

**Studien zur Genomkolinearität
in der Familie der Brassicaceae:
Arabidopsis thaliana und *Capsella rubella***

I n a u g u r a l - D i s s e r t a t i o n

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Adile Acarkan

aus Duisburg

Köln
2001

Berichtersteller:

Prof. Dr. J. Schell
Prof. Dr. U.-I. Flügge

Tag der mündlichen Prüfung:

15.06.2000

Abstract

Comparative genetic and physical mapping experiments were performed for two diploid species belonging to the family of the Brassicaceae, *Arabidopsis thaliana* and *Capsella rubella*.

Markers located on *Arabidopsis thaliana* chromosome IV were used for genetic mapping experiments in *Capsella* and two linkage groups could be established. In respect to marker repertoire and colinearity, a high degree of conservation could be demonstrated for the two species. A 200 kb region located on chromosome IV and a 30 kb region located on chromosome I of *Arabidopsis thaliana* were chosen for microsynteny studies. *Arabidopsis* and *Capsella* showed conservation of gene repertoire, order and orientation. Intergenic distances were also of similar size in both species. Some genes were present in different copy numbers in the analysed regions of the *Arabidopsis* and *Capsella* genomes. Comparative sequence analysis revealed a high degree of conservation for the exon-intron structure of orthologous genes in the two crucifer species. Average sequence identity values of 90% were found for exon sequences at nucleotide and amino acid level.

In summary, on genetic as well as molecular levels, an almost complete colinearity of genomes could be demonstrated for the two closely related species, *Arabidopsis thaliana* and *Capsella rubella*. The results presented here show that it is feasible to transfer information obtained in the *Arabidopsis* genome project to other closely related diploid species.

Inhaltsverzeichnis

1. Einleitung	1
1.1 Variabilität der Genomgrößen im Pflanzenreich	1
1.2 Klassen repetitiver Elemente in Pflanzen	2
1.3 Karten pflanzlicher Genome	5
1.4 Untersuchungen pflanzlicher Genome mittels Sequenzierung	9
1.5 Vergleichende Kartierungsexperimente	10
1.6 Ein Modellsystem für vergleichende Genomanalysen	14
1.7 Zielsetzung	16
1.8 Abkürzungen	18
2. Material und Methoden	19
2.1 Material	19
2.1.1. Chemikalien und Enzyme	19
2.1.2 DNA-Längenstandards	20
2.1.3 Lösungen und Medien	20
2.1.3.1 Medien	20
2.1.3.2 Lösungen	21
2.2 Verwendete Bakterienstämme und Vektoren	21
2.2.1 Bakterienstämme	21
2.2.2 Vektoren	21
2.3 Oligonukleotide	21
2.4 Verwendete Marker für genetische Kartierungsexperimente	22
2.5 Molekularbiologische Techniken	22
2.5.1 Isolierung von Pflanzen-Gesamt-DNA	22
2.5.2 Southern-Blot und Hybridisierungsexperimente	23
2.5.3 Genomische <i>Capsella rubella</i> Cosmid-Bibliothek	23
2.5.4 Herstellung der <i>Capsella rubella</i> Cosmid-Koloniemembranen	24
2.5.5 Koloniehybridisierung der <i>Capsella rubella</i> Cosmid-Membranen	24
2.6 Sequenzierung und Sequenz-Analyse	25
2.6.1 Sequenzierung	25
2.6.2 Sequenz-Analyse	25

2.7 Genetische Analyse	25
2.7.1 Erstellung der Kartierungspopulation	25
2.7.2 Polymorphismus-Analyse	26
3. Ergebnisse	27
3.1 Genetische Kartierung von <i>Arabidopsis thaliana</i> -Markern in <i>Capsella</i>	27
3.1.1 Kartierungspopulation von <i>Capsella</i>	27
3.1.2 Restriktions-Fragment-Längen-Polymorphismus (RFLP)-Analyse von <i>C. grandiflora</i> und <i>C. rubella</i>	27
3.1.3 Genetische Marker für die Kartierung	28
3.1.4 RFLP-Analyse der <i>Arabidopsis thaliana</i> Chromosom IV-Marker	32
3.1.5 Kartierung der Marker in <i>Capsella</i>	35
3.2 Vergleichende Mikrosyntenieanalyse in <i>Arabidopsis thaliana</i> und <i>Capsella rubella</i>	41
3.2.1 Identifizierung kodierender Sequenzen in der 200 Kbp-Region von <i>Arabidopsis thaliana</i>	41
3.2.1.1 Identifizierung der homologen Cosmid-Klone im <i>Capsella</i> <i>rubella</i> -Genom	43
3.2.1.2 Sequenzierung der beiden <i>Capsella rubella</i> Cosmid-„Contigs“	50
3.2.1.3 Analyse der Sequenzierungsdaten	53
3.2.1.3.1 Analyse der Region zwischen den Contigs 0 und I	53
3.2.1.3.2 cDNAs K und 8 in der genomischen <i>C. rubella</i> -Sequenz	60
3.2.1.4 Vergleich der genomischen Sequenzen von <i>A. thaliana</i> und <i>C. rubella</i>	60
3.2.2 Vergleichende Mikrosyntenieanalyse einer 30 Kbp-Region in <i>A. thaliana</i> und <i>C. rubella</i>	76
3.2.2.1 Identifizierung kodierender Sequenzen	76
3.2.2.2 Identifizierung der homologen Region in <i>Capsella rubella</i>	77
3.2.2.3 Subklonierung und Sequenzierung der Cosmide 10 und 20	77
3.2.2.4 Sequenzanalyse von Cosmid 10 und 20 aus <i>Capsella rubella</i>	80
3.2.2.5 Vergleichende Sequenzanalyse in <i>A. thaliana</i> und <i>C. rubella</i>	80
4. Diskussion	87
4.1 Vergleichende genetische Kartierungsexperimente	87
4.1.1 Konservierung des Markerrepertoires in <i>Arabidopsis</i> und <i>Capsella</i>	87
4.1.2 Kolineare Chromosomensegmente in <i>Arabidopsis</i> und <i>Capsella</i>	89

4.2 Mikrosyntenieuntersuchungen in <i>Arabidopsis</i> und <i>Capsella</i>	93
4.2.1 Konservierung des Markerrepertoires und der Markeranordnung in <i>Arabidopsis</i> und <i>Capsella</i>	94
4.2.2 Vergleichende Sequenzuntersuchungen	99
5. Zusammenfassung	103
6. Literaturverzeichnis	104
Anhang	
I Daten der genetischen Kartierungsexperimente in <i>Capsella</i>	I
II Daten der Kopplungsanalyse in <i>Capsella</i>	II
III Sequenzen	III
IV Sequenzierte Subklone und PCR-Produkte	XLIII

1. Einleitung

Die Genome höherer Pflanzen-Arten zeigen erhebliche Unterschiede in ihrer Größe. Aufgrund sehr kleiner Genomgrößen sind *Arabidopsis thaliana* und Reis zu Modellsystemen der pflanzlichen Genomforschung geworden. Das *A. thaliana* Genom stellt zur Zeit das best-untersuchte Pflanzengenom dar (Meinke *et al.*, 1998). Ähnlich detaillierte Genomuntersuchungen wie sie für *A. thaliana* und Reis durchgeführt werden, sind mit der heutigen Technologie nicht für viele andere Pflanzenarten möglich. Daher muß es ein zentrales Ziel der pflanzlichen Genomanalyse sein, die Übertragbarkeit der Daten, die im Rahmen der *A. thaliana* und Reisgenomprojekte gewonnen werden, auf verwandte Arten zu prüfen. Zur Klärung dieser Fragestellung sind vergleichende Genomuntersuchungen unverzichtbar.

1.1 Variabilität der Genomgrößen im Pflanzenreich

Die DNA-Menge im haploiden Genom höherer Pflanzen beträgt zwischen 0,2 pg und >100 pg und unterscheidet sich somit um mehr als das 500-fache (Bennett und Smith, 1976). Während bei den Dikotyledonen *Arabidopsis thaliana* mit ca. 145 Mbp (Arumuganathan und Earle, 1991) und Reis (*Oryza sativa*) (Arumuganathan und Earle, 1991) mit 415-463 Mbp als Vertreter der Monokotyledonen zu den Arten mit einem kleinen Genom gehören, stellen beispielsweise Weizen (*Triticum aestivum*; ca. 16000 Mbp) (Arumuganathan und Earle, 1991) und *Fritillaria assyriaca* (ca. 127400 Mbp) (Bennett und Smith, 1976) Arten mit größeren Genomen dar. Eine große Varianz bezüglich der Genomgröße kann auch innerhalb einer Familie beobachtet werden. In der Familie der Brassicaceae wird zum Beispiel eine Genomgröße von 145 Mbp für *Arabidopsis thaliana* beobachtet, während die von *Brassica napus* mit 1129-1235 Mbp angegeben wird (Arumuganathan und Earle, 1991).

Eine direkte Beziehung zwischen Genomgröße und Komplexität des Organismus wird nicht beobachtet. Unterschiede können nur bedingt auf Polyploidisierung der Genome zurückgeführt werden. In der Familie der Brassicaceae liegen beispielsweise für *Brassica rapa*, *B. oleracea* und *B. nigra* Triplikationen umfangreicher Genombereiche

vor (Lagercrantz und Lydiat, 1996). Auch in der Familie der Poaceae ist für Mais eine Genomvergrößerung durch Duplikationen zu beobachten (Helentjaris *et al.*, 1988).

Analysen zur DNA-Reassoziationskinetik zeigten, daß Genome aus hoch- und mittel-repetitiven Elementen sowie aus Einzelkopie-Sequenzen aufgebaut sind. Der Unterschied in Genomgrößen ist dabei hauptsächlich auf den Anteil repetitiver Sequenzen im Genom zurückzuführen (Flavell *et al.*, 1974). Der Anteil der repetitiven Elemente im Genom variiert sehr stark in den unterschiedlichen Pflanzenarten. Dabei reichen die Anteile im Genom von 10% in *Arabidopsis* (Leutwiler *et al.*, 1984) bis zu 70-75% in Weizen (Flavell und Smith, 1976), Roggen (Smith und Flavell, 1977) und Erbse (Murray *et al.*, 1978). Auch die Verteilung von Gensequenzen kann zwischen relativ eng verwandten Arten sehr verschieden sein. Während im Genom von Erbse Einzelkopiesequenzen in der Regel nicht länger als 1 Kbp sind (Murray *et al.*, 1978), sind in der Mungobohne, die derselben Familie angehört, etwa 50% der Einzelkopiesequenzen in Blöcken größer als 6 Kbp organisiert (Murray *et al.*, 1979). Somit kann ein direkter Zusammenhang zwischen Genomgröße, repetitiven Elementen und Sequenzverteilung im Genom festgestellt werden. Größere Genome haben oft einen größeren Anteil an repetitiven Sequenzen, und Gensequenzen sind als kleinere Blöcke in die repetitiven Bereiche eingefügt (Flavell, 1980). Es handelt sich bei repetitiven Bereichen vielfach um Retrotransposon-ähnliche Elemente (SanMiguel *et al.*, 1996).

Für die Komplexität der Pflanzen scheint dieser repetitive Anteil im Genom nicht von großer Bedeutung zu sein, da Pflanzen mit kleinen wie großen Genomen ähnlich komplexe StoffwechsellLeistungen und morphologische Entwicklungen aufweisen können.

1.2 Klassen repetitiver Elemente in Pflanzen

Repetitive Elemente setzen sich aus einer Vielzahl von Familien zusammen. Der repetitive Anteil des Genoms wird in zwei Gruppen eingeteilt, den in Tandem organisierten und im Genom verteilt vorkommenden repetitiven Elementen.

Zu tandemartig organisierten Elementen, die kodierende Eigenschaften haben, gehören rDNA-Sequenzen. Die 18S-25S rDNA-Sequenzen stellen die Nukleolus-organisierenden Regionen (NOR) des Genoms dar. In *Arabidopsis thaliana* sind die

einzelnen 18S-25S rDNA-Elemente ca. 10 Kbp lang und kommen im Genom in 570 bis 750 Kopien vor (Pruitt und Meyerowitz, 1986; Copenhaver und Pikaard, 1996).

Satelliten-DNA, die aus kurzen in Tandem organisierten Sequenzen aufgebaut ist, wurde erstmals in CsCl-Gradient-Analysen beobachtet. Sie repräsentiert den hochrepetitiven Anteil der Elemente im Genom. Monomere der Satelliten-DNA setzen sich aus 150 bis 350 Bp langen Sequenzabschnitten zusammen und stellen in der Regel das konstitutive Heterochromatin dar (Gerlach und Peacock, 1980), das in Centromer- und Telomer-Regionen vorliegt (Bedbrook *et al.*, 1980; Maluszynska und Heslop-Harrison, 1991). Aber auch in anderen Regionen der Chromosomen wurde Satelliten-DNA lokalisiert (Ganal *et al.*, 1988; Bedbrook *et al.*, 1980). Während einige Mitglieder der Satelliten-DNA, die in Centromernähe lokalisiert sind, Konservierung bezüglich ihrer Sequenz zwischen verschiedenen Arten (Xia *et al.*, 1993) zeigen, ist die Sequenz der repetitiven Elemente in den restlichen Chromosomenregionen sogar in sehr eng verwandten Arten nicht konserviert (Schweizer *et al.*, 1993).

In *A. thaliana* macht Satelliten-DNA ca. 1-2% des Genoms aus. Für *A. thaliana* sind verschiedene Elemente mit einer Monomerlänge von 160 Bp, 180 Bp und 500 Bp bekannt (Martinez-Zapater *et al.*, 1986; Simoens *et al.*, 1988). Das 180 Bp-Element ist in den Centromer-Regionen von *A. thaliana* in repetitiven Blöcken von ca. 1000 Kbp vertreten (Round *et al.*, 1997; Maluszynska und Heslop-Harrison, 1991). Das 500 Bp-Element ist ebenfalls in den Centromerregionen lokalisiert, während die Vertreter der 160 Bp-Familie im Gegensatz dazu verteilt im Genom vorkommen (Bauwens *et al.*, 1991).

Weitere tandemartig organisierte repetitive Elemente im Pflanzen-Genom kommen in Telomerregionen vor. Die Tandemblöcke dieser kurzen Einheiten sind heterogen in ihrer Länge. In Mais beträgt die Länge dieser Blöcke aus Telomer-spezifischen Sequenzen zwischen 1,8 bis 40 Kbp je nach Gewebe und Entwicklungszustand (Burr *et al.*, 1992). Diese Sequenzen spielen eine funktionelle Rolle bei der Chromosomenreplikation. Aufgrund der Funktionalität dieser repetitiven Sequenzen im Genom liegt ein hoher Konservierungsgrad im Vergleich zu anderen repetitiven Sequenzen vor. In *A. thaliana* besteht das Monomer aus einer 7 Bp-Einheit (5'-CCCTAAA-3') (Richards und Ausubel, 1988).

Weitere repetitive Elemente im Genom sind die kurzen Sequenzwiederholungen, sogenannte Mikro- und Mini-Satelliten. Sie kommen verstreut im Genom vor. Minisatelliten-DNA ist aus Kopien von 10-100 Bp-Monomeren aufgebaut (Ramel,

1997), und Mikrosatelliten-DNA setzt sich aus Kopien von Di- oder Tri-Nukleotid-Einheiten zusammen (Akkaya *et al.*, 1992).

Ebenfalls verteilt im Genom vorkommende Elemente sind die Retrotransposons und ihre Derivate. Es sind transponierbare Elemente, deren Transposition über ein RNA-Intermediat erfolgt. Zu der Klasse der LTR-Retrotransposons gehört unter anderem das *Athila*-Element (Péllissier *et al.*, 1995) aus *A. thaliana*. In Mais sind viele verschiedene Retrotransposon-Familien beschrieben, die in sehr hohen Kopienzahlen vorliegen. In der 240 Kbp *Adh-1*-Region von Mais konnten 10 Retroelement-Familien mit 10 bis 50000 Kopien identifiziert werden. Der Anteil der Retroelemente am Mais-Genom wird auf 50% geschätzt (SanMiguel *et al.*, 1996).

Die andere Klasse der transponierbaren Elemente sind die Transposons, die im Gegensatz zu Retrotransposons über ein DNA-Intermediat mobilisiert werden. Transposon-Elemente sind besonders intensiv in Mais (z.B. *Ac* und *Spm*) charakterisiert worden (Fedoroff, 1989). Diese Elemente kommen in geringen Kopienzahlen (<100 Kopien) vor.

Kleine, AT-reiche transponierbare Elemente mit charakteristischen invertierten repetitiven Sequenzen, die in einer viel höheren Kopienzahl vorkommen, sind MITEs (*miniature inverted-repeat elements*). Zu dieser Familie der mobilen Elemente gehören *Tourist*, *Stowaway* und *Emigrant* (Bureau und Wessler, 1992; Bureau und Wessler, 1994; Casacuberta *et al.*, 1998). Während *Stowaway* in dikotylen und monokotylen Pflanzen gefunden wurde, wurde *Tourist* nur in Gräser-Arten identifiziert (Bureau und Wessler, 1994). In *A. thaliana* konnte mit *Emigrant* ebenfalls ein MITE-Element detektiert werden, das auch in verschiedenen anderen Arten der Brassicaceae-Familie vorkommt (Casacuberta *et al.*, 1998). Während Mitglieder der *Tourist*-Familie in Kopienzahlen von >10000 im Genom von Mais vorliegen, beträgt die Kopienzahl der *Emigrant*-Elemente in verschiedenen *Arabidopsis*-Ökotypen und *Brassica*-Arten 500 bis 1000. Die verschiedenen Familien zeigen keine Sequenzhomologien, haben jedoch die gleichen strukturellen Merkmale. Sie besitzen keine kodierenden Sequenzen. Der Transpositionsmechanismus ist bisher nicht bekannt.

Eine weitere Familie repetitiver Elemente, die in *A. thaliana* vorkommt, ist die *AthE1*-Familie, die ca. 1% des Genoms repräsentiert (Surzycki und Belknap, 1999). Die Vertreter dieser Gruppe haben keine Ähnlichkeiten zu bekannten Retrotransposons oder MITEs und sind nicht durch repetitive Sequenzen flankiert. Verschiedene Vertreter der

Familie zeigen sowohl für die 5'- als auch die 3'-Sequenzen Identitäten über 98%, dies stellt das charakteristische Merkmal dieser Familie dar.

1.3 Karten pflanzlicher Genome

Ein wesentliches Ziel der Genom-Analyse ist die Analyse der Genom-Organisation und die Erstellung von Genom-Karten. Dabei können für das Genom einer Pflanze cytogenetische, genetische und physikalische Karten etabliert werden.

Cytogenetische Karten liefern Informationen über die Struktur des Genoms in Bezug auf Chromosomenzahl, Chromosomengröße und Position der Centromere. Für cytogenetische Analysen eignen sich besonders die Pflanzen-Arten mit großen Chromosomen wie *Fritillaria* ssp. oder *Vicia* ssp. In Mais zum Beispiel wurden durch die Verbindung cytogenetischer Techniken mit genetischen Daten erstmals mobile Elemente postuliert (McClintock, 1948).

Mit genetischen Karten wird der relative Abstand zwischen zwei Markern bestimmt. In klassisch genetischen Karten werden phänotypisch sichtbare Mutationen als Marker eingesetzt. Zur Etablierung der Karten werden Pflanzen, die polymorphe Genloci haben, miteinander gekreuzt und der Phänotyp der Nachkommenschaft wird in den Nachfolge-Generationen bestimmt. Dabei wird die Häufigkeit von Rekombinationsereignissen zwischen den phänotypischen Markern untersucht. Die Rekombinationshäufigkeit spiegelt den Kopplungsgrad der beiden Marker wider und wird als Maß für den relativen Abstand angenommen. Bei Rekombinationshäufigkeiten >50% werden die Marker als ungekoppelt bezeichnet. Die Einheit für die relative Distanz wird mit 1 cM angegeben, dies entspricht einer Rekombinationshäufigkeit von 1%. Zur Kartierung werden Zwei- und Drei-Punkt-Analysen eingesetzt. Für *Arabidopsis thaliana* liegt eine genetische Karte mit mehr als 450 Loci vor (Meinke *et al.*, 1998).

Durch Kopplungsanalysen unter Verwendung trisomer Linien ist es möglich, die ermittelten Kopplungsgruppen aus der klassischen Kartierung mit den entsprechenden Chromosomen in Beziehung zu setzen. Die Positionen von 76 Loci auf den fünf Chromosomen wurden in *A. thaliana* ebenso wie die Position der Centromere für Chromosomen I-III und V auf diese Art bestimmt (Koornneef *et al.*, 1983). Das

Centromer von Chromosom IV wurde später durch Anwendung der Tetraden-Analyse lokalisiert (Copenhaver *et al.*, 1998).

Bei den molekularen Markern handelt es sich dagegen um phänotypisch neutrale DNA-Marker. Die Analyse beruht dabei auf der Detektion von Sequenzunterschieden (DNA-Polymorphismen) in homologen Genomregionen zwischen verschiedenen Pflanzen-Arten oder Ökotypen einer Art. Die DNA-Polymorphismen werden durch molekularbiologische Methoden wie Southern-Blot-Hybridisierung oder PCR-Amplifizierung nachgewiesen. Die RFLP (Restriktionsfragment-Längen-Polymorphismus)-Marker gehören zu dem am meisten verbreiteten Markersystem (Chang und Meyerowitz, 1991).

Weitere molekulare Markertypen, die auf PCR-Methodik beruhen, sind RAPDs (*random amplified polymorphic DNAs*, Williams *et al.*, 1990), CAPS (*cleaved amplified polymorphic sequences*, Konieczny und Ausubel, 1993) AFLPs (*amplified fragment length polymorphisms*, Vos *et al.*, 1995) und SNPs (*single nucleotide polymorphisms*, Cho *et al.*, 1999). Es können auch Mikrosatelliten-Sequenzen zur Erzeugung von Markern ausgenutzt werden, da die Di- und Tri-Nukleotid-Einheiten der Mikrosatelliten sich durch die Anzahl der Wiederholungen in einem repetitiven Block in verschiedenen Ökotypen unterscheiden können (Bell und Ecker, 1994). Ein Vorteil der Marker, die auf PCR basieren, ist im Vergleich zur RFLP-Methode der geringere Bedarf an genomischer DNA zur Durchführung der Analysen.

Molekulare Marker erlauben im Gegensatz zu phänotypischen Markern eine viel schnellere Etablierung von genetischen Karten, für ihre Erstellung werden nur eine bzw. wenige Kartierungspopulationen benötigt. Als Kartierungspopulationen werden oft F₂-Generationen einer Kreuzung zweier Arten oder Ökotypen oder aber rekombinante Inzucht (*recombinant inbred=RI*)-Linien (Burr *et al.*, 1988) eingesetzt. Mit Hilfe molekularer Marker wurden genetische Karten für viele verschiedene Pflanzen-Arten aufgestellt, zum Beispiel für Tomate, *Arabidopsis*, Mais und Kartoffel (Bonierbale *et al.*, 1988; Chang *et al.*, 1988; Helentjaris *et al.*, 1988; Gebhardt *et al.*, 1991). Für einige Pflanzen-Arten wie Tomate, Reis oder Mais liegen Karten mit weit über 1000 Markern vor (Tanksley *et al.*, 1992; Harushima *et al.*, 1998; Davis *et al.*, 1999).

Auch können klassisch-genetische Karten mit molekular-genetischen Karten in Beziehung gesetzt werden (Tanksley *et al.*, 1992; Hauge *et al.*, 1993), wenn eine ausreichende Anzahl von Markern sowohl auf der klassisch-genetischen als auch auf der molekular-genetischen Karte verankert sind. Werden klassisch-genetische

Merkmale, welche für die Pflanzenzüchtung von Bedeutung sind, auf diese Weise auf molekularen Markerkarten verankert, so kann die gewonnene Information zur Marker-gestützten Züchtung (*marker-assisted breeding*) von Pflanzensorten eingesetzt werden. Eine molekulare Markerkarte eines Genoms gibt Auskunft über die Abfolge und genetischen Abstände von Markern entlang der verschiedenen Chromosomen. Daher liefert die molekulare Markerkarte ein Bild über die Organisation der Chromosomen, ihre Erstellung ist dabei unabhängig von der Genomgröße der untersuchten Art. Soll die Genomorganisation der Pflanzen dagegen auf der Ebene der Gene untersucht werden, so ist die Erstellung physikalischer Chromosomenkarten eine notwendige Voraussetzung. Die Klonierung von genomischen DNA-Fragmenten und anschließende Anordnung der klonierten DNA-Fragmente resultiert in der Erstellung einer physikalischen Karte, die auf Klonen beruht. Diese Methodik wurde in Pflanzen bisher erfolgreich für kleine Genome eingesetzt sowie für ausgewählte Bereiche größerer Genome.

Die Klonierung von großen genomischen DNA-Fragmenten in entsprechende Vektoren vereinfacht eine solche Analyse der Genome. Zu diesen Vektoren gehören zum Beispiel die YAC (*yeast artificial chromosome*)-Vektoren (Burke *et al.*, 1987), in die bis zu 1000 Kbp große DNA-Fragmente kloniert werden können. BAC (*bacterial artificial chromosome*)-Klone tragen 100-300 Kbp große DNA-Fragmente (Shizuya *et al.*, 1992). Der Vorteil von BAC- gegenüber YAC-Bibliotheken ist die höhere Stabilität der Klone, der viel geringere Anteil chimärer Klone und die höhere Redundanz der Bibliotheken.

Die Klone der BAC-Bibliotheken von Reis und *A. thaliana* wurden durch *Fingerprinting*-Analyse zu großen Blöcken aus überlappenden Klonen (*Contigs*) angeordnet (Hong *et al.*, 1997; Marra *et al.*, 1999). Bei der *Fingerprinting*-Analyse, die erstmals in *C. elegans* erfolgreich zur Erstellung von Chromosomenkarten mit Cosmid-Klonen eingesetzt wurde (Coulson *et al.*, 1986), werden die Klone mit Restriktionsenzymen gespalten und im Anschluß an die gelelektrophoretische Auftrennung nach teilweise identischen Fragmentmustern untersucht. Die Klonbereiche mit teilweise gleichem Fragmentmuster repräsentieren überlappende Regionen. Die erste Karte eines pflanzlichen Genoms mit der *Fingerprinting*-Methode wurde für *A. thaliana* erstellt (Marra *et al.*, 1999).

Im Gegensatz zu einer solchen Strategie, in der alle Regionen des Genoms gleichzeitig untersucht werden, kann auch eine Methode zur Genomkartierung gewählt werden, welche sich an Karten oder Chromosomen orientiert. Auf diese Weise können die

Arbeiten auf ein Chromosom oder einen genomischen Bereich beschränkt werden. Molekulare Sonden, denen eine bestimmte genetische Kartenposition zugewiesen sind, werden eingesetzt, um korrespondierende Klone aus Klonbibliotheken zu isolieren. Dies verankert die identifizierten Klone auf der genetischen Karte. Wird eine große Anzahl molekularer Sonden und Klonbibliotheken mit großen Insertionen zur Kartierung eingesetzt, so werden Klone oft nicht nur durch eine einzige Sonde verankert, sondern durch zwei und mehr Sonden, die eng gekoppelt in einer Region des Genoms vorliegen. Dadurch können Klone zueinander angeordnet und zu Contigs zusammengefaßt werden. Der Einsatz ausreichend vieler molekularer Marker als Sonden resultiert in einer umfassenden physikalischen Karte des untersuchten Genombereichs oder Genoms.

Die ersten auf diese Weise erstellten Klonkarten pflanzlicher Chromosomen waren die Karten für die Chromosomen II und IV von *Arabidopsis thaliana* (Schmidt *et al.*, 1995; Zachgo *et al.*, 1996).

Durch die Verwendung der molekularen Marker zur Klonidentifizierung sind die erstellte physikalische Karte und die genetische Karte unmittelbar miteinander verknüpft. Daher können genetische und physikalische Distanzen zwischen Markern verglichen werden. Dabei variiert der Wert für den genetischen Abstand von 1 cM in verschiedenen Bereichen des Chromosoms IV von *Arabidopsis thaliana* zwischen 30 Kbp und >550 Kbp (Schmidt *et al.*, 1995; Schmidt *et al.*, 1996).

Die physikalischen Klonkarten können ausgenutzt werden, um Daten über Verteilung von Genen und repetitiven Elementen zu erhalten. Für *A. thaliana* konnte auf diese Weise gezeigt werden, daß repetitive Elemente gehäuft in Centromerbereichen vorkommen, während die euchromatischen Regionen kaum repetitive Sequenzen aufweisen (Schmidt *et al.*, 1995).

Die integrierten physikalischen und genetischen Karten von Genombereichen sind eine Voraussetzung für kartierungsgestütztes Klonieren (*map based cloning*) von Loci. Diese Methode wurde beispielsweise erfolgreich zur Isolierung von Genen in *Arabidopsis* (Arondel *et al.*, 1992; Giraudat *et al.*, 1992), Tomate (Martin *et al.*, 1993), Reis (Song *et al.*, 1995) und Gerste (Büschges *et al.*, 1997) verwendet.

1.4 Untersuchungen pflanzlicher Genome mittels Sequenzierung

Die Sequenzierung stellt eine Methode zur Analyse der Genomorganisation auf Nukleotidebene dar, setzt aber Klonkarten der Chromosomen oder der zu untersuchenden Regionen voraus. Klone werden subkloniert, und die ermittelten Sequenzen werden mit Hilfe von Computerprogrammen zueinander angeordnet. Für Sequenzierungsprogramme, welche die Etablierung der gesamten Sequenz eines Organismus zum Ziel haben, eignen sich besonders Arten mit kleinen Genomen wie zum Beispiel *Arabidopsis thaliana*, als Modellorganismus der Dikotyledonen, und Reis (*Oryza sativa*), als Modellorganismus der Monokotyledonen.

Die Genom-Sequenzierung von *Arabidopsis thaliana* wird bereits durch verschiedene Gruppen durchgeführt (Bevan *et al.*, 1997). Es sind 86,5% des Genoms, was 122488 Kbp entspricht, sequenziert (28.2.00; <http://www.arabidopsis.org/agi.html>). Die beiden Chromosomen II und IV sind bis auf die Centromer, Telomer und Nucleolusorganisierenden Regionen sequenziert (Lin *et al.*, 1999; Mayer *et al.*, 1999). Die Sequenzen sind in verschiedenen Datenbanken zugänglich. Auf Chromosom IV wurde der kurze Arm beginnend von der NOR-Region bis zum repetitiven Bereich des Centromers (3Mbp) sequenziert. Auf dem langen Arm wurde die Sequenz des 14,5 Mbp umfassenden Bereichs zwischen Centromer und Telomer erstellt. Für diese Regionen wird geschätzt, daß ca. 3800 Gene vorliegen. Mehr als 10000 ESTs (*expressed sequence tags*) konnten diesen Genen zugeordnet werden (Mayer *et al.*, 1999). Analysen zur Bestimmung der Gendichte basieren auf Vergleichen der genomischen Sequenz mit EST-Daten und der Anwendung geeigneter Computeralgorithmen für Genvorhersagen. Die Analysen dieser Sequenzen ergaben eine Gendichte zwischen 4,1-4,6 Kbp pro Gen für *Arabidopsis thaliana* (Lin *et al.*, 1999; Mayer *et al.*, 1999; Sato *et al.*, 1998). Die ersten Angaben bezüglich der Gendichte wurden durch Bevan *et al.* (1998) für eine 1,9 Mbp große Region auf dem langen Arm von Chromosom IV gemacht. Für diese Region wurde ein Gen pro 4,8 Kbp ermittelt.

Die Genomsequenzierung des zweiten Modellorganismus, Reis, wurde von mehreren Gruppen begonnen (Sasaki und Burr, 2000). Neben der kleinen Genomgröße von ca. 450 Mbp ist die ökonomisch wichtige Rolle dieser monokotylen Pflanze der entscheidende Faktor für die Genomsequenzierung. Die Sequenzierung wird für den

Reis-Ökotyp *Nipponbare* durchgeführt. Die ersten Daten der Sequenzierung und die verfolgte Strategie sind in der Datenbank zugänglich (<http://www.staff.or.jp/>).

Weitere Sequenzierungsprojekte, die unabhängig von der Genomgröße sind, und die eine Etablierung von physikalischen Karten nicht erfordern, sind die EST-Sequenzierungsprojekte (Höfte *et al.*, 1993; Newman *et al.*, 1994; Cooke *et al.*, 1996; Yamamoto und Sasaki, 1997). In EST-Projekten werden durch Sequenzierung von cDNA-Klonen Teilsequenzen ermittelt und in Datenbanken zugänglich gemacht (<http://www.ncbi.nlm.nih.gov/dbEST/index.html>). Die resultierenden Sequenzdaten sind in der Regel nur 300-500 Bp lang. Die ermittelten Sequenzen werden mit Gensequenzen aus der Datenbank verglichen, und anhand der vorliegenden Sequenzhomologie werden den ESTs Funktionen zugeordnet. Sowohl für *Arabidopsis thaliana* als auch Reis sind bereits mehr als 45000 ESTs sequenziert (<http://www.ncbi.nlm.nih.gov/dbEST/index.html>). Durch die Sequenzierung der EST-Klone, die transkribierte Sequenzen im Genom repräsentieren, ist beispielsweise die Identifizierung von Genen möglich, die durch klassische Methoden nicht identifiziert werden können (Delseny *et al.*, 1997). ESTs, die Sequenzhomologien zu bestimmten Genen wie Resistenzgenen besitzen, können zur Kartierung dieser Gene in molekularen Karten eingesetzt werden (Botella *et al.*, 1997).

1.5 Vergleichende Kartierungsexperimente

Besonders detaillierte Genomuntersuchungen können mit der vorliegenden Methodik zur Zeit nur für wenige Arten durchgeführt werden. Daher ist es von besonderem Interesse zu untersuchen, ob Ergebnisse der Genomorganisation, die für Modellorganismen erarbeitet werden, auf andere Arten übertragbar sind. Vergleichende Kartierungen nehmen dabei eine zentrale Rolle ein.

Die Verwendung von RFLP-Markern in verschiedenen Arten ist aufgrund der hohen Konservierung von Gensequenzen während der Evolution möglich. Für die vergleichende Kartierung werden dieselben Marker in zwei verwandten Spezies eingesetzt. Mit dieser Methode werden Chromosomenkarten für die beiden Arten etabliert, die direkt verglichen werden können. Es können Aussagen über den Kolinearitätsgrad ganzer Chromosomen oder Chromosomen-Abschnitte gemacht

werden. Durch einen Vergleich der Markeranordnung in den verschiedenen Arten können außerdem Duplikationen, Inversionen oder Translokationen von Chromosomenbereichen ermittelt werden.

Die ersten Kartierungsexperimente wurden in der Familie der Solanaceae durchgeführt. Zu dieser Familie gehören Tomate, Kartoffel, Paprika und Tabak. Die ersten vergleichenden RFLP-Karten wurden für Tomate und Kartoffel mit 134 Markern erstellt (Bonierbale *et al.*, 1988). Schon nach kurzer Zeit wurden Karten mit höheren Markerdichten etabliert (Tanksley *et al.*, 1992). Die Vergleiche der genetischen Tomaten-Karte mit der Kartoffel-Karte mit Hilfe der RFLP-Kartierung zeigten einen hohen Grad an Konservierung des Genrepertoires und der Genomkolinearität zwischen diesen beiden Arten (Gebhardt *et al.*, 1991; Tanksley *et al.*, 1992). In sieben von zwölf Chromosomen ist die Reihenfolge der verwendeten molekularen Marker in Tomate und Kartoffel identisch. Die restlichen fünf Chromosomen unterscheiden sich in den beiden Arten durch Inversionen von Chromosomenarmen, die durch Rekombinationsereignisse in der Nähe oder direkt am Centromer verursacht wurden (Tanksley *et al.*, 1992). Im Gegensatz zu der ähnlichen Genomorganisation in Tomate und Kartoffel unterscheidet sich das Paprika-Genom vom Tomaten-Genom trotz ähnlichen Genrepertoires und gleicher Chromosomenzahl ($n=12$) durch zahlreiche Inversionen und Translokationen (> 22) (Tanksley *et al.*, 1988; Prince *et al.*, 1992; Livingstone *et al.*, 1999).

Auch die Familie der Poaceae wurden mit vergleichender genetischer Kartierung auf Genomkolinearität hin untersucht. Zu dieser Familie gehören alle wichtigen Getreide-Arten. Mit der Etablierung von vergleichenden Karten für die homeologen Kopplungsgruppen 2 und 3 in Weizen, Roggen und Gerste konnte für diese drei verwandten Arten ein hoher Grad an Kolinearität für die drei Genome (A, B und D) des hexaploiden Weizens und der Genome von Gerste und Roggen etabliert werden (Devos *et al.*, 1993b; Devos und Gale, 1993). Eine Abweichung von der Kolinearität zeigen die kurzen Arme der Chromosomen 2B (Weizen) und 2R (Roggen). Die beiden Chromosomensegmente scheinen in interchromosomale Translokationen (2BS/6BS und 2RS/7RL) involviert zu sein (Devos *et al.*, 1993b). Mit weiteren vergleichenden Kartierungsexperimenten zwischen Weizen und Roggen wurden Translokationen für alle Roggen-Chromosomen ermittelt (Devos *et al.*, 1993a).

Die Analysen in Mais und Hirse zeigen ebenfalls Kolinearität der verwendeten Marker für große Segmente der beiden Genome (Whitkus *et al.*, 1992; Berhan *et al.*, 1993). Viele der *Sorghum*-Loci kartieren zu zwei Loci in der genetischen Mais-Karte (Whitkus

et al., 1992). In neun Fällen wurden für die beiden Spezies Inversionen der Markerreihenfolge festgestellt, die durch intrachromosomale Translokationen oder Inversionen hervorgerufen wurden. Dagegen wurden in Mais und *Sorghum* keine interchromosomalen Translokationen festgestellt.

Mit vergleichender Kartierung in Mais und Reis wurden kollineare Kopplungsgruppen erstellt, die etwa 67% der Genome von Mais und Reis repräsentieren. Die Kopplungsgruppen sind fast identisch in Bezug auf das Genrepertoire und die Genanordnung. Der Anteil der Loci in Mais, die während der Polyploidisierung dupliziert wurden, konnte mit Hilfe der vergleichenden Kartierung identifiziert werden (72%) (Ahn und Tanksley, 1993).

Durch die vergleichende genetische Kartierung vieler Gräser-Arten wird deutlich, daß ihre Genome durch kollineare Blöcke aufgebaut sind (Moore, 1995). Durch einfaches Umordnen der Blöcke ist es möglich, die Chromosomen von Reis, Mais, Weizen, Hirse etc. aufzubauen (Gale und Devos, 1998). Dies ist beachtlich, da die Trennung von Mais, Reis und Weizen ca. 60 Millionen Jahre zurückliegt (Moore *et al.*, 1995) und große Unterschiede in der Genomgröße in den untersuchten Pflanzen, zum Beispiel zwischen Reis und Mais um das 5-6-fache, beobachtet werden können (Arumuganathan und Earle, 1991).

Vergleichende genetische Kartierungsexperimente wurden zur Bestimmung der Genomkollinearität auch in der Familie der Brassicaceae durchgeführt. Die Genome der diploiden Arten *Brassica oleracea*, *Brassica rapa* und *Brassica nigra* sind vervielfacht, und jedes Chromosomensegment liegt dabei in ein bis drei Kopien vor (Lagercrantz und Lydiate, 1996). Die vergleichende Analyse von *Brassica nigra* und *Arabidopsis thaliana* zeigt eine Konservierung von Chromosomensegmenten mit einer durchschnittlichen Größe von 8 cM (Lagercrantz, 1998), was ca. 90 Rearrangement-Ereignissen seit der Trennung der Vorläufer von *Brassica nigra* und *Arabidopsis thaliana* entspricht. Vergleiche der Markeranordnung in *Arabidopsis thaliana* und *Brassica oleracea* deuten ebenfalls auf eine hohe Zahl von Umorganisationsereignissen der Chromosomen hin. Elf Regionen mit konservierter Organisation wurden identifiziert, die 3,7 bis 49,6 cM großen Kopplungssegmenten in *A. thaliana* entsprechen. Mindestens 17 Translokations- und 9 Inversions-Ereignisse führten zu den Unterschieden in der Genomorganisation der beiden Spezies (Kowalski *et al.*, 1994). In *Brassica napus* wurden für eine 30 cM-Region des *A. thaliana* Chromosoms IV sechs kollineare Bereiche mit einer durchschnittlichen Länge von 22 cM identifiziert. Zwei der

kolinearen Segmente zeigten jedoch Inversionen, die ca. 20 cM groß sind (Cavell *et al.*, 1998). Die vergleichende Analyse zwischen *A. thaliana* und *B. nigra* für einen 7,5 cM großen Bereich aus *A. thaliana* konnte drei kolineare Segmente in *B. nigra* identifizieren. Die Kolinearität eines der drei Segmente ist durch eine Insertion unterbrochen (Lagercrantz *et al.*, 1996).

Vergleichende genetische Kartierungsexperimente eng verwandter Arten ergab in allen Fällen Evidenzen für kolineare Chromosomenabschnitte, jedoch variierte die Länge der identifizierten kolinearen Blöcke. Chromosomenkartierungen auf physikalischer und auf Sequenzebene ermöglichen einen viel detaillierteren Vergleich der Genome als vergleichende genetische Kartierungen. Solche Untersuchungen können die Verteilung von Genen und repetitiven Elementen ermitteln. Auch können Aussagen über die Abfolge von einzelnen Genen in einer Region sowie die Abstände zwischen Genen gemacht werden.

Durch vergleichende physikalische Kartierung der Region um den Selbstinkompatibilitäts (SI)-Locus in *B. campestris* wurde ein hoher Grad an Syntenie zwischen *B. campestris* und *A. thaliana* festgestellt. Allerdings liegt das SI-Gen in der orthologen Region von *Arabidopsis thaliana* nicht vor (Conner *et al.*, 1998).

Eine 15 Kbp-Region von *A. thaliana*, die fünf Gene enthält, zeigt identische Anordnung der Gene in den diploiden *Brassica*-Arten *B. rapa*, *B. oleracea* und *B. nigra*. Die Größe der orthologen Genomregionen in den verschiedenen *Brassica*-Arten umfaßt zwischen 121-320 Kbp. Zumindest der Abstand von zwei Genen ist in *B. rapa* und *Arabidopsis thaliana* ähnlich (Sadowski *et al.*, 1996).

Auch die Analyse der orthologen *Adh*-Region in *Sorghum* und Mais zeigte, daß die Anordnung der Gene in dieser Region kolinear ist. Während die Region in *Sorghum* 75 Kbp groß ist, ist die orthologe Region in Mais dreimal so groß (225 Kbp). 74% der Maisregion setzt sich dabei aus verschiedenen Vertretern der repetitiven Elemente wie Retrotransposons und MITEs zusammen. Auch im Genom von *Sorghum* wurden MITEs und Transposon-ähnliche Elemente identifiziert, allerdings zu einem geringeren Anteil. Die Vergrößerung der Genomregion in Mais ist also auf verstärkte Insertion von repetitiven Elementen zurückzuführen (Tikhonov *et al.*, 1999).

Pflanzen mit großen Genomen, wie beispielsweise Mais, zeichnen sich durch vergrößerte intergenische Bereiche aus (Barakat *et al.*, 1998). Jedoch liegen in Mais auch Genombereiche vor, die eine ähnliche Gendichte aufweisen wie Pflanzen mit kleineren Genomen. Der Vergleich der orthologen Region des *Lrk*- bzw. *Tak*-Gens in

Weizen, Gerste, Mais und Reis zeigte eine Konservierung der Genanordnung und der Gendichte in allen vier untersuchten Pflanzen. Dabei ist die Gendichte mit einem Gen pro 4-5 Kbp ähnlich hoch wie in *A. thaliana*, obwohl sich die untersuchten Pflanzen beträchtlich in der Genomgröße unterscheiden (Feuillet und Keller, 1999). Ein möglicher Grund für die Konservierung des geringen Genabstandes zwischen *Lrk* und *Tak* könnte ein bidirektionaler Promotor sein, der zwischen den beiden zueinander invertiert angeordneten Genen vermutet wird (Feuillet und Keller, 1999).

Zur Zeit liegen nur exemplarische Daten zur Mikrosyntenie vor. In diesen Untersuchungen wurde auch auf der Ebene der Gene ein hoher Grad an Kolinearität festgestellt, jedoch werden erhebliche Unterschiede in Bezug auf Gendichte und des Vorkommens repetitiver Elemente beobachtet.

1.6 Ein Modellsystem für vergleichende Genomanalysen

Als Modellsystem für vergleichende Kartierungen auf genetischer wie auf physikalischer Ebene wurde die Familie der Brassicaceae ausgewählt. In der Brassicaceae-Familie wurden 10-fache Unterschiede in der Genomgröße beobachtet. Die Genomgrößen liegen zwischen ca. 130 Mbp (*Arabidopsis thaliana*) und 1129-1235 Mbp (*Brassica napus*) (Arumuganathan und Earle, 1991). Zu dieser Familie gehören die wichtige Kulturpflanze Raps (*Brassica napus*) sowie die verschiedenen *Brassica*-Arten (*Brassica rapa*, *Brassica oleracea* und *Brassica nigra*), die ein diploides Genom besitzen und aus einem hexaploiden Vorläufer hervorgegangen sind (Lagercrantz und Lydiat, 1996). Zu den Vertretern der *Brassica rapa*-Arten gehören die verschiedenen Rüben-Sorten und Chinakohl. In der *B. oleracea*-Gruppe sind alle bekannten Kohlsorten vertreten wie zum Beispiel Broccoli, Blumenkohl oder Rosenkohl. Schwarzer Senf repräsentiert einen der bekannten Vertreter der *B. nigra*-Gruppe. Aus den diploiden *Brassica*-Arten *B. rapa*, *B. oleracea* und *B. nigra* sind die amphidiploiden *Brassica*-Arten *B. napus*, *B. juncea* und *B. carinata* hervorgegangen (zusammengefaßt in Lydiat *et al.*, 1993).

Arabidopsis thaliana gehört zur Familie der Brassicaceae, hat ein sehr gut untersuchtes Genom von ca. 130 Mbp und einem Anteil an repetitiven Elementen von lediglich 10%. Zahlreiche genetische (Chang *et al.*, 1988; Nam *et al.*, 1989; Liu *et al.*, 1996; Alonso-

Blanco *et al.*, 1998) wie molekulare Karten (zusammengefaßt in Schmidt, 1998; Mozo *et al.*, 1999; Marra *et al.*, 1999) wurden für *A. thaliana* etabliert. Über 80% der Genom-Sequenzen sind ermittelt (<http://www.arabidopsis.org/>). Chromosomen II und IV, die 35% des Genoms repräsentieren, wurden bis auf die Centromer, Telomer und NOR-Regionen sequenziert (Lin *et al.*, 1999; Mayer *et al.*, 1999).

Neben der Genomsequenzierung ist die Identifizierung von Genen und Mutanten ein weiterer Aspekt der Genomanalyse in *A. thaliana* (Meinke *et al.*, 1998). Die Analyse von Mutanten, reverse Genetik und Insertionsmutagenese sind die Techniken, die zur Ermittlung der Gene und Genfunktionen eingesetzt werden. Die Identifizierung der Gene, die bisher auf klassische Weise nicht gefunden wurden, erfolgt über die umfangreichen EST-Sequenzierungsprojekte (Delseny *et al.*, 1997). Mehr als 45000 EST-Sequenzen sind zur Zeit für *A. thaliana* in der EST-Datenbank zugänglich (<http://www.ncbi.nlm.nih.gov/dbEST/index.html>). Die *Microarray*-Technik wird eingesetzt, um Daten über Expressionsmuster der Gene und somit über die Genfunktion zu erhalten (Ruan *et al.*, 1998; Richmond und Somerville, 2000).

Ein Ziel des Genomprojektes für den Modellorganismus *Arabidopsis* ist die Übertragung der Information auf Arten mit größeren Genomen, zum Beispiel für Karten-gestütztes Klonieren (*map-based cloning*) oder Marker-unterstützte Züchtungsprogramme in Raps. Um dies zu ermöglichen, muß zunächst die Übertragbarkeit der ermittelten Daten in *A. thaliana* durch vergleichende Genomanalyse in der Brassicaceae-Familie überprüft werden.

Die *Brassica*-Genome sind aus polyploiden Vorläufern hervorgegangen, daher besitzen sie komplexe Genome (Lagercrantz und Lydiate, 1996). Um in vergleichenden Genomanalysen abschätzen zu können, ob und welche Auswirkungen Polyploidie auf den Grad der Genomkolinearität hat, ist es nötig, einerseits verschiedene diploide Arten zu untersuchen, andererseits müssen diploide und polyploide Arten betrachtet werden.

Die *Capsella*-Arten sind weitere Vertreter der Familie der Brassicaceae, die jedoch ökonomisch nicht von Bedeutung sind. Die bekannteste Art ist *Capsella bursa-pastoris*. Ihren Ursprung haben die *Capsella*-Arten in Eurasien. Die Ausbreitung von *C. bursa-pastoris* ist weltweit zu beobachten. Ein wichtiges Merkmal, das für die Verbreitung der Pflanze verantwortlich ist, ist die genetische Flexibilität durch das polyploide Genom. Während *C. bursa-pastoris* tetraploid ist, kommen in diesem Genus auch diploide Vertreter wie *Capsella rubella* oder *Capsella grandiflora* vor. *C. rubella* ist wie *C. bursa-pastoris* selbstbestäubend. *Capsella grandiflora* dagegen ist selbstinkompatibel.

Die Größe des Verbreitungsgebietes von *C. grandiflora* ist auf Griechenland, den Balkan und Nord-Italien beschränkt. Die Chromosomenzahlen für die diploiden Arten sind $n=8$ Chromosomen (Hurka und Neuffer, 1997). Die Genomgröße von *C. bursa-pastoris* ist mit ca. 680 Mbp (Bennett und Smith, 1976) bestimmt.

Die diploiden Vertreter der *Capsella*-Gruppe stellen trotz ihrer geringen ökonomischen Bedeutung gute Objekte zur vergleichenden Genomkartierung mit *Arabidopsis thaliana* dar. Ihr Genom zeigt Veränderungen in der Genomorganisation, was durch die unterschiedliche Zahl der Chromosomen in beiden Arten ($n=5$ *A. thaliana*; $n=8$ *Capsella*) deutlich wird. Die Trennungszeitpunkte von *Brassica* und *A. thaliana* wurden durch einen Vergleich der plastidären Maturase-Gene (*matK*) mit 16-20 Millionen Jahren bestimmt, während *Capsella* und *Arabidopsis* sich vor ca. 8-10 Millionen Jahren getrennt haben (Koch *et al.*, zur Veröffentlichung eingereicht). Neben den unterschiedlichen Trennungszeitpunkten ist beim Vergleich von *Arabidopsis*, *Capsella* und *Brassica* der unterschiedliche Ploidiegrad der Pflanzen der entscheidende Aspekt. Während *Arabidopsis* und *Capsella* die diploiden Vertreter der Brassicaceae-Familie repräsentieren, leiten sich die diploiden *Brassica*-Arten von hexaploiden Vorfahren her. Daher besitzen sie vervielfachte Genome, wobei für verwendete Marker ein bis drei Loci identifiziert wurden (Lagercrantz und Lydiat, 1996). Durch vergleichende Genomanalyse in den Arten *Arabidopsis*, *Capsella* und *Brassica* kann eine Aussage über den Grad der Genomkolarität in der Familie der Brassicaceae getroffen werden. Da sich diese Arten durch verhältnismäßig kleine Genome auszeichnen ist es möglich, die Vergleiche sowohl auf genetischer als auf Sequenzebene durchzuführen. Lediglich Vergleiche auf diesen verschiedenen Ebenen gewährleisten umfassende Kenntnisse über die Genomstruktur in den untersuchten Pflanzen und lassen verlässliche Aussagen über die Übertragbarkeit von Daten von *A. thaliana* auf verwandte Arten zu. Im Rahmen dieser Arbeit wird der Grad der Genomkolarität sowohl auf genetischer und als auch auf Sequenzebene in *Arabidopsis* und *Capsella* untersucht.

1.7 Zielsetzung

Im Rahmen dieser Arbeit soll eine detaillierte Analyse der Genomkolarität für die beiden diploiden Arten *Arabidopsis thaliana* und *Capsella rubella*, die zu der Familie

der Brassicaceae gehören, durchgeführt werden. Dazu muß die vergleichende Genomanalyse auf genetischer, physikalischer wie auf Sequenz-Ebene in beiden Spezies durchgeführt werden.

Für die vergleichenden Analysen stehen ausgehend von *A. thaliana* zahlreiche Chromosomen-Karten, RFLP-Marker, ESTs und genomische Sequenzen zur Verfügung. Für *Capsella* soll eine geeignete Population erstellt werden, die sich zur Etablierung einer genetischen Karte eignet. *Arabidopsis*-Sonden sollen für die Kartierungsexperimente in *Capsella* eingesetzt werden. Durch vergleichende genetische Kartierungsexperimente soll der Grad der Markerkonservierung und Kolinearität zwischen *Arabidopsis thaliana* Chromosom IV und orthologen Chromosomen-Bereichen in *Capsella rubella* untersucht werden. Die ermittelten Daten sollen zeigen, ob in den diploiden *Capsella*-Arten die Kolinearität durch eine ähnlich hohe Anzahl von Genom-Umorganisationsereignissen unterbrochen ist wie sie für die *Brassica*-Arten beschrieben sind, die hexaploiden Ursprungs sind.

Kleine sequenzierte Bereiche des *Arabidopsis*-Genoms sollen für die vergleichende physikalische Kartierung in *Capsella rubella* herangezogen werden. Hierdurch soll ermittelt werden, inwieweit sich die beiden Arten bezüglich des Genrepertoires und der Genreihenfolge in orthologen Regionen gleichen. Durch Sequenzierung der orthologen Bereiche soll die vergleichende Analyse der Genstrukturen und Sequenzidentitäten ermöglicht werden.

Diese Untersuchungen sollen zeigen, ob und inwieweit Daten über die strukturelle Organisation des *A. thaliana* Genoms auf andere diploide Arten der Familie der Familie der Brassicaceae übertragbar und damit nutzbar sind.



Arabidopsis thaliana *Capsella rubella*

1.8 Abkürzungen

Abb.	Abbildung
<i>A. thaliana</i>	<i>Arabidopsis thaliana</i>
bp (Kbp, Mbp)	Basenpaare (Kilobasenpaare, Megabasenpaare)
BSA	Rinderserumalbumin
°C	Grad Celcius
<i>C. grandiflora</i>	<i>Capsella grandiflora</i>
<i>C. rubella</i>	<i>Capsella rubella</i>
CTAB	Cetyltrimethylammoniumbromid
DNA	Desoxyribonukleinsäure
dNTP	Desoxyribonucleosidtriphosphat
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	Ethylendiaminotetraessigsäure
EST	<i>expressed sequence tag</i>
g	Gramm, Erdbeschleunigung
IPTG	Isopropyl-β-D-thiogalaktopyranosid
mM, M	Millimolar, Molar
PCI	Phenol-Chloroform-Isoamylalkohol
PCR	Polymerase-Kettenreaktion
RFLP	Restriktionsfragmentlängenpolymorphismus
RNase	Ribonuklease
SDS	Natriumdodecylsulfat
Tris	Tris-(Hydroxymethyl)-Aminomethan
v/v	Volumen pro Volumen
w/v	Gewicht pro Volumen
X-Gal	5-Brom-4-Chlor-3-Indolyl-β-D-Galaktopyranosid

2. Material und Methoden

2.1 Material

2.1.1 Chemikalien und Enzyme

Agarose	Gibco BRL, USA
ATP	Boehringer Mannheim
Bacto-Agar, Bacto-Trypton, Hefe-Extrakt	Difco Laboratories, Detroit USA
BSA	MBI Fermentas, St. Leon-Rot
Chemikalien	Calbiochem Behring Corp., USA
	Gibco BRL, USA
	J.T.Baker, Holland
	Merck, Darmstadt
	Pharmacia, Schweden
	Serva, Heidelberg
	Sigma Chem. Co., St. Louis USA
DNA-Längenstandards	MBI Fermentas, St. Leon-Rot
dNTPs	MBI Fermentas, St. Leon-Rot
Filterpapiere	Schleicher & Schüll, Dassel
Klenow-Fragment der DNA-Polymerase I	MBI Fermentas, St. Leon-Rot
Lachsspermen-DNA	Sigma Chem. Co., St. Louis USA
DNA modifizierende Enzyme	Boehringer Mannheim
	MBI Fermentas, St. Leon-Rot
	New England Biolabs, Inc., USA
Nylonmembranen	Amersham, England (HybondN ⁺ -Membran)
	Pall, USA (BiodyneA-Membran)
Nukleotide	MBI Fermentas, St. Leon-Rot
Oligonukleotide	Gibco BRL, USA
	Metabion, Martinsried
	MWG, Ebersbach
PCR-Produktaufreinigung	Boehringer Mannheim (High Pure PCR-Purification-Kit TM)
Plasmid-DNA-Präparation	Boehringer Mannheim (High Pure Plasmid-Kit TM)
Radiochemikalien	Amersham, England (α - ³² P-dCTP)

“Random-Primer” (p(dN) ₆)	Boehringer Mannheim
Restriktionsenzyme	Boehringer Mannheim
	MBI Fermentas, St. Leon-Rot
RNase A	Boehringer Mannheim
Röntgenfilme	Kodak, USA (Kodak-X-OMAT AR-5)
Taq-Polymerase	Gibco BRL, USA
T4-DNA-Ligase	New England Biolabs, Inc., USA

2.1.2 DNA-Längenstandards

	Fragmentgrößen (in Kbp)
GeneRuler™ 1Kbp DNA Leiter (MBI Fermentas)	10/ 8/ 6/ 5/ 4/ 3,5/ 3/ 2,5/ 2/ 1,5/ 1/ 0,75/ 0,5/ 0,25
λ-DNA (MBI Fermentas)	48,5
λ-DNA, <i>EcoRI/HindIII</i> gespalten (MBI Fermentas)	21,226/ 5,148/ 4,973/ 4,268/ 3,53/ 2,027/ 1,904/ 1,584/ 1,375/ 0,947/ 0,831/ 0,564/ 0,125

2.1.3 Lösungen und Medien

Alle Lösungen und Medien wurden in destilliertem Wasser angesetzt und vor Gebrauch entweder autoklaviert oder sterilfiltriert.

2.1.3.1 Medien

Medien für die *E. coli*-Anzucht:

LB:	1% (w/v) Bacto-Trypton 0,5% (w/v) Hefe-Extrakt 1% (w/v) NaCl pH 7,0 mit NaOH
LB-Agar:	LB-Medium 1,5% (w/v) Bacto-Agar

2.1.3.2 Lösungen

PCI:	50% (v/v) Tris gesättigtes Phenol 48% (v/v) Chloroform 2% (v/v) Isoamylalkohol
RNase A:	Stammlösung 10 mg/ml in 10 mM Tris/HCl (pH 7,5), 15 mM NaCl
TE:	10 mM Tris/HCl, pH 8,0 1 mM EDTA, pH 8,0

2.2 Verwendete Bakterienstämme und Vektoren

2.2.1 Bakterienstämme

E. coli K12 DH5 α (Hanahan, 1983)

E. coli SURETM2 (Stratagene)

2.2.2 Vektoren

pGEM7Z (Promega)

pGEM-T (Promega)

pCLD04541 (Bancroft *et al.*, 1997)

2.3 Oligonukleotide

Universe: GTA AAA CGA CGG CCA GT

Reverse: AAC AGC TAT GAC CAT G

T7: GTA ATA CGA CTC ACT ATA GGG C

T3: AAT TAA CCC TCA CTA AAG GG

SP6: CAT ACG ATT TAG GTG ACA CTA TAG

Zusätzliche Oligonukleotide wurden bei den Firmen Gibco, Metabion und MWG synthetisiert.

2.4 Verwendete Marker für genetische Kartierungsexperimente

Als Marker für die genetischen Kartierungsexperimente in *Capsella* wurden *Arabidopsis thaliana* RFLP-Marker (mi und ARMS) (Liu *et al.*, 1996; Fabri und Schäffner, 1994) und ESTs (Höfte *et al.*, 1993; Newman *et al.*, 1994) eingesetzt. Diese wurden vom *Arabidopsis thaliana*-Stock-Center, Ohio, USA, bezogen.

2.5 Molekularbiologische Techniken

Alle gängigen molekularbiologischen Techniken wurden entsprechend den Standardprotokollen durchgeführt (Sambrook *et al.*, 1989).

2.5.1 Isolierung von Pflanzen-Gesamt-DNA

Die Präparation von Pflanzen-Gesamt-DNA erfolgte nach der CTAB-Methode (Dellaporta *et al.*, 1983). 1-2 g junge Blätter wurden in flüssigen Stickstoff schockgefroren und anschließend in einem Mörser mit flüssigem Stickstoff zu einem feinen Pulver verarbeitet. Das gefrorene und in pulverisierter Form vorliegende Blattmaterial wurde in 10 ml auf 65°C vorgewärmten CTAB-Extraktionspuffer aufgenommen und unter wiederholtem Mischen 45 Minuten bei 65°C inkubiert. Ein Volumen PCI wurde zugegeben und 10 Minuten bei Raumtemperatur auf einem Überkopfschüttler inkubiert. Anschließend wurde die Lösung 10 Minuten bei 4000 rpm (Heraeus Varifuge 3.OR) zentrifugiert. Der Überstand wurde in ein neues Reaktionsgefäß überführt, mit einem Volumen Chloroform erneut 10 Minuten bei Raumtemperatur auf dem Überkopfschüttler invertiert und dann 10 Minuten bei 4000 rpm (Heraeus Varifuge 3.OR) zentrifugiert. Für die DNA-Fällung wurde der Überstand in einem Reaktionsgefäß mit 0,7 Volumen Isopropanol vermischt. Die Lösung wurde 5 Minuten auf Eis inkubiert und anschließend bei 4000 rpm (Heraeus Varifuge 3.OR) zentrifugiert. Das resultierende DNA-Pellet wurde in 200-500 µl TE-Puffer mit 10 mg/ml RNase A gelöst und bei 4°C gelagert.

CTAB-Extraktionspuffer:	2% CTAB
	1,4 M NaCl
	0,1 M Tris/HCl, pH 8,0
	20 mM EDTA, pH 8,0
	0,2% β -Mercaptoethanol

2.5.2 Southern-Blot und Hybridisierungsexperimente

Die Southernblot- und Hybridisierungs-Experimente wurden entsprechend den empfohlenen Herstellerangaben durchgeführt (HybondN⁺, Amersham; BiotodyneA, Pall). Nach der Hybridisierung der Nylonmembranen über Nacht bei 65°C wurden die Membranen zuerst in kalter Waschlösung und anschließend bei 60°C 20-30 Minuten gewaschen.

Waschlösung:	0,3 M NaCl
	30 mM Natriumcitrat
	0,1% SDS

2.5.3 Genomische *Capsella rubella* Cosmid-Bibliothek

Für die Identifizierung der *Capsella rubella* Cosmid-Klone wurde eine genomische Gesamt-DNA-Bibliothek von *Capsella rubella* eingesetzt. Die Bibliothek besteht aus 46.000 Einzelklonen (Schmidt *et al.*, 1999).

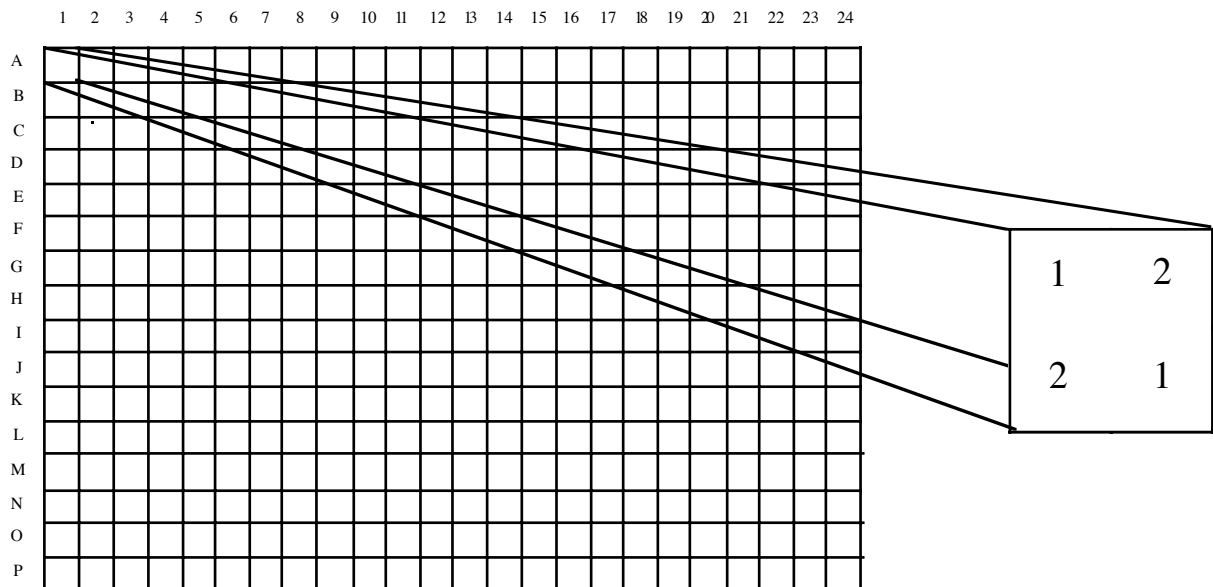
Gesamt-DNA von *Capsella rubella* wurde partiell mit *Mbo*I bzw. *Taq*I gespalten und über einen Saccharose-Gradienten größenfraktioniert. Die DNA-Fractionen, die Fragmente im Bereich von 17-25 Kbp enthielten, wurden in den Cosmidvektor pCLD04541 (Bancroft *et al.*, 1997) kloniert. Die mit *Mbo*I behandelte genomische Pflanzen-DNA wurde in die *Bam*HI-Schnittstelle des Cosmidvektors kloniert, und die mit *Taq*I behandelte DNA in die *Cla*I-Schnittstelle des Vektors. Die durchschnittliche Insertgröße der Cosmidklone beträgt 20 Kbp.

Die Cosmidklone wurden in 384-Kammer-Mikrotiterplatten mit LB-Medium, welches Glycerin und 10 μ g/ml Tetracyclin enthält, bei -70°C gelagert.

2.5.4 Herstellung der *Capsella rubella* Cosmid-Koloniemembranen

Die Cosmidklone wurden mit Hilfe eines Stempels mit 384 Stiften auf LB-Agarplatten mit 10 µg/ml Tetracyclin gestempelt. Auf jeweils eine LB-Agarplatte wurden insgesamt 4x384 Einzelkolonien übertragen. Nach der Inkubation der Platten über Nacht bei 37°C wurde eine Nylonmembran (11x7,5 cm, HybondN⁺; Amersham) luftblasenfrei auf die Agarplatte aufgelegt. Nach 2-3 Minuten wurde die Membran vorsichtig von der Agarplatte abgenommen und mit der Kolonieseite nach oben auf eine LB-Agarplatte mit 10 µg/ml Tetracyclin aufgelegt. Die Weiterbehandlung der Membranen erfolgte nach der Inkubation über Nacht bei 37°C entsprechend der Methode in Sambrook *et al.* (1989).

Für die Erstellung der Koloniemembranen wurden jeweils zwei Mikrotiterplatten zweimal auf eine LB-Agarplatte entsprechend des unten angegebenen Schemas überstempelt.



2.5.5 Koloniehybridisierung der *Capsella rubella* Cosmid-Membranen

Kolonie-Hybridisierungen wurden über Nacht bei 65°C den Herstellerangaben (Amersham) entsprechend durchgeführt. Anschließend wurden die Membranen zuerst mit kalter Waschlösung (2.5.2) bei Raumtemperatur, dann 20-30 Minuten bei 60°C gewaschen.

2.6 Sequenzierung und Sequenz-Analyse

2.6.1 Sequenzierung

Die Sequenzierungen wurden durch die Sequenzierungs-Arbeitsgruppe ADIS (Automatische DNA-Isolierung und Sequenzierung) des Max-Planck-Institutes durchgeführt. Die Bearbeitung der Sequenzen erfolgte auf Unix-Rechnern.

Für die Sequenzierungen wurden die unter 2.3 genannten Oligonukleotide eingesetzt.

2.6.2 Sequenz-Analyse

Für die Bearbeitung der Sequenzen wurden verschiedene Programme aus den GCG-Programm-Paket verwendet (Wisconsin Package Version 10.0, Genetics Computer Group (GCG), Madison, Wisc.).

In den erstellten genomischen *Capsella rubella* Sequenzen sind die kodierenden Regionen in der Regel auf beiden Strängen sequenziert worden. Für Bereiche, die lediglich einmal sequenziert wurden, wurden die Elektropherogramme der jeweiligen Sequenzierungen überprüft.

2.7 Genetische Analyse

2.7.1 Erstellung der Kartierungspopulation

Zur Erstellung einer Kartierungspopulation wurde die selbst-inkompatible Art *Capsella grandiflora* mit Pollen von *Capsella rubella* bestäubt. Um einen guten Samenansatz zu gewährleisten, wurde die Selbstbestäubung der F_1 -Hybridpflanzen mit Pinseln unterstützt.

Ca. 100 Nachkommen einer der F_1 -Hybridpflanzen wurden ausgesät. Für die Kartierung wurden aus dieser Population 50 selbst-kompatible F_2 -Pflanzen ausgewählt. Für jede dieser 50 F_2 -Pflanzen wurde Gesamt-DNA aufgearbeitet.

2.7.2 Polymorphismus-Analyse

Für die Polymorphismus-Analyse wurde Gesamt-DNA von *Capsella rubella* und ein Gemisch von Gesamt-DNA aller F₂-Pflanzen der Kartierungspopulation mit den Restriktionsenzymen *Bgl*II, *Dra*I, *Eco*RI, *Eco*RV bzw. *Xba*I behandelt, auf Agarose-Gelen aufgetrennt, und auf Membranen transferiert.

Das Gemisch von Gesamt-DNA aller F₂-Pflanzen wurde anstelle von Gesamt-DNA von *Capsella grandiflora* eingesetzt, da aufgrund der Selbst-Inkompatibilität dieser *Capsella*-Art nicht ausreichende Mengen an Blattmaterial zur Isolierung von Gesamt-DNA für die Polymorphismus-Analysen zur Verfügung standen. Bei der weiteren Auswertung dieser Experimente mußte daher beachtet werden, daß das Gemisch von Gesamt-DNA der F₂-Pflanzen sowohl den *Capsella grandiflora*- als auch den *Capsella rubella*-Genotyp repräsentiert.

Zur Polymorphismus-Bestimmung wurden für die Hybridisierungsexperimente als DNA-Sonden *Arabidopsis-thaliana*-RFLP-Marker (Fabri und Schäffner, 1994; Liu *et al.*, 1996) oder ESTs (Höfte *et al.*, 1993; Newman *et al.*, 1994) verwendet. Alle Marker wurden sequenziert. Ein Vergleich der ermittelten Sequenzen mit Sequenzen in den Datenbanken diente zur Ermittlung bzw. der Verifizierung der Kartenpositionen der Marker.

Für die Kartierungsdaten der kodominanten Marker wurden Nachkommen, die für einen bestimmten Marker die *C. grandiflora*-Allele homozygot vorliegen hatten, mit 1 bezeichnet. Heterozygote Genotypen für einen bestimmten Marker wurden mit einer 2 benannt, während das Vorliegen homozygoter *C. rubella*-Allele mit 3 angegeben wurde.

Für die Kopplungsanalyse wurde das Programm MAPMAKER mit der Haldane cM-Funktion eingesetzt (Lander *et al.*, 1987). Die genetischen Abstände wurden in centi-Morgan (cM) angegeben.

3. Ergebnisse

3.1 Genetische Kartierung von *Arabidopsis thaliana*-Markern in *Capsella*

Das Ziel der genetischen Kartierungsexperimente in *Capsella* war die vergleichende Untersuchung der Genomorganisation von *Capsella* und *Arabidopsis thaliana*. Die resultierenden Kopplungsgruppen sollten mit der *Arabidopsis* Chromosomenkarte verglichen werden (Lister und Dean, 1993; http://nasc.nott.ac.uk/new_ri_map.html), um Aussagen über den Grad der Genomkolinearität und die Chromosomen-Evolution zu machen. *Arabidopsis thaliana* besitzt 10 Chromosomen (2n), *Capsella rubella*, die auch zur Familie der Brassicaceae gehört, 16 Chromosomen (2n).

Für die genetischen Kartierungsexperimente in *Capsella* wurden *Arabidopsis thaliana* RFLP-Marker (mi-...: Liu *et al.*, 1996; m-...: Fabri und Schäffner, 1994) und ESTs (Höfte *et al.*, 1993; Newman *et al.*, 1994) eingesetzt, die vom *Arabidopsis thaliana*-Stock-Center (Ohio, USA) bezogen wurden.

3.1.1 Kartierungspopulation von *Capsella*

Da keine Informationen über den Polymorphismusgrad in *Capsella* vorlagen, wurde eine interspezifische Kreuzung für die Etablierung der genetischen Karte von *Capsella* gewählt. Hierfür wurden *C. grandiflora*-Blüten mit *C. rubella*-Pollen bestäubt (Marcus Koch, unveröffentlicht). Bei *C. grandiflora* handelt es sich um eine selbstinkompatible Spezies. Nach Selbstbestäubung einer Pflanze, die aus dieser Kreuzung stammte, wurden 100 Pflanzen aus der F₂-Generation angezogen. Aus dieser Population wurden 50 selbstkompatible Pflanzen für die genetischen Kartierungsexperimente ausgewählt.

3.1.2 Restriktions-Fragment-Längen-Polymorphismus (RFLP)-Analyse von *C. grandiflora* und *C. rubella*

Mit genetischen Markern als radioaktiv markierte Sonden sollte der Polymorphismus-Grad zwischen den beiden Spezies *C. grandiflora* und *C. rubella*, die den genetischen Hintergrund der Kartierungspopulation repräsentieren, festgestellt werden, um Aussagen über die Eignung der erstellten Kartierungspopulation zu treffen.

Für die DNA-Polymorphismus-Analyse wurden Membranen mit Gesamt-DNA von *C. rubella* und einem Gemisch von Gesamt-DNA der F₂-Pflanzen, die die

Kartierungspopulation repräsentieren, erstellt. Zur Untersuchung wurde Gesamt-DNA benutzt, die mit den Restriktionsenzymen *Bgl*III, *Dra*I, *Eco*RI, *Eco*RV, *Hind*III bzw. *Xba*I gespalten worden war. Das Gemisch aus Gesamt-DNA der F₂-Pflanzen wurde statt Gesamt-DNA von *C. grandiflora* eingesetzt, da aufgrund der Selbstinkompatibilität dieser *Capsella*-Art keine homozygote Linie herangezogen werden konnte. Für die weiteren Untersuchungen war zu beachten, daß das DNA-Gemisch sowohl die Allele von *C. grandiflora* als auch die von *C. rubella* repräsentiert.

3.1.3 Genetische Marker für die Kartierung

Es wurden 33 Marker für Chromosom IV und 13 Marker für die restlichen vier Chromosomen von *A. thaliana* für die RFLP-Analyse eingesetzt. Es handelt sich bei diesen Markern um RFLP-Marker (mi...-Marker; Liu *et al.*, 1996; ARMS-Marker, m:...; Fabri und Schöffner, 1994) und um EST-Sequenzen (Höfte *et al.*, 1993; Newman *et al.*, 1994). In den folgenden Tabellen 3.1-3.3 und Abbildung 3.1 wurden diese Marker, die zur Überprüfung sequenziert wurden, mit der jeweiligen Position auf der genetischen Karte angegeben. Für Marker mi390, der mit Hilfe von RFLP-Kartierung auf Chromosom 4 von *Arabidopsis thaliana* kartiert worden war (Liu *et al.*, 1996), ergab die Kartierung der ermittelten Sequenz, daß der Marker zu einem Locus auf Chromosom 2 korrespondiert (Tabelle 3.1, Abb. 3.1). Für die Marker, welche in *Arabidopsis* nicht auf der genetischen Karte lokalisiert sind, werden RFLP-Marker angegeben, die physikalisch eng benachbart sind. Diese Informationen wurden mit Hilfe der *Arabidopsis thaliana* Datenbank ermittelt (<http://www.arabidopsis.org/>).

Das Ziel der Arbeit für die vergleichende genetische Analyse von *C. rubella* mit *A. thaliana* war in erster Linie die Erstellung einer vergleichenden genetischen Karte mit den *A. thaliana* Markern von Chromosom IV. Die Kartierung der *A. thaliana* Marker von Chromosom I war erforderlich, da im zweiten Teil der Arbeit die Mikrosyntenie einer Region des Chromosoms I von *A. thaliana* im Vergleich zu *Capsella rubella* analysiert werden sollte (3.2.2). Die Untersuchung der Marker, die auf den *Arabidopsis* Chromosomen I, II, III und V liegen, sollte außerdem zeigen, ob Kopplungen zu Markern von Chromosom IV in *Capsella* nachgewiesen werden können.

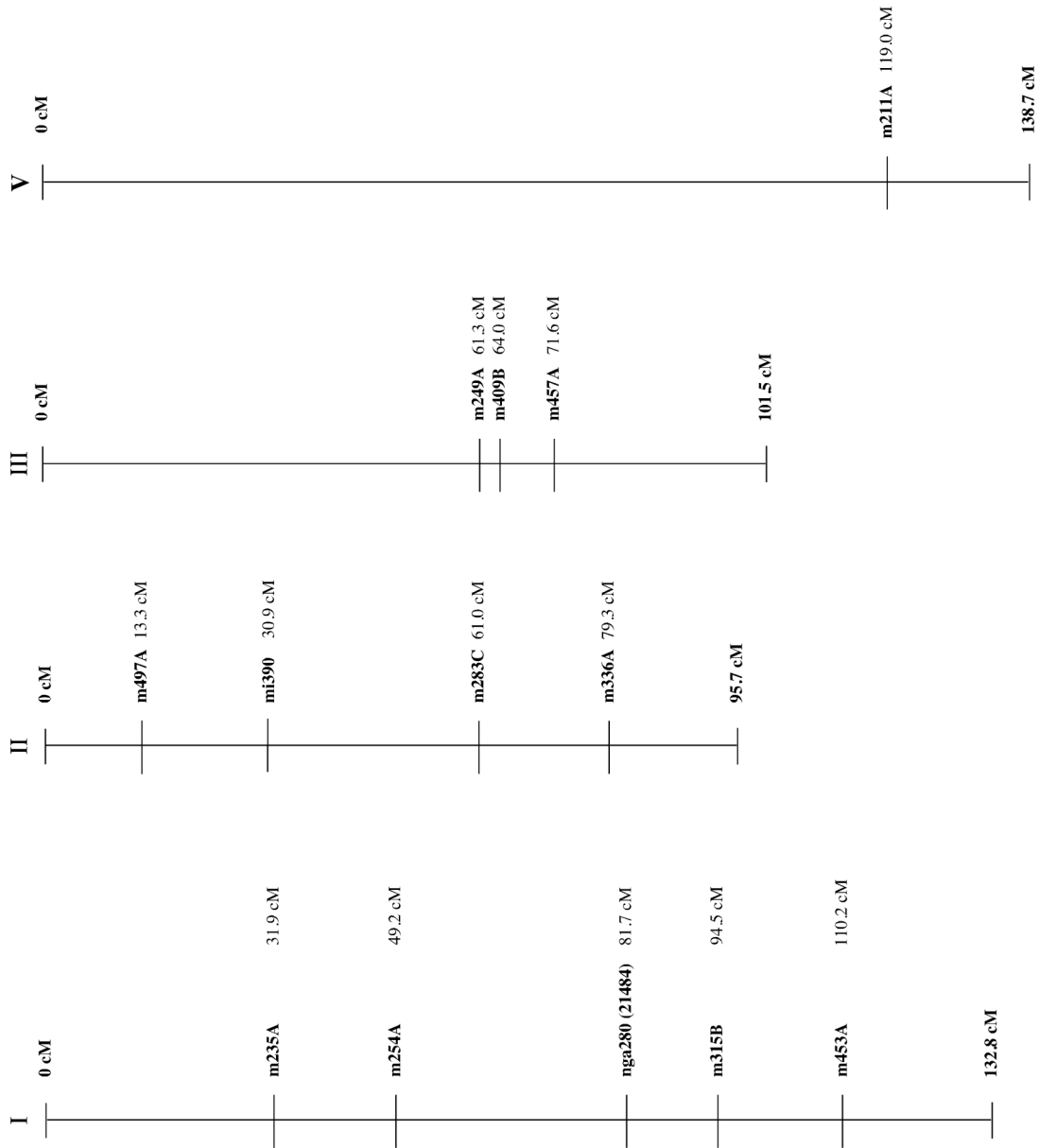


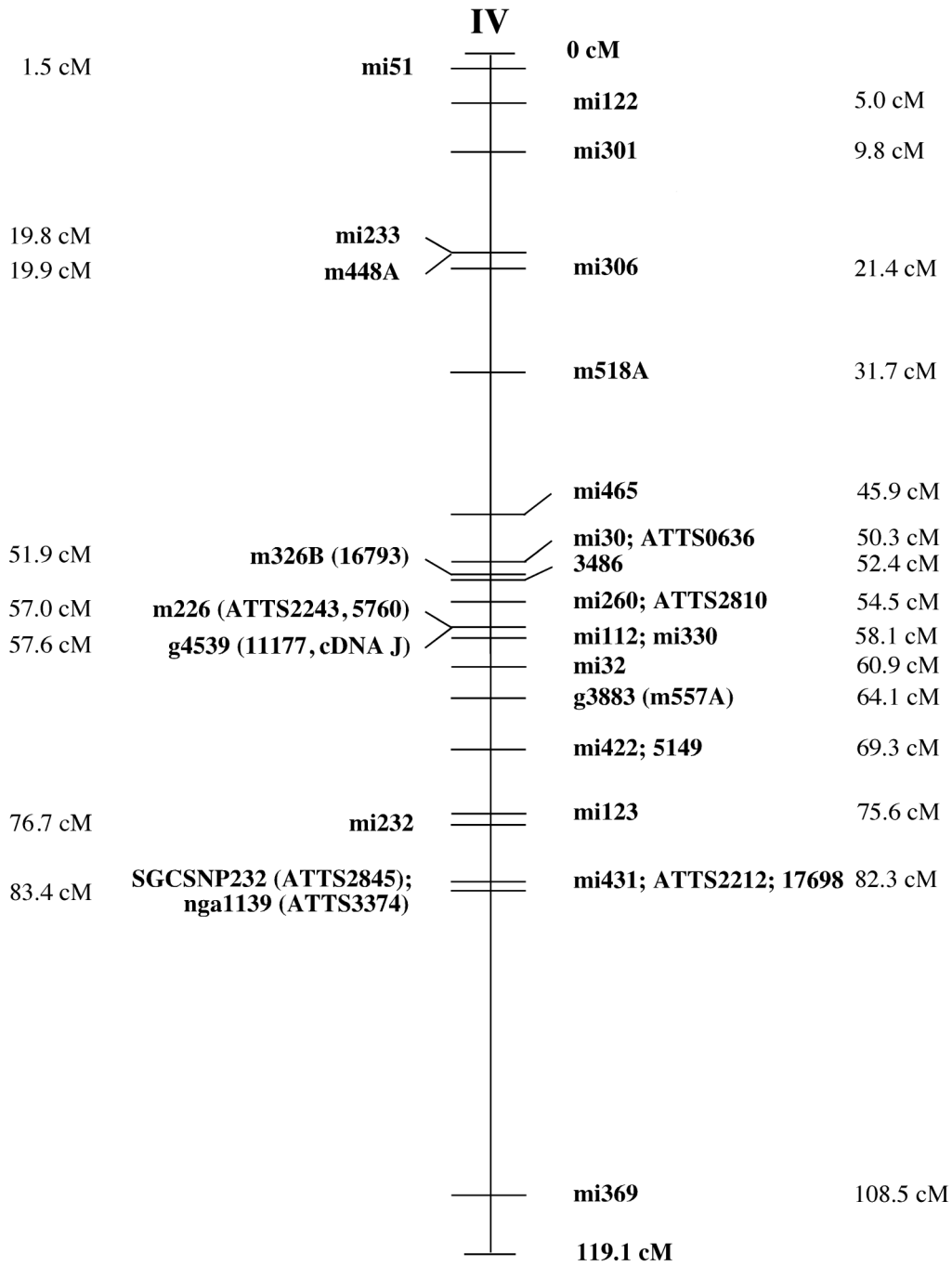
Abbildung 3.1: Die genetische Karte der fünf Chromosomen von *A. thaliana* (I-V) ist abgebildet. Die Informationen zur Erstellung der Karte wurden aus der *A. thaliana*-Datenbank bezogen (http://nasc.nott.ac.uk/new-ri_map.html).

Die genetischen Positionen der für die vergleichende Kartierung verwendeten Marker sind in centi-Morgan (cM) angegeben. Für Marker, die in *Arabidopsis* genetisch nicht kartiert wurden, werden Positionen von Markern gezeigt, die unmittelbar benachbart vorliegen.

Teil A: *A. thaliana* Chromosomen I, II, III und V.

Teil B: *A. thaliana* Chromosom IV.

A



3.1-B

Arabidopsis thaliana Chromosom IV mi...-RFLP-Marker:

Marker	Position in <i>Arabidopsis thaliana</i> RI-Karte (cM)
mi30	50,3
mi32	60,9
mi51	1,5
mi112	58,1
mi122	5,0
mi123	75,6
mi232	76,7
mi233	19,8
mi260	54,5
mi301	9,8
mi306	21,4
mi330	58,1
mi369	108,5
mi390	Chr. II 30,9
mi422	69,3
mi431	82,3
mi465	45,9

Tabelle 3.1: Für die RFLP-Analyse eingesetzte mi...-Marker (Liu *et al.*, 1996), die auf Chromosom IV von *A. thaliana* kartieren (http://nasc.nott.ac.uk/new_ri_map.html), sind in cM angegeben. Es handelt sich bei diesen Markern um genomische Zufallsfragmente. Die Sequenz des Markers mi390 korrespondiert zu einer Kartenposition auf Chromosom II.

Arabidopsis thaliana ARMS-Marker:

Marker	Chromosom-(cM)	Sequenzierter Klon
m235A	I (31,9)	T26J12
m254A	I (49,2)	F28K20
m315B	I (94,5)	F18J1
m453A	I (110,2)	F24J13
m283C	II (61,0)	T6B20
m336A	II (79,3)	F16B22
m497A	II (13,3)	F5G3
m249A	III (61,3)	F18L15
m409B	III (64,0)	T23J7
m457A	III (71,6)	F18B3
m326B	IV (51,9)	ATFCA1
m448A	IV (19,9)	F9H3
m518A	IV (31,7)	T22B4
m557A	IV (64,1)	T5K18
m211A	V (119,0)	MJH22

Tabelle 3.2: Aufgelistet wurden ARMS-Marker von *A. thaliana* (Fabri und Schäffner, 1994). Die Marker-Position auf den jeweiligen Chromosomen sind in cM angegeben (http://nasc.nott.ac.uk/new_ri_map.html) sowie die sequenzierten Klone, auf denen die ARMS-Marker lokalisiert sind (<http://www.arabidopsis.org/maps.html>).

Arabidopsis thaliana ESTs und cDNAs:

Marker	Chromosom (cM)	Genbank-Akzession	Sequenzierter Klon	Physikalischer Abstand zum nächsten kartierten Marker in der RI-Karte
21484	I (81,7)	N96681	F20N2	<100 Kbp nga280
3486	IV(52,4)	T21478	ATFCA2	<110 Kbp mi198
5149	IV (69,3)	T41886	F16G20	<60 Kbp mi422
16793	IV (51,9)	R90438	ATFCA1	<170 Kbp m326
ATTS0636	IV (50,3)	Z18140	ATFCA0	<90 Kbp mi30
ATTS2212	IV (82,3)	Z29768	F26P21	<300 Kbp mi431
ATTS2243	IV (57,0)	Z29799	ATFCA6	<200 Kbp m226
ATTS2810	IV (54,5)	Z33719	ATFCA5	<65 Kbp mi260
ATTS2845	IV (83,4)	Z33754	T4L20	<15 Kbp SGCSNP232
ATTS3374	IV (83,4)	Z34614	M4E13	<230 Kbp nga1139
5760	IV (57,0)	T42497	ATFCA6	<200 Kbp m226
11177	IV (57,6)	T76399	ATFCA7	<150 Kbp g4539
cDNA J	IV (57,6)		ATFCA7	<100 Kbp g4539
17698	IV (82,3)	H75993	F10N7	<300 Kbp mi431

Tabelle 3.3: Angegeben sind die verwendeten *A. thaliana*-ESTs und die cDNA J (3.2.1). Aufgeführt sind die Markerbezeichnung in der Datenbank, die Chromosomenposition (http://nasc.nott.ac.uk/new_ri_map.html), die Genbank-Akzessionsnummer und der sequenzierte Klon, auf welchem der Marker lokalisiert ist.

3.1.4 RFLP-Analyse der *Arabidopsis thaliana* Chromosom IV-Marker

Von den 46 getesteten Markern konnte für zwei Marker, mi32 und ATTS2810, kein Polymorphismus mit den verwendeten sechs verschiedenen Restriktionsenzymen *Bgl*II, *Dra*I, *Eco*RI, *Eco*RV, *Hind*III bzw. *Xba*I gefunden werden (Daten nicht gezeigt). Für weitere sechs Marker, EST11177, mi233, m497A, m283C, m336A und m409B, konnte mit Gesamt-DNA aus *Capsella* keine Hybridisierung nachgewiesen werden (Abb. 3.2, Daten nicht gezeigt). Offensichtlich liegen die entsprechenden Markersequenzen in *Capsella* nicht vor. Das Ergebnis der Hybridisierung einer Membran mit *Dra*I beziehungsweise mit *Eco*RI/*Hind*III gespaltener Gesamt-DNA von *A. thaliana* und *C. rubella* bestätigte, daß mi233 und EST11177 mit *A. thaliana*-DNA hybridisieren, jedoch nicht mit *C. rubella*-DNA (Abbildung 3.2).

Für die monomorphen Marker, mi32 und ATTS2810, wurde die RFLP-Analyse nicht mit weiteren Restriktionsenzymen durchgeführt, da für die restlichen 29 Marker, die auf Chromosom IV lokalisiert sind, mit verschiedenen Restriktionsenzymen Polymorphismen ermittelt werden konnten (Tabelle 3.4).

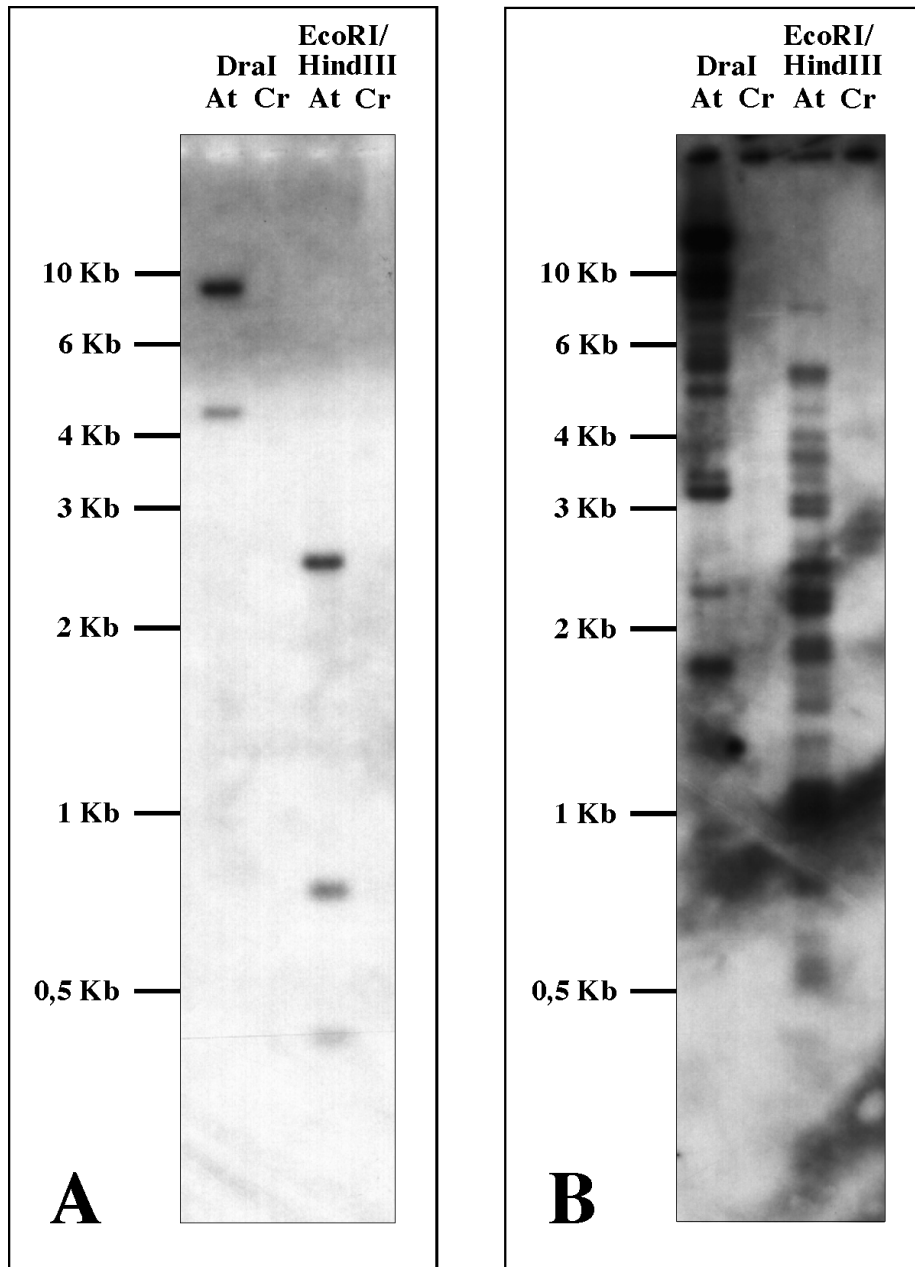


Abbildung 3.2: Die Ergebnisse der Hybridisierungen mit den radioaktiv markierten Sonden der Marker mi233 (**A**) und EST 11177 (**B**) sind gezeigt. An die Membran ist mit *DraI*- bzw. *EcoRI/HindIII*-gespaltene Gesamt-DNA von *A. thaliana* (At) bzw. *C. rubella* (Cr) gebunden.

Marker 11177 (**B**) zeigt ein repetitives Hybridisierungsmuster mit Gesamt-DNA von *A. thaliana*. Die Signalstärke der hybridisierenden DNA-Fragmente liefert Hinweise über den Grad der Sequenzhomologie zwischen der verwendeten Sonde und den hybridisierenden DNA-Fragmenten.

Marker	<i>Bgl</i> III	<i>Dra</i> I	<i>Eco</i> RI	<i>Eco</i> RV	<i>Hind</i> III	<i>Xba</i> I
mi30	+	-	-	-	+	+
mi51	-	+	+	+	-	-
mi112	-	+	-	-	-	-
mi122	+	-	-	+	-	+
mi123	-	+	+	+	+	-
mi232	-	-	+	+	+	-
mi260	-	+	+	+	-	-
mi301	-	+	+	-	-	-
mi306	+	-	-	-	-	+
mi330	+	-	-	+	-	+
mi369	-	+	-	-	-	-
mi390	-	+	+	-	-	+
mi422	-	-	+	-	+	-
mi431	+	-	-	+	+	-
mi465	-	-	+	+	-	+
3486	n.g.	n.g.	+	+	n.g.	+
5149	n.g.	n.g.	+	+	n.g.	+
16793	n.g.	n.g.	n.g.	+	n.g.	+
ATTS2243	n.g.	n.g.	+	+	n.g.	+
ATTS2212	n.g.	n.g.	+	+	n.g.	+
ATTS2243	-	+	-	-	-	-
ATTS2845	-	-	+	+	-	-
ATTS3374	-	-	-	-	+	+
cDNA J	+	+	+	+	+	+
21484	-	+	-	-	-	-
17698	-	+	+	+	-	+
5760	-	+	+	-	-	-
m235A	n.g.	n.g.	+	n.g.	n.g.	n.g.
m254A	n.g.	n.g.	n.g.	+	n.g.	+
m315B	-	-	-	+	-	+
m453A	+	-	-	-	+	-
m249A	n.g.	n.g.	n.g.	n.g.	n.g.	+
m457A	-	+	-	+	+	+
m448A	n.g.	n.g.	+	+	n.g.	n.g.
m518A	n.g.	n.g.	+	n.g.	n.g.	+
m326B	n.g.	n.g.	+	+	n.g.	n.g.
m557A	n.g.	n.g.	n.g.	n.g.	n.g.	+
m211A	+	+	-	-	+	+

Tabelle 3.4: Die Ergebnisse der Polymorphismusanalyse der Marker von *A. thaliana* mit *Capsella*-Gesamt-DNA ist gezeigt. Gesamt-DNA von *C. rubella* und dem Gemisch aus Gesamt-DNA von Pflanzen der Kartierungspopulation wurde mit verschiedenen Restriktionsenzymen gespalten, aufgetrennt und auf Membranen transferiert. Mit “+” bzw. “-“ ist gekennzeichnet, ob ein Polymorphismus zwischen *C. grandiflora* und *C. rubella* mit dem jeweiligen Restriktionsenzym detektiert wurde. Nicht ermittelte Daten sind mit “n.g.” angegeben.

Für 12 Marker wurden bei den Hybridisierungen Membranen verwendet, die nur mit drei verschiedenen Enzymen (*Eco*RI, *Eco*RV und *Xba*I) gespaltene Gesamt-DNA von *C.*

rubella und dem Gemisch aus Gesamt-DNA von 50 Pflanzen der Kartierungspopulation gebunden hatten. Für die restlichen 26 Marker wurden Membranen eingesetzt, die mit sechs verschiedenen Restriktionsenzymen (*Bgl*II-, *Dra*I-, *Eco*RI-, *Eco*RV-, *Hind*III- bzw. *Xba*I-) gespaltene DNA trugen. In Abbildung 3.3 ist exemplarisch das Hybridisierungsergebnis für den Marker cDNA J gezeigt.

Der Polymorphismus-Grad der einzelnen Markern ist sehr unterschiedlich. Der Marker cDNA J zeigt für alle sechs verschiedenen Restriktionsspaltungen Polymorphismen. Für die Marker mi369, ATTS2243 und 21484 dagegen konnte nur für die Spaltung mit einem Restriktionsenzym ein Polymorphismus nachgewiesen werden. In der Regel wurden Polymorphismen am häufigsten mit *Eco*RI, *Eco*RV, *Hind*III und *Xba*I gespaltener Gesamt-DNA ermittelt (Tabelle 3.5).

Restriktions-enzym	Anzahl der Marker, die Polymorphismus zeigen	Gesamt-Zahl der Marker, die mit dem jeweiligen Enzym getestet wurden	Polymorphismus-Häufigkeit (%)
<i>Bgl</i> II	8	26	30,8
<i>Dra</i> I	14	26	53,8
<i>Eco</i> RI	20	34	58,8
<i>Eco</i> RV	21	34	61,8
<i>Hind</i> III	10	26	38,5
<i>Xba</i> I	21	35	60,0

Tabelle 3.5: Berechnung der Polymorphismushäufigkeit

Es ist angegeben, wie viele Marker einen Polymorphismus mit DNA zeigen, die mit einem bestimmten Enzym gespalten wurde. Außerdem ist angegeben, wie viele Marker insgesamt für die Analyse mit dem jeweiligen Enzym verwendet wurden.

Die Marker mi260 und mi422 wurden in weiteren Experimenten nicht eingesetzt, da diese Marker ein repetitives Hybridisierungsmuster zeigten. Der Sequenzhomologie-Grad zwischen *Arabidopsis* und *Capsella* ist nicht für alle getesteten Marker gleich hoch. Für mi198 und mi465 beispielsweise ist die Hybridisierungssignalstärke in *C. rubella* gering, was auf einen niedrigen Homologiegrad zwischen *Arabidopsis* und *Capsella* zurückschließen läßt. Die Marker wurden für die weiteren Experimente nicht benutzt.

3.1.5 Kartierung der Marker in *Capsella*

22 RFLP-Marker (Fabri und Schäffner, 1994; Liu *et al.*, 1996) und ESTs (Höfte *et al.*, 1993; Newman *et al.*, 1994), die auf *Arabidopsis thaliana* Chromosom IV lokalisiert sind und für welche DNA-Polymorphismen ermittelt werden konnten, wurden für die genetische Kartierung eingesetzt (Tabelle 3.6).

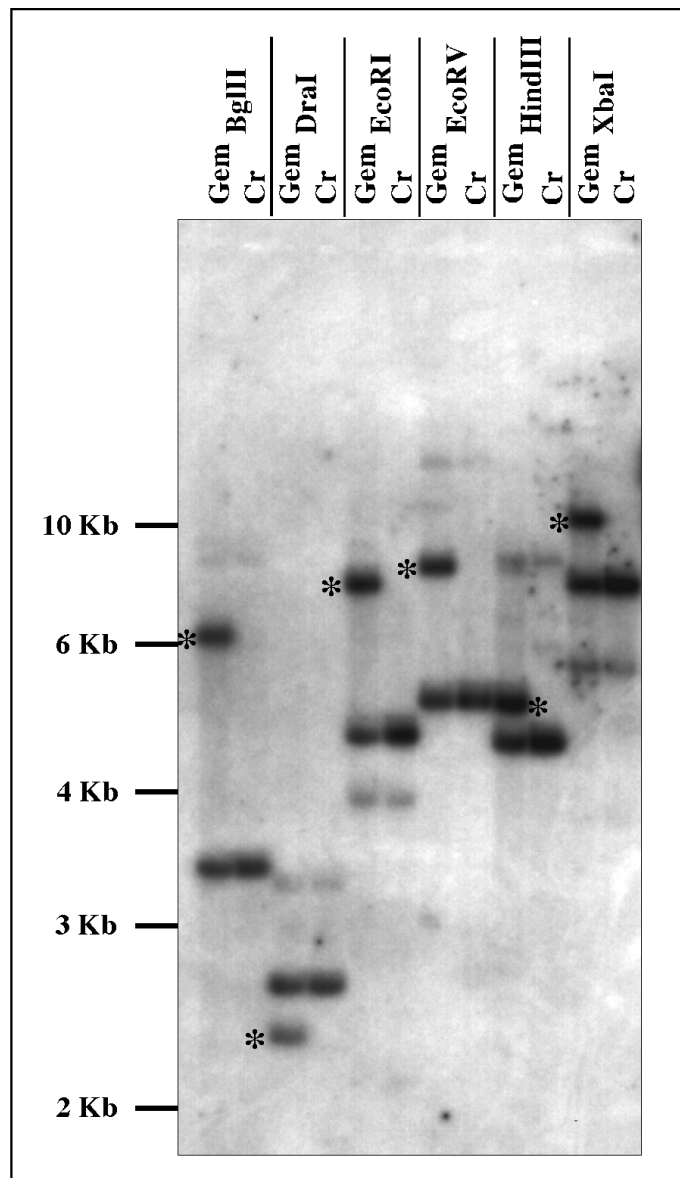


Abbildung 3.3: Das Hybridisierungsergebnis von Marker cDNA J mit einer Membran, auf die mit sechs verschiedenen Restriktionsenzymen gespaltene Gesamt-DNA von *C. rubella* (Cr) sowie einem DNA-Gemisch der 50 F2-Pflanzen der Kartierungspopulation (Gem) transferiert wurde, ist abgebildet. Bei den verwendeten Enzymen handelt es sich um: *BglII*, *DraI*, *EcoRI*, *EcoRV*, *HindIII* und *XbaI*.

An der linken Bildseite ist in Kbp der Längenstandard angegeben. Die Restriktionsfragmente, die einen RFLP anzeigen, sind mit einem Stern (*) gekennzeichnet.

In den Gelspuren mit Gesamt-DNA-Gemisch (Gem) hybridisiert der Marker mit DNA-Fragmenten des *C. rubella*- und des *C. grandiflora*-Allels.

Bei 18 Markern handelt es sich um Marker, die in *Capsella* zu einem Locus im Genom korrespondieren, während für die restlichen vier Marker 3486, ATTS2212, ATTS3374 und mi330 je zwei RFLP-Loci ermittelt wurden. Abbildung 3.4A zeigt das Hybridisierungsergebnis mit Marker mi122 für 25 Pflanzen der *Capsella*-Kartierungspopulation. Marker mi122 hybridisiert in *C. grandiflora* mit einem ca. 10 Kbp langen DNA-Fragment, während es in *C. rubella* mit einem ca. 7 Kbp langen DNA-Fragment hybridisiert. Das Hybridisierungsergebnis für Marker ATTS3374, der wie die drei weiteren Marker 3486, ATTS2212 und mi330 im *Capsella*-Genom mindestens zwei Loci besitzt, ist in Abbildung 3.4B gezeigt.

Zusätzlich wurden die Marker m315B, m235A, 21484 und m254A von *Arabidopsis* Chromosom I, mi390 von Chromosom II, m457A und m249A von Chromosom III und m211A von Chromosom V in *Capsella* kartiert. Für den Chromosom I-Marker m315B konnten im *Capsella*-Genom zwei Loci ermittelt werden.

Die ermittelten Daten in der Kartierungspopulation sind in Tabelle 3.6 und im Anhang aufgeführt. In der Kartierungspopulation wird für die kodominanten Marker eine 1:2:1-Verteilung (*C. grandiflora* homozygot: heterozygot: *C. rubella* homozygot) erwartet. Die Daten einer χ^2 -Analyse zeigen für die Marker mi30, mi306, m211A, m326A und ATTS0636 eine abweichende Verteilung. Für die Allele von *C. rubella* und *C. grandiflora* wird eine 1:1-Verteilung erwartet. Die Daten der χ^2 -Analyse zur Allelverteilung zeigte nur für die beiden Marker m211A und m457A abweichende Werte. Der Signifikanzwert betrug für die Analysen 0,05.

Für die Kopplungsanalyse wurde das Programm MAPMAKER mit einem lod-Wert von 3 verwendet (Lander *et al.*, 1987). Die genetischen Abstände wurden in centi-Morgan (cM) angegeben (Haldane-Funktion). Abbildung 3.5 zeigt den Vergleich zwischen den ermittelten Kopplungsgruppen für *Capsella* und der Karte von *Arabidopsis thaliana* Chromosom IV. Die Kartierungsdaten beziehen sich auf die rekombinante Inzuchtlinien-Population von *A. thaliana* (Lister und Dean, 1993) und wurden der *A. thaliana*-Datenbank entnommen (http://nasc.nott.ac.uk/new_ri_map.html). Für die Chromosom IV-Marker konnten in *Capsella* zwei Kopplungsgruppen ermittelt werden. In beiden Kopplungsgruppen wurde ein hoher Grad an Kolinearität beobachtet. Lediglich Marker 3486 zeigte in *Capsella* keine Kopplung zu den restlichen Chromosom IV-Markern. Einer der zu Marker m315B korrespondierenden Loci (*A. thaliana* Chromosom I-Marker) (m315Bb) wurde in *Capsella* gekoppelt zu *Arabidopsis* Chromosom IV-Marker mi122 und m448A gefunden. Die anderen auf den Chromosomen I, II, III und V von *Arabidopsis* lokalisierten RFLP-Marker zeigten in *Capsella* keine Kopplung zu den Loci, die mit Chromosom IV von *Arabidopsis* identifiziert wurden (siehe Anhang).

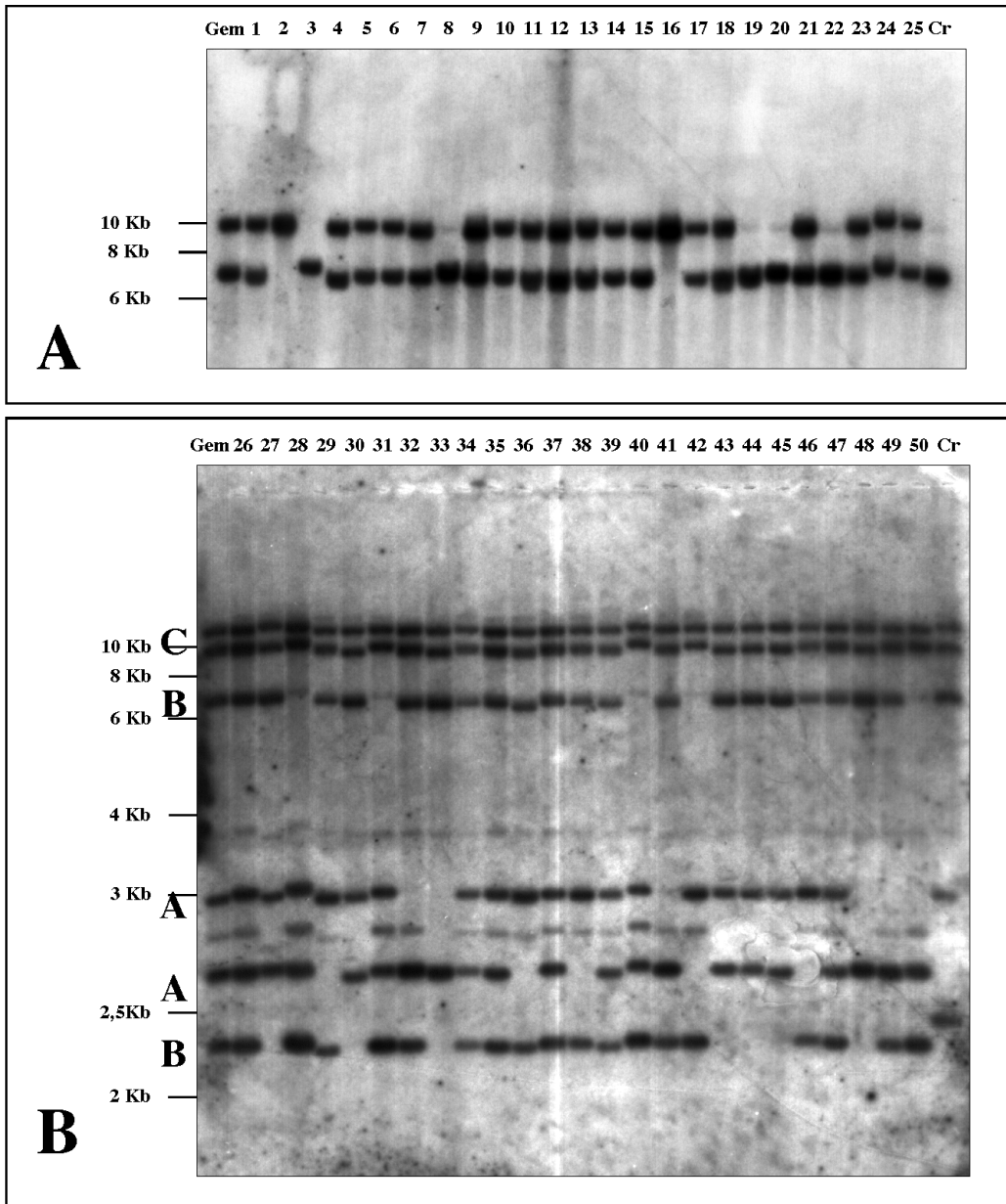


Abbildung 3.4: A: Gezeigt ist das Hybridisierungsergebnis der Sonde mi122 mit einer Membran, die *Bgl*III-gespaltene Gesamt-DNA vom DNA-Gemisch der Kartierungspopulation (Gem), von Pflanzen der Kartierungspopulation (1-25) und von *C. rubella* (Cr) gebunden hat. mi122 korrespondiert in *Capsella* zu einem Locus.

B: Hybridisierung des Markers ATTS3374 mit einer Membran, die *Xba*I-gespaltene Gesamt-DNA vom DNA-Gemisch der Kartierungspopulation (Gem), von Pflanzen der Kartierungspopulation (26-50) und von *C. rubella* (Cr) gebunden hat.

Marker ATTS3374 liegt in *Capsella* an mindestens zwei unterschiedlichen Loci vor. Die polymorphen DNA-Fragmente der jeweiligen Loci sind mit **A** bzw. **B** und die beiden monomorphen DNA-Fragmente wurden mit einem **C** gekennzeichnet.

Bei DNA-Fragmenten mit schwächeren Hybridisierungssignalen handelt es sich um Kreuzhybridisierungen.

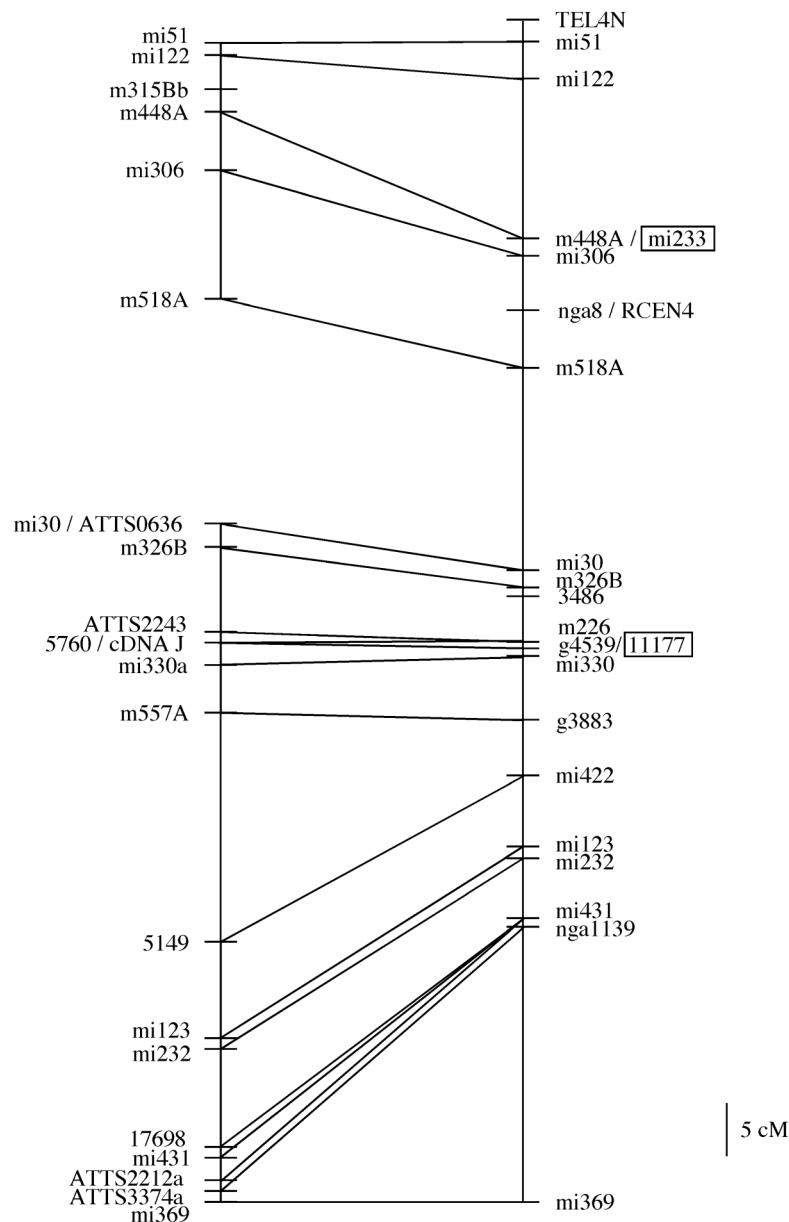


Abbildung 3.5: Vergleichende genetische Karte von Chromosom IV von *Arabidopsis thaliana* und den *Capsella*-Kopplungsgruppen. Auf der rechten Bildhälfte ist die Col x Ler RI-Karte für *A. thaliana* Chromosom IV angegeben. Die Informationen zur Erstellung der Karte wurden aus der *A. thaliana*-Datenbank bezogen (http://nasc.nott.ac.uk/new-ri_map.html). Links sind die beiden ermittelten Kopplungsgruppen für *Capsella* dargestellt. Einige der verwendeten Marker (ESTs und cDNAs) (ATTS0636; ATTS2243; 5760; cDNAJ; 5149; 17698; ATTS2212a; ATTS3374a; 3486), die in *Arabidopsis* bisher genetisch nicht kartiert wurden, sind an Positionen kartierter Marker angegeben, die physikalisch eng benachbart sind. Marker mi233 und 11177 zeigten keine Hybridisierung in *C. rubella* (Abb. 3.2).

Marker	Anzahl der ausgewerteten Individuen	<i>C. grandiflora</i> homozygote Individuen	Heterozygote Individuen	<i>C. rubella</i> homozygote Individuen	χ^2 -Test	Anzahl der <i>C. grandiflora</i> -Allele	Anzahl der <i>C. rubella</i> -Allele	χ^2 -Test
mi30	48	5	34	9	0,01	44	52	0,41
mi51	48	7	32	9	0,06	46	50	0,68
mi122	50	7	33	10	0,06	47	53	0,55
mi123	49	10	29	10	0,44	49	49	1,00
mi232	49	11	28	10	0,59	50	48	0,84
mi306	50	7	35	8	0,02	49	51	0,84
mi330a	50	10	31	9	0,23	51	49	0,84
mi330b	50	14	25	11	0,84	53	47	0,55
mi369	50	11	30	9	0,34	52	48	0,69
mi390	47	9	24	14	0,58	42	52	0,30
mi431	50	11	28	11	0,70	50	50	1,00
m211A	47	13	31	3	0,01	57	37	0,04
m235A	47	10	29	8	0,25	49	45	0,68
m249A	47	13	24	10	0,82	50	44	0,54
m254A	47	13	26	8	0,45	52	42	0,30
m315Ba	50	14	26	10	0,70	54	46	0,42
m315Bb	49	7	32	10	0,08	46	52	0,54
m326B	47	4	34	9	0,01	42	52	0,30
m448A	49	8	33	8	0,05	49	49	1,00
m457A	48	17	24	7	0,12	58	38	0,04
m518A	45	14	24	7	0,30	52	38	0,14
m557A	47	9	28	10	0,41	46	48	0,84
3486a	45	11	24	10	0,88	46	44	0,83
3486b	50	18	23	9	0,17	59	41	0,07
5149	44	10	24	10	0,83	44	44	1,00
ATTS0636	50	5	35	10	0,01	45	55	0,32
ATTS2212a	50	12	28	10	0,64	52	48	0,69
ATTS2212b	46	13	26	7	0,31	52	40	0,21
ATTS2243	50	9	32	9	0,14	50	50	1,00
ATTS3374a	50	12	29	9	0,44	53	47	0,55
ATTS3374b	49	9	26	14	0,55	44	54	0,31
5760	45	7	30	8	0,08	44	46	0,83
cDNA J	50	9	33	8	0,08	51	49	0,84
21484	50	17	25	8	0,20	59	41	0,07
17698	50	11	27	12	0,84	49	51	0,84

Tabelle 3.6: Die Segregationsdaten und die Werte der χ^2 -Analyse der *Capsella*-Kartierungspopulation sind für alle zur vergleichenden Kartierung eingesetzten Marker aufgeführt (siehe Anhang). Der Signifikanzwert für die χ^2 -Analyse liegt bei 0,05. Die Anzahl der Individuen, die für die jeweiligen Marker ausgewertet wurden, die Zahl der homozygoten und heterozygoten Pflanzen für die Allele von *C. grandiflora* und *C. rubella* und die beobachtete Allelverteilung sind aufgeführt.

3.2 Vergleichende Mikrosyntenieanalyse in *Arabidopsis thaliana* und *Capsella rubella*

Die Genomkolinearität zwischen *A. thaliana* und *C. rubella* wurde in diesem Teil der Arbeit auf molekularer Ebene untersucht. Für die Analyse wurde ein 200 Kbp großer Bereich zwischen den genetischen Markern m226 (57,0 cM) und g4539 (57,6 cM) auf dem langen Arm von *A. thaliana* Chromosom IV ausgesucht (Lister und Dean, 1993; http://nasc.nott.ac.uk/new_ri_map.html).

3.2.1 Identifizierung kodierender Sequenzen in der 200 Kbp-Region von *Arabidopsis thaliana*

Zur Identifizierung kodierender Sequenzen in diesem 200 Kbp-Bereich wurden zuerst Fragmente aus *A. thaliana* Cosmid-Klonen (Bancroft *et al.*, 1997), die diese Region repräsentieren, als Sonden für Hybridisierungen einer PRL2 cDNA-Bibliothek von *A. thaliana* (Newman *et al.*, 1994) verwendet (Schmidt *et al.*, 1999). Elf cDNA-Klone, die verschiedene Gene repräsentieren, wurden isoliert (cDNA A-K) (Abbildung 3.6). Die Klone, welche jeweils die längste cDNA-Insertion besaßen, wurden für die Sequenzanalyse eingesetzt (Sequenzen siehe Anhang). Der Vergleich der ermittelten cDNA-Sequenzen mit der genomischen *A. thaliana*-Sequenz (Bevan *et al.*, 1998; Z97341, Z97342) zeigte, daß zehn cDNA-Klone mit Ausnahme von cDNA I eine Sequenzidentität zwischen 97-100% im Vergleich zur genomischen *A. thaliana*-Sequenz haben. Durch einen weiteren Vergleich der ermittelten cDNA-Sequenzen mit den Sequenzen in der *A. thaliana*-Datenbank konnte gezeigt werden, daß in drei Fällen Sequenzhomologien zu bekannten Genen vorliegen. cDNA B kodiert für eine Acyl-CoA-Oxidase (ACX1, AF057044) und cDNA D entspricht dem ATHB-2/HAT-4-Gen von *Arabidopsis thaliana* (Carabelli *et al.*, 1993; X68145; Schena *et al.*, 1993; M90394). Für cDNA I wurde eine hohe Sequenzhomologie zum RPP5-Resistenzgen (U97106) aus *Landsberg erecta* festgestellt (Parker *et al.*, 1997).

Durch einen Vergleich der genomischen *A. thaliana*-Sequenz in dem untersuchten 200 Kbp-Bereich mit den EST-Sequenzdaten (Höfte *et al.*, 1993; Newman *et al.*, 1994) wurden ESTs gefunden, die Sequenzhomologien von 90% oder mehr zur genomischen Sequenz haben. Acht EST-Klone, die nicht homolog zu den bereits ausgewählten elf cDNA-Klonen waren, wurden zur Sequenzanalyse ausgesucht (EST 1-8) (Abbildung 3.6; Tabelle 3.7). Die Analyse der ermittelten Sequenzen zeigte für die acht EST-Klone eine Sequenzidentität zwischen 97-100% zur genomischen *A. thaliana*-Sequenz (Sequenzen siehe Anhang).

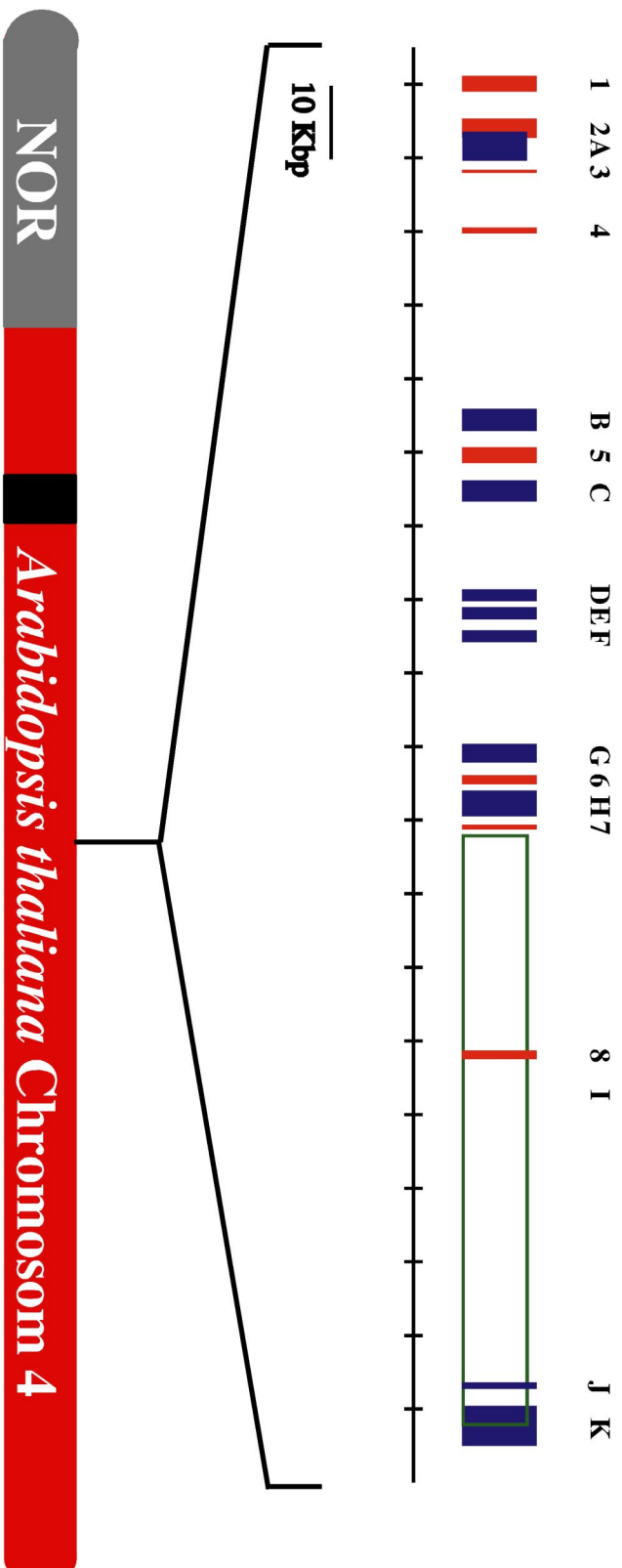


Abbildung 3.6: Gezeigt ist die 200 Kbp-Region, die auf dem langen Arm von *Arabidopsis thaliana* Chromosom IV lokalisiert ist. Elf cDNA-Klone (A-K) wurden durch Hybridisierung einer PRL2 *A. thaliana* cDNA-Bibliothek (Newman *et al.*, 1994) identifiziert (Schmidt *et al.*, 1999). Mehrere Bereiche zeigen Sequenzhomologien zu cDNA I. Die Region ist in der Abbildung in Form eines Kästchens mit grüner Umrandung gezeichnet. Weitere acht cDNA-Klone (1-8) konnten durch Vergleich der genomischen *A. thaliana* Sequenz mit bekannten *Arabidopsis* EST-Sequenzen für die Mikrosyntenie-Analysen ausgesucht werden (Höfte *et al.*, 1993; Newman *et al.*, 1994). Alle 19 identifizierten cDNA-Klone wurden für die weiteren Analysen sequenziert (Sequenzen im Anhang).
 NOR : Nucleolus organisierende Region

Auch hier wurden in der Datenbank für einen EST hohe Homologien zu bereits bekannten Genen gefunden. EST 4 zeigt Homologien zum ribosomalen Protein L15 aus *Petunia hybrida* (AF088912).

Gen	EST-Bezeichnung in der Datenbank	Akzessionsnummern der EST-Sequenzen
EST 1	13004	R30399
EST 2	26868 5760	AA395085 T42497
EST 3	ATTS1135 (VBVYE05)	Z25670
EST 4	ATTS1638 (VCVCB12) 25012	Z26547 AA042756
EST 5	28424 4442	AA585775 T22434
EST 6	12918 34356	R30313 AI099630
EST 7	4420	T22412
EST 8	11177	T76399

Tabelle 3.7: Die acht EST-Klone sind mit den jeweiligen Akzessionsnummern und der Klonbezeichnung aus der *A. thaliana*-Datenbank gezeigt.

3.2.1.1 Identifizierung der homologen Cosmid-Klone im *Capsella rubella*-Genom

Zur Identifizierung von *Capsella rubella* Cosmid-Klonen, welche die homologen Sequenzen zu den 19 cDNA-Sequenzen aus *A. thaliana* besitzen, wurden die cDNA-Klone (cDNA A-K, EST 1-8; Abbildung 3.6) als Sonden in Koloniehybridisierungen mit der *Capsella rubella*-Cosmidbibliothek eingesetzt (Schmidt *et al.*, 1999). In Abbildung 3.7 ist exemplarisch das Ergebnis einer Koloniehybridisierung mit cDNA B als Sonde gezeigt.

Für 17 der 19 cDNA-Sequenzen wurden *C. rubella* Cosmid-Klone identifiziert, die Sequenzhomologie zu den eingesetzten Sonden zeigten. Dies konnte durch Southern-Blot-Analysen bestätigt werden. Das Ergebnis weist auf einen hohen Konservierungsgrad des Genrepertoires der beiden Pflanzen *A. thaliana* und *C. rubella* hin. Für EST 8 (11177) wurde bereits in genetischen Kartierungsexperimenten gezeigt, daß das entsprechende Gen in *C. rubella* nicht vorliegt (Abbildung 3.2). Weitere Hybridisierungsexperimente mit Gesamt-DNA von *C. rubella* und *A. thaliana* und cDNA K als Sonde zeigten, daß auch dieses Gen in *C. rubella* fehlt (Abbildung 3.8). Beide cDNAs hybridisieren mit Gesamt-DNA von *A. thaliana*, jedoch nicht mit *C. rubella*-DNA. Während cDNA K in beiden DNA-Spaltungen von *A. thaliana* mit drei DNA-Fragmenten hybridisiert, zeigt EST 8 ein repetitives Hybridisierungsmuster mit *A. thaliana*-DNA (Abbildung 3.2).

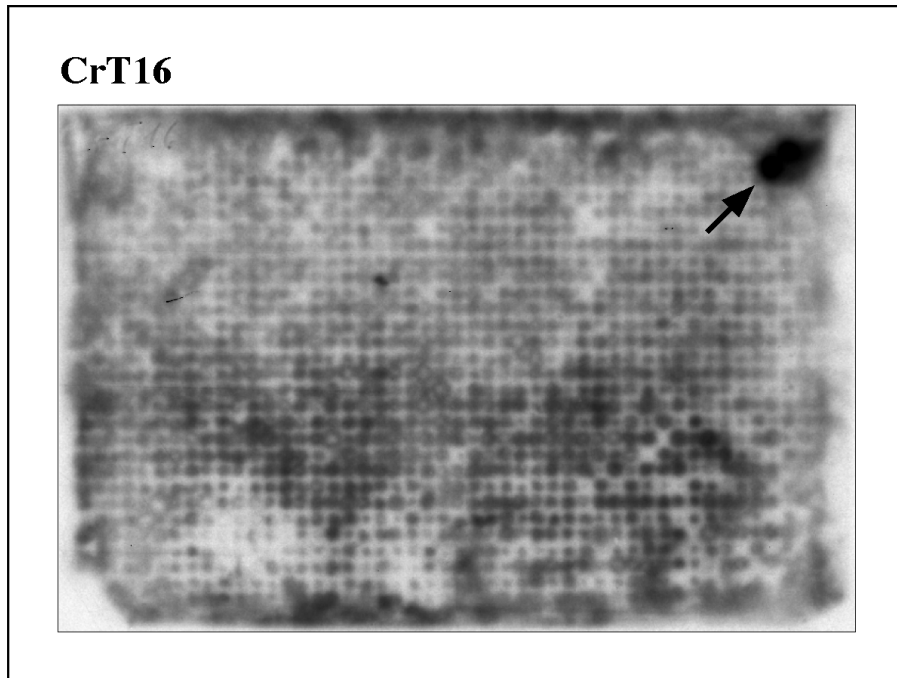


Abbildung 3.7: Das Ergebnis einer Koloniehybridisierung mit der genomischen *Capsella rubella* Cosmid-Bibliothek ist gezeigt. Die Membran hat die Klone der Mikrotiterplatten 31 und 32 der *TaqI*-Klonierung gebunden. Gen B wurde als radioaktiv markierte Sonde eingesetzt. Da jeder Cosmidklon bei der Herstellung der Koloniemembranen doppelt gestempelt wurde, hybridisieren zwei Kolonien mit der verwendeten Sonde, diese entsprechen dem Klon mit den Koordinaten B23 der Mikrotiterplatte 32. Dieser Klon wurde in den späteren Analysen als Cosmid 44 bezeichnet.

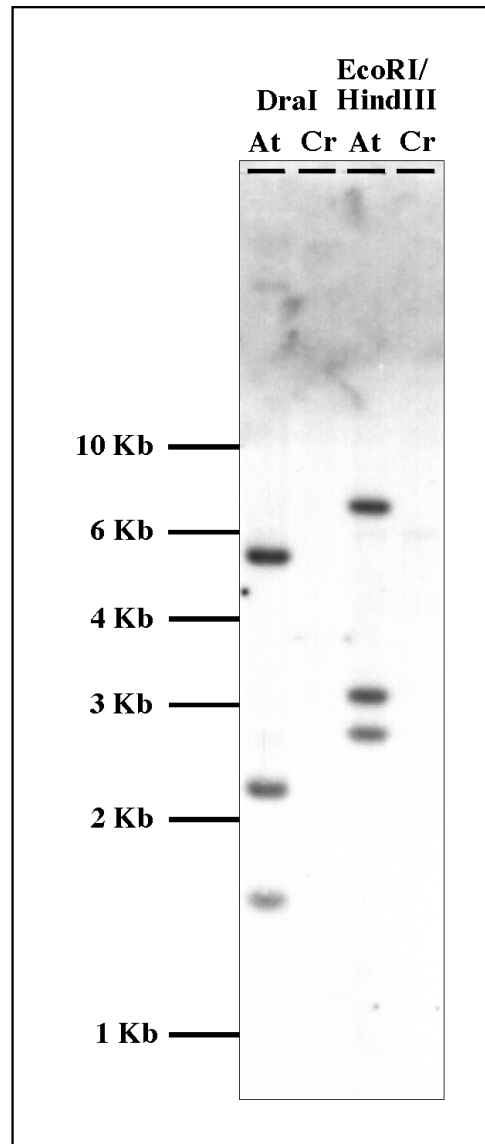


Abbildung 3.8: Das Hybridisierungsergebnis von cDNA K mit Gesamt-DNA von *A. thaliana* und *C. rubella* ist gezeigt. Die Gesamt-DNA in den ersten beiden Spuren wurde mit *DraI* gespalten und in den letzten beiden Spuren mit *EcoRI/HindIII*. Während die Sonde in *A. thaliana* mit mehreren DNA-Fragmenten hybridisiert, ist kein Hybridisierungssignal ähnlicher Stärke für *C. rubella*-DNA zu sehen. Bei dem ca. 6 Kbp großen *EcoRI/HindIII* *C. rubella* DNA-Fragment handelt es sich um Kreuzhybridisierung. Am linken Bildrand ist zur Orientierung der Längenstandard angegeben.

EST 8 korrespondiert zur LTR-Region (*long terminal repeat*) eines Retrotransposon-ähnlichen Elementes in *A. thaliana* (<http://websvr.mips.biochem.mpg.de/proj/thal/>). In der Regel liegt zwischen verschiedenen Spezies kein hoher Konservierungsgrad für repetitive Elemente vor. Dies steht in Übereinstimmung mit dem hier gezeigten Fehlen der EST 8-Sequenz in *Capsella rubella*.

Viele der identifizierten *C. rubella*-Cosmide hybridisieren mit mehreren (2 bis 4) cDNAs. Für Cosmid-Klon 3-II konnten sogar Homologien zu fünf verschiedenen cDNAs (cDNAs G-I) festgestellt werden (Tabelle 3.8; Abbildung 3.9).

Für cDNA E konnte eine große Anzahl von *C. rubella*-Cosmiden identifiziert werden, die nur mit diesem cDNA-Klon hybridisierende DNA-Fragmente besaßen, jedoch keine Homologien mit den unmittelbar benachbart lokalisierten cDNAs aufwiesen. Lediglich für die Cosmide 3-I, 5, 17, 54-I, 98 und 110 konnten Hybridisierungen zu cDNA D, die in *A. thaliana* in enger Nachbarschaft zu Gen E vorliegt, festgestellt werden. Die Cosmide 17, 54-I, 98 und 110 korrespondierten außerdem noch mit Gen F (Tabelle 3.8; Abbildung 3.9). Die anderen Cosmide, die mit cDNA E in *C. rubella* identifiziert wurden, repräsentieren offensichtlich nicht die orthologe Region zu dem untersuchten *A. thaliana*-Bereich.

Die identifizierten *C. rubella* Cosmid-Klone konnten mit Hilfe der ermittelten Southern-Blot Daten zueinander angeordnet und zu zwei Bereichen („Contigs = *contiguous regions*“) zusammengefaßt werden (Tabelle 3.8; Abbildung 3.9).

cDNA/EST	Cosmide, die mit den entsprechenden cDNAs und ESTs aus <i>A. thaliana</i> hybridisieren
1	2, 6, 7-I, 18, 34
2	2, 6, 7-I, 14-I, 18, 34
A	2, 6, 7-I, 14-I, 18, 34
3	14-I, 34
4	12
B	7-II, 44,
5	7-II, 44, 47, 3-I, 5, 49-I
C	7-II, 44, 47, 3-I, 5, 49-I
D	3-I, 5, 49-I, 17, 54-I, 98, 110
E	3-I, 5, 49-I, 17, 54-I, 98, 110
F	17, 54-I, 98, 110
G	49-II, 3-II
6	49-II, 3-II
H	49-II, 3-II, 7-III, 15, 54-II, 63
7	3-II, 7-III, 15, 54-II, 63
I	3-II, 7-III, 15, 54-II, 63
J	7-III, 15, 25, 54-II, 63

Tabelle 3.8: Die Southern-Blot-Hybridisierungsergebnisse der identifizierten *C. rubella* Cosmid-Klone mit den 17 *A. thaliana* cDNAs (A-J, 1-7) sind zusammengefaßt.

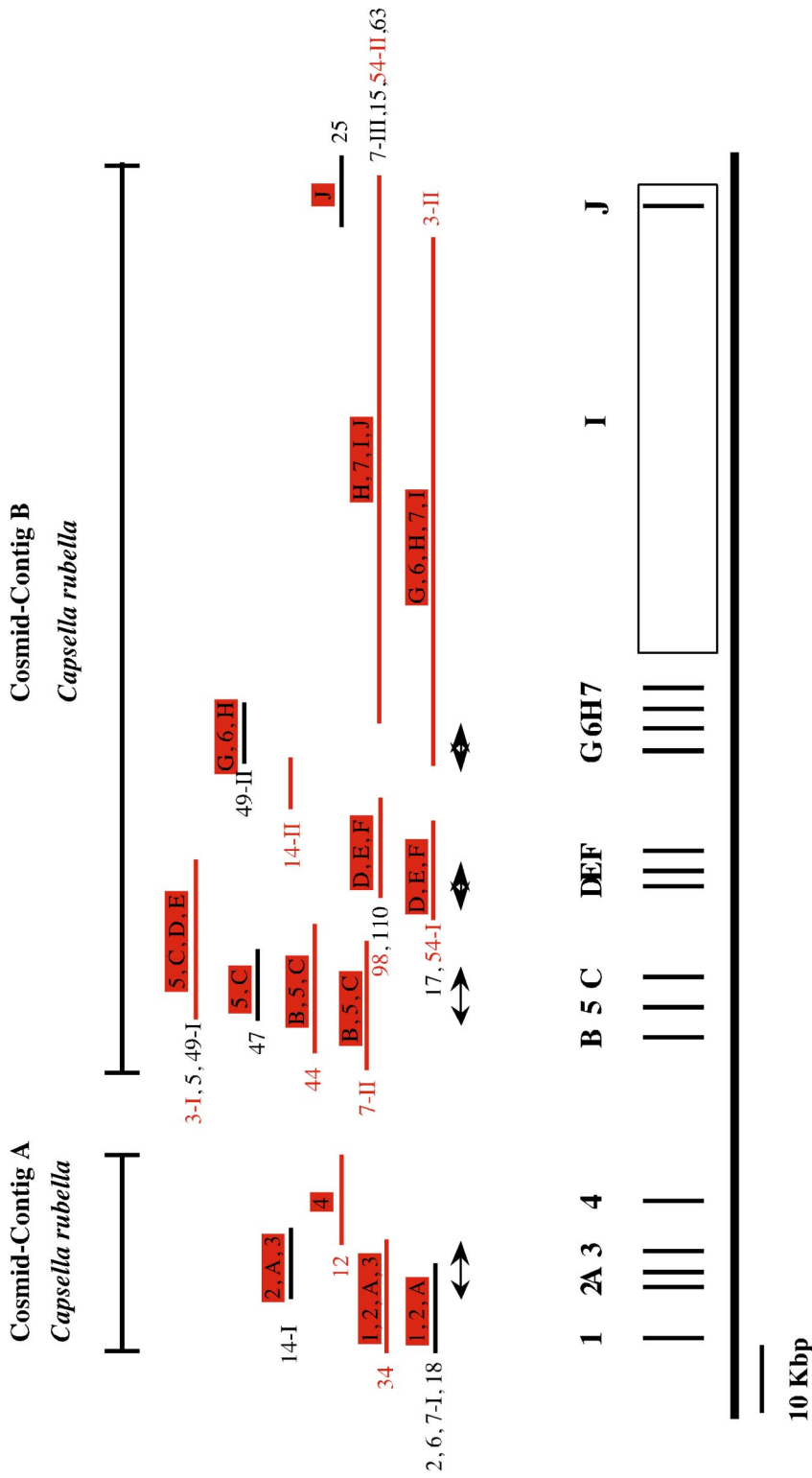


Abbildung 3.9: In der unteren Bildhälfte ist die 200 Kbp *Arabidopsis thaliana* Region gezeigt. Die Lage der cDNAs A-J und 1-7 in *C. rubella* wurde in Form von vertikalen Balken gezeichnet. Im Fall von cDNA I wurden die multiplen Kopien mit Hilfe eines Rechtecks dargestellt. Die *Arabidopsis thaliana* cDNA-Klone wurden als Sonden zur Identifizierung der homologen *Capsella rubella* Cosmid-Klone eingesetzt. Die *C. rubella* Cosmids sind als horizontale Linien dargestellt. Die Cosmids sind in zwei Cosmid-Contigs zusammengefasst. Durch Pfeile sind Gene gekennzeichnet, deren Reihenfolge in *Capsella rubella* nicht eindeutig bestimmt werden konnte. In rot gezeichnete Cosmids wurden für die Subklonierungs- und Sequenzierungsexperimente eingesetzt. Die genaue Position der Cosmid-Enden sowie die genaue Größe der Cosmids wurde nicht ermittelt, daher wurden die Cosmids in dieser Abbildung nicht maßstabsgerecht gezeichnet. In roten Kästchen mit schwarzer Schrift wurden cDNAs angegeben, die mit den jeweiligen *C. rubella*-Cosmids hybridisierende DNA-Fragmente zeigten.

Für einige Gene war es möglich, durch Zusammenstellung der Ergebnisse aller Hybridisierungsexperimente die Reihenfolge der Gene in *C. rubella* zu ermitteln. Gene, deren Reihenfolge in *C. rubella* nicht eindeutig bestimmt werden konnte, sind in Abbildung 3.9 durch Pfeile markiert. Diese Analyse zeigte, daß die Gene in *Arabidopsis thaliana* und *Capsella rubella* in sehr ähnlicher Reihenfolge angeordnet sind.

Exemplarisch sind die Southern-Blot Ergebnisse für die drei *A. thaliana*-cDNAs A, D und J in Abbildung 3.10 gezeigt. cDNA D hybridisiert mit DNA-Fragmenten der Cosmide 3-I, 5, 49-I, 17, 54-I, 98 und 110, wobei bei den Cosmiden 5, 17, 54-I, 98 und 110 ein ca. 4 Kbp großes DNA-Fragment mit Gen D hybridisiert. In den Cosmiden 3-I und 49-I hybridisieren DNA-Fragmente >20 Kbp. Hier wurde die *C. rubella* DNA, welche mit cDNA D hybridisiert, in der verwendeten *Hind*III-Spaltung nicht vom Cosmidvektor abgespalten.

Für die Erstellung der zwei Cosmid-Contigs wurden auch „Chromosom-walking“-Experimente durchgeführt. Hiermit wurden die Ergebnisse der Southern-Blot-Analysen mit den cDNA-Klonen als Sonden bestätigt und zusätzlich weitere Cosmid-Klone identifiziert. Durch Einsatz von Subklonen aus den Cosmiden 12 und 34 konnte in Southern-Blot-Experimenten bestätigt werden, daß diese Cosmide teilweise überlappen (Daten nicht gezeigt). Diese Analyse war erforderlich, da keines der identifizierten Cosmide sowohl mit EST 3 als auch EST 4 hybridisiert (Abbildung 3.9).

In der Koloniehybridisierung der genomischen *C. rubella*-Bibliothek mit den verwendeten cDNA-Klonen als Sonden konnte kein Cosmid, welches sowohl mit cDNA F als auch cDNA G hybridisiert, identifiziert werden. Durch Verwendung von Subklonen aus den Cosmiden 98 und 3-II wurde in einer weiteren Koloniehybridisierung der *C. rubella*-Bibliothek Cosmid 14-II identifiziert. Mit Hilfe von Southern-Blot-Experimenten wurde bestätigt, daß Cosmid 14-II weder mit Gen F noch mit Gen G hybridisiert. In diesem Cosmid-Klon ist also ausschließlich der intergenische Bereich zwischen Gen F und G vertreten. Zur Bestätigung, daß Cosmid 14-II überlappende Bereiche mit den Cosmiden 98 und 3-II hat, wurden Cosmid 14-II und weitere Cosmide, die in der Region vorliegen, mit Sonden aus den Cosmiden 98 und 3-II hybridisiert. In Abbildung 3.11 ist schematisch die Lage der verwendeten Subklone gezeigt.

In Abbildung 3.12 sind die Ergebnisse dieser Analyse dargestellt. Sonde 98-1 hybridisiert mit einem 2 Kbp-Fragment in Cosmid 14-II und in Cosmid 98 mit einem DNA-Fragment >20Kbp. Hier wurde der DNA-Abschnitt, der mit der Sonde hybridisiert, mit der verwendeten *Hind*III-Restriktionsspaltung vom ca. 20 Kbp großen Cosmidvektor nicht abgespalten (Abbildung 3.12A). Mit Sonde 3-3 hybridisieren die Cosmide 14-II, 49-II und 3-II (Abbildung 3.12B). Auch hier hybridisiert die Sonde in den Cosmiden 14-II und 3-II zu DNA-Abschnitten, die vom Cosmidvektor nicht abgespalten wurden. Bei dem markierten DNA-Fragment in Cosmid 98 handelt es sich um eine unspezifische Kreuzhybridisierung mit der verwendeten Sonde.

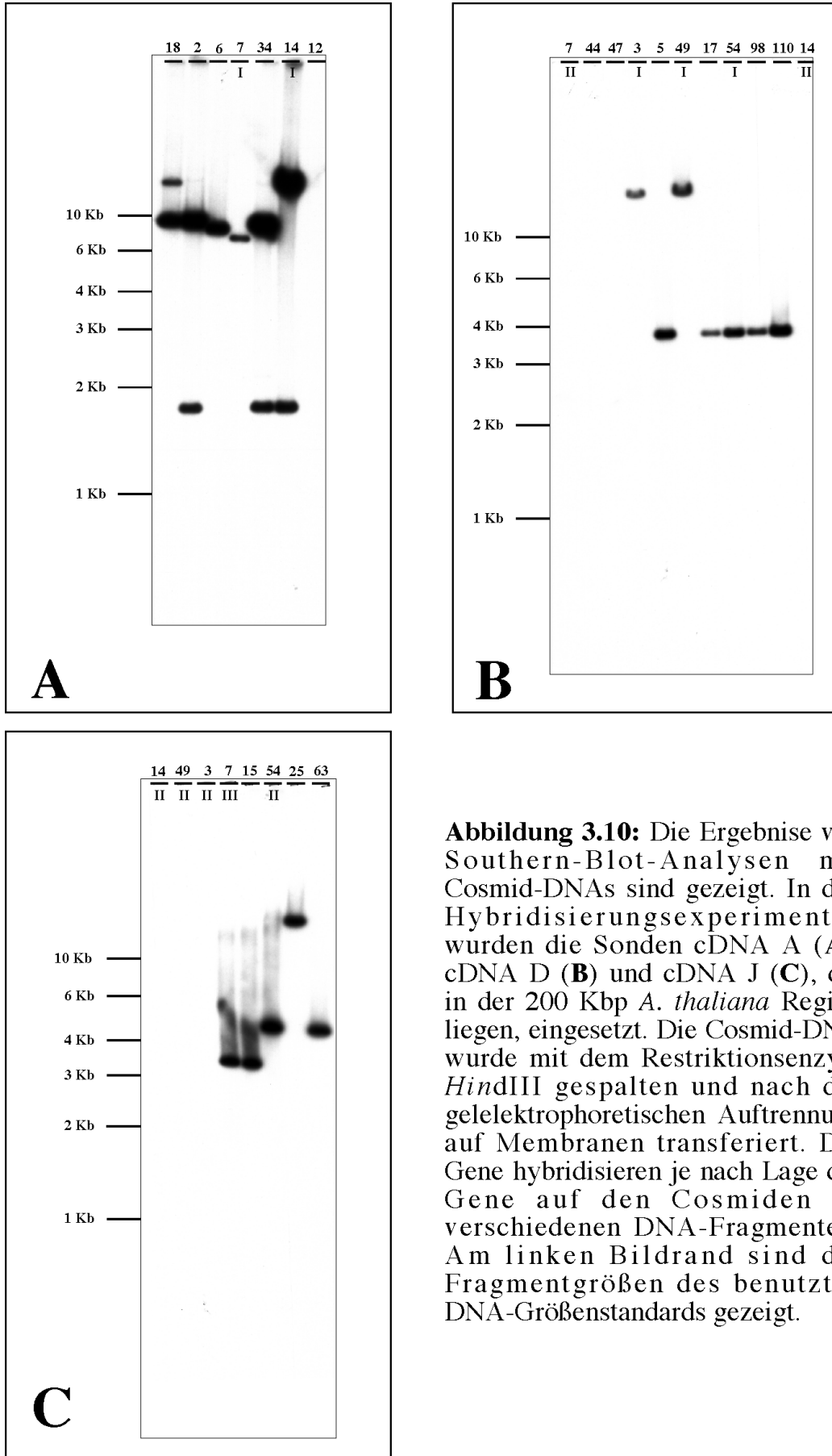


Abbildung 3.10: Die Ergebnisse von Southern-Blot-Analysen mit Cosmid-DNAs sind gezeigt. In den Hybridisierungsexperimenten wurden die Sonden cDNA A (**A**), cDNA D (**B**) und cDNA J (**C**), die in der 200 Kbp *A. thaliana* Region liegen, eingesetzt. Die Cosmid-DNA wurde mit dem Restriktionsenzym *Hind*III gespalten und nach der gelelektrophoretischen Auftrennung auf Membranen transferiert. Die Gene hybridisieren je nach Lage der Gene auf den Cosmiden zu verschiedenen DNA-Fragmenten. Am linken Bildrand sind die Fragmentgrößen des benutzten DNA-Größenstandards gezeigt.

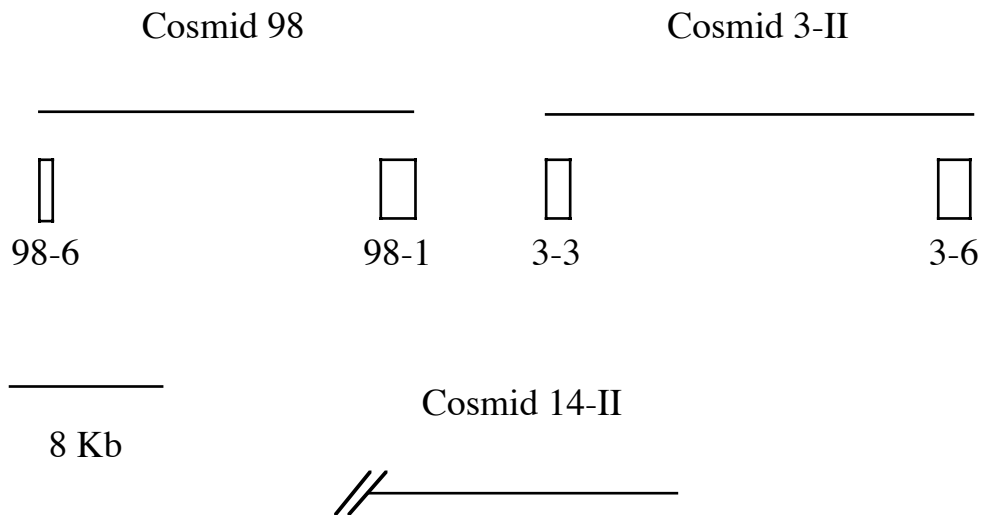


Abbildung 3.11: Schematische Darstellung der Lage der Subklone 3-3, 3-6, 98-1 und 98-6 in den Cosmiden 3-II und 98.

In Abbildung 3.12C und 3.12D sind die Ergebnisse der Hybridisierungen mit den Sonden 98-6 und 3-6 gezeigt. Beide Subklone zeigen kein Hybridisierungssignal mit Cosmid 14-II. Probe 98-6 hybridisiert mit den Cosmiden 3-I, 54-I und 98 und repräsentiert den Cosmidbereich von 98, welcher in der Nähe von cDNA D liegt (Abbildung 3.12C). Sonde 3-3 hybridisiert nur mit Cosmid 3-II (Abbildung 3.12D).

In der Region zwischen Genen 4 und B konnten durch den Einsatz von Subklonen aus den Cosmiden 12 und 7 in der verwendeten genomischen *C. rubella*-Bibliothek keine Cosmid-Klone identifiziert werden, die sowohl mit Cosmid 12 aus Contig A als auch Cosmid 7 aus der Contig B-Region hybridisieren (Abbildung 3.9; Analyse siehe unten).

3.2.1.2 Sequenzierung der beiden *Capsella rubella* Cosmid-„Contigs“

Für die Syntenie-Analyse von *A. thaliana* und *C. rubella* auf Sequenzebene wurden Cosmide aus beiden *C. rubella* Contigs subkloniert und anschließend sequenziert.

Die Cosmide 12 und 34 aus Cosmid-Contig A bzw. Cosmide 7-II, 44, 3-I, 54-I, 98, 14-II, 3-II, 54-II aus Contig B (Abbildung 3.9) wurden hierfür mit geeigneten Restriktionsenzymen gespalten und in den Plasmidvektor pGEM7Z (Promega) kloniert (Tabelle 3.9).

Durch Hybridisierungen mit cDNAs als Sonden wurde festgestellt, welche Subklone kodierende Sequenzen tragen. Cosmid-DNAs wurden ebenfalls als Sonden eingesetzt, um auch diejenigen Subklone zu identifizieren, die den intergenischen Bereichen entsprechen.

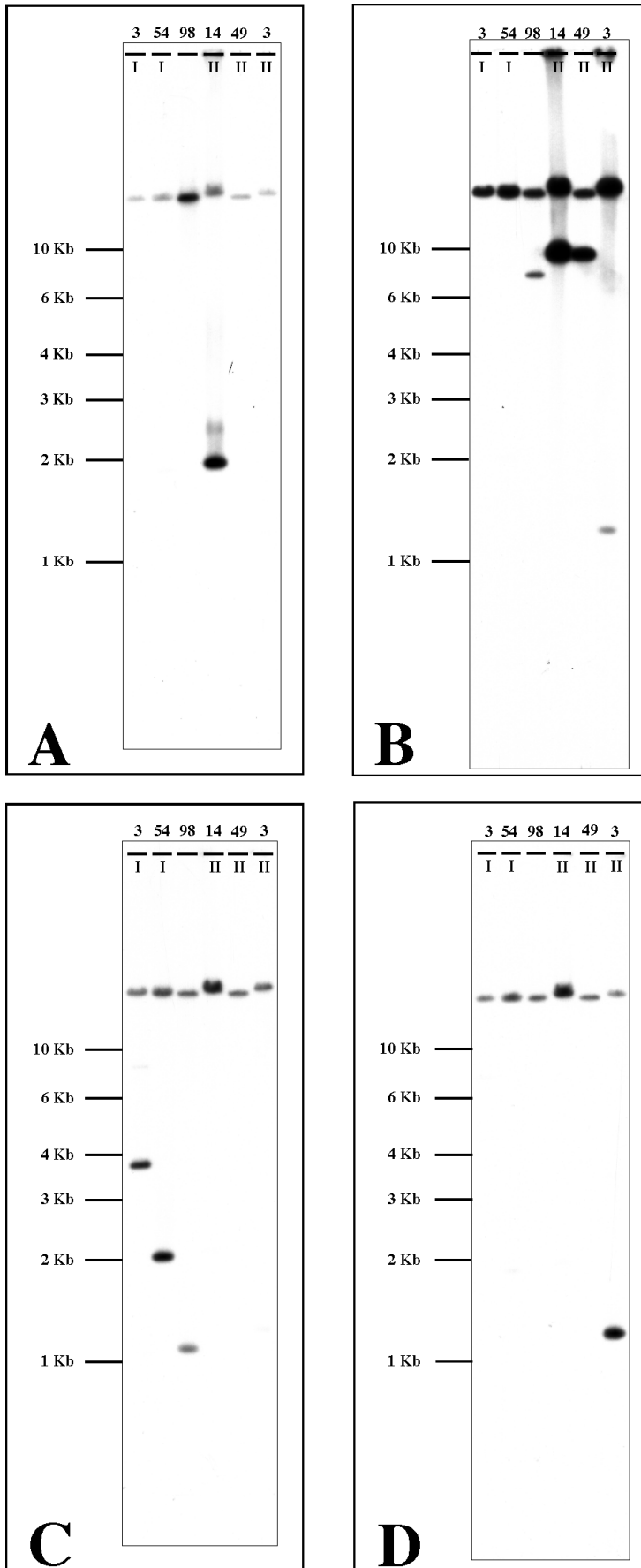


Abbildung 3.12: Die Ergebnisse der Experimente in denen Cosmide, die in Nachbarschaft von Cosmid 14-II lokalisiert sind, mit Subklonen aus Cosmid 98 und 3-II hybridisiert wurden, sind abgebildet. Bild **A**: Sonde 98-1, Bild **B**: Sonde 3-3, Bild **C**: Sonde 98-6 und Bild **D**: Sonde 3-6. Die Cosmide wurden mit *Hind*III gespalten und nach der gelelektrophoretischen Auftrennung auf Membranen transferiert.

Zur Subklonierung eingesetzte Cosmid-Klone	Zur Subklonierung der Cosmide verwendete Restriktionsenzyme
12	<i>EcoRI</i> (E) <i>SacI</i> (S) <i>EcoRI/XhoI</i> (EX)
34	<i>EcoRI</i> (E) <i>SacI</i> (S) <i>HindIII</i> (H)
7-II	<i>EcoRV</i> (V) <i>HindIII</i> (H) <i>SacI/XhoI</i> (SX)
44	<i>EcoRI</i> (E) <i>HindIII</i> (H) <i>EcoRI/SacI</i> (ES)
3-I	<i>EcoRI</i> (E) <i>HindIII</i> (H) <i>EcoRI/XbaI/XhoI</i> (EXX)
54-I	<i>EcoRI</i> (E) <i>EcoRI/XbaI/XhoI</i> (EXX)
98	<i>HindIII</i> (H) <i>HindIII/XhoI</i> (HX) <i>EcoRI/XhoI</i> (EX)
14-II	<i>EcoRI</i> (E) <i>HindIII/XhoI</i> (HX)
3-II	<i>EcoRI</i> (E) <i>SacI</i> (S) <i>EcoRI/XbaI/XhoI</i> (EXX)
54-II	<i>EcoRI</i> (E) <i>HindIII</i> (H) <i>SacI</i> (S) <i>EcoRI/XbaI/XhoI</i> (EXX) <i>EcoRI/XhoI</i> (EX)

Tabelle 3.9: Die Cosmidklone aus den Contigs A und B sind mit den entsprechenden Restriktionsenzymen aufgelistet, die für die Subklonierung der Cosmide verwendet wurden. In Klammern sind die Abkürzungen der Restriktionsenzyme angegeben, die für die Bezeichnung der entsprechenden Subklone verwendet wurden. Bei Verwendung eines Restriktionsenzym für die Spaltung der Cosmid-DNA wurden die DNA-Fragmente in die entsprechende Restriktionsschnittstelle des Plasmidvektors kloniert. Wurde die Cosmid-DNA mit mehreren Restriktionsenzymen gleichzeitig behandelt, wurden die DNA-Fragmente nach Auffüllen der überhängenden Fragment-Enden in die *SmaI*-Schnittstelle des Vektors kloniert. Eine Liste aller sequenzierten Subklone ist im Anhang zu finden.

Alle unterschiedlichen Subklone eines Cosmids wurden zunächst mit den Oligonukleotiden "Universe" und "Reverse" sequenziert (2.3). Bei Subklonen mit größeren Insertionen von genomischer *C. rubella*-DNA (>1 Kbp) wurden für die weiteren Sequenzierungen geeignete Oligonukleotidsequenzen aus den erhaltenen Sequenzierungsdaten ausgewählt.

Bei der Zusammenstellung der Einzelsequenzdaten für die jeweiligen Subklone wurde festgestellt, daß mit den verwendeten Restriktionsenzym-Kombinationen in einigen Bereichen der Cosmide keine überlappenden Sequenzen zwischen den einzelnen Subklonen eines Cosmids erhalten wurde. In diesen Fällen wurde die fehlende Sequenz

zwischen den benachbarten Subklonen mittels PCR amplifiziert und anschließend sequenziert (siehe Anhang).

In Abbildung 3.13 ist exemplarisch die Contig A-Region, im folgenden mit Contig 0 bezeichnet, mit den jeweiligen Positionen der sequenzierten Subklone und der verwendeten PCR-Produkte angegeben. In Abbildung 3.14 ist das Ergebnis der Sequenzierung der *C. rubella* Cosmide zusammengefaßt.

Aufgrund der Ergebnisse der Hybridisierungsexperimente mit den Sonden G, 6, H, 7, I und J wurden die Cosmide 49-II, 3-II, 7-III, 15, 54-II, 63 und 25 zu einem Contig zusammengefaßt (Abbildung 3.9). Wie in Abbildung 3.14 gezeigt wird, sind in der Regel die überlappenden Regionen benachbarter Cosmide groß. Eine Ausnahme in dieser Hinsicht stellt der Sequenzbereich dar, der sowohl von Cosmid 3-II als auch 54-II abgedeckt wird. Hier beträgt der überlappende Bereich lediglich 29 Bp. Durch PCR-Analysen auf den *C. rubella* Cosmidklonen 7-III und 25, die ebenfalls diese Region umfassen (Abbildung 3.9), konnte jedoch nicht sichergestellt werden, daß der gesamte Sequenzbereich durch die Cosmide 3-II und 54-II repräsentiert ist. Daher wurden nur die Sequenzen der Cosmide 7-I, 44, 3-I, 54-I, 98, 14-II und 3-II in Contig I zusammengefaßt, während Cosmid 54-II Contig II repräsentiert (Abbildung 3.14).

Die ermittelten genomischen *C. rubella* Sequenzen betragen für Contig 0 36190 Bp, für Contig I 86006 Bp und für Contig II 22005 Bp. Für die Cosmide 44, 54-I, 98 und 14-II wurden lediglich Teile der genomischen Insertion ermittelt. Insgesamt wurden in *Capsella rubella* ca. 144 Kbp sequenziert, die einem Großteil der 200 Kbp großen Region in *Arabidopsis thaliana* entsprechen.

3.2.1.3 Analyse der Sequenzierungsdaten

3.2.1.3.1 Analyse der Region zwischen den Contigs 0 und I

Mit genetischen Kartierungsexperimenten konnte gezeigt werden, daß die beiden Contigs 0 und II eng gekoppelt in *Capsella* vorliegen. In 45 Individuen der Kartierungspopulation zeigen EST 2 (5760) und cDNA J Kosegregation (Abbildung 3.5). Das heißt, daß die Loci maximal 1,1 cM voneinander entfernt liegen.

Mit Hilfe von „Chromosome-walking“-Experimenten konnten für die Region zwischen den Contigs 0 und I keine Cosmid-Klone identifiziert werden, die die zwei Contigs miteinander verbinden bzw. die Contigs vergrößern (Abbildung 3.9).

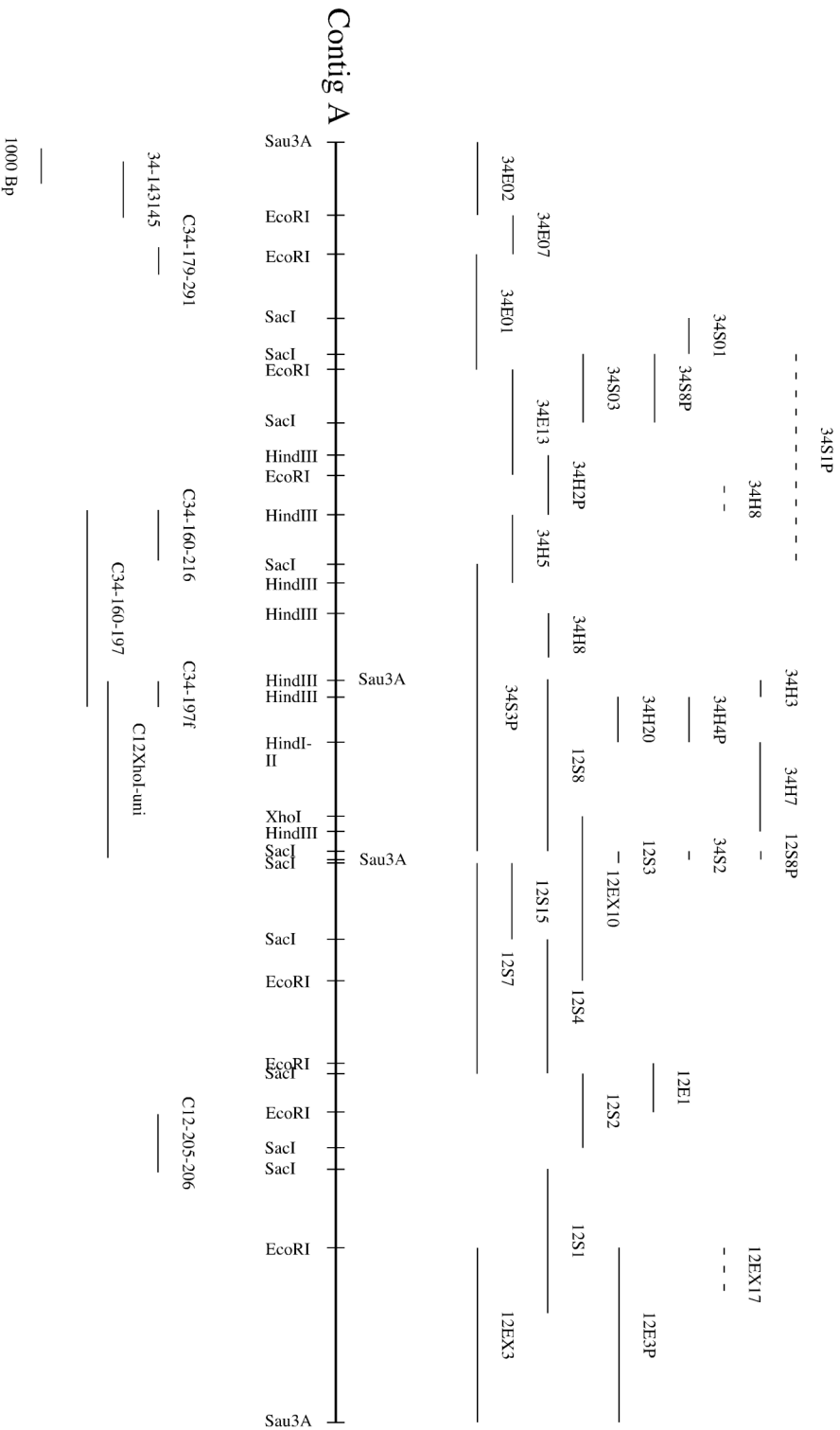
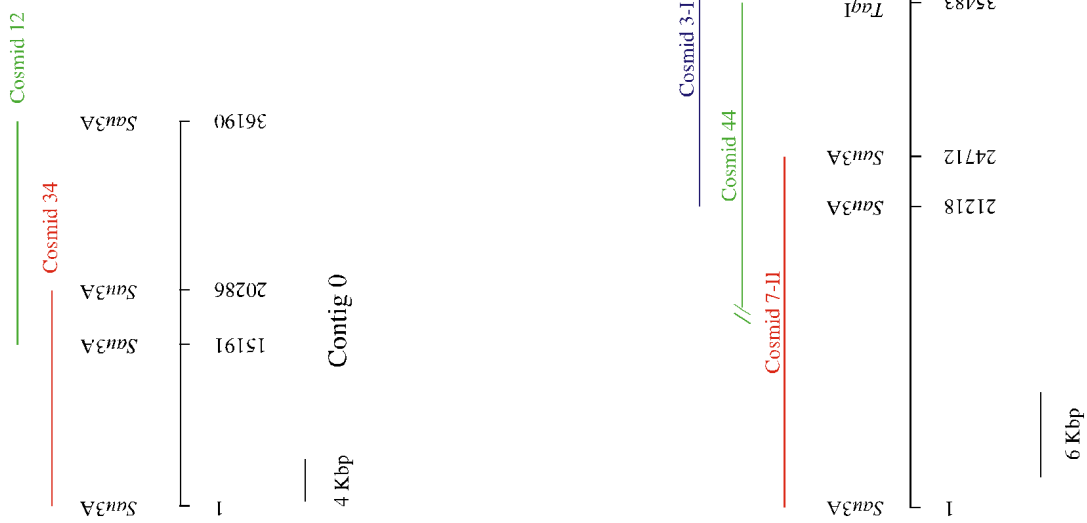


Abbildung 3.13: Die Subklonkarte für die Cosmide 12 und 34 aus Contig A (Contig 0) ist gezeigt. Die Schnittstellen der Restriktionsenzyme in der Contig A-Sequenz sind dargestellt, die für die Subklonierung der Cosmide 12 und 34 benutzt wurden. Außerdem sind die *Sau3A*-Schnittstellen eingezeichnet, die die genomischen Fragmentenden der Cosmide anzeigen. In der oberen Bildhälfte sind die zur Sequenzierung eingesetzten Subklone aus beiden Cosmiden dargestellt. Anhand der Subklonbezeichnung ist erkennbar, welche Restriktionsenzyme für die Subklonierung verwendet wurden und aus welchem Cosmid die Subklone stammen. Diese Bereiche der genomischen Sequenz waren nur teilweise durch die vorliegenden Subklone abgedeckt. In der unteren Bildhälfte sind die PCR-Produkte gezeigt, die zur Sequenzierung eingesetzt wurden. Die PCR-Produkte wurden mit den Namen der Oligonukleotide gekennzeichnet. Die getrichelt abgebildeten Subklone 34S1P, 34H8 und 12EX17 sind chimär.

Abbildung 3.14: Die Positionen der einzelnen *C. rubella*-Cosmide, die zur Sequenzierung dieser Region eingesetzt wurden, sind bezogen zur ermittelten Gesamtsequenz der *C. rubella*-Region abgebildet. Zur Orientierung sind die Positionen der Restriktionsschnittstellen *TagI* und *Sau3A*, die die Insertionen der genomischen *C. rubella*-DNA in den Cosmiden begrenzen, dargestellt. Für die Cosmide 44, 54-I, 98 und 14-II wurden nicht die vollständigen Sequenzen der *C. rubella* DNA Insertionen ermittelt. Dies wurde durch Schrägstriche angegeben.



Zur Bestätigung dieser Ergebnisse wurden der Subklon 12E3P aus Cosmid 12 (Abbildung 3.13) und Subklon 7SX10 aus Cosmid 7 für Hybridisierungen mit Cosmiden aus den Contigs 0 und I eingesetzt. Die Lage der Subklone auf den Cosmiden ist in Abbildung 3.15 dargestellt. Die Subklone hybridisieren nur mit dem Cosmid, aus dem sie stammen (Daten für Subklon 7SX10 nicht gezeigt; Abbildung 3.16). In Cosmid 12 hybridisieren ca. 4 Kbp, 2 Kbp und 1,6 Kbp große DNA-Fragmente mit dem Subklon 12E3P. Aus der Sequenz des Cosmids 12 kann abgeleitet werden, daß auch ein 117 Bp Fragment mit der Sonde nachgewiesen werden sollte (Abbildung 3.15), welches in dem durchgeführten Southern-Blot Experiment jedoch nicht zu erkennen war. Keines der anderen Cosmide aus dem Contig 0- bzw. Contig I-Bereich zeigt eine Hybridisierung mit dem Subklon. Diese Ergebnisse belegen ebenso wie Sequenzvergleiche, daß die beiden Cosmid-Contigs 0 und I keine Überlappung zeigen.

Eine weitere Hybridisierungsanalyse wurde mit PCR-Produkt 288/289, das sich aus Cosmid 7 ableitet, und Subklon 12E3P aus Cosmid 12 durchgeführt. Es sollte ermittelt werden, ob es *C. rubella* DNA-Fragmente gibt, die sowohl mit dem Subklon aus Cosmid 12 als auch mit dem PCR-Produkt aus Cosmid 7 hybridisieren. *C. rubella* Gesamt-DNA wurde jeweils mit den Restriktionsenzymen *EcoRI*, *EcoRV* und *HindIII* gespalten und auf Membranen transferiert. Der Vergleich der in Abbildung 3.18 dargestellten Ergebnisse der Hybridisierungsexperimente zeigt, daß der Subklon und das PCR-Produkt in keinem Fall ein oder mehrere gleiche DNA-Fragmente detektieren. In Abbildung 3.17 sind die genaue Lage der Restriktionsschnittstellen für die drei verwendeten Restriktionsenzyme in der genomischen Sequenz der Contigs 0 und I, die Positionen des verwendeten Subklons sowie des PCR-Produkts und die Mindestgrößen der erwarteten DNA-Fragmente gezeigt. Die Sonden von Subklon 12E3P und PCR-Produkt 288/289 hybridisieren unter anderem mit *C. rubella* Restriktionsfragmenten, die sowohl Cosmidsequenzen als auch dem bisher nicht identifizierten Genom-Abschnitt zwischen Contig 0 und I entsprechen. Die Differenzen der bekannten Mindestgröße dieser Restriktionsfragmente und den tatsächlich ermittelten Fragmentgrößen geben Auskunft über die Position der Restriktionsschnittstellen für *EcoRI*, *EcoRV* und *HindIII* in der genomischen Sequenz von *C. rubella*, die bisher nicht ermittelt wurde. Aus diesen Ergebnissen kann abgeleitet werden, daß der Bereich zwischen Contig 0 und Contig I mindestens 5,5 Kbp umfaßt (Tabelle 3.10).

Durch Vergleich der *A. thaliana* genomischen Sequenz mit den Sequenzdaten für die cDNAs 4 und B (Abbildung 3.9) wurde ermittelt, daß der Bereich zwischen den beiden Genen in *A. thaliana* ca. 24,5 Kbp groß ist. In *C. rubella*-Contig 0 beträgt die Sequenzlänge von cDNA 4 bis zum Contig 0-Ende ca. 6 Kbp und in Contig I beträgt die Sequenzlänge bis zum Gen B ca. 14 Kbp (Tabelle 3.12). In *C. rubella* wurden somit bereits ca. 20 Kbp des intergenischen Bereichs zwischen den Genen 4 und B sequenziert.

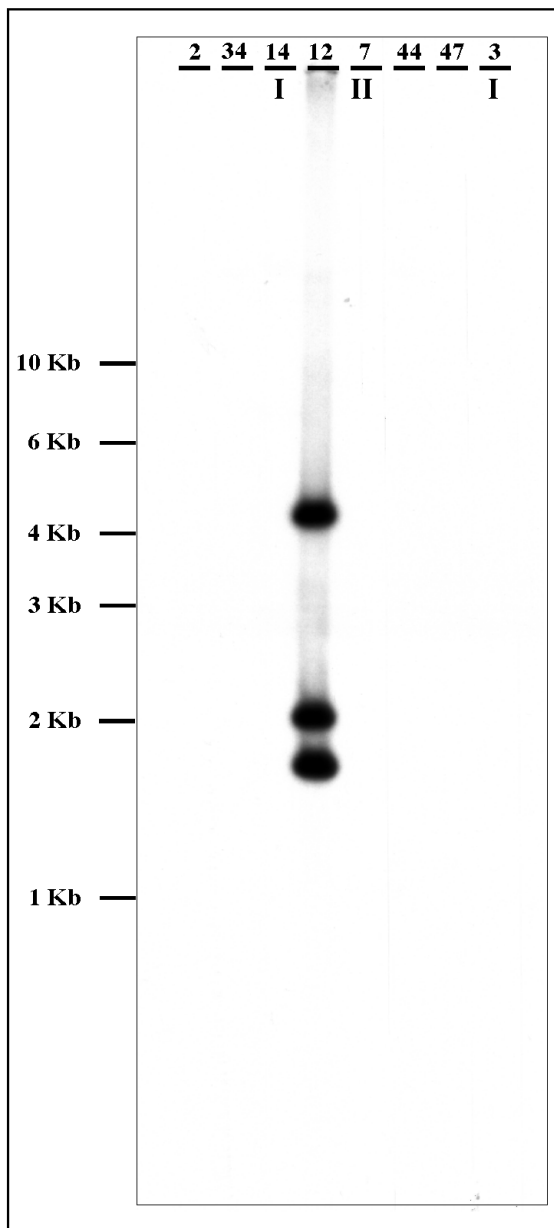
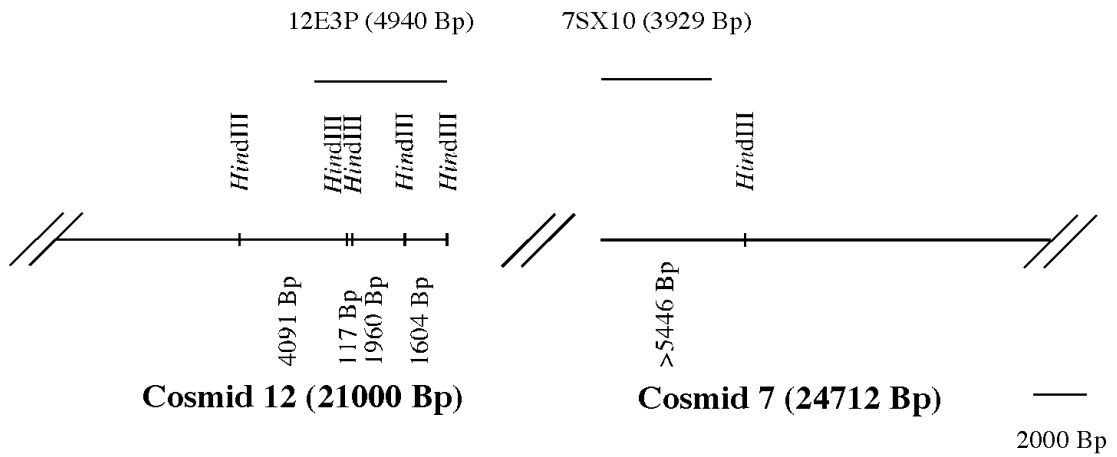


Abbildung 3.15: Schematische Darstellung der Hybridisierungsergebnisse aus Abbildung 3.16.

Die Restriktionsschnittstellen für *Hind*III sind für Cosmid 12 aus Contig 0 bzw. Cosmid 7 aus Contig I für den Teil der Cosmidsequenzen gezeigt, die den Sonden entsprechen.

Im unteren Teil der Abbildung wurden die erwarteten DNA-Fragmentgrößen für die Hybridisierung der *Hind*III-gespaltenen Cosmide mit den Subklonen 7SX10 und 12E3P als Sonden aufgeführt.

Abbildung 3.16: Das Hybridisierungsergebnis mit Subklon 12E3P aus Cosmid 12 als Sonde ist gezeigt. Die Cosmid-DNAs wurden mit *Hind*III gespalten.

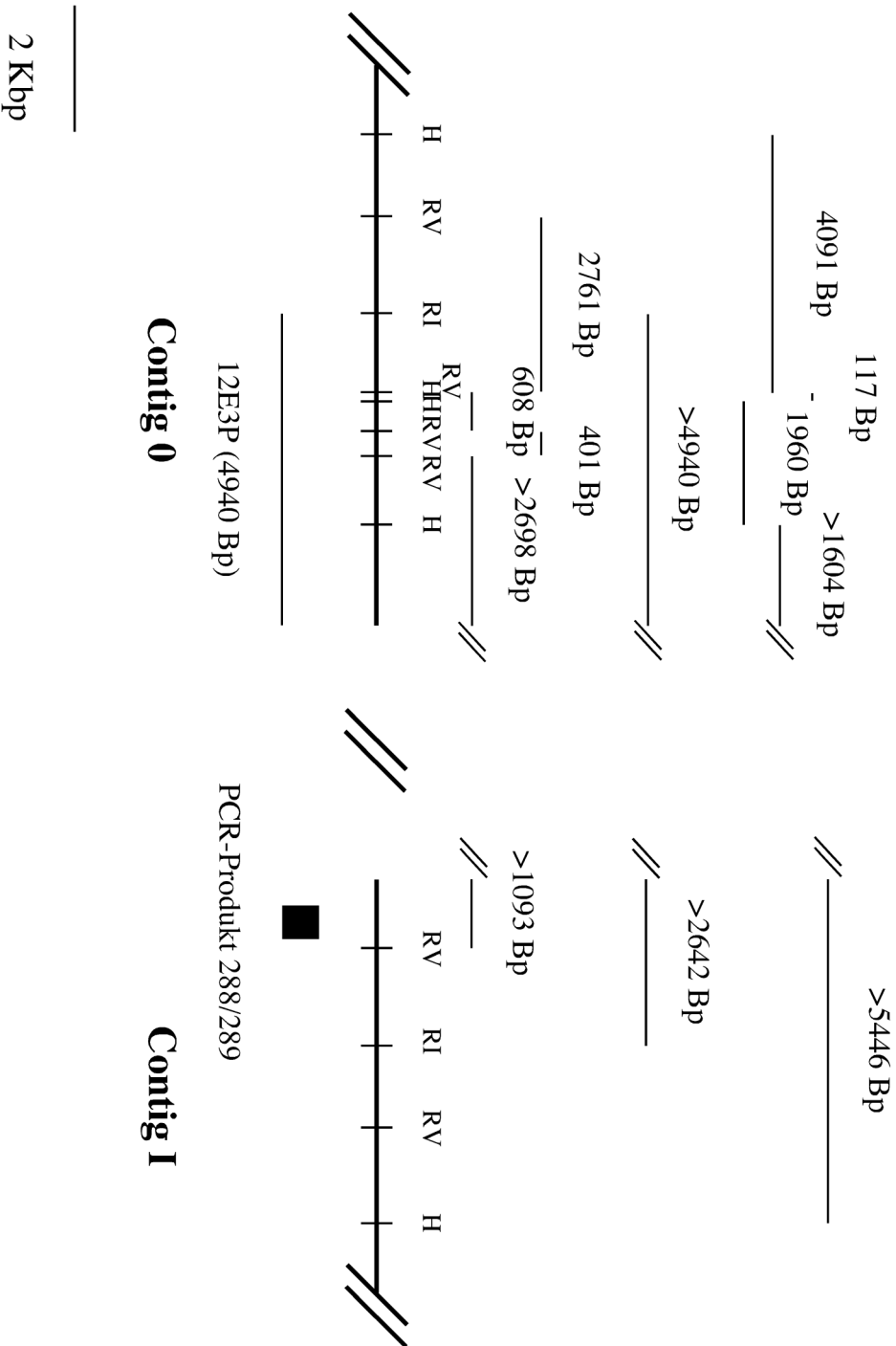


Abbildung 3.17: Im unteren Teil des Schemas ist die Lage der verwendeten Sonden (12E3P und 288/289) gezeigt. Die Position und Größe der erwarteten DNA-Fragmente ist im oberen Teil der Zeichnung angegeben. Für Fragmente bei denen die Restriktionsstelle in der Region zwischen Contig 0 und I nicht bekannt ist, wurde die Mindestgröße angegeben. H: *HindIII*; RI: *EcoRI*; RV: *EcoRV*

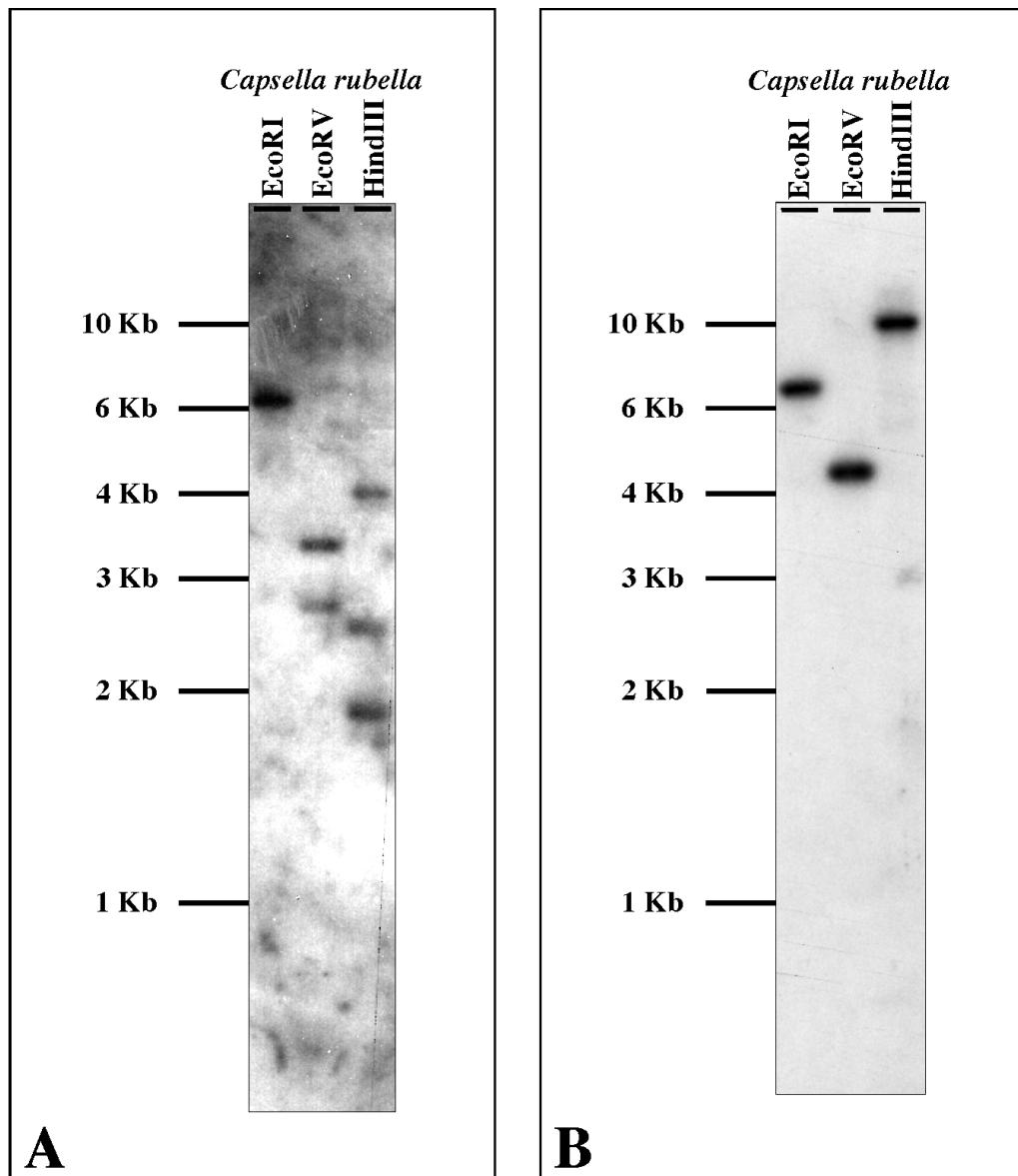


Abbildung 3.18: Die Hybridisierung von gespalteener *C. rubella* Gesamt-DNA mit Subklon 12E3P und einem PCR-Produkt aus Subklon 7SX10 ist abgebildet. Die Gesamt-DNA wurde mit den Restriktionsenzymen *EcoRI*, *EcoRV* und *HindIII* gespalten. In Bild **A** ist das Ergebnis der Hybridisierung gezeigt, in der die Membran mit dem Subklon 12E3P analysiert wurde. Bild **B** zeigt das Ergebnis mit der Sonde aus Subklon 7SX10 (288/289). Der Vergleich der Ergebnisse für die einzelnen Hybridisierungen zeigt, daß die beiden Sonden nicht mit DNA-Fragmenten gleicher Größe hybridisieren.

Da mindestens 5,5 Kbp zwischen den Contigs 0 und I liegen, umfaßt der intergenische Bereich zwischen den Genen 4 und B in *C. rubella* mindestens 25,5 Kbp und ist damit länger als in *A. thaliana*.

	<i>EcoRI</i>		<i>EcoRV</i>		<i>HindIII</i>	
	erwartete Fragmentgrößen	detektierte Fragmentgrößen	erwartete Fragmentgrößen	detektierte Fragmentgrößen	erwartete Fragmentgrößen	detektierte Fragmentgrößen
12E3P	>4,9 Kbp	6,1 Kbp	>2,7 Kbp 2,8 Kbp 0,6 Kbp 0,4 Kbp	3,4 Kbp 2,8 Kbp	>1,6 Kbp 1,9 Kbp 4,1 Kbp 0,1 Kbp	2,5 Kbp 1,9 Kbp 4,1 Kbp
288/289	>2,6 Kbp	6,2 Kbp	>1,1 Kbp	4,2 Kbp	>5,4 Kbp	10 Kbp

Tabelle 3.10: Die Mindestgrößen der erwarteten DNA-Fragmente (Abbildung 3.17) und die im Hybridisierungsexperiment detektierten Banden (Abbildung 3.18) für die drei Spaltungen sind zusammengefaßt.

3.2.1.3.2 cDNAs K und 8 in der genomischen *C. rubella*-Sequenz

Kolonie- und Southern-Blot-Experimente hatten ergeben, daß in den identifizierten Cosmiden keine Homologen zu den cDNAs K und 8 gefunden wurde. Vergleiche der cDNA-Sequenzen K und 8 mit den ermittelten genomischen *C. rubella*-Sequenzen bestätigten, daß die homologen Sequenzen in den untersuchten *C. rubella*-Contigs nicht vorlagen.

3.2.1.4 Vergleich der genomischen Sequenzen von *A. thaliana* und *C. rubella*

Mit der Analyse auf Southern-Blot-Ebene konnte die Anordnung einiger Gene in *C. rubella* nicht genau ermittelt werden (Abbildung 3.9). In der weiteren Analyse wurden die ermittelten *C. rubella*-Sequenzen mit den Sequenzen der *A. thaliana*-cDNAs verglichen, um die genaue Anordnung der Gene auf den *C. rubella* Cosmid-Contigs zu bestimmen.

Die Positionsangaben der Gene in der *A. thaliana*-Region sind bezogen auf einen Sequenzabschnitt, der sich aus Abschnitten der ATFCA6- (Z79742) und ATFCA7-Contigs (Z79743) zusammensetzt. Die Sequenzen von 140001-200952 des ATFCA6-Contigs und 1-140000 des ATFCA7-Contigs entsprechen dabei den Positionen 1-60952 bzw. 56292-196291 im hier verwendeten Contig. Zunächst wurden die ermittelten Sequenzen der verwendeten *A. thaliana* cDNA- und EST-Klone mit vorhergesagten Genen im ATFCA-Contig verglichen (<http://websvr.mips.biochem.mpg.de/proj/thal/>). Damit wurde festgestellt, inwieweit die Genvorhersagen mit den cDNA-Sequenzen für die untersuchten Gene übereinstimmten (Tabelle 3.11).

Für die weiteren Vergleiche mit *C. rubella*-Sequenzen wurden die cDNA-Klone benutzt, wenn die cDNA-Sequenz teilweise keine Übereinstimmung zu den vorhergesagten Gensequenzen zeigten (2, A, 5, C, 6, H, 7, J). Die Genvorhersage At4g16830 und das offene Leseraster der cDNA-Sequenz G sind identisch, auch in diesem Fall wurde die cDNA-Sequenz für die Vergleiche benutzt. In den Fällen 3, 4, E und F wurde die Genvorhersage für die Vergleiche herangezogen, da diese cDNA-Klone nicht vollständig die 5'-kodierenden Sequenzen der Gene umfassen. cDNA 1 zeigt Übereinstimmung zur Sequenz der Genvorhersage, jedoch sind sowohl die Sequenz der Genvorhersage als auch die cDNA-Sequenz nicht vollständig. In diesem Fall wurde die cDNA-Sequenz weiterverwendet.

Ein Vergleich der ermittelten cDNA-Sequenzen mit den genomischen *A. thaliana* Sequenzen zeigte, daß 16 der 17 untersuchten cDNA-Klone Sequenzidentitäten zwischen 97-100% mit der genomischen Sequenz aufwiesen. Lediglich cDNA I zeigte geringere Sequenzidentitäten. Daher wurde die Sequenz von cDNA I für die Vergleiche nicht verwendet. Die stattdessen aus der Datenbank bezogene RPP5-Gensequenz stammt aus dem *A. thaliana* Ökotyp Landsberg *erecta* (U97106, Parker *et al.*, 1997). Auch für Gen D wurde eine Sequenz aus der Datenbank für die Vergleiche herangezogen (Athb-2, X68145), da der charakterisierte cDNA-Klon D einige Sequenz-Unterschiede zur genomischen *A. thaliana*-Sequenz aufweist.

cDNA	Exon-Sequenzen im vorhergesagten Gen und cDNA stimmen überein	Sequenz, die für die vergleichenden Analysen verwendet wurde
1	+	cDNA 13004
2	-	cDNA 5760
A	-	cDNA A
3	+	At4g16690
4	+	At4g16720
B	-	ACX1-cDNA, AF057044
5	-	cDNA 4442
C	-	cDNA C
D	+	Athb-2-cDNA, X68145
E	+	At4g16790
F	+	At4g16800
G	+	cDNA G
6	-	cDNA 12918
H	-	cDNA H
7	-	cDNA 4420
I	-	RPP5-cDNA, U97106
J	-	cDNA J, cDNA ATTS1159

Tabelle 3.11: Die Daten aus dem Vergleich der cDNA-Sequenzen mit den vorhergesagten Genen in der Datenbank (<http://websvr.mips.biochem.mpg.de/proj/thal/>) für die ATFCA-Region sind zusammengefaßt. In der zweiten Spalte der Tabelle wurde angegeben, welche Genvorhersage- und cDNA-Sequenzen in den Exonsequenzen übereinstimmen. In der dritten Spalte ist aufgeführt, welche Sequenzen für die Analyse der Exon-Intron-Strukturen in der *Capsella rubella*-Sequenz benutzt wurde (siehe Anhang).

Ein Vergleich einiger cDNA-Sequenzen mit der genomischen Sequenz von *A. thaliana* gab Hinweise auf mögliche Fehler in der genomischen Sequenz. Denn es wurden für die Gene B, 5 und C Verschiebungen im offenen Leseraster in der cDNA-Sequenz verglichen mit der genomischen Sequenz nachgewiesen. Durch die Sequenzierung von PCR-Produkten, die diesen Bereich des *A. thaliana* Genoms abdeckten, konnte bestätigt werden, daß in der Region von Bp 49400 bis Bp 60400 von ATFCA zahlreiche Sequenzfehler in der genomischen Sequenz vorlagen. In diesen Fällen wurden die PCR-ermittelten Daten für die Vergleiche herangezogen (M. Roßberg, unveröffentlicht; siehe *A. thaliana* Gen 5 und C im Anhang).

Bei den cDNA-Klonen A bzw. B handelt es sich um chimäre Klone. Die 119 Bp, die am 5'-Ende der cDNA B liegen, sowie eine 47 Bp lange Insertion in cDNA-Klon A zeigen keine Homologie zu *A. thaliana*-Sequenzen. Für cDNA B, die für Acyl-CoA-Oxidase (ACX1, AF057044) kodiert, wurde die cDNA-Sequenz von ACX1 aus der *A. thaliana*-Datenbank für die Sequenz-Vergleiche eingesetzt. Für cDNA A wurde aus der genomischen *A. thaliana*-Sequenz eine cDNA-Sequenz ohne den nicht-homologen 47 Bp-Bereich abgeleitet (siehe Anhang).

Weitere cDNA-Klone, für die einige Sequenzunterschiede zur genomischen *A. thaliana*-Sequenz ermittelt wurden, sind cDNA E, 6 und H. Außerdem lagen auch einige Unterschiede in cDNA 5 verglichen mit der PCR-ermittelten genomischen Sequenz vor (siehe oben). In diesen Fällen konnten die Unterschiede nicht auf fehlerhafte genomische oder cDNA-Sequenzen zurückgeführt werden. Daher wurde angenommen, daß in diesen Fällen eine fehlerhafte cDNA-Synthese stattgefunden haben mußte. Für cDNA E wurde die Genvorhersage weiterverwendet (siehe oben), während für die Gene 5, 6 und H cDNA-Sequenzen aus der genomischen *A. thaliana*-Sequenz abgeleitet wurden (siehe Anhang) und für die Vergleiche mit *Capsella rubella* verwendet wurden.

In Tabelle 3.11 sind die Ergebnisse für die Übereinstimmung der Genvorhersagen mit den cDNA-Sequenzen zusammengefaßt. Außerdem wurde in der Tabelle angegeben, mit welcher Sequenz weitergearbeitet wurde, um die Transkriptionsrichtungen und die Exon-Intron-Strukturen der Gene in *Capsella rubella* zu ermitteln. Die zur weiteren Analyse eingesetzten *A. thaliana*-Sequenzen und die ermittelten *Capsella rubella*-Gensequenzen sind im Anhang angegeben.

Die Anordnung der Gene und die Transkriptionsrichtung wurde in *C. rubella* und *A. thaliana* verglichen. Dabei wurden die Bereiche vom Start- bis zum Stop-Codon berücksichtigt. Die Ergebnisse sind in Tabelle 3.12 und Abbildung 3.19 gezeigt. Die Anordnung der 17 Gene in *C. rubella* ist identisch mit der Genreihenfolge in *A. thaliana*. Dies gilt ebenso für die Transkriptionsrichtungen. Von Gen 1 sind lediglich die letzten 136 Basenpaare in der Contig 0-Region vertreten. Gen C ist in *C. rubella* dupliziert und die beiden Kopien sind in Tandem organisiert. Gen 7 liegt ebenfalls in *Capsella rubella* in zwei Kopien vor, während *Arabidopsis thaliana* lediglich eine Kopie besitzt. RPP5-

homologe Sequenzen liegen im *A. thaliana* Ökotyp Columbia in einem Block aus zehn in Tandem organisierten Kopien vor (Bevan *et al.*, 1998). Auf den *C. rubella* Cosmid-Contigs I und II konnten lediglich zwei Teilsequenzen mit Homologie zu RPP5 nachgewiesen werden. Aus der Abbildung ist zu erkennen, daß die intergenischen Bereiche zwischen den Genen 3 und 4 sowie F und G in *C. rubella* verglichen mit *A. thaliana* vergrößert sind. Die restlichen intergenischen Bereiche zeigen eine ähnliche Größe wie in *A. thaliana*. Die Größe der Gene in dieser Region ist ebenfalls in beiden Arten konserviert.

Gen	Position auf ATFCA-Contig	Position auf Contig 0	Position auf Contig I	Position auf Contig II
1	3958 ¹⁾ -5733	1-136		
2	9827-12071	4137-6158		
A*	11562-15369	5616-9491		
3	16906-17355	10701-11109		
4	24657-25816	28897-30203		
B	49431-53189		14084-17873	
5	54474-56348		19203-21206	
C	59069-61639		22266-24939 26000-28757	
D	73792-75105		38755-40183	
E	76248-77669		41510-42922	
F	79432-80810		45462-46866	
G	95122-96809		71596-73261	
6*	98895-100044		75486-76507	
H*	100644-104386		76960-80845	
7	105263-106141		81607-82485	6409-7149
I	107427-120201 120558-123427 125007-129956 136508-141042 143674-150192 154065-154616 157650-162286 163511-168841 170844-175508 184656-186459		84132-86006 (RPP5-cDNA 2323-3993)	999-3937 (RPP5-cDNA 1-2952)
8	133055-134100 136360-137406			liegt nicht vor
J	181588-182082			16469-17065
K	184656-189726			liegt nicht vor

Tabelle 3.12: Die ermittelten Positionen für die Gene in *A. thaliana* und *C. rubella* sind aufgeführt. Die Angaben wurden für die Erstellung der Abbildung 3.19 verwendet. Die Positionsangaben in der Tabelle beziehen sich auf die in dieser Arbeit zusammengefaßten Region von ATFCA6 und ATFCA7. Der Bereich von 140001-200952 des ATFCA6-Contigs (Z79742), und 1-140000 des ATFCA7-Contigs (Z79743) entsprechen den Positionen 1-60952 bzw. 56292-196291 im hier verwendeten Contig. Für die Positionsangaben wurde der Bereich der Gene zwischen Start- und Stop-Codons berücksichtigt. Die Gene, die kein längeres offenes Leseraster haben, sind in der Tabelle mit einem Stern (*) gekennzeichnet. Für Gen 1 ist die Position des Startcodons nicht bekannt (¹⁾). Die Kopien der RPP5-homologen Sequenzen in *A. thaliana* wurden der Datenbank entnommen (<http://websvr.mips.biochem.mpg.de/proj/thal/>).

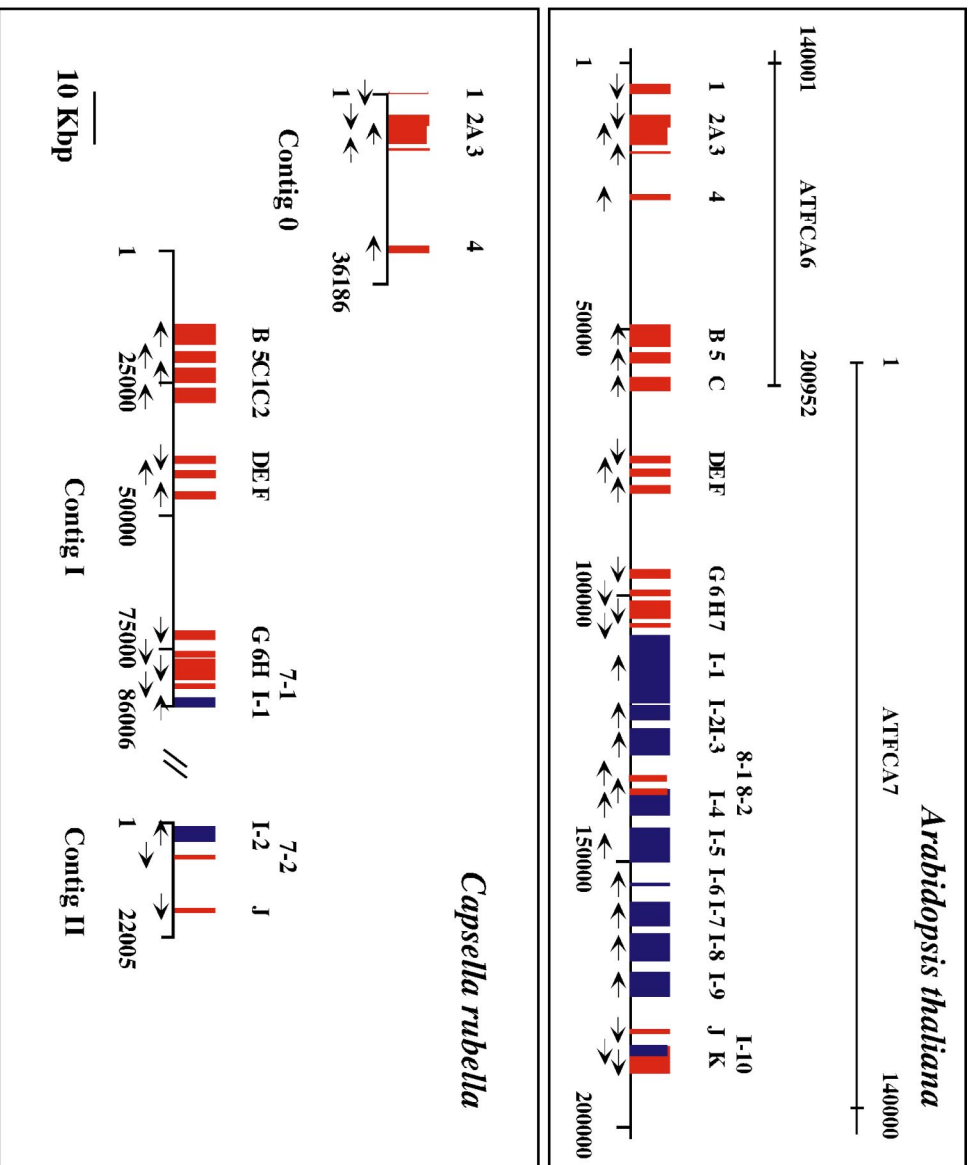


Abbildung 3.19: Die ermittelten Positionen der Gene in der 200 Kbp-Region auf *A. thaliana* Chromosom IV bzw. in der homologen Region in *C. rubella* ist gezeigt. Im oberen Teil der Abbildung ist die Region in *A. thaliana* mit den 19 Genen A-K und 1-8, die in dieser Region lokalisiert sind, abgebildet. Die Positionsangaben beziehen sich auf die in dieser Arbeit zusammengefaßten Regionen von ATFC A6 und ATFC A7. Der Bereich von 140001-200952 des ATFC A6-Contigs (Z79742) und 1-140000 des ATFC A7-Contigs (Z79743) entsprechen den Positionen 1-60952 bzw. 56292-196291 im hier verwendeten Contig. Die Transkriptionsrichtung der Gene wurde mit Pfeilen angegeben.

Ein Vergleich der cDNA-Sequenzen bzw. der vorhergesagten Sequenzen mit den genomischen *A. thaliana* und *C. rubella* Sequenzen wurde durchgeführt, um die Exon-Intron-Struktur der Gene zu bestimmen. Durch Sequenzuntersuchungen wurde für die Gene 2, 4, 5, B-G und J gezeigt, daß ein offenes Leseraster vorliegt. Für Vergleiche wurden lediglich die Bereiche vom Start- bis zum Stop-Codon berücksichtigt. Die Gene besitzen eine konservierte Exon-Intron-Struktur in *A. thaliana* und *C. rubella* (Abbildung 3.20A-F).

Im Fall von Gen 3 wurde ermittelt, daß der 5'-Bereich der Genvorhersage aus *A. thaliana* (At4g16690, Bp 1-440) mit der Sequenz des *C. rubella* Contigs keine Konservierung bezüglich der Exon-Intron-Struktur aufweist. Wird jedoch lediglich der Bereich 457-789 der Genvorhersage sowohl in *A. thaliana* als auch in *C. rubella* für die Genvorhersage benutzt, so kann in beiden Arten ein Gen mit einem Intron nachgewiesen werden (Abbildung 3.20AII).

Für Gen 7 wurde ein cDNA-Klon identifiziert, der nicht das vollständige offene Leseraster des Gens umfaßt. Auch in diesem Fall wurde ermittelt, daß die Genvorhersage (At4g16850) mit der Sequenz des *C. rubella*-Contigs keine Konservierung bezüglich der Exon-Intron-Struktur aufweist. Für Gen 7 liegen in *C. rubella* zwei Kopien vor (siehe Tabelle 3.12 und Abbildung 3.19). Ein Vergleich der genomischen *A. thaliana*-Sequenzen mit den *C. rubella*-Sequenzen ergab Hinweise, daß ein ca. 900 Bp großer Bereich im ATFCA-Contig hohe Homologien zu ähnlich großen Bereichen in den *Capsella*-Contigs I und II aufweist. Die weitere Analyse zeigte, daß auf diese Weise offene Leseraster in beiden Arten identifiziert werden konnten (Abbildung 3.20FI), dabei zeigen die letzten 431 Bp des für *A. thaliana* bestimmten offenen Leserasters vollständige Übereinstimmung mit der cDNA 4420 (siehe Anhang).

Die ermittelten Längen der offenen Leseraster sind für einige der Gene (3, 4, B, C und F) in *Arabidopsis thaliana* und *C. rubella* identisch. Eine der Kopien für Gen 7 (7-1) aus *C. rubella* weist ebenfalls eine identische Länge der Exonsequenzen verglichen mit dem *A. thaliana*-Gen auf, während die andere Kopie (7-2) kürzer ist. In den Genen 2, 5, D und J aus *C. rubella* liegt ein längeres offenes Leseraster als in *A. thaliana* vor, für die Gene E und G kann der umgekehrte Fall beobachtet werden. Bis auf die Gene 2, 3 und G ist die Gesamtlänge der Gene in *C. rubella* durch geringfügig längere Intronsequenzen größer als in *A. thaliana* (Tabelle 3.13).

Für die cDNAs A, 6 und H konnte kein offenes Leseraster identifiziert werden, welches den weitaus größten Teil der cDNA-Klone umfaßte. Die ermittelten cDNA-Sequenzen in *A. thaliana* weisen lediglich offene Leseraster auf, die sich über Teilbereiche der cDNA Sequenzen erstrecken und für Peptide bzw. Proteine mit einer Länge von bis zu 576 Aminosäuren kodieren. In diesen Fällen kann daher durch Sequenzanalysen allein nicht festgestellt werden, ob die cDNA-Sequenzen vollständigen Gensequenzen entsprechen.

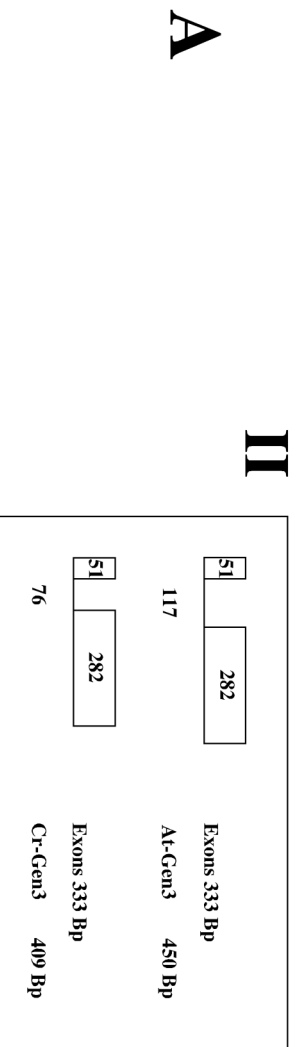
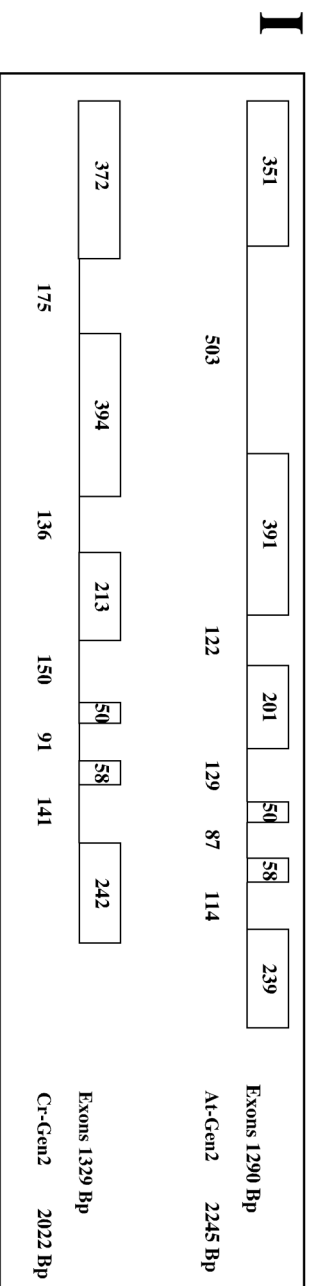


Abbildung 3.20: Genstrukturen der Gene 2-7, A-H; J in *A. thaliana* und *C. rubella*

A: Gen 2 und 3; **B:** Gen 4 und B; **C:** Gen 5 und C; **D:** Gen D und E; **E:** Gen F und G; **F:** Gen 7, J und A; **G:** Gen 6 und H

Die jeweiligen Exon- und Intron-Größen sind angegeben, sowie die Gesamtlängen der Gene und Exons sind rechts bzw. unter den Genstrukturen angezeigt. Die Angaben beziehen sich auf den Bereich zwischen Start- und Stop-Codon außer für Gen A, 6 und H. Da in diesen Fällen nicht bekannt ist, ob die cDNA-Klone den gesamten transkribierten Genbereichen entsprechen, sind die Enden des ersten und letzten Exons gestrichelt dargestellt. In der Abbildung ist *Arabidopsis thaliana* mit At und *Capsella rubella* mit Cr abgekürzt.

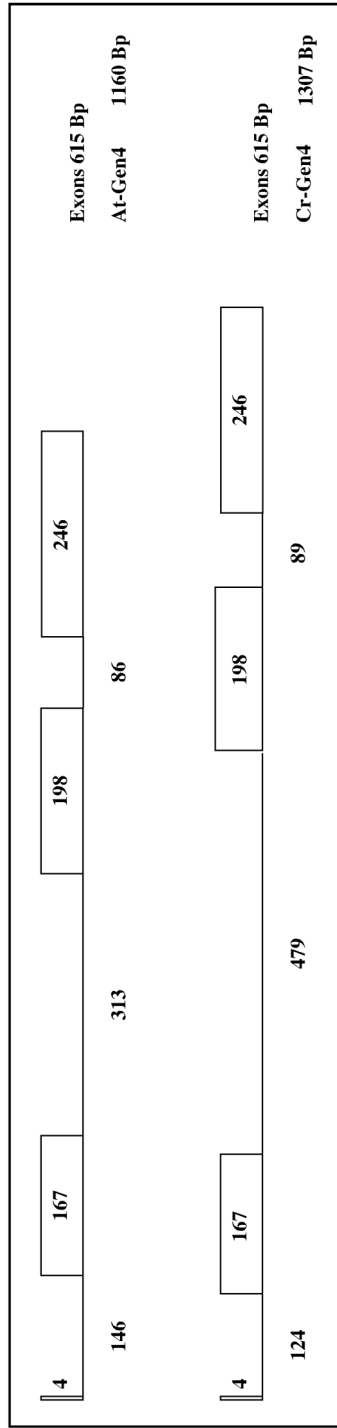
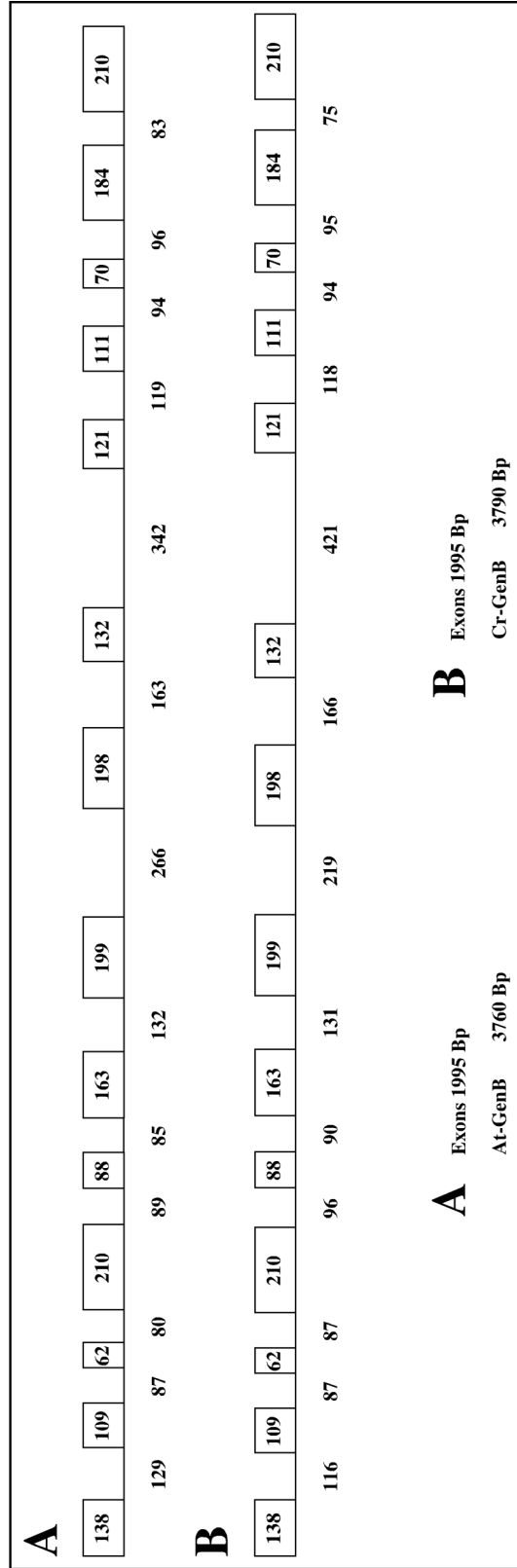


Abbildung 3.20 B



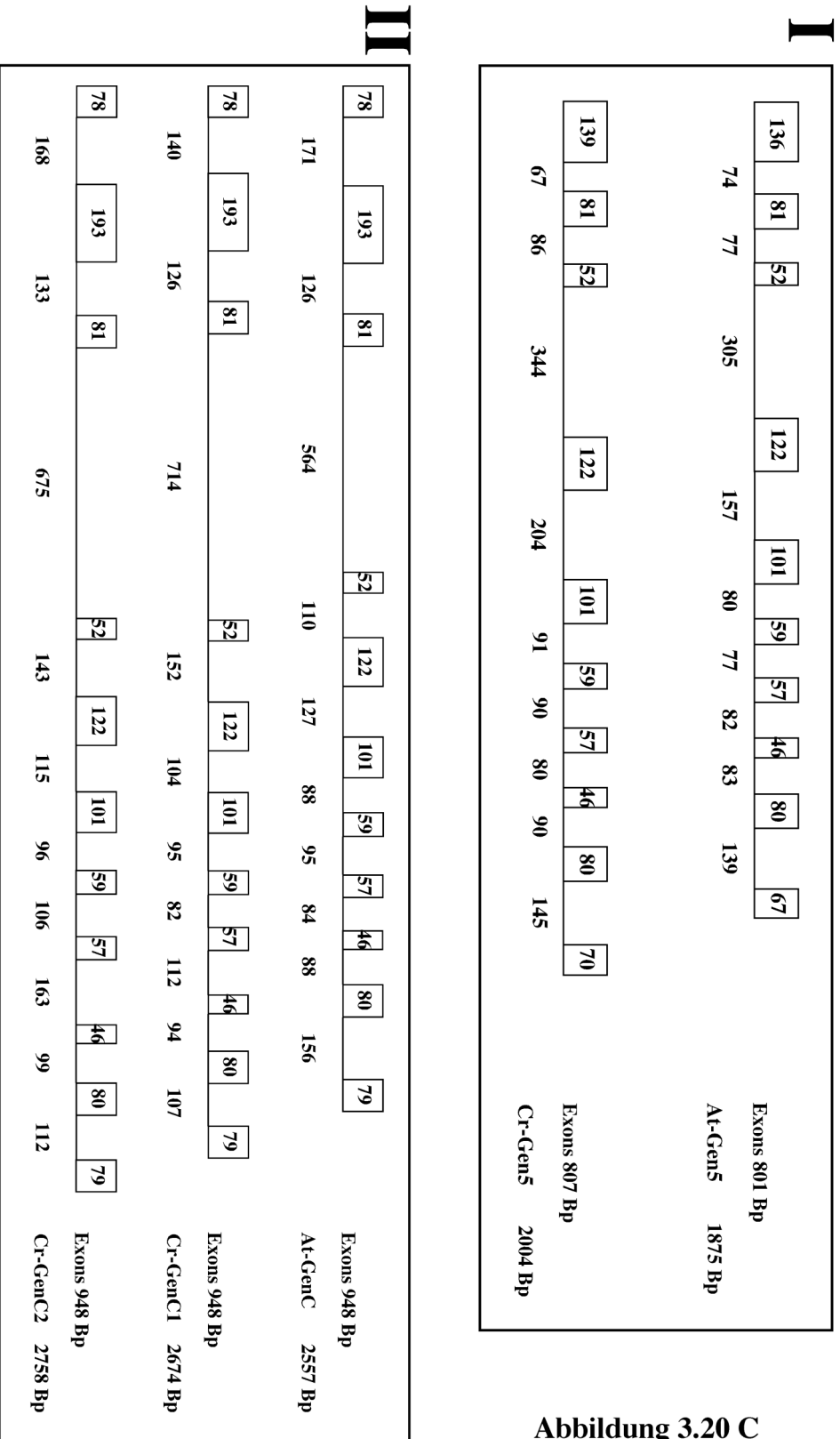


Abbildung 3.20 C

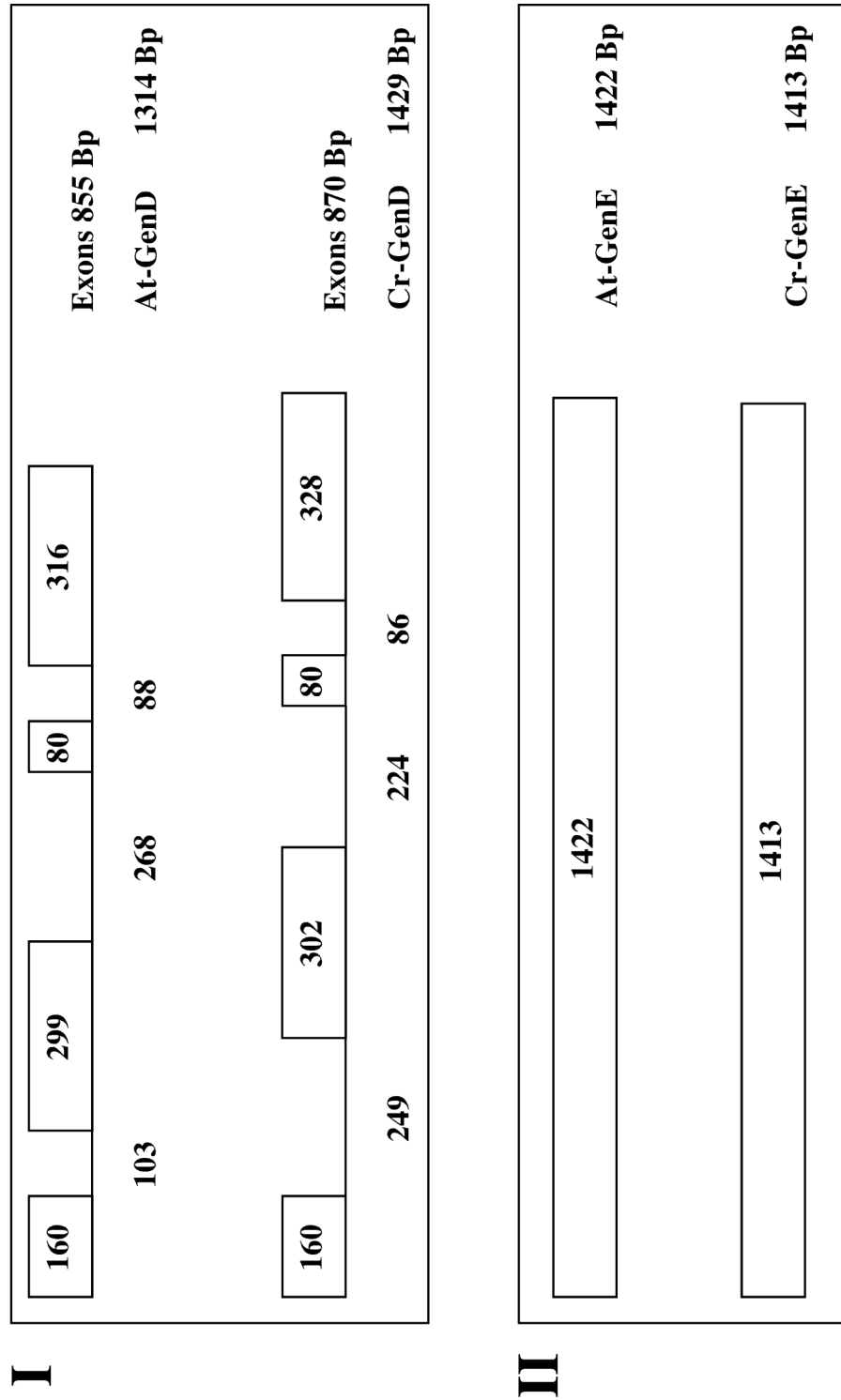


Abbildung 3.20 D

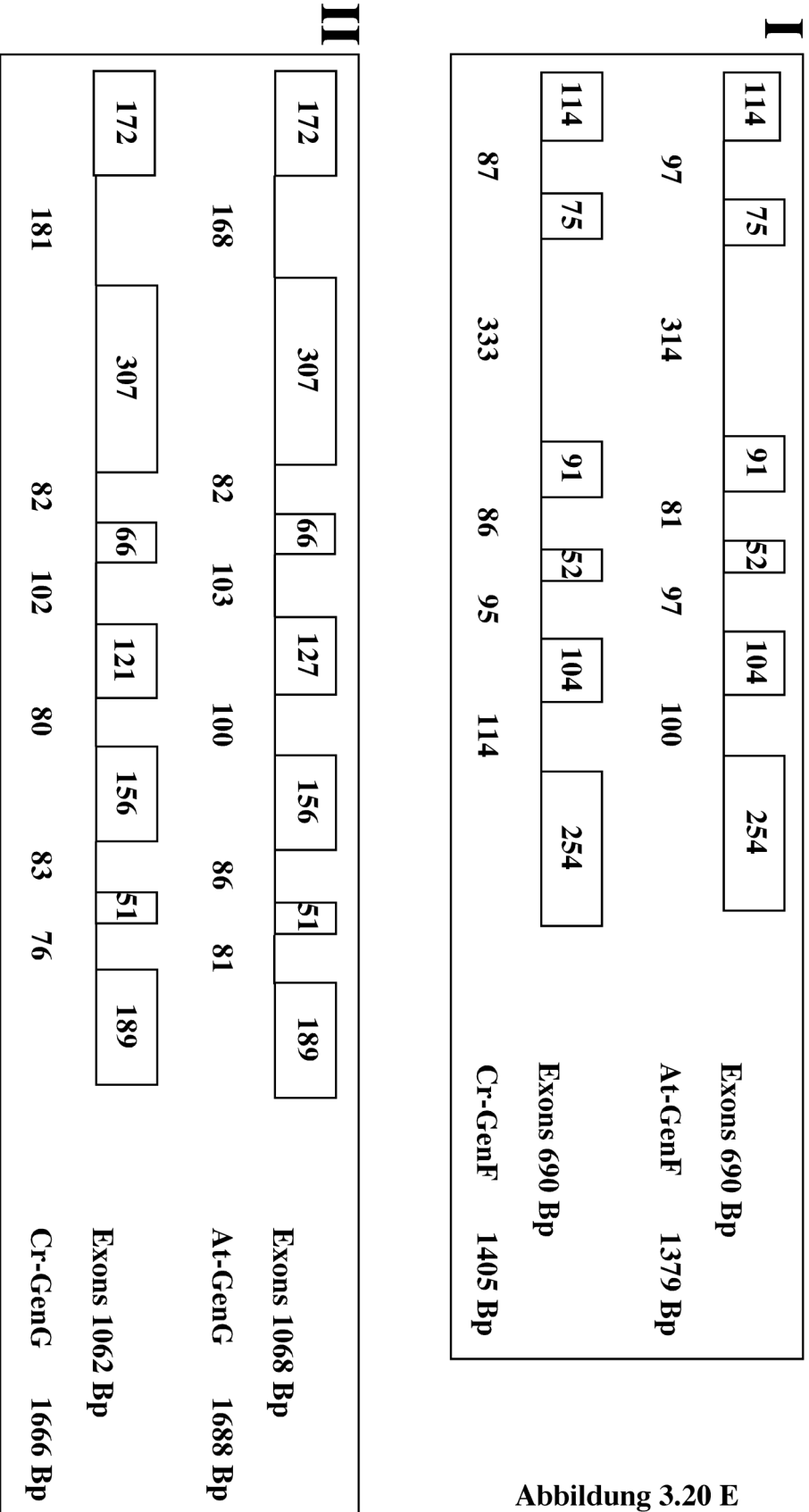


Abbildung 3.20 E

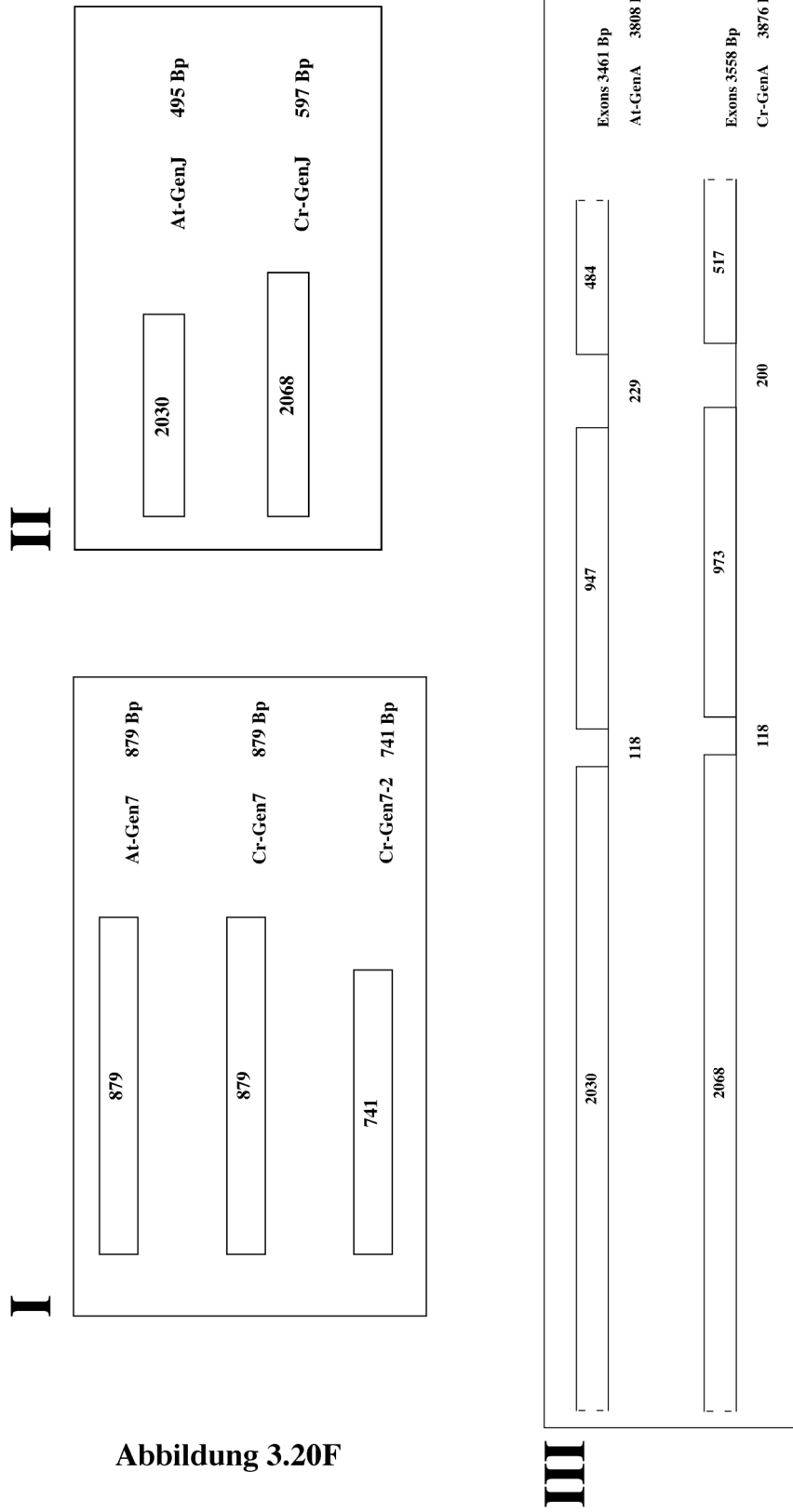


Abbildung 3.20F

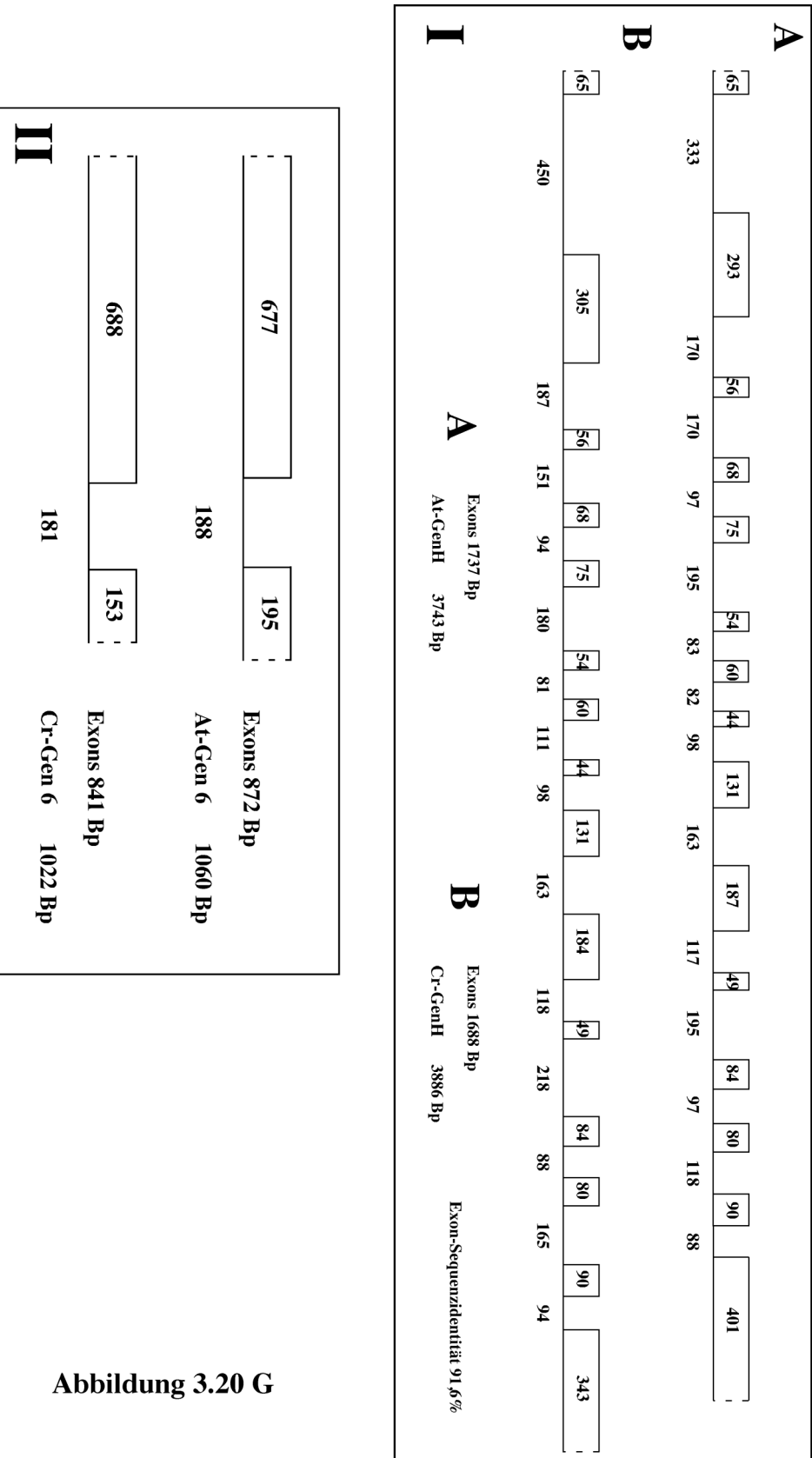


Abbildung 3.20 G

Die Sequenzvergleiche ergaben, daß auch für diese cDNA-Sequenzen konservierte Bereiche in *C. rubella* identifiziert werden konnten. Aus der genomischen *C. rubella*-Sequenz konnten Exon- und Intronsequenzen abgeleitet werden, die darauf hinweisen, daß diese Gene ebenso wie Gene, die für lange offene Leseraster kodieren, die gleiche Anzahl von Exons in beiden Arten aufweisen sowie eine konservierte Struktur zeigen (Abbildung 3.20FIII, 3.20G).

Gen	Exongröße in <i>A. thaliana</i> (Bp)	Exongröße in <i>C. rubella</i> (Bp)	Genlänge in <i>A. thaliana</i> (Bp)	Genlänge in <i>C. rubella</i> (Bp)
2	1290	1329	2245	2022
A*	3461	3558	3808	3876
3	333	333	450	409
4	615	615	1160	1307
B	1995	1995	3760	3790
5	801	807	1875	2004
C	948	C1 - 948 C2 - 948	2557	C1 - 2674 C2 - 2758
D	855	870	1314	1429
E	1422	1413	1422	1413
F	690	690	1379	1405
G	1068	1062	1688	1666
6*	872	841	1060	1022
H*	1737	1688	3743	3886
7	879	7-1 - 879 7-2 - 741	879	7-1 - 879 7-2 - 741
J	495	597	495	597

Tabelle 3.13: Die Exongrößen und Gesamtlängen der Gene in Abbildung 3.20 ist tabellarisch zusammengefaßt gezeigt. Es ist jeweils der Bereich zwischen Start und Stop-Codon berücksichtigt. Die Gene, die eine konservierte Exon-Intron-Struktur jedoch kein offenes Leseraster haben, sind in der Tabelle mit einem Stern (*) gekennzeichnet. C1 bzw. C2: Kopien von Gen C in *C. rubella*; 7-1 bzw. 7-2: Kopien von Gen 7 in *C. rubella*

Die Gene beider Arten wurden anschließend hinsichtlich der Sequenzidentität auf Nukleotidebene und auf Proteinebene untersucht. Für Sequenzidentitätsanalysen auf Nukleotidebene wurden nur die Exonsequenzen im Bereich zwischen Start- und Stop-Codon verglichen. Sie betragen auf Nukleotidebene >87,5% und auf Proteinebene >83,5% (Tabelle 3.14). Wurden die Intronsequenzen in die Sequenzvergleiche mit einbezogen, so ergaben sich geringere Sequenzidentitäten.

Ein weiterer Vergleich von Gen C in *A. thaliana* und *C. rubella* wurde auf der Aminosäuresequenz-Ebene durchgeführt. In beiden Genkopien von *C. rubella* liegen 20 synapomorphe Aminosäuren vor. Dagegen zeigen die *C. rubella*-Genkopie C1 bzw. die *C. rubella*-Genkopie C2 verglichen mit *A. thaliana*-Gen C lediglich 13 und 8 synapomorphe Aminosäuren in der Proteinsequenz. Dieser Unterschied in der Aminosäure-Zusammensetzung bedeutet, daß die Gen-Duplikation in *Capsella* nach der Trennung von *Arabidopsis* und *Capsella* stattgefunden haben muß (Abbildung 3.21).

Gen	Sequenzidentität auf Nukleotidebene (%)	Sequenzidentität auf Proteinebene (%)
2	87,6	83,7
3	88,9	85,5
4	96,4	99,5
B	93,3	96,7
5	93,2	94,7
C (<i>A. thaliana</i>) - C1	91,2	89,8
C (<i>A. thaliana</i>) - C2	91,0	88,2
C1 - C2	93,6	91,7
D	92,7	96,5
E	91,2	88,5
F	93,9	97,4
G	93,2	92,8
7 (<i>A. thaliana</i>) - 7-1	92,6	91,4
7 (<i>A. thaliana</i>) - 7-2	91,4	90,3
7-1 - 7-2	95,7	94,1
J	89,8	85,8

Tabelle 3.14: Die Werte für die Sequenzidentitäten zwischen den homologen Genen in *A. thaliana* und *C. rubella* ist für die Exon- und Protein-Sequenzen angegeben. Dabei wurden die Bereiche von den Start bis zu den Stop-Codons berücksichtigt.
C1 bzw. C2: Kopien von Gen C in *C. rubella*; 7-1 bzw. 7-2: Kopien von Gen 7 in *C. rubella*

Cr-C1	MGTILKLP	II	DLSSPEKLST	SSLIRQACLD	HGFFYLTNHG	VSEELMEGVL
At-C	MGTALKLP	II	DLSSPEKLST	SRLIRQACLD	HGFFYLTNHG	VSEELMEGVL
Cr-C2	MGTVLKLP	II	DLSSPEKIST	SRLIRQACLD	HGFFYLTNHG	VSEELMEGVL
Cr-C1	MESKKLFS	LP	LDEKVMARR	GFRGYSPLYE	EKPHSSTLAK	GDSKEMFTFG
At-C	IESKKLFS	LP	LDEKVMARRH	GFRGYSPLYD	EKLESSTSI	GDSKEMFTFG
Cr-C2	MESKKLFS	LP	LDEKVMARR	GFRGYSPLYD	EKLDPSATSK	GDSKEIFTFG
Cr-C1	SSEGVLAQ	IY	PNEWPREELL	PLWRPTMECY	YKTVMDVGKK	LLGLVALALN
At-C	SSEGVLGQ	LY	PNKWPLEELL	PLWRPTMECY	YKNVMDVGKK	LFGLVALALN
Cr-C2	SSEGVLAQ	IY	PNEWPLEELL	PLWRPTMECY	YKSVMDVGKK	LLGLVALALD
Cr-C1	LEENFFEQ	VG	GFNAQAAVVR	LLRYPGELNS	SGEVTTCGASA	HSDYGMITLL
At-C	LEENYFEQ	VG	AFNDQAAVVR	LLRYSGESNS	SGEETTCGASA	HSDFGMITLL
Cr-C2	LEEKFFEQ	VE	GFNVQAAVVR	LLRYPGELNS	SGEVTTCGASA	HSDYGMITLL
Cr-C1	ATDGVVGL	QV	CRDKDKEPKV	WEDVAGVKGA	FVVNIGDLME	RWTNGLFRST
At-C	ATDGVVGL	QV	CRDKDKEPKV	WEDVAGIKGT	FVVNIGDLME	RWTNGLFRST
Cr-C2	ATDGVVGL	QV	CRDKDKEPRV	WEDVAGIKGA	FLVNIGDLME	RWTNGLFRST
Cr-C1	MHRVVS	VGKE	RYSVAVFVDP	DPNCVVECLE	SCCSETSPPR	FPPVRTKDYF
At-C	LHRVVS	VGKE	RFSVAVFVDP	DPNCVVECLE	SCCSETSPPK	FPPVRARDYF
Cr-C2	MHRVVS	VGKE	RYSVAVFFDP	DPNSVVECLE	SCCSETSPPR	FPPIRTRDYF
Cr-C1	HERFSQ	TLAS	YSGSG			
At-C	HERFSQ	TLAS	YSGSD			
Cr-C2	HERFSQ	TFAS	YSGSG			

Abbildung 3.21: Gezeigt ist der Aminosäuresequenz-Vergleich für Gen C aus *A. thaliana* (At-C) mit den beiden Kopien von *C. rubella* (Cr-C1, Cr-C2). Positionen an denen Aminosäure-Unterschiede vorliegen, sind grau unterlegt.

3.2.2 Vergleichende Mikrosyntenieanalyse einer 30 Kbp-Region in *A. thaliana* und *C. rubella*

Eine weitere Region auf *A. thaliana* Chromosom I wurde für die Syntenieanalyse in den beiden Spezies *A. thaliana* und *C. rubella* ausgesucht. Es sollte festgestellt werden, ob und inwieweit die ermittelten Synteniedaten in der 200 Kbp Region auf *A. thaliana* Chromosom IV für weitere chromosomale Regionen zutreffen. Die ausgewählte Region liegt in *A. thaliana* auf BAC-Klon F20N2 (AC002328) in Nachbarschaft zum genetischen Marker nga280 (81.7 cM; http://nasc.nott.ac.uk/new_ri_map.html). Zur genetischen Kartierung in *C. rubella* wurde EST 21484 verwendet, welcher ca. 100 Kbp von nga280 entfernt lokalisiert ist (siehe 3.1).

3.2.2.1 Identifizierung kodierender Sequenzen

Durch Vergleich genomischer Chromosom I-Sequenzen von *A. thaliana* mit den zugänglichen EST-Sequenzdaten in der Datenbank konnten für diesen Genombereich drei EST-Klone (22916, 21484 und 5787) (Höfte *et al.*, 1993; Newman *et al.*, 1994) identifiziert werden. Die Sequenzidentitäten der drei ESTs betragen 90% oder mehr, verglichen mit der genomischen *A. thaliana*-Sequenz. In Tabelle 3.15 sind diese ESTs mit den jeweiligen Akzessionsnummern aus der Datenbank und ihrer Lage auf dem jeweiligen BAC-Klon angegeben. Die ESTs sind alle auf BAC F20N2 zwischen Position 79956 und 101710 lokalisiert.

EST-Klon	Akzessionsnummer	Position auf BAC F20N2
22916	W43308	79956-83795
21484	N96681	99292 -101649
5787	T42524	101212-101710

Tabelle 3.15: Für die drei identifizierten EST-Klone sind die Akzessionsnummern in der Datenbank angegeben. Alle ESTs sind auf BAC-Klon F20N2 auf *A. thaliana* Chromosom I lokalisiert. Die Bereiche auf BAC F20N2, die den ermittelten cDNA-Sequenzen entsprechen, sind aufgeführt.

Die ESTs wurden sequenziert (Sequenzen siehe Anhang) und zur Identifizierung von Cosmiden in *C. rubella*, die homologe Sequenzen tragen, eingesetzt. Die ermittelten cDNA-Sequenzen zeigen Identitäten zwischen 98% und 100% zur genomischen *A. thaliana*-Sequenz. Außerdem konnte gezeigt werden, daß die ESTs 5787 und 21484 auf dem BAC-Klon F20N2 in der gleichen Region lokalisiert sind (Tabelle 3.15). Es handelt es sich bei diesen beiden ESTs um Klone, die verschiedene Abschnitte eines Gens repräsentieren.

3.2.2.2 Identifizierung der homologen Region in *Capsella rubella*

Die drei ESTs wurden als radioaktiv markierte DNA-Sonden in Koloniehybridisierungsexperimenten mit der genomischen *Capsella rubella* Cosmid-Bibliothek eingesetzt (Schmidt *et al.*, 1999). Es wurden acht *C. rubella* Cosmid-Klone identifiziert (Abbildung 3.22 und 3.23). Southern-Blot-Analysen zeigten, daß auch in dieser Region des *C. rubella*-Genoms die identifizierten Cosmide mit einem oder mehreren Genen, die als Sonden eingesetzt wurden, hybridisierende DNA-Fragmente aufwiesen (Abbildung 3.22). Von acht identifizierten *C. rubella*-Cosmiden hybridisierten die Cosmide 7 und 10 sowohl mit EST 21484 als auch 22916. Wenn berücksichtigt wurde, daß die ESTs 5787 und 21484 cDNA-Klone eines Gens repräsentieren, hybridisieren die restlichen sechs Cosmide zu jeweils einem der als Sonde benutzten Gene in der betrachteten Region.

Die durchschnittliche Insertgröße der *C. rubella*-Cosmide beträgt 20-25 Kbp und der Abstand der ESTs 21484 und 22916, die zu DNA-Fragmenten in zwei verschiedenen *C. rubella*-Cosmiden (Cosmid 7 und 20) hybridisieren, beträgt in *A. thaliana* ca. 16 Kbp (Tabelle 3.15). Daher weisen die Southern-Blot Daten auch für diese Region in *C. rubella* auf eine ähnliche genomische Größe wie in *A. thaliana* hin. Die Cosmide wurden mit Hilfe der ermittelten Southern-Blot Daten zu einem Cosmid-Contig zusammengefaßt (Abbildung 3.23). Aus diesem Cosmid-Contig wurden die Cosmide 10 und 20 für die Sequenzanalysen ausgewählt. Beide Cosmide hybridisieren mit den ESTs 5787 und 21484, und Cosmid 10 hybridisiert zusätzlich mit EST 22916.

3.2.2.3 Subklonierung und Sequenzierung der Cosmide 10 und 20

Zur Subklonierung der Cosmide wurden Einzelspaltungen mit *Hind*III und Mehrfachspaltungen mit *Eco*RI/*Xba*I/*Xho*I eingesetzt. Die *Hind*III-DNA-Fragmente wurden in die entsprechende Schnittstelle im Plasmidvektor pGEM7Z kloniert und die DNA-Fragmente aus der Dreifachspaltung nach Auffüllreaktion der überhängenden Enden mit dem Klenow-Fragment der Polymerase I in die *Sma*I-Schnittstelle desselben Vektors. Die identifizierten Subklone wurden sequenziert und zur weiteren Analyse mit Hilfe des GCG-Programmpaketes bearbeitet.

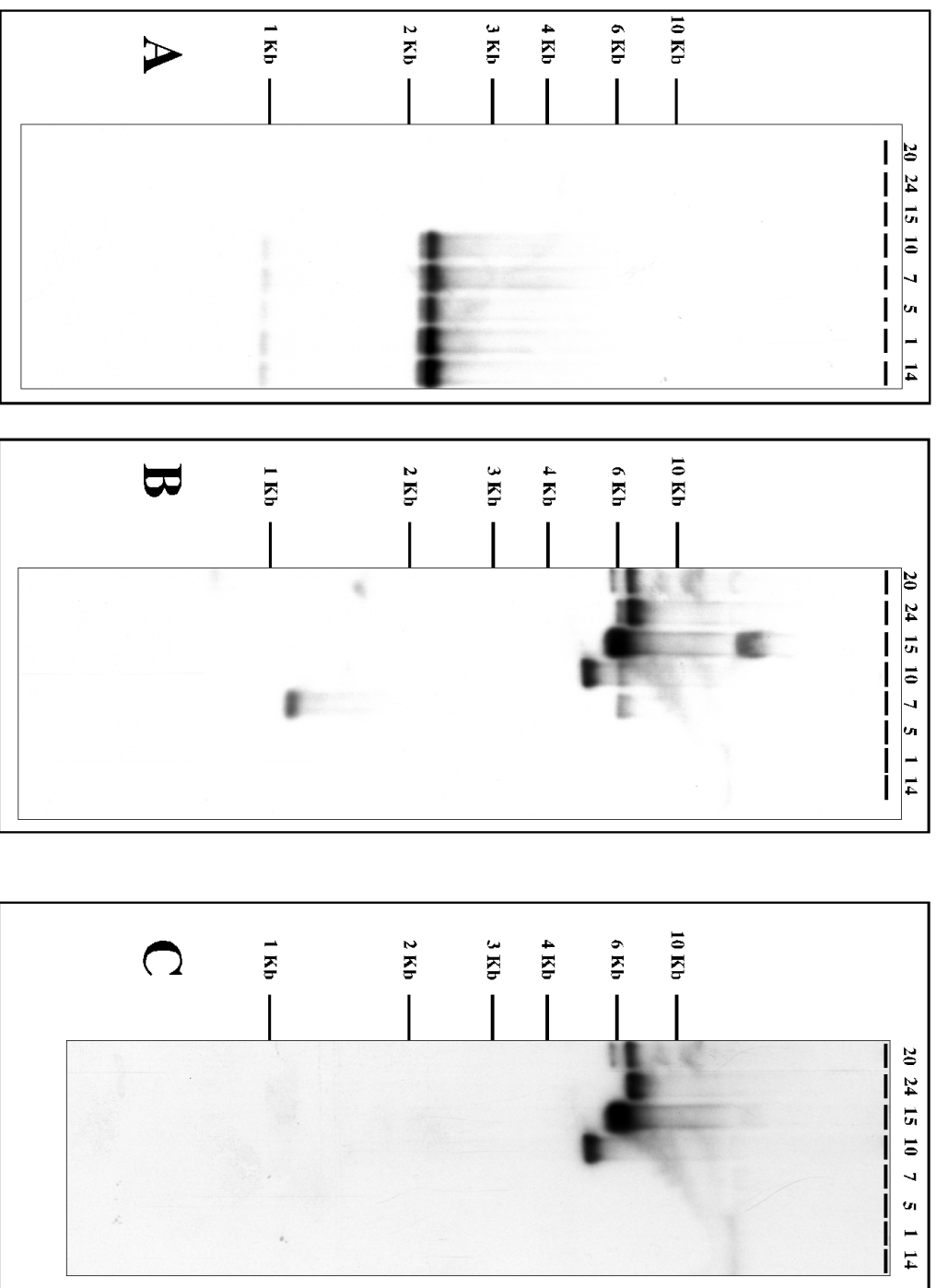


Abbildung 3.22: Die Ergebnisse von Southern-Blot-Experimenten der acht identifizierten *C. rubella*-Cosmide aus der *C. rubella* Cosmid-Bibliothek mit Sonden der ESTs 5787, 21484 und 22916 sind gezeigt. Die *C. rubella*-Cosmide wurden mit *Hind*III gespalten. **A:** Sonde 22916; **B:** Sonde 21484 und **C:** Sonde 5787. Zur Orientierung sind links die Fragmentgrößen des DNA-Längenstandards angegeben.

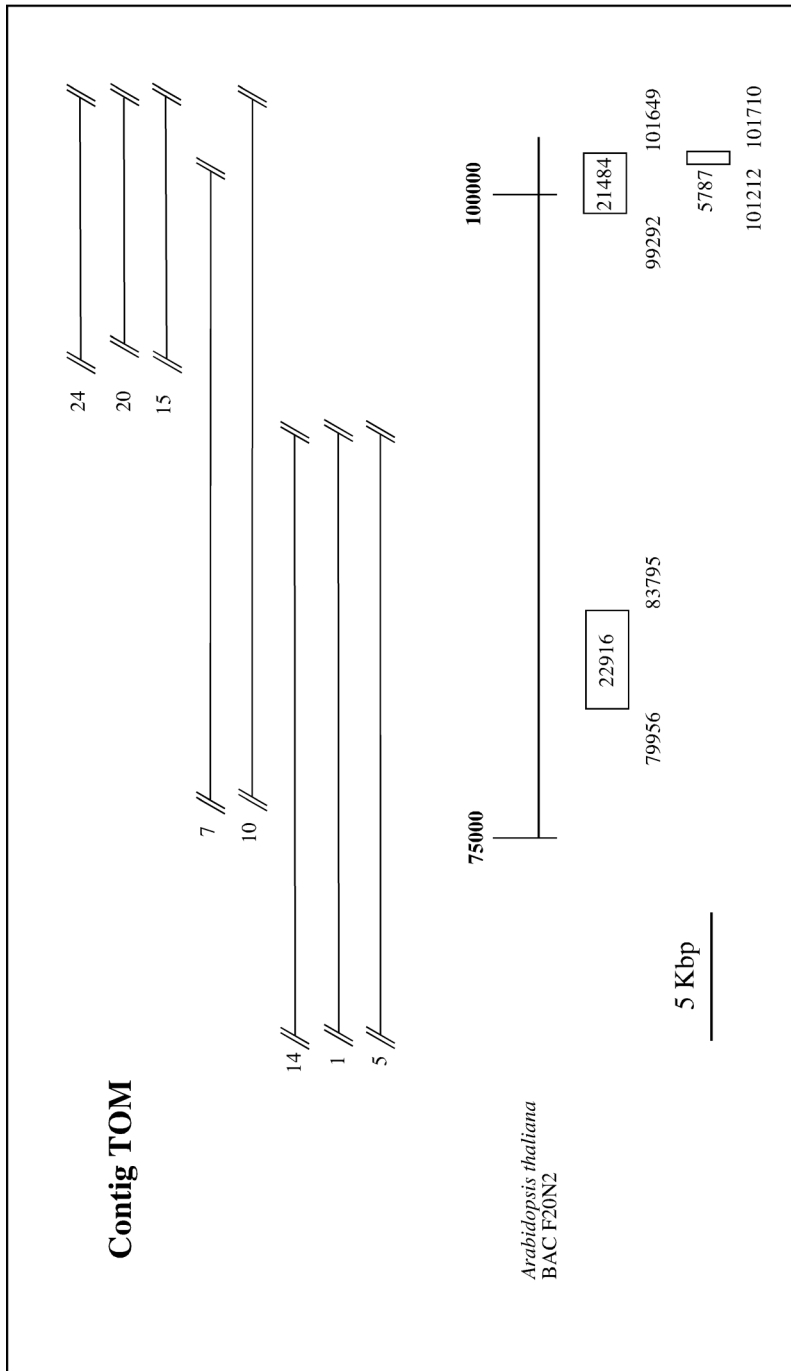


Abbildung 3.23: In der unteren Bildhälfte ist die analysierte Region des BACs F20N2 für den Bereich zwischen den Positionen 75000 und 102000 gezeigt. BAC F20N2 ist auf *A. thaliana*-Chromosom I in der Nähe zum genetischen Marker nga 280 lokalisiert. Die orthologen *C. rubella*-Cosmide sind in der oberen Bildhälfte zu einem Cosmid-Contig, Contig TOM, zusammengefaßt dargestellt. Zur Erstellung des Contigs wurde lediglich die ESTs 22916, 21484 und 5787 als Sonden eingesetzt, die Positionen der Cosmid-Enden sowie die genaue Größe der Cosmide wurde nicht ermittelt. Folglich ist die genaue Anordnung der *C. rubella*-Cosmide zueinander und eine maßstabgerechte Darstellung der Cosmide in dieser Abbildung nicht möglich. In der Abbildung wurden daher die Cosmidenden mit schrägen Linien markiert.

3.2.2.4 Sequenzanalyse von Cosmid 10 und 20 aus *Capsella rubella*

Die Sequenzen der einzelnen Subklone wurden zu einer genomischen Gesamtsequenz von 27056 Bp zusammengefügt. In Abbildung 3.24 ist die Anordnung der einzelnen Subklone relativ zur ermittelten genomischen Sequenz der gesamten *C. rubella*-Region gezeigt. Sequenzbereiche, die durch Subklone nicht repräsentiert waren, wurden mit Hilfe von PCR-Experimenten amplifiziert. Im Fall von Cosmid 10 wurde die gesamte Sequenz der genomischen *C. rubella*-Insertion ermittelt. Die Insertgröße für diesen Cosmid-Klon beträgt ca. 22,5 Kbp. Im Fall von Cosmid 20 wurden ca. 9,8 Kbp der klonierten *C. rubella*-DNA sequenziert. Die sequenzierten Subklone sind im Anhang aufgelistet.

3.2.2.5 Vergleichende Sequenzanalyse in *A. thaliana* und *C. rubella*

Für die weiteren Mikrosyntenie-Analysen wurden sechs vorhergesagte Gene aus der annotierten Datenbanksequenz, im weiteren als TOM 1-6 bezeichnet, herangezogen. Zwei der Gene werden durch die verwendeten EST-Klone 5787, 21484 und 22916 repräsentiert. Bei den ESTs 5787 und 21484 handelt es sich um cDNAs des Gens TOM 2. EST 22916 zeigt Homologien zu Gen TOM 5. Allerdings konnte in diesem Fall gezeigt werden, daß die vorhergesagte Genstruktur falsch ist, da Teile der cDNA-Sequenz nicht den vorhergesagten Exons entsprechen. Die mit Hilfe der cDNA abgeleitete Genstruktur ist in Abbildung 3.26B(E) gezeigt. In Tabelle 3.16 sind die Positionen der vorhergesagten Gene auf der Teilsequenz von BAC F20N2 (34161 Bp) angegeben. Diese Teilsequenz wurde für die weiteren Analysen aus der Gesamtsequenz des BAC-Klons erstellt. Dabei entspricht Nukleotidposition 1 auf der Teilsequenz Position 75000 in der Gesamtsequenz von F20N2 und Position 34161 entsprechend 109160.

Gen	Position auf der Teilsequenz von F20N2	Sequenzidentitätsnummer des vorhergesagten Gens
TOM 1	30588-31925	25033
TOM 2	24275-26555	25032
TOM 3	19738-21982	25031
TOM 4	12657-16157	25030
Gen 22916	5041-8693	W43308
TOM 6	2384-4214	25028

Tabelle 3.16: Die fünf vorhergesagten Gene TOM 1-4, 6 und Gen 22916 in der untersuchten Region auf Chromosom I von *A. thaliana* wurden mit den entsprechenden Positionen auf der Teilsequenz von BAC F20N2 und der Sequenzidentitätsnummer angegeben. Dabei ist der Bereich vom Start- bis zum Stop-Codon berücksichtigt. Die Sequenzidentitätsnummern wurden aus der Datenbank bezogen (<http://www.tigr.org/tdb/at/atgenome/atgenome.html>).

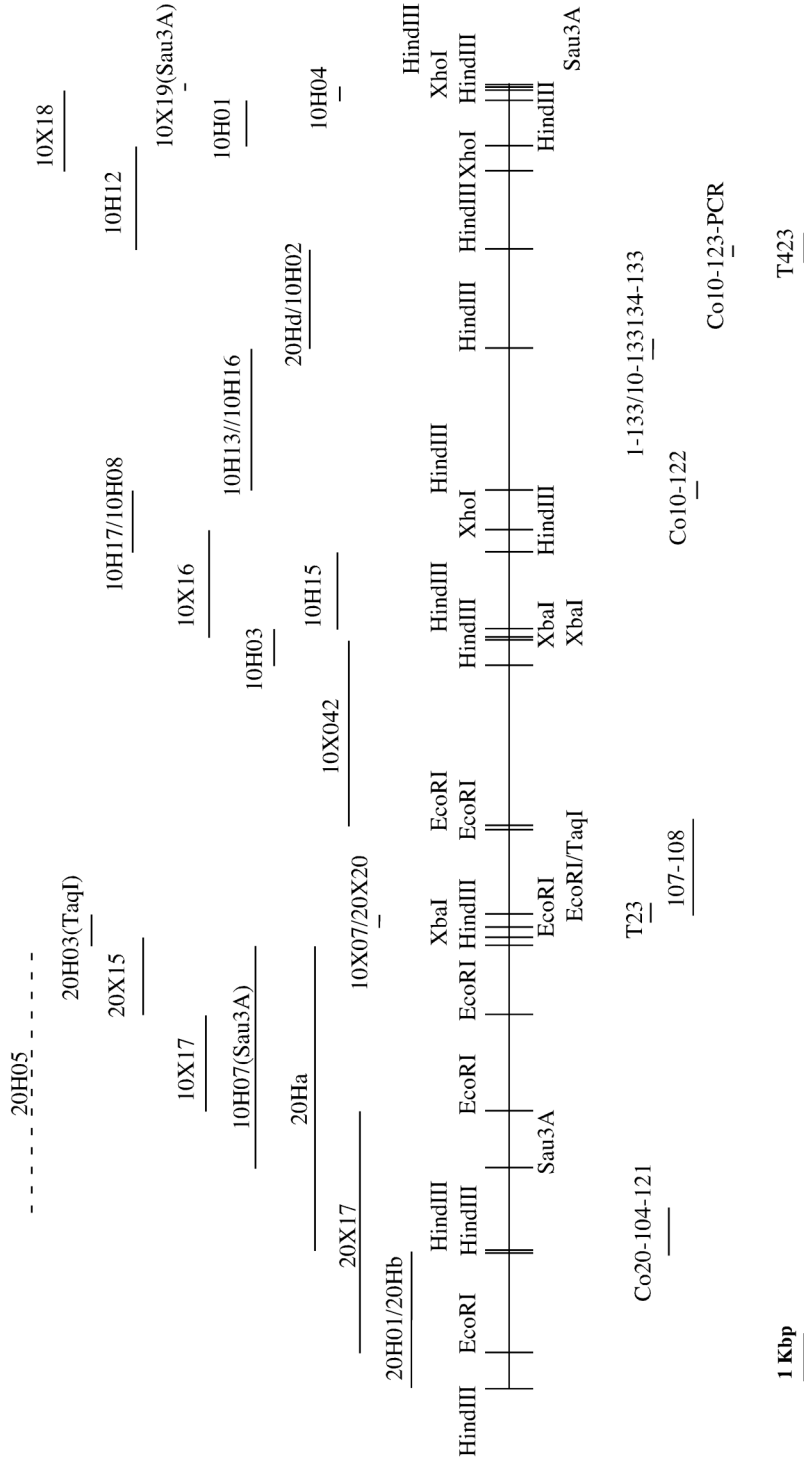


Abbildung 3.24: Die Positionen der einzelnen Subklone aus den Cosmiden 10 und 20, die zur Sequenzierung der Contig TOM eingesetzt wurden, sind in Beziehung zur ermittelten Gesamtsequenz der *C. rubella*-Region gesetzt worden (obere Bildhälfte). Die sequenzierten PCR-Produkte sind im unteren Bildteil dargestellt. Zur Orientierung sind die Positionen der Restriktionsenzymchnittstellen, welche zur Subklonierung von Cosmid 10 und 20 eingesetzt wurden, in der Gesamtsequenz angegeben. Lediglich die jenigen *Sau3A*- und *TaqI*-Erkennungssequenzen sind dargestellt, die die Insertionen der genomischen *C. rubella*-DNA in den Cosmiden begrenzen. Klon 20H05 ist chimär (gestrichelte Darstellung).

Die aus der *A. thaliana*-Datenbank entnommenen Sequenzen von TOM 1-4 und 6 wurden für die anschließenden Analysen verwendet, wobei diese mit dem Start-Codon ATG beginnen und mit einem der drei Stop-Codons enden. Für die Analyse von TOM 5 wurde das offene Leseraster der cDNA-Sequenz des EST-Klons 22916 eingesetzt.

Durch Vergleiche der erstellten Gesamtsequenz von *C. rubella* mit Sequenzen von TOM 1-4, 6 und der cDNA-Sequenz von EST 22916 aus *A. thaliana* wurden die Positionen für sechs homologe Gene in *C. rubella* abgeleitet. Die Ergebnisse sind in Abbildung 3.25 zusammengefaßt. In der sequenzierten Region von *C. rubella* ist das Gen TOM 6 nicht mit der gesamten Gensequenz vertreten. Durch einen Sequenzvergleich von TOM 6 mit der *C. rubella*-Gesamtsequenz konnte festgestellt werden, daß lediglich der kodierende 3'-Bereich des Gens sequenziert wurde. Die restlichen fünf Gene TOM 1-4 und 22916 sind in *C. rubella* mit der gesamten Gensequenz vertreten. Für alle sechs Gene ist die Reihenfolge und die Transkriptionsrichtung in *A. thaliana* und *C. rubella* identisch. Auch die Größe der Gene ist in *C. rubella* und *A. thaliana* sehr ähnlich. Die intergenischen Bereiche zeigen dagegen größere Unterschiede in ihrer Länge. Während in *C. rubella* für den intergenischen Bereich zwischen TOM 1 und 2 eine Vergrößerung der genomischen Sequenzlänge um ca. 740 Bp verglichen mit dem in *A. thaliana* ermittelt wurde, wurde für die restlichen intergenischen Bereiche eine Reduktion der genomischen Sequenzlänge zwischen 200 Bp und 1,8 Kbp festgestellt (Abbildung 3.25). In *C. rubella* beträgt die genomische Größe der Region zwischen TOM 1 (Position 2960) und TOM 6 (Position 26602) 23643 Bp und für den korrespondierenden Bereich in *A. thaliana* 26373 Bp (Position 4215-30587). Für diesen Bereich ist die genomische Sequenz in *A. thaliana* um ca. 2,7 Kbp länger als in *C. rubella*.

In den weiteren Analysen wurde die Struktur und Sequenzhomologie der Gene auf Nukleotid- und Protein-Ebene untersucht (Abbildung 3.26). Dazu wurde ein Sequenzvergleich der vorhergesagten Exon-Sequenzen mit den genomischen *A. thaliana*- und *C. rubella*-Sequenzen vorgenommen. Die Gene TOM 1-4 und 22916, die in der sequenzierten Region von *C. rubella* mit der gesamten Gensequenz vorliegen, haben die gleiche Zahl an Exons und Introns wie die homologen Gene in *A. thaliana*. Die Größe der Exons und Introns variiert jedoch in den meisten Fällen zwischen beiden Spezies. Die Varianz für die Exons beträgt in der Regel zwischen 3 und 15 Basenpaare und für die Intronbereiche 2 bis 109 Basenpaare. Lediglich das dritte Exon von TOM 3 zeigt zwischen *C. rubella* und *A. thaliana* einen Größenunterschied von 90 Bp. Beim Vergleich der Gesamtlängen der Gene TOM 1, 2, 3, 4 und 22916 in *A. thaliana* und *C. rubella* fällt auf, daß vier der Gene in *C. rubella* länger sind. Ein Vergleich der gesamten Exonlängen zeigte, daß in TOM 1 und 3 die kodierenden Bereiche in *C. rubella* größer sind. TOM 4 besitzt in beiden Pflanzen die gleiche Länge von 3501 Bp. Im Fall von TOM 2 und Gen 22916 ist dies umgekehrt.

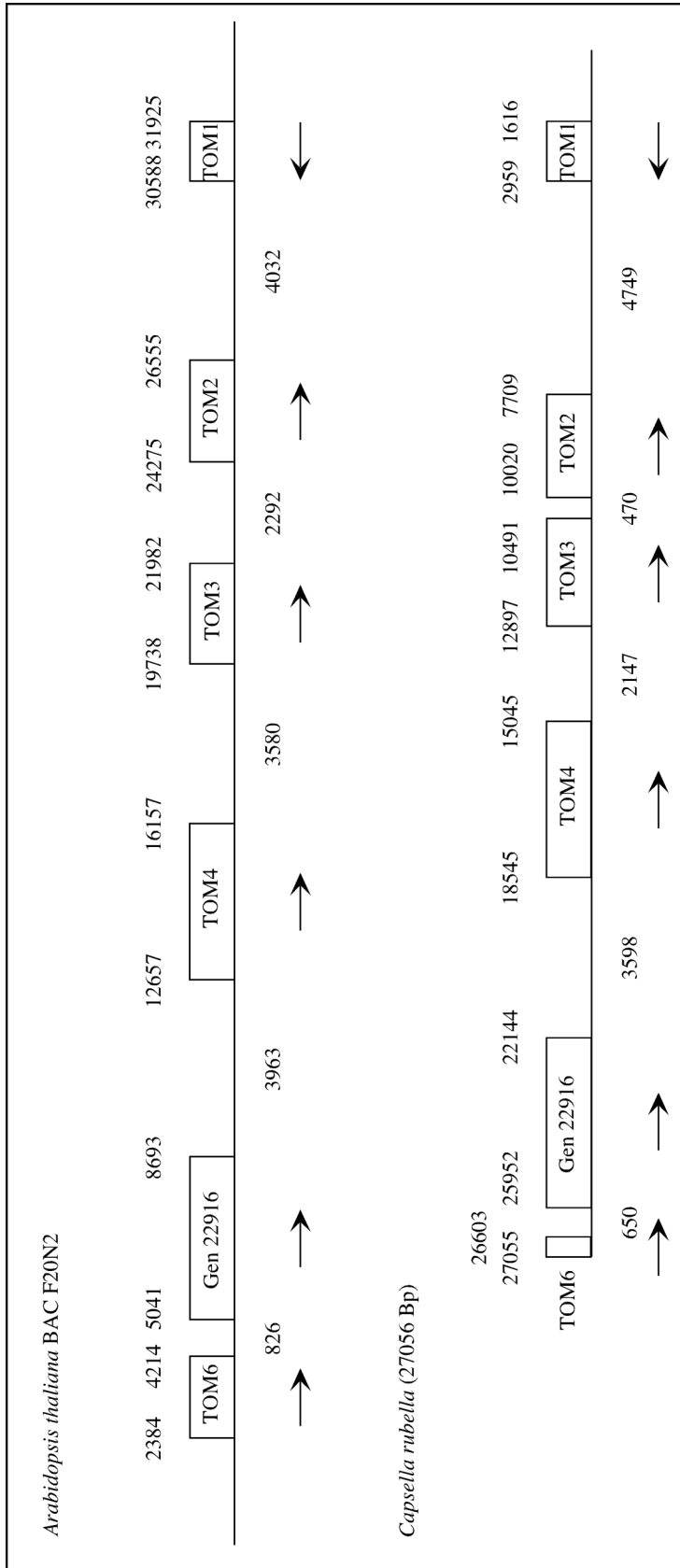


Abbildung 3-25: Anordnung der sechs Gene TOM 1-4, 6 und Gen 22916 in *A. thaliana* und *C. rubella*. Für *A. thaliana* ist in der oberen Darstellung eine 34161 Bp große Teilsequenz von BAC F20N2 angegeben. Position 1 in der Teilsequenz entspricht der in der Datenbank zugänglichen BAC-Sequenz Position 75000. Für *C. rubella* ist die in dieser Arbeit ermittelte Gesamtsequenz des Bereiches angegeben. Zusätzlich wurden die Positionen der Start- und Stop-Codons der Gene sowie die Größen der intergenischen Bereiche gezeigt. Die Transkriptionsrichtung der Gene wurden in der Abbildung mit Pfeilen dargestellt. In *C. rubella* ist die gesamte Gensequenz von Tom 6 nicht ermittelt. Der sequenzierte Bereich umfasst lediglich 453 Bp aus der kodierenden 3'-Region des Gens.

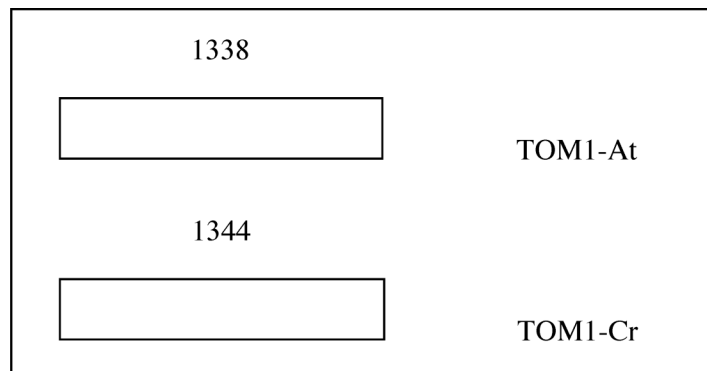
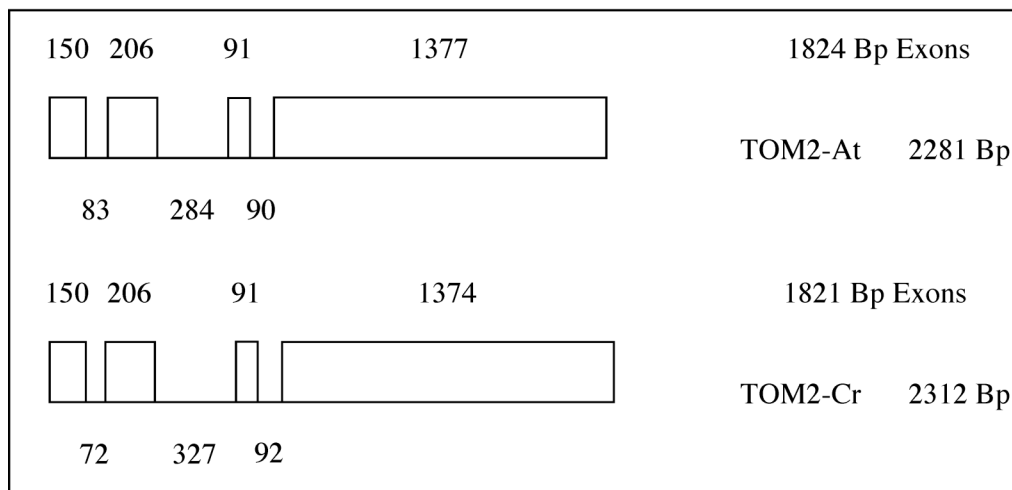
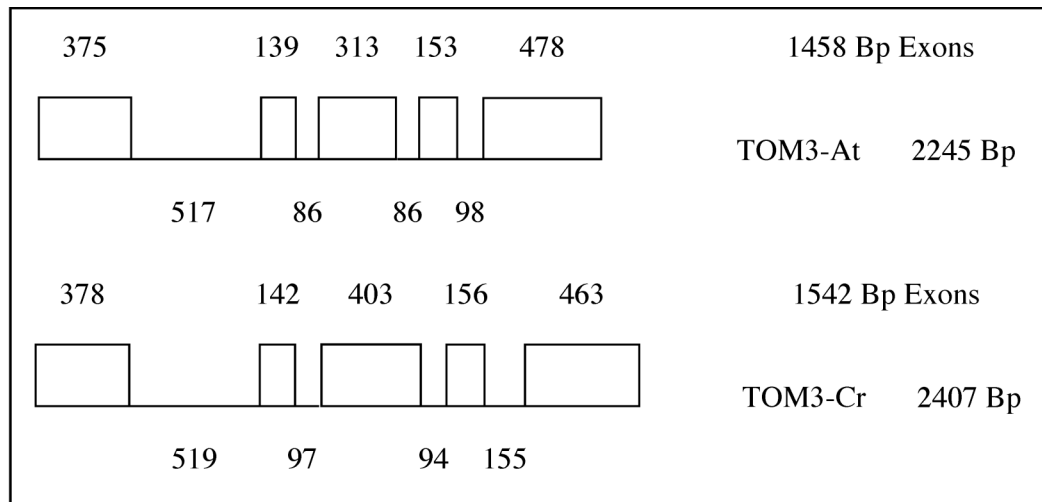
**A****B****Abbildung 3.26A**

Abbildung 3.26: Genstrukturen für TOM1-4 und 22916 in *A. thaliana* und *C. rubella*

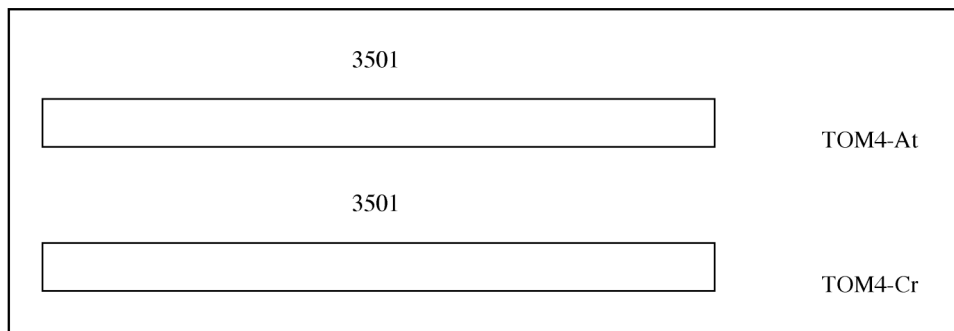
3.26A: A: TOM1; B: TOM2

3.26B: C: TOM3; D: TOM4 und E: 22916

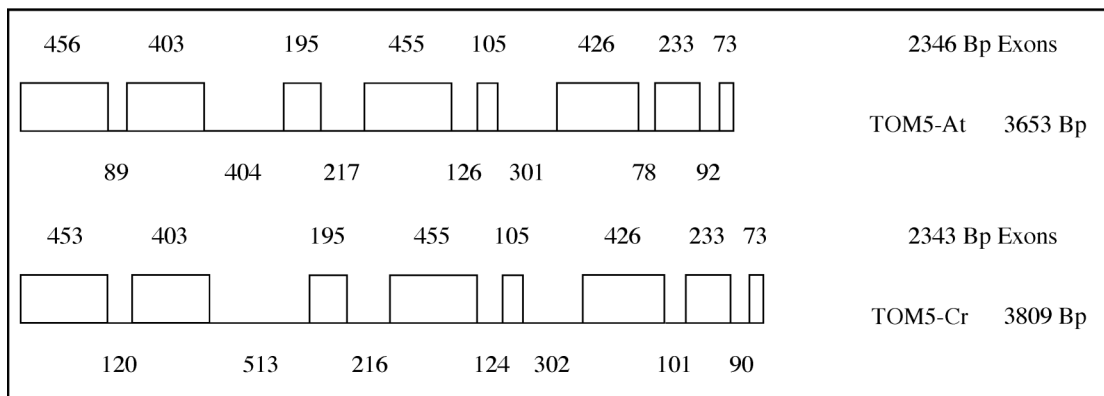
Die jeweiligen Exon- und Intron-Größen sind für den Bereich zwischen Start- und Stop-Codon angegeben. Außerdem sind die Gesamtlängen der Gene und die Summe der Länge der Exons rechts neben den Genstrukturen angezeigt. In der Abbildung ist *Arabidopsis thaliana* mit At und *Capsella rubella* mit Cr abgekürzt.



C



D



E

Abbildung 3.26B

Zusammenfassend wurde für das TOM-Contig in *C. rubella* festgestellt, daß die intergenischen Bereiche kürzer und die Gensequenzen in der Regel länger sind als in *A. thaliana*.

Nachdem ein hoher Konservierungsgrad für die Genstrukturen der Gene TOM 1-4 und 22916 in *C. rubella* im Vergleich zu *A. thaliana* festgestellt wurde, wurde der Identitätsgrad der Gensequenzen ermittelt. Auch hier wurde ein hoher Konservierungsgrad für die Gene in *C. rubella* und in *A. thaliana* festgestellt. Die Daten dieser Analyse sind in Tabelle 3.17 gezeigt.

Gen	Sequenzidentität auf Nukleotidebene	Sequenzidentität auf Proteinebene
TOM 1	92,6%	91,4%
TOM 2	94,1%	95,1%
TOM 3	80,3%	65,2%
TOM 4	92,3%	94,1%
Gen 22916	93,4%	95,4%

Tabelle 3.17: Die ermittelten Daten für die Sequenzidentität auf Nukleotid- und Proteinebene für die Gene TOM 1-4 und 22916 sind angegeben. Während für den Vergleich im Fall von TOM 1-4 die vorhergesagte Gensequenz aus der Datenbank eingesetzt wurde, wurde für TOM 5 die Sequenz von cDNA 22916 verwendet. Berücksichtigt wurden hierfür die Sequenzen zwischen Start- und Stop-Codons.

Die Sequenzidentitäten auf Nukleotid-Ebene liegen zwischen 80,3% und 94,1%, die Werte auf Protein-Ebene zwischen 65,2% und 95,4%. Die geringste Sequenzidentität auf Nukleotid- bzw. Protein-Ebene wurde für das Gen TOM 3 ermittelt und für Gen 22916 die höchsten Identitätswerte. Mit Ausnahme von TOM 3 ist der Identitätsgrad in den einzelnen Genen sowohl für die Nukleotid- als auch die Protein-Sequenzen sehr ähnlich und liegt jeweils über 90%.

4. Diskussion

Die im Rahmen dieser Arbeit durchgeführten Untersuchungen sollten nachweisen, in welchem Maße es möglich ist, Daten über die strukturelle Organisation des *A. thaliana*-Genoms auf andere diploide Arten der Familie der Brassicaceae zu übertragen. Diese Fragestellung erfordert die vergleichende Genomanalyse von zwei Arten. *Arabidopsis* und *Capsella* wurden für die Vergleiche ausgewählt. Es wurden Kopplungsgruppen vergleichend betrachtet, um zu zeigen, ob Kolinearität in den diploiden Brassicaceae-Vertretern über größere Chromosomensegmente vorliegt. Die vergleichende physikalische Genomanalyse kleiner ausgewählter Genombereiche wurde dann eingesetzt, um nachzuweisen, ob der in genetischen Experimenten gefundene Grad der Kolinearität in ähnlicher Weise auf der molekularer Ebene beobachtbar ist.

4.1 Vergleichende genetische Kartierungsexperimente

Vergleichende genetische Kartierungsexperimente geben Auskunft über die Konservierung des Markerrepertoires in den untersuchten Arten. Weiterhin erlauben sie festzustellen, ob kolineare Chromosomensegmente in den Spezies vorliegen. Außerdem kann die Größe der kolinearen Bereiche abgeschätzt werden.

4.1.1 Konservierung des Markerrepertoires in *Arabidopsis* und *Capsella*

Es wurden für die genetische Kartierungsanalyse in *Capsella Arabidopsis thaliana*-RFLP-Marker (mi...-, Liu *et al.*, 1996; m...-Marker, Fabri und Schöffner, 1994) und cDNA-Klone verwendet. Die Hybridisierungsanalysen dieser Marker zeigten, daß für 40 der 46 getesteten Marker in *Capsella* homologe Sequenzen vorliegen.

Zu den Markern, für die keine Hybridisierung mit *Capsella*-DNA nachgewiesen werden konnten gehören mi233 und EST 11177. Für RFLP-Marker mi233 und weitere sieben mi-RFLP-Marker war bekannt, daß diese Sequenzen in einigen Ökotypen von *Arabidopsis thaliana* fehlen. Während die mi233-Sequenz in den Ökotypen Columbia, Rschew und Wassilewskija vorliegt, fehlt sie in Landsberg *erecta*. Es handelt sich

wahrscheinlich bei dieser Gruppe von Markern um nicht-kodierende Bereiche des Genoms, die in einigen Ökotypen deletiert wurden (Liu *et al.*, 1996). Im Rahmen der in dieser Arbeit durchgeführten Mikrosyntenieuntersuchungen wurde festgestellt, daß kodierende Bereiche des Genoms eine sehr viel höhere Sequenzidentität in *Arabidopsis* und *Capsella* aufweisen als intergenische Regionen (siehe unten). Daher ist davon auszugehen, daß Marker die intergenische Regionen des *Arabidopsis*-Genoms repräsentieren, in den durchgeführten Hybridisierungsexperimenten in *Capsella* nicht nachgewiesen werden können.

EST 11177 hat hohe Sequenzhomologien zur LTR (*long terminal repeat*)-Region eines Retrotransposon-ähnlichen Elementes in *Arabidopsis thaliana*. Während dieser Marker in *A. thaliana* zu mehreren Kopien im Genom hybridisiert, konnte in *Capsella rubella* keine homologe Sequenz gefunden werden (Abb. 3.2). Der Konservierungsgrad für repetitive Sequenzen wie die Retrotransposon-ähnlichen Elemente ist zwischen verschiedenen Arten weniger hoch als für die kodierenden Sequenzbereiche des Genoms. Für das Retroelement *Reina*, welches in der Nähe des *Adh*-Gens in Mais lokalisiert ist, konnte in Hirse keine Kreuzhybridisierung gezeigt werden (Avramova *et al.*, 1996). Auch für weitere Retroelemente aus einer 225 Kbp-Region um das *Adh*-Gen in Mais, die zu 74% aus LTR-Retroelementen besteht, wurden in der korrespondierenden Hirse-Region keine homologen Sequenzen gefunden (Tikhonov *et al.*, 1999). Das Fehlen homologer Sequenzen von EST 11177 in *Capsella* steht im Einklang mit diesen Ergebnissen.

Die ermittelten Hybridisierungsdaten für die getesteten Marker zeigen somit einen sehr hohen Konservierungsgrad bezüglich des Markerrepertoires in *A. thaliana* und *Capsella*. Da 14 von 46 Markern cDNA-Sequenzen repräsentieren (Tabelle 3.3) und ein großer Teil der verwendeten 32 RFLP- bzw. ARMS-Marker Sequenzhomologien zu EST-Sequenzen in der *A. thaliana*-Datenbank zeigen (Daten nicht gezeigt), kann bereits aus diesen Daten ein hoher Konservierungsgrad bezüglich des Genrepertoires abgeleitet werden.

Der gefundene, hohe Konservierungsgrad des Markerrepertoires entspricht den Erwartungen, denn dies wurde bereits in anderen vergleichenden Untersuchungen beschrieben. In der Familie der Solanaceae konnte dies für Tomate - Kartoffel (Bonierbale *et al.*, 1988; Gebhardt *et al.*, 1991; Tanksley *et al.*, 1992) und Tomate - Paprika (Tanksley *et al.*, 1988; Livingstone *et al.*, 1999) gezeigt werden. Auch in der Familie der Poaceae wurde ein hoher Konservierungsgrad bezüglich des Genrepertoires

festgestellt (zusammengefaßt in Devos und Gale, 1997). In der Familie der Brassicaceae zeigen die Vergleiche verschiedener *Brassica*-Arten ähnliche Ergebnisse (Lagercrantz und Lydiate, 1996; Bohuon *et al.*, 1996). Kartierungsexperimente, in denen *Arabidopsis thaliana* und die *Brassica*-Arten *B. napus*, *B. nigra*, *B. rapa* und *B. oleracea* verglichen wurden, zeigen auch für diese eng verwandten Arten das Vorliegen eines sehr ähnlichen Genrepertoires (Conner *et al.*, 1998; Cavell *et al.*, 1998; Lagercrantz *et al.*, 1996; Lagercrantz, 1998; Teutonico und Osborn, 1994; Kowalski *et al.*, 1994).

4.1.2 Kolineare Chromosomensegmente in *Arabidopsis* und *Capsella*

Für die genetische Kartierung wurde in *Capsella* eine Kartierungspopulation aus 50 F₂-Pflanzen durch eine interspezifische Kreuzung von *C. grandiflora* und *C. rubella* etabliert. *A. thaliana* Chromosom IV RFLP-Marker und cDNAs wurden als Sonden für genetische Kartierungsexperimente in *Capsella* verwendet. Ein Vergleich der resultierenden Kopplungsgruppen erlaubt Aussagen über den Konservierungsgrad bezüglich der Chromosomen-Organisation in *Capsella* und *A. thaliana*. Während die *Capsella*-Arten im haploiden Chromosomensatz 8 Chromosomen besitzen (Hurka und Neuffer, 1997), hat *A. thaliana* n=5 Chromosomen (Koorneef *et al.*, 1983).

In *Capsella* wurden zur Kartierung 22 *A. thaliana* Chromosom IV-Marker eingesetzt (Tabelle 3.6), von denen vier Marker (3486, ATTS2212, ATTS3374, mi330) zu je zwei verschiedenen polymorphen Loci im *Capsella*-Genom korrespondieren. Zwei Kopplungsgruppen konnten mit 21 der 26 Loci etabliert werden. Drei der Marker (ATTS2212, ATTS3374 und mi330), für die im *Capsella*-Genom zwei polymorphe Loci ermittelt wurden, sind mit einem Locus auf einer der erstellten Kopplungsgruppen repräsentiert. Lediglich Marker 3486, der auch zu zwei verschiedenen polymorphen Loci in *Capsella* korrespondiert, konnte auf keiner der beiden Kopplungsgruppen von *Capsella* lokalisiert werden. Da dieser Marker nicht nur zu polymorphen DNA-Fragmenten, sondern auch mit mehreren monomorphen DNA-Sequenzen in *Capsella* hybridisiert (Daten nicht gezeigt), ist nicht auszuschließen, daß einer dieser monomorphen Loci in einer der ermittelten Kopplungsgruppen von *Capsella* kartiert. Dies könnte durch den Einsatz anderer Restriktionsenzyme für die

Polymorphismusanalyse und weitere genetische Kartierungsexperimente geklärt werden.

Ein weiterer Marker, der auf *A. thaliana*-Chromosom IV lokalisiert ist und keiner der beiden Kopplungsgruppen in *Capsella* zugeordnet werden konnte, ist der RFLP-Marker mi390. Alle Marker, die zur genetischen Kartierung eingesetzt wurden, wurden sequenziert und in der *A. thaliana*-Datenbank auf die Position im *A. thaliana*-Genom hin überprüft. Lediglich die Sequenz von mi390 wurde bei dieser Analyse an einer von der molekularen Markerkarte (Liu *et al.*, 1996) abweichenden Position im Genom kartiert. Den Sequenzvergleichen zufolge ist mi390 auf *A. thaliana*-Chromosom II lokalisiert (Tabelle 3.1). Diese Ergebnisse erklären, warum der eingesetzte Marker nicht auf den für *Capsella* erstellten Kopplungsgruppen gefunden wurde.

In *Capsella* konnte gezeigt werden, daß 5 von 30 Markern zu mindestens zwei verschiedenen Loci korrespondieren. Durch Sequenzvergleiche in der *A. thaliana*-Datenbank für Marker, die zu zwei verschiedenen polymorphen Loci in *Capsella* kartieren, wurde festgestellt, daß jeder dieser Marker ein Mitglied einer Multigenfamilie repräsentiert. Zum Beispiel 3486 zeigt hohe Homologien zu α -Tubulin-Genen (Kopczak *et al.*, 1992) und ATTS3374 zu Ascorbatperoxidase-Genen in *A. thaliana* (Jespersen *et al.*, 1997). Jedoch liegt keine Evidenz für Duplikationen von Chromosomensegmenten vor wie zum Beispiel in den verschiedenen diploiden *Brassica*-Arten *B. oleracea*, *B. rapa* und *B. nigra* (Lagercrantz und Lydiate, 1996; Lagercrantz *et al.*, 1996; Lagercrantz, 1998). Auch in der Familie der Poaceae wurde für viele Arten gezeigt, daß ganze Chromosomen oder Chromosomensegmente dupliziert vorliegen. In Weizen liegen beispielsweise die meisten Gene tripliziert in den drei Weizen-Genomen A, B und D vor (Devos und Gale, 1993). Auch in Mais hybridisieren viele RFLP-Marker zu zwei Loci im Genom, was durch Duplikation von Chromosomenbereichen zu erklären ist (Helentjaris *et al.*, 1988).

Die beiden erstellten Kopplungsgruppen für *Capsella* zeigen Kolinearität zu *A. thaliana*-Chromosom IV für die Bereiche zwischen den Markern mi51 und m518A bzw. mi30 und mi369 (Abbildung 3.5). Die Größe der Kopplungsgruppen beträgt in *Capsella* 23,9 bzw. 63,9 cM und für die entsprechenden Bereiche in *A. thaliana* 30,2 bzw. 58,2 cM. Durch weitere genetische Kartierungsexperimente mit RFLP-Markern aus Centromer-Regionen der *A. thaliana*-Chromosomen wurden zwei Chromosom V-Marker in einer der *Capsella* Kopplungsgruppen gekoppelt zu mi30 gefunden (Oliver

Clarenz, Diplomarbeit Universität zu Köln, 2000). Die Kartierung eines weiteren Markers aus der Centromer-Region von *A. thaliana*-Chromosom IV (N97271) zeigte eine Inversion des Chromosomenbereichs zwischen Marker m518A und N97271 in *Capsella* relativ zu *Arabidopsis* (Oliver Clarenz, Diplomarbeit Universität zu Köln, 2000).

Zusammenfassend kann festgestellt werden, daß durch den Einsatz der *Arabidopsis* Chromosom IV-Marker zwei ausgedehnte kolineare Chromosomensegmente für *Arabidopsis* und *Capsella* ermittelt werden konnten. Es gibt keinerlei Hinweise für das Fehlen größerer Chromosomenbereiche in *Capsella*, denn bis auf die wenigen diskutierten Ausnahmen finden sich alle Markersequenzen sowohl in *Capsella* als auch in *Arabidopsis* wieder. Weiterhin gibt es keine Evidenzen für Duplikationen von Chromosomenbereichen. Da *Arabidopsis* und *Capsella* eine unterschiedliche Chromosomenzahl aufweisen, ist es nicht unerwartet, daß *Arabidopsis*-Chromosom IV in *Capsella* auf zwei verschiedenen Kopplungsgruppen vorliegt. Eine weitere Abweichung von der Genomkolinearität stellt die Inversion des Chromosomenbereichs zwischen Marker m518A und N97271 in *Capsella* relativ zu *Arabidopsis* dar.

Auch die Kopplungsanalysen, die mit Markern der *A. thaliana*-Chromosomen I-III und V durchgeführt wurden, wiesen kolineare Chromosomenbereiche nach. Zum Beispiel die Marker m235A und m254A aus *A. thaliana*-Chromosom I sowie die Marker m249A und m457A aus *A. thaliana*-Chromosom III liegen auch in *Capsella* gekoppelt vor. Für die Chromosomensegmente konnte eine ähnliche Größe wie in *A. thaliana* ermittelt werden (siehe Anhang). Außerdem konnten die verwendeten Marker der *A. thaliana* Chromosomen I-III und V in allen Fällen in die Kopplungsgruppen, die mit weiteren *Arabidopsis*-Markern dieser Chromosomen erstellt wurden, eingeordnet werden (Karine Boivin, unveröffentlicht). Die kolinearen Chromosomenblöcke, die in dieser Studie gefunden wurden (Karine Boivin, unveröffentlicht), sind ähnlich groß wie die, die in dieser Arbeit ermittelt wurden.

Trotz der unterschiedlichen Chromosomenzahl zwischen *A. thaliana* und *Capsella* ist für die beiden diploiden Arten also ein hoher Grad an Kolinearität zu beobachten. Lediglich zwei intrachromosomale Inversionen wurden in *Capsella* relativ zu *Arabidopsis* beobachtet (Oliver Clarenz, Diplomarbeit Universität zu Köln, 2000; Karine Boivin, unveröffentlicht). Diese Ergebnisse stehen im Gegensatz zu Untersuchungen, in denen *Arabidopsis* mit paleoploiden Arten der Brassicaceae verglichen wurde.

Zwischen *A. thaliana* und *B. oleracea*, für die ein konserviertes Genrepertoire festgestellt wurde, ist die Kolinearität als Folge zahlreicher Genom-Umorganisationen, die seit der Trennung der beiden Spezies stattgefunden haben, auf kleinere Chromosomenbereiche von 3,7 bis 49,6 cM begrenzt. Diese Unterschiede in der Anordnung der Marker sind auf mindestens 17 Translokations- und 9 Inversions-Ereignisse zurückzuführen (Kowalski *et al.*, 1994). In *A. thaliana* und einer weiteren *Brassica*-Art, *B. nigra*, wurde auch eine große Anzahl von Chromosomen-Umorganisationen seit der Trennung dieser beiden Spezies festgestellt. Ihre Zahl wird auf ca. 90 geschätzt. Durch die zahlreichen Umstrukturierungen des *B. nigra*-Genoms beträgt die durchschnittliche Größe der kolinearen Chromosomenbereiche nur ca. 8 cM (Lagercrantz, 1998).

Die Trennung der Vorläufer von *Arabidopsis* und *Capsella* erfolgte später als die Trennung von *Brassica* und *Arabidopsis*. Die Analyse der mitochondrialen Gene für die NADH-Untereinheit (*nad4*) ermittelte für die *Brassica*-Arten und *A. thaliana* einen Trennungszeitpunkt zwischen 14,5 bis 20 Millionen Jahren (Yang *et al.*, 1999). Die Trennungszeitpunkte von *Brassica* und *A. thaliana* wurden durch einen weiteren Vergleich der plastidären Maturase-Gene (*matK*) mit 16-20 Millionen Jahren bestimmt, während *Capsella* und *Arabidopsis* sich vor ca. 8-10 Millionen Jahren getrennt haben (M. Koch, B. Haubold und T. Mitchell-Olds, zur Veröffentlichung eingereicht). Dieser spätere Trennungszeitpunkt der Arten bietet eine Erklärung für die geringere Anzahl an Chromosomen-Umorganisationen, die für *Arabidopsis*-*Capsella* verglichen mit *Arabidopsis*-*Brassica* beobachtet wird.

Auch in der Familie der Solanaceae wurde bei genetischen Kartierungsuntersuchungen festgestellt, daß für den Vergleich verschiedener Arten unterschiedliche Anzahlen von Chromosomenumordnungs-Ereignissen gefunden wurden. Die Genome von Kartoffel und Tomate unterscheiden sich lediglich durch fünf intrachromosomale Inversionen, die durch Chromosomenbrüche in der Nähe oder am Centromer hervorgerufen wurden (Tanksley *et al.*, 1992). Im Gegensatz dazu unterscheidet sich das Paprika-Genom vom Tomaten-Genom trotz ähnlichen Genrepertoires und gleicher Chromosomenzahl ($n=12$) durch zahlreiche Inversionen und Translokationen (Prince *et al.*, 1992; Livingstone *et al.*, 1999). Auch diese Ergebnisse können durch unterschiedliche Trennungszeitpunkten für die untersuchten Artenpaare erklärt werden. Die Trennung der Kartoffel- und Tomatenarten erfolgte vor ca. 10 Millionen Jahren, während die

Vorläufer von Tomate und Paprika sich vor ca. 40 Millionen Jahren auseinanderentwickelt haben (Paterson *et al.*, 1996).

Trotz des Divergenzzeitpunktes von über 60 Millionen Jahren und der unterschiedlichen Genomgrößen zeigen genetische Kartierungsexperimente in Weizen, Mais, Reis und weiteren Arten der Poaceae dagegen einen hohen Grad an Konservierung bezüglich des Genrepertoires und der Kolinearität (Ahn *et al.*, 1993; Ahn und Tanksley, 1993; Devos *et al.*, 1993). Obwohl die Trennung der Vorläufer von *Brassica* und *Arabidopsis* später erfolgte als die Trennung der untersuchten Gräserarten (Moore *et al.*, 1995; Yang *et al.*, 1999) zeigen diese Brassicaceae-Arten, verglichen mit den Poaceae-Arten, einen viel höheren Grad an Genom-Umorganisationsereignissen (Lagercrantz, 1998).

In polyploiden Pflanzen muß die Paarung homeologer Chromosomen verhindert werden, um aberrante Chromosomensegregationen in der Meiose zu vermeiden. Arten der Poaceae zeigen einen hohen Anteil repetitiver Elemente im Genom (Gale und Devos, 1998), dies könnte dazu beitragen, die Paarung homeologer Chromosomen zu unterdrücken. Die *Brassica*-Arten besitzen kleinere Genome (Arumuganathan und Earle, 1991) und damit einen sehr viel geringeren Anteil repetitiver Elemente. In diesen Arten könnte die Paarung homeologer Chromosomen durch die hochgradige Umorganisation der Chromosomen gewährleistet sein. Diese Spekulation bietet einen Erklärungsansatz, für die verglichen mit den polyploiden Brassicaceae-Arten, stärker ausgeprägte Kolinearität in der Familie der Poaceae.

4.2 Mikrosyntenieuntersuchungen in *Arabidopsis* und *Capsella*

Die vergleichende physikalische Genomanalyse sollte zeigen, ob die festgestellte Markerkolinearität auf chromosomaler Ebene auch auf physikalischer Ebene vorzufinden ist. Dabei war das Ziel festzustellen, ob beispielsweise in der Region zwischen zwei kolinearen Markern eine Konservierung des Genrepertoires und unter Umständen der Genreihenfolge zu beobachten ist. Der vergleichenden Kartierung auf physikalischer Ebene sind Grenzen in der Auflösung gesetzt. Nur vergleichende Sequenzierungsexperimente von orthologen Bereichen in *C. rubella* können umfassende Daten über den Konservierungsgrad der Genorientierung, der Genstrukturen und der Sequenzidentität zwischen *A. thaliana* und *C. rubella* ermitteln.

Die geplanten detaillierten Mikrosyntenieuntersuchungen konnten im Rahmen dieser Arbeit nur geleistet werden, indem die Experimente auf zwei kleine Genombereiche beschränkt wurden. Die Untersuchung von zwei verschiedenen Regionen wurde durchgeführt, um generellere Aussagen zur Mikrosyntenie in den untersuchten Arten machen zu können. Die erste ca. 200 Kbp große Region kartiert auf dem langen Arm von *A. thaliana* Chromosom IV zwischen den Markern m326A und mi330 (Lister und Dean, 1993; http://nasc.nott.ac.uk/new_ri_map.html) und ist vollständig sequenziert (Bevan *et al.*, 1998; Mayer *et al.*, 1999). Die zweite Region liegt in *A. thaliana* auf Chromosom I auf dem sequenzierten BAC-Klon F20N2 (AC002328) in Nachbarschaft zum genetischen Marker nga280 (81.7 cM; http://nasc.nott.ac.uk/new_ri_map.html). Für die genetische Kartierung in *C. rubella* wurde EST 21484 verwendet, welcher ca. 100 Kbp von nga280 lokalisiert ist.

4.2.1 Konservierung des Markerrepertoires und der Markeranordnung in *Arabidopsis* und *Capsella*

Für die 200 Kbp-Region aus *A. thaliana* wurden durch Hybridisierung einer *A. thaliana* cDNA-Bibliothek (Newman *et al.*, 1994) elf verschiedene cDNA-Klone (A-K) identifiziert, die hohe Homologien zu Genen dieser Region zeigten. Ein Vergleich der EST-Sequenzdaten mit den genomischen Sequenzen beider Regionen ermittelte 11 EST-Klone. Die korrespondierenden Gene für 8 ESTs konnten in der 200 Kbp-Region lokalisiert werden, während für die Chromosom I-Region 3 ESTs, die zwei verschiedenen Genen entsprechen, isoliert wurden.

Alle Klone wurden eingesetzt, um homologe Sequenzen in *Capsella rubella* zu identifizieren. Die Ergebnisse dieser Experimenten zeigten einen ebenso hohen Konservierungsgrad bezüglich des Genrepertoires wie er in den vergleichenden genetischen Kartierungsexperimenten ermittelt wurde (siehe 4.1.1). Mit Ausnahme von cDNA-Klon K und EST 8 konnten für die restlichen 20 cDNA-Klone, die in den ausgewählten *A. thaliana* Regionen lokalisiert sind, die homologen Sequenzen in *C. rubella* identifiziert werden. Der EST-Klon 8 entspricht dem genetischen Marker 11177, für den bereits diskutiert wurde, daß keine homologe Sequenz in *C. rubella*

vorliegt und daß dieser Klon Homologien zur LTR-Sequenz eines Retrotransposon-ähnlichen Elementes aus *A. thaliana* aufweist.

Durch Vergleich der 19 verschiedenen cDNA-Klon-Sequenzen aus der 200 Kbp-Region mit den vorhergesagten Genen in der *A. thaliana*-Datenbank (<http://websvr.mips.biochem.mpg.de/proj/thal/>) konnten für alle 19 cDNA-Klone Sequenz-Homologien zu vorhergesagten Genen ermittelt werden. Eine der cDNA-Sequenzen zeigte dabei Homologie zu zehn in Tandem organisierten homologen Sequenzen (Bevan *et al.*, 1998). In dieser analysierten 200 Kbp-Region sind jedoch weitere acht vorhergesagte Gene lokalisiert, die weder durch Hybridisierung der *A. thaliana* cDNA-Bibliothek noch durch Vergleich der genomischen Sequenz von *A. thaliana* mit den EST-Sequenzdaten identifiziert werden konnten. Da die Gendichte für *A. thaliana* mit einem Gen pro 4,1 bis 4,6 Kbp ermittelt wurde (Lin *et al.*, 1999; Mayer *et al.*, 1999; Sato *et al.*, 1998), ist davon auszugehen, daß nicht alle kodierenden Sequenzen der Region mit Hilfe der EST-Analysen und dem Durchsuchen der cDNA-Bibliotheken identifiziert wurden. Zum Beispiel ist die Region zwischen Gen 4 und B in *A. thaliana* ca. 24,5 Kbp groß. Jedoch konnten für diesen Bereich keine cDNA-Klone oder ESTs identifiziert werden. In einer Region dieser Größe wären statistisch 4 bis 5 Gene lokalisiert. In der *A. thaliana*-Datenbank sind für diesen Bereich drei Gene vorhergesagt (<http://websvr.mips.biochem.mpg.de/proj/thal/>). Der Vergleich der genomischen Sequenz von *C. rubella* mit den Sequenzen der vorhergesagten Gene zeigte, daß zumindest in zwei Fällen die Sequenzbereiche, für die Genvorhersagen vorliegen, in *C. rubella* einen hohen Sequenzhomologiegrad aufweisen.

Es war auch nicht zu erwarten, daß die verwendeten Ansätze zur Identifizierung von cDNA-Sequenzen aller Gene der Region führen würde. Trotz des Umfangs der *Arabidopsis*-EST-Kollektionen liegen für einen großen Teil der transkribierten Sequenzen in *Arabidopsis* keine EST-Sequenzen vor (Mayer *et al.*, 1999; Lin *et al.*, 1999). Für die Identifizierung der cDNA-Sequenzen mittels eines Hybridisierungsansatzes wurde eine ca. 200000 Klone umfassende Bibliothek eingesetzt, daher ist es unwahrscheinlich, daß sehr gering exprimierte Gene nachgewiesen werden (Schmidt *et al.*, 1999).

In der zweiten 30 Kbp-Region auf *A. thaliana* Chromosom I wurden drei ESTs identifiziert. In der *A. thaliana*-Datenbank sind für diese Genomregion sechs Gene vorhergesagt. Daher beträgt die durchschnittliche Gendichte ein Gen pro 5 Kbp. Dieser

Wert liegt im Rahmen der veröffentlichten Daten von einem Gen pro 4,1-4,6 Kbp (Mayer *et al.*, 1999; Lin *et al.*, 1999).

Es konnte gezeigt werden, daß in beiden Regionen die Anordnung der in *C. rubella* vorliegenden Gene identisch zu der in *A. thaliana* beobachteten ist. Auch die intergenischen Bereiche sind in beiden Spezies ähnlich groß. Gene, die in *A. thaliana* physikalisch eng benachbart vorliegen, sind in *C. rubella* ebenfalls physikalisch eng benachbart. Auch in *B. oleracea*, diese Art besitzt eine Genomgröße von ca. 618 Mbp (Arumuganathan und Earle, 1991), ist das Genrepertoire der 200 Kbp-Region aus *A. thaliana* im wesentlichen konserviert. Jedoch sind die intergenischen Bereiche in *B. oleracea* in der Regel größer als in *A. thaliana* oder *C. rubella* (M. Roßberg, unveröffentlicht). Die Analyse der Region um das *RPM1*-Gen in *A. thaliana* und *Brassica napus* zeigte, daß die Reihenfolge der Gene *GTP*, *RPM1* und *M4* identisch ist und die intergenischen Bereiche eine ähnliche Größe aufweisen (Grant *et al.*, 1998). Auch für eine homologe 15 Kbp große Region des *A. thaliana* Chromosoms III ist in den drei *Brassica*-Arten *B. nigra*, *B. rapa* und *B. oleracea* die Reihenfolge der fünf Gene konserviert. Die Größe der homologen Regionen in den *Brassica*-Arten wurde mit Hilfe von Southern-Blot-Analyse von Pulsfeld-aufgetrennter DNA bestimmt und beträgt maximal 320 Kbp (Sadowski *et al.*, 1996).

Wie in dieser Arbeit für zwei verschiedene Regionen in *C. rubella* und *A. thaliana* beobachtet, zeigen auch die Mikrosyntenieanalysen in Mais, Reis und Hirse für die *sh2*-Region Konservierung bezüglich der Genanordnung der beiden Gene *sh2* und *a1* in den homologen Regionen. Die Größe der intergenischen Bereiche in diesen Arten ist dagegen gering konserviert. Während der Abstand zwischen den beiden Genen in Mais ca. 140 Kbp umfaßt, beträgt die Distanz in Reis und Hirse lediglich ca. 19 Kbp (Chen *et al.*, 1997). Auch für die 225 Kbp große *Adh*-Region in Mais wurde in Hirse Kolinearität für die neun bekannten bzw. vorhergesagten Gene, die in dieser Region lokalisiert sind, gezeigt. Die homologe Region in Hirse ist jedoch lediglich 78 Kbp groß und enthält weitere fünf Gene (Tikhonov *et al.*, 1999). Diese Daten lassen vermuten, daß die Länge der intergenischen Bereiche mit der Genomgröße der Pflanzen korreliert ist (Bennetzen *et al.*, 1998), da die Genomgrößen für Mais ca. 2292 Mbp, für Hirse ca. 748 Mbp und für Reis ca. 415 Mbp betragen (Arumuganathan *et al.*, 1991). Doch die Analyse der *Lrk*-Region in den vier Vertretern der Gräser Weizen (Genomgröße ca. 16000 Mbp), Gerste (Genomgröße ca. 5000 Mbp), Mais und Reis zeigte, daß trotz der sehr unterschiedlichen Genomgrößen die intergenischen Bereiche in der *Lrk*-Region in

Weizen, Reis und Gerste eine ähnliche Länge haben (vgl. Einleitung; Feuillet und Keller, 1999). In diesen chromosomalen Bereichen ist auch die Gendichte mit einem Gen pro 4-5 Kbp sehr ähnlich zu der beobachteten Gendichte in *A. thaliana* (Lin *et al.*, 1999; Mayer *et al.*, 1999).

Die Genomgröße von *A. thaliana* wird nach den neuesten Daten auf ca. 130 Mbp geschätzt (Mayer *et al.*, 1999). Für *Capsella rubella*, die eine diploide Verwandte der tetraploiden *C. bursa-pastoris* ist, liegen bisher keine Daten zur Genomgröße vor. Jedoch wurde die Genomgröße der tetraploiden *C. bursa-pastoris* mit ca. 680 Mbp bestimmt (Bennett und Smith, 1976). Unter Berücksichtigung des Ploidiegrades zwischen den beiden *Capsella*-Arten kann die Genomgröße auf ca. 340 Mbp geschätzt werden. Jedoch deuten die Hybridisierungsergebnisse der *C. rubella*-Bibliothek, die zur Identifizierung der *C. rubella* Cosmid-Klone eingesetzt wurde, auf eine 4- bis 5-fache Genomäquivalenz hin. Da die genomische Bibliothek aus 46000 Klonen mit einer durchschnittlichen DNA-Insertionsgröße von ca. 20 Kbp besteht und dies ca. 920 Mbp genomischer *C. rubella*-DNA in der Bibliothek entspricht, wäre die Genomgröße von *C. rubella* mit 340 Mbp überschätzt. Diese Hypothese wird auch dadurch gestützt, daß die untersuchten genomischen Regionen in *A. thaliana* und *C. rubella* eine sehr ähnliche Größe aufweisen. Wahrscheinlicher ist ein Wert für die Genomgröße von *C. rubella* in der Größenordnung von 200 Mbp.

In *C. rubella* ist die homologe Region zur 200 Kbp-Region in *A. thaliana* durch drei Cosmid-Contigs repräsentiert. Genetische Kartierungen der Marker EST 5760 (Contig 0), 54-I/54X06 (Contig I) und cDNA J (Contig II) zeigen, daß alle drei Contigs genetisch eng gekoppelt sind (siehe Anhang). Insgesamt umfaßt der bisher sequenzierte Bereich ca. 144 Kbp. Die 30 Kbp-Region setzt sich aus einem Contig mit zwei Cosmiden zusammen, das eine genomische Gesamt-Sequenzlänge von 27056 Bp aufweist. Die Analyse der genomischen *C. rubella*-Sequenzen zeigt, daß neben der Konservierung des Genrepertoires und der Genreihenfolge auch die Transkriptionsrichtung der Gene in *C. rubella* und *A. thaliana* identisch ist. Unterschiede zwischen den Regionen von *A. thaliana* und *C. rubella* sind neben den fehlenden Sequenzen von cDNA K und EST 8 in *Capsella rubella* die Duplikationen der Gene C und 7 in *Capsella rubella*. Die beiden Kopien von Gen C sind dabei in Tandem organisiert. Der Vergleich der *sh2/a1*-homologen Regionen in Reis, Mais und Hirse zeigt ebenfalls eine Duplikation von *a1* in Reis und Hirse, die auch tandemartig organisiert ist (Chen *et al.*, 1997; 1998). Die Duplikation von Gen C konnte nur in *C.*

rubella beobachtet werden. In der homologen Region in *Brassica oleracea* liegt Gen C ebenso wie in *A. thaliana* als Einzelkopie-Sequenz vor (M. Roßberg, unveröffentlicht). Die Daten deuten auf eine Duplikation von Gen C in *C. rubella* nach der Trennung der Vorläufer von *Capsella* und *Arabidopsis* hin, dies konnte durch die Untersuchung der abgeleiteten Aminosäuresequenzen der Gene erhärtet werden (Abbildung 3.21). Die Duplikation von *a1* in Hirse scheint dagegen vor der Trennung von Mais, Reis und Hirse stattgefunden zu haben, da in Reis ebenfalls eine zweite Kopie von *a1* gefunden wurde (Chen *et al.*, 1998).

Durch Analysen auf Expressionsebene könnten Aussagen zur Funktionalität der beiden Genkopien von Gen C in *Capsella rubella* gemacht werden. Aus den Sequenzdaten alleine ist keine Aussage möglich, ob eventuell eine Genkopie für ein nicht-funktionelles Protein bzw. ein Protein mit veränderter Funktion kodiert.

Homologe Sequenzen von *RPP5* (Parker *et al.*, 1997) liegen in der 200 Kbp-Region von *A. thaliana* Ökotyp Columbia in einem Block aus zehn tandemartig organisierten Kopien vor (Bevan *et al.*, 1998), während in *C. rubella* bisher lediglich zwei Teilsequenzen des Gens in den Contigs I und II identifiziert wurden. Es kann aber nicht ausgeschlossen werden, daß weitere Kopien in *C. rubella* vorliegen, da der Bereich zwischen Contig I und Cosmid 54-II (Contig II) nicht sequenziert wurde. Die beiden Contigs I und II überlappen mit lediglich 29 Bp. Da die PCR-Analysen keine eindeutige Evidenz für den überlappenden Bereich zwischen den beiden Contigs lieferten, ist nicht auszuschließen, daß zwischen Contig I und II weitere Cosmidklone lokalisiert sind, die weitere Kopien des *RPP5*-Homologs tragen. Durch Identifizierung dieser weiteren Cosmid-Klone könnten in *C. rubella* auch Aussagen über die Kopienzahl der *RPP5*-homologen Sequenzen gemacht werden. Eine solche weiterführende Analyse wäre besonders interessant, da sie die vergleichende Analyse von Resistenzgen-ähnlichen Sequenzen in *Arabidopsis* und *Capsella* ermöglichen würde. Für andere Arten wurde gezeigt, daß in solchen Regionen des Genoms ein höherer Grad an Umordnungen stattfindet.

In Reis, *A. thaliana*, Tomate und Kartoffel sind die Resistenz-Gene zusammen mit den Resistenzgen-ähnlichen Sequenzen (RGA) in Blöcken („*Clustern*“) organisiert (Leister *et al.*, 1996; Leister *et al.*, 1998; Botella *et al.*, 1997). Resistenz-Genloci zeigen im Vergleich zum restlichen Genom eine viel schnellere Umorganisation (Leister *et al.*, 1998). Duplikationen, Deletionen bzw. Insertionen scheinen die Mechanismen zu sein, die zu den Unterschieden in der Resistenzgen-Organisation führen. Die Organisation

der Kopien in tandemartigen Wiederholungen kann als ein Hinweis für die Entstehung der Gene durch Genduplikationen mit anschließenden ungleichen Rekombinationen verstanden werden (Ronald, 1998). Im Ökotyp *L. erecta* liegen zum Beispiel zwei Kopien von *RPP8* im Genom vor, von denen *RPP8-Ler* das funktionelle Gen und *RPH8A* ein nicht-funktionelles Homolog ist. Der Ökotyp Columbia, der kein funktionelles *RPP8*-Gen besitzt, trägt ein chimäres Gen, das aus ungleicher Rekombination zwischen den Vorläufern von *RPP8-Ler* und *RPH8A* in einem *Ler*-ähnlichen Haplotyp hervorgegangen ist (McDowell *et al.*, 1998). In *Brassica napus* und *A. thaliana*-Ökotypen wurden unabhängige Deletionsereignisse für die *RPM1*-Gene nachgewiesen. Die Deletionen sind dabei auf das Resistenzgen beschränkt, da die Kolinearität der flankierenden Marker in der Region erhalten ist (Grant *et al.*, 1998).

4.2.2 Vergleichende Sequenzuntersuchungen

In *A. thaliana* wurden die Sequenzen der 17 Gene der 200 Kbp-Region vom Start- bis Stop-Codon ermittelt, um eine weitere Analyse der Gene bezüglich des Konservierungsgrades in der Exon-Intron-Struktur und der Sequenzhomologie zu ermöglichen. Hierfür wurden die Sequenzdaten der *A. thaliana* cDNA-Klone mit vorhergesagten Gensequenzen aus der *Arabidopsis*-Datenbank (<http://websvr.mips.biochem.mpg.de/proj/thal/>) verglichen. Der Vergleich zeigt, daß die vorhergesagten Exon-Intron-Strukturen in 10 von 17 Fällen in *A. thaliana* nicht mit cDNA-Sequenzen übereinstimmen. Die Genvorhersagen konnten aber in einigen Fällen zur Ermittlung der vollständigen kodierenden Sequenz ausgenutzt werden, wenn, wie in dieser Arbeit für 3 cDNAs gezeigt, relativ kleine Sequenzabschnitte in der 5'-Region des Gens fehlen. Eine weitere cDNA-Sequenz, die nicht mit der vorhergesagten Gensequenz Übereinstimmung zeigt, ist Gen 22916 aus der auf *A. thaliana* Chromosom I lokalisierten 30 Kbp-Region. In der Sequenz der Genvorhersage von TOM 5 war ein zusätzlicher Sequenzabschnitt, der in der Sequenz des cDNA-Klons 22916 nicht vorliegt.

Die Ermittlung der Exon-Intron-Strukturen für die homologen Gene in *C. rubella* zeigt einen hohen Konservierungsgrad bezüglich der Exon-Intron-Struktur in *A. thaliana* und *C. rubella*. Dies konnte genutzt werden, um durch Sequenzvergleiche der genomischen

Arabidopsis- und *Capsella*-Sequenzen Genvorhersagen für zwei Gene abzuleiten (Gen 3 und Gen 7), für die *Arabidopsis* cDNA-Klone vorlagen, die lediglich einen Teilbereich der kodierenden Region umfaßten. Diese Ergebnisse belegen, daß die vergleichende Sequenzierung des *C. rubella*-Genoms zur Identifizierung von Genbereichen in *A. thaliana* ausgenutzt werden kann. Ein solcher Ansatz ist umso wichtiger, weil ein Vergleich der cDNA-Sequenzen mit den vorhergesagten Genen zahlreiche Diskrepanzen aufzeigt. Allerdings geben diese Daten nur Hinweise auf das Vorliegen von Gensequenzen und müssen experimentell überprüft werden.

Mit Ausnahme von TOM 3 zeigen die Exonsequenzen der *C. rubella* Gene sehr hohe Sequenzidentität auf Nukleotid- (>87,5%) bzw. Aminosäure-Ebene (>83,5%) zu den homologen Genen in *A. thaliana*. TOM 3 zeigt in *Capsella rubella* im Vergleich zu *Arabidopsis thaliana* lediglich eine Sequenzidentität von 80% auf Nukleotid-Ebene und 65% auf Protein-Ebene. Expressionsstudien für die vorhergesagten TOM 3-Gene in *C. rubella* und *A. thaliana* könnten Auskunft darüber geben, ob es sich um funktionelle Gene handelt. Wenn in beiden Arten Expression der vorhergesagten TOM 3-Gene nachgewiesen würden, so wäre es außerordentlich interessant festzustellen, welche Funktion diese Gene haben und ob sie noch ähnliche oder gleiche Funktionen ausüben, da sie verglichen mit allen anderen untersuchten Sequenzen sehr viel umfangreicheren Sequenzveränderungen unterworfen wurden.

Neben den ermittelten Gensequenzen, die ein offenes Leseraster besitzen, liegen in der untersuchten Region cDNA-Sequenzen (zum Beispiel: cDNAs A, H und 6) vor, für die zwar eine Konservierung der Exon-Intron-Strukturen in *A. thaliana* und *C. rubella* gezeigt wurde, die aber kein längeres offenes Leseraster aufweisen. Für diese cDNA-Klone konnten in der *A. thaliana*-Datenbank keine Genvorhersagen gefunden werden, da diese auf der Identifizierung längerer offener Leseraster beruhen. Dies zeigt deutlich, daß vergleichende Genomanalysen dazu beitragen können, solche Sequenzen zu identifizieren. Bei diesen cDNA-Sequenzen kann es sich um mRNA-ähnliche, nicht-kodierende Sequenzen handeln. Diese Gruppe von transkribierten Sequenzen wurde sowohl bei Pflanzen als auch bei Tieren beobachtet. Es handelt sich dabei um stabile Transkripte, die gespleißt und polyadenyliert werden, jedoch kein längeres offenes Leseraster (ORF) besitzen (Erdmann *et al.*, 1999; Olivas *et al.*, 1997). Es wird vermutet, daß die Transkripte regulatorische Funktionen haben. Zu dieser Gruppe gehören unter anderem die Vertreter der ENOD40-Gene (Crespi *et al.*, 1994), Systemin (Barciszewski und Legocki, 1997) und CR20 (Teramoto *et al.*, 1996), die bereits in vielen Pflanzen-

Arten identifiziert wurden und als Signalträger bei biotischem Stress aktiv sind (Erdmann *et al.*, 1999). In *A. thaliana* sind bisher zwei nicht-kodierende Transkriptsequenzen bekannt, die den CR20- bzw. GUT15-Genen zugeordnet werden. Die Länge der Transkripte beträgt 758 bzw. 1377 Bp, wobei im 1377 Bp-Transkript eine kleine Region mit offenem Leseraster vorhanden ist (Erdmann *et al.*, 1999; <http://www.man.poznan.pl/5Ddata/ncRNA/index.html>). In Gurke wurden mehrere Transkripte des CR20-Gens identifiziert, die durch alternatives Spleißen des zweiten Introns der prä-RNA hervorgehen. Der Vergleich der CR20-RNA aus Gurke und AtCR20-1 aus *A. thaliana* zeigt Konservierung für eine 180 Bp-Region, wobei diese aus sieben Elementen besteht, die das Potential zur Bildung von Sekundärstrukturen besitzen (Teramoto *et al.*, 1996). Das Potential zur Sekundärstrukturbildung ist ein charakteristisches Merkmal der Regulator-Elemente. Aufgrund der niedrigen Sequenzhomologie dieser Transkripte zwischen verschiedenen Arten konnten die cDNA-Sequenzen aus *A. thaliana*, die kein offenes Leseraster besitzen, zu keiner der bereits bekannten Sequenzen in der Datenbank zugeordnet werden.

Zusammenfassend kann gesagt werden, daß die vergleichenden Sequenzanalysen sehr zum Verständnis der Mikrosyntenie beigetragen haben. Nur mit Hilfe einer solch detaillierten Analyse konnte nachgewiesen werden, daß nicht nur homologe Sequenzen zu den untersuchten *Arabidopsis*-Genen in *Capsella rubella* vorliegen, sondern daß die Struktur der Exon-Sequenzen weitgehend erhalten ist. Dies steht interessanterweise im Gegensatz zu Ergebnissen, die der Vergleich einiger Gene in *Arabidopsis thaliana* und *Brassica oleracea* ergab (Mathias Roßberg und K. Boivin, unveröffentlicht). Auch das Studium der duplizierten Gene kann nur auf Sequenzebene erschöpfend untersucht werden. Ebenso würden Sequenzbereiche, die geringere Sequenzidentitäten aufweisen als die meisten anderen untersuchten Genbereiche, wie z. B. TOM 3, mit den angewendeten Stringenzkriterien in Hybridisierungsexperimenten detektiert werden.

Die ermittelten vergleichenden Sequenzdaten bieten die Möglichkeit, auch die untersuchten *Arabidopsis*-Bereiche genauer zu charakterisieren, da in dieser Arbeit gezeigt werden konnte, daß in vielen Fällen vergleichende Daten verlässliche Interpretationen zur Genstrukturvorhersage beitragen können.

Im Gegensatz zu den hoch-konservierten Exonsequenzen (<90%) zeigen die Intronsequenzen in *A. thaliana* und *C. rubella* eine niedrigere Sequenzhomologie (<80%). In intergenischen Bereichen liegt ebenfalls eine geringere Sequenzidentität zwischen beiden Arten vor. Die intergenischen Bereiche in den analysierten

Genomregionen von *C. rubella* zeigen keine deutlichen Größen-Unterschiede zu denen in *A. thaliana*. In einigen Fällen ist die Gesamtgenlänge bestehend aus Exon- und Intron-Sequenzen durch Vergrößerung bzw. Verkleinerung der Intronsequenzen zwischen *A. thaliana* und *C. rubella* verschieden. Die Exonlängen sind in den meisten Fällen in beiden Pflanzen sehr ähnlich. Auch in Mais, Reis und Hirse wurde für die Gene *sh2* und *a1* eine höhere Sequenzhomologie in den Exons als in den Introns beobachtet (Chen *et al.*, 1997).

Die vergleichbar großen intergenischen Bereiche in *Arabidopsis* und *Capsella* deuten an, daß der Anteil repetitiver Elemente in *Capsella* ähnlich niedrig ist wie in *Arabidopsis thaliana* (Lin *et al.*, 1999; Mayer *et al.*, 1999). In einigen Regionen konnten in *Capsella* Mikrosatelliten-Sequenzen, die aus Dinukleotid-Monomeren zusammengesetzt sind, identifiziert werden. Der Genomanteil direkter Sequenz- bzw. indirekter Sequenzwiederholungen in *C. rubella* ist in den untersuchten Regionen sehr niedrig. Für zwei Bereiche in der zu der 200 Kbp-Region homologen Region konnten längere, indirekte Sequenzwiederholungen in den sequenzierten Cosmid-Contigs festgestellt werden. Der erste Bereich liegt in Contig I zwischen den Genen C2 und D bei Position 33450 bis 34000, und der zweite Bereich ist in Contig II zwischen Position 8100 bis 10550 lokalisiert. Im zweiten Bereich ist zwischen den beiden indirekten Sequenzwiederholungen, die eine Länge von ca. 1 Kbp haben, ein 400 Bp langer Sequenzabschnitt eingefügt.

Die Syntenieuntersuchungen in *C. rubella* und *A. thaliana* zeigen einen hohen Grad an Kolinearität sowohl auf molekularer als auch auf genetischer Ebene. Die für diese beiden diploiden Pflanzen ermittelten Daten werden unverzichtbar sein, um Ergebnisse zur Genomstruktur, die in diploiden *Brassica*-Arten, die hexaploiden Ursprungs sind, gewonnen werden, interpretieren zu können. Weiterhin zeigen die Daten, daß die Informationen aus *A. thaliana* unmittelbar auf andere diploide Arten der Brassicaceae übertragen werden können und beispielsweise zur Karten-gestützten Klonierung von Genen eingesetzt werden können. Diese Strategie wurde bereits für die Klonierung des Gai-Gens aus Reis, Mais und Weizen ausgenutzt (Peng *et al.*, 1999).

5. Zusammenfassung

Im Rahmen dieser Arbeit wurden vergleichende genetische und physikalische Kartierungsstudien in zwei diploiden Spezies der Brassicaceae-Familie, im Modellorganismus *Arabidopsis thaliana* und *Capsella rubella*, durchgeführt.

Vergleichende Kartierungsexperimente mit Chromosom IV-Markern von *A. thaliana* etablierten zwei Kopplungsgruppen in *Capsella*. Ein hoher Konservierungsgrad bezüglich des Markerrepertoires und der Kolinearität konnte zwischen *A. thaliana* und *Capsella* ermittelt werden. Für Mikrosyntenie-Analysen wurden eine 200 Kbp-Region auf Chromosom IV bzw. eine 30 Kbp-Region auf Chromosom I von *Arabidopsis thaliana* ausgewählt. Beide Arten zeigen hohe Konservierung des Genrepertoires. Dies gilt ebenso für die Genreihenfolge und die Orientierung der Gene. Einige Gene liegen in unterschiedlichen Kopienzahlen in *Arabidopsis thaliana* und *Capsella rubella* in den untersuchten Genombereichen vor. Auch die intergenischen Bereiche sind in beiden Spezies in der Länge vergleichbar. Durch Sequenzanalysen wurde für die orthologen Gene in *C. rubella* ein hoher Konservierungsgrad für die Exon-Intron-Strukturen der Gene gezeigt. Die Sequenzidentitäten der orthologen Gene betragen in den Exonsequenzen in der Regel über 90% auf Nukleotid- und Aminosäureebene. In dieser Arbeit wurde sowohl auf genetischer wie auf molekularer Ebene zwischen den beiden eng verwandten Spezies, *A. thaliana* und *C. rubella*, eine nahezu vollständige Kolinearität der Genome ermittelt.

Diese Ergebnisse zeigen, daß die unmittelbare Übertragbarkeit von Daten aus dem Genomprojekt von *A. thaliana* auf verwandte diploide Arten möglich und damit nutzbar ist.

6. Literaturverzeichnis

Ahn, S. and Tanksley, S.D. (1993). Comparative linkage maps of the rice and maize genomes. *Proc. Natl. Acad. Sci. USA* **90**, 7980-84.

Ahn, S., Anderson, J.A., Sorrells, M.E. and Tanksley, S.D. (1993). Homoeologous relationships of rice, wheat and maize chromosomes. *Mol. Gen. Genet.* **241**, 483-490.

Akkaya, M.S., Bhagwat, A.A. and Cregan, P.B. (1992). Length polymorphisms of simple sequence repeat DNA in soybean. *Genetics* **132**, 1131-39.

Alonso-Blanco, C., Peeters, A.J.M., Koornneef, M., Lister, C., Dean, C., van den Bosch, N., Pot, J. and Kuiper, M.T.R. (1998). Development of an AFLP based linkage map of *Ler*, *Col* and *Cvi* *Arabidopsis thaliana* ecotypes and construction of a *Ler/Cvi* recombinant inbred line population. *The Plant Journal* **14**, 259-271.

Arondel, V., Lemieux, B., Hwang, I., Gibson, S., Goodman, H.M. and Somerville, C.R. (1992). Map-based cloning of a gene controlling omega-3 fatty acid desaturation in *Arabidopsis*. *Science* **258**, 1353-1355.

Arumuganathan, K. and Earle, E.D. (1991). Nuclear DNA content of some important plant species. *Plant Mol. Biol. Rep.* **9**, 208-218.

Avramova, Z., Tikhonov, A., SanMiguel, P., Jin, Y.-K., Liu, C., Woo, S.-S., Wing, R.A. and Bennetzen, J.L. (1996). Gene identification in a complex chromosomal continuum by local genomic cross-referencing. *The Plant Journal* **10**, 1163-1168.

Bancroft, I., Love, K., Bent, E., Sherson, S., Lister, C., Cobbett, C., Goodman, H. and Dean, C. (1997) A strategy involving the use of high redundancy YAC subclone libraries facilitates the contiguous representation in cosmid and BAC clones of 1.7 Mb of the genome of the plant *Arabidopsis thaliana*. *Weeds World* **4**, 1-9.

Barakat, A., Matassi, G. and Bernardi, G. (1998). Distribution of genes in the genome of *Arabidopsis thaliana* and its implications for the genome organization of plants. *Proc. Natl. Acad. Sci. USA* **95**, 10044-10049.

Barciszewski, J. and Legocki, A.B. (1997). Two plant signalling peptides: systemin and ENOD 40. *Acta Biochimica Polonica* **44**, 795-802.

Bauwens, S., Van Oostveldt, P., Engler, G. and Van Montagu, M. (1991). Distribution of the rDNA and three classes of highly repetitive DNA in the chromatin of interphase nuclei of *Arabidopsis thaliana*. *Chromosoma* **101**, 41-48.

Bedbrook, J.R., Jones, J., O'Dell, M., Thompson, R.D. and Flavell, R.B. (1980). A molecular description of telomeric heterochromatin in *Secale* species. *Cell* **19**, 545-560.

Bell, C.J. and Ecker, J.R. (1994). Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*. *Genomics* **19**, 137-144.

Bennett, M.D. and Smith, J.B. (1976). Nuclear DNA amounts in angiosperms. *Phil. Trans. R. Soc. Lond.* **274**, 227-274.

Bennetzen, J.L., SanMiguel, P., Chen, M., Tikhonov, A., Francki, M. and Avramova, Z. (1998). Grass genomes. *Proc. Natl. Acad. Sci. USA* **95**, 1975-1978.

Berhan, M.A., Hulbert, S.H., Butler, L.G. and Bennetzen, J.L. (1993). Structure and evolution of the genomes of *Sorghum bicolor* and *Zea mays*. *Theor. Appl. Genet.* **86**, 598-604.

Bevan, M., Ecker, J., Theologis, S., Federspiel, N., Davis, R., McCombie, D., Martienssen, R., Chen, E., Waterston, B., Wilson, R., Rounsley, S., Venter, C., Tabata, S., Salanoubat, M., Quetier, F., Cherry, M. and Meinke, D. (1997). Objective: the complete sequence of a plant genome. *The Plant Cell* **9**, 476-478.

Bevan, M., Bancroft, I., Bent, E., Love, K., Goodman, H., Dean, C., Bergkamp, R., Dirkse, W., Van Staveren, M., Stiekema, W., Drost, L., Ridley, P., Hudson, S.-A., Patel,

K., Murphy, G., Piffanelli, P., Wedler, H., Wedler, E., Wambutt, R., Weitzenegger, T., Pohl, T.M., Terryn, N., Gielen, J., Villarroel, R., De Clerck, R., Van Montagu, M., Lechary, A., Auborg, S., Gy, I., Kreis, M., Lao, N., Kavanagh, T., Hempel, S., Kotter, P., Entian, K.-D., Rieger, M., Schaeffer, M., Funk, B., Mueller-Auer, S., Silvey, M., James, R., Montford, A., Pons, A., Puigdomenech, P., Douka, A., Voukelatou, E., Million, D., Hatzopolous, P., Piravandi, E., Obermaier, B., Hilbert, H., Düsterhöft, A., Moores, T., Jones, J.D.G., Eneva, T., Palme, K., Benes, V., Rechman, S., Ansorge, W., Cooke, R., Berger, C., Delseny, M., Voet, M., Volckaert, G., Mewes, H.-W., Klostermann, S., Schueller, C. and Chalwatzis, N. (1998). Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of *Arabidopsis thaliana*. *Nature* **391**, 485-488.

Bonierbale, M.W., Plaisted, R.L., Tanksley, S.D. (1988). RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. *Genetics* **120**, 1095-1103.

Botella, M.A., Coleman, M.J., Hughes, D.E., Nishimura, M.T., Jones, J.D.G. and Somerville, S.C. (1997). Map positions of 47 *Arabidopsis* sequences with sequence similarity to disease resistance genes. *The Plant Journal* **12**, 1197-1211.

Bohuon, E.J.R., Keith, D.J., Parkin, I.A.P., Sharpe A.G. and Lydiate, D.J. (1996). Alignment of the conserved C genome of *Brassica oleracea* and *Brassica napus*. *Theor. Appl. Genet.* **93**, 833-839.

Bureau, T.E. and Wessler, S.R. (1992). *Tourist*: a large family of small inverted repeat elements frequently associated with maize genes. *The Plant Cell* **4**, 1283-1294.

Bureau, T.E. and Wessler, S.R. (1994). *Stowaway*: A new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants. *The Plant Cell* **6**, 907-916.

Burke, D.T., Carle, G.F. and Olson, M.V. (1987). Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* **236**, 806-812.

Burr, B., Burr, F.A., Thompson, K.H., Albertson, M.C. and Stuber, C.W. (1988). Gene mapping with recombinant inbreds in maize. *Genetics* **118**, 519-526.

Burr, B., Burr, F.A., Matz, E.C. and Romero-Severson, J. (1992). Pinning down loose ends: Mapping telomeres and factors affecting their length. *The Plant Cell* **4**, 953-960.

Büschges, R., Hollrichter, K., Panstruga, R., Simons, G., Wolter, M., Frijters, A., van Daelen, R., van der Lee, T., Diergaarde, P., Groenendijk, J., Töpsch, S., Vos, P., Salamini, F. and Schulze-Lefert, P. (1997). The barley *Mlo* gene: a novel control element of plant pathogen resistance. *Cell* **88**, 695-705.

Carabelli, M., Sessa, G., Baima, S., Morelli, G. and Ruberti, I. (1993). The *Arabidopsis Athb-2* and *-4* genes are strongly induced by far-red-rich light. *The Plant Journal* **4**, 469-479.

Casacuberta, E., Casacuberta, J.M., Puigdomenech, P. and Monfort, A. (1998). Presence of miniature inverted-repeat transposable elements (MITEs) in the genome of *Arabidopsis thaliana*: characterisation of the *Emigrant* family of elements. *The Plant Journal* **16**, 79-85.

Cavell, A.C., Lydiate, D.J., Parkin, I.A.P., Dean, C. and Trick, M. (1998). Collinearity between a 30-centimorgan segment of *Arabidopsis thaliana* chromosome 4 and duplicated regions within the *Brassica napus* genome. *Genome* **41**, 62-69.

Chang, C. and Meyerowitz, E.M. (1991). Plant genome studies: restriction fragment length polymorphism and chromosome mapping information. *Curr. Opin. Genet. Dev.* **1**, 112-118.

Chang, C., Bowman, J.L., DeJohn, A.W., Lander, E.S. and Meyerowitz, E.M. (1988). Restriction fragment length polymorphism linkage map for *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **85**, 6856-6860.

Chen, M., SanMiguel, P., de Oliveira, A.C., Woo, S.-S., Zhang, H., Wing, R.A. and Bennetzen, J.L. (1997). Microcolinearity in *sh2*-homologous regions of the maize, rice, and sorghum genomes. *Proc. Natl. Acad. Sci. USA* **94**, 3431-3435.

Chen, M., SanMiguel, P. and Bennetzen, J.L. (1998). Sequence organization and conservation in *sh2/al1*-homologous regions of sorghum and rice. *Genetics* **148**, 435-443.

Cho, R.J., Mindrinos, M., Richards, D.R., Sapolsky, R.J., Anderson, M., Drenkard, E., Dewdney, J., Reuber, T.L., Stammers, M., Federspiel, N., Theologis, A., Yang, W.-H., Hubbell, E., Au, M., Chung, E.Y., Lashkari, D., Lemieux, B., Dean, C., Lipschutz, R.J., Ausubel, F.M., Davis, R.W. and Oefner, P.J. (1999). Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. *Nat. Genet.* **23**, 203-207.

Clarenz, O. (2000). Studien zur Genomkolinearität in der Familie der Brassicaceae. Diplomarbeit, Universität zu Köln.

Conner, J.A., Conner, P., Nasrallah M.E. and Nasrallah, J.B. (1998). Comparative mapping of the *Brassica* S locus region and its homeolog in *Arabidopsis*: Implications for the evolution of mating systems in the Brassicaceae. *The Plant Cell* **10**, 801-812.

Cooke, R., Raynal, M., Laudié, M., Grellet, F., Delseny, M., Morris, P.C., Guerrier, D., Giraudat, J., Quigley, F., Clabault, G., Li, Y.-F., Mache, R., Krivitzky, M., Gy, I.J.-J., Kreis, M., Lecharny, A., Parmentier, Y., Marbach, J., Fleck, J., Clément, B., Philipps, G., Hervé, C., Bardet, C., Tremousaygue, D., Lescure, B., Lacomme, C., Roby, D., Jourjon, M.F., Chabrier, P., Charpentreau, J.-L., Desprez, T., Amselem, J., Chiapello H. and Höfte, H. (1996). Further progress towards a catalogue of all *Arabidopsis* genes: analysis of a set of 5000 non-redundant ESTs. *The Plant Journal* **9**, 101-124.

Copenhaver, G.P. and Pikaard, C.S. (1996). Two-dimensional RFLP analyses reveal megabase-sized clusters of rRNA gene variants in *Arabidopsis thaliana*, suggesting local spreading of variants as the mode for gene homogenization during concerted evolution. *The Plant Journal* **9**, 273-282.

Copenhaver, G.P., Browne, W.E. and Preuss, D. (1998). Assaying genome-wide recombination and centromere functions with *Arabidopsis* tetrads. *Proc. Natl. Acad. Sci. USA* **95**(1), 247-252.

Coulson, A., Sulston, J., Brenner, S. and Karn, J. (1986) Toward a physical map of the genome of the nematode *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. USA* **83**, 7821-7825.

Crespi, M.D., Jurkevitch, E., Poiret, M., d'Aubenton-Carafa, Y., Petrovics, G., Kondorosi, E. and Kondorosi, A. (1994). enod40, a gene expressed during nodule organogenesis, codes for a non-translatable RNA involved in plant growth. *EMBO Journal* **13**, 5099-5012.

Davis, G.L., McMullen, M.D., Baysdorfer, C., Musket, T., Grant, D., Staebell, M., Xu, G., Polacco, M., Koster, L., Melia-Hancock, S., Houchins, K., Chao, S. and Coe, E.H. Jr. (1999). A maize map standard with sequenced core markers, grass genome reference points and 932 expressed sequence tagged sites (ESTs) in a 1736-locus map. *Genetics* **152**, 1137-1172.

Dellaporta, S.L., Wood, J. and Hicks, J.B. (1983) A plant DNA miniprep: version II. *Plant Mol. Biol. Rep.* **1**, 19-21.

Delseny, M., Cooke, R., Raynal, M. and Grellet, F. (1997). The *Arabidopsis thaliana* cDNA sequencing projects. *FEBS Letters* **405**, 129-132.

Devos, K.M. and Gale, M.D. (1993). Extended genetic maps of the homoeologous group 3 chromosomes of wheat, rye and barley. *Theor. Appl. Genet.* **85**, 649-652.

Devos, K.M. and Gale, M.D. (1993). The genetic maps of wheat and their potential in plant breeding. *Outlook on Agriculture* **22**, 93-99.

Devos, K.M., Atkinson, M.D., Chinoy, C.N., Francis, H.A., Harcourt, R.L., Koebner, R.M.D., Liu, C.J., Masojc, P., Xie, D.X. and Gale, M.D. (1993a). Chromosomal rearrangements in the rye genome relative to that of wheat. *Theor. Appl. Genet.* **85**, 673-

680.

Devos, K.M., Millan, T. and Gale MD. (1993b). Comparative RFLP maps of homoeologous group-2 chromosomes of wheat, rye and barley. *Theor. Appl. Genet.* **85**, 784-792.

Devos, K.M. and Gale, M.D. (1997). Comparative genetics in the grasses. *Plant Molecular Biology* **35**, 3-15.

Erdmann, V.A., Szymanski, M., Hochberg, A., de Groot, N. and Barciszewski, J. (1999). Collection of mRNA-like non-coding RNAs. *Nucleic Acids Research* **27**, 192-195.

Fabri, C.O. and Schöffner, A.R. (1994). An *Arabidopsis thaliana* RFLP mapping set to localize mutations to chromosomal regions. *The Plant Journal* **5**, 149-156.

Fedoroff, N. V. (1989). About maize transposable elements and development. *Cell* **56**, 181-191.

Feuillet, C. and Keller, B. (1999). High gene density is conserved at syntenic loci of small and large grass genomes. *Proc. Natl. Acad. Sci. USA* **96**, 8265-8270.

Flavell, R.B., Bennett, M.D., Smith, J.B. and Smith, D.B. (1974). Genome size and the proportion of repeated nucleotide sequence DNA in plants. *Biochem. Genet.* **12**, 257-269.

Flavell, R.B. and Smith, D.B. (1976). Nucleotide sequence organisation in the wheat genome. *Heredity* **37**, 231-252.

Flavell, R. (1980). The molecular characterization and organization of plant chromosomal DNA sequences. *Annu. Rev. Plant Phys.* **31**, 569-596.

Gale, M.D. and Devos, K.M. (1998). Comparative genetics in the grasses. *Proc. Natl. Acad. Sci. USA* **95**, 1971-1974.

Ganal, M.W., Lapitan, N.L.V. and Tanksley, S.D. (1988). A molecular and cytogenetic survey of major repeated DNA sequences in tomato (*Lycopersicon esculentum*). *Mol. Gen. Genet.* **213**, 262-268.

Gebhardt, C., Ritter, E., Barone, A., Debener, T., Walkemeier, B., Schachtschnabel, U., Kaufmann, H., Thompson, R.D., Bonierbale, M.W., Ganal, M.W., Tanksley, S.D. and Salamini, F. (1991). RFLP maps of potato and their alignment with the homoeologous tomato genome. *Theor. Appl. Genet.* **83**, 49-57.

Gerlach, W.L. and Peacock, W.J. (1980). Chromosomal locations of highly repeated DNA sequences in wheat. *Heredity* **44**, 269-276.

Giraudat, J., Hauge, B.M., Valon, C., Smalle, J., Parcy, F. and Goodman, H.M. (1992). Isolation of the Arabidopsis *ABI3* gene by positional cloning. *The Plant Cell* **4**, 1251-1261.

Grant, M.R., McDowell, J.M., Sharpe, A.G., de Torres Zabala, M., Lydiate, D.J. and Dangl, J.L. (1998). Independent deletions of a pathogen-resistance gene in *Brassica* and *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **95**, 15843-15848.

Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* **166**, 557-580.

Harushima, Y., Yano, M., Shomura, A., Sato, M., Shimano, T., Kuboki, Y., Yamamoto, T., Lin, S.Y., Antonio, B.A., Parco, A., Kajiya, H., Huang, N., Yamamoto, K., Nagamura, Y., Kurata, N., Khush, G.S. and Sasaki, T. (1998). A high-density rice genetic linkage map with 2275 markers using a single F2 population. *Genetics* **148**, 479-494.

Hauge, B.M., Hanley, S.M., Cartinhour, S., Cherry, J.M., Goodman, H.M., Koornneef, M., Stam, P., Chang, C., Kempin, S., Medrano, L. and Meyerowitz, E.M. (1993). An integrated genetic/RFLP map of the *Arabidopsis thaliana* genome. *The Plant Journal* **3**, 745-754.

Helentjaris, T., Weber, D. and Wright, S. (1988). Identification of the genomic locations of duplicate nucleotide sequences in maize by analysis of restriction fragment length polymorphisms. *Genetics* **118**, 353-363.

Höfte, H., Desprez, T., Amselem, J., Chiapello, H., Caboche, M., Moison, A., Jourjon, M.-F., Charpentreau, J.-L., Berthomieu, P., Guerrier, D., Giraudat, J., Quigley, F., Thomas, F., Yu, D.-Y., Mache, R., Raynal, M., Cooke, R., Grellet, F., Delseny, M., Parmentier, Y., de Marcillac, G., Gigot, C., Fleck, J., Philipps, G., Axelos, M., Bardet, C., Tremousaygue, D. and Lescure, B. (1993). An inventory of 1152 expressed sequence tags obtained by partial sequencing of cDNAs from *Arabidopsis thaliana*. *The Plant Journal* **4**, 1051-1061.

Hong, G., Qian, Y., Yu, S., Hu, X., Zhu, J., Tao, W., Li, W., Su, C., Zhao, H., Qiu, L., Yu, D., Liu, X., Wu, B., Zhang, X. and Zhao, W. (1997). A 120 kilobase resolution contig map of the rice genome. *DNA Seq.* **7**, 319-335.

Hurka, H. and Neuffer, B. (1997). Evolutionary processes in the genus *Capsella* (*Brassicaceae*). *Pl. Syst. Evol.* **206**, 295-316.

Jespersen, H.M., Kjaersgard, I.V.H., Ostergaard, L. and Welinder, K.G. (1997). From sequence analysis of the three novel ascorbate peroxidases from *Arabidopsis thaliana* to structure, function and evolution of seven types of ascorbate peroxidase. *Biochem. J.* **326**, 305-310.

Koch, M., Haubold, B. and Mitchell-Olds, T. Comparative evolutionary analysis of the chalconesynthase and alcohol dehydrogenase from *Arabidopsis* and *Arabis* relatives. (eingereicht zur Veröffentlichung)

Konieczny, A. and Ausubel, F.M. (1993). A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific PCR-based markers. *The Plant Journal* **4**, 403-410.

Koornneef, M., van Eden, J., Hanhart, C.J., Stam, P., Braaksma, F.J. and Feenstra, W.J. (1983). Linkage map of *Arabidopsis thaliana*. *J. Hered.* **74**, 265-272.

- Kopczak, S.D., Haas, N.A., Hussey, P.J., Silflow, C.D. and Snustad, D.P. (1992). The small genome of *Arabidopsis* contains at least six expressed alpha-tubulin genes. *The Plant Cell* **4**, 539-547.
- Kowalski, S.P., Lan, T.-H., Feldmann, K.A. and Paterson, A.H. (1994). Comparative mapping of *Arabidopsis thaliana* and *Brassica oleracea* chromosomes reveals islands of conserved organization. *Genetics* **138**, 499-510.
- Lagercrantz, U. and Lydiate, D. (1996). Comparative genome mapping in *Brassica*. *Genetics* **144**, 1903-1910.
- Lagercrantz, U., Putterill, J., Coupland, G. and Lydiate, D. (1996). Comparative mapping in *Arabidopsis* and *Brassica*, fine scale genome collinearity and congruence of genes controlling flowering time. *The Plant Journal* **9**, 13-20.
- Lagercrantz, U. (1998). Comparative mapping between *Arabidopsis thaliana* and *Brassica nigra* indicates that *Brassica* genomes have evolved through extensive genome replication accompanied by chromosome fusions and frequent rearrangements. *Genetics* **150**, 1217-1228.
- Lander, E.S., Green, P., Abrahamson, J., Barlow, A., Daly, M.J., Lincoln, S.E. and Newburg, L. (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* **1**, 174-181.
- Leister, D., Ballvora, A., Salamini, F. and Gebhardt, C. (1996). A PCR-based approach for isolating pathogen resistance genes from tomato with potential for wide application in plants. *Nat. Genet.* **14**, 421-429.
- Leister, D., Kurth, J., Laurie, D.A., Yano, M., Sasaki, T., Devos, K., Graner, A., Schulze-Lefert, P. (1998). Rapid reorganization of resistance gene homologues in cereal genomes. *Proc. Natl. Acad. Sci. USA* 1998 **95**, 370-375.

Leutwiler, L.S., Hough-Evans, B.R. and Meyerowitz, E.M. (1984). The DNA of *Arabidopsis thaliana*. *Mol. Gen. Genet.* **194**, 15-23.

Lin, X., Kaul, S., Rounsley, S., Shea, T.P., Benito, M.-I., Town, C.D., Fujii, C.Y., Mason, Y., Bowman, C.L., Barnstead, M., Feldblyum, T.V., Buell, C.R., Ketchum, K.A., Lee, J., Ronning, C.M., Koo, H.L., Moffat, K.S., Cronin, L.A., Shen, M., Pai, G., Van Aken, S., Umayam, L., Tallon, L.J., Gill, J.E., Adams, M.D., Carrera, A.J., Creasy, T.H., Goodman, H.M., Somerville, C.R., Copenhaver, G.P., Preuss, D., Nierman, W.C., White, O., Eisen, J.A., Salzberg, S.L., Fraser, C.M. and Venter, C. (1999). Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. *Nature* **402**, 761-768.

Lister, C. and Dean, C. (1993). Recombinant inbred lines for mapping RFLP and phenotypic markers in *Arabidopsis thaliana*. *The Plant Journal* **4**, 745-750.

Liu, Y.-G., Mitsukawa, N., Lister, C., Dean, C. and Whittier, R.F. (1996) Isolation and mapping of a new set of 129 RFLP markers in *Arabidopsis thaliana* recombinant inbred lines. *The Plant Journal* **10**, 733-736.

Livingstone, K.D., Lackney, V.K., Blauth, J.R., van Wijk, R. and Jahn, M.K. (1999). Genome mapping in Capsicum and the evolution of genome structure in the Solanaceae. *Genetics* **152**, 1183-1202.

Lydiate, D., Sharpe, A., Lagercrantz, U. and Parkin, I. (1993). Mapping the *Brassica* genome. *Outlook Agric.* **22**, 85-92.

Maluszynska, J. and Heslop-Harrison, J.S. (1991). Localization of tandemly repeated DNA sequences in *Arabidopsis thaliana*. *The Plant Journal* **1**, 159-166.

Marra, M., Kucaba, T., Sekhon, M., Hillier, L., Martienssen, R., Chinwalla, A., Crockett, J., Fedele, J., Grover, H., Gund, C., McCombie, W.R., McDonald, K., McPherson, J., Mudd, N., Parnell, L., Schein, J., Seim, R., Shelby, P., Waterston, R. and Wilson, R. (1999). A map for sequence analysis of the *Arabidopsis thaliana* genome. *Nat. Genet.* **22**, 265-270.

Martin, G.B., Brommonschenkel, S.H., Chunwongse, J., Frary, A., Ganai, M.W., Spivey, R., Wu, T., Earle, E.D. and Tanksley, S.D. (1993). Map-based cloning of a protein kinase gene conferring disease resistance in tomato. *Science* **262**, 1432-1436.

Martinez-Zapater, J.M., Estelle, M.A. and Somerville, C.R. (1986). A highly repeated DNA sequence in *Arabidopsis thaliana*. *Mol. Gen. Genet.* **204**, 417-423.

Mayer, K., Schüller, C., Wambutt, R., Murphy, G., Volckaert, G., Pohl, T., Düsterhöft, A., Stiekema, W., Entian, K.-D., Terry, N., Harris, B., Ansong, W., Brandt, P., Grivelli, L., Rieger, M., Weichselgartner, M., de Simone, V., Obermaier, B., Mache, R., Müller, M., Kreis, M., Delseny, M., Puigdomenech, P., Watson, M., Schmidheini, T., Reichert, B., Portatelle, D., Perez-Alonso, M., Boutry, M., Bancroft, I., Vos, P., Hoheisel, J., Zimmermann, W., Wedler, H., Ridley, P., Langham, S.-A., McCullagh, B., Bilham, L., Robben, J., Van der Schueren, J., Grymonprez, B., Chuang, Y.-C., Vandebussche, F., Braeken, M., Weltjens, I., Voet, M., Bastiaens, I., Aert, R., Defoor, E., Weitzenegger, T., Bothe, G., Ramsperger, U., Hilbert, H., Braun, M., Holzer, E., Brandt, A., Peters, S., van Staveren, M., Dirkse, W., Mooijman, P., Klein Lankhorst, R., Rose, M., Hauf, J., Kötter, P., Berneiser, S., Hempel, S., Feldpausch, M., Lamberth, S., Van den Daele, H., De Keyser, A., Buysschaert, C., Gielen, J., Villaruel, R., De Clercq, R., Van Montagu, M., Rogers, J., Cronin, A., Quail, M., Bray-Allen, S., Clark, L., Doggett, J., Hall, S., Kay, M., Lennard, N., McLay, K., Mayes, R., Pettet, A., Rajandream, M.-A., Lyne, M., Benes, V., Rechmann, S., Borkova, D., Blöcker, H., Scharfe, M., Grimm, M., Löhnert, T.-H., Dose, S., de Haan, M., Maarse, A., Schäfer, M., Müller-Auer, S., Gabel, C., Fuchs, M., Fartmann, B., Granderath, K., Dauner, D., Herzl, A., Neumann, S., Argiriou, A., Vitale, D., Liguori, R., Piravandi, E., Massenet, O., Quigley, F., Clabaud, G., Mündlein, A., Felber, R., Schnabl, S., Hiller, R., Schmidt, W., Lecharny, A., Aubourg, S., Chefdor, F., Cooke, R., Berger, C., Montfort, A., Casacuberta, E., Gibbons, T., Weber, N., Vandenbol, M., Bargues, M., Terol, J., Torres, A., Perez-Perez, A., Purnelle, B., Bent, E., Johnson, S., Tacon, D., Jesse, T., Heijnen, L., Schwarz, S., Scholler, P., Heber, S., Francis, P., Bielke, C., Frishman, D., Haase, D., Lemcke, K., Mewes, H.W., Stocker, S., Zaccaria, P., Bevan, M., Wilson, R.K., de la Bastide, M., Habermann, K., Parnell, L., Dedhia, N., Gnoj, L., Schutz, K., Huang, E., Spiegel, L., Sehkon, M., Murray, J., Sheet, P., Cordes, M., Abu-Threideh, J., Stoneking, T., Kalicki, J., Graves, T., Harmon, G., Edwards, J., Latreille, P., Courtney, L., Cloud,

J., Abbott, A., Scott, K., Johnson, D., Minx, P., Bentley, D., Fulton, B., Miller, N., Greco, T., Kemp, K., Kramer, J., Fulton, L., Mardis, E., Dante, M., Pepin, K., Hillier, L., Nelson, J., Spieth, J., Ryan, E., Andrews, S., Geisel, C., Layman, D., Du, H., Ali, J., Berghoff, A., Jones, K., Drone, K., Cotton, M., Joshu, C., Antonoiu, B., Zidanic, M., Strong, C., Sun, H., Lamar, B., Yordan, C., Ma, P., Zhong, J., Preston, R., Vil, D., Shekher, M., Matero, A., Shah, R., Swaby, I.K., O'Shaughnessy, A., Rodriguez, M., Hoffman, J., Till, S., Granat, S., Shohdy, N., Hasegawa, A., Hameed, A., Lodhi, M., Johnson, A., Chen, E., Marra, M., Martienssen, R. and McCombie, W.R. (1999). Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*. *Nature* **402**, 769-777.

McClintock, B. (1948). Mutable loci in maize. *Carnegie Inst. Wash. Yearbook* **47**, 155-169.

McDowell, J.M., Dhandaydham, M., Long, T.A., Aarts, M.G., Goff, S., Holub, E.B. and Dangl, J.L. (1998). Intragenic recombination and diversifying selection contribute to the evolution of downy mildew resistance at the *RPP8* locus of *Arabidopsis*. *The Plant Cell* **10**, 1861-1874.

Meinke, D.W., Cherry, J.M., Dean, C., Rounsley, S.D. and Koornneef, M. (1998). *Arabidopsis thaliana*: a model plant for genome analysis. *Science* **282**, 662, 679-682.

Moore, G. (1995). Cereal genome evolution: pastoral pursuits with „Lego“ genomes. *Curr. Opin. Genet. Dev.* **5**, 717-724.

Mozo, T., Dewar, K., Dunn, P., Ecker, J.R., Fischer, S., Kloska, S., Lehrach, H., Marra, M., Martienssen, R., Meier-Ewert, S. and Altmann, T. (1999). A complete BAC-based physical map of the *Arabidopsis thaliana* genome. *Nat. Genet.* **22**, 271-275.

Murray, M.G., Cuellar, R.E. and Thompson, W.F. (1978). DNA sequence organization in the pea genome. *Biochemistry* **17**, 5781-5790.

Murray, M.G., Palmer, J.D., Cuellar, R.E. and Thompson, W.F. (1979). Deoxyribonucleic acid sequence organization in the mung bean genome. *Biochemistry*

18, 5259-5266.

Nam, H.-G., Giraudat, J., den Boer, B., Moonan, F., Loos, W.D.B., Hauge, B.M. and Goodman, H.M. (1989). Restriction fragment length polymorphism linkage map of *Arabidopsis thaliana*. *The Plant Cell*, **1** 699-705.

Newman, T., de Bruijn, F.J., Green, P., Keegstra, K., Kende, H., McIntosh, L., Ohlrogge, J., Raikhel, N., Somerville, S., Thomashow, M., Retzel, E. and Somerville, C. (1994). Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous *Arabidopsis* cDNA clones. *Plant Phys.* **106**, 1241-1255.

Olivas, W.M., Muhrad, D. and Parker, R. (1997). Analysis of the yeast genome: identification of new non-coding and small ORF-containing RNAs. *Nucleic Acids Research* **25**, 4619-4625.

Parker, J.E., Coleman, M.J., Szabo, V., Frost, L.N., Schmidt, R., van der Biezen, E.A., Moores, T., Dean, C., Daniels, M.J. and Jones J.D.G. (1997). The *Arabidopsis* downy mildew resistance gene *RPP5* shares similarity to the Toll and Interleukin-1 receptors with *N* and *L6*. *The Plant Cell* **9**, 879-894.

Paterson, A.H., Lan, T.-H., Reischmann, K.P., Chang, C., Lin, Y.-R., Liu, S.-C., Burow, M.D., Kowalski, S.P., Katsar, C.S., DelMonte, T.A., Feldmann, K.A., Schertz, K.F. and Wendel, J.F. (1996). Toward a unified genetic map of higher plants, transcending the monocot-dicot divergence. *Nature Genet.* **14**, 380-382.

Pélissier, T., Tutois, S., Deragon, J.M., Tourmente, S., Genestier, S. and Picard, G. (1995). *Athila*, a new retroelement from *Arabidopsis thaliana*. *Plant Mol. Biol.* **29**, 441-452.

Peng, J., Richards, D.E., Hartley, N.M., Murphy, G.P., Devos, K.M., Flintham, J.E., Beales, J., Fish, L.J., Worland, A.J., Pelica, F., Sudhakar, D., Christou, P., Snape, J.W., Gale, M.D. and Harberd, N.P. (1999). 'Green revolution' genes encode mutant gibberellin response modulators. *Nature* **400**, 256-261.

Prince, J.R., Pochard, E. and Tanksley, S.D. (1992). Construction of a molecular linkage map of pepper and a comparison of synteny with tomato. *Genome* **36**, 404-417.

Pruitt, R.E. and Meyerowitz, E.M. (1986). Characterization of the genome of *Arabidopsis thaliana*. *J. Mol. Biol.* **187**, 169-183.

Ramel, C. (1997). Mini- and microsatellites. *Environ. Health Perspect.* **105** Suppl. **4**, 781-789.

Richards, E.J. and Ausubel, F.M. (1988). Isolation of a higher eukaryotic telomere from *Arabidopsis thaliana*. *Cell* **53**, 127-136.

Richmond, T. and Somerville, S. (2000). Chasing the dream: plant EST microarrays. *Curr. Opin. Plant Biol.* **3**, 108-116.

Ronald, C.R. (1998). Resistance gene evolution. *Curr. Opin. Plant Biol.* **1**, 294-298.

Round, E.K., Flowers, S.K. and Richards, E.J. (1997). *Arabidopsis thaliana* centromere regions: genetic map positions and repetitive DNA structure. *Genome Res.* **7**, 1045-1053.

Ruan, Y., Gilmore, J. and Conner, T. (1998). Towards *Arabidopsis* genome analysis: monitoring expression profiles of 1400 genes using cDNA microarrays. *The Plant Journal* **15**, 821-833.

Sadowski, J., Gaubier, P., Delseny, M. and Quiros, C.F. (1996). Genetic and physical mapping in *Brassica* diploid species of a gene cluster defined in *Arabidopsis thaliana*. *Mol. Gen. Genet.* **251**, 298-306.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (eds), 1989. Molecular cloning: a laboratory manual. Second edition. Cold Spring Harbor Laboratory Press, New York.

SanMiguel, P., Tikhonov, A., Jin, Y.-K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., Springer, P.S., Edwards, K., Lee, M., Avramova, Z., Bennetzen, J.L.

(1996). Nested retrotransposons in the intergenic regions of the maize genome. *Science* **274**, 765-768.

Sasaki, T. and Burr, B. (2000). International Rice Genome Sequencing Project: the effort to completely sequence the rice genome. *Curr. Opin. Plant Biol.* **3**, 138-141.

Sato, S., Kotani, H., Hayashi, R., Liu, Y.G., Shibata, D. and Tabata, S. (1998). A physical map of *Arabidopsis thaliana* chromosome 3 represented by two contigs of CIC YAC, P1, TAC and BAC clones. *DNA Res.* **5**, 163-168.

Schena, M., Lloyd, A.M. and Davis, R.W. (1993). The *HAT4* gene of *Arabidopsis* encodes a developmental regulator. *Genes Dev.* **7**, 367-379.

Schmidt, R., West, J., Love, K., Lenehan, Z., Lister, C., Thompson, H., Bouchez, D. and Dean, C. (1995). Physical map and organization of *Arabidopsis thaliana* chromosome 4. *Science* **270**, 480-483.

Schmidt, R., West, J., Cnops, G., Love, K., Balestrazzi, A. and Dean, C. (1996). Detailed description of four YAC contigs representing 17Mb of chromosome 4 of *Arabidopsis thaliana* ecotype Columbia. *The Plant Journal* **9**, 755-765.

Schmidt, R. (1998). Physical mapping of the *Arabidopsis thaliana* genome. *Plant Physiol. Biochem.* **36**, 1-8.

Schmidt, R., Acarkan, A., Koch, M. and Roßberg, M. (1999). A strategy for comparative physical mapping in cruciferous plants. In *Plant Evolution in Man-Made Habitats, Proceedings of the VIIth International IOPB Symposium Amsterdam* (eds. L.W.D. van Raamsdonk and J.C.M. den Nijs), Hugo de Vries Laboratory, Amsterdam, The Netherlands, pp. 183-196.

Schweizer, G., Borisjuk, N., Borisjuk, L., Stadler, M., Stelzer, T., Schilde, L. and Hemleben, V. (1993). Molecular analysis of highly repeated genome fractions in *Solanum* and their use as markers for the characterization of species and cultivars. *Theor. Appl. Genet.* **85**, 801-808.

Shizuya, H., Birren, B., Kim, U.-J., Mancino, V., Slepak, T., Tachiiri, Y. and Simon, M. (1992). Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl. Acad. Sci. USA* **89**, 8794-8797.

Simoens, C.R., Gielen, J., Van Montagu, M. and Inzé, D. (1988). Characterization of highly repetitive sequences of *Arabidopsis thaliana*. *Nucl. Acids Res.* **16**, 6753-6766.

Smith, D.B. and Flavell, R.B. (1977). Nucleotide sequence organisation in the rye genome. *Biochim. Biophys. Acta* **474**, 82-97.

Song, W.-Y., Wang, G.-L., Chen, L.-L., Kim, H.-S., Pi, L.-Y., Holsten, T., Gardner, J., Wang, B., Zhai, W.-X., Zhu, L.-H., Fauquet, C. and Ronald, P. (1995). A receptor kinase-like protein encoded by the rice disease resistance gene, Xa21. *Science* **270**, 1804-1806.

Surzycki, S.A. and Belknap, W.R. (1999). Characterization of repetitive DNA elements in *Arabidopsis*. *J. Mol. Evol.* **48**, 684-691.

Tanksley, S.D., Bernatzky, R., Lapitan, N.L. and Prince, J.P. (1988). Conservation of gene repertoire but not gene order in pepper and tomato. *Proc. Natl. Acad. Sci. USA* **85**, 6419-6423.

Tanksley, S.D., Ganai, M.W., Prince, J.P., de Vicente, M.C., Bonierbale, M.W., Broun, P., Fulton, T.M., Giovannoni, J.J., Grandillo, S., Martin, G.B., Messeguer, R., Miller, J.C., Miller, L., Paterson, A.H., Pineda, O., Röder, M.S., Wing, R.A., Wu, W. and Young, N.D. (1992). High density molecular linkage maps of the tomato and potato genomes. *Genetics* **132**, 1141-1160.

Teramoto, H., Toyama, T., Takeba, G. and Tsuji, H. (1996). Noncoding RNA for CR20, a cytokinin-repressed gene of cucumber. *Plant Molecular Biology* **32**, 797-808.

Teutonico, R.A. and Osborn, T.C. (1994). Mapping of RFLP and qualitative trait loci in *Brassica rapa* and comparison to the linkage maps of *B. napus*, *B. oleracea*, and *Arabidopsis thaliana*. *Theor. Appl. Genet.* **89**, 885-894.

Tikhonov, A.P., SanMiguel, P.J., Nakajima, Y., Gorenstein, N.M., Bennetzen, J.L. and Avramova, Z. (1999). Colinearity and its exceptions in orthologous *adh* regions of maize and sorghum. *Proc. Natl. Acad. Sci. USA* **96**, 7409-7414.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M. and Zabeau, M. (1995). AFLP: a new technique for DNA fingerprinting. *Nucl. Acids Res.* **23**, 4407-4414.

Whitkus, R., Doebley, J. and Lee, M. (1992). Comparative genome mapping of sorghum and maize. *Genetics* **132**, 1119-1130.

Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A. and Tingey, S.V. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucl. Acids Res.* **18**, 6531-6535.

Xia, X., Selvaraj, G. and Bertrand, H. (1993). Structure and evolution of a highly repetitive DNA sequence from *Brassica napus*. *Plant Mol. Biol.* **21**, 213-224.

Yamamoto, K. and Sasaki, T. (1997). Large-scale EST sequencing in rice. *Plant Mol. Biol.* **35**, 135-144.

Yang, Y.W., Lai, K.N., Tai, P.Y. and Li, W.H. (1999). Rates of nucleotide substitution in angiosperm mitochondrial DNA sequences and dates of divergence between *Brassica* and other angiosperm lineages. *J. Mol. Evol.* **48**, 597-604.

Zachgo, E.A., Wang, M.L., Dewdney, J., Bouchez, D., Camilleri, C., Belmonte, S., Huang, L., Dolan, M. and Goodman, H.M. (1996). A physical map of chromosome 2 of *Arabidopsis thaliana*. *Genome Res.* **6**, 19-25.

Anhang

I Daten der genetischen Kartierungsexperimente in *Capsella*

Kartierungspopulation: 50 F2-Pflanzen aus einer Kreuzung zwischen *Capsella grandiflora* und *Capsella rubella*

1 = Genotyp der F2-Pflanze: *Capsella grandiflora* homozygot

2 = Genotyp der F2-Pflanze: heterozygot

3 = Genotyp der F2-Pflanze: *Capsella rubella* homozygot

- = Genotyp der F2-Pflanze nicht ermittelt

m518A	213221222122-11121122121----2232232221212113222333
mi306	213222221222221223222222122232232221212123222333
m448A	21322222122222122322322221-233222221112123222332
m315Bb	2132222212222212233232222-2233222221112123223332
mi122	21322232222221223323222212233222221112123223322
mi51	213222322222212233232222122332222111212322--22
mi30	21-22222221222223222332-331222232222121222332222
ATTS0636	213222222212222232223322331222232222121222332222
m326B	21322222221--2223222322-33122223222221222332222
ATTS2243	213222222212212231223222331223121222221222332221
EST5760	2-3-22222-1-21223-223222331223121222221222232221
cDNAJ	213222222212212231223222331223121222221222232221
mi330a	213222222212212231223222331323121222221122232221
m557A	2232--222221-212231233222332323121222211122232221
EST5149	2---22122221-212131223-2-2323321222323121232233211
mi123	32222112221322213122322222332112232221232-32211
mi232	32222112221322213122322222332112232221232-32111
EST17698	322221122213223231223123222332112232221322232111
mi431	322221122213223221223123222332112232221322232111
ATTS2212a	3221211222213222221223123222332112232221322232111
ATTS3374a	3221211222213222221223123222322112232221322232111
mi369	3221211222213222221223123222322112232222322232111
mi390	2-3-21232122-123321232332231231232222212133322321
ATTS3374b	223121232222-123321232332231231232222212133322321
m211A	3123121122322--1222211122-22221222221222121222212
mi330b	12233312222132223222322233231322211112312112212211
EST3486a	321231221233132132221----22213311-232222122222321
m315Ba	3212312212331321322212111222133112232222122222321
ATTS2212b	213-11321112--22111222-12122222123231322222322232
m235A	-122321222321222223221122-2223222-3112321232132122
m254A	-1222-12222313222231311-11222322223212321232122121
EST3486b	21131111222213223231331111122212213222231312222122
EST21484	21131112222213223332232111112111213222232212222122
m249A	-1322-2113121-12122112132323222222231321322232231
m457A	31222111131112121221121323222221-2-31222322232121

EST5760	2-3-22222-1-21223-2232223312231212222212222323221
54-I/54X06	213222222212212231-232223312231213222212222323221
cDNAJ	2132222222122122312232223312231212222212222323221

II Daten der Kopplungsanalyse in *Capsella*

MAPMAKER, lod-Wert 3, Haldane cM-Funktion

mi30	0 cM	m518A	12,1 cM
ATTS0636	2,1 cM	mi306	5,5 cM
m326B	8,0 cM	m448A	2,1 cM
ATTS2243	1,0 cM	m315Bb	3,2 cM
EST5760	0 cM	mi122	1,1 cM
cDNAJ	2,1 cM	mi51	23,9 cM
mi330a	4,5 cM		
m557A	21,6 cM	mi390	1,1 cM
EST5149	9,1 cM	ATTS3374b	
mi123	1,0 cM		
mi232	9,2 cM	EST3486a	0 cM
EST17698	1,0 cM	m315Ba	
mi431	2,1 cM		
ATTS2212a	1,0 cM	m235A	13,6 cM
ATTS3374a	1,0 cM	m254A	
mi369	63,9 cM		
		EST3486b	11.8 cM
		EST21484	
		m249A	11.8 cM
		m457A	

III Sequenzen

Aufgeführt sind die im Rahmen dieser Arbeit ermittelten cDNA-Sequenzen sowie die genomischen Sequenzen von *Capsella rubella*, die den untersuchten *Arabidopsis thaliana* cDNA-Sequenzen entsprechen. Wurden für die Vergleiche Genvorhersagen oder andere Datenbanksequenzen verwendet, so sind die entsprechenden Sequenzen ebenfalls aufgeführt.

Translatierbare Bereiche sind in Fettdruck dargestellt. In Fällen, in denen kein längeres offenes Leseraster in den cDNAs bzw. den abgeleiteten Genen gefunden wurde, sind die Exonsequenzen durch Großbuchstaben und die Intronsequenzen durch Kleinbuchstaben dargestellt. Unterstrichene Bereiche in den cDNA-Sequenzen zeigen keine Homologie zur entsprechenden genomischen Sequenz in *Arabidopsis thaliana*, sie repräsentieren Klonierungsartefakte.

Sequenz des *Arabidopsis thaliana* cDNA-Klons 13004 (EST 1)

```

1  AGGCAAATCT GCATTTCTCT CTAAGTAGAA GTGGGATTCT CAGTCTAGAT
51  CGAGGAGATG CTGTAATTGA AATCACAGAA TGGGTAGACG TTCCTAAGAA
101  GAACGTGACT ATTGATAGTA ACACAACACTAC ATCAACAGGC AATGCCACTG
151  ACGAGAATTC ACAAGAAAAT AAAGAAGACC TGCAAACCGA TGCTGAAAAC
201  AGCACTGCTT CAAATACAAC AGCAGAAGAG CCAGCTGTTG CTTCTCTGGG
251  CACAGAAAAA AAGCTGAAGA AGCGGACATT CAGAATACCT CTGAAGGTAG
301  TTGAGAAAAC TGTGGACCT GGAGCACCAT TTTCGAAAGA GTCTCTTGCT
351  GAAGCTAAGA TAAAATTAGA AGCCTTGAC AAGAAAGATA GGGAAAGAAG
401  AAGAACGGCT GAGTTAAAAA ACAACCTTGA ATCTTATATA TATGCTACCA
451  AAGAGAAGCT GGAACACCCC GAGTTTGAAA AGATATCCAC CCAAGAAGAG
501  CGCAAGGCGT TTGTTGAAAA GCTTGATGAG GTGCAAGATT GGCTTTACAT
551  GGATGGAGAA GATGCTAATG CCACAGAGTT TGAGAAGCGG CTTGACTCAC
601  TTAAAGCCAT CGGCAGTCCC ATATCTTTCC GATCAGAGGA GCTTACAGCA
651  CGACTAGTAG CAATTGAATA TGCTAGGAAA TACCCTAAGT AACTCAAAGA
701  GATCATAAAA GAATGGGAGA CGAACAAAAC TTGGCTTCCA AAAGAAAAAA
751  TCGACGAGGT CTCAAAGGAA GCAGAGAAAAG TAAAAAGCTG GTTAGATAAG
801  AATGTGGCTG AGCAAGAAAA GANTTCTCTG TGGAGCAAAC CGGTGTTCAC
851  GTCCACTGAA GTGTACGCTA AAGTATTTAC TNTACAAGAC AAGGTCACAA
901  AGGTGAATAA GATCCCAAAG CCAAAGCCAA AGATAGAGAA AGTAACCAAG
951  ACGGAGAACA CAACAAAGGA GGAGGAACAA TCAAAATCCT NTGATGAAGC
1001  TGCCAAAGAA GAAGAAAGTC ACGACGAGCT TTGATTAAGG ATATAATAAA
1051  TAGAGTTTTT GTTAATTTTT GTGTTTTTAA GAGGGAAAAT AGGATTAAAA
1101  GTAGGCAGCT GAGATGAGCA AAAGGATTTT ATTCCTTTCT CCTCAACTTT
1151  TTCACATCAC CATTAGAAAC GGAATNTTCT GTGTTTTTGT TTTTCATTAT
1201  ATCAAGTTTT GGTCTCTCTA TTTCTATAAC TGAATNTTCA ATATATTTAG
1251  TAGAAGTTTA TAACCNAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
1301  A

```

Sequenz des abgeleiteten Genfragments 1 (*Capsella rubella*)

```

1  GATCCCAAAG CCAAAGCCAA AGATAGAGAA AGCAACAAAG AAAGAGAACA
51  CAACGAAGGA GGAGAAAGAG GAACAATCAA AAGCCTCAGA CTCCAATTCT
101  TCTTCCGATG AAGCTGGCAG CAAAGACGAG CTTTGA

```

Sequenz des *Arabidopsis thaliana* cDNA-Klons 5760 (EST 2)

```

1  CAAGTGATCG AATATTTGAT TCTTTTTTCAT GATCAGAGAT CTGATCTAAA
51  TGGAGTATGA CCGTACGTTG AACTTGATTC CACACGGACT ACAAGGAGTA
101  AGAGAGGAGG AAGAAATCGA AGAAGACGAT GAAGAGAGCA TGACGTTGTC

```

151	GTCAGTACCA	GAAAACGAAA	CGTCGGAGTG	TTCTTCGCCT	CCGGCGACGT
201	ATCCTCCAAT	CCCTCCTCGG	CTTAAAACAC	CGAGAGAGCC	GATGGAGTTT
251	CTATGCAGAT	CATGGAGTCT	TTCTACTTCT	GAGATCTCTT	TAGCTTTATC
301	ATCTCAGAAA	TC TGATAAAC	AACTCAACAA	AAACCCTAAT	ATTTCTCAGT
351	TGGCCGATGT	CACTTCTCTG	GCTCCGGTAG	CACCACCACC	GCCGCTACAA
401	ACGGGAAAGC	TAGCGAGTGC	GGTTCACGCG	CGGAGAACGG	GAACGATCGG
451	AAAATGGTTC	CACCACCAGG	AATTCGTCGG	CGGGAAGGTC	TCCGCCGTGA
501	AGAAGAGAGA	TAGAGTTCGC	GTGGAGAAAAG	CTCATCTGCA	TTCCGCCGTG
551	TCCATTGCGT	CTCTGGCGAC	GGCAATCGCC	GCAGTGACCG	CTTCAGGCAA
601	TCAAGACGGC	TTTCGCTGGAT	CAAAGATGAG	CTCAGCCCTT	GCCTCGGCGT
651	CGGAGTTGTT	AGCTTCTCAC	TGCCCTCGAAT	TGGCGGAGCT	TGCCGGCGCC
701	GATCACGATC	GCGTCGTCTC	TGCTGTCCGA	TCAGCTGTTG	ATGTTCTGTTG
751	ACCTGGCGAT	TTGTTGACTC	TAAC TGCTGC	AGCTGCAACA	GCATTGAGAG
801	GAGAAGCAGC	TTTGAGGGTT	AGACTACCAA	AGGAAGCTAA	GAACAGTGCA
851	GCTATAAGCC	CCTGTGAGAG	AGTTTTACCA	GAGACTCATA	GTTGCTCTTC
901	TGAGCTCGAT	CGCACTAGCA	CGACCGATGA	ACTCATATCT	GCAAAAAGGAG
951	TTGAAGAATC	AACTGGAGAG	CTAATGCAGT	GCACACGAAA	CGGTGTTNTG
1001	CGGTGGAAGC	ATGTAAGAGT	GTACATCAAC	AAAAAATCTC	AGGTCATAGT
1051	AGAAATCAAA	AGCAAACAGT	TTGTAGGAGC	ATTCTCCATG	AAAAGCAAG
1101	GCATTGTAAA	TGATGTATGC	GAGACAGTCT	CAGGCCCTGCA	AAATGGAAAA
1151	GACACGGAAA	ACACAGAAGA	AGAGCTCTAT	TTCCGAATCA	GTACAGGGAA
1201	AGGTCTGACA	AAGTTC AAGT	GCAAGAGCAA	GGCTGATAAG	CAGACATGGG
1251	TGGATAGCAT	CCGGAATCTT	CTCCATCGAG	TAAC TGCTGT	TGAGGTCGAC
1301	ACTTCTCTTG	AAACTATAAA	CATTAACGAT	AGCACATAAA	TCATGATGTC
1351	TTGTGTTAAC	GGTGTTAGTG	TTGGTTATTT	TCATTACAAG	CCAATGATAT
1401	GTATTGTTTT	AAGAACATAN	ACATATCTGC	AAGTAGGTGT	AGTGTTTTTTC
1451	ATCTTCGTCA	CTTGTTTTTTA	TTTCAGTCAA	ATTGTAACA	TTGTTTTGTG
1501	GAAATAATGA	AGAAGATTAT	AATAAAAAAA	AAAAAAAAAA	AAAAAAA

Sequenz des abgeleiteten Gens 2 (*Capsella rubella*)

1	ATGGAGCATG	ACCGTACGTT	CGACTTGTTT	CCATACGGAC	TACAAGAAGT
51	CGCCGAGGAA	GAAGAAATCG	AAGAGTTAGA	AAGAGATGAC	GAGAGCATGA
101	CGCTGTCTGTC	GGTGCCGGAA	AACGATACGT	CGGAGTGTTT	TTCATCTCCG
151	GCTATGTATC	CTCCGATCCC	TCC TCCGCC	AAA ACTCCGA	GAGAGCCTAT
201	GGAGTTTCTA	TGCAGATCAT	GGAGCATGTC	TACGTCTGAG	ATCTCTTTAG
251	CTTTGTCTTC	TCAGAAACCT	AATAAACAAA	CCAACAAAGA	ATCTAATATT
301	TCTAAGTTGG	ACGCCGTACC	TTCGCCGGCT	CCGGCACCAT	CACCACCAGT
351	TCCACCACCA	CCGCCGCTAC	TAGTAAGTCT	CTCTTGCTTC	TCCATCTTCA
401	GTTTTAGTGT	CTTTTTATGA	TTTGTATGCT	ACGATCATGC	AAATAAGGTG
451	TTTTATCGTA	CGGGCCAATA	ATTTGTTCAT	AGGCCAAAA	TTATATAATA
501	CCCATATCTT	GGAAAAGCCC	ATAATATTTA	AACTGTTTCT	TCGACAGACG
551	GGAAAGCAAC	CGAGTGTGGT	AAACGCGCGG	AGGACGGGAA	CGCTCGGTAA
601	ATGGTTCCAC	CACCGAGAAA	TTGTCCGGAG	TAAAGGCTCC	GCCGTAAGAA
651	AGAGAGATAA	AGTTCGAGTG	GAGAA TGCGC	ATCTACATTC	CGCCGTATCG
701	ATTGCGTCTC	TGGCGACGGC	CATCGCCGCC	GTGACAGCTT	CGTGCAGCCA
751	AGACGGCTTC	ACGGAATCAA	AGACGATGAG	TTCAGCGCTG	GCTTCGGCTT
801	CGGAGCTGTT	GGCTTCTCAC	TGCCCTCGAAT	TGGCGGAGCT	CTCCGGAGCC
851	GGTCACGATC	GTGTCTCTCT	TGCAGTCCGA	TCTGCCGTAG	ATGTTCTGTTG
901	ACCTGGCGAT	TTGTTGACTT	TAAC TGCTGC	AGCTGCAACT	GGTAAATPCA
951	TCCTTGACCT	TGATCAGATC	TTTCTTTTTG	CTACAATGGC	GTAAACTTAG
1001	TCCAATGCAT	ATTTGAGTTG	ATCTTTCGAT	TTGGTTTCAT	GATAAGATTT
1051	ATGTTTTATGG	CTTTTCTGAT	AATGTAGCAT	TGAGAGGAGA	AGCAGCTTTG
1101	AGGGCAAGAC	TACCAAAGGA	AGCTAAGAAC	AACGCAGCTA	TAAGCCCTTG
1151	CGAAAGAGCT	TTACCAGAAA	CTCATGATTG	TTCTTCTGAG	GTCGATTGCG
1201	CTAGCACGAC	CGATGAACAG	ATCTCTGCAC	AGGGAGATGA	AGAATCAAAA
1251	CTGGATTGTA	CTGGAGATCT	AATGCAGTGC	ACACGAAACG	GTAATATTTT
1301	CAGCAAGAGA	GCTTATCTCT	TGCTGATTGC	TGTGAATGAT	GATATCTCAT
1351	TTCAAGAAAG	TGTATATTCG	GCCATATGTA	TTTGAACAAC	AAATAACTTT
1401	CACCTACCGA	TATTGATAAA	CCGACCTATG	AACCTGTCAG	GAGCTTTTTCG
1451	GTGGAAGCAT	ATAAAAGTGT	ACATCAACCA	GAGATCTCAG	GTAAAAAAA
1501	AAAAGCAAAA	ACTCATCAAA	GCAAAAAAGT	TTTATGGTTG	TCAAATCTTA
1551	GAGTTGACAT	TCAACTCGTT	TTCTCTTGCA	GGTTGTAGTA	GAAATCAAAA

1601	GCAAACACGT	CAGAGGAGCG	TTCTCCATGA	AAAGCAAAGG	TAGTTAACTA
1651	ACTATTGTCA	CCTTCAGATA	TCTCTTCTTT	CTTTCAATGG	TCGTGGTGTT
1701	CATTTGTGTA	AAGCCAAAGT	AGTCTTGAGA	CTTTTTCCAA	TGCTTACTTG
1751	ATCCAATGGT	TATATCACTA	CACACTATAG	GAATTGTAAA	CGATGTATGC
1801	GAGATCGTCT	CGGGCCTGCA	ATATGGAAAA	GAAACAGAAA	ACGCAGAAGA
1851	AGAGCTCTAC	TTCCGGAATCA	GTACTGGGAA	AGGTCTAACG	AAGTTCAAGT
1901	GCAAGAGCAA	GGCTGATAAG	CAGACATGGA	TGGATGACAT	CCGGAATCTT
1951	CTCCATCGGG	TAACTGCTGT	TGAGGTTACC	AACACTTCTC	TTGAAACTAC
2001	AAACATTGCC	GATAGCACAT	AA		

Sequenz des *Arabidopsis thaliana* cDNA-Klons A

1	ATTAGATTTT	AGAGCTAACT	CTGAGAATTA	GGGTTCGTCT	CTCGACAAAC
51	AAAAATCTCA	GCCCCGTGGA	ATCTATTCTG	GTGTAAATCG	AAGAATGGAG
101	ACTAATGTTC	CGGAGAAGAA	GATCCTTGGT	TTCCCGGAAG	AGATGAATTC
151	GACAATTCAG	TTTTTATCTG	TTGAGAAAAAC	AACATGTCTG	CTTCTGTTTT
201	CTATAAAATT	CAATATACCA	AAGAAAATTT	CTACAAAAGAG	AAGAAGAATA
251	TCTGTAGAAG	AAGAAGAAGA	AGAAGAAGAC	GAGAATAAAA	CGAAGTTGAG
301	TTCTTGGGAG	AAGTTTTTGA	GTAGCGATGA	GATTTTGAGG	TTAAGGGAAG
351	TTTCAAGACG	AAAGTATTTG	ACGGATCGAG	AGAATAAGAA	GGTGGAGGAG
401	TTAAGGGATG	AGATCAAAGA	CGAGATTATG	TTGAATGATG	GTGTGAAGAT
451	<u>GCCTAAGAAG</u>	GAAGTAATTG	AGTTTAGATA	TAAGAGAGCA	TTGTATGAGT
501	TGGTGAAGAA	AACAGAGGAT	GATGATGATG	TTGAAGGGTA	TAGGTTTCCA
551	GATGCTTATG	ATCAAGAAGG	GTGTATTGAT	CAGAAGAAGA	GGTTTGATGT
601	TGCCAAGGAG	AGGTATTGTG	AGCGAAGAAG	GAGCGGGAGA	GTTGTTACGG
651	AGCAAGAAGC	TTGGGAGGAT	CATCAGGCTC	AGAAAGCGAG	AGTGAGTTTT
701	GGTGCAAAGG	ATAAGAAGCA	AGTTGTTGAT	GGGTATGAGT	TTGTGTTTGA
751	TGATTTGACA	GGCTTTGTG	AGGAGTCTAG	CGAGGCGGAA	ACTGGGAAAC
801	ATCGAGGTTG	TTATTCTGAAG	ACTGCTGCAG	AGAAGGCTCG	GGAGGGGAGG
851	GAGTTTCTTC	CGATTCATGG	GTACAGGGAG	GAACTGCTGA	AACTTATTGA
901	AGAGAATCAG	GTTCTTGTCA	TTGTTGGAGA	GACAGGATCG	GGTAAAACCTA
951	CGCAGATACC	GCAGTATCTT	CAAGAAGCCG	GTTACACAAA	GCGTGGAAAG
1001	ATCGGTTGTA	CTCAGCCTCG	GAGAGTTGCT	GCAATGAGCG	TTGCTTCCCG
1051	AGTGGCTCAA	GAGGTTGGTG	TTAAACTTGG	ACATGAGGTT	GGATATTCCA
1101	TTAGATTTGA	AGACTGTACT	TCAGAGAAAA	CAGTTATCAA	GTACATGACT
1151	GATGGGATGC	TGCTACGAGA	GCTACTCATA	GAACCGAAAC	TCGATAGCTA
1201	TAGTGTATC	ATTATTGATG	AGGCGCACGA	AAGAACACTG	TCCACTGATA
1251	TTTTGTTTTG	GTTAGTGAAG	GATGTAGCCA	AGGTTAGGCC	TGATCTGAGG
1301	TTAATAATCT	CCAGTGCAAC	GCTTGAAGCT	AAAAAGTTCT	CCGAATACTT
1351	TGATTCGGCT	CGGATTTACC	TGATTCCAGG	AAGAAGATAT	CCAGTTGAAA
1401	AGCTCTTCAG	AAAATGTCC	GAACCTGACT	ACTTGGAGAC	AGTGATACGC
1451	ACTGTGGTTC	AGATACATCA	GACTGAGGCG	ATAGGAGACA	TTTTAGTTTT
1501	CCTCACTGGT	CAAGAAGAGA	TCGAAACAGT	AGAAACGAAT	TTGAAGCGAA
1551	GGATGATGGA	TTTGGGTACA	AAGGGATCTG	AGATCATTAT	CTGTCTTATC
1601	TACTCAAATC	TCCCAACTCC	ACTACAGGCA	AAAGTGTTTG	AACCGGCCCC
1651	TAAAGGGACA	CGGAAAGTTG	TCTTGGCAAC	GAATATTGCA	GAAACATCGT
1701	TAACCATTTGA	TGGGGTTAAA	TACGTCATTG	ATCCAGGATA	TTGCAAGATC
1751	AACTCATACA	ACCCGAGAAC	GGGAATGGAG	TCACTCCTTG	TAACTCCTAT
1801	TTCCAAGGCT	TCCGCAGCGC	AACGAGCTGG	TCGGTCTGGA	AGAACAGGTC
1851	CTGGGAAATG	TTTCCGGTTG	TACAATATCA	AGGATTTAGA	GCCCACTACA
1901	ATACCGGAAA	TCCAAAGGGC	TAACTTGCA	AGTGTGTGTC	TTACTTTGAA
1951	GAGTCTTGGG	ATTCAAGATG	TGTTCAATTT	TGATTTTCATG	GATCCTCCAC
2001	CTGAAAACGC	TCTTTTGAAG	GCTCTAGAGT	TGCTCTATGC	GCTGGGTGCT
2051	TTGGATGAAA	TTGGTGAGAT	TACTAAGGGA	ACTCGGTCTT	CTATCGACCA
2101	AAGAACCAGC	AAGTTTTTCGC	CGACAAGGCT	CGGATGGATT	TCTATGAAGA
2151	TACAGAGAAC	GTCGGCGATC	ACATTGCATT	GTTAAGGGTT	TACAATTCCT
2201	GGAAGGAAGA	AACTACTCA	ACTCAATGGT	GCTGCGAGAA	GTTTCAATCAG
2251	AGCAAGAGTA	TGAAAAGGGC	AAGAGATATA	CGTGACCAGC	TACTTGGACT
2301	TTTAAACAAG	ATTGGAGTAG	AACTTACTTC	CAACCCGAAT	GATTTAGACG
2351	CAATAAAGAA	AGCAATCTTG	GCCGGATTCT	TCCCGCATTC	TGCAAAGTTA

2401	CAAAAGAATG	GATCGTACAG	AAGAGTCAAA	GAACCTCAGA	CAGTTTATGT
2451	ACACCCTAAC	TCAGGCCTTAT	TTGGGGCGTC	ACCTAGTAAG	TGGTTGGTGT
2501	ATCATGAACT	AGTACTTACT	ACCAAGGAAT	ATATGAGGCA	TACAACCTGAA
2551	ATGAAGCCTG	AGTGGCTAAT	TGAAATAGCT	CCTCATTACT	ACAAGCTTAA
2601	AGACATTGAA	GATACTCGGC	CAAAGAAAAC	ACAGAGAAGA	ATCGAAGAAG
2651	CCTCGACGTC	TAAGGTTGAT	ACTAACAAAG	AAACCCGGAC	CTCGAAGGTT
2701	GATACTAACA	AGAAATCAAA	GAGGTAACAA	CCAAACTAGT	CGTTTCTCTT
2751	TACTGTTTTCT	GAACGAAGTT	AAAGCAATTG	TGATACTAGT	ACTTTTTTAT
2801	CCGTTTTTCT	CTGAAAGATT	C'TTATGTTTA	AGTCATCATT	TACTGACTAT
2851	TGTTTTATTTT	TATCTTATAA	ACGCATGATT	CTGGGTTTTT	TCGAGTCTTT
2901	ACTGGAGACT	ATAGACAAC	GATTAGAAAAG	ACAGAAGAAA	ATCAAGTATG
2951	CAGAAGTGAC	AGAGAAACTC	ACCGTGATTG	TCATCGGGGA	AGAAGCCATG
3001	GAGAAACACA	TGGACATTTT	TCCGTTTCAA	GAGAAGTGTC	GACCTCAACA
3051	GCAGTTACTC	GATGGAGAAG	ATTCCGGATG	CTATCCACCC	ATGTCTGCTT
3101	ATCAGCCTTG	CTCTTGCACT	TGAACTTTGT	CAGACCTTTC	CCTGTACCGA
3151	TTCCGAAATA	GAGCTCTTCT	TCTGTGTTTT	CCGTGTCTTT	TCCATTTTGC
3201	AGGCCCTGAGA	CTGTCTCGCA	TACATCATTT	ACAAATGCCG	TAGTGATTCA
3251	ACCACTGGAT	TATTGTAAAG	CATTGGACAA	AGTCCCAAGA	TCACTCTGCT
3301	TATACAAATG	AACCCCTTGG	C'TTTGAAAAG	AAAGAAGAGT	AGCTCACTAA
3351	CCTTTGCTTT	TCATGGAGAA	TGCTCCTACA	ACGTGTTTGC	TTTTGATTTC
3401	TACTATGACC	TGCAAAAGAG	AAACCGAGCA	AAATTCAACT	CTAGAATTCA
3451	ACCAGCATAT	AACTTTGTTA	GCTTCAATGA	GTATTCGATG	TTTTAACCTG
3501	AGATTTTTTTT	GTTAAAAAAA	AAAAAAAAAA	AAAAAAAA	

Sequenz des abgeleiteten Gens A (*Arabidopsis thaliana*)

1	ATTAGATTTT	AGAGCTAACT	CTGAGAATTA	GGGTTCGTCT	CTCGACAAAC
51	AAAAATCTCA	GCCCCGTGGA	ATCTATTCTG	GTGTAAATCG	AAGAATGGAG
101	ACTAATGTTT	CGGAGAAGAA	GATCCTTGGT	TTCCCGGAAG	AGATGAATTC
151	GACAATTCAG	TTTTTATCTG	TTGAGAAAAC	AACATGTCGT	CTTCCGTGTT
201	CTATAAAAT	CAATATACCA	AAGAAAATTT	CTACAAAGAG	AAGAAGAATA
251	TCTGTAGAAG	AAGAAGAAGA	AGAAGAAGAC	GAGAATAAAA	CGAAGTTGAG
301	TTCTGGGGAG	AAGGTTTTGA	GTAGCGATGA	GATTTTGAGG	TTAAGGGAAG
351	TTTCAAGACG	AAAGTATTTG	ACGGATCGAG	AGAATAAGAA	GGTGGAGGAG
401	TTAAGGGATG	AGAGAAGTAA	TTGAGTTTAG	ATATAAGAGA	GCATTGTATG
451	AGTTGGTGAA	GAAAACAGAG	GATGATGATG	ATGTTGAAGG	GTATAGGTTT
501	CCAGATGCTT	ATGATCAAGA	AGGGTGTATT	GATCAGAAGA	AGAGGTTTGA
551	TGTTGCCAAG	GAGAGGTATT	GTGAGCGAAG	AAGGAGCGGG	AGAGTTGTTA
601	CGGAGCAAGA	AGCTTGGGAG	GATCATCAGG	CTCAGAAAGC	GAGAGTGAGG
651	TTTGGTGCAA	AGGATAAGAA	GCAAGTTGTT	GATGGGTATG	AGTTTGTGTT
701	TGATGATTTG	ACAGGC'TTTG	TTGAGGAGTC	TAGCGAGGCG	GAAACTGGGA
751	AACATCGAGG	TTGTTATTCG	AAGACTGCTG	CAGAGAAGGC	TCGGGAGGGG
801	AGGGAGTTTC	TTCCGATTCA	TGGGTACAGG	GAGGAACTGC	TGAAACTTAT
851	TGAAGAGAAT	CAGGTTCTTG	TCAT'TGTTGG	AGAGACAGGA	TCGGGTAAAA
901	CTACGCAGAT	ACCGCAGTAT	C'TTCAAGAAG	CCGGTTACAC	AAAGCGTGGA
951	AAGATCGGTT	GTACTCAGCC	TCGGAGAGTT	GCTGCAATGA	GCGTTGCTTC
1001	CCAGATGGCT	CAAGAGGTTG	GTGTTAAACT	TGGACATGAG	GTTGGATATT
1051	CCATTAGATT	TGAAGACTGT	ACTTCAGAGA	AAACAGTTAT	CAAGTACATG
1101	ACTGATGGGA	TGCTGCTACG	AGAGCTACTC	ATAGAACCGA	AACTCGATAG
1151	CTATAGTGTC	ATCATTATTG	ATGAGGCGCA	CGAAAGAACA	CTGTCCACTG
1201	ATATTTTGT	TGCGTTAGTG	AAGGATGTAG	CCAAGGTTAG	GCCTGATCTG
1251	AGGTTAATAA	TCTCCAGTGC	AACGCTTGAA	GCTAAAAAGT	TCTCCGAATA
1301	C'TTTGATTCG	GCTCGGATTT	ACCTGATTCC	AGGAAGAAGA	TATCCAGTTG
1351	AAAAGCTCTT	CAGAAAATGT	CCTGAACCTG	ACTACTTGGA	GACAGTGATA
1401	CGCACTGTGG	TTCAGATACA	TCAGACTGAG	GCGATAGGAG	ACATTTTAGT
1451	TTTCCTCACT	GGTCAAGAAG	AGATCGAAAC	AGTAGAAACG	AATTTGAAGC
1501	GAAGGATGAT	GGATTTGGGT	ACAAAGGGAT	CTGAGATCAT	TATCTGTCCCT
1551	ATCTACTCAA	ATCTCCCAAC	TCCACTACAG	GCAAAAAGTGT	TTGAACCGGC
1601	CCCTAAAGGG	ACACGAAAAG	TTGTCTTGGC	AACGAATATT	GCAGAAACAT
1651	CGTTAACCAT	TGATGGGGTT	AAATACGTCA	TTGATCCAGG	ATATTGCAAG
1701	ATCAACTCAT	ACAACCCGAG	AACGGGAATG	GAGTCACTCC	TTGTAACCTC
1751	TATTTCCAAG	GCTTCCGCG	CGCAACGAGC	TGGTCGGTCT	GGAAGAACAG
1801	GTCCTGGGAA	ATGTTTCCGG	TTGTACAATA	TCAAGGATTT	AGAGCCCACT

1851	ACAATACCGG	AAATCCAAAG	GGCTAACCTT	GCAAGTGTG	TGCTTACTTT
1901	GAAGAGTCCT	GGGATTC AAG	ATGTGTTC A A	TTTTGATTT C	ATGGATCCTC
1951	CACCTGAAAA	CGCTCTTTTG	AAGGCTCTAG	AGTTGCTCTA	TGCGCTGGGT
2001	GCTTTGGATG	AAATTTGGTGA	GATTACTAAG	gtaggagaaa	gaatggtgga
2051	gtttccagtt	gatccaatgc	tgtcaaagat	gattgttggt	tcggaaaagt
2101	ataagtgcctc	gaaagagata	atcacgatag	ctgccatggt	gtctgtagGG
2151	AACTCGGTCT	TCTATCGACC	AAAGAACCAG	CAAGTTTTCG	CCGACAAGGC
2201	TCGGATGGAT	TTCTATGAAG	ATACAGAGAA	CGTCGGCGAT	CACATTGCAT
2251	TGTTAAGGGT	TTACAATTCC	TGGAAGGAAG	AAA AACTACTC	AACTCAATGG
2301	TGCTGCGAGA	AGTTCATTCA	GAGCAAGAGT	ATGAAAAGGG	CAAGAGATAT
2351	ACGTGACCAG	CTACTTGGAC	TTTTAAACAA	GATTGGAGTA	GA A CTTACTT
2401	CCAACCCGAA	TGATTTAGAC	GCAATAAAGA	AAGCAATCTT	GGCCGGATTC
2451	TTCCCGCATT	CTGCAAAGTT	ACAAAAGAAT	GGATCGTACA	GAAGAGTCAA
2501	AGAACCTCAG	ACAGTTTATG	TACACCCTAA	CTCAGGCTTA	TTTGGGGCGT
2551	CACCTAGTAA	GTGGTTGGTG	TATCATGAAC	TAGTACTTAC	TACCAAGGAA
2601	TATATGAGGC	ATACA A CTGA	AATGAAGCCT	GAGTGGCTAA	TTGAAAATAGC
2651	TCCTCATTAC	TACAAGCTTA	AAGACATTGA	AGATACTCGG	CCAAAAGAAAA
2701	CACAGAGAAG	AATCGAAGAA	GCCTCGACGT	CTAAGGT TGA	TACTAACAAG
2751	AAAACCCGGA	CCTCGAAGGT	TGATACTAAC	AAGAAA TCAA	AGAGGTAACA
2801	ACCAAAC TAG	TCGTTTCTCT	TTACTGTTTC	TGAACGAAGT	TAAAGCAATT
2851	GTGATACTAG	TACTTTTTTTA	TCCGTTTTTTC	TCTGAAAAGAT	TCTTATGTTT
2901	AAGTCATCAT	TTACTGACTA	TTGTTTATTT	TTATCTTATA	AACGCATGAT
2951	TCTGGGTTTT	TTCGAGTCTT	TACTGGAGAC	TATAGACAAC	TGATTAGAAA
3001	GACAGAAGAA	AATCAAGTAT	GCAGAAGTGA	CAGAGAAACT	CACCGTGATT
3051	GTCATCGGGG	AAGAAGCCAT	GGAGAAACAC	ATGGACATTT	CTCCGgtgag
3101	attcatcaag	ccttattata	atcttcttca	ttatttccac	aaaacaatgt
3151	ttacaatttg	actgaaataa	aaacaagtga	cgaagatgaa	aaacactaca
3201	cctacttgca	gatatgtata	tgttcttgaa	acaatacata	tcattggcgt
3251	gtaatgaaaa	taaccaacac	taacaccggt	aacacaagac	atcatgattt
3301	atgtgctatc	gttaatgttt	atagTTTCAA	GAGAAGTGTC	GACCTCAACA
3351	GCAGTTACTC	GATGGAGAAG	ATTCCGGATG	CTATCCACCC	ATGCTGCTT
3401	ATCAGCCTTG	CTCTTGCACT	TGAACTTTGT	CAGACCTTTC	CCTGTACCGA
3451	TTCCGAAAATA	GAGCTCTTCT	TCTGTGTTTT	CCGTGTCTTT	TCCATTTTGC
3501	AGGCCTGAGA	CTGTCTCGCA	TACATCATTT	ACAATGCCTG	TAGTGATTCA
3551	ACCACTGGAT	TATTGTAAAG	CATTGGACAA	AGTCCCAAGA	TCACTCTGCT
3601	TATACAAATG	AACACCATAG	ACATTGAAAG	AAAGAAGAGT	AGCTCACTAA
3651	CCTTTGCTTT	TCATGGAGAA	TGCTCCTACA	ACGTGTTTGC	TTTTGATTTT
3701	TACTATGACC	TGCAAAAGAG	AAACCGAGCA	AAATTCAACT	CTAGAATTCA
3751	ACCAGCATAT	AAC TTTGTTA	GCTTCAATGA	GTATTCGATG	TTTAACTTGA
3801	GATTTTTT				

Sequenz des abgeleiteten Gens A (*Capsella rubella*)

1	AAAATCCCAG	TCCCCTGTTG	TAAATCGAAG	CATGGAGAGC	AATGATCTGG
51	AGAAGAAGAC	CACTGAT TCA	GCGGAAGAGA	TGAATTCGAG	AGATCAGGTA
101	CGAGCATCAT	CATCAGGTGA	GAATCTGGAT	CTGTATGAAC	ATAGGGAAGA
151	AGATGAAGAA	GGGTTTATGT	CGGTGGAGAA	GATATGTCGC	CTTCCTGTTT
201	CTATAAGATT	TAATCTACCA	AAGAAAAACT	CTACTAAGAG	AAGAAGAGAA
251	TCTAGAGAAG	AAGATGGCGA	GAAGAAGAAA	GCGAAAATGA	GTTCTGGAGA
301	TAAGGTTTTG	AGTAGCGATG	CGATTAGGAT	AGTGAGGAGG	CGAGGGTATC
351	TGAAGGATCG	AGAGAAGAA	AAGGTGGAGG	AGTTAAGGGA	CGCGATAGAA
401	GACGAGATTA	TGTTCTTTGA	TGGTGTGAAG	GTTACTGAGA	GAGAAGAAAT
451	TGCGCTTAGA	CATAAGATAG	AGTTGTATGA	ATTGGTGAAG	AAAACAGAGG
501	ATGATGATGA	TGACGATGTT	GAAGGTATA	GGATTCCAGA	TGCTTATGAT
551	CAAGATGGGT	GTATTAATCA	GAAGAAGAGA	TTTGCTGTGG	CCGAGCAGCG
601	TTATAGTGAG	GGGAAGAGGA	CCCGGAGAGT	CGTTACCGAG	CAAGAAGCTT
651	GGGAGGATCA	TCAGGCGCAG	AAAGCGAGAG	TAGGTTTTCGG	TGCAAAAGGAT
701	AAGAAGCAAG	TTTGTTGATGG	GTATGAGTTT	GTGTCTGACG	AGTTAGCAGG
751	TTTGTTGAG	GAT TCTAGTA	GTGAGGTGGA	AACTGGGAAG	CACCGAGGTC
801	GTTACTCGAA	AACAGTTGCA	GAAAAGGCTC	GGGAAGAGAG	GGAGTTTCTT
851	CCGATTCATG	GATACCGGGA	TGAATTGCTG	AATCTCATTG	AAAAAATCA
901	GGTCTTGTG	ATTGTTGGGG	AAACAGGATC	AGGGAAGACT	ACGCAGATAC

951	CGCAGTATCT	TCAGGAAGCC	GGTTACACAA	AGCGTGGAAA	GATCGGTTGC
1001	ACTCAGCCTA	GGAGAGTTGC	TGCTATGAGC	GTTGCTTCCC	GAGTGGCTCA
1051	AGAGGTTGGT	GTTAAACTGG	GACATGAAGT	TGGGTATTCC	ATTAGATTTG
1101	AAGACTGTAC	TTCAGATAAA	ACAGTTGTGA	AGTACATGAC	TGATGGGATG
1151	CTGTTGCGAG	AGCTACTCAC	AGAACCGAAA	CTTGATAGCT	ATAGTGTGAT
1201	CATAATTGAT	GAGGCACACG	AAAGAACCCCT	GTCCACTGAT	ATTTTGTTTG
1251	CCTTAGTTAA	GGATGTAACA	AAGGTTAGGC	CTGATCTGAG	GTTAATAATC
1301	TCAAGCGCAA	CGCTTGAAGC	TAAAAAGTTC	TCCGAATACT	TTGATTCGGC
1351	TCGGATTTAC	CTAATCCCAG	GACGAAGATA	TCCAGTTGAA	AAGCTATACA
1401	GAAGATGGCC	TGACGCTGAC	TACTTGGAGA	CAGCAATACG	TACTGTGTTT
1451	CAGATCCATC	AGACTGAGCC	GACTGGAGAC	ATTTTAGTTT	TCCTCACAGG
1501	TCAAGAAGAG	ATTGAAACAG	CAGAAACAAA	GTTGAAGCGA	AGGATGATGG
1551	ATTTGGGTAC	AAAGAGCTCT	GAGATCATT	TCTGTCCTAT	CTACTCAAAT
1601	CTCCAACCG	AAC TACAGGC	AAAAGTGTTC	GAACCAGCTC	CTAAAGGGTC
1651	GCGGAAAGTT	GTCC TGGCAA	CCAATATCGC	AGAAACATCA	TTAACCATTG
1701	ATGGGATAAG	GTACGTCATT	GATCCGGGAT	ATTGTAAGAT	CAACTCATA
1751	AACCCGAGAA	CCGGAATGGA	GTCATTGCTT	GTAAGTCCCTA	TTTCCAAGGC
1801	CTCCGCAGAG	CAACGAGCTG	GTCGGTCTGG	AAGAACAGGT	CCAGGAAAAAT
1851	GTTTCCGGTT	TACAATGTA	AAGGATTTGG	AACCCACCAC	AATACCGGAA
1901	ATCCAAAGGG	CTAACCTTGC	AAGTGTGTG	CTTACTTTGA	AGAGTCTTGG
1951	GATTCAAGAT	GTGTTTAATT	TTGATTTTCAT	GGATCCTCCA	CCAGAAACCG
2001	CTCTTTTGAA	AGCTCTAGAG	TTGCTCTATG	CTCTGGGTGC	TTTGGATGGC
2051	AATGGTGAGA	TTACTAAGgt	aggagaaaga	atgggtggagt	ttccagttga
2101	tccgatgcta	tcaaagatga	ttggttggtc	ggaaaagtac	aagtgctcga
2151	aagagataat	cacgatagcc	gccatggtgt	ctgtagGGAA	CTCAATCTTT
2201	TATCGACCAA	AGAACCAGCA	ACTTCTCGCC	GACAAATGCTC	GGATGAATTT
2251	CTATGTAGAT	ACAGAGAATG	TAGGCGACCA	CATTGCGTTG	TTAAGGGTCT
2301	ACAATTCCTG	GAAAGAAACA	AACTACTCGA	CTCAATGGTG	CTACGAGAAC
2351	TTCGTTTACA	GCAAGAGTAT	GAAAAGAGCT	AGAGATATAC	GTGACCAGCT
2401	GATTGGACTT	CTAAACAAGA	TTGGAGTAGA	ACTTAATTC	AACCCGAATG
2451	ATTTAGACGC	AATAAAGAAA	GCGATCTTGG	CCGGATTCTT	CCCGCATTC
2501	GCAAAGTTAC	AAAAAGATGG	ATCGTACAGA	AGAGTGAAAG	AACCTCAGC
2551	AGTTTATGTA	CATCCTAACT	CAGGCTTATT	TGGGGCATCA	CCTAGTAAGT
2601	GGTTGGTGTA	TCATGAACTA	GTACTTACTA	CCAAGGAATA	TATGAGGCAT
2651	ACAACCGAGA	TGAAACCCGA	GTGGCTAATT	GAAATAGCTC	CTCACTACTA
2701	CAAGGTTAAA	GACATTGAAG	ACACTCAGCC	AAAGAAAACG	CACAGAAGAA
2751	GCGAAAGAGC	CTCAACGTCA	AAGGTTGATA	CTCACAAGAA	AACCCGGACA
2801	TCAAAGGTCG	ATACGAACAA	GAAAGCAAAG	AGGTAACAAC	CAAACCTAGT
2851	CATATCTCTT	TACTGTTTCT	GAACGAAGTT	AAAAAAATTG	TGCTACTAGA
2901	ACTTTTTTCAT	CCATTCCTCT	CGGAAGATAC	TTTTTATGTTT	AGTCGGCATC
2951	AGTTTTTTTTA	ATCATTACTA	ACTACTGTTT	CTTCCATCT	TATAAACGCA
3001	TGATTCTGAG	TTATTTTCGAG	TTTACTGGGG	ACTATAGACA	ACTGATAAGA
3051	AAGGGGAAAA	TCAAGAATTC	AACAGCGACT	GAAAAACTCA	CCCGGTATAT
3101	ACACGGGGAA	GATGCCAGGA	TTTTACGCGG	TGGTGGAGAA	ACACCTGGAC
3151	ATTCTTCCGg	tgagattcat	taggccttag	aataatcttc	ttcattat
3201	ttagaaaaaa	aaatgtttac	aactgactga	aataaaaaact	gcgaagatga
3251	aaaaacatca	caactactta	cagatatgta	tatgtaatga	aaataaccaa
3301	ctaatactct	tataacaaga	catggcttat	gtgttatgtg	ctatcggcaa
3351	tgtttgtagT	TTCAAGAGAA	GTGTTGGTAA	CCTCAACAGC	AGTTACCCGA
3401	TGGAGAAGAT	TCCGGATGTC	ATCCATCCAT	GTCTGCTTAT	CAGCCTTGCT
3451	CTTGCACTTG	AAC TTCGTTA	GACCTTTCCC	AGTACTGATT	CCGAAGTAGA
3501	GCTCTTCTTC	TGCGTTTTTCT	GTTTTCTTTTC	CATATTGCAG	GCCCGAGACG
3551	ATCTCGCATA	CATCGTTTAC	AATTCCTATA	GTGTGTAGTG	ATATAACCAT
3601	TGGATCAAGT	AAGCATTTGA	AAAAGTCTCA	AGACTACTTT	GGCTTTACAC
3651	AAATGAACAC	CACGACCATT	GAAAGAAAGA	AGAGATATCT	GAAGGTGACA
3701	ATAGTTAGTT	AAC TACCTTT	GCTTTTCATG	GAGAACGCTC	CTCTGACGTG
3751	TTTGCTTTTTG	ATTTCTACTA	CAACCTGCAA	GAGAAAACGA	GTGAATGTC
3801	AACTCTAGAA	TTTGACAACC	ATAAACTTTT	TTTGCTTTGA	TGAGTTTTTG
3851	CTTTTTTTTTT	TTAACCTGAG	ATCTCT		

Sequenz des *Arabidopsis thaliana* cDNA-Klons ATTS1135 (EST 3)

```

1   CTTTCTTCTC TGTCCTCAACT ACCTTATTTCG CGTATCTCCT CCGTGCGGTT
51  TCTTTTCTTC AACGATAAAC TTCAACTTAT CCTTCATACA TTCCGATCTA
101 GATAACAAAC TAATTGGGAA TAATTAGATG ATTTGTTTGG AAGGTGATTG
151 ATATGATATT ATGATTCTAT TAAGTAATTT CCTTTATGTT ATTTGATTTG
201 AGTTTGAATA AGCTTTACAT GCGAAGCAAT CCAGGGTAAAG AGAGATAGAA
251 ATCTCCTTGT AAAGGAATGT TATTTGTATC CTTGGTTTCT TTTTGTCAAT
301 AAATCTATCA TAGTAGCTAA G

```

Sequenz der vorhergesagten cDNA des Gens 3 (*Arabidopsis thaliana*)

```

1   ATGAAACCGG AGTTTATACG CCATTATTAC TATAGCCAAA GCCCTCTTGA
51  GGACGTAAC TGTGCATCTA AGCTGTTGCG TCCTGCTCCT ATGAGGGCCT
101 TTCAAGATCT TGATAAGCTA CCTCCAAATC CCGAGGCCGA GAAAGTTCCCT
151 CGAGTTTACA TCAGACTGC TAAGGATAAT CTATTTGATT CTGTGCGTCA
201 AGACCTTTTG GTGGAGAATT GGCCACCTTC TCAGCTGTAT GTCTTGGAGG
251 ATAGTGACCA TTCTGCTTTC TTCTCTGTCC CAACTACCTT ATTCGCGTAT
301 CTCCTCCGTG CGGTTTCTTT TCTTCAACGA TAA

```

Sequenz des abgeleiteten Gens 3 (*Capsella rubella*)

```

1   ATGAAACAAG AGTTTCTACG CCAATATTAC TATAGCCAAA GCCCTCTTGA
51  GGTACTCTC TTATTTACTT CACAAGCTAA TGTTTCAATA TTACTATTGT
101 GTGGTTATTG AATATGGTTA TTGGTAGGAT GTAAGTTTGG CAACTAAACT
151 GCTGCGTCC TCTCCTATGA GGCATTTC AAGTCTTGAT AAGTCGCCCTC
201 CAAATCCCGA GGTAGAGAAA GTTCCGCGAG TCTACATCAA GACTGGTAAG
251 GATAATCTAT TTAGTTCGGT GCGTCAAGAC CTTTTGGTGA AGAACTGGCC
301 ACCCTCTCAG TTCTATGTCT TGGAGGAGAG TGACCATTCT GCTTTTTTCT
351 CTGTGCCAAC CACCTTATTC GTCTATCTCC TCCGTGCCGT TTCTTTTTCTT
401 CACAAATAA

```

Sequenz des *Arabidopsis thaliana* cDNA-Klons ATTS1638 (EST 4)

```

1   GGGAAAGAAG AACAGAGGTC TCCGCGGAAA GGGTCACAAC AACCACAAGA
51  ACCGTCCATC TCGCAGGGCT ACATGGAAGA AAAACAACCTC TCTCAGCCTT
101 CGTCGTTACC GTTGATTGCC TTTTGCTTTT ATTATTACTG TTATCATCTT
151 CTAGGTTCCA TGTGTTGCA CTTTATCTC TCTCAGGATG ACTTGTTTTT
201 GCAAATTTTG GAGACTATAC AATTATGCTT TGAACCTAAA TTTTATATTT
251 GCTTTGCTTT AA

```

Sequenz des *Arabidopsis thaliana* cDNA-Klons 25012 (EST 4)

```

1   CAAGTATGTA TCTGAGCTAT GGAGGAAGAA ACAGTCCGAT GTGATGAGAT
51  TCCTCCAGAG GGTAGGTGC TGGGAGTACA GACAGCAACC TTTCGATTGTT
101 CGTCTCGTCA GGCCTACTCG TCCTGATAAG GCTCGTCGTT TGGGTTACAA
151 GGCCAAGCAG GGCCTTGTG TGTAACCGTGT ACGTGTGAGA CGTGGTGGAC
201 GCAAGAGGCC AGTGCCTAAG GGTATTGTGT ATGGTAAGCC AACAAACCAG
251 GGAGTGACAC AACTCAAGTT CCAGCGTAGC AAGCGTCTG TTGCTGAGGA
301 GCGTGTGGC AGGAAATTGG GTGGTCTCAG AGTTGTCAAC TCTTACTGGC
351 TCAATGAGGA TTCGACCTAC AAGTACTACG AGATCATCTT GGTGACCCA
401 GCACACAATG CTGTGCGTAA TGACCCGAGA ATCAACTGGA TCTGCAACCC
451 AGTTCACAAG CACCGTGAGC TCAGAGGACT TACCTCAGAG GGAAAGAAGA
501 ACAGAGGTCT CCGCGGAAAG GGTCACAACA ACCACAAGAA CCGTCCATCT
551 CGCAGGGCTA CATGGAAGAA AAACAACCTCT CTCAGCCTTC GTCGTTACCG
601 TTGATGGCCT TTTGCTTTTA TTATTACTGT TATCATCTTC TAGGTCCAT
651 GTTGTGTCAC TTTTATCTCT CTCAGGATGA CTTGTTTTTG CAAATTTTGG
701 AGACTATACA ATTATGCTTT GAACCTAAAT TTTATATTTG CTTTGCTTTT
751 TTGTTGATCT CAGTCTGGTT TTGGAATATA TCGAGG

```

Sequenz der vorhergesagten cDNA des Gens At4g16720 (*Arabidopsis thaliana*)

```

1  ATGGGTGCGT ACAAGTATGT ATCTGAGCTA TGGAGGAAGA AACAGTCCGA
51  TGTGATGAGA TTCCCTCCAGA GGGTTAGGTG CTGGGAGTAC AGACAGCAAC
101 CTTCGATTGT TCGTCTCGTC AGGCCTACTC GTCCTGATAA GGCTCGTCGT
151 TTGGGTTACA AGGCCAAGCA GGGCTTTGTT GTGTACCGTG TACGTGTGAG
201 ACGTGGTGGA CGCAAGAGGC CAGTGCCTAA GGGTATTGTG TATGGTAAGC
251 CAACAAACCA GGGAGTGACA CAACTCAAGT TCCAGCGTAG CAAGCGTTCT
301 GTTGCTGAGG AGCGTGCCTGG CAGGAAATTG GGTGGTCTCA GAGTTGTCAA
351 CTCTTACTGG CTCAATGAGG ATTTCGACCTA CAAGTACTAC GAGATCATCT
401 TGGTTGACCC AGCACACAAT GCTGTGCGTA ATGACCCGAG AATCAACTGG
451 ATCTGCAACC CAGTTCACAA GCACCGTGAG CTCAGAGGAC TTACCTCAGA
501 GGGAAAGAAG AACAGAGGTC TCCGCGGAAA GGGTCACAAC AACCACAAGA
551 ACCGTCCATC TCGCAGGGCT ACATGGAAGA AAAACAACCT TCTCAGCCTT
601 CGTCGTTACC GTTGA

```

Sequenz des abgeleiteten Gens 4 (*Capsella rubella*)

```

1  ATGGGTAAAT TCGATCTTCT CCATCTTTTCG TCTAGATCCA TTCTCCTGCT
51  ATGTTTTTCC AAATGATGTA TGTGTTGCGG ATCTGTTATT ACGAGCTACT
101 TAGCTAAATG GATTTTGGTG AACCTTAGGT GCGTACAAGT ATGTATCGGA
151 GCTATGGAGG AAGAAACAGT CCGATGTGAT GAGGTTCCCT CAGAGGGTGA
201 GGTGCTGGGA GTACAGACAG CAACCTTCGA TTGTTGCTCT CGTTAGGCCT
251 ACTCGTCCTG ATAAGGCTCG TCGTTTGGGT TACAAGGCCA AGCAGGTGAT
301 TGATTTTCATC CTAAATTTCT CGTCCTTGTT GTAGTTTTTA CTAATGTTGC
351 CTCGTATATA ATGAGGCTAG TGAACTTTGT TGCTCAACTG TCAAAATGAGT
401 TTTGTAGGAG TTGATTGATA TTGGTAGAAA TGAGATCCCTG TAAGGTAGAT
451 AGTGAGGTTT TTTTTTTGAT ATTGTTGTGA TATCTGCGTA TGATCTAGCC
501 TCTGTTGGTT GAGTTTGATG TTTTTTTGCG TTGTTAGTAG AGATTTAGAA
551 GCGTTACTGG TTATGTGGTC TCTTGAAACA CGCCATTTCT GAAAATTTGT
601 TGTTAGTGAT TGAATCCCTC CCATGACCAG CTACATGAAA CTAGCCAGTC
651 TGATATGCTG ATTTATACTT TTATACAAAT CTAGCTCTTG TTCTGACTGG
701 TACATGTAAA GATAGTCTT GCTCTAGCTT CAGAATCTGG TACTAATCTT
751 GTTGGGTTGT TGATGTTTGT TTAGGGCTTT GTTGTGTACC GTGTCCGTGT
801 GAGACGTGGT GGACGCAAGA GGCCAGTGCC CAAGGGTATT GTGTATGGTA
851 AGCCTACAAA CCAGGGAGTG ACACAACCTCA AGTTCCAGCG TAGCAAGCGA
901 TCTGTTGCTG AGGAGCGTGC TGGCAGGAAA TTGGGTGGTC TCAGAGTTGT
951 CAACTCTTAC TGGCTCAATG AGGTATACTT ACTACCTATT TTCAGAACTC
1001 CATTTTGGTG ACTCTTGGTT TTACAGATAA CAGATGTTAA TTTACCAATT
1051 GTTTTCAATA GGATTCGACT TACAAGTACT ACGAGATCAT CTTGGTTGAC
1101 CCAGCACACA ATGCTGTCCG TAATGACCCG AGGATCAACT GGATCTGCAA
1151 CCCAGTTCAC AAGCACCGTG AGCTCAGAGG ACTTACTTCA GAGGGAAGA
1201 AGAACAGAGG TCTTCGTGGA AAGGGTCACA ACAACCACAA GAACCGCCCA
1251 TCTCGCAGAG CTACATGGAA GAAGAACAAC TCCATCTCCC TTCGTCGTTA
1301 CCGTTAA

```

Sequenz des *Arabidopsis thaliana* cDNA-Klons B

```

1  GGAGATAGGA AGAGAAGGCG GAGGTGCGATG AAGAAGCGGA GTTCGCATAA
51  GAGAAGGAGT TTGAGTGAGT CTGAGGATGA GGAAGAGGGA AGGAGTAAGA
101 GAAGAAAGGA GAGGAGAGGA AGATGGGAAA TGGAGCATAT AACTCAATGG
151 ACAACGGGTT TCTTATGTTT GATCATGTTC GCATTCCTAG AGATCAAATG
201 CTCATGAGGT TGTGAAAGT TACAAGGGAA GGAAAATATG TTCCATCGGA
251 TGTTCCAAAG CAGCTGGTAT ATGGTACTAT GGTGTATGTG AGACAAGCAA
301 TTGTGGCAGA TGCTTCCATT GCACTGTCTC GAGCAGTTTG CATAGCTACA
351 AGGTACAGCG CAGTGCAGG GCAGTTTGGC GCACATAATG GTGGCATTGA
401 GACACAGGTG ATTGATTATA AAACCTCAGCA GAACAGGCTA TTTCTCTGTC
451 TAGCATCTGC ATATGCATTT CGATTTGTTG GAGAGTGGCT AAAATGGCTG

```

501 TACACGGATG TAACTGAAAG ACTGGCGGCT AGTGATTTTG CAACTTTGCC
551 TGAGGCTCAC GCATGCACTG CAGGATTGAA GTCTCTCACC ACCACAGCCA
601 CTGCGGATGC CATTGAAGAA TGTCGTAAGT TATGTGGTGG ACACGGCTAC
651 TTGTGGTGCA GTGGGCTCCC CGAGCTGTTT GCTGTATATG TTCCTGCCTG
701 CACATACGAA GGAGACAATG TTGTGCTGCA ATTACAGGTT GCTCGATTCC
751 TCATGAAGAC AGTCGCCAG CTGGGATCTG GAAAGGTTC TGTGGCACA
801 ACTGCTTATA TGGGCCGGC AGCACATCTT TTGCAATGTC GTTCTGGTGT
851 TCAAAAGGCT GAGGATTGGT TAAACCCTGA TGTGTACTG GAAGCTTTCG
901 AGGCTAGGGC TCTAAGAATG GCTGTACTT GTGCCAAAAA TCTCAGCAAG
951 TTTGAGAATC AGGAACAAGG ATTCCAAGAG CTCTTGGCTG ATTTGGTTGA
1001 AGCCGCTATT GCTCATTGCC AATTGATTGT TGTTTCCAAG TTCATAGCGA
1051 AACTAGAGCA AGACATAGGT GGCAAAAGG TGAAGAAACA GCTGAACAAT
1101 CTGTGTTACA TTTATGCTCT TTATCTCCTC CACAAACATC TCGGCGATTT
1151 CCTGTCCACT AACTGCATCA CTCCAAAACA AGCCTCTCTT GCTAACGACC
1201 AGCTCCGTTT CTTATACACT CAGGTCCGCC CTAATGCGGT TGCACCTGTG
1251 GACGCCTTCA ATTACACAGA CCATTACTTG AACTCGGTGC TGGGCCGTTA
1301 CGACGGTAAT GTGTACCCAA AGCTCTTTGA GGAAGCGTTG AAGGATCCAT
1351 TGAACGACTC GGTGGTTCCT GATGGGTACC AAGAATACCT TCGACCTGTG
1401 CTTAAGCAGC AACTTCGTAC CGTAGGCTC TGAAGAGTTT TCTTTGCTTG
1451 ATACTCGATA TGGTTAATCA CATTTGACTT GCTTCGTCTT TCTTCTTCGT
1501 CTTCTTCGTC ATCTCGCTTT GAATAATTTT GCAGTTTAAA AACTGGCGAT
1551 GCCCTTATTT ATATGTAGCA ATGTAATAGT TAATGTACGA TCGTCATATG
1601 GCGGAATTTT AGTACTATTT TTCGTTTTCA ATGCAACATT AATACAAAAA
1651 AAAAAAAAAA AA

Sequenz des *Arabidopsis thaliana* Acyl-CoA Oxidase cDNA-Klons (AF057044)

1 TTTTTTTCCT ATCATCTCTG AGAGTTTTCT CGAGAACTT TTGAGTGT
51 AGCTACTAGA TTCTGAATTA CGAATCATGG AAGGAATTGA TCACCTCGCC
101 GATGAGAGAA ACAAAGCAGA GTTCGACGTT GAGGATATGA AGATCGTCTG
151 GGCTGGTTCC CGCCACGCTT TTGAGGTTTC CGATCGAATT GCCCGCCTTG
201 TCGCCAGCGA TCCGGTGTTT GAGAAAAGCA ATCGAGCTCG GTTGAGTAGG
251 AAGGAGCTGT TTAAGAGTAC GTTGAGAAAA TGTGCCCATG CGTTTAAAAAG
301 GATTATCGAG CTTCTCTCA ATGAGGAAGA AGCAGGAAGA TTGAGGCACT
351 TTATCGACCA GCCTGCCTAT GTGGATCTGC ACTGGGGAAT GTTTGTGCCT
401 GCTATTAAGG GGCAGGTAC AGAGGAGCAG CAGAAGAAGT GGTGTGCGCT
451 GGCCAATAAG ATGCAGATTA TTGGGTGTTA TGCACAGACT GAGCTTGGTC
501 ATGGCTCAAA TGTTCAAGGA CTTGAGACAA CTGCCACATT TGATCCCAAG
551 ACTGATGAGT TTGTAATTCA CACTCCAACT CAGACTGCAT CCAAATGGTG
601 GCCTGGTGGT TTGGGAAAAG TTTCTACTCA TGCTGTTGTT TACGCTCGTC
651 TCATAACTAA CGGAAAAGAC TACGGTATCC ATGGATTCAT CGTGCAACTG
701 CGAAGCTTAG AAGATCATT TCCTCTCCG AATATAACTG TTGGTGATAT
751 CGGGACAAAG ATGGGAAATG GAGCATATAA TTCAATGGAC AACGGGTTTC
801 TTATGTTTTGA TCATGTTTCG ATTCCTAGAG ATCAAATGCT CATGAGGCTG
851 TCAAAAGTTA CAAGAGAAGG AGAATATGTT CCATCGGATG TTCCAAAGCA
901 GCTGGTATAT GGTACTATGG TGTATGTGAG ACAAACAATT GTGGCTGATG
951 CTTCCAATGC ACTATCTCGA GCAGTTTGCA TAGCTACAAG ATACAGTGCA
1001 GTGCGGAGGC AATTTGGCGC ACATAATGGT GGCATTGAGA CACAGGTGAT
1051 TGATTATAAA ACTCAGCAGA ACAGGCTATT TCCTCTGCTA GCATCTGCAT
1101 ATGCATTTTCG ATTTGTTGGA GAGTGGCTAA AATGGCTGTA CACGGATGTA
1151 ACTGAAAGAC TGGCGGCTAG TGATTTGCGA ACTTTGCCTG AGGCTCATGC
1201 ATGCACTGCA GGATTGAAGT CTCTCACCAC CACAGCCACT GCGGATGGCA
1251 TTGAAGAATG TCGTAAGTTA TGTGGTGGAC ATGGATACTT GTGGTGCAGT
1301 GGGCTCCCCG AGCTGTTTGC TGTATATGTT CCTGCCTGCA CATA CGAAGG
1351 AGACAATGTT GTGCTGCAAT TACAGGTTGC TCGATTCCCTC ATGAAGACAG
1401 TCGCCCAGCT GGGATCTGGA AAGGTTCCTG TTGGCACAAC TGCTTATATG
1451 GGCCGGGCAG CACATCTTTT GCAATGTCTG TCTGGTGTTT AAAAGGCTGA
1501 GGATTGGTTA AACCTGATG TTGTACTGGA AGCTTTTCGAA GCTAGGGCTC
1551 TCAGAAATGGC TGTACGTGT GCCAAAAATC TCAGCAAGTT TGAGAATCAG
1601 GAACAAGGAT TCCAAGAGCT CTTGGCTGAT TTGGTTGAGG CCGCTATTGC
1651 TCATTGCCAA TTGATTGTTG TTTCCAAGTT CATAGCGAAA CTGGAGCAAG
1701 ACATAGGTGG CAAAGGAGTG AAGAAACAGC TGAATAATCT GTGTTACATT

1751	TATGCTCTTT	ATCTCCTCCA	CAAACATCTC	GGCGATTTCC	TCTCCACTAA
1801	CTGCATCACT	CCCAAACAAG	CCTCTCTTGC	TAACGACCAG	CTCCGTTCCCT
1851	TATACACTCA	GGTCCGGCCT	AATGCGGTTG	CACTTGTGGA	CGCCTTCAAT
1901	TACACCGACC	ATTACTTGAA	CTCGGTTCTT	GGCCGTTACG	ACGGTAATGT
1951	GTACCCAAAG	CTCTTTGAGG	AAGCGTTGAA	GGATCCATTG	AACGACTCGG
2001	TGGTTCCTGA	TGGGTACCAA	GAATACCTTC	GACCTGTGCT	TCAGCAGCAA
2051	CTTCGTACCG	CTAGGCTCTG	AAGAGTTTTC	TTTGCTTGAT	ACTCGATATG
2101	GTTAATCACA	TTAGACTTGC	TTCGTCCTTC	TTCTTCGTCT	TCTTCTTCTT
2151	CTCGCTTTGA	ATAATTTTCG	AGTTTTAAAA	CTGGCGATGC	CCTTATTTAT
2201	ATGTAGCAAT	GTAATAGTTA	ATGTACGATC	GTCATATGGC	GGAATTTTAG
2251	TACTATTTTT	CGTTTTCAAT	GCAACATTAA	TACAATTGAT	CGTTTCTACT

Sequenz des abgeleiteten Gens B (*Capsella rubella*)

1	ATGGAAGGAG	TTGATCACCT	TGCTGATGAG	AGGAACAAGG	CGGAGTTCGA
51	CGTCGACGAG	ATGAAGATCG	TCTGGGCTGG	TTCCCACCAC	GCTTTTGAGG
101	TTTCCGATCG	AATTGCCCGC	CTCGTCGCCT	CCGATCCGGT	TAATTTCCCTT
151	CTTCTTCTCA	ATTTCTCAAT	TCTCCTTGTT	TTTTGGTAGG	ATTAGGGATT
201	TTTCATGGTG	TTCGATTTTG	TGAATTTTCGT	TTTTGTGTTT	TTGTTTATGT
251	ATAGGTCTTT	GAGAAAAGCA	ATAGAGCTAG	GTTGAGTAGG	AAGGAGCTGT
301	TTAAGAGCAC	TTTGAGGAAA	TGTGCTCATG	CTTTTAAGAG	GATTATTGAG
351	CTTCGTCTCA	ATGGTAATGG	AGGATTCCTC	TTCTCACGCT	CCTCCTAATT
401	GTTTTTCGTTG	ATTC AAGGTT	TTTGTATTGA	GAGTCTTTTG	TTGGTTACAG
451	AGGAAGAAGC	AGGAAGATTG	AGGCACTTTG	TAGATCAGCC	TGCCTTTGTG
501	GATCTGCAC T	GGGTACTCTT	TTTTTTTTTTT	CTCAATCAAT	AGATATGGTT
551	TATAAAGACA	CCGTTTTGAT	TTTTTTTATTC	ATGGTATCTC	TTTCTGTAGG
601	GAATGTTTGT	CGCTGCTATT	AAAGGGCAGG	GTACTGAGGA	GCAGCAGCAG
651	AAGTGGTTGT	CGCTGGCTAA	TAAGATGCAG	ATTATTGGGT	GTTATGCACA
701	AACTGAGCTT	GGCCATGGCT	CTAATGTTCA	AGGCCTTGAG	ACA ACTGCCA
751	CTTTTGATCC	CAAGACGGAT	GAGTTTGTTA	TTCATACACC	AACTCAGACT
801	GCATCTAAAG	TGAGTTGACG	ACTAAGAAGA	AGCCTTTCAT	TGAGTTATTA
851	TTATGCACTG	TGACTTAGTT	GTGTAAGAAT	TTTATCTGCT	TTGATTGTTG
901	AATAGTGGTG	GCCTGGTGGC	CTGGGAAAAG	TTTCTACTCA	TGCTGTTGTT
951	TACGCTCGTC	TCATTACTGA	CGGCAAAGAC	TATGGTATCC	ATGGTATGTA
1001	CTGCAAAACA	AAACAAAAAA	GCCAGTATTT	TTTAGTTAAA	TTTTTCTTGT
1051	TCAGGTATTG	ATTGCTTAAA	CTTCCTCTTC	CAGGATTCAT	CGTGCAACTG
1101	CGCAGCTTAG	AAGATCATTC	TCCTCTTCCA	AATATAACTG	TTGGTGATAT
1151	TGGGACGAAG	ATGGGAAATG	GAGCATATAA	TTCAATGGAC	AACGGGTTTC
1201	TTATGTTTGA	TCATGTTTCG	ATTCCTAGAG	ATCAAATGCT	CATGAGGTCC
1251	TCCCTCTTTG	CCTTCTTGTT	CTGAGTTTAA	TATGCATTCA	AGATGTAATG
1301	ATAGGTTTAG	AGAAATCGCT	TGTGCTTAAA	TGCGGGCTTA	TGGTCTACA
1351	TTTTTCTCCT	CTTTCAAAAC	GTTTCAGGCT	GTCCAAAGTT	ACAAGGGAAG
1401	GAAAATATGT	TCCATCGGAT	GTTCCAAAAC	AGCTGGTGTA	CGGTACTATG
1451	GTGTATGTGA	GACAAACAAT	TGTGGCAGAT	GCTTCCAATG	CTCTATCTCG
1501	AGCAGTTTGC	ATAGCTACAA	GGTACAGTGC	AGTGAGGAGG	CAATTTGGCG
1551	CACATAATGG	TGGCATTGAG	ACACAGGTAC	TCATTTAATA	TTCAAGGATA
1601	CTCATGGCTG	CGTACACCTA	TTGATGTTAG	ATCATAACAA	ATTTCAAAGT
1651	ATTTCTACCT	TTTTCTTGTA	TGAATATTTT	CCTGATTTAA	GCAGATTTGT
1701	CTGTAAGTTT	CTGCTGAAC T	GATTTTCTTC	GATTTGGTTG	TGTAGAGTTT
1751	GCTTTTTTTTT	CTTTTGCAAT	CATTAACGTT	TTTTGGTTAT	CACAGGTGAT
1801	TGATTATAAA	ACTCAGCAGA	ACAGGCTATT	TCCTCTGCTA	GCGTCTGCAT
1851	ATGCATTCCG	ATTTGTTGGA	GAGTGGCTAA	AATGGCTGTA	CACGGATGTA
1901	ACTGAAAGAC	TGGCGGCTAG	TGATTTTACA	ACTTTGCCAG	AGGCTCATGC
1951	ATGCACTGCA	GGATTGAAGT	CTCTCACCAC	CACAGCCACT	GCGGTATGTG
2001	TCTGAGCTGA	TCCTTTATGC	TCACTAGACT	AAACAGATGA	ATTTCTTTTC
2051	TCTGAGGGTC	CTACGGCTCG	CATGTTGAAA	ATTTTGTTTG	TTACCTCGAC
2101	ATTGTCTTGG	TGGTTTATTT	AGAATAGAAA	AGTCAGCCTA	AGCGTTATAT
2151	TGGGTGCAGG	ATGGCATTGA	AGAATGTCGT	AAGTTATGTG	GTGGACATGG
2201	CTATTTGTGG	TGCAGTGGGC	TACCTGAGCT	GTTTGCTGTG	TATGTTCCCTG
2251	CCTGCACATA	TGAAGGAGAC	AATGTTGTGC	TGCAATTACA	GGTACTTATT
2301	ATCAGCTACG	ACCTATTTCC	ACTTACTAGA	ATAA TACTGT	CAGCTCAAA T
2351	ATATTTTCGTA	GAACGTGGGT	CTTCCTCTTA	TGTCTCTTTT	GTGTTT TAGC

2401	TGTGAGCTCT	TGATTTAACA	GGCTTATCGT	AGTTTTCTAC	TTTCTTAGAT
2451	ATGATTATTT	GCAAGGTTTT	GTTAATTTCA	CCTGATCCCA	GCTGGGATAC
2501	TGCTTCTCTGA	AGAATTGAGC	AACATATACA	CATAATTTGG	GCTTCGATCA
2551	CCTGCATTTT	GTTTTTGTTT	TTCTGGTTGA	TATGAAAAAC	CACATTTAAC
2601	TGCTTTTGAA	GTTTATGGTT	CATTGTTGTT	TCCCAATAT	TCAATAAGAA
2651	AAGCTCTAAC	CCCCCTTGCC	ATTTGTGGTT	ATTCACGCTA	ATGTTTGAAA
2701	TTGTATGTAT	AGGTTGCTAG	ATTCCTTATG	AAGACAGTCG	CCCAGCTGGG
2751	ATCCGAAAAG	GTTCCGTGTG	GCACAACCTGC	TTATATGGGC	CGGGCAGCAC
2801	ATCTTTTGCA	ATGTCGTTCT	GGTGTTCAAA	AGGGTAAAGTT	TTTAAAGTTTA
2851	TATATTCTAT	TTTCGTA CTG	ATACAATAAC	AAACGAATGA	TCTTTCCTCG
2901	AGAAAATTCA	CGAGTTAATG	GTATAGTTGG	TTTTATTTAT	TGGGTATGTA
2951	GCTGAGGATT	GGTAAAACC	TGGTGCGGTA	CTGGAAGCTT	TTGAAGCTAG
3001	GGCTTTGAGA	ATGGCTGTTA	CTTGTGCCAA	AAATCTCAGC	AAGTTTGAGA
3051	ATCAGGAACA	AGGTACTCTT	CTAACTCCGT	GAATCCAGAA	CAACTGGTCG
3101	TCATGTCACA	TCCTAAGCCA	AAAGACCAA	CTCTCACATA	TTTCTCCTGT
3151	CTATAGGATT	CCAAGAGCTT	TTGGCTGATC	TGGTTGAGGC	TGCAATTGCC
3201	CATTGCCAAT	TGATTGTTGT	TTCCAAGTAA	GCTACTTCAG	AATCCGTTCA
3251	CACTTATCAA	ATAAGTTAGT	GATTCCTCTT	CTAGATCAGC	TCCTGAATCT
3301	TGCCATTTTT	GTCTCGGTTA	GGTTCATAGC	TAAACTAGAG	CAAGACTAG
3351	AAGGCAAAGG	AGTGAAGAAA	CAGCTGAATA	ATCTGTGTTA	CATTTACGCG
3401	CTCTACCTCC	TTCAACAACA	CCTCGGAGAT	TTCTCTCCA	CAAACACCAT
3451	CACTCCCAA	CAAGGGTCGC	TCGCCAACGA	CCAGCTCCGC	TCCTTGTA CT
3501	CTCAGTTTTT	ACATTGCTTT	CTCGCTGCAT	TTGATTCTT	CCAAAAGCTG
3551	AGTCCTAACA	CTTTGTGTTT	CATCATCTAG	GTTCTGCTTA	ATGCAGTTGC
3601	GCTTGTGGAC	GCGTTCAATT	ACACTGACCA	CTACTTGAAC	TCGGTGCTGG
3651	GCCGTTACGA	CGGTAATGTG	TACCCGAAAC	TGTTTGAGGA	AGCGTGGAAAG
3701	GATCCATTGA	ACGACTCCGT	GGTTCCTGAC	GGATACCATG	AATACATCCG
3751	ACTTTTGCTC	AAGCAGCAAC	TTCTGACCGC	TAGACTCTAA	

Sequenz des *Arabidopsis thaliana* cDNA-Klons 4442 (EST 5)

1	ACTTCCGATC	ATCGACCTTT	CTTCGCCGGG	AAAAATTTCC	ACTGCTAAAT
51	TGATTTCGTC	GGCATGTCTT	GAACATGGAT	TCTTCTATGT	AAAGAATCAT
101	GGAATCTCGG	AAGAAATTAAT	GGAAGGGGTT	TTTAAGGAGA	GCAAAGGCTT
151	CTTTAATCTC	CCATTTAGAGG	AAAAGATGGC	TTTACTCCGC	CGTGATTTGC
201	TCGGTTATAC	TCCGTTGTAT	GCTGAGAAAC	TTGACCCGTC	CTTGAGCTCC
251	ATAGGTGATT	CCAAGGAAAG	CTTCTATTTT	GGCTCTTTGG	AAGGTGTTCT
301	AGCTCAACGT	TACCCAAATC	AATGGCCCTT	TGAAGGTATC	TTGCCATCAT
351	GGAGGCAAAC	CATGGAAACT	TACTACAAGA	ATGTCCGTGC	CGTGGGTAGA
401	AAATTGCTTG	GCTTGATTGC	CTTAGCATTG	GATTTAGATG	AGGACTCTTT
451	CGAAAAAGTT	GGAGCCTTGA	ATGATCCTAC	AGCAGTTGTT	CGCCTCTTAC
501	GCTATCCAGG	TGAAGTGATT	TCGTCAGATG	TTGAAACGTA	TGGTGCCTCA
551	GCTCACTCAG	ATTATGGAAT	GGTCACTCTT	CTTTTACTG	ATGGAGTTCC
601	AGGACTTCAG	GTTTGTAGAG	ACAAATCGAA	ACAACCACAC	ATTTGGGAAG
651	ATGTCCCTGG	AATTAAAGGG	GCGTTTATCG	TCAACATTGG	TGATATGATG
701	GAGAGATGGA	CCAATGGATT	GTTCGGTCT	ACATTGCATA	GAGTGATGCC
751	GGTGGGAAAA	GAACGTTACT	CGGTGGTGT	CTTCTTAGAT	CCCAATCCAG
801	ATTGTAATGT	GAAATGCTTG	GAGAGTTGTT	GCAGCGAAAC	TTGTCCCCCT
851	AGATTCCTCGC	CTATCCTCGC	CGGTGACTAT	ATTAAGGAGC	GTTTCAGGCT
901	AACATACGCC	ACTTCCTAGT	CTATGTTTGT	TACGAAATGT	AAGAAAATAA
951	TAATGTTCAA	ACTTCTACGT	ACCATAATGT	TTAATTAGTG	GAAACTTATG
1001	GAACTTGAAA	AAAAAAAAAA	AAAAAAAAAA	AAA	

Sequenz des abgeleiteten Gens 5 (*Arabidopsis thaliana*)

1	ATGGAAGGGG	TTTTTAAGGA	GAGCAAAGGC	TTCTTTAATC	TCCCATTAGA
51	GGAAAAGATG	GCTTTACTCC	GCCGTGATTT	GCTCGGTTAT	ACTCCGTTGT
101	ATGCTGAGAA	ACTTGACCCG	TCCTTGAGCT	CCATAGGTTT	GCATTTTTTG
151	CATATGTGGC	TTGTTTTTTG	TTTGCTGAGC	TAGTCTATAA	AGTTGATATT

201 TCCTTGTTAG **GTGATTCCAA GGAAAGCTTC TATTTTGGCT CTTTGAAGG**
 251 **TGTTCTAGCT CAACGTTACC CTAACCAATG GCCTTCTGAA GGTAATTTCT**
 301 TCCTAACTGT TTCATTAATC TGTTTTAATT TCAATCTCTC ACGGTCACGG
 351 ATTATGTGTA TGTAAAGGT **ATCTTGCCAT CATGGAGGCA AACCATGGAA**
 401 **ACTTACTACA AGAATGTCCT** GTAAGCAGTT GCTCTCTTGT CGTTCCTTCT
 451 CAAAGTGAAT CTCAAAATAA AACTTCATAC CTATATCTC AGTTGAGAAT
 501 AGGTTTCAAG TTCATTTCAA CAGTTAGATT TAAACCTTGA TGGTAGGAAT
 551 AAATATAGTC TATTTTAGCC AAGTACATCG GTAGTACAAA TGAAAAATTC
 601 TTTTCAGGTC ATAAATTGTA AGTAGATTAG GCTGTTTCAT GAGTATTTCC
 651 TCTCTTTTAT TTGCAATTTT CATATTACTT TCCTGGATCA ACTTGCTAGT
 701 TGTGTGTTTG CATTCCTGCT TTAAGGTCCG **TGGGTAGAAA ATTGCTTGGC**
 751 **TTGATTGCCT TAGCATGGAA TTTAGATGAG GACTTCTTCG AAAAAATTGG**
 801 **AGCCTTGAAT GATCCTACAG CAGTTGTTCCG CCTCTTACGA TATCCAGGTG**
 851 TGCTATATGTA CATCAAAAAG TTCCTGGTTGG CTTGATATTA ATCTTAATGA
 901 ACTTTGGTCA GAGTTAGATC GTAACATATG TAATCAGTGG AAACCTAAGC
 951 CACATTCGGG GAGTTTGAGT CAGGTTTTGA TCTCTGCACT TGTATGATT
 1001 ATAGGTGAAG **TGATTTGCTC AGATGTTGAA ACGTATGGTG CCTCAGCTCA**
 1051 **CTCAGATTAT GGAATGGTCA CTCCTTCTTT GACTGATGGA GTTCCAGGAC**
 1101 **TTCAGATCAA TATCTCTGAA TTTCTCGTCA TAATCTTCTA TTAGTAACCT**
 1151 TATGTTCTTA CACTGTCTTT GAACGAACTT CTTAGGTTTG **TAGAGACAAA**
 1201 **TCGAAACAAC CACACATTTG GGAAGATGTC CCTGGAATTA AAGGTAAGT**
 1251 CTAGCTTTTC TTAGCTACTA GTTTATGATG AGAAATGATT GTTGAAATCC
 1301 GGGGACTTTC TGTAACCGCA **GGGCGTTTAT AGTCAACATT GGTGATATGA**
 1351 **TGGAGAGATG GACCAATGGA TTGTTTCGGT AAACCTAAAC CAGTTACCT**
 1401 AGTCGTGAGT GCGGTTTTTA TGCCTTCTTC ACATTCGTA TTTTCTTGA
 1451 ATGGGTTTTAG **GTCTACATTG CACAGAGTGA TGCCGGTGGG AAAAGAACGT**
 1501 **TACTCGGTAG CTTTGTCTTG TTTCTTTCTT TCTAATTTCT ATACAAATTG**
 1551 ATAAAGTTTT GAGTTATTTA AAGGAAAAGT TTTTGTAGG **TGGTGTCTTT**
 1601 **CTTAGATCCC AATCCAGATT GTAATGTGAA ATGCTTGGAG AGTTGTTGCA**
 1651 **GCGAACTTTG TCCCCAAGG TATTTCTTCT TGAATCTTAA TCCAATTTAA**
 1701 GGGTGTGAGG AGAGGCACGT TACACGTTTG TTAAATCAT TGGTCTTCAA
 1751 GTGTAAAAGA ATAAAAGTTT AGTCCCTTCT TCATAAAGAT TTAGTTCCAT
 1801 CGGTGCAGAT **TCCCGCTTAT TCTCGCCGGC GACTATATCA AGGAGCGTTT**
 1851 **CAGGCTAACA TACGCCACTT CCTAG**

Sequenz des abgeleiteten Gens 5 (*Capsella rubella*)

1 **ATGGAAGGG TTTTAAAGGA GAGCAAAGGC TTCTTTAATC TTCCACTAGA**
 51 **GGAGAAGATG GCTTTACTCC GCCGTGATTT GCTTGGTTAT ACTCCTTTGT**
 101 **TTGCTGAGAA GCTTGACCCG TCTTTGACCT CCTCCACAGG TTTTATTCAT**
 151 TTGGTGGTTT GTTTTCTTGC TGGAGGTAGT CTACAGATTT GATCTCTCTT
 201 TGTTAGGTGA **TTCCAAGGAA AGCTTTTATT TTGGCTCTTT GGAAGGCGTG**
 251 **CTTGCTCAAC GATACCCAAA TCAATGGCCT CCTGAAGGTT ATTTCCACT**
 301 ACATGGGTTT TTTCTAATCT GTTTTGATCT CAATCTCTTT TTTTCTCTCT
 351 CACGGATTAT GTGTATGTAT TAGATCTCTT **GCCATCATGG AGACAAACCA**
 401 **TGGAATGTTA CTACAAGAA GTCCCTGTAAG TACCTGCTCT TTAGTCGTTT**
 451 TTCTGCTAAG TTAATCCAAA GCTAAAACCT AATACCATGA TTTCTCTATA
 501 CATCCTCAGC TGATAATAGG TTTTCAAGTT CATATCAACA GTTCAAATTT
 551 GAACCTTGAT GGTAGGGACA AAAATAGTTT ATTGTAGCCA ATTACATTTT
 601 TAGTAGAAGT GGACAATCTT AGCTGCCATG GTGTTCTTGA AATATCTTCT
 651 TTTCAGGTTA TAAATTGTAA GGAGATTATG CTTATTTTCT GAGTAGTTTC
 701 TCATTTTGCA ACTTTCCTAT TACTTTCCTG GACCAACGTG CTAGTTGTGT
 751 TTTGCTTTTC TGCTTTAAGC **TCCGTCGGTA GAAAATTGCT CGGCTTGATT**
 801 **GCCTTGGCAT TGGATTTAGA TGAGGACTTC TTCGAAAAAG TTGGAGCCTT**
 851 **GAATGATCCT ACAGCAGTTG TTCGCCCTTT ACGCTATCCA GGTATCTATG**
 901 TACATTAGAT TCTGGTTGGA TTGATGATTG ATATTAATAT TGATGAACTT
 951 TGGAAGAAT TAGTATCTTA ACAGAAATAG AGGAACATGT ATCTAAACAT
 1001 ATGCAAGAAG TGGAATGTA AGCCATATTC CAGGAGTTCG AGTCAGGCAG
 1051 GTTTTGATCT CTGCACCTGT GATGATATAT ATATATATAT TATAGGTGAA
 1101 **GTGATTATGT CGAATGAAGA AACATATGGT GCCTCAGCTC ACTCAGACTA**
 1151 **TGGAATGGTC ACTCTTCTAA CGACTGATGG AGTTCCAGGG CTTCAGGTAA**
 1201 ATGATTTCTT ATTCCTTCTT GTAACCTCTC TCATACCTT CTCTTTTCTT

1251 TAACCTTATG TTCTTACACC GTCTGTGAAA CTCATAGGTT **TGTAGAGACA**
1301 **AATCGAAACA ACCACACATT TGGGAAGATG TCCCTGGAAT TAAAGGGTAA**
1351 AGTTAGCTTT TTTTAACTA CTGGTTTTG GAGGTTTGTC GAATAGAAAC
1401 AATTGTCGAA ATCGGGGAC TTTCCTATAA CTGTAGGGCG **TTTATCGTCA**
1451 **ACATTGGTGA TATGATGGAG AGATGGACCA ATGGACTGTT CCGGTAAATT**
1501 TTTAATCAAT CCTAGTTTAG AGTGAGTGTG TTTTCTAGGT AACACTTCAC
1551 ATTATTCATG TAATGGGGTT TAGGTCTACG **TTGCACAGAG TGATGCCAGT**
1601 **TGGAAAAGAA CGTTACTCGG TAACTTTGCC TCAATCTTTT CCATTTTAC**
1651 TAACAAATG ATAATGTTTT GGGGGATAAA GTAATTAATT CGAGGAGAAG
1701 TTACTGTAGG **TGGTGTCTT CTTAGATCCC AATCCAGACT GTAATGTGGA**
1751 **ATGCTTGGAG AGTTGTTGCA GCGGAACTTG TCCCCAAGG TATTGGAGAA**
1801 AACTCTTGTT GAATTACTCC AATTTAGGGT GTGAAGGGGT GCATATCTGT
1851 TGTTAAATCG TTTTGGACCT TGAATGGCAA AAGAATAAAA GTTAAGTCCG
1901 TTTCTGATAT ACAACTCTGC TTGTATCGGT GTAGATTCCC **GCCTATTCTC**
1951 **GCAGGCGACT ATATGAAGGA GCGTTTCAGG CTAACATACG CAACATCCTC**
2001 **TTAG**

Sequenz des *Arabidopsis thaliana* cDNA-Klons C

1 AACAAATGG **GAACAGCTCT GAAACTTCCG ATCATCGATC TTTCTTCGCC**
51 **GGAGAAGCTC TCTACTTCCC GATTGATTCG TCAGGCTTGC TTGGATCATG**
101 **GATTCTTCTA TCTTACGAAC CATGGCGTTT CGGAAGAGTT GATGGAAGGA**
151 **GTGTTAATAG AGAGCAAGAA ACTATTCTCC CTTCCCTCTG ATGAGAAGAT**
201 **GGTGATGGCT CGTCATGGTT TTCGTGGTTA CTCACCGTTA TACGACGAGA**
251 **AACTTGAATC ATCTTCCACC TCCATAGGTG ATTCCAAGGA GATGTTCACT**
301 **TTTGGATCTT CAGAAGGAGT TTTGGGTCAA CTTTACCCTA ACAAGTGGCC**
351 **TC TTGAAGAG CTTTGGCCGC TTTGGAGGCC AACCATGGAA TGCTACTATA**
401 **AAAATGCAT GGATGTTGGT AAGAAATTGT TTGGCCTAGT GGCTTTGGCA**
451 **TTGAATTTAG AGGAAACTA CTTTGAACAA GTGGGAGCTT TCAATGATCA**
501 **AGCAGCGGTT GTTCGTCTCT TACGTTATTC AGGAGAATCA AATTCATCTG**
551 **GAGAAGAAAC ATGTGGTGCC TCTGCTCATT CAGATTTTGG AATGATAACT**
601 **CTTCTTGCAA CTGATGGAGT TGCAGGGCTT CAAGTTTGTA GAGATAAAGA**
651 **TAAAGAACCA AAAGTTTGGG AAGATGTCGC TGGTATTAAA GGGACTTTTG**
701 **TTGTCAACAT CGGCGACCTT ATGGAAAGAT GGACTAATGG TCTTTTCCGA**
751 **TCAACATTGC ATAGAGTGGT GTCAGTGGGA AAGGAACGTT TTTCCGTGGC**
801 **TGTATTTGTC GATCCAGACC CAAACTGCGT TG TAGAGTGC TTGGAGAGCT**
851 **GTTGTAGTGA AACATCTCCT CCCAAGTTT CACCGGTCCG GGCCAGAGAT**
901 **TATTTCCACG AGCGATTGAG TCAAACCTC GCCTCGTACT CCGGTTCCGA**
951 **CTAAAATATA TGAGAACTGA GAAGCTCCAG GTTTCTTAA ATAAGACATG**
1001 **TGTCAAATGT TTGAGTTTCA CTTTTCGCTG GACATGAAAT AGCTATTATA**
1051 **AACCAAAAGG TGATAAAACA CTTAAACAG TATAATGTAA TTAGGTGTAT**
1101 **TTTCGATTGA TAAATGTATT AATTGTAGAC TAAAAAATAA AAAAAAATAA**
1151 **AAAAA**

Sequenz des abgeleiteten Gens C (*Arabidopsis thaliana*)

1 **ATGGGAACAG CTCTGAACT TCCGATCATC GATCTTCTT CGCCGGAGAA**
51 **GCTCTCTACT TCCCGATTGA TTCGTCAGGT TCTTCTTCTT CTCCCCCTTG**
101 **TCTTTTTTTTT CATGGTAACA AAACGAAGAT ATTGATTGTA TTATTTATAC**
151 **TCTAAACAGT AAATGGTCT TATATTACTG TAGAAATGTC GGTTGTGCGAT**
201 **AATAGTGT TT TAGAATGATA TTGTGATGCT TAAATATCCA ATTGTTTAGG**
251 **CTTGCTTGGG TCATGGATTC TTCTATCTTA CGAACCATGG CGTTTCGGAA**
301 **GAGTTGATGG AAGGAGTGT AATAGAGAG AAGAACTAT TCTCCCTTCC**
351 **CTTTGATGAG AAGATGGTGA TGGCTCGTCA TGGTTTTCGT GGTTACTCAC**
401 **CGTTATACGA CGAGAACTT GAATCATCTT CCACCTCCAT AGGTTCTTGA**
451 **TAGTATAATA GCTCATGAAA TTGCTTTGAT TTGTCCAATG ACTGGATTCA**
501 **TGTGTTTCAA TTATTTGCGT TTTGGTTCCA AAAAGTGGAA TGTGTCATTT**
551 **GAGTTTTCAT ACAACAGGT GATTCCAAGG AGATGTTTAC TTTTGGATCT**
601 **TCAGAAGGAG TTTTGGGTCA ACTTTACCCT AACCAAGTGG CTCTTGAAGG**

651	TAAACTCTTT	GCCAACCTCTA	CTATGTGGAC	TAGATGAGAG	CAAACAGGAA
701	TTCTGTGATC	ATCTAGTCAG	TAAATCATAA	TTTTTAAAAG	ATAAAGAACT
751	TACTCATGTT	AAATTAACAA	CAATTTTTTTA	AAAAATTTTAA	ATAAATTATG
801	AGCTATCTTG	AATCCAAAAC	ATTTTAAATTT	CAAAACTAAC	ATTAGAAAAAT
851	TCTAACGACT	TCCTCGAAAGT	TATGAACCTAT	AACACTGAAA	AAACAAAAAT
901	ATAAGAAAATA	AAATAAAAAA	CTATATATTTT	AGATGATAAG	TTTATTAAAG
951	CTATAAAATTT	TAAACCGTAT	TTCCCCATAA	CTAAAACCAA	ACTCGACATA
1001	TAAGAAAAAAA	TTGAAGAGCA	TATAAAATTTA	TCACAAAATC	TTAATGAAGT
1051	TGAAGAAAAT	AGTGAGATAT	ATGAAAATGG	TGATAAACCC	TATTTAGCAA
1101	ATAGTAACAA	AACCATTTAG	GCCTCTAAAC	CCATTGTTC	TTTCAAATTA
1151	TAGTCTTTTC	TTTTGTTTAT	GTTGATTATA	TATATATAAC	TTATGTGTTA
1201	TGGATGTTTA	CAGAGCTTTT	GCCGCTTTGG	AGGCCAACCA	TGGAATGCTA
1251	CTATAAAAAAT	GTCATGTAAG	TAGCAGGAAA	TTCTGCTTAT	GCATACCCA
1301	TGCTTAAGTC	TAAATGGTCA	CTCTAGCACC	TTAGATACTT	CTAAGGCTAA
1351	AGATGTCATT	GTTGCTTCTT	AATAGGGATG	TTGGTAAGAA	ATTGTTTGGC
1401	CTAGTGGCTT	TGGCATTGAA	TTTAGAGGAA	AACTACTTTG	AACAAGTGGG
1451	AGCTTTCAAT	GATCAAGCAG	CGGTTGTTCG	TCTCTTACGT	TATTCAGGTC
1501	TTTTACCAA	CCTTGTAACA	TTTTGGGGAT	TGATTAGAGA	CATTTTAGCA
1551	ACTGTCTTAA	AACTGTTCTA	AAAGAAAAAC	CCACTGCTAG	TTTGTTTACT
1601	TTACTCTTTG	TCACATTATT	AAAGGAGAAT	CAAATTCATC	TGGAGAAGAA
1651	ACATGTGGTG	CCTCTGCTCA	TTCAGATTTT	GGAAATGATA	CTCTTCTTGC
1701	AACTGATGGA	GTTGCAGGGC	TTCAAGTATT	CATCTCATAA	TCTTTTTTAA
1751	ACTGAGAATG	TCTTCTTTTG	TTTTCTTCTA	CTTCTTATAA	CAATGTTTAC
1801	GTATACTAAA	AAGGTTTGTG	GAGATAAAGA	TAAAGAACCA	AAAGTTTGGG
1851	AAGATGTCGC	TGGTATTAAA	GGGTAAGTTA	TATACTTATA	TAAATGCTTA
1901	AACATATGCT	TTTTGATTGA	CACATTTTGT	AATTAAGGCC	TAAAACCTGT
1951	GAGAAATTTT	GTGTAAGGAC	TTTTGTTGTC	AACATCGGCG	ACCTTATGGA
2001	AAGATGGACT	AATGGTCTTT	TCCGGTACCG	AAAAAAATAT	TCCTTACTTT
2051	TAGTTTACAT	TTCATATATA	AGGTGATATT	TTGATTCAAG	ATCGCTTCTT
2101	GATTTAAGAT	CAACATTGCA	TAGAGTGGTG	TCAGTGGGAA	AGGAACGTTT
2151	TTCCGTACGG	TACAATTTGT	TTGAAATAGT	CTCTGTTTTA	CTAATCCTCT
2201	TTATTATCGT	GACATTAACT	TATTGGTTTG	ATTACTTTGA	AGGTGGCTGT
2251	ATTTGTGCGAT	CCAGACCCAA	ACTGCGTTGT	AGAGTGCTTG	GAGAGCTGTT
2301	GTAGTGAAC	ATCTCCTCCC	AAGTATGTAA	TAACCTGTTA	TTTTGGTTGT
2351	CTCATGTTAT	AAACTTTTGA	GTTTGACTGT	TTTCATCCAT	CATGCTGGGG
2401	GTGATTTGTG	ATAACTAGAA	AGTTTTTTGTC	TTTTATTGGA	AAAACAAAAA
2451	AAGATGATTT	GTCAATTTCT	TGCCTCAGGT	TTCCACCGGT	CCGGGCCAGA
2501	GATTATTTCC	ACGAGCGATT	CAGTCAAACC	CTCGCCTCGT	ACTCCGGTTC
2551	CGACTAA				

Sequenz des abgeleiteten Gens C1 (*Capsella rubella*)

1	ATGGGAACGA	TTCTGAAACT	TCCAATCATC	GATCTTTCCT	CGCCGGAGAA
51	ACTCTCTACT	TCCAGTTTGA	TTCCGTCAGGT	TCTTCTTCTT	CTTCTCCCTC
101	TAAGATAACA	AAACGACATT	ATTGATTGTA	TTATTAGTTA	ATACTCTTAA
151	AAAAGACAGT	AAATGCTTCT	TTATATGGTT	ACTGTGTCCG	TGGTAAAAAT
201	GTTTTTTTTT	TTTTGAAGGC	TTGTTTGGAT	CATGGATTCT	TCTACTTAAC
251	GAATCATGGC	GTTTCGGAAG	AGTTGATGGA	AGGAGTGTTA	ATGGAGAGCA
301	AGAAACTGTT	CTTCTTCCCT	CTTGACGAGA	AGATGGTGAT	GGCTCGTCTG
351	GGTTTTCGTG	GTTACTCACC	TTTATACGAG	GAGAAACCCC	ACTCCTCTAC
401	CTTAGCCAAA	GGTTCCTAAT	TATATTATAG	CTTATCATAG	TTAGCTTATA
451	CTAACTTTTA	TCATTGGTTT	TGGTTAAGAA	ATGGAAATGT	GTGATATATT
501	GATTTGAGAT	TTTTTTTTTTT	TTCTTTTTCTT	ATAAAAGGTG	ATTCCAAGGA
551	GATGTTTACA	TTTGGATCTT	CAGAAGGAGT	TTTGGCACAA	ATTTACCCTA
601	ATGAGTGGCC	TCGTGAAGGT	AACATTTCTT	GGATAATACT	TTATAATCGA
651	TATTATACTA	TTGAAACTTT	TTATTTAATA	TATGATGATA	GTGAGCAAGT
701	TGATATAGTG	AAACTCCTAT	TAGTCTCAA	TGATTGTTC	TCTTTGATAT
751	AAAACTCTT	TACCCACTCT	TGATCGAGAC	GTGTATATAA	AATTATAAAT
801	AGAAAAGTAG	AATTAGCTAC	AAACAAATTA	TTATATTTTT	AAGACGAATC
851	GGGAAAGCAA	ACCGGATTTT	CCGGACAATC	AGTTCAATAT	TGTATTATAG
901	TTTAATATAG	TTTTTGATAA	TTAATGTTAT	CATAAGAGTT	TGATAACAAT
951	TTTTCTTCTT	CTTCATTTTT	TGGCCTAGAT	CTTAATAACA	AAAAAAAAAA

1001 AAAAAAAAAAT CTACACCTTG TAATCAGTGT TCTAACAGAG TTCTAAACGG
 1051 TCTTCTAGCT AGGTGCTCAT TATCGAAGTA TTTGTTTCTT GAATTCACAA
 1101 TATTTTTTAAA TGCAAACTA ATATTGAGGT TTGGAAAATTT CCTAATCATA
 1151 TTGACCCGGA GAACATAGAA AGAAAAATA GAACATAATGA AATGTTTATT
 1201 TCATCTACCA ATTTAACGGA ATTTCAAGAA ATTTGTTGGCC TTTGAACCCC
 1251 TTGTTTCTAT CAAATAGTAA TGCTTTATTT TGTAATTGTT TATGATAATT
 1301 ATCTATATCT TATGGATTAT GAATGTTTGT AGAGCTTTTG **CCGCTGTGGA**
 1351 **GGCCAACCAT GGAATGTTAC TATAAACTG TAATGTAAGC** AGATGGAAAC
 1401 TTGTTCTGCT TATGCTGACC AAAATACTTG TAAAACTAAT TAATTAATGA
 1451 TGTC AACGTT GTTTCCTTTT CTATATATAT TGGATTGTTG TATGTTGTGG
 1501 TTTTGTGTTA ATGGTCTTTT GTTGCTTCTT AATTAGGGAT **GTTGGTAAAA**
 1551 **AATTGCTTGG CCTAGTGGCT TTGGCACTGA ATTTAGAAGA** **AAACTTCTTT**
 1601 **GAACAAGTGG GAGGCTTCAA TGCTCAAGCA GCAGTTGTTT** **GTCTTTTACG**
 1651 **TTATCCAGGT CTTATAACAT TTTGGGGATT** GATATGAAAA CTCTACTACT
 1701 GCTAGATTTA TTTACTTTTA TTCTTATTAA AAGTCTCTCT TTCTCACATT
 1751 ATATATAAAT AGGAGAGTTG **AATTCATCTG GAGAAGTAA** **ATGTGGAGCG**
 1801 **CTGCTCATT CAGATTATGG AATGATAACT CTTCTTGCAA** **CTGATGGAGT**
 1851 **TCCAGGGCTT CAAGTATTCA TCTCATAATC** TTTAACTTAA AACTCAGTTT
 1901 TAAAGTAAAA TAAAACAAAT ATTTTGTAT ACCTTTAACA ATGTTTACGT
 1951 ACTAAAAGGT **ATGTAGGGAT AAAGATAAAG** **AACCAAAGT TTGGGAAGAT**
 2001 **GTCGCTGGAG TCAAAGGGTA** AGTTACACTA TCATGTACTT TATTATTAAG
 2051 GCCTAAAAAC TTAAGATATG ATATGACTTG GTGTAATAA AAACCGCAGG
 2101 **GCTTTTGTG TCAACATTGG CGACCTAATG GAGAGATGGA** **CTAATGGTCT**
 2151 **TTTCAGGTAC CTACAAAAAC TAAAACATTC** ATACTTTTTA GTTTACAGTT
 2201 TCATTTCTAA ACTAAAAACAT TCATACTTTG AGTTAATTAA GATATATCTC
 2251 TTCTTGATTA TAATTAAGAT **CAACAATGCA TAGAGTGGTG** **TCCGGTGGGAA**
 2301 **AAGAACGTTA TTCGGTACGT** ACGATGTAAT TCGTTTGAAT AAGTCTCTGT
 2351 TTTTGCCTAAT CATCATCATT AAACGTGAAC TAAACAAGTT GGTTTTGATT
 2401 ACTTGAAGGT **GGCTGTATTC GTCGATCCAG** **ACCCAAACTG CGTTGTGGAG**
 2451 **TGCTTGGAGA GCTGTGTAG TGAACATCT CCTCCAGGT** ATATTAAGAC
 2501 AATAATGCAT AGAGTGTGT TGTGATAAC TTAAAAAAT TTGGCTTTCT
 2551 ATTGAAATCA CTTAAATTGG TTGATTTGTC ATTTTGTTC CTCAGGTTTC
 2601 **CACCTGTCCG GACCAAAGAT TATTTCCACG** **AGCGATTCAG TCAAACCTC**
 2651 **GCCTCGTACT CCGTTCGGG CTAA**

Sequenz des abgeleiteten Gens C2 (*Capsella rubella*)

1 **ATGGGAACGG TTCTGAACT TCCGATCATC GATCTTTCTT** **CGCCGGAGAA**
 51 **AATCTCTACT TCCCGATTGA TTCGTCAGGT** TCTTCTCCTT CTCTTCTTC
 101 TTCTTCTTCG GAAACAAAAC GACCGACTA TATTGGTTGT ATTATATCTA
 151 GCTTGAAAAA ATAGTAAATG CTTTCTTTAT ATGGTFACTA TGCAAACGAC
 201 GGTAGTAAAT TGATATTTTG ATGATTGTTT TTTTATTTTT ATTGAGGCTT
 251 **GTTTGGATCA TGGATTCTT TACCTCACGA ATCATGGAGT** **CTCGGAAGAG**
 301 **TTGATGGAAG GAGTGTAAAT GGAGAGCAAG** **AAACTGTTCT CTCTTCTCT**
 351 **TGACGAGAAG ATGGTGATGG CTCGTCGTGG** **TTTTCGTGGT TACTCACCTT**
 401 **TATACGACGA AAAGCTTGAC CCCTCTGCCA CATCCAAAGG** TTCGTAATTA
 451 CATTATGTTA TAGCTTATCC TAATTTTATT ATACTAACTC TTATCATTGG
 501 CTTTGGTTGA GAAAAAGGAA ATGTGTGACA TACAGTATAT TCATTTTACA
 551 TTTTTTTCTT TTCTTATAAA AGGTGATTCC **AAGGAGATTT** **TCACATTTGG**
 601 **ATCATCAGAA GGAGTTTTGG CACAAATTTA** **CCCTAATGAA** **TGGCCTCTTG**
 651 **AAGTAAACAT TTGTTGGATC** ATAATATGTA CTTTATAATG GATATAATAT
 701 AGAAACTTTA TACTTAAATA GATGATAATA GTGAGGAAGT TAATACTAAT
 751 TTAAGACAT TGTAAAAAAT GGGGATCTAA ATGT'TTTTAG GGAGACAGAA
 801 TTCTCTTTTC AAAGATATCC ATTTTTTTTT TAAATATCAT ATATTGAAAG
 851 AGTTTTATTT AAAGACACCC ATAAAAATTA CCATTGTGGA TGCTCATTTT
 901 TTTAATAGCT ATTTTAATAG TTTGATAGTT ATTTTAAAGA TTTATAGTTA
 951 ATAGTTATTT AAATAGTTTG ATAATATGTT TTCTTCTTCC TCATTTTTTT
 1001 TGGCCTTAGT AACAAAAAAT AAATGCATAC CTCTACTAA TAGTTATGAG
 1051 TTATATGAAG TATCTTTTTT CTATAACAA TTATGAACTT ATTGAATCCA
 1101 AAATATTTTG AAATGAAAAA CTAACATTGA GGTGTGAAAA AATCTTAAAC
 1151 ATATTGAATT ATAGAACGTT AAAAAAATTT GATGGAATAT TGTTTGTTTT
 1201 AGCTACCAAT TTTAACCAAA TTGCAACAAA TAAAAATCAA CATTCCAACC

1251	AAAAGAAAA	CATTACAATT	AGGTCTGGCT	TTTGAACCC	TTAATTATTT
1301	ATATTGTAAT	TGTTTTGAAT	GTTTGTAGAG	CTTTTGCCGC	TTTGGAGGCC
1351	AACCATGGAA	TGCTACTATA	AATCTGTCAT	GTAAGTGGCT	GGAACTTGT
1401	TCTGCTTATG	CTGACCAAAA	AAACTACTTTT	AAAAATTTAT	AATGGAGCAA
1451	TTTTTGTTTT	CCTTTTCATA	TATTTAATTG	GATTGTTGTA	AGCATGTTGT
1501	GGTTTTTGT	GCTTCTTAAT	TAGGGATGTT	GGTAAGAAAT	TGCTTGGCCT
1551	AGTGGCTTTG	GCATTGGATT	TAGAGGAAAA	ATTTTTTGAA	CAAGTGGAAG
1601	GCTTCAATGT	TCAAGCAGCA	GTTGTTTCGTC	TTTTACGCTA	TCCAGGTCTT
1651	TCAACATACC	TTGTATAGTC	TTATAACATT	TTGGGGATTG	ATAAGAATGC
1701	TCTAGTACTG	CTAGAGTTCT	TTACTTTTAT	TCTTAATTAA	AGTCCCTCTT
1751	TATCACATAG	GAGAGTTGAA	TTCATCTGGA	GAAGTAACAT	GTGGTGCCTC
1801	TGCTCATTCA	GATTATGGAA	TGCTAACTCT	TCTGGCAACA	GATGGAGTTC
1851	CAGGGCTTCA	AGTATTCATC	TCATATATTA	TATATTTTAA	TTCAGCTTTT
1901	AAAATTAAT	AAAACAAAAT	AGTTGCTATA	CGTTTAACAA	TGTTTACGTA
1951	CTAAAAGGTT	TGTAGGGATA	AAGATAAAGA	ACCAAGAGTT	TGGGAAGATG
2001	TCGCTGGAAT	CAAAGGTAA	GTTACATAAA	ACCATAAACA	TATATGCTTT
2051	TTAATTGACA	CTATTTGTAG	TTAAGGCCA	AAACTCTGA	GATATATATA
2101	TGATTGACAT	TTAAAAATCAC	AGGGCTTTTC	TTGTTAACAT	TGGCGACCTT
2151	ATGGAGAGAT	GGACTAATGG	CCTTTTCCGG	TACCTACAAA	AACTAAAAA
2201	TCTATATATA	TACTTTTTAG	TTTACATTTT	CATTTCTTTT	ATAACACTAC
2251	AATTTTCAAT	CTTTTGAAAC	ACTACAGTTT	CATTTTCTTA	AGGTAATATT
2301	TAAACTTAAG	ATTTTGCATG	TCTCTTCTTG	ATTATAATCA	AGATCAACAA
2351	TGCATAGAGT	GGTGTCCGGT	GGGAAAGAAC	GTTATTCCGGT	ACGTACGATG
2401	TAATTCGTT	GAAGTAGCTA	GTCCTGTTT	TTGCTAATCA	CCATCACATA
2451	AACGTGACAT	AAACAAGTTG	GTTCTGATTA	CTTGAAGGTG	GCTGTATTCT
2501	TCGATCCAGA	CCCAAACAGC	GTTGTGGAGT	GCTTGGAGAG	CTGTTGTAGT
2551	GAAACATCTC	CTCCCCGGTG	TGTGTGTGTA	TATTAATAAA	TCTTGTGTT
2601	TTGTGATAAC	TTAAAAAGTT	TTGGTGTCT	ATCGTAAGAA	CTTAAATTA
2651	ATGGTTGATT	TGTCATTTTG	TTGCATCAGG	TTCCACCCTA	TCCGGACCAG
2701	AGATTATTTT	CACGAGCGAT	TCAGTCAAAC	CTTCGCCTCC	TATCCGGTT
2751	CCGGCTAA				

Sequenz des *Arabidopsis thaliana* cDNA-Klons D

1	CAAAGAAAA	GCAGACAAC	TTATTTGCAA	AACAGAGTTT	TTTTTTCTTT
51	TCTTGAGAAA	GTTC AACAGA	AGATGATGTT	GGAGAAAGAC	GATCTGGGTC
101	TAAGCTTAGG	CTTGAATTTT	CCAAAGAAAC	AGATCAATCT	CAAATCAAT
151	CCATCTGTTT	CTTTACTCC	TTCTTCTTCT	TCTTTTGGAT	TATTCAGAAG
201	ATCTTCATGG	AACGAGGTT	TTACTTCTTC	AGTTCCAAAC	TCAGATTCTG
251	CACAAAAAGA	AACAAGAAC	TTCATCCGAG	GAATCGACGT	GAACAGACCA
301	CCGTCTACAG	CGGAATACGG	CGACGAAGAC	GCTGGAGTAT	CTTCACCTAA
351	CAGTACAGTC	TCAAGCTCTA	CAGGGAAAAG	AAGCGAGAGA	GAAGAAGACA
401	CAGATCCACA	AGGCTCAAGA	GGAATCAGTG	ACGATGAAGA	TGGTGATAAC
451	TCCAGGAAAA	AGCTTAGACT	TTCCAAAGAT	CAATCTGCTA	TTCTTGAAGA
501	GACCTTCAAA	GATCACAGTA	CTCTCAATCC	GAAGCAGAAG	CAAGCATTTG
551	CTAAACAATT	AGGGTTAAGA	GCAAGACAAG	TGGAAGTTTG	GTTTCAGAAC
601	AGACGAGCAA	GAACAAAGCT	GAAGCAAACG	GAGGTAGACT	GTGAGTTCTT
651	ACGGAGATGC	TGCGAGAATC	TAACGGAAGA	AAACCGTCGG	CTACAAAAAG
701	AAGTAACGGA	GTTGAGAGCA	CTTAAGCTCT	CTCCTCAGTT	CTACATGCAC
751	ATGAGCCCAG	CCACTACTTT	GACCATGTG	CCTTCATGTG	AACACGTGTC
801	GGTCCCAGCA	CCACAACCTC	AGGCTGCTAC	GTCAGCACAC	CACCGGACGT
851	TGCCGGTCAA	TGCGTGGGCT	CCTGCGACGA	GGATATCTCA	CGGCTTGACT
901	TTTGACGCTC	TTCGTCTAG	GTCCTAAGTC	TTTTTACTTA	CAATCAAAGG
951	GCATTGTGGT	CGTTTTATTA	AGTTTCAGGG	ACCAGATATG	CATGTAGTTG
1001	TTAACATGTA	TGTATTTTTT	TTAGAAAAGAA	AGAAAAACAG	ATTAATAAAA
1051	AAAAAAAAAA	AAA			

Sequenz der Athb-2 cDNA (X68145)

1	TGCCCCCAGC	TAGTCACATA	CTCATGATTG	CAAAATCTCT	CTCTCTCTCT
---	------------	------------	------------	------------	------------

51	GCCTCTCTAT	ATATTAACCT	TTCTTCTTCC	TTTACTTTCT	CATCTTCTAT
101	CTCTCAAAAAG	AAAAGCAGAC	AACTTTATTT	GCAAAAACAG	AGTTTTTTTT
151	TCTTATCTTG	AGAAAGTTCA	ACAGAAGATG	ATGTTTCGAGA	AAGACGATCT
201	GGGTCTAAGC	TTAGGCTTGA	ATTTTCCAAA	GAAACAGATC	AATCTCAAAT
251	CAAATCCATC	TGTTTCTGTT	ACTCCTTCTT	CTTCTTCTTT	TGGATTATTC
301	AGAAGATCTT	CATGGAACGA	GAGTTTTACT	TCTTCAGTTC	CAAACCTCAGA
351	TTCGTCACAA	AAAGAAACAA	GAACTTTCAT	CCGAGGAATC	GACGTGAACA
401	GACCACCGTC	TACAGCGGAA	TACGGCGACG	AAGACGCTGG	AGTATCTTCA
451	CCTAACAGTA	CAGTCTCAAG	CTCTACAGGG	AAAAGAAAGC	AGAGAGAAGA
501	AGACACAGAT	CCACAAGGCT	CAAGAGGAAT	CAGTGACGAT	GAAGATGGTG
551	ATAACTCCAG	GAAAAAGCTT	AGACTTTCCA	AAGATCAATC	TGCTATTCTT
601	GAAGAGACCT	TCAAAGATCA	CAGTACTCTC	AATCCGAAGC	AGAAGCAAGC
651	ATTGGCTAAA	CAATTAGGGT	TACGAGCAAG	ACAAGTGGAA	GTTTGGTTTC
701	AGAACAGACG	AGCAAGAACA	AAGCTGAAGC	AAACGGAGGT	AGACTGCGAG
751	TTCTTACGGA	GATGCTGCGA	GAATCTAACG	GAAAGAAACC	GTCGGCTACA
801	AAAAGAAGTA	ACGGAATTGA	GAGCACTTAA	GCTCTCTCCT	CAGTTCTACA
851	TGCACATGAG	CCCACCCACT	ACTTTGACCA	TGTGCCCTTC	ATGTGAACAC
901	GTGTCGGTCC	CGCCACCACA	ACCTCAGGCT	GCTACGTCAG	CGCACCACCG
951	TCGTTTGCCG	GTCAATGCGT	GGGCTCCTGC	GACGAGGATA	TCTCACGGCT
1001	TGACTTTTTGA	CGCTCTTCGT	CCTAGGTCCCT	AAGTCTTTTTT	ACTTGCAACC
1051	AAAGGGCATT	TTGGTCGTTT	TTTAAAGTTTC	ATGGACCAGA	TATGCATGTA
1101	GTTGTTAACA	TGTATGTATT	TTCTTAAAAA	AAAAAAAAAA	

Sequenz des abgeleiteten Gens D (*Capsella rubella*)

1	ATGATGTTGG	AGAAGGACGA	CTTGGGTCTA	AGCTTAGGCT	TGAACTACCC
51	TAAGAAACAG	ATGAATCTCA	ATTCAAATCC	CTCTGTTTCT	GTTACTCCTT
101	CTTCTTCTTC	TTTTGGATTA	CTCCGTAGAT	CTTCATTGAA	CGAGAGTTTT
151	ACTTCTTCAG	GTATTTTCAT	ATCTTCTTGG	GTTCTCTGT	TTTTATTATT
201	TTCCCGAGAA	AAAATCAATT	TTTCCAGGAA	AAACTCAGCA	AAAACCTTAA
251	AAGTCTCACT	TTTCTTAACA	TATTCATAAA	ACCCATCTTT	AATCTGTTGT
301	AAAAACCATT	GCTTTTTGAC	AATTTTCTCC	ACTTTCCGGA	GAAAAATAATA
351	TCCATTTTCC	CGGAAAATCA	CATCATATAA	CCTTAAAAGT	CTAAATCTTA
401	TGCATGCAGT	TCCAAACTCA	GATTCATCTC	AAAAAGAAAC	AAGAACTTTC
451	ATCCGTGGAA	TCGACGTGAA	CAGACCACCG	TCGACGGCGG	AATACGGCGA
501	AGAAGAGCC	GGAGTATCTT	CACCGAACAG	CACAGTCTCA	AGCTCAACAG
551	GGAAAAGAAG	CGAGAGAGAA	GAAGACACAG	ATCCACAAGG	CTCAAGAGGA
601	GGAATCAGCG	ACGATGAAGA	CGGTGATAAC	TCCAGGAAAA	AACTCAGACT
651	CTCCAAAGAT	CAATCTGCTA	TTCTCGAAGA	AACCTTCAAA	GATCACAGTA
701	CTCTCAATCC	GGTAAATAAT	TTTTACTCTT	CTTTCAAATG	GACCAATGAT
751	GATGGAAACG	AAAATAACGC	GTAAATGTTT	TATATTCAAC	ATGACTATTA
801	TTACTTGTTG	ACAAGAGAAG	GAAAAAACA	TTTGCTACTA	TTGGGTTTTT
851	GGTCACTCAT	AAATGCAGAT	ATTAATTTGT	AAAATGGTAT	TTTCGTAAAT
901	TTTATACTGA	TTCAAGGTTT	GTTTTTTTGT	AACAGAAAGCA	GAAGCAAGCA
951	TTGGCTAAGC	AATTAGGATT	GAGAGCTAGA	CAAGTGGAAAG	TTTGGTTTTCA
1001	GAACAGACGA	GCTAGGTAAG	ATATAAAACG	AACCAACTCT	GTTTCCCCGA
1051	TTTGCCCCTG	TTTTATCTTA	TGTAATGATA	AGGGTAGTTT	GGTAATAACA
1101	GAACAAAGCT	GAAGCAGACG	GAGGTTGACT	GCGAGTTCTT	ACGGCGATGC
1151	TGCGAGAATC	TAACGGAAGA	GAACCGTAGG	CTACAGAAAG	AAGTAACGGA
1201	GTTAAGAGCA	CTTAAGCTGT	CTCCTCAGTT	CTACATGCAC	ATGAGCCCAC
1251	CAACCACATT	GACCATGTGC	CCTTCGTGCG	AACACGTGTC	GGCCCCCTCA
1301	CCACAACAAC	CACCTCAGGC	TGCAACGTCA	GCGCAGCACC	ACCGAGGGTC
1351	GTTGCCGGTC	AATGCTTGGG	CTCAGGGTAC	GAGGATCTCT	CACGGCTTGA
1401	CTTTTGACGC	TCTTCGTCCT	CGGTCCTAA		

Sequenz des *Arabidopsis thaliana* cDNA-Klons E

1	GGAAGCAAGA	AGCTTGAAGA	AACCAATCCA	ATTGGGTAAT	AAAGAAGATC
51	AAAACCCTAG	AAAATCTTAC	TCTCGATTCA	TCTTTAAAGC	TCTTATCTTC
101	ACCGTGCTCT	GCGCCGTCGT	ACCTGTCTTC	CTTCTCAGA	CACCAGAGCT

151	TGCTAACCAA	ACAAGACTCC	TCGAGCTTCT	CCACCTTGTT	TTCGTGCGTA
201	TCGCAGTCTC	TTACGGTCTC	TTCAGCCGCA	GGAACTACGA	CGGAGGAGGA
251	GGTGGAGGAA	CAAGCAATAG	TGATCACAAC	AAAGCTGATC	ATAGTAATAA
301	TAATTCGCAT	TCATATGTGC	CTAAGATTCT	TGAAGTATCC	TCTGTTTTTA
351	ACGTGGGTCA	CGAGAGTGAA	TCTGAACCGT	CCGATGATTC	CTCCGGTGAT
401	CAACGTAAGT	TTCAGACATG	GAAGAACAAG	TACCACATGA	AAATCCCCGA
451	GGTTGAGACT	CGTTTCGTTG	ATCGAGTTAG	TTCAGAAAAC	AGAGAGAAGC
501	CTCTGCTTTT	GCCGGTTCGG	AGCTTGAATT	ATTCTCGTGT	TTCTGATTCT
551	TCCGGCGATA	ATTCCGGTCCG	ATGGGAGAAA	GTGAGATCTA	AGAGAGAACT
601	TCTGAAGACT	CTTGGCGATG	ATAATAGTGA	TGTGCTTCCT	TCTCCGATTC
651	CATGGAGGTC	AAGATCATCT	TCATCATCAT	CATCATCATC	AAAGGAGGTT
701	GAATCTCTAC	CGTCCGTTAA	GAATCTGACT	ACAGTTGAAT	CACAGCCGTT
751	GATCAAGAAT	CTGACACCAT	CTTCTTCTTT	CTCTTCTCCA	AGAAAGTCGA
801	ATCCTATAACC	TAATCTCGCA	TC TGAGTTCC	ATCCATCTCC	GCCACCGCCT
851	CCTCCGCCGC	CGCCACCCT	ACCGGCGTTT	TATAACTCCT	CGTCGAGAAA
901	AGATCATCCC	GGAATTTACA	GGGTTGAGAG	GAGAGAATCA	TCAGTTCACA
951	AGACGAAATT	TGCAGGAGGT	GAGTTTCATC	CTCCGCCGCC	TCTTCTTCCA
1001	CCACCTCCGG	TGGAGTATTA	TAAGTCACCT	CCGACAAAAT	TCAGACTAAG
1051	TAACGAACGG	AGAAAGTCTT	CGGAGCAAAA	GATGAAAAGA	AACGCTCCTA
1101	AAAAGGTTTT	GTGGTCCGAT	CCAATCGTGG	AATCGAAGGA	ACAAGACACA
1151	GAGAAGAATG	ATCAAAGAAG	TAAC TTGGGA	AGCAAGGCAG	TGGAAGAATC
1201	CGAGAATGGA	GAACAGAGAA	GAGGAGAAAA	TGAAATCCAC	GACGAGGTTG
1251	AGAAGAAGAT	AGTAGAGGAA	GAAGGAGTTA	GTGAGATCAA	CAATGGAAAT
1301	GACGTGGACA	AGAAGGCAGA	TGAGTTCATT	GCAAAGTTCA	GAGAACAGAT
1351	TAGGTTACAA	AGAATCGAGT	CTATCAAGAG	ATCTACTAAT	AAGATCTCTG
1401	CAAATTCCTC	GAGGTAGAAC	TCATTATTTA	TTAATAATAG	GTATACATTT
1451	TAAGTATGTT	TTGGTAATCA	TCATAAGGTT	GTAATAATTA	GGGGAACAAA
1501	TATTTTTGTT	AAAAAAAAAA	AAAAAAA		

Sequenz der vorhergesagten cDNA des Gens At4g16790 (*A. thaliana*)

1	ATGGTGGAAG	CAAGAAGCTT	GAAGAAACCA	ATCCAATTGG	GTAATAAAGA
51	AGATCAAAAC	CCTAGAAAAT	TCTACTCTCG	ATTTCATCTT	AAAGCTCTTA
101	TCCTCACCGT	GCTCTGCGCC	GTCGTACCTG	TCTTCTTTC	TCAGACACCA
151	GAGCTTGCTA	ACCAAACAAG	ACTCCTCGAG	CTTCTCCACC	TTGTTTTCGT
201	CGGTATCGCA	GTCTCTTACG	GTCCTTTCAG	CCGCAGGAAC	TACGACGGAG
251	GAGGATGTGG	AGGAACAAGC	AATAGTGATC	ACAACAAAGC	TGATCATAGT
301	AATAATAATT	CGCATTCATA	TGTGCCTAAG	ATTCTTGAAG	TATCCTCTGT
351	TTTTAACGTG	GGTCACGAGA	GTGAATCTGA	ACCGTCCGAT	GATTCTCTCCG
401	GTGATCAACG	TAAGTTTCAG	ACATGGAAGA	ACAAGTACCA	CATGAAAATC
451	CCCGAGGTTG	AGACTCGTTT	CGTTGATCGA	GTTAGTTCAG	AAAACAGAGA
501	GAAGCCTCTG	CTTTTGCCGG	TTCGGAGCTT	GAATTATTCT	CGTGTTCCTG
551	ATTCTTCCGG	CGATAATTC	GGTCGATGGG	AGAAAGTGAG	ATCTAAGAGA
601	GAACTTCTGA	AGACTCTTGG	TGATGATAAT	AGTGATGTGC	TTCCTTCTCC
651	GATTCCATGG	AGGTCAAGAT	CATCTTCATC	ATCATCATCA	TCATCAAAGG
701	AGGTTGAATC	TCTACCGTCC	GTTAAGAATC	TGACTACAGT	TGAATCACAG
751	CCGTTGATCA	AGAATCTGAC	ACCACCTTCT	TCTTCTCTT	CTCCAAGAAA
801	GTCGAATCCT	ATACCTAATC	TCGCATCTGA	GTTCCATCCA	TCTCCGCCAC
851	CGCCTCCTCC	GCCGCCGCCA	CCACTACCGG	CGTTTTATAA	CTCCTCGTCG
901	AGAAAAGATC	ATCCCAGGAA	TTACAGGGTT	GAGAGGAGAG	AATCATCAGT
951	TCACAAGACG	AAATTTGCAG	GAGGTGAGTT	TCATCCTCCG	CCGCCTCCTC
1001	CTCCACCACC	TCCGGTGGAG	TATTATAAGT	CACCTCCGAC	AAAATTCAGA
1051	CTAAGCAACG	AACGGAGAAA	GTCCTCGGAG	CAAAAGATGA	AAAGAAAACGC
1101	TCCTAAAAAG	GTTTGGTGGT	CCGATCCAAT	CGTGGAATCG	AAGGAACAAG
1151	ACACAGAGAA	GAATGATCAA	AGAAGTAACT	TGGGAAGCAA	GGCAGTGGA
1201	GAATCCGAGA	ATGGAGAACA	GAGAAGAGGA	GAAAATGAAA	TCCACGACGA
1251	GGTTGAGAAG	AAGATAGTAG	AGGAAGAAGG	AGTTAGTGAG	ATCAACAATG
1301	GAAGTGACGT	GGACAAGAAG	GCAGATGAGT	TCATTGCAAA	GTTCAGAGAA
1351	CAGATTAGGT	TACAAAGAAT	CGAGTCTATC	AAGAGATCTA	CTAATAAGAT
1401	CTCTGCAAA	TCTTCGAGGT	AG		

Sequenz des abgeleiteten Gens E (*Capsella rubella*)

```

1   ATGGTGGAA G CAAAAAGCTC CAAGAAACCA ACCCAATTGG G TACTAAAAGA
51  AGATCAAAAC CCTACAAGAT TCTACTCTCG ATTCATATTC AAAGCTCTTA
101 TCCTCACC GTCTGCGCC GTCGTACCTG TCTTCCTTTC TCAGACGCCG
151 GAGCTAGCTA ACCAAACAAG ACTCCTCGAG CTTCTCCACC TTGTTTTCGT
201 CGGCATAGCC GTCTCTACG GCCTCTTCAG CCGCCGTAAC TATGACGGAG
251 GAGGAGCCGG AGGATCAAGC AATAGTGATT ACAACAAAGC TGATCATCAT
301 AACAACAACA ACTCGCACTC TTATGTTTCT AAGCTTCTTG AAGTATCCTC
351 TGTTTTTAAC GTTGATCACG AGAGTGAATC TGAACCGTCC GATGATTCCT
401 CCGGCGACCA CCGTAAGTTT CAGGCATGGA GGAACAAGTA CCACATGAAA
451 ATCCCCGAG TTGAGACTCG TTTCGTTCGAT CGAGTTAGCT CAGAGTTCAG
501 AGAGAAACCT CTGCTTTTAC CTGTTTCGAG CTGAACTAT TATCCTGTTC
551 CTGATTCCTC CGGCGATAAC TCCGGTCGAT GGGATAAAGT GAGATCCAAA
601 AGACAGCTTT TGAAGACTCT TGGTGATGAT AACAGTGATG TGCTTCCTTC
651 TCCGATTCCA TGGAGGTC AA GATCATCATC ATCGTCAAAA GAGATTGAAT
701 CTCCACCGTC CATTAAGAAT CTGACTACAG TTGAATCACA GCCGTTGATC
751 AAGAATCTGA CACCATCTTC TTCTTACTCC TCTCCAGAA AGTCGAATCC
801 TATACCTAAT CTCGCATCTC AGTTTCATCC ATCTCCGCCA CCTCCTCCTC
851 CGCCTCCACC GCTCTACCG GCGTTTTATA AATCATCGTC GAGAAAGGTT
901 CAACCGGGAA TTTACAGAGT TGAGCGGAGG GAATCAGTTC ACAAGACTAA
951 ATTTGAAGGC GGTGAGTTTC ATAATCCTCC GCCGCCTCCG CCACCACCAC
1001 CTCCGGTGGG GTATTATAAG TCACCTCCGA CAAAATTCAG AATAAGTAGC
1051 GAACGGAGAA AGACGTCGGA GCAAAAGATG AAAAGAACT TTCTAAAAA
1101 GGTGTGGTGG TCTGATCCAA TCGTGGAATC CAAGGAACAA GACACGGAGA
1151 AGAATGATCA AAGAAGTTAC TTGGAAAAGCA AGGAAGCTGA AGACTTCGAG
1201 AAAGAAGAAC AGAGAAGAAG AGAAAATGAA ATGCACGAGG AGGTTGAGAA
1251 GAAGATAGTA GAGGAAGAAC GAGCTTGTGA AAGCAACAAT GTAAGTGACG
1301 TGGACAAGAA GGCAGATGAG TTCATTGCAA AGTTCAGAGA ACAGATTAGG
1351 TTACAGAGAA TCGAGTCTAT CAAGAGATCT GCTAGTAAGA TCTCTGCAAA
1401 CACTTCGAGG TAG

```

Sequenz des *Arabidopsis thaliana* cDNA-Klons F

```

1   TACATCAAGA TAATTCCGCT AGGGTTGTGA TGATCAGAAG TCTGGTTCCA
51  GGAGTTTTCT GTGCAGGTGC AGATCTTAAG GAACGTAGAA CTATGAGTCC
101 ATCTGAGGTT CACACATATG TCAACTCGTT GCGCTACATG TTCTCGTTCA
151 TTGAGGCACT AAGTATCCCA ACTATTGCTG CAATAGAAGG TGCAGCACTG
201 GGAGGTGGAC TCGAAATGGC TCTAGCTTGT GATCTTCGAA TCTGTGGCGA
251 AAATGCGGTA TTTGGCTTGC CTGAAACAGG ACTCGCTATA ATCCCTGGAG
301 CTGGTGGGAC ACAGAGGCTC TCAAGGCTGG TTGGGAGATC GGTATCAAAG
351 GAACTTATAT TTACAGGTCG AAAGATTGAT GCAATAGAAG CTGCAAATAA
401 AGGGCTTGTG AACATTTGTG TTACGGCAGG TGAGGCTCAT GAGAAAGCAA
451 TTGAAATGGC ACAGCAGATA AACGAAAAG GTCCATTAGC TATAAAGATG
501 CGGAAGAAAG CGATTGATGA AGGAATTGAA ACAAATATGG CTTTCAGGTTT
551 GGAAGTAGAA GAGATGTGTT ATCAGAAGCT TCTTAATACT CAAGATCGTC
601 TTGAAGGATT AGCTGCTTTT GCTGAGAAGC GTAAGCCTCT GTACTCTGGC
651 AATTGAAAAC CTTATAAGAG AAAGAATCAT TAGTGTCTTA TCAATTTAGT
701 TTCAATAAAT ATTGAGTTGG ATATTAATAA CCGAATGGT CTAAGCAAGG
751 AAAAAAAAC ATCTAAATGT CAAAAA AAAA

```

Sequenz der vorhergesagten cDNA des Gens At4g16800 (*A. thaliana*)

```

1   ATGTTGAAAA GTCTCCAGAA CGCATTGAG TCCATACATC AAGATAATTC
51  CGCTAGGGTT GTGATGATCA GAAGTCTGGT TCCAGGAGTT TTCTGTGCAG
101 GTGCAGATCT TAAGGAACGT AGAACTATGA GTCCATCTGA GGTTCACACA
151 TATGTCAACT CGTTGCGCTA CATGTTCTCG TTCATTGAGG CACTAAGTAT
201 CCAACTATT GCTGCAATAG AAGGTGCAGC ACTGGGAGGT GGACTCGAAA
251 TGGCTCTAGC TTGTGATCTT CGAATCTGTG GCGAAAATGC GGTATTTGGC

```

301 **TTGCCTGAAA CAGGACTCGC TATAATCCCT GGAGCTGGTG GGACACAGAG**
 351 **GCTCTCAAGG CTGGTTGGGA GATCGGTATC AAAGGAACTT ATATTTACAG**
 401 **GTCGAAAGAT TGATGCAATA GAAGCTGCAA ATAAAGGGCT TGTGAACATT**
 451 **TGTGTTACGG CAGGTGAGGC TCATGAGAAA GCAATTGAAA TGGCACAGCA**
 501 **GATAAACGAA AAAGGTCCAT TAGCTATAAA GATGGCGAAG AAAGCGATTG**
 551 **ATGAAGGAAT TGAAACAAAT ATGGCTTCAG GTTTGGAAGT AGAAGAGATG**
 601 **TGTTATCAGA AGCTTCTTAA TACTCAAGAT CGTCTTGAAG GATTAGCTGC**
 651 **TTTTGCTGAG AAGCGTAAGC CTCTGTACAC TGGCAATTGA**

Sequenz des abgeleiteten Gens F (*Capsella rubella*)

1 **ATGCTGAAAA GTCTCCAGAA CACATTTGAG TCCATACACC AAGATAGTTC**
 51 **AGCTCGGGTT GTCATGATCA GAAGTCTGGT TCCAGGAGTT TTCTGTGCTG**
 101 **GTGCAGATCT CAAGGTAGTG TTGTCACTGG CATGTCTCGC GGAATGTTCA**
 151 **ATTTGTGTAG CCATTTCTTA AAAATATCTG TTTTGTTTTG ATTGATGTCA**
 201 **GGAGCGTAGA ACAATGAGTC CATCTGAGGT TCACACATAT GTCAACTCGT**
 251 **TGCGGTACAT GTTCTCGTTC ATAGAGGTAT ATTCATCATCT ATCCTTGATG**
 301 **GTGTTTCATAG ATTAGCAGCT ACTGTCAATG ATGGAAAAAT TGTGGGGTTA**
 351 **TGTTAACGGT CAACCTTAAC TGTGTTTGGC ATTAATGTAT ACAATCTCAC**
 401 **ATCAAATCAT TAGGATGGCC TACTACAATT TGTGAATTA GCTCTAGCTT**
 451 **TATCAAGATT AAGAATTTAC TTATAGTTGT TCTACTTATA TGAAATCTT**
 501 **TTTCTGTTTT CTTTGTTTTG TTCTCCCTTA TCTCAATGTA AAAGCTCTAG**
 551 **GAATTGAATT TCAGTGAGAT TTTTATGCTG TAGTGACTTG AAAGACTTGT**
 601 **GTTTTCTAGG CACTGAGCAT CCCAACTATT GCTGCAATCG AAGGTGTGGC**
 651 **ACTGGGAGGT GGACTCGAAA TGGCTCTAGC TTGTGATCTT CGAATTTGTG**
 701 **GTACACATAT CTGGTCATAC TCTTATTATA ATCAACCGGC CAAGATAAGA**
 751 **AAATATGGTT AATACTCAGG AAATTGAATT TCTCAGGCGA AAATGCAGTA**
 801 **TTTGGCTTGC CTGAAACAGG ACTGGCTATA ATCCCTGGGT ATATAATAAA**
 851 **CTTCATCACT CTTGACAATT GATTTCTCTT ATCACCTTT TTATCATTCT**
 901 **TCTGATAATT CATTCTCTTT TGCTTCTATT TAGAGCTGGT GGGACACAGA**
 951 **GGCTGTGCGAG GCTGGTTGGG AGATCCATAT CAAAGGAACT TATATTCACA**
 1001 **GGTCGAAAGA TTGATGCAAG AGAAGCTGCA AATAAAGGTC TTTGACTATA**
 1051 **CAAGTTGCAT TTCTTATATT GTGGTTTTGT TTTGGCAATC TTCCCAAATA**
 1101 **AGGATTGATT CATATCTCT TTTCATTTTT TTTTCTTTT GTGTGGTGCA**
 1151 **GGGCTTGTGA ACATTTGTGT TACCGCTGGT GAGGCTCATG AGAAAGCAAT**
 1201 **TGAAATGGCA CAGCAGATAA ATGAAAAAGG TCCACTGGCT atAAAAGATGG**
 1251 **CGAAGAAAGC GATTGATGAA GGAATAGAGA CGAATATGGC TTCGGGTTTG**
 1301 **GAGGTAGAAG AGATGTGTTA TCAGAATCTT CTTAATACTC AAGATCGTCT**
 1351 **TGAAGGATTA GCTGCTTTTG CCGAGAAGCG AAAGCCTCTT TACTACTGGCA**
 1401 **AGTGA**

Sequenz des *Arabidopsis thaliana* cDNA-Klons G

1 **ACCCATTTTCG TCTCCTTCAT TTTGACGTTT TCTTTAGATC TGTTAATAAT**
 51 **CCCATCCATG GCTTGGTCTA GTCAC TAGTG GTGTTATCTC TCTGTGTGGA**
 101 **ATAAAAAGCA AGGTTTTTTA ATCACTTGGA TTGTTTTAAG ATATGGCAAC**
 151 **TTTGAACCCCT TTTGATTTGT TGGATGATGA TGCTGAGGAT CCAAGCCAGC**
 201 **TCGCTGTTGC CATCGAGAAG ATTGATAAGT CCAAGAAATC TGGACAGGTT**
 251 **TCGAGCTTGC CTGCTAAGTC AGCTCCTAAG CTTCCATCGA AGCCACTTCC**
 301 **TCCTGCTCAA GCCGTGAGAG AGGCAAGGAG TGATGCTCCA CGTGGTGGTG**
 351 **GAGGCCGTG AGGATTTAAC CGTGGTCGTG GTGGTTACAA CCGTGATGAT**
 401 **GGTAACAATG GATATTCAGG GGGATACACT AAACCCTCAG GTGAAGGAGA**
 451 **TGTTTCAAAG TCTTCTTACG AGAGGCGTGG CGGTGGTGGT GCTCCTCGTG**
 501 **GTTCTTCCG TGGTGAAGGT GGTGGACCTG GTGGTGGTGG TCGTGGTGA**
 551 **TTCAGCAATG AGGGTGGTGA TGGGGAACGT CCTCGAAGGG CCTTTGAGCG**
 601 **TCGTAGTGGA ACTGGCAGAG GGAGTGAAT CAAGCGTGAC GGATCTGGTC**
 651 **GTGGGAATG GGGAACTCCA GGGGAAGAGA TAGCTGCTGA GACTGAAGCA**
 701 **GTAGCTGGTG TTGAGACTGA GAAGGATGTT GGAGAGAAGC CAGCTGTTGA**
 751 **TGATGTAGCT GCTGATGCTA ACAAGGAGGA TACTGTTGTT GAGGAGAAAG**
 801 **AGCCTGAGGA TAAGGAAATG ACTCTTGATG AGTATGAGAA AATACTCGAG**

851	GAGAAGAAA	AGGCACTTCA	ATCATTA	ACT	ACCTCTGAGA	GGAAAGTTGA
901	TACGAAAGTG	TTTGAATCAA	TGCAACA	ACT	GTCAAAACAAG	AAGTCTAATG
951	ATGAAATCTT	CATCAAGCTG	GGTCTGATA	AGGACAAACG	CAAAGATGAC	
1001	AAAGAAGAGA	AGGCTAAGAA	GGCTGTGAGC	ATCAATGAGT	TTCTGAAGCC	
1051	AGCAGAGGGT	GGGAACTACT	ACCGAGGAGG	TCGTGGTGGC	CGTGGACGTG	
1101	GTGGTCGTGG	CCGTGGAGGT	GTTTCTAGTG	GCGAATCTGG	TGGTTACCGT	
1151	AATGAAGCTG	CACCGCAAT	TGGAGATGCT	GCTCAGTTCC	CATCTCTTGG	
1201	GGGCAAGTAA	GATCCATCCA	TGATAACGTC	CACACTCGTG	CATCCTCCTT	
1251	TAGGATTTTG	TGCGAGGATT	TACTGTTTAC	TGGTCTCTCG	TTGTCAGATG	
1301	TAAATAATTA	GGTGTCTGTCG	TCAGTTTTTA	GATTTTATGC	AAAACTTTAC	
1351	ACTTGTGGT	GTTCTTTACT	TTTGAAACAG	TACACTCTTC	TTAATTTTAC	
1401	CATTTTACTC	GTTTTGTGTT	TGTTGTTTTT	TTATTTATGA	AACATGTTTT	
1451	GAGACGGACA	TATCTAATGT	TATCGGGTTT	GAGTCTTTAC	TCTGTCTCCT	
1501	CTTTTCTCCA	AAAAAAAAAA	AAAAAAAAAA	AAAAAAA		

Sequenz des abgeleiteten Gens G (*Capsella rubella*)

1	ATGGCAACTT	TGAACCCTTT	TGATTTGTTG	GGTGATGATG	CTGAGGATCC	
51	GAGCCAGCTT	GCTGTTTCCA	TCGAGAAAAC	TGATAAGTCC	AAGAAATCTG	
101	GACAAGTTTC	GGGCTTGTC	GCTAAGTCAA	CTCCTAAGCT	TCCCTCTAAG	
151	CCACTTCCTC	CTGCTCAAGC	CGGTGAAGTT	TCTTCTTCT	CATAAAAAAA	
201	GAGTCAATAC	TTTAAATATA	AAAATTCGAT	CTTTTACT	GGATAAGCT	
251	GTTAATATCG	TTGTTGGAT	TATGTTGATT	ATTCATGGA	GTTGGGTATT	
301	GTTGTTTGAG	TGATTTGGAT	TGGTATCTGA	TTAGTTTTGG	TGTTGTTGAT	
351	TAGTGAGAGA	GGCAAGGAGT	GATGCTCCGC	GTGGTGGTGG	TGGAGGAGGC	
401	CGTGGAGGAT	TTAGCCGTGG	TCGTGGTGGT	TACAATCGTG	ATGATGGTAA	
451	CAATGGATAT	TCAGGAGGAT	ACTCTAAACC	CTCTGAGGAA	GGAGATGTTT	
501	CTAAGTCTTC	ATACGAGAGG	CGTGCTGGTG	GTGGTGGTGC	TCCTCGTGGT	
551	TCCTTCCGTG	GTGAAGGTGG	TGTTGGTCCG	CGTGGTGGAT	TCGGCAATGA	
601	GGGTGGTGAA	GGTGAACGTC	CTCGAAGGAC	ATTTGAGCGT	CGTAGTGGAA	
651	CTGGCAGAGG	GTCAGTCTTA	GTCTGTGTTT	GCATAATCTA	CAATTGAAAT	
701	GACTGTCCA	ATCTCATCTG	TTGGTGTFTA	TCTTTTGGAT	AGGAGTGACT	
751	TCAAACGTGA	TGGATCTGGC	CGTGGGAATT	GGGGAACCTC	AGGGGAAGAG	
801	TTAGCTTCGT	AAGTATATGT	TTATGGTTTA	TTGCATCTGT	TATAATTGTA	
851	TGTATTTATG	TGATTATGAT	TGGGAATTTG	GACTCAAAGA	AGCTTCTCTG	
901	CATCTTTAAG	TGAGACTGAA	GCAGTAGCTG	GTGTTGACAC	TGAGAAGGAT	
951	GTTGGGGAGA	AACCAGCTGT	TGATGCTGAT	GCTAACAAGG	AGAACACTGC	
1001	TGAGTTGAG	GAGAAAGAGC	CTGAGGATAA	GGTAAACAGA	GAATGGAATA	
1051	GAATTACACT	GGCATCAGTT	TTTCATCATC	AATGATTGCT	TATTGACTTG	
1101	TGTTTCTACA	GGAAATGACT	CTTGATGAGT	ATGAAAAGAT	ACTTGAGGAG	
1151	AAGAAAAGG	CAC TTCAATC	ATTGACCACC	ACTGAGAGGA	AAGTTGATAC	
1201	TAAAGTGT TT	GAATCAATGC	AACAAC TGTC	AAACAAGAA G	TCTAATGATG	
1251	AAATCTTCAT	TAAGTTGGTA	AGATCATGTA	TCATTTCTTT	TTTTGGTCCG	
1301	CTGTAAACTT	GCTTAAACGA	TTCTGAACCT	TTCTTTGATC	TATTTACCAG	
1351	GGTCTGATA	AGGACAAACG	CAAAGACGAC	AAAGAAGAGA	AGGCTAAAAA	
1401	GGTTTGAACC	ATTGTCTGAA	TCTAACATTT	TGTAGAAGTC	AATAAGTTTG	
1451	CTTACTGATG	TGTTTTCTGA	TTTGCAGGCT	GTAAGCATCA	ATGAGTTTCT	
1501	GAAGCCAGCT	GAGGGTCAGA	ACTACTACCG	AGGAGGTCCG	GGTGGCCGTG	
1551	GACGTGGTGG	TCGCGGCCGT	GGAGGTGTTT	CTAGTGGAGA	GCATGGTGGT	
1601	TACCGTAATG	AACCTGCACC	AGCCATTGGA	GATACCACTC	ACTTCCCATC	
1651	TCCTGGCGGA	AAGTAA				

Sequenz des *Arabidopsis thaliana* cDNA-Klons 12918 (EST 6)

1	AGAAACTGAA	GGAGTTGGAG	AAGAAGATGA	AACTAGCTGG	ATACAAAACCG	
51	GAACTAGAGT	TTGCTTTACA	CAATGTAGAG	GAAGAGCAGA	AGGAGAAGCT	
101	ATTGTTATGG	CACAGCGAGA	AGTTAGCTGT	TGCCTTTGGT	TGCATAAAAC	
151	TCCCCTGAAG	TTCACCAATA	CAAGTGTTC	AGAATTTGAG	AATCTGTGGT	
201	GATTGTCATA	AAGCAATCAA	ATTTATTTTCG	GAGATAGAGA	AACGAGAGAT	
251	CATTGTAAGA	GACACCACAA	GGTTTACCA	TTTCAAAGAT	GGGTCTTGCT	


```

301 CTTGTGGCGA TTACTGGTGA AAAGAGAAGA GCTTTGACTC TCTCATTGGT
351 CAAACCTGAC TGTATTTATA TGCGTTATTG TGTGGTAAAG TTTGACCTT
401 TGACTTTACA AGTTGGCGTT AAGAAGAGAG ATGCGTAGAT CAGCGAGTGG
451 TTCTAGATTT TTGGATCATT TTCCGGCGAC TTCAAGGTCT CCGCCTCGAT
501 CTCAGAGTGT TACAGCTATG GAAGATGATG TGGAGCTGCT TTTGCCTAGG
551 TACGATCCGA ATTCACAAGC GGGGAAGAGA GAGAAAGTCAA GATTCAGATT
601 TGCAGAAAAC GTCATCCATT TGATTCCTCT CATTCCTCTT CTCTGTGTG
651 CAATCCTCTG GCTCTCCTCT TATTCAGCAG CGTTAAGGAG TTGAGTTCAA
701 GAAGCAACAT GTTGTCTTGT CTCCATGGAA ACTCATCATA TTCAGTTTTG
751 GGAAAGGAAA CAATTATTTT ACCGCCGGTG ATTATGTGCC GCAAACCATA
801 CGTAACTCTT GTAATTTTTG GTTCTGTAGA CACATAAAAAG GATCTCTCGT
851 TTTTCATGAAA TGTATGTTTA ATAGTTCAC TATAAAAAAA AAAAAAAA
901 AAAAAAAA

```

Sequenz des abgeleiteten Gens 6 (*Capsella rubella*)

```

1 AGAAACTGAA TGAGTTGGAG AAGAAGATGA TATTAGCTGG ATACAATCCG
51 GAGCTAGAGT TTGATTTACA CAATGTAGAA GAAGAGCAA AGGAGAAGCT
101 ATTGTTATGG CATAGTGAGA AGTTAGCTGT TGCATTTGGT TGCATAAAAC
151 TCCCTCAAGG TTCACAGATA CAAGTGTTC AAGAACTTGAG AATCTGTGGT
201 GATTGTCATA AAGCAATCAA ATTTATCTCG GAGATAGAGA AACGAGAGAT
251 CATGGTGAGA GACACCACAA GGTTCACCA TTTCAAAAAT GGGTCTTGTT
301 CTTGTGGCGA TTACTGGTGA AAAGAGAAGA GCTTTGACTC TTCCATTGGT
351 CAAACCTGAC GTAACGACTG TATTTATATG CGTTATTGGG TGAAAAAAGT
401 ATCGACCTTT GACTTGACAA CTTGCGTTAA GGATAGAGAT CCGTAGATCG
451 ACGAGTGGTT CTAGAGTTTC GGATCAGTCT CCGGCGAAGA CCGCCTCC
501 GCCACGATCT CAGAGTGTTA CAGCTATGGA AGATGATGAT GTAGAGCTGC
551 TTTTGCCTCG GTACGACCCG AATCTCAAG CGGGGAAGAG AGAAAAATCA
601 AGATTCAGAT CTGCAGAAA CGTCATCCAT CTCATTCCCT TCATTCTTCT
651 TCTGTGTGTC ATAATCCTCT GGCTCTTCTC TCATTGAGT aaaccgagaa
701 attgattcag tctctatgga agctataatt gatctgtgaa acttaattag
751 gaacttcaca aagaccatt tggatataag gattgggata tctctgtaat
801 gttagaatct tgtgttgatt catttcattg tttcattttg ttcatatgat
851 ggatttggtg atttgtcagC AGCCTTAAGG AGTTGAGTTC AAGAAGCAA
901 AATGTTGTCT TGTCTCCATG GGAATCAT TTAGTTTGA GGGAAGGAAA
951 CAACTGTTTT TTTACTGCCG GTGATTTTGT GCCGCAACGC AAAGCCATAC
1001 GTAACCTTTG CTTTTTTTGG TT

```

Sequenz des *Arabidopsis thaliana* cDNA-Klons H

```

1 CAAGCTTCTT CGATTTTGCT CTCTCTTACA CAGCCAATCG GTGTTTTTCG
51 AGCTTTCAGG CCTCAATCCA AGACATCTA TATAAGCATA TTGCAGAAGA
101 GGCGTTCTA ATTGTTGCAT TGAGTTTATC GCTATGACGT AGGGAAATTC
151 TAATTTAGGG GAGGCTCAG AGTTTGCAC TAACTTCATA TCGGCTCTTG
201 ACGTTGTTGA GTGTAATTGA ACAAGAATGT GTAGGCAGAA TTGTCGCGCG
251 AAATCCTCAC CGGAGGAAGT GATTTCAACT GATGAGAATC TCTTGATATA
301 TTGTAAACCT GTTCGACTAT ATAACATCTT TCACCTTCGC TCTCTAGGCA
351 ACCCATCGTT TCTTCCAAGA TGCTTGAAC ACAAATTTGG AGCAAAGCGC
401 AAAAGAAAGT CAAGATCTAC TGGGATGGTA GTTTTCAACT ATAAGGATTG
451 TAATAACACA TTACAGAAA CTGAAGTTAG GGAGGATTGT TCTTGTCCAT
501 TTTGCTCTAT GCTATGTGGT AGCTTCAAG TGGGCAACTA TTACAACTGA
551 GGGGCTGCAA TTTTATTTGA ATTCATCTCA TGATTTATTT GAATTTGAGT
601 TCAAGCTTTC GGAAGAATAC CAGACAGTTA ATGTTTCTGT AAACTTAAT
651 TCCTTCATAT TTGAGGAAGA AGGAAGTGAT GACGATAAAT TTGAGCCCTT
701 CTTTCTCTG TCGAAACCTC GTAAGCGGAG ACAAAGAGGT GGCAGAAATA
751 ACACCAGGAG ACTTAAAGTA TGCTTTTTAC CGTTGGATTC ACCCAGTTTA
801 ACTAATGGCA CAGAAAATGG AATCACCTTA CTTAATGACG GAAACCGTGG
851 TTTAGGATAT CCCGAGGCAA CAGAGCTTGC TGGACAATTT GAGATGACCA
901 GCAACATTC ACCAGCCATA GCCCACTCTT CTCTGGACGC TGGTGCTAAA
951 GTTATATTGA CAAGCGAAGC TGTGGTCCCT GCTACTAAGA CAAGAAAGTT

```

1001 ATCTGCTGAG CGATCAGAGG CTAGAAGCCA CCTACTTCTT CAGAAACGCC
 1051 AATTCTATCA TTCACACAGA GTCCAGCCAA TGGCGCTTGA GCAAGTAATG
 1101 TCTGACCGGG ATAGCGAGGA TGAAGTCGAT GACGATGTTG CAGATTTTGA
 1151 AGATCGCCAG ATGCTTGATG ACTTTGTGGA TGTGAATAAA GATGAAAAGC
 1201 AATTCATGCA TCTTTGGAAC TCGTTTGTAA GAAAACAAAG GGTATAGCA
 1251 GATGGTCATA TTTCGTGGGC ATGTGAAGCA TTTTCAAGAT TTTACGAGAA
 1301 AGAGTTGCAC CGTACTCAT CACTCTTCTG GTGTTGGAGA TTGTTTTTGA
 1351 TTAAACTATG GAACCATGGA CTGTGCGACT CAGCCACCAT CAACAACGTC
 1401 AATACCATCC TCGAGAATTG CCGTAATAGC TCAGACACCA CCACCACCAA
 1451 CAACAACAAC AGTGTGGATC GTCCCAGTGA CTCAAAACACC AACAAACAATA
 1501 ACATTTGTGGA TCATCCCAAT GACATAAACA ACAAGAACA TGTTGACAAC
 1551 AAGGACAATA ACAGCAGAGA ACAAAGTAAT TAAATAGGAA AATCTCCGGC
 1601 TTAGATGATA CCGATTTATC GGATTGTAAC TTATTCTTCT TTCTTAAAAA
 1651 ATGTTTTAGG AGCAAACAAA TTTTTTATAT GTTAGTGTAT TCAACTGATT
 1701 ACATTTTTAG TTAAAAAAA AAAAGGATTC TGCTTATAAC TAAAAAAA
 1751 AAAAAAAA

Sequenz der abgeleiteten cDNA des Gens H (*Arabidopsis thaliana*)

1 CAAGCTTCTT CAATTTTGCT TGCTCTCTCT CTTACACGGC CAATCGGTGT
 51 TTTCGCAGCT TTCAGGCCTC AATACAAGAC ATTCTATATA AGCATATTGC
 101 AGAAGAGCGG GTTCTAATTG TTGCATGGAG TTGAACAATA TGACGTAGGG
 151 AAATTCCTAAT TTAGGGGAGG CCTCAGAGTT TGCCTAACT TCATAATCAG
 201 CTCTGGACGT TGTGATTTGT ATTTGAACAA GAATGTGTAG GCAGAATTGT
 251 CGCGCGAAAT CCTCACCGGA GGAAGTGATT TCAACTGATG AGAATCTCTT
 301 GATATATTGT AAACCTGTTT GACTATATAA CATCTTTCAC CTTGCTCTC
 351 TAGGCAACCC ATCGTTTCTG CCAAGATGCT TGAACACAA AATTGGGGCA
 401 AAGCGCAAAA GAAAGTCAAG ATCTACTGGG ATGGTAGTTT TCAACTATAA
 451 GGATTGTAAT AATACATTAC AAAGAAGTGA AGTTAGGGAG GATTGTTCTT
 501 GTCCATTTTG CTCTATGCTA TGTGGTAGCT TCAAGGTGGG CAACTATTAC
 551 AACTGAGGGG CTGCAATTTT ATTTGAATTC ATCTCATGAT TTATTTGAAT
 601 TTGAGTTCAA GCTTTTGGAA GAATACCAGA CAGTTAATGT TTCTGTAAAA
 651 CTTAATTCCT TCATAATTTGA GGAAGAAGGA AGTGATGATG ATAAATTTGA
 701 GCCTTCTCT CTCTGCTCGA AACCTCGTAA GCGTAGACAA AGAGGTGGCA
 751 GAAATAACAC CAGGAGACTT AAAGTATGCT TTTTACCGTT GGATTCACCC
 801 AGTTTAGCTA ATGGCACAGA AAATGGAAAT GCCCTGCTGA ATGATGGAAA
 851 CCGTGGTTTA GGATATCCCG AGGCAACAGA GCTTGCTGGA CAATTTGAGA
 901 TGACTAGCAA CATCCACCA GCCATAGCCC ACTCTTCTCT GGACGCTGGT
 951 GCTAAAGTTA TATTAACAAC CGAAGCTGTG GTCCCTGCTA CTAAGACAAG
 1001 AAAGTTATCT GCTGAGCGAT CAGAGGCTAG AAGCCACCTA CTTCTTCAGA
 1051 AACGCCAATT CTATCATCTT CACAGAGTCC AGCCAATGGC GCTTGAGCAA
 1101 GTAATGTCTG ATCGGGATAG CGAGGATGAA GTCGATGACG ATGTTGCAGA
 1151 TTTTGAAGAT CGCCAGATGC TTGATGACTT TGTGGATGTG AATAAAGATG
 1201 AAAAGCAATT CATGCATCTT TGGAACTCGT TTGTAAGAAA ACAAAGGGTT
 1251 ATAGCAGATG GTCATATCTC TTGGGCATGT GAAGTATTTT CAAGATTTTA
 1301 CGAGAAAAGAG TTGCACTGTT ACTCATCACT CTTCTGGTGT TGGAGATTGT
 1351 TTTTGATTAA ACTATGGAAC CATGGACTTG TCGACTCAGC CACCATCAAC
 1401 AACTGCAATA CCATCCTCGA GAATTGCCGT AATACCTCAG TCACTAACAA
 1451 CAACAACAAC AGTGTGGATC ATCCCAGTGA CTCAAAACACC AACAAACAATA
 1501 ACATTTGTGGA TCATCCGAAT GACATAAAAA ACAAGAACA TGTTGACAAC
 1551 AAGGACAATA ACAGCAGAGA CAAGTAATTA AATAGGAAAC ACTCCGGTTT
 1601 AGATGATACC GATCTATCGG ATTGTAACCT ATTCTTCTTT CTTAAAAAAA
 1651 TTGTTTTAGGA GCAAACAAAG ATTTTATTTG TTAGTGTATT CAACTGATTA
 1701 CATTTTTAGT TAAAAAATG GATTCTCCTT AATAACT

Sequenz des abgeleiteten Gens H (*Capsella rubella*)

1 CAAGCTACTT CGATTTTTGC TTGCTCTTCT TTTACACGAC CAATCGGTGT
 51 TTTCGCAGCT TTCAGgtttg tctcaaatct caaattagat cggagtcacg
 101 tcataacaat tgataaaacc taattgcttt cgttatattg taagagtatt
 151 aaatthttgc agtagatccg gaattcgaat tttacctaa ctttcatata
 201 gactaatata gttttagat acggagccta gttctggatt ttttagatttc

251 ttcaatgcac attttttcca taaaattcgt acaatttgat ctccgaactt
 301 gtgtttttaa caaagtttct gtttttagaa ttattgctta aagtgtcatt
 351 ctttcatctc gtcctgagaa agctgtgatt tttctaataa gctatgattc
 401 ttctgctgag tggaattgga taaagttttc gttttttgta cgcctactct
 451 tgtatgtgtc ttgaaaggac tctctgaatt gctaaaaaaaa gtcttgtgct
 501 ttcttttgcta cacagGTCGC AACAGGACA TTCTCTAGAA GGTCTATTG
 551 CAGAATTTAC CCTGAAGAAG CCGGATCTAA TTGCTGCATC AAGTTGATG
 601 CTATGAAATA GAAATTTCTAA TTCCAGAAAAG CCTCAGAGT TTGCACTAAC
 651 CTCATAAATC GGCTGTGGAC ATCGTTGATT GTGAATGAAC AAGAATGTGT
 701 AGACAAAATTT GTCGCGCGAA ATCCTCGCCG GAGGAAGTGA TTTCAACTGA
 751 TGAGAATCTC TTGATATATT GTAAACCTGT TCGACTCTAC AACATCTTTC
 801 ACCTTCGCTC TCTATGCAAC gtatgattta ccttcctctc tcatcatttt
 851 agctcagtaa tgactgtttt cagttttttc tttcatctcc tgtgtagatt
 901 tcccactaat aatttggggt tgtaagctg ataatggcct gattcatggc
 951 gagtgtgtgc ttgttttgtc tcataatggt atttgaactt gttgcttgtt
 1001 gttacagCCA TCGTTTCTTC CAAGATGCTT GAACTACAAA ATTTGGAGCAA
 1051 AGCGCAAAAG AAAGtatgag tttcttctcg tatgtagtgt ctacagtgat
 1101 atgttattat attataagga agctgattat ttatctttgt tgagagatat
 1151 ggacatgatg atgtttttca ccttttcatg ctatacactt ttttcaaaa
 1201 tgtgttttga tttagTCAAG ATCTACTGGG ATGGTAGTTT TCAACTATAA
 1251 GGATTGTAAT AACACGTTAC AGAAAACCTGA AGgtgagtat ttttcggttc
 1301 ttagacatat tcgacgccag tttctatggt tctctagatg aatttttact
 1351 aactattttt ctgtattggt atgcagTTAG GGAGGATTGT TCTTGTCAT
 1401 TTTGCTTTTAT GCTATGTGGT AGCTTCAAGG TGGGCATATA TTACAACCTGA
 1451 Ggttttcttc gaggcctttc atatctaaca ctgtgaaacg ctactaccct
 1501 ttcattgctgt atacctgcag tgtgcggttg catacttttg tgtttgttgg
 1551 ttgtcttctc actcctgaac tgctgagtga gaaaacatgt tccagatgga
 1601 gcttacaatc cattgtcttg tgtctatgca gGGGCTTCAG TTTTCAATTTGA
 1651 ATTCATCTCA TGATTTGTTT GAATTTGAGT TCAAGgtatg tggttttatg
 1701 gaatttctatg ttttactctg ccttttttta atgagattat agttaaaaagg
 1751 gtctttctctg ttgtagCTTT CGGAAGAATA CCAGACAGTT AATGTTTCTG
 1801 TGAAACTTAA TTCCCTCATA TTTGAGGtta gttaccttta actgttaatt
 1851 ccttcatatt tgaggttagt tacctttaac tgggaaatcc tatagctggt
 1901 gaaaatgtag tttatattcc atccttcttt gtactagGAA GAAGGAAGTG
 1951 ATGACGATAA ATTTGAGCCC TTCTCTCTCT Ggtaaatctc aacaccctta
 2001 gatagaattc cttaatagca gtaactcctt acttttcttg tcagtacttc
 2051 tttataaatc caaccataat gttttgcagC TCAAAACCTC GAAAGCGTAG
 2101 ACAAAGAGGT GGCAGAAATA ACACAAGGCG ACTTAAAGTT AGCTTTTTTAC
 2151 CGCTGGATT CACCCAGTTTA GCTAATGGCA CAGAAAATGG AACTTCCCTA
 2201 CTGACTGATG gtaaaatcac atcttcttct gtggcattcc ttgtggtttt
 2251 gaacttcata ttacaggaga agatacaatg gcttgattgt ttagtttttt
 2301 tccttctcct cgcattcttc atgagagggg aattttacca taactgatgt
 2351 acaaaatgaa tgacatgcta cagGAAACCG TGGTCTAGGA TATCCCGAGG
 2401 CAACAGAGT TTGCTGGACAA TTTGAGATGA CTAGCAACAC TCCGCCTGCC
 2451 ATAGCTCATT CTTCTGGCGC TGATGCTAAA GTTGTAATTGA CAAGCGAAGC
 2501 TGTGGTCACT GCTACTAGGA CAAAAAAGTT ATCTGCTGAG CGATCAGAGG
 2551 CTAGAAGggt tgttcatcat gcatcttgtc atcataattt accattcccg
 2601 ttgtttacaaa ttcactcttc tatgatggat aagtgtttac catacttccc
 2651 taataaccga gaaaatttct tccagCCACC TACTTCTTCA GAAACGCCAA
 2701 TTCTATCATT CTCACAGAGT GCAGgtgatt tgagttcttt cacctacttc
 2751 ttaagcattt tctttatatt gcccgctatg atatcttatt aaagcatact
 2801 tgggtcccgat tcacgggtact ggtttgattg ttctcatcca aatctgtata
 2851 ttttattttg tctatccacg aagaagatac taagtccatg tagtctctat
 2901 ttcatttttaa ctaacaaaaa actgcatttt tttctgacaa agCCAATGGG
 2951 TCTTGAGCAA GTAATGTCCTG ATCGGGATAG CGAGGATGAA GTTGATGACG
 3001 ATGTTGCAGA TTTTCGAAGAT GCCCAGgtat tccatgattt ctttctgcta
 3051 actacaggaa atggtatatt cgatataact tgctaatggc ttttgaactt
 3101 taaaaaagct gcagATGCTT GATGATTTTG TGGATGTGAA TAAAGATGAA
 3151 AAGCAATTTCA TGCATCTTTG GAACTCATTT GTAAGAAAAC AAAGGtaact
 3201 acatttctta aacttttaag aaacacacac tcaaaaggct caaagcttat
 3251 tttcttctgt ttctgaacgg aaagacctta tgtgtttacat tacaatcctt
 3301 tgtactgatt ttggttatgg aacaatatgc taaaattcaa ataatgatta
 3351 atgttgcagA GTTATAGCAG ATGGTCATAT CTCATGGGCA TGTGAAGCAT

3401	TTTCAAGATT	TTACGAGAAG	GAATTGCACA	GTTACTCATC	ACTTTTCTGg
3451	taaatacact	gactacacta	catactcatt	tacacacgtc	acattattac
3501	actatcaatt	tggctacgtt	tctgagaaaa	taaaaatttg	cagGTGCTGG
3551	AGATTATTTT	TGATCAAAC	ATGGAACCAT	GGACTTGTGG	ACGCAGCTAC
3601	CATCAACAAC	TGCAGTACCA	TACTCGAGAA	TTGCCGTAAT	AGCTCAGACA
3651	CCAACAACAA	TAACATCGCA	GATCATCCCA	ATGACATGAA	CAACATCAAC
3701	ATTGACAACG	ACGACCATAA	CTGCAGAGAC	AAGTAACCTA	AAATAGGAAG
3751	ATCTCCGTTT	TAGAGGATTA	CTATTGATGG	GATCGTATAA	CTTATTCCCTT
3801	CTTTTATTTCA	AAAAATGTTT	AGGAGCAAAC	AAAGATGTAC	TCAACTGATT
3851	TTTACATTTT	TTAGTTTGAA	AAAAATGGATT	CTGATT	

Sequenz des *Arabidopsis thaliana* cDNA-Klons 4420 (EST 7)

1	CGTTGGTTAT	ACACCTACAT	TGCGTTATCT	AATCACGATC	CTCTTTCTCG
51	GAGTTGAGGT	ATACATAATG	GCTATAACAG	GGCTTGGGTT	TGTCGTCTCT
101	GTATTGGAAG	AAAGATACGG	TTTTGATGCG	ATTAAGGAAG	GAACCGCTCT
151	GATGAAAGGG	AGGAGGATAA	CAGGGTTGGC	CTTGGCGGGT	GTATTTGTGT
201	TTCTATCGAG	TTTCATTGGT	CATGGGATGG	AGAAGTTGGC	GAAGGAGCTA
251	GATATGGACT	TGAGCTCAGG	GTCGTGGTGG	AGGTCGGTGG	TTGTGGCCGG
301	TGGGTGGGAC	GGGTGGAAGC	TGGTTTGTAT	GTATGGGGCT	GAGGTTGTGC
351	TGAGTTATGT	TGTAATCACT	GTTTTTACT	GTGAGTGTAG	GAAACGCCAT
401	GGTAACAGTT	ATGCAATTGT	TGCTGATGAG	GAAGGTCTTG	CTATTTAAAA
451	TTAGTTCCAA	CACTTTTTTTG	TTGTTGATAA	TTCTCTTGTTG	AATCTGTGAA
501	ATACATTTTTT	GTTTACCAAA	AAAAAAAAAAA	AAA	

Sequenz der abgeleiteten cDNA des Gens 7 (*Arabidopsis thaliana*)

1	ATGATTCTTT	TCAGAAACAA	ACACATCTTA	TTACCAATCT	TTGCTTTCAT
51	AGCGATCCCT	CTCGCAGCCT	TACACCCTC	CTTAACTCTT	ACCTCCTTTC
101	GCCTTAAAAA	CCATGTCCTT	CGTTTAGAAG	CCTTAGCTAA	CGTTGTGCAT
151	ACACGGTTTG	AAGCTAGACA	GATTTGGCAA	GAGCTAGAG	AAGATGCTGT
201	CTCGCTTTTA	CATCTCAAAT	TCCGGTATTT	CGTCCCATCC	TTTATCCTCT
251	CCTGTATGGC	CTCCATAACT	GTCATCACAT	CAACTTCCTT	CTCACACCAA
301	GGCATAAACC	CATCTCTGAA	GTCGTCATTT	GCTTCAGTCA	AGTCTTCGTG
351	GATGCGAGTA	ACAGCAACTT	CCATCATCGT	CTACGGTCTG	CTCTTCCTCT
401	ACTCACCAAT	TTCCATGTTT	CTATCAGCTC	TCGTTGGTTA	TACACCTACA
451	TTGCGTTATC	TAATCACGAT	CCTCTTTCTC	GGAGTTGAGG	TATACATAAT
501	GGCTATAACA	GGGCTTGGGT	TTGTCGTCTC	TGTATTGGAA	GAAAAGATACG
551	GTTTTGATGC	GATTAAGGAA	GGAACCGCTC	TGATGAAAGG	GAGGAGGATA
601	ACAGGGTTGG	CCTTGGCGGG	TGTATTTGTG	TTTCTATCGA	GTTTCATTGG
651	TCATGGGTATG	GAGAAGTTGG	CGAAGGAGCT	AGATATGGAC	TTGAGCTCAG
701	GGTCGTGGTG	GAGGTCGGTG	GTTGTGGCCG	GTGGGTGGGA	CGGGTGGGAA
751	CTGGTTTGTG	TGTATGGGGC	TGAGGTTGTG	CTGAGTTATG	TTGTAATCAC
801	TGTTTTTTTAC	TGTGAGTGTA	GGAAACGCCA	TGGTAAACAGT	TATGCAATTG
851	TTGCTGATGA	GGAAGGTCTT	GCTATTTAA		

Sequenz des abgeleiteten Gens 7-1 (*Capsella rubella*)

1	ATGATTCTCT	TCAGAAACAA	ACACATCTTG	TTCCCAGTCT	TTGCTCTCAT
51	AGCGATCCCT	TTAGCTGCCC	TACACCCTATC	CCTAACTCTG	ACCTCCTTTC
101	GCCTTAAAAA	CCATGTCCTT	CGTCTAGAAG	CCTTAGCTAA	CGTCGTGCAC
151	ACACGGTTTG	AAGCTAGACA	GATTTGGCAA	GAGCTAGAG	AAGATGCTGT
201	CTCGCTTTTA	CATCTCAAAT	CCCGGTATTT	TGTCCCATCC	TTTATCCTCT
251	CTAGTATTGC	CTCCATAACC	GTCATTACAT	CTACTTCCTT	CTCACACCAA
301	GGCCTAAACC	CATCTCTTAA	GTCGTCATGG	GCTTCAGTGA	AGTCTTCGTG
351	GATGCGAGTA	ACAGTAACTT	CCGTCATCGT	CTATGGTCTG	CTCTTCCTCT
401	ACTCTCCAAT	TCCCATGTTT	ATATCAGCTC	TCCCTCGGTTA	CACACCCACA
451	TTACGTTATC	TCATCACGAT	CTTCTGTCTC	GTTGTTGAGG	TATACATAAT
501	GGCTATAACA	GGGCTCGGCC	TTGTCGTCTC	TGTATTGGAA	GAAAAGATACG
551	GTTTTGATGC	GATTAAGGAA	GGAACCGCTC	TGATGAAAGG	GAGGAGGATA

601 ACAGGGTTGG CTTTGGCGGG TGTATTTGTG TTCTTATCAA GTTTCATTGG
 651 TCATGGGATG GAGAAGTTGG CAAAGGAGCT AGATATGGAC TCCAGCTCAG
 701 GGTCTGGTGG GAGGTCGGTG GTTGTGGCCG GTGGGTGGGA CGGTGGGAAG
 751 CTGGTTTGTG TGTATGGGGC TGAGGTGGTG CTGAGTTATG TTGTAATCAC
 801 TGTTTTTTAC TGTGAGTGTA GGAATCGCAA TGGTATCAGC GATGCAAATG
 851 TTGCTGATGA CGCAGGTCTT GCTATTTAA

Sequenz des abgeleiteten Gens 7-2 (*Capsella rubella*)

1 ATGATTCCTCT TCGGAAACAA ACACATCTTG TTCCCAGTCT TTGCTTTCAA
 51 AGCGATCCCT TTAGCTGCC TACACCTATC CCTAACTCTT ACCTCCTTTC
 101 GCCTTAAAAA CCATGTCTTT CGTCTAGAAG CCTTCGCTAA CGCCGTGCAC
 151 ACACGGTTTG AAGCTAGACA GATCTGGCAA GAGTCTAGAG AAGACGCTGT
 201 CTCGCTTTTA CATCTCAAGT CCCGGTATTT TGTCCTATCC TCTATCCTCT
 251 CTTGTATTGC CTCCATAACC GTCATTACAT ATACTTCCTT CTCACACCAA
 301 GGCCTAACCC CATCTCTTAA GTCGTCATGG GCTTCAGTGA AGTCCTCGTG
 351 GATGCGAGTA ACAGTAACTT CCGTCATCGT CTATGGTCTG CTCTTCCTCT
 401 ACTCTCCAAT TCCCATGTTT ATATCAGCTC TCGTTGGTTA CACACCCACG
 451 TTGCGTTATC TCATCACCAT CTTCTGTCTC GGTGTTGAGG TATACATAAT
 501 GGCTATAACA GGGATTGGCC TTGTTGTCTC TGTATTGGAA GAAAGATATG
 551 GTTTTGATGC TATTAACGAA GGAACCGCTC TGATGAAAGG GAGGAGGATA
 601 ACAGGGTTGG CTTTGGCGGG TGTATTTGTG TTTTATCGA GTTTCATTGG
 651 TCATGGGATG GAGAAGTTGG CGAAGGAGCT AGATATGGAC TTCAGCTCAG
 701 GGTCTGGTGT GTGGCTGGTG GGTGGGATGG TTGGAAGTTG A

Sequenz des *Arabidopsis thaliana* cDNA RPP5 abgeleitet aus der Gensequenz U97106

1 ATGGCGGCTT CTTCTTCTTC TGGCAGACGG AGATACGACG TTTTCCAAAG
 51 CTTCAAGTGGG GTTGATGTTT GCAAGACGTT CCTCAGCCAT CTTCTCAAGG
 101 CTCTCGACGG CAAATCAATC AATACATTCA TCGATCATGG AATCGAGAGA
 151 AGCCGCACAA TCGCCCTGA GCTTATATCG GCGATTAGAG AAGCTAGGAT
 201 CTCAATCGTC ATCTTCTCTA AGAACTATGC TTCTTCAACG TGGTGCTTAA
 251 ATGAATTGGT TGAGATCCAC AAGTGCTTTA ATGATTTAGG TCAAATGGTG
 301 ATTCCAGTTT TCTACGACGT TGATCCTTCG GAAGTTAGAA AACAGACCGG
 351 CGAATTTGGA AAGGTCCTTG AAAAGACATG CGAGGTCAGC AAGGACAAAC
 401 AACCAGGGGA TCAGAAACAA AGATGGGTGC AAGCTCTCAC AGATATAGCA
 451 AATATAGCCG GAGAGGATCT TCTGAACGGG CCTAATGAAG CGCATATGGT
 501 TGAAAAGATA TCCAATGATG TTTTGAATAA ACTTATCACT CGGTCAAAGT
 551 GTTTTGATGA CTTTCGTCGGA ATTGAAGCTC ATATTGAGGC AATAAAATCA
 601 GTATTGTGCT TGGAATCCAA GGAAGCTAGA ATGGTCGGGA TTTGGGGACA
 651 GTCAGGGATT GGTAAGAGTA CCATCGGAAG AGCTCTTTTC AGTCAACTCT
 701 CTAGCCAGTT CCACCATCGC GCTTTCCTAA CTTATAAAAAG CACCAAGTGGT
 751 AGTGACGTCT CTGGCATGAA GTTGAGTTGG CAAAAAGAGC TTCTCTCGGA
 801 AATCTTAGGT CAAAAGGACA TAAAGATAGA GCATTTTGGT GTGGTGGAGC
 851 AAAGGTTAAA TCACAAGAAA GTTCTTATCC TTCTTGATGA TGTGGATAAT
 901 CTAGAGTTTC TTAAGACCTT GGTGGGAAAA GCTGAATGGT TTGGATCTGG
 951 AAGCAGAATA ATTGTGATCA CTCAAGATAG GCAACTTCTC AAGGCTCATG
 1001 AGATTGACCT TGTATATGAG GTGAAGCTGC CATCTCAAGG TCTTGCTCTT
 1051 AAGATGATAT CCAATATGCT TTTTGGGAAA GACTCTCCAC CTGATGATTT
 1101 TAAGGAACCTA GCATTTGAAG TTGCCGAGCT TGTCGGTAGT CTTCCTTTGG
 1151 GTCTCAGTGT CTTGGGTTCA TCTTTAAAAG GAAGGGACAA AGATGAGTGG
 1201 GTGAAGATGA TGCCTAGGCT TCGAAATGAT TCAGATGATA AAATTGAGGA
 1251 AACACTAAGA GTCGGCTACG ATAGGTTAAA TAAAAAAAAT AGAGAGTTAT
 1301 TTAAGTGATC TGCATGTTTT TTCAATGGTT TTAAGTCAG TAACGTCAA
 1351 GAATTACTTG AAGATGATGT TGGGCTTACA ATGTTGGCTG AGAAGTCCCT
 1401 CATACTGATT ACACCGGGTG GATATATAGA GATGCACAAT TTGCTAGAGA
 1451 AATTGGGTAG AGAAATTGAT CGTGCAAAGT CCAAGGGTAA TCCTGGAAAA
 1501 CGTCAATTTT TGACGAATTT TGAGGATATT CGAGAAGTAT TGACCGAGAA
 1551 AACTGGGACC GAAACTCTTC TTGGAATACG TTTGCCACAC CCGGGATATC

1601 TTACGACAAG GTCGTTCTTA ATAGATGAAA AATCATTCAA AGGCATGCGT
 1651 AATCTCCAAT ATCTAGAAAT TGGTTATTGG TCAGATGGGG TTCTACCTCA
 1701 GAGCCTCGTT TATTTCCCTC GTAAACTCAA AAGGCTATGG TGGGATAAAT
 1751 GTCCATTGAA GCGTTTGCC TCTAATTTTA AGGCTGAGTA TCTGGTTGAA
 1801 CTCAGAATGG TGAATAGTAA GCTTGAGAAG CTGTGGGATG GAACTCAGCC
 1851 CCTTGGAAGT CTCAAGAAGA TGGATTTGTA TAATTCCTAC AAATTGAAAG
 1901 AAATTCAGTA TCTTCTTTA GCCATAAAC TCGAGGAATT AAATCTTGAA
 1951 GAATGCGAAT CTTTGGAGAC ACTTCCCTCC TCGATTCAGA ATGCCATTAA
 2001 ACTGAGGGAG TTAAATTGTT GGGGGGGGCT ATTAATAGAT TTAAATCAT
 2051 TAGAAGGCAT GTGTAATCTC GAATATCTAT CAGTTCCTAG TTGGTCAAGT
 2101 AGGGAATGCA CTCAGGGCAT CGTTTATTTT CCTCGTAAAC TCAAAAGTGT
 2151 ATTGTGGACT AATTGTCCAT TGAAGCGTTT GCCTTCTAAT TTTAAGGCTG
 2201 AGTATCTGGT TGAATCATA ATGGAGTACA GTGAGCTTGA GAAGCTGTGG
 2251 GATGGTACTC AGTCACTTGG AAGTCTCAAG GAGATGAATT TGAGGTATTC
 2301 CAACAATTTA AAAGAAATTC CAGATCTTTC TTTAGCCATA AACCTCGAGG
 2351 AATTAGATCT TTTTGGATGC GTATCTTTGG TGACACTTCC TTCTCGATT
 2401 CAGAATGCCA CTAAACTGAT CTATTTAGAT ATGAGTGAAT GCGAAAATCT
 2451 AGAGAGTTT CCAACCGTTT TCAACTGAA ATCTCTCGAG TACCTCGATC
 2501 TCACCTGGAT CCCGAATTTG AGAAATTTCC CAGCAATCAA AATGGGATGT
 2551 GCCTGGACTA GATTATCTCG AACAAAGATTG TTTCCGGAAG GGAGAAATGA
 2601 GATCGTGGTA GAAGATTGTT TCTGGAACAA GAATCTCCCT GCTGGACTAG
 2651 ATTATCTCGA CTGCCCTTATG AGATGTATGC CTTGTGAATT TCGCTCAGAA
 2701 CAACTCACTT TTCTCAATGT GAGCGGCTGC AAGCTTGAGA AGCTATGGGA
 2751 AGGCATCCAG TCGCTTGGAA GTCTCGAAGA GATGGATCTG TCAGAATCTG
 2801 AAAACCTGAA AGAACTCCA GATCTTTCAA AGGCCACCAA TCTGAAGCTT
 2851 TTATGTCTCA GCGGGTGCAA AAGTTTGGTG ACACTTCCTT CTACAATTGG
 2901 GAATCTTCAA AATTTGAGAC GTTTGTACAT GAACAGATGC ACAGGGCTGG
 2951 AGGTTCTTCC GACCGATGTC AACTTGTTCAT CTCTCGAAAC CCTCGATCTC
 3001 AGTGGTTGCT CAAGTTTGAG AACTTTTCCT CTGATTTCAA CTAATATTGT
 3051 ATGTCTCTAT CTGGAAAACA CCGCCATTGA AGAAATCCA GATCTTTCAA
 3101 AGCCACCAA GCTCGAGTCT TTGATACTCA ACAACTGCAA AAGTTTGGTG
 3151 ACACCTCCTT CTACAATTGG GAATCTTCAA AATTTGAGAC GTTTGTACAT
 3201 GAACAGATGC ACAGGGCTGG AGCTTCTTCC GACCGATGTC AACTTGTTCAT
 3251 CTCTCGAAAC CCTCGATCTC AGTGGTTGCT CAAGTTTGAG AACTTTTCCT
 3301 CTGATTTCAA CTAGAATCGA ATGTCTCTAT CTAGAAAACA CCGCCATTGA
 3351 AGAAGTTCCC TGCTGCATTG AGGATTTTAC GAGGCTCACT GTACTACGGA
 3401 TGTATTGTTG CCAGAGGTTG AAAAACATCT CCCCAAACAT TTTCAGACTG
 3451 ACTAGTCTTA CGCTCGCCGA CTTTACAGAC TGTAGAGGTG TCATCAAGGC
 3501 GTTGAGTGAT GCAACTGTGG TAGCGACAAT GGAAGATCAC GTTTCTTGTG
 3551 TACCATTATC TGAAAACATT GAATATACAT GTGAACGTTT CTGGGATGCG
 3601 TGTTCTGATT ATTACTCTGA TGACTTTGAG GTAAATCGGA ACCCAATTAG
 3651 ATTGTCAACG ATGACTGTCA ACGATGTGGA GTTTAAGTTT TGTGTCTCCA
 3701 TTACGATCAA AGAATGCGGT GTACGACTCT TGATGTCTA TCAAGAAACA
 3751 GAGCACAACC AACAACTAC GAGAAGCAAG AAGCGGATGC GGATGACATC
 3801 GGGGACATCT GAAGAAGATA TCAACTTACC CTATGGCCAA ATTGTAGCGG
 3851 ACACAGGATT GGCCGCTCTA AATACAGAGC TTTGTTAGG GCAGGGAGAA
 3901 GCATCATCAT CAACATCTCT AGAGGGGGAA GCTTTGTGTG TTGATGATTA
 3951 CATGATAAAT GAAGAACAAG ATGAACAAAT ACCTATCTTG TATCCTGTTT
 4001 ATGATATAGA CGATGATATG TGGAGATCAT TATTTTTTGG AGATACAGAC
 4051 GATGATATGT GGAGATCATT ATATTCTGCA GAATGA

Sequenz des *Arabidopsis thaliana* cDNA-Klons 11177 (EST 8)

1 GGAGTTGGCC ATAAGCTCGG ACAAAGGGAA ATCATGCGGG CTAAGCTAGG
 51 ATCGTAAAGG AATCAGCTGG GTCAGCTCGT CCGAAGCTCA AGCTCAGCTG
 101 GAAGGAGTGA CGGTTACTCG GTTGAGTGA CAGTCCACCT CGAGCAGCTG
 151 GGCGAGTGAC GTTGAGCTC GAGGAGCTG GTCGACTGTC TGACCAGCTG
 201 GGCGGTTACC CGGACGGTTA CTGGGCGGT TACGGTCGAA CGGGTGCTCG
 251 CCCAAAGGGT GGATAAGGGC GAGCAGTTAC GGACCGAGAG ATGCAACTCT
 301 TCGGAAAGGT GCAAGGAACG GAGGTGATCC TTCGGGAAGC ATCAATTGCG

```

351 CCCTTATATA AGGGAGGGTG CCAAGACCT TTAGAGGCAC CAGAACACAC
401 AGAAAATACC GGAAACTCAC TATTAACTC TCGTCCTTTC GGGAGAATAA
451 ATCGTCGGAG GAAATCGGTA GTCGGACTGT GACCGGCAAG GAAGAAGAAG
501 GGTCCGATCG ACCGAACATC GTGAAGGAAG GTAAGGGAAA CCTAGACCTC
551 GAGTCTAAGG ATCTTAGAAT ACGTTAATTG AGTCGAGATT CCTTGAAAGG
601 AACGACTAAT GGGCTGGGTA GAATCGGGTC ATGTGCAACC TACACAAGAC
651 TCAGTCCGGG CAAGGATCG GCTAGGGCGG GGACGCGGCT CGGTGGAGGA
701 CCGGAGGGTC TTAATTGACC TACCGAGAAT CGGCTGGACT TATGCGCGAT
751 CTTCCCCCTCT CCAAATCCTC TCCGCGGATC CGAGTAGTGA ACCAATGGTG
801 ATCGTGGGAA GGACTTGGCT TAGGCGCGAG AGACTCAAGG CCGGGATGAT
851 GGGTGGTGAT AGTTTCGGAT AGTCTAGGGA TCAATTATAT TACTTGTAAT
901 AGCTTATATG CAAACTCATA AGTTGTATCG AATCCTTTAC TTAATTAATA
951 AAGTATCTGA TTGTTTGTTT AAAAAAAAAA AAAAAAA

```

Sequenz des *Arabidopsis thaliana* cDNA-Klons J

```

1 AAGAACCCTC ACTTTTCTCT CCCACTCTTT CTTTACTAC TCTCACACAT
51 ATCTCTGTCT ATATATCACT TTACATAAAC GACTATTCCA CACACAAACA
101 CACATAGCCA TGGCCTCTTC TTTCTCTTCA CAAGCCTTCT TCTTGCTCAC
151 ATTGTCTATG GTTTTAATTC CTTTCTCTTT AGCTCAAGCT CCCATGATGG
201 CTCCTTCTGG CTCAATGTCC ATGCCGCCA TGCTAGCGG CGGTGGAAGC
251 TCGGTTCCCTC CTCCAGTGAT GTCTCCGATG CCAATGATGA CTCCACCACC
301 TATGCCATATG ACTCCACCAC CCATGCCCAT GACTCCACCA CCTATGCCTA
351 TGGCTCCACC ACCAATGCCC ATGGCTTCAC CACCAATGAT GCCAATGACT
401 CCATCTACAA GCCCAAGCCC ATTAACAGTT CCGGATATGC CTTCGCCGCC
451 GATGCCATCC GGAATGGAAT CTTACCTTC TCCAGGACCC ATGCCACCGG
501 CAATGGCGATC GCCGGATTCA GGAGCTTTCA ATGTTAGAAA CAACGTCGTA
551 ACACTTTCAT GCGTTGTTGG AGTTGTTGCA GCCCATTTTC TCCTCGTTTG
601 AAATGATTAT TGAATTGGTC AGCCTCGATC GTTTTCTTGT AATTTACCTT
651 CATTTTTTTC CCTCAAATTA TTAGTGGTCA TCATTTTATA ATATTTGAGT
701 TTGTGTTTGA TGTACGATTC AGACATTTGT TTGCATTATG TGCTTAATAA
751 GTTTATCGTT GACTCTAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
801 AAAA

```

Sequenz des *Arabidopsis thaliana* cDNA-Klons ATTS1159

```

1 CTCTTCACAA GCCTTCTTCT TGCTCACATT GTCTATGGTT TTAATTCCTT
51 TCTCTTTAGC TCAAGCTCCC ATGATGGCTC CTTCTGGCTC AATGTCCATG
101 CCGCCTATGT CTAGCGGCGG TGGAAGCTCG GTTCTCCTC CAGTGATGTC
151 TCCGATGCCA ATGATGACTC CACCACCTAT GCCTATGACT CCACCACCCA
201 TGCCCATGAC TCCACCACCT ATGCCATGAG CTCCACCACC AATGCCCATG
251 GCTTACCAC CAATGATGCC AATGACTCCA TCTACAAGCC CAAGCCCATT
301 AACAGTTCCG GATATGCCTT CGCCGCCGAT GCCATCCGGA ATGGAATCTT
351 CACCTTCTCC AGGACCCATG CCACCGGCAA TGGCGGCTTC GCCGGATCA
401 GGAGCTTTCA ATGTTAGAAA CAACGTCGTA ACACTTTCAT GCGTTGTTGG
451 AGTTGTTGCA GCCCATTTTC TCCTCGTTTG AAATGATTAT TGAATTGGTC
501 AGCCTCGATC GTTTTCTTGT AATTTACCTT CATTTTTTTC CCTCAAATTA
551 TTAGTGGTCA TCATTTTATA ATATTTGAGT TTGTGTTTGA TGTACGATTC
601 AGACATTTGT TTGCATTATG TGCTTAATAA GTTTATCGTT GACTCTACTT
651 GAAGAGAG

```

Sequenz des abgeleiteten Gens J (*Capsella rubella*)

```

1 ATGGCCTCTT CTTCTCTTCC ACAAGCCTTC CTCTTGCTTG CATGTCTAT
51 GGTTTTAGTT CTTTCTCTT TAGCTCAAGC TCCCATGATG GCTCCTTCCG
101 GCTCAATGTC CATGCCCCCA ATGCCATGCG GCGGCGGTGG AAGCATGATC
151 CCCCCTCCGG CTATGACTCC ACCACCCATG CCTATGATGA CTCCACCACC
201 TATGCCATATG ATGACTCCAC CACCTATGCC TATGATGACT CCACCACCCA
251 TGCCTATGAT GACTCCACCA CCCATGCCTA TGATGACTCC GCCACCCATG

```

301 TCCATGACTC CACCACCTAT GCCTATGATG ACTCCACCAC CTATGTCCAT
 351 GACTCCACCA CCAATGCCTA TGGCTTCACC ACCAATGATG TCAATGGGCC
 401 CAAGCCCAAG CCCATTAACA GTTCCCAGTA TGGCTTCCCC GCCGATGCCA
 451 TCGGGAATGG AATCAGCAGC ATCTCCGGGA CCCATGCCAC CGCCAATGGC
 501 GGCTCCCCCT GATTCTGGAG CTTTCAATGT TAGAAACAAC GTCGTAGCAG
 551 TGCTTGGCGT TGTGGAGTT GTTGCAGCAC ATCTTCTCCT CGTTTGA

Sequenz des *Arabidopsis thaliana* cDNA-Klons K

1 AGTTTCCATG AAGTCTTCTT CTTCTCAGAG CTACGATGTT TTCCCCAACT
 51 TCAGAGGGGA AGATGTCCGC CACTCATTAG TCAGCCACCT TCGTAAGGAA
 101 CTTGATCGCA AATTCATCAA TACATTTCAAC GACAACAGGA TCGAAAAGAAG
 151 CCGCAAAATA ACCCCAGAGC TGTATTGGC TATAGAAAAAC TCAAGGATCT
 201 CACTTGTCTG CTTCTCTAAA AACTATGCTT CTTCCACGTG GTGCTTAGAT
 251 GAATTGGTGA AGATCCAAGA GTGCTATGAG AAATTGGATC AAATGGTGAT
 301 TCCTATTTTC TACAAAGTAG ATCCTTCTCA TGTAGAAAA CAGACCCGGAC
 351 AGTTTGGCAT GGTCTTTGGG GAGACCTGCA AGGGCAGAAC AGAGAATGAG
 401 AAGCGAAAAT GGATGCGAGC TCTAGCAGAG GTAGCACATC TAGCCGGAGA
 451 AGATCTTCGG AACTGGCGTA GCGAAGCAGA AATGCTTGAA AATATCGCCA
 501 AGGATGTTTC AAACAACTC TTCCCCCAT CAAATAATTT CAGTGACTTC
 551 GTCGGGATTG AAGCTCATAT AGAGGCATTG ATATCAATGT TGCCTTCGA
 601 CTCCAAGAAA GCGCGAATGA TAGGGATTTG TGGCCCTTCT GAAACTGGTA
 651 AGACTACTAT AGGAAGAGCT TTATACAGCA GACTCAAAAAG CGACTTCCAC
 701 CATCGGGCTT TCGTAGCATA TAAGAGAAAA ATACGGAGCG ACTATGACCA
 751 GAAGCTGTAT TGGGAAGAAC AATTTCTATC AGAAATTTCTT TGTGAGAAGG
 801 ATATAAAGAT AGAGGAATGT GGTGCGGTGG AACACGACT AAAGCACACG
 851 AAGTCTCTTA TTGTTCTTGA TGATGTTGAT GATATAGAG TACTAAAAAC
 901 TTGTGGTAGA CGTATCAGAT GGTTTGGATC TGAGAGCAA ATTGTTTGA
 951 TTACTCAAAA AAGGGAACCT CTCAAGGCTC ACAACATTGC ACATGTTTAT
 1001 GAAGTGGGAT TCCCATCAGA AGAACTGGCT CATCAAATGT TTTGTGCGATA
 1051 TGCTTTTGGG AAAAACTCTC CACCTCATGG TTTCAACGAG CTTGCAGATG
 1101 AAGCTGCAA GATTGCCGGT AATCGTCCTA AAGCTCTCAA GTACGTTGGT
 1151 TCGTCTTTTA GAAGACTAGA CAAGGAACAA TGGGTGAAGA TGCTATCGGA
 1201 GTTCCGTAAT AATGGGAATA AACTAAAAAT CAGCTATGAT GAGTTAGATG
 1251 GTAAAGGTCA AGATTACGTT GCGTGTTFAA CCAATGGTTC AAATTCCCAG
 1301 GTGAAGGCAG AGTGGATCCA CTTGGCACTA GCGGTGTCTA TATTACTTAA
 1351 CATCCGCAGT GACGGAACGA CAATTTCTAA ACATTTGAGC TACAATCGAA
 1401 GCATGGCGCA ACAAGCAAAA ATATGGTGGT ATGAAAATTT GGAAAGAGTG
 1451 TGCAAGAAGT ACAATATATG TGGCATAGAT AGCTCAACTG ATGGTGGTGG
 1501 TTCAACGTAT GGACAATGTT CAAATTTCCA ATTTCAAAGA AACATGGACG
 1551 CAAGCCCTCG GGGAAACAAA ACCAGTAATC AATCCACAAA AGATTACCAG
 1601 AGAGCTTCTC AAGTAGAGAA GGAAAAGATC GAGTACTGTG AGCCACATGT
 1651 ATATATAACA CCTGCCATCT TCAGTGACGG AACCAGAGCT CCCAAATACG
 1701 TGGAGTCTAG TAGCCGAAGA GTAACCCAAG TGCACCACGC CAAAACATGG
 1751 TGGCCGGAGA ATTGTGAAAA AGTGTACGAG AACCACAACA ACATATATGG
 1801 CATAGATAGA TCAATTGACG GTGGTGATAA GTTTGAGGGA AAAAGTAAAG
 1851 TTAGTGACGG TGGGTTAGAC GGCAAAGATC AAGGTTCAAT GTATGGACAA
 1901 AGTTCAAAT CCGAGCTTCA AATAAACATG GACGCAGACA ATAGGAGATG
 1951 TGAACCTGTG AGTGAGATGC TTTTCAAAAA TTACAATGTG TGTTCGCTA
 2001 ATGGTTTAA CCGATGTAAC TGTTCAAATC CCCAGTCTCA AAGAAAATTG
 2051 GACGCAAGCC TAAAGAAAGA CAAAATTGTG CATGAGTGGA TTCGAACCGG
 2101 TAGTGGCTTC TTTTTTGATT TTCAGGGCCC AAAATCTATA GTGATCTGTG
 2151 CTCAAGTGGG CGAGAAAAT TTCGAGTACT GTGAACAAGG TGTGTATATA
 2201 ACACTTGGCA TCCTCAGTGG CCGAATTATA GTTCTCAAAC ATTTGGAGTT
 2251 CAGCCGAAGA ATGGCGCAAC AAGCAAAAAG TTGGTGGTCG GAAAATTGGA
 2301 TAAAAGTGTA CCAGGAGCAC AACATATGTG GCATAGATAA ATCATTGAT
 2351 GGTAGGTTTG ATGATAGACG TGTCATTCGT CAATTAAGAC CAACTGAAG
 2401 CACGGCTTGG CCCTTGTAAG AGTCAGGACT TGAGCTAGAT ACTTAAGCCA
 2451 AGCAGATCTG TTTGGCTAAG TCTGGTCGGG TTAGCGCGAA ATTGTGCGAA
 2501 TCTGTTTAAAT AATTGCATTC AGTCAAGTC TTTAATTTGT ACTTCATGGC

2551 TTTAATACAC TAATTAAACA TTGGATCATT TTTAGTTTTT TCTTCTTTTC
 2601 AAAATACATA TTTTTCTAAA AAAAAAAAAA AAAAAAAAAA AA

25033 - Sequenz der vorhergesagten cDNA des Gens TOM1 (*Arabidopsis thaliana*)

1 ATGCTTACTT CCTTCAAATC CTCTAGCTCC TCCTCCGAAG ATGCCACCGC
 51 TACCACCACC GAGAATCCTC CTCCTTTGTG CATCGCCTCC TCCTCGGCCG
 101 CAACCTCCGC CTCACATCAC CTCCTGCGTC TTCTTTTCAC CGCTGCGAAT
 151 TTCGTCTCCC AGTCAAACCT CACCGCCGCT CAAAACCTTAC TCTCAATCCT
 201 CTCCCTTAAC TCTTCTCCTC ACGGCGACTC CACCGAGCGA CTTGTACACC
 251 TCTTCACTAA AGCCTTGTC GTACGAATCA ACCGTCAGCA ACAAGATCAG
 301 ACGGCTGAAA CGGTTGCCAC GTGGACGACG AACGAAATGA CGATGAGTAA
 351 CTCCACGGTG TTCACGAGCA GTGTATGCAA AGAACAGTTC TTGTTTCGAA
 401 CCAAGAACAA CAATTCCTGAC TTCGAGTCTT GTTACTATCT TTGGCTAAAC
 451 CAACTAACGC CGTTTTATTG GTTTCGGTCAT TTAACGGCGA ACCAAGCTAT
 501 CCTCGACGCG ACGGAGACAA ACGATAACGG AGCTCTACAT ATACTTGATT
 551 TAGATATATC ACAAGGACTT CAATGGCCTC CATTGATGCA AGCCCTAGCA
 601 GAGAGGTCAT CAAACCCTAG CAGTCCACCT CCATCTCTCC GCATAACCGG
 651 ATGCGGTCTGA GATGTAACCG GATTAACCG AACTGGAGAC CGGTTAACCC
 701 GGTTTCGCTGA CTCTTTAGGT CTCCAATTCC AGTTTCACAC GCTAGTGATC
 751 GTAGAAGAAG ATCTCGCCGG ACTTTTGCTA CAGATCCGAT TGTAGCTCT
 801 CTCAGCCGTA CAAGGAGAGA CCATTGCCGT CAATTGTGTT CACTTCTCC
 851 ACAAATATTT TAACGACGAT GGAGATATGA TCGGTCACCT CTTGTCAGCG
 901 ATCAAGAGCT TAAACTCTAG AATCGTTACA ATGGCAGAGA GAGAAGCTAA
 951 TCATGGAGAT CACTCGTTC TGAATAGATT CTCTGAGGCA GTGGATCATT
 1001 ACATGGCGAT CTTTGATTCTG TTGGAAGCGA CGTTGCCGCC AAATAGCCGA
 1051 GAGAGACTAA CCTTAGAGCA ACGGTGGTTC GGTAAGGAGA TTTTGATGT
 1101 TGTGGCGGCG GAAGAGACGG AGAGAAAGCA AAGACATCGG AGGTTTGAGA
 1151 TTTGGGAAGA GATGATGAAG AGGTTTGGTT TCGTTAACGT TCCTATTGGA
 1201 AGCTTTGCTT TGTCTCAAG TAAGCTTCTT CTTAGACTTC ATTATCCTTC
 1251 AGAAGGTTAT AATCTTCAGT TCCTTAACAA TTCTTTGTTT CTTGGCTGGC
 1301 AAAATCGTCC CCTCTTCTCC GTTTCGTCTG GAAATGA

Sequenz des abgeleiteten Gens TOM1 (*Capsella rubella*)

1 ATGCTTACTT CTTTCAAATC CTCTAGCTCC TCCTCCGAAG ATGCCACCAC
 51 CGAGACTCCT CCTCCTCCTT TTTGTCAGCT TGCCTCCTCC TCCTCATCCT
 101 CCGCCGCTAC CTCGCCCTCC CATCACCTCC GTCGTCTACT TTTCAACCGT
 151 GCTGATTTCTG TCTCCCAGTC TAACTTCACC GCCGCTCGGA ACTTACTCTC
 201 AATCCTCTCC CTCAACTCTT CTCCTTACGG CGACTCCACC CAGCGACTCG
 251 CCCATCTCTT CACGAAAGCC TTGTCCCTTAC GGATCAACCG TCTACAACAG
 301 GAACAAGATC CAACGGTTGC CACGTGTACG ACGAACGAAA TGAATATGAG
 351 TACCAACTCC ACGGTGTTCA CGAGCAGTGT GTGCAAAGAA CAGTCTTGT
 401 TTCGGACCAA GAACAACAAC TCTGACTTCG AGTCTTGTTA CTATCTTTGG
 451 CTGAATCAAC TAACGCCGTT TATTCGGTTC GGTCATTTAA CGGCGAACCA
 501 GCGGATCCTC GACGCGACCG AGACAAACGA TAATGGAGCT CTACATATAC
 551 TTGATTTAGA TATATCACA GGACTTCAGT GGCTCCATT GATGCAAGCC
 601 CTAGCAGAGA GGTCAACAAG CCTAACAGT CCACCACCAT CTCTCCGAT
 651 AACCGGATGC GGTGAGATG TAACCGGATT AAACCGAACC GGAGATAGGT
 701 TAACCCGGTT TGCTAACTCT TTAGGTCTCC AGTTCAGTT TCACAAGCTG
 751 GTGATCGTAG ACGAAGATCT CCTGGGCTC TTGCTACAGA TCCGATTGTT
 801 AGCTCTCTCC GCCGTACAAG GAGAGACCAT TGCCGTCAAC TGTGTTCACT
 851 TCCTCTACAA GTTCTTTAAC GACGATGCAG ATTTGATCGG TCACTTCTTA
 901 ACAGTGATCA AGAGCTTAAA CCTTAGAATC GTGACAATGG CTGAGAGAGA
 951 AGCTAATCAT GGAGATCACT CGTTCCTGAT TAGATTCTCA GAGGCTTTGG
 1001 ATCATTACGT GCGGATATTT GATTCCTTGG AAGCGACTTT GCCGCCAAAT
 1051 AGCCGAGAGA GACTAACCCT AGAGCAACGG TGGTTCGGTA AGGAGATTAT
 1101 GGATGTCGTG GCGGCGGAAG CGACGGAGAG GAAGCAAAGA CATCGGAGGT
 1151 TTGAGATATG GGGAGAGATG ATGAAGAGAT TTGGTTTCGT TAACGTTCTT
 1201 ATTGGAAGCT TTGCTTTGTC TCAAGCTAAG CTTCCTCTTA GGCTCCATTA

1251 **TCCTTCAGAA GGTATAATC TTCAGTTCCT TAACGATTCT TTGTTTCTTG**
 1301 **GTTGGCAAAA TCGTCTACTC TTCTCCGTTT CATCGTGGAA ATGA**

Sequenz des *Arabidopsis thaliana* cDNA-Klons 5787

1 **GGCTTTAACA GAGCTGAGTA TCAGGAGTTG CTTTCACGTA ACAGATTCTT**
 51 **CGATAGAGTC CCTGGCTACA TGGGAAAGAC AAGCAGAGGG AGGAAGCAAA**
 101 **CAATTGAGGA AGCTCAATGT CCATAACTGT GTGAGCTTAA CTACTIONGAGC**
 151 **ACTGAGATGG TTAAGCAAGC CGTCTTTTGC GGGTCTGCAT TGGCTCGGAA**
 201 **TGGGACAAAC ACGATTCGCG GGTCGAAAAG AGACGGTTAC AGCTATGATT**
 251 **TGTGGACAGA GGCCGTGGCT CACGTTGTGC TTCGATGGAT GTGAGCTTGG**
 301 **ATGCTCTGAT GGTGGGAAT TCCACACGCC ACAACGCCAC TGAAGAATC**
 351 **ATTGTCTCAC TAGTTCATGT ATGTATATAA ATCGTTCAGA AACATGATT**
 401 **AGTAGGATGT GTAATTTTAT TTAAACCTTT TAAAAAATTT GATTGTAAAC**
 451 **TTGTAAAAAGT TAAACTAATT CCAAGAAATG TATTATTATC TTTACTTGCA**
 501 **AAAAAAAAA AAAAAAAAAA A**

Sequenz des *Arabidopsis thaliana* cDNA-Klons 21484

1 **TCTGTGTTAG CAGAAGAAAT ACTGAAGCGT TTAGATCTGG AGAATCTCTG**
 51 **TTCCGTCGCA TCGTTCCTCA CCACTCTCCG CTCCGCCGTT GTTTCCGGCG**
 101 **TTCTTCCTTC TCTTACCTCC CTCGATCTCT CTGTTTTCTC TCCGGACGAT**
 151 **GAAACTCTGA ATCACGTGCT GCGTGGTTGT ATCGGACTCA GCAGCTTAAC**
 201 **TCTGAAATGT CTCCGCCTTA ACGCTGCTTC TGTGCGTGGG GTTCTCGGGC**
 251 **CACATCTCCG AGAACTTCCAC TTGCTCAGGT GTTCACTTTT ATCTCCACC**
 301 **GTCTCACTT ACATTGGCAC CCTCTGCCCA AATCTCCGGG TTCTTACTCT**
 351 **GGAAATGGCA GACCTGGATT CTCCCGATGT TTTTCAGTCT AACTTGACGC**
 401 **AGATGCTTAA CGGGTGTCTT TATCTGGAGT CTCTGCAACT TAATATCCGA**
 451 **GGTATCTTAG TAGATGCCAC CGTTTCCAG TCTGTTAGAT TCTCCTTGCC**
 501 **AGAAACTCTC AAAGCTCTAA GGTACAGCC ACTTCTTGAG TCTGAGGCAA**
 551 **TTCTTCTCAT GAACAGATTC AAAGTGACTG GAACTTATTT ATCTCAGCCT**
 601 **GATTATAATA GTGCCCTGCT ATCTCCTTCA CCTTCTTTTA CTTTGCAAAG**
 651 **CCTGTCTCTT GTCCTGGACT TGATATCTGA CAGGCTTATC ATAGCTATCA**
 701 **CAGGTTCTCT TCCTCAACTT GTCAAACCTGG ACCTTGAAGA TCGTCTGAA**
 751 **AAAGAGCCAT TTCCCGACAA TGACTTGACT TACTACTGGG TTCAGGCTCT**
 801 **CGTTTTTTGT CAGCAACTCA CAAGCCTCTC TCTGGTTCGA ACTTGTTACA**
 851 **ACCGCAAGAT ATCATTCAAG AGAATAAACG ATATGGGTAT ATTTCTTCTC**
 901 **TCCGAGGCTT GCAAGGGTTT AGAGTCTGTG AGACTCGGTG GGTTCCTAAA**
 951 **AGTCAGCGAC GCTGGTTTTG CGTCGCTCCT TCACTCATGT CGGAACTTGA**
 1001 **AAAAGTTTGA AGTTCGAGGT GCCTTCTGTG TGTCTGATTT GGCATTTTAT**
 1051 **GATGTCACAG GGTCTTCTTG TTCTCTTCAA GAGGTCAGAC TCTCGACGTG**
 1101 **CCCTCTCATA ACAAGTGAAG CTGTGAAGAA ACTGGGGTTG TGTGGCAATC**
 1151 **TTGAGGTGCT CACTTGGGCG AGCTGCAAAA GTATATCAGA TTCTTGCTCT**
 1201 **AACTCTGTTT CAGCCCTCAG AAAGTTAACT TCTCTGAATC TCGCTGGAGC**
 1251 **TGATGTAACC GATAGTGGCA TGCTTGCACT TGGTAAGTCA GATGTTCCCA**
 1301 **TCACGCAACT GTCTCTCCGC GGCTGTAGGA GAGTCAGTGA CAGAGGAATA**
 1351 **TCCTATCTGT TGAACAATGA AGGAACAATC AGCAAAACAT TGTCAACACT**
 1401 **TGATCTCGGT CATATGCCAG GAATCTCAGA CAGAGCCATC CACACAATCA**
 1451 **CACACTGTTG CAAGGCTTTA ACAGAGCTGA GTATCAGGAG TTGCTTTTAC**
 1501 **GTAACAGATT CTTGATAGA GTCCCTGGCT ACATGGGAAA GACAAGCAGA**
 1551 **GGGAGGAAGC AAACAATTGA GGAAGCTCAA TGTCATAAC TGTGTGAGCT**
 1601 **TAACTACTGG AGCACTGAGA TGGTTAAGCA AGCCGTCTTT TCGGGTCTG**
 1651 **CATTGGCTCG GAATGGGACA AACACGATTC CGGGTTCGAA AAGAGCAGGT**
 1701 **TACAGCTATG ATTTGTGGAC AGAGGCCGTG GCTCACGTTG TGCTTCGATG**
 1751 **GATGTGAGCT TGGATGCTCT GATGGTTGGG AATTCCACAC GCCACAACGC**
 1801 **CACTGAAGAA CTCATTGTCT CACTAGTTCA TGTATTGTAT AAAATCGTTC**
 1851 **AGAAACATGA TTCAGTAGGA TGTGTAATTT TATTTAAACC TTTTAAAAA**
 1901 **T**

25032 - Sequenz der vorhergesagten cDNA des Gens TOM2 (*Arabidopsis thaliana*)

```

1  ATGGAGGAAG TGACGAGATC TGTGTTAGCA GAAGAAATAC TGAAGCGTTT
51  AGATCTGGAG AATCTCTGTT CCGTCGCATG CGTTTCCACC ACTCTCCGCT
101 CCGCCGTTGT TTCCGGCGTT CTTCCCTTCTC TTACCTCCCT CGATCTCTCT
151 GTTTTCTCTC CGGACGATGA AACTCTGAAT CACGTGTGTC GTGGTTGTAT
201 CGGACTCAGC AGCTTAACTC TGAATTGTCT CCGCCTTAAC GCTGCTTCTG
251 TGCGTGGAGT TCTCGGGCCA CATCTCCGAG AACTTCACTT GCTCAGGTGT
301 TCACTTTTAT CCTCCACCGT CCTCACTTAC ATTGGCACCC TCTGCCAAA
351 TCTCCGGGTT CTTACTCTGG AAATGGCAGA CCTGGATTCT CCCGATGTTT
401 TTCAGTCTAA CTTGACGCAG ATGCTTAAACG GGTGTCCTTA TCTGGAGTCT
451 CTGCAACTTA ATATCCGAGG TATCCTAGTA GATGCCACCG CTTTCCAGTC
501 TGTTAGATTC TCCTTGCCAG AAACTCTCAA AGCTCTAAGG TTACAGCCAC
551 TTCTTGAGTC TGAGGCAATT CTTCTCATGA ACAGATTCAA AGTGACTGGA
601 ACTTATTTAT CTCAGCCTGA TTATAATAGT GCCTTGCTAT CTCCTTCACC
651 TTCTTTTACT TTGCAAAGCC TGTCTCTTGT CCTGGACTTG ATATCTGACA
701 GGCTTATCAT AGCTATCACA GGTTCTCTTC CTCAACTTGT CAAACTGGAC
751 CTTGAAGATC GTCCTGAAAA AGAGCCATTT CCCGACAATG ACTTGACTTA
801 CACTGGGCTT CAGGCTCTCG GTTTTTGTCA GCAACTCACA AGCCTCTCTC
851 TGGTTCGAAC TTGTTACAAC CGCAAGATAT CATTCAAGAG AATAAACGAT
901 ATGGGTATAT TTCTTCTCTC CGAGGCTTGC AAGGGTTTAG AGTCTGTGAG
951 ACTCGTGGG TTCCCAAAAG TCAGCGACGC TGGTTTTGCG TCGCTCCTTC
1001 ACTCATGTCG GAACTTGAAA AAGTTTGAAG TTCGAGGTGC CTTCTGTG
1051 TCTGATTTGG CATTTCATGA TGTACAGGG TCTTCTGTT CTCTTCAAGA
1101 GGTACAGACTC TCGACGTGCC CTCTCATAAC AAGTGAAGCT GTGAAGAAAC
1151 TGGGGTTGTG TGGCAATCTT GAGGTGCTCG ACTTGGGCAG CTGCAAAAAGT
1201 ATATCAGATT CTTGCCCTCA CTCTGTTTCA GCCCTCAGAA AGTTAACTTC
1251 TCTGAATCTC GCTGGAGCTG ATGTAACCGA TAGTGGCATG CTGCACTTG
1301 GTAAGTCAGA TGTTCCCATC ACGCAACTGT CTCTCCGCGG CTGTAGGAGA
1351 GTCAGTGACA GAGGAATATC CTATCTGTTG AACAAATGAAG GAACAATCAG
1401 CAAAACATTG TCAACACTTG ATCTCGGTCA TATGCCAGGA ATCTCAGACA
1451 GAGCCATCCA CACAATCACA CACTGTTGCA AGGCTTTAAC AGAGCTGAGT
1501 ATCAGGAGTT GCTTTCACGT AACAGATTCT TCGATAGAGT CCCTGGCTAC
1551 ATGGGAAAGA CAAGCAGAGG GAGGAAGCAA ACAATTGAGG AAGCTCAATG
1601 TCCATAACTG TGTGAGCTTA ACTACTGGAG CACTGAGATG GTTAAGCAAG
1651 CCGTCTTTTG CGGGTCTGCA TTGGCTCGGA ATGGGACAAA CACGATTCGC
1701 GGGTCGAAAA GAGACGGTTA CAGCTATGAT TTGTGGACAG AGGCCGTGGC
1751 TCACGTTGTG CTTGATGGA TGTGAGCTTG GATGCTCTGA TGGTTGGGAA
1801 TTCCACACGC CACAACGCCA CTGA

```

Sequenz des abgeleiteten Gens TOM2 (*Capsella rubella*)

```

1  ATGGAGGAAG TGACGAGATC TGTGTTAGCG GAGGAGATAC TGAAGCGTTT
51  AGATCTGGAG AATCTCTGTT CCGTCGCCTG TGTTCACC ACCCTCCGCT
101 CCGCCGTTGT CTCCGGCGTT CTTCCCTTCTC TTACGTCTCT CGATCTCTCT
151 GTGAGTTTCA GATTCCTCTGA GATTTTGGAG ATTTTCTCTG GAAAGAGAGT
201 AATTTTATCG AATCTCTGTT AGGTTTTCTC TCCGGACGAG GAGACTCTGA
251 ATCACGTGCT GCGTGGTTGC ATCGGACTTA GTAGCTTAAC TCTGAATGT
301 CTCCGTTTG ACGCTAATTC TGTTCGTGGG GTTCTCGGCC CACATCTCCG
351 GGAAC TTCAC TTGCTCAGGT GTTCTCTTCT ATCCTCCACC GTCCTCACTT
401 CCATCGGTAC CCTCTGCCCA AATCTCCGGT ACTATTTCTT CTTCCCATGA
451 ACTTTTTAAT AGATACCTCA ATGAATTCTC GTGCTTTTGG CCCATGTTT
501 TTTCTGAAA AGGGAGAAGA TTTACGAAAA GTCTCATTTT TTCTGAATTT
551 ATCACTTCAT TGATAGAATT GAATTAAATT GAGTCCCTCTG TCACTCCCT
601 TGAAGGCAAG CCTTTACATA TCTGGGGTTA GCTATGACTT TTGTCAATCA
651 CTCTTCTTGT ATTAGAAGCA TTGAGTTTGT AGCTCTAGAG TTGAAGTTGT
701 GATTGTGACT TTGTGAGTAC TTAAGAGTTT TTTTGGTTT AATATTTTAT
751 TACAGGGTTC TTACCCCTGGA AATGTCAGAC CTGGATTCTC CCGCTGTTTT
801 CCAGTCTAAT TTGACGCAGA TGCTTAAACGG GTGCTCTTAT CTGGAGGTAA
851 TTTGTTAAAA GCTTGTCTGAG GCATTTGGTCA GAAGAAGCGA GAAAGTGTGA
901 AATTTGAATTT GGGTTTTATA TTTCTGTGAT GCTTGCAGTC TCTGCAACTT
951 AATATCCGAG GTATCTAGT AGATGCCACA GCTTTCAGT CTGTTAGATT

```

1001 CTCCTTGCCA GAAACTCTCA AAGCTCTCAG GTTACAGCCA CTGCTTGAGT
1051 CCGAAGCAAT TCTTCTCATG AACAGATTCA AAGTCACTGG AACTTGTTTA
1101 TCCTCAGGCTG ATTATAGTGC CTTGCTTTCT CTTTCACCTT CTTTTACTTT
1151 GCAAAGCCTG TCTCTTGTGC TGGACTTGAT ATCTGACAGG CTAATCATAG
1201 CTATCACAGG TTCTCTTCC TCACTTGTC AACTGGACCT TGAAGACCGT
1251 CCTGAAAAAG AGCCATTTC CGACAGTGAC CTGACTTACA CTGGGCTTCA
1301 AGCTCTGGGT TATTGTCAGC AACTCACAAG CCTCTCTCTG TTCGAACTT
1351 GTTACAACCG CAAGATATCA TTCAAGAGAA TAAACGATAT GGGTATATTT
1401 CTTCTCTCCG AGGCTTGCAA GGGTTTAGAG TCAGTGAGAC TCGGTGGGTT
1451 CCAAAAAGTT AGCGACGCTG GTTTTGCATC GCTCCTGCAC TCATGTCGGA
1501 ACTTGAAAAA GTTTGAAGTA CGAGGTGCCT TCCTCTTGTC TGATTTGGCA
1551 TTCCATGATG TCACAGGGTC TTCTTGTTTT CTTCAAGAGG TCAGACTCTC
1601 GACTTGTCCT CTCATAACCA GCGAAGCTGT GAAGAACTG GGGTTGTGTG
1651 GCAATCTTGA GGTGCTCGAC TTGGGCAGCT GCAAAAAGTAT ATCAGATTCT
1701 TGCCCTCAACT CTGTTTCAGT CCTTAGAAGG TTAACCTTCTC TGAATCTCGC
1751 AGGAGCTGAT GTAACCGATA GTGGCATGCT TGCACCTGGT AAGTCAGATG
1801 TTCCATACAC GCAACTGTCT CTCCGCGGCT GTAAGAGAGT CAGTGATAGG
1851 GGAATATCCC ATCTGTTGAT CAATGAAGGA ACAATCAGCA AAACATTATC
1901 AACACTTGAT CTTGGTCACA TGCCAGGAAT GTCAGACAGA GCCATCCACA
1951 CAATACGCGG TTATTGCAAG GCTTTAACAG AGCTAAGTAT CAGGAGTTGC
2001 TTTTACGTAA CAGATTCTTC GATAGAGTCG CTGGCTACAA GGGAAAGACA
2051 AGTAGAGGGA GGAAGCAAAC AGTTGAGGAA GCTCAATATC CATAACTGTG
2101 TGAGCTTAAC GACTGGAGCA CTGAGATGGC TGAGCAAGCC GTCTTTTGGC
2151 GGTCCTGCAT GGCTCGGTTT GGGACAAACA CGAATTGCTG GTCGAAAAGA
2201 GACGGTCACA GCTACGATT GTGGGCAGAG GCCGTGGCTC ACATTGTGCT
2251 TCGATGGGTG TGAGCTTGA TGCTACGATG GTTGGGAATT CCACACGCCA
2301 CAGCGCCACT GA

25031 - Sequenz der vorhergesagten cDNA des Gens TOM3 (*Arabidopsis thaliana*)

1 ATGAGTGATT TTGATGAAAA CTTTCATCGAA ATGACGTCGT ATTGGGCTCC
51 ACCATCCAGT CCTAGCCCAA GAACGATATT GGCAATGCTG GAGCAAACCG
101 ACAATGGTCT GAATCCAATC AGTGAGATCT TCCCTCAAGA AAGCTTGCCA
151 AGAGATCATA CTGATCAATC TGGACAAAAGA TCTGGTCTTC GTGAGAGACT
201 GGCTGCAAGA GTAGGATTC AATCTTCCAAC ACTCAATACA GAAGAAAACA
251 TGAGTCCTTT GGATGCATTT TTCAGGAGCT CGAATGTTCC TAATTCTCCT
301 GTCGTTGCAA TCTCTCCAGG ATTCAGTCCA TCAGCACTAT TGCATACTCC
351 CAATATGGTC AGTGATTCTT CCCAGATTAT CCCTCCGTCT TCAGCCACCA
401 ATTACGGACC TCTAGAGATG GTGGAAACTT CCGGTGAAGA CAATGCAGCG
451 ATGATGATGT TCAACAACGA TCTTCTTAT CAGCCGTACA ATGTTGATCT
501 GCCTTCTCTA GAAGTCTTTG ATGATATTGC AACGGAAAGAG TCCTTTTATA
551 TCCCATCTTA TGAACCTCAT GTTGACCCAA TTGGAACCTC TTTAGTCACA
601 TCCTTTGAAT CTGAACTCGT TGACGATGCC CATAACGACA TCATCTCCAT
651 TGAGGACAGT GAGAGCGAGG ATGGAAACA AGATGATGAC GACGAGGACT
701 TCCAATACGA AGACGAAGAC GAAGACCAAT ACGACCAAGA TCAAGATGTA
751 GATGAAGATG AAGAGGAAGA AAAAGATGAA GACAATGTTG CATTAGATGA
801 TCCTCAACCT CCACCTCAA AGAGAAGGAG ATATGAGGTA TCAAAATGA
851 TTGGAGCCAC AAGAACAAGC AAGACACAAA GGATCATACT TCAGATGGAA
901 AGCGACGAAG ACAATCCTAA CGATGGTTAT CGCTGGAGAA AATACGGTCA
951 AAAAGTCGTC AAAGGAAATC CTAATCCGAG GAGTACTTC AAGTGCACA
1001 ACATCGAGTG CAGAGTGAAG AAACATGTGG AGAGAGGAGC AGACAATATC
1051 AAGTTGGTTG TGAATACATA CGATGGGATA CACAACCATC CTTACCACC
1101 TGCACGTAGA AGCAATTCCA GTTCAAGGAA CCGGTCTGCA GGGGCAACAA
1151 TACCTCAAAA TCAGAATGAT CGAACCAGTC GGTAGGTTAG GGCTCCTCCT
1201 ACTCCTACTC CTCCTACTCC TCCTCCTTCG TCTTACACAC CTGAGGAGAT
1251 GAGGCCTTTC TCTTCGTTGG CTACAGAAAT TGATCTGACA GAGGTTTATA
1301 TGACCGGAAT CTCTATGCTG CCGAATATAC CCGTTTACGA GAATTCGGGT
1351 TTTATGTACC AGAATGATGA ACCGACGATG AATGCGATGC CCGATGGTTC
1401 AGATGTGTAC GATGGGATCA TGGAACGCCT GTATTTTAAAG TTTGGTGTCTG
1451 ACATGTAG

Sequenz des abgeleiteten Gens TOM3 (*Capsella rubella*)

1 ATGAGTGATT TTGATGGAAA CTTCACGAA TCGACGTCGT GTTGGGCACC
 51 ACCTTTCAGC CCTAGCCCAA GAGCGATATT GGCGATGATT GATCAAGCCG
 101 AAGACGATGG TCTGGATCTA ATCTCTGAGA TGTTCCCTCA TACCAACTTA
 151 CCAAGAGACC ATAGTGCTCA GTCCCTCTGAA CAAAGATCTG GTCTCGGCCA
 201 GAGATTAGCT GCAAGAGTAG GGTTCAGTCC TCCACCACTC GACACAGAAA
 251 ACATAAGTCC TTCGGCTGCA TCTTTAAGGA GCTCGACGAC TGTTCCCTCT
 301 CCTGTCTCTG CAATCCCCTC GGGATTTATC CCATCAACGC TATTGCAATC
 351 TCCCAATATG GTCACGTATT CTTACAGGTT ACGTAGCAAT GTCTTTCGTG
 401 TTTAATCATA TGGTTATAAA TGAATTTATT TATCAAAAAT ATATAAGTTA
 451 AACGAATCTG TCATTGAACC AAATATTTTCG ACGTATGTTA TTTCAATTCT
 501 TTTACGTGTA GCATAGACTA AAAAAATGGTT TTATATACAT CTATTGGTAA
 551 ACACGAAATC TATTATTTCC TTTTGTTTTAC GTTTTGAAAA TTTCCATACC
 601 CCATTTGCGT TTACTCCAAC TTACGAATGA ATGTTTGGTT CATGAATCTG
 651 GTTGTCTCGG ACAGATTTGT TTTGTTTATGA TTGACTGTAT TTTTCTCGGA
 701 TGGATTTTGA TACCAGCTAT ATATATATTT AATTACATTT CATATAATTC
 751 ATTTTAATTT TCGTTTATAT ATATATATAT ATATATATAT ATATATATAT
 801 ACATAATATA CATAAAATTA TTGTTAAAAC CATCGAGATC TTGGTCTCAT
 851 TCTTAAGCCG AAACGTGAAT ATCTTTTTGT AACCTTCTTA TTTCCAGATG
 901 ATCCCAGCTC CGTCTTCAGA CACCAATTAC GGACCTCAAG GTATGGTGGG
 951 AACTTCCGGC GAAGGCAATG CACCAACGTT GATATTCAAC AACGATCACC
 1001 CTTATCAGCC AGTGAATGCT GATCTGCCCTC TTCTTGATGG TAACGGAGTT
 1051 GCTTGTGTCA CAAGTTTGTCT TCCATTTCTC TATAATTTCT CTATCTTTTA
 1101 ATTTTATCAT AATTTTCTTT ACAGAATGTG AGGCAGTCTT TGATGATGTT
 1151 CCAACGGAAG AGTCCGTTTA TCTCCCAACT TATGAAACTG ATCATGTTGA
 1201 GCCAATTGGG GATTCCTTAG TCCCTAATTC CTTTGAACAT GAATTCGTTG
 1251 GAGATGCCAA TGTCAACTTC ATCTCCAATG AGGGCACTGA GAGCGTGAAG
 1301 GACAGCGAGA GCGATGATGA GTTGGACTTC GACTTTGAAT TCGAAGATGA
 1351 AGATGAAGCT GAAGCTGAAG CTGAAGATAA TGATGAAGAT AAAGCTGAAG
 1401 ATGAAGATGA AGATAATGAT GAAGATGGAA ATGAAGATGA AGATGAAGAT
 1451 GATGAGGAAG ATGAAAATGA TGAGGAAGAG GCTGATCAGC ACGACGAAGC
 1501 TGACTTTGAT GATGCTGAAC CTTCATCTCC AAAGAGAAGG TCCAAACAAA
 1551 AGCTAAAGAT CATAAATATC AACAAATAA CTTAAAACAA CATATATATA
 1601 TCTTTTATAA GCTAAACTAT ATATATCTTG CAGGAAGTTT GAGGCATCAT
 1651 CAAACATGAT CGGAGCAACG AGAACAAACA AGGCACAAAG GGTCACTACTT
 1701 CAGATGGAAA GCGATGAAGA TAATCCTGAA GATGGCTTTC GCTGGAGAAA
 1751 ATACGGACAG AAAGTTGTCA AAGGAAATCC TAATCCGAGG TTGTTATTTT
 1801 CCCCCACCCA CCCATCTTCT TTATTTCTCT CTAAGTAGTCA CTTTGGTTTT
 1851 GATGTTCTGT CCGTAGCTGC TTTAATATAT ATATATATGC TTCTTAATTA
 1901 ATACGTTTTT TTGTGTGTGT GGGTATATAT GTTATGAATC GCAGGAGTTA
 1951 CTTCAAATGC ACAACAACG ACTGCAACGT GAAGAAACAT GTGGAGAGAG
 2001 GAGCAGACAA TTTTAAGATA CTTGTGACAT CCTACGATGG GATACACAAC
 2051 CATCCTCCCC CACCTGCACG CTGTAGAATC AATCCGGTC CAAGGAATCG
 2101 GTCCGGTACA ACAACAACGA CTCAAAATCA GAGCTATCGA ACCGACCGGT
 2151 TAGGTAGGTT TCCTGCTCCT TCGTCTGTGA TCACACCTAT GGAGATGATG
 2201 CCTTTACTTT CCTTTACTCC ACCACTTGAT ATGTCTAAGG TTTACATGAC
 2251 CGGACTTAAT AAGCTGCCGA ATTTTTTCAGT TTTCCAGAAT CCCGGTTACA
 2301 TGTACCGGAA TGATGAACCG ATGATGAATG TTGTACCGAA TGGTCTGGG
 2351 ATTTACGGTG GGATCAGGCA TCGCCTATTT GCTGAGTTTG GTGTTAACTT
 2401 GATGTAA

25030 - Sequenz der vorhergesagten cDNA des Gens TOM4 (*Arabidopsis thaliana*)

1 ATGAAGCAGA GATGGCTGTT AGTGTGATA CTCTGTTTCT TTACGACTTC
 51 TCTTGTGATG GGCATTCATG GAAAACACCT AATCAACGAT GACTTCAACG
 101 AAACGTCTCT GTTATTGGCG TTCAAGCAA AACTCTGTAAA ATCTGATCCT
 151 AATAACGTTT TTGGTAACTG GAAATACGAG TCGGGTCGTG GTTCATGCTC
 201 TTGGCGAGGG GTTCTTTGCT CTGACGATGG TCGAATCGTC GGATTAGATC
 251 TCCGAAACAG TGGACTCACC GGAACCCATA ACCTAGTTAA CCTCACGGCC

301 TTGCCAATC TCCAGAATCT TTA CTTGCAA GGAAATTACT TCTCCTCTGG
351 AGGAGATTCC TCTGGTTCG ATTGTTATCT TCAAGTTCTG GATTTGTCTT
401 CGAATCTAT TTCAGACTAT TCAATGGTGG ATTACGTTTT CTCAAAATGT
451 TCGAATCTGG TTTCGGTGAA TATCTCCAAC AACAACTCG TCGGAAAATT
501 AGGTTTTGCT CCGTCGTCTT TACAGAGCTT GACGACTGTT GATCTCTCTT
551 ACAACATCTT ATCAGATAAG ATTCCGGAGA GTTTTATTTT AGATTTTCCG
601 GCGTCTGTA AATATCTCGA TCTAACTCAC AACAACTTAT CCGGCGACTT
651 CTCCGATCTC AGTTTCGGTA TCTGTGGGAA TCTCACCTTC TTCAGTCTAT
701 CACAGAACAA TCTCTCCGGT GATAAAATCC CAATCACTCT ACCAAACTGC
751 AAATTCCTCG AGACATTAAA CATCTCTCGG AACAACTCTG CCGGAAAAAT
801 CCCTAACGGA GAGTATTGGG GAAGTTTCCA GAATCTGAAA CAGCTCTCTC
851 TAGCTCATAA CCGTCTCTCC GCGGAAATCC CACCGGAGCT TTCTCTACTC
901 TGCAAAACCC TAGTGATTCT TGATCTATCC GGAAACACTT TCTCCGGCGA
951 GCTTCCTTCA CAGTTCACCG CCTGCGTCTG GTTACAGAAT CTTAACCTCG
1001 GAAACAACATA CCTCTCCGGA GATTTCTTAA ACACGGTCTG GAGTAAAATC
1051 ACCGGAATCA CTTATCTATA CGTTGCTTAC AACAACTCTT CAGGCTCTGT
1101 TCCGATTTCT CTCACAAACT GTTCAAATCT TCGTGTCTT GATCTAAGCT
1151 CAAATGGTTT CACCGGAAAT GTACCGTCTG GTTCTGCTC TCTGAAAAGC
1201 TCGCCGGTTT TTGAAAAGAT TCTCATAGCT AACAAATACC TCTCAGGAAC
1251 AGTTCCATG GAGCTTGTA AATGCAAGAG CTTGAAGACA ATTGATCTTA
1301 GCTTCAATGA GCTTACTGGT CCGATTCCGA AAGAGATATG GATGTTGCCG
1351 AATCTGTCGG ATTTGGTTAT GTGGGCGAAC AATCTCACTG GGACGATCCC
1401 GGAAGGTGTT TGTGTTAAAG GAGGAAATTT GGAAACTCTT ATCTCAACA
1451 ACAATCTCTT AACCGGTTCA ATCCCTGAGT CGATCTCGAG ATGCACCAAT
1501 ATGATCTGGA TCTCTCTTTC TAGTAATCGT CTCACCGGGA AAATCCCGAG
1551 CGGGATTGGT AATCTTTCCA AGTTAGCAAT CTTGCAGCTA GGGAATAACT
1601 CTCTGTCCGG GAACGTTCCCT CGCCAGCTTG GGAAGTCAA GAGTCTGATC
1651 TGGCTTGATC TGAACAGCAA CAATCTAACC GGGGACCTCC CGGGTGAGCT
1701 AGCTAGCCAA GCCGGGTTAG TAATGCCTGG GAGCGTTTCA GGTAAAACAGT
1751 TTGCGTTTGT GAGAAACGAA GGTGGAACAG ACTGCAGAGG TGCAGGTGGG
1801 TTAGTAGAGT TTGAAGGCAT TCGTGCTGAA CGATTAGAGC GGCTACCAAT
1851 GTTTCATTCCG TGTCCC CGA CACGGATATA CTCAGGCATG ACAATGTACA
1901 CATTCTCAGC AAATGGAAGC ATGATCTATT TCGACATCTC GTATAATGCT
1951 GTTTCAGGTT TTATACCTCC TGGTTATGGT AACATGGGCT ATCTCCAGGT
2001 CTTGAATTTG GGACATAACC GGATAACCGG AACCATCCCG GACAGTTTTG
2051 GAGGATTGAA AGCGATTGGT GTTCTCGATC TATCTCACA CAATCTTCAA
2101 GGGTACTTAC CCGGTCGCT GGGGTCGCTT TCTTCTCTCA GTGATCTTGA
2151 TGCTCTAACC AACCAACCTA CTGGTCCAAT CCCATTTGGA GGTGAGCTTA
2201 CAACGTTCCC TGTCTCGAGA TACGCAACA ACTCAGGCTT CTGTGGTGTT
2251 CCTTTGCGTC CCTGCGGTTT AGCTCCTCGA CGGCCATTA CCTCCCCTAT
2301 CCATGCCAAG AAGCAAACTG TTGCAACCGC TGTGATCGCT GGAATTGCGT
2351 TTCTTTTCAT GTGCTTTGTG ATGCTAGTCA TGGCGCTTTA CAGGGTGAGG
2401 AAAGTTCA GAAGGAACA GAAGAGGAG AAATACATTG AGAGCTTCC
2451 AACTTCCGGA AGCTGCAGCT GGAAGCTCTC TAGCGTTCCCT GAACCGCTTA
2501 GCATCAACGT TGTACATTC GAGAAAACCGC TGAGAAAACT CACTTTTCGCT
2551 CATCTTCTTG AAGCTACAAA CGGGTTTAGC GCAGAGACTA TGGTCCGGTC
2601 TGGTGGGTTT GGAGAAGTAT ACAAGGCCCA ACTCAGAGAC GGATCTGTTG
2651 TAGCGATCAA GAAGTTGATT CGAATCACGG GACAAGCGCA CAGAGAGTTC
2701 ATGGCTGAGA TGGAAACAAT CGGAAAAATC AAACACAGAA ACCTTGTTCC
2751 ACTTTTGGGA TATTGCAAGG TTGGTGAAGA GAGGCTTCTT GTCTACGAAT
2801 ACATGAAATG GGGAAAGCTTA GAAACCGTTC TTCACGAGAA ATCATCGAAG
2851 AAAGGCGGAA TATATCTGAA TTGGGCCGCA AGGAAGAAGA TCGCGATTGG
2901 AGCTGCAAGA GGGCTAGCGT TTTTGCACCA TAGCTGCATT CCTCACATTA
2951 TCCACAGAGA CATGAAATCA AGCAATGTTT TTCTGGACGA AGATTTCGAA
3001 GCACGTGTTT CGGATTTCCG AATGGCGAGG CTGGTCAGCG CTCTACACAC
3051 CCATCTGAGC GTGAGCACGC TCGCGGGTAC ACCAGTTTAC GTTCCAGCGG
3101 AATATTACCA GAGTTTCCGG TGTACAGCTA AAGGGGATGT TTACAGCTAC
3151 GGAGTTTATC TTCTTGAGCT TCTGTCTGGT AAGAAAACCA TTGATCCAGG
3201 GGAGTTTGGG GAAGATAATA ACCTTGTAGG GTGGGCGAAA CAGCTATACA
3251 GAGAGAAAAG AGGAGCTGAG ATTTCTGATC CGGAGCTTGT AACCGATAAA
3301 TCAGGCGATG TTGAGCTGTT TCATTATTTG AAGATCGCGT CTCAATGCTT
3351 GGATGATCGA CCGTTTAAGC GGCCGACAAT GATTCAACTG ATGGCAATGT
3401 TCAAAGAGAT GAAGGCTGAT ACAGAGGAAG ACGAAAGTCT CGATGAGTTT

3451 TCGCTCAAGG AAAC TCCGTT GGTCGAAGAA TCGCGAGATA AGGAGCCTTA
3501 A

Sequenz des abgeleiteten Gens TOM4 (*Capsella rubella*)

1 ATGAAGCAA AATGGCTGTT TGTGTTGATC CTCTGTTTCT TTACGGCGTT
51 GGGCATTTCAT GGAAAAAGAC TAATCAACTC CGACTTCGAC GAAACTGCTC
101 TTCTAATGGC TTTCAAGCAA TTCTCTGTGA AATCTGATCC CAATAACGTT
151 CTTGGTAACT GGATATACGA GTCGGGTCGT GGTCATGTT CTTGGCGAGG
201 CGTTTCCTGC TCTGACGACG GTCGAATCGT CGGACTAGAT CTCCGAAACG
251 GTGGAGTCAC CGGAACACTA AACCTAGCTA ACCTCACGGC GTTGCCTAAT
301 CTCCAGAATC TTTACCTACA AGGAAACTAT TTCTCCTCTT CCTCTGGAGG
351 AGATTTCGTCG TCTGGTCTTT ATTGTTATCT TCAAGTTCG GATTTGTCTT
401 CGAACTTGAT CTCAGACTAC TCATTGGTGG ATTACGTTTT CTCAAAATGT
451 TCGAATCTGG TTTCTGTGAA TTTCTCAAAC AACAAGCTCG TCGGGAAATT
501 AGGTTTTGCT CCGTCGTCGT TGAAGAGCTT GACGACTGTT GATTTCTCTT
551 ACAATATCTT ATCGGAGAAG ATCCCGGAGA GTTTTATCTC AGAGTCCCG
601 GCGTCGTTGA AGTATCTTGA TCTCACTCAT AACAATTTCT CCGGCGATTT
651 CTCCGATTTA AGTTTCGGGA TGTGTGGGAA TCTCAGCTTTC TTCAGTCTAT
701 CACAGAACAA TATCTCCGGC GTTAAATTC CATTCTCTCT ACCGAACTGC
751 AGATTCTCTG AGACGTTGAA CATCTCTCGG AACAATCTCG CCGGAAAAAT
801 CCCC GGCGGA GAATATTGGG GAAGTTTCCA GAATCTAAAA CAGCTATCTC
851 TAGCTCACAA CCGTTTTCTCC GGCGAAATCC CACCGGAGCT TTCTCTACTC
901 TGCAAAACAT TAGAGACTCT TGATCTCTCT GGAAACGCTC TCTCCGCGCA
951 GCTCCCTTCA CAGTTCACCG CCTGCGTCTG GTTACAGAAC CTTAATATCG
1001 GAAACAAC TA CTTCTCCGGC GATTTCTTAA GCACCGTCGT GAGTAAAATC
1051 ACCAGAATCA CTTATCTATA CGTTGCTTTC AACAACATCT CAGGCTCTGT
1101 TCCGATTTCA CTCACCAACT GTACAAACCT TCGTGTCTCT GATTTAAGCT
1151 CAAATGGTTT CACCGGAAAT GTACCGTCTG GTTTATGCTC TCAGCAAAGC
1201 TCGCCGGTTC TTGAAAAGCT TCTCATAGCC AACAATTACC TCTCAGGAAC
1251 AGTTCCTGTG GAGCTTGGTA AATGTAAGAG CTTGAAGACA ATTGATCTCA
1301 GCTTCAATGA GCTTACTGGT CCGATTCCAA AAGATGTATG GATGTTGCCG
1351 AATCTGTCTG ATCTGGTTAT GTGGGCGAAC AATCTCACTG GAAGTATCCC
1401 AGAAGGTCTT TGTGTTAAAG GAGGGAATTT GGAACTATC ATCCTCAACA
1451 ACAATCTCTT ATCCGGTTC ATCCCTCAGT CGATCTCAAG ATGCACCAAC
1501 ATGATCTGGA TCTCTCTTTC AAGTAACCGT CTGACCGGGA AAATCCCTAC
1551 TGGAATCGGT AATCTTTCCA AGTTAGCAAT CCTGCAGCTG GGGAAACAATT
1601 CTTTGTCCGG AAACGTTCCA CGCCAGCTCG GGAAGTCAA GAGCTTGATT
1651 TGGCTCGATC TGAACAGCAA CAATCTAACC GGGGACCTCC CTGGTGAGCT
1701 AGCTAGCCAA GCGGGGTTAG TAATGCCCGG GAGTGTTC A GGTAAACAGT
1751 TTGCGTTTGT GAGGAACGAA GCGGGAACAG ACTGCAGAGG TGCAGGGGGA
1801 CTAGTAGAAT TTGAAGGCAT TCGCGCAGAA CGATTAGAGC GGTTCCTCAAT
1851 GGTTCATTTCG TGTCTTGC GA CACGGATATA CTCAGGCATG ACAATGTACA
1901 CATTCTCTGC AAATGGAAGC ATGATCTACT TCGACATCTC GTACAATGCG
1951 GTTTCGGGTT TAATACCTCC TGGTTATGGT AACATGGGCT ATCTCCAGGT
2001 CTTGAATCTG GGACATAATC GGATAACTGG AAACATCCCT GACAGTCTTG
2051 GAGGACTGAA AGCCATTGGT GTTCTCGATC TATCTACAA CGATCTTCAA
2101 GGCTACTTAC CCGGATCGCT GGGATCGCTT TCTTTCTCTCA GTGATCTCGA
2151 TGTCTCTAAC AACAACCTCA CTGGTCCAAT CCCATTTGGA GGTCAAGCTTA
2201 CAACATTTCC TGTCTCAAGA TACGCAAACA ACTCTGGCCT ATGTGGTGTG
2251 CCTTTGCGCC CGTGGGTTTC AGCTCCTAGG CGGCCCATTA CCTCCAGTGT
2301 CCATGCCAAG AAGCAAACCT TTGCAACCGC TGTGATTGCT GGAATCGCAT
2351 TTTCTTTTCAT GTGCCTTGTG ATGCTATTTA TGGCGCTTTA CAGGGTGAGG
2401 AAAGTTCAGA AGAAGGAAC T GAAGAGGGAG AAATACATTG AGAGCCTTCC
2451 AACTTCGGGA AGTTGCAGCT GGAAGCTCTC TAGCGTTCCT GAACCGCTTA
2501 GCATCAATGT TGCTACGTTT GAAAAACCC T TGAGAAAAC T CACGTTCCGG
2551 CATCTTCTTG AAGCTACAAA CCGGTTTAGC GCAGAGACTA TGGTCCGATC
2601 TGGTGGGTTT GGAGAAGTCT ACAAGGCTCA ACTCAGAGAC GGATCTGTTG
2651 TAGCAATCAA GAAGTTGATT CGGATCACAG GACAAGGCGA TAGAGAGTTC
2701 ATGGCTGAGA TGGAAACAAT TGGAAAAATC AAGCACAGAA ACCTTGTTCC
2751 GCTTTTGGGA TATTGCAAAG TTGGTGAAGA GAGGCTTCTT GTTTATGAAT
2801 ACATGAAATG GGAAGCTTA GAAACCGTTC TTCACGAGAA ATCATCGAAG
2851 AAAGGTGGAA TATTTCTAAA CTGGACAGCA AGGAAGAAGA TTGCGATTGG

2901 AGCTGCAAGA GGGCTAGCGT TTCTGCACCA TAGCTGCATT CCTCACATAA
 2951 TCCACAGAGA CATGAAATCG AGCAATGTTC TTCTAGATGA AGATTTTGGAA
 3001 GCACGTGTCT CGGATTTTGG AATGGCGAGG CTGGTCAGCG CTCTAGACAC
 3051 GCATCTGAGC GTGAGCACGC TCGCGGGTAC ACCAGGATAC GTTCCACCAG
 3101 AATATTACCA GAGTTTCCGG TGTACAGCTA AAGGGGATGT TTACAGCTAC
 3151 GGAGTTATAC TTCTTGAGCT TCTCTCTGGT AAGAAACCCA TTGATCCAGG
 3201 GGAGTTTGGG GAAGACAATA ACCTGGTCGG GTGGGCAAAA CAGCTATATA
 3251 GAGAGAAAAG TGGAACTGAG ATTCTTGACC CGGAGCTTGT AACCGAAAAA
 3301 TCAGGTGATG CTGAGCTATT TCATTACTTG AAGATTGCGT CTCAATGCTT
 3351 GGATGATCGA CCGTTTAAAG GGCCAAACAAT GATTCAAAGT ATGGCAATGT
 3401 TCAAAGAGCT TAAGGCTGAC ACAGAAGAAG ACGAAAGTCT CGATGAGTTT
 3451 TCGCTCAAGG AAACCTCCGT GGTAGAAGAG TCCCGAGATA AGGAGCCTTA
 3501 A

Sequenz des *Arabidopsis thaliana* cDNA-Klons 22916

1 CCCGGGGAGA CTTAAATTGA TGCAGTAGTG AAATATCACG GCGAAGGGAT
 51 GGTTTTGTGG TAAAAGAGCC AATTGAAATC AGTTATGTCA TCGGGAGGAG
 101 CCGGAGAGTA TAACGAAGAC AGACATCTGT TACGATCCAC CGACGGAGAT
 151 GAAGTCGGCA TTGGCGGAGG AGAGGGTGAT CTAGATGTTG AATCTCAGTC
 201 TCCGGCTATC AGAAGTGGAG CTGGAGGAGT TAGGGATCTG TTCAAGCATA
 251 TAGATCGGAG ATTTTCTCTT TCCGGTCGTC GTTTAAGTTT TAAACGGATG
 301 GAGAATATCA GAGTCGATAG AGAGCGCCAT AATCCTTCTT CTTCTTCAGC
 351 GTTTTCGGCT GCTGGAGAAG AAGATGGTGG TGGTATCAGT AATTTACATA
 401 GCGTTGATGA TCGAAATGAC GAGTACGGGT TTGATGAAGA AGTTCCTCGA
 451 GATATGCTC CACCTGAGTG GGCTCTGCTT CTCATTGGCT GTCTTATTGG
 501 TGTTGCCGCT GGAATTTGTG TCGCCGGCTT CAATAAAGGG GTTCATGTCA
 551 TTCATGAGTG GGCATGGGCT GGTACTCCCA ACGAAGGTGC TGCATGGCTT
 601 CGTCTACAGA GACTAGCTGA TACTTGGCAT CGGATTCTTC TAATTCCGGT
 651 CACTGGAGGT GTTATAGTAG GAATGATGCA CGGTTTGCTT GAGATCTTAG
 701 ACCAAATAAG GCAATCTAAT TCTTCTCAA GACAAGGACT AGATTTTCTT
 751 GCTGGTATTT ATCCAGTGAT AAAGGCCATC CAAGCTGCTG TGACCCCTGG
 801 TACAGGATGT TCACTGGGTC CTGAGGGACC TAGCGTTGAC ATTGGAAAAA
 851 CATGTGCCAA CGGTTTTGCA CTCATGATGG AAAACAACAG AGAAAAGAAGA
 901 ATAGCTCTCA CCGCAGCTGG TGCGGCTTCT GGGATTGCAT CTGGTTTCAA
 951 TGCAGCAGTG GCGGGTTGTT TCTTTGCTAT TGAAACTGTT TTAAGACCTT
 1001 TACGTGCCGA GAACTCGCCT CCATTTACCA CTGCAATGAT AATATTGGCC
 1051 TCTTTATAT CATCGACTGT GTCAAATGCT CTGCTGGGA CTCATGAGC
 1101 TTTTACAGTT CCTCATACG CCTGAAGTC TGCTGCAGAG CTCCCTCTGT
 1151 ACCTGATACT AGGCATGCTT TGCGGTGCTG TCAGTGTGT TTTTTCTCGG
 1201 CTCGTACAGT GGTTTACCAA GTCATTCGAT TTCATCAAAG ACAAATTTGG
 1251 TCTTCCGCT ATCGTGTGCC CTGCACTAGG TGGTTTAGGA GCTGGAATTA
 1301 TTGCTCTCAA GTACCCTGGA ATTTTGTATT GGGGGTTAC AAATGTTGAG
 1351 GAAATCCTGC ATACTGGAAA AAGTGCTTCA GCTCCTGGAA TCTGGTTATT
 1401 AGCTCAGTTA GCCGCTGCAA AAGTTGTGGC TACAGCTCTT TGCAAAGGCT
 1451 CTGGGCTTGT AGGTGGTCTA TATGCACCAA GTTTGATGAT TGGTGCTGCT
 1501 GTTGGTGCTG TATTTGGGGG TTCGGCTGCT GAGATTATTA ACAGAGCTAT
 1551 TCCTGGTAAT GCTGCTGTTG CTCAACCACA AGCTTATGCT CTGGTTGGAA
 1601 TGGCAGCGAC ACTAGCTTCA ATGTGCTCTG TTCCCTTAAC ATCAGTATTA
 1651 CTGCTGTTCC AGCTAACGAA AGATTATAGA ATTTTGCTTC CCTCATGGG
 1701 AGCAGTTGGT TTAGCAATAT GGGTTCCCTC TGTGGCAAAC CAAGGCAAAG
 1751 AGAGTGATTC ATCTGAAGGT CGTAGCACAG GAAGAGGATA CTCTTCTCTT
 1801 TCACCTTCCG AACGTAAAAC CGAAGGAGTC TGGAGACATA CGGATAATGC
 1851 TGACTCCCTT GAGCTTACCG TCATAGAGAA CCCTGACCAT AATTCCTTTT
 1901 TGGATGAAGA GACTATTCTG GAAGACTTAA AAGTTATGCG GGTATGTCA
 1951 AAGAACTATG TGAAAGTTT TCCAGGAACA ACTCTAAGAG AAGCAAGAAA
 2001 CATCCTTAAA GAGAGCCACC AAAACTGCAT TATGGTGGTC GATGACGATG
 2051 ATTTTCTAGC TGGAACTCTA ACACACGGTG ACATAAGAAG AACTTTGTCC
 2101 AATAACGCCT CTACAATCTT AGATGAGAAT ACATGTCCGG TTTCTTCTGT
 2151 ATGTACTAAG AAAATAAGCT ATCGAGGCCA AGAACGCGGT TTACTTACTT

2201 **GTTATCCAGA CGCCACAGTT GGTGTAGCCA AGGAGTTGAT GGAGGCTAGA**
 2251 **GGTGTAAGC AATTACCCTGT TGTAAAACGA GGTGAAGTAA TTCACAAAGG**
 2301 **GAAACGAAGG AAACCTGCTTG GCCTCCCTTCA TTACGATTCC ATTTGGACTT**
 2351 **TTCTCAGAGA TGAAATGAGT CGTAGGAGAT CGATCAACGA CCGGAGAAAA**
 2401 **GACAAAGAGG TTGGTACAAA TGGGCATTGA TTGGGTTTGT GCCCATTTTG**
 2451 **GTATAGCTTT TTTTTTATAT TTTTCTCTTT TTAATTTTTT GAGGTCAAAAT**
 2501 **TGATTAGAAC GAGGCAAACA AGATCGTAAA AAT**

Sequenz des abgeleiteten Gens 22916 (*Capsella rubella*)

1 **ATGTCATCGG GAGGAGCCGG AGAGTATAAC GAAGACAGGC ATCTGTTGCG**
 51 **ATCCACCGGC GATGAAGACA GCATTGGTAG AGAGGATGGT GATATCGATG**
 101 **TTGAATCTCA GTCTCCGGCT GTCAGAAGTG GAGCGGGAGG AGTTAGGGAT**
 151 **CTGTTTAAAG ATCTGGATCG GAGATTTTCA CTCTCCGGGC GTCGTTTAAAG**
 201 **CTTCAAACGG ATGGAGAATA ACAGAGTCGA TAGAGAGCGC CACAACCCGT**
 251 **CTTCTTCTTC CTCCGCGTTT TCGGCTGCAG AAGAAGATGG CGGTGGTATC**
 301 **AGTAATTTAC ATAACGTGGA TGATCGAATT GACGAGTACG GATTTGATGA**
 351 **AGAAGTGCTC GGAGATAGTG CTCCACCTGA GTGGGCTTTG CTTCTCATCG**
 401 **GCTGTCTTAT CGGTGTTGCC GCTGGAATTT GTGTCGCTGG CTTCAACAAA**
 451 **GGGGTACGGT CATAGATCAG ATCTCGATCT ATGTTGATTC GATTGAACAC**
 501 **CTTGATCCTG CTAGCTTTAA TTTACGTTAT ACTATTTTGA CTGCTGATTC**
 551 **ATTAGATGAG TGCTATTGTC TAGGTTTCATG TTATTCATGA GTGGGCATGG**
 601 **GCTGGTACTC CCAACGAAGG TGCTGCATGG CTCCGCTTAC AGAGACTAGC**
 651 **TGACACTTGG CACCGGATTC TTCTAATTCC GGTCACTGGG GGTGTTATAG**
 701 **TTGGGATGAT GCACGGTTTG CTCGAGATCT TAGACCAAAT CAGGCAATCC**
 751 **ACTTCTTCTC AGCGACAAGG ACTAGATTTT CTTGCTGGTA TTTATCCCGT**
 801 **GATAAAGGCC ATCCAAGCTG CTGTGACCCT TGGTACAGGA TGTTCACTGG**
 851 **GTCCTGAGGG ACCTAGCGTC GACATTGGAA AGTCTTGTGC CAATGGTTTT**
 901 **GCTCTCATGA TGGAAAACAA CAGAGAAAGG AGAATAGCTC TCACTGCAGC**
 951 **TGGTGC GGCT TCTGGGATTG CATCTGGTAT AATCTACGCC CTACCTTTTTT**
 1001 **TAAAACAGCT TATGCACCTT CTGAATTGAT ATTGATGTAA TATGGATTGC**
 1051 **CGAGGCATGG AACAGCAACA TAGTTTCTTC CTGCACTCAA GTATATTAGC**
 1101 **ACTAGTTAAC ATTAGCTTTA ACCTGGTGAA TTTTCGTTAAT AAATGTCTAT**
 1151 **TCTAGTTTCC TTGTATCATT TTGGCTTATT GCAATCTTTA TGGCCCTCA**
 1201 **TTTTTTTCGCT CACTTATCTG TCCCCATAAT AGGTATGTAT TATATAGCCT**
 1251 **ACAAATGTGC CAATTGATCT TAGGTTCTAC TACTAGAAA AACATCTTGC**
 1301 **TGGGAGTAAT TACTGAATAC TAGTATATTT ATGGCTATTT AAAGTATGAT**
 1351 **TCATGGGCAC TGGTTTTTCT AGCTAAAGAT AAGAGCACTA AGAGAGGCTC**
 1401 **TTTGTTCCTT TGTTCCTCCC AGGTAGGGAC TCAATTCGCT TGTTCGTCGT**
 1451 **AAAAAATCTG ATTTAATCTT ATCTTTATTC GTGTTACAGG TTTCAATGCA**
 1501 **GCAGTGGCAG GTTGTTCCTT TGCTATTGAG ACCGTTTTAA GACCTCTACG**
 1551 **TGCCGAGAAC TCGCCTCCAT TTACCCTGC AATGATAATA TTGGCCTCTG**
 1601 **TTATATCTTC GACTGTGTCA AATGCCTGC TTGGAECTCA ATCAGCTTTC**
 1651 **ACAGTTCCCT CACACGACCT GAAGTCCGCT GCAGGTATGT TTTTTCATTG**
 1701 **TTAATCCGTA TTGCAAACCT TCCAGTTAGT ATTATGGTAT TATTTTGTG**
 1751 **TAAGATCTGT ATTTAGTGTG CTGATGATTA TGGTATAATG TGATATATCA**
 1801 **TAAAGAATCG TGGTAATTAT AAAATAATGA AGGTCAATAT TTATTTGCTT**
 1851 **CATCGAGTAT TTGTGTATAG ATTTACATAT CTGTTTTCAA TTGTCTCCAG**
 1901 **AGCTCCCTCT GTACCTGATA TTAGGCATGC TTTGCGGTGC TGTCAGCGTT**
 1951 **GTTTTTCTC GTCTCGTAC CTGGTTTACC AAGTCATTCG ACTTCATCAA**
 2001 **AGAAAAATTT GGTCTTCCTG CTATTGTGTG CCCTGCACTA GGTGGTTTAG**
 2051 **GAGCTGGAAT TATTGCTCTC AAGTACCCTG GAATTTTGTG TTGGGGGTTT**
 2101 **ACAAATGTTG AGGAAATCCT GCATACCTGGA AAGAGTGCTT CAGCTCCTGG**
 2151 **AATTTGGTTA TTAGCTCAGT TAGCCGCAGC CAAAGTTGTG GCTACAGCTC**
 2201 **TTTGCAAAGG CTCTGGGCTT GTAGGTGGTC TATATGCACC AAGTTTGATG**
 2251 **ATTGGTGCTG CTGTTGGTGC TGTGTTGGG GGTTCGGCTG CTGAGATTAT**
 2301 **TAACAGAGCT ATTCCTGGTA ATGCTGCTGT TGCTCAACCA CAAGCTTACG**
 2351 **CTCTGATTTT GTATTTCCCT TGTGCGAGGC CTGCTTCAG TGCAGTTTTT**
 2401 **TAAAATAAAT TTTGTTACTC TTCAATTTGG ATGTCATGAA ATCTCAGTTG**
 2451 **TTAAGATGCT GTCGTTTTAT GTGATGCAGG TTGGAATGGC AGCTACACTA**
 2501 **GCTTCCATGT GTTCTGTTCC GTTAACATCA GTGTTACTGC TGTTGAGCT**
 2551 **AACAAAAGAT TATAGAATTT TGCTTCCCTT CATGGTAAAT ATTATCTTTC**
 2601 **CCGCTTAGTT TCAATGGTGC ATGTCGTGTA AATTTGCTTA AGAAGGGCGG**

2651 GATTCATAAA GATAGTTCTG ATTTTTCTCG TTTAACTAAT CAAGCAGTAA
2701 AACCGTGGCA AAACATATGAA ATGTACCACT TAAGATGTCG CTGTAACAGG
2751 ATGTTATTAA CTAGAGGTC AGAGCTTGTT TGTATGTAGA CAAGAATATT
2801 CTTTTTTCAC TTAACATGCT GCATTCTGGT TCATATTTTG ATCAAGTTGT
2851 TCTTGATCTG ATCTCTCTGT CGATTTCTAC TAACAGGGAG **CAGTTGGTTT**
2901 **AGCTATATGG GTTCCCTCTG TCGCAAACCA AGGCAAAGAG AGTGAGTCAT**
2951 **CTGAAGGTCG CAACACTGGA AGAGGATACT CTTCTGTTTC ACCTTTAGAG**
3001 **CGTAAAACGG AAGGTGTATG GAGACACACG GATAATGTTG ACTCCCTTGA**
3051 **GCTCACCGTC ATAGAGAGCC CTGACCATAA ATCCTTTTTG GATGAGGAAA**
3101 **CTATTCTGGA GGATTTAAAG GTTTTGCGGG TTATGTCAAA GAACTACGTG**
3151 **AAGGTTTCTC CAGGAATGAC TTTAAGGGAA GCAAGAAATA TCCTTAAAGA**
3201 **CAGCCACCAA AACTGCCTTA TGGTGGTCGA TGAAGATGAT TTTCTAGCCG**
3251 **GGATCCTAAC ACACGGCGAC ATCAGAAGAT ACTTGCCAA CAACGTCTCA**
3301 **ACAATCTTAG ATGTAAGAAG TGAACAAAA CAAACCCCTT CAGTCTCGTA**
3351 CTACTCATAA CTCTGCTCTT TAAATTACAG TGGACTTGAA AATAAGTCTA
3401 TGTCTCTTGG CAGGAGAATA CATGCCAGGT TTCTTCAGTG TGTACTAAGA
3451 ATATAATCTA CCGAGGCCAA GAACGCGGTC TGCTAACTTG TTATCCAGAC
3501 GCGACGGTTG GAGTAGCCAA AGAGTTGATG GAGGCAAGAG GAGTGAAGCA
3551 **ATTACCTTGC GTAAAACGAG GTGAAGTAAT TCACAAAGGG AAAAGAAGGA**
3601 **AATGCTTGG CCTTCTCAC TATGACTCTA TTTGGACTTT TCTCAGGTAA**
3651 CAGAAAAAAG AAAAAAAAT CTGTTTTTAC TCAAAGTTAA CCAAAGCCAA
3701 AGTTGTTAAC CTTTGGTTGG TTGTTTTTTC ACGCAGAGAT **GAAATGAGTC**
3751 **GGAGGAGATC GATCAACGAC CGGAGAAAAG ACAAAGAGGT TGGTGCAAAT**
3801 **GGGCATTGA**

25028 - Sequenz der vorhergesagten cDNA des Gens TOM6 (*Arabidopsis thaliana*)

1 **ATGGGCCAGT TTTATTTTAG GGCCCATAAA AGTCGGGTCG GATCCTCTTC**
51 **AAAATGTGGA TTCGGTTCAG CGATTCTCTT CTCGGTAGTT CTCAAAGCA**
101 **GAGAGCAGCG GCAGAGACGT CTGAGTAGAA CCTGGGATCC AGGTTAGTT**
151 **ATTTTACCAT TATACCCCTT GAACTCGTCT TCTTCGTCAA GCACTAGTGT**
201 **TGCGGTTCTG AAAGCTTCTC GCTTCTCTT CACATCGAGG AAATTTTGCA**
251 **ATGGTAGCAT CGGTGGTAT GTAACCGATA ACGGGACCGA AGAGCCCTTG**
301 **AAGATAACTT GGGAACTTTC GGAGATGGAT TGTGAGTTTG ACCAGGAGGA**
351 **AAATGGTGAG AAAATCTCTG TTAGAAAAAG GTTTATGGAG AGTACTAAAC**
401 **TCAGTGCCTC TCGGGTCTC GATACGTTAC AGCAAGATTG TCCCGGTTTC**
451 **AATACGAAAT CAGCTTTAGA TGAATTGAAT GTTCAATCT CAGGGCTATT**
501 **AGTGAGAGAG GTTCTTGTG GGATCTTAAG AACTCTGAGC TTTGTAATA**
551 **AAACAAGGTG TGCTAAGCTC GCTTACAAGT TCTTTGTGTG GTGTGGTGGG**
601 **CAAGAGAATT TCAGGCACAC TCGGAATTGT TATCATTTAC TTATGAAGAT**
651 **ATTTGCTGAA TGTGGTGAGT ATAAAGCCAT GTGCAGGCTA ATTGATGAGA**
701 **TGATCAAAGA CGGTTATCCA ACAACAGCAT GCACGTTTAA TCTATTGATA**
751 **TGTACTTGCG GGAAGCAGG CCTTGCTAGA GATGTTGTTG AGCAGTTCAT**
801 **CAAGTCGAAA ACTTTCAATT ACCGACCTTA TAAACTCTT TACAATGCCA**
851 **TTTTGCATTC TTTGCTAGGG GTGAAACAGT ACAAGCTGAT TGATTGGGTT**
901 **TACGAGCAGA TGTTAGAAGA CGGGTTTACG CCAGATGTTT GACTTACAA**
951 **CATTGTAATG TTTGCAAAC TTAGGTTAGG GAAAAACAGAT CGGCTTTATA**
1001 **GATTGCTAGA TGAAATGGTT AAAGACGGGT TTTCTCCTGA TTTGTATACA**
1051 **TACAACATCC TCCTCCATCA TCTAGCAACA GGAAACAAGC CTCTCGCTGC**
1101 **CTTTAACCTT CTGAATCATA TGAGGGAAGT AGGAGTTGAG CCCGGTGTCA**
1151 **TCCATTTTTC TACTCTGATA GACGGGCTGA GCCGAGCTGG GAAAGTAGAA**
1201 **GCCTGCAAACT ACTTTATGGA CGAAACGGTG AAAAGTTGGAT GCACGCCAGA**
1251 **TGTTGTTTGC TACTCTGTTA TGATCACAGG ATATATATCA GGTGGGGAGC**
1301 **TCGAGAAAGC CGAAGAAATG TTCAAAAGAAA TGACAGAAAA GGGGCAACTC**
1351 **CCGAATGTTT TCACATACAA TTCGATGATC CGCGGGTTTT GTATGGCGGG**
1401 **GAAATTCAAA GAGGCGTGCG CGTTGCTCAA GGAAATGGAG TCAAGAGGGT**
1451 **GTAACCCTAA TTTTGTAGTG TATAGTACAC TGGTGAACAA TCTAAAAAAC**
1501 **GCGGGGAAGG TTTTAGAGGC TCATGAAGTA GTGAAGGACA TGGTAGAGAA**
1551 **AGGGCATTAT GTTCATCTTA TTTCAAAAGTT GAAAAAATAT AGAAGATCTT**
1601 **AG**

Sequenz des abgeleiteten Genfragments TOM6 (*Capsella rubella*)

```
1  ATCCATTTCA  CTACTCTCGT  AGACGGGCTT  AGCCGAGCTG  GGAAGTTAGA
51  AGCTTGCCAA  TACTTTATGG  ACGAAATGGT  GAAAGTTGGA  TGCACGCCGG
101  ATGTTGTTTG  CTACACTGTT  ATGATTACAG  GATATATTC  AGGCGGGGAG
151  CTCGAGAAAG  CTGAGGAGAT  GTTCAAAGAG  ATGACGGAAA  AGGGACAGCT
201  CCCAAATGTT  TTCACATACA  ATTCCCTGAT  CCGGGGGTTT  TGTATGGCGG
251  GAAAGTTCAA  AGAGGCGTGC  TCGTTGCTCA  AGGAAATGGA  GTC AAGAGGG
301  TGTAACCCTA  ATTTTGTGGT  GTATAGTACA  CTGGTGAACA  ATCTTAGAAA
351  TGCCGGGAAG  CTTTGTAGAG  CTCACGAGGT  AGTTAAGGAC  ATGGTGGAGA
401  AAGGGCATT  TGCTCATT  CTTTCAAAGC  TGAAAAATA  TAGAAGATCC
451  TAG
```

IV Sequenzierte Subklone und PCR-Produkte

Cosmid 34 (Contig0)

34E01, 2, 7, 13
34S01, 1P, 2, 3, 3P, 8, 8P
34H2P, 3, 4P, 5, 7P, 8, 8P, 20

PCR-Produkte:

34-143-145-PCR (526-2162 Contig0)
C34-179-PCR (179-291: 2970-3747 Contig0)
P3-160 (160-216: 10404-11833 Contig0)
C34-197-160-PCR (160-197: 10404-15962 Contig0)
C34-197f-PCR (197-for: 15263-15962 Contig0)

Cosmid 12 (Contig0)

12E01, 3P
12EX3, 10, 17
12S01, 2, 3, 4, 7, 8, 15

PCR-Produkte:

12XhoI-for-uni-PCR (15191-16377 Contig0)
C12-205-206-PCR (205-206: 24470-29122 Contig0)

Cosmid 7-II (ContigI)

7H01, 2, 3, 5, 12, 22
7S10, 11, 27, 39
7RV12

PCR-Produkte:

C7-283-285 (3797- 5930 ContigI)
7292-PCR (292-293: 10660-11993 ContigI)

Cosmid 44 (ContigI)

44E01, 2, 4, 6
44H01, 2, 4, 6, 7, 10
44ES3

PCR-Produkte:

44HIII.rev (ca. 11889-14889 ContigI)
44JK (20812-21158 ContigI)
44C-172 (172-229: 20941-21532 ContigI)
C44E-229 (172-229: 20941-21532 ContigI)

Cosmid 3-I (ContigI)

3E01, 2, 4, 6, 7
3H01, 2, 3, 3del, 4, 4.1, 4.3, 4.5, 4.6, 5
3X02, 4, 5, 6, 7

PCR-Produkte:

P1-229 (172-229: 20941-21532 ContigI)
3I-30-34 (30-34: 22936-23706 ContigI)
10-174 (33-174: 23687-23891 ContigI)
3I-30-31 (30-31:26755-27203 ContigI)
3I-32-35 (27861-32131 ContigI)
3I-28-29 (37591-37891 ContigI)

Cosmid 98 (ContigI)

98H01, 2, 3, 3P, 5, 7, 8, 10, 17, 20
98EX4, 5, 12
98HX01, 13, 15, 18, 25, 28

PCR-Produkte:

11-181 (181-182: 48241-50218 ContigI)
P2-182-245 (48919-50218 ContigI)

Cosmid 54-I (ContigI)

54E02, 6
54X02, 3, 4, 5, 6

PCR-Produkte:

54-22f (44853-51069 ContigI)
54-27f (45595-51069 ContigI)
54I-26-uni (50132-51069 ContigI)

Cosmid 14-II (ContigI)

14E01, 2, 4, 4.5, 7
14HX1, 2, 3, 4, 7, 8, 9, 10, 29

PCR-Produkte:

C14-AB (57791-58207 ContigI)

Cosmid 3-II (ContigI)

3E02, 3, 4, 9, 10, 17, 18
3S01, 3, 4, 18
3X03, 8, 9, 12, 12-2P, 14, 17

PCR-Produkte:

3II-38-uni (63441-64398 ContigI)
3II-37-38 (64219-64398 ContigI)
3II-94-95-PCR (70492-71738 ContigI)
2-177-178 (73981-76123 ContigI)
3II-97-141 (77051-78382 ContigI)
3-175 (175-176: 78465-79087 ContigI)
3II-SpeI (84452-86006 ContigI)
3II-SacI (84827-86006 ContigI)

Cosmid 54-II (ContigII)

54E01, 1a, 2, 2-2P, 3, 4, 4-2P, 7, 15, 16
54H01, 2, 5, 7, 8, 10, 11, 12
54S01, 2, 5, 6, 12, 13, 16, 21
54EX3, 3P, 7, 16
54X06, 8, 10, 11, 12, 14, 16

PCR-Produkte:

54XhoI (1-1548 ContigII)
54S-167-127 (714-2699 ContigII)
54S-168-188 (3045-5418 ContigII)
5440-163 (163-277: 8219-8820 ContigII)

Cosmid 10 (Contig-TOM)

10H01, 2, 3, 4, 7, 8, 12, 13, 15, 16, 17
10X04-2P, 7-2P, 16, 17, 18, 19

PCR-Produkte:

T23 (T7-T8): 9652-10055
107-108: 9759-11765
Co10-122 (110-122): 18437-18805
1-133/10-133134-133: 21292-21759
Co10-123 (122-123): 23451-23693

Cosmid 20 (Contig-TOM)

20H01, 3, 5, a, b, d
20X15, 17, 20

PCR-Produkte:

Co20-104-121: 2740-3743
T423 (T42-T43): 23325-23942

Anmerkungen

Ich versichere, daß ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen in der Arbeit einschließlich Tabellen, Karten und Abbildungen, die anderen Werke im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; daß diese Dissertation in keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; daß sie noch nicht veröffentlicht worden ist sowie, daß ich eine solche Veröffentlichung vor Abschluß des Promotionsverfahrens nicht vornehmen werde.

Die Bestimmungen der Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. J. Schell betreut worden.

Publikationen:

Schmidt, R., Acarkan, A., Koch, M. and Roßberg, M. (1999). A strategy for comparative physical mapping in cruciferous plants. *Plant Evolution in Man-Made Habitats, Proceedings of the VIIth International IOPB Symposium Amsterdam* (eds. L.W.D. van Raamsdonk and J.C.M. den Nijs), Hugo de Vries Laboratory, Amsterdam, The Netherlands, pp.183-196.

Danksagung

Herrn Prof. Dr. J. Schell möchte ich für die Betreuung meiner Doktorarbeit bedanken.

Herrn Prof. Dr. U.-I. Flügge danke ich für die Übernahme des Zweitgutachtens.

Frau Dr. R. Schmidt möchte ich mich für die Bereitstellung des Arbeitsplatzes und für die außerordentliche Betreuung danken. Die zahlreichen Anregungen und Diskussionen trugen zum Gelingen dieser Arbeit bei.

Bei den Mitgliedern meiner Arbeitsgruppe, Karine Boivin, Oliver Clarenz, Alexandra Forsbach, Mario Gils, Berthold Lechtenberg, Dr. Mathias Roßberg, Daniel Schubert, Sandra Stegemann möchte ich für die gute Arbeitsatmosphäre bedanken und vor allem für das Verständnis und die Unterstützung in der Endphase meiner Arbeit.

Mein besonderer Dank gilt Dr. Markus Koch und Dr. Mathias Roßberg Betreuung zu Beginn der Arbeit und die zahlreichen Hilfestellungen bei experimentellen und theoretischen Problemen.

Karine Boivin, Oliver Clarenz, Dr. Markus Koch und Dr. Mathias Roßberg danke ich für die Verwendung ihrer unveröffentlichten Forschungsdaten im Diskussionsteil.

Meinem langjährigen Studien- und Laborkollegen Mario Gils möchte ich für die gute Freundschaft danken, sein stets offenes Ohr bei Problemen jeglicher Art und für unsere „Privatgespräche“ im Kaffeeraum.

Frank Eickelmann danke ich für die Pflege meiner Pflanzen und die zahlreichen Ermahnungen. „Adile! Deine Pflanzen können geerntet werden oder soll ich sie wegschmeissen?“.

Den Mitgliedern der übrigen Arbeitsgruppen des Max-Delbrück-Laboratoriums danke ich für die nette Zusammenarbeit und Hilfsbereitschaft.

Außerdem möchte ich mich bei meinen Bekannten und Freunden bedanken, die stets Verständnis gezeigt haben, wenn „Frau Doktor wieder im Stress? war.

Zuletzt und vor allem möchte ich meiner Familie für die moralische und finanzielle Unterstützung danken, die mir das Studium und die Promotion erst möglich machten. Außerdem möchte ich mich für die Verpflegungspakete bedanken, die oft die Ausübung meiner „hervorragenden“ Kochkünste während der Doktorarbeit überflüssig machte.

Lebenslauf

Name, Vorname	Acarkan, Adile
Geburtsdatum	31.03.1971
Geburtsort	Duisburg
Eltern	Yurdanur Acarkan Altan Acarkan
Staatsangehörigkeit	türkisch
Familienstand	ledig
Schulbildung	1977-1981 Gemeinschafts-Grundschule Hochfeldstr., Duisburg 1981-1983 Freiherr-vom-Stein-Realschule, Duisburg 1983-1990 August-Seeling-Gymnasium, Duisburg
Studium	
Oktober 1990 - September 1996	Studium der Biologie an der Heinrich-Heine- Universität Düsseldorf
Februar 1993	Diplomvorprüfung
April/Mai 1995	mündliche Diplomprüfung
Mai 1995- September 1996	Diplomarbeit in der Arbeitsgruppe von Prof. Dr. Bünemann mit dem Thema: „Versuche zur immunologischen Identifizierung spezifischer Dynein-“heavy chain“-Proteine aus Spermien von <i>Drosophila</i> “
November 1996 - Juni 2000	Dissertation am Max-Delbrück-Laboratorium in der Max-Planck-Gesellschaft Betreuung: Prof. Dr. J. Schell