. AboVici-Referenztranskriptom	7
II. SNP-Analysen und SNP-Kandidaten	7
Die Daten aus dem Pflanzenmaterial von 2019	8
Neubewertung der Daten aus dem Pflanzenmaterial von 2017	16
Synthese aus allen Datensätzen des Pflanzenmaterials von 2017 & 2019	16
Aspekte der Sequenzier-Methodik und Auswirkungen auf die Ergebnislisten	21
Zwischenfazit	24
Neubewertung des evaluierten KASP-Markersets aus 2017	24
Fazit	29
III. Differentielle Expression und Kandidatengene	29
Auswertung der Transkriptomdaten	29
Überprüfung der Expression mit quantitativer PCR	32
Die Schnittmenge aus SNP- und Expressionskandidaten	34
Voraussichtlicher Nutzen und Verwertbarkeit der Ergebnisse	37
Gegenüberstellung der ursprünglich geplanten zu den tatsächlich erreichten Zielen; Hinweise auf weiterführende Fragestellungen	38
Zusammenfassung	38
Literaturverzeichnis	39
Übersicht über alle im Berichtszeitraum vom Projektnehmer realisierten Veröffentlichungen zum Projekt, bisherige und geplante Aktivitäten zur Verbreitung der Ergebnisse	40
Anhang I - Finale Liste der SNP-Kandidaten, lokalisierbar auf Chr.2 von <i>M.truncatula</i>	41
Anhang II- Finale Liste der SNP-Kandidaten, nicht lokalisierbar auf dem Genom von <i>M.truncatula</i>	42
Anhang III - Kurz gefasster Erfolgskontrollbericht	43

Abkürzungsverzeichnis

AG	Arbeitsgruppe
LVC	Niedrig-Vicin/Convicin
HVC	Hoch-Vicin/Convicin
M.truncatula	<i>Medicago truncatula</i>
MACE	M assive A nalysis of c DNA E nds
NGS	Next-Generation Sequencing
VC	Vicin/Convicin
V.faba	<i>Vicia faba</i>
qPCR	q uantitative P olymerase C hain R eaction
SNP	S ingle N ucleotide P olymorphism

Gegenstand des Vorhabens, Ziele und Aufgabenstellung des Projekts

(Beteiligung lt. AboVici-Vorhabenbeschreibung an den „Genetischen Forschungsaufgaben“)

Der Gehalt an den anti-nutritiven Substanzen Vicin und Convicin (=“VC“) in reifen Samen ist eine monogen vererbte Eigenschaft der Ackerbohne (*Vici faba* L.). Eine effiziente Selektion, gestützt auf molekulare Marker, auf Vicin- und Convicin-Armut der Ackerbohne ist in der Züchtung derzeit noch nicht etabliert. Es gibt daher wenig VC-arme Sommer-Ackerbohnen und markfähige VC-arme Winterackerbohnen sind noch nicht verfügbar. Ziel war die Identifizierung des *V.faba*-Gens, welches für die VC-Armut ursächlich verantwortlich ist sowie die Entwicklung möglichst eng gekoppelter Marker um letztlich den Qualitätsaspekt „Vicin- und Convicin-Armut“ mit dem hohen Leistungspotential von Winterackerbohnen in der Züchtung kombinieren zu können. Damit sollen der Landwirtschaft zeitnah neue, attraktive *V.faba* Sorten zur Verfügung gestellt werden, die letztlich zu eine Ausweitung des Anbaus und der Nutzung der Ackerbohne in Deutschland führen werden. Damit passt das Projekt zum dem Ziel der Bekanntmachung im Rahmen der BMEL-Eiweißpflanzenstrategie vom 27.03.2015: „Eine signifikante Ausweitung und Verbesserung des Anbaus und der Verwertung von ... Bohnen ... in Deutschland“ (s. Titel dieser Bekanntmachung Nr. 09/14/31).

Zu diesem Zwecke kooperierte die RLP-AgroScience in Neustadt a.d. Weinstrasse in Aufgabe 6 („Feinkartierung und Lokalisation des V+C-Locus“) des AboVici-Projektes aufs Engste mit der Arbeitsgruppe in Göttingen um den Projektkoordinator Prof. Dr. W. Link, die das für die molekulargenetischen Untersuchungen benötigte Pflanzenmaterial züchtete und für die verschiedenen Teilaufgaben bereitstellte. Das Pflanzenmaterial wurde züchterisch so bearbeitet, des es optimale Voraussetzungen für die molekulargenetischen Analysen lieferte. Die daraus resultierenden Daten wiederum wurden von der AG Göttingen für die Entwicklung und Validierung möglichst eng mit dem VC-Genlocus gekoppelter Selektionsmarker verwendet (s.a. Tacke et al. 2021, Manuskript bei Plant Breeding eingereicht, die aktuelle Version ist diesem Abschlussbericht beigelegt).

Planung und Ablauf des Projektes

Das Vorhaben gliederte sich in mehrere Abschnitte, die teilweise aufeinander aufbauten und von der Verfügbarkeit des Pflanzenmaterials abhingen und vorwiegend auf sog. Next-Generation Sequenzierungen zur genomweiten Transkriptomanalyse von gereinigtem Gesamt-RNA-Material basierten.

Mithilfe zweier hoch-isogener Geschwister-Linienpaare, die dennoch für den VC-Gehalt aufspalten (jeweils ein VC-hoher -d.h. mit normalem oder „Wildtyp“ VC-Gehalt- und ein VC-armer Genotyp), sollte zuerst ein Referenz-Transkriptom für *Vicia faba* erstellt werden. Aus dem gleichen Pflanzenmaterial sollten über genomweite Genexpressionsanalysen signifikante Unterschiede identifiziert werden, die mit einem niedrigen Gehalt an VC korrelieren und daraus ein erster Satz an Kandidatengenen für den bis dato unbekanntem VC-Biosyntheseweg abgeleitet werden (Aufgabe 6.1). Aus dem gleichen Rohdatensatz sollten SNP-Kandidaten für den VC-Gehalt identifiziert werden, aus denen Assays entwickelt und validiert werden sollten, die zur Feinkartierung des VC-Genlocus verwendet werden können (Aufgabe 6.2). Über eine weitere Transkriptomanalyse von zwei aus einem Backcross stammenden, für VC-Gehalt maximal kontrastierende Bulks (Aufgabe 6.3) sowie von weiteren hoch-isogenen *V.faba* Linien (Aufgabe 6.4) sollten in Querschnittsvergleich über alle verwendeten Hoch-VC- und Niedrig-VC-Genotypen den Ergebnissen der ersten Analysen eine Einengung der ursprünglich identifizierten Kandidaten-SNPs und der differentiell exprimierten Kandidatengene erfolgen. Die aus der Schnittmenge resultierenden Kandidatengene sollte bioinformatisch auf ihre mögliche Funktion in der V+C-Biosynthese untersucht und an weiterem, genetisch unverwandtem aber für Vicin/Convicin Gehalt kontrastierendem *V.faba* Pflanzenmaterial auf ihre Expressionsmuster validiert werden (Aufgabe 6.5).

Vom Ablauf her waren die Arbeiten in Neustadt sowohl abhängig von der Bereitstellung des hochwertigen und züchterisch aufwändig hergestellten Pflanzenmaterials durch die AG Göttingen als auch von den Ergebnissen der Next-Generation-Sequenzierungen für die Transkriptomanalysen sowie der zugehörigen Bioinformatik (Algorithmen für Expressionswerte, SNPs und InDel-Analysen), die bei einem externen Dienstleister beauftragt werden mussten. Da das Pflanzenmaterial (Fokus: Samenschalen als vermuteter Ort der VC-Biosynthese und Embryos, in der Verlängerungsphase auch Blätter) abhängig von Entwicklungszustand und Reifegrad der sich entwickelnden Samen war, gab es hier zum ursprünglichen Zeitplan leichte Verschiebungen in der Verfügbarkeit, die der jahreszeitlichen Entwicklung der Ackerbohnenpflanzen im Vergleich zum offiziellen Projektstart geschuldet waren. Merkliche Verzögerungen ergaben sich bei der ersten und zweiten NGS-Sequenzierungsrunde bzw. der daran anknüpfenden Bioinformatik, bei der in der ersten Runde (2017) eine Probenvertauschung durch intensives Prüfen der fertigen Datensätze nachgewiesen werden konnte, die zu einer Wiederholung der gesamten Auswertungen (Expressions & SNP-Datensätze) führte. Im zweiten NGS-Ansatz (2019) führte eine zu strikte Parametrierung der Rohdaten beim Mapping auf das 2017 erstellte Referenztranskriptom zu einer zu geringen Anzahl an verwertbaren „reads“ (deutlich unter 50% unter der angestrebten Sequenziertiefe) in allen RNASeq-Proben, die in die eigentlichen bioinformatischen Analysen einfließen. Die stellte sich in einer ersten kritischen Analyse der Ergebnisse im Nachhinein heraus und konnte anhand von Referenzdaten aus der ersten NGS-Runde nachgewiesen werden. Durch eine Wiederholung der bioinformatischen Prozessierung der Rohdaten mit angepassten Parametern konnte dies jedoch von Sequenzierdienstleister kuriert werden. Die nachfolgenden Expressions- und SNP-Analysen mussten ebenfalls mit den verbesserten Datensätzen wiederholt werden. Dadurch ergaben sich weitere erhebliche Verzögerungen. Zur Validierung der wichtigsten Kandidatengene erschien eine Überprüfung der differentiellen Genexpression an -im Projektverlauf bis dato nicht berücksichtigtem- Blattmaterial kontrastierender VC-Genotypen erforderlich. Gegen Projektende ergaben sich noch Verzögerungen im Ablaufplan der qPCR-basierten Validierungen durch Corona-Maßnahmen-bedingte Einschränkungen der Arbeitsmöglichkeiten, die jedoch durch entsprechende zeitliche Verlängerung der Projektlaufzeit ausgeglichen werden konnten. Letztlich konnten die gestellten Aufgaben durch die genehmigten Laufzeitverlängerungen erfolgreich bearbeitet werden.

Wissenschaftlicher und technischer Stand, an den angeknüpft wurde

Die Samen der einjährigen Faba- oder Ackerbohne (*Vicia faba* L.) werden als Lebens- und Futtermittel genutzt. Sie sind wegen ihres Gehaltes und hochwertigem Protein von ca. 30% als regionale pflanzliche Proteinquelle geschätzt (Link et al. 1994; Crépon et al. 2010). Ackerbohnen werden in der Landwirtschaft zur Verbesserung der Bodenfruchtbarkeit und in Fruchtfolgen zur Unterbrechung von Schädlingen und Krankheiten eingesetzt (Köpke und Nemecek 2010; Kulak et al. 2013). Allerdings ist der Anbau von Faba-Bohnen derzeit in Deutschland auf ca. 50.000 ha begrenzt (Statistisches Bundesamt 2020). Ursachen hierfür sind u.a. ihre Anfälligkeit gegen Wassermangel und das Vorkommen zweier antinutritiver Sameninhaltsstoffe, Vicin [2,6-Diamino-4,5-Dihydroxypyrimidine-5-(β -D-Glucopyranosid)] und Convicin [2,4,5-Trihydroxy-6-aminopyrimidine-5-(β -D-Glucopyranoside (V/C)). Es handelt sich um zwei strukturell verwandte Pyrimidin-Glykoside, die nur aus *Vicia faba*, *Vicia narbonensis* und *Vicia sativa* (Griffith and Ramsay, 1992) bekannt sind. Sie kommen in der Fababohne in allen Pflanzenteilen vor und sind im Ackerbohnen Saatgut mit ca. 0,3% bis 1,5% in der Trockensubstanz zu finden (Goyoaga et al. 2008). Der Biosyntheseweg konnte bisher noch nicht vollständig aufgeklärt werden. Es bestand aber der Verdacht, dass diese Substanzen über den Pyrimidin- bzw. Orotsäure-Biosyntheseweg entstehen könnten (Brown und Roberts 1972).

Durch einen hohen VC-Gehalt sind Anbau und Verwertung der Ackerbohne deutlich eingeschränkt. Dies gilt zum einen in der Tierhaltung bei Nicht-Wiederkäuern, bei denen zu hohe Anteile an Ackerbohne im Futter zu deutlichen Leistungsverlusten (z.B. bei Legehennen oder Masthühnern) führen (I. Halle, 2019 Khamassi et al. 2013). Zum anderen führt auch beim

Menschen mit einem X-chromosomalen, genetisch bedingten Mangel an Glukose-6-Phosphat-Dehydrogenase der Verzehr von VC-haltigen Bohnen bei den Betroffenen zu einer hämolytischen Anämie, die auch als Favismus bezeichnet wird (Cappellini und Fiorelli 2008). Mit ca. 7.5% der Weltbevölkerung handelt es sich um die häufigste Enzym-Krankheit beim Menschen.

Es ist nur eine züchterisch relevante Quelle für einen genetisch bedingt sehr niedrigen Vicin und Convicin-Gehalt (synonym „V/C“; „niedrig-V/C“; „V/C-arm“) von Ackerbohnen-Samen beschrieben (Duc et al., 1989), die in allen aktuell auf dem Markt verfügbaren Sorten eingekreuzt wurde. Die Akzession 1268(4)(1) aus Radzikov (Polen) wies mit 0,46‰ nur etwa ein Zwanzigstel des üblichen V/C-Gehaltes auf. Dieser im Vergleich zum Wildtyp um 95% reduzierte Gehalt vererbt sich monogen rezessiv mit intermediärer Ausprägung bei Heterozygotie („vc-“/“VC+“). Die Samen übernehmen ihren V/C-Gehalt aus dem Genotyp ihrer Mutterpflanze. Vicin und Convicin werden in der Samenschale gebildet und in den Embryo transportiert (Duc et al., 1989; Brown and Roberts, 1972; Ray et al., 2015, Björnsdotter et al. 2020).

Eine assemblierte und annotierte Sequenz des Genoms der Faba-Bohne (ca. 13 Gb) ist derzeit noch nicht publiziert. Daher ist der hohe Synteniegrad von *Vicia faba* mit den sequenzierten Genomen von *Medicago truncatula* und weiteren Leguminosen der Unterfamilie Faboideae wichtig für genomische Analysen (Cruz-Izquierdo et al. 2012, Übersicht in O'Sullivan und Angra 2016;). Dies führte zur Entwicklung von SNP-basierten KASP-Assays u.a. für die Erstellung von genetische Karten (Webb et al 2016). Der VC-Locus wurde mit von *M.truncatula* abgeleiteten SNP-Markern als Haupt-QTL distal auf Chromosom 1 der Faba-Bohne kartiert (Khazaei et al. 2015). Diese Region, die bei *V.faba* den VC-Locus beherbergt, weist eine hohe Kolinearität zu einem syntänen Abschnitt auf Chromosom 2 des *M.truncatula* Genoms auf. In *M.truncatula* wurden in dem Intervall, das durch die den VC-Locus flankierenden Marker begrenzt wird, insgesamt 340 Gene gefunden. Die QTL-Region erstreckte sich über ein Intervall von 3,6 cM, flankiert von co-segregierenden SNP-Markern (Khazai et al 2017, Song 2017). Leider erwiesen sich auch die Marker, die am nächsten beieinander lagen, als nicht zuverlässig diagnostisch für das Merkmal V/C-Armut in einem größeren Satz verschiedener Genotypen und sind daher nur eingeschränkt nutzbar für die gezielte Züchtung auf Niedrig-VC-Sorten.

Mithilfe einer 454-Pyrosequenzierung wurden erste differentiell exprimierte Kandidaten-Contigs für das „vc-“Gens durch Vergleich der Transkriptome einer Niedrig-VC-Sorte mit zwei hoch-VC-Sorten charakterisiert (Ray et al. (2015). Die verwendeten Genotypen waren aber insgesamt bezogen auf den genetischen Hintergrund zu unterschiedlich und es konnte das ursächliche Gen nicht zweifelsfrei identifiziert werden. Weitere Details zum wissenschaftlichen und technischen Standes der Forschung findet sich in einem gemeinsamen, diesem Bericht beigelegten Manuskript, welches zur Veröffentlichung eingereicht wurde (Tacke et al. 2021).

Material und Methoden

Die Untersuchungen fokussierten auf die unreifen Samenschalen der Entwicklungsstadien 4, 5 und 6 sowie auch um gleichzeitig präparierte Embryos von Ackerbohnen. Die Samenschalen gelten als der Ort der Biosynthese von Vicin und Convicin. Daher wurden sie als Material als für die Transkriptom-Analysen am geeignetsten ausgewählt. Die Embryos, in die Vicin und Convicin von der maternalen Samenschale aus eingelagert werden, dienten der -späteren- Kontrolle von Kandidatengenen auf Expressionsebene. Gene, die ursächlich in der VC-Biosynthese involviert sind, sollten in der Samenschale aktiv sein, dagegen in den Embryos nur schwach bis gar nicht exprimiert werden. Die aus Embryos isolierte RNA diente daher als Negativkontrolle. Darüber hinaus dienten junge Blätter als Kontrollmaterial, um Kandidatengene in Ihrem Expressionsmuster in vegetativen Gewebe zu überprüfen.

Die verwendeten Genotypen wiesen hinsichtlich des VC-Locus jeweils die gleiche Allelkonfiguration auf (homozygot Wildtyp/VC-Gehalt: „V+C+“; heterozygot intermediärer VC-Gehalt „V+C+/v-c-“; homozygot niedriger-VC-Gehalt „v-c-“). Dies wurde durch sorgfältige Genotypi-

Abo-Vici - Abschlußbericht - RLP AgroScience GmbH

Züchtung und Agronomie neuartiger, Vicin-armer Ackerbohnen und Einsatz als einheimisches Eiweißfutter
 Förderkennzeichen: 2815EPS066 23 April 2021

sierung anhand vorhandener KASP-Marker und phänotypischer Bestimmung des VC-Gehaltes von der AG Göttingen sichergestellt. Neben Hoch-isogenen Geschwisterpaaren und Linien wurden gepoolte Proben aus dem Rückkreuzungen Hiv*Mél in der BC3-F3 zur Züchtung VC-armer Winterackerbohnen sowie deren Ausgangseltern (letztere nur für Blattmaterial) verwendet. Außerdem wurden F2-Pflanzen von zwei weiteren Kreuzungen -mit der Sorte Ukko und jeweils einer VC-armen hochisogenen Linie als Eltern- genutzt, die in der F2-Population noch für VC-Gehalt aufspalten. Abb.1a und 1b geben einen Überblick über das gesamte verwendete Pflanzenmaterial. Weitere Details finden sich in dem schon erwähnten Manuskript (Tacke et al. Plant Breeding eingereicht) sowie im Abschlussbericht des Koordinators Prof. Dr. W. Link, AG Göttingen (Aufgabe 5.2).

	Hochisogene Geschwisterlinien							
	VC-Phänotyp	Testa		Embryo	VC-Genotyp		Analysen	
		NGS I	qPCR	qPCR	VC-Genotyp	SNP	Expro	
Ernte 7-2017	NIL							
	Gö x 139-1	Low - VC	✓	✓	✓	homozygot	v-/c-	✓
	Gö x 139-2	High - VC	✓	✓	✓	homozygot	V+/C+	✓
	Gö x 201-3	Low - VC	✓	✓	✓	homozygot	v-/c-	✓
	Gö x 201-4	High - VC	✓	✓	✓	homozygot	V+/C+	✓
	NPZ x 848-3	Low - VC	n.a.	✓	✓	homozygot	v-/c-	n.a.
	NPZ x 848-4	High - VC	n.a.	✓	✓	homozygot	V+/C+	n.a.
	NPZ x 843-2	Low - VC	n.a.	✓	✓	homozygot	v-/c-	n.a.
	NPZ x 843-3	High - VC	n.a.	✓	✓	homozygot	V+/C+	n.a.
Ernte 12-2018	F2-Population - Ukko-1-5 x (Mél/2xILB938/2)-139-1-3							
	VC-Phänotyp	Testa		Embryo	VC-Genotyp		Analysen	
	NGS	qPCR	qPCR	VC-Genotyp	SNP	Expro		
	F2-Pools							
	Ukko x „139-	Low - VC	n.a.	✓	✓	homozygot	v-/c-	n.a.
	Ukko x „139-	intermediär	n.a.	✓	✓	heterozygot	V+/C+ v-/c-	n.a.
	Ukko x „139-	High - VC	n.a.	✓	✓	homozygot	V+/C+	n.a.
	F2-Population - Ukko-1-5 x (LowVC.14.8099-843-2)-1-1							
	VC-Phänotyp	Testa		Embryo	VC-Genotyp		Analysen	
	NGS	qPCR	qPCR	VC-Genotyp	SNP	Expro		
F2-Pools								
Ukko x „843-	Low - VC	n.a.	✓	✓	homozygot	v-/c-	n.a.	
Ukko x „843-	intermediär	n.a.	✓	✓	heterozygot	V+/C+ v-/c-	n.a.	
Ukko x „843-	High - VC	n.a.	✓	✓	homozygot	V+/C+	n.a.	
Ernte 1-2019	Hochisogene "Sixdenier"-Linien							
	VC-Phänotyp	Testa		Embryo	VC-Genotyp		Analysen	
	NGS II	qPCR	qPCR	VC-Genotyp	SNP	Expro		
	F8							
	Six-247	Low - VC	✓	✓	✓	homozygot	v-/c-	✓
	Six-248	Low - VC	✓	✓	✓	homozygot	v-/c-	✓
	Six-249	intermediär	n.a.	✓	✓	heterozygot	V+/C+ v-/c-	n.a.
	Six-250	High - VC	✓	✓	✓	homozygot	V+/C+	✓
	Six-251	intermediär	n.a.	✓	✓	heterozygot	V+/C+ v-/c-	n.a.
	Backcross "Hiverna/2 x Mélodie/2"							
VC-Phänotyp	Testa		Embryo	VC-Genotyp		Analysen		
NGS II	qPCR	qPCR	VC-Genotyp	SNP	Expro			
BC-Pool								
BC3(F3)Low	Low - VC	✓	✓	✓	homozygot	v-/c-	✓	
BC3(F3)High	High - VC	✓	✓	✓	homozygot	V+/C+	✓	

Abbildung 1a: Pflanzenmaterial für die molekulargenetischen Aufgaben und seine Verwendung

	Hochisogene, Backcross und weitere Genotypen							
	Linien	VC-Phänotyp	Blätter		VC-Genotyp	Analysen		
			NGS	qPCR		SNP	Expro	
Ernte 4-2020	Gö x 139-1	Low - VC	n.a.	✓	homozygot	v-/c-	n.a.	✓
	Gö x 139-2	High - VC	n.a.	✓	homozygot	V+/C+	n.a.	✓
	NPZ x 848-3	Low - VC	n.a.	✓	homozygot	v-/c-	n.a.	✓
	NPZ x 848-4	High - VC	n.a.	✓	homozygot	V+/C+	n.a.	✓
	BC3(F3)Low	Low - VC	n.a.	✓	homozygot	v-/c-	n.a.	✓
	BC3(F3)High	High - VC	n.a.	✓	homozygot	V+/C+	n.a.	✓
	Mélodie/2*	Low - VC	n.a.	✓	homozygot	v-/c-	n.a.	✓
	Hiverna/2*	High - VC	n.a.	✓	homozygot	V+/C+	n.a.	✓
	Saskatoon	High - VC	n.a.	✓	homozygot	V+/C+	n.a.	✓

Abbildung 1b: Pflanzenmaterial für die molekulargenetischen Aufgaben und seine Verwendung

Aus allen Pflanzenproben wurde mit für das Untersuchungsmaterial optimierten Methoden die Gesamt-RNA isoliert, aufgereinigt und qualitätsgeprüft und dann für die weiteren Analysen (Next Generation Sequenzierung und qPCR-Analytik) verwendet. Die Transkriptomanalysen wurden mit Material aus 2017 („NGS I“) und 2019 („NGS II“) durchgeführt (Abb. 1a). Für das Referenztranskriptom wurde das Pflanzenmaterial aus 2017 („NGS I“) verwendet. Mithilfe der qPCR-Analytik wurden die Ergebnisse aus den Transkriptomdaten (Expressionsverhalten) am gleichen Material, welches für die NGS-Ansätze verwendet wurde, verifiziert und an weiteren, unverwandten Genotypen mit einem definierten VC-Status validiert. Die von einem externen Dienstleister generierten NGS-Datensätze (Expressionsdaten und SNP-Daten) und das Filtern auf VC-assoziierte Kandidaten erfolgten mithilfe von individuell angepassten Excel-Dateien. Ausgewählten Kandidaten-Contigs wurden manuell anhand von Standardsoftware für molekularbiologische Anwendungen (Lasergene Core Suite Software-Paket V.15 DNASTAR Inc. Madison, USA) analysiert und -wenn erforderlich- mithilfe von BLAST-Algorithmen (NCBI, USA) gegen die öffentlich verfügbaren Datenbanken abgeglichen (National Center for Biotechnology Information; <https://blast.ncbi.nlm.nih.gov/Blast.cgi>). Ansonsten wurden Sequenzvergleiche mit dem AboVici-Referenztranskriptom bevorzugt über einen lokalen BLASTN-Algorithmus intern durchgeführt. Weitere methodische Details zur Datenauswertung finden sich in dem angehängten Manuskript (Tacke et al. 2021). Alle RNA-Proben wurden mit dem BIORAD-iScript cDNA Synthesis Kit jeweils in zwei separaten Ansätzen in cDNA-Templates (A und B-Probe) umgewandelt und davon je zwei Aliquote in der qPCR eingesetzt, so dass letztlich 4 „technische“ Replikate für die Auswertung der Expressionsdaten zur Verfügung standen. Die qPCR-Rohdaten wurden mit dem öffentlich verfügbaren LinReg-PCR-Tool (Version 2017.1, Dr. M.J. Ruijter, Academic Medical Center, Amsterdam) prozessiert und die drei besten Cq-Werte pro Probe ermittelt. Die differentielle Expression von Kandidatengen wurde dann mit der Standard- $\delta\delta Cq$ -Methode und drei Referenzgenen kalkuliert. Genespezifische Amplikons wurden manuell mithilfe des frei-verfügbaren, internet-basierten Algorithmus „Primer3Plus“ (Untergasser et al 2007) entworfen. In der Anfangsphase wurde die Genexpression vorwiegend mit BIORAD-iTaq Universal SYBR Green Supermix nach Herstellerangaben bestimmt. Im weiteren Projektverlauf wurde jedoch auf Taqman-basierte Assays (TaqMan® Fast Advanced Master Mix, Applied Biosystems, Thermo-Fisher) umgestellt, um eine höhere Spezifität zu erreichen und PCR-basierte Artefakte zu vermeiden. Die Detektion der qPCR-Signale erfolgte auf zwei CFX96™ Real-Time PCR Systemen (BioRad, München).

Darstellung der wichtigsten Ergebnisse und Diskussion

Die o.g. Aufgaben 6.1 bis 6.5 und die zeitliche Abfolge der Verfügbarkeit des Pflanzenmaterials greifen thematisch ineinander. Daher werden im Folgenden die Ergebnisse entsprechend den entscheidenden Schwerpunkten gruppiert und bezüglich ihrer inhaltlichen Bedeutung zusammengefasst.

I. AboVici-Referenztranskriptom

Die Basis für genomweite Genexpressionsanalysen sowie zur Identifikation von SNP-Kandidaten für einen niedrigen VC-Gehalt stellte ein Testa-spezifisches Referenztranskriptom für *V.faba* dar (Abb. 2), welches aus Testa-Gewebe der Entwicklungsstadien 4 und 5 der beiden hoch-isogenen Geschwisterlinien (Abb. 1a, „NGS I“) im RNASeq-Standardverfahren von einem externen Sequenzierdienstleister erstellt wurde.

Längster Contig	15752 b
Kürzester Contig	201 b
Mittlere Contig-Länge	859 b
N ₅₀	1520 b
Alle assemblierten Contigs	52916
Alle assemblierten Basen	45465399 b

Abb. 2 *Vicia faba* Referenztranskriptom

Alle Gensequenzen (bzw. deren Contigs) wurden über einen Vergleich mit dem nahe verwandten *Medicago truncatula* Genom annotiert. Da *Vicia faba* auch einzigartige Gene enthält, die in *Medicago* nicht vorkommen (so z.B. höchstwahrscheinlich auch Gene für die VC-Biosynthese) wurden diese mit der Annotation „uncharacterized RNA“ in den Datensätzen hinterlegt.

Für die eigentlichen Analysen wurden eine Kombination aus dem etablierten RNASeq-Verfahren und einer 3'-basierten MACE-Sequenzierstrategie (Massive Analysis of cDNA Ends) gewählt. Letztere Methode fokussiert auf die 3'-Enden der mRNAs und benötigt einen geringeren technischen Sequenzieraufwand als das herkömmliche RNASeq-Verfahren. Daher war dieser Ansatz auch kostengünstiger und ermöglichte es, mehrere Testa-Proben (jeweils pro Genotyp für die Entwicklungsstadien 4, 5 und 6 getrennt) einzeln zu analysieren. Technische oder biologische Triplikate -wie in der akademischen Forschung üblich- waren aufgrund des begrenzten Budgets nicht möglich. Ziel war es, möglichst viele unterschiedliche Genotypen mit kontrastierenden VC-Gehalt im Projektverlauf über NGS-Ansätze analysieren zu können, um durch die unabhängigen Rekombinationsereignisse in der Region des VC-Locus zu einer maximalen Einengung auf das oder die Kandidatengene zu gelangen. RNASeq-Daten basierten z.T. auf gepoolten RNA-Proben (für Entwicklungsstadium 4 und 5 je Genotyp). MACE-Daten lieferten hinsichtlich der Expressionswerte ähnliche, aber nicht identische Ergebnisse wie die RNASeq-basierten Expressionswerte. Dabei wurden die separaten MACE-Daten für Stadium 4 und 5 mit den RNASeq-Daten der gepoolten Stadien 4 und 5 verglichen. Da es sich hierbei um zwei unterschiedliche Sequenzieretechniken handelt, mit denen das gleiche RNA-Material analysiert wurde, sollten die Ergebnisse aus beiden Analysen deutliche Übereinstimmungen aber auch Unterschiede zeigen, sowohl was die differentielle Expression als auch die potentiellen homozygot SNPs angeht. Dies wird im Folgenden noch thematisiert, wo es relevant war.

II. SNP-Analysen und SNP-Kandidaten

Bei der SNP-Analyse wurde berücksichtigt, dass die interessanten SNP-Sequenzen auf einer Region von Chromosom 2 von *M.truncatula* kartieren, welche syntän zu dem Chromosom 1 von *V.faba* ist, auf dem die vc-Mutation liegen muss. Des Weiteren wurde nach solchen SNP-haltigen Sequenzen gesucht, die nicht auf das *M.truncatula* Genom kartiert werden konnten. Da die Biosynthese von VC in *M.truncatula* nicht vorkommt, könnte es gut sein, dass sich unter den zugehörigen Contigs Gene befinden, die an der VC-Biosynthese beteiligt sind. Außerdem wurde auf die Schnittmenge derjenigen SNPs gefiltert, die allen Niedrig-VC-Genotypen gemeinsam sind. Dabei wurden nur bi-allelische SNPs identifiziert. D.h. die -entfernt verwandten- Hoch-VC-Genotypen tragen das gleiche Allel. Die beiden niedrig-VC-haltigen Genotypen,

denen der chromosomale Abschnitt der VC-Mutation gemeinsam ist, tragen das korrespondierende Allel.

Die Daten aus dem Pflanzenmaterial von 2019

Basierend auf den Transkriptomdaten aus der zweiten Sequenzierung („NGS II“ „2019“) wurden umfangreiche SNP-Analysen durchgeführt, deren nach Genotypen gestaffeltes methodisches Vorgehen im 3ten Zwischenbericht genauer beschrieben wurde. Die Auswertung war aber mit Ende des letzten Berichtszeitraums (Jan. 2020) nicht abgeschlossen. Es stand noch eine Zusammenschau mit den SNP-Daten aus der ersten Runde der Transkriptomanalysen 2017 aus. Die Erwartung war, dass ein Vergleich aller im Projekt verwendeter V/C-kontrastierenden Genotypen -ähnlich wie bei den Expressionsdaten- wirkungsvoll auf diejenigen SNP-Kandidaten fokussiert, die genetisch möglichst eng mit den V/C-Locus gekoppelt sind.

Bei der separaten Auswertung der SNP-Daten für die Hoch/Niedrig-VC-Paarungen (BC3-Low vs. BC3-High, Sixdenier“247/LowVC“ vs. Sixdenier250/HighVC, Sixdenier“248“ vs. Sixdenier250/HighVC) wurde anfänglich ausschließlich die Kategorie „strikt homozygote SNPs“ ausgewertet, analog zur ersten Projektphase „2017“. Die Redundanz-eliminiert Schnittmenge aus den SNPs, die jeweils in den RNASeq- oder MACE-Datensätzen getrennt identifiziert wurden (s. Zwischenbericht 3, Seite 14, Tabelle 21) ergab eine deutlich geringer Anzahl an SNP-Kandidaten für die Kategorien „auf Chr.2 von *M.truncatula* kartierbar“ sowie nicht auf *M.truncatula* kartierbar als die Auswertung von 2017 (Zwischenbericht 1, Tab. 8 bis 11, Seite 8 und 9).

Bei einer Durchsicht der individuellen SNP-Werte für einzelne Kandidaten fiel im Nachhinein auf, dass die vom SNP-Detektionsalgorithmus definierte Kategorie „strikt homozygot“ sehr strenge Kriterien für einen SNP-Kandidaten an die Zählwerte für die Basen (Referenz bzw. Variante) anlegt. Verdeutlicht wird das (s. Abb.3) anhand dreier Beispiele. Die Referenzsequenz von Contig19580 besitzt in Position 441 einen A/G-Dimorphismus, der in niedrig-VC-Genotypen (Low VC) ein „A“ im Gegensatz zu „G“ im Wildtyp (High VC) aufweist. Das geht aus der Interpretation der Zählwerte für die beiden alternativen Basen „A“ und „G“ an der Position 441 für die drei Einzelvergleiche hervor. Allerdings ist der Zähler für die Gegenbase nicht unbedingt „0 wie es im Idealfall zu erwarten wäre“, sondern 2 oder 3. Bedingt durch unvermeidbare Fehleinbauten bei der biochemischen Herstellung der cDNA-Matrizen für die eigentliche Sequenzierung kann es aber gelegentlich zu falschen Basen d.h. Artefakten in der Sequenzabfolge kommen.

Genotyp	gene_id	description	Position	SNP-Nucleotid		Basen - counts		snp_type	M.truncatula Chr.2 Region (+10000bp)		
				Referenz	Variante	Low VC	High VC				
247	Contig19580	V7BV39_PHAVU Peroxidase	441	A	G	90	2	2	1076	majority_strict	1320001
248	Contig19580	V7BV39_PHAVU Peroxidase	441	A	G	105	0	2	1076	majority_strict	1320001
BC3	Contig19580	V7BV39_PHAVU Peroxidase	441	A	G	174	0	3	737	majority_strict	1320001
247	Contig11262	G7ILA4_MEDTR Glycosyltransferase	1603	T	G	397	1	0	99	majority_strict	1350001
248	Contig11262	G7ILA4_MEDTR Glycosyltransferase	1603	T	G	351	0	0	99	homozygous_strict	1350001
BC3	Contig11262	G7ILA4_MEDTR Glycosyltransferase	1603	T	G	334	0	1	80	homozygous_strict	1350001
247	Contig14304	3,4-dihydroxy-2-butanone 4-phosphate synthase	726	A	T	39	3557	7911	27	homozygous_strict	1850001
248	Contig14304	3,4-dihydroxy-2-butanone 4-phosphate synthase	726	A	T	25	2428	7911	27	majority_strict	1850001
BC3	Contig14304	3,4-dihydroxy-2-butanone 4-phosphate synthase	726	A	T	26	3521	7883	29	homozygous_strict	1850001

Abbildung 3: Beispiele für Kategorisierung von SNP-Kandidaten als „majority_strict“ und „strikt homozygot“ aus den RNASeq-Daten, je Genotypen-Vergleich

Sinkt dadurch die Häufigkeit von Base zu Variante unter das Verhältnis von 1:100 werden in der Konsequenz einige SNP-Konfigurationen statt als „homozygous_strict“ in die Kategorie „majority_strict“ eingeordnet. Allerdings funktioniert diese Kategorisierung nicht immer genau, wie der Fall Contig11262 für den BC3 an Position 1603 zeigte. Hier wurde der SNP-Kandidat vom Algorithmus als „homozygous_strict“ eingeordnet, obwohl in dem High-VC Genotyp nur ein Verhältnis von 1:80 vorliegt. Für Contig14304 liegt in Fall von Sixdenier„248“ das Verhältnis von Referenz zu Variante knapp unter 1:100 (i.e. 25:2428), was zu der Kategorie „majori-

Abo-Vici - Abschlußbericht - RLP AgroScience GmbH

Züchtung und Agronomie neuartiger, Vicin-arter Ackerbohnen und Einsatz als einheimisches Eiweißfutter
 Förderkennzeichen: 2815EPS066 23 April 2021

ty_strict führte. In allen drei Beispielen ist es jedoch nicht unwahrscheinlich, dass der Zähler für die Gegenbase eigentlich „0“ betragen müsste. Alle drei o.g. Contig-SNP-Beispiele würden nach einer Auswahl, die sich ausschließlich auf die Kategorisierung „homozygous_strict“ stützt, aussortiert. Um nicht von vorneherein durch zu stringente Auswahlkriterien mögliche SNP-Kandidaten zu verlieren, wurden aufgrund der oben beschriebenen Beobachtung die Auswertungen um die Kategorie „majority_strict“ erweitert und die Zahlenwerte anschließend kritisch einzeln ausgewertet. Dies wurde rückwirkend auch mit den Datensätzen aus der ersten Transkriptomanalyse („NGS II“, 2019) durchgeführt, um in der abschließenden Gesamtbewertung über alle im Projekt analysierten Genotypen eine möglichst umfassende Liste putativer SNP-Kandidaten zu erhalten.

Zuerst wurden die Paarvergleiche einzeln ausgewertet, wieder unterschieden in SNP-Kandidaten, deren zugehörige Contig-Sequenzen sich auf Chromosom 2 von *M. truncatula* kartieren ließen („Chr.2“) und denjenigen, die keinem Chromosom des *Medicago*-Genoms zugeordnet werden konnten („NA“). Die auf Chr.2 kartierten SNP-Contigs wurden noch nach ihrer Position sortiert: von links nach rechts enthält die erste Spalte alle auf dem gesamten Chr.2 kartierten SNP-Contigs, die zweite Spalte diejenigen, die durch zwei den VC-Locus weitläufig umfassende SNP-Marker (Vf_Mt2g005900_001: 390.181 - 390.203 und Vf_Mt2g013900_001: 3.818.903 - 3.819.080) definiert wird, die dritte Spalte umfasst nur noch SNP-Contigs, deren Position auf Chr.2 innerhalb der bis dato besten SNP-Marker liegt (Vf_Mt2g009320_001: 1.869.538 - 1.869.707 und Vf_Mt2g010740_001: 2.505.174 - 2.505.326), die den VC-Locus lt. AG Prof. Link am effektivsten eingrenzen. Es wurden dabei jeweils die Kategorien „homozygous_strict“ und majority_strict separat gelistet. Für die MACE-Daten (Abb.3) aus 2019 fällt auf, dass in etwa gleich viele SNP-Kandidaten in beide Kategorien fallen.

Six247 vs. Six250				
MACE-2019-SNPs Chr. 2				NA
von	100,001	390001	1,860,001	
bis	3,080,001	3,080,001	2,500,001	
homozygous strict	84	82	26	26
majority strict	91	88	29	21
Σ (Hs&Ms):	175	170	55	47
different contigs:	73	69	25	16
Six248 vs. Six250				
MACE-2019-SNPs Chr. 2				NA
von	100,001	390001	1,860,001	
bis	3,080,001	3,080,001	2,500,001	
homozygous strict	88	87	27	24
majority strict	96	93	31	21
Σ (Hs&Ms):	184	180	58	45
different contigs:	81	78	26	19
BC3-Low vs. BC3-High				
MACE-2019-SNPs Chr. 2				NA
von	360001	860,001	1,870,001	
bis	29,180,001	2,090,001	2,090,001	
homozygous strict	28	25	4	24
majority strict	28	27	3	4
Σ (Hs&Ms):	56	52	7	28
different contigs:	26	22	5	19

Abbildung 3: MACE-Daten „homozygous strict“ und „majority_strict“-SNPs, je Genotypen-Vergleich

Abo-Vici - Abschlußbericht - RLP AgroScience GmbH

Züchtung und Agronomie neuartiger, Vicin-arter Ackerbohnen und Einsatz als einheimisches Eiweißfutter
 Förderkennzeichen: 2815EPS066 23 April 2021

Bei den RNASeq-Daten (Abb.4) fallen ca. vier- bis sechsmal mehr SNP-Kandidaten der Kategorie „homozygous_strict“ an als für die Kategorie „majority_strict“. Die -höchstwahrscheinlich technische- Ursache für dieses Phänomen ist unbekannt, es könnte jedoch an der insgesamt geringeren Sequenziertiefe der MACE-Ansätze im Vergleich zu den RNASeq-Ansätzen liegen. Insgesamt finden sich in den RNASeq-Datensätzen deutlich mehr SNP-Kandidaten als in den korrespondierenden MACE-Datensätzen. Dies und die Größenordnung der SNP-Kandidaten pro Paarvergleich insgesamt stimmen mit den Erfahrungen aus der ersten Transkriptomanalyse überein.

Six247 vs. Six250				
RNASeq-2019-SNPs Chr. 2				NA
von	40,001	390001	1,860,001	
bis	43,710,001	3,080,001	2,500,001	
homozygous strict	691	663	233	126
majority strict	88	85	23	13
Σ (Hs&Ms):	779	748	256	139
different contigs:	144	132	44	35
Sixdenier 248 vs 250				
RNASeq-2019-SNPs Chr. 2				NA
von	540,001	390001	1,870,001	
bis	43710001	3080001	2,500,001	
homozygous strict	645	615	199	116
majority strict	110	108	34	14
Σ (Hs&Ms):	755	723	233	130
different contigs:	147	135	41	37
BC3-Low vs. BC3-High				
RNASeq-2019-SNPs Chr. 2				NA
von	1,320,001	1320001	1,860,001	
bis	1,980,001	1,980,001	1,980,001	
homozygous strict	190	190	34	86
majority strict	41	41	10	14
Σ (Hs&Ms):	231	231	44	100
different contigs:	46	46	9	39

Abbildung 4: RNASeq-Daten „homozygous strict“ und „majority_strict“-SNPs, je Genotypen-Vergleich

Es fällt außerdem auf, dass die Anzahl der SNP-Kandidaten, die in den BC3-Datensätzen zu finden sind, in etwa nur ein Drittel der SNP-Kandidaten beträgt, die aus den „Sixdenier“-Datensätzen gefiltert werden konnten. Die Ursache dürfte in der unterschiedlichen Genetik der Ansätze (Hoch-isogene Paare versus Backcross-Bulks) zu finden sein. Die Einbeziehung der Kategorie „majority_strict“ bei den RNASeq-Daten führte bei dem Paarvergleich Sixdenier 248 (LVC) versus Sixdenier 250 (HVC) zu einer ca. 13% Steigerung der SNP-Kandidaten, bei Sixdenier 249 (LVC) versus Sixdenier 250 (HVC) zu einer ca. 17% Zunahme und bei dem BC3-Bulks zu fast 22% mehr SNP-Kandidaten. Bei den MACE-Datensätzen beträgt der Gewinn ca. 50%.

Aus diesen Primär-Kandidaten der Paarvergleiche wurden nun -getrennt für MACE- und RNASeq-Daten- zuerst für die hochisogenen „Sixdenier“-Linien diejenigen SNP-Kandidaten herausgefiltert, die beiden Low-VC/Hoch-V/C-Paaren bzgl. der Position des Dimorphismus gemeinsam sind. Dabei ergab sich zwangsläufig eine neue Kategorie von SNP-Kandidaten, die als „mixed“ bezeichnet werden und die aus einem als „homozygot_strict“ in einem hochisogenen Paarvergleich und einem als „majority_strict“ bewerteten SNP-Kandidaten im ande-

Abo-Vici - Abschlußbericht - RLP AgroScience GmbH

Züchtung und Agronomie neuartiger, Vicin-arter Ackerbohnen und Einsatz als einheimisches Eiweißfutter
 Förderkennzeichen: 2815EPS066 23 April 2021

ren hoch-isogenen Paarvergleich zusammengesetzt sind. Auch SNP-Kandidaten mit einer Einstufung als „majority_strict“ aus beiden Paarvergleichen wurden detektiert. Bei den MACE-Daten fällt nur ein geringer Teil der SNP-Kandidaten in die Kategorie „mixed“ (Abb.5). Insgesamt erweitern die Kategorien „majority_strict“ und „mixed“ die Zahl an potentiellen SNP-Kandidaten gegenüber der Kategorie „homozygous_strict“ erheblich.

[Six247 & Six248] vs. Six250					
MACE-2019-SNPs Chr. 2				NA	
von	100,001	390001	1,860,001		
bis	3,020,001	3,080,001	2,500,001		
homozygous strict	59	58	18	14	
majority strict	74	71	20	17	
mixed	14	14	9	3	
Σ (Hs&Ms):	147	143	47	34	
diff. contigs:	69	66	24	13	

Abbildung 5: MACE - SNP-Kandidaten Schnittmenge über alle „Sixdenier“-Genotypen 2019

Bei den RNASeq-Daten finden sich dagegen in der Kategorie „mixed“ fast genauso viele SNP-Kandidaten wie in der Kategorie „majority_strict“. Der überwiegende Teil der SNP-Kandidaten zählt jedoch weiter in die Kategorie „homozygous_strict“ (Abb.6).

[Six247 & Six248] vs. Six250					
RNASeq-2019-SNPs Chr. 2				NA	
von	40,001	390001	1,870,001		
bis	43,710,001	3,020,001	2,500,001		
homozygous strict	604	577	198	97	
majority strict	65	64	17	6	
mixed	61	58	21	13	
Σ (Hs&Ms):	730	699	236	116	
diff. contigs:	141	129	42	29	

Abbildung 6: RNASeq - SNP-Kandidaten Schnittmenge über alle „Sixdenier“-Genotypen 2019

Insgesamt ist durch die Schnittmengenbildung über beide hoch-isogenen Sixdenier-Paare eine Reduktion der Menge an SNP-Kandidaten („Σ(H&Ms)“) erkennbar, die aber moderat ausfällt und für die MACE-Daten bei ca. 18% liegt, bei den RNASeq-Daten bei nur 5%. Eine wesentlich deutlichere Einengung der SNP-Kandidaten ergab sich im nächsten Schritt durch die weitere Schnittmengenbildung mit dem Backcross (i.e. BC3-Bulks, „BC3“, Abb.7).

BC3 & Six247 & Six248					
MACE-2019-SNPs Chr. 2				NA	
von	100,001	1330001	1,930,001		
bis	3,020,001	1,950,001	1,950,001		
homozygous strict	4	4	2	1	
majority strict	11	11	1	3	
mixed	4	4	0	0	
Σ (Hs&Ms):	19	19	3	4	
diff. contigs:	11	11	2	2	

Abbildung 7: MACE - SNP-Kandidaten Schnittmenge über BC3 und „Sixdenier“-Genotypen 2019

Von ursprünglich 147 SNP-Kandidaten mit einer Lokalisierbarkeit auf Chr.2 von *M.truncatula* bleiben nur noch 19 Stück als allen LVC/HVC-Paaren gemeinsam übrig. Das entspricht einer

Verringerung um ca. 87%. Ein ähnliches Bild ergibt sich für die nicht-kartierbaren („NA“) SNP-Kandidaten. Von 34 bleiben nur 4 Stück übrig.

Für die RNASeq-Daten reduziert sich der Satz an SNP-Kandidaten ebenfalls um 84 bis 73% („Chr.2“ bzw. „NA“). Allerdings bleiben deutlich mehr SNP-Kandidaten übrig (6- bis 8-fach) als im Vergleich zum MACE-Datensatz (Abb.8).

	BC3 & Six247 & Six248			
	RNASeq-2019-SNPs Chr. 2			NA
von	1,320,001	1,320,001	1,860,001	
bis	1,980,001	1,980,001	1,980,001	
homozygous strict	79	79	51	27
majority strict	8	8	0	1
mixed	29	29	6	4
Σ (Hs&Ms):	116	116	57	32
diff. contigs:	34	34	7	6

Abbildung 8: RNASeq - SNP-Kandidaten Schnittmenge über BC3 und „Sixdenier“-Genotypen 2019

Abschließend wurde eine Redundanz-eliminierte Liste aller SNP-Kandidaten aus den separat analysierten MACE- und RNASeq-Daten 2019 erstellt (Abb.9).

	RNASeq & MACE: 2019		
	BC3&Six247 & Six248		
	Chr. 2-wide	Chr.2-narrow	NA
von	1,320,001	1,860,001	
bis	1,980,001	1,980,001	
homozygous strict	81	14	28
majority strict	10	0	1
mixed	27	5	4
Σ (Hs&Ms):	118	19	33
diff. Contigs:	35	7	6

Abbildung 9: RNASeq&MACE - nicht-redundante SNP-Kandidaten für alle Genotypen 2019

Durch die Einbeziehung der Kategorie „majority_strict“ erweiterte sich die Liste der ursprünglichen SNP-Kandidaten auf Chr.2 vom 81 um 37 Stück auf 118 und bei den nicht-kartierbaren Contigs um 5 auf insgesamt 33 SNP-Kandidaten. An der Tatsache, dass die MACE-Daten deutlich weniger SNP-Kandidaten zur finalen Liste beitrugen änderte sich dadurch nichts (vgl. 3ter Zwischenbericht, Seite 14). Unter den 116 (RNASeq) bzw. 19 (MACE) SNP-Kandidaten waren 17 redundant, d.h. sie waren in beiden Datensätzen vorhanden. Aus beiden Methoden (MACE und RNASeq) war jeweils eine Bewertung der SNP-Zahlenwerte nach „homozygot_strict“, mixed oder „majority_strict“ für die gleiche SNP-Position vorhanden. Bei der Durchsicht der 17 redundanten SNP-Kandidaten fand sich in 8 SNP-Kandidaten (aus 5 verschiedenen Contigs) die Situation, dass derselbe SNP aus den MACE-Daten für alle drei Genotypen als „majority_strict“ kategorisiert wurde, die RNASeq-Daten diesen SNP durchgängig als „homozygot_strict“ identifizierten (Abb.10). Die umgekehrte Situation war jedoch nicht vorhanden.

Abo-Vici - Abschlußbericht - RLP AgroScience GmbH

Züchtung und Agronomie neuartiger, Vicin-arter Ackerbohnen und Einsatz als einheimisches Eiweißfutter
 Förderkennzeichen: 2815EPS066 23 April 2021

Genotyp	gene_id	position	Basen - counts						snp_type	Region (+10C Kongruenz)	Detection
			SNP-Nucleotid		Low VC [4u5u6]		High VC [4u5u6]				
			Referenz	Variante	Referenz	Variante	Referenz	Variante			
247 TRINITY_DN37464_c0_g1_i1_2	468	A	G	157	0	3	246	majority_strict	1330001	✓	MACE-2019
248 TRINITY_DN37464_c0_g1_i1_2	468	A	G	224	2	3	246	majority_strict	1330001	✓	MACE-2019
BC3 TRINITY_DN37464_c0_g1_i1_2	468	A	G	165	1	2	147	majority_strict	1330001	✓	MACE-2019
247 TRINITY_DN37464_c0_g1_i1_2	468	A	G	1560	0	4	1012	homozygous_strict	1330001	✓	RNASeq-2019
248 TRINITY_DN37464_c0_g1_i1_2	468	A	G	875	1	4	1012	homozygous_strict	1330001	✓	RNASeq-2019
BC3 TRINITY_DN37464_c0_g1_i1_2	468	A	G	1030	0	4	1193	homozygous_strict	1330001	✓	RNASeq-2019
247 Contig3485	252	A	G	947	21	21	844	majority_strict	1550001	✓	MACE-2019
248 Contig3485	252	A	G	915	11	21	844	majority_strict	1550001	✓	MACE-2019
BC3 Contig3485	252	A	G	2127	35	36	1512	majority_strict	1550001	✓	MACE-2019
247 Contig3485	252	A	G	505	1	0	445	homozygous_strict	1550001	✓	RNASeq-2019
248 Contig3485	252	A	G	317	3	0	445	homozygous_strict	1550001	✓	RNASeq-2019
BC3 Contig3485	252	A	G	810	3	2	641	homozygous_strict	1550001	✓	RNASeq-2019
247 Contig3485	309	G	A	498	4	11	442	majority_strict	1550001	✓	MACE-2019
248 Contig3485	309	G	A	554	7	11	442	majority_strict	1550001	✓	MACE-2019
BC3 Contig3485	309	G	A	1080	9	3	732	majority_strict	1550001	✓	MACE-2019
247 Contig3485	309	G	A	762	0	0	671	homozygous_strict	1550001	✓	RNASeq-2019
248 Contig3485	309	G	A	476	3	0	671	homozygous_strict	1550001	✓	RNASeq-2019
BC3 Contig3485	309	G	A	1222	1	1	893	homozygous_strict	1550001	✓	RNASeq-2019
247 Contig4459_2	1622	C	T	196	2	7	167	majority_strict	1760001	✓	MACE-2019
248 Contig4459_2	1622	C	T	287	1	7	167	majority_strict	1760001	✓	MACE-2019
BC3 Contig4459_2	1622	C	T	199	1	6	199	majority_strict	1760001	✓	MACE-2019
247 Contig4459_2	1622	C	T	761	0	2	348	homozygous_strict	1760001	✓	RNASeq-2019
248 Contig4459_2	1622	C	T	449	1	2	348	homozygous_strict	1760001	✓	RNASeq-2019
BC3 Contig4459_2	1622	C	T	592	1	2	668	homozygous_strict	1760001	✓	RNASeq-2019
247 Contig4459_2	1729	C	T	710	8	18	372	majority_strict	1760001	✓	MACE-2019
248 Contig4459_2	1729	C	T	679	2	18	372	majority_strict	1760001	✓	MACE-2019
BC3 Contig4459_2	1729	C	T	813	24	18	726	majority_strict	1760001	✓	MACE-2019
247 Contig4459_2	1729	C	T	291	0	0	166	homozygous_strict	1760001	✓	RNASeq-2019
248 Contig4459_2	1729	C	T	176	0	0	166	homozygous_strict	1760001	✓	RNASeq-2019
BC3 Contig4459_2	1729	C	T	229	0	0	253	homozygous_strict	1760001	✓	RNASeq-2019
247 Contig4459_2	1797	A	G	756	28	18	331	majority_strict	1760001	✓	MACE-2019
248 Contig4459_2	1797	A	G	602	15	18	331	majority_strict	1760001	✓	MACE-2019
BC3 Contig4459_2	1797	A	G	883	50	64	875	majority_strict	1760001	✓	MACE-2019
247 Contig4459_2	1797	A	G	256	0	0	100	homozygous_strict	1760001	✓	RNASeq-2019
248 Contig4459_2	1797	A	G	136	0	0	100	homozygous_strict	1760001	✓	RNASeq-2019
BC3 Contig4459_2	1797	A	G	179	0	1	173	homozygous_strict	1760001	✓	RNASeq-2019
247 Contig14304	84	A	G	1353	47	24	5023	majority_strict	1850001	✓	MACE-2019
248 Contig14304	84	A	G	1918	57	24	5023	majority_strict	1850001	✓	MACE-2019
BC3 Contig14304	84	A	G	1619	77	9	3381	majority_strict	1850001	✓	MACE-2019
247 Contig14304	84	A	G	2221	16	10	6235	homozygous_strict	1850001	✓	RNASeq-2019
248 Contig14304	84	A	G	1440	12	10	6235	homozygous_strict	1850001	✓	RNASeq-2019
BC3 Contig14304	84	A	G	1892	2	15	7596	homozygous_strict	1850001	✓	RNASeq-2019
247 Contig4170	1853	C	T	230	7	3	114	majority_strict	1950001	✓	MACE-2019
248 Contig4170	1853	C	T	189	0	3	114	majority_strict	1950001	✓	MACE-2019
BC3 Contig4170	1853	C	T	251	2	3	162	majority_strict	1950001	✓	MACE-2019
247 Contig4170	1853	C	T	304	0	1	149	homozygous_strict	1950001	✓	RNASeq-2019
248 Contig4170	1853	C	T	196	0	1	149	homozygous_strict	1950001	✓	RNASeq-2019
BC3 Contig4170	1853	C	T	226	0	1	241	homozygous_strict	1950001	✓	RNASeq-2019

Abbildung 10: RNASeq&MACE - redundante SNP-Kandidaten für alle Genotypen 2019

Ob die unterschiedlichen SNP-Ergebnisse mit den beiden unterschiedlichen Sequenzierstrategien (insbesondere die cDNA-Template-Herstellung aus der Gesamt-RNA) bzw. der unterschiedlichen Sequenziertiefe bei MACE und RNASeq erklärt werden können, bleibt offen. Vergleicht man die Zahlen zwischen MACE- und RNASeq-Daten, so ist kein direkter Zusammenhang mit der größeren Sequenziertiefe bei RNASeq erkennbar. Die Einbeziehung der Kategorie „majority_strict“ in die SNP-Auswertung erschien sinnvoll, da offenbar die Einordnung eines SNP-Kandidaten in Abhängigkeit von der verwendeten Methode variieren und ggf. zum vorzeitigen Ausschluss in der Auswertung führen kann.

In zwei Fällen lag eine Kombination aus 100% „majority_strict“ (MACE-Daten) und eine gemischte Bewertung (RNASeq-Daten) vor (Abb. 11). Dies betraf den Contig14304 (unseren Riba1-Kandidaten). Eine ausschließliche Kategorisierung nach „homozygous_strict“ hätte in diesem Fall zum Ausschluß und ggf. Datenverlust (?) geführt.

Abo-Vici - Abschlußbericht - RLP AgroScience GmbH

Züchtung und Agronomie neuartiger, Vicin-arter Ackerbohnen und Einsatz als einheimisches Eiweißfutter
 Förderkennzeichen: 2815EPS066 23 April 2021

Genotyp	gene_id	position	SNP-Nucleotid		Basen - counts				snp_type	Region (+10C Kongruenz)	Detection	
			Referenz	Variante	Low VC [4u5u6]	High VC [4u5u6]	Referenz	Variante				Referenz
247	Contig14304	68	T	C	1622	56	47	5110	majority_strict	1850001	✓	MACE-2019
248	Contig14304	68	T	C	2110	50	47	5110	majority_strict	1850001	✓	MACE-2019
BC3	Contig14304	68	T	C	1983	98	21	4078	majority_strict	1850001	✓	MACE-2019
247	Contig14304	68	T	C	1439	7	7	3529	homozygous_strict	1850001	✓	RNASeq-2019
248	Contig14304	68	T	C	956	8	7	3529	homozygous_strict	1850001	✓	RNASeq-2019
BC3	Contig14304	68	T	C	1263	1	23	5841	majority_strict	1850001	✓	RNASeq-2019
247	Contig14304	216	A	T	9	279	2217	12	majority_strict	1850001	✓	MACE-2019
248	Contig14304	216	A	T	19	516	2217	12	majority_strict	1850001	✓	MACE-2019
BC3	Contig14304	216	A	T	18	241	1094	5	majority_strict	1850001	✓	MACE-2019
247	Contig14304	216	A	T	41	6000	7890	50	homozygous_strict	1850001	✓	RNASeq-2019
248	Contig14304	216	A	T	34	3907	7890	50	majority_strict	1850001	✓	RNASeq-2019
BC3	Contig14304	216	A	T	23	5279	7892	32	homozygous_strict	1850001	✓	RNASeq-2019

Abbildung 11: RNASeq&MACE - redundante SNP-Kandidaten für alle Genotypen 2019

In einem Fall gaben die MACE-Daten eine einheitliche Bewertung als „homozygous_strict“ während die RNASeq-Daten eine gemischte Bewertung lieferten (Abb.12).

Genotyp	gene_id	position	SNP-Nucleotid		Basen - counts				snp_type	Region (+10C Kongruenz)	Detection	
			Referenz	Variante	Low VC [4u5u6]	High VC [4u5u6]	Referenz	Variante				Referenz
247	Contig3200_2	306	G	A	20	0	0	16	homozygous_strict	1930001	✓	MACE-2019
248	Contig3200_2	306	G	A	33	0	0	16	homozygous_strict	1930001	✓	MACE-2019
BC3	Contig3200_2	306	G	A	26	0	0	26	homozygous_strict	1930001	✓	MACE-2019
247	Contig3200_2	306	G	A	53	0	0	19	homozygous_strict	1930001	✓	RNASeq-2019
248	Contig3200_2	306	G	A	38	1	0	19	majority_strict	1930001	✓	RNASeq-2019
BC3	Contig3200_2	306	G	A	37	0	0	29	homozygous_strict	1930001	✓	RNASeq-2019

Abbildung 12: RNASeq&MACE - redundante SNP-Kandidaten für alle Genotypen 2019

Zwei SNP-Kandidaten wurden übereinstimmend als „homozygous_strict“ sowohl über MACE als auch über RNASeq identifiziert (Abb.13). Für Contig20221_2 -Pos. 1444 liegen die Zählwerte bei beiden Methoden in der gleichen Größenordnung. Bei Contig4170-Pos.1685 weist die Statistik deutlich höhere Werte für die RNASeq-Daten aus. Das illustriert noch einmal, wie unterschiedliche Methoden zu unterschiedlichen Daten am gleichen RNA-Template führen können.

Genotyp	gene_id	position	SNP-Nucleotid		Basen - counts				snp_type	Region (+10C Kongruenz)	Detection	
			Referenz	Variante	Low VC [4u5u6]	High VC [4u5u6]	Referenz	Variante				Referenz
247	Contig20221_2	1444	A	G	0	17	17	0	homozygous_strict	1540001	✓	MACE-2019
248	Contig20221_2	1444	A	G	0	17	17	0	homozygous_strict	1540001	✓	MACE-2019
BC3	Contig20221_2	1444	A	G	0	38	31	0	homozygous_strict	1540001	✓	MACE-2019
247	Contig20221_2	1444	A	G	0	43	34	0	homozygous_strict	1540001	✓	RNASeq-2019
248	Contig20221_2	1444	A	G	0	21	34	0	homozygous_strict	1540001	✓	RNASeq-2019
BC3	Contig20221_2	1444	A	G	0	27	39	0	homozygous_strict	1540001	✓	RNASeq-2019
247	Contig4170	1685	C	G	19	0	0	32	homozygous_strict	1950001	✓	MACE-2019
248	Contig4170	1685	C	G	39	0	0	32	homozygous_strict	1950001	✓	MACE-2019
BC3	Contig4170	1685	C	G	9	0	0	15	homozygous_strict	1950001	✓	MACE-2019
247	Contig4170	1685	C	G	346	0	0	174	homozygous_strict	1950001	✓	RNASeq-2019
248	Contig4170	1685	C	G	223	0	0	174	homozygous_strict	1950001	✓	RNASeq-2019
BC3	Contig4170	1685	C	G	260	1	0	275	homozygous_strict	1950001	✓	RNASeq-2019

Abbildung 13: RNASeq&MACE - redundante SNP-Kandidaten für alle Genotypen 2019

Zwei Chr.2-lokalisierbare SNP-Kandidaten fanden sich nur in den MACE-Daten (Abb. 14). Contig10165_3 mit nur einer detektierten SNP-Position innerhalb seiner Sequenz wurde ausschließlich durch MACE-Daten identifiziert, für Contig18427 wurden zusätzlich drei weitere SNP-Kandidaten (Pos. 1435, 1634 und 2300) über die RNASeq-Daten gefunden.

Abo-Vici - Abschlußbericht - RLP AgroScience GmbH

Züchtung und Agronomie neuartiger, Vicin-armer Ackerbohnen und Einsatz als einheimisches Eiweißfutter
 Förderkennzeichen: 2815EPS066 23 April 2021

Genotyp	gene_id	position	SNP-Nucleotid		Basen - counts				snp_type	Region (+10C Kongruenz)	Detection	
			Referenz	Variante	Low VC [4u5u6]	High VC [4u5u6]	Referenz	Variante				Referenz
247	Contig10165_3	1737	A	G	0	6	16	0	homozygous_strict	1350001	✓	MACE-2019
248	Contig10165_3	1737	A	G	0	6	16	0	homozygous_strict	1350001	✓	MACE-2019
BC3	Contig10165_3	1737	A	G	0	8	24	0	homozygous_strict	1350001	✓	MACE-2019
247	Contig18427	2892	C	T	0	38	50	0	homozygous_strict	1740001	✓	MACE-2019
248	Contig18427	2892	C	T	0	50	50	0	homozygous_strict	1740001	✓	MACE-2019
BC3	Contig18427	2892	C	T	0	54	43	2	majority_strict	1740001	✓	MACE-2019

Abbildung 14: MACE - SNP-Kandidaten für alle Genotypen 2019 nicht in RNASeq-Daten zu finden

Bei den 33 nicht-kartierbaren SNP-Kandidaten („NA“, vgl. Abb.9) findet sich eine ähnliche Situation (vgl. Abb.10) bzgl. der 3 redundanten, d.h. in MACE-Daten und RNASeq-Daten gefundenen SNPs (Abb.15), die sich ausschließlich innerhalb eines Contigs befinden. MACE lieferte die Kategorie „majority_strict“, RNASeq dagegen „homozygous_strict“. Auffällig hier, dass die Zählerwerte für MACE-Datenpunkte teilweise doppelt bis dreifach höher liegen als für RNASeq, was wiederum gegen die schon zuvor angesprochene Sequenziertiefe als Ursache für Abweichungen vom theoretischen Nullwert nach oben für die entsprechende Base spricht.

Genotyp	gene_id	position	SNP-Nucleotid		Basen - counts				snp_type	Region (+10C Kongruenz)	Detection	
			Referenz	Variante	Low VC [4u5u6]	High VC [4u5u6]	Referenz	Variante				Referenz
247	TRINITY_DN44477_c0_g1_i2	444	C	A	61	775	1422	35	majority_strict	nan	✓	MACE-2019
248	TRINITY_DN44477_c0_g1_i2	444	C	A	68	758	1422	35	majority_strict	nan	✓	MACE-2019
BC3	TRINITY_DN44477_c0_g1_i2	444	C	A	85	1374	1207	104	majority_strict	nan	✓	MACE-2019
247	TRINITY_DN44477_c0_g1_i2	444	C	A	0	432	994	1	homozygous_strict	nan	✓	RNASeq-2019
248	TRINITY_DN44477_c0_g1_i2	444	C	A	2	317	994	1	homozygous_strict	nan	✓	RNASeq-2019
BC3	TRINITY_DN44477_c0_g1_i2	444	C	A	1	485	432	2	homozygous_strict	nan	✓	RNASeq-2019
247	TRINITY_DN44477_c0_g1_i2	593	T	C	50	2309	3032	54	majority_strict	nan	✓	MACE-2019
248	TRINITY_DN44477_c0_g1_i2	593	T	C	42	2611	3032	54	majority_strict	nan	✓	MACE-2019
BC3	TRINITY_DN44477_c0_g1_i2	593	T	C	62	3735	1752	139	majority_strict	nan	✓	MACE-2019
247	TRINITY_DN44477_c0_g1_i2	593	T	C	3	918	1450	3	homozygous_strict	nan	✓	RNASeq-2019
248	TRINITY_DN44477_c0_g1_i2	593	T	C	3	676	1450	3	homozygous_strict	nan	✓	RNASeq-2019
BC3	TRINITY_DN44477_c0_g1_i2	593	T	C	0	984	686	6	homozygous_strict	nan	✓	RNASeq-2019
247	TRINITY_DN44477_c0_g1_i2	668	C	T	2630	18	42	3371	majority_strict	nan	✓	MACE-2019
248	TRINITY_DN44477_c0_g1_i2	668	C	T	3005	22	42	3371	majority_strict	nan	✓	MACE-2019
BC3	TRINITY_DN44477_c0_g1_i2	668	C	T	3800	16	45	1696	majority_strict	nan	✓	MACE-2019
247	TRINITY_DN44477_c0_g1_i2	668	C	T	2237	1	14	2998	homozygous_strict	nan	✓	RNASeq-2019
248	TRINITY_DN44477_c0_g1_i2	668	C	T	1758	4	14	2998	homozygous_strict	nan	✓	RNASeq-2019
BC3	TRINITY_DN44477_c0_g1_i2	668	C	T	2619	1	10	1626	homozygous_strict	nan	✓	RNASeq-2019

Abbildung 15: RNASeq&MACE - redundante SNP-Kandidaten für alle Genotypen 2019 - „NA“

Ein einziger SNP-Kandidat, der ausschließlich in den RNASeq-Daten enthalten war, hatte durchgehend die Bewertung „majority_strict“ (Abb.16).

Genotyp	gene_id	position	SNP-Nucleotid		Basen - counts				snp_type	Region (+10C Kongruenz)	Detection	
			Referenz	Variante	Low VC [4u5u6]	High VC [4u5u6]	Referenz	Variante				Referenz
247	TRINITY_DN49974_c1_g1_i2	68	A	G	489	5	8	2599	majority_strict	nan	✓	RNASeq-2019
248	TRINITY_DN49974_c1_g1_i2	68	A	G	290	3	8	2599	majority_strict	nan	✓	RNASeq-2019
BC3	TRINITY_DN49974_c1_g1_i2	68	A	G	455	1	15	3912	majority_strict	nan	✓	RNASeq-2019

Abbildung 16: RNASeq only - SNP-Kandidat für alle Genotypen 2019 - „majority_strict“

Vier reine RNASeq SNP-Kandidaten (in 2 Contigs) weisen eine gemischte Kategorisierung auf (Abb. 17). Diese und der vorgenannte SNP-Kandidat wären somit bei Filtern auf „homozygous_strict“ nicht in der Auswertung aufgetaucht.

Abo-Vici - Abschlußbericht - RLP AgroScience GmbH

Züchtung und Agronomie neuartiger, Vicin-armer Ackerbohnen und Einsatz als einheimisches Eiweißfutter
 Förderkennzeichen: 2815EPS066 23 April 2021

Genotyp	gene_id	position	SNP-Nucleotid		Basen - counts				snp_type	Region (+10C Kongruenz)	Detection	
			Referenz	Variante	Referenz	Variante	Referenz	Variante				Referenz
247	Contig17671	321	C	T	1	161	102	0	homozygous_strict	nan	✓	RNASeq-2019
248	Contig17671	321	C	T	0	85	102	0	homozygous_strict	nan	✓	RNASeq-2019
BC3	Contig17671	321	C	T	1	110	97	1	majority_strict	nan	✓	RNASeq-2019
247	Contig17671	760	T	A	89	0	0	44	homozygous_strict	nan	✓	RNASeq-2019
248	Contig17671	760	T	A	39	0	0	44	homozygous_strict	nan	✓	RNASeq-2019
BC3	Contig17671	760	T	A	49	0	1	42	majority_strict	nan	✓	RNASeq-2019
247	Contig17671	764	G	T	93	1	0	44	majority_strict	nan	✓	RNASeq-2019
248	Contig17671	764	G	T	39	0	0	44	homozygous_strict	nan	✓	RNASeq-2019
BC3	Contig17671	764	G	T	52	0	0	45	homozygous_strict	nan	✓	RNASeq-2019
247	TRINITY_DN44477_c0_g1_i2	1522	C	G	106	0	1	139	homozygous_strict	nan	✓	RNASeq-2019
248	TRINITY_DN44477_c0_g1_i2	1522	C	G	61	0	1	139	majority_strict	nan	✓	RNASeq-2019
BC3	TRINITY_DN44477_c0_g1_i2	1522	C	G	66	0	0	60	homozygous_strict	nan	✓	RNASeq-2019

Abbildung 17: RNASeq only - SNP-Kandidaten für alle Genotypen 2019 - „mixed“

Die restlichen 24 SNP-Kandidaten, die ausschließlich in den RNASeq-Daten zu finden waren, zählen durchgehend in Kategorie „homozygous_strict“. Ein einziger SNP-Kandidat findet sich ausschließlich in den MACE-Daten und besteht ebenfalls nur aus „homozygous_strict“ Bewertungen für die drei Genotypen-Paare (Abb. 18).

Genotyp	gene_id	position	SNP-Nucleotid		Basen - counts				snp_type	Region (+10C Kongruenz)	Detection	
			Referenz	Variante	Referenz	Variante	Referenz	Variante				Referenz
247	Contig9400	743	C	T	13	0	0	5	homozygous_strict	nan	✓	MACE-2019
248	Contig9400	743	C	T	11	0	0	5	homozygous_strict	nan	✓	MACE-2019
BC3	Contig9400	743	C	T	11	0	0	10	homozygous_strict	nan	✓	MACE-2019

Abbildung 18: MACE only - SNP-Kandidat für alle Genotypen 2019 - „homozygous_strict“

Neubewertung der Daten aus dem Pflanzenmaterial von 2017

Die hier ausführlich dargelegten Auswertungen für das Material aus 2019 wurden für die beiden nah-isogenen Linienpaare (Göx „139“ und NPZx „848“; Material aus 2017) ebenfalls um die Kategorie „majority_strict“ erweitert und die Rohdaten wurden daher analog einer Neubewertung unterzogen. Nur so war eine Synopsis aller gemeinsamen SNP-Kandidaten aus dem verwendeten Pflanzenmaterial beider Transkriptomanalysen über den gesamten Projektverlauf sinnvollerweise möglich.

RNASeq & MACE: 2017			
Göx139 & NPZx848			
	Chr. 2-wide	Chr.2-narrow	NA
von	390,001	1,860,001	
bis	3,620,001	2,500,001	
homozygous strict	431	135	60
majority strict	9	2	2
mixed	74	10	12
Σ (Hs&Ms):	514	147	74
diff. Contigs:	128	36	28

Abbildung 19: RNASeq&MACE - nicht-redundante SNP-Kandidaten für alle Genotypen 2017

Synthese aus allen Datensätzen des Pflanzenmaterials von 2017 & 2019

Im Vergleich zu der Gesamtanalyse für das Pflanzenmaterial aus 2019 (s. Abb.9) ist die Ausbeute an potentiellen SNP-Kandidaten deutlich höher. Ein Blick auf die Koordinaten für Chr.2 (i.e. die SNP-Sequenzen, die auf Chr.2 von *M.truncatula* kartiert werden konnten) zeigt, dass es zwischen dem Pflanzenmaterial aus 2017 (zwei Paaren von nah-isogenen Geschwisterpflanzen) und dem aus 2019 (zwei kontrastierende Backcross-Bulks und drei hoch-isogene Linien) merkliche Unterschiede gibt. Die Chr.2-lokalisierbaren SNP-Kandidaten streuen im

analysierten Pflanzenmaterial von 2017 im fraglichen Intervall deutlich breiter (Abb. 19, „von - bis“) als in dem Pflanzenmaterial von 2019 (Abb. 9 „von - bis“). Damit erweist sich die von Prof. W. Link verfolgte Strategie als aussichtsreich, über möglichst unterschiedliche genetische Hintergründe hinweg (bedingt durch unabhängige Rekombinationsereignisse um den VC-Locus herum) auf das gemeinsame Selektionskriterium, den Locus für VC-Armut, zu „filtern“ und so die Anzahl an SNP-Marker auf ein Minimum einzuengen. SNP-Kandidaten, die über alle betrachteten unterschiedlichen V/C-Genotypen hinweg in einem möglichst engen Bereich auf dem syntänen Abschnitt von Chr.2 von Medicago liegen, sollten von besonderem Interesse sein, aber auch diejenigen allen gemeinsamen SNP-Sequenzen, die sich nicht auf dem Medicago-Genom kartieren lassen. Die Schnittmengenbildung über alle Genotypen führte zu einer deutlichen Reduzierung der Anzahl an SNP-Kandidaten (Abb.20) gegenüber den separaten Vergleichen von 2017 (s.o. Abb. 19) und 2019 (s.o. Abb. 9).

RNASeq & MACE: 2017 & 2019				
Göx139 & NPZx848 & BC3 & Sixdenier				
	Chr. 2-wide	Chr.2-narrow	nicht kartierbar	Σ Chr2-wide & NA
von	1,320,001	1,860,001	"NA"	
bis	1,980,001	1,980,001		
"Hs"-SNPs	45	6	16	61
"Ms"-SNPs	2	0	0	2
mixed-SNPs	24	4	4	28
Σ (Hs&Ms):	71	10	20	91
diff. Contigs:	28	5	4	32

Abbildung 20: RNASeq&MACE - Σ aller SNP-Kandidaten über alle Genotypen 2017 & 2019

Im Vergleich mit der Anzahl an potentiellen SNP-Kandidaten aus der ersten Projektphase (Abb. 21) ergibt sich eine drastische Reduktion auf knapp 14% für Chr.2 kartierbare SNP-Kandidaten sowie auf 27% für nicht-kartierbare SNP-Kandidaten.

2017 => Schnittmenge 2017&2019			
	Chr. 2-wide	Chr.2-narrow	nicht kartierbar
von	1,320,001	1,860,001	"NA"
bis	1,980,001	1,980,001	
"Hs"-SNPs	10.4%	4.4%	26.7%
"Ms"-SNPs	22.2%	0.0%	0.0%
mixed-SNPs	32.4%	40.0%	33.3%
Σ (Hs&Ms):	13.8%	6.8%	27.0%
diff. Contigs:	21.9%	13.9%	14.3%

Abbildung 21: RNASeq&MACE - %-Anteil an finalen SNP-Kandidaten über alle Genotypen im Vergleich zu Daten von 2017

Im Vergleich dazu ist die Schnittmenge mit den SNP-Daten des Pflanzenmaterials der zweiten Analysenrunde in 2019 größer (Abb. 22). Das liegt schon darin begründet, dass hier im Vergleich zu 2017 „nur“ ca. 25% an SNP-Kandidaten (i.e. Chr.2 lokalisierbare und „nicht auf Medicago kartierbare SNP-Kandidaten) und zugehörige Contigs insgesamt identifiziert wurden. D.h. dass ca. 60% der SNP-Kandidaten aus 2019 auch in die Schnittmenge des gesamten Pflanzenmaterials aus 2017 und 2019 fallen.

Abo-Vici - Abschlußbericht - RLP AgroScience GmbH

Züchtung und Agronomie neuartiger, Vicin-arter Ackerbohnen und Einsatz als einheimisches Eiweißfutter
 Förderkennzeichen: 2815EPS066 23 April 2021

		2019 => Schnittmenge 2017&2019		
		Chr. 2-wide	Chr.2-narrow	nicht kartierbar
von		1,320,001	1,860,001	"NA"
bis		1,980,001	1,980,001	
	"Hs"-SNPs	55.6%	42.9%	57.1%
	"Ms"-SNPs	20.0%	0.0%	0.0%
	mixed-SNPs	88.9%	80.0%	100.0%
	Σ (Hs&Ms):	60.2%	52.6%	60.6%
	diff. Contigs:	80.0%	71.4%	66.7%

Abbildung 22: RNASeq&MACE - %-Anteil an finalen SNP-Kandidaten über alle Genotypen im Vergleich zu Daten von 2019

Betrachtet man -wie zu Eingang erläutert- die durch die Auswertung der Kategorie „majority_strict“ hinzugewonnenen SNP-Kandidaten (vgl. S.9), so ergibt sich bei insgesamt 91 SNP-Kandidaten in 32 Contigs eine ca. 33% höhere „Ausbeute“ (Abb. 23, vgl. dazu Abb.21). Der größte Zugewinn liegt bei den Chr.2-lokalisierbaren SNP-Kandidaten.

		Durch "Ms"-SNPs zusätzlich erzeugte SNP-Kandidaten			
		Chr. 2-wide	Chr.2-narrow	"NA"	Σ Chr.2&NA
"+" SNPs		26	4	4	30
		36.6%	40.0%	20.0%	33.0%
"+" Contigs		5	1	0	5
		17.9%	20.0%	0.0%	15.6%

Abbildung 23: RNASeq&MACE - Zuwachs durch Einbeziehung der Kategorie „majority_strict“

Die Zahl der Contigs ändert sich dabei nur in geringem Umfang (Abb. 24). D.h. die zusätzlichen SNP-Kandidaten finden sich in schon vorher durch SNPs charakterisierten Contigs.

		RNASeq & MACE: 2017 & 2019			
		Göx139 & NPZx848 & BC3 & Sixdenier			
		Chr. 2 -wide	Chr. 2 -narrow	NA	Σ Chr2-wide & NA
von		1,320,001	1,930,001	NA	Σ Chr2-wide & NA
bis		1,980,001	1,980,001		
	homozygot strict	45	6	16	61
	different contigs	23	4	4	27

Abbildung 24: RNASeq&MACE-„homozygot_strict“ SNP-Kandidaten über alle Genotypen 2017 & 2019

Außerdem fallen letztlich nur zwei SNP-Kandidaten unter die Kategorie „Majority_strict“, die restlichen 29 SNP-Kandidaten sind Hybride bzw. „mixed“, was die Wahrscheinlichkeit erhöht, dass es sich hier i.d.R. um die schon o.g. technischen Artefakte handelt und sich darunter eigentlich SNP-Kandidaten der „besten“ Kategorie „homozygot_strict“ verbergen. Beispielhaft ist dies an Contig3485 für die SNP-Position 252 zu erkennen (Abb. 25). Im Genotyp Sixdenier 247 und Sixdenier 248 liefern die MACE- und die RNASeq-Daten unterschiedliche Bewertungen, ebenso für den Backcross3. In allen drei Fällen kann man zugunsten der besseren Bewertung entscheiden und damit eine durchgängige Bewertung als „homozygot_strict“ für den SNP-Kandidaten annehmen.

Abo-Vici - Abschlußbericht - RLP AgroScience GmbH

Züchtung und Agronomie neuartiger, Vicin-arter Ackerbohnen und Einsatz als einheimisches Eiweißfutter
 Förderkennzeichen: 2815EPS066 23 April 2021

Genotyp	gene_id	description	position	SNP-Nucleotid		Basen - counts				snp_type	Region (+10000bp)	Detection
				Referenz	Variante	Low [4u5u6]		High [4u5u6]				
						Referenz	Variante	Referenz	Variante			
139	Contig3485	G7INT9_MEDTR Chlorophyll a-b bir	252	A	G	1235	0	3	745	homozygous_strict	1,550,001	RNASeq-2017
247	Contig3485	G7INT9_MEDTR Chlorophyll a-b bir	252	A	G	947	21	21	844	majority_strict	1,550,001	MACE-2019
247	Contig3485	G7INT9_MEDTR Chlorophyll a-b bir	252	A	G	505	1	0	445	homozygous_strict	1,550,001	RNASeq-2019
248	Contig3485	G7INT9_MEDTR Chlorophyll a-b bir	252	A	G	915	11	21	844	majority_strict	1,550,001	MACE-2019
248	Contig3485	G7INT9_MEDTR Chlorophyll a-b bir	252	A	G	317	3	0	445	homozygous_strict	1,550,001	RNASeq-2019
848	Contig3485	G7INT9_MEDTR Chlorophyll a-b bir	252	A	G	804	0	0	606	homozygous_strict	1,550,001	RNASeq-2017
139&848	Contig3485	G7INT9_MEDTR Chlorophyll a-b bir	252	A	G	1479	8	8	829	homozygous_strict	1,550,001	MACE-2017
BC3	Contig3485	G7INT9_MEDTR Chlorophyll a-b bir	252	A	G	2127	35	36	1512	majority_strict	1,550,001	MACE-2019
BC3	Contig3485	G7INT9_MEDTR Chlorophyll a-b bir	252	A	G	810	3	2	641	homozygous_strict	1,550,001	RNASeq-2019
139	Contig14304	G7IPL5_MEDTR 3,4-dihydroxy-2-bu	216	A	T	29	6125	7851	5	homozygous_strict	1,850,001	RNASeq-2017
247	Contig14304	G7IPL5_MEDTR 3,4-dihydroxy-2-bu	216	A	T	9	279	2217	12	majority_strict	1,850,001	MACE-2019
247	Contig14304	G7IPL5_MEDTR 3,4-dihydroxy-2-bu	216	A	T	41	6000	7890	50	homozygous_strict	1,850,001	RNASeq-2019
248	Contig14304	G7IPL5_MEDTR 3,4-dihydroxy-2-bu	216	A	T	19	516	2217	12	majority_strict	1,850,001	MACE-2019
248	Contig14304	G7IPL5_MEDTR 3,4-dihydroxy-2-bu	216	A	T	34	3907	7890	50	majority_strict	1,850,001	RNASeq-2019
848	Contig14304	G7IPL5_MEDTR 3,4-dihydroxy-2-bu	216	A	T	53	6982	7756	10	homozygous_strict	1,850,001	RNASeq-2017
BC3	Contig14304	G7IPL5_MEDTR 3,4-dihydroxy-2-bu	216	A	T	18	241	1094	5	majority_strict	1,850,001	MACE-2019
BC3	Contig14304	G7IPL5_MEDTR 3,4-dihydroxy-2-bu	216	A	T	23	5279	7892	32	homozygous_strict	1,850,001	RNASeq-2019

Abbildung 25: RNASeq&MACE - Beispiel für Chr.2 kartierte SNP-Kandidaten Kategorie „majority_strict“

Für den Contig14304 in SNP-Position 216 lässt sich für zwei Genotypen (Sixdenier 247 und BC3) die Bewertung „majority_strict“ auf „homozygot_strict“ korrigieren, für den Genotyp Sixdenier 248 dagegen bleibt die Kategorisierung für beide Sequenziermethoden bei „majority_strict“.

Genotyp	gene_id	description	position	SNP-Nucleotid		Basen - counts				snp_type	Region (+10000bp)	Detection
				Referenz	Variante	Low [4u5u6]		High [4u5u6]				
						Referenz	Variante	Referenz	Variante			
139	TRINITY_DN44477_c0_g1_i2	AOA1S2XB2_CICAR non-specific li	444	C	A	12	934	1622	4	homozygous_strict	nan	MACE-2017
139	TRINITY_DN44477_c0_g1_i2	AOA1S2XB2_CICAR non-specific li	444	C	A	1	467	1161	3	homozygous_strict	nan	RNASeq-2017
247	TRINITY_DN44477_c0_g1_i2	AOA1S2XB2_CICAR non-specific li	444	C	A	61	775	1422	35	majority_strict	nan	MACE-2019
247	TRINITY_DN44477_c0_g1_i2	AOA1S2XB2_CICAR non-specific li	444	C	A	0	432	994	1	homozygous_strict	nan	RNASeq-2019
248	TRINITY_DN44477_c0_g1_i2	AOA1S2XB2_CICAR non-specific li	444	C	A	68	758	1422	35	majority_strict	nan	MACE-2019
248	TRINITY_DN44477_c0_g1_i2	AOA1S2XB2_CICAR non-specific li	444	C	A	2	317	994	1	homozygous_strict	nan	RNASeq-2019
848	TRINITY_DN44477_c0_g1_i2	AOA1S2XB2_CICAR non-specific li	444	C	A	10	586	995	6	majority_strict	nan	MACE-2017
848	TRINITY_DN44477_c0_g1_i2	AOA1S2XB2_CICAR non-specific li	444	C	A	1	347	1128	0	homozygous_strict	nan	RNASeq-2017
BC3	TRINITY_DN44477_c0_g1_i2	AOA1S2XB2_CICAR non-specific li	444	C	A	85	1374	1207	104	majority_strict	nan	MACE-2019
BC3	TRINITY_DN44477_c0_g1_i2	AOA1S2XB2_CICAR non-specific li	444	C	A	1	485	432	2	homozygous_strict	nan	RNASeq-2019

Abbildung 26: RNASeq&MACE - Bsp. für nicht-kartierbaren SNP-Kandidaten Kategorie „majority_strict“

Das gleiche Phänomen findet sich auch bei den SNP-Kandidaten, die über Sequenzvergleich nicht auf dem *Medicago*-Genom allokiert werden können (Abb. 26). Beispielweise verwandeln sich für Contig TRINITY_DN44477_c0_g1_i2 alle in den MACE-Daten als „majority_strict“ bewerteten SNPs für Position 444 in den korrespondierenden RNASeq-Daten zu „homozygous_strict“.

D.h. bei gleichzeitig vorliegender Information für eine SNP-Position aus beiden Sequenziermethoden (MACE und RNASeq) kann häufig eine höherwertige Kategorisierung von „majority_strict“ zu „homozygous_strict“ erfolgen. Davon betroffen waren 4 SNP-Kandidaten aus 2 Contigs in der Gruppe der nicht-lokalisierbaren Contigs (Abb. 27) und 7 SNP-Kandidaten in der Gruppe der Chr.2-lokalisierbaren Contigs (Abb. 28).

Abo-Vici - Abschlußbericht - RLP AgroScience GmbH

Züchtung und Agronomie neuartiger, Vicin-armer Ackerbohnen und Einsatz als einheimisches Eiweißfutter

Förderkennzeichen: 2815EPS066

23 April 2021

Genotyp	gene_id	description	position	SNP-Nucleotid		Basen - counts				snp_type	Region (+1000bp)	Detection
				Referenz	Variante	Low [4u5u6]		High [4u5u6]				
						Referenz	Variante	Referenz	Variante			
139	Contig17671	G7L410_MEDTR Heavy-metal-assoc	761	A	T	148	0	1	68	homozygous_strict	nan	MACE-2017
139	Contig17671	G7L410_MEDTR Heavy-metal-assoc	761	A	T	107	2	0	96	majority_strict	nan	RNASeq-2017
247	Contig17671	G7L410_MEDTR Heavy-metal-assoc	761	A	C	88	0	0	44	homozygous_strict	nan	RNASeq-2019
248	Contig17671	G7L410_MEDTR Heavy-metal-assoc	761	A	C	39	0	0	44	homozygous_strict	nan	RNASeq-2019
848	Contig17671	G7L410_MEDTR Heavy-metal-assoc	761	A	C	106	2	1	40	majority_strict	nan	MACE-2017
848	Contig17671	G7L410_MEDTR Heavy-metal-assoc	761	A	C	93	0	0	72	homozygous_strict	nan	RNASeq-2017
BC3	Contig17671	G7L410_MEDTR Heavy-metal-assoc	761	A	T	49	0	0	41	homozygous_strict	nan	RNASeq-2019
139	TRINITY_DN44477_c0_g1_i2	AQA1S2XBU2_CICAR non-specific li	444	C	A	12	934	1622	4	homozygous_strict	nan	MACE-2017
139	TRINITY_DN44477_c0_g1_i2	AQA1S2XBU2_CICAR non-specific li	444	C	A	1	467	1161	3	homozygous_strict	nan	RNASeq-2017
247	TRINITY_DN44477_c0_g1_i2	AQA1S2XBU2_CICAR non-specific li	444	C	A	61	775	1422	35	majority_strict	nan	MACE-2019
247	TRINITY_DN44477_c0_g1_i2	AQA1S2XBU2_CICAR non-specific li	444	C	A	0	432	994	1	homozygous_strict	nan	RNASeq-2019
248	TRINITY_DN44477_c0_g1_i2	AQA1S2XBU2_CICAR non-specific li	444	C	A	68	758	1422	35	majority_strict	nan	MACE-2019
248	TRINITY_DN44477_c0_g1_i2	AQA1S2XBU2_CICAR non-specific li	444	C	A	2	317	994	1	homozygous_strict	nan	RNASeq-2019
848	TRINITY_DN44477_c0_g1_i2	AQA1S2XBU2_CICAR non-specific li	444	C	A	10	586	995	6	majority_strict	nan	MACE-2017
848	TRINITY_DN44477_c0_g1_i2	AQA1S2XBU2_CICAR non-specific li	444	C	A	1	347	1128	0	homozygous_strict	nan	RNASeq-2017
BC3	TRINITY_DN44477_c0_g1_i2	AQA1S2XBU2_CICAR non-specific li	444	C	A	85	1374	1207	104	majority_strict	nan	MACE-2019
BC3	TRINITY_DN44477_c0_g1_i2	AQA1S2XBU2_CICAR non-specific li	444	C	A	1	485	432	2	homozygous_strict	nan	RNASeq-2019
139	TRINITY_DN44477_c0_g1_i2	AQA1S2XBU2_CICAR non-specific li	593	T	C	10	2725	3394	7	homozygous_strict	nan	MACE-2017
139	TRINITY_DN44477_c0_g1_i2	AQA1S2XBU2_CICAR non-specific li	593	T	C	2	766	1389	2	homozygous_strict	nan	RNASeq-2017
247	TRINITY_DN44477_c0_g1_i2	AQA1S2XBU2_CICAR non-specific li	593	T	C	50	2309	3032	54	majority_strict	nan	MACE-2019
247	TRINITY_DN44477_c0_g1_i2	AQA1S2XBU2_CICAR non-specific li	593	T	C	3	918	1450	3	homozygous_strict	nan	RNASeq-2019
248	TRINITY_DN44477_c0_g1_i2	AQA1S2XBU2_CICAR non-specific li	593	T	C	42	2611	3032	54	majority_strict	nan	MACE-2019
248	TRINITY_DN44477_c0_g1_i2	AQA1S2XBU2_CICAR non-specific li	593	T	C	3	676	1450	3	homozygous_strict	nan	RNASeq-2019
848	TRINITY_DN44477_c0_g1_i2	AQA1S2XBU2_CICAR non-specific li	593	T	C	12	1656	1713	26	majority_strict	nan	MACE-2017
848	TRINITY_DN44477_c0_g1_i2	AQA1S2XBU2_CICAR non-specific li	593	T	C	1	601	1261	0	homozygous_strict	nan	RNASeq-2017
BC3	TRINITY_DN44477_c0_g1_i2	AQA1S2XBU2_CICAR non-specific li	593	T	C	62	3735	1752	139	majority_strict	nan	MACE-2019
BC3	TRINITY_DN44477_c0_g1_i2	AQA1S2XBU2_CICAR non-specific li	593	T	C	0	984	686	6	homozygous_strict	nan	RNASeq-2019
139	TRINITY_DN44477_c0_g1_i2	AQA1S2XBU2_CICAR non-specific li	668	C	T	2107	6	4	2434	homozygous_strict	nan	MACE-2017
139	TRINITY_DN44477_c0_g1_i2	AQA1S2XBU2_CICAR non-specific li	668	C	T	2427	5	12	3897	homozygous_strict	nan	RNASeq-2017
247	TRINITY_DN44477_c0_g1_i2	AQA1S2XBU2_CICAR non-specific li	668	C	T	2630	18	42	3371	majority_strict	nan	MACE-2019
247	TRINITY_DN44477_c0_g1_i2	AQA1S2XBU2_CICAR non-specific li	668	C	T	2237	1	14	2998	homozygous_strict	nan	RNASeq-2019
248	TRINITY_DN44477_c0_g1_i2	AQA1S2XBU2_CICAR non-specific li	668	C	T	3005	22	42	3371	majority_strict	nan	MACE-2019
248	TRINITY_DN44477_c0_g1_i2	AQA1S2XBU2_CICAR non-specific li	668	C	T	1758	4	14	2998	homozygous_strict	nan	RNASeq-2019
848	TRINITY_DN44477_c0_g1_i2	AQA1S2XBU2_CICAR non-specific li	668	C	T	1280	10	17	1422	majority_strict	nan	MACE-2017
848	TRINITY_DN44477_c0_g1_i2	AQA1S2XBU2_CICAR non-specific li	668	C	T	1926	1	4	2679	homozygous_strict	nan	RNASeq-2017
BC3	TRINITY_DN44477_c0_g1_i2	AQA1S2XBU2_CICAR non-specific li	668	C	T	3800	16	45	1696	majority_strict	nan	MACE-2019
BC3	TRINITY_DN44477_c0_g1_i2	AQA1S2XBU2_CICAR non-specific li	668	C	T	2619	1	10	1626	homozygous_strict	nan	RNASeq-2019

Abbildung 27: RNASeq&MACE - Chr.2-kartierbare SNP-Kandidaten Kategorie „homozygous_strict“ >> „majority_strict“

Abo-Vici - Abschlußbericht - RLP AgroScience GmbH

Züchtung und Agronomie neuartiger, Vicin-armer Ackerbohnen und Einsatz als einheimisches Eiweißfutter
 Förderkennzeichen: 2815EPS066 23 April 2021

Genotyp	gene_id	description	position	SNP-Nucleotid		Basen - counts				snp_type	Region (+1000bp)
				Referenz	Variante	Low [4u5u6]		High [4u5u6]			
						Referenz	Variante	Referenz	Variante		
139	Contig20221_2	G7INT7_MEDTR Spolie-like protein	1475	C	T	0	32	68	0	homozygous_strict	1,540,001
247	Contig20221_2	G7INT7_MEDTR Spolie-like protein	1475	C	T	0	20	21	0	homozygous_strict	1,540,001
247	Contig20221_2	G7INT7_MEDTR Spolie-like protein	1475	C	T	0	40	27	0	homozygous_strict	1,540,001
248	Contig20221_2	G7INT7_MEDTR Spolie-like protein	1475	C	T	0	18	21	0	homozygous_strict	1,540,001
248	Contig20221_2	G7INT7_MEDTR Spolie-like protein	1475	C	T	0	18	27	0	homozygous_strict	1,540,001
848	Contig20221_2	G7INT7_MEDTR Spolie-like protein	1475	C	T	0	53	50	0	homozygous_strict	1,540,001
139&848	Contig20221_2	G7INT7_MEDTR Spolie-like protein	1475	C	T	1	29	33	0	homozygous_strict	1,540,001
BC3	Contig20221_2	G7INT7_MEDTR Spolie-like protein	1475	C	T	1	44	41	2	majority_strict	1,540,001
BC3	Contig20221_2	G7INT7_MEDTR Spolie-like protein	1475	C	T	0	28	27	0	homozygous_strict	1,540,001
139	Contig3485	G7INT9_MEDTR Chlorophyll a-b bir	252	A	G	1235	0	3	745	homozygous_strict	1,550,001
247	Contig3485	G7INT9_MEDTR Chlorophyll a-b bir	252	A	G	947	21	21	844	majority_strict	1,550,001
247	Contig3485	G7INT9_MEDTR Chlorophyll a-b bir	252	A	G	505	1	0	445	homozygous_strict	1,550,001
248	Contig3485	G7INT9_MEDTR Chlorophyll a-b bir	252	A	G	915	11	21	844	majority_strict	1,550,001
248	Contig3485	G7INT9_MEDTR Chlorophyll a-b bir	252	A	G	317	3	0	445	homozygous_strict	1,550,001
848	Contig3485	G7INT9_MEDTR Chlorophyll a-b bir	252	A	G	804	0	0	606	homozygous_strict	1,550,001
139&848	Contig3485	G7INT9_MEDTR Chlorophyll a-b bir	252	A	G	1479	8	8	829	homozygous_strict	1,550,001
BC3	Contig3485	G7INT9_MEDTR Chlorophyll a-b bir	252	A	G	2127	35	36	1512	majority_strict	1,550,001
BC3	Contig3485	G7INT9_MEDTR Chlorophyll a-b bir	252	A	G	810	3	2	641	homozygous_strict	1,550,001
139	Contig3485	G7INT9_MEDTR Chlorophyll a-b bir	309	G	A	1998	1	5	1062	homozygous_strict	1,550,001
247	Contig3485	G7INT9_MEDTR Chlorophyll a-b bir	309	G	A	498	4	11	442	majority_strict	1,550,001
247	Contig3485	G7INT9_MEDTR Chlorophyll a-b bir	309	G	A	762	0	0	671	homozygous_strict	1,550,001
248	Contig3485	G7INT9_MEDTR Chlorophyll a-b bir	309	G	A	554	7	11	442	majority_strict	1,550,001
248	Contig3485	G7INT9_MEDTR Chlorophyll a-b bir	309	G	A	476	3	0	671	homozygous_strict	1,550,001
848	Contig3485	G7INT9_MEDTR Chlorophyll a-b bir	309	G	A	1250	4	1	835	homozygous_strict	1,550,001
139&848	Contig3485	G7INT9_MEDTR Chlorophyll a-b bir	309	G	A	595	0	4	355	homozygous_strict	1,550,001
BC3	Contig3485	G7INT9_MEDTR Chlorophyll a-b bir	309	G	A	1080	9	3	732	majority_strict	1,550,001
BC3	Contig3485	G7INT9_MEDTR Chlorophyll a-b bir	309	G	A	1222	1	1	893	homozygous_strict	1,550,001
139	Contig4459_2	G7IPJ9_MEDTR Phospho-2-dehydro	1729	C	T	570	0	0	446	homozygous_strict	1,760,001
247	Contig4459_2	G7IPJ9_MEDTR Phospho-2-dehydro	1729	C	T	710	8	18	372	majority_strict	1,760,001
247	Contig4459_2	G7IPJ9_MEDTR Phospho-2-dehydro	1729	C	T	291	0	0	166	homozygous_strict	1,760,001
248	Contig4459_2	G7IPJ9_MEDTR Phospho-2-dehydro	1729	C	T	679	2	18	372	majority_strict	1,760,001
248	Contig4459_2	G7IPJ9_MEDTR Phospho-2-dehydro	1729	C	T	176	0	0	166	homozygous_strict	1,760,001
848	Contig4459_2	G7IPJ9_MEDTR Phospho-2-dehydro	1729	C	T	651	1	0	342	homozygous_strict	1,760,001
139&848	Contig4459_2	G7IPJ9_MEDTR Phospho-2-dehydro	1729	C	T	558	3	7	264	majority_strict	1,760,001
BC3	Contig4459_2	G7IPJ9_MEDTR Phospho-2-dehydro	1729	C	T	813	24	18	726	majority_strict	1,760,001
BC3	Contig4459_2	G7IPJ9_MEDTR Phospho-2-dehydro	1729	C	T	229	0	0	253	homozygous_strict	1,760,001
139	Contig4459_2	G7IPJ9_MEDTR Phospho-2-dehydro	1797	A	G	358	0	0	208	homozygous_strict	1,760,001
247	Contig4459_2	G7IPJ9_MEDTR Phospho-2-dehydro	1797	A	G	756	28	18	331	majority_strict	1,760,001
247	Contig4459_2	G7IPJ9_MEDTR Phospho-2-dehydro	1797	A	G	256	0	0	100	homozygous_strict	1,760,001
248	Contig4459_2	G7IPJ9_MEDTR Phospho-2-dehydro	1797	A	G	602	15	18	331	majority_strict	1,760,001
248	Contig4459_2	G7IPJ9_MEDTR Phospho-2-dehydro	1797	A	G	136	0	0	100	homozygous_strict	1,760,001
848	Contig4459_2	G7IPJ9_MEDTR Phospho-2-dehydro	1797	A	G	465	0	0	221	homozygous_strict	1,760,001
139&848	Contig4459_2	G7IPJ9_MEDTR Phospho-2-dehydro	1797	A	G	876	5	9	344	majority_strict	1,760,001
BC3	Contig4459_2	G7IPJ9_MEDTR Phospho-2-dehydro	1797	A	G	883	50	64	875	majority_strict	1,760,001
BC3	Contig4459_2	G7IPJ9_MEDTR Phospho-2-dehydro	1797	A	G	179	0	1	173	homozygous_strict	1,760,001
139	Contig7559_2	G7IPL2_MEDTR 2-oxoglutarate/mal	515	A	G	1	465	630	0	homozygous_strict	1,830,001
247	Contig7559_2	G7IPL2_MEDTR 2-oxoglutarate/mal	515	A	G	2	70	55	0	majority_strict	1,830,001
247	Contig7559_2	G7IPL2_MEDTR 2-oxoglutarate/mal	515	A	G	2	393	318	0	homozygous_strict	1,830,001
248	Contig7559_2	G7IPL2_MEDTR 2-oxoglutarate/mal	515	A	G	0	95	55	0	homozygous_strict	1,830,001
248	Contig7559_2	G7IPL2_MEDTR 2-oxoglutarate/mal	515	A	G	0	240	318	0	homozygous_strict	1,830,001
848	Contig7559_2	G7IPL2_MEDTR 2-oxoglutarate/mal	515	A	G	1	395	384	0	homozygous_strict	1,830,001
139&848	Contig7559_2	G7IPL2_MEDTR 2-oxoglutarate/mal	515	A	G	0	31	37	1	homozygous_strict	1,830,001
BC3	Contig7559_2	G7IPL2_MEDTR 2-oxoglutarate/mal	515	A	G	0	78	84	0	homozygous_strict	1,830,001
BC3	Contig7559_2	G7IPL2_MEDTR 2-oxoglutarate/mal	515	A	G	0	299	521	0	homozygous_strict	1,830,001
139	Contig4170	G7IPPO_MEDTR Chaperone DnaJ-dc	1853	C	T	365	0	0	171	homozygous_strict	1,950,001
247	Contig4170	G7IPPO_MEDTR Chaperone DnaJ-dc	1853	C	T	230	7	3	114	majority_strict	1,950,001
247	Contig4170	G7IPPO_MEDTR Chaperone DnaJ-dc	1853	C	T	304	0	1	149	homozygous_strict	1,950,001
248	Contig4170	G7IPPO_MEDTR Chaperone DnaJ-dc	1853	C	T	189	0	3	114	majority_strict	1,950,001
248	Contig4170	G7IPPO_MEDTR Chaperone DnaJ-dc	1853	C	T	196	0	1	149	homozygous_strict	1,950,001
848	Contig4170	G7IPPO_MEDTR Chaperone DnaJ-dc	1853	C	T	478	0	0	336	homozygous_strict	1,950,001
139&848	Contig4170	G7IPPO_MEDTR Chaperone DnaJ-dc	1853	C	T	210	0	2	113	homozygous_strict	1,950,001
BC3	Contig4170	G7IPPO_MEDTR Chaperone DnaJ-dc	1853	C	T	251	2	3	162	majority_strict	1,950,001
BC3	Contig4170	G7IPPO_MEDTR Chaperone DnaJ-dc	1853	C	T	226	0	1	241	homozygous_strict	1,950,001

Abbildung 28: RNASeq&MACE - Chr.2-kartierbare SNP-Kandidaten Kategorie „homozygous_strict“ >> „majority_strict“

Eine komplette Liste aller 91 SNP-Kandidaten findet sich im Anhang I und II.

Aspekte der Sequenzier-Methodik und Auswirkungen auf die Ergebnislisten

Interessant im Zusammenhang mit der Auswertung ist auch die Frage, mit welcher der beiden Methoden (RNASeq und/oder MACE) die SNP-Kandidaten identifiziert werden konnten. Dabei

muss allerdings -wie schon erwähnt- berücksichtigt werden, dass die beiden Techniken nicht äquivalent bzgl. ihrer potentiellen Daten sind
 Der MACE-Ansatz wurde in der Projektplanung ursprünglich zusätzlich zu dem RNASeq-Ansatz gewählt, um mehr Einzelproben kostengünstiger auf Genexpression untersuchen zu können. Mit RNASeq alleine hätten aufgrund der höheren Kosten nicht so viele Proben bearbeitet werden können. Die dabei -ebenfalls- anfallenden SNP-Daten wurden jedoch mit ausgewertet, auch wenn sie nicht als Replikat der korrespondierenden RNASeq-Daten gewertet werden können. Das wird bei längeren Contigs deutlich, in denen mehrere SNP-Positionen gefunden wurden. MACE-basierte SNP-Daten (**massive analysis of cDNA-ends**) finden sich prinzipiell immer im 3'-Bereich eines Contigs während RNASeq-basierte SNP-Daten über den gesamten Bereich eines Contigs verteilt sind (Abb. 29). Z.B. Contig 4459_2 besitzt laut Referenztranskriptom eine Länge von 1981b. Es befinden sich in diesem Contig vier potentielle SNP-Positionen: Pos.54, Pos.1473, Pos.1729 und Pos.1797. Anhand der Spalte „Detection“ wird erkennbar, dass nur in den beiden dem 3'-Ende der putativen mRNA benachbarten Positionen aus den MACE-Daten Informationen vorliegen. Dies stimmt mit der theoretischen Erwartung überein, dass die Verteilung der SNP-Position entlang einer mRNA vom verwendeten Detektionsverfahren abhängt.

Genotyp	gene_id	description	position	SNP-Nucleotid		Basen - counts				snp_type	Region (+10000bp)	Detection
				Referenz	Variante	Low [4u5u6]		High [4u5u6]				
						Referenz	Variante	Referenz	Variante			
139	Contig4459_2	G7IPJ9_MEDTR Phospho-2-dehydr	54	A	C	130	0	0	96	homozygous_strict	1,760,001	RNASeq-2017
247	Contig4459_2	G7IPJ9_MEDTR Phospho-2-dehydr	54	A	C	92	0	0	80	homozygous_strict	1,760,001	RNASeq-2019
248	Contig4459_2	G7IPJ9_MEDTR Phospho-2-dehydr	54	A	C	67	0	0	80	homozygous_strict	1,760,001	RNASeq-2019
848	Contig4459_2	G7IPJ9_MEDTR Phospho-2-dehydr	54	A	C	143	0	0	149	homozygous_strict	1,760,001	RNASeq-2017
BC3	Contig4459_2	G7IPJ9_MEDTR Phospho-2-dehydr	54	A	C	115	0	0	152	homozygous_strict	1,760,001	RNASeq-2019
139	Contig4459_2	G7IPJ9_MEDTR Phospho-2-dehydr	1473	C	T	1607	0	0	964	homozygous_strict	1,760,001	RNASeq-2017
247	Contig4459_2	G7IPJ9_MEDTR Phospho-2-dehydr	1473	C	T	1079	2	2	494	homozygous_strict	1,760,001	RNASeq-2019
248	Contig4459_2	G7IPJ9_MEDTR Phospho-2-dehydr	1473	C	T	656	1	2	494	homozygous_strict	1,760,001	RNASeq-2019
848	Contig4459_2	G7IPJ9_MEDTR Phospho-2-dehydr	1473	C	T	1711	2	1	1034	homozygous_strict	1,760,001	RNASeq-2017
BC3	Contig4459_2	G7IPJ9_MEDTR Phospho-2-dehydr	1473	C	T	911	0	1	878	homozygous_strict	1,760,001	RNASeq-2019
139	Contig4459_2	G7IPJ9_MEDTR Phospho-2-dehydr	1729	C	T	570	0	0	446	homozygous_strict	1,760,001	RNASeq-2017
247	Contig4459_2	G7IPJ9_MEDTR Phospho-2-dehydr	1729	C	T	710	8	18	372	majority_strict	1,760,001	MACE-2019
247	Contig4459_2	G7IPJ9_MEDTR Phospho-2-dehydr	1729	C	T	291	0	0	166	homozygous_strict	1,760,001	RNASeq-2019
248	Contig4459_2	G7IPJ9_MEDTR Phospho-2-dehydr	1729	C	T	679	2	18	372	majority_strict	1,760,001	MACE-2019
248	Contig4459_2	G7IPJ9_MEDTR Phospho-2-dehydr	1729	C	T	176	0	0	166	homozygous_strict	1,760,001	RNASeq-2019
848	Contig4459_2	G7IPJ9_MEDTR Phospho-2-dehydr	1729	C	T	651	1	0	342	homozygous_strict	1,760,001	RNASeq-2017
139&848	Contig4459_2	G7IPJ9_MEDTR Phospho-2-dehydr	1729	C	T	558	3	7	264	majority_strict	1,760,001	MACE-2017
BC3	Contig4459_2	G7IPJ9_MEDTR Phospho-2-dehydr	1729	C	T	813	24	18	726	majority_strict	1,760,001	MACE-2019
BC3	Contig4459_2	G7IPJ9_MEDTR Phospho-2-dehydr	1729	C	T	229	0	0	253	homozygous_strict	1,760,001	RNASeq-2019
139	Contig4459_2	G7IPJ9_MEDTR Phospho-2-dehydr	1797	A	G	358	0	0	208	homozygous_strict	1,760,001	RNASeq-2017
247	Contig4459_2	G7IPJ9_MEDTR Phospho-2-dehydr	1797	A	G	756	28	18	331	majority_strict	1,760,001	MACE-2019
247	Contig4459_2	G7IPJ9_MEDTR Phospho-2-dehydr	1797	A	G	256	0	0	100	homozygous_strict	1,760,001	RNASeq-2019
248	Contig4459_2	G7IPJ9_MEDTR Phospho-2-dehydr	1797	A	G	602	15	18	331	majority_strict	1,760,001	MACE-2019
248	Contig4459_2	G7IPJ9_MEDTR Phospho-2-dehydr	1797	A	G	136	0	0	100	homozygous_strict	1,760,001	RNASeq-2019
848	Contig4459_2	G7IPJ9_MEDTR Phospho-2-dehydr	1797	A	G	465	0	0	221	homozygous_strict	1,760,001	RNASeq-2017
139&848	Contig4459_2	G7IPJ9_MEDTR Phospho-2-dehydr	1797	A	G	876	5	9	344	majority_strict	1,760,001	MACE-2017
BC3	Contig4459_2	G7IPJ9_MEDTR Phospho-2-dehydr	1797	A	G	883	50	64	875	majority_strict	1,760,001	MACE-2019
BC3	Contig4459_2	G7IPJ9_MEDTR Phospho-2-dehydr	1797	A	G	179	0	1	173	homozygous_strict	1,760,001	RNASeq-2019

Abbildung 29: RNASeq&MACE - Bsp. SNP-Positionen und Detektionsmethode (RNASeq oder MACE)

In der Retrospektive wurde aufgeschlüsselt, wie die Daten zu den 91 SNP-Kandidaten (vgl. Abb. 20) in den beiden verschiedenen methodischen Ansätzen verteilt waren. Betrachtet man die Gesamtmenge der einzelnen Datenpunkte für Chr.2-lokalisierbare und „NA“-SNP-Kandidaten (Abb.30), so kommen ca. 88% bzw. 81% der SNP-Datenpunkte aus den RNASeq-Ansätzen, ohne dabei zwischen „homozygot_strict“ und „majority_strict“ zu unterscheiden.

RNASeq & MACE: 2017 & 2019				
SNP-Daten	Chr.2-wide		NA	
Σ	397	100%	123	100%
RNASeq	350	88.2%	100	81.3%
MACE	47	11.8%	23	18.7%

Abbildung 30: RNASeq&MACE- Statistik aller SNP-Datenpunkte der finalen SNP-Kandidaten

Im gesamten Projektverlauf wurden fünf kontrastierende V/C-Genotypenpaare (Nah-isogene „Göx139“ und „NPZx848“; Hochisogene: „Six247“ und „Six248“ und zwei F3-Bulks aus der Rückkreuzung „BC3“) auf SNP-Kandidaten hin analysiert. Für 91 SNP-Kandidaten, die allen diesen High-VC/Low-VC-Paarungen gemeinsam sind, werden somit mindestens 455 Datenpunkte benötigt. Bezogen auf diese theoretische Mindestanzahl von SNP-Datenpunkten erreichen die gesamten RNASeq-Daten mit 450 eine Abdeckung von ca. 98.9%, die gesamten MACE-Daten mit 70 dagegen nur 15.4%. Zusammengenommen lieferten beide Verfahren 520 Datenpunkt und somit 65 Datenpunkte mehr als unbedingt notwendig, was einer Redundanz von 14.3% an Datenpunkten entspricht. Die Datensätze wurden -getrennt nach dem methodischen Vorgehen- auf die Anzahl an darin enthaltenen SNP-Kandidaten hin ausgezählt (Abb. 31). Im direkten Vergleich zeigt sich, dass von den finalen 91 SNP-Kandidaten die hinreichenden und notwendigen Informationen zu 90 Stück in den RNASeq-Daten gefunden werden können. Die MACE-Daten lieferten die kompletten Informationen für 12 SNP-Kandidaten. Betrachtet man diejenigen SNP-Kandidaten, für die redundante bzw. komplementäre Informationen, d.h. sowohl von RNASeq- wie auch von MACE-Daten, vorhanden sind, gilt dies für 22 SNP-Kandidaten. Dies bedeutet aber nicht, dass in diesen Fällen jeweils die kompletten Informationen aus beiden Detektionsmethoden vorliegen (s.u.).

	RNASeq		MACE		RNASeq&MACE	
	Chr. 2-wide	"NA"	Chr. 2-wide	"NA"	Chr. 2-wide	"NA"
"Hs"-SNPs	44	15	2	0	7	1
"Ms"-SNPs	2	0	2	0	0	0
mixed-SNPs	24	5	5	3	8	6
Σ (Hs&Ms):	70	20	9	3	15	7
Σ:	90		12		22	

Abbildung 31: RNASeq&MACE- Statistik aller SNP-Kandidaten nach Detektionsmethode

In einer anderen Betrachtungsweise wurden in den RNASeq- und MACE-Datensätzen diejenigen Datenpunkte erfasst, die jeweils exklusiv nur mit einer der beiden Methoden detektiert wurden und daher nicht in dem korrespondierenden Sequenzieransatz vorhanden sind (Abb. 32). Ca. 75% der SNP-Kandidaten werden alleine mit den RNASeq-spezifischen Daten abgebildet. Nur für einen einzigen SNP-Kandidaten (Abb. 33) sind die MACE-Daten hinreichend und zwingend erforderlich und es gibt keine korrespondierenden RNASeq-Daten. In den restlichen ca. 24% der Fälle, bei denen Datenpunkte aus beiden Detektionsmethoden existieren, waren die RNASeq-Daten ausreichend, um einen SNP-Kandidaten zu identifizieren (Abb. 32). Für acht (8*; Chr.2-wide) bzw. drei (3*; „NA“) SNP-Kandidaten war auch jeweils ein kompletter Datensatz in den MACE-Daten vorhanden.

	Keine MACE-Daten!		Keine RNASeq-Daten!		gemeinsame Daten	
	nur RNASeq-Daten	"NA"	nur MACE-Daten	"NA"	RNASeq&MACE	
	Chr. 2-wide	"NA"	Chr. 2-wide	"NA"	Chr. 2-wide	"NA"
"Hs"-SNPs	30	11	1	0	7	1
"Ms"-SNPs	2	0	0	0	0	0
mixed-SNPs	23	2	0	0	8	6
Σ (Hs&Ms):	55	13	1	0	15 (8*)	7 (3*)
Σ:	68 [74,7%]		1 [1,1%]		22 [24,2%]	
Σ:	91 [100%]					

Abbildung 32: RNASeq&MACE- Statistik aller SNP-Kandidaten spezifisch für Detektionsmethode

Abo-Vici - Abschlußbericht - RLP AgroScience GmbH

Züchtung und Agronomie neuartiger, Vicin-armer Ackerbohnen und Einsatz als einheimisches Eiweißfutter
Förderkennzeichen: 2815EPS066 23 April 2021

Genotyp	gene_id	description	position	SNP-Nucleotid		Basen - counts				snp_type	Region (+10000bp)	Detection
				Referenz	Variante	Low [4u5u6]		High [4u5u6]				
						Referenz	Variante	Referenz	Variante			
247	Contig10165_3	D2D582_MEDTR GT2 OS=Medicago	1737	A	G	0	6	16	0	homozygous_strict	1,350,001	MACE-2019
248	Contig10165_3	D2D582_MEDTR GT2 OS=Medicago	1737	A	G	0	6	16	0	homozygous_strict	1,350,001	MACE-2019
139&848	Contig10165_3	D2D582_MEDTR GT2 OS=Medicago	1737	A	G	0	9	11	0	homozygous_strict	1,350,001	MACE-2017
BC3	Contig10165_3	D2D582_MEDTR GT2 OS=Medicago	1737	A	G	0	8	24	0	homozygous_strict	1,350,001	MACE-2019

Abbildung 33: einziger SNP-Kandidaten, der exklusiv nur in MACE-Daten zu finden ist

Abschließend wurde für alle MACE- und RNASeq-Datensätze aufgeschlüsselt nach den beiden Kategorien „Chr.2-wide“ und „NA“ - analysiert, welche Datenpunkte minimal notwendig waren, um einen SNP-Kandidaten über das gesamte genetische Material hinweg zu definieren. Interessanterweise waren von den 91 SNP-Kandidaten nur für ein einzigen die MACE-Daten zwingend erforderlich, in allen anderen Fällen waren die RNASeq-Daten ausreichend, um den SNP-Kandidaten zu identifizieren (Abb. 34).

Minimal notwendige Daten für SNP			
SNP-Typ	Chr. 2-wide	Detection	"NA"
homozygous strict	44	RNASeq	15
homozygous strict	1	MACE	0
majority strict	2	RNASeq	0
majority strict	0	MACE	0
mixed-SNPs	24	RNASeq	5
mixed-SNPs	0	MACE	0
Σ Check:	71		20
		91	
"0" nicht notwendig / nicht gefunden			

Abbildung 34: RNASeq&MACE - Minimal notwendige Daten für alle 91 SNP-Kandidaten

Zwischenfazit

Unter dem Aspekt der SNP-Kandidaten-Detektion und der optimalen Ausbeute ist der RNA-Seq-Methode der Vorzug zu geben. Eine Einbeziehung der Kategorie „majority_strict“ führte zu zusätzlichen SNP-Kandidaten. Da die Daten nicht auf der Analyse von genomischer DNA sondern auf den mRNA-Transkripten beruhen, beeinflusst die Abundanz der einzelnen Transkripte die statistische Bewertung erheblich. Insbesondere wenn wenige mRNA-Kopien in den Zellen vorhanden sind, wirken sich auch Artefakte, die durch die Sequenzierung selbst bzw. in der Herstellung der cDNA-Templates entstehen können, ungünstig auf die Bewertung eines potentiellen SNPs aus. Daher ist ein manuelles Überprüfen der SNP-Häufigkeiten und der Verhältnisse zueinander in den zu vergleichenden Niedrig-VC versus Hoch-VC Genotypen angeraten und kann u.U. zu einer anderen Einschätzung führen als es die starre Arithmetik in den Auswertungen vorgibt.

Neubewertung des evaluierten KASP-Markersets aus 2017

Aus den SNP-Analysen der ersten Runde „2017“ wurde von der AG Link ein Satz an potentiellen KASP-Markern entwickelt, experimentell getestet und erfolgreich für die Feinkartierung des VC Locus sowie die Züchtung einer VC-armen Winterackerbohne eingesetzt. 58 von ursprünglich 65 Kandidaten (Tacke et al., 2021) wurden im Projektverlauf verwendet und evaluiert. Durch die danach erfolgte Auswertung der zweiten Transkriptomanalyse „2019“ reduzierte sich die Zahl dieser ursprünglichen SNP-Kandidaten von 65 auf 17 Stück, davon dreizehn kartierbare und vier nicht auf *M.truncatula* kartierbare SNP-Contigs.

Diese drastische Reduzierung fand eine Erklärung in dem Vergleich der SNP-Kandidatenlisten von 2017 mit der finalen Liste der 91 SNP-Kandidaten aus der Auswertung aller Datensätze. Beim Erstellen dieser Listen wurden ja nur die Datensätze berücksichtigt,

die in die zuvor erläuterten Kategorien „homozygot_strict“ und „majority_strict“ fielen. Daten mit schlechteren Bewertungen wurden dabei ausgeblendet, da diese höchstwahrscheinlich nicht mit einer homozygoten Situation in den analysierten diploiden Genotypen kompatibel sind. Um die Ursachen für den Wegfall der ursprünglich als SNP-Kandidaten identifizierten Contigs nachvollziehen zu können, wurden daher zuerst alle verfügbaren SNP-Daten der zugehörigen Contigs aus allen Datensätzen zusammengestellt, unabhängig von ihrer Kategorisierung des SNP-Detektionsalgorithmus. Dies umfasste primär einen Datensatz von 2818 einzelne Datenpunkte. Nach Aussortieren der nicht-relevanten SNP-Daten blieben davon 524 Datenpunkte übrig, die sich auf die 65 SNP-Kandidaten aus 2017 verteilten.

Dies resultierte aber nicht in jedem Fall in einen kompletten Datensatz, der alle 5 verwendeten Genotypen-Vergleiche umfasste (Abb. 35). Zur Vereinfachung -und in Anlehnung an die Haplotypen-Darstellung der experimentell validierten KASP-Marker der AG Link- wurden die Zahlenverhältnisse für jeden SNP interpretiert, so dass -hypothetisch- entweder homozygot nur ein LVC-Allel („A“) oder das korrespondierende HVC-Allel (B) oder beide Allele („H“; heterozygot) in dem untersuchten Genotyp vorliegen. In neun Fällen waren keine SNP-Daten („n.d.“ in Abb.35) in allen drei Vergleichen aus 2019 (hoch-isogene Linien: Sixdenier247 LVC vs. Sixdenier250 HVC, Sixdenier248 LVC vs. Sixdenier250 HVC und Backcross-3F3 LVC-Bulk vs. Backcross-3F3 HVC-Bulk) zu finden.

Abo-Vici - Abschlußbericht - RLP AgroScience GmbH

Züchtung und Agronomie neuartiger, Vicin-armer Ackerbohnen und Einsatz als einheimisches Eiweißfutter
 Förderkennzeichen: 2815EPS066 23 April 2021

					Allelstatus, abgeleitet aus den SNP-Daten von RNASeq und MACE									
					"AA" LowVC vs. "BB" HighVC									
					2017		2019			2017		2019		
VFS18002.A	M.truncatula				2017	2017	2019	2019	2019	2017	2017	2019	2019	
No.	SNP/gene_ID	Position	Region	Methode	139-L	848-L	Six247-L	Six-248-L	BC3-L	139-H	848-H	Six-250	BC3-H	
3	Contig2795	569	390001	MACE&RNASeq	A	A	A	A	n.d.	B	B	B	n.d.	
42	Contig5514	3719	590001	RNASeq only	A	A	A	A	A	B	B	B	H	
43	Contig18076	4998	760001	RNASeq only	A	A	A	A	A	B	B	B	H	
44	Contig7764_2	227	930001	RNASeq only	A	A	A	A	n.d.	B	B	B	n.d.	
45	Contig11152_2	1960	1100001	RNASeq only	A	A	A	A	A	B	B	B	H	
4	Contig2806_2	266	1140001	MACE&RNASeq	A	A	A	A	A	B	B	B	H	
5	Contig20794	964	1150001	MACE&RNASeq	A	A	A	A	n.d.	B	B	B	n.d.	
46	Contig2800	792	1300001	RNASeq only	A	A	A	A	n.d.	B	B	B	n.d.	
6	Contig6812	370	1320001	MACE&RNASeq	A	A	A	A	A	B	B	B	B	
7	Contig11262	408	1350001	MACE&RNASeq	A	A	A	A	A	B	B	B	B	
47	Contig5797	2406	1370001	RNASeq only	A	A	A	A	A	B	B	B	B	
48	Contig9470	1150	1470001	RNASeq only	A	A	A	A	A	B	B	B	B	
8	Contig20221_2	1444	1540001	MACE&RNASeq	A	A	A	A	A	B	B	B	B	
9	Contig3485	309	1550001	MACE&RNASeq	A	A	A	A	A	B	B	B	B	
10	Contig3485	327	1550001	MACE&RNASeq	A	A	A	A	n.d.	B	B	B	n.d.	
49	Contig17017	1461	1590001	RNASeq only	A	A	A	A	A	B	B	B	B	
50	Contig701_2	1170	1650001	RNASeq only	A	A	A	A	A	B	B	B	B	
11	Contig4274	407	1670001	MACE&RNASeq	A	A	A	A	n.d.	B	B	B	n.d.	
51	Contig4459_2	1473	1760001	RNASeq only	A	A	A	A	A	B	B	B	B	
12	Contig7559_2	515	1830001	MACE&RNASeq	A	A	A	A	A	B	B	B	B	
52	Contig14304	216	1850001	RNASeq only	A	A	A	A	A	B	B	B	B	
66	Contig19058	513	1869538 Vf_Mt2g009320_001		A	A	A	A	H	B	B	B	B	
53	Contig18365	1091	1880001	RNASeq only	A	A	A	A	n.d.	B	B	B	n.d.	
13	Contig4170	1853	1950001	MACE&RNASeq	A	A	A	A	A	B	B	B	B	
14	Contig10359_2	824	1970001	MACE&RNASeq	A	A	A	A	H	B	B	B	B	
54	Contig9103_2	3101	1970001	RNASeq only	A	A	A	A	A	B	B	B	B	
55	Contig18949	2498	2030001	RNASeq only	A	A	A	A	H	B	B	B	B	
15	Contig20444	1175	2090001	MACE&RNASeq	A	A	A	A	n.d.	B	B	B	n.d.	
16	Contig20444	1249	2090001	MACE&RNASeq	A	A	A	A	H	B	B	B	H	
56	Contig10690	2070	2190001	RNASeq only	A	A	A	A	H	B	B	B	H	
57	Contig9590	1311	2250001	RNASeq only	A	A	A	A	H	B	B	B	H	
17	TRINITY_DN40503_c0_g1_i1	296	2390001	MACE&RNASeq	A	A	A	A	H	B	B	B	H	
58	Contig5616	1106	2430001	RNASeq only	A	A	A	A	n.d.	B	B	B	n.d.	
18	Contig18779_3	607	2460001	MACE&RNASeq	A	A	A	A	n.d.	B	B	B	n.d.	
19	TRINITY_DN37125_c0_g1_i1	33	2500001	MACE&RNASeq	A	A	A	A	H	B	B	B	H	
67	Contig12520_2	507	2505174 Vf_Mt2g010740_001		A	A	A	A	H	B	B	B	B	
20	Contig3069_2	785	2600001	MACE&RNASeq	A	A	A	A	n.d.	B	B	B	n.d.	
21	Contig4233	866	2610001	MACE&RNASeq	A	A	A	A	H	B	B	B	H	
22	Contig1038	703	2750001	MACE&RNASeq	A	A	A	A	n.d.	B	B	B	n.d.	
23	Contig7104_2	1068	2790001	MACE&RNASeq	A	A	A	A	n.d.	B	B	B	n.d.	
24	Contig7104_2	1076	2790001	MACE&RNASeq	A	A	A	A	n.d.	B	B	B	n.d.	
25	Contig7104_2	1104	2790001	MACE&RNASeq	A	A	A	A	n.d.	B	B	B	n.d.	
59	Contig12988_2	998	2790001	RNASeq only	A	A	A	A	n.d.	B	B	B	n.d.	
60	Contig21357	2062	2850001	RNASeq only	A	A	A	A	H	B	B	B	H	
26	Contig22054_2	452	2910001	MACE&RNASeq	A	A	A	A	n.d.	B	B	B	n.d.	
61	Contig9274	1645	2970001	RNASeq only	A	A	A	A	n.d.	B	B	B	n.d.	
62	Contig4762_2	2508	3010001	RNASeq only	A	A	A	A	H	B	B	B	H	
27	Contig9847	3058	3020001	MACE&RNASeq	A	A	A	A	n.d.	B	B	B	n.d.	
28	Contig2809_2	452	3080001	MACE&RNASeq	A	A	A	A	n.d.	B	B	B	n.d.	
29	Contig2809_2	550	3080001	MACE&RNASeq	A	A	A	A	n.d.	B	B	B	n.d.	
30	Contig7537_2	210	3080001	MACE&RNASeq	A	A	n.d.	n.d.	n.d.	B	B	n.d.	n.d.	
63	Contig60_2	1779	3110001	RNASeq only	A	A	n.d.	n.d.	n.d.	B	B	n.d.	n.d.	
64	Contig20591	638	3180001	RNASeq only	A	A	n.d.	n.d.	n.d.	B	B	n.d.	n.d.	
31	Contig21312_2	1904	3470001	MACE&RNASeq	A	A	n.d.	n.d.	n.d.	B	B	n.d.	n.d.	
32	Contig21312_2	1952	3470001	MACE&RNASeq	A	A	n.d.	n.d.	n.d.	B	B	n.d.	n.d.	
33	Contig15453	419	3520001	MACE&RNASeq	A	A	n.d.	n.d.	n.d.	B	B	n.d.	n.d.	
65	Contig22277	303	3590001	RNASeq only	A	A	n.d.	n.d.	n.d.	B	B	n.d.	n.d.	
34	Contig19763_2	1717	3610001	MACE&RNASeq	A	A	n.d.	n.d.	n.d.	B	B	n.d.	n.d.	
35	TRINITY_DN42642_c0_g1_i2	112	3610001	MACE&RNASeq	A	A	n.d.	n.d.	n.d.	B	B	n.d.	n.d.	
36	Contig8443	1173	n.a.	MACE&RNASeq	A	A	A	A	H	B	B	B	H	
37	Contig8443	1261	n.a.	MACE&RNASeq	A	A	A	A	H	B	B	B	H	
38	TRINITY_DN44477_c0_g1_i2	444	n.a.	MACE&RNASeq	A	A	A	A	A	B	B	B	B	
39	TRINITY_DN44477_c0_g1_i2	593	n.a.	MACE&RNASeq	A	A	A	A	A	B	B	B	B	
40	TRINITY_DN44477_c0_g1_i2	668	n.a.	MACE&RNASeq	A	A	A	A	A	B	B	B	B	
41	TRINITY_DN49974_c1_g1_i2	126	n.a.	MACE&RNASeq	A	A	A	A	A	B	B	B	B	

Abbildung 35: KASP-Markerset 2017 - SNP-Daten Vergleich mit 2019 und hypothetischer Allelstatus

Abo-Vici - Abschlußbericht - RLP AgroScience GmbH

Züchtung und Agronomie neuartiger, Vicin-arter Ackerbohnen und Einsatz als einheimisches Eiweißfutter
 Förderkennzeichen: 2815EPS066 23 April 2021

In insgesamt 30 Fällen lieferten die beiden F3-Backcross-Bulks keine Daten. D.h. in diesen Fällen ist mit sehr hoher Wahrscheinlichkeit der bi-allelische Polymorphismus im untersuchten Material für die beiden SNP-Allele nicht vorhanden und der zugehörige SNP-Kandidat konnte als für eine Niedrig-VC-Selektion nicht geeignet eliminiert werden. Die Möglichkeit, daß die zugehörigen mRNA-Transkripte in den betroffenen Genotypen eventuell nicht vorhanden sind, wurde für alle Fälle von „n.d.“ (in Tab. 35) ausgeschlossen. Denn zu jedem einzelnen Contig waren Expressionswerte sowohl in den Sixdenier-Linien wie auch in den Backcross-Bulks vorhanden (Daten nicht gezeigt). Eine weitergehende Ursachenklärung war nicht möglich, da die zugehörigen Reads nicht in geeigneter Form zur Verfügung standen. Es müsste ggf. geprüft werden, ob in diesen Fällen in der Region um die SNP-Position kein Reads existieren oder ob an der SNP-Position keines der beiden erwarteten Nukleotide vorhanden ist, sondern eines der anderen verbleibenden Nukleotide.

Unter den restlichen 35 SNP-Kandidaten fanden sich in 18 Fällen heterozygote Allel-Konstellationen und zwar ausschließlich im Backcross Material. In 10 Fällen galt dies sowohl für den Niedrig-VC-Bulk („LVC“) wie auch für den Hoch-VC-Bulk („HVC“). In jeweils weiteren 4 Fällen war die heterozygote Allel-Konstellation nur in einem der beiden Bulks zu finden. Für den korrespondierenden Bulk lag eine homozygote Allel-Konstellation vor, die auch mit den restlichen Genotypen übereinstimmte. Da laut der Abo-Vici-Arbeitshypothese die LVC- und HVC-Phänotypen jeweils auf einer durchgängig homozygoten Allel-Verteilung für das VC-Gen beruhen müssen, wurden die SNP-Kandidaten mit einem heterozygoten Zustand unabhängig vom Vorkommen in einem der beiden Backcross-Bulks ebenfalls herausgefiltert. Übrig blieben damit dreizehn KASP-Assay validierte SNPs, die auf Chr.2 von *M.truncatula* kartierbar waren und vier SNPs, deren Contig-Sequenzen nicht kartiert werden konnten. Acht dieser KASP-Marker (Abb. 36, sortiert nach der Tabelle 6, Tacke et al. 2021) konnten erfolgreich in der Marker-gestützten Selektion der Backcross-Strategie zur Einkreuzung der Niedrig-VC Eigenschaft von der AG Link in Winterackerbohnen eingesetzt werden (s.a. AG-Link: Abo-Vici_Göttingen-Schlussbericht zu Aufgabe 6).

				Allelstatus, abgeleitet aus den SNP-Daten von RNASeq und MACE									
				"AA" LowVC vs. "BB" HighVC									
				2017				2019					
				139-L		848-L		Six247-L		Six-248-L		BC3-L	
				139-H		848-H		Six-250		BC3-H			
VFS18002.A		M.truncatula											
No.	SNP/gene_ID	Position	Region	Method	139-L	848-L	Six247-L	Six-248-L	BC3-L	139-H	848-H	Six-250	BC3-H
39	TRINITY_DN44477_c0_g1_i2	593	n.a.	MACE&RNASeq	A	A	A	A	A	B	B	B	B
49	Contig17017	1461	1590001	RNASeq only	A	A	A	A	A	B	B	B	B
51	Contig4459_2	1473	1760001	RNASeq only	A	A	A	A	A	B	B	B	B
41	TRINITY_DN49974_c1_g1_i2	126	n.a.	MACE&RNASeq	A	A	A	A	A	B	B	B	B
12	Contig7559_2	515	1830001	MACE&RNASeq	A	A	A	A	A	B	B	B	B
52	Contig14304	216	1850001	RNASeq only	A	A	A	A	A	B	B	B	B
384	Contig14304	726	1850001	RNASeq only	A	A	A	A	A	B	B	B	B
54	Contig9103_2	3101	1970001	RNASeq only	A	A	A	A	A	B	B	B	B
13	Contig4170	1853	1950001	MACE&RNASeq	A	A	A	A	A	B	B	B	B

Abbildung 36: KASP-Marker 2017 Kernset, hypothet. Allelstatus, abgeleitet aus den Transkriptomdaten

Betrachtet man die 28 kartierbaren Contigs aus der finalen SNP-Kandidaten-Liste (vgl. Abb. 20), in denen SNP-Positionen gefunden wurden (Abb. 37), sind es insgesamt drei Contigs mit nur einem einzigen SNP, die unverändert in der Schnittmenge aus den ursprünglich 65 SNP-Kandidaten und in der finalen Liste zu finden sind. In fünfzehn Fällen kommen neue Contigs mit insgesamt 30 SNP-Positionen dazu. In 13 Contigs erweitert sich die Anzahl an SNP-Positionen um weitere 27 Positionen.

Abo-Vici - Abschlußbericht - RLP AgroScience GmbH

Züchtung und Agronomie neuartiger, Vicin-artermer Ackerbohnen und Einsatz als einheimisches Eiweißfutter
Förderkennzeichen: 2815EPS066 23 April 2021

SNPs-kartierbar auf Chr.2 von <i>M. truncatula</i>				SNPs- nicht kartierbar auf <i>M.truncatula</i>			
# Contig	Position	Chr.2-Region	SNPs	# Contig	Position	Chr.2-Region	SNPs
1 Contig6812	287	1,320,001	2	36 TRINITY_DN38912_c1_g1_i3	259	1,650,001	1
2 Contig6812	370	1,320,001		37 Contig4274	259	1,670,001	
3 Contig5730	770	1,340,001		38 Contig4274	866	1,670,001	2
4 Contig5730	920	1,340,001		39 Contig12307	1111	1,710,001	
5 Contig5730	1151	1,340,001		40 Contig12307	1791	1,710,001	3
6 Contig5730	1415	1,340,001	6	41 Contig12307	1973	1,710,001	
7 Contig5730	1532	1,340,001		42 Contig18427	1435	1,740,001	
8 Contig5730	1760	1,340,001		43 Contig18427	1634	1,740,001	3
9 Contig10165_3	1737	1,350,001	1	44 Contig18427	2300	1,740,001	
10 Contig11262	408	1,350,001		45 Contig4459_2	54	1,760,001	
11 Contig11262	1354	1,350,001		46 Contig4459_2	1473	1,760,001	4
12 Contig11262	1412	1,350,001	4	47 Contig4459_2	1729	1,760,001	
13 Contig11262	1603	1,350,001		48 Contig4459_2	1797	1,760,001	
14 Contig6063	1140	1,360,001		49 Contig19011	621	1,780,001	1
15 Contig6063	1267	1,360,001	2	50 Contig7609	115	1,800,001	
16 Contig5797	257	1,370,001		51 Contig7609	325	1,800,001	
17 Contig5797	362	1,370,001		52 Contig7609	392	1,800,001	4
18 Contig5797	906	1,370,001	4	53 Contig7609	1131	1,800,001	
19 Contig5797	2406	1,370,001		54 Contig7559_2	515	1,830,001	1
20 Contig9470	1150	1,470,001		55 Contig14304	216	1,850,001	
21 Contig9470	1402	1,470,001	2	56 Contig14304	687	1,850,001	
22 Contig20221_2	505	1,540,001		57 Contig14304	726†	1,850,001	
23 Contig20221_2	986	1,540,001		58 Contig14304	1055	1,850,001	7
24 Contig20221_2	1444	1,540,001	4	59 Contig14304	1574	1,850,001	
25 Contig20221_2	1475	1,540,001		60 Contig14304	1598	1,850,001	
26 Contig27050	453	1,540,001		61 Contig14304	1602	1,850,001	
27 Contig27050	536	1,540,001	2	62 Contig18365	1609	1,880,001	1
28 Contig3485	104	1,550,001		63 Contig3200_2	946	1,930,001	
29 Contig3485	252	1,550,001		64 Contig3200_2	1129	1,930,001	2
30 Contig3485	309	1,550,001		65 Contig4170	377	1,950,001	
31 Contig3485	876	1,550,001	4	66 Contig4170	1685	1,950,001	3
32 Contig17017	1461	1,590,001	1	67 Contig4170	1853	1,950,001	
33 Contig1696	2740	1,640,001	1	68 Contig9103_2	463	1,970,001	
34 Contig17637	728	1,650,001	1	69 Contig9103_2	581	1,970,001	3
35 Contig701_2	1170	1,650,001	1	70 Contig9103_2	3101	1,970,001	
		Σ SNPs	35	71 Contig22331	1455	1,980,001	1
					Σ SNPs		36
Σ all SNPs = 71 in 28 Contigs							

Abbildung 37: KASP-Marker 2017 (in gelb) versus finales SNP-Kandidatenset aus 2017 & 2019

†: dieser SNP auf Contig14304 ist identisch mit den „SNP384“ (M. Song 2017 und Khazaei et al. 2017, siehe Tacke et al. 2021)

Bei den nicht-kartierbaren SNPs sind 2 Contigs identisch mit denen aus dem ersten KASP-Markerset von 2017. Für einen von beiden kamen 8 weitere SNP-Kandidaten (d.h. SNP-Positionen innerhalb der Contig-Sequenz) hinzu. Zwei neue Contigs mit insgesamt 8 zusätzlichen SNP-Kandidaten kamen in der finalen Auswertung hinzu. In einem Fall ist der -einzige-SNP innerhalb der zugehörigen Contig-Sequenz konserviert. Dieser 319 b lange, nicht kartierbare Contig TRINITY_DN49974_c1_g1_i2, aus dem der AboVici-KASP-Marker „A41“ abgeleitet wurde (Abb. 36), überlappt auf cDNA-Ebene mit den Contig14304 (RibA/VC1) am 3'-Ende um 117 Basen (= Pos.1 - 117). Davon sind allerdings 7 Positionen unterschiedlich. D.h. es resultiert eine ca. 93% Identität im überlappenden Bereich. Der SNP „A41“ liegt 75 b entfernt (Pos. 194) vom überlappenden Sequenzbereich, also weiter „3'-seitig“ auf dem Contig TRINITY_DN49974_c1_g1_i2. Es könnte daher sein, dass dieser Contig von einer mRNA-Variante von RibA/VC1 stammt. Die 319 b Sequenzinformation von Contig TRINITY_DN49974_c1_g1_i2 ist allerdings nicht ausreichend, bei einem BLAST gegen das *M.truncatula*-Genom einen signifikanten Treffer zu erzeugen, wie eine individuelle Überprüfung bestätigte. Dies war offensichtlich der Grund, der zu der Zuordnung „n.a.“ führte. Auch eine Annotation war über die 319 b lange Sequenz nicht möglich. Erst eine manuelle Feinanalyse der cDNA-Sequenzen untereinander offenbarte eine potentielle Kontinuität beider Sequenzen. Auch die Tatsache, dass in den Feinkartierungen des VC-Locus der AG-Link der KASP-Marker „A41“ in die Nähe von SNP384(=> Pos.726 Contig14304) und A52(=> Pos.216 in Contig14304) positioniert wurde, unterstützt die Annahme, dass der 319 b Contig von einer Variante der RibA/VC1-mRNA stammen könnte. D.h. es besteht eine räumliche Nähe zwischen beiden Markern auf dem *V.faba* Chromosom 1, die sich in der genetischen Koppelung niederschlägt.

Der Terminus „neu“ bezieht sich in allen Fällen auf die ursprünglich 2017 ausgewählten 65 SNP-Kandidaten aus der ersten, größeren Liste an SNP-Kandidaten (insgesamt 463 Chr.2 kartierbare und 57 „nicht auf *M.truncatula* kartierbare putative SNPs). Diese basierte auf der damals ausschließlich berücksichtigten Kategorie „homozygot_strikt“. Insofern waren diese „neuen“ SNPs schon teilweise in dieser ersten Liste vorhanden, kamen aber nicht in die engere Auswahl für das KASP-Markerliste 2017, die Grundlage für die Feinkartierungen des VC-

Genlocus waren (s. Tacke et al 2021, und im Abschlussbericht des Koordinators Prof. Dr. W. Link, AG Göttingen (Aufgabe 6) .

Soweit es anhand der Sequenzinformationen der SNP-haltigen Contigs aus dem Referenztranskriptom möglich war, wurden die Auswirkungen der SNPs auf die putativen oder über Annotation nachgewiesenen Leserahmen, i.e. davon abgeleitete Proteine bioinformatisch untersucht. In keinem der Fälle konnte eine wesentliche Änderung diagnostiziert werden, die einen Ausfall des Translationsproduktes (i.e. Stop-Codon) oder Änderungen der Faltung (z.B. Glycin als Helixbrecher, hydrophiler versus hydrophober Aminosäurerest etc.) vorhersagen ließen.

Fazit

Der Bereich, den die auf Chr. 2 von *M.truncatula* kartierbaren SNP-Kandidaten abdeckten (Abb. 38) verringerte sich von der ersten zur zweiten Transkriptomanalyse signifikant. Der zum *V.faba* Chromosom 1 syntäne Bereich des *M.truncatula* Chromosom 2 zwischen den beiden am weitesten auseinanderliegenden SNP-Kandidaten um den VC-Genlocus herum reduzierte sich von ursprünglich ca. 3,22 Mill. Basen auf 0,66 Mill. Basen, d.h. ca. um den Faktor 4,9. Ursächlich dürfte hierfür die von der AG Link verwendete Strategie bei der Auswahl der unterschiedlichen Genotypen und der damit einhergehenden Rekombinationsereignisse im Bereich des VC-Gens sein. Alle Ansätze zielten ja auf die Selektion von Niedrig-VC-Genotypen, die aber auf unterschiedliche Züchtungsgänge und Genotypen zurückgingen. Das Filtern auf die allen Hoch/Niedrig-VC-Vergleichen gemeinsamen SNPs führte nachweislich zu einer Fokussierung auf den Abschnitt des Chromosom 1 von *V.faba*, der mit großer Wahrscheinlichkeit das VC-Gen beinhaltet und für welches die neuen SNP-Kandidaten eine effizientere Marker-gestützte Selektion ermöglichen sollen.

SNPs-kartierbar auf Chr.2 von <i>M. truncatula</i>					
KASP-Marker 2017			SNP-Kandidaten 2017&2019		
Conitg_ID	SNP-Position	Chr.2-Region	Conitg_ID	SNP-Position	Chr.2-Region
Contig2795	569	390,001	Contig6812	287	1,320,001
TRINITY_DN42642_c0_g1_i2	112	3,610,001	Contig22331	1455	1,980,001

Abbildung 38: äußerste KASP-Marker 2017 versus finale SNP-Kandidaten, gefiltert aus 2017 & 2019

Bzgl. der Qualität bzw. Nutzbarkeit der final ermittelten zusätzlichen SNP-Kandidaten (58 kartierbare und 16 „n.a.“) für die Züchtungspraxis kann zu diesem Zeitpunkt keine abschließende Aussage gemacht werden, da diese Daten im restlichen Projektverlauf nicht mehr für weitere KASP-Assays verwendet werden konnten. Es ist aber zumindest bei den SNP-Kandidaten, die auf *M.truncatula* kartiert werden konnten, erkennbar, das die zusätzlichen SNP-Positionen alle (bis auf #71, Abb. 37) innerhalb der schon von den validierten KASP-Markern des ersten Markersets liegen. Es bleibt zu klären, ob diese zusätzlichen Markerkandidaten in stabile KASP-Assays umgewandelt werden können, und wie gut ihre Koppelung mit dem Niedrig-VC-Gehalt ist, um für eine effizientere Marker-gestützte Selektion genutzt werden zu können. Da für die Identifizierung der SNP-Kandidaten weitere, neue Genotypen (insbesondere durch den Backcross auch Winterackerbohne) verwendet wurden, ist zu hoffen, dass einige der neuen SNPs sich auch auf einem erweiterten Spektrum an unterschiedlichen Genotypen als robust und für eine zielgerichtete Niedrig-VC-Selektion als geeignet erweisen werden.

III. Differentielle Expression und Kandidatengene

Auswertung der Transkriptomdaten

Analog zum Vorgehen bei der SNP-Kandidatensuche wurde bei der Suche nach differentiell exprimierten Kandidatengenen die bzgl. des VC-Gehaltes kontrastierenden Genotypen gegeneinander verglichen. Es standen die gleichen RNASeq- und MACE-Roh-Datensätze zur Verfügung, die schon für die SNP-Kandidatensuche verwendet wurden (s. Abb. 1a).

Abo-Vici - Abschlußbericht - RLP AgroScience GmbH

Züchtung und Agronomie neuartiger, Vicin-armer Ackerbohnen und Einsatz als einheimisches Eiweißfutter
 Förderkennzeichen: 2815EPS066 23 April 2021

Für die Suche nach Kandidatengenen der VC-Biosynthese wurden zwei Fälle unterschieden. Zum einen wurde nach Genen (bzw. deren Contigs) gefiltert, die in den Niedrig-VC Genotypen wesentlich schwächer exprimiert wurden (d.h. Reduktion oder Ausfall einer Genfunktion) als in den VC-Genotypen. Zum anderen wurden Kandidaten aus den Daten gefiltert, die in den Niedrig-VC-Genotypen stärker exprimiert werden (z.B. Überexpression eines negativen Regulators oder eines VC-degradierenden Enzyms).

Der wirksamste Filter war damit die Suche nach differentiell exprimierten Kandidatengenen, die über das gesamte Pflanzenmaterial aus beiden NGS-Runden von 2017 und 2019 eine einheitliche Tendenz zeigten. Diejenigen Kandidaten, die über das gesamte Spektrum der unterschiedlichen Genotypen konstant eine differentielle Expression zwischen V/C-armen und Hoch-V/C-Pflanzen aufweisen, sollten diejenigen sein, die für die VC-Biosynthese eine Rolle spielen. Dabei galt es auch zu berücksichtigen, dass umweltbedingte Schwankungen bei dem zu verschiedenen Jahreszeiten geernteten Material bei der Expressionsstärke eine Rolle spielen könnten, die sich auch teilweise in geringeren Unterschieden bemerkbar machen können. Für eine differentielle Expression wurde daher als Kriterium der recht geringe Faktor zwei angesetzt, sowohl für Kandidaten („Down“), die in VC-armen Genotypen mindestens zweifach geringer als im Wildtyp exprimiert werden, als auch umgekehrt solche Kandidaten („Up“), die in V/C-armen Genotypen mindestens zweifach stärker als im Wildtyp exprimiert werden. Dazu wurden zuerst die RNASeq- und MACE-Daten getrennt analysiert und danach die Schnittmenge der Contigs bestimmt, die durch beide Sequenzieretechniken gleichermaßen als zweifach differentielle identifiziert wurden. Daraus wurde im nächsten Schritt ein nicht-redundantes Set aller Kandidaten-Contigs für die Transkriptomdaten von jeweils 2017 und 2019 erstellt (Abb.39 a und b).

Differentielle Expression - Contigs : High[Stadium 4 & Stadium 5] / Low[Stadium 4 & Stadium 5] > 2											
RNASeq			MACE			RNASeq und MACE					
x-Göttingen	x-NPZ	139 & 848	x-Göttingen	x-NPZ	139 & 848	x-Göttingen	x-NPZ	139 & 848	Σ Contigs "Down" 2017		
139-2/139-1	848-4/848-3		139-2/139-1	848-4/848-3		139-2/139-1	848-4/848-3		MACE only	non-redundant	
6459	6750	1720	2345	2264	316	923	1222	158	158	1878	

Differentielle Expression - Contigs: High[Stadium 4 & Stadium 5] / Low[Stadium 4 & Stadium 5] < 0,5											
RNASeq			MACE			RNASeq und MACE					
x-Göttingen	x-NPZ	139 & 848	x-Göttingen	x-NPZ	139 & 848	x-Göttingen	x-NPZ	139 & 848	Σ Contigs "Up" 2017		
139-2/139-1	848-4/848-3		139-2/139-1	848-4/848-3		139-2/139-1	848-4/848-3		MACE only	non-redundant	
6177	5125	1356	1843	2279	279	656	367	74	205	1561	

Abbildung 39a: zweifach-differentiell exprimierte Contigs für RNASeq- und MACE-Daten aus 2017

Differentielle Expression - Contigs : High[Stadium 4 & Stadium 5] / Low[Stadium 4 & Stadium 5] > 2											
RNASeq					MACE						
x-Göttingen					x-Göttingen					Σ Contigs "Down" 2017	
BC3-H/BC3-L	S50-H/S47-L	S50-H/S48-L	Six-H/Six-L	BC3&Six	BC3-H/BC3-L	S50-H/S47-L	S50-H/S48-L	Six-H/Six-L	BC3&Six	MACE only	non-redundant
10778	13397	8702	4997	660	950	2729	1232	765	18	11	671

Differentielle Expression - Contigs: High[Stadium 4 & Stadium 5] / Low[Stadium 4 & Stadium 5] < 0,5											
RNASeq					MACE						
x-Göttingen					x-Göttingen					Σ Contigs "Up" 2017	
BC3-H/BC3-L	S50-H/S47-L	S50-H/S48-L	Six-H/Six-L	BC3&Six	BC3-H/BC3-L	S50-H/S47-L	S50-H/S48-L	Six-H/Six-L	BC3&Six	MACE only	non-redundant
6959	8805	9335	3839	353	1076	2066	1537	534	15	11	364

Abbildung 39b: 2fach-differentiell exprimierte Contigs für RNASeq- und MACE-Daten aus 2019

Die anfangs noch sehr hohen Zahlen für Kandidaten-Contigs wurden durch die Synthese der Daten aus 2017 („139 & 848“) und 2019 („BC3 & Six“) letztlich erheblich reduziert und die gewünschte Filterwirkung erzielt (Abb. 40). Auch die 454-Pyrosequencing-Daten von Ray et al. (2015) wurden in dieser Analyse und der Statistik zwar berücksichtigt, erwiesen sich bei genauerer Analyse leider als nicht verlässlich und bargen die Gefahr, dass wichtige Kandidaten fälschlicherweise ausgeblendet werden. Ursächlich dafür dürfte die deutlich geringere Sequenzieretiefe sein, die von Ray et al. im 454-Verfahren verwendet wurde.

Abo-Vici - Abschlußbericht - RLP AgroScience GmbH

Züchtung und Agronomie neuartiger, Vicin-arter Ackerbohnen und Einsatz als einheimisches Eiweißfutter
 Förderkennzeichen: 2815EPS066 23 April 2021

High/Low > 2 "Down"			High/Low < 0.5 "Up"		
4			0		
41		6039	18		6925
1878	671		1561	364	
139 & 848	BC3&Six	Fatima/A01155	139 & 848	BC3&Six	Fatima/A01155
Abo-Vici		Ray et al. 2015	Abo-Vici		Ray et al. 2015

Abbildung 40: Finale Statistik 2fach-differentiell exprimierte Contigs für RNASeq- und MACE-Daten

Letztlich findet sich kein „Up“-Contig, der unter Einbeziehung der 454-Daten in der Vicin/Convicin-armen Sorte stärker exprimiert ist. Für die eigenen Abo-Vici-Daten dagegen bleiben 18 Kandidaten übrig.

4 Contigs sind mindestens zweifach in VC-haltigen Genotypen stärker exprimiert, wenn die 454-Daten mitberücksichtigt werden. Allerdings lässt sich drei dieser Contigs anhand ihrer Sequenz keine sinnvolle Funktion zuordnen, sie zählen zu der Kategorie „uncharacterized RNA“. Der vierte mit den 454-Daten übereinstimmende Contig, ist interessanterweise TRINITY_DN44878_c1_g2_i2. Seine Differentialität basiert allerdings auf äußerst geringen Werten (ein „454“-read für die Niedrig-V/C-Sorte gegenüber drei „454“-reads in der Hoch-V/C-Sorte!). Es handelt sich um einen Vertreter aus der RibA-Contig14304-Gruppe, die schon in der ersten Projektphase anhand der Abo-Vici-eigenen Transkriptom-Analyse als möglicher Kandidat der VC-Biosynthese auffiel. Der eigentlich RibA-Hauptvertreter, Contig14304, der ja zudem Kandidaten-Contig4518 von Ray et al. passt (s. 2ter Zwischenbericht), fiel bei dieser statistischen Auswertung der 454-Daten allerdings durch das Raster, da die für ihn ermittelten Werte (194 454-reads/Low VC zu 353 454-reads/High VC) nur einen Quotienten von 1.8 ergaben. Fokussiert man ausschließlich auf die Abo-Vici-Daten bleiben insgesamt 41 „Down“-Contigs übrig, die als mindestens zweifach geringer in VC-armen Sorten exprimiert bewertet wurden.

Das Kriterium der zweifach unterschiedlichen Expression erwies sich in der individuellen Detailanalyse der einzelnen Kandidaten-Contigs als nicht ausreichend. Unter den (41+18=) 59 herausgefilterten Contigs waren viele, deren read-Zahlen im unteren einstelligen Bereich lagen. Z.B. ergibt sich ein differentielles Verhältnis von größer „2“ oder kleiner „0.5“ schon bei einem read-Quotienten von 3 reads :1 read bzw. 1 read :3 reads. Derartig sehr niedrig exprimierte Gene sind über die qPCR mit Standardverfahren nicht mehr nachweisbar. Zusätzliche Ausschlusskriterien waren eine fehlende Annotation („uncharacterized RNA“) oder eine Annotation, die eindeutig keinen Bezug zur Thematik aufwies (z.B. „Retrovirus-related Pol polyprotein from transposon“ oder „Metacaspase-1-like protein“). Nach manueller Durchsicht der Kandidatenliste blieben jeweils 7 Contig-Gruppen übrig (Abb 41), die bisher zum größeren Teil leider nur mit unbekannter Funktion „uncharacterized“ annotiert wurden. Ein Abgleich mit den bis dato verfügbaren Daten der ersten Genomsequenzierung durch das NORFAB-Konsortiums, bei dem auch der Abo-Vici-Projektpartner NPZ-Lemke beteiligt ist, brachte keine neuen Erkenntnisse bzgl. der „uncharacterized“-Kandidaten-Contigs, da zu dem Zeitpunkt nur nicht-annotierte gDNA-Sequenzen zur Verfügung standen.

[+][L] >2x		7		
RNASeq & MACE non-redundant 2017 & 2019 "Up"				
Contig	Description	M.truncatula Chr.	Region	Status / verwandte Contigs
TRINITY_DN44878_c1_g2_i2	A0A2K3MSS0_TRIPR Riboflavin biosynthesis protein ribBA OS=Trifolium pratense	NC_016408.2	1,840,000	RibA
Contig14304	G7IPL5_MEDTR 3,4-dihydroxy-2-butanone 4-phosphate synthase OS=Medicago truncatula	NC_016408.2	1,840,000	RibA
TRINITY_DN49974_c1_g1_i2	NA	NA	NA	RibA
TRINITY_DN74545_c0_g1_i1	uncharacterized / Abgleich mit annotiertem V.faba-Genom erforderlich	NA	NA	unique
Contig11156(1447 b)	G7INU9_MEDTR Peroxidase OS=Medicago truncatula	NC_016408.2	1,580,000	unique
Contig509(570 b)	uncharacterized / Abgleich mit annotiertem V.faba-Genom erforderlich	NA	NA	related to Contig 16886
Contig1024_2(1049 b)	uncharacterized / Abgleich mit annotiertem V.faba-Genom erforderlich	NA	NA	unique
Contig9336	G7ID29_MEDTR Pectin lyase superfamily protein OS=Medicago truncatula	NC_016407.2	45,070,000	related to Contig7856_2
Contig18210_2	LOW QUALITY PROTEIN [Medicago truncatula]? / Abgleich mit annotiertem V.faba-Genom erforderlich	NC_016409.2?	46,240,000	unique
[+][L] <0.5x		7		
RNASeq & MACE non-redundant 2017 & 2019 "Down"				
Contig	Description	M.truncatula Chr.	Region	Status / verwandte Contigs
Contig11162(1826 b)	G7ILA4_MEDTR Glycosyltransferase OS=Medicago truncatula	NC_016408.2	1350000	unique
Contig4411(1859 b)	A0A067XTV5_CICAR Glycosyltransferase OS=Cicer arietinum	NC_016408.2	2060000	unique
Contig6131(1121 b)	uncharacterized / Abgleich mit annotiertem V.faba-Genom erforderlich	NA	NA	related to TRINITY_DN40106_c0_g1_i2
TRINITY_DN40106_c0_g1_i2	uncharacterized / Abgleich mit annotiertem V.faba-Genom erforderlich	NA	NA	related to Contig6131
Contig20626_2(726 b)	uncharacterized / Abgleich mit annotiertem V.faba-Genom erforderlich	NA	NA	unique
Contig15538_2(587 b)	uncharacterized / Abgleich mit annotiertem V.faba-Genom erforderlich	NA	NA	unique
Contig13514(1072 b)	uncharacterized / Abgleich mit annotiertem V.faba-Genom erforderlich	NA	NA	related to TRINITY_DN10433_c0_g2_i1
Contig15766_2	G7JKH5_MEDTR DUF506family protein OS=Medicago truncatula	NC_016410.2	5961000	unique

Abbildung 41: Finale Expressions-Kandidatenliste für Abo-Vici

Eine Überprüfung der NGS-basierten Expressionswerte bekannter Gene des pflanzlichen Pyrimidin-Biosynthesewegs ergab keinerlei Auffälligkeiten (CPSase, ATCase, DHOase, DHODH, OPRTase und ODCase). Somit konnte der hypothetisch als Kandidat postulierte Orotsäure-Biosyntheseweg als Ursache der VC-Mutation ausgeschlossen werden.

Überprüfung der Expression mit quantitativer PCR

Für einige Kandidatengene bzw. deren Referenz-Contigs ergab die Annotation eine möglichen enzymatischen Funktion im Kontext der VC-Biosynthese (wie Peroxidase, Glycosyltransferase oder Pectin-Lyase-Superfamilie). Deren aus den Transkriptomdaten stammenden Expressionswerte wurden experimentell mithilfe quantitativer PCR am gleichen Pflanzenmaterial respektive RNA-Material verifiziert (= Bestätigung der Transkriptomdaten) und an weiterem Pflanzenmaterial validiert (= Überprüfung der Expression an nicht-sequenzierten Genotypen) analysiert. Vier ausgewählte Referenzgene (Daten nicht gezeigt) wurden ebenfalls validiert und mithilfe des Web-basierten Metatools RefFinder (<https://www.heartcure.com.au/reffinder/>) analysiert. Daraus resultierte ein stabiles Referenzgenset über alle verwendeten Genotypen hinweg. Das Set dient zur Normalisierung der Cq-Werte für die differentiell exprimierten Kandidatengene über die verschiedenen Proben hinweg, um so ggf. auftretende technische Schwankungen in den cDNA-Templates zu korrigieren.

Um ein Primerpaar zu identifizieren, welches erfolgreich zum Nachweis für den zu-gehörigen Kandidaten-Contig eingesetzt werden kann, waren i.d.R. Tests mit zwei bis drei verschiedenen Primerkombinationen nötig. Daher wurden diese zuerst an repräsentativen Mixen für die verwendeten Genotypen getestet und zu Artefakten neigende Kombinationen verworfen. Insbesondere in den Proben, in denen das Zielgen sehr schwach exprimiert war, tendierten die Primer offensichtlich zur Artefaktbildung, was in den SYBRGreen basierten qPCR-Assays letztlich eine Quantifizierung des zu messenden Zielgenes unmöglich machte. Daher wurde in vielen Fällen auf ein Sonden-basiertes qPCR-Verfahren (sog. Taqman-Assay) umgestiegen, bei dem zusätzlich zu Vorwärts- und Rückwärtsprimer eine dazwischen liegende, mit einem Fluorophor ausgestattete Sonde für die spezifische Signalerzeugung verantwortlich ist. In diesem Fall wird nur das spezifische Signal der Sonde detektiert, Primer-Artefakte tragen nicht zum Signal bei.

Zusätzlich zur möglichen Primer-Artefaktbildung, die sich in Tests vorab diagnostizieren und ausschließen ließ, wurde mit einer Pearson-Korrelation überprüft, ob die über qPCR ermittelten Expressionsprofile der Kandidatengene mit denen der Transkriptomdaten übereinstimmten. Nur genspezifische Assays mit einem Pearson-Korrelationskoeffizienten deutlich über 0.9 wurden für die detaillierten qPCR-Analysen berücksichtigt.

Hauptkriterium der qPCR-Analytik war ein in der Tendenz konstantes Expressionsverhältnis über alle im Projekt verwendeten Niedrig-VC-Genotypen im Vergleich zu den korrespondie-

renden Hoch-VC-Genotypen in den (maternalen) Samenschalen, dem wahrscheinlichsten Ort der VC-Biosynthese. In den parallel untersuchten Geweben der zugehörigen Embryos und in Blättern war eine identische Situation eher nicht anzunehmen. Ausschlaggebend waren nicht die Absolutwerte sondern die Relation zwischen Hoch-VC und Niedrig-VC Genotypen, die im gleichen Zeitraum angezogen und geerntet wurden, da die Expression der Kandidatengene auch umweltbedingten Schwankungen unterworfen war.

Letztlich wurden in dem Testa-Gewebe zwei Gene identifiziert (grün unterlegt in Spalte Contig-ID), deren mRNA-Spiegel in Hoch-VC-Genotypen durchweg höher waren als in den Niedrig-VC-Genotypen (Abb. 42; „**H** > **L**“)- In beiden Fällen glichen die Expressionswerte in den intermediären Genotypen eher dem Wildtyp-Niveau (**H**=**L**>**L**). Contig18210_2 zeigte allerdings bei der manuellen Überprüfung eine sehr schlechte Annotation im Referenztranskriptom. Trotz intensiver zusätzlicher bioinformatischer Analyse war keine eindeutige Charakterisierung dieser Sequenz und eine Identifizierung über die öffentlichen Datenbanken möglich. Eine Kartierung der Contig18210_2-Sequenz auf das Medicago-Genom blieb ebenso ohne signifikanten Treffer. Contig14304 (und seine beiden partiell überlappenden verwandten Contigs TRINITY_DN49974_c1_g1_i2 und TRINITY_DN44878_c1_g2_i2) zeigt ebenfalls eine durchgängig höhere Expression im Wildtyp. Außerdem kartieren zwei der Contig-Sequenzen in dem syntänen Bereich, in dem der VC-Genlocus vermutet wird. Auf den Contig14304 (3,4-dihydroxy-2-butanone 4-phosphate synthase / Riboflavin biosynthesis protein ribBA, chloroplastic) wird weiter unten noch detaillierter eingegangen.

Zwei Gene, die beide zu den Glykosyltransferasen zählen, wiesen in Testa-Gewebe durchweg einen niedrigeren mRNA-Spiegel in Hoch-VC-Genotypen als in den Niedrig-VC-Genotypen (Abb. 42; „**L** > **H**“). Beide kartieren auf Chr.2 von *M.truncatula*, Contig4411 liegt in dem engeren Intervall, welches den VC-Genlocus definiert.

		qPCR - Testa												
				Ernte 2017				Ernte 2019						
				RNASeq&MACE		qPCR-only		RNASeq&MACE		qPCR-only				
				Gbx139	NFZ848	Gbx201	NFZ843	Sik-247	Sik-248	BC3	Sik-249	Sik-251	UKko139	UKko843
Expression	Contig-ID	Description	M.truncatula	Region										
	Contig3487_2	uncharacterized RNA	No Hit	n.a.	✓	✓	✓	✓	✓	no	H=>L	H=>L	no	no
	TRINITY_DN25206_c0_g1_i1	uncharacterized RNA	No Hit	n.a.	✓	✓	✓	✓	✓	no	H=>L	H=>L	no	no
	Contig4180_2	A2Q4Q6_MEDTR Allantoinase	Chr.2	3.444.512-3.442.360	✓	✓	✓	no	no	no	H=<L	H=<L	✓	✓
	TRINITY_DN33781_c0_g1_i2	uncharacterized RNA	No Hit	n.a.	✓	✓	✓	✓	✓	no	H=>L	H=>L	✓	✓
	Contig18210_2	G7J561_MEDTR Peptide upstream ORF protein	No Hit	n.a.	✓	✓	✓	✓	✓	no	H=>L	H=>L	✓	✓
H > L	Contig10165_3	D2D582_MEDTR GT2	Chr.2	1.356.892-1.358.375	✓	✓	✓	no	no	✓	H=<L	H=<L	no	no
	Contig14304	G7IPL5_MEDTR 3,4-dihydroxy-2-butanone 4-phosphate synthase	Chr.2	1.852.121-1.849.137	✓	✓	✓	✓	✓	✓	H=>L	H=>L	✓	✓
	TRINITY_DN49974_c1_g1_i2	uncharacterized RNA	No Hit	n.a.	✓	✓	✓	✓	✓	✓	H=>L	H=>L	✓	✓
	TRINITY_DN44878_c1_g2_i2	A0A082SHW5_GLYSO Riboflavin biosynthesis protein ribBA, chloroplastic	Chr.2	1.849.916-1.849.125	✓	✓	✓	✓	✓	✓	H=>L	H=>L	✓	✓
	Contig11156	G7INU9_MEDTR Peroxidase	Chr.2	1.584.570-1.584.037 1.592.243-1.593.755	✓	✓	✓	✓	✓	✓	H=>L	H=>L	no	no
	Contig9336	Pectin lyase superfamily protein	Chr.1,6,7,8	n.a.	✓	✓	✓	✓	✓	✓	H=>L	H=>L	✓	?
	Contig20626_2	uncharacterized RNA	No Hit	n.a.	✓	✓	✓	✓	✓	no	L>=H	L>=H	?	✓
	Contig23204	uncharacterized RNA	No Hit	n.a.	✓	✓	✓	no	no	no	L=H	L=H	✓	✓
L > H	TRINITY_DN26401_c0_g1_i2	uncharacterized RNA	No Hit	n.a.	✓	✓	✓	✓	✓	no	L>=H	L>=H	no	no
	Contig4411	A0A067XTV5_CICAR Glycosyltransferase	Chr.2	2.065.969-2.066.611	✓	✓	✓	✓	✓	✓	L>=H	L>=H	✓	✓
	Contig11262	G7ILA4_MEDTR Glycosyltransferase	Chr.2	1.351.463-1.349.454	✓	✓	✓	✓	✓	✓	L>=H	L>=H	✓	✓
H = L	Contig18594	AOA152YHC0_CICAR bifunctional riboflavin biosynthesis protein RIBA 1	Chr.4	37.271.855-37.270.593	✓	✓	✓	✓	✓	✓	L=H	L=H	✓	✓
	Contig6807		Chr.4	37.267.131-37.271.512	✓	✓	✓	✓	✓	✓	L=H	L=H	✓	✓

Abbildung 42: qPCR-Validierung der Expressions-Kandidaten für VC-Armut in Testa-Gewebe

Zwei weitere Contigs (Contig9336 und Contig20626_2) zeigten eine fast konsistente Expression, jedoch waren in beiden Fällen jeweils für eine der für VC-Gehalt spaltenden F2-Populationen (Ukko x „139“ bzw. Ukko x „843“) die Daten widersprüchlich, so dass diese beiden Contigs nicht in die engere Auswahl kamen.

Die Expressionsdaten für die Embryos (Abb. 43) und für die Blattproben (Abb. 44) lassen kein durchgängig einheitliches Expressionsmuster für die vier zuvor genannten Testa-Kandidaten erkennen. Es fällt immer mindestens eine Genotypen-Probe heraus, die nicht zu dem erwarteten Muster passt („no“).

Die zwei Kandidaten Contig18210_2 und Contig14304 zeigen allerdings auch in dem Embryogewebe in 8 von 9 Fällen die gleiche Tendenz wie im Testa-Gewebe. Einzig der „unsichere“ Contig20626_2 hatte in den Embryos eine durchgehend gleichbleibende Tendenz auf (Abb.43).

				qPCR - Embryo											
				Ernte 2017					Ernte 2019						
				qPCR-only					RNASeq&MACE						
Expression	Contig-ID	Description	M.truncatula	Region	G6x139	NPZ843x	G6x201	NPZ843	Slk-247	Slk-248	BC3	Slk-249	Slk-251	Ukoxx139	Ukoxx843
H > L	Contig3487_2	uncharacterized RNA	No Hit	n.a.	✓	✓	✓	✓	✓	✓	no	H=>L	H=>L	no	no
	TRINITY_DN25206_c0_g1_i1	uncharacterized RNA	No Hit	n.a.	✓	✓	✓	✓	✓	✓	no	H=>L	H=>L	no	no
	Contig4180_2	A2Q4Q6_MEDTR Allantoinase	Chr.2	3.444.512-3.442.360	✓	✓	✓	✓	✓	✓	✓	H=>L	H=>L	no	✓
	TRINITY_DN33781_c0_g1_i2	uncharacterized RNA	No Hit	n.a.	✓	✓	✓	✓	✓	✓	✓	H=>L	H=>L	no	✓
	Contig18210_2	G7J561_MEDTR Peptide upstream ORF protein	No Hit	n.a.	✓	✓	✓	✓	✓	✓	✓	H=>L	H=>L	no	✓
	Contig10165_3	D2D582_MEDTR GT2	Chr.2	1.356.892-1.358.375	✓	✓	no	?	no	no	no	n.d.	n.d.	no	no
	Contig14304	G7IPL5_MEDTR 3,4-dihydroxy-2-butanone 4-phosphate synthase	Chr.2	1.852.121-1.849.137	✓	✓	no	✓	✓	✓	✓	H>L	H>L	✓	✓
	TRINITY_DN49974_c1_g1_i2	uncharacterized RNA	No Hit	n.a.	✓	✓	no	✓	✓	✓	✓	H>L	H>L	✓	✓
	TRINITY_DN44878_c1_g2_i2	AOA0B2SHW5_GLYSO Riboflavin biosynthesis protein ribBA, chloroplatic	Chr.2	1.849.916-1.849.125	no	no	no	no	no	no	no	L=H	L=H	no	no
	Contig11156	G7INU9_MEDTR Peroxidase	Chr.2	1.584.570-1.584.037 1.592.243-1.593.755	no	✓	✓	✓	✓	✓	no	H=>L	H=>L	no	no
Contig9336	Pectin lyase superfamily protein	Chr.1,6,7,8	n.a.	no	no	no	no	no	no	no	H=H	L=H	no	no	
L > H	Contig20626_2	uncharacterized RNA	No Hit	n.a.	✓	✓	✓	✓	✓	✓	no	L>H	L>H	✓	✓
	Contig23204	uncharacterized RNA	No Hit	n.a.	✓	✓	✓	✓	no	no	no	L<H	L<H	✓	✓
	TRINITY_DN26401_c0_g1_i2	uncharacterized RNA	No Hit	n.a.	no	no	no	no	no	no	no	L=H	L<H	no	no
	Contig4411	AOA067XTV5_CICAR Glycosyltransferase	Chr.2	2.065.969-2.066.611	no	no	no	no	no	no	no	L<H	L<H	no	no
	Contig11262	G7ILA4_MEDTR Glycosyltransferase	Chr.2	1.351.463-1.349.454	no	no	no	no	no	no	no	L=H	L=H	✓	✓
	Contig18594	AOA152YHC0_CICAR bifunctional riboflavin biosynthesis protein RIBA 1	Chr.4	37.271.855-37.270.593	?	✓	✓	✓	✓	✓	✓	L=H	L=H	✓	✓
Contig6807	Chr.4		37.267.131-37.271.512	?	✓	✓	✓	✓	✓	✓	L=H	L=H	✓	✓	

Abbildung 43: qPCR-Validierung der Expressions-Kandidaten für VC-Armut in Embryo-Gewebe

In den meisten Fällen war die Expressionsstärke im Testa-Gewebe deutlich höher als in den Embryos, was für die Kandidatengene eine präferentielle Expression in der Testa unterstreicht. In Blattproben der verschiedenen Genotypen war keine einheitliche Expression zu finden (Abb.44). In zwei Fällen konnte gar keine Expression nachgewiesen werden (Contig23204 und TRINITY_DN26401_c0_g1_i2).

				qPCR - Blatt					
				Ernte 2020					
				qPCR-only					
Expression	Contig-ID	Description	M.truncatula	Region	G6 139x	NPZ 843x	BC3	Mei / Hv	Mei / Sask
H > L	Contig3487_2	uncharacterized RNA	No Hit	n.a.	✓	✓	?	no	✓
	TRINITY_DN25206_c0_g1_i1	uncharacterized RNA	No Hit	n.a.	✓	✓	no	no	no
	Contig4180_2	A2Q4Q6_MEDTR Allantoinase	Chr.2	3.444.512-3.442.360	✓	✓	no	✓	✓
	TRINITY_DN33781_c0_g1_i2	uncharacterized RNA	No Hit	n.a.	✓	✓	no	✓	✓
	Contig18210_2	G7J561_MEDTR Peptide upstream ORF protein	No Hit	n.a.	✓	✓	no	✓	✓
	Contig10165_3	D2D582_MEDTR GT2	Chr.2	1.356.892-1.358.375	no	no	no	no	no
	Contig14304	G7IPL5_MEDTR 3,4-dihydroxy-2-butanone 4-phosphate synthase	Chr.2	1.852.121-1.849.137	no	?	no	✓	no
	TRINITY_DN49974_c1_g1_i2	uncharacterized RNA	No Hit	n.a.	✓	?	no	✓	no
	TRINITY_DN44878_c1_g2_i2	AOA0B2SHW5_GLYSO Riboflavin biosynthesis protein ribBA, chloroplatic	Chr.2	1.849.916-1.849.125	no	no	no	no	no
	Contig11156	G7INU9_MEDTR Peroxidase	Chr.2	1.584.570-1.584.037 1.592.243-1.593.755	✓	no	no	no	no
Contig9336	Pectin lyase superfamily protein	Chr.1,6,7,8	n.a.	n.d.	n.d.	n.d.	n.d.	n.d.	
L > H	Contig20626_2	uncharacterized RNA	No Hit	n.a.	✓	✓	no	✓	✓
	Contig23204	uncharacterized RNA	No Hit	n.a.	n.a.	n.a.	n.a.	n.a.	
	TRINITY_DN26401_c0_g1_i2	uncharacterized RNA	No Hit	n.a.	n.a.	n.a.	n.a.	n.a.	
	Contig4411	AOA067XTV5_CICAR Glycosyltransferase	Chr.2	2.065.969-2.066.611	?	no	no	no	no
	Contig11262	G7ILA4_MEDTR Glycosyltransferase	Chr.2	1.351.463-1.349.454	✓	✓	no	✓	✓
	Contig18594	AOA152YHC0_CICAR bifunctional riboflavin biosynthesis protein RIBA 1	Chr.4	37.271.855-37.270.593	✓	✓	✓	?	?
Contig6807	Chr.4		37.267.131-37.271.512	✓	✓	✓	?	?	

Abbildung 44: qPCR-Validierung der Expressions-Kandidaten für VC-Armut in Blattgewebe

Letztlich blieben aufgrund ihres Expressionsmusters die vier (in der Spalte Contig-ID grün unterlegt) Kandidaten-Contigs aus den Expressionsanalysen übrig.

Die Schnittmenge aus SNP- und Expressionskandidaten

Abschließend wurde die Frage danach gestellt, welche Contig-Sequenzen über alle im Projekt verwendeten Genotypen sowohl SNP-Kandidaten lieferten als auch durch ihr differentielle Expressionsverhalten aufgefallen waren. Durch die Schnittmengenbildung über alles SNP-Kandidaten und Expressionskandidaten blieben nur noch 3 Contigs übrig, die diese Kriterien erfüllten:

- Contig11262 tr|G7ILA4_MEDTR Glycosyltransferase OS=*Medicago truncatula* (das zugehörige Gen ist durchgängig in Niedrig-VC-Genotypen (LVC) stärker exprimiert als im Wildtyp (HVC) (Vgl. Abb. 42 und Anhang I). Eine Glykosyltransferase-Aktivität ist im Rahmen der VC-Synthese als Übertragung der Aglykone auf einen Glucosemoleküle durchaus vor-

stellbar. Eine erhöhte Aktivität könnte in den Niedrig-VC-Genotypen aber auch zu einem -im Vergleich zum Wildtyp- erhöhten Abbau der beiden Pyrimidin-Glykoside und somit zur phänotypischen VC-Armut beitragen. Ein direkter Zusammenhang ist aber z. Zt. nicht bewiesen sondern rein spekulativ.

- Contig14304 ("G7IPL5_MEDTR 3,4-dihydroxy-2-butanone 4-phosphate synthase OS=*Medicago truncatula*") und
- TRINITY_DN49974_c1_g1_i2 ("uncharacterized RNA")

Das zugehörige Gen ist durchgängig in Niedrig-VC-Genotypen schwächer exprimiert als im Wildtyp mit Hoch-VC-Gehalt.

Zu diesen zwei Contigs gehörte noch eine dritte Contig-Sequenz, TRINITY_DN44878_c1_g2_i2. Die Sequenzdaten aller drei Contigs legten nahe, dass alle drei dieser separat assemblierten Contigs eng verwandt sind oder sogar von einem Gen stammen. Außerdem kartierten Contig 14304; TRINITY_DN44878_c1_g2_i2 und Contig14304 auf dem *Medicago* Chr. 2 in einer Region, die in der Nähe des VC-Lokus liegen muß. Alle drei Contigs sind in den AboVici-RNASeq- und MACE-Daten differentiell exprimiert. Außerdem passte die 370 bp-lange Sequenz eines der 6 Kandidatengene, Contig 4518 aus der Arbeit von Ray et al (2015) perfekt zu diesen drei Contigs (Abb.45, s.a. 2ter Zwischenbericht 2019 AboVici RLP AgroScience).

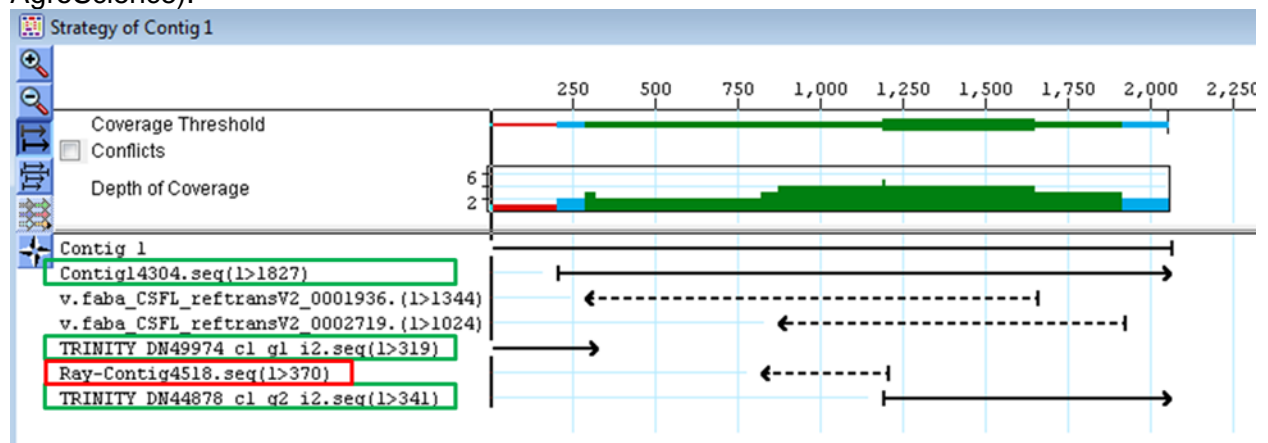


Abbildung 46: Schematischer Sequenzvergleich zu Ray-Kandidat 4518

Die ursprüngliche Annotation des Ray-Contigs 4518 "reticuline oxidase like protein" war aufgrund der geringen Sequenzinformation nicht zutreffend. Die Annotation des längsten Contig14304 mit einem offenen Leseraster von 502 Aminosäuren dagegen lautet "Bifunctional riboflavin biosynthesis protein RibA1". Hierbei handelt es sich um ein Protein mit zwei funktionellen Domänen (GTP Cyclohydrolase II / 3,4-Dihydroxy-2-Butanon-4-phosphat-Synthase), welches den ersten Schritt der Riboflavin-Biosynthese (Vitamin B2) katalysiert. Die GTP Cyclohydrolase II ist die enzymatische Funktion, die in der Vitamin B2-Synthese aus GTP ein Pyrimidin-Intermediat erzeugt. Dabei wird GTP zu 2,5-Diamino-6-Hydroxy-4-(5-Phosphoribosylamino)-Pyrimidin, Formiat und Pyrophosphat umgewandelt (Abb.46).

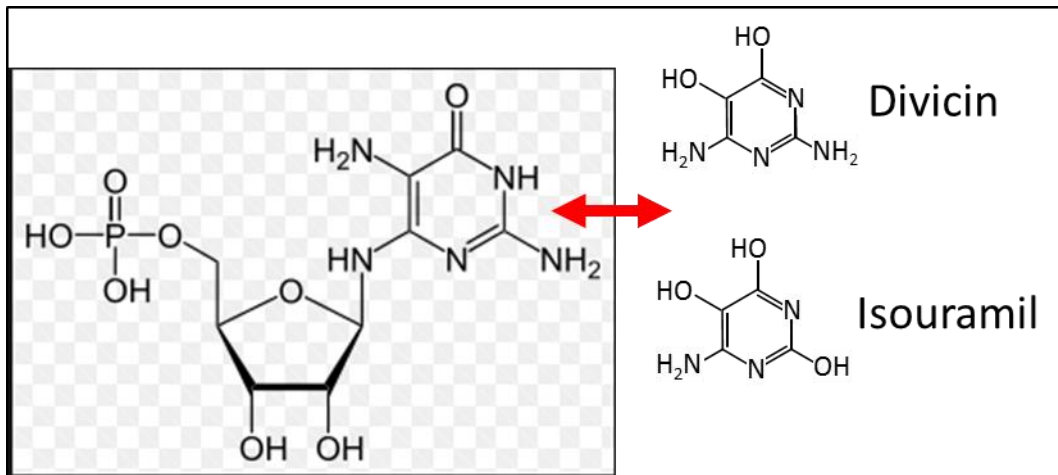


Abbildung 46: Strukturformel zu 2,5-Diamino-6-Hydroxy-4-(5-Phosphoribosylamino)-Pyrimidin

Auffällig war die strukturelle Ähnlichkeit zwischen dem Pyrimidin-Rest und den beiden Aglykonen Divicin und Isouramil. Eine biochemische Umwandlung zwischen den Molekülen war zum Zeitpunkt dieser Entdeckung (2019) allerdings noch nicht dokumentiert und es war hochspekulativ, ob das zu Contig14304 gehörende *Vicia faba* RibA1-Protein in der Lage sein könnte, eine Vorstufe des Vicins bzw. Convicins zu synthetisieren. Hypothetisch könnte eine geringere Expression der mRNA für RibA1/Contig14304/RayContig4518 in VC-armen Genotypen zu weniger Protein und damit zu reduzierten Mengen möglicher Vorstufen für VC führen.

Damit war RIBA1 ein sehr naheliegender Kandidatengenen für die VC-Biosynthese. Allerdings ist RIBA1 essentiell für die Vitamin-B2-Synthese. Sollte RIBA1 in Verbindung zur VC-Biosynthese stehen, müsste sich eine Deregulation des RIBA1-Gens in VC-armen Pflanzen auf die Samenschalen beschränken, denn die Pflanze ist als Ganzes auf Riboflavin als Vorstufe für die essentiellen enzymatischen Cofaktoren FMN und FAD angewiesen. In vegetativen Geweben wie insbesondere im Blatt, dürften keine Unterschiede in der Expression für dieses Gen zwischen VC-contrastierenden Genotypen nachzuweisen sein.

Aus den Transkriptomdaten ergaben sich jedoch zwei weitere, verwandte Contigs, Contig18594 und Contig6807, die ebenfalls die RIBA1-Annotation besaßen. Beide Contigs bilden einen überlappenden Cluster und kartieren auf Chr.4 des Medicago-Genoms in der Region um die Position 37270000. Beide Contigs sind in den NGS-Daten nicht differentiell über alle sequenzierten Genotypen. Außerdem konnten die qPCR-Daten dies für alle weiteren Genotypen bestätigen. Das Expressionsniveau war im Vergleich zu dem Cluster um den Contig14304 geringer aber gleichmäßig. Contig14304, wird im Vergleich zu diesen beiden Contigs 15 bis 20-fach (in Niedrig-VC-Genotypen) und bis zu ca. 100-fach (in Hoch-VC-Genotypen) stärker exprimiert. Somit wäre über dieses zweite RibA1-Gen die Grundversorgung der Vitamin B2-Synthese für die Pflanze auch in den Niedrig-VC-Genotypen sichergestellt.

Anfang 2020 erschien auf der Internetplattform BioRxiv ein bis dato nicht „peer-reviewed“ Manuskript des NORFAB-Konsortium (Björnsdotter et al. 2020), welches das Kandidatengen RibA/Contig14304/RayContig4518 als Ursache der VC-Armut in Ackerbohne beschreibt. Die Autoren präsentieren u.a. einen funktionellen Assay mit einem aus der *V.faba* Sorte Hedin klonierten RibA-Protein, mit dem sie die von unserer Arbeitsgruppe hypothetisch formulierte biochemische Aktivität des nun als „VC1“ bezeichneten Gens zweifelsfrei nachweisen. Die Ursache der VC-Armut vermuteten sie in einer Zwei-Basen-Frame-Shift-Insertion („AT“) in der GTP-Cyclohydrolase II-Domäne des bi-funktionellen Enzyms RIBA1, die exklusiv in der Genvariante des Niedrig-VC Genotyps Mélodie nachgewiesen wurde.

Diese Daten wurden durch Metabolitenstudien untermauert, die beweisen, dass die Pyrimidin-Glucoside Vicin und Convicin über den Purinstoffwechsel mit GTP als Ausgangssubstrat und nicht -wie bisher vermutet- über den Pyrimidin/Orotsäure-stoffwechsel synthetisiert werden.

VC1 ist das erste Gen eines in dem vorab veröffentlichten Manuskript postulierten Biosyntheseweges in dem allerdings vier weitere enzymatische Umwandlungen erfolgen müssen, um zu den Produkten Vicin und Convicin zu gelangen. Zu den korrespondierenden Genen wur-

den in der Publikation keine experimentellen Daten präsentiert, die Autoren formulieren aber einen theoretischen Biosyntheseweg unter Zuhilfenahme enzymatischer Funktionen (Desaminase, Hydrolase, Glycosyltransferase). Zu letzterem -Glycosyltransferase- lieferte das Abo-Vici-Projekt mindestens einen Kandidaten.

InDel-Analysen mit dem GATK-Algorithmus (<https://gatk.broadinstitute.org/hc/en-us>) der Abo-Vici-RNASeq- und MACE-Daten ergab leider kein eindeutiges Bild. In dem ersten Datensatz von 2017 wurde die „AT“-Insertion nicht gefunden, in der Analyse mit dem Material aus 2019 konnte sie dagegen diagnostiziert werden; allerdings waren auch die Daten nicht eindeutig. Ursächlich dafür dürfte sein, dass genau um diese Region (ca. 20 Basen vor und nach der AT-Insertion) die Sequenz zwischen Contig14304 und den beiden nicht-differentiell exprimierten Contigs für die zweite RibA1-Variante identisch sind und daher eine genaue Zuordnung der kurzen NGS-reads ohne eine gut definierte gDNA-Matrize (i.e. ein korrekt assembliertes und annotiertes *V.faba* Genom) nicht möglich ist. Es kommt offenbar aufgrund der lokal begrenzten Sequenzübereinstimmung beider Gene (und damit beider mRNAs) zu einer Überlagerung bzw. Durchmischung passender RNASeq-reads für die beiden RibA1-Gene von Chr.2 und Chr.4. Um solche Effekte in Zukunft zu vermeiden bleibt hier nur der direkte Weg über eine Klonierung der zugehörigen cDNAs und anschließenden Sequenzierung des kompletten Transkriptes.

Während die Insertionsmutation im VC1-Gen zu dem Niedrig-VC-Phänotyp führt, ist vermutlich das putative zweite RibA-Gen für die essentielle Versorgung der Pflanze mit Vitamin B zuständig. Unsere Transkriptomdaten zeigen eine wesentlich niedrigere, dafür aber Genotyp-unabhängige, konstante Expression für die zugehörigen Contigs. Dies könnte auch plausibel erklären, warum die vc1-Mutation zu keiner Null-Mutante sondern zu einer Niedrig-V/C-Mutante führt. Die basale Enzymaktivität des zweiten RIBA-Proteins könnte neben der Synthese von Riboflavin im geringeren Umfang in Samenschalen auch Vorstufen für die VC-Biosynthese liefern. Die dafür benötigten enzymatischen Schritte (Desaminase, Hydrolase, Glycosyltransferase) müssen daher in den Niedrig-VC-Genotypen existieren, um die -in geringem Umfang entstehenden- Vorstufen in die Endprodukte Vicin und Convicin umwandeln zu können. Die Aktivität(en) bzw. mRNA-Meng(en) der betroffenen Enzyme könnten in einer Niedrig-V/C-Mutante zumindest im Testa-Gewebe ebenfalls im Vergleich zum Wildtyp abgesenkt sein, sofern sie nicht auch anderweitige, essentielle metabolische Funktionen übernehmen. Unsere Expressionsdaten haben dafür allerdings keine Kandidaten mit dem passenden Muster und einer geeigneten Annotation geliefert. Eventuell müsste hier auch eine höhere Sequenzierentiefe (verbunden mit höheren Kosten) zur Anwendung kommen, um die Genexpression im niedrig-abundanten Bereich besser abbilden zu können.

Die vorab publizierten Arbeiten von Björnsdotter et al (2020) bestätigen letztlich unseren Haupt-Kandidaten Contig14304.

Voraussichtlicher Nutzen und Verwertbarkeit der Ergebnisse

Die vorgestellten Arbeiten dienen in erster Linie der Identifizierung von neuen SNP-Kandidaten, die eine effizientere Marker gestützte Selektion der Niedrig-Vicin/Convicin Mutation in der Züchtung von VC-armen Ackerbohne ermöglichen bzw. beschleunigen sollen. Die Ergebnisse aus Neustadt flossen direkt in die eigentliche Züchtungsarbeit der AG Prof. W. Link in Göttingen ein, die anhand der zur Verfügung gestellten SNP-Kandidatenliste einen Satz neuer KASP-Marker entwickelte, welche für eine Marker-gestützte Selektion auf VC-arme Ackerbohnen im Abo-Vici Projekt erfolgreich benutzt wurden und in drei unterschiedlichen genetischen Hintergründen gut funktionieren. Sofern sich diese molekularen Marker auch in weiteren für die Ackerbohnen-Züchtung relevanten genetischen Hintergründen als robust erweisen, stehen neue, effiziente Selektionswerkzeuge für die Züchtung neuer, VC-armer Winter- und Sommerackerbohnen zur Verfügung.

Außerdem konnte die für die VC-Armut ursächliche Mutation auf ein Kandidaten-Gen eingegrenzt werden. Hierfür wurden im Projekt Hinweise erarbeitet. Eine externe Arbeitsgruppe des NORFAB-Konsortiums publizierte zu diesem Kandidatengen im Projektverlauf überzeugende experimentelle Beweise, die unseren Verdacht bestätigten. Das Wissen über die Mutation in dem

Gen VC1/Riba1 könnte genutzt werden, um daraus einen besonders zuverlässigen, direkten Selektionsmarker zu entwickeln, der zu 100% mit der VC-Armut gekoppelt sein müsste.

Gegenüberstellung der ursprünglich geplanten zu den tatsächlich erreichten Zielen; Hinweise auf weiterführende Fragestellungen

Die genomweiten Transkriptom-Analysen in Neustadt führten zu der Etablierung eines Testaspezifischen Referenztranskriptoms.

Aus den NGS-Datensätzen wurden SNP-Kandidaten identifiziert, die teilweise erfolgreich für eine Marker-gestützte Selektion der VC-Armut in Kreuzungen von Sommer- und Winterackerbohne eingesetzt wurden. Die Generierung der SNP-Kandidaten erfolgte entsprechend der Verfügbarkeit des Pflanzenmaterials in zwei Teilen 2017 und 2019. Aus dem Teilergebnis 2017 konnten ausgewählte SNP-Marker identifiziert, von der AG in Göttingen validiert und für Züchtungsvorhaben erfolgreich eingesetzt werden. Die SNP-Kandidaten aus dem zweiten Teil 2019 konnten in der Praxis wegen des fortgeschrittenen Projektablaufs für die (Fein-)Kartierungen und Züchtungsgänge nicht mehr berücksichtigt werden. Die Informationen über diese SNP-Kandidaten stehen aber für weitere Züchtungsarbeiten und Projekte zur Verfügung.

Die Auswertung der Expressionsdaten und deren anschließende Validierung in einer größeren Satz an Genotypen führte zu der Identifizierung eines Kandidatengens, VC1/RibA1, welches aufgrund einer Mutation in seiner kodierenden Sequenz ursächlich für die VC-Armut verantwortlich ist. Externe Ergebnisse einer konkurrierenden Arbeitsgruppe lieferten dazu die Bestätigung und experimentelle Belege. Diese Mutation (eine zwei Basen Insertion) könnte der Ausgangspunkt für die Entwicklung eines Assays sein, der noch wesentlich verlässlicher eine VC-Armut vorhersagen könnte, als die bisher entwickelten, eng korrelierten SNP-Marker, da diese Mutation und der davon abgeleitete Assay identisch sind.

Die Kenntnis von einer weiteren, eng verwandten, aber nicht genetisch gekoppelten RibA1-Gen, kann unter Umständen genutzt werden, um den VC-Gehalt in *V.faba* in Samen züchterisch noch weiter zu verringern, sofern sich für diese Gen geeignete Allele identifizieren lassen. Allerdings besteht auch die Möglichkeit, dass eine weitere Reduktion zu im Ertrag beeinträchtigten oder gar letalen Phänotypen führen könnte. Dies wäre Gegenstand eines neu zu beantragenden Projekts.

Auch die drei verbleibenden Kandidatengen aus den Expressionsanalysen kommen für weitergehende Untersuchungen in einem neuen Projekt in Betracht, um ihre mögliche Verbindung zur VC-Armut aufklären zu können. Hier wären auch mehr biochemische Aspekte experimentell zu berücksichtigen.

Anhand der erarbeiteten Transkriptom-Daten war eine vollständige Aufklärung des kompletten VC-Biosynthesewegs bisher nicht möglich. Es konnte aber ein wichtiger Baustein identifiziert werden, das VC1/RibA1-Gen, welches den ersten und wichtigsten Schritt hin zu Vicin und Convicin katalysiert, wie externe Daten gezeigt haben.

Nicht alle neuen Informationen zu SNPs und Kandidatengen konnten im Rahmen des Projektes erschöpfend genutzt werden. Die Hauptziele der Generierung von geeigneten SNP-Kandidaten und Kandidatengen für die VC-Biosynthese wurden jedoch erreicht.

Zusammenfassung

Vicin/Convicin-arme Ackerbohnen stellen ein äußerst erstrebenswertes Zuchtziel dar. Im Rahmen des AboVici-Projektes wurden in Neustadt genomweiten Transkriptomanalysen an sorgfältig ausgewählten, phänotypisch und genetisch bestens charakterisierten Ackerbohnen-Pflanzen durchgeführt. Die verwendeten Genotypen zeichneten sich durch maximalen Unterschieden im VC-Gehalt aus. Damit konnten neue, eng mit dem VC-Locus gekoppelte SNP-Marker identifiziert werden, die Eingang in eine Marker-gestützte Selektion von VC-armen Winterbohnen fanden. Außerdem wurde ein Kandidatengen für die VC-Biosynthese identifiziert, welches weitere Möglichkeiten für eine gezielte Züchtung VC-armer Ackerbohnen liefert.

Literaturverzeichnis

- Björnsdotter**, E.; Nadzieja, M.; Chang, W.; Escobar-Herrera, L.; Mancinotti, D.; Angra, D.i. et al. (2020): VC1 catalyzes a key step in the biosynthesis of vicine from GTP in faba bean. In: bioRxiv, 2020.02.26.966523. DOI: 10.1101/2020.02.26.966523.
- Brown**, E. G.; Roberts, F. M. (1972) Formation of vicine and convicine by *Vicia faba*. In: Phytochemistry 11 (11), S. 3203–3206.
- Cappellini**, M. D., & Fiorelli, G. (2008) Glucose-6-phosphate dehydrogenase deficiency. Lancet, 371, 64–74.
- Crépon**, Katell; Marget, Pascal; Peyronnet, Corinne; Carrouée, Benoit; Arese, Paolo; Duc, Gérard (2010): Nutritional value of faba bean (*Vicia faba* L.) seeds for feed and food. In: Fields Crops Research (115), S. 329–339.
- Cruz-Izquierdo**, S., Avila, C. M., Satovic, Z., Palomino, C., Gutierrez, N., Ellwood, S. R., et al. (2012) Comparative genomics to bridge *Vicia faba* with model and closely-related legume species: stability of QTLs for flowering and yield-related traits. Theor. Appl. Genet. 125, 1767–1782
- Duc** G, Sixdenier G, Lila M, Furstoss V (1989) Search of genetic variability for vicine and convicine content in *Vicia faba* L. A first report of a gene which codes for nearly zero-vicine and zero-convicine contents In: 1. International Workshop on 'Antinutritional Factors (ANF) in Legume Seeds', Wageningen (Netherlands), 23-25 Nov 1988: Pudoc. Wageningen, The Netherlands (Pbs), pp 305–313
- Goyoaga**, C., Burbano, C., Cuadrado, C.,Varela, Al.; Guillamón, E., Pedrosa, M. M., Muzquiz, Mercedes (2008) Content and distribution of vicine, convicine and l-DOPA during germination and seedling growth of two *Vicia faba* L. varieties. In: Eur Food Res Technol 227 (5), S. 1537–1542.
- Griffiths**, D. W., & Ramsay, G. (1992) The concentration of vicine and convicine in *Vicia faba* and some related species and their distribution in mature seeds. Journal of the Science of Food and Agriculture, 59, 463–468.
- Halle**, I. (2019) (Institut für Tierernährung, FLI, Braunschweig) Untersuchungen zum Einfluss eines steigenden Anteils an Winter-Ackerbohnen in der Futtermischung auf Legeleistungsmerkmale und Eiqualität von Legehühnern 15. Tagung Schweine- und Geflügelernährung, 19. - 21. November 2019, Lutherstadt Wittenberg : Tagungsband (2019), S. 139-141.
- Khamassi**, K., Ben Jeddi, F., Hobbs, D., Irigoyen, J., Stoddard, F., O'Sullivan, D. M., Jones, H. (2013) A baseline study of vicine–convicine levels in faba bean (*Vicia faba* L.) germplasm. In: Plant Genet. Resour. 11 (3), S. 250–257.
- Khazaei**, H.; O'Sullivan, D.I M.; Jones, H.; Pitts, N.; Sillanpää, M. J.; Pärssinen, P. et al. (2015): Flanking SNP markers for vicine–convicine concentration in faba bean (*Vicia faba* L.). In: Mol Breeding 35 (1), pp. 1–6. DOI: 10.1007/s11032-015-0214-8.
- Khazaei**, H.; Purves, R. W.; Song, M.; Stonehouse, R.; Bett, K. E.; Stoddard, F. L.; Vandenberg, A. (2017): Development and validation of a robust, breeder-friendly molecular marker for the vc - locus in faba bean. In: Mol Breeding 37 (11). DOI: 10.1007/s11032-017-0742-5.
- Köpke**, U.und Nemecek, T. (2010) Ecological services of faba bean. In: Field Crops Research 115 (3), S. 217–233.
- Kulak**, M., Graves, A., Chatterton, J. (2013) Reducing greenhouse gas emissions with urban agriculture: A Life Cycle Assessment perspective. In: Landscape and Urban Planning 111, S. 68–78.
- Link**, W., Stelling, D., Ebmeyer, E. (1994) Factors determining the performance of synthetics in *Vicia faba* L. 1. Heterogeneity, heterozygosity, and degree of cross-fertilization. In: Euphytica 75 (1-2), S. 77–84.
- O'Sullivan**, D. M., Angra, D. (2016) Advances in Faba Bean Genetics and Genomics Front. Genet., 7,150
- Ray**, H., Bock, C., & Georges, F. (2015) Faba bean: Transcriptome analysis from etiolated seedling and developing seed coat of key cultivars for synthesis of proanthocyanidins, phytate, raffinose family oligosaccharides, vicine, and convicine. The Plant Genome, 8
- Song**, M. (2017): Preventing Favism by Selecting Faba Bean Mutants Using Molecular Markers. In: STEM Fellowship Journal 3 (1), pp. 2–6. DOI: 10.17975/sfj-2017-001.
- Statistisches Bundesamt** (2020) Ackerland nach Hauptfruchtgruppen und Fruchtarten. Online verfügbar unter <https://www.destatis.de/DE/Themen/Branchen-Unternehmen/Landwirtschaft-Forstwirtschaft-Fischerei/Feldfruechte-Gruenland/Tabellen/ackerland-hauptfruchtgruppen-fruchtarten.html>, zuletzt aktualisiert am 03.08.2020, zuletzt geprüft am 27.09.2020.
- Tacke** R., Ecke W., Höfer M., Sass O., Link W. Zooming into the genomic vicinity of the major locus for vicine and convicine in faba bean (*Vicia faba* L.). Plant Breeding Manuskript eingereicht
- Untergasser**, A. Harm Nijveen, Xiangyu Rao, Ton Bisseling, René Geurts, and Jack A.M. Leunissen: Primer3Plus, an enhanced web interface to Primer3 Nucleic Acids Research 2007 35: W71-W74; doi:10.1093/nar/gkm306

Übersicht über alle im Berichtszeitraum vom Projektnehmer realisierten Veröffentlichungen zum Projekt, bisherige und geplante Aktivitäten zur Verbreitung der Ergebnisse

Internetpräsenz des AboVici-Projektes

Ergebnisse und Fortschrittsberichte wurden kontinuierlich gemeinsam mit dem Projektkoordinator Prof. W. Link auf der Website des Abo-Vici-Projektes in Göttingen eingestellt. (<https://www.uni-goettingen.de/de/abo-vici-projekt/559637.html>).

Jahresberichte

Jeweils ein Artikel in den Jahresberichten „2017 und 2018“ sowie „2019“ der RLP AgroScience GmbH, die auf deren Homepage veröffentlicht wurden (<http://aplanta.agroscience.de/index.php/de/presse/jahresberichte>)

Vorträge

Im Rahmen der drei AboVici-Projektmeetings 2017, 2018 und 2019 wurde jeweils von M. Höfer in einem Vortrag über die Fortschritte der Arbeitsgruppe aus Neustadt berichtet.

Höfer, M. Züchtung und Agronomie neuartiger, Vicin-armer Ackerbohnen und Einsatz als einheimisches Eiweißfutter, Vortrag auf Sitzung des wissenschaftlichen Beirats der RLP AgroScience GmbH, 22.02.2019.

Höfer, M. Transkriptomanalyse zur Identifizierung neuer molekularer Marker für Vicin/Convicin-arme Ackerbohnen („Abo-Vici“ Projekt) Workshop Molekulare Pflanzenzüchtung – Gesellschaft für Pflanzenbiotechnologie am 5. und 6. September 2019 in Geisenheim.

Poster

Tacke, **Höfer**, Angra, O'Sullivan, and Link, 2019. ICLGG, Frankreich. “High-resolution map fragments from three bi-parental crosses to zoom into the genomic vicinity of a known major gene for very low vicine and convicine seed content in faba bean (*Vicia faba* L.)”.

Tacke, **Höfer**, Link, 2019. CiBreed Workshop, Deutschland (Göttingen). “Fine-mapping of a major gene for very low vicine & convicine seed content in faba bean (*Vicia faba* L.)”.

Tacke, Ecke, **Höfer**, Link, 2020. GPZ, Tulln, Österreich. Fine-mapping of two bi-parental crosses to zoom into the genomic vicinity of the major QTL for very low vicine & convicine seed content in faba bean (*Vicia faba* L.).

Publikationen - Manuskript eingereicht

Tacke R., Ecke W., **Höfer** M., Sass O., Link W. Zooming into the genomic vicinity of the major locus for vicine and convicine in faba bean (*Vicia faba* L.). Plant Breeding Manuskript eingereicht

Geplante Publikationen / in Vorbereitung

Höfer M., Tacke, W. Link. Gene expression profiling in *Vicia faba* genotypes exhibiting contrasting vicine and convicine content.

Abo-Vici - Abschlußbericht - RLP AgroScience GmbH

Züchtung und Agronomie neuartiger, Vicin-armer Ackerbohnen und Einsatz als einheimisches Eiweißfutter
 Förderkennzeichen: 2815EPS066
 23 April 2021

Anhang I

Finale Liste der SNP-Kandidaten, lokalisierbar auf Chr.2 von *M.truncatula*

SNP-Kandidaten-Schnittmenge: 139 & 848 & BC3 & Sixdenier -RNASeq & MACE-DATASET - Majority_strict & Homozygot_strict -Chr.2 - 2017 & 2020			
SNP/gene_ID	Position	Annotation	Mt.-Chr.2-Region
Contig6812	287	A0A152X9E0_CICAR Peroxidase OS=Cicer arietinum GN=LOC101503097 PE=3 SV=1	1,320,001
Contig6812	370	A0A152X9E0_CICAR Peroxidase OS=Cicer arietinum GN=LOC101503097 PE=3 SV=1	1,320,001
Contig5730	770	A0A152XB81_CICAR Phosphatidylserine decarboxylase proenzyme 2 OS=Cicer arietinum GN=LOC101504195 PE=3 SV=1	1,340,001
Contig5730	920	A0A152XB81_CICAR Phosphatidylserine decarboxylase proenzyme 2 OS=Cicer arietinum GN=LOC101504195 PE=3 SV=1	1,340,001
Contig5730	1151	A0A152XB81_CICAR Phosphatidylserine decarboxylase proenzyme 2 OS=Cicer arietinum GN=LOC101504195 PE=3 SV=1	1,340,001
Contig5730	1415	A0A152XB81_CICAR Phosphatidylserine decarboxylase proenzyme 2 OS=Cicer arietinum GN=LOC101504195 PE=3 SV=1	1,340,001
Contig5730	1532	A0A152XB81_CICAR Phosphatidylserine decarboxylase proenzyme 2 OS=Cicer arietinum GN=LOC101504195 PE=3 SV=1	1,340,001
Contig5730	1760	A0A152XB81_CICAR Phosphatidylserine decarboxylase proenzyme 2 OS=Cicer arietinum GN=LOC101504195 PE=3 SV=1	1,340,001
Contig10165_3	1737	D2D582_MEDTR GT2 OS=Medicago truncatula GN=11429237 PE=2 SV=1	1,350,001
Contig11262	408	G71LA4_MEDTR Glycosyltransferase OS=Medicago truncatula GN=11429235 PE=3 SV=1	1,350,001
Contig11262	1354	G71LA4_MEDTR Glycosyltransferase OS=Medicago truncatula GN=11429235 PE=3 SV=1	1,350,001
Contig11262	1412	G71LA4_MEDTR Glycosyltransferase OS=Medicago truncatula GN=11429235 PE=3 SV=1	1,350,001
Contig11262	1603	G71LA4_MEDTR Glycosyltransferase OS=Medicago truncatula GN=11429235 PE=3 SV=1	1,350,001
Contig6063	1140	G71LA7_MEDTR UDP-glucosyltransferase family protein OS=Medicago truncatula GN=11420634 PE=4 SV=1	1,360,001
Contig6063	1267	G71LA7_MEDTR UDP-glucosyltransferase family protein OS=Medicago truncatula GN=11420634 PE=4 SV=1	1,360,001
Contig5797	257	A0A152X9E4_CICAR beta-xylosidase/alpha-L-arabinofuranosidase 1 OS=Cicer arietinum GN=LOC101506119 PE=4 SV=1	1,370,001
Contig5797	362	A0A152X9E4_CICAR beta-xylosidase/alpha-L-arabinofuranosidase 1 OS=Cicer arietinum GN=LOC101506119 PE=4 SV=1	1,370,001
Contig5797	906	A0A152X9E4_CICAR beta-xylosidase/alpha-L-arabinofuranosidase 1 OS=Cicer arietinum GN=LOC101506119 PE=4 SV=1	1,370,001
Contig5797	2406	A0A152X9E4_CICAR beta-xylosidase/alpha-L-arabinofuranosidase 1 OS=Cicer arietinum GN=LOC101506119 PE=4 SV=1	1,370,001
Contig9470	1150	A0A152XB89_CICAR F-box/kelch-repeat protein At1g22040-like OS=Cicer arietinum GN=LOC101507430 PE=4 SV=1	1,470,001
Contig9470	1402	A0A152XB89_CICAR F-box/kelch-repeat protein At1g22040-like OS=Cicer arietinum GN=LOC101507430 PE=4 SV=1	1,470,001
Contig20221_2	505	G71NT7_MEDTR SpoIIIE-like protein phosphatase domain protein OS=Medicago truncatula GN=11430345 PE=4 SV=2	1,540,001
Contig20221_2	986	G71NT7_MEDTR SpoIIIE-like protein phosphatase domain protein OS=Medicago truncatula GN=11430345 PE=4 SV=2	1,540,001
Contig20221_2	1444	G71NT7_MEDTR SpoIIIE-like protein phosphatase domain protein OS=Medicago truncatula GN=11430345 PE=4 SV=2	1,540,001
Contig20221_2	1475	G71NT7_MEDTR SpoIIIE-like protein phosphatase domain protein OS=Medicago truncatula GN=11430345 PE=4 SV=2	1,540,001
Contig27050	453	A0A152X9E8_CICAR probable protein phosphatase 2C 62 OS=Cicer arietinum GN=LOC101509107 PE=4 SV=1	1,540,001
Contig27050	536	A0A152X9E8_CICAR probable protein phosphatase 2C 62 OS=Cicer arietinum GN=LOC101509107 PE=4 SV=1	1,540,001
Contig3485	104	G71NT9_MEDTR Chlorophyll a-b binding protein, chloroplastic OS=Medicago truncatula GN=11430346 PE=3 SV=1	1,550,001
Contig3485	252	G71NT9_MEDTR Chlorophyll a-b binding protein, chloroplastic OS=Medicago truncatula GN=11430346 PE=3 SV=1	1,550,001
Contig3485	309	G71NT9_MEDTR Chlorophyll a-b binding protein, chloroplastic OS=Medicago truncatula GN=11430346 PE=3 SV=1	1,550,001
Contig3485	876	G71NT9_MEDTR Chlorophyll a-b binding protein, chloroplastic OS=Medicago truncatula GN=11430346 PE=3 SV=1	1,550,001
Contig17017	1461	G71NV3_MEDTR Dentin sialophosphoprotein-like protein, putative OS=Medicago truncatula GN=11430750 PE=4 SV=1	1,590,001
Contig1696	2740	IA0A151T246_CAJCA Potassium transporter OS=Cajanus cajan GN=KK1_023570 PE=3 SV=1	1,640,001
Contig17637	728	A0A152X9F5_CICAR calmodulin-binding transcription activator 2-like isoform X2 OS=Cicer arietinum GN=LOC101514494 PE=4 SV=1	1,650,001
Contig701_2	1170	A0A151T0H3_CAJCA Calmodulin-binding transcription activator 2 OS=Cajanus cajan GN=KK1_022967 PE=4 SV=1	1,650,001
TRINITY_DN38912_c1_g1_i3	259	A0A152YVH8_CICAR calmodulin-binding transcription activator 2-like isoform X2 OS=Cicer arietinum GN=LOC101498639 PE=4 SV=1	1,650,001
Contig4274	259	A0A152XBW2_CICAR Ribosome biogenesis regulatory protein OS=Cicer arietinum GN=LOC101515036 PE=3 SV=1	1,670,001
Contig4274	866	A0A152XBW2_CICAR Ribosome biogenesis regulatory protein OS=Cicer arietinum GN=LOC101515036 PE=3 SV=1	1,670,001
Contig12307	1111	G71NX3_MEDTR ICE-like protease (Caspase) p20 domain protein OS=Medicago truncatula GN=11429238 PE=4 SV=1	1,710,001
Contig12307	1791	G71NX3_MEDTR ICE-like protease (Caspase) p20 domain protein OS=Medicago truncatula GN=11429238 PE=4 SV=1	1,710,001
Contig12307	1973	G71NX3_MEDTR ICE-like protease (Caspase) p20 domain protein OS=Medicago truncatula GN=11429238 PE=4 SV=1	1,710,001
Contig18427	1435	G71PJ6_MEDTR Methyl-CpG-binding domain protein OS=Medicago truncatula GN=11421605 PE=4 SV=2	1,740,001
Contig18427	1634	G71PJ6_MEDTR Methyl-CpG-binding domain protein OS=Medicago truncatula GN=11421605 PE=4 SV=2	1,740,001
Contig18427	2300	G71PJ6_MEDTR Methyl-CpG-binding domain protein OS=Medicago truncatula GN=11421605 PE=4 SV=2	1,740,001
Contig4459_2	54	G71PJ9_MEDTR Phospho-2-dehydro-3-deoxyheptonate aldolase OS=Medicago truncatula GN=11418834 PE=3 SV=1	1,760,001
Contig4459_2	1473	G71PJ9_MEDTR Phospho-2-dehydro-3-deoxyheptonate aldolase OS=Medicago truncatula GN=11418834 PE=3 SV=1	1,760,001
Contig4459_2	1729	G71PJ9_MEDTR Phospho-2-dehydro-3-deoxyheptonate aldolase OS=Medicago truncatula GN=11418834 PE=3 SV=1	1,760,001
Contig4459_2	1797	G71PJ9_MEDTR Phospho-2-dehydro-3-deoxyheptonate aldolase OS=Medicago truncatula GN=11418834 PE=3 SV=1	1,760,001
Contig19011	621	A0A0B2QR29_GLYSO Splicing factor 3B subunit 1 (Fragment) OS=Glycine soja GN=glysoja_049832 PE=4 SV=1	1,780,001
Contig7609	115	A0A152X9G3_CICAR tRNA (Cytosine(38)-C(5))-methyltransferase isoform X1 OS=Cicer arietinum GN=LOC101491158 PE=3 SV=1	1,800,001
Contig7609	325	A0A152X9G3_CICAR tRNA (Cytosine(38)-C(5))-methyltransferase isoform X1 OS=Cicer arietinum GN=LOC101491158 PE=3 SV=1	1,800,001
Contig7609	392	A0A152X9G3_CICAR tRNA (Cytosine(38)-C(5))-methyltransferase isoform X1 OS=Cicer arietinum GN=LOC101491158 PE=3 SV=1	1,800,001
Contig7609	1131	A0A152X9G3_CICAR tRNA (Cytosine(38)-C(5))-methyltransferase isoform X1 OS=Cicer arietinum GN=LOC101491158 PE=3 SV=1	1,800,001
Contig7559_2	515	G71PL2_MEDTR 2-oxoglutarate/malate translocator OS=Medicago truncatula GN=11421607 PE=4 SV=1	1,830,001
Contig14304	216	G71PL5_MEDTR 3,4-dihydroxy-2-butanone 4-phosphate synthase OS=Medicago truncatula GN=11418835 PE=3 SV=1	1,850,001
Contig14304	687	G71PL5_MEDTR 3,4-dihydroxy-2-butanone 4-phosphate synthase OS=Medicago truncatula GN=11418835 PE=3 SV=1	1,850,001
Contig14304	726	G71PL5_MEDTR 3,4-dihydroxy-2-butanone 4-phosphate synthase OS=Medicago truncatula GN=11418835 PE=3 SV=1	1,850,001
Contig14304	1055	G71PL5_MEDTR 3,4-dihydroxy-2-butanone 4-phosphate synthase OS=Medicago truncatula GN=11418835 PE=3 SV=1	1,850,001
Contig14304	1574	G71PL5_MEDTR 3,4-dihydroxy-2-butanone 4-phosphate synthase OS=Medicago truncatula GN=11418835 PE=3 SV=1	1,850,001
Contig14304	1598	G71PL5_MEDTR 3,4-dihydroxy-2-butanone 4-phosphate synthase OS=Medicago truncatula GN=11418835 PE=3 SV=1	1,850,001
Contig14304	1602	G71PL5_MEDTR 3,4-dihydroxy-2-butanone 4-phosphate synthase OS=Medicago truncatula GN=11418835 PE=3 SV=1	1,850,001
Contig18365	1609	G71PM3_MEDTR Phototropic-responsive NPH3 family protein OS=Medicago truncatula GN=11428537 PE=3 SV=1	1,880,001
Contig3200_2	946	G71PN8_MEDTR Carbonic anhydrase OS=Medicago truncatula GN=11433869 PE=3 SV=1	1,930,001
Contig3200_2	1129	G71PN8_MEDTR Carbonic anhydrase OS=Medicago truncatula GN=11433869 PE=3 SV=1	1,930,001
Contig4170	377	G71PP0_MEDTR Chaperone DnaJ-domain protein OS=Medicago truncatula GN=11433870 PE=4 SV=2	1,950,001
Contig4170	1685	G71PP0_MEDTR Chaperone DnaJ-domain protein OS=Medicago truncatula GN=11433870 PE=4 SV=2	1,950,001
Contig4170	1853	G71PP0_MEDTR Chaperone DnaJ-domain protein OS=Medicago truncatula GN=11433870 PE=4 SV=2	1,950,001
Contig9103_2	463	G71S57_MEDTR Ulp1 protease family, carboxy-terminal domain protein OS=Medicago truncatula GN=11424218 PE=4 SV=2	1,970,001
Contig9103_2	581	G71S57_MEDTR Ulp1 protease family, carboxy-terminal domain protein OS=Medicago truncatula GN=11424218 PE=4 SV=2	1,970,001
Contig9103_2	3101	G71S57_MEDTR Ulp1 protease family, carboxy-terminal domain protein OS=Medicago truncatula GN=11424218 PE=4 SV=2	1,970,001
Contig22331	1455	Q2HV44_MEDTR Centromere protein, putative OS=Medicago truncatula GN=11419789 PE=4 SV=1	1,980,001

In gelb sind diejenigen SNP-Kandidaten markiert, die bisher als KASP-Marker verwendet wurden.

Abo-Vici - Abschlußbericht - RLP AgroScience GmbH

Züchtung und Agronomie neuartiger, Vicin-armer Ackerbohnen und Einsatz als einheimisches Eiweißfutter
 Förderkennzeichen: 2815EPS066 23 April 2021

Anhang II

Finale Liste der SNP-Kandidaten, nicht lokalisierbar auf dem Genom von *M.truncatula*

Schnittmenge: 139 & 848 & BC3 & Sixdenier -RNASeq & MACE-DATASET - Majority_strict & Homozygot_strict - "NA" - 2017 & 2019			
SNP/gene_ID	Position	Annotation	Mt.-Genom
Contig17671	206	G7L410_MEDTR Heavy-metal-associated domain protein OS=Medicago truncatula GN=11435142 PE=4 SV=2	n.a.
Contig17671	303	G7L410_MEDTR Heavy-metal-associated domain protein OS=Medicago truncatula GN=11435142 PE=4 SV=2	n.a.
Contig17671	307	G7L410_MEDTR Heavy-metal-associated domain protein OS=Medicago truncatula GN=11435142 PE=4 SV=2	n.a.
Contig17671	321	G7L410_MEDTR Heavy-metal-associated domain protein OS=Medicago truncatula GN=11435142 PE=4 SV=2	n.a.
Contig17671	760	G7L410_MEDTR Heavy-metal-associated domain protein OS=Medicago truncatula GN=11435142 PE=4 SV=2	n.a.
Contig17671	761	G7L410_MEDTR Heavy-metal-associated domain protein OS=Medicago truncatula GN=11435142 PE=4 SV=2	n.a.
Contig17671	764	G7L410_MEDTR Heavy-metal-associated domain protein OS=Medicago truncatula GN=11435142 PE=4 SV=2	n.a.
Contig9400	353	A0A1S2X9G1_CICAR metacaspase-3-like isoform X1 OS=Cicer arietinum GN=LOC101488786 PE=4 SV=1	n.a.
TRINITY_DN44477_c0_g1_i2	444	A0A1S2XBU2_CICAR non-specific lipid-transfer protein-like protein At5g64080 OS=Cicer arietinum OX=3827 GN=LOC101508788 PE=3 SV=1	n.a.
TRINITY_DN44477_c0_g1_i2	593	A0A1S2XBU2_CICAR non-specific lipid-transfer protein-like protein At5g64080 OS=Cicer arietinum OX=3827 GN=LOC101508788 PE=3 SV=1	n.a.
TRINITY_DN44477_c0_g1_i2	668	A0A1S2XBU2_CICAR non-specific lipid-transfer protein-like protein At5g64080 OS=Cicer arietinum OX=3827 GN=LOC101508788 PE=3 SV=1	n.a.
TRINITY_DN44477_c0_g1_i2	1066	A0A1S2XBU2_CICAR non-specific lipid-transfer protein-like protein At5g64080 OS=Cicer arietinum OX=3827 GN=LOC101508788 PE=3 SV=1	n.a.
TRINITY_DN44477_c0_g1_i2	1255	A0A1S2XBU2_CICAR non-specific lipid-transfer protein-like protein At5g64080 OS=Cicer arietinum OX=3827 GN=LOC101508788 PE=3 SV=1	n.a.
TRINITY_DN44477_c0_g1_i2	1442	A0A1S2XBU2_CICAR non-specific lipid-transfer protein-like protein At5g64080 OS=Cicer arietinum OX=3827 GN=LOC101508788 PE=3 SV=1	n.a.
TRINITY_DN44477_c0_g1_i2	1454	A0A1S2XBU2_CICAR non-specific lipid-transfer protein-like protein At5g64080 OS=Cicer arietinum OX=3827 GN=LOC101508788 PE=3 SV=1	n.a.
TRINITY_DN44477_c0_g1_i2	1482	A0A1S2XBU2_CICAR non-specific lipid-transfer protein-like protein At5g64080 OS=Cicer arietinum OX=3827 GN=LOC101508788 PE=3 SV=1	n.a.
TRINITY_DN44477_c0_g1_i2	1508	A0A1S2XBU2_CICAR non-specific lipid-transfer protein-like protein At5g64080 OS=Cicer arietinum OX=3827 GN=LOC101508788 PE=3 SV=1	n.a.
TRINITY_DN44477_c0_g1_i2	1522	A0A1S2XBU2_CICAR non-specific lipid-transfer protein-like protein At5g64080 OS=Cicer arietinum OX=3827 GN=LOC101508788 PE=3 SV=1	n.a.
TRINITY_DN44477_c0_g1_i2	1539	A0A1S2XBU2_CICAR non-specific lipid-transfer protein-like protein At5g64080 OS=Cicer arietinum OX=3827 GN=LOC101508788 PE=3 SV=1	n.a.
TRINITY_DN49974_c1_g1_i2	126	uncharacterized RNA	n.a.

In **gelb** sind diejenigen SNP-Kandidaten markiert, die bisher als KASP-Marker verwendet wurden.

Anhang III

Kurz gefasster Erfolgskontrollbericht

1. Beitrag des Ergebnisses zu den förderpolitischen Zielen

Die Ergebnisse des Projektteil „molekulargenetische Aufgaben“, die in Neustadt bearbeitet wurden, lieferten -in enger Kooperation mit dem Projektkoordinator Prof. Dr. W. Link- mögliche Werkzeuge für eine schneller und effizienter Selektion des Merkmals „Vicin/Convicin-Armut“ in der Ackerbohne. Diese wurden in Göttingen verifiziert und validiert. Die Züchtung einer neuen, VC-armen Winterbohne wurde damit ermöglicht. Die AG Neustadt erarbeitete im Projekt mit dem von der AG Göttingen zur Verfügung gestellten Pflanzenmaterial vorwiegend Daten, deren Nutzen sich erst in der genetisch-züchterischen Arbeit der AG Göttingen erweisen musste. Der mögliche Beitrag der AG Neustadt ist somit ein eher indirekter, da die Ergebnisse erst durch die Arbeiten der AG Göttingen ihre potentielle Wirkung entfalten können. Die darauf basierende Züchtungserfolge der AG Link in Göttingen leisten einen direkten Beitrag zur qualitativen Verbesserung bzw. Beschleunigung der Ackerbohnenzüchtung. Die Etablierung von VC-armen Winterackerbohnen erhöht mittel- bis langfristig den Anbau und die bekannten günstigen Wirkungen von proteinreichen, einheimischen Eiweißpflanzen in Deutschland. Damit unterstützen die in Neustadt erzielten Ergebnisse letztlich die Eiweißpflanzenstrategie des BLEM.

2. Ergebnis des Vorhabens und die gesammelten wesentlichen Erfahrungen

Das Teilprojekt lieferte konkret in der Züchtung verwertbare SNP-Kandidaten und weitere Kandidatengene, ausschließlich basierend auf der NGS-Technologie, die sich damit als sehr nützlich erwies. Voraussetzung für den Erfolg eines derartigen Ansatzes ist allerdings das hervorragend genetisch charakterisierte und züchterisch vorbereitete Pflanzenmaterial, welches die AG Göttingen und Leitung von Prof. Dr. W. Link vorbereite und geeignete Proben zur Verfügung stellte.

Ein kritischer Punkt war die Verfügbarkeit des Pflanzenmaterials, welches die unabdingbare Voraussetzung für den Projektstart der Arbeiten in Neustadt war. Die aktuelle Jahreszeit bzw. der Vegetationsperiode der Ackerbohne spielen eine wichtige Rolle, da -im Gegensatz zu Blättern- die für die Untersuchung benötigten Samenschalen erst gegen Ende des Lebenszyklus einer Ackerbohne nach der Befruchtung entstehen und geerntet werden können. Der aus administrativen Gründen festgelegten Projektstart und die Verfügbarkeit des benötigten Pflanzenmaterials passten für die AG in Neustadt nicht ganz zusammen. Dadurch kam es leichten Verzögerungen im Projektverlauf.

Die Sequenzierung musste schon aus technischen Gründen -wie heutzutage üblich- an einen Sequenzier-Dienstleister als Unterauftrag abgegeben werden. Dieser hatte auch die Expertise für die weitergehende Bioinformatik zu Expressions-, SNP- und INDel-Analysen, was in dieser Kombination nicht oder nur zu erheblichen finanziellen Aufschlägen bei angefragten Firmen angeboten wurde. Allerdings kam es wie immer wieder zu zeitlichen Verzögerungen, teilweise bedingt durch im Nachhinein zu wiederholende Auswertungen der Rohdaten, die sich erheblich auf den Ablauf des Projektes auswirkten. Es kam dadurch mehrfach zur Verschiebung des Arbeitsplans und des Abrufens der geplanten Finanzmittel. Die daraus entstehenden verwaltungstechnischen Probleme konnten mit Anträgen und der großzügiger Unterstützung bzw. Beratung durch das BLE, namentlich Dr. Luiz Massucati, behoben werden. Als Konsequenz aus dieser Erfahrung daraus wurde darauf geachtet für -aktuelle und zukünftige- Aufträge mit Sequenzierdienstleistern immer einen realistischen und definierten Zeitpunkt der Übergabe der Daten und Ergebnislisten vertraglich zu vereinbaren. Wird der vereinbarte Zeitpunkt

überschritten, so reduziert sich die Vergütung um einen vorher mit der Firma ausgehandelten Betrag. Dies hat inzwischen bei einer anderen Arbeitsgruppe der RLP AgroScience erfolgreich funktioniert.

Wie im Abschlußbericht im Kapitel II. SNP-Analysen und SNP-Kandidaten in den Absätzen „Aspekte der Sequenzier-Methodik und Auswirkungen auf die Ergebnislisten“ und „Zwischenfazit“ ausgeführt, hatte die klassische RNASeq-Methode im Vergleich zu der kostengünstigeren MACE-Methode hinsichtlich der Ausbeute an verwertbaren SNP-Kandidaten deutlich bessere Ergebnisse erzielt. Aufgrund dieser Erfahrung würde in einem neuen Projektantrag auf jeden Fall einem -teureren- RNASeq-Ansatz der Vorzug gegeben werden. Auch eine höhere Sequenziertiefe (für die bessere Erfassung seltener Transkripte) wird in Zukunft angestrebt. Je nach der Anzahl an zu analysierenden Einzelproben (oder gar technischen Replikate, die in der akademischen Forschung Voraussetzung für eine Publizierbarkeit der Ergebnisse ist) führt dies eventuell bei Antragstellung zu Konflikten mit einem möglicherweise einzuhaltenden Budget.

3. Erfindungen/Schutzrechtsanmeldungen und erteilte Schutzrechte

Alle im Rahmen des Abo-Vici Projekt erzeugten Erkenntnisse und Daten stehen über die wissenschaftlichen Publikationen, Vorträge und Posterpräsentationen der Wissenschaftsgemeinde und daran interessierten Zuchtfirmen frei zur Verfügung. Es ergeben sich keine Ansprüche auf Patente oder Erfindungen.

4. wirtschaftliche Erfolgsaussichten nach Projektende (mit Zeithorizont)

Da -wie oben schon erwähnt- alle Ergebnisse aus Neustadt über die Arbeiten der AG Göttingen um Prof. W. Link in die Ackerbohnenzüchtung einfließen, ist eine daraus resultierende Wirkung indirekter Natur. Es wird daher auf den Verwertungsplan zum Abschlussbericht (Teilprojekt, Göttinger Arbeiten zu Abo-Vici) von Prof. W. Link verwiesen (S19. Punkt 2) verwiesen.

5. wissenschaftliche und/oder technische Erfolgsaussichten nach Projektende

Es gelten analog die gleichen Überlegungen wie unter 4. Es wird daher auf den Verwertungsplan zum Abschlussbericht (Teilprojekt, Göttinger Arbeiten zu Abo-Vici) von Prof. W. Link verwiesen (S19. Punkt 3) verwiesen.

6. wissenschaftliche und wirtschaftliche Anschlussfähigkeit

Es gelten analog die gleichen Überlegungen wie unter 4 und 5. Es wird daher auf den Verwertungsplan zum Abschlussbericht (Teilprojekt, Göttinger Arbeiten zu Abo-Vici) von Prof. W. Link verwiesen (S19. Punkt 4) verwiesen.

Hervorzuheben wäre noch, dass die minoren Gene, die lt. Prof. W. Link zusätzlich zum vc-Gen bis zu 20%-30% der genetischen Variation des VC-Gehaltes beeinflussen können, auch unter den drei Enzymen zu finden sein könnten, die laut Björnsdotter et al. für einen derzeit noch theoretischen VC-Biosyntheseweg verantwortlich gemacht werden. Sofern geeignetes Pflanzenmaterial zur Verfügung stehen wird, könnten auch für diese Fragestellung NGS-Ansätze in Frage kommen.

Als direkte Folge aus dem AboVic-Projektergebnissen wurde im Rahmen der Bekanntmachung Nr. 07/20/33 der Bundesanstalt für Landwirtschaft und Ernährung (BANz AT 02.09.2020 B5) über die RLP AgroScience GmbH von Dr. M. Höfer am 1.12.2020 eine Projektskizze zu einem Einzel-Projektantrag mit dem Titel „Etablierung eines 100%igen Selektionsmarkers für Vicin-

Armut in der Ackerbohne“ eingereicht. Die Projektidee knüpft direkt an das vorangegangene Abo-Vici-Projekt an unter Verwendung der darin gewonnen Erkenntnisse und unter Zuhilfenahme der von Bjornsdotter et al. (2020) vorab veröffentlichten Daten und wird sowohl von Prof. Dr. W. Link aus auch von der NPZ Lembke (Dr. Olaf Sass) durch Bereitstellung von geeignetem Pflanzenmaterial unterstützt.

7. Präsentationsmöglichkeiten für mögliche Nutzer

Trifft für Neustadt nicht zu. Außer der schon im Abschlußbericht erwähnten Internetpräsenz des AboVici-Projektes und eingereichten und geplanten Publikationen erfolgen Präsentationen für Nutzer ausschließlich über die AG Göttingen unter der Leitung von Prof. W. Link.

8. die Einhaltung der Ausgaben- und Zeitplanung

Wie oben unter Punkt 2 erläutert -sowie aus den Zwischenberichten zu dem berichteten Projektteil im Detail ersichtlich- und kam es in der ursprünglich geplanten Projektlaufzeit vom 1. April 2017 bis zum 31. Januar 2020 mehrfach zu Verzögerungen im Zeitplan, die letztlich zu einer beantragten und genehmigten Verlängerung bis zum 31ten Oktober 2020 führten. Durch diese von der BLE freundlicherweise genehmigten Verlängerungen konnten alle geplanten experimentellen Arbeiten erfolgreich abgeschlossen werden. Die Auswertung der bis dahin erzielten Ergebnisse jedoch (insbesondere die finalen SNP-Auswertungen mit einer Re-Evaluierung unter Berücksichtigung der Kategorie „majority_strict“ s.a. Kapitel II des Abschlussberichtes, Absatz „Die Daten aus dem Pflanzenmaterial von 2019“) war allerdings unerwarteter Weise äußerst umfangreich und zog sich bis März 2021 hin, was zu der erheblichen Verzögerung des Abgabetermins dieses Abschlussberichtes beitrug.

Die von der BLE bereitgestellten finanziellen Mittel wurden über die Projektlaufzeit komplett ausgeschöpft. Dankenswerterweise konnten durch Genehmigung der Mittelaufstockung damit auch die deutlich erhöhten Kosten für eine Taqman-basierten qPCR- Analytik gedeckt werden. Im Laufe des Projektes stellte sich nämlich heraus, dass die ursprünglich geplante und kostengünstige qPCR-Standard-Methode mit dem SYBRGreen-Fluoreszenzfarbstoff nicht geeignet war, einen fehlerfreien Nachweis der Expression von Kandidatengene zu gewährleisten. Ohne die zusätzlichen Mittel wären die qPCR-Validierungen nicht im erforderlichen Umfang möglich gewesen.

Neustadt a.d. Weinstrasse am 23.04.2021



(Dr. Michael Höfer)