

3.2 Segregationsanalyse von Protein-Polymorphismen – Ergebnisse aus 2D-Mustern

Die Genotypisierung der B1-Mäuse anhand von Proteinpolymorphismen wurde bereits in der AG Klose durchgeführt. Der Vergleich der Proteinmuster des Gehirns von B6 und MSP ergab über 1324 polymorphe Proteinspots. Mehr als die Hälfte dieser Spots zeigten ein auswertbares Segregationsmuster in den B1-Tieren und konnten im Folgenden genetisch kartiert werden. Spots, die auf dem Gel in serienartigen Gruppen lagen und gleiches Vererbungsmuster zeigten, wurden aus praktischen Gründen unter dem gleichen Varianten-Namen zusammengefasst. Die Zahl der zu kartierenden Datensätze reduzierte sich damit auf 441. Tab.3.2.a) zeigt eine Übersicht über diese von Klose erstellten Ergebnisse (Gauss et al. 1999).

Tab.3.2.a): Die Analyse der Polymorphismen zwischen *Mus musculus* und *Mus spretus* (J. Klose in Gauss et al. 1999)

Zahl der analysierten Proteinspots der 2-DE Muster von <i>Mus musculus</i> und <i>Mus spretus</i>	8767 Spots (100%)
Zahl der polymorphen Proteinspots zwischen <i>Mus musculus</i> und <i>Mus spretus</i>	1324 Spots (15,1 %)
Proteinspots, die Segregation in den B1-Tieren zeigten und zur genetischen Kartierung eingesetzt wurden	744 Spots (8,5 %)
Spotzahl reduziert auf Varianten	441 Varianten

Wie in Material und Methoden dargestellt, wurden die polymorphen Proteinspots in verschiedene Variationstypen unterteilt. Ferner wurden sie in drei Kategorien entsprechend ihrer genetischen Signifikanz eingeteilt: A: Proteine zeigen Segregation in B1. B: Proteine zeigen ein Hybridmuster in F1, aber keine Segregation in B1. C: Proteine zeigen nur stammesspezifische Variation, aber erlauben keine genetische Interpretation in F1 und B1. Eine detaillierte Darstellung der Häufigkeit der verschiedenen Variationstypen und Kategorien gibt Tab.3.2.b) an.

Tab.3.2.b) Häufigkeit der verschiedenen Variationstypen. Spalte A: Proteine zeigen Segregation in B1. Spalte B: Proteine zeigen ein Hybridmuster in F1, aber keine Segregation in B1. Spalte C: Proteine zeigen nur stammesspezifische Variation, aber erlauben keine genetische Interpretation in F1 und B1.

Variationstypen der polymorphen Proteinspots	Anzahl der varianten Proteinspots Zahl der Spots – Zahl der Varianten			
	Kategorien genetischer Signifikanz (s. Text)			Summe
	A	B	C	
<u>Qualitative Varianten</u>				
Positionsvarianten (mV)	574 – 293	46 – 34	48 – 39	668 – 366
Gesamtzahl in %				7.62%
<u>Quantitative Varianten</u>				
Presence/absence Variants (paV)	137 – 119	132 – 119	72 – 67	341 – 305
Amount Variants (aV)	33 – 29	165 – 132	117 – 104	315 – 265
Gesamtzahl	170 – 148	297 – 251	189 – 171	656 – 570
Gesamtzahl in %				7.49%
Gesamtzahl aller Varianten	744 – 441	343 - 285	237 – 210	1324 – 936
Gesamtzahl aller Varianten in %				15.1%

3.3 Eine genetische Karte aus DNA- und Proteinpolymorphismen

3.3.1 Vorbemerkung zur Genkartierung

Die Analyse von genetischer Kopplung besteht aus drei Teilen: (a) dem Test auf Kopplung (linkage) zwischen Markern (oder in unserem Fall Marker und dem zu kartierenden Proteinpolymorphismus), (b) dem Abschätzen der genetischen Distanz in cM zwischen zwei Markern (oder Marker und Protein) (map distance) und (c) dem Bestimmen der Ordnung mehrerer Marker (oder Marker und Proteinen) einer Kopplungsgruppe (ordering).

Im Test auf Kopplung zwischen zwei Markern oder Protein und Marker wird geprüft, ob die Zahl der Rekombinationen zwischen zwei Markern signifikant abweicht von rein zufälliger Rekombination. Diese beträgt für Loci, die auf verschiedenen Chromosomen liegen, also ungekoppelt vererbt werden, 50%. Um Kopplung zu begründen gibt es drei Methoden: a) die klassische, statistische Methode nach Silver (Silver 1985; Ott und Donis-Keller 1994), b) die Bayes'sche Analyse nach Neumann (Neumann 1990, 1991) sie schätzt Wahrscheinlichkeiten von 95% und 99% für Kopplung zweier Marker ab; c) die dritte Methode ist die (LOD = logarithmic odds ratio) nach Morton (Morton 1955). Die Methode nach Morton vergleicht zwei Wahrscheinlichkeiten: die Wahrscheinlichkeit dafür, dass die beobachteten Daten gefunden werden, wenn die beiden Loci in der aus der Rekombinationsfrequenz errechneten Distanz liegen, und die Wahrscheinlichkeit, genau diese Daten zu erhalten, wenn die Loci nicht gekoppelt wären. Der Logarithmus des Verhältnisses dieser beiden Wahrscheinlichkeiten wird abgekürzt mit LOD-Wert. Ein LOD-Wert von 3 bedeutet eine Wahrscheinlichkeit von $10^3/1$ für Kopplung von zwei Loci. In anderen Worten: die Kopplung zweier Loci ist 1000 x wahrscheinlicher als das zufällige Auftreten dieser Daten zwischen nicht gekoppelten Loci. Wie in Tab. 3.3.1 dargestellt besteht ein Zusammenhang zwischen den Kopplungswerten der drei Verfahren (Manly 1993 und „Handbuch des Mapmanager Programms“ unter <http://mcbio.med.buffalo.edu/pub/MapMgr/>)

Tab.3.3.1: Zusammenhang zwischen der klassischen statistischen Methode, der Bayes'sche Analyse und der LOD-Wert Methode zur Analyse von Kopplung

Bayes'sche Analyse (Neumann)	LOD-Wert der Kopplung (Morton)
95% Wahrscheinlichkeit der Kopplung	3,1
99% Wahrscheinlichkeit der Kopplung	4,0

conventional χ^2 Test (Silver, Ott)	LOD-Wert der Kopplung (Morton)
P = 0.001	2,6
P = 0.0001	3,6

Das Abschätzen der genetischen Distanz in cM zwischen zwei Markern, die signifikante Kopplung zeigen, kann direkt aus der Rekombinationsfrequenz ermittelt werden: z.B. zeigen zwei von 64 Tieren Rekombination zwischen zwei Markern, so beträgt ihr Abstand auf dem Chromosom 3.1cM ($2/64$ Tieren zeigen Rekombination = $0,0313 \times 100 = 3,1\% = 3,1\text{cM}$). Bei Rekombinationsfrequenzen $> 1/5$ (= 20cM) können zur Berechnung des Abstandes in cM andere Algorithmen hinzugezogen werden, wie z.B. der von Haldane, der die Wahrscheinlichkeit von Doppelrekombinationen berücksichtigt oder der Algorithmus von Kosambi, der Interferenz berücksichtigt.

Für diese Arbeit wurde der Haldane'sche Algorithmus verwendet, da das vorhandene Gerüst aus DNA-Markern, sogenannten Ankermarkern, die vom EUCIB-Projekt für Nutzer allgemein zur Verfügung gestellt werden, grosse Lücken aufwies und mit versteckten Doppelrekombinationen gerechnet werden musste (siehe Abb.3.3.2).

Die Wahrscheinlichkeit der Ordnung von Markern einer Kopplungsgruppe wird aus den Abständen und Rekombinationsfrequenzen errechnet.

Bei einer Rückkreuzung kann die Abfolge von Markern auf dem Chromosom durch eine einfache Haplotypanalyse per Hand bestimmt werden (gute Beschreibung des Vorgehens bei Silver Kap 9.4.4.5) oder mit Hilfe des Computerprogramms Mapmanager erstellt werden (Manly 1993).

Für eine F₂-Geschwisterkreuzung oder die Segregationsanalyse in menschlichen Familien wurde die Methode der ‚maximum likelihood analysis‘ (Lander und Green 1987) entwickelt und ist als Computerprogramm namens ‚MapMaker‘ bekannt (Lander et al. 1987). Hier werden für verschiedene Abfolgen von Markern Wahrscheinlichkeiten berechnet (likelihoods). Die Sicherheit der Ordnung mit der höchsten Wahrscheinlichkeit bestimmt sich aus dem Vergleich mit der Wahrscheinlichkeit für die 2. beste Ordnung. Wenn die Wahrscheinlichkeit für die 1. Ordnung 1000 x höher liegt als für die 2. Ordnung (log likelihood von 3) dann ist sie sehr sicher (Lander et al. 1987). Für die Erstellung der genetischen Karte aus Ankermarkern des EUCIB-Projektes, IRS-PCR und Mikrosatelliten - Markern sowie Datensätzen von Proteinpolymorphismen wurde in dieser Arbeit das MapMakerprogramm verwendet, da mit der grossen Anzahl der Rückkreuzungstiere (Ankermarkerdaten für ca. 900 Tiere, IRS-PCR Markerdaten für 200-300 Tiere, Mikrosatelliten- und Proteindaten für 10-64 Tiere) sowie der zu ordnenden Datensätze (88 Ankermarker, 69 IRS-PCR Marker, 6 Mikrosatellitenmarker, 441 Datensätze polymorpher Proteine, 53 MSO Marker des EUCIB Projektes) eine Haplotypanalyse per Hand nicht durchführbar war. Obwohl das Programm die rechenaufwendige Methode der „maximum likelihood analysis“ verwendet und in der Lage ist, nicht-informative Genotypisierungsdaten miteinzubeziehen, beschränkt es sich bei der Analyse einer (voll informativen) Rückkreuzung darauf, Rekombinationen zwischen Markern zu zählen und daraus eine Anordnung der Marker und Proteine zu erstellen (Lander und Green 1987).

3.3.2 Genkartierungsdaten

Der Ausgangspunkt der Kartierung: die Rahmenkarte (framework map)

Als Rahmen für die Integration von Proteinen dienten zum einen die Ankermarker des EUCIB-Projektes [Breen, 1994 #399](Rhodes et al. 1998). Die Polymorphismusdaten der Ankermarker wurden von der Webseite der MBx-Datenbank geladen. Diese Ankermarker sind Mikrosatelliten- oder RFLP-Marker. Sie wurden auf allen Rückkreuzungstieren der EUCIB-Kreuzung (982 Tiere) analysiert und bilden daher mit 4-5 Markern pro Chromosom einen guten Rahmen für jede der 20 Kopplungsgruppen (= Chromosomen) des Mausgenoms. Von den 93 Ankermarkern wurden 88 genutzt, da für diese von einem grossen Anteil unserer 64 zur Proteinanalyse eingesetzten Tiere Genotypisierungsdaten vorhanden waren.

Als Anker für die genetische Karte der EUCIB Kreuzung waren Marker ausgewählt worden, die dem damaligen Stand des Chromosomen Reports 1993 in gleichmässigem Abstand von ca. 20 cM über das Mausgenom verteilt waren (Breen et al. 1994; Rhodes et al. 1998). Nach Kartierung der Ankermarker erwiesen sich die genetischen Abstände zwischen vielen Ankermarkerpaaren jedoch als sehr gross (+/- 50 cM).

Da schon versucht worden war, unsere 440 Proteindatensätze in das Gerüst aus Ankermarkern zu integrieren, jedoch ca. 200 Proteine eingeordnet werden konnten, schien das Markergerüst zu grob, um signifikante Kopplung zu finden (M.Kelly und K. Büssow, unveröffentlicht).

Erweiterung der Rahmenkarte mit IRS-PCR und Mikrosatellitenmarkern

In unserem Fall war der Einsatz zusätzlicher Ankermarker erforderlich, um ein Markergerüst aufzubauen, zu dem möglichst viele der 440 Proteinpolymorphismen signifikante Kopplung zeigen.

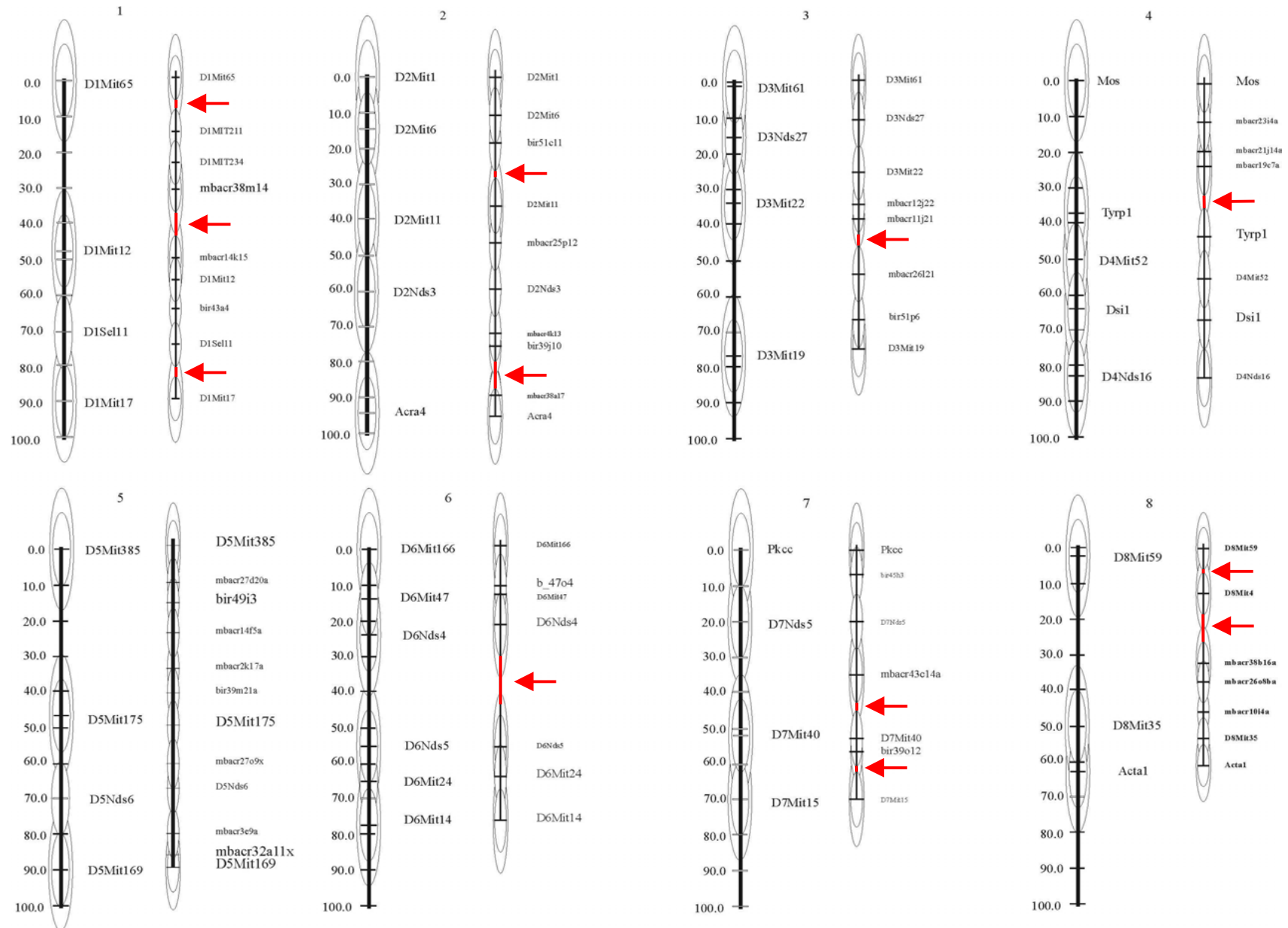
Die Erhöhung der Markernetzwerkdicke war für uns auch deshalb besonders wichtig, weil ein Teil der Proteinpolymorphismusdaten in nur 20 - 30 der 64 Rückkreuzungstiere erhoben werden können.

Als Kriterium für den maximalen Abstand zwischen zwei Markern, über den Kopplung mit 95% Wahrscheinlichkeit gefunden werden kann, gilt der „Swept Radius“. Das Konzept des Swept Radius wurde von Carter und Falconer entwickelt (Carter und Falconer 1951). Der Swept Radius wurde definiert als die Grösse eines Intervalls auf jeder Seite eines Markers innerhalb dessen Kopplung mit einer bestimmten Signifikanz gefunden werden kann (Silver und Buckler 1986; Neumann 1990, 1991; Silver 1995).

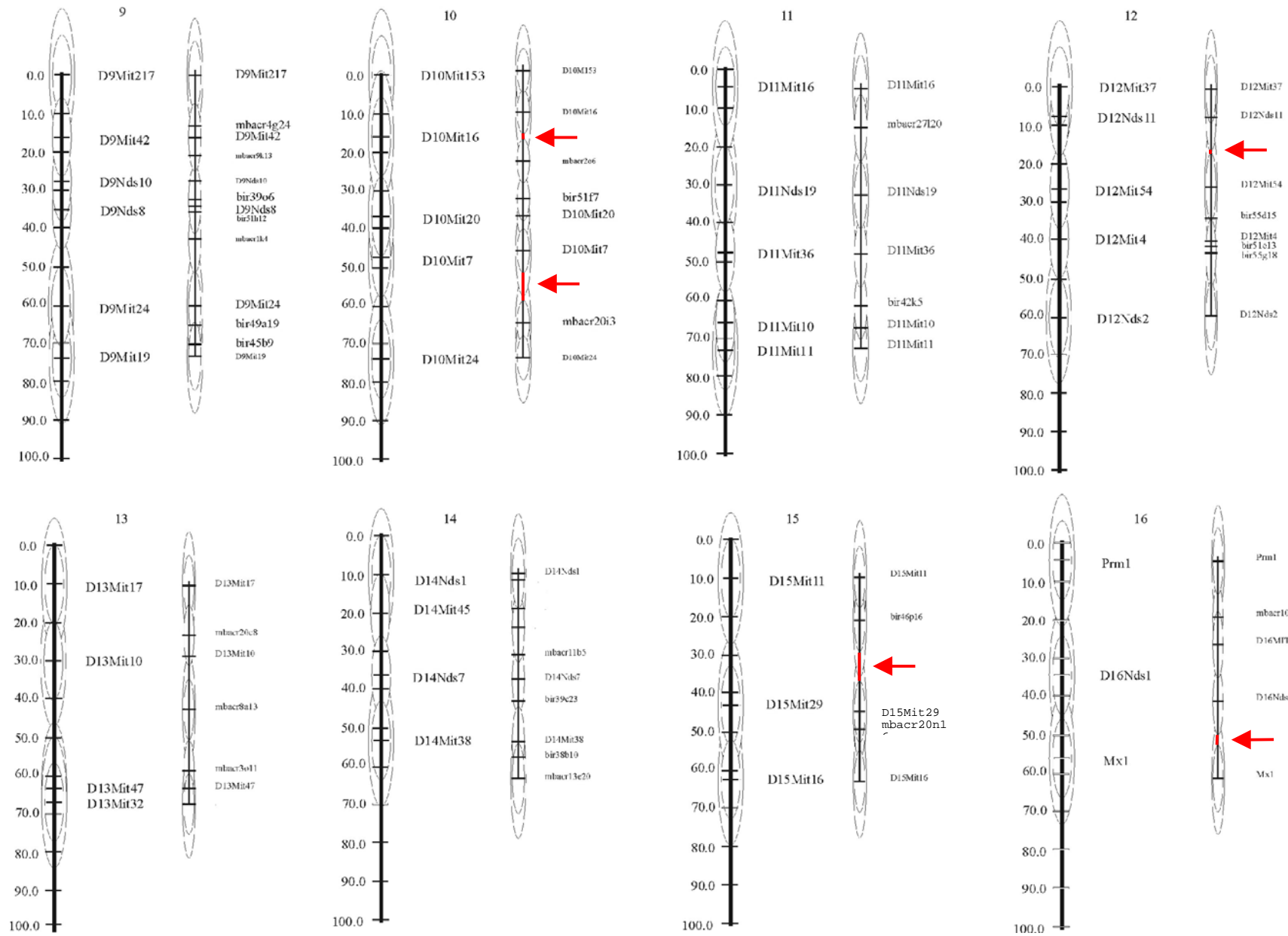
Die Länge des Swept radius auf jeder Seite eines Ankermarkers beträgt nach Silver bei 60 gemeinsam genotypisierten Tieren 17cM (95% confidence level), Kopplung für nur 30 gemeinsam genotypisierte Tiere kann im Umkreis von je 10cM eines Ankers gefunden werden. Abb.3.3.2 zeigt, inwieweit Kopplung gefunden werden kann: links, das Ankergerüst des EUCIB Projekts rechts, das um IRS- und Mikrosatellitenmarker erweiterte Gerüst dieser Arbeit.

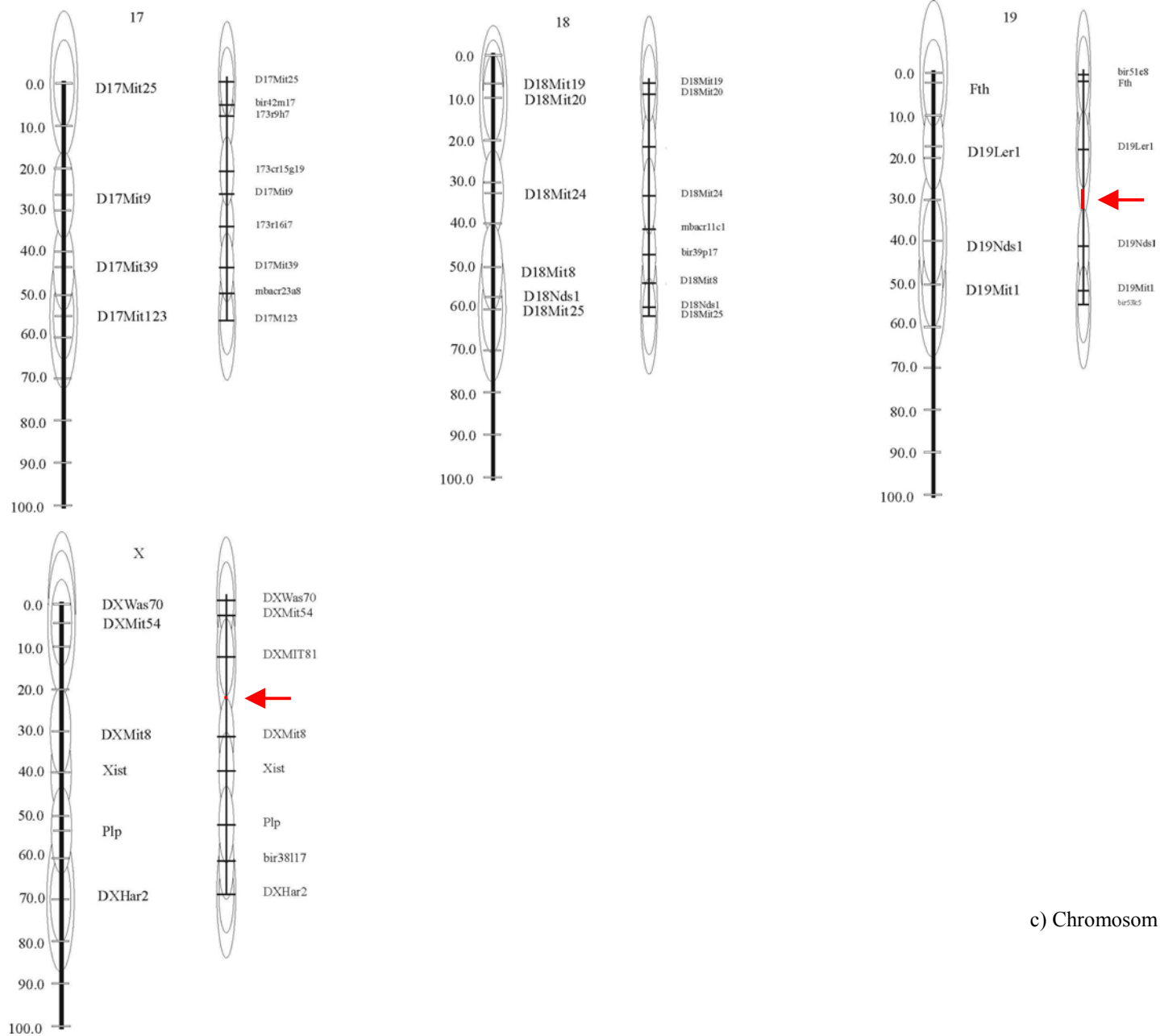
Abb.3.3.2: Bereiche der Chromosomen 1-19 und X, in denen Kopplung nachgewiesen werden kann. Grosse Kreise bezeichnen den Bereich innerhalb dessen mit 60 Genotypisierungsdaten Kopplung gefunden werden kann. Innerhalb der kleinen Kreise kann auch für nur 30 Typisierungen signifikante Kopplung nachgewiesen werden. Das linke Chromosom eines Paares zeigt die Ankermarkerkarte des EUCIB-Projektes, die rechten Chromosomen zeigen, dass nach Einführen der in dieser Arbeit beschriebenen IRS-PCR und Mikrosatellitenmarker die Chancen, Kopplung zu Protein-segregationsdaten zu finden stark erhöht wurden. Zum Beispiel könnten für Chromosom 1 zwischen den Ankermarkern D1Mit65 und D1Mit12 keine Proteinpolymorphismus kartiert werden, da ihr genetischer Abstand zu gross war, um signifikante Kopplung für Daten zu finden, die auf 60 Rückkreuzungstieren erhoben wurden. Nach dem Einführen neuer Marker sieht man, dass die Bereiche um die Marker in denen Kopplung gefunden werden kann, überlappen. Für das gesamte Chromosom kann nun Kopplung zu Protein-segregationsdaten gefunden werden. Rot markiert sind Regionen, in denen mit nur 30 verfügbaren Genotypisierungsdaten keine Kopplung gefunden werden kann.

a) Chromosomen 1-8



b) Chromosomen 9-16





c) Chromosom 17-19, X

Verbleibende Lücken des Markergerüsts, die mehr als 2x17cM betragen, befinden sich auf Chromosom 6 (D6Nds4 - 52,5 cM - D6Nds5) und Chromosom 8 (D8Mit4 - 42,4 cM - bir38b16). In diesen Bereichen kann selbst für Genotypisierungsdaten von 60 Tieren keine Kopplung zum bestehenden Markernetzwerk etabliert werden. Abstände von mehr als 2x10cM bestehen auf den Chromosomen 1, 2, 3, 4, 7, 8, 10, 12, 15, 16, 19 und X (siehe rote Markierungen in Abb.3.3.2). In diesen Intervallen kann keine Kopplung für Proteinpolymorphismen etabliert werden, die auf weniger als 30 Rückkreuzungstieren erhoben wurden.

Tab.3.3.2.a): Anzahl der Proteinpolymorphismen, die Kopplung zum DNA-Markernetzwerk zeigen. Nach Integration von IRS-PCR und Mikrosatellitenmarkern zeigen ca. 30 weitere Proteinpolymorphismen Kopplung zum Markernetzwerk.

Anteil aller polymorphen Varianten, die signifikante Kopplung zeigen zu:			
einem Framework aus 88 EUCIB Ankermarkern		einem Framework aus 88 EUCIB Ankermarkern, 69 IRS-Markern und 6 Mikrosatellitenmarkern	
LOD 3 95% Wahrscheinlichkeit	LOD 4 99% Wahrscheinlichkeit	LOD 3 95% Wahrscheinlichkeit	LOD 4 99% Wahrscheinlichkeit
371/441 = 84%	341/441 = 77%	398/441 = 90%	374/441 = 85%

Integration von Proteindaten in das DNA-Marker Netzwerk

Das Erstellen einer geordneten genetischen Karte aus 374 Proteinpolymorphismen, die mit LOD-Wert 4 Kopplung an das Netzwerk aus 88 EUCIB-Ankermarkern, 69 IRS-PCR Markern und 6 Mikrosatellitenmarkern zeigten, wurde mit dem Programm Mapmaker (Lander et al. 1987) durchgeführt. 359 Proteine konnten mit einem Grenzwert der relativen Wahrscheinlichkeiten von $\geq 10^{2.5} : 1$ geordnet werden. Für das Abschätzen der wahrscheinlichsten Position im Vergleich zur nächstwahrscheinlichen Position eines Proteins im Markergerüst wird von Lander für die sukzessive Einordnung von Markern der Grenzwert von $10^2 : 1$ als ausreichend für die Positionierung angegeben. In der Mikrosatellitenkarte von Dietrich et al. wurden die Marker bis zu einem Wahrscheinlichkeitsgrenzwert von $10^1 : 1$ geordnet (Dietrich et al. 1992).

Wie man den Abbildungen der Chromosomenkarten in Anhang sehen kann, trat oft der Fall auf, dass Proteine zwar in die gleiche Region der Karte fielen, sich aufgrund widersprüchlicher Rekombinationen aber nicht gemeinsam in ein Intervall integrieren liessen. In diesem Fall wurde das Protein (oder die Gruppe von Proteinen) mit der höheren Wahrscheinlichkeit zuerst in das Intervall integriert und restliche Proteine neben die Karte gezeichnet, und zwar über die Intervalle, für die eine Wahrscheinlichkeit der Reihenfolge von $10^{2.5}$ fach : 1 nicht erreicht wurde.

50 Proteine konnten lediglich Chromosomen zugeordnet werden: 32 davon zeigten Kopplung an einen oder mehrere Marker des Chromosoms mit einem LOD-Wert von > 4 ; 18 mit einem LOD-Wert zwischen 3 und 4. Für diese 50 Proteine, sowie für alle Proteine, die über mehrere Intervalle der Karte gezeichnet sind, enthält Tab.1 des Anhangs die Angaben des höchsten LOD-Wertes für die Kopplung zu einem Marker der Karte sowie den Abstand in cM zu diesem Marker. Chromosomenkarten und Tabellen mit Angaben zur Positionierung und Kopplung befinden sich im Anhang.

Integration der MSO-Marker des EUCIB Projektes in die genetische Karte

Die EUCIB Rückkreuzung ist für genomweite Feinkartierung von Markern (global high resolution mapping) konzipiert worden. Im ersten Schritt wurden die Rekombinationen eines jeden Mauschromosoms durch Typisierung aller 1000 Rückkreuzungstiere mit jeweils 4-5 DNA-Markern pro Chromosom lokalisiert. Die genomweite Ankerkarte lokalisiert damit für jedes Rückkreuzungstier die vorhandenen Rekombinationen. Im zweiten Schritt wurden alle Rückkreuzungstiere, die zwischen denselben Ankermarkern Rekombination gezeigt hatten, mit weiteren Markern für dieses Intervall typisiert. Dieses zweite Set von Markern wurde ‚sekundäre Ankermarker‘ oder ‚MSO‘ genannt. Sie wurden zur Genotypisierung von rund 200 Rückkreuzungstieren eingesetzt.

Bei der Erstellung unsere eigenen genetischen Karte aus Protein- und DNA-Polymorphismen wurde in den letzten Schritten der Kartierung auch MSO-Marker miteinbezogen. Die Datensätze der MSO-Marker wurden von MBx bezogen. Sie enthalten Genotypisierungsdaten für ca. 200 der insgesamt 1000 Rückkreuzungstiere, jedoch sind meist nur wenige unserer 64 zur Proteinkartierung eingesetzten Tiere genotypisiert. Die wenigen mit den Proteindaten überlappenden Typisierungen machen eine Einordnung mit Mapmaker oft nur mit einer relativen Wahrscheinlichkeit von 10:1 (log likelihood ratio 1) zwischen die kartierten Proteine möglich. 53 MSO Marker wurden in die Karte integriert, um ausser den Ankermarkern selbst weitere Referenzpunkte zu schaffen, über die unsere Karte mit der EUCIB Karte und anderen Maus-Karten (Copeland et al. 1993; Dietrich et al. 1994) verglichen werden kann.

Die Kartierungsergebnisse für Ankermarker und MSO-Marker, IRS-PCR- und Mikrosatellitenmarker sowie 409 kartierte Varianten sind im Anhang (Chromosomenkarten) dargestellt. Von den 1324 polymorphen Proteinspots der Elternstämme *Mus musculus* und *Mus spretus* (15% aller Spots des 2D-Gels) zeigten etwas mehr als die Hälfte (8.5%) mendelnde Segregation in der Rückkreuzungsgeneration. Die Hälfte der Spots (7.6%) konnte kartiert werden, sie entsprechen 409 Varianten und damit 409 genetisch kartierten Loci (Tab.3.3.2.b).

Tab.3.3.2.b) Kartierte Varianten: 409 der 441 segregierenden Varianten konnten in dieser Arbeit kartiert werden. Sie entsprechen 664 Spots des 2DE-Gels.

genetisch kartierte Varianten: entsprechende Einzelspots:	409 665 (7.6%)
segregierende Varianten: entsprechende Einzelspots:	441 744 (8.5%)
polymorphe Spots der Elternstämme: <i>Mus musculus – Mus spretus</i>	1324 (15,1 %)
Anzahl der Spots eines 2D- Musters (Gehirn, Cytoplasma)	8767 (100%)

3.3.3 Datenqualität

Die automatisierten Analysen mit Hilfe des 'build' Befehls des Programms Mapmaker (Kap.2.4) förderten Unstimmigkeiten zu der „von Hand“ mit dem 'try' Befehl entworfenen Karte zu tage. Abweichungen in der Einordnung von Proteinen in einzelne Intervalle beruhten auf der Tatsache, dass mehrere Proteine jeweils einzeln zwar unzweifelhaft in ein Intervall der Karte eingeordnet werden konnten, jedoch untereinander widersprüchliche Rekombinationen zeigten, weshalb eine gemeinsame Integration in die Karte nicht möglich war. Wie in Kapitel 2.4 beschrieben wurde ein Teil der Proteine in entsprechende Intervalle integriert, während die Kartierungsregion der verbleibenden mit einer Linie neben dem Chromosom bezeichnet wurde.

3.3.4 Verteilung der kartierten Proteine auf die einzelnen Chromosomen – Vergleich mit der MBx-Datenbank

Alle Segregationsdaten, die von Mäusen des EUCIB-Projektes gewonnen werden, können in eine Projektdatenbank eingespeist werden. Sie werden dann automatisch in die genetische Karte des EUCIB-Projektes eingeordnet. Die Tab.3.3.4.a) vergleicht die Kartierungsergebnisse der MBx-Datenbank mit den hiesigen Ergebnissen. Weiterhin werden die 409 kartierten Varianten auf ihre Kartierungsposition hin analysiert: Varianten, die auf dieselbe Stelle oder in das gleiche Intervall kartieren, können (müssen aber nicht!) demselben Locus angehören, daher wurde zwischen gleichen und unterschiedlichen Kartierungspositionen unterschieden. Schlussendlich wurde aufgelistet, welche Variantentypen in einer Gruppe gleichkartierender Varianten vorkommen.

Tab.3.3.4.a): Anzahl der kartierten Varianten der Mbx-Datenbank im Vergleich zu hiesigen Kartierung. Spalte 1 zeigt die Anzahl der kartierten Varianten der Mbx-Datenbank im Vergleich zu hiesigen Kartierung, sowie abweichend kartierte Varianten. Spalte 2 nennt die Zahl der positionierten Varianten, ohne die, die lediglich Kopplung zum Chromosom zeigen. Spalte 3 gibt an, in wieviele Gruppen die kartierten Varianten fallen und Spalte 4 listet die Variantentypen (mV, aV, paV) auf die in einer Gruppe enthalten sind (2mv (3x) bedeutet: eine Gruppe von 2 mV-Varianten findet sich 3x auf dem Chromosom)

Chromosom	Zahl der <u>Varianten</u> pro Chromosom			Zahl der <u>positionierten</u> * ¹ Varianten pro Chromosom (Position oder Intervall) (diese Arbeit)	Zahl der <u>unterschiedlichen</u> Kartierungspositionen* ² (,Gruppen‘)	Variantentypen an einer Kartierungsposition
	diese Arbeit	MBx	Abweichungen			
1	13	9	mV184	10	8	1mV (3x) 1paV (2x) 2mV (1x), 7mV (1x), 8mV (1x)
2	16	12		14	11	1mV (4x) 1paV (4x) 2mV (1x), 2paV(1x), 2mV/1paV(1x)
3	24	12		19	14	1mV (6x) 1paV (4x) 2mV, 3mV, 5mV, 1paV/1aV
4	18	8		16	9	1mV (2x) 1paV 2mV (2x), 7mV, 2mV/1paV (2x) 1mV/1paV
5	21	0		18	12	1mV (4x) 1paV (2x) 1aV 4mV (2x) 2mV, 3mV, 2paV
6	78	16	mV77, mV197	71	19	1mV (4x) 1paV, 1aV 2mV (2x) 2paV (2x) 3mV, 4mV, 6mV 9mV, 1mV/1paV, 3paV, 58mV/1paV 17mV/7paV
7	23	19		20	12	1mV (4x) 1paV (2x) 2mV (2x) 3mV, 6mV, 14mV 1mV/3paV
8	17	5		12	6	1mV (2x) 1paV (2x) 4mV, 5paV
9	24	19		22	17	1mV (4x) 1paV (3x) 2mV (4x) 3mV (2x) 10mV, 12mV

Chromosom	Zahl der <u>Varianten pro Chromosom</u>			Zahl der <u>positionierten</u> * ¹ Varianten pro Chromosom (Position oder Intervall) (diese Arbeit)	Zahl der <u>unterschiedlichen</u> Kartierungspositionen* ² („Gruppen“)	Variantentypen an einer Kartierungsposition
	diese Arbeit	MBx	Abweichungen			
10	13	9		10	7	1mV (2x) 1paV 2mV, 4mV, 6mV 14mV
11	35	13		33	11	1mV (3x) 1paV, 1mV 2paV (2x) 52mV, 2mV/4paV/1aV 1mV/1paV
12	12	5		11	7	1paV (2x) 1aV 2mV, 4paV, 1mV/2paV, 1paV/1aV
13	17	11		16	10	1paV (4x) 3mV, 4mV, 2paV, 14mV 2mV/1paV/1aV, 1mV/1paV
14	10	8		8	6	1mV (3x) 1paV, 2paV, 1mV/1paV
15	18	11	mV66, mV68, paV467	16	11	1mV (4x) 1paV (3x) 1aV 2mV, 3mV, 17mV
16	13	8		11	5	1mV 1paV (2x) 2mV/1aV, 3mV/3paV
17	14	8		14	10	1mV (4x) 1paV (3x) 3mV, 1mV/2paV
18	15	9		13	9	1mV (2x) 1paV (3x) 3mV, 4mV, 5mV, 2mV
19	17	12		17	8	1mV 1aV (2x) 3mV, 2paV, 5mV/1paV, 3mV/1paV/1aV, 3mV/1aV
X	11	6		9	7	1mV (3x) 2mV, 5mV, 7mV, 6mV/2paV
Σ	409	200	6	360	175	

*¹ Varianten, die auf konkreter Position oder in einem Intervall (Linie) kartieren, ohne diejenigen, die nur Kopplung zum Chromosom zeigen

*² Spots, die auf dieselbe Position oder in ein identisches Intervall fallen, werden als gleich kartiert betrachtet (nicht jedoch solche, die in überlappende Intervalle fallen, auch nicht Gruppen, die an den Enden der Chromosomen beginnen)

Wie in Tab.3.3.4 genannt, besteht für einige Varianten ein Unterschied zwischen der im Rahmen dieser Arbeit durchgeführten Kartierung und der Kartierung auf der MBx-Karte. MV184 ist in der MBx Karte zwischen 44cM und 45cM gemeinsam mit mV76 integriert, obwohl zwischen den Varianten mehr als 12cM genetischer Abstand besteht. MV77 und mV197 sind in der MBx Karte in umgekehrter Reihenfolge aufgeführt. Ihre Reihenfolge ist jedoch mit einer relativen Wahrscheinlichkeit von $10^3 : 1$ so wie in dieser Arbeit beschrieben (Chr.Karte). Die Variante mV66 wurde mit Mapmaker mit einer relativen Wahrscheinlichkeit von $> 10^{2,5}$ (log likelihood ratio $>2,5$) in Bezug auf mV68 und paV467 in der hier angegebenen Reihenfolge geordnet.

Die Kartenlänge in Abhängigkeit von der Kartierungsfunktion

Centimorgan-Abstände in genetischen Karten sind von der Anzahl und Verteilung der Rekombinationsereignisse zwischen Markern abhängig. Die Rekombinationshäufigkeit über das Genom hinweg variiert zwischen verschiedenen Mäusestämmen (Seldin et al. 1989). In verschiedenen Kreuzungen erscheinen deshalb Markerabstände in genomischen Regionen mit häufiger Rekombination zu gross im Verhältnis zur wirklichen physikalischen Länge des Chromosoms. Dieses Phänomen beruht darauf, dass Crossover nicht gleichmässig verteilt über das ganze Genom verteilt stattfinden, sondern in kleinen einige Kilobasen grossen Regionen – diese werden hot spots of recombination genannt - stark gehäuft auftreten (Bryda et al. 1992). Weiterhin spielt das Geschlecht der untersuchten Tiere eine Rolle (Reeves et al. 1991).

Die Länge einer genetischen Karte hängt auch von dem zur Berechnung der Abstände verwendeten Algorithmus ab. Tab.3.3.4.b) demonstriert die verschiedenen Längen von Chromosom 1, die man mit denselben Ausgangsdaten erhalten kann. Eine Berechnung der Abstände zwischen den Markern mit der Kartierungsfunktion nach Kosambi (Kosambi 1944) ergibt eine geringere Länge als die Berechnung mit der Haldane'schen Kartierungsfunktion, da Kosambi das Phänomen der Interferenz (Haldane und Waddington 1931) mit einbezieht. Interferenz heisst, dass ein Chiasma (Crossover) zwischen den elterlichen Chromosom während der Meiose weitere Chiasmata in der Nachbarschaft unterdrückt, so dass zwischen zwei Markern mit bis zu 20 % Rekombinationen keine unentdeckte Rekombination zu vermuten ist. Haldane's Algorithmus dagegen berechnet - je grösser die Rekombinationsfrequenz zwischen zwei Markern ist - unentdeckte Doppelcrossover mit ein und schätzt den echten Abstand zwischen den Markern als wesentlich grösser ein. Aufgrund der grossen Abstände zwischen den Markern der EUCIB Ankerkarte haben wir unsere Karte mit der Haldane Funktion berechnet. Tab.3.3.4.c) demonstriert für jedes Chromosom die Länge unseres Gerüsts aus Ankermarkern, IRS-PCR - und Mikrosatelliten - Markern vor und nach der Integration der Proteinpolymorphismen. Das Kriterium, die Längen der Chromosomen durch die Integration von Proteindaten nicht zu vergrössern (map inflation), wurde gut erfüllt.

Die Kartenlänge nimmt mit der Zahl der Rückkreuzungstiere zu: verwendet man nur 90 Tiere des gesamten Datensatzes zur Berechnung der Karte, ergeben sich mit beiden Kartierungsfunktionen

geringere Kartenlängen. Das zeigt, dass in den 900 Tieren im Verhältnis mehr fehlerhafte Rekombinationen zu finden sind als in 90 Tieren.

Tab.3.3.4.b) Berechnung der Länge von Chromosom 1 mit dem Algorithmus nach Haldane bzw. Kosambi. Die Tabelle demonstriert die Unterschiede in der Chromosomenlänge für Chromosom 1, wenn für exakt denselben Datensatz verschiedene Algorithmen verwendet werden.

Chromosom 1		Algorithmus nach Haldane	Algorithmus nach Kosambi
diese Arbeit: Anker (Centromer – IRS- + Mikrosat.Marker – Anker (Telomer)	900 Tiere	145,8 cM	124,3 cM
	90 Tiere	115,8 cM	101,0 cM
MBx Datenbank: Anker (Centromer) – Marker – Anker (Telomer)		Algorithmus (Rhodes et al. 1998) 90 cM	

Tab.3.3.4.c) Die Gesamtlänge der genetischen Karte. Hier soll demonstriert werden, wie sich die Kartenlänge nach Integration der Proteine in die Rahmenkarte aus Ankermarkern, IRS-PCR-Markern und Mikrosatellitenmarker verändert.

	<u>diese Arbeit</u>		<u>diese Arbeit</u>	<u>MBx-Karte</u>
	Anker-Centromer – IRS- und Mikrosat.Marker – Anker-Telomer		Anker-Centromer – IRS- und Mikrosat.Marker – Proteine – – Anker-Telomer	Anker-Centromer – Marker – Anker-Telomer
	Haldane- Algorithmus	Kosambi- Algorithmus	Haldane- Algorithmus	(Rhodes et al. 1998)
Chr1	145,8	124,3	146,3	90
Chr2	140,9	122,3	141,9	95
Chr3	115,3	99,7	113,4	77
Chr4	116,6	100,3	115,6	84
Chr5	143,9	127,8	151,0	92
Chr6	106,1	85,6	105,3	78
Chr7	104,6	89,3	104,2	71
Chr8	121,7	101,6	120,0	62
Chr9	90,0	81,5	89,9	73
Chr10	93,7	82,7	97,0	68
Chr11	91,7	79,6	89,6	72
Chr12	75,2	66,0	76,2	60
Chr13	79,4	69,4	80,1	68
Chr14	62,3	58,1	68,7	53
Chr15	74,6	62,6	80,0	62
Chr16	78,1	66,3	77,8	56
Chr17	76,0	69,1	77,0	59
Chr18	66,4	60,7	66,9	55
Chr19	70,1	59,5	72,4	54
ChrX	89,2	78,8	89,2	69
Summe	1942,4cM	1626,2cM	1962,5cM	1398cM