

**Repetitive DNA-Sequenzen im Genom der Kichererbse
(*Cicer arietinum* L.)**

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften

vorgelegt beim Fachbereich Biologie und Informatik
der Johann Wolfgang Goethe - Universität
in Frankfurt am Main

von
Christina Staginnus
aus Hanau

Frankfurt am Main 2001
(DF1)

vom Fachbereich Biologie und Informatik der Johann Wolfgang Goethe-Universität als
Dissertation angenommen.

Dekan: Prof. Dr. B. Streit

1. Gutachter: Prof. Dr. G. Kahl

2. Gutachter: Prof. Dr. H. D. Osiewacz

Datum der Disputation: 08.04.2002

INHALTSVERZEICHNIS

A.	EINLEITUNG	1
A.I.	Repetitive DNA	1
A.I.1.	Satelliten-DNA	2
A.I.2.	Klasse I: Retrotransposons	6
A.I.3.	Klasse II: DNA-Transposons	13
A.I.4.	Hat repetitive DNA eine Funktion im Genom?	16
A.I.5.	Welche Mechanismen begrenzen die Amplifikation repetitiver Sequenzen?	19
A.II.	Die Kichererbse (<i>Cicer arietinum</i> L.) und die Gattung <i>Cicer</i>	21
A.II.1.	Allgemeines	21
A.II.2.	Cytologische Untersuchungen bei <i>C. arietinum</i>	21
A.II.3.	Die Analyse des Kichererbsengenoms	22
A.II.4.	Verwandte Wildarten der Kichererbse	23
A.II.5.	Repetitive DNA im Genom von <i>C. arietinum</i>	24
A.III.	Die Fragestellung	26
B.	MATERIAL & METHODEN	27
B.I.	Material	27
B.I.1.	Pflanzen- und Pilzmaterial	27
B.I.2.	Bakterienstämme und Vektoren	28
B.I.3.	Restriktionsenzyme	29
B.I.4.	Oligonukleotide	30
B.I.5.	Verwendete Computerprogramme	31
B.II.	Methoden	32
B.II.1.	Isolation von Nukleinsäuren aus verschiedenen Pflanzen	32
B.II.2.	DNA-Konzentrationsbestimmung	35
B.II.3.	Restriktion von genomischer und bakterieller DNA	35
B.II.4.	Gelelektrophoretische Trennung von DNA-Fragmenten	35
B.II.5.	Southern Transfer	38
B.II.6.	Klonierung von DNA	40
B.II.7.	Isolation klonierter Fragmente	42
B.II.8.	Polymerase-Kettenreaktion (PCR)	44
B.II.9.	Sequenzierung von DNA-Fragmenten	48
B.II.10.	Generation von „sequence-specific amplified polymorphisms“ (S-SAP) mit PCR	49
B.II.11.	Southern-Hybridisierung	51
B.II.12.	<i>in-situ</i> Hybridisierung	53

C.	ERGEBNISSE	61
C.I.	Hochrepetitive Sequenzen	61
C.I.1.	Die tandemrepetitiven Elemente <i>CaSat 1/ CaSat 2</i>	61
C.I.1.1.	Genomische Organisation	61
C.I.1.2.	Chromosomale Lokalisation	68
C.I.1.3.	Vorkommen und Abundanz der Satellitenelemente in anderen Spezies	73
C.I.2.	Das disperse repetitive Element <i>CaRep</i>	73
C.I.2.1.	Genomische Organisation der nicht-tandemartig organisierten hochabundanten Sequenzen	73
C.I.2.2.	Sequenzierung der Klone aus den drei dispersen Sequenzgruppen	75
C.I.2.3.	Identifikation des Polyprotein-Leserasters eines vollständigen <i>CaRep</i> -Elements	79
C.I.2.4.	Sequenzierung der langen terminalen Repeats (LTRs) von <i>CaRep</i>	84
C.I.2.5.	Chromosomale Lokalisation der <i>CaRep</i> -Elemente	96
C.I.2.6.	Abundanz von <i>CaRep</i> -Sequenzen in anderen Spezies	99
C.II.	Mittel- bis niedrigrepetitive Sequenzen	100
C.II.1.	Das <i>Ty1-copia</i> -Element <i>CaTy</i>	100
C.II.1.1.	Nachweis von <i>Ty1-copia</i> -Retrotransposonsequenzen in Leguminosen	100
C.II.1.2.	Abundanz und genomische Organisation von <i>Ty1-copia</i> -Retrotransposons in <i>Cicer arietinum</i>	104
C.II.1.3.	Chromosomale Lokalisation der <i>CaTy</i> -Familie in Kichererbse	105
C.II.1.4.	Sequenzen vom <i>CaTy</i> -Typ im Genus <i>Cicer</i> und in verwandten Spezies	107
C.II.2.	<i>CaDis</i> - das Relikt eines Retrotransposons?	108
C.II.2.1.	Sequenzen mit Homologie zum repetitiven Element <i>RPE15</i> (<i>Medicago sativa</i>)	108
C.II.2.2.	Sequenzierung der <i>RPE15</i> -Amplifikate	109
C.II.2.3.	Genomische Organisation der Amplifikate im Genom der Kichererbse	110
C.II.2.4.	Chromosomale Lokalisation der <i>CaDis</i> -Sequenzen	112
C.II.2.5.	Abundanz <i>CaDis</i> -ähnlicher Sequenzmotive im Genus <i>Cicer</i> und verwandten Spezies	114
C.II.3.	<i>CaLin</i> - ein <i>LINE</i>-ähnliches non-LTR-Retrotransposon	116
C.II.3.1.	Identifikation von <i>LINE</i> -ähnlichen Sequenzen in verschiedenen <i>Cicer</i> -Spezies	116
C.II.3.2.	Genomische Organisation der <i>LINE</i> -ähnlichen Sequenzen in <i>Cicer arietinum</i>	120

C.II.3.3.	Chromosomale Lokalisation von <i>CaLin</i> -Kopien in der Kichererbse	120
C.II.3.4.	<i>CaLin</i> -ähnliche Sequenzen im Genus <i>Cicer</i> und verwandten Spezies	122
C.III.	Das DNA-Transposon (Klasse II) <i>CaEn/Spm</i>	124
C.III.1.	Nachweis von Transposons des <i>En/Spm</i> -Typs in pflanzlichen Genomen	124
C.III.1.1.	Degenerierte Oligonukleotide aus der Transposaseregion von 132 <i>En/Spm</i> -Elementen	124
C.III.1.2.	<i>En/Sp132m</i> -Transposons in <i>Cicer</i> und anderen Gattungen	127
C.III.1.3.	Identifikation der Amplifikate durch Sequenzierung	129
C.III.2.	Charakterisierung von <i>CaEn/Spm</i> , einem Klasse II-Transposon im Genom der Kichererbse	132
C.III.2.1.	Sequenzierung einiger Transposonkopien	132
C.III.2.2.	Genomische Organisation und Abundanz der <i>En/Spm</i> -Familie in Kichererbse	133
C.III.2.3.	Chromosomale Lokalisation der <i>CaEn/Spm</i> Transposonkopien	134
C.III.2.4.	Abundanz verwandter Sequenzmotive im Genus <i>Cicer</i> und anderen Genera	134
C.IV.	Transkription von dispersen, niedrig- und mittelrepetitiven Elementen der Kichererbse	137
C.IV.1.	Nachweis von spezifischen „messenger“-RNAs repetitiver Elemente mit RT-PCR	137
C.IV.2.	Identifikation von <i>CaDis</i> -homologen Transkripten durch Sequenzanalyse	139
C.V.	Verwendungsmöglichkeiten charakterisierter repetitiver Elemente aus <i>Cicer arietinum</i>	140
C.V.1.	Phylogenetisch-taxonomische Betrachtungen im Genus <i>Cicer</i>	140
C.V.1.1.	Abundanz und genomische Organisation der Tandemelemente <i>CaSat1</i> und <i>CaSat2</i> in Spezies der Gattung <i>Cicer</i>	140
C.V.2.	Generation von molekularen Markern für die genetische Karte des Kichererbsengenoms	143
C.V.2.1.	Ableitung eines geeigneten Oligonukleotids aus <i>CaRep</i> und seine Verwendung in einem „sequence-specific amplified polymorphism“(S-SAP)-Ansatz	144
C.V.2.2.	Lokalisation der polymorphen Marker in der genetischen Karte	146
C.V.2.3.	Verifikation der Identität der generierten Marker durch Southern-Hybridisierung	151
D.	DISKUSSION	153

D.I.	Die Satellitenelemente <i>CaSat1</i> und <i>CaSat2</i>	153
D.I.1.	<i>CaSat1</i> und <i>CaSat2</i> - dominante Bestandteile des Genoms von <i>C. arietinum</i>	153
D.I.1.1.	Struktur und Sequenz der Monomere von <i>CaSat1</i> und <i>CaSat2</i>	153
D.I.1.2.	<i>CaSat1</i> und <i>CaSat2</i> weisen eine übergeordnete Multimerstruktur auf	154
D.I.1.3.	Enthält <i>CaSat1</i> eingestreute Retrotransposonsequenzen?	155
D.I.1.4.	Die chromosomale Lokalisation gibt Hinweise auf einen möglichen Ursprung von <i>CaSat1</i>	156
D.I.1.5.	Impliziert die Lokalisation von <i>CaSat2</i> einen möglichen funktionellen Aspekt?	158
D.I.1.6.	Die vermutliche Evolution der beiden <i>CaSat</i> -Elemente	159
D.I.1.7.	<i>CaSat1</i> und <i>CaSat2</i> sind unterschiedlich homogenisierte Satellitenfamilien im Genom von <i>C. arietinum</i>	159
D.I.2.	Verwendung der tandemrepetitiven Elemente für phylogenetisch-taxonomische Betrachtungen	160
D.I.2.1.	<i>C. cuneatum</i> ist ein Außenseiter unter den annuellen <i>Cicer</i> -Spezies	160
D.I.2.2.	<i>CaSat1</i> und <i>CaSat2</i> unterstützen molekulare Daten in der Taxonomie	161
D.II.	Das LTR-Retrotransposon <i>CaRep</i>	162
D.II.1.	<i>CaRep</i> - ein hochabundantes Retrotransposon des <i>Ty3-gypsy</i> -Typs	162
D.II.1.1.	PCR-gestützte Isolation einer vollständigen <i>CaRep</i> -Kopie	162
D.II.1.2.	Struktur einer vollständigen Kopie von <i>CaRep</i>	163
D.II.1.3.	Abundanz von <i>CaRep</i> -Kopien im Vergleich zu anderen LTR-Elementen	165
D.II.1.4.	Chromosomale Organisation von <i>CaRep</i>	166
D.II.1.5.	Evolution von <i>CaRep</i> im Genus <i>Cicer</i>	167
D.II.2.	<i>CaRep</i> -Sequenzen als molekulare Marker	168
D.II.2.1.	Auf Retroelementen basierende molekulare Marker	168
D.II.2.2.	Die S-SAP-Technik macht <i>CaRep</i> -Marker sichtbar	168
D.II.2.3.	Die Kartierung von <i>CaRep</i> -Markern	169
D.III.	Die mittel- und niedrigrepetitiven Retroelemente <i>CaTy</i>, <i>CaDis</i> und <i>CaLin</i>	171
D.III.1.	Die LTR-Retrotransposonfamilien <i>CaTy</i> und <i>CaDis</i>	171
D.III.1.1.	<i>Ty1-copia</i> -Sequenzen im Genom von <i>C. arietinum</i>	171
D.III.1.2.	<i>CaTy</i> -Sequenzen aus Kichererbse sind sehr heterogen	172
D.III.1.3.	<i>CaDis</i> -Sequenzen weisen eine höhere Homogenität auf	172
D.III.1.4.	Chromosomale Lokalisation von <i>CaTy</i> und <i>CaDis</i>	173
D.III.1.5.	Vermutliche Evolution der beiden LTR-Retrotransposons <i>CaDis</i> und <i>CaTy</i>	174

D.III.1.6.	Gibt die Transkription von <i>CaDis</i> -Sequenzen einen Hinweis auf die Aktivität des Elements?	175
D.III.2.	Die non-LTR Retrotransposonfamilie <i>CaLin</i>	177
D.III.2.1.	<i>CaLin</i> -Sequenzen weisen die für LINEs typische Sequenzheterogenität auf	177
D.III.2.2.	Die niedrige Abundanz von <i>CaLin</i> wird vermutlich durch seine Heterogenität bedingt	178
D.III.2.3.	<i>CaLin</i> -Sequenzen liegen teilweise in euchromatischen Bereichen	178
D.III.2.4.	<i>CaLin</i> ist vermutlich eine sehr ursprüngliche Retrotransposonfamilie in Leguminosen und zeigt innerhalb der Gattung <i>Cicer</i> keine Veränderung seiner Organisation	179
D.IV.	Nachweis und Charakterisierung des DNA-Transposons <i>CaEn/Spm</i>	180
D.IV.1.	Ein PCR-Ansatz zum Nachweis und zur Isolation von <i>En/Spm</i> -ähnlichen Transposonsequenzen aus Pflanzengenomen	180
D.IV.2.	Eine Transposonfamilie der Kichererbse: <i>CaEn/Spm</i>	181
D.IV.2.1.	Struktur und Abundanz <i>En/Spm</i> -ähnlicher Elemente im Genom von <i>C. arietinum</i>	181
D.IV.2.2.	Gibt es Hinweise auf die Funktionalität von <i>CaEn/Spm</i> -Kopien?	182
D.IV.2.3.	Verbreitung von <i>CaEn/Spm</i> -homologen Sequenzen in verwandten Genomen	184
D.IV.2.4.	Chromosomale Lokalisation der <i>CaEn/Spm</i> -Elemente	184
D.V.	Repetitive Elemente im Genom von <i>C. arietinum</i> - Grundlage für ein vorläufiges Chromosomenmodell von <i>C. arietinum</i>	186
D.V.1.	Übersicht über die Lokalisation der untersuchten repetitiven Familien	186
D.V.2.	Heterochromatische Lokalisation und weitgehende Inaktivität - Resultat eines Abwehrmechanismus durch das Wirtsgenom?	188
D.V.3.	Hinweise auf eine Evolution der untersuchten Elemente	189
D.V.4.	Die Verteilung repetitiver Sequenzen im Kichererbsengenom entspricht eher dem <i>A.thaliana</i> -Genom als Genomen vergleichbarer Größe	190
D.V.5.	Die Groborganisation der repetitiven DNA und die Genomanalyse in <i>C. arietinum</i>	192
E.	ZUSAMMENFASSUNG	195
E.	Summary	197
F.	LITERATUR	199
G.	PUBLIKATIONEN	229
H.	ABKÜRZUNGEN	231

A. EINLEITUNG

A.I. Repetitive DNA

Die Genome höherer Pflanzen sind in ihrer Komplexität sehr variabel und unterscheiden sich in ihrer Größe um mehrere Zehnerpotenzen. *Arabidopsis thaliana* mit ca. 0.2 pg pro haploidem, unrepliziertem Genom (1C) und etwa 125 Mbp steht Genomen von über 120pg /1C oder 16000 Mbp (*Allium cepa*), 30000 Mbp (*Tulipa* sp.) bzw. 123000 Mbp (*Fritillaria assyrica*) gegenüber (Bennett & Smith, 1976; Arumuganathan & Earle, 1991), obwohl alle Angiospermen einen vergleichbar hohen Entwicklungsstand aufweisen und vermutlich etwa 15000-25000 Gene besitzen (Vedel & Delseny, 1987; The *Arabidopsis* Genome Initiative, 2000). Dieses Phänomen wurde schon vor 30 Jahren als das C-Wert Paradoxon beschrieben (Thomas, 1971). Tatsächlich machen genische Sequenzen in den meisten Pflanzenarten nur wenige Prozent des Kerngenoms aus und sind über weite taxonomische Bereiche colinear angeordnet. Der Hauptanteil besteht aus repetitiven Sequenzen, deren 2 bis über 10000 Bp langen Motive mehrere hundert- bis hunderttausendmal im Genom vorliegen und somit hauptsächlich für die unterschiedlichen Genomgrößen verantwortlich sind. Je nach Struktur und Sequenz kann man verschiedene repetitive Familien unterscheiden, die entweder tandemartig organisiert, meist wenige, definierte chromosomale Abschnitte einnehmen oder dispers mehr oder weniger gleichmäßig im gesamten Genom verteilt sind. Obwohl weitgehend alle Hauptklassen repetitiver DNA (Desoxyribonukleinsäure) in allen Pflanzen vorhanden sind, variieren die unterschiedlichen Familien sowohl in Sequenz als auch in ihrer genomischen Lokalisation zwischen nahe verwandten Spezies beträchtlich (Bennetzen, 2000). In großen Genomen existieren oft viele hochabundante repetitive Familien und können bis zu 50% der Kern-DNA ausmachen (SanMiguel et al., 1996), während das kleine Genom von *A. thaliana* nur wenige hochabundante Elemente aufweist und die übrigen repetitiven Familien auf niedrige Kopienzahlen begrenzt sind (The *Arabidopsis* Genome Initiative, 2000). Weshalb einige repetitive Sequenzfamilien in manchen Genomen zu hohen Kopienzahlen amplifiziert wurden, nicht jedoch in anderen, ist unbekannt. Durch Sequenzierungsprojekte, *in-situ* Hybridisierung und Southern-Analysen ist die Organisation von repetitiven Sequenzen in verschiedenen Pflanzengenomen bekannt, so daß ein pflanzliches Chromosomenmodell erstellt werden konnte (Schmidt & Heslop-Harrison, 1998). Während bei Zuckerrübe, Weizen, Gerste und Roggen der größte Teil des Chromosoms von repetitiven Sequenzfamilien mit familienspezifischer Lokalisation bedeckt ist und Gene sich in Clustern dazwischen und besonders an den Chromosomenenden befinden, zeigte die Sequenzierung von *A. thaliana*, daß hier die meisten Repeats im pericentrischen Bereich konzentriert sind (Heslop-Harrison et al., 1997; The *Arabidopsis* Genome Initiative, 2000).

In den letzten 20 Jahren gewann die Struktur, Organisation und Evolution repetitiver DNA zunehmend an Interesse, so daß das Bild der egoistischen „junk“-DNA (Doolittle & Sapienza, 1980; Orgel & Crick, 1980), die sich um ihrer selbst willen vermehrt und propagiert wird,

solange sie nicht nachteilig auf das Genom wirkt, langsam verblaßt. Vielmehr zeigt sich, daß repetitive Sequenzen eine signifikante Rolle in der Evolution von Genomen und Spezies spielen, besonders im Hinblick auf Chromosomenstruktur und Genexpression. Die Genomkartierung und -sequenzierung bei Pflanzen hat zu einem Verständnis von Funktion und Regulation pflanzlicher Gene geführt, das den Kenntnissen über Organisation und Funktion repetitiver DNA weit voraus ist. Um Struktur, Organisation und Evolution von pflanzlichen Genomen zu verstehen, ist daher die Analyse von repetitiver DNA unerlässlich.

A.I.1. Satelliten-DNA

Nach ihrer Organisation können zwei große Klassen repetitiver DNA unterschieden werden. Die erste Klasse bildet sog. Tandemabfolgen, d.h. räumlich definierte Sequenzblöcke aus repetitiven Grundeinheiten, die ohne Zwischenraum hintereinander angeordnet sind. Zu den mittelrepetitiven Elementen dieser Gruppe zählen die **rRNA-Gene**, zu denen die 18S-5.8S-25S rDNA in der NOR (Nukleolus Organisator Region) und die davon getrennt lokalisierten 5SrRNA-Gene gehören. Die codierenden Sequenzen der reifen rRNAs sind hochkonserviert, während die Längen und Sequenzen der intergenischen Spacer (IGS) zwischen verschiedenen Spezies stark divergieren (Hemleben, 1993). Ebenfalls tandemartig angeordnet sind die 4-10 Bp langen **Telomersequenzen**, die spezifische Proteine binden und die Enden eukaryontischer Chromosomen vor Degradation und Fusion schützen. Dazu fügt die sog. Telomerase bei den meisten Pflanzen das Motiv 5`-CCCTAAA-3` komplementär zu einem internen RNA (Ribonukleinsäure)-Templat *de novo* an die Chromosomenenden an (Biessman & Mason, 1994). Je nach Länge der repetitiven Einheiten kann man ferner **Mikro-** und **Minisatelliten** mit Wiederholungseinheiten von 1-10 bzw. 10-100 Bp unterscheiden (Jeffreys et al., 1985; Litt & Luty, 1989). Beide Sequenzklassen sind in eukaryontischen Genomen ubiquitär verbreitet und meistens niedrig bis mittelrepetitiv und dispers im Genom verstreut (Weising et al., 1995).

Als **Satelliten-DNA** galt ursprünglich repetitive DNA, die im CsCl-Dichtegradienten eine von der Hauptfraktion der Kern-DNA getrennte Bande erzeugte, da sie aufgrund der hohen Abundanz ihrer Sequenzmotive einen abweichenden GC-Gehalt aufweist (Barnes et al., 1985). Allerdings beinhalten diese Satellitenbanden oft auch rRNA-Gensequenzen (Hemleben et al., 1977). Heute umfaßt diese Nomenklatur hochrepetitive, tandemartig angeordnete Sequenzen mit Einheiten bis zu mehreren Tausend Basenpaaren, d.h. auch die kryptischen Satelliten, die aufgrund ihres GC-Gehalts in der Hauptbande des CsCl-Gradienten liegen. Im Gegensatz zu Mini- und Mikrosatelliten kann die Satelliten-DNA mit 10^3 bis 10^7 Kopien pro Locus mehrere Prozent der Kern-DNA ausmachen. Monomerlängen von 160-180 oder 350-360 Bp sind bei pflanzlichen Satelliten am häufigsten (Harrison & Heslop-Harrison, 1995; Schmidt & Heslop-Harrison, 1996; Alix et al., 1998; Ananiev et al., 1998a; Heslop-Harrison et al., 1999), es wurden aber auch Monomere von über Tausend Basen detektiert (Hagemann et al., 1993). Die ubiquitär im Pflanzenreich vorhandenen Satelliten sind meist im subtelomerischen oder pericentrischen Heterochromatin konzentriert und können auf

einzelnen, mehreren oder allen Chromosomen verteilt sein (Kamm et al., 1994; Vershinin et al., 1996; De Jeu et al., 1997). Diese Sequenzen sind oft an den für pflanzliche Methylasen zugänglichen Motiven CpG oder CpNpG methyliert. So liegt bei einigen Cucurbitaceen fast jedes dritte Cytosin als 5-Methylcytosin vor (Hemleben et al. 1982). Entsprechend sind Satellitensequenzen größtenteils transkriptionell inaktiv und in der Interphase heterochromatisch organisiert. Die Lage im konstitutiven Heterochromatin bedingt eine späte Replikation in der S-Phase.

A.I.1.1. Evolution und Amplifikation von Satelliten-DNA

Verschiedene Mechanismen sind an der Evolution und Amplifikation von Satelliten-DNA beteiligt. Satellitensequenzen können durch **Strangverrutschen bei der Replikation („slipped strand mispairing“)** amplifiziert werden (Levinson & Gutman, 1987): homologe Sequenzwiederholungen nahe der Replikationsgabel können zu einem Verrutschen („slipping“) von Matrizenstrang und neusynthetisiertem Strang gegeneinander führen, wobei Fehlpaarungen („mispairing“) entstehen (Singer & Berg, 1992). Bei deren Reparatur wird die Sequenzabfolge um wenige Einheiten verlängert oder verkürzt, was z.B. zur Bildung des Satelliten beitragen kann. **Ungleiches „Crossing-over“** kann bei der Rekombination zwischen Schwesterchromatiden oder Chromatiden homologer bzw. nicht-homologer Chromosomen durch Überkreuzung und Bruchstücktausch nicht-alleler, aber homologer Untereinheiten einer Satellitenabfolge entstehen (Smith, 1976). Bei höheren Ordnungen von Wiederholungseinheiten kann dies zur Deletion bzw. Duplikation langer Abschnitte führen. Nach Smith et al. (1976) führt dieser Mechanismus bei Sequenzen, die keiner Selektion unterliegen, zu einer repetitiven Tandemstruktur. Stephan (1989) zeigte durch Simulationen, daß mit niedrigerer Rekombinationsrate - wie etwa in den heterochromatischen Bereichen der Satelliten-DNA - die Komplexität der Substrukturen und die Heterogenität einer Tandemabfolge zunehmen. Ein weiterer Amplifikationsmechanismus ist die sog. **„rolling circle“-Amplifikation**, bei der ein zirkulärer Doppelstrang extrachromosomal repliziert wird (Walsh, 1987). Plasmide mit Tandemwiederholungen stellen möglicherweise ein Templat für eine solche Replikation dar (Okumura et al., 1987). Allerdings konnte bisher keine Reinsertion einer linearen Sequenz beobachtet werden. Möglicherweise sind auch Transpositionereignisse an der Amplifikation von Satelliten beteiligt (Vedel & Delseny, 1987). Im Wechsel mit Amplifikationsereignissen auftretende **Mutationen** (Insertionen, Deletionen, Inversionen und Punktmutationen) sorgen für eine schnelle Divergenz der einzelnen Einheiten sowie die Entstehung höherer Strukturhierarchien (Willard & Waye, 1987a; Morgante et al., 1997). Dem steuern Homogenisierungsmechanismen entgegen, wie etwa die **Genkonversion**, bei der Heteroduplexe aus nichtidentischen Tandemwiederholungen Reparaturmechanismen unterworfen werden, so daß homologe, komplementäre Sequenzen entstehen (Kourilsky, 1986). Auch Rekombinationsereignisse (ungleiches „crossing-over“) können zur Homogenisierung (Smith, 1976) und konzertierten Evolution bestimmter Genombereiche beitragen. Die genannten Divergenz- und

Homogenisationsmechanismen machen Satelliten-DNA zu einer dynamisch evolvierenden Komponente des Genoms. Trotzdem können Satelliten innerhalb bestimmter taxonomischer Gruppen in ihrer Sequenz, chromosomalen Lokalisation und Abundanz konserviert sein (Dover, 1986). Je nach Evolutionsdynamik der Genombereiche können unterschiedlich konservierte Sequenzen von familien-, gattungs- oder speziesspezifischen und genom-spezifischen Satellitenfamilien zusammen innerhalb eines Genoms vorliegen (Hemleben 1993; King et al., 1995; Nagaki et al., 1995; Macas et al., 2000). Einige pflanzliche Satellitenfamilien divergieren selbst zwischen verschiedenen Chromosomen (Cuadrado & Jouve, 1997; Zhao et al., 1998), wie auch die α -Satelliten beim Menschen (Willard & Waye, 1987b). Solche Satelliten können als molekulare Marker in der Pflanzenzüchtung eingesetzt werden, etwa zur Identifikation der Genome in somatischen Hybriden (Schweizer et al., 1988; Schmidt et al., 1990; Fahleson et al., 1997; Zahnke & Hemleben, 1997) oder fremder Chromosomen in monosomatischen Additionslinien (Schmidt et al., 1990; Schmidt & Heslop-Harrison, 1996; Desel, 2001).

Unter einer Additionslinie versteht man eine Zelllinie bzw. eine Linie von Individuen, die zusätzlich zu ihrem natürlichen Chromosomensatz weitere Chromosomen oder Chromosomenarme besitzen. Eine monosomische Additionslinie enthält dementsprechend ein zusätzliches fremdes Chromosom.

Die vergleichende Charakterisierung verschiedener Satelliten innerhalb eines Genoms gibt Einblick in dessen Organisation und Evolution. Die Konservierung bestimmter Sequenzfamilien zwischen verschiedenen Spezies einer Gattung oder Familie läßt Rückschlüsse auf Verwandtschaftsbeziehungen zu und kann phylogenetische oder taxonomische Fragen klären (Hemleben, 1993; Ingham et al., 1993; Kamm et al., 1995).

Die Entstehung von Satelliten-DNA ist vermutlich aus jeder beliebigen Sequenz möglich. Homologievergleiche legen eine Abstammung aus tRNA- (Benslimane et al., 1986), oder rRNA-Sequenzen (Wu & Wu, 1987; Maggini et al., 1991) nahe, während z.B. auch 30 Bp lange Wiederholungseinheiten unbekannter Herkunft dupliziert, trimerisiert und schließlich amplifiziert wurden (Ingham et al., 1993).

Satelliten-DNA läßt sich meist durch Behandlung mit Restriktionsendonukleasen und nachfolgender Southern-Hybridisierung durch ein charakteristisches Leiterbandenmuster nachweisen, weshalb man oft auch von „Restriktionssatelliten“ spricht (Hemleben, 1993). Wechselnde Zyklen von Amplifikation und Divergenz lassen konservierte Restriktionsschnittstellen in einem Teil der Wiederholungseinheiten degenerieren, so daß bei der Restriktion neben Monomeren auch Multimere entstehen. Restringierte genomische DNA zeigt daher nach der Hybridisierung ein Leitermuster, aus dem die Länge der repetitiven Einheit abgeleitet werden kann (Fig.A1).

Die zweite Gruppe bilden **disperse repetitive** Elemente, deren mittel- oder hochrepetitive Einheiten im Genom zwischen singulären Sequenzen oder zwischen repetitiven Elementen eingestreut sind (Smyth, 1991). Ein Großteil von ihnen stellt mobile genetische Elemente („transposable Elemente“) oder Relikte von solchen dar (Flavell, 1986). Diese Elemente

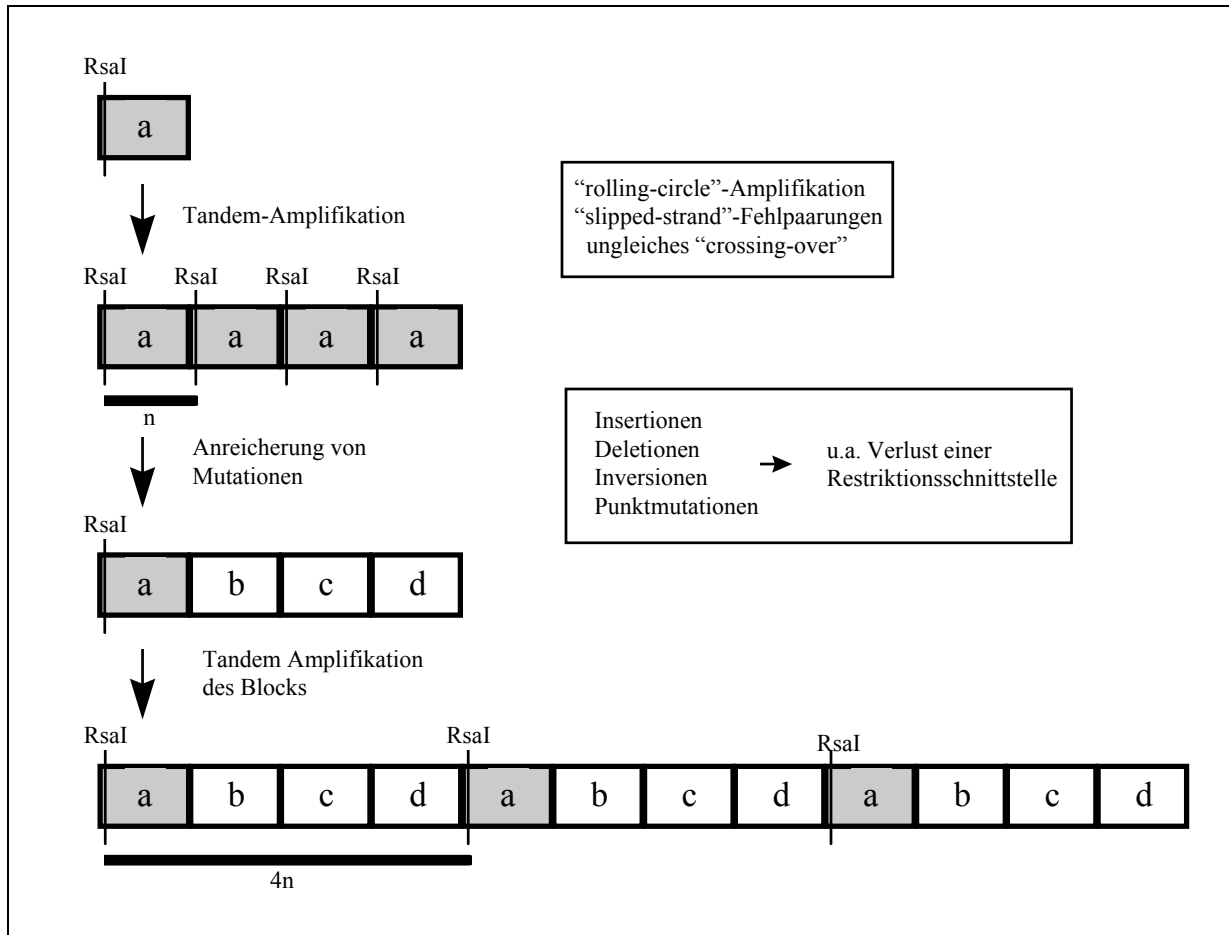


Fig.A1: Evolution einer Tandemfolge und Entstehung des charakteristischen Leiterbandenmusters. Konservierte Restriktionsschnittstellen in einzelnen Monomeren (a) können durch Divergenz verlorengehen, so daß bei weiteren Amplifikationszyklen übergeordnete Einheiten (a-d) entstehen. Nach der entsprechenden Restriktion entstehen neben Monomeren auch Multimere (hier ein Tetramer), die die charakteristische Abfolge von Banden („Leiterbanden“) bilden.

können gleichmäßig im gesamten Genom verteilt sein oder in bestimmten Regionen Cluster bilden. Transposable Elemente lassen sich je nach Art der Transposition in zwei Gruppen einteilen (Klasse I und II). Einige transposable Elemente können nicht einer der beiden Klassen zugeordnet werden, da ihr Transpositionsmechanismus unklar ist. Zu diesen gehören die kurzen (100-500 Bp), **dispersen MITEs** („miniature inverted-repeat transposable elemente“), die TIRs („terminal inverted repeats“) besitzen und zuerst in Gräsern beschrieben wurden (Bureau & Wessler, 1992; 1994). Da bisher keine codierten Proteine nachgewiesen werden konnten, geht man davon aus, daß MITEs wie *Tourist* oder *Stowaway* nicht-autonome Elemente sind, die durch Transposasen anderer Elemente der Klasse II bewegt werden

können (Wessler et al., 1995). Allerdings liegen MITEs in weit höheren Kopienzahlen vor als DNA-Transposons, so daß sie vermutlich einen anderen Mechanismus als die beschriebene direkte Exzision / Insertion zur Transposition verwenden. Im Rahmen dieser Arbeit wurden jedoch keine MITEs analysiert.

A.I.2. Klasse I: Retrotransposons

Elemente der Klasse I werden auch Retrotransposons genannt und gehören mit den Retroviren zu den Retroelementen. Sie wurden zuerst in Tiergenomen und in Hefe gefunden und transponieren durch Reverse Transkription eines RNA-Intermediats. Retrotransposons sind ubiquitär in Pflanzen vorhanden (Flavell et al., 1992a; Voytas et al., 1992; Kumar & Bennetzen, 1999). Ihr replikativer Transpositionsmechanismus führt bei jeder Transposition im Genom zu einer Erhöhung der Kopienzahl, so daß sie große Teile des Genoms ausmachen können (z.B. bis zu 50% in *Zea mays*; SanMiguel et al., 1996). Auch hier lassen sich abhängig von Struktur und Transpositionsmechanismus zwei Gruppen unterscheiden.

A.I.2.1. LTR-Retrotransposons

LTR-Retrotransposons werden von langen terminalen Sequenzwiederholungen („long terminal repeats“, LTRs) flankiert und haben daher eine hohe Ähnlichkeit zu Retroviren. Die 0.1-5 Kbp langen LTRs werden von der kurzen inversen Sequenz 5`-TG...CA-3` begrenzt und enthalten regulatorische Elemente für Promotion und Termination der Transkription sowie deren Initiations- und Endsignale. Im zentralen Teil befinden sich offene Leseraster, die für die strukturellen und enzymatischen Proteinkomponenten codieren: *gag* (group-specific antigen) codiert für die Proteine zur Reifung sowie Verpackung von Retroelement-RNA und Enzymen in eine Kapsidstruktur („virus-like particle“- VLP), während *pol* (Polyprotein) die Gene für Protease (Prot), RNase H, Reverse Transkriptase (RT) und Integrase/Endonuklease (Int oder Endo) beinhaltet (Fig.A2). Reverse Transkriptasen verschiedener Retrotransposons und retroviraler Sequenzen zeigen insgesamt sieben konservierte Peptidregionen, die zusammen 178 Aminosäuren umspannen (Xiong & Eickbush, 1990) und aus denen Oligonukleotidprimer zur Amplifikation verschiedener Retroelementfamilien abgeleitet werden können (Flavell et al., 1992b; Wright et al, 1996). *gag* und *pol* können entweder von demselben oder von verschiedenen Leserastern codiert sein, d.h. ein Leserasterwechsel oder eine Translationsreinitiation müssen erfolgen, um die Expression der 3'-codierten Komponenten zu garantieren. *gag* und *pol* ermöglichen eine autonome Transposition des Elements, jedoch sind für seine Aktivität intakte regulatorische Sequenzen in den LTRs Voraussetzung (Kumar & Bennetzen, 1999). Einige pflanzliche Elemente weisen als dritte Funktion *env* („envelope“) auf, die eine Proteinhülle generiert (Fig.A2). Sie befähigt das Element zur Bildung eines extrazellulären, infektiösen Virions, was die Grenzen zwischen Retrotransposon und Retrovirus fließend macht (Bennetzen, 1996; Kumar, 1998). Allerdings konnten bisher im Gegensatz zu tierischen Genomen bei Pflanzen keine infektiösen Retroviren beobachtet werden (Kumar & Bennetzen, 1999). Zwischen den

Polyproteinleserastern und den LTRs liegen (1) eine kurze Sequenz (PBS; „primer binding site“), die komplementär zum 3'-Ende zellulärer tRNAs ist, und (2) eine weitere purinreiche Sequenz (PPT; „polypurin tract“). Beide dienen als Primerbindestelle für die Synthese des + und - Strangs (Fig.A2).

Ty1-copia Elemente sind im gesamten Pflanzenreich verbreitet und konnten in Vertretern der Bryophyten, Gymnospermen und Angiospermen nachgewiesen werden (Flavell et al., 1992b; Voytas et al., 1992). Auch *Ty3-gypsy* Elemente sind ubiquitär anzutreffen (Sentry & Smyth, 1989; Nozue et al., 1997; Suoniemi et al., 1998; Thomson et al., 1998; Kossack & Kinlaw, 1999; Kumekawa et al., 1999). Sequenzfamilien beider Gruppen treten in großen Genomen oft in hohen Kopienzahlen auf, z.T. bis zu einigen Millionen, und sind mehrere Kilobasen lang (Leeton & Smyth, 1993; Pouteau et al., 1994; Bennetzen, 1996; Kumar, 1996; San Miguel et al., 1996; Kunze et al., 1997). Hochabundante Vertreter sind z.B. die *Ty1-copia* Elemente *BARE1* (12.1 Kbp) aus *Hordeum vulgare* mit 50 000 Kopien (Suoniemi et al., 1996a), *Opie2* (8.7 Kbp) und *Prem2* (9.5 Kbp) mit 30 000 bzw. 10 000 Kopien aus *Zea mays* (SanMiguel et al., 1996, Turcich et al., 1996) oder die *Ty3-gypsy* Retrotransposons *Cinfull* (8.6 Kbp) aus *Zea mays* mit 20 000 und *IFG* (5.9 Kbp) aus *Pinus radiata* mit 10 000 Kopien (SanMiguel & Bennetzen, 1998; Kossack & Kinlaw, 1999). *Ty1-copia* Elemente sind meist dispers im Euchromatin verstreut und in einigen Regionen ausgespart oder in ihrer Kopienzahl reduziert (Moore et al., 1991; Brandes et al., 1997b; Heslop-Harrison et al., 1997; Pearce et al., 1997), jedoch gibt es auch Ausnahmen mit Anhäufungen im Heterochromatin des Telomerbereichs oder des Centromers (Pearce et al., 1996b; Pelissier et al., 1996; Heslop-Harrison et al., 1997). *Ty3-gypsy* Elemente sind bei Gräsern von centromerischen Regionen her bekannt, wie etwa *cereba* (*Hordeum vulgare*; Presting et al., 1998) oder *Grande* und *Zeon*, die in den Heterochromatinköpfen der Pachytänchromosomen von *Zea mays* lokalisiert sind (Ananiev et al., 1998a). Aus der genomischen Sequenz von *A. thaliana* ist bekannt, daß LTR-Elemente im pericentrischen Heterochromatin angereichert sind (The *Arabidopsis* Genome Initiative, 2000), jedoch konnte diese Sequenzklasse auch oft in und nahe bei Genen identifiziert werden (SanMiguel et al., 1996). Sowohl intergenisch als auch im Heterochromatin bilden mehrere Retrotransposonfamilien oft komplexe, verschachtelte Serien aus ineinandergesprungenen Kopien (SanMiguel et al., 1996; The *Arabidopsis* Genome Initiative, 2000).

A.I.2.1.1. Die Transposition von *Ty1-copia*- und *Ty3-gypsy*-Elementen

Abhängig von der Reihenfolge der Polyproteingene (*pol*) und der Aminosäuresequenz konservierter Bereiche (z.B. RT) können zwei Familien der LTR- Retrotransposons unterschieden werden: *Ty3-gypsy* ähnliche Elemente, die mit Ausnahme von *env* die gleiche Struktur wie Retroviren aufweisen, und *Ty1-copia* Elemente, deren Endonukleasefunktion von Sequenzen 5' der RT codiert wird (Kumar & Bennetzen, 1999) (Fig.A2). Die Transposition folgt in beiden Familien sowie in Retroviren dem gleichen Schema: Das Element wird in voller Länge transkribiert. Damit wird sowohl mRNA für die codierten

Komponenten als auch ein RNA-Templat für eine weitere Kopie geliefert. Der Transkriptionsstart liegt zwischen U3 („unique 3'-RNA region“)- und R („repeated RNA region“)-Region des 5'-LTRs, das Transkript endet zwischen R- und U5 („unique 5'-RNA region“)-Region des 3'-LTRs (Fig.A2). Die Promotorsequenzen für die Transkriptionsinitiation durch Polymerase II und die Signale für das 5'-und 3'-Prozessieren der RNA befinden sich im U3-Bereich beider LTRs, d.h. müssen in jeweils einem LTR supprimiert werden (Boeke & Corces, 1989; Manninen & Schulmann, 1993). Auch das typische Polyadenylierungssignal AATAAA bzw. GATAAA ist in der U3-Region der LTRs zu finden und bewirkt eine Polyadenylierung ca. 20-30 Bp stromabwärts im 3'-LTR. Im Cytoplasma wird ein virusähnliches Partikel („virus-like partikel“; VLP) aus *gag*-Komponenten und RNA zusammengesetzt, das ebenfalls die zur Reversen Transkription und Integration notwendigen Proteine beinhaltet. Im VLP wird zelluläre tRNA als Primer für die Reverse Transkriptase an die PBS angelagert und die Reverse Transkription initiiert, wobei Chaperondomänen wie auch Cofaktoren mit der RT assoziiert sind (Kumar & Bennetzen, 1999; Cristofari et al., 2000; Lewin, 2000). Gleichzeitig katalysiert die Protease die Reifung des Polyproteins zu den einzelnen funktionellen Peptiden.

Die Reverse Transkription der fehlenden LTR-Bereiche wird durch temporäre Zirkularisierung des RNA-Templats erreicht, da kurze LTR-Bereiche am 5'- und 3'-Ende identisch sind. Die neue Kopie verfügt so später über zwei identische LTRs (Fig. A2). Die Synthese des zweiten DNA-Strangs erfolgt nach Entfernung der Templat-RNA durch die RNase H mit Hilfe der zweiten Primerbindestelle (PPT). Das VLP wird in den Nukleus transferiert, wo das doppelsträngige DNA-Fragment schließlich durch einen versetzten Doppelstrangsschnitt der Integrase und das Reparatursystem der Zelle in die genomische DNA integriert wird. Außerdem findet eine 3-5 Bp lange Zielsequenzverdoppelung statt (TSD; „target site duplication“).

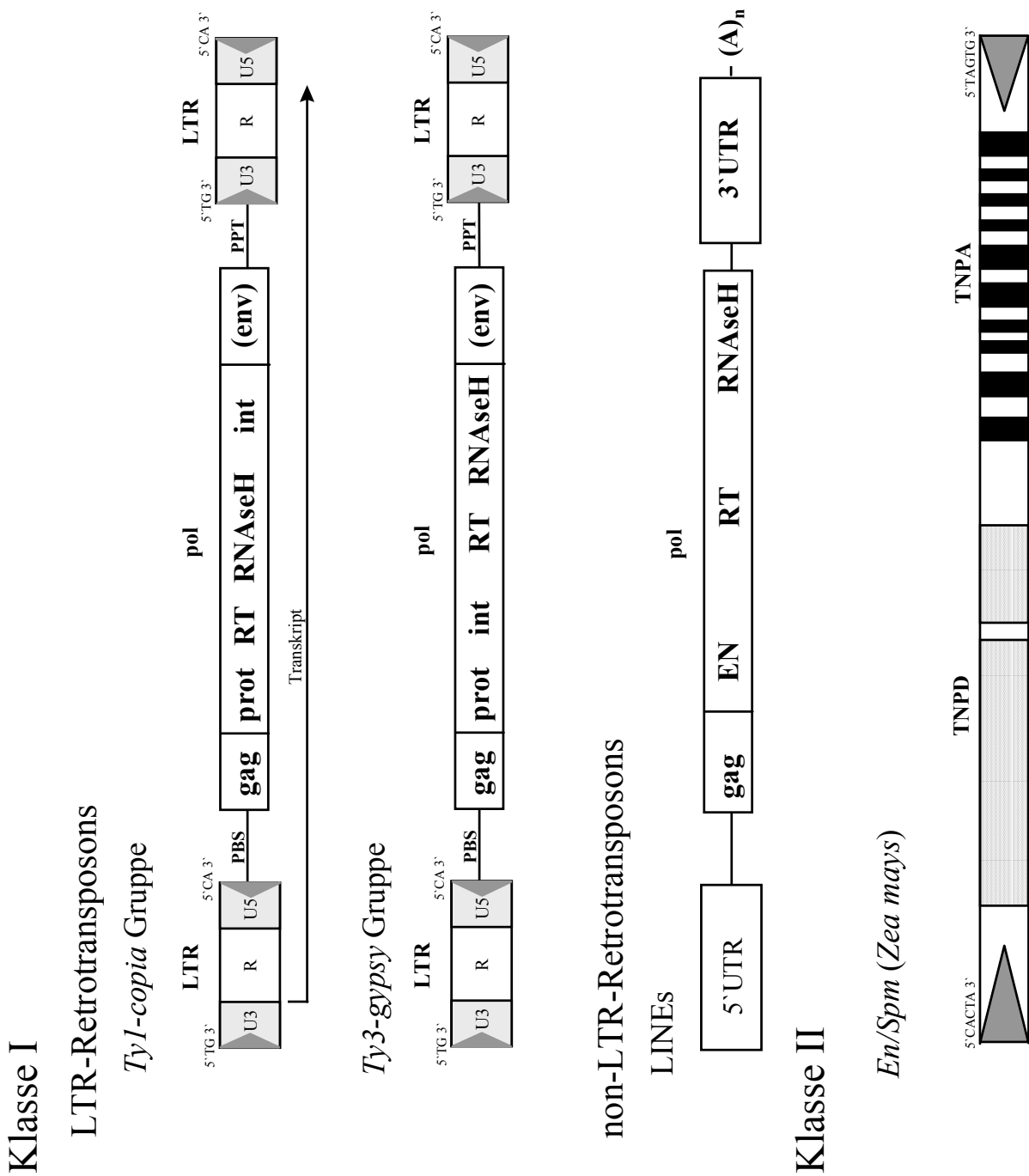
Fig.A2: Allgemeine Struktur von *Ty1-copia*-, *Ty3-gypsy*- und LINE-Retrotransposons sowie *En/Spm*-ähnlicher Transposons der Klasse II.

LTR-Retrotransposons werden von langen terminalen Wiederholungseinheiten (LTRs) flankiert, die durch kurze inverse Sequenzmotive 5'-TG-3' / 5'-CA-3' begrenzt werden. Die U3-, U5- („unique 5'/3'-RNA“) und R-Region („repeated RNA“) enthält Signale für Transkriptionsinitiation und -termination. Länge und Position des Transkripts sind durch einen Pfeil gekennzeichnet. Die *gag*-Gene codieren für Proteine zur Reifung und Verpackung der RNA in eine Kapsidstruktur, das *pol*-Polyprotein für Protease (prot), RNase H, Reverse Transkriptase (RT) und Integrase (int). Einige Elemente enthalten ein drittes Leseraster mit Ähnlichkeit zu retroviralen Hüllproteinen (env). *Ty1-copia*- und *Ty3-gypsy* Elemente unterscheiden sich in der Abfolge der Polyproteinfunktionen. PBS („primer binding site“) und PPT („polypurine tract“) kennzeichnen die Primerbindestellen für Reverse Transkription und Doppelstrangsynthese.

Non-LTR-Elemente der LINE-Gruppe weisen ORFs für *gag* und *pol* auf, letzterer mit Funktionen für Endonuklease (EN), Reverse Transkriptase (RT) und RNase H. Flankierend liegen 5'- und 3'-untranslatierte Regionen (5'-/3'-UTR) und am 3'-Ende ein poly(A)-Abschnitt.

En/Spm-ähnliche Elemente werden von terminalen, inversen Wiederholungseinheiten (TIRs) begrenzt, die mit der Sequenz 5'-CACTA-3' beginnen. Die internen Leseraster für die Proteine TNPD (Transposaseaktivität) und TNPA (Regulatorprotein) sind von Introns unterbrochen.

Die Darstellung ist nicht maßstabsgetreu. Nicht dargestellt sind die in allen Fällen durch die Insertion erzeugten, flankierenden Verdoppelungen der Zielsequenz.



A.I.2.2. Die non-LTR-Retrotransposons

Zu den non-LTR-Retrotransposons gehören LINEs („long interspersed nuclear elements“) und SINEs („short interspersed nuclear elements“). LINEs können mehrere Kilobasen lang und sehr variabel in ihrer Struktur sein. Sie weisen im Gegensatz zu den LTR-Retrotransposons keine begrenzenden LTRs auf, sondern besitzen am 3`-Ende einen poly(A)-Abschnitt. Zwei interne ORFs (offene Leseraster; „open reading frame“) codieren für *gag* bzw. für Reverse Transkriptase, Endonuklease und RNase H (*pol*). Beide ORFs können mit einem Leserasterwechsel überlappen und ermöglichen eine autonome Transposition (Fig.A2). Eine Integrase konnte bisher noch nicht identifiziert werden (Kumar & Bennetzen, 1999). Die Transkription führt zu einer polyadenylierten mRNA in voller Länge des Elements, die der selbstcodierten Reversen Transkriptase als Templat dient (Boeke, 1997).

Die Reverse Transkriptase weist wie alle retroviralen DNA-Polymerasen (auch die der LTR-Elemente) sieben konservierte Domänen auf (Xiong & Eickbush, 1990). Flankierend zu den ORFs befinden sich 5`- und 3`-untranslatierte Bereiche, die im 5`-Bereich Promotorsequenzen für die Transkription beherbergen. SINEs sind im Gegensatz dazu nur mehrere hundert Basenpaare lang und besitzen keine codierenden Kapazitäten. Wie die LINEs besitzen sie einen poly(A)-Abschnitt im 3`-Bereich. Der 5`-Teil zeigt oft Homologien zu tRNAs oder 7SL („spliced leader“-)RNA-Sequenzen, die bei Tieren an Spleißprozessen beteiligt sind (Schmidt, 1999).

Im menschlichen Genom und anderen Säugern machen non-LTR-Elemente, besonders *L1* (LINE) und die *Alu*-Familie (*SINE*) bis zu 35 % der gesamten DNA aus (*Alu* stellt mit 10^6 Kopien fast 10% des menschlichen Genoms dar). *Caenorhabditis elegans* enthält beispielsweise mehr non-LTR- als Retroelemente mit LTRs. LINEs wurden auch in Wirbellosen und Pilzen gefunden (z.B. *R, I, Jockey* in *Drosophila melanogaster*), aber es gibt auch Organismen, die diese repetitive Klasse nicht enthalten, wie z.B. *Saccharomyces cerevisiae* (Kumar & Bennetzen, 1999). Obwohl im Pflanzenreich noch nicht viele LINE-ähnliche Retrotransposons detailliert untersucht wurden, sind sie in verschiedenen, phylogenetisch weit auseinanderliegenden Pflanzengenomen nachgewiesen worden, so daß man davon ausgehen kann, daß diese Elementgruppe genauso wie LTR-Retrotransposons ubiquitär in Pflanzen vorhanden ist (Wright et al., 1996; Kubis et al., 1998a; Noma et al., 1999). LINEs konnten sogar im mitochondrialen Genom von *A. thaliana* nachgewiesen werden (Knoop et al., 1996; Wright et al., 1996). Während viele Elemente in mittlerer Abundanz vorliegen, wie *Cin4* aus *Zea mays* mit 6.6 Kbp Länge und ca. 50-100 Kopien

(Schwarz-Sommer et al., 1987), *Tall-1* aus *Arabidopsis thaliana* (6.1 Kbp) mit 6 Kopien oder das 8.9 Kbp lange *Zepp* aus der Grünalge *Chlorella vulgaris* mit ca. 130 Kopien (Higashiyama et al., 1997), können LINEs aber auch bedeutende Genomkomponenten darstellen: etwa das 4.5 Kbp lange, hochabundante *del2* aus *Lilium speciosum* mit 250 000 Kopien (Leeton & Smyth, 1993), das 4% des Genoms ausmacht. LINEs aus der Gattung *Beta* stellen sehr heterogene Elementfamilien dar (Schmidt et al., 1995; Kubis et al., 1998a). Viele pflanzliche LINE-Elemente sind im 5' Bereich deletiert (Kumar & Bennetzen, 1999). Im Gegensatz zu den bereits bekannten LINE-Familien konnten bisher nur wenige SINEs in Pflanzen identifiziert werden, darunter *pSINE1* aus *Oryza sativa* mit 125 Bp und ca. 130 Kopien (Mochizuki et al., 1992), *TS* aus *Nicotiana tabacum* (Yoshioka et al., 1993) mit 111 Bp und 50 000 Kopien oder das *SI* SINE mit 170 Bp Länge und ca. 500 Kopien aus *Brassica napus* (Deragon et al., 1996). Non-LTR-Elemente sind meist dispers im Genom verstreut, teilweise mit Clustern in subtelomerischen (Kubis et al., 1998a), telomerischen (Higashiyama et al., 1997) oder centromerischen (Goubely et al., 1999) Regionen. Diese Cluster entstehen im Fall von *Zepp* durch Insertion von Elementen ineinander (Noutoshi et al., 1998). Die Loci der rDNA-Gene werden von einigen Elementen ausgespart (Schmidt et al., 1995, Kubis et al., 1998a), während andere bevorzugt mit rDNA-Loci assoziiert sind (Goubely et al., 1999).

A.I.2.3. Die Transposition von non-LTR-Retrotransposons

Die Retrotransposition von non-LTR-Retrotransposons ist ein komplexer Prozeß, der wie bei LTR-Elementen Transkription des vollständigen Elements, Integration in cytoplasmatische Ribonukleopartikel und Transport in den Kern umfaßt, wo die Reverse Transkription und Integration in das Kerngenom stattfinden. Allerdings unterscheidet sich der Ablauf der Initiation der Reversen Transkription und der Reintegration der Sequenzen ohne codierte Integrase ins Genom (Finnegan, 1997). Nach einem Modell von Schwarz-Sommer et al. (1987) für *Cin4* (*Zea mays*) wird die Reverse Transkription durch die Zielsequenz im Genom gestartet, die entweder durch zufällige Strangbrüche während Rekombination oder Reparatur oder durch gezielte Nukleaseaktivität des Elements freie 3'-OH Gruppen aufweisen. Die RT ist mit dem 3'-Ende der LINE-RNA assoziiert und nutzt die 3'-OH Gruppe als Initiationspunkt, wobei die Sequenz zwischen den versetzten Strangbrüchen ebenfalls als Templat dient, was zu einer Zielsequenzverdopplung („target site duplication“; TSD) führt. Integration und Reverse Transkription laufen simultan ab. Eine zelluläre oder elementcodierte RNase entfernt den RNA-Strang und das Reparatursystem der Zelle ergänzt den zweiten DNA-Strang, ebenfalls unter Verwendung der Zielsequenz als Startpunkt. Nur LINEs sind zu autonomer Transposition in der Lage, SINEs nutzen dafür die enzymatische Maschinerie der LINEs, deren Sequenz sie im 3'-Bereich imitieren (Schmidt, 1999).

Die Genomgröße korreliert sowohl mit der Gesamtzahl vorhandener Retroelemente als auch mit der Anzahl der vorhandenen Retrotransposonfamilien. Das Genom von *Saccharomyces cerevisiae* (12 Mbp) besteht zu 3.1% aus LTR-Elementen, während non-LTR-Elemente

fehlen. Der Hefepilz *Magnaporthe grisea* (40 Mbp) besitzt beide Retrotransposon-Großklassen, jedoch in niedrigen Kopienzahlen, und der Nematode *Caenorhabditis elegans* (100 Mbp) weist viele LINEs auf, jedoch nur wenige *Ty3-gypsy*-Elemente und keine bzw. nur sehr wenige *Ty1-copia*-Elemente und *SINEs*. Die Fruchtfliege *Drosophila melanogaster* (165 Mbp) verfügt über LTR- wie non-LTR-Sequenzen und aktive Retroviren, jedoch konnten bisher keine *SINEs* gefunden werden. Auch in Säugern (~3000 Mbp) sind Retroviren und *Ty3-gypsy* Elemente abundant, *Ty1-copia* Elemente fehlen allerdings. Dagegen stellen *SINEs* und *LINEs* fast 35% des Genoms dar. Im Vergleich zu anderen Eukaryonten weisen Pflanzen eine große Diversität und teilweise hohe Abundanz von allen Retroelementklassen mit Ausnahme von Retroviren auf.

A.I.2.4. Transkription und Aktivität von Retrotransposons in Pflanzengenomen

Die Expression von Retrotransposons in Pflanzen - und damit auch ihre Transpositionshäufigkeit - sind streng reguliert. Vermutlich wurde die Entwicklung von Kontrollmechanismen des Wirts für ihre Transkription und Transposition notwendig, um das Risiko schädlicher Mutationen, das immer mit der Mobilität von Retrotransposons verbunden ist, niedrig zu halten. Folglich überrascht es nicht, daß die meisten Elemente in der Entwicklung der meisten Gewebe transkriptionell inaktiv sind. Transkripte konnten bisher nur in wenigen Fällen in einzelnen Geweben bzw. Entwicklungszuständen nachgewiesen werden, wie z.B. für *Tnt1* in Wurzeln von Tabak (Grandbastien et al., 1989; Pouteau et al., 1991), *BARE1* in Blättern von Gerste, Roggen und Weizen (Suoniemi et al., 1996a,b) oder *PREM2* im frühen Mikrosporenstadium von Mais (Turcich et al., 1996). Viele Elemente verhalten sich passiv und werden nur unter biotischem oder abiotischem Stress - auch bei der Protoplastenisolation, in Gewebekulturen oder als Transgene - aktiviert (Wessler, 1996; Grandbastien, 1998; Okamoto & Hirochika, 2000; Courtial et al., 2001). Bei Mais konnte eine erhöhte Beweglichkeit von DNA-Transposons als Reaktion auf intensive UV-Strahlung beobachtet werden (Walbot, 1999), auch können mikroklimatische Einflüsse die Kopienzahl von *BARE1* in *Hordeum spontaneum* erhöhen (Kalendar et al., 2000). Oft ist der regulatorische Schritt hierbei die Transkriptionsinitiation, wie etwa bei Verwundung oder Kontakt mit Methyljasmonat, CuCl_2 und Salicylsäure oder Pilzextrakten bzw. viralen, bakteriellen oder pilzlichen Pathogenen (*Tto1* und *Tnt1* in Tabak: Pouteau et al., 1991; Hirochika, 1993; Pouteau et al., 1994; Mhiri et al., 1997; Takeda et al., 1998; Mhiri et al., 1999; Takeda et al., 1999; bzw. *Tos17* in Reis: Hirochika et al., 1996a; Hirochika, 1997). Die regulatorischen Sequenzen sind dabei „cis“-aktivierende Elemente im U3-Bereich des 5'-LTR wie BI oder BII bei *Tnt1* (Pouteau et al., 1991; Pouteau et al., 1994; Takeda et al., 1998; Takeda et al., 1999), an deren konservierte Motive Proteine binden, z.B. LBM1 bei *Tto1* (Casacuberta & Grandbastien, 1993; Takeda et al., 1999). Die Aktivatorsequenzen von *Tnt1* und *Tto1* sind auch in heterologen Systemen noch aktiv (Pauls et al., 1994; Hirochika et al., 1996b; Moreau-Mhiri et al., 1996) und weisen Homologien zur AC-I- und H-Box aus Promotoren pflanzlicher Abwehrgene gegen Pathogene auf (Grandbastien et al., 1997;

Grandbastien, 1998). Aufgrund ihrer Streßinduktion werden als Ursprung vieler pflanzlicher Retrotransposons Insekten-Retroviren angenommen, für die eine z.B. auf Verwundung basierende Aktivierung (fressendes Insekt) sinnvoll ist, da sie das Bereitstehen des Vektors anzeigt (Grandbastien, 1998; Kumar & Bennetzen, 1999). Darüberhinaus wirken Regulationsmechanismen der Wirtszelle, die die Gewebe- und Entwicklungsspezifität der Transkription steuern (Moreau-Mhiri et al., 1996; Pauls et al., 1994). Im Gegensatz zu diesem Eindruck der strengen Kontrolle über Retroelemente zeigten kürzlich Homologiesuchen in pflanzlichen EST („expressed sequence tags“)-Datenbanken, daß Retrotransposontranskripte besonders in Monocotyledonen weitverbreitet sind, und daß diese Elemente teilweise auch translatiert und an neuen Loci im Genom integriert werden (Vicent et al., 2001).

A.I.2.5. Evolutionäre Zusammenhänge zwischen den Retroelementgruppen

Retroelemente sind möglicherweise Relikte aus einem Zeitalter, in dem ausschließlich RNA als genetische Information diente. Sie waren schon in den ersten Pflanzen vorhanden, was ihr ubiquitäres Vorkommen und die starke Sequenzheterogenität erklären würde (Gesteland & Atkin, 1993). Ebenso ist eine Entwicklung erst nach dem Entstehen der Eukaryonten mit nachfolgender starker vertikaler und horizontaler Verbreitung denkbar (Kumar & Bennetzen, 1999). Allgemein werden LINEs aufgrund ihrer Sequenzdiversität als die ältesten und ursprünglichsten Retroelemente angesehen. Kladistische Studien legen die Vermutung nahe, daß LTR-Retrotransposons sich durch den Zugewinn von LTRs aus LINEs entwickelten (Xiong & Eickbush, 1990). Eine Zwischenform konnte bei *HeT-A* (*Drosophila melanogaster*) gefunden werden, wo eine Tandeminsertion und der Transkriptionsstart im 3'-Ende des benachbarten Elements die mögliche Entstehung flankierender LTRs verdeutlichen (Danilevskaya et al, 1997; Eickbush, 1997). SINEs gelten als Seitenlinie bei der Entwicklung von LINEs zu LTR-Retroelementen. Die Grenze zwischen Retrotransposons und Retroviren ist durch die Entdeckung von *env*-ähnlichen Leserastern, die bisher als Unterscheidungsmerkmal dienten, in zahlreichen pflanzlichen *Ty1-copia*-Elementen wie *SIRE1* aus *Glycine max* oder *Endovir* aus *Arabidopsis thaliana* und aus *Ty3-gypsy*-Elementen wie *Athila* (*A. thaliana*), *Cyclops* (*Pisum sativum*) und *Calypso* (*G.max*) zunehmend unklarer geworden (Pelissier et al., 1996; Chavanne et al., 1998; Laten et al, 1998; Wright & Voytas, 1998; Peterson-Burch et al., 2000). Allerdings konnten bisher noch keine funktionellen *env*-Glycoproteine oder infektiöse Retroviren in Pflanzengenomen nachgewiesen werden (Kumar, 1998). Im Gegensatz zu tierischen Genomen, wo bisher nur *Ty3-gypsy*-ähnliche retrovirale Elemente gefunden wurden, nimmt man daher bei Pflanzen die unabhängige Aufnahme des *env*-Gens durch beide Gruppen der LTR-Retrotransposons an. Allerdings kann alternativ nicht ausgeschlossen werden, daß in einzelnen Fällen oder auch allgemein beide Retroelementgruppen durch Verlust der *env*-Funktion aus den entsprechenden Retrovirenfamilien hervorgingen (Kumar, 1998).

A.I.3. Klasse II: DNA-Transposons

Im Gegensatz zu Retroelementen bewegen sich Elemente der Klasse II ohne RNA-Intermediat im Genom. Während bei Prokaryonten replikative und nicht-replikative Transpositionsmechanismen bekannt sind, findet man bei eukaryontischen Systemen nur die nichtreplikative Transposition durch Ausschneiden und direktes Reintegrieren der DNA-Kopien. Die Wiederholungseinheiten werden von terminalen invertierten Sequenzwiederholungen („terminal inverted repeats“; TIRs) flankiert, die ebenso wie die subterminal angeordneten kurzen Sequenzwiederholungen als „*cis*“-Determinanten für die transpositionelle Aktivität notwendig sind. Die von den internen Abschnitten codierten Proteine können dagegen auch von anderen Transposonkopien *in trans* zur Verfügung gestellt werden. Man unterscheidet drei Überfamilien von DNA-Transposons in Pflanzen, die nach den zuerst in Mais beschriebenen Familien *Activator* (*Ac*), *Enhancer/Suppressor-Mutator* (*En/Spm*) (Peterson, 1953; McClintock, 1954) und *Mutator* (*Mu*) (Kunze et al., 1997) benannt werden und sich durch die Feinstruktur der TIRs bzw. konservierte Sequenzen der Leseraster unterscheiden. Verschiedene dieser Transposonfamilien können in einem Genom coexistieren und umfassen jeweils autonome und nicht-autonome Kopien, die in aktivem oder passivem Zustand vorliegen können. *Mutator*-ähnliche Elemente besitzen längere TIRs und andere Transpositionsmechanismen als *Ac*- oder *En/Spm*-verwandte Sequenzen (Kunze et al., 1997). *En/Spm*-Elemente weisen die Sequenz 5'-CACTA-3' an den Enden der meist 13 Bp langen TIRs („CACTA-Elemente“; Fig.A2), 3 Bp lange flankierende Wiederholungen der Zielsequenz und 8 - 15 Bp lange subterminale Wiederholungsmotive auf.

Die internen Bereiche werden von Introns unterbrochen und codieren für zwei Proteine, die bei *En/Spm* TNPA und TNPB bzw. bei *Tam1* TNP1 und TNP2 genannt werden (Fig.A2). Bei *En/Spm* befinden sich Promotorsequenzen zur Transkription der beiden proteincodierenden Leseraster im subterminalen repetitiven Motivbereich am 5'-Ende, wobei ein langes polyadenyliertes Transkript entsteht, das später durch Spleißen in zwei große, prozessierte Transkripte zerlegt wird, die für TNPA und TNPB codieren, sowie in mehrere kleinere, für die Transposition vermutlich unerheblichen RNAs (Gierl, 1996; Kunze et al., 1997). TNPB/TNP2 besitzt Transposase (TPase)-Aktivität und katalysiert die Excision und Insertion des Elements durch Bindung der TIRs und endonukleolytische Spaltungen. TNPA/TNP1 bindet an die subterminalen, repetitiven Strukturen und dimerisiert, was die Transposasebindung erleichtert. Gleichzeitig wirkt das zweite Protein als Regulator, der für eine gleichmäßig niedrige Aktivität des Elements sorgt (Kunze et al., 1997). Die Expression der Proteine bzw. die Transposition wird durch solche autoregulatorischen und/oder interagierenden epigenetischen Mechanismen kontrolliert und ist unabhängig von ihrer chromosomalen Lokalisation (Fedoroff, 1999).

Die Insertion eines DNA-Transposons beginnt mit einem versetzten Doppelstrangbruch an der Zielsequenz, deren Enden z.B. mit denen des *En/Spm*-Elements ligiert werden. Ein zelleigener Reparaturmechanismus ergänzt die fehlenden Stücke komplementär, was in einer

flankierenden Zielsequenzverdoppelung („target site duplication“; TSD) resultiert, deren typische Länge bei *En/Spm*-Elementen 3 Bp beträgt. Für die Excision existieren zwei Modelle. Nach Saedler & Nervers (1985) finden an den Enden der TSD versetzte Strangbrüche statt, das überhängende Stück der Transposonkopie wird exonukleolytisch abgebaut. Gleichzeitig ergänzt das Reparatursystem die fehlenden genomischen Stränge komplementär. Die Polymerase kann hierbei versehentlich den Überhang des Elements als Templat verwenden, so daß nach Entfernung der ausgeschnittenen Transposonkopie und Verbindung der verbliebenen Stränge kleinere Rearrangements in der TSD entstehen können, sog. Fußspuren des Transposons („footprints“). Dem zweiten Modell nach Coen et al. (1986) liegt derselbe generelle Mechanismus von Ausschneiden und Reparatur der Ausschnittsstelle zu Grunde, jedoch bilden die überhängenden genomischen Strangenden hier „hairpin“-Strukturen, die an verschiedenen Stellen endonukleolytisch gespalten werden können und aufgrund des Reparaturmechanismus ebenfalls in verschiedenen „footprints“ resultieren.

En/Spm-ähnliche Elemente wurden in verschiedenen Spezies nachgewiesen (Gierl, 1996; Kunze et al., 1997). Allerdings wurden bisher nur wenige Elemente auf molekularer Ebene charakterisiert, wie die zuerst identifizierten Elemente *En/Spm* aus *Zea mays* (Pereira et al., 1996), *Tam1* aus *Antirrhinum majus* (Nacken et al., 1991), *Tpn1* („Japanese Morning glory“/ *Ipomoea nil*; Inagaki et al., 1994), *Psl* (*Petunia hybrida*; Snowden & Napoli, 1998) oder *Rim2* und *Tnr3* aus *Oryza sativa* (Motohashi et al., 1996; He et al., 2000), ferner *Cs1* (*Sorghum bicolor*; Chopra et al., 1999) und *Tdc1* (*Daucus carota*; Ozeki et al., 1997). Auch in Leguminosen wurden die *En/Spm*-ähnlichen Elemente *Tgm1* (*Glycine max*; Rhodes & Vodkin, 1988) sowie *Pis1* (*Pisum sativum*; Shirsat, 1988) identifiziert. Die Längen der vollständig sequenzierten Elemente liegen zwischen 5.2 Kbp (*Tdc1*; Ozeki et al., 1997) und 15.2 Kbp (*Tam1*; Nacken et al., 1991), wobei *Cs1* (23 Kbp) das längste bisher sequenzierte Transposon dieser Gruppe darstellt (Chopra et al., 1999). In manchen Fällen wurden unvollständige Derivate sequenziert, die meistens durch interne Deletionen verändert waren (Ozeki et al., 1997; He et al., 2000), z.B. die ungewöhnlich kurzen Elemente *Tnr3* (1.5 Kbp) und *Pis1* (2.5 Kbp).

Seit fast zwanzig Jahren werden Transposon-Insertionen, die phänotypische Unterschiede verursachen, zur Identifikation und Isolation verschiedener Gene verwendet. Heute ist die Genidentifikation mit bestimmten Transposons auch in heterologen Spezies möglich (Fedoroff et al., 1984; Sundaresan, 1996; Kunze et al., 1997). Zusätzlich erlauben diese mobilen Elemente auch funktionelle Studien der identifizierten Gene und können zum Kartieren und Sequenzieren verwendet werden (Devine & Boeke, 1994; Van den Broeck et al., 1998; Tissier et al., 1999; Walbot, 2000). Mit Hilfe einiger Elemente wurden ferner Verwandtschaftsverhältnisse zwischen Pflanzenspezies untersucht (Iwamoto et al., 1999). Aktive Transposons mit hoher Transpositionsrate wie *En/Spm* (*Z. mays*) können auch in heterologen Genomen transponieren (Cardon et al., 1993; Aarts et al., 1995; Speulmann et al., 1999; Tissier et al., 1999). Die molekulare Analyse endogener Transposons ist Grundlage für verschiedene PCR (Polymerase-Kettenreaktion)-Techniken, wie dem sog. „transposon

display“ (Van den Broeck et al., 1998) oder AIMS (amplification of mutagenized sites; Frey et al., 1998), die auf AFLP („amplified fragment length polymorphism“) beruhen. Sie können zur Isolation flankierender Sequenzen von Transposons, zur Kartierung der identifizierten genomischen Loci und zur Kopienzahlbestimmung einer TE (transposablen Element)-Familie oder Bestimmung der Transpositionsfrequenz verwendet werden. Die ligationsvermittelte PCR-Technik des „transposon insertion display“ (TID) zur Isolation von Transposon-flankierenden Sequenzen kann gleichfalls selektiv für Insertionen in transkribierte Gene verwendet werden (ETID; „expressed transposon insertion display“; Yphremov & Saedler, 2000).

A.I.4. Hat repetitive DNA eine Funktion im Genom?

In Anbetracht des energetischen Aufwands bei der Replikation repetitiver Sequenzen, den eine Pflanzelle bei jeder somatischen Zellteilung betreibt, ist es naheliegend, nach einer biologischen Funktion dieser Sequenzen zu suchen. Entgegen den Behauptungen über „junk“ und „selfish“ (egoistische) DNA (Doolittle & Sapienza, 1980; Orgel & Crick, 1980; Charlesworth et al., 1994) wurden in den letzten 20 Jahren verschiedene Hypothesen zur Funktion von Satelliten-DNA und Retrotransposons aufgestellt.

A.I.4.1. Funktionen von Satelliten-DNA

Da die Sequenzen verschiedener Satelliten nicht konserviert ist, geht man von einer strukturabhängigen Funktionalität aus, die auf der Periodizität einer Sequenzwiederholung und der damit verbundenen Biegung der DNA-Helix beruht. Durch Wechselwirkung mit bestimmten Kernproteinen wird der daraus resultierende sog. „chromatin folding code“ fixiert (Vogt, 1990). Zu solchen Funktionen gehört die Bildung und Stabilisierung bestimmter Chromosomenstrukturen wie Centromere oder Telomere. Auch die Bildung bzw. Aufrechterhaltung des Chromatins im Interphasekern wird durch Wechselwirkungen von Proteinen mit subtelomerischen und pericentrischen Satelliten und der damit verbundenen Bildung von konstitutivem Heterochromatin bedingt (Bennett, 1982). Ebenso sind subtelomerische, pericentrische oder interkalare Sequenzen vermutlich für die Erkennung und Paarung homologer Chromosomen während der Mitose und Meiose verantwortlich, wie identisch verteilte heterochromatische Bereiche auf homologen Chromosomen vermuten lassen (Bennett, 1982; Irick, 1994). Unterschiedliche Zusammensetzung hochrepetitiver Satelliten-DNA kann dabei Genominkompatibilitäten verursachen, wie Untersuchungen von nahe verwandten, jedoch nicht mehr kreuzbaren Arten ergaben (Ganal & Hemleben, 1986; Junghans, 1989). Zudem ist die Rekombinationshäufigkeit in heterochromatischen Bereichen deutlich herabgesetzt und korreliert mit einer späten Replikation in der S-Phase (Vogt, 1990; The *Arabidopsis* Genome Initiative, 2000). Eine pflanzliche Nukleosomeneinheit (175-185 Bp) entspricht etwa den Monomereinheiten vieler Satelliten (160-180 Bp), was den Einfluß einiger Tandemrepeats auf die Nukleosomenpositionierung wahrscheinlich macht. Dies kann

durch Protein-DNA-Wechselwirkungen geschehen, die gleichzeitig Einfluß auf die Nukleosomenposition und Chromatinkondensation haben. Aus *Cucumis sativus* konnte ein solches satellitenbindendes Protein isoliert werden (Hemleben, 1993; Fischer et al., 1994).

Auch die Beteiligung von Satelliten-DNA an der Centromerfunktion in Pflanzengenomen wird diskutiert: Die Centromere sind für die Verbindung der beiden Schwesterchromatide verantwortlich und dienen als Bindestelle für die Proteine des Kinetochors, das dem Spindelfaserapparat als Angriffspunkt dient und in Mitose und Meiose eine gleichmäßige Aufteilung der Chromosomen auf die Tochterzellen garantiert. Die Centromere aller bisher untersuchten höheren Eukaryonten erstrecken sich über mehrere Megabasenpaare und bestehen aus verschiedenen repetitiven Elementen. Trotz des Vorhandenseins konservierter centromerbindender Proteine (z.B. CENB-P bei Hefe und Mensch) konnten bisher keine für alle Chromosomen konservierten Sequenzen identifiziert werden. Die centromerischen repetitiven Elemente sind sogar zwischen den Chromosomen einer Spezies variabel (Copenhaver & Preuss, 1999). Allerdings wurden an den Centromeren von *Arabidopsis* (Martinez-Zapater et al., 1986), *O.sativa* (Dong et al., 1998), *Z.mays* (Ananiev et al., 1998b) und *Beta* (Gindullis et al., 2001) geordnete Satellitenabfolgen gefunden, die von weiteren repetitiven Elementen (Copenhaver et al., 1999; Heslop-Harrison et al., 1999; Tutois et al., 1999) flankiert werden. Ob diese centromerischen Satelliten möglicherweise nur als Folge der niedrigen centromerischen Rekombinationsfrequenz akkumulieren bzw. welche Funktion ihnen zukommt, wird kontrovers diskutiert. Widersprüchlich ist vor allem, daß Satelliten nicht auf allen Centromeren höherer Eukaryonten zu finden sind (z.B. *Vicia faba* oder menschliche Neocentromere; Fuchs et al., 1998; Copenhaver & Preuss, 1999). Auch sind die centromerischen repetitiven Elemente nicht auf Centromere beschränkt und zwischen den Chromosomen sehr heterogen. Nach verschiedenen Modellen könnten Satellitenelemente als sequenzspezifische Ansatzpunkte für den Spindelfaserapparat dienen (Copenhaver & Preuss, 1999), sie können Ziel für epigenetische Modulation sein (wie z.B. Methylierung) oder eine übergeordnete Struktur bilden und so die Centromerstruktur aufrechterhalten, die für dessen Aktivität verantwortlich ist (Vig, 1994; Csink & Henikoff, 1998). Auch passive Funktionen, wie etwa als spät replizierende Punkte des Chromosoms und - durch Amplifikation - als Platzhalter, um der Invasion störender Sequenzen (z.B. mobiler Elemente) entgegenzuwirken (Csink & Henikoff, 1998), werden diskutiert (Schubert, 1998; Henikoff, 1998).

A.I.4.2. Funktionen von transposablen Elementen

Retrotransposons und Transposons können durch Insertionen in und nahe bei Genen Mutationen verursachen, die das Transkriptionsmuster oder die Funktionalität des Genprodukts verändern (Kumar & Bennetzen, 1999). Eine Reihe von Krankheiten bei Maus und Mensch werden durch Insertionen von non-LTR-Retrotransposons verursacht, darunter bestimmte Formen von Brust- und Darmkrebs (Kazazian, 1998). Obwohl Retrotransposons bekanntermaßen auch pflanzliche Genome stark aufgebläht haben (Kumar, 1996; SanMiguel et al., 1998) und durch die Struktur ihrer LTRs nicht-reziproke Rekombinationen auslösen,

die zu Genomrearrangements wie Duplikationen, Deletionen, Inversionen und Translokationen führen können (Kumar & Bennetzen, 1999), kann ihre energieraubende und mutagene Anwesenheit über lange evolutionäre Zeiträume jedoch auch vorteilhaft sein. Die Verschleppung und Amplifikation von Genfragmenten (Bureau et al., 1994; Jin & Bennetzen, 1994; Moran et al., 1999) oder die reverse Transkription von mRNA bzw. anderen RNAs (Drouin & Dover, 1987; Drouin, 1999) kann zur Bildung von Pseudogenen führen, die in Kombination mit regulatorischen Elementen unter bestimmten Selektionsbedingungen funktionell und vorteilhaft sein können. Chromosomale Rearrangements werden auch durch Transposition, sog. „alternative Transposition“ und homologe Rekombination zwischen DNA-Transposons begünstigt (Bennetzen, 2000; Fedoroff, 2000; Gray, 2000). Eine durch Transposoninsertionen veränderte Regulation, Expression oder das Spleißen von Genen kann sich in der Selektion positiv auswirken (Walker et al., 1995; May & Dellaporta, 1998; Girard & Freeling, 1999). Ebenso können „trans“-agierende Faktoren eines Transposons in Wirtsgene umgewandelt und mit wirtseigenen regulatorischen Sequenzen ausgestattet werden wie die dem P-Element verwandten Gencluster in *Drosophila* (Miller et al., 1997). Über die Insertion von DNA-Transposons und deren signifikanten Beitrag zur Evolution von pflanzlichen Resistenzgenen wurde ebenfalls kürzlich berichtet (Luck et al., 1998; Song et al., 1998; Henk et al., 1999; He et al., 2000). Aus Vertebratengenomen sind viele Bereiche codierender Regionen bekannt, die durch Retroelemente generiert wurden (Brosius, 1999), und auch sog. Retrogene, die aus reverser Transkription von mRNAs entstanden (Brosius, 2001). Die Mobilität „cis“-regulatorischer Elemente aus Retrotransposonsequenzen und ihre Rekrutierung für zelluläre Gene ist von Säugergenomen (Brosius, 1999) und Pflanzen (McDonald et al., 1997) bekannt. Sie kann vermutlich nur für einen geringen Teil aller Fälle bewiesen werden, da die Retrotransposoninsertionen bis auf kurze, konservierte Regulatorsequenzen degeneriert sind (Brosius, 2001). So weisen „cis“-aktivierende Elemente pflanzlicher LTR-Retrotransposons Homologien zu Promotoren aus pflanzlichen Genen der Pathogenabwehr auf (Pouteau et al., 1991; Pouteau et al., 1994; Grandbastien et al., 1997; Grandbastien, 1998; Takeda et al., 1998; Takeda et al., 1999). Außer durch Domestikation von elementeigenen Kontrollsequenzen kann auch Heterochromatinbildung durch repetitive Elemente lokalen Einfluß auf die Expression in benachbarten euchromatischen Bereichen nehmen (Hultén & Stacey, 1995; Selinger & Chandler, 2001). Darüberhinaus können bei Pflanzen Insertionen von Retroelementen in Intronsequenzen bzw. Aktivierung von Spleißsignalen zu alternativen Spleißmustern, neuen Exons und veränderten Expressionsmustern führen (Varagona et al., 1992; Marillonnet & Wessler, 1997; Henikoff & Comai, 1998; Brosius, 2001).

Abgesehen von den genannten Möglichkeiten, die Variabilität des Genoms und damit das Potential für selektive Vorteile durch Retrotransposonmobilität zu vergrößern, gibt es auch Beispiele für zelluläre Funktionen, die durch Rekrutierung von retroelementeigenen codierenden Funktionen entstanden sind (Miller et al., 1999). Ein prominentes Beispiel dafür ist die Rolle der non-LTR-Elemente *HeT-A* und *TART*, die die Chromosomenenden bei *Drosophila* anstelle von Telomerrepeats schützen (Pardue et al., 1997; Danilevskaya et al.,

1998). Telomerase und die non-LTR Reversen Transkriptasen arbeiten nach demselben Mechanismus, nur mit verschiedenen RNA-Templaten (Pardue et al., 1997). Allerdings ist bisher unklar, ob *Drosophila* die repetitiven Elemente als Ersatz für eine eigene Telomerase domestiziert, d.h. unter strikter zellulärer Kontrolle gestellt hat, oder ob das Retrotransposon aus der zellulären Telomerase entstanden ist (Eickbush, 1997; Pardue et al., 1997). Ein weiteres Beispiel stellt die sequenzspezifische VDJ-Rekombination zur Generation der Immunglobuline und T-Zell-Rezeptoren bei Vertebraten dar, die durch die Rekombinasefunktion der Gene RAG1 und RAG2 gesteuert wird. Der Mechanismus der Rekombinase zeigt auffällige funktionelle Parallelen zu Transposasen (sog. „alternative Transposition“; Gray, 2000) und retroviralen Integrasen bzw. auch ihre Sequenzen weisen strukturelle Ähnlichkeiten auf (Agrawal et al., 1998; Hiom et al., 1998; Gray, 2000). Man nimmt daher an, daß die Schlüsselfunktion des kombinatorischen Immunsystems der Vertebraten sich aus einem transposablen Element entwickelt hat, dessen ORFs RAG1 und RAG2 ähneln und das von TIRs mit Ähnlichkeit zu den flankierenden Rekombinationssignalen begrenzt wurde (Roth & Craig, 1998; Schluter et al., 1999). Die Virusresistenz der Maus gegen MLV („murine leukemia virus“) stellt eine weitere zelluläre Rekrutierung eines Retroelements dar. Sie geht auf das *Fv1*-Gen zurück, das eine virale *gag*-Sequenz unter zellulärer Kontrolle darstellt (Miller et al., 1999). Für Pflanzen werden ebenfalls von Retrotransposons entlehnte zelluläre Funktionen diskutiert: möglicherweise spielt eine Retrotransposonfamilie, die in den Centromeren von Gräsern konserviert ist, eine Rolle bei der Centromerfunktion (Miller et al., 1998; Presting et al., 1998). Ebenso wird *LINE-CS* Sequenzen (*Cannabis sativa*), die im terminalen Abschnitt auf dem langen Arm des Y-Chromosoms akkumuliert sind, eine Funktion zugeordnet. Dieser heterochromatische Bereich macht den Heteromorphismus zwischen X- und Y-Chromosom aus und ist möglicherweise für die Unterdrückung der Paarung und Rekombination der Geschlechtschromosomen dieser diözischen Pflanze verantwortlich (Sakamoto, 2000).

Das in den achtziger Jahren propagierte Bild der egoistischen oder parasitären „junk“-DNA (Orgel & Crick, 1980; Doolittle & Sapienza, 1980) muß offensichtlich zu Gunsten der repetitiven DNA als eines „Symbionten“ im Genom korrigiert werden (Georgiev, 1984; Brosius, 2001), der die energetischen Kosten für Replikation und Expression durch eine erhöhte Anpassungsfähigkeit aufwiegt und aufgrund vorteilhafter Qualitäten vom Wirtsgenom toleriert wird.

A.I.5. Welche Mechanismen begrenzen die Amplifikation repetitiver Sequenzen?

Abgesehen von Rekombinationsereignissen und Strangverrutschen bei der Replikation, die die Zahl tandemartig angeordneter Wiederholungseinheiten verringern können, oder dem Ineinanderspringen von Retroelementen, die um die Wirtsressourcen konkurrieren, gibt es eine Reihe von Mechanismen, mit denen die Wirtszelle (und

möglicherweise das Element selbst) die Amplifikation von transposablen Elementen begrenzen können. DNA-Methylierung steht hierbei im Vordergrund, wie Untersuchungen an interspezifischen Säugerhybriden (*Macropus eugenii* x *Wallabia bicolor*) zeigten, deren genomweite Untermethylierung mit einer starken Amplifikation von Retroelementen korrelierte (O'Neill et al., 1998). Die Methylierung von Cytosin in Pflanzen ist mit der Ausbildung heterochromatischer Strukturen verbunden, und mit der Transkriptionsinaktivität treten auch mehr C→T-Übergänge durch Desaminierung des 5-Methylcytosins auf. Die Inaktivierung der TEs kommt daher durch verringerte Transkription und auch eine erhöhte Mutationsrate der funktionellen Sequenzen zustande. Auch in Genomen, die keine DNA-Methylierung aufweisen, z.B. von *Drosophila* und *Saccharomyces*, ist eine inaktive Chromatinstruktur zu finden. Ein solcher epigenetisch inaktivierter, d.h. reversibel und ohne Veränderung der Nukleotidsequenz herbeigeführter erblicher Zustand kann durch Insertion des TEs in die Nähe bereits existierender Heterochromatinblöcke eintreten, welche sich dann ausdehnen können (Wakimoto, 1998). Bei Pilzen findet man in bestimmten Zellstadien Mechanismen wie RIP („repeat-induced point mutation“) oder MIP („methylation induced premeiotically“), die repetitive Sequenzen aufgrund von Sequenzhomologien mutieren bzw. methylieren (Selker et al., 1987; Matzke et al., 1999). Aus Pflanzen und Vertebraten sind ebenfalls homologieabhängige Inaktivierungsmechanismen bekannt, die sich gegen fremde DNAs, TEs, Transgene oder endogene Gene mit partiellen Fremdsequenzen in diploiden, somatischen Zellen richten. Sie werden aktiv, sofern die genannten Sequenzen repetitiv sind, RNAs mit Sekundärstrukturen erzeugen oder DNA-Rearrangements hervorrufen (Matzke et al., 1999). Man unterscheidet hier TGS („transkriptional gene silencing“) und PTGS („posttranscriptional gene silencing“). Beim TGS werden multiple Promotorsequenzen methyliert und die Transkription der unliebsamen Sequenz so verhindert. Über DNA-DNA- bzw. bei Transkripten DNA-RNA-Interaktionen können auch *in trans* vorliegende Kopien des Elements methyliert werden. Das PTGS richtet sich gegen mehrfach vorliegende codierende Sequenzen, deren RNAs posttranskriptionell über Bildung von dsRNA-Molekülen im Cytoplasma degradiert werden. Auch bei diesem, ursprünglich wohl gegen virale Pathogene gerichteten Mechanismus („Cosuppression“) werden die DNA-Sequenzen im Kern durch ein RNA-Signal methyliert, was auch *in trans* vorhandene Kopien stilllegt. Zusätzliche mobile Signale werden im Gefäßsystem transportiert und können für eine veränderte Expression an genomischen Loci in der Pflanze sorgen (Matzke et al., 1999).

Einige Autoren betrachten sogar die Evolution epigenetischer Mechanismen zur Expressionskontrolle eukaryontischer Gene als Abwandlungen des beschriebenen zellulären Abwehrsystems gegen fremde oder transposable Elemente, da die regulierten Gene oft TE-Reste oder durch TEs verursachte Rearrangements enthalten (Bestor & Tycko, 1996; McDonald, 1998; Matzke et al., 1999).

TEs weichen den beschriebenen Abwehrmechanismen jedoch z.T. aus. Auch ist die Wirtsabwehr durchlässig, so daß manche Elemente zu hohen Kopienzahlen amplifiziert werden können. Die evolutiven Vorteile der genomischen Variabilität, die durch die TE-Aktivität generiert wird, lassen die Abwehrmechanismen oft erst ab einer bestimmten

Kopienzahl der repetitiven Sequenz (Jensen et al., 1999) oder nur in bestimmten Entwicklungsstadien greifen (z.B. in Keimzellen; Dupressior & Heidmann, 1996). Ebenso können kurze und AT-reiche TE-Sequenzen unerkannt bleiben und werden nicht methyliert, wodurch die Proliferation von MITEs oder mancher Retrotransposons erklärbar ist (Matzke et al., 1999). Möglicherweise spielen auch Eigenschaften des Wirtsgenoms eine Rolle: kleine, diploide Genome mit wenigen repetitiven Sequenzfamilien üben aufgrund der durch Mutationen zu erwartenden Schäden evtl. eine rigorosere Kontrolle aus als große, polyploide Genome mit Genduplikationen (Matzke et al., 1999; The *Arabidopsis* Genome Initiative, 2000), wie etwa der allopolyploide Mais mit starken, kürzlich erfolgten Amplifikationen von Retroelementen (SanMiguel et al., 1996; 1998) und einer großen Zahl MITEs in intergenischen Regionen (Wessler, 1998).

A.II. Die Kichererbse (*Cicer arietinum* L.) und die Gattung *Cicer*

A.II.1. Allgemeines

Die Kichererbse (*Cicer arietinum* L.) gehört zu den ältesten kultivierten Leguminosen und wird seit mehr als 8000 Jahren im Nahen Osten angebaut, wie die ältesten archäologischen Funde von Samen in der Nähe von Burdur (Türkei) belegen (Helbaek, 1970). Die erste Erwähnung der Kichererbse findet sich in frühen Schriften von etwa 3000 v Chr. in Mesopotamien. Für den lateinischen Namen *Cicer* werden mehrere Ursprünge vermutet, am wahrscheinlichsten ist „kikar“ (rund) aus dem Hebräischen. Der Artnamen *arietinum* leitet sich aus der Samenform ab, die an den Kopf eines Widders (*aries*) erinnert (Van der Maesen, 1987). Ursprünglich stammt *C. arietinum* aus dem Südosten der Türkei und den angrenzenden Gebieten Syriens, wo nahe verwandte Wildarten gefunden werden (Van der Maesen, 1987). Heute stellt die Kichererbse mit einer jährlichen Produktion von 3000 Mio US \$ die dritt wichtigste Körnerleguminose weltweit dar (FAO, 1995) und wird besonders in den Ländern der sog. Dritten Welt angebaut. Eine Vielzahl von Akzessionen (Landrassen) wird weltweit in semiariden Gebieten des indischen Subkontinents, im Mittelmeerraum, in Nord- und Mittelamerika sowie in Australien kultiviert. Nach morphologischen Kriterien lassen sich bei der Kulturform von *C. arietinum* zwei Typen unterscheiden: der Kabuli-Typ mit großen, hellen Samen und der kleine und dunkelsamige Desi-Typ. Umfangreiche Sammlungen von Kabuli-Akzessionen sind beim ICARDA (International Centre for Agricultural Research in the Dry Areas) und von Desi beim ICRISAT (International Crop Research Institute for Semiarid Tropics) angelegt worden (Singh et al., 1991). Beliebt ist die Pflanze als kostengünstiges Nahrungs- und Futtermittel, unter anderem wegen des hohen Gehalts hochwertiger Nahrungsproteine in den Samen (FAO, 1988). Zusätzlich ist die Kichererbse, die über eine bis zu einem Meter lange Hauptwurzel verfügt, trockentolerant, was sie besonders in semiariden Gebieten zu einem Hauptnahrungsmittel macht. *C. arietinum* ist einjährig und ein sog. „obligater Selbster“, d.h. eine Fremdbefruchtung wird sowohl im Freiland als auch im Gewächshaus fast nie beobachtet, sondern in >99 % aller Fälle findet - auch in verwandten Arten - eine Selbstbestäubung und Selbstbefruchtung der Blüte statt (Tayyar et al., 1996).

A.II.2. Cytologische Untersuchungen bei *C. arietinum*

Die Kichererbse weist diploide Zellkerne mit einer somatischen Chromosomenzahl $2n = 16$ auf und hat eine Genomgröße von 738 Mbp (Arumuganathan & Earle, 1991). Viele cytologische Untersuchungen wurden bei *C. arietinum* bereits durchgeführt (u.a. Sharma & Gupta, 1986; Galasso & Pignone, 1992; Ocampo et al., 1992; Tayyar et al., 1994; Ahmad, 2000). Die acht Chromosomen sind durch C-Bandierungstechniken unterscheidbar (Galasso & Pignone, 1992), und 43% des Kichererbsengenoms wurden als heterochromatisch charakterisiert (Tayyar et al., 1994; Galasso et al., 1996). Die heterochromatischen Bereiche befinden sich vorwiegend in der Centromerregion aller Chromosomen und lassen sich größtenteils mit DAPI (4,6 Diamino-2-phenylindol; ein Fluorochrom, das bevorzugt an AT-reiches Heterochromatin bindet) anfärben (Galasso & Pignone, 1992). Dies weist auf den AT-Reichtum dieser Sequenzen hin (Schweizer, 1979; Sumner, 1990). Nur ein geringer Teil des Heterochromatins ist DAPI-negativ. Die Chromosomen werden ihrer Größe nach geordnet mit den Buchstaben A bis H bezeichnet (Galasso & Pignone, 1992), wobei das größte Chromosom A einen großen Heterochromatinblock und eine sekundäre Einschnürung aufweist, die die einzige Nukleolus Organisator Region (NOR) mit transkriptionell aktiven rRNA-Genen trägt. Das zweitgrößte Chromosomenpaar B hat ebenfalls einen großen heterochromatischen Anteil und ist wie A und C submetacentrisch. Die übrigen mittelgroßen Chromosomen und das kleinste (H) sind metacentrisch, C, D, E, F und G sind nur durch C-Bandierung oder Längenbestimmung der Chromosomenarme unterscheidbar (Galasso & Pignone, 1992; Tayyar et al., 1994, Ocampo et al. 1992, Ahmad, 2000). Die CMA (Chromomycin A3)-positiven Regionen auf der sekundären Einschnürung von A und im subtelomerischen Bereich von B entsprechen den Loci der 18S-5.8S-25S rRNA-Gene, von denen nur die NOR-Region auf A aktiv transkribiert wird (Galasso et al., 1996). Die 5S rRNA-Gene sind auf Chromosom B (proximal zum 18S-5.8S-25S rDNA-Locus) und auf einem der mittelgroßen Chromosomen (E) lokalisiert (Galasso et al., 1996).

A.II.3. Die Analyse des Kichererbsengenoms

Trotz internationaler Zuchtprogramme konnten die mit ca. 800 kg/ha relativ geringen jährlichen Erträge bisher nicht gesteigert werden. Frühere Aussattermine im kühlen und feuchten Winter und zu Beginn des Frühjahrs können zwar die Produktivität steigern, jedoch werden dann oft große Teile der Ernte durch Pathogenbefall vernichtet. Unter den 50 Pathogenen der Kichererbse verursacht in Nordafrika und Westasien vor allem der Ascomycet *Ascochyta rabiei*, der die sog. „Ascochyta Bleiche“ auslöst (Saxena & Singh, 1987; Nene et al, 1991; Kaiser, 1997), große Ernteverluste, während auf dem indischen Subkontinent ein anderer Schadpilz, *Fusarium oxysporum* f. sp. *ciceri*, für die Ertragseinbußen verantwortlich ist (Jiminez-Diaz et al., 1993). Langfristige Zuchtziele streben daher Hohertragsrassen an, die gleichzeitig Toleranz gegen Trockenheit und Frost

sowie Resistenz gegen die beiden Pilzpathogene vereinen. Nachdem die Verwendung molekularer Marker in den letzten Jahren in verschiedenen Bereichen der Pflanzenzucht als potentes Hilfsmittel erkannt wurde (Melchinger, 1990), sollte auch für *C. arietinum* eine markergestützte Selektion und die Klonierung agronomisch wichtiger Gene durch die Erstellung einer genetischen Karte erleichtert werden. Mittlerweile sind mehrere genetische Karten mit morphologischen Markern, Isoenzymmarkern (Gaur & Slinkard, 1990; Kazan et al., 1993), RAPD ("random amplified polymorphic DNA")- und RFLP ("Restriction fragment length polymorphism")-Markern (Simon & Muehlbauer, 1997), ebenso mit STMS (sequence-tagged microsatellite sites)-, AFLP („amplified fragment length polymorphism“)- und DAF ("DNA amplification fingerprinting")-Markern (Winter et al., 1999, 2000) und mit ISSR (inter simple sequence repeat)-Markern (Santra et al., 2000) vorhanden. In einigen Karten wurden die Loci für die Resistenz gegen *A. rabiei* (QTL; "quantitative trait loci"; Santra et al., 2000) und zwei Rassen von *F. oxysporum* (Ratnaparkhe et al., 1998a, b) lokalisiert. Da die kultivierte Kichererbse wenig genetische Variation aufweist (Ahmad & Slinkard, 1992; Udupa et al., 1993; Labdi et al., 1996), werden für Kopplungsanalysen meist verschiedene interspezifische Kreuzungen von *C. arietinum* x *C. reticulatum* verwendet (Simon & Muehlbauer, 1997; Winter et al., 1999, 2000; Santra et al., 2000), deren Nachkommen fast 100% Pollenfertilität aufweisen (Singh & Ocampo, 1993).

A.II.4. Verwandte Wildarten der Kichererbse

Die Gattung *Cicer* gehört zur Familie der Fabaceae, Unterfamilie Papilionoideae, und stellt eine eigene Gattungsgruppe dar, Cicerea Alef., die insgesamt 9 annuelle und 33 perennierende Arten umfaßt (Van der Maesen, 1987). *C. arietinum* stellt die einzige kultivierte Art dieser Gattung dar, allerdings wird auch die perennielle *C. songaricum* in kleinem Umfang in Afghanistan, Tibet und dem Westhimalaya als Lebensmittel angebaut (Ahmad, 2000). Aufgrund morphologischer Merkmale, geographischer Verteilung und ein- bzw. mehrjährigem Lebenszyklus werden die *Cicer*-Spezies in vier Sektionen (*Monocicer*, *Chamaecicer*, *Polycicer* und *Acanthocicer*) eingeteilt (Van der Maesen, 1987). Alle annualen Arten außer *C. chorassanicum* (*Chamaecicer*) gehören zur Sektion *Monocicer*.

Die Wildarten der Sektion *Monocicer* weisen Resistenz gegen verschiedene biotische und abiotische Stressfaktoren auf und sind deshalb als genetische Ressourcen von großer praktischer Bedeutung für Züchtungsprogramme mit *C. arietinum* (Singh et al., 1998). Innerhalb der Sektion ist *C. arietinum* mit *C. reticulatum* und mit *C. echinospermum* kreuzbar (Ladizinski & Adler, 1976a,b; Singh & Ocampo, 1993). Die Kreuzung von *C. reticulatum* mit *C. arietinum* erzeugt hochgradig fertile Nachkommen (Singh & Ocampo, 1993), während zwischen *C. arietinum* und *C. echinospermum* Kreuzungsbarrieren existieren (Singh & Ocampo, 1997). In der Sektion *Monocicer* gibt es insgesamt vier Kreuzungsgruppen, innerhalb der interspezifische Hybride möglich sind. Die zweite Kreuzungsgruppe umfaßt *C. bijugum*, *C. pinnatifidum* und *C. judaicum*, während *C. chorassanicum* und *C. cuneatum* jeweils eigene Gruppen bilden (Ladizinski & Adler, 1976a,b).

Die verwandtschaftlichen Beziehungen zwischen den verschiedenen *Cicer*-Arten wurden mehrfach und nicht immer übereinstimmend mit Hilfe von Isoenzym polymorphismen (Kazan & Muehlbauer, 1991; Tayyar & Waines, 1996; Labdi et al., 1996) und Charakterisierung von Samenspeicherproteinen (Ahmad & Slinkard, 1992), Karyotypisierungen (Ohri & Pal, 1991; Ocampo et al., 1992; Galasso et al., 1996; Ahmad, 2000) sowie mit DNA-Markern (RAPD; Ahmad, 1999) und Kreuzbarkeitsdaten (Ladizinski & Adler, 1976, a,b) beschrieben. Alle Studien unterstützen übereinstimmend die hohe Ähnlichkeit von *C. arietinum*, *C. reticulatum* und *C. echinospermum*, wobei *C. reticulatum* als der engste Verwandte der kultivierten Spezies gilt. Auch die zweite Kreuzungsgruppe (*C. bijugum*, *C. pinnatifidum*., *C. judaicum*) gruppieren in vielen Untersuchungen gemeinsam, während *C. cuneatum* meistens als Außenseiter auftritt. Kazan & Muehlbauer (1991) bestätigten den monophyletischen Ursprung der annualen Spezies durch vergleichende Isoenzymanalyse und zeigten die enge Verwandtschaft der perennuellen *C. anatolicum* zur ersten Kreuzungsgruppe. Verschiedene, auf morphologischen Kriterien basierende Zuordnungen zu Sektionen wurden aufgrund dieser molekularen Daten in Frage gestellt und taxonomische Umgruppierungen vorgeschlagen.

Alle annualen Spezies von *Cicer* sind diploide Pflanzen mit einer somatischen Chromosomenzahl von $2n = 16$ (Van der Maesen, 1987; Ahmad, 2000). Der heterochromatische Genomanteil der Wildarten unterscheidet sich nicht stark von *C. arietinum*, sondern ist nur anders zwischen den verschiedenen Chromosomen verteilt (Tayyar et al., 1994, Galasso et al., 1996).

A.II.5. Repetitive DNA im Genom von *C. arietinum*

Über den prozentualen Anteil von repetitiven Sequenzen am Genom von *C. arietinum* ist bisher nichts bekannt. Nicht immer korreliert die steigende Genomgröße direkt mit dem repetitiven Anteil (Smyth, 1991), jedoch gilt die auf Reassoziationskinetiken vieler Pflanzen-DNAs basierende Faustregel von Vedel & Delseny (1987), wonach bei einem DNA-Gehalt von mehr als 2 pg pro haploidem Kern mit mehr als 75% repetitiven Sequenzen im Kerngenom zu rechnen ist. Kultivierte *Beta*-Spezies weisen pro diploidem Kern mit 714 Mbp (1.48 pg) bei *Beta vulgaris* ssp. *esculenta* und 758 Mbp (1.57 pg) bei *Beta vulgaris* ssp. *saccharifera* eine ähnliche Genomgröße wie Kichererbse (738 Mbp, 1.53 pg) auf und enthalten 63% mittel- und hochrepetitive Sequenzen (Arumuganathan & Earle, 1991; Schmidt & Heslop-Harrison, 1993). In einem ähnlichem Größenbereich dürfte der Anteil an repetitiven Komponenten auch bei *C. arietinum* liegen.

Bisher existierten nur wenige Informationen über tandemartig angeordnete repetitive Sequenzen im Kichererbsengenom. Im Rahmen von „Oligonukleotid-Fingerprinting“-Studien an DNA von *C. arietinum* sowie bei der Generation von molekularen Markern aus Mikrosatellitenmotiven (Weising et al., 1992; Sharma et al., 1995 a, b; Huettel, 1996) wurde die Abundanz der letzteren beschrieben und Polymorphismen sowie unterschiedliche Informativität verschiedener Mikrosatelliten nach Hybridisierung auf genomische DNA

detektiert. Die chromosomale Lokalisation der Mikrosatelliten ist je nach Motiv entweder in der centromerischen Umgebung konzentriert (z.B. im Falle von CA), oder zusätzlich auch auf den Chromosomenarmen (z.B. GATA), oder dispers über die gesamten Chromosomen verteilt, jedoch mit Aussparung der centromerischen Region (z.B. TA, A, AAC; Gortner et al., 1998). Auch die ungewöhnliche Konzentration der repetitiven Telomersequenz (TTTAGGG)_n in interstitiellen Bereichen bzw. auf dem kurzen Arm der Chromosomen A und B wurde in dieser Arbeit beschrieben. Über hochrepetitive Satelliten-DNA liegen dagegen bisher keine Informationen vor. Die chromosomale Lokalisation der 18S-5.8S-25S rRNA-Gene auf den Chromosomen A und B sowie der 5S rRNA-Gene auf den Chromosomen B und E wurde durch *in-situ* Hybridisierung detektiert (Abbo et al., 1994; Galasso et al., 1996). Über disperse repetitive Elemente ist noch weniger bekannt. Die Anwesenheit von *Ty1-copia* ähnlichen Sequenzen in *C. arietinum* ist durch *in-situ* Hybridisierung nachgewiesen worden (Brandes et al., 1997b), als die chromosomale Lokalisation einer konservierten Reversen Transkriptase-Domäne in 12 verschiedenen Pflanzenspezies untersucht wurde. Über Sequenz oder genomische Organisation dieser repetitiven Klasse war jedoch nichts bekannt. Parallel zu den hier beschriebenen Arbeiten wurde die partielle Sequenz eines LTR-Retrotransposons veröffentlicht (Sant et al., 2000), worauf in Kapitel II.1.1. näher eingegangen wird.

A.III. Die Fragestellung

Um Struktur und Funktion des Kichererbsengenoms zu verstehen, ist eine gründliche Analyse der Zusammensetzung der repetitiven Komponenten, ihrer genomischen Verteilung und Abundanz, ihrer Struktur und chromosomalen Organisation notwendig. Ferner können isolierte repetitive Sequenzen eine Verwendung in einer vertieften Genomanalyse finden (z.B. als molekulare Marker) oder taxonomisch-phylogenetische Fragestellungen untersuchen helfen. Vor diesem Hintergrund sollten abundante Komponenten im Genom von *C. arietinum* identifiziert werden.

Repetitive DNA kann mit Hilfe verschiedener Strategien aus einem Genom kloniert und identifiziert werden: klonierte Restriktionsfragmente aus genomischer DNA können durch Hybridisierung mit gesamtgenomischer DNA auf hochrepetitive Fragmente durchsucht werden. Ebenso können Fragmente einer bestimmten Sequenz gezielt aus dem Genom amplifiziert und kloniert werden. Beide Ansätze wurden zur Identifikation von abundanten Sequenzfamilien in *C. arietinum* verfolgt. Die Isolation und partielle Charakterisierung einiger hochrepetitiver Sequenzen lieferten wichtige Vorinformationen für diese Arbeit (Staginnus, 1995).

- Zunächst sollten die **abundantesten Elemente** des Kichererbsengenoms identifiziert werden.
- Ferner sollte gezielt nach **weiteren typischen Vertretern repetitiver Sequenzklassen** gesucht werden, um einen möglichst vollständigen Überblick über die Zusammensetzung abundanter Sequenzen im Kichererbsengenom zu ermöglichen. Dazu gehören LTR und non-LTR-Retrotransposonfamilien, die ubiquitär im Pflanzenreich zu finden sind und einen Großteil mancher Pflanzengenome ausmachen.

- Von allen identifizierten Sequenzfamilien sollten **Sequenz** und **Struktur** sowie ihre **genomische Organisation** und **chromosomale Lokalisation** charakterisiert werden.
- Die charakterisierten Sequenzklassen sollten ferner auf ihr **Vorkommen** und ihre **Organisation** in verwandten **Arten der Gattung *Cicer* und anderer Leguminosen** untersucht werden, woraus nicht nur evolutionäre Aspekte der Entwicklung der repetitiven DNA abgeleitet werden können, sondern möglicherweise auch die Eignung der Elemente für taxonomische Fragestellungen erkennbar sein kann.
- Exemplarisch sollte die **Verwendungsfähigkeit** einzelner Sequenzklassen **als molekulare Marker** in der Genomanalyse von *C. arietinum* und für taxonomische Fragestellungen geprüft werden.

B. MATERIAL & METHODEN

B.I. MATERIAL

B.I.1. Pflanzen- und Pilzmaterial

Das Samenmaterial von *C. arietinum* und den Wildarten der Gattung *Cicer* wurden vom "International Center for Agricultural Research in the Dry Areas" (ICARDA, Aleppo, Syrien) und von Prof. F.J. Muehlbauer, Pullman (Washington State University, U.S.A.) zur Verfügung gestellt (Tab. B1). Das Saatgut für die übrigen Pflanzen wurde vom Botanischen Garten der Universität Frankfurt a.M. bezogen (Tab. B2). DNA von *Tillandsia usneoides* wurde von R. Horres (Universität Frankfurt) zur Verfügung gestellt, DNA und Kulturen des Ascomyceten *Ascochyta rabiei* stammen von J. Geistlinger (Universität Frankfurt). Alle Pflanzen wurden im Gewächshaus des Biozentrums der Universität Frankfurt angezogen. Für die Isolation von mRNA aus Wurzelgewebe wurden die Pflanzen auf einem Netz über einem

Tab. B1: Verwendete Akzessionen der Gattung *Cicer*, ihre taxonomische Zuordnung, der Lebenszyklus und ihre geographische Verbreitung.

No.	Spezies (Varietät)	Akzession	Taxonomische Zuordnung (Sektion)	Lebenszyklus	Geographische Verbreitung
1	<i>C. anatolicum</i> Alef.	PI383626	<i>Polycicer</i>	perennuell	Türkei, Iran, Irak
2	<i>C. arietinum</i> L.	ILC3279	<i>Monocicer</i>	annuell	Syrien
3	<i>C. arietinum</i> L.	ILC482	<i>Monocicer</i>	annuell	Türkei
4	<i>C. arietinum</i> L.	ILC4958	<i>Monocicer</i>	annuell	Indien
5	<i>C. arietinum</i> L.	WR315	<i>Monocicer</i>	annuell	Indien
6	<i>C. arietinum</i> L.	C104	<i>Monocicer</i>	annuell	Indien
7	<i>C. bijugum</i> K.H.Rech	ILWC 32	<i>Monocicer</i>	annuell	Türkei
8	<i>C. bijugum</i> K.H.Rech	ILWC 62	<i>Monocicer</i>	annuell	Türkei
9	<i>C. bijugum</i> K.H.Rech	ILWC 42	<i>Monocicer</i>	annuell	Syrien
10	<i>C. chorassanicum</i> (Bge) M.Pop.	ILWC 146	<i>Chamaecicer</i>	annuell	Afghanistan
11	<i>C. chorassanicum</i> (Bge) M.Pop	ILWC 23	<i>Chamaecicer</i>	annuell	Afghanistan
12	<i>C. cuneatum</i> Hochst.ex Rich	ILWC 37	<i>Monocicer</i>	annuell	Ethiopien, Ägypten, Sudan
13	<i>C. cuneatum</i> Hochst.ex Rich	ILWC185	<i>Monocicer</i>	annuell	Ethiopien
14	<i>C. cuneatum</i> Hochst.ex Rich	ILWC187	<i>Monocicer</i>	annuell	Ethiopien
15	<i>C. cuneatum</i> Hochst.ex Rich	ILWC40	<i>Monocicer</i>	annuell	Ethiopien
16	<i>C. echinospermum</i> P.H.Davis	ILWC39	<i>Monocicer</i>	annuell	Türkei
17	<i>C. echinospermum</i> P.H.Davis	ILWC 35	<i>Monocicer</i>	annuell	Türkei
18	<i>C. judaicum</i> Boiss.	ILWC 56	<i>Monocicer</i>	annuell	Libanon
19	<i>C. judaicum</i> Boiss.	ILWC 57	<i>Monocicer</i>	annuell	Libanon
20	<i>C. judaicum</i> Boiss.	ILWC 59	<i>Monocicer</i>	annuell	Libanon
21	<i>C. pinnatifidum</i> Jaub. & SP.	ILWC 170	<i>Monocicer</i>	annuell	Türkei
22	<i>C. pinnatifidum</i> Jaub. & SP.	ILWC 204	<i>Monocicer</i>	annuell	Syrien
23	<i>C. reticulatum</i> Ladiz.	Pi489777	<i>Monocicer</i>	annuell	Türkei
24	<i>C. reticulatum</i> Ladiz.	ILWC21	<i>Monocicer</i>	annuell	Türkei
25	<i>C. yamashitae</i> Kitamura	ILWC53	<i>Monocicer</i>	annuell	Afghanistan
26	<i>C. yamashitae</i> Kitamura	ILWC214	<i>Monocicer</i>	annuell	Afghanistan

luftdurchströmtem Wassergefäß („Hydrokultur“) angezogen. Für die Präparation von Metaphasechromosomen wurden handelsübliche Kichererbsen verwendet, die auf feuchtem

Filterpapier bei 25°C in 3-5 Tagen zur Keimung gebracht und bis zu einer Primärwurzellänge von 1-2 cm kultiviert wurden.

Tab. B.2: Verwendete Spezies, die nicht zur Gattung *Cicer* gehören, und ihre taxonomische Zuordnung.

Spezies (Varietät)	Taxonomische Zuordnung
<i>Lens culinaris</i>	Leguminosae, Dicotyledoneae
<i>Pisum sativum</i> (Sprinter, Schnabel)	Leguminosae, Dicotyledoneae
<i>Phaseolus vulgaris</i> (Goldfisch, Blauhilde, Ruhm d. Vorgebirges)	Leguminosae, Dicotyledoneae
<i>Phaseolus aureus</i>	Leguminosae, Dicotyledoneae
<i>Glycine max</i>	Leguminosae, Dicotyledoneae
<i>Medicago sativa</i>	Leguminosae, Dicotyledoneae
<i>Lathyrus odoratus</i> (rot, blau)	Leguminosae, Dicotyledoneae
<i>Vicia faba</i>	Leguminosae, Dicotyledoneae
<i>Lycopersicon esculentum</i> (Harzfeuer)	Solanaceae, Dicotyledoneae
<i>Nicotiana tabacum</i>	Solanaceae, Dicotyledoneae
<i>Nicotiana atropurpurea</i>	Solanaceae, Dicotyledoneae
<i>Arabidopsis thaliana</i>	Brassicaceae, Dicotyledoneae
<i>Beta vulgaris</i>	Chenopodiaceae, Dicotyledoneae
<i>Brassica napus</i>	Brassicaceae, Dicotyledoneae
<i>Pelargonium spec.</i>	Geraniaceae, Dicotyledoneae
<i>Tillandsia usneoides</i>	Bromeliaceae, Monocotyledoneae
<i>Hordeum vulgare</i>	Poaceae, Monocotyledoneae
<i>Triticum aestivum</i>	Poaceae, Monocotyledoneae
<i>Triticum urartu</i>	Poaceae, Monocotyledoneae
<i>Zea mays</i>	Poaceae, Monocotyledoneae
<i>Allium sativum</i>	Alliaceae, Monocotyledoneae
<i>Musa spec.</i> (AAA) (Eßbanane)	Musaceae, Monocotyledoneae
<i>Musa acuminata</i> x <i>Musa</i> <i>balbisiana</i> (Kochbanane)	Musaceae, Monocotyledoneae
<i>Ascochyta rabiei</i> (Pass)	Ascomycetidae, Ascomycota (Fungi)

B.I.2. Bakterienstämme und Vektoren

Für die Klonierungsexperimente wurden die Vektoren pBluescriptSK (Stratagene) und pGEM-T (Promega) verwendet. pBluescript ist ein 2.958 Kbp langes pUC19-Derivat mit einem Ampicillinresistenzgen und einem in der MCS („multiple cloning site“) lokalisierten lacZ-Gen, das in Verbindung mit X-Gal/IPTG (5-Brom-4-Chlor-3-Indolyl β -D-Galactopyranosid/ Isopropylthiogalactosid) die Selektion rekombinanter Konstrukte durch eine Farbreaktion ermöglicht. pGEM-T (3 Kbp) ist durch Restriktion linearisiert und mit 3'-überhängenden Thymidinresten ausgestattet, was ihn speziell für die Klonierung von PCR-Produkten geeignet macht, da viele thermostabile Polymerasen ein einzelnes Adenosin an das 3'-Ende der Amplifikationsfragmente anhängen („T-A cloning“).

Als Wirtsbakterium dienten die *Escherichia coli*-Stämme SURE (=Stop Unwanted Rearrangement Events“; Stratagene; Greener, 1990) und DH5 α (Gibco BRL; Hanahan, 1983). Beide haben folgende Genotypen:

SURE: e14⁻(*McrA*⁻), Δ (*mcrCB-hsdSMR-mrr*)171, *endA1*, *supE44*, *thi-1*, *gyrA96*, *relA1*, *lac*, *sbcC*, *recJ*, *recB*, *umuC::Tn5*(Kan^r), *unvrC*, [F' *proAB*, *lacIqZAM15*, Tn10(*tet*^r)]

DH5 α : F⁺, ϕ 80*dlacZDM15*, D(*lacZYA-argF*)U169, *deoR*, *recA1*, *endA1*, *hsdR17*(rk⁻, mk⁺), *phoA*, *supE44*, Γ , *thi-1*, *gyrA96*, *relA1*

B.I.3. Restriktionsenzyme

Zur hydrolytischen Spaltung von 3'-5'-Phosphodiesterbindungen doppelsträngiger DNA wurden folgende, z.T. methylierungssensitive Restriktionsendonukleasen des Typs II verwendet:

Tab. B3: Verwendete Restriktionsendonukleasen, ihre Erkennungssequenzen und Wirkungsoptimum (°C). Sternchen zeigen Methylierungssensitivität am davor liegenden Nukleotid an.

Restriktionsendonuklease	Erkennungssequenz	Wirkungsoptimum
<i>Alu</i> I	5'AGC*T 3'	37°C
<i>Apa</i> I	5'GGGC*CC *3'	37°C
<i>Bam</i> H I	5'GGATC*C 3'	37°C
<i>Bsu</i> R I	5'GGCC 3'	37°C
<i>Dde</i> I	5'C*TNAG 3'	37°C
<i>Dra</i> I	5'TTTAAA 3'	37°C
<i>Eco</i> R I	5'GAATTC* 3'	37°C
<i>Eco</i> R V	5'GATATC 3'	37°C
<i>Hae</i> III	5'GGC*C 3'	37°C
<i>Hind</i> III	5'AAGC*TT 3'	37°C
<i>Hinf</i> I	5'GANTC 3'	37°C
<i>Hpa</i> II	5'C*C*GG 3'	37°C
<i>Mbo</i> I	5'GATC 3'	37°C
<i>Msp</i> I	5'C*CGG 3'	37°C
<i>Pst</i> I	5'C*TGCAG 3'	37°C
<i>Pvu</i> II	5'CAGC*TG 3'	37°C
<i>Rsa</i> I	5'GTAC 3'	37°C
<i>Sau</i> 3AI	5'GATC* 3'	37°C
<i>Taq</i> I	5'TCGA 3'	65°C

B.I.4. Oligonukleotide

Folgende Oligonukleotide wurden als PCR-Primer bzw. zur Herstellung von Adaptorkassetten verwendet. Die Synthese der Oligonukleotide erfolgte durch die Firmen Biospring (Frankfurt), Roth (Karlsruhe) oder Eurogentec (Liège, Belgien). Die Sequenzen wurden entweder mit den in I.5.2. aufgelisteten Programmen abgeleitet oder aus der Literatur entnommen.

Tab. B4: Sequenz der verwendeten Oligonukleotide. A=Adenin, C=Cytosin, G=Guanin, T=Thymin, I=Inosin; N=A/C/G/T, Y=C/T, R=A/G, S=C/G, W=A/T, M=A/C, K=G/T, D=nicht C, B=nicht A, H=nicht G, V=nicht T.

Name	5'→3' Nukleotidsequenz
DOM1	GTG GGC GAG TAT GCT GCG
DOM2	CTC CCG AAA TTC TGG TTC CC
DOM3	ATG TGA TCT TGG ATT GTG CTC G
DOM4	CCC ATG GCG ATA CAC TTG GTC
DOM5	GCA ATT GTA TGC CAA TCT GTC G
DOM6	TGC AAT CCG ATA AGC AAC CGG CC
LTRU3	TGT ATT TTG GAT TTT GGC TCG G
LTRU5	CCT GCC TCT CTG CCA ATC G
RePT	TAG TTT CGA GGA CGA AAC TAA TTT TAG GGG GG
T7	TAA TAC GAC TCA CTA TAG GG
SP6	CAC ATA CGA TTT AGG TGA CAC TA
M13 Revers	GGA AAC AGC TAT GAC CAT G
UbiR	AGT TTT CAC AGG AAC CAG AA
UbiL	AAC GAT GTC TGG TGC TAC TC
TyA (Flavell et al., 1992b)	ACN GCN TTY YTN CAY GG
TyB (Flavell et al., 1992b)	ARC ATR TCR TCN ACR TA
Wri1 (Wright et al., 1996)	GGG ATC CNG GNC CNG AYG GNW T
Wri2 (Wright et al., 1996)	GGN KRY CCN YTN WSN CCG AAT TCC GGA ATT CGG ISW IAR IGG RYM ICC YTG
repL	TMT GTG YTT GGT TCC CGA
repR	GTT GWA CTM RTA WTG CCT CAA
transAF	GGA AAA CTA ATA TGA TTC GAC ATA ATA TTG AYI TIA TGC A
transCR	CAT AGA CAG ATG TCC ATA TCT TTC AAA ASD RTA CAT CCA
EcoNull	GAC TGC GTA CCA ATT C
EcoAdapter1	CTC GTA GAC TGC GTA CC
EcoAdapter2	AAT TGG TAC GCA GTC TAC

B.I.5. Verwendete Computerprogramme

B.I.5.1. Auswertung der Sequenzdaten

- DNASIS (Hitachi Software) zur Zusammensetzung von Einzelsequenzen und Detektion von kurzen repetitiven Strukturen.
- BLASTN, BLASTX (www3.ncbi.nlm.gov) zur Detektion homologer Sequenzen in den Nukleotidsequenzdatenbanken von GenBank, EMBL, DDBJ, PDB bzw. Peptidsequenzdatenbanken non-redundanter GenBank-CDS-Translationen, PDB, SwissProt, PIR und PRF (Altschul et al., 1990; 1997)- auch als „NCBI („national centre of biology and informatics“-Datenbanken“ bezeichnet.
- CLUSTALW (1.74) zum Homologievergleich verschiedener Sequenzen (Higgins et al., 1988) und zur Visualisierung der genetischen Distanz durch Dendrogramme nach der „neighbour-joining“-Methode (Saitou & Nei, 1987).
- TREEVIEW zur graphischen Darstellung von Dendrogrammen (Page, 1996).
- WINCLONE 4.0 (Microsoft) zu Vergleich und Analyse der codierten Leseraster und Detektion von Restriktionsschnittstellen.
- BLOCK MAKER (<http://bioinformatics.weizmann.ac.il/blocks/blockmkr>) zum Vergleich von Aminosäuresequenzen und zur Identifikation konservierter Bereiche zur Ableitung von Oligonukleotiden.

B.I.5.2. Ableitung von Oligonukleotiden

- CODEHOP („Consensus degenerate hybrid oligonucleotide primer“) (<http://bioinformatics.weizmann.ac.il/blocks/codehop>) zur Ableitung degenerierter Oligonukleotide aus Aminosäuresequenzen (Rose et al., 1998).
- OLIGO vs. 3.4 (www.oligo.net; Molecular Biology Insights Inc., West Cascade, USA) zur Ableitung von Oligonukleotiden aus Nukleotidsequenzen.

B.I.5.3. Integration von Markern in die genetische Karte

- MAPMAKER 3.0 zur Kopplungsanalyse der S-SAP Marker (Lander et al., 1987)

B.I.5.4. *in-situ* Hybridisierung

- PHOTOSHOP 5.0 (Adobe) zum Bearbeiten und Drucken des digitalisierten fotografischen Materials.

B.II.Methoden

B.II.1. Isolation von Nukleinsäuren aus verschiedenen Pflanzen

B.II.1.1. Präparation von Gesamt-DNA aus Pflanzengewebe (modifiziert nach Saghai-Marooft et al., 1984)

Genomische DNA wurde durch Zellaufschluß aus Blättern und Sprossen nach der Methode von Saghai-Marooft et al. (1984) isoliert, wobei jedoch auf die Zugabe von β -Mercaptoethanol verzichtet wurde. Größere Mengen von mitisolierten Polysacchariden wurden nachfolgend durch eine selektive Fällung mit 0.25M NaCl von der DNA abgetrennt (Michaels et al., 1994). Bei seneszentem Gewebe kann der Fällungsschritt mehrfach wiederholt werden. Die Methode ist effizient und liefert für alle enzymatischen Reaktionen zugängliche DNA, die nicht mehr im CsCl-Gradienten gereinigt werden muß.

• Arbeitsschritte

- ca. 4 g frisches Pflanzenmaterial mit flüssigem Stickstoff schockgefrieren und durch Mörsern mechanisch aufschließen.
- 30 min bei 60°C in 12 ml vorgewärmtem CTAB (Cetyltrimethylammoniumbromid)-Puffer inkubieren.
- 12 ml Chloroform/Isoamylalkohol [24:1 (v/v)] zugeben und weitere 15 min bei RT schütteln.
- 20 min lang bei 5000 Upm/RT zentrifugieren (Heraeus Sepatech, Biofuge A; Rot. Cat. No. 1378) und wäßrige Oberphase mit 6 μ l DNase-freie RNase (10 mg/ml) 30 min bei RT inkubieren.
- Reaktionsansatz mit 1 Vol. Isopropanol mischen, 15 min bei RT inkubieren, 20 min bei 5000 Upm/RT zentrifugieren.
- Präzipitat in 5 ml 10 mM Tris (Trishydroxymethylaminomethan)-HCl (pH 8,0)/250 mM NaCl lösen und unter Schwenken 0.35 Vol. Ethanol zugeben, 15 min bei 4°C inkubieren (Eis), 20 min bei 5000 Upm/ RT zentrifugieren.
- Überstand mit 1 Vol. Isopropanol versetzen, 15 min bei RT inkubieren.
- DNA durch leichtes Schütteln bzw. Zentrifugieren (20 min bei 5000 Upm/RT) präzipitieren, mit 70% Ethanol waschen, lufttrocknen und in 0,5-1 ml 10 mM Tris-HCl (pH 8,0) lösen.

• Lösungen

CTAB-Puffer

2% (w/v) Cetyltrimethyl-Ammoniumbromid
1.4 M NaCl
0.1 M Tris-HCl (pH 8,0)
20 mM EDTA (Ethylendiamintetraacetat)

B.II.1.2. Isolation von mRNA aus Pflanzengewebe

Für die Präparation von mRNA aus Blättern, Sprossen, Wurzeln und Blüten von verschiedenen *Cicer*-Spezies wurde der „MPG Guanidine Direct mRNA Purification Kit“ (CPG) verwendet. Bei dieser Methode verhindert ein Guanidinthiocyanat-Puffer die RNA-Degradation durch Inaktivierung von RNAsen. Prozessierte mRNA weist an ihrem 3'-Ende eine Poly(A)-Abfolge auf und kann so durch Hybridisierung mit Oligo(dT)-Ankern und magnetischer Trennung selektiv von der restlichen RNA separiert und konzentriert werden. Durch die Effizienz der Methode können selbst aus geringen Gewebemengen ausreichende mRNA-Konzentrationen zur cDNA-Synthese isoliert werden. Sehr wichtig ist steriles Arbeiten sowie die Sterilisation aller Lösungen und Geräte mit 0.1% DEPC (Diethylpyrocarbonat) bzw. durch Autoklavieren (40min/120°C oder 8h/220°C; Sambrook et al., 1989), um eine RNA-Degradation durch RNAsen zu vermeiden.

• Arbeitsschritte

1. Vorbereitung der Oligo (dT)-Magnetkugeln

(Bindung von biotinyliertem Oligo(dT)₂₅ an Streptavidin-Magnetkugeln)

- pro Isolationsansatz 50µl der Streptavidin-Magnetkugellösung in einem Eppendorfgesäß auf dem Magnetständer plazieren.
- nach ca. 30sec Überstand verwerfen, Magnetkugeln mit 100µl Puffer G1 mischen und auf dem Magnetständer vom Überstand trennen; Arbeitsschritt 2x wiederholen.
- Magnetkugeln in 90µl Puffer G1 resuspendieren, 10µl (800 pmol) biotinyliertes Oligo(dT)₂₅ zugeben, 7 min vortexen und Gefäß invertieren.
- Überstand auf dem Magnetständer entfernen und 3x mit je 100µl Puffer G2 waschen.
- in 50µl Puffer G4 resuspendieren und bei RT lagern.

2. Aufschluß der Pflanzengewebe

- 1.5 ml Puffer G3 mit 15µl β-Mercaptoethanol mischen und auf 4°C kühlen; alle Geräte auf 4°C kühlen und weitere Schritte ebenfalls bei 4°C durchführen.
- ca. 150-180 mg frisches Gewebematerial mit kaltem, DEPC-behandeltem Wasser spülen, unter flüssigem Stickstoff zermörsern und zügig in Puffer G3/β-Mercaptoethanol suspendieren.
- 3 ml Puffer G4 zugeben, vortexen und Gemisch gleichmäßig auf 4 Eppendorfgesäße verteilen.
- 2 min bei 12000 rpm/ 4°C zentrifugieren (Heraeus Sepatech, Minifuge RF; Rot. Cat. No. 2150), auf Eis lagern.

3. Anreicherung der mRNA

- Überstand des ersten Gefäßes mit 50µl der vorbereiteten Oligo(dT)-Magnetkugelsuspension vortexen, Gefäß 4 min bei RT invertieren.
- Überstand auf Magnetständer entfernen, Überstand des zweiten Eppendorfgefäßes zugeben, vortexen und Gefäß 5 min bei RT invertieren.
- mit den beiden restlichen Eppendorfgefäßen ebenso verfahren, wobei die Mischzeit sich jedesmal um 1 min verlängert.
- Magnetkugeln mit gebundener mRNA in 500µl Puffer G5 resuspendieren und den Überstand im Magnetständer entfernen, Arbeitsschritt 2x wiederholen.
- in 15µl Puffer G6 resuspendieren, 2min bei 65°C inkubieren und den Überstand mit der abgelösten mRNA auf dem Magnetständer zügig abnehmen; Arbeitsschritt wiederholen.
- mRNA in 5µl Aliquots bei -80°C lagern.

B.II.1.3. Präparation von cDNA aus pflanzlicher mRNA

Isolierte mRNA (II.1.2.) wurde mit Hilfe einer Reversen Transkriptase (Superscript II; Gibco BRL) und eines Oligo(dT)₁₂₋₁₈ Fragments (Gibco BRL) in cDNA umgeschrieben, die direkt als Templat für einen PCR-Ansatz weiterverwendet werden kann. RNase-Inhibitor (MBI Fermentas) und DEPC-Sterilisation des verwendeten H₂O bidest verhinderte während der Reaktion die Degradierung des Templats durch RNAsen. Nach der cDNA-Synthese wurde die überschüssige RNA durch Zugabe von RNase abgebaut. Die cDNA ist stabil und kann mehrere Monate bei -80°C gelagert werden.

• Arbeitsschritte

- ein 25µl dNTP Mix aus je 10mM jedes Desoxyribonucleotids (MBI Fermentas) wird mit H₂O bidest (DEPC-behandelt) angesetzt und bei 4°C gelagert.
- für einen 30µl Reaktionsansatz werden 5µl mRNA (50-500ng), 3µl Oligo(dT) (250µg/ml) und 10.5µl H₂O bidest (DEPC-behandelt) gemischt, 10 min bei 70°C inkubiert und 10 min auf Eis gelagert.
- 6µl 5x „first strand“ Puffer (250mM TRIS-HCl/ pH 8.3, 375 mM KCl, 15mM MgCl₂; Gibco BRL), 3µl 0.1M DTT (Dithiothreitol), 1µl RNase-Inhibitor (30U; MBI Fermentas), 1.5µl des dNTP Mix werden hinzupipettiert und der Ansatz 2 min bei 42°C inkubiert.
- 1.5µl Superscript II (300U; Gibco BRL) werden zügig mit dem Ansatz vermischt und weitere 50 min bei 42°C inkubiert.
- Inaktivierung der aktiven Proteinkomponenten durch 15 min Inkubation bei 70°C.
- optional mit 3µl DNase-freier RNase (10 mg/ ml) 30 min lang bei RT inkubieren; Isolate bei -80°C lagern.

B.II.2. DNA-Konzentrationsbestimmung

Die Konzentrationsbestimmung von Plasmid- und Pflanzen-DNA erfolgte durch parallele Agarosegelelektrophorese mit λ -DNA bekannter Konzentration und hinreichend verdünnten DNA-Proben. Nach Ethidiumbromidfärbung konnte die Konzentration der Proben aufgrund des bei niedrigen Konzentrationen linearen Zusammenhangs zwischen Ethidiumbromidfluoreszenz und DNA-Konzentration mit Hilfe einer Videokamera und der Software „Bioprofile“ (Vilber Lourmat) bestimmt werden. Diese Methode erlaubt innerhalb kürzester Zeit die Konzentrationsbestimmung und den Mengenabgleich einer hohen Probenanzahl.

B.II.3. Restriktion genomischer und bakterieller DNA

Um pflanzliche genomische DNA und bakterielle Plasmide gezielt zu fragmentieren, wurden Restriktionsendonukleasen des Typs II verwendet. Die Reaktionsansätze wurden gemäß den Vorschriften des Herstellers aus 5U/ μ g DNA Enzym, dem vom Enzym abhängigen Reaktionspuffer und DNA gemischt. Für Plasmidrestriktionen wurden 0.15-1 μ g DNA und 1U/ μ g DNA Enzym eingesetzt, für genomische DNA 8-13 μ g DNA und 5U/ μ g Enzym. Die Ansätze wurden im Temperaturoptimum des Enzyms (Tab. B3) 1.5-8 h inkubiert, und die aktiven Proteinkomponenten durch 10 min Erhitzen auf 65°C inaktiviert. Bei Mehrfachrestriktionen wurden vom Hersteller empfohlene Reaktionsbedingungen eingehalten, die eine mindestens 50%ige Aktivität aller Enzyme zulassen. Der Erfolg einer Restriktion wurde durch gelelektrophoretische Trennung der Fragmente kontrolliert.

B.II.4. Gelelektrophoretische Trennung von DNA-Fragmenten

Aufgrund der negativ geladenen Phosphatreste des Desoxyribose-Phosphatgerüsts bewegen sich DNA-Moleküle in allen pH-Bereichen, wenn ein elektrisches Feld angelegt wird. Bei der Bewegung in Richtung der Anode können sie nach ihrem Molekulargewicht getrennt werden.

B.II.4.1. Trennung durch Agarosegelelektrophorese

Die horizontale Agarosegelelektrophorese wurde bei 1-5 V/cm in einer Flachbettapparatur in 1x TBE (Tris-Borat-EDTA)-Puffer durchgeführt. Je nach Größe der zu trennenden DNA-Fragmente wurde eine Agarosekonzentrationen von 0.6-1.8 % gewählt. Der Zusatz von Ethidiumbromid (EtBr) (1 μ g/ml in 1xTBE), das mit doppelsträngiger DNA interkaliert, machte die DNA-Proben im UV-Licht durch Fluoreszenz sichtbar.

- Arbeitsschritte

- in 1x TBE suspendierte Agarose wird durch Aufkochen gelöst.
- das Gel wird mit ca. 60°C warmer Agaroselösung gegossen, nach Erstarren in die Flachbettapparatur gelegt, mit 1x TBE überschichtet und der Probenkamm entfernt.
- DNA-Lösungen werden mit Auftragspuffer versetzt, in die Probentaschen pipettiert, die Elektrophorese gestartet.
- das Gel kann nach Ende der Elektrophorese 15 min mit EtBr-Lösung angefärbt werden, alternativ kann EtBr auch vor dem Lauf direkt dem Gel zugesetzt werden, um das Laufverhalten während der Elektrophorese zu kontrollieren.
- Dokumentation unter UV-Licht.

• Lösungen

6x Auftragspuffer

0,25% (w/v) Bromphenolblau
0,25% (w/v) Xylencyanol
40% (w/v) Glycerin

10x TBE

0.89 M TRIS
0.89 M Borsäure
20mM EDTA, pH 8.0

B.II.4.2. Trennung durch denaturierende Polyacrylamidgele

Vertikale, 4%ige Polyacrylamid-Harnstoff-Gele können Einzelstrangfragmente bis zu einer Größe von 1Kb mit einem Auflösungsvermögen von einem Nukleotid trennen. Die Polymerisierung von Acrylamid wird durch freie Radikale aus APS (Ammoniumpersulfat) nach Zugabe von TEMED (N,N,N',N'-Tetramethyldiamin) gestartet. Die Harnstoffkonzentration und ein Formamidpuffer ermöglichen nach initialer Denaturierung der Probe eine Trennung von einzelsträngiger DNA. Die Elektrophorese erfolgte in 1x TBE bei 100-130 W und 47°C in Sequenziergelapparaturen (Sequi-Gen GT; BioRad), dabei waren die getrennten Fragmente radioaktiv markiert und wurden durch Autoradiographie (II.11.4.) sichtbar gemacht.

• Arbeitsschritte

- die Glasplatten werden mit 70% Ethanol gereinigt, dann wird eine Platte mit Dimethyldichlorsilan-Lösung (Repelsilane ES; Pharmacia) behandelt, um ein Ablösen des hydrophilen Gels zu ermöglichen; die Gelapparatur wird zusammengesetzt und abgedichtet.
- für das 34 x 50 x 0.05 cm große Gel werden 100ml Gellösung gemischt, zwischen die Platten gegossen und 1 h zur Polymerisation ruhen gelassen.
- das fertige Gel wird in die Apparatur gesetzt, mit 1x TBE überschichtet und 30 min einer Präelektrophorese bei 130W und 47°C unterzogen.

- die Proben werden mit 1Vol. Formamidpuffer versetzt, 3 min bei 96°C denaturiert und in die mehrfach freigespülten Probenetaschen pipettiert; die Elektrophorese wird gestartet.
- nach der Elektrophorese werden die Glasplatten getrennt, das Gel durch Auflage eines Whatman 3MM-Papiers (Schleicher & Schüll) abgelöst, mit Cellophan abgedeckt und 1h unter Vakuum bei 80°C getrocknet.
- nach Entfernung des Cellophans Dokumentation durch Autoradiographie.

• Lösungen

Gellösung

100ml Acrylamid-Mischung
100 µl APS
160 µl TEMED (N,N,N',N'-Tetramethylethylen-diamin)

APS (frisch)

20% (w/v) Ammoniumpersulfat in H₂O

Acrylamid-Mischung

0.1 Vol. Acrylamid-Bisacrylamid Konzentrat
7M Harnstoff
1 x TBE

Formamid-Auftragspuffer

98% Formamid
10mM EDTA (pH8.0)
0.05% (w/v) Bromphenolblau
0.05% (w/v) Xylencyanol

Acrylamid-Bisacrylamid Konzentrat (Roth)

38% (w/v) Acrylamid
2% (w/v) N,N'-Methylenbisacrylamid

B.II.4.3. Eluieren separierter Fragmente aus Agarosegelen

Nach der gelelektrophoretischen Trennung wurden DNA-Fragmente aus der Gelmatrix isoliert, gereinigt und für Klonierungs- oder Markierungsreaktionen weiterverwendet. Für Fragmente zwischen 0.1-4 Kbp können der „Qiaex II Agarose Gel Extraction“ Kit (Qiagen) und zwischen 0.07-10 Kbp der „Qiaquick Gel Extraction“ Kit (Qiagen) verwendet werden, wobei letzterer die Konzentrierung der DNA erleichtert. Bei beiden Methoden wird die Agarose aufgelöst, die gelöste DNA bei hohen Salzkonzentrationen an eine Silicagelmatrix (Kugeln oder Membran) adsorbiert, gewaschen und bei niedrigeren Salzkonzentrationen selektiv eluiert.

• Arbeitsschritte

1. QiaexII-Kit

- die Bande wird aus dem ethidiumbromidgefärbten Gel ausgeschnitten, 3Volt Puffer QX1 und 5µl resuspendiertes QiaexII (bis 200 mg Gel) zugegeben; 10 min Inkubation bei 50°C; bis das Gel aufgelöst ist.
- Probe 1 min bei 13 000 Upm zentrifugieren (Heraeus Sepatech, Minifuge RF; Rotor 2150), Überstand entfernen, Sediment mit 500µl Puffer QX1 waschen.
- Zentrifugation wiederholen, Sediment mit 500 µl Puffer PE waschen; Arbeitsschritt wiederholen.

- Sediment lufttrocknen lassen und in 20 µl 10mM Tris, pH 8.5 suspendieren, 10-20 min bei RT inkubieren, 2 min bei 13000 Upm zentrifugieren.
- Überstand abnehmen und bei 4°C lagern.

2. Qiaquick-Kit

- die Bande wird aus dem ethidiumbromidgefärbten Gel ausgeschnitten, 3Vol. Puffer QG zugegeben; 10 min Inkubation bei 50°C; bis das Gel aufgelöst ist.
- 1Vol. Isopropanol zugeben, Gemisch auf eine Qiaquick-Säule pipettieren, 1min bei 13 000 Upm zentrifugieren (Heraeus Sepatech, Minifuge RF; Rotor 2150).
- Filtrat verwerfen, 500µl Puffer QG zugeben, zentrifugieren, Filtrat verwerfen, 750 µl Puffer PE zugeben, zentrifugieren, Filtrat verwerfen.
- erneut zentrifugieren, um restliches Filtrat zu entfernen, Säule in ein Eppendorfgefäß setzen, 12 µl 10 mM TRIS-HCl (pH 8.5) zugeben und 30 min bei RT inkubieren.
- 2 min zentrifugieren, Filtrat bei 4°C lagern.

B.II. 5. Southern-Transfer

Nach der Auftrennung durch Gelelektrophorese wurden denaturierte DNA-Fragmente durch die Kapillarkräfte eines Pufferstroms aus der Gelmatrix auf eine Nylonmembran überführt und dort irreversibel fixiert (Southern, 1975). Die Nylonmembranen konnten mehrmals zu Hybridisierungsexperimenten verwendet werden.

B.II.5.1. Southerntransfer aus Agarosegelen

Die DNA-Fragmente im Agarosegel wurden durch alkalische Vorbehandlung denaturiert, die Einzelstränge durch hohe Salzkonzentrationen stabilisiert und auf der Nylonmembran durch hohe Temperaturen fixiert.

• Arbeitsschritte

- das Agarosegel wird zugeschnitten und nacheinander je 20 min in Denaturierungspuffer 1, Neutralisierungspuffer 1 und Salzpuffer geschwenkt.
- das Gel wird auf ein Whatman 3MM-Papier (Schleicher & Schüll) gelegt, das an beiden Enden in 20x SSC („saline sodium citrat“; Natriumcitratsalz) eintaucht; eine Nylonmembran (Hybond N; Amersham) wird blasenfrei aufgelegt, darüber werden zwei Schichten in 3xSSC angefeuchtetes Whatman 3MM-Papier gelegt und mit trockenem saugfähigem Papier überschichtet.
- nach 4-12h ist der Kapillartransfer beendet, die Nylonmembran wird abgenommen und 90 min bei 180°C inkubiert.

- Lösungen

Denaturierungspuffer 1

0.5 M NaOH
1.5 M NaCl

Neutralisierungspuffer 1

1M TRIS-HCl, pH 7.0

20x SSC

3 M NaCl
0.3 M Na₃Citrat (pH 7.0)

Salzpuffer

0.5 M TRIS-HCl, pH 7.0
1,5 M NaCl

B.II.5.2. Southerntransfer aus Polyacrylamidgelen

Polyacrylamidgele sind aufgrund ihrer Fragilität schwieriger zu handhaben, weshalb die Denaturierungsschritte nach dem Transfer auf die Nylonmembran stattfanden.

- Arbeitsschritte

- Polycarylamidgel nach Trennung der Glasscheiben durch Auflegen der Nylonmembran von der Platte abziehen und auf den Transferaufbau überführen (II.5.1.).
- nach 4-12h ist der Kapillartransfer beendet, die Nylonmembran wird kurz bei RT getrocknet und nacheinander je 1 min mit Denaturierungspuffer 2, Neutralisierungspuffer 2 und Salzpuffer getränkt; zwischen den einzelnen Schritten wird die Membran jeweils bei RT getrocknet.
- die trockene Membran wird 90 min bei 180°C inkubiert.

- Lösungen

Salzpuffer (siehe II.5.1.)

Denaturierungspuffer 2

0.5 M NaOH

Neutralisierungspuffer 2

1 M TRIS-HCl, pH 7.5
0.15 M NaCl

B.II.5.3. Nukleinsäuretransfer aus Bakterienkulturen

Um die Hybridisierung einer großen Zahl von klonierten Fragmenten zu ermöglichen, wurden Bakterienzellen alkalisch lysiert und die denaturierte DNA auf einer Nylonmembran fixiert.

- Arbeitsschritte

- Bakterienkolonien werden auf einer Nylonmembran, die auf LB-Agar aufliegt (s.II.6.2.), angezogen.

- Nylonmembran wird abgezogen und 5 min lang mit der Bakterienseite nach oben auf ein Whatman 3MM-Papier (Schleicher & Schüll) gelegt, das mit Denaturierungspuffer 2 getränkt ist.
- die Membran wird bei RT getrocknet, 10 min lang mit Neutralisierungspuffer 2 inkubiert, getrocknet und mit Salzpuffer inkubiert.
- die getrocknete Membran wird 90 min bei 80°C gebacken, dann mit Waschlösung getränkt und durch leichtes Reiben von Bakterienresten befreit.

• Lösungen

Waschlösung

50 mM EDTA, pH 8.0

1% SDS

B.II.6. Klonierung von DNA

Ausgewählte DNA-Fragmente wurden in Vektor-DNA eingebaut und durch Bakterienzellen vermehrt. Die kovalente Integration des Fragments in die linearisierte Vektor-DNA wird durch T4-Ligase unter Energieaufwand katalysiert (Ligation). Das Konstrukt konnte durch einen Elektroschock, der kurzzeitig die Membranpermeabilität erhöht (Elektroporation), in kompetente Bakterienzellen transferiert werden (Transformation).

B.II.6.1. Ligation von DNA-Fragmenten

Ligationen wurden sowohl zwischen PCR-Fragmenten und dem Vektor pGEM-T (Promega) als auch zwischen *EcoRI*-Restriktionsfragmenten und dem Vektor pBluescript SK (Stratagene) durchgeführt. Für letztere muß der Vektor passende kohäsive Enden aufzuweisen und vorher ebenfalls durch *EcoRI* linearisiert werden. Bei der Ligation kohäsiver Enden sollen die Restriktionsfragmente in leichtem Überschuß zum Vektor vorliegen, um Selbstligation bzw. Ligation mehrerer Fragmente zu vermeiden. Zusätzlich kann die Dephosphorylierung des Vektors vor der Ligation, d.h. die Abspaltung der freien 5'-Phosphatgruppe durch alkalische Phosphatase, seine Selbstligation reduzieren (Sambrook et al., 1989). pGEM-T ist bereits linear und kann aufgrund der 3'-überhängenden Thymidinreste keine Selbstligation eingehen (I.2.), so daß hier auch weniger Fragment-DNA eingesetzt werden kann. Für die Ligation in pGEM-T wurde das „pGEM-T Vector System I“ Kit (Promega) verwendet.

• Arbeitsschritte

1. pBluescriptSK

- die Plasmid-DNA von pBluescriptSK wird durch *EcoRI*-Restriktion linearisiert, mit je 1 Vol. Phenol/Chloroform ausgeschüttelt und mit 1/10 Vol. 3 M Kaliumacetat (pH 5.2); 2.5 Vol. 100% Etanol 20 min bei -80°C gefällt;

- das Sediment wird mit 70% Ethanol gewaschen, in 10 mM TRIS-HCl (pH 8.0) gelöst.
- Plasmid-DNA wird mit 0.5 U alkalischer Phosphatase (Boehringer)/ μg in 0.1 Vol. 10x Puffer [10mM ZnCl_2 , 10 mM MgCl_2 , 100 mM Tris-HCl (pH 8,3)] bei 37°C für 1h inkubiert, das Enzym 10 min bei 65°C inaktiviert.
 - die DNA wird durch Phenol/Chloroform und Ethanol-fällung gereinigt und in 0.1x TE gelöst.
 - dephosphorylierte Plasmid-DNA und genomische Restriktionsfragmente mischen, 5 min bei 45°C inkubieren, auf Eis lagern.
 - 0.1 Vol. 10x Ligasepuffer (50 mM Tris-HCl [pH 7,5], 7 mM MgCl_2 , 1 mM Dithiothreitol), 1mM ATP (Adenosintriphosphat) und 2 U T4-Ligase (pro μg DNA) zugeben, bei 4°C 8-12h inkubieren.
 - das Enzym wird 10 min lang bei 65°C inaktiviert, der Ansatz durch Phenol/Chloroform und Ethanol-fällung gereinigt, das Präzipitat in H_2O bidest gelöst, bei -80°C gelagert.

2. pGEM-T

- für einen 10 μl Ligations-Ansatz werden ca. 50-60 ng Quiaquick gereinigte PCR- Fragmente (II.8.4.) und 50 ng pGEM-T Vektor-DNA auf Eis gemischt; 0.1 Vol. 10x Ligase Puffer, 1mM ATP, H_2O bidest und 5U T4-Ligase auf Eis dazugegeben und 8-12h bei 4°C inkubiert.
- das Enzym wird 10 min lang bei 65°C inaktiviert und ohne weitere Reinigung zur Transformation verwendet; Lagerung bei -80°C.

B.II.6.2. Herstellung kompetenter Bakterienzellen und Transformation

Vor der Elektroporation mußten die *E. coli*-Zellen zur Aufnahme und Propagierung von Plasmiden vorbereitet, d.h. kompetent gemacht werden. Die klonierten Fragmente der transformierten Zellen waren durch Plasmidisolierung, Vektor-PCR oder Kolonie-Hybridisierung analysierbar.

• Arbeitsschritte

1. Herstellung kompetenter Zellen

- 30 ml LB-Medium (bei SURE: mit Antibiotikum) werden mit den *E. coli*-Zellen (I.2.) beimpft und über Nacht bei 37°C geschüttelt.
- 20 ml der Vorkultur werden in 1l LB-Medium überführt, bei 37°C geschüttelt, bis eine OD_{600} von 0,6-0,8 erreicht ist, und 30 min auf Eis inkubiert.
- die Zellen werden 15 min bei 4000 rpm und 4°C abzentrifugiert und in 1l eiskaltem H_2O bidest resuspendiert; die Zentrifugation wird wiederholt und das Bakterienpellet in 500 ml H_2O bidest aufgenommen.

- nach erneuter Zentrifugation wird das Sediment in 20 ml 10% (v/v) Glycerin resuspendiert; der Arbeitsschritt wird wiederholt, die Zellen in insgesamt 2 ml 10% (v/v) Glycerin aufgenommen und aliquotiert (50µl); bei -80°C lagerbar.

2. Transformation

- 50µl Bakterienzellen werden mit 2µl Ligationsansatz (bzw. Plasmid) bei 4°C gemischt und in eine gekühlte Elektroporationsküvette überführt.
- in einer Elektroporationsapparatur (Gene Pulser; BioRad) werden sie 1.6-2.5 kV Spannung ausgesetzt, sofort in 750µl vorgewärmtes SOC Medium überführt und 1h bei 37°C inkubiert.
- 50-200 µl werden auf LB-Agarplatten mit X-Gal, IPTG und Ampicillin ausplattiert und über Nacht bei 37°C inkubiert; der restliche Transformationsansatz ist mit 10-20% (v/v) Glycerin bei -80°C lagerbar (danach max. 50µl ausplattieren).
- gegebenenfalls Bakterien auf frische LB-Agarplatten vereinzeln bzw. einen identischen Abdruck des Koloniemusters auf Nylonmembran abziehen, umgekehrt auf eine frische LB-Agarplatte legen und über Nacht bei 37°C inkubieren (Agarplatten bei 4°C lagerbar).

• Lösungen

Luria-Bertani (LB) Flüssigmedium

1% (w/v) Baktotrypton
0.5% (w/v) Hefeextrakt
1% (w/v) NaCl

LB-Agarplatten mit X-Gal/ IPTG

1.5% Agar
20µg/ml X-Gal
(5-Brom-4-Chlor-3-Indolyl β-D-Galactopyranosid)
20µg/ml IPTG (Isopropylthiogalactosid)
in LB-Flüssigmedium

SOC-Medium

2% (w/v) Bacto-Tryptone
0,5% (w/v) Bacto-Hefeextrakt
10 mM NaCl
2.5 mM KCl
10 mM MgCl₂
10 mM MgSO₄
20 mM Glucose

Antibiotikakonzentration

für rekombinante Bakterien
(Flüssigmedium und Platten)
100µg/ml Ampicillin

B.II.7. Isolation klonierter Fragmente

Klonierte Fragmente konnten durch Isolation des gesamten Plasmids und nachfolgende Restriktion oder durch eine Vektor-PCR mit universellen Oligonukleotidprimern, die flankierend zur multiplen Klonierungsstelle binden (s.II.8.2.), aus den Bakterien isoliert und zur Sondenmarkierung oder Sequenzierung weiterverwendet

werden. Um eine ausreichende Konzentration des Fragments bereitzustellen, mußten die Bakterienzellen vermehrt werden.

B.II.7.1. Bakterienanzucht

Die Anzucht der *E.coli* (*Escherichia coli*)-Zellen zur Vermehrung und Plasmidisolierung erfolgte in LB-Flüssigmedium, während die Kultivierung zur Vereinzelnung oder Hybridisierung auf LB-Agarplatten stattfand (II.5.3.; II.6.2.). *E. coli* SURE-Zellen wurden in LB-Flüssigmedium mit 12.5 µg/ml Tetracyclin und 50 µg/ml Kanamycin vermehrt. DH5α wurde ohne Antibiotikazusatz angezogen. Rekombinante Bakterien wurden mit 100 µg/ml Ampicillin kultiviert. Langfristig waren Bakterienkulturen in LB-Flüssigmedium mit 10-20% (v/v) Glycerin oder auf Glycerinplatten bei -80°C lagerbar.

• Arbeitsschritte

- für eine Standard-Plasmidisolierung 3ml LB-Flüssigmedium in verschließbaren Reagenzgläsern mit einer Einzelkolonie beimpfen und über Nacht bei 37°C schütteln.
- zur Vereinzelnung werden Bakterien auf LB-Agarplatten ausgestrichen bzw. Kolonien mit einem sterilen Zahnstocher in geordneter Reihenfolge überführt und über Nacht bei 37°C inkubiert.

• Lösungen

Glycerin-Agarplatten

2% (w/v) Agar

20% (v/v) Glycerin

in LB-Flüssigmedium

B.II.7.2. Isolation des klonierten Fragments

Die Isolation der Bakterienplasmide erfolgte durch alkalische Lyse nach Birnboim & Doly (1979), modifiziert nach Sambrook et al. (1989). Dabei lysiert das alkalische Milieu die Bakterienzelle, denaturiert Proteine und chromosomale DNA des Bakteriums, welche als Salze gefällt werden. Die als „supercoil“ vorliegende Plasmid-DNA bleibt in Lösung und wird erst nach den Reinigungsschritten mit Ethanol gefällt.

• Arbeitsschritte

- 1.5 ml der LB-Flüssigkultur werden 1.5 min bei 12000 Upm/ RT pelletiert (Heraeus Sepatech, Minifuge RF; Rotor 2150), der Arbeitsschritt mit den zweiten 1.5 ml im selben Gefäß wiederholt.
- 150µl eiskalte Lösung 1 wird zugegeben, das Sediment sorgfältig resuspendiert; 150µl Lösung 2 werden zugegeben und das Gefäß max. 3 min invertiert.

- 300µl Lösung 3 werden zugegeben, das Gefäß mehrfach invertiert, 15 min auf Eis inkubiert und 20 min bei 12 000 rpm/ 4°C zentrifugiert.
- der Überstand wird mit 4µl DNase-freier RNase versetzt und 30 min bei 37°C inkubiert.
- das Gemisch wird 2x mit je 1Vol. Chloroform/Isoamylalkohol (24:1) ausgeschüttelt, die wässrige Oberphase mit 2.5 Vol. Ethanol versetzt, 15 min bei RT inkubiert und 15 min bei 12 000 Upm/ RT abzentrifugiert.
- das Sediment wird mit 500µl 70% Ethanol gewaschen, luftgetrocknet und in 15µl 10mM TRIS-HCl (pH8.0) aufgenommen und bei 4°C gelagert.

• Lösungen

Lösung 1	Lösung 2	Lösung 3
50 mM Glucose	(in Plastikgefäß lagern)	3.3 M Kaliumacetat(pH
25 mM Tris-HCl pH 8.0	0,2 N NaOH	[60% (v/v) 5 M Kaliumacetat;
10 mM EDTA	1% (w/v) SDS	11.5% (v/v) Eisessig]

Die isolierten pBluescript SK-Plasmide wurden mit *EcoRI* restringiert, das Reaktionsgemisch auf einem 1% Agarosegel aufgetrennt und das Fragment aus dem Gel eluiert. Klonierte PCR-Fragmente in pGEM-T konnten am effizientesten durch eine Vektor-PCR mit dem Oligonukleotidpaar SP6/T7 amplifiziert und ebenfalls über ein Agarosegel gereinigt werden. Um die Fragmente für eine Sequenzierungsreaktion weiterzuverwenden, bot sich in beiden Fällen eine Vektor-PCR an.

B.II.8. Polymerase-Kettenreaktion (PCR)

Mit Hilfe der PCR können beliebige DNA-Fragmente *in vitro* selektiv ca. 10^6 - 10^7 fach amplifiziert werden. Dabei synthetisiert eine DNA-Polymerase in Anwesenheit von kurzen Oligonukleotidprimern und freien Nukleotiden einen komplementären zweiten Strang zu einem einzelsträngigen Templat. Jeder Reaktionszyklus besteht aus drei Temperaturschritten: 1. Denaturierung des doppelsträngigen Templats; 2. Anlagerung des Oligonukleotids zu einem kurzen doppelsträngigen Bereich („annealing“); 3. Verlängerung des angelagerten Primers durch eine DNA-Polymerase („extension“). Die Produkte werden exponentiell amplifiziert, da mehrere Reaktionszyklen stattfinden, bei denen neue Produkte wieder als Template dienen können. Die Denaturierung erfolgt bei 94-96°C, weshalb eine hitzestabile DNA-Polymerase aus *Thermus aquaticus* (*Taq*-Polymerase) bzw. modifizierte Enzyme mit einem Optimum bei 72°C und einer kurzzeitigen Toleranz für Temperaturen von 96°C verwendet werden. Die Stringenz einer PCR-Reaktion kann effizient über die Mg^{2+} -Konzentration und die „annealing“-Temperatur reguliert werden. PCR-Amplifikationen werden für genomische Sequenzen und klonierte DNA-Fragmente aus Vektoren verwendet, ferner für die Markierung von Sonden (II.12.3.) und die S-SAP Markertechnik (II.9.).

B.II.8.1. Amplifikation genomischer Sequenzen

Die verwendeten Oligonukleotide waren komplementär zu spezifischen Zielsequenzen im Genom, bzw. degenerierte Positionen wiesen mehreren Bindungsmöglichkeiten auf (I.4.). Als Templat diente isolierte, genomische DNA und cDNA ohne weitere Reinigung. *Taq*-DNA-Polymerasen (Boehringer; Eurogentec) wurden mit den Pufferlösungen des Herstellers verwendet, die Nukleotide (MBI Fermentas) wurden zu je 500 μ M äquimolar gemischt. Die Reaktion erfolgte in den Thermocyclern „Gene Amp 2400, 9600, 9700“ (Perkin Elmer) und „Mastercycler“ (Eppendorf).

• Reaktionsbedingungen und -zyklen

Reaktion	Oligonukleotide	Reaktionsbedingungen
<i>Ty1-copia</i> -Sequenzen	TyA/ TyB	50ng DNA, 1x Puffer, 2.5mM Mg ²⁺ , je 150 μ M dNTPs, 1U <i>Taq</i> -Pol, je 100pmol/50 μ l Primer
<i>CaDis</i> -Amplifikation	repL/repR	100ng DNA, 1x Puffer, 2.5mM Mg ²⁺ , je 150 μ M dNTPs, 2U <i>Taq</i> -Pol, je 100-200pmol/50 μ l Primer
LINE-Sequenzen	Wri1/Wri2	50ng DNA, 1x Puffer, 2.5mM Mg ²⁺ , je 150 μ M dNTPs, 1U <i>Taq</i> -Pol, je 250pmol/50 μ l Primer
<i>En/Spm</i> -Sequenzen	transAF/transCR	100ng DNA, 1x Puffer, 3mM Mg ²⁺ , je 150 μ M dNTPs, 1.25U <i>Taq</i> -Pol, je 50pmol/50 μ l Primer
<i>CaRep</i> - interne Fragmente	DOM1/DOM2 DOM3/DOM4 DOM5/DOM6	125ng DNA, 1x Puffer, 2.5mM Mg ²⁺ , je 150 μ M dNTPs, 1.25U <i>Taq</i> -Pol, je 100pmol/50 μ l Primer

Reaktion	Reaktionszyklen
<i>Ty1-copia</i> -Sequenzen	1 min 94°C
<i>CaDis</i> -Amplifikation	1 min 94°C } 35x
	1 min 45°C }
	1 min 72°C }
	5 min 72°C }
LINE-Sequenzen	2 min 94°C
	1 min 94°C } 35x
	1 min 47°C }
	2 min 72°C }
	5 min 72°C }
<i>En/Spm</i> -Sequenzen	2 min 94°C
	1 min 94°C } 35x
	1 min 58°C }
	1 min 72°C }

	5 min 72°C
--	------------

Reaktion	Reaktionszyklen
CaRep- interne Fragmente	2 min 94°C
	30 sec 94°C } 35x
	1 min 50°C }
	4 min 72°C }
	5 min 72°C

Für die Amplifikation sehr langer genomischer Fragmente (>2 Kbp) wurde das „Expand High Fidelity PCR System“ (Boehringer) verwendet. Es arbeitet mit einem Gemisch aus thermostabiler *Taq*-DNA-Polymerase (KlenTaq, eine N-terminal deletierte Variante der *Taq*-DNA-Polymerase) und einer sog. „proofreading“ Polymerase (*Pfu* oder *Vent*), d.h. einer gentechnisch hergestellten, ursprünglich aus *Pyrococcus furiosus* bzw. *Thermococcus litoralis* stammenden DNA-Polymerase mit 3`-5`-Exonukleaseaktivität. Das Enzymgemisch erzeugt 5-15 mal weniger Fehler als eine *Taq*-DNA-Polymerase und amplifiziert Fragmente von 5 - 12 Kbp effizient, da inkorporierte fehlgepaarte Basen durch die 3`-5`-Exonukleaseaktivität entfernt werden (Barnes et al., 1994). Die verwendeten PCR-Parameter waren der Länge des zu erwartenden Amplicons angepaßt.

• Arbeitsschritte:

Gemäß der Anleitung des Herstellers werden zwei Gemische hergestellt (Mix 1, Mix 2). Pro PCR-Reaktion werden 25µl Mix 1 mit 25µl Mix 2 vereint. Die Endkonzentration von Mg²⁺ beträgt 1.5mM.

Mix 1 (50µl)

150 ng DNA
je 100 pmol Primer
je 200µM dNTPs

Mix 2 (25µl)

2.6 U Enzym Mischung
5µl 10x Puffer (incl. Mg²⁺)

Reaktion	Oligonukleotide	Reaktionszyklen
CaRep- LTR-Fragmente	LTRU3/LTRU5	2 min 94°C
	LTRU5/DOM5	15 sec 94°C } 10x
	LTRU3/DOM2	30 sec 50°C }
		8 min 68°C }
		15 sec 94°C }
		30 sec 50°C }
		8 min }
		+ 5sec pro Cyclus 68°C }
		7 min 72°C

B.II.8.2. Amplifikation klonierter Sequenzen

Um Anwesenheit und Größe des klonierten Fragments zu analysieren bzw. das Fragment aus dem Plasmid zu isolieren und als Sonde oder zum Sequenzieren weiterzuverwenden, wurde Plasmid-DNA als Templat verwendet. Die Reaktion konnte mit isolierter Plasmid-DNA (II.7.2.) oder auch direkt mit lysierten Bakterienzellen durchgeführt werden. Als Oligonukleotidprimer dienten Sequenzen, die komplementär zu den die MCS („multiple cloning site“) flankierenden Vektorsequenzen in pGEM-T und pBluescriptSK waren (I.4.) und/oder interne Sequenzen des Fragments.

• Reaktionsbedingungen und -zyklen

Zur direkten Amplifikation aus der Bakterienzelle wurde anstelle der Plasmid-DNA ein Zahnstocher mit Bakterienmaterial in 20µl H₂O bidest abgestreift, das Gemisch 5 min lang bei 96°C denaturiert, bevor die restlichen Komponenten zugegeben wurden.

Reaktion	Oligonukleotide	Reaktionszyklen	Reaktionsbedingungen
Vektor-PCR	M13 Revers/T7 SP6/T7 „interne Oligos“	2 min 96°C 20 sec 94°C 23 sec 50°C 90 sec 68°C 7 min 72°C	50ng DNA, 1x Puffer, 2.5mM Mg ²⁺ , je 150µM dNTPs, 1U <i>Taq</i> -Pol, je 100pmol/50µl Primer

B.II.8.3. Reinigung von PCR-Produkten

Zur Weiterverwendung der Amplifikate wurden diese entweder auf einem Agarosegel aufgetrennt, die Fragmente eluiert und gereinigt (II.4.3.), oder direkt mit Hilfe des „Qiaquick PCR Purification“ Kits (Qiagen) aufgearbeitet. Dabei wurde die DNA an eine Silicagelmembran adsorbiert, von Salzen, Enzymen und Primer-Dimeren gereinigt und selektiv eluiert.

• Arbeitsschritte

- die PCR-Reaktion wird mit 5Vol. Puffer PB gemischt, auf eine Qiaquick-Säule gegeben und 1 min bei 13 000 Upm zentrifugiert (Heraeus Sepatech, Minifuge RF; Rotor 2150).
- das Filtrat wird verworfen, 750 µl Puffer PE zugegeben, zentrifugiert und das Filtrat erneut verworfen.
- die Säule wird zentrifugiert, um restliches Filtrat zu entfernen, und in ein Eppendorfgefäß gesetzt.
- 12 µl 10 mM TRIS-HCl (pH 8.5) werden zugeben und 30 min bei RT inkubiert.
- die Säule wird 2 min zentrifugiert, das Eluat bei 4°C gelagert.

B.II.9. Sequenzierung von DNA-Fragmenten

Zur Sequenzierung wurde die enzymatische Methode nach Sanger (1977) verwendet, die auf Kettenabbruch der DNA-Polymerase durch den Einbau von Didesoxynukleotiden beruht. Durch Fluoreszenzmarkierung der Didesoxy-Nukleotid-Triphosphate können die entstehenden DNA-Fragmente über einen Laserstrahl zur Lichtemission angeregt und mit Photodetektoren lokalisiert werden. Mit dieser Technik können bis über 700 Basen nach der Primerbindestelle zuverlässig sequenziert werden. Als Templat wurde sowohl isolierte Plasmid-DNA verwendet, deren Salzkonzentration zuvor durch Dialyse verringert wurde, als auch gereinigte PCR-Amplifikate klonierter Fragmente. Die Sequenzierreaktionen wurden größtenteils mit dem „ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction“ Kit (Perkin Elmer) selber durchgeführt und die gereinigten Produkte anschließend durch die Firma SRD (Oberursel) computergestützt auf einer Elektrophoreseapparatur ausgewertet. Alternativ wurden einige Sequenzierungen insgesamt von der Firma MWG-Biotech (Ebersberg) durchgeführt. Die als Textdatei und Farbdigramm vorliegenden Sequenziererergebnisse wurden in beiden Fällen selber mit der in I.5.1. aufgelisteten Software ausgewertet. Alle in dieser Arbeit gewonnenen Sequenzen sind in der Datenbank des „European Molecular Biological Laboratory“ (EMBL) unter den angegebenen Akzessionsnummern (AJ...) zugänglich.

B.II.9.1. Sequenzierreaktion und Reinigung

Als Sequenzier-Primer wurden je nach Vektor die universellen Vektorsequenzen T7, SP6 und M13 Revers verwendet. War das klonierte Fragment trotz einer bidirektionellen Sequenzierreaktion nicht vollständig sequenziert, wurde ein Oligonukleotid aus dem Endbereich der lesbaren Sequenz abgeleitet und für eine weitere Reaktion mit gleichem Templat verwendet. Die überlappenden Sequenzen wurden bei der Computerauswertung zusammengefügt.

• Arbeitsschritte

Dialyse der Plasmid-DNA

- 1µg isolierte Plasmid-DNA wird auf ein Dialyseplättchen (VSWP01300; Millipore) pipettiert und 1h gegen 0.1 mM Tris-HCl (pH 8,0) dialysiert.
- die Probe wird abpipettiert und das Volumen bestimmt.

Sequenzierreaktion

- ein 10µl Reaktionsansatz mit 250 ng gereinigter Plasmid-DNA oder 130-170 ng gereinigten PCR-Amplifikaten, 4 µl Terminator Reaktions-Mix, 5 pmol

des Primers und H₂O bidest wird gemischt und in einem Thermocycler bei folgenden Reaktionszyklen inkubiert:

30 sec 96°C		
10 sec 96°C	}	
5 sec 50°C	}	25x
4 min 60°C	}	

- Proben werden bei -20°C gelagert oder sofort gereinigt.

Reinigung

- der Reaktionsansatz wird mit 10µl H₂O bidest und 2µl 3M NaAcetat (pH 4.6) gemischt, mit 62µl Ethanol versetzt und 20 min bei -80°C inkubiert.
- das Gemisch wird 15 min lang bei 13 000 Upm/4°C zentrifugiert (Heraeus Sepatech, Minifuge RF; Rotor 2150), der Überstand entfernt und das Sediment mit 500µl 70% Ethanol überschichtet.
- der Ansatz wird 10 min bei 13 000 Upm/ 4°C zentrifugiert, das Sediment 10min bei RT unter Vakuum getrocknet.

B.II.10. Generation von „sequence-specific amplified polymorphisms“ (S-SAP) mit PCR

S-SAP ist eine auf PCR basierende, mit AFLP („amplified fragment length polymorphism“; Vos et al., 1995) verwandte Markertechnik, die Polymorphismen in der Nachbarschaft zu bekannten Sequenzen detektiert und modifizierte Restriktionsfragmente als Templat verwendet (Waugh et al., 1997). Für die hier verwendete, modifizierte Form des S-SAP (Ellis et al., 1998) wurde genomische DNA restringiert, an einen Adaptor ligiert und mit einem zum Adaptor komplementären Kassettenprimer und einem spezifischen Oligonukleotid amplifiziert. Letzteres wurde zuvor radioaktiv markiert, das Reaktionsgemisch auf einem denaturierenden Polyacrylamidgel aufgetrennt, und markierte Fragmente durch Autoradiographie sichtbar gemacht.

B.II.10.1. Herstellung der Adaptorkassette, Restriktion der DNA, Ligation und Präamplifikation

Die Adaptorkassette wurde aus zwei Oligonukleotiden hergestellt, die sich während eines Renaturierungsschrittes so aneinanderlagerten, daß die überhängenden Enden jeweils komplementär zu denen bestimmter Restriktionsfragmente waren. Für die Ligation konnten die Produkte der Restriktion ohne Reinigung eingesetzt werden, da ligierte Adaptor-Fragment-Moleküle die entsprechende Erkennungssequenz der Endonuklease nicht mehr aufwiesen und beide Enzyme bei gleichen Salzkonzentrationen arbeiteten. Die Templat-DNA wurde in einem Präamplifikationsschritt angereichert. Die Sequenzen von Kassetten-, Null- und spezifischen Oligonukleotiden sind in I.4. aufgeführt (bzw. Hüttel, pers.Mitteilung).

• Arbeitsschritte

- 250 ng DNA werden mit 3U Restriktionsendonuklease und 0.2 Vol. 5x RL Puffer in einem Volumen von 20µl gemischt und 3 h bei 37°C inkubiert.
- für 60µl Kassettengemisch werden 15-17µg der beiden Oligonukleotide mit 0.1Vol. 5x RL Puffer und H₂O bidest gemischt, auf 95°C erhitzt und innerhalb von 10 min auf 20°C abgekühlt.
- ein Ligationsgemisch aus 0.5µl Kassettengemisch, 1U T4 DNA-Ligase (MBI Fermentas), 0.1Vol. 10mM ATP, 0.2Vol. 5x RL Puffer und H₂O bidest wird in einem Volumen von 5µl hergestellt.
- der gesamte Ligationsgemisch wird mit dem Restriktionsansatz gemischt (gesamt: 25µl) und mind. 3 h oder über Nacht bei 37°C inkubiert; die Reaktionsprodukte sind bei -20°C lagerbar.
- 5µl der Ligationsreaktion werden mit 20µl H₂O bidest verdünnt, 5µl dieser Verdünnung dienen als Templat für die Präamplifikation (20µl) .

Reaktionsbedingungen der Präamplifikation (20µl)

Oligonukleotide	Reaktionszyklen	Reaktionsbedingungen
<i>Eco</i> RI-Nulloligo	1 min 94°C 30 sec 94°C 1 min 56°C 1 min 72°C 2 min 72°C	5µl Templat, 1x Puffer, 1.5mM Mg ²⁺ , je 200µM dNTPs, 0.5 U <i>Taq</i> -Pol. (Gibco BRL), 5pmol/20µl Primer

- die Produkte der Präamplifikation sind bei -20°C lagerbar.

• Lösungen

5x RL Puffer

- 50mM TRIS-Acetat (pH7.5)
- 50mM MgAcetat
- 250mM KAcetat
- 25mM DTT
- 25ng/µl BSA (Rinderserumalbumin)

B.II.10.2. Markierung der Oligonukleotidprimer

Das spezifische Oligonukleotid wurde mit endständigem γ -P³²dATP markiert. Die Markierungsreaktion wurde durch Polynukleotidkinase, eine T4-Kinase (Boehringer), katalysiert, die das radioaktive γ -Phosphat am 5'-Ende des Primers kovalent bindet.

- ein 20µl Gemisch aus 200 pmol Oligonukleotid, 2 µl γ -P³²dATP (3000 Ci/mmol), 0.1Vol. 10x Kinase-Puffer, 4U T4-Kinase und H₂O bidest wird hergestellt und 1h bei 37°C inkubiert.
- die enzymatischen Komponenten werden 10 min lang bei 70°C inaktiviert; die Produkte der Markierungsreaktion sind begrenzt bei -20°C lagerbar.

B.II.10.3. S-SAP-Amplifikation

Für die S-SAP-Reaktion wurden das markierte spezifische Oligonukleotid und ein zur Adaptorkassette komplementärer Nullprimer verwendet. Sichtbar waren daher nur Amplifikate, die das spezifische Oligonukleotid enthielten.

- 5µl der Präamplifikationsreaktion werden mit 45µl H₂O verdünnt, 2µl dieser Verdünnung dienen als Templat für die S-SAP Amplifikation (10µl).

Reaktionsbedingungen der S-SAP-Amplifikation

Oligonukleotid	Reaktionszyklen	Reaktionsbedingungen
<i>Eco</i> RI-Nulloligo-nukleotid	3 min 94°C	2µl Templat, 1x Puffer, 2mM Mg ²⁺ , je 200µM dNTPs, 0.5 U <i>Taq</i> -Pol. (Gibco BRL), 5pmol/10µl Nullprimer 2pmol/10µl spez.Primer
	30sec 94°C	
	30 sec 65°C	
	-0.7°C pro Zyklus	
	1 min 72°C	
	30 sec 94°C	
	30 sec 56°C	
	1 min 72°C	
	2 min 72°C	
	25x	

- die S-SAP Produkte sind begrenzt bei -20°C lagerbar; 1µl der Reaktion wird gelelektrophoretisch aufgetrennt und die Amplifikate durch Autoradiographie detektiert.

B.II.11. Southern-Hybridisierung

Bei partieller oder vollständiger Übereinstimmung der Nukleotidsequenz können zwei einzelsträngige DNA-Moleküle zu einem Doppelstrang reassoziieren (Hybridisierung). Durch Hybridisierungsexperimente mit einzelsträngiger DNA, die auf einer Nylonmembran fixiert ist (Southern-Hybridisierung), kann daher die Lokalisation von DNA-Moleküle mit ähnlicher Sequenz detektiert werden. Ausschlaggebend für die Stringenz einer Hybridisierung sind Temperatur und Salzkonzentration (Meinkoth & Wahl, 1984). Die verwendeten Sonden wurden radioaktiv markiert und durch Autoradiographie sichtbar gemacht.

B.II.11.1. Radioaktive Markierung von DNA-Fragmenten

Zur Markierung *Eco*RI-restringierter, genomischer DNA-Fragmente bzw. aus Bakterienplasmiden isolierter Fragmente wurde die "random primed"-Methode (Feinberg & Vogelstein, 1983) in Form des "Prime-it II Random Primed DNA Labeling" Kits (Stratagene) verwendet. Das Klenow-Fragment der DNA-Polymerase I synthetisiert hier ausgehend von Hexanukleotiden einen zweiten Strang zum einzelsträngig vorliegenden Fragment, wobei α-³²P-dCTP und / oder α-³²P-dATP inkorporiert werden. Nichteingebaute, radioaktive

Nukleotide wurden durch Ausschlußchromatographie auf einer Sephadex G50-Säule von den markierten Fragmenten abgetrennt.

• Arbeitsschritte

- 50-60 ng gereinigte Fragment-DNA werden mit 10 µl der Hexanukleotidmischung und Wasser auf 34 µl Volumen gebracht, 5 min in einem siedendem Wasserbad und anschließend 10 min auf Eis inkubiert.
- 10 µl 5x Primer-Puffer (0.1 mM äquimolares Gemisch der drei übrigen Nukleotide) und 5µl des radioaktiven Nukleotids (3000 Ci/mmol; Amersham) sowie 5U Klenow-Fragment werden zugegeben, der Reaktionsansatz gemischt und 30 min bei 37-40°C inkubiert.
- die Reaktion wird durch Zugabe von 2µl 0.5 M EDTA (Ethylendiamintetraacetat; pH 8.0) beendet.
- das Volumen wird auf 100 µl gebracht, auf eine mit TEN äquilibrierte Sephadex G50-Säule pipettiert und mit TE-Puffer nachgespült.
- die erste radioaktive Fraktion enthält die markierten Fragmente und wird aufgefangen; sie sollte am selben Tag zur Hybridisierung verwendet werden.

• Lösungen

TE		TEN	
10mM TRIS-HCl (pH 8.0)		10 mM TRIS-HCl (pH 7.5)	
1 mM EDTA		1 mM EDTA	
		50 mM NaCl	

B.II.11.2. Hybridisierung von Southern-, „Blots“ und Koloniefiltern

Um die unspezifische Bindung von Sonden-DNA an die Oberfläche der Nylonmembran zu verhindern, wurden Southern-, „Blots“ und Koloniefilter mit denaturierter und gescherter DNA aus Heringssperma oder Kalbsthymus prähybridisiert. Nach dem Hybridisierungsschritt folgten Waschschrte, bis unspezifische Hintergrundsignale nicht mehr sichtbar waren. Hybridisierungs- und Waschbedingungen für die einzelnen Sonden sind gesondert aufgeführt. Die Exposition eines Röntgenfilms vor einer Lanthanoid/Actinoid-Verstärkerfolie, die bei -80°C verlangsamte Elektronen reflektiert, machte die radioaktiven Signale sichtbar (Autoradiographie).

• Arbeitsschritte

- Nylonmembranen werden in Hybridisierungsröhren oder Plastikwannen mit Prähybridisierungspuffer überschichtet und über Nacht (mind. 4h) bei 60-65°C geschüttelt.

- der Prähybridisierungspuffer wird abgeschüttet, die Membranen werden mit vorgewärmtem Hybridisierungspuffer überschichtet.
- die markierten Sondenfragmente (50ng/1600cm² Filteroberfläche) werden 10 min in einem siedenden Wasserbad denaturiert, kurz auf Eis gelagert und in Hybridisierungspuffer pipettiert.
- die Membranen werden unter Schütteln über Nacht bei 60-65°C inkubiert, dann der Hybridisierungspuffer entfernt und das Hybridisierungsgefäß mit etwas Waschlösung gespült.
- die Membranen werden erst 2x 2 min bei 60°C, dann 1-2x 10 min bei 60-65°C mit Waschlösung geschwenkt.
- die Filter werden luftgetrocknet, mit Cellophan bedeckt und in einer lichtundurchlässigen Kassette (mit Verstärkerfolie) bei -80°C mit einem Röntgenfilm belegt.

• Lösungen

100x Denhardts Lösung

2% (w/v) PVP 40 (Polyvinylpyrrolidon 40)
2% (w/v) BSA
2% (w/v) Ficoll

Prähybridisierungspuffer

6x SSC
5x Denhardts Lösung
200 µg/ml „Blocking“-DNA

Hybridisierungspuffer

6x SSC
5x Denhardts Lösung
40 µg/ml „Blocking“-DNA

Waschlösung

1-3x SSC
1% SDS

Sequenzfamilie	Hybridisierungstemperatur	Waschparameter
<i>CaSat1</i>	65°C	2x 10 min, 65°C, 0.1x SSC/1%SDS
<i>CaSat2</i>	65°C	2x 10 min, 65°C, 0.1x SSC/1%SDS
<i>CaRep1-3</i>	65°C	2x 10 min, 65°C, 0.1x SSC/1%SDS
<i>CaTy</i>	65°C	2x10 min, 65°C, 0.1x SSC/1%SDS bzw. 3x SSC/1%SDS
<i>CaDis</i>	65°C	2x10 min, 65°C, 0.1x SSC/1%SDS bzw. 1x SSC/1%SDS
<i>CaLin</i>	65°C bzw. 60°C	1x 10 min, 65°C bzw. 1x 10min, 55-60°C 3x SSC/1%SDS
<i>CaEn/Spm</i>	60°C	1x 10 min, 60°C, 1x SSC/1%SDS

B.II.12. *in-situ*-Hybridisierung

Die *in-situ*-Hybridisierung nutzt DNA aus Gewebeschnitten, Gewebekulturzellen und isolierten Zellkernen als Templat und ergänzt damit Ergebnisse aus Hybridisierungen auf filtergebundene DNA. Die DNA im Zellkern ist durch assoziierte Proteinstrukturen sterisch nicht so leicht für Sondenmoleküle zugänglich wie gereinigte, filtergebundene Template, was beim Vergleich der Hybridisierungsergebnisse beachtet werden muß. Alle Hybridisierungen

erfolgten auf Interphase- und Metaphasechromosomen aus *C. arietinum*. Letztere eignen sich besonders gut zur Lokalisation repetitiver Sequenzen, da die DNA-Fasern im Vergleich zum Interphasechromatin 5-10 fach stärker kondensiert sind, so daß die Umrisse der Chromosomen erkennbar werden.

B.II.12.1. Fixierung von Wurzelmaterial

Metaphasekerne sind besonders in teilungsaktiven Meristemzellen vorhanden, die sich in Wurzelspitzen oder an der Blattbasis befinden. Daher wurden die Spitzen der Primärwurzeln verwendet. Durch 8-Hydroxychinolin, das die Ausbildung der Spindelfasern hemmt, wurden die Zellzyklen innerhalb des Gewebes synchronisiert und Metaphasekerne angereichert. Das synchronisierte Gewebe wurde nachfolgend fixiert und zur Herstellung von Kernpräparaten verwendet.

• Arbeitsschritte

- 1-2 cm lange frische Primärwurzeln von *C. arietinum* werden 2-3 h bei RT in 2 mM 8-Hydroxychinolin inkubiert.
- das Hydroxychinolin wird entfernt, die Wurzeln mit Fixierlösung versetzt und 2h bei RT und danach 24h bei 4°C inkubiert; nach Wechsel der Fixierlösung sind die Wurzeln bei -20°C lagerbar.

• Lösungen

Fixier-Lösung
(frisch ansetzen)
3Vol. Methanol
1Vol. Eisessig

B.II.12.2. Herstellung der Zellkernpräparate

Zellwände und weitere Zellbestandteile wurden enzymatisch aufgeschlossen. Das Aufspreiten der Zellkerne und die Trennung von den Zellresten erfolgte durch Auftropfen auf einen Objektträger. Zur Präparation der Zellkerne wurde ein Protokoll von Schwarzacher & Leitch (1994) verwendet.

• Arbeitsschritte

- die für die Präparate vorgesehenen Objektträger werden mindestens 2h in konz. Chromschwefelsäure inkubiert, 20 min unter fließendem Wasser und 1 min mit H₂O bidest gespült, über Nacht bei 37°C getrocknet und vor Verwendung in 70% Ethanol getaucht.
- die fixierten Wurzeln werden aus der Fixierlösung genommen, abgetropft und 10 min bei RT in 1x Enzympuffer äquilibriert.

- die Wurzeln werden abgetupft und erneut 15 min in 1x Enzympuffer bei RT geschwenkt.
- die hellen meristematischen Wurzelspitzen werden von der Primärwurzel abgetrennt und 30 min bei 37°C in einem Gemisch aus 1Vol. Pektinase/Cellulase-Mischung mit 1Vol. 1x Enzympuffer inkubiert.
- die Wurzelspitzen werden leicht zerdrückt und 10-30 min bei 37°C inkubiert.
- die Suspension wird durch vorsichtiges Auf- und Absaugen mit einer Pipette gemischt, 4min bei 6000 Upm abzentrifugiert (Heraeus Sepatech, Minifuge RF; Rotor 2150) und der Überstand bis auf 100µl abgenommen.
- die verbliebenen 100µl werden mit 1ml 1x Enzympuffer vermischt, erneut zentrifugiert und der Überstand durch Abpipettieren auf 100µl reduziert; dieser Arbeitsschritt wird wiederholt.
- 1ml Fixierlösung wird zugegeben, das Gemisch abzentrifugiert, der Überstand abgenommen, bis der obere Teil des Pellets sich ablöst; durch weiteres Zentrifugieren und Abpipettieren wird die Suspension auf 200µl reduziert.
- die Suspension wird gemischt, kurz stehengelassen und in Portionen von je 12µl aus 50 cm Höhe auf einen Chromschwefelsäure und Ethanol gereinigten Objektträger aufgetropft.
- die Präparate werden unter einem Phasenkontrastmikroskop kontrolliert und über Nacht bei 37°C inkubiert; die Objektträger sind bei RT trocken lagerbar.

• Lösungen

10x Enzympuffer	Pektinase/Cellulase-Mischung
0.04 M Zitronensäure	1.8% (w/v) Cellulase aus <i>Aspergillus niger</i> (Sigma)
0.06 M NaCitrat	0.2% (w/v) Cellulase (Onozuka R10) (Serva)
pH 4.8	20 % (v/v) Pektinase aus <i>Aspergillus niger</i> (Sigma)

B.II.12.3. Fluoreszenzmarkierung der Sondenfragmente

Die nicht-radioaktive Markierung der Sonden erfolgte durch PCR, wobei digoxigenierte- und biotinylierte Nukleotide eingebaut wurden. Für die Reaktion wurden klonierte Fragmente bzw. genomische DNA als Templat und die jeweils flankierenden Oligonukleotide (I.4.) verwendet. Als Sonden für die rRNA-Gene wurden die klonierten Fragmente aus p18S (18S-5.8S-25S rDNA) und pXV1 (5S rDNA und angrenzende IGS) als Templat verwendet (Schmidt et al., 1994).

• Arbeitsschritte

Reaktionsbedingungen

- die Reaktionen werden mit den üblichen Primer- und Templatkonzentrationen (II.8.1.; II.8.2.) in einem Volumen von 25µl

angesetzt. Zusätzlich zum dNTP-Mix (200 μ M Endkonzentration) wird jeweils 70 μ M Biotin-16-dUTP (Boehringer) und/oder 36 μ M Digoxigenin-11-dUTP (Boehringer) zugegeben.

- Reaktionszyklen siehe II.8.1. und II.8.2.
- der Erfolg der Markierungsreaktion kann durch Agarosegelelektrophorese mit einem unmarkierten Amplifikat kontrolliert werden, das eine höhere Mobilität als die markierten Fragmente hat.
- die Reaktion wird mit 0.1 Vol. 3M NaAcetat (pH 4.6) und 2.5 Vol. eiskaltem Ethanol gemischt, 1-2h bei -70°C inkubiert und 30 min bei 13000 rpm/RT abzentrifugiert
- der Überstand wird entfernt, das Sediment luftgetrocknet und in 10 μ l H₂O bidest aufgenommen; die markierte Sonde kann bei -20°C gelagert werden.

B.II.12.4. Vorbehandlung, Hybridisierung und Waschen

Vor der Hybridisierung wurden die Kern-RNA und das Cytoplasma enzymatisch abgebaut und die Chromatinstruktur aufgelockert, so daß die DNA für Sonde und Nachweisreagenzien zugänglich war (Permeabilisierung). Nachfolgend wurde das Präparat durch Paraformaldehyd fixiert (Verringerung der Diffusion und des Verlusts der DNA während der Hybridisierung) und durch eine Ethanolreihe dehydratisiert. Nach der Hybridisierung wurden unspezifisch gebundene Sondenmoleküle durch Waschen entfernt. Die *in-situ*-Hybridisierung folgte einem modifizierten Protokoll von Schmidt et al. (1994). Die Stringenz der Hybridisierung kann über die Formamid- und Salzkonzentration sowie die Temperatur der Hybridisierungs- und Waschschrte kontrolliert werden (Meinkoth & Wahl, 1984) und ist für die einzelnen Sonden aufgeführt. Für alle Inkubationsschritte mit großen Mengen Lösung wurden die Objektträger in einem Coplin-Gefäß fixiert, geringe Mengen Reagenz wurden direkt aufgetragen und das Präparat durch eine Plastikfolie vor Austrocknung geschützt.

• Arbeitsschritte

Vorbehandlung der Chromosomen

- das RNase-Konzentrat (10 μ g/ μ l) wird mit 2xSSC 1:100 verdünnt, 200 μ l der Lösung werden auf den Objektträger pipettiert, mit einem hitzebeständigen Plastikstreifen abgedeckt und 1h bei 37°C in einer Feuchten Kammer (wasserdampfgesättigt) inkubiert.
- der Plastikstreifen wird entfernt und der Objektträger 3x 5 min bei RT in 2x SSC geschwenkt
- der Objektträger wird in 0.01N HCl überführt und 1 min bei RT inkubiert.
- das Pepsin-Konzentrat (0.5mg/ml) wird 1:50 in 0.01 N HCl verdünnt, 200 μ l der Lösung werden auf den Objektträger pipettiert, mit einem Plastikstreifen abgedeckt und 15 min bei 37°C in der Feuchten Kammer inkubiert.

- der Plastikstreifen wird entfernt und der Objektträger 3x 5 min bei RT in 2x SSC geschwenkt.
- eine klare 4% (w/v) Paraformaldehydlösung in H₂O bidest wird durch Erhitzen auf 60°C unter Rühren und Zugabe einiger Tropfen NaOH konz. hergestellt.
- die Objektträger werden in der auf RT abgekühlten Paraformaldehydlösung 10 min bei RT und dann 2-3x 10 min in 2x SSC geschwenkt.
- die Objektträger werden je 3 min lang in 70%, 90% und 100% Ethanol geschwenkt und anschließend luftgetrocknet.

Hybridisierung

- pro Objektträger werden 30 µl eines Hybridisierungsgemisches aus 0.5-2µl der gefällten Sonde(n), 10.5-15µl 100% Formamid (je nach Stringenz), 6µl 50% Dextransulfat, 3µl 20x SSC, 0.5µl 10% SDS (Natrium/Sodiumdodecylsulfat), 2 µl (10µg/µl) Heringssperma-DNA oder *E.coli*-DNA und H₂O bidest hergestellt.
- das Gemisch wird 10 min auf 70°C erhitzt, dann 5 min auf Eis gekühlt, dann auf den Objektträger pipettiert und mit einem Plastikstreifen blasenfrei abgedeckt.
- beides wird in einem Thermocycler (Hybaid) denaturiert und in mehreren Schritten auf 37°C abgekühlt:
 - 8 min bei 70°C
 - 5 min bei 55°C
 - 2 min bei 50°C
 - 3 min bei 45°C
 - ∞ bei 37°C

und über Nacht in der Feuchten Kammer bei 37°C inkubiert.

Waschen

- die Plastikstreifen werden unter 2xSSC entfernt, die Objektträger 2x 5 min bei 42°C mit vorgewärmter Waschlösung (15-20% (v/v) Formamid in 0.2-0.75x SSC- je nach Stringenz) geschwenkt.
- die Objektträger werden 2x 5 min bei 42°C mit vorgewärmtem 2x SSC und 1x 5 min in SSC bei RT geschwenkt; anschließend wird sofort mit der Signaldetektion fortgefahren.

Sequenzfamilie	Stringenz der Hybridisierung (Meinkoth & Wahl, 1984)	Stringenz der Waschparameter (Meinkoth & Wahl, 1984)
<i>CaSat1</i>	76% (37°C/50% Formamid/2xSSC)	79% (42°C/20% Formamid/0.2xSSC)
<i>CaSat2</i>	76% (37°C/50% Formamid/2xSSC)	79% (42°C/20% Formamid/0.2xSSC)
<i>CaRep1-3</i>	76% (37°C/50% Formamid/2xSSC)	79% (42°C/20% Formamid/0.2xSSC)
<i>CaTy</i>	70-76% (37°C/50-40%Formamid/2xSSC)	72-79% (42°C/20% Formamid/0.2-0.5xSSC)

<i>CaLin</i>	67-76% (37°C/50-35% Formamid/2xSSC)	70-79% (42°C/20% Formamid/0.2-0.75xSSC)
<i>CaDis</i>	76% (37°C/50% Formamid/2xSSC)	79% (42°C/20% Formamid/0.2xSSC)
<i>CaEn/Spm</i>	76% (37°C/50% Formamid/2xSSC)	79% (42°C/20% Formamid/0.2xSSC)

B.II.12.5. Signaldetektion durch Antikörperbehandlung

Die markierten Sondenmoleküle wurden durch biotinbindendes Streptavidin bzw. Digoxigeninantikörper, die mit Fluorochromen konjugiert sind, detektiert. Ein weiterer Waschschritt entfernte überschüssige Proteine. DAPI, ein DNA-bindendes Fluorochrom, färbt bevorzugt AT-reiches Heterochromatin und fluoresziert unter UV-Anregung blau, so daß zumindest Teile der Chromosomen sichtbar gemacht werden konnten.

• Arbeitsschritte

- die Objektträger werden 5 min mit 4x SSC/ 0.2% Tween20 äquilibriert, mit 200 µl 5% BSA (Rinderserumalbumin) in 4x SSC/0.2% Tween20 versetzt, mit einem Plastikstreifen bedeckt und 10-30 min bei 37°C in einer Feuchten Kammer inkubiert.
- nach Ablauf der überschüssigen „Blocking“-Lösung werden die Objektträger mit je 50µl einer 1:100 Verdünnung von Antikörper-Lösung A in 3% BSA/4xSSC/0.2% Tween20 oder einer 1:200 Verdünnung von Antikörper-Lösung B in 3% BSA/4xSSC/0.2% Tween20 bzw einem Gemisch beider Komponenten versetzt, mit einem Plastikstreifen bedeckt und 1h bei 37°C in der Feuchten Kammer inkubiert.
- der Plastikstreifen wird entfernt und die Objektträger 3x 10 min in 4xSSC/ 0.2% Tween20 bei 37°C/ 42°C geschwenkt.
- nach Ablauf des restlichen 4xSSC/ 0.2% Tween20 werden 100µl DAPI Lösung (2-4µg/ml) aufpipettiert, mit einem Plastikstreifen abgedeckt und 10-15 min bei RT inkubiert.
- die Objektträger werden kurz in 4xSSC/ 0.2% Tween20 eingetaucht, ablaufen lassen, mit 1 Tropfen Citiflour AF1/Glycerin (UKC, Canterbury; zur Stabilisierung der Farbstoffe) versetzt und mit einem 24x32mm Glasplättchen bedeckt.
- nach Ausquetschen der überschüssigen Stabilisierungsflüssigkeit werden die Präparate über Nacht bei 4°C, danach trocken und kühl gelagert.

• Lösungen

Antikörper-Lösung A

200µg/ml Fluorescein-5-isothiocyanat (FITC) an Anti-Digoxigenin AK (Boehringer)

Antikörper-Lösung B

1mg/ml Cy3 an Streptavidin (Sigma) in 4xSSC/ 0.2% Tween20

in 4xSSC/ 0.2% Tween20
(gelb-grüne Fluoreszenz)

(rote Fluoreszenz)

B.II.12.6. Dokumentation (Mikroskopie/ Photographie)

Die Auswertung und Photographie der Hybridisierungsergebnisse erfolgte an einem UV-Fluoreszenzmikroskop (Zeiss) mit folgenden Filtern: Set A (DAPI/ 340-380 nm Anregung, Emission >420 nm), Set I2/3 (FITC/ 450-490 nm Anregung, >520 nm Emission), Set N2 (Cy3, TRITC, Texasred/ 515-560 nm Anregung, >590 nm Emission), Multipass-Filter mit drei Anregungs- und Emissionsbereichen für DAPI, FITC (Fluorescein-Isothiocyanat) und Cy3. Die Signale wurden durch Photographie auf Farbfilme (Fujicolor, 400ASA) dokumentiert, die Negative durch Scannen digitalisiert und die Bilddateien mit Hilfe der Software „Adobe Photoshop“ (I.5.3.) prozessiert und ausgedruckt.

C. ERGEBNISSE

C.I. Hochrepetitive Sequenzen

Zur Isolation hochrepetitiver DNA-Fragmente aus dem Genom der Kichererbse wurde eine Plasmidbank durch Klonierung *EcoRI*-restringierter, genomischer DNA von *Cicer arietinum* L. hergestellt und mit gesamtgenomischer DNA desselben Genoms unter stringenten Bedingungen hybridisiert. Klониerte Sequenzen, die sehr hochabundant im Kichererbsengenom vorliegen, erzeugen hierbei die stärksten Hybridisierungssignale. Die Klone mit den intensivsten Hybridisierungssignalen wurden durch weitere Hybridisierungen untereinander in Homologieklassen eingeteilt und getrennt weiteruntersucht. Die fünf Homologiegruppen (Tab.C1) mit den prominentesten Signalen werden nachfolgend im einzelnen beschrieben (C.I.1., C.I.2.).

Tab.C1: Hochabundante Sequenzen aus dem Genom von *C. arietinum*. Einteilung genomischer Klone, die mit einer genomischen Sonde die prominentesten Hybridisierungssignale erzeugten, in fünf Homologiegruppen. Die spätere Zuordnung zu repetitiven Familien ist zur besseren Übersicht angemerkt.

Homologiegruppe	Anteil an den als hochrepetitiv detektierten Klonen der genomischen Bank	Identifizierte Klone, die für weitere Untersuchungen als Sonden verwendet und/ oder sequenziert wurden: pCaEr...	Familie
1	27%	250, 292, 362, 1062, 845, 427, 782, 755, 511, 447, 636	<i>CaSat1</i>
2	4%	453, 806, 855	<i>CaSat2</i>
3	1%	1048	<i>CaRep1</i>
4	5%	638, 757, 915, 936, 1034, 467	<i>CaRep2</i>
5	5%	432, 776, 1049	<i>CaRep3</i>

C.I.1. Die tandemrepetitiven Elemente *CaSat1* und *CaSat2*

C.I.1.1. Genomische Organisation

Zur molekularen Charakterisierung der isolierten Sequenzfamilien sollte die Organisation der Wiederholungseinheiten im Genom von *C. arietinum* untersucht werden. Dies ist mit Hilfe der Southern-Hybridisierung möglich, bei der einzelne Klone der Sequenzfamilie als Sonden auf restringierter, genomischer DNA eingesetzt werden. Je nachdem, ob die Wiederholungseinheiten dispers im Genom verstreut oder in lückenlosen Anhäufungen hintereinander angeordnet sind bzw. ob konservierte Restriktionsschnittstellen in der Sequenz oder der genomischen Nachbarschaft existieren, sind unterschiedliche Bandenmuster oder schmierförmige Signale der homologen Restriktionsfragmente zu erwarten. Die Sequenzierung einzelner Klone aus der jeweiligen Homologiegruppe ermöglichte weitere Einblicke in die Primärstruktur der Sequenzwiederholungen, deren Homologie untereinander und zu bereits charakterisierten Elementen sowie die Identität benachbarter Genombereiche. Aus der Anzahl und der Intensität der Hybridisierungssignale

läßt sich bei einer Southern-Analyse gleichzeitig eine Aussage über die ungefähre Abundanz der homologen Sequenzen im Genom machen.

Zwei der hochabundanten Sequenzklassen weisen ein für Satelliten-DNA charakteristisches Leitermuster auf

Die genomische Organisation von zwei der fünf hochabundanten Sequenzfamilien des Kichererbsengenoms wurde zunächst durch Southern-Hybridisierung auf unterschiedlich restringierte, genomische DNA aus *Cicer arietinum* untersucht. Es wurden 18 verschiedene Restriktionsenzyme verwendet, als Sonde diente das Fragment eines Klones aus den Homologiegruppen 1 und 2 der genomischen Bank (Tab.C1).

Die Southern-Hybridisierung mit den radioaktiv markierten Fragmenten der Klone pCaEr292 und pCaEr362 aus der Homologiegruppe 1 zeigte jeweils sehr intensive Hybridisierungssignale, die bei 14 Restriktionen in Form eines durchgehenden Schmiers, bei

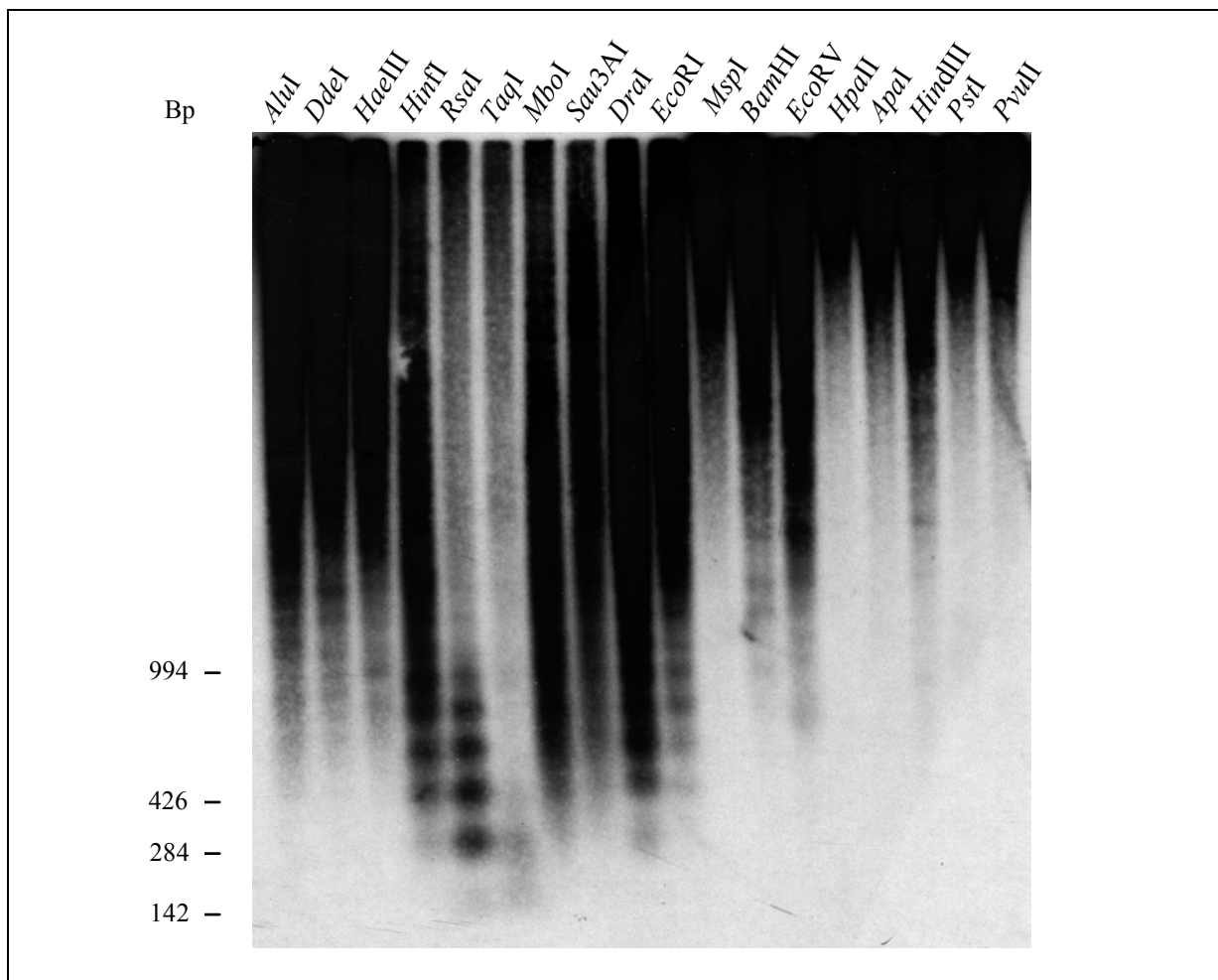


Fig.C1: Genomische Organisation des Satelliten *CaSat1*. Autoradiogramm eines Southern-Blots mit genomischer DNA von *C. arietinum*, die mit verschiedenen Restriktionsendonukleasen behandelt und gegen das Fragment von pCaEr362 (*CaSat1*) hybridisiert wurde. Bei *HinfI*, *DraI*, *EcoRI*, *RsaI* ist das charakteristische Leitermuster erkennbar.

HinfI, *DraI*, *EcoRI*, *RsaI* jedoch als leiterförmig angeordnete, distinkte Banden erschienen (Fig.C1). Die kleinste Bande war bei ca. 170 Bp (*RsaI*) sichtbar, weitere Banden im höheren Molekulargewichtsbereich folgten im Abstand von ca. 170 Bp. Dieses sogenannte Leiterbandenmuster resultiert aus konservierten Restriktionsschnittstellen in tandemartig angeordneten, entweder identischen oder ähnlichen Wiederholungseinheiten eines repetitiven Elementes und ist charakteristisch für Satelliten-DNA.

Die kleinste sichtbare *RsaI*-Bande und der Abstand der Multimerbanden (*DraI*, *EcoRI*, *HinfI*) bestätigen eine Monomerlänge von ca. 170 Bp in dieser Homologiegruppe (Fig.C1). Die Sequenzen der ersten Gruppe gehören demnach zu einer Satelliten-DNA-Familie und werden künftig als *CaSat1* (*Cicer arietinum* **Satellit 1**) bezeichnet.

Die *EcoRI*- und *DraI*-Leiterbanden wiesen zwar den gleichen Abstand wie die *RsaI*- und *HinfI*- Muster auf, waren jedoch in den höheren Molekulargewichtsbereich verschoben. Funktionelle Erkennungssequenzen für *EcoRI* und *DraI* sind demnach nur in einem Teil der 170 Bp-Monomere der Tandemabfolge konserviert.

Zwischen den prominenten Leiterbanden der *RsaI*-restringierten DNA waren weitere feinere Banden sichtbar (Fig.C41 V.1.1.), verursacht durch eine Gruppe von Wiederholungseinheiten, die zu einer Unterfamilie mit divergenter Sequenz gehört - z.B. entstanden durch die Insertion fremder Sequenzmotive in die Monomereinheit oder durch eine Amplifikation mit anderen repetitiven Elementen.

Während *AluI*-, *DdeI*-, *EcoRV*- oder *BamHI*-restringierte DNA einen starken Schmier über den gesamten Molekulargewichtsbereich zeigte, waren die schmierförmigen Hybridisierungssignale der *ApaI*-, *HindIII*-, *PvuII*- und *PstI*-Restriktionen auf den Bereich >5 Kbp beschränkt. Beides deutet auf das Vorkommen von zufällig verteilten, nichtkonservierten Schnittstellen in den Monomereinheiten hin. Je größer die hybridisierenden Fragmente, desto weniger häufig sind die Erkennungssequenzen in den Wiederholungseinheiten vorhanden.

Die Hybridisierungen mit drei klonierten Fragmenten aus der zweiten Homologiegruppe (pCaEr453, 806 und 855) wiesen ebenfalls typische intensive Leiterbandenmuster nach, wobei die Leitersprossen der *RsaI*-Restriktion im Abstand von 100 Bp lagen und die Monomerlänge wiedergeben (Fig.C2). Folglich stellt auch die zweite Sequenzklasse ein tandemrepetitives Element dar und wurde *CaSat2* (*Cicer arietinum* **Satellit 2**) genannt. Leiterbandenmuster mit identischen Abständen zwischen den Leitersprossen wurden bei insgesamt elf Restriktionen (*AluI*, *BamHI*, *DdeI*, *DraI*, *HaeIII*, *HindIII*, *HinfI*, *MboI*, *RsaI*, *Sau3AI*, *TaqI*) sichtbar, was die Anwesenheit vieler konservierter Restriktionsschnittstellen in der Wiederholungseinheit vermuten läßt. Das Auftreten von intensiven Schmierensignalen im hohen Molekulargewichtsbereich der *ApaI*-, *EcoRI*-, *EcoRV*-, *HpaII*-, *MspI*-, *PstI*- und *PvuII*-Restriktionen zeigte darüberhinaus das verstreute Vorkommen der entsprechenden Erkennungssequenzen in den Monomeren (Fig.C2).

Für beide Satellitenfamilien deutet die große Menge homologer, hochmolekularer Fragmente sowie die hohe Signalintensität auf eine hohe Abundanz der Sequenzen im Genom hin.

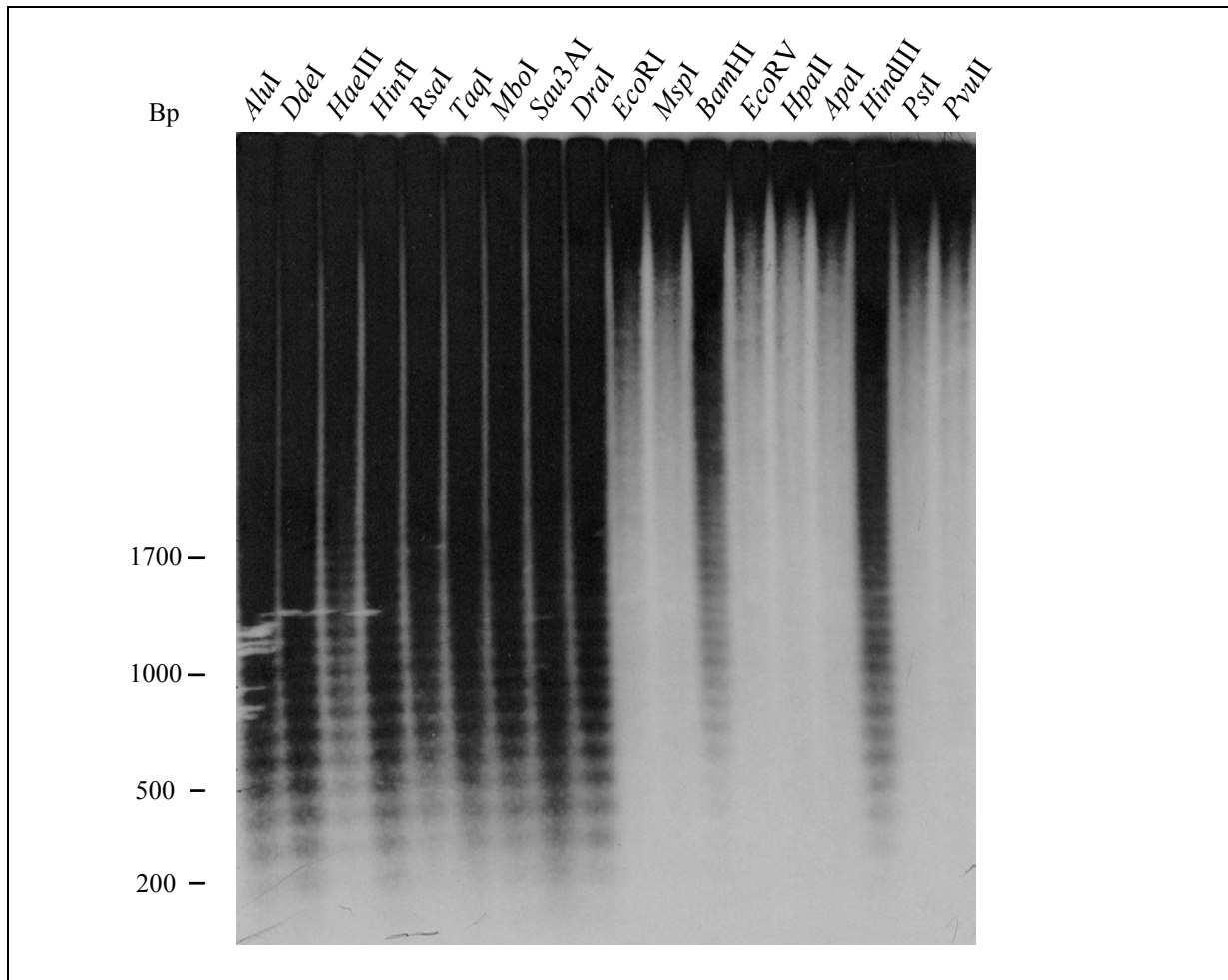


Fig.C2: Genomische Organisation des Satelliten *CaSat2* in *C. arietinum*. Autoradiogramm eines Southern-Blots mit genomischer DNA aus *C. arietinum*, die mit verschiedenen Restriktionsendonukleasen behandelt und gegen das Fragment von pCaEr453 (*CaSat2*) hybridisiert wurde. Leiterbanden im Abstand von 100 Bp lassen Fragmente bis zu 25-meren differenzieren.

***CaSat1* und *CaSat2* zeigen eine tandemartige „head-to-tail“-Orientierung der Monomereinheiten**
Ergänzend zur Southern-Analyse wurden mehrere Klone von *CaSat1* und *CaSat2* sequenziert und die Wiederholungseinheiten untereinander und mit Sequenzen in öffentlichen Datenbanken (B.I.5.1.) verglichen.

Für *CaSat1* konnten insgesamt 28 Monomere aus 11 Klonen sequenziert werden (AJ005998-AJ006001, AJ006004, AJ300290-AJ300299). Jeder Klon enthielt mehrere tandemartig in derselben Orientierung („head-to-tail“) angeordnete Monomereinheiten von 123-184 Bp Länge und einem durchschnittlichen AT-Gehalt von 61,9 % (Fig.C3.a). Die sequenzierten *EcoRI*-Fragmente mit *CaSat1*-Monomeren enthielten bis auf eine Ausnahme (pCaEr782, s.S.65) keine vom Tandemelement unabhängigen Sequenzen. Neben Deletionen und Insertionen einzelner Nukleotide sind besonders zwei große Deletionen von 39 Bp (pCaEr 636, 447) und 24 Bp (pCaEr362), eine Insertion von 21 Bp (pCaEr250) sowie eine kleine 7 Bp Deletion (pCaEr427) in einzelnen Monomeren für die Variabilität der Monomerlänge verantwortlich. Ein direkter Sequenzvergleich aller vollständigen Monomere zeigte eine

relativ einheitliche Monomerlänge von 162-167 Bp bei den übrigen, nicht von den großen Insertionen bzw. Deletionen betroffenen Satelliteneinheiten (Fig.C3a).

Obwohl zahlreiche kurze, direkte und invertierte Sequenzwiederholungen und Poly(A/T)-Abschnitte in den Monomeren vorhanden waren, fehlten größere Wiederholungseinheiten in den ca. 165 Bp Intervallen. Die Position der konservierten *RsaI*-Restriktionsschnittstelle war in den meisten Wiederholungseinheiten erkennbar, ebenso Punktmutationen, die einzelne Erkennungssequenzen verändern (Fig.C3a). Die Wiederholungseinheiten besaßen untereinander 65,5-92 % Sequenzhomologie auf Nukleotidebene, wobei die Homologie zwischen Monomeren aus demselben Klon meistens niedriger war als zwischen Monomeren aus verschiedenen Klonen der Familie (Fig.C3b).

Es konnten weder Homologien zu Sequenzen in öffentlichen Datenbanken noch Ähnlichkeiten zu anderen charakterisierten repetitiven Familien der Kichererbse gefunden werden. Zwei weitere vollständige Monomere des *CaSat1*-Tandemelements wurden unabhängig von den hier beschriebenen Untersuchungen kürzlich von M. Frediani aus der Kichererbse kloniert und sequenziert (AJ012823). Beide Monomere hatten einen AT-Gehalt von 60-64% sowie eine Länge von 159 bzw. 164 Bp und teilten damit die Charakteristika der restlichen *CaSat1*-Wiederholungen, zu denen sie 55-90% Homologie aufwiesen (Fig.C3a/b).

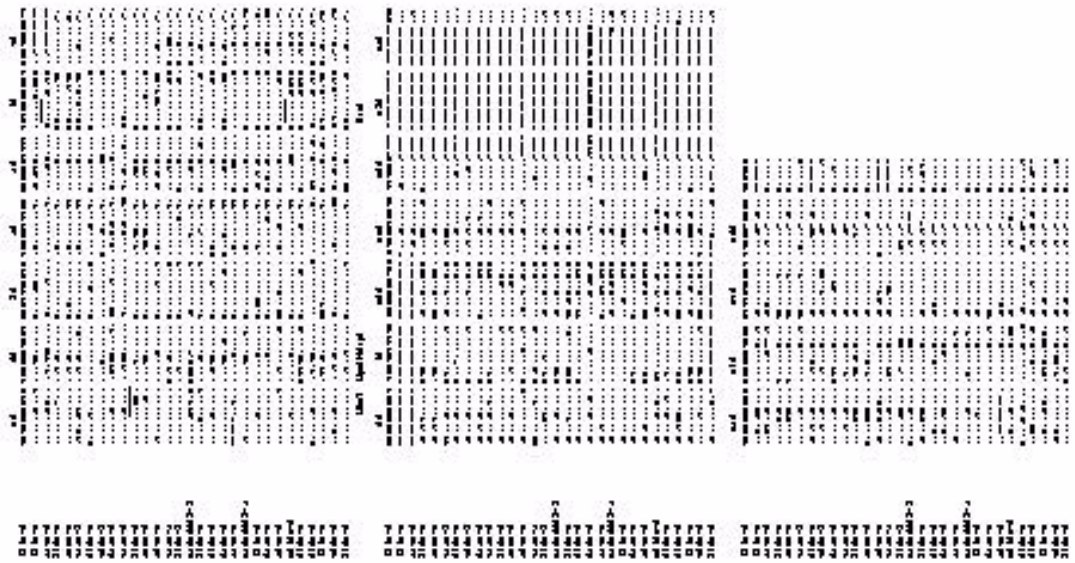
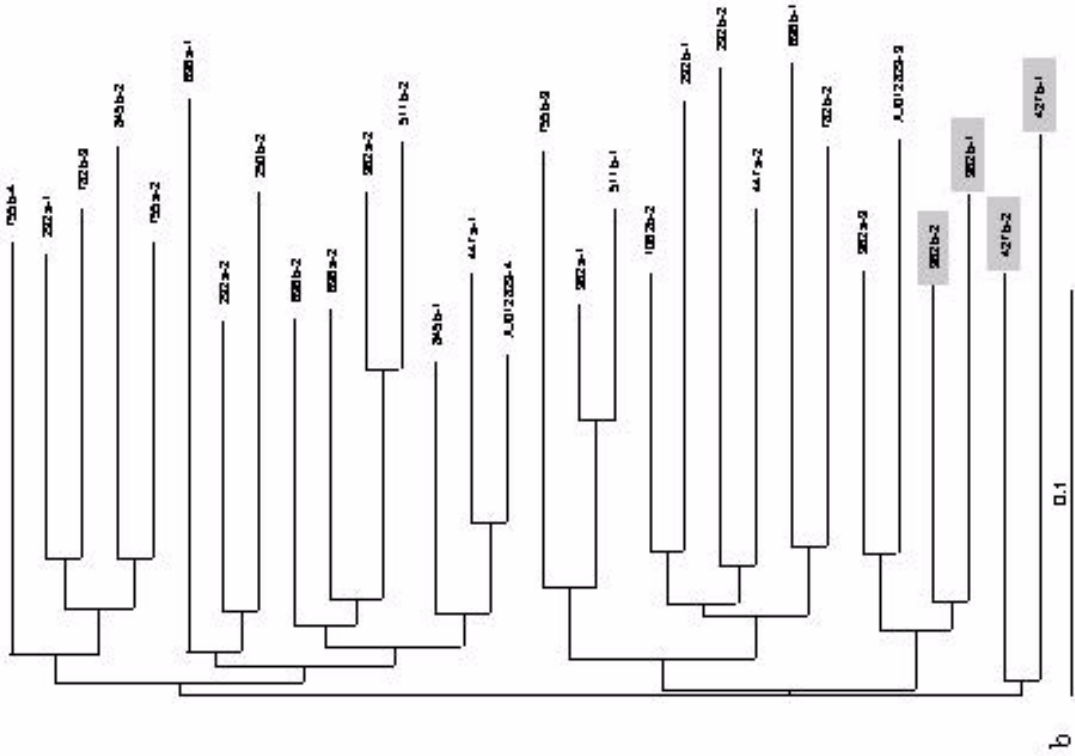
Ein *EcoRI*-Fragment aus pCaEr782 enthielt neben zwei vollständigen *CaSat1*-Monomeren eine unvollständige dritte Wiederholungseinheit, die an eine fremde Sequenz (1159 Bp) angrenzte. Dieser restliche Teil des Fragments zeigte im wahrscheinlichsten ORF 43-59% Ähnlichkeit zu *del1* (vgl. I.2.). Außerdem besaß der Bereich N-terminal zur RNA-Bindestelle im *gag*-Genprodukt und C-terminal zur Protease im Polyprotein Ähnlichkeit zu pflanzlichen Retroelementen. Es konnten keine signifikanten Homologien zu anderen in Kichererbse identifizierten Retroelementen gefunden werden, die Aminosäuresequenz wies nur geringe Ähnlichkeit (40%) zum entsprechenden Bereich in *CaRep* auf. An der Übergangsstelle zu den *CaSat1*-Sequenzen war eine 13 Bp lange Sequenz mit Homologie zu beiden Elementen zu finden. Der Grenzbereich zwischen dem unvollständigen *CaSat1*-Monomer und dem Retrotransposon enthielt außerdem die Sequenz 5`-GAAATC-3`, die möglicherweise eine während des Klonierens degenerierte *EcoRI*- Schnittstelle darstellt, so daß eine Kollokalisierung beider repetitiver Sequenzen unter Vorbehalt diskutiert werden muß.

Die Sequenzen von 20 *CaSat2*-Monomeren aus drei Klonen (AJ006002, AJ006005-AJ006007, AJ300300, AJ300301) zeigten ebenfalls eine „head-to-tail“-Anordnung der Wiederholungseinheiten und bestätigten die vermutete Monomerlänge von 100 Bp (Fig.C4a).

Fig.C3: Sequenzvergleich vollständiger Monomere von *CaSat1*.

(a) Die Nukleotidsequenz von 28 Monomeren aus pCaEr 250, 292, 427, 1062, 636, 447, 362, 511, 755, 782, 845 (AJ005998-006001, AJ006004, AJ300290-AJ300299) und das Fragment von M. Frediani (AJ012823) wurden miteinander verglichen (Abkürzungen hinter dem Namen des Klons bezeichnen das Monomer, z.B. pCaEr511b-1). Eine Monomerlänge von 162-167 Bp ist erkennbar. Bindestriche zeigen die für ein optimales Alignment eingefügten Lücken an, Punkte stehen für identische Nukleotide. Die Erkennungssequenzen der Restriktionsenzyme *HinfI*, *HpaII* und *RsaI* wurden hervorgehoben.

(b) Das mit CLUSTAL W erstellte Dendrogramm reflektiert die genetische Distanz der Sequenzen.



C3

Fig.C4: Sequenzvergleich vollständiger Monomere von *CaSat2*.

(a) Die Nukleotidsequenz von 20 Monomeren aus pCaEr 453, 806, und pCaEr 855 (AJ006002, 006005-AJ006007, AJ300300, AJ300301) wurden miteinander verglichen (Abkürzungen hinter dem Namen des Klon bezeichnen das Monomer, z.B. pCaEr511b-1). Die Monomerlänge von 100 Bp ist stark konserviert. Bindestrichen zeigen die für einen optimalen Sequenzvergleich eingefügten Lücken an, Punkte stehen für identische Nukleotide. Die Erkennungssequenzen der Restriktionsenzyme *DdeI* und *TaqI* sind hervorgehoben.

(b) Das mit CLUSTAL W generierte Dendrogramm reflektiert die genetische Distanz der Sequenzen.

Im Vergleich zu *CaSat1* zeigten die 20 Wiederholungseinheiten nur eine sehr geringe Längenvariabilität, verursacht durch Insertion oder Deletion je eines einzelnen Basenpaares in vier Monomeren. Die Sequenzen wiesen einen mittleren AT-Gehalt von 64,1 % auf und enthielten kurze direkte und inverse Sequenzwiederholungen. Nur wenige Positionen zeigten variable Nukleotide, und die Sequenzhomologie der einzelnen Wiederholungseinheiten lag mit 70-98% deutlich höher als bei *CaSat1*. Obwohl die Southern-Hybridisierungen dies nicht vermuten ließen, konnten nur wenige intakte Erkennungssequenzen von Restriktionsendonukleasen (z.B. *DdeI*, *TaqI*) gefunden werden, daneben etliche kryptische Schnittstellen mit Punktmutationen (FigC4a).

Ein Homologievergleich der *CaSat2*-Monomere untereinander zeigte in drei Fällen die höchste Homologie zwischen direkt benachbarten Einheiten, während bei den übrigen Sequenzen weit auseinanderliegende Monomere eines bestimmten bzw. verschiedener klonierter Fragmente ähnlicher waren (Fig.C4b). Auch für *CaSat2* konnten weder Homologien in öffentlichen Datenbanken noch zu *CaSat1* gefunden werden.

CaSat1 und *CaSat 2* sind demnach die ersten Satelliten-Elemente, die für *Cicer arietinum* beschrieben wurden, und stellen unabhängige AT-reiche und hochrepetitive „head-to-tail“-Tandemabfolgen aus 100 Bp- bzw. ca. 165 Bp-Monomeren dar.

C.I.1.2. Chromosomale Lokalisation

Die physikalische Kartierung der Satellitenfamilien *CaSat1* und *CaSat2* wurde mittels Fluoreszenz *in-situ*-Hybridisierung (FISH) auf Mitosechromosomen durchgeführt. Mitosechromosomen eignen sich sehr gut, um die chromosomale Lokalisation hochrepetitiver Sequenzen sichtbar zu machen, da die DNA-Fasern während der Zellteilung 5-10 mal stärker kondensiert sind als im Interphasechromatin und - besonders in der Metaphase - die Umrisse der einzelnen Chromosomen erkennbar werden. DAPI (4,6 Diamino-2-phenylindol)-gefärbte Metaphasekerne von *Cicer arietinum* weisen $2n = 16$ Chromosomen mit großen heterochromatischen Blöcken auf. Von den beiden größten Chromosomenpaaren A und B trägt A eine sog. sekundäre Einschnürung mit den transkriptionell aktiven rRNA-Gen-Tandemwiederholungen (Nukleolus Organisator Region; NOR; Fig.C5).

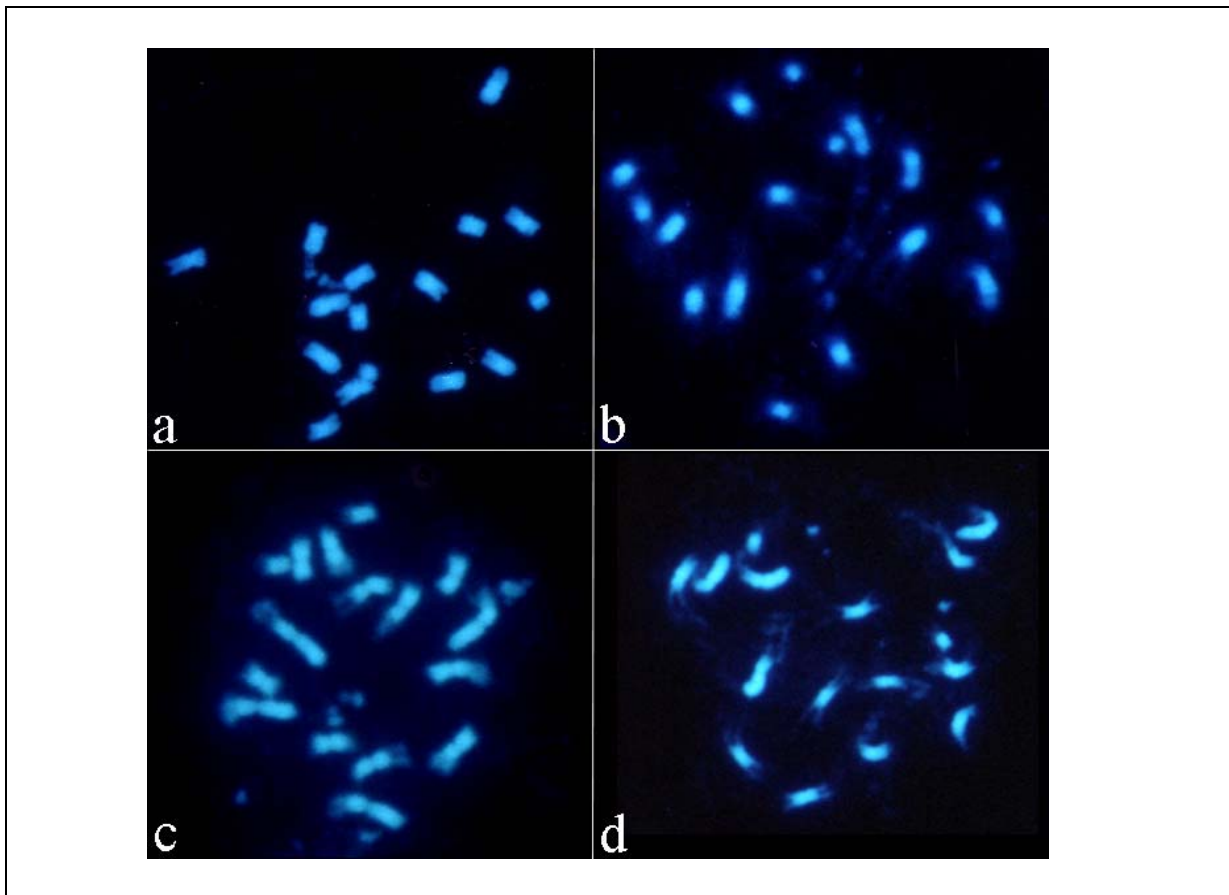


Fig.C5: DAPI-gefärbte Metaphaseplatten aus *C. arietinum*.

(a-d) Hellblaue Fluoreszenz zeigt AT-reiches Heterochromatin, während die interstitiellen euchromatischen Bereiche schwächer gefärbt bzw. unsichtbar sind. Der diploide Kern von *C. arietinum* enthält 8 Chromosomenpaare, von denen hier nur die beiden großen Paare (A und B) sowie das kleinste (H) eindeutig unterscheidbar sind. Das Chromosom A weist eine sekundäre Einschnürung mit der NOR auf.

Chromosomen A und B besitzen hohe Kopienzahlen von *CaSat1* in enger Nachbarschaft zu den 18S-5.8S-25S rDNA-Loci

Für die Fluoreszenz *in-situ*-Hybridisierung (FISH) mit *CaSat1* wurde das Fragment des genomischen Klon pCaEr1062 als Sonde verwendet. Es konnten sehr prominente Fluoreszenzsignale auf zwei Chromosomenpaaren detektiert werden, vermutlich Hauptloci von *CaSat1* mit langen Tandemabfolgen, die interstitiell im Heterochromatin des Chromosomenpaars A nahe der sekundären Einschnürung und auf Chromosomenpaar B im pericentrischen Heterochromatinblock liegen (Fig.C6c). Daneben wurden zahlreiche schwächere Signale auf allen anderen Chromosomenpaaren sichtbar, die kleineren *CaSat1*-Loci mit weniger tandemartig angeordneten Monomerkopien entsprechen (Fig.C6a). Eine gleichzeitige Hybridisierung mit verschiedenfarbig fluoreszenzmarkierten Sonden von *CaSat1* (grün) und der 18S-5.8S-25S rRNA-Genabfolge (pTA71; rot) zeigte die enge Nachbarschaft der rDNA-Loci zu den prominenten *CaSat1*-Orten (Fig.C6b,c). Beide Loci der 18S-5.8S-25S rDNA konnten mit der pTA71 Sonde sichtbar gemacht werden. Dabei grenzten die *CaSat1*-Signale auf Chromosom A an die rDNA-Loci der sekundären

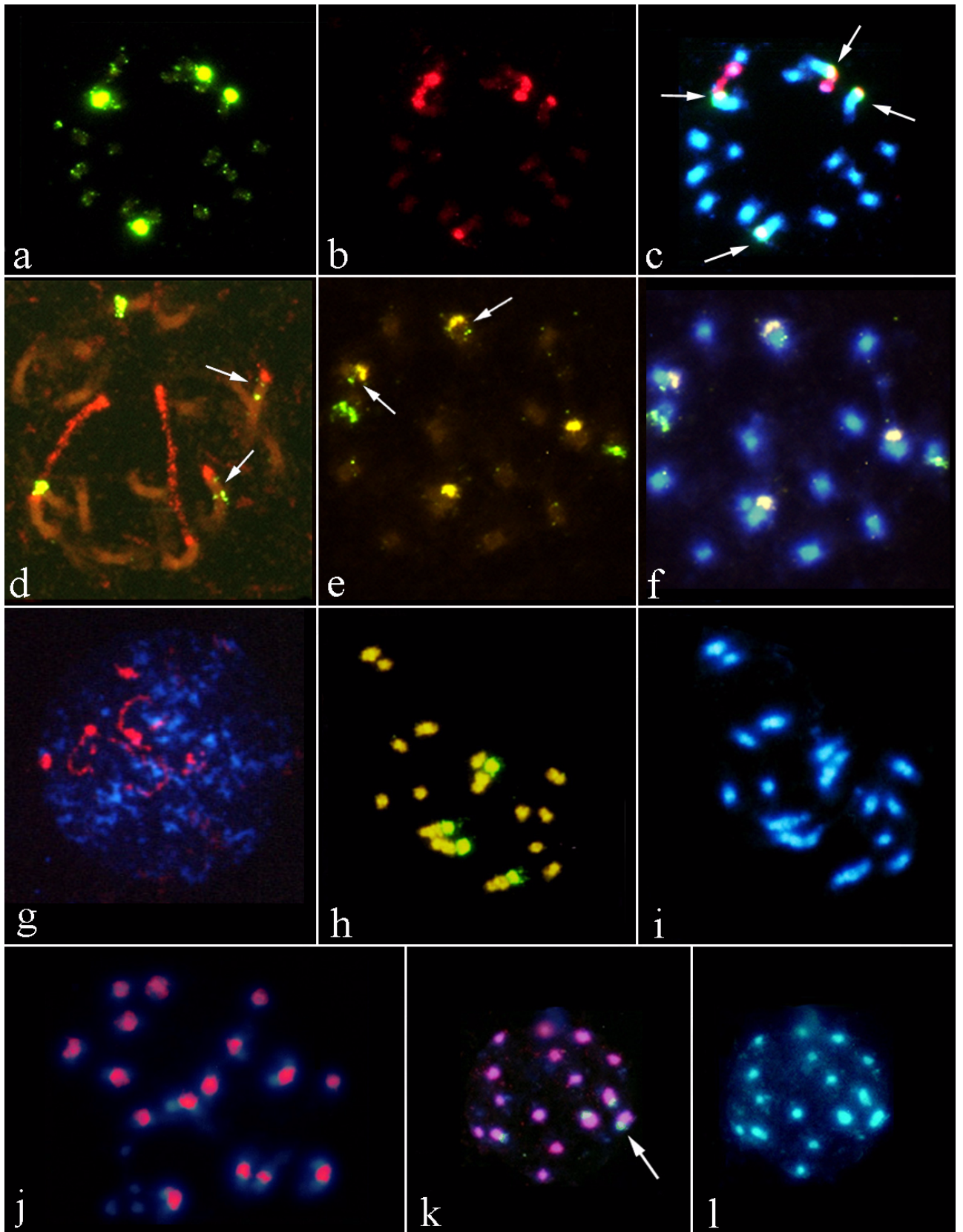
Einschnürung, bedeckten diese aber nicht vollständig (Fig.C6a-c). Auf Chromosom B war der große Satellitenlocus im distalen Bereich des pericentrischen Heterochromatinblocks lokalisiert, d.h. auch hier in unmittelbarer Nähe zu den subterminalen 18S rDNA-Wiederholungen (Fig.C6a-c). Weitere Doppelhybridisierungen mit einer 5S rDNA-Sonde (pXVI) machten die beiden Loci der 5S rDNA auf Chromosom B und einem mittleren Chromosom (E nach Galasso et al., 1996) sichtbar. In Übereinstimmung mit Galasso et al. (1996) war das 5S rDNA-Signal auf Chromosom B proximal zu den 18S rDNA-Loci im distalen Bereich des Heterochromatins lokalisiert (Fig.C6d). Auch die Hauptloci von *CaSat1* erschienen distal und deutlich getrennt von den 5S rDNA-Signalen auf Chromosom B (Fig.C6e,f). Überlappungen von kleineren Loci mit schwächeren Hybridisierungssignalen können jedoch nicht ausgeschlossen werden. Inwieweit *CaSat1*-Sequenzen mit den Sequenzen der 18S rDNA-Transkriptionseinheit überlappen, kann aufgrund der hohen Signalintensitäten bzw. der geringen Auflösung auf stark kondensierten Metaphasechromosomen nicht bestimmt werden.

***CaSat2* bildet den Hauptanteil des pericentrischen Heterochromatinblocks auf allen 8 Chromosomenpaaren**

Für *CaSat2* wurde eine subklonierte partielle Sequenz von pCaEr 855 als Sonde verwendet, die sehr intensive Hybridisierungssignale in den hellblau DAPI-gefärbten pericentrischen Heterochromatinblöcken aller 16 Metaphasechromosomen erzeugte. Die starke Intensität der Hybridisierungssignale, die den größten Teil des pericentrischen Heterochromatins

Fig.C6: Lokalisation der Satelliten *CaSat1*, *CaSat2* und rDNA durch Fluoreszenz *in-situ*-Hybridisierung auf Chromosomen von *C. arietinum*.

(a) Metaphasechromosomen nach *in-situ*-Hybridisierung mit dem Element *CaSat1* (pCaEr1062). Die Hauptloci sind als hellgrüne Fluoreszenz sichtbar und auf Chromosomen A und B konzentriert. Schwächere Signale kleinerer Satellitenanhäufungen befinden sich auf allen Chromosomen. (b) Dieselbe Metaphase nach Hybridisierung mit einer rRNA-Gensonde zeigt die 18S-5.8S-25S rRNA-Genfolge auf zwei Chromosomenpaaren (rote Fluoreszenz). Zwei der Loci sind kondensiert. (c) Computergestützte Kombination aus (a) und (b) mit DAPI-gefärbten Chromosomen derselben Metaphase (blau) zeigt die partielle Überlappung von *CaSat1* und der 18S-5.8S-25S rRNA-Genfolge auf den Chromosomen A und B (hellgelbe Signale, Pfeile). (d) Die Doppel-*in-situ*-Hybridisierung mit den 18S-5.8S-25S rRNA-Genabfolgen (rot) und einer Sonde der 5S rRNA-Gene (gelb-grünes Signal) zeigt die Lokalisation beider Sequenzen auf Chromosom B (Pfeile), wobei die 18S-5.8S-25S rRNA-Genfolge die distalere Position einnimmt. (e) Doppelhybridisierung mit der 5S rRNA-Gensonde (grüne Signale) und *CaSat1* (pCaEr855; gelb) auf einer weiteren Metaphase. (f) Überlagerung derselben Metaphase nach DAPI-Färbung (blau) mit den Signalen der Doppelhybridisierung aus (e). Auch die Satellitensequenz *CaSat1* liegt distal zu den 5S rRNA-Genen (Pfeile). (g) Interphase nach DAPI-Färbung (blau) und Hybridisierung mit der 18S-5.8S-25S rRNA-Sonde (rote Signale). Gut erkennbar sind die in der Interphase teilweise dekondensierten Bereiche der NOR. (h) Metaphasechromosomen nach Doppelhybridisierung mit *CaSat1* (pCaEr1062; gelb-grüne Fluoreszenz) und *CaSat2* (pCaEr855; orange). (i) Metaphase von (h) nach DAPI-Färbung (blaue Fluoreszenz). (j) Eine andere Metaphase mit einer computergestützten Überlagerung von DAPI-Färbung (blau) und Hybridisierungssignalen einer *CaSat2*-Sonde (pCaEr855; rotes Signal). (k) Die simultane Lokalisation von *CaSat1* (pCaEr1062; hellgrün) und *CaSat2* (pCaEr855; rotes Signal) enthüllt die enge physikalische Nachbarschaft der Signale im Interphasekern. (l) Dieselbe Interphase nach DAPI-Färbung zeigt, daß die heterochromatischen Regionen vollständig den Anhäufungen der beiden Satelliten entsprechen.



ausmachten, bestätigt die schon bei der Southern-Hybridisierung erkennbare hohe Abundanz der *CaSat2*-Sequenzen im Kichererbsengenom (Fig.C6j,k,l). Eine Doppelhybridisierung mit den verschiedenfarbig fluoreszenzmarkierten Sonden beider Satellitenelemente zeigte, daß die *CaSat2*-Signale auf den Chromosomen A und B benachbart zu den *CaSat1*-Hauptloci liegen, jedoch deutlich von diesen abgegrenzt sind (Fig.C6h,i).

C.I.1.3. Vorkommen und Abundanz der Satellitenelemente in anderen Spezies

Mit Hilfe von Southern-Hybridisierungen wurden Spezies der Gattung *Cicer*, weitere Leguminosen und andere Dikotyledonen (s.B.I.1.) auf die Anwesenheit von Sequenzen untersucht, die homolog zu den identifizierten Satellitenelementen *CaSat1* und *CaSat2* aus *Cicer arietinum* sind. Gleichzeitig sollten Abundanz und genomische Organisation dieser Sequenzen festgestellt werden. *RsaI*-restringierte, genomische DNA aus einer oder mehreren Akzessionen der jeweiligen Spezies wurde gegen klonierte Fragmente der *CaSat1*- und *CaSat2*-Familien hybridisiert.

CaSat1* und *CaSat2* Sequenzen sind spezifisch für die Gattung *Cicer

Für den Nachweis von *CaSat1* wurden die Fragmente der Klone pCaEr292 und pCaEr1062, für *CaSat2* pCaEr453 sowie eine subklonierte partielle Sequenz dieses Fragments verwendet. Weder für die erste Satellitenfamilie noch für *CaSat2* konnten nach Southern-Hybridisierung Signale in Spezies außerhalb der Gattung *Cicer* detektiert werden (nicht dargestellt). Dagegen wurden intensive Signale innerhalb der *Cicer*-Spezies gefunden, die gesondert und ausführlich im Kapitel V.1.1. beschrieben werden. Beide Satellitenelemente sind damit offenbar spezifisch für die Gattung *Cicer*.

C.I.2. Das disperse repetitive Element *CaRep*

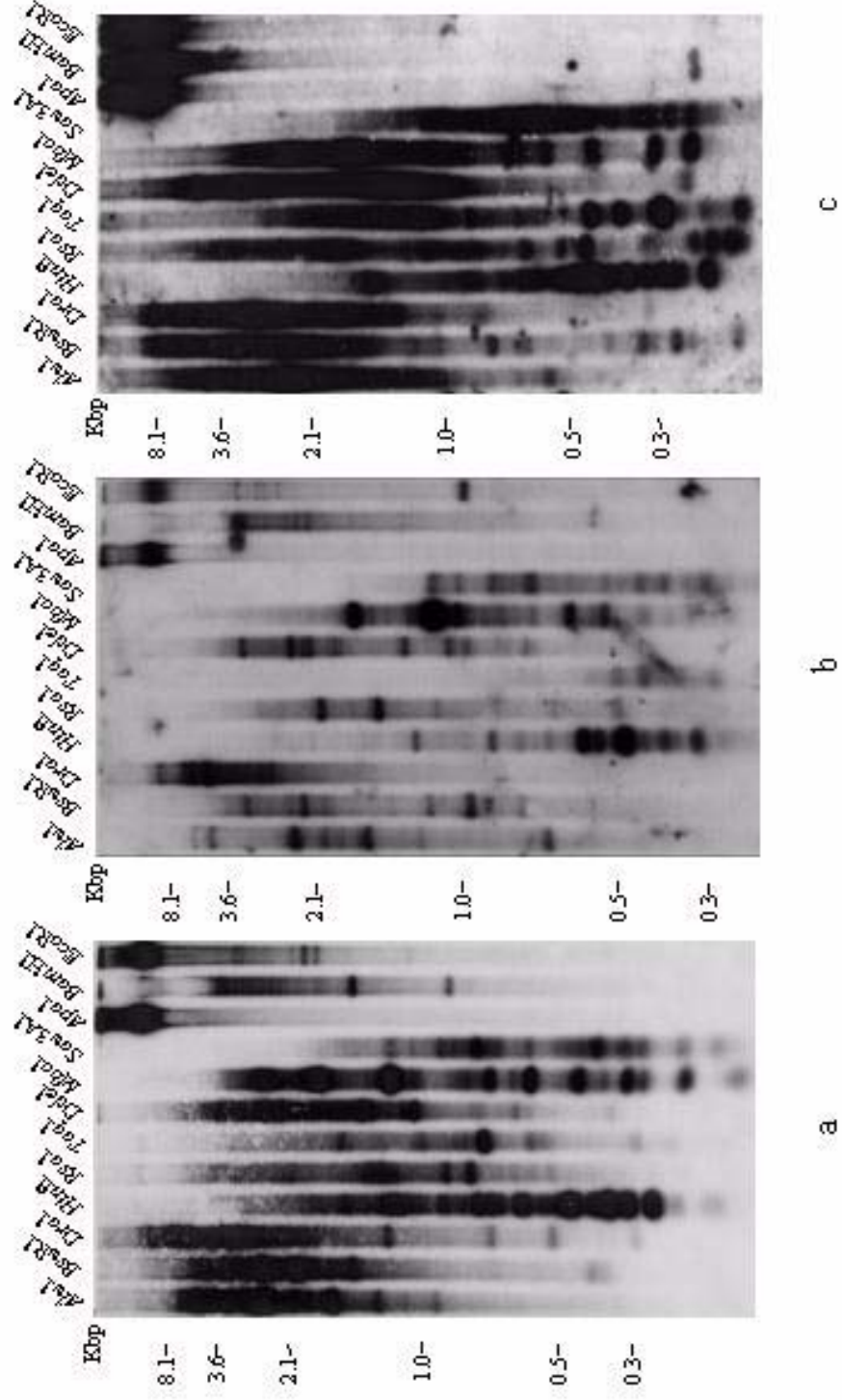
C.I.2.1. Genomische Organisation der nicht-tandemartig organisierten hochabundanten Sequenzen

Drei weitere hochabundante Sequenzgruppen sind dispers im Genom verteilt

Drei der hochrepetitiven Sequenzfamilien konnten nicht den beiden Tandemelementen zugeordnet werden und wiesen eine andere genomische Organisation im Kichererbsengenom auf. Bei einer Southern-Hybridisierung von genomischer DNA aus *Cicer arietinum*, die mit 18 verschiedenen Endonukleasen restringiert wurde, erzeugten die radioaktiv markierten Fragmente der Klone pCaEr1048 (Gruppe 3), pCaEr915 und 936 (beide Gruppe 4) und pCaEr776 (Gruppe 5) intensive Hybridisierungssignale, jedoch keine Leiterbanden. Statt

Fig.C7: Genomische Organisation der Retrotransposonsequenzen *CaRep1-3* im Genom von *C. arietinum*.

(a) Autoradiogramm eines Southern-Blots mit genomischer DNA aus *C. arietinum*, die mit verschiedenen Restriktionsendonukleasen behandelt und gegen das Fragment aus pCaEr1048 (*CaRep1*; Gruppe 3) hybridisiert wurde. (b) Identische Filter wurden gegen das Fragment von pCaEr936 (*CaRep2*; Gruppe 4) und (c) pCaEr776 (*CaRep3*; Gruppe 5) hybridisiert. Unterschiedliche und komplexe Bandenmuster deuten auf eine nicht-tandemartige Organisation dieser repetitiven Sequenzen im Genom hin.



C7

dessen konnten in fast allen Spuren unterschiedliche Muster aus 4-10 Banden im Größenbereich zwischen 0.2-4 Kbp auf einem intensiven Hintergrundsmier detektiert werden. Die einzelnen Homologiegruppen unterschieden sich dabei in ihrem Bandenmuster (Fig.C7a-c). Während der starke Hintergrundsmier auf eine hohe Abundanz homologer Wiederholungseinheiten im Genom hindeutet, verraten die Bandenmuster eine disperse, möglicherweise komplexe genomische Organisation der drei klonierten Sequenzgruppen. Daher muß man das Vorhandensein von drei weiteren dispersen und hochrepetitiven Sequenzfamilien im Genom annehmen.

C.I.2.2. Sequenzierung von Klonen der drei dispersen Sequenzgruppen

Alle drei Sequenzgruppen zeigen Homologien zu verschiedenen Bereichen pflanzlicher *Ty3-gypsy-Retrotransposons*

Insgesamt wurde die Sequenz von 10 klonierten Fragmenten der Homologiegruppen 3-5 ermittelt (AJ411796-AJ411805), von denen alle signifikante Homologien zu Retroelementen aufwiesen. Am häufigsten wurden Homologien zu pflanzlichen LTR-Retrotransposons des *Ty3-gypsy*-Typs, wie etwa aus *Lycopersicon esculentum* (Su & Brown, 1997; AF119040), *Zea mays* (Avramova et al., 1996), *Oryza sativa* (AF111709; Maggy AP000364), *Sorghum bicolor* (AF061282), *Pisum sativum* (AF083074) und *Ananas comosus* (Y12432) gefunden. Ebenso gab es Ähnlichkeiten zu *gypsy*- Elementen aus Pteridophyten (Nozue et al., 1997) und Pilzen (Anaya & Roncero, 1995) sowie LTR- Retrotransposons aus *Drosophila* (Tanda et al., 1994). Die Homologie wurde meistens auf Ebene der wahrscheinlichsten Aminosäuresequenz eines codierten Proteins nachgewiesen, bei vielen Fragmenten zusätzlich auch direkt auf Nukleotidebene. Die verschiedenen Sequenzgruppen wiesen dabei immer Homologien zu unterschiedlichen Bereichen eines Retroelements auf. Besonders deutlich wird dies, wenn man diese Sequenzen mit *dell*, einem *Ty3-gypsy*-Element aus *Lilium henryi* (Smyth et al., 1989) vergleicht, zu dem alle klonierten Fragmente sehr starke Ähnlichkeiten aufweisen (Fig.C8).

Das 2.140 Kbp Fragment von pCaEr1048 (Gruppe3; AJ411796) hatte einen AT-Gehalt von 60% und zeigte keine Homologien zu den Fragmenten der beiden anderen Gruppen. Das für 111 Aminosäuren codierende intakte offene Leseraster zeigte signifikante Homologien zum *gag*-Bereich von *dell*, weiter 5' davon lag eine potentielle „Primer“-Bindestelle (PBS) mit Sequenzverwandtschaft zu tRNA-Motiven. Innerhalb der Sequenz von *dell* umfaßte die Homologie die Positionen 2567-3116 Bp und den Bereich davor bis in den LTR (ab Position 2406) hinein (Fig.C8).

Aus der Gruppe 4 wurden fünf 956 Bp lange Klone (pCaEr 467, 757, 915, 936 und 1034) sowie pCaEr638 mit 1.55 Kbp sequenziert (AJ411797-AJ411802). Diese Fragmente wiesen einen AT-Gehalt von 54-61% auf und zeigten keine Verwandtschaft zu den anderen Klonen. Die ersten 956 Bp waren innerhalb der Fragmente zu 86-97% homolog. Ein durchgehender, für 318 Aminosäuren codierender ORF war bei der Hälfte der Fragmente intakt, wies aber auch bei den restlichen, durch wenige Leserasterwechsel und Stopcodons veränderten

Sequenzen eine deutliche Homologie zum *gag*-Leseraster und der Proteaseregion (3.216-4.064 Kbp) von *del1* auf. Aufgrund leichter Sequenzdivergenz im Bereich der Protease war bei pCaEr638 die nach ca. 956 Bp in allen Fragmenten vorhandene *EcoRI*-Schnittstelle degeneriert, was die Klonierung einer längeren Sequenz aus der gleichen Region ermöglichte. Sie zeigte daher zusätzlich Homologie zur 3' davon gelegenen konservierten Domäne der Reversen Transkriptase (RT) und überspannte den Bereich 3.216-4.667 Kbp in *del1* (Fig.C8).

Die drei Sequenzen aus Gruppe 5 (pCaER776-2058 Bp, pCaEr432-1986 Bp und pCaEr1049-975 Bp) hatten ebenfalls keine Ähnlichkeiten zu Gruppe 3 oder 4 und besaßen einen AT-Gehalt von 61-63% (AJ411803-AJ411805). Bis zur Position 560 Bp waren die Sequenzen auf Nukleotidebene 91-96% homolog zueinander, dann folgten verschiedene große Insertionen und Deletionen sowie ein auffälliges Feld aus komplexen mikrosatellitenartigen Motiven, die einen direkten Sequenzvergleich schwierig machten (s.2.4). Der konserviertere 5'-Bereich der Klone enthielt ein für 139 Aminosäuren codierendes offenes Leseraster mit Homologie zur Integrase und eine Sequenz, die dem Polypurintrakt (PPT) in der 3'-Region des Polyproteinleserasters in *del1* entspricht (Fig.C8). Insgesamt bestand Ähnlichkeit zum Bereich 6.467-6.947 Kbp des Retroelements aus der Lilie, der den Beginn des rechten LTRs miteinbezieht. Die stärker divergierenden und komplex strukturierten 3'-Bereiche der drei

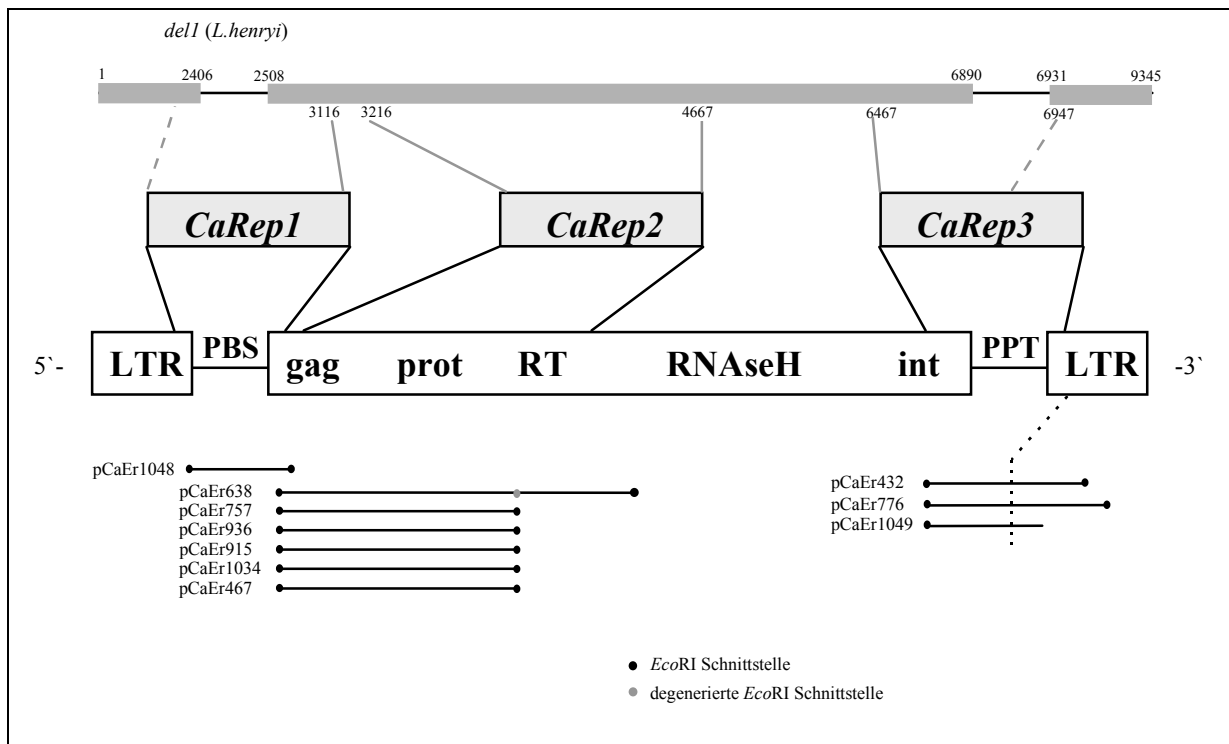


Fig.C8: Die drei Sequenzgruppen *CaRep1-3* und die homologen Bereiche in *del1* aus *Lilium henryi*.

Schematische Darstellung der klonierten Fragmente aus *C. arietinum* (schwarze Linien), homologer Bereiche des Elements *del1* aus *L. henryi* (graues Schema) und der typischen Strukturelementen in *Ty3-gypsy*-Retrotransposons (weiße Kästen). Beginn und Ende der LTRs und des Polyproteinleserasters in *del1* sowie die Begrenzungen der Homologiebereiche zu *CaRep1-3* sind als graue Zahlen in *del1* angegeben. Die Darstellung

ist nicht maßstabsgetreu. Die drei Sequenzklassen weisen deutliche Ähnlichkeiten zu unterschiedlichen Abschnitten von *del1* auf.

Fragmente können dementsprechend Teile eines LTRs aus *C. arietinum* darstellen und weisen aufgrund der Individualität von LTR-Sequenzen keine weiteren Homologien zu *del1* auf. Bei den drei untersuchten Sequenzgruppen aus der genomischen Bank handelt es sich offenbar um Teile von Retrotransposonkopien im Genom der Kichererbse. Ob die klonierten Sequenzen verschiedene Regionen einer einzigen Retrotransposonfamilie in *C. arietinum* repräsentieren oder zu verschiedenen Familien gehören, kann jedoch erst durch die unmittelbare Nachbarschaft der klonierten Sequenzgruppen im Genom bewiesen werden.

Struktur und Sequenz eines Klons aus Gruppe 4 lassen sich dem *Ty3-gypsy*-Typ zuordnen

Vergleicht man die wahrscheinlichste Aminosäuresequenz des von pCaEr638 codierten RT-Proteins mit den entsprechenden Regionen aus der RT anderer Retroelemente, so wird die stärkere Verwandtschaft zu *Ty3-gypsy*-Sequenzen im Vergleich zu *Ty1-copia*-RT-Sequenzen deutlich. Fig.C9a zeigt einen Sequenzvergleich mit dem CLUSTALW-Algorithmus (Higgins & Sharp, 1988), aus dem sich ein Dendrogramm ableiten läßt, das die Sequenzverwandtschaft widerspiegelt (Fig.C9b). Der Baum wurde nach der „neighbour joining“-Methode (Saitou & Nei, 1987) erstellt, wobei die LINE-verwandten Sequenzen als „outgroup“ dienen. Die in pCaEr638 erkennbare Abfolge „Protease-RT“ der Polyproteindomänen ist charakteristisch für den *gypsy*-Typ, während in einem intakten *copia*-Polyprotein beide Domänen durch die Integrasedomäne getrennt werden („Protease-Integrase-RT“; s.A.I.2.).

Die drei hochrepetitiven und dispers organisierten Sequenzgruppen 3-5 stellen demnach Teile von LTR-Retroelementen dar und werden *CaRep1-3* (*Cicer arietinum* Retrotransposon) genannt. Zumindest im Fall von *CaRep2* (Gruppe 4) können sie zweifelsfrei dem *Ty3-gypsy*-Typ zugeordnet werden.

Fig.C9: Sequenzverwandtschaft zwischen *CaRep2* und weiteren LTR- und non-LTR-Retrotransposons.

(a) Sequenzvergleich des wahrscheinlichsten Aminosäureleserasters (ORF1, Nukleotide 1-537) aus pCaEr638 (*CaRep2*) mit den entsprechenden Regionen der *Ty1-copia*-Elemente *Tnt1* aus *Nicotiana tabacum* (X13777), *Tal1* aus *Arabidopsis thaliana* (X53973) und *copia* aus *Drosophila melanogaster* (X01472), *Ty3-gypsy*-Elementen aus *Ananas comosus* (*dea*; Y12432), *del1-46* aus *Lilium henryi* (Smyth et al., 1989) und *Ty3* aus *Saccharomyces cerevisiae* (M34549) sowie den non-LTR-Retrotransposons *BvLi3* aus *Beta vulgaris* (Y13368) und *Ta11-1* aus *Arabidopsis thaliana* (LA47193). Striche zeigen die für ein optimales Alignment eingefügten Lücken an, Sterne markieren Stop- Codons. Homologe Aminosäuren in einer Gruppe sind grau unterlegt. (b) Dendrogramm der Aminosäuresequenzen aus (a). Das Bäumchen wurde mit Hilfe des CLUSTAL W-Programms erstellt (Higgins & Sharp, 1988). Horizontale Linien im Dendrogramm repräsentieren die Divergenz

C. Ergebnisse

zwischen den Sequenzen, der vertikale Abstand ist ohne Belang. Die engere Verwandtschaft von *CaRep2* zu *Ty3-gypsy*-Sequenzen ist in beiden Abbildungen erkennbar.

```

Tnt1      -YKLV-ELPKGKRPLK-CKWVFKLKKDGD-CKLVRYKARLVVKGFEQKKGIDFDEIFSPV
Ta1       -WVLV-DKPQDRKIIG-CRWLFKMKSGIPGVEPVRYKARLVAKGYTOREGVDYQEIFAPV
copia     -WTIT-KRPEKNKIVD-SRWVFSVKYNELGN-PIRYKARLVARGFTQKYQIDYEETFAEV
del1-46   -LNKG-FTRGSTSPWG-AHVLFDPKKDDSD-----KRMCIDYK*LNSVTVKNKYPLPR
dea1      -LDKG-FTRPSVSPWG-APVLFVKKKDGSD-----LRLCVDYRELNKVTIKNKYPLPR
pCAEr638  -LEKK-FTRPSVSPWG-APVLLVKKKDGSD-----SRLCVDYRQLNKATIKNKYPLPR
Ty3       -LDNK-FIVPSKSPCS-SPVVLVPKKDGTD-----FRLCVDYRVLNKATISDPFPLPR
Ta11-1    EIQNFFLLGYFPKSWNFTHLCLLPKKKKPKDMTDLRPISLCSVLYKIIISKIMVRRRL-QPF
BvLi3     IVEDFWASGSLPHGCNSAFTALIPKTQHPRGFNEFRPISMGCLYKIIAKILARRL-QRV

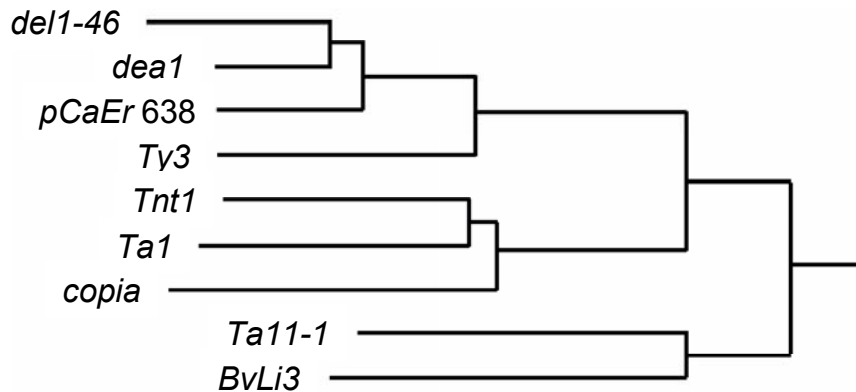
Tnt1      VKMTSI-----RTILSLAASLDLEVEQLDVKTAFLHGDLEE
Ta1       VKHTFI-----RIILMSVVVDQDLELEQMDVKTAFHGELEE
copia     ARISSE-----RFILSLVIQYNLKVHQMVKTAFLNGTLKE
del1-46   TDDLFD-----QLNGA*YFSKIDLRFYHQLRIRA
dea1      TDDLFD-----QLQGSVCVYSKIDLQSGYHQLKIKP
pCAEr638  TDDLMD-----QLRGAALFMSMIDLKSGYHQIRVR
Ty3       TDNLLS-----RIGNAQIFTTLDLHSGYHQIPMEP
Ta11-1    LPDLVSPNQSAFVAERLIFDNILIAHEVVHGLRTHKSVSKGFIAIKSNMSKAFDRVEWNY
BvLi3     MDHLIGPYQSSFIKGRQILDGVLIASELIDSCRR---MKNEAVVLKLDLPHKAFDSDISWNY

Tnt1      EIYMEQPEGFEVAG-----KKHMVCKLNKSLYGLKQAPROWYMK
Ta1       ELYMEQPEGFISED-----GKNKVCLLKKLLYGLKQSPRQWNKR
copia     EIYMRLPQGISCSNS-----DNVCKLNKAIYGLKQAAARCWFEV
del1-46   -DIPKTAFRTR-----YGHYEFVMPFGLTNVPTAFMNL
dea1      EDVSKTAFRTR-----YGHYEFVMPFGLTNAPTAFMDL
pCAEr638  EDIQKTAFRTR-----YGHYELIMMGVNTNAPTIFMDY
Ty3       KDRYKTAFTVP-----SGKYEYTVMPFGLVNAFSTFARY
Ta11-1    VRALLDALGFHQKVVWG-WIMFMISSVSYSVLINDKAFGNIVPSRGLRQGDPLSPFLFVLC
BvLi3     LEWVLCQMNFPVWRS-WIRSCVMSASAAVLINGS PSSVFKLQRGLRQGDPLS-----

Tnt1      FDSFMKSQTYLKYSDPCVYFKRFSNFIIILLYVDDMLIV
Ta1       FNRFMIDQNFIRSEHDACVKQVTEQEHLY-LLLYVDDMLIA
copia     FEQALKECECFVNSVDRCI-----NENIYVLLYVDDVVIA
del1-46   MNRVEREYLDKE-----IVVFVDYVLIY
dea1      MNRVEKPYLDRE-----VVVFIDDILVY
pCAEr638  MNRIFNPFLDKE-----VVVFINDILY
Ty3       MADTER-DL-RE-----VNVYLDLILF
Ta11-1    SEGLTHLMNRAERQGLLSGIRFSENGPAIHLLFADDSLFM
BvLi3     -----

```

a



b

C9

C.I.2.3. Identifikation des Polyprotein-Leserasters eines vollständigen *CaRep*-Elements

Aus den vorliegenden Sequenzdaten ist die Vermutung naheliegend, daß alle drei Sequenzgruppen *CaRep*1-3 zu derselben Retrotransposonfamilie im Kichererbsengenom gehören und verschiedene Abschnitte im und um das Polyproteinleseraster darstellen. Falls dies zutrifft und es sich bei den klonierten Fragmenten aus der genomischen Bank nicht um stark divergierte, vom Großteil der zu postulierenden Kopien abweichende Sequenzen handelt, sollten die fehlenden Zwischenstücke des Polyproteinleserasters isolierbar sein und so der Zusammenhang zwischen den drei Sequenzgruppen auf genomischer Ebene hergestellt werden können.

Fehlende *CaRep*-Sequenzen können durch PCR-Amplifikation ergänzt und identifiziert werden

Um die postulierten, aber noch fehlenden Zwischenstücke selektiv aus dem Genom zu isolieren, wurden Oligonukleotide aus den Randbereichen der bereits vorhandenen klonierten Fragmente abgeleitet. Ihre Orientierung ermöglicht die PCR-Amplifikation der Zwischenstücke

nur, wenn die Sequenzen der Homologiegruppen *CaRep*1, 2 und 3 im Genom tatsächlich flankierend zu den Zielsequenzen vorliegen (Fig.C10). Da die ermittelten Abstände zwischen den konservierten Polyproteindomänen und Bindemotiven im Kichererbsengenom gut mit denen aus *Lilium henryi* korrelierten, boten die Abstände der bereits bekannten homologen Bereiche in der Sequenz von *dell* einen Anhaltspunkt für die zu erwartende Fragmentgröße. Demnach konnten sowohl ein Fragment von etwa 284 Bp für das Oligonukleotidpaar DOM1/2 als auch ein ca. 1971 Bp-Amplifikat für DOM5/6 angenommen werden (Fig.C10). Zusätzlich wurde ein Primerpaar (DOM3/4) für einen 378 Bp langen Bereich der RT-Region abgeleitet, der lediglich von pCaEr638 abgedeckt wird.

Die PCR-Amplifikation wurde mit genomischen Templaten aus verschiedenen Akzessionen von *C. arietinum* und *C. reticulatum* durchgeführt. Mit dem Primerpaar DOM1/2 wurde in allen Fällen eine ca. 300 Bp lange Fragmentklasse amplifiziert, mit den Oligonukleotiden DOM5/6 entstand bei den Kichererbsenakzessionen eine Bande von ca. 2.2 Kbp. Beide Amplifikate entsprachen den erwarteten Größen (Fig.C11). Das Primerpaar DOM3/4 erzielte sowohl aus den genomischen Templaten von *C. arietinum* und *C. reticulatum* als auch dem Fragment von pCaEr638 einheitliche Amplifikate von ca. 400 Bp Länge (Fig.C11). Die einheitliche Länge der mit DOM3/4 produzierten Amplifikate aus genomischer DNA und dem Fragment des Klons pCaEr638 zeigt, daß sich die entsprechenden Oligonukleotidbindestellen im Genom im gleichen Abstand zueinander befinden wie auf dem klonierten Fragment. Daher kann letztere Sequenz als Stellvertreter für homologe Bereiche im Genom dienen und es können dort größere Insertionen oder Deletionen ausgeschlossen werden.

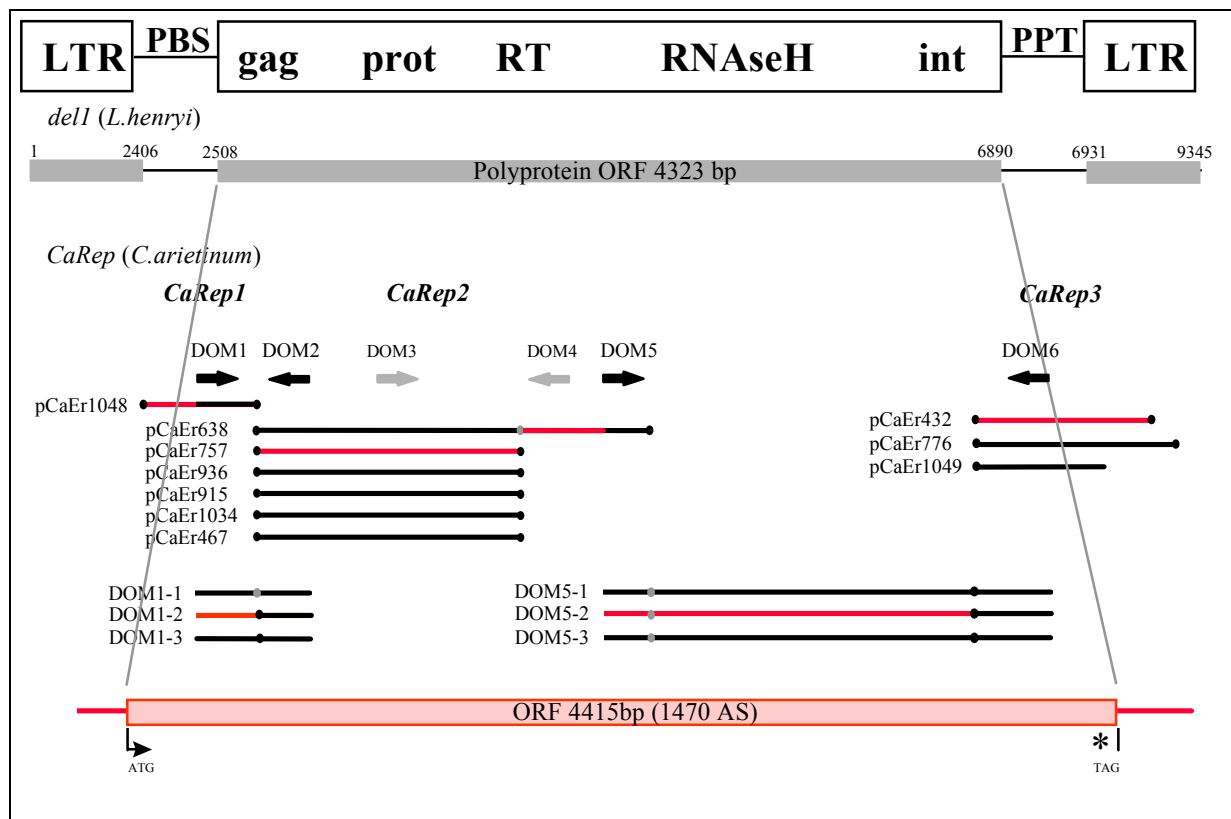


Fig.C10: *CaRep1-3* sind Teile eines Retrotransposons im Genom von *Cicer arietinum*.

Schematische Darstellung der klonierten Zwischenbereiche DOM1(1-3) und DOM5 (1-3) und Ausrichtung der Sequenzgruppen *CaRep1-3* durch Sequenzvergleiche. Die Positionen der Primerpaare DOM 1/2 und DOM 5/6 sind durch schwarze Pfeile an den Sequenzen der klonierten Fragmente *CaRep1-3* (schwarze Linien) markiert, die von DOM3/4 durch graue Pfeile. Amplifikate sind möglich, da die drei Sequenzgruppen im Genom tatsächlich zueinander benachbart sind. Die rot markierten Abschnitte wurden zur Erstellung einer zusammengesetzten Teilsequenz von *CaRep* verwendet (s. Fig.C12).

Die Sequenzierung der Amplifikate bestätigt ihre Zugehörigkeit zur *CaRep*-Familie

Die Sequenzen der mit DOM1/2 amplifizierten und klonierten Fragmente (DOM1-1, 1-2, 1-3) waren 347-351 Bp lang, hatten einen AT-Gehalt von 53- 54 % und waren untereinander auf Nukleotidebene 78-95 % ähnlich (AJ411806-AJ411808). Alle klonierten Fragmente hatten ein durchgehendes ORF, das für 115-117 Aminosäuren codiert, zu 87-95% identische Sequenzen aufwies und das nur bei DOM1-1 durch ein Stop-Codon unterbrochen wurde. Die beiden anderen Fragmente zeigten fast über ihre gesamte Länge hinweg Homologie zur Polyproteinregion der bereits von den *CaRep1-3*-Sequenzen detektierten *Ty3-gypsy*-Retrotransposons aus *Lilium henryi* (*dell*) und *Sorghum bicolor* (AF061282). Der zu 45 % homologe Bereich zu *dell* befand sich erwartungsgemäß im 5'-Bereich des *gag*-Gens (Position 3006-3269 Bp; Fig.C10), zeigte jedoch im Vergleich zum *dell*-ORF 12 zusätzliche Aminosäuren, was die um 50 Bp über der erwarteten Länge liegenden Amplifikate aus der Kichererbse erklärt. Fast über ihre gesamte Länge überlappten die Fragmente mit den Sequenzen aus *CaRep1* und 2, und zeigten, daß diese beiden Sequenzgruppen, nur durch eine *EcoRI*-Schnittstelle getrennt, unmittelbar aneinander grenzen (Fig.C10). Die Homologie auf

Nukleotidebene zu *CaRep1* betrug 90-94% und zu *CaRep2* 83-98 %. Sowohl die Überlappung

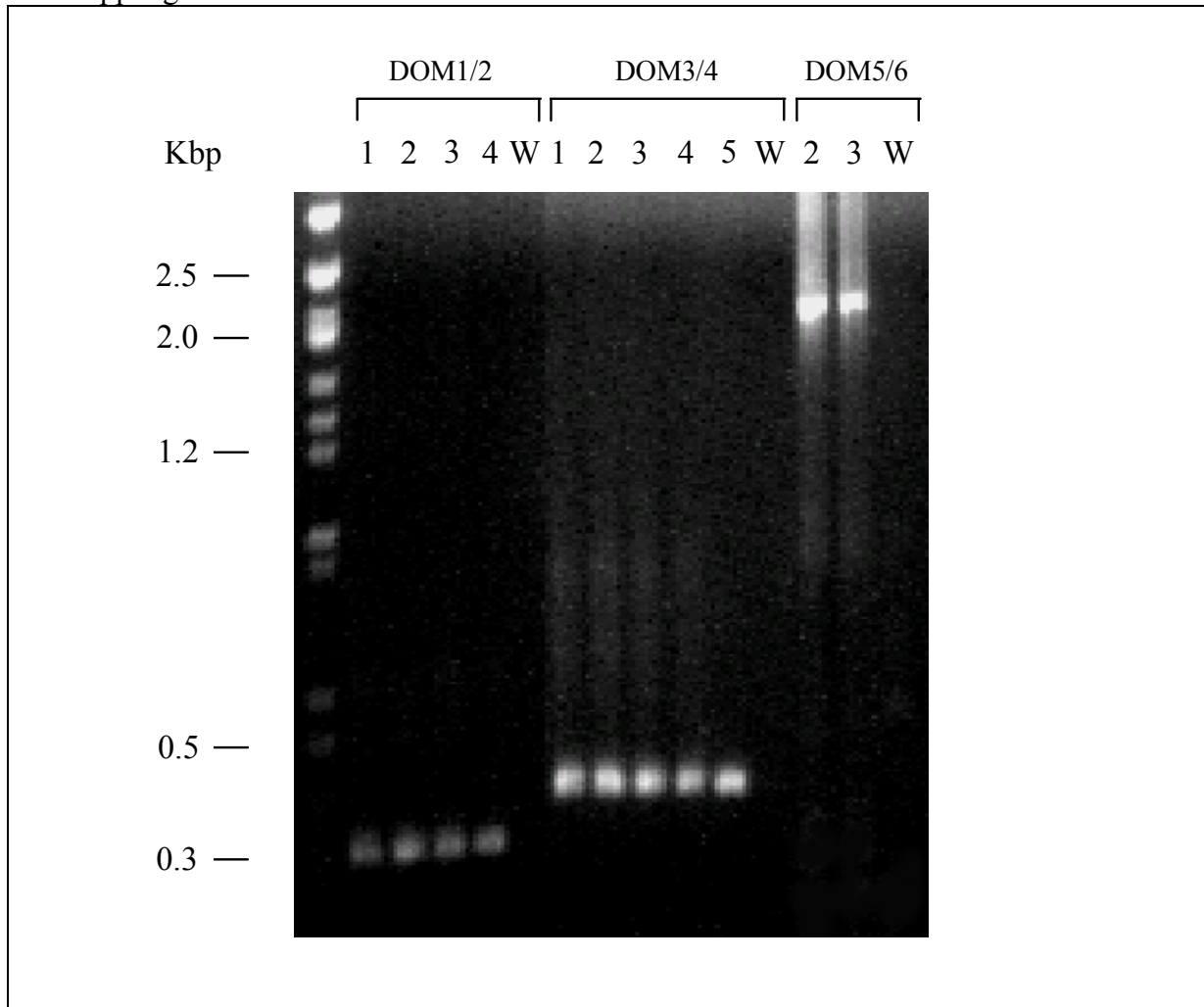


Fig.C11: Amplifikation der genomischen Bereiche zwischen *CaRep1*, 2 und 3.

Ethidiumbromid-gefärbtes Agarosegel mit den Amplifikaten der Primerpaare DOM1/2, DOM3/4 und DOM5/6. Als Templat diente genomische DNA aus *C. arietinum* ILC 482 (1), *C. arietinum* ILC 3279 (2), *C. arietinum* ILC C96 (3), *C. reticulatum* PI489777 (4) und das isolierte Plasmid aus pCaEr638 (5). W= Wasserkontrolle. Die Amplifikate liegen jeweils im erwarteten Größenbereich von 284 Bp (DOM1/2), 378 Bp (DOM3/4) sowie 1971 Bp (DOM5/6).

mit *CaRep1*-und -2-Sequenzen als auch die Position der homologen Abschnitte im *del1*-Leseraster beweisen, daß tatsächlich Sequenzen zwischen *CaRep1* und *CaRep2* aus genomischer DNA amplifiziert wurden.

Die Sequenzen der mit DOM5/6 amplifizierten und klonierten Fragmente (DOM5-1, 5-2, 5-3) waren 1.965-1.966 Kbp lang, enthielten 56-58 % AT-Positionen und waren auf Nukleotidebene zu 93-96 % homolog (AJ411809-AJ411811). Alle drei Fragmente enthielten ein durchgehendes ORF, das für 655 Aminosäuren codiert, welches jedoch in zwei Klonen 1-3 Leserasterverschiebungen oder Stopcodons aufwies. Die Homologie zwischen den ORFs der drei Fragmente betrug 95-97 %. Auch für diese Amplifikatgruppe konnten Homologien

zu denselben pflanzlichen *Ty3-gypsy*-Retrotransposons gefunden werden, die schon durch die *CaRep1-3*-Sequenzen detektiert worden waren, und zwar sowohl auf Nukleotidebene als auch für die dem wahrscheinlichsten Leseraster entsprechenden Aminosäuresequenz. Die Ähnlichkeit der Aminosäuresequenzen zum Leseraster von *del1* (*Lilium henryi*) betrug 68-76% und deckte ohne größere Deletionen oder Insertionen den erwarteten Bereich zwischen 4.641-6.605 Kbp ab, der die konservierten Domänen von RNAse H und Integrase sowie ein dazwischenliegendes Zinkfingermotiv umfaßt (Fig.C10,C12). Die Randbereiche aller Amplifikate überlappten mit den Sequenzen aus *CaRep2* und 3, an deren Enden teils degenerierte, teils intakte *EcoRI*-Schnittstellen lagen. Die Sequenz, die diese Schnittstellen in *C. arietinum* voneinander trennte, war 1.798 Kbp lang, was etwa dem Abstand der entsprechenden Regionen in *del1* (1.804 Kbp) entspricht (Fig.C10). Die Homologie der Überlappungen auf Nukleotidebene betrug 96 % zu *CaRep2* und 92-99 % zu *CaRep3*. Die Position homologer Bereiche in *del1* und die Überlappungen mit *CaRep2* und 3 zeigen, daß auch *CaRep2* und 3 im Genom benachbart lokalisiert sind.

Durch die Isolation der entsprechenden Zwischenstücke wurde bestätigt, daß *CaRep1*, 2 und 3 verschiedene Abschnitte einer hochabundanten dispersen Sequenzfamilie repräsentieren. Sie stellt Kopien eines Retrotransposons des *Ty3-gypsy*-Typs im Kichererbsengenom dar und wurde folglich *CaRep* genannt. Zusätzlich ließ sich auch die gesamte Länge des Polyproteinleserasters von *CaRep* abschätzen. Dieses erstreckte sich über 4.415 Kbp und war damit um 92 Bp länger als der entsprechende Abschnitt von *del1* (4.323 Kbp). Die mit dem Startcodon ATG beginnende und dem Stop-Codon TAG endende Sequenz codierte für ein 1470 Aminosäuren langes Produkt (*del1*: 1461 Aminosäuren). Das konservierte RNA-Bindemotiv des *gag*-Leserasters befand sich 88 Codons oberhalb der Proteasedomäne. Fügte man die einzelnen sequenzierten Abschnitte zusammen, so ließ sich eine zusammengesetzte Sequenz für weite Teile der Wiederholungseinheit ermitteln, die sich über 6.061 Kbp von Bereichen 5' der „Primer“-Bindestelle bis zur Region 3' des Polypurintrakts erstreckte und vermutlich Beginn bzw. Ende der flankierenden langen terminalen Wiederholungseinheiten („long terminal repeats“, LTRs) umfaßte (Fig.C10). Die Reihenfolge der einzelnen Domänen des Polyproteins entsprach mit „*gag*-Protease-RT-RNAse H-Integrase“ der eines *Ty3-gypsy*-Elements (Fig.C12). In den genannten Bereichen betrug die Ähnlichkeit zu *del1* (*Lilium henryi*) 42-75 %. Im Bereich des Polyproteins befanden sich drei Leserasterverschiebungen (Position 292, 555 u.1426; Fig.C12) und ein Stopcodon (Position 662; Fig.C12).

Bei der zusammengesetzten Teilsequenz muß beachtet werden, daß sowohl die aus der genomischen Bank isolierten Teilbereiche *CaRep1-3* als auch die mit PCR ergänzten Regionen (DOM1, DOM5) aus verschiedenen Kopien des Elements stammen können. Etliche der erfaßten Kopien sind offenbar nicht funktionell, wie aus zahlreichen Leserasterverschiebungen und Stopcodons im ORF hervorgeht.

Fig.C12: Zusammengesetzte Teilsequenz des Retrotransposons *CaRep* aus *C. arietinum*. Sequenzvergleich der abgeleiteten Aminosäuresequenz aus den zusammengesetzten Teilstücken von *CaRep* mit der homologen Sequenz von *del1* aus *L. henryi* (Smyth et al., 1989). Typische Domänen für das *gag*-Polyproteins (RNA-

C.I.2.4. Sequenzierung der langen terminalen Repeats (LTRs) von *CaRep*

Da die zusammengesetzte Teilsequenz über das offene Leseraster des Polyproteins hinausreicht und vermutlich den Beginn des rechten LTRs sowie das Ende des linken LTRs enthält, eröffnete sich die Möglichkeit, die LTR-Sequenzen zu identifizieren und so die Gesamtlänge einer intakten *CaRep*-Kopie zu ermitteln. Die Identifikation von LTR-Sequenzen kann nicht über Homologievergleiche erfolgen, da diese Sequenzen nur Promotor- und Terminatorsignale für die Transkription des Elements, jedoch keine offenen Leseraster enthalten und zwischen verschiedenen Retrotransposonfamilien nicht konserviert sind (SanMiguel et al., 1998; Kumewaka et al., 1999). Der einzige Anhaltspunkt zur Erkennung von LTR-Sequenzen ist daher ihre je nach Grad der Divergenz mehr oder weniger starke Konservierung in den flankierenden Bereichen des Polyprotein-ORFs und der „Primer“-Bindestellen (PBS, PPT). Da zunächst keine genomische Bank mit großen Fragmenten von *C. arietinum* vorhanden war, wurden die *CaRep*-LTRs durch PCR-Amplifikation isoliert: unter der Annahme, daß die in der Gesamtsequenz vorliegenden Bereiche bereits den U5-Bereich des 5'-LTRs („linker LTR“) und den U3-Bereich des 3'-LTRs („rechter LTR“) enthalten, wurden aus diesen Bereichen Oligonukleotidprimer abgeleitet (LTRU3, LTRU5; Fig.C13). Dabei war der Primer aus dem U5-Bereich in Richtung des 5'-Endes des LTRs gerichtet

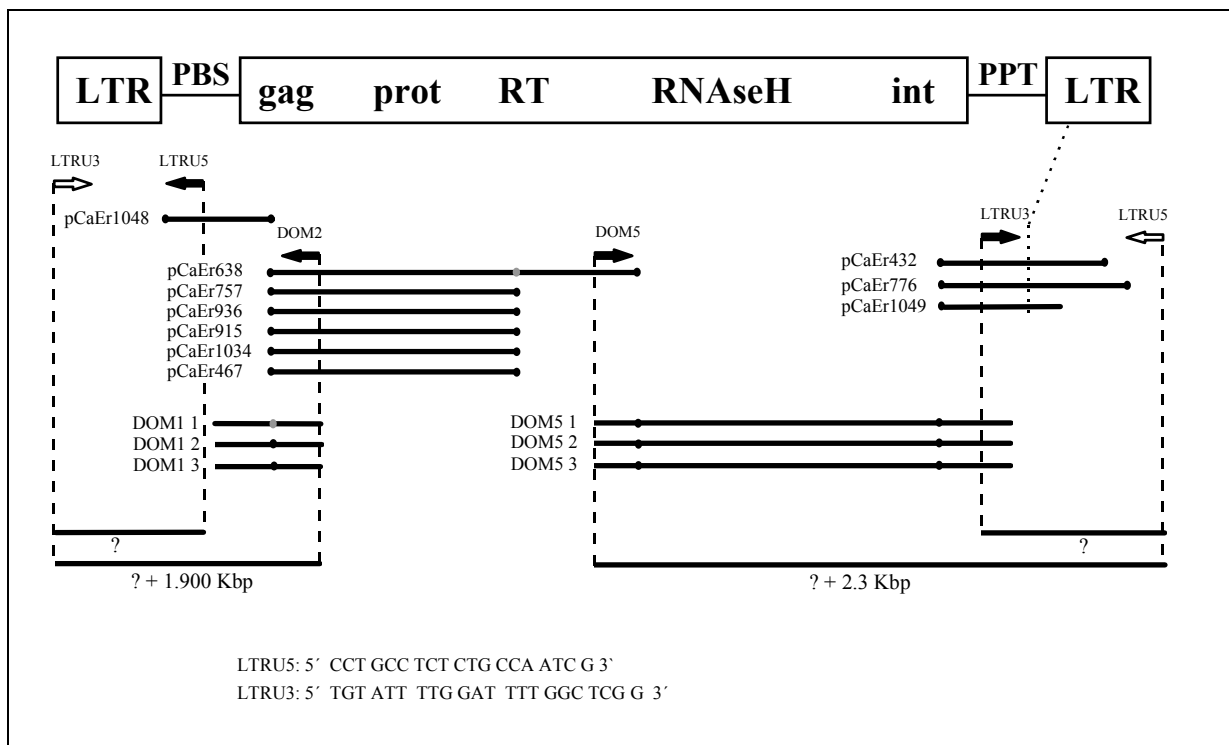


Fig.C13: Identifikation der langen terminalen Sequenzwiederholungen (LTRs) von *CaRep*.

Das Schema zeigt die Sequenz und Position der zur Amplifikation der LTR-Bereiche verwendeten Primer LTRU3 und LTRU5 in der Gesamtstruktur von *CaRep*. Die hypothetische Position der zweiten Oligonukleotidbindestelle innerhalb der direkten Sequenzwiederholungen ist mit einem weißen Pfeil

gekennzeichnet. Die Verwendung von DOM2 und DOM5 sollte ein um ca. 1.9 Kbp bzw. ca. 2.3 Kbp längeres Kontrollfragment liefern als die Kombination LTRU3/LTRU5 (schwarze Linien).

(LTRU5) und der aus dem U3-Bereich in 3'-Richtung (LTRU3) orientiert. Weiterhin wurde postuliert, daß die zum Oligonukleotiddesign genutzten Sequenzen aus LTR-Kopien stammen, die keine hochgradige Variabilität aufweisen, d.h. in einer Anzahl von identischen Wiederholungseinheiten im rechten **und** im linken LTR vorkommen. Nur unter diesen Voraussetzungen sollte eine PCR mit den Oligonukleotiden LTRU3/U5 erfolgreich sein und den gesamten LTR bzw. einen Großteil davon amplifizieren (Fig.C13). Die Amplifikate können hierbei aus dem rechten und linken LTR einer beliebigen Kopie von *CaRep* stammen. Um die Amplifikation von *CaRep*-unabhängigen Loci bzw. PCR-Artefakte auszuschließen, wurden die aus dem Inneren des Polyprotein-ORFs abgeleiteten Primer DOM2 und DOM5 (2.3.) in ein Kontrollexperiment miteinbezogen. Die Verwendung der Oligonukleotidkombination LTRU3/DOM2 sollte außer dem linken LTR einer Kopie noch den angrenzenden 5'-Teil des Polyprotein ORFs und die PBS amplifizieren, während DOM5/LTRU5 nur den rechten LTR und den 3'-Bereich des Leserasters liefern sollten.

Da die Länge der LTRs unbekannt war, oft aber mehrere Kilobasen umfaßte, wurde das „Expand High Fidelity PCR-System“ (Boehringer) zur Amplifikation verwendet. Als Templat diente genomische DNA aus *C. arietinum*. Mit der Oligonukleotidkombination LTRU3/LTRU5 entstanden Amplifikate von ca. 2.4 Kbp Länge, mit LTRU3/DOM2 war neben kleineren Artefakten eine deutliche Bande bei ca. 4.3 Kbp und mit LTRU5/DOM5 nach Herabsetzung der Annealing-Temperatur von 50°C auf 47°C eine Bande bei ca. 4.7 Kbp sichtbar (Fig.C14). Die Größendifferenzen der beiden Kontrollamplifikate zum LTRU3/LTRU5 Amplifikat mit ca. 1.9 Kbp (LTRU3/DOM2) und ca. 2.3 Kbp (LTRU5/DOM5) entsprach etwa dem Abstand des DOM2- bzw. DOM5-Oligonukleotids von der Bindestelle der jeweiligen LTR-Primer (Fig.C13).

Die Sequenz des LTRU3/LTRU5-Fragments war 2.433 Kbp lang, hatte einen AT-Gehalt von 62.64 % und besaß kein offenes Leseraster (AJ411812). Eine BLAST-Analyse konnte keine signifikanten Homologien zu anderen Sequenzen detektieren. Im Bereich des U5-Oligonukleotids und 428 Bp 5' davon bestand auf Nukleotidebene eine 86 %ige Homologie zum entsprechenden Anfangsbereich der *CaRep1*-Sequenz, die im Vergleich zum LTRU3/LTRU5-Amplifikat eine 11 Bp lange Deletion aufwies. Deutliche Homologien zu den *CaRep3*-Sequenzen erstreckten sich über die gesamte Länge des Fragments, jedoch wiesen pCaEr776 und pCaEr432 große Deletionen von 280 Bp und 840 Bp auf. Im 5'-Bereich beherbergten alle vier Sequenzen unterschiedlich lange, imperfekte Mikrosatelliten-ähnliche Motive wie z.B. (TA)_n und (GAAA)_n. Im mittleren Bereich des Fragments existierte eine Region mit mehreren, ca. 150 Bp langen Sequenzwiederholungen (Fig. C15a). Beides zusammen trug zur Divergenz zwischen LTR-Sequenzen verschiedener Kopien bei. Die Homologien bestätigen, daß der flankierende Bereich der codierenden, internen *CaRep*-Sequenzen amplifiziert wurde, der zum 5'- oder 3'-LTR gehört.

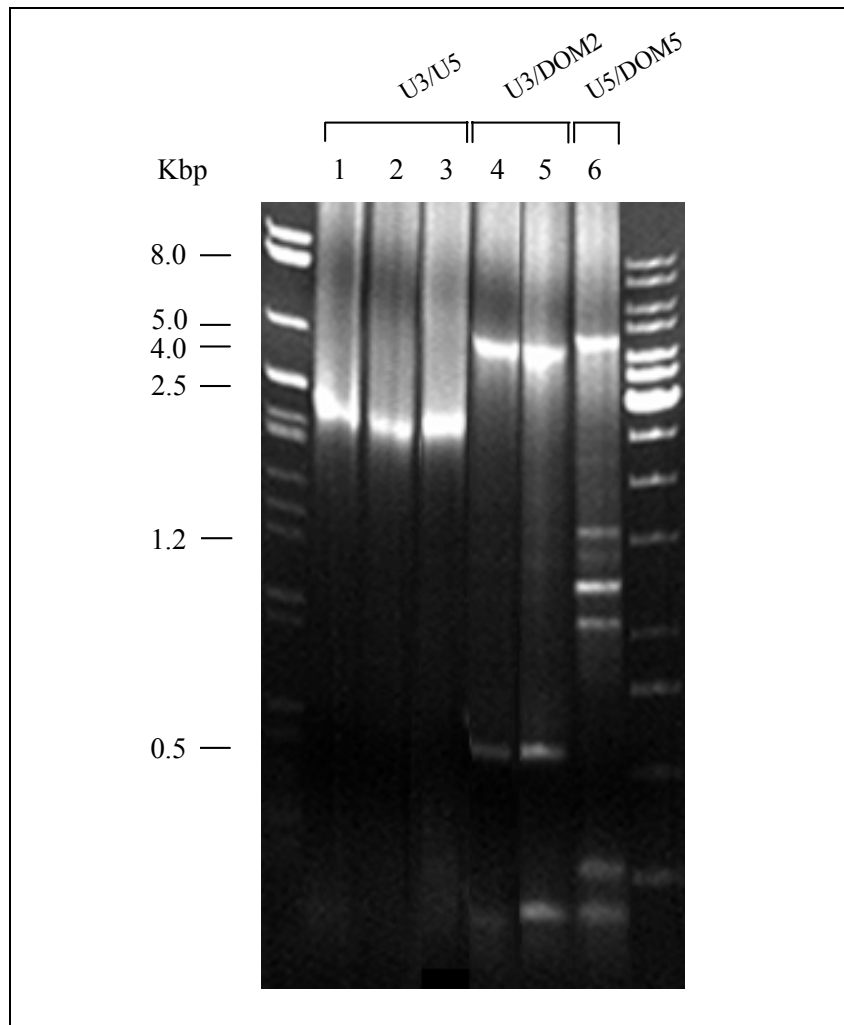


Fig.C14: Amplifikation putativer LTR-Bereiche aus *CaRep*.

Ethidiumbromid-gefärbtes Agarosegel mit Amplifikaten, die mit den Primerpaaren LTRU3/LTRU5, LTRU3/DOM2 und LTRU5/ DOM5 erzeugt wurden. Als Templat diente genomische DNA aus *C. arietinum* ILC 3279, in (2) DNA von *C. arietinum* ILC C96. Die Annealing-Temperatur betrug in (3), (5) und (6) je 47°C, in den übrigen Ansätzen je 50°C. Die Amplifikate der Kontrollexperimente zeigen die erwarteten Größendifferenzen.

Fragment LTRU3/DOM2 war 4.410 Kbp lang und hatte einen AT-Gehalt von 60.75 % (AJ411813). Ein offenes Leseraster mit Homologie zum *gag*- und Protease-Bereich in *del1*, das sich ohne Stop-Codon oder Leserasterwechsel bis zur DOM2-Bindestelle erstreckte, sowie Homologien zu *CaRep1* (87.3%), *CaRep2* und LTRU3/LTRU5 (90.5%) zeigen, daß der (5'-) flankierende Bereich der codierenden *CaRep*-Sequenzen amplifiziert wurde (Fig.C15a). Das Ende des LTRs mit dem charakteristischen Motiv 5'-ACA-3' und - im Abstand von drei Nukleotiden - eine potentielle „Primer“-Bindestelle für die Synthese des (-)-Strangs durch die Reverse Transkriptase (PBS) befanden sich 230 Bp stromaufwärts des Leserasterbeginns (Fig.C15b). Die PBS zeigte Homologie zu dem 12-20 Bp langen, konservierten 3'-OH Ende der tRNA_i^{Meth} aus *Lupinus luteus* (X06458) und *Phaseolus vulgaris* (K00318) bzw. anderen pflanzlichen tRNA-Sequenzen (Sprinzl et al., 1989).

Das Fragment LTRU5/DOM5 war 4.755 Kbp lang (AT-Gehalt: 60.32 %) und wies ein ORF auf, das ab der Position der RNase H Domäne Homologie zum Polyproteinleseraster von *del1* zeigt (zwei Stop-Codons, ein Leserasterwechsel; AJ411814). Die daran anschließende Sequenz zeigte sowohl Homologien als auch große Deletionen und divergente Bereiche zu *CaRep3* bzw. *CaRep1* und den LTRU3/LTRU5-Produkten. Dies weist darauf hin, daß der (3'-)flankierende Bereich zu den bekannten, internen *CaRep*-Sequenzen kloniert wurde (Fig.C15a). Kurz (22 Bp) nach dem Stop-Codon des Polyproteinleserasters war eine 12 Bp lange purinreiche Sequenz zu erkennen, die dem Polypurintrakt (PPT), der „Primer“-Bindestelle für die (+)-Strangsynthese, entsprach. Im Abstand von einem Nukleotid folgte hierauf die charakteristische Basensequenz 5'-TGT-3' am Beginn des LTRs (Fig.C15b). Alle drei Sequenzen zeigten außerhalb des ORFs keine signifikanten Homologien zu anderen Elementen. Da beide Primersequenzen innerhalb des LTRs lagen, d.h. Beginn des 5'- und Ende des 3'-LTRs unbekannt waren (Fig.C15a), konnte die idealisierte Gesamtlänge der *CaRep*-LTRs von 3.364 Kbp erst durch ein Alignment der rechten bzw. linken LTR-Sequenzen bestimmt werden (Fig.C15b). Die einzelnen Sequenzen enthielten Insertionen und Deletionen

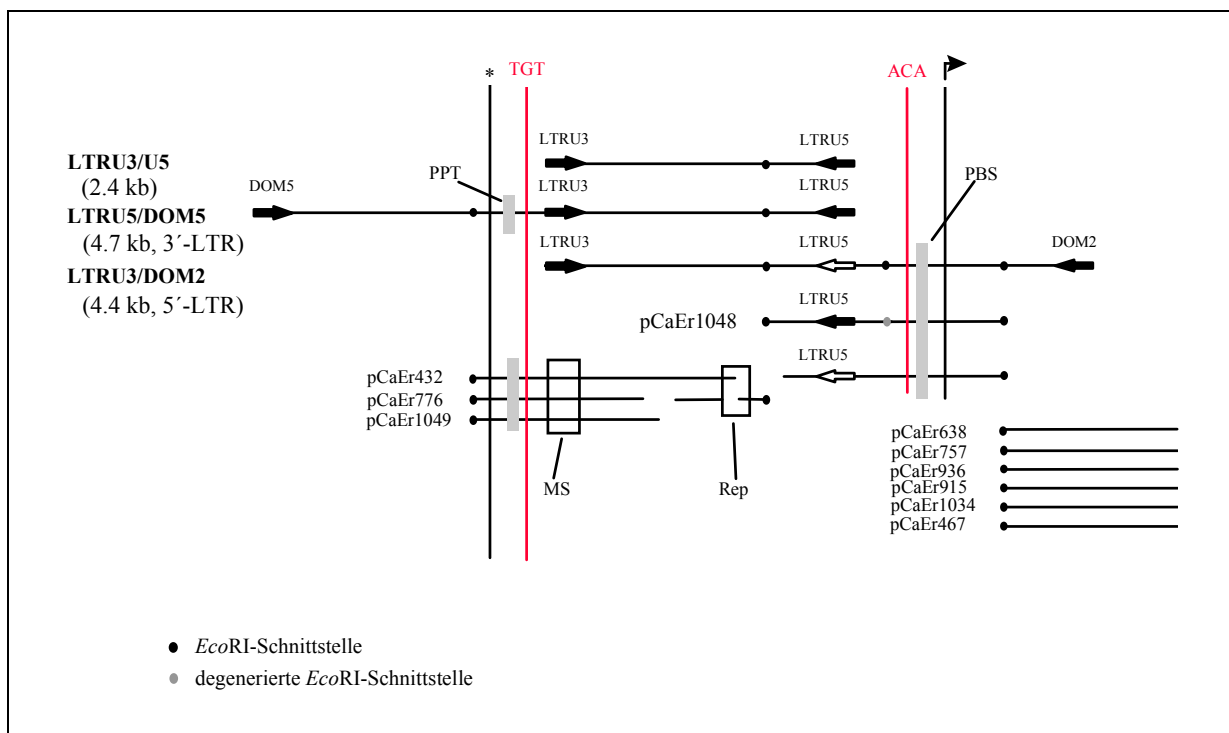


Fig.C15: Identifikation der LTR-Sequenzen von *CaRep*.

(a) Schematischer Vergleich aller Sequenzen aus dem LTR-Bereich (schwarze, horizontale Linien). Dargestellt sind die begrenzenden inversen Motive 5'-TGT-3' und 5'-ACA-3' (rote Linien), die flankierenden PPT- und PBS-Bereiche (graue Kästen), degenerierte und intakte Oligonukleotidbindestellen (weiße/schwarze Pfeile) sowie Beginn (➔) und Ende (*) des Polyproteinleserasters. Innerhalb der LTR-Sequenz befindet sich ein Bereich mit Mikrosatelliten-ähnlichen Motiven (MS) und Sequenzwiederholungen (Rep). Beginn des 5'- und Ende des 3'-LTRs wurden nicht amplifiziert. (b: folgende Seiten) Sequenzvergleich der klonierten Fragmente aus dem 5'-LTR (LTRU3/DOM2), dem 3'-LTR (LTRU5/DOM5) und des Fragments LTRU3/U5. Innerhalb der

C. Ergebnisse

LTR-Sequenz (grauer Kasten) ist das inverse, begrenzende Motiv (ACA bzw. TGT) hervorgehoben, flankierend sind PPT und PBS (fett), sowie Ende (fette*) und Beginn (➔) des Polyproteinleserasters dargestellt. Beginn des 5'- und Ende des 3'-LTRs werden erst durch den Sequenzvergleich ergänzt. [Teil2.doc](#)

3' LTR	PP1	60
	GACGCA* <u>TAG</u> TTTCGAGGACGAAACTAATTTT <u>AGGGCGGTAGTA</u> TGT AAGACCCATA	
LTRU3/U5 5' LTR 3' LTR	-----TGTATTTGGATTTGGCTCGGAGGCTTTTA -----TGTATTTGGATTTGGCTCGGAGGCTTTTA ATTTTAAAGTACCCTTTATGTATTTTGGTGTATTTGGATTTGGCTCGGAGACTTTT *****	120
LTRU3/U5 5' LTR 3' LTR	AGCCAAAGTTAATAATATTTGGAGTTT-ATAATCAAAAAGATATTTGATGCCAATTA TGCCAAAGTTAATAATATTTGGAGTTT-ATGATCAAAAAGTATTTCAATGCCAATTA AGCCAAAGTTAATGATATTTGGAGTTTATATCAAAAAGATATTTGATGCCGATTA *****	180
LTRU3/U5 5' LTR 3' LTR	AAATTTCTCGTCGATAAAATAAATGAATTTAACTGCGATGAATCCTTTACGGAGAATTTA GAATTTCTCGTCGATAAAATAAATGAATTTAACTGCGGTAATAAATTTACGGTTAATTTA GAATTTCTCGTCGATAAAATAAATGAATTTAAGTGCCTGAACTTTACGGAGAATTTA *****	240
LTRU3/U5 5' LTR 3' LTR	ATGCGTTTCGGGTGAAACGGTAATTAACAAATATCTAGAT--TTGAGATATTTGTTAA AGCGTTTCGGGTGAAACGGTAATTAATAAATATCTAGAT--TTGAGATATTTGTTAA ATACGTTACGGGTGAAATGGAATTAACAAATATCTAGATATTTGAGATATTTGTTAA * * * * *	300
LTRU3/U5 5' LTR 3' LTR	GTTATATTTATTATATTTATATATGT---TTGTTGGAGGAAAAAGAAAGAAAACTA GTTATATTTATTATATATATATATATATTTGCTTGGAGGAAAAAGAAAGAAAACTA GTTATATTTATTATATAGATATGT---TTGTTGGAGGAAAAAGAAAGAAAACTA *****	360
LTRU3/U5 5' LTR 3' LTR	GAAGGAAGGAAAAGGAAAAGAAAATAGTATA-AAAGGAAGGAAAAGGAAAAGGAAAAGG GAAGGAAGGAAAAGGAAAAGAAAATAGTATAAAAGGAAGGAAAAGGAAAAGGAAAAGG GAAGGAAGGAAAAGGAAAAGAAAAGAAAATATAAAAGGAAAAGGAAAAGGAAAAGG *****	420
LTRU3/U5 5' LTR 3' LTR	AAAA-CAAAGGAAAAGAAAAGGAAA-GAAAAACCATTTTCCATCGTCTCCTTCC AAAA-CAAAGGAAAAGAAAAGGAAAAGGAAAAGGAAAAGGAAAAGGAAAAGGAAAAGG AAAAACACAGGAAAAGAAAAGGAAAAGGAAAAGGAAAAGGAAAAGGAAAAGGAAAAGG *****	480
LTRU3/U5 5' LTR 3' LTR	TCTGTC-CGCGG-----CCAATTTCTCCTCTTCTCCCATTTTCATTTTCGTCGCTTT TCAGTCACGCAAGAAAACCAAAATTCCTCCTCCTCCTTTTCGTTTCGTTGCTTTC TCTGCCACCGCCG---ACCCATTTTCTCCTCTTCTCCCATTTTCATTTTGTGCTTT * * * * *	540
LTRU3/U5 5' LTR 3' LTR	CCTTCTCGAACTAGTAACCCAAGGTGGGTGGAGTTAGATCAAGGTTTCGCAACT CTTCCCTCAAACCTAGTGACCCAAGGTGGGTGGAGTTAGATCAAGGTTTCGTAAC CTTCTTCAAGACTAGTAACCCAAGGTGGGTGGAGTTAGAACCAAGGATCTCTCGACT * * * * *	600
LTRU3/U5 5' LTR 3' LTR	CCACTCTTCTCTTTGATTCGAAAACGAAGAAAACGGTTTTTGAGATCCAACACAAAAC CCACTCTTCTCTTTGATTCGAAAACGAAGAAAACGGTTTTTAAGATCCAACACAAAAC CCACTCTTCTCTTTGATTCGAAAACGAAGAAAACGGTATTAGAGATCCAACACAAAAC *****	660
LTRU3/U5 5' LTR 3' LTR	CCCTCCGTTTCTCTAGATCCAACCTCATTTCGCGATGAGTCAAGGGAAGGTTGCTC CCCTCCGTTTCTCTAGATCCAACCTCATTTCCTCAAGAGATAGAAGGAAAAGTTGCTC CCCTCCGTTTCTCTAGATCAAGCTTCATTCGCAAGAGTTAGAAGGAAAAGTTACTC *****	720
LTRU3/U5 5' LTR 3' LTR	ACTACCACGCCACGGTGCTAGGGGACCTATTTGGAGGCGAAGTTGCGAGCGGAGATTTTA ACCTCCGTTGCTACAATGCTAGTGCCTAATTTGGAAGCGCGCTTCCGACGGGAGATTTTA ACTACCACGCTACGGTGCTAGGGGACCAATTTGGAGGAAACGTTACGAGCGAAGATTTTA * * * * *	780
LTRU3/U5 5' LTR 3' LTR	CGGAATTTACTCGCGTACCAGAAACGCACGTTAAAGCTAACGTTGAGGTAAGGACTCCT CGGAATTTACTCGCGGACCGGAAACCTCAGATAAAGCTAACGTTGAGGTAAGGACTCCT CGGTTTACTCGCGGTACCGGAAACCCACGTTAAAGCTAACGTTAAGGTAAGGACTCCT ***	840
LTRU3/U5 5' LTR 3' LTR	TCCAACTTCTAGTTGCATT-AGGGACTTATCTGTGTTGTGTGAAAGGAATTTGTTA TCCAACTTTTAGTTGCATTAGGGACTTATCTGTGTTGTGTGAAAGGAATTTGTTA TCCAACTTTTAGTTGCATTAGGGACTTATCTGTGTTGTGTGAAAGGAATTTGTTA *****	900
LTRU3/U5 5' LTR 3' LTR	GGGTTAATGTATCGATTTGGGAAAAACGAACTAATGCGTTTGGGTAACCTTAG GGGTTAATGTATCGATTTGGGAAAAACGAACTAATGCGTTTGGGTAACCTTAG GGGTTAATGTATTTAGGAAAAAT--GAATCTAAGGTTCTGTGGTAAACCTTAG *****	960
LTRU3/U5 5' LTR 3' LTR	AGTTAGGTTTGTATATGATGAATGTTTCGTGTAATGAATTTGAA-GTCATTGTG AGTTAGGTTTGTATATGATGAATGTTTCGTGTAATGAATTTGAAATGTCATTGTG AGTTAGGTTTGTACGATATGATGAATGTTTCGTGTAATGAATTTGAACTACTTTGTG *****	1020
LTRU3/U5 5' LTR 3' LTR	TGTGATTATGAGTAAATGAAATTTGATGATCTGGGCGTTATGTTG-CGTTGATTTGTG TATGATTATGAGTAAATGAAATTTGATGATCTGGGCGTTATGTTGACTTTGATTTGTG TATGATTATGAGTAAATGAACTCTGATGATTTGGGCGTTATGTTG-TGTTGATTTGTG *****	1080
LTRU3/U5 5' LTR 3' LTR	TCGTTGATTTGGCGTCCATACTTGATGTTGTTAATGGTGGTGGTT-GTTGTGAAT TCGTTGATTTGGCGTCCATACTTGATGTTGTTAATGGTGGTGGTT-GTTGTGAAT TCGTCGATTTGGCGTCCATACTTGTTGTTGTTGATGATGATTTGGTGGTGAAT *****	1140

C15b (1)

LTRU3/U5 5' LTR 3' LTR	ATGGTTTGAATGTGAGAGT-----TTGTTTCCGTGGAATTTGAAAATCATTAGAG ATGGTTTGAATGTGAAAGTATGTTTGAATTTGTTTCTGTGGAATTTGAAAACCATAGAG ATGGATTGAATGTGAGAGTATATTGAGTTGTTTCTGTGGAATTTGAAAACCATAGAG *****	1200
LTRU3/U5 5' LTR 3' LTR	T-AAACGAGTATTAATAATGGGAATCTTTAATACCTCGGTTTACTTAATGTGCATTG TAAACGAGTATTAATAATGGGAATCTTTAATACCTCGGTTTACTTAATGTGTATTG TAAACGAGTATTAATAATGGGAATCTTTAATACCTCGGTTTACTTAATGTGCATTG *****	1260
LTRU3/U5 5' LTR 3' LTR	AGGAAAAGTTTAAAGTTAACTGGGCGTTGAAAATGAGGAATCTCTTAATGTCCGGTTAC AGGAAAATGTTTAAAGTTAACTGGGCGTTGAAAATGAGGAATCTCTTAATGTCCGGTTAC AGGAAAATGTTTAAAGTTAACTGGGCGTTAAGAAATGGGAATCTCTTAATGTCCGGTTAC *****	1320
LTRU3/U5 5' LTR 3' LTR	TTAACTATCGTGTGAAAATCGTTCGGTAAATGAGTATTAGAATGGGAATCTCTTA TTAACTATCGTGTGAAAATCGTTCGGTAAATGAGTATTAGAATGGGAATCTCTTA TTAACTATCGTGTGAAAATCGTTCAGTAAATGAGAAATAGAATGGGAATCTCTTA *****	1380
LTRU3/U5 5' LTR 3' LTR	ATGACTCGTTTACTGAATGTTTGTGAGAAAATCGTTTAAAGTCAAAGTATATTAGAAT ATGACTCGTTTACTGCATGTTTTCGAGAAAATCGTTTAAAGTCAAAGTATATTAGAAT ATGCTCTGTTTACTGAATGTTTGTGAGAAAATCGTTTAAAGTCAAAGTATATTAGAAT *****	1440
LTRU3/U5 5' LTR 3' LTR	GGGGAATCTCTTAATATGCTGTTTGTCTTAAACATTTGTTTTGAAAATCATTAAAGATAAA GGGGAATCTCTTAATATGACTGTTTGTCTTAAACATTTGTTTTGAAAATCATTAAAGTAAA GGGGAATCTCTTAATATGCTGTTTGTCTTAAACATTTGTTTTGAAAATCATTAAAGATAAA *****	1500
LTRU3/U5 5' LTR 3' LTR	TCGATATTAGAACGGGAATCTCTTAATGACTTATTTAATTAATGTTGTTTGAAGTC TCGGTATTAGAACGGGAATCTCTTAATGACTTATTTAATTAATGTTGTTTGAAGTC TCAGTATTAGAATGGGAATCTCTTAATGACTTATTTAATTAATGTTGTTTGAAGTAC *****	1560
LTRU3/U5 5' LTR 3' LTR	GTTTAAAGTTAAATGGGTATTAAAAATGGGAATCTTTAATATCTGCATTTGCTTAAACG GTT--AAGTTAAATGGGTATTAAAAATGGGAATCTTTAATATCTGCATTTGCTTAAACG GTTT--AAGTTAAATGGGCGTTAAGAAATGGGAATCTCTTAATGCTTGGTTA--CTTAATG *****	1620
LTRU3/U5 5' LTR 3' LTR	ATTGTTGTAATGGAGTCTGTGTATGCTCCTGCATTTTCAATTTGGTAAATGTGTCGACCC ATTGTTGTAATGGAGTCTGTGAATGCTCATGCATTTTCAATTTGGTAAATGTGTCGACCC TTGGTTGTAAGGTCTTTAAGTTAAATGG--GTATTAATAATGGGAATCTTTAATATC *****	1680
LTRU3/U5 5' LTR 3' LTR	GTGATAGGTGACACCTTGGTAAACTGTACTGACCCGTGATAGTGGTACGTTTACGATTT GTGATAGGTGACACCTTGGTAAACTGTACTGACCCGTGATAGTGGTACGTTTACGATTT TGCAATTTGCTTAAACGATTGTT--GTGATGGAGTC-TG-----TGATGCTCAGCCTT *****	1740
LTRU3/U5 5' LTR 3' LTR	ACATTTTGGTAAATCGTGTGACCCGTGATAGGTGACACCTTGGTAAACTGTACTGACC ACATTTTAG--TAAATCGTGTGACCCGTGATAGGTGACACCATGGTAA--CTGTACTGACC TCATTTTGG--TAAATCGTGTGACCCGTGATAGGTGACACCTTGGTAAACTGTACTGACC *****	1800
LTRU3/U5 5' LTR 3' LTR	CGTGATAGGTGACACCTTGGTAAACTGTACTGACCCGTGATAGTGGTACGTTTACGATTT CGTGATAGGTGACACCTTGGTAAACTGTACTGACCCGTGATAGTGGTACGTTTACGATTT CGTGATAGGTGACACCTTGGTAAACTGTACTGACCCGTGATAGTGGTACGTTTACGATTT *****	1860
LTRU3/U5 5' LTR 3' LTR	TGGCACCTCGGTAACAGTA----- TGGCACCTCGGTAACAGTA----- TGGCACCTCGGTAACAGTA----- *****	1920
LTRU3/U5 5' LTR 3' LTR	-----CTGGTCTGTGATAGG AGTAAATCGTGTGACCCGTGATAGGTGACACCATGGTAACTGTACTGACCCGTGATAGG -----TTGGTCTGTGATAGG *****	1980
LTRU3/U5 5' LTR 3' LTR	CGGTACGTTTACGATTACGCTTTTCTTTAGTAAATCGTGTGACCCGTGATAGGTG TGGTACATTTACGATTACGCTTTTCTTTAGTAAATCGTGTGACCCGTGATAGGTG CGGTACGTTTACGATTACGCTTTTCTTTAGTAAATCGTGTGACCCGTGATAGGTG *****	2040
LTRU3/U5 5' LTR 3' LTR	CACCTCGGTAACACTGTACTGACCCGTGATAGGTGGTACGTTTATGATTTACGCATTTTAC CACCTCGGTAACACTGTACTGACCCGTGATAGGTGGTACGTTTATGATTTACGCATTTTAC CACCTCGGTAACACTGTACTGACCCGTGATAGGTGGTACGTTTATGATTTACGCATTTTAC *****	2100
LTRU3/U5 5' LTR 3' LTR	TAAATCGTGTGACCCGTGATAGGTGGCACCTCGGTAATGGTACTTTGGCCTGCGATAG TAAATCGTGTGACCCGTGATAGGTGGCACCTCGGTAATGGTACTTTGGCCTGCGATAG TAAATCGTGTGACCCGTGATAGGTGGCACCTCGGTAATGGTACTTTGGCCTGCGATAG *****	2160
LTRU3/U5 5' LTR 3' LTR	GCGGTACAATTATGATTTACGGCCCTTCGAGGAGGTTTGGTTTGGAAATCCGAGTCCA GCGGTACAATTATGATTTACGGCCCTTCGAGGAGGTTTGGTTTGGAAATCCGAGTCCA GCGGTACAATTATGATTTACGGCCCTTCGAGGAGGTTTGGTTTGGAAATCCGAGTCCA *****	2220
LTRU3/U5 5' LTR 3' LTR	TGCATTTTGGCATATACGCATTGCATTAGGGTGCCTGGCACCGTGCATGTTTATTGATTCA TGCATTTTGGCATATACGCATTGCATTAGGGTGCCTGGCACCGTGCATGTTTATTGATTG TGCATTTTGGCATATACGCATTGCATTAGGGTGCCTGGCACCGTGCATGTTTATTGATTG *****	2280

C15b (2)

C. Ergebnisse

LTRU3/U5	AGTTTATGATTGAGTGTGTTTGAATTT-GAGTTGTCGTGGTAATTGTGTTGAAATGCTAA	2340
5' LTR	AGTTTATGATTGAGTGTGTTTGAATTTGATTATCGTGATAACTGAGATGAAAGTGTAA	
3' LTR	AGTCTATGATTGAGTGTGTTTGAATTTGATTATCGTGTAATCGCGTTGAAGTGTAA *** ***** ** * * * * * * * * * * * * * * *	
LTRU3/U5	GTGTTATGATTGATTAGCATGTGCGCGTGCATGGATTGCAAACTATTTGCAAAACCA	2400
5' LTR	GTGCTATGTTGTGATTGTTGTTGAGTATGATGGTTATCGAACCTTGTGT-----	
3' LTR	GTGTTATGTTAATTATGTTGTTGTAAGTATGATGGTTTTCGAAATCCCATT----- *** ***** ** * * * * * * * * * * * * * * *	
LTRU3/U5	ATTTCGTCGTGGTATTGTGTGGGTTTTGCTAAGTGTAAAGTTTGGAGTTATGTGTGAGT	2460
5' LTR	----GTCGTAGTAAATCGGTAGAAATGCTAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAGT	
3' LTR	----GCCGTGTAATCGGTAGGAAATGCTAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAGT *	
LTRU3/U5	GTGATTGAATTAGTTTAACTATTTTGGAAAGATAAGCAGGAAATCGATTTCCCAATCG	2520
5' LTR	GAC-TTAGGTTAAATCATGAATTTTGGAAACTGATCAGGAAATCGATTTCCCAATCG	
3' LTR	GAC-TTAGGTTAAATCATGAATTTTGGAAAGATAAGCAGGAAATCGATTTCCCAATCG *	
LTRU3/U5	ATTGGATGAGTTG-CAGGAAGTTCACATTTTAACTAATCGATTTGCCAATCGATTGTA	2580
5' LTR	ATTGGATGGTAAACAGGACTTGGCCAAATCA-CAAAATCGATTAGCCAAATCGATT-TC	
3' LTR	ATTGGATGAGTTG-CAGGAAGTTCACATTTTAACTAATCGATTTGCCAATCGATTGTA ***** ** * * * * * * * * * * * * * * * * * * *	
LTRU3/U5	TAAATATATCTTCAA--AATCTTTAAGAAATCGATTTGGAAATCGATT-GGCAGAGAG	2640
5' LTR	GAAATCAATTTCAAAGGAATCAGTTAAGAAATCGATTTGCCAATCGATTAGGCCTTAAG	
3' LTR	TAAATATACCTTCAA--AATCTTTAAGAAATCGATTTGGAAATCGATT-GGCAGAGAG ***** * * * * * * * * * * * * * * * * * * *	
LTRU3/U5	GCAGG-----	2700
5' LTR	TCAGGGCTCAGGAACCAATCAATCGATTTACCAATCGATTGAATTACGCCAGGATTC	
3' LTR	GCAGG----- ****	
5' LTR	GTTTTTATTAAAAATCTTACCCCAATCGATTGCCAATCGATTGGCTCAGTGAAAAATCC	2760
5' LTR	TCATTTTAGTTGATGATTAATGCTCATGAATCTATCCATGCCCTGTTTATCCTGTG	2820
5' LTR	TTAATTAGGAGATCGAGAACTGCATTTTACCTTCATTTTCATTGGCAATGTAATGTATGA	2880
5' LTR	ACTTTTCGCACATTGAAAATCCTGTTAAGGTGCATGCTTAGTTAAAAAGTTGTTTTGAGC	2940
5' LTR	ATTCAAAAACCTAATGCTATGTGTTATCTGTTATTTCTGGTGGTGACCCCTTACACT	3000
5' LTR	ATTGTGGAATCTGGGCTTTGCCCTCAGATGAGAGTCAGGACGGCCCTACTGGTTCGTAC	3060
5' LTR	CCTACGGACAGGGATGGAGATGGGAATGCTTGACTGCGGTTGTGTTAGGAGATCTCAGC	3120
5' LTR	GGGCGCTGGAGATCTCAGGGTGTATAGTTTTTGGTAGGATGATCAGATTAGGATGAT	3180
5' LTR	GTATAGGACTAGGGCTCTCTTTTGGTTGGAGTATTTTAGTTTGGAAAACGTAC	3240
5' LTR	TTATACAAAATATTATCAGTTTATATCATCTTTGAATGGGTTCCATGTACCATTGTATGT	3300
5' LTR	TTTTGTAGAAATGCTTTGGATTGTAATGGAGAACTTTCCGCTGCTTGTAAATATA	3360
5' LTR	ATGACTCAATTTATCCAAAAGCATTTCCTGATTTATTTCTTGTTCGTTTTTAATTA	3420
5' LTR	CTTTTAAAAAAAAAAAAATATACCCTCGCTTTGAAAACGGGTTT ACAA TTG GG	3480
5' LTR	PBS TATCAGAGCATAGGTT TAGTTTCTTTGGGAGCTGTGGAACCTGGTTATAGATTGTAGG	3540
5' LTR	TCGACCTATAAGTTAGGAGTTGTCATCTGATCATGTGTCGGTTTAGTGACTTGAGTTGT	3600
5' LTR	CAAGTCCAAATTAATGTTGTTAGACGGACCTAGTTTAGAATTTTGAAGTAGT	3660
5' LTR	GTAAAACCTACTTGATGATGGTTCTTATAGGAATCC ATG CCGCC	3707

C15b (3)

(Fig.C15b) und waren zueinander 86-91% homolog. Ein TATA-Box-ähnliches Motiv und das Polyadenylierungssignal 5`-GATAAA-3` konnten in einem Teil der Sequenzen gefunden werden (Fig.C15b).

Schließt man größere Deletionen in den LTR-Sequenzen aus, ergibt sich mit dem 4.415 Kbp langen Polyproteinleseraster, einer 226 Bp langen 5`-untranslatierten Sequenz und 35 Bp im 3`-untranslatierten Bereich eine Gesamtlänge von 11.404 Kbp für eine gesamte *CaRep*-Kopie. Die idealisierte, zusammengesetzte Gesamtsequenz ist in Fig.C16 dargestellt.

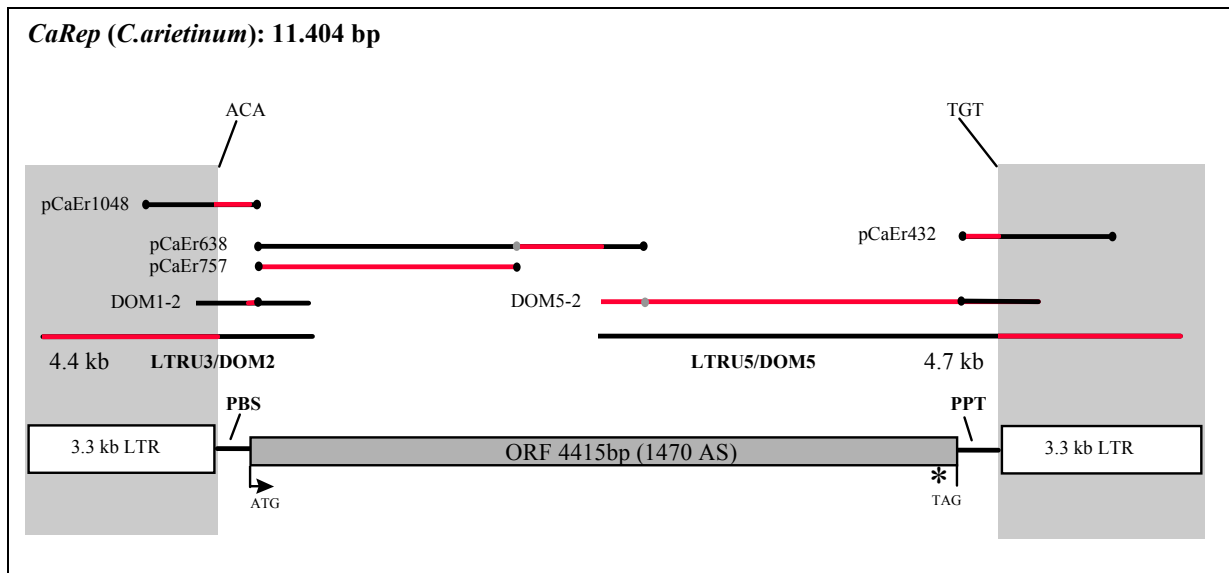


Fig.C16a: Die zusammengesetzte Gesamtsequenz von *CaRep*.

(a) Schematische Übersicht über die klonierten Einzelsequenzen und Auswahl der Teilbereiche (rote Linien) für die zusammengesetzte Gesamtsequenz in (b).

Fig.C16b: Die zusammengesetzte Gesamtsequenz von *CaRep*.

(b) Die 11.404 Kbp lange, zusammengesetzte Gesamtsequenz mit beiden LTRs (hellgrau unterlegt), PBS und PPT (dunkelgraue Kästen) und Aminosäuresequenz des wahrscheinlichsten Polyproteinleserasters, dessen

C.I.2.5. Chromosomale Lokalisation der *CaRep*-Elemente

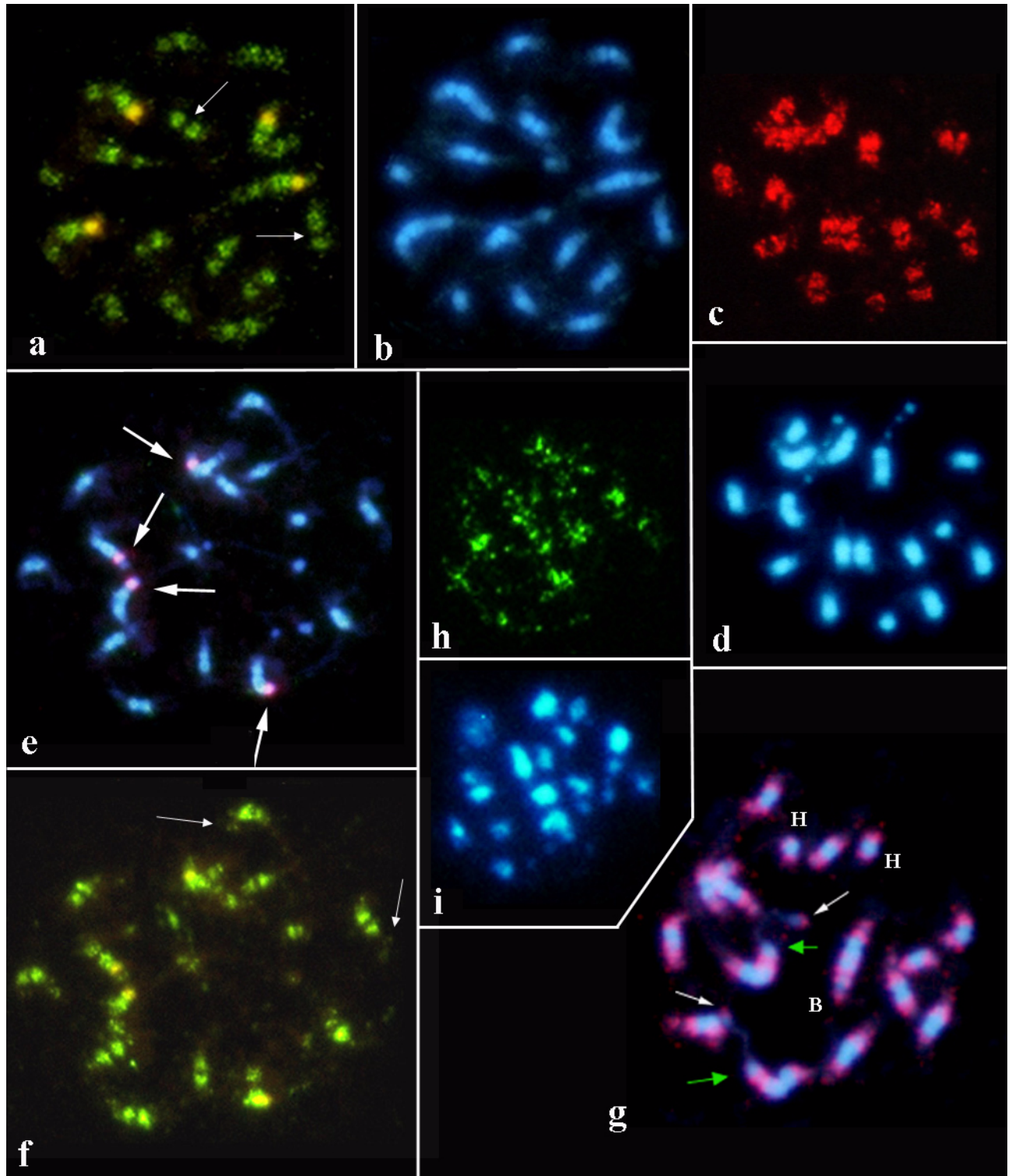
***CaRep*-Sequenzen sind im pericentrischen Heterochromatin aller Chromosomen zu finden**

Für die FISH-Analyse wurden klonierte Sequenzen der Homologiegruppen *CaRep1* (pCaEr1048; 2.140 Kbp), *CaRep2* (pCaEr936; 956 Bp) und *CaRep3* (pCaEr776; 2.058 Kbp) als Sonden verwendet. Alle drei Sonden erzeugten sehr ähnliche, nahezu uniforme Hybridisierungssignale von hoher Intensität, die auf allen Chromosomen größtenteils im pericentrischen heterochromatischen Block lokalisiert waren (Fig.C17). Die starke Signalintensität, die auch bei den dispersen Signalen im Interphasekern zu beobachten war, weist auf die hohe Abundanz aller drei *CaRep*-Abschnitte im Kichererbsengenom hin (Fig.C17h,i). Die Signale von *CaRep1* (Fig.C17 e,f) und *CaRep3* (Fig.C17 c,d,g) erstreckten sich an einigen Stellen über den Heterochromatinblock hinaus in die angrenzenden euchromatischen, nur schwach DAPI-gefärbten Bereiche (Fig.C17 e, Pfeile), während *CaRep2* hauptsächlich auf das Heterochromatin beschränkt war (Fig.C17 a,b). Auf Chromosom A war ein dritter Bereich mit *CaRep*-Sequenzen auf dem Satellitenanhängsel erkennbar (Fig.C17g, weiße Pfeile). Keine oder nur schwache Signale konnten an den Centromeren und ihrer näheren Umgebung detektiert werden (Fig.C17a,b, Pfeile), so daß man davon ausgehen kann, daß *CaRep*-Sequenzen größtenteils von diesen, vermutlich mit *CaSat2*-Sequenzen besetzten Regionen ausgenommen sind. Dies schließt jedoch eine Durchmischung der *CaSat2*-Loci mit *CaRep*-Sequenzen in den distaleren Regionen des Heterochromatinblocks nicht aus. Auf manchen der größeren Chromosomen konnten im Heterochromatin deutliche „Cluster“ von *CaRep*-Signalen festgestellt werden, während dazwischenliegende Bereiche ausgespart waren (Fig.C17g; „B“). Ebenso war die Intensität der Signale auf beiden Seiten des Centromers z.T. unterschiedlich (z.B. kleinstes Chromosom, Fig.C17g; „H“). Unter stringenten Hybridisierungsbedingungen zeigte eine Zweifarben-FISH von *CaSat1* mit *CaRep1* (Fig.C17 e,f) bzw. *CaRep2* (Fig.C17 a,b), daß die Signale dieses Satellitenelements nicht mit denen von *CaRep* überlappen, sondern distale Bereiche des Heterochromatinblocks auf Chromosom A

Fig.C17: Chromosomale Lokalisation der *CaRep*-Sequenzen auf Chromosomen von *C. arietinum*.

(a) Simultane Lokalisation von *CaSat1* (pCaEr1062; gelbes Signal) und *CaRep2* (pCaEr936; grün) auf Metaphasechromosomen. (b) Dieselbe Metaphase nach DAPI-Färbung (blaue Fluoreszenz). Sichtbar ist die Konzentration der Satellitensequenzen im distalen Teil des Heterochromatins sowie die Aussparung der Retrotransposonsequenzen vom centromerischen Bereich (weiße Pfeile). (c) Metaphase, hybridisiert mit *CaRep3* (pCaEr776; rote Signale) sowie (d) dieselbe Metaphase nach DAPI-Färbung (blau). (e) Doppelhybridisierung einer weiteren Metaphase mit *CaRep1* (pCaEr1048; grün) und *CaSat1* (gelbe Fluoreszenz). Die *CaRep*-Signale sind bis in euchromatische Regionen hinein gestreut (Pfeile). (f) Dieselbe Metaphase DAPI-gefärbt: die Hauptbereiche von *CaSat1* (rot) im distalen Heterochromatin sind mit Pfeilen gekennzeichnet. (g) Eine weitere Metaphase nach Hybridisierung mit *CaRep3* (pCaEr776; rote Signale) und DAPI-Färbung (blau). Signale auf den Satellitenanhängseln (weiße Pfeile) sowie deutliche „Cluster“ der *CaRep*-Signale im Heterochromatin der großen Chromosomen („B“) sind erkennbar. Besonders auf dem kleinsten Chromosom ist die ungleiche Intensität der Signale zu beiden Seiten des Centromers zu sehen („H“). *CaRep*-Sequenzen sind auf Chromosomenpaar A von den mit *CaSat1* besetzten distalen heterochromatischen Bereichen neben der

sekundären Einschnürung ausgespart (grüne Pfeile). (h) Interphasekern nach *in-situ*-Hybridisierung mit einer *CaRep2*-Sonde (pCaEr936; grünes Signal) und (i) nach DAPI-Färbung (blaue Fluoreszenz). Die starken Signalintensitäten weisen auf eine hohe Abundanz der Sequenzen im Genom hin.



und B einnehmen. Die *CaRep*-Sequenzen waren im Vergleich dazu eher proximal lokalisiert (Fig.C17g, grüne Pfeile). Die Loci der 18S-5.8S-25S rDNAs auf der sekundären Einschnürung des Chromosoms A waren frei von *CaRep*-Signalen (Fig.C17g), während *CaRep*-Loci auf dem distalen Bereich des Satellitenanhängsels sichtbar waren (weiße Pfeile).

C.I.2.6. Abundanz von *CaRep*-Sequenzen in anderen Spezies

***CaRep*-Sequenzen sind *Cicer*-spezifisch und zeigen innerhalb der Gattung eine unterschiedliche genomische Organisation**

Um herauszufinden, ob und mit welcher Abundanz und genomischer Organisation Sequenzen mit Homologie zu *CaRep* auch in verwandten Pflanzengenomen zu finden sind, wurde eine Southern-Hybridisierung durchgeführt. *RsaI*-restringierte, genomische DNAs aus 1-2 Akzessionen aller annualen *Cicer*-Spezies, verschiedenen Leguminosen und weiteren Dicotyledonen wurden gegen Fragmente der Klone pCaEr1048 (*CaRep1*), pCaEr915 (*CaRep2*) und pCaEr776 (*CaRep3*) hybridisiert. Keine der Sonden detektierte homologe Sequenzen in Leguminosen (außerhalb der Gattung *Cicer*), ebenso wurden in den verwendeten Mono- und Dicotyledonen-Spezies keine Hybridisierungsbanden sichtbar. Dies bedeutet, daß *CaRep* spezifisch für die Gattung *Cicer* ist. Für die Spezies *C. arietinum*, *C. reticulatum* und *C. echinospermum* konnten alle drei Sonden ein intensives und konserviertes Muster im Größenbereich von 0.1-2.2 Kbp vor leichtem Hintergrundschmier produzieren (6-13 Banden), die Bandenmuster der unterschiedlichen *CaRep*-Regionen unterschieden sich jedoch voneinander. In *C. bijugum*, *C. pinnatifidum* und *C. judaicum* war nur ein Teil dieser Muster mit niedrigerer Intensität sichtbar (1-5 Banden), während *C. cuneatum* deutlich abweichende Signale zeigte (lediglich 1 Bande). Im Gegensatz dazu standen die sehr geringen Intensitäten der Banden von *C. chorassanicum* und *C. yamashitae*.

Die starke Signalintensität in *C. reticulatum* und *C. echinospermum* deutet auf eine hohe Abundanz dieser Sequenzfamilie ähnlich der der kultivierten Spezies hin. Auch die genomische Organisation der *CaRep*-Sequenzen erscheint in allen Spezies der ersten Kreuzungsgruppe (*C. arietinum*, *C. echinospermum*, *C. reticulatum*) konserviert, während in der zweiten Kreuzungsgruppe (*C. bijugum*, *C. pinnatifidum*, *C. judaicum*) und in *C. cuneatum* eine abweichende Organisation festzustellen ist. Die *CaRep*-ähnlichen Sequenzen in *C. chorassanicum* und *C. yamashitae* sind deutlich divergent oder liegen in niedrigerer Kopienzahl vor.

C.II. Mittel- und niedrigrepetitive Sequenzen

Neben den beschriebenen hochrepetitiven Sequenzklassen sollte das Vorhandensein weiterer, weniger abundanter repetitiver Sequenzfamilien im Kichererbsengenom untersucht werden. Die Abundanz der klonierten Sequenzen in der genomischen Bank spiegelt - abgesehen von einer Beeinflussung durch die Restriktion (d.h. Erkennungssequenz und Zugänglichkeit für die Endonuklease) - etwa ihre Häufigkeit im Genom wieder. Mittel- und niedrigabundante genomische Sequenzen sind seltener und liefern zudem nur schwächere Hybridisierungssignale, was eine Suche im Gegensatz zu hochrepetitiven Sequenzen sehr viel mühevoller macht. Daher wurde durch gezielte PCR-Amplifikationen und Klonierung der Produkte ein direkter Weg zur Detektion dieser Sequenzklassen gewählt. Um bevorzugt repetitive Sequenzen aus dem Kichererbsengenom zu amplifizieren, wurden zwei Sorten Oligonukleotidprimer verwendet: (1) aus der Literatur bekannte, degenerierte Oligonukleotidprimer, die komplementär zu konservierten Motiven aus codierenden Abschnitten von Retrotransposons sind, was die zu detektierenden Sequenzklassen auf die Gruppe der Retroelemente einschränkte (II.1, II.3); (2) Oligonukleotidmotive, die komplementär zu klonierten Kichererbsenmotiven und repetitiven, homologen Sequenzen aus öffentlichen Datenbanken sind (II.2.).

C.II.1. Das *Ty-copia*-Element *CaTy*

C.II.1.1. Nachweis von *Ty1-copia*-Retrotransposonsequenzen in Leguminosen

***Ty1-copia*-ähnliche Sequenzen sind in Kichererbse und verwandten Leguminosen amplifizierbar**

Um typische Sequenzmotive der ubiquitären *Ty1-copia*-Elemente isolieren und charakterisieren zu können, wurden die von Flavell et al. (1992b) ursprünglich für eine Verwandtschaftsanalyse von *Ty1-copia*-Sequenzen in verschiedenen Pflanzenspezies verwendeten Primer, die komplementär zu Domäne III und V der RT sind und ein ca. 260 Bp langes Fragment aus genomischer DNA amplifizieren sollen, ausgewählt (Fig.C18). Das Oligonukleotid TyB ist in fast allen Retroelementgruppen vorhanden, während TyA für die *Ty1-copia*-Spezifität der amplifizierten Fragmente sorgt (Flavell et al., 1992b).

Als Templat diente genomische DNA annueller *Cicer*-Spezies, weiterer Leguminosen und anderer Dicotyledonen (s.I.1.). Sowohl in allen *Cicer*-Arten als auch in den übrigen untersuchten Pflanzenspezies waren Amplifikate von ca. 260 Bp als klare Banden im Ethidiumbromid-gefärbten Agarosegel sichtbar. Daneben konnten weitere Fragmente von 230, 600 und 690 Bp Länge in verschiedenen Genomen detektiert werden, letztere ausschließlich innerhalb der *Cicer*-Spezies (Fig.C19).

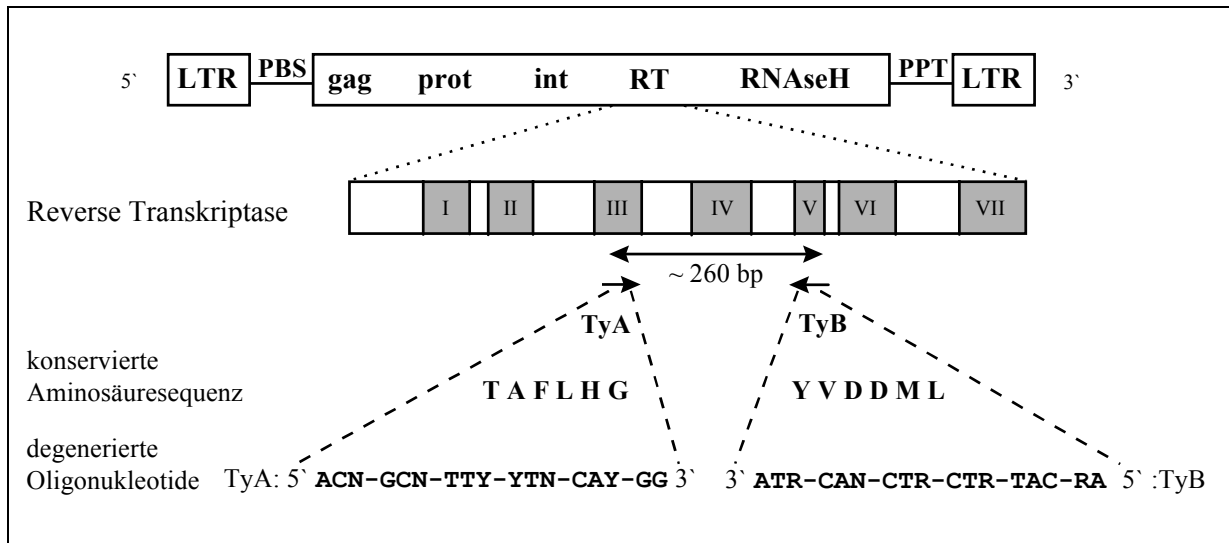


Fig.C18: Position und Sequenz der Oligonukleotide zur Amplifikation von *Ty1-copia*-ähnlichen Elementen.

Das von Flavell et al. (1992b) abgeleitete, degenerierte Oligonukleotidprimerpaar flankiert ein ca. 260 Bp langes Fragment zwischen Domäne III und V der Reversen Transkriptase (Xiong & Eickbush, 1990). Die Sequenz von TyA ist spezifisch für die *Ty1-copia*-Untergruppe.

N = A,C,G,T; Y = C,T; R=A,G.

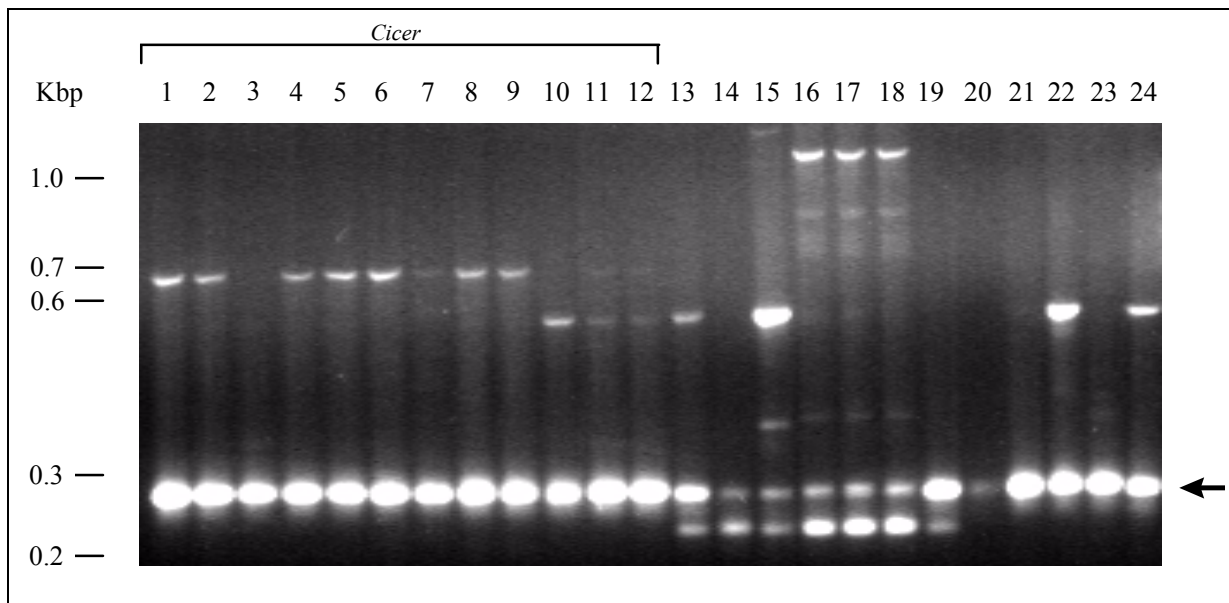


Fig.C19: Amplifikation von *Ty1-copia*-ähnlichen Sequenzen in der Gattung *Cicer* und weiteren Leguminosen.

Im Ethidiumbromid-gefärbten Agarosegel sind Banden zu sehen, die mit den von Flavell et al. (1992b) entwickelten Primern amplifiziert wurden. Folgende genomische DNAs wurden als Templat verwendet: *C. arietinum* (1), *C. reticulatum* (2), *C. chorassanicum* (3), *C. echinospermum* (4,5), *C. bijugum* (6), *C. pinnatifidum* (7,8), *C. judaicum* (9), *C. cuneatum* (10), *C. yamashitae* (11,12), *Pisum sativum* (13), *Vicia faba* (14), *Lens culinaris* (15), *Lathyrus odoratus* (16, 17, 18), *Glycine max* (19), *Phaseolus aureus* (20), *Phaseolus vulgaris* (21), *Nicotiana atropurpurea* (22), *Lycopersicon esculentum* (23) und *Medicago sativa* (24). Amplifikate der erwarteten Größe sind in allen untersuchten Spezies vorhanden (Pfeil).

Die *Ty1-copia*-Sequenzen aus Kichererbse sind zum Teil sehr heterogen

Alle zur Charakterisierung klonierten und sequenzierten Fragmente (TyCa1-TyCa13; AJ411770-AJ411782) zeigten signifikante Ähnlichkeiten zu zahlreichen pflanzlichen Retrotransposons der *Ty1-copia*-Gruppe. Auf Nukleotidebene wiesen sie u.a. hohe Homologien zur entsprechenden Reversen Transkriptase-Region der *Ty1-copia*-Elemente aus *Vicia faba* (AJ239205; 84-93%), *Vigna unguiculata* (Y12763; 85-89%), *Lycopersicon esculentum* (D12844; 83-84%) und *Hordeum vulgare* (AJ241337; 88-91%) auf. Die Ähnlichkeit zu *Tto1* (D83003) und *Tnt1* (X13777) aus *Nicotiana tabacum* lag zwischen 85 und 90%. Die Aminosäuresequenzen der wahrscheinlichsten ORFs zeigten 86-93% bzw. 75-81% Ähnlichkeit zu den RT-Regionen von *Tto1* (T02206) und *Rtr21* (*Oryza sativa*; T03662) bzw. äquivalente Übereinstimmungen mit den entsprechenden Bereichen aus *Solanum tuberosum* (CAA13065; 70-96%), *Lycopersicon esculentum* (CAA13061; 83-92%) und *Picea abies* (CAA11921; 76-83%). Die 260 Bp-Amplifikate repräsentieren somit einen Abschnitt aus *Ty1-copia*-Retrotransposons im Genom der Kichererbse.

Alle dreizehn Fragmente aus *Cicer arietinum* zeigten eine einheitliche Größe von 267 Bp, wiesen ein intaktes offenes Leseraster über die gesamte Länge auf (Ausnahme: TyCa5, TyCa7 und TyCa13, die von je einem Stop-Codon unterbrochen werden) und hatten einen hohen A/T Gehalt von 56-67%. Untereinander zeigten sie 49.8-100% Homologie auf Nukleotidebene, die abgeleiteten Aminosäuresequenzen (Fig.C20a) divergierten bis zu 57% (d.h. 43-100% Ähnlichkeit). Die genetische Distanz der Sequenzen ist aus dem nach der „neighbour-joining“-Methode erstellten Dendrogramm erkennbar (Fig.C20b). Obwohl die Mehrheit der Sequenzen sehr ähnlich war (83-99 % auf Aminosäureebene), wurden auch einige divergente Sequenzen detektiert. So gruppierten TyCa5, TyCa7, die nur 70-75% Homologie zu den übrigen Sequenzen aufwiesen, und TyCa11 mit nur 45-49 % im Dendrogramm weiter entfernt von der Hauptgruppe als Sequenzen fremder Spezies (z.B. *Tto1*, *Tnt1* aus *N.tabacum*). Neben dem Großteil der Sequenzen, der sich zu einer einzigen, relativ homogenen Subfamilie (*CaTy1*) zusammenfassen ließ, stellten TyCa5 und TyCa7 bzw. TyCa11, die untereinander nur zu 43% homolog waren, offenbar zwei weitere divergente Subfamilien dieser Retrotransposonfamilie dar und wurden mit *CaTy2* (TyCa5, TyCa7) und *CaTy3* (TyCa11) bezeichnet. Die allgemeine Bezeichnung für die hier identifizierten *Ty1-copia*-ähnlichen Sequenzen aus *C. arietinum* ist *CaTy* (*Cicer arietinum Ty1-copia*-Familie).

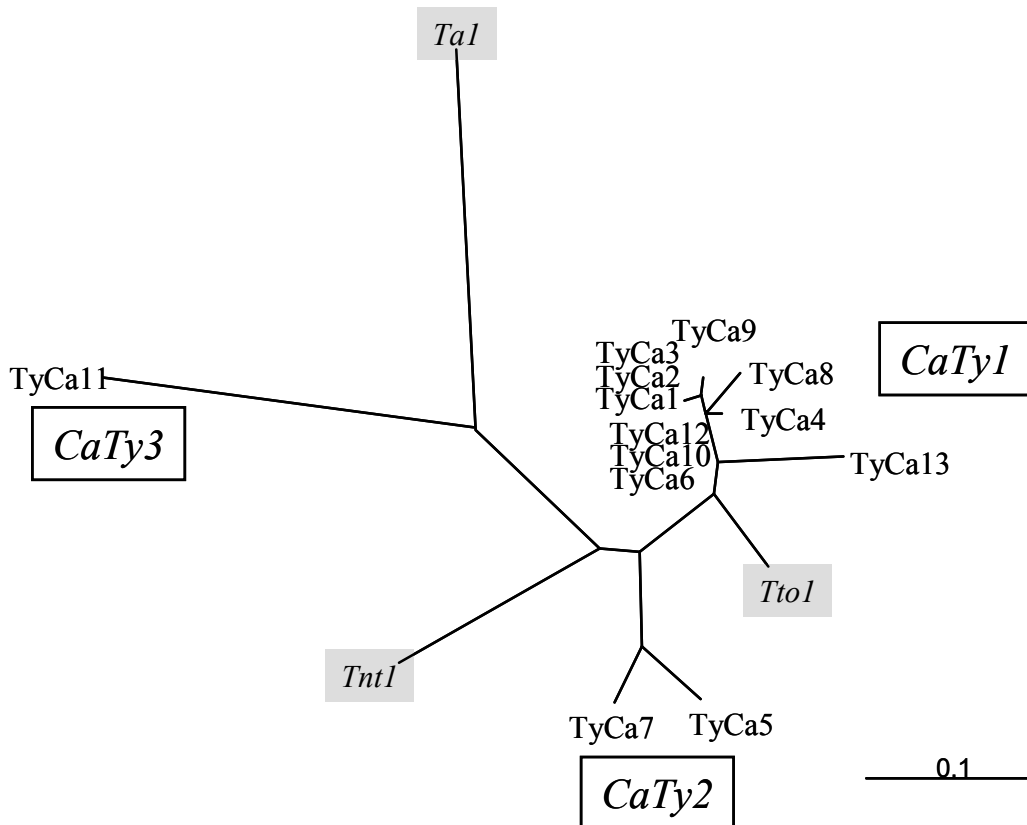
Fig.C20: Sequenzvergleich der wahrscheinlichsten Leseraster aus der klonierten Reverse Transkriptase-Region.

(a) Die sequenzierten PCR-Fragmente aus *C. arietinum* im Vergleich zu den entsprechenden Regionen aus *Tto1* (T02206), *Tnt1* (X13777) (beide aus *Nicotiana tabacum*) und *Tal* (X53975) aus *Arabidopsis thaliana*. Die Oligonukleotidsequenzen wurden nicht in den Sequenzvergleich miteinbezogen. Die Domänen III - V (Xiong & Eickbush, 1990) sind hellgrau unterlegt. Stark konservierte Aminosäuren sind dunkelgrau markiert, Striche zeigen Lücken, die für einen optimalen Sequenzvergleich eingefügt wurden und Sterne symbolisieren Stop-Codons. (b) Genetische Distanz der Aminosäuresequenzen aus (a), als wurzelloses Dendrogramm dargestellt. Es wurde nach der „neighbour-joining“-Methode (Saitou & Nei, 1987) aus dem Sequenzvergleich von (a) erstellt

und betont die Heterogenität der sequenzierten Fragmente aus *C. arietinum* mit den Unterfamilien *CaTy1-3*. Die horizontale Linie repräsentiert Prozent Divergenz (/100).

	III.	IV.	V.
TyCa5	GDLEEEIYMEQPESFKVKG-KKQLVCKFKKSLYGLKQAP*QWYKKFDSFMEKHGYGKTTSDHCVFIKKFSGDGYIILL		LL
TyCa7	GDLEEEIYMEQPEGFEVKG-KDHLVCKLKKRLYGLKQAP*QWYKKFDSFMEKHGYGKTTSYHCVFVKKFSGDGYIILL		LL
TyCa1	GDLEEEIYMKQPDGFLVKG-KEDYVCRRLRKS LYGLKQAPRQWYKKFESVMCEQGYRKTTSDHCVFVKKFADDDFLILL		LL
TyCa3	GDLEEEIYMKQPDGFLVKG-KEDYVCRRLRKS LYGLKQAPRQWYKKFESVMCEQGYRKTTSDHCVFVKKFADDDFLILL		LL
TyCa2	GDLEEEIYMKQPDGFLVKG-KEDYVCRRLRKS LYGLKQAPRQWYKKFESVMCEQGYRKTTSDHCVFVKKFADDDFLILL		LL
TyCa9	GDLEEEIYMKQPDGFLVKG-KEDYVCRRLRKS LYGLKQAPRQWYKKFESVMCEQGYRKTTFDHCVFVKKFADDDFLILL		LL
TyCa10	GDLEEEIYMKQPDGFLVKG-KEDYVCRRLRKS LYGLKQAPRQWYKKFESVMCEQGYRKTTSDHCVFVKKFADDDFLILL		LL
TyCa12	GDLEEEIYMKQPDGFLVKG-KEDYVCRRLRKS LYGLKQAPRQWYKKFESVMCEQGYRKTTSDHCVFVKKFADDDFLILL		LL
TyCa6	GDLEEEIYMKQPDGFLVKG-KEDYVCRRLRKS LYGLKQAPRQWYKKFESVMCEQGYRKTTSDHCVFVKKFADDDFLILL		LL
TyCa8	GDLEEEIYMKQPDGFLVKG-KEDYVCRRLRKS LYGLKQAPRQC YKKFESVMCEQGYRKTTSDQCVFVKKFADDDFIIMLL		LL
TyCa4	GDLEEEIYMKQPDGFLVKG-KEDYVCRRLRKS LYGLKQAPRQWYKKFESIMCEQGYRKTTSDHCVFVKKFADDDFLILL		LL
TyCa13	GDLEEEIYMKQPDGFLVKG-KEDYVCRRLRKS LYGLKQTPM*WYKKFESVIYDQGYMKTTSDHCVFVRKFSKDDFIILL		LL
Tt1	GDLEEEIYMEQPDGFQKKG-KEDYVCRRLRKS LYGLKQAPRQWYKKFESVMGQHGYKKTTSDHCVFAQKFSDDDFIILL		LL
Tnt1	GDLEEEIYMEQPEGFEVAG-KKHMVCKLNKSLYGLKQAPRQWYMKFDSFMKSQTYLKYSDPCVYFKRFSENNFIILL		LL
TyCa11	GDLEEEVYMEQLLGFVAQGESSNMVCR LHRSLYGLKQSPRAWFGREFSSVQEFGMIRSEADHSVFYHHS AQGCIYLVIV-		LL
Tal	GELEEEIYMEQPEGFISED-GKNKVCLLKKLLYGLKQSPRQWNKRFRNRFMIDQNFIRSEHD-ACVKQVTEQEHLYLLL-		LL

a



b

C20

In *C. arietinum* können weitere divergente RT-Sequenzen identifiziert werden

Die auf die Gattung *Cicer* beschränkte 690 Bp-Bande wurde auch aus *C. arietinum* kloniert und sequenziert (AJ411783). Es handelte sich um ein 687 Bp langes PCR-Artefakt, das auf beiden Seiten von Oligonukleotid TyB flankiert wurde. Ein 208 Aminosäuren langes offenes Leseraster (2 Stopcodons, 1 Leserasterwechsel) wies Homologien zu pflanzlichen *Ty1-copia*-Elementen auf (z.B. *Tto1*, *Tnt* aus *N.tabacum*; *Rtr21* aus *O.sativa*). Sowohl der direkte Sequenzvergleich als auch die jeweiligen Positionen homologer Aminosäuren im Retrotransposon *Tto1* (*N.tabacum*; T02206) zeigten, daß das lange Fragment im 3'-Bereich den 267 Bp-Fragmente entsprach, sich jedoch 5' weiter in die RT-Domäne erstreckte. Offenbar enthält das PCR-Artefakt einen degenerierten Bereich der RT-Sequenz um Primer TyA, was die Amplifikation eines längeren Ausschnitts ermöglichte.

C.II.1.2. Abundanz und genomische Organisation von *Ty1-copia*-Retrotransposons in *Cicer arietinum*

***CaTy*-Sequenzen sind mittelrepetitiv und dispers oder in dispersen „Clustern“ organisiert**

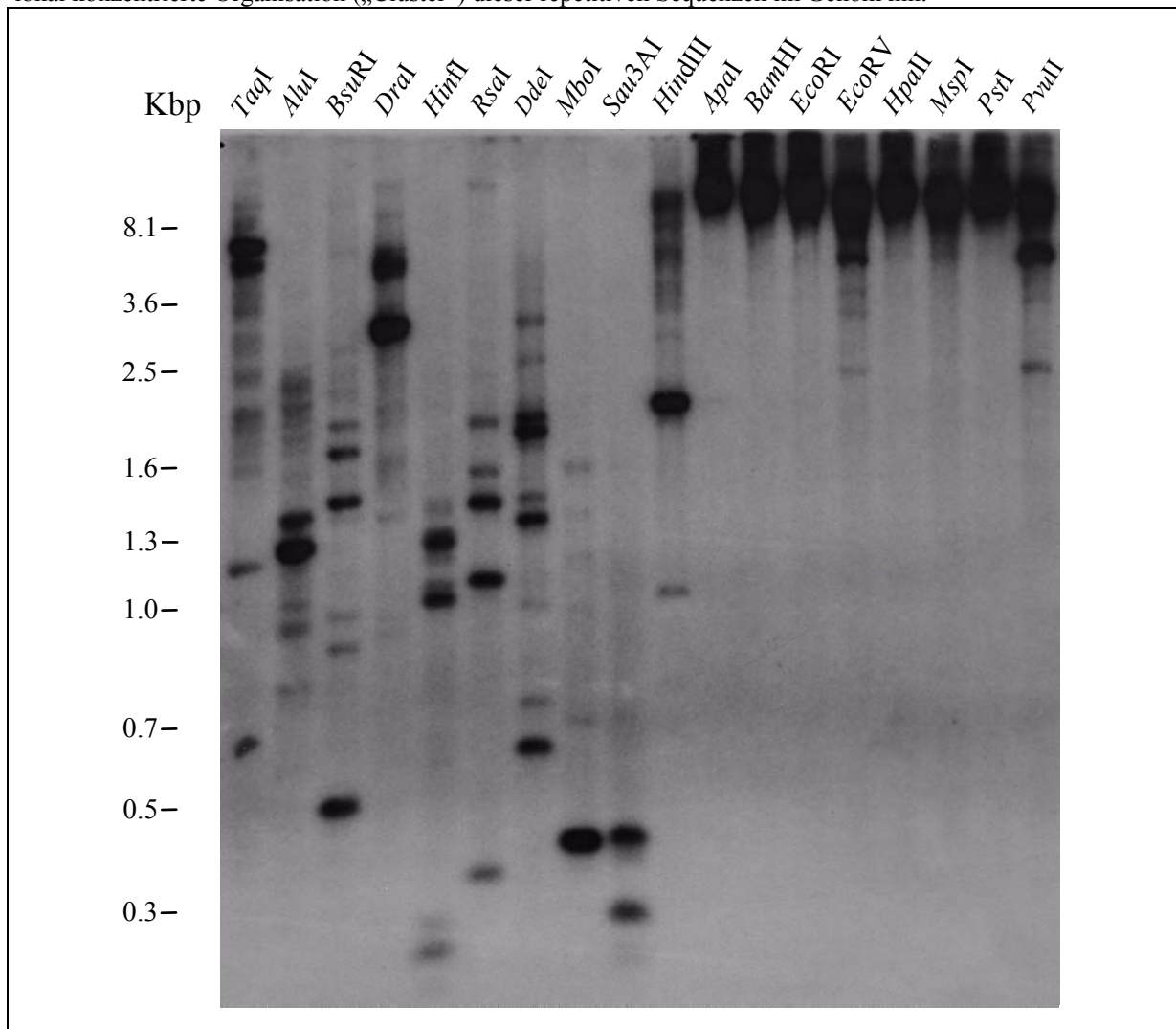
Zur Ermittlung von Abundanz und genomischer Organisation der *Ty1-copia*-Sequenzen in *C. arietinum* wurde eine Southern-Analyse mit restringierter genomischer DNA durchgeführt. Als Sonde diente das klonierte Fragment TyCa1. Ein komplexes Muster aus 4-15 Banden auf einem Hintergrundschmier im Molekulargewichtsbereich von 0.2 bis über 8kb (Fig.C21) wies auf die disperse, möglicherweise „Cluster“-förmige Organisation der Wiederholungseinheiten hin. In einigen Spuren hybridisierten nur hochmolekulare Fragmente (>8-20kb) schmierförmig. Neben der Signalintensität belegt dies die mittelrepetitive Abundanz der homologen Sequenzen im Genom, die geringer als die der genannten hochrepetitiven Elemente *CaSat1*, 2 und *CaRep* ist. Schwächere Signale einzelner Banden können aufgrund der stringenten Hybridisierungsbedingungen (90-100%; Meinkoth & Wahl, 1984) durch Motive zustande kommen, die weniger homolog zur Sonde TyCa1 sind. Dabei ist eine Detektion von Sequenzen der Untergruppe *CaTy2* und *CaTy3* jedoch auszuschließen. Die sehr großen homologen Fragmente der *HpaII*- und *MspI*-Restriktion weisen entweder auf eine hohe Methylierung der Sequenzmotive 5'-CCGG-3' hin oder auf die generelle Seltenheit dieser Schnittstelle, bedingt durch den AT-Reichtum der Sequenzen (vgl.1.1).

CaTy-Elemente stellen offenbar eine mittelrepetitive Sequenzfamilie im Kichererbsengenom dar, die dispers oder in dispersen Anhäufungen („Clustern“) organisiert ist.

Fig.C21: Genomische Organisation der *Ty1-copia*-Sequenzen *CaTy* in *C. arietinum*.

Autoradiogramm eines Southern-Blots mit genomischer DNA aus *C. arietinum*, die mit verschiedenen Restriktionsendonukleasen behandelt und gegen das Fragment aus TyCa1 hybridisiert wurde. Die stringenten Hybridisierungsbedingungen (90-100%; Meinkoth & Wahl, 1984) lassen keine Hybridisierung mit Sequenzen

aus den Untergruppen CaTy2 und CaTy3 zu. Die komplexen Bandenmuster deuten auf eine disperse, eventuell lokal konzentrierte Organisation („Cluster“) dieser repetitiven Sequenzen im Genom hin.



C.II.1.3. Chromosomale Lokalisation der *CaTy*-Familie in Kichererbse

CaTy-Sequenzen sind hauptsächlich im distalen pericentrischen Heterochromatin zu finden

Als Sonde für FISH diente das markierte, 260 Bp lange PCR-Produkt der TyA/B Amplifikation, das ein heterogenes Gemisch von Amplifikaten der *CaTy*-Familie darstellt (wie aus der Sequenzierung desselben Ansatzes hervorgeht; 1.1.), und viele Loci unterschiedlicher Sequenz auf den Metaphasechromosomen nachweisen sollte. Deutliche Hybridisierungssignale konnten auf allen acht Chromosomenpaaren von *C. arietinum* detektiert werden (Fig.C22). Die Loci waren jeweils in den distalen Bereichen des pericentrischen Heterochromatinblocks lokalisiert und sparten den zentralen Teil um das Centromer aus (Fig.C22a-d) - ein Bereich, in dem die größte Konzentration des hochabundanten Satelliten *CaSat2* zu erwarten ist (I.1.2.). Lediglich auf den beiden größten Chromosomenpaaren A und B waren weitere Signale in der Mitte des Heterochromatinblocks angesiedelt (Fig.C22a; Pfeile). Doppelpunktförmige Fluoreszenzsignale in den angrenzenden, nur noch schwach DAPI-gefärbten Regionen einiger Chromosomen wiesen auf weitere *CaTy*-

Kopien in euchromatischen Bereichen hin (Fig.C22c, rote Pfeile). Wie bei *CaRep* (I.2.5.) waren auch in den distalen Regionen der Satellitenanhängsel von Chromosom A Signale detektierbar (Fig.C22c,d, weiße Pfeile). Die Hauptloci von *CaSat1* auf den Chromosomen A und B wie auch die sekundäre Einschnürung der NOR und der angrenzende Locus der 18S-2.5S-25S rDNA (Chromosom A) waren frei von *CaTy*-Signalen.

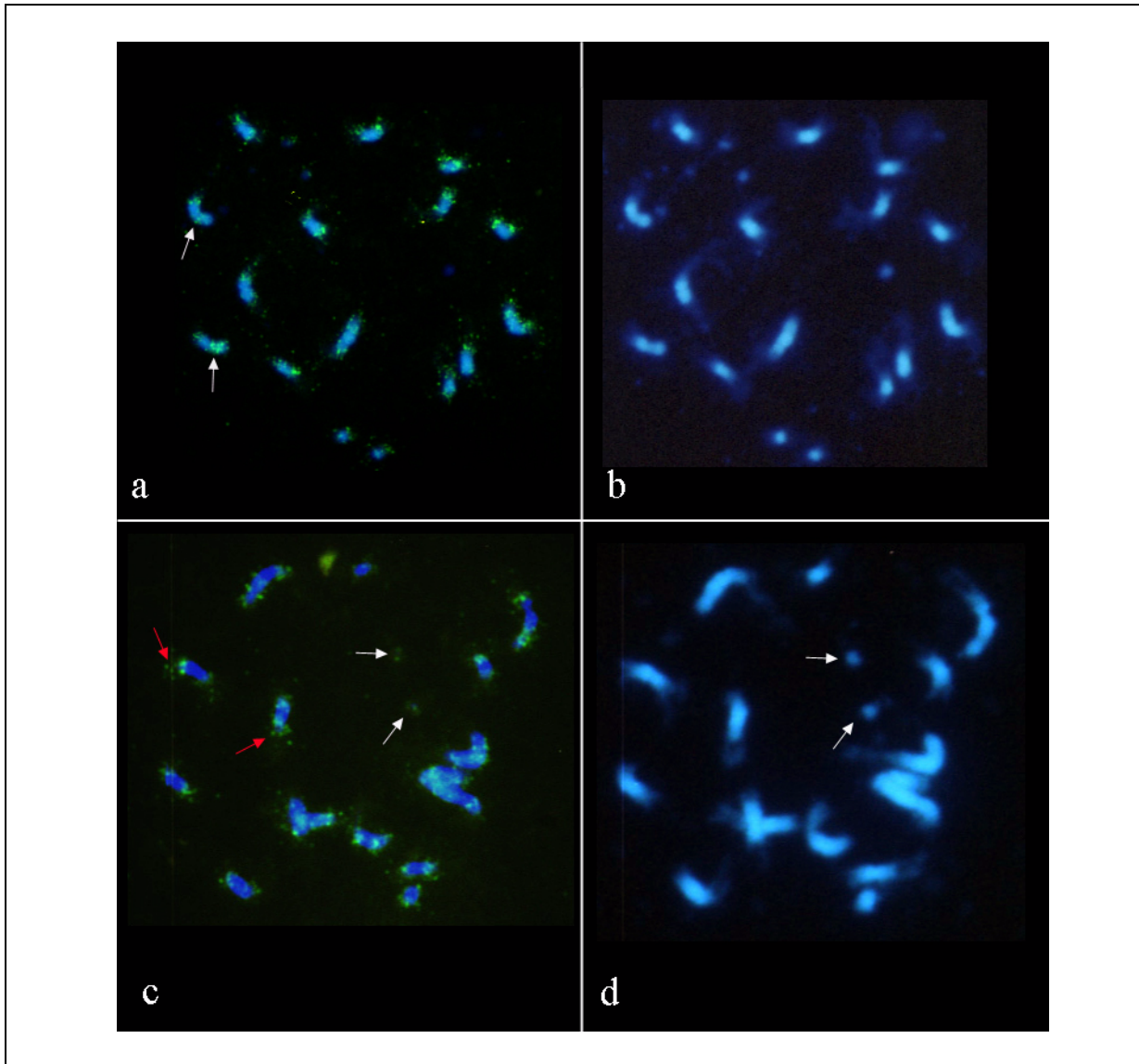


Fig.C22: Lokalisation der *CaTy*-Sequenzen auf Metaphasechromosomen von *C. arietinum*.

a-d) Metaphasechromosomen nach DAPI-Färbung (blaue Fluoreszenz). a,c) Dieselben Metaphasechromosomen nach FISH mit dem markierten TyA/B-Amplifikat (Computerüberlagerung). Die Signale sind als grüne Fluoreszenz in den distalen Bereichen des pericentrischen Heterochromatins sichtbar und sparen den zentralen Teil größtenteils aus. Zusätzlich sind Signale in benachbarten, euchromatischen Bereichen (rote Pfeile in c) und auf den Satellitenanhängseln von Chromosom A sichtbar (weiße Pfeile in c,d).

C.II.1.4. Sequenzen vom *CaTy*-Typ im Genus *Cicer* und in verwandten Spezies

***CaTy*-homologe Motive sind in allen *Cicer*-Spezies außer *C. cuneatum* vorhanden und nicht auf den Genus *Cicer* beschränkt**

Sowohl 260 Bp-Amplifikate der PCR-Analyse als auch Sequenzhomologien zu anderen pflanzlichen *Ty1-copia*-Elementen hatten die Existenz *Ty1-copia*-ähnlicher Sequenzen in nahen Verwandten von *C. arietinum* angezeigt (1.1.). Um zu prüfen, ob diese Sequenzen zur *CaTy*-Familie aus Kichererbse gehören, und um ihre Abundanz und genomische Organisation festzustellen, wurde eine Southern-Analyse mit *RsaI*-restringierter, genomischer DNA aus verschiedenen *Cicer*-Spezies, weiteren Leguminosen und anderen Dicotyledonen durchgeführt. Als Sonde diente das Fragment aus TyCa1. In allen Spezies der Gattung *Cicer* (mit Ausnahme von *C. cuneatum*) wurde ein Muster von 1-4 Banden im Größenbereich von 0.3-2.5 Kbp sichtbar, begleitet von einem schwachen Hintergrundsmier. Restriktionsfragmente von *C. cuneatum*-DNA zeigten auch bei längerer Expositionszeit keinerlei Signale. Die Bandenmuster waren für alle Akzessionen einer Spezies konserviert, ebenso die Muster von *C. reticulatum* und *C. arietinum*, die sich ihrerseits nur unwesentlich von *C. echinospermum* unterschieden. Die Signale der zweiten Kreuzungsgruppe (*C. bijugum*, *C. judaicum* und *C. pinnatifidum*) glichen denen der ersten Gruppe, jedoch wiesen *C. judaicum* und *C. pinnatifidum* schwächere Signale auf. *C. chorassanicum* und *C. yamashitae* waren untereinander ähnlich, wichen jedoch vom Muster der anderen Annuellen ab. Die einzige perennielle Spezies *C. anatolicum* wies mit nur zwei Banden ein völlig anderes Muster auf (ohne Bild).

CaTy-homologe Sequenzen aus dem Kichererbsengenom sind demnach in allen annuellen *Cicer*-Spezies (Ausnahme: *C. cuneatum*) sowie in der perenniellen *C. anatolicum* mit leichten Unterschieden der Abundanzverhältnisse mit einer mittelrepetitiven Kopienzahl vorhanden. Die genomische Organisation der Kopien ist in den ersten beiden Kreuzungsgruppen nahezu einheitlich, zeigt jedoch bei der perenniellen Art sowie *C. chorassanicum* und *C. yamashitae* Unterschiede. *C. cuneatum* enthält *CaTy*-ähnliche Sequenzen entweder gar nicht, in signifikant niedrigerer Kopienzahl oder deutlich von der Sondensequenz divergierend.

Zusätzlich konnten schwache Signale in *Pisum sativum* (vier Banden), beiden *Phaseolus*-Arten (eine Bande) sowie eine schwache Bande in *Vicia faba* detektiert werden, nicht jedoch in den übrigen Leguminosen oder Tomate und Tabak. Die Bandenmuster innerhalb einer Art glichen sich. *CaTy*-ähnliche Motive sind folglich nicht auf den Genus *Cicer* beschränkt, sondern auch bei nah verwandten Leguminosen vorhanden.

Weitere *CaTy*-ähnliche Sequenzen aus anderen Leguminosen oder Tomate und Tabak können offenbar nur durch die degenerierten Oligonukleotidprimer in der PCR detektiert werden und zeigen in der stringenteren Southern-Analyse (90-100%; Meinkoth & Wahl, 1984) keine ausreichende Homologie mehr zur Sonde, die ja nur einen Teil der *CaTy*-Elemente repräsentiert.

C.II.2. *CaDis* - das Relikt eines Retrotransposons?

C.II.2.1. Sequenzen mit Homologie zum repetitiven Element *RPE15* (*Medicago sativa*)

Bei der Analyse von Mikrosatelliten im Genom der Kichererbse wurde in der 5'-benachbarten Region eines (TAA)₄₁-Motivs eine 270 Bp lange Sequenz entdeckt (Ds1, AJ411815), die auf Nukleotidebene 66% Homologie zu einem repetitiven Element (*RPE15*) aus Luzerne (*Medicago sativa*; L39960) aufwies (Winter, pers. Mitteilung; Fig.C23). Dieses Element ist mittelrepetitiv im Genom von *Medicago sativa* vorhanden und wird blütenesspezifisch transkribiert (Xia et al., 1996). Um herauszufinden, ob die homologe Sequenz auch in der Kichererbse in multiplen Kopien vorliegt, wurden die entsprechenden Motive amplifiziert und näher charakterisiert. Dazu wurde aus den Randbereichen der Konsensussequenz der 270 Bp langen homologen Region in *M.sativa* und *C. arietinum* ein degeneriertes Oligonukleotidpaar (*repL/R*) abgeleitet, dessen Sequenz in Fig.C23 dargestellt ist.

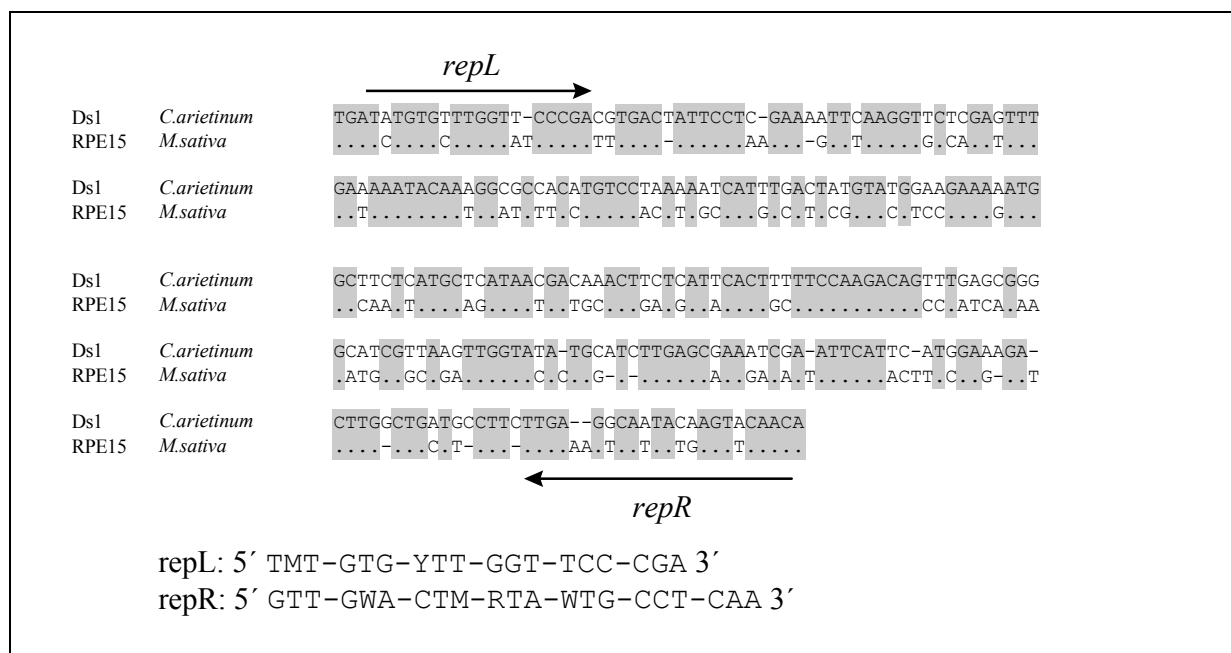


Fig.C23: *RPE15*-homologe Sequenzen aus *C. arietinum* und *M.sativa* und die Position der degenerierten Oligonukleotide *repL/R*.

Sequenzvergleich zwischen dem repetitiven Element *RPE15* aus *M.sativa* und der homologen Sequenz aus *C. arietinum*. Aus dem Randbereich der Konsensussequenz wird ein Oligonukleotidpaar abgeleitet, um weitere homologe Sequenzen zu detektieren. Identische Nukleotide sind durch Punkte dargestellt und grau unterlegt, zum optimalen Sequenzvergleich eingefügte Lücken durch Striche symbolisiert. Die Positionen der Oligonukleotidsequenzen *repL/R* sind als Pfeile dargestellt.

Als Templat diente genomische DNA aus annuellen *Cicer*-Spezies, weiteren Leguminosen und anderen Dicotyledonen. Amplifikate der zu erwartenden Länge von 270 Bp entstanden in allen *Cicer*-Spezies sowie in *Lens culinaris*, *Glycine max* und der als Kontrolle fungierenden *Medicago sativa*, jedoch nicht in den Nichtleguminosen (Fig.C24). Weitere Fragmente von

200, 450 und 850 Bp wurden in verschiedenen Spezies amplifiziert. Offenbar sind verwandte Sequenzmotive des degenerierten Oligonukleotidpaars außer in *Cicer arietinum* auch in anderen Leguminosen vorhanden.

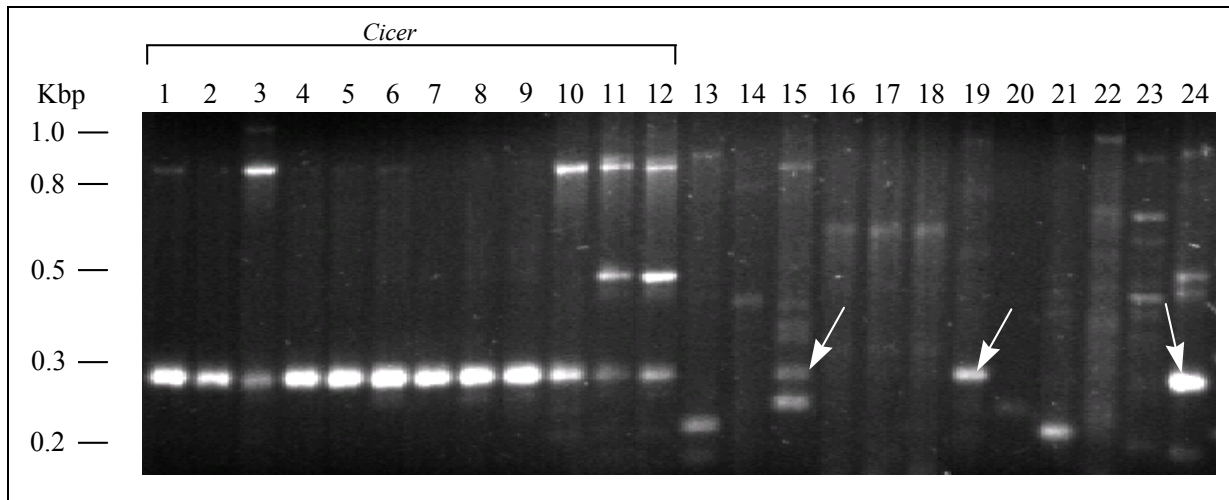


Fig.C24: Amplifikation genomischer DNA aus Pflanzen der Gattung *Cicer* und weiteren Leguminosen mit dem Primerpaar repL/R.

Im Ethidiumbromid-gefärbten Agarosegel sind Banden sichtbar, die mit dem degenerierten Oligonukleotidpaar repL/R amplifiziert wurden. Als Templat wurde genomische DNA folgender Spezies verwendet: *C. arietinum* (1), *C. reticulatum* (2), *C. chorassanicum* (3), *C. echinospermum* (4,5), *C. bijugum* (6), *C. pinnatifidum* (7,8), *C. judaicum* (9), *C. cuneatum* (10), *C. yamashitae* (11,12), *Pisum sativum* (13), *Vicia faba* (14), *Lens culinaris* (15), *Lathyrus odoratus* (16, 17, 18), *Glycine max* (19), *Phaseolus aureus* (20), *Phaseolus vulgaris* (21), *Nicotiana atropurpurea* (22), *Lycopersicon esculentum* (23) und *Medicago sativa* (24). Amplifikate der erwarteten Größe sind in allen *Cicer*-Spezies sowie in *Lens culinaris*, *Glycine max* und *Medicago sativa* (Kontrolle) sichtbar (Pfeile).

C.II.2.2. Sequenzierung der *RPE15*-Amplifikate

Die Amplifikate weisen Homologie zur Polyproteinregion von pflanzlichen Retrotransposons auf

Zehn rekombinante Klone der PCR-Bande aus *C. arietinum* (Ds2-Ds11; AJ411816-AJ411825) wurden sequenziert, so daß zusammen mit dem ursprünglich klonierten Fragment (Ds1; AJ411815) und fünf cDNA-Fragmenten (Dc1-Dc5; AJ411826-AJ411830; s. C.IV.2.) 16 Sequenzen vorlagen. Alle klonierten Fragmente wiesen signifikante Homologien zu codierenden Polyproteinsequenzen pflanzlicher Retrotransposons auf. Die meisten der wahrscheinlichsten Leseraster der klonierten Fragmente zeigten 39-52% Ähnlichkeit zu *RetroSor1* (AF098806), einem *Ty3-gypsy*-Element aus *Sorghum bicolor*, 39-46% zu einem *gag*-Genprodukt aus *Zea mays* (S58484) und 44-62% zum Polyprotein eines Retroelements aus *Arabidopsis thaliana* (AC005957). Zwei ESTs aus *Oryza sativa* (C72786, C26347) mit Homologie zu äquivalenten Retrotransposonregionen - wie etwa *RetroSor* (AF098806) und *Cinful* (*Z.mays*, AF049110) - wiesen jeweils 44-52% Ähnlichkeit zu den potentiellen Leserastern der meisten Fragmente auf. Die äquivalenten Sequenzen gehörten jeweils zum *gag*-Bereich der codierenden Region, der eine Nukleinsäurebindestelle und die codierende Sequenz für ein capsidähnliches Protein umfaßt. Auf Nukleotidebene wiesen einige

Fragmente kurze homologe Basenfolgen zu dem retrovirusähnlichen Element *Calypso* (*Glycine max*, AF186185) und einem dispersen repetitiven Element aus *Vicia melanops* (AJ391794) auf.

Die 16 Fragmente hatten eine relativ einheitliche Länge von 268-269 Bp und einen AT-Gehalt von 58.5-64.2%. Nur Ds6 zeigte eine Deletion von 51 Bp (Fig.C25a). In 11 Sequenzen war ein durchgehendes offenes Leseraster im Bereich zwischen den Primerbindestellen erkennbar, in zwei Fällen war es durch ein Stop-Codon und in drei Fällen durch einen Leserasterwechsel unterbrochen. Untereinander zeigten die 16 Nukleotidsequenzen 85-100% Homologie, die abgeleitete Aminosäuresequenz divergierte bis zu maximal 25%. Die Homologie zum ursprünglich in der Sequenzdatenbank gefundenen Element *RPE15* aus *M.sativa* (L39960) betrug für alle analysierten Sequenzen 59-63%, was die Verwandtschaft zum ursprünglich in der Kichererbse detektierten Locus (Ds1) bestätigt. Die genetische Distanz zwischen den wahrscheinlichsten Leserastern ist in Form eines Dendrogramms (Saitou & Nei, 1987) in Fig.C25b dargestellt. Homologien zu anderen klonierten repetitiven Sequenzen aus *C. arietinum* konnten nicht festgestellt werden.

Mit Hilfe des Oligonukleotidpaares repL/R können also Sequenzen aus dem Kichererbsengenom amplifiziert werden, die Teile der codierenden Region eines Retrotransposons oder dessen Relikten darstellen und in mindestens 13 unabhängigen Kopien existieren. Da die klonierten Sequenzen aus dem *gag*-Bereich der Polyproteinregion stammen, bleibt unklar, ob es sich hierbei um ein Element des *cop*ia- oder des *gypsy*-Typs handelt. Bemerkenswert ist die Nukleotidhomologie zu mehreren dispersen repetitiven Elementen aus anderen Leguminosen (*M.sativa*, L39960; *Vicia melanops*, AJ391794; *Glycine max*, AF186185).

C.II.2.3. Genomische Organisation von *CaDis* im Genom der Kichererbse

***CaDis*-Elemente sind mittelrepetitiv und dispers im Kichererbsengenom organisiert**

Eine Southern-Analyse mit genomischer DNA aus *C. arietinum*, bei der Ds2 als Sonde verwendet wurde, zeigte ein komplexes Muster aus 5-11 diskreten Banden im Molekulargewichtsbereich zwischen 0.2 - >8Kb auf leichtem Hintergrundschmier (Fig.C26). Aus dem komplexen Bandenmuster kann die Existenz mehrerer Kopien und deren disperse Organisation im Genom abgeleitet werden, während die Intensität der Hybridisierungssignale auf eine niedrig- bis mittelrepetitive Abundanz der Motive hindeutet.

Fig.C25: Sequenzvergleich der wahrscheinlichsten Leseraster in klonierten *CaDis*-Fragmenten.

(a) Sequenzen aller klonierten PCR-Fragmente aus *C. arietinum* sind im Vergleich zu den entsprechenden Regionen aus *RetroSor* (*Sorghum bicolor*; AF098806) und einem EST mit Retrotransposonhomologien aus *O.sativa* (E2258) aufgelistet. Die Oligonukleotidprimer wurden nicht in den Sequenzvergleich miteinbezogen. Ds1 ist die erste *CaDis*-Sequenz des Kichererbsengenoms und wird von einem (TAA)₄₁-Motiv flankiert. Konservierte Aminosäuren sind hellgrau unterlegt und Bindestriche zeigen Lücken, die für einen optimalen Sequenzvergleich eingefügt wurden. Sterne kennzeichnen Stop-Codons. In einigen Fällen waren Leserasterwechsel notwendig, um die Sequenzen vergleichen zu können. (b) Genetische Distanz der Aminosäuresequenzen aus (a) als wurzelloses Dendrogramm (Saitou & Nei, 1987). Die horizontale Linie

Die mit dem Oligonukleotidpaar repL/R amplifizierten, retrotransposonartigen Sequenzen stellen eine weitere repetitive Sequenzfamilie dar, die in mittlerer Abundanz im Kichererbsengenom vorliegt. Aufgrund ihrer dispersen genomischen Organisation wurden diese Sequenzen als *CaDis*-Familie (*Cicer arietinum* Disperser Repeat) bezeichnet.

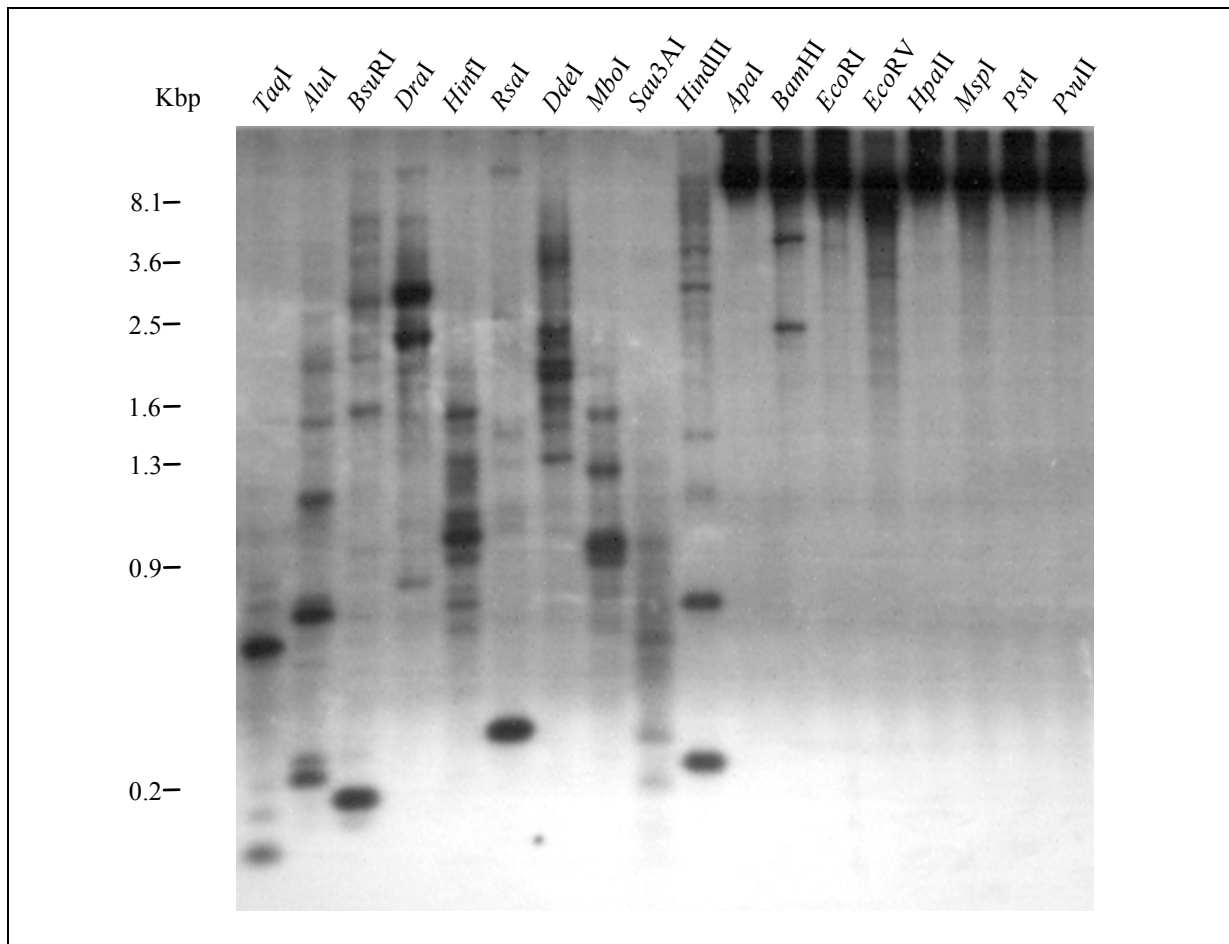


Fig.C26: Genomische Organisation der *CaDis*-Sequenzen im Genom von *C. arietinum*.

Autoradiogramm eines Southern-Blots mit restringierter genomischer DNA aus *C. arietinum*, die gegen das Fragment aus Ds2 hybridisiert wurde. Die Stringenz betrug 90-100% (Meinkoth & Wahl, 1984). Aus dem Bandenmuster kann die Existenz mehrerer Kopien und deren disperse Organisation im Genom abgeleitet werden.

C.II.2.4. Chromosomale Lokalisation der *CaDis*-Sequenzen

CaDis-Signale liegen auch in den euchromatischen Bereichen nahe der pericentrischen Heterochromatinblöcke

Um die chromosomale Lokalisation von *CaDis* zu untersuchen, wurde ein kloniertes *CaDis* Fragment (Ds2) als Sonde für FISH verwendet. Auf mehreren Chromosomenpaaren konnten diskrete *CaDis*-Hybridisierungssignale detektiert werden, die jeweils in den distalen Bereichen der großen pericentrischen Heterochromatinblöcke lokalisiert waren (Fig. C27a-d).

Die Anzahl und Intensität der Fluoreszenzsignale weist auf eine mittelrepetitive Abundanz dieser *CaDis*-Sequenzen hin. Teilweise wurden diese, jeweils auf beiden Chromatiden befindlichen Signale nur an einem Ende des heterochromatischen Blocks sichtbar. In einigen Fällen grenzten sie unmittelbar an die euchromatischen Bereiche der Metaphasechromosomen mit schwächerer DAPI-Färbung an oder erschienen in den distaler gelegenen euchromatischen Regionen (Fig.C27a,c; weiße Pfeile). Auf einem der beiden großen Chromosomenpaare (A) war einer der Loci mitten im pericentromerischen Heterochromatin sichtbar (Fig.C27a; rote Pfeile), während das zweite Signal im distalen Bereich des Blocks lag. Die NOR-Region auf A und die zentrale Region aller pericentrischen Heterochromatinblöcke trugen keine sichtbaren Hybridisierungssignale, so daß *CaDis*-Motive von den Bereichen der hochabundanten repetitiven Familien und rDNA größtenteils ausgespart erscheinen.

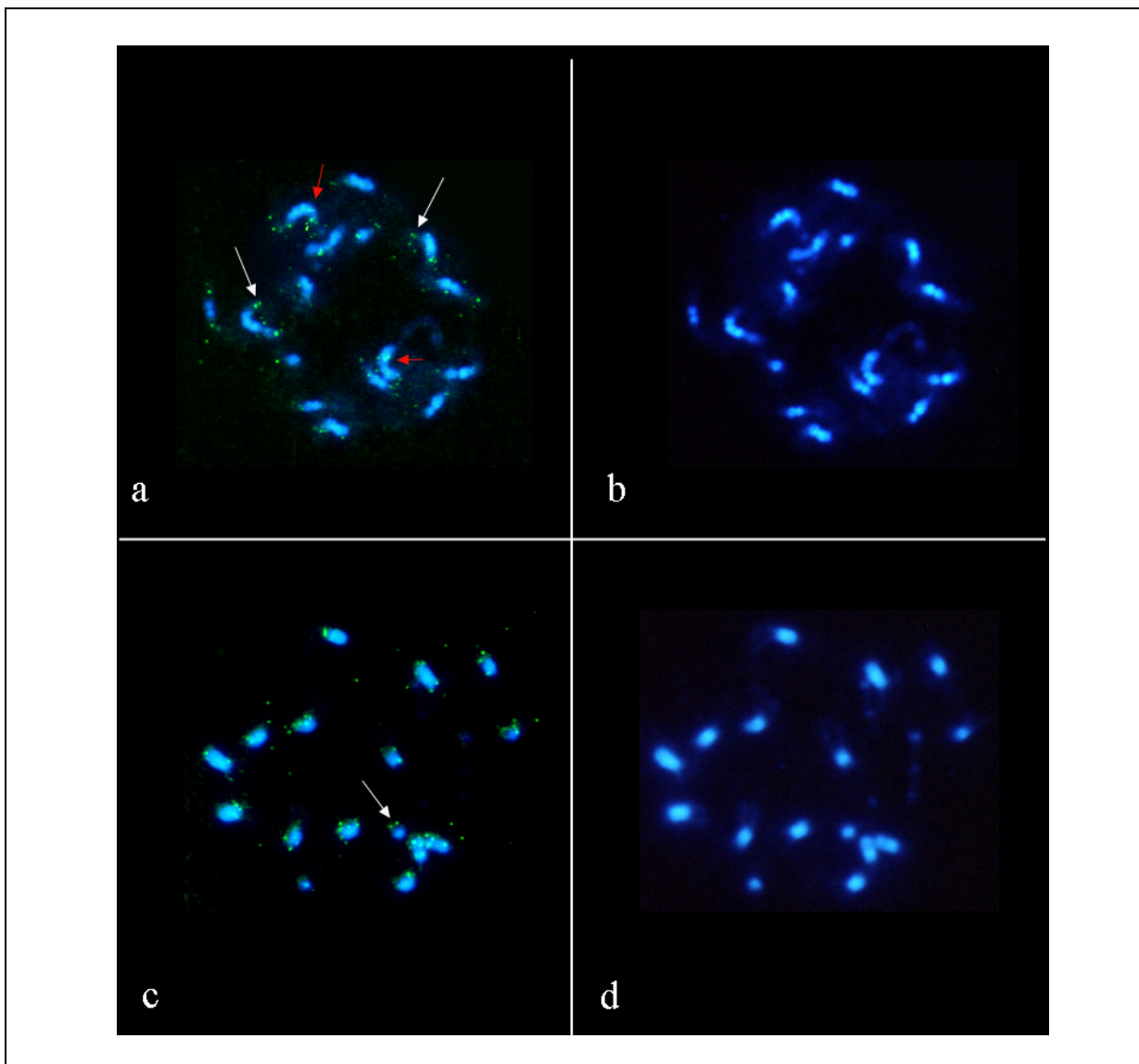


Fig.C27: Chromosomale Lokalisation der *CaDis*-Sequenzen auf Metaphasechromosomen von *C. arietinum*.

a-d) Metaphasechromosomen nach DAPI-Färbung (blaue Fluoreszenz). a,c) Dieselben Metaphasechromosomen nach FISH mit dem markierten Fragment von Ds2 (Computerüberlagerung). Die grünen Fluoreszenzsignale liegen im distalen Bereichen des pericentrischen Heterochromatins von mehreren Chromosomen, teilweise auch in benachbarten, euchromatischen Bereichen (weiße Pfeile). Auf einem der großen Chromosomenpaare (A) liegt ein Signal mitten im Heterochromatinblock (rote Pfeile in a).

C.II.2.5. Abundanz *CaDis*-ähnlicher Sequenzmotive im Genus *Cicer* und verwandten Spezies

***CaDis*-homologe Sequenzen sind in *Cicer* fast einheitlich organisiert und nicht auf die Gattung beschränkt**

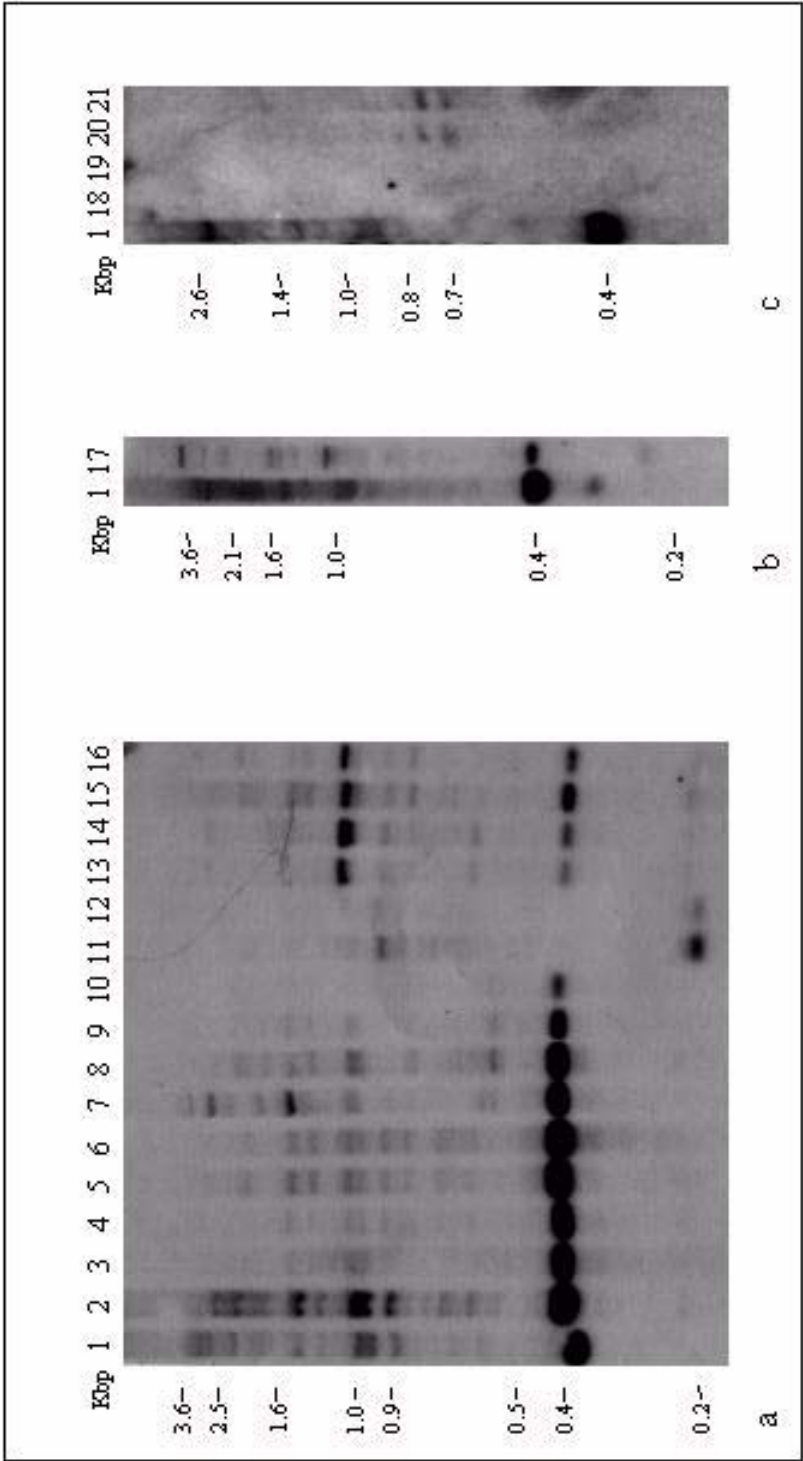
Die Homologie der 270 Bp-Amplifikate in den übrigen annuellen Arten des Genus *Cicer* (2.1) zu den Kichererbsenmotiven sowie ihre Abundanz und genomische Organisation in den Genomen der Wildarten sollten durch eine Southern-Hybridisierung überprüft werden, ebenso die Verbreitung von *CaDis* homologen Sequenzen in anderen Leguminosen und weiteren dikotylen Nichtleguminosen. *RsaI*-restringierte, genomische DNA aus neun *Cicer*-Spezies, weiteren Leguminosen und anderen Dicotyledonen wurde jeweils gegen das klonierte Fragment aus Ds2 hybridisiert. In allen untersuchten Spezies der Gattung *Cicer* (außer *C. cuneatum*) konnte ein monomorphes, intensives Hybridisierungssignal bei 400 Bp nachgewiesen werden, weitere schwache Banden bis zu einem Molekulargewicht von ca. 3.6 Kbp waren auf einem leichten Hintergrundschmier sichtbar (Fig.C28a,b). Dies belegt die Anwesenheit multipler Kopien von *CaDis*-homologen Sequenzen mit disperser Organisation in allen untersuchten *Cicer*-Spezies. Das Muster der schwächeren Banden war jeweils innerhalb der ersten Kreuzungsgruppe (*C. arietinum*, *C. reticulatum*, *C. echinospermum*) und innerhalb der verschiedenen Akzessionen der anderen Spezies konserviert. Es variierte jedoch zwischen den Spezies, so daß man eine leichte spezies- (bzw. kreuzungsgruppen-) spezifische Differenzierung der ansonsten ähnlichen genomischen Organisation annehmen kann. In *C. cuneatum* wiesen zwei Banden von mittlerer Intensität (200 Bp/ 950 Bp) auf die Anwesenheit homologer *CaDis*-Motive hin, die jedoch eine andere genomische Organisation zeigten.

Trotz stringenter Hybridisierungsparameter (1xSSC; 65°C) konnte eine Doppelbande (700 Bp/ 750 Bp) in den *Pisum*-Akkzessionen detektiert werden (Fig.C28c). Homologe Sequenzen zu *CaDis* aus Kichererbse sind demnach nicht auf die Gattung *Cicer* beschränkt, sondern auch in anderen Leguminosen konserviert. In einigen Leguminosen (z.B. *Lens culinaris*, *Glycine max*, *Medicago sativa*) ließen sich keine Southern-Signale mehr detektieren, obwohl die Anwesenheit verwandter Sequenzen durch PCR belegt wurde (2.1.). Zumindest in *Medicago sativa* wird dies durch die starke Divergenz der Motive zur Sondensequenz verursacht (RPE15 hat nur 60% Ähnlichkeit zur Nukleotidsequenz der Sonde).

Fig.C28: Genomische Organisation von *CaDis*-ähnlichen Sequenzen in der Gattung *Cicer* und verwandten Arten.

(a) Autoradiogramm eines Southern-Blots mit *RsaI*-restringierter genomischer DNA aus *C. arietinum* ILC 3279 (1), *C. arietinum* ILC 482 (2), *C. reticulatum* PI 489777 (3), *C. reticulatum* ILWC 21 (4), *C. echinospermum* ILWC 39 (5), *C. echinospermum* ILWC 35 (6), *C. bijugum* ILWC 62 (7), *C. pinnatifidum* ILWC 204 (8), *C. judaicum* ILWC 57 (9), *C. judaicum* ILWC 56 (10), *C. cuneatum* ILWC 40 (11), *C. cuneatum* ILWC 37 (12), *C. chorassanicum* ILWC 146 (13), *C. chorassanicum* ILWC 23 (14), *C. yamashitae* ILWC 53 (15) und *C.*

yamashitae ILWC 214 (16). Als Sonde diente das Fragment aus Ds2. (b) Autoradiogramm eines weiteren Southern-Blots mit *RsaI*-restringierter genomischer DNA aus *C. arietinum* ILC 3279 (1) und *C. anatolicum* PI383626 (17), hybridisiert gegen das Fragment aus Ds2. (c) Gegen dieselbe Sonde wurde ein dritter Southern-Blot mit *RsaI*-restringierter, genomischer DNA aus *C. arietinum* ILC 3279 (1), *Lathyrus odoratus* (18, 19) und *Pisum sativum* (20, 21) hybridisiert.



C28

C.II.3. *CaLin* - ein LINE-ähnliches non-LTR-Retrotransposon

C.II.3.1. Identifikation von LINE-ähnlichen Sequenzen in verschiedenen *Cicer*-Spezies

Neben den bereits identifizierten Retrotransposonfamilien sollten LINE-ähnliche Elemente durch PCR detektiert werden. Der Nachweis wurde ähnlich wie bei den *Ty1-copia*-Elementen mit Hilfe degenerierter Oligonukleotidprimer durchgeführt, die aus den codierenden Regionen des Elements *Ta11-1* (*A. thaliana*) und weiterer pflanzlichen LINEs abgeleitet wurden (Wright et al., 1996). Eines der verwendeten Oligonukleotide (Wri1) ist aus einem non-LTR-Element-spezifischen Motiv des 5'-Bereich vor der ersten RT-Domäne abgeleitet (Wright et al., 1996), das reverse Primer (Wri2) stammt aus der vierten Domäne. Der von diesem Oligonukleotidpaar flankierte Bereich ist etwa 600 Bp lang und liegt im Vergleich zum *Ty1-copia*-Amplifikat zum N-terminalen Ende der RT hin verschoben (Fig.C29).

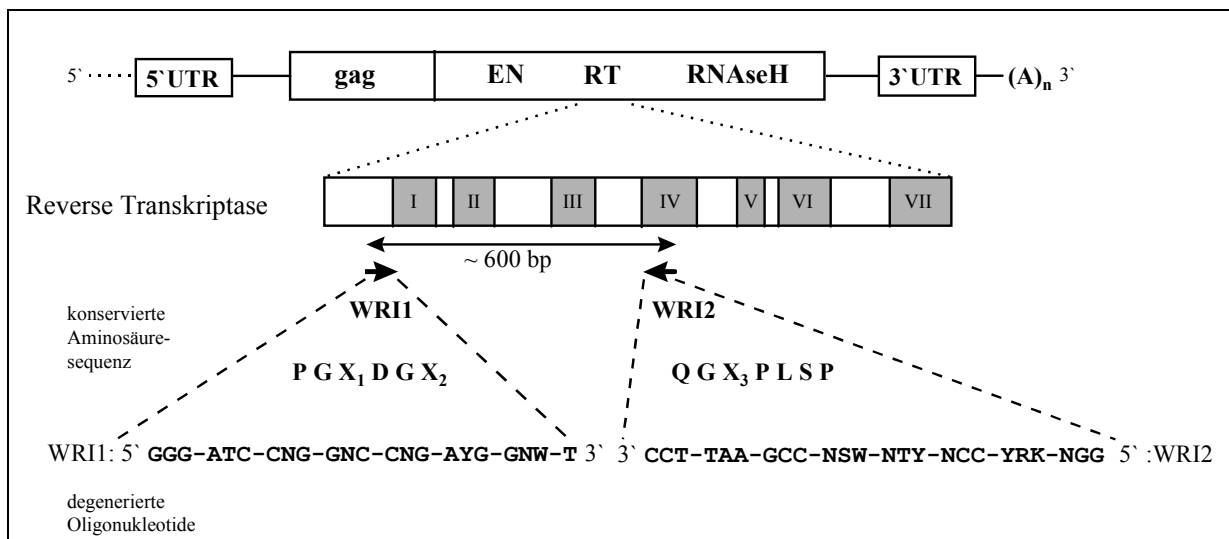


Fig.C29: Position und Sequenz der Oligonukleotidprimer zur Amplifikation von non-LTR-Retrotransposons der LINE-Gruppe.

Struktur eines Retrotransposons der LINE-Gruppe mit *gag*, Endonukleasefunktion (EN), Reverser Transkriptase (RT), RNase H und untranslatierten Regionen (5'/3'-UTR) (oben). Das von Wright et al. (1996) abgeleitete, degenerierte Oligonukleotidpaar flankiert ein ca. 600 Bp langes Fragment aus dem Gen für die Reverse Transkriptase. Die Wri1 Bindestelle liegt im 5'-Bereich vor der ersten Domäne und ist spezifisch für non-LTR-Elemente, während Wri2 aus einer Sequenz der Domäne IV abgeleitet ist (Xiong & Eickbush, 1990).

N = A,C,G,T; Y = C, T; W = A, T; S = C, G; R = A, G; K = G,T

LINE-ähnliche Sequenzen sind in allen untersuchten Leguminosengenomen vorhanden

Mit Hilfe von PCR sollte auch der Nachweis von LINEs in verwandten Arten und weiteren Leguminosen erfolgen. Genomische DNA aus der Kichererbse, weiteren *Cicer*-Spezies sowie andere Leguminosen und zwei Nichtleguminosen, in denen LINE-Elemente bereits nachgewiesen wurden, diente als Templat. In allen untersuchten Pflanzenspezies waren Amplifikate von ca. 600 Bp als Bande erkennbar. Die *Cicer*-Spezies der ersten Kreuzungsgruppe (*C. arietinum*, *C. reticulatum*, *C. echinospermum*) zeigten eine einzelne deutliche 600 Bp Bande, während die zweite Kreuzungsgruppe (*C. bijugum*, *C. pinnatifidum*,

C. judaicum) jeweils nur ein sehr schwaches Produkt aufwies (Fig.C30). Hingegen konnten bei

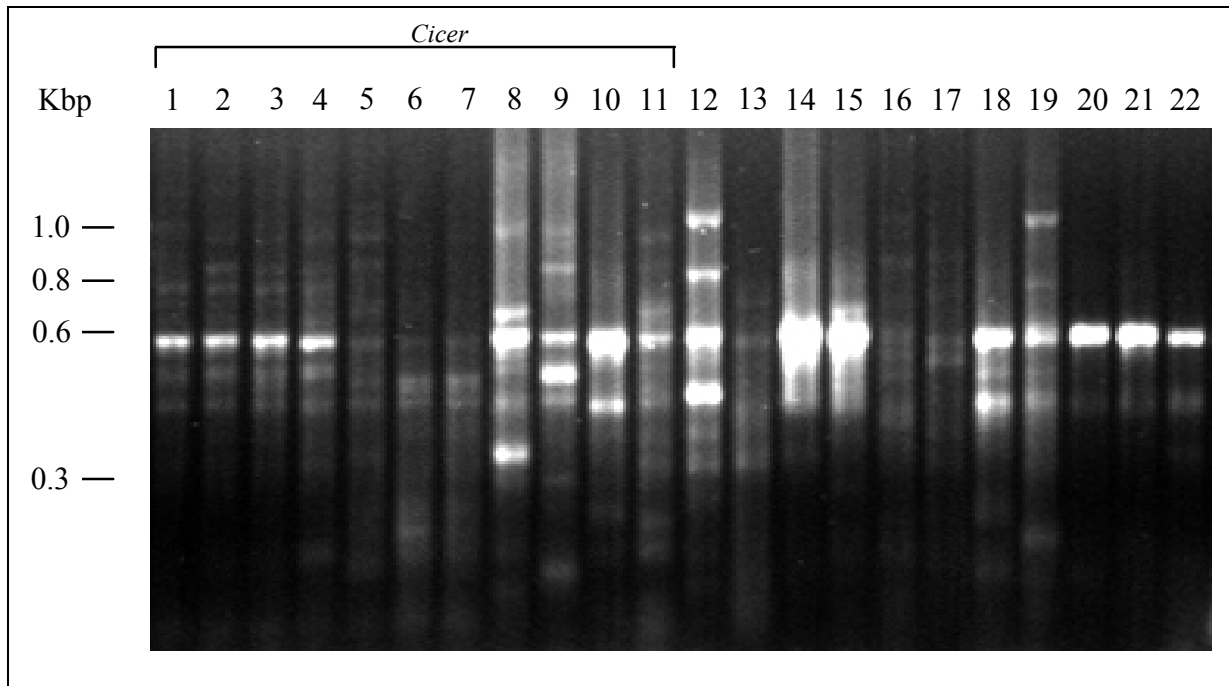


Fig.C30: Amplifikation von LINE-ähnlichen Sequenzen in der Gattung *Cicer* und weiteren Leguminosen. Im Ethidiumbromid-gefärbten Agarosegel sind Banden sichtbar, die mit dem degenerierten Oligonukleotidpaar von Wright et al. (1996) amplifiziert wurden. Als Templat wurde genomische DNA der folgenden Spezies verwendet: *C. arietinum* ILC 482 (1), *C. arietinum* ILC 3279 (2), *C. reticulatum* (3), *C. echinospermum* (4), *C. bijugum* (5), *C. pinnatifidum* (6), *C. judaicum* (7), *C. yamashitae* (8), *C. chorrassanicum* (9), *C. cuneatum* (10), *C. anatolicum* (11), *Lens culinaris* (12), *Vicia faba* (13), *Medicago sativa* (14), *Phaseolus vulgaris* (15), *Phaseolus aureus* (16), *Pisum sativum* (17), *Glycine max* (18), *Lathyrus odoratus* (19), *Arabidopsis thaliana* var. Enkheim (20), *Arabidopsis thaliana* var. Landshut (21) und *Allium cepa* (22). Amplifikate der erwarteten Größe sind in allen untersuchten Spezies erkennbar (Pfeil), jedoch ist die Amplifikation nicht überall optimal.

den übrigen drei Annuellen (*C. yamashitae*, *C. chorrassanicum*, *C. cuneatum*) und der perennuellen Art (*C. anatolicum*) außer einer deutlichen Bande der erwarteten Länge noch weitere 1-3 Signale im Größenbereich von 0.3-1 Kbp detektiert werden. *Vicia faba*, *Pisum sativum* und *Phaseolus aureus* wiesen sehr schwache Signale auf, alle anderen Spezies zeigten entweder deutliche Banden oder außerdem ein bis zwei weitere Amplifikate verschiedener Größen. Alle positiven Kontrollen wiesen das erwartete starke Signal bei 600 Bp auf. Offenbar sind LINE-ähnliche Sequenzen in allen untersuchten Leguminosengenomen vorhanden.

Zehn klonierte Sequenzen der PCR-Amplifikate aus *C. arietinum* (CaLi1-10; AJ411784-AJ411793) und zwei aus *C. reticulatum* (CrLi1-2; AJ411794-AJ411795) wurden näher untersucht. Eine Restriktionsanalyse der Plasmide mit *RsaI* zeigte, daß die ausgewählten Klone, die 75% aller klonierten Fragmente repräsentierten, gemäß ihrer Restriktionsmuster in drei verschiedenen Klassen (I: CrLi1,2; CaLi1,2,10. II: CaLi4,5,7,8. III: CaLi3,6,9) gruppieren. Die Sequenzen der *C. reticulatum*-Klone fügten sich in die Klassen der *C.*

arietinum-Sequenzen ein. Für die AS-Sequenzen der wahrscheinlichsten offenen Leseraster im Bereich der RT wurden ausnahmslos signifikante Homologien zu zahlreichen pflanzlichen non-LTR-Retrotransposons gefunden, wobei alle Sequenzen äquivalente Abschnitte innerhalb der RT-Sequenzen detektierten. Am häufigsten waren Ähnlichkeiten zu Elementen aus *Pinus elliotti* (58-71%; *PeLi5* AJ228333) und *Picea abies* (55-69%; *PaLi1* AJ224369) sowie *Allium cepa* (44-70%; *AcLi3* Y13373), *Antirrhinum majus* (51-80%; *AmLi1* Y13388) und *Beta vulgaris* (46-74%; *BvLi3* Y13368). Zu *Tall-1* aus *A. thaliana* (L47187) bestand 50-66% Ähnlichkeit im entsprechenden Bereich der RT. Auf Nukleotidebene wurden nur vereinzelt und nur sehr kurze homologe Abschnitte zu non-LTR-Retrotransposons gefunden. Die mit dem Oligonukleotidpaar Wri1/2 amplifizierten Sequenzen repräsentieren demnach LINE-ähnliche Elemente oder deren Teilbereiche im Genom der Kichererbse und ihrer nahe verwandten Wildart.

Unter den LINE-ähnlichen Sequenzen aus *Cicer* herrscht große Heterogenität

Die klonierten Fragmente waren 566-602 Bp lang und hatten einen AT-Gehalt von 63-67%, wobei CaLi9 eine Deletion von 12 Bp aufwies und bei CaLi10 im 5'-Bereich 38 Bp fehlten. CaLi 3 und 6 zeigten eine Deletion von 25 Bp, die fehlenden Nukleotide waren durch eine 19 Bp lange fremde Sequenz ersetzt (Fig.C31a). Alle Fragmente enthielten ein offenes Leseraster mit Homologie zum äquivalenten RT-Bereich von non-LTR-Retrotransposons. Die Leseraster waren jedoch jeweils durch 2-4 Stopcodons und 1-6 Leserasterwechsel gekennzeichnet. Verglichen mit den anderen repetitiven Sequenzfamilien der Kichererbse ist die codierende Sequenz nur wenig konserviert. Ein Sequenzvergleich aller Fragmente zeigte eine relativ starke Heterogenität (45-99.8% Homologie auf Nukleotidebene). Drei Gruppen ließen sich bilden, von denen die größte insgesamt 9 Sequenzen umfaßt, die auf Nukleotidebene zu 83-99.8% untereinander homolog sind, darunter auch die beiden *C. reticulatum*-Fragmente. Die zweite Gruppe bestand nur aus CaLi3 und CaLi6, die zu 91% identisch waren, aber nur 47-51% Homologie zur großen ersten Gruppe aufwiesen. Die letzte Sequenz CaLi9 wies eine sehr geringe Ähnlichkeit (45-46%) zu jeder der beiden erstgenannten Gruppen auf. Ein Sequenzvergleich der wahrscheinlichsten Aminosäuresequenzen spiegelte diese Verhältnisse mit einer Spanne von 23-99% wieder (Fig.C31a). Dies bestätigt, daß die sequenzierten Fragmente zwar sehr heterogen erscheinen, jedoch ausnahmslos RT-Sequenzen von non-LTR-Elementen ähneln.

Die genetische Distanz der genannten Sequenzen zueinander und zu den LINE-ähnlichen Elementen *Tall-1* (*A. thaliana*), *del2* (*Lilium speciosum*; Z17425), *BvLi3* (*B. vulgaris*) und

Fig.C31: Sequenzvergleich der wahrscheinlichsten Leseraster aus den klonierten LINE-Sequenzen. (a) Vergleich der sequenzierten PCR-Fragmente aus *C. arietinum* und *C. reticulatum*. Die Oligonukleotidsequenzen wurden nicht in den Sequenzvergleich einbezogen und die Domänen I-IV der Reversen Transkriptase (Xiong & Eickbush, 1990) sind hellgrau unterlegt. Stark konservierte Aminosäuren sind dunkelgrau markiert, Striche zeigen Lücken, die für ein optimales Alignment eingefügt wurden, und unterstrichene Sequenzen stellen Deletionen oder stark abweichende Sequenzen dar. In einigen Fällen waren Leserasterwechsel notwendig, um die Sequenzen vergleichen zu können. (b) Genetische Distanz der Aminosäuresequenzen aus (a) im Vergleich zu weiteren LINEs aus *Beta vulgaris* (*BvLi*; Y13368), *Picea abies* (*PaLi*; AJ224369), *Arabidopsis thaliana*

C. Ergebnisse

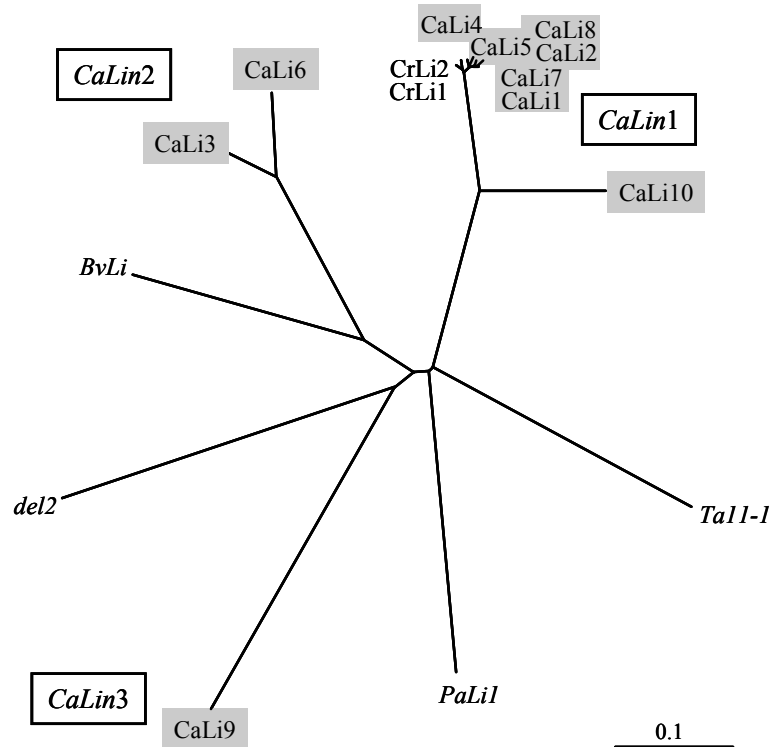
(*Ta11-1*; L47193) und *Lilium speciosum* (*del2*; Z17425) in Form eines wurzellosen Dendrogramms (Saitou & Nei, 1987). Die horizontale Linie repräsentiert Prozent Divergenz (/100). Deutlich wird die Heterogenität der Sequenzen aus *C. arietinum*, die sich in drei Subfamilien (*CaLin1-3*) einteilen lassen.

```

                                I                               II
CaLi1  LNPIFYFYFFVFW--VVKIEN-NYCVWLLIGVFPSTPNITNIALIPKGDQKTVKDWRPALCNMIYKIVTKVLANI LKHVLHKCVSNNQ-AF
CaLi7  LNPIFYFYFFVFW--VVKIEN-NYCVWLLIGVFPSTPNITNIALIPKGDQKTVKDWRPALCNMIYKIVTKVLANI LKHVLHKCVSNNQ-AF
CaLi2  MNPIFYFYFFDFVW--VVKIEN-NYCVWLLIGVFPSTPNITNIALIPKGDQKTVKDWRPALCNMIYKIVTKVLANI LKHVLHKCVSNNQ-AF
CaLi8  LNPIFYFYFFVFW--VVKIEN-NYCVWLLIGVFPSTPNITNIALIPKGDQKTVKDWRPALCNMIYKIVTKVLANI LKHVLHKCVSNNQ-AF
CaLi4  LNPIFYFYFFVFW--VVKIEN-NNCVWLLIGVFPSTPNITNIALIPKGDQKTVKDWRPALCNMIYKIVTKVLANI LKHVLHKCVSNNQ-AF
CaLi5  MNPIFYFYFFVFW--VVKIEN-NYCVWLLIGVFPSTPNITNIALIPKGDQKTVKDWRPALCNMIYKIVTKVHANI LKHVLHKCVSNNQ-AF
CrLi2  LNPIFYFYFFVFW--VVKIEN-NYCVWLLIGVFPSTPNITNIALIPKGDQKTVKDWRPALCNMIYKIVAKVLANI LKHVLHKCVSNNQ-AF
CrLi1  MNPIFYFYFFVFW--VVKIEN-NYCVWLLIGVFPSTPNITNIALIPKGDQKTVKDWRPALCNMIYKIVTKVLANI LKHVLHKCVSNNQ-AF
CaLi10 -----L--TLGFQ--ECCDWLLTGVPSTLNVNITNIALIPKGDKRKTMKDWQPIALCNVNYKIVAKVLSNRLKHLKHLKCIISINQSTF
CaLi3  ----LHFRFVKEFWENIKGDFM-FLCDFHANGRLVKESNCSFIVLIPKKNPQVLYDCRPISLVGRMYKVLKSVLANRLSQVIGSLIVESQSTF
CaLi6  ----LHFGFVKEFWENIKGDFMFLGDFHANGRLVK-SNYSFIVLIPKKNLQVLSDFRMSLVGCMYKALSVDLANRLSQVIGSLIVESQSAF
CaLi9  ---IDGCSFKHSCDMVGLLEN-PMHQFCFNGWIMPKYNSNIVAFILKTPGADMIIECFRPIMAN-FKFKVLSKLVLDMLAIAPKIIEYH-REF

                                III                               IV...
CaLi1  VLNRSILDNAMTTLEIVHHMKTTRS RVGDVALKLDINKTYDCID-SYLREVLVLD-FANQWVNWIIILCVESVEYNIIVNDCVGPITP-RGL-
CaLi7  VLNRSILDNAMTTLEIVHHMKTTRS RVGDVALKLDINKTYDCID-SYLREVLVLD-FANQWVNWIIILCVESVEYNIIVNDCVGPITP-RGL-
CaLi2  VLNRSILDNAMTTLEIVHHMKTTRS RVGDVALKLDINKTYDCID-SYLREVLVLD-FANQWVNWIIILCVESVEYNIIVNDCVGPITP-RGL-
CaLi8  VLNRSILDNAMTTLEIVHHMKTTRS RVGDVALKLDINKTYDCID-SYLREVLVLD-FANQWVNWIIILCVESVEYNIIVNDCVGPITP-RGL-
CaLi4  VLNRSILDNAMTTLEIVHHMKTTRS RVGDVALKLDINKTYDCID-SYLREVLGLD-FANQWVNWIIILCVESVEYNIIVNDCVGPITP-RGL-
CaLi5  VLNRSILDNAMTTLEIVHHMKTTRS RVGDVALKLDINKTYDCID-SYLREVLGLD-FANQWVNWIIILCVESVEYNIIVNDCVGPITP-RGLY
CrLi2  VLNRSILDNAMTTLEIVHHMKTTRS RVGDVALKLDINKTYDCID-SYLREVLGLD-FANQWVNWIIILCVESVEYNIIVNDCVGPITP-RGL-
CrLi1  VLNRSILDNAMTTLEIVHHMKTTRS RVGDVALKLDINKTYDCID-SYLREVLGLD-FANQWVNWIIILCVESVEYNIIVNDCAGPIIP-RGL-
CaLi10 VPRFILDNAMAALTEIVHYMKS KTHGRVGDVALKLDINKAYDCIDREYPREVLNLTGFAD-WVN-IMLCFESTIEYNAIVNDCVGPITP-RRGLR
CaLi3  VKGRQILDGILIANEVV--MRQRRRKE-LILFKVDFEKAYDSVEWDYLELVMTKMGFE-K-RKWILKCVCSASVSVLENGSPTRFVMRKGLR
CaLi6  VKGRQI-VGILIANEVV--MRQRRRKS-LILFKVDFEKAYDSVEWDYLELVMAKMDILEK-RKWILECVGTSVSVLVNESPT-EFVMQRGLR
CaLi9  TKGRHIS-CIFTTSHAINLLD-KSFGCN--LAFKIYIKKAFDTLYRGLFIQTLOAFANNTFCNWILQAKLSLLV----NGKVVDLISLEG
    
```

a



b

C31

[Teil3.doc](#)

PaLi1 (*P.abies*) wurde in Form eines Dendrogramms (Saitou& Nei, 1987) dargestellt (Fig.C31b). Erkennbar war eine homogene, große Gruppe von *Cicer*-Sequenzen, während CaLi3 und CaLi6 bzw. CaLi9 artfremden Elementen ähnlicher waren als anderen Kichererbsen-LINEs. Offenbar wurden drei verschiedene Subfamilien von LINE-ähnlichen Sequenzen aus der Kichererbse isoliert, die untereinander weniger als 36% Ähnlichkeit der wahrscheinlichsten Leseraster aufweisen. Sie wurden unter dem Namen *CaLin* (*Cicer arietinum* LINE-ähnliche-Elemente) zusammengefaßt und mit den Bezeichnungen *CaLin1*,-*2*,-*3* für die einzelnen Subfamilien unterschieden (Fig.C31b). Die beiden Sequenzen aus *C. reticulatum* (CrLi1, CrLi2) stellen enge Verwandte der größten Subfamilie *CaLin1* dar.

C.II.3.2. Genomische Organisation der LINE-ähnlichen Sequenzen in *Cicer arietinum*

***CaLin*-Sequenzen der Kichererbse sind dispers organisiert und von geringer Abundanz**

Um die genomische Organisation und Abundanz der LINE-ähnlichen Sequenzen aus der Kichererbse zu ermitteln, wurde eine Southern-Analyse mit unterschiedlich restringierter genomischer DNA durchgeführt. Als Sonde diente die 600 Bp große PCR-Bande aus *C. arietinum* ILC 3279.

Sichtbar war ein einfaches Muster aus 1-3 Banden pro Restriktion im Molekulargewichtsbereich von 0.3-8 Kbp auf einem leichten hochmolekularen Schmier, was auf die disperse Verteilung multipler Kopien von LINE-ähnlichen Sequenzen hindeutet (nicht dargestellt). Im Vergleich zu den Southern-Analysen der übrigen Sequenzfamilien waren hier die Signalintensitäten relativ niedrig, d.h. die detektierten Sequenzen sind nur mittel- oder niedrigabundant vorhanden. Da das methylierungssensitive Isoschizomerpaar *HpaII/MspI* Fragmente im Größenbereich von 0.5-1 Kbp erzeugte, kann man nicht von einer generell hohen Methylierung der Sequenzen ausgehen.

Die zu den *Wri1/2*-Amplifikaten homologen Sequenzen von *CaLin* stellen eine disperse repetitive Sequenzfamilie mit verhältnismäßig niedriger Abundanz im Kichererbsengenom dar. Da *CaLin*-Sequenzen sehr heterogen sind, ist die Anwesenheit weiterer Kopien mit divergierenden Sequenzen im Genom wahrscheinlich.

C.II.3.3. Chromosomale Lokalisation von *CaLin*-Kopien in der Kichererbse

Um die chromosomale Lokalisation von *CaLin* zu untersuchen, wurde ein Gemisch aller 10 klonierten Fragmente (3.1.) für eine FISH verwendet. Die Fragmente wurden jeweils gesondert markiert, umfaßten alle drei *CaLin*-Subfamilien und sollten so eine maximale Anzahl von Loci auf den Chromosomen detektieren. *CaLin*-Sequenzen konnten nur auf drei der acht Chromosomenpaare nachgewiesen werden und zeigten nur geringe Signalintensitäten. Die Loci lagen auf den beiden größten Chromosomen A und B am Übergang eines distalen Endes des pericentrischen Heterochromatinblocks zum schwächer blaugefärbten Euchromatin sowie auf einem weiteren der mittelgroßen Chromosomenpaare (C-G; Fig.C32a-e). In einzelnen

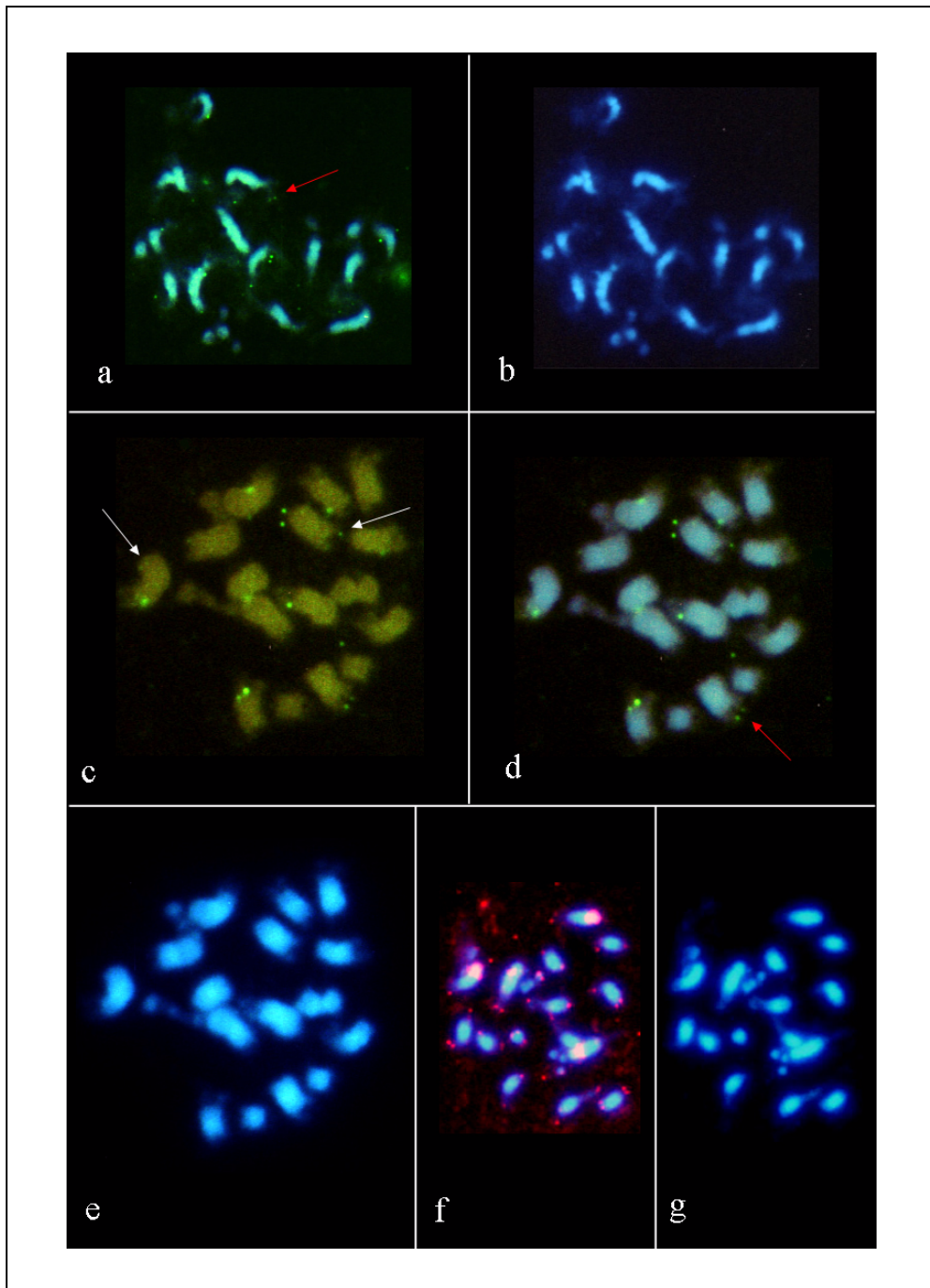


Fig.C32: Lokalisation der *CaLin*-Sequenzen und Telomermotive auf Metaphasechromosomen von *C. arietinum*.

a,b,d,e,f,g) Metaphasechromosomen nach DAPI-Färbung (blaue Fluoreszenz). a,c,d) Dieselben Metaphasechromosomen nach FISH mit einem markierten Gemisch aus 10 klonierten Fragmenten (a,d: Computerüberlagerung). Die Signale sind als grüne Fluoreszenz auf drei Chromosomen im distalen Teil des pericentrischen Heterochromatins und in benachbarten, euchromatischen Bereichen sichtbar (rote Pfeile in a,d). f) Metaphase aus g) nach Hybridisierung mit einer Telomersonde, die das Motiv $(CCCTAAA)_n$ enthält. Neben starken roten Fluoreszenzsignalen im pericentrischen Heterochromatin von A und B sind Doppelsignale im Telomerbereich der meisten Chromosomen erkennbar.

Metaphasen erschien ein zweites Doppelsignal am gegenüberliegenden Ende des Heterochromatinblocks auf Chromosom B (Fig.C32c, weiße Pfeile). Einige der doppelpunktförmigen Signale auf beiden Chromatiden schienen weiter in den benachbarten euchromatischen Bereich verschoben zu sein (rote Pfeile). Die Hauptloci der hochrepetitiven Sequenzen *CaRep*, *CaSat1* und *CaSat2* im zentralen Teil der DAPI-positiven Heterochromatinblöcke sowie die sekundäre Einschnürung auf Chromosom A zeigten keine Signale der LINE-ähnlichen Elemente, jedoch kann eine Durchmischung mit *CaRep* oder den beiden mittelrepetitiven dispersen Sequenzen im distalen Bereich der Blöcke und im angrenzenden Euchromatin nicht ausgeschlossen werden. Die Signalintensität weist auf nur niedrige Abundanz der Zielsequenzen im Kichererbsengenom hin. Trotz der Verwendung eines Fragmentgemischs als Sonde kann hierbei die Heterogenität der *CaLin*-Familie für die geringe Zahl der detektierten Kopien mitverantwortlich sein.

C.II.3.4. *CaLin*-ähnliche Sequenzen im Genus *Cicer* und in verwandten Spezies

Abundanz und Organisation von *CaLin*-Sequenzen sind im Genus *Cicer* größtenteils konserviert - homologe Sequenzen kommen auch in anderen Spezies vor

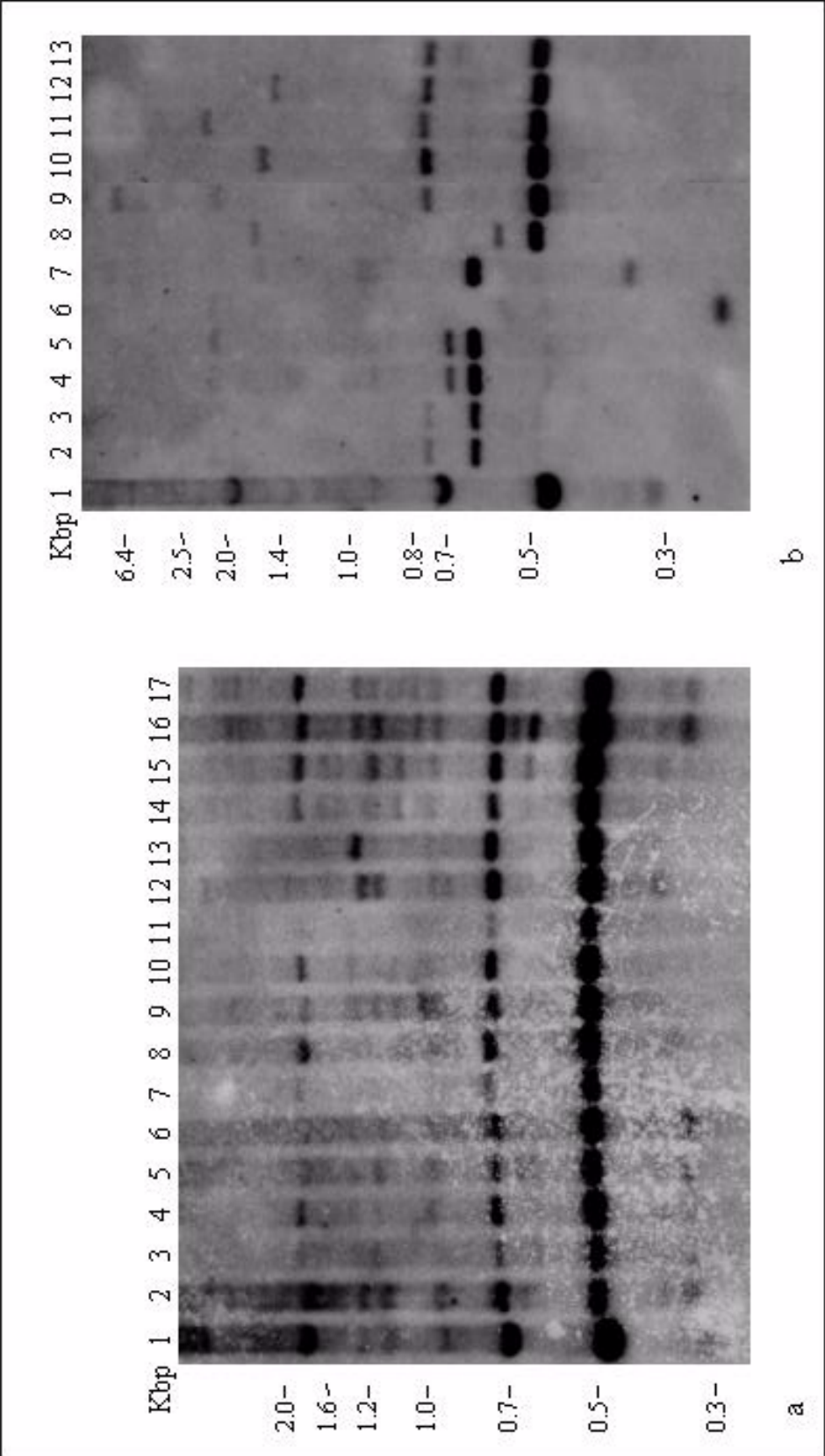
Um herauszufinden, ob die im Genus *Cicer* vorhandenen LINE-Sequenzen homolog zur *CaLin*-Familie aus Kichererbse sind, wie sie organisiert und ob sie auf den Genus *Cicer* beschränkt sind, wurde eine Southern-Analyse durchgeführt. Genomische DNA aus verschiedenen *Cicer*-Spezies, weiteren Leguminosen und anderen Dicotyledonen wurde gegen die 600 Bp-Bande der Wri1/2-PCR aus *C. arietinum* hybridisiert. In allen Spezies der Gattung *Cicer* waren zwei deutliche Banden bei 0.5 und 0.7 Kbp erkennbar (Fig.C33a). Auf einem leichten Hintergrundschmier konnten weitere 2-8 Banden von niedrigerer Intensität detektiert werden, die für einzelne Spezies, Gruppen von Spezies oder die beiden verwendeten Akzessionen (bei *C. cuneatum*, Spur 12 und 13) jeweils unterschiedliche Muster aufwiesen. Die Hybridisierungsmuster von Spezies der zweiten Kreuzungsgruppe mit nur schwachen PCR-Amplifikaten zeigten (3.1.) keine Intensitätsunterschiede zu den übrigen *Cicer*-Spezies.

CaLin-homologe Sequenzmotive sind in allen *Cicer*-Spezies in ähnlicher Abundanz vorhanden und zeigen die gleiche genomische Organisation. Die schwächeren und heterogeneren Signale

Fig.C33: Genomische Organisation von *CaLin*-homologen Sequenzen in der Gattung *Cicer* und verwandten Arten.

(a) Autoradiogramm eines Southern-Blots mit *RsaI*-restringierter genomischer DNA aus *C. arietinum* ILC 3279 (1), *C. arietinum* ILC 482 (2), *C. reticulatum* PI 489777 (3), *C. reticulatum* ILWC 21 (4), *C. echinospermum* ILWC 39 (5), *C. echinospermum* ILWC 35 (6), *C. bijugum* ILWC 42 (7), *C. bijugum* ILWC 62 (8), *C. pinnatifidum* ILWC 204 (9), *C. judaicum* ILWC 57 (10), *C. judaicum* ILWC 56 (11), *C. cuneatum* ILWC 40 (12), *C. cuneatum* ILWC 37 (13), *C. chorassanicum* ILWC 146 (14), *C. chorassanicum* ILWC 23 (15), *C. yamashitae* ILWC 53 (16) und *C. yamashitae* ILWC 214 (17). Als Sonde diente das mit Wri1/2 amplifizierte Fragment aus *C. arietinum* ILC 3279. (b) Autoradiogramm eines weiteren Southern-Blots mit *RsaI*-restringierter

genomischer DNA aus *C. arietinum* ILC 3279 (1), *Lathyrus odoratus* (2,3), *Pisum sativum* (4,5), *Lens culinaris* (6), *Vicia faba* (7), *Medicago sativa* (8), *Glycine max* (9), *Phaseolus vulgaris* (10), *Phaseolus aureus* (11), *Lycopersicon esculentum* (12) und *Beta vulgaris* (13). Die Abundanz und Organisation von *CaLin*-ähnlichen Sequenzen im Genus *Cicer* ist relativ konserviert, die Sequenzfamilie ist ebenfalls in anderen Dikotylen vorhanden.



werden möglicherweise durch von der Sondensequenz divergierende, *CaLin*-ähnliche Sequenzen verursacht und sind speziesspezifisch unterschiedlich organisiert. Alle übrigen Spezies - auch die beiden Nichtleguminosen - wiesen ebenfalls 2-4 deutliche Signale im Größenbereich zwischen 0.25 und 8 Kbp auf einem leichten Hintergrund auf, die innerhalb einer Art konserviert waren (Fig.C33b). *CaLin*-ähnliche Sequenzen waren wie die beiden mittelrepetitiven Familien *CaDis* und *CaTy* nicht auf die Gattung *Cicer* beschränkt, sondern in ähnlicher Abundanz und mit speziesspezifisch abweichender genomischer Organisation auch in anderen Leguminosen vorhanden. Ihr Vorkommen in Tomate und Rübe weist ferner darauf hin, daß verwandte Motive nicht auf Leguminosen beschränkt sind.

C.III. Das DNA-Transposon *CaEn/Spm*

DNA-Transposons sind aufgrund ihres Transpositionsmechanismus meist weit weniger abundant als Retroelemente, deren Kopienzahl sich mit jeder Bewegung im Genom amplifiziert. Daher ist die Wahrscheinlichkeit, Vertreter dieser repetitiven Sequenzklasse unter den hochabundanten Sequenzen in einer genomischen Kichererbsenbank zu finden, relativ gering. Ähnlich wie bei den übrigen mittel- und niedrigrepetitiven Sequenzen (siehe II.) sollte daher eine Detektion von Transposonsequenzen über PCR erfolgen. Im Gegensatz zu den erwähnten Retrotransposonklassen existierten jedoch zum Zeitpunkt dieser Arbeit in der Literatur keine Informationen über degenerierte Oligonukleotidprimer für die Isolation von DNA-Transposons.

C.III.1. Nachweis von Transposons des *En/Spm*-Typs in pflanzlichen Genomen

C.III.1.1. Degenerierte Oligonukleotide aus der Transposaseregion von *En/Spm*-Elementen

Optimales Design eines Primerpaars mit Hilfe des CODEHOP-Prinzips

In Anlehnung an bereits existierende Oligonukleotidpaare für Retrotransposons (siehe II.) sollte ein degeneriertes Primerpaar zur Detektion von DNA-Transposons (Klasse II) hergestellt werden. Oligonukleotide, die möglichst viele divergierende Kopien eines mobilen Elements detektieren sollen, müssen aus dessen konservierten Bereichen abgeleitet werden. Bei einem Transposon der *En/Spm*-Gruppe bieten sich dafür die codierenden Sequenzen der zur Transposition notwendigen Proteine (z.B. der Transposase) aus dem internen Bereich zwischen den TIRs an (Fig.C34a). Die in öffentlichen Datenbanken vorhandenen Aminosäuresequenzen von *En/Spm* (*Zea mays*, S29329) und den *En/Spm*-ähnlichen Elementen aus *Allium cepa* (Zwiebel, AJ223064), *Petunia hybrida* (Petunie, AF009516), *Antirrhinum majus* (Löwenmäulchen, X57297) und *Arabidopsis thaliana* (Ackerschmalwand, AC005897/AC005957) sowie das offene Leseraster einer zufällig klonierten Sequenz der Kichererbse mit hoher Homologie zu *En/Spm*-ähnlichen Transposasen (*Ca3*; AJ406381) wurden verglichen und zeigten drei Bereiche mit konservierten Motiven innerhalb der

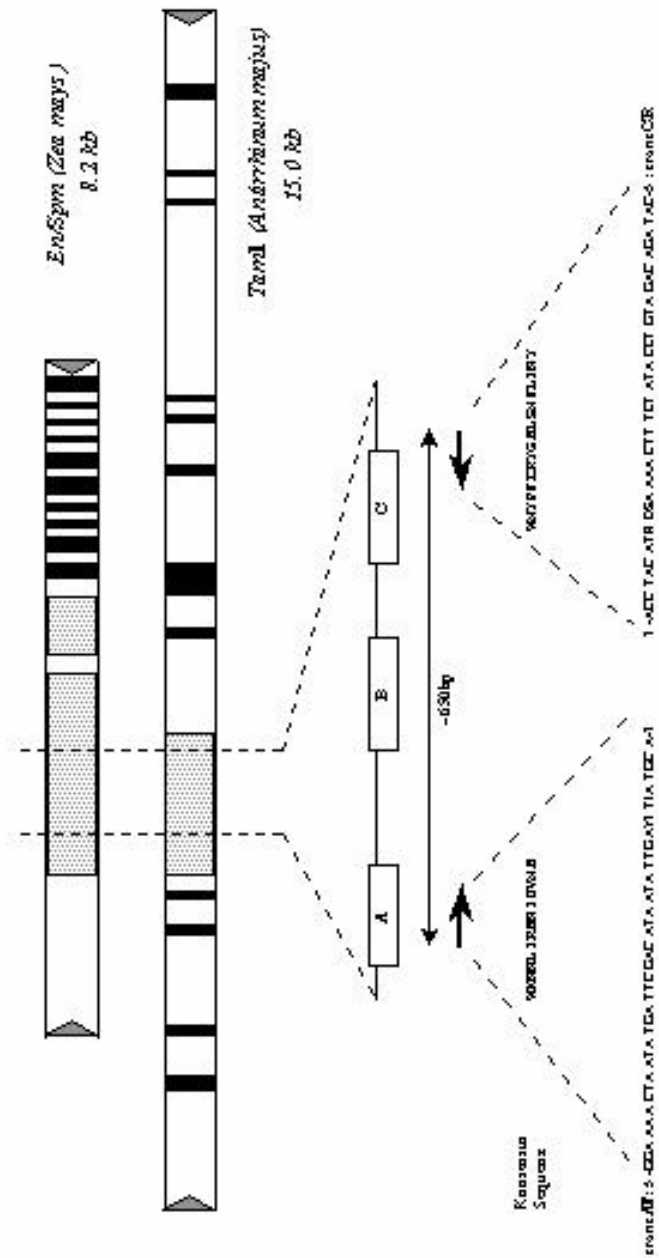
Transposaseregion (Boxen A, B und C in Fig.C34a). Aus den jeweiligen Konsensussequenzen in Box A und C wurde mit Hilfe der CODEHOP-Software je ein Oligonukleotid abgeleitet (Fig.C34b). Dieses Programm wurde speziell zur Ableitung von Primern aus der Konsensussequenz verwandter Proteine mit Hilfe von reverser Translation entwickelt (Rose et al., 1998). Das CODEHOP („**C**onsensus **D**Egenerate **H**ybrid **O**ligonucleotide **P**rimers“-Prinzip berücksichtigt dabei besonders den 3`-Bereich der Oligonukleotidprimer, der für eine optimale Elongation komplementär zum Templat sein muß. Der 3`-Bereich sollte Sequenzen umfassen, die für stark konservierte Motive bzw. Aminosäuren mit möglichst wenig degenerierten Codons codieren (z.B. Methionin, Tryptophan). Die konservierten Bereiche müssen nicht sehr lang sein, eine 3-4 Aminosäuren lange hochkonservierte Abfolge reicht bereits zum Design des sog. „3`-cores“ des Oligonukleotids aus. Das längere sog. „5`-clamp“-Ende des entworfenen Oligonukleotids läßt Fehlpaarungen mit dem Templat zu und dient in den späteren PCR-Zyklen zur Stabilisierung des Oligonukleotids bei dessen Anlagerung, so daß bei hohen Annealing-Temperaturen gearbeitet werden kann. Durch die möglichst geringe Zahl degenerierter Nukleotidpositionen im kritischen 3`-Bereich des Oligonukleotids kann eine hohe Spezifität der Reaktion sichergestellt werden, ohne die Chance zu verringern, auch noch divergente Kopien des Elements zu amplifizieren.

Die mit Hilfe von CODEHOP abgeleitete Sequenz des 40 Bp langen transAF- und des 39 Bp langen transCR-Oligonukleotids ist in Fig.C34a dargestellt. Die hochkonservierte Aminosäuresequenz WMYP aus Box C („3`-core“) erlaubte eine sichere Ableitung des reversen Oligonukleotids, auch wenn in der benachbarten Sequenz eine Insertion in der Kichererbsen- und Petuniensequenz („5`-clamp“) erfolgte (Fig.C34b). Das abgeleitete Oligonukleotidpaar sollte einen ca. 650 Bp langen Bereich aus genomischen *En/Spm*-Elementen amplifizieren können.

Fig.C34: Ableitung von Oligonukleotidsequenzen aus konservierten Regionen des Transposase-Gens in *En/Spm*-ähnlichen Transposons.

(a) Homologe Regionen des offenen Leserasters der Transposase (TPase, graue Kästen) von *En/Spm* (*Zea mays*) und *Tam1* (*Antirrhinum majus*) sind durch senkrechte, gepunktete Linien markiert. Die Exons des zweiten Leserasters sind als schwarze Kästen dargestellt, die terminalen inversen Sequenzen („terminal inverted repeats“; TIRs) durch graue Dreiecke. Konservierte Sequenzmotive der TPase aus fünf *En/Spm*-ähnlichen Elementen (*Allium cepa* [AJ223064], *Petunia hybrida* [AF009516], *Zea mays* [S29329], *Antirrhinum majus* [X57297], *Arabidopsis thaliana* [AC005897, AC005957]) und einer Sequenz aus *C. arietinum* (AJ406381) sind als Kasten A-C schematisiert. Die Position der Oligonukleotidsequenzen transAF/ transCR, die ein ca. 650 Bp langes Fragment flankieren, sind als Pfeile dargestellt, darunter die zu Grunde liegende Aminosäuresequenz und die Nukleotidsequenz der degenerierten Primer. Das Schema ist nicht maßstabsgetreu. (I= Inosin; Y= C, T; S= C, G; R= A, G; D= A, T, G). (b) Vergleich der konservierten Motive aus den Regionen A und C in *Allium cepa*, *Petunia hybrida*, *Zea mays*, *Antirrhinum majus* und *Arabidopsis thaliana* und Bildung einer Konsensussequenz. Die 3`-Sequenz der degenerierten Oligonukleotide liegt im Bereich der stärksten Konservierung (grau unterlegt).

C34



3



b

C.III.1.2. *En/Spm*-Transposons in *Cicer* und anderen Gattungen

Amplifikate einheitlicher Größe wurden in über 30 Pflanzenarten nachgewiesen

Das degenerierte Oligonukleotidpaar (transAF/CR) wurde zur PCR-Amplifikation von genomischer DNA aus 33 verschiedenen Pflanzenspezies verwendet (Tab.C2). Die PCR-Bedingungen waren dabei mit einer Annealing-Temperatur von 58°C stringent gewählt, um die

Tab.C2: Liste der zur PCR-Amplifikation mit dem degenerierten Oligonukleotidpaar transAF/ transCR verwendeten Spezies und das Ergebnis der Amplifikation.

Spezies	Taxonomische Einordnung	Amplifikationsprodukt mit transAF/CR (650 Bp)
<i>Cicer anatolicum</i>	Leguminosae, Dicotyledoneae	+
<i>Cicer arietinum</i> C 104	Leguminosae, Dicotyledoneae	+
<i>Cicer arietinum</i> ILC 482	Leguminosae, Dicotyledoneae	+
<i>Cicer arietinum</i> ILC 3279	Leguminosae, Dicotyledoneae	+
<i>Cicer arietinum</i> WR 315	Leguminosae, Dicotyledoneae	+
<i>Cicer bijugum</i>	Leguminosae, Dicotyledoneae	+
<i>Cicer chorassanicum</i>	Leguminosae, Dicotyledoneae	+
<i>Cicer cuneatum</i>	Leguminosae, Dicotyledoneae	+
<i>Cicer echinospermum</i>	Leguminosae, Dicotyledoneae	+
<i>Cicer judaicum</i>	Leguminosae, Dicotyledoneae	+
<i>Cicer pinnatifidum</i>	Leguminosae, Dicotyledoneae	+
<i>Cicer reticulatum</i>	Leguminosae, Dicotyledoneae	+
<i>Cicer yamashitae</i>	Leguminosae, Dicotyledoneae	+
<i>Glycine max</i>	Leguminosae, Dicotyledoneae	+
<i>Lathyrus odoratus</i>	Leguminosae, Dicotyledoneae	+
<i>Lens culinaris</i>	Leguminosae, Dicotyledoneae	+
<i>Medicago sativa</i>	Leguminosae, Dicotyledoneae	+
<i>Phaseolus aureus</i>	Leguminosae, Dicotyledoneae	+
<i>Phaseolus vulgaris</i>	Leguminosae, Dicotyledoneae	+
<i>Pisum sativum</i>	Leguminosae, Dicotyledoneae	+
<i>Vicia faba</i>	Leguminosae, Dicotyledoneae	+
<i>Nicotiana tabacum</i>	Solanaceae, Dicotyledoneae	+
<i>Nicotiana atropurpurea</i>	Solanaceae, Dicotyledoneae	+
<i>Lycopersicon esculentum</i>	Solanaceae, Dicotyledoneae	+
<i>Arabidopsis thaliana</i>	Brassicaceae, Dicotyledoneae	+
<i>Brassica napus</i>	Brassicaceae, Dicotyledoneae	+
<i>Beta vulgaris</i>	Chenopodiaceae, Dicotyledoneae	+
<i>Pelargonium spec.</i>	Geraniaceae, Dicotyledoneae	+
<i>Triticum aestivum</i>	Poaceae, Monocotyledoneae	+
<i>Triticum urartu</i>	Poaceae, Monocotyledoneae	+
<i>Zea mays</i>	Poaceae, Monocotyledoneae	+
<i>Hordeum vulgare</i>	Poaceae, Monocotyledoneae	+
<i>Allium sativum</i>	Alliaceae, Monocotyledoneae	+
<i>Tillandsia usneoides</i>	Bromeliaceae, Monocotyledoneae	+
<i>Musa spec.</i> (AAA) (Eßbanane)	Musaceae, Monocotyledoneae	+
<i>Musa acuminata</i> x <i>Musa</i>	Musaceae, Monocotyledoneae	-

<i>balbisiana</i> (Kochbanane)		
<i>Ascochyta rabiei</i>	Ascomycetes (Fungi)	-

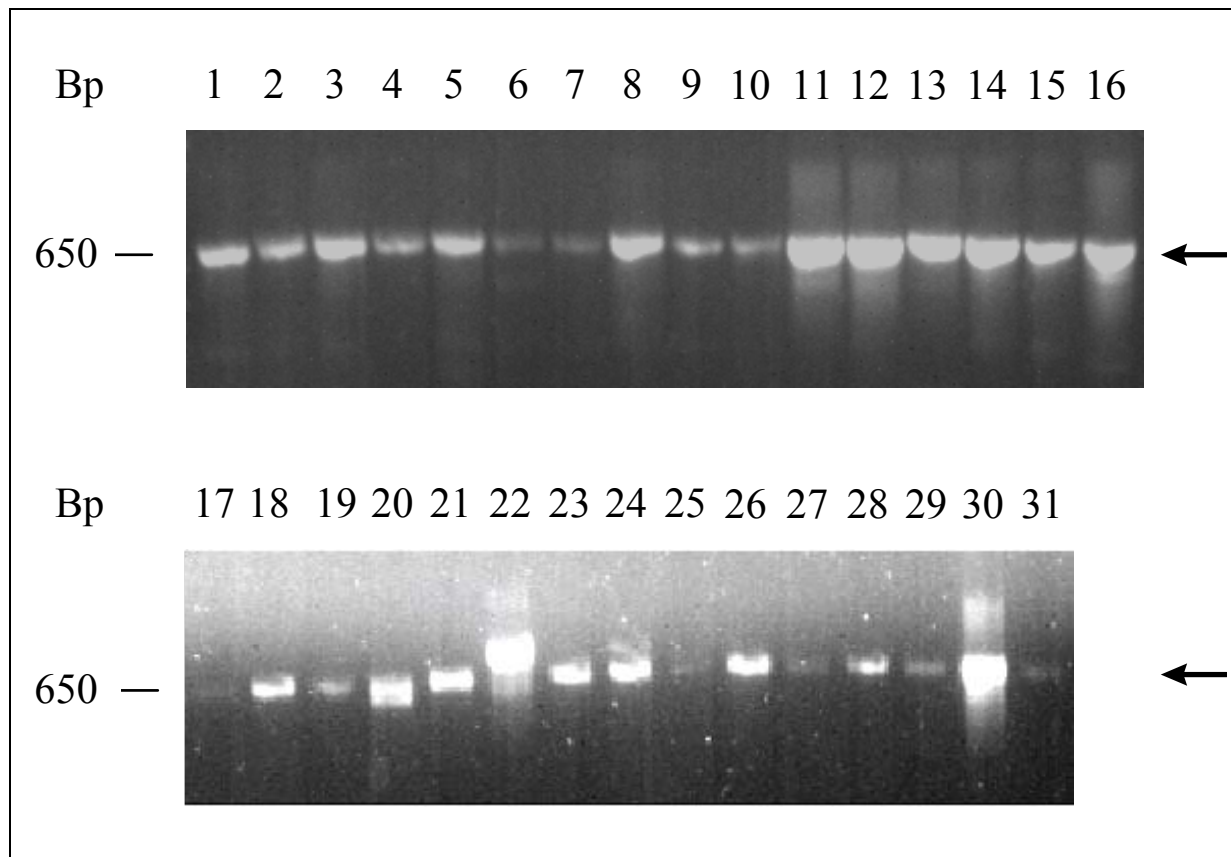


Fig.C35: Amplifikation von *En/Spm*-ähnlichen Sequenzen mit dem degenerierten Oligonukleotidpaar transAF/ transCR in verschiedenen Pflanzenarten.

Ethidiumbromid-gefärbtes Agarosegel mit den durch die degenerierten Primer erzeugten PCR-Produkten von ca. 650 Bp (Pfeile). Als Templat wurde genomische DNA folgender Spezies verwendet: *Cicer arietinum* C104 (1), *C. arietinum* ILC482 (2), *C. arietinum* ILC3279 (3), *C. reticulatum* (4), *C. echinospermum* (5), *C. bijugum* (6), *C. pinnatifidum* (7), *C. judaicum* (8), *C. chorassanicum* (9), *C. yamashitae* (10), *C. cuneatum* (11), *C. anatolicum* (12), *Lens culinaris* (13), *Pisum sativum* (14), *Phaseolus vulgaris* (15), *Phaseolus aureus* (16), *Glycine max* (17), *Medicago sativa* (18), *Lathyrus odoratus* (19), *Lycopersicon esculentum* (20), *Nicotiana tabacum* (21), *Arabidopsis thaliana* (22), *Beta vulgaris* (23), *Brassica napus* (24), *Tillandsia usneoides* (25), *Pelargonium spec.* (26), *Hordeum vulgare* (27), *Triticum aestivum* (28), *Zea mays* (29), *Allium sativum* (30) und *Musa spec.* (AAA; 31). Die hier nicht gezeigte Wasserkontrolle wies keine Signale auf.

Bindung der Primer an weniger homologe Sequenzen auszuschließen. Dieser Ansatz führte in 32 der 33 getesteten Pflanzenarten aus 20 verschiedenen Gattungen zu Amplifikationsprodukten in der erwarteten Länge von etwa 650 Bp. Nach Auftrennung des PCR-Ansatzes in einem Agarosegel konnten die jeweils singulären Banden mit Ethidiumbromid nachgewiesen werden (Fig.C35). Lediglich die DNA der Kochbanane (*Musa acuminata* x *Musa balbisiana*) und des Ascomyceten *Ascochyta rabiei* erzeugten unter den gegebenen Bedingungen keine sichtbare Bande.

C.III.1.3. Identifikation der Amplifikate durch Sequenzierung

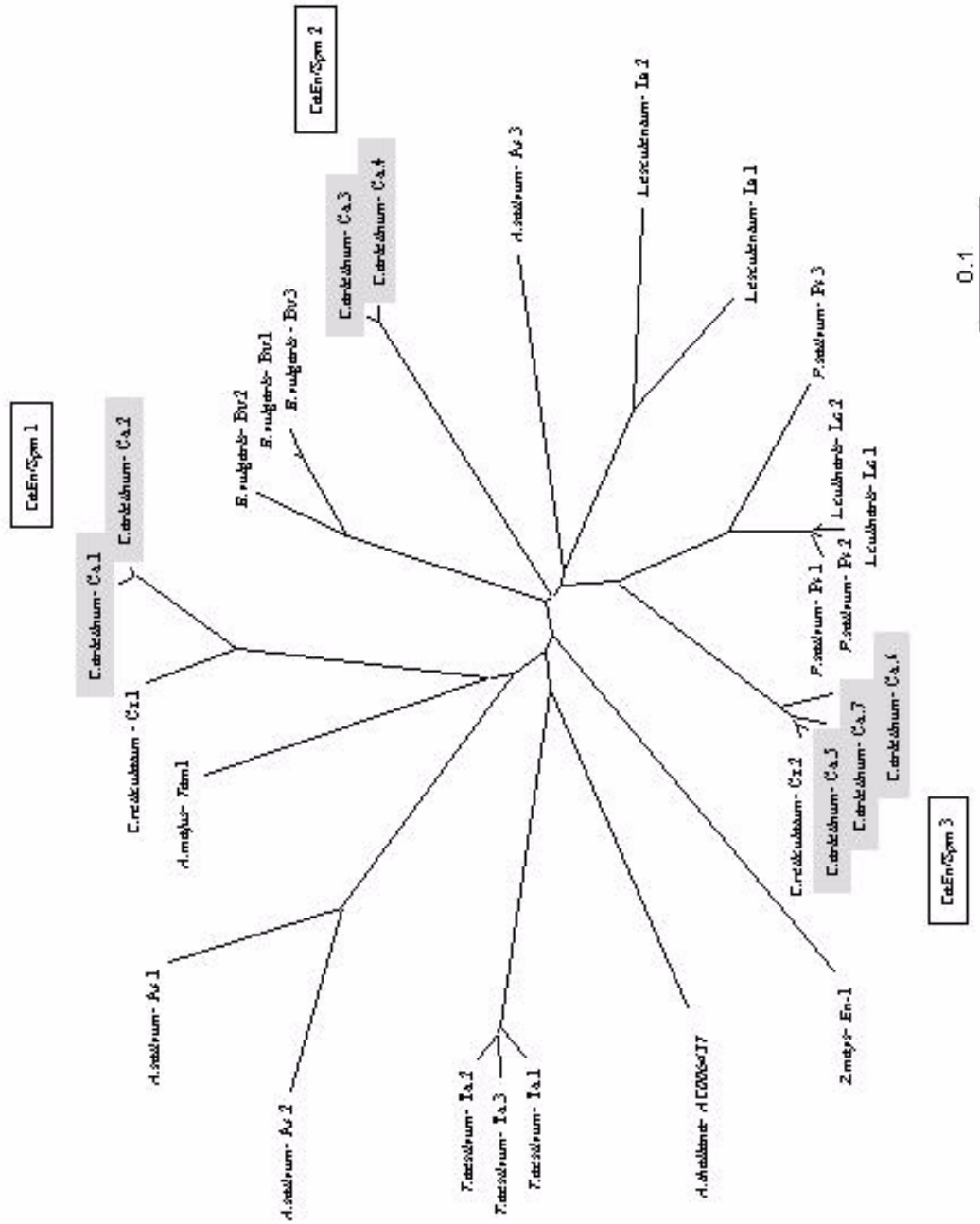
Das degenerierte Oligonukleotidpaar amplifiziert selektiv *En/Spm*-Transposase-Fragmente

Um die Identität der Amplifikate zu klären, wurden die PCR-Produkte von *Cicer arietinum*, *Cicer reticulatum*, *Lens culinaris*, *Pisum sativum*, *Lycopersicon esculentum*, *Beta vulgaris*, *Triticum aestivum* und *Allium sativum* kloniert. Nach einer Restriktionsanalyse der klonierten Fragmente wurden insgesamt 25 rekombinante Klone exemplarisch sequenziert, davon stammten je drei Sequenzen aus *Allium sativum* (AJ406395, AJ406394, AJ297749), *Triticum aestivum* (AJ406396- AJ406398) und *Beta vulgaris* (AJ297748, AJ406392, AJ406391), je zwei aus *Lycopersicon esculentum* (AJ406393, AJ297747) und *Lens culinaris* (AJ406388, AJ406389), drei aus *Pisum sativum* (AJ406390, AJ297746, AJ297745), zwei aus *Cicer reticulatum* (AJ297750, AJ297744) und sieben aus *Cicer arietinum* (AJ406381- AJ406387) (Fig.C36). Die Homologiesuche erbrachte für alle analysierten Sequenzen signifikante Homologien zu den entsprechenden Bereichen verschiedener pflanzlicher *En/Spm*-Elemente, so z.B. 82-94% Homologie zu *Psl* aus Petunie (*Petunia hybrida*) auf Ebene der Nukleotidsequenz. Die Aminosäuresequenzen des jeweils wahrscheinlichsten Leserasters wiesen 37-85% Ähnlichkeit zu *Tam1* aus Löwenmäulchen (*Antirrhinum majus*) und 30-77% Ähnlichkeit zu *En-1* aus Mais (*Zea mays*) auf. Mit Hilfe des degenerierten Oligonukleotidpaares (transAF/CR) können demnach Fragmente der Transposaseregion von *En/Spm*-Transposons in einer Vielzahl von Pflanzenarten selektiv amplifiziert werden.

Die Sequenzen der PCR-Produkte zeigten sowohl auf Nukleotidebene (46-99%) als auch auf Ebene der abgeleiteten Aminosäuresequenz (27- 98%; Fig.C36a) untereinander variable Homologien. Dabei enthielten etwa 44% der Fragmente intakte ORFs, bei den übrigen war das Leseraster -verursacht durch einzelne Punktmutationen - verschoben oder durch Stop-Codons unterbrochen. Das Dendrogramm in Fig.C36b reflektiert die genetische Distanz zwischen den einzelnen Aminosäuresequenzen, wobei zur besseren Übersicht bereits bekannte Transpososasequenzen aus *Z.mays* (*En-1*), *A. majus* (*Tam1*) und *A. thaliana* (AC006437) mit in die Betrachtung einbezogen wurden. Deutlich wird die Verwandtschaft der Leguminosensequenzen mit Ausnahme von *Cicer*, wobei *Cicer arietinum* sehr divergierende Sequenzen zeigt. Auf diese Divergenz wird im folgenden Abschnitt näher eingegangen.

Fig.C36: Sequenzvergleich der wahrscheinlichsten Leseraster aus TPase-Sequenzen verschiedener *En/Spm*- ähnlicher Elemente.

(a) Vergleich der sequenzierten PCR-Fragmente aus verschiedenen Spezies (Tab. C2) mit der entsprechenden Region von *Tam1* (*A. majus*; S23818) und *En-1* (*Z. mays*; S29329). Die Oligonukleotidsequenzen wurden nicht in den Vergleich miteinbezogen und die konservierten Domänen A-C sind grau unterlegt. Stark konservierte Aminosäuren sind in fetten Buchstaben dargestellt, Striche zeigen Lücken an, die für einen optimalen Sequenzvergleich eingefügt wurden, und Sternchen markieren Stop-Codons. In einigen Fällen waren Leserasterwechsel notwendig, um die Sequenzen vergleichen zu können. (b) Ein wurzelloses Dendrogramm zeigt nach Saitou & Nei (1987) die genetischen Distanzen der Aminosäuresequenzen von (a). Deutlich wird die



C36b

C.III.2. Charakterisierung von *CaEn/Spm*, einem Klasse II-Transposon im Genom der Kichererbse

Mit Hilfe der degenerierten Oligonukleotidprimer bietet die PCR eine schnelle und effiziente Möglichkeit, um Fragmente von *En/Spm*-Transposons aus pflanzlichen Genomen zu isolieren. Die Analyse klonierter Amplifikate kann als Ausgangspunkt für die weitere Charakterisierung der Transposonfamilie dienen. Daher wurde die Charakterisierung amplifizierter Transposonsequenzen für das Genom von *Cicer arietinum* exemplarisch durchgeführt. Dabei wurden Sequenz, genomische Organisation und chromosomale Lokalisation der Transposonmotive, sowie zusätzlich ihre Abundanz und Organisation in verwandten *Cicer*-Spezies untersucht.

C.III.2.1. Sequenzierung einiger Transposonkopien

Es gibt mindestens drei Subfamilien von *En/Spm*-ähnlichen Transposons in der Kichererbse

Nach der Amplifikation genomischer DNA aus *Cicer arietinum* mit den degenerierten Primern wurden sechs rekombinante Klone sequenziert, die alle Sequenzen mit signifikanten Homologien zu pflanzlichen *En/Spm*-Transposons enthielten. Gemeinsam mit dem schon zu Beginn der Untersuchungen zufällig klonierten Fragment *Ca3* lagen sieben verschiedene Kopien aus dem Transposasesegen einer *En/Spm*-Transposonfamilie aus Kichererbse vor (*Ca1-Ca7*), die 187-191 AS lange ORFs enthielten. Die Fragmente wiesen auf Ebene der Nukleotidsequenz 88-91% Homologie zu *Psl* aus Petunie (*Petunia hybrida*) auf. Zu *Tam1* (*Antirrhinum majus*) und *En-1* (*Zea mays*) wurden jeweils 60-71% bzw. 52-72% Ähnlichkeit der abgeleiteten Aminosäuresequenz festgestellt. Die Homologiesuche in öffentlichen Datenbanken, u.a. TAIR (*The Arabidopsis Information Resource* - <http://www.arabidopsis.org>) ließ auch signifikante Homologien zu Transposase-ORFs *En/Spm*-ähnlicher Elemente aus *A. thaliana* erkennen (AC006437, AC006586, AF128394, AC005957, AC005897 sowie TAIR-Codes MOD1.14, F6H5_116, F15O4_30, F11I2_2, T17A2_5), wobei die Ähnlichkeit der Aminosäuresequenz des wahrscheinlichsten Leserasters 41- 69% betrug (Fig.C36a,b).

Die Kichererbsensequenzen untereinander divergierten auf Nukleotidebene bis zu 50%, was auch der Divergenz zu anderen Transposons, etwa aus Weizen (49.8%) oder *Beta vulgaris* (49.6%) entsprach. Die Länge der Fragmente war mit 567-574 Bp (ohne Primer) relativ einheitlich. Nur *Ca6* enthielt eine 66 Bp lange Deletion (Fig.C36a). Das Dendrogramm der zu erwartenden Aminosäuresequenzen (Fig.C36b) betont die Heterogenität der Kichererbsensequenzen, die in drei Gruppen zerfallen und möglicherweise evolutiv verschiedene Unterfamilien *En/Spm*-ähnlicher Elemente im Kichererbsengenom darstellen. Sie wurden als *Cicer arietinum En/Spm*-ähnliche Elemente 1-3 (*CaEn/Spm*1-3) bezeichnet. Beide Fragmente aus *C. reticulatum* fügten sich in je eine dieser Gruppen ein (Fig.C36b).

Vier der sieben Kichererbsenfragmente enthielten intakte ORFs mit Homologien zu *En/Spm*-ähnlichen Transposasemotiven, die drei übrigen wiesen durch Punktmutationen verursachte Stop-Codons oder Leserasterwechsel auf.

C.III.2.2. Genomische Organisation und Abundanz der *En/Spm*-Familie

Die *En/Spm*-Familie ist mittelrepetitiv und dispers im Kichererbsengenom verstreut

Die genomische Organisation der *En/Spm*-ähnlichen Elemente aus der Kichererbse wurde mittels einer Southern-Analyse untersucht. Restringierte, genomische DNA aus *Cicer arietinum* wurde gegen das klonierte Fragment *Ca3* hybridisiert. Auf einem Hintergrundschmier waren 12-20 diskrete Banden im Molekulargewichtsbereich von 0.5 bis >8 Kbp erkennbar, woraus auf die Existenz mehrerer Kopien und deren disperse genomische Organisation geschlossen werden kann (Fig.C37). Die Intensität der Hybridisierungssignale deutet auf eine niedrig- bis mittelrepetitive Abundanz der Sequenzen hin. Das Bandenmuster bei den einzelnen Restriktionen war unterschiedlich intensiv, d.h. nicht alle hybridisierenden Fragmente wiesen die gleiche Homologie zur Sonde auf. Die dominanten Banden repräsen-

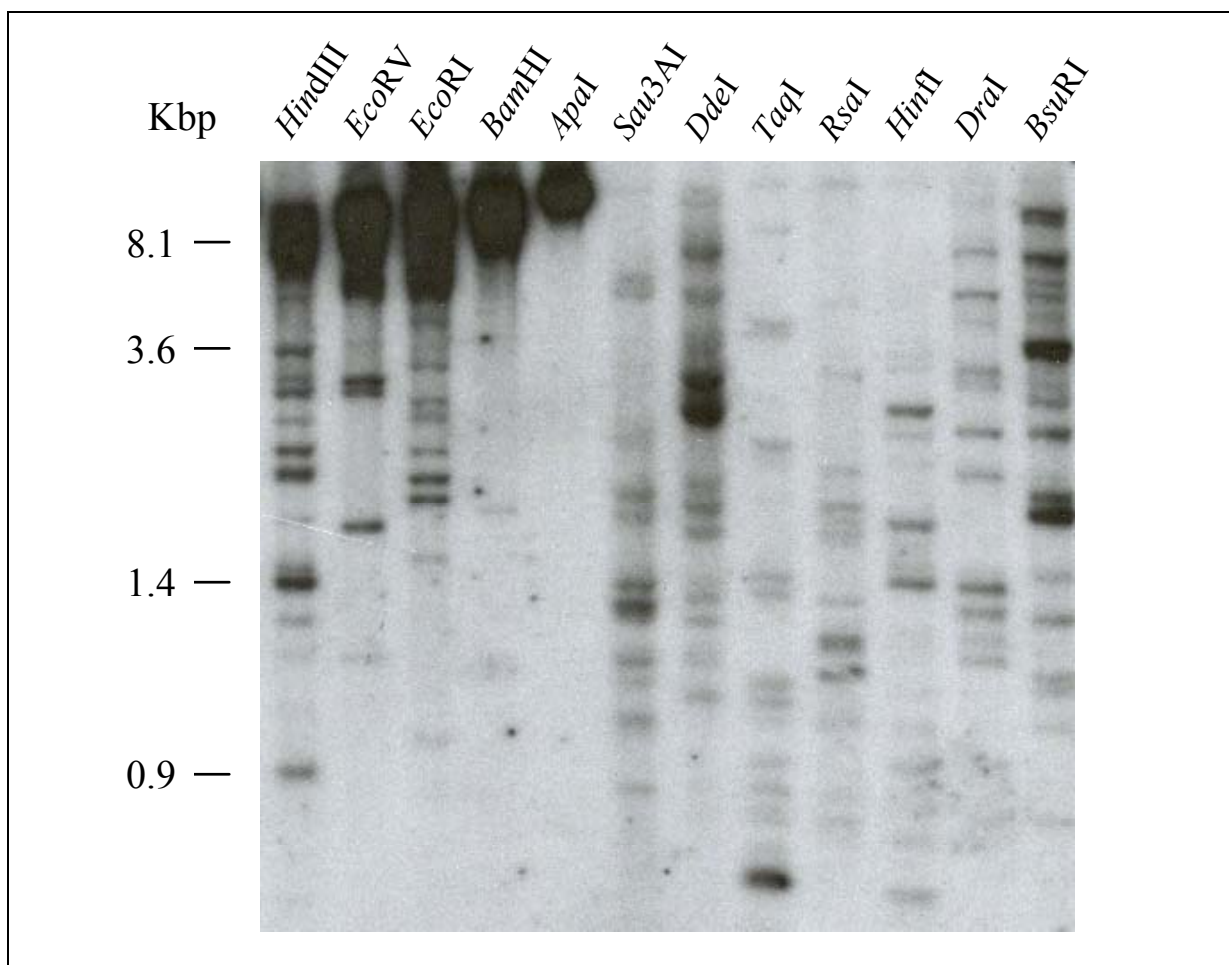


Fig.C37: Genomische Organisation von *En/Spm*-ähnlichen Sequenzen in *Cicer arietinum*.

Autoradiogramm eines Southern-Blots mit restringierter genomischer DNA von *C. arietinum*, hybridisiert mit Ca3. Hybridisierungsmuster und -intensität weisen auf mittlere Abundanz und disperse Verteilung der homologen Sequenzen hin.

tieren konservierte Transposonsequenzen mit hoher Homologie zur Sonde, während die schwächeren Signale möglicherweise durch verwandte, aber divergente Sequenzen verursacht werden. Außerdem kann eine einzelne Bande mehr als ein homologes Restriktionsfragment repräsentieren und so eine höhere Intensität aufweisen. Enzyme mit einer 6 Bp-Erkennungssequenz produzierten einen Schmier im hochmolekularen Bereich, wobei durch die niedrige Auflösung eines Standard-Agarosegels nicht differenziert werden konnte, ob es sich um große Restriktionsfragmente oder ungeschnittene DNA handelte. Kürzere Expositionszeiten ließen in diesem Bereich mehr Banden sichtbar werden, was jedoch nur für einen Teil der Restriktionen zutraf.

CaEn/Spm ist damit das erste DNA-Transposon (Klasse II), das für *Cicer arietinum* beschrieben wurde, und besteht möglicherweise aus mehreren Unterfamilien, die mittelrepetitiv und dispers verteilt im Genom der Kichererbse vorliegen.

C.III.2.3. Chromosomale Lokalisation der *CaEn/Spm*-Transposonkopien

***CaEn/Spm*-Loci sind an den Enden der pericentrischen Heterochromatinblöcke von mindestens sechs Chromosomenpaaren sichtbar**

Die klonierten *En/Spm*-Sequenzen Ca1, Ca3, Ca4, Ca5 und Ca6 aus der Kichererbse dienten als Sonden für FISH. Auf mindestens sechs der acht Chromosomenpaare konnten diskrete Hybridisierungssignale festgestellt werden, deren Anzahl und Intensität auf eine höchstens mittelrepetitive Abundanz der homologen Sequenzmotive hindeuten (Fig.C38a-d). Die Signale wurden im distalen Teil der großen pericentrischen Heterochromatinblöcke identifiziert, unmittelbar angrenzend an die euchromatischen Bereiche der Metaphasechromosomen. Die Signale waren an einem oder beiden Enden des heterochromatischen Blocks angesiedelt und jeweils auf beiden Chromatiden sichtbar (Fig.C38b). Die beiden größten Chromosomenpaare A und B trugen zusätzliche Loci in den pericentromerischen, heterochromatischen Bereichen. Die sekundäre Einschnürung des Chromosoms A, Lokalisation der NOR-Region, und die zentralen Teile des pericentrischen Heterochromatins aller Chromosomen zeigten keine sichtbaren Hybridisierungssignale, was darauf hindeutet, daß Transposase (TPase)-Sequenzen von diesen Bereichen ausgespart sind.

C.III.2.4. Abundanz verwandter Sequenzmotive im Genus *Cicer* und anderen Genera

***CaEn/Spm*-Motive sind nicht auf die Gattung *Cicer* beschränkt**

Verschiedene *Cicer*-Spezies, weitere Leguminosen und einige andere Dicotyledonen wurden durch Southern-Hybridisierung auf das Vorhandensein von *CaEn/Spm*-verwandten Sequenzen überprüft. Ferner sollten Abundanz und genomische Organisation möglicherweise vorhandener Motive analysiert werden. Genomische DNA aus einer oder mehreren

Akzessionen der Spezies wurden gegen das klonierte *CaEn/Spm*-Fragment *Ca3* hybridisiert. In allen untersuchten Genomen der Gattung *Cicer* konnte die Anwesenheit von *CaEn/Spm*-verwandten Sequenzen

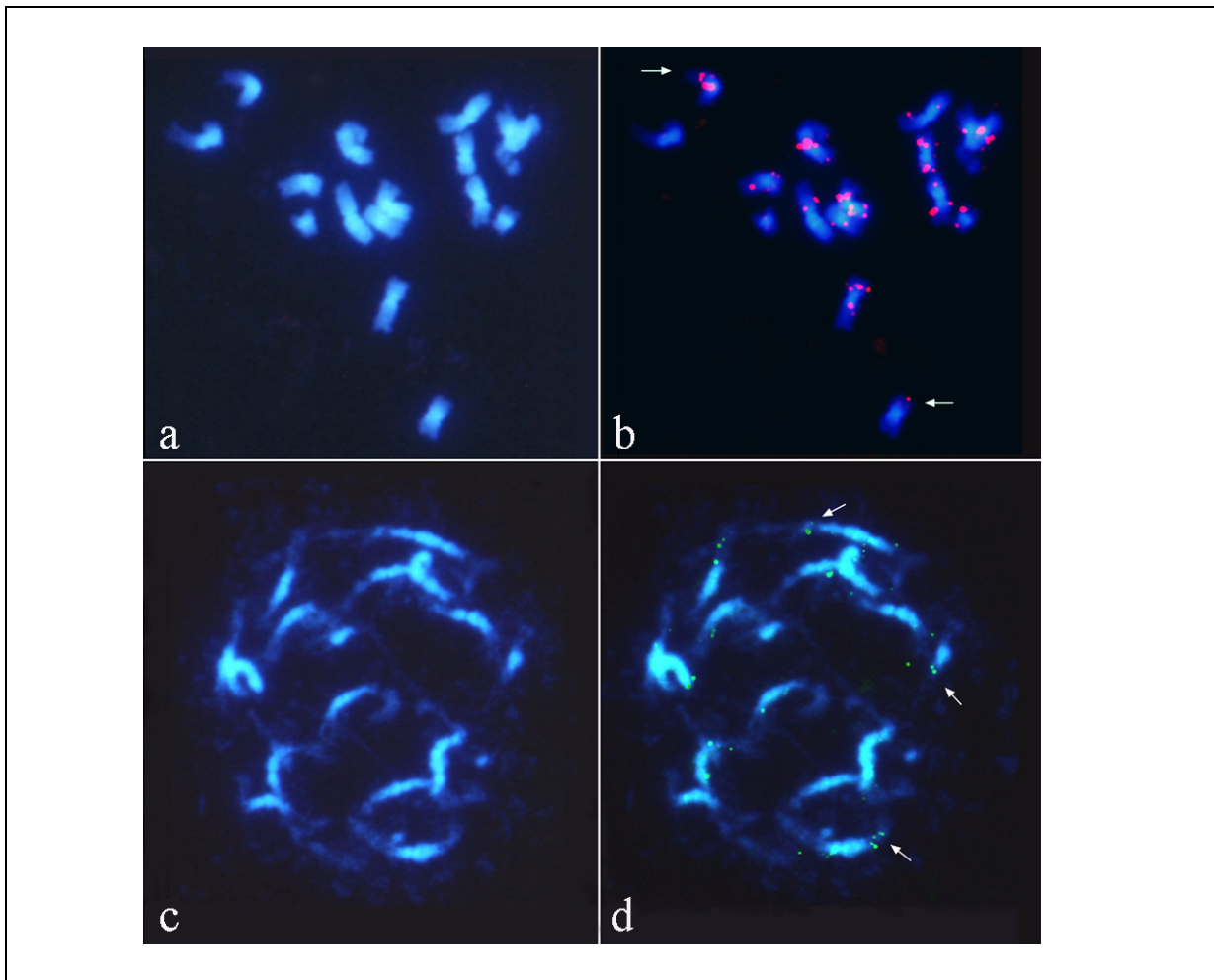


Fig.C38: Lokalisation von *CaEn/Spm*-Sequenzen auf Metaphasechromosomen von *Cicer arietinum* durch FISH.

(a-d) Metaphasechromosomen nach DAPI-Färbung (blau). (b,d) Dieselben Metaphasechromosomen nach Fluoreszenz *in situ*-Hybridisierung mit fünf klonierten Fragmenten von *En/Spm*-ähnlichen Transposonsequenzen aus *C. arietinum* (Computerüberlagerung). Die Signale sind als rote (b) und grüne Fluoreszenz (d) auf mindestens sechs Chromosomenpaaren sichtbar. Sie liegen hauptsächlich in den distalen Teilen der pericentrischen Heterochromatinblöcke und sind eng benachbart zu euchromatischen Regionen (Pfeile). Auf den Chromosomenpaaren A und B können zusätzliche Loci in centromerischer Position innerhalb des Heterochromatins detektiert werden.

nachgewiesen werden (Fig.C39). Das Bandenmuster mit 2-9 Fragmenten im Molekulargewichtsbereich von 0.5 bis 2.1 Kbp kann sowohl Restriktionspolymorphismen innerhalb der Wiederholungseinheit als auch verschiedene Kopien der Transposonfamilie mit internen Deletionen reflektieren. Polymorphismen können auch durch verschiedene Kopien in unterschiedlichem genomischen Kontext verursacht werden (Insertionspolymorphismen). Dies kann ohne Kenntnis der Sequenz einer vollständigen Kopie von *CaEn/Spm* nicht unterschieden werden. Sowohl das Bandenmuster als auch die einheitlichen Signalintensitäten

deuten auf die Anwesenheit multipler Kopien der entsprechenden Sequenzen in allen untersuchten *Cicer*-Arten hin. Obwohl das Hybridisierungsmuster zwischen den einzelnen Spezies variierte, waren bestimmte Fragmente in den Akzessionen einer Spezies konserviert. Es ist daher anzunehmen, daß *CaEn/Spm*-ähnliche Sequenzen in speziesspezifischer genomischer Organisation vorliegen. *En/Spm*-Transposasemotive aus der Gattung *Cicer* sind darüberhinaus auch innerhalb der Leguminosen weitverbreitet, wie exemplarisch an deren Konservierung in der Gattung *Phaseolus* deutlich wird. TPase-ähnliche Sequenzen aus anderen Leguminosen (1.2.) konnten jedoch nicht durch Hybridisierung detektiert werden, was vermutlich durch Sequenzdivergenz zur Sonde oder eine nur geringe Kopienzahl bedingt ist.

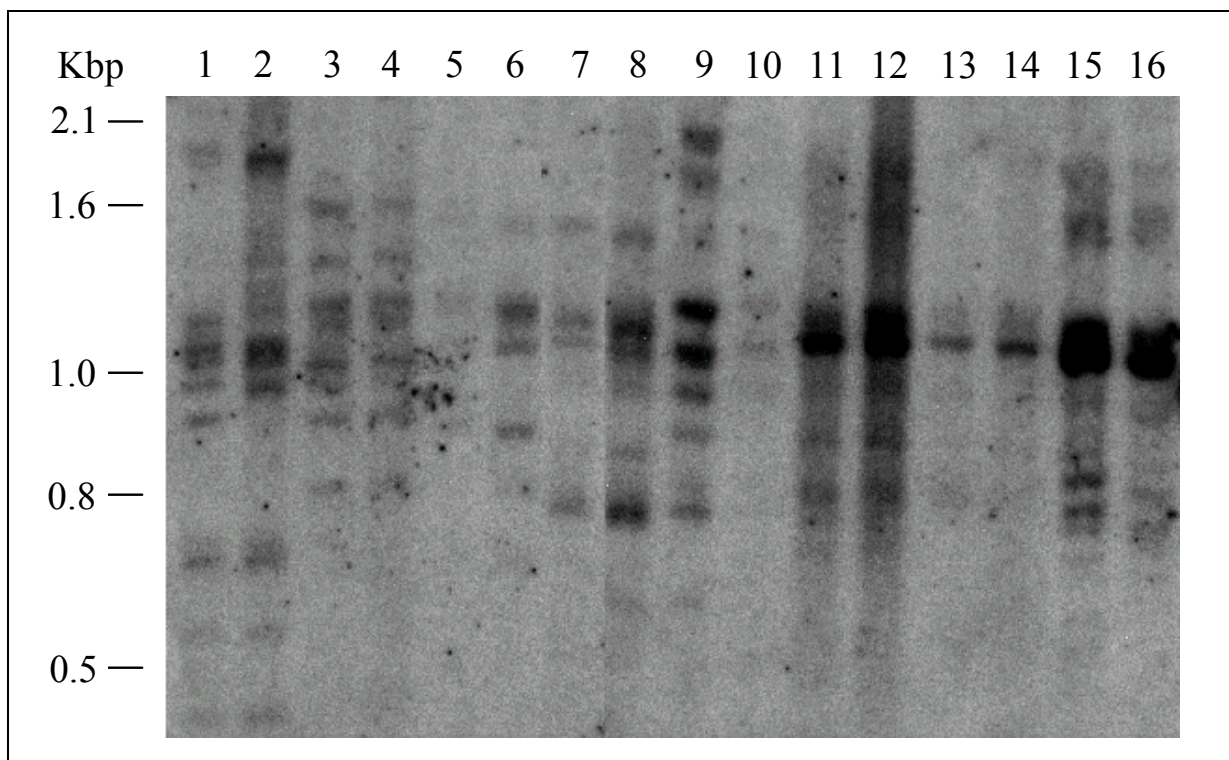


Fig.C39: Genomische Organisation von *CaEn/Spm*-ähnlichen Sequenzen in der Gattung *Cicer*.

Autoradiogramm eines Southern-Blots mit genomischen DNAs aus verschiedenen Spezies und Akzessionen der Gattung *Cicer*, die mit *RsaI* restringiert und gegen *Ca3* hybridisiert wurden. *C. yamashitae* ILWC214 (1), *C. yamashitae* ILWC53 (2), *C. chorassanicum* ILWC23 (3), *C. chorassanicum* ILWC146 (4), *C. cuneatum* ILWC37 (5), *C. cuneatum* ILWC40 (6), *C. judaicum* ILWC57 (7), *C. pinnatifidum* ILWC204 (8), *C. bijugum* ILWC62 (9), *C. bijugum* ILWC42 (10), *C. echinospermum* ILWC 35 (11), *C. echinospermum* ILWC39 (12), *C. reticulatum* ILWC21 (13), *C. reticulatum* PI489777 (14), *C. arietinum* ILC482 (15), *C. arietinum* ILC3279 (16). Es wird nur der Größenbereich mit den meisten homologen Fragmenten in den Wildarten gezeigt. Aufgrund der schwächeren Signale in den Wildarten wurde dieser Southern-Blot länger exponiert als in Fig.C37.

C.IV. Transkription von dispersen, niedrig- und mittelrepetitiven Elementen der Kichererbse

Retrotransposonsequenzen werden sowohl zur Synthese der vom Element codierten Proteine (*gag-pol*-Genprodukte) als auch zur Herstellung eines Templats für die Reverse Transkriptase transkribiert. Obwohl die Transposition auf Transkriptionsebene reguliert wird, kann es in einzelnen Geweben und besonders unter der Einwirkung von biotischen oder abiotischen Streßfaktoren zur Synthese von nachweisbaren Mengen an Transposon-RNA kommen. Vor diesem Hintergrund sollte untersucht werden, ob eine der drei mittel- bis niedrigrepetitiven Retrotransposonsequenzen (*CaTy*, *CaDis*, *CaLin*) oder das Klasse II-Transposon (*CaEn/Spm*) in bestimmten Geweben von *C. arietinum* transkribiert werden. Zum Nachweis der sequenzspezifischen Transkripte wurde RT-PCR verwendet, die im Vergleich zur Northern-Hybridisierung einen sensitiveren Nachweis von geringen RNA-Mengen und divergenten Sequenzen ermöglicht.

C.IV.1. Nachweis von spezifischen „messenger“-RNAs repetitiver Elemente mit RT-PCR

Für die RT-PCR wurde gereinigte, polyadenylierte mRNA in cDNA umgeschrieben und direkt als Templat für die PCR eingesetzt. Verwendet wurden die bereits beschriebenen Oligonukleotidpaare für *CaTy* (II.1.1.), *CaDis* (II.2.1.), *CaLin* (II.3.1.) und das Transposon *CaEn/Spm* (III.1.1.) sowie das spezifische Oligonukleotidpaar UbiL/R für eine Teilsequenz eines Ubiquitin-ähnlichen Gens aus der Kichererbse (*CanSMT3*-Gen; AJ001901). Das *CanSMT3*-Gen ist homolog zu *SMT3*-Genen aus Pflanzen- und Tiergenomen, die vermutlich im Zusammenhang mit Funktion und Struktur eukaryontischer Kinetochoren stehen (Munoz et al., 1998), und enthält insgesamt 1,2 Kbp lange Intronsequenzen. Das PCR-Produkt aus genomischem Templat ist daher deutlich größer als die 340 Bp langen cDNA-Amplifikate, so daß dieser Ansatz in der RT-PCR als Kontrolle für eventuelle genomische Verunreinigungen der cDNA dienen kann. Die cDNA wurde aus Blättern, Wurzeln und Blüten verschiedener Akzessionen von *C. arietinum* sowie aus *C. reticulatum* gewonnen, um möglicherweise existierende gewebespezifische Unterschiede in der Transkription exemplarisch zu erfassen. Verschiedene Streßsituationen wurden simuliert, indem Blätter einzelner Pflanzen verwundet oder mit *Ascochyta rabiei*, einem pathogenen Ascomyceten, infiziert wurden. Als weiterer Ansatz für die RT-PCR diente cDNA von *A. rabiei*, um mögliche Verunreinigungen der pflanzlichen cDNA mit dem Pilzpathogen zu detektieren.

Nur das *CaDis*-spezifische Oligonukleotidpaar detektiert Transkripte in allen untersuchten Geweben

Das spezifische Oligonukleotidpaar UbiR/L amplifizierte nur Sequenzen aus pflanzlichen Templaten (FigC40a), wobei genomische DNA aus *C. arietinum* (Spur1) das erwartete 1.6 Kbp-Fragment lieferte, während in allen cDNAs ein einheitliches, ca. 340 Bp langes Amplifikat sichtbar war. Demnach befinden sich keine genomischen Verunreinigungen in den für alle Ansätze verwendeten cDNA-Templaten. Weder die Transposon- noch *Ty1-copia*-Oligonukleotidprimer erlaubten außer der Produktion der jeweils 650 bzw. 270 Bp langen Amplifikaten aus genomischer DNA irgendeine Amplifikation von *C. reticulatum*- oder *C. arietinum*-cDNA. In den untersuchten Geweben befinden sich daher keine nachweisbaren Mengen homologer RNAs dieser repetitiven Familien. Auch mit den *CaDis*-spezifischen Primern ließ sich in genomischer DNA die bereits bekannte Amplifikatgröße von 280 Bp nachweisen (Fig.C40b, Spur 1). Für die cDNAs aller Gewebe der Kichererbse und aller untersuchten Streßzustände waren außerdem schwache Banden der gleichen Größe sowie 1-2 Banden von ca. 200 Bp und darunter erkennbar (Fig.C40b, Spur 4-10), ebenso bei Blattmaterial von *C. reticulatum* (Fig.C40b, Spur3). Die nachweisbaren Transkripte waren nicht auf *A. rabiei*-Verunreinigungen zurückzuführen, da das Oligonukleotidpaar mit Pilz-

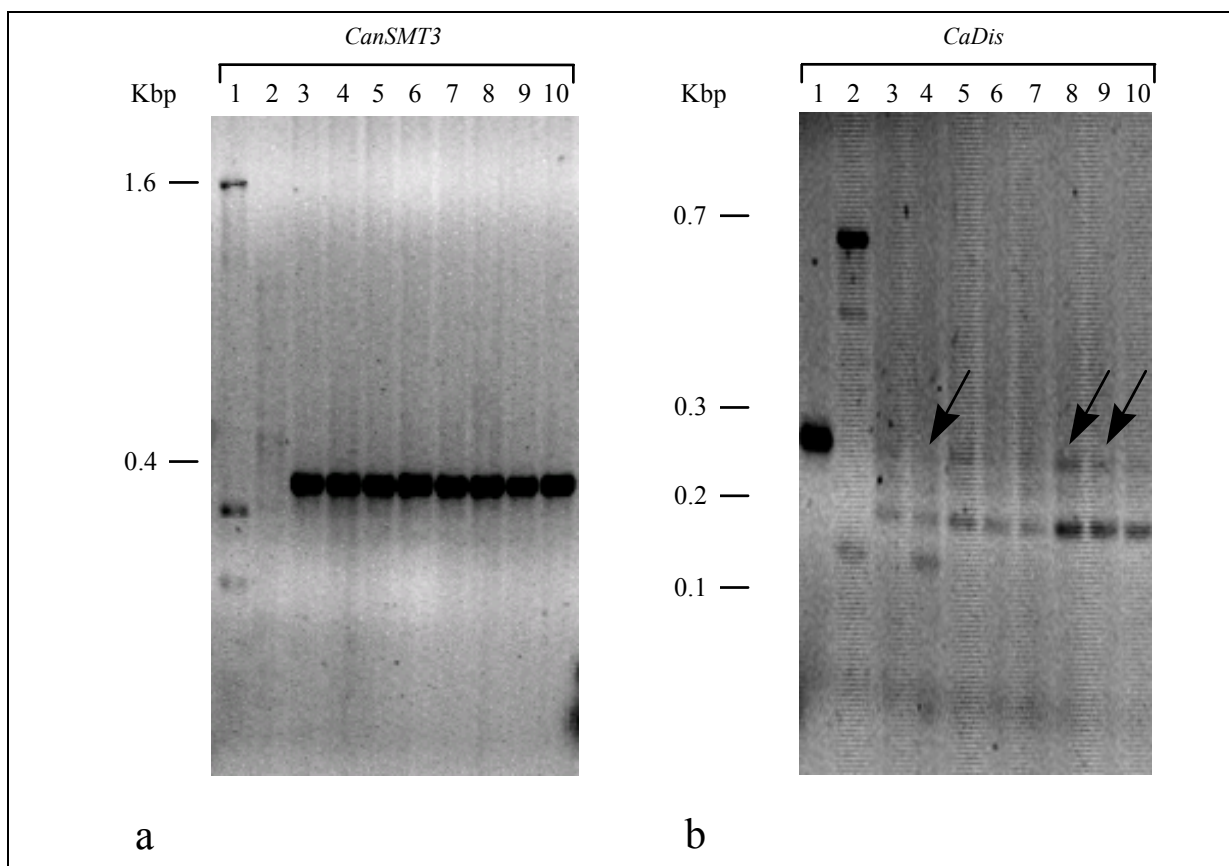


Fig.C40: RT-PCR mit mittelrepetitiven, dispersen Elementen aus *C. arietinum*.

(a) Agarosegel mit amplifizierten Transkripten des *CanSMT3*-Gens, die als Kontrolle dienen und Verunreinigungen durch genomische DNA ausschließen. (b) Ethidiumbromid-gefärbtes Agarosegel mit Banden, die mit dem Primerpaar für *CaDis* amplifiziert wurden. Als Template wurden verwendet: genomische DNA aus *C. arietinum* ILC 3279 (1), genomische DNA aus *Ascochyta rabiei* (2), cDNA aus *C. reticulatum* Pi 489777 (Blatt) (3), cDNA aus *C. arietinum* ILC 4958 (Blatt) (4), cDNA aus *C. arietinum* ILC 482 (Blatt) (5), cDNA aus *C. arietinum* ILC 482 (verwundetes Blatt) (6), cDNA aus *C. arietinum* ILC 482 (infiziertes Blatt) (7), cDNA aus

C. arietinum ILC 3279 (Blüte) (8), cDNA aus *C. arietinum* ILC 4958 (Wurzel) (9), cDNA aus *C. arietinum* ILC 482 (Wurzel) (10). Transkripte von *CaDis* werden in allen untersuchten Geweben von *C. arietinum* und *C. reticulatum* detektiert. Pfeile kennzeichnen die klonierten und sequenzierten Fragmente. Die kleineren Banden beruhen auf PCR-Artefakten. Die Fotografie des Ethidiumbromid-gefärbten Gels wurde zur Verdeutlichung farblich invertiert.

DNA deutlich größere Amplifikate produzierte (ca. 600 Bp, Spur2). Auch mit dem Oligonukleotidpaar für LINE-ähnliche Sequenzen ließen sich eine Vielzahl von schwachen Transkripten in allen Geweben nachweisen (ohne Bild). Die Banden sind jedoch hier vermutlich auf unspezifische PCR-Artefakte zurückzuführen, da keines der Amplikons die Größe der genomischen Bande (ca. 600 Bp) aufweist. Da die Oligonukleotide einen Ausschnitt aus dem RT-Bereich des Polyproteinleserasters flankieren, das nicht von Intronsequenzen unterbrochen wird, sollten transkribierte Sequenzen dieselbe Länge wie das genomische Amplifikat zeigen.

Offenbar sind in mindestens drei Geweben der Kichererbse und unabhängig von der Streßbehandlung Transkripte zu finden, bei denen eine Verwandtschaft zu *CaDis*-Sequenzen naheliegend ist. Das gleiche trifft für weitere Kichererbsenakzessionen und Blattgewebe aus der nah verwandten Wildart *C. reticulatum* zu.

C.IV.2. Identifikation von *CaDis*-homologen Transkripten durch Sequenzanalyse

Die 280 Bp-Transkripte aus Kichererbse stellen mehrheitlich Sequenzen der *CaDis*-Familie dar

Zur Identifikation wurden die 280 Bp langen PCR-Amplifikate aus Blättern, Wurzeln und Blüten von *C. arietinum* kloniert (Pfeile, Fig.C40b). Eine Restriktionsanalyse der rekombinanten Klone wurde durchgeführt, um die Verschiedenartigkeit der klonierten Amplifikate zu erfassen. Von den klonierten Fragmenten wurden 9 exemplarisch sequenziert, mehr als die Hälfte davon wies Homologien zu pflanzlichen Retroelementen auf. Aufgrund der Restriktionsanalyse repräsentieren diese die Mehrzahl der rekombinanten Klone und stellen somit den Hauptanteil der PCR-Bande dar. Ebenfalls sequenzierte Klone aus anderen Restriktionsklassen enthielten genomische Sequenzen für prolinreiche Proteine, einen Translationsinitiationsfaktor und eine Saccharosesynthase.

Die Sequenzen mit den Retrotransposonhomologien hatten mit 228-229 Bp nicht nur exakt dieselbe Länge wie die genomischen Sequenzen der *CaDis*-Familie, sondern wiesen auf Nukleotidebene auch 85-97 % Homologie zu diesen auf, was etwa der Konservierung der cDNA-Sequenzen untereinander (88-95%) entsprach (II.2.2.). Die Transkripte sind demnach eindeutig der *CaDis*-Familie zuzuordnen. Der AT-Gehalt der fünf cDNA-Fragmente lag mit 58.5-62.9% leicht unter dem der genomischen Sequenzen (60.3-64.2%). Sofern mehrere Transkripte eines Gewebes sequenziert worden waren, zeigten sie untereinander keine höhere Homologie als zu den übrigen Sequenzen. Auf Nukleotidebene konnte 60-62% Homologie zum transkribierten Element aus *M.sativa* festgestellt werden (L39960). Drei der fünf cDNA-Sequenzen enthielten ein intaktes offenes Leseraster über ihre gesamte Länge, die beiden übrigen zeigten einen Leserasterwechsel. Die cDNA-Klone fügten sich auch auf Ebene der wahrscheinlichsten AS-Sequenz in die Gruppe der genomischen *CaDis*-Sequenzen ein, zu

denen sie 75-95 % Homologie aufwiesen (Fig.C25 in II.2.2.). Homologe Retrotransposonsequenzen in öffentlichen Datenbanken, entsprachen den detektierten Sequenzen aus dem *gag*-Bereich darunter zwei EST-Sequenzen aus *O.sativa* (II.2.2.).

Die in den verwendeten Geweben der Kichererbse und unter verschiedenen Streßzuständen der Pflanze gefundenen Transkripte stellen also mehrheitlich Sequenzen der *CaDis*-Familie dar.

C.V. Verwendungsmöglichkeiten charakterisierter repetitiver Elemente aus *Cicer arietinum*

Neben der Isolation und Charakterisierung repetitiver Komponenten des Kichererbsengenoms war auch deren Verwendung zur Genomanalyse von *C. arietinum* von Interesse. Durch ihre Fähigkeit zu dynamischen Veränderungen, ihre hohe Abundanz und ihr Vorkommen in weiten Bereichen des Genoms eignen sich viele repetitive Sequenzen zu phylogenetisch-taxonomischen Untersuchungen oder zur Entwicklung molekularer Marker. Zwei Anwendungsaspekte der hochabundanten Elemente *CaSat1*, *CaSat2* und *CaRep* werden im folgenden vorgestellt.

C.V.1. Phylogenetisch-taxonomische Betrachtungen im Genus *Cicer*

Hochrepetitive DNA kann zur Untersuchung phylogenetischer und taxonomischer Fragestellungen verwendet werden, vorausgesetzt daß homologe Sequenzen innerhalb einer Gattung oder Familie konserviert sind. Im Anschluß an die Charakterisierung der hochrepetitiven Satellitenelemente *CaSat1* und *CaSat2* wurde untersucht, ob homologe Sequenzen in weiteren Spezies der Gattung vorhanden sind und in welcher genomischer Organisation und Abundanz sie vorliegen.

C.V.1.1. Abundanz und genomische Organisation der Tandemelemente *CaSat1* und *CaSat2* in Spezies der Gattung *Cicer*

Die Abundanz und genomische Organisation der beiden hochabundanten Sequenzfamilien *CaSat1* und *CaSat2* im Genom weiterer *Cicer*-Spezies wurde durch Southern-Hybridisierung gegen *RsaI*-restringierte, genomische DNA aus neun annuellen und einer perenniellen Art der Gattung untersucht. Für die annuellen Spezies wurden jeweils mehrere Akzessionen verwendet, als Sonden dienten klonierte Fragmente der beiden repetitiven Familien.

Beide Satellitenelemente sind mit hoher Abundanz und konservierter Tandemorganisation in allen annuellen *Cicer*-Spezies außer *C. cuneatum* vorhanden

Beide Satellitensoniden, pCaEr1062 (*CaSat1*) und pCaEr453 bzw. ein partielles, subkloniertes Fragment davon (*CaSat2*) produzierten starke Signale mit der DNA aller annuellen *Cicer*-

Spezies. Eine Ausnahme bildete *Cicer cuneatum*, bei der in keiner der vier verwendeten Akzessionen eine Hybridisierung nachgewiesen werden konnte. In den übrigen annuellen Spezies war jeweils ein konserviertes Leiterbandenmuster sichtbar, welches eine Monomergröße von ca. 170 Bp (*CaSat1*) und 100 Bp (*CaSat2*) widerspiegelte (Fig.C41,C42).

Die Intensität der Hybridisierungssignale von *CaSat1* variierte deutlich zwischen den einzelnen Spezies. Die stärksten Signale waren bei *C. chorassanicum* (Fig.C41, Spur 14 und 15) sichtbar, nur wenig schwächer erschienen die Restriktionsfragmente der ersten Kreuzungsgruppe (*C. arietinum*, *C. reticulatum*, *C. echinospermum*; Fig.C41, Spur 1-5). Deutlich schwächer war das Leitermuster von *C. yamashitae* (Fig.C41, Spur 16 und 17), während *C. bijugum*, *C. pinnatifidum* und *C. judaicum* (zweite Kreuzungsgruppe) die schwächsten Signale zeigten (Fig.C41, Spur 6-11). Bei *C. echinospermum*, *C. reticulatum*, *C. chorassanicum* und *C. yamashitae* waren noch schwächere Banden zwischen den 170 Bp-Oligomeren sichtbar. Dabei handelt es sich vermutlich um eine Untergruppe von abgewandelten, komplex aufgebauten *CaSat1*-Wiederholungseinheiten mit niedrigerer Abundanz.

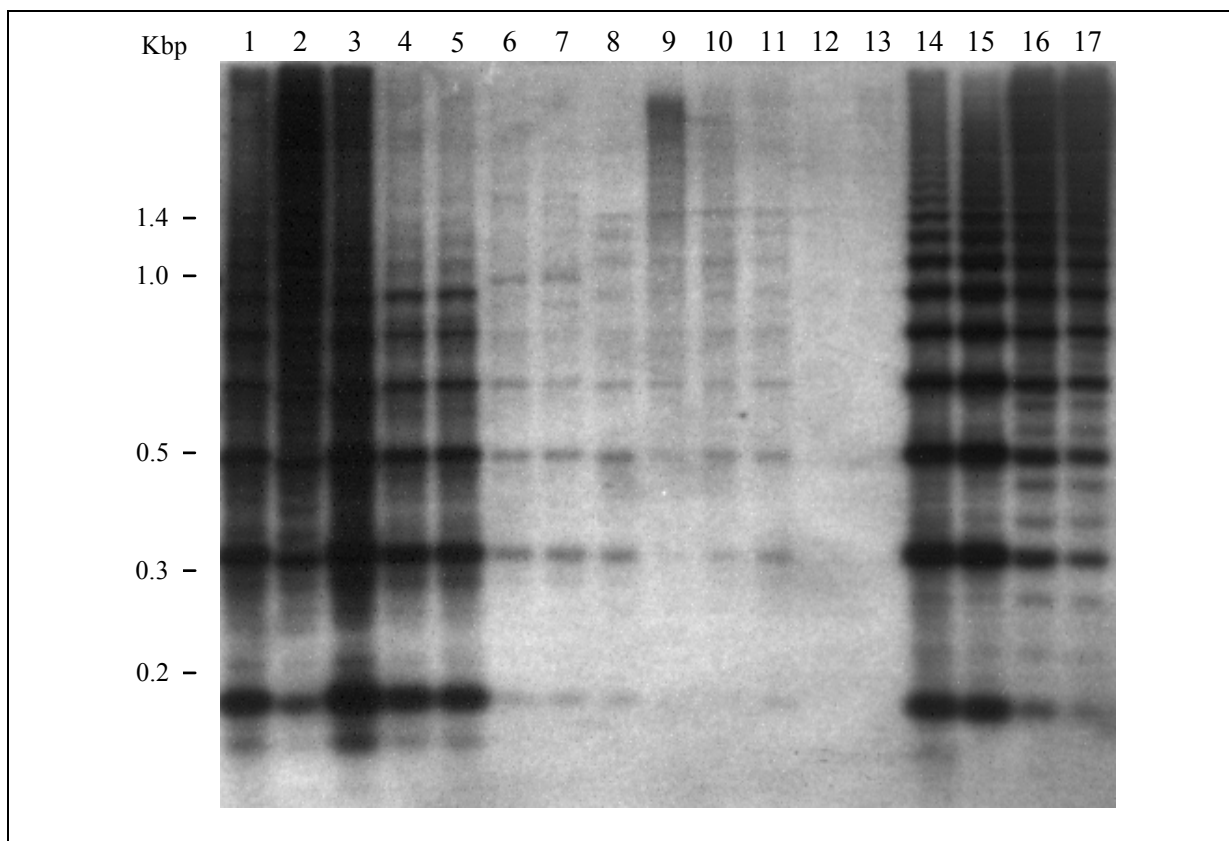
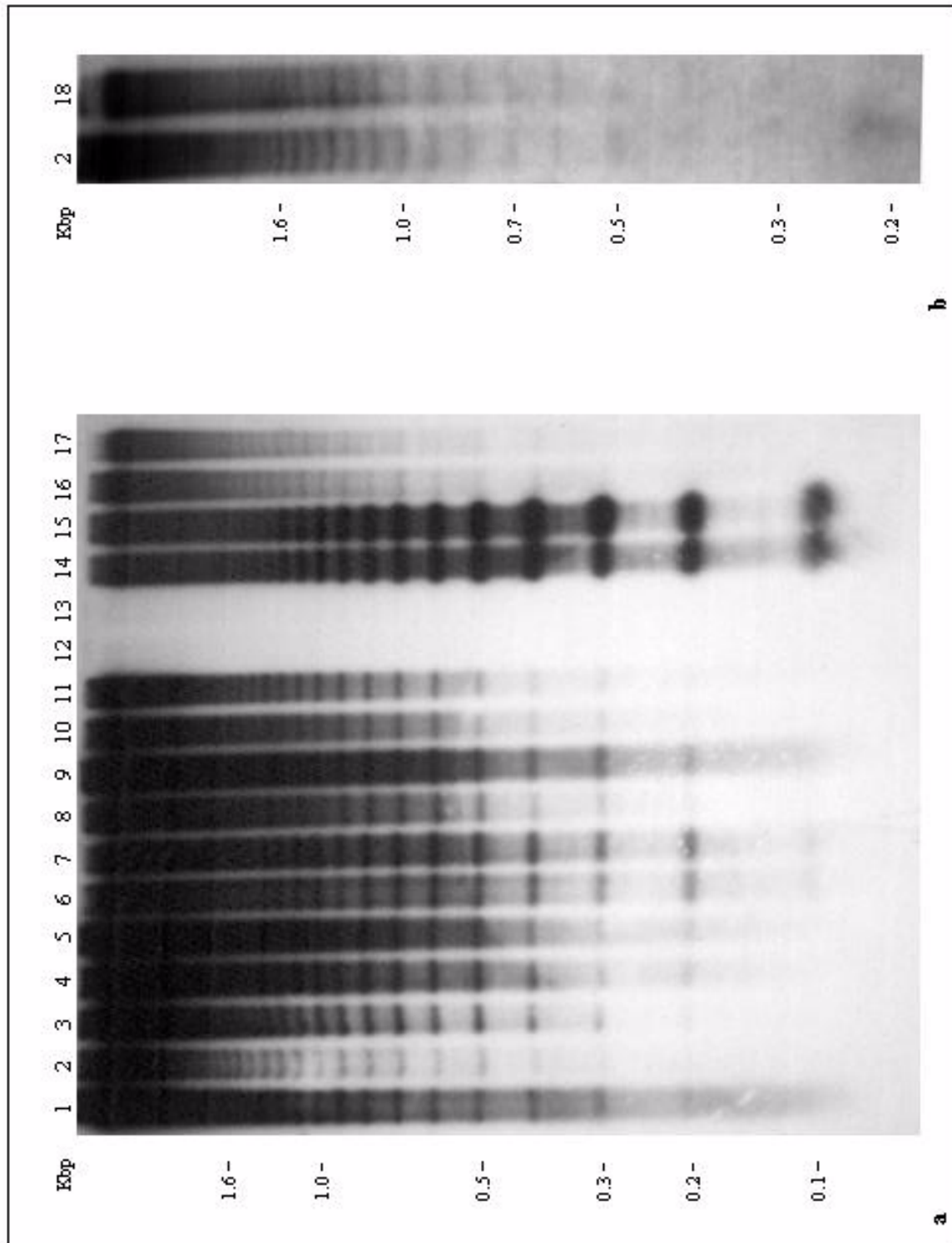


Fig.C41: Genomische Organisation der Satellitenfamilie *CaSat1* in der Gattung *Cicer*.

Autoradiogramm eines Southern-Blots mit *RsaI*-restringierter genomischer DNA verschiedener Spezies, die gegen pCaEr1062 hybridisiert wurde. Folgende Spezies wurden untersucht: *C. reticulatum* PI489777 (1), *C. arietinum* ILC3279 (2), *C. arietinum* ILC482 (3), *C. echinospermum* ILWC39 (4), *C. echinospermum* ILWC 35 (5), *C. bijugum* ILWC32 (6), *C. bijugum* ILWC62 (7), *C. pinnatifidum* ILWC204 (8), *C. pinnatifidum* ILWC170 (9), *C. judaicum* ILWC56 (10), *C. judaicum* ILWC59 (11), *C. cuneatum* ILWC40 (12), *C. cuneatum* ILWC37

(13), *C. chorassanicum* ILWC146 (14), *C. chorassanicum* ILWC23 (15), *C. yamashitae* ILWC53 (16), *C. yamashitae* ILWC214 (17). Die aufgetragenen DNA-Mengen sind in allen Spuren identisch.



C42

Fig.C42: Genomische Organisation der Satellitenfamilie *CaSat2* in der Gattung *Cicer*.

Autoradiogramm zweier Southern-Blots mit *RsaI*-restringiert genomischer DNA verschiedener Spezies, die gegen pCaEr855 (a) bzw. pCaEr453 (b) hybridisiert wurde. Folgende Spezies wurden untersucht: *C. reticulatum* PI489777 (1), *C. arietinum* ILC3279 (2), *C. arietinum* ILC482 (3), *C. echinospermum* ILWC39 (4), *C. echinospermum* ILWC 35 (5), *C. bijugum* ILWC32 (6), *C. bijugum* ILWC62 (7), *C. pinnatifidum* ILWC204 (8), *C. pinnatifidum* ILWC170 (9), *C. judaicum* ILWC56 (10), *C. judaicum* ILWC59 (11), *C. cuneatum* ILWC40 (12), *C. cuneatum* ILWC37 (13), *C. chorassanicum* ILWC146 (14), *C. chorassanicum* ILWC23 (15), *C. yamashitae* ILWC53 (16), *C. yamashitae* ILWC214 (17) und *C. anatolicum* PI383626 (18). Die aufgetragenen DNA-Mengen sind in allen Spuren identisch.

Die Hybridisierungsintensitäten von *CaSat2* zeigten im Vergleich zu *CaSat1* weit geringere Unterschiede (mit Ausnahme von *C. cuneatum*), und die Leiterbanden waren über die Länge von 20meren hinaus klar zu unterscheiden (Fig.C42a). In *C. chorassanicum* spiegelten die Hybridisierungssignale ein schon im Ethidiumbromid- gefärbten Agarosegel sichtbares Leitermuster wieder. Auffällig war die stärkere Intensität der kleineren Oligomer- und der Monomerbande bei *C. chorassanicum* im Vergleich zu *C. arietinum*, obwohl die Sonde aus *C. arietinum* stammte (Fig.C42a, Spur 14 und 15). Die im Gegensatz zur Kichererbse sichtbare 100 Bp-Bande weist auf die starke Konservierung der *CaSat2*-Monomersequenz in dieser Spezies hin. In der perennuellen Art *C. anatolicum* konnte mit einer *CaSat2*-Sonde das gleiche intensive Leitermuster wie in den annualen Spezies detektiert werden (Fig.C42b, Spur 18)

Sequenzen beider Satellitenfamilien sind in allen annualen *Cicer*-Spezies außer *C. cuneatum* in hohen Kopienzahlen vorhanden und bezüglich der Tandemorganisation und Länge der Monomereinheit konserviert. Die Anwesenheit stark divergierter *CaSat1*- oder *CaSat2*-Kopien im Genom von *C. cuneatum*, die unter den gegebenen Stringenzbedingungen keine Hybridisierungssignale mehr erzeugen, kann allerdings nicht vollständig ausgeschlossen werden. Leichte Variabilität in Feinstruktur und Abundanz machen beide Elemente für taxonomische und phylogenetische Betrachtungen im Genus *Cicer* verwendbar.

C.V.2. Generation von molekularen Markern für die genetische Karte des Kichererbsengenoms

Die Wiederholungseinheiten der Retrotransposonfamilie *CaRep* sind hochabundant und dispers über große Teile des Genoms von *C. arietinum* verteilt, weshalb sie zur Generation von molekularen Markern geeignet erscheinen. Darüberhinaus sind diese Sequenzen in sehr ähnlicher genomischer Organisation auch in nah verwandten Genomen wie etwa dem von *C. echinospermum* und *C. reticulatum* vorhanden, so daß auf *CaRep* basierende molekulare Marker sich auch zur Integration in eine bereits existierende genetische Karte

eignen, die auf einer interspezifischen Kreuzung beruht. Im Anschluß an die Charakterisierung der *CaRep*-Sequenzen wurde daher untersucht, ob diese sich zur Ableitung molekularer Marker eignen, welche in die genetische Karte des Kichererbsengenoms integriert werden können.

C.V.2.1. Ableitung eines geeigneten Oligonukleotids aus *CaRep* und seine Verwendung in einem „sequence specific amplified polymorphism“ (S-SAP)-Ansatz

Polymorphismen, die auf Variabilität der den einzelnen *CaRep*-Kopien benachbarten genomischen Sequenzen beruhen („insertion site polymorphism“), können ebenso wie Polymorphismen, die innerhalb der Struktur der verschiedenen *CaRep*-Wiederholungseinheiten auftreten („internal polymorphism“), zur Konstruktion oder Ergänzung einer genetischen Karte oder für phylogenetische Studien verwendet werden. Es gibt verschiedene, sowohl auf Hybridisierung als auch auf PCR beruhende Methoden, um solche Polymorphismen sichtbar zu machen. Eine sehr effiziente Möglichkeit bietet die S-SAP („sequence-specific amplification polymorphism“)-Methode, ein PCR-Ansatz, der durch Amplifikation von Restriktionsfragmenten unter Verwendung eines spezifischen Oligonukleotids die hohe Sensitivität der AFLP („amplified fragment length polymorphism“)-Technik mit einer hohen Spezifität verknüpft (Waugh et al., 1997).

Das spezifische Oligonukleotid RePT aus dem Polypurintrakt (PPT) von *CaRep* eignet sich für eine S-SAP-Analyse

Bei der hier verwendeten modifizierten Form des S-SAP (Ellis et al., 1998) wurde genomische DNA zunächst restringiert und die Restriktionsfragmente an einen passenden Adaptor ligiert. Die Produkte dienen in der nachfolgenden PCR als Template, wobei ein zum Adaptor komplementäres Kassettenoligonukleotid in Verbindung mit einem sequenzspezifischen Primer für *CaRep* verwendet wurde (Fig.C43a). Die Amplikons sollten daher nur aus solchen *CaRep*-Kopien bestehen, die in amplifizierbarem Abstand zur spezifischen Primerbindestelle eine Restriktionsschnittstelle aufweisen. Das spezifische Oligonukleotid RePT wurde aus den Randbereichen der konservierten *CaRep*-Sequenzen, d.h. aus der Konsensussequenz um den putativen Polypurintrakt (PPT) abgeleitet, der sich unmittelbar 5'-„upstream“ des rechten LTRs befindet (Fig.C43b). Fig.C43c zeigt zwei benachbarte Restriktionsschnittstellen entweder innerhalb des Elements („internal polymorphism“) oder die Insertionsstelle in der genomischen DNA flankierend („insertion site polymorphism“), die zu unterschiedlichen Amplifikatlängen führen können.

Die S-SAP-Amplifikate wurden in einem denaturierenden Polyacrylamidgel aufgetrennt, das den Molekulargewichtsbereich zwischen 200 und 1200 Bp hoch auflöste. Die radioaktive Markierung des spezifischen Oligonukleotids machte die aus der Kombination RePT-

Fig.C43: Generation von molekularen Markern mit Hilfe des Retrotransposons *CaRep* in *C. arietinum* und *C. reticulatum*.

(a) Schematische Darstellung der modifizierten S-SAP-Technik (Ellis et al., 1998). Restringierte genomische DNA wird an einen Adaptor ligiert und als Templat für eine PCR mit einem Kassetten- und spezifischen Oligonukleotid verwendet. (b) Position und Sequenz des *CaRep*-spezifischen Oligonukleotids RePT, das aus dem PPT-Bereich abgeleitet wurde. (c) Benachbarte Restriktionsschnittstellen innerhalb des Elements oder im flankierenden genomischen Bereich können Grundlage für das Auftreten von Polymorphismen sein.

/Kassettenoligonukleotid generierten Amplifikate selektiv sichtbar. Nebenprodukte mit zwei flankierenden Kassettenprimern blieben dabei unsichtbar.

Die Kombination mit dem *EcoRI*-Kassettenoligonukleotiden detektierte die meisten Polymorphismen

Insgesamt wurden 9 Adapter-/Kassettenoligonukleotidpaare in Kombination mit RePT getestet, wobei bis zu 40 verschiedene Amplifikationsprodukte im Polyacrylamidgel aufgetrennt wurden. Die Mehrzahl dieser Kombinationen (d.h. *EcoRI*-, *HindIII*-, *BamHI*-, *AluI*-, *MseI*- und *XhoI*-Kassetten) erzeugte polymorphe Banden zwischen den Eltern der interspezifischen Kreuzung von *C. arietinum* (ICC4958) x *C. reticulatum* (Pi489777; s.2.2.). Die meisten Polymorphismen konnten mit der Kombination RePT-/*EcoRI*-Kassette detektiert werden. Hier konnten insgesamt 12 polymorphe Banden auf dem Gel getrennt werden (FigC44; Spur 1,2).

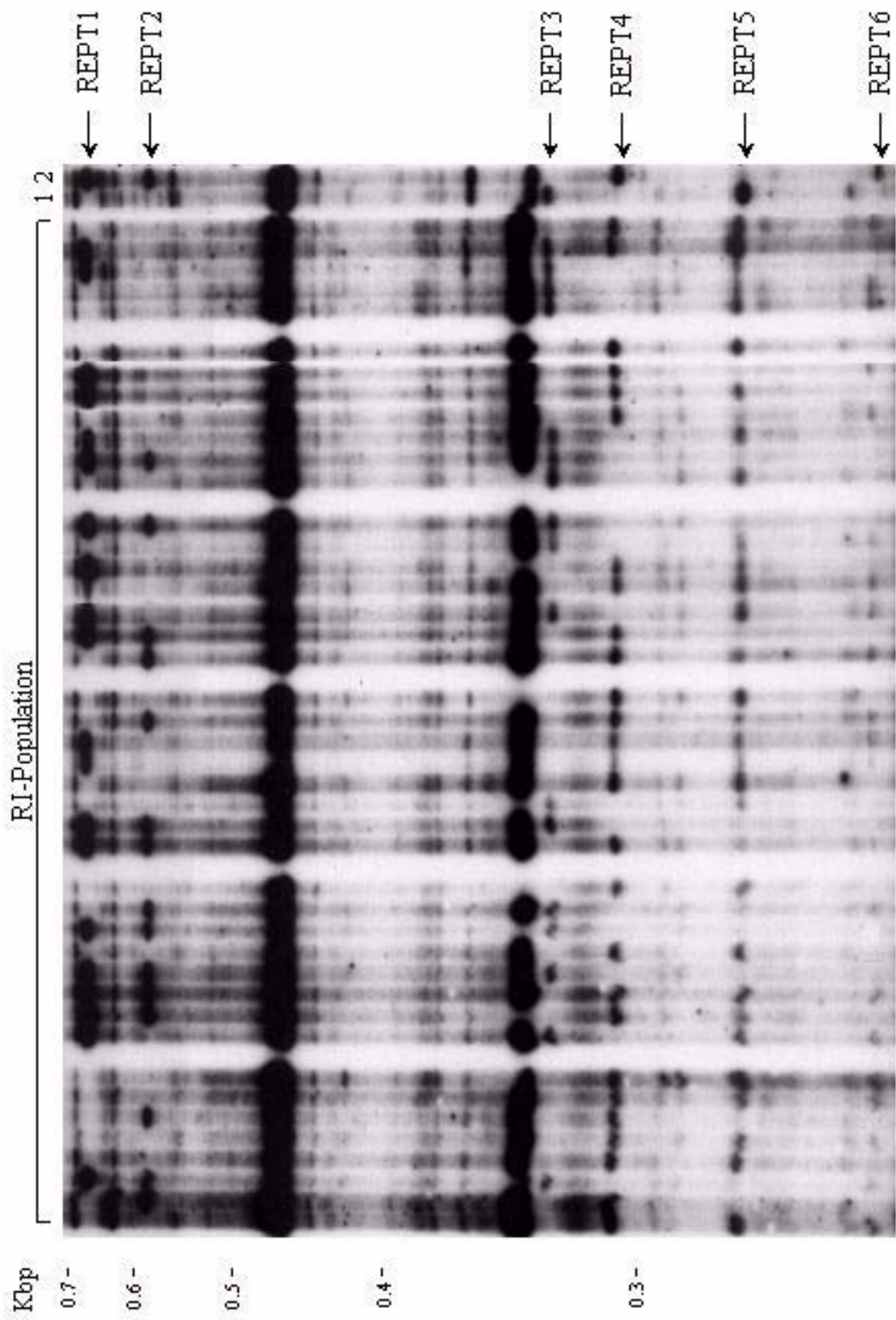
C.V.2.2. Lokalisation der polymorphen Marker auf der genetischen Karte

CaRep*-Marker zeigen Kopplung zu den Resistenzgenloci gegen *Fusarium oxysporum

Die existierende genetische Karte des Kichererbsengenoms (Winter et al., 1999) beruht auf der interspezifischen Kreuzung *C. arietinum* (ICC4958) x *C. reticulatum* (Pi489777). Die Nachkommenschaft aus 131 rekombinanten Inzuchtlinien (RILs) in der F7/F8-Generation ist nach den Mendelschen Regeln durch die obligate Selbstung theoretisch nahezu homozygot und segregiert für die Resistenz gegen das Pilzpathogen *Fusarium oxysporum f.sp. ciceri*. Die genetische Karte wurde durch Kopplungsanalyse von 364 DNA-Markern, Isoenzymen und Resistenzgenen konstruiert. Die verschiedenen DNA-Marker umfassen STMS (sequence tagged microsatellite sites)-, DAF („DNA amplified fingerprinting“)-, ISSR (inter simple sequence repeat)-, AFLP- und SCAR (sequence-characterized amplified region)- sowie cDNA-Loci und Loci von Resistenzgenanalogen (RGAs). Um die zwischen den Eltern polymorphen S-SAP-Produkte in die Karte integrieren zu können, wurde ihre Kopplung zu den vorhandenen Markern in 94 Individuen der RILs (rekombinante Inzuchtlinien) untersucht (Fig.C44). Von der Gesamtheit der polymorphen Banden segregierten sechs in dieser Population, die im folgenden REPT1-6 genannt werden. Die An- bzw. Abwesenheit von zwei dieser Banden (REPT 2, REPT 5) wich signifikant vom Verhältnis 1:1 ab. Die Integration in die Kopplungsgruppen der Karte wurde mit Hilfe von MAPMAKER v 3.0 bei einem „LOD score“ von 4 durchgeführt.

Fig.C44: Modifizierte S-SAP Analyse mit dem Oligonukleotid RePT aus dem PPT-Bereich des *Ty3-gypsy*-Elements *CaRep*.

Autoradiogramm mit S-SAP-Produkten, die das ³²P-markierte Oligonukleotid RePT inkorporiert haben und in einem Polyacrylamidgel aufgetrennt wurden. Als Kassettenprimer wurde das *EcoRI*-Oligonukleotid ohne selektive Basen verwendet. *C. reticulatum* Pi 489777 (1) und *C. arietinum* ILC 4958 (2) stellen die Elternlinien der interspezifischen Kreuzung dar, deren Nachkommen als rekombinante Inzuchtlinien verwendet wurden (RI-Population). Polymorphe Banden, die zur Kartierung verwendet wurden, sind durch Pfeile gekennzeichnet (REPT1-6).



C44

[MAPTIV1ö.CDR](#)

C45(1)

[MAPTIV2ö.CDR](#)

C45(2)

Fig.C45: Genetische Karte des *C. arietinum*-Genoms mit aus *CaRep* entwickelten Retrotransposon-Markern. Die genetische Karte (verändert nach Winter et al., 1999) wurde durch Kopplungsanalyse von 364 DNA-Markern, Isoenzymen und Resistenzgenen in einer interspezifischen *C. arietinum* x *C. reticulatum*-Kreuzung erstellt. Die bunten Balken stehen für STMS- (rot), AFLP- (blau), DAF- (schwarz), ISSR- (orange), SCAR- (gelb), cDNA- (grau), Isoenzym- (dunkelgrün) und RGA-(Resistenzgenanaloge) Loci (hellgrün). Grün und blau umrandete Felder symbolisieren Marker mit anonymer Sequenz aus repetitiver DNA. Zwei der *Fusarium oxysporum*-Resistenz-Loci sind mit Foc4 und Foc5 gekennzeichnet. Marker, die auf S-SAP mit dem Retrotransposonprimer RePT basieren, sind mit REPT 1-6 bezeichnet und über vier Kopplungsgruppen verstreut.

Die genetische Distanz zwischen zwei Loci basiert auf einer Kopplungsanalyse und kann mit der sog. „LOD score“- Methode abgeschätzt werden (Lander et al., 1987). Bei der Ermittlung der genetischen Distanz werden verschiedene Rekombinationshäufigkeiten verglichen. Die Wahrscheinlichkeit der genetischen Kopplung zweier Sequenzen wird jeweils durch die Wahrscheinlichkeit ihrer getrennten Vererbung geteilt, der Logarithmus aus diesem Wert entspricht dem „LOD score“. Ein „LOD score“ von z.B. 3 zeigt eine 1000-fach höhere Wahrscheinlichkeit der gekoppelten Vererbung im Vergleich zur Nicht-Kopplung der entsprechenden Loci an.

$$\text{LOD score} = z = \log \left[\frac{\text{Wahrscheinlichkeit eines gegebenen Kopplungswerts}}{\text{Wahrscheinlichkeit der Nicht-Kopplung}} \right]$$

Die genetische Distanz mit dem höchsten „LOD score“ wird für die weitere Rechnung verwendet, in der Praxis gelten Werte >3 als signifikant. Die „LOD score“- Methode liegt auch dem Programm MAPMAKER zu Grunde, das für viele Kopplungsanalysen in Pflanzengenomen verwendet wird.

Die Reihenfolge der Marker innerhalb einer Kopplungsgruppe wurde durch die „compare & try“- Funktion des Programms ermittelt. Alle sechs Marker zeigten Kopplung zu den vorhandenen Markern und konnten in die Karte integriert werden. Die einzelnen Loci waren über vier der acht großen Kopplungsgruppen (LG1, 2, 5, 6) verstreut und wurden mit dem Code REPT 1-6 benannt (Fig.C45). Die auf einer Kopplungsgruppe kartierenden Loci lagen nicht in Gruppen, sondern vereinzelt vor. Zwei der Retrotransposonmarker (REPT 2, REPT 5) wiesen Kopplung zu den Resistenzgenloci gegen *Fusarium oxysporum* Rasse 4 (Foc4) und Rasse 5 (Foc5) in der zweiten Kopplungsgruppe auf. Der kleinste genetische Abstand zu diesen Loci betrug jedoch immerhin noch 12.3 cM (Fig.C45).

Das disperse Element *CaRep* kann demnach erfolgreich zur Generation von DNA-Markern und deren Integration in die genetische Karte der Kichererbse verwendet werden.

C.V.2.3. Verifikation der Identität der generierten Marker durch Southern-Hybridisierung

Polymorphe S-SAP-Amplifikate werden ausschließlich durch *CaRep* generiert

Nachteilig bei der S-SAP-Methode (und anderen PCR-Techniken) zur Detektion von Polymorphismen ist die Anonymität der erzeugten Banden im Polyacrylamidgel. Eine Möglichkeit, um die Identität der Amplifikate nachzuweisen, ist ihre Sequenzierung. Schneller und billiger ist bei dem hier verwendeten, modifizierten Retrotransposon-S-SAP jedoch eine Southern-Hybridisierung mit einer retrotransposonspezifischen Sonde, da bereits klonierte verwandte Sequenzen vorhanden sind. Dazu wurde eine S-SAP-PCR mit unmarkierten RePT-Primern durchgeführt, im Polyacrylamidgel aufgetrennt und auf eine Nylonmembran übertragen. Durch Hybridisierung mit dem radioaktiv markierten Fragment eines *CaRep3*-Klons (pCaEr776), sollten die Banden ermittelt werden, die tatsächlich von *CaRep* generiert wurden. Vom spezifischen Oligonukleotid RePT amplifizierte *CaRep*-Fragmente überlappen ca. 1600 Bp mit dem 3'-Bereich des *CaRep3*-Klons und sollten auch bei sehr stringenten Hybridisierungsbedingungen zu spezifischen Signalen führen.

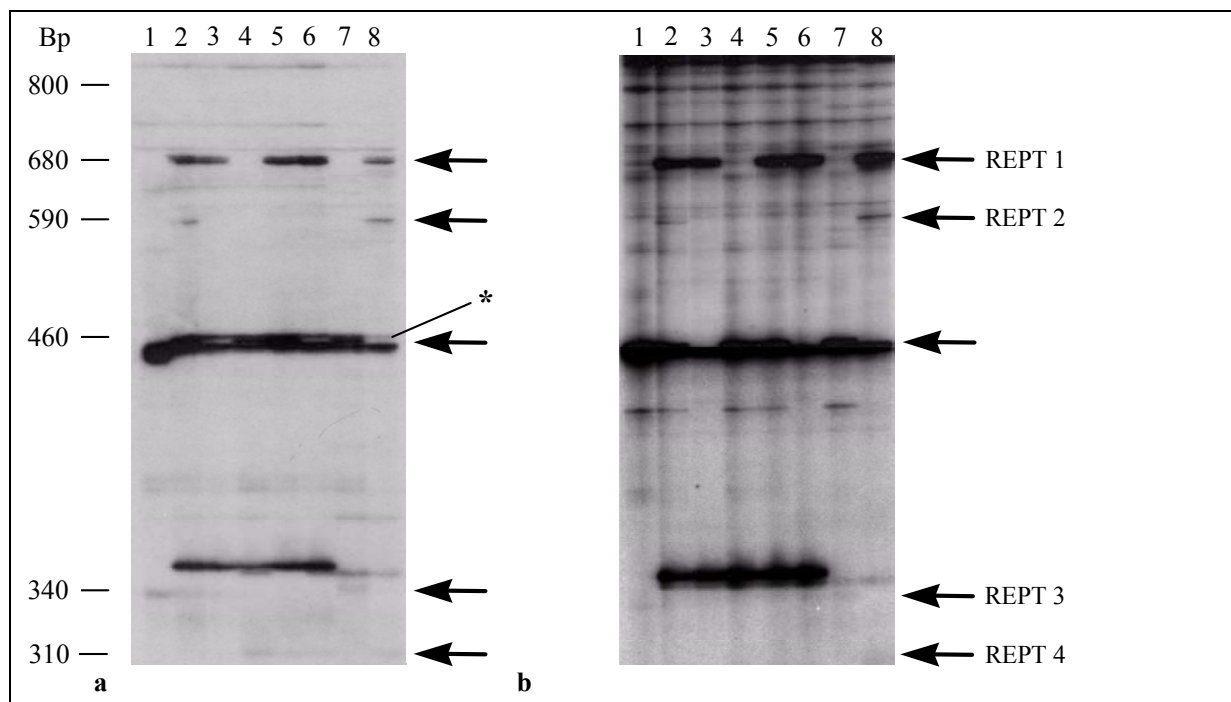


Fig.C46: Kartierung von Retrotransposonkopien durch S-SAP: Verifikation der Sequenzidentität durch Southern-Analyse.

(a) Autoradiogramm mit S-SAP Produkten, die das ³²P-markierte Oligonukleotid RePT enthalten und in einem Polyacrylamidgel aufgetrennt wurden. Als Templat diente die DNA ausgewählter RI-Linien (1-6), *C. reticulatum* Pi 489777 (7) und *C. arietinum* ILC 4958 (8). Einige der polymorphen Banden, die zur Kartierung verwendet wurden, sind durch Pfeile gekennzeichnet (REPT1-4). (b) Autoradiogramm eines Southern-Blots mit identischer Beladung. Diese S-SAP Produkte enthalten jedoch unmarkierte RePT-Primer. Der Blot wurde gegen das Fragment aus pCaEr776 (*CaRep3*) hybridisiert. Alle polymorphen Produkte zeigen Signale (REPT1-4), nur eine monomorphe Bande ließ sich nicht detektieren (*).

Ein Vergleich der S-SAP-Bandenmuster aus Eltern und ausgewählten Individuen der RI-Linien (Fig.C46a) mit dem Hybridisierungsmuster der jeweils nicht-radioaktiven Reaktionen (Fig.C46b) zeigte, daß alle starken S-SAP-Amplifikate auch auf dem Southern-Blot Signale erzeugten. Alle im Bereich zwischen 300-1000 Bp aufgetrennten polymorphen Banden waren sichtbar, lediglich ein monomorphes Produkt (460 Bp) ließ sich nicht durch Hybridisierung detektieren, was auf eine signifikant von pCaEr776 abweichende Sequenz oder ein völlig *CaRep*-unabhängiges purinreiches (da mit dem PPT hybridisierend) Fragment hindeutet. Demnach werden alle polymorphen S-SAP Banden durch *CaRep*-ähnliche Sequenzen generiert. Die deutlich höhere Zahl sichtbarer Banden und der leichte Hintergrundschmier im Southern-Blot kommen möglicherweise durch die ebenfalls detektierten reinen *EcoRI*-Kassettenoligonukleotid-Produkte zustande, sofern sie verwandte Sequenzen zur hochrepetitiven *CaRep*-Familie enthalten.

Das disperse *CaRep*-Element generiert demnach Polymorphismen zwischen den Eltern einer interspezifischen Kreuzung, die durch PCR detektierbar sind und als Marker die genetische Karte des Kichererbsengenoms ergänzen können. Die Integration der Marker in die Karte ist identisch mit der Kartierung einiger *CaRep*-Kopien, da diese DNA-Marker tatsächlich Loci des repetitiven Elements im Genom darstellen.

D. DISKUSSION

D.I. Die Satellitenelemente *CaSat1* und *CaSat2*

D.I.1. *CaSat1* und *CaSat2* - dominante Bestandteile des Genoms von *C. arietinum*

Zwei hochrepetitive Sequenzfamilien aus dem Genom von *C. arietinum* konnten aufgrund ihrer genomischen Organisation als Satellitenelemente identifiziert werden und wurden *CaSat1* und *CaSat2* genannt. Es handelt sich um die hier zuerst beschriebenen Familien von Satelliten-DNA bei der Kichererbse (Staginnus et al., 1999). Beide Elementfamilien sind spezifisch für die Gattung *Cicer* und machen einen großen Teil des Heterochromatins der Chromosomen von *C. arietinum* aus. Gemeinsam mit dem Retroelement *CaRep* stellen sie vermutlich die abundantesten Sequenzen im Kichererbsengenom dar, wie man aus der vollständigen Bedeckung des DAPI (4,6 Diamino-2-phenylindol)-positiven Heterochromatins in Interphase- und Metaphasekernen mit den Signalen aller drei repetitiven Sequenzfamilien schließen kann.

D.I.1.1. Struktur und Sequenz der Monomere von *CaSat1* und *CaSat2*

Sowohl *CaSat1* als auch *CaSat2* bilden lange Tandemwiederholungen, in denen die einzelnen Monomere in gleicher Orientierung hintereinander ausgerichtet sind („head-to-tail arrangement“). Die Länge und Sequenz dieser Monomere - auch die unabhängig von dieser Untersuchung isolierten *CaSat1*-Monomere von M. Frediani (AJ012823)- sind jeweils innerhalb der Familien konserviert und zeigen keine Verwandtschaft zu anderen Sequenzen in den Datenbanken bzw. im Kichererbsengenom.

Mit einer Länge von 162-167 Bp liegen die *CaSat1*-Monomere im bevorzugten Größenbereich vieler pflanzlicher Tandemelemente (160-180 Bp bzw. 320-370 Bp; Ingham et al., 1993; Harrison & Heslop-Harrison, 1995; Kamm et al., 1995; Schmidt & Heslop-Harrison, 1996; Alix et al., 1998; Ananiev et al., 1998a; Heslop-Harrison et al., 1999). Die Monomergröße der zweiten Satellitenfamilie liegt mit 100 Bp deutlich unter diesem Bereich, was ebenfalls für etliche andere Tandemelemente zutrifft (Anamthawat-Jónsson & Heslop-Harrison, 1993; Kolchinsky & Gresshoff, 1995; Morgante et al., 1997; Brown et al., 1998; Jakowitsch et al., 1998; Nagaki et al., 1998). Die Monomerlänge von Tandemelementen wird oft mit der nukleosomalen Struktur in heterochromatischen Bereichen in Zusammenhang gebracht. Die Abstände der Nukleosomen sind zwar fixiert, können aber in unterschiedlichen Regionen des pflanzlichen Genoms variieren. Innerhalb einer Monomersequenz konservierte d(A)_n-Trakte verursachen im Zusammenspiel mit der Helixstruktur eine gebogene DNA-Konformation, die die Bildung von Nukleosomen ermöglicht (Vershinin et al., 1998). Die sequenzabhängige, durch die Monomerabfolge verursachte Periodizität dieser Konformation wurde außer in Tabak auch in *Brassica* nachgewiesen (Matýasek et al., 1996) und stellt auch bei Tieren den strukturellen Code für den Nukleosomenabstand dar (Radic et al., 1987). Dieser wiederum beeinflusst die Chromatinkondensation und damit den Aufbau

heterochromatischer Bereiche. Satelliten mit der häufigen Monomerlänge von 160-180 Bp umspannen dabei genau ein Nukleosom. Für Grasgenome konnte dagegen gezeigt werden, daß mehrere der kleineren Wiederholungseinheiten von ~118 Bp je zwei Nukleosomen umfaßten (Vershinin et al., 1998). Satelliten mit gleichen Monomerlängen diktiert demnach immer den gleichen Nukleosomenabstand und damit eine charakteristische Chromatinkondensation. Möglicherweise beeinflussen auch im Kichererbsengenom die beiden AT-reichen, deutlich voneinander getrennten Heterochromatinabschnitten lokalisierten Satelliten die Heterochromatinbildung über eine Nukleosomenpositionierung.

Abgesehen von kurzen direkten und inversen Sequenzwiederholungen fehlen in den Monomersequenzen beider Satelliten längere Substrukturen. Man kann daher nicht, wie etwa bei anderen Tandemelementen mit 60 Bp-Wiederholungen in *Vicia faba* und *Brassica* auf eine Ableitung aus tRNA-Genen schließen (Kato et al., 1984; Benslimane et al., 1986; Harrison & Heslop-Harrison, 1995). Beide Monomersequenzen aus Kichererbse enthalten einen hohen Adenosin- und Thyminanteil, was von anderen Tandemelementen aus Leguminosen ebenfalls bekannt ist (Xia & Erickson, 1993, Schmidt & Kudla, 1996; Zhao et al., 1998). Der hohe AT-Anteil stimmt mit der chromosomalen Lokalisation von *CaSat1* und *CaSat2* überein, die einen großen Teil des pericentrischen Heterochromatins darstellen, das mit DAPI färbbar und demnach AT-reich ist. Eine starke Methylierung der Sequenzen - wofür ihre geringfügige Restriktion mit dem methylierungssensitiven Isoschizomerpaar *HpaII/MspI* spricht - ist zwar naheliegend (Hemleben, 1993), kann aber aufgrund ihres hohen AT-Gehalts ohne Antimethylcytosin-Antikörpernachweis (South-Western Blot) nicht bestätigt werden (Kamm et al., 1994).

D.I.1.2. *CaSat1* und *CaSat2* weisen eine übergeordnete Multimerstruktur auf

Sowohl bei *CaSat1* als auch bei *CaSat2* konnte eine nur geringe Sequenzverwandtschaft zwischen den meisten direkt benachbarten Monomeren festgestellt werden. Dagegen sind Sequenzen nicht benachbarter Einheiten, die von verschiedenen Klonen und möglicherweise im Genom auseinanderliegenden Bereichen des Satelliten stammen, enger miteinander verwandt. Diese Beobachtungen passen zum Modell einer übergeordneten Hierarchie in der Struktur von langen Tandemabfolgen, das von Willard & Wayne (1987a) für die Monomere der α -Satelliten im Menschen zu Grunde gelegt wurde. Demnach existieren innerhalb der Tandemfolge Multimere aus mehreren Monomeren, die untereinander divergieren, jedoch zum äquivalenten Monomer des folgenden Multimers hohe Homologie aufweisen. Solche übergeordneten Hierarchien homologer Bausteine können durch Amplifikation oder Sequenzhomogenisation in einer Gruppe von Wiederholungseinheiten in nicht sehr weit zurückliegenden Entwicklungsphasen des Satelliten entstehen (Willard & Wayne, 1987a). In Soja wird sowohl höhere, als auch niedrigere Ähnlichkeit benachbarter Monomere im Vergleich zu Wiederholungseinheiten von verschiedenen klonierten Fragmenten desselben Tandemelements berichtet. Dies legt die Existenz von Monomer- und Dimer-tandemwiederholungen nahe (Morgante et al., 1997). Hinweise auf eine übergeordnete Multimerstruktur der Kichererbsensatelliten konnten nicht

nur beim Sequenzvergleich der Monomere gefunden werden. Das ausgeprägte Leiterbandenmuster von elf verschiedenen Restriktionen läßt auf das Vorhandensein vieler konservierter Restriktionsschnittstellen innerhalb der *CaSat2*-Monomere schließen. Bei der Sequenzierung von 20 Monomeren konnten jedoch nur wenige vollständig erhaltene sowie einige durch Punktmutationen degenerierte Schnittstellensequenzen pro Monomer gefunden werden. Man muß daher davon ausgehen, daß die klonierten Monomere von *CaSat2* aus einer divergierten Multimerregion stammen, die eine der höhermolekularen Banden darstellt und in der viele Erkennungssequenzen bereits degeneriert sind. Für beide Satelliten wurden demnach Bereiche aus einer übergeordneten Multimerstruktur mit leicht divergierenden Einheiten sequenziert. Die Southern-Analyse bestätigt die Konservierung und Amplifikation von größeren Multimeren bei beiden Satelliten. Für eine genauere Analyse der Periodizität von Mono- und Multimeren ist jedoch die Zahl der hier analysierten, direkt benachbarten Monomere zu gering.

D.I.1.3. Enthält *CaSat1* eingestreute Retrotransposonsequenzen?

Für keinen Satelliten konnten angrenzende genomische Bereiche sequenziert werden. Die einzige Ausnahme davon bildet Klon pCaEr782 (*CaSat1*), vorausgesetzt, die Klonierung zweier unabhängiger *EcoRI*-Fragmente kann ausgeschlossen werden (C.I.1.). Die benachbarte Retrotransposonsequenz könnte in diesem Fall aus dem Randbereich einer Tandemkette stammen. Dabei lassen die kleineren, in den distalen Bereichen des Heterochromatinblocks versprengten Sequenzanhäufungen eine Klonierung benachbarter genomischer Bereiche wahrscheinlicher erscheinen als die langen Multimerabfolgen der Hauptloci. Eine weitere Erklärung für die fremde Sequenz in unmittelbarer Nachbarschaft der Satellitenmonomere wäre die Vermischung der Satelliteneinheiten mit anderen repetitiven Elementen. Im Tomatengenom sind z.B. Minisatellitenstrukturen in bestimmten Bereichen eines längeren Satelliten vorhanden (Broun & Tanksley, 1993). Ananiev et al. (1998a) berichten von der Einstreuung vollständiger Kopien der Retrotransposons *Zeon*, *Prem2* und *Grande* zwischen die Monomere eines 180 Bp-Satelliten im sog. knopfförmigen („knob“) Heterochromatin von *Zea mays*. Auch aus *Arabidopsis* ist die Durchmischung des centromerischen 180 Bp langen Satellitenelements *pAL* mit degenerierten Kopien des LTR-Retrotransposons *Athila* bekannt (Pelissier et al., 1995; Pelissier et al., 1996). Die Entstehung der Satellitensequenz des Klons pCaEr782 aus einem Retrotransposon (Langdon et al., 2000b) ist aufgrund fehlender Sequenzhomologien unwahrscheinlich. Eine Assoziation von Retrotransposon- mit Satellitensequenzen durch Insertion des mobilen Elements in die Tandemabfolge ist durch die unvollständige Retrotransposonstruktur im flankierenden Bereich (5'-LTR) und seine damit verbundene Immobilität auszuschließen. Da nur ein Ende des Satellitentandems kloniert wurde, ist das Vorhandensein der restlichen Retrotransposonsequenz auf der anderen Seite ungewiß, und auch eine durch „rolling circle“-Replikation (Walsh, 1987) denkbare Satelliteninsertion in das Retrotransposon bleibt zweifelhaft. Am wahrscheinlichsten erscheint eine Aneinanderlagerung beider Sequenzen durch nichthomologe Rekombination zwischen kurzen ähnlichen Sequenzen im Satellitentandem und innerhalb des

Retrotransposons. Ein solcher Mechanismus wurde für den unmittelbaren Übergang von PSR2-Satellitelementen in Sequenzen des Retroelements *NATE* in der parasitischen Wespe *Nasonia vitripennis* postuliert (McAllister et al., 1999). Die Ansammlung mobiler Elemente in nicht-codierenden Regionen wie etwa Satelliten-DNA ist vorteilhaft, da hier letale Mutationen unwahrscheinlich sind (Charlesworth et al., 1994). Nach dem sog. Zuwachs-Modell („accretion model“; McAllister & Werren, 1999), akkumulieren inserierte fremde Sequenzen durch Amplifikation und ungleichen Strangaustausch („unequal crossing-over“) des Satelliten entweder in Inseln oder am Ende eines Tandemstrangs.

D.I.1.4. Die chromosomale Lokalisation gibt Hinweise auf einen möglichen Ursprung von *CaSat1*

Obwohl *CaSat1*-Sequenzen hauptsächlich im pericentrischen Heterochromatinblock der Chromosomen A und B lokalisiert sind, findet man sie auch in den distalen Regionen des Heterochromatins von A und B sowie der meisten anderen Chromosomen. Daher kann man annehmen, daß im Laufe der Evolution dieser Sequenzfamilie entweder eine Verteilung von *CaSat1*-Monomeren im Genom stattgefunden hat oder einige der verstreuten Loci eine starke Amplifikation erfahren haben. Auch in *Brassica* kommen chromosomenspezifische repetitive Elemente vor, die in pericentrischer Umgebung lokalisiert sind, jedoch leicht divergente Sequenzen in anderen chromosomalen Bereichen besitzen (Harrison & Heslop-Harrison, 1995). Kleine Anhäufungen von Satellitensequenzen, die abseits von ihren centromerischen oder subtelomerischen Hauptloci im Genom verstreut sind, wurden ebenfalls bei Gräsern (Anamthawat-Jónsson & Heslop-Harrison, 1993), *Arabidopsis arenosa* (Kamm et al., 1995) und *Beta procumbens* (Schmidt & Heslop-Harrison, 1996) gefunden. Ihr Verbreitungsmechanismus ist unbekannt, kann aber neben nicht-homologem Austausch zwischen verschiedenen Chromosomen oder Schwesterchromatiden auch extrachromosomale, sog. „rolling-circle“-Amplifikation (Walsh, 1987) mit nachfolgender Insertion an neuen Genomorten beinhalten (McAllister & Werren, 1999). Für *TrsA*, eine Tandemelement-Familie aus Reis, dessen Monomere von direkten Sequenzwiederholungen flankiert werden, wurde ein Transpositionsmechanismus für die Verbreitung beschrieben, der sich von der typischen Transposition bei Retro- und Klasse II-Elementen unterscheidet (Ohtsubo & Ohtsubo, 1994). Der Transpositionsprozess soll ähnlich dem bei SV40 und alphoid-ähnlichen Sequenzen ablaufen, der direktes Ausschneiden oder die Bildung von cDNA nach Transkription einiger Monomere und nachfolgende Amplifikation während der Replikation des Chromosoms umfaßt.

Für die beiden Hauptloci von *CaSat1* auf den Chromosomen A und B ist die enge Nachbarschaft zu den 18S-5.8S-25S rDNA-Abfolgen auffällig. Ob die beiden Signale direkt aneinandergrenzen oder voneinander getrennt liegen, kann aufgrund der relativ niedrigen Auflösung der Hybridisierungssignale auf den kondensierten Metaphasechromosomen nicht festgestellt werden. Eine partielle oder - wie auf Chromosom B - großflächige Überlappung der Signale kann nicht ausgeschlossen werden. Beide Sequenzen nehmen jedoch nicht

identische Loci ein, da die Signale auf keinem der beiden Chromosomen vollständig überlappen. Die auffällige räumliche Kopplung des Satelliten mit den rDNA-Sequenzen kann durch Sequenzverwandtschaft von *CaSat1* mit den Wiederholungseinheiten der „intergenic spacer region“ (IGS) erklärt werden. Die IGS liegt zwischen den tandemartig angeordneten codierenden Regionen der 18S-5.8S-25S rRNA und enthält transkribierte („5'-3'-external transcribed spacer“; ETS) sowie nichttranskribierte Abschnitte („non-transcribed spacer“; NTS). Neben den Sequenzen zur Transkriptionsregulation kann die IGS verschiedene, in singuläre Sequenzen eingebettete, repetitive Sequenzen enthalten und spielt vermutlich eine Rolle bei der Bildung übergeordneter Chromatinstrukturen im rDNA-Bereich (Griess et al. 1993). Im Gegensatz zu den codierenden Regionen sind Länge und Sequenz der IGS selbst zwischen nahe verwandten Arten sehr heterogen (Hemleben, 1993), verursacht durch die Anwesenheit bzw. Amplifikation verschiedener repetitiver Sequenzfamilien (Gerstner et al., 1988, Schiebel et al., 1989; Hemleben, 1993). In der IGS von *Vigna radiata* konnte ein 174 Bp-Tandemelement identifiziert werden, der unabhängig von der IGS als prominentes Satellitenelement existiert und in bestimmten Bereichen des Genoms konzentriert ist (Unfried et al., 1991). In der verwandten Art *Vigna angularis* existiert ein dazu homologes IGS-Motiv, das jedoch keine unabhängigen Wiederholungseinheiten bildet. Neben *V. radiata* weisen weitere Arten der Untergattung *Ceratotropis* (Gattung *Vigna*) eng mit IGS-Motiven verwandte, aber unabhängige Satellitenelemente auf (Hemleben, 1993). Auch andere Leguminosen zeigen ähnliche Homologien: Die *Vicia faba* IGS-Region enthält ein 336 Bp-Tandemelement, das außerdem in dispersen Blöcken im Genom verteilt ist (Kato et al., 1985; Maggini et al., 1991). *Phaseolus vulgaris* weist Sequenzen auf, die mit einem 166 Bp-Element im IGS verwandt sind (Falquet et al., 1997). In tierischen Genomen wie Maus (Arnheim et al., 1980), *Chironomus* (Schmidt, 1984) und *Triturus* (DeLuccini et al., 1988) existieren ebenfalls unabhängige Satelliten mit starker Verwandtschaft zu IGS-Wiederholungseinheiten. In einigen Fällen wird von der Auswanderung von IGS-Wiederholungseinheiten in andere genomische Bereiche ausgegangen (Unfried et al., 1991; Hemleben, 1993), während Sequenzvergleiche in *Chironomus* und *Vicia faba* nahelegen, daß die IGS-Wiederholungseinheiten durch Transposition unabhängiger Sequenzen in die rDNA-Einheiten entstanden sind. Als Mechanismus für dieses Phänomen wird Transposition vermutet, wobei z.B. eine Wiederholungseinheit mit einem Teil des IGS transkribiert, durch „self-priming“ interner Strukturen revers transkribiert und nach Transfer zu anderen genomischen Orten dort durch „ungleiches Crossing-over“ amplifiziert werden kann. Andere Sequenzen zeigen Homologien zu „hot spots“ der Rekombination (DeLucchini et al., 1988). Eine Sequenzverwandtschaft zwischen *CaSat1* und einem IGS-Wiederholungsmotiv in Kichererbse ist angesichts ähnlicher Phänomene in nah verwandten Arten denkbar. Keine der genannten Leguminosensequenzen zeigte signifikante Homologien zu *CaSat1*-Sequenzen, was wegen der hohen Variabilität der IGS-Wiederholungseinheiten (Falquet et al., 1997) und der Spezifität der Satellitensequenz für die Gattung *Cicer* auch nicht zu erwarten ist. Aufschluß über die Entstehung von *CaSat1* durch starke Amplifikation eines IGS-Wiederholungsmotivs und anschließende Dispersion weiterer Tandem-„Cluster“ im Genom (oder umgekehrt: die Einwanderung verstreuter Sequenzen und nachfolgende Amplifikation

in der rDNA-Region) kann letztendlich nur die Sequenzierung von IGS-Bereichen aus *C. arietinum* geben.

D.I.1.5. Impliziert die Lokalisation von *CaSat2* einen möglichen funktionellen Aspekt?

CaSat2 ist auf allen acht Chromosomen von *C. arietinum* in centromerischer Umgebung lokalisiert. Repetitive Sequenzen wie Retrotransposons und Satellitenelemente sind bei vielen Pflanzen als dominante Bestandteile des centromerischen Heterochromatins bekannt (Kamm et al., 1994; Harrison & Heslop-Harrison, 1995; Kamm et al. 1995; Ananiev et al., 1998b; Dong et al., 1998; Kubis et al., 1998b). Das Centromer von *Arabidopsis* wird vom 180 Bp-Satellitenelement pAL (Martinez-Zapater et al., 1986) eingenommen, Reis und Mais weisen centromerischen Satelliten auf (Ananiev et al., 1998b; Dong et al., 1998), und das Centromermodell für *Beta* umfaßt verschiedene Satelliten (Gundillis et al., 2001), jedoch wird ihre Funktion bei der Centromeraktivität kontrovers diskutiert (Csink & Henikoff, 1998; Henikoff, 1998; Schubert, 1998; Copenhaver & Preuss, 1999). Die Monomere von *CaSat2* sind kürzer als die der meisten untersuchten Centromersatelliten, die zur 160-180 Bp-Gruppe mit strenger nukleosomaler Organisation gehören, allerdings werden aus *Saccharum* und *Sorghum* auch Centromer-assoziierte 140 Bp-Satelliten beschrieben (Nagaki et al., 1998; Zwick et al., 2000). In mehreren pflanzlichen Centromeren konnten Homologien zwischen dem Bindemotiv für CENB-P aus *homo sapiens* und konservierten Boxen aus Satellitenmonomeren gefunden werden (Nagaki et al., 1998; Heslop-Harrison et al., 1999; Zwick et al., 2000; Gundillis et al., 2001). Auch centromerische *Ty3-gypsy*-Elemente und andere centromerische repetitive Elemente enthalten homologe Motive (Aragon-Alcaide et al., 1996; Presting et al., 1998; Nonomura & Kurata, 1999). Die *CaSat2*-Monomere enthalten Motive mit geringer Ähnlichkeit zu einer zum CENB-P-Bindemotiv homologen Box aus CCS1, einer centromerischen Sequenz in *Brachypodium sylvaticum* (Waldzenke, Poaceae; Aragon-Alcaide et al., 1996). Zum CENB-P- Bindemotiv aus *H. sapiens* selbst oder den homologen Boxen aus *Beta procumbens* (Gundillis et al., 2001), *Oryza sativa*, *Hordeum vulgare* oder *Zea mays* (Aragon-Alcaide et al., 1996) bestehen keine signifikanten Homologien. Von allen untersuchten Kichererbsenelementen nimmt *CaSat2* auf allen Chromosomen die dem Centromer am nächsten liegende Position ein. Allerdings läßt sich durch die niedrige Auflösung der *in-situ*-Hybridisierung auf Metaphasechromosomen nicht feststellen, ob die sehr intensiven Signale die primäre Einschnürung bedecken oder nur überstrahlen. Weniger abundante und bisher unbekannte repetitive Elemente der Kichererbse, die zwischen die *CaSat2*-Monomere eingestreut sind oder dichter am Centromer liegen, sind denkbar. Insgesamt ist nach dem bisherigen Kenntnisstand nicht ausgeschlossen, daß *CaSat2* die zentrale Centromersequenz darstellt. Um dies festzustellen, ist jedoch eine physikalische Feinkartierung und Sequenzierung der centromerischen Region Voraussetzung.

D.I.1.6. Die vermutliche Evolution der beiden *CaSat*-Elemente

Sowohl die Tandemorganisation als auch die Monomerlänge der beiden gattungsspezifischen Satelliten ist zwischen allen annuellen *Cicer*-Spezies konserviert, lediglich *C. cuneatum* bildet eine Ausnahme. Beide Satellitenfamilien waren demnach schon in einem Vorläufer der nach Kazan & Muehlbauer (1991) monophyletischen annuellen Spezies vorhanden. Die Trennung eines Vorläufers von *C. cuneatum* von der gemeinsamen Linie der annuellen Spezies mit der perenniellen Art *C. anatolicum* ist nach Mutationsraten in flankierenden Sequenzen von Microsatellitenloci vor ca. 125 000 Jahren anzunehmen (Choumane et al., 2000). Nach dieser Trennung und vor der Abspaltung der *C. chorassanicum*-Linie (vor ca. 110 000 Jahren) kann eine Amplifikation der beiden Satellitenelemente *CaSat1* und *CaSat2* zu dominanten genomischen Komponenten angenommen werden. Zumindest bei *CaSat1* deuten die Intensitäten der Leiterbandenmuster auf unterschiedlich starke Amplifikationen und teilweise auch auf Divergenz eines Teils der Wiederholungseinheiten während der weiteren Evolution der verschiedenen *Cicer*-Arten hin.

D.I.1.7. *CaSat1* und *CaSat2* sind unterschiedlich homogenisierte Satellitenfamilien im Genom von *C. arietinum*

CaSat2 macht - wie aus den starken *in-situ*-Signalen hervorgeht - den Hauptteil des pericentrischen Heterochromatinblocks aller Chromosomen aus und stellt mit großer Wahrscheinlichkeit die häufigere Satellitenfamilie dar. Das Fehlen zusätzlicher, verstreuter Hybridisierungssignale sowie die stärkere Konservierung von Sequenz und Länge der Monomere unterscheidet diese Familie von *CaSat1*. Obwohl die höhere Anzahl sequenzierter Klone die Statistik der Sequenzhomogenität von *CaSat1* möglicherweise verzerrt, können hier Wiederholungseinheiten mit divergenteren Sequenzen ausgemacht werden, die eventuell den schwächeren Banden im Leitermuster zuzuordnen sind. Dies gibt Anlaß zur Hypothese, daß beide Satellitenfamilien in unterschiedlichen Homogenisationszuständen im Kichererbsengenom vorliegen. Verantwortlich für die Sequenz- und Längenhomogenität der Mono- bzw. Multimere einer Tandemabfolge ist die Verbreitung und Fixierung aufgetretener Sequenzvariationen („molecular drive“), die sich automatisch innerhalb einer Population einstellen. Mechanismen wie ungleiches „crossing-over“ zwischen Chromatiden oder Chromosomen sowie Genkonversion und Transposition sorgen hierbei je nach Sequenzfamilie innerhalb einer bestimmten Zeit für eine Fixierung der Sequenzveränderung unter den Kopien (Dover 1982; Strachan et al., 1985). Beim Menschen konnte eine Homogenisation von Monomerlängen bei α -Satelliten inter- und intrachromosomal nachgewiesen werden (Willard, 1991). Im Wechsel zu den Homogenisationsmechanismen führen Mutationen und andere Rearrangements zu einem heterogenen Übergangszustand, aus dem sich möglicherweise neue Unterfamilien des Satelliten bilden, die wiederum homogenisiert werden können. Einige gut untersuchte Satellitenfamilien aus der Gattung *Cucurbita* zeigen, daß homogene und heterogene Satellitenfamilien in einem Genom nebeneinander vorliegen können, während deren Verhältnis in nah verwandten Arten genau umgekehrt ist und auch die Kopienzahlen deutliche Differenzen aufweisen (King et al., 1995). King und Mitarbeiter stellten fest, daß zumindest ein Satellit pro Spezies stark amplifiziert

und homogenisiert vorlag. Dies trifft bei der Kichererbse für *CaSat2* zu, der relativ homogen erscheint und auf allen Chromosomen mit ähnlich hoher Abundanz vorliegt. *CaSat1*, der sich durch heterogenere Monomere, starke Amplifikation auf zwei Chromosomen sowie versprengten Anhäufungen auf den übrigen auszeichnet, ist offenbar nicht auf allen Chromosomen vollständig homogenisiert. Es läßt sich dabei nicht entscheiden, ob *CaSat1* ein jüngeres, sich neu entwickelndes Element ist oder die Relikte einer älteren Tandemfolge darstellt, deren Wiederholungseinheiten teilweise verlorengegangen sind. Letzteres wurde für einen 350 Bp Satelliten von *Cucurbita maxima* und *C. lundelliana* berichtet (King et al., 1995).

D.I.2. Verwendung der tandemrepetitiven Elemente für phylogenetisch-taxonomische Betrachtungen

Hochrepetitive Tandemelemente stellen dynamische Komponenten pflanzlicher Genome dar (Charlesworth et al., 1994), die durch Akkumulation von Mutationen schnell verändert werden können. Nachfolgende Amplifikation und Homogenisierung (Genkonversion) führen zu ihrer Unterrepräsentation, Abwesenheit oder starker Divergenz in nah verwandten Genomen, die zu phylogenetischen Studien - auch bei Fabaceen (Calderini et al., 1997) - herangezogen werden können (Kamm et al., 1995; Vershinin et al., 1996, Alix et al., 1998; Brown et al., 1998; Takeda et al., 2000). Unterschiedlich starke Amplifikation von Satelliten bestätigt die evolutiven Verwandtschaftsverhältnisse zwischen Spezies der Gattung *Cucumis* (Helm & Hemleben, 1997), *Cucurbita* (King et al., 1995) und *Beta* (Kubis et al., 1997), ebenso bestärken genusspezifische Satelliten taxonomische Klassifizierungen (Schmidt et al., 1991; Schmidt & Heslop-Harrison, 1993).

D.I.2.1. *C. cuneatum* ist ein Außenseiter unter den annualen *Cicer*-Spezies

Die Abwesenheit von *CaSat1*- und *CaSat2*-Hybridisierungssignalen in allen untersuchten Akzessionen der Wildart *C. cuneatum* ist auffällig. Dies bedeutet jedoch nicht zwingend, daß beide Satellitenfamilien in dieser Art völlig fehlen. Es ist auch möglich, daß sie im Genom von *C. cuneatum* in vergleichsweise niedrigerer Kopienzahl oder stark divergierender Sequenz vorliegen. Eine solche Unterrepräsentation einiger, in nahen Verwandten hochamplifizierter Elemente konnte in verschiedenen Spezies der Gattung *Cucurbita* gezeigt werden, wo sich die entsprechenden Sequenzen zwar nicht in einer Southern-Hybridisierung nachweisen ließen, jedoch mit PCR aus dem Genom amplifizierbar waren (King et al., 1995). Choumane et al. (2000) vermuten eine sehr frühe Abspaltung der ebenfalls einjährigen Art *C. cuneatum* von einem gemeinsamen Vorläufer der restlichen Annualen. Ein von den übrigen einjährigen Arten getrennter Ursprung wird aufgrund von Isoenzym polymorphismen (Kazan & Muehlbauer, 1991; Tayyar & Waines, 1996; Labdi et al., 1996) und Kreuzbarkeitsdaten (Ladizinski & Adler, 1976a,b) angenommen. Morphologische Merkmale (van der Maesen, 1987) sowie die endemische Verbreitung dieser Art in Äthiopien, Ägypten und im Sudan bekräftigen diese Hypothese. Aufgrund dieser

Unterschiede wurde sogar eine Trennung dieser Spezies von der Sektion *Monocicer* vorgeschlagen (Tayyar & Waines, 1996), in der fast alle weiteren annuellen Arten zusammengefaßt werden. Die Abwesenheit oder die starke Divergenz der beiden Satellitenelemente *CaSat1* und *CaSat2* im Genom von *C. cuneatum*, die in der Kichererbse und den übrigen Annuellen eine dominante, heterochromatische Komponente des Genoms darstellen, unterstützt diese Annahme deutlich.

D.I.2.2. *CaSat1* und *CaSat2* unterstützen molekulare Daten in der Taxonomie

Verglichen mit den anderen annuellen Arten ist die Monomereinheit von *CaSat2* bei *C. chorassanicum* bemerkenswert konserviert. Offenbar muß man eine abweichende Evolution dieses Satelliten in *C. chorassanicum* annehmen, obwohl er wie bei den anderen annuellen Spezies aus der Sektion *Monocicer* ein prominentes genomisches Element darstellt. *C. chorassanicum* wurde als einzige einjährige Art aufgrund morphologischer Merkmale der Sektion *Chamaecicer* zugeordnet. Der längere Satellit *CaSat1* ist im Genom von *C. chorassanicum* mit ähnlich konservierten Monomereinheiten vorhanden wie in den übrigen Annuellen. Diese Tatsache unterstützt die Ergebnisse der Isoenzymanalysen (Kazan & Muehlbauer, 1991; Labdi et al., 1996, Tayyar & Waynes, 1996), die zu der Forderung führten, *C. chorassanicum* zu den übrigen Annuellen in die Sektion *Monocicer* zu integrieren (Kazan & Muehlbauer, 1991, Tayyar & Waynes, 1996). Allerdings wurden keine weiteren perennierenden *Chamaecicer*arten vergleichend untersucht, so daß eine generelle Konservierung von *CaSat1* in dieser Sektion nicht ausgeschlossen werden kann.

Als Vorläufer der drei miteinander kreuzbaren annuellen Spezies (*C. arietinum*, *C. reticulatum* und *C. echinospermum*) wird die perennierende Art *C. anatolicum* aufgrund von Isoenzymdaten vermutet (Kazan & Muehlbauer, 1991). Auch in dieser Spezies sind Sequenzen der *CaSat2*-Familie mit der gleichen hohen Abundanz wie in den annuellen Arten vorhanden. Auch die genomische Tandemorganisation und die Monomerlänge sind konserviert. Dies unterstreicht die enge Verwandtschaft von *C. anatolicum* mit den annuellen Arten, jedoch nicht mit der ersten Kreuzungsgruppe im speziellen, was die dominanten Bestandteile des Genoms anbetrifft. Für eine genauere Analyse des vermutlichen Vorläufers von *C. arietinum*, *C. reticulatum* und *C. echinospermum* müßten nicht nur weitere perennielle Arten in die Untersuchung miteinbezogen, sondern auch Sequenzdaten aus den drei Wildarten miteinander verglichen werden.

Insgesamt zeigt sich bei der Untersuchung der beiden Satelliten in den Wildarten der Gattung *Cicer*, daß die Analyse hochrepetitiver Bereiche im Kichererbsengenom phylogenetische Fragen und taxonomische Klassifizierungen auf molekularer Ebene unterstützen und ergänzen können. Abundanz und Organisation von *CaSat1* und *CaSat2* bestärken weitgehend die schon vorliegenden molekularen Daten (Isoenzym- und Speicherproteinanalysen). Auch die Konservierung von Primerbindestellen für STMS („sequence-tagged microsatellite sites“-) Marker im Genom der kultivierten Art und innerhalb der Wildarten ist hinsichtlich der Klassifizierung von *C. cuneatum* bzw. der engen Verwandtschaft von *C. anatolicum* zu den Annuellen mit den Daten der Satellitenanalysen konsistent (Choumane et al., 2000).

D.II. Das LTR-Retrotransposon *CaRep*

D.II.1. *CaRep* - ein hochabundantes Retrotransposon des *Ty3-gypsy*-Typs

Die Familie *CaRep* stellt eine dominante, dispers-repetitive Sequenzklasse im Genom der Kichererbse dar. Sequenzen dieses Typs nehmen große Teile des pericentrischen Heterochromatins ein und gehören zu einem Retrotransposon des *Ty3-gypsy*-Typs bzw. zu Relikten davon. *Ty3-gypsy*-Elemente, wie die *Ty1-copia*-Gruppe, sind als evolutiv ursprüngliche genomische Bestandteile ubiquitär im Pflanzenreich zu finden (Voytas et al., 1992; Suoniemi et al., 1998). Elemente mit Homologie zu den hier isolierten Fragmenten sind nur innerhalb der Gattung *Cicer* konserviert, liegen jedoch in variabler Abundanz und Organisation vor.

D.II.1.1. PCR-gestützte Isolation einer vollständigen *CaRep*-Kopie

Für die Identifikation von *CaRep* stand keine genomische Bank mit großen Fragmenten von *C. arietinum* zur Verfügung. Um eine wegen der vermuteten großen Anzahl deletierter *CaRep*-Kopien aufwendige Suche nach intakten Wiederholungseinheiten zu umgehen, wurde PCR zur Isolation der fehlenden Bereiche und der flankierenden LTRs verwendet. Ein ähnlicher Ansatz mit klonierten Fragmenten und PCR-Ergänzungen wurde auch zur Isolation von *RIRE3* verwendet (Kumewaka et al., 1999), und für die Sequenzierung der *BARE1*-LTRs wurde inverse PCR eingesetzt (Suoniemi et al., 1997). Da die Gesamtsequenz von *CaRep* aus mehreren im Genom voneinander getrennten Teilbereichen rekonstruiert wurde, repräsentiert sie keine tatsächlich existierende Kopie. Die Gesamtsequenz stammt vielmehr aus verschiedenen Kopien und unterliegt einer primerabhängigen Auswahl bestimmter Zielsequenzen. Daher läßt sich aus der rekonstruierten Kopie keine Information über ihre Funktionalität ableiten. Die verwendete Methode ermöglicht lediglich die Ermittlung von Gesamtlänge und Struktur für *CaRep* und eine Ableitung geeigneter Oligonukleotidprimer für die Generation von molekularen Markern.

Da nun jedoch alle codierenden Domänen sowie die LTRs als Sondensequenzen zur Verfügung stehen, wird eine seit kurzem zur Verfügung stehende BAC („bacterial artificial chromosome“) -Bank mit langen, genomischen Fragmenten die Isolation vollständiger Kopien ermöglichen.

Vergleicht man die aus der genomischen Bank von *C. arietinum* isolierten, hochrepetitiven Sequenzen (*CaRep1-3*) mit der rekonstruierten Gesamtkopie, so fällt eine Häufung klonierter Sequenzen aus dem Bereich des *gag*- und Protease-ORFs sowie dem PPT-Bereich 3' („downstream“) der Integrasesequenz auf. Für die Region zwischen RT (Reverse Transkriptase) und PPT konnten jedoch keine hochrepetitiven Klone aus der Bank isoliert werden. Dies wird vermutlich durch die *EcoRI*-Restriktion der klonierten Fragmente verursacht, da im genannten Teilbereich offensichtlich ein Mangel an (funktionellen)

Schnittstellen besteht. Große Fragmente sind weniger gut klonierbar und daher in der Bank unterrepräsentiert. Eine weitere Erklärung für die Häufung klonierter Bereiche liegt in der möglicherweise unterschiedlichen Konservierung verschiedenener Teilbereiche. Aus dem *Ty1-copia*-Element *BARE1* (*Hordeum vulgare*) ist der 5'-untranslatierte Bereich als eine schneller evolvierende Komponente bekannt und zeigt eine im Vergleich zu den codierenden Domänen und den LTRs unterschiedliche chromosomale Lokalisation (Suoniemi et al., 1996a). Ferner müssen nicht alle *CaRep*-Kopien hochabundant sein. Für *BARE1* wurden niedrigere Kopienzahlen des *gag*-Bereichs im Vergleich zu RT- und LTR-Sequenzen ermittelt (Manninen & Schulmann, 1993), *Zeon1* in Mais und *skippy* aus dem Genom von *Fusarium oxysporum* zeigen mehr Kopien der LTRs und *gag*-Leseraster als der codierenden *pol*-Domänen. Zumindest bei *Fusarium* ist dies auf solo-LTRs zurückzuführen, die durch Rekombination entstanden (Anaya & Roncero, 1995; Hu et al., 1995).

D.II.1.2. Struktur einer vollständigen Kopie von *CaRep*

Neben den Sequenzhomologien der Reversen Transkriptase unterstreicht auch die typische Struktur der rekonstruierten, kompletten Kopie von *CaRep*, daß es sich um ein LTR-Retrotransposon des *Ty3-gypsy*-Typs handelt.

Im 5'-Bereich der codierenden Sequenzen befindet sich das für die Verpackung der Retroelement-RNA in ein Capsid verantwortliche *gag*-Leseraster mit der konservierten RNA-Bindeprotein-Domäne. Diese Sequenz, die mit der Positionierung des tRNA-Motivs auf der Retrotransposon-RNA in Verbindung gebracht wird, befindet sich bei *CaRep* ähnlich wie bei *IFG7* (*Pinus*), *Ta1* (*A. thaliana*) oder *Ty3* (*S.cerevisiae*) etwa 88 Codons 5' der Proteasedomäne (Hansen et al., 1988; Voytas et al., 1990; Kossack & Kinlaw, 1999). Das folgende Polyproteinleseraster zeigt die für *Ty3-gypsy*-Elemente und Retroviren charakteristische Abfolge der Protease-, Reversen Transkriptase-, RNase H- und Integrasedomänen. Die Länge der *gag*- und *pol*-Region entspricht mit 1470 Codons etwa der in *dell* (*L. henryi*) oder *IFG* (*Pinus glauca*; Smyth et al., 1989, Kossack & Kinlaw, 1999), d.h. die Leseraster enthalten keine großen Deletionen, Insertionen oder zusätzlichen Leseraster, wie sie bei anderen *Ty3-gypsy*-Kopien gefunden wurden (Hu et al., 1995, Martinez-Izquierdo et al., 1997; Kumewaka et al., 1999).

Die *gag*- und *pol*-Genprodukte werden oft als ein Fusionsprodukt entweder von einem Leseraster translatiert und sind durch ein suprimierbares Stop-Codon voneinander getrennt, oder von zwei verschiedenen Leserastern gelesen, wobei die Regulation der Genprodukte durch einen Leserasterwechsel bei der Translation erfolgt (Kumar & Bennetzen, 1999). Die rekonstruierte Kopie von *CaRep* weist weder einen Leserasterwechsel zwischen RNA-Bindemotiv und der konservierten Proteasedomäne noch ein Stopcodon auf. Auch andere pflanzlichen LTR-Elemente wie *hopscotch* (*Z.mays*) oder *Tto1* und *Tnt1* (*N.tabacum*) weisen ein durchgehendes Leseraster auf (Grandbastien et al., 1989; White et al., 1994; Hirochika et al., 1996b), wobei die gesamte RNA möglicherweise durch einen Spleißmechanismus geteilt wird bzw. einen Teil des Fusionsproteins degradiert wird (Hirochika et al., 1996b). Das

durchgehende Leseraster von *CaRep* kann jedoch auch durch das Zusammenfügen verschiedener Sequenzen erklärt werden.

Charakteristische Elemente wie die beiden Primer-Bindestellen (PPT, PBS) mit dem für LTR-Retrotransposons typischen Abstand zu den LTRs (Chavanne et al., 1998) vor dem *gag*-Leseraster (PBS) bzw. vor dem rechten LTR (PPT) sind vorhanden. Die Homologie der PBS-Sequenz zum 3'-OH-Ende pflanzlicher Methionin-tRNAs (Sprinzl et al., 1989) bzw. die Verwendung von tRNA^{Meth} als „Primer“ für die (-)-Strang DNA-Synthese ist bei Pflanzen vorherrschend (Puruggannan et al., 1994; Kossack & Kinlaw, 1999; Kumewaka et al., 1999). Es gibt nur wenige Ausnahmen, z.B. *Zeon1* (*Z.mays*), bei dem tRNA^{Lys} verwendet wird, oder *cyclops* (*Pisum sativum*) und *Athila* (*A. thaliana*) mit tRNA^{Glu} (Hu et al., 1995; Pelissier et al., 1995; Chavane et al., 1998). Im Gegensatz zu tierischen Retroelementen, die diverse Zelltypspezifische tRNAs verwenden, wird dies mit der Vielzahl von pflanzlichen Zelltypen in Zusammenhang gebracht, die sich zu Keimzellen entwickeln können: pflanzliche Retroelemente können in einer Vielzahl von Zelltypen erfolgreich weitergegeben werden, so daß die ubiquitär vorhandene Initiator-tRNA^{Meth} einen „zuverlässigen“ Primer für die reverse Transkription darstellt (Kossack & Kinlaw, 1999).

Verglichen mit den meisten LTR-Retrotransposons sind die 3.3 Kbp umfassenden *CaRep*-LTRs relativ lang und verursachen die ebenfalls auffällig große Gesamtlänge von *CaRep* (11.4 Kbp). Einige *Ty3-gypsy*-Elemente weisen ebenfalls große Längen auf, wie etwa *cyclops* (*Pisum sativum*) mit 12.3 Kbp (Chavanne et al., 1998) oder *Athila* (*A. thaliana*) mit 10 Kbp (Pelissier et al., 1995), enthalten jedoch oft Insertionen fremder DNA (Martinez-Izquierdo et al., 1997). Da LTRs durch einen Templatwechsel der Reversen Transkriptase gebildet werden, ist ihre Länge nicht festgelegt (Varmus, 1983). Die LTRs von *RIRE3* mit 2316 Bp, *RIRE8A* mit 2948 Bp, *dell* (2406-2415 Bp) oder *Levil* aus *Sorghum* (4.6 Kbp) zählen zu den längsten bekannten LTR-Sequenzen (Smyth et al., 1989; Bennetzen, 1996; Kumewaka et al., 1999). Die meisten pflanzlichen LTR-Elemente besitzen 0.3-1.5 Kbp lange LTRs (Puruggannan et al., 1994; Bennetzen, 1996; Kumar & Bennetzen, 1999; Marin & Llorens, 2000). Bis auf einige Nukleotide der terminalen Regionen weisen die LTRs verschiedener Familien meist nur sehr schwache Homologien untereinander auf (SanMiguel et al., 1998; Kumewaka et al., 1999), so daß fehlende signifikante Homologien der *CaRep*-LTRs zu anderen Sequenzen der NCBI („national centre of biology and informatics“-)Datenbanken nicht ungewöhnlich sind. Die LTRs von *CaRep* werden von dem für Retroviren typischen und möglicherweise für die Reintegration in genomische DNA bedeutenden 4 Bp langen inversen Motiv 5'-TG...CA-3' flankiert (Temin, 1980; Kumar & Bennetzen, 1999). Meistens befinden sich eine TATA-Box oder Polyadenylierungssignale in der zentralen R-Region („repeated RNA region“) beider LTRs (Kumar & Bennetzen, 1999). In den sequenzierten *CaRep*-LTRs konnte ein TATA-Box-ähnliches Motiv detektiert werden, das durch den 30 Bp unterhalb gelegenen Transkriptionsstart eine ca. 960 Bp lange U3 („unique 3' RNA“-)Region abgrenzen würde. Das in Pflanzen konservierte Polyadenylierungssignal 5'-AATAAA- 3' (Mogen et al., 1990) fehlt an adäquater Position, jedoch befindet sich die aus BARE1 bekannte Sequenz 5'-GATAAA- 3' (Manninen et al., 1993) etwa 430 Bp hinter dem Beginn der R-Region. Beide Motive sind nicht in allen LTR-Sequenzen konserviert, so daß die

Zuordnung unsicher ist. Die resultierenden Größen der U3- (960 Bp), R- (430 Bp) und U5 („unique 5'RNA“-)Regionen (1970 Bp) übertrifft die anderer pflanzlicher LTRs bei weitem (Manninen et al., 1993; Kumewaka et al., 1999). Möglicherweise sind die Promotorsequenzen für RNA-Polymerase II in den sequenzierten LTRs degeneriert und weisen inaktive Kopien aus, da ein funktionsfähiges Element vollständige LTR-Abschnitte benötigt. Direkt nach der Insertion einer Kopie in genomische DNA weisen beide LTRs aufgrund des Integrationsmechanismus eine identische Sequenz auf, die jedoch durch Mutationen divergieren kann, bis dies zum Verlust der Mobilität führt. In Mais wurden aufgrund der Divergenz zwischen linkem und rechten LTRs Altersschätzungen der ineinander inserierten Kopien getroffen (SanMiguel et al., 1998). Bei den für *CaRep* identifizierten 5'-und 3'-LTRs bleibt jedoch unklar, ob sie zu derselben Kopie gehören bzw., ob es sich um neuinserierte Kopien handelt. Die 86-91 % Identität zwischen beiden Sequenzen zeigt demnach nur exemplarisch die generelle Divergenz zwischen LTRs derselben Familie. Variationen von bis zu 50% innerhalb einer Familie sind jedoch bekannt (SanMiguel et al., 1998).

D.II.1.3. Abundanz von *CaRep*-Kopien im Vergleich zu anderen LTR-Elementen

Die fast vollständige Bedeckung des Heterochromatins von Inter- und Metaphasechromosomen mit Signalen von *CaRep* und den beiden Satelliten zeigt, daß es sich hierbei um die Elemente mit der höchsten Kopienzahl im Kichererbsengenom handelt. Wie aus den Signalintensitäten abzuschätzen ist, stellt *CaRep*, verglichen mit den beiden anderen LTR-Retrotransposons *CaDis* und *CaTy*, die Familie mit den am häufigsten vorkommenden, dispers verteilten Elementen dar, die im Rahmen dieser Untersuchung analysiert wurde. Auch aus anderen Pflanzen wie *Picea glauca* (L'Homme et al., 2000), *Oryza sativa* (Kumewaka et al., 1999) oder *Lycopersicon esculentum* (Su & Brown, 1997) wird von mehreren LTR-Retrotransposonfamilien berichtet. Mehr als die Hälfte des ~2500 Mbp großen Genoms von Mais besteht aus verschiedenen LTR-Elementen (SanMiguel et al., 1996). Hierzu zählen die bekannten *Ty1-copia*-Elemente *BsI*, *Opie-1*, und *PREM2*, die sich in ihrer Kopienzahl um den Faktor 10^4 voneinander unterscheiden (Jin & Bennetzen, 1989; SanMiguel et al., 1996; Turcich et al., 1996), während zwei weitere *Ty3-gypsy*-ähnliche Familien (*Cinful-1*, *Grande1*) mit 20 000 und 1300 Kopien hochabundant sind (SanMiguel et al., 1996; Martinez-Izquierdo et al., 1997). Unterschiedliche Abundanz kann auf verschiedene Mechanismen zur Kopiezahlbegrenzung der einzelnen Familien hinweisen. Selbst für das kleine Genom von *A. thaliana* konnte eine Datenbanksuche zusätzlich zu den 3 bereits bekannten *Ty3-gypsy*-Elementen *Athila*, *Tat* und *Tma* sechs weitere Sequenzfamilien gleichen Typs ermitteln (Marin et al., 2000). Es ist folglich nicht auszuschließen, daß außer den hier genannten Familien weitere, bislang noch nicht detektierte LTR-Retrotransposons im Kichererbsengenom vorhanden sind.

D.II.1.4. Chromosomale Organisation von *CaRep*

Sowohl der für *Ty3-gypsy*-Elemente nicht ungewöhnliche hohe AT-Gehalt (Smyth et al., 1989; Pelissier et al., 1995) als auch die aus der Southern-Analyse anzunehmende disperse

Verteilung einerseits und teilweise konzentrierte Anhäufung („Cluster“) der *CaRep*-Sequenzen im Kichererbsengenom andererseits korrelieren gut mit ihrer chromosomalen Verteilung: der größte Teil der Retrotransposonsequenzen ist im pericentrischen Heterochromatin aller acht Chromosomen verteilt, einzelne Signale liegen in euchromatischen, vermutlich genreichen Bereichen.

Während *Ty1-copia*-Retrotransposons meist auf den Chromosomenarmen dispers angeordnet sind und heterochromatische Bereiche aussparen (Schmidt et al., 1996; Brandes et al., 1997b, Heslop-Harrison et al. 1997), findet man Vertreter der *Ty3-gypsy*-Gruppe oft in heterochromatischer Umgebung. Die aus Gerste klonierte Sequenz *Cereba* ist in vielen Gräsern, darunter Weizen und Roggen, an den Centromeren konserviert, was möglicherweise auf einen Zusammenhang mit der Centromerfunktion hinweist (Presting et al., 1998). Das *monkey*-Element aus Banane (*Musa* sp.) ist - von weiteren verstreuten Kopien abgesehen - größtenteils in der NOR konzentriert und wurde vermutlich durch Rekombinationsereignisse zwischen den rDNA-Gen-Wiederholungen mit diesen hochamplifiziert (Balint-Kurti et al., 2000). Im Gegensatz dazu sind die *CaRep*-Kopien deutlich von der engeren centromerischen Umgebung ausgespart, die vermutlich von *CaSat2* besetzt wird, und auch die NOR und Telomerbereiche sind frei von *CaRep*-Signalen. Ursache für diese Aussparungen können Amplifikations- und Homogenisationsmechanismen innerhalb einer Tandemabfolge repetitiver Sequenzen (z.B. durch Eliminierung von einzelnen Retrotransposonkopien) sein.

Trotz allem ist die Mehrzahl der *CaRep*-Kopien im pericentrischen Heterochromatin konzentriert. Dies wird z.B. auch von *Grande* und *Zeon1* berichtet, *Ty3-gypsy*-Elementen also, die gemeinsam mit weiteren Retrotransposons den 180 Bp-Satelliten im knopfförmigen Heterochromatin („knob heterochromatin“) in Mais unterbrechen (Ananiev et al., 1998a). Ebenso ist *Athila*, ein *Ty3-gypsy*-Element, im pericentrischen Heterochromatin von *A. thaliana* in den dominanten centromerischen Satelliten *pAL1* eingestreut (Pelissier et al., 1996). Für *CaRep* ist eine Vergesellschaftung mit *CaSat2*-Sequenzen im distaleren Bereich des Heterochromatinblocks nicht ausgeschlossen, kann jedoch mit der nur gering auflösenden Hybridisierung an Metaphasechromosomen nicht zweifelsfrei belegt werden. Eine „Cluster“-Bildung kann, wie bei *Athila* vermutet, durch gezieltes Inserieren in Tandemelemente (Pelissier et al., 1996), eine Coamplifikation mit einem Satelliten oder durch gegenseitige Insertion verschiedener Retrotransposonkopien zustande kommen. Ein Ineinanderspringen von Retrotransposons sowie resultierende verschachtelte „Cluster“ sind bisher von *Z. mays*, *A. thaliana* und *H. vulgare* bekannt (San Miguel et al., 1996; Martinez-Izquierdo et al., 1997; Suoniemi et al., 1997; Kumar & Bennetzen, 1999). Die Konzentration von Retroelementen im Heterochromatin ist besonders für kleinere Genome vorteilhaft, da damit die mutagene Wirkung auf codierende Bereiche reduziert wird. Neben der Kontrolle der Transpositionsaktivität durch Heterochromatinbildung oder Insertion der Elemente in bzw. nahe bereits existierender Heterochromatinblöcke inaktivieren sich die Elemente durch Insertionen auch gegenseitig (Kumar et al., 1997).

Einige *CaRep*-Sequenzen sind dagegen im genreichen Euchromatin lokalisiert. Eine enge Nachbarschaft zu Genen ist von vielen Pflanzen her bekannt, in denen Retroelemente oft erst durch ihre mutagene Wirkung isoliert wurden (Kumar & Bennetzen, 1999). Die

Sequenzierung großer genomischer Klone aus Mais hat jedoch gezeigt, daß die euchromatischen Retrotransposons meistens in intergenischen Regionen (intergenischen „Spacern“ oder anderen repetitiven Elementen) und nicht direkt in codierenden Bereichen liegen (SanMiguel et al., 1996).

Nicht alle Teilbereiche der *CaRep*-Kopien zeigen die gleiche chromosomale Lokalisation. So können z.B. euchromatische Loci nicht mit allen Sonden (*CaRep*1-3) nachgewiesen werden, und auch die durch Southern-Analyse generierten Bandenmuster waren je nach verwendeter Domäne des Elements unterschiedlich. Die zugrundeliegende, unterschiedlich starke Konservierung und Amplifikation einzelner *CaRep*-Teilbereiche im Genom wurde bereits erwähnt und konnte auch bei *in-situ*-Hybridisierungen mit codierenden Domänen und LTR-Sequenzen von *BARE1* (*Hordeum vulgare*) festgestellt werden (Suoniemi et al., 1996a).

D.II.1.5. Evolution von *CaRep* im Genus *Cicer*

Die Sequenzfamilie *CaRep* ist in allen untersuchten Spezies der Gattung *Cicer* vorhanden, variiert jedoch in Abundanz und genomischer Organisation zwischen einzelnen Gruppen von Spezies. Bestimmte Strukturen sind innerhalb der ersten Kreuzungsgruppe (*C. arietinum*, *C. reticulatum*, *C. echinospermum*; ICARDA, 1990; Ladizinski & Adler 1976a,b) konserviert, andere innerhalb der zweiten Kreuzungsgruppe (*C. bijugum*, *C. pinnatifidum*, *C. judaicum*). In den Genomen von *C. chorassanicum* und *C. yamashitae* sind *CaRep*-Kopien entweder sehr divergent oder nur in niedriger Kopienzahl vorhanden. *C. cuneatum* weist eine völlig andere genomische Organisation auf als die übrigen Spezies, enthält jedoch homologe Sequenzen zu *CaRep* - im Gegensatz zu den beiden Satelliten. Die Organisation der *CaRep*-Kopien weist auf mehrfache Rearrangements dieser Sequenzklasse im Laufe der Entwicklung der annuellen Spezies in den letzten 100 000 Jahren hin (Choumane et al., 2000), die möglicherweise mit Transposition und Amplifikation der Elemente einhergingen. Die Weitergabe der *CaRep*-Sequenzen erfolgte dabei wohl in vertikaler Richtung, da die Konservierung der genomischen Organisation die vermutlichen Verwandtschaftsverhältnisse zwischen den annuellen Arten widerspiegelt (Kazan & Muehlbauer, 1991).

Divergente, aber mit *CaRep* verwandte Sequenzen, die unter den stringenten Hybridisierungsbedingungen nicht mehr detektierbar sind, sind in anderen Leguminosen und Nichtleguminosen nicht ausgeschlossen. In den NCBI-Datenbanken konnten zwar verwandte Elemente in weiter entfernten Spezies (z.B. *Lilium henryi*, *Lycopersicon esculentum*, *Z.mays*, *O.sativa*) gefunden werden, die jedoch nur zu stark konservierten codierenden Sequenzen Homologien aufwiesen.

D.II.2. *CaRep*-Sequenzen als molekulare Marker

D.II.2.1. Auf Retroelementen basierende molekulare Marker

Da Retrotransposons oft einen Großteil des pflanzlichen Genoms ausmachen und dispers die Chromosomen verstreut sind, eignen sie sich als Werkzeuge für die pflanzliche Genomanalyse, besonders zur Generation molekularer Marker und zur funktionellen Analyse

von Genen (Kumar & Hirochika, 2001). Durch ihre replikative Transposition erzeugen Retroelemente Diversität, entweder direkt durch Insertionsmutationen von Genen oder indirekt, indem sie die Genomgröße und -organisation durch ihre Beweglichkeit variieren. Während Markertechnologien, die auf SNPs („single nucleotide polymorphisms“) oder SSRs („simple sequence repeats“) basieren, zufällige und relativ kurze Unterschiede in der Nukleotidsequenz detektieren, verursachen Retrotransposons Veränderungen von Hunderten von Basenpaaren bis zu mehreren Kilobasen, die im Gegensatz zu den SNPs und SSRs irreversibel sind. In vielen Pflanzen konnten Retrotransposon-dirigierte Marker bereits zur Erstellung und Ergänzung genetischer Karten verwendet werden oder erlaubten Diversitäts- und Phylogeniestudien durch die Feststellung der zeitlichen Abfolge von neuen Insertionen (Kumar & Hirochika, 2001). In Gerste konnte ein Resistenzgen gegen das Pilzpathogen *Pyrenophora teres* („net blotch“) durch Marker aus den Elementen *BARE1* und *Sukkula* auf der genetischen Karte lokalisiert werden (Manninen et al., 2000). Auch ein Resistenzgen für *Chalara elegans* („black root rot“) in *Nicotiana debneyi* und *N. tabacum* mit einem Marker assoziiert, der eine Retrotransposonsequenz (*Tnd1*) repräsentiert (Kenward et al., 1999).

D.II.2.2. Die S-SAP-Technik macht *CaRep*-Marker sichtbar

Mit der von AFLP („amplified fragment length polymorphism“) abgeleiteten S-SAP („sequence-specific amplified polymorphisms“-) Technik werden durch Kombination eines konventionellen AFLP- und eines elementspezifischen Oligonukleotidprimers individuelle Insertionen sichtbar gemacht (Waugh et al., 1997). Neben diesen Insertionspolymorphismen kann auch die Heterogenität innerhalb der Sequenzen zur Generation von S-SAP-Markern dienen, etwa mit Hilfe der RIVP-Methode („retrotransposon internal variation polymorphism“; Vershinin & Ellis, 1999). Oft ist die Betrachtung von mehreren Retroelementen mit unterschiedlicher Transpositionsvergangenheit informativ. So konnten z.B. verschiedene *Ty1-copia*-Elemente ergänzend in Diversitätsstudien der Gattung *Pisum* verwendet werden (Pearce et al., 2000).

Eine Modifikation der S-SAP-Methode wurde von Ellis et al. (1998) mit dem *Ty1-copia*-Element *PDR1* aus *Pisum sativum* entwickelt, die informative, auf Insertionspolymorphismen basierende Marker generierte, da hierbei das Element nur sehr kurze LTR-Sequenzen besitzt. Vergleicht man die bekannte Länge des LTRs von *CaRep* (ca. 3.3 Kbp) mit dem unter den gewählten Bedingungen amplifizierbaren Primerabstand (max. 1200-1300 Bp), so zeigt sich, daß die von *CaRep* generierten Polymorphismen nur auf interne Variabilität (d.h. Funktionalität und Position einer Restriktionsschnittstelle; s.Fig.C16b) des rechten LTRs zurückzuführen sind, sofern diese in intakten Kopien von *CaRep* lokalisiert sind. Daher kann die vorhandene Kopienzahl nicht aus der Anzahl der generierten Amplifikate abgeleitet werden, wie etwa bei der Erbse (Ellis et al., 1998), was die Differenz zwischen hohen Intensitäten bei Southern- und *in-situ*-Hybridisierung und der relativ niedrigen S-SAP-Bandenanzahl erklärt. Erstens bindet das aus drei klonierten *CaRep*-Sequenzen abgeleitete PPT-Oligonukleotid vermutlich nur an einen Teil der insgesamt vorhandenen *CaRep*-Kopien, weiterhin wird ein Teil der Wiederholungseinheiten aufgrund der gewählten PCR-

Bedingungen und des Trennvermögens des PAA-Gels möglicherweise nicht nachgewiesen, da die nächste benachbarte *EcoRI*-Schnittstelle zu weit entfernt liegt. Der alternative Einsatz von Primern aus dem LTR-Bereich könnten dagegen Insertionspolymorphismen sichtbar machen, diese binden aber aufgrund der variableren LTR-Sequenzen vermutlich ebenfalls nicht an alle Kopien.

D.II.2.3. Die Kartierung von *CaRep*-Markern

Von zwölf mit der PPT/*EcoRI*-Primerkombination erzeugten polymorphen Banden zwischen den verwandten Elternlinien konnten sechs zur Kartierung verwendet werden. Die restlichen Amplifikate waren in der verwendeten Kreuzungspopulation nicht polymorph. Unterschiede zwischen Eltern und Population sind bei dieser Kreuzung möglich, da die individuellen Elternpflanzen nicht vorliegen, und nur Pflanzen derselben Akzession verwendet werden. Diese können jedoch Unterschiede zu den eigentlichen Elternpflanzen aufweisen (Hüttel, 1996).

Die Ergebnisse der Kartierung sind nicht direkt mit der physikalischen Lokalisation der *CaRep*-Sequenzen auf Metaphasechromosomen von *C. arietinum* zu vergleichen, da die genetische Karte auf einer interspezifischen Kreuzung zwischen *C. arietinum* und *C. reticulatum* basiert. Diese Kreuzung ist für Segregationsstörungen bekannt, die zu einer Abweichung von der erwarteten 1:1-Verteilung parentaler Allele führt (Winter et al., 1999; Winter et al., 2000). Dies könnte z.B. auf eine Inkompatibilität bei der Paarung von Chromosomen aus *C. arietinum* und *C. reticulatum* hindeuten, die in der Meiose zu einer anormalen Verteilung der Chromosomen führt. Die Karyotypen beider Pflanzen sind relativ ähnlich (Galasso et al., 1996; Ahmad, 2000). Trotzdem gibt es Unterschiede, die auf chromosomale Rearrangements zurückzuführen sind, wie etwa den Wegfall eines Satellitenanhängsels und damit verbunden den Verlust einer NOR auf Chromosom B bei *C. arietinum* (Galasso et al., 1996). Obwohl Hybride aus *C. arietinum* und *C. reticulatum* normale Chromosomenpaarungen und Fertilität zeigen (Ladizinski & Adler, 1976; Singh & Ocampo, 1993), sind Translokationen oder Deletionen, die das Hybridgenom vom Kichererbsengenom unterscheiden, in bestimmten Bereichen nicht auszuschließen. Darüberhinaus könnten bei der auf PCR beruhenden S-SAP-Markertechnik möglicherweise *CaRep*-Loci detektiert werden, die bei der *in-situ*-Hybridisierung mit einer relativ langen Sonde nicht zugänglich sind. Auf Differenzen bei der Detektion von Mikrosatelliten mit Hybridisierungs- bzw. PCR-Techniken wurde bereits hingewiesen (Gortner et al., 1998). Aus den genannten Gründen ist es daher nicht möglich, die Verteilung der *CaRep*-Marker über mehrere Kopplungsgruppen direkt mit ihrer dispersen Verteilung bzw. mit der Konzentration eines Großteils der Kopien im Heterochromatin in Beziehung zu setzen.

Der *CaRep*-Marker REPT2 ist auf der genetischen Karte relativ eng bei einem Marker lokalisiert, der ein Resistenzgenalog (RGA) darstellt (Hüttel, pers. Mitteilung). Diese Nachbarschaft kann jedoch nur unter Vorbehalt auf physikalischen Abstand übertragen werden, da chromosomale Bereiche mit repetitiven Elementen oft mit einer niedrigen Rekombinationsrate korrelieren (Martienssen, 1998; The *Arabidopsis* Genome Initiative,

2000) und in einer auf Rekombination basierenden genetischen Karte sehr enge Abstände aufweisen.

Die Verwendung von *CaRep* in einem S-SAP-Ansatz erweist sich als eine Methode zur schnellen Produktion molekularer Marker für die Kichererbse. Die S-SAP-Technik generiert wie auch AFLP (Winter et al., 2000) eine große Zahl polymorpher Produkte in der interspezifischen Kreuzung *C. arietinum* x *C. reticulatum*. Eine weitere Detektionsmethode für Retrotransposon-Marker ist RBIP („retrotransposon-based insertion polymorphism“), ein codominantes Markersystem, bei dem einzelne Insertionen durch Oligonukleotidprimer gegen interne und flankierende Sequenzen durch PCR sichtbar gemacht werden (Flavell et al., 1998). Neben einfachen RFLPs („restriction fragment length polymorphism“) können Restriktionsschnittstellenvariationen interner Sequenzen durch PCR-Amplifikation als sog. „Transposon-Signaturen“ sichtbar gemacht werden (Puruggannan & Wessler, 1995; Shcherban et al., 2000). Andere PCR-Techniken detektieren Polymorphismen zwischen invers angeordneten LTRs zweier benachbarter Elemente (IRAP: „inverse retrotransposon amplified polymorphism“) oder zwischen Amplikons von LTRs und nahegelegenen Mikrosatelliten (REMAP; „retrotransposon microsatellite amplified polymorphism“; Kalendar et al., 1999; Provan et al., 1999), was durch die Entwicklung von schnellen Isolierungstechniken für LTR-Sequenzen zunehmend erleichtert wird (Pearce et al., 1999). Aufgrund der vermutlichen Vermischung der *CaRep*-Sequenzen mit Mikrosatellitenmotiven (Gortner et al., 1998) und weiteren Retrotransposonfamilien (*CaLin*, *CaTy* und *CaDis*), die ebenfalls in den distalen Bereichen des pericentrischen Heterochromatins und den angrenzenden euchromatischen Bereichen von *C. arietinum*-Chromosomen lokalisiert sind, bieten sich z.B. REMAP und IRAP in Kombination mit den benachbarten Elementen als ergänzende Markertechniken für weitere Kartierungen an.

D.III. Die mittel- und niedrigrepetitiven Retroelemente *CaTy*, *CaDis* und *CaLin*

D.III.1. Die LTR-Retrotransposonfamilien *CaTy* und *CaDis*

D.III.1.1. *Ty1-copia*-Sequenzen im Genom von *C. arietinum*

Die *CaTy*-Sequenzen aus Kichererbse repräsentieren eine mittelrepetitive Retrotransposonfamilie der *Ty1-copia*-Untergruppe bzw. deren defekte Fragmente, von denen

eine RT-Region erhalten geblieben ist. Die sehr heterogenen Vertreter dieser Familie sind dispers oder in dispersen „Clustern“ im Genom verstreut und hauptsächlich in den distalen Bereichen des pericentrischen Heterochromatins aller acht Chromosomen lokalisiert, einige Kopien liegen aber auch im angrenzenden Euchromatin. Die Sequenzfamilie ist in allen annuellen Spezies der Gattung *Cicer* in vielen Kopien vorhanden. Darüberhinaus konnten homologe Sequenzen auch in anderen Leguminosen detektiert werden.

Die ubiquitäre Verbreitung von Retrotransposons der *Ty1-copia*-Untergruppe in Pflanzengenomen ist vielfach beschrieben (Flavell et al., 1992a; Flavell et al. 1992b; Voytas et al., 1992; Kumar, 1996, Kumar & Bennetzen, 1999). *Ty1-copia*-Elemente sind aus Leguminosengattungen wie *Vicia*, *Glycine* und *Phaseolus* (Pearce et al., 1996a; Bhattacharyya et al., 1997; Garbner et al., 1999) und auch aus *C. arietinum* bekannt (Brandes et al., 1997b, Sant et al., 2000). Allerdings untersuchten Brandes et al. (1997b) nur die chromosomale Lokalisation von *Ty1-copia*-ähnlichen Sequenzen. In der vorliegende Arbeit wurde darüberhinaus erstmalig die genomische Organisation von *Ty1-copia*-Sequenzen im Kichererbsengenom und nahen Verwandten beschrieben sowie eine vergleichende Sequenzanalyse durchgeführt. Es ist unwahrscheinlich, daß es sich hier um die bereits von Brandes et al. (1997b) analysierten Sequenzen handelt, da sich trotz Verwendung identischer Primersequenzen in Abhängigkeit von den PCR-Bedingungen in den ersten Amplifikationszyklen unterschiedliche Oligonukleotidbindungen ergeben. Die *in-situ*-Hybridisierung von *CaTy* zeigt daher im Gegensatz zur Arbeit von Brandes und Mitarbeitern zusätzliche euchromatische Loci und die Aussparung des zentralen Teils der pericentrischen Heterochromatinblöcke, was zusätzliche Informationen liefert.

Sant et al. (2000) beschreiben zwei weitere RT-Sequenzen aus *C. arietinum* (C1, C10: AF264002, AF264003), deren wahrscheinlichste Leseraster viele Stop-Codons aufweisen und nur 29-39% Homologie zu den hier beschriebenen Fragmenten zeigen. Die Autoren verwendeten Primer, deren Sequenz von den in der vorliegenden Arbeit benutzten Oligonukleotiden abweicht, so daß die Amplifikate C1 und C10 zwar äquivalente Abschnitte aus der RT-Region pflanzlicher Retroelemente darstellen, aber aufgrund der fehlenden konservierten Aminosäuren der TyA-Bindestelle nicht eindeutig der *Ty1-copia*-Untergruppe zugeordnet werden können. Lediglich die Sequenzhomologie läßt eine Verwandtschaft zu dieser Großfamilie annehmen. Die Kopienzahlbestimmung und Verwendung der Sequenzen zur Verwandtschaftsanalyse innerhalb der Gattung *Cicer* basieren demnach nicht zwingend auf *Ty1-copia*-Elementen. Auch bei dem in dieser Arbeit beschriebenen 690 Bp langen PCR-Artefakt läßt die potentielle Aminosäuresequenz NVFLNG (statt TAFLHG) an der 5'-Primerposition die Zugehörigkeit zur *Ty1-copia*-Familie nur vermuten (vgl. Xiong & Eickbush, 1990). Beides macht deutlich, daß außer den beschriebenen Familien *CaRep*, *CaDis* und *CaTy* weitere LTR-Retrotransposonfamilien bzw. Relikte davon im Genom von *C. arietinum* vorhanden sind, und daß sie - quantitativ gesehen - nicht unbedeutende Bestandteile des Genoms darstellen.

D.III.1.2. *CaTy*-Sequenzen aus Kichererbse sind sehr heterogen

Ty1-copia-Elemente sind als Populationen mit sehr heterogener Sequenz von den Genomen vieler höherer Pflanzen her bekannt. Verglichen damit sind *Ty1-copia*-Sequenzen aus Tieren und niederen Eukaryonten wesentlich homologer (Mount & Rubin, 1985; Kim et al., 1998). Im Kartoffelgenom zeigen die verschiedeneklontierten RT-ORFs untereinander Ähnlichkeiten von 5-75% (Flavell et al., 1992b), vergleichbare Sequenzen aus verschiedenen anderen Mono- und Dikotyledonen sind sehr heterogen (Flavell et al., 1992a; Asins et al., 1999; Gribbon et al., 1999; Matsuoka & Tsunewaki, 1999). Die Heterogenität unter den *Ty1-copia*-ähnlichen Sequenzen im Kichererbsengenom läßt sich teilweise innerhalb der *CaTy*-Kopien nachweisen. Diese Sequenzvariabilität wird durch die von der RT bedingten relativ hohen Mutationsrate während eines Transpositionszyklus verursacht (etwa 2.5×10^{-5} bei *Ty1* in *S. cerevisiae*; Gabriel et al., 1996). Bei der durch Pathogenbefall aktivierten Familie *Tnt1* (*N. tabacum*) scheint die Variabilität der an der Expressionsregulation beteiligten U3-Region eine adaptive Kapazität der Elemente zu erhalten, was der Anpassungsfähigkeit von Retroviren ähnelt (Cascuberta et al., 1997). Pflanzliche Genome erlauben eine Ansammlung defekter Kopien über lange Zeiträume hinweg, und bestimmte pflanzliche Meristeme können zu generativen Zellen ausdifferenzieren. D.h. im Gegensatz zu Tiergenomen kann eine Vielzahl von Zelltypen aktive und mutierte Kopien vertikal weitergeben (Kumar & Bennetzen, 1999). Dies trifft auf die übrigen Retrotransposongruppen wie LINEs, SINEs oder *Ty3-gypsy*-Elemente ebenfalls zu, ist jedoch bei diesen nicht detailliert untersucht worden (Kumar & Bennetzen, 1999).

D.III.1.3. *CaDis*-Sequenzen weisen eine höhere Homogenität auf

Die *CaDis*-Familie aus *C. arietinum* ist, wie auch *CaTy*, eine mittelrepetitive und dispers im Genom organisierte Sequenzfamilie, die auf mehreren Chromosomen in den distalen Bereichen der heterochromatischen Blöcke und im angrenzenden Euchromatin lokalisiert ist. Aufgrund von Sequenzhomologien kann davon ausgegangen werden, daß es sich auch bei dieser Familie um LTR-Retrotransposons bzw. Relikte davon handelt. Allerdings wurde im Gegensatz zu *CaTy* und *CaLin* eine Region aus dem *gag*-Leseraster amplifiziert. *gag* liegt 5' des *pol*-ORFs und ist für die Verpackung der RNA in ein virusähnliches Partikel („virus-like particle“; VLP) verantwortlich. Da die Reihenfolge der einzelnen *pol*-Domänen nicht erkennbar ist, läßt sich nicht entscheiden, ob *CaDis* zur *Ty1-copia*- oder der *Ty3-gypsy*-Untergruppe gehört. Das hier amplifizierte Motiv liegt nicht im Bereich konservierter *gag*-Domänen wie der „Cys-His-Box“ (Hu et al., 1995), so daß die relative Homogenität der Sequenzen mit einer weniger starken Konservierung der Oligonukleotidbindestellen erklärbar ist.

D.III.1.4. Chromosomale Lokalisation von *CaTy* und *CaDis*

Die mittlere Abundanz und disperse genomische Organisation von *CaTy*-Sequenzen sowie ihre Methylierung bzw. ihr AT-Reichtum korreliert gut mit ihrer Lokalisation im pericentrischen Heterochromatin. Der zentrale Teil der Heterochromatinblöcke, vermutlich

von langen Tandemwiederholungen von *CaSat2*-Sequenzen besetzt, wird von *CaTy* und *CaDis* größtenteils ausgespart. Wie bei *CaTy* erstrecken sich auch bei *CaDis* die in den distalen Teilen der pericentrischen Heterochromatinblöcke liegenden Signale bis in euchromatische Regionen und sind in den Bereichen mit hochabundanten Satelliten oder in der NOR abwesend. Eine Durchmischung von *CaSat2* mit den beschriebenen Retrotransposons (*CaRep*, *CaDis*, *CaLin*) in den distalen Bereichen der Heterochromatinblöcke sowie auf den Satellitenanhängseln (*CaTy*) und dem centromerischen Bereich von Chromosom A und B ist zu erwarten. In *Arabidopsis* sowie bei Mais zeigte die Sequenzierung der centromerischen Bereiche, daß Retroelementsequenzen ineinander verschachtelt sind (San Miguel et al., 1996; Kumar & Bennetzen, 1999), was ohne höher auflösende *in-situ*-Hybridisierungen oder Sequenzierung langer genomischer Fragmente für *C. arietinum* allerdings nicht geklärt werden kann.

In den meisten untersuchten Pflanzengenomen sind *Ty1-copia*-Sequenzen mehr oder weniger gleichmäßig im Euchromatin entlang der Chromosomenarme verteilt, wobei Centromere, rDNA-Loci oder interstitielle bzw. terminale heterochromatische Regionen ausgespart werden oder verringerte Signalintensitäten zeigen (Brandes et al., 1997b; Heslop-Harrison et al., 1997; Hanson et al., 1999; Linares et al., 1999). *Allium cepa* bildet mit einer starken Konzentration von *Ty1-copia*-Sequenzen im terminalen Heterochromatin eine Ausnahme (Pearce et al., 1996b), ebenso *A. thaliana*, in dessen Genom diese Sequenzen im pericentrischen Heterochromatin konzentriert und mit dem 180 Bp-Satelliten *pAL* assoziiert sind (Martinez-Zapater et al., 1986; Brandes et al., 1997a). Brandes et al. (1997b) zeigten zuerst eine solche pericentrisch heterochromatische Konzentration auch für *Ty1-copia*-Sequenzen in *C. arietinum* und postulierten deren Amplifikation durch replikative Mechanismen in Verbindung mit einem damals unbekanntem pericentrischen Tandemelement. Die höhere Auflösung der *in-situ*-Hybridisierung in der vorliegenden Arbeit sowie die vorliegenden Kenntnisse über die hochrepetitiven Sequenzen im pericentrischen Heterochromatin (Staginnus et al., 1999) zeigen, daß eine gemeinsame Amplifikation mit einem Teil der *CaSat2*-Sequenzen in den distaleren Bereichen des Heterochromatins nicht ausgeschlossen werden kann. Allerdings muß man angesichts der deutlichen Aussparung von *CaTy*-Sequenzen im Zentrum der Blöcke auch andere Mechanismen diskutieren: die schon erwähnte Insertionspräferenz der Retroelemente kann z.B. eine Akkumulation disperser Elemente in *CaSat2*-besetzten Heterochromatinbereichen erklären (SanMiguel et al., 1996; Tikhonov et al., 1999; Vicient et al., 1999). Eine hypothetische, starke Amplifikation von *CaSat2* könnte die Retrotransposonkopien nachfolgend distal abgedrängt haben. Die Abwesenheit von *CaTy* in zentralen *CaSat2*-Bereichen ist ferner durch das Fehlen einer für die Integrase zugänglichen Insertionsstelle bedingt (durch Sequenz- oder DNA-Konformation) oder durch schnelle Homogenisierungsmechanismen in tandemorganisierten Bereichen zu erklären. Die Aussparung von Retroelementen aus evolutionär jungen centromerischen Sequenzen im Gegensatz zur Anwesenheit dieser Elemente in älteren, interkalaren Tandemsequenzen ist von der Gattung *Beta* her bekannt (Schmidt et al., 1994; Schmidt et al., 1995). Denkbar ist, daß großflächige Insertionen in euchromatische Regionen

zu letalen Mutationen führen oder Deletionen von Retroelementen, wie sie bei BARE1 vermutet werden, verhindert werden (Heslop-Harrison et al., 1997; Vicent et al., 1999).

Obwohl *C. arietinum* ein größeres Genom besitzt als *Arabidopsis*, könnte doch die untypische Konzentration von *CaTy*- und *CaDis*-Sequenzen die Gefahr ihrer Insertion in codierende Sequenzen verringern und selektiv vorteilhaft gewesen sein. Parallel dazu kann auch die Konkurrenz mehrerer Retroelemente im Wirtsgenom Einfluß auf das Ineinanderspringen bzw. die gegenseitige Inaktivierung und damit die Konzentration von Retroelementen haben (Kumar et al., 1997).

D.III.1.5. Vermutliche Evolution der beiden LTR-Retrotransposons *CaDis* und *CaTy*

Die relative Ursprünglichkeit der *CaDis*- und *CaTy*-Sequenzen (sowie von *CaLin*) wird durch deren Abundanz über die Gattungs- bzw. Familiengrenzen hinaus unterstrichen, die in der Southern-Analyse und den Homologievergleichen erkennbar ist: Sequenzen mit hoher Homologie zu *CaTy*-Elementen in anderen Leguminosen wie *Phaseolus*, *Vicia*, *Pisum* sowie auch *Vicia faba* und *Vigna unguiculata* weisen wie schon bei *CaDis* darauf hin, daß diese Familien evolutiv älter sind als die *Cicer*-spezifischen Elemente *CaSat1*, *CaSat2* oder *CaRep* und schon früh in der Familie der Leguminosen vorhanden waren.

Die genomische Organisation der *CaDis*-homologen Sequenzen innerhalb der verwandten annuellen Spezies ist - abgesehen von leichten spezies- und kreuzungsgruppenspezifischen Variationen - relativ konserviert. Diese Invariabilität kann einerseits durch eine geringe Homologie im *gag*-Bereich verursacht werden, andererseits aber darauf zurückzuführen sein, daß *CaDis*-Sequenzen bereits in einem Vorgänger der annuellen Wildarten und *C. anatolicum* vorhanden waren und seitdem keine Mobilität gezeigt haben, die zu einer Veränderung von Kopienzahl oder der genomischen Organisation geführt hätte. Lediglich *C. cuneatum* weist wie bei den übrigen Retroelementen eine abweichende genomische Organisation der Sequenzen auf und unterstreicht damit seine phylogenetische Distanz zu den übrigen Annuellen.

CaTy-Elemente, die in allen annuellen *Cicer*-Spezies (außer *C. cuneatum*) sowie in der perenniellen Art *C. anatolicum* detektiert wurden, sind ebenfalls schon in einem Vorgänger der annuellen und perenniellen Arten repetitiv vorhanden. Allerdings gibt es in der genomischen Organisation und Abundanz Unterschiede zwischen einzelnen Spezies oder Speziesgruppen, was im Gegensatz zu *CaDis* auf Phasen der Aktivität während der Evolution des Genus *Cicer* hindeutet. Da die perennielle Spezies *C. anatolicum* eine deutlich andere genomische Organisation als die Annuellen zeigt, ist eine bedeutende Umorganisation von *CaTy*-Kopien in einem Vorgänger der Annuellen anzunehmen (Kazan & Muehlbauer, 1991). Das Fehlen von entsprechenden Signalen bei *C. cuneatum* muß nicht die Abwesenheit von *CaTy*-Elementen im Genom bedeuten, sondern kann auf stark divergierende Sequenzen oder deutlich niedrigere Kopienzahl zurückgehen. Dies unterstützt Vermutungen, nach denen *C. cuneatum* einen separaten Ursprung des einjährigen Lebenszyklus besitzt oder sehr früh von den übrigen annuellen Spezies abgespalten wurde, was auch durch Isoenzymanalysen, Kreuzbarkeitsdaten und Konservierung von Mikrosatelliten-flankierenden Sequenzen

(Ladizinsky & Adler, 1976a,b; Tayyar & Waines, 1996; Choumane et al., 2000) vermutet und auch durch die Organisation der Satelliten-DNA widergespiegelt wird.

Da die genomische Organisation von *CaTy*-Sequenzen zwischen enger verwandten Spezies ähnlicher ist (z.B. innerhalb der Kreuzungsgruppen), kann man wie bei den meisten pflanzlichen Retrotransposons von einer bevorzugt vertikalen Transmission dieser Elemente ausgehen (Flavell et al., 1992a; Matsuoka et al., 1999). Horizontaler Transfer von LTR-Retrotransposons wurde bisher nur von einem *Ty1-copia*-Element aus der Gattung *Drosophila* berichtet, wo vermutlich Milben als Vektoren dienten (Flavell, 1999; Jordan et al., 1999). Für Pflanzen dagegen fehlen Hinweise auf eine horizontale Übertragung, obwohl sie durch interspezifische Kreuzungen oder Parasiten mit breitem Wirtsspektrum möglich wäre und der Ursprung einiger pflanzlicher Retroelemente auf Retroviren aus Insekten zurückgeführt wird (Kumar & Bennetzen, 1999).

D.III.1.6. Gibt die Transkription von *CaDis*-Sequenzen einen Hinweis auf die Aktivität des Elements?

Einige Loci der *CaDis*-Sequenzfamilie liegen in euchromatischen Regionen verschiedener Chromosomen und korrespondieren möglicherweise mit den sequenzierten cDNA-Klonen, die - unter Vorbehalt der kleinen Stichprobenzahl - einen geringeren AT-Gehalt aufweisen als die genomischen Klone. Ein großer Teil der Sequenzen enthält ununterbrochene ORFs. Somit ist eine Transkription dieser Sequenzen nicht ausgeschlossen. Die Transkription *CaDis*-verwandter Sequenzen in anderen Spezies ist bekannt: zwei homologe *gag*-Bereiche von Retrotransposons aus *Oryza sativa* konnten in EST-Datenbanken gefunden werden (C72786, C26347). Die eng verwandte Sequenz RPE15 aus *Medicago sativa* ist mittelrepetitiv und wird in späten Entwicklungsstadien der Blütenknospe transkribiert (Xia et al., 1996). Da jedoch nur ein Teilbereich des *gag*-Leserasters von *CaDis* isoliert wurde, kann keine Aussage über die gesamte Sequenz der *gag*- und *pol*-Gene gemacht werden. Mehr als die codierenden Sequenzen, die auch von anderen autonomen Kopien bereitgestellt werden können, sind die regulatorischen Sequenzen im LTR-Bereich für die Funktionsfähigkeit eines Elements ausschlaggebend. Diese sind ebenso wie ihr Methylierungsstatus für *CaDis* derzeit unbekannt, so daß keine weiterreichende Aussage über seine Funktionalität möglich ist.

Die Transkription des *gag*-Bereichs von *Zeon*, einem LTR-Retrotransposon aus Mais, konnte in geringem Maß im Endosperm nachgewiesen werden (Hu et al., 1995). Obwohl nur nichtautonome Kopien isoliert werden konnten, beweisen doch DNA-Rearrangements, daß das Element aktiv war. Bei *CaDis* läßt dagegen die einheitliche genomische Organisation eine Mobilität in der jüngeren phylogenetischen Entwicklung der Gattung *Cicer* unwahrscheinlich erscheinen. Die Transkription eines Retrotransposons ist allerdings nicht immer gleichbedeutend mit seiner Transposition oder Kopienzahlexpansion. So konnten für das in Blättern von *Hordeum vulgare* stark transkribierte BARE1 auch VLPs („virus-like particles“) isoliert werden. Jedoch vermutet man posttranslationale Mechanismen, die eine Aufblähung des Gerstengenoms durch unkontrollierte Insertion verhindern (Suoniemi et al.,

1996a; Jääskeläinen et al., 1999). Eine Reduktion der Kopien nach Integration in das Genom durch homologe Rekombination der LTRs konnte für das gleiche Element gezeigt werden (Vicent et al., 1999). In *A. thaliana* konnte „anti-sense“-RNA zu einem Transkript des *Ty1-copia*-Elements *AtRel* isoliert werden, die vermutlich an dessen post-transkriptionaler Kontrolle beteiligt ist (Kuwahara et al., 2000). Für *Ty1* aus *S. cerevisiae* konnten Wirtsgene isoliert werden, die eine Anreicherung von *Ty1*-Proteinen verhinderten, die Degradierung von cDNA beschleunigten und so die Proliferation des Elements regulierten (Curcio & Garfinkel, 1999). Obwohl solche Mechanismen für Pflanzen noch nicht bekannt sind, ist eine posttranskriptionelle Regulation von *CaDis*-Transkripten nicht ausgeschlossen.

CaDis-Transkripte kommen unabhängig von einer Vorbehandlung in allen untersuchten Geweben der Kichererbse vor, was ungewöhnlich ist. Die Transkription von Retrotransposons wird meistens nur in einzelnen Geweben und Entwicklungszuständen bzw. in bestimmten Streßzuständen beobachtet. So wird z.B. die homologe Sequenz aus Luzerne nur gewebespezifisch exprimiert (Xia et al., 1996; Grandbastien, 1998). Die euchromatische Lokalisation deutet darüberhinaus an, daß sich aktiv transkribierte Gene in unmittelbarer Nachbarschaft zu *CaDis*-Kopien befinden können, deren Polymerase bis in den Bereich des sequenzierten *gag*-Leserasters durchliert. In diesem Fall wären die schwachen Signale der RT-PCR als „Artefakte“ der Polymerase zu erklären. Aus *Nicotiana glauca* (Royo et al., 1996) ist das *Ty3-gypsy*-ähnliche Element *Tna1* bekannt, das nach Bestäubung oder Berührung der Griffel transkribiert wird. Da das Transkript jedoch keine vollständige Kopie des Elements repräsentiert, wird vermutet, daß die *Tna1*-Sequenz in ein exprimiertes Gen inseriert ist. Solche chimären Transkripte sind auch aus Mais und Tabak bekannt (Pouteau et al., 1991; Turcich & Mascarenhas, 1994) und können für die Transkripte von *CaDis* nicht ausgeschlossen werden.

Letztendlich lassen die bisherigen Kenntnisse über die Struktur und Transkription von *CaDis* nicht erkennen, ob die Transkription lediglich ein Durchlesen aus benachbarten genischen Bereichen darstellt oder ob *CaDis* selbst transkribiert und postranskriptionell reguliert wird. Eine aktive Transposition von *CaDis*-Kopien erscheint eher unwahrscheinlich. Weitere Kenntnisse über die Struktur einer vollständigen *CaDis*-Kopie, Nachweise der Translation von *CaDis*-Transkripten (z.B. Nachweis eines Capsidproteins) oder die Untersuchung einer intraspezifischen Kreuzung auf Insertion neuer *CaDis*-Kopien könnten die Problematik der Funktionsfähigkeit bzw. Mobilität von *CaDis* möglicherweise erhellen.

D.III.2. Die non-LTR-Retrotransposonfamilie *CaLin*

Bisher konnten mit Hilfe degenerierter Oligonukleotidprimer LINE-Sequenzen aus etlichen Pflanzengenomen isoliert werden (Wright et al., 1996; Kubis et al., 1998; Noma et al., 1999). Die hier identifizierten Sequenzen bilden die erste für *C. arietinum* beschriebene LINE-ähnliche Elementfamilie und wurden unter dem Namen *CaLin* (*Cicer arietinum* LINE-ähnliche Elemente) zusammengefaßt. Für *CaLin* wurde bisher aufgrund der PCR-gestützten Isolation nur ein kleiner Bereich der RT sequenziert, so daß keine Aussagen über Vollständigkeit oder Länge der Kopien gemacht werden kann. Die dispers organisierte

Familie kann aber aufgrund starker Sequenzheterogenität in drei Subfamilien (*CaLin1-3*) unterteilt werden und gehört gemeinsam mit *CaTy* und *CaDis* - verglichen mit den hochrepetitiven *CaRep*-Elementen - zu den weniger abundanten Retroelementfamilien der Kichererbse. Die *CaLin*-Sequenzen sind mindestens auf drei der acht Chromosomen von *C. arietinum* lokalisiert. Verwandte Sequenzen sind in ähnlicher Abundanz und größtenteils mit der gleichen genomischen Organisation in allen anderen annuellen Spezies der Gattung *Cicer* zu finden. Darüberhinaus ist die Sequenzfamilie nicht auf die Gattung *Cicer* beschränkt, sondern auch in anderen Leguminosen und sogar einigen Nichtleguminosen nachweisbar.

D.III.2.1. *CaLin*-Sequenzen weisen die für LINEs typische Sequenzheterogenität auf

Auffällig ist die starke Sequenzheterogenität zwischen den einzelnen *CaLin*-Kopien im Kichererbsengenom (45-99% Homologie). Ähnlich divergente LINE-Sequenzen wurden auch von anderen Pflanzen berichtet. So weist eine Homologie von 43-96% zwischen den RT-Bereichen von verschiedenen LINEs aus *Beta vulgaris* auf mehrere Subfamilien hin (Schmidt et al., 1995; Kubis et al., 1998). Auch in *A. thaliana* liegt die Ähnlichkeit von RT-Abschnitten verschiedener LINE-Elemente nur zwischen 48-88% (Wright et al., 1996). Im Bereich der Endonuklease (EN) herrscht eine vergleichbare Heterogenität, so daß innerhalb der entsprechenden Sequenzen aus *Oryza sativa* nur 38-98% der Nukleotide konserviert sind (Noma et al., 1999). Eine solch starke Sequenzheterogenität ist wie bei allen Retrotransposons in pflanzlichen Genomen mit der erwähnten fehlerbehafteten Arbeitsweise der Reversen Transkriptase zu erklären, die keine 5'-3'-Exonukleasefunktion besitzt, aber für jede replikative Transposition verantwortlich ist. Die dadurch erhöhte Mutationsrate macht sich bei LINEs um so stärker bemerkbar, da sie evolutiv sehr alte Elemente im Pflanzengenom darstellen und divergente, funktionslose Kopien sich dort über lange Zeiträume akkumulieren konnten (Eickbush, 1994; Kumar & Bennetzen, 1999; Schmidt, 1999). Häufige Leserasterwechsel und Stopcodons im codierenden Bereich der RT von *CaLin*-Elementen sind ebenfalls kein Einzelfall, sondern werden auch aus RTs und ENs von LINEs in *O.sativa*, *B. vulgaris* und *Cannabis sativa* berichtet (Kubis et al., 1998; Noma et al., 1999; Sakamoto et al., 2000) und unterstreichen die Sequenzheterogenität. Die entsprechenden Kopien stellen Pseudogene dar, d.h. degenerierte Kopien, die nicht mehr autonom beweglich sind. Eine „trans“-Aktivierung solcher Sequenzen ist nicht ausgeschlossen, jedoch sind die benötigten regulatorischen Elemente im 5'-untranslatierten Bereich für *CaLin* unbekannt.

Die hohe Anzahl bekannter Pseudogene aus dem EN- und RT-Bereich wie die auch die größere Verwandtschaft der analysierten pflanzlichen LINE-Fragmente zwischen taxonomisch enger verwandten Genomen spricht für eine hauptsächlich vertikale Weitergabe von LINEs (Wright et al., 1996; Kubis et al., 1998; Noma et al., 1999). Die engere Verwandtschaft einiger *CaLin*-Sequenzen zu fremden LINEs ist daher eher ein Ausdruck der starken Heterogenität zwischen den Subfamilien als ein Hinweis auf horizontalen Transfer zwischen verschiedenen taxonomischen Gruppen, obwohl letzteres nicht ausgeschlossen werden kann.

D.III.2.2. Die niedrige Abundanz von *CaLin* wird vermutlich durch seine Heterogenität bedingt

Die Signalintensitäten von *in-situ*- und Southern-Hybridisierungen weisen übereinstimmend auf eine niedrige Abundanz der *CaLin*-Sequenzen im Kichererbsengenom hin. Dies kann zumindest zum Teil auf die starke Heterogenität dieser Sequenzen zurückgeführt werden, d.h. trotz Verwendung eines Gemischs aus verschiedenen *CaLin*-Fragmenten weist nur ein Teil der Kopien ausreichende Homologie zur dieser Sonde auf und wird detektiert. Generell sind LINE-Elemente in pflanzlichen Genomen nur in mittlerer Kopienzahl zu finden. So weist z.B. *Cin4* aus *Z.mays* <100 Kopien auf (Schwarz-Sommer et al., 1987), in der gleichen Größenordnung liegt *Zepp* (*Chlorella vulgaris*), und auch in *Cannabis sativa* ist die LINE-ähnliche Familie mit 300-400 Kopien nur mittelrepetitiv (Higashiyama et al., 1997, Sakamoto et al., 2000). *Ta11* (*A. thaliana*) und *LINE-CS* (*O.sativa*) weisen dagegen niedrigere Kopienzahlen auf (Wright et al., 1996; Noma et al., 1999). Eine Ausnahme bildet die *del2*-Familie aus *Lilium speciosum*, die mit 250 000 Kopien 4 % des gesamten Genoms ausmacht (Leeton & Smith, 1993).

D.III.2.3. *CaLin*-Sequenzen liegen teilweise in euchromatischen Bereichen

CaLin-Sequenzen liegen größtenteils an der Peripherie der pericentrischen Heterochromatinblöcke von mindestens drei Chromosomen. Möglicherweise treten LINE-Sequenzen in bestimmten Bereichen gehäuft (als „Cluster“) auf, wie etwa die LINEs in zwei *Beta*-Spezies (Schmidt et al., 1995; Kubis et al., 1998). Wie in *Beta* sind auch im Kichererbsengenom die Bereiche der hochrepetitiven Tandemelemente und der 18S-5.8S-25S rRNA-Genabfolgen von LINE-Signalen ausgespart, was möglicherweise auf einen Verlust der Retroelemente durch schnelle Amplifikations- und Homogenisierungsmechanismen innerhalb der Tandemwiederholungseinheiten zurückzuführen ist (Schmidt, 1999). In *Hordeum spontaneum* sind im Gegensatz dazu LINE-Anhäufungen direkt in den sekundären Einschnürungen zweier Chromosomen lokalisiert (Belyayev et al., 2001). *Zepp* aus *Chlorella vulgaris* ist im Telomerbereich angehäuft (Higashiyama et al., 1997), und auch die beiden non-LTR-Elemente *TRASI* und *SART1* (*Bombyx mori*) sind mit Telomersequenzen assoziiert (Okazaki et al., 1995; Takahashi et al., 1997).

Der größte Teil einer LINE-Familie aus *Cannabis sativa* ist in langen Tandemabfolgen auf dem heterochromatischen, langen Arm des Y-Chromosoms konzentriert (Sakamoto et al., 2000), und auch die meisten LINE-ähnlichen *Ta11*-Elemente (*A. thaliana*) sind heterochromatisch in pericentrischer Umgebung lokalisiert (The *Arabidopsis* Genome Initiative, 2000). Die Kartierung von LINE-tragenden YAC („yeast artificial chromosome“)-Klonen in *O.sativa* weist ebenfalls auf die Lokalisation eines Teils der Sequenzen im centromerischen Heterochromatin hin (Noma et al., 1999). Obwohl größtenteils im Heterochromatin lokalisiert, sind einige *CaLin*-Signale im euchromatischen Bereich der Kichererbsenchromosomen zu finden, so daß ein Teil dieser Sequenzen in möglicherweise

genhaltigen Regionen liegt, wofür auch die nicht besonders hohe Methylierung einiger *CaLin*-Sequenzen spricht. Einige Elemente aus Tiergenomen zeigen Präferenzen für bestimmte Zielsequenzen, die von der EN erkannt und geschnitten werden (Feng et al., 1998), was zu einer Insertion in genische Bereiche führen kann. SINEs aus *Brassica napus*, die vermutlich die Endonukleasefunktion eines LINEs nutzen, bevorzugen Zielregionen, die Kernmatrixproteine binden können und codierende Sequenzen flankieren (Tikhonov et al., 2001). Durch Sequenzierung größerer DNA-Abschnitte konnten auch bei Mais LINE-Elemente in Introns von Genen nachgewiesen werden (Thikonov et al., 1999).

D.III.2.4. *CaLin* ist vermutlich eine sehr ursprüngliche Retrotransposonfamilie in Leguminosen und zeigt innerhalb der Gattung *Cicer* keine Veränderung seiner Organisation

Die genomische Organisation und Abundanz *CaLin*-ähnlicher Sequenzen sind in allen annuellen *Cicer*-Spezies nahezu einheitlich. Offensichtlich ist diese Sequenzfamilie bereits in einem Vorgänger aller Annuellen vorhanden gewesen. In Anbetracht der Heterogenität der Sequenzen kann dies als Hinweis darauf gewertet werden, daß *CaLin*-Sequenzen evolutiv sehr alte Elemente darstellen und möglicherweise die ursprünglichsten aller untersuchten repetitiven Familien in Kichererbse sind. Die evolutionäre Ursprünglichkeit der *CaLin*-ähnlichen Sequenzen wird durch ihre Anwesenheit in allen untersuchten Leguminosen und weiteren Dicotyledonen unterstützt.

Die Konservierung von Abundanz und Organisation zeigt, daß seit der Trennung der verschiedenen annuellen Spezies keine massiven Rearrangements bzw. Amplifikationen dieser Elemente mehr stattgefunden haben. Dies wird durch die starke Degeneration der Leseraster unterstützt. In der Gattung *Beta* konnten dagegen verschiedene Kultivare durch eine Hybridisierung mit *BvLi3* unterschieden werden, und in nahverwandten *Beta*-Spezies konnten Punktmutationen nachgewiesen werden, die eine Variabilität dieser LINE-Familie während der Domestikation und Züchtung von *Beta* zeigen, obwohl Transpositionen nicht direkt beobachtet werden konnten (Schmidt, 1999).

Obwohl die Hybridisierungssignale in der Gattung *Cicer* und anderen Leguminosen von gleicher Intensität waren, hatte die PCR- Amplifikation nicht in allen Spezies zu einheitlichen Ergebnissen geführt. Die Problematik der Amplifizierbarkeit von LINEs ist auch in anderen Pflanzengenomen beschrieben worden (Noma et al., 1999; Schmidt, 1999) und wird auf die für LINEs typische Heterogenität der Sequenzen zurückgeführt. Möglicherweise sind die hier beschriebenen schwachen Amplifikationsprodukte auf viele divergierende bzw. nur wenige wirklich homologe Primerbindestellen zurückzuführen, während die internen Bereiche dem Hybridisierungssondengemisch konserviertere Angriffspunkte bieten.

D.IV. Nachweis und Charakterisierung des DNA-Transposons *CaEn/Spm*

D.IV.1. Ein PCR-Ansatz zum Nachweis und zur Isolation von *En/Spm*-ähnlichen Transposonsequenzen aus Pflanzengenomen

Bisher basierten Entdeckung und molekulare Analyse von Transposonsequenzen größtenteils auf phänotypischen Mutanten. In jüngster Zeit eröffnet die wachsende Zahl der Genomsequenzierungsprojekte neue Einblicke in die Vielfalt und Verteilung von Transposonsequenzen innerhalb eines Genoms. Bei einer solchen Auswertung der Sequenzdaten, auch „transposon mining“ genannt, werden potentielle Transposonsequenzen über Homologiesuchen mit bekannten Transposonsequenzen aus der Fülle der Sequenzdaten isoliert und identifiziert (Le et al., 2000).

Die in der vorgelegten Arbeit entwickelte Amplifikation von Transposonsequenzen mit Hilfe degenerierter Oligonukleotide, die aus der Transposaseregion pflanzlicher Transposons abgeleitet wurden, stellt eine Strategie zum schnellen und einfachen Nachweis von *En/Spm*-ähnlichen Elementen in pflanzlichen Genomen dar. Sie kann unabhängig von phänotypischen Mutanten angewendet werden und erfordert keine vorherige Kenntnis von Sequenzen des betreffenden Genoms. Die einfache und schnelle Durchführung dieser PCR-Methode erlaubt darüberhinaus die parallele Untersuchung einer großer Anzahl von Genomen.

Degenerierte Oligonukleotidprimer wurden bereits zur Amplifikation verschiedener Retrotransposonsequenzen entwickelt und verwendet, so zur Identifikation von LTR-Elementen des *Ty1-copia*- (Flavell et al., 1992b) und *Ty3-gypsy*-Typs (Suoniemi et al., 1998) sowie von non-LTR-LINEs (Wright et al., 1996; Kubis et al., 1998; Noma et al., 1999). Darüberhinaus ist ein einfacher Nachweis von transposablen Elementen der Klasse II zur Untersuchung dieser repetitiven Sequenzfamilie an sich, vor allem aber wegen der vielen Verwendungsmöglichkeiten eines Transposons als Werkzeug in der Genomanalyse wünschenswert. Nachdem seine Existenz nachgewiesen worden ist, kann das amplifizierte Transposasefragment als Ausgangspunkt für die weitere Analyse seiner genomischen Organisation, seiner Abundanz und möglicherweise auch seiner Aktivität dienen und zur Isolation des gesamten Elements aus dem jeweiligen Genom verwendet werden.

Die Verwendung eines solchen degenerierten Oligonukleotidpaars war - bis auf Kochbanane (*Musa balbisiana* x *Musa acuminata*) - in allen untersuchten Pflanzengenomen, darunter wichtigen Kulturpflanzen wie Zuckerrübe, Weizen und Erbse erfolgreich. Die Detektion von *En/Spm*-Elementen in 20 Pflanzengattungen unterstreicht ihr fast ubiquitäres Vorkommen in Pflanzen (Nacken et al., 1991). Außerdem erlaubt dieses Primerpaar den schnellen Nachweis eines *En/Spm*-Elements in jedem beliebigen Pflanzengenom und - wie für Kichererbse gezeigt - dessen weitere Charakterisierung und Isolation. Die Ähnlichkeit zwischen den sequenzierten TPase-Fragmenten aus verschiedenen Pflanzenarten ist teilweise gering, so daß man davon ausgehen kann, daß das abgeleitete Oligonukleotidpaar in der Lage ist, auch sehr divergente Mitglieder der *En/Spm*-Familie zu detektieren. Die in dieser Arbeit aus *Pisum sativum* klonierte Sequenz zeigte ebenfalls keine signifikante Homologie zu *Pis1*, dem für Erbse bereits beschriebenen Transposon dieser Kategorie (Shirsat, 1988), und gehört damit zu einer von *Pis1* separaten *En/Spm*-ähnlichen Sequenzfamilie. In *Glycine max* wurde in der

vorliegenden Arbeit nur ein schwaches Amplifikationsprodukt beobachtet, obwohl in dieser Art bereits eine *En/Spm*-Familie, *Tgm*, beschrieben wurde (Rhodes & Vodkin, 1998).

D.IV.2. Eine Transposonfamilie der Kichererbse: *CaEn/Spm*

D.IV.2.1. Struktur und Abundanz *En/Spm*-ähnlicher Elemente im Genom von *C. arietinum*

Der oben skizzierte PCR-Ansatz ermöglichte den Nachweis *En/Spm*-ähnlicher Elemente im Genom von *Cicer arietinum* und nah verwandter Arten der Gattung *Cicer*. Die Charakterisierung von Sequenz, genomischer Organisation und chromosomaler Lokalisation dieser Elemente stellte das Potential der Methode unter Beweis. Diese Sequenzen sind die ersten in der Kichererbse identifizierten Transposons der Klasse II und wurden *Cicer arietinum En/Spm*-ähnliche Elemente genannt (*CaEn/Spm*). Southern-Analyse und *in-situ*-Hybridisierung belegen übereinstimmend, daß diese Sequenzen im Kichererbsengenom eine niedrige bis höchstens mittlere Abundanz aufweisen und dispers im Kichererbsengenom organisiert sind.

Die vergleichsweise hohe Divergenz unter den *En/Spm*-ähnlichen Elementen deutet auf die Anwesenheit mehrerer Subfamilien hin (*CaEn/Spm1-3*). Auch in anderen Pflanzen wurden - wie in *C. arietinum* - ganze Familien verwandter, *En/Spm*-ähnlicher Sequenzen vorgefunden (Rhodes & Vodkin, 1988; Ozeki et al., 1997). Bei *Tpn1* („Japanese Morning Glory“/ *Ipomoea nil*) weist eine große Zahl der Kopien Deletionen im internen Bereich auf (Inagaki et al., 1994), was zur Heterogenität einer Transposonfamilie beitragen kann. Die *Ac/Ds*-ähnliche Familie *Tam3* aus *Antirrhinum majus*, die sonst ungewöhnlich hoch konservierte interne Strukturen und eine große Zahl aktiver Kopien aufweist, umfaßt nicht-funktionelle TPase-Regionen mit nur 88% Homologie zu einer aktiven Kopie (Kishima et al., 1999). Da nur ein Teil des codierenden Transposasebereichs der *CaEn/Spm*-Elemente bekannt ist, kann die Isolation nicht-autonomer Kopien nicht ausgeschlossen werden, wozu etwa die deletierte Sequenz *Ca6* aus Kichererbse zählen könnte. Die vollständige Länge der *CaEn/Spm*-Elemente wurde bisher nicht ermittelt.

Anzahl und Intensität der Southern- und *in-situ*-Hybridisierungssignale zeigen trotz Verwendung unterschiedlicher Sonden übereinstimmend, daß die Kopienzahl der *CaEn/Spm*-Elemente signifikant niedriger ist als die der meisten anderen repetitiven Elemente im Kichererbsengenom. Ähnlich wie bei den anderen pflanzlichen *En/Spm*-Elementen (z.B. *Rim2*, *Cs1*, *Psl* und *Tgm*) muß von einer niedrig repetitiven Abundanz ausgegangen werden. Lediglich *Pis1* (Shirsat, 1988) aus der Erbse weist mehr als 100 Kopien pro haploidem Genom auf. Die mindestens vorhandenen sieben sequenzierten Kichererbsenkopien sind über sechs der acht Chromosomen verteilt. Die niedrige Abundanz wird möglicherweise durch den nichtreplikativen Transpositionsmechanismus von Klasse II-Transposons in Pflanzen bedingt, bei dem die Sequenzen nicht bei jeder Transposition amplifiziert werden.

D.IV.2.2. Gibt es Hinweise auf die Funktionalität von *CaEn/Spm*-Kopien?

D.IV.2.2.1. Aktive Transposons sind wertvolle Werkzeuge in der Genomanalyse

Aktive Transposons können für verschiedene Techniken der Genomanalyse verwendet werden. Beim sog. „transposon tagging“ (Fedoroff et al., 1984; Sundaresan, 1996; Kunze et al., 1997) und der „saturation mutagenesis“ (Walbot, 2000) dient die Insertionsmutagenese zur Identifikation flankierender Sequenzen. Auch die biologische Rolle bekannter Gensequenzen aus EST-Projekten kann durch „knock-out“-Mutanten (reverse Genetik) ermittelt werden (Maes et al., 1999; Tissier et al., 1999; Walbot, 2000). Die ligationsvermittelte PCR-Technik des „transposon insertion display“ (TID) kann selektiv für Insertionen in transkribierte Gene verwendet werden (ETID; „expressed transposon insertion display“; Yphremov & Saedler, 2000).

D.IV.2.2.2. Transposontranskripte fehlen in *C. arietinum*

In vielen Pflanzen wurden Transkripte der zur Transposition notwendigen codierenden Sequenzen durch Northern-Analysen oder RT-PCR nachgewiesen (Pereira et al., 1986; Nacken et al., 1991; Ozeki et al., 1997; Snowden & Napoli, 1998). Für *Tdc1*, *Cs1* und *Psl* konnte eine Transposition auch direkt nachgewiesen werden (Ozeki et al., 1997; Snowden & Napoli, 1998; Chopra et al., 1999). Ähnlich wie bei Retroelementen (Hirochika, 1993; Pouteau et al., 1994; Hirochika et al., 1996a; Mhiri et al., 1997, Takeda et al., 1998) gibt es auch bei *En/Spm*-Transposons durch Pathogenbefall induzierte (He et al., 2000) oder entwicklungsabhängige Aktivität (Levy & Walbot, 1990) sowie Elemente, die in Zellkulturen zur Transposition angeregt werden (Ozeki et al., 1997). Mit Hilfe von RT-PCR wurde gezielt nach Transkripten verschiedener Transposons gesucht (Snowden & Napoli, 1998; Henk et al., 1999). Ähnliche Untersuchungen konnten bei Kichererbse sowie *C. reticulatum* jedoch keine Transkripte der Transposase-kodierenden Genfragmente der *CaEn/Spm*-Familie nachweisen. Durch interagierende autoregulatorische und/oder epigenetische Mechanismen werden Expression und Transposition von *En/Spm*-Elementen relativ unabhängig von ihrer chromosomalen Lokalisation reguliert (Fedoroff, 1999). Aufgrund der Cytosinmethylierung regulatorischer Sequenzen kann die Transposase eines Elements inaktiviert werden (Fedoroff et al., 1995). Das Fehlen von Transkripten muß daher nicht zwingend die Abwesenheit funktionsfähiger *CaEn/Spm*-Kopien bedeuten, möglicherweise kann die sehr niedrige bzw. nicht vorhandene Transkriptionsrate durch besondere (hier nicht simulierte) Einflüsse oder in bestimmten Entwicklungszuständen induziert bzw. verstärkt werden.

D.IV.2.2.3. Selbst intakte Transposaseleseraster erlauben keine Rückschlüsse auf die Funktionalität

Da direkte Beweise für die Transpositionsfähigkeit von *CaEn/Spm*-Transposons bisher fehlen, ist der Nachweis autonomer und funktionsfähiger Kopien ein vordringliches Ziel, um zu prüfen, ob das Element sich für transpositionsabhängige Methoden der Genomanalyse eignet. Dabei ist nicht nur die Vollständigkeit der beiden Leseraster (Autonomie), sondern auch die terminale und subterminale Struktur der Kichererbselemente sowie ihr

Methylierungszustand ausschlaggebend, da diese cis-regulatorische Elemente darstellen und für die Funktionalität unabdingbar sind.

Codierende Bereiche für die Transposase sind nur von wenigen *En/Spm*-Elementen vollständig charakterisiert. Soweit überhaupt untersucht, überlappen die codierenden Bereiche der beiden zur Transposition notwendigen Proteine. Die Exons des ersten Gens sind dabei über die ganze Länge des Elements verstreut, während das zweite Gen in einem Intron des ersten lokalisiert ist (Pereira et al., 1986; Nacken et al. 1991; Snowden & Napoli, 1998). Bei den meisten Elementen ist nur das zweite Leseraster konserviert, dessen Ablesung zu einem längeren, aber weniger abundanten Transkript als das des ersten führt: *Tam1* und *En/Spm* enthalten ORF für die entsprechenden Proteine *Tnp2* (752 Aminosäuren) bzw. *TnpD* (950 Aminosäuren), das Transposon aus Karotte (*Tdc1*) zeigt ein teilweise konserviertes, kürzeres Leseraster von 520 Aminosäuren für die Transposase. *Psl* und *Tgm* enthalten ORFs mit signifikanten Homologien zu diesen Leserastern. In einigen Fällen konnten nur Leseraster mit Deletionen identifiziert werden oder es wurden gar keine homologen Sequenzen gefunden (Chopra et al., 1999; He et al., 2000). Die Identifikation der *CaEn/Spm*-Elemente durch PCR eröffnet den Zugang zu einem 187-191 Aminosäuren langen Ausschnitt aus dem zu *TnpD* bzw. *Tnp2* äquivalenten zweiten Leseraster, der vergleichsweise hohe Homologie zu anderen Transposons wie *Tam1*, *En-1* und *Psl* aufweist und in vielen Kopien vollständig ist. Der limitierte Ausschnitt aus der codierenden Region läßt allerdings noch keine Aussage über deren Länge und Erhaltung bei *Ca/EnSpm*-Elementen zu. Ebenfalls unklar bleibt die Struktur der terminalen invertierten Sequenzmotive (TIRs) und der subterminalen Regionen mit cis-regulatorischen Elementen für die Transposition.

Eine Übersicht über Gesamtlänge und -struktur von *CaEn/Spm*-Elementen und damit ihrer Funktionalität ist nur durch die Isolation einer vollständigen Kopie dieser Familie zu ermitteln, die durch eine kürzlich erstellte BAC („bacterial artificial chromosome“) -Bank mit großen Genomfragmenten der Kichererbse möglich wird.

D.IV.2.3. Verbreitung von *CaEn/Spm*-homologen Sequenzen in verwandten Genomen

Ähnlich wie die beschriebenen Retrotransposons von *Cicer arietinum* sind die mit *CaEn/Spm*-verwandten Sequenzen nicht auf die Gattung *Cicer* beschränkt. Die vorliegenden Sequenzen der nah verwandten Art *C. reticulatum* z.B. fügen sich gut in zwei der drei *CaEn/Spm*-Subfamilien ein. Homologe Sequenzen in verwandten Spezies sind auch für eine Anzahl anderer pflanzlicher *En/Spm*-Elemente belegt (Snowden & Napoli, 1998; Chopra et al., 1999), wohingegen *Tnr3* in vielen Reisvarietäten, aber nicht außerhalb der Speziesvorkommt (Motohashi et al., 1996). Offenbar sind die *CaEn/Spm*-Elemente bereits in einem hypothetischen Vorgänger der Kichererbse und nahe verwandten Leguminosen vorhanden. Innerhalb der Gattung *Cicer* ist eine unterschiedliche genomische Organisation der *CaEn/Spm*-ähnlichen Sequenzen zu erkennen, was auf eine Transpositionsaktivität nach der Abspaltung der verschiedenen annuellen Spezies von einem gemeinsamen Vorgängergenom hindeutet. Allerdings ist für Transposons der Klasse II auch horizontaler Transfer bekannt, etwa für *mariner*- oder *P-Elemente* in der Gattung *Drosophila* und

verwandter Dipteren (Robertson & Lampe, 1995; Clark et al., 1998). Zumindest für die *P-Elemente* ist bekannt, daß Milben der Art *Proctolaelaps regalis* als Vektoren eines solchen horizontalen Transfers zwischen den *Drosophila*-Arten dienen (Houck et al., 1991). Auch kann die horizontale Weitergabe der *En/Spm*-Elemente durch semiparasitische Insekten innerhalb der Leguminosen nicht ausgeschlossen werden.

D.IV.2.4. Chromosomale Lokalisation der *CaEn/Spm*-Elemente

Eine ganze Reihe Transposons wurde über phänotypische Auffälligkeiten isoliert. Dabei sind diese Transposons stets in codierende oder regulatorische Bereiche von Genen inseriert (Bonas et al., 1984; Inagaki et al., 1994; Ozeki et al., 1997; Chopra et al., 1999). Sequenzierungsprojekte zeigen aber, daß Transposons der *En/Spm*-Gruppe überwiegend in nicht-codierenden, repetitiven Bereichen zu finden sind (Le et al., 2000; The *Arabidopsis* Genome Initiative, 2000). Die bisher untersuchten pflanzlichen Elemente liegen zumindest teilweise konzentriert mit anderen repetitiven Elementen vor, etwa mit *Tourist*- oder MITE-Elementen („miniature inverted-repeat transposable elements“; Chopra et al., 1999) oder in SINEs (Motohashi et al., 1996). Allerdings gibt es auch Beispiele, bei denen die Insertion unabhängig von der Abundanz der Zielsequenzen ist (Nacken et al., 1991). Die detaillierteste Studie über die genomische Verteilung von *En/Spm*-Elementen wurde bisher im Rahmen der Genomsequenzierung von *A. thaliana* durchgeführt (The *Arabidopsis* Genome Initiative, 2000). In *A. thaliana* machen transposable Elemente insgesamt 14% des Genoms aus, und im Rahmen der Sequenzierung konnten insgesamt 2203 mobile Elemente der Klasse II gefunden werden. Dabei sind Transposons vom CACTA (*En/Spm*-) Typ besonders in den heterochromatischen Bereichen um die Centromere anzutreffen. Sie sind - relativ zu den Retroelementen - eher in der Peripherie lokalisiert, während erstere dichter am Centromer sitzen (The *Arabidopsis* Genome Initiative, 2000). Le et al. (2000) untersuchten 17.2 Mbp eines zumeist euchromatischen Bereichs des *Arabidopsis*-Genoms nach Transposonsequenzen, wozu sie Homologievergleiche durchführten („transposon mining“). Sie fanden darunter drei CACTA-Elemente. Die Transposasesequenzen aus dem Genom von *Arabidopsis thaliana*, die die höchsten Homologien zu *CaEn/Spm* aufweisen (AC006437, AC006586, AF128394, AC005957, AC005897 sowie TAIR („The Arabidopsis Information Resource“) Codes MOD1.14, F6H5_116, F1504_30, F11I2_2, T17A2_5), sind über vier der fünf Chromosomen verstreut und auch innerhalb eines Chromosoms nicht unmittelbar benachbart, sondern mindestens 0.5 Mbp voneinander getrennt (*The Arabidopsis Information Resource*- <http://www.arabidopsis.org>). In der Nachbarschaft homologer Bereiche in *Arabidopsis* konnten weitere Transposasegene oder Pseudogene detektiert werden (z.B. in AC006437, AC006586, AC005897). Auch liegen andere codierende Sequenzen und Teile von Retrotransposons, insbesondere des pericentrischen Elements *Athila* (Pelissier et al., 1995) in unmittelbarer Nähe der *TNP2*-ähnlichen Sequenzen von *Arabidopsis*. Die genannten Sequenzen sind nur maximal 1.2-2.1 Mbp von den centromerischen Sequenzen entfernt und gehören vermutlich zu den genannten pericentrischen, repetitiven Sequenzen.

Verglichen mit Retrotransposons ist über die chromosomale Lokalisation von DNA-Transposons mit Hilfe von FISH (Fluoreszenz in-situ-Hybridisierung) relativ wenig bekannt. Im Pufferfisch (*Tetraodon fluviatilis*) konnten *mariner*-ähnliche Elemente bevorzugt im rDNA-assoziierten Heterochromatin zweier Chromosomenpaare gefunden werden (Mandrioli et al., 2000). Das *Hoppel*-(ähnliche) Transposon in *Drosophila melanogaster* stellt eine wichtige Komponente des pericentrischen Heterochromatins in Mitosechromosomen dar, wo es vorwiegend mit Tandemelementen co-lokalisiert. Allerdings konnten zusätzlich auch euchromatische Loci ausgemacht werden (Coelho et al., 1998). Die meisten durch FISH darstellbaren *CaEn/Spm*-Kopien sind sowohl von der NOR als auch von den Anhäufungen der hochrepetitiven Sequenzen ausgenommen und liegen in der Peripherie der pericentrischen Heterochromatinblöcke, soweit dies die relativ niedrige Auflösung auf Metaphasechromosomen erkennen läßt. Dies entspricht etwa der durch Sequenzierung nachgewiesenen Lokalisation im *Arabidopsis*-Genom (The *Arabidopsis* Genome Initiative, 2000). Einzelne Transposonsignale, die möglicherweise Anhäufungen aus multiplen Kopien repräsentieren, konnten an der Grenze zum pericentrischen Heterochromatin in euchromatischen Regionen detektiert werden. Dies widerspricht nicht einer möglichen transkriptionellen Aktivität von einigen wenigen Kopien. Soweit bekannt ist, berichtet die vorliegende Arbeit zum ersten Mal über die chromosomale Lokalisation von pflanzlichen DNA-Transposons mit Hilfe von FISH (Staginnus et al., 2001).

D.V. Repetitive Elemente im Genom von *C. arietinum* - Grundlage für ein vorläufiges Chromosomenmodell von *C. arietinum*

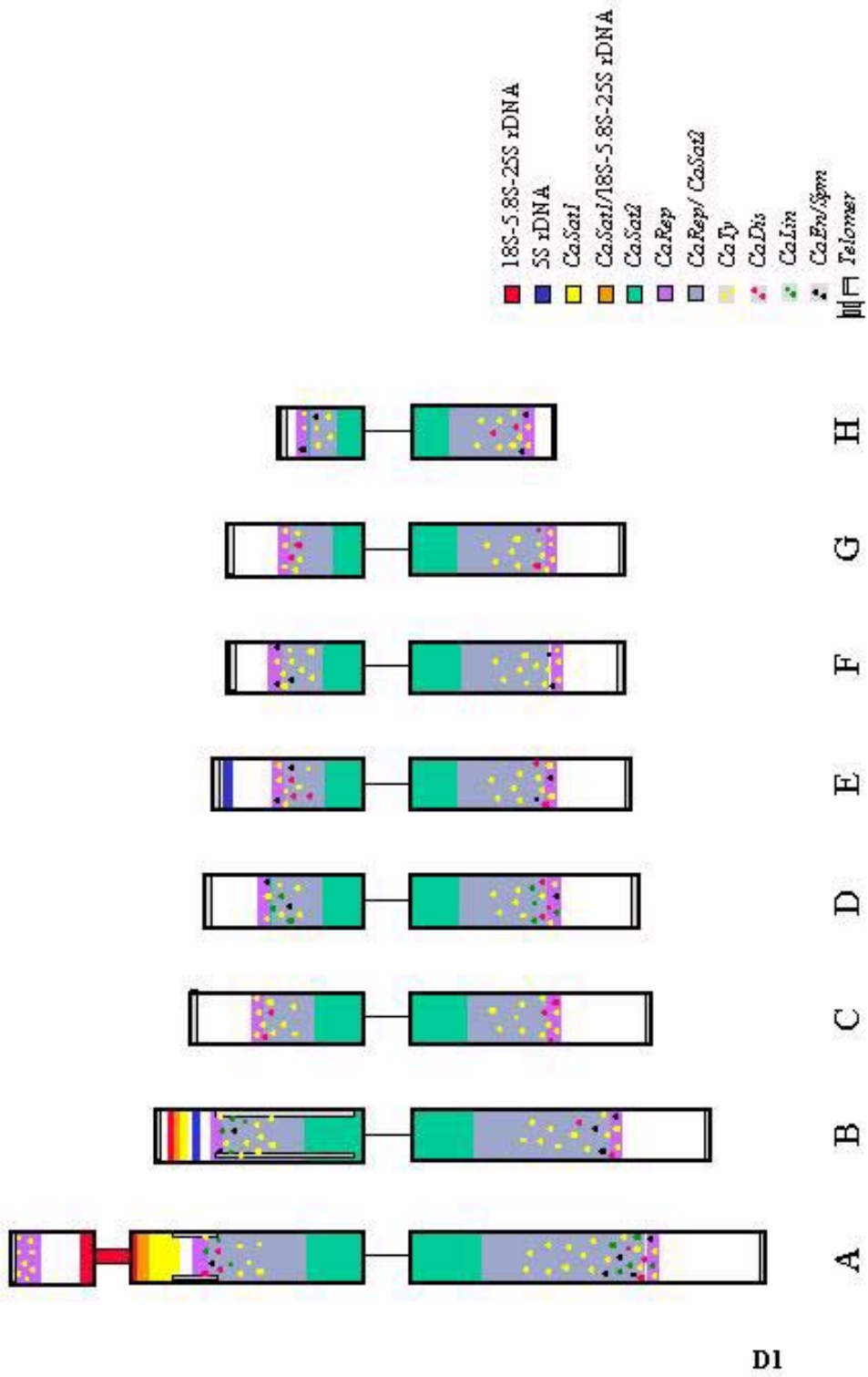
D.V.1. Übersicht über die Lokalisation der untersuchten repetitiven Familien

Die Ergebnisse der *in-situ*-Hybridisierung auf Metaphasechromosomen lassen sich zu einem vorläufigen Überblick über die Verteilung der untersuchten repetitiven Familien verwenden, der vereinfacht, nicht maßstabsgetreu und schematisch in Fig.D1 dargestellt ist und auch Ergebnisse vorangegangener Arbeiten beinhaltet (Galasso et al., 1996; Gortner et al., 1998).

Das hochabundante *Ty3-gypsy*-Element *CaRep* und der Satellit *CaSat2* sind hauptsächlich im pericentrischen Heterochromatin aller acht Chromosomen lokalisiert. *CaRep* weist darüberhinaus auch euchromatische Loci auf und ist vom zentralen Teil der Heterochromatinblöcke ausgeschlossen, der vom Tandemelement *CaSat2* bedeckt ist. In den distaleren Bereichen des Heterochromatins ist eine Vermischung beider Elemente nicht auszuschließen. Das zweite Tandemelement *CaSat1* ist auf Chromosomen A und B in direkter Nachbarschaft oder sogar überlappend mit den Loci der 18S-5.8S-25S rRNA-Gene

konzentriert und weist außerdem weitere Loci auf allen übrigen Chromosomen auf. Neben den beiden 18S-5.8S-25S rRNA-Genabfolgen auf Chromosom A und B existieren zwei Loci der 5S rRNA-Gene auf Chromosom B und E. Die weniger abundanten Retroelemente liegen wie *CaRep* im distalen Bereich der pericentrischen Heterochromatinblöcke aller (*CaTy*) oder einiger Chromosomen (*CaDis*, *CaLin*). Ob sie in diesen Bereichen ineinander verschachtelte Abfolgen bilden wie bei Mais (SanMiguel et al., 1996), ist ohne Sequenzierung oder höher auflösende Hybridisierungstechniken nicht auszumachen. Alle drei Sequenzfamilien weisen wie *CaRep* auch Signale in den angrenzenden euchromatischen Regionen auf. Sowohl *CaRep* als auch *CaTy* sind ebenfalls auf dem Satellitenanhängsel des größten Chromosomenpaars (A) vorhanden. Im distalen Heterochromatin und dem benachbarten Euchromatin von mindestens sechs Chromosomen befinden sich außer den verschiedenen Retroelementen auch Kopien des Transposons *CaEn/Spm*. Mikrosatellitenmotive sind je nach Sequenz im pericentrischen

Fig.D1: Übersicht über die chromosomale Lokalisation repetitiver Elemente in *C. arietinum*. Schematische und nicht maßstabsgetreue Darstellung der acht Metaphasechromosomen (verändert nach Galasso et al., 1996). Die Hauptloci der hochabundanten Elemente und rRNA-Gene sind als farbige Flächen dargestellt. Der mögliche Überlappungsbereich von 18S-5.8S-25S rDNA (rot) und *CaSat1* (gelb) ist orange dargestellt, der von *CaSat2* (grün) und *CaRep* (lila) in graublau. Die 5S rRNA-Gene sind blau gehalten. Die ungefähre Position der weniger abundanten Sequenzfamilien ist durch gelbe (*CaTy*), rote (*CaDis*), dunkelgrüne (*CaLin*) und schwarze (*CaEn/Spm*) Punkte dargestellt. Telomersequenzen existieren in Übereinstimmung mit Gortner et al. (1998) an den Chromosomenenden und im zentralen heterochromatischen Bereich von Chromosom A und B (zur Übersicht als graue Balken). Nicht aufgeführt ist die Lokalisation der verschiedenen Mikrosatellitenmotive (siehe dazu Gortner et al., 1998), zur besseren Übersicht wurde ebenfalls auf die Darstellung euchromatischer Loci einzelner Sequenzfamilien verzichtet.



Heterochromatin lokalisiert (CA, GATA) oder dispers entlang der Chromosomenarme verstreut (TA, A, AAC, GATA). Die Enden der acht Chromosomenpaare werden durch Telomersequenzen vom *A. thaliana*-Typ (TTTAGGG)_n abgeschlossen.

Eine *in-situ*-Hybridisierung auf Metaphasechromosomen wird vermutlich nicht alle Kopien der hier beschriebenen Elemente sichtbar machen können, da für konsistente Signale mehrere Kilobasen der Zielsequenz im Genom notwendig sind (Jiang & Gill, 1994). Eine Detektion von Zielsequenzen unter 1 Kbp wird nur selten berichtet (Desel et al., 2001). Auch die Feinauflösung der Signale ist durch die Kondensation der Metaphasechromosomen auf etwa 2-5 Mbp begrenzt (Trask, 1991; Jiang et al., 1996), so daß für detailliertere Aussagen besonders über die Zusammensetzung der distaleren Heterochromatinbereiche andere *in-situ*-Hybridisierungstechniken wie FISH auf Pachytän-Chromosomen oder „extended-fibre FISH“ (Franz et al., 1996a; de Jong et al., 1999) nötig sind.

D.V.2. Heterochromatische Lokalisation und weitgehende Inaktivität - Resultat eines Abwehrmechanismus durch das Wirtsgenom?

Die meisten der hier nachgewiesenen Kopien von *CaRep*, dem abundantesten Retroelement der Kichererbse, liegen im pericentrischen Heterochromatin. Fast alle der untersuchten mittelrepetitiven dispersen Elemente aus *C. arietinum* zeigen keinen Hinweis auf Transkription, und selbst bei *CaDis* ist eine Transkription bzw. Aktivität des Elements fraglich. Darüberhinaus weisen viele der untersuchten Elemente Leserasterwechsel und Stopcodons innerhalb ihrer codierenden Sequenzen auf und müssen daher zu den defekten Kopien gerechnet werden. Mangelnde Hinweise auf Transpositionsaktivität sind von vielen anderen Pflanzen her ebenfalls bekannt und auf gezielte Abwehr- bzw. Regulationsmechanismen der Wirtszellen - möglicherweise auch der Retroelemente selbst - gegen die mutagene Mobilität von transposablen Elementen zurückzuführen (Grandbastien, 1998, Kumar & Bennetzen, 1999). Diese Kontrolle von transposablen Elementen und Viren erfolgt größtenteils über Cytosinmethylierung, wie Beobachtungen am hypomethylierten Genom eines interspezifischen Säugerhybriden (*Macropus eugenii* x *Wallabia bicolor*) zeigten, wobei Hypomethylierung mit der gleichzeitigen starken Amplifikation von Retroelementen einherging (O'Neill et al., 1998). Aus *Arabidopsis* ist die Mobilisierung einer *En/Spm*-ähnlichen Transposonfamilie in Mutanten mit Hypomethylierung repetitiver Sequenzen bekannt. Dabei scheint das Gen DDM1 („decrease in DNA methylation“) für die Stabilisierung der Transposonaktivität verantwortlich zu sein (Miura et al., 2001). Cytosinmethylierung an 5'-CG-3'- bzw. 5'-CNG-3'-Sequenzen im Pflanzengenom wird von Heterochromatinbildung begleitet und führt zu Transkriptionsinaktivität sowie einer erhöhten Rate von C→T-Übergängen, d.h. es erfolgt zusätzlich eine Inaktivierung der Elemente durch eine erhöhte Mutationsrate (San Miguel et al., 1998; Hirochika et al., 2000). Diese Inaktivierung tritt oft erst auf, wenn die Kopienzahl des Elements erhöht ist und ähnelt daher einer durch ein repetitives Element induzierten Geninaktivierung (Matzke & Matzke, 1998). Zusätzlich zur epigenetischen Kontrolle durch Methylierung wirken posttranskriptionelle

Mechanismen, die sequenzspezifische RNA-Degradation im Cytoplasma beinhalten (Matzke et al., 2000). Außerdem führt die Präferenz von Retroelementen, ineinander zu inserieren ebenfalls zu gegenseitiger Inaktivierung der Elemente (SanMiguel et al., 1996), was möglicherweise durch Konkurrenz mehrerer Retroelemente um die Ressourcen des Wirts begünstigt wird (Kumar et al., 1997). Obwohl ohne Sequenzierung längerer DNA-Abschnitte unklar bleibt, ob ein Ineinanderspringen von Retroelementen auch bei Kichererbse stattgefunden hat, ist doch eine Häufung der dispersen, transposablen Elemente bzw. ihrer Relikte im pericentrischen Heterochromatin offensichtlich - und möglicherweise die Folge eines solchen Abwehrmechanismus, der auch die mangelnden Hinweise auf transpositionelle Aktivität der untersuchten Elemente erklären würde.

D.V.3. Hinweise auf eine Evolution der untersuchten Elemente

Bisher ist wenig bekannt über die Frequenz und Periodizität, mit der die Transposition von Retroelementen in der Evolution einzelner Pflanzen auftritt (Kumar & Bennetzen, 1999). Wie aus Sequenzvergleichen von LTRs hervorgeht, waren die meisten Retrotransposons in Mais innerhalb der letzten 2-6 Millionen Jahre aktiv (SanMiguel et al., 1998). Untersuchungen der Insertionspolymorphismen von *Ty1-copia*-Elementen in Gräsern (*BARE1*) oder in *Pisum* zeigen, daß die Transpositionsaktivität verschiedener Elemente im Laufe der Zeit stark variiert. Einige Familien bewegen sich seit langer Zeit nicht mehr im Genom, während andere in jüngerer Vergangenheit deutliche Aktivität aufwiesen (Waugh et al., 1997; Ellis et al., 1998; Gribbon et al., 1999; Pearce et al., 1999). Alle sieben untersuchten Elemente der Kichererbse waren offenbar schon in einem Vorgänger der annualen *Cicer*-Spezies vorhanden. Sieht man von der möglicherweise unterschiedlich starken Konservierung der verschiedenen Sondenbereiche ab, stellen die drei weniger abundanten Retrotransposons *CaTy*, *CaDis* und *CaLin* dabei die ursprünglicheren Elemente dar, die im Gegensatz zu den genusspezifischen Familien *CaSat1*, *CaSat2* und *CaRep* bereits früh in der Entwicklung der Leguminosen vorhanden waren. Dabei kann man das non-LTR-Element *CaLin*, das seit der Trennung der annualen Arten vermutlich nicht mehr aktiv war, als die ursprünglichste der drei Familien betrachten. Dies entspricht der Hypothese, LINEs seien die urtümlichsten Retrotransposons, aus denen LTR-Elemente durch Bildung der terminalen Wiederholungen hervorgegangen sind (Xiong & Eickbush, 1990; Danilevskaya et al., 1997; Eickbush, 1997). *CaTy* zeigt im Gegensatz zu *CaDis* und *CaLin* innerhalb der annualen Spezies deutliche Unterschiede in der genomischen Organisation und war daher vermutlich in jüngerer Zeit aktiv. *CaRep* ist vermutlich ebenfalls innerhalb der letzten 100 000 Jahre innerhalb der annualen Spezies aktiv gewesen und in *C. arietinum* zu hohen Kopienzahlen amplifiziert worden, was Organisation und Kopienzahlunterschiede belegen. Die beiden Satellitenelemente *CaSat1* und *CaSat2* wurden offenbar nach Abspaltung einer mit *C. cuneatum* gemeinsamen Vorgängerart hochamplifiziert, wobei *CaSat1* während der weiteren Entwicklung der einjährigen Arten eine unterschiedlich starke Amplifikation erfuhr. Auffällig ist die zentrale Lokalisation eines Großteils der *CaSat2*-Sequenzen im pericentrischen Heterochromatinblock, während die übrigen Retroelemente möglicherweise

untereinander und mit dem Satelliten vermischt in den distaleren Bereichen vorliegen. Hieraus könnte man schließen, daß nach der letzten Amplifikation von *CaRep* eine Akkumulation der evolutiv älteren Retroelemente im Heterochromatin durch gegenseitiges Ineinanderspringen erfolgte bzw. eine nachfolgende Heterochromatinbildung stattfand, was beides zu einer weitgehenden Inaktivierung dieser Elemente geführt hat. Die jüngste starke Amplifikation eines Teil der *CaSat2*-Sequenzen (nach Abspaltung des mit *C. cuneatum* gemeinsamen Vorgängers) könnte danach die Retroelemente in den distaleren Bereich des Heterochromatins abgedrängt haben. Allerdings bleibt dies ohne weitere Informationen über die chromosomale Lokalisation der Sequenzen in Genomen anderer aktueller Spezies weitgehend hypothetisch.

D.V.4. Die Verteilung repetitiver Sequenzen im Kichererbsengenom entspricht eher dem *A.thaliana*-Genom als Genomen vergleichbarer Größe

Verschiedene Pflanzengenome besitzen eine unterschiedliche Anordnung und Zusammensetzung von repetitiver DNA und Genen. Neben der unterschiedlichen Komposition und Lokalisation von Satellitenelementen, die durch speziesspezifische Amplifikationsereignisse oder unterschiedliche Stabilität der von der Länge der Monomere abhängigen Packung und Chromatinstruktur der Sequenzen zustandekommen kann (Kubis et al., 1998), haben auch transposable Elemente starken Einfluß auf die Genomkomposition. Renaturierungsstudien zeigten, daß große Pflanzengenome große Mengen repetitiver DNA enthalten, die größtenteils transposable Elemente darstellen und mit Genen durchsetzt sind (Übersicht bei Bennetzen, 2000). Die bisher sequenzierten Genombereiche zeigten neben genreichen Abschnitten mit wenigen eingestreuten Retrotransposons auch Blöcke aus ineinander verschachtelten Retroelementen, die mit größeren genischen Blöcken vermischt waren (Übersicht bei Bennetzen, 2000). Die genischen Bereiche kleinerer Genome (z.B. *A. thaliana*) zeigen dagegen nur sehr wenige eingestreute Retrotransposons. *Arabidopsis* besitzt jedoch nicht - wie zunächst vermutet - weniger, sondern nur konzentrierter lokalisierte und niedriger abundante repetitive Elemente (Kapitonov & Jurka, 1999).

Die heterochromatischen, pericentrischen Blöcke in Kichererbsenchromosomen nehmen einen beträchtlichen Teil der Chromosomenmasse ein (Galasso & Pignone, 1992). Auffällig ist die starke Konzentration der repetitiven Sequenzen in diesem Bereich: interstitielle Satelliten fehlen, und die verschiedenen Retrotransposonfamilien zeigen nur eine schwache Abdeckung euchromatischer Bereiche. Folglich sind genreiche Abschnitte im distalen Bereich der Chromosomenarme zu erwarten. Damit entspricht das hier vorgestellte, vorläufige Verteilungsschema repetitiver DNA grob den Verhältnissen in *A. thaliana*, zumal die Chromosomen von *C. arietinum* mit den großen, pericentrischen, DAPI-positiven Heterochromatinbereichen den *A. thaliana*-Chromosomen auch cytologisch sehr ähnlich sind (Brandes et al., 1997b). In *A. thaliana* ist der größte Teil der Chromosomenarme aus niedrigrepetitiven bzw. einzelnen Sequenzen zusammengesetzt und weist keine interstitiellen Satellitenelemente auf (Brandes et al., 1997a). Bedeutende Tandemelemente sind in den großen centromerischen und pericentrischen Heterochromatinblöcken konzentriert. Die dort

hauptsächlich lokalisierte AT-reiche Satellitenfamilie *pAL1* (Martinez-Zapater et al., 1986) enthält eingestreute mittelrepetitive Sequenzen, Retrotransposons und Mikrosatellitenmotive (Pelissier et al., 1996; Brandes et al., 1997a.; Heslop-Harrison et al., 1997). Auch die untypische Konzentration von *Tyl-copia*-Sequenzen im Heterochromatinblock von *A. thaliana* findet man bei *C. arietinum* (Brandes et al., 1997b; diese Arbeit). Im Gegensatz dazu sind bei *Beta vulgaris* - ebenfalls diploid und von ähnlicher Genomgröße wie Kichererbse (758 Mbp; Arumuganathan & Earle, 1991) - die Chromosomen größtenteils mit repetitiven Sequenzen besetzt, Gene sind auf Inseln zwischen den verschiedenen Blöcken repetitiver Elemente nahe den Chromosomenenden lokalisiert (Schmidt & Heslop-Harrison, 1998). Der centromerische Bereich wird vorwiegend aus Satellitenelementen gebildet (Schmidt & Heslop-Harrison, 1998), weitere Satelliten liegen interkalar gehäuft vor, ebenso wie LINE-Elemente. LTR-Retrotransposonsequenzen sind entlang der Chromosomenarme gleichmäßig verteilt, und Mikrosatellitenmotive liegen dispers oder lokal konzentriert dazwischen oder sind in Tandemelemente eingestreut (Kubis et al., 1998; Schmidt & Heslop-Harrison, 1998). Auch die grobe Chromosomenorganisation in Weizen und verwandter Gräser folgt diesem Chromosomenmodell, obwohl die Chromosomen z.T. zehnmal größer sind (Schmidt & Heslop-Harrison, 1998).

Die begrenzte Abundanz und konzentrierte Lokalisation repetitiver Familien in *A. thaliana* wird möglicherweise durch eine stärkere Kontrolle des Wirts über transposable Elemente, deren Insertionspräferenzen oder ihre Amplifikation gemeinsam mit anderen Elementen verursacht, was durch das verringerte Risiko, schädliche Mutationen zu erzeugen, evolutiv vorteilhaft war. Langdon et al. (2000) betrachten die Konzentration von Retrotransposons sogar als Strategie der Elemente, um in „toleranten“ genomischen Bereichen (z.B. heterochromatische Centromer- oder Telomerbereiche) weiter zu existieren und möglicherweise für den Wirt nützliche Funktionen zu übernehmen, ohne dessen Lebensfähigkeit bzw. Fertilität und damit die eigene Existenzgrundlage zu gefährden. Eine weniger starke Regulation von TEs und damit verbunden eine hohe Kopienzahl und dispersere Verteilung von transposablen Elementen können sich nur große, oft polyploide Genome mit einem gewissen Puffer gegen Insertionsmutagenese „leisten“ (Matzke et al., 2000). Die Kichererbse folgt nach den bisherigen Erkenntnissen - wie bereits von Brandes et al. (1997b) bei der Analyse der *Tyl-copia*-Sequenzen vermutet - offensichtlich dem Weg der kleinen Genome und konzentriert die transposablen Elemente in der Nähe von oder vermischt mit Satelliten-DNA im Heterochromatin. Dies kann durch gezielte Insertion der TEs in die heterochromatischen Bereiche, durch deren Methylierung und Chromatinkondensation oder im Zuge der Amplifikation eines Tandemelements geschehen. Auch sind aus der Gerste Mechanismen bekannt, mit denen Retrotransposonkopien deletiert und die Elemente so inaktiviert werden können (Vicient et al., 1999; Kalendar et al., 2000). Wieso sich *C. arietinum* nicht wie andere Genome vergleichbarer Größe entwickelt hat, bleibt unbekannt. Die Konzentration der TEs war aber offenbar evolutiv vorteilhaft.

Es ist sehr unwahrscheinlich, daß die untersuchten Sequenzfamilien das vollständige Spektrum repetitiver DNA im Genom von *C. arietinum* darstellen. Einerseits wurden bisher

noch nicht alle mittelrepetitiven Sequenzen aus der genomischen Bank untersucht, aus der *CaRep* und die beiden Satelliten isoliert wurden. Ebenso stellt diese Bank durch die Restriktion mit *EcoRI* eine möglicherweise einseitige Auswahl an repetitiven Sequenzen bereit. Weiterhin ist keine der beschriebenen Sequenzen im subtelomerischen, CMA (Chromomycin A3)-positiven Heterochromatin des Chromosoms B lokalisiert, das keine NOR trägt (Galasso & Pignone, 1992), so daß auch hier weitere repetitive Sequenzen zu erwarten sind. Galasso et al. (2001) beschreiben die chromosomenspezifische, nicht-tandemförmig organisierte Sequenzfamilie *pLc7*, die in niedrigerer Kopienzahl auch in Kichererbse nachgewiesen werden konnte, jedoch keine Homologie zu einem der hier beschriebenen Elemente aufwies. Auch Retrotransposonsequenzen in zufällig aus dem Genom amplifizierten Fragmenten, die keine Verwandtschaft zu den identifizierten Familien aufweisen, deuten auf die Anwesenheit weiterer repetitiver Elemente hin (Benko-Iseppon, Hüttel; pers. Mitteilung). Ferner gibt es weitere Sequenzklassen, die ubiquitär im Pflanzenreich sind und deren Anwesenheit im Kichererbsengenom noch nicht untersucht wurde, wie etwa *Mutator*- oder *hAT*-(*hobo/Activator/Tam3*) bzw. *Ac/Ds*-ähnliche DNA-Elemente (Kunze et al., 1997) oder die kleineren, vermutlich zu den nicht-autonomen DNA-Elementen gehörenden MITEs, zu denen *Tourist* und *Stowaway* zählen (Wessler et al., 1995). Ebenso wurde das Kichererbsengenom noch nicht nach non-LTR-Elementen der SINEs (short interspersed nuclear element; Übersicht bei Schmidt, 1999) untersucht. Auch bei der Genomsequenzierung von *A. thaliana* konnten über die bereits bekannten Sequenzfamilien hinaus (Martinez-Zapater et al., 1986; Konieczny et al., 1991; Tsay et al., 1993; Pelissier et al., 1995; Wright et al., 1996; Wright & Voytas, 1998) noch etliche, weniger abundante Elemente identifiziert werden (Kapitonov & Jurka, 1999; Le et al., 2000; The *Arabidopsis* Genome Initiative, 2000). Auch bei Kichererbse ist eine zukünftige Detektion weiterer mittelrepetitiver Sequenzfamilien anzunehmen, allerdings ist es unwahrscheinlich, daß diese einen großen Anteil des Genoms ausmachen. Insgesamt kann man davon ausgehen, daß *CaRep* und die beiden Satellitenelemente die abundantesten Sequenzfamilien im *C. arietinum*-Genom darstellen. Unterstützt wird dies durch die Abundanz dieser Sequenzen in zufälligen PCR-Fragmenten (Benko-Iseppon, Hüttel; pers. Mitteilung).

D.V.5. Die Groborganisation der repetitiven DNA und die Genomanalyse von *C. arietinum*

Die Erkenntnisse über die grobe Chromosomenorganisation der Kichererbse sind vor allem für die Genomanalyse nützlich. Das jetzt vorhandene Wissen über euchromatische, möglicherweise zu genischen Bereichen benachbarten Kopien einiger Retrotransposons (*CaDis*- oder *CaLin*) kann gezielt zur Verfeinerung der genetischen Karte und zum markergestützten Klonieren bestimmter Gene genutzt werden. Dabei eröffnen Bereiche, in denen verschiedene Retroelemente und Mikrosatelliten gemeinsam vorliegen, besondere Möglichkeiten für kombinierte Markertechniken, zumal die LTR-Sequenzen von *CaRep* bekannt sind. Ferner ist unsere Kenntnis über die Konzentration der repetitiven Sequenzen im pericentrischen Heterochromatin hilfreich, um die Proportionen der genetischen Karte von *C.*

arietinum einschätzen zu können. Die physikalischen Abstände zwischen Genen und Markern entsprechen in Gräsern (Heslop-Harrison, 1991; Lukazewski, 1992; Schwarzacher, 1996), in *Arabidopsis* (Lin et al., 1999; Mayer et al., 1999) und beim Menschen (Dunham et al., 1999) nur wenig den Distanzen auf der genetischen Karte. Untersuchungen mit Mais zeigen, daß Gene allgemein als „hot spots“ für Rekombinationsereignisse dienen (Dooner & Martinez-Ferez, 1997), während die Rekombinationsfrequenz in den heterochromatischen Bereichen mit transposablen Elementen von *A. thaliana* deutlich herabgesetzt ist (The *Arabidopsis* Genome Initiative, 2000) und die Korrelation zwischen niedriger Genexpression, hoher Transposondichte und niedriger Rekombination unterstreicht (Martienssen, 1998). Die großen pericentrischen Heterochromatinblöcke von *C. arietinum* stellen daher vermutlich nur einen kleinen, gering aufgelösten Abschnitt in der auf Rekombination basierenden genetischen Karte dar, in dem die aus repetitiver DNA generierten Marker dicht gedrängt sitzen, obwohl sie einen großen Teil mancher Chromosomen ausmachen. Die genischen Bereiche in den distaleren Abschnitten der Chromosomenarme werden demnach eine bessere Auflösung auf der genetischen Karte zeigen, jedoch durch den physikalisch nicht unbedeutenden Bereich der pericentrischen repetitiven Sequenzen unterbrochen werden.

E. ZUSAMMENFASSUNG

Die Kichererbse (*C. arietinum* L.) ist eine der ältesten kultivierten Leguminosen und stellt eine Nahrungsgrundlage in vielen Ländern mit semiaridem Klima dar. Seit einigen Jahren unterstützen molekulare Marker sowohl die Züchtungsforschung als auch die Genomanalyse dieser Pflanze, so daß bereits eine genetische Karte des Kichererbsengenoms erstellt werden konnte. Da repetitive Komponenten einen Großteil des Genoms ausmachen können, ist ihre Charakterisierung zum Verständnis der Organisation des Kichererbsengenoms unerlässlich.

Insgesamt konnten sieben repetitive Sequenzfamilien aus *C. arietinum* isoliert und identifiziert werden, von denen einige direkt in der Genomanalyse bzw. für taxonomische Fragen Anwendung fanden. Außerdem wurde ein universelles Primerpaar zur schnellen Detektion von DNA-Transposons entwickelt und verwendet.

1. Die beiden Satellitenelemente *CaSat1* und *CaSat2* sowie das *Ty3-gypsy*-ähnliche Retrotransposon *CaRep* stellen vermutlich die abundantesten Familien repetitiver Sequenz dar. Die beiden AT-reichen Satelliten bestehen aus 162-167 bzw. 100 Bp langen, hintereinander („head-to-tail“) angeordneten Monomeren. *CaSat2* ist ein dominanter Bestandteil des pericentrischen Heterochromatins, während *CaSat1* hauptsächlich in der Nachbarschaft der 18S-5.8S-25S rDNA-Loci lokalisiert ist. Eine Verwandtschaft von *CaSat1* zu IGS-Sequenzen ist nicht auszuschließen. Eine vollständige Kopie des dispersen Retroelements *CaRep*, das im pericentrischen Heterochromatin konzentriert ist, jedoch den zentralen Bereich von *CaSat2* ausspart, konnte rekonstruiert werden. Das Element ist 11.4 Kbp lang und hat relativ lange, flankierende LTRs von je 3.3 Kbp. Alle drei hochrepetitiven Elemente sind spezifisch für die Gattung *Cicer*. Die genomische Organisation von *CaRep* wurde offenbar während der Evolution der annualen *Cicer*-Spezies mehrfach rearrangiert.
2. Da die beiden Satelliten *CaSat1* und *CaSat2* in allen untersuchten Spezies der Gattung *Cicer* (außer *C. cuneatum*) vorhanden sind, lassen sich ihre Abundanz und Organisation zur Klärung taxonomischer Fragen verwenden. Die Ergebnisse unterstützen die Außenseiterrolle von *C. cuneatum* innerhalb der annualen *Cicer*-Spezies und damit Thesen zu einem separaten Ursprung dieser Annualen sowie die Forderung nach der taxonomischen Trennung von der Sektion *Monocicer*. Auch unterstützen und ergänzen die Ergebnisse über die Satellitenelemente bereits vorliegende molekulare Daten, Karyotypisierungen sowie Kreuzbarkeitsdaten und bestätigen die Gruppierung von *C. chorassanicum* sowie der ersten Kreuzungsgruppe.
3. Das hochabundante Retrotransposon *CaRep* kann für eine S-SAP („sequence-specific amplified polymorphisms“)-ähnliche Markertechnik verwendet werden und liefert polymorphe molekulare Marker, die sich in die genetische Karte integrieren lassen. Die damit detektierten Polymorphismen werden vermutlich durch Variabilität innerhalb der LTR-Sequenzen erzeugt. Ein Teil der so generierten Marker ist mit den

Resistenzgenloci gegen *Fusarium oxysporum* f.sp. *ciceri* gekoppelt. Die Hybridisierung bestätigt, daß durch die Marker tatsächlich Loci des Retrotransposons auf der genetischen Karte lokalisiert werden.

4. Unabhängig davon konnten durch PCR-Ansätze die drei mittelrepetitiven, dispersen Retrotransposonfamilien *CaDis*, *CaTy* und *CaLin* identifiziert werden. Von den beiden LTR-Retrotransposons *CaDis* und *CaTy* gehört letztere Familie zur *Ty1-copia*-Untergruppe, *CaLin* repräsentiert ein non-LTR-Element der LINE-Klasse. Alle drei Familien sind weniger abundant als *CaRep* und in den distalen Bereichen einiger oder aller Heterochromatinblöcke bzw. in den angrenzenden euchromatischen Regionen zu finden. Im Gegensatz zu den hochabundanten Familien kommen homologe Sequenzen der drei Familien auch außerhalb der Gattung *Cicer* vor und deuten auf eine evolutiv frühe Anwesenheit innerhalb der Fabaceen hin. Die sehr heterogenen Sequenzen von *CaLin* sind vermutlich sehr ursprünglich und haben während der Evolution der Annuellen keine großen Rearrangements erfahren. *CaDis*-Sequenzen werden in vielen Geweben transkribiert, jedoch ist unklar, ob dies auf eine Aktivität des Elements hinweist oder auf das Durchlesen benachbarter Gene zurückzuführen ist.
5. Um auch transposable Elemente der Klasse II (DNA-Transposons) nachweisen zu können, wurde ein degeneriertes Primerpaar aus dem Bereich der Transposase abgeleitet, mit dessen Hilfe pflanzliche *En/Spm*-Elemente in über 30 Pflanzenarten aus 20 Gattungen, darunter auch wichtigen Kulturpflanzen wie Zuckerrübe, Weizen und Erbse nachgewiesen werden können. Dieser Ansatz stellt einen Ausgangspunkt für die Charakterisierung solcher Elemente in beliebigen Pflanzengenomen und ihre Verwendung bei der Genomanalyse sowie zur Genidentifikation und -isolation dar.
6. Die niedrig- bis mittelrepetitive, disperse *En/Spm*-Familie aus *C. arietinum* (*CaEn/Spm*) umfaßt drei Subfamilien und ist auf mindestens sechs Chromosomen im distalen Teil der Heterochromatinblöcke angrenzend an euchromatische Bereiche lokalisiert. Die hier vorgelegten Befunde umfassen die erste chromosomale Lokalisation eines DNA-Transposons in Pflanzen durch FISH. Die genomische Organisation homologer Sequenzen in anderen *Cicer*-Spezies deutet möglicherweise auf transposable Aktivität während der Evolution der Annuellen hin. *CaEn/Spm*-ähnliche Elemente scheinen bereits in einem Vorgänger von *C. arietinum* und nahe verwandten Leguminosen vorhanden gewesen zu sein.

Durch die vorliegenden Daten war es möglich, ein vorläufiges Modell der Verteilung repetitiver Sequenzen auf den acht Kichererbsenchromosomen zu entwerfen, das für die weitere Genomanalyse von *C. arietinum* genutzt werden kann. Obwohl vermutlich nur ein Teil aller repetitiven Komponenten dieses Genoms charakterisiert wurde, deutet sich - im Gegensatz zu Pflanzen mit vergleichbarer Genomgröße - durch die Konzentration repetitiver

Sequenzen im pericentrischen Heterochromatin eine Ähnlichkeit zur Chromosomenorganisation in *Arabidopsis thaliana* an.

E. Summary

Chickpea (*C. arietinum* L.) is one of the oldest food legumes and represents a basic source for nutrition in many countries of semi-arid areas. Molecular markers have supported breeding research as well as genome analysis of this plant for several years and a genetic map of the chickpea genome has already been established. Since repetitive components can make up the majority of a plant genome, their characterization is essential for the understanding of the chickpea genome organization.

Altogether, seven repetitive sequence families have been identified in *C. arietinum*, some of which could be directly used for genome analysis or to elucidate taxonomic questions. Moreover a universal primer pair has been developed for the fast detection of class II transposons.

1. The satellite elements *CaSat1* and *CaSat2* as well as the *Ty3*-gypsy-like retrotransposon *CaRep* probably represent the most abundant classes of repetitive DNA. The AT-rich satellites consist of 162-167 bp- (*CaSat1*) and 100 bp- (*CaSat2*) monomers in head-to-tail arrangement. *CaSat2* is a major component of the pericentric heterochromatin, whereas *CaSat1* is mainly localized adjacent to the 18S-5.8S-25S rDNA clusters on chromosomes A and B. Hence a relationship of the *CaSat1* sequences to the IGS subrepeats cannot be ruled out. A full copy of the dispersed retroelement *CaRep* was reconstructed. *CaRep* sequences are mainly concentrated in the pericentric heterochromatin but are excluded from the central areas which are probably occupied by *CaSat2*. The element is 11.2 kb long and contains large LTRs of 3.3 kb. The three highly repetitive elements are specific for the genus *Cicer* and the organization of *CaRep* has been rearranged several times during the evolution of the annual *Cicer* species.
2. The presence of *CaSat1* and *CaSat2* in all *Cicer* species investigated (except *C. cuneatum*) as well as their variation in structure and copy number provide useful tools for taxonomic studies. The results support the special status of *C. cuneatum* within the annual *Cicer* species and a separate origin as well as the suggestion to separate *C. cuneatum* from the section *Monocicer*. They also complete and support already existing molecular data, karyotyping and crossability results concerning the taxonomic position of *C. chorassanicum* and the first crossability group.
3. The highly abundant retrotransposon *CaRep* can be used for an S-SAP („sequence-specific amplified polymorphisms“)-like marker technique and produces polymorphic markers, which can be integrated into the genomic map. The detected polymorphisms

seem to be generated by the variability within the LTR sequences. Some of these markers are genetically linked to the resistance loci against *Fusarium oxysporum* f.sp. *ciceri*. Hybridization proves them in fact to represent retrotransposon loci on the genetic map.

4. Independently, three middle-repetitive dispersed retrotransposon classes *CaDis*, *CaTy* and *CaLin* could be identified by PCR. *CaDis* and *CaTy* represent LTR-retrotransposons, the latter belonging to the *Ty1-copia* subgroup, whereas *CaLin* is a non-LTR-retrotransposon of the LINE class. All three families are less abundant than *CaRep* and localized in the distal parts of some or all pericentric heterochromatic blocks. Sometimes they also reside in adjacent euchromatic regions. In contrast to the highly abundant sequence classes, homologous sequences of these families are not restricted to the genus *Cicer* which indicates their early existence in the evolution of the Fabaceae. The heterogeneous sequences of *CaLin* seem to be of ancient origin and have not been rearranged severely during the evolution of the annual species. *CaDis* sequences are transcribed in several tissues, but it is not clear whether this reflects the activity of the element or the transcriptional read-through from neighbouring genes.
5. In order to prove the existence of class II transposable elements, degenerate primers were deduced from the TPase region of plant *En/Spm*-like transposons. The primer pair allowed the amplification of similar sequences from more than 30 plant species of 20 genera including important crops like sugar beet, wheat and pea. This approach can be used for the detection and characterization of DNA transposable elements in plant species, their complete isolation and use for genome analysis or gene identification.
6. The low-to-middle-repetitive dispersed *En/Spm*-family from *Cicer arietinum* (*CaEn/Spm*) comprises three subfamilies and is located on at least six chromosomes in the distal part of the heterochromatic blocks adjacent to euchromatic areas. The present work contains the first report about the chromosomal localization of class II elements in plants by FISH. The genomic organization of homologous sequences in other *Cicer* species probably indicates transpositional activity during the evolution of the annual species. *CaEn/Spm*-like elements have probably been already present in a common ancestor of *C. arietinum* and closely related legume species.

The present data allow the construction of a preliminary model of the distribution of repetitive sequences on the chickpea chromosomes, which can be used for the genome analysis of *C. arietinum*. Though probably only part of the repetitive components of this genome has been characterized, the concentration of repetitive sequences in the pericentric heterochromatin indicates similarity to the organization of *Arabidopsis thaliana* chromosomes.

F. LITERATUR

- AARTS, MGM., CORZAAN, P., STIEKEMA, WJ., PEREIRA, A. (1995) A two-element *Enhancer-Inhibitor* transposon system in *Arabidopsis thaliana*. *Mol Gen Genet* 247: 555-564.
- ABBO, S., MILLER, TE., READER, SM., DUNFORD, RP., KING, IP. (1994) Detection of ribosomal RNA sites in lentil and chickpea by fluorescent in-situ hybridization. *Genome* 37: 713-716.
- AGRAWAL, A., EASTMANN, Q., SCHATZ, D. (1998) Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature* 394: 744-751.
- AHMAD, F., GAUR, PM., SLINKARD, AE. (1992) Isozyme polymorphism and phylogenetic interpretations in the genus *Cicer* L.. *Theor Appl Genet* 83: 620- 627.
- AHMAD, F. (1999) Random amplified polymorphic DNA (RAPD) analysis reveals genetic relationships among the annual *Cicer* species. *Theor Appl Genet* 98: 657-663.
- AHMAD, F. (2000) A comparative study of chromosome morphology among nine annual species of *Cicer* L. *Cytobios* 101: 37-53.
- AHMAD, F., SLINKARD, AE. (1992) Genetic relationships in the genus *Cicer* L. as revealed by polyacrylamid electrophoresis of seed storage proteins. *Theor Appl Genet* 84: 688-692.
- ALIX, K., BAURENS, F., PAULET, F., GLASZMANN, J., D'HONT, A. (1998) Isolation and characterization of a satellite DNA family in the *Saccharum* complex. *Genome* 41: 854-864.
- ALTSCHUL, SF., GISH, W., MILLER, W., MYERS, EW., LIPMAN, DJ. (1990) Basic local alignment search tool. *J Mol Biol* 215: 403-410
- ALTSCHUL, SF., MADDEN, TL., SCHÄFFER, AA., ZHANG, J., ZHANG, Z., MILLER, W., LIPMAN, DJ. (1997) Gapped BLAST and PSI BLAST: a new generation of protein database search programs. *Nucl Acids Res* 25: 3389-3402.
- ANAMTHAWAT-JÓNSSON, K., HESLOP-HARRISON, JS. (1993) Isolation and characterization of genome-specific DNA sequences in *Triticeae* species. *Mol Gen Genet* 240: 151-158.
- ANANIEV, EV., PHILLIPS, RL., RINES, HW. (1998b) Chromosome-specific molecular organization of maize (*Zea mays* L.) centromeric regions. *Proc Natl Acad Sci USA* 95: 13073-13078.
- ANANIEV, EV., PHILLIPS, RL., RINES, HW. (1998a) Complex Structure of knob DNA on maize chromosome 9: Retrotransposon invasion into heterochromatin. *Genetics* 149: 2025-2037.
- ANAYA, N., RONCERO MI. (1995) *Skippy*, a retrotransposon from the fungal plant pathogen *Fusarium oxysporum*. *Mol Gen Genet* 249: 637-647.
- ARAGON-ACLAIDE, L., MILLER, T., SCHWARZACHER, T., READER, S., MOORE, G. (1996) A cereal centromeric sequence. *Chromosoma* 105: 261- 268.

- ARNHEIM, N., SEPERACK, P., BANERJI, J., LANG, RB., MIESFELD, R., MARCU, KB. (1980) Mouse rDNA nontranscribed spacer sequences are found flanking immunoglobulin C_H genes and elsewhere throughout the genome. *Cell* 22: 179- 185.
- ARUMUGANATHAN, K., EARLE, ED. (1991) Nuclear DNA content of some important plant species. *Plant Mol Biol Rep* 9: 208-218.
- ASÍNS, MJ., MONFORTE, AJ., MESTRE, PF., CARBONELL, EA. (1999) *Citrus* and *Prunus copia*-like retrotransposons. *Theor Appl Genet* 99: 503-510.
- AVRAMOVA, Z., TIKHONOV, A., SAN MIGUEL, P., JIN, YK., LIU, C., WOO, SS., WING, RA., BENNETZEN, JL. (1996) Gene identification in a complex chromosomal continuum by local genomic cross-referencing. *Plant J* 10:1163-1168.
- BALINT-KURTI, PJ., CLENDENNEN, SK., DELEZELOVA, M., VALARIK, M., DOLEZEL, J., BEETHAM, PR., MAY, GD. (2000) Identification and chromosomal localization of the *monkey* retrotransposon in *Musa* sp.. *Mol Gen Genet* 263: 908-915.
- BARNES, SR., JAMES, AM., JAMIESON, G. (1985) The organization, nucleotide sequence and chromosomal distribution of a satellite DNA from *Allium cepa*. *Chromosoma* 92: 185-192.
- BARNES, W.M. (1994) PCR amplification of up to 35kb DNA with high fidelity and high yield from λ -bacteriophage templates. *Proc Natl Acad Sci USA* 91: 2216-2220.
- BELYAYEV, A., RASKINA, O., NEVO, E. (2001) Chromosomal distribution of reverse transcriptase-containing retroelements in two Triticeae species. *Chrom Res* 9: 129-136.
- BENNETT, MD. (1982) Nucleotypic basis of the spatial ordering of chromosomes in eukaryotes and the implications of the order for genome evolution and phenotypic variation. In: Dover, G., Flavell, R. (eds.): *Genome evolution*, Academic Press, London.
- BENNETT, MD., SMITH, JB. (1976) Nuclear DNA amounts in angiosperms. *Phil Trans R Soc Lond B* 274: 227-274.
- BENNETZEN, J. (2000) Transposable element contributions to plant gene and genome evolution. *Plant Mol. Biol.* 42: 251-269.
- BENNETZEN, JL. (1996) The contributions of retroelements to plant genome organization, function and evolution. *Trends Microbiol* 4: 347-353.
- BENSLIMANE, AA., DRON, M., HARTMANN, C., RODE, A. (1986) Small tandemly repeated DNA sequences of higher plants likely originate from a tRNA gene ancestor. *Nucl Acids Res* 14: 8111-8119.
- BESTOR, T., TYCKO, B. (1996) Creation of genomic methylation patterns. *Nat Genet* 12: 363-367.
- BHATTACHARYYA, MK., GONZALES, RA., KRAFT, M., BUZZELL, RI. (1997) A *copia*-like retrotransposon *Tgmr* closely linked to the *Rps1-k* allele that confers race-specific resistance of soybean to *Phytophthora sojae*. *Plant Mol Biol* 34: 255-264.

- BIESSMANN, H., MASON, JM. (1994) Telomeric repeat sequences. *Chromosoma* 103: 154-161.
- BIRNBOIM, HC., DOLY, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucl Acids Res* 7, 1513-1523.
- BOEKE, JD. (1997) LINEs and *Alus*- the polyA connection. *Nature Genet* 16: 6-7.
- BOEKE, JD., CORCES, VG. (1989) Transcription and reverse transcription of retrotransposons. *Annu Rev Microbiol* 43: 403-434.
- BONAS, U., SOMMER, H., SAEDLER, H. (1984) The 17 kb *Tam1* element of *Antirrhinum majus* induces a 3bp duplication upon integration into the chalcon synthase gene. *EMBO J* 3: 1015-1019.
- BRANDES, A., THOMPSON, H., DEAN, C., HESLOP-HARRISON, JS. (1997a) Multiple repetitive sequences in the paracentromeric regions of *Arabidopsis thaliana* L. *Chrom Res* 5: 238-246.
- BRANDES, A., HESLOP-HARRISON, JS., KAMM, A., KUBIS, S., DOUDRICK, RL., SCHMIDT, T. (1997b) Comparative analysis of the chromosomal and genomic organization of *Ty1-copia*-like retrotransposons in pteridophytes, gymnosperms and angiosperms. *Plant Mol Biol* 33: 11-21.
- BROSIUS, J. (1999) Genomes were forged by massive bombardments with retroelements and retrosequences. *Genetica* 107: 209-238.
- BROSIUS, J. (2001) Das Säugergenom - Schrottplatz der Evolution? *Laborjournal* 5: 18-23.
- BROUN, P., TANKSLEY, SD. (1993) Characterization of tomato DNA clones with sequence similarity to human minisatellites 33.6 and 33.15. *Plant Mol Biol* 23: 231-242.
- BROWN, GR., NEWTON, CH., CARLSON, JE. (1998) Organization and distribution of a *Sau3a* tandem repeated DNA sequence in *Picea* (Pinaceae) species. *Genome* 41: 560-565.
- BUREAU, TE., WESSLER, SR. (1992) *Tourist*: a large family of small inverted repeat elements frequently associated with maize genes. *Plant Cell* 4: 1283-1294.
- BUREAU, TE., WESSLER, SR. (1994) *Stowaway*: a new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants. *Plant Cell* 6: 907-916.
- BUREAU, TE., WHITE, SE., WESSLER, SR. (1994) Transduction of a cellular gene by a plant retroelement. *Cell* 77: 479-480.
- CALDERINI, O., PUPILLI, F., PAOLOCCI, F., ARCIONI, S. (1997) A repetitive and species-specific sequence as a tool for detecting the genome contribution in somatic hybrids in the genus *Medicago*. *Theor Appl Genet* 95: 734-740.
- CARDON, GH., FREY, M., SAEDLER, H., GIERL, A. (1993) Mobility of the maize transposable element *En/Spm* in *Arabidopsis thaliana*. *Plant J* 3: 773-784.

- CASACUBERTA, JM., GRANDBASTIEN, MA. (1993) Characterization of LTR sequences involved in the protoplast specific expression of the tobacco *Tnt1* retrotransposon. Nucl Acids Res 21: 2087-2093.
- CASACUBERTA, JM., VERNHETTES, S., AUDEON, C., GRANDBASTIEN, MA. (1997) Quasispecies in retrotransposons: a role for sequence variability in *Tnt1* evolution. Genetica 100: 109-117.
- CHARLESWORTH, B., SNIEGOWSKI, P., STEPHAN, W. (1994) The evolutionary dynamics of repetitive DNA in eukaryotes. Nature 371: 215- 220.
- CHAVANE, F., ZHANG, DX., LIAUD, MF., CERFF, R. (1998) Structure and evolution of *Cyclops*: a novel giant retrotransposon of the *Ty3-gypsy* family highly amplified in pea and other legume species. Plant Mol Biol 37: 363-375.
- CHOPRA, S., BRENDDEL, V., ZHANG, J., AXTELL, DA. AND PETERSON, T. (1999) Molecular characterization of a mutable pigmentation phenotype and isolation of the first active transposable element from *Sorghum bicolor*. Proc Natl Acad Sci USA, 96: 15330-15335.
- CHOUMANE, W., WINTER, P., WEIGAND, F., KAHL, G. (2000) Conservation and variability of sequence-tagged microsatellite sites (STMSs) from chickpea (*Cicer arietinum* L.) within the genus *Cicer*. Theor Appl Genet 101: 269-278.
- CLARK, J., KIM, P. AND KIDWELL, M. (1998) Molecular evolution of P transposable elements in the genus *Drosophila*. III. The melanogaster species group. Mol Biol Evol 15: 746-755.
- COELHO, P., QUEIROZ-MACHADO, J., HARTL, D. AND SUKEL, C. (1998) Pattern of chromosomal localization of the *hoppe1* transposable element family in the *Drosophila melanogaster* subgroup. Chrom Res 6: 385-395.
- COEN, ES., CARPENTER, R., MARTIN, C. (1986) Transposable elements generate novel spatial patterns of gene expression in *Antirrhinum majus*. Cell 47: 285-296.
- COPENHAVER, GP., NICKEL, K., KUROMORI, T., BENITO, M., KAUL, S., LIN, X., BEVAN, M., MURPHY, G., HARRIS, B., PARNELL, LD., MCCOMBIE, WR., MARTIENSSSEN, RA., MARRA, M., PREUSS, D. (1999) Genetic definition and sequence analysis of *Arabidopsis* centromeres. Science 286: 2468-2474.
- COPENHAVER, GP., PREUSS, D. (1999) Centromeres in the genomic era: unraveling paradoxes. Curr Opin Plant Biol 2: 104-108.
- COURTIAL, B., FEUERBACH, F., EBERHARD, S., ROHMER, L., CHIAPELLO, H., CAMILLERI, C., LUCAS, H. (2001) *Tnt1* transposition events are induced by in-vitro transformation of *Arabidopsis thaliana*, and transposed copies integrate into genes. Mol Genet Genomics 265: 32-42.
- CRISTOFARI, G., FICHEUX, D., DARLIX, JL. (2000) The *gag*-like protein of the yeast *Ty1* retrotransposon contains a nucleic acid chaperone domain analogous to retroviral nucleocapsid proteins. J Biol Chem 275: 19210-19217.

- CSINK, AK., HENIKOFF, S. (1998) Something for nothing: the evolution and utility of satellite repeats. *Trends in Genomics* 14: 200-204.
- CUADRADO, A., JOUVE, N. (1997) Distribution of highly repeated DNA sequences in species of the genus *Secale*. *Genome* 40: 309- 317.
- CURCIO, MJ., GARFINKEL, DJ. (1999) New lines of host defense: inhibition of *Ty1* retrotransposon by *Fus3p* and *NER/TFIIH*. *Trend Genet* 15: 43-45.
- DANILEVSKAYA, ON., ARKHIPOVA, IR., TRAVERSE KL., PARDUE, ML. (1997) Promoting in tandem: the promotor for telomere transposons *Het-A* and implications for the evolution of retroviral LTRs. *Cell* 88: 647-655.
- DANILEVSKAYA, ON., TAN, C., WONG, J., ALIBHAI, M., PARDUE, ML. (1998) Unusual features of the *Drosophila melanogaster* telomere transposable element *Het-A* are conserved in *Drosophila yakuba* telomere elements. *Proc Natl Acad Sci USA* 95: 3770-3775.
- DEJEU, MJ., LASSCHUIT, J., KUIPPERS, AGJ., KAMSTRA, SA., VISSER, RGF. (1997) Characterization and localization of repetitive DNA sequences in the ornamental *Alstroemeria aurea* Graham. *Theor Appl Genet* 94: 982-990.
- DEJONG, JH., FRANSZ, P., ZABEL, P. (1999) High resolution FISH in plants: techniques and applications. *Trends Plant Sci* 4: 258–263.
- DELUCCINI, S., ANDRONICO, F., ANDREAZZOLI, M., GIULIANI, M., SAVINO, R., NARDI, I. (1988) Extra-ribosomal spacer sequences in *Triturus*. *J Mol Biol* 204: 805-813.
- DERAGON, JM., GILBERT, N., ROUQUET, L., LENOIR, A., ARNAUD, P., PICARD, G. (1996) A transcriptional analysis of the *SIBn* (*Brassica napus*) family of SINE. *Plant Mol Biol* 32: 869-878.
- DESEL, C., JUNG, C., CAI, D., KLEINE, M., SCHMIDT, T. (2001) High-resolution mapping of YACs and the single-copy gene *Hsl pro-1* on *Beta vulgaris* chromosomes by multi-colour fluorescence *in situ* hybridization. *Plant Mol Biol* 45: 113–122.
- DEVINE, S. AND BOEKE, J. (1994) Efficient integration of artificial transposons into plasmid targets *in vitro*: a useful tool for DNA mapping, sequencing and genetic analysis. *Nucl Acids Res* 22: 3765-3772.
- DONG, F., MILLER, JT., JACKSON, SA., WANG, GL., RONALD, PC., JIANG, J. (1998) Rice (*Oryza sativa*) centromeric regions consist of complex DNA. *Proc Natl Acad Sci USA* 95: 8135-8140.
- DOOLITTLE, WF., SAPIENZA, C. (1980) Selfish genes, the phenotype paradigm and genome evolution. *Nature* 284: 601-603.
- DOONER, HK., MARTINEZ-FEREZ, IM. (1997) Recombination occurs uniformly within the *bronze* gene, a meiotic recombination hotspot in the maize genome. *Plant Cell* 9: 1633-1646.
- DOVER, G. (1982) Molecular drive: a cohesive mode of species evolution. *Nature* 299: 111-117.

- DOVER, GA. (1986) Molecular drive in multigene families: how biological novelties arise, spread and are assimilated. *Trends Genet* 2:159-165.
- DROUIN, G. (1999) Expressed retrotransposed 5S rRNA genes in the mouse and rat genomes. *Genome* 43: 213-215.
- DROUIN, G., DOVER, GA. (1987) A plant processed pseudogene. *Nature* 328: 557-558.
- DUNHAM, I., SHIMIZU, N. ET AL. (1999) The DNA sequence of human chromosome 22. *Nature* 402: 489-495.
- DUPRESSION, A., HEIDMANN, T. (1996) Germ line-specific expression of intracisternal A-particle retrotransposons in transgenic mice. *Mol Cell Biol* 16: 4495-4503.
- EICKBUSH, TH. (1994) Origin and evolutionary relationships of retroelements. In: Morse, SS. (ed) *The evolutionary biology of viruses*, pp. 121-157. Raven Press, New York.
- EICKBUSH, TH. (1997) Telomerase and Retrotransposons: which came first? *Science* 277:911-912.
- ELLIS, TH., POYSER, SJ., KNOX, MR., VERSHININ, AV., AMBROSE, MJ. (1998) Polymorphism of insertion sites of *Ty1-copia* class retrotransposons and its use for linkage and diversity studies in pea. *Mol Gen Genet* 260: 9-19.
- FAHLESON, J., LAGERCRANTZ, U., MOURAS, A., GLIMELIUS, K. (1997) Characterization of somatic hybrids between *Brassica napus* and *Eruca sativa* using species-specific repetitive sequences and genomic in-situ hybridization. *Plant Science* 123: 133-142.
- FALQUET, J., CREUSOT, F., DRON, M. (1997) Molecular analysis of rDNA unit and characterization of a satellite DNA homologous to IGS subrepeats. *Plant Physiol Biochem* 35: 611-622.
- FAO (Food and Agricultural Organization of the United Nations Organization) (1988). *Quart Bull Statistics* 1: 50.
- FAO (Food and Agriculture Organization of the United Nations Organization) (1995) *Quart Bull Statistics* 8: 51,54.
- FEDOROFF, N. (1999) The *Suppressor-mutator* element and the evolutionary riddle of transposons. *Genes to Cells* 4: 11-19.
- FEDOROFF, N. (2000) Transposons and genome evolution in plants. *Proc. Natl. Acad. Sci. USA*, 97: 7002-7007.
- FEDOROFF, N., FURTEK, D. AND NELSON, O. (1984) Cloning of the *bronze* locus in maize by a simple and generalizable procedure using the transposable element *Activator* (Ac). *Proc Natl Acad Sci USA*, 81: 3825-3829.
- FEDOROFF, N., SCHLÄPPLI, M. AND RAINA, R. (1995) Epigenetic regulation of the maize *Spm* transposon. *Bio Essays* 17: 291- 297.

- FEINBERG, AP., VOGELSTEIN, B. (1983) A technique for radiolabeling DNA restriction endonuclease fragments th high specific activity. *Anal Biochem* 132, 6-13.
- FENG, Q., SCHUHMANN, G., BOEKE, JD. (1998) Retrotransposon *R1Bm* endonuclease cleaves the target sequence. *Proc Natl Acad Sci USA* 95: 2083-2088.
- FINNEGAN, DJ. (1997) Transposable elements: how non-LTR retrotransposons do it. *Curr Biol* 7: 245-248.
- FISCHER, TC., GRONER, S., ZENTGRAF, U., HEMLEBEN, V. (1994) Evidence for nucleosomal phasing and a novel protein specifically binding to cucumber satellite DNA. *Z Naturforsch* 49: 79-86.
- FLAVELL, RB. (1986) Repetitive DNA and chromosome evolution in plants. *Phil Trans R Soc Lond B* 312: 227-242.
- FLAVELL, A., SMITH, D., KUMAR, A. (1992b) Extreme heterogeneity of *Ty1-copia* group retrotransposons in plants. *Mol Gen Genet* 231: 233-242.
- FLAVELL, AJ. (1999) Long terminal repeat retrotransposons jump between species. *Proc Natl Acad Sci USA* 96: 12211-12212.
- FLAVELL, AJ., DUNBAR, E., ANDERSON, R., PEARCE, SR., HARTLEY, R., KUMAR, A. (1992a) *Ty1-copia* group retrotransposons are ubiquitous and heterogeneous in higher plants. *Nucl Acids Res* 20: 3639-3644.
- FLAVELL, AJ., KNOX, MR., PEARCE, SR., ELLIS, TH. (1998) Retrotransposon-based insertion polymorphisms (RBIP) for high throughput marker analysis. *Plant J* 16: 643-650.
- FRANSZ, PF., ALONSO-BLANCO, C., LIHARSKA, TB., PEETERS, AJM., ZABEL, P., DE JONG, J.H. (1996) High-resolution physical mapping in *Arabidopsis thaliana* and tomato by fluorescence *in situ* hybridization to extended DNA fibres. *Plant J* 9: 421-430.
- FREY, M., STETTNER, C., GIERL, A. (1998) A general method for gene isolation in tagging approaches: amplification of insertion mutagenized sites (AIMS). *Plant J* 13: 717-721.
- FUCHS, J., STREHL, S., BRANDES, A., SCHWEIZER, D., SCHUBERT, I. (1998) Molecular-cytogenetic characterization of the *Vicia faba* genome - heterochromatin differentiation, replication patterns and sequence localization. *Chrom Res* 6: 219- 230.
- GABRIEL, A., WILLEMS, M., MULES, EH., BOEKE, JD. (1996) Replication infidelity during a single cycle of *Ty1* retrotransposition. *Proc Natl Acad Sci USA* 93: 7767-7771.
- GALASSO, I., FREDIANI, M., MAGGIANI, M., CREMONINI, R., PIGNONE, D. (1996). Chromatin characterization by banding techniques, in-situ hybridization, and nuclear DNA content in *Cicer L.* (Leguminosae). *Genome* 39: 258-265.
- GALASSO, I., PIGNONE, D. (1992) Characterization of chickpea chromosomes by banding techniques. *Genet Res Crop Evol* 39: 115-119.

- GALASSO, I., SCHMIDT, T., PIGNONE, D. (2001) Identification of *Lens culinaris* ssp. *culinaris* chromosomes by physical mapping of repetitive DNA sequences. *Chrom Res* 9: 199-209.
- GANAL, M., HEMLEBEN, V. (1986) Different AT-rich satellite DNAs in *Cucurbita pepo* and *Cucurbita maxima*. *Theor Appl Genet* 73: 129-135.
- GARBNER, K., BILIC, I., PUSCH, O., TOHME, J., BACHMAIR, A., SCHWEIZER, D., JANTSCH, V. (1999) The *Tpv2* family of retrotransposons of *Phaseolus vulgaris*: structure, integration characteristics, and use for genotype classification. *Plant Mol Biol* 39: 797-807.
- Gaur, PM., SLINKARD, AE. (1990) Inheritance and linkage of isoenzyme coding genes in chickpea. *J Hered* 81: 455-461.
- GEORGIEV, GP. (1984) Mobile genetic elements in animal cells and their biological significance. *Eur J Biochem* 145: 203-220.
- GERSTNER, J., SCHIEBEL, K., VON WALDBURG, G., HEMLEBEN, V. (1988) Complex organization of the length heterogeneous 5' external spacer of mung bean (*Vigna radiata*) ribosomal DNA. *Genome* 30: 723- 733.
- GESTELAND, RF., ATKIN, JF. (1993) The RNA world. New York. Cold Spring Harbour Press.
- GIERL, A. (1996) The *En/Spm* transposable element of maize. In: Saedler, H. & Gierl, A (eds.) Transposable elements. Springer, pp. 145-159.
- GINDULLIS, F., DESEL, C., GALASSO, I., SCHMIDT, T. (2001) The large-scale organization of the centromeric region in *Beta* species. *Genome Res* 11: 253-265.
- GIRARD, L., FREELING, M. (1999) Regulatory changes as a consequence of transposon insertion. *Dev Genet* 25: 291-296.
- GORTNER, G., NENNO, M., WEISING, K., ZINK, D., NAGL, W., KAHL G. (1998) Chromosomal localization and distribution of simple sequence repeats and the *Arabidopsis*-type telomere sequence in the genome of *Cicer arietinum* L.. *Chrom Res* 6: 97-104.
- GOUBELY, C., ARNAUD, P., TATOUT, C., HESLOP-HARRISON, JS., DERAGON, JM. (1999) *Si* SINE retrotransposons are methylated at symmetrical positions in *Brassica napus*: identification of preferred target site for asymmetrical methylation. *J Mol Biol* 39: 243-255.
- GRANDBASTIEN, MA. (1998) Activation of plant retrotransposons under stress conditions. *Trends Plant Sci* 3:181-187.
- GRANDBASTIEN, MA., LUCAS, H., MOREL, JB., MHIRI, C., VERNHETTES, S., CASACUBERTA, JM. (1997) The expression of the tobacco *Tnt1* retrotransposon is linked to plant defense responses. *Genetica* 100: 241-252.
- GRANDBASTIEN, MA., SPIELMANN, A., CABOCHE, M. (1989) *Tnt1*, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics. *Nature* 377: 376-380.
- GRAY, Y. (2000) It takes two transposons to tango. *Trends Genet.* 16: 461-468.

- GREENER, A. (1990) *E.coli* SURE™ strain: clone “unclonable“ DNA. *Strategies* 3: 5-6.
- GRIFFON, BM., PEARCE, SR., KALENDAR, R., SCHULMAN, A., PAULIN, L., JACK, P., KUMAR, A., FLAVELL, AJ. (1999) Phylogeny and transpositional activity of *Ty1-copia* group retrotransposons in cereal genomes. *Mol Gen Genet* 261: 883-891.
- GRIESS, EA., GRASSER, KD., FEIX, D. (1993) Repeat units from maize rDNA external spacer region exhibit DNA curvature and interact with high-mobility-group proteins. *Planta* 191: 524- 531.
- HAGEMANN, S., SCHEER, B., SCHWEIZER, D. (1993) Repetitive sequences in the genome of *Anemone blanda*: identification of tandem arrays and of dispersed repeats. *Chromosoma* 102: 312-324.
- HANAHAH, D. (1983) Studies on transformation of *E. coli* with plasmids. *J Mol Biol* 166: 557-580.
- HANSEN, LJ., CHALKER, DL., SANDMEYER, SB. (1988) *Ty3*, a yeast retrotransposon associated with tRNA genes, has homology to animal retroviruses. *Mol Cell Biol* 8: 5245-5256.
- HANSON, RE., ISLAM-FARIDI, MN., CRANE, CF., ZWICK, MS., CZESCHIN, DG., WENDEL, JF., MCKNIGHT, TD., PRICE, HJ., STELLY, DM. (1999) *Ty1-copia* retrotransposon behaviour in a polyploid cotton. *Chrom Res* 8: 73-76.
- HARRISON, GE., HESLOP-HARRISON, JS. (1995) Centromeric repetitive DNA sequences in the genus *Brassica*. *Theor Appl Genet* 90: 157-165.
- HE, ZH., DONG, HT., DONG, JX., LI, DB., RONALD, PC. (2000) The rice *Rim2* transcript accumulates in response to *Magnaporthe grisea* and its predicted protein product shares similarity with *TNP2*-like proteins encoded by CACTA transposons. *Mol Gen Genet* 264: 2-10.
- HELBAEK, H. (1970) The plant husbandry at Hacillar. In: Excavation at Halicar, J. Mellaart (ed.) pp. 189-244. Edinburgh University Press.
- HELM, MA., HEMLEBEN, V. (1997) Characterization of a new prominent satellite DNA of *Cucumis metuliferus* and differential distribution of satellite DNA in cultivated and wild species of *Cucumis* and in related genera of Cucurbitaceae. *Euphytica* 94: 219-226.
- HEMLEBEN, V. (1993) Repetitive and highly repetitive DNA components as molecular markers for evolutionary studies and in plant breeding. *Curr Top Mol Genet (Life Sci Adv)* 1: 173- 185.
- HEMLEBEN, V., GRIERSON, D., DERTMANN, H. (1977) The use of equilibrium centrifugation in actinomycin-caesium chloride for the purification of ribosomal DNA. *Plant Sci Lett* 9: 129-135.
- HEMLEBEN, V., LEWEKE, B., ROTH, A., STADLER, I. (1982) Organization of highly repetitive satellite DNA of two Cucurbitaceae species (*Cucumis melo* and *Cucumis sativus*). *Nucl Acids Res* 10: 631-644.

- HENIKOFF, S. (1998) Reply. Trends Genet 14: 386- 387.
- HENIKOFF, S., COMAI, L. (1998) Trans-sensing effects. Cell 93: 329-332.
- HENK, A., WARREN, R., INNES, R. (1999) A new *Ac*-like transposon of *Arabidopsis* is associated with a deletion of the RPS5 disease resistance gene. Genetics 151: 1581-1589.
- HESLOP-HARRISON, JS., MURATA, M., OGURA, Y., SCHWARZACHER, T., MOTOYOSHI, F. (1999) Polymorphisms and genomic organization of repetitive DNA from centromeric regions of *Arabidopsis* chromosomes. Plant Cell 11: 31-42.
- HESLOP-HARRISON, JS. (1991) The molecular cytogenetics of plants. J Cell Sci 100: 15-21.
- HESLOP-HARRISON, JS., BRANDES, A., TAKETA, S., SCHMIDT, T., VERSHININ, AV., ALKHIMOVA, EG., KAMM, A., DOUDRICK, RL., SCHWARZACHER, T., KATSIOTIS, A., KUBIS, S., KUMAR, A., PEARCE, SR., FLAVELL, AJ., HARRISON, GE. (1997) The chromosomal distribution of *Tyl-copia* group retrotransposable elements in higher plants and their implications for genome evolution. Genetica 100: 197-204.
- HIGASHIYAMA, T., NOUTOSHI, Y., FUJIE, M., YAMADA, T. (1997) *Zepp*, a LINE-like retrotransposon accumulated in the *Chlorella* telomeric region. EMBO J 16: 3715-3723.
- HIGGINS, D., SHARP, P. (1988) CLUSTAL: a package for performing multiple sequence alignments on a microcomputer. Gene 73: 237-244.
- HIOM, K., MELEK, M., GELLERT, M. (1998) DNA transposition by the RAG1 and RAG2 proteins: a possible source of oncogenetic translocations. Cell 94: 463-470.
- HIROCHIKA, H. (1993) Activation of tobacco retrotransposons during tissue culture. EMBO J 12: 2521-2528.
- HIROCHIKA, H. (1997) Retrotransposons of rice: their regulation and use for genome analysis. Plant Mol Biol 35: 231-240.
- HIROCHIKA, H., OKAMOTO, H., KAKUTANI, T. (2000) Silencing of retrotransposons in *Arabidopsis* and reactivation by the *ddm1* mutation. Plant Cell 12: 357-368.
- HIROCHIKA, H., OTSUKI, H., YOSHIKAWA, M., OTSUKI, Y., SUGIMOTO, K., TAKEDA, S. (1996b) Autonomous transposition of the tobacco retrotransposon *Tto1* in rice. Plant Cell 8: 725-734.
- HIROCHIKA, H., SUGIMOTO, K., OTSUKI, Y., TSUGAWA, H., KANDA, M. (1996a) Retrotransposons of rice involved in mutations induced by tissue culture. Proc Natl Acad Sci USA 93: 7783-7788.
- HOUCK, M., CLARK, J., PETERSON, K., KIDWELL, M. (1991) Possible horizontal transfer of *Drosophila* genes by the mite *Proctolaelaps regalis*. Science 253: 1125-1128.
- HU, W., DAS, OP., MESSING, J. (1995) *Zeon-1*, a member of a new maize retrotransposon family. Mol Gen Genet 248: 471-480.

- HULTÉN, MA., STACEY, M. (1995) Does junk DNA regulate gene expression in humans? *J Clin Pathol Mol Pathol* 48: M118-M123.
- HÜTTEL, B. (1996) Mikrosatelliten in der Genomanalyse: Abundanz, Variabilität und Vererbung in der Kichererbse (*Cicer arietinum* L.). Dissertation, Frankfurt/M.
- ICARDA (International Centre for Agricultural Research in Dry Areas) (1990) Food legume improvement program. In: ICARDA Annual Report, Aleppo, Syria, pp.61-64.
- INAGAKI, Y., HISATOMI, Y., SUZUKI, T., KASAHARA, K., IIDA, S. (1994) Isolation of a *Suppressor-Mutator/Enhancer*-like transposable element, *Tpn1*, from Japanese Morning Glory bearing variegated flowers. *Plant Cell* 6: 375-383.
- INGHAM, LD., HANNA, WW., BAIER, JW., HANNAH, LC. (1993) Origin of the main class of repetitive DNA within selected *Pennisetum* species. *Mol Gen Genet* 238: 350-356.
- IRICK, H. (1994) A new function for heterochromatin. *Chromosoma* 103: 1-3.
- IWAMOTO, M., NAGASHIMA, H., NAGAMINE, T., HIGO, H. AND HIGO, K. (1999) A *Tourist* element in the 5' flanking region of the catalase gene *CatA* reveals evolutionary relationships among *Oryza* species with various genome types. *Mol Gen Genet* 262: 493-500.
- JÄÄSKELÄINEN, M., MYKKÄNEN, AH., ARNA, T., VICIENT, CM., SUONIEMI, A., KALENDAR, R., SAVILAHTI, H., SCHULMAN, AH. (1999) Retrotransposon *BARE1*: expression of encoded proteins and formation of virus-like particles in barley cells. *Plant J* 20: 413-422.
- JAKOWITSCH, J., PAPP, I., MATZKE, MA., MATZKE, AJ. (1998) Identification of a new family of highly repetitive DNA, NTS9, that is located predominantly on the S9 chromosome of tobacco. *Chrom Res* 6:649- 659.
- JEFFREYS, AJ., WILSON, V., THEIN, SL. (1985) Hypervariable "minisatellite" regions in human DNA. *Nature* 314: 67-73.
- JENSSEN, S., GASSAMA, MP., HEIDMANN, T. (1999) Taming of transposable elements by homology dependent gene silencing. *Nat Genet* 21: 94-96.
- JIANG, J., GILL, BS. (1994) Nonisotopic *in-situ* hybridization and plant genome mapping: the first 10 years. *Genome* 37: 717-725.
- JIANG, J., HULBERT, SH., GILL, BS., WARD, DC. (1996) Interphase fluorescence in situ hybridization mapping: A physical mapping strategy for plant species with large complex genomes. *Mol Gen Genet* 252: 497-502.
- JIMINÉZ-DIAZ, RM., CRINO, P., HALILA, MH., MASCONI, C., TRAPERO-CASAS, AT. (1993) Screening for resistance to fusarium wilt and ascochyta blight in chickpea. In: Singh KB, Saxena MC (eds.) *Breeding for stress tolerance in cool-season food legumes*. John Wiley and Sons, New York, pp 77-95.
- JIN, YK., BENNETZEN, JL. (1989) Structure and coding properties of *Bs1*, a maize retrovirus-like transposon. *Proc Natl Acad Sci USA* 86: 6235-6239.

- Jin, YK., Bennetzen, JL. (1994) Integration and non-random mutation of a plasma membrane proton ATPase gene fragment within the *BsI* retroelement of maize. *Plant Cell* 6: 1177-1186.
- JORDAN, IK., MATYUNINA, LV., McDONALD, JF. (1999) Evidence for the recent horizontal transfer of long terminal repeat retrotransposon. *Proc Natl Acad Sci USA* 96: 12621-12625.
- JUNGHANS, H.(1989) Charakterisierung genomspezifischer repetitiver DNA von *Hordeum vulgare*. Möglichkeiten des Einsatzes von repetitiven DNA Sonden und Restriktionslängenpolymorphismen in der Genomanalyse. Dissertation, Martin-Luther Universität, Halle-Wittenberg.
- KAISER, WJ. (1997) Inter- and intra-national spread of *Ascochyta* pathogens of chickpea, faba bean, and lentil. *Can J Plant Pathol* 19: 25–224.
- KALENDAR, R., GROB, T., REGINA, M., SUONIEMI, A., SCHULMAN, AH. (1999) IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques. *Theor Appl Genet* 98: 704-711.
- KALENDAR, R., TASKANEN, J., IMMONEN, S., NEVO, E., SCHULMANN, AH. (2000) Genome evolution of wild barley (*Hordeum spontaneum*) by *BARE1* retrotransposon dynamics in response to sharp micropclimatic divergence. *Proc Natl Acad Sci USA* 97: 6603-6607.
- KAMM, A., GALASSO, I., SCHMIDT, T., HESLOP-HARRISON, JS. (1995) Analysis of a repetitive DNA family from *Arabidopsis arenosa* and relationships between *Arabidopsis* species. *Plant Mol Biol* 27: 853-862.
- KAMM, A., SCHMIDT, T., HESLOP-HARRISON, JS. (1994) Molecular and physical organization of highly repetitive, undermethylated DNA from *Pennisetum glaucum*. *Mol Gen Genet* 244: 420-425.
- KAPITONOV, VV., JURKA, J. (1999) Molecular paleontology of transposable elements from *Arabidopsis thaliana*. *Genetica* 107: 27-37.
- KATO, A., YAKURA, K., TANIFUJI, S. (1985) Repeated DNA sequences found in the large spacer of *Vicia faba* rDNA. *Biochimica et Biophysica Acta* 825: 411-415.
- KATO, A., YAKURA, K., TANIFUJI, S. (1984) Sequence analysis of *Vicia faba* repeated DNA, the *FokI* repeat element. *Nucl Acids Res* 12: 6415-6426.
- KAZAN, K., MUEHLBAUER, FJ. (1991) Allozyme variation and phylogeny in annual species of *Cicer* (*Leguminosae*). *Pl Syst Evol* 175: 11-21.
- KAZAN, K., MUEHLBAUER, FJ., WEEDEN, NF., LADIZINSKY, G. (1993) Inheritance and linkage relationships of morphological and isozyme loci in chickpea (*Cicer arietinum* L.). *Theor Appl Genet* 86: 417-426.
- KAZAZIAN, HH. (1998) Mobile elements and disease. *Curr Opin Genet Dev* 8: 343-350.

- KENWARD, KD., BAI, D., BAN, MR., BRANDLE, JE. (1999) Isolation and characterization of *Tnd-1*, a retrotransposon marker linked to black root rot resistance in tobacco. *Theor Appl Genet* 98: 387-395.
- KIM, JM., VANGURI, S., BOEKE, JD., GABRIEL, A., VOYTAS, DF. (1998) Transposable elements and genome organization: a comparative survey of retrotransposons revealed by the complete *Saccharomyces cerevisiae* genome sequence. *Genome Res* 8: 464-478.
- KING, K., JOBST, J., HEMLEBEN, V. (1995) Differential homogenization and amplification of two satellite DNAs in the genus *Cucurbita* (Cucurbitaceae). *J Mol Evol* 41: 996- 1005.
- KISHIMA, Y., YAMASHITA, S., MARTIN, C., MIKAMI, T. (1999) Structural conservation of the transposon Tam3 family in *Antirrhinum majus* and estimation of the number of copies able to transpose. *Plant Mol Biol* 39: 299-308.
- KNOOP, V., UNSELD, M., MARIENFELD, J., BRANDT, P., SÜNKEL, S., ULLRICH, H., BRENNICKE, A. (1996) *Copia*-, *gypsy*- and LINE-like retrotransposon fragments in the mitochondrial genome of *Arabidopsis thaliana*. *Genetics* 142: 579-585.
- KOLCHINSKY, A., GRESSHOFF, PM. (1995) A major satellite DNA of soybean is a 92 base pairs tandem repeat. *Theor Appl Genet* 90: 621-626.
- KONIECZNY, A., VOYTAS, DF., CUMMINGS, MP., AUSUBEL, FM. (1991) A superfamily of *Arabidopsis thaliana* retrotransposons. *Genetics* 127: 801-809.
- KOSSACK, DS., KINLAW, CS. (1999) IFG, a *gypsy*-like retrotransposon in *Pinus* (Pinaceae), has an extensive history in pines. *Plant Mol Biol* 39: 417-426.
- KOURILSKY, P. (1986) Molecular mechanism for gene conversion in higher cells. *Trends Genet* 2: 60-63.
- KUBIS, S., HESLOP-HARRISON, JS., SCHMIDT, T. (1997) A family of differentially amplified repetitive DNA sequences in the genus *Beta* reveals genetic variation in *Beta vulgaris* subspecies and cultivars. *J Mol Evol* 44: 310- 320.
- KUBIS, S., SCHMIDT, T., HESLOP-HARRISON, JS. (1998) Repetitive DNA as a major component of plant genomes. *Annals of Bot (suppl.)* 82: 45-55.
- KUBIS, S., HESLOP-HARRISON, J., DESEL, C., SCHMIDT, T. (1998) The genomic organization of non-LTR retrotransposons (LINEs) from three *Beta* species and five other angiosperms. *Plant Mol Biol* 36: 821-831.
- KUMAR, A. (1996) The adventures of the *Ty1-copia* group of retrotransposons in plants. *Trends Genet* 12: 41-43.
- Kumar, A. (1998) The evolution of plant viruses: moving to green pastures. *Trends Plant Sci.* 3: 371-374.
- KUMAR, A., BENNETZEN, JL. (1999) Plant retrotransposons. *Ann Rev Genet* 33: 479-532.

- KUMAR, A., HIROCHIKA, H. (2001) Applications of retrotransposons as genetic tools in plant biology. *Trends Plant Sci* 6: 127-134.
- KUMAR, A., PEARCE, SR., MCLEAN, K., HARRISON, G., HESLOP-HARRISON, JS., WAUGH, R., FLAVELL, AJ. (1997) The *Ty1-copia* group of retrotransposons in plants: genomic organisation, evolution, and use as molecular markers. *Genetica* 100: 205-217.
- KUMEKAWA, N., OHTSUBO, H., HORIUCHI, T., OHTSUBO, E. (1999) Identification and characterization of novel retrotransposons of the *gypsy*-type in rice. *Mol Gen Genet* 260: 593-602.
- KUNZE, R., SAEDLER, H., LÖNNIG W. (1997) Plant transposable elements. *Adv Bot Res* 27: 332-470.
- KUWAHARA, A., KATO, A., KOMEDA, Y. (2000) Isolation and characterization of *copia*-type retrotransposons in *Arabidopsis thaliana*. *Gene* 244: 127-136.
- L'HOMME, Y., SEGUIN, A., TREMBLAY, FM. (2000) Different classes of retrotransposons in coniferous spruce species. *Genome* 43: 1084-1089.
- LABDI, M., ROBERTSON, LD., SINGH, KB., CHARRIER, A. (1996) Genetic diversity and phylogenetic relationships among the annual *Cicer* species as revealed by isozyme polymorphism. *Euphytica* 88: 181-188.
- LADIZINSKY, G., ADLER, A. (1976b) Genetic relationships among the annual species of *Cicer* L.. *Theor Appl Genet* 48: 197-203.
- LADIZINSKY, G., ADLER, A. (1976a) The origin of chickpea *Cicer arietinum* L.. *Euphytica* 25: 211-217.
- LANDER, E., GREEN, P., ABRAHAMSON, J., BARLOW, A., DALY, M., LINCOLN, S., NEWBURG, L. (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1: 174-181.
- LANGDON, T., SEAGO, C., MENDE, M., LEGGETT, M., THOMAS, H., FORSTER, JW., THOMAS, H., JONES, RN., JENKINS, G. (2000a) Retrotransposon evolution in diverse plant genomes. *Genetics* 156: 313-325.
- LANGDON, T., SEAGO, C., JONES, RN., OUGHAM, H., THOMAS, H., FORSTER, JW., JENKINS, G. (2000b) *De Novo* evolution of satellite DNA on the rye B chromosome. *Genetics* 154: 869-884.
- LATEN, HM., MAJUMDAR, A., GAUCHER, EA. (1998) SIRE1, a *copia/Ty1*-like retroelement from soybean, encodes a retroviral envelope-like protein. *Proc Natl Acad Sci USA* 95: 6897-6902.
- LE, Q., WRIGHT, S., YU, Z., BUREAU T. (2000) Transposon diversity in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA*, 97: 7376-7381.
- LEETON PRJ., SMYTH, DR. (1993) An abundant LINE-like element amplified in the genome of *Lilium speciosum*. *Mol Gen Genet* 237: 97-104.

- LEVINSON, G., GUTMAN, GA. (1987) Slipped strand mispairing: a major mechanism for DNA sequence evolution. *Mol Biol Evol* 4: 203-221.
- LEVY, AA., WALBOT, V. (1990) Regulation of the timing of transposable element excision during maize development. *Science* 248: 1534-1537.
- LEWIN, B. (2000) *Genes VII*. Oxford University press, New York, pp. 485-506.
- LIN, XY., KAUL, SS. ET AL. (1999) Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. *Nature* 402: 761-768.
- LINARES, C., SERNA, A., FOMINAYA, A. (1999) Chromosomal organisation of a sequence related to LTR-like elements of *Ty1-copia* retrotransposons in *Avena* species. *Genome* 42: 706-713.
- LITT, M., LUTY, JA. (1989). A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am J Hum Genet* 44: 397-401.
- LUCK, J., LAWRENCE, G., FINNEGAN, E., JONES, D., ELLIS, G. (1998) A flax transposon identified in two spontaneous mutant alleles of the *L6* rust resistance gene. *Plant J.* 16: 365-369.
- LUKASZEWSKI, AJ. (1992) A comparison of physical distribution of recombination in chromosome 1R in diploid rye and in hexaploid triticale. *Theor Appl Genet* 83: 1048-1053.
- MACAS, J., POZÁKOVÁ, D., NAVRÁTILOVÁ, A., NOUZOVÁ, M., NEUMANN, P. (2000) Two new families of tandem repeats isolated from the genus *Vicia* using genomic self-priming PCR. *Mol Gen Genet* 263: 741- 751.
- MAES, T., DE KEUKELEIRE, P., GERATS, T. (1999) Plant tagnology. *Trends Plant Sci.* 4: 90-96.
- MAGGINI, F., CREMONINI, R., ZOLFINO, C., TUCCI, GF., D'OVIDIO, R., DELRE, V., DEPACE, C., SCARASCIA MUGNOZZA, GT., CIONINI, PG. (1991) Structure and chromosomal localization of DNA sequences related to ribosomal subrepeats in *Vicia faba*. *Chromosoma* 100: 229-234.
- MANDRIOLI, M. (2000) *Mariner*-like transposable elements are interspersed within the rDNA-associated heterochromatin of the pufferfish *Tetraodon fluviatilis* (Osteichthyes). *Chrom Res* 8: 177-179.
- MANNINEN, I., SCHULMANN, AH. (1993) *BARE1*, a *copia*-like retroelement in barley (*Hordeum vulgare* L.) *Plant Mol Biol* 22: 829-846.
- MANNINEN, O., KALENDAR, R., ROBINSON, J., SCHULMAN, AH. (2000) Application of *BARE1* retrotransposon markers to the mapping of a major resistance gene for net blotch in barley. *Mol Gen Genet* 264: 325-334.
- MARILLONNET, S., WESSLER, SR. (1997) Retrotransposon insertion into the maize *waxy* gene results in tissue-specific RNA processing. *Plant Cell* 9: 967-978.

- MARIN, I., LLORENS, C. (2000) *Ty3-gypsy* retrotransposons: description of new *Arabidopsis thaliana* elements and evolutionary perspectives derived from comparative genomic data. *Mol Biol Evol* 17: 1040-1049.
- MARTIENSSSEN, R. (1998) Transposons, DNA methylation and gene control. *Trends Genet* 14: 263-264.
- MARTINEZ-ZAPATER, JM., ESTELLE, MA., SOMMERVILLE, CR. (1986) A highly repeated DNA sequence in *Arabidopsis thaliana*. *Mol Gen Genet* 204: 417-423.
- MARTINEZ-IZQUIERDO, JA., GARCIA- MARTINEZ, J., VICIENT, CM. (1997) What makes *Grande1* retrotransposon different? *Genetica* 100: 15-28.
- MATSUOKA, Y., TSUNEWAKI, K. (1999) Evolutionary dynamics of *Ty1-copia* group retrotransposons in grass shown by reverse transcriptase domain analysis. *Mol Biol Evol* 16: 208-217.
- MATÝASEK, R., FULNECEK, J., FAJUJKS, J., BEZDEK, M. (1996) Evidence for a sequence-directed conformation periodicity in the genomic highly repetitive DNA detectable with single strand specific chemical probe potassium permanganate. *Chrom Res* 4: 340-349.
- MATZKE, AJM., MATZKE, MA. (1998) Position effects and epigenetic silencing of plant transgenes. *Curr Opin Plant Biol* 1: 142-148.
- MATZKE, MA., METTE, MF., AUFSATZ, W., JAKOWITSCH, J., MATZKE, AJM. (1999) Host defenses to parasitic sequences and the evolution of epigenetic control mechanisms. *Genetica* 107: 271-287.
- MATZKE, MA., METTE, MF., MATZKE, AJM. (2000) Transgene silencing by the host genome defense: implications for the evolution of epigenetic control mechanisms in plants and vertebrates. *Plant Mol Biol* 43: 401-415.
- MAY, B., DELLAPORTA, S. (1998) Transposon sequences drive tissue-specific expression of the maize regulatory gene *R-s*. *Plant J* 13: 241-247.
- MAYER, K., SCHULLER, C., THE EUROPEAN UNION ARABIDOPSIS SEQUENCING CONSORTIUM AND THE COLD SPRING HARBOUR, WASHINGTON UNIVERSITY IN ST. LOUIS AND PE BIOSYSTEMS ARABIDOPSIS SEQUENCING CONSORTIUM (1999) Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*. *Nature* 402: 769-777.
- MCALLISTER, BF., WERREN, JH. (1999) Evolution of tandemly repeated sequences: what happens at the end of an array? *J Mol Evol* 48: 469-481.
- MCCLINTOCK, B. (1954) Mutations in maize and chromosomal aberrations in *Neurospora*. *Carnegie Institution of Washington Year Book* 53: 254-260.
- MCDONALD, JF. (1998) Transposable elements, gene silencing and macroevolution. *Trends Ecol Evol* 13: 94-95.

- MCDONALD, JF., MATYUNINA, LV., WILSON, S., JORDAN, IK., BOWEN, NJ., MILLER, WJ. (1997) LTR-retrotransposons and the evolution of eukaryotic enhancers. *Genetica* 100: 3-13.
- MEINKOTH, J., WAHL, G. (1984) Hybridization of nucleic acids immobilized on solid supports. *Anal Biochem* 138: 267-284.
- MELCHINGER, AE. (1990) Use of molecular markers in breeding for oligogenic disease resistance. *Plant Breed* 104:1-19.
- MHIRI, C., DEWIT, PJ., GRANDBASTIEN, MA. (1999) Activation of the promotor of the *Tnt1* retrotransposon in tomato after inoculation with the fungal pathogen *Cladosporium fulvum*. *Mol Plant-Microbe Interact* 12: 592- 603.
- MHIRI, C., MOREL, JB, VERNHETTES, S., CASACUBERTA, JM., LUCAS, H., GRANDBASTIEN, MA. (1997) The promotor of the tobacco *Tnt1* retrotransposon is induced by wounding and by abiotic stress. *Plant Mol Biol* 33: 257-266.
- MICHAELS, SD., JOHN, MC., AMASINO, RM. (1994) Removal of polysaccharides from plant DNA by ethanol precipitation. *BioTechniques* 17, 274-276.
- MILLER, JT., DONG, F., JACKSON, FA., SONG, J., JIANG, J. (1998) Retrotransposon-related DNA sequences in the centromeres of grass chromosomes. *Genetics* 150: 1615-1623
- MILLER, W., MCDONALD, J., PINSKER W. (1997) Molecular domestication of mobile elements. *Genetica* 100: 261-270.
- MILLER, WJ., MCDONALD, JF., NOUAUD, D., ANXOLABÉHÈRE, D. (1999) Molecular domestication- more than a sporadic episode in evolution. *Genetica* 107: 197-207.
- MIURA, A., YONEBAYASHI, S., WATANABE, K., TOYAMA, T., SHIMADA, H., KAKUTANI, T. (2001) Mobilization of transposons by a mutation abolishing full DNA methylation in *Arabidopsis*. *Nature* 411: 212-214.
- MOCHIZUKI, K., UMEDA, M., OHTSUBO, H., OHTSUBO, E. (1992) Characterization of a plant SINE: *pSINE1*, in rice genomes. *Japan J Genet* 57: 155-166.
- MOGEN, BD., MACDONALD, MH., GRAYBOSCH, R., HUNT, AG. (1990) Upstream sequences other than AAUAAA are required for an efficient messenger RNA 3'-end formation in plants. *Plant Cell* 2: 1261-1272.
- MOORE, G., BATTY, N., FLAVELL, R. (1991) A family of retrotransposons and associated genomic variation in wheat. *Genomics* 10: 461-468.
- MORAN, JV., DEBERARDINIS, RJ., KAZAZIAN, HH. (1999) Exon shuffling by *L1* retrotransposition. *Science* 283: 1530-1534.
- MOREAU-MHIRI, C., MOREL, JB., AUDEON, C., FERAULT, M., GRANDBASTIEN, MA., LUCAS, H. (1996) Regulation of expression of the tobacco *Tnt1* retrotransposon in heterologous species following pathogen-related stresses. *Plant J* 9: 409-419.

- MORGANTE, M., JURMAN, I., SHI, L., ZHU, T., KEIM, P., RAFALSKI, JA. (1997) The STR120 satellite DNA of soybean: organization, evolution and chromosomal specificity. *Chrom Res* 5: 363-373.
- MOTOHASHI, R., OHTSUBO, E., OHTSUBO, H. (1996) Identification of *Tnr3*, a *Suppressor-Mutator/Enhancer*-like transposable element from rice. *Mol Gen Genet* 250: 148-152.
- MOUNT, SM., RUBIN, GM. (1985) Complete nucleotide sequence of the *Drosophila* transposable element *copia*: homology between *copia* and retroviral proteins. *Mol Cell Biol* 5: 1630-1638.
- MUNOZ, FJ., DOPICO, B., LABRADOR, E. (1998) Isolation and characterization of a cDNA sequence CanSTM3 (AJ001901) from *Cicer arietinum* L. epicotyls (PGR-069). *Plant Physiol* 116: 1605.
- NACKEN, W., PIETROWIAK, R., SAEDLER, H., SOMMER, H. (1991) The transposable element *Tam1* from *Antirrhinum majus* shows structural homology to the maize transposon *En/Spm* and has no sequence specificity of insertion. *Mol Gen Genet* 228: 201-208.
- NAGAKI, K., TSUJIMOTO, H., ISONO, K., SASAKUMA, T. (1995) Molecular characterization of a tandem repeat, *Afa* family, and its distribution among Triticeae. *Genome* 38: 479- 486.
- NAGAKI, K., TSUJIMOTO, H., SASAKUMA, T. (1998) A novel repetitive sequence of sugar cane, SCEN family, locating on centromeric regions. *Chrom Res* 6: 295-302.
- NENE, YL., REDDY, MV., HAWARE, MP., GHANEKAR, AM., AMIN, KS. (1991) Field diagnosis of chickpea diseases and their control. In: ICRISAT Information Bulletin no. 28, Andhra Pradesh, India, pp.1-6.
- NOMA, K., OHTSUBO, E., OHTSUBO, H. (1999) Non-LTR retrotransposons (LINEs) as ubiquitous components of plant genomes. *Mol Gen Genet* 261: 71-79.
- NONOMURA, KL., KURATA, N. (1999) Organization of the 1.9 kb repeat unit RCE1 in the centromeric region of rice chromosomes. *Mol Gen Genet* 261: 1-10.
- NOUTOSHI, Y., ARAI, R., FUJIE, M., YAMADA, T. (1998) Structure of the *Chlorella Zepp* retrotransposon: *Zepp* clusters in the genome. *Mol Gen Genet* 259: 256-263.
- NOZUE, K., KANEGAE, T., WADA, M. (1997) A full length *Ty3/gypsy*-type retrotransposon in the fern *Adiantum*. *J Plant Res* 110: 495-499.
- O'NEILL, RJW., O'NEILL, MJ., GRAVES, MA. (1998) Undermethylation associated with retroelements activation and chromosomes remodelling in an interspecific mammalian hybrid. *Nature* 393: 68-72.
- OCAMPO, B., VENORA, G., ERRICO, A., SINGH, KB., SACCARDO, F. (1992) Karyotype analysis in the genus *Cicer*. *J Genet Breed* 46: 229-240.
- OHRI, D., PAL, M. (1991) The origin of chickpea (*Cicer arietinum* L.): karyotype and nuclear DNA amount. *Heredity* 66: 367-372.

- OHTSUBO, H., OHTSUBO, E. (1994) Involvement of transposition in dispersion of tandem repeat sequences (*TrsA*) in rice genomes. *Mol Gen Genet* 245: 449-455.
- OKAMOTO, H., HIROCHIKA, H. (2000) Efficient insertion mutagenesis of *Arabidopsis* by tissue culture-induced activation of the tobacco retrotransposon *Tto1*. *Plant J* 23: 291-304.
- OKAZAKI, S., ISHIKAWA, H., FUJIWARA, H. (1995) Structural analysis of *TRAS1*, a novel family of telomeric repeat-associated retrotransposons in the silkworm, *Bombyx mori*. *Mol Cell Biol*: 15: 4545-4552.
- OKUMURA, K., KIYAMA, R., OISHI, M. (1987) Sequence analysis of extrachromosomal *Sau3a* and related family DNA: analysis of recombination in the excision event. *Nucl Acids Res* 15: 7477-7489.
- ORGEL, T., CRICK, FHC. (1980) Selfish DNA: the ultimate parasite. *Nature* 284: 604-607.
- OZEKI, Y., DAVIES, E., TAKEDA, J. (1997) Somatic variation during long term subculturing of plant cells caused by insertion of a transposable element in a phenylalanine ammonia lyase (PAL) gene. *Mol Gen Genet* 254: 407-416.
- PAGE, R. (1996) Treeview: An application to display phylogenetic trees on personal computers. *Computer applications in Biosciences* 12: 357-358.
- PARDUE, ML., DANILEVSKAYA, ON., TRAVERSE, KL., LOWENHAUPT, K. (1997) Evolutionary links between telomeres and transposable elements. *Genetica* 100: 73-84.
- PAULS, PK., KUNERT, K., HUTTNER, E., GRANDBASTIEN, MA. (1994) Expression of the tobacco *Tnt1* retrotransposon promoter in heterologous species. *Plant Mol Biol* 26: 393-402.
- PEARCE, SR., HARRISON, G., HESLOP-HARRISON, JS., FLAVELL AJ., KUMAR, A. (1997) Characterization and genomic organization of *Ty1-copia* group retrotransposons in rye (*Secale cereale*). *Genome* 40: 617-625.
- PEARCE, SR., HARRISON, G., LI, D., HESLOP-HARRISON, JS., KUMAR, A., FLAVELL, AJ. (1996a) The *Ty1-copia* retrotransposons in *Vicia* species: copy number, sequence heterogeneity and chromosomal localization. *Mol Gen Genet* 250: 305-315.
- PEARCE, SR., KNOX, M., ELLIS, TH., FLAVELL, AJ., KUMAR, A. (2000) Pea *Ty1-copia* group retrotransposons: transpositional activity and use as markers to study genetic diversity in *Pisum*. *Mol Gen Genet* 263: 898-907.
- PEARCE, SR., PICH, U., HARRISON, G., FLAVELL, AJ., HESLOP-HARRISON, JS., ET AL. (1996b) The *Ty1-copia* group retrotransposons of *Allium cepa* are distributed throughout the chromosomes but are enriched in the terminal heterochromatin. *Chrom Res* 4: 357-364.
- PEARCE, SR., STUART-ROGERS, CM., KUMAR, A., ELLIS, THN., FLAVELL, AJ. (1999) Rapid isolation of plant *Ty1-copia* group retrotransposon LTR sequences for molecular marker studies. *Plant J* 19: 711-717.
- PELISSIER, T., TUTOIS, S., DERAGON, JM., TOURMENTE, S., GENESTIER, S., PICARD, G. (1995) *Athila*, a new retroelement from *Arabidopsis thaliana*. *Plant Mol Biol* 29: 441-452.

- PELISSIER, T., TUTOIS, S., TOURMENTE, S., DERAGON, JM., PICARD, G. (1996) DNA regions flanking the major *Arabidopsis thaliana* satellite are principally enriched in *Athila* retroelement sequences. *Genetica* 97: 141-151.
- PEREIRA, A., CUYPERS, H., GIERL, A., SCHWARZ-SOMMER, Z., SAEDLER, H. (1986) Molecular analysis of the *En/Spm* transposable element system in *Zea mays*. *EMBO J.* 5: 835-841.
- PETERSON, P.A. (1953) A mutable pale green locus in maize. *Genetics* 45: 115-133.
- PETERSON-BURCH, BD., WRIGHT, DA., LATEN, HM., VOYTAS, DF. (2000) Retroviruses in plants? *Trends Genet* 16: 151-152.
- POUTEAU, S., BOCCARA, M., GRANDBASTIEN, MA. (1994) Microbial elicitors of plant defense responses activate transcription of a retrotransposon. *Plant J* 5: 535-542.
- POUTEAU, S., HUTTNER, E., GRANDBASTIEN, MA., CABOCHE, M. (1991) Specific expression of the tobacco *Tnt1* retrotransposon in protoplasts. *EMBO J* 10: 1911-1918.
- PRESTING, GG., MALYSHEVA, L., FUCHS, J., SCHUBERT, I. (1998) A *Ty3/gypsy* retrotransposon-like sequence localizes to the centromeric regions of cereal chromosomes. *Plant J* 16: 721- 728.
- PURUGGANAN, MD., WESSLER, SR. (1994) Molecular evolution of *magellan*, a maize *Ty3-gypsy*-like retrotransposon. *Proc Natl Acad Sci USA* 91:11674-11678.
- PROVAN, J., THOMAS, WTB., FORSTER, BP., POWELL, W. (1999) *Copia*-SSR: A simple marker technique which can be used on total genomic DNA. *Genome* 42: 363-366.
- PURUGGANAN, MD., WESSLER, SR. (1995) Transposon signatures: species-specific molecular markers that utilize a class of multi-copy nuclear DNA. *Mol Ecol* 4: 265-269.
- RADIC, MZ., LUNDGREN, K., HAMKALO, BA. (1987) Curvature of mouse satellite DNA and condensation of heterochromatin. *Cell* 50: 1101-1108.
- RATNAPARKHE, MB., SANTRA, DK., TULLU, A., MUEHLBAUER, FJ. (1998a) Inheritance of inter-simple-sequence polymorphisms and linkage with a fusarium wilt resistance gene in chickpea. *Theor Appl Genet* 96: 348–353.
- RATNAPARKHE MB, TEKEOGLU M, MUEHLBAUER FJ (1998b) Inter-simple-sequence-repeat (ISSR) polymorphisms are useful for finding markers associated with disease resistance gene clusters. *Theor Appl Genet* 97: 515–519.
- RHODES, P., VODKIN, L. (1988) Organization of the *Tgm* family of transposable elements in soybean. *Genetics* 120: 597-604.
- ROBERTSON, H., LAMPE D. (1995) Recent horizontal transfer of a *mariner* transposable element among and between Diptera and Neuroptera. *Mol Biol Evol* 12: 850-862.

- ROSE, T., SCHULTZ, E., HENIKOFF, J., PIETROKOVSKI, S., MCCALLUM, C., HENIKOFF, S. (1998) Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences. *Nucl Acids Res* 26: 1628-1635.
- ROTH, DB., CRAIG, NL. (1998) VDJ Recombination: a transposase goes to work. *Cell* 94: 411-414.
- ROYO, J., NASS, N., MATTON, DP., OKAMOTO, S., CLARKE, AE., NEWBIGIN, E. (1996) A retrotransposon-like sequence linked to the S-locus of *Nicotiana glauca* is expressed in styles in response to touch. *Mol Gen Genet* 250: 180-188.
- SAEDLER, H., NEVERS, P. (1985) Transposition in plants: a molecular model. *EMBO J* 4: 585-590.
- SAGHAI-MAROOF, M., SOLIMAN, K., JORGENSEN, R., ALLARD, R. (1984) Ribosomal DNA spacer-length polymorphism in barley: Mendelian inheritance, chromosomal localization and population dynamics. *Proc Natl Acad Sci USA*, 81: 8014-8018.
- SAITOU, N., NEI, M. (1987). The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4: 406-425.
- SAKAMOTO, K., OHMIDO, N., FUKUI, K., KAMADA, H., SATOH, S. (2000) Site-specific accumulation of a LINE-like retrotransposon in a sex chromosome of the dioecious plant *Cannabis sativa*. *Plant Mol Biol* 44: 723-732.
- SAMBROOK, J., FRITSCH, E.F., MANIATIS, T. (1989). *Molecular cloning - a laboratory manual*. Cold Spring Harbor laboratory Press, New York, USA.
- SANGER, F., MICKLEN, S., COULSEN, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Sci USA* 74: 5463-5467.
- SANMIGUEL, P., BENNETZEN, JL. (1998) Evidence that a recent increase in maize genome size as caused by the massive amplification of intergene retrotransposons. *Ann Bot* 81: 37-44.
- SANMIGUEL, P., BRANDON, SG., TIKHONOV, A., NAKAJIMA, Y., BENNETZEN, JL. (1998) The paleontology of intergene retrotransposons of maize. *Nat Genet* 20: 43-45
- SANMIGUEL, P., TIKHONOV, A., JIN, YK., MOTCHOULSKAIA, N., ZAKHAROV, D., MELAKE-BERHAN, A., SPRINGER, PS., EDWARDS, KJ., LEE, M., AVRAMOVA, Z., BENNETZEN, JL. (1996) Nested retrotransposons in the intergenic regions of the maize genome. *Science* 274: 765-768.
- SANT, VJ., SAINANI, MN., SAMI-SUBBU, R., RANJEKAR, PK., GUPTA, VS. (2000) *Ty1-copia* retrotransposon-like elements in the chickpea genome: their identification, distribution and use for diversity analysis. *Gene* 257: 157-166.
- SANTRA, DK., TEKEOGLU, M., RATNAPARKHE, M., KAISER, WJ., MUEHLBAUER, FJ. (2000) Identification and mapping of QTLs conferring resistance to *Ascochyta blight* in chickpea. *Crop Science* 40: 1606-1612.
- SAXENA, MC., SINGH, KB. (1987) *The chickpea*. CAB International, Wallingford, UK

- SCHIEBEL, K., VON WALDBURG, G., GERSTNER, J., HEMLEBEN, V. (1989) Termination of transcription of ribosomal RNA genes of mung bean occurs with a 175 bp repetitive element of the spacer region. *Mol Gen Genet* 218: 302- 307.
- SCHLUTER, SF., BERNSTEIN, RM., BERNSTEIN, H., MARCHALONIS, JJ. (1999) „Big Bang“ emergence of the combinatorial immune system. *Dev Comp Immunol* 23: 107-111.
- SCHMIDT, T., KUDLA, J. (1996) The molecular structure, chromosomal organization and interspecific distribution of a family of tandemly repeated DNA sequences of *Antirrhinum majus* L.. *Genome* 39: 243- 248.
- SCHMIDT, T., HESLOP-HARRISON, JS. (1996) High-resolution mapping of repetitive DNA by in-situ hybridization: molecular and chromosomal features of prominent dispersed and discretely localized DNA families from the wild beet species *Beta procumbens*. *Plant Mol Biol* 30: 1099-1119.
- SCHMIDT, T., HESLOP-HARRISON, JS. (1993) Variability and evolution of two satellite DNAs in the genus *Beta*. *Genome* 36: 1074-1079.
- SCHMIDT, T., JUNG, C., METZLAFF, M. (1991) Distribution and evolution of two satellite DNAs in the genus *Beta*. *Theor Appl Genet* 82: 793-799.
- SCHMIDT, T., JUNGHANS, H., METZLAFF, M. (1990) Construction of *Beta procumbens*-specific DNA probes and their application for the screening of *B.vulgaris* x *B.procumbens* (2n=19) addition lines. *Theor Appl Genet* 79: 177-181.
- SCHMIDT, T., SCHWARZACHER, T., HESLOP-HARRISON, JS. (1994) Physical mapping of rRNA genes by fluorescent in-situ hybridization and structural analysis of 5S rRNA genes and intergenic spacer sequences in sugar beet (*Beta vulgaris*). *Theor Appl Genet* 88: 629-636.
- SCHMIDT, T. (1999) LINEs, SINEs and repetitive DNA: non-LTR retrotransposons in plant genomes. *Plant Mol Biol* 40: 903-910.
- SCHMIDT, T., HESLOP-HARRISON, JS. (1998) Genomes, genes and junk: the large-scale organization of plant chromosomes. *Trends Plant Sci* 3: 195-199.
- SCHMIDT, T., KUBIS, S., HESLOP-HARRISON, JS. (1995) Analysis and chromosomal localization of retrotransposons in sugar beet (*Beta vulgaris* L.): LINEs and *Ty1-copia* elements as major components of the genome. *Chrom Res* 3: 335-345.
- SCHMIDT, T., SCHWARZACHER, T., HESLOP-HARRISON, J.S. (1994) Physical mapping of rRNA genes by fluorescent in-situ hybridization and structural analysis of 5S rRNA genes and intergenic spacer sequences in sugar beet (*Beta vulgaris*). *Theor Appl Genet* 88: 629-636.
- SCHMIDT, ER. (1984) Clustered and interspersed repetitive DNA sequence family of *Chironomus*. The nucleotide sequence family of *Cla*-elements and various flanking sequences. *J Mol Biol* 178: 1-15.
- SCHUBERT, I. (1998) Late-replicating satellites: something for all centromeres? *Trends in Genomics* 14: 385- 386.

- SCHWARZACHER, T., LEITCH, AR. (1994) Enzymatic treatment of plant material to spread chromosomes for in-situ hybridization. In Isaac PG (ed) Protocols for nucleic acid analysis by non-radioactive probes, Methods in Molecular Biology 28: 153-160, Humana Press Inc., Totowa, NJ.
- SCHWARZACHER, T. (1996) The physical organization of Triticeae chromosomes. In: JS. Heslop-Harrison (ed.), Unifying plant genomes, Symposia of the Society for Experimental Biology, 50: 71-75. The Company of Biologists Ltd, Cambridge.
- SCHWARZ-SOMMER, Z., LECLERCQ, L., GOEBEL, E., SAEDLER, H. (1987) *Cin4*, an insert altering the structure of the A1 gene in *Zea mays*, exhibits properties of non-viral retrotransposons. EMBO J 13: 3873-3880.
- SCHWEIZER, G., GANAL, M., NINNEMANN, H., HEMLEBEN, V. (1988) Species-specific DNA sequences for identification of somatic hybrids between *Lycopersicon esculentum* and *Solanum acaule*. Theor Appl Genet 75: 679- 684.
- SCHWEIZER, D. (1979) Fluorescent chromosome banding in plants: applications, mechanisms, and implications for chromosome structure. Proc Fourth John Innes Symp., Norwich, U.K., pp. 61-72.
- SELINGER, DA., CHANDLER, VL. (2001) B-Bolivia, an allele of the maize *b1* gene with variable expression, contains a high copy retrotransposon-related sequence immediately upstream. Plant Physiol 125: 1363-1379.
- SELKER, EU., CAMBARERI, EB., JENSEN, BC., HAACK, KR (1987) Rearrangement of duplicated DNA inspecialized cells of *Neurospora*. Cell 51: 741-752.
- SENTRY, JW. SMYTH, DR. (1989) An element with long terminal repeats and its variant arrangements in the genome of *Lilium henryi*. Mol Gen Genet 215: 349-354.
- SHARMA, PC., HÜTTEL, B., WINTER, P., KAHL, G., GARDNER, RC., WEISING, K. (1995) The potential of microsatellites for hybridization- and PCR-based fingerprinting of chickpea (*Cicer arietinum* L.) and related species. Electrophoresis 16: 1755-1761.
- Sharma, PC., Winter, P., Bünger, T., Hüttel, B., Weigand, F., Weising, K., Kahl, G. (1995) Abundance and polymorphism of di-, tri- and tetranucleotide tandem repeats in chickpea (*Cicer arietinum* L.). Theor Appl Genet 90: 90-96.
- SHARMA, PC., GUPTA, PK. (1986) Cytogenetics of legume genera *Cicer* L. and *Lens* L. In: genetics and Crop Improvement. pp.321-340, Gupta, PK., Bahl, JR. (eds.), Rastogi and Co, Meerut, India.
- SHCHERBAN, AB., VAUGHAN, DA., TOMOOKA, N. (2000) Isolation of a new retrotransposon-like DNA sequence and its use in analysis of diversity within the *Oryza officinalis* complex. Genetica 108: 145-154.
- SHIRSAT, A. (1988) A transposon-like structure in the 5' flanking sequence of the legumin gene from *Pisum sativum*. Mol Gen Genet 212: 129-133.

- SIMON, C.J., MUEHLBAUER, F.J. (1997) Construction of a chickpea linkage map and its comparison with maps of pea and lentil. *J Hered* 88: 115-119.
- SINGER, M., BERG, P. (1992) *Gene und Genome*, Spektrum, Heidelberg.
- SINGH, K.B., HOLLY, L., REDDY, M.V. (1991) A catalog of Kabuli chickpea germplasm. ICARDA, p. 398.
- SINGH, K.B., OCAMPO, B. (1993) Interspecific hybridization in annual *Cicer* species. *J Genet Breed* 47: 199-204.
- SINGH, K.B., OCAMPO, B. (1997) Exploitation of wild *Cicer* species for yield improvement in chickpea. *Theor Appl Genet* 95: 418-423.
- SINGH, K.B., OCAMPO, B., ROBERTSON, L.D. (1998) Diversity for abiotic and biotic stress resistance in the wild annual *Cicer* species. *Genet Res Crop Evol* 45: 9-17.
- SMITH, G.P. (1976) Evolution of repeated DNA sequences by unequal crossover. *Science* 191: 528-535.
- SMYTH, D.R. (1991) Dispersed repeats in plant genomes. *Chromosoma* 100: 355-359.
- SMYTH, D.R., KALITSIS, P., JOSEPH, J.L., SENTRY, J.W. (1989) Plant retrotransposon from *Lilium henryi* is related to *Ty3* of yeast and the *gypsy* group of *Drosophila*. *Proc Natl Acad Sci USA* 86: 5015-5019.
- SNOWDEN, K., NAPOLI, C. (1998) *PsI*: a novel *Spm*-like transposable element from *Petunia hybrida*. *Plant J* 14: 43-54.
- SONG, W.Y., PI, L.Y., BUREAU, T.E., RONALD, P.C. (1998) Identification and characterization of 14 transposon-like elements in the noncoding regions of members of the *Xa21* family of disease resistance genes in rice. *Mol Gen Genet* 258: 449-456.
- SOUTHERN, E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* 98: 503-517.
- SPEULMANN, E., METZ, P.L.J., VAN ARKEL, G., TELINTEL HEKKERT, B., STIEKEMA, W.J., PEREIRA, A. (1999) A two-component *Enhancer-Inhibitor* transposon mutagenesis system for functional analysis of the *Arabidopsis* genome. *Plant Cell* 11: 1853-1866.
- SPRINZL, M., HARTMANN, T., WEBER, J., BLANK, J., ZEIDLER, R. (1989) Compilation of tRNA sequences and sequences of tRNA genes. *Nucl Acids Res* 17: 1-72.
- STAGINNUS, C., WINTER, P., DESEL, C., SCHMIDT, T., KAHL, G. (1999) Molecular structure and chromosomal localization of major repetitive DNA families in the chickpea (*Cicer arietinum* L.) genome. *Plant Mol Biol* 39: 1037-1050.
- STAGINNUS, C. (1995) Genomanalyse der Gattung *Cicer*: Vorkommen und Abundanz repetitiver DNA, Staatsexamensarbeit, Frankfurt a.M.

- STAGINNUS, C., HÜTTEL, B., DESEL, C., SCHMIDT, T., KAHL, G. (2001) A PCR-based assay to detect *En/Spm*-like transposon sequences in plants. *Chrom Res* 9: 591-605.
- STEPHAN, W. (1989) Tandem-repetitive noncoding DNA: forms and forces. *Mol Biol Evol* 6: 198-212.
- STRACHAN, T., WEBB, D., DOVER, GA. (1985) Transition stages of molecular drive in multi-copy DNA families in *Drosophila*. *EMBO J* 4: 1701- 1708.
- SU, PY., BROWN, TA. (1997) *Ty3-gypsy*-like retrotransposon sequences in tomato. *Plasmid* 38: 148-157.
- SUMNER, AT. (1990) Chromosome banding. Unwin Hayman, London 1990, pp. 155-186.
- SUNDARESAN, V. (1996) Horizontal spread of transposon mutagenesis: new uses for old elements. *Trends Plant Sci.* 1: 184-190.
- SUONIEMI, A., ANAMTHAWAT-JÓNSSON, K., ARNA, T., SCHULMANN, AH. (1996a) Retrotransposon *BARE1* is a major, dispersed component of the barley (*Hordeum vulgare* L.) genome. *Plant Mol Biol* 30: 1321-1329.
- SUONIEMI, A., NARVANTO, A., SCHULMAN, AH. (1996b) The *BARE1* retrotransposon is transcribed in barley from an LTR promoter active in transient assays. *Plant Mol Biol* 31: 295-306.
- SUONIEMI, A., SCHMIDT, D., SCHULMANN, AH. (1997) *BARE1* insertion site preferences and evolutionary conservation of RNA and cDNA processing sites. *Genetica* 100: 219-230.
- SUONIEMI, A., TASKANEN, J., SCHULMAN A. (1998) *Gypsy*-like retrotransposons are widespread in the plant kingdom. *Plant J.* 13: 699-705.
- TAKAHASHI, H., OKAZAKI, S., FUJIKAWA, H. (1997) A new family of site-specific retrotransposons, *SART1*, is inserted into telomeric repeats of the silkworm, *Bombyx mori*. *Nucl Acids Res* 25: 578-584.
- TAKEDA, S., ANDO, H., TAKEDA, K., HARRISON, GE., HESLOP-HARRISON, JS. (2000) The distribution, organization and evolution of two abundant and widespread DNA sequences in the genus *Hordeum*. *Theor Appl Genet* 100: 169- 176.
- TAKEDA, S., SUGIMOTO, K., OTSUKI, H., HIROCHIKA, H. (1998) Transcriptional activation of the tobacco retrotransposon *Tto1* by wounding and methyl jasmonate. *Plant Mol Biol* 36: 365-376.
- TAKEDA, S., SUGIMOTO, K., OTSUKI, H., HIROCHIKA, H. (1999) A 13-bp cis-regulatory element in the LTR promoter of the tobacco retrotransposon *Tto1* is involved in responsiveness to tissue culture, wounding, methyl jasmonate and fungal elicitors. *Plant J* 18: 383-393.
- TANDA, S., MULLOR, JL., CORCES, VG. (1994) The *Drosophila* retrotransposon *tom* encodes an envelope protein. *Mol Cell Biol* 14: 5392-5401.

- TAYYAR, RI., WAINES, JG. (1996) Genetic relationships among annual species of *Cicer* (Fabaceae) using isozyme variation. *Theor Appl Genet* 92: 245-254.
- TAYYAR, RI., LUKASZEWSKI, AJ., WAINES, JG. (1994) Chromosome banding patterns in the annual species of *Cicer*. *Genome* 37: 656-663.
- TEMIN, HM. (1980) Origin of retroviruses from cellular movable elements. *Cell* 21: 599-600.
- The *Arabidopsis* Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796-815.
- THOMAS, CA. (1971) The genetic organization of chromosomes. *Ann Rev Genet* 5: 237-256.
- THOMSON, KG., THOMAS, JE., DIETZGEN, RG. (1998) Retrotransposon-like sequences integrated into the genome of pineapple, *Ananas comosus*. *Plant Mol Biol* 38: 461-465.
- TIKHONOV, AP., LAVIE, L., TATOUT, C., BENNETZEN, JL., AVRAMOVA, Z., DERAGON, JM. (2001) Target sites for SINE integration in *Brassica* genomes display nuclear matrix binding activity. *Chrom Res* 9: 325-337.
- TIKHONOV, AP., SANMIGUEL, PJ., NAKAJIMA, Y., GORENSTEIN, ND., BENNETZEN, JL, AVRAMOVA, Z. (1999) Colinearity and its exceptions in the orthologous *adh* regions of maize and sorghum. *Proc Natl Acad Sci USA* 96: 7409-7414.
- TISSIER, A., MARILLONNET, S., KLIMYUK, V. ET AL. (1999) Multiple independent defective *Suppressor-mutator* transposon insertions in *Arabidopsis*: a tool for functional genomics. *Plant Cell* 11: 1841-1852.
- TRASK, BJ. (1991) Fluorescence in situ hybridization: Applications in cytogenetics and gene mapping. *Trends Genet* 7: 149-154.
- TSAY, YF., FRANK, MJ., PAGE, T., DEAN, C., CRAWFORD, NM. (1993) Identification of a mobile endogenous transposon in *Arabidopsis thaliana*. *Science* 260: 342-344.
- TURCICH, MP., BOKHARRIRIZA, A., HAMILTON, DA., HE, CP., MESSIER, W., ET AL. (1996) *PREM2*, a *copia*-type retroelement in maize is expressed preferentially in early microspores. *Sexual Plant Reprod* 9: 65-74.
- TURCICH, MP., MASCARENHAS, JP. (1994) *PREM-1*, a putative maize retroelement has LTR (long terminal repeat) sequences that are preferentially transcribed in pollen. *Sex Plant Reprod* 7: 2-11.
- TUTOIS, S., CLOIX, C., CUVILLIER, C., ESPAGNOL, MC., LAFLEURIEL, J., PICARD, G., TOURMENTE, S. (1999) Structural analysis and physical mapping of a pericentromeric region of chromosome 5 of *Arabidopsis thaliana*. *Chrom Res* 7: 143- 156.
- UDUPA, SM., SHARMA, A., SHARMA, AP., PAI, RA. (1993) Narrow genetic variability in *Cicer arietinum* L. as revealed by RFLP analysis. *J Plant Biochem Biotechnol* 2: 83-86.

- UNFRIED, K., SCHIEBEL, K., HEMLEBEN, V. (1991) Subrepeats of rDNA intergenic spacer present as prominent independent satellite DNA in *Vigna radiata* but not in *Vigna angularis*. *Gene* 99: 63-68.
- VAN DEN BROECK, D., MAES, T., SAUER, M., ZETHOF, J., DE KEUKELEIRE, P., D'HAUW, M., MONTAGU, M., GERATS, T. (1998) Transposon display identifies individual transposable elements in high copy number lines. *Plant J* 13: 121-129.
- VAN DER MAESEN, LJG. (1987) Origin history and taxonomy of chickpea. In: Saxena, MC., Singh, KB. (eds): *The chickpea*, pp11-34. CAB Int Publ, UK.
- VARAGONA, MJ., PURUGGANAN, M., WESSLER, SR. (1992) Alternative splicing induced by insertion of retrotransposons into the maize *waxy* gene. *Plant Cell* 4: 811-820.
- VARMUS, HE. (1983) Retroviruses. In: Shapiro, JA. (ed) *Mobile Genetic Elements*. Academic Press. New York, pp.411-505.
- VEDEL, F., DELSENY, M. (1987) Repetivity and variability of higher plant genomes. *Plant Physiol Biochem* 25: 191-210.
- VERSHININ, AV., ALKHIMOVA, EG., HESLOP-HARRISON, JS. (1996) Molecular diversification of tandemly organized DNA sequences and heterochromatic chromosome regions in some Triticeae species. *Chrom Res* 4: 517-525.
- VERSHININ, AV., HESLOP-HARRISON, JS. (1998) Comparative analysis of the nucleosomal structure of rye, wheat and their relatives. *Plant Mol Biol* 36: 149-161.
- VERSHININ, AV., ELLIS, TH. (1999) Heterogeneity of the internal structure of *PDR1*, a family of *Ty1-copia*-like retrotransposons in pea. *Mol Gen Genet* 262: 703-713.
- VICIENT, CM., JÄÄSKELÄINEN, MJ., KALENDAR, R., SCHULMAN, AH. (2001) Active retrotransposons are a common feature of grass genomes. *Plant Physiol* 125: 1283-1292.
- VICIENT, CM., SUONIEMI, A., ANAMTHAWAT-JÓNSSON, K., TASKANEN, J., BEHARAV, A., NEVO, E., SCHULMANN, AH. (1999) Retrotransposon *BARE1* and its role in genome evolution in the genus *Hordeum*. *Plant Cell* 11: 1769-1784.
- VIG, BK. (1994) Do specific nucleotide bases constitute the centromere? *Mutat Res* 309: 1-10.
- VOGT, P. (1990) Potential genetic functions of tandem repeated DNA sequence blocks in the human genome are based on a highly conserved „chromatin folding code“. *Hum Genet* 84: 301-336.
- VOS, P., HOGERS, R., BLEEKER, M., REIJANS, M., VAN DE LEE, T., HORNES, M., FRIJTER, BA., POT, J., PELEMAN, J., KUIPER, M., ZABEAU, M. (1995) AFLP. A new technique for DNA fingerprinting. *Nucl Acids Res* 23: 4407-4414.
- VOYTAS, DF., CUMMINGS, MP., KONIECZNY, A., AUSUBEL, FM., RODERMEL, SR. (1992) *Copia*-like retrotransposons are ubiquitous among plants. *Proc Natl Acad Sci USA* 89: 7124-7128.

- VOYTAS, DF., KONIESZNY, A., CUMMINGS, MP, AUSUBEL, FM. (1990) The structure, distribution and evolution of the *Tal* retrotransposable element family in *Arabidopsis thaliana*. *Genetics* 126: 713-721.
- WAKIMOTO, B. (1998) Beyond the nucleosome: epigenetic aspects of position effect variegation in *Drosophila*. *Cell* 93: 321-324.
- WALBOT, V. (1999) UV-B damage amplified by transposons in maize. *Nature* 397: 398-399.
- WALBOT, V. (2000) Saturation mutagenesis using maize transposons. *Curr Opin Plant Biol* 3: 103-107.
- WALKER, EL., ROBBINS, TP., BUREAU, TE., KERMICLE, J., DELLAPORTA, SL. (1995) Transposon-mediated chromosomal rearrangements and gene duplications in the formation of the maize *R-r* complex. *EMBO J* 14: 2350-2363.
- WALSH, JB. (1987) Persistence of tandem arrays: implications for satellite and simple-sequence DNAs. *Genetics* 115: 553- 567.
- WAUGH, R., MCLEAN, K., FLAVELL, AJ., PEARCE, SR., KUMAR, A., THOMAS, BBT., POWELL, W. (1997) Genetic distribution of *BARE1*-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP). *Mol Gen Genet* 253: 687-694.
- WEISING, K., KAEMMER, D., WEIGAND, F., EPPLIN, JT., KAHL, G. (1992) Oligonucleotide fingerprinting reveals various probe-dependent levels of informativeness in chickpea (*Cicer arietinum* L.). *Genome* 35: 436-442.
- WEISING, K., NYBOM, H., WOLFF, K., MEYER, W. (1995) DNA fingerprinting in plants and fungi. CRC Press, Boca Raton, 1.Aufl.
- WESSLER, S. (1998) Transposable elements associated with normal plant genes. *Physiologia Plantarum* 103: 581-586.
- WESSLER, SR. (1996) Plant retrotransposons: turned on by stress. *Curr Biol* 6: 959-961.
- WESSLER, SR., BUREAU, TE., WHITE, SE. (1995) LTR-Retrotransposons and MITES: important players in the evolution of plant genomes. *Curr Opin Genet Dev* 5: 814-821.
- WHITE, SE., HABERA, LF., WESSLER, SR. (1994) Retrotransposons in the flanking regions of normal plant genes: A role for *cop*ia-like elements in the evolution of gene structure and expression. *Proc Natl Acad Sci USA* 91: 11792-11796.
- WILLARD, HF., WAYE, JS. (1987b) Chromosome-specific subsets of human alpha satellite DNA: analysis of sequence divergence within or between chromosomal subsets and evidence for an ancestral pentameric repeat. *J Mol Evol* 25: 207-214.
- WILLARD, HF., WAYE, JS. (1987a) Hierarchical order in chromosome-specific human alpha satellite DNA. *Trends Genet* 3: 192-198.

- WILLARD, HF. (1991) Evolution of alpha satellite. *Curr Opin Genet Dev* 1: 509- 514.
- WINTER, P., BENKO-ISEPPON, AM., HÜTTEL, B., RATNAPARKHE, M., TULLU, A., SONNATE, G., PFAFF, T., TEKEOGLU, M., SANTRA, D., SANT, VJ., RAJESH, PN, KAHL, G., MUEHLBAUER, FJ. (2000) A linkage map of the chickpea (*Cicer arietinum* L.) genome based on the recombinant inbred lines from a *C.arietinum* x *C.reticulatum* cross: localization of resistance genes for *Fusarium* races 4 and 5. *Theor Appl Genet* 101: 1155-1163.
- WINTER, P., PFAFF, T., UDUPA, S.M., SHARMA, P.C., SAHI, S., ARREGUIN-ESPINOZA, R., WEIGAND, F., MUEHLBAUER, F.J., KAHL, G. (1999) Characterization and mapping of sequence-tagged microsatellite sites in the chickpea (*Cicer arietinum* L.) genome. *Mol Gen Genet* 262: 90-101.
- WRIGHT, D., KE, N., SMALLE, J., HAUGE, B., GOODMAN, H. AND VOYTAS, D. (1996) Multiple non-LTR retrotransposons in the genome of *Arabidopsis thaliana*. *Genetics* 142: 569-578.
- WRIGHT, DA., VOYTAS, DF. (1998) Potential retroviruses in plants: *Tal* is related to a group of *Arabidopsis thaliana* *Ty3-gypsy* retrotransposons that encode envelope-like proteins. *Genetics* 149: 703-715.
- WU, T., WU, R. (1987) A new rice repetitive DNA shows sequence homology to both 5S RNA and tRNA. *Nucl Acids Res* 15: 5913-5923.
- XIA, X., ERICKSON, L. (1993) An AT-rich DNA sequence, E180, in alfalfa (*Medicago sativa*). *Genome* 36: 427-432.
- XIA, X., DU, S., ERICKSON, L. (1996) A moderately repetitive DNA sequence in alfalfa is transcribed in a floral-specific manner. *Genome* 39: 9-16.
- XIONG, Y., EICKBUSH, TH. (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J* 9: 3353-3362.
- YOSHIOKA, Y., MATSUMOTO, S., KOJIMA, S., OHSHIMA, K., OKADA, N., MACHIDA, Y. (1993) Molecular characterization of a short interspersed repetitive element from tobacco that exhibits sequence homology to specific tRNAs. *Proc Natl Acad Sci USA* 90: 6562-6566.
- YPHREMOV, A., SAEDLER, H. (2000) Display and isolation of transposon-flanking sequences strating from genomic DNA or RNA. *Plant J* 21: 495-505.
- ZAHNKE, C., HEMLEBEN, V. (1997) A new *Solanum* satellite DNA containing species-specific sequences which can be used for identification of genome parts in somatic hybrids of potato. *Plant Science* 126: 185- 191.
- ZHAO, X., JI, Y., DING, X., STELLY, DM., PATERSON, AH. (1998) Macromolecular organization and genetic mapping of a rapidly evolving chromosome-specific repeat family (B77) in cotton (*Gossypium*). *Plant Mol Biol* 38: 1031-1042.
- ZWICK, MS., ISLAM-FARIDI, MN., ZHANG, HB., HODNETT, GL., GOMEZ, MI., KIM, JS., PRICE, HJ., STELLY, DM. (2000) Distribution and sequence analysis of the centromere-associated repetitive element CEN 38 of *Sorghum bicolor* (Poaceae). *Am J Bot* 87: 1757-1764.

G. PUBLIKATIONEN

STAGINNUS, C., WINTER, P., DESEL, C., SCHMIDT, T., KAHL, G. (1999) Molecular structure and chromosomal localization of major repetitive DNA families in the chickpea (*Cicer arietinum* L.) genome. *Plant Mol Biol* 39: 1037-1050.

STAGINNUS, C., HÜTTEL, B., DESEL, C., SCHMIDT, T., KAHL, G. (2001) A PCR-based assay to detect *En/Spm*-like transposon sequences in plants. *Chrom Res* 9: 591-605.

WINTER, P., STAGINNUS, C., SHARMA, P.C., KAHL, G. (2001) Organisation and genetic mapping of the chickpea genome. In: *Biotechnology for the Improvement of Legumes* (P.K. Jaiwal and R.P. Singh, eds.) Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 285-360.

BENKO-ISEPPON, A.M., WINTER, P., HÜTTEL, B., STAGINNUS, C., MUEHLBAUER, F.J., KAHL, G. (2002) Markers closely linked to *Fusarium* resistance genes in chickpea show homology to pathogenesis-related genes located on *Arabidopsis* chromosome 5 and 1. *Theor Appl Genet*: (in Druck)

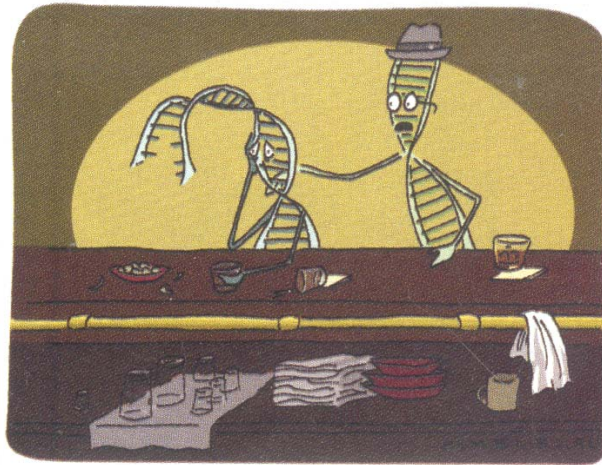
H. ABKÜRZUNGEN

(v/v)	Volumen pro Volumen
(w/v)	Gewicht pro Volumen
°C	Grad Celsius
λ-DNA	DNA des Bakteriophagen Lambda
A	Adenin
AFLP	„Amplified fragment length polymorphism“
AJxxxxxx	sechsstellige Akzessionsnummer in EMBL-Datenbank
APS	Ammoniumpersulfat
ATP	Adenosintriphosphat
B	nicht A (= C-, G- oder T)
BAC	„Bacterial artificial chromosome“
bidest	doppelt destilliert
bp (Mbp, kbp)	Basenpaare (Mega-, Kilo-)
BSA	„Bovine serum albumin“ (Rinderserumalbumin)
C	Cytosin
cDNA	DNA-Kopie einer RNA-Matrize
Ci	Curie
cM	centiMorgan
CMA	Chromomycin A3
CODEHOP	„Consensus degenerate hybrid oligonucleotide primer“
CTAB	Cetyltrimethylammoniumbromid
Cy3	Cyanin, Fluorochrom
D	nicht C (= A, G oder T)
DAF	„DNA amplification fingerprinting“
DAPI	4,6 Diamino-2-phenylindol
dATP	Deoxyadenosintriphosphat
dCTP	Deoxycytidintriphosphat
dGTP	Deoxyguanosintriphosphat
D	nicht C (= A, T oder G)
DEPC	Diethylpyrocarbonat
DNA	Desoxyribonukleinsäure
dNTPs	Desoxynukleotidtriphosphate
ds	Doppelstrang-
dT	Desoxythymidin
DTT	Dithiothreitol
dTTP	Desoxythymidintriphosphat
dUTP	Desoxyuridintriphosphat
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	Ethylendiamintetraacetat
EMBL	„European Molecular Biological Laboratory“
Endo, EN	Endonuklease
<i>env</i>	„envelope“, Leseraster mit Ähnlichkeit zu retroviralen Hüllproteinen
EST	„Expressed sequence tag“
EtBr	Ethidiumbromid
ETID	„Expressed transposon insertion display“
ETS	„External transcribed spacer“
Fig	Abbildung
FISH	Fluoreszenz <i>in-situ</i> Hybridisierung
FITC	Fluorescein-Isothiocyanat, Fluorochrom
G	Guanin

H. Abkürzungen

g (mg, µg, ng, pg)	Gramm (Milli-, Mikro-, Nano-, Piko-)
<i>gag</i>	„group-specific antigen“, Leseraster in Retroelementen
H	nicht G (= A-, C- oder T)
h	Stunde(n)
I	Inosin
ICARDA	"International Center for Agricultural Research in Dry Areas"
ICRISAT	"International Crop Research Institute for Semi-Arid Tropics"
Arid	
IGS	„Intergenic spacer region“
ILC	„International Legume Cicer“
ILWC	„International Legume Wild Cicer“
Int	Integrase
IPTG	Isopropyl-β-D-Thio-Galaktopyranosid
IRAP	„Inverse retrotransposon amplified polymorphism“
ISSR	„Inter simple sequence repeat“
K	G oder T
kV	Kilovolt
l (ml, µl)	Liter (Milli-, Mikro-)
LINE	„Long interspersed nuclear elements“
LTR	„Long terminal repeat“
M	A oder C
M (µM) (Maßeinheit)	molar (mikro-)
MCS	"Multiple cloning site", (multiple Klonierungsstelle)
min	Minute(n)
MIP	„Methylation induced premeiotically“
MITEs	„Miniature inverted-repeat transposable element“
mmol (pmol)	Millimol (Piko-)
mRNA	„messenger“-RNA
N	A-, C-, G- oder T
NCBI	„National Center for Biotechnology Information“
NOR	Nukleolus Organisator Region
NTS	„non-transcribed spacer“
OD _x	Optische Dichte bei der Wellenlänge x nm
ORF	offenes Leseraster („open reading frame“)
PBS	Primerbindestelle („primer binding site“)
PCR	Polymerasekettenreaktion
<i>pol</i>	„polyprotein“, Leseraster in Retroelementen
PPT	Polypurintrakt
Prot	Protease
PTGS	„Posttranscriptional gene silencing“
PVP 40	Polyvinylpyrrolidon 40
QTL	„Quantitative trait loci“
R	A oder G
RAPD	"Random amplified polymorphic DNA"
RBIP	„Retrotransposon-based insertion polymorphism“
rDNA,-RNA	ribosomale DNA,-RNA
REMAP	„Retrotransposon microsatellite amplified polymorphism“
RFLP	„Restriction fragment length polymorphism“
RGA	Resistenzgenanaloge
RIL	rekombinante Inzuchtslinie
RIP	„Repeat-induced point mutation“
RNA	Ribonukleinsäure

RNase	Ribonuklease
R-Region	„Repeated RNA region“ (Teil des LTR)
RT (Temperaturangabe)	Raumtemperatur
RT	Reverse Transkriptase
RT-PCR	Herstellung einer cDNA-Kopie aus einer RNA-Matrize mittels RT und anschließender PCR
RIVP	„Retrotransposon internal variation polymorphism“
S	C oder G
SCAR	„Sequence-characterized amplified region“
SDS	Natrium(Sodium)dodecylsulfat
sec	Sekunde(n)
SINE	„Short interspersed nuclear elements“
7SL RNA	„Spliced leader“-RNA
SNP	„Single nucleotide polymorphism“
S-SAP	„Sequence-specific amplified polymorphisms“
SSC	„Saline sodium citrate“ (Natriumcitratsalz)
SSR	„Simple sequence repeats“
STMS	„Sequence-tagged microsatellite sites“
T	Thymin
Tab	Tabelle
TAIR	„The Arabidopsis Information Resource“
TBE	Tris-Borat-EDTA
TE	transposables Element
TEMED	N,N,N',N'-Tetramethyldiamin
TGS	„Transcriptional gene silencing“
TID	„Transposon insertion display“
TIR	„Terminal inverted repeat“
TPase	Transposase
Tris	Tris-(hydroxymethyl)-aminomethan
tRNA	transfer RNA
TSD	Zielsequenzverdoppelung („target site duplication“)
U	Units
U3-, U5-Region	„Unique 5'/3'-RNA region“ (Teil des LTR)
5'-, 3'-UTR	5', 3'-untranslatierte Region
Upm	Umdrehungen pro Minute
UV	ultraviolettes Licht
V	nicht T (= A, C oder G)
V/cm	Volt pro Zentimeter
VLP	„Virus-like particle“
Vol	Volumen
W	A oder T
X-Gal	5-Brom-4-Chlor-3-Indolyl- β -Galaktopyranosid
Y	C oder T
YAC	„Yeast artificial chromosome“



*She's a transposon, Doug . . .
You knew she'd leave you some day.*

DANKSAGUNG

Ich danke Prof. Dr. Günter Kahl für die Bereitstellung eines Arbeitsplatzes und die Betreuung dieser Arbeit, ebenso für sein Interesse und seine Diskussionsbereitschaft.

Ferner möchte ich mich bei der gesamten Arbeitsgruppe „Pflanzliche Molekularbiologie“ für die gute Arbeitsatmosphäre, Hilfsbereitschaft und Diskussionswilligkeit bei wissenschaftlichen Problemen bedanken. Das gilt besonders für T. Pfaff, P. Winter, R. Horres, G. Gortner, J. Geistlinger und S. Kost.

Besonderen Dank verdient B. Hüttel, der sehr ausdauernd alle möglichen Probleme mit mir diskutierte, mich immer unterstützte und außerdem mit diverser Software vertraut machte.

Dr. T. Schmidt danke ich sehr herzlich für die Möglichkeit, FISH in seinem Labor zu erlernen, ebenso für seine kompetenten Ratschläge und seine unerschöpflichen Kenntnisse über repetitive DNA.

Dank gilt auch der gesamten Arbeitsgruppe „Molekulare Cytogenetik“ an der Uni Kiel für ihre Unterstützung und Gastfreundschaft, besonders Christine Desel, die mir beim FISHen stets mit Rat und Tat zur Seite stand und immer zu Diskussionen bereit war.

Meinen Eltern und der restlichen Familie möchte ich sehr für ihre finanzielle und ideelle Unterstützung danken.

Vielen Dank ebenso an alle, die mir beim Korrekturlesen dieser Arbeit geholfen haben.

Außerdem danke ich der FAZIT-Stiftung für die finanzielle Unterstützung dieser Arbeit.

LEBENS LAUF

Name:	Christina Staginnus	
Geburtstag:	17.05.1971	
Geburtsort:	Hanau	
Eltern:	Götz und Hanne-Lore Staginnus, geb. Meyer	
Ausbildung:	1977 - 1981	Grundschule in Hanau
	1981 - 1983	Förderstufe in Hanau
	1983 - 1990	Gymnasium in Hanau, Abitur: Juni 1990
	10/1990 - 11/1995	Studium des Lehramts an Gymnasien mit den Fächern Biologie und Chemie an der Johann Wolfgang Goethe-Universität, Frankfurt a.M.
	Oktober 1992	Zwischenprüfung für Kandidaten des wissenschaftlichen Lehramts an Höheren Schulen (Gesamtnote sehr gut)
	November 1995	Erste Staatsprüfung für Lehramtskandidaten an Gymnasien, Staatsexamensarbeit in Botanik (Arbeitsgruppe Pflanzliche Molekularbiologie) bei Prof. Dr. G. Kahl „Genomanalyse der Gattung <i>Cicer</i> : Vorkommen und Abundanz repetitiver DNA“ (Gesamtnote sehr gut)
	5/1996 - 10/1996	Pädagogischer Vorbereitungsdienst für das Lehramt an Gymnasien im Studienseminar Offenbach, Unterbrechung zum 31.10.1996 zu Gunsten einer Promotionsarbeit
	ab 11/1996	Promotion an der Universität Frankfurt a.M. in der Arbeitsgruppe Pflanzliche Molekularbiologie bei Prof. Dr. G. Kahl, Titel: „Repetitive DNA-Sequenzen im Genom der Kichererbse (<i>Cicer arietinum</i> L.)“
Berufstätigkeit:	seit 11/1998	Wissenschaftlicher Mitarbeiter des Botanischen Instituts der Johann Wolfgang Goethe-Universität, Frankfurt a.M.

