

Methodentransfer für ein allgemeingültiges SOP; DETECT Gerste

Dieses SOP dient dem Nachweis einer durch CRISPR/Cas9 induzierten Mutation (-1 bp) im Promotor des Anfälligkeitgens *HvPDIL5-1* auf Chromosom 4H in einer homozygoten Mutante, welche aus der Sommergerste Golden Promise DH1-6 erzeugt wurde.

Das Protokoll für den Nachweis mittels Illumina short-read Technologie umfasst:

I.) DNA-Isolation

Die Isolation der DNA erfolgt aus 1-1,5g einer frisch vermahlenden homogenisierten Probe.

Soweit die notwendigen Qualitätskriterien an das Ergebnis eingehalten werden, ist die exakte Ausführung der DNA-Isolation weitgehend beliebig.

Eine photometrische Messung der isolierten DNA sollte eine minimale Konzentration von 40ng/μL ergeben (bei einem Volumen von 100μL und einem A260/280-Quotienten von ~1.5-1.85).

II.) Amplifikation der Zielregionen

pmNILP.IPKv2.2 (75bp)

NILP-IPKv2-F3 TCGTCGGCAGCGTC AGATGTGTATAAGAGACAG cggaccggccccagagaat

NILP-IPKv2-R3 GTCTCGTGGGCTCGG AGATGTGTATAAGAGACAG gaagacggggagacgaggag

Die Ausführung sollte unter Verwendung von SybrGreen an real-time PCR-Geräten erfolgen. Die Konzentration von SybrGreen muss jedoch auf 1/3 der von Hersteller vorgegebenen Konzentration reduziert werden (d.h. 1/30.000x anstelle von 1/10.000x des Reaktionsvolumens).

Für die Amplifikation müssen alle Proben die gleiche Qualität und Ausgangskonzentration aufweisen. Empfohlen wird 200ngDNA/PCR [5μl einer 40ng/μL-DNA-Lösung].

20μL PCR-Ansatz:

- 5μl 40ng/μL-Template-DNA
- 10μL 2x Mastermix **Platinum Superfi II**
- 0,1μL Forward-Primer [100uM]
- 0,1μL Reverse-Primer [100uM]
- 0,2μL SybrGreen (einer 1zu300 Vorverdünnung in HPLC-Wasser)
- 4,6μL HPLC-Wasser

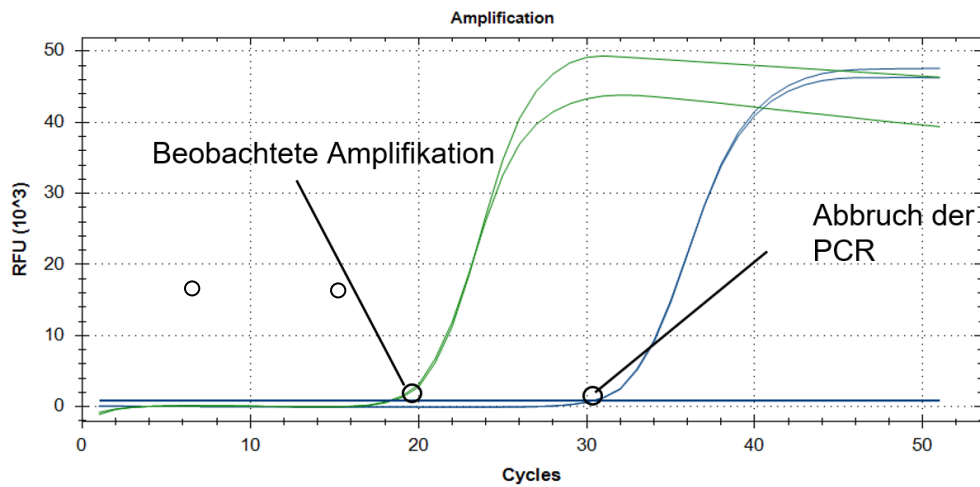
PCR-Programm

98°C/30Sek; [98°C/10sek; 60°C/10Sek.; 72°C/30Sek+Fluoreszenzaufnahme]-max.35 Zyklen*
+5min/72°C; 16°C /Pause

*Aufgrund möglicher Schwankungen in der Quantität der amplifizierbaren Template-DNA, unterschiedlicher Sensitivität der real time PCR-Maschinen und unterschiedlicher Anforderungen an das Limit of Detection (LOD) der Analysen, lässt sich keine absolute Anzahl der PCR-Zyklen definieren. Im Standardprogramm wird (wie abgebildet) von einer maximalen Zyklen-Anzahl von 35 Ct ausgegangen. In der Ausführung muss die Amplifikation jedoch überwacht und beim Erreichen einer notwendigen Zyklenzahl unterbrochen und bei der

finalen Elongation (72°C/5min) fortgesetzt werden. Unnötig lange Amplifikation erhöhen drastisch die Fehlerrate der Analysen.

Beispiel:



Bei der Ausführung der Assays wird der logarithmische Anstieg der PCR nach 19 Zyklen (am Threshold) beobachtet. Beabsichtigt wird eine Sensitivität von 0,01% (1 zu **1000** = $\log(1000)/\log(2) \approx 10Ct$). Damit wird die zyklische Amplifikation nach $19Ct + 10Ct \approx 29$ ten PCR-Zyklus unterbrochen. Für eine Sensitivität von 0,03% gilt dann $19Ct + (\log(3000)/\log(2)) = 19 + 11.6 \approx 31Ct$.

Entscheidend für die quantitative Ausführung ist aber auch, dass die Schwankung der einzelnen Proben am Threshold nicht mehr als $\pm 1,5$ Zyklen beträgt (hier: 17.5 bis 20.5Ct). Bei Beobachtung einer höheren Schwankung soll die Konzentration der Proben neu eingestellt werden und die Amplifikation der Proben muss dann wiederholt werden.

III.) Aufreinigung der PCR-Produkte

Die PCR-Amplifikate werden chemisch mittels ExoI/CIP-System von NEB aufgereinigt, wodurch die Oligos degradiert und dephosphoryliert werden. Hierfür werden 6µL PCR-Produkt und 2µL ExoI/CIP zusammen für 15min bei 37°C inkubiert. Die Deaktivierung der Enzyme erfolgt dann für 10min bei 90°C.

IV.) Indexierung der Amplifikate

Die Indexierung der chemisch aufgereinigten Amplifikate erfolgt mittels einer zweiten PCR.

Indexierungs-Kit: *Nextera XT Index Kit v2* (beinhaltet ausschließlich Primer für die Indexierung)

Polymerase NEB Q5-Polymerase

Ansatz: 50µl

5µl aufgereinigtes Template

5µl S-Oligos

5µl N-Oligos

25µl 2x Q5-Mastermix

10µl HPLC-Wasser

PCR-Programm: 95°C/5min; [95°C/30Sek; 55°C/30Sek; 72°C/30Sek]x7; 72°C; 5min

V.) Aufreinigung der Indexierungsprodukte

Alle Amplifikate aus der Indexierung werden zu einer einzelnen „Prä-Library“ zusammengeführt. Ein Aliquot der Prä-Library (mind. 60µL) wird mittels 2-3% Agarose-Elektrophorese aufgetrennt. Die DNA-Bande wird scharf ausgeschnitten und mittels eines Silica-Membran-Systems aufgereinigt.

VI.) Erstellung der Library

Für die Beladung einer MiSeq-Kassette ist die Gesamtmolarität der Probe und Anteil der IPC (hier: PhiX) entscheidend.

Die Molarität der Library ist von der Kapazität der verwendeten Illumina MiSeq-Kassette abhängig. Hier wird die Vorbereitung der Library für eine MiSeq 300-V2 Chemie beschrieben. Die über Agarose-Elektrophorese aufgereinigte „Prä-Library“ wird mittels QBit fluorometrisch quantifiziert. Die Molarität der Illumina-PhiX-Kontrolle muss nicht berechnet werden, sie ist bekannt.

Nun wird die Konzentration der DNA in die Anzahl der Amplifikate umgerechnet:

Hierfür wird folgende Formel verwendet:

$$([A]/([660 \text{ g/mol}] * [B]) * 10^6 = \text{nM Amplikon}$$

A = Konzentration der Prä-Library nach QBit-Messung in ng/µL

B = mittlere Amplikonlänge (PCR + Adapter und Index) ~ 230bp

Beispiel: Eine Library mit 18,5ng/µl liefert somit 121,9 nM Lösung.

Die Library wird zunächst auf 4nM eingestellt. Hierbei erfolgt folgende Berechnung:

$$(([A]/[B]) - 1) * [C] = [D]$$

A = Molarität vorhandener Lösung

B = Soll-Molarität (hier = 4nM)

C = Volumen der Prä-Library in µL

D = Volumen HPLC-Wasser in µL, das notwendig ist, um die erwünschte Konzentration zu erreichen

Beispiel: 10µL Prä-Library mit der Konzentration 121,9nM müssen zusammen mit 294,8 µL HPLC-Wasser gemischt werden, um eine Konzentration von 4nM zu erreichen. Auch die Konzentration der PhiX-Lösung wird von 10nM auf 4nM eingestellt.

Die Prä-Library und PhiX müssen denaturiert und verdünnt werden:

Vorbereitung: aus einer 2N NaOH (frisch hergestellt! oder zuvor bei -20°C aufbewahrt) werden 1000µl einer 0.2N NaOH Lösung mit HPLC-Wasser angesetzt. Die 0.2N NaOH-Lösung wird bis zur Anwendung auf Eis gelagert. Jeweils 5µL 4nM Prä-Library und 5µL PhiX werden in 2ml Caps mit jeweils 5µL 0.2N NaOH für 5min (gut mischen!) bei Raumtemperatur inkubiert. Die Inkubation wird durch Zugabe von 990µL HT1 Puffer (aus dem Illumina MiSeq 500-V2 Kit) gestoppt. Die Proben werden gründlich gemischt und auf Eis gelagert. Die resultierende Konzentration der Prä-Library und PhiX beträgt nun jeweils 20pM. Für ein MiSeq-Run sollte der Anteil von PhiX an der Gesamt-Library bei circa 30-50% liegen. Dazu werden 700µL der denaturierten Prä-Library (20pM) und 300µL der denaturierten PhiX-Lösung (20pM) vermischt. Anschließend wird diese Lösung mittels HT1-Puffer auf eine finale Konzentration von circa 8pM verdünnt. Hierfür werden 400µL der mit PhiX dotierten Prä-Library (20pM) und 600µL HT1-Puffer vermischt und bis zur Beladung der Kassette lichtgeschützt und auf Eis gelagert. Nach Vorbereitung des Runs wird 1ml der Library in die ausgewiesene Position der Kassette pipettiert.

VII.) NGS-Run

Bei einem MiSeq-Run sollten nicht mehr als 80 Proben gleichzeitig analysiert werden. Die Planung des MiSeqRuns sollte mit der Software Illumina Experiment Manager (unter Windows) durchgeführt werden.

Der MiSeq-Run wird mit 2x 76 Basen dimensioniert. Die Planung mittels Software erfolgt an einem MiSeq-unabhängigen Rechner. Die Software ist benutzerfreundlich-intuitiv und Menü-geführt. Die Planung führt mit der Erstellung eines „Sample-Sheets“ zu einer CSV-Datei, die alle notwendigen Informationen zur Kontrolle des MiSeqs und für das spätere Demultiplexing der Proben enthält.

VIII.) Qualitätskriterien an den NGS-Run

Ein erfolgreicher MiSeq-Run sollte folgende Kriterien erfüllen:

- eine Cluster-Dichte von 700-1300kCluster/mm²
- einen PhiX-Anteil zwischen 30-50%
- einen Anteil von >90% PF Cluster
- mehr als 14 Millionen Reads PF
- Jede Indexierung erreichte weitgehend die gleiche Anzahl an Reads ($\pm 50\%$). Negative Ausreißer sollten wiederholt werden.

IX.) Umwandlung von bcl-Files ins FastQ-Format

Es erfolgt eine manuelle Umwandlung der bcl-Files zum FastQ-Format (unter Linux, mit Illumina bcl2fastq v2.2). Die Software arbeitet unter RedHat-Implementierungen von Linux: z.B.: CentOS7. Bei der Umwandlung ist darauf zu achten, dass:

- nur das SampleSheet des aktuellen MiSeq-Runs verwendet wird
- die Fehlertoleranz für die Indexierung beim Demultiplexing bei 0 liegt
- komprimierte FastQ-Dateien entstehen, die dann mittels gzip-ähnlicher Programme entpackt werden müssen

Syntax (nach Installation; unter CentOS in usr/local/bin/bcl2fastq)

```
bcl2fastq --runfolder-dir /.../ --output-dir /.../ --sample-sheet /.../*.csv --barcode-mismatches 0
```

X.) Rohdaten-Workflow

Die bioinformatische Bearbeitung der Reads wird in zwei Stufen unterteilt: a.) Bereinigung der Rohdaten (Reads) und b.) Auswertung der Sequenzen in Hinblick auf die Anteile der gesuchten Mutation. Als „belastbare Sequenzen“ wird ein bereinigter, hochqualitativer Sequenz-Datensatz bezeichnet. Die Sequenzen des Datensatzes können als „wahrscheinlich fehlerfrei“ und amplikonspezifisch betrachtet werden. Es gibt unterschiedliche Möglichkeiten beide bioinformatische Schritte sachgerecht zu realisieren. Aus diesem Grund wurden nachfolgend nur die qualitätssichernden Vorgaben für die beiden bioinformatischen Workflows beschrieben.

X.a) Bioinformatische Bereinigung der Reads.

1. Das Produkt der Sequenzierung sind zwei Datensätze pro Probe („R1“- und „R2“-Reads) in „fastq“-Format. Die Sequenzen der R1- und R2-Datensätze besitzen eine Länge, die eine Überlappung der Sequenzen ermöglichen soll. Das Zusammenführen der Sequenzen (Bildung einer Consensus Sequenz) hat eine Filterfunktion, daher sollten hier nachfolgende Vorgaben beachtet werden:
 - Der Bereich der R1 und R2-Überlappung sollte mindestens 40nt betragen;
 - Bei mehr als 20% unterschiedlicher Basen im Überlappungsbereich sollte keine Zusammenführung der Reads erfolgen.

- Bei der Zusammenlegung der Sequenzen muss der „Q-Score“ für jede Base des Überlappungsbereich neu berechnet werden. Ausgabeformat ist eine „fastq“-Datei.
 - Nicht zusammengeführte Reads sollten verworfen werden. Die Prozentzahl der zusammengeführten Reads ist ein Qualitätsmerkmal der Sequenzierung und soll überwacht werden.
2. Beschneidung der Oligo-Sequenzen: Die Oligonukleotide für die Target-spezifische Amplifikation wurden künstlich synthetisiert und stellen somit eine signifikante Fehlerquelle dar. Aus diesem Grund ist es ratsam, die zusammengeführten Reads beidseitig, um die Sequenz der Oligonukleotide zu beschneiden. Die Beschneidung kann entweder durch das Kürzen der Sequenz an beiden Enden um die Anzahl der Basen der Oligonukleotide (nicht empfohlen) oder durch das Wiederfinden der Oligos mittels Homologie-Überprüfung (empfehlenswert) erfolgen. Diese Aktion erfüllt eine Filterfunktion. Das Auffinden der Oligo-Sequenzen und die Beschneidung an dieser Stelle selektiert spezifische, vollständige Amplikons in korrekter Länge. Es sollten keine Reads weiterverarbeitet werden, die nicht beide Oligonukleotide enthalten. Nicht zugeschnittene Reads sollten verworfen werden. Ausgabeformat dieser Aktion ist eine „fastq“-Datei.
 3. Beseitigung von potenziell inkorrekten bzw. fehlerbehafteten Reads. Die einzelnen PCR-Schritte vor der Sequenzierung und die Sequenzierung mittels „Sequencing-by-Synthesis“ (SbS) verursacht Artefakte und zeichnet sich durch abnehmende Qualität (der Basen) mit zunehmender Länge der Sequenzierung aus. Aus dem Datensatz sollten daher alle Reads entfernt werden die kürzer sind als 100nt und/oder ein „N“ enthalten und/oder einen potenziellen Fehler enthalten können (empfohlen wird ein „mee=0,05“). Nach der Bereinigung ist das Mitführen der Qualitätsmerkmale pro Base nicht mehr notwendig, daher ist das Ausgabeformat eine „fasta“-Datei. Der prozentuale Anteil an „unbrauchbaren“ bzw. verworfenen Reads ist ein Qualitätsmerkmal der Sequenzierung und sollte überwacht werden.
 4. Die meisten Sequenzen des Datensatzes einer Probe sind zueinander identisch und können auch als solche (mit einer 100%-Homologie) ausgezählt werden. Eine Gruppe identischer Sequenzen wird als „Cluster“ bezeichnet und durch eine Leitsequenz („Uniq“) repräsentiert. Die Leitsequenzen müssen durchnummeriert werden. Die Leitsequenz-Nummer sollte im Kopf der Sequenzen (nach „>“) definiert werden. Die Kopfzeile der Leitsequenzen muss auch die Anzahl der identischen Sequenzen des Clusters aufweisen (size=). Die Ausgabe erfolgt in einer „fasta“-Datei. Anzahl der Cluster pro Anzahl der Sequenzen ist ein Qualitätsmerkmal und sollte überwacht werden.
 5. Man darf davon ausgehen, dass eine „Leitsequenz“, die nur einmal im Experiment vorkommt (ein Singleton) nicht richtig sein kann. Die Singletons können jedoch einen großen Anteil der Cluster bilden, tragen aber nicht wesentlich zum Informationsgehalt des Experimentes bei. Aus diesem Grund sollen alle Singletons verworfen werden. Hieraus resultiert die minimale Clustergröße =2. (min. Cluster-size=2). Die Ausgabe des Datensatzes erfolgt in einer „fasta“-Datei. Die Sequenzen des Datensatzes können nur als „bereinigt“ bzw. als „belastbar“ angesehen werden. Die Gesamtanzahl der belastbaren Sequenzen ist ein Qualitätsmerkmal und sollte überwacht werden.

X.b) Auswertung

Für die Auswertung der Sequenzen gibt es unterschiedliche Möglichkeiten. Die einfachste Art ist die Anwendung eines „multiple sequence alignment“ (MSA)-Algorithmus (z.B. muscle, mafft, clustalw...) und das anschließende Auszählen der Mutationsereignisse an der Position der Mutation. Die Verwendung von diversen Programmen für eine automatische Auszählung von Mutationen an einer bestimmten Stelle ist auch möglich und vielfältig.

Die Sequenzen können jedoch auch mittels „FASTX-Toolkit“ (FASTA Formatter; http://hannonlab.cshl.edu/fastx_toolkit/index.html) linearisiert und anschließend in Excel importiert und untersucht bzw. ausgezählt werden.

Beachte: Die Sensitivität der Analysen hängt sowohl von der Anzahl der Zyklen bei der Amplifikation (im Verhältnis zu der Menge der amplifizierbaren DNA) als auch von der Anzahl der verfügbaren belastbaren Sequenzen ab. Unzureichende Amplifikation bei ausreichender Anzahl an belastbaren Sequenzen kann falschnegative Ergebnisse induzieren - genauso wie eine korrekte Amplifikation bei zu niedriger Anzahl an belastbaren Sequenzen.

XI; Einhaltung der Qualitätsmerkmale der Analysen

Der zulässige Fehler in der Sequenzierung (die Feststellung einer falschen Base an einer beliebigen Position einer belastbaren Sequenz) liegt bei $\leq 0,01\%$.

Sequenzierfehler können in jedem Stadium der Analysen entstehen: bei der Ausführung der PCR, der Indexierung, der Cluster-Anreicherung, durch die Interpretation der „Farbe einer Base“ im SbS; durch Index „Hopping“ und Kontamination. Eine sachgerechte Umsetzung des Protokolls liefert eine reale Sequenzierfehlerrate von $\leq 0,008\%$. Es wird dringend empfohlen, bei der Verifizierung des Verhaftens die umgebungsspezifische Sequenzierfehlerrate zu bestimmen.

Als technische Nachweisgrenze (und Vertrauensgrenze des Assays) wurde ein Wert von $0,03\%$ (mut-DNA in Wt-DNA) festgelegt.

Das „Limit of Detection“ (LOD95) beträgt $0,1\%$ für „Saatgut in Saatgut“-Gemische bzw. $0,05\%$ für DNA in DNA-Gemische.

Die minimale zulässige Clustergröße für belastbare Sequenzen beträgt 2.

Das Bezugssystem für quantitative Angaben ist die ausgezählte Anzahl der mutierten Sequenzen prozentual bezogen auf die Anzahl der belastbaren Sequenzen.

Interne positive Kontrolle (IPK): Die Gesamt-Anzahl der belastbaren Sequenzen stellt eine gleichwertige Äquivalente zur Anwendung einer IPK dar.

Die minimale zulässige Anzahl belastbarer Sequenzen für die Sicherstellung der technischen Nachweisgrenze in der Auswertung beträgt 18.000. Die Anzahl der Sequenzen ergibt sich unter Berücksichtigung des technischen LOD (3000 Sequenzen), der minimalen Mehrfachabdeckung (3x) und der minimalen Clustergröße (2).

Die Konzentrationsabhängigkeit und Linearität des Nachweisverfahrens gilt im Bereich von $0,1\%$ bis 5% (mut-DNA in wt-DNA) beim Einsatz von 200ng gDNA/PCR . Beachte: Eine Überamplifikation der Probe kann die Linearität der Analysen deutlich einschränken.

Die RDSr (short-term precision) beträgt $< 26\%$ im Bereich $0,1\%$ - 5% GVO (mut-DNA in wt-DNA).

Begleitend zu Proben sollten immer Kontrollreaktionen amplifiziert und sequenziert werden:

- Negative Kontrolle (NC), d.h. genomische Wildtyp-DNA (200ng/PCR)
- Template-freie Kontrolle (NTC);
- Positive Kontrolle (PC) am „Limit of Detection“ d.h. wildtypische DNA der analysierten Spezies mit einem Anteil von $0,1\%$ (Saatgut in Saatgut) an gesuchter Mutation; weitere positive Kontrolle am Schwellenwert der Handlungsnotwendigkeit – z.B. $0,9\%$ (Saatgut in Saatgut).

Die Analyse darf als valide betrachtet werden, wenn:

NTC keine auswertbaren Sequenzen liefert. Sequenzen im NTC-Datensatz liefern einen Hinweis auf mögliche Kontamination der Analyseumgebung und/oder „Index-Hopping“.

Im NC-Datensatz sollte der Anteil der Sequenzen gesuchter Mutation einen Wert von $0,01\%$ nicht überschreiten.

Eine Abweichung in der Wiederfindung der positiven Kontrollen sollte einen Wert von 25% nicht überschreiten.