

Molekulare Analyse des Kolumnar-Gens beim Apfel (*Malus x domestica*)

Dissertation

zur Erlangung des Grades
Doktor der Naturwissenschaften
(Dr. rer. nat.)



am Fachbereich Biologie der
Johannes Gutenberg-Universität in Mainz
Institut für Molekulargenetik,
Gentechnische Sicherheitsforschung und -beratung

vorgelegt von

Dominik Otto

Geboren am 06. Januar 1983

in Bad Hersfeld

Mainz, 2013

Dekan:

1. Berichtstatter:

2. Berichtstatter:

Tag der mündlichen Prüfung: 24.01.2014

INHALTSVERZEICHNIS

INHALT

1	Einleitung	1
1.1	Der Kulturapfel, <i>Malus x domestica</i>	1
1.2	Kolumnarwachstum beim Apfel	3
1.3	Pflanzengenome und transposable Elemente.....	6
1.4	Positionelle Klonierung.....	12
1.5	Molekulare Marker	14
1.6	Zielsetzung	19
2	Material und Methode.....	21
2.1	Allgemeine Methoden zum Arbeiten mit DNA.....	21
2.1.1	Elektrophoretische Auftrennung von DNA	21
2.1.2	Wiedergewinnung von DNA-Fragmenten aus Agarosegelen.....	22
2.1.3	Phenol-Chloroform-Extraktion	23
2.1.4	Fällung von DNA.....	23
2.1.5	Polymerase-Ketten-Reaktion (PCR)	24
2.1.6	Aufreinigung von PCR-Produkten	25
2.1.7	Klonierung von PCR-Produkten	25
2.1.8	Southern Blot.....	25
2.1.9	„Pirota“ Blot.....	27
2.2	Konstruktion einer Apfel-BAC-Bibliothek	27
2.2.1	Versuchsmaterial.....	27
2.2.2	Isolation von Zellkernen	27
2.2.3	Kern-Lyse und Proteinase K-Verdau	28
2.2.4	Restriktion	28
2.2.5	Größenfraktionierung.....	29
2.2.6	Test zur Überprüfung der Integrität-Enden	30
2.2.7	Ligation.....	31
2.2.8	Transformation mittels Elektroporation.....	31
2.2.9	Herstellung elektrokompenter Zellen.....	32
2.2.10	BAC-Minipräparation	32
2.3	Screening einer Apfel-BAC-Bank.....	33
2.3.1	Konstruktion PCR-basierender Sonden.....	33
2.3.2	Koloniefilter-Hybridisierung	34

2.3.2.4	Randsequenzierung	35
2.4	Sequenzierung relevanter Klone mittels shotgun-Bibliotheken	35
2.3.1	Plasmid-Isolierung nach Whitehead	36
2.3.2	Nebulisieren der Plasmid-DNA und Endfilling.....	36
2.3.3	Größenfraktionierung und Elektroelution.....	37
2.3.4	Ligation und Transformation	37
2.3.5	Überprüfung der shotgun-Bibliothek und Sequenzierung	37
2.3.6	Auswertung von DNA-Sequenzen	38
2.5	Sequenzierung von BAC-Klonen mittels Illumina Next-Generation Sequencing	38
2.5.1	Plasmid-Isolierung nach Promega	39
2.5.2	Auswertung der Illumina Next-Generation Sequencing Sequenzdaten.....	39
2.5.3	Überprüfung der Kandidaten der BAC-Klone	40
2.6	Illumina Next-Generation Sequencing genomischer Apfel-DNA	41
2.6.1	Erstellung einer „mate pair“-Bibliothek.....	41
2.6.2	Sequenzierung und Filterung der Rohdaten	43
2.7	Erstellung eines Gesamtcontigs der Co-Zielregion	44
2.7.1	Erstellung von Metacontigs	44
2.7.2	Konstruktion eines Gesamtcontigs	44
2.8	Annotation von Genen und transposablen Elementen	44
2.8.1	Annotation von Genen.....	44
2.8.2	Annotation von transposablen Elementen	45
2.9	Herstellung Co-Gen gekoppelter Marker.....	45
2.9.1	Versuchsmaterial	45
2.9.2	Primerdesign für PCR-basierende Marker	46
2.9.3	Testen von Kandidaten für Co-Gen gekoppelter Marker.....	46
2.9.4	Klonierung Co-Gen gekoppelter Marker und Sequenzierung	46
2.10	Standardlösungen und Materialien	47
3	Ergebnisse.....	49
3.1	Festlegung der Co-Zielregion	49
3.2	Erstellung von Apfel-BAC-Bibliotheken.....	50
3.3	Screening der Apfel-BAC-Bibliotheken	55
3.4	BAC-Klone aus der Co-Zielregion und Illumina Next-Generation Sequencing.....	58
3.4.1	Sequenzierung und Assemblierung der BAC-Klone.....	58
3.4.2	Genomische Illumina-Sequenzdaten der Apfelsorte P28.....	60
3.4.3	Analyse chimärer BAC-Klone.....	62
3.5	Analyse der genomischen Co-Gen-Region	66

3.5.1 Konstruktion von Metacontigs und eines Gesamtcontigs	66
3.5.2 Sequenzvergleiche der Apfelsorten P28 und GD	69
3.5.3 Annotation von Genen und transposablen Elementen	75
3.7 Molekulare Co-Gen gekoppelte Marker	85
3.7.1 Identifikation Co-Gen gekoppelter Marker	85
3.7.2 Feinkartierung der Zielregion	88
3.7.3 Markeranalysen anderer Apfelsorten	91
3.7.4 Transposon-basierende Marker	94
4 Diskussion	97
4.1 Bestimmung der Co-Zielregion	97
4.2 Genomische Apfel-BAC-Bibliotheken	99
4.3 Erstellung einer Referenzsequenz der Co-Zielregion	101
4.4 Analyse der Co-Zielregion.....	107
4.5 Identifizierung der Co-Mutation.....	109
5 Zusammenfassung	117
Abkürzungsverzeichnis.....	118
Abbildungsverzeichnis	121
Tabellenverzeichnis.....	122
Webverzeichnis.....	123
Literaturverzeichnis	124
Anhang	137
A1 Vektorkarten	137
A2 Sonden für das Screening von genomischen Apfel-BAC-Bibliotheken.....	139
A3 Markeranalysen mittels QIAxcel	141
A4 Markertests diverser Apfelsorten mittels selbst erstellter Marker	142
A5 Inhaltsverzeichnis des elektronischen Anhangs.....	143
Curriculum Vitae	144
Danksagung.....	146
Eidesstattliche Erklärung.....	147

1 Einleitung

1.1 Der Kulturapfel, *Malus x domestica*

Im Jahr 2011 waren Äpfel hinter Bananen (107 Mt) und Wassermelonen (102 Mt) mit einer Jahresproduktion von über 75 Mio. Tonnen die am häufigsten angebaute Obstart weltweit (FAO, 2013). Nach dem Pro-Kopf-Verbrauch zu urteilen, waren im Jahr 2011 Apfelsaft und Apfelsaftschorle mit zusammen 17,4 Litern in Deutschland die beliebtesten Fruchtsaftgetränke (Vdf, 2013) und Äpfel mit 19,5 kg pro Privathaushalt das meist geschätzte Obst (AMI GmbH, 2013). Der Apfel verdankt seine Beliebtheit vor allem der guten Lagerfähigkeit und seinem enormen Spektrum an Zubereitungsmöglichkeiten (Luby 2003). Darüber hinaus gelten Äpfel als gesundheitsfördernd und vorbeugend gegen Herz-Kreislauf-Erkrankungen, Asthma, Diabetes und Krebs (Boyer und Liu 2004; Gerhauser 2008). Äpfel werden in 93 Ländern auf einer Fläche von über 4,7 Mio. Hektar angebaut. Der Hauptproduzent im Jahr 2011 war China mit 36 Mio. Tonnen, gefolgt von den USA und Indien mit 4,3 und 2,9 Mio. Tonnen. Deutschland lag auf Platz 14 mit fast 900.000 Tonnen (FAO, 2013).

Der Apfel gehört systematisch betrachtet in der Familie der Rosengewächse (Rosaceae) zur Unterfamilie der Spiraeoideae. Innerhalb dieser Unterfamilie zählt der Apfel neben Birne, Quitte und Mispel zum Tribus Pyreae, genauer zum Untertribus Pyrinae (Kernobstgewächse). Die Kernobstgewächse wurden lange Zeit als eigene Unterfamilie (Maloideae) der Rosengewächse geführt; neuere molekulargenetische Untersuchungen zeigen jedoch eine Zugehörigkeit zu der Unterfamilie Spiraeoideae (Potter *et al.* 2007). Die Gattung Äpfel (*Malus*) umfasst zwischen 25 und 47 Arten (Robinson *et al.* 2001), die weltweit bekannteste und wirtschaftlich bedeutendste Art ist der Kulturapfel (*Malus x domestica*). Die Anzahl der bekannten Apfelsorten wird weltweit auf über 20.000 geschätzt. Der haploide Chromosomensatz (x) der meisten Rosengewächse beträgt 7, 8 oder 9. Im Gegensatz dazu haben die Pyreae einen haploiden Chromosomensatz von 17 (Chevreau *et al.* 1985). Es existieren zwei Hypothesen zur Entstehung dieser außergewöhnlich hohen Chromosomenzahl in Pyreae, zum einen durch Allopolyploidisierung zwischen spiraeoiden ($x=9$) und amygdaloiden ($x=8$) Vorfahren (Phipps *et al.* 1991), zum anderen durch Autopolyploidisierung eines Vorfahren mit $x=9$ wie etwa *Gillenia* durch Diploidisierung und Aneuploidisierung (Evans und Campbell 2002). Neuere Ergebnisse unterstützen die Hypothese der Entstehung der Pyreae durch Autopolyploidisierung. Maliepaard *et al.*

(1998) fanden anhand von molekularen Markern bereits große Homologien zwischen unterschiedlichen Chromosomen, beispielsweise zwischen „Linkage Group“ (LG) 5 und LG 10. Velasco *et al.* (2010) bestätigten dies durch paarweise Vergleiche der Apfelchromosomen und identifizierten große homologe Bereiche zwischen den Chromosomen 3 und 11, 5 und 10, 9 und 17 und 13 und 16, sowie weitere kleinere Homologien, was insgesamt für eine genomweite Duplikation (GWD) infolge einer Autopolyploidisierung spricht. Die molekulare Phylogenie der *Wx*-Gene („waxy genes“) belegt eine enge Verwandtschaft zwischen *Gillenia* und *Pyraea*, ein weiteres Indiz für eine Autopolyploidisierung (Velasco *et al.* 2010).

Der Kulturapfel entstand vermutlich aus interspezifischen Hybridisierungen. Daher ist der Name *Malus x domestica* wissenschaftlich anerkannt (Korban und Skirvin 1984). Taxonomisch betrachtet sind die am nächsten verwandten Arten zu *Malus x domestica* die asiatischen Arten *Malus x asiatica*, *Malus baccata*, *Malus micromalus*, *Malus orientalis*, *Malus prunifolia* und *Malus sieversii*, sowie die europäische Spezies *Malus sylvestris*. Diese Arten sind vermutlich mehr oder weniger an der Entstehung des Genpools des Kulturapfels beteiligt (Robinson *et al.* 2001). *Malus sieversii* ist jedoch die einzige wild vorkommende Art, die alle Qualitäten des Kulturapfels in Bezug auf Frucht- und Baummorphologie aufweist (Forsline *et al.* 2002). Juniper und Mabberley (2006) postulierten, dass es sich daher bei *Malus sieversii* und *Malus x domestica* um die gleiche Spezies handele und führten die einheitliche Bezeichnung *Malus pumila* ein. Die Sequenzierung des Genoms eines Kulturapfels (Golden Delicious, GD) belegte nicht nur den hohen Verwandtschaftsgrad zwischen *Malus x domestica* und *Malus sieversii*, sondern bestätigte darüber hinaus, dass es sich um dieselbe Art handelt (Velasco *et al.* 2010). Obwohl *Malus pumila* die anerkannte Bezeichnung ist, wird in dieser Arbeit die Bezeichnung *Malus x domestica* für den Kulturapfel verwendet. Neuste molekulare Analysen belegen, dass der europäische Wildapfel *Malus sylvestris* nach *Malus sieversii* den größten Anteil am Genpool des Kulturapfels ausmacht, was dazu führt, dass einige Sorten des Kulturapfels mit *Malus sylvestris* näher verwandt sind als mit *Malus sieversii* (Cornille *et al.* 2012). *Malus sieversii* stammt ursprünglich aus dem Gebiet um das Tian Shan-Gebirge, das sich über das heutige Staatsgebiet von China, Kasachstan, Kirgisistan, Usbekistan und Tadschikistan erstreckt (Morgan und Richards 1993). Um Almaty, früher Alma-Ata (übersetzt: „Großvater des Apfels“), liegt das Zentrum der größten Diversität und der Ursprung des Kulturapfels (Vavilov 1926; Janick *et al.* 1996). Von dort wurde der Kulturapfel vermutlich durch Römer

und Griechen entlang der Seidenstraße nach Europa und Nordafrika gebracht, ehe er sich weltweit verbreitete (Juniper und Mabberley 2006). Archäologische Funde über das Sammeln von Wildäpfeln in Europa gehen zurück in das Neolithikum (vor 11.200 Jahren) und die Bronzezeit (vor ca. 4.500 Jahren) (Hopf 1973; Schweingruber und Hopf 1979). Vor rund 4.000 Jahren traten im Nahen Osten bereits Formen des Apfels auf, die mit dem heutigen Kulturapfel vergleichbar sind (Zohary und Hopf 2000).

Der Ursprung der kontrollierten Züchtung wird Thomas Andrew Knight im Jahr 1806 zugeschrieben, der den ersten Kultivar mit bekannten Eltern züchtete (Janick *et al.* 1996). Der hohe Grad an Selbstinkompatibilität und der stark heterozygote Charakter des Genoms sowie eine lange juvenile Phase von sechs bis acht Jahren erschweren die konventionelle Züchtung des Kulturapfels seit jeher (Korban und Chen 1992; Hemmat *et al.* 1994). Abhilfe schaffen könnte jedoch die Hochgeschwindigkeitszüchtung, eine Errungenschaft neuester Forschungen, bei der die juvenile Phase stark verkürzt ist. Dies gelingt durch die Erstellung transgener Apfelinien mit früher Blüte durch Überexpression des Gens *BpMADS4* aus der Birke (Flachowsky *et al.* 2011). Darüber hinaus können durch molekularen Gentransfer, etwa *Agrobacterium tumefaciens*-vermittelt, gezielt Gene in das Genom der gewünschten Apfelsorte integriert werden (Malnoy *et al.* 2007; Flachowsky *et al.* 2010).

1.2 Kolumnarwachstum beim Apfel

Das Kolumnarwachstum geht auf eine Zufallsmutante der Apfelsorte McIntosh zurück, welche erstmals 1969 von Fisher beschrieben wurde. Sie wurde in den frühen 1960er Jahren in einer Obstanlage im Okanagan Valley (British Columbia, Kanada) beobachtet und nach ihrem Entdecker, dem aus Polen stammenden Obstbauern Wijcik, als „McIntosh Wijcik“ benannt (Fisher 1970). Die Mutation trat in einem einzelnen Trieb an der Spitze eines 50 Jahre alten normalwüchsigen Apfelbaums der Sorte McIntosh auf (Looney und Lane 1984). Diese als Knospen- oder Sprossmutationen bezeichneten vegetativen Veränderungen („Sports“) sind von großer züchterischer Bedeutung und werden seit jeher zur Etablierung neuer Sorten genutzt (Schmidt 1937). Der kolumnare Phänotyp stellt eine extreme Form des Kurztriebwuchses („spur type growth“) dar (Blažek 1992), charakterisiert durch verkürzte Internodien, eine sehr kompakte Wuchsform, eine geringe Anzahl an Seitentrieben bis zum vollständigen Fehlen dieser und eine damit verbundene erhöhte Anzahl an Fruchtspießen (Lapins 1969). In Abbildung 1.1 sind Fotos solcher kolumnaren Apfelbäume abgebildet. In seltenen Fällen treten längere Seitentriebe auf, welche in der

Regel parallel zum Hauptspross wachsen (Hemmat *et al.* 1997). Der Ausprägungsgrad des Kolumnarwachstums hängt von zahlreichen Faktoren ab, etwa dem Alter der Pflanzen und Umwelteinflüssen wie beispielsweise dem Klima und der Unterlage (Kenis und Keulemans 2007).



Abbildung 1.1 Kolumnare Apfelbäume

Die Fotos zeigen kolumnare Apfelbäume zur Zeit des Fruchtstandes. Die Bilder wurden 2003 in der Forschungsanstalt Geisenheim aufgenommen. Die Seitentriebe sind in Fruchtspitze umgewandelt, sodass die Früchte direkt am Stamm wachsen (A). Der sehr kompakte Wuchscharakter der Bäume ermöglicht eine sogenannte Doppelreihpflanzung, wodurch der Ertrag erheblich gesteigert werden kann (B).

Kolumnare Apfelbäume sind von großem wirtschaftlichem Interesse. Die kompakte Wuchsform ermöglicht ein sehr dichtes Pflanzen der Bäume (Abbildung 1.1 B), sodass der Ertrag von durchschnittlich 30 bis 60 t/ha bei normalwüchsigen Bäumen auf über 100 t/ha gesteigert werden kann (Jacob 2010). Weitere Aspekte sind die Kosten- und Zeitersparnisse durch ein Ausbleiben bzw. eine Minimierung des Beschneidens der Bäume und eine Ernte der Früchte mittels Vollerntemaschinen (Tobutt 1985). Darüber hinaus führt der Säulenwuchs zu einer gleichmäßigen Sonneneinstrahlung auf die Früchte, was eine homogene Entwicklung, Reifung und Färbung der Früchte begünstigt (Kenis und Keulemans 2007). Neben diesen positiven Aspekten weisen kolumnare Apfelbäume jedoch auch einige negative Eigenschaften auf. So ist die Sorte McIntosh Wijk stark anfällig gegen Apfelschorf, die Lagerfähigkeit der Früchte ist stark begrenzt und die Fruchtqualität kann nicht mit der weit verbreiteter Sorten wie etwa Golden Delicious oder Gala mithalten (Meulenbroek *et al.* 1998, P. Braun, pers. Kommunikation). Aus diesem Grund wurden weltweite Züchtungsprogramme initiiert, um das kolumnare Wachstum mit den positiven

Fruchteigenschaften der kommerziell erfolgreichen Apfelsorten und deren Resistenzen gegen Schorf und weitere Krankheiten zu kombinieren (Petersen und Krost 2013). Dies ist jedoch aufgrund der bereits genannten Schwierigkeiten bei der Züchtung sehr zeit- und kostenintensiv.

Der stark veränderte Wuchscharakter eines kolumnaren im Vergleich zu einem normalwüchsigen Apfelbaum legt die Vermutung nahe, dass veränderte Phytohormonkonzentrationen eine entscheidende Rolle bei der Ausprägung des kolumnaren Phänotyps spielen. In mehreren Studien konnte gezeigt werden, dass kolumnare Bäume im Vergleich zu normalwüchsigen Bäumen höhere Konzentrationen an freien Auxinen und Cytokininen in der Sprossspitze aufweisen (Looney und Lane 1984; Watanabe *et al.* 2004; 2008). Hingegen ist die Konzentration an Abscisinsäure und polaren Gibberellinen in kolumnaren Apfelsorten im Vergleich niedriger (Lee und Looney 1977; Looney und Lane 1984). Krost *et al.* (2012) konnten durch Transkriptomanalysen des Sprossapikalmeristems der kolumnaren Sorte Procats 28 (P28) im Vergleich zur nicht kolumnaren Sorte A14-190-93 (A14) diese Resultate größtenteils bestätigen. Gene für die Biosynthese und Signaltransduktion des Auxins IAA sind aktiv, gegenteiliges gilt für die Biosynthese und Signaltransduktion von Gibberellinen.

Kreuzungsanalysen zwischen kolumnaren und normalwüchsigen Eltern führten zu dem Schluss, dass ein dominantes Allel des Kolumnargens (*Co*-Gen) für die Ausprägung des Säulenwuchses verantwortlich ist (Lapins 1969; Lapins und Watkins 1973). Allerdings spalten sich die Nachkommen einer Kreuzung zwischen *Co/co* und *co/co* nicht wie man es für einen Mendelschen Erbgang erwarten würde im Verhältnis 1:1 (kolumnar:nicht kolumnar) auf. Meist liegt die Zahl der kolumnaren Nachkommen mit rund 44 % knapp unter den erwarteten 50 % (Lapins 1969). Aus diesem Grund geht man von einem oder mehreren untergeordneten Modifikatoren aus, die neben dem *Co*-Gen eine entscheidende Rolle bei Vererbung und Ausprägung des Kolumnarwachstums spielen (Lapins 1976). Bisher sind alle untersuchten kolumnaren Apfelsorten heterozygot für *Co* (Tian *et al.* 2005). Allerdings ist es der Universität Geisenheim (Deutschland) gelungen, eine homozygote kolumnare Apfelsorte zu züchten. Kreuzungen mit dieser Apfelsorte führen ausschließlich zu kolumnarwüchsigen Nachkommen (P. Braun, pers. Kommunikation).

Für die Züchtung ist es von erheblicher Bedeutung, möglichst früh detektieren zu können, ob ein Nachkomme kolumnar- oder normalwüchsig ist. Phänotypisch können erst nach zwei bis drei Jahren zuverlässige Aussagen über den Wuchscharakter getroffen werden (Blažek

1992; Baldi *et al.* 2012). Eine große Schwierigkeit stellen dabei die vielen existierenden Zwischenformen dar (Ikase und Dumbravs 2004; Moriya *et al.* 2009). Aus diesen Gründen ist die Verfügbarkeit zuverlässiger molekularer Marker für die Züchtung von erheblicher Bedeutung. Darüber hinaus ermöglichen molekulare Marker ein Lokalisieren des *Co*-Gens auf genomischer Ebene – ein essentieller Schritt, um die bisher unbekannte Natur der *Co*-Mutation zu untersuchen.

Neben dem Kolumnarwachstum im Apfel sind ähnliche Wuchsformen in der Sauerkirsche (Schuster 2009) und im Pfirsich (Yamazaki *et al.* 1987) beschrieben. Allerdings sind die molekularen Mechanismen weitgehend unerforscht. Während für die Sauerkirsche keine molekularen Hintergründe bekannt sind, wird das Kolumnarwachstum im Pfirsich möglicherweise durch ein rezessives Gen (*broomy*, *br*) verursacht. Da es sich beim Apfel jedoch um eine dominante Mutation handelt, sind vermutlich verschiedene Mechanismen für die Ausprägung eines ähnlichen Phänotyps verantwortlich (Sajer *et al.* 2012).

1.3 Pflanzengenome und transposable Elemente

Höhere Pflanzen besiedeln nahezu jedes Ökosystem der Erde und sind als Nahrungs-, Roh- und Sauerstofflieferant essentiell für das Leben von Mensch und Tier. Die Lebensformen reichen von sehr kleinen Pflanzen mit einer Länge von weniger als einem Millimeter wie die Zwergwasserlinse bis hin zu Mammutbäumen mit einer Höhe von über 100 Metern und einem Stammdurchmesser von mehr als 10 Metern (Paterson *et al.* 2010). Bedingt durch ihre sessile Lebensweise entwickelten Pflanzen zahlreiche Strategien zum Schutz vor Fressfeinden, Pathogenen und wechselnden Umweltbedingungen (Sterck *et al.* 2007). Möglicherweise ist dies einer der Gründe für die meist größeren Genome und komplexeren Genomstrukturen von Pflanzen im Vergleich zu denen der Tiere (Gregory 2005). Ein weiterer Grund liegt in mehreren genomweiten Duplikationen während der Evolution von Pflanzen (Proost *et al.* 2011). Zur Komplexität der Pflanzengenome tragen darüber hinaus starke Heterozygotien sowie Polyploidien bei (Hamilton und Robin Buell 2012).

Die Publikation des ersten sequenzierten Pflanzengenoms im Jahr 2000 der Pflanze *Arabidopsis thaliana* stellte einen wichtigen Schritt in der Analyse von Genomstrukturen in Pflanzen dar (TAG1 2000). Im Laufe der vergangenen Jahre folgten, begünstigt durch den Aufstieg des Next-Generation Sequencings, zahlreiche weitere Genomsequenzierungen von vor allem wirtschaftlich relevanten Kulturpflanzen (Hamilton und Robin Buell 2012), von denen eine Auswahl in Tabelle 1.1 dargestellt ist.

Tabelle 1.1 Eine Auswahl sequenzierter Pflanzengenome

Spezies	Genomgröße	Gene	TE ¹	Quelle
<i>Arabidopsis thaliana</i> (Acker-Schmalwand)	125 Mbp	27.228	18,5	TAGI (2000)
<i>Glycine max</i> (Sojabohne)	1,1 Gbp	46.430	50,3	Schmutz <i>et al.</i> (2010)
<i>Malus x domestica</i> (Apfel)	742,3 Mbp	57.386	42,4	Velasco <i>et al.</i> (2010)
<i>Oryza sativa</i> (Reis)	466 Mbp	40.577	39,5	Yu <i>et al.</i> (2002)
<i>Populus trichocarpa</i> (Balsam-Pappel)	485 Mbp	45.654	35,0	Tuskan <i>et al.</i> (2006)
<i>Sorghum bicolor</i> (Mohrenhirse)	730 Mbp	34.496	62,0	Paterson <i>et al.</i> (2009)
<i>Vitis vinifera</i> (Wein)	487 Mbp	33.514	21,5	Jaillon <i>et al.</i> (2007)
<i>Zea mays</i> (Mais)	2,3 Gbp	32.540	84,2	Schnable <i>et al.</i> (2009)

¹ Prozentualer Anteil transposabler Elemente (TE) am Gesamtgenom

Die Sequenzierung von Genomen beruht auf zwei verschiedenen Herangehensweisen (Green 2001). Bei der „clone by clone“-Strategie wird das Genom kloniert und ausgewählte Klone sequenziert. Im Gegensatz dazu steht die „whole genome shotgun“-Strategie, bei der das komplette Genom in kleine Fragmente zerlegt wird und diese sequenziert werden. Ein entscheidender Vorteil bei der Verwendung von klonierter genomischer DNA ist, dass ein Klon nur einen begrenzten chromosomalen Bereich enthält und die Komplexität im Vergleich zum Gesamtgenom stark reduziert ist (Imelfort und Edwards 2009). Neben der traditionellen Sanger-Sequenzierung ermöglicht die Verwendung von Next-Generation Sequencing, in Form von beispielsweise einer Roche 454 Pyro- oder einer Illumina-Sequenzierung, die Generierung von großen Datenmengen in kürzerer Zeit bei geringeren Kosten (Schatz *et al.* 2010). Dies hat dazu geführt, dass die Sequenzierung von Pflanzengenomen seit 2006 rasant zugenommen hat (Tabelle 1.1). Allerdings verursacht zum Beispiel die Illumina-Technologie Plattform, dass kürzere Sequenzen generiert werden, was wiederum erhebliche Probleme bei der Assemblierung der Daten erzeugt (Imelfort und Edwards 2009). Lediglich die Genome von *Arabidopsis thaliana* und Reis (Sanger-Sequenzierung in Kombination mit einer „clone by clone“-Strategie) sind tatsächlich weitgehend vollständig sequenziert, wohingegen es sich bei den anderen in Tabelle 1.1 aufgeführten Genomen um Entwürfe, sogenannte „draft genomes“, mit kleineren oder größeren Lücken handelt, die sich meist in repetitiven Elementen begründen (Claros *et al.* 2012).

Pflanzen enthalten, bedingt durch genomweite Duplikationen, eine große Zahl an Genen (Sterck *et al.* 2007). Die Anzahl der vorhergesagten Gene ist Tabelle 1.1 zu entnehmen. Der Apfel besitzt im Vergleich zu anderen Kulturpflanzen mit 57.386 Genen, darunter 11.444 Apfel-spezifische Gene, die höchste Anzahl (Velasco *et al.* 2010). Die Ursache hierfür ist noch unklar, allerdings zeigt der Vergleich mit anderen sequenzierten Genomen, dass die

Gruppen von Genen, die im Apfel stärker vertreten sind, in direktem Zusammenhang mit der Reaktion auf Stress und Anpassung an die Umwelt stehen, sowie beteiligt an der Bildung von Holz und der Fruchtentwicklung sind (Troggio *et al.* 2012).

In der Regel machen den größten Anteil pflanzlicher Genome aber nicht kodierende Bereiche aus, sondern repetitive Elemente, vor allem in Form von transposablen Elementen (Feschotte *et al.* 2002). Das Apfelgenom beispielsweise besteht zu 67 % aus repetitiven Elementen, wobei 42,4 % des Gesamtgenoms auf Transposons entfallen (Velasco *et al.* 2010). Im Mais machen transposable Elemente sogar 84 % des Genoms aus (Schnable *et al.* 2009). Aufgrund ihres hohen genomischen Anteils und der Fähigkeit an neuen Stellen im Genom zu integrieren, spielen transposable Elemente eine wichtige Rolle für die Struktur und Evolution von Pflanzengenomen (Bennetzen 2000). Oft verursachen sie weitreichende phänotypische Effekte in den Wirtspflanzen, die gewissermaßen auch der Auslöser für die Entdeckung der transposablen Elemente waren. So wurden sie erstmals im Mais von Barbara McClintock beschrieben (McClintock 1944). Sie untersuchte den C Locus, welcher für die Färbung von Maiskörnern verantwortlich ist. Bei Anwesenheit des Dissociation (Ds) Locus innerhalb des C Locus entsteht eine gelbe Färbung. Springt Ds aus diesem Locus, was die Anwesenheit eines Activator (Ac) Locus voraussetzt, entsteht wiederum das Wildtyp-Allel mit violetter Färbung (McClintock 1946, 1950; 1951; 1953). Mittlerweile ist bekannt, dass es sich bei Ds und Ac um Mitglieder einer Transposonfamilie (Ac/Ds) handelt (Bennetzen 2000). Ein wichtiges Merkmal bei der Charakterisierung transposabler Elemente einer Familie ist die Unterscheidung in autonome und nicht-autonome Elemente (Wicker *et al.* 2007). Ac besitzt alle benötigten Domänen zur Transposition und ist somit ein autonomes Element. Ds hingegen ist ein nicht-autonomes Element, da es nicht zur eigenständigen Transposition befähigt ist. Die Transposition nicht-autonomer Elemente durch Anwesenheit autonomer Elemente beschränkt sich nicht nur auf Mitglieder einer Familie, sondern gilt auch familienübergreifend, was als „cross-mobilisation“ bezeichnet wird (Feschotte *et al.* 2003).

Das erste Klassifizierungssystem für Transposons wurde 1989 von Finnegan eingeführt. Es unterscheidet anhand der Zwischenstufe bei der Transposition. Die Zwischenstufe ist bei der Klasse I Transposons (Retrotransposons) ein RNA-Molekül, bei Klasse II Transposons (DNA-Transposons) ein DNA-Molekül. Der Mechanismus der Transposition wird bei den Retrotransposons als „copy-and-paste“ bezeichnet (Finnegan 1989). Retrotransposons lassen sich aufgrund ihres Integrationsmechanismus weiter in LTR- und Non-LTR-

Retrotransposons unterteilen (Lisch 2013). Anhand des Grads ihrer Sequenzähnlichkeit (Xiong und Eickbush 1990) und der Reihenfolge der kodierten Gene (Bennetzen 1996) werden LTR-Retrotransposons in die Subklassen *Ty1-Copia* und *Ty3-Gypsy* untergliedert. Die Non-LTR-Retrotransposons lassen sich in LINEs (long interspersed nuclear elements) und SINEs (short interspersed nuclear elements) unterteilen. LINEs enthalten eine Vielzahl der Gene der LTR-Retrotransposons, sind aber ansonsten einfacher aufgebaut (Kumar und Bennetzen 1999). SINEs hingegen besitzen keines der für die Transposition benötigten Gene. Sie stammen bis auf wenige Ausnahmen (z. B. die Alu-Elemente) von tRNA-Genen ab (Okada 1991a, b). DNA-Transposons sind meist durch kurze „terminal inverted repeats“ (TIRs) und ein Transposase-Gen (TPase) charakterisiert (Jiang *et al.* 2004). Eine besondere Gruppe der DNA-Transposons stellen die sogenannten MITEs (miniature inverted-repeat transposable elements) dar. Diese nicht-autonomen Elemente sind besonders kurz (< 600 bp) und treten bevorzugt in nicht-kodierenden Bereichen von Genen auf (Bureau und Wessler 1992, 1994). Zudem stellen sie meist den Hauptanteil der DNA-Transposons. In *Arabidopsis* wurden im Vergleich zu ungefähr 1.000 anderen DNA-Transposons etwa 1.200 MITEs identifiziert (TAGI 2000). In Abbildung 1.2 sind die Hauptgruppen der transposablen Elemente dargestellt, die in Pflanzen vorkommen.

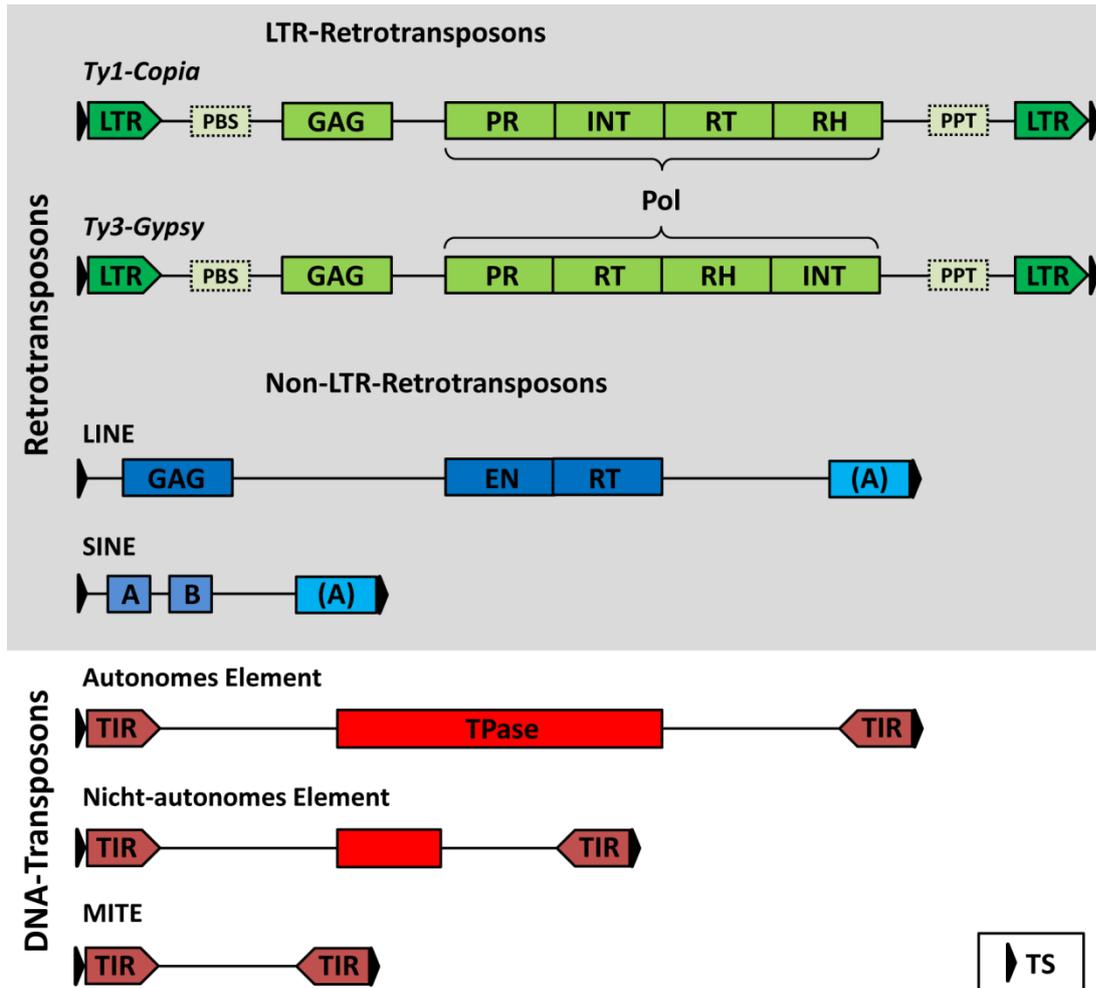


Abbildung 1.2 Transposable Elemente in Pflanzen

Die Abbildung zeigt schematisch die Struktur transposabler Elemente in Pflanzen, aufgeteilt in Retrotransposons (grau hinterlegt) und DNA-Transposons (weiß hinterlegt). Die Retrotransposons sind unterteilt in LTR-Retrotransposons mit den Hauptvertretern *Copia* und *Gypsy* und in Non-LTR-Retrotransposons mit den Vertretern LINE und SINE. Die DNA-Transposons gliedern sich in ein allgemeines autonomes Element, ein allgemeines nicht-autonomes Element mit nicht funktionsfähiger Transposase und in ein MITE. TS: „target site“; LTR: „long terminal repeat“; PBS: „primer binding site“; PPT: „polypurine tract“; PR: Protease; INT: Integrase; RT: reverse Transkriptase; RH: RNaseH; EN: Endonuklease; A, B: A- und B-Promotor Erkennungsstelle für die RNA Polymerase III; (A): Abfolge der Base Adenin; TIR: „terminal inverted repeat“; TPase: Transposase; GAG: Capsid Protein

Die Grafik wurde in Anlehnung an die Publikationen von Casacuberta und Santiago (2003), Kumar und Bennetzen (1999) und Wicker *et al.* (2007) erstellt.

Charakteristisch für transposable Elemente ist die „target site“ (TS), bzw. die bei der Integration eines Transposons an einen neuen Locus entstehende Duplikation dieser Sequenz („target site duplication“, TSD). So besitzen LTR-Retrotransposons typischerweise eine TSD von 4 bis 6 bp, hAT Elemente, eine Superfamilie der DNA-Transposons, hingegen eine TSD von 8 bp (Wicker *et al.* 2007). Eine Ausnahme bilden die Helitrons, eine weitere

Superfamilie der DNA-Transposons. Diese transposablen Elemente bedienen sich im Fall einer Transposition im Vergleich zum „cut-and-paste“-Mechanismus der meisten DNA-Transposon einem „rolling circle“-Mechanismus, bei dem keine TSD entsteht (Kapitonov und Jurka 2001).

Transposable Elemente sind aufgrund ihrer Transpositionsmöglichkeiten der variabelste Teil eines Genoms und besitzen einen großen Einfluss auf dessen Architektur (Lisch 2013). Die Art und Weise, wie Transposons zu genotypischen und phänotypischen Veränderungen führen, ist sehr unterschiedlich. Im Folgenden sollen einige Möglichkeiten anhand von Beispielen erläutert werden. Der einfachste Weg ist die Inaktivierung von Genen durch Erzeugung einer Nullmutation. Beispielsweise zerstört die Integration eines LTR-Retrotransposons in ein Intron des *MdPI* Gens bei den Apfelsorten Rae Ime, Spencer Seedless und Wellington Bloomless dessen Funktionsfähigkeit, was zu kernlosen Früchten führt (Yao *et al.* 2001). Die Insertion von Transposons in regulatorische Elemente wie Enhancer oder Silencer kann zu einer veränderten Genexpression führen. Ein besonders eindrucksvolles Beispiel ist der *Vegetative to generative transition 1 (Vgt1)*-Locus im Mais. Die Integration eines MITEs in eine CNS („conserved non-coding sequence“) des *Vgt1*, der sich rund 70 kb „upstream“ des Transkriptionsfaktors *ZmRap2.7* befindet, führt zu dessen Überexpression, was sich phänotypisch in einer späteren Blütezeit ausprägt (Salvi *et al.* 2007). Neben dem Einfluss auf die Genexpression durch Beeinflussung bereits vorhandener regulatorischer Elemente gibt es Beispiele, in denen Transposons durch Einbringen neuer Enhancer oder Promotoren die Genexpression verändern. Dies wird dadurch begünstigt, dass einige transposable Elemente bevorzugt in 5'-Bereiche von Genen springen, nicht aber in Exons (Naito *et al.* 2009). Die Enhancer Aktivität des Retrotransposons *Hopscotch*, welches im Mais rund 60 kb „upstream“ des *teosinte branched 1 (tb1)*-Locus integrierte, führt zu einer Überexpression von *tb1* (Studer *et al.* 2011). Es handelt sich bei *tb1* um einen Transkriptionsfaktor, der die Ausbildung von Verzweigungen unterdrückt. Dies führt zu einer starken Verringerung der Seitentriebe im Mais im Vergleich zum Teosinte, der als Urvater des Kulturmais gilt (Doebley *et al.* 2006). Neben regulatorischen Informationen enthalten autonome DNA-Transposons ein Transposase-Gen, welches neben seiner Funktion für die Transposition möglicherweise auch eine wichtige Rolle für die Wirtspflanze spielt (Muehlbauer *et al.* 2006). In *Arabidopsis thaliana* konnte gezeigt werden, dass das Gen *DAYSLEEPER*, welches essentiell für eine normale Wuchsform der Pflanzen ist, von dem Transposase-Gen eines hAT-Elements abstammt (Bundock und Hooykaas 2005). Darüber

hinaus kann es durch die Integration eines Transposons auch zu epigenetischen Veränderungen kommen. Im Fall des *FLOWERING WAGENINGEN (FWA)*-Locus in *Arabidopsis thaliana* bewirkt ein SINE, dass dieses Gen in vegetativen Geweben epigenetisch ausgeschaltet wird – ein Prozess in dem sowohl kleine RNAs wie auch DNA-Methylierung beteiligt sind (Kinoshita *et al.* 2007; Fujimoto *et al.* 2008). Mutanten, in denen das Gen nicht ausgeschaltet ist, zeigen eine späte Blüte. Ein epigenetisches Ausschalten von vor allem autonomer transposabler Elemente mittels kleiner RNAs, DNA-Methylierung und Histon-Modifizierung dient Pflanzen zur Kontrolle von und zum Schutz vor Transposons (Lisch 2009). Die bisher beschriebenen Einflüsse transposabler Elemente beziehen sich meist auf einzelne Gene, also begrenzte Loci. Transposons können jedoch auch zu größeren Veränderungen in der Architektur von Chromosomen führen. Dieses Phänomen ist beispielsweise beim Mais beschrieben, wo es durch Rekombination zwischen Ac-Elementen zu Deletionen, Inversionen und Translokationen kommt (Yu *et al.* 2011).

1.4 Positionelle Klonierung

Die Schwierigkeit in der Sequenzierung von Pflanzengenomen liegt, wie bereits erwähnt, in der starken Ausprägung von Polyploidie, Heterozygotie und Paralogie, gepaart mit großen Genomen und einem hohen Anteil repetitiver Elemente (Hamilton und Robin Buell 2012). Oftmals sind jedoch nicht komplette Genome, sondern nur einzelne Gene oder Chromosomenbereiche von Interesse. Zur strukturellen und funktionellen Analyse meist unbekannter Gene ist die positionelle Klonierung nach wie vor eine weit verbreitete Methode. Grundvoraussetzungen hierfür sind die Verfügbarkeit genomischer DNA-Bibliotheken mit einer hohen Genomabdeckung und das Vorhandensein gekoppelter Marker, die das Gen von Interesse flankieren (Tanksley *et al.* 1995). Neben der Isolierung hochmolekularer DNA (Zhang *et al.* 1995) ist die Wahl eines geeigneten Vektors für die Konstruktion genomischer DNA-Bibliotheken von entscheidender Bedeutung. Weit verbreitete Vektoren sind „yeast artificial chromosome (YAC)“ (Burke *et al.* 1987), „bacterial artificial chromosome (BAC)“ (Shizuya *et al.* 1992) und „P1-derived artificial chromosome (PAC)“ (Ioannou *et al.* 1994). YACs ermöglichen im Vergleich zu BACs und PACs sehr große Integrate. Während BACs und PACs DNA-Moleküle einer Größe von bis zu 300 kb aufnehmen können, sind bei der Verwendung von YACs Integratgrößen bis 2 Mb möglich (Monaco und Larin 1994). Allerdings weisen YACs auch erhebliche Nachteile auf: Sie sind oft instabil und neigen zur Bildung von Chimären (Green *et al.* 1991). Der Anteil an chimären

YACs beträgt bis zu 40 % (Umehara *et al.* 1995). Dies erschwert im Besonderen „Chromosome Walk“-Strategien zur Analyse eines größeren Genomabschnitts. Im Vergleich dazu machen bei BAC-Bibliotheken chimäre Klone lediglich 4 bis 11 % der Gesamtklone aus (Volik *et al.* 2003). Zudem ist die Isolierung von YAC-DNA kompliziert. Hauptfaktoren sind die enorme Größe und die Linearität der YACs und die darin begründete hohe Ähnlichkeit zum Wirtsgenom (Monaco und Larin 1994). Im Vergleich dazu können die zirkulären BACs und PACs mittels standardisierter Plasmidpräparationsprotokolle in ausreichender Menge gewonnen werden (Vinatzer *et al.* 1998). Die Linearität der YACs führt zu einem weiteren Nachteil. So ist die Effektivität der Transformation linearer Moleküle erheblich niedriger als zirkulärer (Woo *et al.* 1994). Darüber hinaus werden bei der Transformation Hefe-Sphäroplasten benötigt, deren Transformation relativ ineffizient ist. Somit werden große Mengen DNA zur Herstellung der Bibliothek benötigt (Monaco und Larin 1994). Die hohe Stabilität, die einfache Isolierung und der niedrige Anteil an chimären Klonen machen BACs zu idealen Vektoren für die Erstellung von genomischen DNA-Bibliotheken (Choi *et al.* 1995).

Aus einer BAC-DNA-Bibliothek können dann mit Hilfe flankierender Marker Klone, die das gewünschte Gen bzw. Teile des Gen-Locus enthalten, identifiziert werden. Für das Screening einer DNA-Bibliothek gibt es prinzipiell zwei Strategien. Beim PCR-basierenden Screening werden die Plasmide der rekombinanten Klone isoliert und positive Klone durch PCR ermittelt (Kim *et al.* 1996). Durch die Verwendung von Pools ist dies eine schnelle Methode, die zudem sehr spezifisch ist (Xu *et al.* 2001). Eine Alternative stellt die Methode der Koloniefilter-Hybridisierung nach Grunstein und Hogness (1975) dar. Die Klone einer gesamten Genbank werden dazu auf Membranen übertragen, die anschließend mit radioaktiv- oder nicht-radioaktiv-markierten Sonden hybridisiert werden, um positive Klone zu identifizieren (Patocchi *et al.* 1999). Die Verwendung radioaktiv-markierter Sonden führt in der Regel schnell und eindeutig zu Hybridisierungssignalen von Klonen, die die zur Sonde homologe Sequenz enthalten (Kaufmann *et al.* 1997; Patocchi *et al.* 1999). Die Methode der Koloniefilter-Hybridisierung ist im Vergleich zur PCR-basierenden Methode weniger kostenintensiv, da nur die Plasmide hybridisierender Klone isoliert werden.

Die anschließende Sequenzierung identifizierter Klone kann traditionell mittels Sanger-Sequenzierung, mittels neuerer Sequenziermethoden wie einer Roche 454 Pyro- oder einer Illumina-Sequenzierung oder aber durch eine Kombination der Methoden erfolgen (Imelfort und Edwards 2009).

1.5 Molekulare Marker

Gehölzpflanzen zeigen einen hohen Grad an genetischen Polymorphismen (Hamrick *et al.* 1990), was bei Kreuzungen zu einem hohen Maß an Heterozygotie in den einzelnen Individuen führt (Hemmat *et al.* 1994). Heterozygotien sind eine gute Voraussetzung für die Erstellung von molekularen Markern (Hemmat *et al.* 1994). In Kreuzungsexperimenten können anhand von molekularen Markern Kopplungskarten erstellt werden. Diese sind entscheidend bei der Aufklärung der genetischen Basis phänotypischer Merkmale und letztlich der Lokalisierung und Analyse von Schlüsselgenen (Tanksley *et al.* 1989), etwa bei der Kartierungs-gestützten Klonierung (Wing *et al.* 1994). Darüber hinaus sind molekulare Marker im Bereich der Züchtung von besonderer Bedeutung (Liebhard *et al.* 2002). Im Falle der Selektion hinsichtlich der kolumnaren Wuchsform dauert es etwa zwei bis drei Jahre, bis man anhand des Phänotyps geeignete Nachkommen erkennen kann (Moriya *et al.* 2009). Mithilfe von molekularen Markern kann man bereits deutlich früher selektieren und spart neben Zeit auch Kosten und Platz für die Kultivierung der Pflanzen (Conner *et al.* 1997). Die wichtigsten Arten von molekularen Markern sind im Folgenden kurz charakterisiert.

Bei **RFLP-Markern** (restriction fragment length polymorphism) wird genomische DNA mit verschiedenen Restriktionsenzymen sequenzspezifisch geschnitten und die Restriktionsfragmente gelelektrophoretisch aufgetrennt (Beek *et al.* 1992). Infolge von Mutationen einzelner Basen, die Restriktionsschnittstellen generieren oder zerstören, sowie von Insertionen und Deletionen (Indels) können Unterschiede in der Größe der Restriktionsfragmente zwischen Individuen auftreten. Diese Unterschiede in der Größe der Restriktionsfragmente können durch Southern Blot-Hybridisierungen mit geeigneten „single copy“-Sonden detektiert werden (Botstein *et al.* 1980). Nachteile an RFLP-Markern sind der hohe Zeit- und Kostenaufwand bei der Analyse großer Populationen und die limitierte Verfügbarkeit geeigneter Restriktionsschnittstellen (Beckmann 1988).

Die Verwendung von **RAPD-Markern** (randomly amplified polymorphic DNA) ist im Vergleich dazu wesentlich schneller und benötigt minimale Mengen an DNA (Williams *et al.* 1990). Durch PCR mit Primern zufälliger Basenzusammensetzung werden DNA-Fragmente unterschiedlicher Länge amplifiziert, welche anschließend mittels Gelelektrophorese aufgetrennt werden (Welsh und McClelland 1990). Polymorphismen treten durch das Vorhandensein beziehungsweise Fehlen von Banden auf, sowie in unterschiedlicher Größe der amplifizierten Bereiche.

Bei **AFLP-Markern** (amplified fragment length polymorphism) wird die DNA typischerweise mit den Restriktionsenzymen *EcoRI* und *MseI* verdaut und anschließend Adapter, die spezifisch für die Schnittstellen der Restriktionsenzyme sind, angefügt. Es folgt eine Amplifikation der Restriktionsfragmente durch PCR mit zur Adaptersequenz komplementären Primern. Zunächst wird mit Primern, die am 3'-Ende eine zusätzliche Base enthalten, eine selektive Präamplifikation durchgeführt, gefolgt von einer selektiven Amplifikation mit Primern, welche am 3'-Ende zwei bis drei Basen in das zu amplifizierende Fragment hineinreichen. Durch das Anfügen der selektiven Basen am 3'-Ende der Primer und die Verwendung zweier Restriktionsenzyme wird die Menge an verschiedenen Amplifikaten reguliert, sodass die Komplexität des resultierenden Musters gesteuert werden kann und auf diese Weise eine Art Fingerprint entsteht (Vos *et al.* 1995). Durch eine End-Markierung eines Primers (Fluoreszenz oder Radioaktivität) können die mittels Gelelektrophorese aufgetrennten Amplifikate detektiert werden. Anwendung finden die AFLP-Marker bei dem „genetischem Fingerabdruck“ in der Kriminalistik und beim Vaterschaftstest (Prochazka *et al.* 2001).

Bei **SSR-Markern** (simple sequence repeat) dienen Polymorphismen in Mikrosatelliten als Marker (Litt und Luty 1989; Weber und May 1989). Mikrosatelliten sind kurze repetitive Sequenzen, welche tandemartig hintereinander angeordnet sind. Die Länge der sich wiederholenden Einheiten liegt meist zwischen 1 und 6 bp. Mikrosatelliten kommen sehr häufig, statistisch betrachtet etwa alle 50 kb, im Pflanzengenom vor (Morgante und Olivieri 1993), sind hypervariabel hinsichtlich der Zahl der „tandem repeats“ und über das komplette Genom verstreut (Wu und Tanksley 1993). Diese Eigenschaften machen Mikrosatelliten zu einer vielversprechenden Quelle für molekulare Marker. Die Entstehung der Hypervariabilität kann durch „replication slippage“ bei der mitotischen Vermehrung von Soma- und Keimzellen erklärt werden (Seyffert 2003). Die Fragmentlängenpolymorphismen der SSR-Marker werden durch PCR und anschließende Gelelektrophorese nachgewiesen.

SCAR-Marker (sequence characteristic amplified region) sind sehr spezifische Marker, die einem genetischen Locus bekannter Sequenz entstammen (Paran und Michelmore 1993). Durch PCR mit spezifischen Primern wird der Locus amplifiziert und das Amplifikat gelelektrophoretisch aufgetrennt. Meist werden SCAR-Marker aus bekannten RAPD- oder AFLP-Amplifikaten entwickelt (Paran und Michelmore 1993; Kim *et al.* 2003). Die entsprechenden Fragmente werden dazu kloniert und sequenziert, um spezifische Primer erstellen zu können. Die Vorteile eines SCAR-Markers sind die Reproduzierbarkeit, eine

hohe Spezifität und eine einfache Handhabung (Tartarini *et al.* 1999; Tian *et al.* 2004). **Indel-Marker** (insertion deletion polymorphism) basieren auf Sequenzunterschieden zweier oder mehrerer Individuen in Form von Insertionen oder Deletionen, die durch direkten Sequenzvergleich detektiert werden (Rickert *et al.* 2003). Die entsprechenden Unterschiede werden mit flankierenden Primern mittels PCR amplifiziert und die Amplifikate anschließend gelelektrophoretisch aufgetrennt (Vali *et al.* 2008).

SNPs (single nucleotide polymorphism) sind die am häufigsten auftretende Art an Sequenzvariation im Genom und damit hochinteressant für die Verwendung als molekulare Marker (Newcomb *et al.* 2006). Drei Arbeitsschritte sind üblicherweise notwendig, um SNP-Marker zu etablieren. Zunächst wird ein SNP bioinformatisch durch Sequenzvergleiche oder durch molekularbiologische Techniken wie die Allel-spezifische Hybridisierung (Kwok 2001) nachgewiesen. In einem zweiten Schritt wird der SNP validiert, das bedeutet beispielsweise den Ausschluss eines Sequenzierfehlers. In einem dritten und letzten Schritt werden eine Vielzahl an Individuen im Idealfall durch „high-throughput genotyping“-Verfahren auf den SNP hin untersucht (Chagné *et al.* 2008).

Bis heute gibt es eine Vielzahl molekularer Marker, deren Kopplung mit dem Kolumnargen nachgewiesen ist. Bereits 1997 lokalisierten Conner *et al.* das *Co*-Gen auf Kopplungsgruppe 10 einer McIntosh Wijcik Kopplungskarte anhand der dominanten RAPD-Marker P459z-800 und P255a-1145. Allerdings bestand die Kopplungskarte noch aus 20 Kopplungsgruppen. Hemmat *et al.* (1997) verwendeten die Marker P459z-800, P255a-1145 und einen neuen Marker (SSR^{Co}) in Kreuzungsexperimenten der Sorten McIntosh Wijcik und NY 75441-58. Anhand der Rekombinationsfrequenzen wurden die genetischen Abstände auf 24 cM für *Co*/P255a-1145, 8,9 cM für *Co*/P459z-800, 10 cM für SSR^{Co}/P459z-800 und 6 cM für *Co*/SSR^{Co} bestimmt. Das *Co*-Gen wird nach diesen Erkenntnissen von den Markern SSR^{Co} und P459z-800 flankiert, mit Tendenz zu dem Marker SSR^{Co}. Ein Jahr später bestätigten Maliepaard *et al.* (1998) mittels des SSR-Markers USA-SSR11 die Lokalisation des *Co*-Gens auf Kopplungsgruppe 10 einer Kopplungskarte der Sorten Prima und Fiesta (je 17 Kopplungsgruppen). Kim *et al.* (2003) untersuchten eine Nachkommenschaft der normalwüchsigen Sorte Fuji und kolumnaren Sorte Tuscan mit den molekularen Markern SSR^{Co}, P459z-800 und P255a-1145, sowie dem selbst erstellten RAPD-Marker WB82₆₇₀ beziehungsweise dem daraus resultierenden SCAR-Marker SCB82₆₇₀. Der genetische Abstand für P459z-800 beträgt 10,8 cM, für SSR^{Co} 19,1 cM und für SCB82₆₇₀ lediglich 1,8 cM. Der RAPD-Marker P255a-1145 wies hingegen keine Kopplung mit dem *Co*-Gen auf. Tian *et*

al. (2005) untersuchten die Nachkommenschaft einer Kreuzung zwischen der normalwüchsigen Sorte Fuji und der kolumnaren Sorte Telamon mit den Markern SSR^{Co}, P459z-800, P255a-1145 und SCB82₆₇₀. Die genetische Distanz der Marker zu *Co* beträgt für SSR^{Co} 26,5 cM und für P459z-800 13,2 cM. Sowohl für den Marker P255a-1145 wie auch für SCB82₆₇₀ konnte keine Kopplung gefunden werden. Als Grund für den unterschiedlichen Kopplungsgrad der Marker SSR^{Co} und P459z-800 werden die unterschiedlichen Populationsgrößen genannt (Kim *et al.* 2003; Tian *et al.* 2005). Tian *et al.* (2005) wiesen zudem eine starke Kopplung (3,9 cM) des Markers Ch03d11 (Liebhard *et al.* 2002) mit dem *Co*-Gen nach. Zudem etablierten sie den RAPD-Marker S1142₆₈₂ und den AFLP-Marker E-ACT/M-CTA₃₄₆, welche sie in die SCAR-Marker SCAR₆₈₂ und SCAR₂₁₆ konvertierten. SCAR₆₈₂ und SCAR₂₁₆ flankieren das *Co*-Gen mit genetischen Abständen von 2,9 und 12,3 cM. Moriya *et al.* (2009) zeigten für den SSR-Marker Hi01a03 (Silfverberg-Dilworth *et al.* 2006) eine enge Kopplung zum *Co*-Gen, darüber hinaus bestätigten sie die Kopplung der Marker SCAR₆₈₂, SCAR₂₁₆ und Ch03d011, widerlegten aber die Kopplung des Markers SCB82₆₇₀ mit *Co*. SCB82₆₇₀ scheint sortenspezifisch für Tuscan zu sein (Fernández-Fernández *et al.* 2008). Moriya *et al.* (2009) wiesen, wie bereits Tian *et al.* (2005), abweichende Rekombinationsfrequenzen ausgewählter Marker in unterschiedlichen Populationen nach. In einer Population der Kreuzung Fuji und 8H-9-45 cosegregieren Ch03d11 und SCAR₆₈₂ mit *Co*. In einer Population der Kreuzung Fuji und 5-12786 hingegen beträgt der genetische Abstand zu *Co* für Ch03d11 und für SCAR₆₈₂ 1,4 cM. Bai *et al.* (2012) verwendeten neben den bereits genannten Markern SCAR₆₈₂, Ch03d11 und Hi01a03, die SSR-Marker EMPc105 (Fernández-Fernández *et al.* 2006) und AU223548 (Silfverberg-Dilworth *et al.* 2006) zur groben Eingrenzung der *Co*-Gen-Region. Zur Feinkartierung wurden weitere SSR-Marker etabliert, die auf eine Nachkommenschaft von 528 Bäumen (4 Kreuzungen) und 290 weitere kolumnare Bäume getestet wurden. Der Marker C18470-25831 cosegregiert in allen getesteten Pflanzen mit *Co*, der Marker C1753-3520 weist zwei Rekombinante auf, die Marker C7629-22009 und C6835.383-2 einen beziehungsweise vier Rekombinante (Bai *et al.* 2012). Moriya *et al.* (2012) etablierten weitere SSR-Marker durch Kreuzungsanalysen an 1000 Nachkommen (31 Populationen). Sie fanden drei SSR-Marker, die in allen Nachkommen mit *Co* cosegregieren; dies sind Mdo.chr10.12, Mdo.chr10.13 und Mdo.chr10.14. Flankiert wird die vermutete *Co*-Gen-Region mit den Markern Mdo.chr10.11 und Mdo.chr10.15 mit einem genetischen Abstand zu *Co* von lediglich 0,1 bzw. 0,2 cM. Weitere Kreuzungsanalysen wurden von Baldi *et al.* (2012) durchgeführt. Sie untersuchten

1250 Nachkommen zweier Kreuzungen zwischen GD und McIntosh Wijcik und etablierten unter anderem die Co-Gen gekoppelten SSR-Marker Co04R11 (0,4 cM), Co04R12 (gekoppelt) und Co04R13 (0,16 cM). Abbildung 1.3 zeigt die Co-Gen gekoppelten Marker angeordnet nach ihrer Lokalisation im GD-Genom auf Chromosom 10 (Velasco *et al.* 2010). Die RAPD-Marker P459z-800 und P255a-1145 (Conner *et al.* 1997) konnten nicht lokalisiert werden, da die Sequenzen nicht zur Verfügung stehen und die verwendeten Primer mit 10 Nukleotiden für eine erfolgreiche BLAST-Suche zu kurz sind. Der SCAR-Marker SCB82₆₇₀ (Kim *et al.* 2003) ist ebenfalls nicht abgebildet, da er durch BLAST-Suchen auf Chromosom 3 Position 31,293 Mb im GD-Genom lokalisiert wurde.

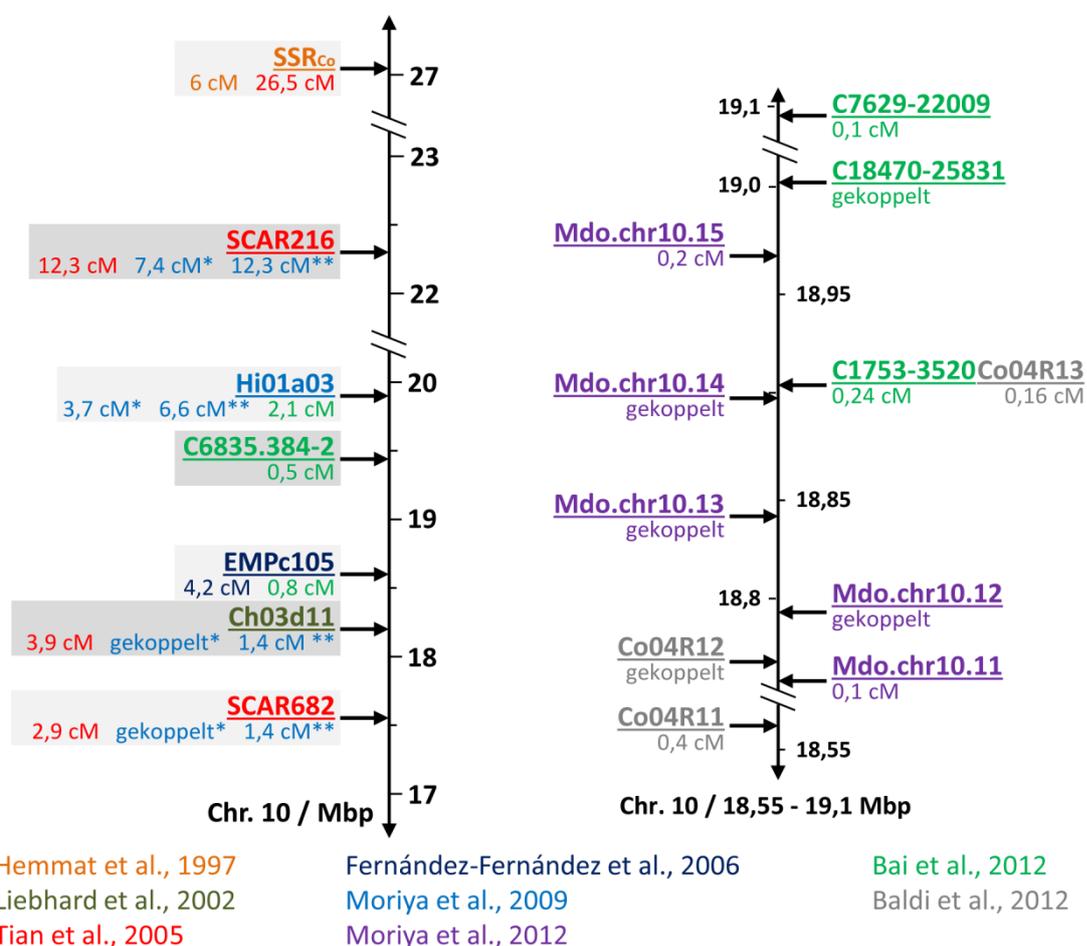


Abbildung 1.3 Co-Gen gekoppelte Marker

Die Grafik zeigt eine Auswahl molekularer Marker, die Co-Gen gekoppelt sind. Die Anordnung der Marker erfolgte nach deren genomischer Lage im GD-Genom. Unterhalb der Namen der Marker sind ihre genetischen Abstände zum Kolumnenrand aufgeführt (* und ** kennzeichnen Ergebnisse aus zwei verschiedenen Populationen). Die Schriftfarbe gibt Aufschluss über die Herkunft des Markers und den genetischen Abstand. Der Bereich 18,55 bis 19,1 Mbp auf Chromosom 10 ist im rechten Teil der Abbildung vergrößert dargestellt, da er vermutlich den Co-Locus enthält und dementsprechend eine hohe Markerdichte aufweist.

1.6 Zielsetzung

Das Ziel der vorliegenden Arbeit ist die Aufklärung der molekularen Ursache der *Co*-Mutation, die beim Apfel zum Kolumnarwachstum führt. Trotz der großen wirtschaftlichen Bedeutung sind die bisherigen Erkenntnisse über diese Mutation sehr begrenzt. Die Untersuchung von *Co* soll zunächst durch ein klassisches, positionelles Klonierungsexperiment erfolgen. Neben der Lokalisation der *Co*-Mutation und die Erstellung geeigneter Sonden soll mit Hilfe der Konstruktion von genomischen BAC-Bibliotheken mit mehrfacher Genomabdeckung die *Co*-Region kloniert und deren Sequenz bestimmt werden. Das Material stellt die Hochschule Geisenheim in Form von *in-vitro*-Kulturen der heterozygot kolumnaren Apfelsorte Procats 28 zur Verfügung. Genomische Illumina „mate pair“-Bibliotheken dieser Sorte sollen zudem sequenziert und bioinformatisch analysiert werden, um die genomische Region vollständig zu annotieren und, wenn möglich, die Sequenzen den jeweils homologen Chromosomen zuzuordnen. Somit soll eine möglichst vollständige genomische Referenz der *Co*-Region einer kolumnaren Apfelsorte erstellt werden, die die Grundlage für weitere Analysen bildet. Zusammenfassend soll in der vorliegenden Arbeit die „traditionelle“ Methode der positionellen Klonierung mit den neueren Techniken des Next-Generation Sequencings kombiniert werden, um auf diese Weise eine verlässliche und möglichst vollständige genomische Struktur der Zielregion zu etablieren.

Des Weiteren sollen anhand der gewonnenen Sequenzinformationen neue mit der *Co*-Mutation gekoppelte Marker erstellt und getestet werden. Das Ziel der Generierung von eng gekoppelten Markern ist die Feinkartierung der genomischen Region, sowie die Möglichkeit, Apfelbäume bereits in einem frühen Entwicklungsstadium zuverlässig zu genotypisieren, was für die Apfelzüchtung von großer Bedeutung ist. Dazu steht eine Nachkommenschaft von etwa 100 Individuen einer Kreuzung der Apfelsorten Procats 28 und A14 zur Verfügung, sowie eine Sammlung genomischer DNAs von weiteren kolumnaren und nicht kolumnaren Apfelsorten. Die Arbeit ist ein Teilprojekt des vom Bundesministerium für Ernährung, Landwirtschaft und Verbraucherschutz geförderten Projekts (Nr. 511-06.01-28-1-43.042-07) zur Identifikation und Funktionsaufklärung des *Co*-Gens.

2 Material und Methode

2.1 Allgemeine Methoden zum Arbeiten mit DNA

2.1.1 Elektrophoretische Auftrennung von DNA

Die Auftrennung von DNA-Molekülen erfolgte je nach Anwendung in unterschiedlichen Agarosegelelektrophorese-Systemen. Die verwendeten Protokolle wurden in Anlehnung an das Protokoll von Sambrook *et al.* (1989) erstellt. Als Marker dienten die in Abbildung 2.1 dargestellten Molekulargewichtsstandards der Fa. Fermentas (St. Leon-Rot, Deutschland). Die Proben wurden mit ca. 1/6 Volumen Bromphenolblau-DNA-Ladepuffer versetzt, um diese zu beschweren und um die Lauffront sichtbar zu machen. Lagen die zu erwartenden Banden in der Lauffront von Bromphenolblau, wurde als Ladepuffer Orange Dye oder 30 %-ige Saccharose-Lösung verwendet.

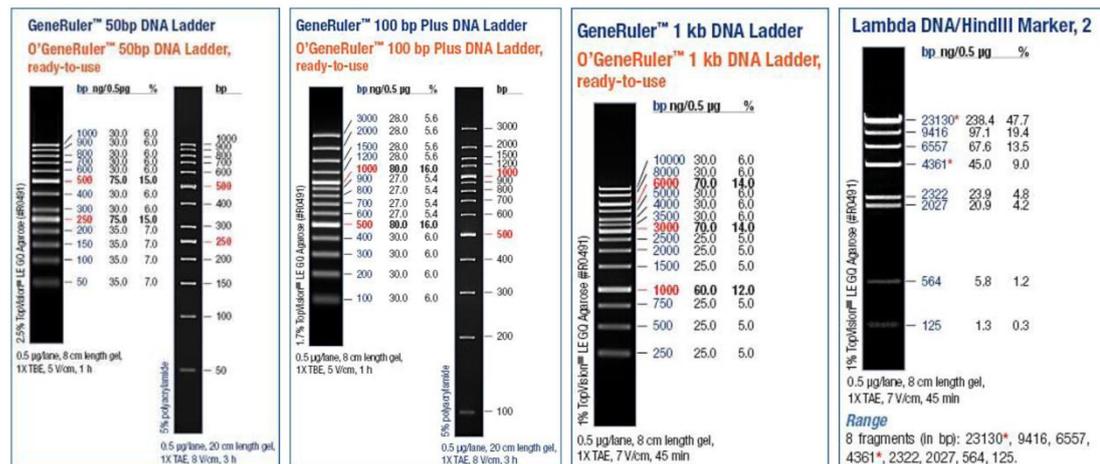


Abbildung 2.1 Verwendete Molekulargewichtsstandards der Fa. Fermentas

Abgebildet sind die in dieser Arbeit verwendeten Molekulargewichtsstandards der Fa. Fermentas.

2.1.1.1 Testgele

Zur qualitativen Überprüfung von DNA-Proben wurde die vertikale GENTERPHORESE™-Kammer der Fa. GENTERPRISE (Mainz, Deutschland) verwendet. Die Elektrophorese erfolgte in Gelen (7 cm x 6,5 cm x 0,1 cm) mit Agarosekonzentrationen von 0,8 % bis 2 % in 1 x TBE-Puffer im TBE-Puffer-System bei 100 bis 150 mA für 30 bis 45 min. Als Molekulargewichtsstandards wurden 2 µl der oben genannten Marker verwendet. Im Anschluss an die Elektrophorese wurden die Gele für 3 min in einer Ethidiumbromid-(EtBr)-Färbelösung (5 µg/ml in 1 x E-Puffer) inkubiert, kurz gewässert und auf einem Transilluminator der Fa. BACHOFER (Reutlingen, Deutschland) bei 312 nm sichtbar gemacht. Die Dokumentation der

Gele erfolgte über das Carestream Gel Logic 112 System (Carestream Health, Rochester, USA). Alle in dieser Arbeit gezeigten Gelbilder sind invertiert dargestellt.

2.1.1.2 Analytische/präparative Gele

Zur besseren Auftrennung von DNA-Fragmenten aufgrund einer längeren Laufstrecke wurde die Elektrophorese in vertikalen „Maxi“-Gekassetten angewendet. Die „Maxi“-Gele (18 cm x 13 cm x 0,4 cm) wurden hierfür mit einer Agarosekonzentration von 0,8 % bis 1,5 % in 1 x E-Puffer bzw. 1 x TBE-Puffer hergestellt. Die Auftrennung erfolgte in einem speziellen Natriumphosphatpuffer-System (E-Puffer) bei 15 mA für 16 bis 20 h (Über-Nacht-Gele) oder im TBE-Puffer-System bei 30 mA für 3 bis 4 h (Über-Tag-Gele). Als Molekulargewichtsstandards wurden 4 µl der zuvor genannten Marker eingesetzt. Zur besseren Auftrennung kleinerer DNA-Fragmente (unter 1000 bp) wurden vertikale „Midi“-Gele (10,4 cm x 10,2 cm x 0,4 cm) mit einer Agarosekonzentration von 1 bis 2 % in 1 x TBE-Puffer verwendet. Die Auftrennung erfolgte bei 25 mA für 1 bis 2 h. Um größere DNA-Mengen (> 10 µg) gelelektrophoretisch aufzutrennen, beispielsweise bei der Größenfraktionierung zur BAC-Bank-Herstellung, wurde das „Horizontal System for Submerged Gel Electrophoresis“ der Fa. GIBCO/BRL, heute Life Technologies (Gaithersburg, USA), verwendet. Diese horizontalen „Mega“-Gele (25 cm x 20 cm x 0,7 cm) wurden mit einer Agarosekonzentration von 0,6 % in 1 x E-Puffer hergestellt. Die Parameter für die Auftrennung waren 100 mA für 16 bis 20 h in 1 x E-Puffer. Als Molekulargewichtsstandards wurden 10 µl Lambda/HindIII eingesetzt. Zur Auftrennung von sehr kleinen DNA-Fragmenten (< 200 bp) sowie zur parallelen Auftrennung vieler Proben (96-iger Format) wurde das horizontale Gelelektrophorese System „Sunrise™ 96“ der Fa. GIBCO/BRL, heute Life Technologies (Gaithersburg, USA), verwendet. Die Gele (24 cm x 12,5 cm x 0,5 cm) wurden mit Agarosekonzentrationen von 1 % bis 5 % in 1 x TBE hergestellt. Die Parameter eines Laufs waren 100 mA für 2 bis 3 h.

2.1.2 Wiedergewinnung von DNA-Fragmenten aus Agarosegelen

2.1.2.1 Wiedergewinnung durch Elektroelution

Der gewünschte DNA-Bereich wurde mittels Skalpell aus dem präparativen Gel ausgeschnitten und das entsprechende Gelstück mit 600 µl 0,5 x TBE in einen Dialyseschlauch der Fa. Medicell International (London, UK) überführt. Der Dialyseschlauch wurde mittels Kunststoff-Clips luftblasenfrei verschlossen und in eine Elektrophoresekammer mit 0,5 x TBE gelegt, die zur Kühlung auf Eis stand. Die

Elektroelution erfolgte für 2 h bei 100 mA. Um am Schlauch haftende DNA wieder in Lösung zu bringen, wurde die Polung für 5 min umgekehrt und anschließend für 1 min wieder in die ursprüngliche Richtung gestellt. Dieser Schritt wurde zweimal wiederholt. Die gelöste DNA wurde aus dem Dialyseschlauch in ein 2 ml Eppendorfgefäß überführt und der Dialyseschlauch mit 300 µl 0,5 x TBE nachgespült. Der Puffer aus dem Dialyseschlauch wurde mit dem ersten Eluat vereint.

2.1.2.2 Wiedergewinnung mittels Kit

Die DNA wurde mit dem NucleoSpin® Extract II Kit der Fa. Macherey-Nagel (Düren, Deutschland) nach dem Protokoll „DNA extraction from agarose gels“ nach Angaben des Herstellers wiedergewonnen. Dieses Kit fand vor allem bei der Wiedergewinnung von PCR-Produkten aus präparativen Agarosegelen Anwendung.

2.1.3 Phenol-Chloroform-Extraktion

Die Methode der Extraktion mit Phenol, Chloroform und Isoamylalkohol (PCI, 25:24:1; v:v:v) dient zur Abtrennung von Proteinen. Dazu wurden die Proben mit gleichem Volumen PCI versetzt und nach fünfminütigem Invertieren in der Mikroliterzentrifuge Mikro 200 der Fa. Hettich (Tuttlingen, Deutschland) zur Phasentrennung zentrifugiert (RT, 14000 Upm, 10 min). Die obere Nukleinsäure-haltige Phase wurde in ein neues Reaktionsgefäß überführt. Zur Entfernung von Phenolresten wurde 1 Volumen Chloroform-Isoamylalkohol (CI, 24:1; v:v) zugegeben, invertiert und erneut zentrifugiert. Abschließend wurde die obere, wässrige Phase abgezogen und in ein neues Reaktionsgefäß überführt.

2.1.4 Fällung von DNA

Die Fällung der DNA dient dem Reduzieren des Volumens sowie der Entfernung von Salzurückständen (Sambrook *et al.* 1989). Die DNA-Lösung wurde mit 1/10 Volumen 10 x Dialysepuffer versetzt und gut gemischt. Anschließend wurde das 2 bis 2,5-fache Volumen Ethanol abs. hinzugegeben und invertiert. Es folgte eine Inkubation bei -20 °C für mindestens 1 h. Durch die anschließende Zentrifugation in der Kühlzentrifuge 5804R der Fa. Eppendorf (Hamburg, Deutschland) bei 4 °C, 14000 Upm für mindestens 30 min wurde die ausgefällte DNA präzipitiert. Der Überstand wurde verworfen und die DNA durch Zugabe von 70 %-igem Ethanol und Zentrifugation in der Mikroliterzentrifuge Mikro 200 der Fa. Hettich (Tuttlingen, Deutschland) bei RT, 14000 Upm für 5 min gewaschen. Der Überstand

MATERIAL UND METHODE

wurde erneut verworfen und das DNA-Pellet luftgetrocknet. Das trockene DNA-Pellet wurde in HPLC-H₂O gelöst.

2.1.5 Polymerase-Ketten-Reaktion (PCR)

2.1.5.1 Standard PCR

Die Polymerase-Ketten-Reaktion ist eine Methode zur Amplifikation eines spezifischen DNA-Abschnittes (Mullis *et al.* 1992). Die PCRs wurden mit der GoTaq[®] DNA Polymerase (Promega, Madison, USA) in einem TProfessional Thermocycler (Biometra, Göttingen, Deutschland) durchgeführt. Alle in dieser Arbeit verwendeten Primer wurden von Life Technologies (Darmstadt, Deutschland) bezogen. Eine Liste der in dieser Arbeit verwendeten Primer befindet sich im elektronischen Anhang. Ein beispielhafter PCR-Ansatz sowie das entsprechende Programm sind den Tabellen 2.1 und 2.2 zu entnehmen.

Tabelle 2.1 Standard PCR-Ansatz

5x GoTaq [®] Flexi Puffer	10 µl
MgCl ₂ (25 mM)	4 µl
dNTPs (10 mM each)	1 µl
Forward Primer (0,01 mM)	2 µl
Reverse Primer (0,01 mM)	2 µl
GoTaq [®] DNA Polymerase (5 U/µl)	0,2 µl
Template DNA (20 ng/µl)	1 µl
HPLC-H ₂ O	29,8 µl
Gesamtvolumen	50 µl

Tabelle 2.2 PCR-Programm

Initiale	94 °C	4:00 min	
Denaturierung	94 °C	0:30 min	
Denaturierung	94 °C	0:30 min	40 x
Annealing	56 - 65 °C	0:30 min	
Elongation	72 °C	1 min/kb	
Finale	72 °C	10 min	
Elongation	72 °C	10 min	
	4 °C	∞	

2.1.5.2 „long range“ PCR

Die „long range“ PCRs (I_r-PCR) wurden mit der TaKaRa LA Taq[®] Hot Start Version (TaKaRa, Otsu, Japan) durchgeführt. Der PCR-Ansatz sowie das PCR-Programm sind in den Tabellen 2.3 und 2.4 aufgeführt. Die PCRs wurden in einem TProfessional Thermocycler (Biometra, Göttingen, Deutschland) durchgeführt.

Tabelle 2.3 „long range“ PCR-Ansatz

10x LA PCR Buffer	5 µl
dNTPs (2,5 mM each)	8 µl
Forward Primer (0,01 mM)	2 µl
Reverse Primer (0,01 mM)	2 µl
TaKaRa LA Taq [®] HS (5 U/µl)	0,2 µl
Template DNA (20 ng/µl)	2 µl
HPLC-H ₂ O	30,8 µl
Gesamtvolumen	50 µl

Tabelle 2.4 „long range“ PCR-Programm

Initiale	94 °C	1:00 min	
Denaturierung	94 °C	0:30 min	
Denaturierung	94 °C	0:30 min	40 x
Annealing	58 - 68 °C	0:30 min	
Elongation	68 °C	1 min/kb	
Finale	72 °C	10 min	
Elongation	72 °C	10 min	
	10 °C	∞	

2.1.6 Aufreinigung von PCR-Produkten

Um überschüssige Primer und Nukleotide, die bei Folgereaktionen stören könnten, zu entfernen, wurden die PCR-Ansätze mit dem Kit NucleoSpin® Extract II Kit der Fa. Macherey-Nagel (Düren, Deutschland) aufgereinigt. Es wurde das Protokoll „PCR clean up“ nach Angaben des Herstellers durchgeführt.

2.1.7 Klonierung von PCR-Produkten

2.1.7.1 pGEM®-T Easy Vector

Für die Klonierung von PCR-Produkten wurde das „pGEM®-T (Easy) Vector System“ der Fa. Promega (Madison, USA) verwendet (Vektorkarte siehe Anhang A1). Die Durchführung erfolgte nach Angaben des Herstellers. Die Integratgröße konnte durch einen Verdau mit den Enzymen *EcoRI* (pGEM®-T Easy Vector) oder *SacI* und *SacII* (pGEM®-T Vector) und anschließender gelelektrophoretischer Auftrennung überprüft werden. Die Sequenzierung erfolgte mit den Primern Sp6 und T7.

2.1.7.2 pCR™ 4-TOPO® Vector

Für große PCR-Produkte, bspw. die Ir-PCR-Produkte, sowie bei schwierig zu klonierenden PCR-Produkten wurde das „TOPO® TA Cloning® Kit for Sequencing“ von Invitrogen™ (Carlsbad, USA) verwendet (Vektorkarte siehe Anhang A1). Die Durchführung erfolgte nach Angaben des Herstellers.

2.1.8 Southern Blot

Beim Southern Blot wird gelelektrophoretisch aufgetrennte DNA auf einer Membran fixiert, um diese anschließend mit markierten Sonden zu hybridisieren und somit spezifisch einzelne DNA-Fragmente nachzuweisen (Southern 1975).

2.1.8.1 Restriktion genomischer DNA

Genomische DNA wurde mit verschiedenen Restriktionsenzymen möglichst vollständig geschnitten. Die enzymatischen Reaktionen wurden in einem Volumen von 200 µl für insgesamt 20 h bei 37 °C durchgeführt. Eine Ausnahme bildeten Ansätze mit *Sau3AI*; diese wurden aufgrund der hohen Aktivität der Restriktionsendonuklease nur 1,5 h bei 37 °C inkubiert. Mittels Testgel wurden die Verdau auf Vollständigkeit überprüft. Abschließend erfolgte eine Ethanol-Fällung.

2.1.8.2 Southern Transfer

Die gefällten Restriktionsansätze wurden elektrophoretisch auf einem Über-Tag-Gel aufgetrennt. Das Gel wurde anschließend mit EtBr angefärbt und dokumentiert. Die Markerbanden wurden mit einem Spatel durchgestochen. Anschließend wurde das Gel in Denaturierungslösung (0,5 M NaOH/1,5 M NaCl) für 15 min bei RT inkubiert. Nach kurzem Wässern wurde es in Neutralisierungslösung (3 M NaCl/0,5 M Tris pH 7,0) überführt und dort für 20 min bei RT inkubiert. Der Transfer der DNA auf Nitrozellulosemembranen (Protran® BA85 der Fa. Whatman®, New Jersey, USA) erfolgte ü.N. uni- oder bidirektional. Beim Sandwichblots (bidirektional) ist das Gel umgeben von in 2 x SSC angefeuchteten Nitrozellulosemembranen, gefolgt von drei Lagen ebenfalls in 2 x SSC angefeuchtetem Whatman 3MM-Papier. Umschlossen werden diese von etwa 5 cm hohen Stapeln an Papierhandtüchern. Eine Plexiglasplatte und ein weiteres Gewicht beschweren die komplette Konstruktion. Beim unidirektionalen Blot liegt das Gel auf Whatman 3MM-Papier, dessen Enden in ein Reservoir an 20 x SSC reichen. Auf dem Gel liegt die Nitrozellulosemembran, gefolgt von vier Lagen in 2 x SSC angefeuchtetem Whatman 3MM-Papier und einem etwa 5 cm hohem Stapel an Papierhandtüchern. Nach erfolgtem Transfer wurden die Nitrozellulosemembranen kurz in 2 x SSC gewaschen. Mit einem Fettstift konnten unter UV-Licht die Probestaschen, die Gelumrisse sowie die Banden des Markers auf der Nitrozellulosemembran markiert werden, ehe diese zum Fixieren der DNA bei 80 °C für 2 h inkubiert wurden.

2.1.8.3 Hybridisierung

Die Hybridisierung des Southern Blots (Meinkoth und Wahl 1984) erfolgte mit einer radioaktiv markierten Sonde (siehe 2.3.1.1). Die Blot-Filter wurden in Präinkubationsmedium nach Denhardt (1966) für 3 h bei 60 °C inkubiert. Die Hybridisierungslösung aus radioaktiv markierter Sonde, „Carrier“-DNA und Hybridisierungspuffer wurde durch Erhitzen denaturiert (10 min, 95 °C) und anschließend rasch in Eiswasser abgekühlt. Die Hybridisierung erfolgte in Glas-Hybridisierungsröhren ü.N. bei 60 °C im Hybridisierungssofen. Nach Abgießen der Hybridisierungslösung wurden die Membranen mit 2 x SSC bei 60 °C für 30 min gewaschen. Dieser Waschschrift wurde zweimal wiederholt. Abschließend wurden die Filter kurz in 2 x SSC bei RT gewaschen, auf Filterpapier getrocknet und auf 3MM-Filterpapier fixiert mit Haushaltsfolie glatt abgedeckt.

2.1.8.4 Autoradiografie

Das radioaktive Isotop ^{32}P emittiert β -Teilchen. Diese bewirken eine Schwärzung eines Röntgenfilms. Die Verwendung einer Verstärkerfolie verkürzt die Expositionszeit bei einer Temperatur von $-80\text{ }^{\circ}\text{C}$. Die Expositionszeit liegt je nach Strahlungsintensität zwischen 2 h und einer Woche. Der Röntgenfilm wurde für die Entwicklung für 4 min in Entwicklerlösung inkubiert, anschließend kurz gewässert und dann für 4 min in Fixierlösung getaucht. Abschließend wurde der Film gewässert und getrocknet.

2.1.9 „Pirota“ Blot

Der „Pirota“ Blot wird auch als reverser Southern Blot bezeichnet und dient der Identifizierung von Sonden mit hoch- bzw. mittelrepetitiven Elementen (Buhariwalla *et al.* 1995). Im Gegensatz zum Southern Blot wird nicht die genomische DNA, sondern die Sonde auf einer Nitrozellulosemembran fixiert. Diese wird anschließend mit radioaktiv markierter genomischer DNA hybridisiert. Die Durchführung entsprach ansonsten der des Southern Blots (siehe 2.1.8).

2.2 Konstruktion einer Apfel-BAC-Bibliothek

2.2.1 Versuchsmaterial

Als Ausgangsmaterial dienten *in-vitro*-Kulturen der kolumnar-wüchsigen Apfelsorte P28. P28 stammt aus Kreuzungen der kolumnaren Sorte Flamenco und der normal-wüchsigen Sorte Topaz. Das Material stellte die Universität Geisenheim zu Verfügung.

2.2.2 Isolation von Zellkernen

Die Isolation der Zellkerne erfolgte nach einem modifizierten Protokoll von L.A. Burgoyne (Burgoyne *et al.* 1970). Zunächst wurde das Ausgangsmaterial (etwa 1,5 g) mittels Skalpell zerkleinert und in 10 ml Homogenisierungspuffer aufgenommen. Dieser wurde zuvor mit 1/10 Volumen 10 %-igem Triton X-100 versetzt. Die Verwendung von Triton X-100 dient dem chemischen Zellaufschluss und bewirkt zudem, dass die isolierten Zellkerne keine äußere Membran mehr besitzen (Blobel und Potter 1966). Der mechanische Zellaufschluss erfolgte in einem Homogenisator für 15 min auf Eis. Anschließend wurde das Gemisch über Gaze filtriert. Der Filterkuchen wurde erneut in 10 ml Homogenisierungspuffer aufgenommen, homogenisiert und filtriert. Die Filtrate wurden vereinigt. In einem Zentrifugenbecher wurde Lösung 1 vorgelegt und mit dem Filtrat überschichtet. Das

Volumenverhältnis betrug 1:1. Es wurde für 20 min bei 4 °C und 14000 Upm in der Kühlzentrifuge 3K30 der Fa. Sigma (Osterode am Harz, Deutschland) zentrifugiert. Der Überstand wurde verworfen und das Pellet in 2,5 ml Lösung 2 resuspendiert. In zwei Ultrazentrifugen-Tubes wurden jeweils 650 µl Lösung 2 vorgelegt und mit der Suspension im Verhältnis 1:2 überschichtet. Es folgte ein Zentrifugationsschritt bei 35000 Upm und 4 °C für 1,5 h in der Ultrazentrifuge Sorvall Discovery M120SE der Fa. Thermo Scientific (Waltham, USA). Die Überstände wurden verworfen und die Pellets in je 300 µl L-Puffer aufgenommen. Bei zeitnaher Verwendung erfolgte die Lagerung bei 4 °C, ansonsten wurden die Suspensionen bei -20 °C gelagert.

Als Alternative zum Homogenisator erfolgte der mechanische Zellaufschluss mit einem Stabmixer. Etwa 30 g Ausgangsmaterial wurden unter Zugabe von 40 ml mit 10%-igem Triton X-100 versetzten Homogenisierungspuffer in einem vorgekühlten 1-Liter-Becherglas auf Eis mittels Stabmixer zerkleinert. Das Gemisch wurde über Gaze filtriert und der Filterkuchen in 40 ml Homogenisierungspuffer aufgenommen, erneut zerkleinert und filtriert. Dieser Schritt wurde ein weiteres Mal wiederholt. Die Filtrate wurden vereinigt. Aufgrund des hohen Volumens (insgesamt 120 ml) wurden die folgenden, oben beschriebenen Zentrifugationsschritte sechs Mal parallel durchgeführt.

Als zweite Alternative wurden die *in-vitro*-Kulturen unter flüssigem Stickstoff mit Mörser und Pistill zerkleinert. Das weitere Vorgehen entspricht dem Ansatz bei Verwendung eines Homogenisators.

2.2.3 Kern-Lyse und Proteinase K-Verdau

Die Zellkern-Suspensionen wurden mit je 80 µl 10 %-iger SDS-Lösung und je einer Spatelspitze Proteinase K versetzt und bei 60 °C für 10 min inkubiert. Anschließend erfolgte eine Inkubation bei 37 °C für 60 min. Die einzelnen Ansätze wurden vereint und ü.N. gegen autoklaviertes VE-Wasser dialysiert, um die Proteinase K zu verdünnen und SDS zu entfernen. Zum Entfernen der Proteine und die damit verbundene Aufreinigung der DNA wurden zwei PCI- und zwei CI-Extraktionen durchgeführt (siehe 2.1.3). 1/250 des Probenvolumens wurde zur Überprüfung auf Qualität und Quantität der DNA auf ein 1 %-iges Über-Tag-Gel aufgetragen.

2.2.4 Restriktion

Mittels Restriktionsendonukleasen kann die DNA an definierten Stellen geschnitten werden. Die DNA sollte jedoch nicht vollständig restringiert werden, da für die Konstruktion

einer BAC-Bibliothek möglichst große DNA-Fragmente benötigt werden. Allerdings muss gewährleistet sein, dass durch den Verdau genügend passende Enden generiert werden, um eine Ligation mit pBeloBAC11 zu ermöglichen. Aus diesem Grund erfolgte die Restriktion in einem großen Volumen bei kurzer Inkubationszeit und geringer Enzymkonzentration. Die Restriktion wurde in Ansätzen zu je 5 ml durchgeführt. Dazu wurden 4,5 ml DNA-Lösung mit 500 µl Restriktionspuffer A und 1 µl des Enzyms *Sau3AI* (1 U/µl, Roche Diagnostics, Mannheim, Deutschland) versetzt. Die Inkubation erfolgte bei 37 °C. Pro Durchgang wurden ein Ansatz mit 1 min, zwei Ansätze mit 1,5 min und ein Ansatz mit 2 min Restriktionsdauer angesetzt. Da das Enzym weder mit EDTA noch mittels Hitze inaktiviert werden kann, wurde im Anschluss eine PCI-Extraktion durchgeführt. 1/250 des Probenvolumens wurden zur Überprüfung der Restriktion auf ein 1 %-iges Über-Tag-Gel aufgetragen. Bei erfolgreicher Restriktion wurden die verschiedenen Ansätze vereint und durch Wasserentzug mittels PEG auf etwa 3 ml eingeeengt.

2.2.5 Größenfraktionierung

Die gelelektrophoretische Auftrennung von DNA ist Grundlage dieser Methode, bei der kleine DNA-Fragmente (< 23 kb) von der gewünschten hochmolekularen DNA abgetrennt werden. Für ein effektives Abtrennen der kleinen Fragmente waren zwei aufeinanderfolgende Größenfraktionierungen erforderlich. Für die erste Größenfraktionierung wurden 0,8 %-ige Über-Nacht-Gele in einem speziellen Natriumphosphatpuffer-System (E-Puffer) verwendet. Die Proben wurden vor dem Auftragen für 5 Minuten auf 65 °C erhitzt, um kleine DNA-Fragmente, die an den großen DNA-Molekülen anhaften, zu lösen. Nach Ende der Gelelektrophorese wurden die an den Rändern aufgetragenen Markerspuren mit einem geringen Teil der präparativen Spur mittels Skalpell abgeschnitten und mit EtBr angefärbt. Unter UV-Licht wurde der gewünschte Bereich (> 23 kb) markiert. Anschließend wurde das Gel wieder zusammengesetzt und der markierte Bereich in der präparativen Spur „blind“ ausgeschnitten. Dadurch verhinderte man ein Anfärben der DNA mit EtBr, sowie eine mögliche Schädigung der DNA durch UV-Strahlung. Der präparative Teil des Gels wurde erst nach dem Ausschneiden der Banden angefärbt und dokumentiert. Die DNA im ausgeschnittenen Agarose-Stück wurde mittels Elektroelution wiedergewonnen (siehe 2.1.2.1). Die eluierte DNA wurde zur zweiten Größenfraktionierung erneut auf ein 0,8 %-iges Über-Nacht-Gel aufgetragen. Die zweite Fraktionierung entsprach in der

Methodik der ersten. Als Alternative zu den Über-Nacht-Gelele wurde das „Horizontal System for Submerged Gel Electrophoresis“ der Fa. GIBCO/BRL, heute Life Technologies (Gaithersburg, USA), verwendet. Die Durchführung entsprach ansonsten der Größenfraktionierung mittels Über-Nacht-Gelele. Ein „Mega“-Gel ersetzte bis zu 25 Über-Nacht-Gelele.

2.2.6 Test zur Überprüfung der Integrat-Enden

Vor der Ligation sollte überprüft werden, ob die Restriktion mit *Sau3AI* erfolgreich war und die überhängenden einzelsträngigen DNA-Enden trotz zweier Größenfraktionierungen intakt waren. Abbildung 2.2 zeigt die palindromische Restriktionsschnittstelle des Enzyms.

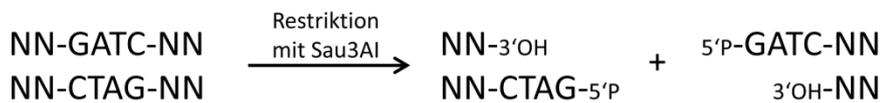


Abbildung 2.2 Restriktionsschnittstelle *Sau3AI*

Um zu überprüfen, ob die Integrat-Enden intakt waren, wurden vier verschiedene Endfilling-Reaktionen angesetzt (Tabelle 2.5). Dabei wurden teilweise radioaktiv markierte Nukleotide verwendet. Nach einer Ethanol-fällung konnte mittels Geiger-Müller-Zähler überprüft werden, ob die radioaktiv markierten Nukleotide eingebaut wurden und die Überhänge demnach intakt waren. Die Reaktion erfolgte bei RT für 30 min, zum Abstoppen wurde für 5 min bei 65 °C inkubiert.

Tabelle 2.5 Endfilling (radioaktiv)

Ansatz A*		Ansatz GA*	
DNA-Lösung	10 µl	DNA-Lösung	10 µl
10 x Klenow-Puffer	3 µl	10 x Klenow-Puffer	3 µl
Klenow-Fragment	2,5 µl	Klenow-Fragment	2,5 µl
HPLC-H ₂ O	12,5 µl	HPLC-H ₂ O	11,5 µl
α- ³² P-dATP	2 µl	dGTP	1 µl
		α- ³² P-dATP	2 µl
Ansatz GATC*		Ansatz GA*TC*	
DNA-Lösung	10 µl	DNA-Lösung	10 µl
10 x Klenow-Puffer	3 µl	10 x Klenow-Puffer	3 µl
Klenow-Fragment	2,5 µl	Klenow-Fragment	2,5 µl
HPLC-H ₂ O	9,5 µl	HPLC-H ₂ O	8,5 µl
dGTP	1 µl	dGTP	1 µl
dATP	1 µl	α- ³² P-dATP	2 µl
dTTP	1 µl	dTTP	1 µl
α- ³² P-dCTP	2 µl	α- ³² P-dCTP	2 µl

Klenow-Puffer und Klenow-Fragment (NEB, Frankfurt am Main, Deutschland)

Nukleotide (Roche Diagnostics, Mannheim, Deutschland)

radioaktiv markierte Nukleotide (Hartmann Analytic, Braunschweig, Deutschland)

2.2.7 Ligation

Für die Ligation wurde das Enzym T4 DNA Ligase (Roche Diagnostics, Mannheim) verwendet. Als Vektor diente mit dem Restriktionsenzym BamHI geschnittener und AP-behandelter pBeloBAC11 (Shizuya *et al.* (1992), Vektorkarte siehe Anhang A1). Der fertig behandelte Vektor wurde von Thomas Herold (2012) bereitgestellt. Die Ligation erfolgte mindestens ü.N. bei RT. Ein Ligationsansatz ist Tabelle 2.6 zu entnehmen.

Tabelle 2.6 Ligationsansatz

DNA-Lösung	15 µl
pBeloBAC11	1 µl
10 x DNA Ligase Puffer	2 µl
T4 DNA Ligase	2 µl
Gesamtvolumen	20 µl
10 x DNA Ligase Puffer u. T4 DNA Ligase (Roche Diagnostics, Mannheim, Deutschland)	

Anschließend wurde der Ligationsansatz gefällt und das Pellet in 20 µl HPLC-H₂O gelöst. Die Nomenklatur der Ligationen (L) ergab sich aus dem Ausgangsmaterial, der Art der Größenfraktionierung (bspw. „M“ für „Mega“-Gel) und einer fortlaufenden Nummer. Beispielsweise erhielt die zweite Ligation aus P28 *in-vitro*-Kulturen mit Größenfraktionierungen über „Mega“-Gele den Namen P28ivML2. Die Nomenklatur der Ligationen wurde auf die daraus resultierende DNA-Bibliothek übertragen.

2.2.8 Transformation mittels Elektroporation

Mit der Methode der Elektroporation können rekombinante Plasmide in elektrokompente Zellen transformiert werden (Dower *et al.* 1988). Für die Transformation wurden 2 µl des Ligationsansatzes zu 40 µl der elektrokompenten Zellen gegeben und in eine gekühlte Küvette (Gene Pulser® Cuvette) mit einer Spaltbreite von 0,1 cm der Fa. Bio-Rad Laboratories (Hercules, USA) überführt. Die Transformation erfolgte bei 1,8 kV im Elektroporator Micro Pulser™ der Fa. Bio-Rad (München, Deutschland) für 3-5 ms. Nach der Transformation wurde umgehend 1 ml vorgewärmtes SOC-Medium (Sambrook *et al.* 1989) zu den Zellen gegeben und diese darin resuspendiert. Es folgte eine Inkubation unter Schütteln in einem 15 ml-Reaktionsgefäß bei 37 °C für 1 h. Im Anschluss wurden pro Agarplatte 250 µl des Transformationsansatzes ausplattiert und für 16 bis 20 h bei 37 °C bebrütet. Die bewachsenen Agarplatten wurden bei 4 °C gelagert.

2.2.9 Herstellung elektrokompetenter Zellen

Für die Herstellung elektrokompetenter Zellen wurden 50 ml LB-Medium mit einer Einzelkolonie des *E. coli*-Stammes DH10B (Grant *et al.* 1990) angeimpft und unter Schütteln bei 37 °C ü.N. inkubiert. Im Anschluss wurden 10 ml der Kultur in 500 ml LB-Medium überimpft (insgesamt 4 Ansätze) und erneut unter Schütteln bei 37 °C inkubiert, bis mithilfe des BioPhotometers der Fa. Eppendorf (Hamburg, Deutschland) eine OD₆₀₀ von 0,71 bis 0,72 ermittelt wurde. War dieser Wert erreicht, wurden die Bakterienkulturen für 45 min auf Eis inkubiert. Es folgte eine Zentrifugation (4 °C, 2500 Upm, 15 min) in vorgekühlten 500 ml PET-Zentrifugenbechern in der Zentrifuge 4K15C der Fa. Sigma (Osterode am Harz, Deutschland). Anschließend wurden die Pellets in je 50 ml sterilem, eiskaltem HPLC-H₂O resuspendiert, jeweils zwei Ansätze vereint und erneut zentrifugiert (4 °C, 2500 Upm, 20 min). Die Pellets wurden in 30 ml sterilem, eiskaltem HPLC-H₂O resuspendiert, in vorgekühlte 80 ml Zentrifugenbecher überführt und in der Kühlzentrifuge 3K30 der Fa. Sigma (Osterode am Harz, Deutschland) zentrifugiert (4 °C, 2500 Upm, 20 min). Die Pellets wurden in je 5 ml eiskaltem 10 %-igem Glycerin resuspendiert und erneut zentrifugiert (s.o.). Der Überstand wurde sorgfältig abgeschüttet und die Pellets im verbleibenden restlichen Überstand resuspendiert und anschließend vereint. Im Anschluss wurden die elektrokompetenten DH10B-Zellen aliquotiert. Hierfür wurden je 40 µl Aliquots in bei -80 °C vorgekühlte 1,5 ml Eppendorfgefäße auf Eis überführt. Die kompetenten Zellen wurden bei -80 °C gelagert.

2.2.10 BAC-Minipräparation

Um die statistische Genomabdeckung einer BAC-Bibliothek zu ermitteln, wurde ein Restriktionsverdau zur Integratgrößenbestimmung einiger zufällig ausgewählter Plasmide durchgeführt. Die BAC-Minipräparation dient zur Isolierung von Plasmiden (modifiziert nach Sheng *et al.*/Whitehead). Die gewünschte Kolonie wurde in 5 ml L-Medium ü.N. bei 37 °C angezogen. Die ü.N.-Kultur wurde in der Kühlzentrifuge 4K15C der Fa. Sigma (Osterode am Harz, Deutschland) bei RT mit 5000 Upm für 20 min zentrifugiert und der Überstand verworfen. Das Pellet wurde in 100 µl kalter Lösung 1 resuspendiert. Nach Zugabe von 1 µl RNaseH (Peqlab, Erlangen, Deutschland) wurde die Lösung in ein 1,5 ml-Eppendorfgefäß überführt und auf Eis für 2 min inkubiert. Anschließend wurden 200 µl Lösung 2 hinzugegeben und 8 bis 10-mal invertiert. Nach Zugabe von 150 µl eiskalter Lösung 3 wurde erneut invertiert und 10 min auf Eis inkubiert. Des Weiteren wurde die Probe für 10 min bei

14000 Upm und RT in der Mikroliterzentrifuge Mikro 200 der Fa. Hettich (Tuttlingen, Deutschland) zentrifugiert. Der Überstand wurde in ein neues 1,5 ml-Eppendorfgefäß überführt und der Zentrifugationsschritt wiederholt. Es folgte eine Fällung mit 300 µl Isopropanol und anschließender Zentrifugation in der Mikroliterzentrifuge Mikro 200 der Fa. Hettich (Tuttlingen, Deutschland) bei RT und 14000 Upm für 15 min. Der Überstand wurde verworfen, das Pellet in 300 µl TE-Puffer gelöst. Nach Zugabe von 150 µl 7,5 M Kaliumacetat-Lösung wurde die Lösung für mindestens 15 min bei -80 °C inkubiert. Die aufgetaute Probe wurde in der Mikroliterzentrifuge Mikro 200 der Fa. Hettich (Tuttlingen, Deutschland) zentrifugiert (RT, 14000 Upm, 10 min). Der Überstand wurde mit 900 µl Ethanol abs. gefällt. Nach abschließender Zentrifugation (s.o.) wurde der Überstand verworfen, das Pellet getrocknet und in 20 µl HPLC-H₂O aufgenommen. Für die Restriktion wurde der in Tabelle 2.7 abgebildete Ansatz pipettiert und bei 37 °C für mindestens 4 h inkubiert. Der Restriktionsansatz wurde komplett auf ein 1 %-iges Über-Tag-Gel aufgetragen.

Tabelle 2.7 Restriktionsansatz BAC-Minipräparation

Plasmid-Lösung	10 µl
10 x RE-Puffer	2 µl
Acetyliertes BSA	0,2 µl
Restriktionsenzym <i>EcoRI</i> (12 U/µl)	1 µl
HPLC-H ₂ O	6,8 µl
Gesamtvolumen	20 µl

RE Puffer, *EcoRI* und BSA (Promega, Madison, USA)

2.3 Screening einer Apfel-BAC-Bank

Das Verfahren der Koloniefilter-Hybridisierung nach Grunstein und Hogness (1975) dient der Identifizierung von Bakterienklonen, die einen Klonierungsvektor mit dem eingebauten, gesuchten DNA-Abschnitt enthalten.

2.3.1 Konstruktion PCR-basierender Sonden

Die Transkriptomdaten der Apfelsorten P28 und A14 des Roche-454- und des Illumina-Laufes von Clemens Krost (2012) dienten als Ausgangssequenzen zur Konstruktion der Sonden in der Zielregion. Durch BLASTn-Suchen gegen das veröffentlichte Apfelgenom (Velasco *et al.* 2010) wurden diejenigen Contigs der Transkriptomdaten identifiziert, die in der Zielregion liegen. Diese Contigs wurden mit dem Programm Spidey (NCBI) auf Exongrenzen untersucht, um möglichst große Exons (> 300 bp) zu finden. Durch das Entwerfen spezifischer PCR-Primer in diesen Exonbereichen konnten „single-copy“-Sonden

erstellt werden. Anhand von Southern und „Pirotta“ Blot-Analysen wurden die PCR-Produkte abschließend auf repetitive Elemente untersucht (siehe Kapitel 2.1.7 u. 2.1.8). In Bereichen der Zielregion, die nicht von den Transkriptomdaten abgedeckt waren, wurden die Sequenzinformationen des veröffentlichten Genoms zur Herstellung PCR-basierender Sonden genutzt.

2.3.1.1 Random Primed Oligo Labeling

Dieses Verfahren beschreibt die Hybridisierung von Oligonukleotiden zufälliger Sequenz an eine einzelsträngige Matrizen-DNA und die anschließende Elongation durch das Klenow-Fragment unter Einbau radioaktiv markierter Nukleotide (Feinberg und Vogelstein 1983). Zur radioaktiven Markierung mit α -³²P-dATP (10 μ Ci/ μ l) wurde das Random Primed DNA Labeling Kit der Fa. Roche Diagnostics (Mannheim, Deutschland) nach Angaben des Herstellers verwendet. Das Labeling erfolgte mit 7 μ l α -³²P-dATP bei 37 °C für 1,5 h. Abschließend wurde die DNA gefällt, der Einbau des radioaktiv markierten Nukleotids mittels Geiger-Müller-Zähler überprüft und das Pellet in 100 μ l HPLC-H₂O gelöst.

2.3.2 Koloniefilter-Hybridisierung

2.3.2.1 Vorbereitung der Filter

Auf die Agarplatten der Transformation wurden Nitrozellulose-Filter (Protran® BA85 der Fa. Whatman®, New Jersey, USA) gelegt und mit einer Impfnadel asymmetrische Markierungen zur späteren Orientierung auf dem Filter angebracht. Die Filter wurden für 7 min auf mit 0,5 M NaOH getränktes Filterpapier gelegt und anschließend getrocknet. Des Weiteren wurden die Filter für 5 min auf mit 1,5 M Tris pH 7,4 getränktes Filterpapier gelegt und danach für 5 min in 0,5 M Tris/1,5 M NaCl pH 7,4 getaucht. Im nächsten Schritt wurden die getrockneten Filter für 30 min bei RT in Proteinase K-Lösung (1 mg/ml in 1 x SSC) inkubiert. Die Filter wurden kurz durch Ethanol abs. gezogen und getrocknet. Es folgte eine Inkubation der Filter für 5 min in 0,3 M NaCl. Durch kräftiges Drücken der nassen Filter zwischen Filterpapier konnten Koloniereste entfernt werden. Abschließend wurde die Plasmid-DNA durch Anbacken bei 80 °C für 2 h auf der Filtermembran fixiert.

2.3.2.2 Hybridisierung

Die Hybridisierung erfolgte wie in Kapitel 2.1.8.3 beschrieben. Allerdings wurden für die Hybridisierung Petrischalen verwendet, die in eine feuchte Kammer gestellt und bei 60 °C ü.N. inkubiert wurden. Es wurde zudem nur ein Waschschrift mit vorgewärmtem 2 x SSC für

5 min durchgeführt. Abschließend wurden die Filter zum Trocknen ausgelegt, auf 3MM-Filterpapier fixiert und mit Haushaltsfolie glatt abgedeckt.

2.3.2.3 Autoradiografie

Die Autoradiografie wurde analog zu den Southern Blots wie in Kapitel 2.1.8.4 beschrieben durchgeführt.

2.3.2.4 Zweite Runde der Koloniefilter-Hybridisierung

Da die Agarplatten oft dicht bewachsen waren, war eine Zuordnung zwischen einem Signal auf dem Autoradiogramm und der entsprechenden Kolonie auf der Agarplatte meist nicht eindeutig. Aus diesem Grund wurden alle Kolonien, welche sich im engeren Umfeld zu einem starken Signal des Autoradiogramms befanden, gepickt und auf zwei gerasterten Nitrozellulose-Filtern identisch ausgestrichen. Die Nitrozellulose-Filter wurden ü.N. bei 37 °C bebrütet. Mit einem der beiden 2. Runde-Filter wurden die Schritte unter 2.3.2.1 – 2.3.2.3 durchgeführt. Als Sonde wurde die der ersten Runde verwendet. Auf den gerasterten Filtern war eine eindeutige Zuordnung zwischen Signal des Autoradiogramms und ausgestrichener Kolonie gewährleistet. Von der unbehandelten Kopie konnten die positiven Kolonien gepickt und angezogen werden.

2.3.2.4 Randsequenzierung

Um die genomische Lage eines BAC-Klons zu bestimmen, wurden die Ränder des Integrates mit den dafür vorgesehenen Primern Sp6 und T7 nach Sanger *et al.* (1977) sequenziert. Die Sequenzierungen wurden von der Fa. GENterprise (Mainz, Deutschland) durchgeführt. Mithilfe des Suchalgorithmus BLASTn (Altschul *et al.* 1990) wurde die genomische Lage der Sequenzen untersucht. Klone, die in der Zielregion lagen, wurden anschließend vollständig sequenziert.

2.4 Sequenzierung relevanter Klone mittels shotgun-Bibliotheken

Diese Methode ermöglichte die Sequenzierung großer BAC-Klone. Dazu wurden die isolierten Plasmide zufällig in kleinere Fragmente zerlegt, welche wiederum kloniert und anschließend sequenziert wurden. Durch die Sequenzierung vieler dieser kleinen Fragmente konnte letztlich die komplette BAC-Sequenz zusammengesetzt werden.

2.3.1 Plasmid-Isolierung nach Whitehead

Das verwendete Verfahren zur Plasmid-Präparation wurde am Whitehead Institute in Massachusetts entwickelt. Zunächst wurde eine 5 ml Über-Tag-Kultur angelegt. Die Über-Tag-Kultur wurde in 500 ml antibiotikahaltiges L-Medium überführt und bei 37 °C unter Schütteln für 16 bis 18 h inkubiert (ü.N.-Kultur). Die ü.N.-Kultur wurde in ein 500 ml-Zentrifugengefäß überführt und in der Kühlzentrifuge 4K15C der Fa. Sigma (Osterode am Harz, Deutschland) zentrifugiert (4 °C, 5000 Upm, 15 min). Der Überstand wurde verworfen und das Pellet in 10 ml Lösung I resuspendiert und in einen 100 ml-Zentrifugenbecher überführt. Nach Inkubation für 5 min bei RT erfolgte die Zugabe von 20 ml Lösung II. Nach gründlichem Invertieren wurde für 10 min bei RT inkubiert. Schließlich wurden 15 ml kalte Lösung III hinzugegeben, invertiert und die Probe für 10 min auf Eis gestellt. Es folgte ein Zentrifugationsschritt in der Kühlzentrifuge 3K30 der Fa. Sigma (Osterode am Harz, Deutschland) bei 4 °C und 14000 Upm für 20 min, welcher mit dem Überstand direkt im Anschluss wiederholt wurde. Für die Isopropanol-Fällung wurde die Probe auf zwei 50 ml-Reaktionsgefäße aufgeteilt und jeweils mit 15 ml Isopropanol versetzt. Nach Zentrifugation in der Kühlzentrifuge 4K15C der Fa. Sigma (Osterode am Harz, Deutschland) bei RT mit 5000 Upm für 20 min wurden die Überstände verworfen und die Pellets in jeweils 3 ml TE-Puffer resuspendiert. Nach Zugabe von jeweils 1,5 ml 7,5 M Kaliumacetat wurden die Proben vereint und bei -80 °C für mindestens 45 min inkubiert. Die aufgetaute Probe wurde erneut zentrifugiert (s. o.), ehe der Überstand einer Ethanol-fällung unterzogen wurde. Das Pellet wurde in 700 µl 50T/50E-Puffer resuspendiert. Für den anschließenden RNase-Verdau wurde 1 µl RNaseH (Peqlab, Erlangen, Deutschland) hinzugegeben und für 45 min bei 37 °C inkubiert. Es folgte eine PCI- und eine CI-Extraktion (siehe 2.1.3). Die obere Phase wurde mit 700 µl Isopropanol gefällt und in 200 µl HPLC-H₂O aufgenommen.

2.3.2 Nebulisieren der Plasmid-DNA und Endfilling

Das Scheren von 50 µg Plasmid-DNA erfolgte in 2 ml TE-Puffer für 25 sec bei 1 bar in einem Nebulizer der Fa. Invitrogen™ (Carlsbad, USA). Abschließend wurde die gescherte DNA-Lösung gefällt. Mittels Endfilling wurden die Enden der nebulisierten DNA-Fragmente aufgefüllt, sodass einheitliche „glatte“ Enden entstehen. Der Endfilling-Ansatz ist Tabelle 2.8 zu entnehmen. Die Reaktion erfolgte bei 37 °C für 30 min.

Tabelle 2.8 Endfilling-Ansatz

Nebulisierte DNA-Lösung	31,5 µl
5 x T4-DNA Polymerase Puffer	10 µl
dNTPs (10 mM each)	1 µl
BSA (100 x)	0,5 µl
Klenow Fragment (2 U/µl)	3 µl
T4-DNA Polymerase (1 U/µl)	4 µl
Gesamtvolumen	50 µl
alle Reagenzien (Roche Diagnostics, Mannheim, Deutschland)	

2.3.3 Größenfraktionierung und Elektroelution

Nach erfolgreichem Endfilling wurde die DNA über ein präparatives „Midi“-Gel aufgetrennt. Es wurden zwei Fraktionen ausgeschnitten (1-2 kb und 2-4 kb) und die DNA mittels Elektroelution (siehe 2.1.2.1) wiedergewonnen und gefällt. Mittels Testgel wurde die größenfraktionierte DNA überprüft.

2.3.4 Ligation und Transformation

Die größenfraktionierte DNA wurde in *Sma*I-restringierten, AP-behandelten pUC19 kloniert. Pro Fraktion wurde eine Ligation angesetzt. Die Reaktion erfolgte für 60 h bei RT nach dem in Tabelle 2.9 zu entnehmenden Ansatz.

Tabelle 2.9 Ligations-Ansatz

Größenfraktionierte DNA-Lösung	6 µl
10 x Ligase Puffer	2 µl
pUC19 <i>Sma</i> I AP	1 µl
HPLC-H ₂ O	9 µl
T4-DNA Ligase (1 U/µl)	2 µl
Gesamtvolumen	50 µl
alle Reagenzien (Roche Diagnostics, Mannheim, Deutschland)	

Nach erfolgter Ligation wurden die Ansätze gefällt. Die Transformation erfolgte wie unter 2.2.8 beschrieben. Pro Ansatz wurden zweimal 100 µl und dreimal 10 µl des Transformationsansatzes auf Agarplatten ausplattiert.

2.3.5 Überprüfung der shotgun-Bibliothek und Sequenzierung

Pro Transformationsansatz wurden 8 Klone gepickt und die Plasmide mit dem Kit „peqGOLD Plasmid Miniprep Kit I“ der Fa. Peqlab (Erlangen, Deutschland) nach Angaben des Herstellers isoliert. Die isolierten Plasmide wurden in 50 µl HPLC-H₂O eluiert. Durch einen Restriktionsverdau und anschließende gelelektrophoretische Auftrennung wurde die

Integratgröße der Plasmide überprüft. Der Ansatz ist Tabelle 2.10 zu entnehmen. Die Restriktion erfolgte für 3 h bei 37 °C.

Tabelle 2.10 Restriktionsansatz pUC

Plasmid-Lösung	5 µl
10 x Restriktionspuffer R+	2 µl
HPLC-H ₂ O	11 µl
Restriktionsenzym <i>HindIII</i> (10 U/µl)	1 µl
Restriktionsenzym <i>EcoRI</i> (12 U/µl)	1 µl
Gesamtvolumen	20 µl

alle Reagenzien (Fermentas, St. Leon-Rot, Deutschland)

Enthielten mindestens 12 der 16 untersuchten Klone ein Integrat in erwarteter Größe, wurden pro Transformationsansatz 192, insgesamt folglich 384 Klone gepickt. Die Anzucht der Klone erfolgte in 96er-„Deep-Well“-Platten der Fa. ABgene (Epsom, UK). Hierfür wurde pro Klon 1 ml ampicillinhaltiges 2 x LB Medium vorgelegt. Die Klone wurden für 18 h bei 37 °C bebrütet. Die Plasmid-Isolierung und anschließende Sequenzierung wurden von der Fa. GENterprise (Mainz, Deutschland) durchgeführt.

2.3.6 Auswertung von DNA-Sequenzen

Mit den von der Fa. GENterprise (Mainz, Deutschland) erhaltenen Sequenzdaten sollten die BAC-Klone vollständig zusammengesetzt werden. Die Editierung der Sequenzdaten (EditSeq™) und die Erstellung eines Sequenz-Alignments (SeqMan™) wurden mithilfe des Software-Pakets Lasergene® der Fa. DNASTAR Inc. (Madison, USA) durchgeführt. Einzelne Chromatogramme wurden mit dem Programm FinchTV v1.4.0 der Fa. Geospiza Inc. (Seattle, USA) visuell kontrolliert. Traten bei den Alignments Lücken oder schlecht abgedeckte Bereiche auf, so wurden diese mit speziell entworfenen Primern nochmals sequenziert. Abschließend wurde der Vektoranteil des BAC-Klons entfernt, sodass die reine Integratsequenz vorlag.

2.5 Sequenzierung von BAC-Klonen mittels Illumina Next-Generation Sequencing

Sollen mehrere BAC-Klone sequenziert werden, ist die Illumina-Sequenzierung eine kostengünstigere Alternative zur Sequenzierung mittels shotgun-DNA-Bibliotheken.

Da für die Illumina-Sequenzierung DNA-Mengen von 1 µg ausreichen, erfolgte die Plasmid-Isolierung nach Promega. Die Erstellung der „paired end“-Bibliotheken wurde von der Fa. GENterprise (Mainz, Deutschland) durchgeführt. Die Sequenzierung erfolgte mittels eines

100 bp „paired end“-Laufs auf dem Illumina HiSeq 2000 durch das Nukleinsäure-Analyse-Zentrum (IMSB, Mainz, Deutschland).

2.5.1 Plasmid-Isolierung nach Promega

Zur Plasmid-Isolierung wurde das Kit „PureYield Plasmid Maxiprep System“ der Fa. Promega (Madison, USA) verwendet. Die Durchführung erfolgte nach Angaben des Herstellers. Da es sich bei den zu isolierenden Plasmiden um BAC-Konstrukte handelte, wurde die Menge des Ausgangsmaterials von im Protokoll angegebenen 300 ml auf 450 ml erhöht.

2.5.2 Auswertung der Illumina Next-Generation Sequencing Sequenzdaten

2.5.2.1 Aufbereitung und Filterung der Rohdaten

Die Aufbereitung der Rohdaten wurde mit zahlreichen Perl-Skripten umgesetzt. Diese wurden von Benjamin Rieger (IMSB) in Zusammenarbeit mit Dr. Steffen Rapp (IMSB) geschrieben. Aus Tabelle 2.11 sind die verwendeten Perl-Skripte mit einer kurzen Beschreibung zu entnehmen.

Tabelle 2.11 Verwendete Perl-Skripte

qseq2fastq.pl	Umwandlung von „qseq“-Dateien in das „fastq“-Format bei gleichzeitiger Filterung nach „Chastity“ und Sequenzen mit mehr als 3 uneindeutigen Basen („N“)
fastq_integrity.pl	Aussortieren von Sequenzen, deren Länge des Sequenzstrings mit der Länge des Qualitystrings nicht übereinstimmt
fastq_splitup.pl	Aufteilung der Sequenzen einer „fastq“-Datei in Datenpakete einstellbarer Größe
fasta_check_iupac.pl	Ausgabe der Positionen von IUPAC-Basen
fasta_extract.pl	Extraktion von „fasta“-Sequenzen anhand einer Liste mit „fasta“-IDs.
seq_info.pl	Ausgabe von: Sequenzanzahl, Basenzahl, min. und max. Länge, Basenverteilung, n-stats und quality stats (nur für „fastq“-Dateien)
fasta_n_stats.pl	Ausgabe der Positionen uneindeutiger Basen

alle Skripte: Benjamin Rieger und Steffen Rapp (IMSB)

Die aufbereiteten Rohdaten wurden mit der Funktion „NGS Import“ in die CLC Genomics Workbench (CLC bio, Aarhus, Denmark) als „paired end“-Sequenzen eingeladen und mittels „Trim Sequences“ gefiltert. Sequenzen mit schlechter Qualität ($\leq 0,005$) und nicht eindeutigen Basen (N) wurden aussortiert. Zudem wurden am 5'-Ende jeder Sequenz 5 Basen, am 3'-Ende jeder Sequenz 20 Basen entfernt.

2.5.2.2 Assemblierung der aufbereiteten Rohdaten

Mit der Funktion „de novo assembly“ der CLC Genomics Workbench wurden die aufbereiteten und getrimmten Datenpakete assembliert. Im Anschluss wurden die erhaltenen Contigs auf Vektor- und Randsequenzen überprüft und zudem untersucht, ob

ein Ringschluss vorliegt. Trafen diese drei Kriterien auf einen oder mehrere Contigs zu, waren dies Kandidaten für einen korrekt assemblierten Klon, welche, wie unter 2.5.3 beschrieben, weiter überprüft wurden. Waren ein oder mehrere der Kriterien nicht erfüllt, wurden die erhaltenen Contigs mittels NGEN2 (DNASTAR, Madison, USA) erneut assembliert, um weitere Kandidaten zu erhalten. Parallel wurden die Rohdaten mithilfe der Assembly Cell (CLC bio, Aarhus, Denmark) assembliert. Die verwendeten Parameter sind Tabelle 2.12 zu entnehmen.

Tabelle 2.12 Parameter der Assembly Cell

-w	k-mer size (variiert von 25 - 31)
-b	bubble size (variiert von 200 - 300 in 10er Schritten)
-p	paired end info : fb ss 200 500 (forward-backward, start-start, 200-500 distance)
-cpus	Anzahl der Kerne für threads (variiert 1-16)
-q -i	paired end files (file1 file2)

Die aus den unterschiedlichen Assemblierungen erhaltenen Kandidaten wurden mittels NGEN2 (DNASTAR, Madison, USA) erneut assembliert, um doppelte Kandidaten auszusortieren. Auf diese Weise wurde für jeden BAC-Klon ein finaler Contig erstellt.

2.5.3 Überprüfung der Kandidaten der BAC-Klone

Die assemblierten Klone wurden weiteren Tests unterzogen, um sie auf Richtigkeit zu überprüfen.

2.5.3.1 Mapping der aufbereiteten Rohdaten

Zunächst wurden die aufbereiteten Rohdaten mit der „Map Reads to Reference“-Funktion der CLC Genomics Workbench gegen die zusammengesetzte BAC-Sequenz kartiert. Signifikant stärker und schwächer abgedeckte Bereiche wurden mit Primer flankiert, um sie anschließend mittels Sanger-Sequenzierung zu überprüfen.

2.5.3.2 Restriktionskartierung

Nach erfolgreichem Mapping wurden Restriktionskartierungen durchgeführt, um die BAC-Sequenzen abschließen zu verifizieren. Mithilfe der „Restriction Site Analysis“-Funktion der CLC Main Workbench wurden Restriktionsenzyme für jeden BAC-Klon ausgewählt, sodass bei dem anstehenden Verdau bevorzugt Restriktionsfragmente mit einer Größe unter 10 kb entstanden. Sollten größere Fragmente entstehen, wurden, wenn möglich, weitere Enzyme verwendet, die in diesen Bereichen ihre Erkennungssequenz besaßen. Für jeden Klon wurden drei bis vier Restriktionsenzyme bzw. Enzymkombinationen ausgewählt. Die verwendeten Enzyme sind Tabelle 2.13 zu entnehmen. Die Restriktionen wurden mit ungefähr 600 ng Plasmid bei 37 °C ü.N. durchgeführt. Anschließend wurden die

Restriktionsansätze gefällt und in 10 µl HPLC-H₂O aufgenommen. Die gelelektrophoretische Auftrennung erfolgte über 1 %-ige Über-Tag-Gele. Anschließend wurden die mittels CLC Main Workbench ermittelten Fragmentgrößen mit den tatsächlichen Fragmentgrößen verglichen. Abweichende Bereiche wurden mittels Sanger-Sequenzierung überprüft.

Tabelle 2.13 Restriktionsenzyme und Puffer

Enzym	Puffer	Enzym	Puffer
<i>ApaI</i> ²	Buffer A ²	<i>PstI</i> ⁵	H Buffer ⁵
<i>AgeI</i> ⁴	NEBuffer 1 ⁴	<i>PvuI</i> ^{2,3}	Buffer M ² , One-Phor-All Buffer Plus ³
<i>BamHI</i> ²	Buffer B ²	<i>PvuII</i> ²	Buffer M ²
<i>BclI</i> ⁴	NEBuffer 3 ⁴	<i>SacI</i> ³	One-Phor-All Buffer Plus ³
<i>BglII</i> ³	One-Phor-All Buffer Plus ³	<i>SalI</i> ²	Buffer H ²
<i>Clal</i> ⁴	NEBuffer 4 ⁴	<i>SmaI</i> ^{2,3}	Buffer A ² , One-Phor-All Buffer Plus ³
<i>EcoRI</i> ¹	Buffer R ⁺¹	<i>XbaI</i> ³	One-Phor-All Buffer Plus ³
<i>EcoRV</i> ¹	Buffer R ⁺¹	<i>XhoI</i> ^{2,3}	Buffer B ² , One-Phor-All Buffer Plus ³
<i>HindIII</i> ^{1,2}	Puffer R ⁺¹ , Buffer B ²	<i>XmaI</i> ⁴	NEBuffer 4 ⁴
<i>HpaI</i> ¹	Buffer B ¹		

¹ Fermentas, St. Leon-Rot, Deutschland

² Roche Diagnostics, Mannheim, Deutschland

³ ehemals Pharmacia, heute GE Healthcare, Little Chalfont, UK

⁴ NEB, Ipswich, USA

⁵ ehemals Amersham, heute GE Healthcare, Little Chalfont, UK

2.6 Illumina Next-Generation Sequencing genomischer Apfel-DNA

Zur Überprüfung der sequenzierten BAC-Klone und als Ergänzung zu diesen, wurden „mate pair“-Bibliotheken der Apfelsorte P28 erstellt.

2.6.1 Erstellung einer „mate pair“-Bibliothek

Die Erstellung der „mate pair“-Bibliothek aus genomischer DNA der Apfelsorte P28 wurde mit einigen Modifikationen nach dem Illumina-Protokoll „Mate Pair Library v2 - Sample Preparation Guide“ (Illumina, San Diego, USA) durchgeführt.

Fragmentierung genomischer DNA

Die Isolierung hochmolekularer DNA wurde wie in Kapitel 2.2 „Konstruktion einer Apfel-BAC-Bibliothek“ durchgeführt. Allerdings wurde die DNA nach Aufreinigung durch 3 PCI- und 2 CI-Extraktionen mit 2 Volumen Ethanol abs. überschichtet und mittels Glasstab aufgewickelt. Nach einem Waschschrift mit 70 %-igem Ethanol wurde sie in HPLC-H₂O abgewickelt. Die DNA-Konzentration wurde mittels Nanodrop ND-1000 (PeqLab, Erlangen, Deutschland) bestimmt. Anstelle der laut Hersteller vorgeschlagenen DNA-Menge von 10 µg wurden etwa 14 µg eingesetzt. 10 mM TE-Puffer diente als Nebulisierungspuffer. Die Fragmentierung der DNA wurde mittels Nebulizer bei 0,5 bar für 12 sec durchgeführt. Für

die anschließende Aufreinigung wurde das „Wizard SV Gel and PCR Clean-Up System“ der Fa. Promega (Madison, USA) nach Angaben des Herstellers verwendet. Die DNA wurde mit auf 60 °C vorgewärmtem QIAGEN EB Buffer (QIAGEN, Hilden, Deutschland) eluiert.

End Repair und Biotinylierung

Die Durchführung geschah nach Protokoll. Die abschließende Aufreinigung wurde wie oben beschrieben durchgeführt.

Größenfraktionierung fragmentierter DNA

Die Größenfraktionierung wurde mittels 0,8 %-igem TBE-„Maxi“-Gel durchgeführt. Die ausgeschnittenen Bereiche entsprachen einer Größe von 3-3,5 kb, 3,5-4 kb, 4-5 kb, 5-6 kb, 7-8 kb und 8-10 kb. Die Wiedergewinnung erfolgte mittels des „Wizard SV Gel and PCR Clean-Up System“ der Fa. Promega (Madison, USA). Die DNA wurde mit zweimal 50 µl auf 60 °C vorgewärmtem QIAGEN EB Buffer (QIAGEN, Hilden, Deutschland) eluiert. Die eluierte DNA wurde anschließend mithilfe des Qubit Fluorometers (Invitrogen™, Carlsbad, USA) und des Agilent 2100 Bioanalyzers („High Sensitivity DNA Assay“, Agilent, Santa Clara, USA) quantifiziert und auf Qualität überprüft.

Zirkularisierung größenfraktionierter DNA

Die Zirkularisierung wurde nach Protokoll durchgeführt. Allerdings wurde der Ansatz an die verfügbare DNA-Menge angepasst. Für die Ansätze mit einer DNA-Menge von 600 ng wurde der im Protokoll angegebene Ansatz verwendet. Für eine DNA-Menge von 400 ng wurde ein Ansatz mit 200 µl Volumen gewählt.

Verdau linearer DNA

Die Durchführung erfolgte nach Protokoll, lediglich das zugegebene Volumen der Reagenzien wurde dem Probenvolumen angepasst.

Scheren zirkularisierter DNA

Die Fragmentierung wurde mit dem Covaris S2 (Covaris, Woburn, USA) in den „MicroTubes AFA Fiber with Snap-Cap“ durchgeführt. Die ausgewählten Parameter waren: Frequency sweeping, Intensity 4, Duty Cycle 10 %, Burst per sec 200 und Time 55 sec bei 4 °C. Da das Probenvolumen der verwendeten Tubes bei maximal 130 µl lag, wurden die Proben auf ein Vielfaches davon aufgefüllt und in mehreren Schritten fraktioniert.

Aufreinigung biotinylierter DNA

Das Vorgehen entsprach dem des Protokolls. Als „magnetic beads“ wurden 20 µl der „Dynabeads M-280 Streptavidin“ der Fa. Invitrogen™ (Carlsbad, USA) verwendet.

End Repair und A-Tailing

Die Durchführung entsprach den Angaben des Protokolls.

Ligation der Adapter

Im Gegensatz zu den Angaben des Protokolls wurden TruSeq-Adapter (Illumina, San Diego, USA) verwendet. Für die Durchführung ergab sich daraus jedoch keine Veränderung.

Anreicherung der Adapter-modifizierten DNA Fragmente durch PCR

Die Primer für die PCR wurden den verwendeten Adaptern angepasst. Es wurden die Enrichment-Primer P1 und P2 (Illumina, San Diego, USA) verwendet. Die PCR wurde in einem TProfessional Thermocycler (Biometra, Göttingen, Deutschland) nach Protokoll durchgeführt. Lediglich die Annealing-Temperatur wurde auf 60 °C angepasst. Nach erfolgter PCR wurden die Proben mittels des Agilent 2100 Bioanalyzers („High Sensitivity DNA Assay“, Agilent, Santa Clara, USA) überprüft.

Größenselektion der Bibliothek

Zur Größenfraktionierung wurden 1,5 %-ige TBE-Midi-Gele verwendet. Der Bereich von 300 bis 650 bp wurde jeweils ausgeschnitten und die DNA nach Protokoll wiedergewonnen. Abschließend diente ein „High Sensitivity DNA Assay“ des Agilent 2100 Bioanalyzers (Agilent, Santa Clara, USA) zur Quantifizierung und Qualitätsüberprüfung der „mate pair“-Bibliothek. Bis zur Sequenzierung wurde diese bei -20 °C gelagert.

2.6.2 Sequenzierung und Filterung der Rohdaten

Die Sequenzierung der „mate pair“-Bibliotheken erfolgte durch das Nukleinsäure-Analyse-Zentrum (IMSB, Mainz, Deutschland) mittels 100 bp „paired-end“-Läufen auf dem Illumina®HiSeq 2000. Die erhaltenen Rohdaten wurden mithilfe des Perl-Skriptes „seq_filter“ (Benjamin Rieger, IMSB) bereinigt. Das Skript entfernt Adapter-Sequenzen. Zudem werden am 5'-Ende einer jeden Sequenz 6 bp, am 3'-Ende einer Sequenz 5 bp entfernt. Es werden darüber hinaus Basen entfernt, die einen Phred-Wert unter 20 aufweisen, was einer Richtigkeit der ermittelten Base von unter 99 % entspricht. Sequenzen, die nach erfolgtem Bereinigen eine Länge unter 30 bp aufweisen, wurden aussortiert. Durch die Filterung der Rohdaten treten „single end reads“ auf, also Sequenzen, deren Partner aussortiert wurde. Diese wurden für die weitere Analyse verworfen.

2.7 Erstellung eines Gesamtcontigs der Co-Zielregion

2.7.1 Erstellung von Metacontigs

Die Sequenzen der BAC-Klone wurden mit dem Programm SeqMan™ des Software-Pakets Lasergene® der Fa. DNASTAR Inc. (Madison, USA) assembliert, um überlappende Klone zu Metacontigs zusammenzuführen. Anhand der selbst erstellten Marker (siehe 2.9) konnten die Metacontigs bzw. BAC-Klone dem kolumnaren und dem nicht kolumnaren Chromosom zugeordnet werden.

2.7.2 Konstruktion eines Gesamtcontigs

Zur Erstellung eines Gesamtcontigs wurden die Metacontigs assembliert. Auftretende Lücken zwischen Metacontigs wurden mit „long range“ PCR und anschließender Sequenzierung geschlossen. Der Gesamtcontig ist eine Consensus-Sequenz, wobei keine Unterscheidung zwischen kolumnaren und nicht kolumnaren Chromosom erfolgte.

2.8 Annotation von Genen und transposablen Elementen

2.8.1 Annotation von Genen

Aus der GDR-Datenbank wurden die im GD-Genom annotierten „Coding Sequences“ (CDS) heruntergeladen. Es handelt sich dabei um Konsensus CDS, die mit den Programmen Glimmer, FGENESH, Genewise, GMAP und Twinscan vorhergesagt wurden (Velasco *et al.* 2010). Durch ein Mapping sollten die CDS herausgefiltert werden, die mit dem erstellten Gesamtcontig übereinstimmen. Ein Problem für ein stringentes Mapping stellen IUPAC-Basen dar, die sich sowohl im Gesamtcontig, vor allem aber in den CDS befinden. Um dies zu umgehen, wurden mithilfe des Perl-Skriptes „fasta_kmer“ der Gesamtcontig in k-mers einer Länge von 100 bp und die CDS in k-mers einer Länge von 30 bp zerlegt. Mittels „fasta_iupac-permut“ wurden die IUPAC-Basen permutiert, das bedeutet, es wurde jede IUPAC-Base durch jede mögliche Base ersetzt. Da Bereiche mit mehreren aufeinanderfolgenden „Ns“ zu 4^n Möglichkeiten und damit verbunden zu einer enormen Menge an Daten führen würden, wurden diese Bereiche mittels „fasta_ncut“ herausgeschnitten und die verbleibenden Teilstücke als „eigenständige“ CDS behandelt. Die auf diese Weise erstellten k-mers der CDS wurden mittels Bowtie (Langmead *et al.* 2009) mit einer Stringenz von 100 % gegen die k-mers des Gesamtcontigs aligniert. Abschließend wurde quantifiziert, wie viele k-mers einer jeden CDS im Gesamtcontig treffen, wobei jedes k-mer auch bei mehreren Treffern nur einmal gezählt wurde. Für diese Berechnung wurde die

ursprüngliche Anzahl an k-mers verwendet und nicht die Anzahl nach der Permutierung. Es wurden die CDS extrahiert, bei denen mindestens 50 % der k-mers in dem Gesamtcontig trafen. Die Erstellung und Verwendung der Perl-Skripte sowie die Verwendung von Bowtie wurde von Benjamin Rieger (IMSB) durchgeführt. Die extrahierten CDS wurden durch das „Large Gap Mapping“ Plug-in der CLC Genomics Workbench mit Standardeinstellungen gegen den Gesamtcontig kartiert. Anhand des Mappings wurden die CDS mithilfe des Programms SeqBuilder™ der Fa. DNASTAR Inc. (Madison, USA) annotiert.

Zur Annotation weiterer Gene sowie zur Überprüfung der vorhergesagten CDS wurden zwei Transkriptomdatensätze des Sprossapikalmeristems der Apfelsorte P28 aus der Doktorarbeit von Clemens Krost (2012) gegen den Gesamtcontig mithilfe des „Large Gap Mapping“ Plug-ins der CLC Genomics Workbench kartiert.

2.8.2 Annotation von transposablen Elementen

Zur Identifikation transposabler Elemente wurde das Programm CENSOR (Kohany *et al.* 2006) genutzt, sowie für die genauere Charakterisierung von Vollängen LTR-Retrotransposons das Programm LTR-Finder (Xu und Wang 2007). CENSOR ist ein BLAST-Programm, wobei als Datenbank die „Rebase Update“ (Jurka *et al.* 2005) dient, eine Datenbank eukaryotischer repetitiver Elemente. Zur Identifizierung von LTRs und TSDs via Dotplot wurde das Programm MegAlign™ des Software-Pakets Lasergene® der Fa. DNASTAR Inc. (Madison, USA) verwendet. Die Annotationen wurden mit den Programmen SeqBuilder™ und CLC Genomics Workbench durchgeführt.

2.9 Herstellung Co-Gen gekoppelter Marker

2.9.1 Versuchsmaterial

Als Versuchsmaterial dienten genomische DNAs der normal-wüchsigen Apfelsorte A14 und der kolumnar-wüchsigen Apfelsorte P28 sowie genomische DNAs der Nachkommenschaft aus Kreuzungen zwischen A14 und P28. P28 stammt aus Kreuzungen der kolumnaren Sorte Flamenco und der normal-wüchsigen Sorte Topaz; A14 aus der kolumnar-wüchsigen Sorte Waltz mit dem normal-wüchsigen Kreuzungspartner GD. Des Weiteren stellte die Forschungsanstalt Geisenheim genomische DNAs der kolumnaren Apfelsorten Procats 1, Procats 4, Procats 11, Procats 13, Procats 27, Greencats, Pomforyou, Pomfital, Kordonia, A10-28, A68-173, A73-19-93K, McIntosh Wijcik, Pomgold und HL 4 K, sowie der nicht kolumnaren Apfelsorten McIntosh, Elswout, Gala, Jonagold, Pinova und Topaz zur

Verfügung. Die Apfelsorte A73-19-93K ist im Gegensatz zu den anderen kolumnaren Apfelsorten homozygot für *Co*.

2.9.2 Primerdesign für PCR-basierende Marker

Die Identifizierung von Insertionen und Deletionen (Indels) durch Sequenzvergleiche zwischen den BAC-Sequenzen der kolumnar-wüchsigen Apfelsorte P28 und Sequenzen der veröffentlichten normal-wüchsigen Apfelsorte GD (Velasco *et al.* 2010) sowie Chromosomen-spezifische Unterschiede zwischen den BAC-Sequenzen waren Grundlage für die Herstellung geeigneter Marker. Die Sequenzvergleiche wurden mit dem Software-Paket Lasergene® der Fa. DNASTAR Inc. (Madison, USA) durchgeführt. Die Primer für entsprechende Marker wurden so gewählt, dass das erwartete PCR-Produkt möglichst eine Größe unter 300 bp aufwies, damit auch Polymorphismen geringer Nukleotidanzahl gelelektrophoretisch detektiert werden konnten. Für die elektrophoretische Auftrennung wurde ein 4 %-iges Agarose-Gel verwendet.

2.9.3 Testen von Kandidaten für Co-Gen gekoppelter Marker

Die entworfenen Primer wurden zunächst auf genomische DNA der Apfelsorten A14 und P28 angewandt. Waren eindeutige Unterschiede zwischen den Eltern erkennbar, wurde das Primerpaar auf 95 Einzelnachkommen (48 kolumnar, 47 nicht kolumnar) überprüft. Abschließend erfolgten Tests des Markers auf genomische DNA weiterer Apfelsorten, um eine mögliche Sortenunabhängigkeit zu untersuchen.

2.9.4 Klonierung Co-Gen gekoppelter Marker und Sequenzierung

Abschließend wurden die Marker (PCR-Produkte) in pGEM®-T Easy kloniert (siehe 2.1.7) und mit den Primern T7 und Sp6 sequenziert. Die Sequenzierung wurde von der Fa. GENterprise (Mainz, Deutschland) durchgeführt. Anhand der Sequenzen konnte überprüft werden, ob der hergestellte Marker tatsächlich von dem untersuchten Locus auf Chromosom 10 stammt.

2.10 Standardlösungen und Materialien

Wenn nicht anders beschrieben, wurden die Lösungen mit VE-Wasser hergestellt.

Agarosegel (1 %)	1 % (w/v) Agarose in TBE-Puffer (1 x)
Agar-Platten	7 g Agar-Agar 5 ml Ampicillin-Lösung (10 mg/ml) oder 5 ml Chloramphenicol-Lösung (1,5 mg/ml) 25 mg IPTG und 50 mg X-Gal in 1 ml DMF in 500 ml 1x LB-Medium
DNA-Ladepuffer	4 M Harnstoff 0,1 M Na ₂ EDTA 50 % (w/v) Saccharose 0,1% (w/v) BPB
Denaturierungspuffer (Southern Blot)	50 mM NaOH 1,5 M NaCl
Dialysepuffer (10 x)	3 M NaCl 0,25 M Tris-HCl 1 M Na ₂ EDTA
Elektrophoresepuffer (10 x) (E-Puffer)	0,36 M Tris 0,3 M NaOH 0,1 M Na ₂ EDTA
EtBr-Färbelösung	0,001 % (v/v) EtBr-Stammlösung in 1 x E-Puffer
EtBr-Stammlösung	0,5 % (w/v) EtBr in 1x E-Puffer
Homogenisierungspuffer (500 ml)	58 g Saccharose 10 ml 0,1 M Na ₂ EDTA, pH 7,4 2,5 ml 0,1 M Na ₂ EGTA, pH 7,4 50 ml Puffer A
Hybridisierungspuffer	1 x PM (in 3 x SSC) 0,5 % SDS
LB-Medium	0,005 % (w/v) NaCl 0,005 % (w/v) Hefe-Extrakt 0,01 % (w/v) Trypton
Lösung 1 (Chromatingewinnung, 250 ml)	117,5 g Saccharose 2,5 ml 0,1 M Na ₂ EDTA 2,5 ml 0,1 M Na ₂ EGTA 25 ml Puffer A
Lösung 2 (Chromatingewinnung, 100 ml)	71,88 g Saccharose 100 µl 0,1 M Na ₂ EDTA 100 µl 0,1 M Na ₂ EGTA 10 ml Puffer A
Lösung I (Whitehead), pH 8	10 mM Na ₂ EDTA
Lösung I, pH 8 (BAC Plasmid-Minipräparation)	25 mM Tris 10 mM Na ₂ EDTA
Lösung II (Whitehead/ BAC Plasmid-Minipräparation)	0,2 M NaOH 1 % (w/v) SDS

MATERIAL UND METHODE

Lösung III, pH 5,2 (Whitehead/ 96er Plasmid-Präparation)	1,875 M Kaliumacetat
Lösung III, pH 5,2 (BAC Plasmid-Minipräparation)	4,5 M Kaliumacetat
L-Puffer	0,1 M Na ₂ EDTA 0,01 M Tris, pH 7,6 0,02 M NaCl
Neutralisierungslösung (Southern Blot)	0,5 M Tris/HCl, pH 7,4 3 M NaCl
PBS/BSA	10 mM MgCl ₂ 10 mg/ml BSA in 1 x PBS
Phenol/Chloroform/ Isoamylalkohol (PCI) (25:24:1)	50 % 50 mM Phenol 48 % Chloroform 2 % Isoamylalkohol
Phosphate Buffered Saline (10 x) (PBS)	1,37 M NaCl 30 mM KCl 10 mM Kaliumhydrogenphosphat 0,1 M Dinatriumhydrogenphosphat
Präinkubationsmedium (PM) (nach DENHARDT 1966)	0,02 % (w/v) Ficoll 400 0,02 % (w/v) PVP 0,02 % (w/v) BSA in 3 x SSC
Puffer A (10 x), pH 7,4	0,6 M KCl 0,15 M NaCl 1,5 mM Spermin 5 mM Spermidin 0,15 M Tris-HCl
Saccharose-Lösung (30 %-ig) (250 ml)	75 g Saccharose in 1 x TBE 0,1 % (w/v) SDS
Standard-Saline-Citrate-Puffer (20x SSC)	3 M NaCl 0,3 M Natriumcitrat
Tris-Borat-EDTA-Puffer (1 x TBE)	90 mM Tris 90 mM Borsäure 25 mM Na ₂ EDTA
TE-Puffer, pH 8,0 (Whitehead)	10 mM Tris 50 mM Na ₂ EDTA
TE-Puffer, pH 7,6 (Chromatingewinnung)	10 mM Tris 1 mM Na ₂ EDTA
TE-Puffer, pH 8,0 (shotgun-Bank)	10 mM Tris 1 mM Na ₂ EDTA
50T/50E-Puffer, pH 8,0	50 mM Tris 50 mM Na ₂ EDTA

3 Ergebnisse

3.1 Festlegung der Co-Zielregion

Um die Ursache für das Kolumnarwachstum auf molekularer Ebene zu charakterisieren, wurde die ungefähre genomische Lage des Co-Gens zunächst berechnet. Als Grundlage dienten die von Tian *et al.* (2005) publizierten Marker SCAR₂₁₆ und SCAR₆₈₂, die sich in einem genetischen Abstand von 12,3 cM und 2,9 cM zum Co-Gen befinden.

Unter Verwendung der Gleichung

$$P_{xz} = P_{xy} + P_{yz} - 2P_{xy}P_{yz}$$

(x: SCAR₆₈₂, y: Co Gen und z: SCAR₂₁₆; Abbildung 3.1) nach Lynch und Walsh (1998) ergibt sich daraus eine Rekombinationsfrequenz zwischen SCAR₂₁₆ und SCAR₆₈₂ (P_{xz}) von 0,1448, also 14,48 cM. Mit BLASTn-Suchen der Marker-Sequenzen in dem veröffentlichten GD-Genom (Velasco *et al.* 2010) konnte der physikalische Abstand der Marker berechnet werden. SCAR₂₁₆ zeigt einen signifikanten BLASTn-Treffer auf Chromosom 10, Position 22.271.333 bp bis 22.271.508 bp, SCAR₆₈₂ auf Chromosom 10, Position 17.542.255 bp bis 17.542.639 bp. Dies ergibt einen physikalischen Abstand von 4,729 Mbp. Setzt man den genetischen und den physikalischen Abstand in Beziehung zueinander, entsprechen 14,48 cM etwa 4,729 Mbp (Abbildung 3.1).

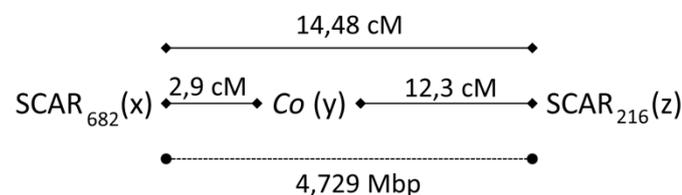


Abbildung 3.1 Vereinfachte genetische bzw. physikalische Karte der Co-Gen-Region

Abgebildet sind die zur Berechnung der Zielregion verwendeten Marker SCAR₆₈₂ und SCAR₂₁₆ mit den genetischen Abständen zum Co-Gen. Darüber hinaus ist die genetische und physikalische Distanz dieser Marker zueinander gezeigt.

Daraus ergibt sich, dass 1 cM in etwa 326.000 bp entspricht. SCAR₆₈₂ liegt bei ungefähr 17,5 Mbp auf Chromosom 10 mit einem Abstand von 2,9 cM zu Co. Das Co-Gen ist somit auf Chromosom 10 um Position 18,5 Mbp zu lokalisieren. Da es sich bei der Berechnung lediglich um eine grobe Abschätzung handelt, wurde die Zielregion auf Chromosom 10, Position 18 bis 19 Mbp, festgelegt. Eine Feinkartierung der Region sollte anhand selbst hergestellter molekularer Marker erfolgen (Kapitel 3.7.2).

3.2 Erstellung von Apfel-BAC-Bibliotheken

Die Isolierung hochmolekularer DNA ist Grundlage für die Konstruktion genomischer BAC-DNA-Bibliotheken. Als Ausgangsmaterial dienten *in-vitro*-Kulturen der heterozygot kolumnaren Apfelsorte P28. Der mechanische Gewebe- und Zellaufschluss durch die Verwendung eines Stabmixers ermöglichte im Vergleich zu Homogenisator bzw. Mörser und Pistill die parallele Verarbeitung einer höheren Anzahl an *in-vitro*-Kulturen. Die isolierten DNAs der unterschiedlichen Ansätze wurden gelelektrophoretisch auf ihre Qualität hin untersucht. Wie exemplarisch in Abbildung 3.2 A gezeigt, hat die Hauptmasse der DNAs aller Ansätze eine elektrophoretische Mobilität deutlich kleiner als die der 23 kb-Bande des verwendeten Molekulargewichtsmarkers. Ungeachtet der verwendeten Methode zum Zell- und Gewebeaufschluss konnten große Mengen qualitativ hochwertiger DNA isoliert werden. Lediglich ein leichter Schmier ist ein Anzeichen für eine minimale Degradierung der DNA.

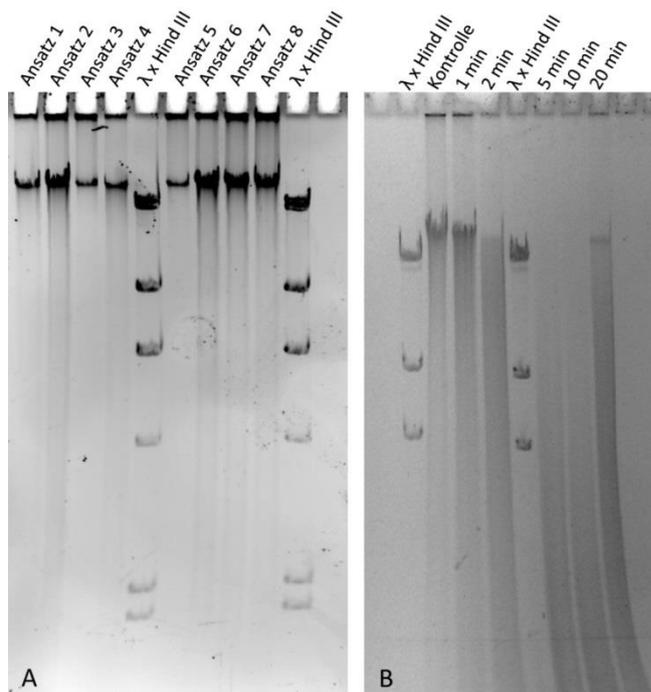


Abbildung 3.2 Gelelektrophoretische Auftrennung genomischer P28-DNA

In Teil A) ist die gelelektrophoretische Auftrennung hochmolekularer genomischer DNA gezeigt. Bei den verschiedenen Ansätzen wurden unterschiedliche Methoden für den Gewebe- und Zellaufschluss gewählt. 1/2/6: Stabmixer; 3: Mörser und Pistill; 4/5/7/8: Homogenisator
 In Teil B) ist eine elektrophoretische Auftrennung von DNA nach unterschiedlich starkem Verdau mit dem Restriktionsenzym *Sau3AI* abgebildet. Die einzelnen Spuren enthalten DNA, die variierenden Restriktionszeiten ausgesetzt war.

Durch einen unvollständigen Restriktionsverdau sollten möglichst große klonierbare DNA-Fragmente generiert werden. Eine geeignete Inkubationsdauer wurde durch die Erstellung einer Restriktionskinetik ermittelt (Abbildung 3.2 B). Die Kontrolle zeigt unrestringierte genomische DNA der Sorte P28. Bereits nach zweiminütiger Inkubationsdauer ist ein deutlicher Verdau der DNA zu erkennen. Nach weiteren drei Minuten befindet sich die Hauptmasse der DNA bereits unter der 6 kb-Bande des Molekulargewichtsmarkers. Problematisch sind oft Schwankungen in der Effizienz des Restriktionsverdaus durch eine inhomogene Verteilung hochmolekularer DNA in Lösung. Dies ist anhand des 20-minütigen Verdaus in Abbildung 3.2 B zu sehen. Neben einem sehr starken Schmier ist eine Bande oberhalb der 23 kb-Bande des Molekulargewichtsmarker zu erkennen. Obwohl jeder Ansatz das gleiche Volumen an DNA-Lösung enthielt, sind deutliche Unterschiede in der Konzentration zu beobachten. Letztlich wurden Restriktionszeiten von einer, eineinhalb und zwei Minuten gewählt, da bei der Restriktionsdauer von einer und zwei Minuten ein eindeutiger Verdau stattgefunden hat, die Hauptmasse der DNA aber dennoch über der 23 kb-Bande des Molekulargewichtsmarkers liegt. Um die Auswirkungen von Unterschieden in der DNA-Konzentration zu minimieren, wurden jeweils Verdaus mit den drei oben genannten Zeiten durchgeführt und die Ansätze gepoolt.

Kleine DNA-Fragmente (< 23 kb) sollten anschließend von der Hauptmasse der DNA abgetrennt werden, da sie bei einer Ligation im Vergleich zu den gewünschten großen Fragmenten bevorzugt ligieren. Aus diesem Grund wurden zwei aufeinanderfolgende Größenfraktionierungen durchgeführt. Dazu wurden zunächst „Maxi“-Gele im speziellen Natriumphosphatpuffer-System verwendet. Durch die stark begrenzte Ladekapazität dieses vertikalen Gelsystems waren zahlreiche Gele (etwa 50) notwendig, um ausreichend DNA für eine Ligation zu erhalten. Hingegen konnte durch die parallele Verwendung zweier „Mega“-Gele bereits ausreichend DNA für eine Ligation bereitgestellt werden.

Nach erfolgter Größenfraktionierung wurde mittels Endfilling-Reaktion mit radioaktiv markierten Nukleotiden unterschiedlicher Zusammensetzung überprüft, ob die überhängenden DNA-Enden intakt waren und eine Ligation in mit *BamHI* restringierten pBeloBAC11 möglich war (Tabelle 3.1).

Tabelle 3.1 Strahlungsintensitäten zur Überprüfung der Integrat-Enden

Ansatz	Pellet 1	Überstand 1	Pellet 2	Überstand 2
A*	10 Bq	750 Bq	1,7 Bq	4,5 Bq
GA*	38 Bq	760 Bq	5 Bq	18 Bq
GATC*	70 Bq	770 Bq	42 Bq	10 Bq
GA*TC*	115 Bq	1210 Bq	70 Bq	25 Bq

NN-3'OH
NN-CTAG-5'P
 Integrat-Ende

1: Fällung; 2: Waschschrift; *: radioaktiv markiert

Bei intakten Integrat-Enden erfolgt der Einbau der Nukleotide in der Reihenfolge dGTP, dATP, dTTP und dCTP. Ansatz A* zeigt die Hintergrundstrahlung an, also die Strahlungsintensität, die messbar ist, obwohl kein Einbau des radioaktiv markierten Nukleotids stattgefunden hat, da das radioaktive Nukleotid α -³²P-dATP ohne die Anwesenheit von dGTP nicht eingebaut werden kann. Im Vergleich dazu enthält der Ansatz GA* neben dem radioaktiv markierten dATP auch dGTP. Nach der Fällung zeigt das Pellet eine deutliche Strahlung und damit den Einbau der Nukleotide. Der Ansatz GATC* weist nach der Fällung ebenfalls eine deutliche Strahlung auf. Da in diesem Ansatz radioaktiv markiertes dCTP verwendet wurde, lässt die gemessene Strahlung die Schlussfolgerung zu, dass Moleküle vorlagen, deren kompletter Überhang intakt war. Der Ansatz GA*TC* bestätigt dies. Die unterschiedlich hohe Strahlung zwischen den Ansätzen GA* und GATC* ist darauf zurückzuführen, dass das radioaktiv markierte dCTP eine höhere spezifische Strahlung aufweist als das markierte dATP. Die Messungen nach dem Waschschrift bestätigen die Messungen der Fällungen, lediglich Ansatz GA* zeigt eine deutlich schwächere Strahlung im Pellet. Möglicherweise führte hier der Waschschrift zum Verlust des Pellets. Die Endfilling-Reaktionen zeigten, dass ausreichend DNA-Enden generiert wurden und diese trotz der zahlreichen Arbeitsschritte erhalten blieben.

Die überprüften Fraktionen wurden kloniert. Für die Transformationen von BAC-Banken wurden elektrokompente Zellen mit einer Transformationsrate von 5×10^9 bis 1×10^{10} verwendet. Es wurden vier BAC-Bibliotheken hergestellt (P28ivL4, P28ivL10, P28ivML1 und P28ivML2), die sich durch die Ausgangsmenge des eingesetzten Materials, die Art des Gewebeaufschlusses und das für die Größenfraktionierung verwendete Gelelektrophorese-system unterschieden. Der Gewebeaufschluss der Banken P28ivL4 und P28ivL10 wurde mittels Homogenisator durchgeführt, ehe dieser durch Verwendung eines Stabmixers erfolgte. Die Größenfraktionierungen erfolgten bei den Banken P28ivML1 und P28ivML2 anstelle von „Maxi“-Gelen mit „Mega“-Gelen. Die Banken P28ivL10 und P28ivML2 wurden nur mit der Hälfte (15 g) an Ausgangsmaterial hergestellt.

Um die statistische Genomabdeckung der erstellten BAC-Bibliotheken näherungsweise zu ermitteln, wurden die Plasmide zufällig ausgewählter Klone isoliert. Nach erfolgreichem Restriktionsverdau wurden die Plasmide gelelektrophoretisch aufgetrennt und die Integratgröße bestimmt (Abbildung 3.3).

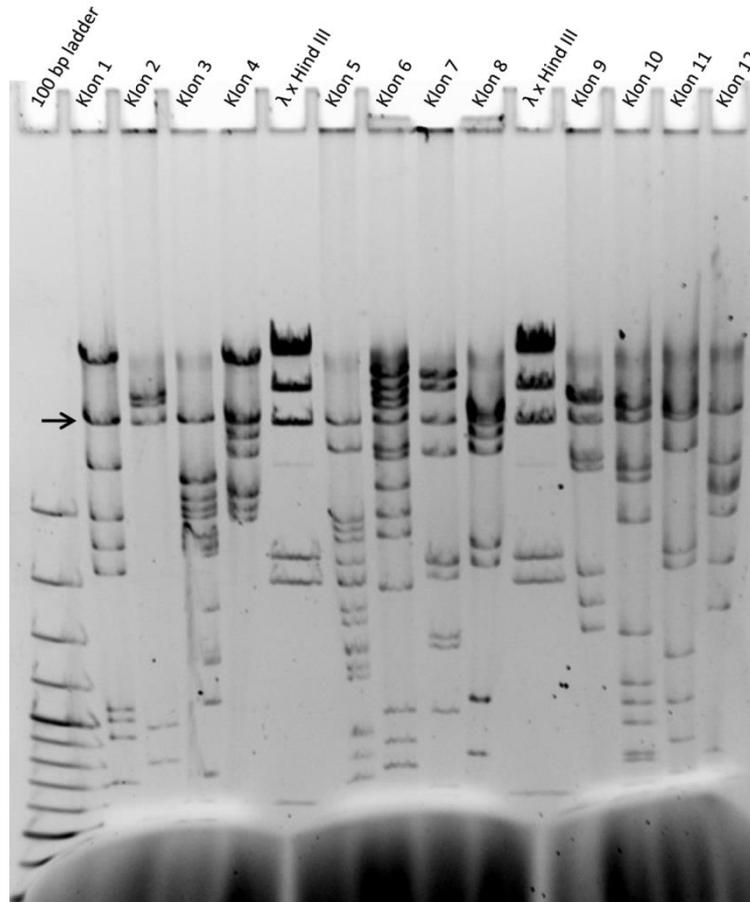


Abbildung 3.3 Restriktionsverdau der Plasmide ausgewählter BAC-Klone

Die Abbildung zeigt die gelelektrophoretische Auftrennung von *EcoRI* geschnittener Plasmid-BAC-DNA. Bei allen Klonen lässt sich eine Vektorbande in Höhe von etwa 6600 bp erkennen (schwarzer Pfeil). Da die Ansätze nicht entsalzt wurden, ist im unteren Bereich des Gelbildes eine starke Salzfront sichtbar. Exemplarisch sind in dieser Abbildung die Klone der Transformationen P28ivML1 gezeigt.

Der verwendete Vektor pBeloBAC11 besitzt zwei Restriktionsschnittstellen für *EcoRI*. Es entstehen ein großes Fragment von 6633 bp und zwei weitere Fragmente (852 bp und 22 bp), die jedoch mit Teilen des Integrates verbunden sind und somit nicht als einzelne Banden bei einer gelelektrophoretischen Auftrennung zu erkennen sind. Anhand des gezeigten Gelbildes konnten die ungefähren Integratgrößen der BAC-Klone abgeschätzt werden (Tabelle 3.2).

Tabelle 3.2 Integratgrößen ausgewählter BAC-Klone

Klon	Größe	Klon	Größe	Klon	Größe
1	36 kb	5	24 kb	9	20 kb
2	18 kb	6	55 kb	10	22 kb
3	30 kb	7	30 kb	11	18 kb
4	38 kb	8	22 kb	12	16 kb

∅ der Integratgröße: 27 kb

Anhand der ermittelten durchschnittlichen Integratgröße der BAC-Klone von 27 kb und der Gesamtanzahl an Kolonien der BAC-Bibliothek konnte die statistische Genomabdeckung bestimmt werden. In Tabelle 3.3 sind die erfolgreich konstruierten BAC-Bibliotheken mit Anzahl der Kolonien, durchschnittlicher Integratgröße und daraus resultierender Gesamtintegratgröße aufgelistet. Unter Verwendung der Carbon-Clarke-Formel wurde die Wahrscheinlichkeit (P), dass ein einmaliger („single-copy“) DNA-Abschnitt in der BAC-DNA-Bibliothek enthalten ist, berechnet (Clarke und Carbon 1976).

Die Formel lautet:

$$P = 1 - (1-f)^N$$

(N: Anzahl der Kolonien; f: Klongröße dividiert durch die Genomgröße)

Tabelle 3.3 BAC-DNA-Bibliotheken der Apfelsorte Procats 28

Bank	Anzahl an Kolonien	∅ Integratgröße	Gesamtintegratgröße	P (Carbon-Clarke)
P28ivL4	26.000	27 kb	702 Mbp	0,61
P28ivL10	13.000	31 kb	403 Mbp	0,42
P28ivML1	79.000	27 kb	2.133 Mbp	0,94
P28ivML2	35.000	24 kb	840 Mbp	0,68
Gesamt	153.000	26,5 kb	4.078 Mbp	0,99

Die erstellten BAC-DNA-Bibliotheken umfassten insgesamt rund 150.000 Klone, die zusammengenommen rund 4.000 Mbp genomische Apfel-DNA enthielten. Geht man von einer Genomgröße des Apfels von 742,3 Mbp aus (Velasco *et al.* 2010), entspricht dies einer statistischen Genomabdeckung von etwa 5,4-fach. Die Wahrscheinlichkeit, dass ein „single-copy“ DNA-Abschnitt in den erstellten BAC-DNA-Bibliotheken vorhanden ist, beträgt nach der Carbon-Clarke-Formel 99 %.

3.3 Screening der Apfel-BAC-Bibliotheken

Um die Apfel-BAC-Bibliotheken nach Klonen aus der *Co*-Region durchsuchen zu können, mussten zunächst geeignete Sonden für das Screening hergestellt werden. Die genomische Region des *Co*-Gens konnte grob auf Chromosom 10 Position 18 bis 19 Mbp eingegrenzt werden (siehe 3.1). Die verwendeten Sonden wurden in insgesamt vier Runden erstellt. Der erste und zweite Teil der Sonden beruht auf Roche 454- bzw. Illumina-Transkriptomdaten von Clemens Krost (2012), die gegen die Contigs der Zielregion des GD-Genoms kartiert wurden (Anhang Tabelle A2.1 und Tabelle A2.2). In einer dritten Runde wurden Sonden in die Bereiche gelegt, in denen noch keine BAC-Klone lokalisiert werden konnten. Als Sequenzgrundlage dienten die annotierten Contigs des GD-Genoms (Anhang Tabelle A2.3). Es wurden bevorzugt Sonden erstellt, die in Exon-Bereichen der Gene in der Zielregion liegen, um auf diese Weise die Chance auf einen „single-copy“-Charakter zu erhöhen. In einer vierten Runde wurde speziell die Region um 18,7 Mbp auf Chromosom 10 untersucht, da dieser Bereich bisher von keinem Klon abgedeckt wurde. Die Schwierigkeit lag vor allem darin, dass im GD-Genom keine Contigs an dieser Stelle annotiert sind. Sonden wurden in die Randbereiche von „long range“ PCR-Produkten gelegt, die in die Lücke hineinreichten (Anhang Tabelle A2.4). Alle erstellten Sonden wurden zunächst mit Southern oder „Pirota“ Blot-Analysen auf repetitive Elemente überprüft. In Abbildung 3.4 ist exemplarisch die gelelektrophoretische Auftrennung einiger Sonden (A) und das dazugehörige Autoradiogramm (B) des „Pirota“ Blots zu sehen.

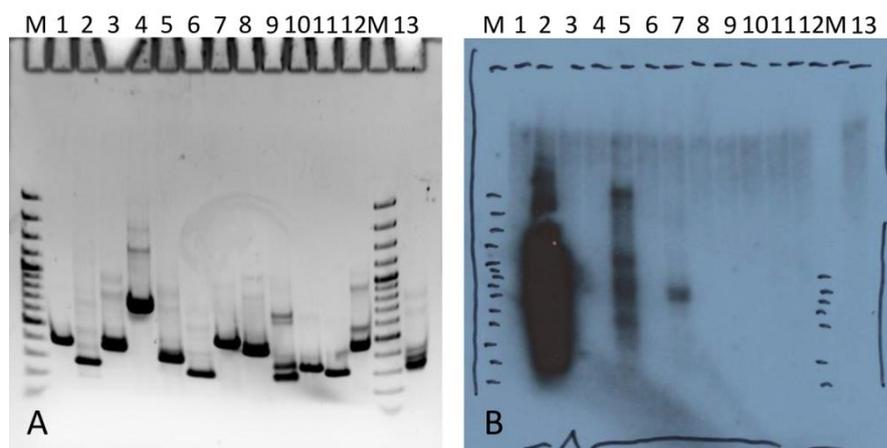


Abbildung 3.4 „Pirota“ Blot potenzieller Sonden

In A) ist die gelelektrophoretische Auftrennung 13 potenzieller Sonden abgebildet. Als Marker (M) wurde die 100 bp Ladder verwendet. B) zeigt den „Pirota“ Blot dieses Gels. Als Sonde diente radioaktiv markierte genomische P28-DNA. Sonden in den Spuren 2 und 5 enthalten vermutlich hoch- und mittelrepetitive Elemente und wurden daher für ein Screening ausgeschlossen.

Die Sonden in den Spuren 2 und 5 zeigen im Autoradiogramm des „Pirota“ Blots ein starkes bzw. sehr starkes Signal, sodass sie vermutlich mittel- oder hochrepetitive Elemente beinhalten und für ein Screening somit nicht infrage kommen. Die Sonde der Spur 7 zeigt ein Signal im „Pirota“ Blot, jedoch nicht auf Höhe des spezifischen PCR-Produktes. Aus diesem Grund konnte diese Sonde nach Aufreinigung durch eine präparative Gelelektrophorese verwendet werden. Der Großteil der Sonden, wie bspw. die in Spur 10, zeigt eine distinkte Bande nach der gelelektrophoretischen Auftrennung und kein Signal im „Pirota“ Blot. Hoch- und mittelrepetitive Elemente konnten somit ausgeschlossen werden, und die Sonden sind demnach für ein Screening geeignet. Die Sonden für die Region um 18,7 Mbp auf Chromosom 10 zeigten alle ein schwaches Signal im „Pirota“ Blot. Da es jedoch nicht möglich war, „single-copy“ Sonden in diesem Bereich zu erstellen, wurden sie dennoch für ein Screening verwendet.

Mittels Koloniefilter-Hybridisierung wurden die Apfel-BAC-Bibliotheken nach Klonen aus der Zielregion durchsucht. Sondenpools von bis zu 14 Sonden wurden für das Screening eingesetzt (Anhang A2). Mit den Sondenpools 1, 2 und 3 konnten insgesamt 61, mit dem Sondenpool 4 51 positive Klone identifiziert werden (Tabelle 3.4). Die genomweite Duplikation (u.a. Chromosom 5 und Chromosom 10) sowie die Tatsache, dass Sondenpool 4 repetitive Sonden enthielt, gaben Anlass zu der Vermutung, dass nicht alle positiven Klone von Chromosom 10 stammen. Um die genomische Lage der BAC-Klone zu ermitteln, wurden Randsequenzierungen der Klone durchgeführt. In Tabelle 3.4 sind die Ergebnisse der Randsequenzierungen dargestellt. Zudem sind die positiven Klone der entsprechenden BAC-Bibliothek sowie dem verwendeten Sondenpool zugeordnet.

Tabelle 3.4 Lokalisierung positiver BAC-Klone der Koloniefilter-Hybridisierung

Bank	Gesamt-integratgröße	Sondenpool 1	Sondenpool 2	Sondenpool 3	Sondenpool 4
P28ivL4	702 Mbp	3/3/0/0	-	-	-
P28ivL10	403 Mbp	2/1/1/0	-	-	-
P28ivML1	2.133 Mbp	25/14/9/2	16/8/2/6	15/10/1/4	-
P28ivML2	840 Mbp	-	-	-	51/1/2/48

Zu den Sondenpools: Gesamtzahl pos. Klone/Klone Chr.10/Klone Chr.5/Klone anderes Chr.

Von den 61 positiven Klonen, die mithilfe der Sondenpools 1, 2 und 3 detektiert wurden, stammen 36 von Chromosom 10, 13 von Chromosom 5 und 12 von anderen Chromosomen. Von den 51 positiven Klonen bei Verwendung von Sondenpool 4 konnte lediglich ein Klon auf Chromosom 10 lokalisiert werden. Zwei weitere stammen von Chromosom 5 und 48 Klone von diversen anderen Chromosomen. Durch die Randsequenzierungen konnten

somit insgesamt 37 Klone identifiziert werden, die auf Chromosom 10 in der Region 18 bis 19 Mbp lokalisiert werden konnten. Diese Klone sind in Tabelle 3.5 in der Reihenfolge ihrer genomischen Lage aufgelistet. Aus der genomischen Lage der Randsequenzen und den Restriktionsverdau der isolierten Plasmide konnten zunächst die erwarteten Integratgrößen abgeschätzt werden. Die tatsächliche Integratgröße wurde durch die vollständige Sequenzierung (Kapitel 3.4) ermittelt. Die Zuordnung zu einem Metacontig ist in Kapitel 3.5 beschrieben.

Tabelle 3.5 BAC-Klone aus der Co-Zielregion

Klon	BLAST-Treffer (Sp6)	BLAST-Treffer (T7)	Berechnete Integratgröße	Tatsächliche Integratgröße	Metacontig
D2	18,088 Mbp	18,139 Mbp	51 kb	50.708 bp	1nc
2D8	18,130 Mbp	18,166 Mbp	36 kb	37.644 bp	1c
K4	18,139 Mbp	18,173 Mbp	34 kb	32.824 bp	1c
3I9	18,167 Mbp	Chr. 6			1c
F7	18,198 Mbp	18,225 Mbp	27 kb	27.410 bp	2nc
A4	18,20 Mbp	18,245 Mbp	45 kb	41.914 bp	2nc
C5	18,20 Mbp	18,246 Mbp	46 kb	38.600 bp	2nc
B3_3	18,227 Mbp	18,263 Mbp	36 kb	34.206 bp	2nc
J7_3	18,237 Mbp	18,269 Mbp	32 kb	35.082 bp	2nc
2G4	18,261 Mbp	18,302 Mbp	41 kb	42.491 bp	2nc
3G6	18,263 Mbp	18,293 Mbp	30 kb	29.226 bp	2nc
K25	18,292 Mbp	18,312 Mbp	20 kb	24.397 bp	3c
I3	18,293 Mbp	18,352 Mbp	59 kb	44.784 bp	3c
A8	18,297 Mbp	18,340 Mbp	43 kb	29.002 bp	3c
2J6	unbekannt	18,408 Mbp			3nc
J2	18,387 Mbp	18,425 Mbp	38 kb	35.096 bp	4c
H3	18,420 Mbp	18,460 Mbp	40 kb	51.371 bp	4c
I2_2	18,465 Mbp	18,507 Mbp	42 kb	44.881 bp	4c
L2	18,500 Mbp	18,535 Mbp	35 kb	34.525 bp	4nc
B9	18,512 Mbp	18,542 Mbp	30 kb	29.256 bp	4nc
H1	18,472 Mbp	18,534 Mbp	62 kb	62.343 bp	4c
H2	18,541 Mbp	18,571 Mbp	30 kb	30.259 bp	4nc
H4	18,559 Mbp	18,579 Mbp	20 kb	19.886 bp	4nc
G5	18,550 Mbp	Chr. 13			5c
D4	Chr. 14	18,600 Mbp			5c
2F3	18,608 Mbp	18,645 Mbp	37 kb	37.004 bp	6nc
2I8	18,622 Mbp	18,653 Mbp	31 kb	31.427 bp	6c_a
3B8	18,641 Mbp	18,665 Mbp	24 kb	25.455 bp	6nc
2I4	18,661 Mbp	Chr. 4			6c_b
K65	Chr.4	Chr.4	34 kb	34.006 bp	6c_b
I2_3	18,749 Mbp	18,800 Mbp	51 kb	59.881 bp	7c
G4_3	18,775 Mbp	18,804 Mbp	29 kb	29.444 bp	7nc
F1_3	18,798 Mbp	18,850 Mbp	52 kb	54.162 bp	7nc
A4_3	18,806 Mbp	18,869 Mbp	63 kb	60.065 bp	7c
K2_3	Chr. 0	18,836 Mbp			7c
A9_3	18,837 Mbp	18,859 Mbp	22 kb	21.609 bp	7c
1C3	18,884 Mbp	18,923 Mbp	39 kb	39.293 bp	8c

rot: vermutlich chimärer BAC-Klon (Kapitel 3.4.3)

Da es sich bei dem verwendeten Referenzgenom um das GD-Genom (Velasco *et al.* 2010) handelt, die BAC-Klone aber der Apfelsorte P28 entstammen, waren Unterschiede zwischen der aufgrund der publizierten Apfelgenomsequenz abgeschätzten und tatsächlichen Integratgröße durchaus zu erwarten. So liegen bspw. bei dem Klon H3 die Randsequenzen nach BLASTn-Suchen etwa 40 kb auseinander, die tatsächliche Integratgröße beträgt jedoch etwa 51 kb. Diese Differenz und damit ein möglicher Unterschied zwischen dem kolumnaren und nicht kolumnaren Chromosom ist Gegenstand der detaillierten Sequenzauswertung in Kapitel 3.5. Sechs BAC-Klone sind in Tabelle 3.5 in rot dargestellt. Bei diesen liegt einer der BLAST-Treffer der Randsequenzen eindeutig auf Chromosom 10, der andere weist einen signifikanten Treffer auf einem anderen Chromosom auf. In diesen Fällen handelt es sich höchstwahrscheinlich um *in-vitro*-Ligate, also um ein Vektormolekül, das zwei Integratmoleküle aufgenommen hat (chimäres Plasmid). Diese Klone wurden dennoch vollständig sequenziert und anschließend genauer analysiert (Kapitel 3.4.3). Der Klon K65 zeigt bei BLASTn-Suchen der Randsequenzen eindeutige Treffer auf Chromosom 4. Wie sich allerdings später herausstellte, wurde der entsprechende Contig von Velasco *et al.* (2010) falsch annotiert und K65 liegt auf Chromosom 10 (Kapitel 3.4.3).

3.4 BAC-Klone aus der Co-Zielregion und Illumina Next-Generation Sequencing

3.4.1 Sequenzierung und Assemblierung der BAC-Klone

Die Klone K25, H2, H4 und K4 wurden mittels Shotgun-Sequenzierung vollständig sequenziert. Die Sequenzen der Subklone wurden mit dem Software-Paket Lasergene™ der Fa. DNASTAR Inc. in einen Kloncontig assembliert. Der BAC-Klon K25 besitzt ein Integrat mit einer Länge von 24.397 bp. Die Längen der Integrate von H2, H4 und K4 betragen 30.259, 19.886 und 32.824 bp.

Der Klon K65 wurde mittels „Primer-Walk“ und Sanger-Sequenzierung vollständig sequenziert und besitzt eine Integratgröße von 34.006 bp. Aus Zeit- und Kostengründen wurden die verbleibenden 32 BAC-Klone mittels Illumina Next-Generation Sequencing sequenziert. Hierzu wurden „paired-end“-Bibliotheken der Plasmide erstellt und diese mit dem Illumina HiSeq 2000 sequenziert. Pro Klon wurden 3,5 bis 12 Mio. Sequenzen (Reads) generiert. Diese Rohdaten wurden nach erfolgter Integritätsprüfung („fastq_integrity.pl“) für die *de novo*-Assemblierung der BAC-Klone verwendet. Insgesamt konnte für alle 32

Klone ein Contig erstellt werden, der die geforderten Kriterien (Einschluss der Sp6/T7-Randsequenzen und Ringschluss) erfüllte. Zur bioinformatischen Überprüfung wurden die gefilterten Reads gegen die Contigs mittels CLC Genomics Workbench kartiert. Bereiche auffälliger Sequenzabdeckung wurden durch Sanger-Sequenzierungen verifiziert. Zur abschließenden Untersuchung wurden Restriktionskartierungen durchgeführt. In Abbildung 3.5 und Tabelle 3.6 sind exemplarisch die gelelektrophoretische Auftrennung der Restriktionsverdau des Klons 3G6 und die *in silico* berechneten Fragmentgrößen für zwei der vier Restriktionsverdau dargestellt.

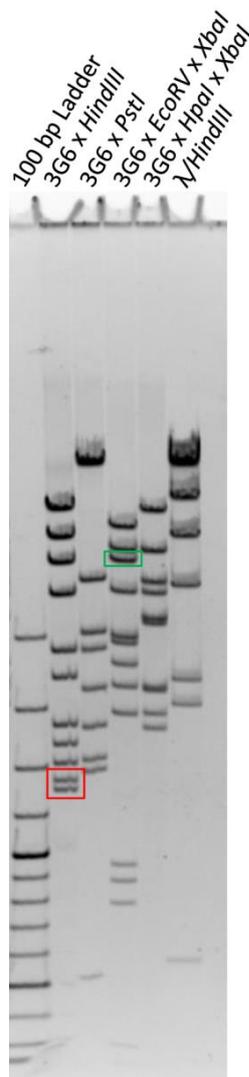


Tabelle 3.6 Berechnete Fragmentgrößen des Klons 3G6

<i>HindIII</i>	<i>EcoRV x XbaI</i>
8279	6805
6268	5596
5105	4706
4000	3862
2679	2954
2302	2834
1804	2455
1649	2204
1505	1903
1326	936
1320	864
233	776
173	498
133	229
17	143

Abbildung 3.5 Restriktionskartierung des BAC-Klons 3G6

Die Abbildung zeigt das Fragmentmuster von vier Restriktionsverdau des BAC-Klons 3G6 mit den Enzymen *HindIII*, *PstI*, *EcoRV* und *XbaI* sowie *HpaI* und *XbaI* nach gelelektrophoretischer Auftrennung. Zur Größenbestimmung dienen die in den beiden äußeren Spuren aufgetragenen Molekulargewichtsstandards. Kritische Stellen in den Restriktionsverdau sind durch farbige Rahmen markiert.

Die anhand des Gels abgeschätzte Größe der Banden wurde mit den *in silico* berechneten Größen verglichen. Bei einer Übereinstimmung kann man von einer korrekt zusammengesetzten BAC-Sequenz ausgehen. Allerdings handelt es sich bei der

Größenbestimmung der Fragmente aufgrund des Gelbildes lediglich um eine Abschätzung, sodass kleine Abweichungen nicht berücksichtigt werden können. Zudem sind Fragmente unter 300 bp von der Lauffront überlagert und wurden daher ignoriert. In dem ausgewählten Beispiel des BAC-Klons 3G6 treten zwei Phänomene auf, die vereinzelt beobachtet werden konnten. Zum einen wurde im Doppelverdau mit *EcoRV* und *XbaI* eine *XbaI*-Schnittstelle vermutlich nicht geschnitten. Die erwarteten Banden bei 498 bp und 4706 bp treten nicht auf, dafür ist eine Bande auf Höhe von ungefähr 5200 bp sichtbar (Abbildung 3.5, Tabelle 3.6, grün). Die gleiche *XbaI*-Schnittstelle ist auch im Doppelverdau mit *HpaI* und *XbaI* nicht geschnitten. Eine Sequenzanalyse zeigt jedoch eine intakte *XbaI*-Schnittstelle. Das zweite Phänomen zeigt sich im Restriktionsverdau mit *HindIII*. Die berechneten Fragmente ergeben u.a. ein DNA-Molekül von 1320 bp und eines von 1326 bp Größe. Der minimale Größenunterschied sollte in einer Bande doppelter Intensität resultieren. Das Agarosegel zeigt jedoch zwei klare Einzelbanden (Abbildung 3.5, Tabelle 3.6, rot). Die Bereiche wurden aus diesem Grund einer PCR-Analyse unterzogen. Das PCR-Produkt, welches der 1320 bp-Bande entspricht, zeigte erneut eine langsamere Laufgeschwindigkeit in der Gelelektrophorese. Die anschließende Sequenzierung der Banden bestätigte jedoch die vermuteten Sequenzen. Die langsamere Laufgeschwindigkeit der 1320 bp-Bande muss somit in der Sequenz begründet sein. Eine mögliche Erklärung dieser zwei Phänomene wird in Kapitel 4.3 ausführlich diskutiert. Insgesamt wurden auf diese Art und Weise alle Illumina-sequenzierten BAC-Klone überprüft und die Sequenzen bestätigt. Die mittels Sequenzierung bestimmte Integratgröße der BACs ist in Tabelle 3.5 aufgelistet.

3.4.2 Genomische Illumina-Sequenzdaten der Apfelsorte P28

Als Ergänzung zu den BAC-Klonen und zur Identifizierung chimärer BAC-Klone sowie zur Verifizierung der im Laufe dieser Arbeit erstellten Metacontigs und des Gesamtcontigs (Kapitel 3.5.1) wurden genomische Illumina „mate pair“-Bibliotheken der Apfelsorte P28 erstellt und sequenziert. In Tabelle 3.7 sind Umfang, Zusammensetzung und Qualität der Illumina-Rohdaten und der gefilterten (gf) Illumina-Datensätze der „mate pair“-Bibliotheken (mp) gezeigt. Die statistische Auswertung erfolgte mit Hilfe des Perl-Skriptes „seq_info.pl“. Der „forward“- und „reverse“-Datensatz einer Sequenzierung wurden jeweils zusammengefasst und die durch das Filtern der Rohdaten entstandenen „single end“-Reads verworfen. Wurde eine Bibliothek mehrfach sequenziert, so ist dies durch römische Ziffern

gekennzeichnet. Die bei der Herstellung der Banken verwendete Fragmentgröße ist jeweils angegeben.

Tabelle 3.7 Statistische Auswertung der genomischen Illumina „mate pair“-Reads

Bibliothek	Anzahl der Sequenzen	Anzahl der Basen	Ø-liche Sequenzlänge	% AT	% GC	% N	Ø-liche Qualität [Phred]
mp_3,5-4 kb	42.195.980	4.261.793.980	101,00	59,55	40,30	0,15	34,66
mp_3,5-4 kb_gf	38.316.692	3.321.641.056	86,72	60,49	39,51	0	37,47
mp_5-6 kb	62.874.970	6.350.371.970	101,00	59,90	39,95	0,15	34,35
mp_5-6 kb_gf	56.039.022	4.871.150.479	86,95	60,85	39,15	0	37,43
mp_7-8 kb_I	67.204.562	6.787.660.762	101,00	60,30	39,52	0,18	34,78
mp_7-8 kb_I_gf	60.624.108	5.336.986.428	88,04	61,02	38,98	0	37,59
mp_7-8 kb_II	169.925.944	17.162.520.344	101,00	60,74	39,24	0,02	34,94
mp_7-8 kb_II_gf	154.347.190	13.578.500.132	87,98	61,20	38,80	0	37,23
mp_7-8 kb_III	173.633.588	17.536.992.388	101,00	60,71	39,27	0,02	35,19
mp_7-8 kb_III_gf	157.782.392	13.911.790.073	88,17	61,17	38,83	0	37,40
Gesamt: gf	467.109.404	41.020.068.168	87,83	61,07	38,93	0	37,38

Nach erfolgter Bereinigung der Rohdaten standen etwa 467 Mio. „mate pair“-Reads zur Verfügung, was einer statistischen Genomabdeckung von rund 55-fach entspricht. In Abbildung 3.6 ist die Anzahl der Reads gegen die Abstände der zusammengehörigen „forward“- und „reverse“-Sequenzierung eines exemplarischen Mappings der „mate pair“-Bibliotheken mit maximaler Stringenz dargestellt. Die tatsächlichen Abstände der „mate pair“-Reads in dem Mapping spiegeln die Größe der eingesetzten Fragmente für die erstellten Bibliotheken wider, was für erfolgreich erstellte „mate pair“-Banken spricht.

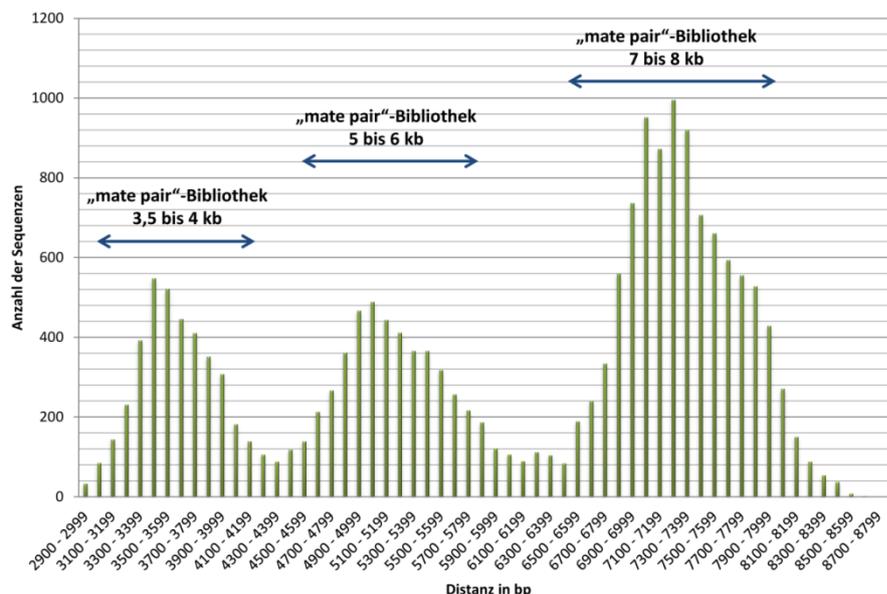


Abbildung 3.6 Tatsächliche Abstände der „mate pair“-Reads

Gezeigt ist die grafische Auftragung der Anzahl der Sequenzen gegenüber der Distanz der „forward“- und „reverse“-Sequenzierung eines Moleküls.

Bei der Kartierung der Sequenzen fiel jedoch auf, dass außergewöhnlich viele „mate pair“-Reads identisch waren. Aus diesem Grund wurden mit der Funktion „Remove Duplicate Reads“ die kartierten „mate pair“-Reads des exemplarischen Mappings überprüft. Die Ergebnisse sind in Tabelle 3.8 aufgeführt. Auffällig ist die extrem hohe Anzahl an Duplikaten in der Bibliothek mp_7-8 kb. Diese liegt je nach Sequenzierung bei knapp 50 bis über 70 %.

Tabelle 3.8 Statistische Analyse von Duplikaten in den „mate pair“-Bibliotheken

Datensätze	Kartierte Reads	Duplikate	% Duplikate	Verbleibende Reads
mp_3,5-4 kb_gf	7.752	306	3,95	7446
mp_5-6 kb_gf	11.536	1.508	13,07	10.028
mp_7-8 kb_I_gf	11.880	5.860	49,33	6020
mp_7-8 kb_II_gf	24.312	17.354	71,38	6958
mp_7-8 kb_III_gf	28.840	20.924	72,55	7916
Gesamt	84.320	45.952	54,5	38.368

Leider konnte die Funktion „Remove Duplicate Reads“ nicht auf die aufbereiteten Rohdaten angewandt werden, da dazu nicht ausreichend Rechenkapazität zur Verfügung stand. Allerdings ist davon auszugehen, dass die in Tabelle 3.7 errechnete Anzahl von 41 Milliarden Basen nicht einer statistischen Genomabdeckung von etwa 55-fach entspricht, sondern diese bei über 50 % Duplikaten nur etwa 26-fach beträgt.

3.4.3 Analyse chimärer BAC-Klone

Die in Tabelle 3.5 in rot dargestellten BAC-Klone (3I9, 2J6, G5, D4, 2I4 und K2_3) weisen bei BLASTn-Suchen der Randsequenzen Treffer auf unterschiedlichen Chromosomen auf, was ein Hinweis auf ein chimäre Plasmide sein kann. Bei einem *in-vitro*-Ligat sollte eine mögliche Verbindungsstelle zwischen den DNA-Fragmenten unterschiedlicher chromosomaler Herkunft in einem Mapping der „mate pair“-Datensätze von keinen „mate pair“-Read überspannt werden. Zudem sollte eine vermeintliche Verbindungsstelle mittels BLASTn-Suchen zu identifizieren sein. In Abbildung 3.7 sind Ausschnitte der Mappings am Beispiel von K2_3 und 2I4 dargestellt. Im Falle von K2_3 ist ein deutlicher Bruchpunkt (Abbildung 3.7 A, roter Pfeil) zu erkennen. Kein „paired read“ überspannt diese Stelle, weshalb man davon ausgehen kann, dass diese Verbindungsstelle im Genom so nicht existiert. Der Bruchpunkt beinhaltet die Sequenz „GATC“, was ein weiterer Hinweis dafür ist, dass hier ein *in-vitro*-Ligat vorliegt, bei dem zwei *Sau3AI*-Fragmente miteinander über die *Sau3AI*-überhängenden Enden verbunden wurden.

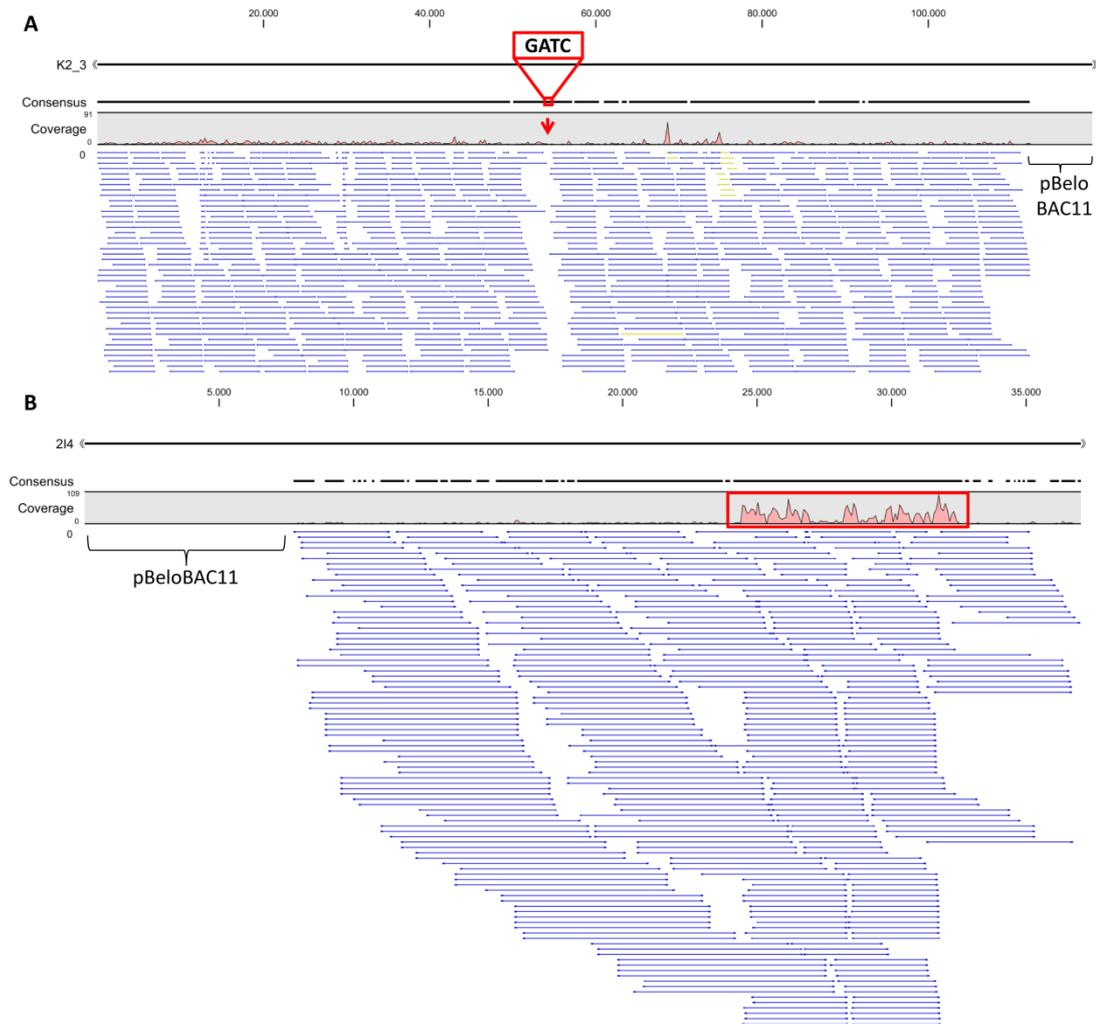


Abbildung 3.7 Mapping der „mate pair“-Reads gegen die BAC-Klone K2_3 und 214

A) zeigt das Mapping der genomischen „mate pair“-Reads gegen den BAC Klon K2_3. Die Verbindungsstelle zwischen den zwei DNA-Molekülen unterschiedlicher chromosomaler Herkunft ist mit einem roten Pfeil markiert und die dort liegende Basenabfolge „GATC“ vergrößert dargestellt. In B) ist das Mapping des Klons 214 abgebildet. Hier ist keine Verbindungsstelle zu identifizieren. Der Bereich um 24.000 bis 33.000 bp fällt durch eine wesentlich höhere Abdeckung als der restliche Teil des Klones auf, was auf repetitive Strukturen in diesem Bereich hindeutet (rot umrandet). In beiden Mappings ist nur ein Ausschnitt der kartierten Reads gezeigt. Zudem wurden ausschließlich „paired reads“ kartiert.

Auf diese Weise wurde der von Chromosom 10 stammende Anteil bei den Klone 319, 2J6, G5, D4 und K2_3 identifiziert und verifiziert (Tabelle 3.9).

Tabelle 3.9 Zusammensetzung chimärer BAC-Klone

Klon	Integratgröße	Integrat von Chr. 10
319	56.310 bp	29.257 bp
2J6	nicht bestimmt	29.422 bp
G5	70.861 bp	35.737 bp
D4	76.546 bp	46.805 bp
K2_3	112.223 bp	57.867 bp

Im Gegensatz dazu wird der BAC-Klon 214 in dem durchgeführten Mapping der „mate pairs“ komplett überspannt. Auffällig ist jedoch die hohe Abdeckung im Bereich um 24.000 bis 33.000 bp, die auf repetitive Strukturen hindeutet (Abbildung 3.7 B). In Abbildung 3.8 ist ein Schema des Klonen inklusive der BLASTn-Treffer dargestellt.



Abbildung 3.8 Schema des BAC-Klons 214 inklusive BLASTn-Treffern

Dargestellt sind der BAC-Klon 214 mit Transposon, sowie den BLASTn-Treffern von Chromosom 10 (MDC038765.9) und Chromosom 4 (MDC023034.28). Der Grafik sind zudem die Positionen der Sonden 214_S1 und 34.28_S1 zu entnehmen. Die Primer für eine überspannende „long range“ PCR sind mit „a“ und „b“ gekennzeichnet; die Primer der inversen PCR mit „c“ und „d“.

Zur Klärung der Frage, ob es sich bei dem Klon 214 um ein *in-vitro*-Ligat handelt, wurde eine „long range“ PCR auf genomische P28 DNA durchgeführt, die das Transposon (Abbildung 3.7, roter Kasten) und damit die mögliche Verbindungsstelle überspannt. Die verwendeten Primer sind in Abbildung 3.8 mit „a“ und „b“ gekennzeichnet. Das erwartete Produkt von rund 11 kb konnte amplifiziert und mittels zweier Sequenzierungen verifiziert werden. Des Weiteren wurden genomische Southern Blot Analysen mit den Sonden 214_S1 und 34.28_S1 durchgeführt. Die verwendeten Restriktionsenzyme und die damit verbundenen erwarteten Fragmentgrößen auf dem Autoradiogramm sind in Tabelle 3.10 aufgelistet

Tabelle 3.10 Erwartete Fragmentgrößen der Southern Blot Analysen des BAC-Klons 214

Enzym	<i>EcoRI</i>	<i>HpaI</i>	<i>SacI</i>	<i>SacII</i>	<i>PstI</i>
214_S1	17.529 bp	7.361 bp	6.433 bp	48.199 bp	22.265 bp
34.28_S1	17.529 bp	8.314 bp	10.628 bp	48.199 bp	22.265 bp

Das Gelbild (A) und die Autoradiogramme der Sonden 214 (B) und 34.28_S1 (C) sind in Abbildung 3.9 dargestellt. Das Agarosegel wurde unidirektional geblottet und die Nitrozellulosemembran zunächst mit der Sonde 214_S1 und anschließend mit der Sonde 34.28_S1 hybridisiert. Der *EcoRI*-Verdau zeigt auf beiden Autoradiogrammen ein definiertes Signal erwarteter, identischer Größe. Die Signale des *HpaI*-Verdaus sind auf den Autoradiogrammen ebenfalls gleich, sodass vermutlich die *HpaI*-Schnittstelle im Transposon nicht geschnitten wurde. Der *SacI*-Ansatz zeigt auf beiden Autoradiogrammen ein Signal in der erwarteten Höhe. Die Signale des Ansatzes *SacII* liegen in einem

Größenbereich, der mit dem verwendeten Gelelektrophorese-System nicht zufriedenstellend aufgetrennt werden konnte. Der Verdau von *PstI* scheint zudem unvollständig zu sein. Dennoch bestätigen die Signale im Autoradiogramm des *EcoRI*-, *HpaI*- und *PstI*-Verdaus die angenommene genomische Struktur.

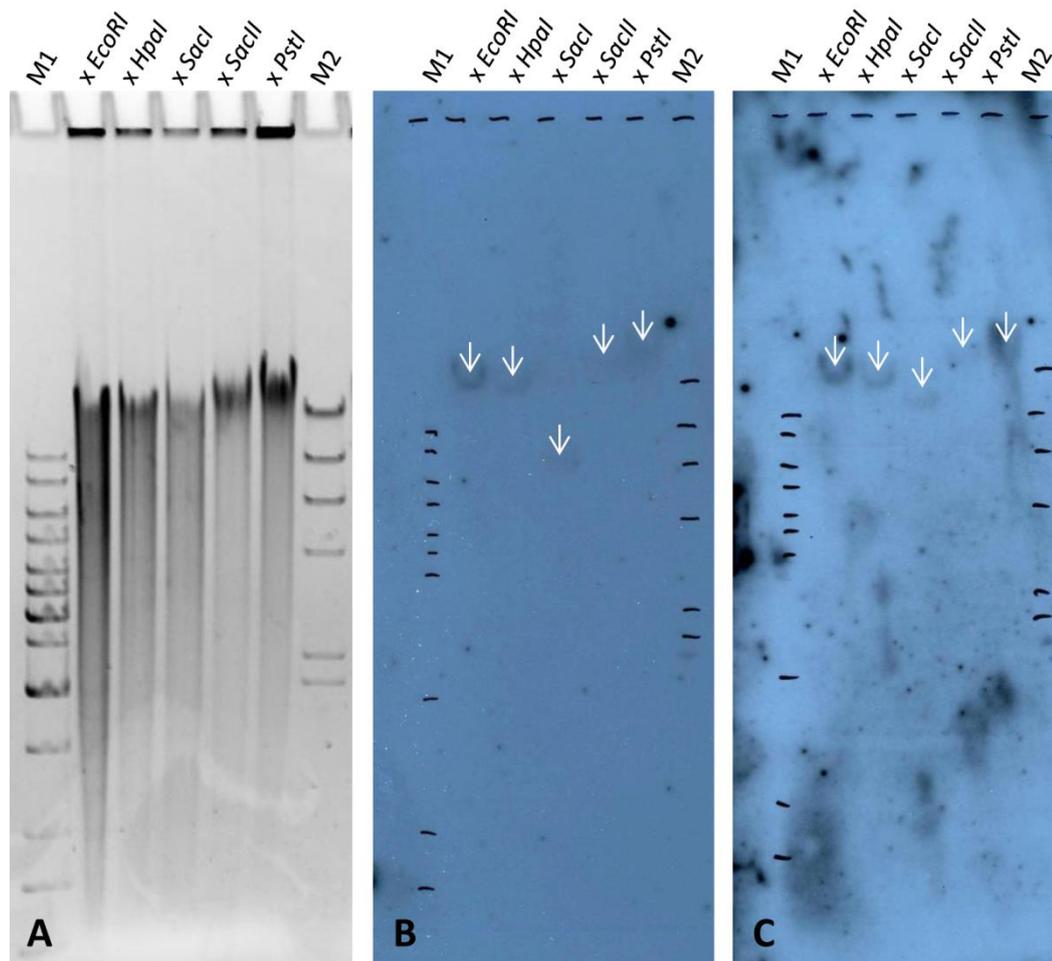


Abbildung 3.9 Southern Blots des BAC-Klons 214

A) zeigt die gelelektrophoretische Auftrennung restringierter genomischer DNA, B) und C) die Autoradiogramme der Sonden 214_S1 und 34.28_S1. Die Expositionszeit betrug sieben (B) bzw. sechs (C) Tage. Die Signale des Autoradiogramms sind mit weißen Pfeilen markiert. Als Molekulargewichtsstandards wurde die 1 kb Ladder (M1) und Lambda/*HindIII* DNA (M2) verwendet.

Als drittes und letztes Verfahren wurde eine inverse PCR mit den in Abbildung 3.8 gekennzeichneten Primern „c“ und „d“ durchgeführt. Das PCR-Produkt wurde durch Sequenzierung verifiziert und die in Abbildung 3.8 gezeigte Struktur von 214 bestätigt. Durch das Mapping der „mate pairs“, den Analysen durch „long range“ PCR, den genomischen Southern Blots und der inversen PCR konnte gezeigt werden, dass der Klon 214 kein *in-vitro*-Ligat ist und somit der Apfelcontig MDC023034.28 auf Chromosom 4 nicht korrekt annotiert ist. Da der BAC-Klon K65 über eine Länge von 7,3 kb mit dem BAC-Klon 214

mit 100 %-iger Übereinstimmung überlappt, konnte auch dieser Klon, trotz BLAST-Treffern der Randsequenzen auf Chromosom 4, auf Chromosom 10 positioniert werden.

3.5 Analyse der genomischen *Co-Gen*-Region

3.5.1 Konstruktion von Metacontigs und eines Gesamtcontigs

Als Grundlage für die Rekonstruktion der mit *Co* gekoppelten genomischen Region in der kolumnaren Apfelsorte P28 wurden die bereits vorhandenen genomischen Sequenzen von GD (Velasco *et al.* 2010) verwendet. Die Sequenzen der BAC-Klone wurden anhand ihrer BLASTn-Treffer in der Zielregion positioniert. Bereiche, in denen nach den Screenings keine BAC-Klone zur Verfügung standen, wurden mittels „long range“ PCR amplifiziert und anschließend sequenziert. Besonders schwierig gestaltete sich dies an Position 18,68 bis 18,71 Mbp (Chromosom 10) im GD-Genom, da in diesem Bereich keine Contigs annotiert sind und die Lücke zu groß für eine direkte „long range“ PCR war. Daher wurden die „mate pair“-Reads gegen die Ränder der bekannten Sequenz kartiert. Es wurden anschließend die Partner-Reads extrahiert, die in der Lücke liegen. Die Sequenzinformation dieser Partner-Reads war Grundlage zur Erstellung der Primer für die „long range“ PCRs, die letztlich zum Schließen der Lücke führten. Mithilfe von *Co-Gen* gekoppelten Markern (Kapitel 3.7) konnten einzelne BAC-Klone dem kolumnaren und dem nicht kolumnaren Chromosom zugeordnet werden. Überlappende BAC-Klone wurden entsprechend ihres chromosomalen Ursprungs zu Metacontigs zusammengefügt. Insgesamt konnten 18 Metacontigs gebildet werden, darunter acht des kolumnaren und sechs des nicht kolumnaren Chromosoms. Darüber hinaus wurden die Metacontigs und „long range“ PCR-Produkte in einen etwa 850 kb langen Gesamtcontig zusammengefasst.

In Abbildung 3.10 sind die identifizierten und sequenzierten BAC-Klone aus P28, die anhand der GD Referenz und überlappenden Bereichen in Metacontigs und letztlich in einen vorläufigen Gesamtcontig angeordnet werden konnten, dargestellt. Zudem zeigt die Abbildung die Anordnung der Contigs des GD-Genoms und eine grobe Übersicht der dort annotierten Gene.

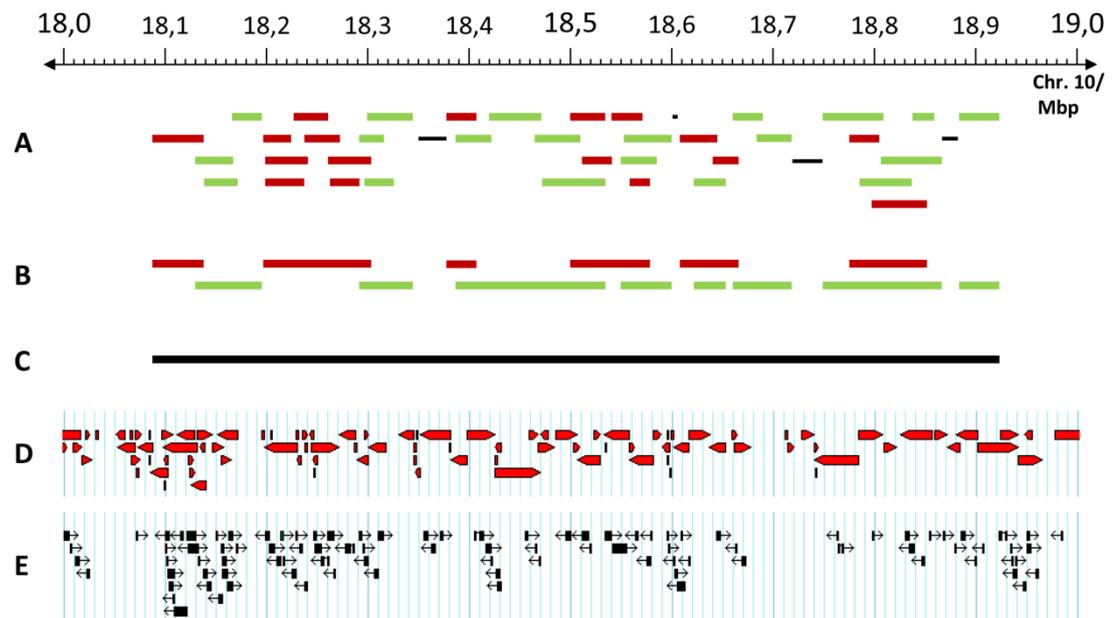


Abbildung 3.10 Co-Zielregion auf Chromosom 10, Position 18 bis 19 Mbp

A) zeigt alle BAC-Klone und „long range“ PCR-Produkte aus der Zielregion angeordnet nach ihren BLAST-Treffern. grün: kolumnares Chromosom (BAC); rot: nicht kolumnares Chromosom (BAC); schwarz: kolumnares oder nicht kolumnares Chromosom („long range“ PCR-Produkt); In B) sind die BAC-Klone entsprechend ihrer chromosomalen Zuordnung zusammengefasst als Metacontigs dargestellt. Diese wurden unter anderem für den Vergleich der BAC-Sequenzen mit den Sequenzen des veröffentlichten Apfelgenoms verwendet. C) zeigt den Gesamtcontig, also die Consensus-Sequenz, die entsteht, wenn man alle Sequenzen der BAC-Klone und „long range“ PCR-Produkte beider Chromosomen gemeinsam assembliert. D) stellt die Anordnung der Contigs des GD-Genoms in der Zielregion dar, sowie die dort annotierten Gene (E). Die Grafiken D und E stammen aus dem „apple genome browser“.

Anhand genomischer DNA-Sequenzen der Apfelsorte P28 in Form von Illumina „mate pair“-Reads wurde überprüft, ob die Metacontigs und der Gesamtcontig korrekt zusammengesetzt sind und somit die genomische Struktur widerspiegeln. Die zur Verfügung stehenden Reads mit einer statistischen Genomabdeckung von etwa 26-fach, reichten nicht aus, um den Gesamtcontig bei maximaler Mapping-Stringenz lückenlos abzudecken. Allerdings kann durch Verwendung der „mate pair“-Reads eine Aussage darüber getroffen werden, ob die grundsätzliche Zusammensetzung korrekt ist. Insgesamt konnten alle Metacontigs verifiziert werden. Einer genaueren Überprüfung bedurfte es der Region zwischen 240 und 250 kb im Gesamtcontig, die in Abbildung 3.11 dargestellt ist.

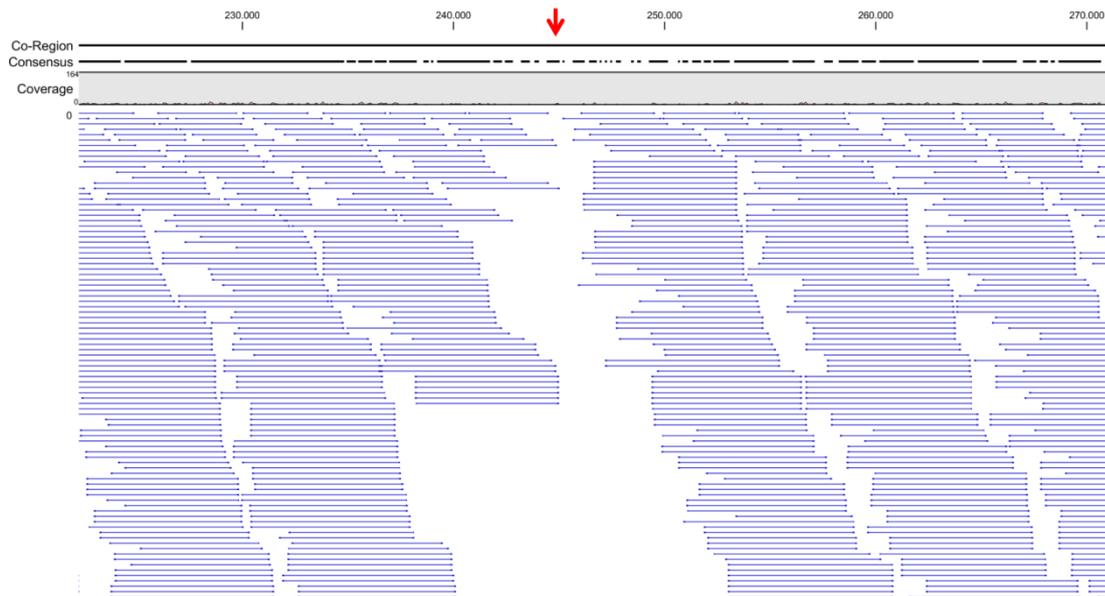


Abbildung 3.11 Mapping der „mate pair“-Reads gegen den Gesamtcontig

Abgebildet ist ein Ausschnitt des Mappings der „mate pair“-Reads gegen den Gesamtcontig. Der rote Pfeil markiert die einzige Position des Mappings, die von keinem „mate pair“-Read überspannt wird.

Es handelt sich dabei um die einzige Position im Gesamtcontig, die von keinen „mate pair“-Reads überspannt wird. Dies ist ein Anzeichen dafür, dass diese Stelle im Genom möglicherweise so nicht existiert. An der Position handelt es sich um die etwa 150 bp lange Überlappung zwischen dem Metacontig_3c und dem „long range“ PCR-Produkt LR_Mc_3c_3nc. Das „long range“ PCR-Produkt zeigt starke Übereinstimmung mit dem im GD-Genom annotierten Contig MDC018604.401, was die Vermutung nahe legt, dass es sich um ein Amplifikat des nicht kolumnaren Chromosoms handelt. Der Metacontig_3c stammt hingegen vom kolumnaren Chromosom und weist daher im Vergleich zu den Contigs des GD-Genoms starke Unterschiede auf (Abbildung 3.12). Möglichweise ist dies eine Ursache für die fehlenden Reads. Bisherige Versuche, den Bereich mittels Southern Blot Analysen oder überspannenden „long range“ PCRs zu überprüfen, schlugen fehl, sodass die genomische Organisation an dieser Position nicht abschließend geklärt werden konnte.

Der erstellte und abgesehen von der einen unklaren Position verifizierte Gesamtcontig hat eine Länge von 848.024 bp. Insgesamt stammen 539.688 bp (63,64 %) vom kolumnaren Chromosom, 400.968 bp (47,28 %) vom nicht kolumnaren Chromosom und 93.074 bp (10,98 %) aus „long range“ PCR-Produkten, bei denen keine chromosomale Zuordnung möglich ist.

3.5.2 Sequenzvergleiche der Apfelsorten P28 und GD

Um Sequenzunterschiede zwischen dem kolumnaren und dem nicht kolumnaren Chromosom zu identifizieren, wurden die Metacontigs der heterozygot kolumnaren Apfelsorte P28 (Abbildung 3.10 B) untereinander und mit den Contigs des GD-Genoms (Abbildung 3.10 D) verglichen. Die Ergebnisse sind in den Abbildungen 3.12 bis 3.14 grafisch dargestellt.

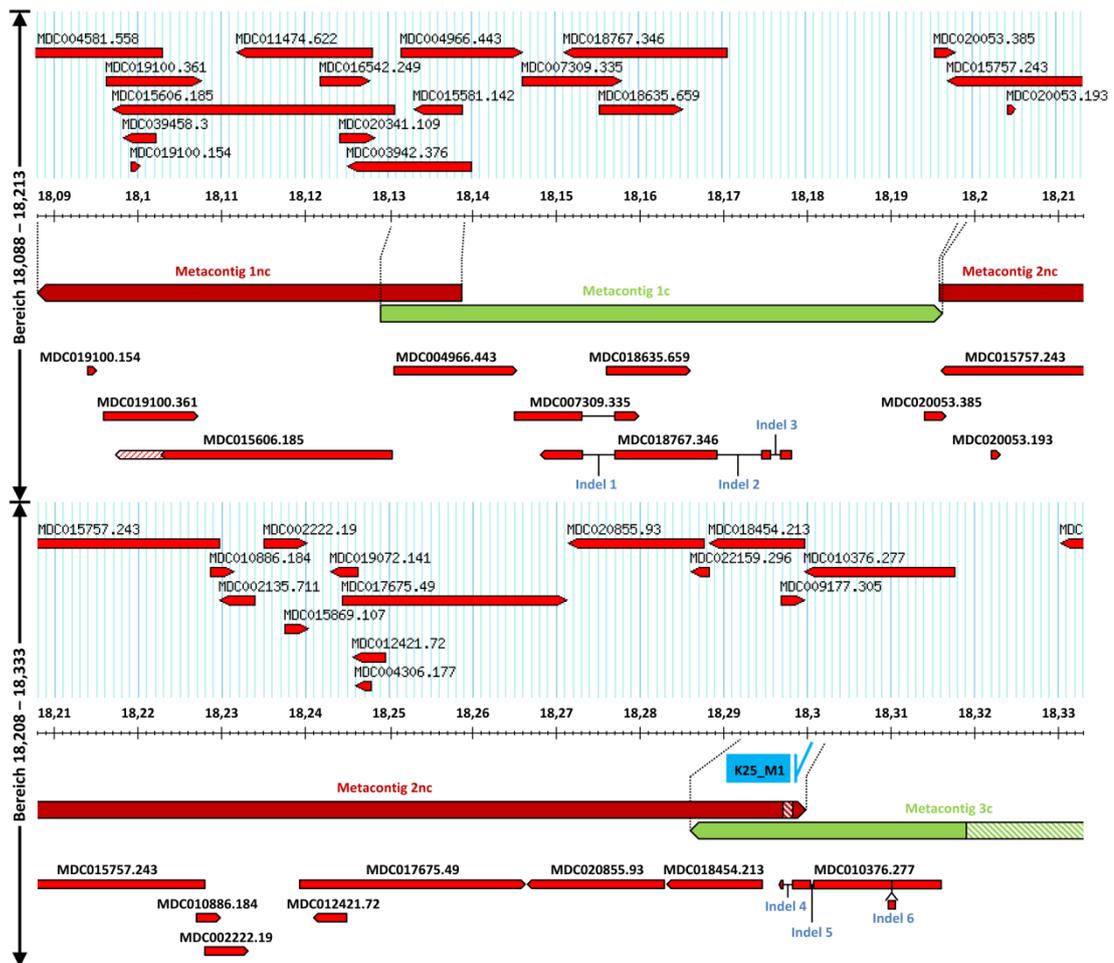


Abbildung 3.12 Metacontigs vs. GD-Genom Teil I

Gezeigt sind jeweils 125 kb große Ausschnitte aus der Co-Zielregion, die um jeweils 5 kb überlappen. Jeder Bereich besteht aus einem Auszug aus dem „apple genome browser“, in dem die annotierten Contigs (MDCs) und die entsprechenden Positionen im GD-Genom zu sehen sind. Darunter befinden sich die selbst erstellten Metacontigs, deren Lage in Bezug auf das GD-Genom durch gestrichelte Linien gekennzeichnet ist. Indels zwischen den Metacontigs in überlappenden Bereichen sind schraffiert dargestellt. Im untersten Teil sind die Contigs des GD-Genoms entsprechend ihrer Lage in Bezug auf die Metacontigs abgebildet. Schraffierte Flächen in Bereichen der MDCs stehen für fehlende Übereinstimmungen zwischen den MDCs und den dazugehörigen Metacontigs. Darüber hinaus sind die selbst erstellten Co-Gen gekoppelten Marker in den Grafiken in blau dargestellt (Kapitel 3.7). Im Fall von Metacontig 3c handelt es sich bei dem schraffierten Bereich nicht um ein Indel, da eine mögliche Referenz fehlt (Position 18,318 bis 18,33 Mbp).

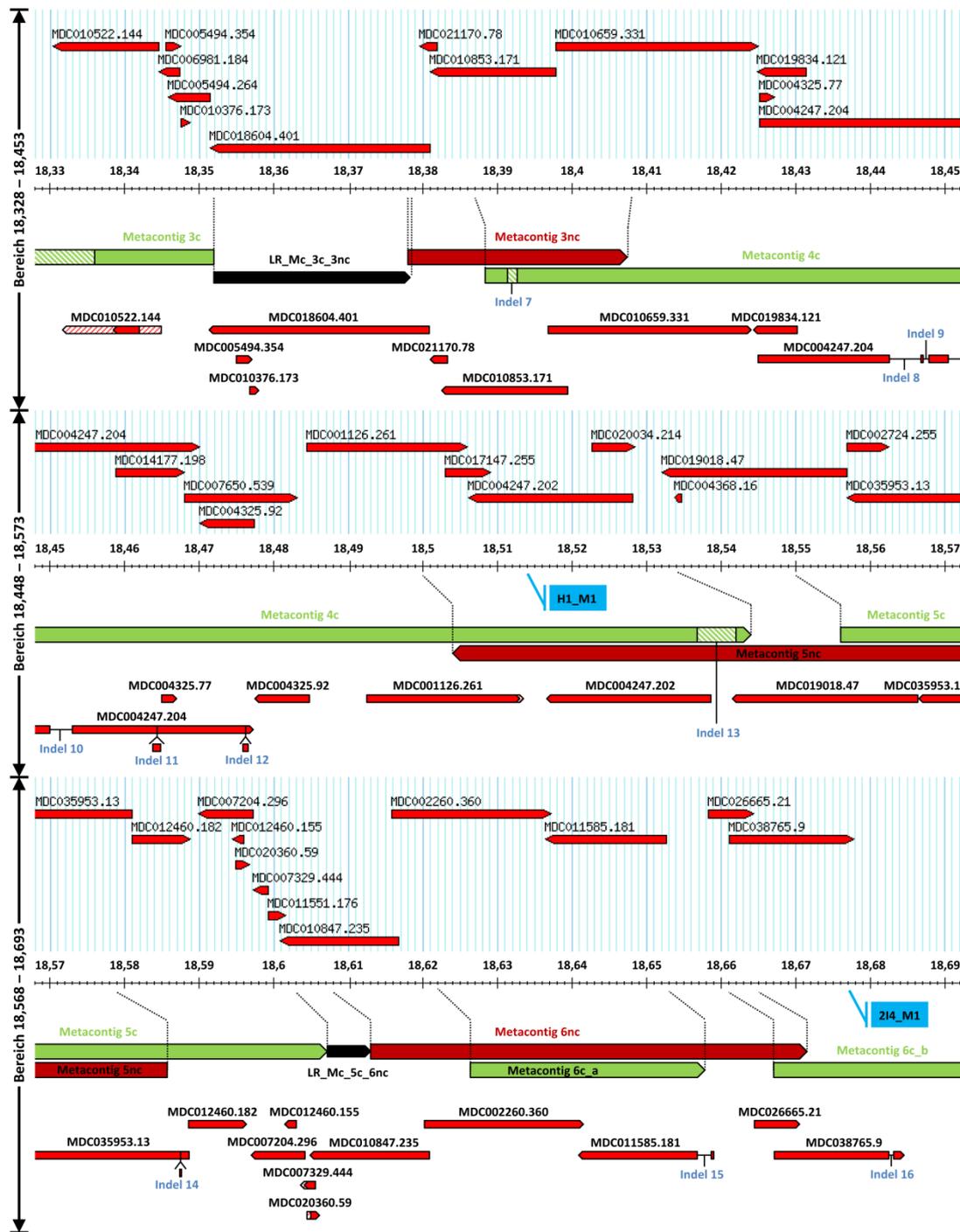


Abbildung 3.13 Metacontigs vs. GD-Genom Teil II

Die Erklärung der Abbildung ist der Beschriftung der Abbildung 3.12 zu entnehmen.

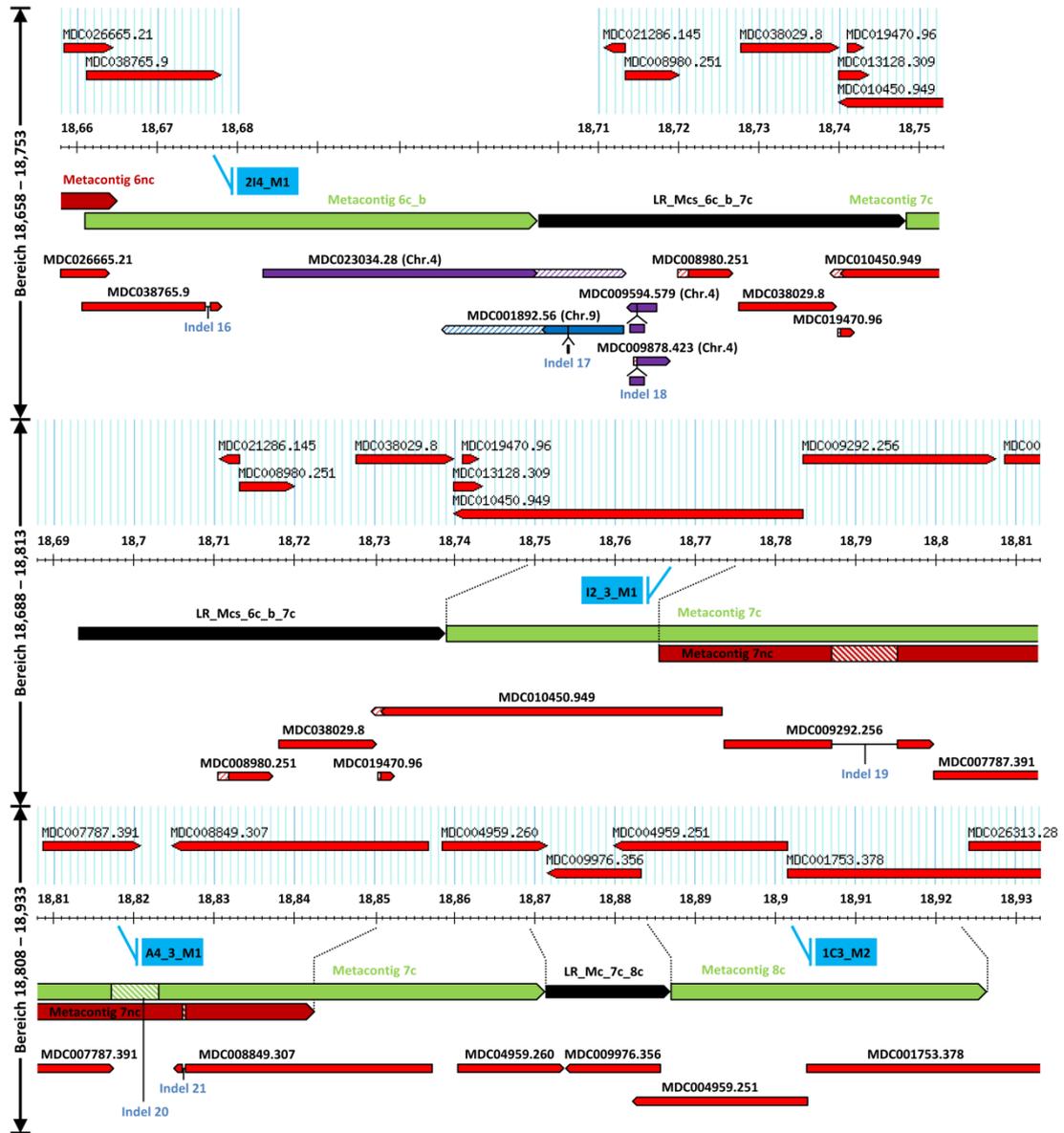


Abbildung 3.14 Metacontigs vs. GD-Genom Teil III

Die Erklärung der Abbildung ist der Beschriftung der Abbildung 3.12 zu entnehmen. Der obere Teil der Abbildung zeigt eine Vergrößerung der Region 18,658 bis 18,753 Mbp, da dort im GD-Genom kaum Contigs annotiert sind und in diesem Bereich drei MDCs neu kartiert werden konnten, die fälschlicherweise Chromosom 4 zugeordnet sind sowie ein MDC, der irrtümlich auf Chromosom 9 annotiert ist. Besonderes Augenmerk liegt auf Indel 19, da dies der einzige Unterschied zwischen dem kolumnaren und nicht kolumnaren Chromosom ist, der durch genomische Illumina-Datensätze der Apfelsorten McIntosh und McIntosh Wijkik bestätigt werden konnte.

Insgesamt konnten drei Contigs (MDCs) des GD-Genoms die auf Chromosom 4 und ein Contig, der auf Chromosom 9 annotiert war, nun eindeutig als auf Chromosom 10 lokalisiert bestimmt werden (Abbildung 3.14). Darüber hinaus wurden in der untersuchten Region (18,088 bis 18,923 Mbp) 21 auf Chromosom 10 annotierte GD-Contigs gefunden, die keine Sequenzübereinstimmung mit den Sequenzen der Metacontigs besitzen und somit vermutlich irrtümlich Chromosom 10 zugeordnet wurden. 11 GD-Contigs weisen lediglich Abschnitte auf, die keine Übereinstimmung mit den Metacontigs zeigen; diese Bereiche sind in den Abbildungen 3.12 bis 3.14 schraffiert dargestellt. Insgesamt konnten 21 Indels (> 100 bp) identifiziert werden, wobei 11 einen signifikanten Treffer bei BLASTn-Suchen gegen eine Datenbank transposabler Elemente (CENSOR) ergeben. Bei zwei weiteren Indels handelt es sich um Duplikationen, 8 Indels bleiben uncharakterisiert. Indel 15 ist der einzige Unterschied, der auf Sequenzvergleichen zwischen einem Metacontig des nicht kolumnaren Chromosoms und einem Contig des GD-Genoms beruht. Alle weiteren Indels beruhen entweder auf Unterschieden zwischen kolumnaren und nicht kolumnaren Metacontigs oder auf Unterschieden zwischen kolumnaren Metacontigs und Contigs des GD-Genoms (Tabelle 3.11). Diese Indels könnten potenziell mit der *Co*-Mutation assoziiert sein. Zur Überprüfung möglicher Unterschiede zwischen kolumnarem und nicht kolumnarem Chromosom wurden die Apfelsorten McIntosh und McIntosh Wjick verwendet, da der einzige genomische Unterschied zwischen diesen beiden Sorten sehr wahrscheinlich die *Co*-Mutation ist. Genomische Illumina-Datensätze der Apfelsorten McIntosh und McIntosh Wjick (Doktorarbeit Romina Petersen, IMSB) wurden gegen den Gesamtcontig kartiert. Anschließend wurden die Integrationsstellen der Indels untersucht. Werden diese von „paired reads“ überspannt, so kann man davon ausgehen, dass das Indel an dieser Position in der entsprechenden Apfelsorte vorhanden ist. Im gegenteiligen Fall ist von einem Fehlen des Indels auszugehen. Von Interesse sind Indels, die Unterschiede zwischen McIntosh und McIntosh Wjick aufweisen. In Tabelle 3.11 sind die Ergebnisse der Mappings und der Charakterisierung der Indels aufgeführt.

Tabelle 3.11 Identifizierte Indels in der Co-Zielregion

Indel	CENSOR Treffer	Quelle	Mapping McIntosh	Mapping McIntosh Wijcik
1	-	c ↔ GD	vorhanden	vorhanden
2	RTE-1B_Mad	c ↔ GD	vorhanden	vorhanden
3	DNA9-3_Mad	c ↔ GD	vorhanden	vorhanden
4	-	c ↔ nc u. GD	vorhanden	vorhanden
5	(Duplikation)	c ↔ GD	vorhanden	vorhanden
6	hAT-4_Mad	c ↔ GD	vorhanden	vorhanden
7	-	c ↔ nc	nicht vorhanden	nicht vorhanden
8	-	c ↔ GD	vorhanden	vorhanden
9	(Duplikation)	c ↔ GD	vorhanden	vorhanden
10	-	c ↔ GD	vorhanden	vorhanden
11	<i>Gypsy-36_Mad</i> LTR	c ↔ GD	vorhanden	vorhanden
12	DNA3-3D_Mad	c ↔ GD	vorhanden	vorhanden
13	RTE-1_Mad	c ↔ nc	nicht vorhanden	nicht vorhanden
14	RTE-1_Mad	c ↔ GD	vorhanden	vorhanden
15	-	nc ↔ GD	vorhanden	vorhanden
16	<i>Copia-59_Mad</i>	c ↔ GD	vorhanden	vorhanden
17	-	lr ↔ GD	vorhanden	vorhanden
18	<i>Gypsy-18_Mad</i>	lr ↔ GD	vorhanden	vorhanden
19	<i>Gypsy-44_Mad</i>	c ↔ nc u. GD	nicht vorhanden	vorhanden
20	RTE-1_Mad	c ↔ nc	nicht vorhanden	nicht vorhanden
21	-	c ↔ nc u. GD	vorhanden	vorhanden

c: kolumnarer Metacontig (Mc); nc: nicht kolumnarer Mc; lr: „long range“ PCR-Produkt

Der einzige Unterschied, der anhand der genomischen Daten von McIntosh und McIntosh Wijcik bestätigt wird, liegt in der Ab- bzw. Anwesenheit des Transposons *Gypsy-44* (Indel 19) an Position 18,79 auf Chromosom 10 (Abbildung 3.14). Nur in der Apfelsorte McIntosh Wijcik ist das Transposon an dieser Stelle lokalisiert. Die genomische Struktur und die Integrationsstelle von *Gypsy-44* liegen in Form von Metacontig_7c und Metacontig_7nc für das kolumnare und nicht kolumnare Chromosom vor und wurden in Zusammenarbeit mit Romina Petersen (IMSB) genauer analysiert. In Abbildung 3.15 ist die genomische Situation dieses Locus schematisch dargestellt. *Gypsy-44* liegt in 3'-5'-Orientierung im 5'-LTR eines *Gypsy-33* Retrotransposons.

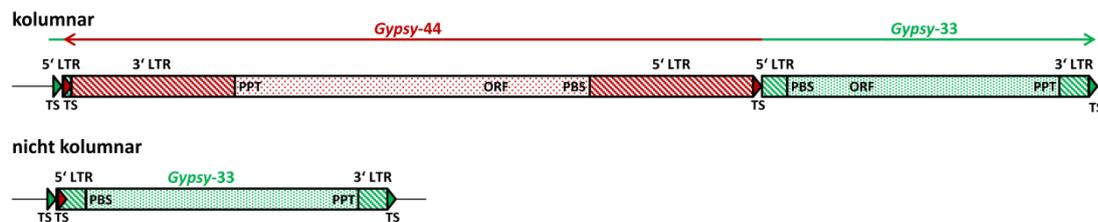


Abbildung 3.15 Gypsy-44 in der Region um 18,79 Mbp auf Chromosom 10

Der obere Teil der Abbildung zeigt die genomische Lage des Transposons *Gypsy-44* (rot) auf Chromosom 10 für das kolumnare Chromosom. Das Transposon befindet sich im 5'-LTR eines weiteren transposablen Elements (*Gypsy-33*, grün). Weitere Elemente der Transposons sind eingezeichnet, wie „target site (TS)“, „primer binding site (PBS)“, „polypurine tract (PPT)“, „open reading frame (ORF)“ und „long terminal repeat (LTR)“. Der untere Teil der Abbildung spiegelt die genomische Situation der Region um 18,79 Mbp im nicht kolumnaren Chromosom wider.

Bei *Gypsy-44* und *Gypsy-33* handelt es sich um LTR-Retrotransposons. Die typischen Elemente der Transposons wie „long terminal repeats“, „target site duplications“, „primer binding sites“ und „polypurine tracts“ konnten identifiziert werden. Beide Transposons enthalten einen ORF. Dabei handelt es sich im Fall von *Gypsy-33* vermutlich um ein gag-verwandtes Gen. Bei dem ORF von *Gypsy-44* konnte jedoch keine Funktion oder Domäne identifiziert werden. Da die für die Transposition benötigten Gene nur teilweise oder gar nicht vorhanden sind, handelt es sich bei *Gypsy-44* und *Gypsy-33* vermutlich um nicht-autonome Elemente. Um zu verifizieren, dass das Transposon *Gypsy-44* nur bei kolumnaren Apfelsorten vorkommt, wurden Primer entworfen, die die rechte und linke Integrationsstelle sowie das komplette Transposon überspannen. Es konnte gezeigt werden, dass *Gypsy-44* an dieser Position ausschließlich bei den heterozygot kolumnaren Apfelsorten McIntosh Wijcik und P28 auftritt, bei den nicht kolumnaren Apfelsorten McIntosh und A14 jedoch fehlt. Die detaillierten Ergebnisse der PCRs sind in Kapitel 3.7.4 dargestellt.

Alle anderen Indels aus Tabelle 3.11 sind an den entsprechenden Positionen sowohl in McIntosh wie auch in McIntosh Wijcik vorhanden oder nicht vorhanden und somit vermutlich nicht mit der Co-Mutation assoziiert.

3.5.3 Annotation von Genen und transposablen Elementen

Zur Annotation der Gene im Gesamtcontig wurden die CDS der GDR-Datenbank, sowie zwei Transkriptomdatensätze des Sprossapikalmeristems der Apfelsorte P28 von Clemens Krost (2012) verwendet. In Abbildung 3.16 sind die annotierten Gene im Gesamtcontig abgebildet.

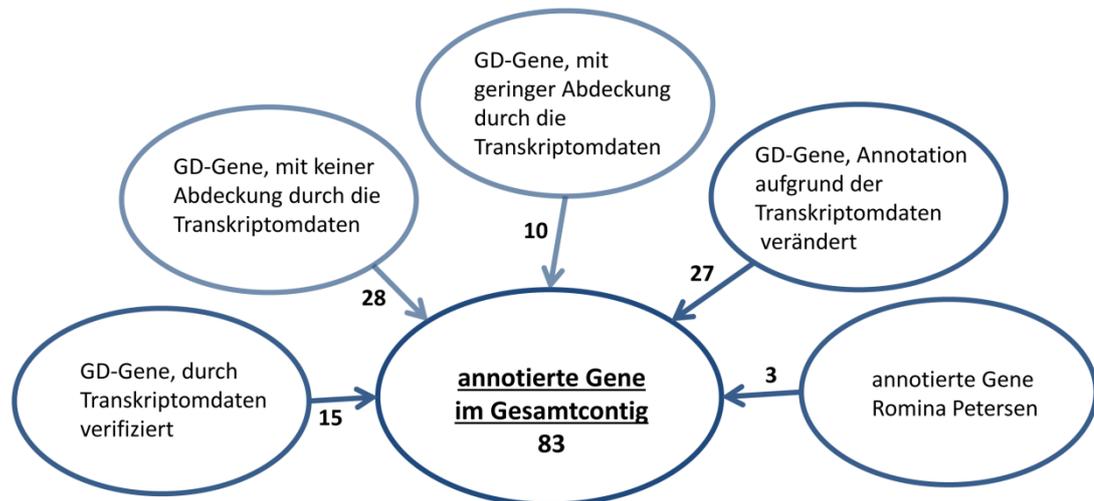


Abbildung 3.16 Annotierte Gene im erstellten Gesamtcontig

Die Grafik zeigt, welche Gruppen von Genen in welcher Anzahl im Gesamtcontig identifiziert werden konnten. Die einzelnen Gene sind gruppiert in solche, die durch die Transkriptomdaten (TD) nur gering oder gar nicht abgedeckt sind, in Gene, die durch die TD verifiziert werden konnten und in solche, deren Annotation aufgrund der TD geändert wurde. Drei Gene wurden von Romina Petersen (IMSB) identifiziert und stammen somit nicht aus den im GD-Genom annotierten CDS.

Insgesamt konnten 83 Gene im Gesamtcontig annotiert werden. 38 dieser Gene sind durch die Transkriptomdaten gar nicht oder zu gering abgedeckt, um Aussagen über die Struktur der Gene zu machen zu können. Aus diesem Grund wurden diese Gene entsprechend ihrer Annotation im GD-Genom übernommen. Bei 15 Genen konnte die Annotation des GD-Genoms durch die Transkriptomdaten verifiziert werden. Drei Gene wurden durch Romina Petersen im Rahmen ihrer Doktorarbeit charakterisiert. Bei 27 Genen wurde anhand der Transkriptomdaten die Exon-Intron-Struktur im Vergleich zur GD-Annotation korrigiert und entsprechend annotiert. In einigen Fällen besteht die Vermutung, dass es sich bei den annotierten CDS des GD-Genoms möglicherweise um zwei oder sogar drei getrennte Transkripte handelt, da die 3'-Bereiche der Gene in den Mappings überproportional stark abgedeckt sind. Dies ist sehr wahrscheinlich auf die zweimalige Selektion mittels oligo(dT)-Affinitätsaufreinigung (Poly-A⁺-RNA und anschließende Amplifikation) bei der cDNA-

Synthese zurückzuführen (Krost 2012). Eine vollständige Liste der selbst annotierten Gene befindet sich im elektronischen Anhang.

In Bezug auf das GD-Genom erstreckt sich der erstellte Gesamtcontig von Position 18,088 bis 18,923 Mbp auf Chromosom 10. In diesem Bereich sind 84 Gene vorhergesagt. Detaillierte Angaben zu diesen Genen sind Tabelle 3.12 zu entnehmen. Die ersten beiden Spalten beziehen sich auf die CDS-Datensätze der GDR-Datenbank, die Spalten drei bis fünf auf die Überprüfung der Annotationen basierend auf den zwei Transkriptomdatensätzen des Sprossapikalmeristems der Apfelsorte P28. Die detaillierte Analyse der Gene in der Zielregion und Genexpressionsstudien zwischen P28 und A14 sowie zwischen McIntosh und McIntosh Wjck waren nicht Teil dieser Arbeit, sondern der Arbeiten von Clemens Krost (2012) und Romina Petersen.

Tabelle 3.12 Im GD-Genom annotierte Gene in der Zielregion

CDS	Bemerkung 1	TD	geänderte Annotation	Bemerkung 2
MDP0000423953		x		
MDP0000423954		x		
MDP0000244450	Pseudogen	✓x	MDP0000244450-like	
MDP0000320862	identisch mit MDP0000244450, zusätzlicher Teil nicht Chr.10	✓		
MDP0000423958		✓		
MDP0000320861	Pseudogen	✓x	MDP0000320861-like, -like2 und -like3	vermutlich 3 getrennte Transkripte
MDP0000232611	3'-Splice-Site vor Exon 14 fehlt	✓x	MDP0000232611-like	Exon 1 bis 7 aus GD-Annotation, Intron in 3'-UTR
MDP0000286153				
MDP0000303750	Contigs im GD-Genom falsch annotiert (MDC011474.622 und			
MDP0000169714	MDC003942.376), Gene nicht annotiert			
MDP0000169715				
MDP0000320859		✓x	MDP0000320859-like	
MDP0000326734		x		
MDP0000169210		✓		
MDP0000436968	Gene identisch	✓		
MDP0000706741		✓		
MDP0000304842	5'-Splice-Site nach Exon 2: CG,	✓x		
MDP0000303255	3'-Splice-Site vor Exon 4: AC, Splice-Variante	✓x		
MDP0000238679	Splice-Variante	✓x	MDP0000304842-like	2 Introns im 3'-UTR
MDP0000301481	3'-Splice Site vor Exon 4: AC, BLAST-Treffer Chr. 0, Splice-Variante	✓x		
MDP0000303254		✓x	MDP0000303254-like	
MDP0000304843	Splice-Variante zu MDP0000303254	✓x		
MDP0000208936	Pseudogen	✓		
MDP0000350168		✓		
MDP0000205935		✓		
MDP0000205933	3'-Splice-Site vor Exon 2: AT	x		
MDP0000908448		✓		
MDP0000301011		x		
MDP0000301010		✓x	MDP0000301010-like	Splice-Variante identifiziert
MDP0000205928		✓x	MDP0000205928-like	
MDP0000165198	Gene identisch			
MDP0000265747		x		

x: nicht exprimiert ✓: exprimiert, geringe Abdeckung ✓✓: exprimiert, wie annotiert ✓x: exprimiert, geänderte Annotation TD: Transkriptomdaten

CDS	Bemerkung 1	TD	geänderte Annotation	Bemerkung 2
MDP0000214784		✓ x	MDP0000214784-like	Intron in 3'-UTR, 2 Varianten an 3'-UTRs
MDP0000294006	Gen nicht annotiert, kein Start- und Stopp-Codon			
MDP0000641747		✓ x	MDP0000641747-like	
MDP0000307618		✓ x	MDP0000307618-like	
MDP0000307619		x		
MDP0000308878		✓		
MDP0000711589		✓		
MDP0000343332	Gen nicht annotiert, kein Start- und Stopp-Codon			
MDP0000771587		✓ x	MDP0000771587-138428-like	vermutlich ein langes Transkript
MDP0000138428	SNP im Stopp-Codon: TTA	✓ x		
MDP0000138427		✓ x	MDP0000138427-like	
MDP0000214259		x		
MDP0000214258		✓		
MDP0000214257		✓		
MDP0000255817	Pseudogen (BLAST-Treffer Chr. 11)	✓		
MDP0000697030		✓ x	MDP0000697030-like	
MDP0000191925		✓		
MDP0000191926		✓		
MDP0000253723		x		
MDP0000170185	Splice-Variante zu MDP0000253723			
MDP0000342057	BLAST-Treffer auf Chr. 5	✓ x	MDP0000342057-like	Pseudogen
MDP0000513356	Exon 3 (~50% des Gens) nicht vorhanden, Gen nicht annotiert			
MDP0000784187		✓		
MDP0000136351	Contig im GD-Genom falsch annotiert (MDC007650.539), Gen nicht annotiert			
MDP0000157859	Pseudogen	✓ x	MDP0000157859-like	
MDP0000706530		✓		
MDP0000271350		✓ x	MDP0000271350-like und -like2	vermutlich 2 getrennte Transkripte
MDP0000143705		✓ x	MDP0000143705-like	
MDP0000254096		✓ x	MDP0000254096-like und -like2	vermutlich 2 getrennte Transkripte
MDP0000234557		✓		
MDP0000326311	Gen nicht annotiert, kein Start- und Stopp-Codon			
MDP0000897594		✓ x	MDP0000897594-like	
MDP0000136858		x		

x: nicht exprimiert ✓: exprimiert, geringe Abdeckung ✓✓: exprimiert, wie annotiert ✓ x: exprimiert, geänderte Annotation TD: Transkriptomdaten

CDS	Bemerkung 1	TD	geänderte Annotation	Bemerkung 2
MDP0000524262		x		
MDP0000284965	Pseudogen	✓ x	MDP0000284965-like und -like2	vermutlich 2 getrennte Transkripte; beide PolyA-Signale vorhanden
MDP0000367163		✓✓		
MDP0000329966		✓✓		
MDP0000508371		✓✓		
MDP0000942873		✓✓		
MDP0000186457		x		
MDP0000324960	Pseudogen (BLAST-Treffer Chr. 7)	✓		
MDP0000855671		x		
MDP0000154627		x		
MDP0000369268	BLAST-Treffer Chr. 11	x		
MDP0000273398	MDP von Chromosom 9	x		
MDP0000286915		x		
MDP0000187369		x		
At1g08530-like	von Romina Petersen annotiert	✓✓		
MDP0000766466	Stopp-Codon: TAG (nicht kolumnares Chr.), CAG (kolumnares Chr.)	x		
Downy Mildew Resistant 6-like	von Romina Petersen annotiert	✓		
MDP0000362327	BLAST-Treffer auf Chr. 0	x		
MDP0000464120	BLAST-Treffer Chr. 1	x		
MDP0000927098		x		
MDP0000187760	SNP verändert Stopp-Codon	x		
MDP0000927091		✓✓		
MDP0000287209	Start-Codon: ATG (nicht kolumnares Chr.), ATA (kolumnares Chr.)	x		Exon 1 in TD nicht abgedeckt
Auxin-induced protein 5NG4-like	von Romina Petersen annotiert	✓✓		Splice-Variante identifiziert
MDP0000912172		✓		
MDP0000139773		x		
MDP0000934869		x		
MDP0000934866		✓ x	MDP0000934866-like	
MDP0000163720		✓		
MDP0000269126	nur ersten zwei Exons vorhanden	x		

x: nicht exprimiert ✓: exprimiert, geringe Abdeckung ✓✓: exprimiert, wie annotiert ✓ x: exprimiert, geänderte Annotation TD: Transkriptomdaten

Transposable Elemente machen 42,4 % des Apfelgenoms aus (Velasco *et al.* 2010). Zur Identifizierung transposabler Elemente im Gesamtcontig wurden die Programme CENSOR und LTR-Finder verwendet. Insgesamt wurden im Gesamtcontig 160 Transposons und ein Virus detektiert und annotiert. Die Klassifizierung transposabler Elemente ist variabel, meist werden sie in die zwei Klassen Retrotransposons (Klasse I) und DNA-Transposons (Klasse II) unterteilt (Finnegan 1989). Die Retrotransposons lassen sich wiederum in die Gruppen LTR-Retrotransposons und Non-LTR-Retrotransposons untergliedern. Die 160 identifizierten transposablen Elemente unterteilen sich in 66 DNA-Transposons, 68 LTR-Retrotransposons und 26 Non-LTR-Transposons (Abbildung 3.17).

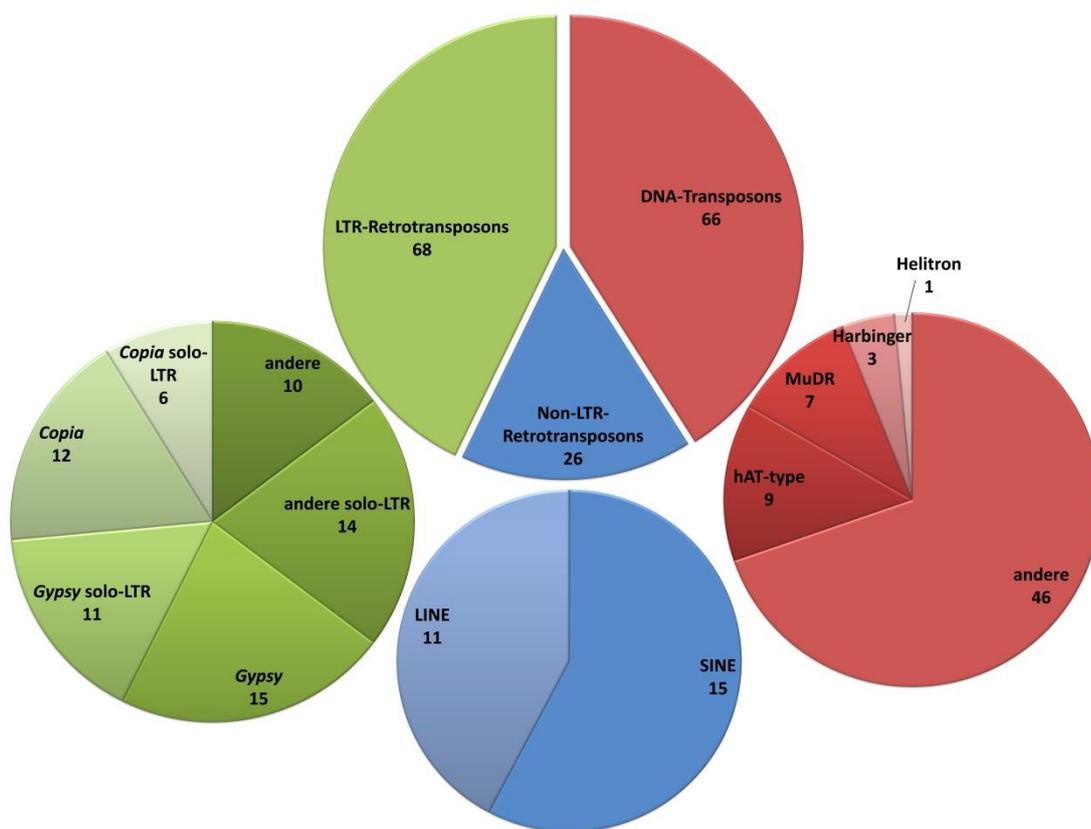


Abbildung 3.17 Annotierte transposable Elemente in der Zielregion

Die annotierten transposablen Elemente sind in dieser Abbildung in die Klasse der DNA-Transposons (rot) und Retrotransposon unterteilt, wobei die Retrotransposons nochmals in LTR- (grün) und Non-LTR-Retrotransposons (blau) untergliedert werden. Zudem sind aus der Abbildung die Vertreter der einzelnen Klassen nochmals in Untergruppen eingeteilt. Die jeweilige Anzahl der annotierten transposablen Elemente ist hinter der Bezeichnung gezeigt.

Die LTR-Retrotransposons setzen sich aus den Hauptvertretern *Copia* und *Gypsy* sowie einer Gruppe uncharakterisierter LTR-Retrotransposons (andere) zusammen. *Gypsy*-LTR-Retrotransposons machen mit 15 den größten Teil dieser Gruppe aus, gefolgt von *Copia*-

LTR-Retrotransposons mit 12 Vertretern. Darüber hinaus wurden 31 solo-LTRs identifiziert. Diese entstehen bei Rekombinationsereignissen zwischen den LTRs der LTR-Retrotransposons (Vitte und Panaud 2005). Die Non-LTR-Retrotransposons lassen sich untergliedern in SINEs und LINEs. Es konnten 15 SINEs und insgesamt 11 LINEs identifiziert werden. Der Hauptvertreter der LINEs sind „retrotransposable elements (RTEs)“ (Malik und Eickbush 1998). Die DNA-Transposons bestehen zum größten Teil (46 von 66) aus uncharakterisierten transposablen Elementen (andere). Es konnten aber auch Vertreter der Gruppen hAT-type, MuDR, Harbinger und Helitrons identifiziert werden.

Die annotierten transposablen Elemente machen mit 337.972 bp etwa 40 % des Gesamtcontigs aus, dabei stellen die LTR-Retrotransposons mit rund 74 % den größten Anteil dar. In Abbildung 3.18 ist die Verteilung der annotierten transposablen Elemente bezüglich ihrer Größe dargestellt.

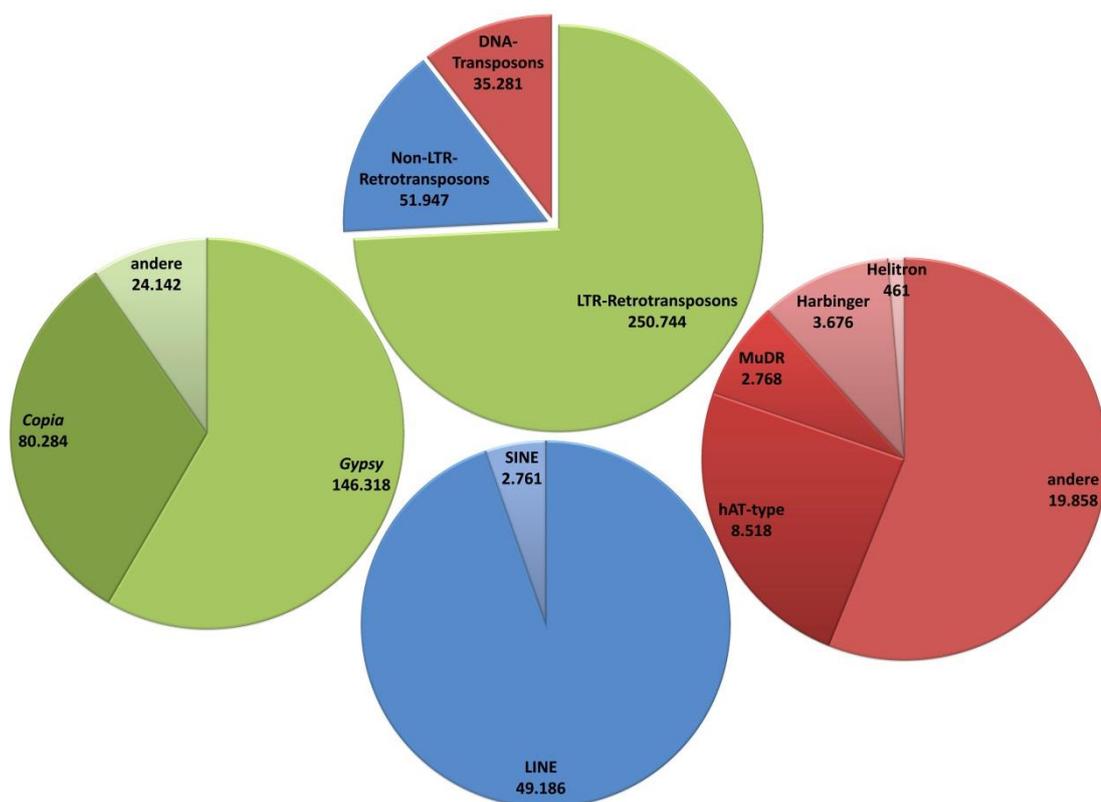


Abbildung 3.18 Größenverteilung der annotierten transposablen Elemente

Die Diagramme zeigen die Verteilung der transposablen Elemente bezüglich ihrer Größe untergliedert in die Klassen DNA-Transposons, LTR-Retrotransposons und Non-LTR-Retrotransposons. Die Größe der einzelnen Gruppen ist in Basenpaaren angegeben. Insgesamt machen die transposablen Elemente 337.972 bp und damit 40 % des Gesamtcontigs aus.

Von den 160 transposablen Elementen handelt es sich bei 12 lediglich um Teilfragmente transposabler Elemente, die annotiert wurden. Solo-LTRs, die durch ungleiche homologe Rekombination entstehen, spielen bei Betrachtung der „target site duplication (TSD)“ eine gesonderte Rolle. Findet das Rekombinationsereignis zwischen den LTRs eines Retrotransposons statt, entsteht ein solo-LTR mit konservierter TSD des Retrotransposons. Bei einem Ereignis zwischen den LTRs zweier Retrotransposons hingegen entsteht ein solo-LTR mit unterschiedlichen „target sites“ (Vitte und Panaud 2005). Bei den annotierten solo-LTRs ist lediglich in sechs von 32 Fällen eine konservierte TSD zu identifizieren. Abzüglich der Teilfragmente transposabler Elemente und der solo-LTRs verbleiben 116 annotierte transposable Elemente, von denen zehn keine TSD aufweisen, 14 zeigen eine TSD mit einem Nukleotidaustausch und 92 besitzen eine perfekte TSD. Außer der TSD sind insbesondere die LTRs bei LTR-Retrotransposons eine auffällige Struktureigenschaft. Die Sequenzen der LTRs sind bei der Integration des Transposons in der Regel identisch, weichen aber mit der Zeit unter anderem durch Substitution einzelner Basen immer mehr voneinander ab (SanMiguel *et al.* 1998). Den exakten Zeitpunkt der Integration zu bestimmen ist sehr schwierig, allerdings kann die Sequenzidentität genutzt werden, um ungefähr abschätzen zu können, ob ein LTR-Retrotransposon schon vor längerer Zeit oder erst kürzlich ins Genom integrierte. Die Sequenzübereinstimmung der LTRs der in dieser Arbeit annotierten Retrotransposons liegt durchschnittlich bei 91 %. Interessant ist jedoch das Retrotransposon *Gypsy-44* an Position 18,79 auf Chromosom 10: Die LTRs sind 100 %-ig identisch, was vermuten lässt, dass die Integration an dieser Position erst vor kurzem stattgefunden hat.

Im elektronischen Anhang befinden sich die detaillierten Ergebnisse der BLAST-Suchen mittels CENSOR gegen die Repeat-Datenbank „Repbse“ sowie weitere Details zu den annotierten transposablen Elementen. Zusammenfassend ist in den Abbildungen 3.19 und 3.20 der Gesamtcontig mit den annotierten transposablen Elementen und Genen dargestellt, um einen Überblick über deren Verteilung zu erhalten. Auffällig ist, dass transposable Elemente bevorzugt in genarmen Regionen vorkommen (Abbildung 3.19). Transposable Elemente, die in Genbereiche integrierten, liegen überwiegend in Intron-Bereichen. Zudem kommt es häufig vor, dass Transposons in andere Transposons integrieren (Abbildung 3.20).

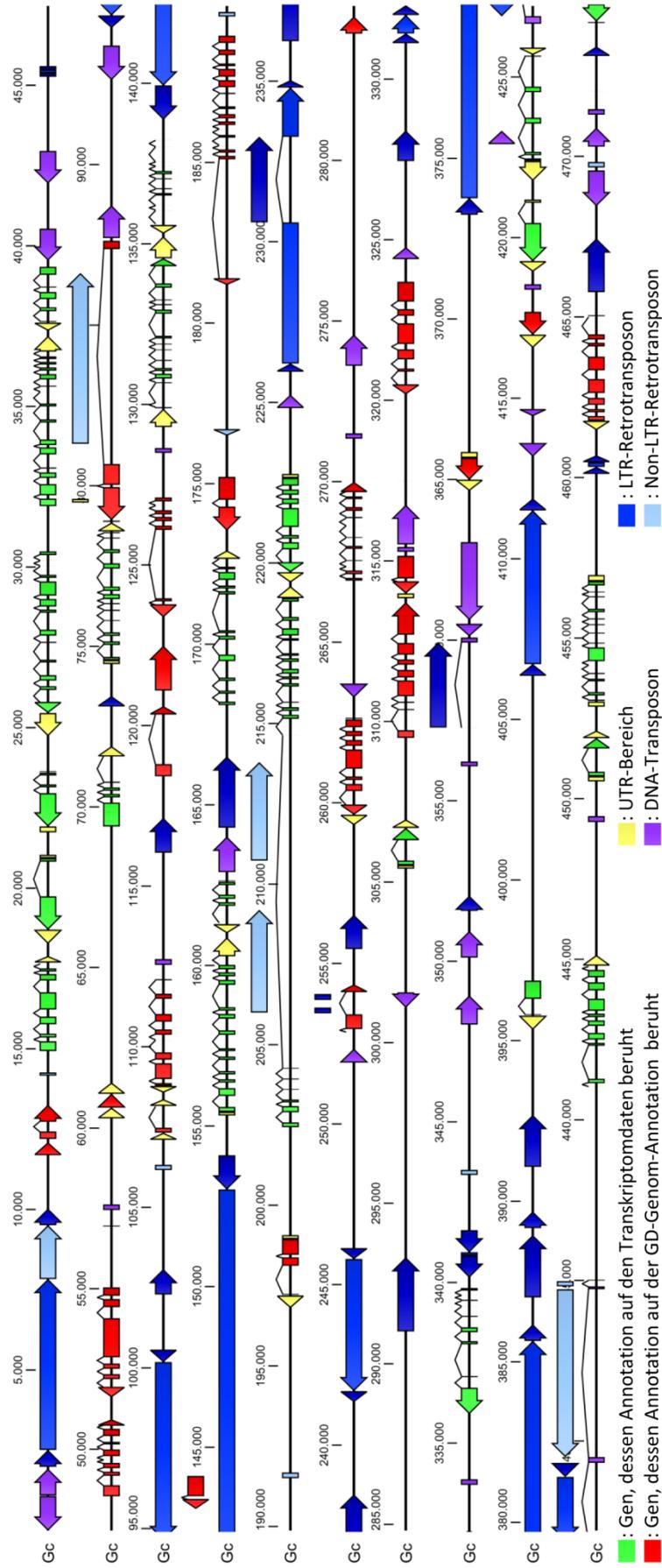


Abbildung 3.19 Annotierte Gene und transposable Elemente im Gesamtcontig Teil I

Der dargestellte erste Teil des Gesamtcontigs zeigt eine hohe Dichte an transposablen Elementen. Bei den LTR-Retrotransposons sind die LTRs gesondert annotiert.

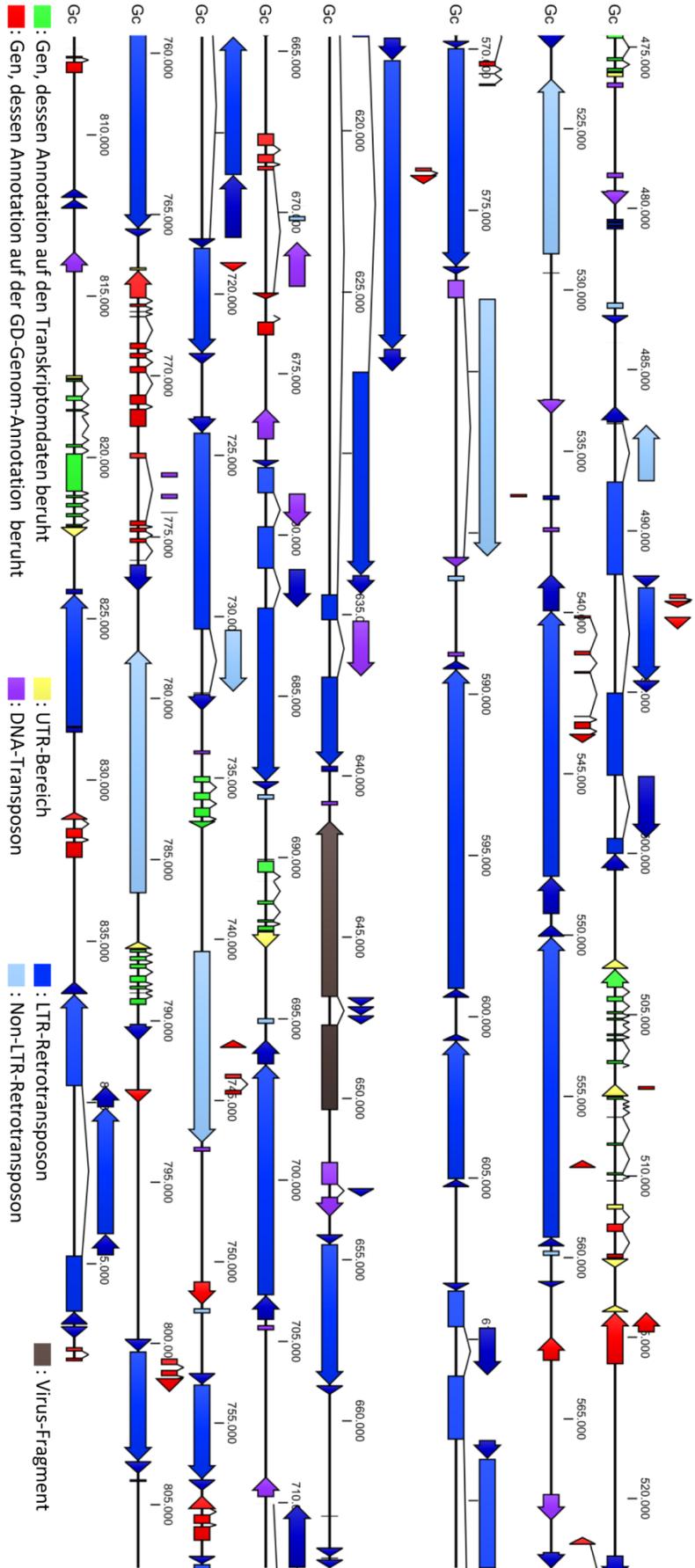


Abbildung 3.20 Annotierte Gene und transposable Elemente im Gesamtcontig Teil II

Der zweite Teil des Gesamtcontigs ist überwiegend durch Transposons charakterisiert. Auffällig ist zudem der Bereich von 608 bis 640 kb, in dem drei Retrotransposons ineinander verschachtelt liegen. Bei den LTR-Retrotransposons sind die LTRs gesondert annotiert.

3.7 Molekulare Co-Gen gekoppelte Marker

3.7.1 Identifikation Co-Gen gekoppelter Marker

Durch Sequenzvergleiche überlappender BAC-Sequenzen der heterozygot kolumnaren Apfelsorte P28 und Vergleiche dieser mit Sequenzen des GD-Genoms konnten zahlreiche Indels identifiziert werden. Diese können als potenzielle Marker zur Unterscheidung von kolumnaren und nicht kolumnaren Apfelbäumen dienen. Indel-flankierende Primer wurden erstellt und PCRs auf genomische DNAs der Sorten P28 und A14 durchgeführt. Die Benennung der Marker erfolgte gemäß den BAC-Klonen, in denen die Indels identifiziert wurden. In Tabelle 3.13 sind die potenziellen Marker mit erwarteter Fragmentgröße für das kolumnare und das nicht kolumnare Chromosom und die genomische Lage aufgelistet.

Tabelle 3.13 Potenzielle Co-Gen gekoppelte Marker

Marker	genomische Region	erwartete Fragmente	
		kolumnar	nicht kolumnar
1C3_M1	18,909 Mbp	348 bp	307 bp
1C3_M2	18,902 Mbp	414 bp	248 bp
A4_3_M1	18,818 Mbp	120 bp	152 bp
K2_3_M1	18,806 Mbp	642 bp	519 bp
I2_3_M1	18,767 Mbp	143 bp	126 bp
I2_3_M2	18,750 Mbp	286 bp	256 bp
2I4_M1	18,677 Mbp	945 bp	333 bp
2I8_M1	18,637 Mbp	238 bp	204 bp
D4_M1	18,562 Mbp	149 bp	179 bp
G5_M1	18,550 Mbp	225 bp	241 bp
H1_M1	18,514 Mbp	128 bp	206 bp
H1_M2	18,512 Mbp	261 bp	246 bp
K25_M1	18,304 Mbp	136 bp	157 bp

In der gelelektrophoretischen Auftrennung der PCR-Produkte sollte die heterozygot kolumnare Apfelsorte P28 zwei Banden zeigen, eine in Höhe des erwarteten kolumnar-spezifischen Produkts und eine Bande in Höhe des nicht kolumnar-spezifischen Produkts. Die homozygot nicht kolumnare Apfelsorte A14 sollte nur eine Bande in Höhe des nicht kolumnar-spezifischen Fragments aufweisen. Die Ergebnisse sind in Abbildung 3.21 dargestellt.

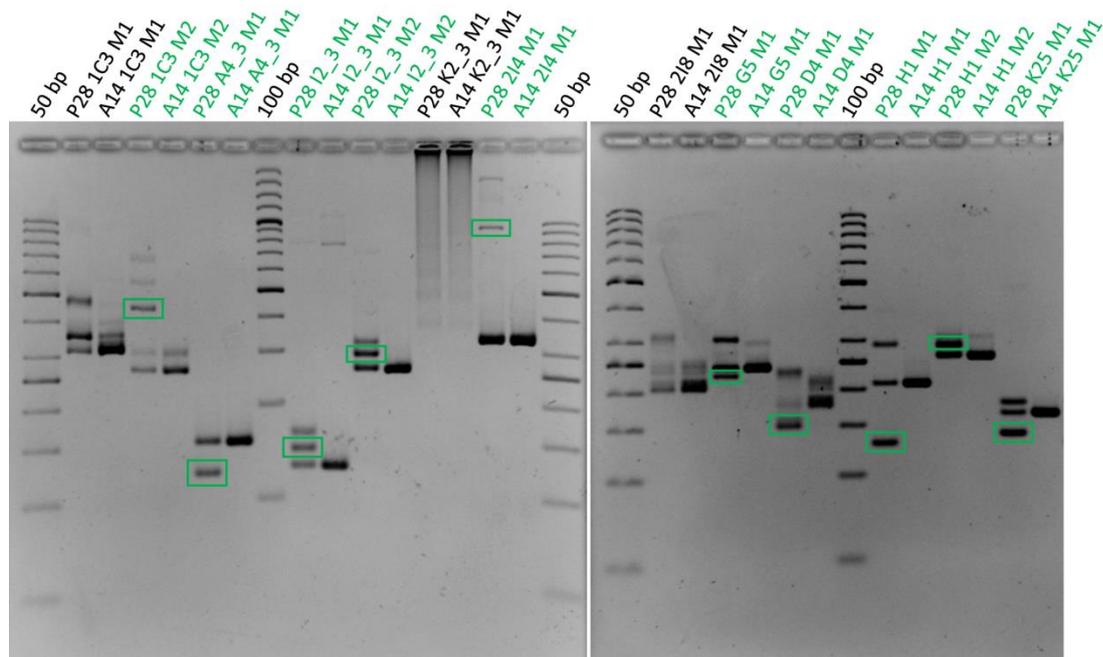


Abbildung 3.21 Potenzielle Co-Gen gekoppelte Marker

Die Bilder zeigen die Ergebnisse von Agarosegelelektrophoresen der PCR-Produkte der potenziellen Marker. Marker, die eindeutige Unterschiede zwischen A14 und P28 aufzeigen und sich somit möglicherweise zur Differenzierung zwischen kolumnarem und nicht kolumnarem Chromosom eignen, sind in grün dargestellt. Das kolumnar-spezifische PCR-Produkt ist grün eingerahmt. Als Molekulargewichtsstandards wurde die 50 bp und 100 bp Ladder genutzt.

Es konnten insgesamt 10 der 13 potenziellen Marker verifiziert werden. Sie zeigen die erwarteten eindeutigen Unterschiede zwischen A14 und P28. Allerdings tritt bei PCRs auf genomische P28-DNA mehrfach neben den erwarteten zwei Amplifikaten und möglichen unspezifischen Produkten eine weitere, distinkte dritte Bande auf (Abbildung 3.21). Da die Chromosomen-spezifischen Unterschiede in der Regel sehr gering sind (< 50 bp), liegt die Vermutung nahe, dass es sich bei der dritten Banden um ein Hybrid aus den anderen zwei Banden handelt. Die homologen Bereiche der zwei Amplifikate lagern sich während der Abkühlphase des PCR-Zyklus zusammen und die Insertion bildet eine Schleife. Dies wurde mittels S1-Nuklease-Verdau untersucht. Die S1-Nuklease entfernt solche einzelsträngigen Schleifen. In Abbildung 3.22 ist der Verdau für drei ausgewählte Marker gezeigt. Die in PCRs auf genomische P28-DNA auftretende dritte Bande ist nach S1-Nuklease-Behandlung nicht mehr zu detektieren. Damit konnte gezeigt werden, dass es sich bei der dritten Bande höchstwahrscheinlich um ein Hybrid der anderen zwei Banden gehandelt hat.

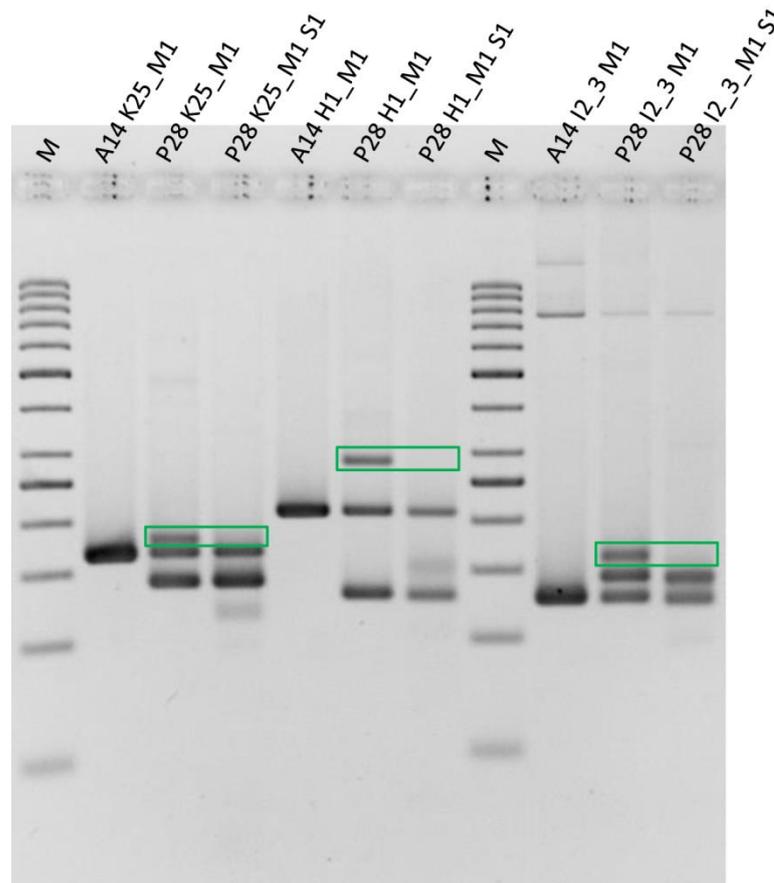


Abbildung 3.22 S1-Nuklease-Verdau ausgewählter Co-Gen gekoppelter Marker

Das Gelbild zeigt die Auftrennung der PCR-Produkte Co-Gen gekoppelter Marker nach erfolgtem S1-Nuklease-Verdau (S1) und anschließender Fällung. Zudem sind die PCR-Produkte ohne S1-Nuklease-Verdau gelelektrophoretisch aufgetrennt. Als Molekulargewichtsstandard (M) diente die 50 bp Ladder. Die Hybridbande und die entsprechende Stelle nach S1-Nuklease-Verdau sind grün eingerahmt.

Zur weiteren Markeranalyse wurden die sechs Marker K25_M1, H1_M1, I2_3_M1, A4_3_M1 und 1C3_M2 verwendet. Um abschließend zu überprüfen, ob die erhaltenen PCR-Produkte tatsächlich von den zu untersuchenden Loci stammen, wurden die Amplifikate einer PCR auf genomische DNA der Apfelsorten P28 und A14 in pGEM-T Vektor kloniert und anschließend sequenziert. Durch BLASTn-Suchen der Sequenzen gegen das GD-Genom konnten alle Amplifikate der sechs Marker der entsprechenden Region auf Chromosom 10 zugeordnet werden. Um zu überprüfen, ob es sich bei den identifizierten Markern tatsächlich um kolumnar-spezifische Unterschiede handelt, wurde eine Nachkommenschaft einer Kreuzung der Apfelsorten P28 und A14 analysiert. Entscheidend dabei war, ob die zweite Bande aus P28 nur bei Nachkommen mit kolumnarem Phänotyp in der PCR auftritt und somit kolumnar-spezifisch ist.

3.7.2 Feinkartierung der Zielregion

Die genomischen DNAs von 95 Nachkommen (48 kolumnare, 47 nicht kolumnare) einer Kreuzung der Apfelsorten P28 und A14 dienten der Analyse der *Co*-Gen gekoppelten Marker. Zunächst wurde jedoch die grobe Abschätzung der Zielregion (Kapitel 3.1) durch Tests der 95 Nachkommen mit diversen Markern verfeinert. Dazu wurden PCRs mit den publizierten Markern Ch03d11, Hi01a03, SCAR₂₁₆, SCAR₆₈₂ (Abbildung 1.3) und dem selbst erstellten Marker K25_M1 durchgeführt und mittels QIAxcel elektrophoretisch aufgetrennt. Die Ergebnisse sind im Anhang A3 abgebildet, die kompletten Daten befinden sich im elektronischen Anhang. Der Marker SCAR₆₈₂ wurde für die Auswertung verworfen, da die PCRs keine zuverlässig reproduzierbaren Ergebnisse lieferten. Der Marker Ch03d11 zeigt für alle kolumnaren und nicht kolumnaren Nachkommen das erwartete Bandenmuster mit Ausnahme des Nachkommen 8.21 N. Dieser Nachkomme zeigt einen normalwüchsigen Phänotyp (N), allerdings das kolumnare Bandenmuster. Dies entspricht einer Rekombinationsrate des Markers mit dem *Co*-Gen von 1,05 %. Der Marker Hi01a03 zeigt drei (7.99 C, 6.40 N und 7.17 N), der Marker SCAR₂₁₆ sechs (7.99 C, 6.40 N, 7.17 N, 8.17 C, 7.2 N und 8.8 N) Nachkommen, bei denen das Bandenmuster nicht dem Phänotyp entspricht. Dies kommt einer Rekombinationsfrequenz der Marker von 3,16 % und 6,32 % gleich. Der selbst hergestellte Marker K25_M1 zeigt hingegen bei allen PCRs das erwartete Bandenmuster und weist somit 100 % Kopplung mit dem *Co*-Gen auf.

Neben der Analyse des Markers K25_M1 wurde der Marker 1C3_M2 auf insgesamt 115 Nachkommen (55 kolumnare, 60 nicht kolumnare) einer Kreuzung der Apfelsorten P28 und A14 durch PCR und anschließende gelelektrophoretische Auftrennung getestet (Gelbilder nicht gezeigt). Alle Nachkommen zeigen das erwartete Bandenmuster, sodass der Marker 1C3_M2 wie bereits K25_M1 eine 100 %-ige Kopplung mit *Co* zeigt. Die noch verbleibenden Marker H1_M1, 214_M1, I2_3_M1 und A4_3_M1 wurden aufgrund des begrenzten Materials sowie aus Zeit- und Kostengründen auf insgesamt nur zehn Einzelnachkommen (5 kolumnare, 5 nicht kolumnare) getestet. Unter den zehn Nachkommen befinden sich die „Ausreißer“ der Marker Ch03d11 und Hi01a03 und zwei des Markers SCAR₂₁₆. Alle Marker zeigen bei den getesteten Nachkommen das erwartete Bandenmuster, wodurch man von einer 100 %-igen Kopplung mit dem *Co*-Gen ausgehen kann (Gelbilder nicht gezeigt). Die Ergebnisse bzw. die daraus abgeleitete Kopplungskarte sind in Abbildung 3.23 zusammengefasst.

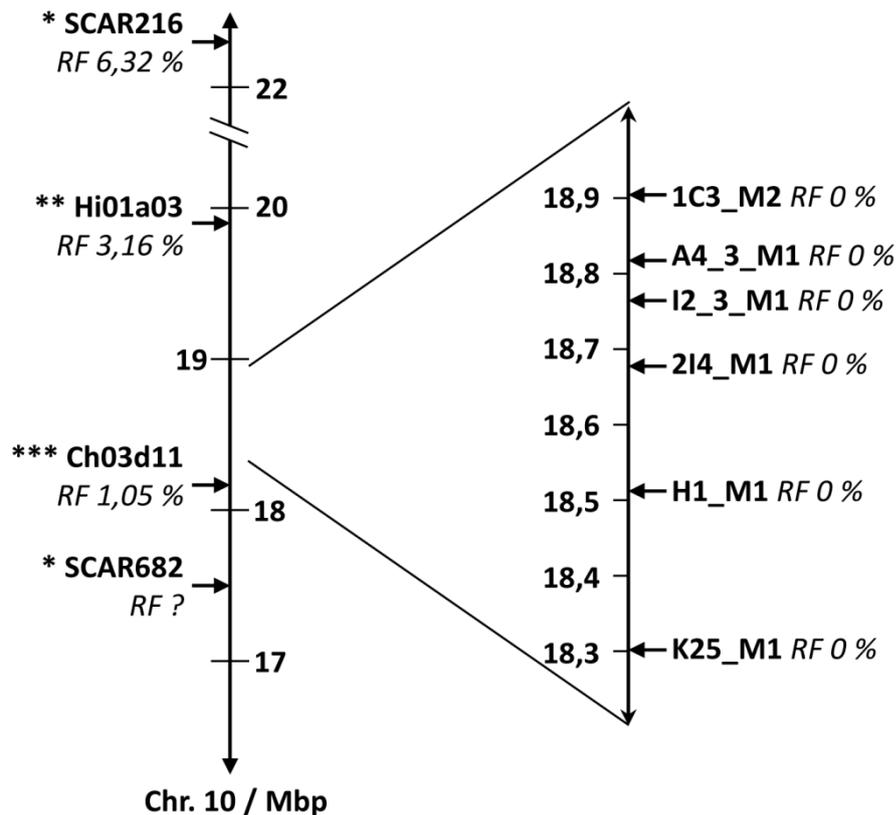


Abbildung 3.23 Feinkartierung der Co-Zielregion

Die Abbildung zeigt die Ergebnisse der Markeranalysen der Nachkommenschaft einer Kreuzung der Apfelsorten A14 x P28. Neben der chromosomalen Lage sind die Rekombinationsfrequenzen (RF) der Marker abgebildet. * Tian *et al.* (2005) ** Moriya *et al.* (2009) *** Liebhard *et al.* (2002)

Obwohl die zur Verfügung stehende Nachkommenschaft mit den getesteten 95 Nachkommen sehr klein ist, konnte anhand der Markeranalysen eine Grenze der Zielregion verfeinert werden. Wie in Abbildung 3.23 zu sehen, zeigt der Marker Ch03d11 eine Rekombinationsfrequenz von 1,05 %, der Marker K25_M1 hingegen eine vollständige Kopplung mit Co. Die Grenze der Co-Region in Richtung 18 Mbp auf Chromosom 10 liegt somit bei 18,228 Mbp (Ch03d11). Um die Zielregion in Richtung 19 Mbp zu verfeinern, wurden PCRs mit den Markern C6835.384-2 (Pos. 19,457 Mbp), C7629-22009 (Pos. 19,095) und C18470-25831 (Pos. 19,004) von Bai *et al.* (2012) sowie mit dem Marker Mdo.chr10.15 (Pos. 18,962) von Moriya *et al.* (2012) durchgeführt (Abbildung 1.3). Die gelelektrophoretische Auftrennung ausgewählter PCR-Produkte ist in Abbildung 3.24 dargestellt.

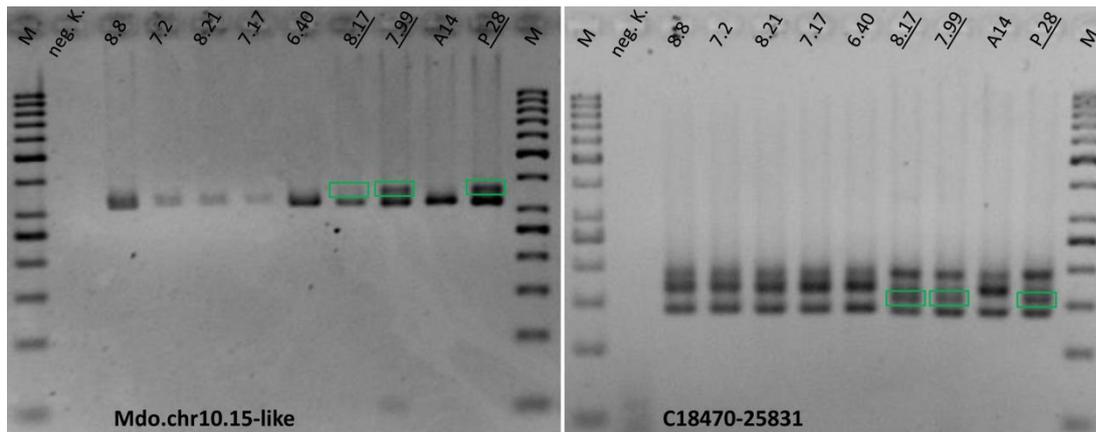


Abbildung 3.24 Markertests zur Feinkartierung der Zielregion

Gezeigt sind Gelbilder der Markertests zur Feinkartierung der Zielregion in Richtung 19 Mbp. Als Vergleich zu den „Ausreißern“ dient genomische P28- und A14-DNA. Die Proben der Apfelbäume mit kolumnarem Phänotyp sind unterstrichen. Das kolumnar-spezifische PCR-Produkt ist grün eingerahmt.

Die Marker C7629-22009 und C6835.384-2 zeigen keine Unterschiede zwischen kolumnarem und nicht kolumnarem Genotyp und wurden somit verworfen (Gelbilder nicht abgebildet). Bei dem Marker Mdo.chr10.15 ist eine eindeutige Zuordnung der PCR-Produkte zu dem kolumnaren Chromosom, selbst nach mehreren Sequenzierungen, nicht möglich. Daher wurden neue Primer erstellt (Mdo.chr10.15-like), die den gewünschten Locus auf Chromosom 10 spezifisch amplifizieren. Auch wenn bei dem Nachkommen 8.17 das kolumnar-spezifische PCR-Produkt nur sehr schwach ist, zeigen alle untersuchten „Ausreißer“ den ihrem Phänotyp entsprechenden Genotyp. Gleiches gilt für den Marker C18470-25831. Sowohl der Marker Mdo.chr10.15-like wie auch der Marker C18470-25831 zeigen somit 100 % Kopplung mit dem *Co*-Gen, sodass keine weitere Feinkartierung in Richtung 19 Mbp erreicht werden konnte.

3.7.3 Markeranalysen anderer Apfelsorten

Neben den Apfelsorten P28 und A14 wurden die neu entwickelten Marker auf fünfzehn kolumnare und sechs nicht kolumnare Apfelsorten getestet. Durch die Verwendung weiterer Apfelsorten sollte untersucht werden, ob die erstellten Marker spezifisch für P28 bzw. A14 oder ob sie sortenunabhängig anwendbar sind. In Tabelle 3.14 sind die Ergebnisse der Markertests zusammengefasst. Eine Darstellung der PCR-Analysen befindet sich im Anhang in Abbildung A4.1.

Tabelle 3.14 Markertests auf diverse Apfelsorten

Apfelsorte	K25_M1	H1_M1	I2_3_M1	1C3_M2
kolumnar:				
Procats 1	ho c	he c	ho c	ho c
Procats 4	he c	he c	he c	he c
Procats 11	he c	he c	he c	he c
Procats 13	he c	he c	he c	he c
Procats 27	he c	he c	he c	he c
Procats 28	he c	he c	he c	he c
Greencats	he c	he c	he c	he c
Pomforyou	he c	he c	he c	he c
Pomfital	he c	he c	he c	he c
Kordonia	ho c	he c	ho c	ho c
A10-28	he c	he c	he c	he c
A68-173	ho c	he c	ho c	ho c
A73-19-93K	ho c	ho c	ho c	ho c
McIntosh Wijcik	ho c	he c	ho c	ho c
Pomgold	he c	he c	he c	he c
HL 4 K	ho c	he c	ho c	ho c
nicht kolumnar:				
McIntosh	ho c	he c	ho c	ho c
Elswout	ho nc	ho nc	ho nc	ho nc
Gala	ho nc	ho nc	ho nc	ho nc
Jonagold	he c	ho nc	ho nc	he c
Pinova	he c	ho nc	ho nc	he c
Topaz	he c	he c	he c	he c
A14-190-93	ho nc	ho nc	ho nc	ho nc

he: heterozygot; ho: homozygot; c: kolumnar; nc: nicht kolumnar

Die Apfelsorten sind in Tabelle 3.14 anhand ihres Phänotyps in kolumnar und nicht kolumnar unterteilt. In den Spalten zwei bis vier sind die Genotypen der Apfelsorten in Bezug auf den jeweilig verwendeten Marker angegeben. Die homozygote Apfelsorte A73-19-93K (*Co/Co*) zeigt für jeden der verwendeten Marker erwartungsgemäß das homozygot kolumnare Bandenmuster. Die Sorten McIntosh und McIntosh Wijcik zeigen für die jeweiligen Marker dasselbe Bandenmuster. Dies verwundert nicht, da es sich bei McIntosh Wijcik um eine somatische Mutante von McIntosh handelt und der einzige Unterschied

zwischen den beiden Sorten höchstwahrscheinlich die *Co*-Mutation ist. Die verwendeten Marker detektieren bei zahlreichen Apfelsorten den erwarteten Genotyp, wie etwa bei Gala und Elswout. Allerdings gibt es auch einige Apfelsorten, bei denen die Marker nach PCR nicht den erwarteten Genotyp zeigen, was gegen eine Sortenunabhängigkeit der Marker spricht. Eine besondere Rolle spielt die Apfelsorte Topaz. Für jeden der vier Marker zeigt sie den heterozygot kolumnaren Genotyp, obwohl der Phänotyp von Topaz eindeutig nicht kolumnar ist. Ausgewählte *Co*-Gen gekoppelte Marker anderer Arbeitsgruppen wurden daraufhin auf genomische DNA des Kultivars Topaz getestet (Abbildung 3.25). Für jeden dieser Marker zeigt die Sorte Topaz den heterozygot kolumnaren Genotyp. Damit ist keiner der verfügbaren *Co*-Gen gekoppelten Marker sortenunabhängig und absolut mit *Co* gekoppelt.

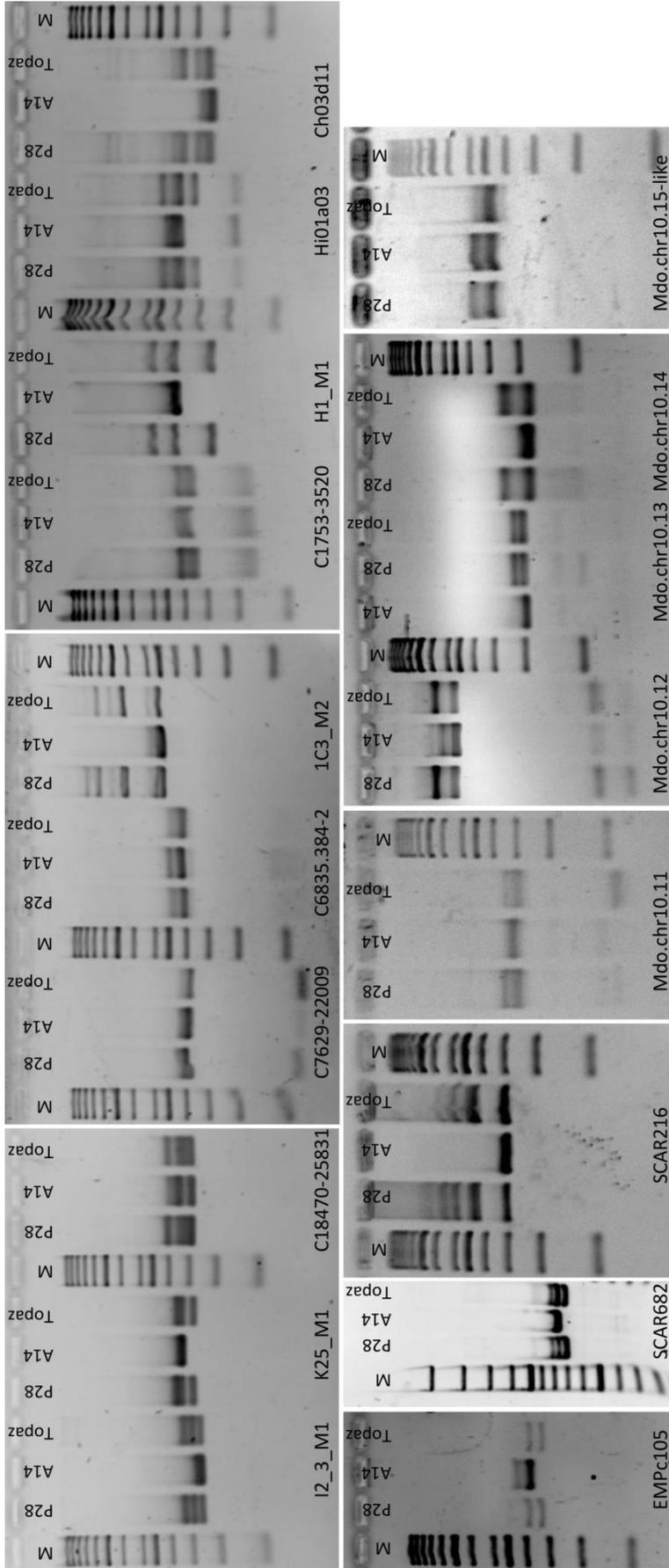


Abbildung 3.25 Genotypisierung der Apfelsorte Topaz mittels verschiedener Marker

Eine Auswahl Co-Gen gekoppelter Marker (Liebhard *et al.* 2002; Tian *et al.* 2005; Fernández-Fernández *et al.* 2006; Moriya *et al.* 2009; Moriya *et al.* 2012; Bai *et al.* 2012) sowie die selbst erstellten Marker wurden auf genomische DNA von Topaz getestet. Abgebildet ist eine Zusammenstellung der Gelbilder der PCRs. Alle Marker zeigen den heterozygoten (Co/-) Genotyp, obwohl Topaz eindeutig einen nicht kolumnaren Phänotyp besitzt. Als Vergleich dienen genomischen DNAs der Sorten P28 und A14.

3.7.4 Transposon-basierende Marker

Die Identifizierung des Transposons *Gypsy-44* in der Region um 18,79 Mbp auf Chromosom 10 als der einzige Unterschied zwischen den Apfelsorten McIntosh und McIntosh Wijcik war die Basis für weitere Markeranalysen (Kapitel 3.5.2). Es wurden drei Primerpaare entworfen: 1. „Linker Rand“, das Primerpaar überspannt das linke Ende von *Gypsy-44* und einen Teil der linken flankierenden DNA; 2. „Rechter Rand“, die beiden Primer überspannen das rechte Ende von *Gypsy-44* sowie einen Teil der rechten flankierenden DNA; 3. „überspannend“, die Primer überspannen das komplette Transposon sowie einen Teil der rechten und linken flankierenden DNA (Abbildung 3.26).

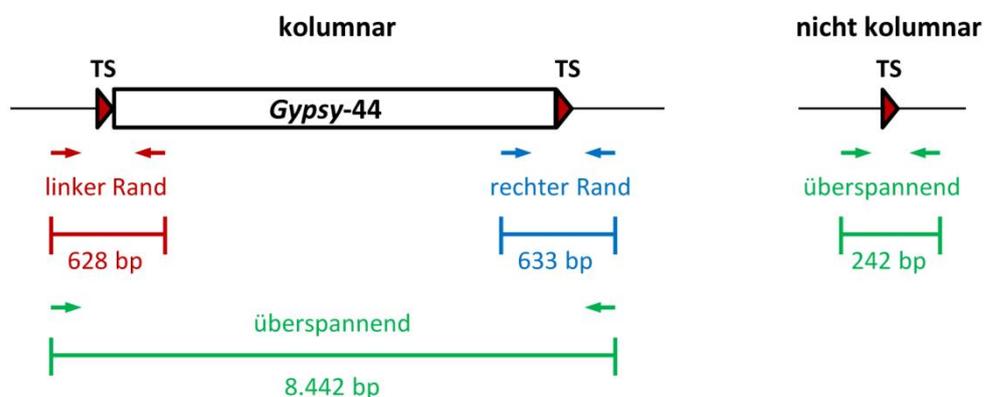


Abbildung 3.26 *Gypsy-44*-basierende Primer

Die Abbildung zeigt schematisch die Lage der Primer für die Transposon-basierenden Marker sowie die erwarteten Größen der PCR-Produkte. Im linken Teil der Abbildung ist das kolumnare Chromosom dargestellt. Es sind die Positionen der drei erstellten Primerpaare eingezeichnet. Das Primerpaar „Linker Rand“ überspannt das linke Ende von *Gypsy-44* und einen Teil der linken flankierenden DNA (rot). Die beiden Primer „Rechter Rand“ überspannen das rechte Ende von *Gypsy-44* sowie einen Teil der rechten flankierenden DNA (blau). Die Primer „überspannend“ überspannen das komplette Transposon sowie einen Teil der rechten und linken flankierenden DNA (grün). Im rechten Teil der Abbildung ist das nicht kolumnare Chromosom mit der „leeren“ „target site“ (TS) dargestellt. *Gypsy-44* ist nicht vorhanden. Mit dem Primerpaar „überspannend“ wird bei einer PCR lediglich die „leere“ „target site“ und ein Teil der rechten und linken flankierenden DNA amplifiziert.

Als Template für die PCRs dienten genomische DNAs der nicht kolumnaren Apfelsorten McIntosh, A14 und Topaz, der heterozygot kolumnaren Sorten P28 und McIntosh Wijcik und der homozygot kolumnaren Sorte A73-19-93K (A73). Die gelelektrophoretische Auftrennung der PCR-Produkte ist in Abbildung 3.27 dargestellt.

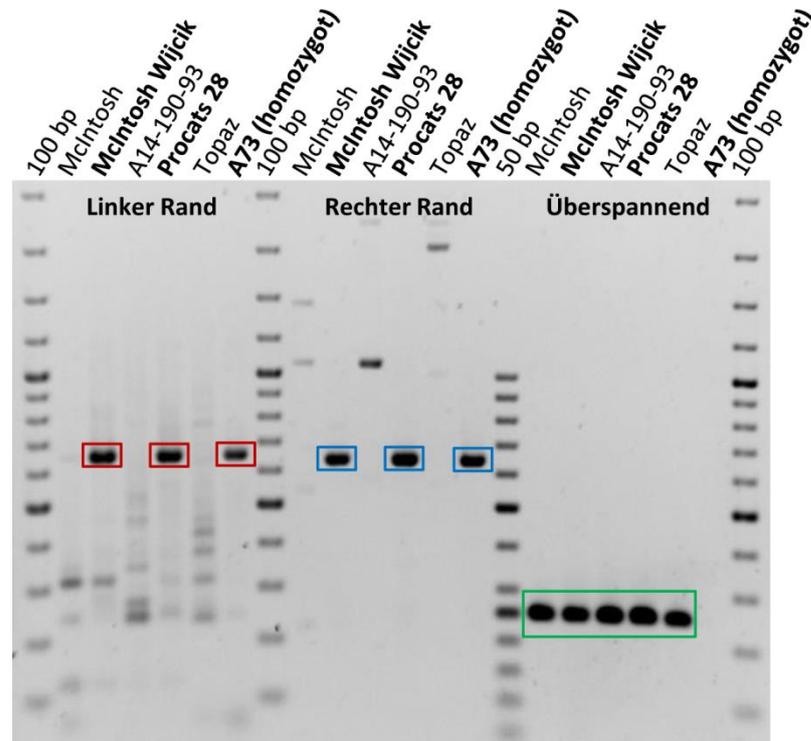


Abbildung 3.27 Gypsy-44-spezifische PCR auf genomische DNAs diverser Apfelsorten

Mit Primern, die den linken (rot) und rechten (blau) Rand sowie das komplette Transposon bzw. die Integrationsstelle (grün) überspannen, wurden PCRs auf genomische DNAs kolumnarer (fett) und nicht kolumnarer Sorten durchgeführt. Die Abbildung zeigt die gelelektrophoretische Auftrennung der PCR-Produkte der Gypsy-44-spezifischen Marker. Als Molekulargewichtsmarker dienten die 50 bp und 100 bp Ladder.

Die heterozygot und homozygot kolumnaren Apfelsorten zeigen ein PCR-Produkt in erwarteter Höhe für den linken und rechten Rand des Transposons. Dies beweist, dass das Transposon an der untersuchten Position vorhanden ist. Bei den nicht kolumnaren Apfelsorten tritt hingegen kein PCR-Produkt in entsprechender Höhe auf. Die Integrationsstelle-überspannende PCR zeigt für die heterozygot kolumnaren und nicht kolumnaren Sorten ein Produkt, welches die Anwesenheit der „leeren“ Integrationsstelle (ohne Transposon) anzeigt. Die homozygot kolumnare Sorte zeigt dabei kein Produkt. Ein mögliches PCR-Produkt eines homozygot kolumnaren Genotyps sowie des kolumnaren Chromosoms der heterozygot kolumnaren Sorten entsteht vermutlich aufgrund seiner enormen Größe (etwa 8,5 kb) unter den gewählten Amplifikationsbedingungen mit genomischer DNA als Template nicht. Verwendet man hingegen den entsprechenden BAC als Template, gelingt die Amplifikation des erwarteten DNA-Fragments. Neben den gezeigten Apfelsorten wurde das Primerpaar, welches den rechten Rand des Transposons überspannt, auf die nicht kolumnaren Apfelsorten Jonagold, Gala, Elswout und Pinova,

sowie auf die kolumnaren Sorten Procats 1, Procats 4, Procats 11, Procats 13, Greencats, Kordonia, Pomforyou, Pomfital, Pomgold, A10-28 und HL 4 K getestet. In allen kolumnaren Apfelsorten tritt die erwartete PCR-Bande auf, wohingegen sie bei den nicht kolumnaren Sorten fehlt (Daten Doktorarbeit Romina Petersen, IMSB). Somit konnte ein molekularer Marker etabliert werden, der sortenunabhängig die Differenzierung zwischen kolumnarem und nicht kolumnarem Genotyp ermöglicht.

4 Diskussion

Kolumnare Apfelbäume können aufgrund ihrer sehr kompakten Wuchsform erheblich dichter gepflanzt werden als normalwüchsige Bäume, was in einer Steigerung des Ertrags um bis zu 300 % pro Hektar resultiert (Jacob 2010). Das wirtschaftliche Interesse wird darüber hinaus dadurch geweckt, dass lediglich minimaler Aufwand für den Rückschnitt der Bäume betrieben werden muss und die Ernte mittels Vollerntemaschinen durchgeführt werden kann (Tobutt 1985). Allerdings weisen kolumnare Apfelbäume auch einige negative Eigenschaften wie schlechte Fruchtqualität und verminderte Lagerfähigkeit der Früchte, sowie eine hohe Anfälligkeit gegen Krankheiten wie beispielsweise Apfelschorf auf (Meulenbroek *et al.* 1998; P. Braun, pers. Kommunikation). Die Erzeugung von Apfelsorten mit kolumnarem Wachstum und hoher Fruchtqualität durch konventionelle Kreuzung gestaltet sich aufgrund des stark heterozygoten Charakters und der Selbstinkompatibilität des Apfels sowie der langen juvenilen Phase als besonders schwierig (Hemmat *et al.* 1994). Aus diesem Grund ist die Identifizierung und Charakterisierung der *Co*-Mutation für die Apfelzüchtung von großer Bedeutung. Allerdings ist die molekulare Ursache des Kolumnarwachstums weitgehend unerforscht. Die Motivation zu dieser Arbeit ist daher die Aufklärung der molekularen Natur der *Co*-Mutation. Das Ziel sollte klassisch über positionelles Klonieren in Kombination mit Next-Generation Sequencing erreicht werden. Als Ausgangsmaterial für die Konstruktion der BAC-Bibliotheken diente die heterozygot kolumnare Apfelsorte P28, da für diese *in-vitro*-Kulturen in ausreichender Menge und jederzeit zur Verfügung standen. Darüber hinaus war genügend Material aus einer Nachkommenschaft einer Kreuzung zwischen P28 und A14 vorhanden, was als Grundlage für die Feinkartierung der *Co*-Region und die Erstellung *Co*-Gen gekoppelter Marker diente. Die Apfelsorten McIntosh Wjick und McIntosh waren darüber hinaus in Form genomischer Illumina Next-Generation Sequencing Reads verfügbar, um identifizierte Unterschiede zwischen dem kolumnaren und dem nicht kolumnaren Chromosom auf Zusammenhänge mit der *Co*-Mutation zu überprüfen. Zur weiteren Unterstützung der Ergebnisse diente eine Auswahl kolumnarer und nicht kolumnarer Apfelsorten.

4.1 Bestimmung der *Co*-Zielregion

Für die molekulare Analyse der *Co*-Mutation ist eine möglichst genaue genomische Lokalisierung von entscheidender Bedeutung. Bereits 1997 lokalisierten Conner *et al.* die *Co*-Region auf Chromosom 10. Im Laufe der Jahre wurde die genomische Lage von *Co* durch

immer neue und enger gekoppelte Marker verfeinert (Liebhard *et al.* 2002; Tian *et al.* 2005; Bai *et al.* 2012; Baldi *et al.* 2012; Moriya *et al.* 2012). Ein wichtiger Schritt bei der Identifizierung der *Co*-Region stellte die Veröffentlichung des Apfelgenoms im Jahr 2010 dar (Velasco *et al.* 2010). Die Berechnung der *Co*-Zielregion auf der Basis dieses Referenzgenoms diente in der vorliegenden Arbeit als erster Anhaltspunkt zur Identifizierung von *Co*. Im weiteren Verlauf konnte diese durch selbst erstellte Marker verifiziert und verfeinert werden. Allerdings war die stark begrenzte Nachkommenschaft von etwa 100 Nachkommen aus einer Kreuzung der Sorten P28 und A14 der limitierende Faktor bei der Feinkartierung der Zielregion. Dennoch konnte eine Grenze der Zielregion bei 18,228 Mbp auf Chromosom 10 mithilfe des Markers Ch03d11 bestimmt werden. Als experimentell gesicherte andere Grenze der Zielregion gilt der Marker Hi01a03 bei 19,911 Mbp. Die selbst erstellten Marker zeigten bei der verwendeten Nachkommenschaft alle eine 100 %-ige Kopplung mit *Co*, ebenso wie die Marker von Moriya *et al.* (2012) und Bai *et al.* (2012). Die Marker von Baldi *et al.* (2012) wurden nicht verwendet, da in dem entsprechenden Bereich eigene Marker zur Verfügung standen. In Abbildung 4.1 sind die Zielregionen von Baldi *et al.* (2012), Bai *et al.* (2012) und Moriya *et al.* (2012) sowie die selbst ermittelte Zielregion dargestellt.

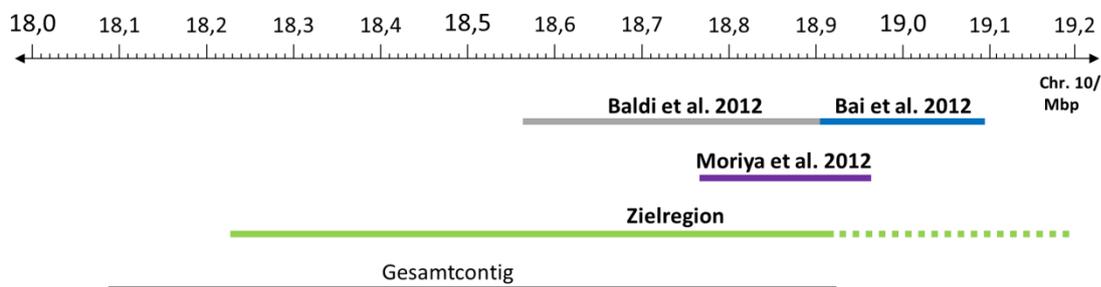


Abbildung 4.1 Co-Zielregion

Die Abbildung zeigt die Zielregionen der Publikationen von Baldi *et al.* (2012), Bai *et al.* (2012) und Moriya *et al.* (2012) sowie die Lage der selbst festgelegten Zielregion. Zudem ist der erstellte Gesamtcontig eingezeichnet.

Ausschlaggebend für die feinere Kartierung der Zielregion anderer Forschergruppen ist die Verfügbarkeit deutlich größerer Nachkommenschaften, wie etwa 1250 Nachkommen bei Baldi *et al.* (2012). Der erstellte Gesamtcontig, der die Grundlage für die Charakterisierung der *Co*-Region bildet, beinhaltet jedoch die komplette Region von Baldi *et al.* (2012) sowie den größten Teil der Zielregion von Moriya *et al.* (2012). Wagt man einen Vergleich zwischen den Grundlagen, auf denen die Zielregionen der genannten Forschergruppen

erstellt wurden, so zeichnen sich Baldi *et al.* (2012) durch die größte analysierte Nachkommenschaft mit 1250 Nachkommen im Vergleich zu 528 (Bai *et al.* 2012) und 1000 Nachkommen (Moriya *et al.* 2012) aus. Baldi *et al.* (2012) wiesen darüber hinaus bei Verwendung des Markers C18470-25381, für den Bai *et al.* (2012) eine absolute Kopplung mit *Co* angeben, in der verwendeten Nachkommenschaft drei Rekombinante nach. Neben der Größe der Nachkommenschaft hat die Verwendung unterschiedlicher Apfelsorten großen Einfluss auf die exakte Lokalisierung der *Co*-Region. Moriya *et al.* (2009) wiesen, wie bereits Tian *et al.* (2005), abweichende Rekombinationsfrequenzen ausgewählter Marker in unterschiedlichen Populationen nach. In einer Population der Kreuzung Fuji und 8H-9-45 cosegregieren Ch03d11 und SCAR₆₈₂ mit *Co*. In einer Population der Kreuzung Fuji und 5-12786 hingegen beträgt der genetische Abstand zu *Co* für Ch03d11 und für SCAR₆₈₂ 1,4 cM. Baldi *et al.* (2012) nutzten zwei Kreuzungen zwischen GD und McIntosh Wijcik, wohingegen Bai *et al.* (2012) vier Kreuzungen mit unterschiedlichen Kreuzungspartnern und Moriya *et al.* (2012) sogar 31 verschiedene Nachkommenschaften verwendeten.

Eine Übertragung der Zielregionen auf andere Apfelsorten ist somit immer mit Vorsicht zu betrachten. Allerdings lässt sich zusammenfassend festhalten, dass der Bereich der selbst eingegrenzten Zielregion, der sowohl mit Baldi *et al.* (2012) und Moriya *et al.* (2012) übereinstimmt (670 – 830 kb im Gesamtcontig), als gesichert gelten und für weitere Analysen als Grundlage dienen kann.

4.2 Genomische Apfel-BAC-Bibliotheken

Die Konstruktion genomischer DNA-BAC-Bibliotheken mit großen Integratmolekülen und einer hohen Anzahl rekombinanter Plasmide ist Voraussetzung für die positionelle Klonierung. Im Verlauf dieser Arbeit konnte die Herstellung der BAC-Bibliotheken durch einige Neuerungen erheblich verbessert werden. Zu Beginn der Arbeit wurde der Gewebeaufschluss der *in-vitro*-Kulturen mittels Homogenisator durchgeführt, wodurch die verwendete Menge an Ausgangsmaterial mit 1,5 g pro Durchgang stark begrenzt war und mehrere Durchläufe nötig waren, um ausreichend genomische DNA zu gewinnen. Durch die Verwendung eines Stabmixers für den Gewebeaufschluss konnte das pro Durchgang verwendete Ausgangsmaterial auf 30 g gesteigert werden. Dies führte zu einer enormen Zeitersparnis ohne Qualitätsverlust (Abbildung 3.2). Durch das Ansetzen mehrerer Restriktionsverdau unterschiedlicher Restriktionsdauer und das anschließende Poolen dieser Ansätze, wurde das Problem minimiert, dass hochmolekulare DNA in Lösung

inhomogen verteilt ist. Anfangs gelang aufgrund des sehr langen und aufwendigen Herstellungsprozesses nur etwa jede fünfte Bibliothek. Durch die Verwendung sogenannter „Mega“-Gele zur Größenfraktionierung, die eine etwa 25-fache Ladekapazität im Vergleich zu dem bisher verwendeten Gelsystem besitzen, konnte neben einer enormen Zeitersparnis auch die Effektivität der Konstruktion von BAC-Banken gesteigert werden. So gab es zum Ende dieser Arbeit keine Ausfälle bei der Konstruktion der Banken mehr. Zudem wurde durch ein Endfilling mittels radioaktiv markierter Nukleotide ein Test etabliert, mit dem die Konstruktion einer BAC-Bank bereits während des Herstellungsprozesses überprüft werden konnte. Die Qualität der Banken zu vergleichen, gestaltete sich jedoch schwierig, da für die Abschätzung der Integratgröße nur 12 Klone pro BAC-Bank aufgearbeitet wurden. Bei der BAC-Bank P28ivML1 aus der 32 der 37 BAC-Klone aus der Zielregion stammen, zeigte sich beispielsweise, dass die aufgrund der Restriktionsverdau der Testklone angenommene durchschnittliche Integratgröße der Klone mit etwa 27 kb zu niedrig angesetzt war, da durch die Sequenzierung eine durchschnittliche Integratgröße von 40 kb bestimmt wurde (chimäre BAC-Klone ausgenommen). Dies könnte auch erklären, dass nach dem Screening der Bank P28ivL4 mit Sondenpool 1 nur drei positive Klone identifiziert wurden, bei der Bank P28ivML1 hingegen 25, obwohl die Gesamtintegratgrößen der beiden Banken lediglich um Faktor drei variieren.

Neben der Integratgröße ist die Anzahl rekombinanter BAC-Klone ausschlaggebend für die Qualität einer BAC-Bibliothek. Durch die angesprochenen Neuerungen konnte die Anzahl der Klone von 26.000 auf 79.000 gesteigert werden (Tabelle 3.3). Zusammengefasst enthielten die vier erstellten genomischen Bibliotheken rund 150.000 rekombinante BAC-Klone mit etwa 26,5 kb durchschnittlicher Integratgröße, was einer statistischen Genomabdeckung von 5,4-fach entspricht. Unter Verwendung der Carbon-Clarke-Formel liegt die Wahrscheinlichkeit, in den erstellten Bibliotheken einen gesuchten einmaligen DNA-Abschnitt zu identifizieren, bei 99 %. Der Vergleich der selbst erstellten Banken mit anderen Apfel-BAC-Bibliotheken beispielsweise von Vinatzer *et al.* (1998) mit 36.864 Klonen und einer durchschnittlichen Integratgröße von 120 kb oder Xu *et al.* (2001) mit 31.584 Klonen und einer durchschnittlichen Größe von 125 kb zeigt, dass die Anzahl rekombinanter BAC-Klone in derselben Größenordnung liegt, die Integratgrößen jedoch in den selbst erstellten BAC-Bibliotheken erheblich kleiner sind. Die angesprochenen Arbeiten verwendeten die Methode nach Zhang *et al.* (1995), bei der die isolierten Nuklei in Agarose gegossen werden, um die DNA vor Scherkräften zu schützen. In diesen Agarose-„Plugs“

wird auch der anschließende Restriktionsverdau durchgeführt, ehe die Größenfraktionierungen mithilfe des „Pulsed-field gel electrophoresis“-Systems (Schwartz und Cantor 1984) folgen. Die Verwendung von Agarose-„Plugs“ und „Pulsed-field gel electrophoresis“-System zur Größenfraktionierung wurden im Rahmen einer Diplomarbeit (Otto 2008) ausführlich getestet. Es konnten jedoch keine zufriedenstellenden Resultate erzielt werden, sodass in der vorliegenden Arbeit von der Verwendung dieser Methode abgesehen wurde.

4.3 Erstellung einer Referenzsequenz der Co-Zielregion

Die erstellten Apfel-BAC-Bibliotheken wurden mittels Koloniefilter-Hybridisierung analysiert, um rekombinante Plasmide aus der Zielregion zu identifizieren. Für das Screening wurden PCR-basierende Sonden erstellt (Anhang A2). Das Screenen der BAC-Bibliotheken führte zur Detektion von 61 Klonen mit den Sondenpools 1, 2 und 3. Von diesen 61 Klonen stammen 36 Klone von Chromosom 10, 13 Klone von Chromosom 5 und 12 von anderen Chromosomen. Der Grund, warum so viele Klone von Chromosom 5 stammen, liegt vermutlich darin, dass Chromosom 5 und Chromosom 10 eine große Anzahl paraloge Sequenzen enthalten, die durch eine genomweite Duplikation während der Evolution des Apfelgenoms entstanden sind (Abbildung 4.2 A). Im „Circle Plot“, der die Colinearität genomischer Bereiche des Chromosoms 10 mit den anderen Apfelchromosomen darstellt, lässt sich die enge Verwandtschaft der Chromosomen 5 und 10 deutlich erkennen (Abbildung 4.2 B).

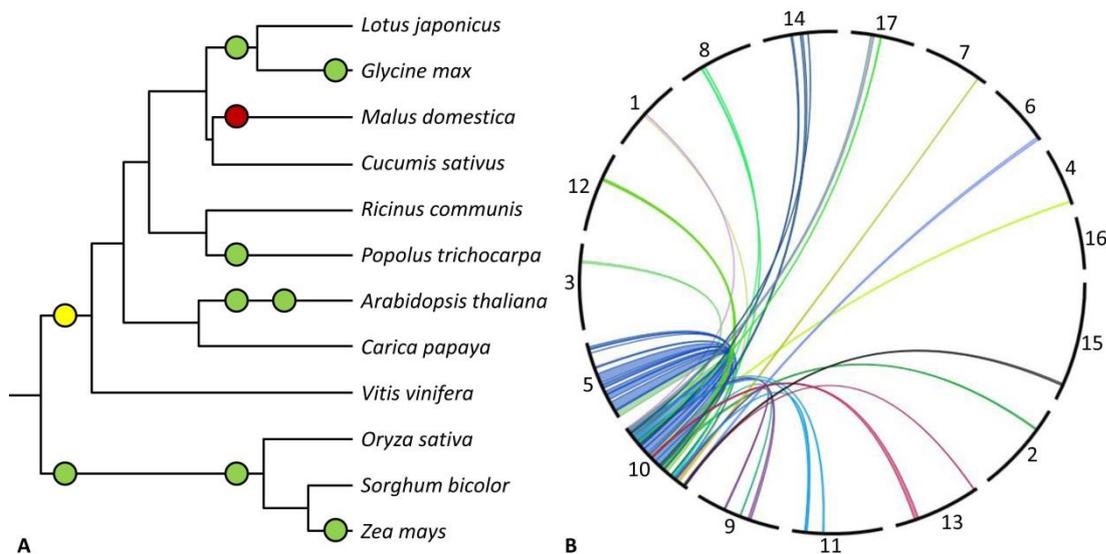


Abbildung 4.2 Evolution des Apfelgenoms

A) zeigt einen phylogenetischen Stammbaum ausgewählter Landpflanzen mit als Kreise eingezeichneten genomweiten Duplikationen. Die genomweite Duplikation, aus der sich die enge Verwandtschaft zwischen Chromosom 5 und 10 begründet, ist in rot dargestellt, die genomweite Duplikation in der Evolution der Eudikotyledonen in gelb. Der Stammbaum wurde in Anlehnung an Proost *et al.* (2011) erstellt. B) stellt den Sequenzvergleich von Chromosom 10 des Apfels mit allen anderen Apfelchromosomen dar. Auffällig ist die große Übereinstimmung zwischen Chromosom 5 und Chromosom 10 infolge einer genomweiten Duplikation. Der „Circle Plot“ wurde mit der Plattform PLAZA erstellt (Proost *et al.* 2009).

Die 12 nur scheinbar positiven BAC-Klone von anderen Chromosomen resultieren vermutlich aus den moderaten Hybridisierungsbedingungen (60 °C in 3 x SSC) der Koloniefilter-Hybridisierung. Eine weitere Ursache sind möglicherweise homologe Bereiche zwischen Chromosom 10 und anderen Apfelchromosomen durch eine ursprünglichere genomweite Duplikation, die in der Evolution der Eudikotyledonen stattgefunden hat, aber bereits sehr lange zurückliegt (Abbildung 4.2 A). Ein stringenteres Verfahren für das Screening von BAC-Bibliotheken ist das PCR-basierte Screening (Xu *et al.* 2001). Allerdings ist dieses Verfahren sehr kostenintensiv, da das rekombinante Plasmid jeder Kolonie vor dem Screening isoliert werden muss. Der Sondenpool 4 führte zu 51 vermeintlich positiven Klonen, wobei letztlich nur ein Klon auf Chromosom 10 in der Zielregion positioniert werden konnte. Ausschlaggebend dafür ist der leicht repetitive Charakter der verwendeten Sonden, der mittels „Pirota“ Blot nachgewiesen wurde (Kapitel 3.3). Die repetitive Struktur der Sonden liegt in ihrer genomischen Herkunft begründet. Die Region, in der die Sonden aus Sondenpool 4 liegen (615 bis 665 kb im Gesamtcontig), enthält überwiegend transposable Elemente und nur ein Gen, welches in einem Retrotransposon lokalisiert ist

(Abbildung 3.20). Dies veranschaulicht die Schwierigkeit, in dieser Region eine „single-copy“-Sonde zu erstellen.

Die Randsequenzierung der Klone stellte ein adäquates Mittel zur endgültigen Lokalisierung der BAC-Klone dar (Tabelle 3.5). Die vollständige Sequenzierung erfolgte bei einigen Klonen mittels Shotgun-Sequenzierung (K25, H2, H4 und K4), bei der Mehrzahl jedoch mittels Illumina Next-Generation Sequencing. Während die Sequenzierung mittels Shotgun- bzw. Sangersequenzierung Sequenzen bis zu 1.000 bp Länge liefert, werden beim Illumina Next-Generation Sequencing Reads mit einer Länge von lediglich 101 bp generiert. Nach Aufbereitung der Rohdaten liegt die durchschnittliche Readlänge bei etwa 86 bp. Die erheblich kürzere Readlänge wird durch eine wesentlich höhere Abdeckung ausgeglichen. Allerdings bleiben repetitive Sequenzen der limitierende Faktor bei einer Assemblierung. Ist ein repetitives Element länger als die verfügbare Sequenzlänge, führt dies zu Lücken in der Assemblierung (Schatz *et al.* 2010). Die Unterschiede zwischen den verschiedenen Sequenziermethoden hinsichtlich Länge und Zahl der erhaltenen Sequenzen erfordern die Verwendung unterschiedlicher Assemblierungsstrategien. Während das Programm SeqMan™ auf der „Overlap-Layout-Consensus“-Strategie (OLC) basiert, bedient sich CLC des „de Bruijn“-Graphen (DBG) (Kumar und Blaxter 2010). Die OLC-Strategie beruht stark vereinfacht dargestellt auf dem paarweisen Vergleich der Sequenzen zur Identifizierung von Überlappungen und letztlich zur Erstellung einer Consensus-Sequenz (Myers 1995). Sie eignet sich besonders für die Assemblierung von Sequenzen die mit dem Sangerverfahren erstellt wurden. Wichtige Parameter bei dieser Strategie sind die Festlegung der minimalen Länge einer Überlappung sowie deren minimale Sequenzidentität (Miller *et al.* 2010). Bei Verwendung eines DBGs werden die Sequenzen hingegen in kurze Fragmente definierter Länge, sogenannte k-mers, zerlegt (de Bruijn und Erdos 1946). Dies ermöglicht die Bearbeitung großer Datensätze, wie sie bei Illumina-Sequenzierungen der Fall sind, da nicht jede Sequenz gespeichert werden muss, sondern lediglich die unterschiedlichen k-mers und deren Anzahl. Bei der Assemblierung werden k-mers miteinander verknüpft, die eine Überlappung der Länge k-1 besitzen, und so zu Consensus-Sequenzen zusammengefügt. Diese Methode ist sehr sensitiv gegenüber repetitiven Strukturen, die länger als die verwendete k-mer Länge sind, und Sequenzierfehlern (Miller *et al.* 2010). Durch eine Verfeinerung der Algorithmen werden diese Probleme jedoch minimiert (Pevzner *et al.* 2004). Darüber hinaus wird zur Verbesserung der Assemblierung in repetitiven Bereichen verstärkt die Information der „paired end“-Sequenzierung für die Assemblierung genutzt.

Die Verwendung der „paired end“-Sequenzierung ermöglicht es, dass man zwei zueinander gehörige Reads eines DNA-Fragmentes mit bekannter Distanz erhält, was zum Überspannen von repetitiven Bereichen führen kann, insbesondere bei der Verwendung von „mate pair“-Bibliotheken (Chaisson *et al.* 2009).

Die Assemblierung der BAC-Klone, die mittels shotgun-Sequenzierung sequenziert wurden, erfolgte problemlos mithilfe des Programms SeqMan™. Allerdings ist diese Sequenziermethode im Vergleich zum Illumina Next-Generation Sequencing sehr kosten- und zeitintensiv (Schatz *et al.* 2010), weswegen die Mehrzahl der Klone Illumina sequenziert wurden. Wie aber bereits geschildert stellt die Assemblierung der Illumina-Reads eine Herausforderung dar. Letztlich war die Kombination der beiden Assemblierungsstrategien in Form von OLC und DBG die Methode der Wahl. Dazu wurden zunächst *de novo*-Assemblierungen der Illumina-Reads mittels DBG mit der CLC Genomics Workbench und der Assembly Cell durchgeführt. Die Verwendung variierender k-mer-Längen ist ein wichtiger Schritt bei der Assemblierung von Next-Generation Sequencing-Datensätzen (Surget-Groba und Montoya-Burgos 2010). Die Wahl der optimalen k-mer Länge hängt von mehreren Faktoren ab, etwa der Gesamtabdeckung, der Fehlerrate der Reads und der Komplexität des zu assemblierenden DNA-Moleküls (Simpson *et al.* 2009). Da die optimale k-mer Länge schwer vorherzusagen ist und von BAC-Klon zu BAC-Klon variiert, wurden für jeden BAC-Klon Assemblierungen mit verschiedenen k-mer Längen durchgeführt. Die entstehenden Contigs wurden anschließend mittels OLC in SeqMan™ erneut assembliert. Von entscheidender Bedeutung für die Auswahl des korrekten Contigs war das Vorhandensein der Randsequenzen, der Vektorsequenz und des Ringschlusses. Die BAC-Sequenzen wurden darüber hinaus durch ein Mapping der Illumina-Reads überprüft. Homopolymer-Runs und Bereiche mit geringem GC-Gehalt zeigen eine geringe Abdeckung in den Mappings (Minoche *et al.* 2011). Diese Bereiche wurden daher mittels Sanger-Sequenzierung überprüft. Möglicherweise kommen diese Schwankungen in der Abdeckung durch den Amplifikationsschritt in der Herstellung der Illumina-Bibliotheken zustande (Aird *et al.* 2011).

Mit der Erstellung von Restriktionskartierungen konnten die BAC-Sequenzen abschließend verifiziert werden (Abbildung 3.5). Eine Besonderheit bei den Restriktionsverdau war das vereinzelte Auftreten nicht geschnittener *Xba*I-Schnittstellen, trotz intakter Erkennungssequenz. Die Ursache dafür liegt in der Verwendung eines *dam*⁺ *E.coli*-Stammes (DH10B) für die Konstruktion einer BAC-Bibliothek in Kombination mit der Verwendung

eines dam-sensitiven Restriktionsenzym (*XbaI*) für die Restriktionskartierungen. Die Dam-Methylase methyliert in *E.coli* die N6-Position des Adenins in der Erkennungssequenz „GATC“ (Hattman *et al.* 1978). Die Restriktionsendonuklease *XbaI* besitzt die Erkennungssequenz „TCTAGA“. Statistisch betrachtet folgt in 1/16 der Fälle der Erkennungssequenz „TCTAGA“ die Basenfolge „TC“, was zur Bildung einer Dam-Methylierungsstelle führt („TCTAGATC“). Die daraus resultierende Methylierung der Restriktionsschnittstelle verhindert die Spaltung der DNA durch das Enzym *XbaI* (Luo *et al.* 2003). An allen untersuchten Positionen, an denen eine *XbaI*-Schnittstelle nicht geschnitten wurde, liegt eine Dam-Methylierungsstelle vor. Für Restriktionsverdau mit *Clal* konnte selbiges Phänomen festgestellt werden. Das *dam*-sensitive Enzym besitzt die Erkennungssequenz „ATCGAT“, wodurch eine Methylierung sogar in 1/4 der Fälle auftritt. Die zweite Besonderheit bei den Restriktionsverdau war die deutliche gelelektrophoretische Trennung zweier DNA-Moleküle, deren Molekülgröße nur um 6 bp variierte (Abbildung 3.5). PCR-Analysen und anschließende Sequenzierungen zeigten, dass die Sequenzen korrekt assembliert wurden, die 1320 bp-Bande jedoch eine deutlich langsamere Wanderungsgeschwindigkeit aufwies als die Größe vermuten lassen würde. Eine mögliche Erklärung für dieses Phänomen ist eine Krümmung der DNA, die Einfluss auf die gelelektrophoretische Auftrennung hat (Hagerman 1984). Die Krümmung wird zum Beispiel durch Homopolymere von Adenin bzw. Thymin, die in bestimmten Wiederholungen auftreten, verursacht (Koo *et al.* 1986). Trotz der weit verbreiteten Meinung, dass dieses Phänomen nur bei Polyacrylamid-Gelen auftritt (u.a. Marini *et al.* 1982), konnten Calladine *et al.* (1991) zeigen, dass dies auch in Agarose-Gelen zu beobachten ist und die entsprechenden DNA-Moleküle nur etwa 92 % ihrer Wanderungsgeschwindigkeit aufweisen.

Ein Problem bei Klonierungsexperimenten besteht in dem Auftreten chimärer BAC-Klone, die durch Ligation zweier DNA-Moleküle unterschiedlicher chromosomaler Herkunft in ein Vektor-Molekül entstehen, da somit eine Verknüpfung von Sequenzen stattfindet, die auf genomischer Ebene nicht existiert. Nach erfolgreicher Assemblierung der Sequenzen der BAC-Klone und Verifizierung der Sequenzen sollte daher bei chimären BAC-Klonen der Anteil extrahiert werden, der aus der Zielregion stammt. In der BAC-Bibliothek P28ivML1 konnten von den 32 Klonen aus der Zielregion fünf chimäre Klone identifiziert werden (Tabelle 3.5). Dies entspricht einem prozentualen Anteil von 13,5 % der BAC-Klone und liegt damit über dem in der Regel beobachteten Anteil chimärer Klone von 4 bis 11 % (Volik *et al.*

2003). Mögliche Ursache für deren Entstehung ist eine hohe DNA-Konzentration bei der Ligation (Osoegawa *et al.* 1998). Die detektierten Klone enthalten zwei etwa gleich große DNA-Moleküle unterschiedlichen chromosomalen Ursprungs (Tabelle 3.9). Meist beinhalten chimäre Klone jedoch ein sehr großes und ein sehr kleines DNA-Fragment (Osoegawa *et al.* 2000). Diese Beobachtung bezieht sich jedoch auf BAC-Bibliotheken mit durchschnittlichen Integratgrößen von 130 bis 200 kb. Chimäre BACs aus etwa gleich großen DNA-Molekülen würden sich aufgrund der enormen Größe mit erheblich schlechterer Effektivität transformieren lassen, womit ihr Auftreten sehr selten ist (Sheng *et al.* 1995). Im Gegensatz dazu besitzen die hier erstellten BAC-Bibliotheken im Durchschnitt eine Integratgröße von lediglich 30 kb, sodass selbst ein chimärer BAC mit ungefähr 60 kb Integrat kaum schlechter transformierbar wäre. Sheng *et al.* (1995) verzeichneten eine erhebliche Abnahme der Transformationseffizienz erst bei Plasmiden über 80 kb. Dies könnte somit eine Erklärung für den höheren Anteil an chimären BAC-Klonen in der untersuchten BAC-Bibliothek sein. Allerdings standen mit der BLAST-Suche gegen das GD-Genom und der Verfügbarkeit der „mate pair“-Datensätze für ein Mapping (Abbildung 3.7) zwei Methoden zur Verfügung, um chimäre BAC-Klone schnell zu identifizieren. Aufwendiger war die Untersuchung des Klons 214, bei dem die BLAST-Suche und das Mapping der „mate-pair“-Datensätze widersprüchliche Ergebnisse lieferten. Allerdings konnte durch „long range“ PCR, Southern Blot-Analysen und inverse PCR zweifelsfrei nachgewiesen werden, dass es sich bei 214 nicht um einen chimären BAC-Klon handelt, sondern der BLAST-Treffer in einem GD-Contig liegt, der fälschlicherweise auf Chromosom 4 anstelle von Chromosom 10 positioniert wurde. Insgesamt wurden drei Contigs von Chromosom 4 und ein Contig von Chromosom 9 letztlich auf Chromosom 10 neu kartiert, sowie 21 Contigs identifiziert, die keine Übereinstimmung zu Sequenzen aus der untersuchten Region aufweisen und somit vermutlich von anderen Chromosomen stammen. Ursache dafür könnten mögliche Unterschiede zwischen der untersuchten Sorte P28 und GD sein. Wahrscheinlicher ist jedoch, dass es sich um Fehler bei der Assemblierung und Annotation des Apfelgenoms handelt, was bei einem sogenannten „draft genome“, also eine Art Entwurf des Genoms, nicht ausgeschlossen ist.

Aufgrund der erstellten Co-Gen gekoppelten Marker konnten alle BAC-Sequenzen dem kolumnaren und nicht kolumnaren Chromosom zugeordnet und Chromosomen-spezifisch zu Metacontigs zusammengefügt werden (Abbildung 3.10). Die Metacontigs und „long range“ PCR-Produkte wurden zu einem Gesamtcontig von 848.024 bp Länge

zusammengefasst. Die Metacontigs und der Gesamtcontig wurden anhand genomischer Illumina „mate pair“-Datensätzen der Apfelsorte P28 auf Richtigkeit überprüft. Die Mappings der „mate pair“-Reads gegen den Gesamtcontig zeigten im Besonderen in den Datensätzen der großen Fraktion (7 bis 8 kb) mit 50 bis 72 % einen großen Anteil an Duplikaten (Tabelle 3.8). Grund für den hohen Anteil ist vermutlich der Amplifikationsschritt in Form einer Anreicherungs-PCR bei der Herstellung der „mate pair“-Bibliothek. Die hohe Zahl an Duplikaten scheint bei „mate pair“-Bibliotheken nicht außergewöhnlich zu sein. Bei der Sequenzierung des Pandagenoms liegt der durchschnittliche Anteil an Duplikaten für Banken mit 5 kb Fragmentgröße bei 39 %, bei Banken mit 10 kb Fragmentgröße sogar bei 77 % (Li *et al.* 2010). Durch die Funktion „Remove Duplicate Reads“ der CLC Genomics Workbench konnten die Duplikate jedoch problemlos entfernt werden, sodass die Verfügbarkeit der „mate pair“-Reads für Mappings eine einfache Möglichkeit zur Überprüfung des Gesamtcontigs darstellte. Lediglich eine Position, zwischen 240 und 250 kb im Gesamtcontig, zeigt eine Lücke (Abbildung 3.11). Es handelt sich dabei um den Übergang eines Metacontigs des kolumnaren Chromosoms (3c) zu einem „long range“ PCR-Produkt (LR_Mc_3c_3nc), welches aufgrund der hohen Übereinstimmung zu dem dort annotierten GD Contig vermutlich das nicht kolumnare Chromosom widerspiegelt. Größere Unterschiede zwischen den beiden Chromosomen könnten ausschlaggebend dafür sein, dass ein Mapping mit maximaler Stringenz zu keiner Überspannung dieser Position führt. Von einer umfangreichen Analyse dieser Region wurde abgesehen, da die betreffende Stelle mit einer Entfernung von fast 500 kb zur Integrationsstelle von *Gypsy-44* höchstwahrscheinlich in keinem Zusammenhang mit der *Co*-Mutation steht. Obwohl die genomische Situation an dieser einen Position nicht abschließend geklärt werden konnte, wurde der Gesamtcontig als zusammenhängende Consensus-Sequenz für weitere Analysen verwendet.

4.4 Analyse der Co-Zielregion

Die Referenz der *Co*-Zielregion hat eine Länge von 848.024 bp. Den größten Anteil stellen transposable Elemente mit rund 40 %, was in guter Korrelation mit ihrem Anteil von rund 42,4 % des Gesamtgenoms des Apfels steht (Velasco *et al.* 2010).

Betrachtet man die transposablen Elemente in Bezug auf die Klassifizierung nach Wicker *et al.* (2007), lassen sie sich zunächst grob in die Klassen Retrotransposons und DNA-Transposons unterteilen. Retrotransposons machen mit 89,5 % den Hauptanteil

transposabler Elemente des Gesamtcontigs aus, was ebenfalls in guter Übereinstimmung mit ihrem Anteil von 88,7 % an transposablen Elementen im Apfelgenom steht (Velasco *et al.* 2010). Aufgrund des hohen genomischen Anteils verwundert es nicht, dass die Größe eines Genoms mit der Gesamtmasse an Retrotransposons und deren Anzahl an verschiedenen Familien korreliert (Kumar und Bennetzen 1999). Das 12 Mbp große Genom von *Saccharomyces cerevisiae* besitzt nur fünf verschiedene Familien von LTR-Retrotransposons, die etwa 3 % des Genoms ausmachen (Kim *et al.* 1998). Das Genom von *Arabidopsis thaliana* (125 Mbp) besteht zu 10 % aus Retrotransposons mit etwa 100 Familien (TAGI 2000). *Zea mays* mit einer Genomgröße von 2,3 Gbp setzt sich zu rund 70 % aus Retrotransposons zusammen, die mehreren 100 Familien angehören (SanMiguel und Bennetzen 1998). Retrotransposons lassen sich in die Unterklassen LTR- und Non-LTR-Retrotransposons unterteilen. Die Non-LTR-Retrotransposons machen nur etwa 15,3 % der Retrotransposons im Gesamtcontig aus, wobei davon 14,5 % auf LINEs und nur 0,82 % auf SINEs entfallen. Den Hauptanteil der Retrotransposons bilden die LTR-Retrotransposons mit 74,2 % in Form der Superfamilien *Gypsy* (43,3 %) und *Copia* (23,8 %). Während der prozentuale Anteil an LINEs sehr gut mit dem Anteil am Gesamtgenom von 15,3 % übereinstimmt, ist die Superfamilie *Gypsy* im Gesamtcontig etwas unter- (59,5 % in GD) *Copia* hingegen überrepräsentiert (12,9 % in GD). Für SINEs steht kein Vergleichswert zur Verfügung. Die Unterschiede der prozentualen Anteile von *Gypsy* und *Copia* hängen vermutlich damit zusammen, dass Retrotransposons nicht gleichmäßig über das komplette Genom verteilt sind, sondern oftmals geclustert auftreten (Kumar und Bennetzen 1999).

Neben transposablen Elementen konnten im Gesamtcontig insgesamt 83 Gene annotiert werden. Von diesen 83 Genen beruhen 53 auf Annotationen im GD-Genom, wobei 15 durch Transkriptomdaten aus dem Sprossapikalmeristem verifiziert wurden. Bei 27 Genen wurde die Annotation geändert und weitere 3 Gene wurden neu annotiert. Eine repräsentative Anreicherung von Genen ähnlicher Funktion in der *Co*-Zielregion wurde mittels Blast2GO (Conesa und Götze 2008) überprüft, konnte aber nicht festgestellt werden. Allerdings befinden sich durchaus Gene oder Gengruppen in dieser Region, die potenzielle Kandidaten für *Co* sind, wie beispielsweise Transkriptionsfaktoren, Methyltransferasen/Demethylasen, Spleißfaktoren, Gene, die mit Phytohormonen in Verbindung stehen, sowie Kinasen/Phosphatasen. Die Gene wurden auf Unterschiede zwischen dem kolumnaren und nicht kolumnaren Chromosom untersucht. Die Analysen stützen sich auf genomische und transkriptomische Illumina-Datensätze in Kombination mit PCRs der Apfelsorten McIntosh

und McIntosh Wijcik, sowie P28 und A14 (Doktorarbeit Romina Petersen, IMSB). In den potenziellen Kandidatengenomen für *Co* in dieser Region konnten jedoch keine kolumnar-spezifischen Unterschiede festgestellt werden, wodurch eine direkte Wirkung der überprüften Gene als *Co*-Gen ausgeschlossen werden konnte. Die Analyse beschränkte sich auf die kodierenden Bereiche der Gene, sodass mögliche regulatorische Sequenzen nicht überprüft wurden.

Zur Identifizierung der *Co*-Mutation waren neben Unterschieden in den Genen strukturelle Variationen zwischen dem kolumnaren und nicht kolumnaren Chromosom von besonderem Interesse. Dies wurde durch Sequenzvergleiche zwischen den chromosomen-spezifischen Metacontigs der heterozygot kolumnaren Apfelsorte P28 und durch Sequenzvergleiche zwischen dem Gesamtcontig von P28 und dem GD-Genom umgesetzt und auf diese Weise 21 größere Indels (> 100 bp) identifiziert (Abbildung 3.12 bis 3.14). Um zu überprüfen, ob diese Unterschiede tatsächlich kolumnar-spezifisch sind und nicht der stark heterozygote Charakter des Apfelgenoms oder der Vergleich unterschiedlicher Apfelsorten (P28 gegen GD) die Ursachen dieser Unterschiede sind, wurden genomische Illumina-Reads der Apfelsorten McIntosh und McIntosh Wijcik für Mappings verwendet (Tabelle 3.11). Damit konnte eine Assoziation von 20 Indels mit der *Co*-Mutation ausgeschlossen werden. Lediglich die 8,2 kb große Insertion (Indel 19), die eine Kopie des *Gypsy-44* Retrotransposons darstellt, unterscheidet den kolumnaren vom nicht kolumnaren Haplotyp (Position 710.0063 bis 718.262 im Gesamtcontig).

4.5 Identifizierung der *Co*-Mutation

Die Integration des *Gypsy-44* Retrotransposons konnte als einziger Unterschied zwischen dem kolumnaren und dem nicht kolumnaren Chromosom in P28 festgestellt und anhand der genomischen Illumina-Daten von McIntosh Wijcik bestätigt werden. Um die Verbindung zwischen der Insertion von *Gypsy-44* und dem kolumnaren Phänotyp zu überprüfen, wurden genomische PCRs, die dessen Integrationsstellen an Position 18,79 auf Chromosom 10 überspannen, durchgeführt (Abbildung 3.27). Jede bisher auf diese Weise getestete DNA aus einer kolumnaren Apfelsorte erzeugt die PCR-Produkte, die für den rechten und linken Rand des Retrotransposons charakteristisch sind. Damit ist die Anwesenheit von *Gypsy-44* an dieser Position 100 % korreliert mit dem kolumnaren Phänotyp. Bei jeder normalwüchsigen Apfelsorte fehlen hingegen diese PCR-Produkte. Nur das Amplifikat der „leeren“ „target site“ in der Transposon-überspannenden PCR wird erzeugt. Ein besonders

aufschlussreicher Fall ist die normalwüchsige Sorte Topaz. Hier zeigen alle nicht auf dem Retrotransposon basierenden Markertests den Haplotyp des kolumnaren Genotyps (Abbildung 3.25). Die Transposon-basierenden Marker, die die Integrationsstellen von *Gypsy-44* überspannen, sind hingegen die ersten und einzigen Marker, die für die Apfelsorte Topaz nicht den kolumnaren, sondern den korrekten nicht kolumnaren Genotyp zeigen. Die Untersuchung der Apfelsorten McIntosh und McIntosh Wijcik, die sich lediglich in der *Co*-Mutation unterscheiden, bestätigen die Hypothese, dass die Insertion von *Gypsy-44* die entscheidende Mutation ist, die zum kolumnaren Phänotyp führt. Die selbst erstellten Indel-basierenden Marker zeigen in PCRs bei beiden Sorten dasselbe Bandenmuster (Tabelle 3.14), wohingegen bei Verwendung der Transposon-basierenden Marker ein unterschiedliches Bandenmuster auftritt.

Die Sortenunabhängigkeit der Transposon-basierenden Marker – insbesondere im Hinblick auf die Sorte Topaz – sowie das unterschiedliche Bandenmuster zwischen McIntosh und McIntosh Wijcik beweisen, dass es sich im Gegensatz zu allen bisher erstellten Markern bei den Transposon-basierenden Markern nicht nur um Unterschiede zwischen dem kolumnaren und dem nicht kolumnaren Chromosom handelt, sondern um die *Co*-Mutation selbst.

Eine genauere Untersuchung der Integrationsstelle von *Gypsy-44* zeigt, dass sie sich in einem weiteren LTR-Retrotransposon (*Gypsy-33*) befindet (Abbildung 3.15). Die Tatsache, dass Retrotransposons in andere Retrotransposons integrieren, ist keine Seltenheit. Im Gesamtcontig konnten insgesamt neun solcher Ereignisse beobachtet werden, wobei in fünf Fällen nur noch ein solo-LTR Hinweise auf die Integration eines Retrotransposons liefert. Zudem wurde dieses Phänomen im Mais am Beispiel einer 280 kb Region, die das *Adh1-F* Gen enthält, beschrieben, in der zehn von 20 Retrotransposons in anderen Retrotransposons liegen (SanMiguel *et al.* 1996). Darüber hinaus findet man beim Apfel eine negative Korrelation zwischen Gen- und Retrotransposondichte (Troggio *et al.* 2012). Bereiche, in denen eine hohe Dichte an Genen vorliegt, enthalten wenige Retrotransposons und umgekehrt, wie es in den Abbildungen 3.19 und 3.20 deutlich zu erkennen ist. Das gleiche Phänomen tritt auch bei anderen Pflanzen auf, zum Beispiel bei der Hirse (Paterson *et al.* 2009). Zurückzuführen ist dies vermutlich auf die Tatsache, dass die Integration vor allem von LTR-Retrotransposons in Gene oder Genbereiche aufgrund ihrer Größe von 2000 bp bis 18.000 bp fatale Auswirkung auf die entsprechenden Gene und somit für die Wirtspflanze selbst hätte (Vitte und Panaud 2005; Bennetzen 2000).

Die Sequenzidentität zwischen den beiden LTRs eines Retrotransposons lässt Rückschlüsse über den Zeitpunkt der Integration zu, da die LTRs bei der Integration identisch sind, im Laufe der Zeit jedoch durch Substitution einzelner Basen immer mehr voneinander abweichen (SanMiguel *et al.* 1998). Allerdings wird zur Berechnung der Zeitspanne die Substitutionsrate benötigt, die jedoch von Spezies zu Spezies unterschiedlich und für den Apfel nicht verfügbar ist. Bei Gräsern liegt sie bei $6,5 \times 10^{-9}$ Substitutionen pro synonyme Stelle pro Jahr (Gaut *et al.* 1996). Auch ohne Berechnung des ungefähren Zeitpunktes der Integration fällt auf, dass die Sequenzidentität der LTRs bei verschachtelten Retrotransposons von innen nach außen abnimmt und damit das innere transposable Element in das außenliegende gesprungen ist. Im Fall von *Gypsy-44* und *Gypsy-33* betragen die Sequenzübereinstimmungen 94,6 % für das außenliegende *Gypsy-33* und 100 % für das in *Gypsy-33* integrierte Retrotransposon *Gypsy-44*. Die identischen LTRs von *Gypsy-44* lassen den Schluss zu, dass die Integration erst kürzlich stattgefunden haben kann, was gut damit übereinstimmt, dass ein Auftreten der *Co*-Mutation erstmalig vor rund 50 Jahren beschrieben wurde.

Wolters *et al.* (2013) wiesen parallel zu der vorliegenden Arbeit eine 1956 bp große Insertion bei McIntosh Wijcik als einzigen Unterschied in der *Co*-Zielregion von Baldi *et al.* (2012) im Vergleich zu McIntosh auf Grundlage von BAC-Banken der beiden Apfelsorten nach. Bei der Insertion, deren Position im Apfelgenom identisch mit dem in der vorliegenden Arbeit beschriebenen Integrationsort von *Gypsy-44* ist, handelt es sich höchstwahrscheinlich um einen solo-LTR des *Gypsy-44* Retrotransposons. Es sind flankierende „target sites“ vorhanden und diese identisch, sodass von einer Rekombination zwischen den LTRs einer Transposon-Kopie auszugehen ist. Ob die Rekombination jedoch tatsächlich im Genom von McIntosh Wijcik oder lediglich im untersuchten BAC während des Klonierungsprozesses stattgefunden hat bleibt offen. Damit bleibt zudem die Frage ungeklärt, ob die komplette Kopie von *Gypsy-44* für den kolumnaren Phänotyp verantwortlich ist oder ob letztlich der solo-LTR an dieser Position ausreicht. Als gesichert gelten kann jedoch, dass in der Apfelsorte P28 an der untersuchten Position eine vollständige Kopie von *Gypsy-44* vorhanden ist.

Das Auftreten einer Knospen- bzw. Sprossmutation, wie im Fall von McIntosh und McIntosh Wijcik, die durch eine Insertion eines Retrotransposons zur Entwicklung einer neuen Sorte führt, ist kein Einzelfall. So entstanden neue Apfelsorten aus Knospenmutationen der Sorten Gala und Braeburn durch Insertion von Retrotransposons (Venturi *et al.* 2006).

Darüber hinaus sind Sprossmutationen, oft durch Retrotransposons ausgelöst, entscheidend für die Sortenbildung bei Zitrusfrüchten (Asíns *et al.* 1999). Der Zusammenhang zwischen der Integration von *Gypsy-44* und der Ausprägung des kolumnaren Phänotyps ist jedoch schwierig zu ermitteln, da die möglichen Wirkmechanismen von Retrotransposons äußerst vielfältig sind (Lisch 2013). In den Apfelsorten Rae Ime, Spencer Seedless und Wellington Bloomless erzeugt die Integration eines LTR-Retrotransposons in ein Intron des *MdPI*-Gens eine Nullmutation des Gens, was zur Bildung kernloser Früchte führt (Yao *et al.* 2001). Ein weiteres Beispiel wurde beim Wein beschrieben; dort führt die Integration eines Retrotransposons in den Promotorbereich des Gens *Vvmyb1A*, welches die Anthocyanin-Biosynthese reguliert, zu einer „loss of function“-Mutation, was sich phänotypisch in der Ausbildung weißer Beeren zeigt (Kobayashi *et al.* 2004).

An Position 719.006 bis 719.299 und damit direkt neben der Integrationsstelle von *Gypsy-44* liegt das Gen für das Protein MDP0000766466. Es handelt sich aber vermutlich um das *gag*-homologe Gen des *Gypsy-33* Retrotransposons, da es ein typisches Cys/His-Finger Motiv enthält (Otto *et al.* 2013). Ansonsten liegen keine weiteren Gene in unmittelbarer Nähe. Damit kann die Inaktivierung von Genen durch Erzeugung einer Nullmutation durch die Integration von *Gypsy-44* in Gene oder deren unmittelbaren Promotorbereiche als sehr unwahrscheinlich angesehen werden. Möglicherweise hat das Transposon Einfluss auf weiter entfernt liegende Gene, beispielsweise durch die Insertion regulatorischer Elemente. Die Enhancer-Aktivität des Retrotransposons *Hopscotch*, welches im Mais rund 60 kb „upstream“ des *teosinte branched 1 (tb1)*-Locus integrierte, führt zu einer Überexpression von *tb1* (Studer *et al.* 2011). *Tb1* ist ein Transkriptionsfaktor, der die Ausbildung von Verzweigungen unterdrückt. Dies führt zu einer starken Verringerung der Seitentriebe im Mais im Vergleich zu der Vorläuferspezies *Teosinte* (Doebley *et al.* 2006). Betrachtet man die von der Integrationsstelle weiter entfernt liegenden Gene, befinden sich dort einige ursprüngliche Kandidaten für das *Co*-Gen, die in Zusammenhang mit der Entwicklung von Pflanzen stehen (Baldi *et al.* 2012). Es handelt sich dabei um Transkriptionsfaktoren der Familien „APETALA2/ethylene response factor“ (AP2/ERF), „basic helix-loop-helix“ (bHLH) und „No Apical Meristem“ (NAM). Drei in dieser Region annotierte Proteine (MDP0000855671, MDP0000286915 und MDP0000187369) zeigen signifikante Übereinstimmungen mit ERF-Transkriptionsfaktoren. Mitglieder der Familie AP2/ERF spielen Schlüsselrollen sowohl in der Entwicklung von Pflanzen, wie auch in der Reaktion

der Pflanzen auf Stress (Yu *et al.* 2012). Das Protein MDP0000139773 liegt im Bereich um 800 kb im Gesamtcontig und ist laut InterPro-Domänensuche vermutlich ein NAM-Transkriptionsfaktor. Diese Proteine sind unter anderem an der Entwicklung des Sprossapikalmeristems beteiligt (Souer *et al.* 1996). An Position 817 kb im Gesamtcontig und damit etwa 100 kb von *Gypsy-44* entfernt, befindet sich das Gen für das Protein MDP0000934866-like, ein bHLH-Transkriptionsfaktor. Die Funktion dieser Proteine ist sehr vielfältig und reicht von der Zell- und Gewebsentwicklung bis hin zu Stoffwechselprozessen (Heim *et al.* 2003). MDP0000934866-like ist das einzige der genannten für Transkriptionsfaktoren kodierenden Gene, für das eine Expression in den Transkriptomdaten des Sprossapikalmeristems der Sorte P28 nachgewiesen wurde (Tabelle 3.12). Weitere Gene in diesem Bereich sind die von Romina Petersen neu identifizierten Gene, die im GD-Genom nicht annotiert sind. Es handelt sich dabei um At1g08530-like, dessen Funktion unbekannt ist, um Downy Mildew Resistant 6-like, einem Regulator in der Abwehrreaktion von Pflanzen (Van Damme *et al.* 2005; 2008) und Auxin-induced protein 5NG4-like, das eine Funktion in der Wurzelbildung besitzt (Busov *et al.* 2004). Downy Mildew Resistant 6-like zeigt in Knospen eine signifikante Überexpression in McIntosh Wijkik im Vergleich zu McIntosh (Wolters *et al.* 2013). Welchen Einfluss die Integration von *Gypsy-44* allerdings tatsächlich auf die dort lokalisierten Gene hat, lässt sich nicht abschließend klären. Dazu wären weitere Experimente in Form einer Überexpression der Gene in nicht kolumnaren Apfelbäumen oder ein Ausschalten der Gene in kolumnaren Apfelbäumen erforderlich. Neben der Wirkung als Enhancer wäre auch ein gegenteiliger Effekt verursacht durch *Gypsy-44* als Insulator denkbar. Ein Insulator verhindert die Wechselwirkungen zwischen Enhancer und Promotor eines Gens (Bell und Felsenfeld 1999). Die Wirkung eines *Gypsy* Transposons als Insulator wurde beispielsweise in *Drosophila* beschrieben (Gause *et al.* 2001). Darüber hinaus wurde in Tabakzellen und in *Arabidopsis thaliana* nachgewiesen, dass DNA-Fragmente durch die Integration in den Bereich zwischen Enhancer und Promotor als Insulatoren in Pflanzen fungieren können (Nagaya *et al.* 2001; Hily *et al.* 2009).

Durch die Integration eines transposablen Elementes können neben regulatorischen Elementen auch die für die Transposition benötigten Gene durch zusätzliche Funktion Einfluss auf den Organismus haben (Muehlbauer *et al.* 2006). Das Retrotransposon *Gypsy-44* besitzt zwar einen transkribierten ORF von ungefähr 600 bp, allerdings konnten weder BLAST- noch Domänen-Suchen Aufschluss über die Funktion eines potenziellen Genproduktes geben. Damit besitzt *Gypsy-44* nicht die für die Transposition typischen ORFs

für GAG und Pol. Es stellt sich jedoch die Frage, wie *Gypsy-44* zur Transposition befähigt ist, obwohl es offensichtlich nicht autonom ist. Möglicherweise verhält es sich ähnlich wie bei dem Retrotransposon *Dasheng* im Reis. Bei *Dasheng* wurde ebenfalls eine kürzliche Transposition festgestellt und das, obwohl es wie *Gypsy-44* zwar wichtige cis-regulatorische Elemente wie PPT und PBS besitzt, die entscheidenden ORFs jedoch fehlen (Jiang *et al.* 2002a). Im Fall von *Dasheng* geht man davon aus, dass die Transposition bzw. die dafür notwendigen Proteine durch das autonome Retrotransposon *RIRE2* bereitgestellt wurden, dessen cis-regulatorischen Elemente hohe Ähnlichkeit zu denen von *Dasheng* aufweisen (Jiang *et al.* 2002b). Ob es sich im Apfel in Bezug auf *Gypsy-44* auch so verhält und ebenfalls ein autonomes Transposon mit ähnlichen cis-regulatorischen Elementen die Transposition von *Gypsy-44* ermöglichte, oder ob sich unter den zahlreichen weiteren Kopien von *Gypsy-44* im Genom möglicherweise autonome Vertreter befinden, bleibt offen.

In Pflanzen werden Retrotransposons in der Regel stark methyliert (Bennetzen *et al.* 1994). Diese Cytosin-Methylierungen haben starken Einfluss auf die Transkription, da sie unter anderem die Bindung regulatorischer Proteine verhindern (Ehrlich und Ehrlich 1993). Vermutlich ist die Methylierung eine Art Schutzmechanismus der Pflanze, die die Transposition der Retrotransposons verhindern soll (Hirochika *et al.* 2000). Da diese bei Retrotransposons über eine RNA-Zwischenstufe funktioniert, ist dieser Mechanismus sehr effektiv. Allerdings hat die Methylierung der Retrotransposons auch einen negativen Effekt auf die Expression benachbarter Gene (Finnegan *et al.* 1998). Die Komplexität der Methylierung wird am Beispiel des Tabak Retrotransposons *Tto1* deutlich. In *Arabidopsis* wird das aktive Retrotransposon *Tto1* methyliert und dadurch inaktiviert, im Tabak hingegen bleibt es selbst bei einer hohen Kopienzahl von über 300 transkriptionell aktiv (Hirochika *et al.* 2000). Aufgrund der transkriptionellen Aktivität von *Gypsy-44* in Blättern und im Sprossapikalmeristem in der untersuchten Region ist eine Methylierung zumindest in diesen Geweben unwahrscheinlich (Otto *et al.* 2013). Eine genaue Untersuchung der Region sowie anderer Gewebe bedarf spezieller Analyseverfahren, beispielsweise durch die AFLP-basierte Detektion von DNA-Methylierungen (Xu *et al.* 2000), da nicht absolut ausgeschlossen werden kann, dass die transkriptionelle Aktivität von *Gypsy-44* an dieser Position von anderen *Gypsy-44*-Kopien im Apfelgenom stammt.

Die Ursachen für die Aktivität von Retrotransposons sind in der Regel Stresssituationen für die Wirtspflanze (Casacuberta und Santiago 2003). Die Tabak-Retrotransposons *Tnt1* und *Tto1* sind unter Einwirkung von biotischen und abiotischen Stressfaktoren, die eine

Abwehrreaktion der Pflanze hervorrufen, transkriptionell aktiv (Beguiristain *et al.* 2001; Takeda *et al.* 1999). Auslöser für den Stress können beispielsweise Pilz- und Virusbefall oder eine Verletzung der Pflanze sowie UV-Strahlung oder die Herstellung von Zellkulturen sein (Pouteau *et al.* 1994). Die Verbindung zwischen der Abwehrreaktion einer Pflanze und der Aktivität von Retrotransposons rührt daher, dass die Promotoren einiger Retrotransposons denen der Abwehrgene der Pflanze ähneln (Wessler 1996). Möglicherweise liegt somit die Ursache für die transkriptionelle Aktivität von *Gypsy-44* in der identifizierten deutlichen Hochregulation von Genen für die Pflanzenabwehr und Stressreaktion in Blättern von McIntosh Wjick im Vergleich zu Blättern von McIntosh (Otto *et al.* 2013).

In Abbildung 4.3 sind die möglichen Wirkmechanismen des LTR-Retrotransposons *Gypsy-44* in der Zielregion zusammenfassend dargestellt. So könnte *Gypsy-44* in Form eines Enhancers auf eine Vielzahl benachbarter Gene wirken. Zudem wäre eine Rolle als Insulator möglich, der Auswirkungen auf die Expression eines etwa 40 kb „downstream“-liegenden Transkriptionsfaktors hätte. Darüber hinaus könnten die Transkripte von *Gypsy-44*, die sowohl in Blättern von McIntosh Wjick und im Sprossapikalmeristem von P28 identifiziert wurden, eine Funktion als Protein oder non-coding RNAs haben, die wiederum Auswirkungen auf die genomische Region haben. Über den genauen Wirkmechanismus kann nach den bisherigen Erkenntnissen nur spekuliert werden.

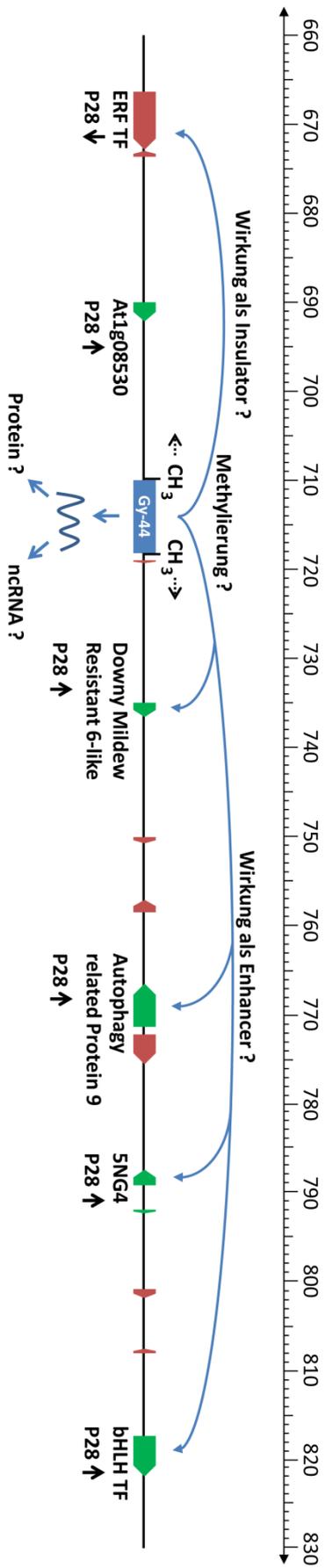


Abbildung 4.3 Mögliche Wirkmechanismen von *Gypsy-44* in der Co-Zielregion

Die Abbildung zeigt die Lokalisierung des LTR-Retrotransposons *Gypsy-44* (Gy-44) im erstellten Gesamtcontig. In grün und rot sind die umliegenden Gene dargestellt. Die Farben beziehen sich auf die Expression im Sprossapikalmeristem von P28 (rot = nicht exprimiert; grün = exprimiert). Der genaue Wirkmechanismus von *Gypsy-44*, der letztlich zur Ausprägung des kolumnaren Phänotyps führt, konnte nicht endgültig geklärt werden. Ein interessantes Gen in der unmittelbaren Nähe zur Integrationsstelle ist nicht vorhanden. Möglicherweise wirkt das Retrotransposon auf das hochregulierte Gen für einen basischen Helix-Loop-Helix Transkriptionsfaktor (bHLH TF) oder andere dargestellte Gene in Form eines Enhancers, was die gesteigerte Expression in P28 erklären würde. Eine eventuelle Methylierung des Retrotransposons könnte ebenfalls Auswirkung auf die Gene in dieser Region haben. Da das Retrotransposon in der Apfelsorte Wjicik (Blatt) und P28 (Sprossapikalmeristem) transkriptionell aktiv ist, könnten die Transkripte in Form eines möglichen Proteins oder in Form von small RNAs eine Wirkung auf die dargestellte Region haben. Unterhalb der exprimierten Gene ist die Information über deren Regulation (hoch- oder herunterreguliert) in Form eines Pfeiles dargestellt. Die Daten zur Genanordnung und zu den Expressionsniveaus stammen zum Teil aus der Doktorarbeit von Romina Petersen und beziehen sich auf das Sprossapikalmeristem von P28 im Vergleich zu A14.

5 Zusammenfassung

Im Rahmen dieser Arbeit sollten die molekularen Grundlagen der *Co*-Mutation, die für die kolumnare Wuchsform von Apfelbäumen verantwortlich ist, aufgeklärt werden. Die bisherigen Erkenntnisse hinsichtlich der Lokalisierung und der Struktur der *Co*-Mutation sind trotz der großen wirtschaftlichen Bedeutung begrenzt.

Durch eine klassische positionelle Klonierungsstrategie in Kombination mit genomischen Illumina Next-Generation Sequencing-Datensätzen konnte eine etwa 850 kb große Referenzsequenz der *Co*-Zielregion der kolumnaren Apfelsorte Procats 28 erstellt und verifiziert werden. Auf Basis dieser Referenz konnte die *Co*-Mutation in Form der Integration des LTR-Retrotransposons *Gypsy-44* im kolumnaren Chromosom an Position 18,79 Mbp auf Chromosom 10 lokalisiert und die Integration von *Gypsy-44* als einziger Unterschied zwischen McIntosh und McIntosh Wijcik identifiziert werden. Die Existenz eines „*Co*-Gens“ über das *Gypsy-44* Transposon hinaus kann aufgrund der Analyse der Gene in der überlappenden Zielregion von Baldi *et al.* (2012) und Moriya *et al.* (2012) sowie den angrenzenden Bereichen ausgeschlossen werden (Doktorarbeit Romina Petersen, IMSB). Zudem ist ein weiteres Ereignis neben der Integration von *Gypsy-44* innerhalb eines so begrenzten Chromosomenabschnittes zur etwa gleichen Zeit als höchst unwahrscheinlich einzustufen.

Anhand der durchgeführten Markeranalysen konnte eine direkte Kopplung zwischen der Integration des Retrotransposons *Gypsy-44* und dem kolumnaren Phänotyp hergestellt werden, allerdings ist der direkte Zusammenhang mit dem kolumnaren Phänotyp noch ungeklärt. Die Aufklärung des Wirkmechanismus gestaltet sich als schwierig, da durch die Integration kein Gen direkt betroffen ist. Zudem ist eine Integration des *Gypsy-44* Retrotransposons in eine nicht kolumnare Apfelsorte durch Transformationsexperimente nahezu unmöglich, da höchstwahrscheinlich nicht das Retrotransposon an sich, sondern die Integration an Position 18,79 auf Chromosom 10 für den *Co*-Phänotyp verantwortlich ist. Dennoch könnten durch die Überexpression benachbarter Gene in normalwüchsigen Apfelbäumen, oder das Ausschalten dieser in kolumnaren Bäumen, mögliche Wirkmechanismen identifiziert werden.

Durch die Etablierung der Transposon-basierenden Marker ist es gelungen, die bisher ersten und einzigen molekularen Marker zu erstellen, die in Bezug auf die *Co*-Mutation sortenunabhängig sind und somit eine verlässliche Genotypisierung von Apfelbäumen in Bezug auf das Kolumnarwachstum ermöglichen.

Abkürzungsverzeichnis

%	Prozent
A14	A14-190-93
abs.	absolut
Ac	Activator
AFLP	amplified fragment length polymorphism
AMI GmbH	Agrarmarkt Informations-Gesellschaft mbH
BAC	Bacterial Artificial Chromosome
BLAST	Basic Local Alignment Search Tool
bp	Basenpaare
BPB	Bromphenolblau
<i>BpMADS4</i>	<i>Betula pendula MADS4</i>
Bq	Becquerel
<i>br</i>	<i>broomy</i>
BSA	bovine serum albumin
bspw.	beispielsweise
bzw.	beziehungsweise
ca.	circa
Cl	Chloroform/Isoamylalkohol
cM	centiMorgan
CNS	conserved non-coding sequence
Co-Gen	Kolumnargen
Dam	DNA adenine methylase
dATP	Desoxyadenosintriphosphat
dCTP	Desoxycytosintriphosphat
dGTP	Desoxyguanosintriphosphat
DMF	Dimethylformamid
DNA	Desoxyribonukleinsäure
dNTP	Desoxynukleosidtriphosphat
Ds	Dissociator
dTTP	Desoxythymidintriphosphat
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	Ethylendiamintetraacetat
EN	Endonuklease
E-Puffer	Natriumphosphatpuffer-System
<i>et al.</i>	<i>et alii/aliae/alia</i>
EtBr	Ethidiumbromid
Fa.	Firma
FAO	Food and Agriculture Organization of the United Nations
<i>FWA</i>	<i>FLOWERING WAGENINGEN</i>
GAG	Capsid Protein
GD	Golden Delicious
GDR	Genome Database for Rosaceae

GWD	genomweite Duplikation
h	Stunde
ha	Hektar
hAT	hobo-Ac-Tam3
HPLC	high-performance liquid chromatography
IAA	Indol-3-essigsäure
INT	Integrase
IMSB	Institut für Molekulargenetik, gentechnologische Sicherheitsforschung und Beratung
IPTG	Isopropyl-1-thio-beta-D-Galactopyranosid
kb	Kilobase
KFH	Koloniefilter-Hybridisierung
kg	Kilogramm
kV	Kilovolt
l	Liter
LB	lysogeny broth
LG	Linkage Group
LINE	long interspersed nuclear element
lr	long range
LTR	long terminal repeat
M	Molar
mA	Milliampere
Mbp	Megabase
MCS	multiple cloning site
<i>MdPI</i>	<i>Malus x domestica PISTILLATA</i>
mg	Milligramm
µg	Mikrogramm
min	Minute
Mio.	Million
MITE	Miniature inverted-repeat transposable element
ml	Milliliter
µl	Mikroliter
mM	Millimolar
ms	Millisekunde
Mt	Megatonne
NaOH	Natriumhydroxid
NCBI	National Center for Biotechnology Information
nm	Nanometer
OD	Optische Dichte
P28	Procats 28
PAC	P-1 derived Artificial Chromosome
PBS	primer binding site
PCI	Phenol/Chloroform/Isoamylalkohol
PCR	Polymerase-Ketten-Reaktion
PEG	Polyethylenglycol

ABKÜRZUNGSVERZEICHNIS

pers.	persönlich
pg	Picogramm
pH	potentia Hydrogenii
PM	Präinkubationsmedium
PPT	polpurine tract
PR	Protease
PVP	Polyvinylpyrrolidon
RAPD	random amplified polymorphic DNA
RFLP	restriction fragment length polymorphism
RH	RNaseH
RNA	Ribonukleinsäure
RNase	Ribonuklease
RT	Reverse Transkriptase
RTE	retrotransposable element
s	Sekunde
SCAR	sequence characteristic amplified region
SDS	Sodiumdodecylsulfat
SINE	short interspersed nuclear element
SNP	single nucleotide polymorphism
s.o.	siehe oben
SSC	Standard-Salz-Citrat
SSR	simple sequence repeat
TAGI	The Arabidopsis Genome Initiative
<i>tb1</i>	<i>teosinte branched 1</i>
TBE	Tris-Borat-EDTA
TE	Tris-EDTA
TIR	terminal inverted repeat
Tpase	Transposase
TS	Target Site
TSD	Target Site Duplication
u.a.	unter anderem
ü.N.	über Nacht
Upm	Umdrehungen pro Minute
UV	ultraviolett
V	Volt
Vdf	Verband der deutschen Fruchtsaft-Industrie e.V.
VE	voll entsalzt
vergl.	vergleiche
<i>Vgt1</i>	<i>Vegetative to generative transition 1</i>
<i>Wx</i>	<i>Waxy</i>
X-Gal	5-Brom-4-chlor-3-indolyl-alpha-D-galactopyranosid
YAC	Yeast Artificial Chromosome
<i>ZmRAP2.7</i>	<i>Zea mays RAP2.7</i>

Abbildungsverzeichnis

Abbildung 1.1 Kolumnare Apfelbäume.....	4
Abbildung 1.2 Transposable Elemente in Pflanzen.....	10
Abbildung 1.3 <i>Co</i> -Gen gekoppelte Marker	18
Abbildung 2.1 Verwendete Molekulargewichtsstandards der Fa. Fermentas	21
Abbildung 2.2 Restriktionsschnittstelle <i>Sau3AI</i>	30
Abbildung 3.1 Vereinfachte genetische bzw. physikalische Karte der <i>Co</i> -Gen-Region	49
Abbildung 3.2 Gelelektrophoretische Auftrennung genomischer P28-DNA	50
Abbildung 3.3 Restriktionsverdau der Plasmide ausgewählter BAC-Klone.....	53
Abbildung 3.4 „Pirota“ Blot potenzieller Sonden	55
Abbildung 3.5 Restriktionskartierung des BAC-Klons 3G6.....	59
Abbildung 3.6 Tatsächliche Abstände der „mate pair“-Reads.....	61
Abbildung 3.7 Mapping der „mate pair“-Reads gegen die BAC-Klone K2_3 und 2I4.....	63
Abbildung 3.8 Schema des BAC-Klons 2I4 inklusive BLASTn-Treffern	64
Abbildung 3.9 Southern Blots des BAC-Klons 2I4	65
Abbildung 3.10 <i>Co</i> -Zielregion auf Chromosom 10, Position 18 bis 19 Mbp	67
Abbildung 3.11 Mapping der „mate pair“-Reads gegen den Gesamtcontig	68
Abbildung 3.12 Metacontigs vs. GD-Genom Teil I.....	69
Abbildung 3.13 Metacontigs vs. GD-Genom Teil II	70
Abbildung 3.14 Metacontigs vs. GD-Genom Teil III	71
Abbildung 3.15 <i>Gypsy-44</i> in der Region um 18,79 Mbp auf Chromosom 10	74
Abbildung 3.16 Annotierte Gene im erstellten Gesamtcontig	75
Abbildung 3.17 Annotierte transposable Elemente in der Zielregion	80
Abbildung 3.18 Größenverteilung der annotierten transposablen Elemente.....	81
Abbildung 3.19 Annotierte Gene und transposable Elemente im Gesamtcontig Teil I.....	83
Abbildung 3.20 Annotierte Gene und transposable Elemente im Gesamtcontig Teil II.....	84
Abbildung 3.21 Potenzielle <i>Co</i> -Gen gekoppelte Marker	86
Abbildung 3.22 <i>S1</i> -Nuklease-Verdau ausgewählter <i>Co</i> -Gen gekoppelter Marker	87
Abbildung 3.23 Feinkartierung der <i>Co</i> -Zielregion.....	89
Abbildung 3.24 Markertests zur Feinkartierung der Zielregion.....	90
Abbildung 3.25 Genotypisierung der Apfelsorte Topaz mittels verschiedener Marker	93
Abbildung 3.26 <i>Gypsy-44</i> -basierende Primer	94
Abbildung 3.27 <i>Gypsy-44</i> -spezifische PCR auf genomische DNAs diverser Apfelsorten	95
Abbildung 4.1 <i>Co</i> -Zielregion.....	98
Abbildung 4.2 Evolution des Apfelgenoms	102
Abbildung 4.3 Mögliche Wirkmechanismen von <i>Gypsy-44</i> in der <i>Co</i> -Zielregion.....	116
Abbildung A1.1 Vektorkarten Teil I.....	137
Abbildung A1.2 Vektorkarten Teil II.....	138
Abbildung A3.1 QIAxcel-Lauf	141
Abbildung A4.1 Markertests diverser Apfelsorten	142

Tabellenverzeichnis

Tabelle 1.1 Eine Auswahl sequenzierter Pflanzengenome	7
Tabelle 2.1 Standard PCR-Ansatz	24
Tabelle 2.2 PCR-Programm.....	24
Tabelle 2.3 „long range“ PCR-Ansatz.....	24
Tabelle 2.4 „long range“ PCR-Programm	24
Tabelle 2.5 Endfilling (radioaktiv).....	30
Tabelle 2.6 Ligationsansatz	31
Tabelle 2.7 Restriktionsansatz BAC-Minipräparation.....	33
Tabelle 2.8 Endfilling-Ansatz	37
Tabelle 2.9 Ligations-Ansatz	37
Tabelle 2.10 Restriktionsansatz pUC.....	38
Tabelle 2.11 Verwendete Perl-Skripte	39
Tabelle 2.12 Parameter der Assembly Cell.....	40
Tabelle 2.13 Restriktionsenzyme und Puffer	41
Tabelle 3.1 Strahlungsintensitäten zur Überprüfung der Integrat-Enden	52
Tabelle 3.2 Integratgrößen ausgewählter BAC-Klone	54
Tabelle 3.3 BAC-DNA-Bibliotheken der Apfelsorte Procats 28	54
Tabelle 3.4 Lokalisierung positiver BAC-Klone der Koloniefilter-Hybridisierung	56
Tabelle 3.5 BAC-Klone aus der <i>Co</i> -Zielregion	57
Tabelle 3.6 Berechnete Fragmentgrößen des Klons 3G6	59
Tabelle 3.7 Statistische Auswertung der genomischen Illumina „mate pair“-Reads.....	61
Tabelle 3.8 Statistische Analyse von Duplikaten in den „mate pair“-Bibliotheken	62
Tabelle 3.9 Zusammensetzung chimärer BAC-Klone.....	63
Tabelle 3.10 Erwartete Fragmentgrößen der Southern Blot Analysen des BAC-Klons 2I4	64
Tabelle 3.11 Identifizierte Indels in der <i>Co</i> -Zielregion	73
Tabelle 3.12 Im GD-Genom annotierte Gene in der Zielregion	77
Tabelle 3.13 Potenzielle <i>Co</i> -Gen gekoppelte Marker.....	85
Tabelle 3.14 Markertests auf diverse Apfelsorten.....	91
Tabelle A2.1 Sonden in der Zielregion für die KFH Runde I (Sondenpool I)	139
Tabelle A2.2 Sonden in der Zielregion für die KFH Runde II (Sondenpool II)	139
Tabelle A2.3 Sonden in der Zielregion für die KFH Runde III (Sondenpool III)	140
Tabelle A2.4 Sonden in der Zielregion für die KFH Runde IV (Sondenpool IV)	140
Tabelle A3.1 Plattenbelegung PCR und QIAxcel.....	141

Webverzeichnis

AMI GmbH	ami-informiert.de
„apple genome browser“	genomics.research.iasma.it/gb2/gbrowse/apple
BLAST	genomics.research.iasma.it/blast/blast.html
CENSOR	girinst.org/censor
FAO	faostat.fao.org
GDR	rosaceae.org
PLAZA	bioinformatics.psb.ugent.be/plaza
Repbase	girinst.org/repbase
Spidey	ncbi.nlm.nih.gov/spidey/
Vdf	fruchtsaft.org
Vecscreen	ncbi.nlm.nih.gov/tools/vecscreen/

Literaturverzeichnis

- Aird D, Ross MG, Chen WS, Danielsson M, Fennell T, Russ C, Jaffe DB, Nusbaum C, Gnirke A (2011) Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome biology* 12 (2):R18
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215 (3):403-410
- Asíns MJ, Monforte AJ, Mestre PF, Carbonell EA (1999) Citrus and Prunus-like retrotransposons. *Theor Appl Genet* 99 (3-4):503-510
- Bai T, Zhu Y, Fernández-Fernández F, Keulemans J, Brown S, Xu K (2012) Fine genetic mapping of the Co locus controlling columnar growth habit in apple. *Molecular Genetics and Genomics* 287 (5):437-450
- Baldi P, Wolters P, Komjanc M, Viola R, Velasco R, Salvi S (2012) Genetic and physical characterisation of the locus controlling columnar habit in apple (*Malus × domestica* Borkh.). *Molecular Breeding*:1-12
- Beckmann JS (1988) Oligonucleotide Polymorphisms: A New Tool for Genomic Genetics. *Nat Biotech* 6 (9):1061-1064
- Beek JG, Verkerk R, Zabel P, Lindhout P (1992) Mapping strategy for resistance genes in tomato based on RFLPs between cultivars: Cf9 (resistance to *Cladosporium fulvum*) on chromosome 1. *TAG Theoretical and Applied Genetics* 84 (1):106-112
- Beguiristain T, Grandbastien MA, Puigdomenech P, Casacuberta JM (2001) Three Tnt1 subfamilies show different stress-associated patterns of expression in tobacco. Consequences for retrotransposon control and evolution in plants. *Plant physiology* 127 (1):212-221
- Bell AC, Felsenfeld G (1999) Stopped at the border: boundaries and insulators. *Current opinion in genetics & development* 9 (2):191-198
- Bennetzen J (2000) Transposable element contributions to plant gene and genome evolution. *Plant Mol Biol* 42 (1):251-269
- Bennetzen JL (1996) The contributions of retroelements to plant genome organization, function and evolution. *Trends in Microbiology* 4 (9):347-353
- Bennetzen JL, Schrick K, Springer PS, Brown WE, SanMiguel P (1994) Active maize genes are unmodified and flanked by diverse classes of modified, highly repetitive DNA. *Genome* 37 (4):565-576
- Blažek J (1992) Segregation and general evaluation of spur type or compact growth habits in apples. *Acta Hort* 317:71-79
- Blobel G, Potter VR (1966) Nuclei from Rat Liver: Isolation Method That Combines Purity with High Yield. *Science* 154 (3757):1662-1665
- Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American journal of human genetics* 32 (3):314-331
- Boyer J, Liu R (2004) Apple phytochemicals and their health benefits. *Nutrition Journal* 3 (1):5
- Buhariwalla H, Greaves S, Magrath R, Mithen R (1995) Development of specific PCR primers for the amplification of polymorphic DNA from the obligate root pathogen *Plasmodiophora brassicae*. *Physiological and Molecular Plant Pathology* 47 (2):83-94

- Bundock P, Hooykaas P (2005) An Arabidopsis hAT-like transposase is essential for plant development. *Nature* 436 (7048):282-284
- Bureau TE, Wessler SR (1992) Tourist: a large family of small inverted repeat elements frequently associated with maize genes. *The Plant cell* 4 (10):1283-1294
- Bureau TE, Wessler SR (1994) Stowaway: a new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants. *The Plant cell* 6 (6):907-916
- Burgoyne LA, Waqar MA, Atkinson MR (1970) Calcium-dependent priming of DNA synthesis in isolated rat liver nuclei. *Biochemical and Biophysical Research Communications* 39 (2):254-259
- Burke D, Carle G, Olson M (1987) Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* 236 (4803):806-812
- Busov V, Johannes E, Whetten R, Sederoff R, Spiker S, Lanz-Garcia C, Goldfarb B (2004) An auxin-inducible gene from loblolly pine (*Pinus taeda* L.) is differentially expressed in mature and juvenile-phase shoots and encodes a putative transmembrane protein. *Planta* 218 (6):916-927
- Calladine CR, Collis CM, Drew HR, Mott MR (1991) A study of electrophoretic mobility of DNA in agarose and polyacrylamide gels. *J Mol Biol* 221 (3):981-1005
- Casacuberta JM, Santiago N (2003) Plant LTR-retrotransposons and MITEs: control of transposition and impact on the evolution of plant genes and genomes. *Gene* 311:1-11
- Chagné D, Gasic K, Crowhurst RN, Han Y, Bassett HC, Bowatte DR, Lawrence TJ, Rikkerink EHA, Gardiner SE, Korban SS (2008) Development of a set of SNP markers present in expressed genes of the apple. *Genomics* 92 (5):353-358
- Chaisson MJ, Brinza D, Pevzner PA (2009) De novo fragment assembly with short mate-paired reads: Does the read length matter? *Genome Research* 19 (2):336-346
- Chevreau E, Lespinasse Y, Gallet M (1985) Inheritance of pollen enzymes and polyploid origin of apple (*Malus x domestica* Borkh.). *TAG Theoretical and Applied Genetics* 71 (2):268-277
- Choi S, Creelman R, Mullet J, Wing R (1995) Construction and characterization of a bacterial artificial chromosome library of *Arabidopsis thaliana*. *Plant Mol Biol Rep* 13 (2):124-128
- Clarke L, Carbon J (1976) A colony bank containing synthetic CoI EI hybrid plasmids representative of the entire *E. coli* genome. *Cell* 9 (1):91-99
- Claros MG, Bautista R, Guerrero-Fernández D, Benzerki H, Seoane P, Fernández-Pozo N (2012) Why Assembling Plant Genome Sequences Is So Challenging. *Biology* 1 (2):439-459
- Conesa A, Götz S (2008) Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics. *International Journal of Plant Genomics* 2008
- Conner PJ, Brown SK, Weeden NF (1997) Randomly Amplified Polymorphic DNA-based Genetic Linkage Maps of Three Apple Cultivars. *J Amer Soc Hort Sci* 122 (3):350-359
- Cornille A, Gladieux P, Smulders MJM, Roldán-Ruiz I, Laurens F, Le Cam B, Nersesyan A, Clavel J, Olonova M, Feugey L *et al.* (2012) New Insight into the History of Domesticated Apple: Secondary Contribution of the European

- Wild Apple to the Genome of Cultivated Varieties. *PLoS Genet* 8 (5):e1002703
- de Bruijn NG, Erdos P (1946) A combinatorial problem. *Koninklijke Nederlandse Akademie v Wetenschappen* 49 (49):758-764
- Denhardt DT (1966) A membrane-filter technique for the detection of complementary DNA. *Biochemical and Biophysical Research Communications* 23 (5):641-646
- Doebley JF, Gaut BS, Smith BD (2006) The Molecular Genetics of Crop Domestication. *Cell* 127 (7):1309-1321
- Dower WJ, Miller JF, Ragsdale CW (1988) High efficiency transformation of E.coli by high voltage electroporation. *Nucleic acids research* 16 (13):6127-6145
- Ehrlich M, Ehrlich KC (1993) Effect of DNA methylation on the binding of vertebrate and plant proteins to DNA. *Exs* 64:145-168
- Evans RC, Campbell CS (2002) The origin of the apple subfamily (Maloideae; Rosaceae) is clarified by DNA sequence data from duplicated GBSSI genes. *American journal of botany* 89 (9):1478-1484
- Feinberg AP, Vogelstein B (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal Biochem* 132 (1):6-13
- Fernández-Fernández F, Evans K, Clarke J, Govan C, James C, Marič S, Tobutt K (2008) Development of an STS map of an interspecific progeny of *Malus*. *Tree Genetics & Genomes* 4 (3):469-479
- Fernández-Fernández F, Harvey NG, James CM (2006) Isolation and characterization of polymorphic microsatellite markers from European pear (*Pyrus communis* L.). *Molecular Ecology Notes* 6 (4):1039-1041
- Feschotte C, Jiang N, Wessler SR (2002) Plant transposable elements: where genetics meets genomics. *Nat Rev Genet* 3 (5):329-341
- Feschotte C, Swamy L, Wessler SR (2003) Genome-Wide Analysis of mariner-Like Transposable Elements in Rice Reveals Complex Relationships With Stowaway Miniature Inverted Repeat Transposable Elements (MITEs). *Genetics* 163 (2):747-758
- Finnegan DJ (1989) Eukaryotic transposable elements and genome evolution. *Trends in Genetics* 5 (0):103-107
- Finnegan EJ, Genger RK, Peacock WJ, Dennis ES (1998) DNA Methylation in Plants. *Annual review of plant physiology and plant molecular biology* 49:223-247
- Fisher DV (1969) Spur-type strains of McIntosh for high density plantings. *British Columbia Fruit Growers' Association Quarterly Report* 14 (2):3-10
- Fisher DV (1970) Spur strains of 'McIntosh' discovered in British Columbia, Canada. *Fruit Var Hort Dig* 24:27-32
- Flachowsky H, Hättasch C, Höfer M, Peil A, Hanke M-V (2010) Overexpression of LEAFY in apple leads to a columnar phenotype with shorter internodes. *Planta* 231 (2):251-263
- Flachowsky H, Le Roux P-M, Peil A, Patocchi A, Richter K, Hanke M-V (2011) Application of a high-speed breeding technology to apple (*Malus × domestica*) based on transgenic early flowering plants and marker-assisted selection. *New Phytologist* 192 (2):364-377
- Forsline PL, Aldwinckle HS, Dickson EE, Luby JJ, Hokanson SC (2002) Collection, Maintenance, Characterization, and Utilization of Wild Apples of Central Asia. *Horticultural Reviews*. John Wiley & Sons, Inc.,

- Fujimoto R, Kinoshita Y, Kawabe A, Kinoshita T, Takashima K, Nordborg M, Nasrallah ME, Shimizu KK, Kudoh H, Kakutani T (2008) Evolution and Control of Imprinted *FWA* Genes in the Genus *Arabidopsis*. *PLoS Genet* 4 (4):e1000048
- Gause M, Morcillo P, Dorsett D (2001) Insulation of enhancer-promoter communication by a gypsy transposon insert in the *Drosophila cut* gene: cooperation between suppressor of hairy-wing and modifier of *mdg4* proteins. *Molecular and cellular biology* 21 (14):4807-4817
- Gaut BS, Morton BR, McCaig BC, Clegg MT (1996) Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcl*. *Proceedings of the National Academy of Sciences* 93 (19):10274-10279
- Gerhauser C (2008) Cancer chemopreventive potential of apples, apple juice, and apple components. *Planta Med* 74 (13):1608-1624
- Grant SG, Jessee J, Bloom FR, Hanahan D (1990) Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants. *Proceedings of the National Academy of Sciences* 87 (12):4645-4649
- Green ED (2001) Strategies for the systematic sequencing of complex genomes. *Nature reviews Genetics* 2 (8):573-583
- Green ED, Riethman HC, Dutchik JE, Olson MV (1991) Detection and characterization of chimeric yeast artificial-chromosome clones. *Genomics* 11 (3):658-669
- Gregory TR (2005) The C-value Enigma in Plants and Animals: A Review of Parallels and an Appeal for Partnership. *Annals of Botany* 95 (1):133-146
- Grunstein M, Hogness DS (1975) Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. *Proceedings of the National Academy of Sciences* 72 (10):3961-3965
- Hagerman PJ (1984) Evidence for the existence of stable curvature of DNA in solution. *Proceedings of the National Academy of Sciences* 81 (15):4632-4636
- Hamilton JP, Robin Buell C (2012) Advances in plant genome sequencing. *The Plant Journal* 70 (1):177-190
- Hamrick J, Godt M, Brown A, Clegg M, Kahler A, Weir B (1990) Allozyme diversity in plant species. *Plant population genetics, breeding, and genetic resources*:43-63
- Hattman S, Brooks JE, Masurekar M (1978) Sequence specificity of the P1 modification methylase (M.Eco P1) and the DNA methylase (M.Eco dam) controlled by the *Escherichia coli* *dam* gene. *J Mol Biol* 126 (3):367-380
- Heim MA, Jakoby M, Werber M, Martin C, Weisshaar B, Bailey PC (2003) The Basic Helix-Loop-Helix Transcription Factor Family in Plants: A Genome-Wide Study of Protein Structure and Functional Diversity. *Molecular Biology and Evolution* 20 (5):735-747
- Hemmat M, Weeden NF, Conner PJ, Brown SK (1997) A DNA Marker for Columnar Growth Habit in Apple Contains a Simple Sequence Repeat. *J Amer Soc Hort Sci* 122 (3):347-349
- Hemmat M, Weeden NF, Manganaris AG, Lawson DM (1994) Molecular marker linkage map for apple. *The Journal of heredity* 85 (1):4-11
- Herold T (2012) Vergleichende Analyse der Gene und der Genomstruktur der geschlechtsbestimmenden Chromosomenregion von *Chironomus* und *Camptochironomus*. Dissertation, Johannes Gutenberg Universität, Mainz

- Hily JM, Singer SD, Yang Y, Liu Z (2009) A transformation booster sequence (TBS) from *Petunia hybrida* functions as an enhancer-blocking insulator in *Arabidopsis thaliana*. *Plant cell reports* 28 (7):1095-1104
- Hirochika H, Okamoto H, Kakutani T (2000) Silencing of retrotransposons in *Arabidopsis* and reactivation by the *ddm1* mutation. *The Plant cell* 12 (3):357-369
- Hopf M (1973) *Apfel (Malus communis L.); Aprikose (Prunus armeniaca L.)*, vol 1. *Reallexikon der Germanischen Altertumskunde*. Walter de Gruyter,
- Ikase L, Dumbravs R (2004) Breeding of columnar apple-trees in Latvia. *Biologija* 2:8-10
- Imelfort M, Edwards D (2009) De novo sequencing of plant genomes using second-generation technologies. *Briefings in Bioinformatics* 10 (6):609-618
- Ioannou PA, Amemiya CT, Garnes J, Kroisel PM, Shizuya H, Chen C, Batzer MA, de Jong PJ (1994) A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nature genetics* 6 (1):84-89
- Jacob HB (2010) Breeding experiments of apple varieties with columnar growth and low chilling requirements. *Acta Hort* 872 (159-164)
- Jaillon O, Aury J, Noel B, Policriti A, Clepet C (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449 (7161):463-467
- Janick J, Cummins JN, Brown SK, Hemmat M (1996) *Apples*. In *Fruit Breeding, Volume I: Tree and Tropical Fruits*. Wiley & Sons, Inc.
- Jiang N, Bao Z, Temnykh S, Cheng Z, Jiang J, Wing RA, McCouch SR, Wessler SR (2002a) Dasheng: A Recently Amplified Nonautonomous Long Terminal Repeat Element That Is a Major Component of Pericentromeric Regions in Rice. *Genetics* 161 (3):1293-1305
- Jiang N, Feschotte C, Zhang X, Wessler SR (2004) Using rice to understand the origin and amplification of miniature inverted repeat transposable elements (MITEs). *Current Opinion in Plant Biology* 7 (2):115-119
- Jiang N, Jordan IK, Wessler SR (2002b) Dasheng and RIRE2. A Nonautonomous Long Terminal Repeat Element and Its Putative Autonomous Partner in the Rice Genome. *Plant Physiology* 130 (4):1697-1705
- Juniper BE, Mabblerley DJ (2006) *The story of the apple*. Timber Press,
- Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J (2005) Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and genome research* 110 (1-4):462-467
- Kapitonov VV, Jurka J (2001) Rolling-circle transposons in eukaryotes. *Proceedings of the National Academy of Sciences* 98 (15):8714-8719
- Kaufmann P, Pfefferkorn A, Teuber M, Meile L (1997) Identification and quantification of *Bifidobacterium* species isolated from food with genus-specific 16S rRNA-targeted probes by colony hybridization and PCR. *Applied and Environmental Microbiology* 63 (4):1268-1273
- Kenis K, Keulemans J (2007) Study of Tree Architecture of Apple (*Malus × domestica* Borkh.) by QTL Analysis of Growth Traits. *Mol Breeding* 19:193-208
- Kim JM, Vanguri S, Boeke JD, Gabriel A, Voytas DF (1998) Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete *Saccharomyces cerevisiae* genome sequence. *Genome research* 8 (5):464-478

- Kim MY, Song KJ, Hwang J-H, Shin Y-U, Lee HJ (2003) Development of RAPD and SCAR markers linked to the Co gene conferring columnar growth habit in apple (*Malus pumila* Mill.). *Journal of Horticultural Science & Biotechnology* 78 (4):512-517
- Kim U-J, Birren BW, Slepak T, Mancino V, Boysen C, Kang H-L, Simon MI, Shizuya H (1996) Construction and Characterization of a Human Bacterial Artificial Chromosome Library. *Genomics* 34 (2):213-218
- Kinoshita Y, Saze H, Kinoshita T, Miura A, Soppe WJJ, Koornneef M, Kakutani T (2007) Control of FWA gene silencing in *Arabidopsis thaliana* by SINE-related direct repeats. *The Plant Journal* 49 (1):38-45
- Kobayashi S, Goto-Yamamoto N, Hirochika H (2004) Retrotransposon-Induced Mutations in Grape Skin Color. *Science* 304 (5673):982-982
- Kohany O, Gentles A, Hankus L, Jurka J (2006) Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinformatics* 7 (1):474
- Koo HS, Wu HM, Crothers DM (1986) DNA bending at adenine . thymine tracts. *Nature* 320 (6062):501-506
- Korban SS, Chen H (1992) Apple. In: *Biotechnology of perennial fruit crops, Biotechnology in agriculture, vol 8*. CAB International, Wallingford, UK, pp 203-227
- Korban SS, Skirvin RM (1984) Nomenclature of the domesticated apple. *Hort Science* 19:177-180
- Krost C (2012) *Transkriptom des Meristemgewebes des Kolumnarapfels*. Dissertation, Johannes Gutenberg Universität, Mainz
- Krost C, Petersen R, Schmidt ER (2012) The transcriptomes of columnar and standard type apple trees (*Malus x domestica*) - A comparative study. *Gene* 498:223-230
- Kumar A, Bennetzen JL (1999) Plant Retrotransposons. *Annual Review of Genetics* 33 (1):479-532
- Kumar S, Blaxter ML (2010) Comparing de novo assemblers for 454 transcriptome data. *BMC genomics* 11:571
- Kwok P-Y (2001) Methods for Genotyping Single Nucleotide Polymorphisms. *Annual Review of Genomics and Human Genetics* 2 (1):235-258
- Langmead B, Trapnell C, Pop M, Salzberg S (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* 10 (3):R25
- Lapins KO (1969) Segregation of compact growth types in certain apple seedling progenies. *Canadian Journal of Plant Science* 49:765-768
- Lapins KO (1976) Inheritance of compact growth type in apple. *J Am Soc Hort Sci* 101:133-135
- Lapins KO, Watkins R (1973) Genetics of compact growth habit. Report of East Malling Research Station for 1972 136
- Lee JM, Looney NE (1977) Absisic Acid Levels and Genetic Compaction in Apple Seedlings. *Canad J Plant Sci* 57 (1):81-85
- Li R, Fan W, Tian G, Zhu H, He L, Cai J, Huang Q, Cai Q, Li B, Bai Y *et al.* (2010) The sequence and de novo assembly of the giant panda genome. *Nature* 463 (7279):311-317
- Liebhart R, Gianfranceschi L, Koller B, Ryder CD, Tarchini R, Van de Weg E, Gessler C (2002) Development and characterisation of 140 new

- microsatellites in apple (*Malus x domestica* Borkh.). *Molecular Breeding* 10:217-241
- Lisch D (2009) Epigenetic regulation of transposable elements in plants. *Annual review of plant biology* 60:43-66
- Lisch D (2013) How important are transposons for plant evolution? *Nat Rev Genet* 14 (1):49-61
- Litt M, Luty JA (1989) A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *American journal of human genetics* 44 (3):397-401
- Looney NE, Lane WD (1984) Spur-type growth mutants of McIntosh apple: A review of their genetics, physiology and field performance. *Acta Hort* 146:31-46
- Luby JJ (2003) *Apples: Botany, Production and Uses*. CABI Publishing,
- Luo M-C, Thomas C, You FM, Hsiao J, Ouyang S, Buell CR, Malandro M, McGuire PE, Anderson OD, Dvorak J (2003) High-throughput fingerprinting of bacterial artificial chromosomes using the snapshot labeling kit and sizing of restriction fragments by capillary electrophoresis. *Genomics* 82 (3):378-389
- Lynch M, Walsh JB (1998) *Genetics and Analysis of Quantitative Traits*. Sinauer Assocs., Inc., Sunderland, MA. ,
- Maliepaard C, Alston FH, van Arkel G, Brown LM, Chevreau E, Dunemann F, Evans KM, Gardiner S, Guilford P, van Heusden AW *et al.* (1998) Aligning male and female linkage maps of apple (*Malus pumila* Mill.) using multi-allelic markers. *TAG Theoretical and Applied Genetics* 97 (1):60-73
- Malik HS, Eickbush TH (1998) The RTE class of non-LTR retrotransposons is widely distributed in animals and is the origin of many SINEs. *Molecular Biology and Evolution* 15 (9):1123-1134
- Malnoy M, Jin Q, Borejsza-Wysocka EE, He SY, Aldwinckle HS (2007) Overexpression of the Apple MpNPR1 Gene Confers Increased Disease Resistance in *Malus x domestica*. *Molecular Plant-Microbe Interactions* 20 (12):1568-1580
- Marini JC, Levene SD, Crothers DM, Englund PT (1982) Bent helical structure in kinetoplast DNA. *Proceedings of the National Academy of Sciences* 79 (24):7664-7668
- McClintock B (1944) The Relation of Homozygous Deficiencies to Mutations and Allelic Series in Maize. *Genetics* 29 (5):478-502
- McClintock B (1946) *Maize genetics*. Year book - Carnegie Institution of Washington 45:176-186
- McClintock B (1950) The origin and behavior of mutable loci in maize. *Proceedings of the National Academy of Sciences of the United States of America* 36 (6):344-355
- McClintock B (1951) *Chromosome Organization and Genic Expression*. Cold Spring Harbor Symposia on Quantitative Biology 16:13-47
- McClintock B (1953) Induction of Instability at Selected Loci in Maize. *Genetics* 38 (6):579-599
- Meinkoth J, Wahl G (1984) Hybridization of Nucleic Acids Immobilized on Solid Supports. *Analytical Biochemistry* 138:267-284
- Meulenbroek B, Verhaegh J, Janse J (1998) Inheritance Studies with Columnar Type Trees. *Acta Hort* 484:255-260

- Miller JR, Koren S, Sutton G (2010) Assembly algorithms for next-generation sequencing data. *Genomics* 95 (6):315-327
- Minoche AE, Dohm JC, Himmelbauer H (2011) Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems. *Genome biology* 12 (11):R112
- Monaco AP, Larin Z (1994) YACs, BACs, PACs and MACs: Artificial chromosomes as research tools. *Trends in Biotechnology* 12 (7):280-286
- Morgan J, Richards A (1993) *The book of apples*. Ebury Press, London,
- Morgante M, Olivieri AM (1993) PCR-amplified microsatellites as markers in plant genetics. *The Plant Journal* 3 (1):175-182
- Moriya S, Iwanami H, Kotoda N, Takahashi S, Yamamoto T, Kazuyuki A (2009) Development of a Marker-assisted Selection System for Columnar Growth Habit in Apple Breeding. *J Japan Soc Hort Sci* 78 (3):279-287
- Moriya S, Okada K, Haji T, Yamamoto T, Abe K (2012) Fine mapping of Co, a gene controlling columnar growth habit located on apple (*Malus domestica* Borkh.) linkage group 10. *Plant Breeding*:no-no
- Muehlbauer G, Bhau B, Syed N, Heinen S, Cho S, Marshall D, Pateyron S, Buisine N, Chalhoub B, Flavell A (2006) A hAT superfamily transposase recruited by the cereal grass genome. *Molecular Genetics and Genomics* 275 (6):553-563
- Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H (1992) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. 1986. *Biotechnology* 24:17-27
- Myers EW (1995) Toward simplifying and accurately formulating fragment assembly. *Journal of computational biology : a journal of computational molecular cell biology* 2 (2):275-290
- Nagaya S, Yoshida K, Kato K, Akasaka K, Shinmyo A (2001) An insulator element from the sea urchin *Hemicentrotus pulcherrimus* suppresses variation in transgene expression in cultured tobacco cells. *Molecular Genetics and Genomics* 265 (3):405-413
- Naito K, Zhang F, Tsukiyama T, Saito H, Hancock CN, Richardson AO, Okumoto Y, Tanisaka T, Wessler SR (2009) Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. *Nature* 461 (7267):1130-1134
- Newcomb RD, Crowhurst RN, Gleave AP, Rikkerink EHA, Allan AC, Beuning LL, Bowen JH, Gera E, Jamieson KR, Janssen BJ *et al.* (2006) Analyses of Expressed Sequence Tags from Apple. *Plant Physiology* 141 (1):147-166
- Okada N (1991a) SINEs. *Current opinion in genetics & development* 1 (4):498-504
- Okada N (1991b) SINEs: Short interspersed repeated elements of the eukaryotic genome. *Trends in Ecology & Evolution* 6 (11):358-361
- Osoegawa K, Tateno M, Woon PY, Frengen E, Mammoser AG, Catanese JJ, Hayashizaki Y, de Jong PJ (2000) Bacterial artificial chromosome libraries for mouse sequencing and functional analysis. *Genome research* 10 (1):116-128
- Osoegawa K, Woon PY, Zhao B, Frengen E, Tateno M, Catanese JJ, de Jong PJ (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 52 (1):1-8
- Otto D (2008) *Analyse des Kolumnar-Gens*. Diplomarbeit, Johannes Gutenberg Universität, Mainz

- Otto D, Petersen R, Brauksiepe B, Braun P, Schmidt ER (2013) The columnar mutation ("Co gene") of apple (*Malus x domestica*) is associated with an integration of a Gypsy-like retrotransposon. *Molecular Breeding*:in press
- Paran I, Michelmore RW (1993) Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. *TAG Theoretical and Applied Genetics* 85 (8):985-993
- Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberler G, Hellsten U, Mitros T, Poliakov A *et al.* (2009) The Sorghum bicolor genome and the diversification of grasses. *Nature* 457 (7229):551-556
- Paterson AH, Freeling M, Tang H, Wang X (2010) Insights from the comparison of plant genome sequences. *Annual review of plant biology* 61:349-372
- Patocchi A, Vinatzer BA, Gianfranceschi L, Tartarini S, Zhang HB, Sansavini S, Gessler C (1999) Construction of a 550 kb BAC contig spanning the genomic region containing the apple scab resistance gene Vf. *Mol Gen Genet* 262 (4-5):884-891
- Petersen R Molekulargenetische Ursachen des Kolumnarwachstums bei Äpfeln. Dissertation, Johannes Gutenberg Universität, Mainz
- Petersen R, Krost C (2013) Tracing a key player in the regulation of plant architecture: the columnar growth habit of apple trees (*Malus x domestica*). *Planta*:1-22
- Pevzner PA, Tang H, Tesler G (2004) De Novo Repeat Classification and Fragment Assembly. *Genome Research* 14 (9):1786-1796
- Phipps JB, Robertson KR, Rohrer JR, Smith PG (1991) Origins and Evolution of Subfam. Maloideae (Rosaceae). *Systematic Botany* 16 (2):303-332
- Potter D, Eriksson T, Evans RC, Oh S, Smedmark JEE, Morgan DR, Kerr M, Robertson KR, Arsenault M, Dickinson TA *et al.* (2007) Phylogeny and classification of Rosaceae. *Plant Systematics and Evolution* 266 (1):5-43
- Pouteau S, Grandbastien M-A, Boccara M (1994) Microbial elicitors of plant defence responses activate transcription of a retrotransposon. *The Plant Journal* 5 (4):535-542
- Prochazka M, Walder K, Xia J (2001) AFLP fingerprinting of the human genome. *Hum Genet* 108 (1):59-65
- Proost S, Pattyn P, Gerats T, Van de Peer Y (2011) Journey through the past: 150 million years of plant genome evolution. *The Plant journal : for cell and molecular biology* 66 (1):58-65
- Proost S, Van Bel M, Sterck L, Billiau K, Van Parys T, Van de Peer Y, Vandepoele K (2009) PLAZA: a comparative genomics resource to study gene and genome evolution in plants. *The Plant cell* 21 (12):3718-3731
- Rickert AM, Kim JH, Meyer S, Nagel A, Ballvora A, Oefner PJ, Gebhardt C (2003) First-generation SNP/InDel markers tagging loci for pathogen resistance in the potato genome. *Plant Biotechnology Journal* 1 (6):399-410
- Robinson JP, Harris SA, Juniper BE (2001) Taxonomy of the genus *Malus* Mill. (Rosaceae) with emphasis on the cultivated apple, *Malus domestica* Borkh. *Plant Systematics and Evolution* 226 (1):35-58
- Sajer O, Scorza R, Dardick C, Zhebentyayeva T, Abbott AG, Horn R (2012) Development of sequence-tagged site markers linked to the pillar growth type in peach (*Prunus persica*). *Plant Breeding* 131 (1):186-192
- Salvi S, Sponza G, Morgante M, Tomes D, Niu X, Fengler KA, Meeley R, Ananiev EV, Svitashv S, Bruggemann E *et al.* (2007) Conserved noncoding genomic

- sequences associated with a flowering-time quantitative trait locus in maize. *Proceedings of the National Academy of Sciences* 104 (27):11376-11381
- Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular cloning*, vol 2. Cold spring harbor laboratory press New York,
- Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences* 74 (12):5463-5467
- SanMiguel P, Bennetzen JL (1998) Evidence that a Recent Increase in Maize Genome Size was Caused by the Massive Amplification of Intergene Retrotransposons. *Annals of Botany* 82 (suppl 1):37-44
- SanMiguel P, Gaut BS, Tikhonov A, Nakajima Y, Bennetzen JL (1998) The paleontology of intergene retrotransposons of maize. *Nat Genet* 20 (1):43-45
- SanMiguel P, Tikhonov A, Jin Y-K, Motchoulskaia N, Zakharov D, Melake-Berhan A, Springer PS, Edwards KJ, Lee M, Avramova Z *et al.* (1996) Nested Retrotransposons in the Intergenic Regions of the Maize Genome. *Science* 274 (5288):765-768
- Schatz MC, Delcher AL, Salzberg SL (2010) Assembly of large genomes using second-generation sequencing. *Genome Research* 20 (9):1165-1173
- Schmidt M (1937) Somatische Mutationen beim Kern- und Steinobst und ihre züchterische Bedeutung. *Der Züchter* 9 (4):81-91
- Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J *et al.* (2010) Genome sequence of the palaeopolyploid soybean. *Nature* 463 (7278):178-183
- Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA *et al.* (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science* 326 (5956):1112-1115
- Schuster M (2009) Sour cherries *Prunus Cerasus* L. with columnar tree habit. *Acta Hort* 814:325-328
- Schwartz DC, Cantor CR (1984) Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell* 37 (1):67-75
- Schweingruber FH, Hopf M (1979) *Wildäpfel und prähistorische Äpfel*. Rheinland-Verlag,
- Seyffert W (2003) *Lehrbuch der Genetik*, vol 2. Spektrum,
- Sheng Y, Mancino V, Birren B (1995) Transformation of *Escherichia coli* with large DNA molecules by electroporation. *Nucleic acids research* 23 (11):1990-1996
- Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiri Y, Simon M (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proceedings of the National Academy of Sciences of the United States of America* 89 (18):8794-8797
- Silfverberg-Dilworth E, Matasci C, Van de Weg W, Van Kaauwen M, Walser M, Kodde L, Soglio V, Gianfranceschi L, Durel C, Costa F *et al.* (2006) Microsatellite markers spanning the apple (*Malus x domestica* Borkh.) genome. *Tree Genetics & Genomes* 2 (4):202-224
- Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Birol I (2009) ABySS: A parallel assembler for short read sequence data. *Genome Research* 19 (6):1117-1123
- Souer E, van Houwelingen A, Kloos D, Mol J, Koes R (1996) The No Apical Meristem Gene of *Petunia* Is Required for Pattern Formation in Embryos and

- Flowers and Is Expressed at Meristem and Primordia Boundaries. *Cell* 85 (2):159-170
- Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* 98 (3):503-517
- Sterck L, Rombauts S, Vandepoele K, Rouzé P, Van de Peer Y (2007) How many genes are there in plants (... and why are they there)? *Current Opinion in Plant Biology* 10 (2):199-203
- Studer A, Zhao Q, Ross-Ibarra J, Doebley J (2011) Identification of a functional transposon insertion in the maize domestication gene *tb1*. *Nat Genet* 43 (11):1160-1163
- Surget-Groba Y, Montoya-Burgos JI (2010) Optimization of de novo transcriptome assembly from next-generation sequencing data. *Genome Research* 20 (10):1432-1440
- TAGI (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408 (6814):796-815
- Takeda S, Sugimoto K, Otsuki H, Hirochika H (1999) A 13-bp cis-regulatory element in the LTR promoter of the tobacco retrotransposon *Tto1* is involved in responsiveness to tissue culture, wounding, methyl jasmonate and fungal elicitors. *The Plant journal : for cell and molecular biology* 18 (4):383-393
- Tanksley SD, Ganai MW, Martin GB (1995) Chromosome landing: a paradigm for map-based gene cloning in plants with large genomes. *Trends in Genetics* 11 (2):63-68
- Tanksley SD, Young ND, Paterson AH, Bonierbale MW (1989) RFLP Mapping in Plant Breeding: New Tools for an Old Science. *Nat Biotech* 7 (3):257-264
- Tartarini S, Gianfranceschi L, Sansavini S, Gessler C (1999) Development of reliable PCR markers for the selection of the *Vf* gene conferring scab resistance in apple. *Plant Breeding* 118 (2):183-186
- Tian Y-K, Wang C-H, Zhang J-S, James C, Dai H-Y (2005) Mapping *Co*, a gene controlling the columnar phenotype of apple, with molecular markers. *Euphytica* 145 (1):181-188
- Tian YK, Wang CH, Dai HY, Zhang JS (2004) Screening of a RAPD marker tightly linked to *Co* gene in apple and the SCAR marker conversion. *Yi chuan xue bao = Acta genetica Sinica* 31 (9):919-925
- Tobutt K (1985) Breeding columnar apples at East Malling. *Acta Horticulturae* 159:63-68
- Troggio M, Gleave A, Salvi S, Chagné D, Cestaro A, Kumar S, Crowhurst RN, Gardiner SE (2012) Apple, from genome to breeding. *Tree Genetics & Genomes* 8 (3):509-529
- Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A *et al.* (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313 (5793):1596-1604
- Umehara Y, Inagaki A, Tanoue H, Yasukochi Y, Nagamura Y, Saji S, Otsuki Y, Fujimura T, Kurata N, Minobe Y (1995) Construction and characterization of a rice YAC library for physical mapping. *Mol Breeding* 1 (1):79-89
- Vali U, Brandstrom M, Johansson M, Ellegren H (2008) Insertion-deletion polymorphisms (indels) as genetic markers in natural populations. *BMC Genetics* 9 (1):8

- Van Damme M, Andel A, Huibers RP, Panstruga R, Weisbeek PJ, Van den Ackerveken G (2005) Identification of Arabidopsis Loci Required for Susceptibility to the Downy Mildew Pathogen *Hyaloperonospora parasitica*. *Molecular Plant-Microbe Interactions* 18 (6):583-592
- Van Damme M, Huibers RP, Elberse J, Van den Ackerveken G (2008) Arabidopsis DMR6 encodes a putative 2OG-Fe(II) oxygenase that is defense-associated but required for susceptibility to downy mildew. *The Plant Journal* 54 (5):785-793
- Vavilov NI (1926) Studies on the origin of cultivated plants. *Trudy Byuro Prikl Bot* 16:139-245
- Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, Fontana P, Bhatnagar SK, Troggio M, Pruss D *et al.* (2010) The genome of the domesticated apple (*Malus x domestica* Borkh.). *Nat Genet* 42 (10):833-839
- Venturi S, Dondini L, Donini P, Sansavini S (2006) Retrotransposon characterisation and fingerprinting of apple clones by S-SAP markers. *Theor Appl Genet* 112 (3):440-444
- Vinatzer BA, Zhang HB, Sansavini S (1998) Construction and characterization of a bacterial artificial chromosome library of apple. *TAG Theoretical and Applied Genetics* 97 (7):1183-1190
- Vitte C, Panaud O (2005) LTR retrotransposons and flowering plant genome size: emergence of the increase/decrease model. *Cytogenetic and genome research* 110 (1-4):91-107
- Volik S, Zhao S, Chin K, Brebner JH, Herndon DR, Tao Q, Kowbel D, Huang G, Lapuk A, Kuo W-L *et al.* (2003) End-sequence profiling: Sequence-based analysis of aberrant genomes. *Proceedings of the National Academy of Sciences* 100 (13):7696-7701
- Vos P, Hogers R, Bleeker M, Reijans M, Lee Tvd, Hornes M, Friters A, Pot J, Paleman J, Kuiper M *et al.* (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic acids research* 23 (21):4407-4414
- Watanabe M, Bessho H, Suzuki A, Komori S (2008) Seasonal Changes of IAA and Cytokinin in Shoots of Columnar Type Apple Trees. *Acta Hort* 774:75-80
- Watanabe M, Suzuki A, Komori S, Bessho H (2004) Comparison of Endogenous IAA and Cytokinins in Shoots of Columnar and Normal Type Apple Trees. *J Japan Soc Hort Sci* 73 (1):19-24
- Weber JL, May PE (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *American journal of human genetics* 44 (3):388-396
- Welsh J, McClelland M (1990) Fingerprinting genomes using PCR with arbitrary primers. *Nucleic acids research* 18 (24):7213-7218
- Wessler SR (1996) Plant retrotransposons: Turned on by stress. *Current Biology* 6 (8):959-961
- Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O *et al.* (2007) A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* 8 (12):973-982
- Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic acids research* 18 (22):6531-6535

- Wing RA, Zhang H-B, Tanksley SD (1994) Map-based cloning in crop plants. Tomato as a model system: I. Genetic and physical mapping of jointless. *Molecular and General Genetics MGG* 242 (6):681-688
- Wolters PJ, Schouten HJ, Velasco R, Si-Ammour A, Baldi P (2013) Evidence for regulation of columnar habit in apple by a putative 2OG-Fe(II) oxygenase. *New Phytologist* 200 (4):993-999
- Woo S-S, Jiang J, Gill BS, Paterson AH, Wing RA (1994) Construction and characterization of bacterial artificial chromosome library of *Sorghum bicolor*. *Nucleic acids research* 22 (23):4922-4931
- Wu K-S, Tanksley SD (1993) Abundance, polymorphism and genetic mapping of microsatellites in rice. *Molecular and General Genetics MGG* 241 (1):225-235
- Xiong Y, Eickbush TH (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences. *The EMBO journal* 9 (10):3353-3362
- Xu M, Li X, Korban S (2000) AFLP-Based detection of DNA methylation. *Plant Mol Biol Rep* 18 (4):361-368
- Xu M, Song J, Cheng Z, Jiang J, Korban SS (2001) A bacterial artificial chromosome (BAC) library of *Malus floribunda* 821 and contig construction for positional cloning of the apple scab resistance gene Vf. *Genome / National Research Council Canada = Genome / Conseil national de recherches Canada* 44 (6):1104-1113
- Xu Z, Wang H (2007) LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Research* 35 (suppl 2):W265-W268
- Yamazaki K, Okabe M, Takahashi E (1987) New broomy flowering peach cultivars. *Bulletin of the Kanagawa Horticultural Experiment Station*
- Yao J-L, Dong Y-H, Morris BAM (2001) Parthenocarpic apple fruit production conferred by transposon insertion mutations in a MADS-box transcription factor. *Proceedings of the National Academy of Sciences* 98 (3):1306-1311
- Yu C, Zhang J, Peterson T (2011) Genome Rearrangements in Maize Induced by Alternative Transposition of Reversed Ac/Ds Termini. *Genetics* 188 (1):59-67
- Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X *et al.* (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). *Science* 296 (5565):79-92
- Yu Z-X, Li J-X, Yang C-Q, Hu W-L, Wang L-J, Chen X-Y (2012) The Jasmonate-Responsive AP2/ERF Transcription Factors AaERF1 and AaERF2 Positively Regulate Artemisinin Biosynthesis in *Artemisia annua* L. *Molecular Plant* 5 (2):353-365
- Zhang H-B, Zhao X, Ding X, Paterson AH, Wing RA (1995) Preparation of megabase-size DNA from plant nuclei. *The Plant Journal* 7 (1):175-184
- Zohary D, Hopf M (2000) Domestication of plants in the Old World. 3rd edition edn. Oxford University Press,

Anhang

A1 Vektorkarten

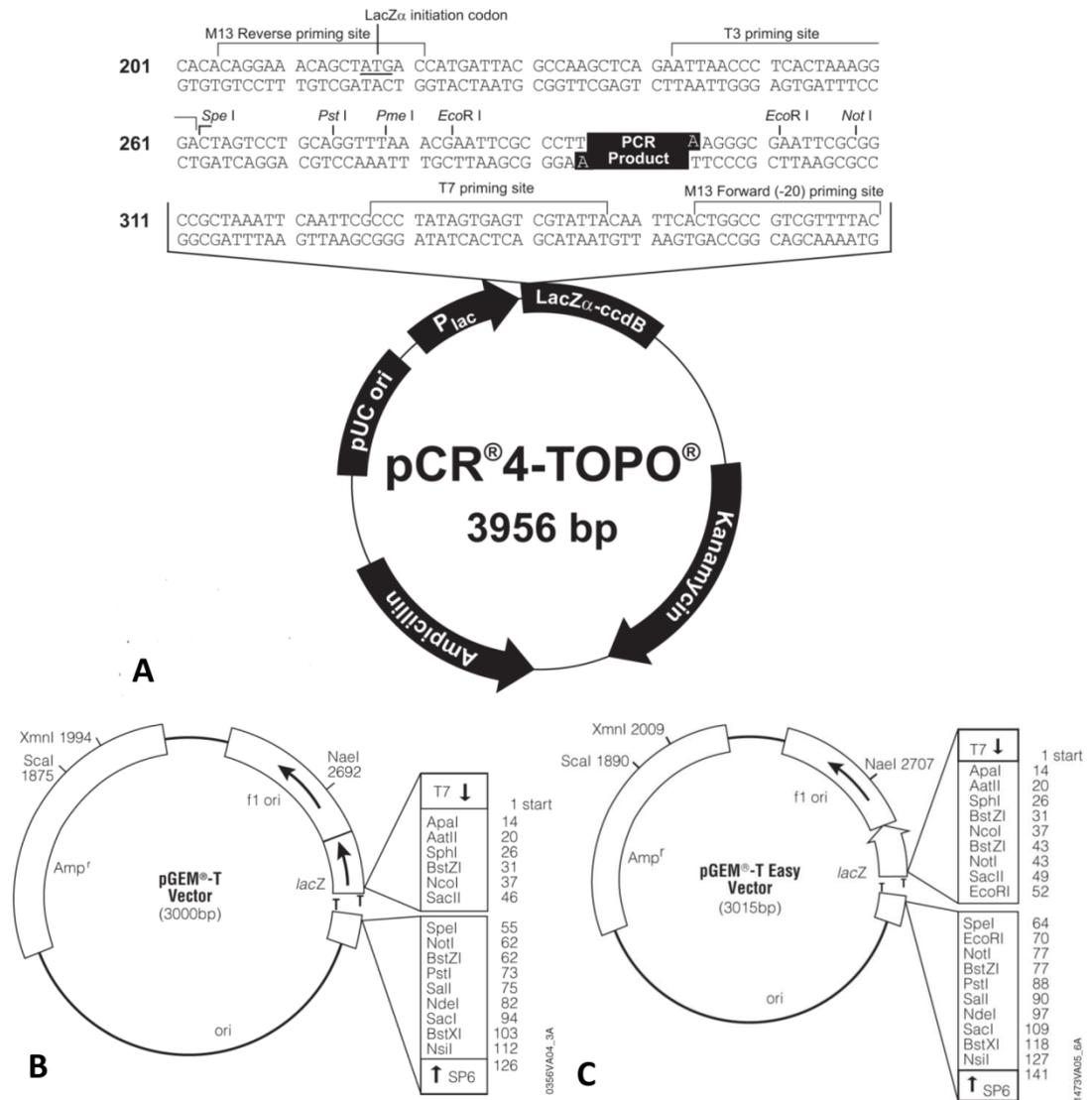


Abbildung A1.1 Vektorkarten Teil I

Die Grafik zeigt die Vektoren, die für die Klonierung von PCR-Produkten verwendet wurden. In A) ist die Vektorkarte des pCR[®]4-TOPO[®] Vektors der Fa. Invitrogen[™] (Carlsbad, USA) abgebildet. B) und C) zeigen Vektorkarten des pGEM[®]-T bzw. des pGEM[®]-T Easy Vektors der Fa. Promega (Madison, USA).

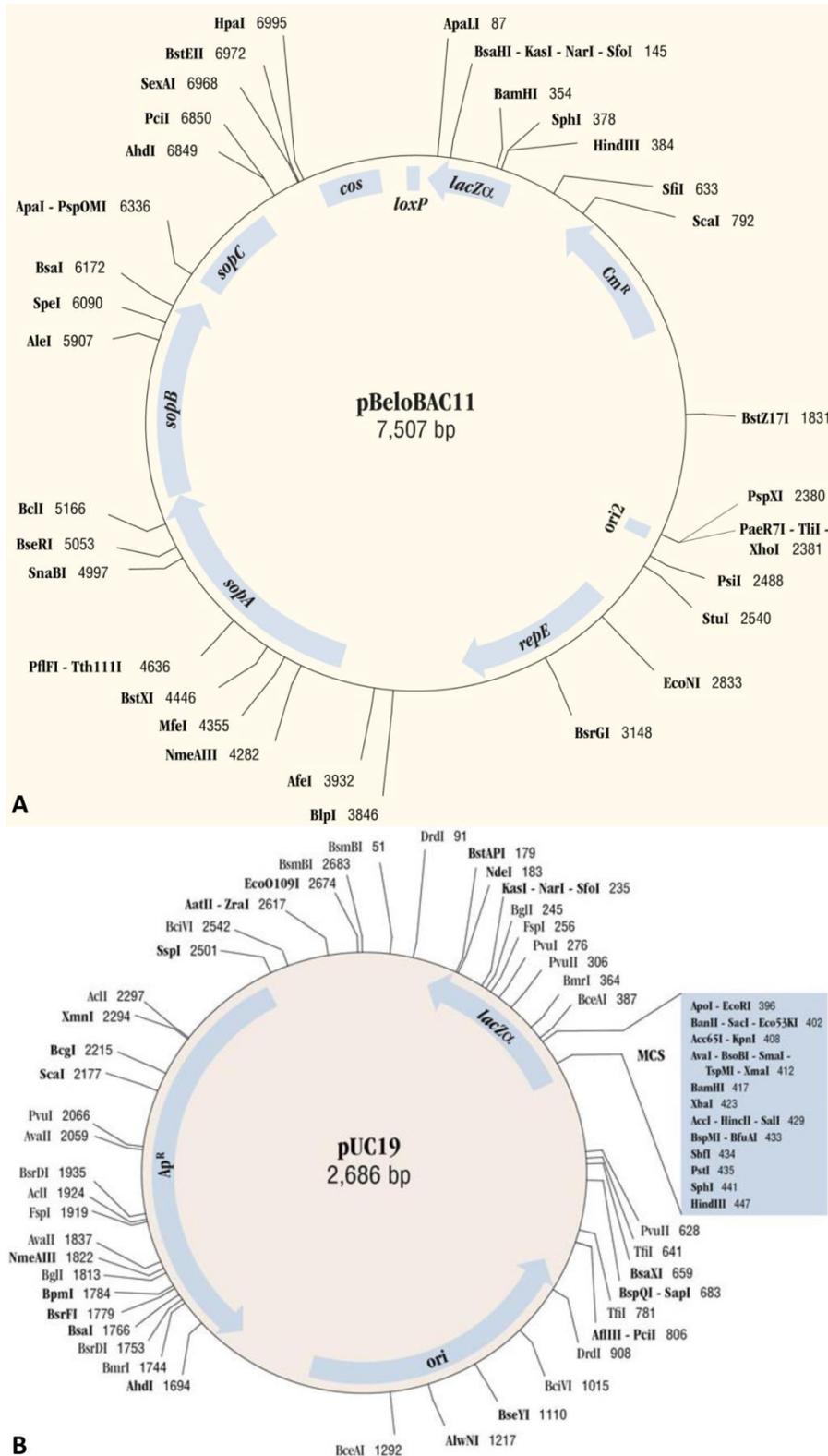


Abbildung A1.2 Vektorkarten Teil II

Die Grafik zeigt die Vektorkarten der Vektoren pBelobAC11 (A) und pUC19 (B) der Fa. NEB (Ipswich, USA). Der Vektor pBelobAC11 wurde für die Konstruktion von BAC-DNA-Bibliotheken, der Vektor pUC19 bei der Herstellung der shotgun-Bibliotheken verwendet.

A2 Sonden für das Screening von genomischen Apfel-BAC-Bibliotheken

Die Sonden für das Screening der genomischen Apfel-BAC-Banken wurden in vier aufeinanderfolgenden Runden generiert. Für die ersten beiden Runden standen Transkriptomdatensätze der Apfelsorten P28 und A14 (Krost 2012) zur Verfügung. Die dritte Runde basiert auf Contigs des veröffentlichten GD-Genoms (Velasco *et al.* 2010). Abschließend wurden Sonden auf selbst generierte und sequenzierte „long range“ PCR-Produkte erstellt. In den folgenden vier Tabellen sind die jeweiligen Sonden aufgelistet.

Tabelle A2.1 Sonden in der Zielregion für die KFH Runde I (Sondenpool I)

Sonde	Contig Transkriptomdaten	Apfelsorte	Region		Sondengröße (bp)
			von	bis	
1	GJVGIEO02HFQT7	P28	18102354	18102656	303
2	GDSYFPM02CZF0X	A14	18113754	18114115	362
3	GDSYFPM02DA9IW	A14	18140560	18140942	383
4	GDSYFPM02DCQ1P	A14	18202028	18202466	439
5	GDSYFPM02D6PQF	A14	18294439	18294687	249
6	GJVGIEO02HUZG7	P28	18294750	18295075	326
7	Contig1921	A14	18308944	18309255	312
8	contig5251	P28	18420127	18420535	409
9	Contig19753	P28	18497033	18497451	419
10	Contig4278	A14	18514614	18515126	513
11	Contig10083	P28	18525098	18525503	406
12	GDSYFPM02D552Q	A14	18536006	18536399	394
13	Contig9945	P28	18545869	18546134	266
14	GDSYFPM02DFS67	A14	18611803	18612115	313

Tabelle A2.2 Sonden in der Zielregion für die KFH Runde II (Sondenpool II)

Sonde	GD Contig	Region		Sondengröße (bp)
		von	bis	
15	MDC041557.8	18065865	18066190	326
16	MDC017675.49	18251945	18252272	328
17	MDC004247.204	18464761	18465155	395
18	MDC010450.949	18779415	18779917	503
19	MDC007787.391	18815994	18816349	356
20	MDC008849.307	18843567	18843963	397
21	MDC008849.307	18828005	18828161	157
22	MDC004959.260	18863056	18863193	138
23	MDC001753.378	18906442	18906597	156
24	MDC027247.53	18946350	18946533	184

Tabelle A2.3 Sonden in der Zielregion für die KFH Runde III (Sondenpool III)

Sonde	GD Contig	Region		Sondengröße (bp)
		von	bis	
25	MDC018767.346	18165396	18165857	462
26	MDC020053.385	18197605	18198043	439
27	MDC020855.93	18284923	18285549	627
28	MDC018604.401	18376876	18377281	406
29	MDC018604.401	18366513	18366956	444
30	MDC010853.171	18382704	18383131	428
31	MDC002260.360	18621825	18622143	319
32	MDC038765.9	18661361	18661852	492
33	MDC004959.260	18861541	18861935	395
34	MDC004959.251	18892713	18893056	344
35	MDC001753.378	18941089	18941369	281

Tabelle A2.4 Sonden in der Zielregion für die KFH Runde IV (Sondenpool IV)

Sonde	„long range“ PCR- Produkt	Sondengröße (bp)
36	LR R	434
37	LR T	385
38	LR T	379
39	LR V	340
40	LR X	1212
41	LR Z	1153
42	LR AE	314

A3 Markeranalysen mittels QIAxcel

Die Tabelle A3.1 zeigt die Belegung der 96-Well Platte mit genomischen DNAs der Einzelnachkommen einer Kreuzung zwischen P28 und A14.

Tabelle A3.1 Plattenbelegung PCR und QIAxcel

	1	2	3	4	5	6
A	6.5 C	6.8 C	6.10 C	6.14 C	6.19 C	6.24 C
B	7.10 C	7.18 C	8.59 C	8.63 C	8.68 C	8.74 C
C	7.50 C	7.60 C	7.71 C	7.77 C	7.85 C	7.89 C
D	8.6 C	8.7 C	8.16 C	8.17 C	8.31 C	8.36 C
E	6.7 N	6.18 N	6.20 N	6.23 N	6.25 N	6.29 N
F	6.52 N	6.54 N	6.60 N	6.62 N	8.4 N	8.9 N
G	6.91 N	6.95 N	6.98 N	6.101 N	6.102 N	6.103 N
H	8.18 N	8.19 N	8.20 N	8.21 N	7.17 N	7.19 N

	7	8	9	10	11	12
A	6.32 C	6.41 C	6.43 C	6.72 C	6.76 C	7.1 C
B	7.21 C	7.25 C	7.26 C	7.28 C	7.42 C	7.43 C
C	7.91 C	7.95 C	7.99 C	7.102 C	8.76 C	8.5 C
D	8.44 C	8.45 C	8.46 C	8.55 C	8.58 C	8.2 C
E	6.31 N	6.34 N	6.37 N	6.39 N	6.40 N	6.44 N
F	8.14 N	8.15 N	6.73 N	6.74 N	6.88 N	6.90 N
G	6.105 N	6.106 N	7.2 N	7.3 N	7.6 N	8.8 N
H	7.31 N	7.37 N	7.87 N	7.93 N	7.98 N	NK

C: kolumnarer Phänotyp; N: normalwüchsiger Phänotyp; NK: Negativkontrolle

In Abbildung A3.1 ist exemplarisch die Grafik eines QIAxcel-Laufes gezeigt. Die vollständigen Daten befinden sich im elektronischen Anhang.



Abbildung A3.1 QIAxcel-Lauf

Die Abbildung zeigt einen Ausschnitt der Auftrennung der PCR-Produkte des Markers Ch03d11 mittels QIAxcel. Während sich in Spur H12 die Negativkontrolle befindet, handelt es sich in allen anderen Spuren um PCRs auf genomische DNAs von Nachkommen mit normalwüchsigem Phänotyp. Der Nachkomme in Spur H4 zeigt im Gegensatz zu den anderen Nachkommen den kolumnaren Genotyp (rot eingerahmt). Es handelt sich somit um einen „Ausreißer“. In der Spur H4 befindet sich der Nachkomme 8.20.

A4 Markertests diverser Apfelsorten mittels selbst erstellter Marker

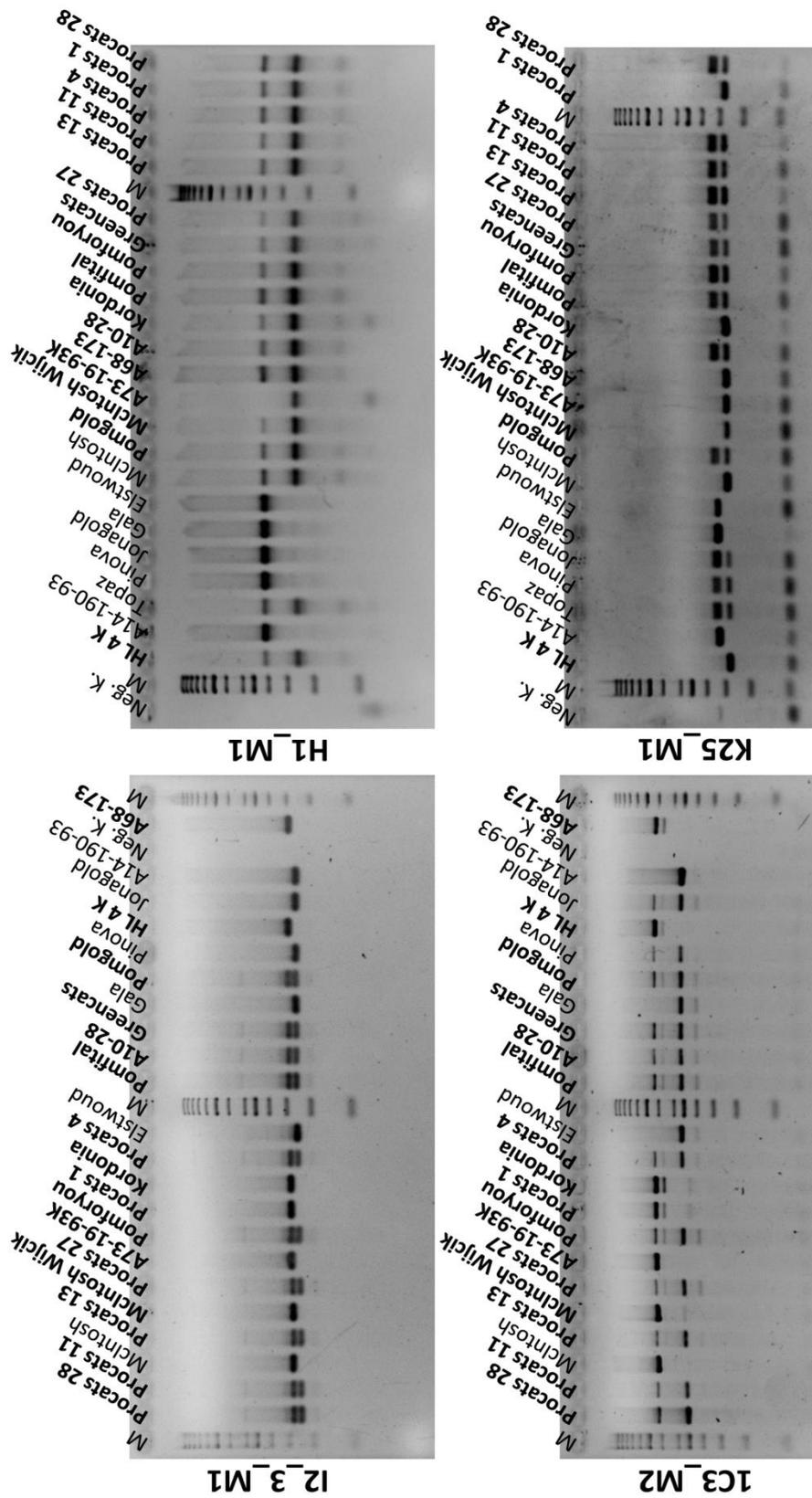


Abbildung A4.1 Markertests diverser Apfelsorten

Die Grafik zeigt eine Zusammenstellung der gelelektrophoretisch aufgetrennten PCR-Produkte der selbst erstellten molekularen Marker 12_3_M1, K25_M1, 1C3_M2 und H1_M1 nach erfolgtem S1-Nuclease-Verdau. Als Template dienten genomische DNAs diverser Apfelsorten, wobei die Namen kolmunarer Apfelsorten fett dargestellt sind. M: 50 bp Ladder

A5 Inhaltsverzeichnis des elektronischen Anhangs

Folgende Dateien befinden sich im beigefügten elektronischen Anhang:

- Verwendeten Perl-Skripte:
 - fasta_check_iupac.pl
 - fasta_extract.pl
 - fasta_iupac_permut.pl
 - fasta_kmer.pl
 - fasta_n_stats.pl
 - fastq_integrity.pl
 - fastq_splitup.pl
 - qseq2fastq.pl
 - seq_filter.pl
 - seq_info.pl

- Ergebnisse des QIAxcel_Laufes der Markertests

- Primerliste

- Liste aller Annotationen im Gesamtcontig

- Gesamtcontig als gbk-Datei

- Liste der Gene im Gesamtcontig

- Liste der Transposons im Gesamtcontig (inklusive CENSOR-Daten)

- Vorliegende Doktorarbeit im PDF-Format

Curriculum Vitae

Danksagung

Eidesstattliche Erklärung

Hiermit erkläre ich, die vorliegende Dissertation selbstständig und ohne fremde Hilfe angefertigt zu haben. Weiterhin versichere ich, örtlich übernommene Ausführungen anderer Autoren und an Gedankengänge Anderer anlehrende eigene Formulierungen entsprechend gekennzeichnet und die Quellen zitiert zu haben.

(Ort, Datum)

(Dominik Otto)