

Optimierung der Entwicklung von Introgressionspopulationen

Optimizing the development of introgression populations

Eva Herzog^{1*}

Abstract

Introgression populations are valuable genetic resources which are developed by marker-assisted backcrossing. Their development is labor-intensive, costly and time consuming, as commonly numerous backcrosses and large population sizes are required. Crossing and selection schemes which are suitable for a limited number of individuals and high-throughput marker analyses would therefore greatly improve the efficiency of this process. We employed computer simulations to design selection strategies for the development of maize introgression populations of 100 lines with population sizes of 360 individuals per backcross generation for DH and S_2 crossing schemes. The number of simultaneous backcross programs was reduced by pre-selection for complete donor chromosomes or donor chromosome halves. Selection for complete donor chromosomes in generations BC_1 and BC_2 was the most efficient strategy if no selection was conducted in generation BC_3 . With an additional selection step in generation BC_3 , selection for donor chromosome halves improved the separation of target segments. Gradually increasing population sizes over backcross generations was advantageous for both DH and S_2 crossing schemes. For DH schemes, this approach reduced the total number of required high-throughput marker analyses. For S_2 crossing schemes, large population sizes in the final backcross generation enabled selection for short target segments and thus reduced the fixation of large donor chromosome segments. The suggested crossing and selection schemes provide guidelines for breeders which help to make the development of introgression populations more efficient.

Keywords

Computer simulation, high-throughput marker analysis, introgression library, marker-assisted backcrossing, *Zea mays*

Einleitung

Introgressionsbibliotheken sind Sets nahe-isogener Linien (NILs), die kurze, durch Marker kontrollierte Segmente des Genoms eines exotischen Donors in einem gemeinsamen genetischen Hintergrund tragen (ESHED und ZAMIR

1994). Sie sind nützliche Ressourcen sowohl für die Identifikation neuer Allele mit vorteilhaften Eigenschaften, als auch für den anschließenden Zuchtprozess, da die neu gefundenen Allele bereits in einem adaptierten Hintergrund vorliegen.

Introgressionsbibliotheken werden für gewöhnlich durch marker-gestützte Rückkreuzung entwickelt, der sich Selbstung oder die Induktion doppelt haploider (DH) Linien anschließt. Die Rückkreuzungsprogramme für ihre Entwicklung sind arbeits- und zeitaufwendig, wenn eine komplette Abdeckung des Donorgenoms durch kurze, gleichmäßig verteilte Chromosomensegmente angestrebt wird, da dann separate Rückkreuzungsprogramme für jedes der Zielsegmente erforderlich sind. Trotz des hohen Aufwands bei der Erstellung wird häufig nur eine unvollständige Abdeckung des Donorgenoms erreicht (SCHMALENBACH et al. 2008, FALKE et al. 2008).

Ein möglicher Ansatz zur Kosteneinsparung ist, die Anzahl der simultan durchgeführten Rückkreuzungsprogramme zu reduzieren, indem während der Rückkreuzung zunächst Individuen vorselektiert werden, die jeweils ein komplettes Donorchromosom tragen (FALKE et al. 2009). Wird dieses Konzept angepasst auf Rückkreuzungsprogramme mit begrenzten Ressourcen, lassen sich Introgressionspopulationen entwickeln, die aber noch einige zusätzliche Donorsegmente im genetischen Hintergrund tragen. Solche Introgressionspopulationen können jedoch mit weniger Individuen und Markeranalysen entwickelt werden. Für diese Introgressionspopulationen sollte als Mindeststandard möglichst komplette Abdeckung des Donorgenoms und gleichmäßig Verteilung der Zielsegmente angestrebt werden.

Die Entwicklung von Linien durch *in vivo* Induktion doppelt haploider (DH) Linien wird routinemäßig in Maiszüchtungsprogrammen angewandt. Dennoch gibt es bislang keine Richtlinien, wie sich diese Technologie bei der Entwicklung von Introgressionsbibliotheken effizient einsetzen lässt, und was im Unterschied zur immer noch kostengünstigeren Selbstung zu beachten ist. Die Ziele unserer Simulationsstudie waren daher die Entwicklung geeigneter Kreuzungs- und Selektionsschemata für die Entwicklung von Introgressionspopulationen mit begrenztem Kontingent an Rückkreuzungsindividuen und Hochdurchsatzchips, und die Erstellung von Richtlinien für das optimale Versuchsdesign für DH- und S_2 -Kreuzungsschemata.

¹ Institut für Pflanzenbau und Pflanzenzüchtung II, Justus-Liebig-Universität GIESSEN, Deutschland
* Ansprechpartner: Eva HERZOG, eva.herzog@uni-giessen.de



Material und Methoden

Genetisches Modell

Computersimulationen wurden durchgeführt für ein genetisches Maismodell mit 10 Chromosomen von 200 cM Länge. Molekulare Marker für die Selektion waren gleichmäßig mit einer Dichte von einem Marker alle 1 cM im Genom verteilt. Es wurde angenommen, dass die Genotypisierung mit Hochdurchsatzchips durchgeführt wird und alle Marker im Genom in einem Analyseschritt genotypisiert werden können.

Kreuzungsschemata

Es wurden BC₃DH- und BC₃S₂-Kreuzungsschemata untersucht. In den Generationen BC₁-BC₃ erfolgte die Selektion einer festen Anzahl bester Individuen anhand eines Selektionsindex. Diese Individuen wurden, je nach Generation, rückgekreuzt oder zur Linienentwicklung mittels Selbstung oder *in vivo* DH-Induktion verwendet. Die resultierenden Introgressionspopulationen bestanden jeweils aus 100 NILs.

Selektionsindizes

Die NILs sollten je ein Zielsegment des Donorgenoms von 20 cM Länge tragen. Um den Selektionsindex i für ein Individuum in Bezug auf ein bestimmtes Zielsegment zu bestimmen, wurde mit t_c der Anteil Donorgenom in Prozent auf dem Chromosom, auf dem das Zielsegment sich befindet, bezeichnet. Mit t_h wurde der Anteil Donorgenom in Prozent auf der Chromosomenhälfte, auf der das Zielsegment sich befindet, bezeichnet. Mit t_s wurde der Anteil Donorgenom in Prozent im Zielsegment selbst bezeichnet. Die Werte für den genetischen Hintergrund b_c, b_h, b_s sind komplementär zu t_c, t_h und t_s und bezeichnen den Anteil an Rezipientengenom außerhalb der jeweiligen Selektionsregion. Bei der Selektion auf ganze Chromosomen wurde für jedes Chromosom $c = 1, 2, \dots, 10$ eine feste Anzahl bester Individuen für den Selektionsindex $i = t_c + b_c$ selektiert. Für Selektion auf Chromosomenhälften $h = 1, 2, \dots, 20$ oder Zielsegmente

$s = 1, 2, \dots, 100$ wurde anhand der Indizes $i = t_h + b_h$ und $i = t_s + b_s$ analog verfahren.

Selektionsstrategien

Selektion auf ganze Chromosomen (C), auf Chromosomenhälften (H) und auf Zielsegmente (S) wurden in den Generationen BC₁-BC₃ kombiniert, um verschiedene Selektionsstrategien zu bilden. Die untersuchten Kombinationen von Kreuzungsschema und Selektionsstrategie sind in der ersten Spalte von *Tabelle 1* aufgelistet. Für alle Selektionsstrategien wurden die besten 100 NILs für den Selektionsindex $i = t_s + b_s$ in Generation DH oder S₂ selektiert.

Simulationsreihen

In der ersten Simulationsreihe wurde Selektion nur in den Generationen BC₁ und BC₂ durchgeführt. In der zweiten Simulationsreihe wurde zusätzlich Selektion in Generation BC₃ durchgeführt. Die Gesamtpopulationsgröße pro Generation in der ersten und zweiten Simulationsreihe war konstant und betrug $n_{\text{tot}} = 360$ Individuen. In der dritten Simulationsreihe wurden optimierte Kreuzungsschemata mit Selektion in allen drei Rückkreuzungsgenerationen und ansteigenden Populationsgrößen untersucht.

Ergebnisse und Diskussion

Mit fast allen untersuchten Kombinationen aus Kreuzungsschema und Selektionsstrategie wurde eine Abdeckung des Donorgenoms $\geq 99\%$ erzielt (*Tabelle 1*). Diese Werte bezogen sich allerdings nicht nur auf die Zielsegmente, sondern beinhalteten auch zusätzliche unerwünschte Donorsegmente. Die Abdeckung des Donorgenoms war daher nicht ausreichend, um die Qualität einer Introgressionspopulation zu beschreiben. Erhebliche Unterschiede zwischen den Introgressionspopulationen wurden deutlich, wenn der Gesamtdonorgenomanteil, der Donorgenomanteil auf dem Chromosom, welches das Zielsegment trägt, und der durchschnittliche Donorgenomanteil in den Zielsegmenten betrachtet wurden.

Tabelle 1: Donorgenomabdeckung (Cov), Gesamtdonorgenomanteil (DGP_{tot}), Donorgenomanteil auf den Trägerchromosomen (DGP_{CC}), Donorgenomanteil in den Zielsegmenten (DGP_{TS}) und benötigte Anzahl von Hochdurchsatzchips (HT) für verschiedenen Strategien zur Entwicklung von Introgressionspopulationen bei Mais (Mittelwerte über 1000 Replikationen der Simulation)

Table 1: Coverage of the donor genome (Cov), total donor genome proportion (DGP_{tot}), donor genome proportion of the carrier chromosomes (DGP_{CC}), donor genome proportion of the target segments (DGP_{TS}) and required number of high-throughput chips (HT) for different strategies for developing introgression populations in maize (arithmetic means over 1000 replications of the simulation)

Simulationsreihe	Strategie	Cov	DGP _{tot} (%)	DGP _{CC} (%)	DGP _{TS} (%)	HT
1	BC ₃ DH-CC	99.2	5.0	35.2	97.8	1080
	BC ₃ DH-HH	99.8	5.1	33.3	94.2	1080
	BC ₃ DH-CH	99.6	5.1	35.1	93.7	1080
2	BC ₃ DH-CCC	99.1	4.8	35.8	98.0	1800
	BC ₃ DH-HHH	99.9	4.7	33.9	98.9	1800
3	BC ₃ DH-HHH*	99.9	5.0	33.3	98.4	1440
1	BC ₃ S ₂ -CC	99.3	5.7	39.2	97.0	1080
	BC ₃ S ₂ -HH	99.9	5.3	33.3	90.3	1080
	BC ₃ S ₂ -CH	99.7	5.3	35.0	89.9	1080
2	BC ₃ S ₂ -CCC	98.3	5.7	44.6	97.5	1440
	BC ₃ S ₂ -HHH	99.7	5.0	38.1	98.6	1440
3	BC ₃ S ₂ -HHS*	99.8	4.3	30.4	96.2	1400

Ein hoher Donorgenomanteil auf den Trägerchromosomen deutet auf unerwünschte Donorgenomsegmente hin, die nicht von den Zielsegmenten getrennt werden konnten. Solche großen Donorsegmente in Nachbarschaft zu den Zielsegmenten können die Detektion von QTL beeinträchtigen. Darüber hinaus bergen große Donorgenomsegmente das Risiko, dass unerwünschte Eigenschaften des Donors übertragen werden, und erfordern oft zusätzliche Rückkreuzungsschritte zur Abtrennung der Zielsegmente. Der Donorgenomanteil auf den Trägerchromosomen sollte daher niedrig sein. Innerhalb der Zielsegmente sollte der Donorgenomgehalt hingegen so hoch wie möglich sein, um die gesamte Bandbreite an genetischer Variation im Donor zu erfassen.

Die Selektionsstrategie ist der wichtigste Faktor für die Verteilung und Länge der Donorsegmente in Introgressionspopulationen. Selektionsstrategien, die zunächst Individuen vorselektieren, die ganze Donorchromosomen tragen, reduzieren zwar die Anzahl der Rückkreuzungsprogramme im Vergleich zu Selektionsstrategien, die direkt auf Zielsegmente selektieren. Bei langen Chromosomen von 200 cM Länge oder mehr bergen sie jedoch das Risiko, dass große Teile des Donorgenoms auf dem Zielchromosom bis zur Linienentwicklung konserviert und dann fixiert werden. Zusätzlich zu Selektion auf ganze Chromosomen mit den Selektionsstrategien CC und CCC wurde daher mit den Selektionsstrategien HH und HHH auch Selektion auf Chromosomenhälften untersucht. Damit wurde eine Reduktion des Donorgenomanteils auf den Trägerchromosomen von bis zu 6,5% erreicht, während die Maßzahlen für den genetischen Hintergrund in etwa gleich blieben. Selektion auf Chromosomenhälften in den Rückkreuzungsgenerationen ist daher von Vorteil für Kulturarten mit langen Chromosomen, wie z.B. Mais, Weizen oder Raps, da dadurch das Risiko der Übertragung unerwünschter Eigenschaften vom Donor verringert wird.

Wenn keine Selektion in Generation BC_3 durchgeführt wurde, führte Selektion auf Chromosomenhälften jedoch zu einer beträchtlichen Reduktion des Donorgenomanteils in den Zielsegmenten von 3,6-7,1%, was auf häufigeren Verlust der Zielsegmente hindeutet. Dies ist vermutlich den sehr kleinen Populationsgrößen von $n = 18$ Individuen zuzuschreiben, die mit der Unterteilung der konstanten Gesamtpopulationsgröße von $n_{tot} = 360$ Individuen in 20 Subpopulationen für die Chromosomenhälften einhergeht.

Der gleiche Verlust von Zielsegmenten wurde auch für die kombinierte Selektionsstrategie CH beobachtet, bei der in Generation BC_1 auf ganze Donorchromosomen und in Generation BC_2 auf Chromosomenhälften selektiert wurde. Zusätzlich wurde bei dieser Selektionsstrategie ein ähnlich hoher Wert für den Donorgenomanteil auf dem Trägerchromosom beobachtet wie bei ausschließlicher Selektion auf ganze Chromosomen. Dies ist vermutlich auf die effiziente Selektion auf ganze Chromosomen aus der verhältnismäßig großen BC_1 Population von $n_{tot} = 360$ Individuen zurückzuführen. Die selektierten ganzen Donorchromosomen blieben zum großen Teil bis zur Linienentwicklung erhalten. Die Selektionsstrategie CH vereint daher die Nachteile von Selektion auf ganze Donorchromosomen mit den Nachteilen von Selektion auf Donorchromosomenhälften und ist

nicht empfehlenswert für Rückkreuzungsprogramme mit kleinen, konstanten Gesamtpopulationsgrößen. Für kleine Rückkreuzungsprogramme mit Gesamtpopulationsgrößen von $n_{tot} = 360$ Individuen ohne Selektion in Generation BC_3 ist daher Selektion auf ganze Donorchromosomen in den Generationen BC_1 und BC_2 die günstigste Lösung.

Für effektive Reduktion des Donorgenomanteils auf den Trägerchromosomen durch Selektion auf Chromosomenhälften ohne Verlust der Zielsegmente muss die Anzahl der Träger von Zielsegmenten für die Linienentwicklung erhöht werden. Ein möglicher Lösungsansatz für dieses Problem wäre ein zusätzlicher Selektionsschritt in Generation BC_3 . Während dieser für die S_2 -Kreuzungsschemata ohne Probleme möglich ist, da durch Selbstung viele Linien aus einem Rückkreuzungsindividuum erzeugt werden können, kann mit *in vivo* Induktion von DH Linien oft nur eine Linie aus einem Rückkreuzungsindividuum gewonnen werden. In den Kreuzungsschemata ohne Selektion in Generation BC_3 war ein direkter Vergleich der Ergebnisse zwischen DH- und S_2 -Kreuzungsschemata bei gleichem Einsatz von Individuen und Hochdurchsatzchips möglich. In diesem Fall waren die DH-Kreuzungsschemata den S_2 -Kreuzungsschemata überlegen, was vor allem auf die komplette Homozygotie der DH Linien zurückzuführen ist.

Für Selektion in Generation BC_3 wurde in den DH-Kreuzungsschemata eine Verdopplung der Populationsgröße vorgenommen, und nur die bessere Hälfte der Individuen wurde für die DH-Induktion selektiert. In den S_2 -Kreuzungsschemata wurde die konstante Populationsgröße beibehalten und nur ein bestes Individuum für den jeweiligen Selektionsindex selektiert. Dieser Unterschied zwischen DH- und S_2 -Kreuzungsschemata zeigt sich nicht nur in der Anzahl der benötigten Hochdurchsatzchips, sondern führt auch zu einem Unterschied in der Selektionsintensität. Es ist daher zu erwarten, dass unterschiedliche Selektionsstrategien für diese Formen der Linienentwicklung optimal sind.

Durch den zusätzlichen Selektionsschritt in Generation BC_3 wurden die Maßzahlen für den genetischen Hintergrund kaum verändert, der Donorgenomanteil in den Zielsegmenten wurde jedoch verbessert. Dies war besonders deutlich für Selektion auf Chromosomenhälften mit Selektionsstrategie HHH, bei der eine Verbesserung von bis zu 8,6% beobachtet wurde. Insgesamt lieferte Selektion auf Chromosomenhälften jetzt die besseren Ergebnisse. Für das beste DH-Kreuzungsschema BC_3 DH-HHH fiel die Verbesserung nur sehr gering aus im Vergleich zu Schema BC_3 DH-CC ohne Selektion in Generation BC_3 . Dafür mussten jedoch 720 zusätzliche Hochdurchsatzchips aufgewendet werden. In den S_2 -Kreuzungsschemata ging der zusätzliche Selektionsschritt in Generation BC_3 mit einem stark erhöhten Donorgenomanteil auf den Trägerchromosomen von 38-48% einher. Dies deutet auf Fixierung der Selektionsregionen der letzten Rückkreuzungsgeneration hin. Selektion in Generation BC_3 hat daher das Potential, die Abdeckung der Zielsegmente zu verbessern. Für DH-Kreuzungsschemata ist sie jedoch nur effizient, wenn die Anzahl der benötigten Hochdurchsatzchips reduziert werden kann. Für die S_2 -Kreuzungsschemata muss die Fixierung großer Donorchromosomensegmente vermieden werden.

Bei konstanten Gesamtpopulationsgrößen von $n_{tot} = 360$ Individuen pro Generation ist die Populationsgröße in Generation BC_1 , in der noch keine Unterteilung in Subpopulationen vorgenommen wird, relativ groß im Vergleich zum Selektionsgewinn, der durch Selektion einer relativ kleinen Fraktion von 10 oder 20 Individuen erzielt werden kann. Es liegt daher nahe, die Populationsgröße in dieser Generation zu reduzieren und die freiwerdenden Ressourcen fortgeschrittenen Generationen zuzuschlagen. Für die DH-Kreuzungsschemata werden so größere Populationsgrößen in Generation BC_3 möglich, ohne dass insgesamt mehr Ressourcen aufgewendet werden müssen als für die S_2 -Kreuzungsschemata. Für die S_2 -Kreuzungsschemata ermöglichen größere Populationsgrößen die Selektion auf Zielsegmente schon in Generation BC_3 .

Das Kreuzungsschema $BC_3DH-HHH^*$ mit ansteigenden Populationsgrößen führte zu ähnlichen Ergebnissen wie das bislang beste Schema $BC_3DH-HHH$, benötigte aber 360 Individuen und Hochdurchsatzchips weniger. Im Vergleich zum billigeren Schema BC_3DH-CC wurde der Donorgenomanteil auf den Trägerchromosomen verringert und der Donorgenomanteil in den Zielsegmenten erhöht. Die Investition in die zusätzlichen Hochdurchsatzchips scheint daher lohnenswert. Das Kreuzungsschema $BC_3S_2-HHS^*$ mit ansteigenden Populationsgrößen resultierte in der gewünschten Reduktion des Donorgenomanteils auf den Trägerchromosomen von 38% auf 30% im Vergleich zum bislang besten Schema BC_3S_2-HHH , obwohl 40 Hochdurchsatzchips weniger verwendet wurden. Der

Donorgenomanteil in den Zielsegmenten wurde zwar um 2,4% reduziert, dafür wurde allerdings das Verhältnis von Donorgenomanteil auf den Trägerchromosomen zu Donorgenomanteil in den Zielsegmenten erheblich verbessert. Für S_2 -Kreuzungsschemata ist Selektion auf Zielsegmente in der letzten Rückkreuzungsgeneration daher zu empfehlen. Im Vergleich zum insgesamt besten aber auch teuersten DH-Kreuzungsschema $BC_3DH-HHH$, erreichte das Kreuzungsschema $BC_3S_2-HHS^*$ nur leicht schlechtere Werte, brauchte jedoch 400 Hochdurchsatzanalysen weniger. S_2 -Kreuzungsschemata bieten daher ökonomische Alternativen zu DH-Kreuzungsschemata.

Literatur

- ESHED Y, ZAMIR D, 1994: A genomic library of *Lycopersicon pennellii* in *L. esculentum*: a tool for fine mapping of genes. *Euphytica* 79, 175-179.
- FALKE KC, SUŠIĆ Z, HACKAUF B, KORZUN V, SCHONDELMAIER J, WILDE P, WEHLING P, WORTMANN H, MANK R, VAN DER VOORT JR, MAURER HP, MIEDANER T, GEIGER HH, 2008: Establishment of introgression libraries in hybrid rye (*Secale cereale* L.) from an Iranian primitive accession as a new tool for rye breeding and genomics. *Theor Appl Genet* 117, 641-652.
- FALKE KC, MIEDANER T, FRISCH M, 2009: Selection strategies for the development of rye introgression libraries. *Theor Appl Genet* 119, 595-603.
- SCHMALENBACH I, KÖRBER N, PILLEN K, 2008: Selecting a set of wild barley introgression lines and verification of QTL effects for resistance to powdery mildew and leaf rust. *Theor Appl Genet* 117, 1093-1106.