

**Analyse von Zentromer-DNA
für die
Erzeugung künstlicher Chromosomen**

Tobias Schwarz

2008

Aus der Abteilung Medizinische Genetik
der Ludwig-Maximilians-Universität München
Leiter: Prof. Dr. med. Jan Murken

**Analyse von Zentromer-DNA
für die
Erzeugung künstlicher Chromosomen**

Dissertation
zum Erwerb des Doktorgrades der Medizin
an der Medizinischen Fakultät der
Ludwig-Maximilians-Universität zu München

vorgelegt von
Tobias Schwarz
aus Dachau
2008

Mit Genehmigung der Medizinischen Fakultät
Der Universität München

1. Berichterstatter:	Prof. Dr. J. Murken
2. Berichterstatterin:	Prof. Dr. O. Steinlein
Mitberichtersteller:	Prof. Dr. S. Stengel-Rutkowski Priv. Doz. Dr. R. Kappler Prof. Dr. G. W. Bornkamm
Mitbetreuung durch den promovierten Mitarbeiter:	Dr. D. Schindelhauer
Dekan:	Prof. Dr. D. Reinhardt
Tag der mündlichen Prüfung:	19.06.2008

Inhaltsverzeichnis**Abkürzungsverzeichnis..... VI**

A. Einleitung	1
1. Künstliche Chromosomen.....	1
1.1. Funktionen und Bestandteile von (künstlichen) Chromosomen	1
1.2. „Top Down“ und „Bottom Up“	3
1.3. Erste Erfolge in der Entwicklung chromosomaler Vektoren	3
1.4. Anwendungsbereiche künstlicher Chromosomen	4
2. Die Erforschung der Zentromer-DNA besitzt eine vergleichsweise lange Geschichte...	6
3. Aufbau und Funktion der Zentromere	7
3.1. Zentromere sind essentielle Bestandteile eukaryotischer Chromosomen	7
3.2. Die Protein-Struktur der menschlichen Zentromere	8
3.3. α -Satelliten-DNA an menschlichen Zentromeren.....	9
3.4. Der DXZ1-Lokus des menschlichen X-Chromosoms.....	12
4. Sequenz-Funktions Relation von α -Satelliten-DNA	14
4.1. Die Zentromer-Regionen aller höherer Eukaryoten enthalten repetitive DNA, sind jedoch evolutionär nicht konserviert	14
4.2. Die CENP-B Box besitzt proteinbindende Eigenschaften	15
4.3. Verschiedene Mechanismen der epigenetischen Markierung der Zentromer- DNA werden diskutiert	16
4.4. Dizentrische Chromosomen	17
4.5. Neozentromere	19
5. Die Verwendung künstlicher Chromosomen zur Erforschung der Zentromere	20
6. Die Evolution der α -Satelliten-DNA	21
6.1. Die Phylogenese der humanen Zentromer-Arrays	21
6.2. Concerted Evolution.....	23
6.3. Homogenisierung und Fixierung von Varianten	24
6.4. Die Mechanismen von Molecular Drive	25
6.4.1. Unequal crossing over	26
6.4.2. Genkonversion.....	28
6.5. Der experimentelle Nachweis von Molecular Drive.....	29

6.6. α -Satelliten-DNA ist sehr gut geeignet um die Mechanismen von Concerted Evolution zu untersuchen	29
7. Zielsetzung	30
B. Ergebnisse.....	31
1. Die X-chromosomalen α -Satelliten-PACs	31
2. Klonierung und Endsequenzierung von 2 kb und 3 kb <i>Bam</i> -HI Higher order repeats ..	33
2.1. Sequenz-Homologien innerhalb und zwischen den PACs	34
2.2. Die Divergenz zwischen den Higher order repeats wird im Wesentlichen durch variable Positionen verursacht	38
2.3. Ausschluss von Klonierungs- und Sequenzierungsartefakten	39
3. Nachweis der Fixierung der globalen Varianten in der Bevölkerung.....	41
3.1. PCR und Restriktionsverdau humaner X-chromosomaler α -Satelliten-DNA	41
3.2. Klonierung und Sequenzierung von PCR-Fragmenten	42
3.3. Restriktionsverdau und Southern Blot kompletter genomischer DNA	44
4. Die variablen Basen treten in den Higher order repeats unabhängig voneinander auf..	45
5. Kartierung von Varianten innerhalb der α -Satelliten-PACs.....	47
5.1. Bestätigung der Homogenität der PACs durch Kartierung mittels <i>Bam</i> -HI	48
5.2. Die Varianten treten in den PACs in Clustern auf	49
5.3. Die PACs repräsentieren mindestens 0,51 Mb des DXZ1 Arrays eines einzelnen menschlichen X-Chromosoms	51
5.4. Bestätigung des menschlichen Ursprungs der Cluster sowie der Stabilität der α -Satelliten-DNA	51
6. Weitere variable Basen treten unabhängig von den <i>Hind</i> -III und <i>Dra</i> -III Clustern auf	52
6.1. Klonierung von Higher order repeats aus den <i>Hind</i> -III Clustern des PAC B11	52
6.2. Klonierung von Higher order repeats aus den <i>Dra</i> -III Clustern des homogenen PAC A7	53
6.2.1. Konstruktion des Vektors BsD	53
6.2.2. Auch in den <i>Dra</i> -III Clustern des homogenen PAC A7 treten die variablen Basen unabhängig voneinander auf.....	54
6.3. Klonierung weiterer <i>Bam</i> -HI Higher order repeats des PAC A7 ohne <i>Dra</i> -III Schnittstelle	54
7. Echte Einzelmutationen kommen in den Higher order repeats kaum vor	56

8. Die im menschlichen X-chromosomalen α -Satelliten-Array vorhandenen Varianten sind in der Spezies <i>Gorilla gorilla</i> nicht nachweisbar.....	58
C. Methoden.....	60
1. DNA-Präparationsmethoden.....	60
1.1. Plasmidisolierung mittels Mini-Präparation.....	60
1.2. Präparation intakter PAC-DNA in Agarose-Blöckchen.....	60
1.2.1. Herstellung von Bakterien-Agarose-Blöckchen.....	60
1.2.2. Reinigung der Blöckchen von Proteinen.....	61
1.2.3. Aufreinigung intakter PAC-DNA.....	62
1.3. Plasmid-Schnellpräparationen aus Bakterienkolonien.....	63
2. Aufreinigung von DNA.....	63
2.1. Phenolextraktion.....	63
2.2. Ethanol-fällung.....	64
2.3. Aufreinigung von PCR-Produkten mittels Mikrozentrifugiersäule.....	64
2.4. Gelextraktion mittels Mikrozentrifugiersäule.....	64
2.5. Elektroelution von DNA aus Agarosegelen.....	64
3. Restriktionsenzymatische DNA-Spaltung.....	65
3.1. Spaltung von DNA in wässriger Lösung.....	65
3.2. Spaltung von DNA in Agarose-Blöckchen.....	65
3.3. Partialverdau von PAC-DNA.....	66
4. Gelelektrophorese.....	68
4.1. Auftrennung von DNA-Fragmenten mittels konventioneller Gelelektrophorese..	68
4.2. Auftrennung langer DNA-Fragmente mittels Pulsfeld-Gelelektrophorese (PFGE).....	68
5. DNA-Klonierung.....	69
5.1. Dephosphorylierung von DNA-Enden.....	69
5.2. Ligierung.....	70
5.3. Herstellung elektrokompenter Bakterien.....	70
5.4. Transformation von <i>E. coli</i> mittels Elektroporation.....	71
5.5. Antibiotische Selektion und blau/weiß Selektion von Klonen mittels X-Gal.....	71
5.6. Anlegen von Übernachtskulturen und Aufbewahrung von Klonen als Glycerinkultur.....	72

6. Verdau überhängender 5'-Enden mittels S1-Endonuklease	72
7. Einklonierung eines doppelsträngigen Oligonukleotids	73
8. Polymerasekettenreaktion (PCR).....	74
9. DNA-Sequenzierung.....	75
9.1. Korrekturlesen der DNA-Sequenzen	76
10. Blotting von DNA auf Nylonmembranen.....	76
11. Radioaktive Markierung von DNA-Proben.....	77
12. Hybridisierung	78
D. Material	80
1. DNA.....	80
1.1. Humane α -Satelliten-PACs.....	80
1.2. Humane genomische DNA.....	80
1.3. <i>Gorilla gorilla</i> genomische DNA	80
1.4. DNA-Sequenzen aus Datenbanken	80
2. Vektoren.....	81
3. Elektrokompente Bakterien	81
4. Kits.....	81
5. Oligonukleotide	81
6. Längenstandards	82
7. Enzyme	82
8. Chemikalien	82
9. Radioaktive Substanzen	83
10. Geräte und sonstige Materialien	84
11. Computerprogramme	85
12. Datenbanken	85
E. Diskussion.....	86
1. Die stabil klonierten PACs repräsentieren mindestens 500 kb des α -Satelliten-Arrays eines einzelnen, menschlichen X-Chromosoms	87
2. Die α -Satelliten-DNA der PACs weist eine sehr hohe Homogenität auf und macht eine Sequenzierung mit konventionellen Methoden unmöglich.....	87

3. Die Ergebnisse ermöglichen weiterführende Untersuchungen von <i>de novo</i> erzeugten künstlichen Chromosomen	89
4. Die variablen Basen treten unabhängig von ihrer Lokalisation im Higher order repeat auf	90
5. Die häufigen Varianten sind in der Bevölkerung fixiert.....	90
6. Die schnelle Ausbreitung der Varianten erfolgt maßgeblich durch Genkonversionen .	91
7. Genkonversionen sind häufige Ereignisse in tandem-repetitiver DNA.....	93
8. Die Ausbildung der Cluster erfolgt durch intrachromosomale Genkonversionen.....	94
9. Die relativen Geschwindigkeiten der intra- und interchromosomalen Ausbreitung von Varianten.....	95
10. Genkonversionen und die gemeinsame Evolution der α -Satelliten-DNA.....	97
11. Ursache und Geschwindigkeit des Molecular Drive der α -Satelliten-DNA	98
12. Ausblick	99
F. Zusammenfassung	101
G. Literaturverzeichnis	103
H. Anhang.....	121
1. Am Aufbau der Zentromere beteiligte Proteine.....	121
2. Karten der verwendeten Vektoren	123
3. Auflistung der klonierten Higher order repeats und Einzelsequenzen mit EMBL/Genbank Accession numbers.....	124
3.1. Endsequenzen der in die Analyse einbezogenen Higher order repeats.....	124
3.2. PAC-Endsequenzen.....	126
3.3. Nicht in die Analyse einbezogene Klone	127
Publikationen.....	128
Danksagung.....	129
Lebenslauf.....	130

Abkürzungsverzeichnis

A	Adenin
Abb.	Abbildung
ACA	Antizentromer-Antikörper
bp	Basenpaare
BSA	Rinderserumalbumin
C	Cytosin
cDNA	complementary DNA
CENP	Centromere protein
CIP	Calf intestine phosphatase
DNA	Desoxyribonukleinsäure
dNTP	2'-Desoxyribonukleosid-5'-triphosphat (dATP, dCTP, dTTP, dGTP)
ds	double stranded
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	Ethylendiamintetraessigsäure
EMBL	European Molecular Biology Laboratory
et al.	und andere
EthBr	Ethidium-Bromid
FISH	Fluoreszenz <i>in situ</i> Hybridisierung
G	Guanin
HAC	Human artificial chromosome
HAT	Hypoxanthin, Aminopterin, Thymidin
HPRT	Hypoxanthin-Guanin-Phosphoribosyltransferase
INCENP	Inner centromere protein
IPTG	Isopropyl-1-thio- β -D-Galactopyranosid
Kap.	Kapitel
kb	Kilobasenpaare
kDa	Kilodalton
LB-Medium	Luria-Bertani Medium
LINE	Long interspersed element
MAC	Mammalian artificial chromosome
Mb	Megabasenpaare
mRNA	messenger-RNA
PAC	P1 derived artificial chromosome
PCR	Polymerase chain reaction
PFGE	Pulsfeld-Gelelektrophorese
RNA	Ribonukleinsäure
rRNA	ribosomale RNA
SDS	Sodium-Dodecyl-Sulfat
SINE	Short interspersed element
snRNA	small nuclear RNA
ss	single stranded
T	Thymin
Tab.	Tabelle

TACF	Telomere associated chromosome fragmentation
TAE	Tris-Acetat-EDTA
TBE	Tris-Borat-EDTA
TE	Tris-EDTA
Tris	2-Amino-2-(hydroxymethyl)-1,3-propandiol
tRNA	transfer RNA
ü.N.	über Nacht
UpM	Umdrehungen pro Minute
X-Gal	5-Bromo-4-chloro-3-indolyl- β -D-Galactopyranosid
X-SCID	X-linked severe combined immunodeficiency
YAC	Yeast artificial chromosome

A. Einleitung

Ein von mehreren Arbeitsgruppen weltweit verfolgtes Ziel ist die Konstruktion von künstlichen Chromosomen. Diese sollen als universelle Vektoren einen stabilen, regulierten und sicheren Transfer von genetischem Material in beliebige eukaryotische Zielzellen ermöglichen. Daraus würde sich eine Vielzahl von Anwendungsmöglichkeiten ergeben, von der Erforschung zellulärer und chromosomaler Funktionen, über Analysen einzelner Gene und Sequenzabschnitte, bis hin zu transgenen Tieren oder der Etablierung einer somatischen Gentherapie.

Ein für die Funktion von Chromosomen essentieller Bestandteil ist dabei das Zentromer. Die molekularen Grundlagen der natürlichen menschlichen Zentromere, sowie die Bedingungen zur Ausbildung eines aktiven Zentromers, sind bislang jedoch kaum definiert. Primäres Ziel dieser Arbeit war es daher mehrere aus einer PAC Bank (P1 derived artificial chromosome) isolierte DNA-Abschnitte des Zentromers eines einzelnen menschlichen X-Chromosoms zu charakterisieren. In weiteren Studien sollten diese zur Konstruktion von künstlichen Chromosomen verwendet werden, um aus deren Bereitschaft Zentromere zu formieren und aus der mitotischen Stabilität der entstandenen Chromosomen Rückschlüsse auf die Bedeutung der primären DNA-Sequenz für die Funktion der Zentromere ziehen zu können.

1. Künstliche Chromosomen

1.1. Funktionen und Bestandteile von (künstlichen) Chromosomen

Künstliche Chromosomen haben mehrere Aufgaben zu erfüllen. Sie sollen (1) in nur einer bis wenigen Kopien autonom im Zellkern vorliegen, (2) dort repliziert und bei der Zellteilung zuverlässig auf beide Tochterzellen verteilt werden, (3) die zeitlich und quantitativ kontrollierte Expression der von ihnen kodierten Gene erlauben, sowie (4) zu einer möglichst geringen Beeinträchtigung der Zellphysiologie führen.

In Analogie zu den künstlichen Chromosomen der Hefe sind für die Konstruktion sogenannter „Mammalian artificial chromosomes“ (MACs) respektive „Human artificial chromosomes“ (HACs) drei Bestandteile gefragt: Replikationsursprünge, Telomere und Zentromere (Abb. 1). Die Replikationsursprünge ermöglichen die Verdoppelung der DNA in der S-Phase des Zellzyklus. Während bei prokaryotischen Organismen und der Hefe hierfür exakt definierte Startsequenzen existieren, konnten bei den höheren Eukaryoten keine derartigen Konsensussequenzen identifiziert werden (Gilbert 2001). Jedoch scheinen Replikationsursprünge im menschlichen Genom regelmäßig vorzukommen, so dass eine zusätzliche Bereitstellung von

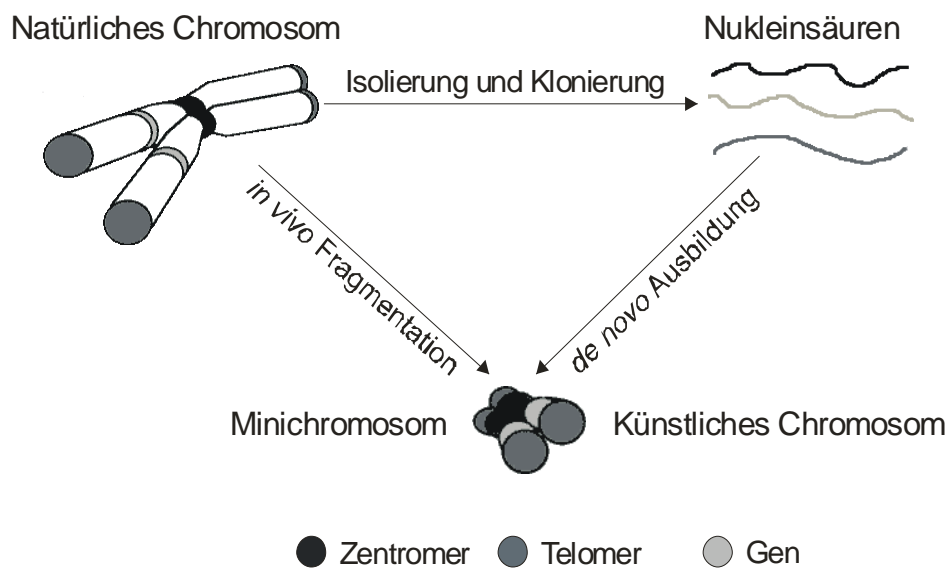


Abb. 1: Minichromosomen und künstliche Chromosomen: Die meisten eukaryotischen Chromosomen besitzen drei wesentliche Bestandteile: Zentromere, Telomere und Replikationsursprünge. Durch Fragmentation natürlicher Chromosomen *in vivo* lassen sich Minichromosomen herstellen. Künstliche Chromosomen entstehen *de novo* aus „nackter“ DNA.

spezifischen DNA-Elementen bei der Konstruktion von künstlichen Chromosomen nicht essentiell ist.

Die Aufgabe der Telomere ist es, die Enden der Chromosomen zu „schützen“. Damit können sie von freien DNA-Enden unterschieden werden, welche zum Beispiel nach Doppelstrangbrüchen auftreten oder von exogener, wie zum Beispiel viraler DNA stammen. In Bezug auf künstliche Chromosomen besteht ihre Bedeutung somit darin, die Integration der transfizierten DNA in die natürlichen Chromosomen der Wirtszelle oder den Abbau durch Nukleasen zu verhindern. Die den Telomeren zugrunde liegende DNA ist dabei relativ gut charakterisiert, und bereits wenige hundert Basenpaare dieser repetitiven TTAGGG-Hexanukleotide genügen, um die Funktion eines Telomers hervorrufen zu können (Hanish et al. 1994). Es konnte sogar gezeigt werden, dass die Konstruktion von künstlichen Chromosomen bei Verwendung zirkulärer DNA auch ohne Telomere möglich ist (Ebersole et al. 2000), wobei es zur Ausbildung von Ringchromosomen ohne freie DNA-Enden kommt.

Den einzig wirklich essentiellen Bestandteil für die Konstruktion von künstlichen Chromosomen stellen somit die Zentromere dar. Sie ermöglichen während der Zellteilung die geordnete Weitergabe jeweils einer der zwei Chromatiden eines replizierten Chromosoms an die beiden Tochterzellen.

1.2. „Top Down“ und „Bottom Up“

Prinzipiell lassen sich zwei Methoden zur Herstellung von künstlichen Chromosomen unterscheiden (Abb. 1): Im sogenannten „Top-Down“ Verfahren werden natürliche Chromosomen *in vivo* mittels Telomer assoziierter Chromosomen-Fragmentation (TACF) schrittweise verkürzt, bis nur noch die für die Funktion des Chromosoms essentiellen Bestandteile vorhanden sind (Barnett et al. 1993). Dadurch ist es gelungen, sogenannte Minichromosomen mit Größen von nur einigen hundert Kilobasen zu erzeugen (Farr et al. 1995, Heller et al. 1996, Mills et al. 1999, Yang et al. 2000). In diese Minichromosomen können dann durch Rekombination, zum Beispiel mittels Cre/loxP-Rekombinase, Gene integriert werden (Voet et al. 2003, Ayabe et al. 2005). Nachteile dieser Methode sind die nur schlecht kontrollierbaren Rekombinationsvorgänge, sowie die immer noch relativ „großen“ Minichromosomen, welche nur schwer exakt zu analysieren sind. Außerdem können diese Minichromosomen nur in einer eukaryotischen Zellkultur vermehrt werden, wodurch eine Produktion in größeren Mengen, wie sie zum Beispiel zur Durchführung einer somatischen Gentherapie erforderlich wäre, praktisch ausgeschlossen ist.

Beim „Bottom-Up“ Verfahren wird isolierte, in Konstrukten präformierte, „nackte“ DNA in die Zielzellen transformiert. Dabei kommt es in einigen Zellen zur Ausbildung einer geeigneten Chromatinstruktur und Formierung stabiler *de novo* Chromosomen (Harrington et al. 1997, Ikeno et al. 1998, Ebersole et al. 2000, Grimes et al. 2002a, Mejía et al. 2002, Laner et al. 2005, Moralli et al. 2006). Diese *de novo* Chromosomen werden auch als künstliche Chromosomen im engeren Sinn bezeichnet. Die „Bottom-Up“ Methode bietet den Vorteil, dass die verwendete DNA flexibel wählbar und modifizierbar sowie vor der Transfektion vollständig sequenzierbar ist. Außerdem lassen sich die DNA-Konstrukte in bakteriellen Wirtssystemen in großer Menge herstellen. Allerdings sind die Vorgänge bei der Ausbildung der Chromatinstruktur, ausgehend von einem nackten DNA-Strang, bislang weitestgehend unverstanden. So ist die Effizienz der HAC-Ausbildung noch gering und die entstehenden Minichromosomen, die in der Regel konkatemerierte Multimere der Input-DNA enthalten (Mejía et al. 2001, Grimes et al. 2002), entsprechen strukturell nicht den ursprünglich eingesetzten DNA-Konstrukten.

1.3. Erste Erfolge in der Entwicklung chromosomaler Vektoren

Das erste *de novo* erzeugte künstliche Chromosom entstand durch Transformation einer Mischung aus α -Satelliten-DNA, Telomer-DNA und genomischer DNA in die menschliche

Sarkom-Zelllinie HT1080 (Harrington et al. 1997). Kurz darauf gelang auch die Konstruktion von HACs durch *in vitro* Ligation der Bestandteile in einem YAC (Yeast artificial chromosome) und anschließende Transfektion in HT1080 Zellen (Ikeno et al. 1998). Diese ersten erfolgreichen Experimente wurden durch eine Reihe weiterer Versuche mit verschiedenen Zentromer-DNAs und unterschiedlichen Vektoren bestätigt (Ebersole et al. 2000, Grimes et al. 2002a, Mejía et al. 2002, Laner et al. 2005, vgl. Kap. 5).

Ein weiterer wichtiger Schritt, hin zu chromosomalen Vektoren, war die Konstruktion von künstlichen Chromosomen, welche humane Genloci enthielten, wie das beim Lesch-Nyhan-Syndrom defekte Hypoxanthin-Guanin-Phosphoribosyltransferase-Gen (HPRT-Gen, Grimes et al. 2001, Mejía et al. 2001) oder das Guanosin-Triphosphat-Cyclohydrolase I-Gen (Ikeno et al. 2002). Wie in den früheren Studien konnte gezeigt werden, dass die Zellen in der Regel nur ein bis zwei künstliche Chromosomen enthielten und keine nachweisbare Menge an Wirts-DNA akquiriert hatten. Die Ausbildung funktioneller Zentromere wurde durch den Immunfluoreszenz-Nachweis von Proteinen wie CENP-A, CENP-C oder CENP-E, welche nur an aktiven Zentromeren vorkommen (Sullivan und Schwartz 1995, vgl. Kap. 4.4), nachgewiesen. Es konnte ferner gezeigt werden, dass die HACs in über 98% der Mitosen an beide Tochterzellen weitergegeben werden und nach zweimonatigem Wachstum (entsprechend circa 60 Zellteilungen) ohne Selektion konnten noch in 35-90% der Wirtszellen künstliche Chromosomen nachgewiesen werden, was die Stabilität der HACs als autonome chromosomale Vektoren unterstreicht. Die Expression des HPRT-Gens wurde sowohl quantitativ mittels Northern-Hybridisierung der HPRT-mRNA als auch funktionell durch HAT-Selektion (Hypoxanthin, Aminopterin, Thymidin) nachgewiesen, was bei fehlerhafter Expression zum Untergang der ursprünglich HPRT-defizienten Zellen führen würde.

Diese Erfolge lassen hoffen, dass es eines Tages möglich sein wird, künstliche Chromosomen als Vektoren für die Transfektion und Expression von Genen zu nutzen. Dabei kommt es zum einen auf einen effizienten Transfer der DNA-Konstrukte in die Wirtszellen an, aber auch auf die Konstruktion aus geeigneten funktionellen chromosomalen Elementen um eine effiziente *de novo* Formierung von Chromosomen zu ermöglichen, ohne dass seltene aktivierende Vorgänge oder Konkatemerisierungen stattfinden müssen.

1.4. Anwendungsbereiche künstlicher Chromosomen

Künstliche Chromosomen besitzen eine Vielzahl von Anwendungsbereichen: Einerseits ermöglichen sie, zusätzlich zur Beobachtung natürlicher Chromosomen, einen direkten

Eingriff in die DNA-Sequenzen um die Struktur und Funktion von Chromosomen sowie die Funktion und Regulation der mit ihnen transfizierten Gene in unterschiedlichen Wirtszellen zu untersuchen. Durch Verwendung geeigneter DNA-Konstrukte können somit die physiologischen und pathologischen Prozesse von Genen, Chromosomen, Zellen, Organen oder ganzen Organismen studiert werden.

Auf der anderen Seite besteht die Möglichkeit künstliche Chromosomen als eukaryotische Vektoren für die Transfektion von kodierender DNA einzusetzen. Die Verwendung geeigneter Gene soll zum Beispiel die Erzeugung transgener Tiere zur Synthese von therapeutischen Proteinen ermöglichen (Kuroiwa et al. 2002, Suzuki et al. 2006). Die transiente Expression bestimmter Proteine könnte aber auch zur Entwicklung neuartiger Impfstoffe auf DNA-Basis genutzt werden (Srivastava und Liu 2003). Die medizinisch verlockendste Einsatzmöglichkeit künstlicher Chromosomen besteht jedoch in der Entwicklung einer somatischen Gentherapie gegen Erbkrankheiten wie zum Beispiel die Mukoviszidose (Zystische Fibrose) (Brown et al. 2000, Choo 2001, Grimes et al. 2002a, Saffery und Choo 2002, Laner et al. 2005).

Die Vorteile der künstlichen Chromosomen gegenüber den bisher hauptsächlich verwendeten viralen Vektoren ergeben sich unter anderem aus deren größerer Kapazität: Bisherige Versuche einer somatischen Gentherapie mittels Vektoren, welche lediglich die cDNA eines Gens sowie einen viralen Promoter enthielten, führten in der Regel nur zu einer kurzzeitig andauernden, unregulierten und schwachen Expression der vielen transfizierten Genkopien (Somia und Verma 2000, Rosenecker und Rudolph 2001, Griesenbach et al. 2002, Schindelhauer und Laner 2002, Klink et al. 2004). Nach kurzer Zeit kam es entweder zum Abbau der transfizierten DNA durch Nukleasen, zur Integration in das Wirtsgenom mit Inaktivierung der Genexpression oder zur Apoptose der transfizierten Zellen. Da mehr als 50% der Sequenzen des humanen Genoms transkribiert werden, führt die zufällige Integration häufig zu einer Beeinträchtigung von Wirtsgenen, oder die transfizierten cDNAs werden unter der Kontrolle von endogenen Promotoren exprimiert. Dies kann im schlimmsten Fall zur Aktivierung von Protoonkogenen und somit zur Entstehung von Neoplasien führen, wie das Auftreten von T-Zell Leukämien bei zwei Patienten nach Gentherapie eines X-chromosomalen schweren kombinierten Immundefekts (X-SCID) zeigt (Hacein-Bey-Abina et al. 2003, Kohn et al. 2003).

Von künstlichen Chromosomen erhofft man sich hingegen, durch Transfektion der kompletten Gene inklusive aller Introns und regulatorischer Sequenzen, eine kontinuierlich

andauernde und physiologisch regulierte Genexpression von einer stabil vererbten Transgen-Kopie. Der Schutz durch Telomere und die autonome Replikation und Segregation der künstlichen Chromosomen verhindern dabei den Abbau der transfizierten DNA und vermindern die Gefahren einer insertionellen Mutagenese (Basu und Willard 2005, Glover et al. 2005).

2. Die Erforschung der Zentromer-DNA besitzt eine vergleichsweise lange

Geschichte

Das Zentromer ist der wichtigste Bestandteil für die Konstruktion von künstlichen Chromosomen. So ist es bisher nicht gelungen die DNA-Sequenz eines Zentromers zu entschlüsseln, und die Zentromerregionen sind nach wie vor „schwarze Löcher“ in der Karte des menschlichen Genoms (Henikoff 2002).

Dabei besitzt die Erforschung der Zentromer-DNA eine für molekulargenetische Verhältnisse lange Geschichte: Bereits Ende der fünfziger Jahre des letzten Jahrhunderts, das Vorliegen der DNA in Form einer Doppelhelix war gerade erst entschlüsselt, hatten theoretische Berechnungen ergeben, dass eine spontane Reassoziaton der komplementären Einzelstränge von thermisch denaturierter, genomischer DNA zur Doppelhelix so gut wie unmöglich ist (Marmur et al. 1963). Dennoch zeigten die Experimente, dass es bei der Abkühlung denaturierter DNA innerhalb kürzester Zeit zur Neubildung von DNA-Doppelsträngen kommt. Die Schlussfolgerung war, dass die Genome der höheren Organismen einen großen Anteil an repetitiven DNA-Sequenzen besitzen müssen (Britten und Kohne 1968).

Etwa zur gleichen Zeit wurde festgestellt, dass die Zentrifugation von kompletter genomischer DNA in einem Cäsiumchlorid-Gradienten neben einer breiten Hauptbande zur Ausbildung schmaler sogenannter Satelliten-Banden führt, deren DNA eine geringere Dichte besitzt als die der Hauptbande (Meselson et al. 1957, Sueoka et al. 1959, Kit 1961). Die Isolierung und Messung der Reassoziations-Kinetik dieser DNA ergab, dass die Satelliten-Banden fast ausschließlich aus schnell reassoziierender repetitiver DNA bestehen (Waring und Britten 1966).

Einige Jahre später zeigten Hybridisierungsexperimente mit radioaktiv markierten Nukleinsäuren, dass ein großer Teil dieser repetitiven Satelliten-DNA an den Zentromerregionen aller menschlicher Chromosomen zu finden ist (Pardue und Gall 1970, Jones und Corneo 1971).

Aufgrund ihrer spezifischen physikalischen Eigenschaften, des Vorkommens in hunderten bis tausenden beinahe identischer Kopien pro Genom, und der Möglichkeit diese durch Verdau mit geeigneten Restriktionsenzymen zu isolieren (Manuelidis 1976, Horz und Zachau 1977, vgl. Kap. 3.3), war die Satelliten-DNA der Zentromere dann eine der ersten Nukleinsäuren, welche einer Sequenz-Analyse unterzogen werden konnte (Southern 1970, Rosenberg et al. 1977).

Seit diesen ersten DNA-Sequenzierungen haben sich die molekularbiologischen Techniken rapide weiterentwickelt. Am Höhepunkt des Wettrennens um die Sequenzierung des Humangenoms im Juni 2000 wurden allein durch die am öffentlichen Humangenomprojekt beteiligten Center circa 1000 Basen pro Sekunde bestimmt (Lander et al. 2001). Im Oktober 2004 wurde die endgültige Version des Humangenoms veröffentlicht (Lander et al. 2004) und auch die Bestimmung von Polymorphismen im menschlichen Genom sowie die Sequenzierungen der Genome vieler weiterer Organismen schreiten rasch voran. Trotz all der Fortschritte ist es jedoch bisher nicht gelungen die Sequenz eines Zentromers zu bestimmen (Eichler et al. 2004, Rudd und Willard 2004), und auch die evolutionäre Entwicklung, die Organisation des Chromatins oder die Bedeutung der primären DNA-Sequenz für die Funktion der Zentromere sind weiterhin ungeklärt.

3. Aufbau und Funktion der Zentromere

3.1. Zentromere sind essentielle Bestandteile eukaryotischer Chromosomen

Die Zentromere besitzen essentielle Funktionen sowohl während der mitotischen als auch während der meiotischen Zellteilung. Durch sie wird die exakte Aufteilung der Chromosomensätze auf die Tochterzellen und somit die Vermehrungsfähigkeit der Zellen gewährleistet.

Das Zentromer wurde erstmals 1880 von Walther Flemming als die während der Metaphase der Zellteilung sichtbare „primäre Konstriktion“ der Chromosomen beschrieben (Flemming 1880). Sie stellt denjenigen Punkt dar, an dem die Chromatiden eines Chromosoms nach der DNA-Replikation noch miteinander verbunden sind bis die Zellen in die Mitose eintreten. In der Prometaphase der Mitose werden die noch verbundenen Chromosomen dann durch Zug von an den Kinetochoren anhaftenden Mikrotubuli in die Mitte zwischen die beiden Spindelpole verlagert, wodurch die sogenannte Metaphasenplatte entsteht. Erst wenn alle Chromosomen in der Äquatorialebene angeordnet sind, kommt es zur raschen Trennung der Chromatiden. Die Zentromere besitzen somit eine wichtige Funktion als „Cycle-Checkpoint“

im Ablauf des Zellzyklus, wodurch sichergestellt wird, dass bei der Zellteilung immer vollständige Chromosomensätze auf die Tochterzellen verteilt werden. In der Anaphase kommt es schließlich zu einer entgegengesetzten Bewegung der Chromatiden, hin zu den Spindelpolen, bevor die Zellteilung mit der Telophase und der Zytokinese abgeschlossen wird (Dobie et al. 1999).

Kommt es bei der Aufteilung der Chromatiden in der Mitose oder der Aufteilung der Chromosomen in der Meiose zu Fehlern, entstehen sogenannte aneuploide Chromosomensätze. Man vermutet, dass die Rate an numerischen Chromosomenaberrationen von befruchteten menschlichen Eizellen bei über 20% liegt. Dieses Auftreten zusätzlicher oder Fehlen einzelner Chromosomen hat meist fatale Folgen: Circa 50% aller spontanen Aborte nach der siebten Schwangerschaftswoche sind auf Aneuploidien zurückzuführen (Jacobs und Hassold 1995). Bei 0,53% aller Neugeborenen liegt eine numerische Anomalie des Chromosomensatzes vor (Burgoyne et al. 1991). Die häufigste derartige chromosomale Aberration ist die Trisomie 21 (Down-Syndrom), welche bei einem von 700 Neugeborenen auftritt. Aber auch Erkrankungen wie zum Beispiel das Klinefelter- oder das Ullrich-Turner-Syndrom sind die Folge von Fehlverteilungen einzelner Chromosomen.

Auch bei der Entwicklung von Neoplasien kommt es manchmal zur Ausbildung aneuploider oder polyploider Chromosomensätze. Deren Vorliegen hat häufig einen signifikanten Einfluss auf den Verlauf und die Therapiemöglichkeiten einer Tumorerkrankung und somit deren Prognose (Hartwell und Kastan 1994). Diese Beispiele zeigen deutlich, wie entscheidend die Zentromere für die Zellteilung und somit für die Überlebensfähigkeit aller mehrzelligen Organismen sind.

3.2. Die Protein-Struktur der menschlichen Zentromere

Eine der frühesten Markierungen der replizierten Zentromer-DNA im Bereich der primären Konstriktion stellt die Beladung mit CENP-A, einem Histon H3-Homolog, dar. Daran lagern sich im Verlauf des Zellzyklus weitere Proteine an, bis sich in der M-Phase die vollständigen Kinetochore mit einer Größe von 0,3-0,5 μm ausgebildet haben (Benninghoff 1994). Diese Proteinkomplexe bilden den Ansatzpunkt für die Mikrotubuli des Spindelapparats, sowie den „Motor“ durch den sich die Chromatiden entlang der Mikrotubuli fortbewegen. Die Kinetochore bestehen morphologisch aus drei Lamellen, die als äußere, mittlere und innere Lamelle bezeichnet werden.

Die ersten am Aufbau der Zentromere beteiligten Proteine wurden über Antikörper im Serum von Patienten identifiziert, welche am CREST-Syndrom litten, einer seltenen limitierten Verlaufsform der systemischen Sklerodermie. Dieses Syndrom manifestiert sich unter anderem in einer Calcinosis cutis, Raynaud-Symptomatik, Ösophagus (engl. esophagus)-Motilitätsstörungen, Sklerodaktylie sowie Teleangiektasien. Die beim CREST-Syndrom auftretenden Antizentromer-Antikörper (ACA) sind hochspezifisch, wohingegen die für die diffuse Form der systemischen Sklerodermie typischen anti-Scl 70 Antikörper nicht vorkommen (Haustein 2002). Durch Analyse der Antigene, gegen welche die ACAs gerichtet sind, wurde eine Reihe von Zentromer-Proteinen (CENP) mit zunächst noch unbekannter Funktion identifiziert (Moroi et al. 1980, Earnshaw und Rothfield 1985, Saitoh et al. 1992). Funktionell lassen sich die Zentromer-Proteine zwei Gruppen zuordnen: Einerseits diejenigen, welche konstitutiv mit dem Zentromer assoziiert sind und vor allem am Aufbau des Chromatins und der inneren Lamelle beteiligt sind, wie zum Beispiel die Zentromer-Proteine CENP-A, -B, -C und -G; andererseits die Proteine, welche nur während der Mitosephase an der Bildung der Kinetochore beteiligt sind, wie zum Beispiel die Proteine CENP-E, CENP-F, INCENP oder Dynein (Saffery et al. 2000).

Eine Übersicht über einige der am Aufbau der Zentromere beteiligten Proteine und deren vermutliche Funktion findet sich in Anhang 1 (s. auch Maiato et al. 2004).

3.3. α -Satelliten-DNA an menschlichen Zentromeren

Die DNA im Bereich der Zentromere aller menschlicher Chromosomen besteht aus sogenannten α -Satelliten welche circa 2-5% des menschlichen Genoms ausmachen (Pardue et al. 1970, Jones und Corneo 1971, Manuelidis 1976, Wu und Manuelidis 1980). Die α -Satelliten-Monomere besitzen eine Länge von circa 171 bp und bilden tandemförmig aneinander gereiht Arrays aus tausenden von Repeats (Abb. 2). Durch Sequenzvergleiche

Tabelle 1: Zugehörigkeit der α -Satelliten-Arrays der menschlichen Chromosomen zu den suprachromosomalen Familien 1-5

Suprachromosomale Familie	Vorläuferrepeat	Typ-A Monomere	Typ-B Monomere	Chromosomen
1	J1-J2	J1	J2	1, 3, 5, 6, 7, 10, 12, 16, 19
2	D1-D2	D2	D1	2, 4, 8, 9, 13, 14, 15, 18 (2 Arrays), 20, 21, 22
3	W1-W2-W3-W4-W5	W4, W5	W1, W2, W3	1, 11, 17, X
4	M1	M1		Y, viele weitere
5	R1, R2 (irregulär)	R2	R1	viele

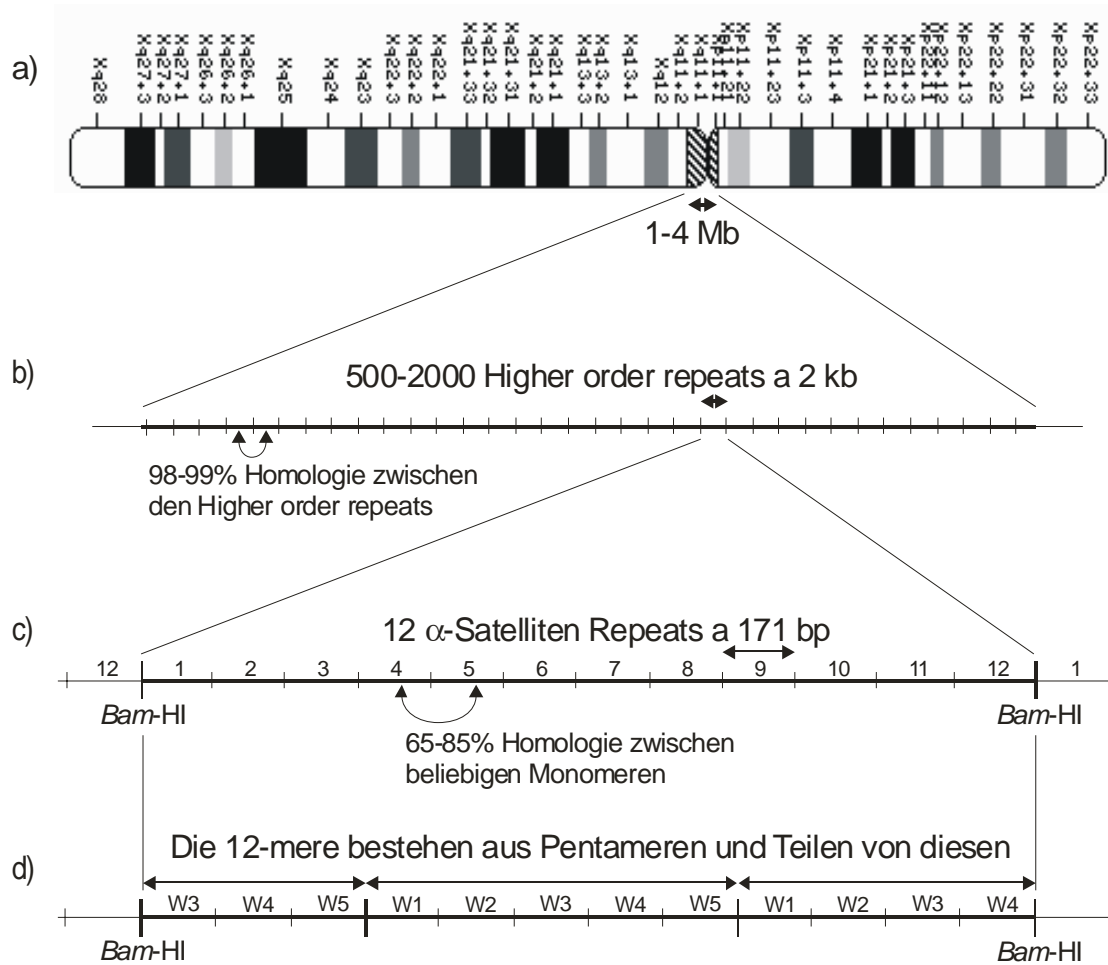


Abb. 2: Der Aufbau der Zentromer-DNA am Beispiel des humanen, X-chromosomalen α -Satelliten Arrays DXZ1: (a) Die Zentromerregion des X-Chromosoms ist 1-4 Mb groß und besteht aus (b) tandemartig aneinandergereihten, circa 2 kb langen Higher order repeats mit Homologien von 98-99%. (c) Diese wiederum bestehen aus 12 circa 171 bp langen α -Satelliten Repeats mit 65-85% Homologie. (d) Der α -Satelliten Array des X-Chromosoms gehört zur suprachromosomalen Familie 5, deren Higher order repeats aus Pentameren und Teilen von diesen bestehen.

lassen sich die Monomere zwei Gruppen zuordnen und werden dementsprechend als Typ-A und Typ-B Monomere bezeichnet, wobei lediglich die Typ-B Monomere eine CENP-B Box, also eine Bindungsstelle für das Zentromer-Protein B besitzen (vgl. Kap. 4.2). Homologievergleiche ermöglichen ferner ein Einordnen der α -Satelliten in dimerische und pentamerische Familien. Je nach Aufbau aus Mono- Di- oder Pentameren lassen sich die Zentromer-Arrays so einer von fünf sogenannten suprachromosomalen Familien zuordnen. Die Arrays der suprachromosomalen Familien eins und zwei sind aus α -Satelliten-Dimeren aufgebaut, diejenigen der suprachromosomalen Familie drei aus Pentameren, die der Familie vier aus Monomeren, und die Arrays der Familie fünf aus zwei unterschiedlichen, irregulär angeordneten Repeats (Tab. 1, vgl. Kap. 6.1). Die Zentromer-Regionen der meisten

Chromosomen enthalten mehrere α -Satelliten-Arrays, wobei es sich in der Regel um einen evolutionär neueren Array der suprachromosomalen Familien 1-3 sowie „Überreste“ von älteren Arrays der suprachromosomalen Familien 4 und 5 handelt

α -Satellit (Monomer): Kleinste Einheit an repetitiver DNA im Bereich der menschlichen Zentromere mit einer Länge von circa 170 bp. Die α -Satelliten Monomere weisen Übereinstimmungen von 60-90% auf.

Higher order repeat: Einheit der beinahe identischen Wiederholung in einem α -Satelliten Array. Bestehen beim Menschen aus 2-34 Monomeren und weisen typische Homogenitäten von >97% auf.

Tandem-Repeat Array: Ununterbrochene Aneinanderreihung von unidirektional angeordneten α -Satelliten, meist hunderte bis tausende von Higher order repeats.

(Waye et al.1987, Finelli et al 1996, Alexandrov et al. 2001, Rosandic et al. 2006).

Mehrere Mono-, Di- oder Pentamere lassen sich wiederum zu den eigentlichen Repeat-Strukturen der heutigen Zentromere zusammenfassen, welche als Higher order repeats bezeichnet werden (Abb. 2c). Diese bestehen je nach α -Satelliten-Array aus 2-34 Monomeren und besitzen demnach Längen zwischen 0,34 und 6,0 kb. Die Sequenzen der einzelnen α -Satelliten-Repeats innerhalb eines Chromosoms sind zu 60-90% identisch. Die Higher order repeats dagegen stimmen in den homogenen Bereichen der Zentromere zu über 97% überein (Waye und Willard 1987, Willard und Waye 1987, Alexandrov et al. 1993, Rudd und Willard 2004). Die Längen der Higher order repeats lassen sich durch bestimmte Restriktionsendonukleasen ermitteln, welche regelmäßig genau einmal innerhalb der Repeats schneiden (Abb. 2c, Horz und Zachau 1977). Übersichten über die Higher order repeat Längen der menschlichen Zentromere sowie deren charakteristische Restriktionsenzyme sind bei Choo et al. (1991), Lee et al. (1997), Choo (1997) oder Alexandrov et al. (2001) zu finden.

Die Länge der gesamten α -Satelliten-Arrays variiert zwischen 200 kb und 5 Mb (Lo et al. 1999), wobei signifikante Unterschiede zwischen den nicht-homologen Chromosomen bestehen. So beträgt die Länge des Zentromer-Arrays des Chromosoms 11 zwischen 2,0 und 4,8 Mb (Mahtani und Willard 1990), die des Y-Chromosoms dagegen lediglich 0,3-1,0 Mb (Tyler-Smith und Brown 1987). Zwischen den homologen Chromosomen nicht-verwandter Individuen variieren die Array-Längen ebenfalls relativ stark. Bestimmt man jedoch die Längen der Zentromer-Arrays in mehreren aufeinander folgenden Generationen einer Familie, zeigt sich (im Rahmen der Messgenauigkeit der Pulsfeld-Gelelektrophorese) eine weitgehend stabile Vererbung (Jabs et al. 1989, Wevrick und Willard 1989, Mahtani und Willard 1990).

Dies lässt darauf schließen, dass es relativ selten zu messbaren Veränderungen der Array-Längen kommt, diese jedoch sehr drastisch ausfallen können.

Insbesondere das Chromosom 21 ist vergleichsweise häufig von extremen Verkürzungen seines Zentromer-Arrays auf bis zu 50 kb betroffen, und bei Patienten mit einer Trisomie 21 wurden derartige Verkürzungen häufiger beobachtet als in den Kontrollen (Lo et al. 1999). Ob jedoch die Paarung von homologen Chromosomen 21 mit stark unterschiedlichen Array-Längen in der Meiose I einen ätiologischen Faktor für die Entstehung der Trisomie 21 darstellt ist unklar.

3.4. Der DXZ1-Lokus des menschlichen X-Chromosoms

Der α -Satelliten-Array des menschlichen X-Chromosoms DXZ1 in seiner heutigen Form entstand vor circa sieben Millionen Jahren zur Zeit der Trennung der Orangutan- und der Gorilla-Linien (Malik und Henikoff 2002). Er besitzt eine Länge von 1 - 4 Mb (Mahtani und Willard 1990) und besteht fast ausschließlich aus unidirektional aneinandergereihten Higher order repeats von circa 2065 bp Länge, welche aus 12 α -Satelliten-Monomeren bestehen (Abb. 2, Yang et al. 1982, Wayne und Willard 1985). Durch die Restriktionsenzyme *Bam*-HI, *Pst*-I sowie *Sst*-I ist das DXZ1-Array in seine charakteristischen 12-mere spaltbar (Choo et al. 1991). Der Higher order repeat gehört der suprachromosomalen Familie drei an und besteht aus zwei kompletten Pentameren sowie einem Pentamerfragment aus zwei Monomeren (Monomerstruktur (W3-W4-W5-W1-W2) - (W3-W4-W5-W1-W2) - (W3-W4)). Die Monomere 1, 4, 9 und 10 (bei Beginn der Zählung an der *Bam*-HI Schnittstelle der Higher order repeats) enthalten in der Regel vollständige CENP-B Boxen (vgl. Kap. 4.2). In Hybridisierungsversuchen konnte gezeigt werden, dass der DXZ1 2 kb Higher order repeat spezifisch für das X-Chromosom ist. Die dem X-chromosomalen Higher order repeat am nächsten verwandten menschlichen α -Satelliten finden sich auf den Chromosomen 1, 11 und 17 (Yang et al. 1982).

Alle bisher untersuchten humanen X-chromosomalen Zentromere enthalten ferner einen geringen aber konstanten Anteil an 3 kb Higher order repeats. Die Analyse dieses Repeat-Typs durch direkte Sequenzierung von PCR-Produkten aus Maus/Mensch-Zellhybriden zeigte, dass der 3 kb Higher order repeat vermutlich durch Unequal crossing over zwischen einem Monomer 2 oder 3 und einem Monomer 7 oder 8 zweier 12-mere entstanden ist (Warburton und Willard 1992). Er besitzt damit exakt ein Pentamer mehr als die 2 kb Repeats

und weist die Monomerstruktur (W3-W4-W5-W1-W2) - (W3-W4-W5-W1-W2) - (W3-W4-W5-W1-W2) - (W3-W4) auf. Es kann allerdings nicht ausgeschlossen werden, dass 17-mer Higher order repeats existieren, welche aus anderen Rekombinationsereignissen hervorgingen.

Sequenzanalysen einzelner 2 kb Higher order repeats ergaben, dass die Monomere untereinander Homologien von 65-85% aufweisen (Waye und Willard 1985), während die Higher order repeats in der Regel zu >97% identisch sind (Durfy und Willard 1989, Warburton und Willard 1990). Die Variabilität zwischen den Higher order repeats wird im Wesentlichen durch variable Basen verursacht, Positionen also, an denen zwei oder mehr Nukleotide alternativ vorkommen (Durfy und Willard 1987, Durfy und Willard 1989, Warburton und Willard 1992). Die Sequenzierung von PCR-Produkten der α -Satelliten-Repeats mehrerer humaner X-chromosomaler Zentromere zeigte, dass die Positionen und relativen Verhältnisse der Varianten in den untersuchten Individuen konserviert sind. Eine exakte Untersuchung der Verteilung und Häufigkeit der Varianten innerhalb eines kontinuierlichen α -Satelliten-Arrays konnte bislang jedoch nicht durchgeführt werden.

2001 veröffentlichten Schueler et al. eine physikalische Karte der Region zwischen dem auf dem kurzen Arm des X-Chromosoms am weitesten proximal gelegenen bekannten Gen ZXDA und dem homogenen Bereich des DXZ1-Arrays. Zentromerwärts des Zink-Finger-Gens ZXDA folgt ein circa 150 kb langer Abschnitt, welcher zu 90% aus repetitiven DNA-Sequenzen, vor allem L1-Elementen, besteht. Daran schließen sich ca. 200 kb divergierte α -Satelliten an, welche am ehesten der suprachromosomalen Familie 4 zugeordnet werden können. Es folgen circa 50 kb an γ -Satelliten (isolinear angeordnete Tandem-Repeats von circa 220 bp Länge) und weitere 175 kb α -Satelliten-DNA, bevor das eigentliche DXZ1-Array mit seinen charakteristischen 2 kb Higher order repeats beginnt. Am Übergang der divergierten α -Satelliten-Monomere zu den 12-meren des DXZ1-Arrays nimmt die Homologie der α -Satelliten-Repeats über einen Bereich von circa 20 kb kontinuierlich von etwa 70% auf die DXZ1 typischen 98-99% zu. Die distalen 265 kb dieses insgesamt 575 kb langen DNA-Abschnitts mit α - und nicht- α -Satelliten-Sequenzen wurden im Rahmen der Sequenzierung des Humangenoms durch die Firma Celera sequenziert (Venter et al. 2001). Dabei ergaben sich keine Hinweise auf bislang unbekannte Gene oder single-copy Sequenzen, welche eine Rolle bei der Funktion der Zentromere spielen könnten.

4. Sequenz-Funktions Relation von α -Satelliten-DNA

Eine kontrovers geführte Debatte betrifft die Bedeutung der primären DNA-Sequenz der α -Satelliten für die Lokalisation und Funktion der Zentromere. Es stellt sich die Frage, ob Zentromere primär durch bestimmte Sequenz-Eigenschaften der ihnen zugrunde liegenden DNA definiert werden, oder epigenetisch durch ihre spezifische Chromatinstruktur (Karpen und Allshire 1997, Choo 1998, Murphy und Karpen 1998, Choo 2000, Cleveland et al. 2003, Schueler und Sullivan 2006).

4.1. Die Zentromer-Regionen aller höherer Eukaryoten enthalten repetitive DNA, sind jedoch evolutionär nicht konserviert

Für eine spezifische Funktion der DNA-Sequenz bei der Ausbildung eines Zentromers spricht, dass sowohl alle höheren Eukaryoten als auch viele einfachere Organismen wie zum Beispiel die Spalthefe *Schizosaccharomyces pombe* im Bereich der Zentromere Arrays aus sich wiederholenden Sequenzen besitzen (Koch 2000, Amor und Choo 2002). Dies trifft auch auf Organismen zu deren Genom sonst relativ arm an repetitiven Sequenzen ist, wie zum Beispiel die Ackerschmalwand *Arabidopsis thaliana* (Copenhaver et al. 1999). Im menschlichen Genom ist α -Satelliten-DNA der einzige DNA-Typ, welcher ausschließlich an allen natürlichen Zentromeren vorkommt (Murphy und Karpen 1998).

Demgegenüber unterscheiden sich sowohl die Sequenzen als auch die Längen der Zentromer-Repeats von Spezies zu Spezies sehr stark, teilweise sogar zwischen eng verwandten Spezies und zwischen den heterologen Chromosomen innerhalb einer Spezies (vgl. Kap. 6.2). Dabei konnten in den Zentromer-Repeats mit Ausnahme der CENP-B Box keine zwischen mehreren Spezies homologe und somit evolutionär konservierte Abschnitte gefunden werden, welche auf eine direkte Sequenz-Funktions Relation hinweisen würden (Choo 1997a, Koch 2000, Henikoff 2002).

Da es jedoch bislang nicht gelungen ist den α -Satelliten-Array eines Zentromers vollständig zu sequenzieren, konnte die Existenz einer „magischen Sequenz“, welche als Ausgangspunkt für die Entstehung des Kinetochors dienen könnte, nicht mit Sicherheit ausgeschlossen werden. Sie könnte als relativ kurze, evolutionär konservierte und für die Zentromer-Funktion essentielle single-copy Sequenz bislang unerkannt innerhalb der α -Satelliten-Arrays liegen. Zum Beispiel enthalten alle 16 Zentromere der Hefe *Saccharomyces cerevisiae*, welche nur 125 bp lang sind, das 25 bp lange Element CDEIII mit Proteinbindungskapazität. Bereits einzelne Basensubstitutionen in diesem DNA-Abschnitt können die Zentromerfunktion

vollständig inaktivieren (Clarke 1990). Jedoch unterscheiden sich die Punktzentromere der Hefe mit einer einzigen Spindelfaser wesentlich von den regionalen Zentromeren der Säuger, bei welchen 20-30 Spindelfasern an ein über bis zu mehrere Mb reichendes Kinetochor binden. Dieser Größenunterschied, *in situ* Hybridisierungen humaner Zentromere (Finelli et al. 1996), die vollständige Sequenzierung menschlicher Neozentromere (Barry et al. 2000a, Barry et al. 2000b, Maggert und Karpen 2000, vgl. Kap. 4.5) sowie die *de novo* Konstruktion von künstlichen Chromosomen unter Verwendung von synthetischen α -Satelliten-Repeats (Harrington et al. 1997, Basu et al. 2005), machen die Existenz einer derartigen „magischen“ Einzelsequenz in den Zentromeren der höheren Organismen jedoch sehr unwahrscheinlich.

Ein weiterer Hinweis darauf, dass eine derartige Sequenz nicht existiert, und dass die primäre DNA-Sequenz keine wesentliche Rolle für die Ausbildung von Zentromeren spielt, sind die holozentrischen Chromosomen des Fadenwurms *Caenorhabditis elegans* (Albertson und Thomson 1982): Obwohl das Genom dieses Nematoden vollständig sequenziert ist (The *C. elegans* sequencing consortium 1998) konnten keine besonderen Satelliten-DNA Abschnitte identifiziert werden, welche eine spezifische Anheftung von Mikrotubuli induzieren könnten (Pluta et al. 1995, Tyler-Smith und Florida 2000).

Dass perizentrisch lokalisierte DNA-Sequenzen bei der Zentromer-Bildung wesentlich beteiligt sind, konnten unter anderem Schueler et al. (2001) anhand von Isochromosomen bei Patientinnen mit Ulrich-Turner-Syndrom widerlegen. Sie beschrieben jeweils drei Patientinnen, welche neben einem karyotypisch unauffälligen ein X-Chromosom aus zwei langen respektive zwei kurzen Armen besaßen. Die „Spiegelebene“ der Chromosomen lag dabei innerhalb des α -Satelliten-Arrays DXZ1. Die scheinbar unbeeinträchtigte Replikation dieser Isochromosomen deutet darauf hin, dass keine essentielle Sequenz in einer der perizentrischen Regionen liegt (s. auch Tyler-Smith et al. 1993).

4.2. Die CENP-B Box besitzt proteinbindende Eigenschaften

Der bislang einzig bekannte Sequenzabschnitt in den menschlichen α -Satelliten-Repeats mit einer spezifisch proteinbindenden Eigenschaft ist die CENP-B Box, eine 17 bp lange Sequenz mit dem Motiv PyTT CGT TGG AAPu CGG GA welche als Bindungsstelle für das Zentromer-Protein B (CENP-B) dient (Masumoto et al. 1989, Muro et al. 1992). Diese CENP-B Box findet sich in den α -Satelliten-Repeats des Menschen und der höheren Affen (Haaf et al. 1995) sowie in der 120 bp langen Minor Satellite DNA der Mausspezies *Mus musculus* und *Mus caroli* (Kipling et al. 1995).

Die genaue Funktion des 80 kDa großen Proteins CENP-B, welches mit seinem aminoterminalen Ende an die CENP-B Box bindet und mit dem carboxyterminalen Ende Homodimere bildet (Muro et al. 1992), ist noch unbekannt. So werden Homologe dieses Proteins auch von den Zellen der afrikanischen grünen Meerkatze *Cercopithecus aethiops* (Goldberg et al. 1996) oder der Hefe *Schizosaccharomyces pombe* (Irelan et al. 2001, Nakagawa et al. 2002) exprimiert, obwohl die Zentromere der niederen Affen oder der Hefe keine CENP-B Boxen enthalten. Und auch die α -Satelliten-Repeats des menschlichen Y-Chromosoms (Tyler-Smith und Brown 1987, Masumoto et al. 1989) oder Neozentromere (vgl. Kap. 4.5) enthalten weder CENP-B Boxen, noch das Protein CENP-B als Bestandteil ihres Chromatins (Saffery et al. 2000). In zwei Studien mit CENP-B knock-out Mäusen konnte keinerlei Effekt auf die Entwicklung der Mäuse oder auf die Zellteilung festgestellt werden (Kapoor et al. 1998, Perez-Castro et al. 1998). In einer dritten Studie wurden ein geringgradig vermindertes Körpergewicht sowie eine verminderte Größe der Testes festgestellt (Hudson et al. 1998), während in einer vierten Studie Dysfunktionen des Genitaltrakts der CENP-B defizienten Mäuse beobachtet wurden, die möglicherweise auf den CENP-B Mangel zurückzuführen sind (Fowler et al. 2000).

Auch die Proteine CENP-A und CENP-C sind in der Lage DNA zu binden, es existiert jedoch keine spezifische Bindungssequenz (Yang et al. 1996, Choo 2000, Politi et al. 2002).

4.3. Verschiedene Mechanismen der epigenetischen Markierung der Zentromer-DNA werden diskutiert

Unter einer „epigenetischen Determination“ versteht man eine Vererbung von Gen- oder chromosomalen Funktionen ohne korrespondierende Veränderungen der primären DNA Sequenz (Sullivan et al. 2001). Beispiele für derartige epigenetische Einflüsse sind die X-Inaktivierung eines der beiden X-Chromosomen in Zellen weiblicher Individuen oder Genomic imprinting bei Erkrankungen wie zum Beispiel dem Angelman- und dem Prader-Willi-Syndrom.

Verantwortlich für die Ausbildung eines Zentromers an einem bestimmten chromosomalen Lokus im Sinne einer epigenetischen Determination könnte zum Beispiel eine spezifische Chromatin-Struktur unter Beteiligung bestimmter Histone oder zentromer-spezifischer Proteine sein. Ein Protein, welches hierbei eine Schlüsselrolle spielt ist CENP-A, ein Histon H3-Homolog, welches ausschließlich an aktiven Zentromeren vorkommt und an der Verbindung der Kinetochore mit dem DNA-Strang beteiligt ist (Sullivan et al. 1994, Vafa und

Sullivan 1997, Sullivan 2001, Masumoto et al. 2004). Neuere Untersuchungen haben gezeigt, dass lediglich künstliche Chromosomen mit einer Größe von über 3 Mb Heterochromatin enthalten, wohingegen auch kleine künstliche Chromosomen CENP-A enthaltendes Euchromatin besitzen. Dies deutet darauf hin, dass nicht das in großen Mengen an allen natürlichen humanen Chromosomen vorkommende Heterochromatin für die Zentromer-Funktion ausschlaggebend ist, sondern kleine, CENP-A enthaltende euchromatische Bereiche (Grimes et al. 2004, Lam et al. 2006, Okamoto et al. 2007).

Aber auch chemische Modifikationen der DNA wie Methylierung (Mitchell et al. 1996), Acetylierung (Ekwall et al. 1997, Nakano et al. 2003), Phosphorylierung, Ubiquitinierung oder Poly-ADP-Ribosylierung (Choo 2000) wurden als Mechanismen der epigenetischen Markierung von Zentromer-DNA diskutiert.

Als weiterer potentieller Mechanismus für die chromosomale Lokalisation der Kinetochor-Bildung wurde die späte Replikation der Zentromer-DNA vorgeschlagen (Csink und Henikoff 1998). Neuere Ergebnisse zeigten jedoch, dass die Replikation zumindest der Hauptmenge der Mb-langen α -Satelliten-Arrays gemeinsam mit vielen weiteren chromosomalen Regionen in der mittleren bis späten S-Phase erfolgt (Ten Hagen et al. 1990, O' Keefe et al. 1992, Shelby et al. 2000, Sullivan und Karpen 2001), die Replikation von lediglich Euchromatin enthaltenden, kleinen künstlichen Chromosomen gar in der frühen S-Phase (Grimes et al. 2004). Demgegenüber liegt der Höhepunkt der Expression von CENP-A in der G2-Phase des Zellzyklus (Shelby et al. 2000) was darauf hindeutet, dass die Ausbildung der zentromer-typischen Chromatinstruktur ebenfalls in der G2-Phase erfolgt und somit nach der DNA-Replikation stattfindet.

4.4. Dizentrische Chromosomen

Hinweise auf eine epigenetische Determination der Zentromer-Aktivität ergeben sich unter anderem aus dem Vorkommen dizentrischer Chromosomen. Durch Rekombination entstehen hierbei Chromosomen mit zwei Arrays aus repetitiver α -Satelliten-DNA, getrennt durch nicht-repetitive euchromatische DNA-Sequenzen (Abb. 3a). Derartige Rekombinationen liegen beispielsweise bei 15% aller Patienten mit Ullrich-Turner Syndrom vor (Sullivan und Willard 1998). Bei Aktivität beider Zentromer-Regionen kommt es in der Anaphase der Mitose zur Ausbildung von Chromatin-Brücken, was in der Regel zum Verlust dieser Chromosomen führt. Wird jedoch eines der beiden Zentromere inaktiviert können stabile dizentrische Chromosomen entstehen, welche auch als pseudodizentrische Chromosomen

bezeichnet werden. Diese Inaktivierung muß durch epigenetische Modifikationen erfolgen, da die DNA-Sequenz der α -Satelliten-Arrays hierbei offenbar nicht verändert wird. Die zentromer-bindenden Proteine CENP-A, CENP-C und CENP-E sind bei pseudodizentrischen Chromosomen nur am aktiven Zentromer nachweisbar, wohingegen sich CENP-B sowohl am aktiven als auch am inaktiven Zentromer anlagert (Sullivan und Schwartz 1995, Sullivan und Willard 1998, Sullivan et al. 2001; Anhang 1). Auf diese Weise konnte gezeigt werden, dass bei aufeinander folgenden Zellteilungen in der Regel derselbe α -Satelliten-Array den Ausgangspunkt für die Kinetochorbildung und die primäre Konstriktion bildet (Earnshaw et al. 1989, Fisher et al. 1997). Bei einem geringen Abstand der beiden α -Satelliten-Arrays kommt es jedoch auch zur Ausbildung von stabilen, funktionell monozentrischen Chromosomen mit zwei aktiven Zentromerregionen, (Wandall 1994, Sullivan und Willard 1998) und in seltenen Fällen gar zu einem Wechseln der Zentromeraktivität zwischen zwei α -Satelliten-Arrays (Dewald et al. 1979, Ing und Smith 1983, Rivera et al. 1989).

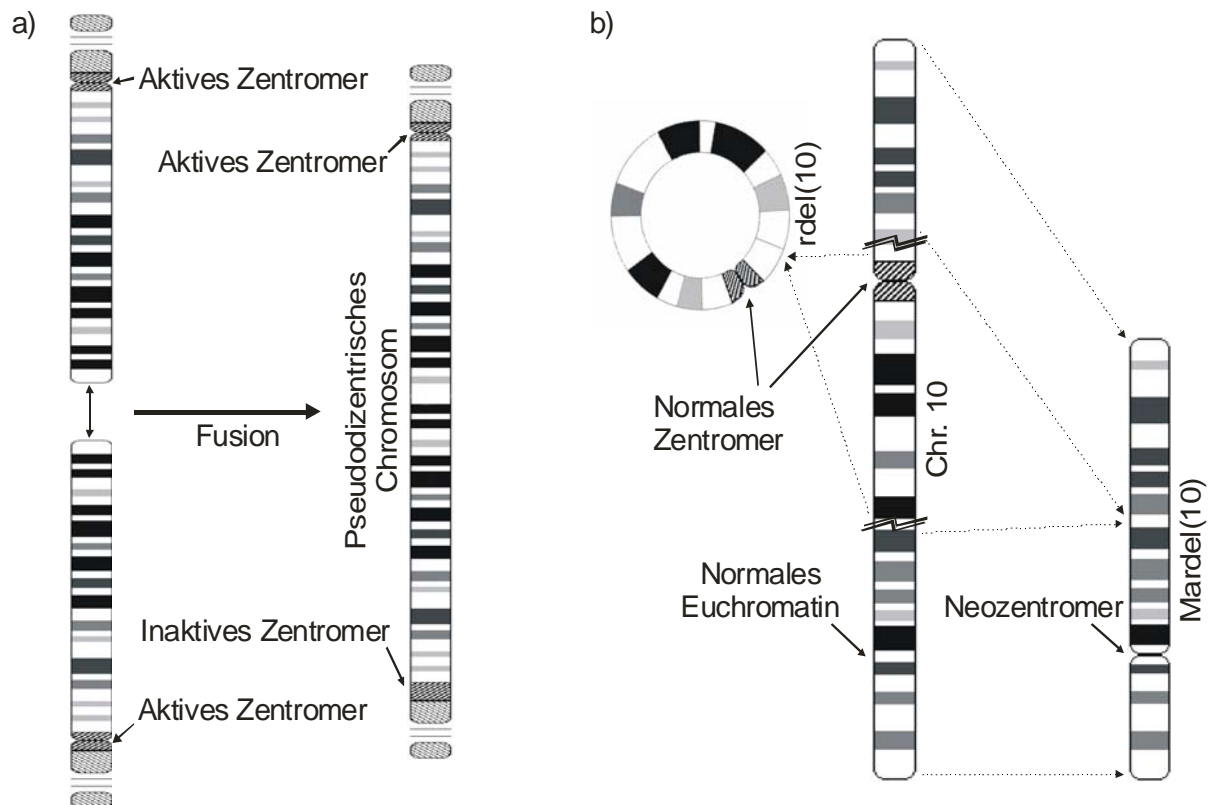


Abb. 3: Beispiele für die epigenetische Determination der Zentromer-Aktivität

- a) Ausbildung eines pseudodizentrischen Chromosoms durch Fusion zweier Chromosomen mit epigenetischer Inaktivierung eines Zentromers bei unverändertem α -Satelliten Array.
- b) Ausbildung eines Neozentromers ohne α -Satelliten DNA in der Region 10q25.2 nach Bruch eines Chromosoms 10 mit Ausbildung eines Ringchromosoms.

4.5. Neozentromere

Ein weiteres Argument für die epigenetische Determination der Zentromerlokalisierung ist die Existenz von Neozentromeren. Dabei kommt es, meist als Folge von Chromosomenbrüchen oder Rearrangements größerer DNA-Abschnitte, zur Ausbildung von aktiven Zentromeren an Chromosomenarmsequenzen ohne α -Satelliten-DNA. Es wurden bereits über 70 Neozentromere auf nahezu allen menschlichen Chromosomen beschrieben (Voullaire et al. 1993, Choo 1997a, Warburton et al. 2000, Amor und Choo 2002, Warburton 2004). Tyler-Smith et al. (1999) berichteten über die Ausbildung eines Neozentromers auf dem langen Arm eines Y-Chromosoms, welches in drei Generationen einer phänotypisch unauffälligen Familie gefunden wurde. Im Jahr 2004 beschrieben Amor et al. drei Personen aus zwei Generationen einer Familie, welche bei karyotypisch unauffälligem Chromosomensatz ein inaktives natürliches Zentromer auf dem Chromosom 4 bei Ausbildung eines Neozentromers im Bereich 4q21.3 besaßen. Dies stellt einen bislang einzigartigen Bericht der Ausbildung eines mitotisch stabilen Neozentromers im Bereich euchromatischer Sequenzen ohne vorhergehende chromosomale Rearrangements und bei weitestgehend unauffälligem Phänotyp der Betroffenen dar.

Das erste und bisher am besten untersuchte menschliche Neozentromer wurde bei der routinemäßigen Karyotypisierung eines Jungen mit Lernschwierigkeiten entdeckt (Voullaire et al. 1993). Es entstand durch die Deletion des normalen Zentromers des Chromosoms 10, was zur Ausbildung eines Ringchromosoms rdel(10) mit dem ursprünglichen Zentromer sowie dem Chromosom mardel(10) mit einem Neozentromer am Locus 10q25.3 führte (du Sart et al. 1997, Abb. 3b). Die Sequenzierung dieses DNA-Abschnitts des Patienten-Chromosoms wies eine völlige Übereinstimmung mit der DNA-Sequenz des väterlichen Ursprungsallels und somit das Fehlen jeglicher α -Satelliten oder CENP-B Boxen nach (Barry et al. 2000a, Barry et al. 2000b). Saffery et al. (2000) konnten mittels Immunfluoreszenz-Markierung von 20 zentromerspezifischen Proteinen alle, mit Ausnahme von CENP-B, am Kinetochor dieses Neozentromers nachweisen (s. Anhang 1).

Auch die Sequenzierung weiterer Neozentromere, deren exakte genomische Position durch die Lokalisation von CENP-A bestimmt wurde, ergab keinen Unterschied zur entsprechenden DNA-Sequenz in normalen Chromosomen (Sullivan et al. 2001, Lo et al. 2001a, Lo et al. 2001b, Amor und Choo 2002). Und die Lokalisation der CENP-A enthaltenden Chromatin-Domänen dreier unabhängig voneinander im Bereich 13q32 aufgetretener Neozentromere ergab, dass diese unterschiedliche Sequenzen im Abstand von 1 bzw. 5 Mb zugrunde liegen

haben (Alonso et al. 2003, Warburton et al. 2004). Somit scheint die Existenz sogenannter „Neozentromer-Hotspots“ nicht durch eine spezifische DNA-Sequenz bedingt zu sein. Was die Ausbildung eines aktiven Zentromers im Bereich heterogener euchromatischer DNA jedoch verursacht, und wie dieser Prozess vonstatten geht ist bislang unklar.

5. Die Verwendung künstlicher Chromosomen zur Erforschung der Zentromere

Eine Möglichkeit, die Voraussetzungen zur Ausbildung aktiver Zentromere und den Zusammenhang zwischen der primären DNA-Sequenz und der Zentromerfunktion zu studieren, besteht in der Konstruktion von künstlichen Chromosomen.

Das erste *de novo* erzeugte menschliche künstliche Chromosom (Harrington et al. 1997) enthielt einen α -Satelliten-Array von über 100 kb Länge, welcher durch wiederholte unidirektionale Ligation eines sequenzierten 2.7 kb Higher order repeats des Chromosoms 17 synthetisiert wurde. Vorausgesetzt es kam dabei nicht zu einer unerkannten Rekombination endogener Zentromer-DNA in die künstlichen Chromosomen, kann durch dieses Experiment die Existenz einer „magischen Sequenz“ als Grundlage für die Funktion von Zentromeren ausgeschlossen werden (s. auch Basu et al. 2005, Ebersole et al. 2005).

Ikeno et al. (1998) gelang es, künstliche Chromosomen mit α -Satelliten-DNA des α 21-I Locus des Chromosoms 21 zu erzeugen. Dieser Array ist aus homogenen Higher order repeats aufgebaut, welche aus 11 Monomeren bestehen und CENP-B Boxen enthalten. Mit der divergierten α -Satelliten-DNA des Locus α 21-II, dessen Repeats keine CENP-B Boxen besitzen, gelang die Erzeugung künstlicher Chromosomen hingegen nicht. In Übereinstimmung mit diesen Ergebnissen zeigten Grimes et al. (2002b), Mejía et al. (2002) sowie Basu et al. (2005), dass mit CENP-B Box enthaltenden α -Satelliten-Repeats des Chromosoms 17 eine effiziente Bildung von HACs möglich ist, während mit denselben Repeats bei mutierten CENP-B Boxen beziehungsweise mit Repeats des Y-Chromosoms, welche keine CENP-B Boxen enthalten, kaum stabile künstliche Chromosomen erzeugt werden konnten. Ohzeki et al. (2002) erzeugten Konstrukte aus vier verschiedenen Typen an repetitiver DNA: α -Satelliten des Chromosoms 21 mit intakten CENP-B Boxen, dieselbe α -Satelliten-DNA nach Inaktivierung der CENP-B Boxen durch Punktmutationen, sowie ein G/C-reiches Fragment des Plasmids pBR322 gleicher Länge jeweils mit und ohne CENP-B Boxen. Dabei erhielten sie nur unter Verwendung eines Arrays aus α -Satelliten-DNA mit intakten CENP-B Boxen künstliche Chromosomen. Ob auch die Homogenität der eingesetzten Repeats von Bedeutung ist, lässt sich bislang nicht mit Sicherheit sagen.

Im Gegensatz zur Notwendigkeit von CENP-B Boxen zur *de novo* Ausbildung von Zentromeren konnten Yang et al. (2000) mittels Telomer assoziierter Chromosomen Fragmentation (TACF) eines menschlichen Y-Chromosoms Minichromosomen erzeugen, welche bei voller mitotischer Stabilität nur 100 kb an α -Satelliten-DNA ohne CENP-B Boxen enthielten. Ebenso gelang es Minichromosomen durch Fragmentation eines Markerchromosoms zu erzeugen, welches ausschließlich ein Neozentromer und keinerlei α -Satelliten-DNA enthielt. Eine *de novo* Ausbildung von HACs unter Verwendung ebendieser Neozentromer DNA gelang hingegen nicht (Saffery et al. 2001, Wong et al. 2002). Zusammengefasst deuten diese Ergebnisse darauf hin, dass sowohl CENP-B Boxen als auch eine weitere Eigenschaft der α -Satelliten-DNA, möglicherweise eine Sequenz-Struktureigenschaft höherer Ordnung, eine Rolle bei der *de novo* Ausbildung von Zentromeren spielen. Für letzteres spricht, dass die Effizienz der *de novo* Formierung von HACs von der Länge der Input DNA abhängt (Grimes et al. 2002, Okamoto et al. 2007). Diese Anforderungen treffen jedoch nicht für die Aufrechterhaltung bereits bestehender Zentromere zu, was auf deren teilweise epigenetische Determination zurückzuführen ist (Cleveland et al. 2003, Masumoto et al. 2004): Die „top down“ Konstruktion von Minichromosomen findet in einer Zellkultur und somit *in vivo* statt, so dass die epigenetische Markierung der Zentromerregionen niemals verloren geht. Dadurch kann die Ausbildung von Kinetochoren auch an nicht- α -Satelliten-DNA beziehungsweise α -Satelliten-DNA ohne CENP-B Boxen stattfinden. Bei der *de novo* Konstruktion von künstlichen Chromosomen wird hingegen „nackte“ DNA transfiziert. Damit deren epigenetische Kennzeichnung als Zentromer-Region effizient abläuft, sind auch sequenzabhängige Eigenschaften wie CENP-B Boxen erforderlich. Die physiologische und evolutionäre Bedeutung einer *de novo* Ausbildung von Zentromeren ist bislang jedoch unklar.

6. Die Evolution der α -Satelliten-DNA

6.1. Die Phylogenese der humanen Zentromer-Arrays

Arrays aus repetitiver DNA finden sich an den Zentromeren aller höherer Eukaryoten. Die α -Satelliten-DNA des Menschen mit einer Repeat-Länge von circa 170 bp ist hingegen primatenspezifisch. Man geht davon aus, dass sich die Arrays der höheren Primaten aus einer ursprünglichen monomerischen Repeatsorte entwickelt haben. Ein Relikt dieser ursprünglichen Repeatsorte ist heute noch im Genom der Grünen Meerkatze *Cercopithecus aethiops* zu finden. Wie auch in dieser Spezies sind die α -Satelliten der niederen Primaten

nicht chromosomenspezifisch, das heißt die Zentromere aller Chromosomen haben die gleiche α -Satelliten-Sequenz (Goldberg et al. 1996, Haaf und Willard 1998). Alexandrov et al. (1988, 2001) haben ein Modell vorgeschlagen, wie die Multimere der suprachromosomalen Familien eins bis drei sowie die heutigen Chromosomen-spezifischen Higher order repeats aus diesen einfachen Arrays homogener Monomere entstanden sein könnten. Demnach bestanden die auf allen Chromosomen identischen Arrays des letzten gemeinsamen Vorfahren der höheren Primaten aus homogenen Monomeren, welche keine CENP-B Boxen besaßen und als Typ A Monomere bezeichnet werden. Ein derartiges Array der suprachromosomalen Familie vier bildet heute noch das Zentromer des menschlichen Y-Chromosoms (Choo et al. 1991, Lee et al. 1997).

Durch Mutationen entstanden hiervon divergierende α -Satelliten-Repeats, wobei sich, möglicherweise bedingt durch Selektionsdruck (Romanova et al. 1996), die CENP-B Box enthaltenden Typ B Monomere ausbildeten. Arrays aus irregulär angeordneten Typ A und Typ B Monomeren gehören der suprachromosomalen Familie fünf an. Durch die Amplifikation von Multimeren entstanden dann aus diesen „alten“ Arrays der suprachromosomalen Familien vier und fünf die neuen Array-Typen der Familien eins bis drei (vgl. Tab. 1 u. Kap. 3.3). Vergleiche der α -Satelliten unterschiedlicher Spezies legen nahe, dass sich der Ur-Typ B Repeat vor circa 15-25 Millionen Jahren entwickelt haben muß (Haaf et al. 1995), während die neueren suprachromosomalen Familien vor der Auftrennung der Menschen, Schimpansen und Gorillas vor circa 7 Millionen Jahren entstanden sein müssen (Haaf und Willard 1998, Alexandrov et al. 2001).

Die bisher scheinbar letzte Stufe der Ausbildung von Multimeren stellt die Entwicklung der Higher order repeats durch Amplifikation mehrerer Mono-, Di- oder Pentamere dar. So besteht zum Beispiel das Zentromer des Chromosoms 11 aus Higher order repeats, welche direkt den Pentameren der suprachromosomalen Familie drei entsprechen (W1-W2-W3-W4-W5). Durch Amplifikation der drei hinteren Monomere eines Pentamers, des kompletten folgenden Pentamers sowie der ersten vier Monomere des nächsten Pentamers entstanden daraus die 12-mere des X-Chromosoms mit der Struktur W3-W4-W5-W1-W2-W3-W4-W5-W1-W2-W3-W4. Durch Duplikationen, Amplifikationen und den Transfer auf nicht-homologe Chromosomen entwickelten sich aus diesen 12-meren die Higher order repeats der Chromosomen 1 und 17. Einhergehend mit dieser Multimerisation kam es zu einer weiteren Diversifizierung der Repeats, so dass heute die meisten Zentromere mit Ausnahme der

Chromosomen 5 und 19, 13 und 21 sowie 14 und 22 spezifische Higher order repeats besitzen (Willard 1985, Willard und Waye 1987, Choo et al. 1991).

6.2. Concerted Evolution

Ursprünglich beschreibt der Begriff „Concerted Evolution“ die Beobachtung, dass die Homologien repetitiver DNA-Sequenzen innerhalb einer Spezies höher sind als zwischen verschiedenen Spezies (Dover 1982). Weiter verallgemeinert führt die gemeinsame Evolution dazu, dass repetitive Sequenzen innerhalb eines Chromosoms, einer Population oder einer Spezies eine hohe Übereinstimmung aufweisen, während zwischen heterologen Chromosomen beziehungsweise verschiedenen Populationen oder Spezies häufig beträchtliche Unterschiede bestehen (Abb. 4, Hennig und Walker 1970, Hood et al. 1970, Walker 1971, Dover et al. 1993, Elder und Turner 1995). Eine derartige evolutionäre Entwicklung ist nicht allein durch natürliche Selektion und genetische Drift erklärbar, sondern erfordert molekulare Mechanismen, wie zum Beispiel Transpositionen, Duplikationen, Deletionen, Genkonversionen, Unequal crossing over, Rolling circle replication oder Replication slippage, welche eine nicht-mendelische Verbreitung von Mutationen ermöglichen. Die Gesamtheit der durch diese Mechanismen verursachten Veränderungen wird auch als „Molecular Drive“ bezeichnet, welcher eine Evolution von Organismen in nicht-mendelischer und nicht-darwinistischer Weise ermöglicht (Dover 1982, Dover 2000, Dover 2002).

Eine gemeinsame evolutionäre Entwicklung wurde sowohl bei tandemartig angeordneten Repeats wie Minisatelliten, α -Satelliten (Waye und Willard 1989) oder den Genen der ribosomalen- und der small nuclear RNAs beobachtet (Arnheim et al. 1980, Pavelitz et al. 1995), als auch bei verstreut im Genom liegenden repetitiven Sequenzen wie den Genen der transfer RNAs (Amstutz et al. 1985), den Globingenen (Maeda und Smithies 1986) oder den LINES und SINEs (Kass et al.

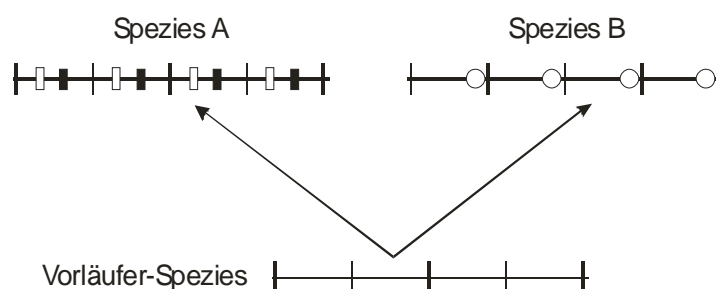


Abb. 4: Definition des Begriffs „Concerted Evolution“: Der Begriff „Concerted Evolution“ beschreibt ursprünglich den Vorgang der zu einer höheren Homogenität von repetitiven DNA-Sequenzen innerhalb einer Spezies als zwischen verschiedenen Spezies führt. Concerted Evolution kann jedoch auch innerhalb von isolierten Populationen oder, wie zum Beispiel bei der humanen α -Satelliten DNA, auf einzelnen Chromosomen einer Spezies auftreten.

1995, Lenoir et al. 1997). Sie findet sich außerdem sowohl bei hochrepetitiver, nicht kodierender DNA wie den Minisatelliten oder α -Satelliten als auch bei mittelrepetitiven, kodierenden DNA-Sequenzen. Diese können zum Beispiel in funktionelle RNAs wie ribosomale- und transfer RNAs transkribiert werden oder für eine Vielzahl von Proteinen wie Globine, Histone, Immunglobuline, MHC-Proteine oder Aktine kodieren (Dover 1982, Kawamura et al. 1992, Charlesworth et al. 1994).

Die hohe Effizienz der gemeinsamen Evolution zeigt eine Untersuchung der Satelliten-DNA des Edelsteinkärpflings *Cyprinodon variegatus* (s. Internet-Links) durch Elder und Turner (1994): Die Analyse von zwölf Populationen dieser Fische von der Ostküste der USA und dem Golf von Mexiko ergab innerhalb der einzelnen Individuen und innerhalb der Populationen keine Unterschiede in den Repeat-Sequenzen. Zwischen den Populationen divergierten die Repeats jedoch in zehn der zwölf Fälle mit einer mittleren Übereinstimmung von lediglich 61,6% (35,0% - 95,9%). Dies ist umso erstaunlicher, als einige der Populationen nur wenige Kilometer voneinander entfernt leben, die Fische lediglich minimale morphologische Unterschiede aufweisen und Fische von unterschiedlichen Populationen sich paaren und fruchtbare Nachkommen hervorbringen können. Allerdings ist der Zeitpunkt der Separation der einzelnen Populationen nicht bekannt.

Aber auch die im Vergleich zum sonstigen Genom hohe Divergenz der α -Satelliten zwischen den nahe verwandten Spezies Mensch, Gorilla und Schimpanse, sowie deren Divergenz zu den niederen Primaten, bei gleichzeitig hoher Homogenität der Repeats innerhalb dieser Spezies zeigt eindrücklich die Effizienz mit der Concerted Evolution stattfindet (Waye und Willard 1989, Archidiacono et al. 1995).

6.3. Homogenisierung und Fixierung von Varianten

Damit repetitive Sequenzen sich einerseits im Laufe der Evolution verändern können und andererseits ihre Homogenität gewahrt bleibt, müssen in einem einzelnen Repeat auftretende Neumutationen entweder rasch entfernt werden oder sich auf weitere Repeats ausbreiten. Dabei müssen prinzipiell zwei Arten der Ausbreitung einer Variante unterschieden werden: Erstens die Homogenisierung, das heißt die Ausbreitung auf weitere Repeats innerhalb eines Genoms und zweitens die Fixierung, das heißt die Ausbreitung auf die homologen Chromosomen aller Individuen einer Spezies. Die Fixierung einer Variante kann nur durch interchromosomale Rekombinationen in der Keimbahn erfolgen. Ob diese nur in der Meiose oder auch bei mitotischen Zellteilungen auftreten ist nicht endgültig geklärt (Bollag et al.

1989). Die Homogenisierung einer Variante kann sowohl durch interchromosomale als auch durch intrachromosomale Rekombinationen erfolgen. Intrachromosomale Rekombinationen wiederum können entweder zwischen den zwei Chromatiden eines Chromosoms (Schwester-Chromatid-Austausch) oder innerhalb einer Chromatide

Concerted Evolution: Die voneinander abhängige Evolution repetitiver DNA-Sequenzen, so dass die Repeats innerhalb einer Spezies (Population, Array) eine höhere Homogenität aufweisen, als Repeats von verschiedenen Spezies (Populationen, Arrays).

Molecular Drive: Evolutionärer Prozess, zusätzlich zu natürlicher Selektion und neutraler Drift, welcher über molekulare Vorgänge wie Transpositionen, Genkonversionen, Unequal crossing over, Replication slippage oder Rolling circle amplification eine nicht-mendelische und nicht-darwinistische Verbreitung von Mutationen ermöglicht.

Homogenisierung: Die Ausbreitung einer Variante über alle in einem Organismus (Chromosom, Array) vorkommende DNA-Repeats.

Fixierung: Die Ausbreitung einer Variante über alle Individuen einer Spezies (Population).

stattfinden. Somit ist eine Homogenisierung von repetitiver DNA prinzipiell in jeder Phase des Zellzyklus möglich.

Wie häufig intra- und interchromosomale Rekombinationen im menschlichen Genom auftreten, und welchen Beitrag sie jeweils zur Homogenisierung von repetitiver DNA leisten ist bislang unklar. Untersuchungen von α -Satelliten-DNA, rRNA und snRNA verschiedener Spezies lassen jedoch vermuten, dass in diesen Tandem-Arrays intrachromosomale Rekombinationen erheblich häufiger sind als interchromosomale (Warburton und Willard 1990, Dover et al. 1993, Schlötterer und Tautz 1994, Warburton und Willard 1995, Liao et al. 1997, Roizès 2006, Rudd et al. 2006).

6.4. Die Mechanismen von Molecular Drive

Es wurde das Auftreten einer Vielzahl von molekularen Mechanismen im Bereich von repetitiver DNA beschrieben deren Bedeutung für die Evolution dieser Regionen seit vielen Jahren kontrovers diskutiert wird. Dazu zählen Unequal crossing over (Smith 1974, Tartoff 1975, Roizès 2006), Genkonversion, replikative Transposition (Thompson-Stewart et al. 1994), Rolling Circle Replication (Hourcade et al. 1973, Spradling 1981) oder Replication slippage (Levinson und Gutman 1987, Stephan 1989). Für alle diese Mechanismen konnten Computersimulationen zeigen, dass sie zur Homogenisierung einer repetitiven DNA beitragen können. Eine exakte experimentelle Analyse der Veränderungen einer repetitiven DNA im zeitlichen Verlauf, um daraus Rückschlüsse auf die zugrunde liegenden Mechanismen ziehen zu können, ist bislang jedoch nicht möglich. Durch die vor allem bei hohen

Sequenzhomologien auftretenden Schwierigkeiten einer Kartierung und Sequenzierung ist es bisher nicht gelungen, die Ausbreitung von Varianten in einem Array aus DNA-Repeats zu verfolgen.

Zwei Mechanismen denen die größte Bedeutung bei der Evolution repetitiver DNA beigemessen wird sind Unequal crossing over und Genkonversion.

6.4.1. Unequal crossing over

Beim Unequal crossing over kommt es zu einer versetzten Aneinanderlagerung zweier (homologer) DNA-Stränge mit anschließendem Austausch an einem bestimmten Kreuzungspunkt (Abb. 5a). Ein Unequal crossing over zwischen zwei identischen DNA-Strängen führt dabei zur Duplikation eines Abschnittes auf dem einen Strang, während es auf dem anderen Strang zum Verlust dieses Abschnittes kommt. Im Fall einer tandem-repetitiven DNA entstehen so ein verkürzter und ein verlängerter Array mit einer, durch die Duplikation bedingt, höheren Homogenität der Repeats. Es kommt jedoch nicht zur Vermehrung einer Sequenz-Variante oder eines Repeats sondern lediglich zu einer Verschiebung von einer Chromatide auf eine andere. Ein solches Ereignis wird daher auch als „reziproke Rekombination“ bezeichnet (Bollag et al. 1989). In einem Modell von Concerted Evolution durch Unequal crossing over muß die Ausbreitung einer Variante in einem Array aus repetitiver DNA durch sich wiederholende Unequal crossing over Ereignisse mit Verlust des

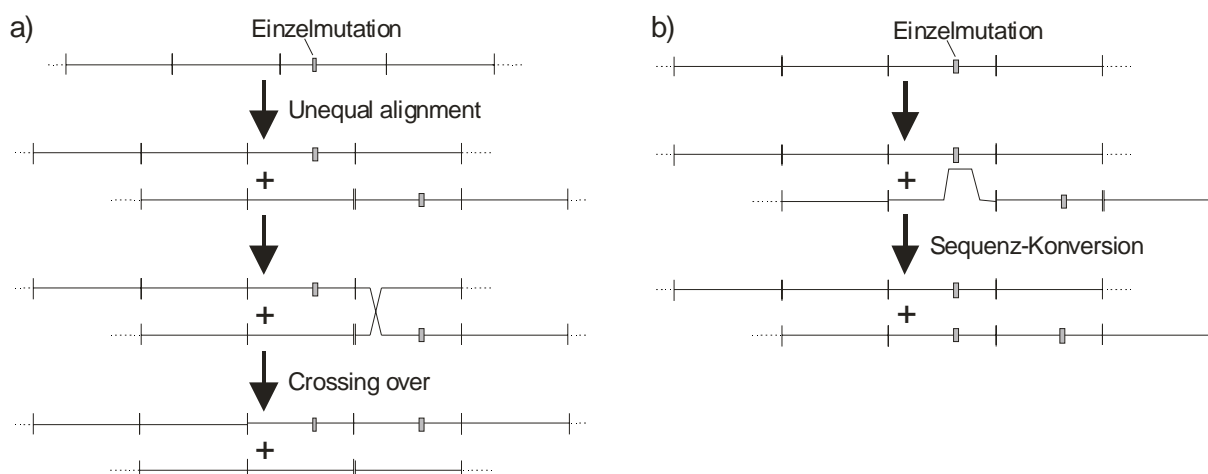


Abb. 5: Unequal crossing over und Genkonversion

a) Unequal crossing over: Die ungleiche Aneinanderlagerung zweier homologer DNA-Stränge mit anschließendem Austausch an einem bestimmten Kreuzungspunkt führt zur Duplikation eines Sequenz-Abschnittes auf dem einen und zum Verlust dieses Abschnittes auf dem anderen DNA-Strang.

b) Genkonversion (Sequenzkonversion): Ein Sequenzabschnitt wird von einem DNA-Strang auf einen anderen übertragen, wobei die Sequenzinformationen auf dem Akzeptor-Strang verloren gehen.

jeweils verkürzten Arrays (Brutlag 1980) oder durch die zusätzliche Amplifikation von Repeats über andere Mechanismen wie zum Beispiel Replication slippage erfolgen (Stephan 1989). Unequal crossing over innerhalb einer Chromatide führt zum Verlust der DNA zwischen den Kreuzungspunkten und somit zur Verkürzung eines repetitiven DNA-Arrays.

Computersimulationen zeigten, dass wiederholtes Unequal crossing over sowohl zur Ausbildung von repetitiven Strukturen, ausgehend von einer beliebigen nicht-repetitiven DNA-Sequenz führen kann, als auch den homogenen Zustand existierender Tandem-Arrays aufrecht erhalten kann (Smith 1976, Ohta 1976, Walsh 1987, Stephan 1989, Stephan und Cho 1994, Alkan et al. 2004). Die Simulationen wiesen außerdem eine Reihe weiterer, für repetitive DNA-Arrays charakteristische Eigenschaften auf (Smith 1976): So entstanden wiederholt übergeordnete Repeat-Strukturen, ähnlich den Higher order repeats der menschlichen α -Satelliten-DNA. Außerdem hatten die resultierenden Arrays, analog den menschlichen α -Satelliten-Arrays (Mashkova et al. 1998, Schueler et al. 2001), zu den Rändern hin abnehmende Sequenz-Homologien. Die logische Folge einer Homogenisierung durch Unequal crossing over, dass in einem Array nahe beieinander liegende Repeats größere Übereinstimmungen aufweisen als entfernt liegende (Brutlag 1980), konnte ebenfalls experimentell gezeigt werden (Warburton und Willard 1990).

Die Computersimulationen zeigten jedoch auch, dass sehr viele Rekombinationsereignisse erforderlich sind um die gemeinsame Evolution von Tandem-Arrays von der Größe der menschlichen α -Satelliten-Arrays erklären zu können. So ergaben Berechnungen durch Ohta (1976), dass die Ausbildung eines Arrays, vergleichbar dem menschlichen α -Satelliten-Array des Chromosoms 17 mit 1000 Higher order repeats von 2,7 kb Länge, circa 10^6 Rekombinationsereignisse erfordern würde. Demgegenüber sind seit der Auftrennung der höheren Affen lediglich 250.000 Generationen vergangen, was die Entstehung des menschlichen D17Z1 Arrays durch ausschließlich meiotisches Unequal crossing over unwahrscheinlich macht. Außerdem ergaben Simulationen durch Charlesworth et al. (1986) dass ein Array aus repetitiver DNA umso stabiler sein sollte, je seltener Unequal crossing over, Amplifikationen oder Selektion auftreten, wohingegen häufige Rekombinationen tendenziell eher zu einer Verkürzung und zum Verlust von Arrays führen.

Ferner ist die gemeinsame Evolution von verstreut im Genom liegenden DNA-Repeats nicht durch Unequal crossing over erklärbar, da Rekombinationen zwischen Repeats, welche durch single-copy Sequenzen voneinander getrennt sind oder gar auf verschiedenen Chromosomen liegen, zu chromosomaler Instabilität führen würden. Ebenfalls nicht erklärbar ist die

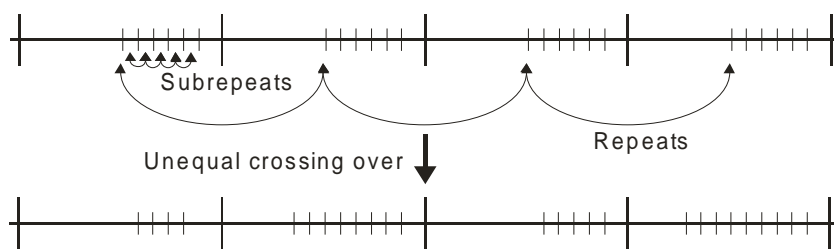


Abb. 6: DNA-Repeats, welche aus einmaligen Sequenzen und homogenen Subrepeats bestehen, widersprechen einer gemeinsamen Evolution durch Unequal crossing over, da dies zu einer ständigen Verkürzung und Verlängerung der Subrepeats führen würde.

gemeinsame Evolution von Repeats, welche sowohl einmalige Sequenzen als auch homogene Subrepeats enthalten (Abb. 6): Unequal crossing over führt zu kontinuierlichen Verlängerungen und Verkürzungen der Repeat-Arrays, wohingegen beispielsweise die Repeats der 5S rRNA eine konstante Anzahl an homogenen Subrepeats enthalten (Nielsen et al. 1990).

6.4.2. Genkonversion

Bei Genkonversionen (Sequenzkonversionen) handelt es sich im Gegensatz zu Unequal crossing over um den unidirektionalen Transfer von genetischer Information. Ein wenige Basen bis einige Kilobasen langer Sequenzabschnitt wird von einem DNA-Strang auf einen anderen „kopiert“ und dieser dabei „überschrieben“ (Abb. 5b, Dover 1993). Dieser Mechanismus ist somit nichtreziprok, da ein Teil der genetischen Information auf dem Empfängerstrang verloren geht, während der Geberstrang unverändert bleibt. Marker außerhalb des konvertierten Bereichs bleiben im Gegensatz zu Unequal crossing over ebenfalls unverändert.

Genkonversionen können innerhalb eines Arrays aus repetitiver DNA, zwischen homologen Chromosomen oder zwischen ähnlichen Abschnitten nicht homologer Chromosomen auftreten. Dabei kann die Ausbreitung einer Mutation gerichtet sein, das heißt die Konversion einer Variante A in die Variante B tritt häufiger auf als umgekehrt (Nagylaki und Petes 1982, Hillis et al. 1991). Ein derartiger Konversions-Bias kann zum Beispiel durch Selektionsdruck bedingt sein, aber auch eine Bevorzugung bestimmter Basenfolgen durch den zugrunde liegenden molekularen Mechanismus ist denkbar.

Wie für Unequal crossing over wurden theoretische Modelle entwickelt, welche zeigten, dass durch Genkonversionen eine effektive Homogenisierung von repetitiver DNA möglich ist (Nagylaki und Petes 1982, Nagylaki 1984a, Nagylaki 1984b). Dies gilt im Gegensatz zu Unequal crossing over auch für die gemeinsame Evolution nicht-tandemförmiger Repeats (Ohta und Dover 1983).

6.5. Der experimentelle Nachweis von Molecular Drive

Das Auftreten von Genkonversionen konnte in verschiedenen repetitiven Sequenzen wie Minisatelliten (Jeffreys et al. 1994), SINEs (Kass et al. 1995, Lenoir et al. 1997), den für rRNAs und tRNAs kodierenden Genen (Amstutz et al. 1985, Hillis et al. 1991, Linares et al. 1994, Gangloff et al. 1996, Liao 2000) oder in Protein-kodierenden Genen (Kawamura et al. 1992, Bettencourt und Feder 2002) nachgewiesen werden. Auch bei der gemeinsamen Evolution der menschlichen α -Satelliten-DNA wurde Genkonversionen eine wesentliche Rolle zugeschrieben (Alexandrov et al. 1993). Es ist jedoch bislang nicht gelungen ihr Auftreten in humaner Zentromer-DNA direkt nachzuweisen.

Im Gegensatz hierzu konnten im divergierten Randbereich des α -Satelliten-Arrays des Chromosoms 21 Monomere gefunden werden, welche die reziproken Produkte nach Unequal crossing over repräsentieren (Mashkova et al. 1998). Ferner kann das Vorkommen von Higher order repeats mit unterschiedlichen Monomerzahlen auf dem Chromosom 17 gut durch Unequal crossing over, ausgehend von einem ursprünglichen 16mer, erklärt werden (Warburton et al. 1993). Die Beobachtung von extrachromosomaler zirkulärer α -Satelliten-DNA (Okumura et al. 1987, Walsh 1987) wurde ebenfalls als das Ergebnis von intrachromosomalem Unequal crossing over gedeutet. Sie wurde jedoch auch als Hinweis auf eine extrachromosomale Rolling circle Amplifikation mit anschließender Integration der entstandenen homogenen Tandem-Arrays interpretiert (Charlesworth et al. 1994, Hourcade et al. 1973). Allen experimentellen Nachweisen von Unequal crossing over gemeinsam ist jedoch ihre Beobachtung in Bereichen mit niedrigen Sequenz-Homologien, in denen die Häufigkeit von Rekombinationen und die Homogenisierung der DNA stark reduziert ist (Mashkova et al. 1998, Schueler et al. 2001).

6.6. α -Satelliten-DNA ist sehr gut geeignet um die Mechanismen von Concerted Evolution zu untersuchen

Die am besten geeigneten Sequenzen um die molekularen Mechanismen der gemeinsamen Evolution zu studieren stellen die α -Satelliten-Repeats der Zentromere dar, da man davon ausgeht, dass ihre evolutionäre Entwicklung ohne direkten Selektionsdruck vonstatten geht (Brutlag 1980, Durfy und Willard 1989). Das Fehlen von konservierten Sequenzabschnitten, mit Ausnahme der CENP-B Box, welche für die Zentromerfunktion jedoch nicht essentiell ist (vgl. Kap. 4.2), legt die Vermutung nahe, dass einzelne Mutationen keinen gravierenden Einfluss auf die Funktion eines α -Satelliten-Arrays für die Aufrechterhaltung eines

funktionellen Zentromers besitzen. Somit lassen sich durch Analyse der DNA-Sequenzen der einzelnen Repeats Rückschlüsse auf die der gemeinsamen Evolution zugrunde liegenden molekularen Mechanismen ziehen, ohne die Selektion berücksichtigen zu müssen.

7. Zielsetzung

Grundlage dieser Arbeit waren fünf PACs (P1 derived artificial chromosomes) welche zwischen 120 und 160 kb an α -Satelliten-DNA des X-Chromosoms eines einzelnen männlichen Individuums enthielten. Durch Klonierung und Endsequenzierung von jeweils zehn zufällig ausgewählten Higher order repeats aus den Inserts dieser PACs sollte die Homogenität der α -Satelliten-DNA innerhalb und zwischen den PACs ermittelt werden. In weiteren Versuchen sollten diese PACs dann für die Konstruktion von künstlichen Chromosomen verwendet werden. Die Häufigkeit der *de novo* Ausbildung von Chromosomen sowie deren mitotische Stabilität hätten somit Rückschlüsse von der Primärsequenz und der Homogenität der α -Satelliten-Repeats auf die Zentromerfunktion erlaubt.

Bei der Endsequenzierung der Higher order repeats fiel jedoch eine Verteilung der variablen Basen innerhalb der Repeats auf, welche durch das häufig angeführte Modell der gemeinsamen Evolution von α -Satelliten-DNA durch Unequal crossing over nicht erklärt werden kann. Daher sollte eine Methode etabliert werden, welche durch Partialverdau der PAC-DNA, Auftrennung mittels Pulsfeld-Gelelektrophorese (PFGE), Southern Blotting und anschließende Hybridisierung von Vektorsequenzen eine Kartierung von Restriktionsschnittstellen und somit von einigen der variablen Basen ermöglicht. Diese Kartierungen sowie weitere Untersuchungen sollten Rückschlüsse auf die Beteiligung von Unequal crossing over und Genkonversionen bei der Evolution der α -Satelliten-DNA ermöglichen.

B. Ergebnisse

Ein Ziel dieser Arbeit war die Charakterisierung von fünf PACs, welche α -Satelliten-DNA des menschlichen X-Chromosoms enthalten, um daraus Erkenntnisse über den Aufbau der homogenen α -Satelliten-Arrays der menschlichen Zentromere zu gewinnen. Im weiteren Verlauf sollten diese PACs zur Konstruktion von künstlichen Chromosomen verwendet werden, um daraus Rückschlüsse von der Sequenz der α -Satelliten-DNA auf die Funktionalität der Zentromere in Bezug auf die Rate an erzeugten künstlichen Chromosomen und deren mitotische Stabilität ziehen zu können. Daher standen zunächst Aspekte wie die Homogenität der Higher order repeats innerhalb der PACs und die Häufigkeit von Varianten und „Neumutationen“ (im Folgenden „Einzelmutationen“ genannt) im Vordergrund. Ferner sollte untersucht werden, ob die Higher order repeats innerhalb der klonierten Array-Segmente der PACs ähnlicher zueinander sind als zwischen den PACs, und ob ihre Varianten gruppierte Verteilungen („Haplotypen“) aufweisen. Dafür wurden von jedem PAC 10 zufällig ausgewählte Higher order repeats subkloniert und endsequenziert. Hierbei fiel eine Verteilung der variablen Basen auf, welche mit dem gängigen Modell der Evolution von α -Satelliten-DNA nicht vereinbar war. In der Folge wurden umfangreiche Analysen und Kartierungen der Varianten durchgeführt, um deren unerwartete Verteilung und damit die molekularen Prozesse, welche der gemeinsamen Evolution der als genetisch „kalt“ geltenden Zentromere (Mahtani und Willard 1998) zugrunde liegen, besser beschreiben zu können. Diese Erkenntnisse könnten entscheidend zu einem besseren Verständnis der Evolution und der normalen Turnover-Vorgänge an chromosomalen Sequenzen beitragen.

1. Die X-chromosomalen α -Satelliten-PACs

Die fünf untersuchten X-chromosomalen PACs A6, A7, A8, A10 und B11 stammen aus Segment I der humanen PAC-Bank RPCI-I, welche durch Mbo-I Partialverdau des Genoms eines einzelnen männlichen Individuums und Klonierung der DNA-Fragmente in den Vektor pCYPAC2 erzeugt wurde (Ioannou et al. 1994, s. Internet-Links). Die Klone wurden durch Hybridisierung der PAC-Bank mit der 2 kb langen X-chromosomalen α -Satelliten-Probe pBamX5 (von Peter Warburton freundlicherweise zur Verfügung gestellt) identifiziert, welche unter strengen Hybridisierungsbedingungen vornehmlich an die Zentromer-Arrays der Chromosomen 11, 17 und X bindet (Willard et al. 1983). Restriktionsanalysen von 48 der insgesamt 347 hybridisierenden Klone mit *Bam*-HI, *Eco*-RI und *Xba*-I ergaben, dass 12 der 48 Klone vom α -Satelliten-Array des X-Chromosoms stammen, 14 von dem des

Chromosoms 17 und 11 von dem des Chromosoms 11. Die verbleibenden 11 Klone konnten anhand der Restriktionsanalysen nicht eindeutig zugeordnet werden.

Da hochrepetitive DNA in Plasmiden und YACs (Yeast artificial chromosomes) häufig instabil ist und zu Deletionen neigt (Neil et al. 1990), wurden zunächst Längenbestimmungen von je zwei Einzelkolonien von je 6 PACs der Chromosomen 17 und X durchgeführt, welche keine erkennbaren Unterschiede ergaben. Die durchschnittliche Insert-Länge aller 48 isolierten α -Satelliten-Klone von 109 kb stimmt mit der durchschnittlichen Insert-Länge dieses Segments der PAC-Bank von 110 kb überein. Bei einem α -Satelliten-Anteil von 1-3% am gesamten menschlichen Genom kann bei der festgestellten Anzahl und Länge der hybridisierenden Klone somit davon ausgegangen werden, dass die Zentromer-Arrays der Chromosomen 11, 17 und X zum größten Teil in der PAC-Bank repräsentiert sind und mit der Probe *pBamX5* hybridisieren (Schindelhauer und Schwarz 2002).

Die durchschnittliche Insert-Länge der fünf für die weiteren Analysen verwendeten X-chromosomalen PACs beträgt 136 kb (Tab. 2). Der vollständige Verdau dieser PACs mit dem Restriktionsenzym *Bam*-HI und anschließende Längenbestimmung der Fragmente auf Ethidium-Bromid gefärbten Agarosegelen, beziehungsweise Southern-Blotting und Hybridisierung mit der α -Satelliten-Probe *pBamX5*, zeigten, dass die PAC-Inserts fast ausschließlich aus 2 kb α -Satelliten-DNA Fragmenten bestehen. Die Längen der von 2 kb abweichenden Fragmente sowie deren Anzahl, welche durch quantitative Schätzung der DNA-Mengen in der Ethidium-Bromid Färbung und durch Kartierung der PACs (s. Kap. 5.1) bestimmt wurde, sind in Tab. 2 aufgelistet. Durch PAC-Endsequenzierung konnten die Klonierstellen der α -Satelliten-Inserts in den PAC-Vektor bestimmt werden und dadurch einige der von 2 kb abweichenden *Bam*-HI Fragmente als Klonierungsränder identifiziert werden.

Restriktionsverdau der PAC-DNA mit den selten schneidenden Enzymen *Not*-I, *Asc*-I, *Mlu*-I,

Tabelle 2: Insert-Längen sowie Längen und Anzahl der von 2 kb abweichenden *Bam*-HI Fragmente der für die Analysen verwendeten X-chromosomalen α -Satelliten-PACs.

PAC	Insert-Länge	Längen der von 2 kb abweichenden <i>Bam</i> -HI Fragmente und deren geschätzte Anzahl. Mit Sternchen versehene Fragmente stellen Klonierungsränder dar.
A6	160 kb	0.9 kb (1*), 1.7 kb (2), 2.2 kb (1), 2.9 kb (5), 4.9 kb (1)
A7	145 kb	0.9 kb (1*)
A8	120 kb	0.9 kb (1*), 1.7 kb (1)
A10	130 kb	0.9 kb (1*)
B11	125 kb	0.55 kb (1*), 2.9 kb (2), 4.9 kb (1)

BssH-II, *Pvu*-I, *Pac*-I und *Sal*-I ergab lediglich eine *Sal*-I Restriktionsschnittstelle im PAC B11. Klonierung und Endsequenzierung eines circa 60 kb langen *Sal*-I Fragments, welches den PAC-Vektor und einen Teil des α -Satelliten-Inserts enthielt, zeigte, dass die *Sal*-I Restriktionsschnittstelle durch Mutation einer einzigen Base eines ansonsten typischen Higher order repeats entstand.

Alle bislang erwähnten Analysen waren Vorarbeiten die von Dr. D. Schindelbauer durchgeführt wurden.

2. Klonierung und Endsequenzierung von 2 kb und 3 kb *Bam*-HI Higher order repeats

Zunächst sollte von den fünf α -Satelliten-PACs des X-Chromosoms eine Reihe von 2 kb und 3 kb Higher order repeats kloniert werden um deren Endsequenzen zu bestimmen. Dazu wurde die in Agarose-Blöckchen präparierte und aufgereinigte PAC-DNA mit dem Restriktionsenzym *Bam*-HI verdaut und die Fragmente mittels Pulsfeld-Gelelektrophorese (PFGE) der Länge nach aufgetrennt (Abb. 7). Die 2 kb und 3 kb Fragmente wurden aus dem

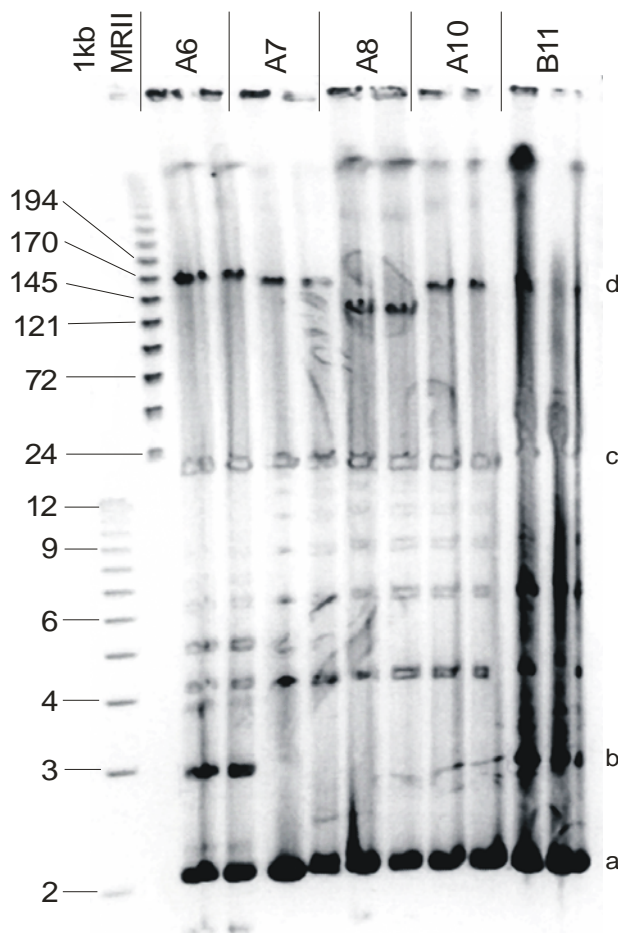


Abb. 7: PFGE der α -Satelliten PACs A6, A7, A8, A10 und B11 nach Verdau mit *Bam*-HI. Jeweils zwei Blöckchen mit kompletter PAC-DNA wurden aufgetragen. Aufgrund eines nur partiellen Verdau der DNA sind sowohl die linearisierten PACs mit Längen von 135-175kb als auch eine Reihe weiterer DNA-Banden mit Abständen von 2kb zu sehen. (a) 2 kb *Bam*-HI α -Satelliten Fragmente. (b) 3 kb *Bam*-HI α -Satelliten Fragmente in den PACs A6 und B11. (c) 16 kb PAC-Vektor. (d) Linearisierte PACs vollständiger Länge. Längenstandards: 1kb: Gibco BRL 1 kb DNA ladder, MRII: NEBioLabs MidRange II PFG Marker. PFGE-Bedingungen: Feldstärke 6 V/cm, Switchzeit 1-20 sek, Laufzeit 14h.

Gel ausgeschnitten und in den weiß/blau selektionierbaren Vektor pBluescript II SK kloniert. Durch Verdau von Minipräparationen der erhaltenen Klone mit *Bam*-HI respektive *Eco*-RI wurden die Insert-Längen sowie die Gesamtlängen der Plasmide bestimmt.

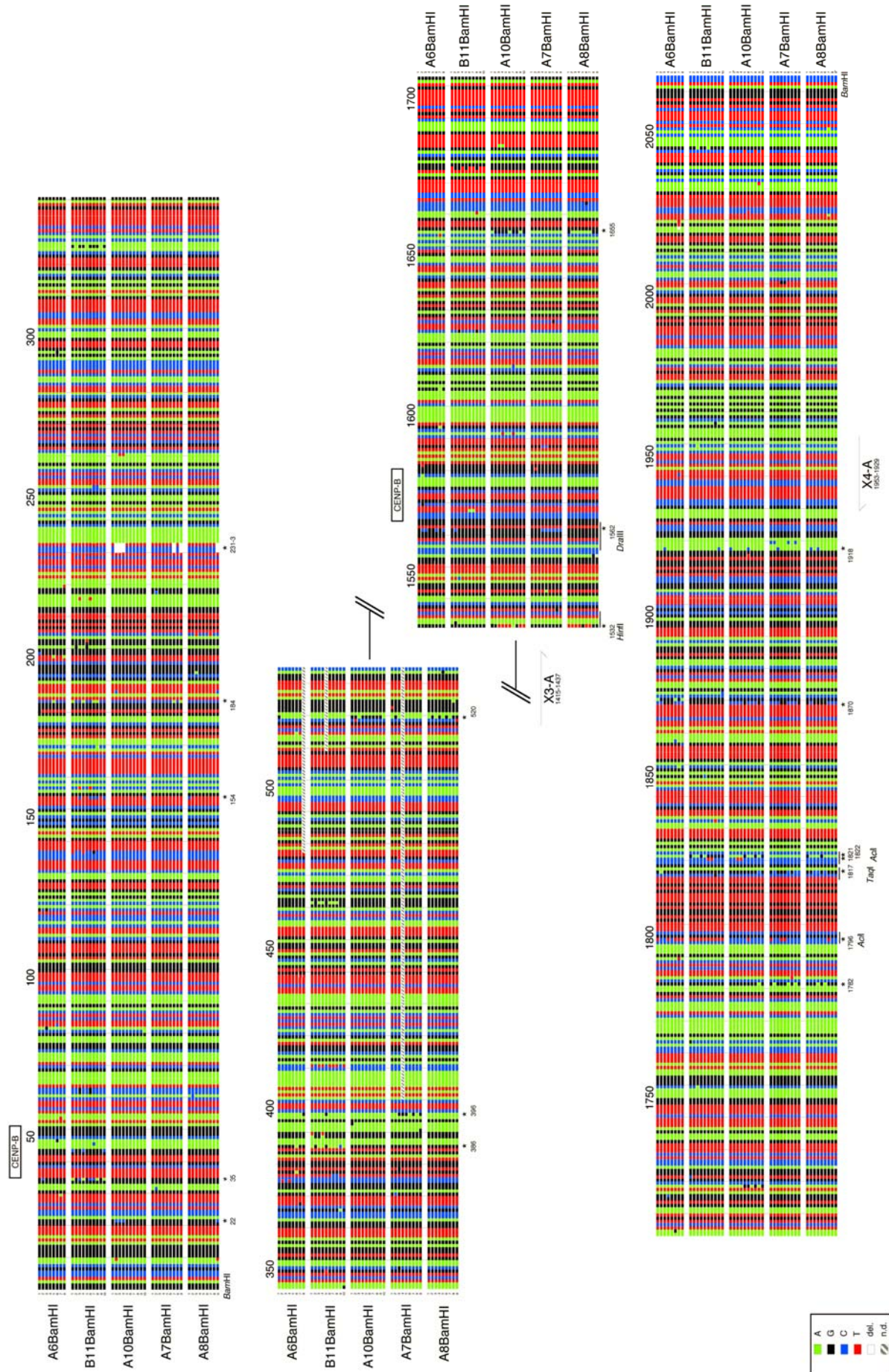
Für die weiteren Analysen wurden jeweils zehn Klone der fünf X-chromosomalen PACs mit 2 kb *Bam*-HI Insert, sowie acht Klone des PAC B11 und drei Klone des PAC A6 mit 3 kb *Bam*-HI Insert ausgewählt. Von diesen insgesamt 61 klonierten Fragmenten wurden mit den Primern T7 und T3 (s. Anhang 2) die Endsequenzen bestimmt, wie unter Methoden beschrieben korrektur gelesen und in einem „4-Farben Alignment“ verglichen (Abb. 8). Schlecht auswertbare Sequenzen wurden entweder erneut mit den Primern T7 beziehungsweise T3 oder mit den Primern X3-A respektive X4-A sequenziert, welche spezifisch im Bereich der Monomere 9 und 12 der 12-meren Higher order repeats der X-chromosomalen α -Satelliten-DNA binden (Abb. 9a, Warburton et al. 1991, Warburton und Willard 1992). Insgesamt wurden von jedem Higher order repeat jeweils circa 500 bp ausgehend von den *Bam*-HI Klonierstellen bestimmt (Abb. 9a). Die EMBL/GenBank Zugriffsnummern aller Sequenzen sind in Anhang 3 aufgelistet.

Entsprechend der 1985 von Wayne und Willard veröffentlichten Sequenz des Higher order repeats HSSATAX wird im Folgenden das 5'-Ende der Erkennungssequenz des Restriktionsenzym *Bam*-HI (5'-GGATCC-3') als erste Base des Repeats sowie die Orientierung vom Primer X-3A zum Primer X-4A festgelegt. Die Endsequenzen werden demnach willkürlich als „vordere“ und „hintere“ Sequenz bezeichnet. Bei der Analyse wurden die vordere und die hintere Sequenz eines *Bam*-HI Fragments stets gemeinsam beurteilt.

2.1. Sequenz-Homologien innerhalb und zwischen den PACs

Das Alignment der Sequenzen ergab vier identische Sequenz-Paare: Zwei in PAC A6 sowie je eines in den PACs A7 und A8. Diese identischen Sequenzen können entweder durch die vollständige Übereinstimmung zweier verschiedener Higher order repeats im sequenzierten

Abb. 8 (folgende Seite): Sequenz-Alignment der klonierten 2 kb *Bam*-HI Fragmente. Dargestellt sind die Endsequenzen von je 8-10 Higher order repeats der X-chromosomalen α -Satelliten PACs A6, A7, A8, A10 und B11. Jede Base ist durch ein dem Nukleotid entsprechend gefärbtes Viereck dargestellt. Jede Reihe entspricht zwei 171 bp Monomeren. Die Positionen der CENP-B Boxen sowie der Primer X-3A und X-4A sind angegeben. Mit einem Stern gekennzeichnet sind diejenigen variablen Basen, welche mindestens drei mal sowie in mindestens zwei PACs vorkommen. Ebenfalls angegeben sind die Spaltstellen von Restriktionsenzymen, die abhängig vom Auftreten einer variablen Base schneiden.



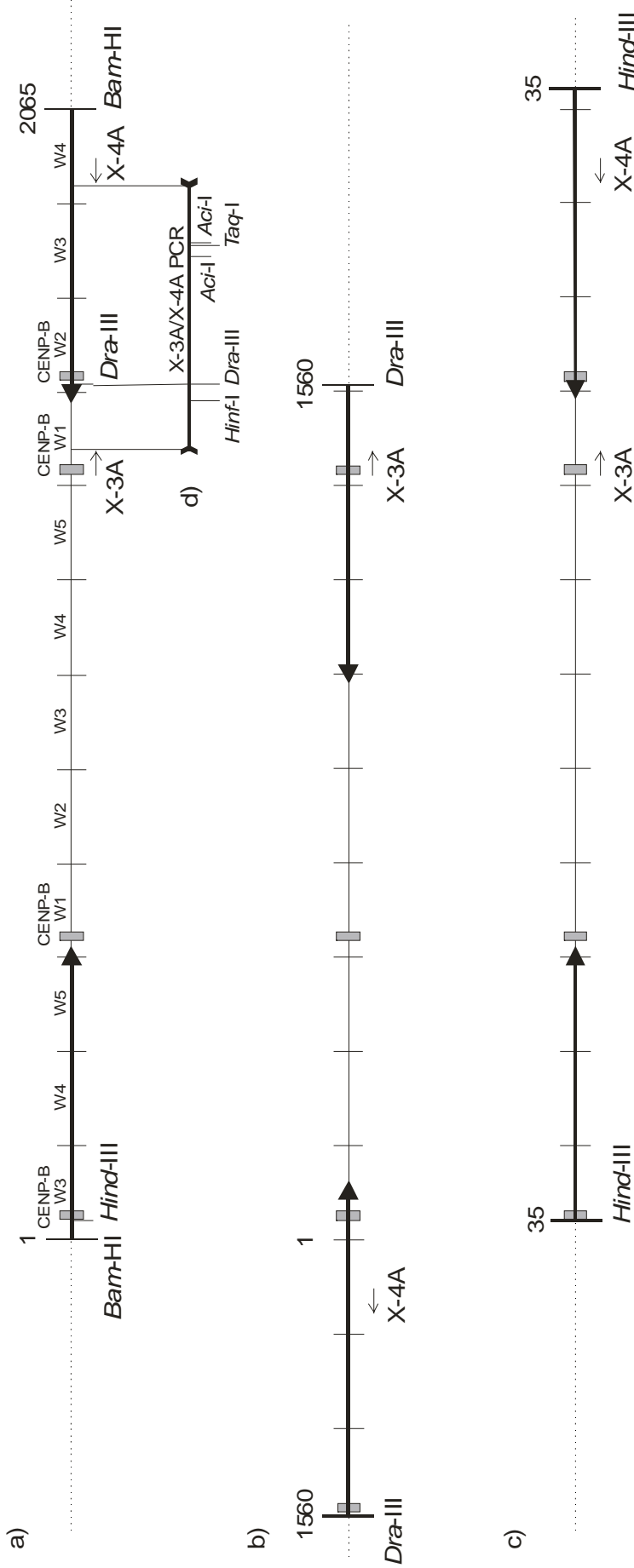


Abb. 9: Karten der klonierten X-chromosomalen α -Satelliten Fragmente. Die fettgedruckten Pfeile stellen die jeweils von den meisten Fragmenten sequenzierten Bereiche dar, die kleinen Pfeile geben die Lokalisation und Orientierung der Primer X-3A und X-4A an. In den Monomeren 1, 4, 9 und 10 sind die hier meist vorhandenen CENP-B Boxen eingezeichnet.
 a) Karte der klonierten Bam-HI Fragmente.
 b) Karte der klonierten Dra-III Fragmente.
 c) Karte der klonierten Hind-III Fragmente.
 d) Karte des mit den Primern X-3A / X-4A spezifisch amplifizierbaren, X-chromosomalen α -Satelliten Fragments. Die Lokalisationen der Restriktionsschnittstellen, welche durch variable Basen beeinflusst werden, sind angegeben.

Bereich oder durch doppeltes Klonieren eines einzelnen Repeats bedingt sein. Geht man von einer PAC-Insertlänge von 140 kb und somit $n=70$ 2 kb Higher order repeats aus, errechnet sich die Wahrscheinlichkeit, dass bei der zufälligen Klonierung von $k=10$ Repeats mindestens ein Repeat doppelt kloniert wird zu:

$$p = 1 - \frac{n \times (n-1) \times (n-2) \times \dots \times (n-k+1)}{n^k} = 0,49$$

Da die vier identischen Sequenzpaare somit vermutlich durch doppelte Klonierung eines Higher order repeats bedingt sind, wurde jeweils nur ein Klon für die weiteren Analysen verwendet. Die Sequenz eines Klons des PAC A8 stimmte vollständig mit der Sequenz eines Klons des PAC A10 überein, möglicherweise bedingt durch ein Überlappen der PACs A8 und A10. Da diese Sequenzen jedoch von unterschiedlichen PACs stammen, wurden sie beide in die folgenden Analysen einbezogen.

Zwischen allen möglichen Sequenzpaaren wurden die Homologien bestimmt (alle Werte finden sich unter http://www.tobias-schwarz.net/gr/comparison_all.html). Die mittleren Homologien der Sequenzen innerhalb der PACs liegen dabei zwischen 98,6% (PAC B11) und 99,3% (PAC A8), die minimalen Homologien zwischen 97,9% (PACs A6 und B11) und

Tabelle 3: Mittlere Homogenität und minimale Homogenität der analysierten Higher order repeats und Häufigkeit von Einzelmutationen in den sequenzierten Bereichen der fünf PACs.

	Gesamter sequenzierter Bereich				In allen Klonen sequenzierter 365 bp-Abschnitt		
	Klonzahl	Mittlere Homogenität (%)	Minimale Homogenität (%)	Einzelmutationen (%)	Mittlere Homogenität (%)	Minimale Homogenität (%)	Einzelmutationen (%) (Klonzahl)
A6	8	98,83	97,9	0,33	98,54	97,3	0,45 (n=8)
A7, zufällig gewählte Klone	9	99,12	98,5	0,14	98,94	98,1	0,09 (n=9)
A8	9	99,25	98,7	0,08	99,14	98,6	0,03 (n=9)
A10	10	99,11	98,7	0,07	98,78	97,8	0,08 (n=10)
B11	10	98,58	97,9	0,20	98,92	97,8	0,25 (n=10)
A7 <i>Dra</i>III-Fragmente	15	99,04	98,2		98,79	97,3	0,14 (n=18)
A7 nicht-<i>Dra</i>III-Fragmente	18	99,22	98,3		99,21	98,4	0,03 (n=18)
PCR-Fragmente	13	97,98	96,8		97,96	96,2	0,38 (n=13)
365 bp Abschnitt von 9 zufällig gewählten Repeats aus allen 5 PACs	9				98,81	97,3	0,12 (n=9)

98,7% (PACs A8 und A10, s. Tab. 3 sowie Abb. 10). Die geringste Übereinstimmung zwischen zwei Higher order repeats überhaupt findet sich zwischen zwei Klonen der PACs B11 und A6 beziehungsweise B11 und A10 und liegt bei jeweils 97,2%. Die sequenzierten Higher order repeats weisen somit insgesamt sehr hohe Homologien auf.

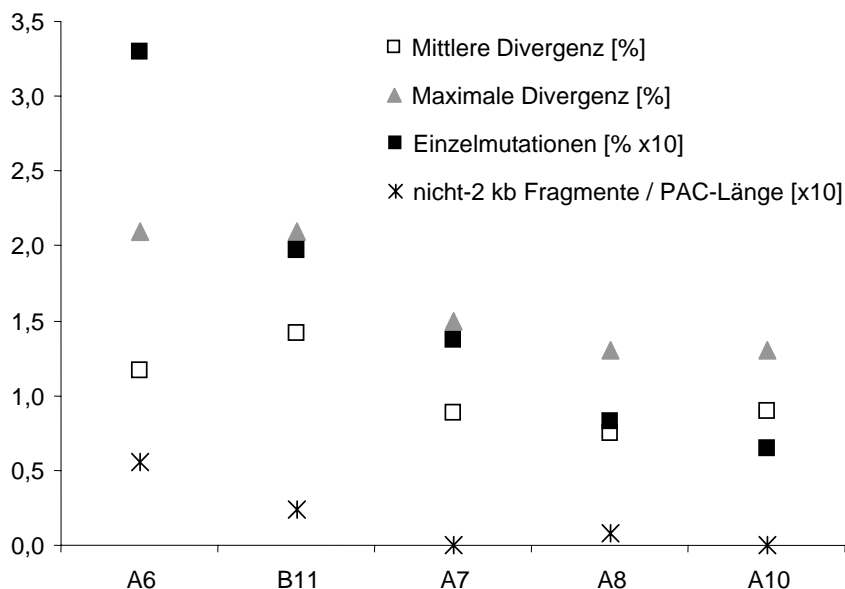


Abb. 10: Homologievergleich der 5 α -Satelliten PACs: Die PACs A6 und B11 sind geringfügig divergenter als die PACs A7, A8 und A10. Bestimmt wurden die mittlere und die maximale Divergenz der klonierten Higher order repeats, die Häufigkeit von Einzelmutationen sowie die Anzahl an nicht-2 kb Fragmenten.

2.2. Die Divergenz zwischen den Higher order repeats wird im Wesentlichen durch variable Positionen verursacht

Wie in Abb. 8 zu erkennen ist, wird der größte Teil der Divergenz zwischen den Higher order repeats durch in einem oder mehreren PACs vorkommende variable Positionen verursacht. Definiert man jedes Auftreten eines bestimmten, von der Konsensussequenz abweichenden Nukleotids in mindestens zwei Higher order repeats als variable Base, so finden sich in den initial sequenzierten 46 Higher order repeats insgesamt 52 solcher Positionen. Dabei handelt es sich um 25 Transitionen, 24 Transversionen, 1 Deletion von drei Basenpaaren sowie 2 Positionen, an welchen drei verschiedene Nukleotide vorkommen. Unter der Annahme, dass die PACs repräsentativ für den gesamten α -Satelliten-Array des menschlichen X-Chromosoms sind, wird im Folgenden davon ausgegangen, dass Varianten welche in mehreren PACs vorkommen eine große Verbreitung innerhalb des Arrays besitzen. Dementsprechend werden Varianten, welche in mindestens drei Higher order repeats sowie in mindestens zwei PACs auftreten als globale Varianten bezeichnet. Insgesamt finden sich in den sequenzierten Abschnitten somit 18 globale Varianten, 30 Varianten welche in nur einem PAC vorkommen, sowie 77 Einzelmutationen.

Die Häufigkeit dieser Einzelmutationen in den einzelnen PACs liegt zwischen 0,07% (PAC A10) und 0,33% (PAC A6). Verglichen mit den mittleren Divergenzen von 0,7% (PAC A8) bis 1,4% (PAC B11) zeigt dies, dass der weitaus größte Teil der Unterschiede zwischen den Higher order repeats durch variable Basen verursacht wird.

Ein Vergleich der mittleren

Homologien, der maximalen Abweichungen, des Auftretens von Einzelmutationen und des Vorkommens von *Bam*-HI Fragmenten, welche von der typischen 2 kb Higher order repeat Länge abweichen, ergibt ein konsistentes Bild der Homogenität der fünf untersuchten PACs (Tab. 3 und Abb. 10): Während die PACs A6 und B11 etwas divergenter sind, weisen die PACs A7, A8 und A10 sehr hohe Homogenitäten auf, was eine kürzer zurückliegende oder stärkere Beteiligung dieser Abschnitte des Zentromers am Prozess der Homogenisierung der α -Satelliten-DNA vermuten lässt.

Die statistische Analyse der Lokalisationen der Sequenzabweichungen innerhalb der Higher order repeats mittels Kolmogoroff-Smirnow-Test für Einstichproben-Probleme ergibt sowohl für die Varianten als auch für die Einzelmutationen mit einer asymptotischen Signifikanz von eins eine Poisson-Verteilung. Eine signifikante Anhäufung von Varianten in bestimmten Bereichen der Higher order repeats, insbesondere im Bereich der CENP-B Boxen, welchen eine Rolle bei Rekombinationen im α -Satelliten-Array zugeschrieben wurde (Warburton et al. 1993, Kipling und Warburton 1997), findet sich somit nicht (Abb. 11).

2.3. Ausschluss von Klonierungs- und Sequenzierungsartefakten

Um auszuschließen, dass die ermittelten Einzelmutationen das Resultat von Fehlern in der Klonierung oder Sequenzierung der Higher order repeats darstellen, wurden acht 3 kb *Bam*-HI Fragmente des PACs B11 in pBluescript II SK kloniert und endsequenziert wie oben

In dieser Arbeit verwendete Klassifikation von Sequenzabweichungen:

Einzelmutation: In den 46 in dieser Arbeit initial ansequenzierten 2 kb *Bam*-HI Fragmenten nur ein mal auftretende Sequenzabweichung. Da hier nur wenige der circa 500-2000 Higher order repeats eines mittleren X-chromosomalen Zentromer-Arrays analysiert wurden, stellt vermutlich nur ein sehr kleiner Teil der hier als Einzelmutationen deklarierten Veränderungen tatsächliche Neumutationen dar.

Variable Base: Auftreten eines bestimmten, von der Konsensussequenz abweichenden Nukleotids in mindestens 2 der ansequenzierten Higher order repeats eines oder mehrerer PACs.

Globale Variante: In den 46 in dieser Arbeit initial sequenzierten 2 kb *Bam*-HI Fragmenten in mindestens drei Repeats sowie in mindestens zwei PACs auftretende variable Base. Es stellte sich im Laufe der vorliegenden Arbeit heraus, dass diese Definition mit einer hohen Wahrscheinlichkeit Varianten betrifft, die in der Population fixiert sind.

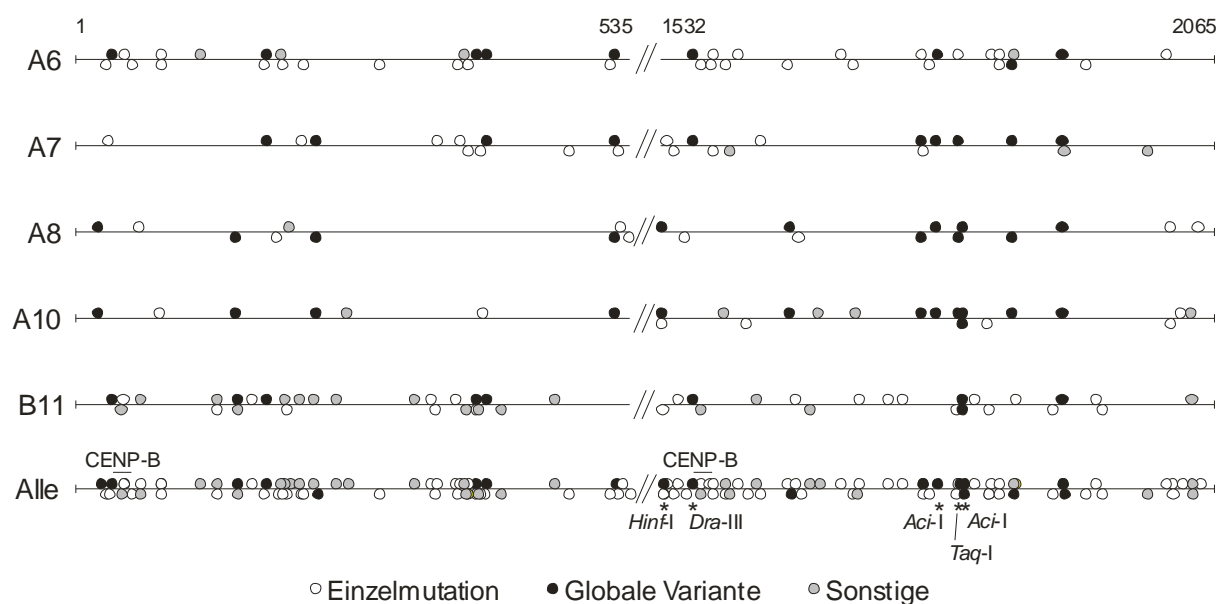


Abb. 11: Sowohl die Einzelmutationen als auch die Varianten sind gleichmäßig über die 12 Monomere des X-chromosomalen Higher order repeats verteilt. Als „Einzelmutation“ wurden Abweichungen von der Konsensussequenz bezeichnet, welche in den 46 initial sequenzierten Higher order repeats nur einmal vorkommen. Als „Global“ wurden Varianten definiert, welche mindestens dreimal sowie in mindestens zwei PACs auftreten. Alle anderen variablen Basen sind als „Sonstige“ angegeben. In den PACs A7, A8 und A10 sind die meisten Unterschiede durch Varianten und nicht durch „neu aufgetretene“ Einzelmutationen verursacht.

beschrieben. Die Sequenzanalyse ergab zwei unterschiedliche 3 kb Repeats, was mit der quantitativen Schätzung der Repeat-Häufigkeit nach *Bam*-HI Verdau der PAC-DNA sowie den durch Partialverdau ermittelten Restriktionskarten (s. Kap. 5.1) übereinstimmt. Darüber hinausgehende Abweichungen in den unabhängig voneinander bestimmten Sequenzen bestanden nicht. Von den 2 Repeats, die als PACs in vielen Bakterien vorlagen, wurden also 8 Subklone isoliert, vermutlich aus 8 individuellen Bakterien, welche exakt 2 unterschiedliche Higher order repeat-Sequenzen ergaben. Dies schließt nicht nur Sequenzierfehler als Ursache für das Auftreten von Einzelmutationen aus, sondern spricht auch gegen häufige Mutationen im Rahmen der PAC-Replikation in den Wirtsbakterien und während der Subklonierungen.

Als weitere Kontrolle wurden die klonierten *Bam*-HI Higher order repeats des PAC A7 mit dem Restriktionsenzym *Dra*-III verdaut, welches entsprechend einer variablen Base an Position 1532 (s. Kap. 3.1) drei der neun Inserts spaltete. Außerdem ergab der Doppel-Verdau der PAC-DNA mit *Bam*-HI und *Dra*-III ein Verhältnis von *Dra*-III gespaltenen zu ungespaltenen 2 kb *Bam*-HI Fragmenten welches sehr gut mit der jeweiligen Häufigkeit der variablen Base an Position 1532 der sequenzierten Higher order repeats übereinstimmt.

Diese Tests bestätigen, dass weder die variablen Basen noch die Einzelmutationen durch Klonier- oder Sequenzierartefakte bedingt sind.

3. Nachweis der Fixierung der globalen Varianten in der Bevölkerung

3.1. PCR und Restriktionsverdau humaner X-chromosomaler α -Satelliten-DNA

Um die relative Häufigkeit der globalen Varianten in den gesamten α -Satelliten-Arrays und ihre Fixierung in der Bevölkerung zu untersuchen, wurden sechs variable Basen ausgewählt, welche im Bereich der Erkennungssequenz eines der Restriktionsenzyme *Hinf*-I, *Dra*-III, *Taq*-I oder *Aci*-I liegen. Je nach Vorliegen der einen oder anderen Variante an der variablen Position schneidet das Enzym den Higher order repeat oder nicht (Abb. 8). Tab. 4 zeigt die Erkennungssequenzen der verwendeten Restriktionsenzyme sowie die variablen Basen, welche die DNA-Spaltung beeinflussen.

Tabelle 4: Erkennungssequenzen der Restriktionsenzyme welche die Higher order repeats im Bereich variabler Basen spalten. Die in der α -Satelliten-DNA variablen Nukleotide sind jeweils unterstrichen.

Restriktionsenzym	Erkennungssequenz	Position der variablen Base im Higher order repeat
<i>Hinf</i> -I	5'- <u>G</u> AN TC-3'	1532
<i>Taq</i> -I	5'-TC <u>G</u> A-3'	1817
<i>Dra</i> -III	5'-CAC NNN <u>G</u> TG-3'	1562
<i>Aci</i> -I	5'-CC <u>G</u> C-3'	1796
<i>Aci</i> -I	5'-CC <u>G</u> C-3'	1821 u. 1822
<i>Hind</i> -III	5'-AAG <u>C</u> TT-3'	35

Alle diese Varianten liegen in einem 489 bp langen Abschnitt der 2 kb Higher order repeats, welcher durch eine PCR mit den Primern X-3A und X-4A spezifisch amplifiziert werden kann (Abb. 9d). Diese Primer binden selektiv an die α -Satelliten-DNA des X-Chromosoms und die PCR-Produkte sind repräsentativ für das gesamte DXZ1-Array (Warburton et al. 1991, Warburton und Willard 1992).

Um einseitige PCR Artefakte zu erkennen, wurden PCRs aus kompletter genomischer DNA von zehn nicht verwandten, männlichen Individuen mit steigenden Template-Mengen von 50 ng, 150 ng, 300 ng und 1000 ng bei abnehmender Anzahl an PCR-Zyklen (35, 25, 20, 15) durchgeführt. Der Verdau dieser PCR-Produkte mit den Enzymen *Hinf*-I, *Dra*-III, *Taq*-I und *Aci*-I zeigte ein annähernd gleiches Verhältnis von gespaltener zu ungespaltener DNA in allen untersuchten Individuen für alle gewählten PCR-Bedingungen (Abb. 12). Dies zeigt, dass die sechs identifizierten variablen Positionen in allen Individuen in annähernd gleichem Verhältnis vorkommen und somit in der Bevölkerung fixiert sind. Die identischen Ergebnisse für unterschiedliche Template-Mengen und Zykluszahlen belegen ferner die Repräsentation des gesamten Zentromer-Arrays durch die PCR-Amplifikation mit den Primern X-3A und X-4A.

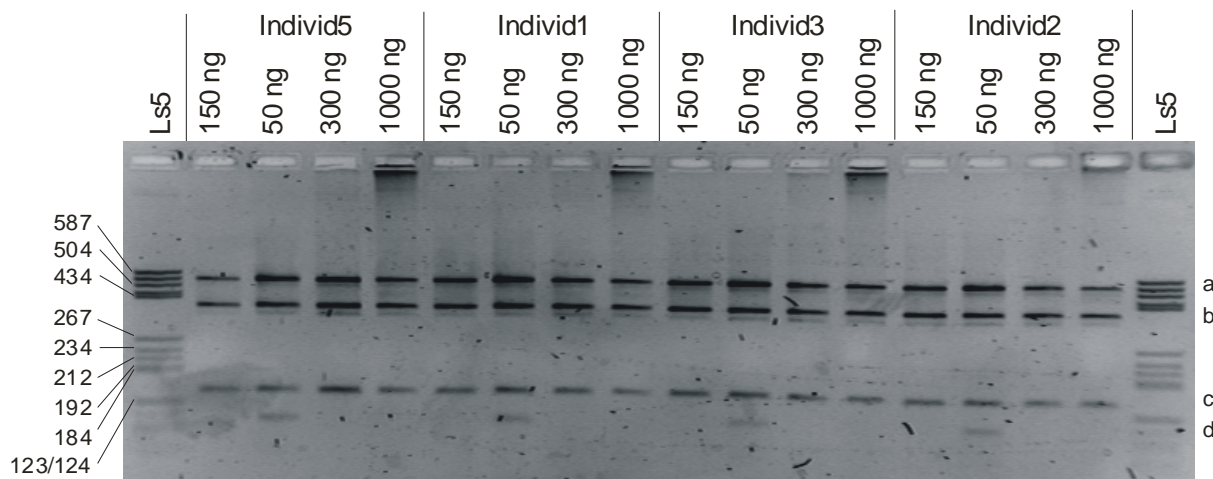


Abb. 12: Die globalen Varianten sind in der Bevölkerung fixiert: Von insgesamt 10 nicht-verwandten, männlichen Individuen wurde ein 489 bp langer Abschnitt der X-chromosomalen α -Satelliten Repeats mittels spezifischer PCR amplifiziert und mit dem variabel spaltenden Enzym *Dra*-III verdaut. Das in allen Individuen jeweils ähnliche Verhältnis von gespaltener zu ungespaltenener DNA zeigt die Fixierung der untersuchten Variante. Abgebildet ist der *Dra*-III Verdau des PCR Produkts von 4 Individuen. Um die Amplifikation von nur einigen wenigen Higher order repeats auszuschließen wurden PCRs mit steigenden Template-Mengen von 50 ng, 150 ng, 300 ng und 1000 ng bei abnehmender Anzahl an PCR-Zyklen (35, 25, 20, 15) durchgeführt. Agarosegelelektrophorese (1,2%), EthBr gefärbt. (a) Ungespaltenes PCR-Produkt. (b, c) Spaltprodukte bei Vorhandensein einer *Dra*-III Restriktionsschnittstelle. (d) Unspezifisches PCR-Produkt bei hoher Zyklenzahl.

Um die Ergebnisse zu verifizieren wurde diese Untersuchung mehrfach durchgeführt. Außerdem wurde eine entsprechende PCR-Amplifikation mit PAC-DNA als Template durchgeführt. Der anschließende Restriktionsverdau ergab eine weitgehende Übereinstimmung zwischen dem Verhältnis von ungespaltenem zu gespaltenem PCR-Produkt und dem Verhältnis der Varianten in den sequenzierten 2 kb *Bam*-HI Fragmenten (Abb. 13).

3.2. Klonierung und Sequenzierung von PCR-Fragmenten

Zur weiteren Bestätigung dieser Ergebnisse wurden die PCR-Produkte von vier Individuen nach 20 Amplifikations-Zyklen von 500 ng Template DNA über das Quiagen PCR Purification Kit aufgereinigt und in den pGEM-T Vektor kloniert. Dreizehn Klone mit dem 500 bp langen Insert wurden mit den Primern SP6 und T7 doppelstrangsequenziert. Dabei wurden die beiden Endsequenzen zunächst unabhängig voneinander bestimmt und korrektur gelesen. Der anschließende Vergleich zeigte keinen einzigen Unterschied in den 6,4 kb der doppelt sequenzierten Bereiche, was die Zulässigkeit der für die 2 kb-Klone verwendeten Einzelstrang-Sequenzierungen unterstreicht.

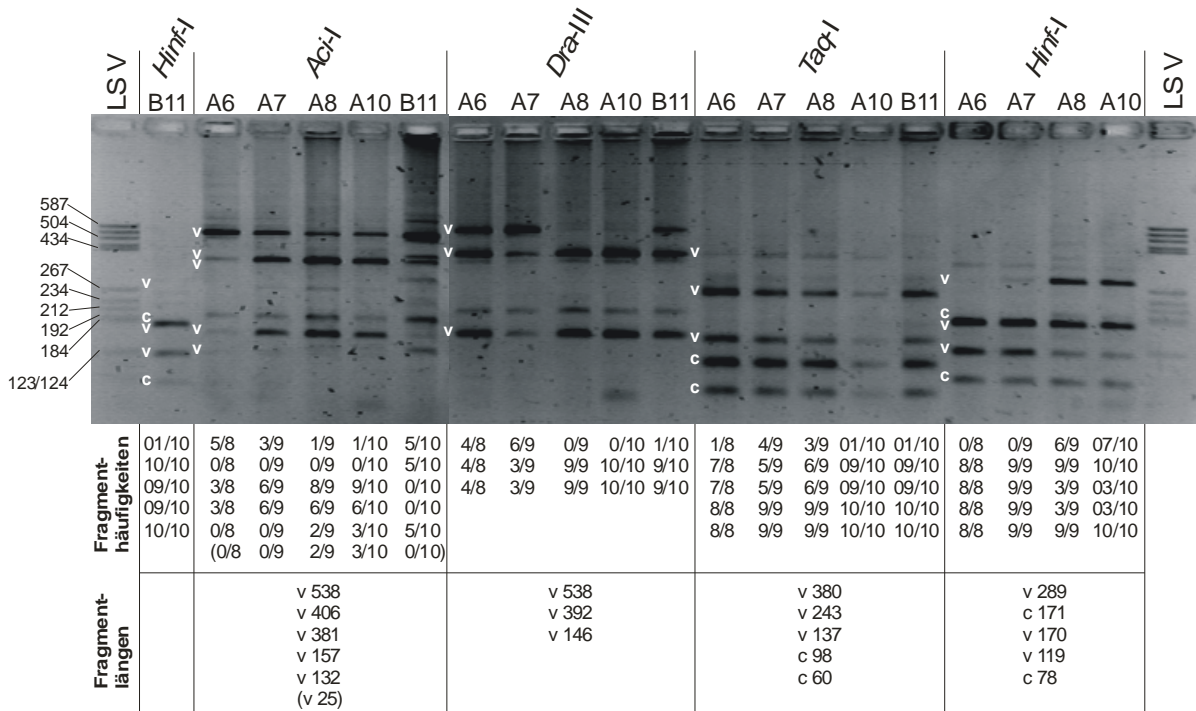


Abb. 13: Die spezifische α -Satelliten PCR mit anschließendem Restriktionsverdau erlaubt die Bestimmung der Häufigkeit von Varianten: Von den fünf X-chromosomalen α -Satelliten PACs wurden PCR-Amplifikationen mit den Primern X-3A und X-4A durchgeführt. Der anschließende Verdau der PCR-Produkte mit den Restriktionsenzymen *Hinf-I*, *Dra-III*, *Taq-I* und *Aci-I* ergab Verhältnisse von gespaltener zu ungespaltener DNA, welche mit den durch Klonierung und Sequenzierung ermittelten Häufigkeiten der variablen Basen weitgehend übereinstimmen. Die aufgrund der DNA-Sequenzen erwarteten variablen und konstanten Banden sind mit v und c bezeichnet. Agarosegelelektrophorese (1,2%), EthBr gefärbt. Die Tabelle unterhalb des Gelphotos gibt die in den Sequenzierungen ermittelten Fragmenthäufigkeiten sowie die exakten Längen der erwarteten Fragmente wieder.

Die Analyse von neun Klonen eines Individuums sowie von vier Klonen dreier weiterer Individuen ergab dreizehn unterschiedliche Sequenzen. Die mittlere Homologie der neun Klone von Individ1 in dem 365 bp langen Sequenzabschnitt, welcher sowohl in den klonierten *Bam*-HI Higher order repeats als auch in den PCR-Fragmenten sequenziert wurde (Abb. 9a und d), liegt bei 97,9% mit einer minimalen Homologie von 96,8% (Tab. 3). Diese, verglichen mit den Homologien des entsprechenden Abschnittes von neun zufällig aus allen fünf PACs ausgewählten Higher order repeats (98,8% bzw. 97,3%), leicht divergiertere Sequenzen deuten auf eine breitere Repräsentation des gesamten DXZ1-Arrays durch die PCR verglichen mit den begrenzten Segmenten der PAC-Klone hin. Ein Teil der höheren Divergenz könnte auf PCR-Fehler der *Taq*-Polymerase zurückzuführen sein, was bei einer geschätzten Fehlerrate von 0,2% (Warburton und Willard 1992) aber nicht die alleinige Ursache sein kann.

Die Sequenzanalysen ergaben darüber hinaus, dass die Erkennungssequenzen der Restriktionsenzyme durch dieselben variablen Basen beeinflusst werden wie in den PACs,

was die Eignung der PCR zur Analyse der Homogenisierung und Fixierung dieser Varianten unterstreicht. Von den zehn durch die Analyse der PACs als global definierten Varianten wurden acht auch in den sequenzierten PCR-Produkten als variabel identifiziert. Drei der variablen Basen in den PCR-Produkten traten in nur einem der fünf PACs auf. Eine PCR-Variante wurde in den PACs nicht identifiziert, was darauf hindeutet, dass die weitaus meisten, aber nicht alle häufigen Varianten durch die Analyse der PACs eines einzigen Individuums identifiziert wurden.

3.3. Restriktionsverdau und Southern Blot kompletter genomischer DNA

Um die relative Häufigkeit einer Variante in den gesamten Zentromer-Arrays verschiedener Individuen noch besser abschätzen zu können und um deren Fixierung in der Bevölkerung ohne PCR-Technik zu bestätigen, wurde die genomische DNA von sechs der zehn mittels PCR untersuchten Individuen mit *Bam*-HI alleine sowie mit *Bam*-HI und *Dra*-III verdaut. Nach Auftrennung der DNA auf einem Agarosegel, Southern Blotting und Hybridisierung mit der spezifisch X-chromosomalen Probe *pBamX5* konnte der Anteil der 2 kb *Bam*-HI Fragmente bestimmt werden, welcher durch *Dra*-III in zwei 1,5 und 0,5 kb lange DNA-

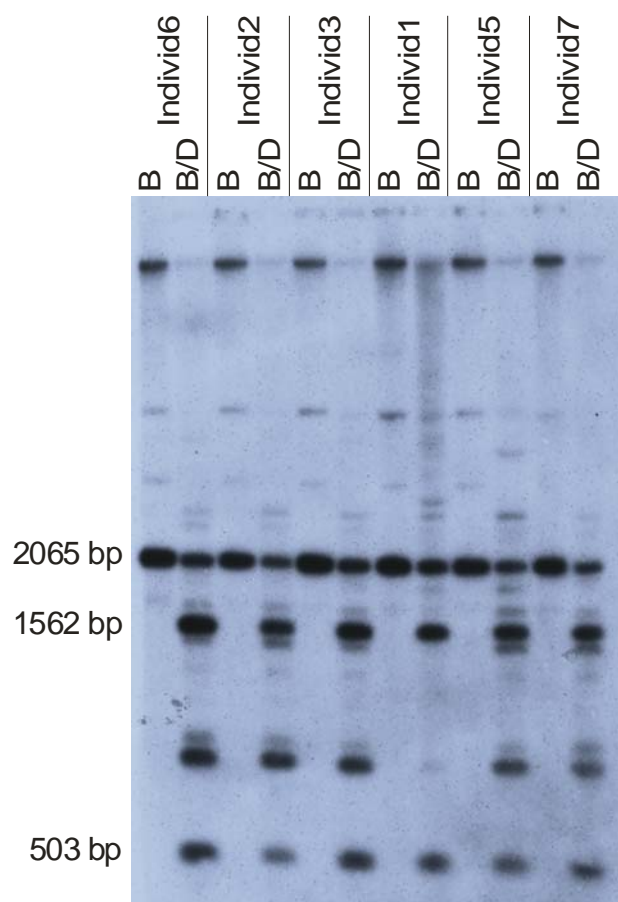


Abb. 14: Die Bestimmung der Häufigkeit der *Dra*-III Variante an Position 1562 durch den Verdau von kompletter genomischer DNA mit anschließender Hybridisierung bestätigt deren Fixierung in der Bevölkerung: Die komplette genomische DNA von 6 nichtverwandten männlichen Individuen wurde mit *Bam*-HI alleine (B) sowie mit *Bam*-HI und *Dra*-III (B/D) verdaut. Nach Auftrennung auf einem Agarose-Gel (1,0%) und Southern Blotting auf eine Nylon-Membran erfolgte die Hybridisierung mit einer spezifisch X-chromosomalen α -Satelliten Probe. Dabei zeigte sich, dass bei allen Individuen der gleiche Anteil an 2065 bp langen *Bam*-HI Fragmenten in zwei 1562 und 503 bp lange Fragmente gespalten wird, und somit die Variante an Position 1562 der X-chromosomalen Higher order repeats in allen Individuen im gleichen Verhältnis vorliegt. Autoradiographie mit ^{32}P .

Fragmente gespalten wurde (Abb. 14). Dies zeigte, dass alle sechs untersuchten Individuen die G/C Variante an Position 1562 in einem Verhältnis von circa 3/1 besitzen, in etwa übereinstimmend mit dem Verhältnis, welches durch Verdau der X-3A/X-4A PCR-Produkte ermittelt wurde. Dadurch konnten sowohl die Anwendbarkeit der spezifischen α -Satelliten-PCR zur Analyse von variablen Basen als auch die Fixierung der G/C Variante in der Bevölkerung bestätigt werden.

4. Die variablen Basen treten in den Higher order repeats unabhängig voneinander auf

Viele Untersuchungen von α -Satelliten-DNA, häufig an den stärker divergierten und somit einfacher zu analysierenden Rändern der Zentromer-Arrays, deuteten auf die Existenz von einer begrenzten Zahl an Haplotypen der Higher order repeats hin, was mit dem Auftreten von Unequal crossing over im Bereich der Zentromer DNA in Verbindung gebracht wurde (Willard et al. 1987, Durfy und Willard 1989, Warburton und Willard 1992, Warburton und Willard 1995). Um zu evaluieren, ob sich die in dieser Arbeit analysierten Higher order repeats aus 5 Regionen eines DXZ1-Arrays bestimmten Haplotypen zuordnen lassen würden, wurden sie entsprechend dem Auftreten der globalen Varianten geordnet (Abb. 15). Dabei zeigte sich, dass die 18 globalen Varianten in den 46 untersuchten Higher order repeats in 40 verschiedenen Kombinationen vorkommen. Identische Kombinationen fanden sich innerhalb der PACs A7 (2x) und B11 sowie zwischen den PACs A6 und A7 beziehungsweise A8 und A10 (zusätzlich zu den 2 Higher order repeats mit identischen Endsequenzen). Versuche, die 40 Kombinationen zu gruppieren, indem man sie entsprechend dem Auftreten der Varianten vom vorderen oder vom hinteren Ende der Repeats ausgehend sortiert, lassen weder die Existenz von Haplotypen noch die Zugehörigkeit eines Repeats zu einem PAC anhand eines spezifischen Musters von Varianten erkennen. Es können im Gegenteil ähnliche Anordnungen von Varianten in allen möglichen Kombinationen zweier PACs gefunden werden. Da diese wiederum, mit Ausnahme der PACs A8 und A10, kaum überlappen können (vgl. Kap. 5.3), kann somit auf das Vorkommen ähnlicher Varianten-Kombinationen auch in größeren Abständen innerhalb des Zentromer Arrays geschlossen werden (Abb. 15).

Innerhalb der Higher order repeats scheinen die einzelnen Varianten völlig unabhängig voneinander aufzutreten. Zur Veranschaulichung können die variablen Basen an den Positionen 1821 und 1822 dienen: Obwohl diese zwei Varianten direkt nebeneinander liegen, kommen sie in den PACs A10 und B11 in allen vier möglichen Kombinationen vor. Dieses

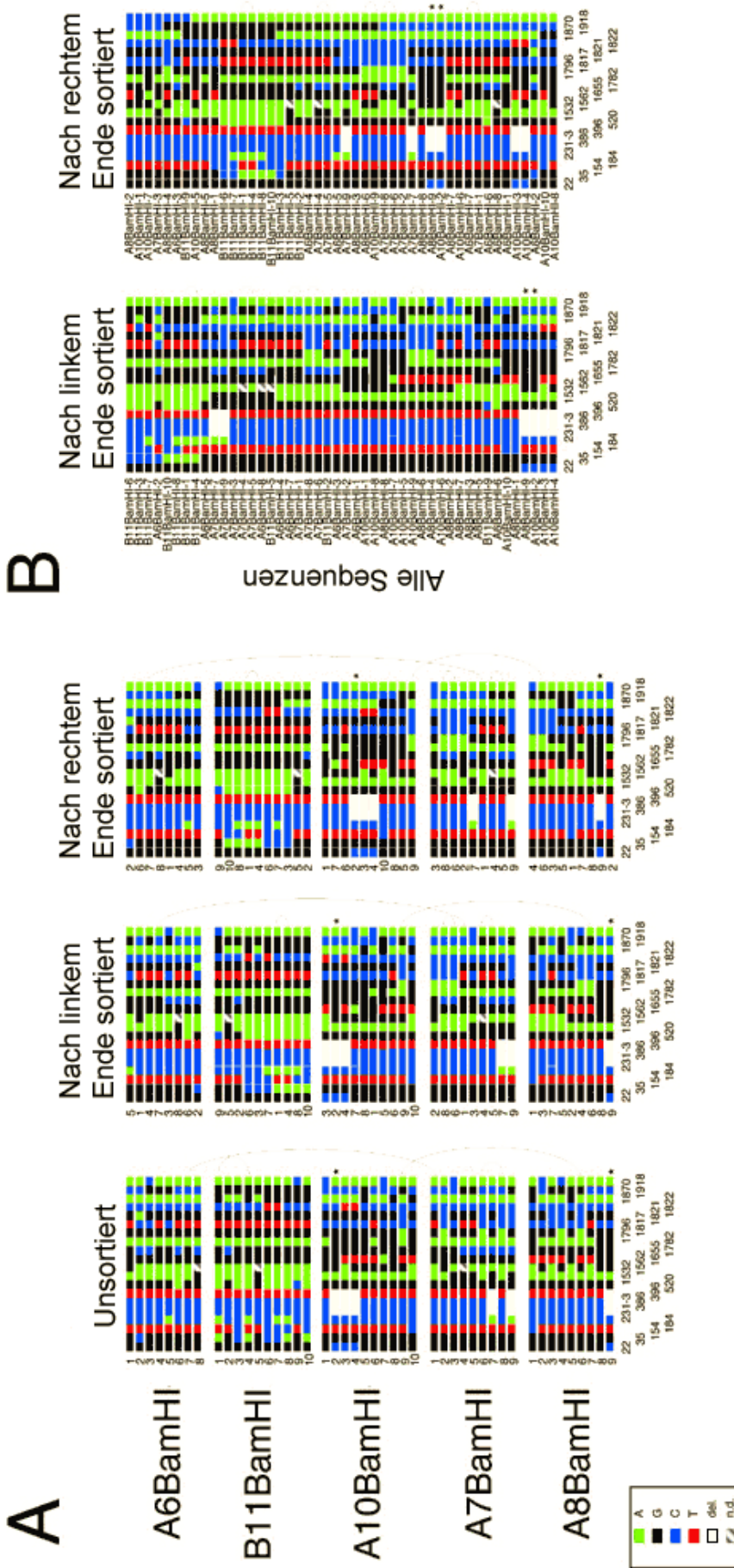


Abb. 15: Das Fehlen von Haplotypen widerspricht einer Homogenisierung der α -Satelliten DNA durch die Amplifikation ganzer Higher order repeats: Dargestellt sind die 18 als „global“ definierten Varianten in den 46 sequenzierten Higher order repeats, (a) aufgetrennt nach dem Ursprungs-PAC sowie (b) unabhängig von der Herkunft. Jedes Nukleotid ist durch ein entsprechend gefärbtes Viereck dargestellt, die Positionen der Varianten in den Higher order repeats sind unten angegeben. Die variable Deletion von drei Basenpaaren an den Positionen 231-233 ist durch drei Vierecke wiedergegeben. Die fünf in je zwei Higher order repeats identischen Varianten-Kombinationen sind durch Linien miteinander verbunden, die zwei Bam-HI Fragmente mit identischen Endsequenzen mit Sternen gekennzeichnet. Wie der Versuch die Kombinationen von Varianten vom linken oder vom rechten Ende her zu sortieren zeigt, lassen sich die Higher order repeats weder bestimmten Haplotypen noch ihrem Ursprungs-PAC zuordnen. Es treten im Gegenteil ähnliche Anordnungen von Varianten in allen möglichen Kombinationen zweier PACs auf.

Fehlen von Haplotypen, sowohl innerhalb der PACs mit Längen um 140 kb als auch bei Betrachtung aller sequenzierten Fragmente, die auch Entfernungen von über 260 kb aufweisen, macht eine Homogenisierung des X-chromosomalen Zentromers durch Amplifikationen ganzer Higher order repeats höchst unwahrscheinlich. Das voneinander unabhängige Auftreten der variablen Basen lässt hingegen am ehesten auf Genkonversionen mit sehr kurzen Sequenzübertragungen als deren Ursprung schließen. Durch die bisherigen Ergebnisse nicht auszuschließen sind jedoch Amplifikationen von Higher order repeats in kleinerem Maßstab von wenigen Kilobasen, die in den analysierten Stichproben nicht erfasst worden wären.

5. Kartierung von Varianten innerhalb der α -Satelliten-PACs

Um die Homogenität der 2 kb Tandemrepeat-Arrays der X-chromosomalen α -Satelliten-DNA bestimmen zu können und um die Ausbreitung von Restriktionsvarianten innerhalb der PACs genauer untersuchen zu können, wurde eine Methode etabliert mit der die über 100 kb langen Arrays mittels partialer Restriktionsspaltung durchgängig kartiert werden konnten. Dazu wurde die aufgereinigte, intakte PAC-DNA mit dem nur innerhalb des Vektors schneidenden Restriktionsenzym *Pac-I* vollständig linearisiert und anschließend mit den Restriktionsenzymen *Bam-HI*, *Dra-III* oder *Hind-III* partial gespalten. Die Hybridisierung eines Southernblots mit Vektor-Endproben macht dann die Fragmenteleitern sichtbar (s. Abb. 16, 17). Integrierte Karten der fünf PACs, welche durch komplette und partielle Restriktions-

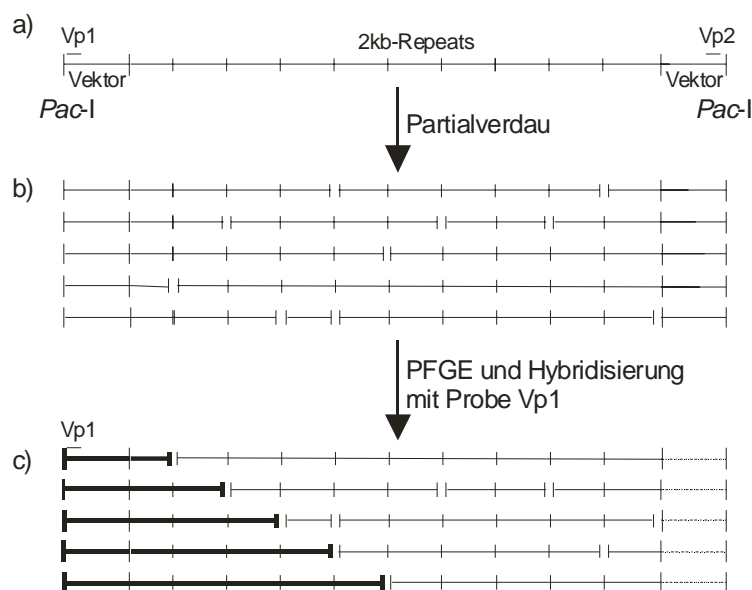


Abb. 16: Die Kartierung von Restriktionsschnittstellen erfolgte mittels Partialverdau von PAC-DNA: Für die Kartierung der Restriktionsschnittstellen wurde die aufgereinigte PAC-DNA zunächst innerhalb des Vektors vollständig mit dem Enzym *Pac-I* gespalten (a). Nach Partialverdau mit einem geeigneten Restriktionsenzym (b) folgte die Auftrennung auf einem Pulsfeldgel und die Übertragung auf eine Nylon-Membran mittels Southern-Blotting. Die Hybridisierung mit einer radioaktiv markierten Probe, welche spezifisch die Vektor-DNA im Bereich der *Pac-I* Restriktionsschnittstelle bindet (VP1 bzw. VP2), ermöglichte die exakte Kartierung der *Bam-HI* Fragmente sowie von zwei variablen Basen (c).

Kartierungen, sowie die Endsequenzierungen der PAC-Inserts gewonnen wurden, sind in Abb. 18 dargestellt.

5.1. Bestätigung der Homogenität der PACs durch Kartierung mittels *Bam*-HI

Die Kartierung der *Bam*-HI Restriktionsschnittstellen ergab, dass die PACs A7 und A10 ausschließlich aus 2 kb *Bam*-HI Fragmenten bestehen, und dass der PAC A8 nur ein einziges davon abweichendes Fragment besitzt, welches anhand des kompletten Verdaus der PAC-DNA auf 1,7 kb geschätzt wurde. Die PACs A6 und B11 besitzen hingegen neun respektive

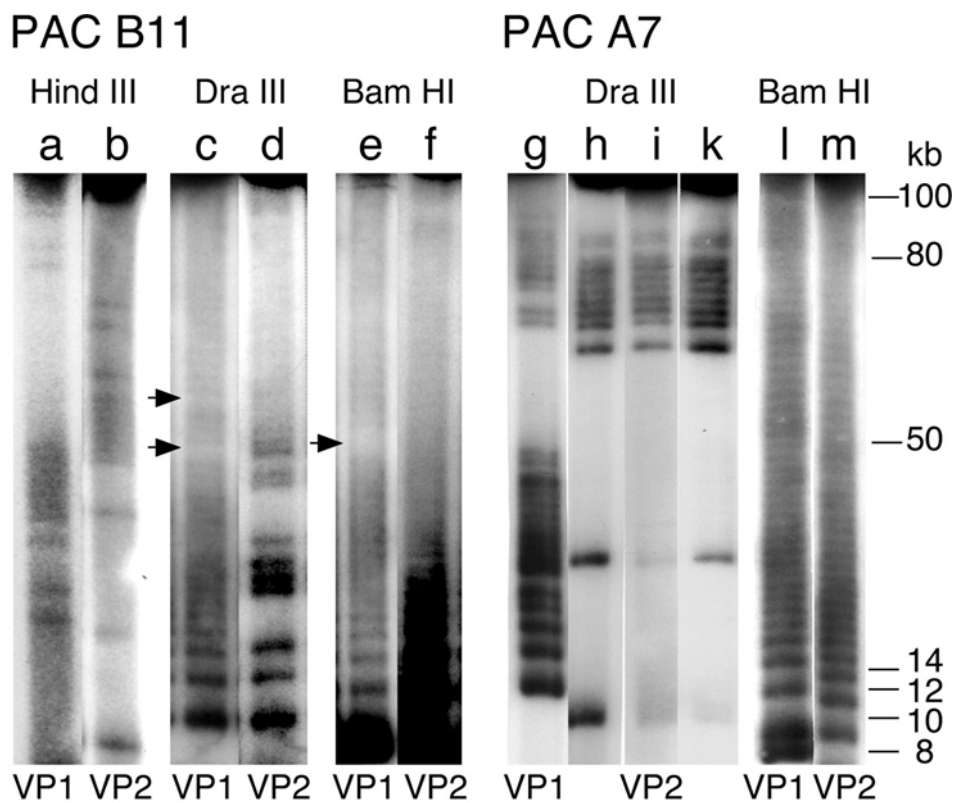


Abb. 17: Partiale Restriktionskartierungen der Higher order repeats und variablen Basen: Dargestellt sind die Kartierungen der *Bam*-HI, *Dra*-III und *Hind*-III Restriktionsschnittstellen in den PACs A7 und B11. VP1 und VP2 stellen die beidseits der *Pac*-I Restriktionsschnittstelle im PAC-Vektor gelegenen Hybridisierungs-Proben dar. Während der sehr homogene PAC A7 ausschließlich aus 2 kb *Bam*-HI Fragmenten besteht (l, m), zeigt der etwas stärker divergente PAC B11 geringfügige Abweichungen in den *Bam*-HI und *Dra*-III Mustern (Pfeile in c und d). Die *Hind*-III Schnittstellen im PAC B11 sowie die *Dra*-III Schnittstellen im PAC A7 treten in Clustern von 7-10 Higher order repeats auf (a, b, g, h). Dabei sind die beiden *Hind*-III Cluster im PAC B11 vermutlich durch eine Duplikation von 35-50 kb entstanden (s. Abb. 17). Als Kontrolle wurden zwei Kolonien des PACs A7 für 6 Tage (~400 Generationen) täglich neu auf Agar ausplattiert. Die unveränderten Restriktionsmuster (i, k) schließen Rekombinationsmechanismen im *E. coli* Wirtsbakterium als Ursache der Clusterbildung aus und belegen die hohe Stabilität der PAC-Insertlänge und der Anordnung der einzelnen Repeats in *E. coli*. Autoradiographie mit ^{32}P .

drei Fragmente mit einer von 2 kb abweichenden Länge, die nicht den Klonierungsrändern zugeordnet werden können. Diese Fragmente, deren Genese und exakte Sequenz unbekannt sind, liegen unregelmäßig in den PACs verteilt. Die Kartierungen zeigen somit eine geringere Struktur-Homogenität bei denjenigen PACs, die auch eine etwas geringere Sequenz-Homogenität aufweisen (Abb. 17e, f, l, m, Abb. 18, Tab. 2).

5.2. Die Varianten treten in den PACs in Clustern auf

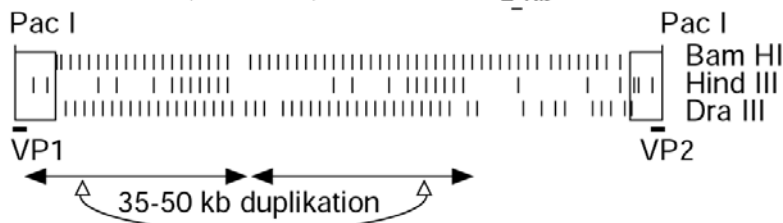
Globale Varianten, welche die Erkennungssequenz von Restriktionsenzymen beeinflussen und in den Higher order repeats nur einmal auftreten, ermöglichten die Kartierung von zwei dieser Polymorphismen nach der oben beschriebenen Methode. So konnten die G/C Variante an Position 1562 des Higher order repeats durch Spaltung mittels *Dra*-III (5'-CAC NNN GTG-3') sowie die C/G beziehungsweise C/A Variante an Position 35 mittels *Hind*-III (5'-AAG CTT-3') kartiert werden (s. Tab. 4 und Abb. 8). Die Variante an Position 35 und die davon abhängige DNA-Spaltung durch *Hind*-III wurde bereits 1987 durch Durfy und Willard erstmals beschrieben. Der vollständige Verdau von genomischer DNA mit anschließender Hybridisierung durch die spezifisch X-chromosomale Probe *pBamX7* hatte ein Verhältnis von 2 kb *Hind*-III zu 2 kb *Bam*-HI Fragmenten von 1 / 10 bis 1 / 100 ergeben (Willard et al. 1986). Ein Screening unabhängiger Individuen hatte auch gezeigt, dass der Anteil an *Hind*-III Restriktionsschnittstellen auf verschiedenen X-chromosomalen Zentromeren relativ variabel ist, und die Analyse der Mitglieder dreier Familien bewies die mendelische Vererbung des spezifischen Musters von unterschiedlich langen *Hind*-III Fragmenten. Ähnlich variable Restriktionsmuster wurden auch durch Verdau mit den Enzymen *Xba*-I, *Eco*-RI, *Bgl*-II und *Msp*-I erzielt.

Obwohl diese großen Unterschiede im Vorkommen von Restriktionsschnittstellen nicht mit den Ergebnissen der direkten Sequenzierung von PCR-Produkten durch dieselbe Arbeitsgruppe (Warburton und Willard 1992, Warburton und Willard 1995) sowie unseren Resultaten einer weitestgehenden Fixierung aller untersuchten Varianten vereinbar sind (s. Kap. 3), war es uns nicht möglich die Verbreitung der *Hind*-III Variante in der Bevölkerung genauer zu bestimmen. Die *Hind*-III Restriktionsschnittstelle liegt außerhalb des verwendeten PCR-Bereichs und nur 31 bp von der *Bam*-HI Schnittstelle entfernt, wodurch eine Unterscheidung auf einem Agarosegel nicht möglich ist. Der insgesamt relativ geringe Anteil an *Hind*-III spaltbaren α -Satelliten-Repeats (Willard et al. 1986) sowie das ausschließliche Auftreten dieser Variante in den divergierten PACs B11 und A6 deuten jedoch darauf hin,

PAC A6 (160 kb)



PAC B11 (125 kb)



PAC A10 (130 kb)



PAC A7 (145 kb)



PAC A8 (120 kb)



Abb. 18: Integrierte Karten der X-chromosomalen α -Satelliten PACs A6, A7, A8, A10 und B11. Während vor allem die PACs A7, A8 und A10 bezüglich der *Bam*-HI Restriktionsschnittstellen hochgradig homogen sind, treten die *Dra*-III und *Hind*-III Schnittstellen und somit die Varianten an den Positionen 1562 beziehungsweise 35 der 2 kb Higher order repeats in Clustern von unterschiedlicher Größe auf. Der PAC B11 enthält vermutlich eine Duplikation von 35-50 kb (Doppelpfeile).

dass sie lediglich in einem begrenzten und stärker divergierten Bereich des X-chromosomalen Zentromers vorkommt. Auch alle anderen von Willard et al. (1986) untersuchten Varianten mit variabler Häufigkeit in der Bevölkerung waren selten, und es ist unklar in welchen Bereichen des Zentromer-Arrays diese variablen Basen vorkommen, was einen relativen Ausschluss von den aktivsten Prozessen der Homogenisierung und Fixierung erklären könnte. Abb. 17 zeigt die partialen Restriktionskartierungen der PACs A7 und B11 mit den Restriktionsenzymen *Bam*-HI, *Dra*-III und *Hind*-III. Auffallend sind jeweils zwei Cluster aus 7-10 *Dra*-III respektive *Hind*-III enthaltenden Higher order repeats, umgeben von einzelnen Restriktionsschnittstellen im Abstand von einem bis drei Repeats. Zwischen den 24-30 kb großen Clustern liegen ebenso lange Abschnitte welche diese Restriktionsschnittstellen nicht enthalten (Abb. 18). Die Verteilung der *Hind*-III Varianten im PAC B11 (entsprechend einem Cytosin an Position 35 des *Bam*-HI Repeats) in Form zweier identischer Cluster deutet dabei

auf die Tandem-Duplikation eines 35-50 kb langen Abschnitts hin, möglicherweise verursacht durch ein Unequal crossing over Ereignis.

5.3. Die PACs repräsentieren mindestens 0,51 Mb des DXZ1 Arrays eines einzelnen menschlichen X-Chromosoms

Weitere Partialverdau zeigen, dass die PACs A8 und A10 durchgehend eine *Dra*-III Restriktionsschnittstelle und damit ein Guanin an Position 1562 der *Bam*-HI Repeats aufweisen, während der PAC A6 einzelne, zufällig verteilte Repeats besitzt, welche nicht *Dra*-III spaltbar sind. Da ein Doppel-Verdau mit *Bam*-HI und *Hind*-III zeigte, dass die PACs A6, A7, A8 und A10 in Übereinstimmung mit den Sequenzdaten keine oder nur sehr wenige *Hind*-III Schnittstellen besitzen, wurde auf deren Kartierung verzichtet.

Die partialen Kartierungen zeigen klar, dass die PACs A8 und A10, bis auf ein am Rand von A8 gelegenes 1,7 kb Fragment, keinerlei Differenzen in ihren Restriktionsmustern aufweisen, womit ein fast vollständiges Überlappen dieser PACs nicht auszuschließen ist (Abb. 18). Da die PACs A6, A7 und B11 aufgrund der *Dra*-III Varianten jedoch kaum überlappen können, lässt dies den Schluss zu, dass die fünf PACs einen Bereich von mindestens 0,51 Mb abdecken, was 12,7 - 50% eines normalen DXZ1-Arrays von 1-4 Mb Länge entspricht.

5.4. Bestätigung des menschlichen Ursprungs der Cluster sowie der Stabilität der α -Satelliten-DNA

Um die Entstehung der Cluster während des Wachstums des Wirts-Bakteriums *E. coli* durch Rekombinationen der PACs selbst auszuschließen, wurden zwei Kolonien des PAC A7 über sechs Tage täglich neu auf Agarplatten ausgestrichen. Bei einer angenommenen Verdopplungszeit von 20 Minuten für den *E. coli* Stamm DH10B ergibt sich somit ein Wachstum über circa 400 Generationen pro Kolonie mit jeweils 6 „bottlenecks“ der Bakterien pro PAC Vereinzelung. Anschließend wurden von beiden Kolonien die Längen der PAC-Inserts bestimmt und erneute Kartierungen der *Dra*-III Restriktionsschnittstellen durchgeführt. Diese zeigten für beide Klone eine gegenüber der ursprünglichen DNA-Präparation des PACs A7 unveränderte Verteilung der G/C Variante (Abb. 17 h-k). Dies beweist einerseits, dass die Cluster menschlichen Ursprungs sind, und unterstreicht andererseits erneut die Stabilität der α -Satelliten-DNA im Wirts-Bakterium.

6. Weitere variable Basen treten unabhängig von den *Hind*-III und *Dra*-III Clustern auf

Die Existenz zweier identischer Cluster von *Hind*-III Schnittstellen im PAC B11 lässt vermuten, dass es hier zur Duplikation eines 35 bis 50 kb langen DNA-Abschnittes durch Unequal crossing over oder replikative Transposition gekommen ist (Abb. 18). Wenn seit dieser Duplikation keine weiteren Rekombinationsereignisse mehr eingetreten sind, dürften die insgesamt zwölf im PAC B11 vorkommenden 2 kb *Hind*-III Fragmente maximal sechs verschiedene Kombinationen von variablen Basen aufweisen. Die Existenz von mehr als sechs in den Varianten unterschiedlichen *Hind*-III Fragmenten würde hingegen auf zusätzliche Rekombinationsvorgänge hinweisen, welche nicht mit einer DNA-Amplifikation einhergehen, wie zum Beispiel Genkonversionen. Das Auftreten von Einzelmutationen würde ferner einen Hinweis auf das Alter der Duplikation geben.

6.1. Klonierung von Higher order repeats aus den *Hind*-III Clustern des PAC

B11

Um zu bestimmen, ob jeweils zwei Higher order repeats aus den beiden *Hind*-III Clustern des PAC B11 tatsächlich identisch sind, wurden die 2 kb *Hind*-III Fragmente mehrfach kloniert. Dazu wurde aufgereinigte DNA des PAC B11 mit *Hind*-III vollständig verdaut, die 2 kb-Bande aus dem Agarosegel eluiert und in den pBluescript II SK-Vektor kloniert. Die Längen der Inserts wurden durch Doppelverdau im Vektor mit *Eco*-RI und *Sal*-I bestimmt und 18 der resultierenden 2 kb-Fragmente mit den Primern T3 und T7 endsequenziert.

Das Alignment dieser 18 Sequenzen ergab sieben verschiedene Kombinationen von globalen Varianten. Unter der Annahme, dass die identischen *Hind*-III Cluster durch Tandem-Duplikation eines längeren DNA-Abschnittes entstanden sind, zeigt dies, dass seitdem weitere Rekombinationen von Varianten durch Genkonversionen stattgefunden haben müssen.

Zusätzlich zu den variablen Basen traten auch Einzelmutationen in den *Hind*-III Fragmenten auf. Diese gehen jedoch stets mit einer individuellen Kombination von Varianten einher, womit eine Unterscheidung der Fragmente unabhängig von den variablen Basen nicht möglich ist. Dies könnte entweder darauf zurückgeführt werden, dass die Rekombinationsvorgänge erst kürzlich stattfanden und somit nicht ausreichend Zeit für die Ansammlung von Mutationen war, oder dass die aufgetretenen Einzelmutationen effektiv wieder entfernt wurden.

6.2. Klonierung von Higher order repeats aus den *Dra*-III Clustern des homogenen PAC A7

Da es sich bei dem PAC B11 um einen stärker divergierten PAC handelt, der eine hohe Anzahl an lokalen Varianten aufweist und somit möglicherweise am Rand des DXZ1-Arrays liegt, und außerdem die *Hind*-III Variante scheinbar nicht in der Bevölkerung fixiert ist, konnte nicht mit Sicherheit gesagt werden, dass die Ergebnisse der *Hind*-III Sequenzierungen auch für diejenigen Bereiche des α -Satelliten-Arrays gelten, in denen aktuell eine effizientere Homogenisierung stattfindet. Daher führten wir zusätzlich eine Analyse der *Dra*-III Cluster des homogenen PAC A7 durch.

6.2.1. Konstruktion des Vektors BsD

Da kein geeigneter Vektor für die Klonierung der *Dra*-III Fragmente mit ihrem drei-Basen 5'-Überhang zur Verfügung stand, musste der Vektor pBluescript II SK entsprechend modifiziert werden. Dazu wurde zunächst eine *Dra*-III Schnittstelle mit nicht kompatiblen 5'-Überhang im fl-ori Bereich entfernt. Hierfür wurde der Vektor mit *Dra*-III linearisiert und die überhängenden 5'-Enden mit S1-Exonuklease entfernt. Unter Verlust der drei mittleren Basen der ursprünglichen *Dra*-III Erkennungssequenz wurde die DNA blunt religiert und der Vektor Bs Δ D in den *E. coli* Stamm DH10B kloniert.

Danach wurde die mittels Quiagen Spin Miniprep Kit präparierte Bs Δ D-DNA mit den Enzymen *Hind*-III und *Bam*-HI verdaut, welche im Bereich der Multiple cloning site des Vektors spalten und zueinander inkompatible 5'-Überhänge produzieren. In den so linearisierten Vektor wurde ein doppelsträngiges Oligonukleotid kloniert, welches durch Hybridisierung der einzelsträngigen synthetischen Oligonukleotide DraF und DraR erhalten wurde. Dieses Oligonukleotid enthält eine *Dra*-III Restriktionsschnittstelle mit der Basenfolge 5'-CAC TGG GTG-3'. Da die mittleren drei Basen der Erkennungssequenz von *Dra*-III (5'-CAC NNN GTG-3') und somit auch der drei-Basen 5'-Überhang nach einem Restriktionsverdau variabel sind, könnte dies zum Ausschluss von *Dra*-III Fragmenten aus der Klonierung führen. Da jedoch alle sequenzierten Higher order repeats in diesen drei Basen identisch waren, wurde nur ein Vektor mit der entsprechenden Sequenz kloniert. Das Oligonukleotid wurde außerdem so synthetisiert, dass das lacZ-Gen des Vektors pBluescript weiterhin „in frame“ translatiert wird und somit eine blau/weiß Selektion entsprechender Klone nach Zugabe von X-Gal möglich ist. Der resultierende Vektor mit einer *Dra*-III Schnittstelle im Bereich der Multiple cloning site wurde als BsD bezeichnet.

6.2.2. Auch in den *Dra*-III Clustern des homogenen PAC A7 treten die variablen Basen unabhängig voneinander auf

Um die 2 kb Fragmente aus den *Dra*-III Clustern des PAC A7 zu klonieren, wurde die PAC-DNA mit *Dra*-III verdaut und durch Ethanol-fällung gereinigt. Die resultierenden Fragmente wurden direkt in den mit *Dra*-III linearisierten Vektor BsD kloniert. Nach Minipräparation von einer Reihe von Klonen und Bestimmung der Insertlängen durch Verdau mit *Dra*-III wurden 19 2 kb Repeats mit den Vektor-Primern T3 und T7 endsequenziert (Abb. 19).

Das Alignment der Sequenzen ergab vier Paare von identischen Klonen, vermutlich bedingt durch doppelte Klonierung von Higher order repeats. Von den verbleibenden 15 Klonen unterscheiden sich drei Paare ausschließlich anhand von Einzelmutationen, weisen jedoch identische Muster der variablen Basen auf (Abb. 20). Somit besitzen die 21 2 kb *Dra*-III Repeats des PAC A7 mindestens zwölf verschiedene Kombinationen von globalen Varianten. Versuche, diese Kombinationen bestimmten Haplotypen zuzuordnen, ergaben erneut eine voneinander völlig unabhängige Verteilung der variablen Basen in den Repeats (Abb. 20). Dies zeigt, dass die Ausbreitung der Varianten nicht nur innerhalb der Länge eines PAC von circa 140 kb oder des gesamten untersuchten Bereichs des DXZ1-Arrays von mindestens 500 kb, sondern auch im Bereich der Cluster von circa 20 kb unabhängig voneinander erfolgt. Eine derartige Ausbreitung lässt sich nur durch Genkonversionen, nicht jedoch durch Amplifikationen längerer DNA-Abschnitte, zum Beispiel im Rahmen von Unequal crossing over erklären. Würde die Ausbreitung von Varianten im Zentromer-Array durch Unequal crossing over erfolgen, könnte die Amplifikationslänge nur ein Vielfaches der Higher order repeat-Länge betragen, da häufige Rekombinationen zwischen ungleich alignierten 171 bp Monomeren zur Auflösung der Higher order repeat Strukturen führen würden. Eine Amplifikationslänge von mindestens 2 kb würde jedoch eine derart autonome Ausbreitung von Varianten, wie sie die Higher order repeats aufweisen, nicht zulassen.

6.3. Klonierung weiterer *Bam*-HI Higher order repeats des PAC A7 ohne *Dra*-III Schnittstelle

Um Inseln mit Higher order repeats entfernterer Verwandtschaft außerhalb der *Dra*-III Cluster des PAC A7 ausschließen zu können, welche eine gerichtete Konversion erklären könnten, wurden weitere 15 *Bam*-HI Higher order repeats kloniert, welche keine *Dra*-III Schnittstelle enthalten. Dazu wurden erneut eluierte 2 kb *Bam*-HI Fragmente des PACs A7 in den Vektor

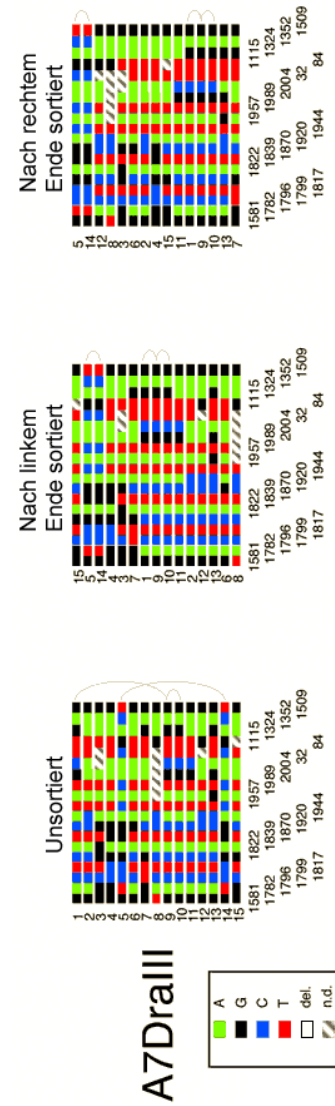
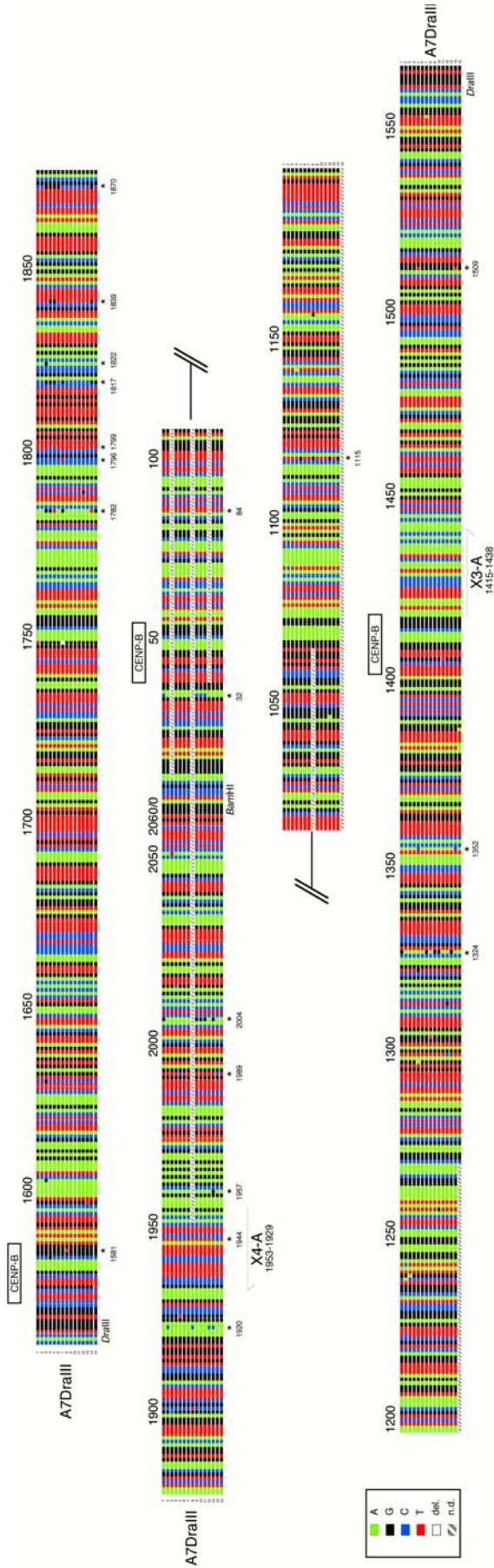


Abb. 19: Sequenz-Alignment von 15 der zwei *Dra*-III Cluster des homogenen PAC A7 bildenden 21 Higher order repeats. Die für die Haplotyp-Analyse verwendeten globalen Varianten sind mit Sternen gekennzeichnet. Die Positionen der CENP-B Boxen sowie der Primer X-3A und X-4A sind angegeben.

Abb. 20: Haplotyp-Analyse der Higher order repeats aus den *Dra*-III Clustern des PAC A7. Dargestellt sind die 19 in den *Dra*-III Fragmenten vorkommenden Varianten, welche in den 15 individuellen Higher order repeats in 12 verschiedenen Kombinationen auftreten. Die weitgehend unabhängige Verteilung der Varianten beweist, dass es innerhalb der Cluster nicht zur Amplifikation bestimmter Haplotypen gekommen ist.

pBluescript II SK kloniert. Von den erhaltenen Bakterienkolonien wurden PCRs mit den Primern X-3A und X-4A aus Plasmid-Schnellpräparationen durchgeführt. Das PCR-Produkt wurde mit *Dra*-III verdaut und auf einem Agarosegel aufgetrennt. Von den nicht *Dra*-III spaltbaren Klonen wurden Minipräparationen angefertigt und anschließend die Plasmide durch Verdau mit *Dra*-III respektive *Bam*-HI erneut auf das Fehlen der *Dra*-III Schnittstelle im Insert sowie eine Insert-Länge von 2 kb kontrolliert. 15 non-*Dra*-III Klone wurden endsequenziert wie unter Methoden beschrieben.

Das Alignment der insgesamt 21 Higher order repeats des PAC A7 ohne *Dra*-III Schnittstelle ergab 18 individuelle Klone aus einem anhand der Kartierungen geschätzten Pool von insgesamt 40 non-*Dra*-III Higher order repeats (Abb. 21). Diese weisen 16 unterschiedliche Kombinationen von globalen Varianten auf, erneut ohne erkennbare Zugehörigkeit zu bestimmten Haplotypen (Abb. 22). Die mittlere Homogenität in dem 365 bp langen Abschnitt zwischen der *Dra*-III Schnittstelle und dem Primer X-4A, welcher für alle Sequenzen (*Bam*-HI Klone, *Dra*-III Klone, *Hind*-III Klone, PCR-Klone) vergleichbar ist (Abb. 9d), beträgt 98,8% für die *Dra*-III und 99,2% für die non-*Dra*-III Klone des PAC A7 (Tab. 3). Die maximale Divergenz zwischen den *Dra*-III und den non-*Dra*-III Klonen von 2,2% ist niedriger als die insgesamt höchste im PAC A7 beobachtete Divergenz von 2,7% zwischen zwei Repeats innerhalb der *Dra*-III Gruppe. Somit besteht kein signifikanter Unterschied in den Higher order repeat Sequenzen zwischen den *Dra*-III und den non-*Dra*-III Bereichen. Die durch die Kartierung ermittelte Homogenität der *Bam*-HI Schnittstellen im PAC A7 (Abb. 18) schließt außerdem die Existenz von inserierter nicht- α -Satelliten-DNA im Array als Ursache der Clusterbildung weitgehend aus. Somit muss es sich bei dem Prozess der Clusterbildung um einen lokalen Vorgang handeln, wohingegen ein „gelenkter“ Fernvorgang äußerst unwahrscheinlich erscheint.

7. Echte Einzelmutationen kommen in den Higher order repeats kaum vor

In den Sequenzen der zuerst analysierten neun *Bam*-HI Higher order repeats des PAC A7 traten 13 Abweichungen von der Konsensussequenz nur einmal auf und wurden daher als Einzelmutationen bezeichnet. Die Sequenzierung der 27 zusätzlichen Higher order repeats (15 *Dra*-III Repeats sowie zwölf *Bam*-HI Repeats ohne *Dra*-III Schnittstelle) führte jedoch dazu, dass echte Einzelmutationen in den homogenen Bereichen der α -Satelliten-Repeats kaum existieren, so dass ihre Häufigkeit vermutlich deutlich unterhalb der ursprünglich für die homogenen PACs bestimmten 0,07-0,14% liegt. Da echte Einzelmutationen somit

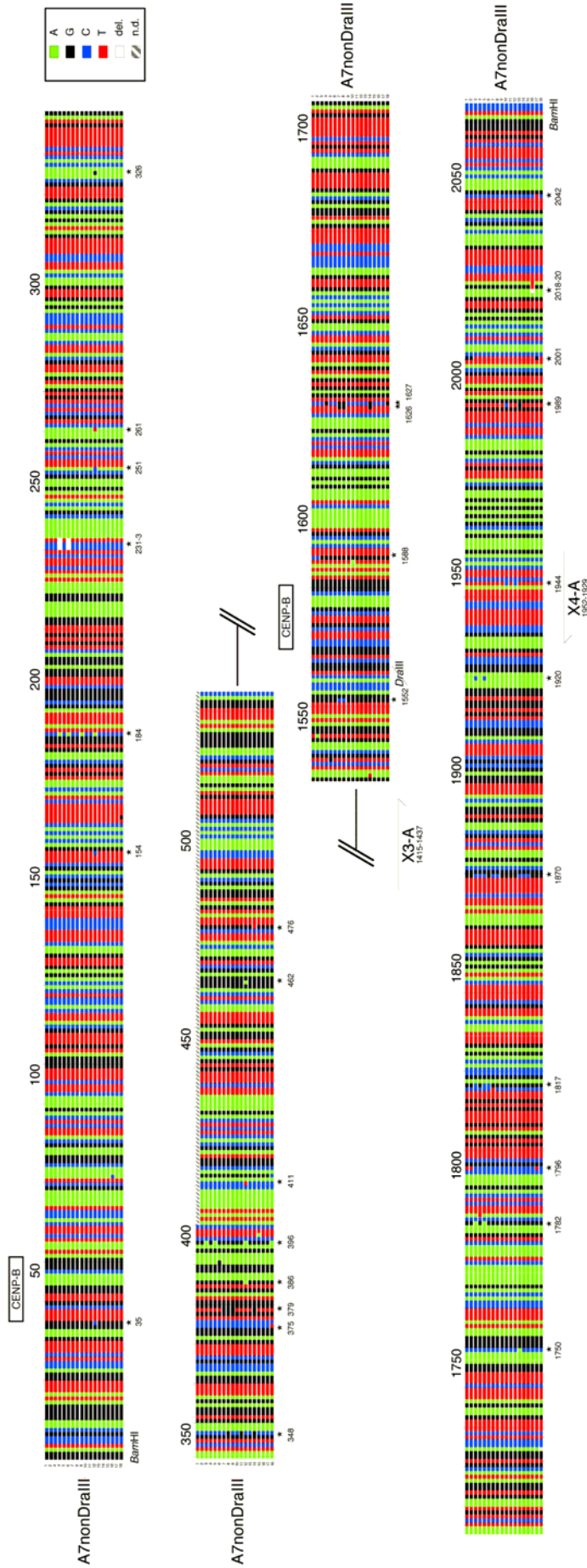
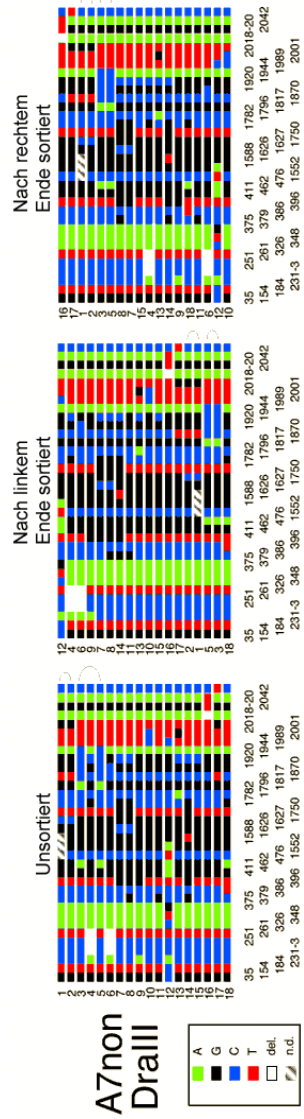


Abb. 21: Sequenz-Alignment von 18 der 40 existierenden non-Dra-III Higher order repeats des PAC A7. Die für die Haplotyp-Analyse verwendeten globalen Varianten sind mit Sternen gekennzeichnet. Die Positionen der CENP-B Boxen sowie der Primer X-3A und X-4A sind angegeben.

Abb. 22: Haplotyp-Analyse der non-Dra-III Higher order repeats des PAC A7. Dargestellt sind die 30 als "global" definierten Varianten, welche in den 18 individuellen Higher order repeats in 16 verschiedenen Kombinationen auftreten. Der Versuch die Higher order repeats zu sortieren zeigt, dass auch die Varianten in den non-Dra-III Regionen des PAC A7 unabhängig voneinander auftreten und somit keine Amplifikation von Haplotypen stattgefunden hat.



höchstens jede tausendste Base auftreten, zeigt dies, dass eine Sequenzierung von Zentromeren mit der heute gängigen Methode eines Assembly von zufällig verteilten Sequenzfragmenten nicht möglich ist.

8. Die im menschlichen X-chromosomalen α -Satelliten-Array vorhandenen

Varianten sind in der Spezies *Gorilla gorilla* nicht nachweisbar

Um zu überprüfen, ob im X-chromosomalen α -Satelliten-Array der Spezies *Gorilla gorilla*, welcher der zurzeit bekannte am engsten mit einem menschlichen Zentromer-Array verwandte α -Satelliten-Typ ist, dieselben variablen Basen vorkommen wie im menschlichen Zentromer, wurde eine PCR-Amplifikation des zur X-3A/X-4A PCR korrespondierenden Higher order repeat-Abschnittes durchgeführt. Das Auftreten identischer variabler Basen in den humanen und den gorilla α -Satelliten-Repeats hätte somit einen Hinweis auf die Geschwindigkeit der Homogenisierung und Fixierung der Varianten erlaubt. Außerdem würde das Vorkommen von variablen Basen an ähnlichen Lokalisationen innerhalb der Higher order repeats der beiden Spezies auf eine Abhängigkeit von der primären α -Satelliten-Sequenz hinweisen.

Um die Spezifität zu erhöhen wurde der Primer X-4A entsprechend der Sequenz des 2 kb Higher order repeats GGSATG eines Gorilla X-Chromosoms (Durfy und Willard 1990) modifiziert. Mit dem resultierenden Primer X-4Ag und dem unveränderten X-3A Primer wurden PCRs aus der genomischen DNA dreier männlicher Flachlandgorillas durchgeführt. Der Verdau mit den Restriktionsenzymen *Hinf-I*, *Dra-III*, *Taq-I* und *Aci-I* zeigte, dass nur durch *Dra-III* eine teilweise Spaltung des PCR-Produkts erfolgte, wohingegen die anderen Enzyme die DNA entweder vollständig oder überhaupt nicht spalteten.

Zur genaueren Analyse der α -Satelliten-Repeats wurde das X-3A/X-4Ag PCR-Produkt zweier Gorillas in den Vektor pGEM-T kloniert und mit den Primern T7 und SP6 doppelstrangsequenziert. Das Alignment von elf Sequenzen ergab drei identische Klone von zwei verschiedenen Gorillas. Die verbleibenden neun Klone weisen Übereinstimmungen zur Konsensus-Sequenz aus allen in dieser Arbeit sequenzierten humanen α -Satelliten-Repeats von 92-95% auf. Dabei besitzen jedoch zwei Gorilla-Sequenzen nur Übereinstimmungen von 91-92% mit den übrigen sieben Gorilla-Sequenzen. Dies bedeutet, dass durch die X-3A/X-4Ag PCR zwei unterschiedliche Arrays von X-chromosomaler Gorilla α -Satelliten-DNA (Durfy und Willard 1990) amplifiziert werden, was eine Varianten-Analyse durch den Restriktionsverdau des PCR-Produkts nicht zulässt. So resultiert zum Beispiel der

unvollständige *Dra*-III Verdau aus einer zwischen den zwei Array-Typen differierenden Base, welche zwei Nukleotide von der humanen *Dra*-III Variante entfernt liegt. Im Bereich der Erkennungssequenzen der übrigen Restriktionsenzyme sind, in Übereinstimmung mit den Restriktionsspaltungen, alle Sequenzen beider Array-Typen identisch. Die sieben Sequenzen des einen Array-Typs weisen nur eine variable Base auf, welche jedoch in den menschlichen α -Satelliten-Repeats nicht existiert. Von den acht Einzelmutationen in den Gorilla-Sequenzen tritt eine als Variante im PAC A7 auf.

Obwohl also die X-3A/X-4Ag PCR, bedingt durch die Repräsentation zweier unterschiedlicher X-chromosomaler α -Satelliten-Arrays, nur eingeschränkt Aussagen zulässt, deuten die Ergebnisse darauf hin, dass die Varianten der menschlichen α -Satelliten-Repeats in den Gorilla-Zentromeren praktisch nicht existieren. Somit muss die Zeitdauer der Homogenisierung und Fixierung einer Variante relativ kurz sein, vereinbar mit den Ergebnissen anderer Studien zur Geschwindigkeit von Concerted Evolution (Elder und Turner 1994). Außerdem bestätigt die Analyse, dass die Lokalisationen der variablen Basen unabhängig von der primären α -Satelliten-Sequenz sind, was eine Abhängigkeit von bestimmten Sequenzmotiven, wie zum Beispiel CENP-B Boxen, oder einer lokalisierten, sequenzabhängigen Topoisomerase II-Aktivität (Florida et al. 2000, Spence et al. 2005), ausschließt.

Zusammengefasst belegen die Experimente erstmals, dass in den homogensten Abschnitten der α -Satelliten-Arrays der menschlichen Zentromere relativ schnelle, lokalisierte Genkonversionen stattfinden müssen.

C. Methoden

1. DNA-Präparationsmethoden

1.1. Plasmidisolierung mittels Mini-Präparation

Für die Gewinnung von Plasmid-DNA mit bis zu 8 kb Länge wurde das QIAprep Spin Miniprep Kit von Quiagen verwendet. Es wurden in der Regel 6 ml einer Übernachtskultur pro Säule eingesetzt. Die Aufreinigung der DNA erfolgte nach dem Protokoll des Herstellers unter Verwendung einer Mikrozentrifuge. Abweichend vom Protokoll wurde die Säule nach dem letzten Waschschrift einige Minuten bei Raumtemperatur stehen gelassen, damit letzte Ethanolreste verdunsten. Anschließend wurde die DNA zunächst mit 30 µl Puffer EB (vom Hersteller) eluiert, worauf eine weitere Elution mit 30 µl TE folgte. Die Eluate wurden vereinigt und eine semiquantitative Bestimmung der DNA-Konzentration über ein Agarosegel durchgeführt.

TE-Puffer: 10 mM Tris, pH 7,4
 1 mM EDTA

1.2. Präparation intakter PAC-DNA in Agarose-Blöckchen

Die Präparation von langer DNA in Agaroseblöckchen erfolgte nach Smith et al. (1988), ergänzt durch einen Pulsfeld-Reinigungsschritt nach Schindelbauer und Cooke (1997). Hierbei wird ausgenutzt, dass intakte, zirkuläre PAC-DNA mit einer Größe von über 40 kb in der Pulsfeld-Gelelektrophorese nicht wandert, gebrochene PACs und das linearisierte *E. coli*-Genom jedoch die Gelblöckchen verlassen. Auf diese Weise können quantitativ zirkuläre DNA-Moleküle mit Längen bis über 300 kb aufgereinigt werden, die weder Doppel- noch Einzelstrangbrüche aufweisen (= intakte DNA). Dadurch werden zum Beispiel effiziente Subklonierungen, partielle Restriktionskartierungen oder der intakte Transfer dieser langen DNA-Moleküle möglich.

1.2.1. Herstellung von Bakterien-Agarose-Blöckchen

Für die Präparation von PAC-DNA wurden 20 µl einer Glycerinkultur auf einer Kanamycin-Agarplatte ausgestrichen und ü.N. bei 37°C inkubiert. Am nächsten Tag wurde eine Kolonie mit einer sterilen Impföse in 75 ml LB_{Kana} (Kanamycin 30 mg/ml, 1/1000 in LB₀) überführt und im Schüttler bei 37°C, 280 UpM ü.N. inkubiert. Daraufhin wurden 300 ml LB_{Kana} hinzugefügt und für weitere circa 2,5 h inkubiert, um die Bakterien aus der stationären Phase wieder in die Wachstumsphase zu überführen. Mittels IPTG (1 M in H₂O_{bidest.}, 1/1000) wurde das lytische Replikon des PAC-Vektors induziert und somit dessen Kopiezahl pro Bakterium

erhöht. Nach erneuter Inkubation über 1 h wurde der Kultur Chloramphenicol (84 mg/ml in Ethanol absolut, 1/1000) hinzugefügt. Durch Blockade der Proteinsynthese an den Ribosomen stoppt dies die Neusynthese des Startkomplexes der Replikation, ermöglicht jedoch die Fertigstellung der laufenden Replikation des PAC-Vektors.

Die Bakterien wurden durch wiederholte Zentrifugation in 50 ml Röhrchen (Rotixa/RP-Zentrifuge, 3000 UpM, 4°C, 15 min) pelletiert und anschließend in 10 ml Pett IV gewaschen. Das Pellet wurde, je nach gewünschter DNA-Konzentration der Gel-Blöckchen, in 1-5 Volumen Pett IV resuspendiert. Die Suspension wurde auf 42°C erwärmt und im Verhältnis 1:1 mit geschmolzener, auf 42°C abgekühlter LMP-Agarose (2% LMP-Agarose ultrapure, Gibco BRL in H₂O_{bidest}) vermischt. Anschließend wurde die Bakterien-Agarose-Mischung in mit Tesa-Krepp abgeklebte und erwärmte Gelblöckchenformen pipettiert. Nach dem Erstarren der Agarose konnten die Blöckchen durch Druckluft aus den Gelformen in 50 ml Röhrchen überführt werden.

LB ₀ -Medium:	1% Trypton / Pepton 0,5% Hefeextrakt 1% NaCl
LB-Agar:	20 g Agar auf 1 l LB-Medium 20 ml pro Platte (Ø 10 cm)
Pett IV:	10 mM Tris, pH 7,4 1 M NaCl

1.2.2. Reinigung der Blöckchen von Proteinen

Um die DNA in den Gelblöckchen von den Proteinen und sonstigen Bestandteilen der Bakterien zu reinigen, musste zunächst die Bakterienwand zerstört werden. Dazu wurden die Blöckchen ü.N. bei 37°C in 10 ml EC-Lysis inkubiert, wobei Lysozym und RNase A jeweils frisch zugegeben wurden.

Anschließend wurde der EC-Lysis Puffer durch 10 ml NDS ersetzt und für 1-3 Tage bei 50°C inkubiert. Durch die dem NDS stets frisch hinzugefügte Proteinase K wurden die Bakterienproteine, oder zum Beispiel auch Restriktionsendonukleasen nach DNA-Spaltungen verdaut. Anschließend wurden die Blöckchen mehrere Tage bei 37°C in NDS-Puffer ohne Proteinase K inkubiert, was schließlich zu einem Selbstverdau der verbleibenden Proteinase K führt. Derart aufgereinigt sind die Blöckchen in NDS-Puffer bei Raumtemperatur über viele Jahre lagerbar. Für die weitere Verarbeitung der Agarose-Blöckchen wurde der NDS-Puffer durch mehrmaliges Äquilibrieren in TEE bei 4°C vollständig herausgewaschen.

EC-Lysis:	6 mM Tris, pH 7,4 1 M NaCl 100 mM EDTA, pH 7,8 0,5 % Brij 58 0,2 % Deoxycholat 0,5 % N-Lauroyl-Sarkosin 1 mg/ml Lysozym 20 µg/ml RNase A
NDS:	0,5 M EDTA, pH 9,0 1 % N-Lauroyl-Sarkosin 1 mg/ml Proteinase K
TEe-Puffer:	10 mM Tris, pH 7,4 1 mM EDTA 0,1 mM EGTA

1.2.3. Aufreinigung intakter PAC-DNA

Um die PAC-DNA vom Genom des *E. coli* Wirtsbakteriums zu isolieren, wurde ein Restriktionsverdau mit dem Enzym *Asc*-I durchgeführt. Dieses besitzt eine 8-Basen Erkennungssequenz, welche ausschließlich aus GC-Paaren besteht (5'-GGCGCGCC-3'), so dass es im GC-reichen *E. coli* Genom durchschnittlich alle 28 kb, im AT-reichen humanen Genom jedoch nur alle 670 kb schneidet (NEBiolabs Katalog, 2002/03). Die Verwendung von PACs ohne *Asc*-I Restriktionsschnittstelle ermöglicht somit eine selektive Linearisierung des *E. coli* Genoms. Da sich zirkuläre, supercoiled DNA-Moleküle mit einer Größe von über 30-40 kb, im Gegensatz zu linearen und offen zirkulären DNA-Fragmenten, in der Pulsfeld-Gelelektrophorese nicht bewegen, können das *E. coli* Genom und gebrochene PAC-DNA selektiv entfernt werden, wohingegen die intakte PAC-DNA in den Blöckchen verbleibt (Schindelhauer und Cooke 1997).

Die Restriktionsspaltung mit *Asc*-I wurde wie in Kap. 3.2 beschrieben durchgeführt. Das Restriktionsenzym wurde durch erneute Inkubation mit NDS entfernt und anschließend der NDS-Puffer durch äquilibrieren in TEe herausgewaschen. Daraufhin wurde eine Pulsfeld-Gelelektrophorese durchgeführt wie in Kap. 4.2 beschrieben. Im Unterschied zum dort beschriebenen Vorgehen mussten die Blöckchen jedoch nach der Elektrophorese wieder aus dem Gel zurückgewonnen werden, weshalb ein Eingießen mit Agarose nicht möglich war. Die Blöckchen wurden daher mit TAE-Puffer in die Geltaschen geladen, und ein Herausschwimmen durch einen „Gel-Deckel“ verhindert, welcher mit LMP-Agarose festgossen wurde. Nach der Elektrophorese wurde der Deckel wieder entfernt und die Blöckchen aus den Geltaschen in 50 ml Röhrchen überführt. Anschließend wurde das

Pulsfeld-Gel mit Ethidiumbromid gefärbt, um den Erfolg des *Asc*-I Verdaus zu überprüfen. Die Blöckchen wurden nach der Aufreinigung der DNA bei 4°C in TEE aufbewahrt.

Um die Effizienz der *Asc*-I Spaltung zu erhöhen wurde zum Teil eine zusätzliche PFGE vor dem Restriktionsverdau durchgeführt. Dadurch wurden bereits gebrochene *E. coli*- und PAC-DNA sowie verbleibender NDS-Puffer vollständig aus den Blöckchen entfernt.

1.3. Plasmid-Schnellpräparationen aus Bakterienkolonien

Um Bakterienkolonien mit einem korrekt klonierten Plasmid zu identifizieren wurden Polymerase-Ketten-Reaktionen mit DNA aus Plasmid-Schnellpräparationen als Template durchgeführt. Dazu wurden die Bakterien einer Kolonie mit einer Impföse in 25 µl TTE überführt und bei 92°C für 10 min lysiert. Die Zellreste wurden mit der Tischzentrifuge bei 14 000 UpM, 4°C für 10 min abzentrifugiert. Vom Überstand wurden 3 µl als Template für eine 50 µl PCR-Reaktion eingesetzt. Das Bakterienlysat konnte für spätere PCR-Reaktionen bei -20°C gelagert werden.

TTE-Puffer: 0,1 ‰ Triton X-100
 20 mM Tris
 2 mM EDTA

2. Aufreinigung von DNA

2.1. Phenolextraktion

Um kurze DNA-Fragmente mit Längen unter 10-20 kb von Enzymen und anderen Proteinen zu reinigen wurden Phenolextraktionen durchgeführt. Dazu wurde der DNA-Lösung 1 Volumen Phenol zugegeben, gevortext und das Gemisch anschließend in der Tischzentrifuge bei 14 000 UpM für 2 min zentrifugiert. Es bilden sich zwei Phasen aus, wobei sich die Proteine im Bereich der Phasengrenze ansammeln. Die obere, wässrige Phase wurde abpipettiert und in ein neues Reaktionsgefäß überführt. Es wurde 1 Volumen PCI (Phenol-Chloroform-Isoamylalkohol, 25:24:1) zugegeben und der Vorgang wiederholt. Um verbliebene Phenolreste zu entfernen wurde abschließend 1 Volumen Chloroform zugegeben, gevortext, erneut zentrifugiert und der Überstand in ein neues Reaktionsgefäß überführt. Anschließend wurde die DNA durch eine Ethanol-fällung weiter gereinigt und aufkonzentriert.

2.2. Ethanolfällung

Zur Reinigung und Aufkonzentrierung von DNA-Lösungen wurden Ethanolfällungen durchgeführt. Dazu wurde die Salzkonzentration der Lösung mit NaCl (5 M) auf 200-500 mmol/l eingestellt. Anschließend wurden 2 Volumen Ethanol 100% (-20°C) zugegeben und die Lösung 15 min in Trockeneis eingefroren. Nach Zentrifugation in der Tischzentrifuge (14 000 UpM, 4°C, 15 min) wurde der Überstand verworfen und das DNA-Pellet bei Raumtemperatur getrocknet. Die DNA wurde in TE-Puffer gelöst und ein Aliquot für eine semiquantitative Mengen-Bestimmung auf einem Agarosegel analysiert.

2.3. Aufreinigung von PCR-Produkten mittels Mikrozentrifugiersäule

Um die PCR-Produkte von den Enzymen, Primern, Salzen und Nukleotiden des Reaktionsansatzes zu reinigen wurde das QIAquick PCR Purification Kit verwendet. Dabei wird die DNA an die Membran einer Mikrozentrifugiersäule gebunden. In einem Waschschrift werden die unerwünschten Bestandteile des PCR-Ansatzes herausgewaschen und verworfen. Anschließend wird die DNA mit EB- oder TE-Puffer wieder von der Membran eluiert. Die Aufreinigung der PCR-Produkte erfolgte nach Anleitung des Herstellers unter Verwendung einer Mikrozentrifuge.

2.4. Gelextraktion mittels Mikrozentrifugiersäule

DNA-Fragmente mit einer Länge bis 5 kb wurden nach dem Ausschneiden der gewünschten Bande aus einem präparativen Agarose-Gel mittels QIAquick Gel Extraction Kit aufgereinigt. Die Prozedur wurde nach dem Protokoll des Herstellers unter Verwendung einer Mikrozentrifuge durchgeführt. Überstieg das Volumen des gelösten Gelstückes die Kapazität der Säule, wurde diese mehrfach beladen und abzentrifugiert. Vor der Elution der DNA wurde die Säule einige Minuten bei Raumtemperatur stehen gelassen, damit letzte Ethanolreste auf der Membran verdunsten. Die Elution der DNA wurde zweifach mit 20 µl Puffer EB (vom Hersteller) und 20 µl TE durchgeführt und die Eluate vereinigt.

2.5. Elektroelution von DNA aus Agarosegelen

Lange DNA kann nach Auftrennung in einem präparativen Agarosegel und Ausschneiden der gewünschten Bande aufgrund ihrer Größe nicht mittels einer Mikrozentrifugiersäule aus dem Gel extrahiert werden. Daher wurden Elektroelutionen in 0,5x TAE-Puffer durchgeführt. Das Gelstück wurde in eine von zwei Kammern eines Eluiergefäßes gelegt, die andere Kammer

mit einer für DNA undurchlässigen Dialysemembran verschlossen. Das Eluiergefäß wurde bis über die Kammern mit Puffer gefüllt. Durch Anlegen einer Spannung von 35 V ü.N. wandert die DNA aus der Agarose in Richtung Anode und sammelt sich an der Membran der zweiten Kammer. Durch Umpolen der Spannung für einige Sekunden wurde die DNA von der Membran gelöst und anschließend der Puffer aus dem Eluiergefäß bis auf ein geringes Restvolumen abgesaugt. Die verbleibenden circa 250 µl DNA-Lösung wurden mit einer abgeschnittenen Pipettenspitze in ein Reaktionsgefäß überführt und bei 14 000 UpM und 4°C für 15 min zentrifugiert um Staub und Bakterien zu sedimentieren. Die Lagerung der DNA erfolgte bei 4°C.

TAE(e)-Puffer:	40 mM Tris
	20 mM Eisessig
	1 mM EDTA
	(0,1 mM EGTA)

3. Restriktionsenzymatische DNA-Spaltung

3.1. Spaltung von DNA in wässriger Lösung

Zur restriktionsenzymatischen Spaltung von DNA in wässriger Lösung wurden die den Enzymen zugeordneten 10x Puffer der Firmen Roche beziehungsweise NEBioLabs sowie die vom Hersteller empfohlenen Temperaturen verwendet. Das Reaktionsvolumen, die eingesetzte Enzymmenge sowie die Inkubationszeit variierten je nach Verwendungszweck der DNA und Aktivität des Restriktionsenzym. So spalten die Enzyme *Bam*-HI oder *Dra*-III sehr schnell, verlieren bei 37°C jedoch auch rasch ihre Aktivität (NEBiolabs Katalog 2002/03). Insbesondere *Dra*-III führt bei längerer Inkubation zu einem Überverdau mit unspezifischer Spaltung der DNA, so dass hier die Inkubationszeit kurz gehalten werden sollte. Für quantitative DNA-Spaltungen wurde jedoch teilweise auch über Nacht verdaut, oder dem Reaktionsansatz nach einiger Zeit neues Enzym zugegeben.

Beim Doppelverdau mit Enzymen, welche unterschiedliche Puffer erfordern, wurde zunächst die DNA-Spaltung mit dem Puffer niedrigerer Salzkonzentration durchgeführt, anschließend das Reaktionsvolumen erhöht und die Salzkonzentration für den Verdau mit dem zweiten Enzym angepasst.

3.2. Spaltung von DNA in Agarose-Blöckchen

Testspaltungen sowie quantitative Spaltungen der PAC-DNA wurden in den Agarose-Blöckchen durchgeführt. Für Längenbestimmungen wurden dabei $\frac{1}{6}$ - $\frac{1}{8}$ eines Blöckchens

eingesetzt, für quantitative Spaltungen die gesamten Blöckchen. Die gereinigten Blöckchen wurden für mehrere Stunden auf Eis in 1x Restriktionspuffer des Herstellers inkubiert. Der Puffer wurde mehrmals gewechselt, um die Ionenkonzentration in den Blöckchen zu äquilibrieren. Anschließend wurden die Blöckchen mit der gerade notwendigen Menge 1x-Puffer überschichtet und das Restriktionsenzym dazu gegeben. Der Reaktionsansatz wurde über Nacht auf Eis inkubiert, damit das Enzym in die Agarose-Blöckchen hinein diffundieren kann. Dabei ist zu beachten, dass einige Enzyme wie z.B. *Eco*-RI sehr schlecht in die Agarose hinein diffundieren (s. NEBiolabs Katalog, 2002/03). Am darauf folgenden Tag wurden die Blöckchen für mehrere Stunden bei der empfohlenen Restriktionstemperatur inkubiert. Wurde die DNA im Anschluss an den Restriktionsverdau sofort mittels einer Pulsfeld-Gelelektrophorese aufgetrennt, wurde die Reaktion lediglich mit EDTA gestoppt. Andernfalls wurde das Restriktionsenzym in NDS mittels Proteinase K verdaut, wonach sich die Blöckchen aufbewahren lassen.

3.3. Partialverdau von PAC-DNA

Um die *Bam*-HI, *Dra*-III und *Hind*-III Restriktionsschnittstellen kartieren zu können, wurde ein Partialverdau der PAC-DNA durchgeführt. Dazu wurden die Agaroseblöckchen zunächst aufgereinigt wie unter 1.2 beschrieben. Danach wurde die PAC-DNA durch quantitativen Verdau mit dem Enzym *Pac*-I linearisiert (s. Kap. 3.2) und mittels einer PFGE aufgetrennt (s. Kap. 4.2). Die linearisierte Vektor-DNA wurde aus dem Gel ausgeschnitten und durch Gelelution aus der Agarose isoliert (s. Kap. 2.5).

Nach Kontrolle der erhaltenen DNA-Menge wurden die Partialverdaus in 20 µl Lösung mit 2 µl 10x Restriktionspuffer bei 37°C durchgeführt. Die Enzymmengen und Restriktionszeiten variierten je nach Aktivität des verwendeten Enzyms (s. Tab. 5).

Tabelle 5: Enzymmengen und Restriktionszeiten für den Partialverdau der α -Satelliten-PACs mit den Enzymen *Bam*-HI, *Dra*-III und *Hind*-III.

<i>Bam</i> -HI und <i>Dra</i> -III		<i>Hind</i> -III	
Enzymmenge [u]	Restriktionsdauer [min]	Enzymmenge [u]	Restriktionsdauer [min]
0	1	0	1
0,1	1	0,4	1
0,4	1	1,6	1
0,8	1	3,2	2
1,6	1	10	5
3	1		

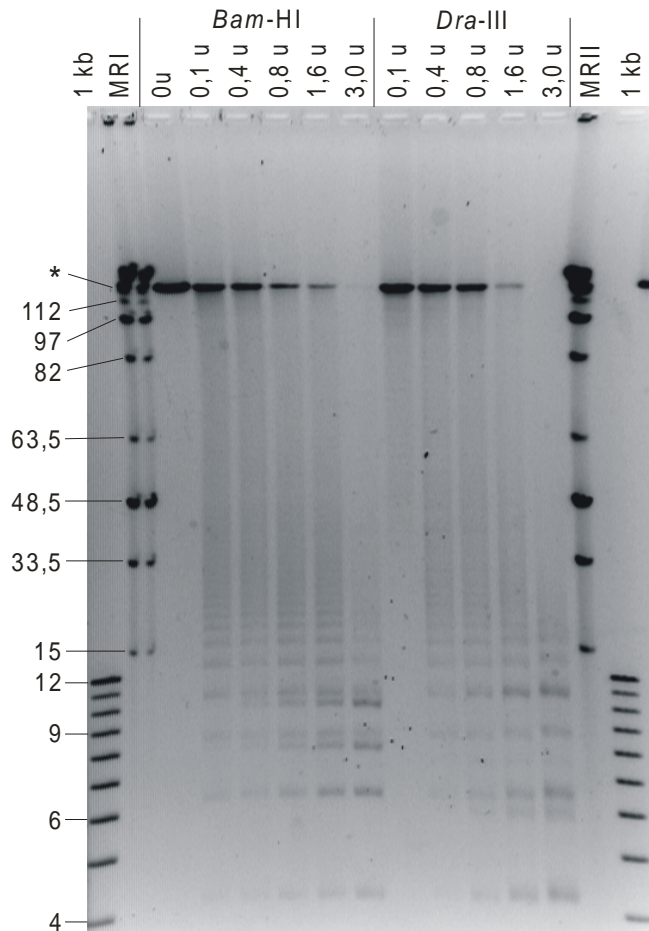


Abb. 23: Pulsfeld-Gelelektrophorese nach Partialverdau des linearisierten PAC B11 mit den Enzymen *Bam*-HI und *Dra*-III. Links sind die Längen der verwendeten DNA-Marker angegeben (1kb: Gibco BRL 1 kb DNA ladder, MRI: NEBioLabs MidRange I PFG Marker). Bei circa 125 kb erkennt man den mit *Pac*-I linearisierten vollständigen PAC-Vektor (Stern). Oberhalb des Gels sind die verwendeten Enzymmengen angegeben, die Restriktionszeit betrug jeweils eine Minute. PFGE-Bedingungen: Feldstärke 6 V/cm, Switchzeit 2,8 sek, Laufzeit 25 h.

Die Reaktionen wurden nach der angegebenen Zeit durch Hinzufügen von 1 μ l EDTA 0,5 M und Abkühlen auf Eis gestoppt. Die DNA wurde mit 5 μ l Auftragspuffer auf ein Pulsfeld-Gel geladen (Laufzeit 25 h, Feldstärke 6 V/cm, Initial und final switch time 2,8 sek, 0,5x TAE-Puffer, Puffertemperatur 12°C). Im Anschluss an die Elektrophorese wurde das Gel mit Ethidiumbromid gefärbt und unter der UV-Lampe ausgewertet. Bereits jetzt war eine Beurteilung des Partialverdaus, der Auftrennung der DNA-Fragmente und der Abwesenheit einer unspezifischen DNA-Degradierung möglich (Abb. 23).

Southern-Blotting auf eine Nylon-Membran und Hybridisierungen mit radioaktiv markierten DNA-Sonden, welche direkt beidseits der *Pac*-I Schnittstelle des PAC-Vektors liegen, ermöglichten im Folgenden die Kartierung der Restriktionsschnittstellen von *Bam*-HI, *Dra*-III und *Hind*-III abhängig von ihrer Entfernung von der *Pac*-I Restriktionsschnittstelle (Abb. 16).

4. Gelelektrophorese

4.1. Auftrennung von DNA-Fragmenten mittels konventioneller

Gelelektrophorese

Zur Auftrennung von DNA-Fragmenten mit Längen bis 20 kb wurden konventionelle Agarose-Gelelektrophoresen durchgeführt. Dabei bewegt sich die DNA in einem elektrischen Feld von 1-5 V/cm durch ein Agarosegel zur Anode, wobei lange Fragmente langsamer wandern.

Für DNA mit einer Länge über 1000 bp wurden 0,8%ige Agarosegele verwendet, für kürzere Fragmente 1-1,2%ige oder höherprozentige Gele. Dazu wurde das Agarosepulver in 1x TBE-Puffer in der Mikrowelle geschmolzen. Nachdem die flüssige Agarose unter Rühren auf circa 50°C abgekühlt war, wurden 2 µl Ethidiumbromid (10 mg/ml) pro 100 ml Gel zugegeben, und das Gel blasenfrei in einen Gelträger mit Kamm gegossen. Das erstarrte Gel wurde in einer Gelkammer mit 1x TBE-Puffer überschichtet und der Kamm gezogen. Da Ethidiumbromid entgegen der Laufrichtung der DNA aus dem Gel heraus wandert, wurde der Puffer im Bereich der Anode je nach Verwendungszweck der Elektrophorese ebenfalls damit versetzt. Die DNA-Proben wurden mit $\frac{1}{5}$ - $\frac{1}{2}$ Volumen Auftragspuffer vermischt und in die Geltaschen geladen. Bei jedem Gellauf wurde ein geeigneter Längenstandard mit aufgetragen. Durch das in die Doppelhelix interkalierende Ethidiumbromid konnten die DNA-Banden unter ultraviolettem Licht sichtbar gemacht und zu Dokumentationszwecken photographiert werden. Der Vergleich der Intensitäten der DNA-Banden mit den bekannten Markerbanden ermöglichte außerdem eine semiquantitative Abschätzung der aufgetragenen DNA-Mengen.

TBE-Puffer:	90 mM Tris 90 mM Borsäure 2 mM EDTA
Auftragspuffer orange:	25% Ficoll in TE-Puffer Orange G
Auftragspuffer blau:	Auftragspuffer orange Xylencyanol Bromphenolblau

4.2. Auftrennung langer DNA-Fragmente mittels Pulsfeld-Gelelektrophorese

(PFGE)

Die elektrophoretische Auftrennung von DNA mit Längen über 20 kb erfordert die Durchführung einer Pulsfeld-Gelelektrophorese (PFGE). Entgegen der konventionellen Gelelektrophorese, bei welcher Nukleinsäuren von über 20 kb Länge in einer einzigen Bande

laufen, ermöglicht sie die Auftrennung von DNA-Fragmenten mit Längen bis zu mehreren Megabasen. Dabei ändert das elektrische Feld in bestimmten Zeitintervallen die Richtung, wobei die Wanderungsgeschwindigkeit unterschiedlich langer DNA-Fragmente unter anderem abhängig von der Frequenz der Richtungsänderung ist (Chu et al. 1986).

Die Pulsfeld-Gelelektrophoresen wurden in einem 1%igen Standard-Agarosegel in 0,5x TAE-Puffer durchgeführt. Damit die DNA-Agaroseblöckchen nicht aus den Geltaschen heraus schwimmen wurden sie mit 1%iger LMP-Agarose (Gibco BRL) in 0,5x TAE-Puffer im Gel eingegossen. DNA in wässriger Lösung wurde wie bei einer konventionellen Gelelektrophorese mit $\frac{1}{5}$ Volumen Auftragspuffer aufgetragen. Die Gelelektrophoresen wurden in 0,5x TAE-Puffer bei 12°C und einer Feldstärke von 6 V/cm durchgeführt. Switch-Zeit und Laufzeit variierten je nach Länge der erwarteten DNA-Banden. Für die Auswertung und photographische Dokumentation wurde das Gel nach Abschluss der Elektrophorese in einer Wanne mit Ethidiumbromid (10 mg/ml, 2 µl/100 ml TAE-Puffer) gefärbt.

Die Herstellung von Agarosegelen mit einer Dicke von circa 15 mm ermöglichte auch das Laden kompletter DNA-Blöckchen zur Aufreinigung und quantitativen Auftrennung langer DNA. Dazu wurden mehrere Agarose-Blöckchen quer in eine große Geltasche gelegt und mit flüssiger LMP-Agarose eingegossen. Zu beiden Seiten der Blöckchen wurde ein MidRange-DNA-Marker aufgetragen. Nach dem Gellauf wurden beidseits die Markerbahnen sowie die Ränder der DNA-Bahnen abgeschnitten und mit Ethidiumbromid gefärbt. Unter UV-Licht konnte so die gewünschte DNA-Bande identifiziert und markiert werden. Anhand der Markierung konnte dann die DNA aus dem ungefärbten Teil des Gels ausgeschnitten werden. Somit wurden Brüche in der PAC-DNA durch die Ethidiumbromid-Färbung und UV-Bestrahlung vermieden. Anschließend wurde die DNA mittels Elektroelution aus der Agarose isoliert.

5. DNA-Klonierung

5.1. Dephosphorylierung von DNA-Enden

Um bei DNA-Klonierungen die Religierung des linearisierten Vektors und somit eine Vielzahl von Antibiotikum-resistenten Bakterienkolonien ohne Insert im Vektor zu verhindern, wurde die Vektor-DNA dephosphoryliert. Dazu wurden zu einem 200 µl-Ansatz aus DNA-Lösung und CIP-Puffer 2 u CIP (Calf Intestine Phosphatase) gegeben und zunächst bei 37°C für 15 min, anschließend bei 50°C für 15 min inkubiert. Danach wurden nochmals 2 u CIP hinzugefügt und für jeweils 15 min bei 37°C und 50°C inkubiert. Die Phosphatase

wurde durch Zugabe von 1/10 Volumen STE-Puffer, 1/40 Volumen 20%igem SDS und 15 minütiger Inkubation bei 68°C inaktiviert. Daran schlossen sich eine Phenolextraktion sowie eine Ethanolfällung an.

CIP-Puffer: 50 mM Tris, pH 9,0
 1 mM MgCl₂
 0,1 mM ZnCl₂
 1 mM Spermidin

STE-Puffer: 0,1 M NaCl
 1 mM Tris, pH 8,3
 1 mM EDTA

5.2. Ligierung

Die Ligierung der Insert-DNA in den linearisierten und dephosphorylierten Vektor erfolgte mit T4 DNA-Ligase. In einem 20 µl Ansatz wurden Insert und Vektor im molaren Verhältnis von 3-5:1 gemischt und 2 µl 10x T4 DNA-Ligase-Puffer hinzugefügt. $\frac{1}{3}$ dieses Ansatzes wurde auf ein Agarosegel aufgetragen, zu den verbleibenden $\frac{2}{3}$ wurden 5 u T4 DNA-Ligase gegeben. Die Ligierung wurde bei 12°C über Nacht durchgeführt. Anschließend wurde der Ansatz für 15 min bei 37°C inkubiert um verbleibende kohäsive Enden zu vereinigen. Vor der Transformation in *E. coli* wurde die Ligase bei 68°C für 20 min inaktiviert. Die Hälfte der DNA wurde zum Vergleich mit dem unligierten Ansatz auf ein Agarosegel aufgetragen, die andere Hälfte in die Elektroporation eingesetzt.

5.3. Herstellung elektrokompenter Bakterien

Zur Herstellung von elektrokompenten Bakterien wurden 20 µl einer Glycerinkultur des *E. coli* Stammes DH10B auf einer Agarplatte ausgestrichen und nach 24 h Inkubation bei 37°C 100 ml LB₀ mit einer Kolonie beimpft. Nach Vermehrung der Bakterien bei 37°C ü.N. wurden 2 l LB₀ hinzu gegeben und für weitere vier Stunden im Schüttler inkubiert, so dass die Bakterien sich vor den folgenden Arbeitsschritten in der exponentiellen Wachstumsphase befanden. Die Zellen wurden abzentrifugiert (Hettich Rotixa/RP Zentrifuge, 2500 UpM, 4°C, 15 min) und das Medium abgeschüttet. Das resultierende Pellet wurde drei mal in je 50 ml destilliertem Wasser gewaschen (Beckman-Zentrifuge, 4000 UpM, 4°C, 10 min). In einem vierten Waschschriff wurden die Zellen in 50 ml 10%igem Glycerin resuspendiert und abzentrifugiert. Das entstandene Pellet wurde in 500 µl 10%igem Glycerin gelöst, in 50 µl-Portionen verteilt und bis zur Verwendung bei -70°C aufbewahrt. Alle Arbeitsschritte wurden

auf Eis bzw. bei 4°C durchgeführt um eine Lyse der Bakterien im destillierten Wasser zu verhindern.

Teilweise wurden für die Elektroporationen auch fertig aufbereitete elektrokompente Bakterien „ElektroMAX“ der Firma Gibco BRL verwendet.

5.4. Transformation von *E. coli* mittels Elektroporation

Die Transformation der ligierten DNA in den *E. coli* Stamm DH10B erfolgte durch Elektroporation mit dem Gene Pulser II der Firma BioRad. In einem typischen Elektroporationsexperiment wurden 1-1,5 µl des Ligationsansatzes sowie 20 µl der selbst hergestellten elektrokompenten Bakterien beziehungsweise 2 µl fertig aufbereitete Bakterien der Firma Gibco BRL verwendet. Der Ansatz wurde auf Eis in eine Elektroporationsküvette pipettiert und das Gesamtvolumen mit destilliertem Wasser auf 40 µl aufgefüllt. Die Elektroporationen wurden bei einem Elektrodenabstand von 1 mm mit 1,8 kV, 25 µF durchgeführt. Sofort nach dem Elektroporationspuls wurden die Bakterien in 1 ml SOC-Medium aufgenommen und bei 37°C für 30 min inkubiert. Anschließend wurden sie in absteigender Menge auf drei Selektionsagarplatten ausplattiert.

SOC-Medium:	2% Trypton
	0,5% Hefe-Extrakt
	10 mM MgCl ₂
	10 mM MgSO ₄
	2,5 mM KCl
	10 mM NaCl
	20 mM Glucose

5.5. Antibiotische Selektion und blau/weiß Selektion von Klonen mittels X-Gal

Zur Selektion auf Bakterien, welche die transformierten Vektoren und damit die von diesen kodierten Resistenzgene enthalten, wurden die Agarplatten und Kulturmedien mit Antibiotika versetzt. Für die Vektoren pBluescript II SK und pGEM-T wurde Ampicillin (50 mg/ml, 1/500-1/1000) eingesetzt, für den PAC-Vektor Kanamycin (30 mg/ml, 1/1000).

Um bei den Klonierungen in die Vektoren pBluescript II SK und pGEM-T Klone mit leerem Vektor von denjenigen mit Insert unterscheiden zu können, wurde eine Blau/Weiß-Selektion mittels X-Gal durchgeführt. Dazu liegt die Multiple-Cloning-Site der beiden Vektoren innerhalb des für das α-Peptid des Enzyms β-Galaktosidase kodierenden Bereichs. Bei Transformation von ungespaltenem oder leer religiertem Vektor mit intakter Multiple-Cloning-Site führt die Induktion mit IPTG (1 M, 1/20000) somit zur Expression der β-Galaktosidase, was bei Anreicherung der Agarplatten mit X-Gal (40 mg/ml in

Dimethylformamid, 1/1000) eine Blaufärbung der Bakterienkolonien bewirkt. Wurde jedoch ein Insert in die Multiple-Cloning-Site inkloniert, kann die β -Galaktosidase nicht exprimiert werden und die Kolonien bleiben weiß.

5.6. Anlegen von Übernachtkulturen und Aufbewahrung von Klonen als Glycerinkultur

Enthielten die Selektionsagarplatten nach Inkubation bei 37°C ü.N. eine ausreichende Anzahl an Kolonien, wurden wenn möglich PCR-Amplifikationen aus Plasmid-Schnellpräparationen zur Kontrolle der klonierten Inserts durchgeführt. War dies nicht möglich, wurde sofort eine größere Anzahl an Mini-Präparationen durchgeführt, und die korrekten Klone anschließend über Restriktionsspaltungen identifiziert.

Für die Amplifikation der Plasmid-DNA wurden die Kolonien mit einer sterilen Impföse in je 7 ml LB-Selektionsmedium überführt. Dieses wurde im Schüttler bei 280 UpM und 37°C ü.N. inkubiert. Am darauf folgenden Tag wurde 1 ml des Mediums in einem 1,5 ml Reaktionsgefäß abzentrifugiert (Tischzentrifuge, 4000 UpM, 4°C, 1 min). Der Überstand wurde abgekippt, das Bakterienpellet in den verbleibenden circa 150 μ l Medium gelöst und mit 150 μ l 86%igem Glycerin vermischt. Die Glycerinkulturen wurden für spätere DNA-Präparationen bei -70°C gelagert. Die verbleibenden 6 ml Medium wurden ebenfalls abzentrifugiert (Schwenkbecherzentrifuge Heraeus, 3500 UpM, 10 min) und das Bakterienpellet in Mini-Präparationen eingesetzt.

6. Verdau überhängender 5'-Enden mittels S1-Endonuklease

Um die *Dra*-III Restriktionsschnittstelle im fl-ori des Vektors pBluescript II SK zu entfernen, wurde der Vektor mit *Dra*-III verdaut und die überhängenden 5'-Enden mittels S1-Endonuklease entfernt, welche spezifisch einzelsträngige DNA und RNA in 5'-Mononukleotide degradiert. Anschließend wurde der Vektor unter Verlust der drei mittleren Basen der Erkennungssequenz von *Dra*-III blunt religiert und in *E. coli* transformiert.

Der pBluescript II SK-Vektor wurde zunächst in einem Gesamtvolumen von 100 μ l mit *Dra*-III verdaut. Anschließend wurden 50 μ l 10x S1-Nuklease-Puffer des Herstellers hinzugefügt und das Gesamtvolumen auf 500 μ l erhöht. Dazu wurden 150 u S1-Nuklease gegeben und bei 37°C für 5 min inkubiert. Die Reaktion wurde durch Zugabe von 10 μ l EDTA (500 mM) gestoppt und die DNA mittels Phenolextraktion und Ethanol-fällung von den

Enzymen gereinigt. Nach dem Lösen des erhaltenen Pellets in 20 µl TE wurde die DNA mittels T4 DNA-Ligase religiert. Anschließend wurde das Volumen wieder auf 200 µl erhöht und ein erneuter *Dra*-III Verdau durchgeführt. Dadurch wurden Vektor-Moleküle, welche die *Dra*-III Schnittstelle noch enthielten, linearisiert und somit aus der Klonierung ausgeschlossen. Nach erneuter Ethanol-fällung des Restriktionsansatzes wurde die DNA in 0,25x TAE gelöst und in *E. coli* elektroporiert. Von den resultierenden Kolonien wurden Mini-Präparationen angefertigt und die Plasmide durch Verdau mit *Dra*-III auf das Fehlen der Restriktionsschnittstelle kontrolliert.

7. Einklonierung eines doppelsträngigen Oligonukleotids

Um in die Multiple-Cloning-Site des so entstandenen Vektors BsΔD eine *Dra*-III Schnittstelle mit einem 5'-TGG Überhang einzufügen, wurde ein doppelsträngiges Oligonukleotid einkloniert, welches durch Hybridisierung zweier einzelsträngiger synthetischer Oligonukleotide erzeugt wurde.

Dazu wurde zunächst der Vektor BsΔD durch Doppelverdau mit den Restriktionsenzymen *Bam*-HI und *Hind*-III linearisiert, wobei zueinander inkompatible 5'-Überhänge entstanden. Die DNA wurde durch Auftrennung auf einem Agarosegel, Ausschneiden der entsprechenden Bande aus dem Gel und Elution über eine Mikrozentrifugiersäule von den Enzymen und den herausgeschnittenen Oligonukleotiden gereinigt. Die Sequenzen der ss-Oligonukleotide wurden so gewählt, dass nach der Hybridisierung zu einem ds-Oligonukleotid komplementäre *Hind*-III und *Bam*-HI Überhänge resultierten. Außerdem war die Länge des ds-Oligonukleotids identisch mit der Länge des durch den Doppelverdau aus dem Vektor herausgeschnittenen DNA-Fragments. Dadurch ist die *in frame* Translation des β-Galaktosidase-Gens, und somit eine blau/weiß-Selektion von Kolonien nach Klonierungen in die neu entstandene *Dra*-III Schnittstelle, weiterhin möglich. Je 1,6 nm der ss-Oligonukleotide wurden zu einem 10 µl Reaktionsansatz aus linearisierter BsΔD-DNA, 1x Ligase-Puffer und 5 u T4 DNA-Ligase gegeben und bei 12°C über Nacht inkubiert. Anschließend wurde das Volumen auf 50 µl erhöht und erneut ein Doppelverdau mit *Hind*-III und *Bam*-HI durchgeführt. Da durch die Insertion des ds-Oligonukleotids die entsprechenden Restriktionsschnittstellen zerstört wurden, konnte so die Klonierung von ungespaltenem BsΔD-Vektor vermieden werden. Die DNA wurde ethanolgefällt, in 20 µl TE gelöst und in *E. coli* transformiert. Von 16 der erhaltenen blauen Kolonien wurden Mini-Präparationen

angefertigt und der Vektor durch Verdau mit *Dra*-III auf das Vorhandensein des korrekten Inserts überprüft.

8. Polymerasekettenreaktion (PCR)

Zur spezifischen Amplifikation von DNA-Abschnitten wurden Polymerasekettenreaktionen nach Saiki (1988) durchgeführt. Damit können Nukleinsäuren von bis zu einigen Kilobasen Länge exponentiell vervielfältigt werden, vorausgesetzt die den Abschnitt flankierenden Sequenzen sind bekannt. Die Reaktion findet in 25-30 Zyklen statt, welche jeweils aus drei Schritten bestehen:

Im ersten Schritt wird die DNA durch 60 sekündiges Erhitzen auf 94°C denaturiert, das heißt die DNA-Doppelstränge werden in Einzelstränge aufgetrennt. Vor dem ersten Zyklus wird die Denaturierung fünf Minuten lang durchgeführt, um die gesamte Template-DNA in Einzelstränge zu überführen.

Im zweiten Schritt wird die Temperatur für 60 Sekunden abgesenkt, so dass zwei Oligonukleotidprimer spezifisch an ihre komplementären Sequenzen in der einzelsträngigen DNA binden können. Die Annealingtemperatur kann abhängig von den Primern durch folgende Formel berechnet werden:

$$T_A = [(N_A + N_T) \times 2^\circ C + (N_G + N_C) \times 4^\circ C] - 4^\circ C$$

T_A : Annealingtemperatur [°C]

N_A, N_T, N_G, N_C : Anzahl der Basen Adenin, Thymin, Guanin bzw. Cytosin im Primer

Ausschlaggebend ist die niedrigere der für beide Primer getrennt ermittelten Temperaturen.

Im dritten Schritt erfolgt bei 72°C die von den Primern ausgehende Extension der DNA-Stränge in 5' -> 3' Richtung. Dabei wird durch die hitzestabile Polymerase des Bakteriums *Thermus aquaticus* ein zur einzelsträngigen Matrize komplementärer DNA-Strang synthetisiert. Die Extensionszeit beträgt eine Minute pro 500-1000 bp. Im letzten PCR-Zyklus wird die Extensionszeit auf 5 min verlängert, damit die verbleibende einzelsträngige DNA vervollständigt werden kann.

Ab dem zweiten PCR-Zyklus können sich die Primer auch an die neu synthetisierten DNA-Stränge anlagern, wobei die Extension der DNA in 3'-Richtung dann nur noch bis zum reversen Primer erfolgt. Dadurch kommt es zu einer exponentiellen Amplifikation des durch die beiden Primer begrenzten DNA-Abschnittes.

Tabelle 6: Die in dieser Arbeit durchgeführten Polymerasekettenreaktionen mit den Längen der amplifizierten DNA-Fragmente sowie den optimalen Annealingtemperaturen.

Vorwärts-Primer	Rückwärts-Primer	Lokus	Länge des PCR-Produkts	Annealing-Temperatur
X-3A	X-4A	Humane α -Satelliten-DNA	535 bp	52°C
X-3A	X-4Ag	Gorilla α -Satelliten-DNA	520 bp	50°C
SP6	T7	Längenbestimmung des α -Satelliten-Inserts im Klonierungsvektor pGEM-T	643-658 bp	48°C
VP1F	VP1R	Synthese der Hybridisierungsprobe VP1 auf dem PAC-Vektor pCYPAC2	1262 bp	62°C
VP2F	VP2R	Synthese der Hybridisierungsprobe VP2 auf dem PAC-Vektor pCYPAC2	1019 bp	58°C

PCR-Ansatz:

- 5 μ l 10x PCR-Puffer
- 4 μ l dNTP-Mix (je 160 μ M dATP, dTTP, dGTP, dCTP)
- je 1 μ l vorwärts- und rückwärts-Primer (20 pM)
- 0,3 μ l *Taq*-Polymerase (15 u)
- Matrizen-DNA
- H₂O bidest ad 50 μ l

10x PCR-Puffer:

- 500 mM KCl
- 100 mM Tris, pH 8,3
- 15 mM MgCl₂
- 0,1 % Gelatine

9. DNA-Sequenzierung

Die DNA-Sequenzierungen erfolgten nach der Kettenabbruchmethode von Sanger et al. (1977). Dabei werden 25 Zyklen einer Reaktion aus vier Schritten durchgeführt: Im ersten Schritt wird die Template-DNA durch 45 sekündiges Erhitzen auf 94°C in ihre Einzelstränge denaturiert. Danach wird der Ansatz für 30 Sekunden auf 50°C abgekühlt, damit ein Primer sich spezifisch an das einzelsträngige DNA-Template anlagern kann. Im dritten Schritt erfolgt die Extension der DNA durch eine thermostabile Polymerase bei 60°C. Der zufällige Einbau eines Didesoxynukleotids führt im vierten Schritt zum Abbruch der DNA-Synthese. Die vier Didesoxynukleotide ddATP, ddTTP, ddGTP und ddCTP sind dabei mit unterschiedlichen Fluoreszenzfarbstoffen markiert, wodurch sich die zuletzt eingebaute Base identifizieren lässt. Durch Auftrennung der DNA-Einzelstränge in einem vertikalen, denaturierenden Polyacrylamidgel kann dann anhand der Längen der fluoreszenzmarkierten DNA-Fragmente die Basenabfolge bestimmt werden.

Die Sequenzreaktionen wurden mit dem BigDye Terminator Cycle Sequencing Ready Reaction Kit von Perkin Elmer durchgeführt. Der BigDye Terminatoren-Mix enthält die DNA-Polymerase, einen Puffer sowie die Desoxy- und Didesoxynukleotide. Das Reaktionsprodukt der 25 Synthese-Zyklen wurde in einem Volumen Ethanol 100% gefällt,

abzentrifugiert (Tischzentrifuge, 14 000 UpM, 4°C, 10 min), der Überstand verworfen und das DNA-Pellet bei Raumtemperatur getrocknet. Die Elektrophoresen der DNA-Fragmente mit dem ABI PRISMTM-System 377 von Perkin Elmer erfolgten am Institut für Biochemie der LMU München beziehungsweise durch die Firma Medigenomix, München.

Sequenzieransatz: 0,2 µl Primer (0,4 pM)
 2 µl BigDye Terminatoren-Mix
 Template-DNA
 H₂O bidest ad 10 µl

DNA-Fällung: 10 µl Sequenzieransatz
 90 µl H₂O bidest
 10 µl NaCl 5 M
 220 µl Ethanol 100%

9.1. Korrekturlesen der DNA-Sequenzen

Die von dem ABI PRISMTM-System automatisch aus den Elektropherogrammen erzeugten Sequenzen wurden überprüft und die Längen der auswertbaren Sequenzabschnitte bestimmt. Bei schlechter Auswertbarkeit der Elektropherogramme wurden die entsprechenden Klone unter Verwendung der Primer T3 / T7 oder der α -Satelliten-Primer X-3A / X-4A resequenziert. Die Sequenzen wurden aligned und eine Konsensussequenz über alle Higher order repeats erstellt. Anschließend wurden alle Abweichungen von der Konsensussequenz erneut überprüft und nicht eindeutige Sequenzabschnitte von den weiteren Analysen ausgeschlossen.

10. Blotting von DNA auf Nylonmembranen

Für die spezifische Hybridisierung von DNA-Fragmenten nach ihrer Auftrennung in einem Agarosegel wurde die DNA im Southern-Blotting Verfahren auf eine Nylonmembran übertragen (Southern 1975). Dazu wurde das Ethidiumbromid-gefärbte Agarosegel zunächst unter UV-Licht mit einem Lineal als Maßstab photographiert um den Abstand der Markerbanden von den Geltaschen bestimmen zu können. Um den Transfer der DNA zu erleichtern wurde das Gel anschließend 2x 15 min in 0,25 M HCl geschwenkt, danach 2x 15 min in Denaturierungspuffer. Der Blot erfolgte mit 0,4 N NaOH, welche mittels Filterpapier durch Schlitze in einer Kunststoffplatte aus einer Blotwanne gesaugt wurde. Unter das zurecht geschnittene Gel wurde ein 3 mm dickes Filterpapier gelegt, auf das Gel die mit NaOH befeuchtete Nylonmembran. Darauf kamen ein in NaOH eingeweichtes DAE-Filterpapier, ein trockenes DAE-Filterpapier sowie drei dicke Filterpapiere. Um die NaOH aus der Blotwanne durch das Gel, die Nylonmembran und die Filterpapiere zu saugen, wurde

Zellstoff auf den Stapel gelegt und mit einem Gewicht beschwert. Die negativ geladene DNA wurde ü.N. auf die positiv geladene Nylonmembran transferiert und dort gebunden. Nach dem Blotten wurde die Membran zum pH-Ausgleich für 10 min in Post-Blot-Waschlösung geschwenkt, anschließend in 2x SSC-Lösung gewaschen und 2 h bei 60-70°C getrocknet.

Denaturierungspuffer:	0,5 M NaOH 1,5 M NaCl
SSC-Lösung:	150 mM NaCl 15 mM Na ₃ Citrat
Post-Blot-Waschlösung:	2x SSC 200 mM Tris, pH 7,4

11. Radioaktive Markierung von DNA-Proben

Für die Hybridisierungen der Southern-Blots wurden DNA-Fragmente durch Einbau des Nukleotids [$\alpha^{32}\text{P}$]-dCTP radioaktiv markiert. Dazu wurden circa 100 ng der PCR-Produkte VP1 und VP2 beziehungsweise der Probe *pBamX5* in einem Gesamtvolumen von 36 μl bei 95°C für 5 min denaturiert und anschließend sofort auf 0°C abgekühlt, um eine Reassoziationsreaktion der einzelsträngigen DNA zu verhindern. Danach wurden 2 μl BSA, 0,5 μl Klenow-Polymerase, 2-3 μl [$\alpha^{32}\text{P}$]-dCTP sowie 10 μl Oligo-labeling-Puffer, welcher für dCTP depletiert ist und Hexanukleotide als Zufallsprimer enthält, hinzugefügt. Die Inkubation dieses Ansatzes bei Raumtemperatur führt, ausgehend von den Zufallsprimern, zur Neusynthese von komplementären DNA-Strängen durch die Klenow-Polymerase mit Einbau des radioaktiven [$\alpha^{32}\text{P}$]-dCTP. Die Reaktion wurde nach circa 30 min durch 150 μl TES-Puffer gestoppt.

Um die überschüssigen radioaktiven Nukleotide aus der Probe zu entfernen, wurde eine Sephadex-G50 Säule hergestellt. Dazu wurde eine 1 ml-Spritze mit autoklavierter Glaswolle nach unten hin abgedichtet und mit in TE-Puffer aufgekochtem Sephadex-G50 aufgefüllt. Die Säule wurde für 10 min abzentrifugiert (Schwenkbecherzentrifuge Heraeus, 2000 UpM) um überschüssigen TE-Puffer zu entfernen. Anschließend wurde die Probe auf die Säule pipettiert und erneut für 10 min (Tischzentrifuge Universal 2S) zentrifugiert. Die überschüssigen radioaktiven Nukleotide bleiben dabei in der Säule zurück, während die gereinigte, radioaktiv markierte Probe als Durchfluss aufgefangen werden kann.

Oligo-labeling-Puffer (OLB-C):	Lösung 0:	1,25 M Tris, pH 8,0 125 mM MgCl ₂
	Lösung A:	1 ml Lösung 0 18 µl β-Mercaptoethanol je 5 µl 100 mM dATP, dGTP, dTTP
	Lösung B:	2 M Hepes pH 6,6 (4 M NaOH)
	Lösung C:	5'pd(N)6-Oligonukleotide (Pharmacia) TE-Puffer
	Die Lösungen A, B und C wurden im Verhältnis 1:2,5:1,5 gemischt	
TES-Puffer:	20 mM NaCl 20 mM Tris, pH 7,5 2 mM EDTA 0,25% SDS 0,25% Bromphenolblau	

12. Hybridisierung

Vor den Hybridisierungen wurden die Nylonmembranen 2x für je 15 min in 2x SSC geschwenkt. Die Filter wurden aufgerollt und in einem verschließbaren Glaskolben mit 10 ml Prähybridisierungslösung bei 65°C unter ständigem Rotieren für 2 h prähybridisiert.

Anschließend wurde die Prähybridisierungslösung durch 10 ml Hybridisierungslösung ersetzt, welcher 1 mg thermisch denaturierter Lachsspermien-DNA hinzugefügt wurde. Durch das kompetitive Binden dieser DNA und der radioaktiv markierten Probe wird die unspezifische Bindung der DNA-Sonde an einfache Repeatsequenzen vermindert. Die von den überschüssigen Nukleotiden gereinigte Probe wurde durch Erhitzen auf 95°C in ihre Einzelstränge denaturiert und in den Hybridisierungskolben gegeben.

Nach der Hybridisierung bei 65°C ü.N. wurden die Filter zunächst kurz mit 2x SSC gespült. Danach wurden die Membranen in Waschlösungen mit abnehmender Salzkonzentration bei steigender Temperatur gewaschen, bis sie eine nur mehr geringe, aber noch messbare Radioaktivität aufwiesen (z.B. Lösung 1 bei Raumtemperatur, Lösung 2 bei 50°C, Lösung 3 bei 68°C). Die Filter wurden in Klarsichtfolie verpackt und in eine Filmkassette eingelegt. Für eine optimale Bandenschärfe erfolgte die autoradiographische Darstellung der α-Satelliten-Leitern größtenteils ohne Verstärkerfolien bei -70°C. Die Expositionszeit betrug in diesen Fällen bis zu mehrere Wochen, bei Verwendung von Verstärkerfolien einige Stunden bis Tage. Bei mehrmaliger Verwendung eines Filters mit unterschiedlichen Proben wurde zwischen den Hybridisierungen gewartet, bis keine Radioaktivität mehr nachweisbar war.

Für die Photodokumentation wurden die entwickelten Filme auf einer Leuchtplatte abphotografiert und die Diapositive mit einem Nikon CoolScan III Diascanner eingelesen.

Prähybridisierungslösung:	1% SDS 0,5% Magermilchpulver 1,5x SSPE
20x SSPE-Lösung:	3,6 M NaCl 0,2 M Na ₂ HPO ₄ 20 mM EDTA, pH 7,4
Hybridisierungslösung:	10% Dextransulfat in Prähybridisierungslösung
Waschlösungen:	Lösung 1: 2x SSC 0,1% SDS Lösung 2: 0,5x SSC 0,1% SDS Lösung 3: 0,1x SSC 0,1% SDS

D. Material

1. DNA

1.1. Humane α -Satelliten-PACs

Die untersuchten α -Satelliten-PACs (P1-derived artificial chromosomes) stammen aus Segment I der humanen PAC-Bank RPCI-I (Ioannou et al. 1994, s. Internet-Links) und wurden über das UK HGMP Resource Center, Hinxtonhall, bezogen. Die Klonnamen wurden freundlicherweise von Charlotte Cole und Dr. Mark Ross konvertiert.

Labor-interne Bezeichnung	Klonkoordinaten
A6	N16/2
A7	P9/3
A8	H19/1
A10	O12/2
B11	E8/5

1.2. Humane genomische DNA

Für die Analyse der Fixierung der Varianten wurde genomische DNA von 10 nichtverwandten, männlichen Individuen verwendet. Die Proben entstammen der DNA-Bank des Humangenetischen Instituts der Universität München. Die Spender mit größtenteils ausländischen Familiennamen, die nicht für das bayerische Kern-Einzugsgebiet der DNA-Bank typisch sind, waren gesunde Angehörige von Patienten, welche an unterschiedlichen, genetisch bedingten Erkrankungen litten.

1.3. *Gorilla gorilla* genomische DNA

Die Proben genomischer DNA dreier männlicher Flachlandgorillas wurden freundlicherweise von Christian Roos vom deutschen Primatenzentrum, Göttingen zur Verfügung gestellt.

Nummer	Gene Bank of Primates Kürzel	Herkunft
014	GGOR002	München TP
165	GGOR010	Basel ZG
167	GGOR012	Basel ZG

1.4. DNA-Sequenzen aus Datenbanken

Name	Acc. Number	Lokus	Länge
HSSATAX	X02418	<i>Homo sapiens</i> DXZ1	2055 bp
GGSATG	X56887	<i>Gorilla gorilla</i> α -Satelliten-DNA	2036 bp

2. Vektoren

Name	Hersteller	Acc. Number	Länge	Sonstiges
pBluescript II SK	Stratagene	X52328	2961 bp	Ampicillin-Resistenz, Blau/Weiß-Selektion
pGEM-T	Promega	X65308	3000 bp	Ampicillin-Resistenz, Blau/Weiß-Selektion
pCYPAC2		U09128	16011 bp	Kanamycin-Resistenz

3. Elektrokompente Bakterien

ElectroMAX DH10B (*E. coli*)

Gibco BRL

4. Kits

QIAprep Spin Miniprep Kit

Qiagen

QIAquick PCR Purification Kit

Qiagen

QIAquick Gel Extraction Kit

Qiagen

BigDye Terminator Cycle Sequencing Ready Reaction Kit

Perkin Elmer

5. Oligonukleotide

Oligonukleotide wurden im Genzentrum München, Abteilung Dr. Arnold synthetisiert. Für die Sequenzierungs- und PCR-Reaktionen wurden folgende Primer verwendet:

Name	Sequenz	Lokus	Position	Referenz
T3	5'- AAT TAA CCC TCA CTA AAG GG -3'	pBluescript II SK	791-772	www.stratagene.com
T7	5'- GTA ATA CGA CTC ACT ATA GGG C -3'	pBluescript II SK pGEM-T	625-646 2983-4	www.stratagene.com
SP6	5'- ATT TAG GTG ACA CTA TAG -3'	pGEM-T	143-126	www.neb.com
X-3A	5'- ATA ATT TCC CAT AAC TAA ACA CA -3'	DXZ1 α -Satelliten GGSATG	1415-1437 1389-1411	Warburton et al. 1991
X-4A	5'- TGT GAA GAT AAA GGA AAA GGC TT -3'	DXZ1 α -Satelliten	1953-1929	Warburton et al. 1991
X-4Ag	5'- TGT GAA GAT AAA GCC TTT TCC -3'	GGSATG	1907-1887	diese Arbeit
VP1F	5'- CGA TGT CAA TTC AGA ACA TCA TTG -3'	pCYPAC2	6501-6524	diese Arbeit
VP1R	5'- TAG CCC GTC TAA CAC CTA TTG C -3'	pCYPAC2	7763-7742	diese Arbeit
VP2F	5'- TCG CTA AAG CCT GTG GTT TCC -3'	pCYPAC2	5175-5195	diese Arbeit
VP2R	5'- ATG TAT GCG TAG ATG CTT GTA C -3'	pCYPAC2	6194-6173	diese Arbeit

Folgende Oligonukleotide wurden nach ihrer Hybridisierung als doppelsträngige DNA in den Vektor BsAD kloniert:

Name	Sequenz
DraF	5'- GAT CGC CCG GCA CTG GGT GAT TCG -3'
DraR	5'- AGC TCG AAT CAC CCA GTG CCG GGC -3'

6. Längenstandards

DNA Molecular Weight Marker V (LS V) (8-587 bp)	Roche
1 kb DNA ladder (75-12216 bp)	Gibco BRL
MidRange I PFG Marker (15-300 kb)	NEBioLabs
MidRange II PFG Marker (24-300 kb)	NEBioLabs

Die PFG (Pulsfeldgel)-Marker liegen in Gelform vor und wurden mit LMP-Agarose in die Geltaschen eingegossen. Der LS V- und der 1 kb-Marker wurden mit TE-Puffer und Auftragspuffer im Verhältnis 1:5:4 verdünnt. In der Regel wurden zu beiden Seiten der aufzutrennenden DNA je 3 µl aufgetragen.

7. Enzyme

Die Enzyme wurden, soweit nicht anders beschrieben, mit dem vom Hersteller gelieferten Puffer unter den empfohlenen Bedingungen verwendet.

Restriktionsenzyme	Roche, NEBioLabs
Alkalische Phosphatase	Roche
T4 DNA-Ligase	Roche
Lysozym	Sigma
Proteinase K	Roche
RNase A	Sigma
Taq-DNA-Polymerase	Amersham
Klenow-Polymerase	Roche
S1-Nuklease	Amersham

8. Chemikalien

Die verwendeten Chemikalien entsprachen soweit nicht anders angegeben dem Reinheitsgrad *pro analysi*.

Agar	Fluka
Agarose „low melting point“ (LMP)	BioRad
Agarose „low melting point“ (LMP), „ultrapure“	Gibco BRL
Agarose (Sea Kem)	Biozym

Ampicillin	Sigma
ATP	Sigma
Borsäure	Fluka
Brij 58	Sigma
Bromphenolblau	Sigma
BSA	Sigma
Chloramphenicol	Sigma
Chloroform	Merck
Deoxycholat	Sigma
Dihydrofuran	Merck
dNTP	Amersham
EGTA	Fluka
Eisessig	Merck
Ethanol absolut	Merck
Ethidiumbromid	Sigma
Ficoll 400	Pharmacia
Formamid	Sigma
Gelatine	Sigma
Glaswolle	Serva
Glycerin (etwa 87 %)	Merck
Hefeextrakt	Fluka
IPTG	Sigma
Kaliumchlorid	Merck
Kanamycin	Sigma
Magnesiumchlorid	Sigma
Natriumchlorid	Merck
Natriumcitrat	Calbiochem
Natriumhydroxid (Plätzchen)	Merck
N-Lauroyl-Sarkosin	Sigma
Orange G	Sigma
Phenol (gelöst)	Appligene oncor
Salzsäure	Merck
Sephadex-G50 DNA Grade	Pharmacia
SDS	Serva
Spermidin	Sigma
Titriplex III (EDTA)	Merck
Tris	USB
Triton X-100	Sigma
Trypton/Pepton	Fluka
X-gal	Sigma
Xylencyanol	Merck
Zinkchlorid	Sigma

9. Radioaktive Substanzen

$[\alpha^{32}\text{P}]$ -dCTP	Amersham
-------------------------------	----------

10. Geräte und sonstige Materialien

Plastikwaren allgemein:

Einmalimpfösen	Greiner
Einwegpipetten 2 ml, 5 ml, 10 ml, 25 ml	Sarstedt
Gelblöckchenform CHEF Disposable Plug Mold	BioRad
Mikroreaktionsgefäße 1,5 ml	Sarstedt
Mikroreaktionsgefäße 1,5 ml mit Schraubverschluß	Sarstedt
Mikroreaktionsgefäße 0,5 ml für PCR (dickwandig)	Eppendorf
Petrischalen (Ø 10 cm)	Greiner
Pipettenspitzen gelb, blau	Sarstedt
Pipettenspitzen weiß	Biozym
Röhrchen 14 ml	Greiner
Röhrchen 15 ml, 50 ml	Sarstedt
Spritzen 1 ml	

Sonstiges Verbrauchsmaterial:

Röntgenfilme Fuji Medical X-Ray	Fuji Film
Farbdiafilm Ektachrome E100S	Kodak
Einmalskalpell Cutfix	Braun
Elektroporationsküvetten 1 mm, 2 mm	BioRad
Nylonmembran Biodyne B	Pall
Handschuhe, Latex	Safeskin
Handschuhe, Nitril	Ansell Medical

Geräte:

Zentrifuge mit Kühlung J-21B	Beckman
Schwenkbecherzentrifuge Labofuge	Heraeus
Schwenkbecherzentrifuge mit Kühlung Rotixa /RP	Hettich
Tischzentrifuge Mikro 12-24	Hettich
Tischzentrifuge mit Kühlung Mikro 22R	Hettich
Tischzentrifuge Universal 2S (Isotopenlabor)	Hettich
SpeedVac Vacuum Concentrator	Bachofer
Autoklav Bioclav	Schütt
Sterilbank für Bakterienkultur	BDK
Brutschrank für Bakterienkultur	National Appliance Co.
Schüttler für Bakterienkultur Unitron	Infors HT
Elektroporationsapparatur Gene Pulser II	BioRad
Elutionskammer	BioRad
Elutionskammer Biotrap* BT1000	Schleicher & Schüll
Elutionsmembranen BT1, BT2	Schleicher & Schüll
Gelelektrophoresekammern, Gelträger und -kämme	Pharmacia, Hauswerkstatt
Power Supply E425 und E502	Bachofer
Power Supply Power Pac 300	BioRad
UV-Handlampe HL-6-M (Wellenlänge 312 nm)	Bachofer

UV-Transilluminator UVT-20 S/M (Wellenlänge 254 nm)	Herolab
Kamera für Gelfotos EASY 429K	Herolab
Kolbenhubpipetten	Gilson
Pipetboy mit Akku	Integra Biosciences
PCR-Maschine DNA Engine	Biozym
PCR-Maschine Thermocycler 60	Bachofer
PF-Control Module Chef-DR II	BioRad
PF-Drive Module Chef-DR II	BioRad
PFG-Elektrophoresekammer Interlock (mit Pumpe)	BioRad
Glaskolben GL45	Schott
Hybridisierofen MiniHybi	Bachofer
Entwicklermaschine Compact 35	Protec
Diascanner CoolScan III	Nikon
pH-Meter pH 530	WTW
Sequenzierer ABI PRISM 377	Applied Biosystems
Vortex REAX 2000	Heidolph
Waage AC100 und P1210	Mettler
Wärmeschrank (37°C)	Heraeus
Wasserbad T/MA6 und RMS/RM6	Lauda

11. Computerprogramme

EasyWin32, Herolab: Aufnahme und Speicherung von Gelfotos

ABI PRISM, Perkin Elmer: Sequenzauswertung

SeComP (<http://www.tobias-schwarz.net/>): Sequenz-Alignment und Homologie-Bestimmungen

Blast (<http://www.ncbi.nlm.nih.gov/BLAST/>): Suche von homologen Sequenzen

Blast2Sequences (<http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi>): Sequenz-Alignment

HUSAR (<http://genius.embnet.dkfz-heidelberg.de/menu/w2h/w2hdkfz/>): Sequence Analysis Package

Microsoft Excel: Berechnung der Homologie-Mittelwerte

SPSS: Kolmogoroff-Smirnow-Test für Einstichproben-Probleme

12. Datenbanken

EMBL-Nucleotide Sequence Database (<http://www.ebi.ac.uk/embl/index.html>)

Entrez Nucleotide database: (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Nucleotide>)

E. Diskussion

Durch die Entwicklung von künstlichen Chromosomen als nicht integrierende, autonom replizierende genomische Vektoren, hofft man viele der Probleme umgehen zu können, die bei den bislang hauptsächlich für die Expression von Genen in eukaryotischen Zellen verwendeten, in das Genom der Wirtszelle integrierenden Vektoren oder transient exprimierenden cDNAs beobachtet werden (Ineffizienz, Apoptose, Fehlregulation, Amplifikation, virale Umorganisation des Zellkerns, Tumorinduktion). Dabei sind die bisher einzigen Sequenzen, die imstande waren eine *de novo* Zentromer-Bildung zu induzieren, homogene α -Satelliten-Repeats. Diese Tandem-Repeats sind allerdings nicht konserviert, sondern unterliegen raschen Veränderungen im Rahmen einer gemeinsamen Evolution, und gelegentlich können funktionelle Zentromere auch an nicht-Satelliten-Sequenzen gefunden werden. Weil also die Sequenz-Funktions-Relation der α -Satelliten-DNA nach wie vor ungeklärt ist und verschiedene Prozesse an den Zentromersequenzen ablaufen (epigenetic maintenance, *de novo* formation), wäre ein entscheidender Schritt, die molekularen Vorgänge, die an der Zentromer-DNA stattfinden und zur Evolution der Tandemrepeat-Sequenzen beitragen, zu verstehen. Solche Mechanismen könnten zum Beispiel eng mit der Zentromerfunktion im Zusammenhang stehen, könnten aber auch die grundsätzlichen Vorgänge der Vererbung genetischen Materials wie Replikation, Reparatur und Segregation widerspiegeln, und im Bereich der „neutral“ evolvierenden Zentromersequenzen besonders gut sichtbar werden.

Zu Beginn dieser Arbeit war jedoch die Verteilung von Varianten in der Satelliten-DNA der Zentromere, welche die Spuren der gesamten molekularen Mechanismen darstellen, nur näherungsweise bekannt, da es bislang nicht gelungen ist einen längeren, zusammenhängenden Abschnitt von homogenen α -Satelliten-Repeats zu sequenzieren (Henikoff 2002). Alle bisher durchgeführten Studien hatten vergleichsweise wenige, kurze DNA-Fragmente mit unbekannter Lokalisation innerhalb der α -Satelliten-Arrays, teilweise aber auch von deren divergierten Rändern (Schueler et al. 2001, Mashkova et al. 1998) oder von unterschiedlichen Chromosomen zugrunde liegen (Waye und Willard 1986, Warburton und Willard 1990).

Primäres Ziel dieser Arbeit war daher die Charakterisierung von fünf X-chromosomalen α -Satelliten-PACs mit Insert-Längen von 120-160 kb, welche aus einem männlichen Probanden isoliert wurden. Die Verwendung dieser charakterisierten α -Satelliten-DNAs für die Konstruktion von künstlichen Chromosomen sollte in folgenden Experimenten

Rückschlüsse von der primären α -Satelliten-Sequenz auf die Häufigkeit der *de novo* Ausbildung von Zentromeren sowie deren mitotische Funktion erlauben. Die Charakterisierung der PACs ergab innerhalb und untereinander eine sehr hohe Homogenität, von der angenommen wird, dass sie für eine *de novo* Zentromer-Formierung essentiell ist. Zusätzlich fiel besonders in den homogensten Abschnitten eine bislang unbekannte und unerwartete Verteilung der variablen Basen auf. Die vorliegende Arbeit diene im Folgenden der Analyse dieser Varianten, um die zugrunde liegenden Vorgänge besser verstehen zu können.

1. Die stabil klonierten PACs repräsentieren mindestens 500 kb des α -Satelliten-Arrays eines einzelnen, menschlichen X-Chromosoms

Die in dieser Arbeit untersuchten fünf α -Satelliten-PACs entstammen dem X-Chromosom eines einzelnen, männlichen Individuums. Vorarbeiten von Dr. D. Schindelhauer wiesen darauf hin, dass der gesamte DXZ1-Array dieses Chromosoms in der humanen PAC-Bank RPCI-I, aus welcher die verwendeten PACs isoliert wurden, repräsentiert ist (Schindelhauer und Schwarz 2002). Durch wiederholte Längenbestimmungen, sowie die Kartierungen zweier PACs nach Wachstum des Wirtsbakteriums über circa 400 Generationen konnten ferner die Stabilität der repetitiven DNA im PAC-Vektor belegt und häufige Rekombinationen der α -Satelliten-Sequenzen im Wirtsbakterium ausgeschlossen werden.

Die genaue Lage der PACs innerhalb des Zentromer-Arrays des X-Chromosoms ist nicht bekannt. Anhand der Kartierungen der variablen Basen konnte jedoch ein wesentliches Überlappen der Klone, mit Ausnahme der PACs A8 und A10, ausgeschlossen werden. Somit repräsentieren die PACs mindestens 500 kb des α -Satelliten-Arrays des X-Chromosoms. Damit bieten sie ideale Voraussetzungen für die Analyse der α -Satelliten-DNA eines einzelnen Zentromer-Arrays.

2. Die α -Satelliten-DNA der PACs weist eine sehr hohe Homogenität auf und macht eine Sequenzierung mit konventionellen Methoden unmöglich

Für die Charakterisierung der PACs wurden zunächst jeweils zehn 2 kb *Bam*-HI Fragmente subkloniert und endsequenziert. Dabei zeigte sich, dass die Higher order repeats durchschnittliche Sequenz-Homologien von über 98,5% aufweisen mit einer maximalen Divergenz innerhalb eines PAC von lediglich 2,1%. Der bei weitem größte Teil dieser Divergenz wird durch variable Basen verursacht (Abb. 8), welche bimorphe Positionen in den Higher order

repeats eines Arrays darstellen. Abweichungen von der Konsensussequenz, welche in nur einem Repeat auftreten und somit potentielle Neumutationen darstellen, sind, selbst in der vergleichsweise kleinen Anzahl der von uns untersuchten Higher order repeats, bereits sehr selten. Somit muss die Häufigkeit von „echten“ Einzelmutationen, welche also im gesamten Zentromer-Array nur ein einziges mal vorkommen, auf weit unter 0,1% geschätzt werden.

Diese hohe Homogenität der Repeats bringt mit sich, dass die Kartierung und Sequenzierung eines kompletten Zentromers mit den heute üblichen Methoden kaum möglich ist. Auf der einen Seite wird die Erstellung von physikalischen Karten der Zentromer-Regionen sowie die Konstruktion von PAC-Contigs durch das Fehlen von seltenen Restriktionsschnittstellen erschwert, auf der anderen Seite wird die Sequenzierung der PACs mittels der heute üblichen Shotgun-Methode durch die Seltenheit von Einzelmutationen verhindert. Und auch wenn die variablen Basen in einer Vielzahl von unterschiedlichen Kombinationen auftreten, kann nicht abgesehen werden, inwieweit diese „einzigartig“ sind: So könnten die Varianten an mehreren Stellen des Arrays identische Kombinationen bilden oder aber durch somatische Rekombinationsvorgänge in verschiedenen Zellen eines Organismus voneinander abweichen. Für die Sequenzierung von Zentromeren müssen also Methoden entwickelt werden, welche es erlauben die Position eines Sequence-Reads unabhängig von seiner Sequenz zu bestimmen.

Für die Bestimmung der Homogenität der PACs wurden jedoch nicht nur die Sequenz-Homologien der Higher order repeats und die Häufigkeit von Einzelmutationen berücksichtigt, sondern auch das Vorkommen von Strukturvarianten der üblicherweise 12-meren X-chromosomalen Higher order repeats. Derartige Fragmente fanden sich vor allem in den PACs A6 und B11, sowie eines im PAC A8 (Tab. 2). Nicht- α -Satelliten-DNA, wie zum Beispiel LINE-Elemente, welche im Randbereich des X-chromosomalen Zentromers gefunden wurden und somit das höhere Alter dieser Region belegen (Schueler et al. 2001), fand sich in keinem der PACs.

Somit ergibt die Charakterisierung der α -Satelliten-DNA ein konsistentes Bild bezüglich der hohen Homologie zwischen den Higher order repeats und der gleichzeitigen Seltenheit von Einzelmutationen und Higher order repeats mit abweichenden Längen (Abb. 10): Die PACs A6 und B11 sind etwas divergenter mit 98,8 bzw. 98,6% mittlerer Repeat-Homologie, 0,33% bzw. 0,20% Einzelmutationen und 9 bzw. 3 nicht-2 kb Fragmenten. Die sehr homogenen PACs A7, A8 und A10 weisen dagegen mittlere Repeat-Homologien von 99,1-99,3%, Einzelmutationen von 0,07-0,14% und nur ein abweichendes *Bam*-HI Fragment im PAC A8 auf.

3. Die Ergebnisse ermöglichen weiterführende Untersuchungen von *de novo* erzeugten künstlichen Chromosomen

Die in dieser Arbeit beschriebenen PACs stellen die am besten charakterisierten zusammenhängenden Abschnitte von homogener, menschlicher α -Satelliten-DNA dar und ermöglichen damit weiterführende Studien zur *de novo* Ausbildung von künstlichen Chromosomen und der Funktion ihrer Zentromere. Dazu wurden die PAC-Inserts in den Vektor pTAT-Bs kloniert, welcher neben einem Blastocidin-Resistenzgen Telomer-Sequenzen von 800 bp Länge enthält. Durch Lipofektion einiger dieser Konstrukte konnten bereits künstliche Chromosomen in geeigneten eukaryotischen Zelllinien etabliert werden.

Ein Vergleich der sehr homogenen PACs A7, A8 und A10 mit den etwas stärker divergierten PACs A6 und B11 könnte möglicherweise Unterschiede in der *de novo* Ausbildung von Zentromeren oder deren Funktion aufdecken. Beispielsweise könnte durch Kultur der Wirtszellen ohne Blastocidin-Selektion die Stabilität der HACs und somit die mitotische Funktion ihrer Zentromere ermittelt werden. Ferner könnte durch Immunfluoreszenz-Untersuchungen die Anlagerung von Zentromer-spezifischen Proteinen und somit die Ausbildung einer Zentromer-spezifischen Chromatinstruktur untersucht werden (Grimes et al. 2004, Nakashima et al. 2005, Lam et al. 2006). In weiteren Studien könnte die Häufigkeit von verspäteten Trennungen und Nondisjunktionen der Chromatiden in Abhängigkeit von der α -Satelliten-Sequenz ermittelt werden (Rudd et al. 2003, Tsuduki et al. 2006). Dies würde möglicherweise zu einem besseren Verständnis der Sequenz-Funktions Relation der Zentromere beitragen.

Allerdings sind die Sequenz-Unterschiede zwischen den PACs relativ gering und die *de novo* Ausbildung eines künstlichen Chromosoms stellt ein vergleichsweise seltenes und von vielen Faktoren abhängiges Ereignis dar. Da ferner die Effizienz der Ausbildung von künstlichen Chromosomen und die Funktion ihrer Zentromere nur mit hohem technischem Aufwand und einer gewissen auswertungsbedingten Unschärfe zu erfassen sind, erscheint fraglich, ob der direkte Vergleich der hier untersuchten PACs quantitativ messbare Unterschiede ergeben würde.

Nichtsdestotrotz bieten die PACs ideale Voraussetzungen für die Konstruktion und Analyse von künstlichen Chromosomen.

4. Die variablen Basen treten unabhängig von ihrer Lokalisation im Higher order repeat auf

Eine Analyse der Zentromer-DNA ist jedoch nicht nur hinsichtlich der Konstruktion von künstlichen Chromosomen erforderlich. Da die α -Satelliten-Repeats scheinbar keiner direkten Selektion unterliegen (Brutlag 1980, Durfy und Willard 1989), eignen sie sich hervorragend um die Mechanismen zu untersuchen, welche der Homogenisierung und Fixierung von neuen Varianten, und somit dem Prozess der gemeinsamen Evolution dieser repetitiven DNA zugrunde liegen. Dabei könnte die etwas niedrigere Homogenität der PACs A6 und B11 bedeuten, dass sie nur noch eingeschränkt am Prozess der Homogenisierung teilnehmen. Die PACs A7, A8 und A10 hingegen weisen eine sehr hohe Homogenität und nur sehr wenige Einzelmutationen auf, was darauf hindeutet, dass diese Bereiche des α -Satelliten-Arrays aktuell und/oder effektiv am Prozess der Homogenisierung teilhaben. Somit liefert vor allem die Analyse der variablen Basen dieser PACs Hinweise auf die Ursachen, Mechanismen und treibenden Kräfte der gemeinsamen Evolution dieser tandem-repetitiven DNA.

Wie Abb. 11 zeigt, treten sowohl die variablen Basen als auch die potentiellen Einzelmutationen gleichmäßig verteilt über die sequenzierten Abschnitte der Higher order repeats auf. Es findet sich keine Abhängigkeit der Varianten von bestimmten Sequenzeigenschaften, wie zum Beispiel eine Bevorzugung bestimmter Nukleotide, eine Konsensussequenz im Bereich der variablen Basen, oder das Vorkommen von CENP-B Boxen. Diese Ergebnisse wurden auch durch die PCR-Analysen von α -Satelliten-DNA der dem Menschen nahe verwandten Spezies *Gorilla gorilla* bestätigt, die ein komplett anderes Set an variablen Positionen in ihren, dem Menschen vergleichsweise ähnlichen, X-chromosomalen α -Satelliten-Repeats aufweisen. Damit ist eine Abhängigkeit von der primären DNA-Sequenz unwahrscheinlich, was mit den wahllos erscheinenden, drastischen Unterschieden in den Zentromersequenzen der höheren Eukaryoten gut vereinbar ist.

5. Die häufigen Varianten sind in der Bevölkerung fixiert

Um für die weiteren Analysen die im α -Satelliten-Array häufigen und somit möglicherweise fixierten Varianten zu identifizieren, wurden diejenigen 18 variablen Positionen ausgewählt, welche in mindestens drei Higher order repeats sowie in mindestens zwei PACs vorkommen. Für die Bestimmung der Verbreitung dieser Varianten in der Bevölkerung wurde die Beeinflussung der Erkennungssequenzen von Restriktionsenzymen durch einige der variablen Basen genutzt. Es wurden Polymerase-Kettenreaktionen durchgeführt, welche spezifisch und

repräsentativ einen 489 bp langen Abschnitt der Higher order repeats des X-chromosomalen Zentromers amplifizieren (Warburton und Willard 1992). Durch Restriktionsverdau des PCR-Produkts mit geeigneten Enzymen konnte so das Vorkommen von fünf der häufigen variablen Basen in der Bevölkerung untersucht werden.

Wie die Restriktionsanalysen der α -Satelliten-DNA von zehn nicht-verwandten Europäern zeigten, liegen alle untersuchten Varianten in jeweils ähnlichen Verhältnissen vor und sind somit in der Bevölkerung fixiert. Diese Ergebnisse wurden durch die Klonierung und Sequenzierung von PCR-Fragmenten, sowie durch den Restriktionsverdau von kompletter genomischer DNA mit anschließender Hybridisierung durch eine spezifisch X-chromosomale DNA-Sonde bestätigt. Sie entsprechen ferner den Ergebnissen von Warburton und Willard (1992, 1995), die bei der direkten Sequenzierung von X-3A/X-4A PCR-Produkten dreier Maus/Mensch-Zellhybride fünf variable Stellen beschrieben, die in allen drei Hybriden in gleichen Verhältnissen vorlagen.

Die Tatsache, dass die Varianten nicht nur in allen untersuchten Individuen existieren, sondern auch in jeweils ähnlichen Verhältnissen vorliegen, bedeutet auch, dass die Ausbreitung einer neuen Variante auf die homologen Chromosomen einer Population schneller erfolgt als die Ausbreitung über den gesamten α -Satelliten-Array. Die Fixierung einer Variante in der Bevölkerung erfolgt somit schneller als die Homogenisierung eines Arrays.

6. Die schnelle Ausbreitung der Varianten erfolgt maßgeblich durch Genkonversionen

Die Existenz der variablen Basen, welche Ausdruck der gemeinsamen Evolution der α -Satelliten-DNA zu sein scheinen, ist bereits seit langer Zeit bekannt (Durfy und Willard 1987, Warburton und Willard 1992). Die Verteilung der Varianten innerhalb der Zentromer-Arrays wurde jedoch noch nie detailliert untersucht. Das vermeintlich gemeinsame Auftreten bestimmter Gruppen von variablen Basen wurde als Existenz von Haplotypen interpretiert, welche als das Ergebnis der Amplifikation von Higher order repeats gedeutet wurden (Warburton und Willard 1992). Aufgrund der vorliegenden Daten vermutete man, dass diese Amplifikationen eine Folge von wiederholtem Crossing over zwischen ungleich alignen Repeats darstellen. Diese Vermutung stützte sich weitgehend auf Computersimulationen (Smith 1976), ein direkter Nachweis der Beteiligung von Unequal crossing over an der

gemeinsamen Evolution der Zentromer-DNA ist bislang nicht hinreichend gelungen (vgl. Kap. 6.4.1).

Da die in dieser Arbeit sequenzierten Varianten einen „Schnappschuss“ der Vorgänge im Zentromer-Array repräsentieren, können anhand ihrer Verteilung Rückschlüsse auf die sie beeinflussenden molekularen Mechanismen gezogen werden. Versuche, die 18 in den PACs verbreiteten Varianten zu gruppieren und somit bestimmten Haplotypen zuzuordnen, ergaben eine voneinander unabhängige Verbreitung der einzelnen variablen Basen im Zentromer-Array (Abb. 15). Eine Ableitung von Haplotypen anhand des gemeinsamen Auftretens bestimmter Varianten war nicht möglich. Diese individuell unabhängige Ausbreitung der variablen Basen im Zentromer-Array deutet auf Genkonversion als den für die Verbreitung einer Variante verantwortlichen Mechanismus hin. Würde die Ausbreitung von Neumutationen durch Unequal crossing over oder durch die Vermehrung kompletter Higher order repeats wie zum Beispiel über eine Rolling circle Amplifikation erfolgen, so würden sich nahe beieinander liegende Varianten gemeinsam verbreiten, was eine Ausbildung von Haplotypen zur Folge hätte. Ein derartiges Muster ließ sich jedoch, gerade in den homogensten Segmenten, nicht erkennen.

Um die Hypothese der Ausbreitung von Varianten über Genkonversionen zu bestätigen, wurde erneut die Beeinflussung der Erkennungssequenzen von Restriktionsenzymen durch einige der variablen Basen benutzt, um diese innerhalb der PACs zu kartieren. So konnte durch Partialverdau der PAC-DNA und Hybridisierung mit einer radioaktiv markierten DNA-Sonde die exakte Verteilung einzelner Varianten innerhalb der PACs ermittelt werden. Die Kartierungen der variablen Basen an den Positionen 35 und 1562 mit den Enzymen *Hind*-III und *Dra*-III zeigten, dass sich Varianten in Form von Clustern ausbreiten, also Abschnitten von mehreren Higher order repeats, die an einer Position ein bestimmtes Nukleotid aufweisen, getrennt durch Gruppen von Higher order repeats, die an dieser Position ein anderes Nukleotid besitzen (Abb. 18).

Derartige Cluster sind zwar mit Unequal crossing over Ereignissen mit einem ungleichen Alignment von jeweils einem bis zwei Higher order repeats vereinbar, ihr Auftreten würde jedoch auch zu einer synchronen Ausbreitung benachbarter Varianten führen und somit die Higher order repeats eines Clusters zu einem Haplotyp gruppieren. Um dies auszuschließen wurde eine Reihe weiterer 2 kb Fragmente sowohl aus den *Dra*-III-Clustern des PACs A7 als auch aus den dazwischen liegenden Bereichen ohne *Dra*-III Restriktionsschnittstellen kloniert und endsequenziert. Dies zeigte, dass auch hier eine Zuordnung der Repeats zu den beiden

Clustern oder eine Ableitung von Haplotypen nicht möglich ist. Somit treten die variablen Basen nicht nur in den zufällig aus allen PACs ausgewählten Higher order repeats, mit Abständen von bis zu 500 kb, sowie innerhalb der PACs mit Abständen von 2 bis 160 kb unabhängig voneinander auf, sondern auch innerhalb der Cluster mit Abständen von unter 30 kb. Dieses individuell unabhängige Auftreten der Varianten und das Fehlen von Haplotypen belegen, dass die Cluster durch Genkonversionen entstanden sein müssen.

7. Genkonversionen sind häufige Ereignisse in tandem-repetitiver DNA

Die identischen Strukturen der beiden *Hind*-III Cluster im PAC B11 (Abb. 18) lassen vermuten, dass es hier zur Verdoppelung eines 35-50 kb langen DNA-Abschnittes kam, zum Beispiel durch ein Unequal crossing over mit einem ungleichen Alignment von 17-24 Higher order repeats. Die Sequenzanalysen der 2 kb *Hind*-III Fragmente aus diesen Clustern, sowie die Kartierung einer *Dra*-III Schnittstelle in einem der Cluster (Abb. 18) belegen jedoch, dass nach dieser Duplikation weitere Veränderungen stattgefunden haben müssen, welche nur durch Genkonversionen zu erklären sind. Auch wenn die genauen zeitlichen Abläufe der einzelnen Rekombinationsereignisse unbekannt sind, lässt dies vermuten, dass Genkonversionen in der α -Satelliten-DNA erheblich häufiger auftreten, in der Gesamtheit also effektiver Spuren hinterlassen, als Unequal crossing over Ereignisse. Das Fehlen von Einzelmutationen, durch welche sich die in den Varianten identischen Higher order repeats

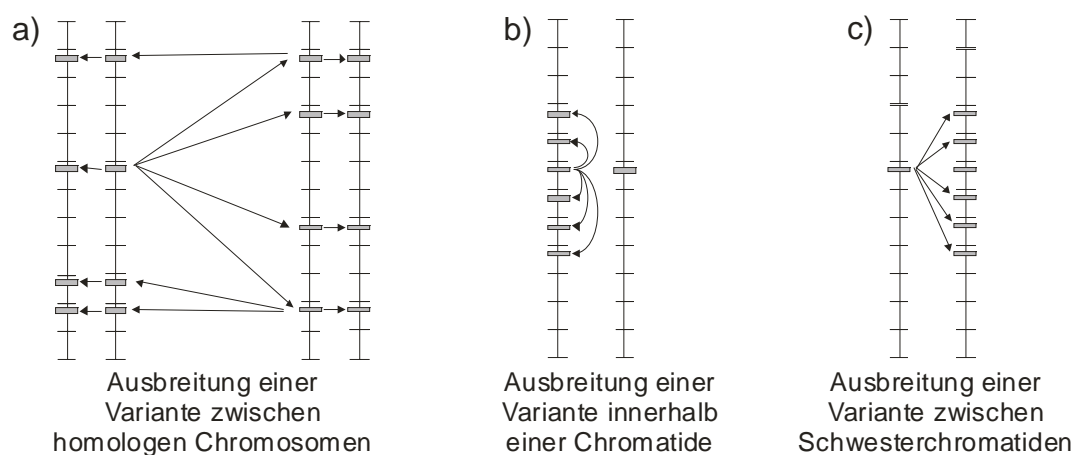


Abb. 24: Intrachromosomale Rekombinationen treten erheblich häufiger auf als interchromosomale: Bei einem Überwiegen der interchromosomalen Ausbreitung der Varianten würde es nicht zur Ausbildung von Clustern kommen, da ein Prozess, welcher gezielt einige wenige Higher order repeats innerhalb eines Arrays aus tausenden von fast identischen Repeats identifizieren kann höchst unwahrscheinlich ist (a). Somit muss die Ausbildung der Cluster durch einen intrachromosomalen Prozess, entweder innerhalb einer Chromatide (b) oder zwischen den Schwesterchromatiden eines Chromosoms (c) erfolgen.

aus den beiden Clustern unterscheiden ließen, ist entweder durch eine Verdoppelung des DNA-Fragments vor relativ kurzer Zeit erklärbar, was das häufige Auftreten von Genkonversionen unterstreichen würde, oder durch eine sehr effektive Elimination aller aufgetretenen Neumutationen, was wiederum für eine starke Beteiligung von Genkonversionen sprechen würde.

8. Die Ausbildung der Cluster erfolgt durch intrachromosomale Genkonversionen

Wie die Kartierungen zeigen, verbreiten sich Varianten von einem Higher order repeat ausgehend nicht zufällig im Array, sondern bevorzugt auf benachbarte Repeats. Würde die Ausbildung dieser Cluster durch einen interchromosomalen Prozess zwischen homologen Chromosomen erfolgen, müsste ein Mechanismus existieren, welcher die Varianten von einer Chromatide gezielt an eine bestimmte Position innerhalb eines Arrays aus tausenden von Higher order repeats einer anderen Chromatide überträgt. Da sich die Repeats der einzelnen Cluster jedoch in lediglich einer einzigen Base unterscheiden, ist ein derartiger Mechanismus nur schwer vorstellbar (Abb. 24a). Daraus lässt sich schließen, dass die Ausbildung der Cluster durch intrachromosomale Genkonversionen erfolgen muss, entweder innerhalb einer Chromatide (Abb. 24b) oder zwischen den Schwesterchromatiden eines Chromosoms (Abb. 24c). Diese intrachromosomalen Rekombinationen müssen häufiger stattfinden als interchromosomale Rekombinationen, welche die Cluster durch ungezielte Sequenzveränderungen wieder zerstören würden. Eine wesentlich höhere Rate an intra- als an interchromosomalen Ereignissen entspricht auch den Ergebnissen früherer Studien (Bollag et

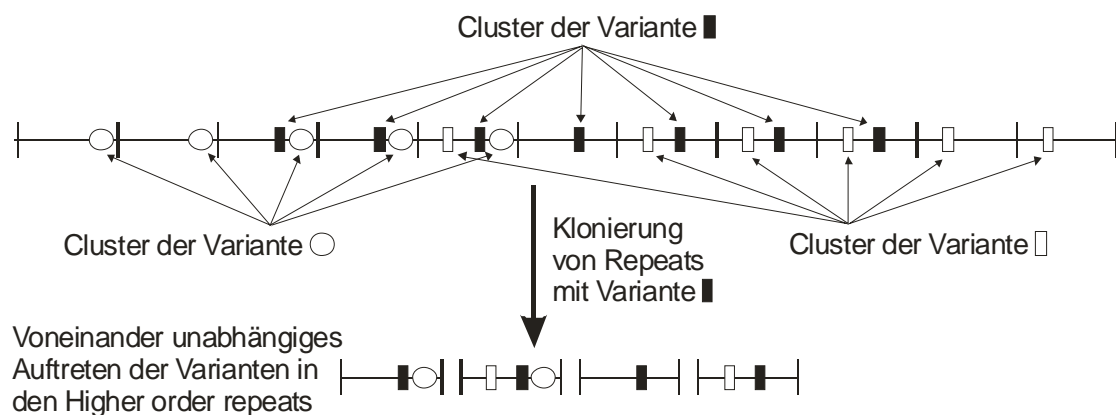


Abb. 25: Die Ausbreitung der Varianten durch Genkonversionen führt zur Ausbildung von sich überlappenden Clustern. Somit kommt es zwar zum gehäuft gemeinsamen Auftreten von Varianten, diese können jedoch auf den Higher order repeats durch andere, davon unabhängige Varianten getrennt sein.

al. 1989, Warburton und Willard 1992, Warburton und Willard 1995). Die Schlussfolgerung aus der Varianten-Analyse der α -Satelliten-Repeats ist somit, dass sich Neumutationen, sofern sie nicht sofort wieder beseitigt werden, durch häufige, intrachromosomale Genkonversionen in der Form von Clustern im Array ausbreiten.

Betrachtet man den gesamten α -Satelliten-Array, so führt die lokale Ausbreitung der Varianten durch Genkonversionen zu einem Bild von sich überlappenden Clustern (Abb. 25). Obwohl sich die einzelnen Varianten unabhängig voneinander verbreiten, hat dies ein gehäuftes gemeinsames Auftreten bestimmter Nukleotide zur Folge. Durch dieses Überlappen von Clustern kann die sowohl in einer früheren Studie (Durfy und Willard 1989), als auch in dieser Untersuchung beobachtete größere Homologie nahe benachbart liegender Higher order repeats (aus PACs klonierte Repeats) verglichen mit zufällig aus einem α -Satelliten-Array ausgewählten Sequenzen (zufällige PCR Amplifikation aus genomischer DNA) erklärt werden.

9. Die relativen Geschwindigkeiten der intra- und interchromosomalen Ausbreitung von Varianten

Die Beobachtungen, dass einerseits intrachromosomale Rekombinationen häufiger sind als interchromosomale und andererseits die Varianten in allen Zentromer-Arrays der Bevölkerung gleich weit verbreitet sind, zeigen, dass es sehr häufig zu lokalen Veränderungen von Varianten kommen muss (Abb. 26a) während die Ausbreitung einer Neumutation über einen gesamten α -Satelliten-Array vergleichsweise viel Zeit in Anspruch nimmt. Ob diese Ausbreitung ausschließlich durch Genkonversionen erfolgt, kann anhand der vorliegenden Daten nicht bestimmt werden. Es ist möglich, dass weitere Mechanismen, welche noch seltener als die interchromosomalen Rekombinationen auftreten (Abb. 26c), zu größeren „Sprüngen“ von Varianten innerhalb einer Chromatide führen (Abb. 26b). Dies könnte zum Beispiel durch Unequal crossing over erfolgen, worauf die zum Teil großen Unterschiede in der Gesamtlänge der α -Satelliten-Arrays von homologen Chromosomen (Wevrick und Willard 1989, Mahtani und Willard 1990) oder die von uns beobachtete Duplikation von mindestens 17 Higher order repeats hindeuten.

Vermutlich noch seltener kommt es zu Rekombinationen zwischen den unterschiedlichen Array-Typen eines Chromosoms wie zum Beispiel des Chromosoms 17 (Abb. 26d). Hier führen wahrscheinlich die zunehmenden Unterschiede in der primären α -Satelliten-Sequenz zu einer Reduktion der Rekombinationsereignisse. Die rapide Abnahme der Häufigkeit von

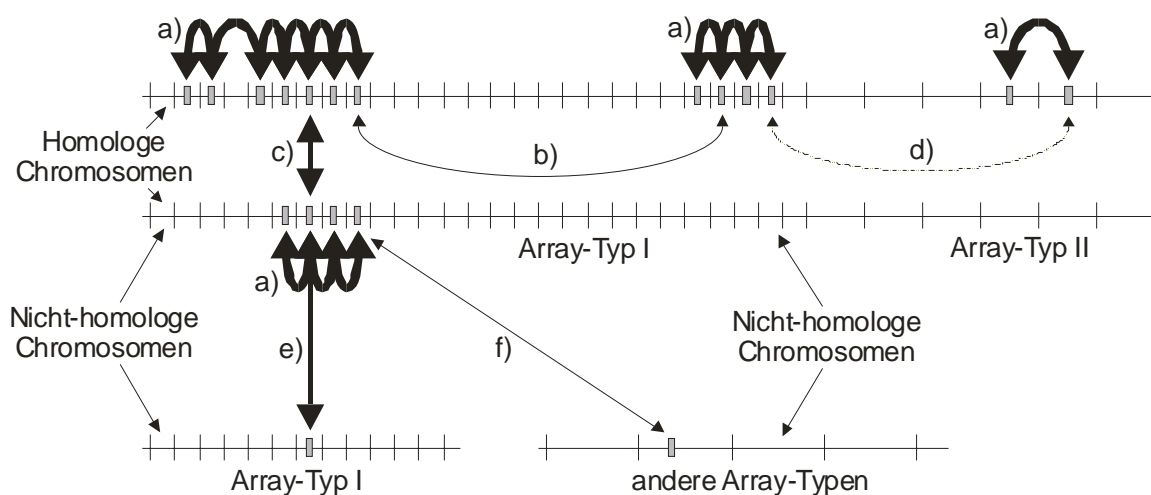


Abb. 26: Die relativen Geschwindigkeiten der intra- und interchromosomalen Ausbreitung der Varianten. Die Dicke der Pfeile repräsentiert die Häufigkeit von Rekombinationsereignissen. Eine Beschreibung der einzelnen Mechanismen findet sich im Text.

Genkonversionen, bereits bei einer Divergenz der homologen Sequenzen von nur wenigen Prozent (Elliott et al. 1998, Johnson und Jasin 2000), könnte so zu einer getrennten Entwicklung von mehreren Array-Typen auf einem Chromosom beitragen.

Zwischen nicht-homologen Chromosomen kann es bei einer hohen Übereinstimmung der α -Satelliten-Arrays möglicherweise zu relativ häufigen Rekombinationsereignissen und damit zu einer gemeinsamen Evolution kommen. So besitzen die akrozentrischen Chromosomen 13 und 21 sowie die Chromosomen 14 und 22 des Menschen die gleichen α -Satelliten-Repeats. Interessanterweise besitzen die Chromosomen 14 und 22 des Schimpansen *Pan troglodytes* (homolog zu den Chromosomen 13 und 21 des Menschen), sowie dessen Chromosomen 15 und 23 (homolog zu den menschlichen Chromosomen 14 und 22) ebenfalls sehr ähnliche α -Satelliten-Sequenzen. Allerdings gehören die Repeats der humanen Chromosomen 13 und 21 beziehungsweise 14 und 22 jeweils anderen suprachromosomalen Familien an, als die Repeats der homologen Chromosomen 14 und 22 beziehungsweise 15 und 23 des Schimpansen (Jorgensen et al. 1992). Dies könnte dadurch erklärt werden, dass eine effektive Homogenisierung zwischen nicht-homologen Chromosomen stattfinden kann, vorausgesetzt die Unterschiede in den Sequenzen sind nicht zu groß, und es besteht eine räumliche Beziehung, wie zum Beispiel bei den akrozentrischen Chromosomen, die über ihre Nucleolus organizer-Regionen im Zellkern oft in unmittelbarer Nähe vorliegen (Abb. 26e).

10. Genkonversionen und die gemeinsame Evolution der α -Satelliten-DNA

Angesichts der Vielzahl an unterschiedlichen Formen von repetitiver DNA in den verschiedensten Organismen (tandemartig vs. verteilt im Genom vorkommend, kodierend und damit der Selektion unterliegend vs. nichtkodierend) muss davon ausgegangen werden, dass sich deren Evolution nicht durch einen einzigen für alle Repeat-Formen identischen Mechanismus beschreiben lässt. Vielmehr muss angenommen werden, dass die erwähnten Rekombinations-Mechanismen wie Genkonversionen, Unequal crossing over, replikative Transpositionen oder Rolling circle Amplifikationen in unterschiedlichen Bereichen des Genoms in unterschiedlichem Ausmaß und mit unterschiedlichen Geschwindigkeiten auftreten, und damit abhängig von der Art der repetitiven-DNA unterschiedlichen Einfluss auf deren evolutionäre Entwicklung haben. So wurden beispielsweise in den Tandem-Repeats der ribosomalen RNA der Hefe sowohl mitotisches und meiotisches Unequal crossing over als auch Genkonversionen nachgewiesen (Liao et al. 1997). Es wurde sogar beobachtet, dass meiotische Genkonversionen häufig zu einem Crossing over der beteiligten DNA-Stränge führen (Bollag et al. 1989), und es gibt Modelle der homologen Rekombination, in denen Unequal crossing over oder Genkonversion die alternativen Resultate ein und desselben Prozesses darstellen (Holliday 1964, Szostak et al. 1983). Entsprechend findet während der Meiose nur in einer kleinen Zahl der synaptonemalen Komplexe ein tatsächliches Crossing over statt (in der Regel ein mal pro Chromosom), während Genkonversionen an den Rekombinationsstellen viel häufiger auftreten könnten.

Bezüglich der α -Satelliten-DNA der menschlichen Zentromere gibt es eine Reihe von überzeugenden Argumenten für das Auftreten von Unequal crossing over: So ist die Existenz von unterschiedlichen Higher order repeat-Längen in den Zentromer-Arrays des Chromosoms 17 eindeutig auf Rekombinationen zwischen ungleich alignen α -Satelliten-Monomeren zurückzuführen (Warburton et al. 1993). Außerdem wurden Unequal crossing over Ereignisse in den divergierten Randbereichen der Zentromere nachgewiesen (Mashkova et al. 1998), worauf auch die identischen Strukturen der beiden *Hind*-III Cluster im weniger homogenen PAC B11 hinweisen. Derartige Ereignisse müssen jedoch entweder als extreme Seltenheit angesehen werden, wie die Entstehung und Ausbreitung von Higher order repeats mit abweichender Monomerzahl, oder treten in Bereichen der Zentromere auf, welche scheinbar nicht mehr am Prozess der Homogenisierung teilnehmen. In dieser Studie hingegen wurden erstmals die homogensten Abschnitte eines α -Satelliten-Arrays untersucht, diejenigen Bereiche also, in denen aktuell eine Homogenisierung der Repeats stattfindet. Dabei ergaben

sich keinerlei Hinweise auf Unequal crossing over Ereignisse. Stattdessen konnten häufige, intrachromosomale Genkonversionen nachgewiesen werden.

11. Ursache und Geschwindigkeit des Molecular Drive der α -Satelliten-DNA

Das von Dover 1982 eingeführte „Molecular Drive“ umfasst alle Mechanismen, die zu einer nicht-mendelischen Vererbung von Mutationen innerhalb repetitiver DNA, und damit zu deren gemeinsamen Evolution beitragen. Dabei stehen nicht die einzelnen Mechanismen im Vordergrund, sondern die Ursache, Richtung und Geschwindigkeit der Veränderungen. Nachdem es keine direkte Sequenz-Funktions Relation der zentromerischen Satelliten-DNA zu geben scheint, und somit ein evolutionärer Vor- oder Nachteil durch einzelne, abweichende Nukleotide sehr unwahrscheinlich ist, stellt sich die Frage, weshalb sich eine bestimmte Base effektiv über mehrere tausend Higher order repeats ausbreitet, während eine andere vollständig verschwindet, und zwar ungeachtet der sonstigen Varianten in den gleichen Repeats. Solange es jedoch nicht möglich ist, die Verbreitung einer Variante zu verschiedenen evolutionären Zeitpunkten zu bestimmen, kann keine Aussage darüber gemacht werden, welche die ursprüngliche Base und welche die sich ausbreitende neue Variante ist. Bei den in dieser Arbeit untersuchten Varianten ließ sich keine Bevorzugung bestimmter Nukleotide, oder ein Überwiegen von Transitionen oder Transversionen, ermitteln.

Schwierig zu beurteilen sind auch die absoluten Häufigkeiten der einzelnen Rekombinationsmechanismen, obwohl die Seltenheit von Einzelmutationen eindrucksvoll zeigt, wie effizient Neumutationen entweder entfernt oder innerhalb des Zentromer-Arrays verbreitet werden. Prinzipiell muss zwischen der Häufigkeit des Neuauftretens von Varianten einerseits und der Geschwindigkeit der Ausbreitung dieser Varianten andererseits unterschieden werden. Die Häufigkeit einer vollständigen Homogenisierung und Fixierung eines neuen Nukleotids lässt sich Anhand der Übereinstimmung der α -Satelliten-Sequenzen unterschiedlicher Spezies abschätzen. So sind die circa 2000 bp langen X-chromosomalen Higher order repeats des Menschen und der Gorillas zu circa 93% identisch. Nachdem sich die menschliche und die Gorilla Linie vor circa 15 Mio. Jahren auftrennten, ergibt dies eine

Rate von $\frac{0,07 \times 2.000}{(15.000.000a \times 2)} = 4,6 \times 10^{-6}$ Basen Abweichung pro Jahr, beziehungsweise eine

komplette Fixierung eines neuen Nukleotids alle 214.000 Jahre. Hierbei sind jedoch nur diejenigen Varianten berücksichtigt, die sich über den gesamten Array und in der gesamten

Population verbreitet haben. Es ist sehr wahrscheinlich, dass neue Varianten erheblich häufiger auftreten, sich über einen Teil des Zentromers ausbreiten, und danach wieder verdrängt werden. Ebenso ist denkbar, dass an einer bestimmten Position innerhalb der 15 Mio. Jahre mehrere Varianten existierten. Somit stellt dies lediglich einen Mindestwert für das Auftreten von vollständig homogenisierten und fixierten neuen Varianten dar. Die exakte Geschwindigkeit, mit der sich neue Varianten ausbreiten, ließe sich nur durch eine Analyse der α -Satelliten sehr nahe verwandter Spezies oder isolierter Populationen ermitteln, in denen fixierte Varianten mit unterschiedlichen Häufigkeiten im Array vorkommen. Die Analyse der X-chromosomalen α -Satelliten-Repeats der Spezies *Gorilla gorilla* in dieser Arbeit hat keine derartige Variante ergeben. Für eine genaue Bestimmung müssten daher näher verwandte Individuen untersucht werden, wie zum Beispiel die archäologischen Funde der direkten Vorfahren des Menschen oder isolierter Naturvölker.

12. Ausblick

Einer der wichtigsten Schritte hin zu einem umfassenderen Verständnis von Struktur, Funktion und Evolution der menschlichen Zentromere wäre die Sequenzierung eines vollständigen α -Satelliten-Arrays. Wie in dieser Arbeit gezeigt werden konnte, ist dies jedoch mit den heute üblichen Techniken nicht möglich, so dass neue Methoden für die Sequenzierung von hochrepetitiver DNA entwickelt werden müssen. Dies kann zum Beispiel auf der Subklonierung der, mittels Partialverdau von intakt präparierter PAC-DNA erzeugten, „ α -Satelliten-Leitern“ beruhen.

Bei der in dieser Arbeit festgestellten, relativ hohen Rate an Genkonversionen innerhalb der Arrays, stellt sich jedoch die Frage, ob Zellen während der Entwicklung eines Individuums Konversionen durchlaufen können. Ein formaler Ausschluss unterschiedlicher Sequenzentwicklungen scheint jedenfalls nicht möglich. Daraus könnte sich ergeben, dass DNA-Fäden aus verschiedenen Zellen eines Individuums (wie die Abschnitte aus einer PAC-Bank aus Millionen von Fibroblasten) unterschiedliche Sequenzen aufweisen und somit die Erstellung eines eindeutigen Contigs unmöglich ist.

Gelingt jedoch die Sequenzierung vollständiger Zentromere, würden sich möglicherweise neue Erkenntnisse bezüglich der *de novo* Konstruktion von künstlichen Chromosomen ergeben.

Die Kombination von künstlichen Chromosomen mit der Sequenzierung kompletter α -Satelliten-Arrays würde schließlich eine genaue Analyse der dem Molecular Drive

zugrunde liegenden Mechanismen ermöglichen. Die Reklonierung und Sequenzierung des Zentromers eines künstlichen Chromosoms nach einer bestimmten Anzahl an Zellteilungen würde so den direkten Vergleich mit der Sequenz der ursprünglich eingesetzten DNA erlauben. Die hier erarbeiteten Methoden zur Feinkartierung sind dabei unerlässlich für die erfolgreiche Reklonierung des charakterisierten Arrays aus Zellkulturzellen. Die Sequenzierung würde schließlich direkt die stattgefundenen Rekombinationen, deren Häufigkeiten, Richtungen und räumliche Verteilung aufdecken. Die daraus gewonnenen Erkenntnisse wären nicht nur für die Evolution der Zentromere von großer Bedeutung, sondern, so schlossen bereits Elder und Turner 1995:

„The prime biological significance of satellite DNAs is not whether they have a function in the genome, but rather that they reveal so much about the molecular evolution of genomes. [...] It is here that we will begin to understand just how much of genomic evolution is a consequence of natural selection and how much is due solely to events occurring as a result of the nature of the genetic material itself.“

F. Zusammenfassung

In den vergangenen Jahren wurden große Fortschritte bei der *de novo* Konstruktion von künstlichen Chromosomen als sichere, flexible, eukaryotische Vektoren erzielt. Dabei stellen die Zentromere den entscheidenden Faktor für die kontrollierte Weitergabe dieser Chromosomen in sich teilenden Zellen dar. Die optimalen Voraussetzungen für die Ausbildung eines Zentromers aus einem nackten DNA-Strang und für eine maximale mitotische Stabilität der *de novo* Chromosomen sind jedoch unbekannt, da weder die Struktur der α -Satelliten-Arrays im Bereich der natürlichen menschlichen Zentromere, noch die Sequenz-Funktions Relation dieser repetitiven DNA abschließend geklärt ist.

Ziel dieser Arbeit war es daher, fünf PACs (P1 artificial chromosomes) mit 120-160 kb an humaner α -Satelliten-DNA als Insert zu charakterisieren, um diese als Grundlage für die Konstruktion von künstlichen Chromosomen verwenden zu können und um Erkenntnisse über den Aufbau der Zentromer-DNA zu gewinnen.

In Vorarbeiten konnte gezeigt werden, dass die verwendeten α -Satelliten-PACs repräsentativ für den üblicherweise 1-4 Mb langen, X-chromosomalen Zentromer-Array eines einzelnen männlichen Individuums sind. Die Klonierung und Endsequenzierung von jeweils zehn 2 kb Higher order repeats aus den fünf PACs ergab eine sehr hohe Homogenität der α -Satelliten-DNA innerhalb der PACs von über 98,6%, sowie nur wenige vom typischen X-chromosomalen 12-mer abweichende Higher order repeats.

Die geringen Unterschiede zwischen den Sequenzen sind im Wesentlichen durch variable Basen verursacht, also Abweichungen von der α -Satelliten-Konsensussequenz, welche identisch in mehreren Higher order repeats auftreten. Diese variablen Basen stellen ein Abbild der Vorgänge dar, die zur „gemeinsamen Evolution“ der α -Satelliten-Repeats führen, welche durch eine höhere Homogenität von repetitiver DNA innerhalb einer Spezies als zwischen verwandten Spezies gekennzeichnet ist. Die Gesamtheit aller nicht-mendelischen Mechanismen, welche zu dieser „gemeinsamen Evolution“ beitragen, wird auch als „Molecular Drive“ bezeichnet, welcher zusätzlich zu Selektion und genetischer Drift evolutionäre Veränderungen in einer nicht-darwinistischen Weise verursacht.

Die Analyse der in dieser Arbeit sequenzierten variablen Basen ergab eine Verteilung, welche nicht mit dem gängigen Modell von Unequal crossing over-Ereignissen als für die gemeinsame Evolution der α -Satelliten-DNA entscheidenden Mechanismus zu vereinbaren ist. Daher wurde eine Methode etabliert, welche durch den Partialverdau von intakter PAC-

DNA und Hybridisierung mit einer radioaktiv markierten Vektor-Sonde die Kartierung von Restriktionsschnittstellen und somit auch von einigen der variablen Basen ermöglichte.

Nach den Erkenntnissen aus diesen Kartierungen, der Sequenzierung weiterer Higher order repeats und der Analyse der Fixierung von Varianten in der Bevölkerung können ein typischer Abschnitt eines homogenen α -Satelliten-Arrays und die maßgeblichen Prozesse, die seine Evolution vorantreiben, folgendermaßen beschrieben werden:

Die Higher order repeats aus den homogensten Regionen eines α -Satelliten-Arrays enthalten praktisch keine Einzelmutationen und weisen eine hohe Dichte an fixierten Varianten auf. Die Varianten verbreiten sich durch häufige Genkonversionen in die benachbarten Higher order repeats. Dabei breiten sich in einem Higher order repeat unmittelbar nebeneinander liegende, wie auch weiter entfernt liegende Nukleotidvarianten voneinander unabhängig aus, so dass die bei den Genkonversionen übertragenen Sequenzen vermutlich nur wenige Basenpaare lang sind. Weitere Verteilungsprozesse, wie Unequal crossing over-Ereignisse, hinterlassen in den homogensten Abschnitten der α -Satelliten-Arrays keine dauerhaften Spuren, so dass diese Vorgänge vermutlich mit einer geringeren Rate ablaufen als die Konversionen zwischen den benachbarten Repeats. Die Positionen der Varianten erscheinen zufällig und hängen nicht von der Primärsequenz oder von bekannten Proteinbindungsstellen wie der CENP-B Box ab. Die absolute Häufigkeit der Genkonversionen ist unklar. Es besteht die Möglichkeit, dass bereits zwischen den Zellen innerhalb eines Individuums Unterschiede auftreten, wodurch ein exaktes Sequenzassembly aus einer Klonbank unmöglich werden würde. Ein Nukleotid einer Variante benötigt weniger als 200.000 Jahre um sich in allen Higher order repeats einer Population durchzusetzen.

Die in dieser Arbeit charakterisierten α -Satelliten-PACs stehen nun für die *de novo* Konstruktion von künstlichen Chromosomen bereit. Während der PAC A6 einen mit 160 kb sehr langen, ununterbrochenen α -Satelliten-Array enthält, zeichnen sich die PACs A7, A8 und A10 vor allem durch ihre sehr hohe Homogenität aus. Der PAC A7 stellt durch seine charakteristischen *Dra-III* Cluster ferner den idealen Kandidaten für eine Reklonierung und Sequenzierung des Zentromers eines künstlichen Chromosoms dar. Dies eröffnet möglicherweise weitere Einblicke in den Turnover der Zentromersequenzen, was zu einem besseren Verständnis der allgemeinen Vererbung von Sequenzen beitragen könnte.

G. Literaturverzeichnis

Albertson DG, Thomson JN (1982). The kinetochores of *Caenorhabditis elegans*. *Chromosoma* **86**: 409-428

Alexandrov IA, Mitkevich SP, Yurov YB (1988). The phylogeny of human chromosome specific alpha satellites. *Chromosoma* **96**: 443-453

Alexandrov IA, Medvedev LI, Mashkova TD, Kisselev LL, Romanova LY, Yurov YB (1993). Definition of a new alpha satellite suprachromosomal family characterized by monomeric organization. *Nucleic Acids Res* **21**: 2209-2215

Alexandrov I, Kazakov A, Tumereva I, Shepelev V, Yurov Y (2001). Alpha-satellite DNA of primates: old and new families. *Chromosoma* **110**: 253-266

Alkan C, Eichler EE, Bailey JA, Sahinalp SC, Tuzun E (2004). The role of unequal crossover in alpha-satellite DNA evolution: a computational analysis. *J Comput Biol* **11**: 933-944

Alonso A, Mahmood R, Li S, Cheung F, Yoda K, Warburton PE (2003). Genomic microarray analysis reveals distinct locations for the CENP-A binding domains in three human chromosome 13q32 neocentromeres. *Hum Mol Genet* **12**: 2711-2721

Amor DJ, Choo KHA (2002). Neocentromeres: Role in human disease, evolution, and centromere study. *Am J Hum Genet* **71**: 695-714

Amor DJ, Bentley K, Ryan J, Perry J, Wong L, Slater H, Choo KH (2004). Human centromere repositioning "in progress". *Proc Natl Acad Sci USA* **101**: 6542-6547

Amstutz H, Munz P, Heyer WD, Leupold U, Kohli J (1985). Concerted evolution of tRNA genes: intergenic conversion among three unlinked serine tRNA genes in *S. pombe*. *Cell* **40**: 879-886

Archidiacono N, Antonacci R, Marzella R, Finelli P, Lonoce A, Rocchi M (1995). Comparative mapping of human alphoid sequences in great apes using fluorescence in situ hybridization. *Genomics* **25**: 477-484

Arnheim N, Krystal M, Schmickel R, Wilson G, Ryder O, Zimmer E (1980). Molecular evidence for genetic exchanges among ribosomal genes on non-homologous chromosomes in man and apes. *Proc Natl Acad Sci USA* **77**: 7323-7327

Auriche C, Carpani D, Conese M, Caci E, Zegarra-Moran O, Donini P, Ascenzioni F (2002). Functional human CFTR produced by a stable minichromosome. *EMBO Rep* **3**: 862-868

Ayabe F, Katoh M, Inoue T, Kouprina N, Larionov V, Oshimura M (2005). A novel expression system for genomic DNA loci using a human artificial chromosome vector with transformation-associated recombination cloning. *J Hum Genet* **50**: 592-599

Barnett MA, Buckle VJ, Evans EP, Porter AC, Rout D, Smith AG, Brown WR (1993). Telomere directed fragmentation of mammalian chromosomes. *Nucleic Acids Res* **21**: 27-36

- Barry AE, Howman EV, Cancilla MR, Saffery R, Choo KHA (2000a). Sequence analysis of an 80 kb human neocentromere. *Hum Mol Genet* **8**: 217–227
- Barry AE, Bateman M, Howman EV, Cancilla MR, Tainton KM, Irvine DV, Saffery R, Choo KH (2000b). The 10q25 neocentromere and its inactive progenitor have identical primary nucleotide sequence: further evidence for epigenetic modification. *Genome Res* **10**: 832-838
- Basu J, Stromberg G, Compitello G, Willard HF, Van Bokkelen G (2005). Rapid creation of BAC-based human artificial chromosome vectors by transposition with synthetic alpha-satellite arrays. *Nucleic Acids Res* **33**: 587-596
- Basu J, Willard HF (2005). Artificial and engineered chromosomes: non-integrating vectors for gene therapy. *Trends Mol Med* **11**: 251-258
- Bettencourt BR, Feder ME (2002). Rapid concerted evolution via gene conversion at the *Drosophila hsp70* genes. *J Mol Evol* **54**: 569-586
- Bollag RJ, Waldman AS, Liskay RM (1989). Homologous recombination in mammalian cells. *Annu Rev Genet* **23**: 199–225
- Britten RJ, Kohne DE (1968). Repeated sequences in DNA. Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms. *Science* **161**: 529-40
- Brown WRA, Mee PJ, Shen MH (2000). Artificial chromosomes: Ideal vectors? *Trends Biotechnol* **18**: 218-223
- Brutlag DL (1980). Molecular arrangement and evolution of heterochromatic DNA. *Annu Rev Genet* **14**: 121-44
- Burgoyne PS, Holland K, Stephens R (1991). Incidence of numerical chromosome anomalies in human pregnancy estimation from induced and spontaneous abortion data. *Hum Reprod* **6**: 555-565
- Cavalli-Sforza LL, Kidd JR, Kidd KK, Bucci C, Bowcock AM, Hewlett BS, Freidlaender JS (1986). DNA markers and genetic variation in the human species. *Cold Spring Harbor Symp Quant Biol* **51**: 411-417
- Charlesworth B, Langley CH, Stephan W (1986). The evolution of restricted recombination and the accumulation of repeated DNA sequences. *Genetics* **112**: 947-962
- Charlesworth B, Sniegowski P, Stephan W (1994). The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* **371**: 215–220
- Choo KH, Vissel B, Nagy A, Earle E, Kalitsis P (1991). A survey of the genomic distribution of α -satellite DNA on all the human chromosomes, and derivation of a new consensus sequence. *Nucleic Acids Res* **19**: 1179–1182
- Choo KH (1997a). Centromere DNA dynamics: Latent centromeres and neocentromere formation. *Am J Hum Genet* **61**: 1225–1233

- Choo KHA (1997b). The centromere. pp. 98–108. Oxford University press, Oxford, UK
- Choo KHA (1998). Turning on the centromere. *Nat Genet* **18**: 3-4
- Choo KHA (2000). Centromerization. *Trends Cell Biol* **10**: 182-188
- Choo KHA (2001). Engineering human chromosomes for gene therapy studies. *Trends Mol Med* **7**: 235-237
- Chu G, Vollrath D, Davis RW (1986). Separation of large DNA molecules by contour-clamped homogeneous electric fields. *Science* **234**: 1582-1585
- Clarke L (1990). Centromeres of budding and fission yeasts. *Trends in Genet* **6**: 150-154
- Cleveland DW, Mao Y, Sullivan KF (2003). Centromeres and Kinetochores: From Epigenetics to Mitotic Checkpoint Signaling. *Cell* **112**: 407–421
- Copenhaver GP, Nickel K, Kuromori T, Benito MI, Kaul S, Lin X, Bevan M, Murphy G, Harris B, Parnell LD, McCombie WR, Martienssen RA, Marra M, Preuss D (1999). Genetic definition and sequence analysis of Arabidopsis centromeres. *Science* **286**: 2468-2474
- Csink AK, Henikoff S (1998). Something from nothing: the evolution and utility of satellite repeats. *Trends Genet* **14**: 200-204
- Dewald GW, Boros SJ, Conroy MM, Dahl RJ, Spurbeck JL, Vitek HA (1979). A tdc(5;15)(p13;p11) chromosome showing variation for constriction in the centromeric region in a patient with cri-du-chat syndrome. *Cytogenet Cell Genet* **16**: 144-153
- Dobie KW, Hari KL, Maggert KA, Karpen GH (1999). Centromere proteins and chromosome inheritance: a complex affair. *Curr Opin Genet Dev* **9**: 206-217
- Dover GA, Linares AR, Bowen T, Hancock JM (1993). Detection and quantification of concerted evolution and molecular drive. *Methods Enzymol* **224**: 525-541
- Dover G (1993). Evolution of genetic redundancy for advanced players. *Curr Opin Genet Dev* **3**: 902-910
- Dover G (1982). Molecular drive: A cohesive mode of species evolution. *Nature* **299**: 111–117
- Dover G (2000). How genomic and developmental dynamics affect evolutionary processes. *BioEssays* **22**: 1153–1159
- Dover G (2002). Molecular drive. *Trends in Genet* **18**: 587-589
- Drenckhahn D, Zenker W (Herausgeber) (1994). Benninghoff, Anatomie. Makroskopische Anatomie, Embryologie und Histologie des Menschen. Urban und Schwarzenberg, München-Wien-Baltimore

Durphy SJ, Willard HF (1987). Molecular analysis of a polymorphic domain of α -satellite from the human X chromosome. *Am J Hum Genet* **41**: 391–401

Durphy SJ, Willard HF (1989). Patterns of intra- and interarray sequence variation in α -satellite from the human X chromosome: Evidence for short-range homogenization of tandemly repeated DNA sequences. *Genomics* **5**: 810–821

Durphy SJ, Willard HF (1990). Concerted evolution of primate α -satellite DNA. Evidence for an ancestral sequence shared by gorilla and human X chromosome α -satellite. *J Mol Biol* **216**: 555–566

du Sart D, Cancilla MR, Earle E, Mao JI, Saffery R, Tainton KM, Kalitsis P, Martyn J, Barry AE, Choo KH (1997). A functional neo-centromere formed through activation of a latent human centromere and consisting of non- α -satellite DNA. *Nat Genet* **16**: 144–153

Earnshaw WC, Rothfield N (1985). Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma. *Chromosoma*. **91**: 313-321

Earnshaw WC, Ratrie H, Stetten G (1989). Visualization of centromere proteins CENP-B and CENP-C on a stable dicentric chromosome in cytological spreads. *Chromosoma* **98**: 1-12

Ebersole TA, Ross A, Clark E, McGill N, Schindelbauer D, Cooke H, Grimes BR (2000). Mammalian artificial chromosome formation from circular alphoid input DNA does not require telomere repeats. *Hum Mol Gen* **9**: 1623-1631

Ebersole T, Okamoto Y, Noskov VN, Kouprina N, Kim JH, Leem SH, Barrett JC, Masumoto H, Larionov V (2005). Rapid generation of long synthetic tandem repeats and its application for analysis in human artificial chromosome formation. *Nucleic Acids Res* **33**: e130

Eichler EE (1999). Repetitive conundrums of centromere structure and function. *Hum Mol Genet* **8**: 151–155

Eichler EE, Clark RA, She X (2004). An assessment of the sequence gaps: unfinished business in a finished human genome. *Nat Rev Genet* **5**: 345-354

Ekwall K, Olsson T, Turner BM, Cranston G, Allshire RC (1997). Transient inhibition of histone deacetylation alters the structural and functional imprint at fission yeast centromeres. *Cell* **91**: 1021-1032

Elder JF, Turner BJ (1994). Concerted evolution at the population level: Pupfish HindIII satellite DNA sequences. *Proc Natl Acad Sci USA* **91**: 994-998

Elder Jr. JF, Turner BJ (1995). Concerted evolution of repetitive DNA sequences in eukaryotes. *Quart Rev Biol* **70**: 297–320

Elliot B, Richardson C, Winderbaum J, Nickoloff JA, Jasin M (1998). Gene conversion tracts from double-strand break repair in mammalian cells. *Mol Cell Biol* **18**: 93-101

- Farr CJ, Bayne RA, Kipling D, Mills W, Critcher R, Cooke HJ (1995). Generation of a human X-derived minichromosome using telomere-associated chromosome fragmentation. *EMBO J* **14**: 5444-5454
- Finelli P, Antonacci R, Marzella R, Lonoce A, Archidiacono N, Rocchi M (1996). Structural organization of multiple alphoid subsets coexisting on human chromosomes 1, 4, 5, 7, 9, 15, 18, and 19. *Genomics* **38**: 325-330
- Fisher AM, Al-Gazali L, Pramathan T, Quaife R, Cockwell AE, Barber JC, Earnshaw WC, Axelman J, Migeon BR, Tyler-Smith C (1997). Centromeric inactivation in a dicentric human Y;21 translocation chromosome. *Chromosoma* **106**: 199-206
- Flemming F (1880). Beitrag zur Kenntnis der Zelle und ihrer Lebenserscheinungen, Teil II. *Archiv Mikrosk Anat* **18**: 151-259
- Florida G, Zatterale A, Zuffardi O, Tyler-Smith C (2000). Mapping of a human centromere onto the DNA by topoisomerase II cleavage. *EMBO Rep* **1**: 489-493
- Fowler KJ, Hudson DF, Salamonsen LA, Edmondson SR, Earle E, Sibson MC, Choo KH (2000). Uterine dysfunction and genetic modifiers in centromere protein B-deficient mice. *Genome Res* **10**: 30-41
- Galtier N (2002). Gene conversion drives GC content evolution in mammalian histones. *Trends Genet* **19**: 65-68
- Gangloff S, Zou H, Rothstein R (1996). Gene conversion plays the major role in controlling the stability of large tandem repeats in yeast. *EMBO J* **15**: 1715-1725
- Gilbert DM. (2001) Making sense of eukaryotic DNA replication origins. *Science* **294**: 96-100
- Glover DJ, Lipps HJ, Jans DA (2005). Towards safe, non-viral therapeutic gene expression in humans. *Nat Rev Genet* **6**: 299-310
- Goldberg IG, Sawhney H, Pluta AF, Warburton PE, Earnshaw WC (1996). Surprising deficiency of CENP-B binding sites in African green monkey alpha-satellite DNA: implications for CENP-B function at centromeres. *Mol Cell Biol* **16**: 5156-5168
- Griesenbach U, Ferrari S, Geddes DM, Alton EW. (2002) Gene Therapy Progress and Prospects: Cystic fibrosis. *Gene Therapy* **9**: 1344-1350
- Grimes BR, Schindelbauer D, McGill NI, Ross A, Ebersole TA, Cooke HJ. (2001). Stable gene expression from a mammalian artificial chromosome. *EMBO-Reports* **21**: 910-914
- Grimes BR, Rhoades AA, Willard HF (2002a). α -satellite DNA and vector composition influence rates of human artificial chromosome formation. *Molecular Therapy* **5**: 798-805
- Grimes BR, Warburton PE, Farr CJ. (2002b). Chromosome engineering: prospects for gene therapy. *Gene Therapy* **9**: 713-718

Grimes BR, Babcock J, Rudd MK, Chadwick B, Willard HF (2004). Assembly and characterization of heterochromatin and euchromatin on human artificial chromosomes. *Genome Biol* **5**: R89

Grimes BR, Monaco ZL (2005). Artificial and engineered chromosomes: developments and prospects for gene therapy. *Chromosoma* **114**: 230-241

Haaf T, Mater AG, Wienberg J, Ward DC (1995). Presence and abundance of CENP-B box sequences in great ape subsets of primate-specific alpha-satellite DNA. *J Mol Evol* **41**: 487-491

Haaf T, Willard HF (1998). Orangutan alpha-satellite monomers are closely related to the human consensus sequence. *Mamm Genome* **9**: 440-447

Hacein-Bey-Abina S, von Kalle C, Schmidt M, McCormack MP, Wulffraat N, Leboulch P, Lim A, Osborne CS, Pawliuk R, Morillon E, Sorensen R, Forster A, Fraser P, Cohen JI, de Saint Basile G, Alexander I, Wintergerst U, Frebourg T, Aurias A, Stoppa-Lyonnet D, Romana S, Radford-Weiss I, Gross F, Valensi F, Delabesse E, Macintyre E, Sigaux F, Soulier J, Leiva LE, Wissler M, Prinz C, Rabbitts TH, Le Deist F, Fischer A, Cavazzana-Calvo M (2003). LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science* **302**: 415-419

Hanish JP, Yanowitz JL, de Lange T (1994). Stringent sequence requirements for the formation of human telomeres. *Proc Natl Acad Sci USA* **91**: 8861-8865

Harrington JJ, Van Bokkelen G, Mays RW, Gustashaw K, Willard HF (1997). Formation of de novo centromeres and construction of first-generation human artificial microchromosomes. *Nat Genet* **15**: 345-355

Hartwell LH, Kastan MB (1994). Cell cycle control and cancer. *Science* **266**: 1821-1828

Haustein UF (2002). Systemic sclerosis – scleroderma. *Dermatology online journal* **8**: 3

Heller R, Brown KE, Burgtorf C, Brown WR (1996). Mini-chromosomes derived from the human Y chromosome by telomere directed chromosome breakage. *Proc Natl Acad Sci USA* **93**: 7125-7130

Henikoff S (2002). Near the edge of a chromosome's "black hole". *Trends Genet* **18**: 165-167

Hennig W, Walker PMB (1970). Variations in the DNA from two rodent families (*Cricetidae* and *Muridae*). *Nature* **225**: 915-919

Hillis DM, Moritz C, Porter CA, Baker RJ (1991). Evidence for biased gene conversion in concerted evolution of ribosomal DNA. *Science* **251**: 308-310

Holliday R (1964). A mechanism for gene conversion in fungi. *Genet Res* **5**: 282-304

Hood L, Eichmann K, Lackland H, Krause RM, Ohms JJ (1970). Rabbit antibody light chains and gene evolution. *Nature* **228**: 1040-1044

- Horz W, Zachau HG (1977). Characterization of distinct segments in mouse satellite DNA by restriction nucleases. *Eur J Biochem* **73**: 383–392
- Hourcade D, Dressler D, Wolfson J (1973). The amplification of ribosomal RNA genes involving a rolling circle intermediate. *Proc Natl Acad Sci USA* **70**: 2926-2930
- Hudson DF, Fowler KJ, Earle E, Saffery R, Kalitsis P, Trowell H, Hill J, Wreford NG, de Kretser DM, Cancilla MR, Howman E, Hii L, Cutts SM, Irvine DV, Choo KH (1998). Centromere protein B null mice are mitotically and meiotically normal but have lower body and testis weights. *J Cell Biol* **141**: 309-319
- Ing PS, Smith SD (1983). Cytogenetic studies of a patient with mosaicism of isochromosome 13q and a dicentric (Y;13) translocation showing differential centromeric activity. *Clin Genet* **24**: 194-199
- International Human Genome Sequencing Consortium (2004). Finishing the euchromatic sequence of the human genome. *Nature* **431**: 931-945
- Ioannou PA, Amemiya CT, Garnes J, Kroisel PM, Shizuya H, Chen C, Batzer MA, de Jong PJ (1994). A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nat Genet* **6**: 84–89
- Ikeno M, Grimes BR, Okazaki T, Nakano M, Saitoh K, Hoshino H, McGill NI, Cooke H, Masumoto H (1998). Construction of YAC-based mammalian artificial chromosomes. *Nat Biotech* **16**: 431-439
- Ikeno M, Inagaki H, Nagata K, Morita M, Ichinose H, Okazaki T (2002). Generation of human artificial chromosomes expressing naturally controlled guanosine triphosphate cyclohydrolase I gene. *Genes to Cells* **7**: 1021–1032
- Irelan JT, Gutkin GI, Clarke L (2001). Functional redundancies, distinct localizations and interactions among three fission yeast homologs of Centromere Protein-B. *Genetics* **157**: 1191-1203
- Jabs EW, Goble CA, Cutting GR (1989). Macromolecular organization of human centromeric regions reveals high-frequency, polymorphic macro DNA repeats. *Proc Natl Acad Sci USA* **86**: 202-206
- Jacobs PA, Hassold TJ (1995). The origin of numerical chromosome abnormalities. *Adv Genet* **33**: 101-133
- Jeffreys AJ, Tamaki K, MacLeod A, Monckton DG, Neil DL, Armour JA (1994). Complex gene conversion events in germline mutation at human minisatellites. *Nature Genet* **6**: 136-145
- Jones KW, Corneo G (1971). Location of satellite and homogeneous DNA sequences on human chromosomes. *Nat New Biol* **233**: 268-271
- Johnson RD, Jasin M (2000). Sister chromatid gene conversion is a prominent double-strand break repair pathway in mammalian cells. *EMBO J* **19**: 3398–3407

Jorgensen AL, Laursen HB, Jones C, Bak AL (1992). Evolutionarily different alphoid repeat DNA on homologous chromosomes in human and chimpanzee. *PNAS* **89**: 3310-3314

Kapoor M, Montes de Oca Luna R, Liu G, Lozano G, Cummings C, Mancini M, Ouspenski I, Brinkley BR, May GS (1998). The cenpB gene is not essential in mice. *Chromosoma* **107**: 570-576

Karpen GH, Allshire RC (1997). The case for epigenetic effects on centromere identity and function. *Trends Genet* **13**: 489-496

Kass DH, Batzer MA, Deininger PL (1995). Gene conversion as a secondary mechanism of short interspersed element (SINE) evolution. *Mol Cell Biol* **15**: 19-25

Kawamura S, Saitou N, Ueda S (1992). Concerted evolution of the primate immunoglobulin alpha-gene through gene conversion. *J Biol Chem* **267**: 7359-7367

Kipling D, Mitchell AR, Masumoto H, Wilson HE, Nicol L, Cooke HJ (1995). CENP-B binds a novel centromeric sequence in the Asian mouse *Mus caroli*. *Mol Cell Biol* **15**: 4009-4020

Kipling D, Warburton PE (1997). Centromeres, CENP-B and *Tigger* too. *Trends Genet* **13**: 141-145

Kit S (1961). Equilibrium sedimentation in density gradients of DNA preparations from animal tissues. *J Mol Biol* **3**: 711-716

Klink D, Schindelbauer D, Laner A, Tucker T, Bebok Z, Schwiebert EM, Boyd AC, Scholte BJ (2004). Gene delivery systems – gene therapy vectors for cystic fibrosis. *J Cyst Fibros* **3 Suppl 2**: 203-12

Koch J (2000). Neocentromeres and α -satellite: a proposed structural code for functional human centromere DNA. *Hum Mol Genet* **9**: 149-154

Kohn DB, Sadelain M, Glorioso JC (2003). Occurrence of leukaemia following gene therapy of X-linked SCID. *Nature Reviews Cancer* **3**: 477-488

Kuroiwa Y, Yoshida H, Ohshima T, Shinohara T, Ohguma A, Kazuki Y, Oshimura M, Ishida I, Tomizuka K (2002). The use of chromosome-based vectors for animal transgenesis. *Gene Therapy* **9**: 708-712

Lam AL, Boivin CD, Bonney CF, Rudd MK, Sullivan BA (2006). Human centromeric chromatin is a dynamic chromosomal domain that can spread over noncentromeric DNA. *PNAS* **103**: 4186-4191

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al.; International Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome. *Nature* **409**: 860-921

Laner A, Goussard S, Ramalho AS, Schwarz T, Amaral MD, Courvalin P, Schindelbauer D, Grillot-Courvalin C (2005). Bacterial transfer of large functional genomic DNA into human cells. *Gene Therapy* **12**: 1559-1572

Lee C, Wevrick R, Fisher RB, Ferguson-Smith MA, Lin CC (1997). Human centromeric DNAs. *Hum Genet* **100**: 291-304

Lenoir A, Cournoyer B, Warwick S, Picard G, Deragon JM (1997). Evolution of SINE S1 retroposons in Cruciferae plant species. *Mol Biol Evol* **14**: 934-941

Levinson G, Gutman GA (1987). Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol Biol Evol* **4**: 203-221

Liao D, Pavelitz T, Kidd JR, Kidd KK, Weiner AM (1997). Concerted evolution of the tandemly repeated genes encoding human U2 snRNA (the RNU2 locus) involves rapid intrachromosomal homogenization and rare interchromosomal gene conversion. *EMBO J* **16**: 588-598

Liao D (2000). Gene conversion drives within genic sequences: concerted evolution of ribosomal RNA genes in bacteria and archaea. *J Mol Evol* **51**: 305-317

Linares AR, Bowen T, Dover GA (1994). Aspects of nonrandom turnover involved in the concerted evolution of intergenic spacers within the ribosomal DNA of *Drosophila melanogaster*. *J Mol Evol* **39**: 151-159

Lo AWI, Liao GCC, Rocchi M, Choo KHA (1999). Extreme reduction of chromosome-specific α -satellite array is unusually common in human chromosome 21. *Genome Res* **9**: 895-908

Lo AW, Craig JM, Saffery R, Kalitsis P, Irvine DV, Earle E, Magliano DJ, Choo KHA (2001a). A 330 kb CENP-A binding domain and altered replication timing at a human neocentromere. *EMBO J* **20**: 2087-2096

Lo AW, Magliano DJ, Sibson MC, Kalitsis P, Craig JM, Choo KHA (2001b). A novel chromatin immunoprecipitation and array (CIA) analysis identifies a 460-kb CENP-A-binding neocentromere DNA. *Genome Res* **11**: 448-457

Maeda N, Smithies O (1986). The evolution of multigene families: human haptoglobin genes. *Annu Rev Genet* **20**: 81-108

Maggert KA, Karpen GH (2000). Acquisition and metastability of centromere identity and function: sequence analysis of a human neocentromere. *Genome Res* **10**: 725-728

Mahtani MM, Willard HF (1990). Pulsed-field gel analysis of α -satellite DNA at the human X chromosome centromere: High-frequency polymorphisms and array size estimate. *Genomics* **7**: 607-613

Mahtani MM, Willard HF (1998). Physical and genetic mapping of the human X chromosome centromere: repression of recombination. *Genome Res* **8**: 100-110

Maiato H, DeLuca J, Salmon ED, Earnshaw WC (2004). The dynamic kinetochore-microtubule interface. *J Cell Sci* **117**: 5461-5477

Malik HS, Henikoff S (2002). Conflict begets complexity: the evolution of centromeres. *Cur Opin Genet Dev* **12**: 711–718

Manuelidis L (1976). Repeating restriction fragments of human DNA. *Nucleic Acids Res* **3**: 3063-3076

Marmur J, Rownd R, Schildkraut CL (1963). Denaturation and renaturation of deoxyribonucleic acid. *Progr Nucleic Acid Res* **1**: 231-300

Mashkova T, Oparina N, Alexandrov I, Zinovieva O, Marusina A, Yurov Y, Lacroix MH, Kisselev L (1998). Unequal cross-over is involved in human alpha satellite DNA rearrangements on a border of the satellite domain. *FEBS Letters* **441**: 451-457

Masumoto H, Masukata H, Muro Y, Nozaki N, Okazaki T (1989). A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite. *J Cell Biol* **109**: 1963–1973

Masumoto H, Nakano M, Ohzeki J (2004). The role of CENP-B and alpha-satellite DNA: de novo assembly and epigenetic maintenance of human centromeres. *Chromosome Res* **12**: 543-556

Mejía JE, Willmott A, Levy E, Earnshaw WC, Larin Z (2001). Functional complementation of a genetic deficiency with human artificial chromosomes. *Am J Hum Genet* **69**: 315-326

Mejía JE, Alazami A, Willmott A, Marschall P, Levy E, Earnshaw WC, Larin Z (2002). Efficiency of *de novo* centromere formation in human artificial chromosomes. *Genomics* **79**: 297-304

Meraldi P, McAinsh AD, Rheinbay E, Sorger PK (2006). Phylogenetic and structural analysis of centromeric DNA and kinetochore proteins. *Genome Biol* **7**: R23

Meselson M, Stahl FW, Vinograd J (1957). Equilibrium Sedimentation of Macromolecules in Density Gradients. *Proc Natl Acad Sci USA* **43**: 581-588

Mills W, Critcher R, Lee C, Farr CJ (1999). Generation of an approximately 2.4 Mb human X centromere-based minichromosome by targeted telomere-associated chromosome fragmentation in DT40. *Hum Mol Genet* **8**: 751-761

Mitchell AR, Jeppesen P, Nicol L, Morrison H, Kipling D (1996). Epigenetic control of mammalian centromere protein binding: does DNA methylation have a role? *J Cell Sci* **109**: 2199-2206

Moralli D, Simpson KM, Wade-Martins R, Monaco ZL (2006). A novel human artificial chromosome gene expression system using herpes simplex virus type 1 vectors. *EMBO Rep* **7**: 911-918

Moroi Y, Peebles C, Fritzler MJ, Steigerwald J, Tan EM (1980). Autoantibody to centromere (kinetochore) in scleroderma sera. *Proc Natl Acad Sci USA* **77**: 1627-1631

Muro Y, Masumoto H, Yoda K, Nozaki N, Ohashi M, Okazaki T (1992). Centromere protein B assembles human centromeric α -satellite DNA at the 17bp sequence, CENP-B box. *J Cell Biol* **116**: 585-596

Murphy TD, Karpen GH (1998). Centromeres take flight: α -satellite and the quest for the human centromere. *Cell* **93**: 317-320

Nagylaki T, Petes TD (1982). Intrachromosomal gene conversion and the maintenance of sequence homogeneity among repeated genes. *Genetics* **100**: 315-337

Nagylaki T (1984a). The evolution of multigene families under intrachromosomal gene conversion. *Genetics* **106**: 529-548

Nagylaki T (1984b). Evolution of multigene families under interchromosomal gene conversion. *Proc Natl Acad Sci USA* **81**: 3796-3800

Nakagawa H, Lee JK, Hurwitz J, Allshire RC, Nakayama J, Grewal SI, Tanaka K, Murakami Y (2002). Fission yeast CENP-B homologs nucleate centromeric heterochromatin by promoting heterochromatin-specific histone tail modifications. *Genes Dev* **16**: 1766-1778

Nakano M, Okamoto Y, Ohzeki J, Masumoto H (2003). Epigenetic assembly of centromeric chromatin at ectopic α -satellite sites on human chromosomes. *J Cell Sci* **116**: 4021-4034

Nakashima H, Nakano M, Ohnishi R, Hiraoka Y, Kaneda Y, Sugino A, Masumoto H (2005). Assembly of additional heterochromatin distinct from centromere-kinetochore chromatin is required for de novo formation of human artificial chromosome. *J Cell Sci* **118**: 5885-5898

Neil DL, Villasante A, Fisher RB, Vetrie D, Cox B, Tyler-Smith C (1990). Structural instability of human tandemly repeated DNA sequences cloned in yeast artificial chromosome vectors. *Nucleic Acids Res* **18**: 1421-1428

Nielsen L, Schmidt ER, Edström JE (1990). Subrepeats result from regional DNA sequence conservation in tandem repeats in *Chironomus* telomeres. *J Mol Biol* **216**: 577-584

Ohta T (1976). Simple model for treating evolution of multigene families. *Nature* **263**: 74-76

Ohta T, Dover GA (1983). Population genetics of multigene families that are dispersed into two or more chromosomes. *Proc Natl Acad Sci USA* **80**: 4079-4083

Ohzeki J, Nakano M, Okada T, Masumoto H (2002). CENP-B box is required for de novo centromere chromatin assembly on human alphoid DNA. *J Cell Biol* **159**: 765-775

Okamoto Y, Nakano M, Ohzeki J, Larionov V, Masumoto H (2007). A minimal CENP-A core is required for nucleation and maintenance of a functional human centromere. *EMBO J* **26**: 1279-1291

O'Keefe RT, Henderson SC, Spector DL (1992). Dynamic organization of DNA replication in mammalian cell nuclei: spatially and temporally defined replication of chromosome-specific alpha-satellite DNA sequences. *J Cell Biol* **116**: 1095-1110

Okumura K, Kiyama R, Oishi M (1987). Sequence analyses of extrachromosomal Sau3A and related family DNA: analysis of recombination in the excision event. *NAR* **15**: 7477-7489

Pardue ML, Gall JG (1970). Chromosomal localization of mouse satellite DNA. *Science* **168**: 1356-1358

Pavelitz T, Rusche L, Matera AG, Scharf JM, Weiner AM (1995). Concerted evolution of the tandem array encoding primate U2 snRNA occurs *in situ*, without changing the cytological context of the RNU2 locus. *EMBO J* **14**: 169-177

Perez-Castro AV, Shamanski FL, Meneses JJ, Lovato TL, Vogel KG, Moyzis RK, Pedersen R (1998). Centromeric protein B null mice are viable with no apparent abnormalities. *Dev Biol* **201**: 135-143

Pluta AF, Mackay AM, Ainsztein AM, Goldberg IG, Earnshaw WC (1995). The centromere: hub of chromosomal activities. *Science* **270**: 1591-1594

Politi V, Perini G, Trazzi S, Pliss A, Raska I, Earnshaw WC, Valle GD (2002). CENP-C binds the alpha-satellite DNA *in vivo* at specific centromere domains. *J Cell Sci* **115**: 2317-2327

Rivera H, Zuffardi O, Maraschio P, Caiulo A, Anichini C, Scarinci R, Vivarelli R (1989). Alternate centromere inactivation in a pseudodicentric (15;20)(pter;pter) associated with a progressive neurological disorder. *J Med Genet* **6**: 626-630

Roizès G (2006). Human centromeric alphoid domains are periodically homogenized so that they vary substantially between homologues. Mechanism and implications for centromere functioning. *Nucleic Acids Res* **34**: 1912-1924

Romanova LY, Deriagin GV, Mashkova TD, Tumeneva IG, Mushegian AR, Kisselev LL, Alexandrov IA (1996). Evidence for selection in evolution of alpha-satellite DNA: the central role of CENP-B/pJ alpha binding region. *J Mol Biol* **261**: 334-340

Rosandic M, Paar V, Basar I, Gluncic M, Pavin N, Pilas I (2006). CENP-B box and pJalpha sequence distribution in human alpha satellite higher-order repeats (HOR). *Chromosome Res* **14**: 735-753

Rosenberg M, Segal S, Kuff EL, Singer MF (1977). The nucleotide sequence of repetitive monkey DNA found in defective simian virus 40. *Cell* **11**: 845-857

Rosenecker J, Rudolph C (2001). Somatische Gentherapie der Mukoviszidose. *Monatsschr Kinderheilkunde* **149**: 270-280

Rudd MK, Mays RW, Schwartz S, Willard HF (2003). Human artificial chromosomes with alpha satellite-based *de novo* centromeres show increased frequency of nondisjunction and anaphase lag. *Mol Cell Biol* **23**: 7689-7697

Rudd MK, Willard HF (2004). Analysis of the centromeric regions of the human genome assembly. *Trends Genet* **20**: 529-533

- Rudd MK, Wray GA, Willard HF (2006). The evolutionary dynamics of alpha-satellite. *Genome Res* **16**: 88-96
- Saffery R, Irvine DV, Griffiths B, Kalitsis P, Wordeman L, Choo KHA (2000). Human centromeres and neocentromeres show identical distribution patterns of >20 functionally important kinetochore-associated proteins. *Hum Mol Gen* **9**: 175-185
- Saffery R, Wong LH, Irvine DV, Bateman MA, Griffiths B, Cutts SM, Cancilla MR, Cendron AC, Stafford AJ, Choo KHA (2001). Construction of neocentromere-based human minichromosomes by telomere-associated chromosomal truncation. *Proc Natl Acad Sci USA* **98**: 5705-5710
- Saffery R, Choo KHA (2002). Strategies for engineering human chromosomes with therapeutic potential. *J Gene Med* **4**: 5-13
- Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA (1988). Primer-directed enzymatic amplification of DNA with thermo-stable DNA polymerase. *Science* **348**: 252-254
- Saitoh H, Tomkiel J, Cooke CA, Ratrie H 3rd, Maurer M, Rothfield NF, Earnshaw WC (1992). CENP-C, an autoantigen in scleroderma, is a component of the human inner kinetochor plate. *Cell* **70**: 115-125
- Sanger F, Nicklen S, Coulson AR (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* **74**: 5463-5467
- Schindelhauer D, Cooke HJ (1997). Efficient combination of large DNA in vitro: In gel site specific recombination (IGSSR) of PAC fragments containing α -satellite DNA and the human HPRT gene locus. *Nucleic Acids Res* **25**: 2241-2243
- Schindelhauer D (1999). Construction of mammalian artificial chromosomes: prospects for defining an optimal centromere. *BioEssays* **21**: 76-83
- Schindelhauer D, Schwarz T (2002). Evidence for a fast, intrachromosomal conversion mechanism from mapping of nucleotide variants within a homogeneous α -satellite DNA array. *Gen Res* **12**: 1815-1826
- Schindelhauer D, Laner A (2002). Visible transient expression of EGFP requires intranuclear injection of large copy numbers. *Gene Ther* **9**: 727-730
- Schlötterer C, Tautz D (1994). Chromosomal homogeneity of *Drosophila* ribosomal DNA arrays suggests intrachromosomal exchanges drive concerted evolution. *Curr Biol* **4**: 777-783
- Schueler MG, Higgins AW, Rudd MK, Gustashaw K, Willard HF (2001). Genomic and genetic definition of a functional human centromere. *Science* **294**: 109-115
- Schueler MG, Sullivan BA (2006). Structural and functional dynamics of human centromeric chromatin. *Annu Rev Genomics Hum Genet* **7**: 301-313

- Shelby RD, Monier K, Sullivan KF (2000). Chromatin assembly at kinetochores is uncoupled from DNA replication. *J Cell Biol* **151**: 1113-1118
- Smith GP (1974). Unequal crossover and the evolution of multigene families. *Cold Spring Harb Symp Quant Biol* **38**: 507-513
- Smith GP (1976). Evolution of repeated DNA sequences by unequal crossover. *Science* **191**: 528-535
- Smith CL, Klcso SR, Cantor CR (1988). Pulsed-field gel electrophoresis and the technology of large DNA molecules. In: Davies KE (ed.): Genome analysis. IRL Press, Oxford, pp.41-72
- Somia N, Verma IM (2000). Gene therapy: trials and tribulations. *Nature Rev Genet* **1**: 91-99
- Southern EM (1970). Base sequence and evolution of guinea-pig alpha satellite DNA. *Nature* **227**: 794-798
- Southern EM (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* **98**: 503-517
- Spence JM, Fournier RE, Oshimura M, Regnier V, Farr CJ (2005). Topoisomerase II cleavage activity within the human D11Z1 and DXZ1 alpha-satellite arrays. *Chromosome Res* **13**: 637-648
- Spradling AC (1981). The organization and amplification of two chromosomal domains containing Drosophila chorion genes. *Cell* **27 (1 Pt 2)**: 193-201
- Srivastava IK, Liu MA (2003). Gene vaccines. *Ann Intern Med* **138**: 550-559
- Stephan W (1989). Tandem-repetitive noncoding DNA: forms and forces. *Mol Biol Evol* **6**: 198-212
- Stephan W, Cho S (1994). Possible role of natural selection in the formation of tandem-repetitive noncoding DNA. *Genetics* **136**: 333-341
- Sueoka N, Marmur J, Doty P (1959). Dependence of the density of deoxyribonucleic acids on guanine-cytosine content. *Nature* **183**: 1429-1431
- Sullivan B, Hechenberger M, Masri K (1994). Human CENP-A contains a histone H3 related histone fold domain that is required for targeting to the centromere. *J Cell Biol* **127**: 581-592
- Sullivan BA, Schwartz S (1995). Identification of centromeric antigens in dicentric Robertsonian translocations: CENP-C and CENP-E are necessary components of functional centromeres. *Hum Mol Genet* **5**: 2189-2198
- Sullivan BA, Willard HF (1998). Stable dicentric X chromosomes with two functional centromeres. *Nat Genet* **20**: 227-228
- Sullivan BA, Blower MD, Karpen GH (2001). Determining centromere identity: Cyclical stories and forking paths. *Nature Genet Rev* **2**: 584-596

Sullivan BA, Karpen GH (2001). Centromere identity in *Drosophila* is not determined in vivo by replication timing. *J Cell Biol* **154**: 683-690

Sullivan KF (2001). A solid foundation: functional specialization of centromeric chromatin. *Curr Opin Genet Dev* **11**: 182-188

Suzuki N, Nishii K, Okazaki T, Ikeno M (2006). Human artificial chromosomes constructed using the bottom-up strategy are stably maintained in mitosis and efficiently transmissible to progeny mice. *J Biol Chem* **281**: 26615-26623

Szostak JW, Orr-Weaver TL, Rothstein RJ, Stahl FW (1983). The double-strand-break repair model for recombination. *Cell* **33**: 25-35

Tartof KD (1975). Redundant genes. *Annu Rev Genet* **9**: 355-385

Ten Hagen KG, Gilbert DM, Willard HF, Cohen SN (1990). Replication timing of DNA sequences associated with human centromeres and telomeres. *Mol Cell Biol* **10**: 6348-6355

The *C. elegans* sequencing consortium (1998). Genome sequence of the nematode *C. elegans*: A platform for investigating biology. *Science* **282**: 2012-2018

Thompson-Steward D, Karpen GH, Spradling AC (1994). A transposable element can drive the concerted evolution of tandemly repetitive DNA. *Proc Natl Acad Sci USA* **91**: 9042-9046

Tsuduki T, Nakano M, Yasuoka N, Yamazaki S, Okada T, Okamoto Y, Masumoto H (2006). An artificially constructed de novo human chromosome behaves almost identically to its natural counterpart during metaphase and anaphase in living cells. *Mol Cell Biol* **26**: 7682-7695

Tyler-Smith C, Brown WRA (1987). Structure of the major block of alphoid satellite DNA on the human Y chromosome. *J Mol Biol* **195**: 457-470

Tyler-Smith C, Oakey RJ, Larin Z, Fisher RB, Crocker M, Affara NA, Ferguson-Smith MA, Muenke M, Zuffardi O, Jobling MA (1993). Localization of DNA sequences required for human centromere function through an analysis of rearranged Y-chromosomes. *Nature Genet* **5**: 368-375

Tyler-Smith C, Gimelli G, Giglio S, Florida G, Pandya A, Terzoli G, Warburton PE, Earnshaw WC, Zuffardi O (1999). Transmission of a fully functional human neocentromere through three generations. *Am J Hum Genet* **64**: 1440-1444

Tyler-Smith C, Florida G (2000). Many paths to the top of the mountain: Diverse evolutionary solutions to centromere structure. *Cell* **102**: 5-8

Vafa O, Sullivan KF (1997). Chromatin containing CENP-A and α -satellite DNA is a major component of the inner kinetochore plate. *Curr Biol* **7**: 897-900

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, et al. (2001). The sequence of the human genome. *Science* **291**: 1304-1351

Voet T, Schoenmakers E, Carpentier S, Labaere C, Marynen P (2003). Controlled transgene dosage and PAC-mediated transgenesis in mice using a chromosomal vector. *Genomics* **82**: 596-605

Voullaire LE, Slater HR, Petrovic V, Choo KHA (1993). A functional marker centromere with no detectable alpha-satellite, satellite III, or CENP-B protein: activation of a latent centromere? *Am J Hum Genet* **52**: 1153-1163

Walker PM (1971). Origin of satellite DNA. *Nature* **229**: 306-308

Walsh JB (1987). Persistence of tandem arrays: implications for satellite and simple sequence DNAs. *Genetics* **115**: 553-567

Wandall A (1994). A stable dicentric chromosome: both centromeres develop kinetochores and attach to the spindle in monocentric and dicentric configuration. *Chromosoma* **103**: 56-62

Warburton PE, Willard HF (1990). Genomic analysis of sequence variation in tandemly repeated DNA. Evidence for localized homogeneous sequence domains within arrays of α -satellite DNA. *J Mol Biol* **216**: 3-16

Warburton PE, Greig GM, Haaf T, Willard HF (1991). PCR amplification of chromosome-specific alpha satellite DNA: definition of centromeric STS markers and polymorphic analysis. *Genomics* **11**: 324-333

Warburton PE, Willard HF (1992). PCR amplification of tandemly repeated DNA: Analysis of intra- and interchromosomal sequence variation and homologous unequal crossing-over in human α -satellite DNA. *Nucleic Acids Res* **20**: 6033-6042

Warburton PE, Waye JS, Willard HF (1993). Nonrandom localization of recombination events in human alpha satellite repeat unit variants: implications for higher-order structural characteristics within centromeric heterochromatin. *Mol Cell Biol* **13**: 6520-6529

Warburton PE, Willard HF (1995). Interhomologue sequence variation of alpha satellite DNA from human chromosome 17: Evidence for concerted evolution along haplotypic lineages. *J Mol Evol* **41**: 1006-1015

Warburton PE, Dolled M, Mahmood R, Alonso A, Li S, Naritomi K, Tohma T, Nagai T, Hasegawa T, Ohashi H, Govaerts LC, Eussen BH, Van Hemel JO, Lozzio C, Schwartz S, Dowhanick-Morrisette JJ, Spinner NB, Rivera H, Crolla JA, Yu C, Warburton D (2000). Molecular cytogenetic analysis of eight inversion duplications of human chromosome 13q that each contain a neocentromere. *Am J Hum Genet* **66**: 1794-1806

Warburton PE (2001). Epigenetic analysis of kinetochore assembly on variant human centromeres. *Trends Genet* **17**: 243-247

Warburton PE (2004). Chromosomal dynamics of human neocentromere formation. *Chromosome Res* **12**: 617-626

Waring M, Britten RJ (1966). Nucleotide sequence repetition: a rapidly reassociating fraction of mouse DNA. *Science* **154**: 791-794

Waye JS, Willard HF (1985). Chromosome-specific α -satellite DNA: Nucleotide sequence analysis of the 2.0 kilobase pair repeat from the human X chromosome. *Nucleic Acids Res* **13**: 2731–2743

Waye JS, Willard HF (1986). Structure, organization, and sequence of α -satellite DNA from human chromosome 17: Evidence for evolution by unequal crossing-over and an ancestral pentamer repeat shared with the human X chromosome. *Mol Cell Biol* **6**: 3156–3165

Waye JS, Willard HF (1987). Nucleotide sequence heterogeneity of alpha satellite repetitive DNA: a survey of aliphoid sequences from different human chromosomes. *Nucleic Acids Res* **15**: 7549-7569

Waye JS, England SB, Willard HF (1987). Genomic organization of alpha satellite DNA on human chromosome 7: evidence for two distinct aliphoid domains on a single chromosome. *Mol Cell Biol* **7**: 349-356

Waye JS, Willard HF (1989). Concerted evolution of alpha satellite DNA: evidence for species specificity and a general lack of sequence conservation among aliphoid sequences of higher primates. *Chromosoma* **98**: 273-279

Wevrick R, Willard HF (1989). Long-range organization of tandem arrays of α -satellite DNA at the centromeres of human chromosomes: High-frequency array-length polymorphism and meiotic stability. *Proc Natl Acad Sci USA* **86**: 9394–9398

Willard HF, Smith KD, Sutherland J (1983). Isolation and characterization of a major tandem repeat family from the human X chromosome. *Nucleic Acids Res* **11**: 2017-2033

Willard HF (1985). Chromosome-specific organization of human alpha satellite DNA. *Am J Hum Genet* **37**: 524-532

Willard HF, Waye JS, Skolnick MH, Schwartz CE, Powers VE, England SB (1986). Detection of restriction fragment length polymorphisms at the centromeres of human chromosomes by using chromosome-specific alpha satellite DNA probes: implications for development of centromere-based genetic linkage maps. *Proc Natl Acad Sci USA* **83**: 5611-5615

Willard HF, Waye JS (1987). Chromosome-specific subsets of human alpha satellite DNA: analysis of sequence divergence within and between chromosomal subsets and evidence for an ancestral pentameric repeat. *J Mol Evol* **25**: 207-214

Willard HF, Greig GM, Powers VE, Waye JS (1987). Molecular organization and haplotype analysis of centromeric DNA from human chromosome 17: implications for linkage in neurofibromatosis. *Genomics* **1**: 368-373

Willard HF (1998). Centromeres: the missing link in the development of human artificial chromosomes. *Curr Opin Genet Dev* **8**: 219-225

Wong LH, Saffery R, Choo KHA (2002). Construction of neocentromere-based human minichromosomes for gene delivery and centromere studies. *Gene Therapy* **9**: 724-726

Wu JC, Manuelidis L (1980). Sequence definition and organization of a human repeated DNA. *J Mol Biol* **25**: 363-386

Yang JW, Pendon C, Yang J, Haywood N, Chand A, Brown WRA (2000). Human mini-chromosomes with minimal centromeres. *Hum Mol Genet* **9**: 1891-1902

Yang CH, Tomkiel J, Saitoh H, Johnson DH, Earnshaw WC (1996). Identification of overlapping DNA-binding and centromere-targeting domains in the human kinetochore protein CENP-C. *Mol Cell Biol* **16**: 3576–3586

Yang TP, Hansen SK, Oishi KK, Ryder OA, Hamkalo BA (1982). Characterization of a cloned repetitive DNA sequence concentrated on the human X chromosome. *Proc Natl Acad Sci USA* **79**: 6593–6597

Links:

Zusätzliche Informationen und Daten:

http://www.tobias-schwarz.net/genome_research.html

Edelsteinkärpfling *Cyprinodon variegatus*:

http://www.gma.org/fogm/Cyprinodon_variegatus.htm

Grüne Meerkatze *Cercopithecus aethiops*:

<http://www.tierenzyklopaedie.de/tiere/meerkatze.html>

Biographie Walther Flemming:

<http://scienceworld.wolfram.com/biography/FlemmingWalther.html>

PAC-Bank RPCI-I:

<http://bacpac.chori.org/humanmalepac.htm>

PAC-Vektor pCYPAC2:

<http://bacpac.chori.org/pcypac2.htm>

Zusätzliche Informationen:

NEBiolabs Katalog, Auflage 2002/03

H. Anhang

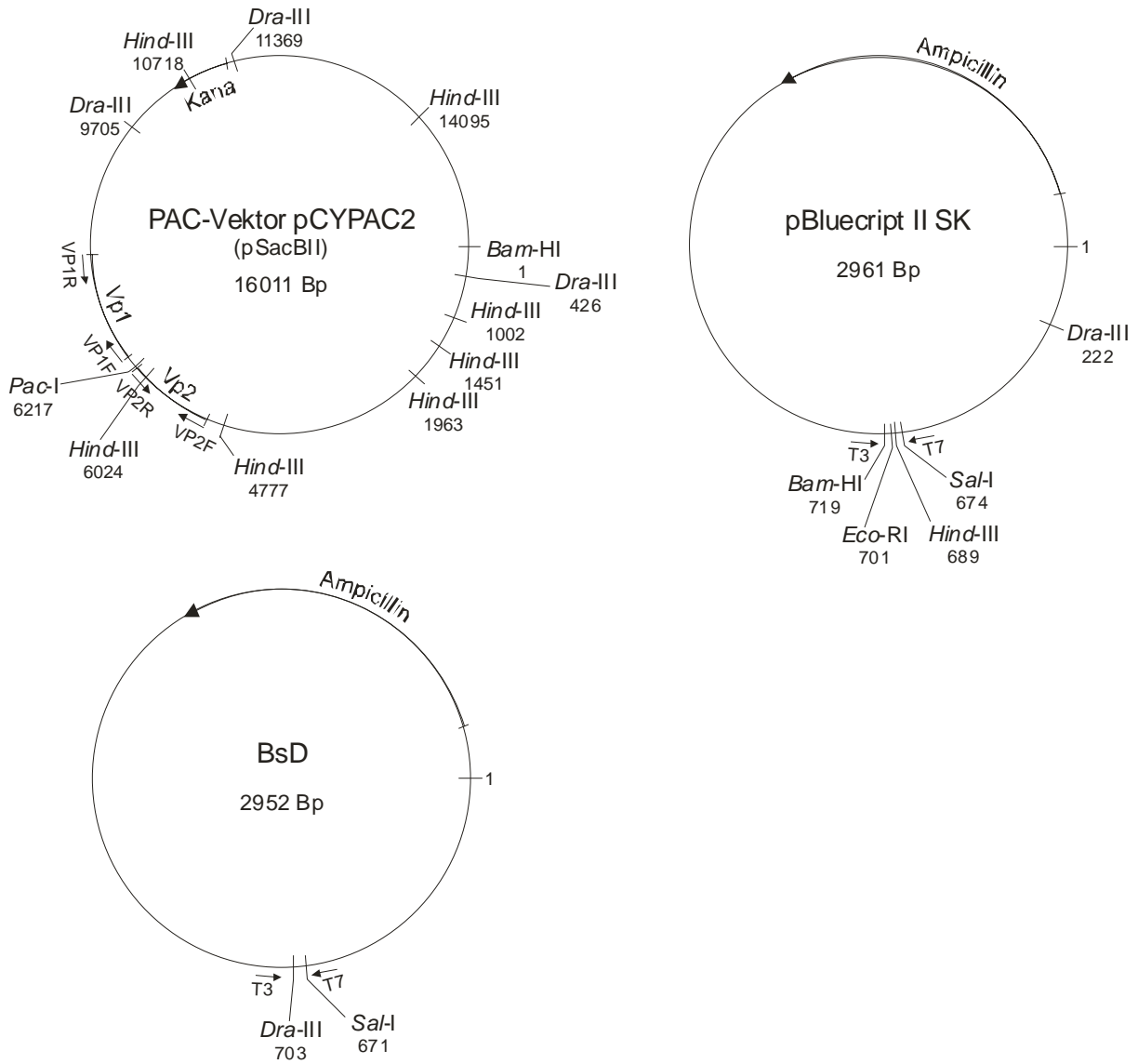
1. Am Aufbau der Zentromere beteiligte Proteine

Protein	Vorkommen	Vermutete Funktion	Knockout-Maus	Nur an aktiven Zentromeren	An Neozentromeren nachweisbar
3F3/2 Antigen	Transient, Mittlere Lamelle	Checkpoint Kontrolle; Tensions-Rezeptor		X	X
APC / Kar9	Äußere Lamelle	Verknüpfung mit Mitose-Spindel			
Arp1		Verknüpfung mit Mitose-Spindel und Chromosomenfindung		X	X
Aurora B	Heterochromatin				
Borealin / Dasra B	Heterochromatin				
BUB1	Transient, Äußere Lamelle	Checkpoint Kontrolle		X	X
BUB3	Transient, Äußere Lamelle	Checkpoint Kontrolle			X
BUBR1	Transient, Äußere Lamelle	Checkpoint Kontrolle		X	X
CDC20	Äußere Lamelle	Checkpoint Kontrolle			
CeKNL-1	Innere Lamelle				
CENP-A / CenH3	Konstitutiv, Innere Lamelle	Histon H3-Homolog; Struktureller Aufbau des Kinetochors; epigenetische Markierung eines Chromosomen-Abschnitts als Zentromer	Letal 6,5 Tage p.c.	X	X
CENP-B	Konstitutiv, Heterochromatin	Struktureller Aufbau des Kinetochors; bindet an CENP-B Box (17 bp Motiv)	In einer von drei Studien Abnahme von Körpergewicht und Hodengröße	O	O
CENP-C	Konstitutiv, Innere Lamelle	Struktureller Aufbau des Kinetochors	Letal 3,5 Tage p.c., Missegregation und Metaphasenarrest	X	X
CENP-E	Transient, Äußere Lamelle	Verknüpfung mit Mitose-Spindel; Chromosomenfindung; molekularer Motor		X	X
CENP-F	Transient, Äußere Lamelle			X	X
CENP-G	Konstitutiv, Innere Lamelle	Struktureller Aufbau des Kinetochors			X
CENP-H	Innere Lamelle			X	X
CENP-I / Mis6	Innere Lamelle				
Chl4R					
CLASPs	Äußere Lamelle				
CLIP-170 / Bik1	Transient, Äußere Lamelle	Verknüpfung mit Mitose-Spindel			X
Dynactin	Transient	Verknüpfung mit Mitose-Spindel; Chromosomenfindung		X	
Dynamitin				X	
Dynein	Transient, Äußere Lamelle	Verknüpfung mit Mitose-Spindel; Chromosomenfindung		X	X
EB1 / Bim1	Äußere Lamelle	Verknüpfung mit Mitose-Spindel			

ERK1	Transient	Checkpoint Kontrolle		X	X
FtaR					
HMG-1		Chromosomen-Kondensation			
HP1	Transient, Heterochromatin	Unterdrückung der Transkription			
ICIS	Heterochromatin				
INCENP	Transient, Heterochromatin	Chromosomen-Segregation; Cytokinese	Letal 2,5 Tage p.c.		X
LIS1	Äußere Lamelle				
M31	Konstitutiv ?				
M33	Transient				
MAD1	Transient, Äußere Lamelle	Checkpoint Kontrolle			
MAD2	Transient, Äußere Lamelle	Checkpoint Kontrolle		X	X
MCAK	Transient, Heterochromatin	Verknüpfung mit Mitose-Spindel; Übergang in Anaphase		X	X
Mcm21R					
Mis12 / Mtw1	Innere Lamelle	Chromosomen-Segregation			
MPS1	Äußere Lamelle				
Ndc80 / Hec1	Äußere Lamelle	Chromosomen-Segregation			
Nnf1R / PMF1					
Nsl1R / DC8 / DC31					
Nuf2	Äußere Lamelle				
p150		Verknüpfung mit Mitose-Spindel; Chromosomenfindung		X	X
p55CDC	Transient	Anaphase promoting complex		X	X
PARP1		Protein modifier		X	X
PcG	Transient				
Plk	Äußere Lamelle				
PP1	Äußere Lamelle				
RanBP2	Äußere Lamelle				
RanGAP1	Äußere Lamelle				
RAD21	Heterochromatin	Kohäsion der Schwesterchromatiden			
Rod	Äußere Lamelle	Verknüpfung mit Mitose-Spindel; Checkpoint Kontrolle			
SCC1	Heterochromatin	Kohäsion der Schwesterchromatiden			
Sim4R / Solt					
Spc24	Äußere Lamelle				
Spc25	Äußere Lamelle				
Survivin	Heterochromatin				
SUV39H				X	X
SUVAR39H1	Transient, Heterochromatin	Unterdrückung der Transkription			
TD60				X	
Topoisomerase II	Konstitutiv ?	Chromosomen-Kondensation und Schwester-Chromatid Trennung		X	X
Tsg24	Transient	Anaphase promoting complex		X	X
Zw10	Transient, Äußere Lamelle	Verknüpfung mit Mitose-Spindel; Checkpoint Kontrolle		X	X
Zwilch		Verknüpfung mit Mitose-Spindel			
Zwint1	Äußere Lamelle			X	X

Nach Pluta et al. 1995, Dobie et al. 1999, Saffery et al. 2000, Warburton 2001, Maiato et al. 2004, Meraldi et al. 2006 u.a.

2. Karten der verwendeten Vektoren



3. Auflistung der klonierten Higher order repeats und Einzelsequenzen mit EMBL/Genbank Accession numbers

3.1. Endsequenzen der in die Analyse einbezogenen Higher order repeats

Ursprung	Subklon	Subklon Laborname	Primer linke Sequenz	EMBL/Genbank Accession number	Anzahl sequenzierter Nukleotide	Primer rechte Sequenz	EMBL/Genbank Accession number	Anzahl sequenzierter Nukleotide	Primer bei zusätzlicher Sequenzierung	Anzahl sequenzierter Nukleotide	Anzahl überlappend sequenzierter Nukleotide	Insgesamt analysierter Sequenzabschnitt
A6	A6BamHI-1	A6_2_4	T7	AJ509874	535	T3	AJ509938	534				1069
	A6BamHI-2	A6_2_10	T3	AJ509875	535	T7	AJ509939	534				1069
	A6BamHI-3	A6_2_12	T3	AJ509876	535	T7	AJ509940	534				1069
	A6BamHI-4	A6_2_13	T3	AJ509877	535	T7	AJ509941	534	X-4A	377	342	1069
	A6BamHI-5	A6_2_17	T3	AJ509878	535	T7	AJ509942	534	X-4A	377	342	1069
	A6BamHI-6	A6_2_18	T3	AJ509879	535	T7	AJ509943	534	X-4A	377	342	1069
	A6BamHI-7	A6_2_22	T3	AJ509880	535	T7	AJ509944	534	X-3A	394	359	1069
	A6BamHI-8	A6_2_24	T7	AJ509881	477	T3	AJ509945	534	X-4A	377	342	1011
	Summe:											8494
A7	A7BamHI-1	A7_4	T3	AJ509882	535	T7	AJ509946	534	X-3A	369	293	1069
	A7BamHI-2	A7_14	T3	AJ509883	535	T7	AJ509947	534				1069
	A7BamHI-3	A7_17	T3	AJ509884	535	T7	AJ509948	534				1069
	A7BamHI-4	A7_1	T7	AJ509885	400	T3	AJ509949	534				934
	A7BamHI-5	A7_5	T7	AJ509886	535	T3	AJ509950	534	X-3A	269	22	1069
	A7BamHI-6	A7_6	T7	AJ509887	535	T3	AJ509951	534	X-3A	231	99	1069
	A7BamHI-7	A7_8	T7	AJ509888	532	T3	AJ509952	534	X-4A	377	190	1066
	A7BamHI-8	A7_15	T3	AJ509889	535	T7	AJ509953	534	X-4A	377	342	1069
	A7BamHI-9	A7_16	T7	AJ509890	532	T3	AJ509954	534				1066
Summe:											9480	
A8	A8BamHI-1	A8_2	T7	AJ509903	535	T3	AJ509967	534	X-3A	355	285	1069
	A8BamHI-2	A8_3	T7	AJ509904	535	T3	AJ509968	534	X-3A	464	383	1069
	A8BamHI-3	A8_4	T7	AJ509905	535	T3	AJ509969	534				1069
	A8BamHI-4	A8_6	T3	AJ509906	535	T7	AJ509970	534				1069
	A8BamHI-5	A8_7	T7	AJ509907	535	T3	AJ509971	534				1069
	A8BamHI-6	A8_8	T3	AJ509908	535	T7	AJ509972	534				1069
	A8BamHI-7	A8_10	T7	AJ509909	535	T3	AJ509973	534				1069
	A8BamHI-8	A8_13	T7	AJ509910	535	T3	AJ509974	534				1069
	A8BamHI-9	A8_14	T7	AJ509911	532	T3	AJ509975	534				1066
Summe:											9618	
A10	A10BamHI-1	A10_2	T3	AJ509912	535	T7	AJ509976	534				1069
	A10BamHI-2	A10_3	T7	AJ509913	532	T3	AJ509977	534				1066
	A10BamHI-3	A10_4	T7	AJ509914	532	T3	AJ509978	534				1066
	A10BamHI-4	A10_5	T7	AJ509915	532	T3	AJ509979	534				1066
	A10BamHI-5	A10_11	T7	AJ509916	535	T3a	AJ509980	534	T3	465	465	1069
	A10BamHI-6	A10_12	T7	AJ509917	535	T3	AJ509981	534	X-3A	328	196	1069
	A10BamHI-7	A10_8	T3	AJ509918	535	T7	AJ509982	534				1069
	A10BamHI-8	A10_9	T7	AJ509919	535	T3	AJ509983	534				1069
	A10BamHI-9	A10_10	T3	AJ509920	535	T7	AJ509984	534				1069

	A10BamHI-10	A10_1	T7	AJ509921	535	T3	AJ509985	534				1069
	Summe:											10681
B11	B11BamHI-1	B11_2_4	T7	AJ509922	535	T3a	AJ509986	534	T3	298	298	1069
	B11BamHI-2	B11_3_1	T7	AJ509923	535	T3	AJ509987	534				1069
	B11BamHI-3	B11_3_2	T7	AJ509929	535	T3	AJ509993	534				1069
	B11BamHI-4	B11_3_4	T3	AJ509924	535	T7	AJ509988	534				1069
	B11BamHI-5	B11_3_5	T3	AJ509925	509	T7	AJ509989	534				1043
	B11BamHI-6	B11_3_7	T3a	AJ509930	535	T7	AJ509994	534	T3	344	344	1069
	B11BamHI-7	B11_3_8	T3	AJ509931	535	T7	AJ509995	534				1069
	B11BamHI-8	B11_3_10	T7	AJ509926	535	T3	AJ509990	534				1069
	B11BamHI-9	B11_3_16	T3	AJ509927	535	T7	AJ509991	534				1069
	B11BamHI-10	B11_3_17	T7	AJ509928	535	T3	AJ509992	534				1069
	Summe:											10664
B11	B11HindIII-7	BH_7	T7	AJ509932	337	T3	AJ509996	414				751
	B11HindIII-8	BH_8	T7	AJ509933	504	T3	AJ509997	565				1069
	B11HindIII-9	BH_9	T7	AJ509934	504	T3	AJ509998	565				1069
	B11HindIII-15	BH_15	T3	AJ509935	504	T7	AJ509999	291				795
	B11HindIII-17	BH_17	T7	AJ509936	504	T3	AJ510000	565				1069
	B11HindIII-18	BH_18	T7	AJ509937	504	T3	AJ510001	565				1069
	Summe:											5822
A7	A7DraIII-1	DraB_1	T3	AJ510002	609	T7	AJ510017	549				1158
	A7DraIII-2	DraB_3	T3	AJ510003	609	T7	AJ510018	549				1158
	A7DraIII-3	DraB_4	T3	AJ510004	513	T7	AJ510019	549				1062
	A7DraIII-4	DraB_6	T3	AJ510005	609	T7	AJ510020	549				1158
	A7DraIII-5	DraB_8	T3	AJ510006	609	T7	AJ510021	549				1158
	A7DraIII-6	DraB_12	T3	AJ510007	609	T7	AJ510022	549				1158
	A7DraIII-7	DraB_14	T3	AJ510008	608	T7	AJ510023	549				1157
	A7DraIII-8	DraB_15	T3	AJ510009	393	T7	AJ510024	500				893
	A7DraIII-9	DraB_33	T3	AJ510010	609	T7	AJ510025	549				1158
	A7DraIII-10	DraB_37	T3	AJ510011	609	T7	AJ510026	549				1158
	A7DraIII-11	DraB_38	T3	AJ510012	609	T7	AJ510027	549				1158
	A7DraIII-12	DraB_40	T3	AJ510013	539	T7	AJ510028	549				1088
	A7DraIII-13	DraB_41	T3	AJ510014	609	T7	AJ510029	549				1158
	A7DraIII-14	DraB_43	T3	AJ510015	609	T7	AJ510030	549				1158
	A7DraIII-15	DraB_44	T3	AJ510016	609	T7a	AJ510031	297	T7	153	153	906
	Summe:											16686
A7	A7nonDraIII-1	A7_oD_4	T3	AJ509891	5 35	T7	AJ509955	534				1069
	A7nonDraIII-2	A7_oD_5	T3	AJ509892	535	T7	AJ509956	534				1069
	A7nonDraIII-3	A7_oD_9	T3	AJ509893	535	T7	AJ509957	534				1069
	A7nonDraIII-4	A7_oD_10	T7	AJ509894	535	T3	AJ509958	534				1069
	A7nonDraIII-5	A7_oD_18	T7	AJ509895	535	T3	AJ509959	534				1069
	A7nonDraIII-6	A7_oD_19	T3	AJ509896	535	T7	AJ509960	534				1069
	A7nonDraIII-7	A7_oD_23	T7	AJ509897	535	T3	AJ509961	534				1069
	A7nonDraIII-8	A7_oD_25	T7	AJ509898	535	T3	AJ509962	534				1069
	A7nonDraIII-9	A7_oD_27	T7	AJ509899	535	T3	AJ509963	534				1069
	A7nonDraIII-10	A7_oD_28	T3	AJ509900	535	T7	AJ509964	534				1069
	A7nonDraIII-11	A7_oD_29	T3	AJ509901	535	T7	AJ509965	534				1069
	A7nonDraIII-12	A7_oD_33	T3	AJ509902	535	T7	AJ509966	534				1069
	Summe:											12828

Individ 1	Individ1-1		SP6	AJ509829	469	T7	AJ509829	489			469	489
	Individ1-2		T7	AJ509830	313	SP6	AJ509830	444			268	489
	Individ1-3		SP6	AJ509831	416	T7	AJ509831	489			416	489
	Individ1-4		T7	AJ509832	489	SP6	AJ509832	292			292	489
	Individ1-5		SP6	AJ509833	472	T7	AJ509833	489			472	489
	Individ1-6		SP6	AJ509834	271	T7	AJ509834	489			271	489
	Individ1-7		SP6	AJ509835	248	T7	AJ509835	489			248	489
	Individ1-9		T7	AJ509836	363	SP6	AJ509836	165			39	489
	Individ1-10		SP6	AJ509837	416	T7	AJ509837	489			416	489
Individ 2	Individ2-7		SP6	AJ509838	354	T7	AJ509838	400			265	489
	Individ2-8		T7	AJ509839	489	SP6	AJ509839	332			332	489
Individ 3	Individ3-7		T7	AJ509840	489	SP6	AJ509840	489	SP6a	489	489	489
Individ 4	Individ4-10		T7	AJ509841	363	SP6	AJ509841	283			157	489
	Summe:										6357	
Gor 014	Gor014-1		T7	AJ509815	271	SP6	AJ509815	329			78	522
	Gor014-3		SP6	AJ509822	396	T7	AJ509822	522			396	522
	Gor014-4		T7	AJ509816	304	SP6	AJ509816	450			232	522
	Gor014-6		T7	AJ509817	522	SP6	AJ509817	267			267	522
	Gor014-7		T7	AJ509818	294	SP6	AJ509818	410			182	522
	Gor014-8		SP6	AJ509819	173	T7	AJ509819	501			152	522
Gor 165	Gor165-2		SP6	AJ509823	387	T7	AJ509823	293			158	522
	Gor165-4		T7	AJ509820	345	SP6	AJ509820	498			321	522
	Gor165-7		T7	AJ509821	522	SP6	AJ509821	501			501	522
Doppelstrangsequenzierte Nukleotide:										6421		
Summe:										4698		
Insgesamt sequenzierte Nukleotide:										95328		

3.2. PAC-Endsequenzen

Ursprung	Primer linke Sequenz	EMBL/Genbank Accession number	Anzahl sequenzierter Nukleotide	Primer rechte Sequenz	EMBL/Genbank Accession number	Anzahl sequenzierter Nukleotide	Insgesamt analysierter Sequenzabschnitt
A6	T7	AJ509845	482	SP6	AJ509844	360	842
A7	SP6	AJ509846	461	T7	AJ509847	475	936
A8	T7	AJ509849	321	SP6	AJ509848	492	813
A10	SP6	AJ509850	482	T7	AJ509851	133	615
B11	SP6	AJ509842	339	T7	AJ509843	368	707
B11.60	P86	AJ509852	228				228
Summe:							4141

3.3. Nicht in die Analyse einbezogene Klone

Ursprung	Subklon Laborname	Primer linke Sequenz	Anzahl sequenzierter Nukleotide	Primer rechte Sequenz	Anzahl sequenzierter Nukleotide	Primer bei zusätzlicher Sequenzierung	Anzahl sequenzierter Nukleotide	Anzahl überlappend sequenzierter Nukleotide	Insgesamt sequenzierte Nukleotide	Identisch mit Subklon
A6	A6_2_21	T3	535	T7	534	X-3A	357	203	1069	A6BamHI-3
	A6_2_23	T3	477	T7	534	X-4A	378	345	1011	A6BamHI-4
A7	A7_7	T7	535	T3	534	X-3A	397	214	1069	A7BamHI-1
A8	A8_9	T7	535	T3	534				1069	A8BamHI-2
B11	BH_1	T7	453	T3	397				850	B11HindIII-8
	BH_2	T3	504	T7	468				972	B11HindIII-17
	BH_3	T3	504	T7	469				973	B11HindIII-17
	BH_10	T7	288						288	Schlechte Sequenzqualität
	BH_11	T3	504	T7	565				1069	B11BamHI-6
	BH_12	T7	504	T3	517				1021	B11BamHI-6
	BH_14	T3	504	T7	565				1069	B11BamHI-6
	BH_16	T7	504	T3	565				1069	B11HindIII-9
	BH_19	T7	248	T3	468				716	B11HindIII-17
	BH_21	T3	447	T7a	333	T7	191	191	780	B11HindIII-17
	BH_22	T7	252	T3	536				788	B11HindIII-9
BH_23	T7	504	T3	565				1069	B11HindIII-17	
BH_24	T3	504	T7	333				837	B11BamHI-6	
A7-oD	A7_oD_11	T7	535	T3	534				1069	A7nonDraIII-2
	A7_oD_30	T3	535	T7	534				1069	A7BamHI-9
	A7_oD_35	T7	535	T3	534				1069	A7nonDraIII-9
DraB	DraB_2	T3	487	T7	549				1036	A7DraIII-1
	DraB_11	T3	397	T7	549				946	A7DraIII-10
	DraB_13	T3	411	T7	148				559	Schlechte Sequenzqualität
	DraB_16	T3	374	T7	549				923	A7DraIII-3
	DraB_35	T3	609	T7	481				1090	A7DraIII-7
Gor 014	Gor014-9	T7	522	SP6	508			508	522	Gor014-7
Gor 165	Gor165-1			SP6	326				326	Unbekannte Sequenz
	Gor165-3	SP6	207						207	Unbekannte Sequenz
	Gor165-5	T7	500	SP6	166			144	522	Gor014-7
	Gor165-8			SP6	105				105	Divergierte α -Satelliten-Seq.
	Gor165-9			SP6	359				359	Divergierte α -Satelliten-Seq.
	Gor165-10	T7	522						522	Divergierte α -Satelliten-Seq.
Nicht berücksichtigt aufgrund vermutlich doppelter Klonierung:									23677	

Publikationen

Schindelhauer D, Schwarz T (2002). Evidence for a fast, intra-chromosomal conversion mechanism from mapping of nucleotide variants within a homogeneous alpha satellite DNA array. *Genome Res* **12**: 1815-1826

Laner A, Schwarz T, Christan S, Schindelhauer D (2004). Suitability of a CMV/EGFP cassette to monitor stable expression from human artificial chromosomes but not transient transfer in the cells forming viable clones. *Cytogenet Genome Res* **107**: 9-13

Laner A, Goussard S, Ramalho AS, Schwarz T, Amaral MD, Courvalin P, Schindelhauer D, Grillot-Courvalin C (2005). Bacterial transfer of large functional genomic DNA into human cells. *Gene Ther* **12**: 1559-1572

Kraner S, Laner A, Schwarz T, Cengizeroglu A, Wagner S, Christan S, Schindelhauer D. Efficient *de novo* centromere formation and maintenance of endogenous centromeres on the dimeric but not the adjacent 13-meric alpha satellite DNA of chromosomes 5 and 19.
Zur Publikation eingereicht bei Am J Hum Genet

Schwarz T, Schindelhauer D. Functional centromere sequences in humans - members of all modern suprachromosomal alpha satellite families efficiently seed centromeres *de novo*.
In Vorbereitung

Poster

Schwarz T, Schindelhauer D (14.-18. Oktober 2000). Tracing the concerted evolution of the homogeneous centromeric tandem repeat array of the human X chromosome. EMBO workshop on centromeres, kinetochores and spindle interactions, Heidelberg (org. Gary Karpen, Tony Hyman, Robin Allshire)

Schindelhauer D, Schwarz T (13.-17. Juni 2002). Evidence for a fast, intrachromosomal conversion mechanism from mapping of nucleotide variants within a homogeneous alpha satellite DNA array. Molecular Evolution, Sorrento (org. G. Bernardi)

Danksagung

Mein besonderer Dank gilt Herrn Dr. Dirk Schindelhauer für die Überlassung des Themas, die Einführung in die molekularbiologischen Techniken, die hervorragende Betreuung über die gesamte Zeit, die intensive Durchsicht des Manuskripts und vor allem für die stets angenehme Atmosphäre.

Bei Herrn Prof. Dr. Jan Murken bedanke ich mich für die Übernahme der offiziellen Betreuung und der damit verbundenen Aufgaben.

Mein Dank gilt auch Susie Christan, Sonja Cattani und Andreas Laner, welche die Arbeit in der AG Schindelhauer so angenehm machten und sich stets bereit zeigten, mir „Teilzeit-Wissenschaftler“ unter die Arme zu greifen.

Ebenso herzlich bedanken möchte ich mich bei Heide Hellebrand und Helene Achatz für die geduldige Unterstützung und Hilfsbereitschaft, aber auch bei allen anderen Mitarbeitern des Humangenetischen Instituts (Alfons, Konstanze, Ingrid, Juliane, Joseph, Claus, Celia, Marco, Steffi, Anne, Moni und allen anderen) für die freundschaftliche Zusammenarbeit.

Harald Brunnhofer, Stephan Haug sowie meinem Bruder Hans-Jörg danke ich für die Unterstützung bei den statistischen und stochastischen Analysen.

Brettmeister M, Sucker DJ und MC „The Mighty“ Güno haben es mir ermöglicht nach arbeitsreichen Tagen den Kopf wieder frei zu bekommen und auch in schwierigen Zeiten neue Motivation zu finden, vielen Dank hierfür.

Nicht zuletzt danke ich meinen Eltern, ohne deren Unterstützung diese Arbeit nicht möglich gewesen wäre.

Lebenslauf

Persönliche Angaben:

- Name: Tobias Schwarz
- Geboren: 14.09.1977 in Dachau
- Familienstand: Ledig, keine Kinder
- Staatsangehörigkeit: Deutsch

Schulbildung:

- 1983 - 1987 Grundschule Dachau
- 1987 - 1996 Josef-Effner-Gymnasium-Dachau
- Juni 1996 Allgemeine Hochschulreife

Zivildienst:

- Sept. 1996 - Sept. 1997 Zivildienst auf einer unfallchirurgischen Station an den Kreiskliniken Dachau-Indersdorf

Studium:

- Nov. 1997 - Okt. 2003 Studium der Humanmedizin an der Ludwig-Maximilians-Universität München
- Okt. 2003 Abschluß mit dem dritten Abschnitt der Ärztlichen Prüfung

Berufstätigkeit:

- Jan. 2004 - Sept. 2004 AiP an der Universitätskinderklinik Würzburg
- Seit Okt. 2004 Stipendiat des MD/PhD-Programms der Universität Würzburg
- Okt. 2004 - Juli 2005 Aufbaustudium am Biozentrum der Universität Würzburg
- Seit Aug. 2005 Biologische Promotion am Institut für molekulare Infektionsbiologie der Universität Würzburg