

Genomanalyse

Kernfach 2. & 3. Studienjahr

2000 / 2001



Gaudenz Dolf
Institut für Tierzucht
Universität Bern

dolf.gaudenz@itz.unibe.ch

Inhaltsverzeichnis

1. Einführung in die Genomanalyse	4
1.1. Zusammenfassung des Stoffes vom 1. Jahr	4
1.1.1. DNA-Sequenzierung	4
1.1.2. Physische Kartierung	4
1.1.2.1. In Situ Hybridisation	4
1.1.2.2. Somatische Zellhybridisation.....	5
1.1.3. Genetische Kartierung	5
1.1.4. Kartierung mit Radiationshybriden	6
2. Segregationsanalyse	6
2.1. Familienmaterial.....	6
2.1.1. Umfang und Struktur.....	6
2.1.2. Vollständigkeit.....	7
2.2. Definition von Merkmalen.....	8
2.3. Einfache Merkmale	8
2.4. Komplexe Merkmale	10
2.4.1. Modelle der Vererbung.....	11
2.4.1.1. Hauptgen	11
2.4.1.2. Polygene Vererbung	12
2.4.1.3. Hauptgen und polygene Vererbung	12
2.4.1.4. Umwelt, keine genetische Komponente	12
2.4.2. Interpretation der Vererbungsmodelle	13
3. Kopplungsanalyse	13
3.1. Lod Score Methode.....	13
3.1.1. Entscheidungskriterien.....	14
3.1.2. Genauigkeit der geschätzten Rekombinationsrate	14
3.1.3. Wieviele Rekombinationsereignisse braucht es um Kopplung festzustellen ?	15
3.1.4. Faktoren, die die Rekombinationsrate beeinflussen	15
3.1.4.1. Geschlecht	15
3.1.4.2. Alter.....	16
3.1.4.3. Heterogenität.....	16
3.1.5. Multipoint Analyse.....	17
3.2. Interpretation von Resultaten der Kopplungsanalyse.....	18
3.2.1. Zwei-Punkt Analyse	18
3.2.2. Multipoint Analyse.....	19
3.3. Nicht-parametrische Methoden	19
4. Markerkarten	20
4.1. Erstellen von Markerkarten	20
4.2. Integrierte Markerkarten.....	21
4.3. Präsentation von Markerkarten	22
4.4. Kontrollmöglichkeiten	22
4.4.1. Überprüfen der Position eines Locus.....	23
4.4.2. Methode der Kartierung	23
4.4.3. Verifikation der Identität eines Locus	23
5. Suche nach einem Gen	24
5.1. Whole Genome Scan	24

5.2. Kandidaten Gene	25
6. Index	26
7. Aufgaben für das 3. Jahr	28
7.1. Web-Sites, die für die Genomanalyse bei Haus- und Nutztieren interessant sind	28
7.2. Kupfertoikose beim Bedlington Terrier	28
7.3. Manuskript Genomanalyse	28

1. Einführung in die Genomanalyse

1.1. Zusammenfassung des Stoffes vom 1. Jahr

Unter den Begriff Genomanalyse fallen alle Untersuchungen die zum Verständnis des Aufbaus und der Funktion des Genoms beitragen. Im engeren Sinn befasst sich die Genomanalyse mit zytogenetischen und molekulargenetischen Analysen, die letztendlich der genetischen Beschreibung eines **Phänotyps** dienen. Die aus den Analysen resultierenden Daten werden in der Regel in **Gen- oder Markerkarten** zusammengefasst. Solche Karten präsentieren Gene und anonyme **Marker** in linearer Abfolge auf den **Chromosomen**, und oft auch relative Distanzen zwischen den **Loci**. In der Tierzucht und der veterinärmedizinischen Diagnostik sind vor allem Loci von Interesse, die **polymorph** sind, denn nur diese lassen einen direkten Bezug zu einem Phänotyp zu.

1.1.1. DNA-Sequenzierung

Mit der **DNA-Sequenzierung** kann die lineare Abfolge der **Nukleotide** eines DNA-Einzelstranges ermittelt werden. Bei der Methode nach Sanger wird durch den Einbau von ddNTP's bei der Synthese ein Kettenabbruch erreicht, was dazu führt, dass die DNA-Vorlage in sämtlichen möglichen Längen, von einem Nukleotid bis zur ihrer Gesamtlänge repliziert wird. Diese Fragmente werden in der Regel mit Sequenziergeräten elektrophoretisch aufgetrennt und analysiert. Indem **homologe** DNA-Abschnitte von verschiedenen **Chromosomen** bzw. Individuen einer Spezies sequenziert werden, lassen sich **Mutationen**, wie Substitutionen, Deletionen oder Insertionen feststellen. In der Regel lassen sich aus der Kenntnis von Mutationen **polymorphe Marker** entwickeln, die sich zur Kartierung und für die Diagnostik eignen.

1.1.2. Physische Kartierung

Mit der **physischen Kartierung** werden **Loci** mit verschiedenen Methoden physisch bestimmten **Chromosomen** oder Chromosomenabschnitten zugewiesen.

1.1.2.1. *In Situ Hybridisation*

Mit der **In Situ Hybridisation** werden DNA-Sonden an denaturierte **Chromosomen**, meistens in der Metaphase, hybridisiert. Heute wird dazu fast ausschliesslich eine nicht-radioaktive Markierung der Sonde verwendet. Dabei hat sich die Markierung mit fluoreszierenden Farbstoffen durchgesetzt, weshalb man auch von **FISH** (Fluorescent In Situ Hybridization) spricht. Die Auflösung bewegt sich in der Grössenordnung von 10^6 bis 10^7 bp. Die Chromosomen werden

entweder vor oder nach der Hybridisierung mit Bänderungsfärbungen identifiziert. Es ist auch möglich mit gewissen Verfahren die Hybridisationssignale und die Chromosomen gleichzeitig zu identifizieren. Eine spezielle Anwendung stellt **Fiber-FISH** dar. Die DNA-Sonden werden an lineare, nicht kondensierte, DNA-Stränge hybridisiert, was zu einer Auflösung im Bereich von wenigen kb bis maximal 500 kb führt. Fiber-FISH ermöglicht die lineare Anordnung von **Loci**, die sehr nahe beieinander liegen, und sich mit andern Methoden nicht ordnen lassen.

1.1.2.2. Somatische Zellhybridisation

Bei der **somatischen Zellhybridisation** werden somatische Zellen, der zu kartierenden Spezies (**Donor**) mit Zellen von etablierten Zelllinien (**Rezipient**, in der Regel chinesischer Hamster oder Maus) fusioniert. Bei der Selektion der fusionierten Zellen macht man sich Enzymdefizienzen der Rezipienten und Unterschiede in Eigenschaften der Donoren und Rezipienten, wie z.B. Resistenzen gegenüber toxischen Substanzen, zu Nutze. Hybridzellen, bzw. die daraus abgeleiteten **Hybridzelllinien** haben die Eigenschaft einen Teil der Donorchromosomen zu verlieren, sodass sich verschiedene Hybridzelllinien in der Zusammensetzung der Donorchromosomen unterscheiden. Beim Erstellen eines Panels von Hybridzelllinien ist es das Ziel mit möglichst wenig Hybridzelllinien alle Donorchromosomen eindeutig aus der Kombination der Chromosomeninformation aller beteiligten Hybridzelllinien ableiten zu können. Aus der Kombination der Chromosomenkonstellation und dem Vorhandensein oder Fehlen von **Loci** oder Genprodukten kann abgeleitet werden auf welchem **Chromosom** ein bestimmtes **Gen** oder ein **Marker** liegen muss. Mit der Analyse von somatischen Zellhybriden können **Syntäniegruppen** erstellt werden, wobei über die lineare Anordnung der Loci auf einem Chromosom keine Informationen gewonnen werden können.

1.1.3. Genetische Kartierung

Die **genetische Kartierung** beruht auf dem Auftreten von **Rekombinationsereignissen** zwischen zwei **Loci**, die sich in Familien beobachten lassen. Eine Voraussetzung dafür ist, dass mindestens ein Elternteil für diese Loci doppelt **heterozygot** ist. Werden die beiden Loci voneinander unabhängig vererbt, erwartet man gleichviele rekombinierte und nicht-rekombinierte **Nachkommen**. Sind zwei Loci gekoppelt ist die **Rekombinationsrate** θ kleiner als 50%. Es lassen sich also mehr nicht-rekombinierte als rekombinierte Nachkommen beobachten. Mit einer **Kopplungsanalyse** will man einerseits die Rekombinationsrate schätzen und andererseits statistisch prüfen, ob die geschätzte Rekombinationsrate von 50% abweicht. Bei unseren Haus- und Nutztieren wird dafür in der Regel die **Lod Score** Methode verwendet. Der Hypothese $\theta < 50\%$ wird die **Nullhypothese** $\theta = 50\%$ gegenübergestellt. Für das aus der Maximierung des Lod Scores gefundene θ kann ein Vertrauensintervall konstruiert werden, in welchem das wahre θ mit etwa 95% Wahrscheinlichkeit liegt. Mit Computersimulationen kann man zeigen, dass Umfang und Struktur des **Familienmaterials** einen grossen Einfluss auf die Kosten und die Aussagekraft eines Projektes hat. Heute stehen eine ganze Reihe von Computerprogrammen für die Kopplungsanalyse zur Verfügung, die alle ihre

Vor- und Nachteile haben. Die Auswahl der Programme hängt einerseits von der Zielsetzung des Projektes, und andererseits von der zur Verfügung stehenden Infrastruktur für die Informatik.

1.1.4. Kartierung mit Radiationshybriden

Die Kartierung mit **Radiationshybriden** stützt sich wie die Kartierung mit somatischen Zellhybriden auf die Fusion von Donor-Zellen mit Rezipienten Zellen. Allerdings werden die Donor-Zellen vor der Fusion unter definierten Bedingungen Röntgenstrahlen oder γ -Strahlen ausgesetzt, was bewirkt, dass die **Chromosomen** in Stücke brechen. Damit ist eine chromosomale Zuweisung der **Loci** nicht mehr möglich, denn die Chromosomenstücke lassen sich zytogenetisch nicht identifizieren. Es ist aber möglich einen Panel von Radiationshybriden zusammenzustellen, der es erlaubt zwischen den Loci Distanzen zu schätzen, ähnlich wie bei der **genetischen Kartierung**. Allerdings entsprechen die mit dieser Methode gefundenen Distanzen nicht oder nur annäherungsweise genetischen Distanzen. Genetische Distanzen beruhen auf biologischen Ereignissen, nämlich **Rekombinationsereignissen**, wogegen mit Radiationshybriden gefundene Distanzen auf physikalische Ereignisse zurückzuführen sind. Im Gegensatz zur genetischen Kartierung müssen die Loci bei der Kartierung mit Radiationshybriden nicht **polymorph** sein.

2. Segregationsanalyse

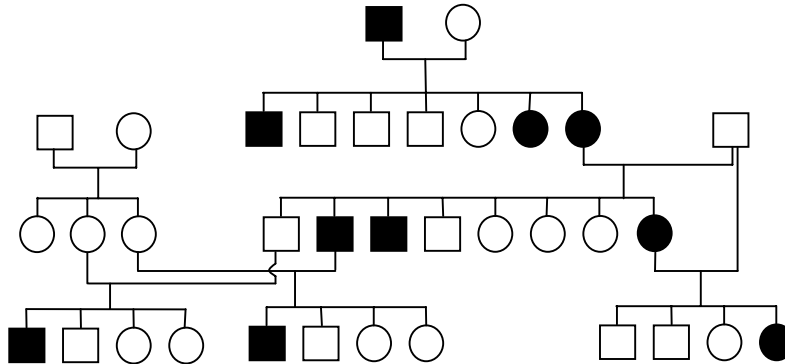
Die **Segregationsanalyse** hat zum Ziel den **Vererbungsmodus** eines **Phänotyps** und damit des darunterliegenden **Genotyps** festzustellen. Dazu braucht es ein möglichst vollständiges **Familienmaterial**, in welchem von möglichst vielen Mitgliedern der Phänotyp bekannt ist. Ein unvollständiges Familienmaterial, in welchem die verwandtschaftlichen Beziehungen ungenügend abgeklärt sind, Mitglieder fehlen oder nur bestimmte Phänotypen miteinbezogen sind, erschwert oder verunmöglicht eine korrekte und zuverlässige Segregationsanalyse. Der Erhebung der Phänotypen kommt eine zentrale Bedeutung zu. Fehlklassierungen von Phänotypen können eine Segregationsanalyse zum Scheitern bringen. Für Fehlklassierungen können **Phänokopien** verantwortlich sein, aber wahrscheinlich häufiger schlecht definierte Merkmale oder oberflächliche **Diagnosen**.

2.1. Familienmaterial

2.1.1. Umfang und Struktur

Ein kleines **Familienmaterial**, das nur eine oder wenige Familien, mit wenig Individuen, umfasst, lässt sich in der Regel problemlos aufzeichnen. Aus den resultierenden Stammbäumen ist es in Fällen von einfachen, monogenen

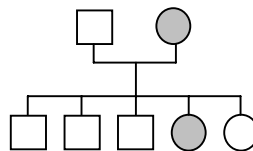
Eigenschaften möglich einen **Erbgang** abzuleiten, der die vorliegenden **Phänotypen** ausreichend erklärt. In der Praxis ist ein Familienmaterial oft komplex und wesentlich grösser, sodass es sehr schwierig wird **Stammbäume** aufzuzeichnen und sich auf einen bestimmten Erbgang festzulegen. Komplex heisst in diesem Fall, dass die Stammbäume sogenannte **Loops** aufweisen. Diese können durch Inzucht oder durch die Anpaarung entstehen.



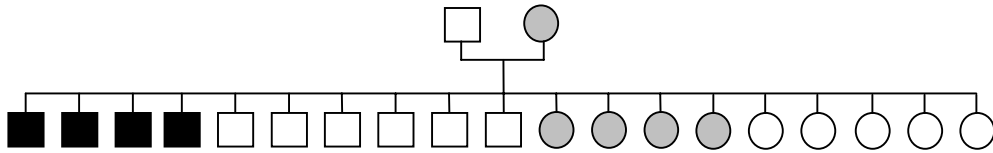
Auf der rechten Seite des obigen Stammbaumes findet sich ein Loop, der durch eine Vater-Tochter Paarung zustande gekommen ist. Auf der linken Seite wurde der Loop durch die Anpaarung von zwei Vollbrüder an zwei Vollschwestern verursacht. Ein Loop besteht immer dann, wenn man in einem Stammbaum, von einem Individuum ausgehend, den Linien folgt und wieder beim gleichen Individuum landet. Ein Familienmaterial bei unseren Haus- und Nutztieren kann ohne weiteres 500 oder mehr Tiere in verschiedenen Familien, und mehrere Loops aufweisen. In solchen Fällen ist man auf den Einsatz von Computerprogrammen und leistungsfähigen Rechnern angewiesen.

2.1.2. Vollständigkeit

Eine Forderung, die an jedes Tiermaterial gestellt werden muss, ist, dass die Familien möglichst vollständig sein müssen. Fehlende Familienmitglieder können eine Segregationsanalyse erschweren oder zu falschen Resultaten führen.



Bei der obigen **Familie** könnte man den Vater mit aa , die Mutter mit Aa bezeichnen oder umgekehrt. Die **Nachkommen** sind Aa und aa im Verhältnis 4 : 1, oder umgekehrt. Die Abweichung vom Verhältnis 1 : 1 könnte man mit der geringen Anzahl von Nachkommen erklären. Als **Erbgang** käme sowohl **autosomal dominant/rezessiv** als auch autosomal **intermediär** in Frage. Die Familie ist aber nicht vollständig, denn es wurden nur einige Nachkommen gesammelt. Ergänzt man die Familie mit den fehlenden Nachkommen ergibt sich ein ganz anderes Bild.



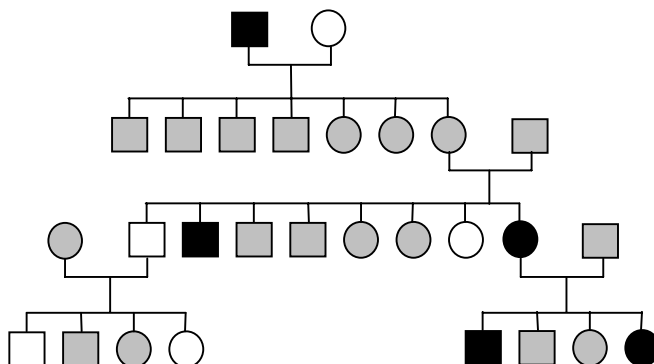
Die vollständige **Familie** deutet auf einen **X-chromosomal** intermediären **Erbgang** hin. Der Vater ist a, die Mutter Aa, die schwarzen männlichen **Nachkommen** A, die weissen männlichen Nachkommen a, die grauen weiblichen Nachkommen Aa, und die weissen weiblichen Nachkommen aa.

2.2. Definition von Merkmalen

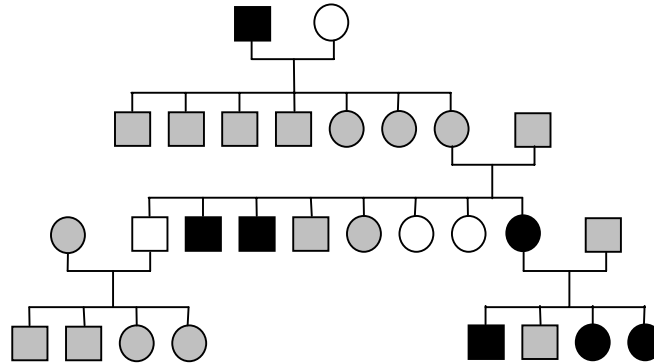
In einer Studie kommt der sorgfältigen Erhebung eines Merkmals allergrösste Bedeutung zu. Damit eine qualitativ gute Erhebung durchgeführt werden kann muss ein Merkmal klar definiert sein. Die Beurteilungskriterien müssen so festgelegt werden, dass sie praktisch durchführbar sind und die verschiedenen **Phänotypen** klar unterscheiden. Dabei muss berücksichtigt werden, ob sich ein Merkmal erst in einem bestimmten Lebensabschnitt manifestiert. In einem solchen Fall müsste der Zeitpunkt der Diagnose festgelegt werden. Man muss auch abklären ob es sehr ähnliche Merkmale gibt (**Phänokopien**), die mit dem interessierenden Merkmal verwechselt werden könnten, und versuchen, diese abzugrenzen. Ferner muss sichergestellt werden, dass die Erhebung für das ganze Tiermaterial gleich durchgeführt wird.

2.3. Einfache Merkmale

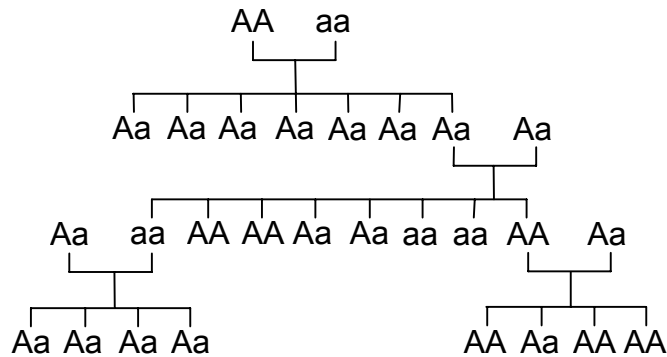
Einfache Merkmale werden scheinbar nur von einem einzigen **Gen** gesteuert. Sie sind aber eher die Ausnahme, denn oft zeigt sich, dass sie mit anderen Merkmalen verknüpft oder nur ein Teilaspekt eines Merkmalkomplexes sind. Aber in einer ersten Näherung ist es durchaus sinnvoll und oft auch ausreichend anzunehmen, dass ein Gen für ein Merkmal verantwortlich ist. In solchen Fällen kann es genügen, sich einen **Stammbaum** mit den **Phänotypen** aufzuzeichnen und einen **Erbgang** daraus abzuleiten. Einfache Erbgänge sind **autosomal rezessiv**, autosomal **dominant**, autosomal **intermediär** oder **codominant**, aber auch X-chromosomale Erbgänge kann man dazurechnen.



Der obenstehende **Stammbaum** lässt einen **autosomal** intermediären **Erbgang** vermuten, denn die Paarung schwarz x weiss ergibt nur einheitlich graue **Nachkommen**, die Paarung grau x grau ergibt weisse, graue und schwarze Nachkommen im Verhältnis 1 : 2 : 1, und die Farbverteilung in den beiden Geschlechtern ist ausgeglichen.



Die erste **Familie** im obenstehenden **Stammbaum** deutet noch auf einen **autosomal** intermediären **Erbgang** hin, aber die **Phänotypen** der **Nachkommen** aus den nachfolgenden Paarungen scheinen diesem nicht ganz zu entsprechen.



Für die nachfolgenden Betrachtungen wollen wir annehmen, dass jede Paarung in diesem **Stammbaum** viermal so viele **Nachkommen** im gleichen Phänotyp-Verhältnis gebracht hätte. Diese Annahme treffen wir, damit die Erwartungswerte im statistischen Test genügend gross werden. Für jeden Paarungstyp können wir die beobachteten Nachkommen in jeder Phänotypklasse zählen. In diesem Stammbaum kommen die Paarungen AA x aa, Aa x Aa, Aa x aa und AA x Aa je einmal vor. Die Nachkommen aus der Paarung AA x aa entsprechen vollständig einem **autosomal** intermediären **Erbgang**, aber diejenigen aus den anderen Paarungen scheinen davon abzuweichen. Wir wollen nun überprüfen, ob diese Abweichungen nur zufälliger Natur sind.

Folgende **Phänotypen**, bzw. Genotypen, unter der Annahme einer **autosomal** intermediären Vererbung, haben wir beobachtet:

Paarung	AA	Aa	aa
Aa x Aa	12	8	12
Aa x aa		16	0
AA x Aa	12	4	

Nach Mendel erwarten wir für einen **autosomal** intermediären **Erbgang** folgende Genotyp-Verteilung:

Paarung	AA	Aa	aa
Aa x Aa	8	16	8
Aa x aa		8	8
AA x Aa	8	8	

$$\chi^2 = \sum \frac{(\text{beobachtete} - \text{erwartete Häufigkeit})^2}{\text{Erwartete Häufigkeit}}$$

Aa x Aa:	$\chi^2 = 2.0 + 4.0 + 2.0 = 8.0$	$\chi^2_{\text{Tabelle}} = 9.21$ (FG=2, P=0.01) $\chi^2_{\text{Tabelle}} = 5.99$ (FG=2, P=0.05)
Aa x aa:	$\chi^2 = 8.0 + 8.0 = 16.0$	$\chi^2_{\text{Tabelle}} = 6.63$ (FG=2, P=0.01) $\chi^2_{\text{Tabelle}} = 3.84$ (FG=1, P=0.05)
AA x Aa:	$\chi^2 = 2.0 + 2.0 = 4.0$	$\chi^2_{\text{Tabelle}} = 6.63$ (FG=2, P=0.01) $\chi^2_{\text{Tabelle}} = 3.84$ (FG=1, P=0.05)

Bei einer **Irrtumswahrscheinlichkeit** von 5% weichen alle drei Paarungstypen von einer **autosomal** intermediären Vererbung ab. Bei einer Irrtumswahrscheinlichkeit von 1% weicht nur die Paarung Aa x aa von unseren Erwartungen ab. Es scheint also, dass wir die Vererbung von schwarz, grau und weiss in diesem **Stammbaum** nicht mit einem einzigen **Gen** erklären können, das autosomal **intermediär** vererbt wird. Um unsere Beobachtungen in diesem Stammbaum zu erklären müssen wir andere, komplexere Vererbungsmodelle überprüfen.

2.4. Komplexe Merkmale

Die Segregation eines **komplexen Merkmales** lässt sich nicht mit einem monogenen **Erbgang** erklären. Sobald Begriffe, wie unvollständige **Penetranz**, unvollständige **Expressivität** oder polygene Komponente zuhilfe genommen werden müssen, um die Vererbung eines Merkmales zu erklären, kann man davon ausgehen, dass es sich um ein mehr oder weniger komplexes Merkmal handelt. Bei komplexen Merkmalen ist der **Vererbungsmodus** nicht direkt aus den Daten ersichtlich, wie das zum Teil bei einfachen Merkmalen der Fall ist. Man muss also

verschiedene Vererbungsmodelle überprüfen, wobei nicht immer ein Modell die Situation deutlich besser erklärt als ein anderes Modell.

2.4.1. Modelle der Vererbung

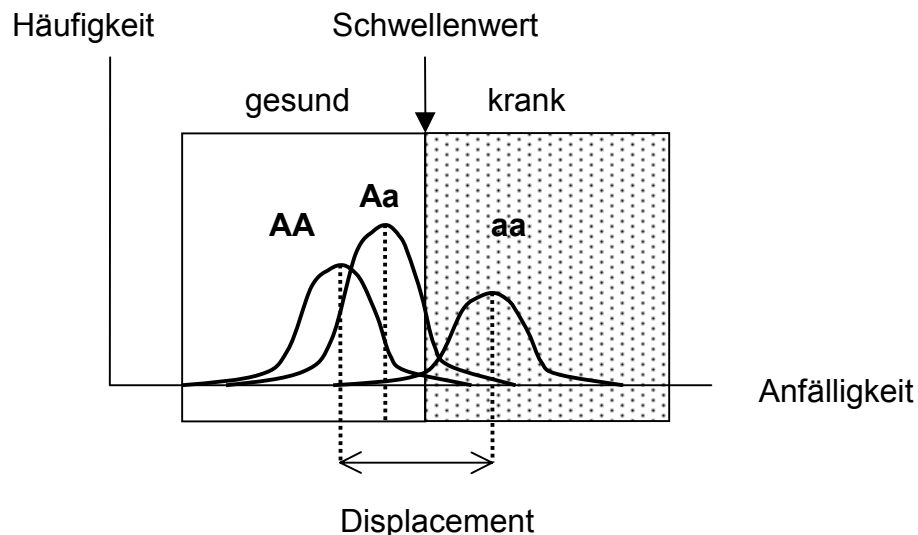
Bevor Modelle verglichen werden können müssen diese definiert werden. Folgende Modelle sollten geprüft werden:

- **Hauptgen**
- **Polygene Vererbung**
- Hauptgen und polygene Vererbung (**mixed inheritance**)
- Umwelt, keine genetische Komponente

Je nach Modell müssen verschiedene Parameter vorgegeben oder geschätzt werden. Zeigt es sich, dass ein Modell mit einer Hauptgenkomponente anderen Modellen überlegen ist, kann das **Familienmaterial** der **genetischen Kartierung** zugeführt werden. Dabei sind vor allem die geschätzten **Allelfrequenzen**, **Dominanzeffekt** und **Penetranzen** für die **Kopplungsanalyse** von Interesse.

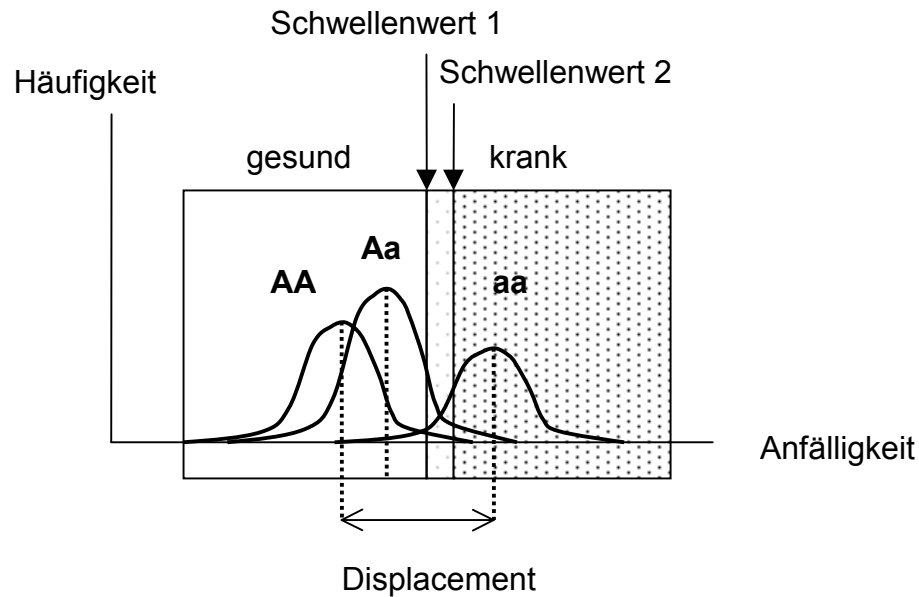
2.4.1.1. Hauptgen

Bei einem qualitativen Merkmal wird für jeden **Genotyp** eine normalverteilte Dichtefunktion auf einer kontinuierlichen Anfälligkeitsskala angenommen. Der **Phänotyp**, z.B. eine Krankheit, wird exprimiert sobald die Anfälligkeit (**Liability**) einen **Schwellenwert (threshold)** überschreitet. Der Schwellenwert wird so gelegt, dass die Fläche unter der Kurve rechts vom Schwellenwert dem Anteil der Tiere vom jeweiligen Genotyp entspricht, die den Phänotyp (z.B. krank) zeigen.



Der Hauptgeneffekt wird mit dem **Dominanzeffekt** und dem **Displacement** beschrieben. Der Dominanzeffekt besagt wo die Heterozygoten auf der Anfälligkeitsskala relativ zu den **Homozygoten** liegen und das Displacement gibt den Abstand zwischen den Verteilungen der Homozygoten in Standardabweichungen an. Dabei wird angenommen, dass die

Standardabweichungen für alle Genotyp-Verteilungen gleich sind. Lassen sich erkrankte Tiere noch weiter unterteilen, z.B. in leicht und schwer erkrankt, kann dies mit dem Einführen zusätzlicher **Schwellenwerte** Rechnung berücksichtigt werden.



2.4.1.2. Polygene Vererbung

Bei einem quantitativen Merkmal nimmt man an, dass die **Phänotypen** innerhalb eines **Genotyps** normalverteilt sind. Die Parameter des Modells sind der Mittelwert und die Standardabweichung für jeden Genotyp. Auch in diesem Modell können **Schwellenwerte** berücksichtigt werden. Dieses Modell wird geprüft um sicherzustellen, dass nicht nach einem **Hauptgen** gesucht wird, das nur sehr wenig zur Erklärung der beobachteten Phänotypen beiträgt.

2.4.1.3. Hauptgen und polygene Vererbung

Dieses Modell vereinigt die beiden vorangehenden Ansätze. Mit ihm man kann gleichzeitig Parameter für das **Hauptgen** und die polygene Komponente schätzen, allerdings ist es mit gängigen Computerprogrammen nicht möglich, den Beitrag eines Hauptgens zum **Phänotyp** abzuschätzen. Wenn die **Heritabilität** sehr klein wird, nähert sich dieses Modell dem Hauptgen-Modell. Eine hohe Heritabilität kann als Hinweis auf weitere Hauptgene gewertet werden.

2.4.1.4. Umwelt, keine genetische Komponente

Unter diesem Modell werden die Allele unabhängig vom **Genotyp** der Eltern weitergegeben, das heisst die Transmissionswahrscheinlichkeit für ein Allel ist gleich für alle Genotypen. Dieses Modell wird in die Evaluation miteinbezogen, um zu zeigen, dass tatsächlich eine genetische Komponente vorhanden ist. Sollte

wider Erwarten ein Umwelt-Modell besser oder gleich gut wie ein genetisches Modell abschneiden, bedeutet das, dass die Daten schlecht sind oder, dass der Anwender sich mit den Grundlagen der Genetik befassen sollte.

2.4.2. Interpretation der Vererbungsmodelle

Modelle, mit denen man unter den gleichen Voraussetzungen Parameter schätzt, können miteinander verglichen werden. Gleiche Voraussetzungen bedeutet, dass die Daten in der gleichen Form, z.B. zwei Klassen, gesund und krank, verwendet werden. Für jedes Modell kann dann eine Likelihood berechnet werden, die besagt wie wahrscheinlich das Modell unter den gegebenen Daten ist. Kann man sich aufgrund der Analysen für eines von mehreren Vererbungsmodellen entscheiden, heisst das nicht, das es das richtige Modell ist, sondern nur, dass es das beste der geprüften Modelle ist.

3. Kopplungsanalyse

Mit einer **Kopplungsanalyse** will man die genetische Distanz zwischen zwei **Loci** schätzen. Man unterscheidet zwischen parametrischen und nicht-parametrischen Ansätzen. Bei den ersteren müssen Parameter, wie Dominanzeffekte, **Allelfrequenzen** und **Penetranzen** bekannt sein, bei den letzteren werden bezüglich **Vererbungsmodus** keine Annahmen getroffen. Im Gegensatz zu parametrischen Methoden, brauchen nicht-parametrische Methoden kein vollständiges **Familienmaterial**, sondern kommen im Extremfall mit Vollgeschwisterpaaren aus. Grundsätzlich sollte immer ein vollständiges Familienmaterial angestrebt werden, denn dieses lässt sämtliche Analyse-Möglichkeiten zu. Steht ein vollständiges Familienmaterial mit sämtlichen Daten zur Verfügung, sind **parametrische Verfahren nicht-parametrischen** immer überlegen.

3.1. *Lod Score Methode*

Mit der **Lod Score** Methode wird gleichzeitig die **Rekombinationsrate** $\Theta < 0.5$ geschätzt und geprüft, ob diese von $\Theta = 0.5$ verschieden ist. Die grundlegenden Hypothesen sind also freie **Rekombination** (H_0) und **Kopplung** (H_1). Üblicherweise wird der Zehnerlogarithmus des Likelihoodquotienten als Entscheidungskriterium verwendet, ob Kopplung vorliegt oder nicht.

$$Z(\Theta) = \log_{10}[L(\Theta)/L(1/2)]$$

Wenn unsere Beobachtungen aus einer Paarung m rekombinate und n nicht-rekombinante **Nachkommen** umfassen, dann lässt sich der entsprechende **Lod Score** wie folgt berechnen:

$$Z(\Theta) = m \cdot \log(\Theta) + n \cdot \log(1-\Theta) + (m+n) \cdot \log(2), \text{ wenn } \Theta > 0$$

$$Z(\Theta) = (m+n) \cdot \log(2), \text{ wenn } \Theta = 0$$

wobei üblicherweise gilt $0 \leq \Theta \leq \frac{1}{2}$

Allerdings gibt es Computerprogramme, die $Z(\Theta)$ über den ganzen Bereich $0 \leq \Theta \leq 1$ maximieren, da $\Theta > \frac{1}{2}$ verbunden mit einem hohen Lod Score auf Fehler in den Daten deutet.

3.1.1. Entscheidungskriterien

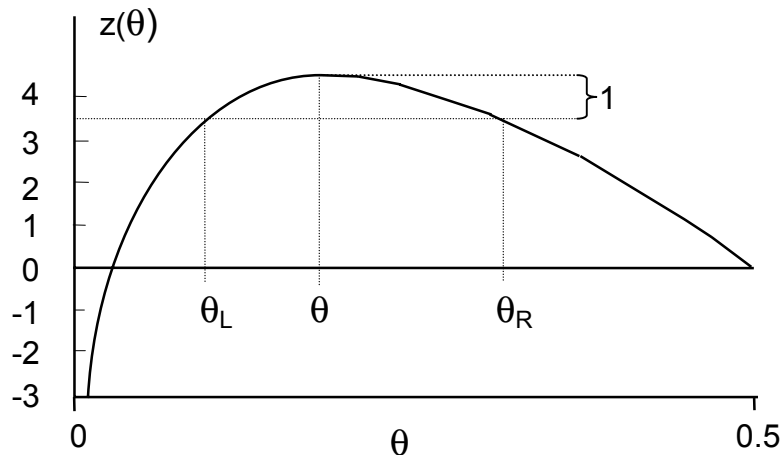
Morton hat 1955 das sequentielle Testen von Likelihoodquotienten für fixierte Rekombinationsraten für die **Kopplungsanalyse** eingeführt und die Signifikanzschwelle auf 0.001 und die Macht des Testes auf 0.99 gesetzt. Er schlug vor, immer weitere Familien zu untersuchen solange

$$Z_1 < Z(\Theta) < Z_0, \text{ wobei } Z_0 = 3 \text{ und } Z_1 = -2$$

$Z(\Theta)$ ist in diesem Fall der **Lod Score**, aufsummiert über alle Familien. Sobald $Z_0 = 3$ (im Fall von **X-chromosomalen Loci** $Z_0 = 2$) erreicht oder überschritten wird, wird die Hypothese der freien Rekombination zurückgewiesen. Diese wird aber angenommen, sobald $Z_1 = -2$ erreicht oder unterschritten wird. Diese Grenzwerte ergeben sich aus der Forderung nach einer Signifikanzschwelle von 0.001 und der Macht des Testes von 0.99, wobei auch eine a priori Wahrscheinlichkeit für **Kopplung** gegeben werden muss. Heute wird der sequentielle Test nicht mehr in seiner ursprünglichen Form durchgeführt. Man testet nicht mehr für verschiedene fixierte Werte der Rekombinationsrate, sondern maximiert $Z(\Theta)$ im Bereich der Rekombinationsrate. Aber auch heute werden immer noch, auch bei unseren Haus- und Nutztieren, die gleichen Testkriterien angewendet. Man kann zeigen, dass die mit den Lod Scores, $Z_0 = 3$ und $Z_0 = 2$ (im Fall von X-chromosomalen Loci), verbundenen Signifikantsschwellen nicht grösser als 10^{-3} , bzw. 10^{-2} , sind. Im Fall, wo der Lod Score über Θ maximiert wird Z_{\max} , ist die damit verbundene empirische Signifikantsschwelle höchstens $10^{-Z_{\max}}$.

3.1.2. Genauigkeit der geschätzten Rekombinationsrate

Die Genauigkeit der geschätzten **Rekombinationsrate** Θ kann mit einem **Supportintervall** angegeben werden. Das $Z_{\max} - 1$ Supportintervall findet man graphisch, indem man eine horizontale Linie 1 **Lod Score** Einheit unter Z_{\max} zieht. Die Voraussetzung ist allerdings, dass $Z_{\max} \geq 3$ ist, sonst ist ein Supportintervall nicht zulässig. Werden die Schnittpunkte der Horizontalen mit der Lod Score Kurve auf die Θ -Achse projiziert, erhält man die beiden Endpunkte Θ_L und Θ_R des Supportintervalls. Nimmt Z_{\max} zu, wenn sich Θ Null nähert, dann wird auf diese Weise nur Θ_R definiert, da $\Theta_L = 0$ wird.



Das **Supportintervall** umfasst alle Θ 's, bei denen Z nicht signifikant von Z_0 , dem kritischen Wert, verschieden ist. Mit mindestens 90% Wahrscheinlichkeit liegt das wahre Θ zwischen Θ_L und Θ_R . Bei einem komplexen **Familienmaterial** kann es hilfreich sein, den kombinierten **Lod Score** über alle Familien für Rekombinationswerte, die grösser sind als 0.5 zu betrachten. In einem Intervall $0 \leq \Theta \leq 1$ zeigt die Lod Score Kurve für Familien mit unbekannter Phase zwei Maxima, da nicht festgelegt ist, welcher Nachkommentyp der rekombinierte ist. Ist das Maximum im Bereich $\Theta > 0.5$ fast so gross wie das Maximum im Bereich $\Theta < 0.5$, bedeutet das, dass bei einem grossen Anteil der Familien die Phase nicht bekannt ist. Ist das Maximum im Bereich $\Theta > 0.5$ grösser als im Bereich $\Theta < 0.5$, bedeutet das, dass Fehler in den Daten vorliegen.

3.1.3. Wieviele Rekombinationsereignisse braucht es um Kopplung festzustellen ?

Wieviele informative **Nachkommen** braucht es, damit man bei einer wahren **Rekombinationsrate** von r mit einer hohen Wahrscheinlichkeit signifikante **Kopplung** ($Z_{\max} \geq Z_0$) findet ? Diese Anzahl hängt von der Familienstruktur, bzw. den Paarungstypen, aber auch von r selber, und weiteren Faktoren, ab. Je näher r bei 0 liegt, desto mehr Nachkommen braucht es, um überhaupt ein Rekombinationsereignis zu beobachten. Je näher r bei 0.5 liegt, desto schwieriger wird es die Alternativhypothese (Kopplung) von der **Nullhypothese** (keine Kopplung) zu unterscheiden. Mit einem **Familienmaterial**, mit 50 bis 200 Nachkommen und einer komplexen Struktur, wie es uns in der Tiergenetik üblicherweise zur Verfügung steht, lassen sich erfahrungsgemäss Rekombinationsraten im Bereich von etwa 5% bis 25% zuverlässig schätzen.

3.1.4. Faktoren, die die Rekombinationsrate beeinflussen

3.1.4.1. *Geschlecht*

Die **Rekombinationsrate** hängt vom **Geschlecht** ab, wobei das Verhältnis Θ_f / Θ_m je nach Chromosomenregion variieren kann. In Fällen wo zwischen Familien unterschieden werden kann, in denen nur der Vater bzw. die Mutter informativ ist,

können geschlechtsspezifische **Lod Scores** für geschlechtsspezifische Rekombinationsraten berechnet werden. Der kombinierte Lod Score berechnet sich dann als Summe der geschlechtsspezifischen Lod Scores, wobei Θ_f und Θ_m verschieden sind:

$$Z(\Theta_f, \Theta_m) = Z_m(\Theta_m) + Z_f(\Theta_f)$$

Im allgemeinen ist es einfacher einen hohen Lod Score zu finden, wenn man Unterschiede in Θ_f und Θ_m zulässt. Deshalb sollte man den kritischen Wert von 3 auf 3.3 anheben. Lässt sich eine Rekombination nicht dem Vater oder der Mutter zuordnen, sind Θ_f und Θ_m abhängig voneinander, d.h. sie können nicht unabhängig voneinander geschätzt werden.

3.1.4.2. *Alter*

Es gibt Hinweise, dass die **Rekombinationsrate** altersabhängig sein könnte. Diese Abhängigkeit scheint sich auf das männliche Geschlecht zu beschränken und mit zunehmendem Alter zu einer Abnahme von Θ_m zu führen.

3.1.4.3. *Heterogenität*

Ein **Phänotyp** ist genetisch heterogen, wenn bei verschiedenen Individuen der genetische Hintergrund unterschiedlich ist. In einigen Familien wird z.B. eine Erbkrankheit durch ein einziges **Gen** verursacht und in anderen Familien wird die scheinbar gleiche Krankheit durch Umweltfaktoren hervorgerufen. In der **Kopplungsanalyse** unterscheidet man zwischen allelischer **Heterogenität** und **Locus** Heterogenität. Allelische Heterogenität liegt dann vor, wenn am gleichen Locus verschiedene **Mutationen** vorliegen, die zur Krankheit führen. Im Normalfall kann allelische Heterogenität mit einer Kopplungsanalyse nicht festgestellt werden. Wird in einigen Familien die gleiche Krankheit durch andere Gene hervorgerufen als in anderen Familien, handelt es sich um Locus Heterogenität, die normalerweise mit der Kopplungsanalyse festgestellt werden kann. Oft lassen sich Familien nach bestimmten Kriterien, wie z.B. Herkunft oder Anzahl betroffener **Nachkommen**, in Klassen einteilen. Um zu Prüfen ob Heterogenität vorliegt vergleicht man die Likelihood der **Rekombinationsrate** unter Homogenität (Θ für alle Klassen gleich) mit derjenigen unter Heterogenität (Θ unterscheidet sich von Klasse zu Klasse). Oft lassen sich aber Familien nicht einfach in Klassen einteilen, und es ist sinnvoller anzunehmen, dass sich das **Familienmaterial** aus zwei Typen (Typ 1 und Typ 2) zusammensetzt, nämlich diejenigen, die **Kopplung** zeigen ($\Theta_1 < 0.5$), und diejenigen, die keine Kopplung zeigen ($\Theta_2 = 0.5$), wobei sich die Familien nicht eindeutig dem einen oder anderen Typ zuordnen lassen. Diese Situation findet man z.B. in einem Material, in dem bei einem Teil der Familien das für die Krankheit verantwortliche Gen neben dem **Marker** liegt, aber im anderen Teil der Familien ganz woanders liegt. Anstatt zu versuchen die Familien den beiden Typen zuzuordnen, analysiert man das gesamte Familienmaterial unter einem Modell, das ein Gemisch der beiden Typen zulässt, indem man z.B. den Anteil der Typ 1 Familien mit α bezeichnet, und denjenigen

der Typ 2 Familien mit $1 - \alpha$. Auch in diesem Fall lässt sich die Hypothese Homogenität gegen die Hypothese Heterogenität ($\alpha < 1$) testen.

3.1.5. Multipoint Analyse

Mit einer **Multipoint Analyse** will man gleichzeitig drei oder mehr **Loci** ordnen und die Rekombinationsraten zwischen benachbarten Loci schätzen. Die Multipoint Analyse ist vor allem wichtig für die Entwicklung von **Markerkarten** und kann entscheidend zur Verbesserung von Markertests beitragen. Im Unterschied zur **Kopplungsanalyse** mit zwei Loci spielt bei der Multipoint Analyse die lineare **Ordnung** der Loci eine Rolle, und die Anzahl der Parameter, **Haplotypen** und Genotypen steigt beträchtlich mit der Anzahl der involvierten Loci.

Die Anzahl der verschiedenen Haplotypen in Bezug auf die Anzahl der Loci ist gegeben durch das Produkt der Anzahl Allele dieser Loci

$$H = \prod n_i$$

Haben wir 5 loci mit 4, 7, 3, 9, und 6 Allelen, ergibt das 4536 Haplotypen. Daraus ergibt sich für jedes Individuum die Anzahl der möglichen kombinierten Genotypen, nämlich

$$H \cdot (H + 1) / 2$$

bzw. mit 4536 Haplotypen ergibt das 10289916 mögliche Genotypen. Diese sehr grossen Zahlen deuten darauf hin, dass das Berechnen einer Likelihood für verschiedene Haplotypen, bzw. Genotypen, an Programme und Computer höchste Ansprüche stellt. Die Anzahl der möglichen linearen Anordnungen von Loci ($n_{\text{Ordnungen}}$) ist gegeben durch die Anzahl der möglichen Permutationen ($n!$) dividiert durch 2, da spiegelbildliche Ordnungen sich nicht unterscheiden.

$$n_{\text{Ordnungen}} = n! / 2 = n \cdot (n - 1) \cdot (n - 2) \cdot \dots \cdot (3) \cdot (2) \cdot (1) / 2$$

Bei 3 Loci ergibt das 3 mögliche Ordnungen, bei 4 Loci 12 mögliche Ordnungen, und bei 5 Loci schon 60 mögliche Ordnungen. Bei der **Multipoint Analyse** unterscheidet man drei Möglichkeiten Ordnungen statistisch abzusichern, bei denen analog zum **Lod Score**, wo nur zwei Loci involviert sind, jeweils ein Likelihood Quotient gebildet wird, nämlich

$$\log (L_1 / L_2)$$

Je nach Support sind L_1 und L_2 verschieden definiert. Der **Global Support** besagt ob ein **Locus** überhaupt in eine **Kopplungsgruppe** gehört. L_1 ist die Maximum Likelihood, wenn der Locus in die bestehende **Ordnung** eingeführt wird und L_2 die Likelihood, wenn der Locus nicht in diese Ordnung gehört. Ein Locus gehört in die bestehende Ordnung, wenn der Global Support 3 oder grösser ist. Der Intervall Support besagt ob ein Locus in einem bestimmten Intervall einer bestehenden Ordnung liegt. L_1 ist die maximum Likelihood, wenn der Locus in diesem Intervall liegt und L_2 die Likelihood, wenn der Locus in irgendeinem anderen Intervall liegt.

Will man den Support für eine bestimmte Ordnung berechnen, ist L_1 die maximum Likelihood für die beste Ordnung, und L_2 die maximum Likelihood für diese bestimmte Ordnung. Loci, die einen Intervall Support von mindestens 3 haben, werden **Framework** Loci genannt. Eine Karte, die nur Framework Loci umfasst, wird Framework Karte genannt. Im Gegensatz dazu wird eine Karte **Comprehensive** genannt, wenn sie alle syntänischen Loci, ungeachtet ihrem Support, umfasst. Oft wird auch neben den drei oben erwähnten Supports ein **Generalized Lod Score** berechnet.

$$\text{Log}(L_1 / L_0) = \log[L(\Theta_{12}, \Theta_{23}, \dots, \Theta_{n-1, n}) / L(\frac{1}{2}, \frac{1}{2}, \dots, \frac{1}{2})]$$

Dieser Generalized Lod Score, der besagt, dass eine Gruppe von **Loci** gekoppelt ist, ist nicht einfach zu interpretieren, da die verschiedenen Rekombinationsraten voneinander nicht unabhängig sind.

Für die verschiedenen Positionen x , die ein Locus in einer Karte einnehmen kann, lässt sich ein kartenspezifischer Multipoint Lod Score berechnen.

$$Z(x) = \log[L(x) / L(\infty)]$$

Mit x wird irgendeine Position in einer definierten Karte bezeichnet. Mit $x = \infty$ wird angedeutet, dass der Locus nicht Teil dieser Karte ist, d.h. $\Theta = \frac{1}{2}$ für diesen Locus und jeden anderen Locus in dieser Karte. Der **Location Score** $S(x)$ stellt eine Transformation von $Z(x)$ dar.

$$S(x) \approx 4.6 \cdot Z(x)$$

Sobald eine **Kopplungsgruppe**, bzw. eine Karte, etwas weiter entwickelt ist, ist es nicht mehr sinnvoll einen **Generalized Lod Score** zu berechnen, da jedesmal, wenn ein neuer **Locus** eingebettet wird, auf eine bestehende lineare **Ordnung**, nämlich die Karte, zurückgegriffen wird.

3.2. Interpretation von Resultaten der Kopplungsanalyse

3.2.1. Zwei-Punkt Analyse

In der Regel wird bei einer **Zwei-Punkt Analyse** die **Rekombinationsrate** mit dem maximalen **Lod Score** angegeben. Da der Lod Score aber nicht alle Eigenschaften der Daten reflektiert ist eine schlüssige Interpretation dieser Angaben nicht möglich. Es wird deshalb empfohlen, Lod Scores für fixierte Werte von Θ anzugeben, nämlich bei 0, 0.001, 0.05, 0.1, 0.2, 0.3, und 0.4. Aus diesen Angaben liesse sich der Verlauf der Lod Score Kurve ableiten. Dies wird jedoch meistens nicht gemacht. Eine weitere Möglichkeit wäre, das Vertrauensintervall mit Θ_L und Θ_R anzugeben. Für die Interpretation von Zwei-Punkt Analyse Resultaten ist am wichtigsten, ob die angegebene Rekombinationsrate durch weitere, unabhängige Analysen bestätigt sind oder nicht. Dabei geht es nur um die Grössenordnung von Θ , z.B. 5% oder 10%, da Θ ja von verschiedenen Faktoren, wie z.B. **Geschlecht** und Individuum, beeinflusst wird.

3.2.2. Multipoint Analyse

Das wichtigste Resultat einer **Multipoint Analyse** ist die relative **Ordnung** der **Loci**. Wie zuverlässig sich Loci ordnen lassen hängt vom **Familienmaterial** und der relativen Lage der Loci zueinander ab. Liegen die zu ordnenden Loci sehr nahe oder sehr weit auseinander, wird es schwierig eine beste Lösung zu finden, die sich deutlich von der nächst besten Lösung unterscheidet. In solchen Situationen findet man oft mehrere gleich gute Lösungen. Auch hier gilt, dass eine Ordnung erst dann als gesichert gilt, wenn sie durch weitere, unabhängige Analysen bestätigt werden kann. Die Distanzen zwischen den Loci müssen vorsichtig interpretiert werden, denn auch hier werden sie von verschiedenen Faktoren, wie z.B. **Geschlecht** und Individuum beeinflusst. Die Distanzen zwischen den Loci lassen sich nicht aus der Addition der Rekombinationsraten aus den **Zwei-Punkt Analysen** herleiten, da zwischen zwei Loci multiple Crossovers auftreten können. Ob ein Locus Teil einer **Framework** Karte ist, oder nur Teil einer **Comprehensive** Karte, hängt auch davon ab, welche Loci den Ausgangspunkt für das Ordnen bilden.

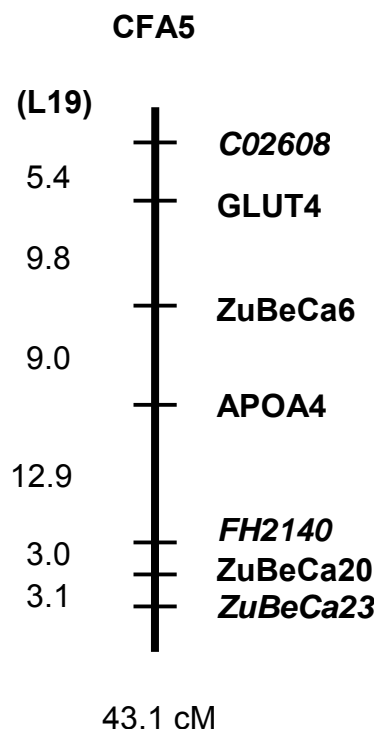
3.3. *Nicht-parametrische Methoden*

Bei einer **Kopplungsanalyse** muss der **Vererbungsmodus** der **Loci** bekannt sein. Es gibt aber Situationen, in denen es nicht möglich ist, sich für einen **Erbgang** zu entscheiden, oder ein Untersuchungsmaterial eine parametrische Analyse nicht zulässt. Es gibt aber Methoden, die eine Kopplungsanalyse erlauben, ohne dass ein Erbgang definiert werden muss. Nicht-parametrische Methoden werden vor allem in der Humangenetik eingesetzt, wo es vorkommen kann, dass wohl erkrankte Individuen zur Verfügung stehen, aber nicht ein vollständiges **Familienmaterial**. Von den verschiedenen Methoden sollen hier lediglich die **Affected Sib-Pair Methoden** erwähnt werden. Dabei geht man von folgender Hypothese aus: Ist ein Locus mit einem für eine Krankheit verantwortlichen **Gen** eng gekoppelt, wird mit dem krankmachenden Allel des Gens immer auch ein bestimmtes Allel des Markers an die **Nachkommen** weitergegeben. Man untersucht also, wie oft zwei erkrankte Vollgeschwister dasselbe Markerallel von den Eltern erhalten haben. Ein solches Markerallel wird "identical by descent", **IBD**, vererbt. Indem nur kranke Individuen analysiert werden, hat man keine Probleme mit **Penetranzen**, muss aber davon ausgehen, dass keine **Heterogenität** vorliegt. Zwei erkrankte Vollgeschwister können 0, 1, oder 2 Allele gemeinsam haben. Liegt keine **Kopplung** vor erwartet man diese Ereignisse (0, 1, oder 2) im Verhältnis 1 : 2 : 1. Bei Kopplung erwartet man eine grössere Anzahl von gemeinsamen Allelen IBD. In der Tiergenetik kommen nicht-parametrische Verfahren eher selten zur Anwendung, da es einfacher ist, ein Familienmaterial zusammenzustellen, als in der Humangenetik. Ein Familienmaterial hat den Vorteil, das es sowohl mit parametrischen, als auch mit nicht parametrischen Methoden analysiert werden kann. Lässt sich ein Erbgang genau definieren, sind parametrische den nicht-parametrischen Methoden überlegen.

4. Markerkarten

4.1. Erstellen von Markerkarten

Moderne **Markerkarten** beinhalten Daten sowohl aus der physischen wie auch aus der **genetischen Kartierung**. Am Anfang kann eine Karte aus einzelnen Syntänie- und **Kopplungsgruppen** bestehen. Indem einzelne **Loci** chromosomal zugewiesen werden, können auch die betreffenden Syntänie- und Kopplungsgruppen bestimmten **Chromosomen** zugewiesen werden. Die dabei auftretenden Widersprüche zwischen genetischer und physischer Kartierung können mit einer zunehmenden Anzahl von Informationen bereinigt werden. Mit der Weiterentwicklung einer Karte wird einmal ein Zustand erreicht, bei dem die Anzahl der Kopplungsgruppen genau der Anzahl der Chromosomen im haploiden Zustand entspricht.

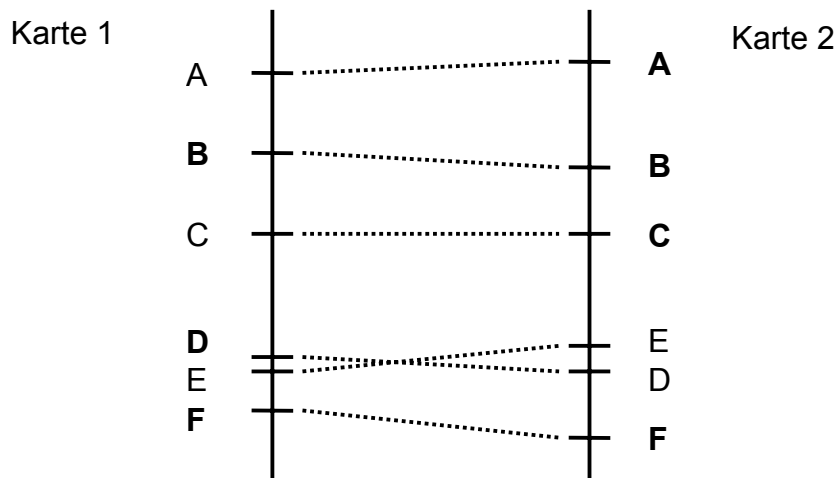


Das oben gezeigte Hundechromosom 5 ist der DogMap **Markerkarte** entnommen. In den DogMap Referenzfamilien wurden alle gezeigten **Loci** genetisch kartiert. Ausserdem wurden ZuBeCa6, ZuBeCa20 und ZuBeCa23 mit **FISH** chromosomal zugewiesen, und zwar auf 5q14, 9q26 und 17q17. Hier stehen also genetische und physische Daten im Widerspruch. Wir haben diese **Kopplungsgruppe** trotzdem CFA5 zugewiesen weil ein anderes Labor GLUT4 ebenfalls CFA5 zugewiesen hat. Da GLUT4 und ZuBeCa6 nicht nur genetisch, sondern auch physisch relativ nahe beieinander liegen, ist es noch nicht möglich zuverlässig das Zentromerende des Chromosomes zu bestimmen. Die normal- und fettgedruckten

Loci bilden hier die **Framework** Karte mit einem **Global Support** > 3. Zusammen mit den schräg- und fettgedruckten ($3 > \text{Global Support} > 2$) Loci bilden sie eine **Comprehensive** Karte von CFA5. Der Widerspruch zwischen der physischen und **genetischen Kartierung** der ZuBeCa Marker wird bereinigt werden, indem sie auch in einem RH-Panel kartiert werden. In dieser Karte wurde die **Ordnung** der Loci mit einer **Multipoint Analyse** festgelegt, wobei es hier keine Rolle spielt mit welchem Framework Markerpaar begonnen wird. Die Distanzen sind allerdings nicht mit der Multipoint Analyse geschätzt worden, sondern es wurden die Daten der **Zwei-Punkt Analyse** verwendet. Die Überlegung dabei ist, dass Zweipunkt Analyse Ergebnisse einfacher zu reproduzieren sind, da Multipoint Distanzen in der Regel vom Vorgehen beim Ordnen abhängig sind. Die Rekombinationsraten wurden in Kosambi-cM umgewandelt, was üblich, aber eigentlich falsch ist. Bei der **Kosambi** Kartierungsfunktion werden Annahmen bezüglich Interferenz getroffen, die man aber in der Regel nicht auf Daten abstützen kann. Richtigerweise müsste man die Haldane Kartierungsfunktion verwenden, da diese keine Interferenz zulässt. Dies wird aber nicht gemacht, da die Karten mit Kosambi-cM kompakter werden.

4.2. Integrierte Markerkarten

Unter **integrierten Markerkarten** versteht man das gleichzeitige Darstellen von verschiedenen Markerkarten, wobei üblicherweise gemeinsame **Loci** mit einem Strich verbunden werden.



Die obenstehende, konstruierte, integrierte Markerkarte besteht aus Karte 1 und Karte 2, in denen die **Framework Marker** jeweils fettgedruckt sind. Handelt es sich bei beiden um genetische Karten, können auch die Distanzen miteinander verglichen werden. Unterschiede in den Distanzen können verschiedene Ursachen haben. Am ehesten dürften Unterschiede im Umfang des **Familienmaterials** und in der Geschlechterverteilung der informativen **Nachkommen** dafür verantwortlich sein. Leider werden in der Regel dazu keine erklärenden Angaben in den Publikationen oder **Datenbanken** gemacht. Karte 1 und Karte 2 stimmen recht gut überein, mit Ausnahme der Plazierung von **Locus E**. Da beide Karten den Locus E nicht mit genügender Sicherheit einordnen

können, heisst das, dass E nahe bei D liegt, wobei man nicht weiss auf welcher Seite.

4.3. Präsentation von Markerkarten

Der Entwicklungsstand der Karten bei unseren Nutz- und Haustieren ist im Vergleich zu Mensch und Maus bescheiden. Am weitesten vorangeschritten sind die Karten beim Rind und beim Schwein. Diejenigen beim Hund und der Katze stehen noch in den Anfängen. Erste einfache Karten wurden bei allen Spezies immer zuerst publiziert. Es hat sich aber gezeigt, dass Zeitschriften nicht das geeignete Medium sind, um **Markerkarten** zu präsentieren. Zum einen nehmen Markerkarten sehr schnell einen Umfang an, der zu riesigen Manuskripten führt, und zum andern dauert es zu lange bis solche Manuskripte schliesslich publiziert werden. Deshalb werden heute Markerkarten fast ausschliesslich über das World Wide Web (**WWW**) präsentiert. Das hat den Vorteil, dass jederman sofort Zugriff auf die Daten hat, und dass solche Karten jederzeit auf den neuesten Stand gebracht werden können. Ein nicht zu unterschätzender Nachteil ist, dass ein Peer Review entfällt, sodass die Qualitätssicherung bei den Daten allein den zuständigen Editoren überlassen ist.

Die Form der Präsentation im WWW fällt für verschiedene **Markerkarten** recht unterschiedlich aus. Die Palette reicht von solchen, die lediglich mit HTML und evt. Java programmiert sind, bis zu solchen, die direkten Zugriff auf relationale **Datenbanken** erlauben. Der Unterschied zwischen diesen beiden Extremen liegt darin, dass bei den ersteren bei jeder Änderung oder Erweiterung die betroffenen Webseiten ergänzt oder neu gestaltet werden müssen. Bei Datenbanken genügt dagegen eine Änderung der betreffenden Daten oder ein Neueintrag. Allerdings ist die Entwicklung und der Unterhalt von Datenbanken recht aufwendig.

Auch in Bezug auf den Informationsgehalt der verschiedenen Präsentationen von Markerkarten gibt es beträchtliche Unterschiede. Datenbanken umfassen in der Regel eine grafische Darstellung der Markerkarten für einzelne **Chromosomen** oder **Kopplungsgruppen** und erlauben es gezielt nach den verschiedensten Informationen zu suchen, wie z.B. nach **Loci**, Chromosomen, **Primer**, Literatur, Methode der Kartierung oder Referenzen. Einfachere Präsentationen beschränken sich in der Regel auf die absolut notwendigen Informationen und verweisen zum Teil auf zusätzliche Informationen, die auf anderen Web Seiten einzusehen sind.

4.4. Kontrollmöglichkeiten

Sucht man bestimmte Informationen in einer **Markerkarte**, die im **WWW** angeboten wird, stellt sich immer die Frage, wie zuverlässig diese sind. Man sollte also, wenn immer möglich, die gefundenen Angaben überprüfen. Besonders in Situationen, in denen man die gefundenen Informationen in der eigenen Arbeit einbinden möchte, ist eine sorgfältige Kontrolle der Daten notwendig. Dabei stehen bei anonymen Markern, wie **Mikrosatelliten**, die Position in der Karte und die Methode, mit der diese kartiert wurden, im Vordergrund. Bei Genen und **ESTs**

kommt noch die Verifikation der Identität hinzu. Vor allem bei der vergleichenden Kartierung ist es unabdingbar die Homologien zu überprüfen.

4.4.1. Überprüfen der Position eines Locus

Wird ein **Locus** in einer **Markerkarte** als "confirmed" präsentiert, heisst das, dass dieser Locus in mindestens zwei unabhängigen Experimenten mit dem gleichen Resultat kartiert wurde. Ist ein Locus nur in einem Experiment kartiert worden, wird er als "provisional" angegeben. In einem solchen Fall muss man andere Karten überprüfen, ob sie diesen Locus nicht auch schon enthalten. Will man einen Locus als **Marker** für einen **Phänotyp** verwenden, um einen **Markertest** zu entwickeln, muss man auf jeden Fall den Marker in einem **Familienmaterial** überprüfen, in dem dieser Phänotyp segregiert. Die Rekombinationsraten müssen möglichst zuverlässig geschätzt werden, damit Testergebnisse korrekt interpretiert werden können.

4.4.2. Methode der Kartierung

Will man einen kartierten **Locus** in einer eigenen Arbeit verwenden, muss man überprüfen, ob die Angaben zur Kartierungsmethode korrekt sind. Am häufigsten schleichen sich Fehler bei der Angabe von Sequenzdaten ein. Vor allem bei Primersequenzen empfiehlt es sich, auf die Originalarbeit zurückzugreifen. Es kann auch sein, dass für einen Locus die Primersequenzen zwar korrekt wiedergegeben sind, aber das Primerpaar nicht optimal für die Analyse dieses Locus ist. In einem solchen Fall kann man anhand der "**accession number**" in die entsprechende Sequenz-Datenbank gehen und die Primersequenzen aufgrund der deponierten **DNA-Sequenz** überprüfen. Allenfalls müssen die **Primer** auch neu gestaltet werden. Leider wird nicht in allen Fällen die zu einem Locus gehörende DNA-Sequenz in einer Sequenz-Datenbank deponiert.

4.4.3. Verifikation der Identität eines Locus

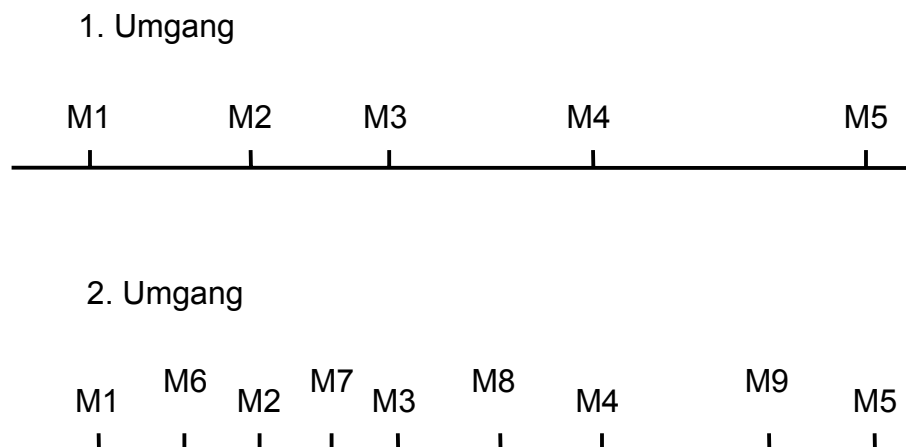
Es kann vorkommen, dass ein anonymer **Locus** in einer **Markerkarte** verwechselt wurde, indem der Locus zwar korrekt kartiert wurde, aber nicht der Locus ist, für den man ihn hält. Weit wichtiger ist aber das überprüfen der Identität von **ESTs** und Genen. Um zu zeigen, dass ein EST oder **Gen** z.B. beim Hund dasselbe ist wie beim Schwein, muss man die Position und die Sequenz überprüfen. Die Position wird anhand von Informationen aus **Chromosome Painting** Experimenten und der Position von benachbarten ESTs und Genen überprüft. Ein Vergleich der betreffenden DNA-Sequenzen zwischen den beiden Spezies sollte eine **Homologie** von mindestens 70% aufweisen. Es gilt zu beachten, dass auch eine hohe Homologie nicht als Beweis gilt, dass ein Gen in einer Spezies die gleiche Funktion ausübt, wie in der anderen Spezies. Bei der Angabe von Homologien muss immer darauf geachtet werden, ob diese auf der Funktion, der **DNA-Sequenz**, oder der Aminosäuresequenz beruhen. Eine hohe DNA-Sequenz Homologie zusammen mit einer übereinstimmenden Position ist aber ein guter Hinweis, dass es sich um dasselbe Gen oder EST handeln könnte.

5. Suche nach einem Gen

In der Tiergenetik bleibt die Suche nach bestimmten Genen oft in der Nähe solcher **Gene** stecken. Das hängt vor allem mit den beschränkten Mitteln, aber auch damit zusammen, dass oft ein guter **Marker** ausreicht, um ein Problem, wie z.B. eine Erbkrankheit, züchterisch zu bearbeiten. Je nach **Phänotyp**, den man untersuchen möchte, bieten sich verschiedene Möglichkeiten, ein Gen zu finden, oder in dessen Nähe zu gelangen. In jedem Fall braucht man aber dazu die verschiedenen Phänotypen in Form von Familien oder Tiergruppen, wie z.B. eine Gruppe krank und eine Gruppe gesund.

5.1. *Whole Genome Scan*

Ist man an einem **Phänotyp** interessiert, der in einer anderen Spezies nicht kartiert ist oder nicht kartiert werden kann, muss man nach dem verantwortlichen **Gen** im ganzen Genom suchen. Möchte man z.B. das Gen für Behornung beim Rind kartieren, sucht man vergeblich nach diesem Gen in Karten beim Mensch oder der Maus. In Situationen, in denen man nicht auf Informationen von anderen Spezies zurückgreifen kann, muss man ein **Familienmaterial** sammeln, in welchem der interessierende Phänotyp segregiert. Zeigt eine **Segregationsanalyse**, dass ein **Hauptgen** mit definiertem **Erbgang** vorhanden ist, kann man einen **Whole Genome Scan** durchführen. Dabei werden anhand bestehender **Markerkarten polymorphe Marker** ausgewählt, die möglichst gleichmässig über das autosomale Genom verteilt sind. Würde der Phänotyp, bzw. das Hauptgen, geschlechtsgebunden vererbt, würde sich ein Genome Scan auf die Geschlechtschromosomen beschränken. In einem ersten Umgang der Typisierung des Familienmaterials werden die Marker so ausgewählt, dass die Distanzen zwischen ihnen z.B. möglichst 20 cM betragen. In einem zweiten Umgang versucht man weitere Marker möglichst in der Mitte von jeweils zwei Markern zu plazieren, sodass sich ein durchschnittlicher Abstand von 10 cM ergibt. Das Netz der Marker wird mit jedem Durchgang enger gelegt, bis man eine oder mehrere Regionen identifiziert hat, die mit einer **Zwei-Punkt Kopplungsanalyse** eine bestimmte **Lod Score** Schwelle erreicht oder überschritten haben.



Ein Problem, speziell bei unseren Haus- und Nutztieren, ist, dass die **Markerkarten** zu wenig **Marker** aufweisen, um die gewünschten Distanzen genau einzuhalten. Sogar bei Karten vom Rind und vom Schwein gibt es immer noch grössere Lücken. Deuten die Untersuchungsergebnisse darauf hin, dass das gesuchte **Hauptgen** in einer solchen Lücke liegen könnte, müssen zusätzliche Marker entwickelt werden. Ein anderes, allgemeines Problem, liegt darin, dass bei einem **Genome Scan** mehrfach getestet wird. Je mehr Marker man typisiert, desto wahrscheinlicher wird es, dass die festgelegte **Lod Score** Schwelle auch rein zufällig erreicht oder überschritten wird. Deshalb muss die Lod Score Schwelle in Abhängigkeit von der Anzahl untersuchter Marker nach oben korrigiert werden. Eine elegante Art, diese Korrektur für das mehrfache Testen, ist das Bereitstellen eines zweiten, unabhängigen **Familienmaterials**, in dem nur noch die wenigen Marker typisiert werden, bei denen die festgelegte Lod Score Schwelle erreicht oder überschritten wurde. Findet man einen Marker, der tatsächlich mit dem Hauptgen gekoppelt ist, müssen weitere Marker in der unmittelbaren Umgebung auch gekoppelt sein.

5.2. Kandidaten Gene

Kandidaten Gene sind Gene, die bei anderen Spezies, wie z.B. Mensch oder Maus, kartiert sind, und von denen man aufgrund der zur Verfügung stehenden Informationen glaubt, dass sie dem gesuchten Gen **homolog** sind. Zeigt es sich, z.B. bei einer Erbkrankheit, dass beim Hund und beim Mensch die klinischen und pathologischen Befunde beinahe identisch sind, liegt die Vermutung nahe, dass diese Krankheit den gleichen genetischen Hintergrund hat. In einem solchen Fall kann man direkt nach dem vermuteten Gen suchen. Dabei kann man das kartierte Gen als Sonde in einer DNA-Library einsetzen, um einen Klon zu identifizieren, der das homologe Gen aufweist. Man kann auch die Sequenz des kartierten Gens dazu verwenden, **Primer** zu entwickeln, die kodierende Regionen amplifizieren. Damit kann man direkt das homologe Gen im Erbgut suchen, ohne dieses vorher zu klonieren. Auf jeden Fall müssen gefundene Klone oder **PCR-Produkte** sequenziert werden, um einerseits die Sequenz-Homologie zu etablieren, und andererseits die für die Krankheit verantwortliche **Mutation** aufzuzeigen. In einem ersten Schritt braucht man also lediglich kranke und gesunde Individuen. In einem zweiten Schritt ist ein **Familienmaterial** notwendig, um zu zeigen, dass die gefundene Mutation auch tatsächlich mit dem **Phänotyp** segregiert. In der Regel wird aber einer Untersuchung mit Kandidaten Genen ein **Genome Scan** vorangestellt, denn oft stehen mehrere Kandidaten Gene zur Verfügung, was diesen Ansatz sehr aufwendig macht. Mit einem Genome Scan ist es möglich Kandidatenregionen zu identifizieren, die es erlauben, zusammen mit **Chromosome Painting** Informationen, eine Selektion unter den Kandidaten Genen zu treffen.

6. Index

accession number	23
Affected Sib-Pair Methoden	19
Allelfrequenz	11, 13
autosomal	7, 8, 9, 10
Chromosome Painting	23, 25
Chromosom	4, 5, 6, 20, 22
codominant	8
Comprehensive	18, 19, 21
Datenbank	21, 22
Diagnose	6
Displacement	11
DNA-Sequenzierung	4, 23
dominant	7, 8
Dominanzeffekt	11
Donor	5
Erbgang	7, 8, 9, 10, 19, 24
EST	22, 23
Expressivität	10
Familie	7, 8, 9
Familienmaterial	5, 6, 11, 13, 15, 16, 19, 21, 23, 24, 25
Fiber-FISH	5
FISH	4, 20
Framework	18, 19, 21
Gen	5, 8, 10, 16, 23
Generalized Lod Score	18
genetische Kartierung	5, 6, 11, 20, 21
Genkarte	4
Genome Scan	24, 25
Genotyp	6, 11, 12
Geschlecht	15, 18, 19
Global Support	17, 21
Haplotyp	17
Hauptgen	11, 12, 24, 25
Heritabilität	12
Heterogenität	16, 19
heterozygot	5
Homologie	4, 23, 25
homozygot	11
Hybridzelllinie	5
IBD	19
In Situ Hybridisation	4
integrierte Markerkarte	21
intermediär	7, 8, 10
Irrtumswahrscheinlichkeit	10
Kandidaten Gen	25
komplexe Merkmale	10

Kopplung	13, 14, 15, 16, 19
Kopplungsanalyse	5, 11, 13, 14, 16, 17, 19
Kopplungsgruppe	17, 18, 20, 22
Kosambi	21
Liability	11
Location Score	18
Locus	4, 5, 6, 13, 14, 16, 17, 18, 19, 20, 21, 22, 23
Lod Score	5, 13, 14, 15, 16, 18, 24, 25
Loop	7
Marker	4, 5, 16, 21, 23, 24, 25
Markerkarte	4, 17, 20, 21, 22, 23, 24, 25
Markertest	23
Merkmal	10
Mikrosatelliten	22
mixed inheritance	11
Multipoint Analyse	17, 19, 21
Mutation	4, 16, 25
Nachkomme	5, 7, 8, 9, 13, 15, 16, 19, 21
nicht- parametrische Methoden	13
Nukleotide	4
Nullhypothese	5, 15
Ordnung	17, 18, 19
parametrische Methoden	13
PCR	25
Penetranz	10, 11, 13, 19
Phänokopie	6, 8
Phänotyp	4, 6, 7, 8, 9, 10, 11, 12, 16, 24, 25
physische Kartierung	4
Polygene Vererbung	11
Polymorphismus	4, 6, 24
Primer	22, 23, 25
Radiationshybriden	6
Rekombination	5, 6, 13
Rekombinationsrate	5, 13, 14, 15, 16, 18
rezessiv	7, 8
Rezipient	5
Schwellenwert	11, 12
Segregationsanalyse	6, 24
somatische Zellhybridisation	5
Stammbaum	7, 8, 9, 10
Supportintervall	14, 15
Syntäniegruppe	5
threshold	11
Vererbungsmodus	6, 10, 13, 19
WWW	22
X-chromosomal	8, 14
Zwei-Punkt Analyse	18, 19, 21, 24

7. Aufgaben für das 3. Jahr

Im 3. Jahr sollen drei Seminare durchgeführt werden, in denen jeweils ein Thema vorgestellt und diskutiert wird. Ein Seminar dauert 1 $\frac{3}{4}$ Stunden, wovon maximal die Hälfte für das Vorstellen des Themas verwendet werden soll. Das Vorbereiten der Präsentationen soll in Absprache mit dem ITZ erfolgen. Entsprechende Termine werden mit den Personen, die ein Thema bearbeiten, vereinbart.

7.1. *Web-Sites, die für die Genomanalyse bei Haus- und Nutztieren interessant sind*

Zusammenstellen von Web-Sites, die für die Genomanalyse bei unseren Haus- und Nutztieren unentbehrlich oder interessant sind. Die Sites sollen nach Art der Informationen und Art der Präsentation gruppiert werden. Anhand von einem Beispiel soll gezeigt werden, welche Sites konsultiert werden sollen, um ein Projekt zu bearbeiten, in dem ein Marker oder ein Gen für eine Erbkrankheit gefunden werden soll. Anhand ausgewählter Beispiele soll die inhaltliche und gestalterische Qualität von solchen Web-Sites beurteilt werden.

7.2. *Kupfertoixikose beim Bedlington Terrier*

Die Erbkrankheit Kupfertoixikose soll klinisch und pathologisch, mit Schwergewicht Hund, beschrieben werden. Die Entwicklung eines Markertests beim Bedlington Terrier soll vorgestellt und die Anwendung dieses Tests in der Zucht beurteilt werden. Es soll gezeigt werden, wie Test-Resultate zu interpretieren sind.

7.3. *Manuskript Genomanalyse*

Das vorliegende Manuskript soll kritisch beurteilt werden, in Bezug auf das Lernziel, den Inhalt und die Präsentation. Positive und negative Punkte aufführen und Verbesserungsvorschläge machen.