

Familien-basierte Assoziationsanalyse mit Haplotypen -  
eine Simulationsstudie zum Vergleich statistischer  
Methoden

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Meinhard Mende

aus

Burgstädt

Bonn, März 2009

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät  
der Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter PD Dr. Michael Knapp
2. Gutachter Prof. Dr. Wolfgang Alt

Tag der mündlichen Prüfung: 09. September 2009

Diese Dissertation ist auf dem Hochschulschriftenserver der ULB Bonn unter

[http://hss.ulb.uni-bonn.de/diss\\_online](http://hss.ulb.uni-bonn.de/diss_online)

elektronisch publiziert.

Erscheinungsjahr: 2009

# Inhaltsverzeichnis

<b>1</b>	<b>Einführung</b>	<b>7</b>
1.1	Grundbegriffe . . . . .	8
1.1.1	Einige genetische Grundlagen . . . . .	8
1.1.2	Wichtige Grundbegriffe der genetischen Epidemiologie . . . . .	10
1.2	Einordnung und Abgrenzung des Themas . . . . .	14
1.2.1	Wert und Probleme Familien-basierter Studien . . . . .	14
1.2.2	Geschichtlicher Abriss . . . . .	16
1.2.3	Eingrenzung der genetisch-statistischen Methoden . . . . .	16
<b>2</b>	<b>Planung der Simulation</b>	<b>18</b>
2.1	Wozu Simulationsverfahren? . . . . .	18
2.2	Modellbildung für die Simulation . . . . .	18
2.2.1	Vorüberlegungen . . . . .	18
2.2.2	Modelle für Alternativhypothesen . . . . .	19
2.2.3	Modelle für die Nullhypothese . . . . .	23
2.2.4	Modelle für fehlende Genotypen . . . . .	24
2.2.5	Modelle für Stratifikation . . . . .	24
2.3	Funktionsweise des Programms zur Simulation der Familiendaten . . . . .	26
2.3.1	Mathematischer Hintergrund . . . . .	26
2.3.2	Implementierung als SAS-Makro . . . . .	27
<b>3</b>	<b>Auswahl der Assoziationsanalyse-Verfahren</b>	<b>28</b>
3.1	Recherche . . . . .	28
3.2	Kurzvorstellung der Methoden und Programme . . . . .	29
3.2.1	Übersicht über die Verfahren . . . . .	29
3.2.2	Multimarker-Tests . . . . .	30
3.2.3	TDT-Erweiterungen . . . . .	30
3.2.4	FBAT . . . . .	31
3.2.5	Auf bedingter Likelihood basierende Ansätze . . . . .	31
3.2.6	Log-lineare Modelle . . . . .	32
3.2.7	Genie . . . . .	32
3.2.8	Sichtung weiterer Programme . . . . .	32
3.2.9	Programme der engeren Wahl . . . . .	34
<b>4</b>	<b>Charakteristika der Programme für die Studie</b>	<b>35</b>
4.1	FBAT . . . . .	35
4.1.1	Grundlagen . . . . .	35
4.1.2	Multi-Marker-Test . . . . .	36
4.1.3	Haplotyp-Test . . . . .	36
4.1.4	Permutationstest . . . . .	37
4.1.5	Test auf Assoziation bei vorhandener Kopplung . . . . .	37
4.1.6	Weitere Optionen . . . . .	37

4.1.7	Ausgabe von FBAT	38
4.2	Famhap	38
4.2.1	Grundlagen	38
4.2.2	Algorithmus	40
4.2.3	Test auf Kopplung	40
4.2.4	Optionen	40
4.2.5	Famhap-Ausgabe	41
4.3	Hotel_Fam	41
4.3.1	Grundlagen	41
4.3.2	Optionen	42
4.3.3	Ausgabe von Hotel_Fam	43
4.4	Unphased	43
4.4.1	Modellannahmen	43
4.4.2	Verfahren	43
4.4.3	Test auf Kopplung	45
4.4.4	Optionen	45
4.4.5	Ausgabe von Unphased	45
4.5	Übersicht zur Ausgabe der vier Programme	46
<b>5</b>	<b>Realisierung und Kontrolle der Simulation</b>	<b>47</b>
5.1	Ablauf	47
5.2	Probelaufe	47
5.2.1	FBAT	49
5.2.2	Famhap	50
5.2.3	Hotel-Fam	50
5.2.4	Unphased	51
5.2.5	Abschätzung der Laufzeiten	52
5.2.6	Test der Synchronizität	52
5.3	Administration der Jobs auf Cluster	53
5.3.1	Start der Rechenjobs	53
5.3.2	Wiederaufnahme abgebrochener Jobs	54
5.3.3	Abschluss eines Jobs	54
5.4	Modifikationen für Familiendaten mit fehlenden Genotypen	56
5.4.1	Veränderungen des Ablaufs	56
5.4.2	Erzeugung der Daten	57
5.5	Analyse in stratifizierten Populationen	57
<b>6</b>	<b>Auswertung der Analysen</b>	<b>58</b>
6.1	Laufzeitbetrachtungen	58
6.2	Validität unter der Nullhypothese mit kompletten Genotyp-Daten	59
6.2.1	Ergebnisse für einzelne Markerkombinationen (Mk)	60
6.2.2	Test auf Kopplung	62
6.2.3	Einfluss des Fehlerniveaus	62
6.2.4	Overall-p-Wert über alle Markerkombinationen	63
6.2.5	Besonderheiten	65
6.3	Power der Tests	69
6.3.1	Power für einzelne Markerkombinationen	69
6.3.2	Tests auf Kopplung	73
6.3.3	Power der verschiedenen Statistiken	74
6.3.4	Vergleich der overall-Power über alle Markerkombinationen	76
6.3.5	Ergebnisse für $\alpha = 0.01$	80
6.4	Resultate mit fehlenden Genotypen	83
6.4.1	Validität der Tests	83
6.4.2	Overall-Power in der Analyse mit fehlenden Genotypen	85

6.5	Ergebnisse unter Stratifikation . . . . .	88
6.6	Weiterführende Analysen . . . . .	90
6.6.1	FBAT: Anzahl Permutationen . . . . .	90
6.6.2	Die Optionen 'minfreq' und 'rare' . . . . .	91
6.6.3	Vergleich verschiedener Familienstrukturen . . . . .	93
6.6.4	Verteilung der p-Werte - Klassifikation der Verfahren . . . . .	97
6.7	Anwendung der vier Verfahren auf reale Familiendaten . . . . .	100
<b>7</b>	<b>Zusammenfassung und Ausblick</b>	<b>102</b>
7.1	Kritische Anmerkungen . . . . .	102
7.2	Zusammenfassung . . . . .	104
7.3	Ausblick . . . . .	105
	<b>Literaturverzeichnis</b>	<b>106</b>
	<b>Abbildungsverzeichnis</b>	<b>117</b>
	<b>Tabellenverzeichnis</b>	<b>119</b>
	<b>Anhang</b>	<b>120</b>
<b>A</b>		<b>121</b>
A.1	Übersicht über zusätzliches Material . . . . .	121
A.2	Zusätzliche Tabellen . . . . .	122
A.3	Abkürzungsverzeichnis . . . . .	139



# Kapitel 1

## Einführung

Aber jetzt, so spricht der HERR, der dich geschaffen ... hat:  
„Weil du teuer bist in meinen Augen und wertvoll bist  
und ich dich liebhabbe ...“ [3]

Die Würde des Menschen ist unantastbar. [5]

Die Einmaligkeit und der unermessliche Wert jedes Menschen war den Gründungsvätern der Bundesrepublik Deutschland als Leitgedanke so wichtig, dass sie damit vor 60 Jahren das Grundgesetz einleiteten. Dieser uralte Gedanke hat auf den noch jungen Wissenschaftszweig 'Genetische Epidemiologie' gleich zweifache Auswirkung: Einerseits motiviert er Forscher, nach Ursachen von Krankheiten zu suchen, um ihnen vorzubeugen, sie zu heilen oder Betroffenen helfen zu können. Andererseits steckt die Würde des Menschen den Rahmen ab, in dem sich Ziel und Methoden jeder Forschung bewegen müssen. Es ist gut sich dessen im wissenschaftlichen Alltag bewusst zu sein.

Die Entwicklung genetisch-statistischer Verfahren wurde von Erkenntnissen der Zellgenetik, Biochemie und medizinischen Genetik voran getrieben. Nachdem viele monogenische Krankheiten erfolgreich durch Kopplungsmethoden aufgeklärt wurden, sind bei komplexen Krankheiten Assoziationsverfahren stärker in den Fokus gerückt. Diese suchen nach Allelen, die gemeinsam mit Krankheitsmerkmalen gehäuft auftreten. Da jede Methode für eine bestimmte Situation konzipiert wurde und von spezifischen Modellannahmen ausgeht, entfaltet sie in diesen Situationen individuelle Stärken. Wie aber ist ihr Verhalten, wenn diese Bedingungen verletzt sind? Um die eher mittleren bis schwachen genetischen Risikoeffekte bei komplexen Krankheiten zu entdecken, bedarf es passend gewählter, ausgereifter Algorithmen. Dazu dienen vergleichende Studien, wie sie z. B. kürzlich Nikodemus et al. [117] für Familien-basierte Single-Marker-Methoden durchgeführt haben. Für Haplotyp-basierte Methoden gibt es bis jetzt einen solchen Vergleich nicht. Vorliegende Arbeit ist dazu gedacht, diese Lücke teilweise zu schließen:

Eine Simulationsstudie soll die Eigenschaften Familien-basierter Assoziationsverfahren mit Haplotypen offenlegen. Damit erhält der Anwender Hinweise, welche Methode er in welcher Situation optimal einsetzen kann. D. h. mit vorliegender Studie wird ein Teil des umfangreichen Instrumentariums der genetischen Epidemiologie gesichtet, geordnet und geschärft. Nach einer kurzen Vorstellung von Grundbegriffen werden die in der Simulation verwendeten Modelle erläutert. Aus einer Übersicht über Assoziationsprogramme werden die für die Studie passenden Verfahren ausgewählt und vorgestellt. Ein Kapitel beschäftigt sich mit Einzelheiten der Simulations-Durchführung und -Überwachung. Schließlich werden die Ergebnisse der Analysen dargestellt und interpretiert und Hinweise für Anwender gegeben, wie die Methoden optimal einzusetzen sind. Ein Ausblick rundet die Studie ab.

# 1.1 Grundbegriffe

## 1.1.1 Einige genetische Grundlagen

Nach NE Morton [113] beschäftigt sich die genetische Epidemiologie mit der Ätiologie und Verteilung von Krankheiten in Gruppen von Verwandten sowie mit vererbten Krankheitsursachen in Populationen. Es geht also z. B. darum, wie genetische Varianten das Risiko von Krankheiten beeinflussen. Das waren in der Vergangenheit Varianten im HLA (*human leukin antigen*)-System, Mikrosatelliten oder RFLPs (*repeat fragment length polymorphisms*) und sind heute vor allem SNPs (*single nucleotide polymorphisms*), d. h. Varianten eines Nucleotids, die mit einer Häufigkeit von mindestens 1% vorkommen. Ihre Ausprägungen - bei SNPs sind das die vier organischen Basen A, C, G und T - werden Allele genannt.

### Chromosomen, Allele, Genotypen und Haplotypen

Jeder Mensch gehört mit 23 Paaren homologer Chromosomen zu den *diploiden* Lebewesen:  $2 \times 22$  Chromosomen, die *Autosomen*, kommen paarig vor. Das gilt für weibliche Personen auch hinsichtlich der Geschlechtschromosomen: Sie besitzen 2 X-Chromosomen. Männliche Personen dagegen haben ein X- und ein Y-Chromosom.

Moderne Genotypisierungstechniken liefern i. d. R. nur den Genotyp, d. h. ein Paar  $G = (A_1/A_2)$  von Allelen. An *diallelischen* Loci, das sind Genorte mit zwei Allelen A und B, kann es also drei verschiedene Genotypen geben: A/A, A/B und B/B. Stimmen die beiden Allele eines Locus überein, ist das Individuum dort *homozygot*, andernfalls *heterozygot*.

Je ein Tripel aus drei Basen codiert jeweils eine bestimmte Aminosäure. Somit verschlüsselt die Abfolge der Basen auf einem Chromosom die Synthese spezifischer Proteine. Ob zwei Allele benachbarter Loci auf demselben Chromosom liegen oder nicht, kann also die biologische Funktion entscheidend beeinflussen. Bei gegebenen Genotypen mehrerer Loci kann man nicht ohne weiteres erkennen, welche auf einem Chromosom und welche auf dem anderen liegen. Mit anderen Worten, man weiß nicht, welche von der Mutter und welche vom Vater vererbt wurden. Man spricht von *unbestimmter Phase*. Labormethoden zur Phasenbestimmung sind auch heute noch aufwändig und teuer, so dass man alternativ dazu durch statistische Verfahren die Wahrscheinlichkeiten für verschiedene Haplotypen ermittelt. Die Anzahl möglicher Haplotyperklärungen aber wächst exponentiell mit der Anzahl heterozygoter Loci. Hat eine Person beispielsweise an drei Markern die Genotypen (A/C,G/T,C/T), so gibt es hier  $2^2 = 4$  verschiedene Möglichkeiten, diese auf zwei Chromosomen anzuordnen. Das ist in Abb. 1.1 verdeutlicht:

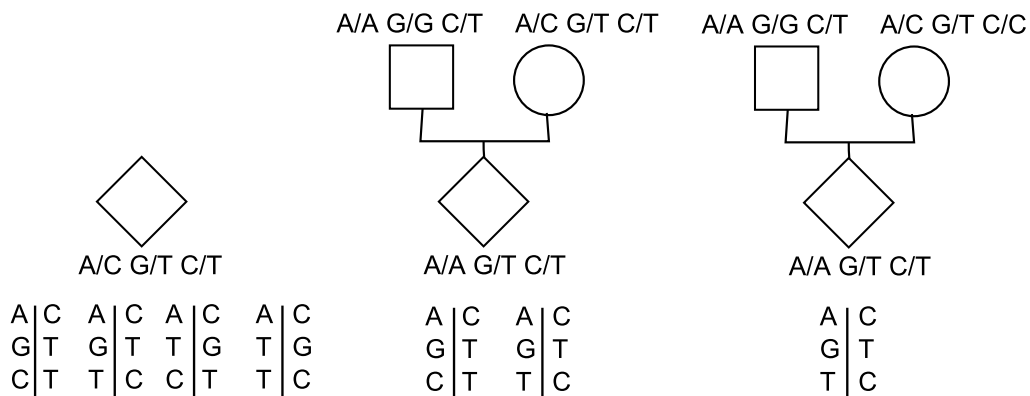


Abbildung 1.1: Beispiel: Phasenaufklärung ohne und mit elterlicher Genotypinformation

Links sieht man die vier möglichen Haplotyperklärungen. Angenommen, die Genotypen der Eltern seien wie in Bildmitte gegeben. Weil der Vater (dargestellt durch das Quadrat oben) an den ersten beiden Loci homozygot ist, überträgt er von dort den Haplotyp (A,G). Damit gibt es



nur noch zwei mögliche Haplotyperklärungen. Bei der Familie rechts hingegen ist zusätzlich die Mutter (durch einen Kreis repräsentiert) am dritten Locus homozygot, so dass die Phase eindeutig bestimmt ist. D. h. das Vorhandensein elterlicher Genotyp-Informationen ermöglicht viel besser, die Phase aufzulösen, als das mit den individuellen Genotypen allein möglich wäre.

Seien A und B die beiden Allele eines diallelischen Locus. Dann kombinieren sich die drei Genotypen (AA, AB und BB) in der Population zu sechs verschiedenen *mating types* (MT): AA × AA, AA × AB, AA × BB, AB × AB, AB × BB, BB × BB. Vorausgesetzt ist, dass nicht zwischen mütterlichen und väterlichen Genotypen unterschieden werden muss. Bezeichnen  $p_A$  bzw.  $p_B$  die Häufigkeiten der Allele A und B in der Population, so ergeben sich die Häufigkeiten der Genotypen bei der Modellvorstellung einer *random union of gametes*:

$$P(AA) = p_A^2 \quad P(AB) = 2 \cdot p_A \cdot p_B \quad P(BB) = p_B^2, \quad (1.1)$$

Diese Relation heißt *Hardy-Weinberg-Gleichgewicht* (HWE) nach den beiden Autoren, die sie vor 100 Jahren unabhängig voneinander veröffentlichten. Bevölkerungs-Fluktuation (*genetic drift*) erzeugt z. B. ein Ungleichgewicht. Eine Verletzung des HWE kann aber auch auf systematische Fehler bei der Genotypisierung hinweisen.

### Crossover, Rekombination und LD

Bei der Bildung von Keimzellen (Meiose) werden durch *crossover*-Ereignisse homologe Teile der Chromosomen auf einem Chromosom neu arrangiert. Das ist schematisch in Abb. 1.2 verdeutlicht:

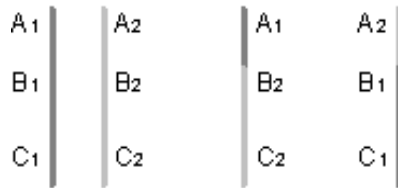


Abbildung 1.2: Crossover zwischen dem 1. und 2. Locus

Links sind homologe Chromosomen mit symbolischen Allelen  $A_i, B_i, C_i$  ( $i = 1, 2$ ) an den Loci  $M_1 M_2 M_3$  dargestellt. Bei der Meiose fand zwischen  $M_1$  und  $M_2$  ein Crossover statt, so dass in verschiedenen Keimzellen (rechts)  $A_2, B_1$  und  $C_1$  (bzw.  $A_1, B_2$  und  $C_2$ ) auf je einem Haplotyp zu finden sind. Von *Rekombination* zwischen zwei Loci spricht man bei einer ungeraden Anzahl von Crossovers.

Das Kopplungsungleichgewicht (*linkage disequilibrium*, LD) misst die Abhängigkeit zwischen Allelen zweier Loci. Seien  $M_1$  und  $M_2$  diallelische Loci mit Allelen  $A_i, B_i$  ( $i = 1, 2$ ) mit den Häufigkeiten  $p_{A_i}$  bzw.  $p_{B_i}$ . Bei Unabhängigkeit ergäben sich die Häufigkeiten der Haplotypen einfach als Produkt der Einzelhäufigkeiten (Kopplungsungleichgewicht):

$$P(A_i B_j) = p_{A_i} \cdot p_{B_j} \quad i, j \in \{1, 2\} \quad (1.2)$$

Die Abweichungen von diesem Produkt nennt man Kopplungs-Ungleichgewicht, also

$$D_{ij} = P(A_i B_j) - p_{A_i} \cdot p_{B_j} \quad i, j \in \{1, 2\} \quad (1.3)$$

Bei diallelischen Loci sind durch ein LD  $D = D_{11}$  in Bezug auf den ersten Haplotypen  $A_1 B_1$  die Abweichungen für die drei anderen Haplotyp-Frequenzen (s. Tab. 1.1) festgelegt:

Dabei bedeutet ein positives LD  $D_{11} = P(A_1 B_1) - p_{A_1} \cdot p_{B_1}$ , dass die Allele  $A_1$  und  $B_1$  häufiger gemeinsam auftreten, als das bei Unabhängigkeit der Loci zu erwarten wäre. Nach Slatkin [142] trifft deshalb die Bezeichnung *non-random association* besser den Kern der Sache.

Die Definition des LD für mehrallelische Marker erfolgt analog. Da das LD stark von den Allelfrequenzen abhängt, normierte Lewontin D mittels Division durch den maximal möglichen

		A	
		A <sub>1</sub>	A <sub>2</sub>
B	B <sub>1</sub>	$p_{A_1} \cdot p_{B_1} + D$	$p_{A_2} \cdot p_{B_1} - D$
	B <sub>2</sub>	$p_{A_1} \cdot p_{B_2} - D$	$p_{A_2} \cdot p_{B_2} + D$

Tabelle 1.1: Kopplungs-Ungleichgewicht (LD)

Absolutbetrag [106]:

$$\text{Mit } D_{Max} = \begin{cases} \min(p_{A_1} \cdot p_{B_2}, p_{A_2} \cdot p_{B_1}) & \text{falls } D_{11} > 0 \\ \min(p_{A_1} \cdot p_{B_1}, p_{A_2} \cdot p_{B_2}) & \text{falls } D_{11} < 0 \end{cases} \quad (1.4)$$

$$\text{definiert man } D' = \frac{D_{11}}{D_{Max}} \quad (1.5)$$

Damit zeigt  $D' = 1$  für obige zwei diallelische Marker A und B an, dass einer der vier möglichen Haplotypen nicht vorkommt [142].

Auch das alternative LD-Maß  $r^2$  (z. B. [45], [142]) leitet sich als Korrelationsmaß vom obigen D ab:

$$r^2 = \frac{[P(A_1B_1) \cdot P(A_2B_2) - P(A_1B_2) \cdot P(A_2B_1)]^2}{p_{A_1} \cdot p_{B_1} \cdot p_{A_2} \cdot p_{B_2}} \quad (1.6)$$

Rekombinationsereignisse schwächen das LD ab, und zwar umso stärker, je größer die Rekombinationsfrequenz  $\theta$  ist. Sei in Abb. 1.2 links in Generation  $t$  ein positives LD  $D_{11}(t)$  zwischen Marker A und B angenommen: Dann verringert Rekombination zwischen A und B diese Häufigkeit und das LD sinkt:

$$D_{11}(t+1) = (1 - \theta) \cdot D_{11}(t) \quad (1.7)$$

Die Rekombinationsfrequenz variiert sehr über das Genom. Bereiche mit hohem LD, sog. *LD-Blöcke* werden durch *recombination hotspots*, Gebiete mit hoher Rekombinationsneigung, getrennt (z. B. [41], [62], [118]).

## 1.1.2 Wichtige Grundbegriffe der genetischen Epidemiologie

### Monogenische und komplexe Krankheiten

Bei der Entstehung einer vererbaren Krankheit geht man oft von folgendem Modell aus: Im Laufe der Geschichte hat an einem spezifischen Genort eine Mutation stattgefunden, die das Risiko für diese Krankheit erhöht. Der Austausch einer Base an einem Locus der DNA könnte z. B. die Synthese eines anderen Proteins hervorrufen, wodurch im Körper ein wichtiges Protein fehlt. Dieser Genort wird deshalb Krankheitslocus (*disease locus*, DL) bzw. Risikolocus genannt. Krankheiten, wie z. B. Cystische Fibrose oder Sichelzellanämie, die sich auf Varianten in einem Gen zurückführen lassen, bezeichnet man deshalb als *monogenisch*. Da der DL nicht bekannt ist, versucht man ihn durch vorgegebene Genorte, Marker, schrittweise zu lokalisieren.

Angenommen am DL gibt es das Risikoallel D in einer Häufigkeit  $p$  und das *wild type allele* d mit der Häufigkeit  $1 - p$ . Die Penetranz  $f_n$  ist als bedingte Wahrscheinlichkeit definiert, an der betreffenden Krankheit zu erkranken, wenn man  $n$  Kopien des Allels D besitzt:

$$f_n = P\{K | n \text{ Kopien des Allels } D\} \quad (n = 0, 1, 2). \quad (1.8)$$

Dabei bezeichne  $K$  bzw.  $\bar{K}$  das Ereignis krank bzw. gesund zu sein. Dann berechnet sich die *Prävalenz*, die Häufigkeit der Krankheit in der Population, nach der Formel ([153], S.71):

$$K_P = p^2 f_2 + 2p(1 - p) f_1 + (1 - p)^2 f_0, \quad (1.9)$$

Die Stärke des genetischen Effekts kann man z. B. durch folgende Kenngrößen angeben:

Das *Risikoverhältnis* ist der Faktor, um den das Risiko von Personen mit Genotyp  $g = (d/D)$  bzw.  $g = (D/D)$  gegenüber Personen ohne Risikoallel (d. h. mit Genotyp  $g = (d/d)$ ) erhöht ist:

$$\psi_1 = P[K|g = (d/D)] / P[K|g = (d/d)] \quad (1.10)$$

$$\psi_2 = P[K|g = (D/D)] / P[K|g = (d/d)] \quad (1.11)$$

Das *odds ratio* (OR, Chancenverhältnis) definiert man z. B. für das dominante Modell (s. u.) für Personen mit Risikoallel gegenüber solchen ohne:

$$OR = \frac{P[K|g \in \{(d/D), (D/D)\}] \cdot P[\bar{K}|g = (d/d)]}{P[K|g = (d/d)] \cdot P[\bar{K}|g \in \{(d/D), (D/D)\}]} \quad (1.12)$$

(Dabei wird in den Formeln 1.10 bis 1.12 eine positive *Phänokopie*-Rate vorausgesetzt, das ist die Häufigkeit, mit der Personen ohne Risikoallel erkranken.)

Man unterscheidet verschiedene Vererbungsmodi (*modes of inheritance*, MOI):

**dominant**, abgekürzt **DOM** : Die Krankheit tritt bereits auf, wenn nur eine Kopie des Allels D vorkommt. D. h. für die Penetranzen gilt:  $f_2 = f_1 \gg f_0$

**rezessiv**, **REZ** : Die Krankheit tritt nur auf, wenn das Risikoallel D homozygot vorkommt, d. h.  $f_2 \gg f_1 = f_0$

**additiv**, **ADD** : Die Penetranzen unterscheiden sich um eine konstante Differenz:  $f_1 = \frac{f_0 + f_2}{2}$

**multiplikativ**, **MUL** : Die Penetranzen unterscheiden sich um einen konstanten Faktor:

$$f_1 = \sqrt{f_2 \cdot f_0}$$

Monogenische Krankheiten zeichnen sich durch eine hohe Penetranz am DL aus. Ein Beispiel ist eine rezessive Krankheit, die dadurch determiniert ist, wenn die Person homozygot mit dem Risikoallel ist. Dann gilt für die Penetranzen:  $f_2 = 1, f_1 = f_0 = 0$ . Bei monogenischen Krankheiten überwiegt der genetische Anteil an der Entstehung und ihre Prävalenz ist i. a. niedrig.

Dem gegenüber stehen viele komplexe Krankheiten, die in der Bevölkerung recht häufig vorkommen, deren Entstehung man aber nicht allein auf die Variation eines Gens zurückführen kann. Vielmehr wird vermutet, dass genetische Varianten an mehreren Loci das Krankheitsrisiko erhöhen und weitere Faktoren zum Ausbruch der Krankheit beitragen. Das resultiert in moderaten oder schwachen Penetranzen (bzw. Risikoeffekten), aber auch in einer großen Uneinheitlichkeit des Phänotyps. So sind obige Vererbungsmodi in allgemeinerer, abgeschwächter Form definiert: Im Beispiel der rezessiven Krankheit kann es sein, dass nicht alle Personen mit zwei Kopien des Risikoallels erkranken (*unvollständige Penetranz*). Andererseits gibt es kranke Personen, die gar kein Risikoallel besitzen (*Phänokopie*). So komplex die Entstehung von Krankheiten ist, so vielgestaltig sind die genetisch-statistischen Verfahren, die im Laufe der Zeit für die Analyse entwickelt wurden. Sie reichen von einfachen  $\chi^2$ -Tests [130] bis hin zu ausgeklügelten Verfahren wie *random forests* oder der Anwendung neuronaler Netze [114].

## Kopplung und Assoziation

Der Name genetische Epidemiologie bringt zum Ausdruck, dass hier zwei Ströme zusammenfließen [26]: Diese Wissenschaftsdisziplin ist eng mit der klassischen Epidemiologie verbunden und ihr Fokus liegt auf familiären, insbesondere genetischen Faktoren von Krankheit und deren gemeinsamen Effekten mit nicht-genetischen Faktoren. So spiegeln die wichtigen Paradigmen Kopplung und Assoziation den genetischen bzw. epidemiologischen Ansatz wider.

Die Lokalisation von Genorten (*gene mapping*) basiert auf dem Schlüsselgedanken, dass ein Risikoallel gemeinsam mit Marker-Varianten von Generation zu Generation weiter gegeben wird. Dieses biologische Konzept findet seinen Niederschlag im Begriff der

**Kopplung (*linkage*)** Zwei genetische Loci sind gekoppelt, wenn sie von Eltern an Nachkommen häufiger gemeinsam weitervererbt werden als das bei unabhängiger Vererbung wahrscheinlich wäre. Das ist gleichbedeutend damit, dass die Loci so benachbart liegen, dass die Rekombinationsfrequenz  $\theta < \frac{1}{2}$  ist.

Geprüft wird folglich die Hypothese  $H_0 : \theta = \frac{1}{2}$  gegen die Alternative  $H_1 : \theta < \frac{1}{2}$ .

Der Lokalisierung von Risiko-Genorten durch Linkage-Analysis ist mit einer Auflösung in der Größenordnung von 1 bis 10 cM [24] eine natürliche Grenze gesetzt: In Regionen enger Kopplung werden kaum noch Rekombinationen beobachtet. Außerdem variiert z. T. die Rekombinationsrate auf engem Raum sehr [15]. So würde eine weitere Verfeinerung durch Kopplungsmethoden unmöglich hohe Stichprobenumfänge erfordern [128]. Nach Tabor et al. [152] setzt auch die phänotypische und kausale Heterogenität komplexer Krankheiten der Kopplungsanalyse Grenzen. Assoziationsverfahren dagegen besitzen größeres Potential als Kopplungsmethoden, Genotyp-Effekte zu erkennen.

Das Konzept Assoziation repräsentiert die statistische Seite der genetischen Epidemiologie:

**Assoziation (*association*)** Allele eines bestimmten Locus sind mit einem Phänotyp (z. B. Krankheit) assoziiert, wenn sie in einer Population häufiger unter den Merkmalsträgern auftreten, als unter Personen, die nicht dieses Krankheitsmerkmal aufweisen [161].

Eine spontane Mutation an einem Locus erzeugt LD (und damit Assoziation) zwischen diesem Genort und benachbarten. Arbeiten, wie z. B. [36], [142], untersuchen genauer, unter welchen Bedingungen Assoziation entstehen kann. Balding [15] gibt darüber hinaus eine Übersicht über Ansätze, wie man diese Assoziation zu entdecken sucht.

Man unterscheidet verschiedene Arten von Assoziation:

***direct association*** Ein Allel am Markerlocus verursacht ein erhöhtes Krankheitsrisiko, d. h. der Marker ist selbst ein DL. Das ist eher ein theoretischer Fall, da die Position des Risikolocus erst bestimmt werden soll.

***indirect association*** Steht ein Marker in hinreichend starkem LD zum Krankheitslocus, wird man dort ebenfalls Assoziation zur Krankheit beobachten. Diese Assoziation steht im Focus von 'Linkage-Disequilibrium-Methoden'.

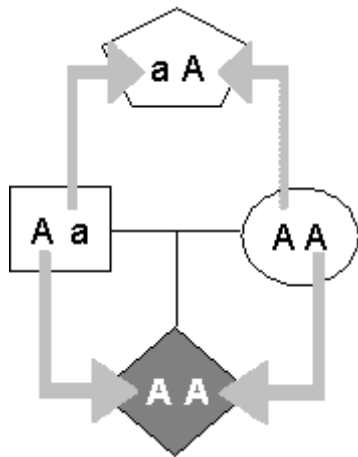
***spurious association*** Viele Artikel bezeichnen mit diesem Begriff Assoziation, die nichts mit der Krankheit zu tun hat. Ursachen dafür können z. B. Bevölkerungsschichtung (*population stratification*), Selektion, Inzucht oder *Bottleneck*-Ereignisse (wie z. B. Kriege oder Epidemien [142]) sein.

Der einfachste Ansatz für Tests auf Assoziation ist eine Fall-Kontroll-Studie. D. h. Allelhäufigkeiten werden zwischen Erkrankten und unverwandten Kontrollen verglichen. Dieses Vorgehen aber steht und fällt mit einer geeigneten Wahl der Kontrollgruppe. Weil sich Menschen verschiedener Nationalitäten in ihren Allelfrequenzen oft deutlich unterscheiden, ergibt sich das Problem der Stratifikation, siehe Absatz 2.2.5.

Familien-basierte Verfahren [24] waren ein Ansatz, dem Problem der Populationsstratifikation zu begegnen: In Abb. 1.3 ist als typische Familienstruktur ein Trio aus Indexfall (Rhombus unten) und seinen Eltern dargestellt. Dabei bezeichnen A und a die Allele des diallelischen Markers M. Falk und Rubinstein folgten nun der Idee, aus den nicht übertragenen Allelen der Eltern passende Pseudo-Kontrollen zu bilden [55], im Bild das Fünfeck oben. In einer Vierfeldertafel wird für Fälle und Kontrollen erfasst, ob das Allel A vorhanden ist bzw. nicht. Damit kann man die *Haplotype Relative Risk*-Statistik (HRR) berechnen:

$$HRR = \frac{c \cdot f}{d \cdot e} \quad (1.13)$$

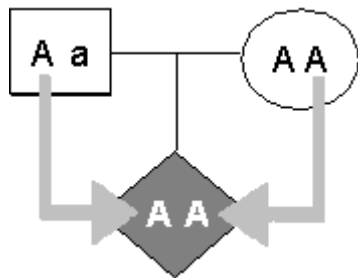
Weiterhin lässt sich ein  $\chi^2$ -Test auf der Basis der Vierfeldertafel durchführen.



	Allel A vorhanden	
	+	-
Kranke	c	d
Pseudo-Kontrollen	e	f

Abbildung 1.3: Haplotype Relative Risk

Auch Spielman et al. [145] gingen beim 1993 publizierten TDT (*transmission disequilibrium test*) von den Genotypen kranker Personen und deren Eltern aus. Sie entwickelten den TDT als Test auf Kopplung zwischen einem Marker M und dem DL [54]: Sei M ein Marker mit Allelen A und a im LD mit dem Risikolocus. Sind Markerlocus und DL ungekoppelt, so übertragen heterozygote Eltern mit jeweils Wahrscheinlichkeit 1/2 die Allele A sowie a an das kranke Kind. In einer Vierfeldertafel wird für jeden elterlichen Genotyp erfasst, welches Transmissionsmuster vorliegt: Z. B. überträgt in Abb. 1.4 der Vater das A-Allel, nicht aber das a-Allel. Bei der homozygoten Mutter ist die Übertragung determiniert: ein A-Allel wird übertragen, das andere nicht. D. h. der Vater liefert den Beitrag '1' für die *u*-Zelle der 2x2-Tafel, die Mutter dagegen für die *s*-Zelle.



		übertragen	
		A	a
nicht	A	s	t
	a	u	v

Abbildung 1.4: TDT: Vergleich der Anzahl übertragener / nicht übertragener Allele A auf das kranke Kind

Die zum McNemar-Test äquivalente [180] Teststatistik

$$T_{TDT} = \frac{(t - u)^2}{t + u} \quad (1.14)$$

hat unter der Nullhypothese den Erwartungswert 0 und ist asymptotisch  $\chi^2$ -verteilt mit einem Freiheitsgrad. Diese Statistik verwendet nur die Häufigkeiten der Nebendiagonalen. Mit anderen Worten: Homozygote Eltern sind für den TDT nicht informativ. Spielman et al. [145] erweiterten ihn auch auf mehrere kranke Kinder.

Für einen DL und einen Markerlocus M gibt es die in Abb. 1.5 dargestellten Möglichkeiten: Die Bereiche I und II kennzeichnen 'keine Kopplung', d. h. DL und M rekombinieren frei ( $\theta = 1/2$ ). Dann erfolgt die Übertragung von Markerallelen und Risikoallel unabhängig. In diesen Fällen wird der TDT nur zufällig signifikant sein.

Ein signifikanter TDT dagegen zeigt an, dass DL und M sowohl gekoppelt sind als auch im LD stehen (IV).

Am kompliziertesten ist die Situation III: Besteht zwischen Marker und DL Kopplung, aber kein LD, so werden die M-Allele von den Eltern an das Kind in erwarteten Häufigkeiten übertragen.

In Familien mit mehreren kranken Kindern allerdings ist die Allelübertragung von Eltern auf die Geschwister nicht unabhängig. Weil der Term  $(t + u)$  in (1.14) nur unter der Nullhypothese -  $H_0$ : keine Kopplung - ein unverfälschter Varianzschätzer für  $(t - u)$  ist, fungiert der TDT in diesem Fall als Test auf Kopplung [145].

		LD	
		nein	ja
Kopplung	nein	I	II
	ja	III	IV

Abbildung 1.5: Kopplung und Assoziation (LD)

Da ein signifikanter TDT sowohl Kopplung als auch Assoziation zwischen Marker und DL anzeigt, gibt es drei mögliche Formulierungen der Nullhypothese [97]. Der TDT kann also in verschiedener Weise angewandt werden: z. B. als Test auf Assoziation nach einem positiven Kopplungsbefund oder als Follow-up einer Fall-Kontroll-Studie (vgl. auch [54]).

Für die Familien-basierten Verfahren dieser Studie werden nur zwei Nullhypothesen betrachtet:

- $H_0^L$ : keine Kopplung (I  $\cup$  II)  
gegen die Alternative  $H_1^L$ : Kopplung (III  $\cup$  IV)
- $H_0^A$ : keine Assoziation oder keine Kopplung (I  $\cup$  II  $\cup$  III)  
gegen die Alternative  $H_1^A$ : Kopplung und Assoziation (IV)

## 1.2 Einordnung und Abgrenzung des Themas

### 1.2.1 Wert und Probleme Familien-basierter Studien

Moderne Technologie ermöglicht es heute, Hunderttausende von Loci auf einem Chip zu genotypisieren. Bei genomweiten Assoziationsstudien überdeckt man das ganze Genom mit einem Raster von Markern, um potentielle Risikoloci zu finden. Nach Wang et al. [157] ist das die pro wahren positivem Fund kostengünstigste Methode - verglichen mit Kandidatengen- bzw. Kopplungsbasiertem Ansatz. Unter diesem Aspekt stellten sich Laird und Lange [97] der Herausforderung, die Position Familien-basierter Verfahren neu zu bestimmen. Sie betrachten sie nicht als Konkurrenz zu populationsbasierten Verfahren, sondern als wichtige Ergänzung. U. a. schlagen diese Autoren vor, FBAT (s. 3.2.4) auf 10 - 20 Marker anzuwenden, die durch genomweites Screening gefunden wurden. Im Folgenden sollen Vor- und Nachteile dieser beiden Ansätze einander gegenübergestellt werden:

#### Vorteile genomweiter Assoziationsstudien (GWASs)

**Keine a-priori-Kenntnisse nötig:** Da eine GWAS keine Vorkenntnisse benötigt, z. B. über das Krankheitsmodell, ist sie für ein erstes Screening prädestiniert.

**Realisierung einfach und kostengünstig:** Meist wird eine GWAS als Fall-Kontroll-Studie durchgeführt. Dann ist die Rekrutierung der Probanden relativ einfach, selbst für spät auftretende Krankheiten. Verglichen damit braucht der TDT etwa einen anderthalbfachen Stichprobenumfang, weil dort Trios statt eines case-control-Paares benötigt werden. Bei spät auftretenden Krankheiten sind die elterlichen Genotypen kaum verfügbar, so dass man dort z. B. auf Geschwister [78], Kinder oder gar Ehepartner der Erkrankten zurückgreifen muss [103].

**Informationsgehalt:** Beim TDT sind nur heterozygote Eltern informativ, was insbesondere bei seltenen Allelen die Lage verschlechtert.

## Vorteile des Familien-basierten Ansatzes

**Robustheit gegen Stratifikation** Familien-basierte Studien sind vom Ansatz her robust gegen Populations-Inhomogenität, da Familien-Kontrollen genetisch passen. Die Qualität von Fall-Kontroll-Studien hängt substantiell von der geeigneten Wahl der Kontrollen ab.

**Phasenauflösung** Die Genotypisierung liefert normalerweise Genotypdaten. Heterozygote Loci bewirken, dass die Phase nicht eindeutig aufgelöst werden kann. Wie aber Abb. 1.1 zeigt, kann die Genotypinformation der Eltern teilweise oder sogar ganz die Eindeutigkeit herstellen. Das ist ein Vorteil insbesondere für Haplotyp-basierte Tests.

**Genotypisierungsfehler erkennen** Inkonsistenzen der Genotypen von Eltern und Kind (nicht ganz präzise 'Mendel-Fehler' genannt) lassen auf Genotypisierungsfehler schließen. Weitere statistische Verfahren, die solche Fehler ziemlich wahrscheinlich nachweisen, wurden ausschließlich für Familiendaten entwickelt, z. B. [21]. Das wiegt den Einwand, dass Genotypisierungsfehler die Ergebnisse Familien-basierter Studien weit mehr verzerren als Fall-Kontroll-Studien [70], zumindest teilweise auf.

**Imprinting** Manchmal hängt die Wirkung eines Allels davon ab, ob es von der Mutter oder vom Vater ans Kind weiter gegeben wurde. Solche Imprinting-Effekte können nur mit familien-basierten Daten ermittelt werden (z. B. [17]).

**Copy Number Variations** treten auf, wenn Chromosom-Abschnitte ausfallen oder in mehreren Kopien vorliegen. Das äußert sich darin, dass an einer Reihe von Loci nicht zwei Allele, sondern drei, eins oder eine andere Anzahl zu finden sind. Wiederum sind Familiendaten hilfreich, diese CNV zu identifizieren und statistisch zu analysieren (siehe z. B. [86]).

Familien-basierte Tests sind *eine* wirksame Möglichkeit, sich gegen den verfälschenden Einfluss von Populationsstratifikation zu schützen. Um aber dem hohen Aufwand zu entgehen, hat man nach Alternativen gesucht. Diese wurden durch Tiwari & al. [155] systematisiert:

**Genomic Control** Devlin und Roeder [46] gingen davon aus, dass Populationsstratifikation Fall-Kontroll-Statistiken (z. B. Armitage's Trend Test) verfälscht. Sie schlugen vor, anhand zusätzlicher Genotypinformation einen Faktor  $\lambda$  zu berechnen, mit dem die  $\chi^2$ -Statistik korrigiert wird. Insbesondere in Populationen mit geringer Stratifikation erreicht Genomic Control höhere Power als der TDT [13].

**Matched Controls** Ein naheliegender Ansatz ist, Kontrollen passend auszuwählen. Lee [104] schlug vor, nach einer Anzahl von Variablen  $s$  Schichten zu definieren. Auf dieser Basis kann ein MDT (*matching disequilibrium test*) berechnet werden. Lee verglich diesen Test mit dem 'genomic control'-Verfahren.

**Structured Association Testing** Sehr vereinfacht dargestellt bedeutet dieser Ansatz, die Assoziationsanalyse stratifiziert in verschiedenen Bevölkerungsgruppen durchzuführen (z. B. Pritchard et al. [121]).

**Regional Admixture Mapping** Dieses Verfahren basiert auf dem Gedanken, dass die Vorfahren von Kranken einer Mischpopulation eher aus einer Bevölkerungsgruppe mit höherer Frequenz des Risikoallels stammen (z. B. [56], [112]).

Mehrere Autoren sichteten und verglichen diese Methoden, z. B. [28],[15]. Die Bedeutung von Populationsstratifikation wurde in der Diskussion sowohl über- als auch unterbewertet. Eine ausführliche Studie von Marchini [108] sorgte hier für mehr Klarheit: In kleinen Stichproben senkt Stratifikation die Power, Effekte zu entdecken. Dagegen wächst die Gefahr falsch positiver Befunde mit wachsendem Stichprobenumfang. Marchini untersuchte weiterhin, wie viele Loci zur Berechnung des Korrekturfaktors  $\lambda$  nötig sind. Schließlich verglichen Köhler und Bickeböller [89] in einer Simulationsstudie Genomic Control und Structured Association Testing.

## 1.2.2 Geschichtlicher Abriss

### Single-Marker-Methoden

Der von Spielman et al. 1993 als TDT [145] publizierte Test wurde zum Prototyp Familien-basierter Verfahren. Da als Stichprobeneinheiten nicht immer Trios zur Verfügung stehen, erfuhr er bald Erweiterungen für verschiedene Familienstrukturen [180]: für ASPs [40], sib-TDT (Geschwister als Kontrollen [144]), STD (mehrere kranke und gesunde Kinder [78]), 1-TDT (Indexfall und ein Elternteil [151]), PDT (erweiterte Kernfamilien [109]) und MDT (Ehepartner oder Kinder als Kontrollen [103]). Zhao [176] und Whittaker & Morris [161] systematisierten Familien-basierte Tests von Assoziation und / oder Linkage (*FBTLA*), darunter obige TDT-Erweiterungen. Zou [180] gelang es, diese Verfahren als Sonderfälle eines retrospektiven logistischen Regressionsmodells einzuordnen.

### Haplotyp-basierte Methoden

Autoren wie Akey et al. [8] oder Zaykin et al. [170] hatten das Potential von Haplotypen bei der Entdeckung von Assoziationen mit komplexen Krankheiten erkannt. Clark fasste 2004 wesentliche Gründe in drei Punkten zusammen [31]:

- 1) Haplotypen definieren funktionelle Einheiten von Genen. Weil man den Ursachen von Krankheiten auf die Spur kommen will, erscheint es verheißungsvoller, Haplotypen anstelle einzelner Marker zu betrachten.

- 2) Genetische Variabilität in Populationen - Mutation, Selektion, Migration usw. - schlägt sich unmittelbar in Haplotypen nieder. D. h. in Haplotypstrukturen mehrerer Loci ist eine gemeinsame Vorgeschichte aufgezeichnet.

- 3) Indem Allele mehrerer Marker zu Haplotypen zusammengefasst werden, wird die Dimensionalität verringert. Statistische Tests haben somit weniger Freiheitsgrade und die Power ist u. U. höher.

Diesen Vorzügen steht der Einwand gegenüber, dass mit wachsender Markerzahl die Anzahl der Haplotypen im Quadrat wächst. Die große Anzahl von Freiheitsgraden wirkt somit einem evtl. Powergewinn entgegen [176]. Wilson [163] bzw. 1999 Clayton & Jones [33] waren die ersten, welche den TDT auf Haplotypen erweiterten, zunächst bei bekannter Phase, dann auch bei unbestimmter Phase [32]. Bereits 2004 hatte sich ein ganzes Spektrum von Haplotyp-Methoden etabliert, so dass das Journal Genetic Epidemiology die Ausgabe 27(4) dieser Problematik widmete [133].

### Verfahrensvergleiche

In den letzten Jahren sind verschiedene Anstrengungen unternommen worden, Verfahren mit Haplotypen zu vergleichen. So evaluierten Zhang und Zhao [173] sowie Marchini et al. [108] die Effizienz verschiedener Schätzverfahren für Haplotypen, darunter Genehunter, Merlin und Phase.

Mehrere Autoren widmeten sich der Frage der Effektivität von Tests mit multiplen Markern auf der Basis von Fall-Kontroll-Daten: So verglichen vor kurzem Chapman & Whittaker [29] sowie Bacanu et al. [14] die Power mehrerer Verfahren. Ebenso verglichen Motsinger-Reif et al. [114] verschiedene Ansätze für den Test auf Wechselwirkungen.

Einen methodologischen Überblick populationsbasierter Verfahren gab D. Balding [15] und stellte dabei fest, dass einer effektiven Analyse der Daten bisher recht wenig Beachtung geschenkt wurde. Ein Review robuster Verfahren gegenüber Populations-Stratifikation [155] ordnet zahlreiche Methoden nach ihrem konzeptionellen Ansatz ein. Haplotyp-Verfahren sind hierbei aber eher die Ausnahme.

## 1.2.3 Eingrenzung der genetisch-statistischen Methoden

Da bis heute ein Vergleich der Effizienz von Familien-basierten Assoziations-Tests mit Haplotypen fehlt, soll die vorliegende Studie diese Lücke teilweise schließen. Im Fokus sind Verfahren, die Genotyp-Daten eng gekoppelter Marker analysieren: Z. B. könnte es sein, dass ein gefundenes



Kandidatengen weiter untersucht werden soll. Oder eine Chromosomen-Region ist durch ein hohes Kopplungssignal eines vorhergehenden Linkage-Scans ins Interesse gerückt und der Krankheits-Genort (DL) soll feiner lokalisiert werden [75]. Vor allem wegen ihrer Robustheit gegenüber Populationsstratifikation sollen hier nur Familien-basierte Verfahren untersucht werden. Quantitative Krankheitsmerkmale, wie z. B. Blutdruck oder Cholesterinspiegel, erfordern statistisch völlig andere Ansätze als binäre Merkmale. So beschränkt sich diese Arbeit ausschließlich auf diese - normalerweise den Krankheitsstatus. Im Absatz 3.1 werden solche Verfahren kurz vorgestellt und es wird aus ihnen eine Auswahl für die Simulationsstudie getroffen.

# Kapitel 2

## Planung der Simulation

### 2.1 Wozu Simulationsverfahren?

Computersimulation wird inzwischen in vielen Wissenschaftszweigen erfolgreich eingesetzt. In der genetischen Epidemiologie spielen Simulationsverfahren gleich an mehreren Stellen eine wichtige Rolle:

1. Die Verteilungsfunktion von Teststatistiken kann in vielen Fällen nicht explizit oder nur asymptotisch über den Zentralen Grenzwertsatz angegeben werden. Das kann den Fehler 1. Art besonders bei kleineren Stichproben inflationieren. Als Ausweg bietet sich an, die Verteilung der Statistik (z. B. durch eine Permutation der Genotypen) per Simulation zu erzeugen und auf diese Weise den p-Wert zu bestimmen ([20], [49] u. a.). In analoger Weise kann man auch die Varianz der Teststatistik ermitteln [80].
2. Bei Haplotyp-basiertem Vorgehen werden  $m$  Marker nicht mehr isoliert betrachtet. Um den Haplotyp mit dem größten Effekt zu finden, müssten also Tests für alle Haplotypen  $B \in \mathfrak{B}$ , der Menge tatsächlich vorkommender Haplotypen, durchgeführt werden. Dabei ist  $\mathfrak{B}$  eine Untermenge der  $2^m$  theoretisch möglichen Haplotypen. Weil die Anzahl der Tests mit wachsendem  $m$  schnell ansteigt, kann die Korrektur für multiples Testen (z. B. nach Bonferroni [77]) die Power untergraben, weil sie die Abhängigkeit der Marker untereinander nicht berücksichtigt. Programme wie Unphased oder Famhap beschreiten deshalb den Weg, aus dem minimalen p-Wert aller Markerkombinationen durch Simulation einen overall-p-Wert zu bestimmen. Wie das sehr effektiv ohne geschachtelte Simulation durchgeführt werden kann [20], wird im Abschnitt 4.2.1 beschrieben.
3. Zum geplanten Test der Assoziationsverfahren sind Stichproben von Kernfamilien mit definierten Eigenschaften notwendig. Auch diese können durch Simulationsprogramme erzeugt werden.
4. Schließlich kann man die genetische Entwicklung ganzer Populationen, ausgehend von wenigen *foundern* (Familiengründern), simulieren. Auf diese Weise kann auch die Entstehung und Verbreitung einer Krankheit mit dem Computer nachempfunden werden. Populationsgenetik ist nicht Thema dieser Studie.

### 2.2 Modellbildung für die Simulation

#### 2.2.1 Vorüberlegungen

Modellbildung heißt Abstraktion von der Wirklichkeit: Die Konzentration auf Wesentliches ist einerseits der Komplexität der Problemstellung geschuldet. Andererseits kann man erwarten, dass Effekte besonders deutlich zu Tage treten, wenn sie anhand 'reiner' Versuchsbedingungen getestet

werden. So wurde z. B. auf gemischte Populationen hinsichtlich verschiedener Familienstrukturen verzichtet. Weiterhin wurde von genau einem Krankheitslocus ausgegangen.

Power und Fehler erster Art Haplotyp-basierter Tests auf Assoziation sollen für verschiedene Datensituationen bestimmt werden. In einem Simulations-Szenarium waren deshalb folgende Komponenten festzulegen:

- das Krankheitsmodell: Häufigkeit des Risikoallels, Vererbungsmodus (gegeben durch Penetranzen der Genotypen am DL),
- das genetische Modell: Anzahl der Marker und ihrer Allele, Rekombinationsfrequenzen, Position des DL innerhalb der Marker, LD-Struktur (gegeben durch Haplotyphäufigkeiten an den Marker-Loci in Bezug zum DL),
- Anzahl und Art der zu analysierenden Familien
- Replikationsanzahl

Um Haplotyp-Effekte untersuchen zu können, benötigt man eine Mindestzahl von Markern, andererseits wächst die Anzahl möglicher Markerkombinationen und damit der Rechenaufwand exponentiell mit der Anzahl der Marker. Somit fiel die Wahl auf Konfigurationen aus fünf diallelischen Markern  $M_1 \dots M_5$ . Der DL befinde sich zwischen  $M_2$  und  $M_3$  (siehe Abbildung 2.1). Marker und DL seien eng gekoppelt (keine Rekombination, außer Modell NMod3, s. u.).

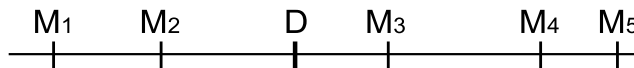


Abbildung 2.1: Krankheitslocus und Markerloci

Was die Krankheits-Allelfrequenzen anbelangt, so bestimmen zwei Auffassungen die kontroverse Diskussion [157], [179]: Nach der *common disease / common variant* (CDCV)-Hypothese werden häufige Krankheiten durch das Zusammenspiel häufiger genetischer Varianten hervorgerufen. Die *multiple rare-variant hypothesis* dagegen geht davon aus, dass unterschiedliche seltene Risikoallele (mit einer *minor allele frequency*,  $MAF < 1\%$  in der Population) Personen für die Krankheit disponieren.

Ohne sich in diese Diskussion zu verlieren sei erwähnt, dass die Häufigkeitsverteilung aller bekannten SNPs extrem schief ist mit einer Häufung im niedrigen Bereich. Funktionale SNPs in kodierenden Regionen haben oft sehr geringe Allelfrequenzen [24]. Andererseits steigen die notwendigen Stichprobenumfänge mit fallender Krankheits-Allelfrequenz drastisch an. So wird als minimale MAF 5% genannt, um Risikoeffekte überhaupt noch entdecken zu können [15]. Somit liegt es auf der Hand, dass Effekte seltener Varianten nur sehr schwer aufzudecken sind.

## 2.2.2 Modelle für Alternativhypothesen

Da hier Methoden für mehrere, eng gekoppelte Marker untersucht werden, kann man die zusammengesetzte Hypothese: 'keine Assoziation' nach McIntyre et al. [110] verschieden formulieren:

- $H_{01}$ : keine Assoziation von Allelen irgendeines Markers und des DL
- $H_{02}$ : keine Assoziation irgendeines Haplotypen mit dem Risikoallel

Wie man leicht erkennen kann, gilt  $H_{02} \supseteq H_{01}$ , aber nicht umgekehrt. D. h. es sind Fälle denkbar, wo keine einzelnen Marker mit der Krankheit assoziiert sind, wohl aber irgendwelche Haplotypen. Durch die Haplotyp-Konfigurationen KMod2 bzw. KMod3 (s. u.) werden genau solche Situationen modelliert. Es wird sich zeigen, dass eine Reihe von Algorithmen in diesen Fällen keine Power haben.

## Familienstruktur

Eine übliche Basiseinheit für Familien-basierte Analysen ist das Trio aus einer erkrankten Person, dem Indexfall, und ihren Eltern. Der 1993 von Spielman et al. publizierte TDT [145] verwendet Daten solcher Triaden. Im Weiteren wird diese Familienstruktur auch mit F10 bezeichnet: ein erkranktes und null gesunde Kinder mit Eltern. Als weitere wichtige Einheit sollen zwei erkrankte Personen (*affected sibpairs*, ASP) mit ihren Eltern betrachtet werden, in der obigen Bezeichnung: F20 oder auch 'AA-Familien'. Schließlich sollen Familien mit je zwei kranken und je zwei gesunden Kindern, d. h. F22 bzw. AAUU-Familien, einbezogen werden.

## Genetisches Modell

Die vier in 1.1.2 vorgestellten Vererbungsmodi sollen untersucht werden: multiplikativ, rezessiv, dominant und additiv.

Die Häufigkeit des Risikoallels und die Penetranzen definieren eindeutig ein Krankheitsmodell, z. B. [179]. Eine andere Möglichkeit wäre, es durch Rekurrenz-Risikoverhältnisse unter Paaren von Verwandten vorzugeben [22]. Die Modelle sollten so gestaltet sein, dass eine Kopplungsanalyse selbst einen voll informativen Marker am Krankheits-Genort kaum entdecken würde. Sei ein *allele sharing*  $\tau = 0.52$  für ASPs und ein Fehler 1. bzw. 2. Art  $\alpha = 0.05$  und  $\beta = 0.80$  angenommen. Nach [92], Formel (3) ergibt sich für den *mean test* [23] mit ASPs ein Stichprobenumfang von 1884.

Weiterhin wurden die Häufigkeit des Krankheitsallels  $p$  und die Penetranzen  $f_2, f_1, f_0$  so gewählt, dass die Prävalenz  $K_P \approx 6\%$  beträgt (siehe 1.9). Das entspricht der Lebenszeit-Prävalenz typischer 'Volkskrankheiten', wie z. B. Angina pectoris (5,8%), Asthma bronchiale (5,7%) und Diabetes (5,8%) [1],[2].

Die Wahrscheinlichkeiten, dass ASPs zwei, ein bzw. null Allelkopien gleicher Herkunft besitzen (*identical by descent*, IBD), berechnen sich nach den Beziehungen [150]:

$$z_2 = \frac{1}{4} + \frac{\frac{V_A}{2} + \frac{3}{4} \frac{V_D}{4}}{4(K_P^2 + \frac{V_A}{2} + \frac{V_D}{4})} \quad (2.1)$$

$$z_1 = \frac{1}{2} - \frac{\frac{V_D}{2}}{4(K_P^2 + \frac{V_A}{2} + \frac{V_D}{4})} \quad (2.2)$$

$$z_0 = \frac{1}{4} - \frac{\frac{V_A}{2} + \frac{V_D}{4}}{4(K_P^2 + \frac{V_A}{2} + \frac{V_D}{4})} \quad (2.3)$$

$$\text{mit } V_A = 2p(1-p)[p(f_2 - f_1) + (1-p)(f_1 - f_0)]^2 \quad (2.4)$$

$$\text{und } V_D = p^2(1-p)^2[f_2 - 2f_1 + f_0]^2 \quad (2.5)$$

Der Erwartungswert für das allele sharing unter ASPs ist

$$\tau = 0.5 \cdot z_1 + z_2. \quad (2.6)$$

Schließlich ist das Risiko an der betrachteten Krankheit zu erkranken, wenn bereits die Schwester oder der Bruder krank ist, das *recurrence risk ratio* für Geschwister, nach [127]:

$$\lambda_S = \frac{0.25}{z_0} \quad (2.7)$$

Wichtige Parameter der vier Modelle sind in Tabelle 2.1 zusammengefasst:

## Haplotyp-Häufigkeiten

Durch die Haplotyp-Struktur wird das LD zwischen Risikolocus und Markern sowie innerhalb der Marker modelliert. Im Modell KMod1 wird der Fall erfasst, dass ein einzelner Markerlocus mit dem DL assoziiert ist. Seien die beiden Allele der Marker  $M_i$  ( $i = 1, \dots, 5$ ) mit  $A_i$  bzw.  $B_i$

MOI <sup>a</sup>	p	Penetranzen				IBD-Wahrscheinlichkeiten				
		$f_2$	$f_1$	$f_0$	$K_P$ <sup>b</sup>	$z_2$	$z_1$	$z_0$	$\tau$ <sup>c</sup>	$\lambda_S$ <sup>d</sup>
MUL <sup>a</sup>	0.118	0.192	0.096	0.048	0.060	0.2704	0.4992	0.2304	0.520	1.08
REZ	0.147	0.224	0.056	0.056	0.060	0.2773	0.4852	0.2375	0.520	1.05
DOM	0.484	0.076	0.076	0.019	0.061	0.2732	0.4936	0.2332	0.520	1.07
ADD	0.056	0.204	0.1275	0.051	0.060	0.2701	0.5000	0.2300	0.520	1.09

Tabelle 2.1: Genetische Modelle

<sup>a</sup> Vererbungsmodus, Abkürzungen s. o.

<sup>b</sup> Prävalenz

<sup>c</sup> mittlere Anzahl Allele IBD unter Geschwistern

<sup>d</sup> Rekurrenzzisiko für Geschwister

bezeichnet. Dann sind die Häufigkeiten der Allele am Marker  $M_3$  auf Chromosomen mit bzw. ohne Risikoallel in Tabelle 2.2 gegeben. Für die verbleibenden vier Marker wurden je zwei gleich häufige Allele vorausgesetzt, die im Kopplungs-Gleichgewicht untereinander und mit Marker  $M_3$  stehen. Somit ergeben sich die bedingten Häufigkeiten der Haplotypen auf Chromosomen mit bzw. ohne Risikoallel D am DL z. B. als  $P(A_1A_2A_3A_4A_5|D) = 0.84 \cdot 0.5^4 = 0.0525$  bzw.  $P(A_1A_2A_3A_4A_5|d) = 0.16 \cdot 0.5^4 = 0.01$  usw.

Allele (a)	P(a D) <sup>a</sup>	P(a d) <sup>b</sup>
$A_3$	0.84	0.16
$B_3$	0.16	0.84

Tabelle 2.2: Markerkonfiguration KMod 1

<sup>a</sup> Häufigkeit des Allels a auf elterlichen Chromosomen mit Allel D am Krankheits-Genort

<sup>b</sup> Häufigkeit des Allels a auf elterlichen Chromosomen mit Allel d am Krankheits-Genort

Haplotypen (h)	P(h D)	P(h d)
$A_2A_3$	0.42	0.08
$A_2B_3$	0.08	0.42
$B_2A_3$	0.08	0.42
$B_2B_3$	0.42	0.08

Tabelle 2.3: Markerkonfiguration KMod 2

Beim zweiten Modell KMod 2 wurde angenommen, dass die Haplotypen an  $M_2$  und  $M_3$  mit den Allelen am Krankheitslocus im Linkage-Disequilibrium stehen. Die bedingten Häufigkeiten der Marker-Haplotypen an  $M_2$  und  $M_3$  bei gegebenem Allel am Krankheits-Genort sind in Tabelle 2.3 gegeben: Damit gelten folgende Beziehungen:  $P(A_2|D) = P(A_2|d)$  und  $P(A_3|D) = P(A_3|d)$ . D. h. sowohl  $A_2$  und  $B_2$ , als auch  $A_3$  und  $B_3$  kommen auf elterlichen Chromosomen mit (bzw. ohne) das Risikoallel D am Krankheits-Genort gleich häufig vor. Das bedeutet, es gibt keine Assoziation zwischen der Krankheit und den Einzelmarkern. Für die verbleibenden Marker  $M_1$ ,  $M_4$  und  $M_5$  wurden wie oben jeweils gleich häufige Allele vorausgesetzt, die im vollständigen Kopplungs-gleichgewicht untereinander und mit den Haplotypen an  $M_2M_3$  stehen. Wiederum errechnen sich die Haplotypfrequenzen als bedingte Häufigkeit, z. B.  $P(A_1A_2A_3A_4A_5|D) = 0.5^3 \cdot P(A_2A_3|D) = 0.0525$  und analog für die anderen Haplotypen.

Für Konfiguration KMod 3 sind die bedingten Häufigkeiten der Haplotypen aus den Markern  $M_2$ ,  $M_3$  und  $M_4$  in Tabelle 2.4 gegeben. Sie sind wiederum so gewählt, dass weder die Einzelmarker

Haplotypen (h)	P(h D)	P(h d)
$A_2A_3A_4$	0.21	0.04
$A_2A_3B_4$	0.04	0.21
$A_2B_3A_4$	0.04	0.21
$A_2B_3B_4$	0.21	0.04
$B_2A_3A_4$	0.04	0.21
$B_2A_3B_4$	0.21	0.04
$B_2B_3A_4$	0.21	0.04
$B_2B_3B_4$	0.04	0.21

Tabelle 2.4: Markerkonfiguration KMod 3

$M_2, M_3, M_4$  noch Haplotypen aus je zwei Markern aus dieser Menge mit der Krankheit assoziiert sind. Es gilt also für die bedingten Häufigkeiten:  $P(A_i|D) = P(A_i|d)$  für  $i = 2, 3, 4$  und ebenso  $P(A_iA_j|D) = P(A_iA_j|d)$  für  $2 \leq i < j \leq 4$ . Wie oben besitzen die Marker  $M_1$  und  $M_5$  je zwei gleich häufige Allele, die im Kopplungs-Gleichgewicht untereinander und mit dem drei-Locus-Haplotyp an  $M_2M_3M_4$  stehen.

Vererbungsmodus, Markerkonfiguration und Familienstruktur definieren damit eins von 36 'Szenarien' unter  $H_1$ . REZ, KMod 2, AAUU-Familien ist ein Beispiel dafür, wie ein Szenarium im weiteren Verlauf bezeichnet wird.

### Stichprobenumfänge

Die Stichprobenumfänge sollten so gewählt sein, dass die simulierten Effekte mit einer Power von ca. 80% entdeckt werden. Der dafür notwendige Stichprobenumfang für Assoziationsstudien hängt wesentlich vom Vererbungsmodus ab (s. u.). Weiterhin steigt er nach Wang & al. [157], je kleiner der genetische Effekt (z. B. das OR) und je kleiner die Häufigkeit des selteneren Allels (MAF) ist.

Mit einer Pilotstudie wurde deshalb die Power abgeschätzt: Für jedes der 4 (Vererbungsmodi)  $\times$  3 (Markerkonfigurationen)  $\times$  3 (Familienstrukturen) = 36 Szenarien wurden je 100 Dateien mit Famhap analysiert. Der Anteil signifikanter Ergebnisse gibt einen groben Schätzer für die Power. Dabei wurde schrittweise die Anzahl von Familien der Stichproben erhöht, bis die Power  $\approx$  80% erreichte. Die Ergebnisse sind in den Tabellen A.1 und A.2 im Anhang dargestellt.

Damit ergeben sich für die einzelnen Szenarien die Stichprobenumfänge (Anzahl Familien) in Tabelle 2.5.

Modell	MOI	Familientypen		
		F10	F20	F22
KMod 1	MUL	700	200	200
	REZ	2500	500	600
	DOM	400	300	300
	ADD	1100	300	300
Kmod 2	MUL	1200	300	300
	REZ	6000	900	900
	DOM	1000	500	500
	ADD	1800	400	400
Kmod 3	MUL	1700	500	400
	REZ	11000	1400	1000
	DOM	1500	700	600
	ADD	2500	500	500

Tabelle 2.5: Stichprobenumfänge für  $H_1$

### 2.2.3 Modelle für die Nullhypothese

Auch für die Tests auf Validität (Einhalten des Fehlerniveaus 1. Art,  $\alpha$ ) wurden Szenarien als Kombination dreier Faktoren konzipiert: die Struktur der Kernfamilien, drei Null-Modelle: NMod 1, NMod 2 und NMod 3 und drei verschiedene Stichprobenumfänge.

#### Struktur der Familien

Wie unter  $H_1$  wurden drei Familienstrukturen getestet: Trios, AA- und AAUU-Familien.

#### $H_0$ -Modelle

Es galt, „Krankheitsmodelle“ zu entwerfen, bei denen *keine* Assoziation bzw. Kopplung zwischen den Haplotypen  $M_1 \dots M_5$  und der Krankheit besteht.

1. Parallel zum Modell KMod 1 wurde im Modell NMod 1 vorausgesetzt, dass das Allel  $A_3$  am Marker  $M_3$  zu 84% auf elterlichen Chromosomen *mit* dem Risikoallel D am Krankheits-Genort und in 16% auf elterlichen Chromosomen *ohne* Risikoallel vorkommt. Die verbleibenden Loci  $M_1, M_2, M_4, M_5$  mit je zwei gleich häufigen Allelen stehen im kompletten Kopplungs-Gleichgewicht untereinander und mit  $M_3$ . Bei gegebenen Rekombinationsraten  $\theta = 0$  sind alle Marker untereinander und mit dem Krankheitslocus eng gekoppelt. Weiterhin sind die drei Penetranzen als identisch festgelegt:  $f_2 = f_1 = f_0 = 0.06$ , d. h. der Locus trägt nicht zu einem höheren Krankheitsrisiko bei. Deshalb gibt es zwischen den Markern oder irgendeiner Kombination und der Krankheit keine Assoziation. Die Häufigkeit des Risikoallels sei mit 0.1 vorgegeben. Nach Formel (1.9) ergibt sich die Prävalenz zu 6%.
2. Im Modell NMod 2 sind die Penetranzen wie folgt festgelegt:  $f_2 = 1$  und  $f_1 = f_0 = 0$  (rezessiver Vererbungsmodus) bei einer Risikoallelfrequenz  $P(D) = 0.01$ . Weiterhin wurden alle 32 möglichen Haplotypen als gleich häufig angenommen, so z. B.  $P(A_1A_2A_3A_4A_5|D) = P(A_1A_2A_3A_4A_5|d) = 0.03125$  usw. Die Rekombinationsraten seien wiederum als  $\theta = 0$  vorausgesetzt. D. h. hier sind die Marker untereinander und mit dem DL eng gekoppelt, aber sie stehen nicht im Kopplungsungleichgewicht zum Risikolocus. Folglich sind sie nicht mit der Krankheit assoziiert.
3. Im Modell NMod 3 sind die Allelhäufigkeiten an  $M_3$  durch Tabelle 2.2 vorgegeben. Die Penetranzen seien:  $f_2 = 0.204$ ,  $f_1 = 0.1275$ ,  $f_0 = 0.051$  (additives Vererbungsmodell). Daraus ergibt sich eine Prävalenz von  $\approx 6\%$ . Wiederum haben die Marker  $M_1, M_2, M_4$  und  $M_5$  je zwei gleich häufige Allele. Die Marker  $M_1, M_2$  bzw.  $M_3, M_4, M_5$  sind untereinander perfekt gekoppelt und stehen im Kopplungs-Gleichgewicht. Allerdings rekombinieren die Marker  $M_2$  bzw.  $M_3$  und der Krankheitslocus frei ( $\theta = 0.5$ ). Das hat zur Folge, dass es zwischen den Markern und dem DL weder Kopplung noch Assoziation gibt.

Das kann durch folgende Grafik veranschaulicht werden.

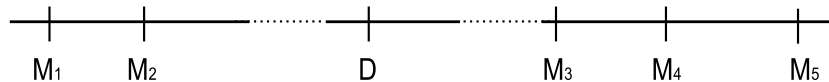


Abbildung 2.2:  $H_0$ -Modell NMod3

#### Stichprobenumfänge für $H_0$

Weil die für verschiedene Alternativen ermittelten Stichprobenumfänge zwischen 200 und 11000 liegen (s. Tab. 2.5), lag es nahe, 200 Familien als kleinsten Stichprobenumfang zu wählen. Voruntersuchungen mit dem Stichprobenumfang 11000 ließen extrem lange Rechenzeiten (Unphased) erwarten bzw. deuteten auf Speicherplatz-Probleme hin (FBAT). Außerdem ist es zurzeit kaum

praktikabel, 11000 Trios zu rekrutieren und zu genotypisieren. Deshalb wurden 2500 als oberer und 1000 als mittlerer Stichprobenumfang für die Nullhypothese gewählt.

Somit definieren Nullmodell, Stichprobenumfang und Familienstruktur eins von 27 Szenarien unter  $H_0$ , z. B. 2500 Trios, NMod 1.

#### 2.2.4 Modelle für fehlende Genotypen

Ziel der Analysen mit unvollständigen Genotyp-Daten war einerseits zu prüfen, ob die Algorithmen gültige Tests sind bzw. bleiben. Andererseits sollte getestet werden, wieviel Power durch fehlende Genotypinformation verloren geht. Moderne Genotypisierungsverfahren haben Fehlraten von weniger als 1% je Locus und Person (z. B. [12]). Also wurde eine *missing rate* von 1% angenommen. Daraus folgt, dass sich in Trios ca. 14%, in AA-Familien ca. 18% und in AAUU-Familien immerhin ca. 26% der Personen mit fehlenden Genotypen finden werden: Das kann man leicht aus der Wahrscheinlichkeit des komplementären Ereignisses 'komplette Daten' errechnen, bei der  $(1 - 0.01)$  mit der Anzahl der Marker (5) und dann noch einmal mit der Anzahl Personen (bei Trios z. B. 3) potenziert wird.

$$P(\text{Genotyp komplett} \mid \text{Trio}) = [(1 - .01)^5]^3 = 0.99^{15} = 0.86$$

Diese Studie betrachtet nur *uninformative missings*, wo die fehlenden Werte nicht von den Genotypen abhängen. Das ist in der Praxis nicht immer gegeben, z. B. wenn einzelne Allele bestimmter Loci schwerer zu genotypisieren sind als andere. Um die Resultate der Analysen mit denen kompletter Daten vergleichen zu können, wurde ein Anteil von 1% Genotypen aus den vollständigen Stichproben fallen gelassen (vgl. 5.4.2).

#### 2.2.5 Modelle für Stratifikation

Robustheit gegenüber Populationsstratifikation war ein wesentlicher Antrieb für die Entwicklung Familien-basierter Assoziationsverfahren [155]. Dass F. Dudbridge kürzlich die Validität von Famhap bei stratifizierter Population in Frage stellte [51], war der Anlass, diesen Aspekt extra abzuklären. Um den Aufwand überschaubar zu halten, wurde dabei ein einfaches Modell mit zwei diallelischen Markern in der Nähe des Krankheitslocus gewählt. Die Stichprobe sei zu unterschiedlichen Anteilen aus zwei Populationen gezogen, die sich sowohl durch die Risiko-Allelfrequenz als auch die Penetranzen unterscheiden. Die Häufigkeiten elterlicher Haplotypen  $P(h|D$  bzw.  $P(h|d)$  auf Chromosomen mit bzw. ohne Risikoallel sind gleich innerhalb der Teilpopulationen, aber nicht zwischen ihnen. Die Marker seien vollständig gekoppelt ( $\theta = 0$ ). In einem 'moderaten' Modell S1 wurde ein rezessiver, im 'extremen' Modell S2 ein multiplikativer Erbgang vorausgesetzt. Die Vorgaben sind in Tabelle 2.6 zusammengefasst.

Seit längerem ist bekannt, dass der Score-Test von TRANSMIT nicht robust gegenüber Stratifikation ist [109]. Der Fehler 1. Art kann sich noch erhöhen, wenn die Genotypen der Eltern nicht vorliegen. Deshalb wurde ein drittes Modell S3 entworfen, das aus S1 durch Auslassen von je 5% Genotypen je Marker entsteht.

Wie man in Tab. 2.6 leicht nachprüfen kann, gibt es in keiner der beiden Teilpopulationen Assoziation zwischen den zwei Markern und der Krankheit. Unterschiedliche Bevölkerungs-Anteile und verschiedene Frequenzen des Risikoallels D lassen aber in der Gesamtpopulation Assoziation entstehen. Die entstandenen Effekte für die Eltern wurden für jeden Marker einzeln in Fall-Kontroll-Analysen mit Armitage's Trendtest geprüft. Tabelle 6.7 enthält die Häufigkeiten signifikanter Tests. Die hohen Anteile signifikanter Tests bestätigen, dass die ausgewählten Verfahren einen inflationären Fehler 1. Art zeigen müssten, falls sie nicht robust gegenüber Stratifikation sind.



Moderates Modell S1								
Haplotyp h		Population 1		Population 2		Gesamtpopulation		
$M_1$	$M_2$	$P(h D)^a$	$P(h d)$	$P(h D)$	$P(h d)$	$P(h D)$	$P(h d)$	
1	1	0.16	0.16	0.24	0.24	0.200	0.224	
1	2	0.64	0.64	0.36	0.36	0.500	0.418	
2	1	0.04	0.04	0.16	0.16	0.100	0.135	
2	2	0.16	0.16	0.24	0.24	0.200	0.224	
Allelfrequenz 1( $M_1$ )		0.8		0.6		0.7		
Allelfrequenz 1( $M_2$ )		0.2		0.4		0.3		
Krankheitsallelfrequenz		0.3		0.8				
Anteil an Population		0.25		0.75				
		$f_2$	0.8	0.5				
Penetranzen		$f_1$	0	0.05				
		$f_0$	0	0.05				
Prävalenz		0.072		0.0545		0.0589		

Extremes Modell S2								
Haplotyp h		Population 1		Population 2		Gesamtpopulation		
$M_1$	$M_2$	$P(h D)^a$	$P(h d)$	$P(h D)$	$P(h d)$	$P(h D)$	$P(h d)$	
1	1	0.09	0.09	0.25	0.25	0.135	0.232	
1	2	0.81	0.81	0.25	0.25	0.652	0.314	
2	1	0.01	0.01	0.25	0.25	0.078	0.223	
2	2	0.09	0.09	0.25	0.25	0.135	0.232	
Allelfrequenz 1( $M_1$ )		0.9		0.5		0.787		
Allelfrequenz 1( $M_2$ )		0.1		0.5		0.213		
Krankheitsallelfrequenz		0.51		0.05				
Anteil an Population		0.2		0.8				
		$f_2$	0.6	0.27				
Penetranzen		$f_1$	0.2	0.09				
		$f_0$	0.0667	0.03				
Prävalenz		0.272		0.0363		0.0834		

Tabelle 2.6: Modelle S1 und S2 unter Stratifikation

<sup>a</sup> Häufigkeiten des Haplotyps h auf elterlichen Chromosomen mit dem Allel D bzw. d am DL

## 2.3 Funktionsweise des Programms zur Simulation der Familiendaten

Der SAS-Makro SIMFAM (von Michael Knapp) kann Genotypen an  $m$  Markerloci  $M_1, \dots, M_m$  mit vorgegebener LD-Struktur in Familien mit  $a$  kranken und  $u$  gesunden Kindern simulieren. Seine Arbeitsweise sei hier kurz skizziert [91]:

### 2.3.1 Mathematischer Hintergrund

**Modell** Gegeben sei ein diallelischer DL mit Allelen  $\{D, d\}$ . Der DL befinde sich zwischen  $M_b$  und  $M_{b+1}$  innerhalb von  $m$  Markern  $M_1, \dots, M_m$ . Dabei seien die Rekombinationsfrequenzen  $\theta_\beta$  ( $\beta = 1, \dots, m$ ) für die aus DL und Markern bestehende *Map* gegeben. Weiterhin seien  $p = P(D)$  die Risikoallelfrequenz und  $f_2, f_1$ , und  $f_0$  die Penetranzen am DL.

**Bezeichnungen** Sei  $a$  bzw.  $u$  die Anzahl kranker bzw. gesunder Kinder und  $M \in \{1, \dots, 6\}$  der *Mating Type* (MT) einer Familie. Mutter und Vater haben je zwei Haplotypen. Dann codiere die Zufallsgröße  $Z_k$  den Transmissions-/Krankheitsstatus des  $k$ -ten Kindes, und zwar  $Z_k \in \{1, \dots, 4\}$ , wenn das Kind krank und  $Z_k \in \{5, \dots, 8\}$ , wenn es gesund ist. So bedeutet  $Z_k = 1$  beispielsweise, dass das Kind den ersten Haplotypen von beiden Eltern erhalten hat und krank ist.  $s_{ij} = P(Z_k = j | M = i)$  bezeichne bei gegebenem Matingtyp  $i$  die bedingte Wahrscheinlichkeit, dass das  $k$ -te Kind den  $j$ -ten Transmissions-/Krankheitsstatus aufweist. Die  $s_{ij}$  ergeben sich aus den Penetranzen. Z. B. ist beim ersten MT  $DD \times DD$  das Kind homozygot und wird mit Wahrscheinlichkeit  $f_2$  krank sein. So ergibt sich für jedes der vier Transmissionsmuster  $s_{1j} = f_2/4$  ( $j = 1, \dots, 4$ ) für kranke und  $s_{1j} = (1 - f_2)/4$  ( $j = 5, \dots, 8$ ) für gesunde Kinder. Damit ist  $s_{ia} := \sum_{j=1}^4 s_{ij}$  die Wahrscheinlichkeit dafür, dass ein Kind des  $i$ -ten MTs krank ist. Die Zufallsgröße  $Y_j$  zählt die Kinder der Familie mit dem Transmissions-/Krankheitsstatus  $j$ :  $Y_j := |\{k : Z_k = j\}|$ , ( $j = 1, \dots, 8$ ). Schließlich sind die Zufallsgrößen:  $A := \sum_{j=1}^4 Y_j$  bzw.  $U := \sum_{j=5}^8 Y_j$  die Anzahlen kranker bzw. gesunder Kinder in der Familie.

**Grundidee** Bei der Simulation ist es sinnvoll, nicht vom Indexfall auszugehen, sondern von den Eltern, deren Mating Type in einem bestimmten Anteil in der Population vorkommt. Sei also eine Familie mit  $a + u$  Kindern gegeben: Dann lässt sich die Wahrscheinlichkeit für MT  $i$  und Transmissions-/Krankheitsstatus  $y$  bei gegebener Anzahl  $a$  kranker Kinder als Produkt dreier Faktoren darstellen:

$$\begin{aligned} P(M = i, Y = y | A = a, U = u) &= P(M = i | A = a, U = u) \\ &\cdot P((Y_1, \dots, Y_4) = (y_1, \dots, y_4) | A = a, U = u, M = i) \quad (2.8) \\ &\cdot P((Y_5, \dots, Y_8) = (y_5, \dots, y_8) | A = a, U = u, M = i) \end{aligned}$$

D. h. der erste Faktor ist die bedingte Wahrscheinlichkeit für den  $i$ -ten MT, gegeben, dass  $a$  Kinder krank sind. Der zweite Faktor enthält die Wahrscheinlichkeit der Transmissionsmuster  $(Y_1, \dots, Y_4)$ , die sich aus der Multinomialverteilung  $((s_{ij}/s_{ia}))_{j=1, \dots, 4}$  bestimmen lassen, wenn der  $i$ -te MT und  $a$  kranke Kinder gegeben sind. Analog enthält der dritte Faktor die multinomiale Wahrscheinlichkeit der Transmissionsmuster für gesunde Kinder, gegeben wiederum der  $i$ -te MT und  $u$  kranke Kinder.

### Algorithmus

1. In einem ersten Schritt wird die Matrix  $S = ((s_{ij}))$  ( $i = 1, \dots, 6; j = 1, \dots, 8$ ) von bedingten Wahrscheinlichkeiten des Transmissions-/Krankheitsstatus berechnet. Weiterhin wird aus  $p$  der Vektor  $m = (m_1, \dots, m_6)$  der Wahrscheinlichkeiten der MTs nach HWE kalkuliert.

2. Danach wird der MT am DL gemäß der MT-Verteilung bei Familien mit  $a$  kranken und  $u$  gesunden Kindern bestimmt. (Das entspricht dem ersten Faktor in (2.8).)
3. Das Transmissionsmuster für ein krankes Kind wird entsprechend Faktor zwei in (2.8) für jedes kranke Kind bestimmt.
4. Analog wird das Transmissionsmuster für jedes gesunde Kind (Faktor drei) simuliert.
5. Gemäß der Allele D bzw. d am Risikolocus und der vorgegebenen LD-Struktur werden jetzt die elterlichen Allele an den Markergenorten bestimmt.
6. Schließlich wird die Allel-Transmission von beiden Eltern auf die Kinder in Abhängigkeit von den Rekombinationsraten  $\theta_\beta$  simuliert. Hat z. B. ein krankes Kind nach dem im Schritt 3 bestimmten Transmissionsmuster das erste Allel am DL vom Vater erhalten, so erhält es mit Wahrscheinlichkeit  $(1 - \theta_b)$  auch das erste Allel am benachbarten Marker  $M_b$ . Ganz analog werden die Allele an den anderen Markern simuliert.

### 2.3.2 Implementierung als SAS-Makro

Der Makro %SIMFAM hat folgende Eingabe-Parameter:

<b>anzf</b>	Anzahl zu erzeugender Familien
<b>n_a</b>	Anzahl kranker Kinder ( $n_a \geq 1$ )
<b>n_u</b>	Anzahl gesunder Kinder ( $n_u \geq 0$ )
<b>p_D</b>	Häufigkeit des Risikoallels D in der Population
<b>penet</b>	Penetranzen $f_2 f_1 f_0$ am DL
<b>anzh</b>	Anzahl Haplotypen
<b>anzm</b>	Markeranzahl (hier: 5)
<b>posd</b>	Position des DL (hier: 3)
<b>thetas</b>	paarweise Rekombinationsfrequenzen zwischen den Markern bzw. mit dem DL

Ein Makro %DISEASE\_PARA berechnet anhand der Häufigkeit des Risikoallels,  $p_D$ , die Häufigkeiten der sechs MTs unter HWE. Für jeden der sechs Mating Types werden mit Hilfe der Penetranzen die Wahrscheinlichkeiten kalkuliert, dass ein Kind aus dem  $i$ -ten Mating krank ist. Schließlich wird die Verteilung der MTs in Familien mit  $n_a$  kranken und  $n_u$  gesunden Kindern berechnet.

Weiterhin wird ein Dataset **freq** benötigt, das für alle Haplotypen die Häufigkeiten **freq\_dis** bzw. **freq\_ndis** auf elterlichen Chromosomen mit bzw. ohne Risikoallel D enthält. D. h. in jeder der **anzh** Beobachtungen ist in **anzm** Spalten der Haplotyp codiert und zwei weitere Spalten enthalten die entsprechenden Häufigkeiten. Damit können die in den Tabellen 2.2 bis 2.4 definierten Häufigkeiten eingegeben werden.

Schließlich muss der Startwert für den Zufallszahlengenerator in einem Dataset mit einer Beobachtung und einer Variablen übergeben werden.

Mit diesen drei Datasets kann der Makro %SIMFAM aufgerufen werden:

```
%SIMFAM(infile, anzf, anzm, anzh, freq, posd, thetas, rand, out1, out2);
```

Dabei bedeuten außer den oben definierten Parametern:

<b>infile</b>	Dataset mit den durch %DISEASE_PARA ermittelten Kenngrößen
<b>freq</b>	Name des Datasets mit den Haplotypfrequenzen
<b>rand</b>	Dataset mit dem Startwert für den Zufallszahlengenerator, nach Programmende wird diese Variable aktualisiert.
<b>out1</b>	Ausgabe-Dataset mit den erzeugten Familien
<b>out2</b>	ASCII-Ausgabedatei mit simulierten Familien

Nach dem oben skizzierten Ablauf simuliert der Makro **anzf** Familien mit den vorgegebenen Eigenschaften. Die simulierten Familiendaten stehen danach als SAS-Dataset und als ASCII-Datei zur Verfügung.

## Kapitel 3

# Auswahl der Assoziationsanalyse-Verfahren

### 3.1 Recherche

Im Fokus dieser Simulationsstudie liegen Tests auf Assoziation

1. zwischen einem binären Merkmal
2. und Genotypdaten mehrerer Marker,
3. die eng gekoppelt sind,
4. für allgemeine Kernfamilien;
5. Verfahren, die keine speziellen Vorkenntnisse benötigen
6. und technische Grundvoraussetzungen erfüllen.

1. Das betrachtete binäre Krankheits-Merkmal (*trait*) ist in der Regel der Krankheitsstatus.

2./3. Es geht darum, den höheren Informationsgehalt von Multi-Marker-Daten zu nutzen. Gruppen eng gekoppelter Marker erhält man z. B., wenn eine Kopplungsanalyse eine Risikoregion identifiziert hat. Dann wählt man in dieser Region einige Marker aus, um den vermuteten Risiko-Genort feiner zu lokalisieren. Auch Kandidaten-Gen-Studien, wo Voruntersuchungen Variationen an einem Gen als potentielle Krankheitsursache ermittelt haben, führen auf diese Situation.

4. Die klassische Familienstruktur des TDT ist das Trio aus kranker Person - dem *Indexfall* - und seinen Eltern. Die Verfahren sollten aber auch aus den Informationen weiterer kranker oder gesunder Geschwister Nutzen ziehen.

5. Viele Verfahren, die von - teilweise sehr spezifischen - Annahmen oder auch Vorkenntnissen ausgehen, z. B. einem Vererbungsmodus der Krankheit, sind in ihrer Anwendung eingeschränkt. Hier untersuchte Methoden sollten auch ohne spezifische a-priori-Kenntnisse hinreichend effektiv funktionieren.

6. Für eine Simulationsstudie muss der Algorithmus selbstverständlich in ein Computerprogramm implementiert sein. Weiterhin sollte er so praktikabel zu handhaben sein, dass er einen automatisierten Ablauf nicht behindert. Das schloss insbesondere mehrstufige Verfahren aus, die auf verschiedene Programme angewiesen sind.

Im Frühjahr 2008 wurden durch das Sonderheft 'Advances in Family-Based Association Analysis', Human Heredity 66(2), 15 Jahre Familien-basierte Assoziationsverfahren gewürdigt. Das bezieht sich auf den Transmission Disequilibrium Test (s. 1.2.2), der zwar nicht das erste dieser Verfahren war, aber viele Autoren zu Erweiterungen inspiriert hat. Da sich Assoziation zwischen Marker und DL als Linkage Disequilibrium widerspiegelt (s. 1.1.2), wurde folgerichtig nach LD-Methoden für Haplotypen gesucht. Die Website <http://www.nslj-genetics.org/ld/> sammelt seit

Jahren Artikel zu dieser Problematik. Die Seite <http://linkage.rockefeller.edu/soft/> trägt dazu korrespondierend Programme zur Linkage- und Assoziationsanalyse zusammen. Weiterhin enthält auch die Site <http://csg.sph.umich.edu/> etliche Links zu statistischen Verfahren für genetische Fragestellungen.

So wurden im Zeitraum von Frühjahr bis Sommer 2006 o. g. Webseiten und typische Zeitschriften der genetischen Epidemiologie ab etwa 2000 recherchiert. PubMed-Recherchen sowie Rück- und Querverweise rundeten das Gesamtbild ab. Die gefundenen Methoden und Programme werden im folgenden Abschnitt aufgelistet.

## 3.2 Kurzvorstellung der Methoden und Programme

### 3.2.1 Übersicht über die Verfahren

Nur wenige der recherchierten Verfahren kamen als Kandidaten der Simulationsstudie in die Vorauswahl. Sie werden im Folgenden durch [V] gekennzeichnet. Daneben werden diese Kürzel gebraucht:

- [F] Die Methode ist anders fokussiert. Z. B. wurde sie für Case-Control-Daten entwickelt oder sie hat ein genomweites Screening zum Ziel.
- [I] Einige Verfahren, wie z. B. CPG (*conditioning on parents genotypes*) bzw. CEPG (*conditioning on exchangeable parental genotypes*) [37, 38], stehen nach aktuellem Kenntnisstand nicht öffentlich als Software zur Verfügung. Wieder andere benötigen zur Analyse zusätzliche Software.
- [D] Bei Programmen, wie z. B. PLINK war die Wirkungsweise der zu Grunde liegenden Algorithmen nicht genau dokumentiert oder referenziert.
- [N] Von Programmen wie TRANSMIT sind aus der Literatur Einschränkungen bzw. Nachteile bekannt. Bei anderer Software offenbarten Probeläufe Mängel, die sie von der Studie ausschlossen.
- [Z] Schließlich wurden eine Reihe von Algorithmen und Programmen erst nach Beginn der Simulation publiziert.

Die Systematisierung der Verfahren ist kompliziert, weil diese sich nach ganz verschiedenen Gesichtspunkten einordnen lassen und es keine überschneidungsfreie Klassifikation gibt. Dennoch sind gewisse Hauptlinien erkennbar:

- Alle hier untersuchten Verfahren sind 'Multi-Marker-Methoden' im weiteren Sinne. Im engeren Sinne werden von Rakovski & al. [126] jene Methoden so bezeichnet, die eine Statistik auf mehrere Marker gleichzeitig anwenden, aber ohne die Auflösung der Phase auskommen. Dass sie die Korrelationsstruktur, das LD, zwischen den Markern nicht auswerten, kann u. U. einen Powerverlust bedeuten.
- Sehr viele Algorithmen lassen sich auf TDT-Verallgemeinerungen zurückführen. Sie basieren auf dem Gedanken, die Anzahl transmittierter und nicht transmittierter Allele zu vergleichen.
- Das Konzept von Rabinowitz & Laird [125] - Bedingen auf suffiziente Statistiken - initiierte eine Reihe von sog. FBATs (*family based association tests*), die Eingang ins gleichnamige Programm fanden.
- F. Dudbridge verwendet eine *retrospective likelihood*, d. h. er bedingt die Likelihood der Kindes-Genotypen auf elterliche Genotypen und den Krankheitsstatus [48]. Sein Modell ermöglicht, sowohl relative Risiken von Haplotypen als auch von Kovariablen zu schätzen und zu testen [49].

- Schließlich basieren mehrere Verfahren auf log-linearen Ansätzen: LCPT [148], HAPLIN [64], TRIMM [141] und PCPH [9].
- Konzepte, die von sehr spezifischen Modellannahmen ausgehen, wurden nicht einbezogen: Das sind z. B. evolutionsbasierte Ansätze und Bayessche Verfahren. Auch *haplotype sharing methods* wurden hier nicht integriert. Der Vollständigkeit halber wurden einige Standardprogramme der Kopplungsanalyse hinsichtlich integrierter Assoziationsverfahren untersucht.

### 3.2.2 Multimarker-Tests

McIntyres Vorschlag, den p-Wert für die maximale TDT-Statistik über k diallelische Marker nach Bonferroni zu korrigieren, ist noch nicht als Multimarker-Verfahren zu werten [110].

#### HOTEL\_FAM

Die multivariate Verallgemeinerung des t-Tests ist Hotelling's  $T^2$ -Test. Xiong et al. [165] wandten diesen auf den Genotyp-Vektor zweier diallelischer Marker an und zeigten, dass diese Statistik mindestens so gute Power wie der  $\chi^2$ -Test auf Assoziation hat. Fan & Knapp erweiterten sie auf *haplotype blocks* für qualitative Merkmale sowohl in Fall-Kontroll- [58] als auch in Familienbasierten Studien [57]. Der Algorithmus liegt auf <http://www.uni-bonn.de/~umt70e/soft.htm> als SAS[4]-Makro implementiert vor [V].

#### TRIMM

Bezeichnen  $C_i$ ,  $M_i$  und  $F_i$  die Anzahl eines Risikoallels von erkranktem Kind, Mutter, bzw. Vater am Marker i, so ist  $D_i = 2C_i - (M_i + F_i)$  die gepaarte Differenz aus dem Genotyp des Kindes und dem nicht transmittierten Komplement. Daraus definierten Shi et al. [141] eine  $\max Z^2$ -Statistik, deren Verteilung durch Permutation bestimmt wird. Mit modifizierten Differenztermen  $D_i$  sind auch Tests auf maternale Effekte bzw. Imprinting möglich.

Aus der Überlegung, dass  $\max Z^2$  weniger wirkungsvoll ist, wenn der unbekannte Krankheitslocus nicht unter den Markern ist, kombinierten die Autoren diese Statistik mit Hotellings  $T^2$  zu einer  $\sum \log(P)$ -Statistik.

Diese Verfahren liegen als R-Paket TRIMM (**TRI**ad **MULTI** **MARKER**) implementiert auf der Website <http://www.niehs.nih.gov/research/resources/software/triad/index.cfm> vor, allerdings erst nach Beginn der Simulation [Z].

#### APRICOT

Der *Adaptive PRINCIPAL COmponent Test* wertet die Kovarianzstruktur der Allelübertragung über alle M Marker durch eine Hauptkomponentenanalyse aus. Indem die Statistik  $T_k$  nur auf Basis der ersten k Hauptkomponenten definiert wird, die den Hauptanteil genetischer Variation enthalten, wird die Anzahl der Freiheitsgrade reduziert [102]. Wie Lee feststellt, kann der Risikolocus nicht präzise lokalisiert werden. Simulationen bestätigten den Einwand, dass APRICOT versagt, wenn gerade die letzten Hauptkomponenten im wesentlichen den Risikolocus enthalten [141]. Von APRICOT wurde keine Implementation gefunden [I].

### 3.2.3 TDT-Erweiterungen

#### TDTHAP

Clayton & Jones [33] hatten 1999 einen auf Haplotypen erweiterten TDT vorgestellt. Diese Methode hat den Nachteil, dass die Phasenauflösung vorausgesetzt wird. TDTHAP lag lange Zeit als R-Paket implementiert vor, wurde aber vor kurzem von der Webseite entfernt [N].

## TRANSMIT

D. Clayton erweiterte den TDT außerdem durch einen Score-Test für Multimarker-Daten [32]. Die Varianz wird dabei mit einem robusten Varianzschätzer nach Huber [84] berechnet. Die Methode liegt auf <http://www-gene.cimr.cam.ac.uk/clayton/software/> als C/C++-Programm TRANSMIT vor. Das Verfahren setzt HWE und eine genetisch homogene Population voraus. Verletzung dieser Annahmen hat z. B. zur Folge, dass das Fehlerniveau 1. Art nicht eingehalten wird [109], [51] [N].

## APL

In einer Kopplungs-Region ist die Allelübertragung von Eltern auf mehrere kranke Kinder korreliert (vgl. Formel (1.14)). Das führt z. B. bei ASP-Daten mit fehlenden elterlichen Genotypen zu verfälschten Varianzschätzungen und in der Konsequenz zu einem erhöhten Fehler 1. Art. APL (*Association in the Presence of Linkage*) [109] stellt sich diesem Problem. Es nutzt als Statistik  $T_i$  die Differenz der Häufigkeiten des Allels 1 zwischen Eltern und ASPs. Dabei wird die Varianz durch Bootstrap-Verfahren berechnet [V].

## FAMHAP

Famhap schätzt Haplotyp-Frequenzen eng gekoppelter Marker durch einen EM (*expectation maximization*)-Algorithmus auf der Basis der Likelihood [19]. Eine TDT-ähnliche Statistik vergleicht die Häufigkeiten übertragener vs. nicht übertragener Haplotypen. Da beim Test auf Haplotyp-Effekte von vornherein nicht klar ist, welche Kombination starke Effekte zeigt, bietet Famhap den Test aller Kombination aus  $m$  Markern an. Aus dem minimalen Haplotyp-spezifischen p-Wert wird durch Simulation ein hinsichtlich multiplen Testens korrigierter 'globaler' p-Wert für jede Markerkombination kalkuliert. Auch ein overall-p-Wert über alle  $M_k$  wird mittels Monte Carlo-Methoden ermittelt.

Von <http://famhap.meb.uni-bonn.de/> kann das Programm heruntergeladen werden [V].

### 3.2.4 FBAT

Das Acronym FBAT wird im weiteren Sinne für den Typus *family based association tests* verwendet [105]. Im engeren Sinne bezeichnet es ein auf <http://biosun1.harvard.edu/~fbat/fbat.htm> verfügbares Programm [V]. Es enthält Tests auf Linkage bzw. Assoziation für Familiendaten mit qualitativen und quantitativen Krankheits-Merkmalen sowie Überlebenszeitdaten.

## PBAT

Das auf <http://www.biostat.harvard.edu/~clange/default.htm> verfügbare Programm PBAT vermag die Power für FBAT zu berechnen. Weiterhin enthält es ein Modul für Assoziationsanalyse mit multiplem Phänotyp (FBAT-GEE [100]). Die Software *P<sup>2</sup>BAT* (<http://www.people.fas.harvard.edu/~tjhoffm/pbatR.html>) verbindet die Funktionalität von PBAT mit einer grafischen Benutzeroberfläche. Außerdem bietet sie die Möglichkeit paralleler Berechnung auf Computersystemen mit mehreren Knoten an [76] [F].

### 3.2.5 Auf bedingter Likelihood basierende Ansätze

## UNPHASED

Eine retrospektive Likelihood reflektiert die Wahrscheinlichkeit der Genotypen, bedingt auf die Krankheitsmerkmale aller Nachkommen und ggf. auf Kovariablen. Die Nullhypothese  $H_0 : \beta = 0$  ist gleichbedeutend mit: Alle Haplotypen haben gleiches relatives Risiko. Auf Basis dieser Likelihood können die Parameter geschätzt und Score-Tests bzw. Likelihood-Quotiententests durchgeführt werden [51]. Dabei erhält man p-Werte aus der asymptotischen  $\chi^2$ -Verteilung bzw. durch Permutation. Die Software ist auch in der Lage, Haplotyp-Daten erweiterter Familien auf Assoziation zu testen.

Das Programm kann von <http://www.mrc-bsu.cam.ac.uk/personal/frank/software/unphased/> heruntergeladen werden. [V]

### 3.2.6 Log-lineare Modelle

#### PCPH

Die *Projection Conditional on Parental Haplotypes Method* arbeitet mit einem log-linearen Modell für das Krankheitsrisiko. Auf der Basis eines Ansatzes von Allen et al. [10] werden die Parameter über einen *Projektions-Maximierungs-Algorithmus* geschätzt. Anhand einer Simulation zeigen die Autoren, dass die zwei daraus abgeleiteten Tests höhere Power als FBAT haben, auch wenn sie keine exakten p-Werte liefern [9]. Das Verfahren wurde ins Programm PCPH auf <http://www.duke.edu/~asallen> implementiert.

Leider kam diese Veröffentlichung zu spät, um noch in die Studie einbezogen zu werden [Z].

#### HAPLIN

Im Programm HAPLIN werden Effekte einer Kopie / zweier Kopien eines Allels in einem log-linearen Modell parametrisiert und geschätzt [64]. Beim Konzept der 'reziproken Referenz' wird ein Haplotyp gegen die Summe der anderen getestet.

Auf <http://www.uib.no/smis/gjessing/genetics/software/haplin/> ist die Software erhältlich. Vermutlich weil HAPLIN als R-Programm die Daten im Speicher halten muss (vgl. [34]), kam es bei Testläufen schnell an Ressourcen-Grenzen. Außerdem ist es nur für Trios und nicht für allgemeine Kernfamilien geeignet [N, F].

### 3.2.7 Genie

Das Java-Programmpaket **Genie** mit verschiedenen Single-Marker- und Haplotyp-Assoziationstests für Stammbäume ist auf <http://www.genepi.med.utah.edu/Genie/index.html> verfügbar. Die Algorithmen folgen einer einfachen Idee [11]:

1. Zunächst werden die Allelfrequenzen der interessierenden Marker geschätzt.
2. Die Allele werden den Foundern im vorliegenden Verhältnis zugeordnet und zufällig den Nachkommen weitergegeben.
3. Die interessierende Statistik wird anhand der Genotyp-Verteilung unter  $H_0$  und der beobachteten Phänotypen berechnet. Durch N-malige Wiederholung der Schritte 2 und 3 wird die Nullverteilung (bedingt auf den Stammbaum und die Phänotypen) erzeugt.
4. Damit kann man den p-Wert für die Statistik der Beobachtung ermitteln.

Genie ist eins der wenigen Verfahren, das in der Lage ist, auf Assoziation in allgemeinen Stammbäumen zu testen. Die strenge Forderung des Moduls **PedGenie**, die Auflösung der Phase zu kennen, wurde erst nach Beginn der Simulation im kürzlich erschienenen Modul **hapMC** fallen gelassen [N,Z].

### 3.2.8 Sichtung weiterer Programme

#### Assoziationstests in Programmen der Kopplungsanalyse

Standardsoftware der genetischen Epidemiologie wurde daraufhin geprüft, ob sie Familien-basierte Assoziationstests mit Haplotypen enthält.



**Genehunter und Allegro** Genehunter (<http://www.broad.mit.edu/ftp/distribution/software/genehunter/>) enthält neben vielen Kopplungsverfahren mit den Kommandos `tdt`, `tdt2`, ..., `tdt5`, `perm1` und `perm2` auch (erweiterte) Transmission Disequilibrium Tests. Leider sind die dahinterliegenden Verfahren nicht genau dokumentiert [D]. Das einzige Zitat in diesem Zusammenhang verweist auf einen älteren Artikel von Dudbridge et al. [53]. ALLEGRO 2 [71] ist eine schnelle Version von Genehunter.

**Merlin** Das auf <http://www.sph.umich.edu/csg/abecasis/Merlin/> zu findende Programm wird als *fastest pedigree analysis package* empfohlen. Es basiert auf einer Arbeit von Abecasis et al. [6] und enthält keine Assoziationsverfahren mit Haplotypen. [F]

**Mendel** Mendel wurde als gemeinsame Plattform für viele Probleme der genetischen Epidemiologie und Populationsgenetik konzipiert [101]. Neben vielen anderen Features wurde auch eine als LLDM (*linkage-linkage disequilibrium mapping*) bezeichnete Methode implementiert [164]. LLDM wird als verheißungsvolles Werkzeug für genomweites Screening angesehen [27]. Allerdings ist es ein Fall-Kontroll-Verfahren [F]. Weiterhin muss man a-priori das genetische Modell kennen.

### Genomweite Assoziationsverfahren

GWASs sind auf ein genomweites Screening (bei nur wenigen Vorkenntnissen) gerichtet. Ihre Sichtung brachte keinen Zugewinn an Methoden für eng gekoppelte Marker.

**snpmatrix** Clayton & Leung [34] implementierten einige Standardmethoden zu genomweiten Assoziationsstudien in R, wie z. B. Cochran-Armitage's Trendtest oder  $\chi^2$ -Statistiken. Diese Verfahren beruhen auf Populations-basierten Studien [F].

**PLINK** Die auf <http://pngu.mgh.harvard.edu/~purcell/plink/> erhältliche Software ist durch den aktuellen Boom genomweiter Assoziationsstudien bekannt geworden. PLINK umfasst Linkage- und Assoziationsverfahren [122] und enthält eine ganze Reihe Features, z. B. den  $T^2$ -Test, den TDT für Einzelmarker und Haplotypen sowie den parenTDT. Leider sind im Handbuch keine Referenzen zu Methoden zu finden, weshalb eine genaue Zuordnung zu in der Literatur bekannten Verfahren kaum möglich ist [D].

Beim Testlauf gab es wiederholt Programmabbrüche. Nach einer diesbezüglichen Anfrage per E-Mail wurde auf die nächste Version verwiesen, die allerdings für den Start der Simulationen zu spät kam.

**GWAVA** Das auf <http://www.sph.umich.edu/csg/abecasis/gwava/index.html> vorliegende Programm (**Genome-Wide Association Visual Analyzer**) ist auf die extremen Anforderungen von genomweiten Assoziationsstudien zugeschnitten. Die durch einen hocheffizienten EM-Algorithmus gewonnenen Haplotyp-Frequenzschätzer können weiter für den Test auf Assoziation genutzt werden (logistisches Regressionsmodell) [F].

### Haplotyp-Sharing-Methoden

Die folgenden Verfahren gehen von der Intuition aus, dass heute beobachtbare Haplotypen in der Nähe eines Krankheitslocus die Entwicklung einer Krankheit widerspiegeln. Vielfach nutzen diese Methoden Maße für die Ähnlichkeit von Haplotypen, wie z. B. MILC (*maximum identity length contrast*) [25]. Dazu gehören z. B. TRIMHAP [107], HS-TDT (*haplotype sharing TDT*, [140]), HSC (*haplotype sharing correlation*, [123, 124]), SP-TDT (*sequential peeling TDT*, [169]) und Haplotyp TDT [168].

Nach Whittemore & Halpern [162] können Haplotyp-basierte Assoziationsverfahren wesentlich höhere Power erreichen als Haplotyp-Sharing-Methoden. Diese wurden nicht in die vorliegende Studie einbezogen [F].

Die beiden Programme ET-TDT und eHap führen die Analyse auf der Basis evolutionärer Beziehungen zwischen Haplotypen durch [138, 139]. Sie benötigen zur Erzeugung eines Haplotyp-Stammbaumes zusätzliche Software, wie z. B. EB oder Entropy Blocker [I].

### R-Packages

Programme wie TRIMM und HAPLIN waren der Anlass, unter Paketen für R / S-plus® nach Familien-basierten Assoziationsverfahren zu suchen. Auf <http://www.bioconductor.org/packages/release/Software.html> stehen die meisten Packages zur Verfügung. Allerdings sind Familien-basierte Assoziationsmethoden (im Gegensatz zu Verfahren zur Genexpressionsanalyse) eher die Ausnahme.

Die oben genannten Verfahren TDTHAP (3.2.3) und FBAT (3.2.4, aber ohne Haplotyp-Analyse) gibt es inzwischen auch für R.

Die beiden Pakete `haplo.stats` bzw. `multigene` z. B. gehen auf D. Schaid zurück [134], betreffen aber beide Fall-Kontroll-Studien [F].

Auf Bayes-Verfahren basiert das Programm BARS (<http://wpicr.wpic.pitt.edu/WPICCompGen>). Es enthält einen effektiven und robusten Screening-Test [175] [F].

Das Programmpaket `gap` (`genetic analysis package` [178]) bietet eine ganze Reihe von Features, allerdings keine Assoziationsverfahren mit Haplotypen [F].

### 3.2.9 Programme der engeren Wahl

Die Programme FBAT, Famhap, Hotel\_Fam, Unphased und APL kamen in die engere Wahl.

APL benötigt für die Analyse eine `.dat`-Datei mit recht speziellen Angaben, wie z. B. Penetranzen und Allelfrequenzen. Bei Pilot-Analysen deuteten teilweise unmögliche p-Werte auf Singularitäten hin. Wiederholte Analysen mit identischen Ausgangsdaten lieferten teilweise deutlich differierende p-Werte. Diese sind offensichtlich auf Abweichungen der Varianzschätzungen (bis 10%!) zurückzuführen. Aus diesem Grund wurde APL nicht in die Studie einbezogen [N].

# Kapitel 4

## Charakteristika der Programme für die Studie

### 4.1 FBAT

#### 4.1.1 Grundlagen

Durch Bedingen auf potentiell 'verzerrende Faktoren' hatten Rabinowitz & Laird [125] die Plattform für eine ganze Reihe von Tests geschaffen: Bedingen auf die elterlichen Genotypen schaltet den Einfluss von Populationsstratifikation aus, Bedingen auf kindliche Phänotypen eliminiert die Sensitivität gegenüber deren Fehlspezifikation [80]. FBAT verwendet (wie auch Unphased, vgl. 3.2.5) einen retrospektiven Ansatz, d. h. die Genotypen werden als abhängige Variablen bei gegebenem Krankheitsstatus betrachtet.

#### Modellannahmen

1. keine Rekombination zwischen den genetischen Markern [80], [95],
2. Rekrutierung der Probandenfamilien allein auf Basis des Krankheits-Merkmals, also ohne Berücksichtigung des Genotyps.

Zur Verteilung des Krankheitsmerkmals und der elterlichen Genotypen sowie zum genetischen Modell müssen keine Annahmen gemacht werden [94]. FBAT testet standardmäßig die Hypothese  $H_0^I$ , aber man kann auch die Hypothese  $H_0^A$  prüfen.

#### Verfahren

Seien  $i = 1, \dots, n$  Kernfamilien mit  $j = 1, \dots, n_i$  Nachkommen gegeben. Sei weiterhin  $X_{ij}$  eine Funktion des Genotypen des j. Kindes in Familie i (bzw. für simultane Tests mehrerer Genotypen ein Vektor). Bei additiver Codierung z. B. zählt  $X_{ij}$  einfach die Anzahl eines Allels. Horvath et al. [79] zeigten, dass diese Codierung auch bei anderen Vererbungsmodi noch eine gute Power liefert. Deshalb verwendet FBAT diese als Voreinstellung.

$T_{ij}$  sei als Funktion des Phänotypen  $Y_{ij}$  angenommen: Beispielsweise kann man durch die Verschiebung  $T_{ij} = Y_{ij} - \mu_{ij}$  quantitative Merkmale zentrieren oder für qualitative Merkmale erreichen, dass gesunde Geschwister einbezogen werden.

Die allgemeine FBAT-Statistik basiert auf der Kovarianz zwischen Krankheitsmerkmal und

Genotyp [97], [96]:

$$S_i = \sum_j T_{ij} X_{ij} \quad (4.1)$$

$$\text{bzw.} \quad S = \sum_{ij} T_{ij} X_{ij} \quad (4.2)$$

$$\text{und damit} \quad U = S - E[S] \quad (4.3)$$

Unter  $H_0$  hat  $U$  den Erwartungswert 0. Durch Nutzung der bedingten Verteilung (s. o.) kann man die Varianz  $V[S_i]$  berechnen und eine asymptotisch  $N(0, 1)$ -normalverteilte Testgröße  $Z$  bzw., wenn  $X_{ij}$  ein Vektor ist, eine asymptotisch  $\chi^2$ -verteilte Testgröße  $C^2$  ableiten:

$$Z = \frac{U}{\sqrt{V}} \quad (4.4)$$

$$\text{bzw.} \quad C^2 = \mathbf{U}' \mathbf{V}^{-1} \mathbf{U} \quad (4.5)$$

$$\text{mit} \quad V = \sum_i V[S_i] \quad (4.6)$$

Dabei ist  $\mathbf{V}^{-1}$  eine (verallgemeinerte) Inverse von  $\mathbf{V}$ . Die Statistik  $C^2$  hat soviel Freiheitsgrade wie der Rang von  $\mathbf{V}$ .

Diese allgemeine Definition der Testgröße (4.2) ermöglicht, FBAT sowohl für dichotome und quantitative Merkmale sowie *time-to-onset*-Daten als auch die  $X_{ij}$  je nach genetischem Modell anzupassen. Da die Tests auf der minimalen suffizienten Statistik basieren, halten sie unabhängig von der Modellwahl das Fehlerniveau 1. Art ein [125]. Standardmäßig bezeichnet  $T_{ij}$  den Krankheitsstatus, d. h.  $T_{ij} = 1$  für ein krankes bzw.  $T_{ij} = 0$  für ein gesundes Kind.

### 4.1.2 Multi-Marker-Test

Liegen  $m = 1, \dots, M$  eng gekoppelte Marker vor, so bilden die  $U$ -Statistiken nach (4.1), (4.3) einen Vektor  $U_{MM} = (\sum_i U_i^1, \dots, \sum_i U_i^M)^T$  [126]. Mit der empirischen Kovarianz- bzw. Korrelationsmatrix  $V_E$  bzw.  $C_E$  und der Diagonalmatrix  $D_T$  mit den wahren Varianzen der  $U_{MM}$ -Komponenten wird eine korrigierte Varianzmatrix definiert

$$V_A = D_T^{1/2} C_E D_T^{1/2} \quad (4.7)$$

und damit die asymptotisch  $\chi_r^2$ -verteilte Multi-Marker-Statistik

$$S_{MM} = U_{MM}^T V_A^{-1} U_{MM} \quad (4.8)$$

Wie oben bedeutet  $V_A^{-1}$  eine verallgemeinerte Inverse von  $V_A$  und  $r$  ist der Rang von  $V_A$ . Den Multimarker-Test ist für beide Hypothesen  $H_0^L$  bzw.  $H_0^A$ , aber nur für das additive Modell formuliert [166], [126].

### 4.1.3 Haplotyp-Test

Angenommen die Phasenauflösung der kindlichen Genotypen  $G_{ij}$  sei bekannt. Wie oben sei  $T_{ij}$  der Krankheitsstatus und  $X_{ij}$  eine Funktion der Haplotypen des  $j$ . Nachkommens der  $i$ . Familie. Zählt z. B.  $X(G_{ij})$  die Anzahl eines bestimmten Haplotyps, erhält man exakt die FBAT-Statistik für einen mehrallelischen Marker, indem jeder Haplotyp sozusagen als Allel behandelt wird. Analog (4.3) berechnet man die Teststatistik [80]:

$$U = \sum_{ij} T_{ij} [X(G_{ij}) - E(X(G_{ij}))]. \quad (4.9)$$

Da normalerweise die Phase unbekannt ist, behilft man sich dadurch, dass in obiger Formel die mit den beobachteten Genotypen kompatiblen Haplotypen mit der Wahrscheinlichkeit ihres Auftretens  $w_{G_{ijk}}$  gewichtet werden:

$$X(G_{ij}) = \sum_k X(G_{ijk})w_{G_{ijk}} \quad (4.10)$$

Dabei wird über die möglichen Haplotypen summiert, wobei  $\sum_k w_{G_{ijk}} = 1$  gilt. Wiederum wird die erweiterte bedingte Verteilung auf der Basis der suffizienten Statistik für die Phasenauflösung des kindlichen Genotyps verwendet, um Erwartungswert und Kovarianzmatrix von  $U$  unter  $H_0$  zu berechnen. Da  $\chi^2$ -Werte für jeden häufigen Haplotyp der Kombination ausgegeben werden, kann man aus dem minimalen p-Wert durch Korrektur für multiples Testen einen overall-p-Wert für die Mk gewinnen.

#### 4.1.4 Permutationstest

Anstelle den p-Wert aus der asymptotischen  $\chi^2$ -Verteilung zu erhalten, kann man ihn durch Monte-Carlo-Methoden bestimmen. Ruft man `hbat` mit der Option `-p#R` auf, wird R-mal über alle kompatiblen kindlichen Haplotypen permutiert. Der Anteil simulierter p-Werte nicht größer als der beobachtete ergibt dann den p-Wert. Standardmäßig werden `#R=100.000` Permutationen durchgeführt, allerdings hört die Prozedur bereits auf, wenn mindestens je 100 Monte-Carlo- $\chi^2$ -Werte  $\geq$  bzw.  $\leq Z_{obs}$  (dem beobachteten  $\chi^2$ -Wert) sind. Das hat zur Folge, dass für relativ große p-Werte weniger Zyklen, für sehr kleine p-Werte viele Zyklen benötigt werden. Dafür sind die Schätzungen der p-Werte umso präziser, je kleiner p ist [95]. Für diese Studie wurden `#R=4.000` Permutationszyklen festgesetzt. Obwohl das wenig erscheint, schöpfte FBAT diese Zahl in vielen Fällen gar nicht aus (vgl. 6.6.1). FBAT berechnet analog zum asymptotischen Test Monte-Carlo-p-Werte sowohl für die  $\chi^2$ -Summen-Statistik als auch für den minimalen p-Wert über alle Haplotypen. Permutationstests gibt es nur für die Hypothese  $H_0^L$ , nicht aber für  $H_0^A$ .

#### 4.1.5 Test auf Assoziation bei vorhandener Kopplung

Bei Kopplung sind die Genotypen von Geschwistern nicht unkorreliert [95]. Deshalb muss man dann die Varianz der Teststatistik anders schätzen: Mit der Option `-e` wird in dieser Situation ein empirischer Schätzer verwendet, den Lake et al. [98] auf der Basis des robusten Varianzschätzers nach White [160] abgeleitet hatten. Ist also davon auszugehen - z. B. durch einen vorangegangenen Linkage-Screen - dass Marker und Risikolocus gekoppelt sind, so muss man `fbat` oder `hbat` mit der Option `-e` starten. Das ist gleichbedeutend mit dem Test der Hypothese  $H_0^A$ .

#### 4.1.6 Weitere Optionen

##### Linear Combination Test

Xu et al. [166] leiteten auf der Grundlage eines ursprünglich für *quantitative trait loci* entwickelten *conditional mean models* [99] einen Test der 'besten Linearkombination'  $T_{LC}$  ab: Seien  $\hat{z}_m$  die Statistiken gemäß (4.4) am Marker  $m$  ( $m = 1, \dots, M$ ). Dann wird eine Linearkombination  $S = \sum_k w_m \hat{z}_m$  mit Gewichten  $w_m$  definiert, die so aus den Daten berechnet werden, dass sie die Statistik  $S$  maximieren. Dieser Test wird durch den Befehl `fbat -1` aufgerufen. In einer Simulationsstudie dieser Autoren hat dieser Test größere Power als der Multimarker- (4.1.2) und der Haplotyp-Test (4.1.3). Weil im Gegensatz zum Haplotyp-Test keine Phasenauflösung erfolgt, verkürzt sich auch die Rechenzeit wesentlich. Die Veröffentlichung kam zu spät, um den LC-Test in die Analyse einzubeziehen [Z].

##### minfreq und minsiz

Die Option `minfreq #fwert` legt die untere Schranke für Haplotypfrequenzen fest. Seltener Haplotypen werden nicht in die Analyse einbezogen. Auf Empfehlung des Handbuchs ([94], 4.1.2.2.)

wurde hier die Einstellung `-minfreq 0.05` verwendet. Mit der Option `minsize #swert` legt man die minimale Anzahl Familien fest, die für die Berechnung der Testgröße nötig ist. Es gab keinen Grund, den Default-Wert `minsize 10` zu ändern.

## model und mode

Durch die Option `model a|d|r|g` kann man die Genotyp-Scores 4.1.1 für  $X_{ij}$  festlegen: additive, dominante, rezessive oder Genotyp-Codierung. So ist z. B. bei der rezessiven Codierung  $X_{ij} = 1$ , falls der Genotyp homozygot hinsichtlich des betrachteten Allels ist und sonst  $X_{ij} = 0$ . Da der Vererbungsmodus vor der Analyse eher unbekannt ist, wurde hier die Standard-Einstellung (additive Codierung) gelassen, wo  $X_{ij}$  einfach die Kopien dieses Allels zählt.

Die Einstellung `-mode b|m|a` schaltet um zwischen biallelischer und multiallelischer Testprozedur (bzw. beide).

### 4.1.7 Ausgabe von FBAT

Entsprechend der Komplexität der Fragestellung liefert die Programmausgabe Werte auf verschiedenen Ebenen: Statistiken und p-Werte

**markerspezifisch** einzeln für jeden Marker,

**haplotypspezifisch** für alle Haplotypen einer Markerkombination,

**global** ein Mk-spezifischer p-Wert über die Haplotypen jeder Mk und

**overall** ein p-Wert über alle 31 Mk, der teilweise nach der Analyse ermittelt werden musste.

Im weiteren Verlauf werden die Markerkombinations-spezifischen p-Werte stets als 'global' bezeichnet, der eine über alle Mk ermittelte Wert dagegen als 'overall' p-Wert.

Für die Multi-Marker-Tests gibt FBAT  $\chi^2$ -Statistiken für jeden Marker aus, die jeweils für alle Kombinationen identisch sind, in denen der Marker vorkommt. Der Haplotyp-Test liefert Statistiken für jeden der  $h$  vorhandenen Haplotypen, so dass man aus deren minimalem p-Wert durch Multiplikation mit  $h$  (Bonferroni-Korrektur) einen Mk-spezifischen p-Wert gewinnen kann. Parallel dazu liefert dieses Verfahren die  $\chi^2$ -Summenstatistik mit  $h - 1$  Freiheitsgraden (globaler p-Wert). Gibt es mit Ausnahme eines Haplotyps nur solche, die seltener sind als in der Option `minfreq #fwert` vorgegeben, so werden beide Statistiken nicht berechnet.

Der Permutationstest ermittelt durch Monte-Carlo-Methoden globale p-Werte speziell für eine Mk ebenfalls für die  $\sum \chi^2$  bzw. die  $\min p$ -Statistik.

In einer Tabelle im Absatz 4.5 am Ende des Kapitels ist dargestellt, welche Werte die Ausgabe der vier Programme jeweils enthält.

## 4.2 Famhap

### 4.2.1 Grundlagen

#### Modellannahmen

1. Für die Haplotyp-Frequenzschätzung werden die elterlichen Genotypen als zufällige Stichprobe unverwandter Personen einer Population im HWE angenommen [93].
2. Zwischen den Markern sei keine Rekombination vorausgesetzt.

## Verfahren

Zhao et al. [177] hatten für Familien mit einem erkrankten Kind eine TDT-Erweiterung vorgeschlagen, die Kombinationen mehrerer Marker auf Assoziation mit Krankheit testet. Knapp & Becker [93] erweiterten diesen Test auf allgemeine Kernfamilien:

Angenommen, für  $M$  Marker liegen die väterlichen Haplotypen  $H_i, H_j$  bzw. mütterlichen Haplotypen  $H_k, H_l$  vor und es bezeichne  $P_{ik,jl}$  die Wahrscheinlichkeit, dass der Vater die Haplotypen  $H_i$  und  $H_j$  hat und  $H_i$  überträgt, die Mutter  $H_k$  und  $H_l$  besitzt und  $H_k$  an das Kind weitergibt, unter der Bedingung, dass das Kind krank ist. Es sei  $h$  die Zahl möglicher Haplotypen. Auf Basis der Anzahlen übertragener / nicht übertragener Haplotypen in folgender Kontingenztafel  $\mathbf{T}$

$$\begin{array}{cccc} & 1 & 2 & \cdots & h \\ 1 & t_{11} & t_{12} & \cdots & t_{1h} \\ 2 & t_{21} & t_{22} & \cdots & t_{2h} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ h & t_{h1} & t_{h2} & \cdots & t_{hh} \end{array} \quad (4.11)$$

hatten Spielman/Ewens [143] folgende TDT-Erweiterung eingeführt:

$$T = \frac{h-1}{h} \sum_{\gamma=1}^h \frac{(t_{\gamma\cdot} - t_{\cdot\gamma})^2}{t_{\gamma\cdot} + t_{\cdot\gamma} - 2t_{\gamma\gamma}} \quad (4.12)$$

$$\text{mit den Randsummen} \quad t_{\gamma\cdot} = \sum_{\delta=1}^h t_{\gamma\delta} \quad (4.13)$$

$$\text{sowie} \quad t_{\cdot\gamma} = \sum_{\delta=1}^h t_{\delta\gamma} \quad (4.14)$$

Um die Statistik (4.12) für allgemeine Kernfamilien nutzen zu können, waren noch drei Schritte notwendig:

- I. Da in der Regel nur Genotypen beobachtet werden, ist die Phase nicht immer eindeutig auflösbar (s. 1.1.1).

Zhao & al. [177] ersetzten nun obige Transmissionstabelle (4.11) durch eine rekonstruierte Tabelle  $\hat{\mathbf{T}}$  mit der geschätzten Anzahl  $\hat{t}_g^{ik,jl}$  von Eltern mit den Haplotypen  $H_i, H_j$  bzw.  $H_k, H_l$ , wobei  $H_i, H_k$  an den Nachwuchs übertragen werden. Unter  $H_0^L$  ist diese Matrix symmetrisch und man kann darauf die Statistik (4.12) anwenden. Wie von Knapp & Becker gezeigt wurde, ist die Symmetrie der Transmissionstabelle nicht notwendig für die Gültigkeit des Tests [93].

- II. Da die Statistik  $T$  nicht notwendig  $\chi^2$ -verteilt sein muss, wird die Verteilung durch Monte-Carlo-Verfahren bestimmt. Dabei werden für allgemeine Kernfamilien je nach Hypothese  $H_0^L$  bzw.  $H_0^A$  die Allele innerhalb verschiedener 'Äquivalenzklassen' permutiert [93]. Sei für eine Markerkombination  $B$   $T_0^B$  der beobachtete Wert der Teststatistik und  $T_r^B$  jener der  $r$ . Permutation, so ergibt sich der p-Wert durch

$$P_0^B = \frac{|r : T_r^B \geq T_0^B|}{R}, \quad (4.15)$$

wobei  $R$  die Anzahl Permutationen und  $|A|$  die Kardinalzahl einer Menge  $A$  bezeichnet.

- III. Aus den durch 2. erhaltenen unkorrigierten p-Werten für alle  $M_k$  muss nun - ebenfalls durch Permutation - ein hinsichtlich multiplen Testens korrigierter p-Wert bestimmt werden. Um eine geschachtelte Permutation [159] zu vermeiden, kommt eine Methode von Ge [63] zum Einsatz: Ein totaler p-Wert wird aus dem minimalen globalen p-Wert der Markerkombinationen auf der Basis aller Minimalwerte  $P_i^B$  der  $i$ . Permutation ( $i = 1, \dots, R$ ) ähnlich (4.15) bestimmt [20].

## 4.2.2 Algorithmus

1. Zunächst werden die Häufigkeiten jedes Haplotyps per EM (*expectation maximization*)-Algorithmus aus der Likelihood geschätzt [19].
2. Falls die beobachteten Genotypen eine mehrfache Haplotyperklärung zulassen, werden diese mit den Häufigkeiten aus 1. gewichtet.
3. Für jede Familie wird der Beitrag zu einer Kontingenztafel übertragener / nicht übertragener Haplotypen berechnet. Daraus wird für die beobachteten Daten eine erweiterte TDT-Statistik  $T_0^B$  nach Spielman & Ewens [143] für eine Markerkombination B kalkuliert.
4. Für jede Markerkombination wird der p-Wert  $P_0^B$  durch Simulation bestimmt. Dabei wird in jeder Wiederholung ein neues Replikat erzeugt, indem die Genotypen aller Kinder mit gleicher Wahrscheinlichkeit gelassen oder durch die nicht transmittierten Allele ersetzt werden.
5. Aus der Liste der minimalen rohen p-Werte jeder Replikation wird der totale p-Wert über alle Mk ermittelt.

## 4.2.3 Test auf Kopplung

Standardmäßig wird die Hypothese  $H_0^A$  getestet. Mit der Option `linkage` kann auch auf Kopplung geprüft werden. Dabei wird abweichend zu 4.2.2 bei jeder Permutation der Transmission-/Non-Transmission-Status der Genotypen für jedes Kind der Familie *unabhängig* permutiert [18].

## 4.2.4 Optionen

- Famhap führt einen Test auf Assoziation nach Zhao [93] mit allen Kernfamilien durch, die an den Markern der betrachteten Kombination voll genotypisiert sind. Vorher wurden die Häufigkeiten  $t_h$ ,  $n_h$  transmittierter bzw. nicht transmittierter Haplotypen  $h = 1, \dots, H$  berechnet. Dann verwendet die Option `zhao` (ab Version 17 `haptdt`) die Summen- $\chi^2$ -Statistik

$$TDT_{sum} = \sum_{\gamma=1}^h \frac{(t_{\gamma.} - t_{. \gamma})^2}{(t_{\gamma.} + t_{. \gamma})} \quad (4.16)$$

Die Option `zhaomax` (ab Version 17 `haptdtmax`) benutzt stattdessen die maximale TDT-Statistik

$$TDT_{max} = \max_{\gamma=1}^h \frac{(t_{\gamma.} - t_{. \gamma})^2}{(t_{\gamma.} + t_{. \gamma})} \quad (4.17)$$

- Beim Aufruf von Famhap kann man die Marker bzw. Markerkombinationen spezifizieren, die getestet werden sollen. Die Option `allcombi` veranlasst den Test aller  $2^M - 1$  Markerkombinationen.
- Nicht selten gibt es nicht nur einen über alle Replikate minimalen p-Wert  $P^{min}$ , sondern mehrere:  $P_r^{min} = P^{min}$ . Mit der Option `secondon` werden für die Festlegung der Reihenfolge auch die zweitbesten rohen p-Werte herangezogen.
- Die Option `SIMULATIONS #s` legt die Anzahl Permutationen fest, die für die Bestimmung des globalen p-Wertes durchgeführt werden. Wie für FBAT bzw. Unphased wurde der Wert `#s = 4000` gesetzt.
- Für den Test auf Assoziation an einem Marker werden die Familien ausgeschlossen, bei denen die Genotypinformationen nicht zumindest für ein Trio komplett vorliegen. Aber mit der Option `u` kann man die Information solcher Familien wenigstens zur Haplotyp-Frequenzschätzung heranziehen.



## 4.2.5 Famhap-Ausgabe

Zunächst meldet Famhap, welche Familien mit inkonsistenten Eltern- und Kinder-Genotypen ausgeschlossen werden mussten. Dann werden für alle vorhandenen Haplotypen der Markerkombination die Haplotypfrequenzen ausgegeben.

Für jede Mk erhält man für alle Haplotypen die Anzahlen, wie oft sie übertragen bzw. nicht übertragen wurden und dafür die  $\chi^2$ -Statistiken. Die Markerkombinationen sind dabei binär sortiert, d. h. sie beginnen hier mit  $M_5$  allein (Mk 1), danach  $M_4$  (Mk 2) usw. und enden mit der Kombination aller 5 Marker (Mk 31). D. h. eine binäre '1' an Stelle  $s$  codiert dabei das Vorhandensein des Markers  $M_{6-s}$  in der Kombination. Somit werden die Markerkombinationen eineindeutig auf die Zahlen 1 bis 31 abgebildet. In Abb. 4.1 sind alle zu einer Kombination gehörenden Marker grau dargestellt.

Wert	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
M1																															
M2																															
M3																															
M4																															
M5																															

Abbildung 4.1: Binäre Codierung der Markerkombinationen

Zum Schluss werden die Teststatistiken für alle Markerkombinationen aufgelistet. Das ist für die Option `zhao` die TDT-Summen-Statistik lt. (4.16) und für `zhaomax` die Maximum-TDT-Statistik nach (4.17). Schließlich wird der aus der maximalen Statistik durch Permutation bestimmte overall-p-Wert ausgegeben.

## 4.3 Hotel\_Fam

### 4.3.1 Grundlagen

#### Modellannahmen

Seien in der Region eines Risikolocus  $J$  Marker oder Haplotyp-Blocks  $H_j$  gegeben. (Ein Haplotyp-Block fungiert dabei wie ein hochpolymorpher Marker.) Ihre  $n_j$  Allele sind mit  $H_{j1}, \dots, H_{jn_j}$  bezeichnet. Die Stichprobe bestehe aus  $N$  Fällen und ihren Eltern (Trios). Weiterhin bezeichnen  $X_i^{(O)}, Y_i^{(m)}, Y_i^{(f)}$  Diployp-Code-Vektoren (s. u.) des kranken Nachkommen, der Mutter und des Vaters der  $i$ . Familie. Weiterhin sei vorausgesetzt:

1. HWE bezüglich der Blocks,
2. *Random Mating* in der Population [57] und
3. hinreichender Stichprobenumfang für asymptotische  $\chi^2$ -Verteilung der Testgröße

#### Verfahren

Der Test basiert auf der Differenz mittlerer Code-Vektoren der Eltern und Nachkommen: Der mittlere Code-Vektor der Eltern in Familie  $i$  ergibt sich zu:

$$Y_i^{(P)} = [Y_i^{(f)} + Y_i^{(m)}]/2. \quad (4.18)$$

Die Mittelwerte aller Code-Vektoren der Nachkommen bzw. ihrer Eltern sind dann:

$$\bar{X}^{(O)} = \sum_{i=1}^N X_i^{(O)} / N \quad (4.19)$$

bzw.

$$\bar{Y}^{(P)} = \sum_{i=1}^N Y_i^{(P)} / N. \quad (4.20)$$

Analoge Mittelwerte  $\bar{Y}^{(m)}$  bzw.  $\bar{Y}^{(f)}$  können auch separat für Mütter bzw. Väter berechnet werden. Mit der Varianzmatrix

$$\begin{aligned} S_P = \frac{1}{N-1} & \left[ \sum_{i=1}^N \left[ \left( X_i^{(O)} - \bar{X}^{(O)} \right) - \left( X_i^{(O)} - \bar{X}^{(O)} \right) \right]^\tau \right. \\ & - \sum_{i=1}^N \left[ \left( X_i^{(O)} - \bar{X}^{(O)} \right) - \left( Y_i^{(P)} - \bar{Y}^{(P)} \right) \right]^\tau \\ & - \sum_{i=1}^N \left[ \left( Y_i^{(P)} - \bar{Y}^{(P)} \right) - \left( X_i^{(O)} - \bar{X}^{(O)} \right) \right]^\tau \\ & \left. + \sum_{i=1}^N \left[ \left( Y_i^{(P)} - \bar{Y}^{(P)} \right) - \left( Y_i^{(P)} - \bar{Y}^{(P)} \right) \right]^\tau \right] \end{aligned} \quad (4.21)$$

kann man dann eine gepaarte Hotelling's  $T^2$ -Statistik definieren:

$$T_P^2 = N(\bar{X}^{(O)} - \bar{Y}^{(P)})^\tau S_P^{-1} (\bar{X}^{(O)} - \bar{Y}^{(P)}) \quad (4.22)$$

Unter  $H_0^A$  und bei Annahme eines hinreichend großen Stichprobenumfangs ist  $T_P^2$  asymptotisch  $\chi^2$ -verteilt. Die Anzahl der Freiheitsgrade ergibt sich je nach Codierung (s. u.). Unter der Alternativhypothese 'Assoziation' ist  $T_P^2$  asymptotisch nichtzentral  $\chi^2$ -verteilt. Weiterhin kann der p-Wert der Statistik durch Monte-Carlo-Simulation ermittelt werden (s. u.). Momentan werden Familien mit fehlenden Genotypen nicht zur Analyse herangezogen. [57].

### 4.3.2 Optionen

Hotel\_Fam testet nur die Nullhypothese  $H_0^A$ .

Gemäß einer Arbeit von D. Schaid [131] werden in Hotel\_Fam sowohl die 'Genotyp-Codierung' als auch die 'Haplotyp- (oder Allel-)Codierung' verwendet, die hier kurz charakterisiert werden: Für beispielsweise  $n_j = 3$  Haplotyp-Blöcke (Allele) an einem Locus gibt es genau  $n_j(n_j+1)/2 (= 6)$  verschiedene paarweise Kombinationen (Genotypen), deren Codierung in Tab. 4.1 dargestellt ist:

Haplotyp	Genotype Coding	Haplotype Coding
$H_1H_1$	10000	20
$H_2H_2$	01000	02
$H_3H_3$	00000 <sup>a</sup>	00 <sup>b</sup>
$H_1H_2$	00100	11
$H_1H_3$	00010	10 <sup>c</sup>
$H_2H_3$	00001	01

Tabelle 4.1: Genotyp- und Haplotyp-Codierung

<sup>a</sup> alles '0' zeigt das Paar  $H_3H_3$  an

<sup>b</sup> Anzahl des 3. Haplotypblocks: 2-0=2, also  $H_3H_3$

<sup>c</sup> '1' steht für  $H_1$ , 2-1=1-mal  $H_3$ , also  $H_1H_3$

- Bei der Genotyp-Codierung steht jede Stelle für einen Genotyp im  $(n_j - 1)$ -dimensionalen Codevektor. Entweder zeigt (genau) eine '1' das Vorhandensein eines bestimmten Pairs an

bzw. der Vektor aus lauter Nullen steht für das letzte mögliche Paar (hier:  $H_3H_3$ ). Bei  $J$  Markern (Haplotyp-Blocks) ergibt das die Vektorlänge  $\sum_{j=1}^{n_j} n_j(n_j + 1)/2 - J$ .

- Bei der Haplotyp- (oder Allel-)Codierung steht jede Stelle für einen bestimmten Haplotyp-Block / ein Allel. D. h. die Vektorkomponenten enthalten die Anzahlen des entsprechenden Allels. Da der Mensch diploid ist, addieren sich diese zu 2 (CNV seien hier ausdrücklich ausgeklammert). Damit kann die  $n_j$ . (hier: 3.) Stelle wegen Redundanz wegfallen, weil sich diese Zahl als Differenz aus zwei und der bisherigen Summe ergibt. Die Dimension des Codevektors bei  $J$  Markern:  $\sum_{j=1}^J n_j - J$  ist deutlich geringer als die bei Genotyp-Codierung.

Zwei Eingabeparameter betreffen die Anzahl der zu testenden Marker und ihre Namen. Ein Test aller Markerkombinationen muss durch ein Rahmenprogramm realisiert werden.

Entsprechend der Codierung berechnet `Hotel_Fam` zwei verschiedene Teststatistiken und gibt je zwei p-Werte aus. Dabei bestimmt die Dimension des Codevektors die Anzahl der Freiheitsgrade der  $\chi^2$ -Verteilung der Testgröße. Standardmäßig berechnet `Hotel_Fam` die asymptotischen p-Werte des Tests. Durch den Parameter `perm_number` kann man die Replikationszahl für die Bestimmung des p-Werts durch Permutation festlegen. Dabei erhält in einer Replikation jeweils mit Wahrscheinlichkeit 1/2 der Nachkomme den beobachteten bzw. nicht übertragenen Haplotyp-Block. (Die Eingabe `perm_number=0` unterdrückt das Permutationsverfahren.)

Zunächst ist Hotelling's  $T^2$  nur für Trios definiert, aber es lässt sich leicht verallgemeinern: Gibt es in Familie  $i$   $k_i (> 1)$  kranke Geschwister, wird ein mittlerer Codevektor durch  $X_i^{(O)} = \sum_{j=1}^{k_i} X_{ij}^{(O)} / k_i$  definiert.

### 4.3.3 Ausgabe von `Hotel_Fam`

Die Ausgabedatei des Programmes enthält Testergebnisse, getrennt für mütterliche bzw. väterliche Allele (bzw. Haplotype-Blocks) und für beide Eltern zusammen. D. h. der Test ermöglicht Aussagen über Unterschiede zwischen maternaler bzw. paternaler Transmission. Es werden die Statistiken, die Freiheitsgrade und p-Werte ausgegeben. Die p-Werte auf Permutationsbasis gibt es nur für beide Eltern zusammen.

## 4.4 Unphased

### 4.4.1 Modellannahmen

Gegeben seien Kernfamilien mit den Genotypen  $M$ ,  $F$  und  $\mathbf{C} = (C_1, \dots, C_k)$  über mehrere Loci von Mutter, Vater und  $k$  Geschwistern. Weiterhin bezeichne  $\mathbf{Y}$  den Merkmals-Vektor der Nachkommen und  $\mathbf{Z}$  einen Kovariablenvektor. Kleinbuchstabige Variablennamen stehen für die Realisierungen dieser Zufallsgrößen. Dabei wird im allgemeinen Modell Folgendes vorausgesetzt:

1. Rekrutierung der Probandenfamilien allein auf Basis des kindlichen Krankheitsmerkmals,
2. HWE und *Random Mating* in der Population [50],
3. Krankheitsmerkmal der Kinder unabhängig vom elterlichen Genotyp,
4.  $\mathbf{Y}$  multivariat normalverteilt mit gemeinsamem Erwartungswert  $\mu$  und Varianz  $\sigma^2$ . Die Krankheitsmerkmale der Geschwister seien unkorreliert.

### 4.4.2 Verfahren

Bei der Ableitung des Algorithmus geht F. Dudbridge von der bedingten Wahrscheinlichkeit der Genotypen aus, bedingt auf Krankheitsmerkmal (und hier nicht dargestellt: Kovariablen). Diese

retrospektive likelihood kann dabei in eine elterliche und eine bedingte Komponente faktorisiert werden. Die Darstellung folgt dabei dem Manual, Version 3.0.6 [50]:

$$P(F = f, M = m, \mathbf{C} = \mathbf{c} | \mathbf{Y} = \mathbf{y}) = P(f, m | \mathbf{y}) \cdot P(\mathbf{c} | f, m, \mathbf{y}) \quad (4.23)$$

In einem logistischen Regressionsmodells werden additive Genotyp-Effekte  $\beta$  und die multinomiale Verteilung der Genotypen  $\alpha$  parametrisiert.  $\beta[\mathbf{c}]$  sei der Vektor  $(\beta_{c_1}, \dots, \beta_{c_k})$  und  $S(f^*, m^*)$  die Menge möglicher kindlicher Genotypen von Eltern mit den Genotypen  $f^*$  bzw.  $m^*$ :

$$P(f, m) = \frac{\exp \left\{ \frac{1}{\sigma^2} (\mathbf{y}'\beta[\mathbf{c}] - \mu\beta[\mathbf{c}] - \frac{1}{2}\beta[\mathbf{c}]\beta[\mathbf{c}]) \right\}}{\sum_{\mathbf{c}^* \in S(f, m)} \exp \left\{ \frac{1}{\sigma^2} (\mathbf{y}'\beta[\mathbf{c}^*] - \mu\beta[\mathbf{c}^*] - \frac{1}{2}\beta[\mathbf{c}^*]\beta[\mathbf{c}^*]) \right\}} \cdot \frac{\sum_{\mathbf{c}^* \in S(f, m)} \exp \left\{ \alpha_f + \alpha_m + \frac{1}{\sigma^2} (\mathbf{y}'\beta[\mathbf{c}] - \mu\beta[\mathbf{c}] - \frac{1}{2}\beta[\mathbf{c}]\beta[\mathbf{c}]) \right\}}{\sum_{f^*, m^*} \sum_{\mathbf{c}^* \in S(f, m)} \exp \left\{ \alpha_f + \alpha_m + \frac{1}{\sigma^2} (\mathbf{y}'\beta[\mathbf{c}] - \mu\beta[\mathbf{c}] - \frac{1}{2}\beta[\mathbf{c}]\beta[\mathbf{c}]) \right\}} \quad (4.24)$$

Die Innovation von Unphased ist, die Genotyp-Effekte in den beiden Termen  $\beta$  bzw.  $\tilde{\beta}$  separat zu modellieren [50].

$$P(f, m) = \frac{\exp \left\{ \frac{1}{\sigma^2} (\mathbf{y}'\beta[\mathbf{c}] - \mu\beta[\mathbf{c}] - \frac{1}{2}\beta[\mathbf{c}]\beta[\mathbf{c}]) \right\}}{\sum_{\mathbf{c}^* \in S(f, m)} \exp \left\{ \frac{1}{\sigma^2} (\mathbf{y}'\beta[\mathbf{c}^*] - \mu\beta[\mathbf{c}^*] - \frac{1}{2}\beta[\mathbf{c}^*]\beta[\mathbf{c}^*]) \right\}} \cdot \frac{\sum_{\mathbf{c}^* \in S(f, m)} \exp \left\{ \alpha_f + \alpha_m + \frac{1}{\sigma^2} (\mathbf{y}'\tilde{\beta}[\mathbf{c}] - \mu\tilde{\beta}[\mathbf{c}] - \frac{1}{2}\tilde{\beta}[\mathbf{c}]\tilde{\beta}[\mathbf{c}]) \right\}}{\sum_{f^*, m^*} \sum_{\mathbf{c}^* \in S(f, m)} \exp \left\{ \alpha_f + \alpha_m + \frac{1}{\sigma^2} (\mathbf{y}'\tilde{\beta}[\mathbf{c}] - \mu\tilde{\beta}[\mathbf{c}] - \frac{1}{2}\tilde{\beta}[\mathbf{c}]\tilde{\beta}[\mathbf{c}]) \right\}} \quad (4.25)$$

Das ist notwendig, um mit fehlenden Genotypen oder unbestimmter Phase umzugehen, hat aber noch weitere Vorteile:

- Die  $\tilde{\beta}$  modellieren elterliche Genotyp-Effekte.
- Weiterhin ist der allgemeinere Ansatz auch angepasst für eine bedingte Analyse oder die Untersuchung von Gen-Gen-Wechselwirkungen [50].

Im Standardmodell wird  $\tilde{\beta} = \mathbf{0}$  und  $\sigma^2 = 1$  gesetzt. Codiert  $y$  den Krankheitsstatus, so ist  $y_i = 1$  ( $i = 1, \dots, k$ ) für kranke Kinder und 0 für gesunde. Der Beitrag der  $i$ . Familie zum Score des Genotyps  $g$  ergibt sich aus der partiellen Ableitung der log-Likelihood:

$$U_g^{(i)} = \left( \frac{\partial \log L^{(i)}}{\partial \beta_g}, \frac{\partial \log L^{(i)}}{\partial \alpha_g} \right)^T \quad (4.26)$$

Dann kann man die Kovarianzmatrix des Total-Scores  $\mathbf{U} = \sum_i \mathbf{U}^{(i)}$  durch  $\mathbf{V} = \sum_i \mathbf{U}^{(i)T} \mathbf{U}^{(i)}$  schätzen.

Seien  $\mathbf{U}_\beta$  z. B. der zum Parameter  $\beta$  gehörende Teilvektor von  $\mathbf{U}$  und  $\mathbf{V}_{\beta\beta}$  die dazugehörige Teilmatrix von  $\mathbf{V}$ . Seien weiterhin  $\mathbf{V}_{\alpha\alpha}$ ,  $\mathbf{V}_{\alpha\beta}$  und  $\mathbf{V}_{\beta\alpha}$  die  $\alpha$  betreffenden Teilmatrizen von  $\mathbf{V}$ . Dann ist

$$\mathbf{V}_\beta = \mathbf{V}_{\beta\beta} - \mathbf{V}_{\beta\alpha} \mathbf{V}_{\alpha\alpha}^{-1} \mathbf{V}_{\alpha\beta} \quad (4.27)$$

die Kovarianzmatrix von  $\mathbf{U}_\beta$  und

$$C_g^2 = (\mathbf{U}_\beta)_g (\mathbf{V}_\beta)_{gg}^{-1} (\mathbf{U}_\beta)_g^T \quad (4.28)$$

definiert die asymptotisch  $\chi^2$ -verteilte Score-Statistik des Genotyps  $g$ . Weiterhin kann man den Maximum-Likelihood-Schätzwert  $\hat{\beta}_g$  zur Konstruktion des Konfidenzintervalls verwenden:

$$\hat{\beta}_g \pm 1.96 (\mathbf{V}_\beta)_{gg}^{1/2} \quad (4.29)$$

Zur vereinfachten Darstellung bezogen sich die letzten Formeln auf Genotypen. Der Hauptfokus von Unphased liegt aber auf Haplotypen.

### 4.4.3 Test auf Kopplung

Ähnlich wie in 4.1.5 wird durch die Option `-nolinkage` auf Kopplung getestet:  $H_0^L$ . Das resultiert einerseits in einer schnelleren Analyse und höherer Power. Andererseits muss man genau prüfen, ob die Voraussetzungen dieses Tests erfüllt sind.

### 4.4.4 Optionen

- Mit den Optionen `-allwindows` `-allcombinations` werden alle Markerkombinationen getestet. Leider konnten nur die asymptotischen Verfahren in die Studie integriert werden, da Vortests für das Permutationsverfahren unpraktikable Rechenzeiten ergaben. So benötigte Unphased für 16 Replikationen mit Samples von 300 AA-Familien bei nur 4000 Permutationen eine Rechenzeit von 97 Tagen!
- Da das Modell (4.25) allgemein formuliert ist, eröffnen sich eine Reihe von Möglichkeiten: Als Standardmodell wird ein volles Modell mit je einem  $\log(\text{OR})$ -Parameter für jeden Haplotyp angepasst. Das entspricht einem Aufruf mit der Option `-model full`. In der Tabelle unten sind die Haplotyp- sowie Allel-Codierung am Beispiel zweier diallelischer Marker dargestellt. Das volle Modell benötigt vier Parameter:

Haplotype	Haplotype Coding	Allele Coding	Allele Main Effects
1-1	$\beta_{11}$	$\beta_0$	$\beta_0$
1-2	$\beta_{12}$	$\beta_0 + \beta_{-2}$	$\beta_0 + \beta_{-2}$
2-1	$\beta_{21}$	$\beta_0 + \beta_{2-}$	$\beta_0 + \beta_{2-}$
2-2	$\beta_{22}$	$\beta_0 + \beta_{-2} + \beta_{2-} + \beta_{22}$	$\beta_0 + \beta_{-2} + \beta_{2-}$

- Bei der Option `-model allelemain` ergeben sich die OR der Haplotypen aus multiplikativen Beiträgen der einzelnen Marker. In der Tabelle oben sieht man, dass dann nur noch drei Parameter anstelle von vier nötig sind. Der daraus resultierende Test hat weniger Freiheitsgrade, vermag aber keine reinen Haplotyp-Effekte aufzudecken.
- Auf die Modellierung parentaler Genotyp-Risiken, vgl. (4.25) durch die Option `-parentrisk` konnte in der Simulation nicht eingegangen werden.
- Mehrere Optionen betreffen die Behandlung fehlender Genotypen, seltener bzw. unbestimmter Haplotypen: Bei der Option `-missing` wird die maximale Information (alle Personen mit Genotypen) einbezogen und über alle konsistenten Ergänzungen der Daten gemittelt. Das schnellere, aber weniger effektive Verfahren `-certain` nutzt nur Familien mit sicherer Phasenauflösung (d. h. höchstens ein heterozygoter Locus). Es kam hier nicht zum Einsatz.
- Haplotypen mit einer Häufigkeit unter der Schranke `-rare #r` werden gepoolt. D. h. bei der Parametrisierung wird für sie einunddasselbe Risiko (OR) vorausgesetzt. Analog zu FBAT (4.1.6) wurde die Studie mit der Einstellung `-rare 0.05` durchgeführt. Diese Option sollte nach Aussage des Handbuchs nicht mit der Option `-model allelemain` zusammen aufgerufen werden.

Als Häufigkeitsschranke für seltene Haplotypen (`-zero`), unter der sie ignoriert werden, wurde die Voreinstellung ( $10^{-8}$ ) gelassen.

### 4.4.5 Ausgabe von Unphased

Zunächst wiederholt Unphased zur Kontrolle alle eingegebenen Optionen und die Anzahl Kernfamilien.

Für jede Markerkombination werden danach die log-Likelihood unter  $H_0$  und  $H_1$ , die Statistik des Likelihood-Quotiententests, die Freiheitsgrade und der p-Wert ausgegeben. (Im Gegensatz zum Handbuch [50] werden Ergebnisse von Likelihood-Quotiententests und nicht von Score-Tests ausgegeben.) Weiterhin erhält man die Allel- bzw. Haplotyp-Frequenzen für Fälle und 'Kontrollen' sowie das Odds-Ratio mit 95%-Konfidenzgrenzen.

Dabei sind die Markerkombinationen nach der Anzahl der Marker sortiert, d. h. hier: Die Ausgabe beginnt mit den Einzelmarkern  $M_1, \dots, M_5$ , setzt mit Zweierkombinationen  $M_1M_2, M_1M_3, \dots, M_4M_5$  fort und schließt mit der Kombination aller fünf Marker.

## 4.5 Übersicht zur Ausgabe der vier Programme

In der folgenden Tabelle sind die ausgegebenen Werte der Verfahren noch einmal zusammengefasst:

Programm	Verfahren	Marker-spezifisch	Haplotyp-spezifisch	Mk-spezifisch 'global'	Über alle Mk 'overall'
FBAT	multi marker test	x <sup>a</sup>		x	
	haplotype test		x	x	
	– “ – permutation test		x	x	
Famhap	zhao		(x) <sup>b</sup>	x	x
	zhaomax		(x)	x	x
Hotel_Fam	genotype coding			x	
	haplotype coding			x	
Unphased	model full		(x) <sup>b</sup>	x	(x) <sup>c</sup>
	model allelemain		(x)	x	(x)

Tabelle 4.2: Ausgabe der vier Programme

<sup>a</sup> D. h. insbesondere, dass die einen Marker betreffenden Statistiken für verschiedene Mk, die diesen Marker enthalten, identisch sind.

<sup>b</sup> Es werden nur Anzahlen übertragener / nicht übertragener Haplotypen ausgegeben, die TDT-Statistik kann berechnet werden.

<sup>c</sup> Permutationstest konnten aus Gründen der Laufzeit hier nicht angewandt werden.

# Kapitel 5

## Realisierung und Kontrolle der Simulation

### 5.1 Ablauf

Um das geplante Ziel zu erreichen, ergaben sich ganz natürlich die folgenden Schritte: Für jedes Programm und jedes Szenarium sollten 1000 Replikationen aus jeweils diesen drei Bestandteilen durchgeführt werden:

- Simulation von Stichproben (*pedigree files* im pre-linkage-Format) gemäß den Modellvorgaben (s. 2.2)
- Durchführung der Assoziationsanalyse mit dem Programm und
- Auslesen relevanter Daten aus den Ausgabe-Dateien und Speichern in Zusammenfassungs-Dateien.

Weiterhin erschien es sinnvoll, diese drei Bestandteile in ein SAS-Rahmenprogramm einzubetten, das die Makros bzw. Skripte aufruft und kontrolliert.

Sowohl für die Modelle unter der Null- als auch unter der Alternativhypothese ergibt sich jedes Szenarium als Kombination je dreier Faktoren (vgl. 2.2.2, 2.2.3):

$H_0$ : Stichprobenumfänge (200, 1000, 2500), Modelle (NMod 1, NMod 2, NMod 3) sowie Familienstrukturen (F10, F20, F22),

$H_1$ : Vererbungsmodi (MUL, REZ, DOM, ADD), Krankheitsmodelle (KMod 1, KMod 2, KMod 3) sowie Familienstrukturen (F10, F20, F22).

Deshalb lag es nahe, die Algorithmen in vier ineinandergeschachtelten Schleifen ablaufen zu lassen (siehe Abb. 5.1).

Laufzeit-Untersuchungen mit je einer Replikation pro Szenarium zeigten allerdings, dass die benötigte Rechenzeit zwischen den einzelnen Verfahren und Szenarien gewaltig differiert (siehe 5.2.5). Um weiterhin die Kapazität des Cluster-Rechners mit 16 Knoten, auf denen (theoretisch) je zwei Jobs laufen können, optimal auszunutzen, lag eine Aufteilung der Gesamtarbeit in überschaubare Einzelpakete nahe. Das konnte z. B. einfach durch die Fixierung der Laufindizes geschehen: z. B. für  $H_1$ : H=2 (rezessiver Modus), M=3 (Modell KMod 3), F=1 (Trios).

### 5.2 Probeläufe

Um die Eigenschaften der Programme kennenzulernen, wurden Testläufe mit verschiedenen Familiendaten durchgeführt. Außerdem musste die Struktur der Programmausgabe bekannt sein, um diese automatisiert auslesen zu können.

Bei den Testläufen wurden folgende Besonderheiten der Programme festgestellt:

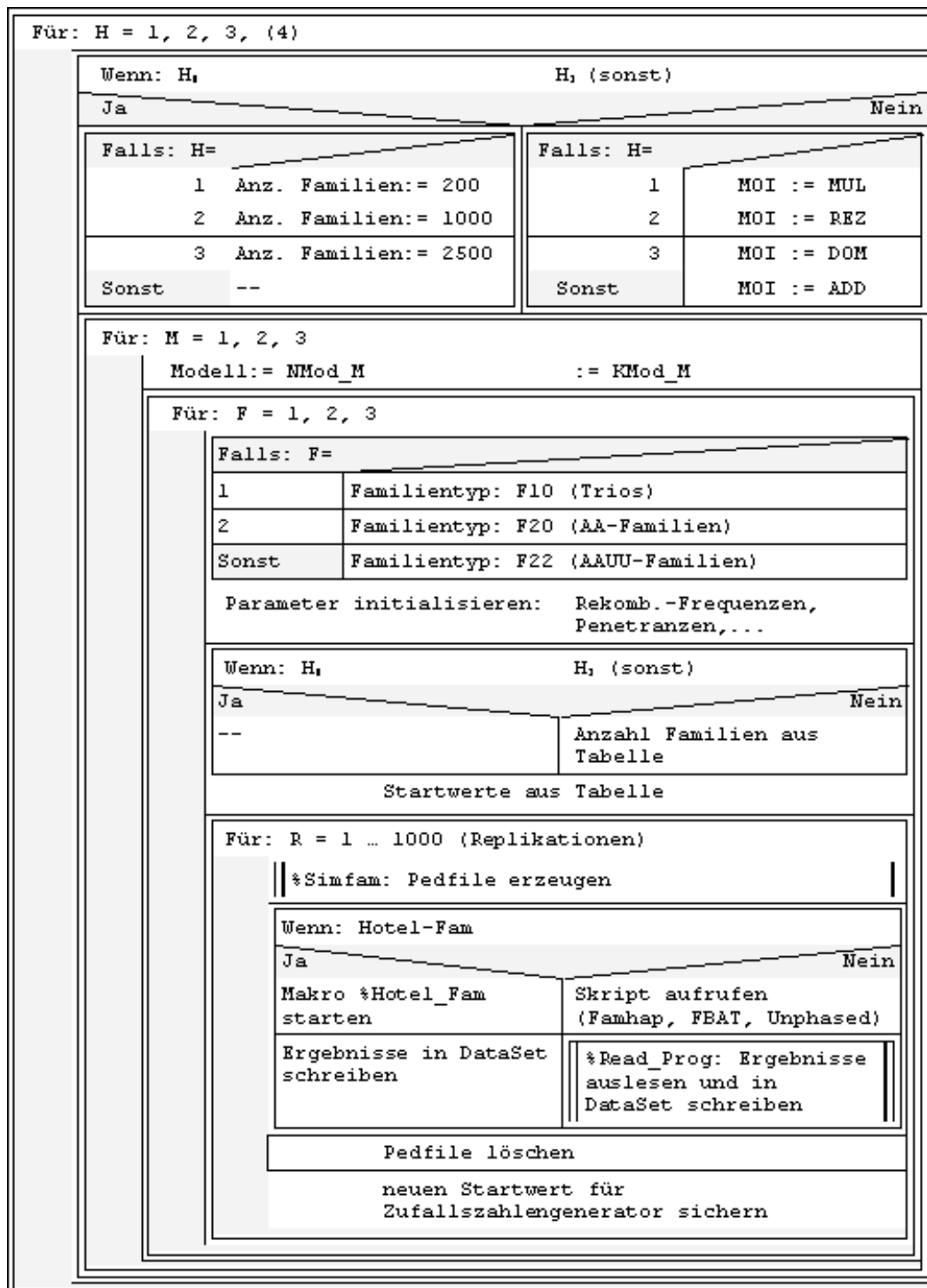


Abbildung 5.1: Ursprünglicher Ablauf der Simulation



### 5.2.1 FBAT

Nach dem Aufruf von FBAT (hier: Version 1.7.2, April 2006) wird eine Oberfläche (Shell) geöffnet, auf der FBAT-Kommandos eingegeben werden können. Für den automatisierten Ablauf muss die Shell umgangen werden, indem die Befehle in einem Skript dem FBAT-Aufruf übergeben werden.

FBAT sollte in fünf verschiedenen Varianten aufgerufen werden: mit dem Multimarker-Test für  $H_0^L$  und  $H_0^A$ , dem Permutationstest und dem Haplotypstest, wiederum für beide Nullhypothesen.

Weil FBAT nicht die Möglichkeit enthält, nacheinander alle Markerkombinationen aus den vorgegebenen Markern zu testen, musste das durch ein spezielles Befehlsskript erreicht werden. Den Empfehlungen im Handbuch [95]: minimale Haplotyphäufigkeit 5% in Verbindung mit mindestens 10 informativen Familien (d. h. also die Optionen `minfreq=0.05 minsize=10`) wurde gefolgt.

In den ersten Versuchen stellte sich heraus, dass FBAT bei Stichproben mit mehreren Tausend Familien immensen Speicherplatzbedarf hat. (Der `top`-Befehl von Linux zeigt die laufenden Prozesse eines Knotens mit ihrem Ressourcenbedarf an. Auf diese Weise war nachweisbar, dass mehrere FBAT-Analysen bei einem Speicherverbrauch von  $> 2,8$  GB abstürzten.) Um diesem Problem zu begegnen, wurde FBAT mit großen Stichproben separat auf einem Knoten gestartet, damit der gesamte Hauptspeicher dieses Knotens diesem Job zur Verfügung stand. Weiterhin wurde das FBAT-Skript in vier Skripte aufgesplittet, mit denen das Programm jedes Mal neu aufgerufen wurde.

Damit wurde das Speicherproblem tatsächlich behoben. Das lässt darauf schließen, dass FBAT nach Beenden einer Analyse den Speicher nicht komplett wieder frei gibt, sondern nur, wenn es mit `quit` komplett beendet wird.

#### FBAT-Skript

Der Test aller Markerkombinationen musste durch ein Skript nachempfunden werden. Dabei bezeichnen `snpaa`, ..., `snpae` die fünf Marker. FBAT wurde mit folgendem Eingabeskript gestartet:

```
fbat < befehlsskript.script > /dev/null
# die Ausgabe erfolgte in ein Logfile, deshalb hier keine Ausgabe
```

Das Eingabeskript hatte diese Struktur:

```
log /pfad1/fbat.out      # Ausgabe ins Logfile festlegen
minsize 10              # minimal 10 Familien
minfreq 0.05           # minimale Haplotypfrequenz 5%
load /pfad2/fbat.ped    # Eingabe-Datei laden
mode b                  # biallelische Testprozedur
                        # 1) Multimarker-Test, HO_L

fbat -m snpae
# ...
# danach Test aller Markerkombinationen bis
# ...
fbat -m snpaa snpab snpac snpad snpae
                        # 2) Multimarker-Test, HO_A

fbat -e -m snpae
# ...
fbat -e -m snpaa snpab snpac snpad snpae
                        # 3) Permutations-Test mit 4000 Wiederholungen, HO_L

hbat -p4000 snpae
# ...
hbat -p4000 snpaa snpab snpac snpad snpae

mode a                  # mehrallelische Testprozedur
hbat snpae              # 4) Haplotyp-Test, HO_L
```

```

# ...
hbat  snpaa snpab snpac snpad snpae
      # 5) Haplotyp-Test, HO_A

hbat  -e snpae
# ...
hbat  -e snpaa snpab snpac snpad snpae
quit      # Ende FBAT

```

## 5.2.2 Famhap

Da im voraus meist unklar ist, *welche* Markerkombination beim Test die größte Assoziation ergibt, ist es in Standard-Situationen sinnvoll, alle Markerkombinationen zu testen. Famhap enthält zu diesem Zweck die Option `allcombi`, die bei  $m$  Markern den Test aller  $2^m - 1$  möglichen Markerkombinationen zur Folge hat. Weiterhin liefert Famhap in diesem Fall einen totalen p-Wert über alle Mk, der hinsichtlich des multiplen Testens korrigiert wurde.

Wird die Option `linkage` gesetzt, testet Famhap die Hypothese  $H_0^L$ , sonst standardmäßig die Hypothese  $H_0^A$ .

Als Algorithmen empfehlen sich die auf Haplotypen erweiterte Methode nach Zhao [93] (Option `zhao`) bzw. mit der Option `zhaomax` die Statistik auf Basis der maximalen Chiquadrat-Statistik der einzelnen Haplotypen.

### Famhap-Skript

Aus oben Genanntem ergab sich für die Analyse mit Famhap17 folgendes Skript:

```

famhap17 /inpfad/famh.ped /outpfad/famh.out zhao allcombi SIMULATIONS 4000
famhap17 /inpfad/famh.ped /outpfad/famh.out zhao allcombi
      SIMULATIONS 4000 linkage
famhap17 /inpfad/famh.ped /outpfad/famh.out zhaomax allcombi
      SIMULATIONS 4000
famhap17 /inpfad/famh.ped /outpfad/famh.out zhaomax allcombi
      SIMULATIONS 4000 linkage

# zhao, zhaomax      - alternative Verfahren
# allcombi           - alle Markerkombinationen testen
# SIMULATIONS 4000  - 4000 Permutationen für den totalen p-Wert
# linkage            - Test auf Kopplung

```

Die Analyse mit fehlenden Genotypen muss mit Option `u` gestartet werden. Sie bewirkt, dass auch Familien in die Haplotyp-Frequenzschätzung einbezogen werden, für die der Genotyp nicht für jeden vorliegt. In den Test werden allerdings nur die Familien einbezogen, für die zumindest ein Trio mit kompletter Genotyp-Information vorhanden ist.

## 5.2.3 Hotel-Fam

Da `Hotel_Fam` ein SAS[4]-IML-Makro ist, kann man sich in diesem Fall komplett auf der SAS-Ebene bewegen, d. h. die Stichproben wurden mit dem Makro `%SimFam` als Datasets erzeugt, temporär abgelegt und sofort mit dem (leicht modifizierten) Makro `%Hotel_Fam` (Version 0.2 vom 15.03.2005) analysiert. Anstelle einer Text-Ausgabe, die hätte extrahiert werden müssen, wurden die charakteristischen Parameter (wie z. B. die p-Werte des asymptotischen bzw. Permutationstests bei Genotyp- und Haplotypcodierung) direkt als eine Beobachtung in ein Dataset geschrieben.

Auch in `Hotel_Fam` gibt es keine Option 'alle Markerkombinationen'. Aber das konnte leicht dadurch emuliert werden, indem der Aufruf des Makros in 5 ineinandergeschachtelte Schleifen eingebettet wurde, die nacheinander alle 31 möglichen Markerkombinationen erzeugen. Einen overall p-Wert erhält man dann durch Bonferroni-Korrektur, wie sie in 6.3.4 näher erläutert wird.

## Makro-Aufruf

Damit ergab sich für jede Markerkombination folgende Makro-Befehlszeile:

```
%Hotel_Fam(Infile, MAnz, MKombi, Outfile, PermNumber);  
/* Parameter:  
    Infile      - Eingabedatei  
    Outfile     - Ausgabe-Dataset  
    MAnz        - Markeranzahl in der Kombination  
    MKombi      - Markernamen in der Kombination  
    PermNumber  - Permutationszahl, hier 4000  
*/
```

## 5.2.4 Unphased

Unphased benötigt schon mit relativ kleinen Stichproben lange Laufzeiten. So musste leider auf den permutationsbasierten Test komplett verzichtet werden (siehe 4.4.4).

Um Rechenzeit zu sparen, empfiehlt das Handbuch eine minimale Haplotypfrequenz mit der Option `-rare` zu setzen. D. h. für die Risikoschätzung werden alle selteneren Haplotypen gepoolt. Allgemeine Empfehlungen zur Wahl dieses Parameters sind sehr schwer zu geben (persönliche Kommunikation mit dem Autor, F. Dudbridge). Deshalb wurde wie bei FBAT diese Häufigkeit `-rare 0.05` gesetzt.

Der Befehl `unphased` wird standardmäßig als Test auf Assoziation (d. h. Hypothese  $H_0^A$ ) durchgeführt, als Test auf Kopplung ( $H_0^L$ ) mit der Option `-nolinkage`.

Alle Markerkombinationen lassen sich durch die Kombination der Optionen `-allwindows` `-allcombinations` testen. Der overall p-Wert wurde aus Rechenzeitgründen nicht durch Permutation, sondern aus dem minimalen p-Wert durch Bonferroni-Korrektur ermittelt.

Bei Analysen mit diesem Programm gab es regelmäßig die Fehlermeldung

```
[Prozessnummer] Speicherzugriffsfehler  
[unphased-Befehlsaufruf]
```

die auch nicht durch Verwenden einer neuen Version beseitigt werden konnte. Nach persönlicher Kommunikation mit dem Autor scheint der Fehler von Komponenten der Linux-Distribution abzuhängen (und damit auf anderen Computern kaum reproduzierbar zu sein). Da sich nachweisen ließ, dass die Fehlermeldung erst *nach* korrektem Abschluss der Analyse und Datenausgabe auftritt, lief Unphased unter diesen Bedingungen und lieferte plausible Ergebnisse.

## Unphased-Script

Unphased (Version 3.0, August 2006 - 2007) wurde mit folgendem Skript gestartet:

```
unphased.lnx infile.ped -allw -allc -rare 0.05 > outfile.out  
unphased.lnx infile.ped -allw -allc -rare 0.05 -nolinkage >> outfile.out  
unphased.lnx infile.ped -allw -allc -model allelemain >> outfile.out  
unphased.lnx infile.ped -allw -allc -model allelemain -nolinkage >> outfile.out  
# -allw -allc - alle Markerkombinationen  
# -rare 0.05 - seltene Haplotypen <5% gepoolt  
# -nolinkage - Test auf Kopplung  
# -model allelemain - nur Parametrisierung der Haupteffekte,  
# - sonst standardmäßig volles Modell
```

Für Familien mit fehlenden Genotypen musste die Option `-missing` ergänzt werden, wodurch über alle konsistenten Genotyp-Ergänzungen der Daten gemittelt wird.

### 5.2.5 Abschätzung der Laufzeiten

Die Analysen wurden teilweise parallel auf einem Cluster-Computer durchgeführt. Jeder der 16 Knoten arbeitet mit einem Pentium-IV-Prozessor (3 GHz) und 2 GB RAM. Für besonders zeitraubende Jobs wurde auch ein um einen Faktor  $>3$  schnellerer Computer mit Dual Core Prozessor (AMD Opteron 8218, 2.6 GHz und 4 GB RAM) eingesetzt.

Um die Laufzeiten abzuschätzen, wurde für jedes Programm je eine Replikation jedes Szenariums (siehe 2.2) durchgeführt. Es wurde dabei darauf geachtet, dass jedem Job genau ein Knoten des Clusters zur Verfügung stand. Neben der groben Abschätzung des zu erwartenden Rechenzeitbedarfs ergab sich die Möglichkeit, Jobs möglichst optimal zu verwalten: Einerseits sollten die Cluster-Knoten gut ausgelastet sein. Andererseits war es sinnvoll, sehr lange laufende Jobs in kleinere Pakete zu splitten.

Sowohl zwischen den einzelnen Programmen als auch für die verschiedenen Stichprobenumfänge gab es sehr große Differenzen. Famhap ist bei weitem das schnellste Programm, auch wenn fünf Befehlsaufrufen bei FBAT je vier bei Famhap und Unphased gegenüberstehen. Für die  $H_0$ -Szenarien bewegt sich die Laufzeit zwischen 11 Sekunden (Famhap, 200 AAUU-Familien für Modell NMod 3 (vgl. Tabelle A.3 im Anhang) und fast 9 Stunden (Unphased, 2500 AAUU-Familien, Modell NMod 1). Am Schluss wurde der Faktor auf Famhap bezogener Laufzeiten berechnet:

Famhap	FBAT	Unphased	Hotel_Fam
1.0	4.1	24.1	6.7

Tabelle 5.1: Relative Laufzeiten unter  $H_0$ ,  
ermittelt aus je einer Replikation der 27 Szenarien

Unter der Alternativhypothese ist die Situation ähnlich: Die Laufzeiten differieren einerseits sehr zwischen den Programmen, andererseits auch zwischen Vererbungsmodi, Familientypen und Krankheitsmodellen. Das ist natürlich auch den sehr unterschiedlichen Stichprobenumfängen (siehe 2.5) geschuldet. In Tabelle A.4 im Anhang finden sich die Laufzeiten für je eine Replikation für Daten mit fehlenden Genotypen. So überspannen hier die Werte den Bereich von 25 Sekunden (Famhap, multiplikatives Modell KMod1, AAUU-Familien) bis knapp 3 Stunden (Unphased, rezessives Model KMod3, AAUU-Familien). Die relativen Laufzeiten unterscheiden sich allerdings nicht ganz so stark:

Famhap	FBAT	Unphased	Hotel_Fam
1.0	3.5	15.0	6.9

Tabelle 5.2: Relative Laufzeiten unter  $H_1$ ,  
ermittelt aus je einer Replikation der 36 Szenarien

Für die Analyse mit kompletten Genotypen ergaben sich ähnliche Aussagen.

Rechnet man die Zeiten für je 1000 Replikationen hoch, so ergibt das unter  $H_0$  ca. 3600 Tage und unter  $H_1$  über 1700 Tage Rechenzeit. Schwankungen auf Grund wechselnder Auslastung der Knoten und des Speichers wurden hierbei ignoriert. Trotzdem verdeutlicht diese sehr grobe Abschätzung, dass es auf eine geschickte Organisation der Computer-Jobs ankam, um die Studie in überschaubarer Zeit abschließen zu können.

### 5.2.6 Test der Synchronizität

Alle eingesetzten Verfahren sollten unter möglichst identischen Bedingungen getestet werden. Das schloss insbesondere die Identität des zu analysierenden Datenbestandes ein. Die Identität der in der  $r$ . Replikation analysierten Ausgangsdateien wurde bei deren Erzeugung durch den Startwert des Zufallszahlengenerators gewährleistet. Um das zu prüfen, wurde für ausgewählte Szenarien der Nullhypothese getestet, ob die Verfahren tatsächlich identische Stichproben (*pedigree files*)

analysierten. Zu diesem Zweck wurden je 100 Replikationen unter  $H_0$  mit 200 Trios und dem Modell NMod 1 durch Famhap, FBAT, Hotel\_Fam und Unphased analysiert und verglichen, ob die jeweils 100. Stichprobe übereinstimmt. Ein zweiter Test verglich die Samples bei Analysen von 200-AAUU-Familien zum Modell NMod 3 nach je 50 Replikationen. Die SAS-Vergleichs-Prozedur Proc COMPARE stellte völlige Übereinstimmung der Dateien für die vier Analyseprogramme fest.

Die im Abschnitt 5.4 beschriebene 'Entkopplung' der drei Arbeitsblöcke sicherte auf natürliche Weise, dass alle vier Methoden mit denselben Familiendaten arbeiten: Bei der Analyse mit fehlenden Genotypen und unter Stratifikation griffen alle Programme auf denselben (vorher erzeugten) Datenbestand zu.

## 5.3 Administration der Jobs auf Cluster

### 5.3.1 Start der Rechenjobs

Alle Rechenjobs wurden mit Hilfe eines PBS-Systems auf dem Cluster verwaltet. Durch den Befehl

```
qsub -l nodes=nodeX [startjob.script]
```

wird auf Knoten nodeX ein Skript startjob.script gestartet. (Z. B. rief dieses das entsprechende SAS-Rahmenprogramm auf.) Vom PBS-System erhält jeder Job eine fortlaufende Nummer jobnr und mit

```
qstat -f [jobnr]
```

können wesentliche Parameter dieses Jobs (z. B. Startdatum und -uhrzeit, ausführender Knoten, Status usw.) angezeigt werden. Weiterhin werden zwei Dateien

```
[jobnr].[startjob.scr].er  
[jobnr].[startjob.scr].ou
```

im Homeverzeichnis des Benutzers angelegt, welche die Ausgabe (ou = Output) bzw. die Fehlermeldungen (er = Error) enthalten. Da hier alle Programm-Ausgaben sofort in Dateien umgeleitet wurden, blieben diese Dateien bei ordnungsgemäßem Ablauf leer, was die Jobüberwachung erleichterte. Einzig bei Unphased erschien je Replikation vier Mal die in 5.2.4 beschriebene Fehlermeldung in der ER-Datei. Eine Protokolldatei mit dem Namen

```
[jobnr].[programm_szenarium].log
```

wurde manuell angelegt mit der Statusanzeige des Jobs, Hinweisen auf Fehler und weiteren wichtigen Bemerkungen.

Die Abarbeitung eines Jobs verlief wie folgt:

**Famhap, FBAT und Unphased:** Ein Skript startete das SAS-Rahmenprogramm. Dieses wiederum rief

- den Makro %SimFam zur Erzeugung der Pedfiles auf,
- das Skript auf, welches die Analyse steuerte (s. 5.2.1 - 5.2.4),
- den Makro %ReadProg auf, der die Kenngrößen aus der Programmausgabe extrahierte und sie als je eine Beobachtung an ein Zusammenfassungs-Dataset anhing.

**Hotel\_Fam:** Ein Skript startete das SAS-Rahmenprogramm, welches

- den Makro %SimFam zur Erzeugung des Pedfiles aufrief,
- die Analyse mit dem Makro %Hotel\_Fam für alle 31 Markerkombinationen startete
- und die charakteristischen Werte ins Summary-Dataset schrieb.

Ergebnis einer korrekt gelaufenen Analyse mit 1000 Replikationen war ein Summary-Dataset mit 1000 Beobachtungen und einer Vielzahl von p-Werten sowie anderer Kenngrößen.

### 5.3.2 Wiederaufnahme abgebrochener Jobs

Angenommen, ein Job war nach  $R_1$  erfolgreichen Replikationen abgebrochen. Dann galt es, die korrekten Anfangsbedingungen wiederherzustellen, um den Job an dieser Stelle authentisch fortzusetzen. Da der Startwert des Zufallszahlengenerators die Erzeugung der Pedfiles determinierte, musste also der für die  $R_1 + 1$ . Replikation gültige Startwert erzeugt werden. Dafür wurde ein SAS-Makro %Neustart(r) geschrieben, der - ausgehend vom Anfangsstartwert - nur die entsprechende Anzahl Aufrufe der Prozedur Call RANUNI enthält, ohne die Pedfiles zu erzeugen. (Hinweis: Call RANUNI gibt außer dem [0,1]-gleichverteilten Pseudo-Zufallszahlenwert einen neu kalkulierten Startwert zurück.) Nach Setzen der Replikationsnummer und des neuen Initialwertes im DataSet nquelle konnte das SAS-Programm neu gestartet werden (vgl. Tab. 5.1). Bei einer Reihe von Jobs wurde parallel eine SAS-Liste mit den Startwerten ausgegeben. In diesem Falle konnte überprüft werden, dass die auf verschiedene Weise gewonnenen Startwerte auch wirklich identisch waren.

Für den zweiten Teil der Analyse (Familiendaten mit fehlenden Genotypen) wurden die Arbeitsschritte wie im Abschnitt 5.4.1 beschrieben entkoppelt. So musste für die Neuaufnahme eines abgebrochenen Jobs im Befehlsskript nur der Startindex  $R_1 + 1$  für die nächste Replikation neu gesetzt werden.

### 5.3.3 Abschluss eines Jobs

Ein beendeter Job ist äußerlich daran erkennbar, dass das PBS-System die OU- und ER-Files (siehe 5.3.1) in das Verzeichnis verschoben hat, aus dem der Job gestartet wurde. Datum und Uhrzeit des Jobendes wurden manuell in die o. g. Protokolldatei übernommen sowie Bemerkungen zum erfolgreichen / fehlerhaften Lauf ergänzt. Beim Neustart eines abgebrochenen Jobs wurden die neuen Angaben angehängt und der Dateiname um die neue Jobnummer erweitert:

```
[jobnr1+jobnr2].[programm_szenario].log
```

Bei der Analyse von unvollständigen Daten wurden die (je 1000) Ergebnisdateien durch einen Makro %Read\_Prog ausgelesen und die p-Werte sowie andere Parameter in ein Zusammenfassungs-Dataset mit je 1000 Replikationen geschrieben. Falls dieses Summary auf den ersten Blick korrekte Daten enthielt (1000 Beobachtungen, keine fehlenden oder mehrfach geschriebenen Werte, keine fehlenden Variablen usw.), wurde es gesichert. Nachdem die komprimierten Ergebnisdateien sowie das Logfile gesichert wurden, konnten die Original-Ausgabedateien gelöscht werden.

Die Teststatistiken der verschiedenen Programmoptionen sind noch einmal in Tabelle 5.3 zusammengefasst:

Name	Programm	Befehl	Beschreibung	Testklassen <sup>a</sup>
Assoziationstests				
mm	FBAT	fbat -m -e	Multi-Marker Test	[M, a] <sup>b</sup>
hbat_b		hbat -e	Haplotyp-Test mit Bonferroni-Korrektur	[H, 1, a] <sup>c</sup>
hbat_g		hbat -e	Haplotyp-Test, globaler p-Wert	[H, 2, a]
zhao	Famhap	famhap zhao	Hap-TDT-Statistik	[H, 2, p] <sup>d</sup>
zhaomax		famhap zhaomax	max Hap-TDT	[H, 1, p]
full	Unphased	unphased -model full	volles Modell	[H, 2, a]
all_m		unphased -model allelemain	nur Allel-Haupteffekte	[M, a]
gen_as	Hotel_Fam	–	asymptotische Statistik, Genotypcodierung	[M, a]
gen_pm		–	Permutationsstatistik, Genotypcodierung	[M, p]
hap_as		–	asymptotische Statistik, Haplotypcodierung	[M, a]
hap_pm		–	Permutationsstatistik, Haplotypcodierung	[M, p]
Tests auf Kopplung				
mm_l	FBAT	fbat -m	Multimarker Test auf Kopplung	[M, a]
hbat_bl		hbat	Haplotyptest auf Kopplung, Bonferroni	[H, 1, a]
hbat_gl		hbat	Haplotyptest auf Kopplung, globales p	[H, 2, a]
perm_m		hbat -p#R	Permutationstest, minimal-p-Statistik	[H, 1, p]
perm_s		hbat -p#R	Permutationstest, Chiquadrat-Summenstatistik	[H, 2, p]
zhao_l	Famhap	famhap zhao linkage	erweiterter TDT auf Kopplung	[H, 2, p]
zhaomax_l		famhap zhaomax linkage	maxTDT auf Kopplung	[H, 1, p]
full_l	Unphased	unphased -model full -nolinkage	volles Modell, Test auf Kopplung	[H, 2, a]
allM_l		unphased -model allelemain -nolinkage	Allel-Haupteffekte, Test auf Kopplung	[M, a]

Tabelle 5.3: Liste der 20 getesteten Statistiken, die durch verschiedene Programmbefehle bzw. Optionen ausgewählt werden können

<sup>a</sup> s. Einteilung der Tests 6.6.4<sup>b</sup> [M] Multi-Marker-Test, [a] asymptotisch<sup>c</sup> [H] Haplotyp-Test, [1] verwendet nur 'besten' Haplotypen, [a] asymptotisch<sup>d</sup> [H] Haplotyp-Test, [2] verwendet Information aller Haplotypen, [p] Permutations-basiert

## 5.4 Modifikationen für Familiendaten mit fehlenden Genotypen

### 5.4.1 Veränderungen des Ablaufs

Mitte 2007 ergab sich die Möglichkeit, testweise 2 - 4 Knoten eines schnelleren Computers (AMD Opteron 8218, Dual Core Processor, 2.6 GHz) für die Analysen zu nutzen. Nach ersten Abschätzungen verkürzte dieser die Rechenzeiten auf 1/3 bis 1/5. Da nicht abzusehen war, wie lange und in welchem Umfang diese zusätzliche Ressource zur Verfügung steht, wurden zunächst die Stichproben für das zeitraubendste Szenarium kopiert und die Unphased-Analyse für diese Familiendaten gestartet. Regelmäßig wurden (manuell) die fertigen Ergebnisdateien kopiert und nach Abschluss aller 1000 Replikationen durch ein SAS-Makro `%Read_Unphased` ausgewertet. Weitere freie Kapazitäten wurden für andere zeitraubende Jobs genutzt.

Der erfolgreiche Lauf dieser Szenarien motivierte vor Beginn der Simulation mit unvollständigen Familiendaten dazu, die drei Hauptblöcke: Erzeugung der Pedfiles / Analyse / Auswertung der Ausgabe generell voneinander abzukoppeln. Folgende Argumente sprechen dagegen bzw. dafür:

**Nachteile:** Der 'entkoppelte' Ablauf ist weniger elegant, weil der nichtautomatisierte (= manuelle) und damit auch der Administrations-Aufwand wachsen.

Das komplexe Verfahren benötigt deutlich geringeren Speicherplatz, da weder die Pedfiles noch die Ausgabe-Files dauerhaft gesichert werden. Umgekehrt erfordert der entkoppelte Ablauf Überlegungen zu Speicherplatz-Verwaltung und -Minimierung.

Die Formate der Eingabe-Dateien für Famhap / Unphased einerseits und FBAT andererseits unterscheiden sich. Hotel\_Fam wiederum arbeitet nicht mit (Text-)Pedfiles sondern mit Datasets.

**Vorteile:** Der Gesamtprozess ist übersichtlicher und weniger komplex, was insbesondere das Testen einzelner Programm-Module verkürzt.

Ein weniger automatisierter Ablauf ermöglicht natürlich, die Prozesse viel detaillierter zu überwachen und ggf. bei Fehlern einzugreifen.

Da alle Ausgabedateien mit Datum und Uhrzeit gespeichert werden, hat man einen sehr guten Überblick über das Laufzeitverhalten in den einzelnen Szenarien.

Weil die originalen Ausgabedateien komprimiert archiviert werden, besteht jederzeit die Möglichkeit, bei Auffälligkeiten noch auf die Originalausgabe zurückzugreifen.

Durch Abstürze u. a. Fehler abgebrochene Jobs sind viel leichter und ohne Fehlerrisiko wieder aufzunehmen (siehe 5.3.2).

Auf Grund dieser Vorteile wurden die in Abschnitt 5.1 erwähnten drei Hauptabschnitte bei der Analyse der Familiendaten mit fehlenden Genotypen generell separat realisiert. Tab. 5.4 mit einem Auszug aus dem Fehlerprotokoll belegt, dass gerade der letzte Vorteil wegen vieler Ausfälle und Abstürze von Cluster und einzelner Knoten zum Tragen kam.

05.02.2008	Wartungsarbeiten, alle Jobs neu gestartet
06.03.2008	Stromausfall, alle Jobs neu aufgesetzt
03.04.2008	Stromausfall, alle Jobs neu aufgesetzt
16.04.2008	node13 abgestürzt, Famh_rez10m2m neu gestartet
17.04.2008	Fehler node5, Famh_rez10m3m neu gestartet
25.04.2008	Absturz node5, Famh_dom10m3m neu gestartet
14.05.2008	Absturz node14, Unph_rez22m2m neu aufgesetzt
15.05.2008	Absturz node2, Neuaufnahme Unph_dom22m3m
04.06.2008	Überhitzung node2, Neuaufnahme Unph_dom22m3m

Tabelle 5.4: Job-Unterbrechungen durch äußere Einflüsse



## 5.4.2 Erzeugung der Daten

Um sicherzustellen, dass die Stichproben kompletter und unvollständiger Genotypen bis auf ausgefallene Werte identisch sind, wurden Genotypen aus vollständigen Daten gemäß 2.2.2, 2.2.3 zufällig fallen gelassen. Das ließ sich leicht durch einen SAS-Makro realisieren, der jedem Genotypen kompletter Stichproben einen auf dem Einheitsintervall gleichverteilten Zufallswert  $x$  zuordnete und den Genotyp dann eliminierte, wenn  $x$  ins Intervall  $[0.75, 0.76)$  fiel.

Bei der Wahl des Intervalls lag folgende Vorüberlegung zu Grunde: Die SAS-Funktion `RANUNI(-1)` liefert für große Anzahlen auf  $[0, 1]$  ziemlich homogen verteilte Werte: In einer Vorstudie wurden  $N = 100.000$  Pseudozufallszahlen auf 20 Klassen im Einheitsintervall verteilt. Die niedrigste (4.94%) bzw. höchste Klassenhäufigkeit (5.108%) wich nur wenig vom theoretischen Wert ab. Für kleinere  $N$  sind die Abweichungen von der Gleichverteilung schon erheblicher. Das Intervall  $[0.75, 0.80)$  war jenes, wo die beobachtete Häufigkeit am wenigsten um die erwartete Anzahl streute.

Weiterhin wurden für die simulierten Familiendaten stichprobenartig für je 1000 Dateien eines Szenariums die mittleren Anteile fehlender Genotypen je Marker einschließlich Standardabweichung ermittelt. Dass die auf einer Normalapproximation basierenden Konfidenzintervalle der Differenz zum Sollwert 0.01 alle die Null enthalten, spricht dafür, dass der Anteil von 1% gut eingehalten wird. Exemplarisch dafür stehe das H0-Szenarium 1000 AAUU-Familien, NMod 1:

Marker	Mittelwert	Std <sup>a</sup>	Konfidenzintervall	
$M_1$	0.01002	0.0012	-0.00005	0.00010
$M_2$	0.00999	0.0013	-0.00009	0.00008
$M_3$	0.01003	0.0013	-0.00005	0.00011
$M_4$	0.01006	0.0013	-0.00002	0.00014
$M_1$	0.01004	0.0013	-0.00004	0.00012
all	0.04915	0.0028		

Tabelle 5.5: Anteil fehlender Genotypen

<sup>a</sup> Standardabweichung

Um den Rechenaufwand zu reduzieren, wurde die Analyse mit fehlenden Werten nicht für jedes Szenarium durchgeführt, sondern nur für jedes zweite (siehe Tab. 5.6)

	model	MUL			REZ			DOM			ADD		
		1	2	3	1	2	3	1	2	3	1	2	3
Fam.- typ	F10	X		X		X		X		X		X	
	F20		X		X	X			X		X		X
	F22	X		X		X		X		X		X	

Tabelle 5.6: Analyse mit 'Missings': Auswahl der Szenarien für die Analyse

## 5.5 Analyse in stratifizierten Populationen

Nach den Vorgaben in 2.2.5 wurden analog zu Absatz 5.1 wurden mit dem SAS-Makro `SIMFAM` (2.3) Familiendaten erzeugt. Diese wurden mit den vier Programmen und zur Kontrolle auch mit `TRANSMIT` (asymptotischer und Bootstrap-Test) analysiert und die Ausgabe ausgelesen. Den Fehler 1. Art erhält man wiederum, indem man die Anzahl zum Niveau  $\alpha$  signifikanter Tests durch die Anzahl Replikationen ( $N = 1000$ ) dividiert.

# Kapitel 6

## Auswertung der Analysen

### 6.1 Laufzeitbetrachtungen

Für die Untersuchung der Laufzeiten empfahl sich die Auswertung der Analyse mit fehlenden Genotypen, da hier die Assoziationsanalyse von der Simulation der Stichproben und dem Auslesen der Ergebnisse getrennt ist (vgl. 5.4.1): So ergab sich bei fehlerfrei beendeten Jobs die Rechenzeit einfach als Differenz zwischen Start- und Endzeit im Jobprotokoll (siehe 5.3.3). Nur im Falle von Hotel\_Fam müsste die Zeit subtrahiert werden, die für das Extrahieren der Resultate in ein Ergebnis-Dataset verbraucht wurde. Man darf aber davon ausgehen, dass diese gegenüber der reinen Analysezeit unerheblich ist.

Selbstverständlich hängt die Laufzeit auf den Cluster-Knoten substantiell davon ab, wie stark Prozessor und Speicher u. U. durch einen parallelen Job auf demselben Knoten belastet sind. In etlichen Szenarien wurden die Analysen durch äußere Einwirkungen (Rechnerabsturz, Stromausfall usw., s. Tab. 5.4) unterbrochen, so dass hier eine exakte Bestimmung der Laufzeiten aufwändig bzw. sogar unmöglich wäre. Die Häufigkeit solcher unvorhersehbarer Abbrüche war natürlich für zeitraubende Jobs - und damit für Unphased - im Durchschnitt höher. Das ist ein Grund, sich bei der Auswertung auf  $H_0$ -Szenarien mit 200 Familien zu beschränken.

Schließlich wurde die Analyse bei fehlenden Genotypen mit Hotel\_Fam nur für für das Modell KMod 1 durchgeführt, da das  $T^2$ -Verfahren keine Power für reine Haplotypeneffekte besitzt. So repräsentieren die Laufzeiten von 7 Szenarien unter der Nullhypothese (200 Familien) bzw. vier Szenarien unter der Alternative die Analyse mit unvollständigen Daten. Hier in Tab. 6.1 sind nur die Summen der Rechenzeiten sowie die Faktoren im Vergleich mit Famhap gegeben; die vollständigen Tabellen findet man im Anhang, Tab. A.5 und A.6.

	Famhap	FBAT	Hotel_Fam	Unphased
$H_0$				
Summe	64:49:48	221:57:16	759:00:09	785:08:16
relativ	1	3.4	11.7	12.1
$H_1$				
Summe	100:01:46	415:56:52	514:37:15	972:13:45
Faktor	1	4.2	5.1	9.7

Tabelle 6.1: Benötigte Laufzeiten (hh:mm:ss):  $H_0$  und  $H_1$  mit fehlenden Genotypen

Da die Laufzeiten von unbeeinflussbaren Faktoren abhingen, sind hinsichtlich der Rechenzeiten nur einige allgemeinere Aussagen möglich:

- Wie zu erwarten wächst die Laufzeit mit wachsendem Stichprobenumfang wie auch mit der

Familiengröße. Zwischen den Analysen mit kompletten sowie unvollständigen Genotypdaten gibt es keine nennenswerten Unterschiede.

- Die in 1000 Durchläufen benötigten realen Laufzeiten liegen in der Größenordnung im Bereich der Abschätzungen vor den Analysen.
- Famhap bewältigte die Analysen in der kürzesten Zeit, gefolgt von FBAT, Hotel\_Fam und Unphased. Die Zeiten für 1000 Analysen ausgewählter Szenarien unter der Null- bzw. Alternativhypothese mit teils fehlenden Genotypen sind in Tab. 6.1 zusammengefasst.
- Ob das genetische Modell die Rechenzeit beeinflusst, lässt sich nicht sagen, da die Unterschiede auch auf die sehr unterschiedlichen Stichprobenumfänge zurückzuführen sein können.
- Es fällt auf, dass der Laufzeitbedarf von Unphased besonders für AAUU-Familien extreme Werte erreicht.
- Hotel\_Fam dagegen scheint sensibel auf die Stichprobengröße zu reagieren, denn die beiden größten Samples (Rezessivmodell, KMod 3 bzw. 2, 11000 bzw. 6000 Trios) benötigen für die Analyse überdurchschnittlich viel Zeit.

Dass sich die realen Laufzeiten der vier Programme im Vergleich zur Vorabschätzung (vgl. Tab. 5.1) nicht so extrem unterscheiden, hat folgende Ursachen:

Die Rechenzeit von Unphased steigt mit wachsender Kinderzahl stark an (vgl. Tab. A.3 im Anhang). Dass für die Auswertung hier die beiden Szenarien (NMod 2 und NMod 3 mit AAUU-Familien) nicht einbezogen werden konnten, stellt dieses Programm in ein günstiges Licht. Ebenso wird die Bilanz von Unphased dadurch geschönt, dass als zeitraubend zu erwartende Jobs von vornherein separat auf einem Knoten gestartet wurden, währenddessen sich andere Jobs die Ressourcen teilen mussten.

Für Hotel\_Fam hingegen steigt die Rechenzeit stark an, wenn die Stichproben größer werden. Die vier hier zusammengefassten  $H_1$ -Szenarien betreffen aber Stichprobenumfänge von nur 200 bis 600 Familien. Damit verzerrt die Auswahl der Szenarien die Statistik auch zugunsten von Hotel\_Fam.

## 6.2 Validität unter der Nullhypothese mit kompletten Genotyp-Daten

In der Simulationsstudie liefert jeder Algorithmus für jedes Szenarium für 1000 Replikationen je eine Teststatistik und den zugehörigen p-Wert. Dabei ist zu beachten, dass die p-Werte nicht im Sinne R. Fishers als Evidenz für Effekte [60] zu betrachten sind, sondern als Basis für Hypothesentests nach dem Neyman-Pearson-Ansatz [116], [67]. Leider hat in der Vergangenheit die Vermischung dieser beiden Ansätze vielfach zu Fehlinterpretationen geführt [68].

Ist das Fehlerniveau  $\alpha$  vorgegeben, so erhält man unter Nullsituationen den Fehler 1. Art und unter der Alternative die statistische Power als Anteil Replikationen mit einem  $p \leq \alpha$ . Entsprechend der Komplexität der untersuchten Situationen ist die Auswertung in ganz verschiedenen Richtungen möglich:

1. Wie gut halten die Algorithmen den Fehler 1. Art für die Markerkombinationen ein?
2. Welche Auswirkungen hat das Fehlerniveau 1. Art,  $\alpha = 0.05$  und  $\alpha = 0.01$ ?
3. Welchen Einfluss hat der Stichprobenumfang?
4. Wie unterscheiden sich die Tests der Hypothesen  $H_0^L$  bzw.  $H_0^A$ ?
5. Wie schneiden die Verfahren im Vergleich der overall-p-Werte ab?
6. Welche Auffälligkeiten und Besonderheiten gibt es?

## 6.2.1 Ergebnisse für einzelne Markerkombinationen (Mk)

Wie schon in Abb. 4.1 dargestellt, sind die Markerkombinationen binär angeordnet. In Diagrammen des Fehlers 1. Art bilden die 31 einzelnen Mk die x-Achse. Darüber wurden als Punktdiagramm die falsch-positiv-Raten verschiedener Tests in unterschiedlichen Farben aufgetragen. Zur besseren Sichtbarkeit wurden die Punkte verbunden. Die Strichlinie kennzeichnet das Soll-Fehlerniveau  $\alpha = 0.05$ . Indem die Werte der 1000 Replikationen als unabhängige Wiederholungen eines Bernoulli-Versuchsschemas aufgefasst werden, sind die gepunktet eingezeichneten Konfidenzgrenzen einer Binomialverteilung  $\text{Bin}(1000, 0.05)$  ein sinnvoller Anhaltspunkt.

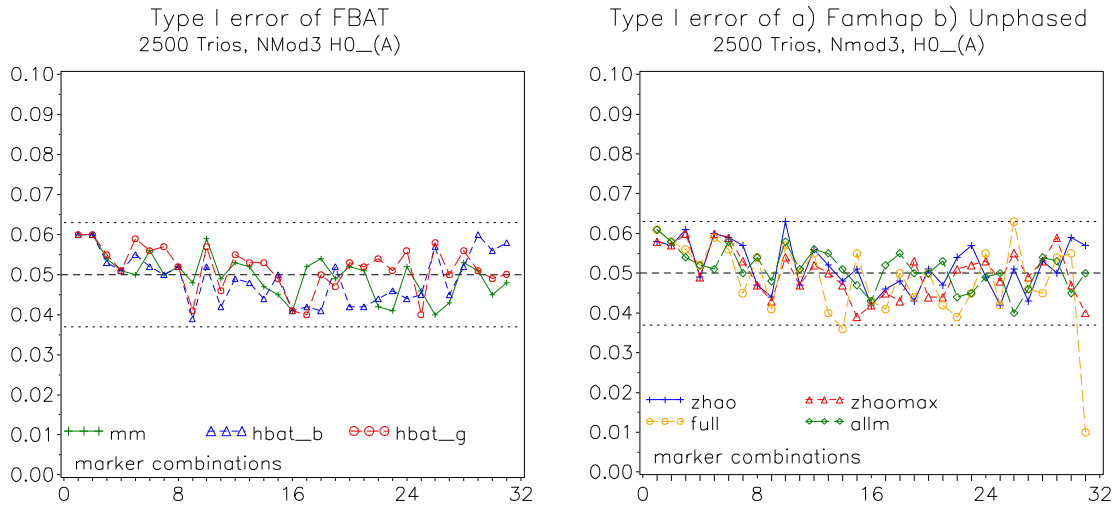


Abbildung 6.1: Fehler 1. Art: 2500 Trios, NMod 3, Tests auf Assoziation, links: FBAT, rechts: Famhap / Unphased

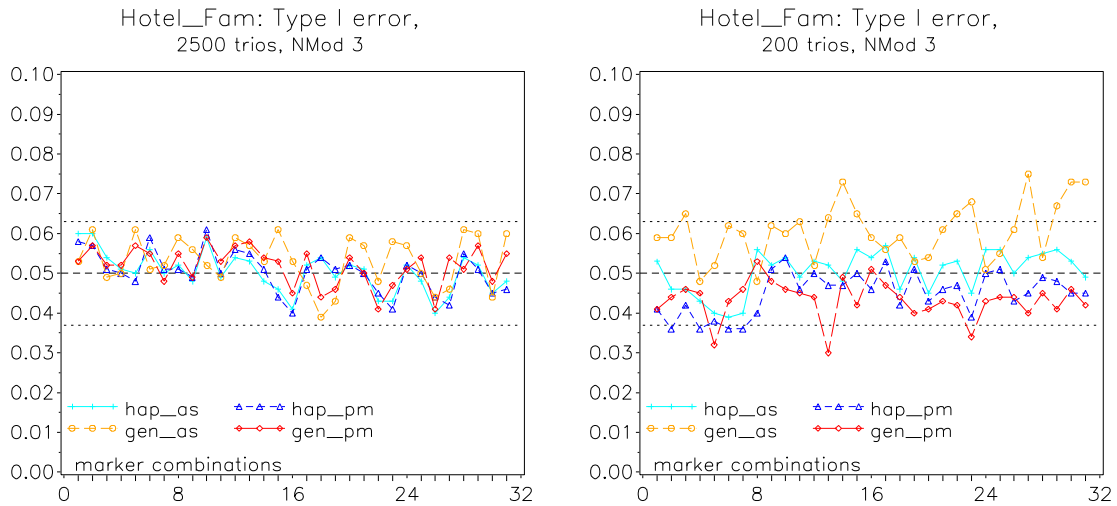


Abbildung 6.2: Fehler 1. Art: Hotel\_Fam, 2500 bzw. 200 Trios, NMod 3

Abb. 6.1 für das Szenarium NMod 3 mit 2500 Trios zeigt einen typischen Verlauf: Links sieht man Mk-spezifische Fehlerwerte der drei Assoziationstests von FBAT: grün der Multimarker-Test und rot bzw. blau der Haplotyp-Test (globale Summenstatistik über alle Haplotypen bzw. Bonferroni-korrigierter minimaler p-Wert). Rechts findet man die analoge Darstellung für je zwei Optionen von Famhap bzw. Unphased und in Abb. 6.2 links die von Hotel\_Fam.

Auf den ersten Blick erkennt man folgende Eigenschaften: Die meisten Verfahren halten das Fehlerniveau 1. Art relativ gut ein. Aber als konservativ ist auch kein Verfahren einzustufen. Selbst wenn sich die Statistiken für viele Mk unterhalb des Fehlerniveaus bewegen, gibt es meist einzelne Kombinationen, wo  $\alpha$  überschritten wird. Ausnahme ist der asymptotische  $T^2$ -Test mit Genotypcodierung für 200 Familien (Abb. 6.2 rechts) mit einer deutlich erhöhten falsch-positiv-Rate. Die Werte streuen meist unregelmäßig um die Fehlerschranke (z. B. für 1000 Trios im Modell NMod 2 von ca. 0.03 bis 0.07, vgl. Abb. 6.3). Die Vermutung, dass die Werte bei höheren Stichprobenumfängen weniger fluktuieren, wurde nicht bestätigt. Die niedrigen Werte für Mk 15, 23, 27 und 31 von Unphased (volles Modell) in Abb. 6.4 und Abb. 6.1 rechts werden später erörtert.

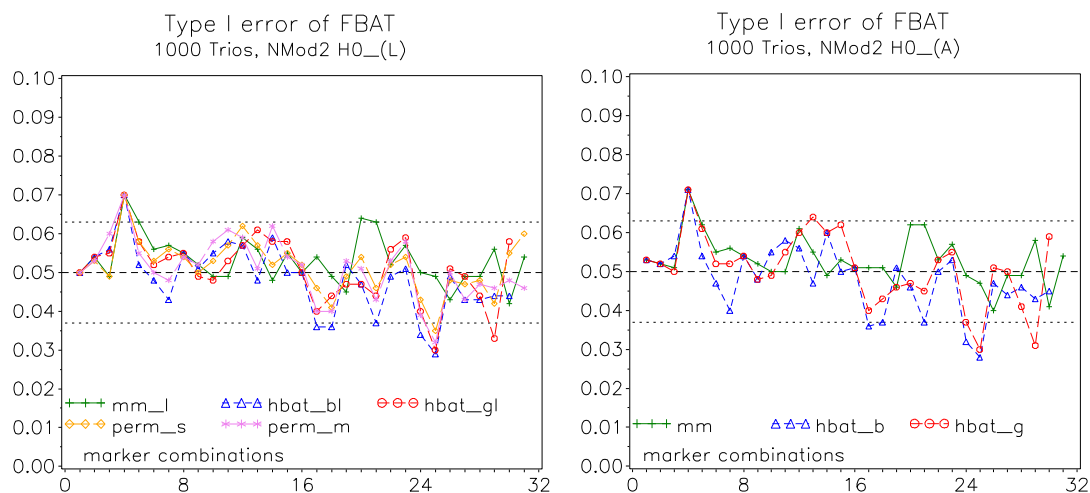


Abbildung 6.3: Fehler 1. Art, FBAT, NMod 2, 1000 Trios;  
 Test von  $H_0^L$  (links) vs.  $H_0^A$  (rechts)

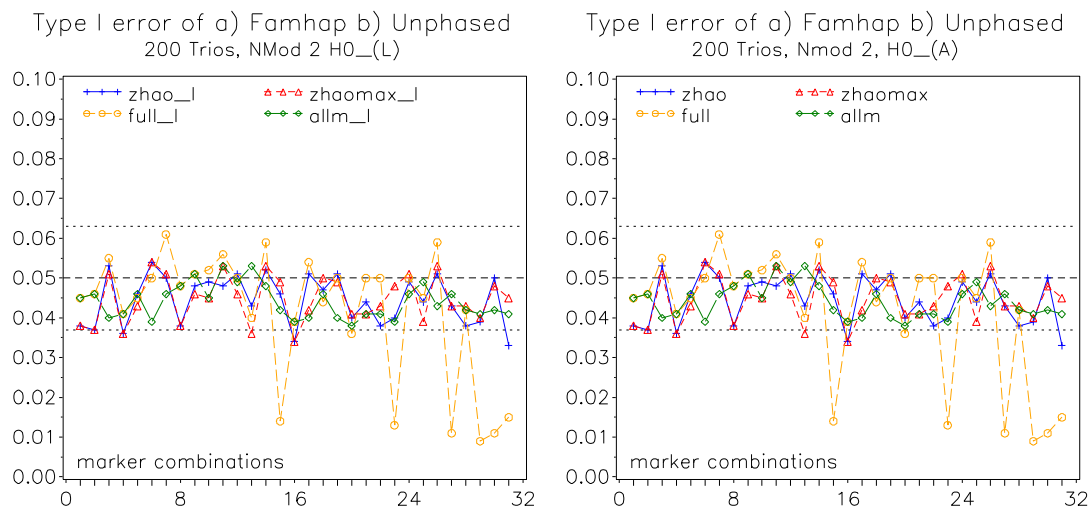


Abbildung 6.4: Fehler 1. Art, Famhap und Unphased: NMod 2, 200 Trios;  
 Test auf  $H_0^L$  (links) vs.  $H_0^A$  (rechts)

## 6.2.2 Test auf Kopplung

Einander entsprechende Tests auf Kopplung bzw. auf Assoziation (z. B. für FBAT `mm` und `mm_1` bzw. `hbat_g` und `hbat_g1`) sind für Trios identisch (Famhap, Unphased) bzw. ähnlich (FBAT). Die Identität ist anhand der vollständigen Übereinstimmung der Resultate erkennbar. Entsprechende Tests, die sich im Namen nur durch den Anhang '\_1' unterscheiden, wurden in gleichen Farben dargestellt. Die Analogien sind in den Abb. 6.3 und 6.4 gut zu erkennen.

Die Konfiguration NMod 2 modelliert die Situation 'keine Assoziation bei vorhandener Kopplung' zwischen Markern und Risikolocus (s. Absatz 2.2.3). Unter Geschwistern ist dann die Übertragung der Allele von den Eltern nicht unabhängig. Das resultiert in signifikanten Tests der Hypothese  $H_0^L$  für Familien mit mehreren Kindern. Abb. 6.5 zeigt deutlich, wie der Fehler 1. Art in diesem Fall inflationiert. Die Ähnlichkeit der Grafiken für FBAT (links) bzw. Famhap bzw. Unphased (rechts) rührt offensichtlich daher, dass die falsch-positiv-Rate mit wachsender Anzahl von Markern in der Kombination zunimmt. So treten die Maxima für Mk 15, 23, 27, 29, 30 und 31 auf, also solchen Kombinationen mit 4 bzw. 5 Markern.

Am wenigsten inflationiert der Fehler 1. Art für die Multimarker-Tests: `mm_1` (FBAT) bzw. `allm_1` (Unphased). Die meisten falsch Positiven gibt es sowohl für das `zhao_1`-Verfahren (Famhap) als auch für `perm_s` und `hbat_g1` von FBAT. Die Werte von `zhaomax_1` (Famhap) bzw. `perm_m` und `hbat_b1` (FBAT) liegen dazwischen.

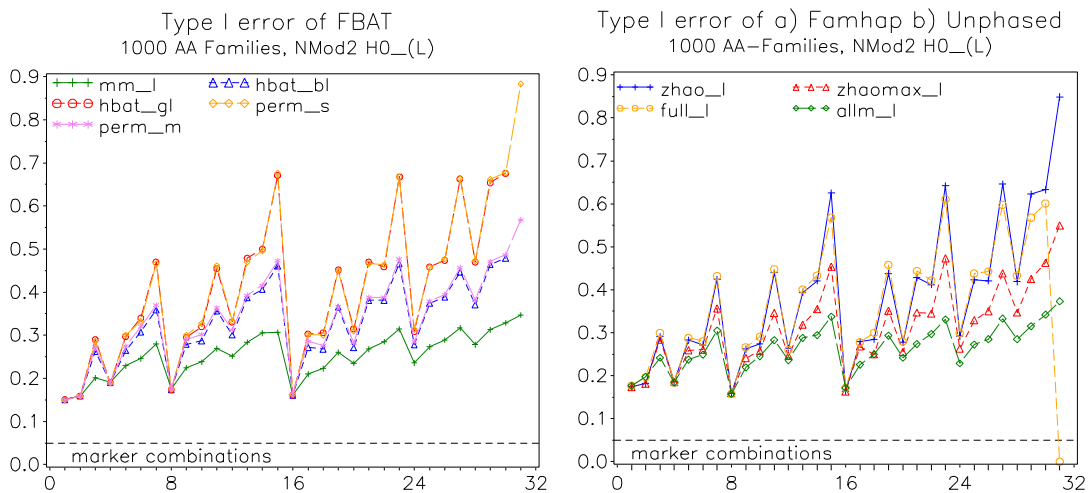


Abbildung 6.5: Test auf Kopplung: NMod 2, 1000 AA-Familien;  
links: FBAT, rechts: Famhap und Unphased

## 6.2.3 Einfluss des Fehlerniveaus

Testet man zum Fehlerniveau  $\alpha = 0.01$ , verändern sich die Aussagen nicht grundlegend. Das wird deutlich, wenn man Abb. 6.6 mit Abb. 6.3 bzw. 6.2 rechts vergleicht: Die falsch-positiv-Raten bewegen sich im wesentlichen innerhalb der Konfidenzgrenzen. Gelegentlich gibt es leicht erhöhte Werte. Unphased (volles Modell) und die `gen_as`-Statistik von `Hotel_Fam` sind besonders für  $N=200$  teilweise antikonservativ. Auch die Inflation des Fehlers 1. Art der Tests auf  $H_0^L$  für NMod 2 bei Familien mit mehreren kranken Kindern gilt ganz analog. Wie zu erwarten liegen die extremen Werte gegenüber  $\alpha = 5\%$  etwas niedriger.

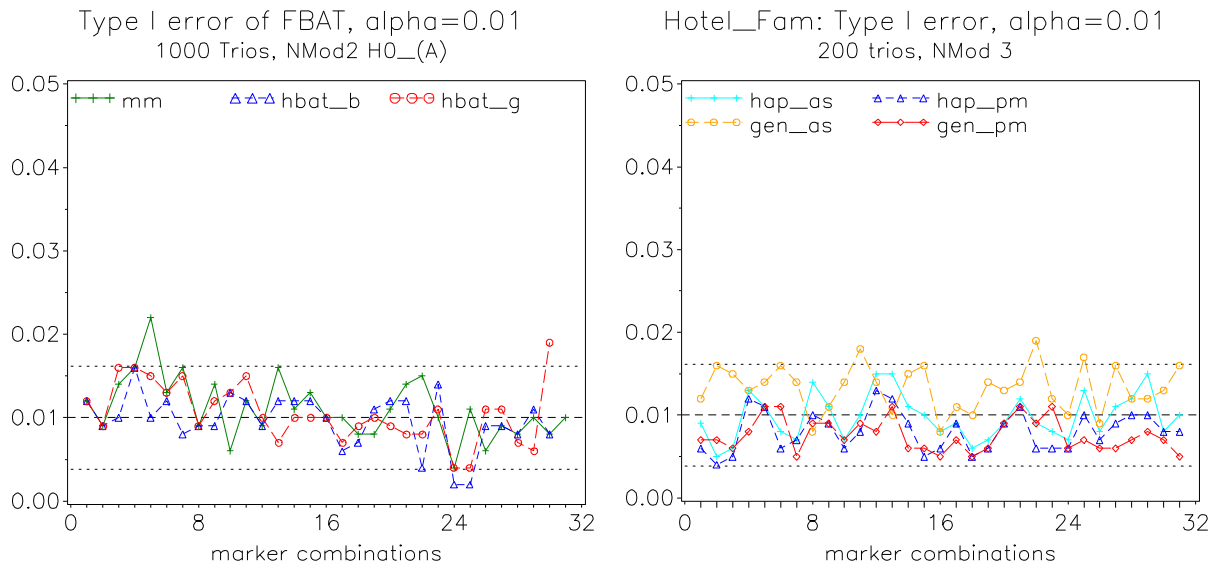


Abbildung 6.6: Nullhypothese: Test zum Fehlerniveau  $\alpha = 0.01$ , links: FBAT, 1000 Trios, NMod 2; rechts: Hotel\_Fam, 200 Trios NMod 3

### 6.2.4 Overall-p-Wert über alle Markerkombinationen

In der Praxis ist meist gefragt, ob *irgendeine* Markerkombination mit der Krankheit assoziiert ist. Deshalb ist eine Aussage über alle Mk in der Regel wertvoll. So wurden den totalen Werten von Famhap Bonferroni-korrigierte minimale p-Werte über alle 31 Mk der anderen Programme gegenübergestellt (s. auch 6.3.4). In Abbildung 6.7 sind die Fehler 1. Art für alle 20 Statistiken gemäß Tab. 5.3 für Modell NMod 1 als Balkendiagramm dargestellt. Im oberen Teil findet man die Tests auf  $H_0^A$ , im unteren die Tests auf Kopplung. Das  $\alpha$ -Fehlerniveau wurde wiederum durch eine gepunktete Linie gekennzeichnet.

Die overall-p-Werte von Famhap streuen um dieses Fehlerniveau. In 7 von 21 Szenarien übersteigen sie ein wenig (0.2 bis 1.3%) die 5%-Grenze, in vier Fällen schöpfen sie den vorgegebenen Rahmen nicht aus. Dabei ist keine Regelmäßigkeit - weder hinsichtlich des Modells, des Stichprobenumfangs, noch der Optionen *zhao* und *zhaomax* erkennbar (s. Abb. 6.8). Die Bonferroni-Korrektur für FBAT, Unphased und Hotel\_Fam überträgt ihren konservativer Charakter auf die overall-p-Statistiken.

In Abb. 6.9 ist das Modell NMod 2 mit AAUU-Familien dargestellt. Im oberen Teil sieht man, dass alle Tests auf Assoziation das Niveau gut einhalten, währenddessen der Fehler 1. Art für die Kopplungs-Tests inflationiert. Das belegt eindrucksvoll, dass man in Situationen mit Kopplung entweder mit Trio-Daten rechnen oder für Familien mit mehr als einem Kind den Test auf  $H_0^A$  verwenden muss. Sonst sind falsch positive Befunde zu erwarten.

Type 1 Error: alpha=0.05, 200 trios, NMod2

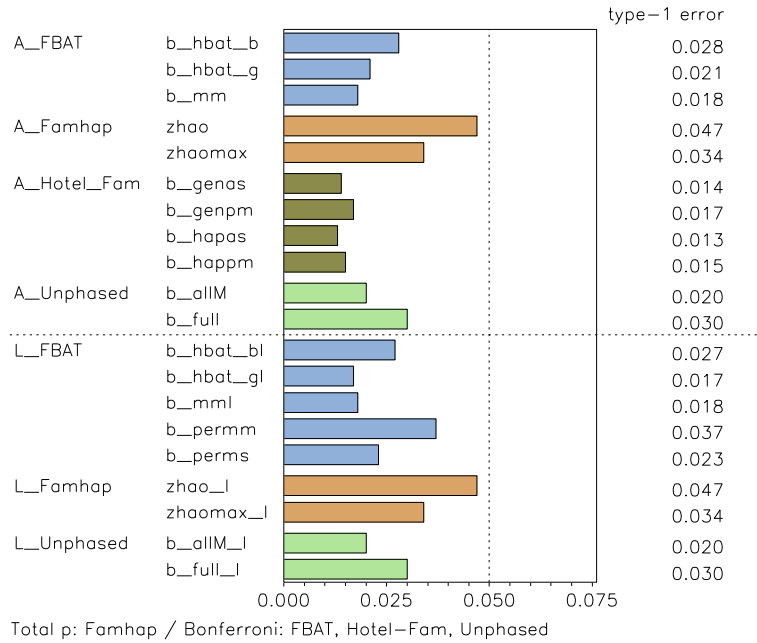


Abbildung 6.7: Fehler 1. Art, overall; 200 Trios, NMod 2  
Type 1 Error: alpha=0.05, 1000 AA-families, NMod1

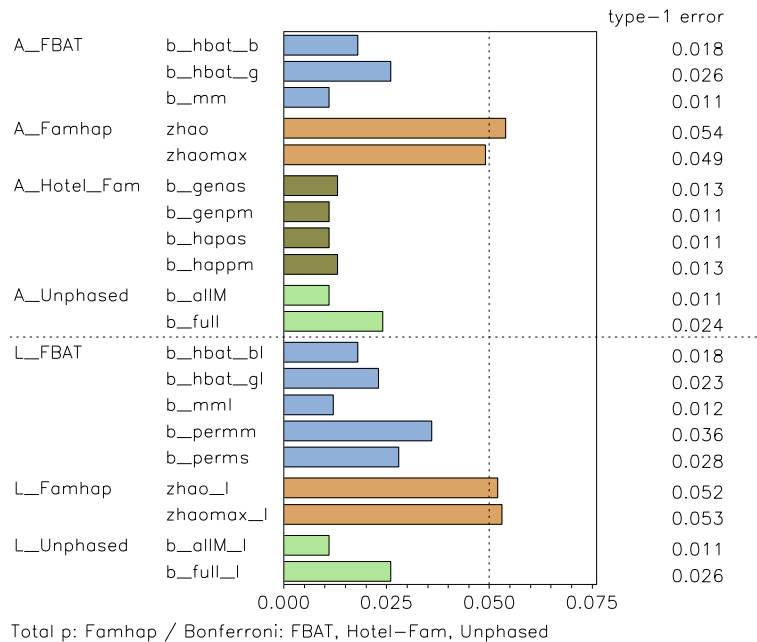


Abbildung 6.8: Fehler 1. Art, overall; 1000 AA-Familien, NMod 1



Type 1 Error: alpha=0.05, 2500 AAUU-families, NMod2

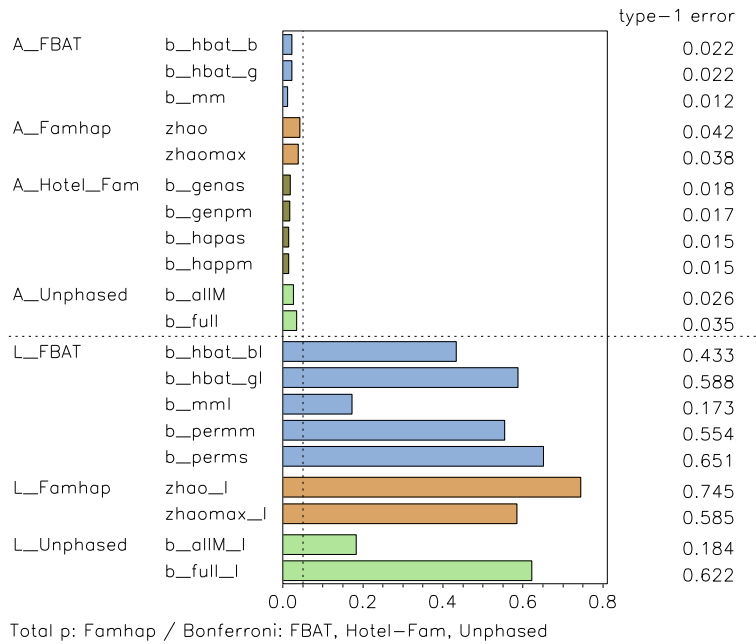


Abbildung 6.9: Fehler 1. Art, overall; 2500 AAUU-Familien, NMod 2

## 6.2.5 Besonderheiten

### FBAT

Multimarker-Test und der globale Haplotyp-Test sowie die beiden Permutationstests bewegen sich unregelmäßig meist innerhalb des Konfidenzbandes. Systematische Unterschiede zwischen den drei Tests auf Assoziation bzw. den fünf Tests auf Kopplung sind nicht identifizierbar.

Links in den Abb. 6.3 und 6.5 erkennt man bei genauem Hinsehen, dass die rot / blau dargestellten Haplotypstests von FBAT für Mk 31 keine Werte besitzen. Das liegt an der Option `minfreq` 0.05. Das wird im Abschnitt 6.6.2 noch ausführlich erörtert.

### Famhap

Die Fehler 1. Art der Optionen `zhao` bzw. `zhaomax` schwingen unabhängig voneinander um  $\alpha = 5\%$ . Nur selten übersteigen Werte das Konfidenzband. Die Fluktuationen um das  $\alpha$ -Niveau sind bei Famhap und FBAT vergleichbar.

### Hotel\_Fam

Der asymptotische Test mit Genotyp-Codierung von Hotel\_Fam ist für 200 Familien recht antikonservativ (8 von 9 Szenarien). Das bestätigt Ergebnisse von R. Fan et al. [57] und kann darauf zurückzuführen sein, dass bei Genotypcodierung in Formel (4.22) der Erwartungswert  $E(\bar{Y}^{(P)} - \bar{X}^{(O)})$  nicht 0 ist. Der permutationsbasierte Test hält bis auf wenige Ausnahmen das Fehlerniveau gut ein. Mit wachsendem Stichprobenumfang nähern sich asymptotische und Permutations-Test weitgehend an.

Für AA-Familien, NMod 2 (Abb. 6.10 links) fällt ein rhythmisches Verhalten auf: Der Fehler 1. Art des `gen_as`-Verfahrens erreicht Spitzenwerte für alle Mk, welche die Marker  $M_4M_5$  enthalten, das sind Mk 3, 7, 11, 15 usw. Die Grafik rechts zeigt die Analyse mit anderen Stichproben, die

zur Kontrolle durchgeführt wurde. Hierbei wurden durch einen anderen Startwert des Zufallszahlengenerators völlig andere Familien mit denselben Eigenschaften erzeugt (vgl. Abb. 5.1). In der zweiten Simulation kann man einen regelmäßigen Rhythmus nicht mehr finden, die falsch-positiv-Rate bleibt allerdings zu hoch.

Die Ergebnisse bestätigen, dass der asymptotische Test mit Genotypcodierung antikonservativ ist [57]. Mehrere Autoren hatten bereits darauf hingewiesen, dass die Näherung von Hotelling's  $T^2$ -Statistik an die  $\chi^2$ -Verteilung - insbesondere bei Verletzung der Normalverteilungsannahme - ungenügend sein kann (z. B. [154], [87], [146], [115]). Aus diesem Grund sollte dem Permutationsbasierten Test generell der Vorrang gegeben werden.

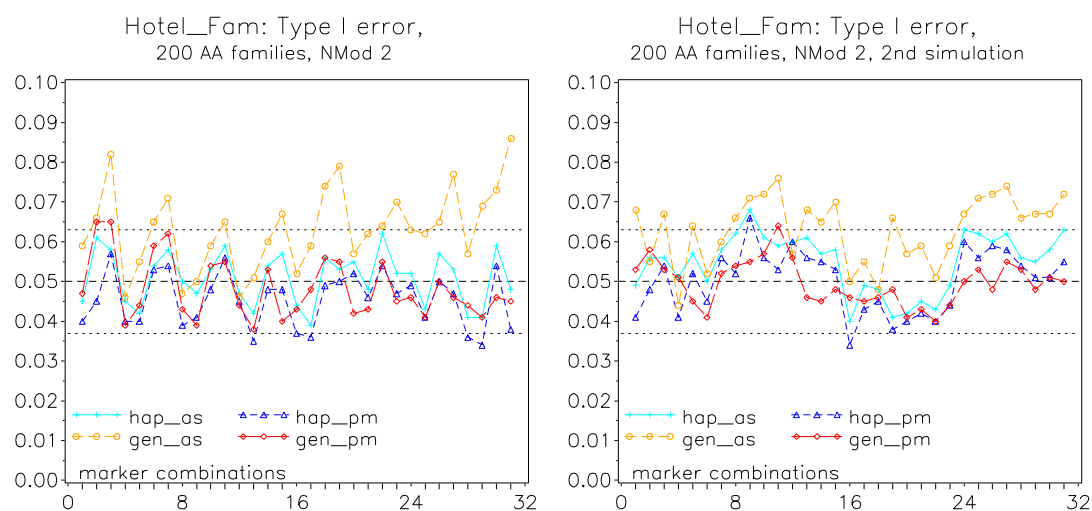


Abbildung 6.10: Fehler 1. Art, Hotel\_Fam, NMod 2, 200 AA-Familien: systematisches Verhalten? Wiederholung der Analyse mit anderen Stichproben

Ähnlich liegen die Dinge im Szenarium 200 Trios, NMod 1 (nicht abgebildet): Hier erreichte der Fehler 1. Art Maxima für alle ungeraden  $M_k$ , also solche, die Marker  $M_5$  enthalten. Eine Wiederholung der Analyse mit anderen Stichproben bestätigte zwar den erhöhten Fehler 1. Art, nicht aber den gleichmäßigen Rhythmus. Weiterhin sprechen die folgenden Argumente dafür, dass es sich hier nicht um Artefakte bei der Erzeugung oder Analyse der Daten handelt, sondern um zufällige Erscheinungen:

Erstens wurden die Analysen der neun Szenarien mit 200 Familien hintereinander durch *einen* SAS-Makro, also unter gleichbleibenden Bedingungen, durchgeführt. So wäre es wahrscheinlich, dass sich programmbedingte Effekte in mehreren Szenarien gleichförmig wiederholen und nicht variiert.

Weiterhin sollten gewisse 'Muster' ähnlich beim asymptotischen Test mit Haplotypcodierung zu finden sein. Zwar lassen sich Ähnlichkeiten nicht leugnen (s. Abb. 6.10), aber bei genauem Hinsehen fällt auf, dass ein monotoner Rhythmus des orangen bzw. türkisfarbenen Verlaufs immer wieder durchbrochen wird.

Drittens erwartet man, dass sich systematische Effekte bei höherer Stichprobenzahl verstärken. Das ist nicht beobachtbar: In Abb. 6.2 z. B. kann man erkennen, dass sich der erhöhte Fehler 1. Art für 200 Familien bei 2500 Familien normalisiert hat.

## Das Modell NMod 2

Abb. 6.10 und Abb. 6.11 weisen mit relativ hohen Fehlerwerten für verschiedene  $M_k$  darauf hin, dass das Modell NMod 2 in gewisser Hinsicht extrem ist: ein rezessives Modell mit einer Häufigkeit des Risikoallels von 1%. Die Penetranzen  $f_2 = 1$  und  $f_1 = f_0 = 0$  bedeuten, dass nur homozygote Nachkommen krank sind. Da Rekombination ausgeschlossen wurde, überträgt also jedes Elternteil

den Haplotyp mit dem Risikoallel. Für AA-Familien bedeutet das aber, dass bei heterozygoten Eltern beide kranken Geschwister identische Haplotypen der Marker  $M_1 \dots M_5$  erhalten. Nur wenn mindestens ein Elternteil selbst krank und damit homozygot ist, können verschiedene Haplotypen an die Nachkommen transmittiert werden. Tatsächlich zeigt sich, dass die Stichproben von AA-Familien des Modells NMod 2 nur ca. 2% Familien enthalten, in denen sich die Geschwister genotypisch unterscheiden. (Der theoretische Anteil beträgt  $2 \cdot 0.01 + 0.01^2 = 0.0201$ ). Durch die Simulation entstandene zufällige 'Haplotyp-Effekte' werden also bei zwei kranken Kindern wegen der identischen Kopien noch verstärkt, was die Schwankungsbreite des Fehlers 1. Art erhöht.

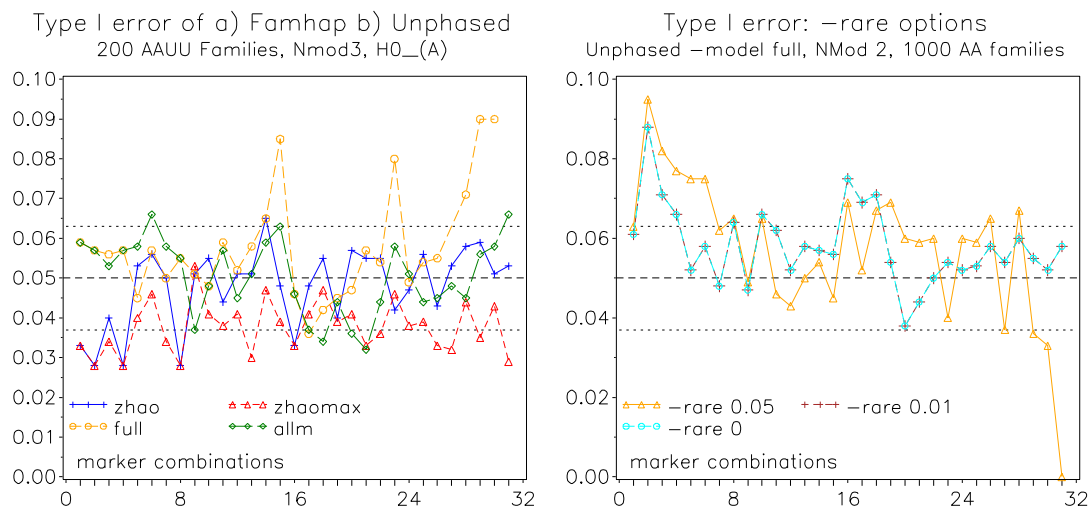


Abbildung 6.11: Fehler 1. Art: Famhap / Unphased, links: NMod 3, 200 AAUU-Familien; rechts: NMod 2, 1000 AA-Familien, unphased -model full -rare 0.05, 0.01, 0

### Unphased / die Option -rare

In etwa der Hälfte der Szenarien zeigen die Optionen `-model full` und `-model allelemain` von Unphased für wenige Markerkombinationen 'Ausreißer', insbesondere für den Stichprobenumfang  $N=200$ . Besonders extrem ist das Szenarium (NMod 3, 200 AAUU-Familien, vgl. Abb. 6.11 links). Hier liegt der Fehler 1. Art für die MK 15, 23, 29 und 30 zwischen 0.08 und 0.09 und verschwindet bei Mk 31 mit dem Wert 0.315 aus dem Zeichenbereich! `allm` ist hier ebenso etwas antikonservativ, aber keinesfalls extrem. Auch für dasselbe Modell mit AA-Familien liegt die falsch-positiv-Rate der `full`-Methode für Mk 31 bei 10%. Mit wachsendem Stichprobenumfang gibt es weniger extreme Werte, aber sowohl das volle als auch das Haupteffekt-Modell sind gelegentlich antikonservativ. Das gilt unabhängig vom gewählten Fehlerniveau  $\alpha = 0.05$  bzw.  $\alpha = 0.01$ .

Übrigens bestätigt auch die Analyse mit fehlenden Genotypen die hohen Fehler 1. Art für einige Mk. So erreicht der Fehlerwert in den Szenarien (NMod 3, 200 AA- bzw. AAUU-Familien) z. B. 0.103 bzw. 0.33.

Oben war schon festgestellt worden, dass der Fehler 1. Art von Unphased (volles Modell) für Kombinationen aus 5 und teilweise auch 4 Markern in vielen Szenarien stark abfällt bzw. 0 wird. Diesen Effekt sieht man in Abb. 6.11 rechts an der orangenen Linie. Es wurde vermutet, dass dieser Effekt durch die Option `-rare 0.05` hervorgerufen wird. Diese Schranke für seltene Haplotypen bedeutet, dass Haplotypen mit einer geringeren Häufigkeit gepoolt, d. h. mit einunddemselben Risikoparameter parametrisiert werden. Betrifft das alle vorkommenden Haplotypen, wird nur ein einziges gemeinsames Risiko modelliert, was einen p-Wert  $p = 1$  zur Folge hat. Durch eine niedrigere Schranke 1% (bzw. durch das Weglassen der Option) müsste dieser ungewollte Effekt abgestellt werden. Um das nachzuweisen, wurde die Analyse mit der Option `-rare 0.01` bzw. ohne diese Option wiederholt. Abb. rechts zeigt die Graphen für verschiedene Optionen: Zunächst fällt auf,

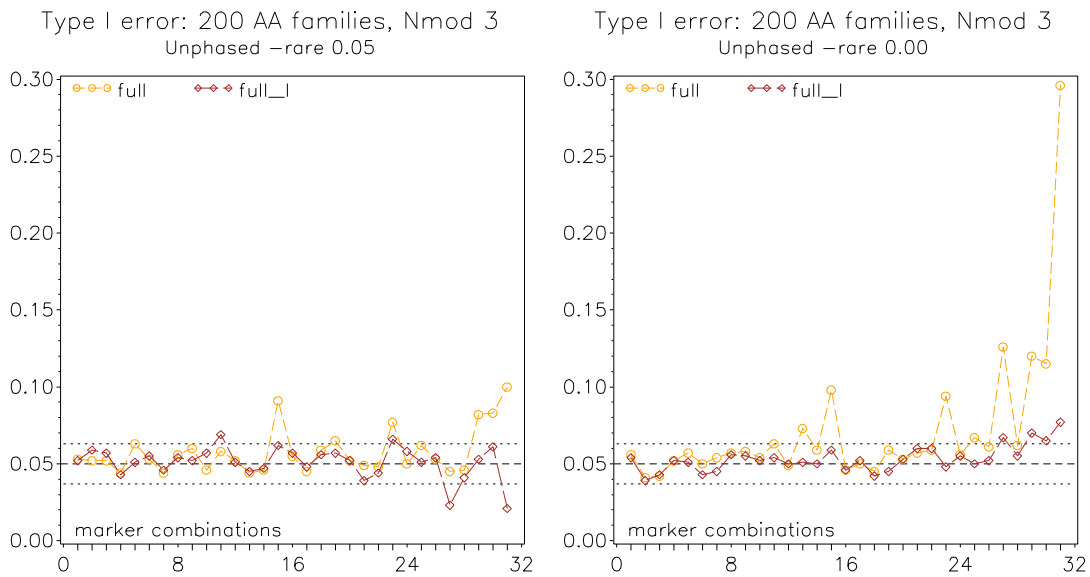


Abbildung 6.12: Unphased mit Option `-rare 0.05` (links) bzw. ohne diese Option (rechts)

dass die braune und türkisfarbene Linie zusammenfallen. D. h. ein Absenken der Haplotypschranke von 1% auf 0 bewirkt hier keine Veränderung mehr. Außerdem bewegt sich der Fehler für Mk 31 tatsächlich im Konfidenzbereich.

Unter Nullsituationen haben aus der Option `-rare 0.05` resultierende  $p$ -Werte  $p = 1$  keine negativen Auswirkungen, weil sie keine falsch positiven Befunde erzeugen.

Zu den wenigen Szenarien mit 200 Familien, bei denen der Fehler 1. Art nach oben ausreißt, gehört NMod 3 mit AA-Familien. In Abb. 6.12 wurden nur die beiden Tests für das volle Modell von Unphased dargestellt: Links mit Option `-rare 0.05`, rechts ohne diese Option. (Das Haupteffekt-Modell fehlt hier, weil es nach Empfehlung des Handbuchs nicht mit der Option `-rare` angewandt werden soll.) Schon in der Standard-Analyse erreicht der Fehler 10% für Mk 31. Eine wiederholte Analyse mit der Option `-rare 0.01` bzw. ohne diese Option verschlimmerte die Situation dramatisch. Es ist nicht klar, was den hohen Anteil falsch Positiver für Mk 31 (`-rare 0.01`: 24.5%, ohne die Option: 29.6%) verursacht. Der Test auf Kopplung ist hier für Mk mit hohen Ordnungsnummern auch leicht antikonservativ, aber keinesfalls außergewöhnlich.

Weil dieser Fall der einzige extreme aller Szenarien bleibt, ist eine Deutung schwierig: Möglicherweise ist die asymptotische Näherung der LRT-Statistik an die  $\chi^2$ -Verteilung bei 200-Familien-Stichproben noch ungenügend. Ebenfalls ist denkbar, dass es bei der Maximierung der Likelihood Instabilitäten gibt. In diesem speziellen Fall verhinderte die Option `-rare 0.05` das 'Ausreißen' des Fehlers 1. Art. Weitere Gedanken zur Schranke 5% für seltene Haplotypen finden sich im Absatz 6.6.2.

## 6.3 Power der Tests

Hinsichtlich der Power unter den Alternativ-Modellen werden folgende Fragen untersucht:

1. Welche Power erzielen die Algorithmen für die einzelnen Markerkombinationen?
2. Wie unterscheiden sich die verschiedenen Statistiken?
3. Welche Unterschiede gibt es zwischen Tests auf Kopplung bzw. Assoziation?
4. Wie schneiden die Algorithmen beim Vergleich der overall-p-Werte ab?
5. Welche Veränderungen gibt es beim Test zum Niveau  $\alpha = 0.01$  anstelle von  $\alpha = 0.05$ ?
6. Was fällt besonders auf?

### 6.3.1 Power für einzelne Markerkombinationen

Die Power für die 31 Mk hängt substantziell von den Markerkonfigurationen (s. 2.2.2) ab. Im Modell KMod 1 steht nur Marker  $M_3$  - das entspricht Mk 4 - im LD zum Krankheitslocus. So sind in Grafik 6.13, mit Mk 4 beginnend, alle Markerkombinationen grün dargestellt, die  $M_3$  enthalten. Vergleicht man nun z. B. damit die Power-Diagramme für FBAT bzw. Hotel\_Fam (s. Abb. 6.14) wird die Analogie offensichtlich.

#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
M1																x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
M2								x	x	x	x	x	x	x	x									x	x	x	x	x	x	x	x
M3			x	x	x	x	x					x	x	x	x					x	x	x	x				x	x	x	x	x
M4		x	x				x	x		x	x			x	x			x	x			x	x			x	x			x	x
M5	x		x		x		x		x		x		x		x		x		x		x		x		x		x		x		x

Abbildung 6.13: Markerkombinationen Konfiguration KMod 1

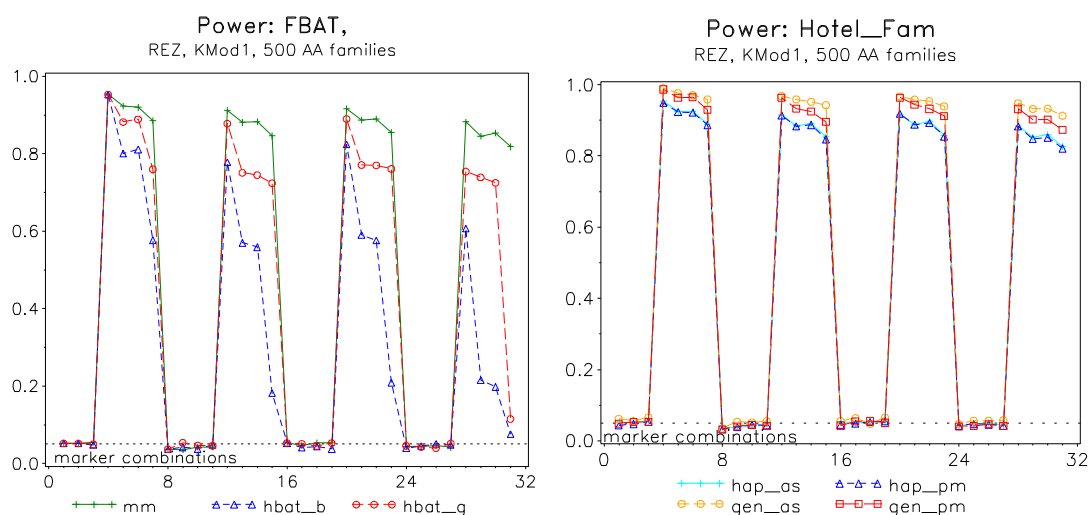


Abbildung 6.14: Rezessives Modell KMod 1, AA-Familien

Wie schon im vorangegangenen Abschnitt bilden die 31 Mk die horizontale Achse. Darüber ist in verschiedenen Farben die Power unterschiedlicher Statistiken aufgetragen und durch Linien verbunden. Es war zu erwarten, dass für Konfiguration KMod 1 die Power der Tests ihr Maximum für Mk 4 erreicht. Auch für alle anderen Kombinationen, die Marker 3 enthalten, wird hohe

bis mittlere Power erzielt. Der Abfall ist umso stärker, je mehr Marker zur Kombination hinzukommen. Die Powerverläufe der verschiedenen Algorithmen (siehe Tab. 5.3) ähneln sich in ihrer Form, unterscheiden sich aber im Niveau teilweise deutlich. Dass zwischen  $M_3$  und den anderen Markern kein LD angenommen wurde, wirkt sich so aus, dass Kombinationen ohne  $M_3$  überhaupt keine Power haben, den Assoziationseffekt zu entdecken. Hier bewegen sich die Werte auf dem  $\alpha$ -Fehlerniveau. Die Diagramme von Famhap und Unphased ähneln dem von FBAT sehr. Favorit in diesem Szenarium ist zweifellos Hotel\_Fam (Abb. 6.14 rechts), das für Mk 4 eine Power nahe 1 erreicht. Auch für Kombinationen mit mehr Markern fällt die Power langsamer ab.

Ein anderer Blickpunkt ist der Power-Vergleich ausgewählter Verfahren jeweils für eine Markerkombination. Dafür wurden die Kombinationen 4, 14 bzw. 30 ausgewählt. Mk 4 entspricht dem Marker  $M_3$  und genau dieser ist in Konfiguration KMod 1 mit der Krankheit assoziiert. Die Power ausgewählter Verfahren für diese Markerkombination findet man in Tab. 6.2 aufgelistet: links Tests auf Kopplung, rechts Tests auf Assoziation. KMod 2 und 3 sind an dieser Stelle uninteressant, da die Algorithmen dort keine Power haben (s. u.). So beschränkt sich die Tabelle auf die 12 KMod-1-Szenarien. Außerdem wurden etliche Verfahren weggelassen, da sie in dieser speziellen Situation identische Werte liefern: Weil alle Verfahren für Mk 4 ihr Maximum erreichen und ein Single-Marker-Effekt vorliegt, stimmen `allM` mit `full` und `allM_1` mit `full_1`, `zhaomax_1` mit `zhao_1` und `hbat_b1` mit `hbat_g1` überein. Vermutlich aus ähnlichen Gründen sind auch die Werte von `perm_m` und `perm_s` gleich. Wegen ihres teilweise antikonservativen Charakters wurden die beiden asymptotischen  $T^2$ -Tests weggelassen. Der jeweils höchste Power-Wert der Tests auf  $H_0^A$  bzw.  $H_0^L$  eines Szenariums wurde fett markiert.

Alle Algorithmen erzielen eine Power über 90%, d. h. sie entdecken bei entsprechendem Stichprobenumfang zuverlässig Single-Marker-Assoziation. Die Unterschiede sind teilweise sehr gering und kein Verfahren ist deutlich überlegen. Der  $T^2$ -Test mit Genotypcodierung kann sich für das rezessive und dominante Modell um ca. 3 bis 5% abheben. Für das additive und multiplikative Modell stellt sich das volle Modell von Unphased gut dar. Bei den Tests auf Kopplung können sich auch das `zhao`-Verfahren und der Permutationstest von FBAT platzieren.

Scenarium	Tests of $H_0^A$						Tests of $H_0^L$				
	zhao <sup>a</sup>	mm	hbat_g <sup>b</sup>	hap_pm	gen_pm	full <sup>c</sup>	zhao_l <sup>a</sup>	mm_l	perm_s <sup>d</sup>	hbat_g <sup>b</sup>	full_l <sup>c</sup>
mul1f10	0.962	<b>0.968</b>	<b>0.968</b>	0.964	0.944	<b>0.968</b>	0.962	0.967	0.964	0.967	<b>0.968</b> <sup>e</sup>
mul1f20	0.916	0.934	0.934	0.912	0.908	<b>0.937</b>	0.929	0.938	<b>0.939</b>	0.938	0.938
mul1f22	0.901	0.916	0.916	0.903	0.905	<b>0.919</b>	0.917	<b>0.926</b>	0.922	<b>0.926</b>	<b>0.926</b>
rez1f10	0.963	0.963	0.963	0.962	<b>0.996</b>	0.964	0.963	0.964	<b>0.965</b>	0.964	0.964
rez1f20	0.959	0.953	0.953	0.949	<b>0.987</b>	0.957	<b>0.966</b>	0.963	0.963	0.963	0.963
rez1f22	0.961	0.973	0.973	0.976	<b>0.994</b>	0.986	0.968	0.982	0.981	0.982	<b>0.986</b>
dom1f10	0.926	0.931	0.931	0.926	<b>0.958</b>	0.924	<b>0.926</b>	0.924	0.924	0.924	0.924
dom1f20	0.914	0.926	0.926	0.915	<b>0.972</b>	0.922	<b>0.927</b>	0.925	0.925	0.925	0.926
dom1f22	0.934	0.941	0.941	0.933	<b>0.984</b>	0.939	<b>0.941</b>	0.940	<b>0.941</b>	0.940	0.940
add1f10	0.956	0.957	0.957	0.955	0.943	<b>0.959</b>	0.956	<b>0.959</b>	0.958	<b>0.959</b>	<b>0.959</b>
add1f20	0.947	<b>0.956</b>	0.955	0.944	0.936	0.955	0.954	<b>0.956</b> <sup>h</sup>	0.955	<b>0.956</b> <sup>h</sup>	<b>0.956</b> <sup>h</sup>
add1f22	0.923	0.930	0.930	0.929	0.917	<b>0.932</b>	0.935	0.939	<b>0.940</b>	0.939	0.939
Total <sup>f</sup>	0	1.3 <sup>g</sup>	0.3 <sup>g</sup>	0	6	4.3 <sup>g</sup>	3.5 <sup>g</sup>	1	3.5 <sup>g</sup>	1	3

Tabelle 6.2: KMod 1: Power der Verfahren an der Markerkombination 4

<sup>a</sup> zhao und zhaomax sowie zhao\_l und zhaomax\_l sind hier identisch.

<sup>b</sup> hbat\_b und hbat\_g sowie hbat\_b1 und hbat\_g1 stimmen überein.

<sup>c</sup> Volles und Haupteffektmodell von Unphased erzielen hier gleiche Ergebnisse.

<sup>d</sup> Auch die beiden Haplotyp-Permutationstests liefern gleiche Resultate.

<sup>e</sup> Die besten Werte eines Szenariums sind fett markiert.

<sup>f</sup> Anzahl 1. Plätze unter den Assoziations- bzw. Kopplungstests

<sup>g</sup> mm, hbat\_b und full teilen sich einen Spitzenplatz; dasselbe gilt z. B. für zhao\_l und perm\_s.

<sup>h</sup> mm\_l, hbat\_g1 und full\_l teilen sich drei 1. Plätze

Im Modell KMod 2 steht die Kombination  $M_2M_3$  (das ist Mk 12) im LD zum Risikolocus. Die Mk, welche diese Kombination enthalten, sind in Abb. 6.15 in türkisfarbenen Tönen dargestellt.

#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	
M1																x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	
M2								x	x	x	x	x	x	x	x									x	x	x	x	x	x	x	x	
M3				x	x	x	x					x	x	x	x					x	x	x	x					x	x	x	x	
M4		x	x				x	x			x	x			x	x			x	x			x	x			x	x			x	x
M5	x		x				x				x			x		x			x			x			x			x			x	

Abbildung 6.15: Markerkombinationen Modell KMod 2

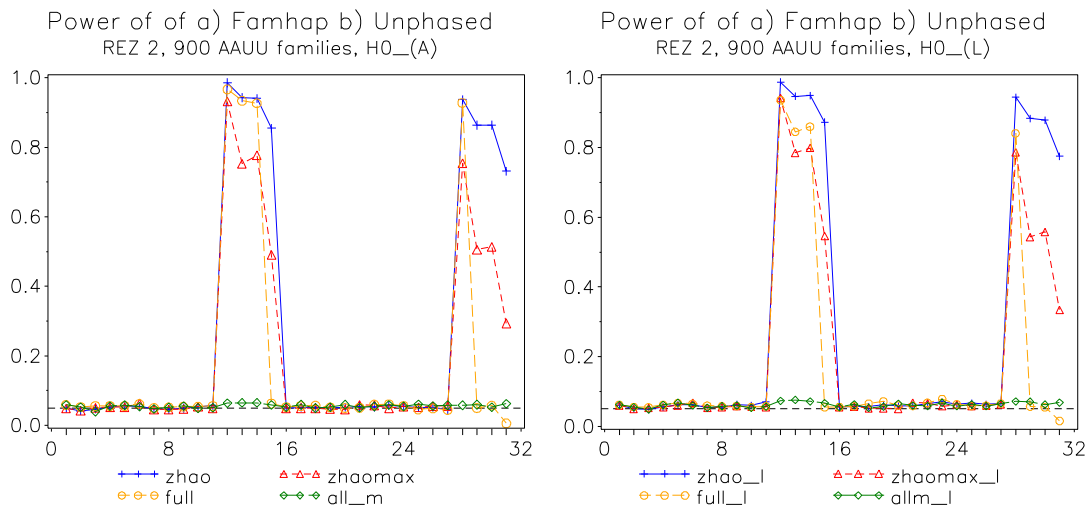


Abbildung 6.16: Power von Famhap und Unphased, rezessives Modell KMod 2, AAUU-Familien; links: Test auf Assoziation, rechts: Test auf Kopplung

Repräsentativ für dieses Modell ist Abb. 6.16. Wiederum sind die Power-Spitzen für die Kombinationen, die  $M_2M_3$  enthalten, deutlich zu sehen. Gegenüber Famhap fällt auf, dass die Power von Unphased (volles Modell, orange) für die Kombinationen 15 und 29 bis 31 einbricht. Das kann für Mk 31 analog oben auf die Option `-rare 0.05` zurückzuführen sein, was im Abschnitt 6.6.2 näher diskutiert wird. Das Allel-Haupteffekt-Modell von Unphased (grün dargestellt) versagt völlig, da es Haplotyp-Effekte nicht modellieren kann.

Im Modell KMod 3 schließlich stehen  $M_2M_3M_4$  - das entspricht Mk 14 - im LD zum Risikolocus. Dementsprechend sind in Abb. 6.17 alle Mk, die diese Kombination enthalten, orange markiert. Neben dem zu erwartenden Maximum bei Mk 14 gibt es nur wenige Mk, an denen die Statistiken überhaupt Assoziation nachweisen können (vgl. Abb. 6.20 rechts). Ansonsten gelten die für KMod 2 getroffenen Aussagen hier ganz analog.

#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	
M1																x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	
M2								x	x	x	x	x	x	x	x									x	x	x	x	x	x	x	x	
M3				x	x	x	x					x	x	x	x					x	x	x	x					x	x	x	x	
M4		x	x				x	x			x	x			x	x			x	x			x	x			x	x			x	x
M5	x		x				x				x			x		x			x			x			x			x			x	

Abbildung 6.17: Markerkombinationen Modell KMod 3



Die Tabellen A.7 und A.8 im Anhang listen die Power für alle Szenarien und Mk 14 auf. Genau für diese Kombination  $M_2M_3M_4$  wurde in KMod 3 Assoziation der Haplotypen mit der Krankheit modelliert. Dementsprechend erreicht die Power der Verfahren für KMod 3 ihr Maximum gerade an dieser Mk. Im Modell KMod 2, wo die Methoden für Mk 12 maximale Power erzielen, besitzen sie für Mk 14 noch gute Power, wie man beispielhaft an Abb. 6.16 erkennen kann. Bei der Konfiguration KMod 1 haben die Tests an Mk 12 ( $M_2M_3$ ) ein (lokales) Maximum, das nur wenig unter dem globalen liegt. Wie die Abbildungen 6.16 und 6.18 belegen, geht die Power für Mk 13 ( $M_2M_3M_5$ ) und Mk 14 von Methode zu Methode sehr unterschiedlich zurück.

In Tab. A.7 und A.8 sind die Spitzenwerte jedes Szenariums fett markiert. Unter den Tests auf  $H_0^A$  dominieren für KMod 1 Multi-Marker-Methoden: **allM** beim multiplikativen, **gen\_pm** beim rezessiven und dominanten und **hap\_pm** beim additiven Vererbungsmodus. Auch der Multi-Marker-Test von FBAT erzielt mit 81 bis 91% eine hohe Power. Wie bereits oben festgestellt wurde, sind diese Verfahren nicht in der Lage, reine Haplotyp-Effekte (KMod 2 und 3) zu entdecken. Bei diesen Konfigurationen teilen sich das volle Modell von Unphased (multiplikativer und rezessiver Erbgang), **zhao** und **hbat\_g** in die Plätze. Dabei fällt auf, dass Famhap vor allem für AAUU-Familien das Verfahren mit der höchsten Power ist, der globale Haplotyp-Test von FBAT für Trios und AA-Familien im dominanten und additiven Fall.

Um die Verteilungen der Werte zu veranschaulichen, wurden Power-Werte über 90% grau und zwischen 80 und 90% hellgrau hinterlegt. Außerdem wird am Fuß der Tabelle gezählt, wie oft die Power in den Intervallen  $(0.9, 1]$ ,  $(0.8, 0.9]$ ,  $(0.7, 0.8]$  liegt bzw.  $\leq 0.7$  ist. Das **zhao**-Verfahren ist gleichmäßig gut: Seine Power übersteigt nicht nur in 19 Fällen 90%, sondern ist auch nie wesentlich unter 70%. Der globale Haplotyp-Test von FBAT folgt mit einigem Abstand. Unphased offenbart beim additiven Vererbungsmodus Schwächen sowohl bei den Tests auf  $H_0^A$  als auch  $H_0^L$ .

Unter den Tests auf Kopplung ist **hbat\_g1** in 1/3 der Fälle der machtvollste. Das Multi-Marker-Verfahren von FBAT (7-mal) und das Haupteffekt-Modell von Unphased (5-mal) teilen sich die Plätze für KMod 1. Wiederum ist das **zhao**-Verfahren für AA- und AAUU-Familien, KMod 2 und 3 (9-mal) das beste. Die Power-Verteilung der Verfahren **hbat\_g1**, **perm\_s** und **zhao\_1** ähneln sich sehr.

Tab. A.9 und A.10 im Anhang enthalten die Power für Markerkombination 30. Sie belegen, dass die Verfahren eine sehr unterschiedliche Gütefunktion haben. Für Mk 30 dominiert der globale asymptotische Haplotyp-Test von FBAT. Die Permutations-basierten  $T^2$ -Tests sowie das Allel-Haupteffekt-Modell von Unphased besitzen gute Power für KMod 1. Die Power des **zhao**-Verfahrens fällt offensichtlich stärker, wenn zur 'optimalen' Kombination weitere Marker hinzukommen. (Bei der Konfiguration KMod 3 z. B. ist  $M_2M_3M_4$  mit der Krankheit assoziiert. Mk 30 aber besteht aus  $M_1M_2M_3M_4$ .) Sowohl das **zhaomax**-Verfahren als auch Unphased mit dem vollen Modell (außer dem dominanten Vererbungsmodus) haben sehr mäßige bis geringe Power. Die Ergebnisse für die Tests auf Kopplung sind ähnlich. Hier teilen sich in Konfiguration KMod 1 die Methoden **mm\_1** für Trios bzw. **allM\_1** für AA- und AAUU-Familien unabhängig vom Vererbungsmodus die Spitzenplätze. Der asymptotische Haplotyp-Test dominiert für die Konfigurationen KMod 2 und 3. Die Verfahren **zhao\_1** bzw. **perm\_s** folgen mit einigem Abstand.

### 6.3.2 Tests auf Kopplung

Für Trios sind Tests auf  $H_0^L$  bzw.  $H_0^A$  identisch (Famhap und Unphased) bzw. ähnlich (FBAT). Bei Familien mit mehreren Geschwistern erzielen Tests auf Kopplung etwas höhere Power. In den Abbildungen 6.16 und 6.18 sind analoge Tests auf Assoziation (links) bzw. auf Kopplung (rechts) in jeweils gleichen Farben dargestellt. So hat die Option **zhao\_1** gegenüber **zhao** (blau) bzw. **zhaomax\_1** vs. **zhaomax** (rot) einen, wenn auch geringen, Vorteil. Allerdings muss man prüfen, ob der Test auf  $H_0^L$  überhaupt gültig ist (s. Absatz 6.2.2). Analoge Tests haben für alle Konfigurationen KMod 1 bis 3 einen sehr ähnlichen 'Verlauf' über alle Mk.

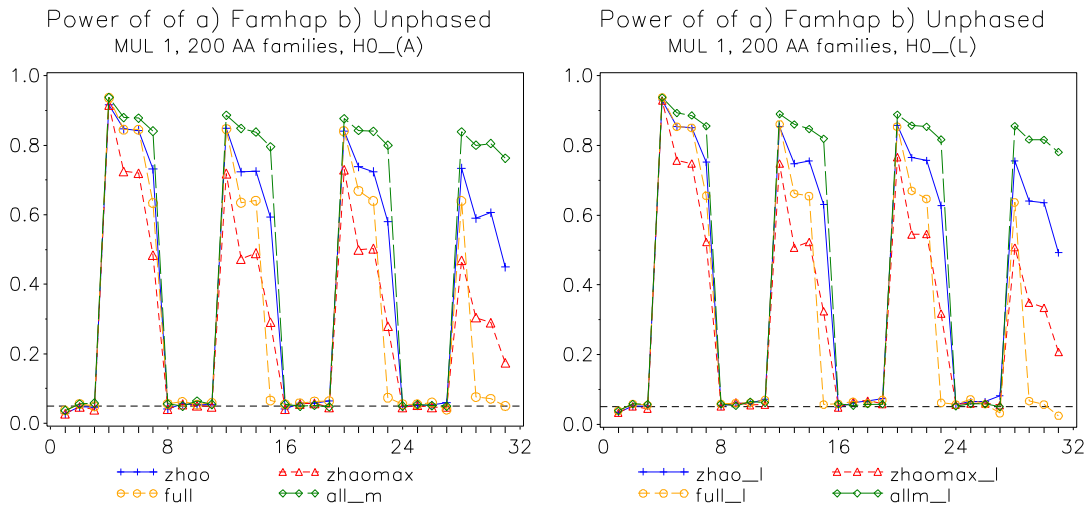


Abbildung 6.18: Power von Famhap und Unphased, multiplikatives Modell KMod 1, AA-Familien; links: Test auf Assoziation, rechts: Test auf Kopplung

### 6.3.3 Power der verschiedenen Statistiken

Zwischen den Algorithmen gibt es teilweise große Powerunterschiede. Die Form der 'Powerverläufe' einzelner Algorithmen ähnelt sich aber. Im Folgenden werden die vier Programme einzeln untersucht:

**FBAT** In Markerkonfiguration KMod 1 sind die Allele eines Markers ( $M_3$ ) mit der Krankheit assoziiert. Multimarker-Tests können Assoziationen mit hoher Power in Kombinationen entdecken, die diesen Marker enthalten. Das spiegelt in Abb. 6.19 links die grüne Linie des Multimarker-Tests `mm` wider und Gleiches gilt auch für `mm_1`. Es fällt auf, dass die Power der Haplotyp-Tests `hbat_b` und `hbat_g` (rot und blau) viel stärker mit zunehmender Zahl der Marker in der Kombination abnimmt. Dabei hat die Mk-spezifische, 'globale' Statistik `hbat_g` i. a. gute Power, währenddessen die Konservativität der Bonferroni-Korrektur für `hbat_b` die Power stark reduziert. Für die Modelle KMod 2 und KMod 3 versagen die Multimarker-Tests völlig (Abb. 6.19 rechts und 6.20).

Die Power der FBAT-Tests auf Kopplung sind in der Regel absteigend so geordnet: `hbat_g1`, `perm_s`, `perm_m` und `hbat_b1` (vgl. Abb. 6.20). Dabei sind einerseits die Werte von `hbat_b1` (blau) und `perm_m` (violett) sowie andererseits `perm_s` (orange) und `hbat_g1` (rot) für viele Mk ähnlich. D. h. die asymptotischen Statistiken konvergieren bereits gut gegen die  $\chi^2$ -Verteilung. Im Absatz 6.6.4 werden die p-Werte der Statistiken noch näher untersucht. Typisch nicht nur für die in Abb. 6.20 dargestellten Szenarien ist auch, dass die Power der drei weniger effizienten Verfahren bei Hinzunahme weiterer Marker zur optimalen Kombination gegenüber `hbat_g1` stärker abfällt.

Schließlich entdeckt man bei genauem Hinsehen in Abb. 6.19, dass die Haplotyp-tests (rot und blau) für Mk 31 keine Werte besitzen. Bei dieser Mk gibt es 32 seltene Haplotypen mit einer Häufigkeit  $< 5\%$ , die durch die Option `-minfreq 0.05` von der Analyse ausgeschlossen werden. Dann kann die  $\chi^2$ -Statistik nicht berechnet werden (vgl. 4.1.7).

**Famhap** Mit der blau dargestellten `zhao`-Option wird durchweg höhere Power als mit `zhaomax` (rot) erzielt (s. Abbildungen 6.16 und 6.18). Außerdem ist der Abfall nach dem Maximum, wenn man zur 'optimalen Kombination' weitere Marker hinzunimmt, steiler. Zur Analyse wird deshalb `zhao` empfohlen, weil es nicht nur die Information des 'besten Haplotyps', sondern auch die der anderen einbezieht. Mit Tests auf Kopplung anstelle von Assoziationstests ist zwischen 1 und 5% Power zu gewinnen.

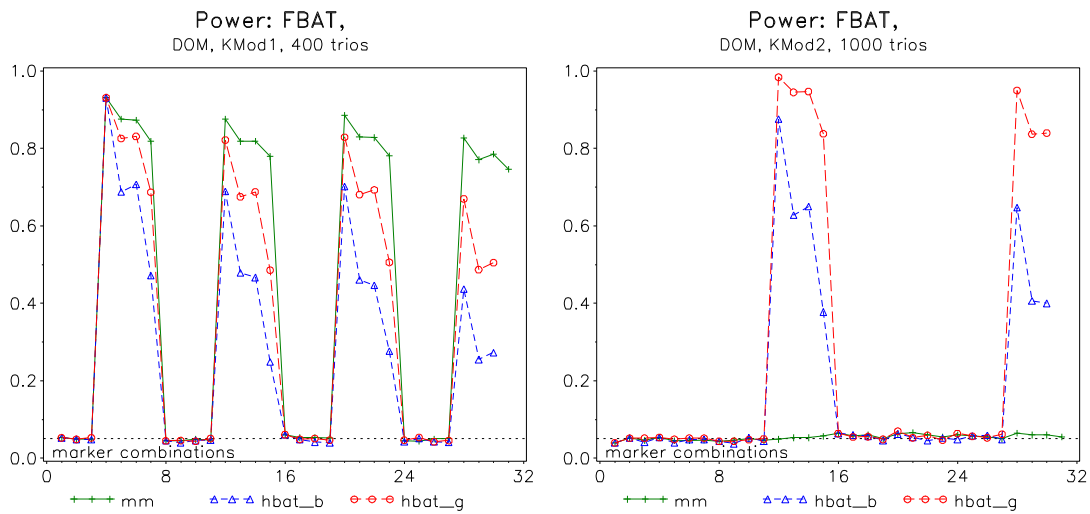


Abbildung 6.19: Power von FBAT: Dominante Modelle KMod 1 und 2, Trios

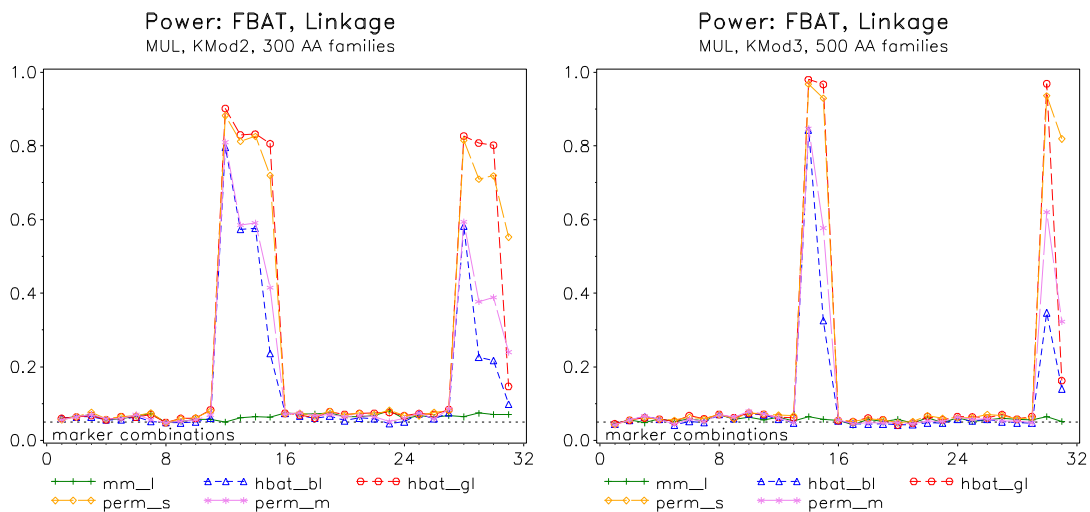


Abbildung 6.20: Power von FBAT, multiplikative Modelle KMod 2/3, AA-Familien, Test auf  $H_0^L$

**Hotel\_Fam** Autoren, wie Rakovski et al. [126], Chapman & Cooper [30] behaupten, dass Multimarker-Methoden den Verfahren überlegen sind, die Phaseninformation nutzen. Hotel\_Fam ist ein solches Multimarker-Verfahren und Fan et al. [57] erhielten am Beispiel einer Asthma-Studie [65] wesentlich kleinere p-Werte als die  $\chi^2$ -Statistik des Zhao-Verfahrens [177]. Diese Überlegenheit wurde für das rezessive und dominante Modell KMod 1 bestätigt, wo ein Marker stark mit der Krankheit assoziiert ist (s. Abb. 6.14). In den Markerkonfigurationen KMod 2 und 3 sind nicht die Allele der Marker, sondern nur Haplotypen mit der Krankheit assoziiert (vgl. 2.2.2). Dann versagen Multimarker-Verfahren völlig.

Für KMod 1 bringt die Haplotyp-Codierung (s. 4.3.2) einen leichten Powergewinn für das additive und multiplikative Vererbungsmodell, während die Genotyp-Codierung für den dominanten bzw. rezessiven Vererbungsgang vorteilhaft ist. Allerdings bewegen sich die Differenzen nur im Bereich von 5 bis 10%. Weil die Power von asymptotischen und Permutationstests nur um ca. 5% abweicht, sind Permutationstests generell zu empfehlen, weil sie den Fehler 1. Art einhalten.

**Unphased** Wüsste man im Voraus, dass Einzelmarker mit dem Risikolocus assoziiert sind (wie für KMod 1), wäre es vorteilhaft, nur Haupteffekte zu modellieren (`-model allelemain`). Die verringerte Anzahl von Freiheitsgraden bewirkt in diesem Fall eine erhöhte Power. Das ist anhand der grünen Linie in Abb. 6.18 gut zu erkennen.

Außerdem soll die geringere Zahl von zu schätzenden Parametern die Rechenzeit verkürzen. Aber bei einer komplexeren Risikostruktur (hier: KMod 2 und 3) kann nur das volle Modell die Assoziation aufdecken. Die Option `-model allelemain` von Unphased ist offensichtlich auch den Multimarker-Methoden zuzurechnen, weil Wechselwirkungen unter den Markern nicht parametrisiert werden (vgl. 4.4.4). Wie man in den Abbildungen 6.16 bzw. 6.18 erkennt, bewegen sich die grün dargestellten Werte nur auf dem  $\alpha$ -Niveau.

Das Setzen einer zu hohen Schranke für seltene Haplotypen (`-rare`) hat bei Unphased noch mehr als bei FBAT negative Auswirkungen (s. 6.6.2).

### 6.3.4 Vergleich der overall-Power über alle Markerkombinationen

#### Verfahren zur Kombination der Mk-spezifischen p-Werte

Wegen der extrem langen Rechenzeiten des Permutationstests von Unphased war Famhap das einzige Programm, das totale p-Werte lieferte (s. 4.4.4). Für die anderen Methoden musste ein Algorithmus gefunden werden, der die p-Werte von  $L$  verschiedenen Mk  $p_l$  ( $l = 1, \dots, L$ ) zu einem 'overall p' kombiniert.

Die einfachste Möglichkeit ist, den minimalen p-Wert  $p_{min}$  aller Mk nach Bonferroni zu korrigieren: Overall-Signifikanz in einer Replikation ergab sich genau dann, wenn  $p_{min}$  zum Niveau  $\alpha/L$  signifikant blieb. Diese Korrektur ist konservativ, besonders wenn - wie hier - die p-Werte korrelieren. Weiterhin ist die Bonferroni-Korrektur nachteilig, wenn aus kleinen oder moderaten Effekten nicht zu kleine p-Werten folgen [110].

Aber die Suche nach besseren Alternativen blieb leider ergebnislos: Fisher's Produktmethode, das Šidák-Verfahren u. a. bringen keine Verbesserung, weil sie wie die Bonferroni-Korrektur unabhängige p-Werte voraussetzen. Da die p-Werte in der Simulationsstudie durch die Formel

$$p = \frac{|\{\text{signifikante Tests}\}|}{|\{\text{Replikationen}\}|} \quad (6.1)$$

berechnet werden, entstehen durch die Diskretisierung gar nicht selten die Werte  $p = 0$  oder  $p = 1$ . Bei vielen Verfahren werden die p-Werte durch inverse Verteilungsfunktionen transformiert, z. B. der Probit-Transformation [43], [172], [73] oder der inversen Gamma-Verteilungsfunktion [171]. Diese Methoden sind nicht anwendbar, weil für Werte 0 bzw. 1 die Inverse einer Verteilungsfunktion nicht definiert ist. Ähnliches gilt für das *rank truncated product* von Dudbridge & Koeleman [52] und die *truncated product method* von Zaykin [172]. Ein  $p = 0$  hat zur Folge, dass das Produkt der p-Werte  $w = 0$  wird. Dann ist aber die 'overall'-Wahrscheinlichkeit nicht berechenbar, weil die Formel für die Verteilungsfunktion den Logarithmus von  $w$  enthält ([172], Formeln (13), (18) im Anhang 1):

$$P(W \leq w) = w \sum_{k=1}^L \binom{L}{k} (1 - \tau)^{L-k} \frac{A(\tau, k, w)^s}{s!} \quad (6.2)$$

$$\text{mit } A(\tau, k, w) = k \ln \tau - \ln w. \quad (6.3)$$

Dabei ist  $\tau$  ein vorgegebener Wert,  $w$  das Produkt der p-Werte  $\leq \tau$  und  $L$  die Gesamtzahl der zusammenzufassenden p-Werte.

Andere Algorithmen (z. B. Monte-Carlo-Methoden) wiederum sind so aufwändig, dass sie sich als Routinekorrektur für tausende Replikationen nicht eignen.

## Vergleich

Wegen der unterschiedlichen Situationen, in denen Tests auf  $H_0^A$  bzw.  $H_0^L$  durchgeführt werden, ist es sinnvoll, die beiden Verfahrensgruppen separat zu betrachten. Einen Vergleich der Assoziationsverfahren auf den ersten Blick ermöglicht Tab. 6.3. (Die ausführlichen Ergebnisse findet man als Tab. A.11 und A.12 im Anhang. Dort ist die Power des jeweils besten Verfahrens jedes Szenariums fett markiert worden.) Die Zusammenfassung in der untenstehenden Tabelle listet auf, in wieviel der 36 Szenarien die Algorithmen die beste Power der 11 Tests auf  $H_0^A$  erzielten. Darunter findet man die Ergebnisse der 9 Tests auf  $H_0^L$ . Zur besseren Übersicht wurden einige Verfahren weggelassen, wie z. B. die beiden asymptotischen Tests von Hotel\_Fam wegen ihrer Antikonservativität für kleine Stichproben:

Tests auf Assoziation	zhao	zhaomax	hbat_b	hbat_g	gen_pm	full
Famhap (total p)	25	0	0	4	6	1
Famhap (Bonferroni)	5	0.5 <sup>a</sup>	2.5 <sup>a</sup>	15	6	7
Tests auf Kopplung	zhao_1	zhaomax_1	perm_m	hbat_b1	hbat_g1	full_1
Famhap (total p)	29	0	0	7	0	0
Famhap (Bonferroni)	2	3.5 <sup>b</sup>	6	2.5 <sup>b</sup>	22	0

Tabelle 6.3: Vergleich Bonferroni-korrigierter min-p-Werte, Anzahl Szenarien, in denen die Verfahren die höchste Power erreichen

<sup>a</sup> Im additiven Modell KMod 1, Trios, teilen sich **zhaomax** und **hbat\_b** einen Spitzenplatz.

<sup>b</sup> Im additiven Modell KMod 1, AAUU-Familien, teilen sich **zhaomax\_1** und **hbat\_b1** einen ersten Platz.

Keine Methode dominiert gleichmäßig über alle Szenarien. Die Fähigkeit von Famhap, einen overall-p-Wert zu berechnen, resultiert in einem Powervorteil. Wird auf die unkorrigierten p-Werte der Markerkombinationen von Famhap dieselbe Bonferroni-Prozedur angewandt, schneidet Famhap wesentlich schlechter ab. So bezieht sich die jeweils erste Zeile auf die totalen p-Werte von Famhap, die zweite auf durch Bonferroni-Korrektur gewonnene Werte. Auch die Unterschieden zu Tab. 6.2 sind teilweise dadurch erklärbar.

Für die Tests beider Hypothesen kann der totale p-Wert der **zhao**-Option von Famhap in etwa 3/4 der Fälle die höchste Power erzielen. Unter den Assoziationstests liegt der  $T^2$ -Permutationstest mit Genotypcodierung in 6 Szenarien (dominantes / rezessives Modell KMod 1) deutlich vorn. Der 'globale' Haplotyp-Test von FBAT erreicht für KMod 3 mit Trios (außer dem rezessiven Modus) dreimal die beste Power. Bei den Tests auf Kopplung gelingt es nur dem globalen Haplotyp-Test von FBAT, die **zhao\_1**-Option sieben mal zu übertreffen. Dagegen würden die FBAT-Haplotyp-tests im Vergleich mit Bonferroni-korrigierten Famhap-Statistiken immerhin in 18 (Tests auf Assoziation) bzw. 31 Fällen (Tests auf Kopplung) den Spitzenplatz einnehmen! Unter den Assoziationstests könnte sich dann auch das volle Modell von Unphased in sieben Fällen an erster Stelle platzieren.

Tab. A.11 und A.12 zeigen nicht nur die Spitzenwerte, sondern auch die Verteilung der Power. Liegen die Werte über 0.85, wurden sie grau unterlegt, liegen sie im Intervall (0.7, 0.85], hellgrau. So sieht man auf den ersten Blick, dass die overall-Power des **zhao**-Verfahrens für alle Szenarien gleichmäßig über 70% ist. Vergleichbar damit ist nur noch **hbat\_g**, dessen Power nicht unter 61% fällt.

Die overall-p-Werte aller 20 Algorithmen wurden für jedes Szenarium in einem Balkendiagramm dargestellt. Dabei befinden sich im oberen Bereich die Tests auf Assoziation (durch ein 'A' markiert), darunter die Tests auf Kopplung ('L'). Ein 'b\_' vor dem Namen des Verfahrens kennzeichnet, dass der overall-Wert durch Bonferroni-Korrektur erhalten wurde. Typisch dafür ist Abb. 6.21.

Für KMod 1 ähneln sich die Werte der FBAT- und Unphased-Optionen sehr. Für AA- bzw. AAUU-Familien differenzieren sich die 20 Statistiken zunehmend. Abgesehen von wenigen Ausnahmen haben die Tests auf  $H_0^L$  einen kleinen Powervorteil gegenüber denen auf  $H_0^A$ , und zwar

Overall power: alpha=0.05, DOM, KMod1, 400 trios

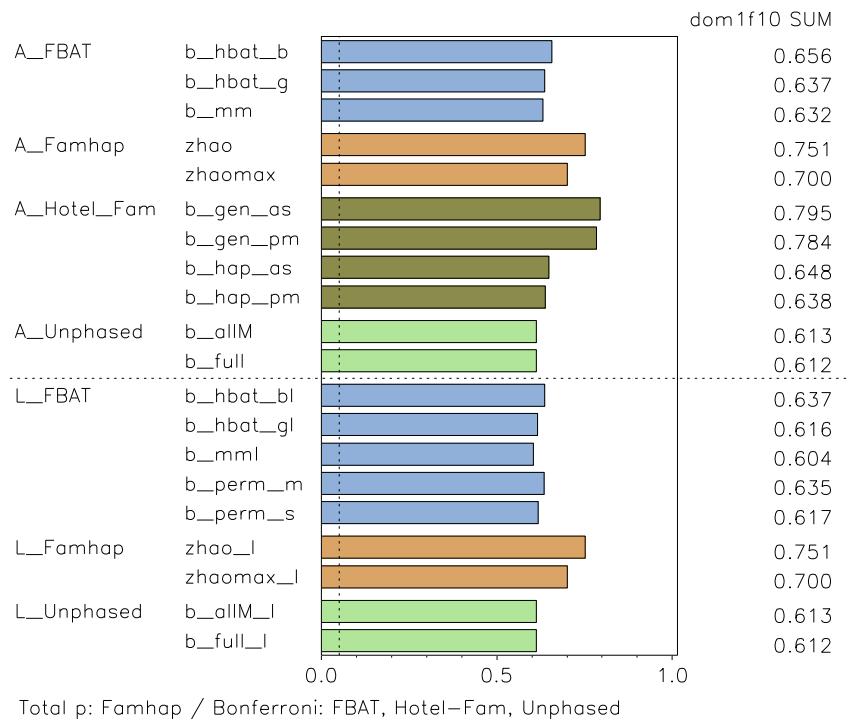


Abbildung 6.21: Powervergleich unter  $H_1$ , KMod 1

zunehmend mit wachsender Kinderzahl. Allerdings ist das u. U. kein wirklicher Gewinn, da dann in Kopplungsregionen Tests auf  $H_0^L$  wegen des inflationären Fehlers 1. Art nicht angewandt werden dürfen. Auch das Versagen der Multimarker-Tests spiegelt sich natürlich deutlich wider (Abb. 6.23 und 6.22). Im Folgenden sind detaillierte Beobachtungen festgehalten:

**FBAT** Normalerweise erzielt der globale Haplotyp-Test die höchste Power. In vielen Szenarien kann der Permutationstest mit  $\chi^2$ -Summenstatistik gut mithalten bzw. ist für KMod 1 teilweise sogar leicht überlegen. Wie schon bei den einzelnen Markerkombinationen fallen der Bonferroni-korrigierte Haplotypstest und der Permutationstest (min  $p$ -Statistik) stark ab. Obwohl der Multimarker-Test für Effekte an einzelnen Markern konstruiert wurde, wird dieser Vorteil im adäquaten Modell KMod 1 nicht sichtbar.

FBAT verwendet standardmäßig die Codierung für ein additives Modell [95]. Trotzdem beobachtet man nicht, dass FBAT in den 9 ADD-Szenarien gegenüber den anderen Vererbungsmodi erhöhte Power hätte. Vielmehr bestätigt sich die Aussage von Horvath [79], dass die additive Codierung auch für andere Modi gute Ergebnisse liefert.

**Famhap** Die in 6.3.1 festgestellte Überlegenheit der **zhao**-Option gegenüber **zhaomax** spiegelt sich auch in der overall-Power wider. Dabei hängt der Powerabfall entscheidend vom Modell ab: Während er bei KMod 1 mit weniger als 5% noch gering ist, erreicht er bei KMod 2 bereits 25-30% und bei KMod 3 über 50%. Offensichtlich gelingt es dem **zhao**-Algorithmus viel besser, den Informationsgehalt komplexerer Familien auszubeuten.

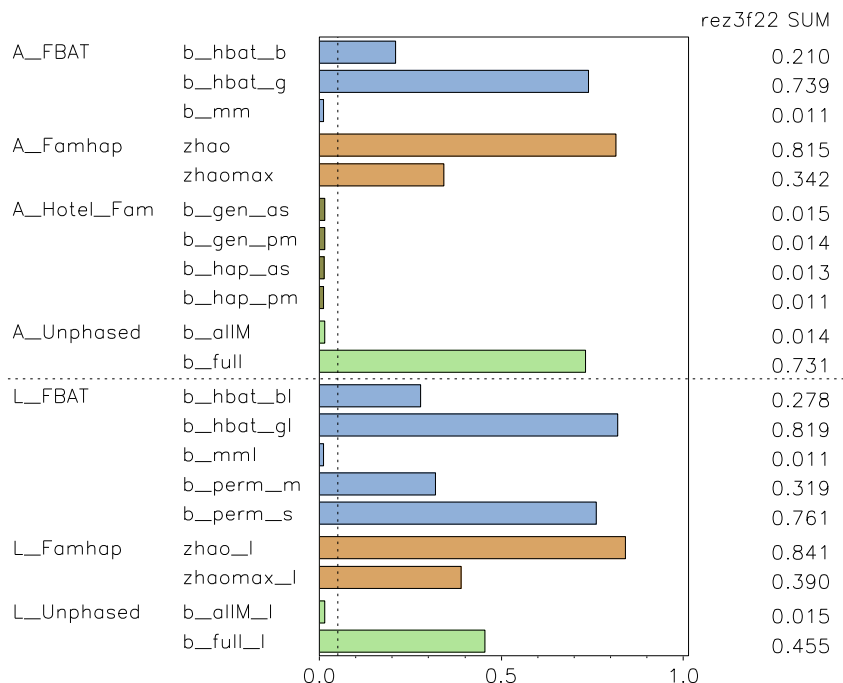
Tests auf  $H_0^L$  und  $H_0^A$  sind für Trios identisch.

**Unphased** Die Identität der Tests auf Kopplung bzw. Assoziation für Trios ist auch klar zu erkennen. Unphased bildet aber gegenüber FBAT und Famhap eine Ausnahme, indem für die Modelle KMod 2 und 3 die Tests auf  $H_0^L$  teilweise deutlich niedrigere Power haben als die auf  $H_0^A$ . Ein Beispiel dafür ist das Szenarium (REZ, KMod 3, AAUU-Familien, vgl. Abb. 6.22). Währenddessen die beste Unphased-Statistik in vielen Szenarien mit den Alternativprogrammen gut mithalten kann, sind die Powereinbrüche in vier Szenarien des additives Modells KMod 2 bzw. 3 besonders deutlich:

MOI	KMod	Fam.-typ	Power
ADD	3	F22	58.0%
ADD	3	F20	36.3%
ADD	2	F22	28.4%
ADD	3	F10	13.8%

Tabelle 6.4: Power von Unphased für ausgewählte Modelle

Overall power: alpha=0.05, REZ, KMod3, 1000 AAUU-families

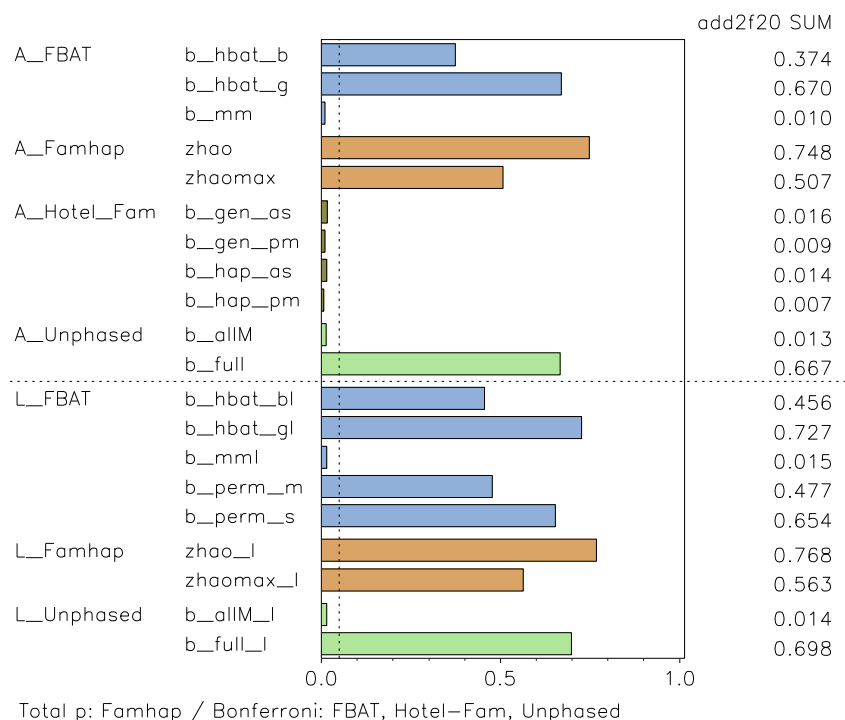


Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Abbildung 6.22: Powervergleich unter  $H_1$ , Modelle KMod 3

**Hotel\_Fam** In sechs Szenarien (KMod 1, REZ und DOM) führt Hotel\_Fam mit dem asymptotischen Test bei Genotyp-Codierung sehr deutlich. Allerdings ist das gerade der Algorithmus, der für 200 Familien in 8 von 9 Szenarien antikonservativ war. Deswegen wird empfohlen, generell auf den nur wenig schwächeren Permutationstest mit Genotyp-Codierung zurückzugreifen. Im additiven und multiplikativen Modus ist regelmäßig folgende Reihenfolge der Verfahren zu beobachten: asymptotischer Test+Haplotypcodierung > Permutationstest+Haplotypcodierung  $\approx$  Permutationstest+Genotypcodierung > asymptotischer Test+Genotypcodierung.

Overall power: alpha=0.05, ADD, KMod2, 400 AA-families



Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Abbildung 6.23: Powervergleich unter  $H_1$ , Modelle KMod 2

### 6.3.5 Ergebnisse für $\alpha = 0.01$

Ein strengeres Fehlerniveau 1. Art bewirkt, dass die Power in Abhängigkeit vom Stichprobenumfang sinkt. Dabei tritt der stärkste Rückgang in Szenarien mit kleinen Stichproben auf. Sonst sind im Vergleich der Verfahren keine prinzipiellen Änderungen zu erwarten.

Tests auf Assoziation	zhao <sup>a</sup>	zhaomax	hbat_b <sup>b</sup>	hbat_g	gen_pm	full
	16	0	0	10	6	4
Tests auf Kopplung	zhao_l	zhaomax_l	perm_s	perm_m	hbat_gl	full_l
	14	1	1	2	17	1

Tabelle 6.5: Overall-Power: Anzahl Szenarien, in denen die Verfahren die höchste Power erreichen

<sup>a</sup> totale p-Werte über alle Mk von Famhap

<sup>b</sup> Bonferroni-korrigierte kleinste p-Werte aller Mk

In Tab. 6.5 ist aufgelistet, in wie viel von 36 Szenarien die jeweiligen Verfahren die höchste



Power erzielen. Auch hierfür findet man ausführliche Power-Tabellen (Tab. A.13 und A.14) im Anhang. Famhap hat seinen deutlichen Vorteil verloren. Vermutlich wirkt sich die Bonferroni-Korrektur für kleinere p-Werte weniger negativ aus. Insbesondere für die Konfigurationen KMod 2 und 3 mit Trios ist die Power der Assoziations- und Kopplungstests des **zhao**-Verfahrens um 5% bis 15% gegenüber **hbat\_g** unterlegen. Trotzdem bleibt die **zhao**-Methode am gleichmäßigsten, denn die Power von **hbat\_g** fällt teilweise bis unter 40%.

Wie Abb. 6.24 zeigt, kann Hotel\_Fam (Genotyp-Codierung) seine führende Rolle für das dominante bzw. rezessive Modell KMod 1 gegenüber den anderen Verfahren noch ausbauen. Die Power nimmt gegenüber  $\alpha = 0.05$  deutlich weniger ab: Für den permutationsbasierten Test und das rezessive Modell z. B. von 93.7% auf 87.0% gegenüber 78.6% / 63.3% für **hbat\_g** und 86.0% / 69.2% für **zhao**.

Es fällt auch auf, dass das volle Modell von Unphased sich unter den Assoziationstests viermal platzieren kann. Beispielhaft dafür steht Abb. 6.25. Ausnahmsweise hat Unphased die höchste Power für einen geringen Stichprobenumfang.

Overall power: alpha=0.01, REZ, KMod1, 2500 trios

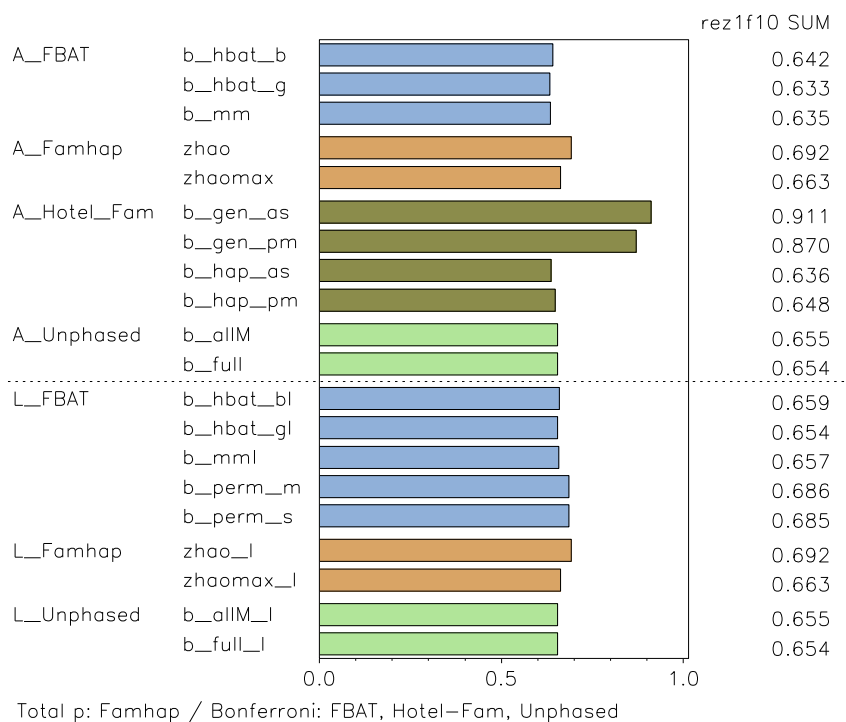


Abbildung 6.24: Powervergleich,  $\alpha = 0.01$ : rezessives Modell KMod 1, Trios

Overall power: alpha=0.01, MUL, KMod1, 200 AAU–families

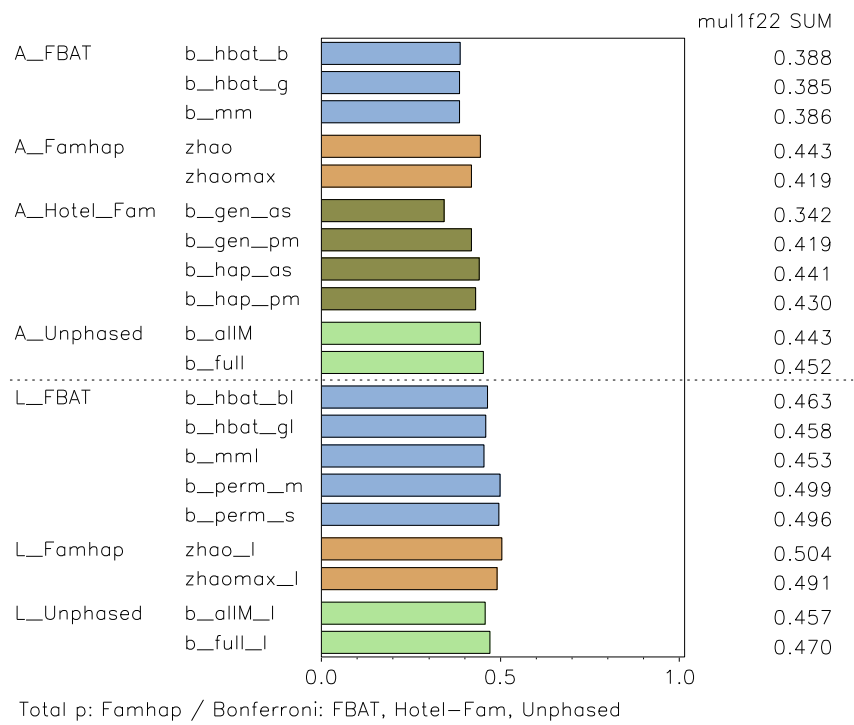


Abbildung 6.25: Powervergleich,  $\alpha = 0.01$ : multiplikatives Modell KMod 1, AAU–Familien

## 6.4 Resultate mit fehlenden Genotypen

Die vier betrachteten Programme behandeln fehlende Genotypen verschieden. FBAT bedingt auf die minimal suffiziente Statistik. Das funktioniert auch noch, wenn der Genotyp eines Elternteils fehlt. Auch Hotel\_Fam kann einen Genotyp-Ausfall in bestimmtem Maße kompensieren. Wenn z. B. der mütterliche Genotyp fehlt, fällt die Familie nur für die mütterliche Statistik  $T_f^2$  aus, nicht aber für die elterliche Statistik  $T_P^2$  [57].

Unphased berechnet, wenn Genotyp-Daten fehlen, die Wahrscheinlichkeiten für alle konsistenten Ergänzungen. Für den Test kann man durch die Option `-missing` über alle diese Beiträge mitteln oder - durch die Option `-certain` - nur die sicheren Haplotyp-Erklärungen verwenden. Für die Simulationsstudie wurde konsequent die erste Option angewandt, weil sie die meiste Information ausschöpft [49]. Famhap kann nur Familien in die Teststatistik einbeziehen, für die an der Markerkombination die Genotypen eines kompletten Trios vorliegen. Mit anderen Worten, fehlende elterliche Genotypen reduzieren automatisch den Stichprobenumfang des Tests. Zur Haplotyp-Frequenzschätzung allerdings werden auch unvollständige Genotyp-Daten herangezogen.

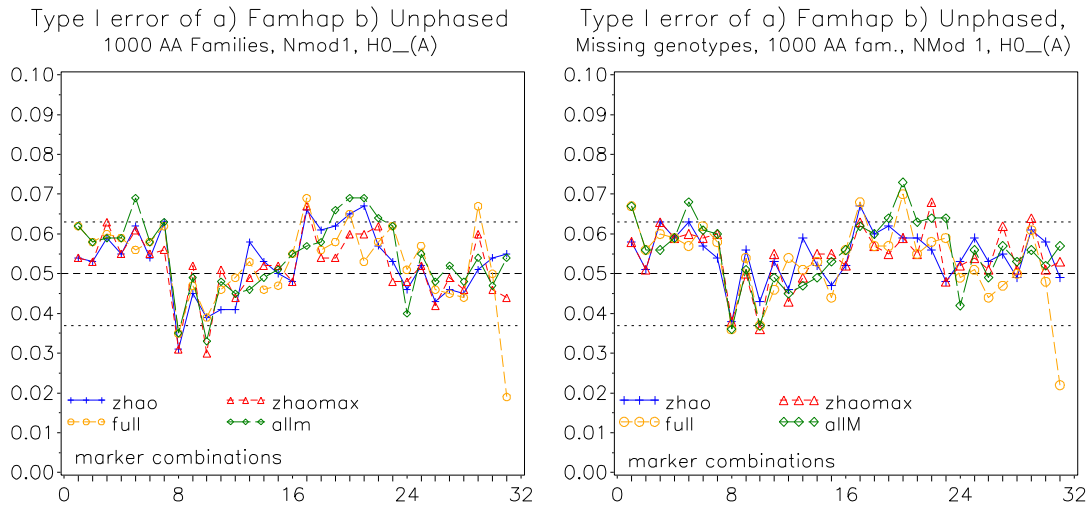


Abbildung 6.26: Fehler 1. Art: Famhap / Unphased, H0\_A, NMod 1, 1000 AA-Familien; links: Analyse mit kompletten, rechts: mit unvollständigen Genotyp-Daten

### 6.4.1 Validität der Tests

Zu den Analysen der vier Programme mit unvollständigen Daten zum Fehlerniveau  $\alpha = 0.05$  kann man folgende allgemeine Aussagen treffen:

- In den meisten Szenarien verändert ein geringer Anteil zufällig fehlender Genotypen das Fehlerverhalten der Algorithmen nur unwesentlich. Oft spiegeln sich Analogien sogar in einem ähnlichen Fehlerverlauf wider, wie in Abb. 6.26 gut zu erkennen ist.
- So wiederholt sich die Antikonservativität des asymptotischen  $T^2$ -Tests mit Genotypcodierung für  $N=200$  und die stärkere Fluktuation der Fehlerwerte des vollen Modells von Unphased. Auch der inflationäre Fehler 1. Art beim Test auf  $H_0^L$  im Modell NMod 2 für AA- und AAUU-Familien bleibt unverändert.
- Der Ausfall von Genotypen bewirkt in ca. der Hälfte der Szenarien eine stärker um den Sollwert schwankende Fehlerrate 1. Art. Dabei liegen meist nur wenige Werte über dem Konfidenzband.

Zu den wenigen Fällen, wo fehlende Genotypen die Fluktuationshöhe verringerten, gehört NMod 3 mit 200 AAUU-Familien (s. Abb. 6.27). Dabei verhalten sich die verschiedenen Algorithmen unterschiedlich: Die relativ hohen Fehlerwerte `gen_as` bei kompletten Daten (links) sind bei unvollständigen Daten (rechts) deutlich verringert. Ebenso liegen die Werte von `gen_pm` im Durchschnitt deutlich niedriger. Dafür hat `hap_as` für einige Mk einen erhöhten Fehler. Die falsch-positiv-Raten von `hap_pm` blieben etwa im gleichen Bereich.

Weiterhin fällt auf, dass sich eine gewisse 'Periodizität' verändert hat: Bei kompletten Daten gibt es (zumindest für `gen_as`) lokale Maxima für die Mk 7, 11, 15, ..., 31, also solche, die  $M_4M_5$  enthalten. Im Bild rechts (unvollständige Daten) zeigen die vier Algorithmen einen Zickzack-Rhythmus. Hier ist der Fehler 1. Art für alle ungeraden Kombinationen mit dem Marker  $M_5$  erhöht. Der Wechsel der Periodizität ist ein weiteres Indiz dafür, dass das rhythmische Verhalten der Hotel\_Fam-Statistiken zufällig und nicht durch rechnerische Artefakte entstanden ist (vgl. 6.2.5).

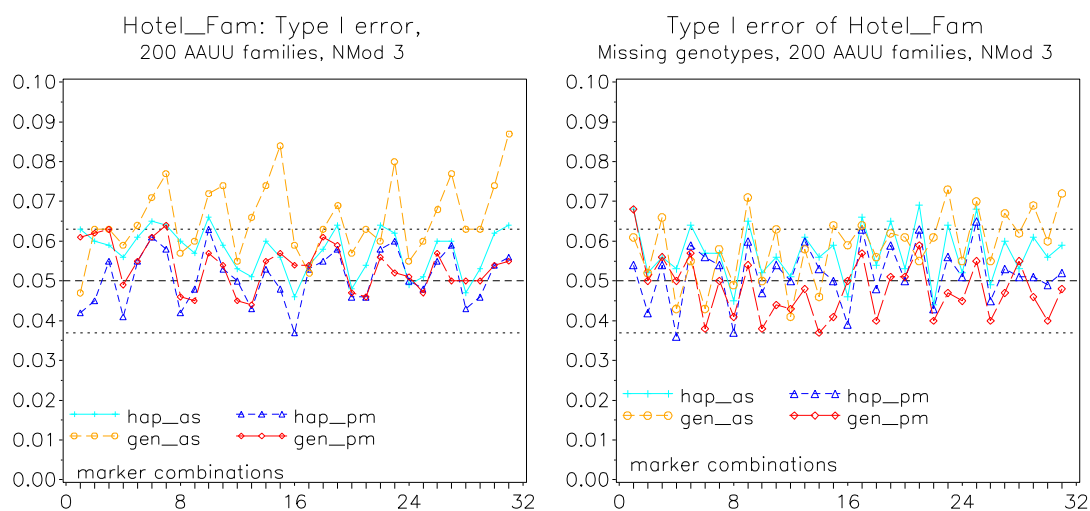


Abbildung 6.27: Fehler 1. Art: Hotel\_Fam, NMod 3, 200 AAUU-Familien; links: Analyse mit kompletten, rechts: mit unvollständigen Genotyp-Daten

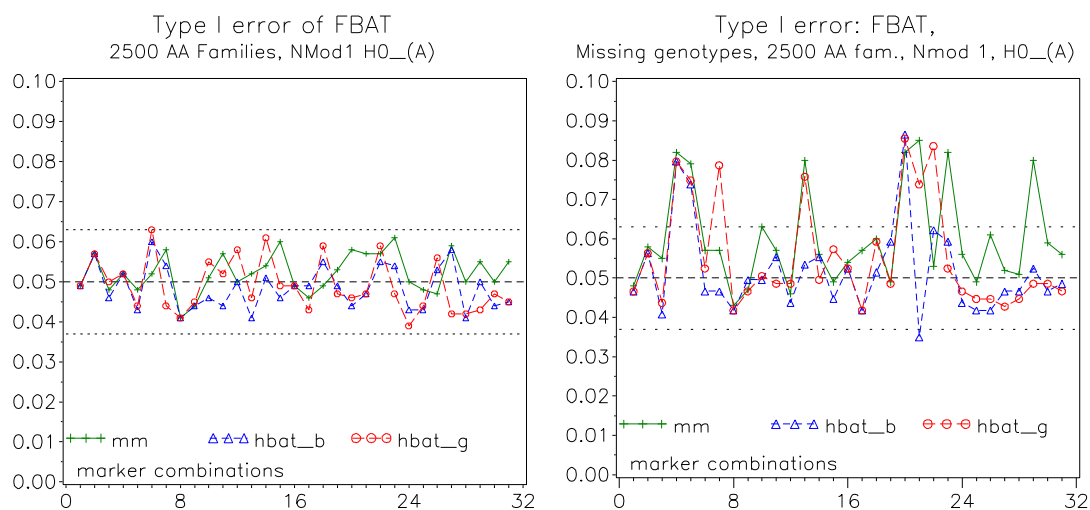


Abbildung 6.28: Fehler 1. Art: FBAT, H0\_A, NMod 1, 2500 AA-Familien; links: Analyse mit kompletten, rechts: mit unvollständigen Genotyp-Daten

Die meisten Erscheinungen in der Analyse mit 'Missings' wurden analog, teilweise in abgeschwächter Form, schon in der Analyse kompletter Daten beobachtet. Eine Ausnahme ist das in Abb. 6.28 dargestellte Szenarium NMod 1, 2500 AA-Familien. Wiederum ist links die Analyse mit kompletten, rechts mit unvollständigen Genotyp-Daten dargestellt. Für etliche Markerkombinationen haben alle FBAT-Algorithmen hier einen überraschend hohen Fehler 1. Art. Das Verhalten der Tests auf Kopplung ist ganz analog. Die Fehlerraten der Verfahren der drei anderen Programme sind in diesem Szenarium unauffällig.

Zusammenfassend kann festgehalten werden, dass ein geringer Anteil ausgefallener Genotypen in der Regel das Verhalten der betrachteten Algorithmen wenig beeinflusst. Im schlimmsten Fall allerdings können fehlende Werte auch falsch positive Befunde begünstigen.

## 6.4.2 Overall-Power in der Analyse mit fehlenden Genotypen

### Powerverlust durch 'Missings'

Fehlende Genotypen bedeuten einen Informationsverlust, der die Power herabsetzt. Deshalb ist es sinnvoll, die Power der verschiedenen Algorithmen für komplette bzw. unvollständige Daten gegenüberzustellen. Basis der Diagramme sind die totalen p-Werte von Famhap und Bonferroni-korrigierte kleinste p-Werte aller Mk der anderen Programme. Weil Multimarker-Tests keine Power besitzen, reine Haplotyp-Effekte zu entdecken, wurden sie für KMod 2 bzw. 3 weggelassen (s. o.). Wiederum wurden alle Analysen zum Niveau  $\alpha = 0.05$  durchgeführt. In Abb. 6.29 bilden die 11 Tests auf  $H_0^A$  die Rubriken der X-Achse. Darüber sind die Power für komplette bzw. unvollständige Daten als Intervall dargestellt und durch den Wert der Differenz ergänzt.

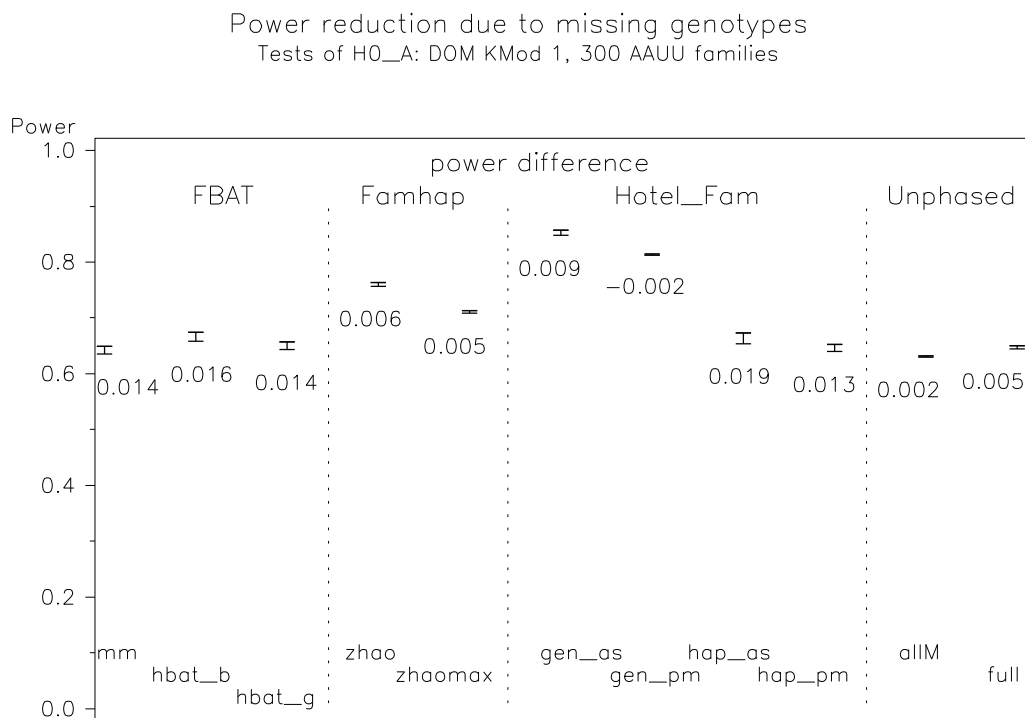


Abbildung 6.29: Powerverlust durch 'Missings': DOM KMod 1, AAUU-Familien

Der Abfall der Power für einzelne Markerkombinationen (hier nicht dargestellt) als auch total über alle Mk ist i. a. relativ klein. In einzelnen Fällen (wie hier der permutationsbasierte  $T^2$ -Test mit Genotypcodierung) nimmt die Power zufällig ein wenig ( $< 1\%$ ) zu.

Unphased hat die geringsten Powerverluste. Für KMod 1 können Hotel\_Fam und Famhap gut

mithalten. Bei den anderen Modellen KMod 2 und 3 verliert Famhap erwartungsgemäß die meiste Power, allerdings bei weitem nicht in der Größenordnung fehlender Genotypen. Dass die Verluste von Modell KMod 1 zu KMod 2 und 3 zunehmen, deutet darauf hin, dass sich der Informationsverlust bei komplexeren Konfigurationen am meisten auswirkt. Irgendwelche Regelmäßigkeiten unter den einzelnen Statistiken sind über die Szenarien nicht erkennbar.

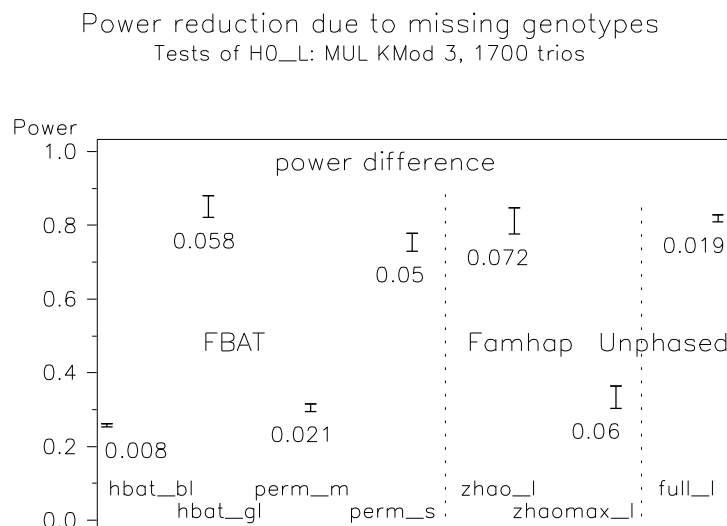


Abbildung 6.30: Powerverluste, Test auf H0\_L: FBAT, Famhap und Unphased, MUL KMod 3, Trios

In Abb 6.30 sind interessante Details zu sehen: Die Verfahren mit hoher Power haben auch deutliche Verluste (zhao: 7.2%, hbat\_gl: 5.8%, perm\_s: 5%). Unphased (volles Modell) wird hier durch seinen geringen Powerabfall zur besten Methode.

Tab. 6.6 zeigt, dass die Power vor allem der FBAT-Verfahren im Durchschnitt für AA- und AAUU-Familien gegenüber Trios weniger abnimmt. Für Famhap und Unphased mit dem vollen Modell gilt dies nicht uneingeschränkt. Genotyp-Informationen zusätzlicher Kinder können also fehlende Genotypen bei Eltern bzw. beim Indexfall zumindest teilweise kompensieren.

Familienstruktur	Tests auf $H_0^A$			Tests auf $H_0^L$			
	hbat_g	zhao	full	hbat_gl	perm_s	zhao_l	full_l
Trios	0.033	0.043	0.011	0.037	0.033	0.042	0.011
AA-Familien	0.025	0.037	0.009	0.019	0.021	0.027	0.014
AAUU-Familien	0.014	0.037	0.006	0.015	0.008	0.038	0.016

Tabelle 6.6: mittlerer Powerverlust durch 'Missings' für ausgewählte Verfahren

### Vergleich der Powerwerte

Knapp und Becker [93] regten 2003 an, die Power des zhao-Verfahrens (Famhap) mit FBAT (als Realisierung des Rabinowitz-Ansatzes [125]) für einen Anteil fehlender Genotypen zu vergleichen. Die vorliegenden Untersuchungen verwirklichen diese Idee. Es interessiert insbesondere, ob sich Famhap hinsichtlich der Overall-Power behaupten kann, wenn durch den Ausfall von Genotypen Familien aus der Analyse herausfallen. Für den hier getesteten 'Missing'-Anteil von 1% kann das bejaht werden. Repräsentativ stehen dafür die neun Tests auf Kopplung im Szenarium (REZ, KMod 1, AA-Familien; Abb. 6.31).

Power reduction due to missing genotypes  
 Tests of  $H_{0\_L}$ : REZ KMod 1, 500 AA families

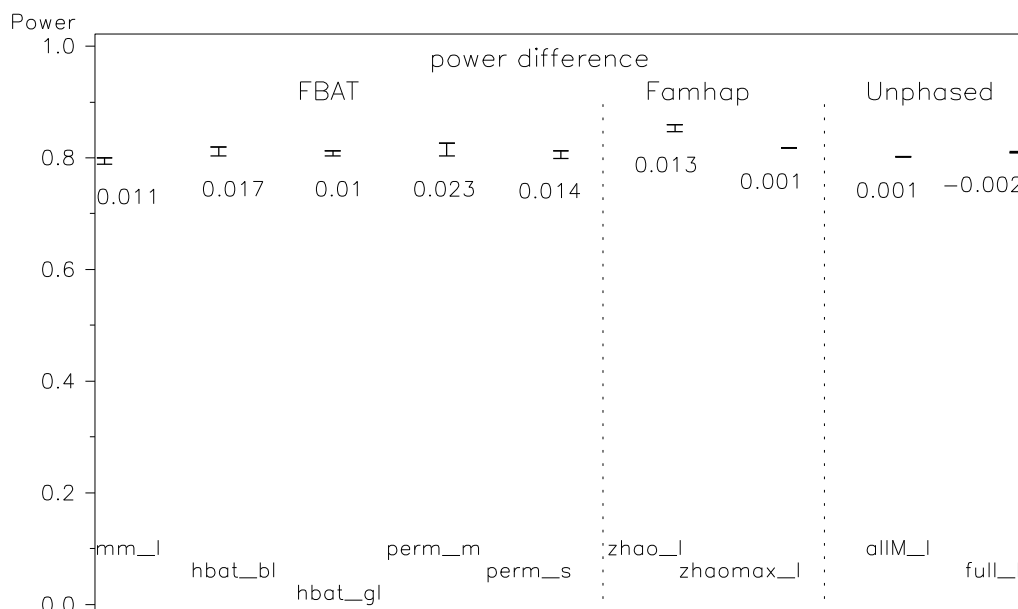


Abbildung 6.31: Analyse mit 'Missings', Test auf  $H_{0\_L}$ , rezessives Modell KMod 1, 500 AA-Familien

Wie schon bei der Analyse mit kompletten Daten dominiert unter den Tests auf Assoziation das **zhao\_l**-Verfahren, auch wenn die Power von 0.86 auf 0.847 zurückgeht. Bei den FBAT-Verfahren reduziert sich die Power in gleicher Größenordnung und das volle Modell von Unphased erzielt gegenüber kompletten Daten sogar einen winzigen Gewinn: von 0.809 auf 0.811.

Die Tab. A.15 und A.16 im Anhang listen die Power für relevante Verfahren auf. Im oberen Bereich findet man Tests auf  $H_0^A$ , im unteren Tests auf  $H_0^L$ . Dabei wurden die beiden asymptotischen  $T^2$ -Tests weggelassen. Die Power der jeweils besten Methode jedes Szenariums wurde fett markiert. Übersteigt die Power 80%, wurde der Wert grau unterlegt, übersteigt sie nur 70%, hellgrau. Am Tabellenende findet man die Anzahl, wie oft ein Verfahren die beste Power erzielt bzw. wie oft die Werte in die Intervalle  $[0, 0.7]$ ,  $(0.7, 0.8]$  bzw.  $(0.8, 1.0]$  fallen.

Obwohl Familien ohne ein genotypisch vollständiges Trio von Famhap nicht zur Teststatistik herangezogen werden können, bleibt die **zhao**-Option dreimal der beste Assoziationstest. Ebenso kann die **gen\_pm**-Statistik in den drei Szenarien im dominanten bzw. rezessiven Modell KMod 1 die Spitzenposition behalten. Unter den Tests auf Kopplung kann das **zhao\_l**-Verfahren seine führende Stellung behaupten.

In Tab. A.16 für die Konfigurationen KMod 2 bzw. 3 wurden die Multi-Marker-Methoden weggelassen. Das **zhao**-Verfahren erzielt nicht nur die meiste Anzahl von Spitzenplätzen, sondern die Power ist gleichmäßig gut. **hbat\_g** und **full** folgen mit einigem Abstand. Es fällt auf, dass das volle Modell von Unphased vor allem in Szenarien mit großen Stichproben (KMod 3, REZ und MUL für Trios) recht kraftvoll ist. Die Verfahren **zhaomax**, **hbat\_b**, **zhaomax\_l**, **perm\_m** sowie **hbat\_bl** sind wegen ihrer geringen Power kaum zu empfehlen.

## 6.5 Ergebnisse unter Stratifikation

Eine erste Analyse prüfte, ob die durch Stratifikation erzeugte Assoziation nachweisbar ist. In einer Fall-Kontroll-Studie mit kranken vs. gesunden Eltern wurden an beiden Markern einzeln die Genotypen mit Armitage's Trendtest getestet. Die hohen Anteile signifikanter Tests von 1000 Replikationen lassen in allen drei Modellen auf Assoziation schließen (s. Tab. 6.7).

Modell	Fam.-typ	Fälle	Kontrollen	$M_1$	$M_2$
S 1	Trios	229	1771	55.9%	83.9%
	AA-Familien	295	1705	85.1%	98.1%
	AAUU-Familien	201	1799	38.3%	64.0%
S 2	Trios	114	886	96.9%	100.0%
	AA-Familien	117	883	98.7%	100.0%
	AAUU-Familien	112	888	93.9%	100.0%
S 3	Trios	229	1771	52.5%	81.1%
	AA-Familien	295	1705	84.0%	97.7%
	AAUU-Familien	201	1799	36.7%	61.8%

Tabelle 6.7: Assoziation durch Stratifikation: Anteil signifikanter Tests

Die sechs Grafiken der Abbildungen 6.32 bis 6.34 unterscheiden sich etwas von denen im Absatz 6.2.1: Auf der horizontalen Achse sind für jedes Verfahren hintereinander die Markerkombinationen  $M_2, M_1$  bzw.  $M_2M_1$  angeordnet. (Für Famhap wurden die 3 Mk durch den Wert der totalen Statistik ergänzt.) Die Fehler 1. Art sind für die drei Familienstrukturen verschiedenfarbig aufgetragen und zur besseren Sichtbarkeit verbunden. Wiederum wurden das Fehlerniveau  $\alpha = 0.05$  und binomiale Konfidenzgrenzen durch gepunktete Linien dargestellt.

FBAT hält die falsch-positiv-Rate sehr gut ein (siehe Abb. 6.32). Mit zwei Ausnahmen liegen alle Werte innerhalb der Konfidenzbereiche. Es ist gut zu erkennen, dass wie oben Tests auf  $H_0^L$  für Familien mit mehreren Kindern einen erhöhten Fehler 1. Art erreichen, währenddessen Trios davon nicht betroffen sind.

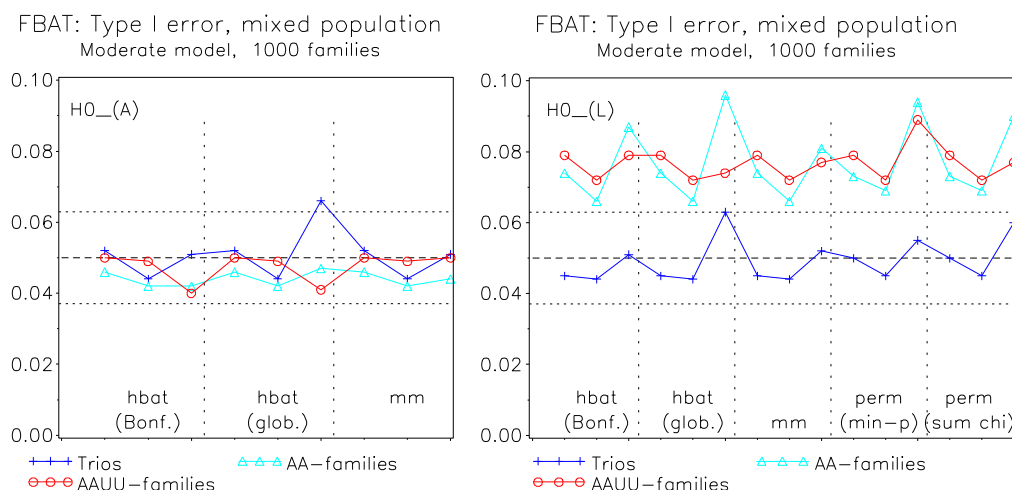


Abbildung 6.32: Fehler 1. Art unter Stratifikation: FBAT, links  $H_{0\_L}$ , rechts  $H_{0\_A}$

Auch die vier Famhap-Algorithmen sind unter Stratifikation gültige Tests. Sowohl für die einzelnen Markerkombinationen als auch den overall-Test wird das Fehlerniveau  $\alpha = 5\%$  gut eingehalten. Tests auf Kopplung geben wie schon in 6.2.2 für AA- und AAUU-Familien einen erhöhten



Anteil falsch positiver Ergebnisse. Als Beispiel seien hier die Grafiken zum extremen Modell S2 abgebildet (s. Abb. 6.33 links). F. Dudbridges Aussage in [51] gilt also nicht für Famhap generell, sondern nur für den Likelihood-Quotiententest mit Fall-Kontroll-Daten.

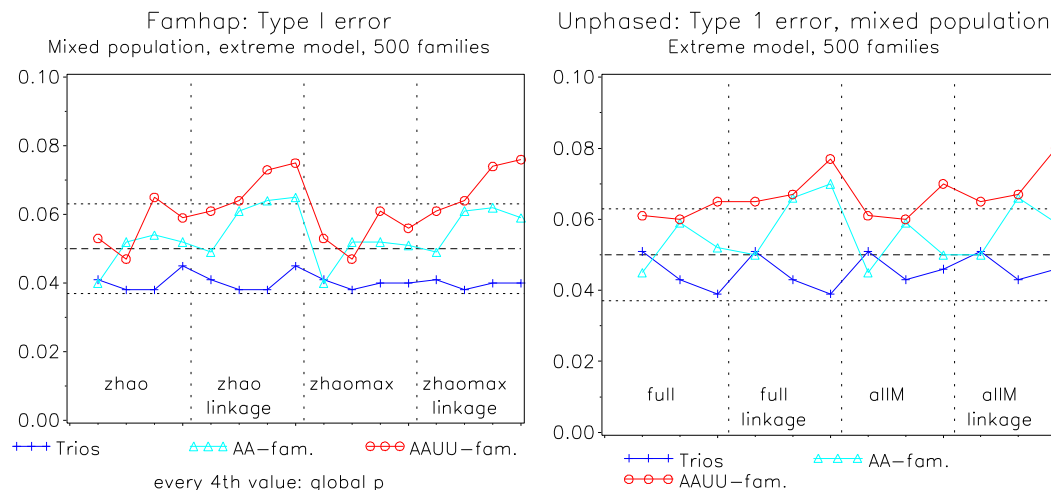


Abbildung 6.33: Fehler 1. Art unter Modell S 2: links Famhap, rechts Unphased

Die betrachteten Algorithmen von Unphased sind robust gegen Populationsstratifikation (s. Abb. 6.33 rechts). Auch Hotel\_Fam hält das Fehlerniveau  $\alpha = 0.05$  gut ein. Zum Vergleich wurde noch das Programm TRANSMIT herangezogen, weil bekannt ist, dass der Score-Test von TRANSMIT mit fehlenden elterlichen Genotypen bei Anwesenheit von Kopplung verzerrt ist (z. B. [109]). Das ist in Abb. 6.34 im ersten Drittel rechts für die  $\chi^2$ -Statistik zu sehen, währenddessen der Test bei Verwendung des Bootstrap-Schätzers für die Varianz gültig ist.

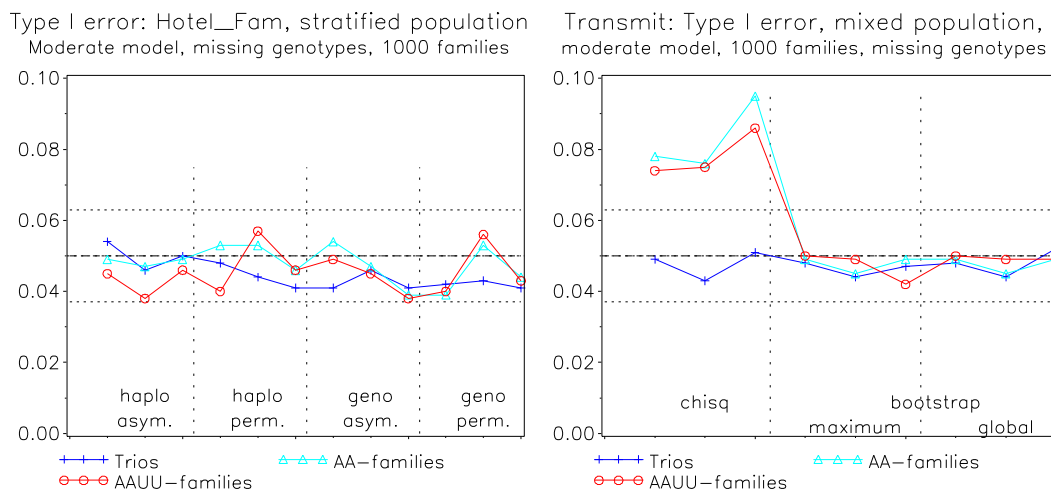


Abbildung 6.34: Fehler 1. Art unter Modell S 3: links Hotel\_Fam, rechts Transmit

## 6.6 Weiterführende Analysen

### 6.6.1 FBAT: Anzahl Permutationen

Für alle Permutationstests waren  $R = 4000$  Replikationen festgelegt worden. Wie bereits in 4.1.4 beschrieben, schöpft FBAT diese Maximalzahl nicht aus, wenn 100 Monte-Carlo-Samples erzeugt wurden, deren Statistik jene der beobachteten Stichprobe entweder überschreiten bzw. andernfalls unterschreiten. Um sich über den tatsächlichen Aufwand ein Bild zu verschaffen, wurden exemplarisch für die 27 Szenarien der Analyse unter  $H_0$  mit kompletten Familiendaten Histogramme erzeugt. Basis dafür waren je 31000 ( $Mk \times$  Replikationen) Permutations-Anzahlen.

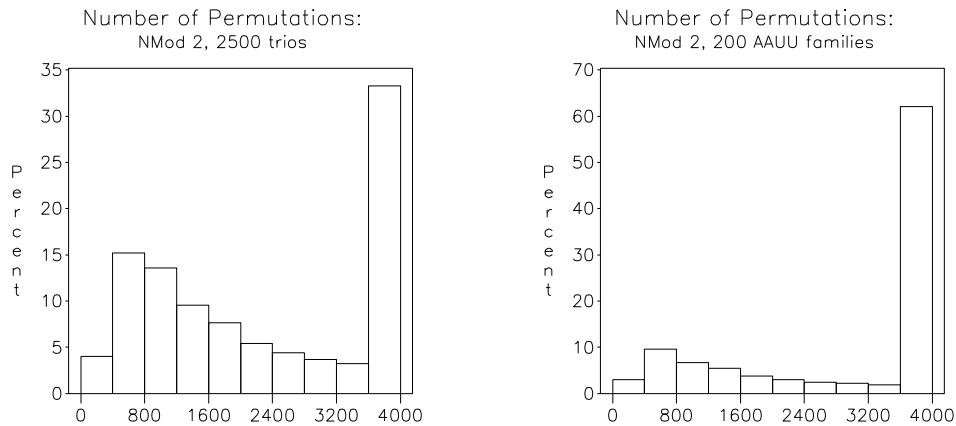


Abbildung 6.35: FBAT: Anzahl benötigter Permutationen unter  $H_1$ , NMod 2; links: 2500 Trios, rechts: 200 AAUU-Familien

Abb. 6.35 ist dafür typisch: In allen Szenarien ist die Klasse  $(3600, 4000]$  die mit der größten Häufigkeit. Nach einem lokalen Maximum in Klasse  $(400, 800]$  fallen die Häufigkeiten monoton bis 3600. Die Grafik links repräsentiert die Modelle NMod 1 und NMod 3 und die Konfiguration NMod 2 mit Trios. Hier wird nur in ca.  $1/3$  aller Fälle die volle Zahl von 4000 Permutationen benötigt. Im Szenarium NMod 2, 2500 Trios z. B. liegt das Minimum bei 274, der Median bei 2000 und der Mittelwert bei 2289. Das rechte Diagramm ist exemplarisch für die Konfiguration NMod 2 mit AA- und AAUU-Familien. Hier braucht knapp  $2/3$  der Läufe die volle Zahl. Im Szenarium NMod 2 mit 200 AAUU-Familien z. B. liegt der Median bei 4000 (Mittelwert: 3012). Immerhin genügen hier in einer Replikation bereits 262 Monte-Carlo-Samples.

Für die Modelle unter der Alternativhypothese gelten ähnliche Aussagen: Die Verhältnisse bewegen sich zwischen den in Abb. 6.35 dargestellten Extremen. Die drei Markerkonfigurationen sind allerdings deutlich verschieden: Während im Modell KMod 1 knapp  $2/3$  aller Fälle in die höchste Klasse fallen, sind es im Modell KMod 2 weniger als 50% und im Modell KMod 3 weniger als 40%. Die Häufigkeiten verteilen sich dementsprechend analog auf die übrigen Klassen.

## 6.6.2 Die Optionen 'minfreq' und 'rare'

Die FBAT-Option `minfreq 0.05` bewirkt, dass bei einer hohen Haplotypanzahl solche mit einer Häufigkeit  $< 5\%$  ausgemustert werden. Das führt dazu, dass im Extremfall keine Statistik berechnet werden kann, weil nur noch ein bzw. gar kein Haplotyp übrig bleibt. Analog kann die Option `-rare 0.05` bei Unphased nach sich ziehen, dass alle Haplotypen als selten klassifiziert werden und für sie ein gemeinsamer Parameter modelliert wird. Dann liefert die Statistik einen p-Wert von 1. Diese Erscheinungen wurden bereits oben im Absatz 6.2 beschrieben.

Die relativ hohe Schranke 5% entspricht der CDCV-Hypothese. Sie ist dann von Vorteil, wenn es wenige Haplotypen in der Markerkombination gibt, die assoziiert sind. Der Wegfall seltener Haplotypen reduziert die Zahl der Freiheitsgrade und erhöht damit die Power. Diese Festlegung nach dem Handbuch [95] kann in bestimmten Situationen nachteilig sein: Im Zusammenhang mit der Wahl und Analyse von *tagSNPs* fanden deBakker et al. [42] heraus, dass die Überprüfung aller Haplotypen auf Assoziation anstelle nur einiger als Stellvertreter die Power erhöht, *seltene* kausale Allele zu finden. Dafür sinkt die Power, mit der man häufigere kausale Allele entdeckt. Das motivierte dazu, die Analyse für folgende Szenarien mit der niedrigeren Schranke (`minfreq 0.01` bzw. `-rare 0.01`) zu wiederholen: ADD, KMod 3; DOM, KMod 3; MUL, KMod 2 und REZ, KMod 1, jeweils mit AA-Familien. Im Fokus stand dabei die Frage, ob das bzw. das Weglassen der Optionen o. g. negative Erscheinungen beseitigt bzw. einen Powergewinn bringt.

Die Auswertung erfolgte analog Absatz 6.3, aber die Diagramme unterscheiden sich in zwei Punkten: Zur Erhöhung der Übersicht wurden auf der X-Achse nur die Markerkombinationen aufgetragen, für die veränderte Optionen auch die Werte der Statistiken veränderten. Weiterhin wurde nur eine Methode je Diagramm dargestellt und darin die verschiedenen Optionen durch andere Farben. An vielen Stellen fällt die grüne Linie (Schranke 1%) mit der roten (ohne Schranke) zusammen. Wie erwartet weichen die Ergebnisse der Analysen teilweise nur in Markerkombinationen mit drei und mehr Markern ab.

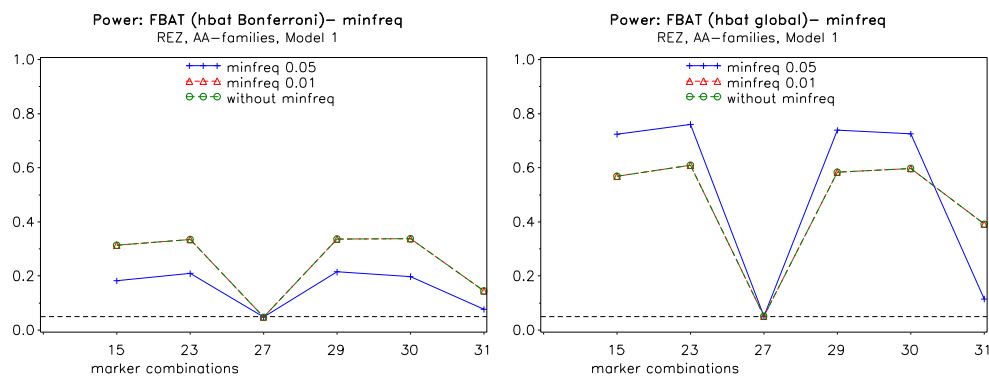


Abbildung 6.36: FBAT, minfreq-Optionen, REZ, KMod 1, AA-Familien

### FBAT

Abb. 6.36 ist typisch für die Resultate: Das Senken bzw. Weglassen der Schranke für seltene Haplotypen ist ambivalent: Es bringt der Bonferroni-korrigierten Statistik (links) Powergewinne von ca. 5 - 15% für mehrere Markerkombinationen. Umgekehrt verringert `minfreq 0.01` bzw. das Weglassen dieser Option in unserem Beispiel die Power des globalen Tests (rechts) für Mk 15, 23, 29 und 30, aber erhöht sie für Mk 31. Die Identität der roten und grünen Linien lässt erkennen, dass eine Analyse mit 1% bzw. ohne diese Schranke dieselben Werte lieferte. Für die entsprechenden Befehle mit Linkage-Option gilt dasselbe.

Analysen im additiven Modell verhalten sich ähnlich. Da sich im dominanten Fall die drei Optionen nur sehr gering unterscheiden, d. h. Differenzen  $\leq 2.3\%$ , wurde auf eine grafische Darstellung

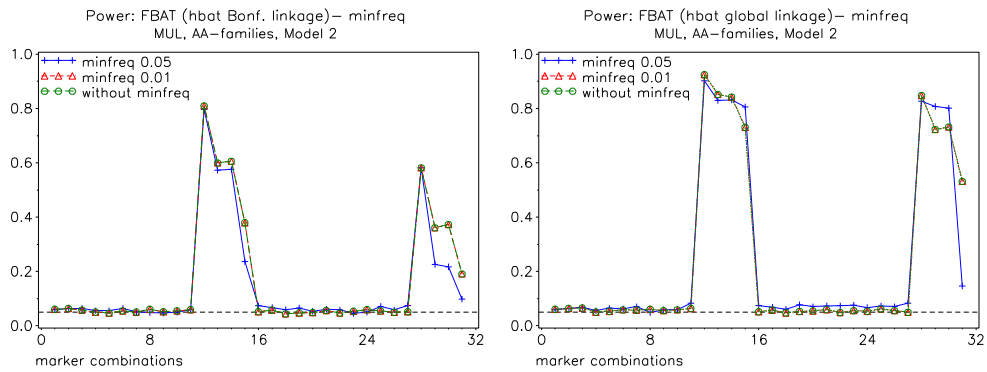


Abbildung 6.37: FBAT, minfreq-Optionen, MUL, KMod 2, AA-Familien

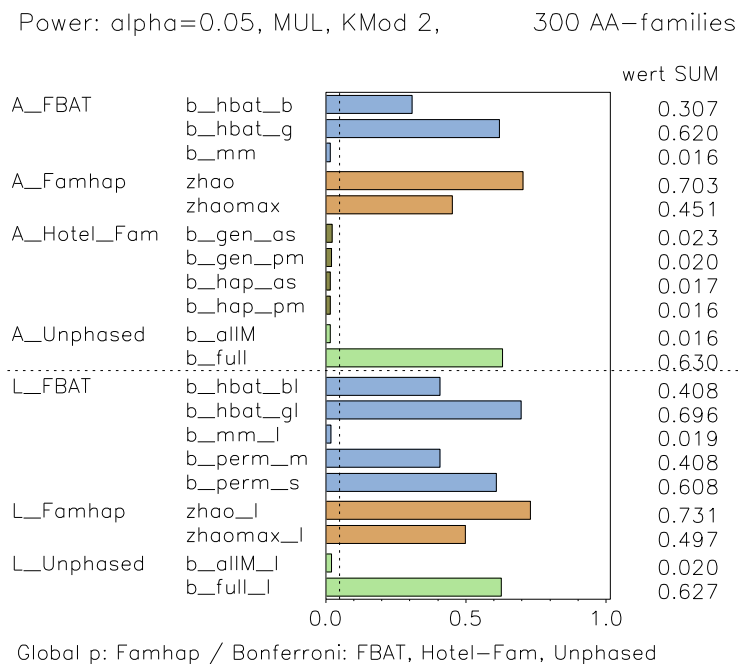


Abbildung 6.38: Overall-Vergleich: FBAT mit Option -minfreq 0.01

verzichtet. Im multiplikativen Modus gibt es für viele Mk vernachlässigbare Differenzen und nur in den Kombinationen mit 4 und 5 Markern sieht man deutliche Abweichungen (s. Abb. 6.37). Die höhere Schranke 5% bringt hier nur für die globale Statistik an Mk 15, 29 und 30 Vorteile. Für dieselbe Statistik sind 1% bzw. das Weglassen der Schranke vorteilhaft für Mk 12, 13, 14, 28 und 31, aber nachteilig für MK 15, 29 und 30. Für `hbat_b` und `hbat_bl` bringt eine niedrigere Schranke immer einen Gewinn (vgl. Abb. 6.37 und 6.36 links).

Die Powergewinne verändern die Aussagen kaum, die oben im Abschnitt 6.3.4 getroffen wurden: Da für die overall-Betrachtung Bonferroni-korrigierte Werte von FBAT einbezogen wurden, sind die Maxima wesentlich. Einerseits wurden mit den neuen Optionen die Spitzenwerte des globalen Haplotyp-Tests nicht übertroffen. Andererseits können die Gewinne die eher mäßige Power des Haplotyp-Tests mit Bonferroni-Korrektur nicht kompensieren. Für das Szenarium MUL, KMod 2, AA-Familien ist in Abb. 6.38 der overall-Vergleich mit den besten Werten dargestellt. FBAT (globaler Test `hbat -minfreq 0.01`) hat zwar gegenüber der `zhao`-Option von Famhap 0.5% (`hbat_g`) bzw. 2.9% (`hbat_gl`) aufgeholt, ist aber von der Spitzenposition noch ein Stück entfernt.

Der Hauptvorteil einer niedrigeren Haplotypschranke ist, dass FBAT auch für Markerkombi-

nationen mit 5 und mehr Markern die Teststatistik berechnen kann. Umgekehrt ist für `minfreq 0.01` gegenüber einer Analyse ohne diese Option die Gefahr geringer, dass zu viele Freiheitsgrade die Power der Tests untergraben.

## Unphased

Im Unphased-Manual [50] wird argumentiert, dass durch das Zusammenlegen seltener Haplotypen mit der Option `-rare` die Laufzeit verkürzt wird. Die vorliegenden Analysen können dies nicht belegen. Weil die Bedingungen auf den Knoten des Cluster-Rechners schwer zu kontrollieren sind, sind nur solche Läufe repräsentativ, bei denen jedem Job ein Knoten separat zur Verfügung stand. Unter  $H_0$  mit 200 AA-Familien im Modell NMod 2 und 3 waren die Analysen ohne Option `-rare` sogar um weniger als 1% schneller als mit der Option `-rare 0.01`. In anderen Szenarien ist die Situation uneinheitlich. Von einem relevanten Rechenzeitvorteil durch das Setzen einer Schranke für seltene Haplotypen kann aber nicht die Rede sein.

Im Programm Unphased ist der Wert 5% der Option `-rare` zu hoch gewählt: Denn ein Absenken oder Weglassen der Schranke brachte in jedem der Szenarien deutliche Vorteile. Beim Test auf Kopplung gewinnt die Power für Mk 31 noch in drei von vier Fällen, wenn man die Option `-rare 0.01` ganz weglässt. Exemplarisch sei das an zwei Diagrammen in Abb. 6.39 dargestellt.

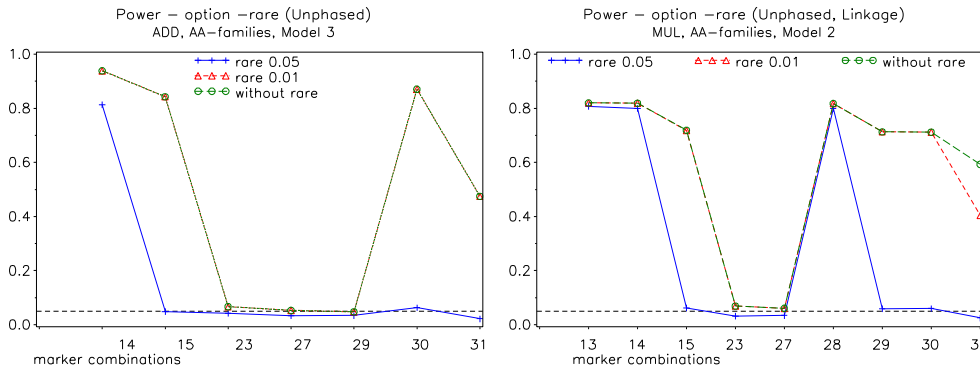


Abbildung 6.39: Unphased, rare-Optionen

## 6.6.3 Vergleich verschiedener Familienstrukturen

### Vorbetrachtungen

Die Effizienz von Haplotyp-Frequenzschätzung hatte D. Schaid [132] durch Vergleich der Likelihoods bzw. mittels Simulation untersucht. Dabei ist die Schätzung für Trios gegenüber Fall-Kontroll-Daten hoch effizient. Die Hinzunahme der Genotypinformation weiterer Geschwister hingegen bringt nur wenig Gewinn. Der Effizienz von Assoziationstests für verschiedene Familienstrukturen sind Arbeiten von Risch & Teng [129] sowie Fingerlin et al. [59] gewidmet. Obwohl die Situationen, die diese Autoren betrachten, recht unterschiedlich sind, stimmen sie darin überein, dass zusätzliche kranke Geschwister die Power substantiell erhöhen. Die Genotypen gesunder Geschwister sind bei niedriger Risikoalleletrate nützlich. Auch Becker & Knapp [20] hatten beobachtet, dass Haplotypmethoden mit Kernfamilien mit mehreren kranken Kindern gegenüber Trios höhere Power erzielen.

Von zwölf untersuchten Modellen (4 Vererbungsmodi  $\times$  3 Markerkonfigurationen, s. 2.2.2) gibt es immerhin acht, für welche die Stichprobenumfänge für AA- bzw. AAUU-Familien (F22 bzw. F20) übereinstimmen (vgl. Tab. 2.5). Deshalb lag es nahe, für diese acht Modelle die Power repräsentativer Statistiken für beide Familienstrukturen zu vergleichen. Eine Übersicht vermittelt Tabelle 6.8. Dort ist der im Paarvergleich F22 - F20 jeweils höhere Wert fett hervorgehoben. Das Bild ist uneinheitlich: Teilweise erzielen AA-, in anderen Szenarien AAUU-Familien die höhere

Power. So sollte genauer untersucht werden, ob größere Familien mit zwei kranken Kindern und ggf. gesunden Geschwistern gegenüber Trios einen wesentlichen Effizienzvorteil bringen.

MOI KMod <sup>a</sup> s-size <sup>b</sup> Statistik	ADD 1		DOM 1		ADD 2		MUL 2		ADD 3	
	300		300		400		300		500	
	F20	F22	F20	F22	F20	F22	F20	F22	F20	F22
hbat_g	<b>0.73</b>	0.68	0.62	<b>0.66</b>	<b>0.67</b>	0.64	<b>0.62</b>	0.61	<b>0.71</b>	0.67
zhao	<b>0.81</b>	0.74	0.73	<b>0.76</b>	<b>0.75</b>	0.71	0.7	<b>0.73</b>	<b>0.74</b>	0.72
full	<b>0.74</b>	0.68	0.61	<b>0.65</b>	<b>0.67</b>	0.28	<b>0.63</b>	0.63	0.36	<b>0.58</b>
genpm	<b>0.71</b>	0.64	0.77	<b>0.81</b>	0.01	0.01	0.02	0.03	0.02	0.01

<sup>a</sup> Vererbungsmodus, Markerkonfiguration

<sup>b</sup> Stichprobenumfang

Tabelle 6.8: Powervergleich für AA- und AAUU-Familien für einige Szenarien

Für sechs Modelle (ADD REZ und MUL: KMod 1; ADD, DOM: KMod 2 und MUL: KMod 3) wurde folgende Analyse für Trios, AA- oder AAUU-Familien durchgeführt: Aus Familien-Stichproben mit zwei kranken und zwei gesunden Kindern (F22) nach 2.2.2 wurden je zwei neue Samples mit der gleichen Anzahl Kernfamilien erzeugt. Durch Weglassen der zwei gesunden Nachkommen entstanden AA-Familien (F20) und daraus durch zufälliges Fallenlassen eines kranken Geschwisters Trios (F10). Diese neuen Stichproben wurden mit allen Programmen analysiert und jeweils zusammengehörige Resultate miteinander verglichen. Für die Konfigurationen KMod 2 und 3 konnten die Multimarker-Verfahren weggelassen werden, weil sie dann sowieso keine Power haben (s. 6.3.3). Mit dieser Versuchsanordnung wird also der Fall geprüft, wo aus Kernfamilien mit zwei kranken und zwei gesunden Kindern nur ein Trio bzw. die ASPs mit ihren Eltern in die Studie einbezogen werden.

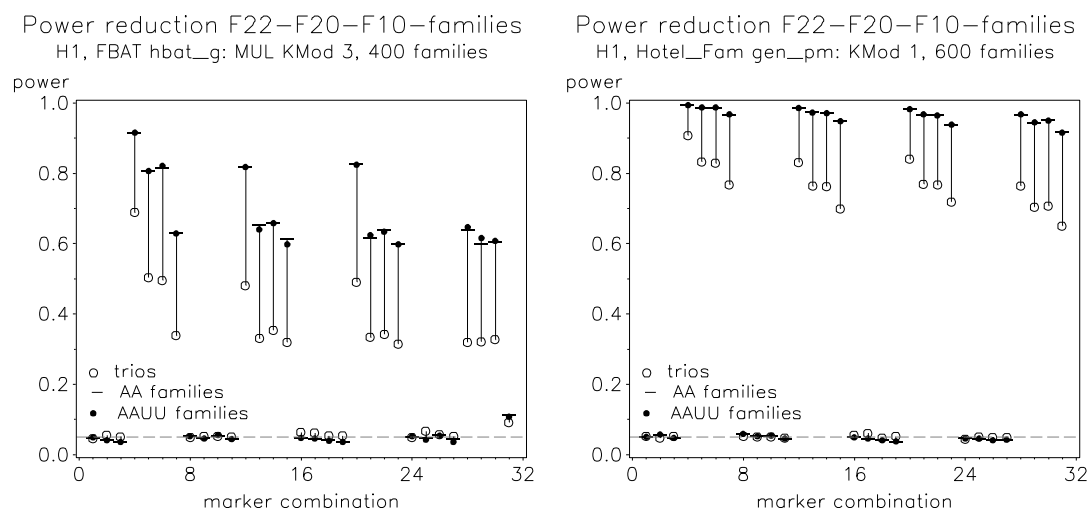


Abbildung 6.40: Vergleich der Familienstrukturen F22 - F20 und F10, links: FBAT: MUL KMod 3, rechts: Hotel\_Fam: REZ KMod 1

## Power für die Markerkombinationen

Abb. 6.41 ist repräsentativ für diese Analyse: Die 31 Markerkombinationen bilden wiederum die x-Achse. Die Power für F22-, F20- und F10-Familien wurden separat für jede Statistik als Kursdiagramm dargestellt. Dabei kennzeichnen leere Kreise Trios, gefüllte Kreise AAUU-Familien und Striche AA-Familien.

Folgende Gemeinsamkeiten fallen auf:

- In der Regel erreichen alle Verfahren mit AA-Familien bzw. AAUU-Familien die höchste Power und meist sind die Werte dafür ähnlich.
- Dagegen fällt die Power für Trios deutlich (teilweise über 50%) ab.
- Tests auf Kopplung und Tests auf Assoziation verhalten sich sehr analog.
- Im Fall des 'Single-Marker-Modells' KMod 1 erzielt F20 an manchen Markerkombinationen eine leicht höhere Power gegenüber F22. Dabei sind die Differenzen für Famhap deutlich höher als für FBAT, und zwar besonders für Kombinationen aus 4 und 5 Markern (Mk 15, 27, 29, 30, 31, Abb. 6.41 links). Mit anderen Worten: Die Hinzunahme zweier weiterer gesunder Geschwister verringert sogar die Power gegenüber der Analyse von AA-Familien.
- Dagegen gibt es für die Markerkonfigurationen KMod 2 und 3 für F22 ein deutliches Plus gegenüber F20 (s. Abb. 6.41 rechts). D. h., hier kann sich die Genotypisierung weiterer gesunder Kinder lohnen.
- Bei Hotel\_Fam (nur für KMod 1) liegen die Power-Werte für F20 und F22 fast gleich. Auch der Abfall für Trios ist bei weitem nicht so groß wie bei den anderen Verfahren (Abb. 6.40, rechts).

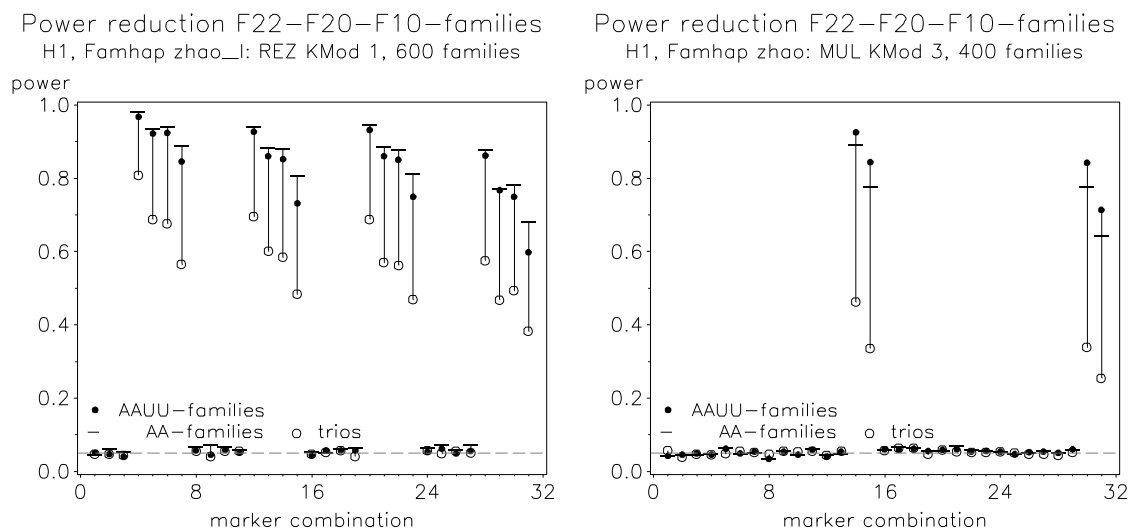


Abbildung 6.41: Vergleich der Familienstrukturen F22 - F20 und F10, Famhap; links: rezessives Modell NMod 1, rechts: multiplikatives Modell NMod 3

Bei Unphased beobachtet man im Modell KMod 1 (ADD / MUL) zunächst den Powereinbruch für die Mk 7, 13, 14, 21, 22 und 28 für Trios im vollen Modell (Abb. 6.42). Für die gleichen Mk ist im vollen Modell dann auch die Power für F20 gegenüber F22 deutlich erniedrigt. Im Fall des additiven Vererbungsmodus liegt sie sogar etwa in der Mitte zwischen Trios und AAUU-Familien. Beide Erscheinungen treten aber nicht im rezessiven Modus auf.

Ganz aus dem Rahmen fällt das additive Modell KMod 2: Abgesehen davon, dass das full-Verfahren hier nur Power für die Mk 12 - 14 und 28 hat, liegt die Power für AAUU-Familien

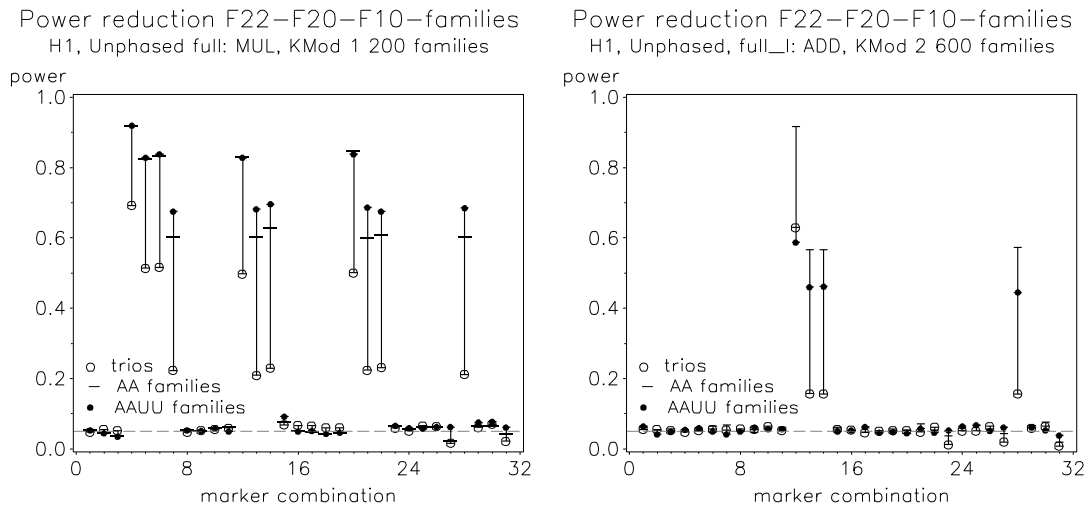


Abbildung 6.42: Vergleich der Familienstrukturen F22 - F20 und F10: Unphased, links: Rezessivmodell KMod 1, rechts: additives Modell KMod 2

deutlich unter der von AA-Familien, für den Test auf  $H_0^L$  an der Mk 12 sogar unter der für Trios (Abb. 6.42 rechts)!

### Overall-p-Werte

Es wurde wie in 6.2.4 verfahren, um zu einem kombinierten p-Wert über alle Markerkombinationen zu gelangen. Die Grafiken zu overall-p-Werten bestätigen die obigen Aussagen: AA-Familien erreichen im Modell KMod 1 die beste Power bzw. sind nur um wenige Prozente im Nachteil (s. Abb. 6.43). Bei den Modellen KMod 2 und 3 dagegen überflügeln AAUU-Familien die Familien mit ASPs deutlich. Abb. 6.44 spiegelt den Powereinbruch des vollen Modells von Unphased für F22 wider. Dazu korrespondiert übrigens Tab. 6.4.

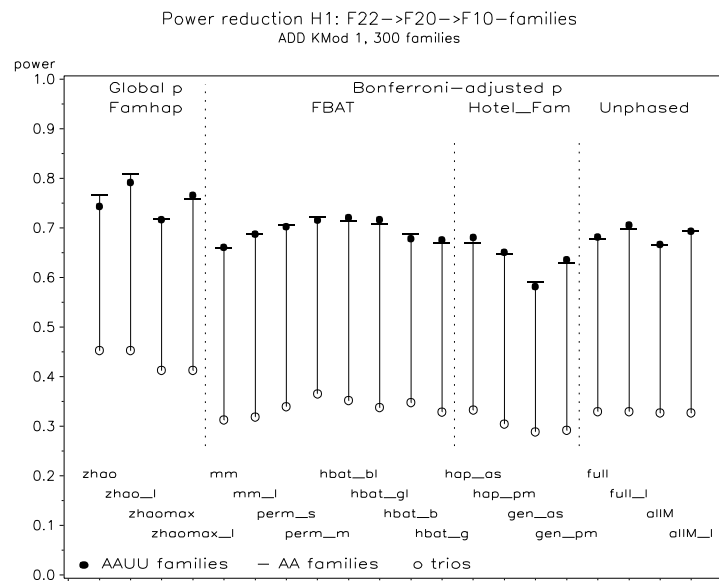


Abbildung 6.43: Vergleich der overall-Power für 300 F22- F20 bzw. F10-Familien, additives Modell KMod 1



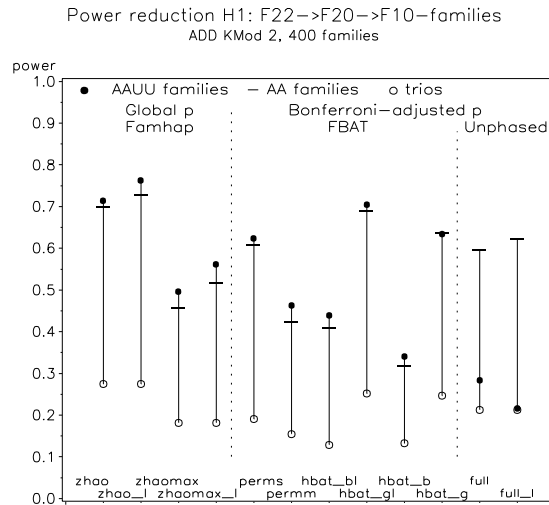


Abbildung 6.44: Vergleich der overall-Power für 500 F22- F20 und F10-Familien, additives Modell KMod 2

Es ist also durchaus vorteilhaft, aus größeren Kernfamilien nicht nur Trios, sondern ein ASP in die Studie einzubeziehen. Der Aufwand, gesunde Kinder zu genotypisieren, scheint sich im Fall der Konfiguration KMod 1 nicht zu lohnen und ist für KMod 2 und 3 zumindest fraglich.

#### 6.6.4 Verteilung der p-Werte - Klassifikation der Verfahren

Anhand der Grafiken in 6.3 ist bereits aufgefallen, dass jeweils mehrere Verfahren ähnlich reagieren. Auch wenn folgende Einteilung der 20 Statistiken in verschiedene Gruppen nach äußeren Merkmalen erfolgt und fast trivial erscheint, hat sie praktische Bedeutung:

- Nach der Komplexität unterteilt eine erste Klassifikation in 'Multi-Marker-Verfahren' [**M**] und 'Haplotyp-Verfahren' [**H**], welche auf Phasenauflösung basieren. Die Studie hat deutlich gezeigt, dass erstere nicht in der Lage sind, reine Haplotyp-Effekte nachzuweisen. Die Methoden `mm`, `mm_l`, `hap_as`, `hap_pm`, `gen_as`, `gen_pm` sowie `allm`, `allm_l` gehören zu ihnen.
- Die Haplotyp-Verfahren wiederum lassen sich danach einteilen, ob die globale Statistik für eine Markerkombination nur auf dem 'besten' Haplotypen, der größten Statistik bzw. dem kleinsten p-Wert, basiert [**1**] oder nicht [**2**]. So gehören zur Gruppe [1] die Verfahren `hbat_b`, `hbat_bl`, `zhaomax`, `zhaomax_l` und `perm_m`. Methoden der Gruppe [2] nutzen die gesamte Information, also z. B. die Summe der  $\chi^2$ -Werte über alle Haplotypen (`hbat_g`, `hbat_gl`, `zhao`, `zhao_l`, `perm_s`) bzw. eine Score- bzw. LRT-Statistik (`full`).
- Eine dritte, triviale Einteilung ist die in asymptotische [**a**] und Permutations-basierte [**p**] Verfahren. D. h. entweder, man vertraut darauf, dass die Näherung an eine  $\chi^2$ -Verteilung bereits ausreichend ist, oder man setzt Monte-Carlo-Methoden ein, um die Verteilung der Teststatistik zu erzeugen und damit den p-Wert zu bestimmen.
- Die Unterscheidung in Tests auf  $H_0^A$  bzw.  $H_0^L$  wurde oben bereits ausreichend untersucht.

Die Verteilung der p-Werte unter  $H_1$  wurde beispielhaft für einzelne Szenarien an ausgewählten Markerkombinationen untersucht. Abb. 6.45 zeigt Histogramme der jeweils 1000 p-Werte verschiedener Tests für das rezessive Modell KMod 2 mit Trios an der Markerkombination 12. (Entsprechend KMod 2 erzielen die Verfahren hier die höchste Power.) Die Häufigkeitsverteilungen für andere Szenarien unter  $H_1$  und an anderen Mk unterscheiden sich davon zwar quantitativ, aber

nicht qualitativ. Wegen der besseren Vergleichbarkeit wurden acht Tests auf  $H_0^L$  gewählt. Links findet man die asymptotischen sowie die beiden auf Permutation beruhenden Haplotypstests von FBAT, rechts dagegen jeweils die beiden Tests auf Kopplung von Famhap bzw. Unphased. Entsprechend dem (negativen) Zehnerlogarithmus der p-Werte entspricht die Klassenbreite '1' einer Zehnerpotenz. Um die Häufigkeit der Werte  $p = 0$  auf der logarithmischen Skala darstellen zu können, wurde dieser Wert durch  $p = 10^{-9}$  substituiert. Die relativ hohe Belegung der Klasse  $[10^{-9}, 10^{-10})$  verdeutlicht, wie oft der Nullwert vorkommt. Die linke Punktlinie markiert das Fehlerniveau  $\alpha = 0.05$ , die rechte den Ersatzwert für die Null.

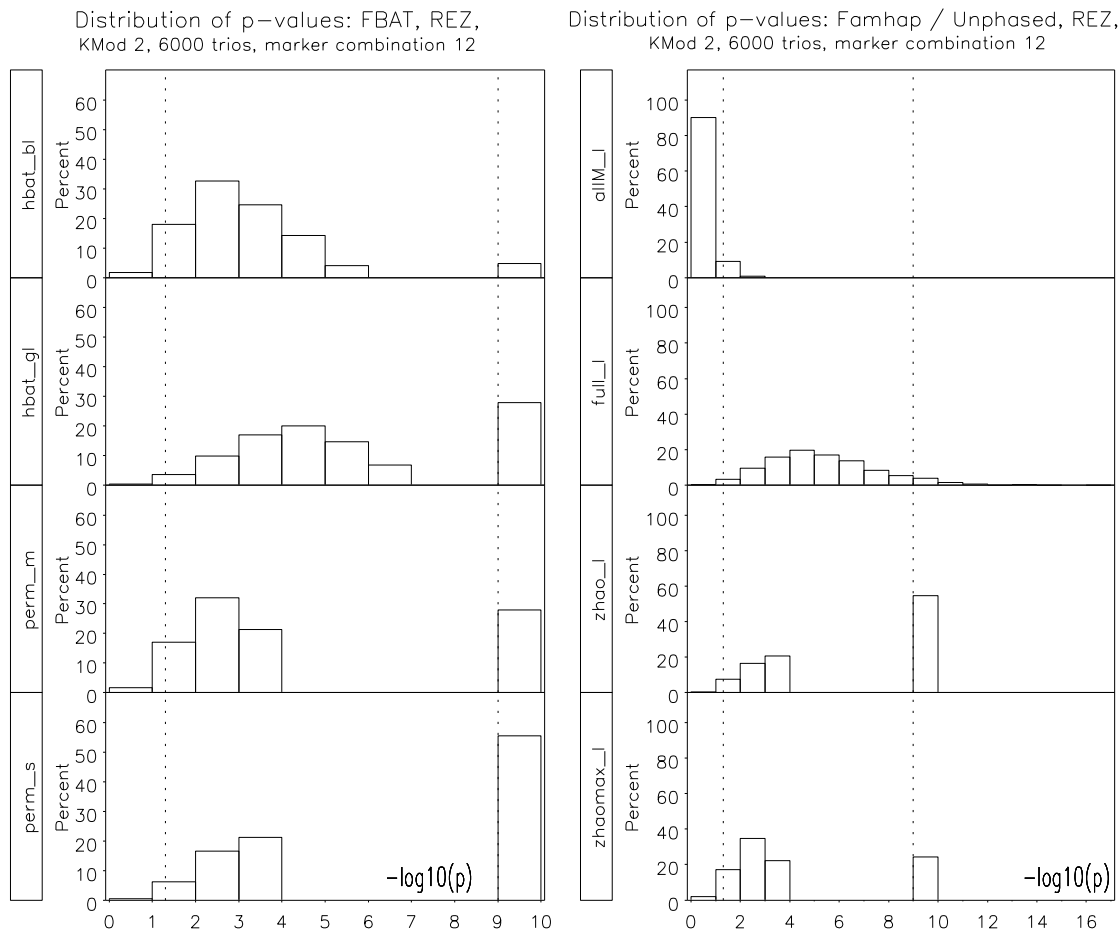


Abbildung 6.45: Verteilung der p-Werte für ausgewählte Verfahren

Die SAS-Prozedur PROC CAPABILITY musste für das Histogramm von Unphased (volles Modell) den Bereich erweitern. Offensichtlich kommen, wenngleich selten, Werte  $p < 10^{-16}$  vor.

Weiterhin fällt auf, dass die Permutations-basierten Verfahren [p] bei  $10^{-4}$  gestutzt sind. Eine Replikationszahl  $R=4000$  hat außer 0 als kleinsten möglichen p-Wert  $p = 0.00025$ . Ist  $R$  die Anzahl Permutationen, so haben Permutationstests nur eine 'Auflösung' von  $1/R$ . Zwar kann man die Anzahl der Replikationen je nach Speicherplatz-Bedarf bzw. Rechenzeit erhöhen (z. B.  $10^8$ ), aber irgendwo gibt es diese Schwelle, ab der die Verteilung auf die 0 abgebildet wird. In Abb. 6.45 sieht man das für perm\_m, perm\_s, zhao und zhaomax\_l.

Asymptotische Tests [a] haben nicht diese Begrenzung, d. h. es sind auch sehr kleine p-Werte möglich. Das ist für full\_l sehr gut erkennbar und trifft im Prinzip auch für hbat\_gl und hbat\_b1 zu. Durch den Befehl hbat werden asymptotisch normalverteilte Z-Statistiken für jeden häufigen Haplotyp berechnet. Mittels Bonferroni-Korrektur hinsichtlich der Anzahl häufiger Haplotypen (vgl. 6.6.2) wurde daraus die Statistik hbat\_b bzw. hbat\_b1 kalkuliert. Übersteigt allerdings der

Z-Wert für irgendeinen Haplotyp bzw. die globale  $\chi^2$ -Statistik einen gewissen Schwellwert, gibt das Programm den p-Wert 0 aus. Aus der Programmausgabe lässt sich die Schranke  $Z = 5$  vermuten. Mit anderen Worten: Haplotyp-spezifische p-Werte  $p < 5.733 \cdot 10^{-7}$  werden auf die Null abgebildet. Im betrachteten Szenarium betrifft das immerhin über 20%. Die Grenze für die globale  $\chi^2$ -Statistik hängt vom Freiheitsgrad ab. Aus der Programmausgabe kann man diese nicht genau rekonstruieren. Sie liegt in der Größenordnung des  $2 \cdot 10^{-7}$ -Quantils. Die Ausgabe  $p = 0$  für die Verfahren `hbat_g` bzw. `hbat_g1` ist also etwa gleichbedeutend mit  $p < 2 \cdot 10^{-7}$ . So sind asymptotische Statistiken sinnvoll, wenn nach Markerkombinationen mit kleinen p-Werten gesucht wird. (Die FBAT-Tests haben zusätzlich noch den Vorteil, dass sie wegen des Bedingens auf minimale suffiziente Statistiken den Fehler 1. Art einhalten.)

Die Verteilung der p-Werte von `hbat_b1` ist gegenüber der von `hbat_g1` um ein bis zwei Klassen nach links verschoben. Das rührt daher, dass der Mk-spezifische p-Wert hier aus dem kleinsten p-Wert durch Bonferroni-Korrektur ermittelt wird. Eine leichte Verschiebung ist auch für `zhaomax` gegenüber `zhao` zu beobachten. Zwar erfolgt dabei keine Multiplikation mit der Markeranzahl, aber der Test basiert ebenso allein auf der maximalen Statistik über alle Haplotypen. Das kann vorteilhaft sein, wenn es *genau einen* Haplotyp mit einem sehr starken Effekt gibt. Die Simulationen aber zeigen, dass die Einbeziehung der Information aller Haplotypen in einer z. T. deutlich höheren Power resultiert.

Schließlich stellt das Histogramm des `a11M_1`-Verfahrens einen Test dar, dessen Power sich auf dem  $\alpha$ -Niveau bewegt.

## 6.7 Anwendung der vier Verfahren auf reale Familiendaten

**Ausgangsdaten** Zur Abrundung dieser Studie wurden die vier Programme auf echte Versuchsdaten angewandt. Dazu standen Genotypen aus einer Assoziationsstudie von 2003 - Schizophrenie und das Gen 6p22.3 - zur Verfügung [137]. Die 196 deutschen, israelischen und ungarischen Familien setzen sich aus 124 Trios, 61 AA-Familien sowie 5 Familien mit drei, 3 Familien mit vier und je eine Familie mit fünf, sechs bzw. sieben Kindern zusammen. Eine vorausgegangene Untersuchung hatte Kopplung im Gen 6p22.3 (Dysbindin) nachgewiesen [136]. Durch eine Assoziationsstudie sollte das Ergebnis bestätigt werden. Deshalb waren 2003 von Schwab et al. sechs der SNPs mit den stärksten Effekten einer Assoziations-Studie von Straub [149] als Marker gewählt und genotypisiert worden: rs3213207, rs1011313, rs2619528, rs760761, rs2619522, rs1018381. Die Daten waren bereits 2004 mit Famhap auf Assoziation getestet worden [20].

**Analyse** Die Genotypen dieser sechs SNPs wurden mit jedem der vier Programme ausgewertet. Dabei wurden wegen des vorliegenden Kopplungsbefunds nur die 11 Tests auf  $H_0^A$  zum Niveau  $\alpha = 0.05$  angewandt. Die Haplotyp-Tests von FBAT wurden mit den Optionen `minfreq 0.05`, `0.01` bzw. `0` durchgeführt. Ebenso kam das volle Modell von Unphased mit den Optionen `-rare 0.05`, `0.01` bzw. `0` zum Einsatz. Als Replikationszahl für die Permutationsverfahren wurde für FBAT und das `zhao`-Verfahren von Famhap  $R=99.999.999$  sowie für `Hotel_Fam`  $R=999.999$  gewählt. Weil der Speicherplatzbedarf mit der Replikationszahl wächst, war die Analyse mit dem `zhaomax`-Verfahren nur mit  $R=100.001$  möglich. Bei sechs Markern gibt es 63 Kombinationen. Hinsichtlich multiplen Testens wurden also die Mk-spezifischen p-Werte mit dem Faktor 63 nach Bonferroni korrigiert. Pendant zum korrigierten minimalen p-Wert der anderen Programme ist wiederum der totale p-Wert für Famhap.

**Ergebnisse** Alle elf Algorithmen stimmen darin überein, dass die Krankheit mit Haplotypen vieler der 63 Markerkombinationen assoziiert ist. Bei den Permutationsverfahren kommen bei einzelnen Markerkombinationen die Werte  $p=0$  vor, nämlich genau dann, wenn es keine Replikation gibt, deren Statistik die beobachtete überschreitet. Die Bonferroni-Korrektur verändert den Wert  $p=0$  nicht. Somit bewegen sich die overall-p-Werte zwischen 0 und 0.0006.

Im Blick auf einzelne Markerkombinationen sind die Tests an vielen Stellen signifikant. Tab. 6.9 enthält für neun der elf Algorithmen die Anzahlen signifikanter Mk sowie die overall-p-Werte. Die beiden asymptotischen  $T^2$ -Tests wurden weggelassen.

Program	Methode	Signif. <sup>a</sup>	Overall p <sup>b</sup>
FBAT	<code>mm</code>	47	0.0003
	<code>hbat_b</code>	51	0.0000 <sup>c</sup>
	<code>hbat_g</code>	48	0.0006
Famhap	<code>zhao</code>	41	0.0001
	<code>zhaomax</code>	39	0.0001
Hotel_Fam	<code>hap_pm</code>	18	0.0000 <sup>d</sup>
	<code>gen_pm</code>	29	0.0000 <sup>e</sup>
Unphased	<code>full</code>	60	0.0006
	<code>allM</code>	57	0.0001

<sup>a</sup> Anzahl signifikanter Mk

<sup>b</sup> Famhap: total p; FBAT, Hotel\_Fam, Unphased: Bonferroni-korrigierter min-p-Wert

<sup>c</sup>  $p=0$  für Mk 29 und 53

<sup>d</sup>  $p=0$  für Mk 20, 49, 53 und 61

<sup>e</sup>  $p=0$  für 11 Markerkombinationen

Tabelle 6.9: Analyse realer Daten

Methode	Markerkombination											
mm	20	22	52	21	23	28	24	18	4	54	29	30
hbat_b	29	53	20	21	52	60	61	28	23	25	49	55
hbat_g	20	4	53	52	61	25	21	29	49	60	28	57
zhao	20	21	52	29	61	53	28	25	49	57	60	24
zhaomax	20	21	28	29	52	53	60	61	49	57	23	25
hap_pm	20	49	53	61	4	18	21	28	29	48	24	25
gen_pm	20	21	29	37	48	49	52	53	57	61	63	25
full	20	60	52	61	54	4	63	26	55	28	21	62
allM	20	52	22	21	28	60	54	53	51	23	29	62

Tabelle 6.10: Markerkombinationen mit den 12 kleinsten p-Werten

Bei allen neun Verfahren sind die Kombinationen 20 und 21 unter den elf mit den kleinsten p-Werten zu finden (vgl. Tab. 6.10). Auch für die Mk 28, 29 und 52 (8-mal), 53 bzw. 61 (7-mal) gibt es gute Übereinstimmung. Offensichtlich ist es kein Zufall, dass diese ersten sieben Mk alle  $M_2M_4$  enthalten. Dabei ist für  $M_2$  allein (Mk 16) kein Effekt nachweisbar! Tab. A.17 im Anhang enthält die detaillierte Auflistung, welche Markerkombinationen signifikant sind: Im oberen Teil der Tabelle findet man die Markerkombinationen 1 bis 31 und darunter die Mk 32 bis 63. Auch hier sind die Mk binär codiert: Graue Rechtecke in den beiden Kopf-Bereichen zeigen an, welche Marker ( $M_1 \dots M_6$ ) jeweils in der Kombination enthalten sind. Darunter werden für die 9 Verfahren nach Bonferroni-Korrektur zum 5%-Niveau signifikante Markerkombinationen durch ein '+' gekennzeichnet.

Der vorliegende Datensatz ist ein gutes Beispiel dafür, dass Kombinationen mehrerer Marker deutliche Effekte zeigen, während für einzelne Marker keine Assoziation nachgewiesen wird.: Richtet man zunächst die Aufmerksamkeit auf die Mk 1, 2, 4, 8, 16 und 32, erhält man Hinweise auf Assoziation der einzelnen Marker  $M_6, \dots, M_1$ . So gibt es für  $M_6$  und  $M_2$  keine Evidenz für Assoziation und bei  $M_1$  (Mk 32) sind nur die beiden Unphased-Verfahren signifikant. Kombination 49 dagegen besteht gerade aus  $M_1M_2M_6$  und wird von allen neun Verfahren als signifikant erkannt.

Am Schluss der Tabelle ist der Quotient aus dem Maximum und Minimum der 9 p-Werte jeder Mk angegeben. Für zwei Mk übersteigt dieser Faktor den Wert 500: So liefert das allM-Verfahren von Unphased für Mk 62  $p = 5.339 \cdot 10^{-04}$  und der  $T^2$ -Test mit Haplotyp-Codierung  $p = 0.2709$ . Ähnliches gilt für MK 46, wo fünf Verfahren nicht signifikant zum 5%-Niveau sind und das full-Verfahren  $p = 1.89 \cdot 10^{-3}$  erreicht. Dieses Ergebnis unterstreicht deutlich, dass p-Werte nur sehr vorsichtig als Effektstärken zu interpretieren sind (vgl. z. B. [35]). Das gilt insbesondere, wenn es sich um so unterschiedliche Methoden wie die hier betrachteten handelt. Die Verschiedenheit der Algorithmen ist weiterhin auch an der unterschiedlichen Reihung der 12 'besten' Markerkombinationen in Tab. 6.10 gut zu erkennen.

Die Permutations-basierten Verfahren liefern nicht selten Werte  $p = 0$ . Offensichtlich gibt es bei diesem Beispiel eine Reihe von Haplotyp-Effekten analog KMod 2 und 3. Das spiegelt sich darin, dass Hotel\_Fam nur 18 bzw. 29 signifikante Markerkombinationen entdeckt. Umgekehrt fragt man sich, ob nicht unter den zahlreichen Signifikanzen der Unphased-Verfahren falsch-positive Befunde sind. (Die beiden Haplotyp-Permutationstests perm\_s bzw. perm\_m geben übrigens 61 bzw. 60 signifikante Markerkombinationen aus. Allerdings ist hier zu berücksichtigen, dass sie als Tests auf Kopplung inadäquat sind, weil bereits Kopplung nachgewiesen wurde und ca. 1/3 der Familien mindestens zwei Kinder enthalten.)

# Kapitel 7

## Zusammenfassung und Ausblick

### 7.1 Kritische Anmerkungen

**Das Modell KMod 2** Um die Inflation des Fehlers 1. Art beim Test auf  $H_0^L$  mit AA- bzw. AAUU-Familien offenzulegen, war im Modell NMod 2 die Risikoallelfrequenz  $p_D = 0.01$  niedrig gewählt (vgl. 2.2.3). Wegen des rezessiven Modells entspricht das einer Prävalenz von 1 auf 10.000 Personen. War dieses Modell extrem oder zu extrem gewählt? Immerhin kann positiv gewertet werden, dass der Fehler 1. Art trotz mancher Fluktuation nicht völlig ausartete. Das spricht für die Robustheit der Verfahren gegenüber extremen Situationen.

Die Ausreißer der Methode `full` von `Unphased` (s. 6.2.5) lassen sich damit nicht restlos erklären. Denkbar sind numerische Instabilitäten oder mangelnde asymptotische Näherung für  $N=200$ . Immerhin wird bei der Herleitung des Verfahrens sehr idealisiert davon ausgegangen, dass der Merkmalsvektor  $\mathbf{Y}$  multivariat normalverteilt ist (vgl. 4.4.1).

**Schranke für seltene Haplotypen** Es wäre vorteilhaft gewesen, die Haplotyp-Schranke (Optionen `minfreq 0.01` bzw. `-rare 0.01`) von vornherein auf 1% anstelle von 5% zu setzen. Die im FBAT-Manual [94], [95] angegebene höhere Schwelle entspricht der CDCV-Hypothese. Sie erhöht die Effizienz, Assoziation einer häufigeren Variante mit der Krankheit nachzuweisen. Seltene Haplotypen mit moderaten Effekten fallen aber damit durch das Sieb. Umgekehrt muss man jedoch annehmen, dass die Stichprobenumfänge vieler Assoziationsstudien zu klein sind, um Effekte solch seltener Varianten überhaupt aufdecken zu können [157]. Optimal ist es, den Schwellwert von Fall zu Fall in Abhängigkeit von der Markeranzahl, der zu erwartenden Haplotyp-Diversität, aber auch vom Stichprobenumfang festzulegen.

**Anzahl von Replikationen** Der Einwand, dass 1000 Replikationen für die Abschätzung des Fehlers 1. Art und der Power zu wenig sind, ist nicht von der Hand zu weisen. Das gilt insbesondere, je kleiner das Fehlerniveau  $\alpha$  gewählt wird. Vermutlich variieren die Werte deshalb auch stärker (vgl. 6.2.5). Allerdings waren mehr Replikationen aus Gründen der Rechenzeit nicht realisierbar. Ähnlich gelagert ist die Frage nach der Anzahl von Wiederholungen bei den Permutationstests. Hier wurde einheitlich  $R = 4000$  gewählt. Das kann insbesondere für kleine  $\alpha$  etwas wenig sein. Allerdings wurde, wie im Absatz 6.6.1 untersucht, diese Höchstzahl bei FBAT nicht einmal ausgeschöpft. Ein anderer kritischer Punkt ist die Wahl der Replikationsanzahl  $R = 4000$ . Sei  $T^0$  die beobachtete Statistik und  $T^r$  die der  $r$ . Replikation ( $r = 1, \dots, R$ ). Weil der globale p-Wert über alle Haplotypen nach der Formel

$$p = \frac{|\{r : T^r \geq T^0\}|}{R}$$

bestimmt wird (vgl. (4.15)), ergeben sich als p-Werte nur Vielfache von  $1/R$ . Ideal wären nun solche  $R$ , für die es kein ganzzahliges  $k$  mit  $p = k/R = \alpha$  gibt, so dass die Fehlerschranke mit

Wahrscheinlichkeit 0 getroffen wird. Andernfalls kann u. U. der Fehler 1. Art leicht nach oben verzerrt werden. (Für  $R = 4000$  wird der Wert  $\alpha = 0.05$  gerade für  $k = 200$  erreicht.)

**LD zwischen den Markern** Bei der Definition der Modelle wurde angenommen, dass die Marker  $M_1, \dots, M_5$  im Kopplungs-Gleichgewicht stehen (vgl. 2.2.2, 2.2.3). Diese Bedingung ist für eng gekoppelte Marker eher die Ausnahme. So ist diese Festlegung einerseits dem Umstand geschuldet, die Komplexität der Modelle nicht zu groß werden zu lassen. Kopplungsgleichgewicht zwischen den Markern lässt die Effekte deutlich hervortreten, währenddessen LD die Effekte modifiziert und verschleiert.

**Das Fehlerniveau 5%** Die Festlegung des Fehlerniveaus  $\alpha = 0.05$  folgte üblichen Gepflogenheiten und rein praktischen Erwägungen. Ein Testniveau  $\alpha = 0.05$  kann für ein Screening genutzt werden, wo Kandidaten für weitere Studien gefunden werden sollen [88]. Viele Autoren empfinden diese Schranke als zu hoch, was einen Teil nicht reproduzierbarer Assoziationsbefunde erklärt (z. B. [68], [35]). Colhoun et al. [35] zeigten anhand eines einfachen Beispiels, dass in Abhängigkeit zur Gesamthäufigkeit positiver Befunde das Verhältnis zwischen wahren positiven zu falsch positiven Befunden viel ungünstiger sein kann, als es die Zahl 5% ausdrückt. Deshalb schlugen diese Autoren für Kandidatengen-Studien komplexer Krankheiten z. B. Signifikanzschwellen zwischen  $5 \times 10^{-4}$  bis  $5 \times 10^{-5}$  vor. Ist die a-priori-Wahrscheinlichkeit für Assoziation wesentlich höher, kann diese strenge Forderung gelockert werden [88].

**Die Verwendung von p-Werten** Die zur Analyse des Dysbindin-Datensatzes eingesetzten Verfahren gaben p-Werte aus, die bis zu einem Faktor  $> 500$  differieren (s. 6.7). Das belegt klar, wie wenig aussagekräftig p-Werte sind. Dieser Befund leitet einen direkt zu einer jahrzehntelangen Debatte zwischen Anhängern von Signifikanztests und Hypothesentests. Obwohl eine ausführliche Diskussion dieser Problematik den Rahmen dieser Studie sprengt, sind wenige Sätze trotzdem notwendig:

R. A. Fisher definierte den p-Wert als ein (induktives) Maß für die Evidenz eines Effekts und schlug  $p = 0.05$  als sinnvolle Schranke vor [60]. Neyman und Pearson dagegen sahen Hypothesentests als Verhaltensregel an, die 'auf lange Sicht garantiert, wenig Fehler zu begehen', aber keine Aussagekraft für den konkreten Fall hat [116], [67]. Die Ähnlichkeit des p-Wertes zur Fehlerschwelle  $\alpha$  hat in der Geschichte zur Vermischung zweier Konzepte geführt [68]. Man kann das nicht als dogmatischen Streit abtun, weil zahlreiche Fehlinterpretationen des p-Wertes durchaus praktische Konsequenzen haben [66]. Z. B. wird die Bedeutung des Wertes  $p = 0.05$  oft überschätzt. Zweifelsohne sind p-Werte auch in der genetischen Epidemiologie ein weit verbreitetes Mittel, Versuchsergebnisse zu präsentieren. Dementsprechend geben die untersuchten (und die anderen recherchierten) Programme in der Regel p-Werte aus. Auch die Suche nach adäquaten Signifikanzschranken für GWAS rückt die Problematik der p-Werte neu in den Blickpunkt.

S. N. Goodman stellt mit dem 'Bayes Factor' ein Konzept vor, das die Evidenz für Effekte misst und Nachteile des p-Wert-Ansatzes vermeidet [69]. Folgende Gleichung beschreibt den Bayes Factor:

$$\text{Prior Odds of } H_0 \times \text{Bayes Factor} = \text{Posterior Odds of } H_0 \quad (7.1)$$

Dabei sind der

$$\text{Bayes Factor} = \frac{P\{\text{Daten}, H_0 \text{ gegeben}\}}{P\{\text{Daten}, H_1 \text{ gegeben}\}} \quad (7.2)$$

und die

$$\text{Prior Odds of } H_0 = \frac{\pi_0}{1 - \pi_0} \quad \text{mit} \quad \pi_0 = P\{H_0\}. \quad (7.3)$$

Damit gibt der Bayes Factor an, um wieviel sich die Evidenz gegen die Nullhypothese und für die Alternative durch den Versuch verbessert. Natürlich hat das seinen Preis: Es ist nicht einfach, aus dem Vorwissen die Prior Odds festzulegen. Außerdem vermittelt das den Eindruck von Subjektivität. Weiterhin kann die Berechnung der Chancen nach dem Versuch ('Posterior Odds') auf komplizierte Mehrfachintegrale führen. Schließlich hängt der Bayes Factor von der

Alternativhypothese ab und nicht nur von  $H_0$  wie der p-Wert. J. Wakefield leitete kürzlich einen handhabbaren asymptotischen Bayes-Faktor für GWAS her [156]. Das zeigt, dass dieser Ansatz für künftige Analysen ernst zu nehmen ist. Außerdem ist der Bayes-Faktor viel besser als der p-Wert geeignet, um Resultate verschiedener Studie zu kombinieren [69].

## 7.2 Zusammenfassung

Vorliegende Studie hat sich der Aufgabe gestellt, die Eigenschaften Familien-basierter Assoziations-tests zu untersuchen, um dem Anwender Entscheidungshilfen bei der Wahl des Verfahrens zu geben. Alle vier untersuchten Programme enthalten valide und effiziente Verfahren zum Test auf Assoziation. Einmal mehr zeigte sich, dass die Auswahl entscheidend vom Blickpunkt und von den Vorkenntnissen abhängt:

- Eine entscheidende Frage ist, wie komplex das gewählte Modell gewählt werden muss: Dafür ist der TDT ein gutes Beispiel: Er ist robust gegen Verzerrungen durch Populationsstratifikation. Der Preis dafür ist aber, dass er in homogenen Populationen z. B. gegenüber Fall-Kontroll-Verfahren Power verliert (z. B. [28]). Analog geht es hier darum, ob bereits Allele einzelner Marker mit der Krankheit assoziiert sind. Dann sind die 'Multi-Marker-Methoden' am machtvollsten. In Situationen allerdings, wo nur Assoziation von Haplotypen vorliegt, versagen sie und Methoden müssen eingesetzt werden, welche die Phase auflösen. Schließlich sind mehrere Methoden in der Lage, Kovariablen in die Analyse einzubeziehen. Aber das wird in gleicher Weise nachteilig sein, wenn die Kovariablen keinen klaren Einfluss auf das Erkrankungsrisiko haben.
- Geht es darum, Assoziation *irgendeines* Haplotypen *irgendeiner* Markerkombination zu entdecken, ist das **zhao**-Verfahren von Famhap vorteilhaft: Die Berechnung eines totalen p-Wertes über alle Mk bringt gegenüber den alternativen Methoden ein Plus an Power. Außerdem gibt das Programm auch detaillierte Informationen zu einzelnen Haplotypen aus. In den untersuchten Szenarien offenbarte die Methode kaum Schwächen. Zwar waren andere Verfahren für einzelne Markerkombinationen in manchen Szenarien besser, **zhao** scheint aber insbesondere bei moderaten Effekten vorzuziehen zu sein.
- Weiterhin ist entscheidend, ob sich die Marker in bekannten Kopplungsregionen befinden. Absatz 6.2.2 unterstreicht eindrücklich, dass dann Tests auf  $H_0^L$  für Familien mit mehreren kranken Kindern den Fehler 1. Art inflationieren. Entweder man führt die Studie mit Trios durch oder benutzt Tests der Hypothese  $H_0^A$ .
- Auch die Struktur der Probandenfamilien hat einen Einfluss: Wenn es außer dem Indexfall noch ein weiteres krankes Geschwister gibt, bringt deren Genotypisierung in der Regel einen relevanten Powergewinn. Der Aufwand, weitere - insbesondere gesunde - Geschwister in die Studie einzubeziehen, will dagegen genau überlegt sein.
- Der 'globale' Haplotyp-Test von FBAT zeigt keine offensichtlichen Schwächen. Wegen der Invarianzeigenschaft minimal suffizienter Statistiken hält er das Fehlerniveau 1. Art gut ein. Die Power des asymptotischen Tests ist über alle Szenarien gut und in vielen Fällen etwas höher als die des Permutations-basierten. Einen geringen Anteil fehlender Genotypen kann das Verfahren gut kompensieren. Zusammen mit dem Programm PBAT, welches u. a. Stichprobenumfänge für Familien-basierte Studien berechnen kann, bildet FBAT ein vielseitiges und komfortables Paket.
- Das volle Modell von Unphased offenbarte leider einige Schwachpunkte: Einerseits streut die falsch-positiv-Rate stärker, andererseits ist die Power vor allem für den additiven Vererbungsmodus nur mäßig. Nach den Ergebnissen der Simulationsstudie kann man eine Analyse mit



dem Haupteffekt-Modell kaum empfehlen: `allM` kann mit guter Power nur Single-Marker-Assoziation nachweisen. In diesen Fällen sind aber die Permutations-basierten  $T^2$ -Tests überlegen. `Unphased` zeichnen die Fähigkeiten aus, sowohl Kovariablen in die Analyse einzubeziehen als auch Stammbäume über mehr als zwei Generationen zu analysieren. Schließlich kann `Unphased` bei einem höheren Anteil fehlender Genotypen vorteilhaft sein, wenn die Power anderer Verfahren durch die 'Missings' zusammenbricht.

- Die Verfahren `perm_m`, `hbat_b`, `zhaomax`, `hap_as` und `gen_as` sind kaum zu empfehlen.
- `Hotel_Fam` dagegen ist in der Lage, Genotypen von Markern zu analysieren, die u. U. in verschiedenen Blöcken liegen. Wegen der Annahme der Rekombinationsfrequenz 0 zwischen den Markern sind die anderen Programme dafür nicht anwendbar.
- `Unphased` und `Famhap` ab Version 18 ermöglichen eine bedingte Analyse. Diese ist nützlich, um Effekte von Haplotypen irgendwelcher Mk von bekannten Effekten eines anderen Markers zu separieren.

### 7.3 Ausblick

**Weitere Verfahrensvergleiche** Eine Reihe von Verfahren wurde zu spät veröffentlicht, um in die Studie einbezogen zu werden. Dazu gehören `TRIMM` [141], `PCPH` [9] und das Modul `hapMC` von 'Genie' [39]. Auch `PLINK` [122], das sich als Standardpaket für genomweite Analysen etablieren konnte, lohnt sich weiter zu verfolgen. Es wäre interessant, wie sich diese Algorithmen im Vergleich zu den vier ausgewählten Programmen verhalten, zumal sie auf anderen statistischen Ansätzen basieren.

Die in dieser Studie betrachteten Familien-basierten Methoden sind auf eng gekoppelte Marker gerichtet. In der Literatur wurde immer mal angeregt, Familien-basierte mit Fall-Kontroll-Methoden zu vergleichen. Obwohl dieser Gedanke verlockend klingt, sprechen mehrere Gründe dagegen: Erstens werden beide Verfahrenstypen mit unterschiedlicher Zielstellung eingesetzt. So sind sie weniger als Konkurrenten, sondern als wichtige Ergänzung zu betrachten [132], [97]. Außerdem ist eine Modellierung aufwändig und schwierig.

Alle vier hier getesteten Programme sind robust gegen Populationsstratifikation. Von großem Interesse ist ein Vergleich mit den Ansätzen 'Genomic Control' oder 'Structured Association Testing' (vgl. 1.2.1). Dieser gäbe Hinweise, ob sie als echte Alternativen den Aufwand Familien-basierter Assoziationsstudien vermeiden könnten.

**Kombination von Familien- und Fall-Kontroll-Daten** In letzter Zeit mehren sich Artikel, in denen es um die Kombination von Fall-Kontroll- und Familiendaten geht: So kann die Kombination der Ansätze die Schwächen beider Verfahren mildern. In der Arbeit von Ackerman et al. [7] z. B. wird aus den Anzahlen transmittierter / nicht transmittierter Allele ein OR berechnet und mit dem Populations-OR verglichen. Weiterhin leiten sich aus dem Vergleich nicht übertragener Allele mit den Populations-Kontrollen Hinweise auf Stratifikation ab. Auch Pfeiffer et al. [119], Guo et al. [72] sowie Hsu et al. [82] entwickelten verschiedene Ansätze zu diesem Problem. Schließlich regten Laird & Lange folgendes Vorgehen an, GWAS für die Planung einer nachfolgenden Familien-basierten Studie auszuwerten [97]: Von den signifikanten Markern der GWAS werden jene 10 bis 20 mit der höchsten zu erwartenden FBAT-Power als Basis der künftigen Studie ausgewählt.

**Fehlende Genotypen und totaler p-Wert** Von der Analyse mit 1% fehlenden Genotypen je Marker und Person kann man nur vorsichtig auf das Verhalten mit höheren Fehlerraten schließen. Offensichtlich leidet die Power von `Unphased` durch Datenausfälle am wenigsten. In dieser Studie ging es weiterhin nur um *uninformative missings*, d. h. der Datenausfall hing nicht von den Genotypen ab. Weitere Analysen waren aus Gründen der Rechenzeit nicht realisierbar. Analysen z. B. mit höheren missing-Raten (5%, 10% usw.) und evtl. *informative missings* wären notwendig, um

auszutesten, wie weit die Verfahren valide bleiben oder ob die Algorithmen durch fehlende Daten instabil werden. Leider mindern fehlende elterliche Genotypen die Power von Famhap, da es bisher aus konzeptionellen Gründen nicht möglich war, unvollständige Trios in die Testprozedur einzubeziehen. FBAT bietet noch keine Option an, alle Markerkombinationen zu testen und darüber eine totale Statistik zu ermitteln. Es müsste geprüft werden, ob das in Famhap implementierte Verfahren von Ge [63] analog zur Kalkulation eines overall-p-Wertes für den Permutations-basierten Haplotypentest angewandt werden kann.

**Eine Antwort und viele neue Fragen** Der p-Wert einer Assoziationsstudie sagt relativ wenig über Risikoeffekte aus. Der Bayes Factor könnte ein praktikablerer Ansatz sein, die mit dem p-Wert verbundenen Probleme hinter sich zu lassen. Vielleicht eignet sich dieser Ansatz sogar, um Ergebnisse von L Markerkombinationen zu einem Wert zu kombinieren. (Der Mangel an einem Verfahren, das aus L nicht notwendig von 0 verschiedenen p-Werten einen 'totalen' Wert berechnet, ist auch durch vorliegende Studie bewusst geworden.)

Ein positiver Assoziationsbefund kann bei der Aufklärung der Kausalität komplexer Krankheiten ein wichtiger Mosaikstein sein. Erfolgreiche Anwendung dieser Verfahren motiviert, sie anzuwenden, auch wenn sie unvollkommen sind. Mit Goodmans Worten:

Standard frequentist methods are most problematic when used to draw conclusions in a single experiment. Their denial of a formal role for external information in inference poses serious practical and logical problems. But Bayesian methods, designed for inductive inference in single experiments, do not guarantee that in the long run, conclusions in which we have 95% confidence will turn out to be true 95% of the time. This is because Bayesian prior probability distributions are not ideal quantitative descriptors of what we know (or what we don't know), and Bayes theorem is an imperfect model for human learning. [69]

For we know in part and we prophesy in part. But when that which is perfect has come then that which is in part will be done away. [74]

# Literaturverzeichnis

- [1] Lebenszeitprävalenzen von ausgewählten Krankheiten in Prozent nach Alter, Geschlecht und sozialer Schicht 2002/2003. [http://www.gbe-bund.de/gbe10/ergebnisse.prc\\_tab?fid=9781&suchstring=Pr%E4valenz\\_Asthma\\_bronchiale&query\\_id=&sprache=D&fund\\_typ=TAB&methode=1&vt=1&verwandte=1&page\\_ret=0&seite=1&p\\_lfd\\_nr=4&p\\_news=&p\\_sprachkz=D&p\\_uid=gast&p\\_aid=51897606&hlp\\_nr=3&p\\_janein=J](http://www.gbe-bund.de/gbe10/ergebnisse.prc_tab?fid=9781&suchstring=Pr%E4valenz_Asthma_bronchiale&query_id=&sprache=D&fund_typ=TAB&methode=1&vt=1&verwandte=1&page_ret=0&seite=1&p_lfd_nr=4&p_news=&p_sprachkz=D&p_uid=gast&p_aid=51897606&hlp_nr=3&p_janein=J). Gesundheitsberichterstattung des Bundes, 06.06.2008.
- [2] Lebenszeitprävalenzen von Herz-Kreislaferkrankungen in Prozent nach Alter, Geschlecht und sozialer Schicht 2002/2003. [http://www.gbe-bund.de/gbe10/ergebnisse.prc\\_tab?fid=9779&suchstring=Pr%E4valenz\\_Angina\\_pectoris&query\\_id=&sprache=D&fund\\_typ=TAB&methode=1&vt=1&verwandte=1&page\\_ret=0&seite=1&p\\_sprachkz=D&p\\_uid=gast&p\\_lfd\\_nr=1&p\\_news=&p\\_aid=51897606&hlp\\_nr=3&p\\_janein=J](http://www.gbe-bund.de/gbe10/ergebnisse.prc_tab?fid=9779&suchstring=Pr%E4valenz_Angina_pectoris&query_id=&sprache=D&fund_typ=TAB&methode=1&vt=1&verwandte=1&page_ret=0&seite=1&p_sprachkz=D&p_uid=gast&p_lfd_nr=1&p_news=&p_aid=51897606&hlp_nr=3&p_janein=J). Gesundheitsberichterstattung des Bundes, 06.06.2008.
- [3] *Die Bibel*, chapter Jesaja 43, 1+4a. R. Brockhaus Verlag, Wuppertal, revidierte Elberfelder Übersetzung, 9. edition, 1991.
- [4] SAS Institute Inc. SAS Institute, Cary, NC, 2004.
- [5] *Grundgesetz der Bundesrepublik Deutschland*, chapter Artikel 1 (1). Deutscher Bundestag - Verwaltung, [http://www.bundestag.de/parlament/funktion/gesetze/gg\\_jan2007.pdf](http://www.bundestag.de/parlament/funktion/gesetze/gg_jan2007.pdf), textausgabe edition, Januar 2007. 23.09.2008.
- [6] G. R. Abecasis, S. S. Cherny, W. O. Cookson, and L. R. Cardon. Merlin-rapid analysis of dense genetic maps using sparse gene flow trees. *Nature Genetics*, 30:97–101, 2002.
- [7] H. Ackerman et al. A comparison of case-control and family-based association methods: The example of sickle-cell and malaria. *Ann Hum Genet*, 69(5):559–565, 2005.
- [8] J. Akey, L. Jin, and M. Xiong. Haplotypes vs single marker linkage disequilibrium tests: what do we gain? *European Journal of Human Genetics*, 9(4):291–300, 2001.
- [9] A. S. Allen and G. A. Satten. Inference on haplotype/disease association using parent-affected-child data: the projection conditional on parental haplotypes method. *Genet Epidemiol*, 31:211–223, 2007.
- [10] A. S. Allen, G. A. Satten, and A. A. Tsiatis. Locally-efficient robust estimation of haplotype-disease association in family-based studies. *Biometrika*, 92:559–571, 2005.
- [11] K. Allen-Brady, J. Wong, and N. J. Camp. PedGenie: an analysis approach for genetic association testing in extended pedigrees and genealogies of arbitrary size. *BMC Bioinformatics*, 7:#209, 2006.
- [12] D. Altshuler, M. J. Daly, and E. S. Lander. Genetic mapping in human disease. *Science*, 322:881–888, 2008.

- [13] S. A. Bacanu, B. Devlin, and K. Roeder. The power of genomic control. *Amer J Hum Genet*, 66:1933–1944, 2000.
- [14] S.-A. Bacanu, M. R. Nelson, and M. G. Ehm. Comparison of association methods for dense marker data. *Genet Epidemiol*, 32(8):791–799, 2008.
- [15] D. J. Balding. A tutorial on statistical methods for population association studies. *Nature Reviews Genetics*, 7(10):781–791, 2006.
- [16] P. Bauer. On the robustness of hotelling’s  $t^2$ . *Biometrical Journal*, 23(4):405–412, 1981.
- [17] T. Becker, M. P. Baur, and M. Knapp. Detection of parent-of-origin effects using haplotype analysis. *Hum Hered*, 62:64–76, 2006.
- [18] T. Becker and M. Knapp. Haplotype frequency estimation and association analysis with FAMHAP. Technical report, Institute for Medical Biometry, Informatics and Epidemiology, University of Bonn, Germany, 03/2007. Manual, Version 17.
- [19] T. Becker and M. Knapp. Maximum-likelihood estimation of haplotype frequencies in nuclear families. *Genet Epidemiol*, 27(1):21–32, 2004.
- [20] T. Becker and M. Knapp. A powerful strategy to account for multiple testing in the context of haplotype analysis. *Amer J Hum Genet*, 74(4):561–570, 2004.
- [21] T. & al. Becker. Identification of probable genotyping errors by considering haplotypes. *European Journal of Human Genetics*, 14:450–458, 2006.
- [22] M. Bishop, editor. *Guide to Human Genome Computing*, chapter Sham, P. and Zhao, J.-H.: Linkage analysis using affected sib-pairs. Academic Press, 2nd edition, 1998.
- [23] W. C. Blackwelder and R. C. Elston. A comparison of sib-pair linkage tests for disease susceptibility loci. *Genet Epidemiol*, 2:85–97, 1985.
- [24] D. Botstein and N. Risch. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nature Genetics*, 33 suppl:228–237, 2003.
- [25] C. Bourgain, E. Genin, H. Quesneville, and F. Clerget-Darpoux. Search for multifactorial disease susceptibility genes in founder populations. *Ann Hum Genet*, 64:255–265, 2000.
- [26] P. R. Burton, M. D. Tobin, and J. L. Hopper. Genetic Epidemiology: Key concepts in genetic epidemiology. *Lancet*, 366(10):941–951, 2005.
- [27] R. M. Cantor, G. K. Chen, P. Pajukanta, and K. Lange. Association testing in a linked region using large pedigrees. *Amer J Hum Genet*, 76:538–542, 2005.
- [28] L. R. Cardon and L. J. Palmer. Population stratification and spurious allelic association. *Lancet*, 361:598–604, 2003.
- [29] J. Chapman and J. Whittaker. Analysis of multiple snps in a candidate gene or region. *Genet Epidemiol*, 32(6):560–566, 2008.
- [30] J. M. Chapman, J. D. Cooper, J. A. Todd, and D. G. Clayton. Detecting disease associations due to linkage disequilibrium using haplotype tags: a class of tests and the determinants of statistical power. *Hum Hered*, 56:18–31, 2003.
- [31] A. G. Clark. The role of haplotypes in candidate gene studies. *Genet Epidemiol*, 27:321–333, 2004.
- [32] D. Clayton. A generalization of the transmission/disequilibrium test for uncertain-haplotype transmission. *Amer J Hum Genet*, 65(4):1170–1177, 1999.

- [33] D. Clayton and H. Jones. Transmission/disequilibrium tests for extended marker haplotypes. *Amer J Hum Genet*, 65(4):1161–1169, 1999.
- [34] H.-T. Clayton, D. Leung. An R package for analysis of whole-genome association studies. *Hum Hered*, 64:45–51, 2007.
- [35] H. M. Colhoun, P. M. McKeigue, and G. D. Smith. Problems of reporting genetic associations with complex outcomes. *Lancet*, 361(9360):865–872, 2003.
- [36] D. N. Cooper and J. F. Clayton. Dna polymorphism and the study of disease associations. *Human Genetics*, 78:299–312, 1988.
- [37] H. J. Cordell. Properties of case/pseudocontrol analysis for genetic association studies: effects of recombination, ascertainment, and multiple affected offspring. *Genet Epidemiol*, 26(3):186–205, 2004.
- [38] H. J. Cordell, B. J. Barratt, and D. G. Clayton. Case/pseudocontrol analysis in genetic association studies: a unified framework for detection of genotype and haplotype associations, gene-gene and gene-environment interactions, and parent-of-origin effects. *Genet Epidemiol*, 26(3):167–185, 2004.
- [39] K. Curtin, J. Wong, K. Allen-Brady, and N. J. Camp. Pedgenie 2:0: Meta genetic association testing in mixed family and case-control designs. In *Abstract, International Genetic Epidemiology Society 15th Annual Meeting*. , St. Pete Beach, FL, November 2006.
- [40] D. Curtis. Use of siblings as control in case-control association studies. *Ann Hum Genet*, 61(4):319–333, 1997.
- [41] M. J. Daly, J. D. Rioux, S. F. Schaffner, T. J. Hudson, and E. S. Lander. High-resolution haplotype structure in the human genome. *Nature Genetics*, 29(2):229–232, 2001.
- [42] P. I. W. de Bakker, R. Yelensky, I. Pe’er, et al. Efficiency and power in genetic association studies. *Nature Genetics*, 37(11):1217–1223, 2005.
- [43] R. DeLongchamp, T. Lee, and C. Velasco. A method for computing the overall statistical significance of a treatment effect among a group of genes. *BMC Bioinformatics*, 7(Suppl 2):S11, 2006.
- [44] B. Devlin, S. A. Bacanu, and K. Roeder. Genomic control to the extreme. *Nature Genetics*, 36(11):1129–1130, 2004.
- [45] B. Devlin and N. Risch. A comparison of linkage disequilibrium measures for fine-scale mapping. *Genomics*, 29:311–322, 1995.
- [46] B. Devlin and K. Roeder. Genomic control for association studies. *Biometrics*, 55:997–1004, 1999.
- [47] F. Dudbridge. Pedigree disequilibrium tests for multilocus haplotypes. *Genet Epidemiol*, 25:115–121, 2003.
- [48] F. Dudbridge. Robust tests of association for multilocus haplotypes in nuclear families. Technical report, MRC Biostatistics Unit, Cambridge, UK, 2005/7.
- [49] F. Dudbridge. Unphased user guide. Technical report, MRC Biostatistics Unit, Cambridge, UK, 2006/5.
- [50] F. Dudbridge. *UNPHASED, user guide*, version 3.0.6 edition, Feb 2007.
- [51] F. Dudbridge. Likelihood-based association analysis for nuclear families and unrelated subjects with missing genotype data. *Hum Hered*, 66:87–98, 2008.

- [52] F. Dudbridge and B. P. C. Koeleman. Rank truncated product of p values, with application to genomewide association scans. *Genet Epidemiol*, 25(4):360–366, 2003.
- [53] F. Dudbridge, BPC Koeleman, J. A. Todd, and D. G. Clayton. Unbiased application of the transmission/disequilibrium test to multilocus haplotypes. *Amer J Hum Genet*, 66:2009–2012, 2000.
- [54] W. Ewens and R. S. Spielman. What is the significance of a significant TDT. *Hum Hered*, 60:206–210, 2005.
- [55] C. T. Falk and P. Rubinstein. Haplotype relative risks: an easy reliable way to construct a proper control sample for risk calculations. *Ann Hum Genet*, 51(3):227–233, 1987.
- [56] D. Falush, M. Stephens, and J. K. Pritchard. Inference of population structure: extensions to linked loci and correlated allele frequencies. *Genetics*, 164:1567–1587, 2003.
- [57] R. Fan, M. Knapp, M. Wjst, C. Zhao, and M. Xiong. High resolution  $T^2$  association tests of complex diseases based on family data. *Ann Hum Genet*, 69(2):187–208, 2005.
- [58] R. Z. Fan and M. Knapp. Genome association studies of complex diseases by case-control designs. *Amer J Hum Genet*, 72:850–868, 2003.
- [59] T. E. Fingerlin, M. Boehnke, and G. R. Abecasis. Increasing the power and efficiency of disease-marker case-control association studies through use of allele-sharing information. *Amer J Hum Genet*, 74(3):432–443, 2004.
- [60] R. A. Fisher. *Statistical methods and scientific inference*. New York: Macmillan, 3rd edition, 1973.
- [61] N. Freimer and C. Sabatti. The use of pedigree, sib-pair and association studies of common diseases for genetic mapping and epidemiology. *Nature Genetics*, 36(10):1045–1051, 2004.
- [62] S. B. Gabriel et al. The structure of haplotype blocks in the human genome. *Science*, 296:2225–2229, 2002.
- [63] Y. Ge, S. Dudoit, and T. P. Speed. Resampling-based multiple testing for microarray data analysis. *Test*, 12, 2003.
- [64] H. K. Gjessing and R. T. Lie. Case-parent triads: Estimating single- and double-dose effects of fetal and maternal disease gene haplotypes. *Ann Hum Genet*, 70(3):382–396, 2006.
- [65] H. Gohlke, T. Illig, M. Bahnweg, N. Klopp, et al. Association of the interleukin 1 receptor antagonist gene. *Am Rev Res Crit Care Med*, 169:1217–1223, 2004.
- [66] Goodman. A dirty dozen: Twelve p-value misconceptions. *Seminars in Hematology*, 45(3):135–140, 2008.
- [67] S. N. Goodman. P values, hypothesis tests, and likelihood: implications for epidemiology of a neglected historical debate. *American Journal of Epidemiology*, 137(5):485–501, 1997.
- [68] S. N. Goodman. Toward evidence-based medical statistics. 1: The p value fallacy. *Annals of Internal Medicine*, 130(12):995–1004, 1999.
- [69] S. N. Goodman. Toward evidence-based medical statistics. 2: The bayes factor. *Annals of Internal Medicine*, 130(12):1005–1013, 1999.
- [70] D. Gordon, C. Haynes, C. Johnnidis, S. B. Patel, A. M. Bowcock, and J. Ott. A transmission disequilibrium test for general pedigrees that is robust to the presence of random genotyping errors and any number of untyped parents. *European Journal of Human Genetics*, 12(9):752–761, 2004.

- [71] D. F. Gudbjartsson, T. Thorvaldsson, A. Kong, G. Gunnarsson, and A. Ingolfsdottir. Allegro version 2. *Nature Genetics*, 37:1015–1016, 2005.
- [72] C.-Y. Guo, K. L. Lunetta, A. L. DeStefano, and L. A. Cupples. Combined haplotype relative risk (CHRR): a general and simple genetic association test that combines trios and unrelated case-controls. *Genet Epidemiol*, 33:54–62, 2009.
- [73] J. Hartung. A note on combining dependent tests of significance. *Biometrical Journal*, 41(7):849–855, 1999.
- [74] M. W. Hayford and editors, editors. *Spirit filled live bible*, chapter 1 Corinthians 13, 9-10. Thomas Nelson Publishers, new King James Version edition, 1991.
- [75] J. N. Hirschhorn and M. J. Daly. Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics*, 6:95–108, 2005.
- [76] T. Hoffmann and C. Lange.  $P^2$ BAT: a massive parallel implementation of PBAT for genome-wide association studies in R. *Bioinformatics*, 22(24):3103–3105, 2007.
- [77] S. Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6:65–70, 1979.
- [78] S. Horvath and N. M. Laird. A discordant-sibship test for disequilibrium and linkage: no need for parental data. *Amer J Hum Genet*, 63:1886–1897, 1998.
- [79] S. Horvath, X. Xu, and N. M. Laird. The family based association test method: strategies for studying general genotype-phenotype associations. *European Journal of Human Genetics*, 9:301–306, 2001.
- [80] S. Horvath, X. Xu, S. L. Lake, Silverman E. K., S. T. Weiss, and N. M. Laird. Family-based tests for associating haplotypes with general phenotype data: Application to asthma genetics. *Genet Epidemiol*, 26(1):61–69, 2004.
- [81] J. M. M. Howson, B. J. Barratt, J. A. Todd, and H. J. Cordell. Comparison of population- and family-based methods for genetic association analysis in the presence of interacting loci. *Genet Epidemiol*, 29(1):51–67, 2005.
- [82] L. Hsu, J. R. Starr, Y. Zheng, and S. M. Schwartz. On combining triads and unrelated subjects data in candidate gene studies: an application to data on testicular cancer. *Hum Hered*, 67:88–109, 2009.
- [83] R. Hubbard, M. J. Bayarri, K. N. Berk, and M. A. Carlton. Confusion over measures of evidence (p’s) versus errors (?’s) in classical statistical testing author(s). *The American Statistician*, 57(3):171–182, 2003.
- [84] P. Huber. The behaviour of maximum likelihood estimates under non-standard conditions. In *Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics and Probability*, volume 1, pages 221–233. University of California Press, Berkeley, 1967.
- [85] J. P. A. Ioannidis. Effect of formal statistical significance on the credibility of observational associations. *American Journal of Epidemiology*, 168:374–383, 2008.
- [86] I. Ionita-Laza, G. H. Perry, B. A. Raby, B. Klanderman, C. Lee, N. M. Laird, S. T. Weiss, and C. Lange. On the analysis of copy-number variations in genome-wide association studies: a translation of the family-based association test. *Genet Epidemiol*, 32(3):273–284, 2008.
- [87] T. Iwashita. Asymptotic null and nonnull distribution of Hotelling’s  $t^2$ -statistic under the elliptical distribution. *Journal of Statistical Planning and Inference*, 61(1):85–104, 1997.

- [88] H. A. Katki. Invited commentary: Evidence-based evaluation of p values and bayes factors. *American Journal of Epidemiology*, 168:384–388, 2008.
- [89] K. Köhler and H. Bickeböller. Case-control association tests correcting for population stratification. *Ann Hum Genet*, 70:98–115, 2006.
- [90] M. C. King, G. M. Lee, N. B. Spinner, G. Thomson, and M. R. Wrensch. Genetic epidemiology. *Annual Review of Public Health*, 5:1–52, 1984.
- [91] M. Knapp. Simulation von Kernfamilien mit einer vorgegebenen Anzahl kranker und gesunder Kinder. Mathematischer Hintergrund für das SAS-Makro SIMFAM, 05. 2003.
- [92] M. Knapp. Power comparison of generalizations of the mean test for affected sib pairs in case of incompletely informative markers. *Genet Epidemiol*, 30(4):314–319, 2006.
- [93] M. Knapp and T. Becker. Family based association analysis with tightly linked markers. *Hum Hered*, 56:2–9, 2003.
- [94] N. M. Laird. *Family-based association tests and the FBAT-toolkit, User’s manual*, version 1.7.2 edition, April 2006.
- [95] N. M. Laird. *Family-Based Association Tests and the FBAT toolkit*. [http://www.biostat.harvard.edu/fbat/fbat\\_manual Dec 2007.doc](http://www.biostat.harvard.edu/fbat/fbat_manual_Dec_2007.doc), version 2.02c edition, December 2007.
- [96] N. M. Laird, S. Horvath, and X. Xu. Implementing a unified approach to family-based tests of association. *Genet Epidemiol*, 19:S36–S42, 2000.
- [97] N. M. Laird and C. Lange. Family-based designs in the age of large-scale gene-association studies. *Nature Reviews Genetics*, 7:385–394, 2006.
- [98] S. L. Lake, D. Blacker, and N. M. Laird. Family-based tests of association in the presence of linkage. *Amer J Hum Genet*, 67(6):1515–1525, 2000.
- [99] C. Lange, D. DeMeo, E. K. Silverman, S. T. Weiss, and N. M. Laird. Using the noninformative families in family-based association tests: a powerful new testing strategy. *Amer J Hum Genet*, 73(4):801–811, 2003.
- [100] C. Lange, S. K. Silverman, X. Xu, S. T. Weiss, and N. M. Laird. A multivariate family-based association test using generalized estimating equations: FBAT-GEE. *Biostatistics*, 4(2):195–206, 2003.
- [101] K. Lange, J. S. Sinsheimer, and E. Sobel. Association testing with mendel. *Genet Epidemiol*, 29(1):36–50, 2005.
- [102] W.-C. Lee. Testing for candidate gene linkage disequilibrium using a dense array of single nucleotide polymorphisms in case-parents studies. *Epidemiology*, 13(5):545–551, 2002.
- [103] W.-C. Lee. Genetic association studies of adult-onset diseases using the casespouse and case-offspring designs. *American Journal of Epidemiology*, 158:1023–1032, 2003.
- [104] W.-C. Lee. Case-control association studies with matching and genomic controlling. *Genet Epidemiol*, 27(1):1–13, 2004.
- [105] J. P. Lewinger and S. B. Bull. Validity, efficiency, and robustness of a family-based test of association. *Genet Epidemiol*, 30:62–76, 2006.
- [106] R. C. Lewontin. The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics*, 49(1):49–67, 1964.



- [107] C. J. MacLean, R. B. Martin, P. C. Sham, R. E. Straub, and K. S. Kendler. The trimmed-haplotype test for linkage disequilibrium. *Amer J Hum Genet*, 66:1062–1075, 2000.
- [108] J. Marchini et al. A comparison of phasing algorithms for trios and unrelated individuals. *Amer J Hum Genet*, 78:437–450, 2006.
- [109] E. R. Martin, M. P. Bass, E. R. Hauser, and N. L. Kaplan. Accounting for linkage in family-based tests of association with missing parental genotypes. *Amer J Hum Genet*, 73(5):1016–1026, 2003.
- [110] L. M. McIntyre, E. R. Martin, K. L. Simonsen, and N. L. Kaplan. Circumventing multiple testing: a multilocus Monte Carlo approach to testing for association. *Genet Epidemiol*, 19(1):18–29, 2000.
- [111] G. Mendel. Versuche über Pflanzenhybriden. *Verhandlungen des Naturforschenden Vereines in Brünn*, IV:3–47, 1866.
- [112] G. Montana and J. K. Pritchard. Statistical tests for admixture mapping with case-control and cases-only data. *Amer J Hum Genet*, 75(5):771–789, 2004.
- [113] N. E. Morton. Fifty years of genetic epidemiology with special reference to japan. *Journal of Human Genetics*, 51(4):269–277, 2006.
- [114] A. A. Motsinger-Reif, D. M. Reif, T. J. Fanelli, and M. D. Ritchie. A comparison of analytical methods for genetic association studies. *Genet Epidemiol*, 32:767–778, 2008.
- [115] G. S. Mudholkar and D. K. Srivastava. Robust analogs of hotelling’s two-sample  $t^2$ . *Communication in Statistics - Theory and Methods*, 29(12):2717–2749, 2000.
- [116] J. Neyman and E. Pearson. On the use and interpretation of certain test criteria for purposes of statistical inference. *Philos Trans R Soc Lond A*, 231:289–337, 1933.
- [117] K. K. Nicodemus, A. Luna, and Y. Y. Shugart. An evaluation of power and type I error of single SNP transmission disequilibrium based statistical methods under varying family structures, missing parental data and population stratification. *Amer J Hum Genet*, 80(1):178–185, 2007.
- [118] M. Nothnagel. *The Definition of Multilocus Haplotype Blocks and Common Diseases*. PhD thesis, Humboldt-Universität, Berlin, 2004.
- [119] R. M. Pfeiffer, D. Pee, and M. T. Landi. On combining family and case-control studies. *Genet Epidemiol*, 32(7):638–646, 2008.
- [120] J. K. Pritchard and P. Donnelly. Case-control studies of association in structured or admixed populations. *Theoretical Population Biology*, 60(3):215–220, 2001.
- [121] J. K. Pritchard, M. Stephens, N. A. Rosenberg, and P. Donnelly. Association mapping in structured populations. *Amer J Hum Genet*, 67:170–181, 2000.
- [122] S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, and P. C. Sham. PLINK: a toolset for whole-genome association and population-based linkage analysis. *Amer J Hum Genet*, 81(3):559–575, 2007.
- [123] D. Qian and D. C. Thomas. Genome scan of complex traits by haplotype sharing correlation. *Genet Epidemiol*, 21(S1):S582–S587, 2001.
- [124] D. J. Quian. Haplotype sharing corellation analysis using family data: A comparison with family-based association test in the presence of allelic heterogeneity. *Genet Epidemiol*, 27(1):43–52, 2004.

- [125] D. Rabinowitz and N. M. Laird. A unified approach to adjusting association tests for population admixture with arbitrary pedigree structure and arbitrary missing marker information. *Hum Hered*, 50(4):227–233, 2000.
- [126] C. S. Rakovski, X. Xu, R. Lazarus, D. Blacker, and N. M. Laird. A new multimarker test for family-based association studies. *Genet Epidemiol*, 31(1):9–17, 2007.
- [127] N. Risch. Linkage strategies for genetically complex traits. II. The power of affected relative pairs. *Amer J Hum Genet*, 46:229–241, 1990.
- [128] N. Risch and K. Merikangas. The future of genetic studies of complex human diseases. *Science*, 273:1516–1517, 1996.
- [129] N. Risch and J. Teng. The relative power of family-based and case-control designs for linkage disequilibrium studies of complex human diseases I. DNA pooling. *Genome Research*, 8(12):1273–1288, 1998.
- [130] P. D. Sasieni. From genotypes to genes: doubling the sample size. *Biometrics*, 53:1253–1261, 1997.
- [131] D. J. Schaid. General score tests for associations of genetic markers with disease using cases and their parents. *Genet Epidemiol*, 13:423–449, 1996.
- [132] D. J. Schaid. Relative efficiency of ambiguous vs. directly measured haplotype frequencies. *Genet Epidemiol*, 23(4):426–443, 2002.
- [133] D. J. Schaid. Genetic epidemiology and haplotypes. *Genet Epidemiol*, 27:317–320, 2004.
- [134] D. J. Schaid, S. K. McDonnell, S. J. Hebring, J. M. Cunningham, and S. N. Thibodeau. Nonparametric tests of association of multiple genes with human disease. *Amer J Hum Genet*, 76:780–793, 2005.
- [135] N. J. Schork, L. R. Cardon, and X. Xu. The future of genetic epidemiology. *Trends in Genetics*, 14(7):266–272, 1998.
- [136] S. G. Schwab et al. Evaluation of a susceptibility gene for schizophrenia on chromosome 6p by multipoint affected sib-pair linkage analysis. *Nature Genetics*, 11:325–327, 1995.
- [137] S. G. Schwab et al. Support for association of schizophrenia with genetic variation in the 6p22.3 gene, dysbindin, in sib-pair families with linkage and in additional sample of triad families. *Amer J Hum Genet*, 72:185–190, 2003.
- [138] H. Seltman, K. Roeder, and B. Devlin. Transmission/disequilibrium test meets measured haplotype analysis: family-based association analysis guided by evolution of haplotypes. *Amer J Hum Genet*, 68:1250–1263, 2001.
- [139] H. Seltman, K. Roeder, and B. Devlin. Evolutionary-based association analysis using haplotype data. *Genet Epidemiol*, 25(1):48–58, 2003.
- [140] Q. Y. Sha, J. P. Dong, R. F. Jiang, H.-S. Chen, and S. Zhang. Haplotype sharing transmission/disequilibrium tests that allow for genotyping errors. *Genet Epidemiol*, 28(4):341–351, 2005.
- [141] M. Shi, D. M. Umbach, and C. R. Weinberg. Identification of risk-related haplotypes with the use of multiple SNPs from nuclear families. *Amer J Hum Genet*, 81:53–66, 2007.
- [142] M. Slatkin. Linkage disequilibrium - understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics*, 9:477–485, 2008.

- [143] R. S. Spielman and W. J. Ewens. The TDT and other family-based tests for linkage disequilibrium and association. *Amer J Hum Genet*, 59:983–989, 1996.
- [144] R. S. Spielman and W. J. Ewens. A sibship test for linkage in the presence of association: the sib transmission/disequilibrium test. *Amer J Hum Genet*, 62:450–458, 1998.
- [145] R. S. Spielman, R. E. McGinnis, and W. J. Ewens. The transmission test for linkage disequilibrium: the insulin gene and insuline dependent diabetes mellitus (IDDM). *Amer J Hum Genet*, pages 506–516, 1993.
- [146] D. K. Srivastava and G. S. Mudholkar. Trimmed  $\tilde{T}^2$ : A robust analog of hotelling’s  $t^2$ . *Journal of Statistical Planning and Inference*, 97:343–358, 2001.
- [147] A. E. Stark. Stages in the evolution of the Hardy-Weinberg law. *Genetics and Molecular Biology*, 29(4):589–594, 2006.
- [148] J. R. Starr, L. Hsu, and S. M. Schwartz. Performance of the log-linear approach to case-parent triad data for assessing maternal genetic associations with offspring disease: type i error, power, and bias. *Am J Epidemiol*, 161(2):196–204, 2005.
- [149] R. E. Straub et al. Genetic variation in the 6p22.3 gene DTNBP1, the human ortholog of the mouse dysbindin gene, is associated with schizophrenia. *Amer J Hum Genet*, 71:337–348, 2002.
- [150] B. K. Suarez, J. P. Rice, and T. Reich. The generalized sib-pair IBD distribution: Its use in the detection of linkage. *Ann Hum Genet*, 42:87–94, 1978.
- [151] F. Sun, W. D. Flanders, Q. Yang, and M. J. Khoury. A new method for estimating the risk ratio in studies using case-parental control design. *American Journal of Epidemiology*, 148:902–909, 1999.
- [152] H. K. Tabor, N. J. Risch, and R. M. Myers. Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nature Reviews Genetics*, 3:391–397, 2002.
- [153] J. D. Terwilliger and J. Ott. *Handbook of human genetic linkage*. The John Hopkins University Press, 1994.
- [154] M. L. Tiku and N. Balakrishnan. Robust Hotelling-type  $t^2$ -statistics. *Biometrical Journal*, 30(3):283–293, 1988.
- [155] H. K. Tiwari, J. Barnholtz-Sloan, N. Wineinger, M. A. Padilla, L. K. Vaughan, and D. B. Allison. Review and evaluation of methods correcting for population stratification with a focus on underlying statistical principles. *Hum Hered*, 66:67–86, 2008.
- [156] J. Wakefield. Bayes factors for genome-wide association studies: Comparison with p-values. *Genet Epidemiol*, 33(1):79–86, 2009.
- [157] W. Y. S. Wang, B. J. Barrat, D. G. Clayton, and J. A. Todd. Genome-wide association studies: theoretical and practical concerns. *Nature Reviews Genetics*, 6:109–118, 2005.
- [158] C. R. Weinberg. It’s time to rehabilitate the p-value. *Epidemiology*, 12:288–290, 2001.
- [159] P. H. Westfall and S. S. Young. *Resampling-based multiple testing: examples and methods for P-value adjustment*. Wiley and Sons, New York, 1993.
- [160] H. White. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48:474–486, 1980.
- [161] J. C. Whittaker and A. P. Morris. Family-based tests of association and/or linkage. *Ann Hum Genet*, 65:407–419, 2001.

- [162] A. S. Whittemore and J. Halpern. Genetic association tests for family data with missing parental genotypes: A comparison. *Genet Epidemiol*, 25(1):80–91, 2003.
- [163] S. R. Wilson. On extending the transmission/disequilibrium test (TDT). *Ann Hum Genet*, 61(2):151–161, 1997.
- [164] M. Xiong and L. Jin. Combined linkage and linkage disequilibrium mapping for genome screens. *Genet Epidemiol*, 19(3):211–234, 2000.
- [165] M. Xiong, J. Zhao, and E. Boerwinkle. Generalized  $T^2$  test for genome association studies. *Amer J Hum Genet*, 70(5):1257–1268, 2002.
- [166] X. Xu, C. Rakovski, X. Xiping, and N. Laird. An efficient family-based association test using multiple markers. *Genet Epidemiol*, 30(7):620–626, 2006.
- [167] K. Yu, J. Xu, D. C. Rao, and M. Province. Using tree-based recursive partitioning methods to group haplotypes for increased power in association studies. *Ann Hum Genet*, 69:577–589, 2005.
- [168] K. Yu, J. Xu, C. Xiong, P. An, and M. A. Province. Global Transmission/Disequilibrium Tests based on haplotype sharing in multiple candidate genes. *Genet Epidemiol*, 29:323–335, 2005.
- [169] K. Yu, S. L. Zhang, I. Borecki, A. Kraja, C. Xiong, R. Myers, and M. Province. A haplotype similarity based transmission/disequilibrium test under founder heterogeneity. *Ann Hum Genet*, 69(4):455–467, 2005.
- [170] D. V. Zaykin et al. Testing association of statistical inferred haplotypes with discrete and continuous traits in samples of unrelated individuals. *Hum Hered*, 53(2):79–91, 2002.
- [171] D. V. Zaykin, L. A. Zhivotovsky, W. Czika, S. Shao, and R. D. Wolfinger. Combining p-values in large-scale genomics experiments. *Pharmaceutical Statistics*, 3:217–226, 2007.
- [172] D. V. Zaykin, L. A. Zhivotovsky, P. H. Westfall, and B. S. Weir. Truncated product method for combining P-values. *Genet Epidemiol*, 22(2):170–185, 2002.
- [173] K. Zhang and H. Zhao. A comparison of several methods for haplotype estimation and haplotype reconstruction for tightly linked markers from general pedigrees. *Genet Epidemiol*, 30(5):423–437, 2006.
- [174] S. Zhang, Q. Sha, H.-S. Chen, J. Dong, and R. Jiang. Transmission/disequilibrium test based on haplotype sharing for tightly linked markers. *Amer J Hum Genet*, 73(3):566–579, 2003.
- [175] X. Zhang, K. Roeder, G. Wallstrom, and B. Devlin. Integration of association statistics over genomic regions using Bayesian adaptive regression splines. *Human Genomics*, 1(1):20 – 29, 2003.
- [176] H. Zhao. Family-based association studies. *Statistical Methods in Medical Research*, 9:563 – 587, 2000.
- [177] H. Zhao, S. Zhang, K. R. Merikangas, M. Trixler, D. B. Wildenauer, F. Sun, and K. K. Kidd. Transmission/disequilibrium tests using multiple-tightly linked markers. *Amer J Hum Genet*, 67:936–946, 2000.
- [178] J. H. Zhao. gap: genetic analysis package. *Journal of Statistical Software*, 23(8):1–18, 2007.
- [179] K. T. Zondervan and L. R. Cardon. The complex interplay among factors that influence allelic association. *Nature Reviews Genetics*, 5:89–100, 2004.
- [180] G. Y. Zou. Statistical methods for the analysis of genetic association studies. *Ann Hum Genet*, 70(2):262–276, 2006.

# Abbildungsverzeichnis

1.1	Beispiel: Phasenauflösung ohne und mit elterlicher Genotypinformation . . . . .	8
1.2	Crossover zwischen dem 1. und 2. Locus . . . . .	9
1.3	Haplotype Relative Risk . . . . .	13
1.4	TDT: Vergleich der Anzahl übertragener / nicht übertragener Allele A auf das kranke Kind . . . . .	13
1.5	Kopplung und Assoziation (LD) . . . . .	14
2.1	Krankheitslocus und Markerloci . . . . .	19
2.2	$H_0$ -Modell NMod3 . . . . .	23
4.1	Binäre Codierung der Markerkombinationen . . . . .	41
5.1	Ursprünglicher Ablauf der Simulation . . . . .	48
6.1	Fehler 1. Art: 2500 Trios, NMod 3, Tests auf Assoziation, links: FBAT, rechts: Famhap / Unphased . . . . .	60
6.2	Fehler 1. Art: Hotel_Fam, 2500 bzw. 200 Trios, NMod 3 . . . . .	60
6.3	Fehler 1. Art, FBAT, NMod 2, 1000 Trios; Test von $H_0^L$ (links) vs. $H_0^A$ (rechts) . . . . .	61
6.4	Fehler 1. Art, Famhap und Unphased: NMod 2, 200 Trios; Test auf $H_0^L$ (links) vs. $H_0^A$ (rechts) . . . . .	61
6.5	Test auf Kopplung: NMod 2, 1000 AA-Familien; links: FBAT, rechts: Famhap und Unphased . . . . .	62
6.6	Nullhypothese: Test zum Fehlerniveau $\alpha = 0.01$ , links: FBAT, 1000 Trios, NMod 2; rechts: Hotel_Fam, 200 Trios NMod 3 . . . . .	63
6.7	Fehler 1. Art, overall; 200 Trios, NMod 2 . . . . .	64
6.8	Fehler 1. Art, overall; 1000 AA-Familien, NMod 1 . . . . .	64
6.9	Fehler 1. Art, overall; 2500 AAUU-Familien, NMod 2 . . . . .	65
6.10	Fehler 1. Art, Hotel_Fam, NMod 2, 200 AA-Familien: systematisches Verhalten? Wiederholung der Analyse mit anderen Stichproben . . . . .	66
6.11	Fehler 1. Art: Famhap / Unphased, links: NMod 3, 200 AAUU-Familien; rechts: NMod 2, 1000 AA-Familien, <code>unphased -model full -rare 0.05, 0.01, 0</code> . . . . .	67
6.12	Unphased mit Option <code>-rare 0.05</code> (links) bzw. ohne diese Option (rechts) . . . . .	68
6.13	Markerkombinationen Konfiguration KMod 1 . . . . .	69
6.14	Rezessives Modell KMod 1, AA-Familien . . . . .	69
6.15	Markerkombinationen Modell KMod 2 . . . . .	72
6.16	Power von Famhap und Unphased, rezessives Modell KMod 2, AAUU-Familien; links: Test auf Assoziation, rechts: Test auf Kopplung . . . . .	72
6.17	Markerkombinationen Modell KMod 3 . . . . .	72
6.18	Power von Famhap und Unphased, multiplikatives Modell KMod 1, AA-Familien; links: Test auf Assoziation, rechts: Test auf Kopplung . . . . .	74
6.19	Power von FBAT: Dominante Modelle KMod 1 und 2, Trios . . . . .	75
6.20	Power von FBAT, multiplikative Modelle KMod 2/3, AA-Familien, Test auf $H_0^L$ . . . . .	75

6.21	Powervergleich unter $H_1$ , KMod 1 . . . . .	78
6.22	Powervergleich unter $H_1$ , Modelle KMod 3 . . . . .	79
6.23	Powervergleich unter $H_1$ , Modelle KMod 2 . . . . .	80
6.24	Powervergleich, $\alpha = 0.01$ : rezessives Modell KMod 1, Trios . . . . .	81
6.25	Powervergleich, $\alpha = 0.01$ : multiplikatives Modell KMod 1, AAUU-Familien . . . . .	82
6.26	Fehler 1. Art: Famhap / Unphased, H0_A, NMod 1, 1000 AA-Familien; links: Analyse mit kompletten, rechts: mit unvollständigen Genotyp-Daten . . . . .	83
6.27	Fehler 1. Art: Hotel_Fam, NMod 3, 200 AAUU-Familien; links: Analyse mit kompletten, rechts: mit unvollständigen Genotyp-Daten . . . . .	84
6.28	Fehler 1. Art: FBAT, H0_A, NMod 1, 2500 AA-Familien; links: Analyse mit kompletten, rechts: mit unvollständigen Genotyp-Daten . . . . .	84
6.29	Powerverlust durch 'Missings': DOM KMod 1, AAUU-Familien . . . . .	85
6.30	Powerverluste, Test auf H0_L: FBAT, Famhap und Unphased, MUL KMod 3, Trios . . . . .	86
6.31	Analyse mit 'Missings', Test auf H0_L, rezessives Modell KMod 1, 500 AA-Familien . . . . .	87
6.32	Fehler 1. Art unter Stratifikation: FBAT, links H0_L, rechts H0_A . . . . .	88
6.33	Fehler 1. Art unter Modell S 2: links Famhap, rechts Unphased . . . . .	89
6.34	Fehler 1. Art unter Modell S 3: links Hotel_Fam, rechts Transmit . . . . .	89
6.35	FBAT: Anzahl benötigter Permutationen unter $H_1$ , NMod 2; links: 2500 Trios, rechts: 200 AAUU-Familien . . . . .	90
6.36	FBAT, minfreq-Optionen, REZ, KMod 1, AA-Familien . . . . .	91
6.37	FBAT, minfreq-Optionen, MUL, KMod 2, AA-Familien . . . . .	92
6.38	Overall-Vergleich: FBAT mit Option -minfreq 0.01 . . . . .	92
6.39	Unphased, rare-Optionen . . . . .	93
6.40	Vergleich der Familienstrukturen F22 - F20 und F10, links: FBAT: MUL KMod 3, rechts: Hotel_Fam: REZ KMod 1 . . . . .	94
6.41	Vergleich der Familienstrukturen F22 - F20 und F10, Famhap; links: rezessives Modell NMod 1, rechts: multiplikatives Modell NMod 3 . . . . .	95
6.42	Vergleich der Familienstrukturen F22 - F20 und F10: Unphased, links: Rezessivmodell KMod 1, rechts: additives Modell KMod 2 . . . . .	96
6.43	Vergleich der overall-Power für 300 F22- F20 bzw. F10-Familien, additives Modell KMod 1 . . . . .	96
6.44	Vergleich der overall-Power für 500 F22- F20 und F10-Familien, additives Modell KMod 2 . . . . .	97
6.45	Verteilung der p-Werte für ausgewählte Verfahren . . . . .	98

# Tabellenverzeichnis

1.1	Kopplungs-Ungleichgewicht (LD) . . . . .	10
2.1	Genetische Modelle . . . . .	21
2.2	Markerkonfiguration KMod 1 . . . . .	21
2.3	Markerkonfiguration KMod 2 . . . . .	21
2.4	Markerkonfiguration KMod 3 . . . . .	22
2.5	Stichprobenumfänge für $H_1$ . . . . .	22
2.6	Modelle S1 und S2 unter Stratifikation . . . . .	25
4.1	Genotyp- und Haplotyp-Codierung . . . . .	42
4.2	Ausgabe der vier Programme . . . . .	46
5.1	Relative Laufzeiten unter $H_0$ , ermittelt aus je einer Replikation der 27 Szenarien . . . . .	52
5.2	Relative Laufzeiten unter $H_1$ , ermittelt aus je einer Replikation der 36 Szenarien . . . . .	52
5.3	Liste der 20 getesteten Statistiken, die durch verschiedene Programmbefehle bzw. Optionen ausgewählt werden können . . . . .	55
5.4	Job-Unterbrechungen durch äußere Einflüsse . . . . .	56
5.5	Anteil fehlender Genotypen . . . . .	57
5.6	Analyse mit 'Missings': Auswahl der Szenarien für die Analyse . . . . .	57
6.1	Benötigte Laufzeiten (hh:mm:ss): $H_0$ und $H_1$ mit fehlenden Genotypen . . . . .	58
6.2	KMod 1: Power der Verfahren an der Markerkombination 4 . . . . .	71
6.3	Vergleich Bonferroni-korrigierter min-p-Werte, Anzahl Szenarien, in denen die Verfahren die höchste Power erreichen . . . . .	77
6.4	Power von Unphased für ausgewählte Modelle . . . . .	79
6.5	Overall-Power: Anzahl Szenarien, in denen die Verfahren die höchste Power erreichen	80
6.6	mittlerer Powerverlust durch 'Missings' für ausgewählte Verfahren . . . . .	86
6.7	Assoziation durch Stratifikation: Anteil signifikanter Tests . . . . .	88
6.8	Powervergleich für AA- und AAUU-Familien für einige Szenarien . . . . .	94
6.9	Analyse realer Daten . . . . .	100
6.10	Markerkombinationen mit den 12 kleinsten p-Werten . . . . .	101
A.1	Pilot-Powerstudie . . . . .	123
A.2	Pilot-Powerstudie, Fortsetzung . . . . .	124
A.3	Laufzeitabschätzung (1 Replikation) unter $H_0$ . . . . .	125
A.4	Laufzeitabschätzung (1 Replikation) unter $H_1$ mit fehlenden Genotypen . . . . .	126
A.5	Benötigte Rechenzeit unter $H_0$ (hh:mm:ss) mit fehlenden Genotypen . . . . .	127
A.6	Ermittelte Laufzeiten unter $H_1$ (hh:mm:ss) mit fehlenden Genotypen . . . . .	127
A.7	Power: Markerkombination 14, Tests auf Assoziation zum Niveau $\alpha = 0.05$ . . . . .	128
A.8	Power: Markerkombination 14, Tests auf Kopplung zum Niveau $\alpha = 0.05$ . . . . .	129
A.9	Power: Markerkombination 30, Tests auf Assoziation zum 5%-Niveau . . . . .	130

A.10 Power: Markerkombination 30, Tests auf Kopplung zum 5%-Niveau . . . . .	131
A.11 Vergleich der overall-Power: Tests auf Assoziation, $\alpha = 0.05$ . . . . .	132
A.12 Vergleich der overall-Power: Tests auf Kopplung, $\alpha = 0.05$ . . . . .	133
A.13 Vergleich der overall-Power: Tests auf Assoziation, $\alpha = 0.01$ . . . . .	134
A.14 Vergleich der overall-Power: Tests auf Kopplung, $\alpha = 0.01$ . . . . .	135
A.15 Overall-Power der Verfahren mit 'Missings', KMod 1 . . . . .	136
A.16 Power der Verfahren mit 'Missings', KMod 2 und 3 . . . . .	137
A.17 Analyse realer Daten: Signifikante Markerkombinationen . . . . .	138
A.18 Abkürzungen . . . . .	139



# Anhang A

## A.1 Übersicht über zusätzliches Material

Im Hauptteil der Arbeit wurden aus der Fülle von Grafiken nur solche ausgewählt, die charakteristische Eigenschaften darstellen. Der Leser soll aber in die Lage versetzt werden, sich in den angefügten pdf-Dokumenten weitere Diagramme ansehen zu können. Das zusätzliche Material der Dateien `H0_mk+Bonf.pdf`, `H1_mk+bonf.pdf`, `Miss+Strat.pdf` sowie `Weitere.pdf` folgt dem Aufbau dieser Arbeit. Die Struktur der Dokumente ist folgende:

### **H0\_mk+Bonf.pdf** Fehler 1. Art für die 27 Szenarien

- Markerkombinationen
  - Diagramme zum Fehlerniveau  $\alpha = 0.05$ 
    - Stichprobenumfang  $N = 200$ : je 45 Grafiken (je 5 zu jedem Szenarium)
    - $N = 1000$
    - $N = 2500$
  - Grafiken zum Fehlerniveau  $\alpha = 0.01$ 
    - $N = 200$ : je 45 Grafiken
    - $N = 1000$
    - $N = 2500$
  - Ausgewählte Grafiken zur Unphased-Option `-rare`
- Fehler 1. Art; „Overall“-p-Werte
  - 27 Diagramme zum Fehlerniveau  $\alpha = 0.05$
  - 27 Diagramme zum Fehlerniveau  $\alpha = 0.01$

### **H1\_mk+Bonf.pdf** Power für die 36 Szenarien der Alternativhypothese

- Tests zum 5%-Niveau
  - Power für alle Markerkombinationen
    - Konfiguration KMod 1: je 5 Grafiken für jedes der 12 Szenarien
    - analog KMod 2
    - ebenso KMod 3
  - „Overall-Power“
- Tests zum 1%-Niveau
  - Power für alle Markerkombinationen
    - KMod 1, KMod 2 und 3 wie oben (je 60 Grafiken)
  - Overall-Power
- Overall-Power: alle 4 Programme Bonferroni-korrigierte „beste“ p-Werte
  - je 36 Diagramme (je Szenarium eins)

### **Miss\_Strat** Analyse mit fehlenden Genotypen bzw. unter Stratifikation

- Analyse unvollständiger Daten
  - Fehler 1. Art für alle Markerkombinationen, wie Analyse mit kompletten Daten, je 5 Grafiken für 27 Szenarien,  $\alpha = 0.05$
  - Darstellung des Powerverlusts durch fehlende Genotypen
    - FBAT: je 6 FBAT-Diagramme für die drei Konfigurationen KMod 1 - 3
    - Famhap / Unphased: analog 18 Grafiken
    - Hotel\_Fam: nur 6 Diagramme für KMod 1
  - Overall-Power: je 1 Grafik für 18 Szenarien
- Analyse unter Stratifikation, nur Fehler 1. Art
  - FBAT: 3 Diagramme (eine für jedes Modell)
  - analog je 3 Grafiken für Famhap, Unphased, Hotel\_Fam und Transmit

### **Weitere.pdf**

- Analyse der Familienstrukturen
  - Famhap: 24 Diagramme (je 4 für 6 Szenarien)
  - FBAT:  $3 \times 8$  Grafiken für KMod 1,  $3 \times 6$  Diagramme für KMod 2 bzw. 3
  - je 3 KMod 1-Szenarien mit je 4 Hotel\_Fam-Statistiken
  - Unphased: 3 KMod 1-Szenarien mit je 4, 3 KMod 2/3-Szenarien mit je 2 Unphased-Statistiken
  - Overall-Statistiken: Powervergleich (6 Szenarien)
- Optionen `minfreq` bzw. `-rare`
  - FBAT `minfreq 0.05 0.01 0`: 12 Power-Diagramme
  - Unphased `-rare 0.05 0.01 0`: 8 Diagramme
- 82 Histogramme der p-Werte (ausgewählte Szenarien, typische Markerkombinationen)
- Anzahl FBAT-Permutationen: 36 Histogramme (`-rare`-Optionen)

## **A.2 Zusätzliche Tabellen**

Modell	Parameter	200	300	400	500	600	700	800	900	1000	1100	1200	1300	1400	1500	1600	1700	1800
KMod1	MUL10				0.60	0.75	0.81	0.87		0.96								
	MUL20	0.81	0.97		1.00					1.00								
	MUL22	0.76	0.90	0.98	0.98					1.00								
KMod2	MUL10									0.66	0.77	0.81	0.86		0.88			
	MUL20	0.44	0.74	0.91	0.98					1.00								
	MUL22	0.45	0.72	0.87	0.97					1.00								
KMod3	MUL10									0.46					0.72	0.75	0.74	0.83
	MUL20		0.58	0.74	0.84	0.92				0.99								
	MUL22	0.40	0.60	0.77	0.89					1.00								
KMod1	REZ10									0.31					0.45			
	REZ20			0.73	0.80	0.94				1.00								
	REZ22				0.70	0.87	0.91			1.00								
KMod2	REZ10									0.12					0.20			
	REZ20				0.49			0.64	0.83	0.84	0.93							
	REZ22				0.37			0.71	0.80	0.83	0.87							
KMod3	REZ10									0.06					0.13			
	REZ20									0.56			0.73	0.84	0.85	0.90		
	REZ22									0.74	0.84	0.89	0.92		0.94			

Tabelle A.1: Pilot-Powerstudie

Modell	Parameter	200	300	400	500	600	700	800	900	1000	1100	1200	1300	1400	1500	1600	1700	1800
KMod1	DOM10		0.63	0.71	0.86	0.87				0.97								
	DOM20	0.55	0.77	0.89	0.96					1.00								
	DOM22	0.48	0.73	0.91	0.96					1.00								
KMod2	DOM10								0.75	0.79	0.82	0.90			0.97			
	DOM20			0.59	0.79	0.86	0.92			0.99								
	DOM22			0.68	0.79	0.90	0.95			1.00								
KMod3	DOM10									0.45				0.75	0.79	0.83	0.84	
	DOM20				0.65	0.70	0.81	0.87		0.95								
	DOM22				0.66	0.81	0.86	0.96		0.98								
KMod1	ADD10									0.77	0.80	0.83	0.84		0.96			
	ADD20	0.64	0.86	0.92	0.96					1.00								
	ADD22	0.56	0.76	0.86	0.93					1.00								
KMod2	ADD10									0.45					0.71	0.75	0.77	0.81
	ADD20	0.34	0.61	0.79	0.94					1.00								
	ADD22		0.50	0.67	0.86	0.91				1.00								
KMod3	ADD10									0.24					0.48			
	ADD20			0.62	0.77	0.85	0.90			0.99								
	ADD22				0.72	0.88	0.97	0.98		1.00								

Modell	Parameter	1900	2000	2500	3000	4000	6000	8000	10000	11000
KMod1	REZ10		0.67	0.78	0.86					
KMod2	REZ10				0.35		0.80			
KMod3	REZ10					0.29		0.68		0.87
KMod2	ADD10	0.83	0.86							
KMod3	ADD10		0.65	0.81	0.87					

Tabelle A.2: Pilot-Powerstudie, Fortsetzung

Modell NMod 1		Famhap	FBAT	Unphased	Hotel_fam
<b>F10</b>	Anzahl Familien				
	200	0:00:18 <sup>a</sup>	0:00:54	0:01:45	0:03:21
	1000	0:01:40	0:05:24	0:09:20	0:16:15
	2500	0:04:33	0:12:44	0:24:04	0:39:58
<hr/>					
<b>F20</b>	200	0:00:33	0:01:55	0:02:47	0:03:25
	1000	0:03:07	0:09:47	0:14:46	0:16:26
	2500	0:08:03	0:25:53	0:39:44	0:40:06
	<hr/>				
<b>F22</b>	200	0:00:28	0:03:00	0:36:00	0:03:23
	1000	0:02:58	0:16:45	3:22:15	0:16:35
	2500	0:08:11	0:45:35	<b>8:54:57</b>	0:40:31
	<hr/>				
Modell NMod 2					
<b>F10</b>	200	0:00:24	0:00:53	0:01:34	0:03:21
	1000	0:01:53	0:04:53	0:08:05	0:16:19
	2500	0:04:58	0:11:54	0:20:23	0:39:52
	<hr/>				
<b>F20</b>	200	0:00:40	0:01:53	0:02:24	0:03:26
	1000	0:03:38	0:10:48	0:11:39	0:16:25
	2500	0:09:56	0:25:27	0:30:03	0:40:12
	<hr/>				
<b>F22</b>	200	0:00:27	0:03:38	0:25:42	0:03:25
	1000	0:03:04	0:18:21	2:09:33	0:16:34
	2500	0:08:42	0:45:53	5:50:42	0:40:32
	<hr/>				
Modell NMod 3					
<b>F10</b>	200	0:00:19	0:00:55	0:01:50	0:03:22
	1000	0:01:49	0:04:57	0:09:08	0:16:16
	2500	0:04:39	0:12:13	0:24:50	0:39:55
	<hr/>				
<b>F20</b>	200	0:00:20	0:01:32	0:02:24	0:03:25
	1000	0:01:34	0:08:07	0:12:34	0:16:27
	2500	0:04:51	0:20:32	0:34:43	0:40:12
	<hr/>				
<b>F22</b>	200	<b>0:00:11</b>	0:01:51	0:18:47	0:03:23
	1000	0:00:50	0:09:19	1:47:30	0:16:36
	2500	0:02:31	0:23:26	4:46:24	0:40:24
	<hr/>				
<b>Summe</b>		1:20:37	5:28:29	32:23:53	9:00:06
<b>Faktor</b>		1	4.1	24.1	6.7

Tabelle A.3: Laufzeitabschätzung (1 Replikation) unter  $H_0$

<sup>a</sup> Laufzeit in hh:mm:ss

	H1-Modell	Fam-Typ	Famhap	FBAT	Unphased	Hotel_Fam
<b>MUL</b>	KMod1	F10	00:01:06 <sup>a</sup>	00:03:36	00:06:16	00:10:33
		F20	00:00:28	00:02:00	00:02:29	00:03:14
		F22	<b>00:00:25</b>	00:03:16	00:35:28	00:03:13
	KMod2	F10	00:02:25	00:05:45	00:11:06	00:17:55
		F20	00:00:45	00:02:56	00:04:00	00:04:46
		F22	00:00:38	00:05:05	00:49:05	00:04:52
	KMod3	F10	00:03:13	00:08:02	00:15:58	00:25:05
		F20	00:01:19	00:05:10	00:06:20	00:07:55
		F22	00:00:54	00:06:12	01:05:26	00:06:16
<b>REZ</b>	KMod1	F10	00:04:45	00:12:51	00:25:16	00:36:48
		F20	00:01:20	00:05:19	00:06:57	00:07:59
		F22	00:01:30	00:10:07	01:43:17	00:09:18
	KMod2	F10	00:11:56	00:28:42	01:02:27	01:28:54
		F20	00:02:49	00:09:21	00:12:06	00:13:48
		F22	00:02:36	00:14:49	02:39:10	00:14:00
	KMod3	F10	00:22:49	01:11:14	01:58:07	02:40:52
		F20	00:04:40	00:14:19	00:19:25	00:21:36
		F22	00:02:56	00:16:17	<b>02:55:24</b>	00:15:34
<b>DOM</b>	KMod1	F10	00:00:36	00:01:53	00:03:22	00:06:00
		F20	00:00:44	00:02:58	00:03:47	00:04:38
		F22	00:00:39	00:05:14	00:47:28	00:04:50
	KMod2	F10	00:01:47	00:04:54	00:08:19	00:15:07
		F20	00:01:24	00:05:11	00:06:02	00:07:49
		F22	00:01:11	00:08:16	01:16:13	00:07:45
	KMod3	F10	00:02:51	00:07:23	00:12:53	00:22:12
		F20	00:02:08	00:07:04	00:08:10	00:10:44
		F22	00:01:26	00:10:23	01:37:18	00:09:22
<b>ADD</b>	KMod1	F10	00:01:45	00:05:31	00:10:36	00:16:09
		F20	00:00:47	00:02:42	00:04:20	00:04:39
		F22	00:00:38	00:04:54	00:53:53	00:04:50
	KMod2	F10	00:03:28	00:08:42	00:17:34	00:26:39
		F20	00:01:01	00:03:51	00:12:57	00:06:08
		F22	00:00:52	00:06:34	01:11:22	00:06:15
	KMod3	F10	00:04:53	00:10:53	00:23:53	00:36:53
		F20	00:01:21	00:04:56	00:06:33	00:07:55
		F22	00:01:11	00:08:27	01:25:57	00:07:48
<b>Summe</b>			01:34:56	05:34:47	23:41:25	10:58:21
<b>Faktor</b>			1	3.5	15.0	6.9

Tabelle A.4: Laufzeitabschätzung (1 Replikation) unter  $H_1$  mit fehlenden Genotypen

<sup>a</sup> Laufzeit in hh:mm:ss

Modell	Famlientyp	Famhap	FBAT	Hotel_Fam	Unphased
NMod 1	F10	08:17:02	18:37:19	94:25:09	51:41:45
	F20	12:33:01	45:01:41	100:02:17	75:01:09
	F22	– <sup>a</sup>	–	–	–
NMod 2	F10	08:53:00	20:17:15	135:40:49	47:27:52
	F20	14:46:26	39:50:22	53:15:02	66:14:36
	F22	– <sup>b</sup>	–	–	–
NMod 3	F10	08:25:56	19:30:39	135:17:04	48:23:52
	F20	07:28:57	35:30:25	137:02:43	67:08:46
	F22	04:25:26	43:09:35	103:17:05	429:10:16
Summe		64:49:48	221:57:16	759:00:09	785:08:16
Faktor <sup>c</sup>		1	3.4	11.7	12.1

Tabelle A.5: Benötigte Rechenzeit unter  $H_0$  (hh:mm:ss) mit fehlenden Genotypen

<sup>a</sup> Zeiten für Unphased wegen mehrerer Unterbrechungen nicht zu ermitteln

<sup>b</sup> Zeiten für Unphased nicht verfügbar

<sup>c</sup> Faktor gegenüber Famhap

Modell	Famlientyp	Famhap	FBAT	Hotel_Fam	Unphased
MUL	F22	10:35:29	212:21:06	78:18:27	607:14:33
REZ	F20	32:58:25	87:47:00	172:39:40	179:28:09
DOM	F10	37:38:01	43:31:39	161:15:10	87:36:12
ADD	F20	18:49:51	72:17:07	102:23:58	97:54:51
Summe		100:01:46	415:56:52	514:37:15	972:13:45
Faktor		1	4.2	5.1	9.7

Tabelle A.6: Ermittelte Laufzeiten unter  $H_1$  (hh:mm:ss) mit fehlenden Genotypen

Szenarium	N <sup>a</sup>	zhao	mm	hbat_b	hbat_g	hap_pm	gen_pm	full	allM
add1f10	1100	0.839 <sup>b</sup>	0.889	0.595	0.811	<b>0.893<sup>c</sup></b>	0.861	0.161	0.890
add1f20	300	0.786	0.879	0.516	0.752	<b>0.884</b>	0.858	0.481	0.881
add1f22	300	0.733	0.851	0.488	0.715	<b>0.855</b>	0.824	0.677	0.852
add2f10	1800	0.908 <sup>b</sup>	0.056	0.717	<b>0.923</b>	0.056	0.046	0.366	0.054
add2f20	400	<b>0.860</b>	0.045	0.578	0.845	0.048	0.038	0.664	0.043
add2f22	400	<b>0.841</b>	0.042	0.554	0.839	0.041	0.040	0.551	0.046
add3f10	2500	0.969	0.035	0.836	<b>0.982</b>	0.036	0.036	0.533	0.035
add3f20	500	0.936	0.061	0.731	<b>0.941</b>	0.061	0.058	0.813	0.064
add3f22	500	<b>0.919</b>	0.044	0.683	0.912	0.046	0.047	0.902	0.046
dom1f10	400	0.738	0.819	0.467	0.688	0.823	<b>0.885</b>	0.667	0.799
dom1f20	300	0.709	0.816	0.462	0.657	0.817	<b>0.902</b>	0.659	0.801
dom1f22	300	0.701	0.847	0.422	0.696	0.848	<b>0.939</b>	0.695	0.836
dom2f10	1000	0.907	0.053	0.650	<b>0.947</b>	0.055	0.049	0.936	0.055
dom2f20	500	0.837	0.050	0.538	<b>0.844</b>	0.048	0.052	0.836	0.047
dom2f22	500	<b>0.904</b>	0.054	0.631	0.898	0.055	0.050	0.888	0.055
dom3f10	1500	<b>0.957</b>	0.055	0.765	0.985	0.059	0.055	0.978	0.055
dom3f20	700	<b>0.937</b>	0.038	0.671	0.947	0.037	0.052	0.943	0.044
dom3f22	600	<b>0.949</b>	0.042	0.697	0.944	0.041	0.042	0.941	0.043
mul1f10	700	0.836	0.891	0.620	0.794	0.891	0.853	0.691	<b>0.896</b>
mul1f20	200	0.725	0.827	0.447	0.655	0.831	0.781	0.641	<b>0.838</b>
mul1f22	200	0.698	0.811	0.430	0.659	0.812	0.746	0.696	<b>0.829</b>
mul2f10	1200	0.919	0.050	0.705	<b>0.934</b>	0.049	0.044	0.916	0.052
mul2f20	300	<b>0.813</b>	0.053	0.539	0.799	0.056	0.058	<b>0.813</b>	0.056
mul2f22	300	<b>0.830</b>	0.052	0.536	0.798	0.054	0.049	0.824	0.054
mul3f10	1700	0.962	0.047	0.812	0.975	0.047	0.041	<b>0.977</b>	0.048
mul3f20	500	0.969	0.055	0.800	0.973	0.054	0.052	<b>0.978</b>	0.055
mul3f22	400	0.926	0.048	0.700	0.929	0.047	0.050	<b>0.930</b>	0.048
rez1f10	2500	0.848	0.913	0.628	0.798	0.912	<b>0.974</b>	0.822	0.916
rez1f20	500	0.830	0.883	0.559	0.745	0.887	<b>0.925</b>	0.806	0.905
rez1f22	600	0.835	0.916	0.607	0.803	0.929	<b>0.972</b>	0.864	0.940
rez2f10	6000	0.952	0.056	0.786	0.960	0.054	0.056	<b>0.967</b>	0.057
rez2f20	900	0.911	0.048	0.678	0.898	0.046	0.037	<b>0.922</b>	0.049
rez2f22	900	<b>0.941</b>	0.059	0.715	0.903	0.061	0.050	0.927	0.065
rez3f10	11000	0.994	0.057	0.914	0.996	0.059	0.052	<b>0.997</b>	0.058
rez3f20	1400	0.979	0.043	0.839	0.981	0.044	0.044	<b>0.988</b>	0.045
rez3f22	1000	<b>0.960</b>	0.053	0.748	0.932	0.052	0.048	0.952	0.052
1. Rang <sup>d</sup>		8.5	0	0	8	3	6	7.5	3
∈ (0.9, 1.0] <sup>e</sup>		19	2	1	17	2	5	15	3
∈ (0.8, 0.9]		10	10	3	7	10	5	8	8
∈ (0.7, 0.8]		6	0	8	7	0	2	0	1
≤ 0.7		1	24	24	5	24	24	13	24

Tabelle A.7: Power: Markerkombination 14, Tests auf Assoziation zum Niveau  $\alpha = 0.05$

<sup>a</sup> Stichprobenumfang

<sup>b</sup> hellgrau: Power ∈ (0.8, 0.9], grau: Power in ∈ (0.9, 1.0]

<sup>c</sup> Wert des jeweils besten Verfahrens fett markiert

<sup>d</sup> Anzahl erster Plätze unter den Verfahren

<sup>e</sup> Anzahl Szenarien, in denen Power in das angegebene Intervall fällt



Szenarium	zhao_l	mm_l	perm_s	perm_m	hbat_gl	full_l	allM_l
add1f10	0.839 <sup>a</sup>	0.887	0.842	0.627	0.811	0.161	<b>0.890</b>
add1f20	0.807	0.892	0.812	0.603	0.789	0.463	<b>0.896</b>
add1f22	0.756	<b>0.866</b>	0.761	0.556	0.752	0.694	0.864
add2f10	<b>0.908<sup>a</sup></b>	0.053	0.911	0.727	<b>0.924</b>	0.366	0.054
add2f20	<b>0.875</b>	0.051	0.861	0.635	<b>0.875</b>	0.671	0.052
add2f22	0.865	0.049	0.861	0.637	<b>0.873</b>	0.462	0.044
add3f10	0.969	0.035	0.971	0.849	<b>0.980</b>	0.533	0.035
add3f20	0.945	0.072	0.945	0.795	<b>0.953</b>	0.788	0.073
add3f22	<b>0.932</b>	0.052	0.924	0.753	<b>0.932</b>	0.893	0.054
dom1f10	0.738	0.797	0.731	0.481	0.683	0.667	<b>0.799</b>
dom1f20	0.731	<b>0.818</b>	0.730	0.511	0.709	0.692	0.817
dom1f22	0.730	<b>0.848</b>	0.733	0.483	0.712	0.718	<b>0.848</b>
dom2f10	0.907	0.055	0.906	0.661	<b>0.942</b>	0.936	0.055
dom2f20	0.855	0.056	0.836	0.590	<b>0.856</b>	0.784	0.055
dom2f22	<b>0.916</b>	0.061	0.898	0.671	0.906	0.788	0.065
dom3f10	0.957	0.055	0.958	0.772	<b>0.981</b>	0.978	0.055
dom3f20	0.948	0.049	0.939	0.717	<b>0.953</b>	0.878	0.052
dom3f22	<b>0.958</b>	0.048	0.943	0.756	0.948	0.821	0.049
mul1f10	0.836	<b>0.896</b>	0.836	0.631	0.799	0.691	<b>0.896</b>
mul1f20	0.756	<b>0.848</b>	0.761	0.528	0.722	0.655	0.847
mul1f22	0.733	<b>0.846</b>	0.743	0.512	0.726	0.709	0.839
mul2f10	0.919	0.052	0.925	0.714	<b>0.934</b>	0.916	0.052
mul2f20	<b>0.837</b>	0.064	0.826	0.591	0.832	0.795	0.066
mul2f22	<b>0.851</b>	0.060	0.830	0.582	0.843	0.798	0.063
mul3f10	0.962	0.046	0.964	0.822	0.976	<b>0.977</b>	0.048
mul3f20	0.971	0.065	0.968	0.847	<b>0.980</b>	0.969	0.066
mul3f22	<b>0.934</b>	0.052	0.929	0.771	0.933	0.877	0.051
rez1f10	0.848	<b>0.917</b>	0.854	0.641	0.802	0.822	0.916
rez1f20	0.856	0.914	0.847	0.638	0.819	0.833	<b>0.915</b>
rez1f22	0.853	0.935	0.869	0.668	0.859	0.875	<b>0.947</b>
rez2f10	0.952	0.056	0.957	0.797	0.963	<b>0.967</b>	0.057
rez2f20	0.922	0.049	0.918	0.723	<b>0.924</b>	0.905	0.050
rez2f22	<b>0.949</b>	0.071	0.930	0.775	0.939	0.860	0.072
rez3f10	0.994	0.058	0.994	0.922	0.996	<b>0.997</b>	0.058
rez3f20	0.984	0.050	0.983	0.879	<b>0.988</b>	0.974	0.053
rez3f22	<b>0.968</b>	0.056	0.949	0.799	0.954	0.835	0.056
1. Rang	9	7	0	0	12	3	5
$\in (0.9, 1.0]^a$	19	3	18	1	19	9	3
$\in (0.8, 0.9]$	11	8	12	4	9	9	8
$\in (0.7, 0.8]$	6	1	6	12	7	7	1
$\leq 0.7$	0	24	0	19	1	1	24

Tabelle A.8: Power: Markerkombination 14, Tests auf Kopplung zum Niveau  $\alpha = 0.05$

<sup>a</sup> Legende wie in Tab. A.7

Szenarium	zhao	zhaomax	mm	hbat_g	hap_pm	gen_pm	full	allM
add1f10	0.718 <sup>b</sup>	0.393	0.876 <sup>b</sup>	0.780	<b>0.877<sup>a</sup></b>	0.827	0.059	0.876
add1f20	0.671	0.360	0.849	0.728	<b>0.858</b>	0.819	0.060	0.846
add1f22	0.584	0.302	0.808	0.678	<b>0.809</b>	0.773	0.056	0.815
add2f10	0.822	0.500	0.059	<b>0.900</b>	0.060	0.057	0.059	0.059
add2f20	0.728	0.383	0.050	<b>0.823</b>	0.049	0.052	0.059	0.050
add2f22	0.711	0.371	0.041	<b>0.820</b>	0.044	0.037	0.061	0.040
add3f10	0.921	0.536	0.040	<b>0.977</b>	0.041	0.042	0.053	0.038
add3f20	0.847	0.503	0.066	<b>0.925</b>	0.067	0.052	0.063	0.067
add3f22	0.821	0.454	0.051	<b>0.907</b>	0.054	0.050	0.055	0.051
dom1f10	0.604	0.304	0.785	0.505	0.789	<b>0.862</b>	0.307	0.758
dom1f20	0.572	0.282	0.767	0.501	0.775	<b>0.876</b>	0.502	0.762
dom1f22	0.543	0.257	0.809	0.512	0.815	<b>0.907</b>	0.534	0.789
dom2f10	<b>0.819</b>	0.429	0.060	0.840	0.061	0.050	0.791	0.062
dom2f20	0.724	0.346	0.049	0.724	0.049	0.052	<b>0.731</b>	0.048
dom2f22	0.789	0.396	0.044	<b>0.794</b>	0.045	0.048	0.785	0.047
dom3f10	0.892	0.525	0.056	<b>0.940</b>	0.053	0.060	0.929	0.056
dom3f20	0.850	0.442	0.039	0.869	0.038	0.040	<b>0.870</b>	0.042
dom3f22	0.864	0.475	0.040	<b>0.883</b>	0.040	0.044	0.872	0.041
mul1f10	0.734	0.377	0.858	0.759	<b>0.859</b>	0.818	0.049	0.854
mul1f20	0.606	0.290	0.793	0.602	0.801	0.719	0.071	<b>0.805</b>
mul1f22	0.520	0.252	0.777	0.608	0.787	0.696	0.078	<b>0.789</b>
mul2f10	0.836	0.486	0.052	<b>0.920</b>	0.051	0.051	0.056	0.052
mul2f20	0.698	0.351	0.060	<b>0.769</b>	0.062	0.058	0.066	0.064
mul2f22	0.696	0.369	0.065	<b>0.784</b>	0.067	0.072	0.079	0.069
mul3f10	0.918	0.559	0.048	<b>0.962</b>	0.050	0.047	0.046	0.046
mul3f20	0.925	0.564	0.050	<b>0.960</b>	0.049	0.053	0.062	0.052
mul3f22	0.843	0.465	0.053	<b>0.902</b>	0.056	0.058	0.049	0.055
rez1f10	0.755	0.431	0.891	0.792	0.891	<b>0.963</b>	0.049	0.899
rez1f20	0.730	0.388	0.854	0.725	0.852	<b>0.902</b>	0.050	0.875
rez1f22	0.716	0.373	0.884	0.793	0.889	<b>0.951</b>	0.063	0.916
rez2f10	0.903	0.528	0.062	<b>0.953</b>	0.062	0.059	0.054	0.062
rez2f20	0.841	0.453	0.046	<b>0.887</b>	0.046	0.042	0.056	0.047
rez2f22	0.864	0.514	0.049	<b>0.911</b>	0.050	0.052	0.057	0.053
rez3f10	0.973	0.724	0.058	<b>0.996</b>	0.060	0.054	0.052	0.057
rez3f20	0.944	0.568	0.050	<b>0.980</b>	0.052	0.046	0.057	0.047
rez3f22	0.894	0.521	0.045	<b>0.931</b>	0.042	0.040	0.049	0.056
1. Rang <sup>c</sup>	1	0	0	21	4	6	2	2
∈ (0.85, 1.0] <sup>d</sup>	11	0	4	18	6	7	4	5
∈ (0.7, 0.85]	17	1	7	13	6	4	4	7
∈ (0.55, 0.7]	6	3	0	2	0	1	0	0
≤ 0.55	2	32	25	3	24	24	28	24

Tabelle A.9: Power: Markerkombination 30, Tests auf Assoziation zum 5%-Niveau

<sup>a</sup> höchster Wert fett markiert

<sup>b</sup> hellgrau: Power ∈ (0.7, 0.85], grau: Power in ∈ (0.85, 1.0]

<sup>c</sup> Anzahl erster Plätze unter den Verfahren

<sup>d</sup> Anzahl Szenarien, in denen Power in das angegebene Intervall fällt

Szenarium	zhao_l	zhaomax_l	mm_l	perm_s	hbat_gl	full_l	allM_l
add1f10	0.718 <sup>a</sup>	0.393	<b>0.878<sup>a</sup></b>	0.724	0.777	0.059	0.876
add1f20	0.708	0.409	0.860	0.698	0.782	0.056	<b>0.860</b>
add1f22	0.628	0.340	0.830	0.623	0.719	0.057	<b>0.832</b>
add2f10	0.822	0.500	0.059	0.827	<b>0.904</b>	0.059	0.059
add2f20	0.759	0.433	0.057	0.746	<b>0.849</b>	0.058	0.056
add2f22	0.766	0.422	0.044	0.761	<b>0.850</b>	0.053	0.046
add3f10	0.921	0.536	0.038	0.920	<b>0.976</b>	0.053	0.038
add3f20	0.876	0.544	0.076	0.872	<b>0.942</b>	0.070	0.076
add3f22	0.843	0.517	0.060	0.837	<b>0.922</b>	0.052	0.061
dom1f10	0.604	0.304	<b>0.759</b>	0.577	0.510	0.307	0.758
dom1f20	0.610	0.318	0.775	0.599	0.587	0.508	<b>0.783</b>
dom1f22	0.587	0.299	0.815	0.585	0.576	0.566	<b>0.820</b>
dom2f10	0.819	0.429	0.059	0.789	<b>0.825</b>	0.791	0.062
dom2f20	0.758	0.380	0.054	0.739	<b>0.767</b>	0.657	0.054
dom2f22	<b>0.819</b>	0.442	0.053	0.801	0.810	0.657	0.055
dom3f10	0.892	0.525	0.055	0.884	<b>0.935</b>	0.929	0.056
dom3f20	0.871	0.473	0.043	0.861	<b>0.889</b>	0.745	0.046
dom3f22	<b>0.895</b>	0.528	0.049	0.886	<b>0.895</b>	0.677	0.050
mul1f10	0.734	0.377	<b>0.857</b>	0.731	0.765	0.049	0.854
mul1f20	0.635	0.335	0.805	0.622	0.676	0.057	<b>0.816</b>
mul1f22	0.578	0.298	0.797	0.572	0.666	0.077	<b>0.807</b>
mul2f10	0.836	0.486	0.051	0.834	<b>0.922</b>	0.056	0.052
mul2f20	0.737	0.393	0.071	0.719	<b>0.802</b>	0.052	0.075
mul2f22	0.733	0.400	0.076	0.727	<b>0.830</b>	0.073	0.076
mul3f10	0.918	0.559	0.048	0.916	<b>0.962</b>	0.046	0.046
mul3f20	0.943	0.612	0.065	0.936	<b>0.969</b>	0.059	0.060
mul3f22	0.867	0.508	0.058	0.862	<b>0.922</b>	0.049	0.060
rez1f10	0.755	0.431	<b>0.900</b>	0.762	0.790	0.049	0.899
rez1f20	0.754	0.428	0.891	0.747	0.789	0.047	<b>0.893</b>
rez1f22	0.750	0.414	0.917	0.760	0.844	0.065	<b>0.925</b>
rez2f10	0.903	0.528	0.062	0.906	<b>0.956</b>	0.054	0.062
rez2f20	0.863	0.501	0.051	0.858	<b>0.908</b>	0.049	0.053
rez2f22	0.879	0.558	0.051	0.880	<b>0.940</b>	0.055	0.063
rez3f10	0.973	0.724	0.058	0.974	<b>0.997</b>	0.052	0.057
rez3f20	0.958	0.620	0.057	0.954	<b>0.988</b>	0.054	0.052
rez3f22	0.909	0.557	0.052	0.904	<b>0.960</b>	0.052	0.066
1. Rang <sup>a</sup>	1.5	0	4	0	22.5	0	8
∈ (0.85, 1.0] <sup>a</sup>	14	0	6	14	17	1	6
∈ (0.7, 0.85]	16	1	6	15	14	2	6
∈ (0.55, 0.7]	6	5	0	7	4	4	0
≤ 0.55	0	30	24	0	1	29	24

Tabelle A.10: Power: Markerkombination 30, Tests auf Kopplung zum 5%-Niveau

<sup>a</sup> Legende wie in Tab. A.9

Szenarium	zhao	zhaomax	mm	hbat_b	hbat_g	hap_pm	gen_pm	full	allM
add1f10	<b>0.843<sup>b</sup></b>	0.814	0.768	0.772	0.769	0.765	0.727	0.758	0.765
add1f20	<b>0.814</b>	0.777	0.715	0.742	0.726	0.718	0.713	0.737	0.732
add1f22	<b>0.744</b>	0.717	0.661	0.679	0.676	0.651	0.636	0.682	0.667
add2f10	<b>0.824</b>	0.614	0.015	0.536	0.817	0.017	0.018	0.785	0.015
add2f20	<b>0.748</b>	0.507	0.010	0.374	0.670	0.007	0.009	0.667	0.013
add2f22	<b>0.714</b>	0.497	0.007	0.341	0.635	0.009	0.007	0.284	0.010
add3f10	0.832	0.360	0.015	0.288	<b>0.860<sup>b</sup></b>	0.015	0.016	0.138	0.014
add3f20	<b>0.740</b>	0.303	0.011	0.191	0.706	0.012	0.011	0.363	0.012
add3f22	<b>0.717</b>	0.284	0.008	0.172	0.665	0.014	0.015	0.580	0.009
dom1f10	0.751	0.700	0.632	0.656	0.637	0.638	<b>0.784</b>	0.612	0.613
dom1f20	0.732	0.686	0.611	0.648	0.617	0.612	<b>0.771</b>	0.614	0.604
dom1f22	0.763	0.713	0.649	0.674	0.657	0.653	<b>0.813</b>	0.650	0.632
dom2f10	<b>0.832</b>	0.519	0.012	0.427	0.860	0.012	0.014	0.820	0.011
dom2f20	<b>0.747</b>	0.471	0.014	0.320	0.672	0.015	0.018	0.651	0.017
dom2f22	<b>0.840</b>	0.571	0.017	0.389	0.731	0.017	0.017	0.733	0.016
dom3f10	0.815	0.297	0.014	0.241	<b>0.866</b>	0.018	0.018	0.829	0.014
dom3f20	<b>0.766</b>	0.247	0.013	0.175	0.734	0.011	0.022	0.726	0.013
dom3f22	<b>0.787</b>	0.28	0.01	0.186	0.709	0.011	0.015	0.706	0.010
mul1f10	<b>0.834</b>	0.802	0.753	0.764	0.759	0.748	0.734	0.758	0.761
mul1f20	<b>0.753</b>	0.704	0.617	0.627	0.617	0.611	0.615	0.665	0.656
mul1f22	<b>0.709</b>	0.669	0.596	0.611	0.603	0.584	0.579	0.642	0.633
mul2f10	<b>0.838</b>	0.595	0.012	0.520	0.827	0.018	0.012	0.824	0.013
mul2f20	<b>0.703</b>	0.451	0.016	0.309	0.615	0.016	0.020	0.630	0.016
mul2f22	<b>0.725</b>	0.490	0.027	0.319	0.614	0.025	0.029	0.629	0.029
mul3f10	0.848	0.364	0.013	0.260	<b>0.872</b>	0.020	0.018	0.828	0.012
mul3f20	<b>0.859</b>	0.354	0.010	0.233	0.833	0.009	0.013	0.803	0.011
mul3f22	<b>0.753</b>	0.288	0.009	0.174	0.686	0.012	0.015	0.641	0.012
rez1f10	0.860	0.823	0.780	0.802	0.786	0.783	<b>0.937</b>	0.800	0.800
rez1f20	0.829	0.785	0.724	0.748	0.734	0.721	<b>0.850</b>	0.793	0.783
rez1f22	0.859	0.821	0.772	0.792	0.783	0.780	<b>0.905</b>	0.837	0.827
rez2f10	0.899	0.665	0.020	0.599	0.898	0.019	0.024	<b>0.905</b>	0.020
rez2f20	<b>0.846</b>	0.599	0.008	0.458	0.766	0.011	0.010	0.805	0.008
rez2f22	<b>0.880</b>	0.659	0.012	0.465	0.782	0.015	0.013	0.806	0.012
rez3f10	0.957	0.548	0.015	0.468	<b>0.964</b>	0.018	0.021	0.954	0.015
rez3f20	<b>0.892</b>	0.373	0.011	0.258	0.860	0.012	0.016	0.876	0.011
rez3f22	<b>0.815</b>	0.342	0.011	0.210	0.739	0.011	0.014	0.731	0.014
1. Rang <sup>c</sup>	25	0	0	0	4	0	6	1	0
∈ (0.85, 1.0] <sup>d</sup>	7	0	0	0	7	0	2	3	0
∈ (0.7, .85]	29	9	6	6	16	6	7	18	6
≤ 0.7	0	27	30	30	13	30	27	15	30

Tabelle A.11: Vergleich der overall-Power: Tests auf Assoziation,  $\alpha = 0.05$

<sup>a</sup> fett: Verfahren mit der jeweils höchsten Power

<sup>b</sup> grau: Power im Intervall (0.85, 1], hellgrau: Power in (0.7, 0.85]

<sup>c</sup> Anzahl erster Plätze unter den Verfahren

<sup>d</sup> Anzahl Szenarien, für die Power ins angegebene Intervall fällt

Szenarium	zhao _l	zhaomax _l	mm _l	perm _s	perm _m	hbat _bl	hbat _gl	full _l	allM _l
add1f10	<b>0.843<sup>a</sup></b>	0.814 <sup>a</sup>	0.766	0.771	0.775	0.773	0.765	0.758	0.765
add1f20	<b>0.832</b>	0.815	0.752	0.768	0.785	0.782	0.764	0.768	0.756
add1f22	<b>0.792</b>	0.766	0.688	0.703	0.716	0.721	0.717	0.706	0.694
add2f10	<b>0.824</b>	0.614	0.015	0.730	0.561	0.541	0.818	0.785	0.015
add2f20	<b>0.768</b>	0.563	0.015	0.654	0.477	0.456	0.727	0.698	0.014
add2f22	<b>0.763</b>	0.562	0.009	0.624	0.463	0.440	0.705	0.217	0.011
add3f10	0.832	0.360	0.014	0.776	0.333	0.278	<b>0.860<sup>a</sup></b>	0.138	0.014
add3f20	0.760	0.349	0.017	0.698	0.305	0.255	<b>0.764</b>	0.353	0.015
add3f22	<b>0.756</b>	0.334	0.013	0.663	0.292	0.253	0.735	0.543	0.014
dom1f10	<b>0.751</b>	0.700	0.604	0.617	0.635	0.637	0.616	0.612	0.613
dom1f20	<b>0.759</b>	0.725	0.627	0.654	0.679	0.670	0.642	0.645	0.633
dom1f22	<b>0.804</b>	0.755	0.653	0.664	0.683	0.690	0.672	0.676	0.655
dom2f10	<b>0.832</b>	0.519	0.011	0.712	0.428	0.428	0.830	0.820	0.011
dom2f20	<b>0.776</b>	0.515	0.018	0.633	0.395	0.372	0.697	0.576	0.018
dom2f22	<b>0.859</b>	0.612	0.018	0.729	0.454	0.441	0.752	0.613	0.018
dom3f10	0.815	0.297	0.014	0.708	0.250	0.234	<b>0.849</b>	0.829	0.014
dom3f20	<b>0.791</b>	0.300	0.014	0.696	0.232	0.219	0.758	0.533	0.014
dom3f22	<b>0.818</b>	0.328	0.015	0.729	0.252	0.246	0.749	0.448	0.014
mul1f10	<b>0.834</b>	0.802	0.758	0.758	0.772	0.765	0.765	0.758	0.761
mul1f20	<b>0.781</b>	0.743	0.679	0.696	0.702	0.704	0.691	0.691	0.683
mul1f22	<b>0.748</b>	0.698	0.643	0.659	0.676	0.678	0.660	0.669	0.648
mul2f10	<b>0.838</b>	0.595	0.012	0.742	0.542	0.515	0.831	0.824	0.013
mul2f20	<b>0.731</b>	0.497	0.019	0.608	0.408	0.379	0.675	0.627	0.020
mul2f22	<b>0.756</b>	0.538	0.031	0.636	0.437	0.419	0.683	0.595	0.030
mul3f10	0.848	0.364	0.012	0.779	0.316	0.261	<b>0.880</b>	0.828	0.012
mul3f20	<b>0.878</b>	0.395	0.010	0.811	0.355	0.315	0.877	0.750	0.011
mul3f22	<b>0.780</b>	0.323	0.014	0.701	0.286	0.244	0.765	0.524	0.013
rez1f10	<b>0.860</b>	0.823	0.800	0.799	0.812	0.813	0.800	0.800	0.800
rez1f20	<b>0.860</b>	0.818	0.800	0.813	0.826	0.820	0.813	0.809	0.802
rez1f22	<b>0.878</b>	0.859	0.841	0.845	0.857	0.855	0.851	0.852	0.845
rez2f10	0.899	0.665	0.020	0.844	0.622	0.609	<b>0.909</b>	0.905	0.020
rez2f20	<b>0.865</b>	0.641	0.010	0.773	0.561	0.549	0.835	0.773	0.010
rez2f22	<b>0.900</b>	0.697	0.014	0.798	0.580	0.559	0.855	0.696	0.013
rez3f10	0.957	0.548	0.015	0.923	0.514	0.473	<b>0.967</b>	0.954	0.015
rez3f20	0.901	0.426	0.017	0.847	0.381	0.337	<b>0.915</b>	0.787	0.016
rez3f22	<b>0.841</b>	0.390	0.011	0.761	0.319	0.278	0.819	0.455	0.015
1. Rang <sup>a</sup>	29	0	0	0	0	0	7	0	0
$\in (0.85, 1.0]^a$	10	1	0	1	1	1	8	3	0
$\in (0.7, 0.85]$	26	9	6	22	7	7	20	14	6
$\leq 0.7$	0	26	30	13	28	28	8	19	30

Tabelle A.12: Vergleich der overall-Power: Tests auf Kopplung,  $\alpha = 0.05$

<sup>a</sup> Legende wie in Tab. A.11

Szenarium	zhao	zhaomax	mm	hbat_b	hbat_g	hap_pm	gen_pm	full	allM
add1f10	<b>0.651</b> <sup>a</sup>	0.599 <sup>b</sup>	0.596	0.587	0.588	0.619	0.586	0.583	0.595
add1f20	<b>0.597</b>	0.578	0.557	0.555	0.556	0.564	0.553	0.579	0.577
add1f22	<b>0.513</b>	0.500	0.475	0.485	0.479	0.485	0.476	0.510	0.499
add2f10	0.611	0.375	0.005	0.337	<b>0.675</b>	0.012	0.009	0.637	0.005
add2f20	<b>0.508</b>	0.271	0.001	0.178	0.494	0.001	0.002	0.479	0.002
add2f22	<b>0.470</b>	0.265	0	0.144	0.445	0.001	0.001	0.146	0.002
add3f10	0.637	0.141	0.002	0.120	<b>0.748</b> <sup>b</sup>	0.004	0.005	0.057	0.002
add3f20	0.508	0.108	0.005	0.053	<b>0.514</b>	0.007	0.003	0.199	0.006
add3f22	0.480	0.107	0.003	0.053	<b>0.489</b>	0.005	0.007	0.418	0.004
dom1f10	0.539	0.493	0.450	0.456	0.454	0.480	<b>0.658</b>	0.419	0.414
dom1f20	0.494	0.467	0.432	0.445	0.432	0.471	<b>0.656</b>	0.421	0.416
dom1f22	0.508	0.495	0.466	0.472	0.461	0.508	<b>0.694</b>	0.458	0.454
dom2f10	0.598	0.276	0.002	0.236	<b>0.710</b>	0.004	0.006	0.650	0.002
dom2f20	<b>0.533</b>	0.213	0.006	0.131	0.483	0.007	0.008	0.460	0.005
dom2f22	<b>0.671</b>	0.309	0.003	0.183	0.575	0.008	0.008	0.570	0.005
dom3f10	0.598	0.128	0.004	0.090	<b>0.749</b>	0.007	0.007	0.699	0.004
dom3f20	0.538	0.099	0.004	0.053	<b>0.556</b>	0.006	0.007	0.541	0.004
dom3f22	<b>0.577</b>	0.119	0.001	0.061	0.531	0.001	0.004	0.532	0.002
mul1f10	<b>0.654</b>	0.618	0.592	0.595	0.595	0.621	0.591	0.598	0.600
mul1f20	<b>0.487</b>	0.455	0.405	0.403	0.398	0.461	0.448	0.461	0.470
mul1f22	0.443	0.419	0.386	0.388	0.385	0.430	0.419	<b>0.452</b>	0.443
mul2f10	0.612	0.334	0.002	0.291	<b>0.687</b>	0.005	0.006	0.684	0.002
mul2f20	<b>0.472</b>	0.219	0.006	0.126	0.418	0.008	0.006	0.451	0.007
mul2f22	<b>0.513</b>	0.258	0.007	0.141	0.418	0.011	0.012	0.467	0.008
mul3f10	0.645	0.141	0.001	0.107	<b>0.759</b>	0.006	0.003	0.710	0.001
mul3f20	0.669	0.161	0.004	0.098	<b>0.688</b>	0.006	0.004	0.669	0.004
mul3f22	<b>0.539</b>	0.110	0.002	0.059	0.519	0.002	0.008	0.494	0.002
rez1f10	0.692	0.663	0.635	0.642	0.633	0.648	<b>0.870</b>	0.654	0.655
rez1f20	0.642	0.601	0.553	0.563	0.562	0.579	<b>0.724</b>	0.634	0.625
rez1f22	0.661	0.640	0.622	0.620	0.621	0.657	<b>0.779</b>	0.687	0.687
rez2f10	0.732	0.399	0.004	0.390	0.783	0.007	0.009	<b>0.816</b>	0.004
rez2f20	<b>0.657</b>	0.352	0.002	0.237	0.599	0.003	0.002	0.653	0.001
rez2f22	<b>0.728</b>	0.426	0.004	0.242	0.600	0.005	0.007	0.650	0.006
rez3f10	0.855	0.274	0.003	0.250	0.905	0.009	0.009	<b>0.912</b>	0.003
rez3f20	0.727	0.160	0.003	0.101	0.724	0.003	0.001	<b>0.747</b>	0.004
rez3f22	<b>0.620</b>	0.150	0.004	0.089	0.564	0.006	0.003	0.588	0.003
1. Rang <sup>c</sup>	16	0	0	0	10	0	6	4	0
$\in (0.7, 1.0]$ <sup>d</sup>	4	0	0	0	10	0	3	4	0
$\in (0.5, 0.7]$	26	6	6	6	7	7	6	18	6
$\leq 0.5$	6	30	30	30	17	29	27	14	30

Tabelle A.13: Vergleich der overall-Power: Tests auf Assoziation,  $\alpha = 0.01$

<sup>a</sup> fett: Verfahren mit der jeweils höchsten Power

<sup>b</sup> grau: Power im Intervall (0.7, 1], hellgrau: Power in (0.5, 0.7]

<sup>c</sup> Anzahl erster Plätze unter den Verfahren

<sup>d</sup> Anzahl Szenarien, für die Power ins angegebene Intervall fällt

Szenarium	zhao _l	zhaomax _l	mm _l	perm _s	perm _m	hbat _bl	hbat _gl	full _l	allM _l
add1f10	<b>0.651<sup>a</sup></b>	0.599	0.595	0.618	0.613	0.592	0.591	0.583	0.595
add1f20	<b>0.650</b>	0.647	0.595	0.633	0.637	0.607	0.610	0.607	0.606
add1f22	0.558	<b>0.570</b>	0.524	0.562	0.564	0.539	0.544	0.531	0.525
add2f10	0.611	0.375	0.004	0.599	0.385	0.340	<b>0.675</b>	0.637	0.005
add2f20	0.544	0.311	0.002	0.514	0.310	0.251	<b>0.571</b>	0.500	0.003
add2f22	0.513	0.293	0.001	0.471	0.268	0.220	<b>0.527</b>	0.099	0.003
add3f10	0.637	0.141	0.002	0.647	0.184	0.125	<b>0.748<sup>a</sup></b>	0.057	0.002
add3f20	0.552	0.144	0.006	0.556	0.156	0.106	<b>0.609</b>	0.209	0.006
add3f22	0.548	0.138	0.005	0.530	0.137	0.099	<b>0.569</b>	0.373	0.005
dom1f10	<b>0.539</b>	0.493	0.412	0.466	0.456	0.423	0.421	0.419	0.414
dom1f20	<b>0.546</b>	0.530	0.444	0.512	0.513	0.473	0.467	0.463	0.454
dom1f22	<b>0.561</b>	0.548	0.477	0.512	0.514	0.502	0.484	0.495	0.479
dom2f10	0.598	0.276	0.002	0.545	0.265	0.226	<b>0.670</b>	0.650	0.002
dom2f20	<b>0.551</b>	0.261	0.005	0.484	0.210	0.159	0.524	0.418	0.005
dom2f22	<b>0.688</b>	0.360	0.004	0.588	0.281	0.237	0.609	0.436	0.005
dom3f10	0.598	0.128	0.004	0.551	0.117	0.089	<b>0.718</b>	0.699	0.004
dom3f20	0.567	0.124	0.007	0.544	0.114	0.094	<b>0.599</b>	0.361	0.007
dom3f22	<b>0.628</b>	0.132	0.002	0.578	0.121	0.091	0.585	0.263	0.002
mul1f10	<b>0.654</b>	0.618	0.601	0.636	0.636	0.597	0.598	0.598	0.600
mul1f20	0.526	0.508	0.483	<b>0.527</b>	0.520	0.496	0.495	0.488	0.498
mul1f22	<b>0.504</b>	0.491	0.453	0.496	0.499	0.463	0.458	0.470	0.457
mul2f10	0.612	0.334	0.002	0.607	0.342	0.291	<b>0.693</b>	0.684	0.002
mul2f20	<b>0.516</b>	0.256	0.007	0.463	0.246	0.190	0.502	0.447	0.006
mul2f22	<b>0.542</b>	0.304	0.008	0.484	0.250	0.212	0.508	0.438	0.009
mul3f10	0.645	0.141	0.001	0.636	0.167	0.110	<b>0.758</b>	0.710	0.001
mul3f20	0.700	0.199	0.005	0.687	0.202	0.147	<b>0.763</b>	0.607	0.005
mul3f22	0.583	0.129	0.002	0.561	0.131	0.087	<b>0.595</b>	0.369	0.003
rez1f10	<b>0.692</b>	0.663	0.657	0.685	0.686	0.659	0.654	0.654	0.655
rez1f20	0.682	0.654	0.655	0.697	<b>0.702</b>	0.674	0.672	0.673	0.659
rez1f22	0.703	0.690	0.710	0.737	<b>0.739</b>	0.702	0.709	0.717	0.719
rez2f10	0.732	0.399	0.004	0.724	0.444	0.406	0.810	<b>0.816</b>	0.004
rez2f20	0.698	0.404	0.001	0.644	0.380	0.328	<b>0.714</b>	0.639	0.002
rez2f22	<b>0.762</b>	0.463	0.006	0.686	0.383	0.325	0.731	0.548	0.007
rez3f10	0.855	0.274	0.003	0.854	0.318	0.259	<b>0.915</b>	0.912	0.003
rez3f20	0.764	0.197	0.004	0.757	0.211	0.138	<b>0.820</b>	0.651	0.005
rez3f22	0.669	0.177	0.005	0.639	0.180	0.126	<b>0.696</b>	0.298	0.004
1. Rang <sup>a</sup>	14	1	0	1	2	0	17	1	0
$\in (0.7, 1.0]$ <sup>a</sup>	5	0	1	4	2	1	10	4	1
$\in (0.5, 0.7]$	31	10	6	26	8	7	21	14	6
$\leq 0.5$	0	26	29	6	26	28	5	18	29

Tabelle A.14: Vergleich der overall-Power: Tests auf Kopplung,  $\alpha = 0.01$

<sup>a</sup> Legende wie in Tab. A.13

KMod 1 scenario	Tests auf $H_0^A$								
	zhao	zhaomax	mm	hbat_b	hbat_g	hap_pm	gen_pm	full	allM
add1f20	<b>0.793<sup>a</sup></b>	0.753 <sup>b</sup>	0.693	0.708	0.702	0.703	0.703	0.725	0.724
dom1f10	0.720	0.678	0.605	0.635	0.617	0.613	<b>0.777</b>	0.605	0.597
dom1f22	0.757	0.708	0.635	0.658	0.643	0.640	<b>0.815</b>	0.645	0.630
mul1f10	<b>0.831</b>	0.788	0.737	0.747	0.743	0.742	0.711	0.748	0.748
mul1f22	<b>0.686</b>	0.652	0.560	0.596	0.582	0.576	0.554	0.631	0.623
rez1f20	0.818	0.773	0.708	0.728	0.722	0.715	<b>0.829</b>	0.790	0.781
1. Rang	3	0	0	0	0	0	3	0	0
$\in (0.8, 1.0]^c$	2	0	0	0	0	0	2	0	0
$\in (0.7, 0.8]$	3	4	2	3	3	3	1	3	3
$\leq 0.7$	1	2	4	3	3	3	3	3	3

Szenarium	Tests auf $H_0^L$								
	zhao_l	zhaomax_l	mm_l	hbat_bl	hbat_gl	perm_s	perm_m	full_l	allM_l
add1f20	<b>0.821</b>	0.792	0.728	0.758	0.746	0.748	0.759	0.753	0.749
dom1f10	<b>0.720</b>	0.678	0.577	0.610	0.596	0.603	0.613	0.605	0.597
dom1f22	<b>0.784</b>	0.731	0.643	0.674	0.656	0.670	0.679	0.657	0.642
mul1f10	<b>0.831</b>	0.788	0.738	0.746	0.742	0.749	0.756	0.748	0.748
mul1f22	<b>0.727</b>	0.676	0.622	0.655	0.647	0.648	0.665	0.657	0.639
rez1f20	<b>0.847</b>	0.817	0.789	0.803	0.803	0.799	0.803	0.811	0.801
1.Rang	6	0	0	0	0	0	0	0	0
$\in (0.8, 1.0]^c$	3	1	0	1	1	0	1	1	1
$\in (0.7, 0.8]$	3	3	3	2	2	3	2	2	2
$\leq 0.7$	0	2	3	3	3	3	3	3	3

Tabelle A.15: Overall-Power der Verfahren mit 'Missings', KMod 1

<sup>a</sup> beste Werte fett markiert<sup>b</sup> grau: Power in  $(0.8, 1.0]$ , hellgrau: Power in  $(0.7, 0.8]$ <sup>c</sup> Anzahl der Powerwerte im angegebenen Intervall



KMod 2 / 3		Tests auf $H_0^A$				
Szenarium	N <sup>a</sup>	zhao	zhaomax	hbat_b	hbat_g	full
add2f10	1800	0.786 <sup>b</sup>	0.554	0.498	<b>0.789<sup>b</sup></b>	0.779
add3f20	500	<b>0.673</b>	0.258	0.158	0.669	0.345
add2f22	400	<b>0.670</b>	0.458	0.335	0.621	0.278
dom3f10	1500	0.748	0.262	0.204	<b>0.823<sup>b</sup></b>	0.814
dom2f20	500	<b>0.714</b>	0.400	0.283	0.647	0.639
dom3f22	600	<b>0.736</b>	0.252	0.172	0.700	0.704
mul3f10	1700	0.776	0.304	0.252	<b>0.819</b>	0.809
mul2f20	300	<b>0.686</b>	0.427	0.290	0.601	0.637
mul3f22	400	<b>0.683</b>	0.254	0.161	0.673	0.635
rez2f10	6000	0.851	0.621	0.564	0.860	<b>0.898</b>
rez3f20	1400	0.821	0.353	0.234	0.822	<b>0.858</b>
rez3f22	900	<b>0.850</b>	0.599	0.448	0.768	0.799
1. Rang		7	0	0	3	2
$\in (0.8, 1.0]^c$		3	0	0	4	4
$\in (0.7, 0.8]$		5	0	0	2	3
$\leq 0.7$		4	12	12	6	5

Szenarium	Tests auf $H_0^L$						
	zhao_l	zhaomax_l	perm_s	perm_m	hbat_bl	hbat_gl	full_l
add2f10	<b>0.786<sup>a</sup></b>	0.554	0.702	0.525	0.504	0.782	0.779
add3f20	0.723	0.301	0.652	0.295	0.234	<b>0.735</b>	0.324
add2f22	<b>0.703</b>	0.514	0.621	0.447	0.429	0.692	0.212
dom3f10	0.750	0.262	0.654	0.215	0.205	0.798	<b>0.814</b>
dom2f20	<b>0.738</b>	0.442	0.613	0.358	0.345	0.676	0.569
dom3f22	<b>0.776</b>	0.296	0.710	0.253	0.241	0.734	0.419
mul3f10	0.776	0.304	0.729	0.295	0.253	<b>0.822</b>	0.809
mul2f20	<b>0.719</b>	0.483	0.615	0.411	0.378	0.669	0.628
mul3f22	0.720	0.286	0.684	0.273	0.231	<b>0.746</b>	0.499
rez2f10	0.858	0.611	0.801	0.587	0.574	0.878	<b>0.898</b>
rez3f20	0.849	0.390	0.817	0.373	0.312	<b>0.883</b>	0.749
rez3f22	<b>0.873</b>	0.662	0.792	0.569	0.551	0.841	0.688
1. Rang	6	0	0	0	0	4	2
$\in (0.8, 1.0]^c$	3	0	2	0	0	4	3
$\in (0.7, 0.8]$	9	0	4	0	0	5	2
$\leq 0.7$	0	12	6	12	12	3	7

Tabelle A.16: Power der Verfahren mit 'Missings', KMod 2 und 3

<sup>a</sup> Stichprobenumfang

<sup>b</sup> grau: Power > 80%, hellgrau: Power in (0.7, 0.8]

<sup>c</sup> Anzahl der Powerwerte im angegebenen Intervall

Mk	1 <sup>a</sup>	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31			
mm		+ <sup>b</sup>			+	+	+	+	+				+	+	+				+	+	+	+	+	+	+	+	+	+	+	+	+	+		
hbat_b		+			+	+	+	+	+	+			+	+	+				+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
hbat_g		+			+	+	+	+	+				+	+	+				+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
zhao					+	+	+			+			+	+					+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
zhaomax					+	+	+					+	+	+					+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
hap_pm					+	+													+		+	+	+		+	+			+	+				
gen_pm					+	+													+	+	+	+		+	+	+	+	+	+	+	+	+	+	+
full		+	+	+	+	+	+	+	+	+	+	+	+	+	+				+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
allM		+			+	+	+	+	+	+			+	+	+				+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Mk	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63				
mm				+			+	+	+	+				+	+		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		
hbat_b			+	+			+	+	+	+				+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
hbat_g			+				+	+	+					+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
zhao			+				+	+					+	+			+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
zhaomax							+	+					+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
hap_pm																																				
gen_pm							+	+												+	+				+	+			+	+	+				+	
full		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
allM		+	+	+			+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	

Tabelle A.17: Analyse realer Daten: Signifikante Markerkombinationen

<sup>a</sup> Markerkombinationen sind gemäß 4.2.5 binär codiert

<sup>b</sup> Signifikanz des mit 63 korrigierten p-Wertes für diese Mk

Abk.	ausgeschrieben	Übersetzung	Referenz
ADD	additive	additiver Vererbungsmodus	1.1.1
ASP	affected sib pairs	Datenstruktur: zwei kranke Geschwister	1.1.2
CDCV	common disease - common variant	–	2.2.1
CVN	copy number variation	–	1.2.1
DL	disease locus	Risiko- / Krankheitslocus	1.1.2
DOM	dominant	dominanter Vererbungsmodus	1.1.1
EM	expectation maximization	Expectation-Maximization-Algorithmus	4.2.2
FBAT	family based association test	Familien-basierter Assoziationstest	3.2.4
GWAS	genome wide association study	genomweite Assoziationsstudie	1.2.1
HLA	human leukin antigen	wichtige Region für das Immunsystem	1.1.1
HWE	Hardy-Weinberg equilibrium	Hardy-Weinberg-Gleichgewicht	Formel (1.1)
$H_0^L$	–	Nullhypothese: keine Kopplung	1.1.2
$H_0^A$	–	Nullhypothese: keine Kopplung oder keine Assoziation	1.1.2
IBD	identical by descent	identisch nach der Herkunft	1.1.1
LRT	likelihood ratio test	Likelihood-Quotienten-Test	4.4.2
LD	linkage disequilibrium	Kopplungs-Ungleichgewicht	1.1.1
MAF	minor allele frequency	Häufigkeit des seltenen Allels	2.2.1
Mk	–	Markerkombination	6.2.1
MOI	mode of inheritance	Vererbungsmodus	1.1.2
MUL	multiplicative	multiplikativer Vererbungsmodus	1.1.1
MT	mating type	–	1.1.1
OR	odds ratio	Risikoverhältnis	Formel (1.12)
REZ	recessive	rezessiver Vererbungsmodus	1.1.1
SNP	single nucleotide polymorphism	–	1.1.1
TDT	Transmission Disequilibrium Test	–	1.1.2

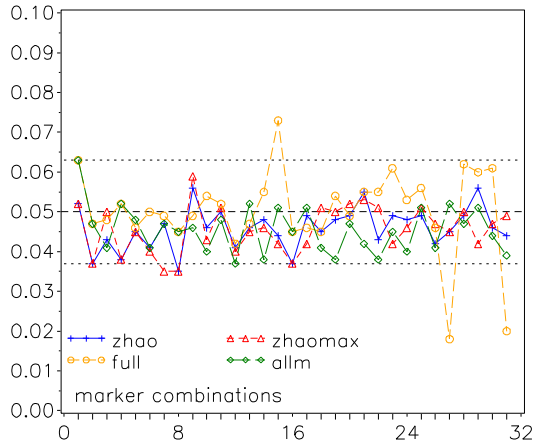
Tabelle A.18: Abkürzungen

# Fehler 1. Art: Grafiken für alle Markerkombinationen

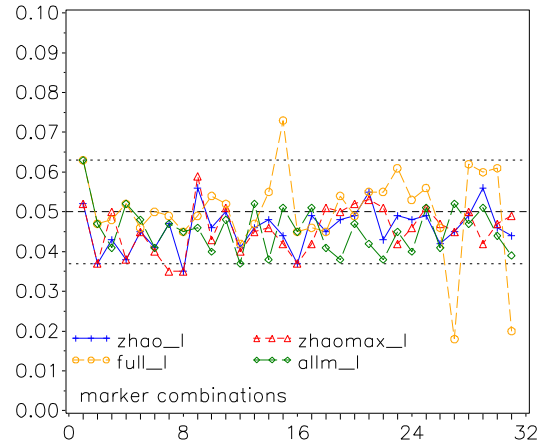
Testniveau  $\alpha = 0.05$ , 200 Familien

NMod 1

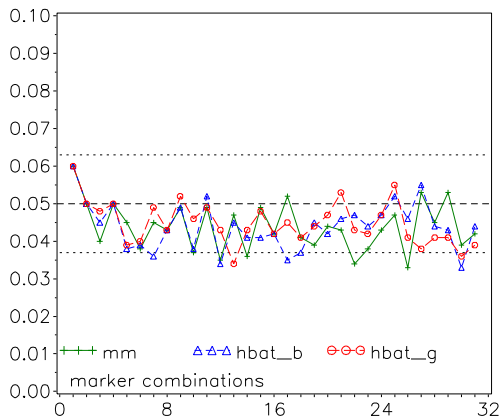
Type I error of a) Famhap b) Unphased  
200 Trios, Nmod1, H0\_(A)



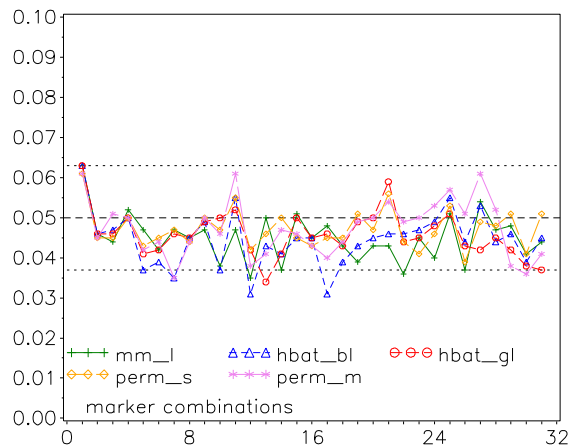
Type I error of a) Famhap b) Unphased  
200 Trios, NMod1 H0\_(L)



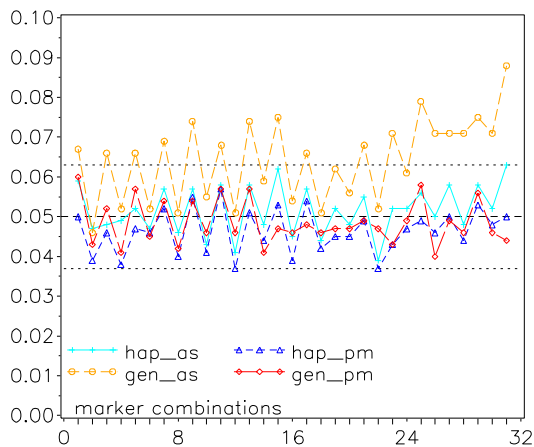
Type I error of FBAT  
200 Trios, NMod1 H0\_(A)



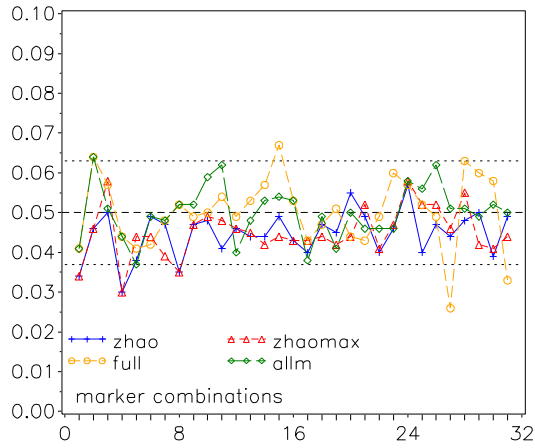
Type I error of FBAT  
200 Trios, NMod1 H0\_(L)



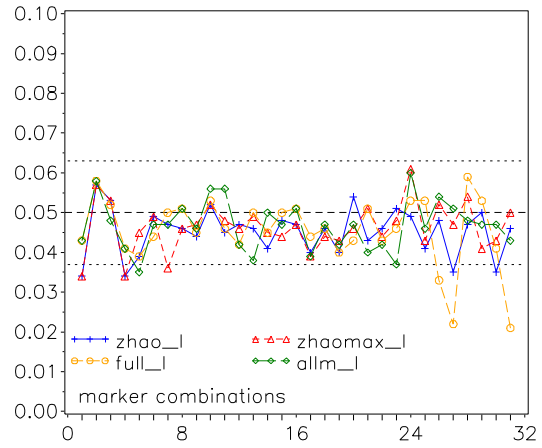
Hotel\_Fam: Type I error,  
200 trios, NMod 1



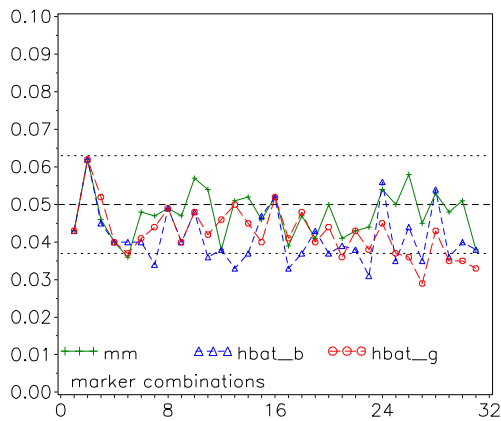
Type I error of a) Famhap b) Unphased  
200 AA Families, Nmod1, H0\_(A)



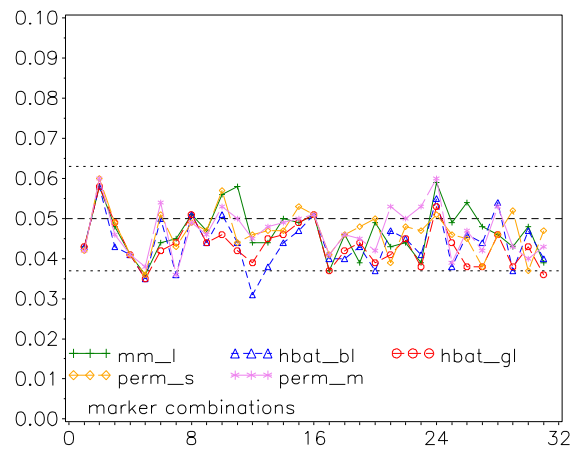
Type I error of a) Famhap b) Unphased  
200 AA Families, NMod1 H0\_(L)



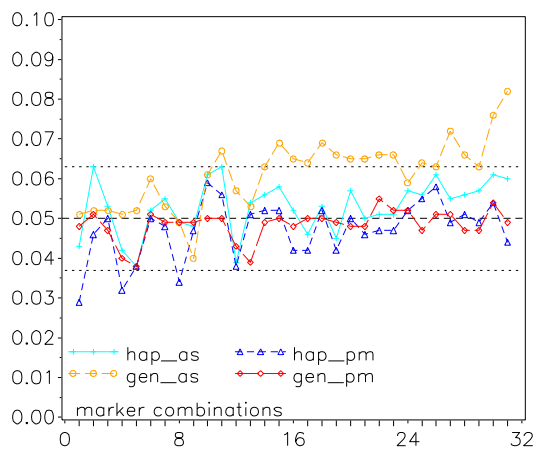
Type I error of FBAT  
200 AA Families, NMod1 H0\_(A)



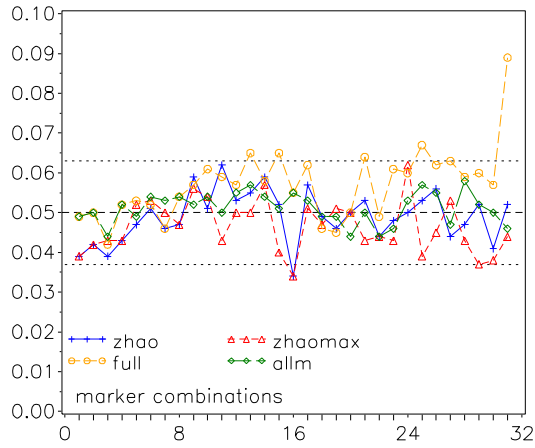
Type I error of FBAT  
200 AA Families, NMod1 H0\_(L)



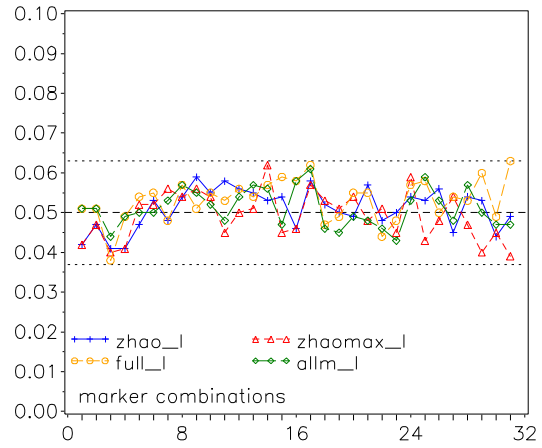
Hotel\_Fam: Type I error,  
200 AA families, NMod 1



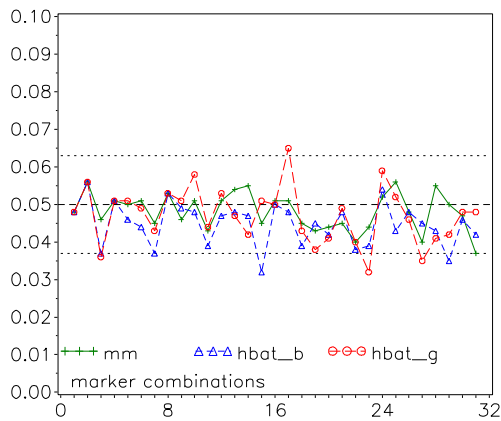
Type I error of a) Famhap b) Unphased  
200 AAUU Families, Nmod1, H0\_(A)



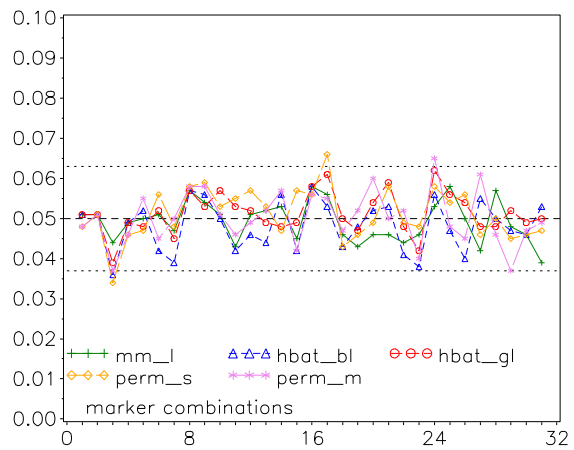
Type I error of a) Famhap b) Unphased  
200 AAUU Families, NMod1 H0\_(L)



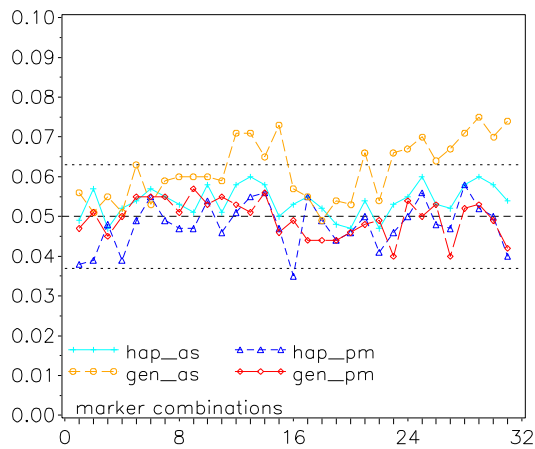
Type I error of FBAT  
200 AAUU Families, NMod1 H0\_(A)



Type I error of FBAT  
200 AAUU Families, NMod1 H0\_(L)

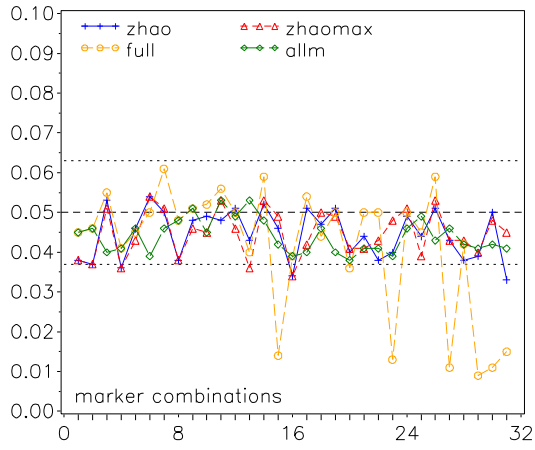


Hotel\_Fam: Type I error,  
200 AAUU families, NMod 1

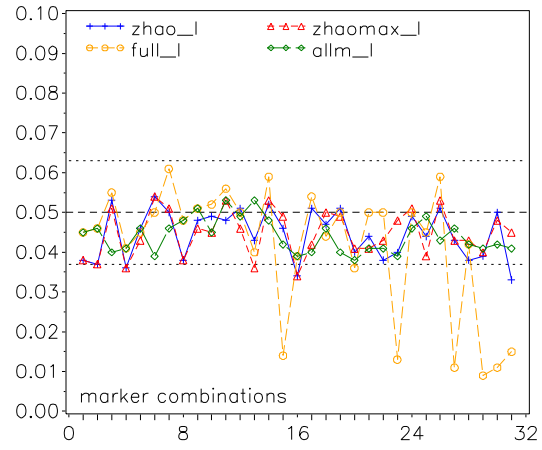


## NMod 2

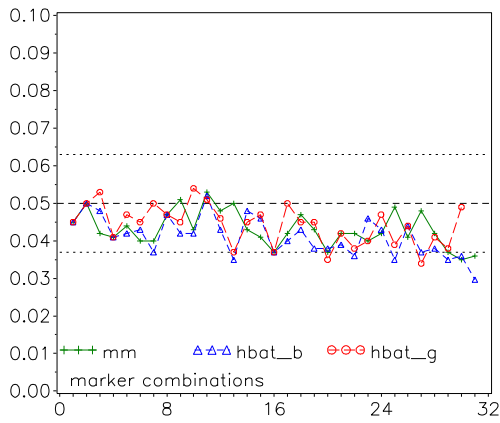
Type I error of a) Famhap b) Unphased  
200 Trios, Nmod 2, H0\_(A)



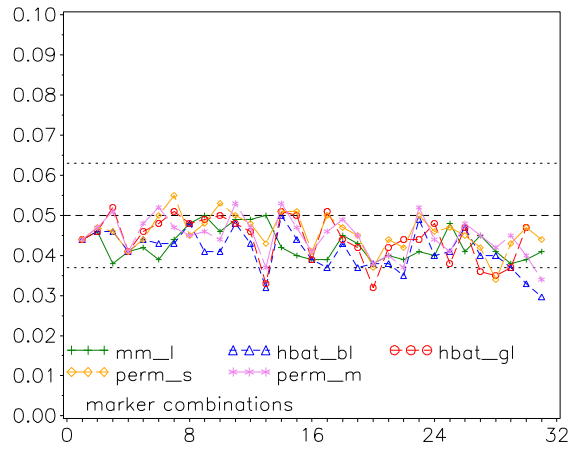
Type I error of a) Famhap b) Unphased  
200 Trios, NMod 2 H0\_(L)



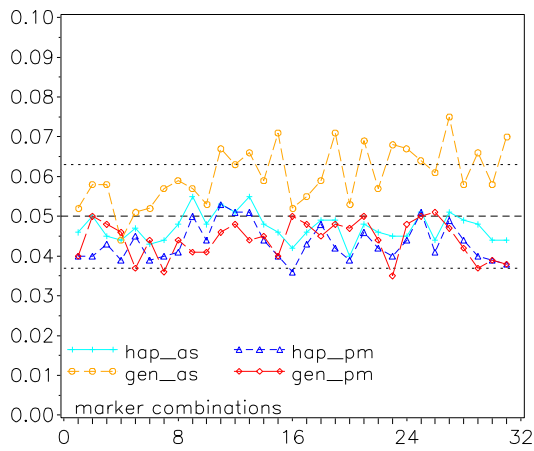
Type I error of FBAT  
200 Trios, NMod2 H0\_(A)



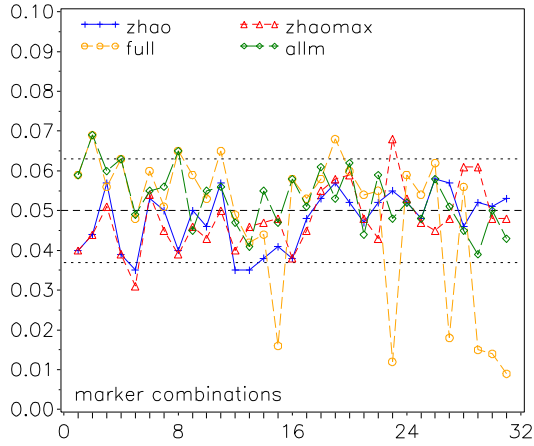
Type I error of FBAT  
200 Trios, NMod2 H0\_(L)



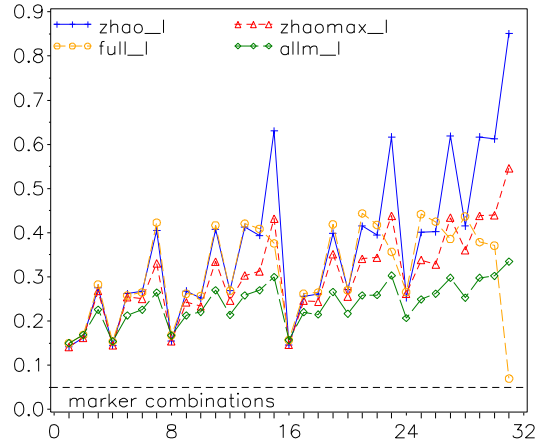
Hotel\_Fam: Type I error,  
200 trios, NMod 2



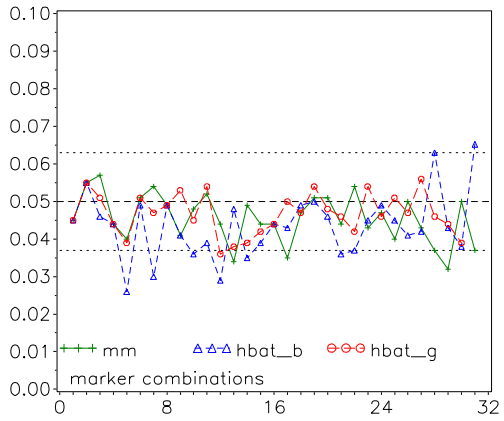
Type I error of a) Famhap b) Unphased  
200 AA Families, NMod 2, H0\_(A)



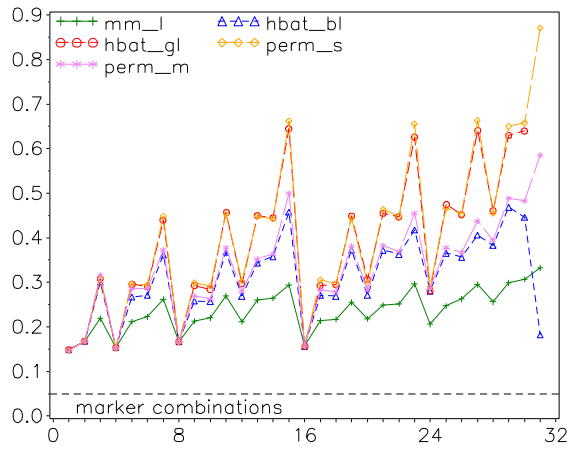
Type I error of a) Famhap b) Unphased  
200 AA Families, NMod2 H0\_(L)



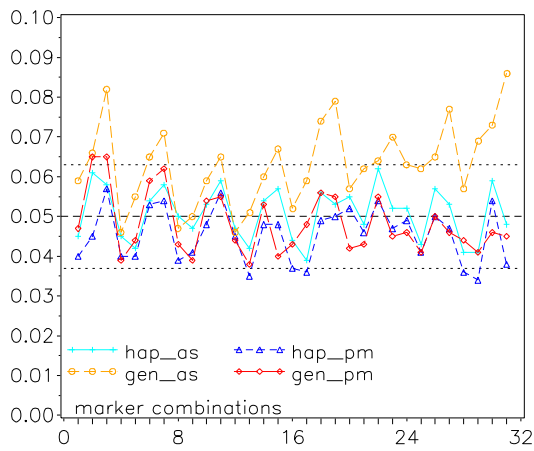
Type I error of FBAT  
200 AA Families, NMod2 H0\_(A)



Type I error of FBAT  
200 AA Families, NMod2 H0\_(L)

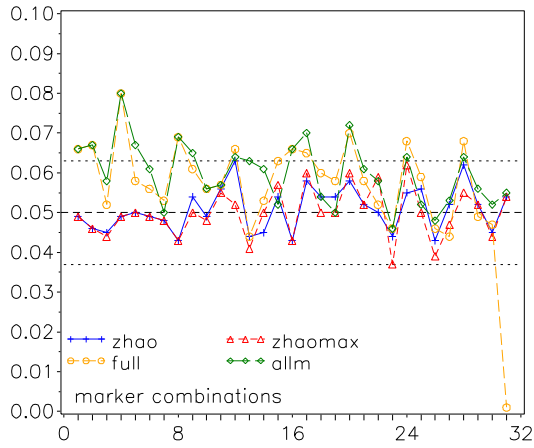


Hotel\_Fam: Type I error,  
200 AA families, NMod 2

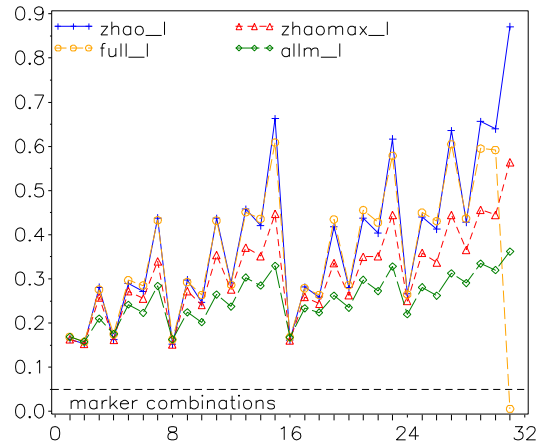




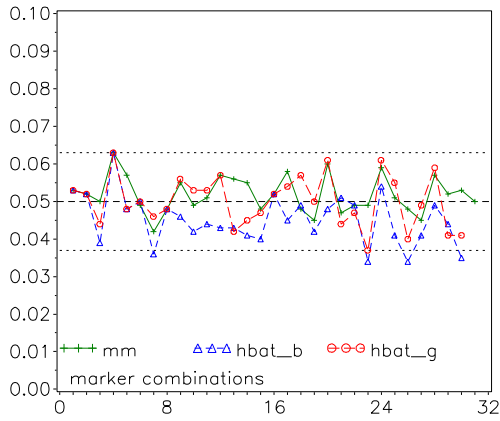
Type I error of a) Famhap b) Unphased  
200 AAUU Families, Nmod2, H0\_(A)



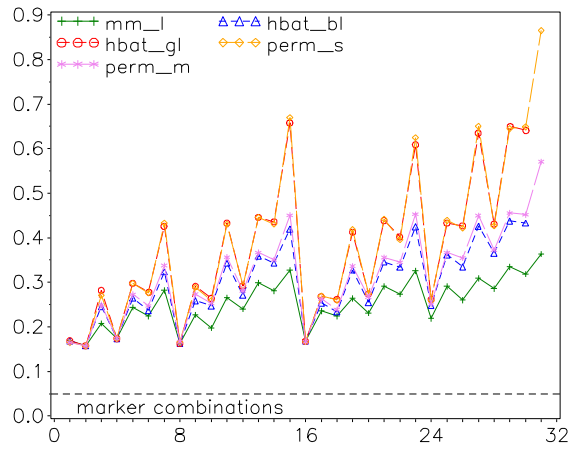
Type I error of a) Famhap b) Unphased  
200 AAUU Families, NMod2 H0\_(L)



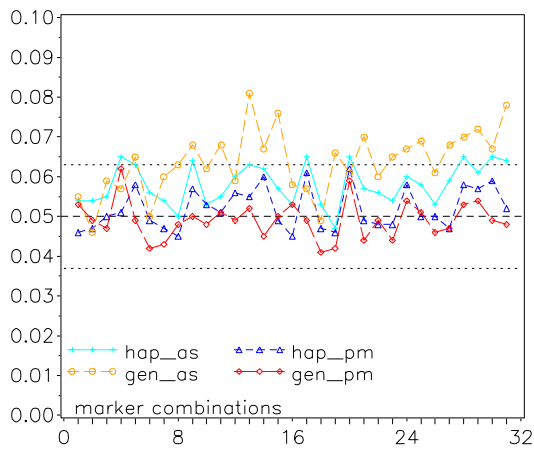
Type I error of FBAT  
200 AAUU Families, NMod2 H0\_(A)



Type I error of FBAT  
200 AAUU Families, NMod2 H0\_(L)

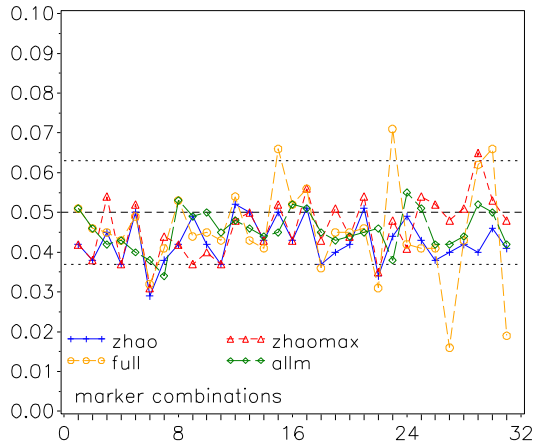


Hotel\_Fam: Type I error,  
200 AAUU families, NMod 2

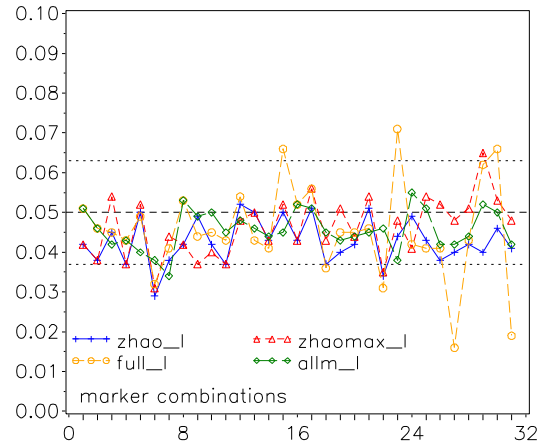


### NMod 3

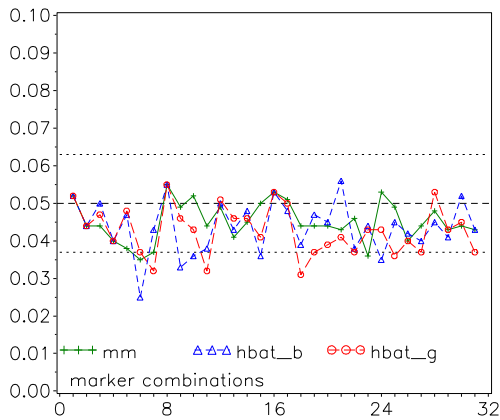
Type I error of a) Famhap b) Unphased  
200 Trios, Nmod3, H0\_(A)



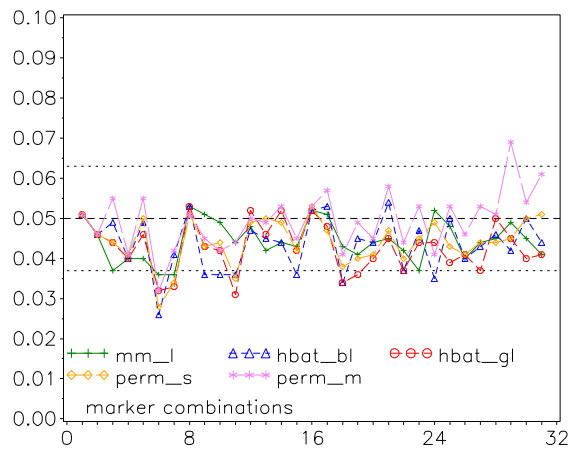
Type I error of a) Famhap b) Unphased  
200 Trios, NMod3 H0\_(L)



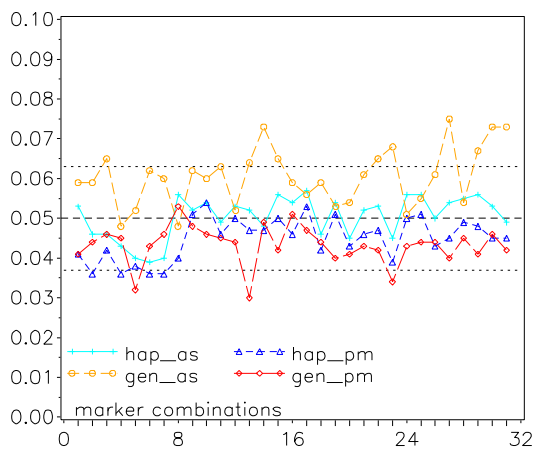
Type I error of FBAT  
200 Trios, NMod3 H0\_(A)



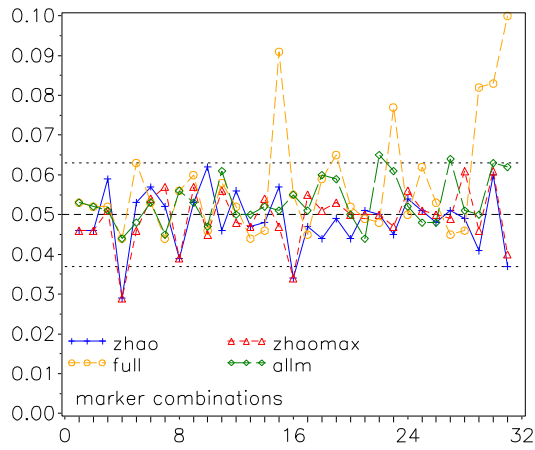
Type I error of FBAT  
200 Trios, NMod3 H0\_(L)



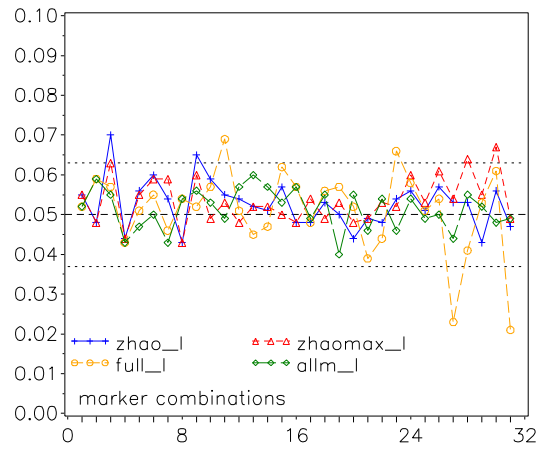
Hotel\_Fam: Type I error,  
200 trios, NMod 3



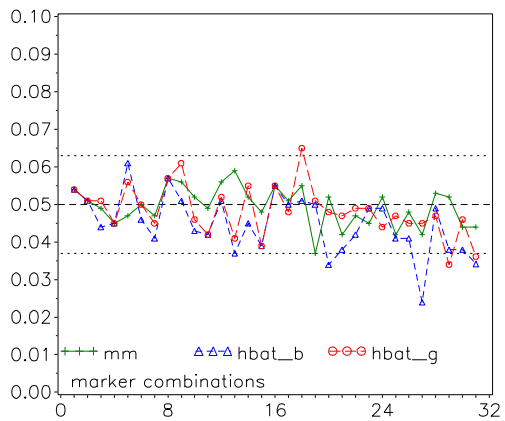
Type I error of a) Famhap b) Unphased  
200 AA Families, Nmod3, H0\_(A)



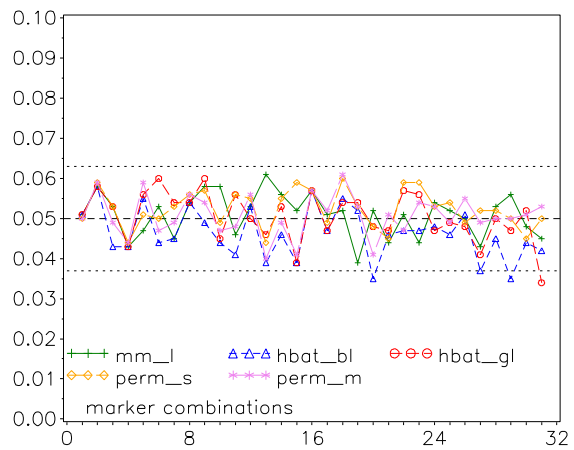
Type I error of a) Famhap b) Unphased  
200 AA Families, NMod3 H0\_(L)



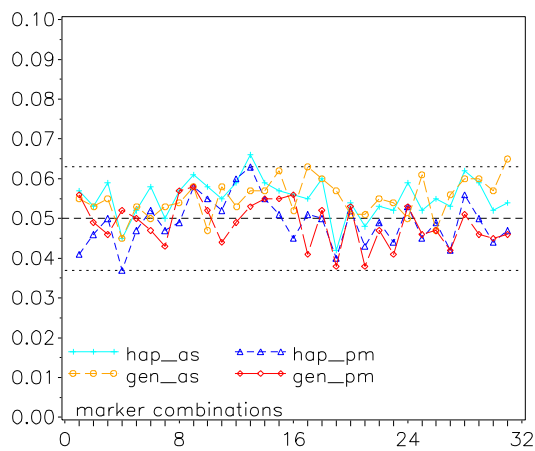
Type I error of FBAT  
200 AA Families, NMod3 H0\_(A)



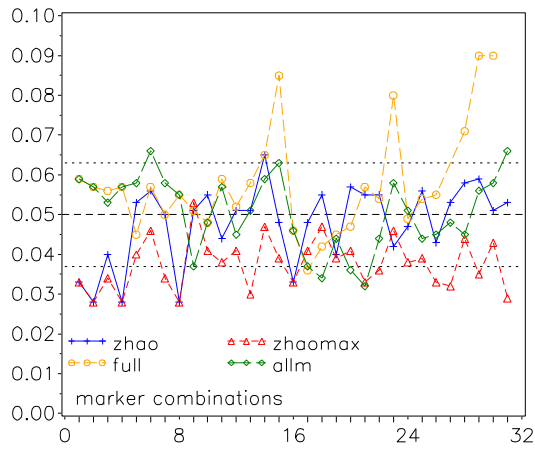
Type I error of FBAT  
200 AA Families, NMod3 H0\_(L)



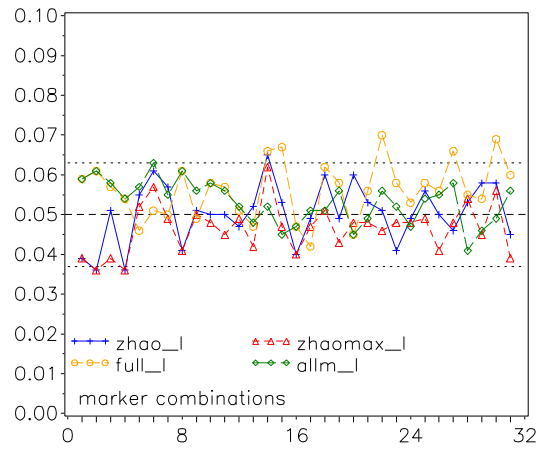
Hotel\_Fam: Type I error,  
200 AA families, NMod 3



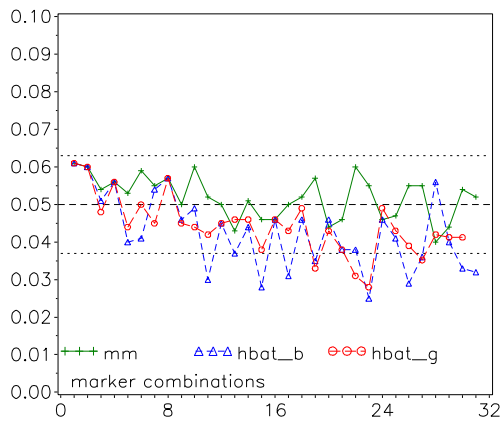
Type I error of a) Famhap b) Unphased  
200 AAUU Families, Nmod3, H0\_(A)



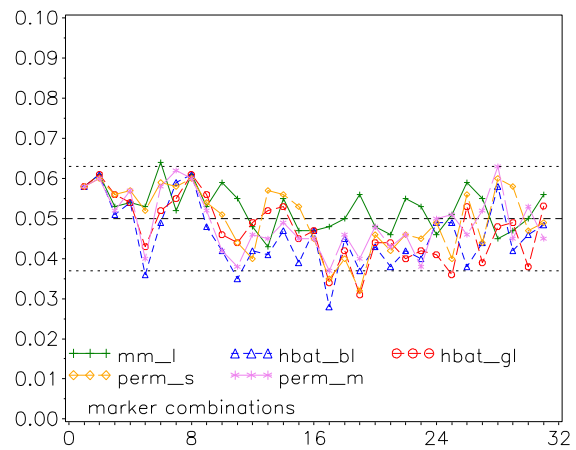
Type I error of a) Famhap b) Unphased  
200 AAUU Families, NMod3 H0\_(L)



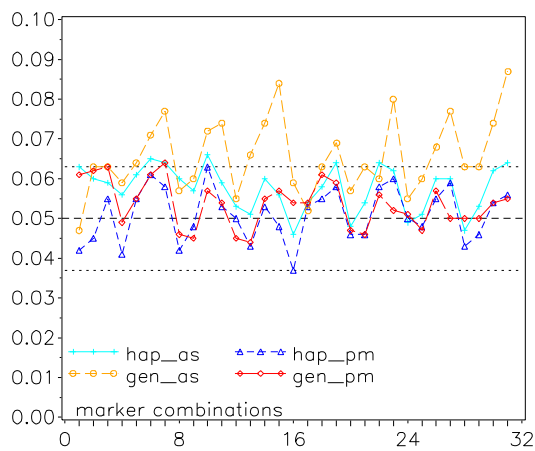
Type I error of FBAT  
200 AAUU Families, NMod3 H0\_(A)



Type I error of FBAT  
200 AAUU Families, NMod3 H0\_(L)



Hotel\_Fam: Type I error,  
200 AAUU families, NMod 3

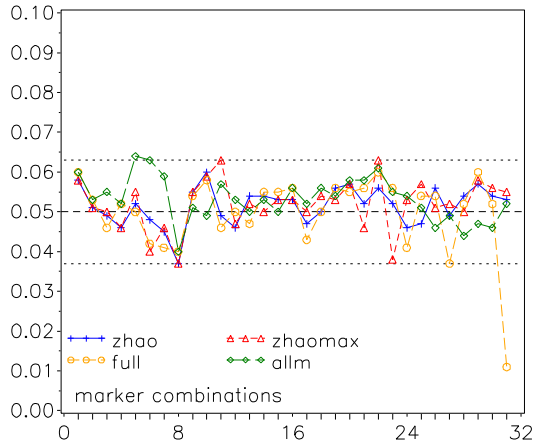


# Fehler 1. Art: Grafiken für alle Markerkombinationen

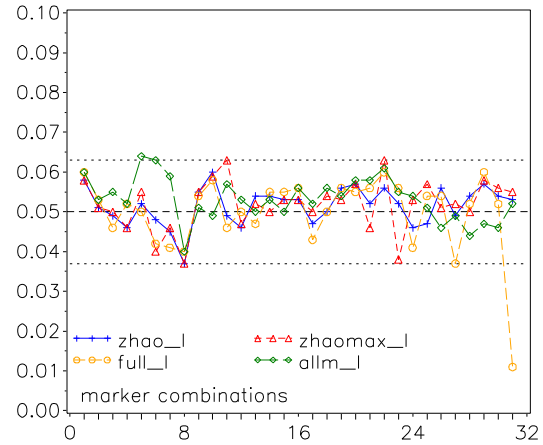
Testniveau  $\alpha = 0.05$ , 1000 Familien

NMod 1

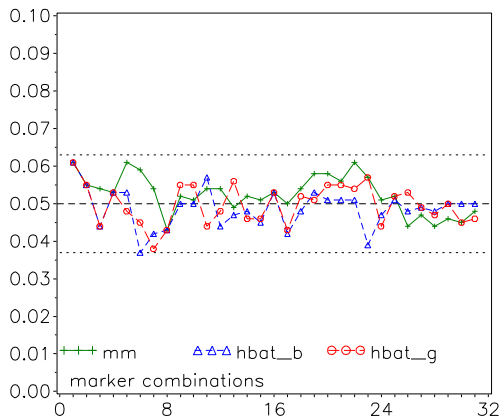
Type I error of a) Famhap b) Unphased  
1000 Trios, Nmod1, H0\_(A)



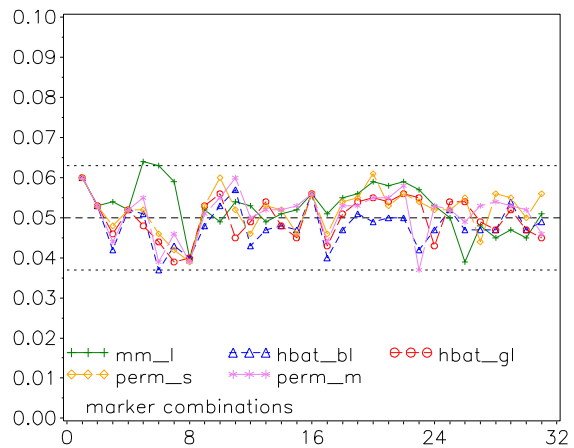
Type I error of a) Famhap b) Unphased  
1000 Trios, NMod1 H0\_(L)



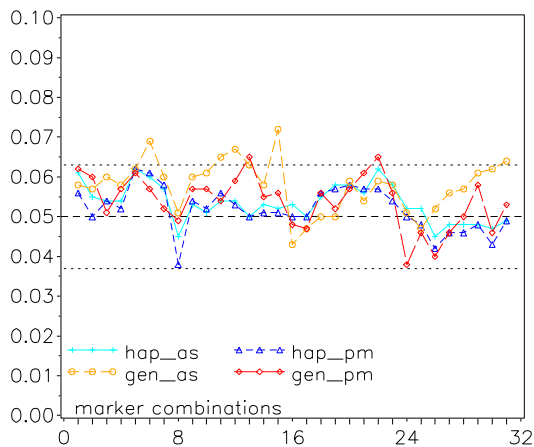
Type I error of FBAT  
1000 Trios, NMod1 H0\_(A)



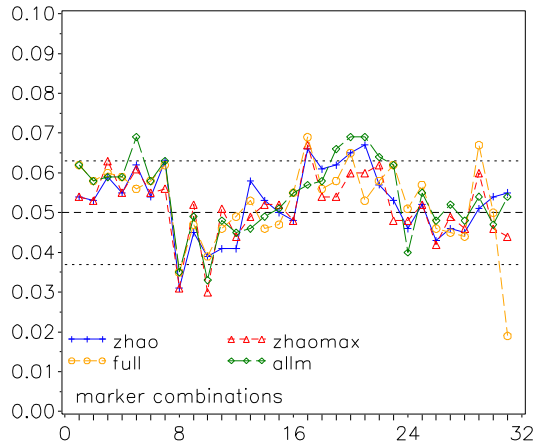
Type I error of FBAT  
1000 Trios, NMod1 H0\_(L)



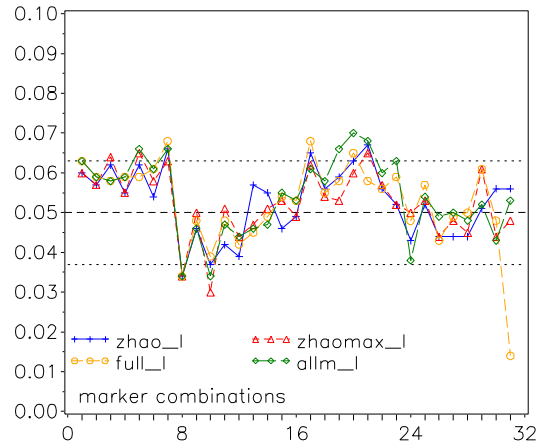
Hotel\_Fam: Type I error,  
1000 trios, NMod 1



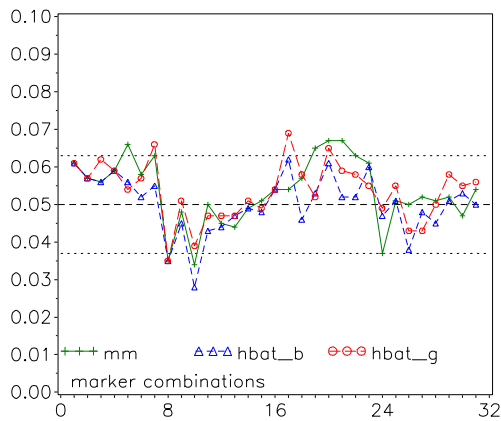
Type I error of a) Famhap b) Unphased  
1000 AA Families, Nmod1, H0\_(A)



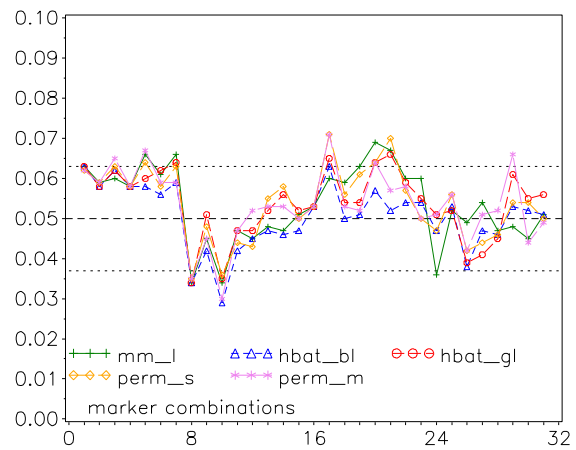
Type I error of a) Famhap b) Unphased  
1000 AA Families, NMod1 H0\_(L)



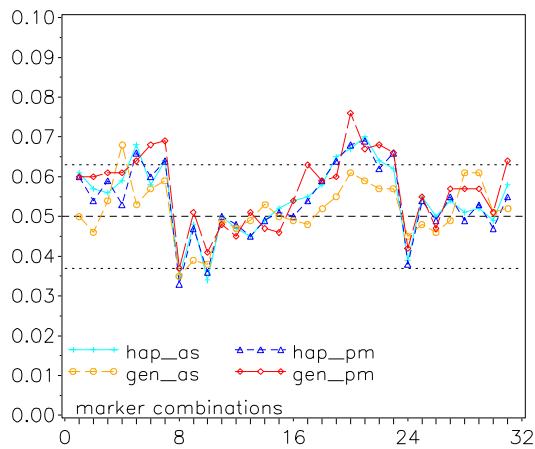
Type I error of FBAT  
1000 AA Families, NMod1 H0\_(A)



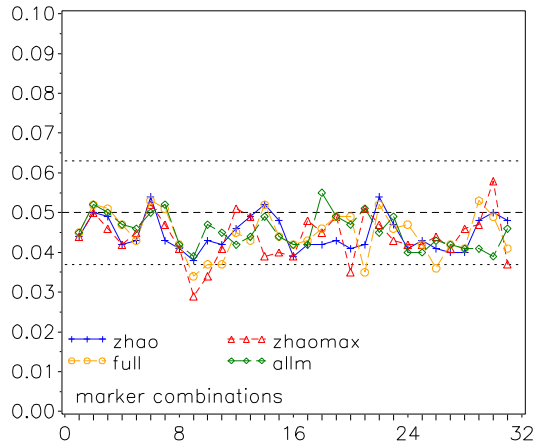
Type I error of FBAT  
1000 AA Families, NMod1 H0\_(L)



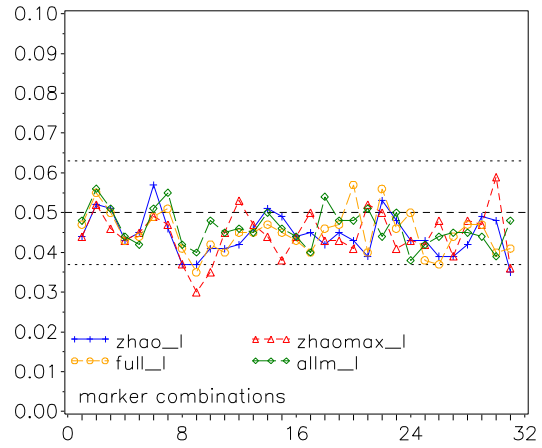
Hotel\_Fam: Type I error,  
1000 AA families, NMod 1



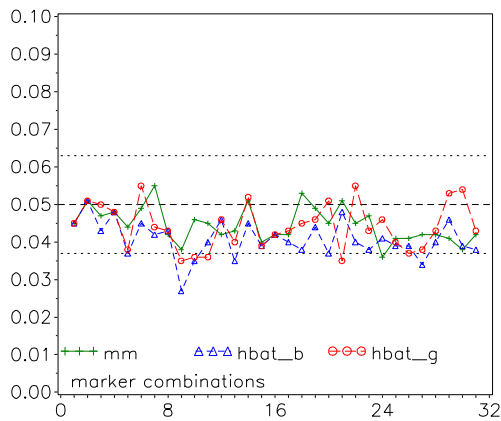
Type I error of a) Famhap b) Unphased  
1000 AAU Families, Nmod1, H0\_(A)



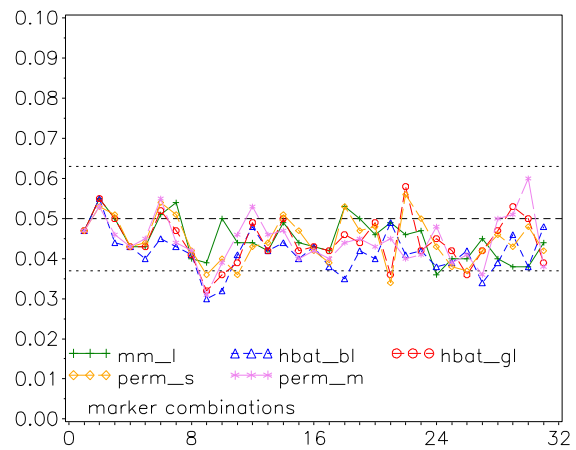
Type I error of a) Famhap b) Unphased  
1000 AAU Families, NMod1 H0\_(L)



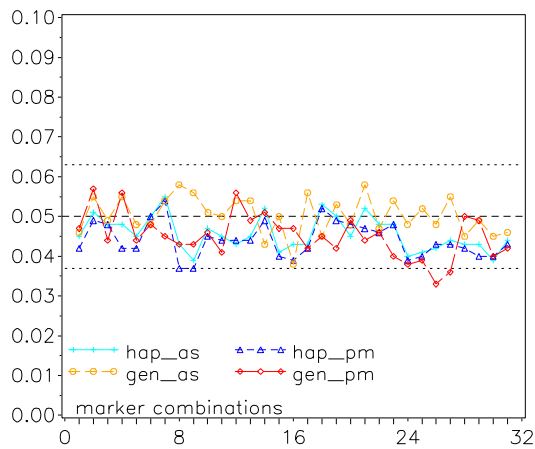
Type I error of FBAT  
1000 AAU Families, NMod1 H0\_(A)



Type I error of FBAT  
1000 AAU Families, NMod1 H0\_(L)

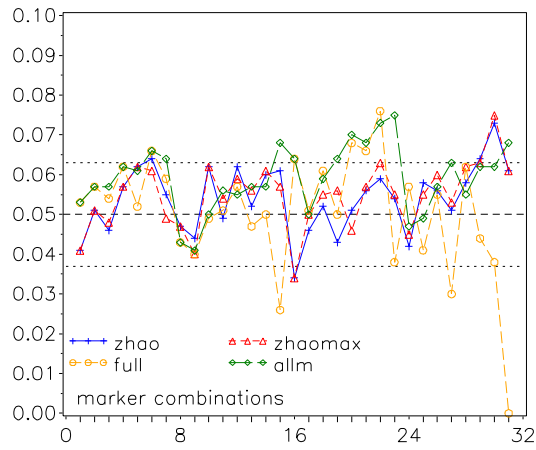


Hotel\_Fam: Type I error,  
1000 AAU families, NMod 1

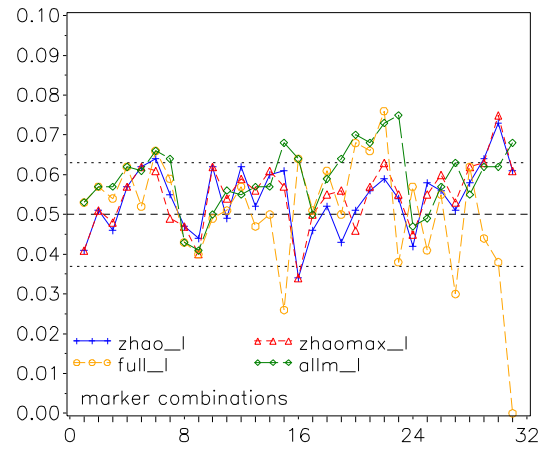


## NMod 2

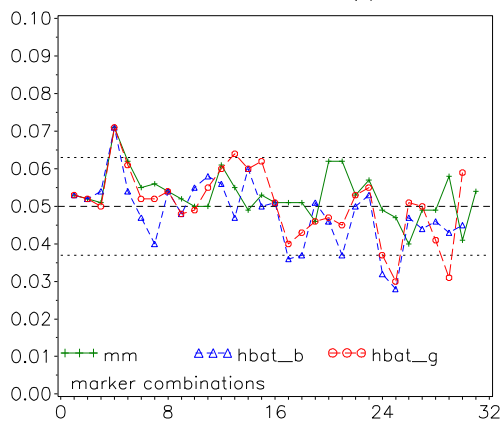
Type I error of a) Famhap b) Unphased  
1000 Trios, Nmod2, H0\_(A)



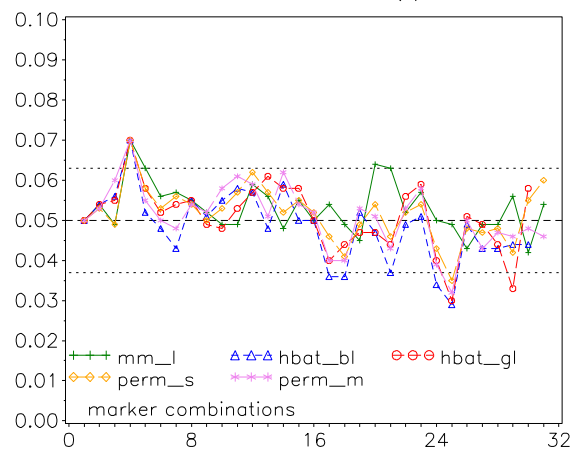
Type I error of a) Famhap b) Unphased  
1000 Trios, NMod2 H0\_(L)



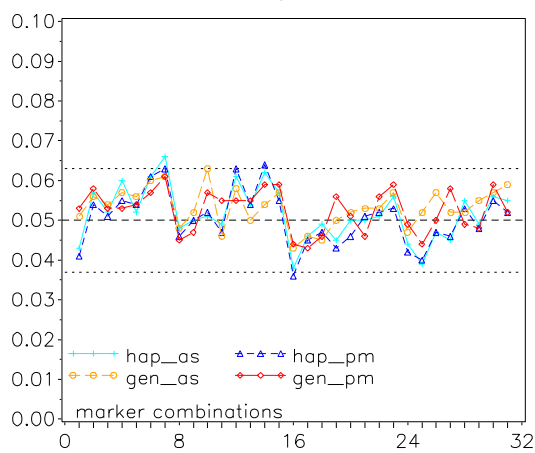
Type I error of FBAT  
1000 Trios, NMod2 H0\_(A)



Type I error of FBAT  
1000 Trios, NMod2 H0\_(L)

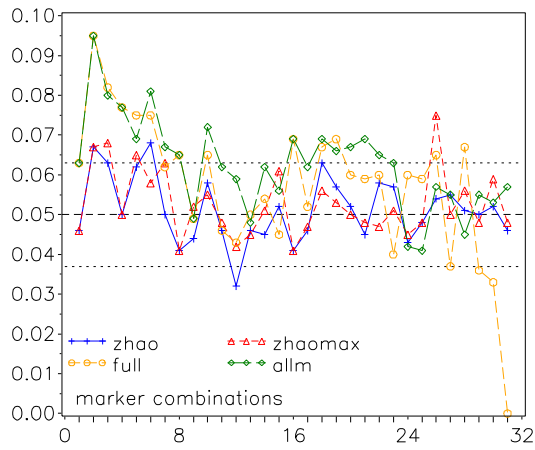


Hotel\_Fam: Type I error,  
1000 trios, NMod 2

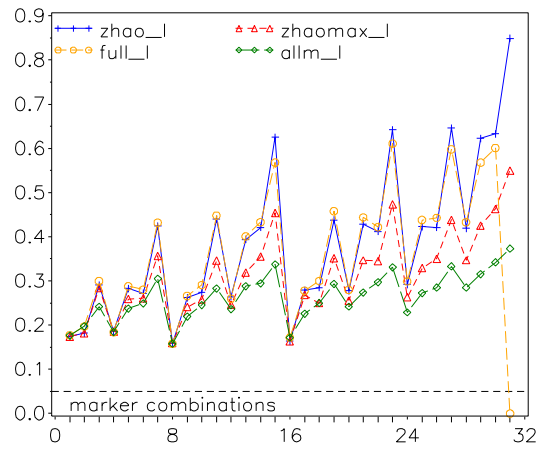




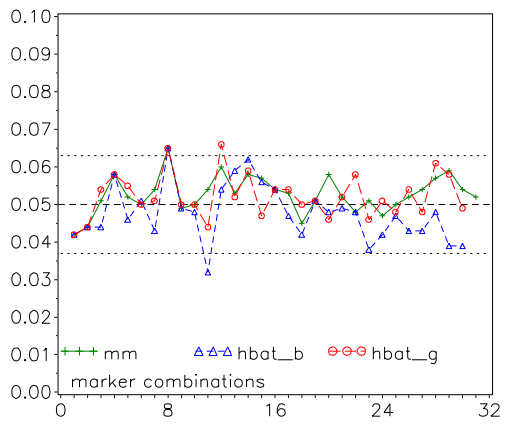
Type I error of a) Famhap b) Unphased  
1000 AA Families, Nmod2, H0\_(A)



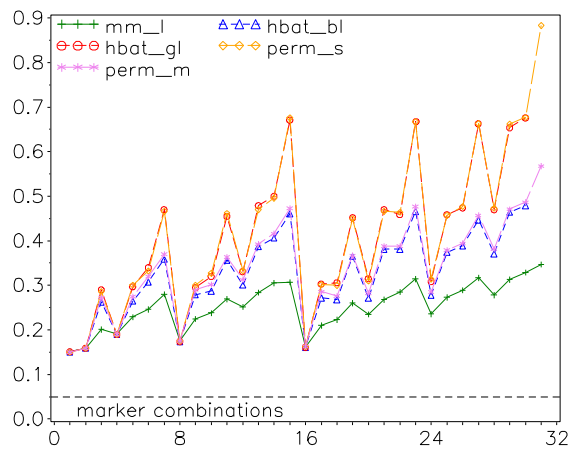
Type I error of a) Famhap b) Unphased  
1000 AA Families, NMod2 H0\_(L)



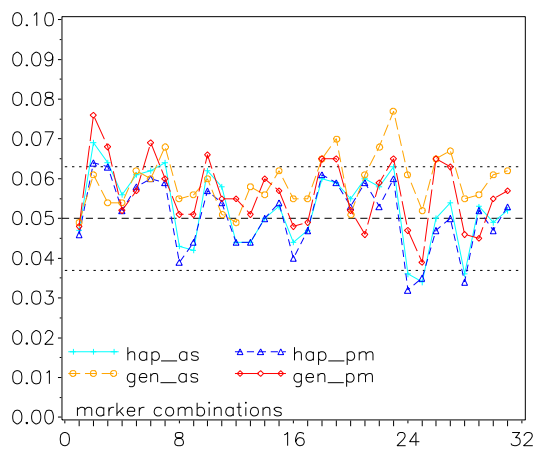
Type I error of FBAT  
1000 AA Families, NMod2 H0\_(A)



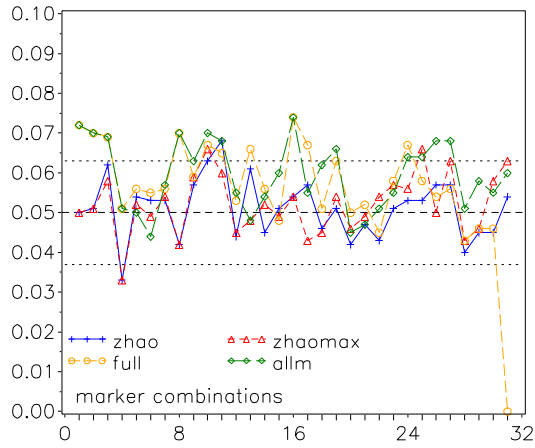
Type I error of FBAT  
1000 AA Families, NMod2 H0\_(L)



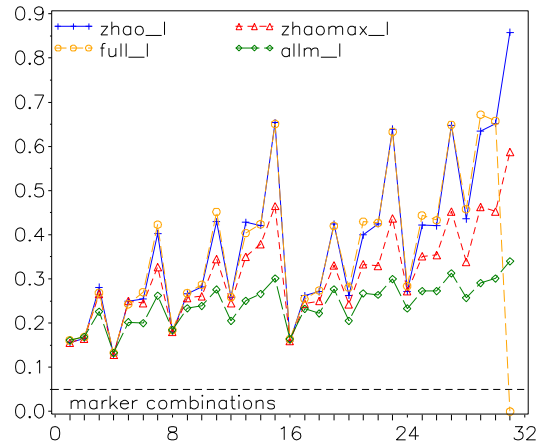
Hotel\_Fam: Type I error,  
1000 AA families, NMod 2



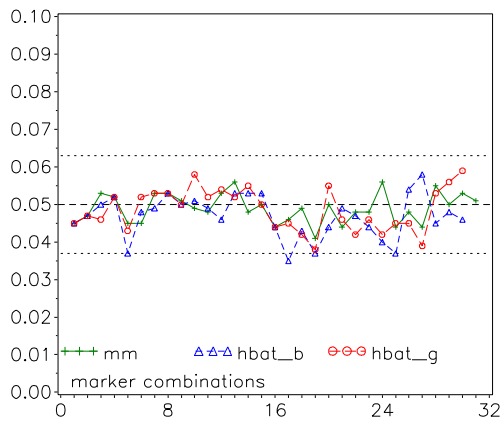
Type I error of a) Famhap b) Unphased  
1000 AAU Families, Nmod2, H0\_(A)



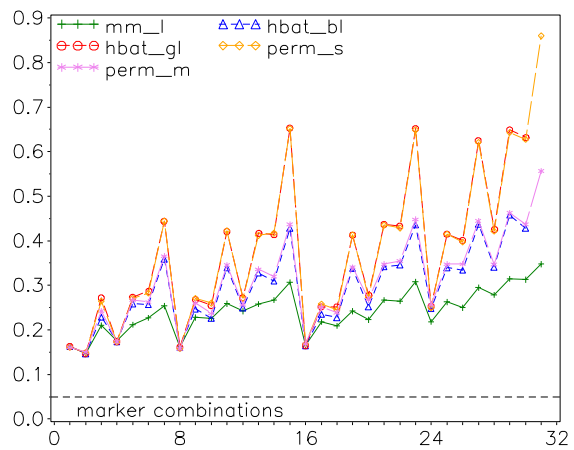
Type I error of a) Famhap b) Unphased  
1000 AAU Families, NMod2 H0\_(L)



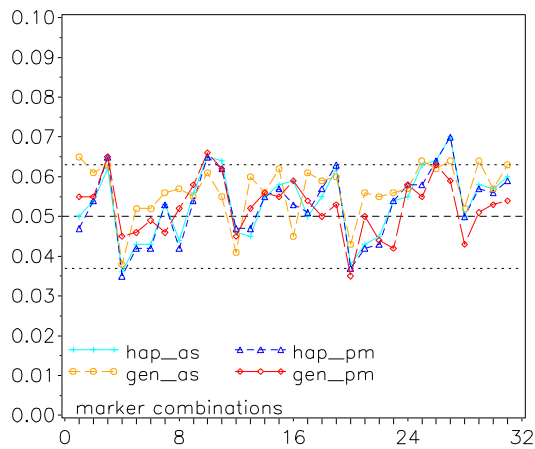
Type I error of FBAT  
1000 AAU Families, NMod2 H0\_(A)



Type I error of FBAT  
1000 AAU Families, NMod2 H0\_(L)

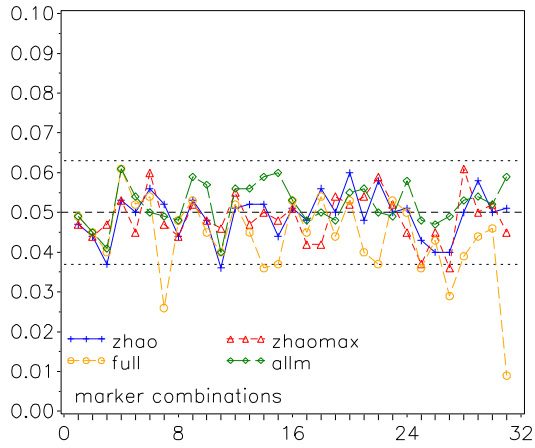


Hotel\_Fam: Type I error,  
1000 AAU families, NMod 2

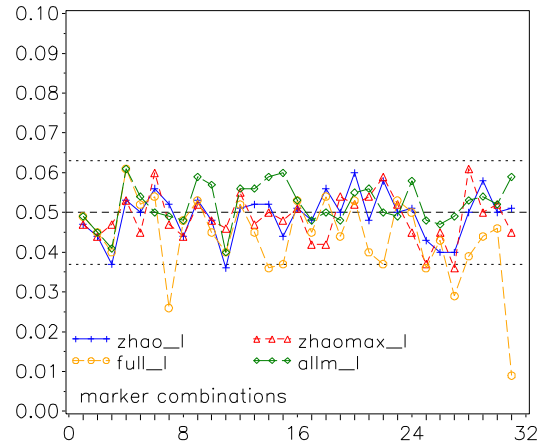


### NMod 3

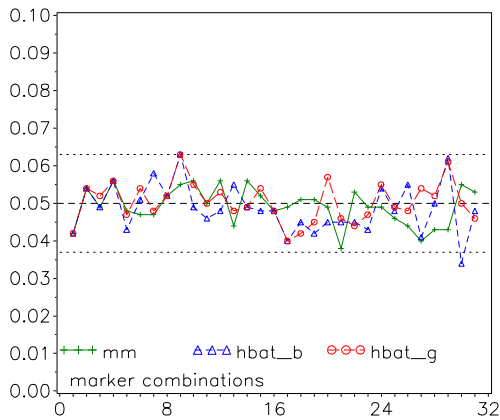
Type I error of a) Famhap b) Unphased  
1000 Trios, Nmod3, H0\_(A)



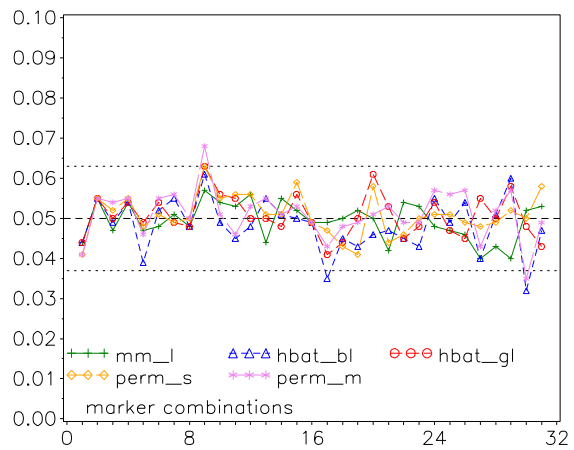
Type I error of a) Famhap b) Unphased  
1000 Trios, NMod3 H0\_(L)



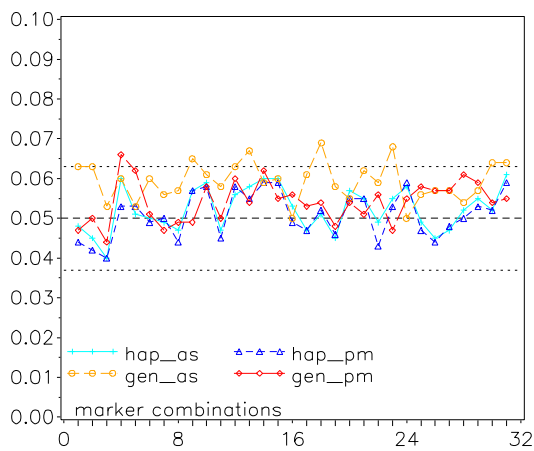
Type I error of FBAT  
1000 Trios, NMod3 H0\_(A)



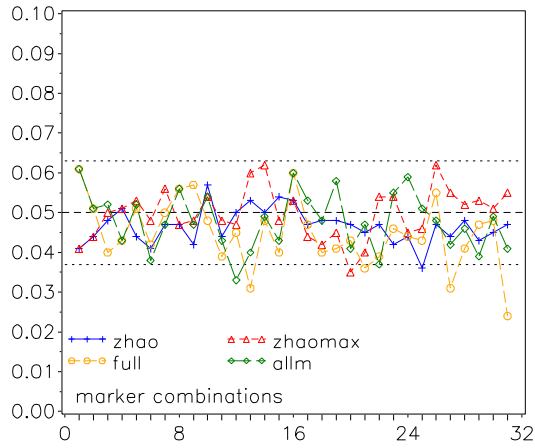
Type I error of FBAT  
1000 Trios, NMod3 H0\_(L)



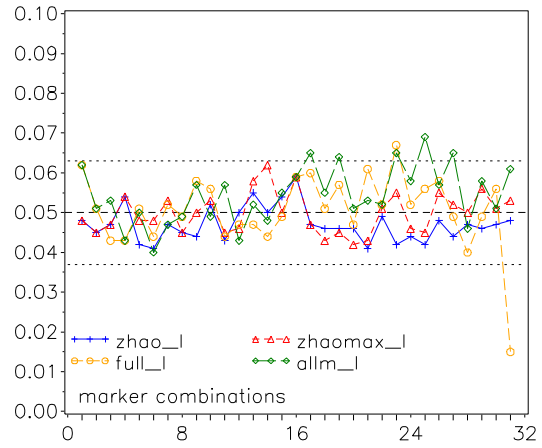
Hotel\_Fam: Type I error,  
1000 trios, NMod 3



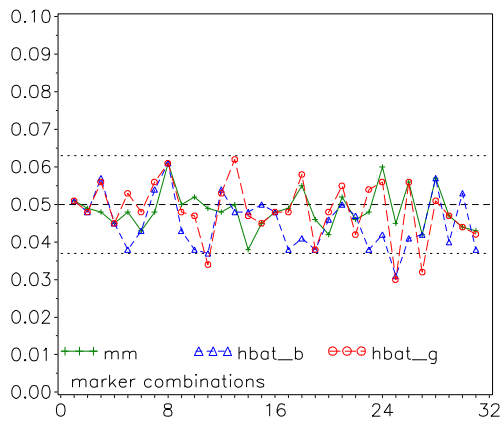
Type I error of a) Famhap b) Unphased  
1000 AA Families, Nmod3, H0\_(A)



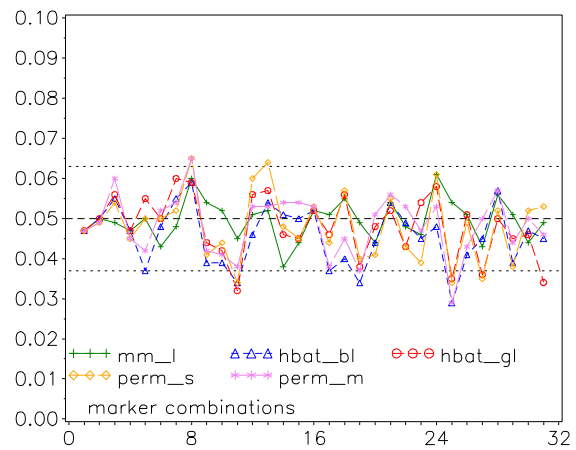
Type I error of a) Famhap b) Unphased  
1000 AA Families, NMod3 H0\_(L)



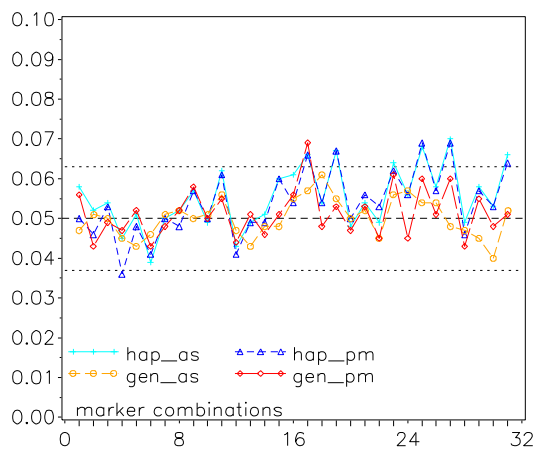
Type I error of FBAT  
1000 AA Families, NMod3 H0\_(A)



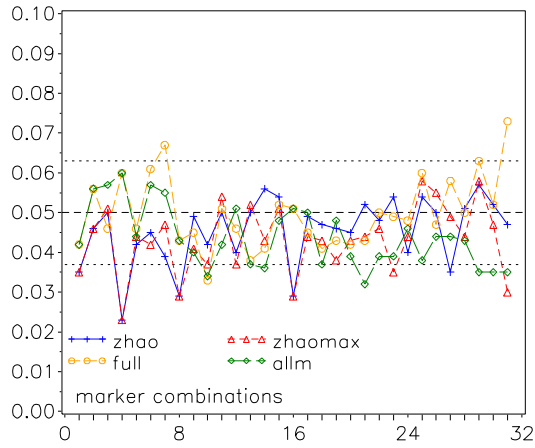
Type I error of FBAT  
1000 AA Families, NMod3 H0\_(L)



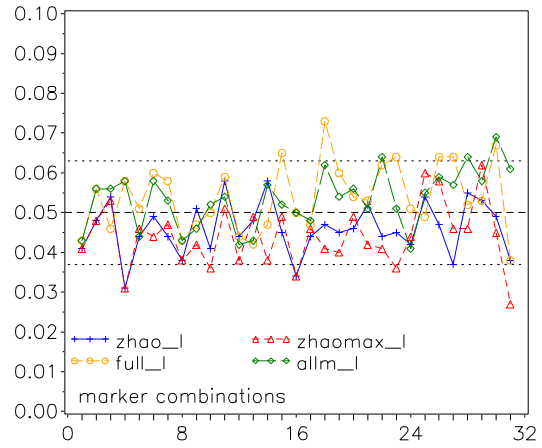
Hotel\_Fam: Type I error,  
1000 AA families, NMod 3



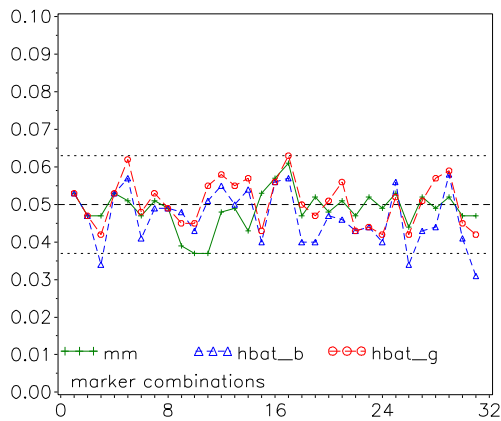
Type I error of a) Famhap b) Unphased  
1000 AAU Families, Nmod3, H0\_(A)



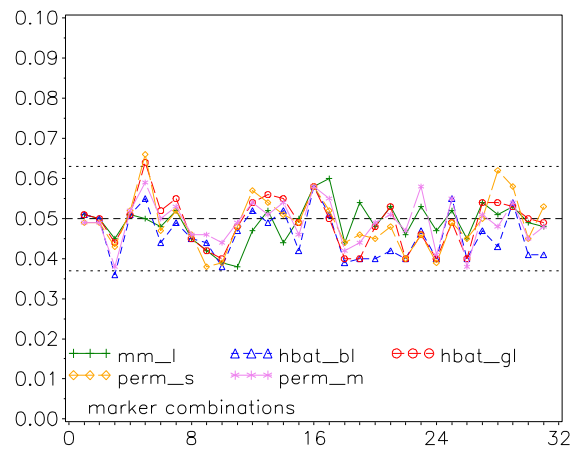
Type I error of a) Famhap b) Unphased  
1000 AAU Families, NMod3 H0\_(L)



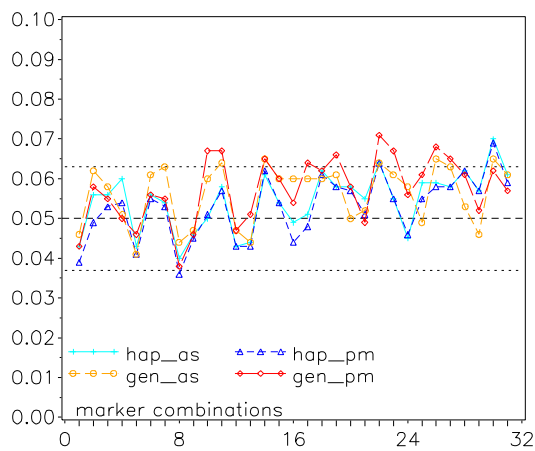
Type I error of FBAT  
1000 AAU Families, NMod3 H0\_(A)



Type I error of FBAT  
1000 AAU Families, NMod3 H0\_(L)



Hotel\_Fam: Type I error,  
1000 AAU families, NMod 3

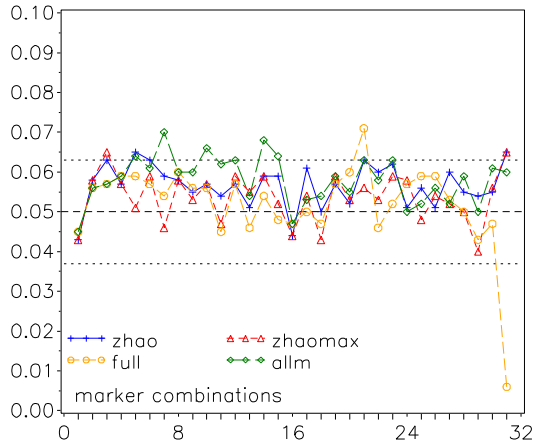


# Fehler 1. Art: Grafiken für alle Markerkombinationen

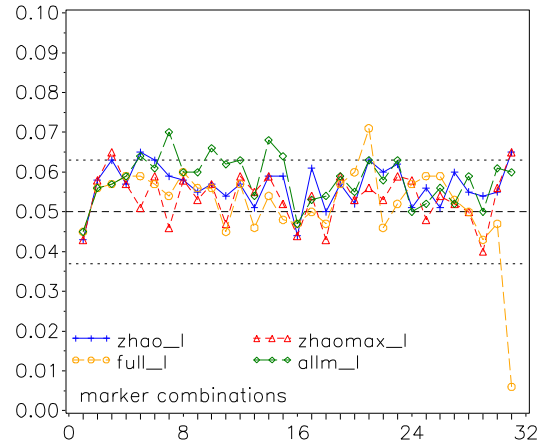
Testniveau  $\alpha = 0.05$ , 2500 Familien

NMod 1

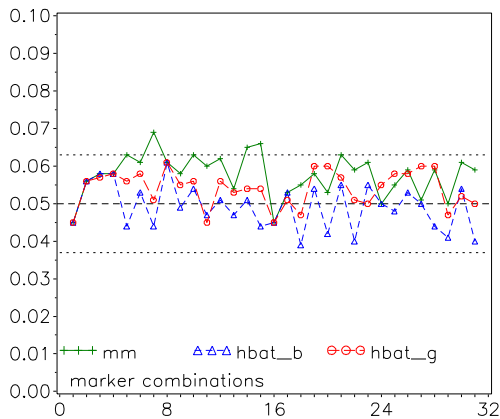
Type I error of a) Famhap b) Unphased  
2500 Trios, Nmod1, H0\_(A)



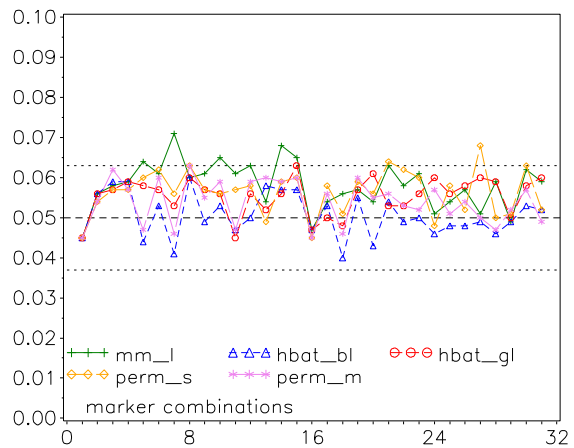
Type I error of a) Famhap b) Unphased  
2500 Trios, NMod1 H0\_(L)



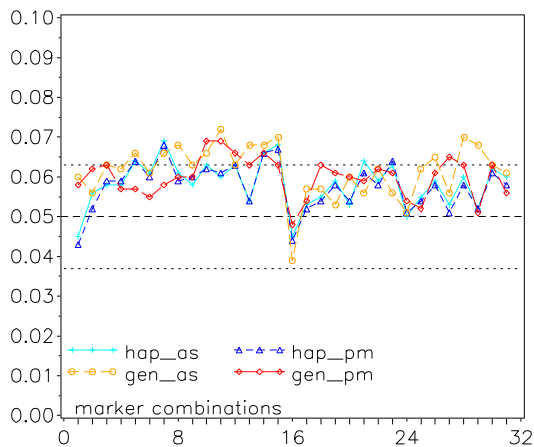
Type I error of FBAT  
2500 Trios, NMod1 H0\_(A)



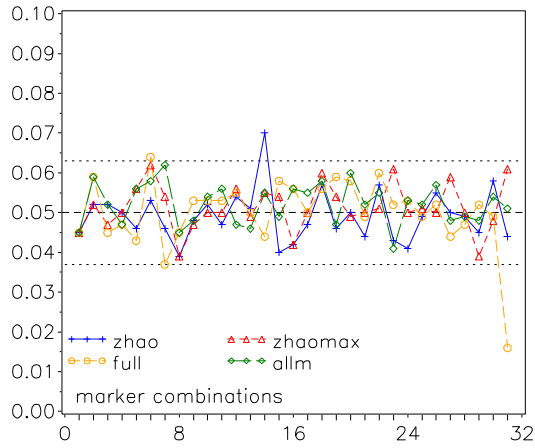
Type I error of FBAT  
2500 Trios, NMod1 H0\_(L)



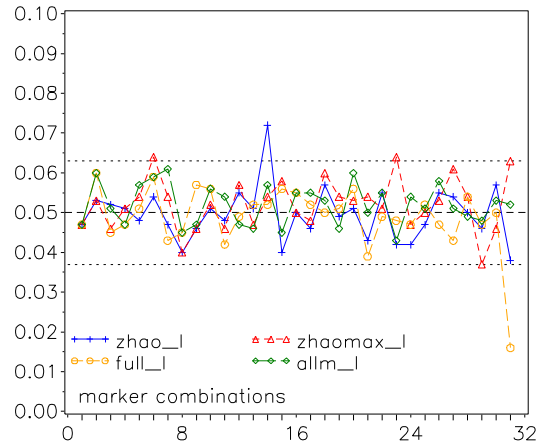
Hotel\_Fam: Type I error,  
2500 trios, NMod 1



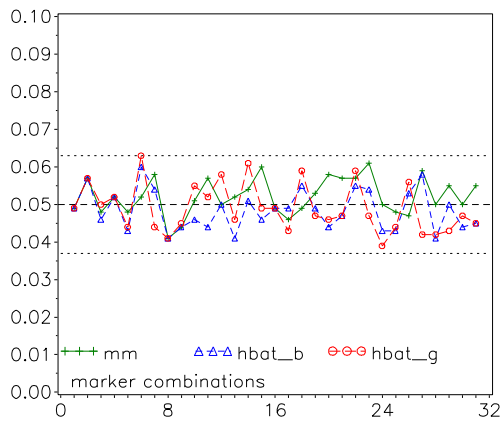
Type I error of a) Famhap b) Unphased  
2500 AA Families, Nmod1, H0\_(A)



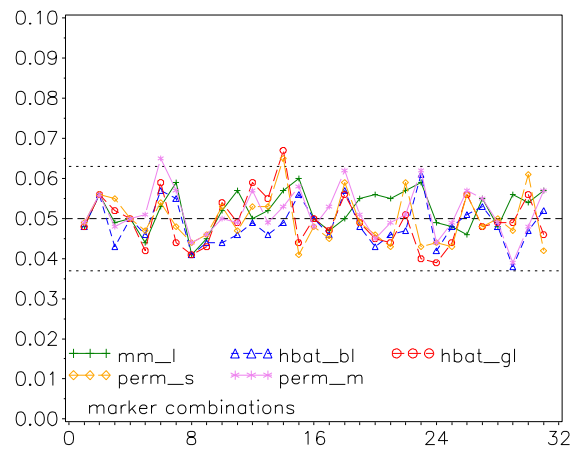
Type I error of a) Famhap b) Unphased  
2500 AA Families, NMod1 H0\_(L)



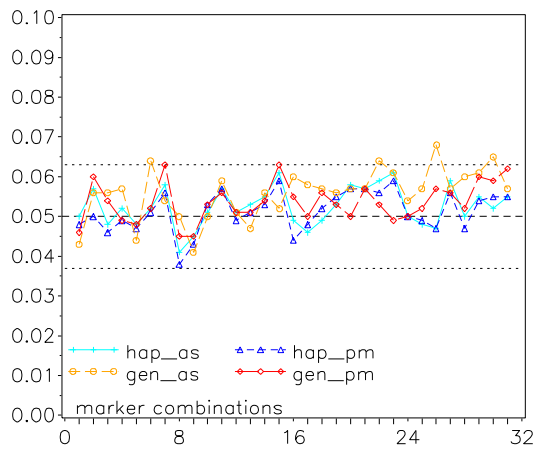
Type I error of FBAT  
2500 AA Families, NMod1 H0\_(A)



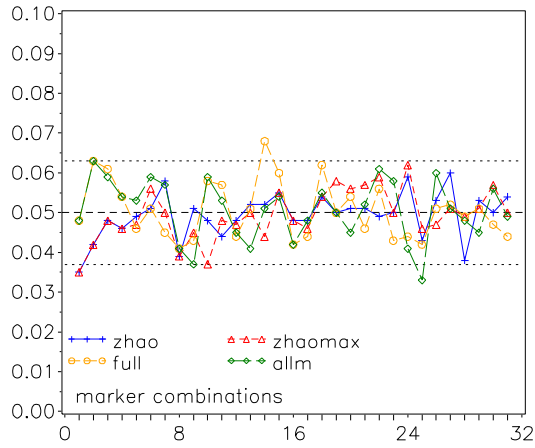
Type I error of FBAT  
2500 AA Families, NMod1 H0\_(L)



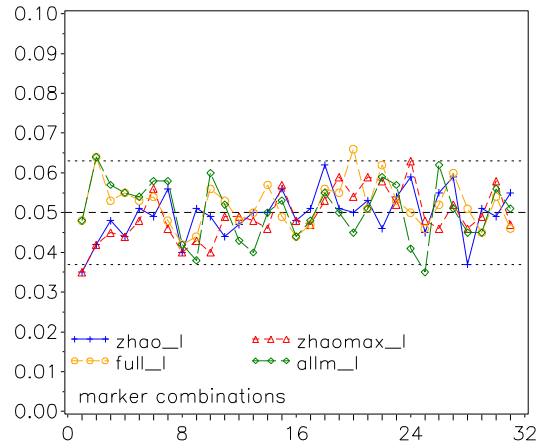
Hotel\_Fam: Type I error,  
2500 AA families, NMod 1



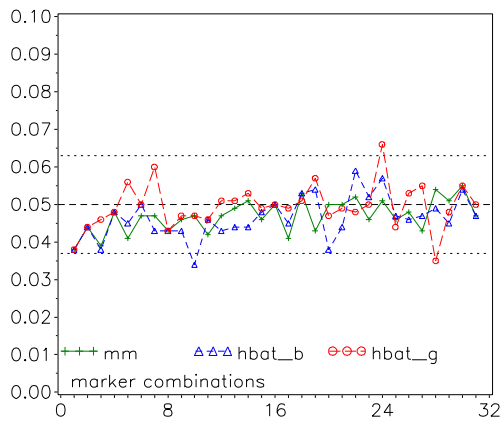
Type I error of a) Famhap b) Unphased  
2500 AAU Families, NMod1, H0\_(A)



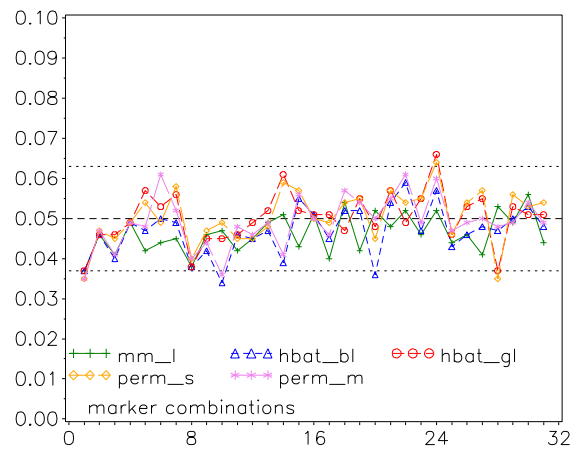
Type I error of a) Famhap b) Unphased  
2500 AAU Families, NMod1 H0\_(L)



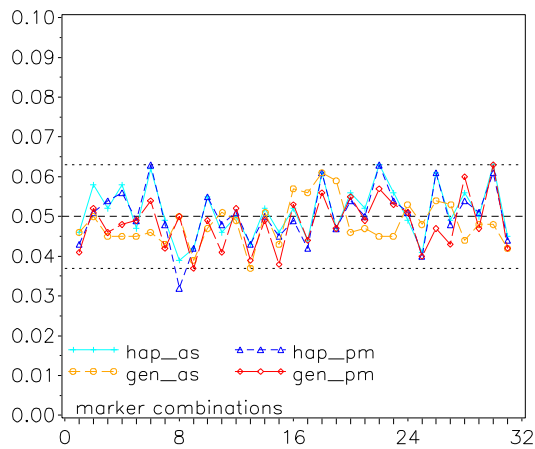
Type I error of FBAT  
2500 AAU Families, NMod1 H0\_(A)



Type I error of FBAT  
2500 AAU Families, NMod1 H0\_(L)



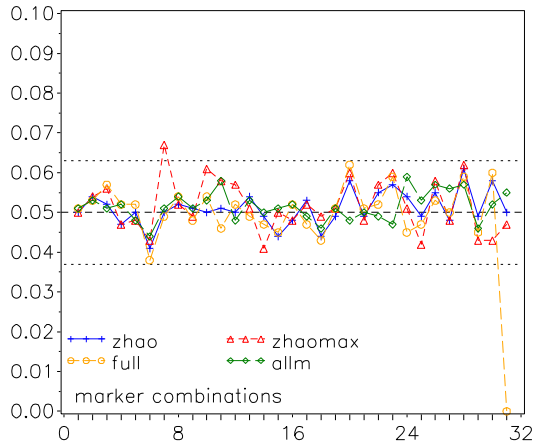
Hotel\_Fam: Type I error,  
2500 AAU families, NMod 1



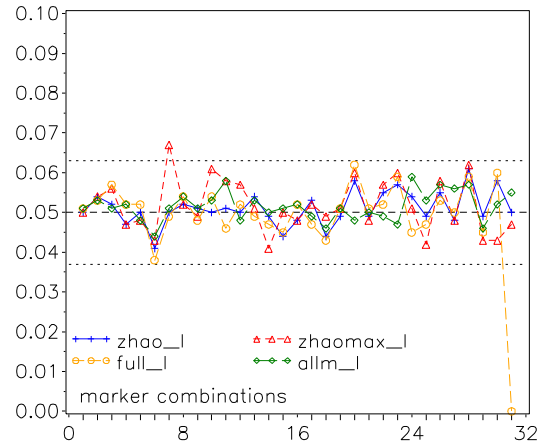


## NMod 2

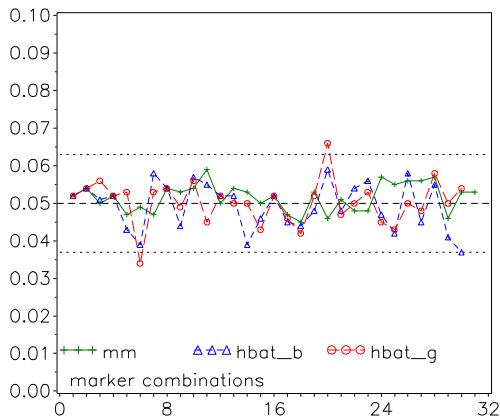
Type I error of a) Famhap b) Unphased  
2500 Trios, Nmod2, H0\_(A)



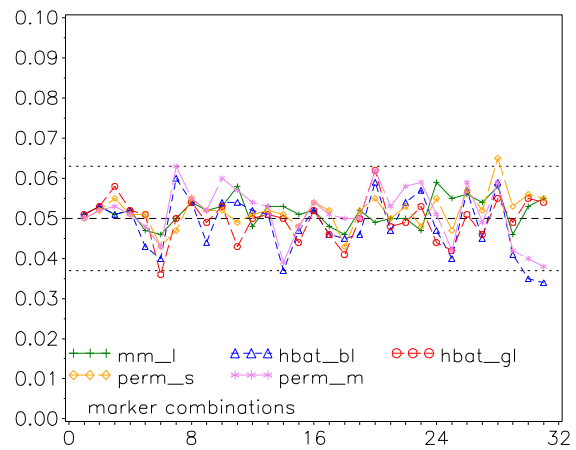
Type I error of a) Famhap b) Unphased  
2500 Trios, NMod2 H0\_(L)



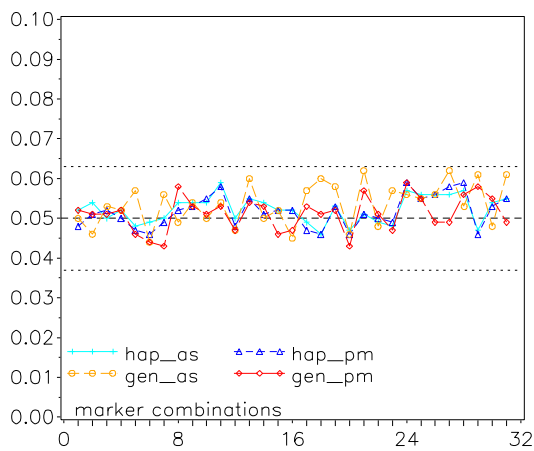
Type I error of FBAT  
2500 Trios, NMod2 H0\_(A)



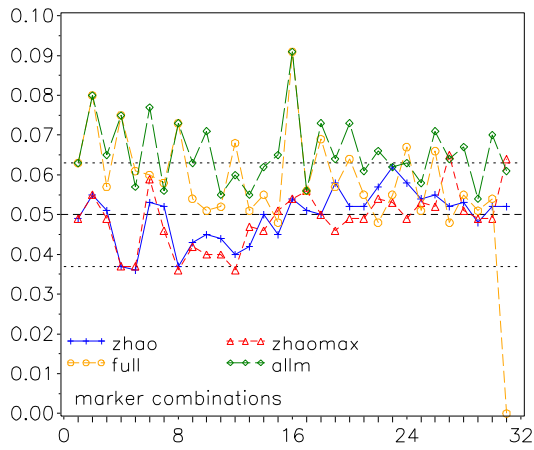
Type I error of FBAT  
2500 Trios, NMod2 H0\_(L)



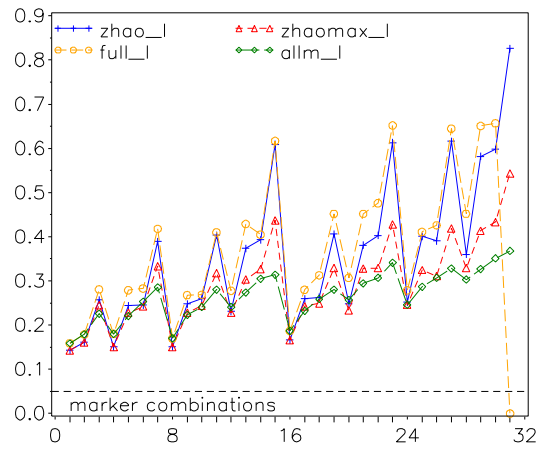
Hotel\_Fam: Type I error,  
2500 trios, NMod 2



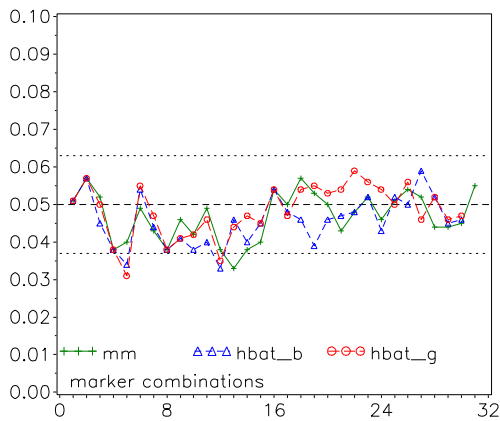
Type I error of a) Famhap b) Unphased  
2500 AA Families, Nmod2, H0\_(A)



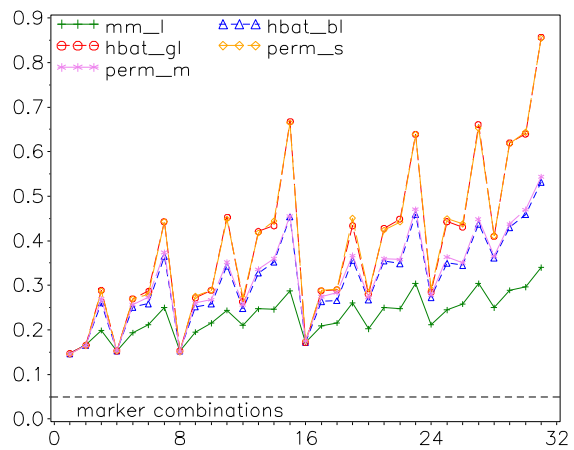
Type I error of a) Famhap b) Unphased  
2500 AA Families, NMod2 H0\_(L)



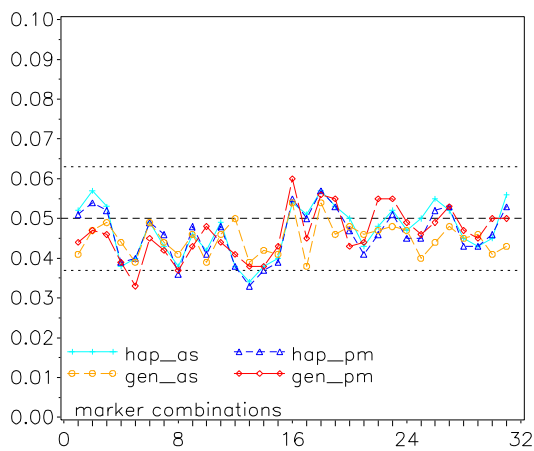
Type I error of FBAT  
2500 AA Families, NMod2 H0\_(A)



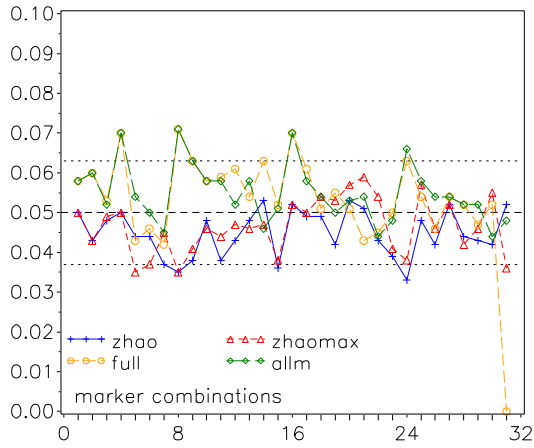
Type I error of FBAT  
2500 AA Families, NMod2 H0\_(L)



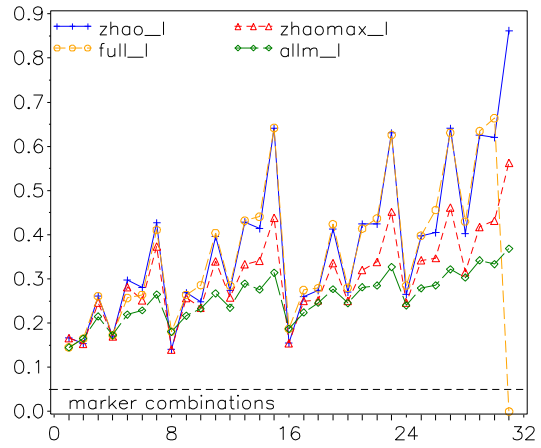
Hotel\_Fam: Type I error,  
2500 AA families, NMod 2



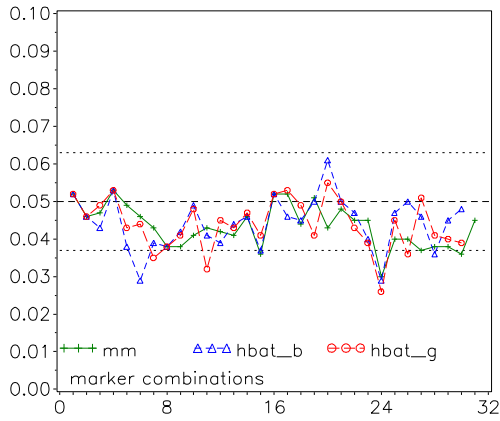
Type I error of a) Famhap b) Unphased  
2500 AAU Families, NMod2, H0\_(A)



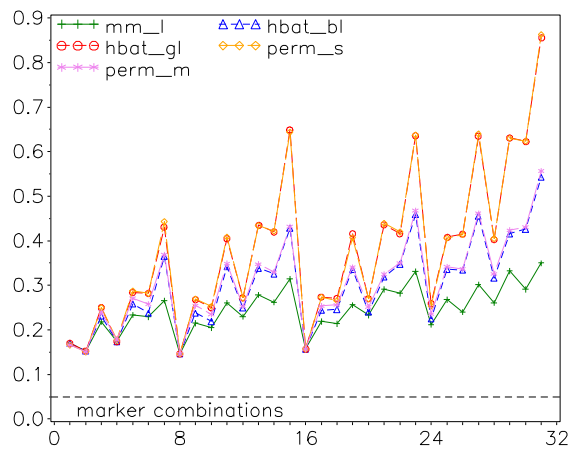
Type I error of a) Famhap b) Unphased  
2500 AAU Families, NMod2 H0\_(L)



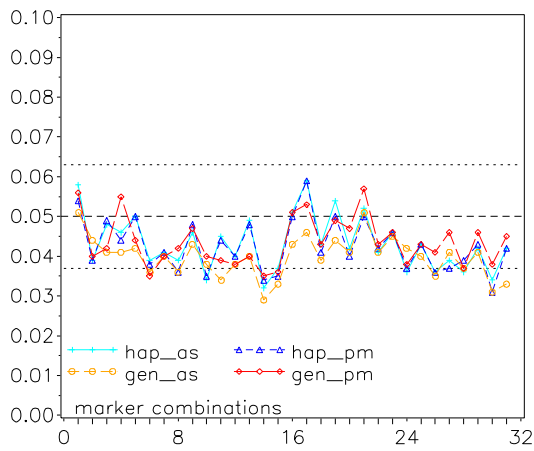
Type I error of FBAT  
2500 AAU Families, NMod2 H0\_(A)



Type I error of FBAT  
2500 AAU Families, NMod2 H0\_(L)

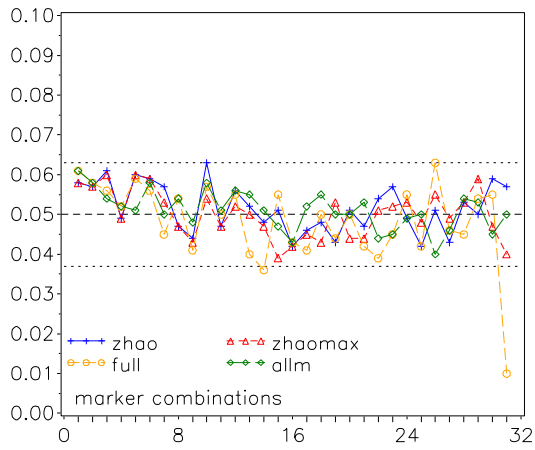


Hotel\_Fam: Type I error,  
2500 AAU families, NMod 2

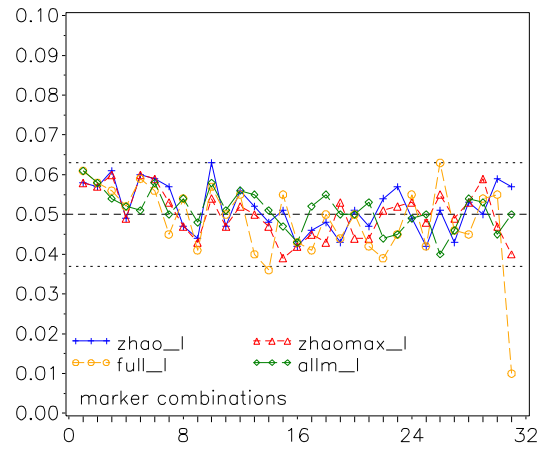


### NMod 3

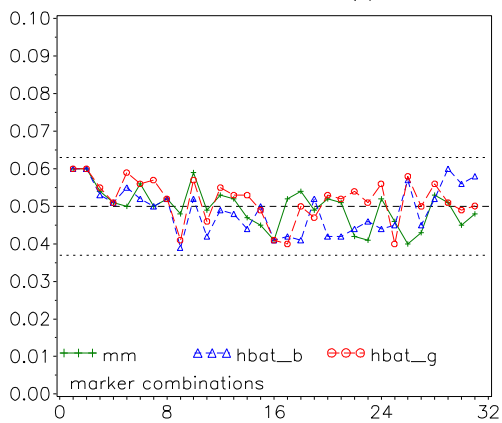
Type I error of a) Famhap b) Unphased  
2500 Trios, Nmod3, H0\_(A)



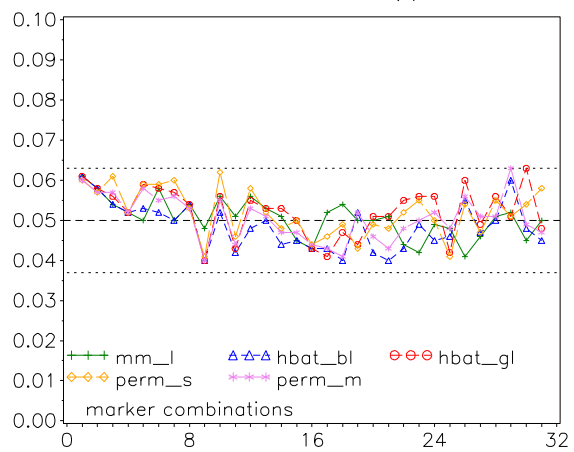
Type I error of a) Famhap b) Unphased  
2500 Trios, NMod3 H0\_(L)



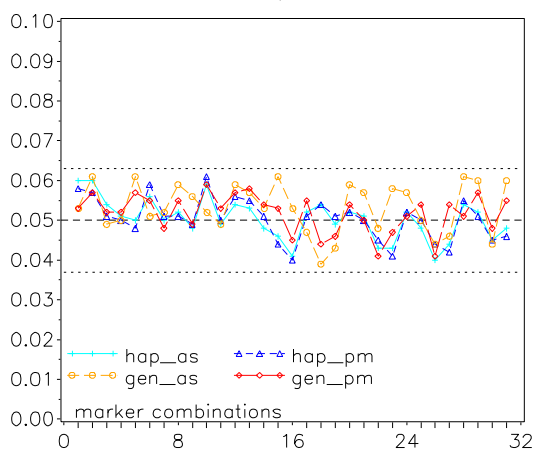
Type I error of FBAT  
2500 Trios, NMod3 H0\_(A)



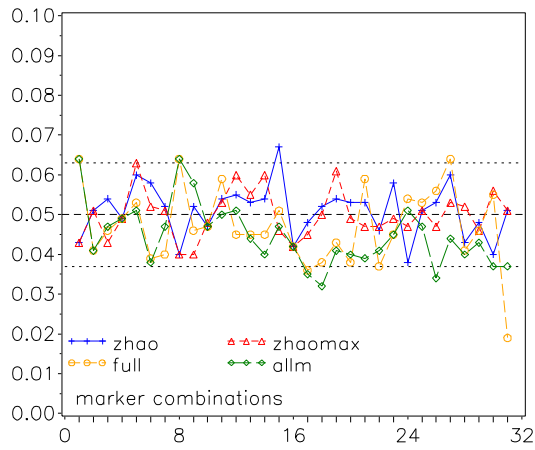
Type I error of FBAT  
2500 Trios, NMod3 H0\_(L)



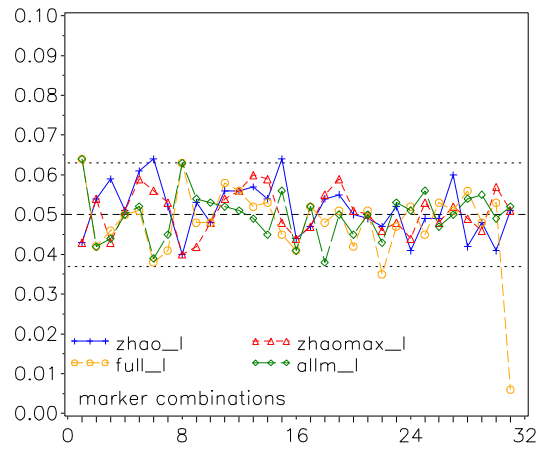
Hotel\_Fam: Type I error,  
2500 trios, NMod 3



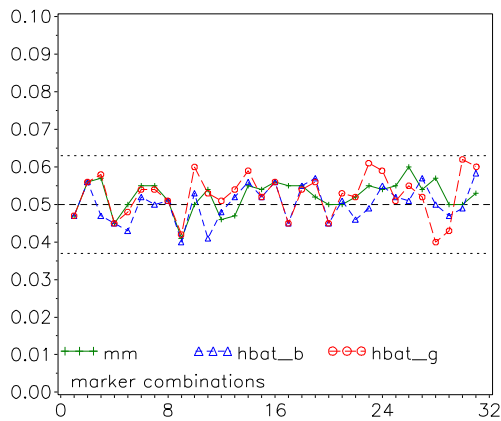
Type I error of a) Famhap b) Unphased  
2500 AA Families, Nmod3, H0\_(A)



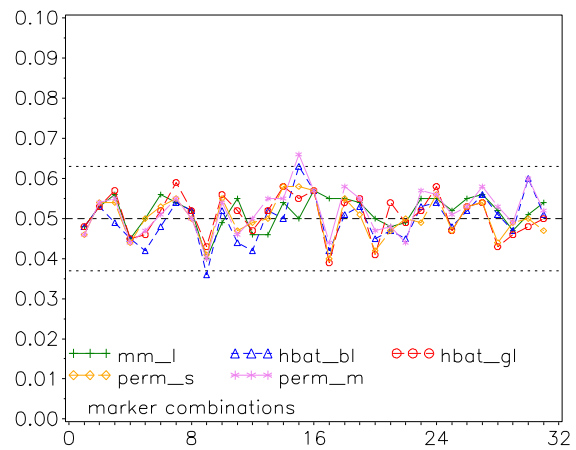
Type I error of a) Famhap b) Unphased  
2500 AA Families, NMod3 H0\_(L)



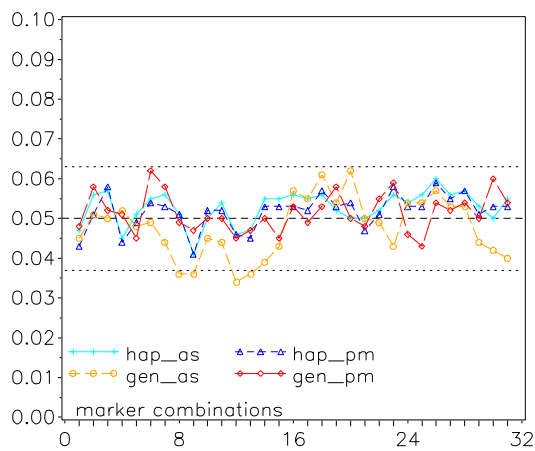
Type I error of FBAT  
2500 AA Families, NMod3 H0\_(A)



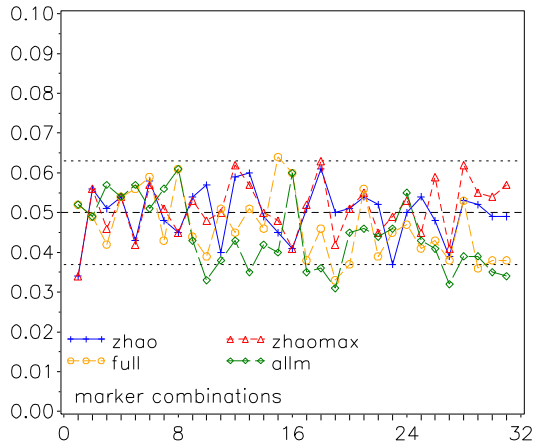
Type I error of FBAT  
2500 AA Families, NMod3 H0\_(L)



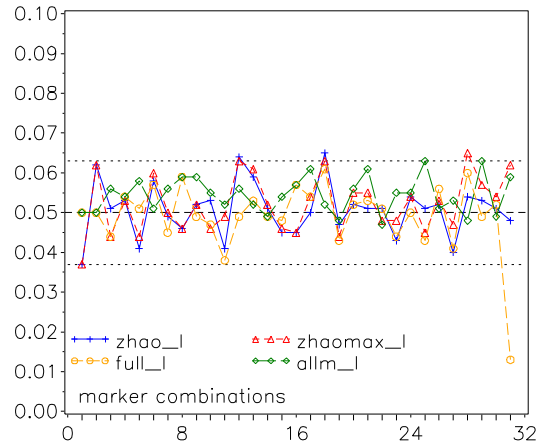
Hotel\_Fam: Type I error,  
2500 AA families, NMod 3



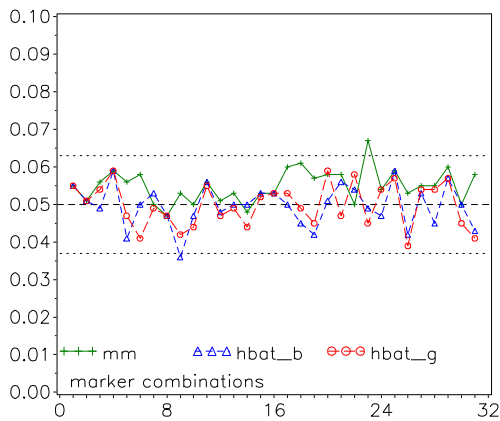
Type I error of a) Famhap b) Unphased  
2500 AAUU Families, Nmod3, H0\_(A)



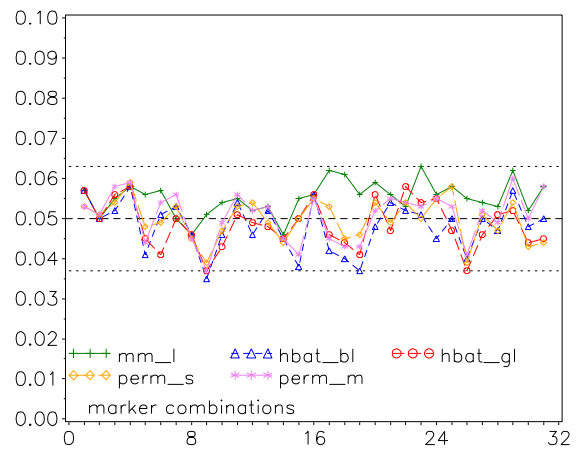
Type I error of a) Famhap b) Unphased  
2500 AAUU Families, NMod3 H0\_(L)



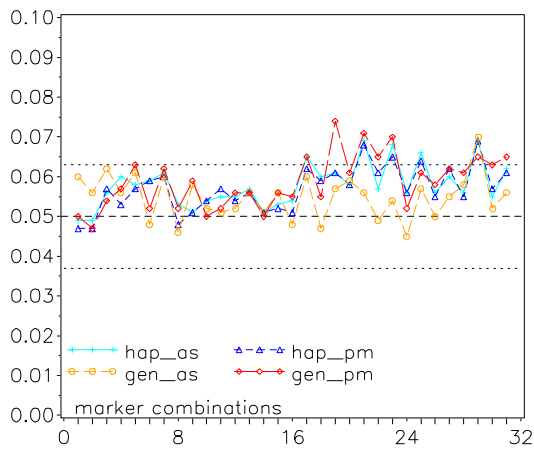
Type I error of FBAT  
2500 AAUU Families, NMod3 H0\_(A)



Type I error of FBAT  
2500 AAUU Families, NMod3 H0\_(L)



Hotel\_Fam: Type I error,  
2500 AAUU families, NMod 3

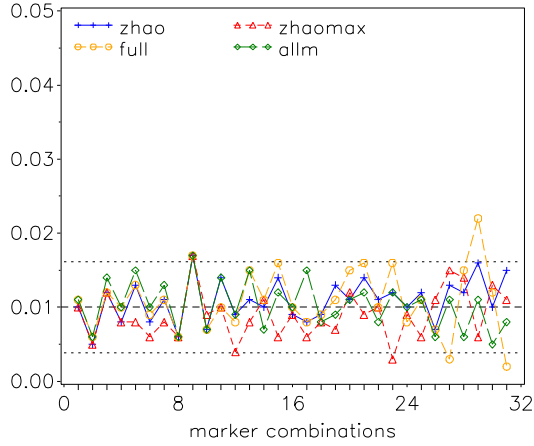


# Fehler 1. Art: Diagramme für alle Markerkombinationen

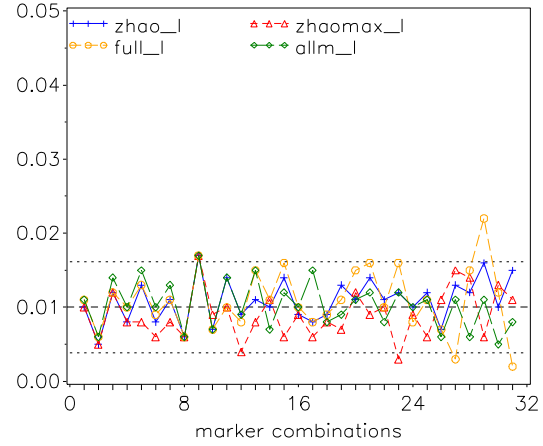
Testniveau  $\alpha = 0.01$ , 200 Familien

NMod 1

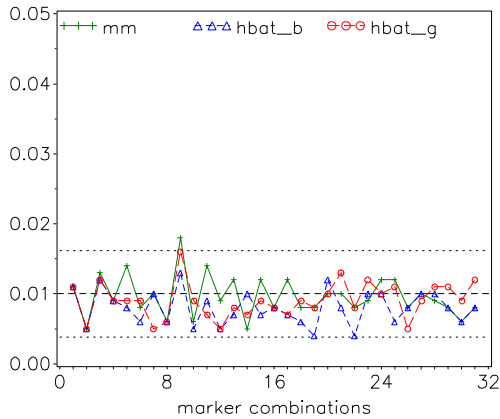
Type I error of a) Famhap b) Unphased  
200 Trios, Nmod1, H0\_(A)



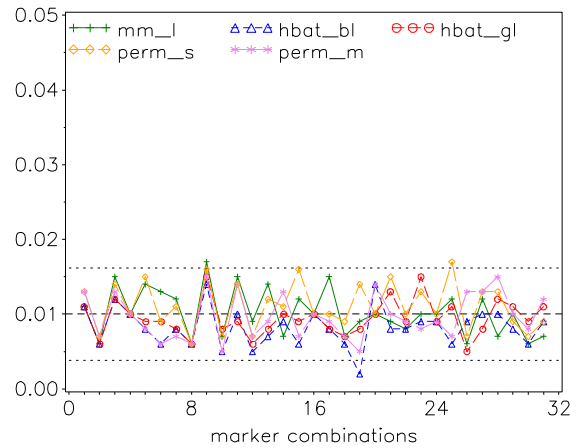
Type I error of a) Famhap b) Unphased  
200 Trios, NMod1 H0\_(L)



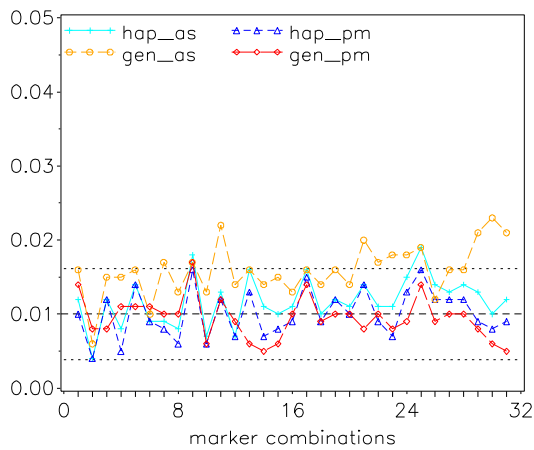
Type I error of FBAT, alpha=0.01  
200 Trios, NMod1 H0\_(A)



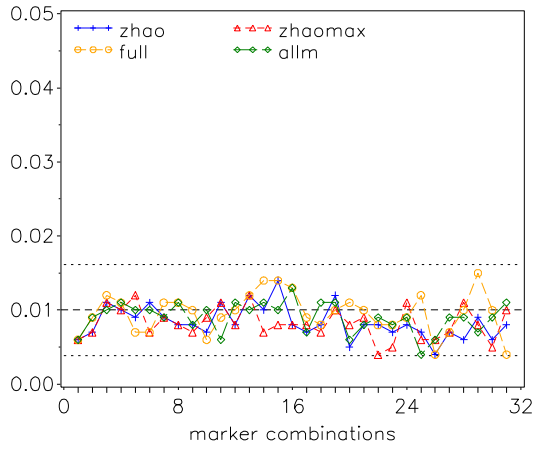
Type I error of FBAT, alpha=0.01  
200 Trios, NMod1 H0\_(L)



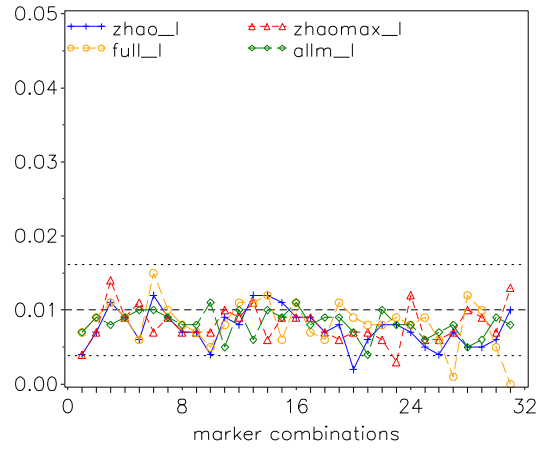
Hotel\_Fam: Type I error, alpha=0.01  
200 trios, NMod 1



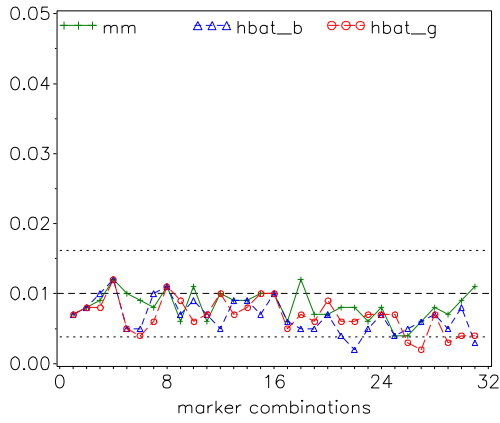
Type I error of a) Famhap b) Unphased  
200 AA Families, NMod1, H0\_(A)



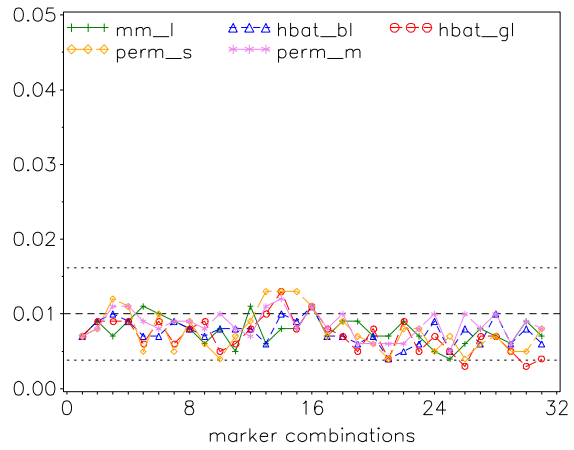
Type I error of a) Famhap b) Unphased  
200 AA Families, NMod1 H0\_(L)



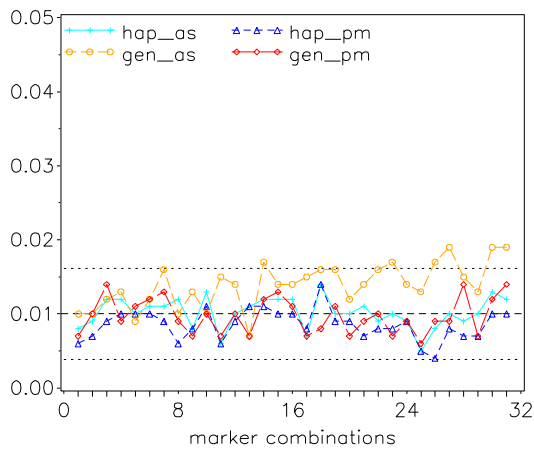
Type I error of FBAT, alpha=0.01  
200 AA Families, NMod1 H0\_(A)



Type I error of FBAT, alpha=0.01  
200 AA Families, NMod1 H0\_(L)

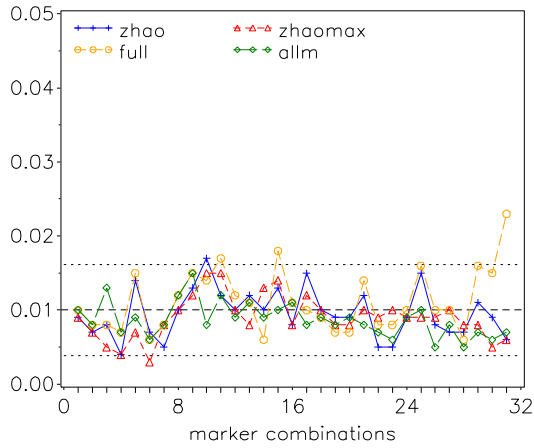


Hotel\_Fam: Type I error, alpha=0.01  
200 AA families, NMod 1

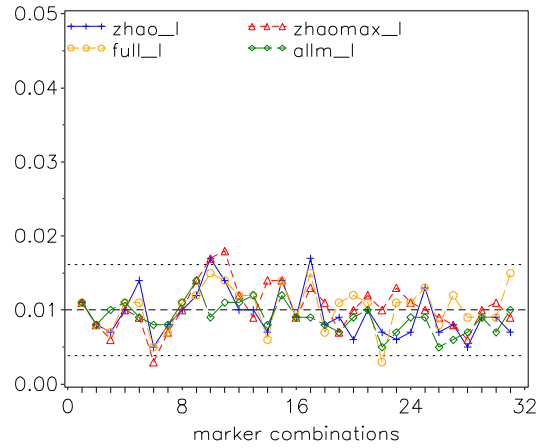




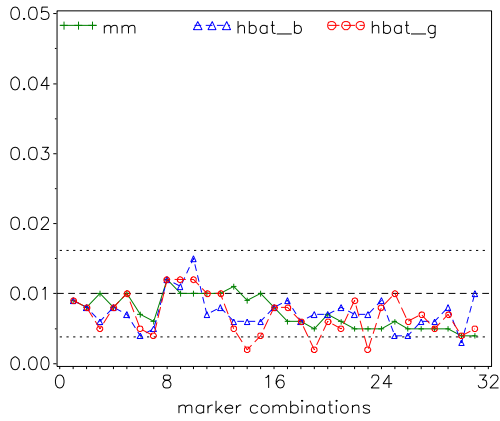
Type I error of a) Famhap b) Unphased  
200 AAU Families, NMod1, H0\_(A)



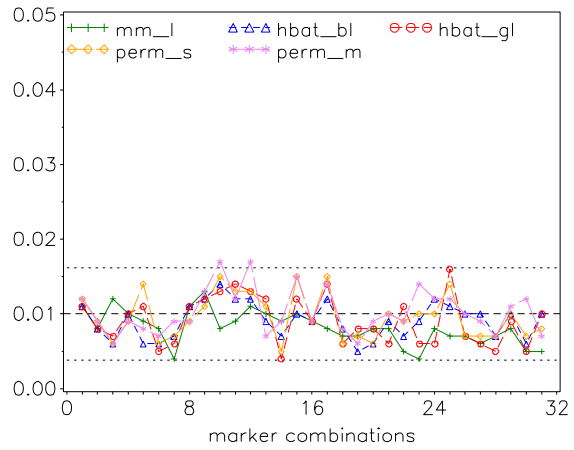
Type I error of a) Famhap b) Unphased  
200 AAU Families, NMod1 H0\_(L)



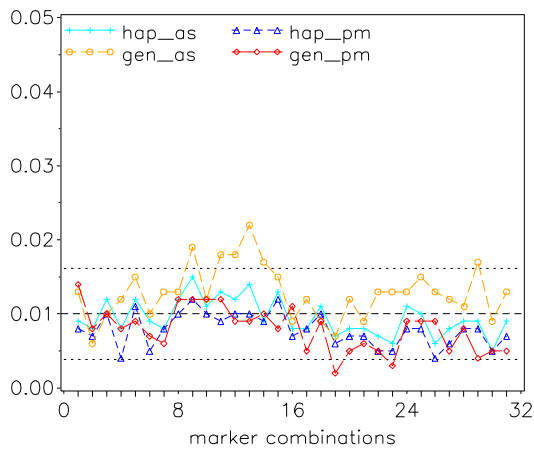
Type I error of FBAT, alpha=0.01  
200 AAU Families, NMod1 H0\_(A)



Type I error of FBAT, alpha=0.01  
200 AAU Families, NMod1 H0\_(L)

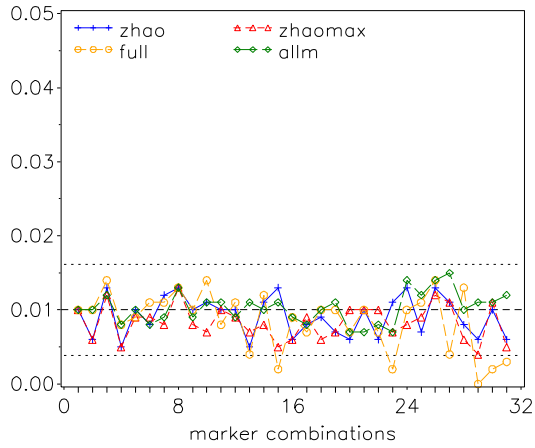


Hotel\_Fam: Type I error, alpha=0.01  
200 AAU families, NMod 1

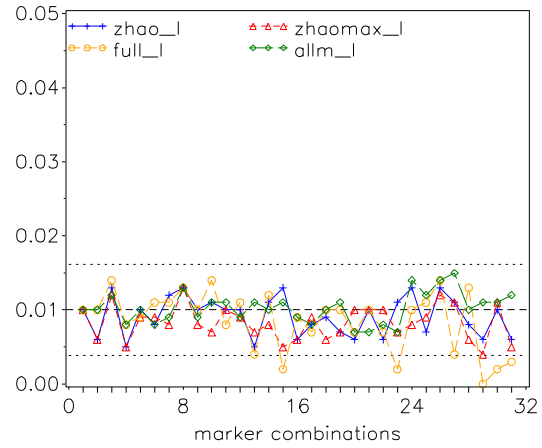


## NMod 2

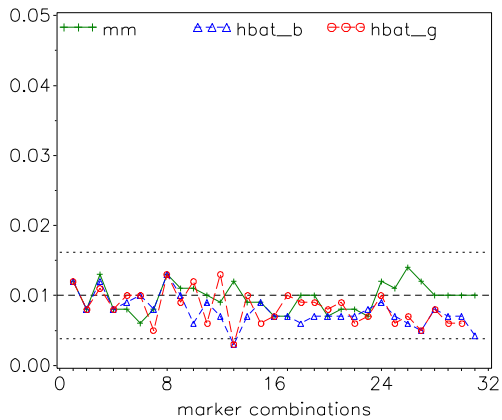
Type I error of a) Famhap b) Unphased  
200 Trios, NMod2, H0\_(A)



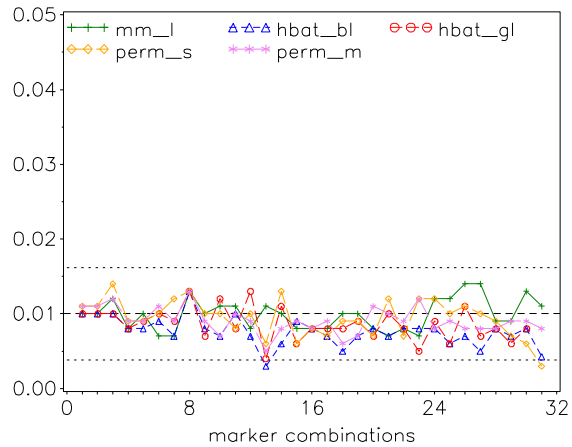
Type I error of a) Famhap b) Unphased  
200 Trios, NMod2 H0\_(L)



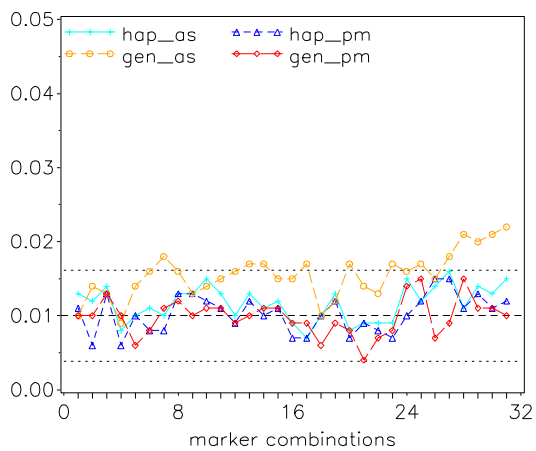
Type I error of FBAT, alpha=0.01  
200 Trios, NMod2 H0\_(A)



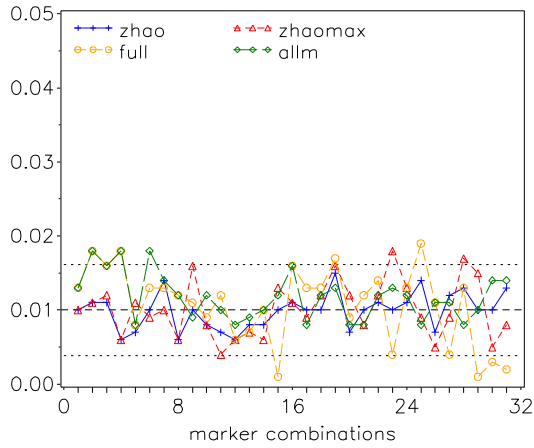
Type I error of FBAT, alpha=0.01  
200 Trios, NMod2 H0\_(L)



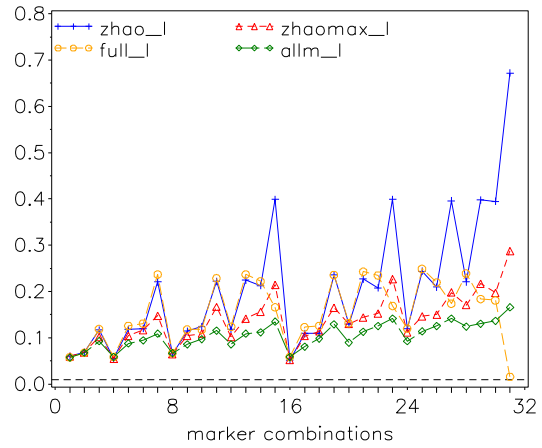
Hotel\_Fam: Type I error, alpha=0.01  
200 trios, NMod 2



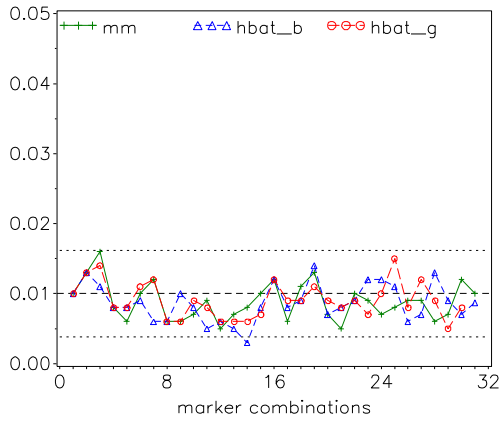
Type I error of a) Famhap b) Unphased  
200 AA Families, NMod2, H0\_(A)



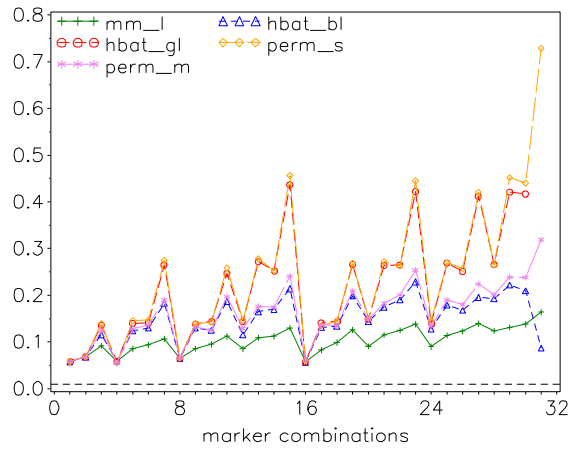
Type I error of a) Famhap b) Unphased  
200 AA Families, NMod2 H0\_(L)



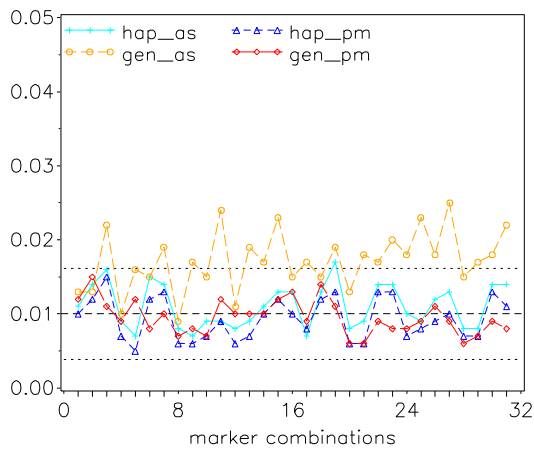
Type I error of FBAT, alpha=0.01  
200 AA Families, NMod2 H0\_(A)



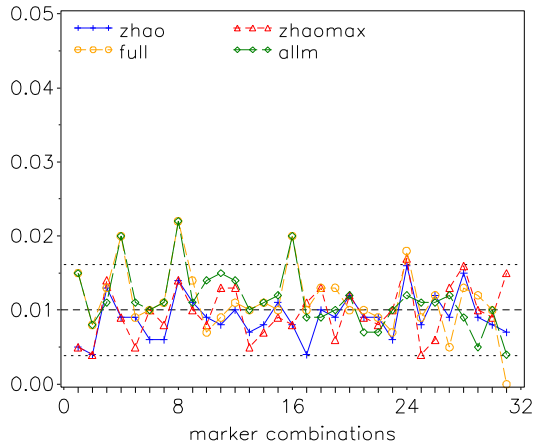
Type I error of FBAT, alpha=0.01  
200 AA Families, NMod2 H0\_(L)



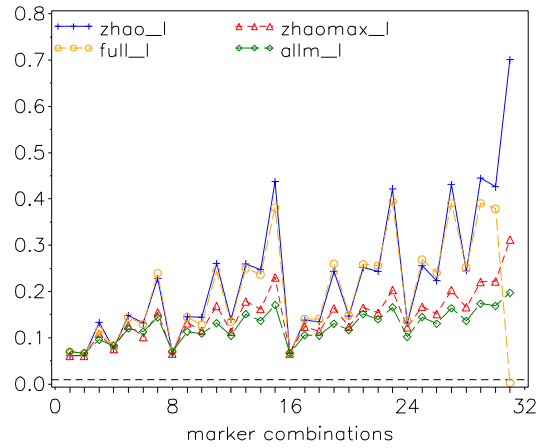
Hotel\_Fam: Type I error, alpha=0.01  
200 AA families, NMod 2



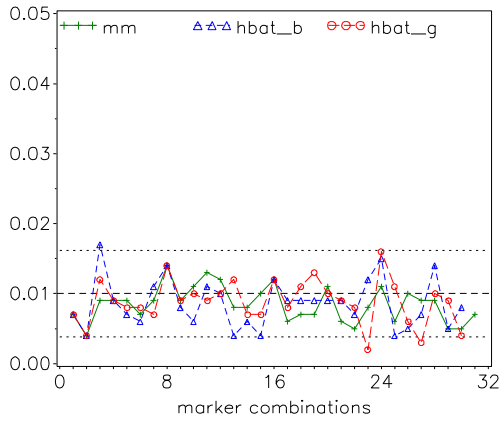
Type I error of a) Famhap b) Unphased  
200 AAU Families, NMod2, H0\_(A)



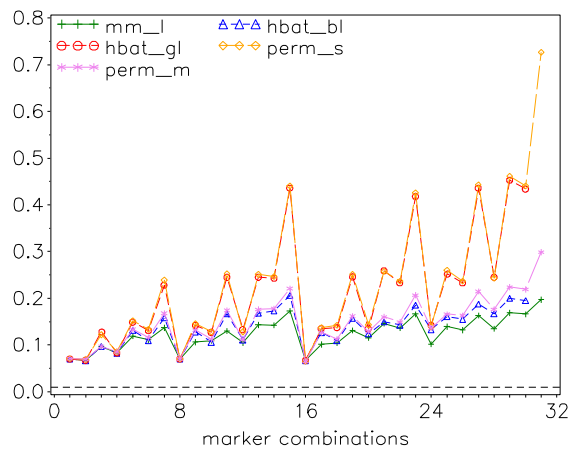
Type I error of a) Famhap b) Unphased  
200 AAU Families, NMod2 H0\_(L)



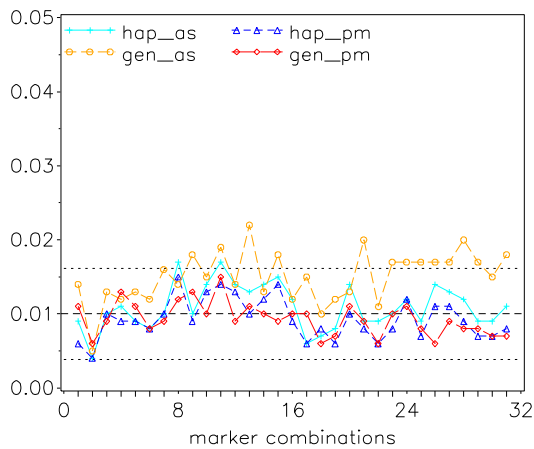
Type I error of FBAT, alpha=0.01  
200 AAU Families, NMod2 H0\_(A)



Type I error of FBAT, alpha=0.01  
200 AAU Families, NMod2 H0\_(L)

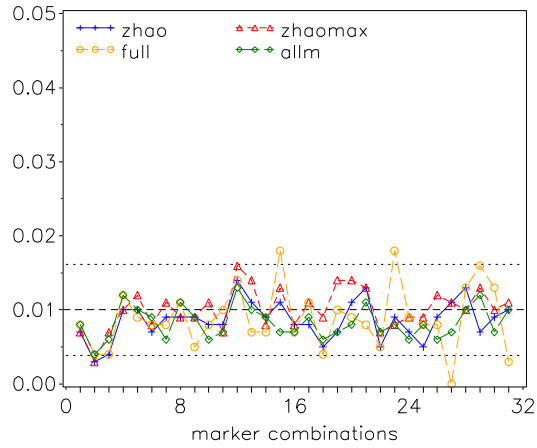


Hotel\_Fam: Type I error, alpha=0.01  
200 AAU families, NMod 2

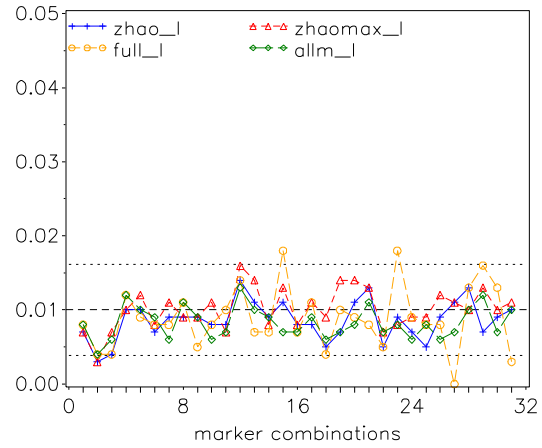


### NMod 3

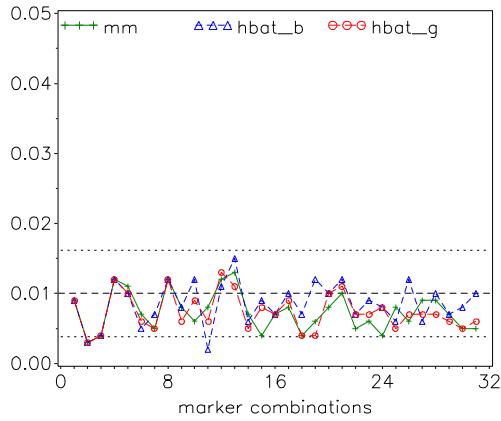
Type I error of a) Famhap b) Unphased  
200 Trios, Nmod3, H0\_(A)



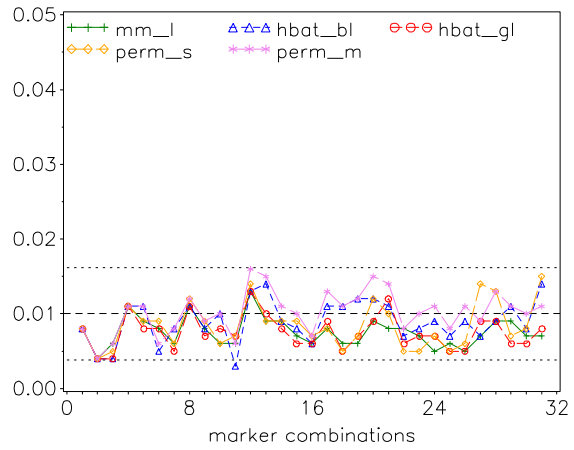
Type I error of a) Famhap b) Unphased  
200 Trios, NMod3 H0\_(L)



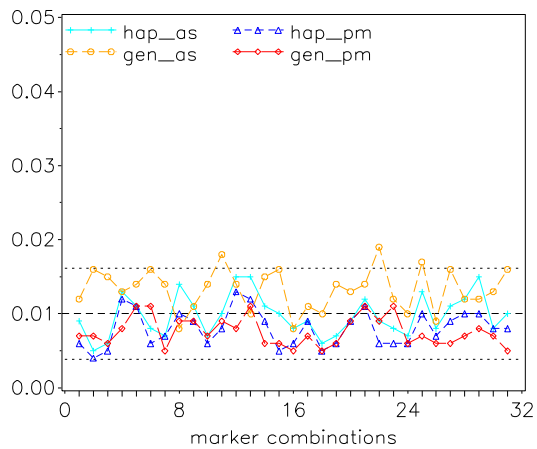
Type I error of FBAT, alpha=0.01  
200 Trios, NMod3 H0\_(A)



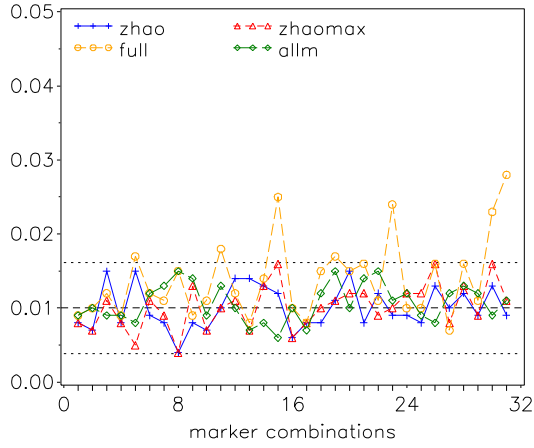
Type I error of FBAT, alpha=0.01  
200 Trios, NMod3 H0\_(L)



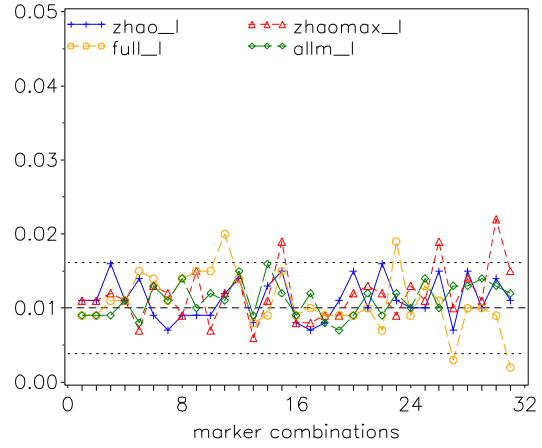
Hotel\_Fam: Type I error, alpha=0.01  
200 trios, NMod 3



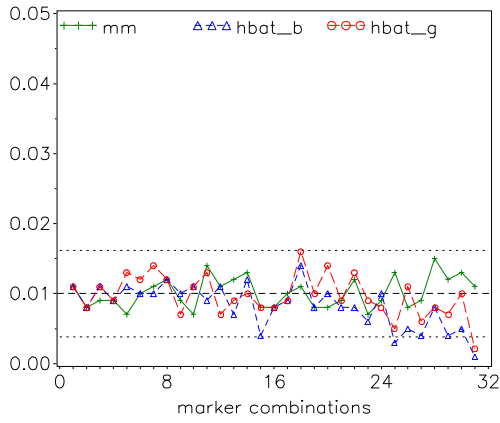
Type I error of a) Famhap b) Unphased  
200 AA Families, NMod3, H0\_(A)



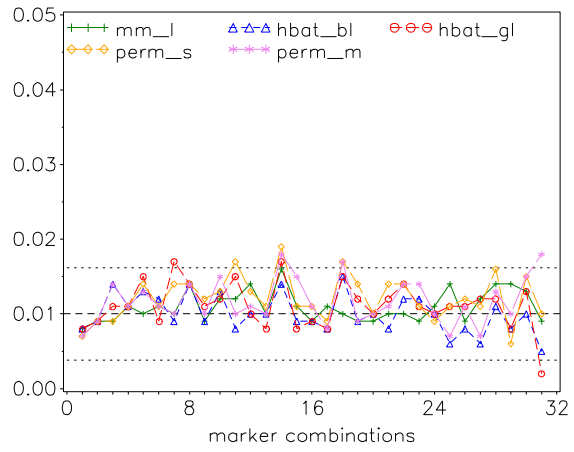
Type I error of a) Famhap b) Unphased  
200 AA-Families, NMod3 H0\_(L)



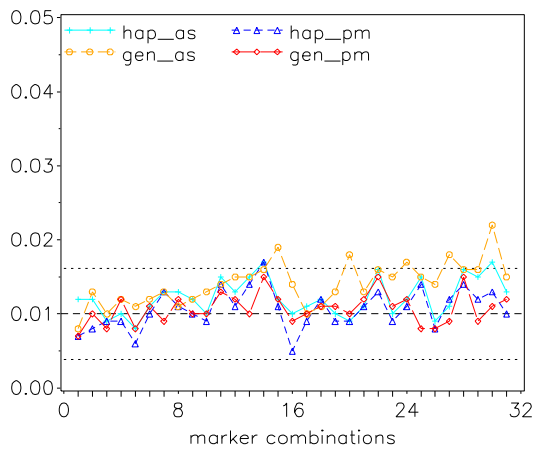
Type I error of FBAT, alpha=0.01  
200 AA Families, NMod3 H0\_(A)



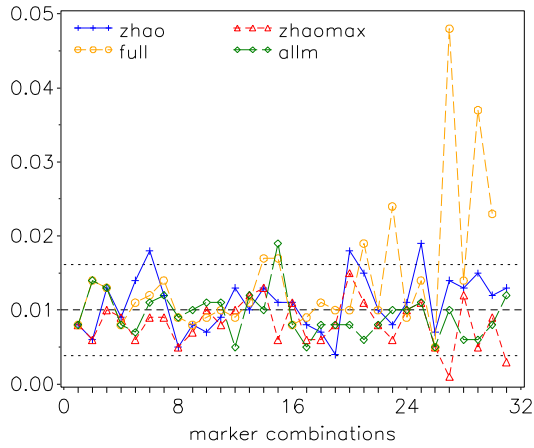
Type I error of FBAT, alpha=0.01  
200 AA Families, NMod3 H0\_(L)



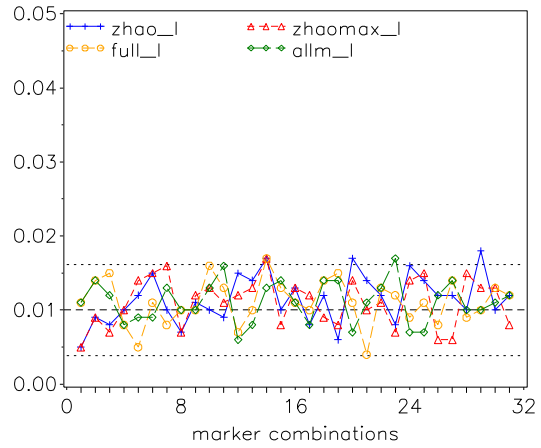
Hotel\_Fam: Type I error, alpha=0.01  
200 AA families, NMod 3



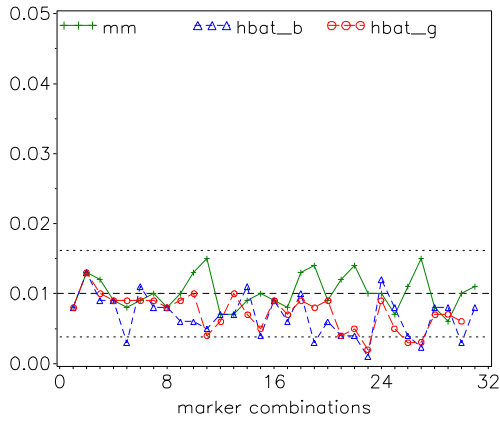
Type I error of a) Famhap b) Unphased  
200 AAU Families, NMod3, H0\_(A)



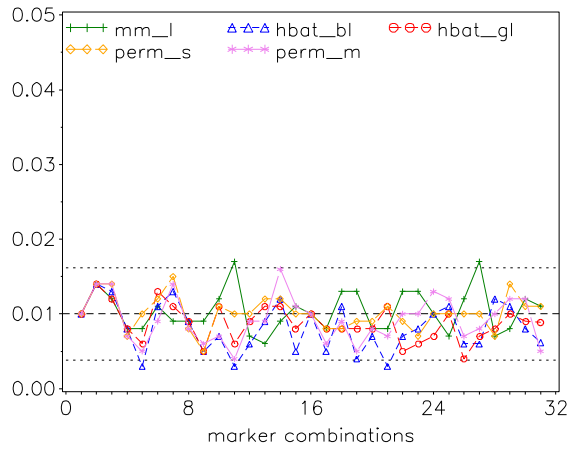
Type I error of a) Famhap b) Unphased  
200 AAU Families, NMod3 H0\_(L)



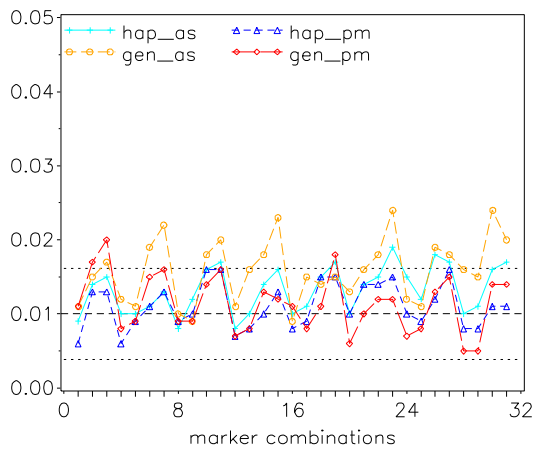
Type I error of FBAT, alpha=0.01  
200 AAU Families, NMod3 H0\_(A)



Type I error of FBAT, alpha=0.01  
200 AAU Families, NMod3 H0\_(L)



Hotel\_Fam: Type I error, alpha=0.01  
200 AAU families, NMod 3

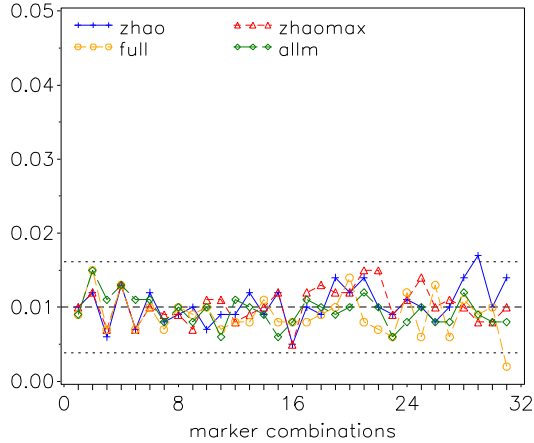


# Fehler 1. Art: Diagramme für alle Markerkombinationen

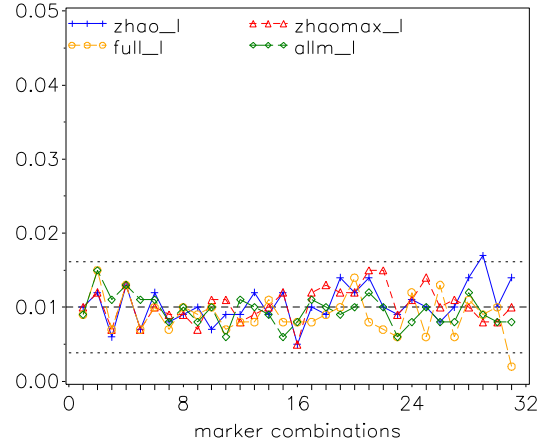
Testniveau  $\alpha = 0.01$ , 1000 Familien

NMod 1

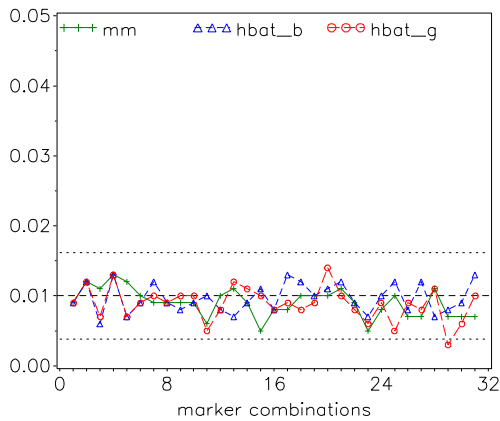
Type I error of a) Famhap b) Unphased  
1000 Trios, Nmod1, H0\_(A)



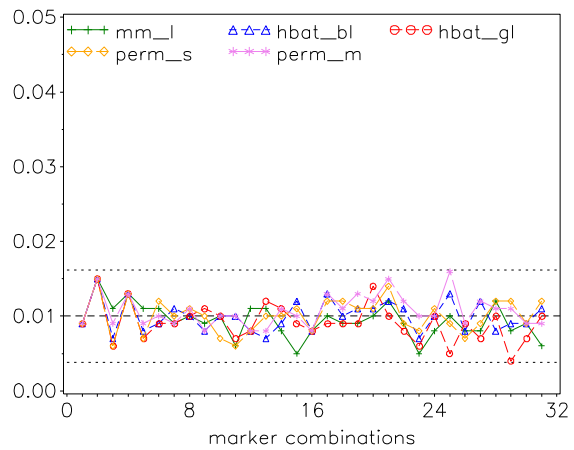
Type I error of a) Famhap b) Unphased  
1000 Trios, NMod1 H0\_(L)



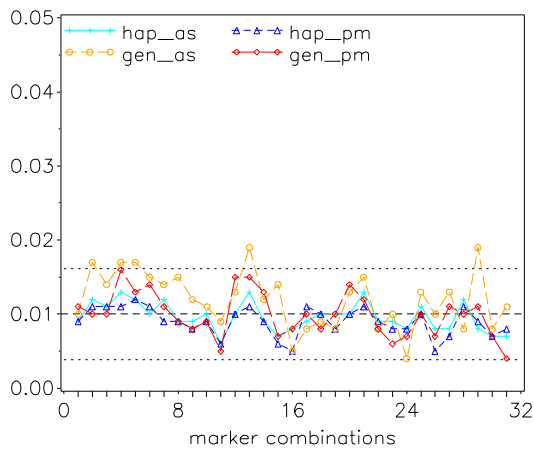
Type I error of FBAT, alpha=0.01  
1000 Trios, NMod1 H0\_(A)



Type I error of FBAT, alpha=0.01  
1000 Trios, NMod1 H0\_(L)

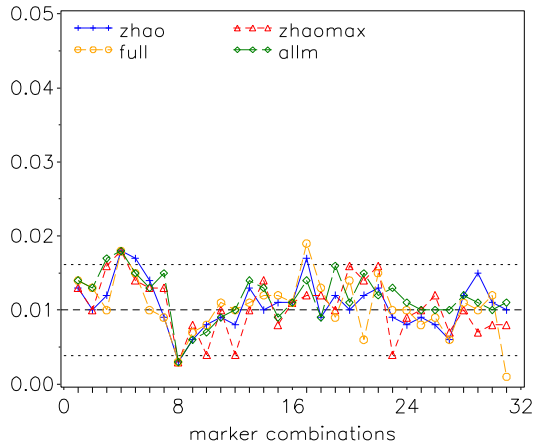


Hotel\_Fam: Type I error, alpha=0.01  
1000 trios, NMod 1

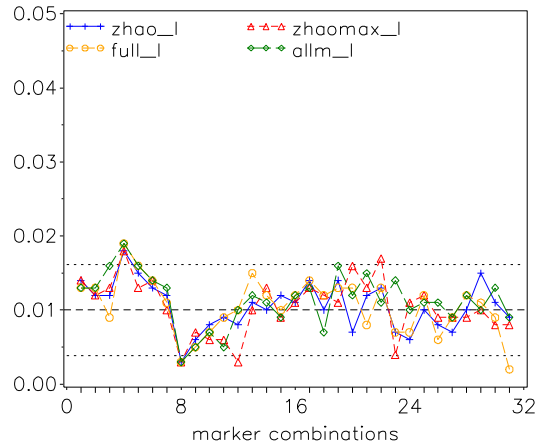




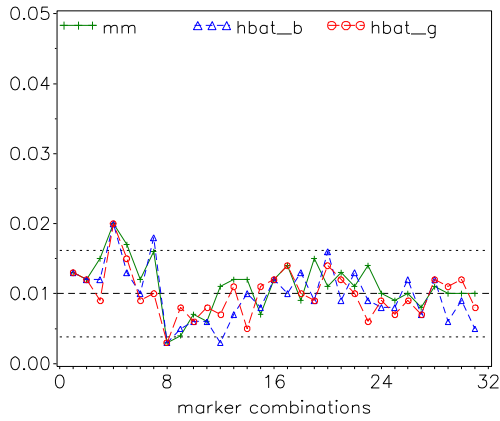
Type I error of a) Famhap b) Unphased  
1000 AA Families, NMod1, H0\_(A)



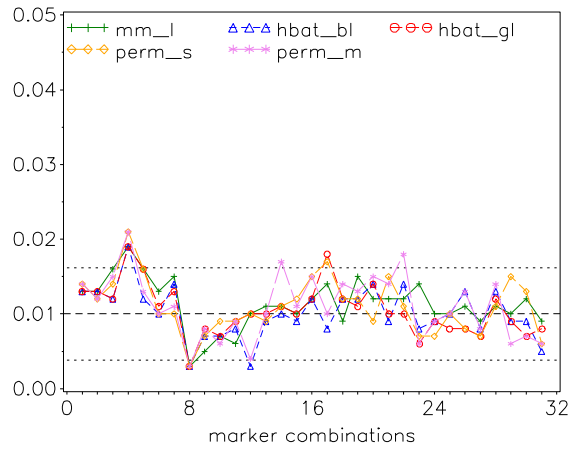
Type I error of a) Famhap b) Unphased  
1000 AA-Families, NMod1 H0\_(L)



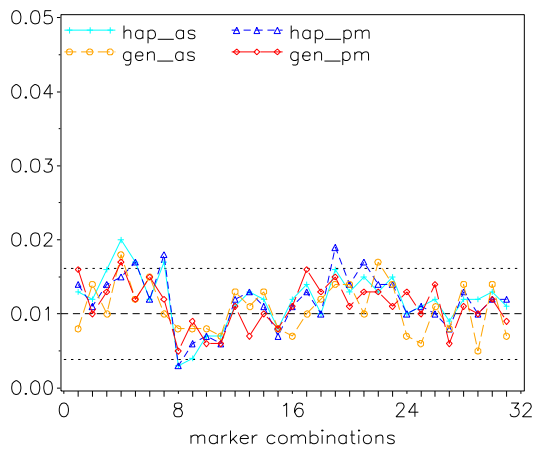
Type I error of FBAT, alpha=0.01  
1000 AA Families, NMod1 H0\_(A)



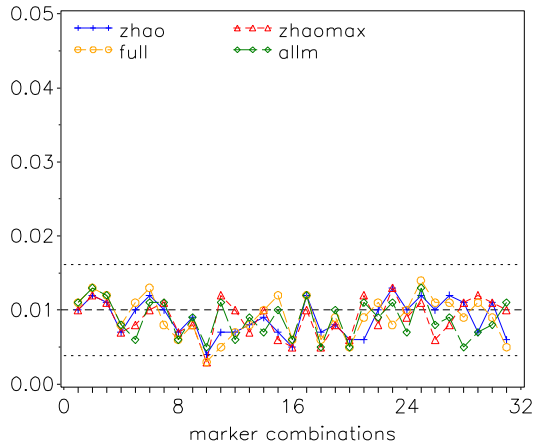
Type I error of FBAT, alpha=0.01  
1000 AA Families, NMod1 H0\_(L)



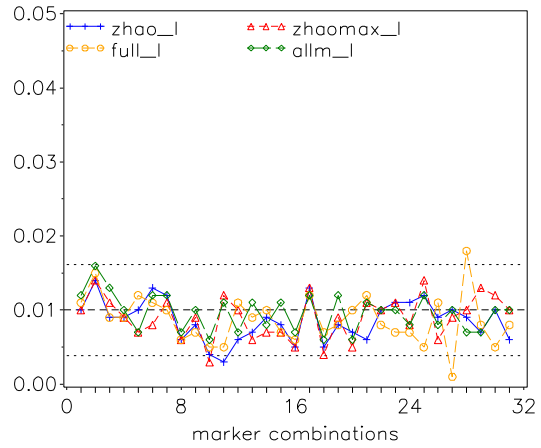
Hotel\_Fam: Type I error, alpha=0.01  
1000 AA families, NMod 1



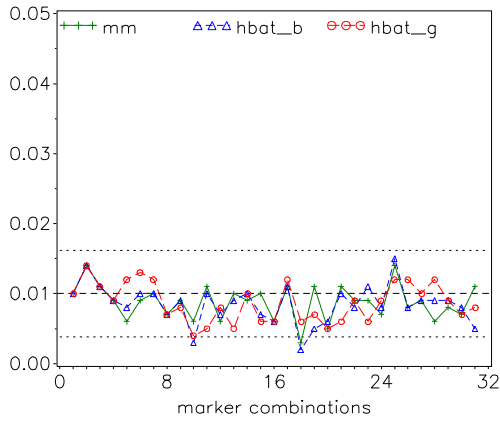
Type I error of a) Famhap b) Unphased  
1000 AAU Families, NMod1, H0\_(A)



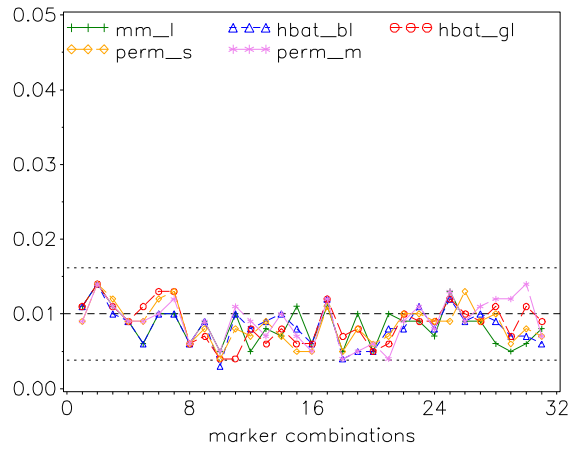
Type I error of a) Famhap b) Unphased  
1000 AAU Families, NMod1 H0\_(L)



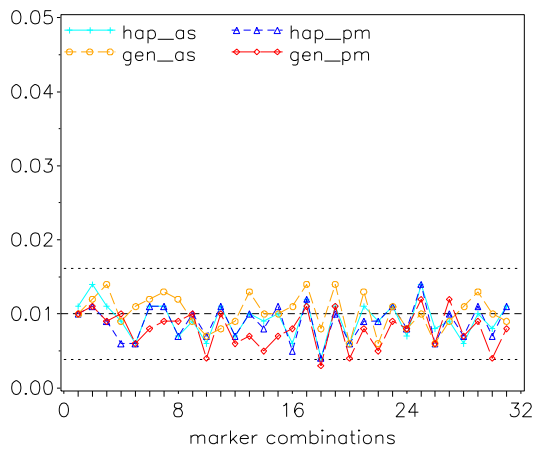
Type I error of FBAT, alpha=0.01  
1000 AAU Families, NMod1 H0\_(A)



Type I error of FBAT, alpha=0.01  
1000 AAU Families, NMod1 H0\_(L)

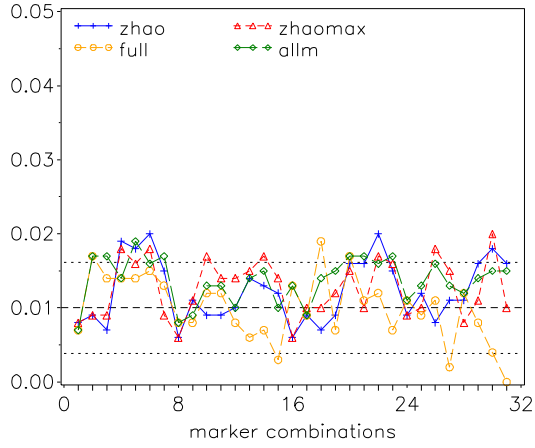


Hotel\_Fam: Type I error, alpha=0.01  
1000 AAU families, NMod 1

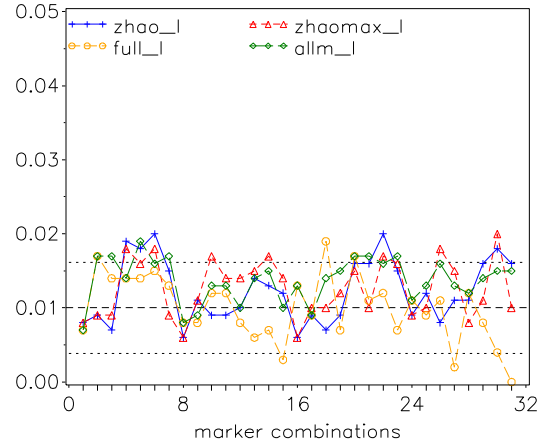


## NMod 2

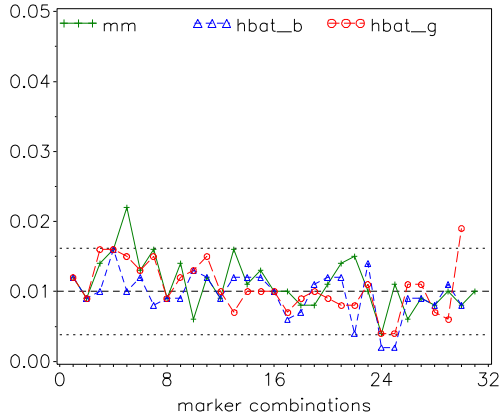
Type I error of a) Famhap b) Unphased  
1000 Trios, NMod2, H0\_(A)



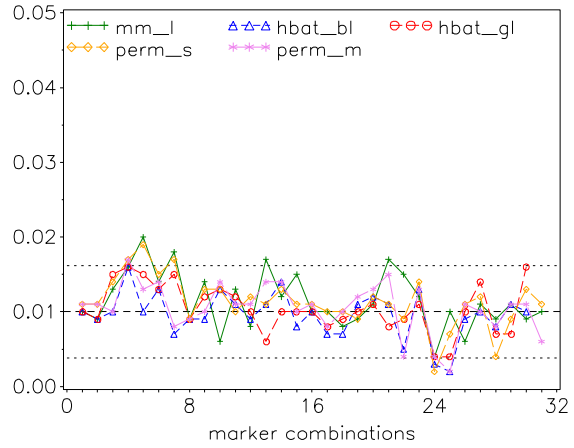
Type I error of a) Famhap b) Unphased  
1000 Trios, NMod2 H0\_(L)



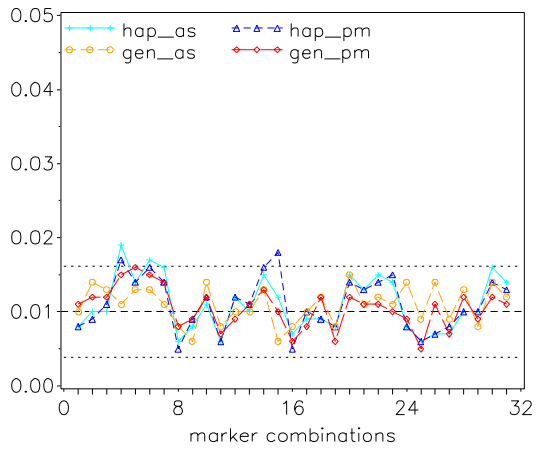
Type I error of FBAT, alpha=0.01  
1000 Trios, NMod2 H0\_(A)



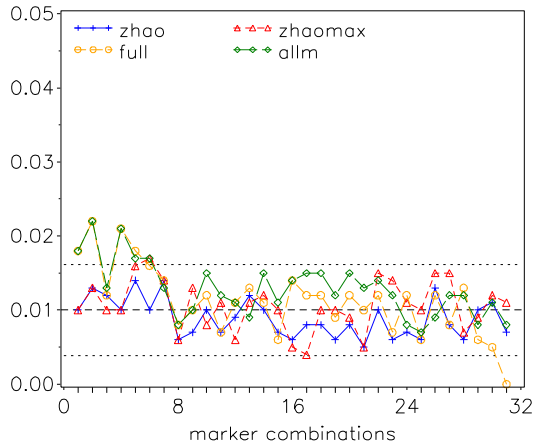
Type I error of FBAT, alpha=0.01  
1000 Trios, NMod2 H0\_(L)



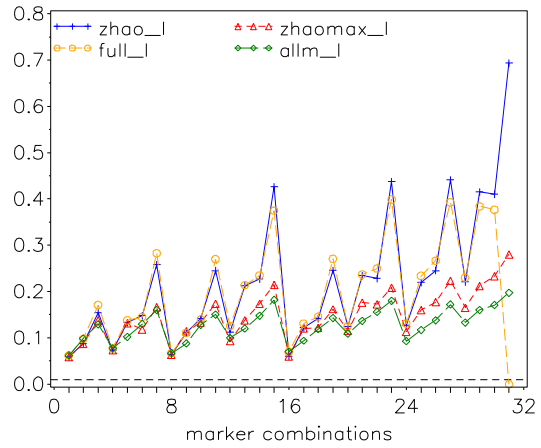
Hotel\_Fam: Type I error, alpha=0.01  
1000 trios, NMod 2



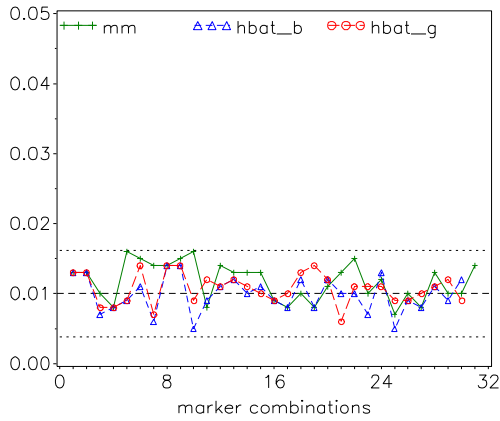
Type I error of a) Famhap b) Unphased  
1000 AA Families, NMod2, H0\_(A)



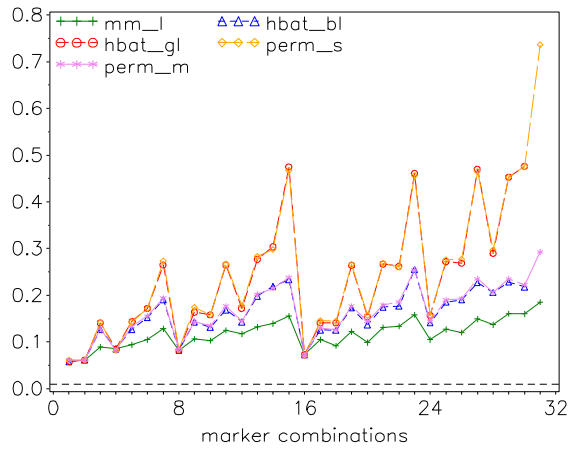
Type I error of a) Famhap b) Unphased  
1000 AA-Families, NMod2 H0\_(L)



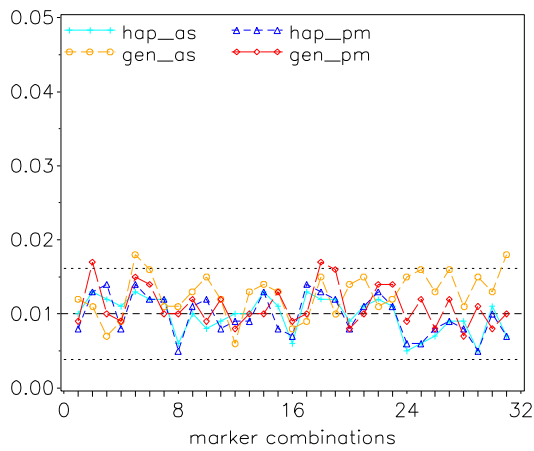
Type I error of FBAT, alpha=0.01  
1000 AA Families, NMod2 H0\_(A)



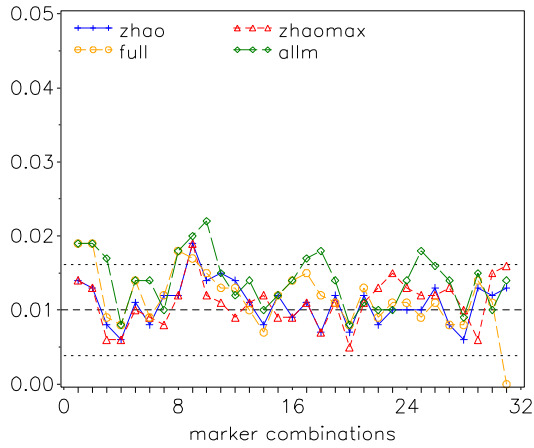
Type I error of FBAT, alpha=0.01  
1000 AA Families, NMod2 H0\_(L)



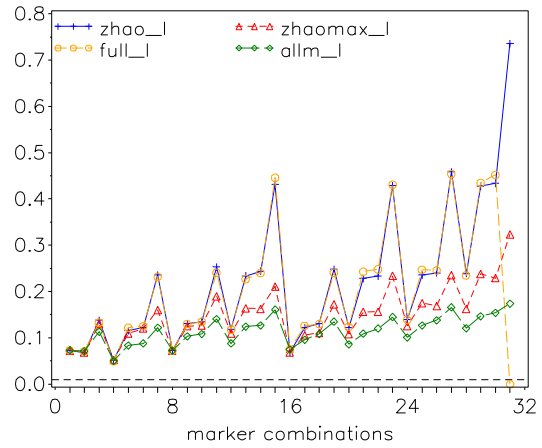
Hotel\_Fam: Type I error, alpha=0.01  
1000 AA families, NMod 2



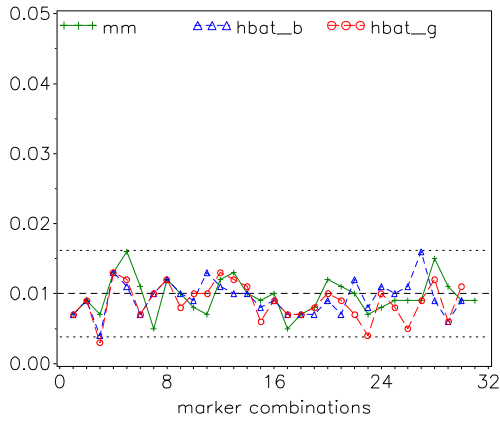
Type I error of a) Famhap b) Unphased  
1000 AAU Families, NMod2, H0\_(A)



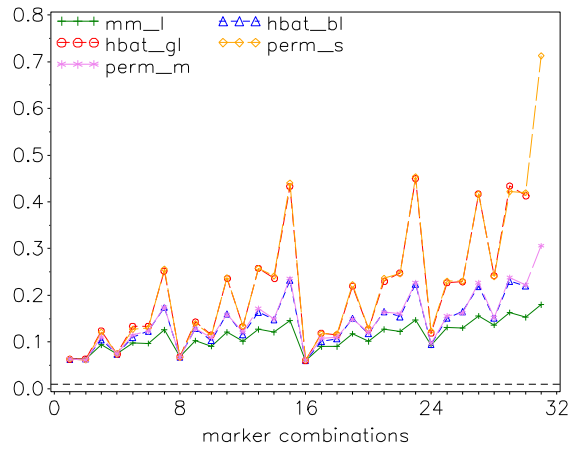
Type I error of a) Famhap b) Unphased  
1000 AAU Families, NMod2 H0\_(L)



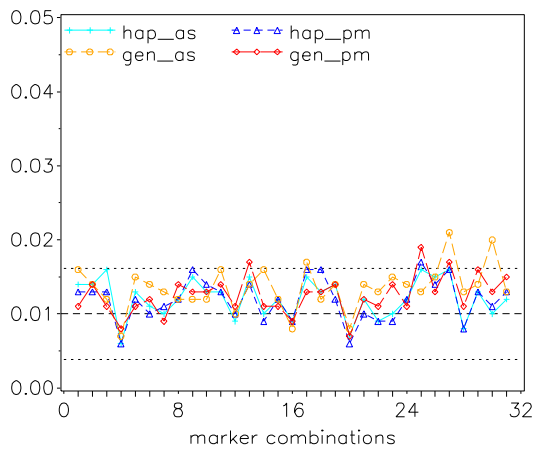
Type I error of FBAT, alpha=0.01  
1000 AAU Families, NMod2 H0\_(A)



Type I error of FBAT, alpha=0.01  
1000 AAU Families, NMod2 H0\_(L)

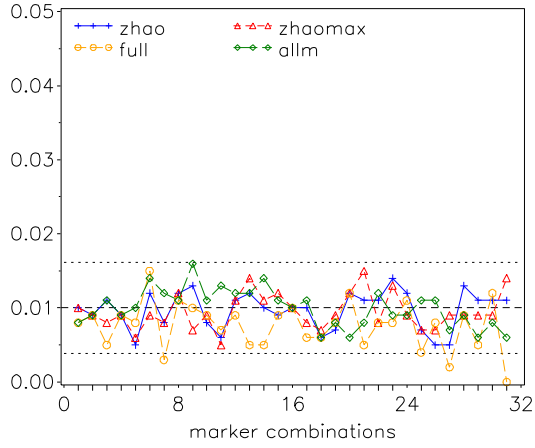


Hotel\_Fam: Type I error, alpha=0.01  
1000 AAU families, NMod 2

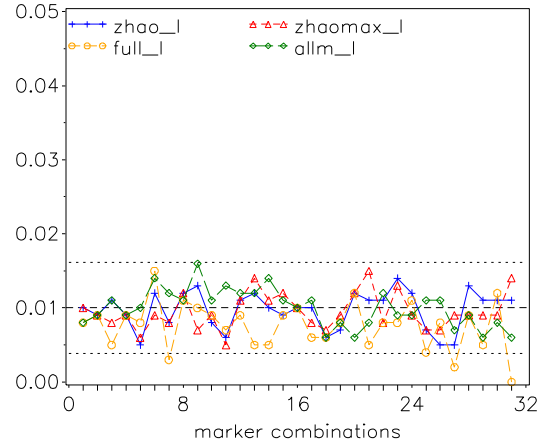


### NMod 3

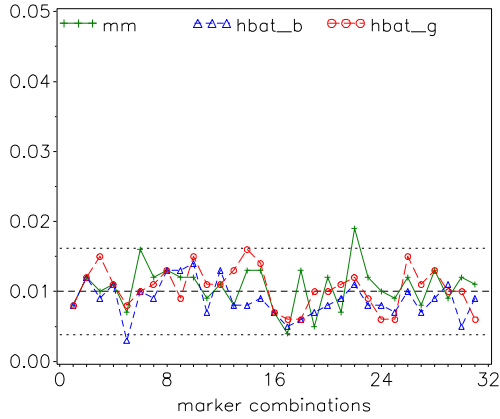
Type I error of a) Famhap b) Unphased  
1000 Trios, Nmod3, H0\_(A)



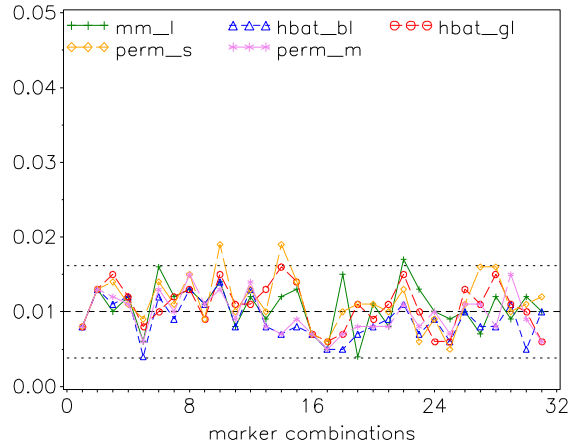
Type I error of a) Famhap b) Unphased  
1000 Trios, NMod3 H0\_(L)



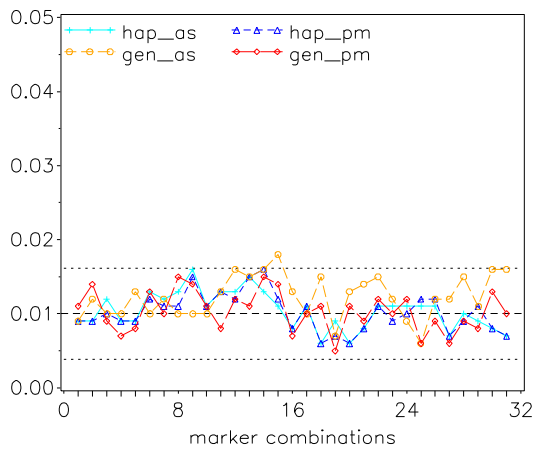
Type I error of FBAT, alpha=0.01  
1000 Trios, NMod3 H0\_(A)



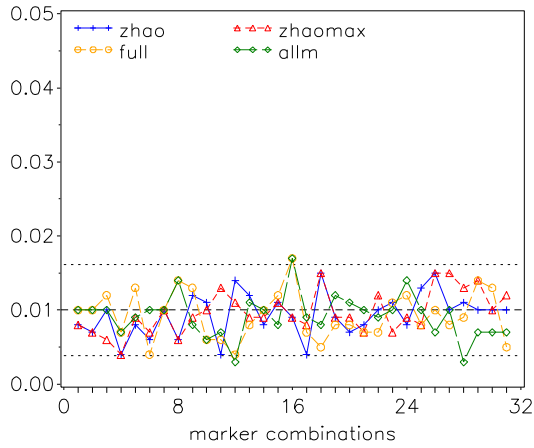
Type I error of FBAT, alpha=0.01  
1000 Trios, NMod3 H0\_(L)



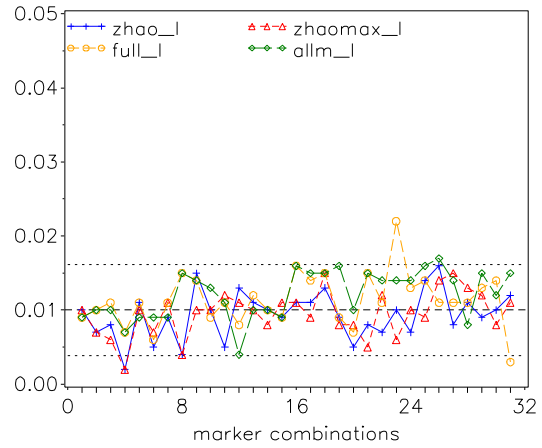
Hotel\_Fam: Type I error, alpha=0.01  
1000 trios, NMod 3



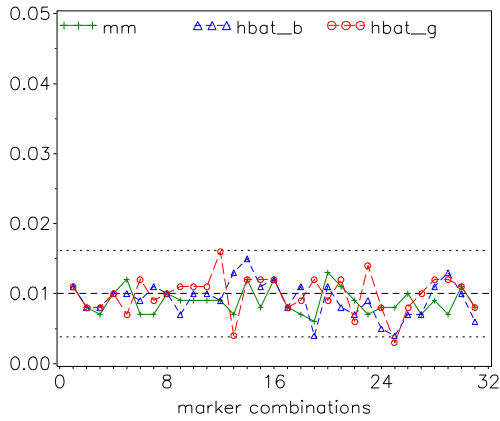
Type I error of a) Famhap b) Unphased  
1000 AA Families, NMod3, H0\_(A)



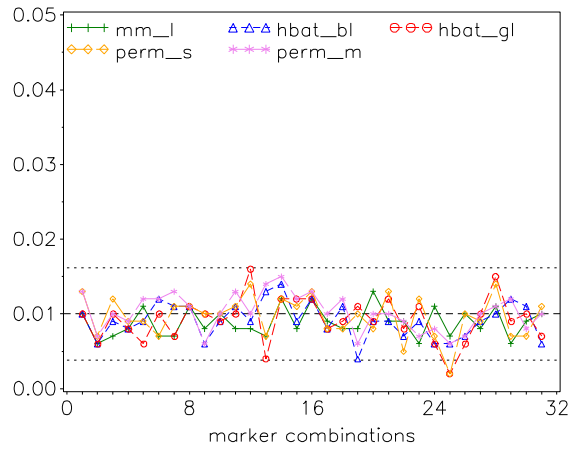
Type I error of a) Famhap b) Unphased  
1000 AA-Families, NMod3 H0\_(L)



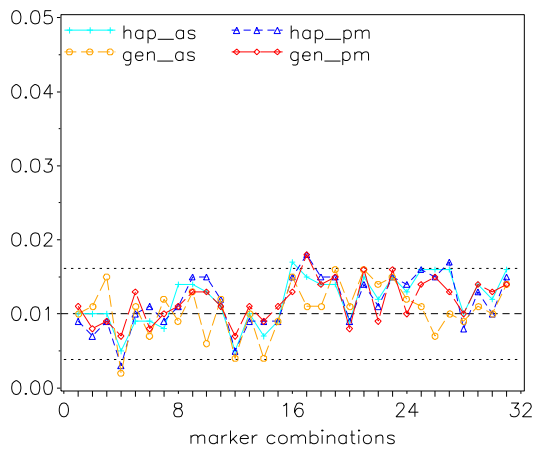
Type I error of FBAT, alpha=0.01  
1000 AA Families, NMod3 H0\_(A)



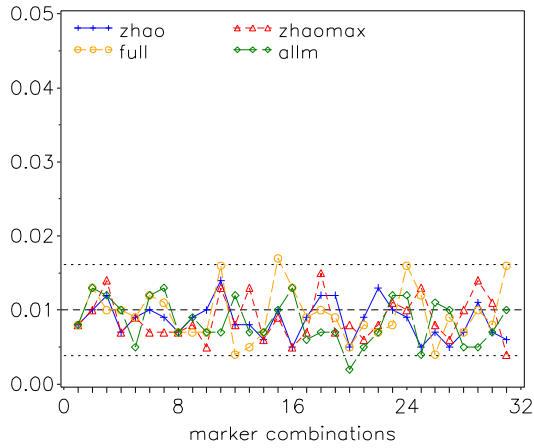
Type I error of FBAT, alpha=0.01  
1000 AA Families, NMod3 H0\_(L)



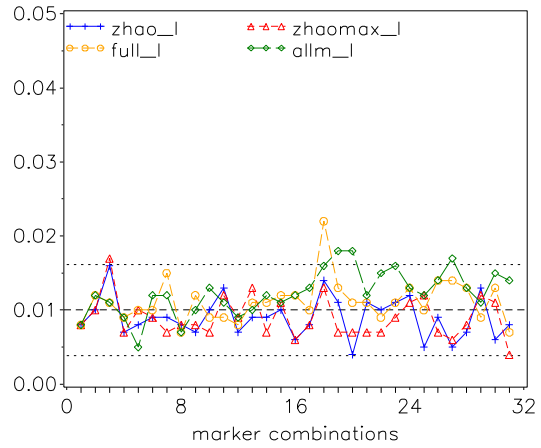
Hotel\_Fam: Type I error, alpha=0.01  
1000 AA families, NMod 3



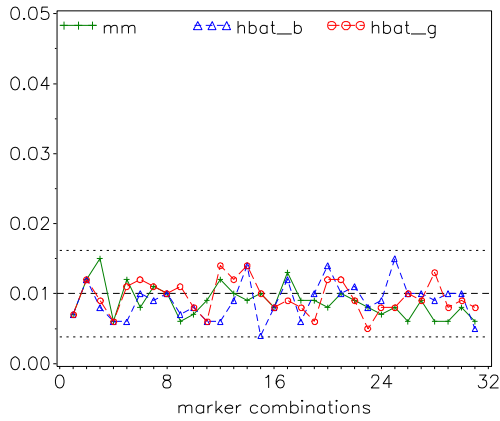
Type I error of a) Famhap b) Unphased  
1000 AAU Families, NMod3, H0\_(A)



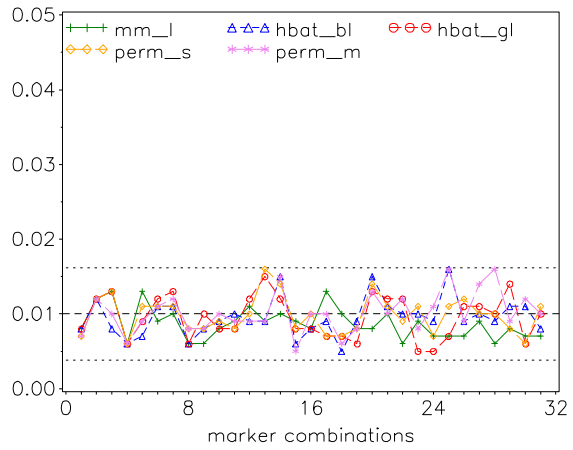
Type I error of a) Famhap b) Unphased  
1000 AAU-Families, NMod3 H0\_(L)



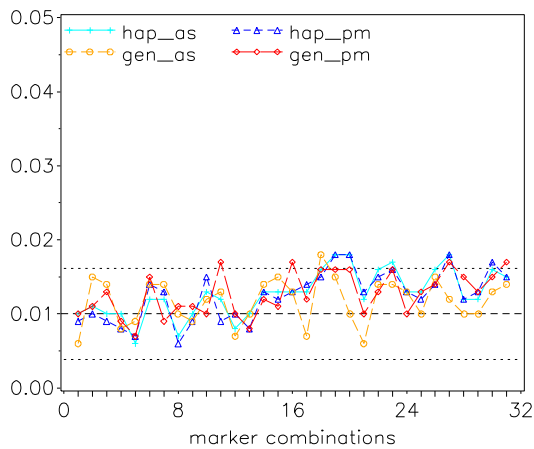
Type I error of FBAT, alpha=0.01  
1000 AAU Families, NMod3 H0\_(A)



Type I error of FBAT, alpha=0.01  
1000 AAU Families, NMod3 H0\_(L)



Hotel\_Fam: Type I error, alpha=0.01  
1000 AAU families, NMod 3



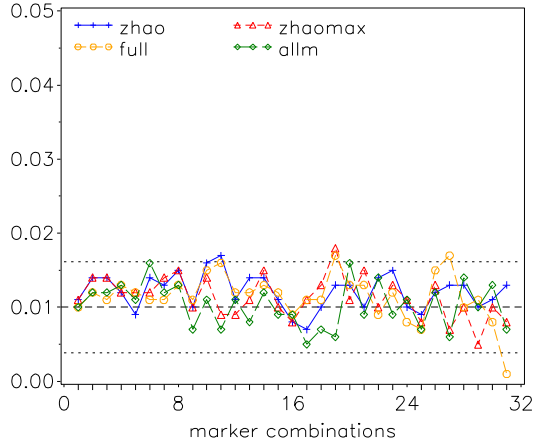


# Fehler 1. Art: Diagramme für alle Markerkombinationen

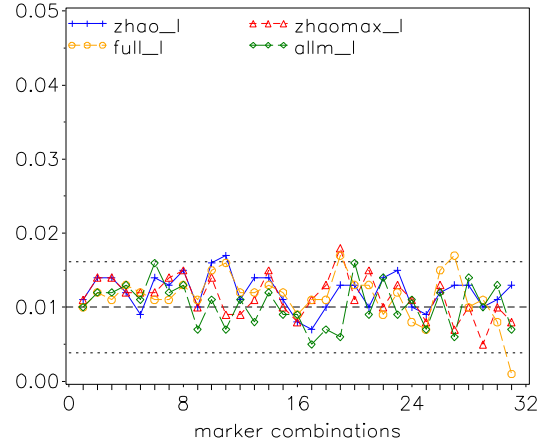
Testniveau  $\alpha = 0.01$ , 2500 Familien

NMod 1

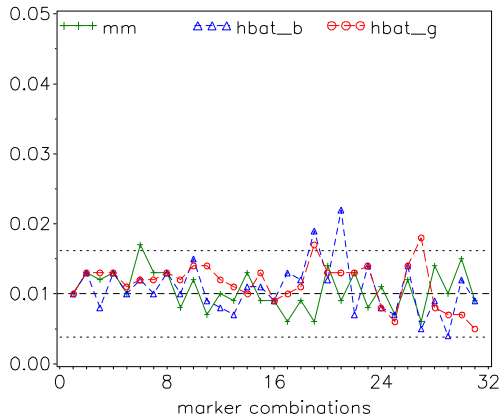
Type I error of a) Famhap b) Unphased  
2500 Trios, Nmod1, H0\_(A)



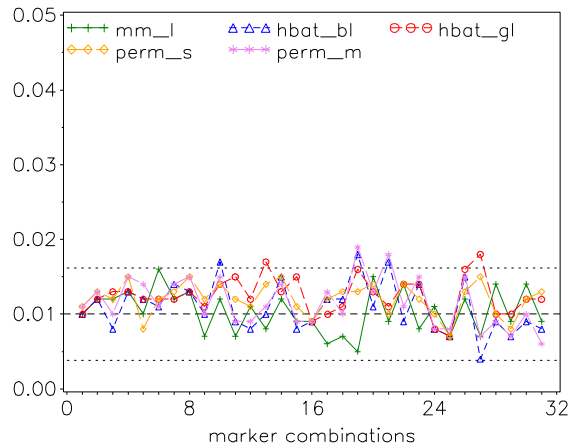
Type I error of a) Famhap b) Unphased  
2500 Trios, NMod1 H0\_(L)



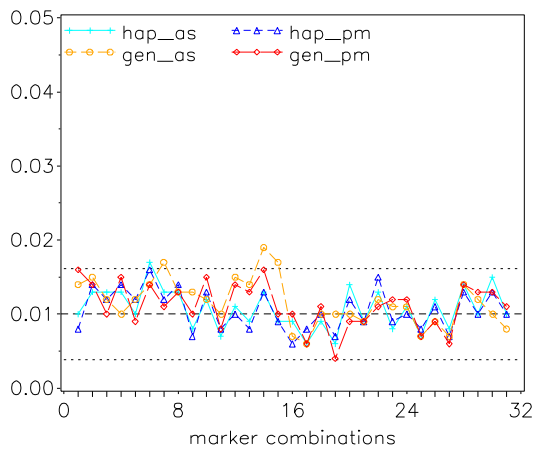
Type I error of FBAT, alpha=0.01  
2500 Trios, NMod1 H0\_(A)



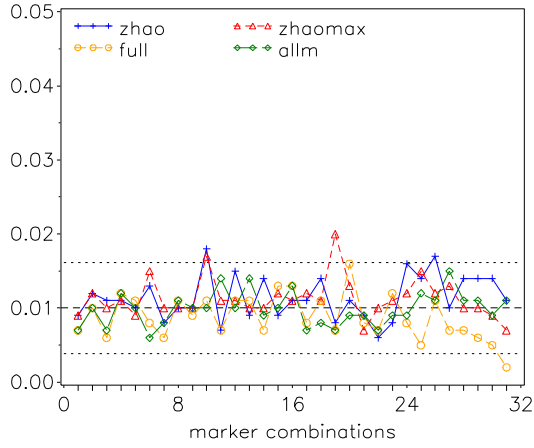
Type I error of FBAT, alpha=0.01  
2500 Trios, NMod1 H0\_(L)



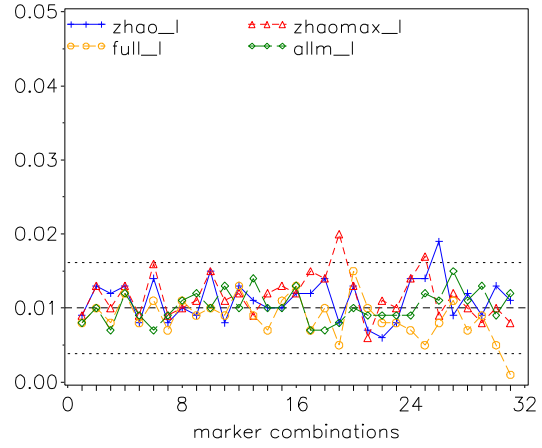
Hotel\_Fam: Type I error, alpha=0.01  
2500 trios, NMod 1



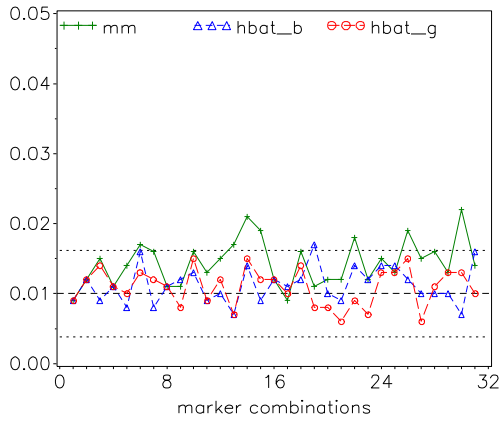
Type I error of a) Famhap b) Unphased  
2500 AA Families, Nmod1, H0\_(A)



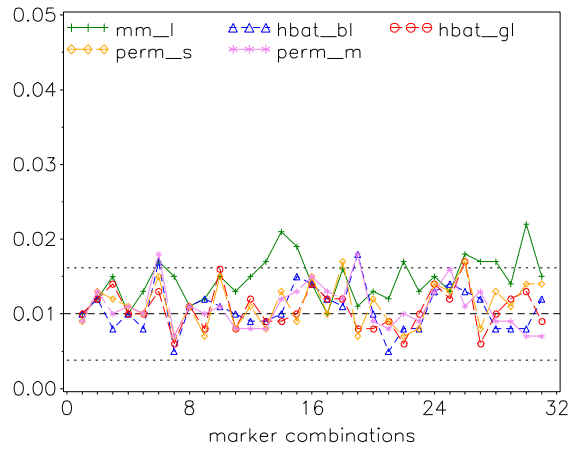
Type I error of a) Famhap b) Unphased  
2500 AA-Families, NMod1 H0\_(L)



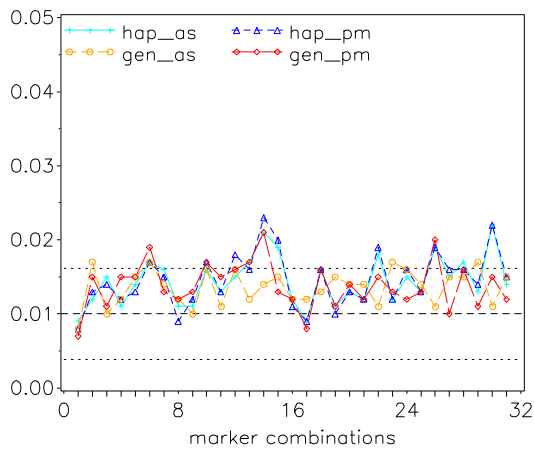
Type I error of FBAT, alpha=0.01  
2500 AA Families, NMod1 H0\_(A)



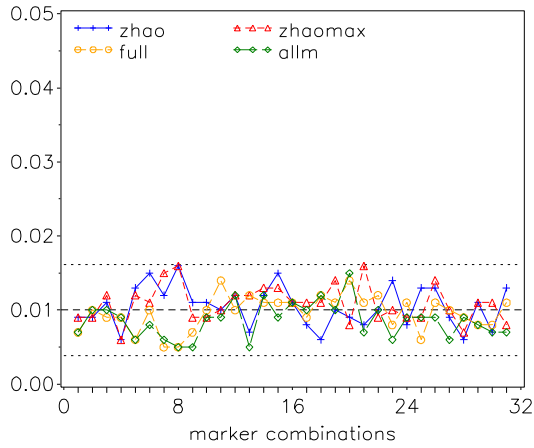
Type I error of FBAT, alpha=0.01  
2500 AA Families, NMod1 H0\_(L)



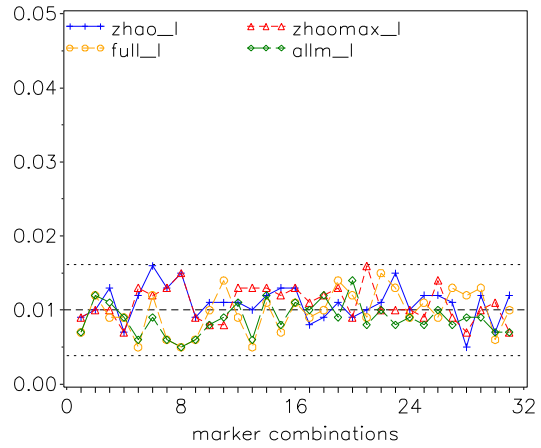
Hotel\_Fam: Type I error, alpha=0.01  
2500 AA families, NMod 1



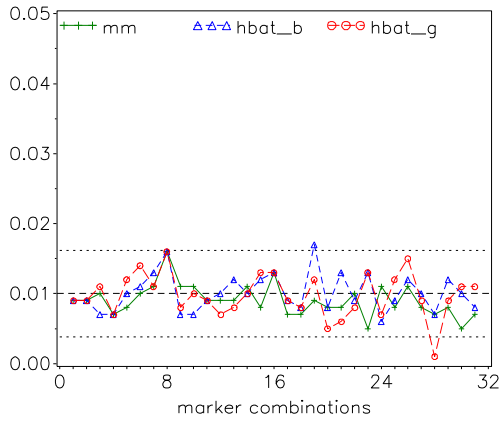
Type I error of a) Famhap b) Unphased  
2500 AAUU Families, Nmod1, H0\_(A)



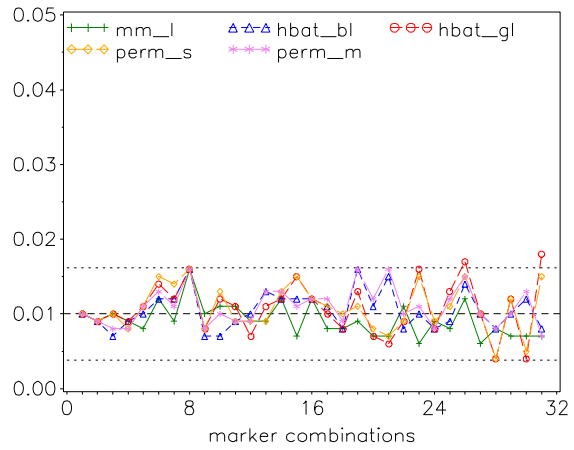
Type I error of a) Famhap b) Unphased  
2500 AAUU-Families, NMod1 H0\_(L)



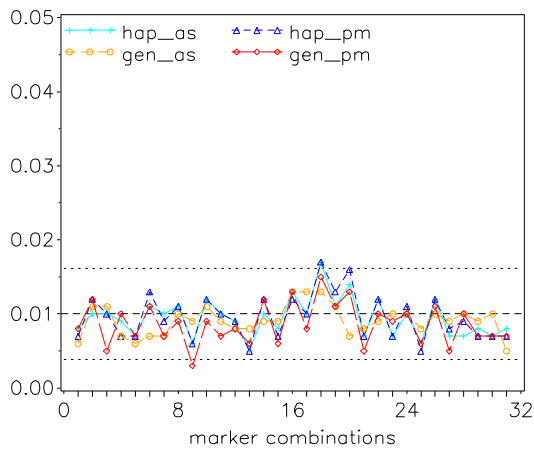
Type I error of FBAT, alpha=0.01  
2500 AAUU Families, NMod1 H0\_(A)



Type I error of FBAT, alpha=0.01  
2500 AAUU Families, NMod1 H0\_(L)

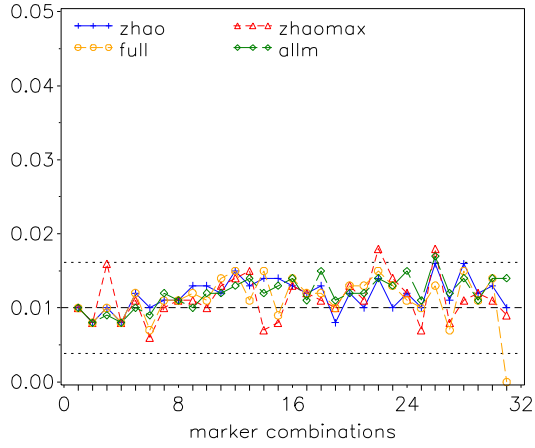


Hotel\_Fam: Type I error, alpha=0.01  
2500 AAUU families, NMod 1

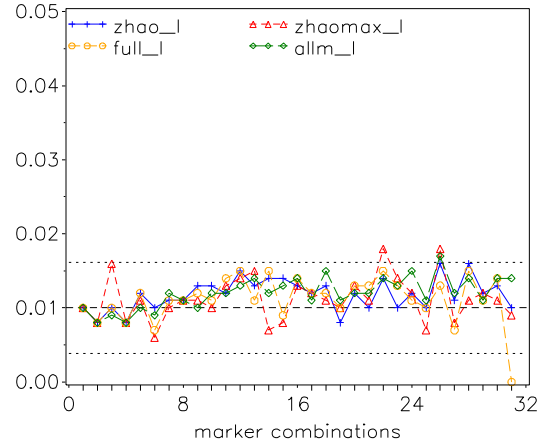


## NMod 2

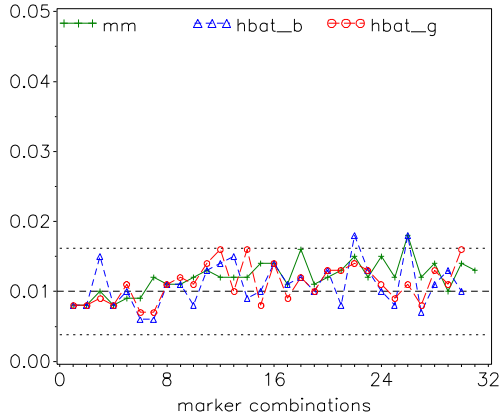
Type I error of a) Famhap b) Unphased  
2500 Trios, Nmod2, H0\_(A)



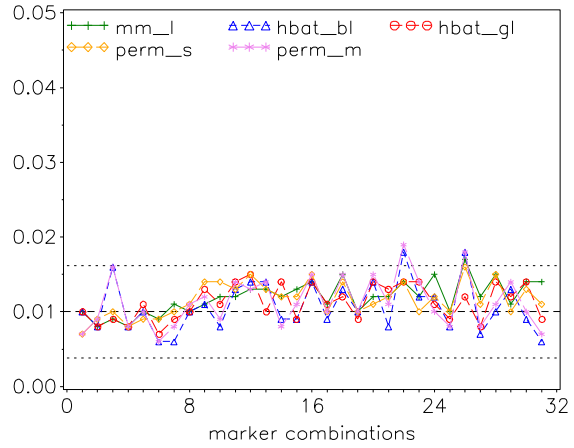
Type I error of a) Famhap b) Unphased  
2500 Trios, NMod2 H0\_(L)



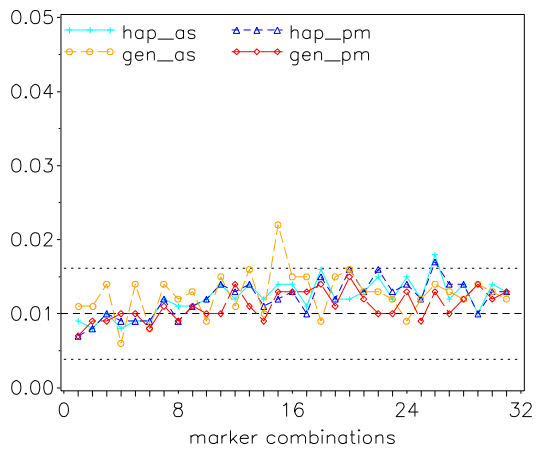
Type I error of FBAT, alpha=0.01  
2500 Trios, NMod2 H0\_(A)



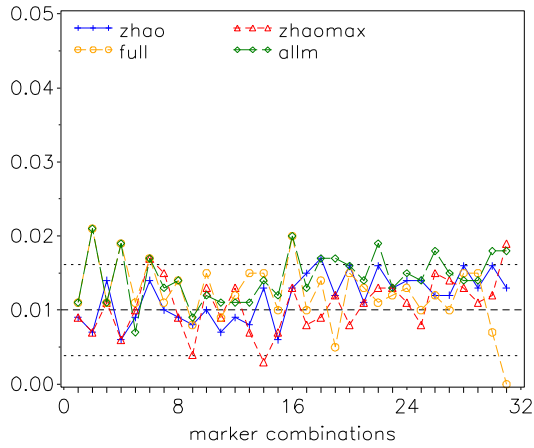
Type I error of FBAT, alpha=0.01  
2500 Trios, NMod2 H0\_(L)



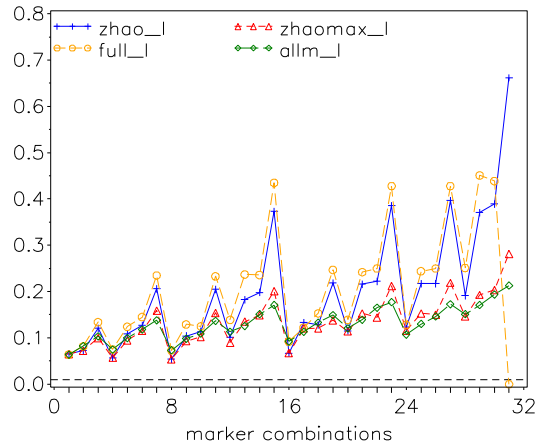
Hotel\_Fam: Type I error, alpha=0.01  
2500 trios, NMod 2



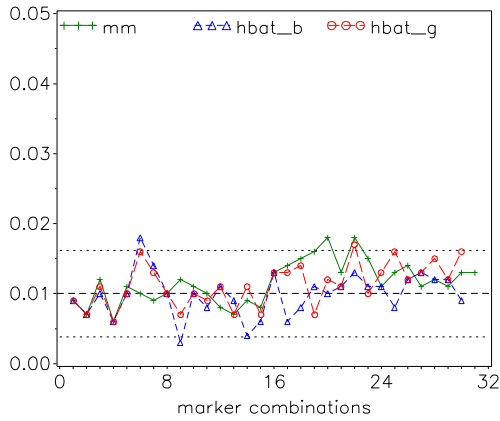
Type I error of a) Famhap b) Unphased  
2500 AA Families, Nmod2, H0\_(A)



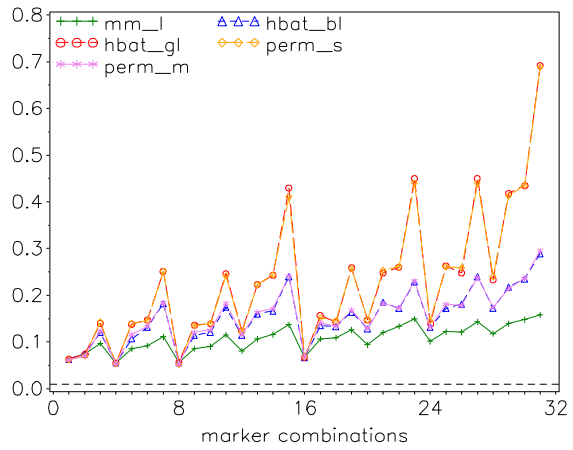
Type I error of a) Famhap b) Unphased  
2500 AA-Families, NMod2 H0\_(L)



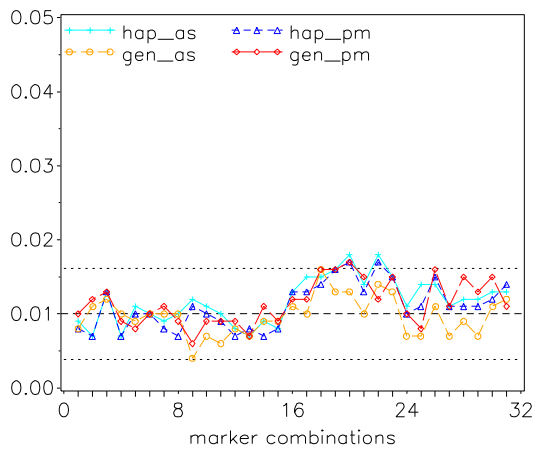
Type I error of FBAT, alpha=0.01  
2500 AA Families, NMod2 H0\_(A)



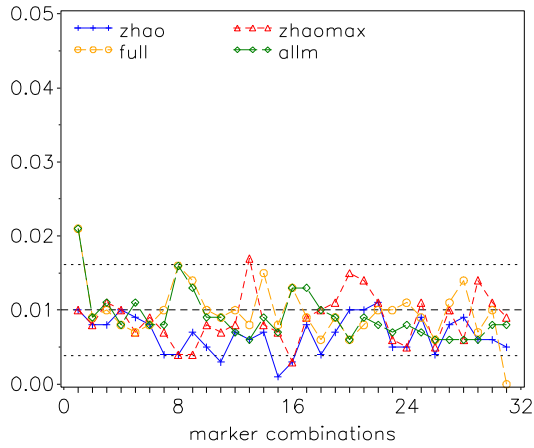
Type I error of FBAT, alpha=0.01  
2500 AA Families, NMod2 H0\_(L)



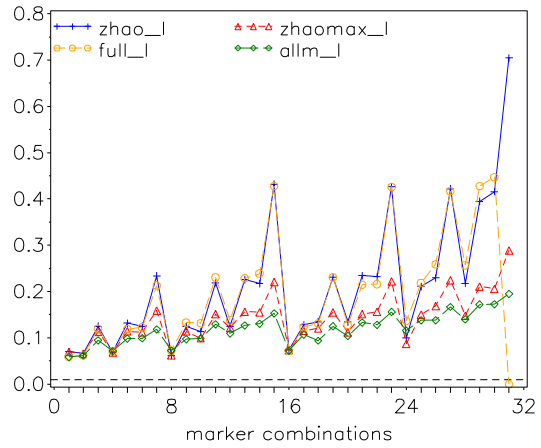
Hotel\_Fam: Type I error, alpha=0.01  
2500 AA families, NMod 2



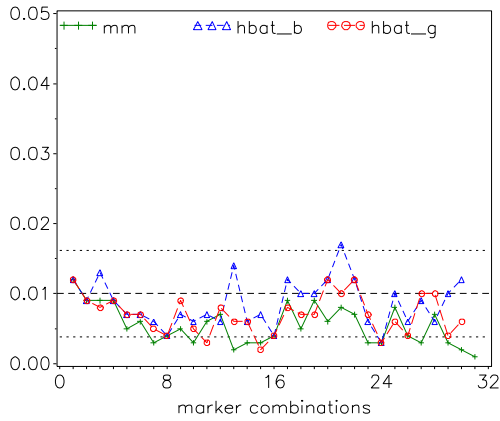
Type I error of a) Famhap b) Unphased  
2500 AAU Families, NMod2, H0\_(A)



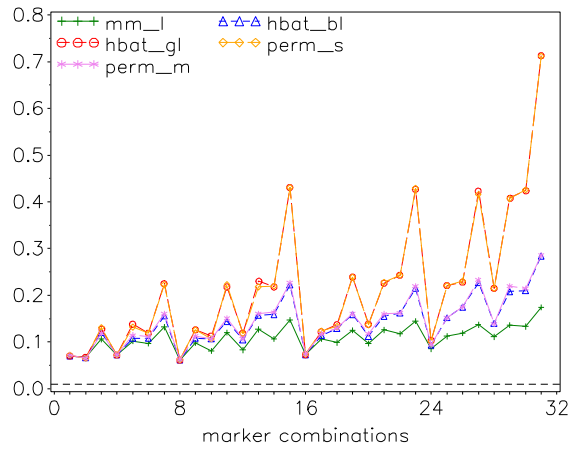
Type I error of a) Famhap b) Unphased  
2500 AAU Families, NMod2 H0\_(L)



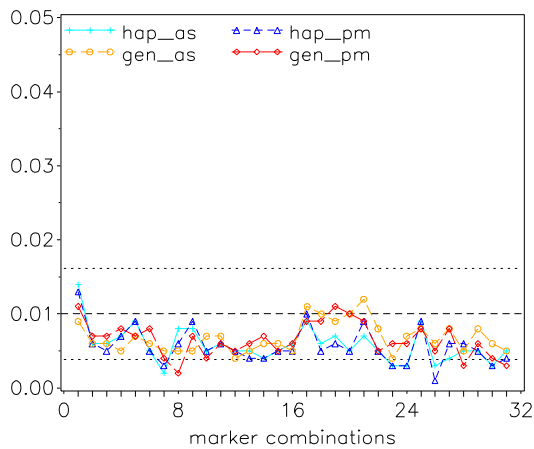
Type I error of FBAT, alpha=0.01  
2500 AAU Families, NMod2 H0\_(A)



Type I error of FBAT, alpha=0.01  
2500 AAU Families, NMod2 H0\_(L)

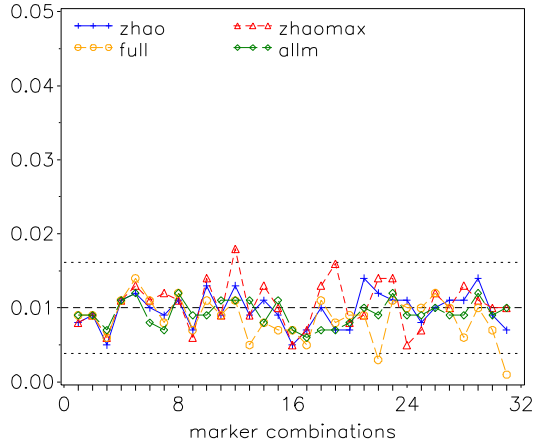


Hotel\_Fam: Type I error, alpha=0.01  
2500 AAU families, NMod 2

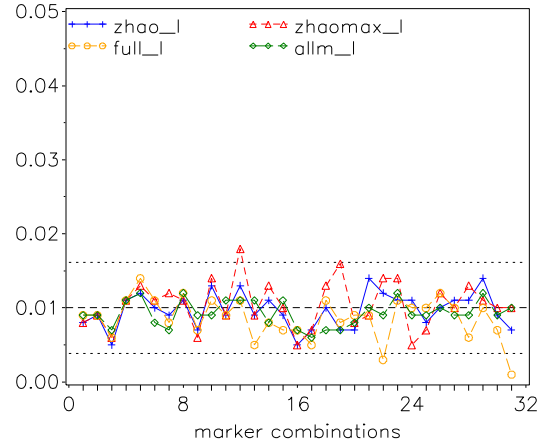


### NMod 3

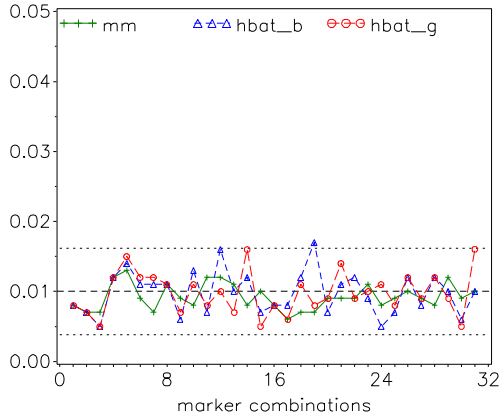
Type I error of a) Famhap b) Unphased  
2500 Trios, NMod3, H0\_(A)



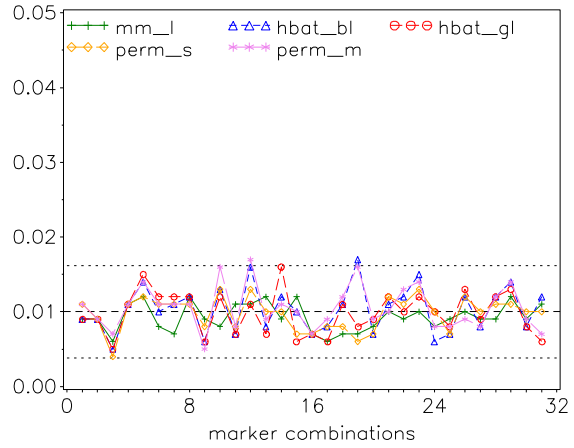
Type I error of a) Famhap b) Unphased  
2500 Trios, NMod3 H0\_(L)



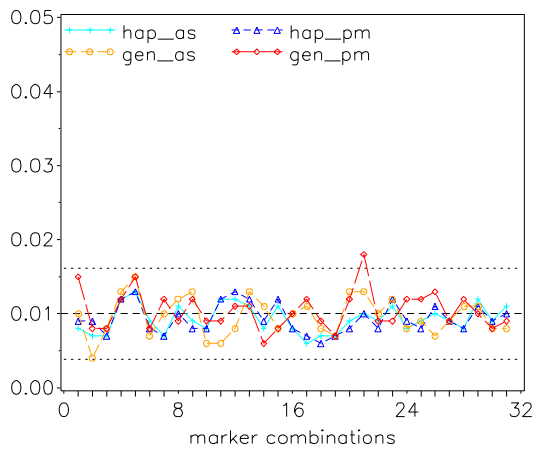
Type I error of FBAT, alpha=0.01  
2500 Trios, NMod3 H0\_(A)



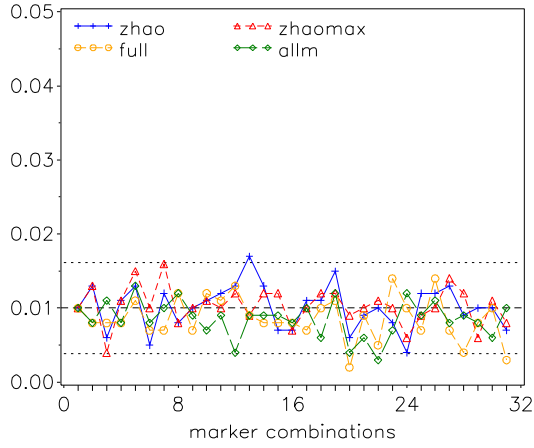
Type I error of FBAT, alpha=0.01  
2500 Trios, NMod3 H0\_(L)



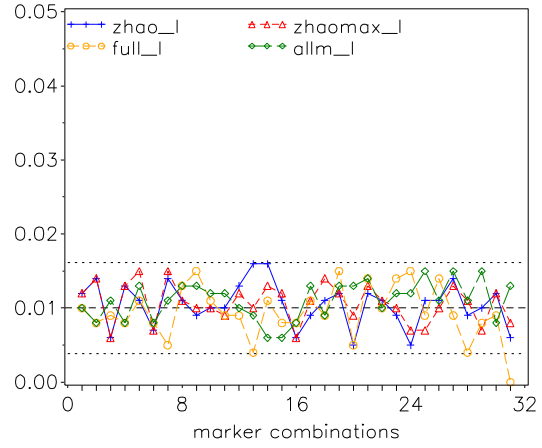
Hotel\_Fam: Type I error, alpha=0.01  
2500 trios, NMod 3



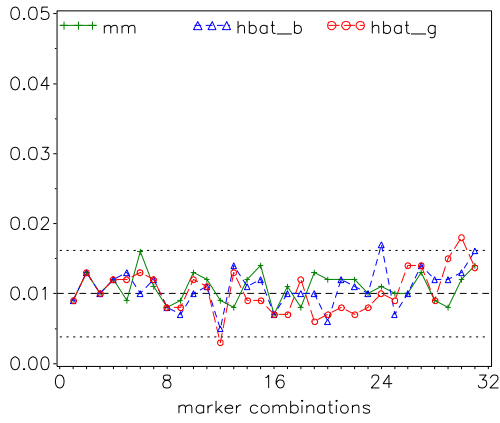
Type I error of a) Famhap b) Unphased  
2500 AA Families, Nmod3, H0\_(A)



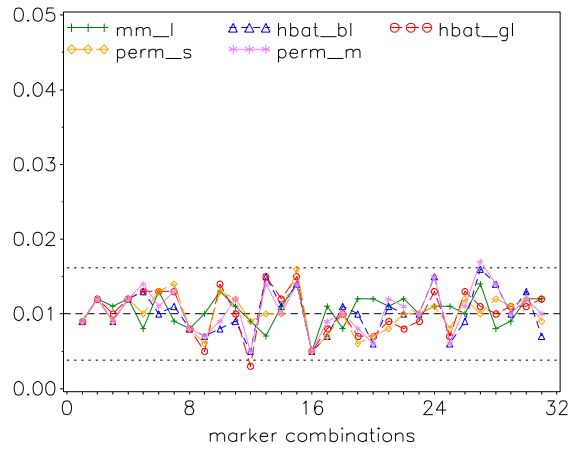
Type I error of a) Famhap b) Unphased  
2500 AA-Families, NMod3 H0\_(L)



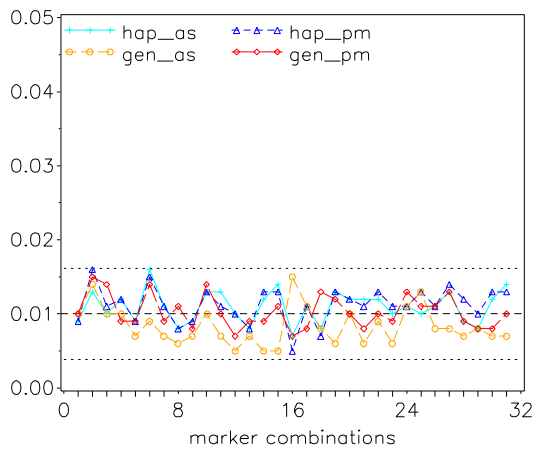
Type I error of FBAT, alpha=0.01  
2500 AA Families, NMod3 H0\_(A)



Type I error of FBAT, alpha=0.01  
2500 AA Families, NMod3 H0\_(L)

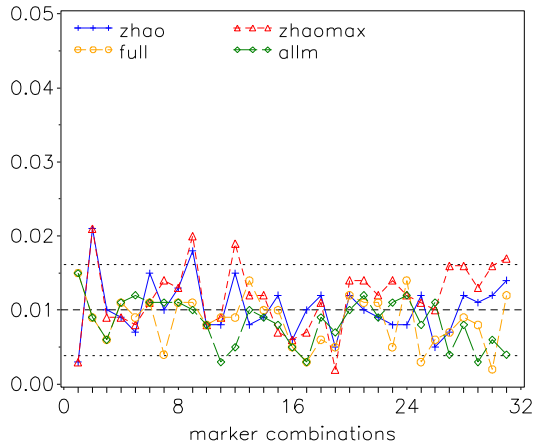


Hotel\_Fam: Type I error, alpha=0.01  
2500 AA families, NMod 3

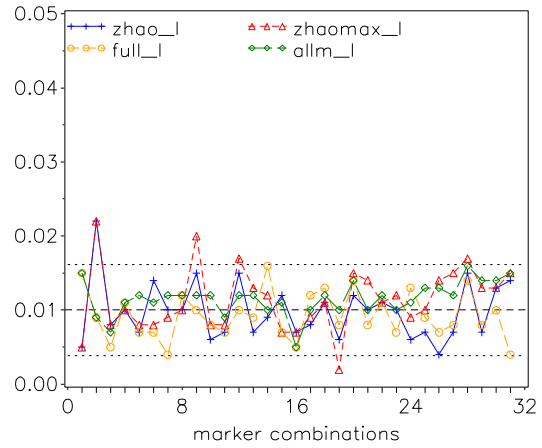




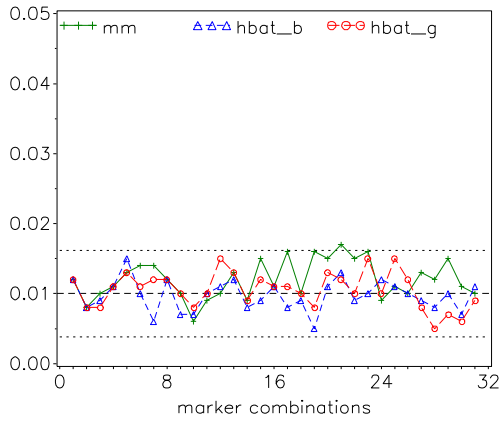
Type I error of a) Famhap b) Unphased  
2500 AAUU Families, Nmod3, H0\_(A)



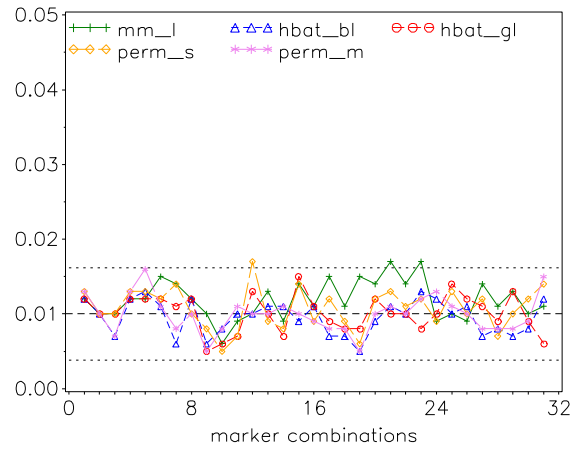
Type I error of a) Famhap b) Unphased  
2500 AAUU-Families, NMod3 H0\_(L)



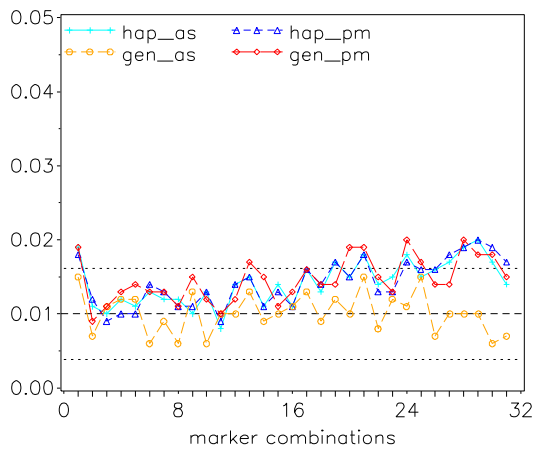
Type I error of FBAT, alpha=0.01  
2500 AAUU Families, NMod3 H0\_(A)



Type I error of FBAT, alpha=0.01  
2500 AAUU Families, NMod3 H0\_(L)



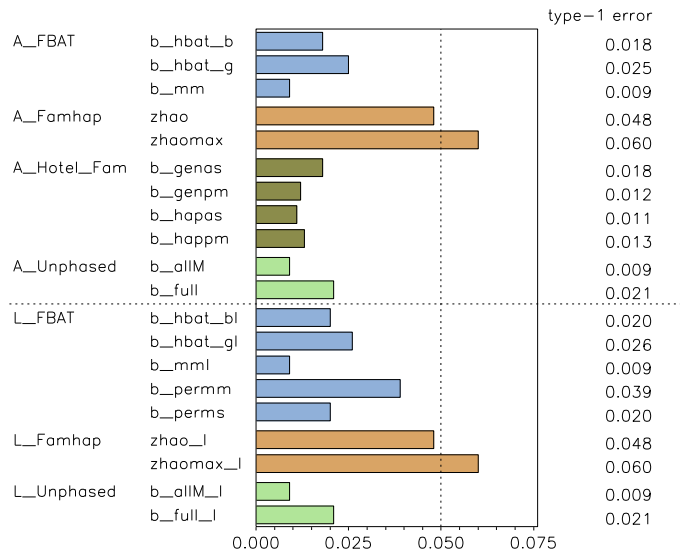
Hotel\_Fam: Type I error, alpha=0.01  
2500 AAUU families, NMod 3



# Fehler I. Art, globale Betrachtung, $\alpha = 0.05$

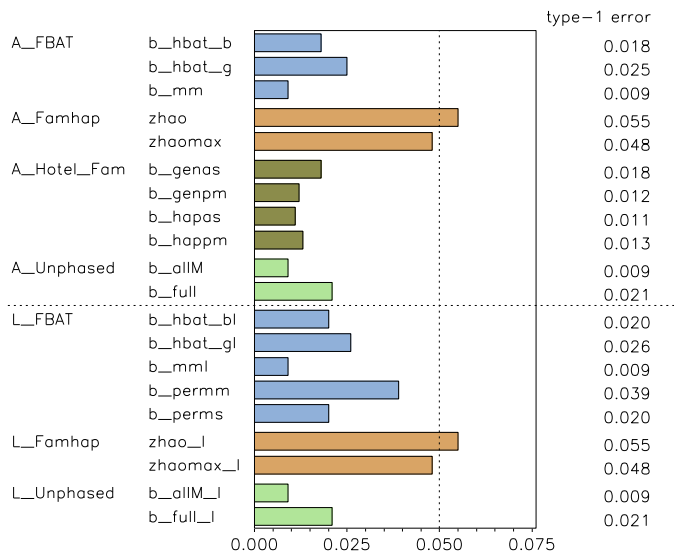
## NMod 1, Trios

Type 1 Error: alpha=0.05, 200 trios, NMod1



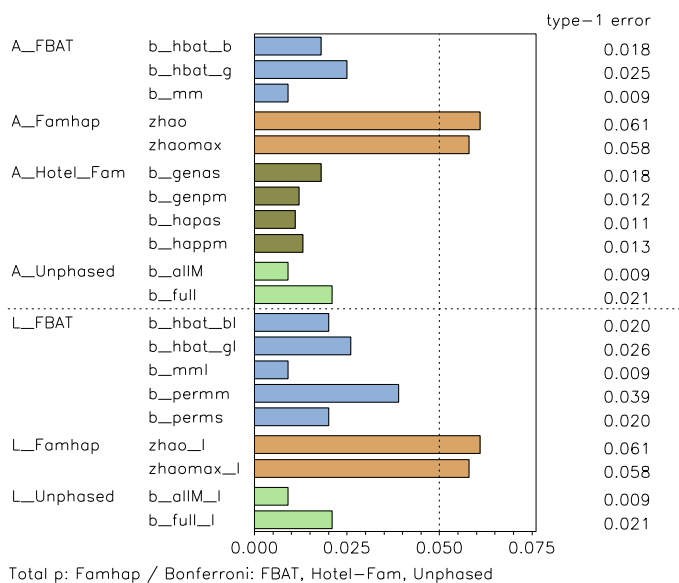
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Type 1 Error: alpha=0.05, 1000 trios, NMod1



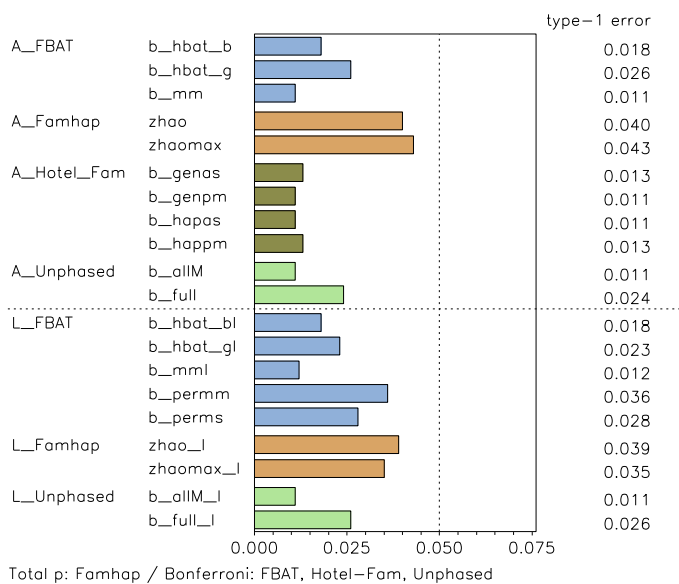
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Type 1 Error: alpha=0.05, 2500 trios, NMod1

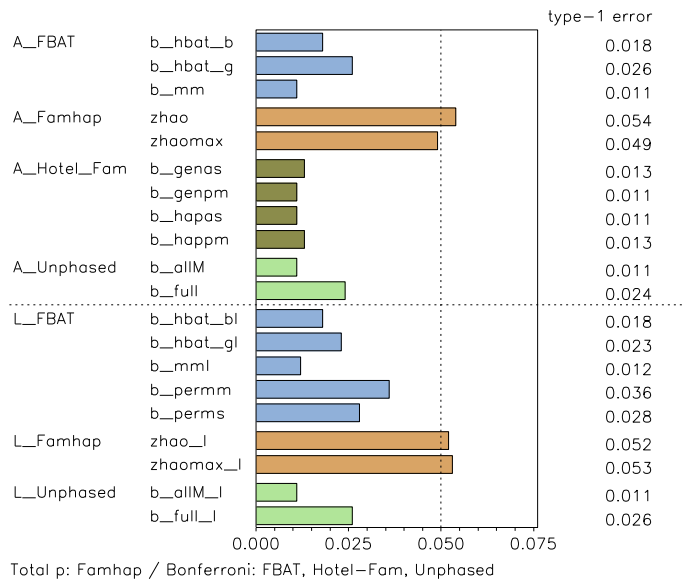


## NMod 1, AA-Familien

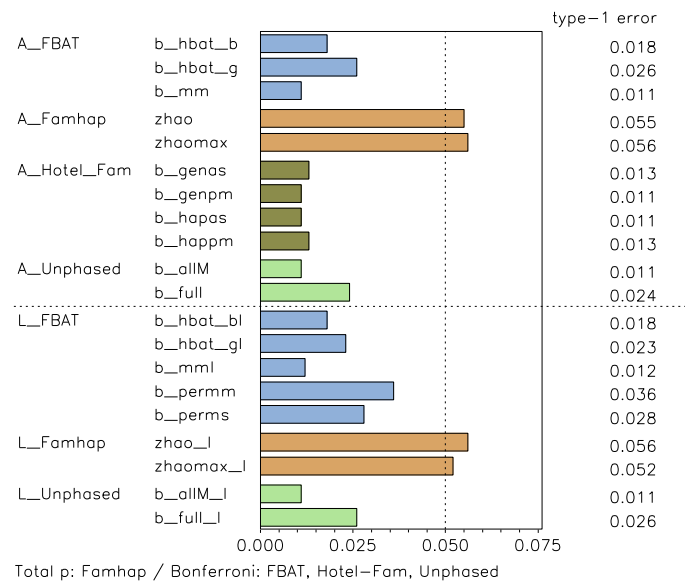
Type 1 Error: alpha=0.05, 200 AA-families, NMod1



Type 1 Error: alpha=0.05, 1000 AA-families, NMod1



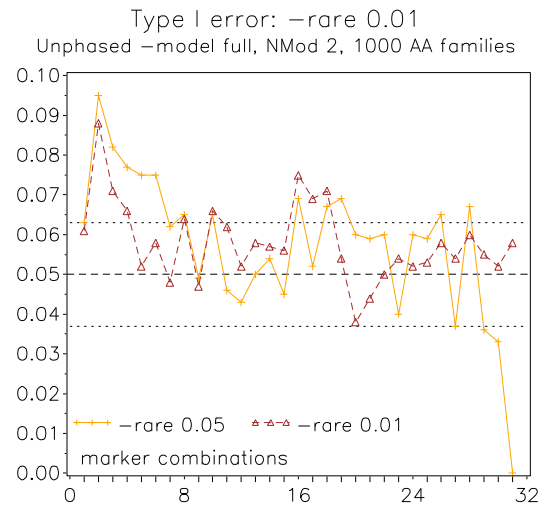
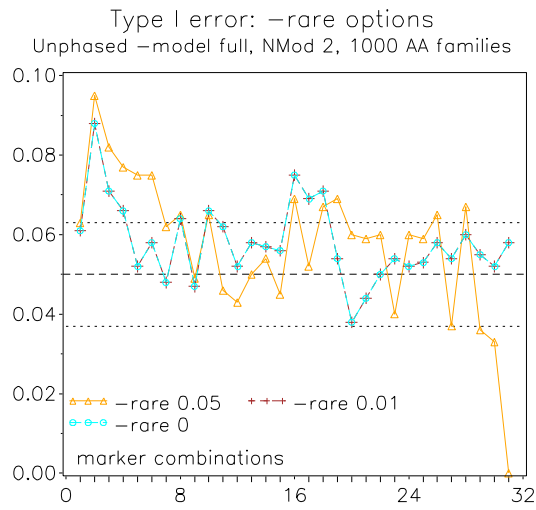
Type 1 Error: alpha=0.05, 2500 AA-families, NMod1



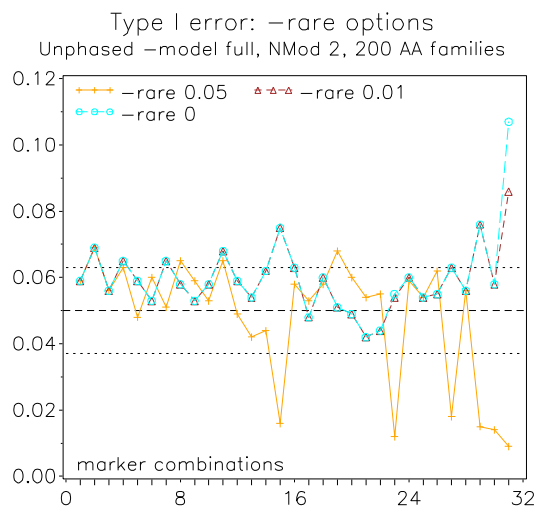
# Fehler 1. Art: Diagramme für alle Markerkombinationen

Testniveau  $\alpha = 0.05$ , Unphased, Option `-rare`

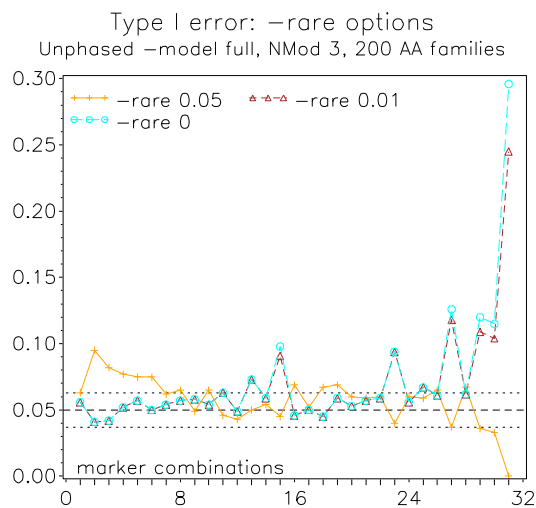
## NMod 2, 1000 AA Familien



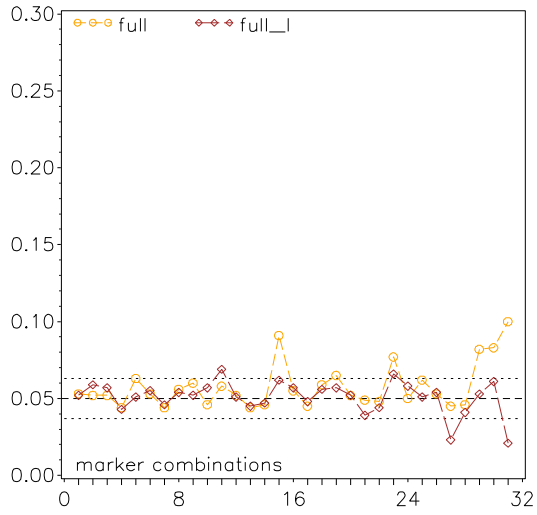
## NMod 2, 200 AA Familien



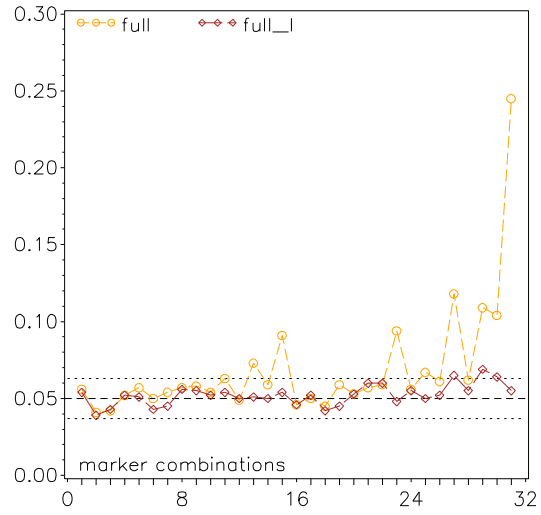
## NMod 3, 200 AA Familien



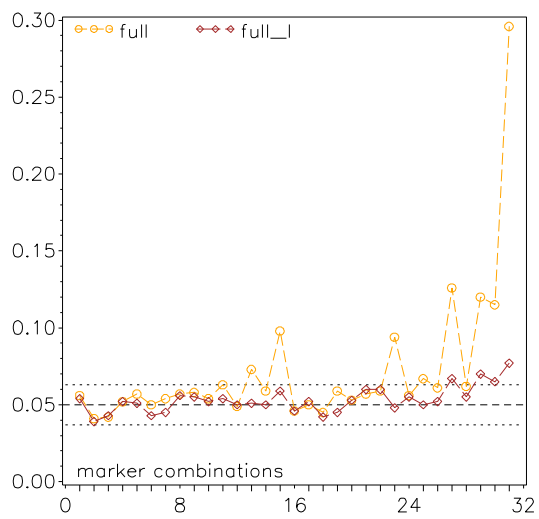
Type I error: 200 AA families, Nmod 3  
Unphased -rare 0.05



Type I error: 200 AA families, Nmod 3  
Unphased -rare 0.01

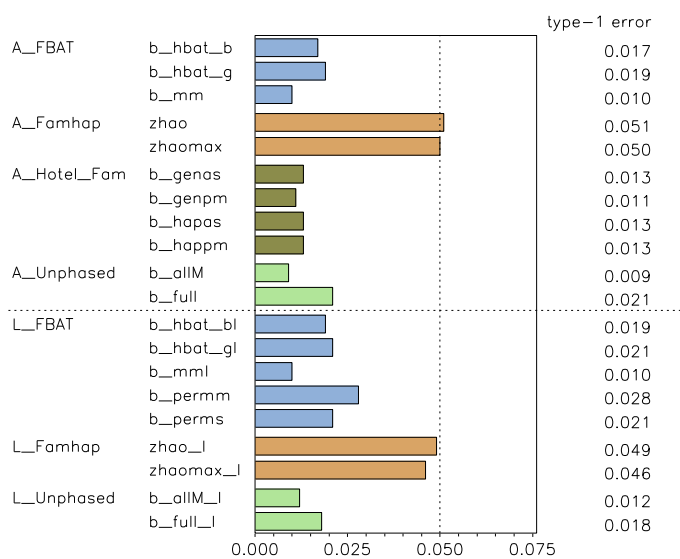


Type I error: 200 AA families, Nmod 3  
Unphased -rare 0.00



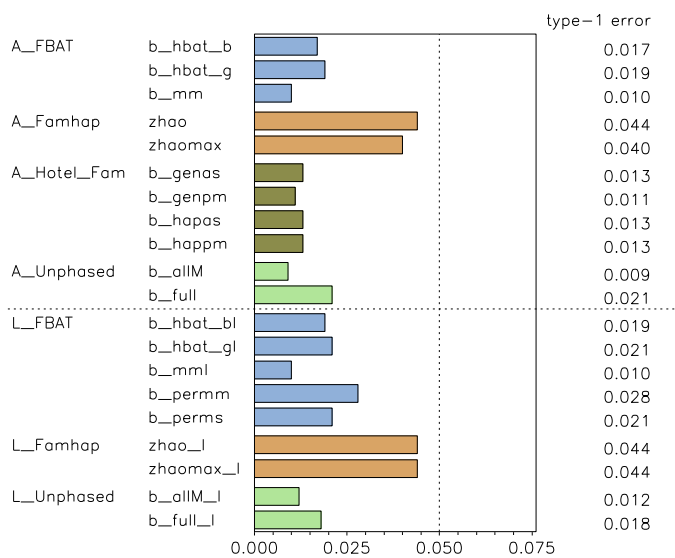
## NMod 1, AAUU-Familien

Type 1 Error: alpha=0.05, 200 AAUU-familien, NMod1



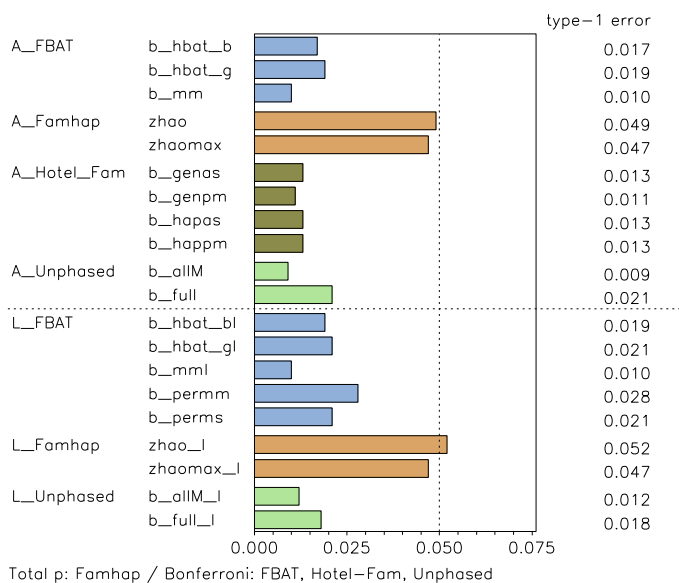
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Type 1 Error: alpha=0.05, 1000 AAUU-familien, NMod1



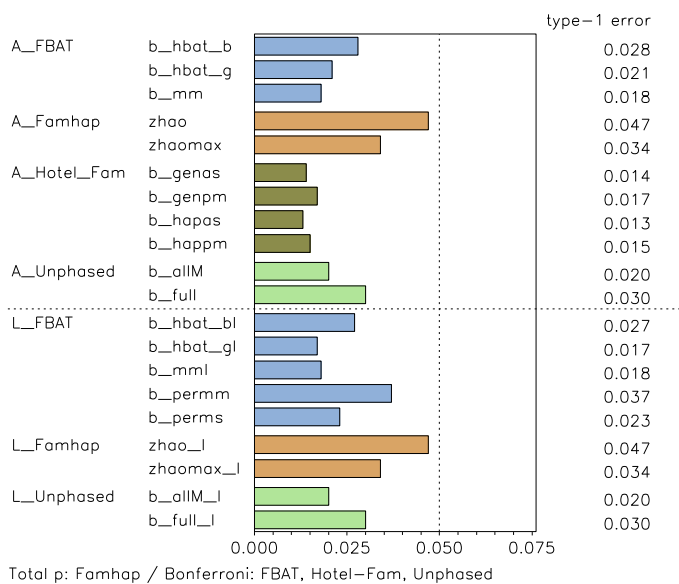
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Type 1 Error: alpha=0.05, 2500 AAU-family, NMod1



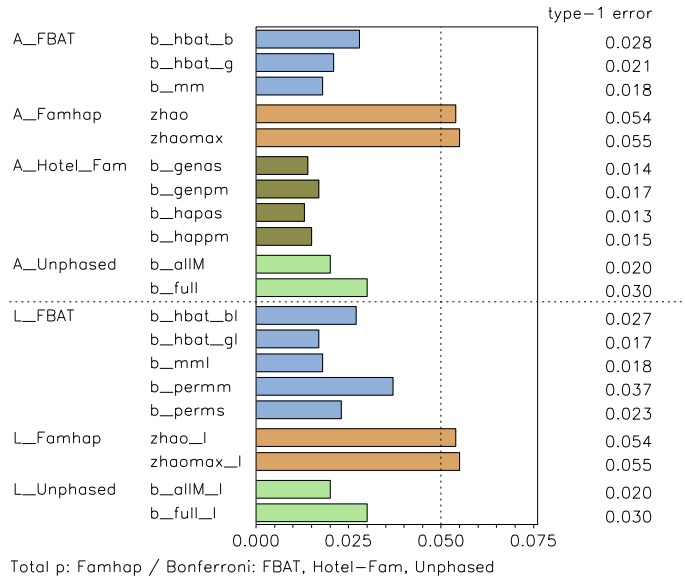
NMod 2, Trios

Type 1 Error: alpha=0.05, 200 trios, NMod2

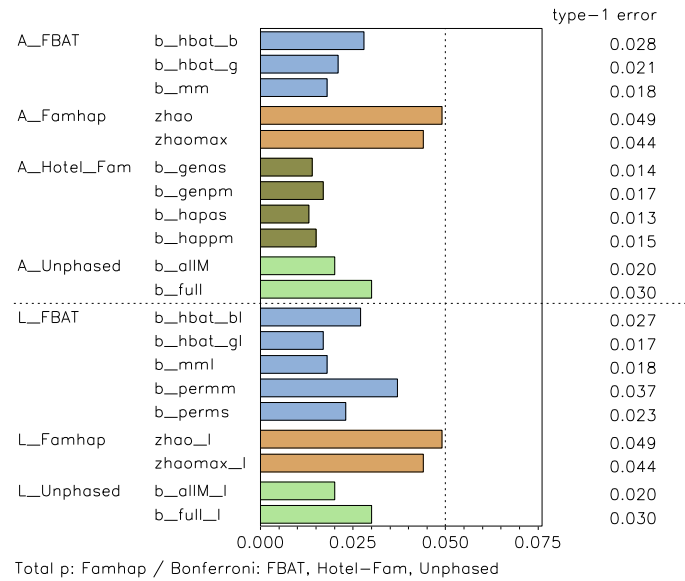




Type 1 Error: alpha=0.05, 1000 trios, NMod2

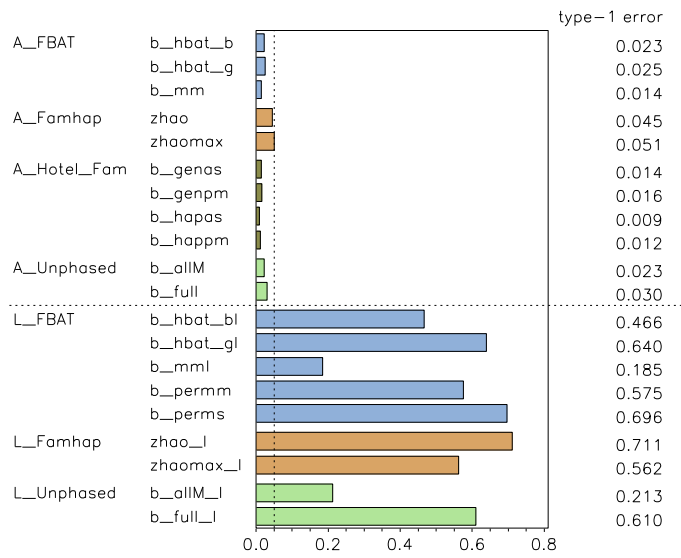


Type 1 Error: alpha=0.05, 2500 trios, NMod2



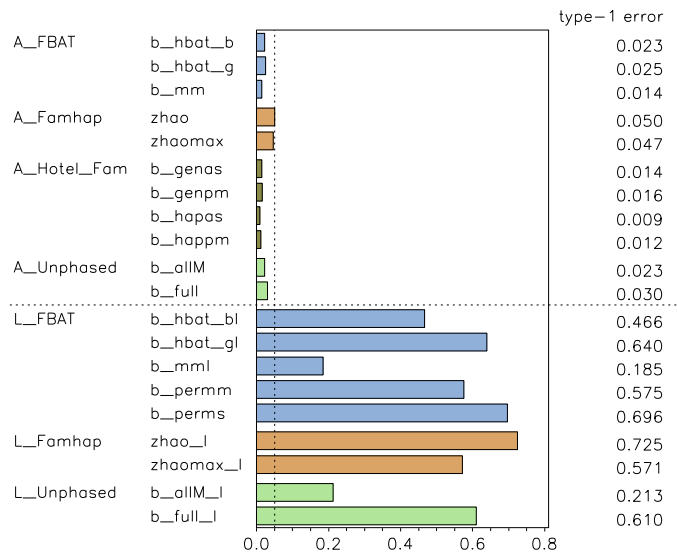
## NMod 2, AA-Familien

Type 1 Error: alpha=0.05, 200 AA-families, NMod2



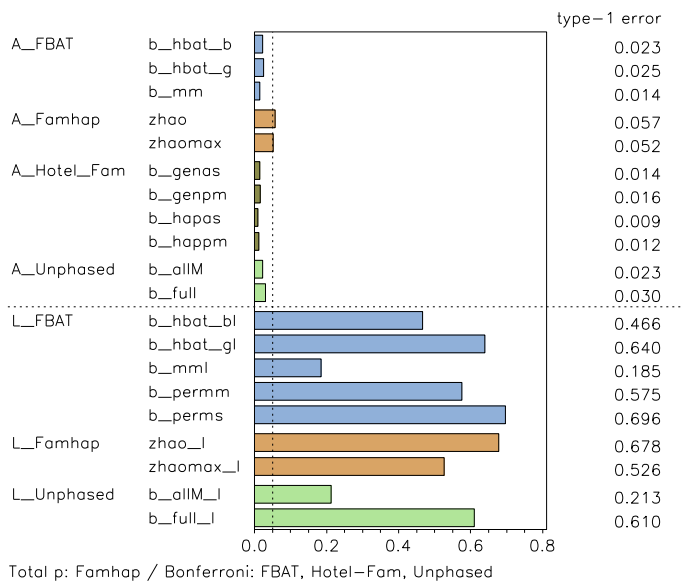
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Type 1 Error: alpha=0.05, 1000 AA-families, NMod2



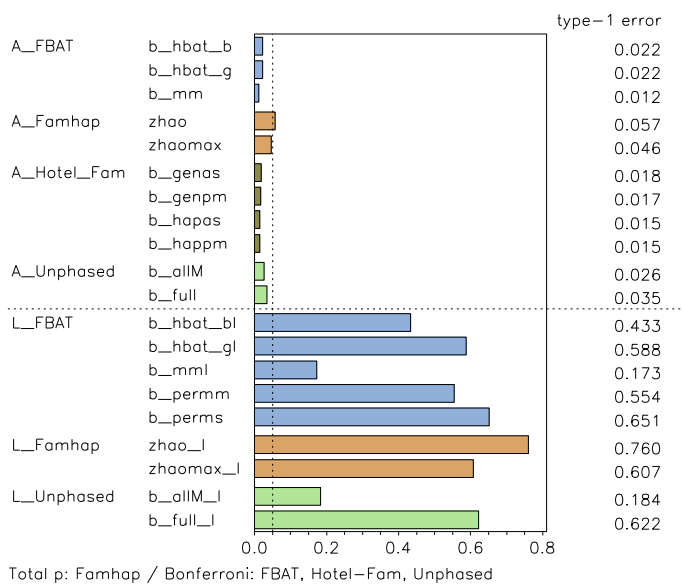
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Type 1 Error: alpha=0.05, 2500 AA-families, NMod2

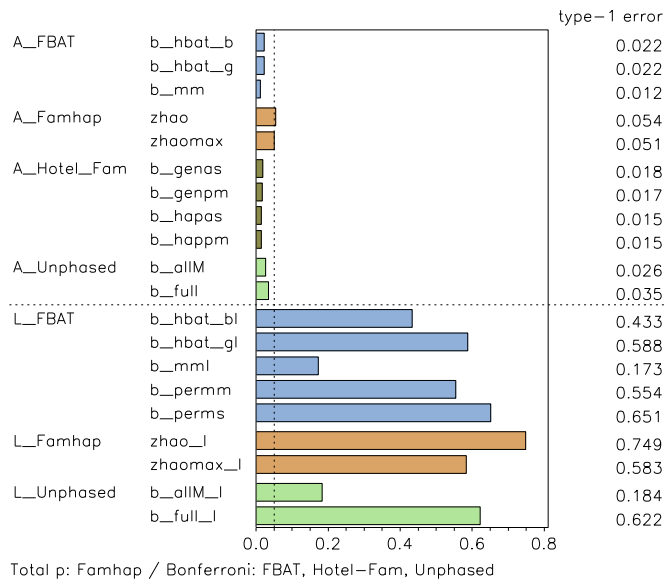


## NMod 2, AAUU-Familien

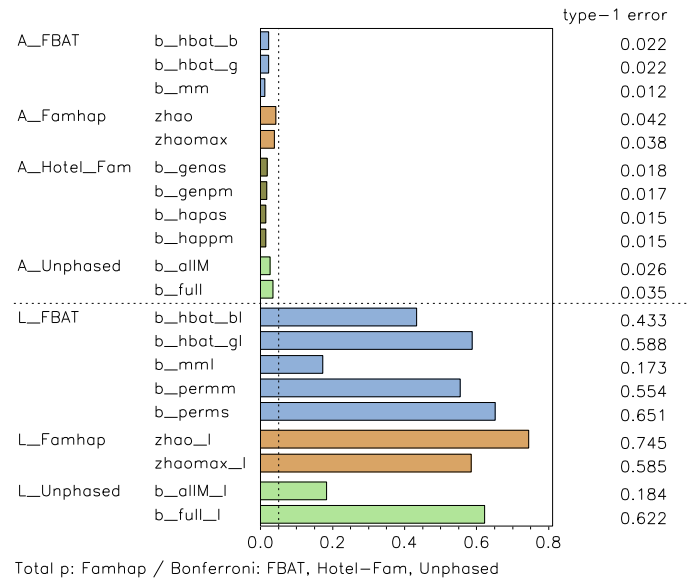
Type 1 Error: alpha=0.05, 200 AAUU-families, NMod2



Type 1 Error: alpha=0.05, 1000 AAU-families, NMod2

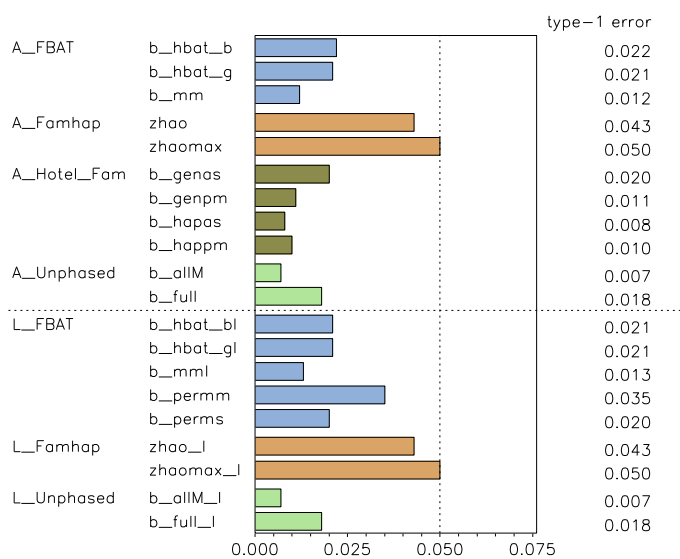


Type 1 Error: alpha=0.05, 2500 AAU-families, NMod2



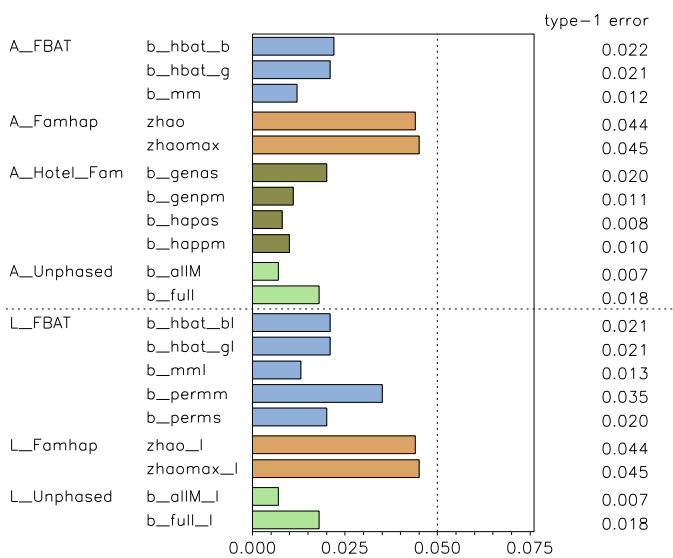
### NMod 3, Trios

Type 1 Error: alpha=0.05, 200 trios, NMod3



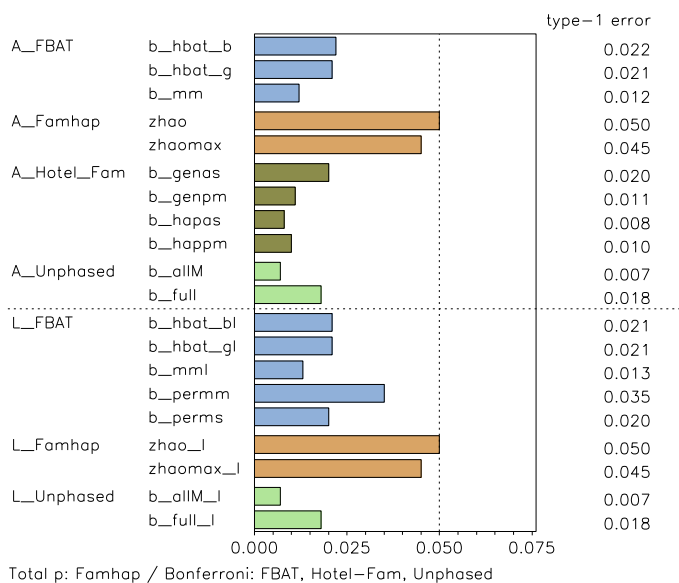
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Type 1 Error: alpha=0.05, 1000 trios, NMod3



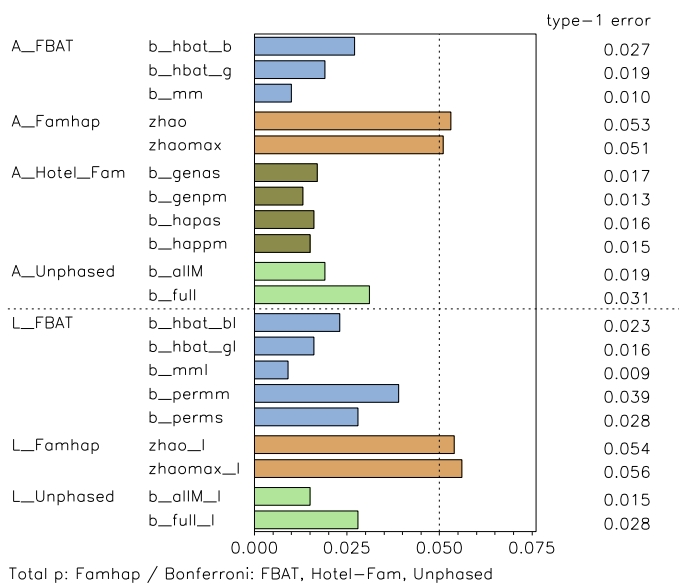
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Type 1 Error: alpha=0.05, 2500 trios, NMod3

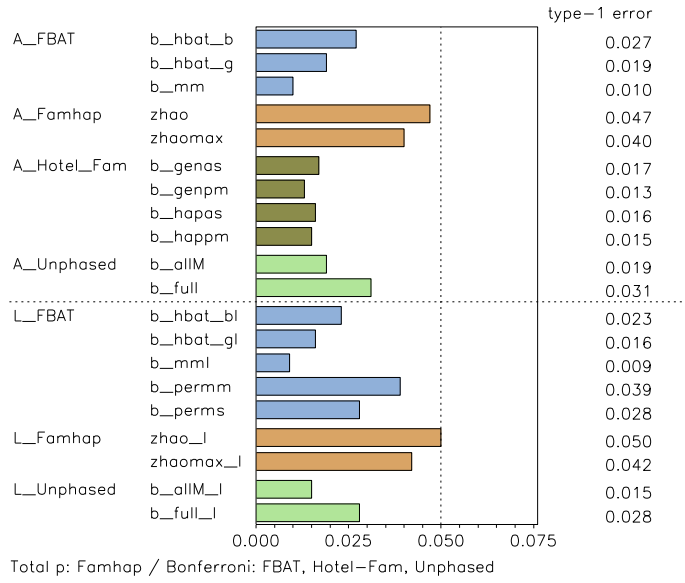


### NMod 3, AA-Familien

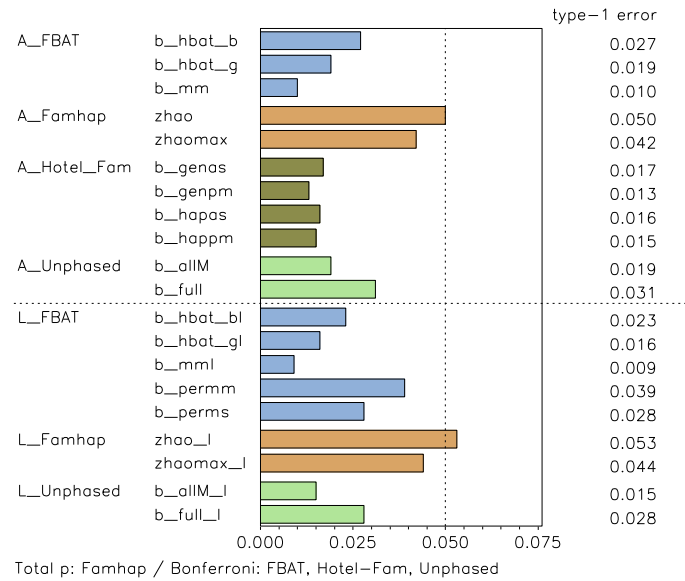
Type 1 Error: alpha=0.05, 200 AA-families, NMod3



Type 1 Error: alpha=0.05, 1000 AA-families, NMod3

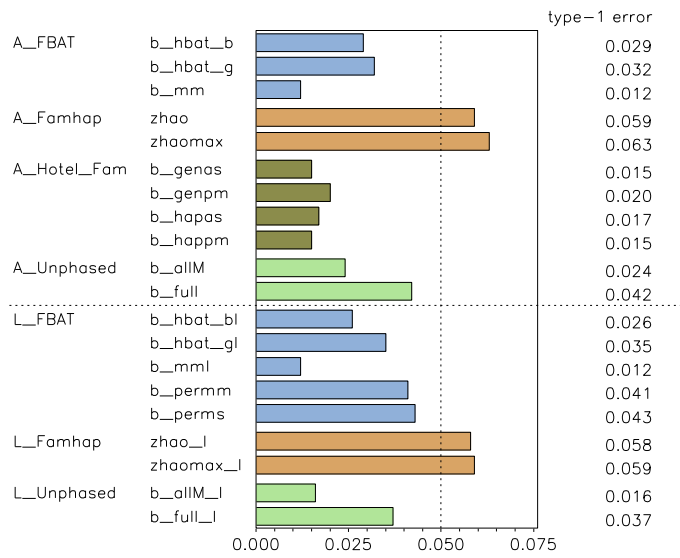


Type 1 Error: alpha=0.05, 2500 AA-families, NMod3



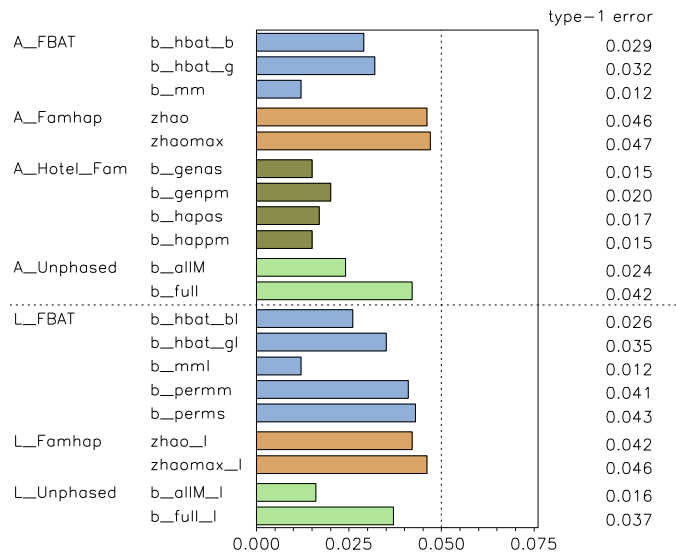
## NMod 3, AAU-Familien

Type 1 Error:  $\alpha=0.05$ , 200 AAU-families, NMod3



Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

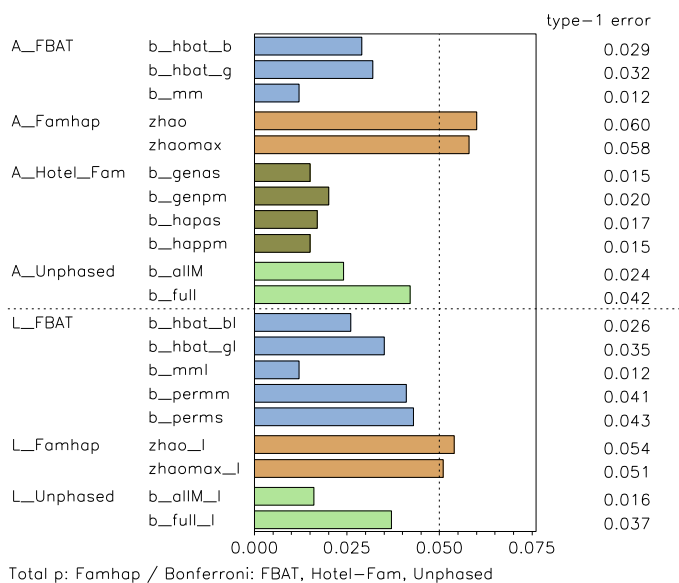
Type 1 Error:  $\alpha=0.05$ , 1000 AAU-families, NMod3



Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased



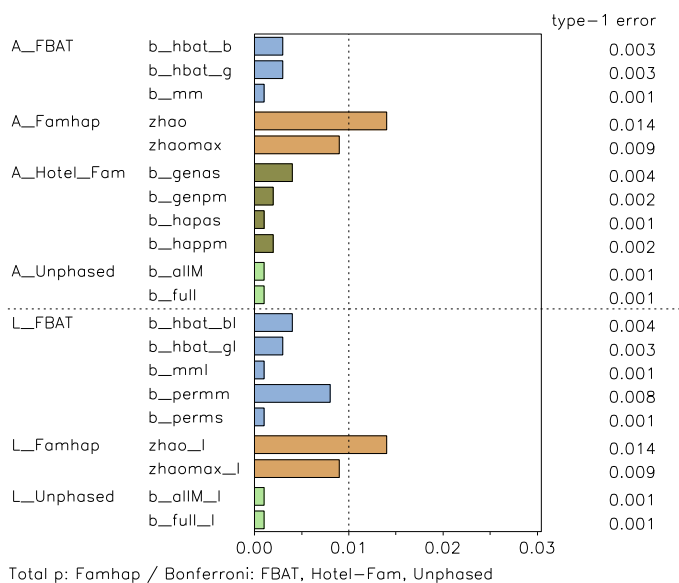
Type 1 Error: alpha=0.05, 2500 AAU-families, NMod3



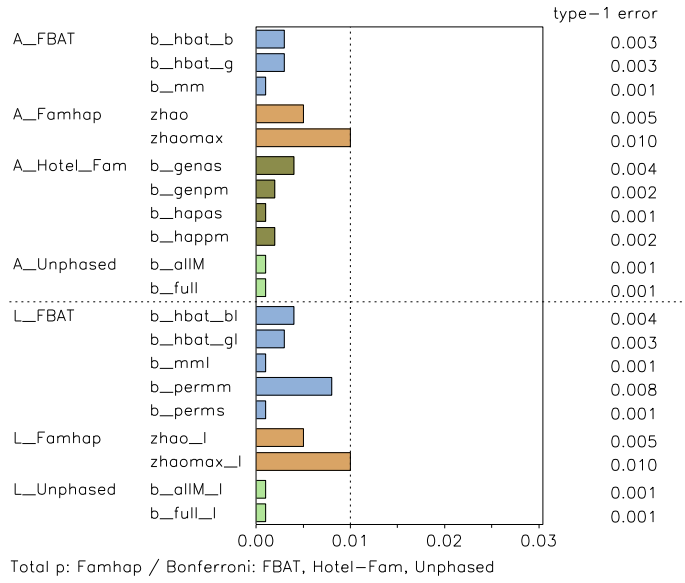
## Fehler I. Art, globale Betrachtung, $\alpha = 0.01$

### NMod 1, Trios

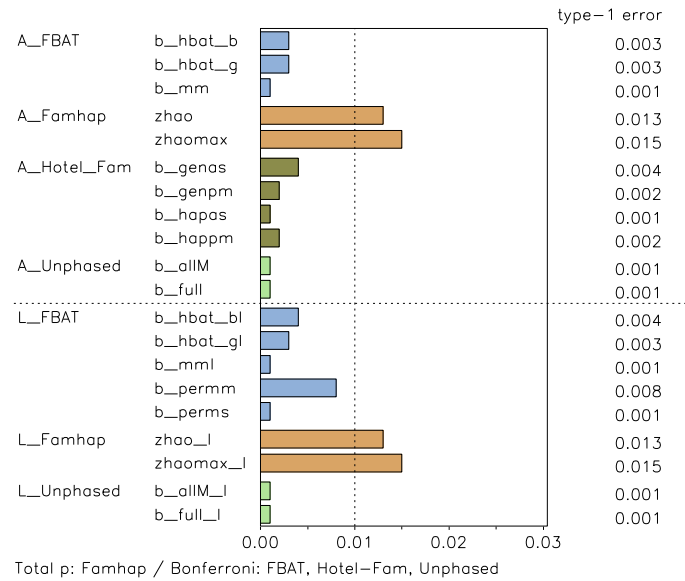
Type 1 Error: alpha=0.01, 200 trios, NMod1



Type 1 Error: alpha=0.01, 1000 trios, NMod1

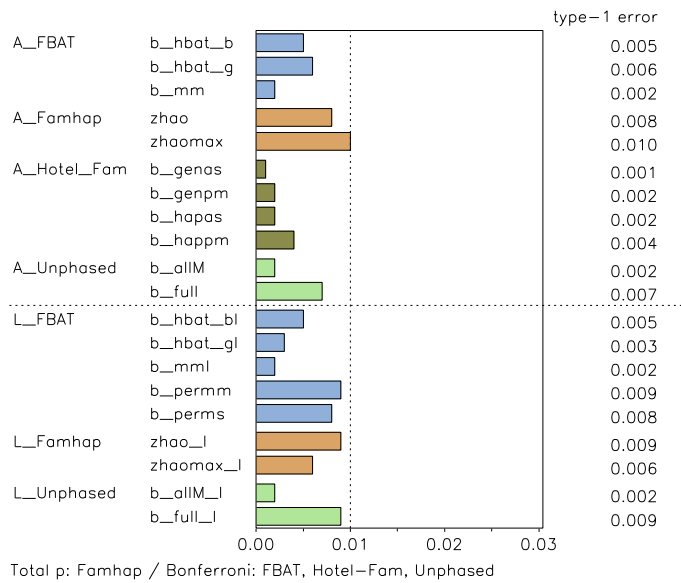


Type 1 Error: alpha=0.01, 2500 trios, NMod1

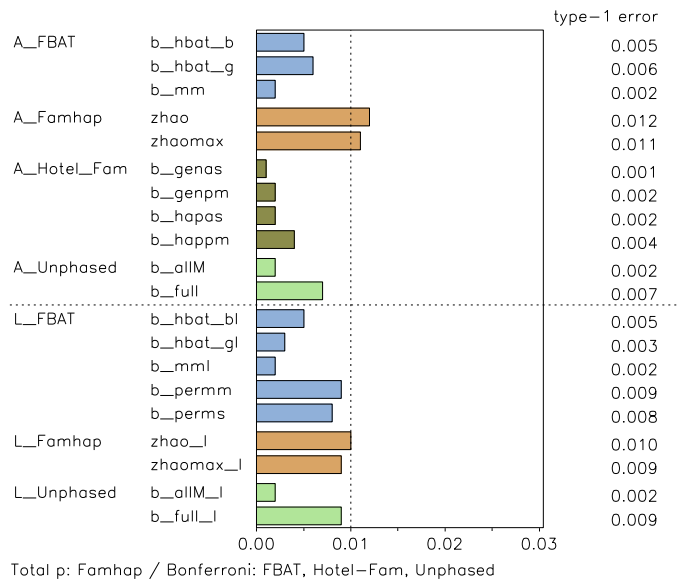


## NMod 1, AA-Familien

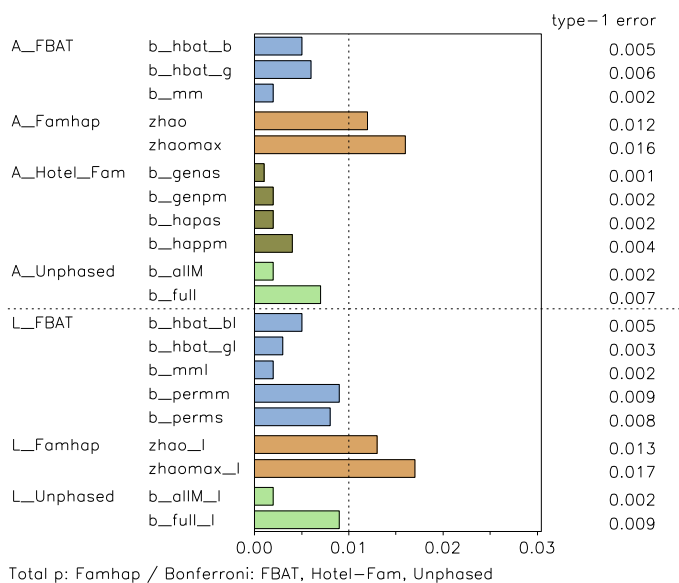
Type 1 Error: alpha=0.01, 200 AA-families, NMod1



Type 1 Error: alpha=0.01, 1000 AA-families, NMod1

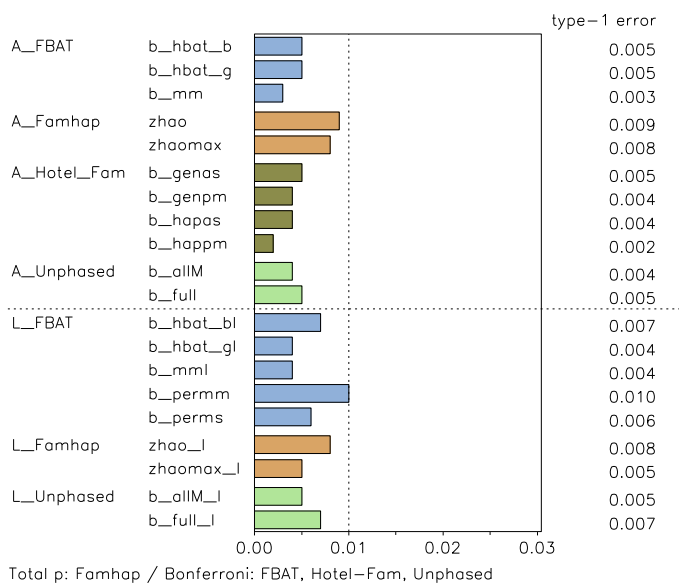


Type 1 Error: alpha=0.01, 2500 AA-families, NMod1

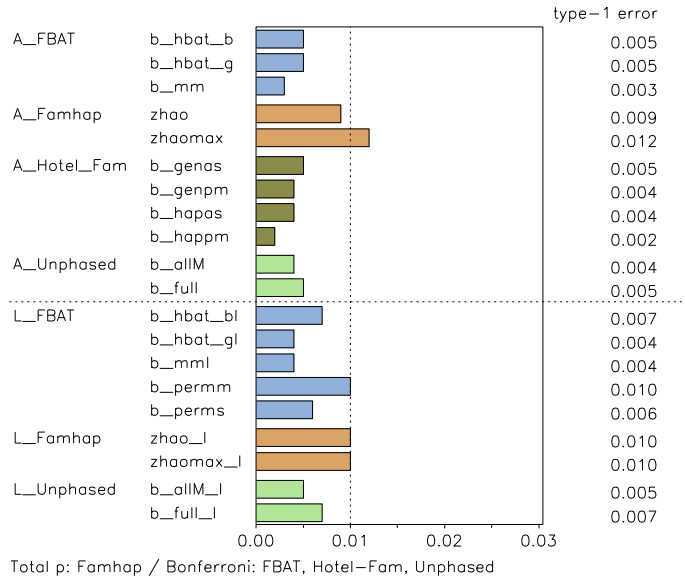


### NMod 1, AAUU-Familien

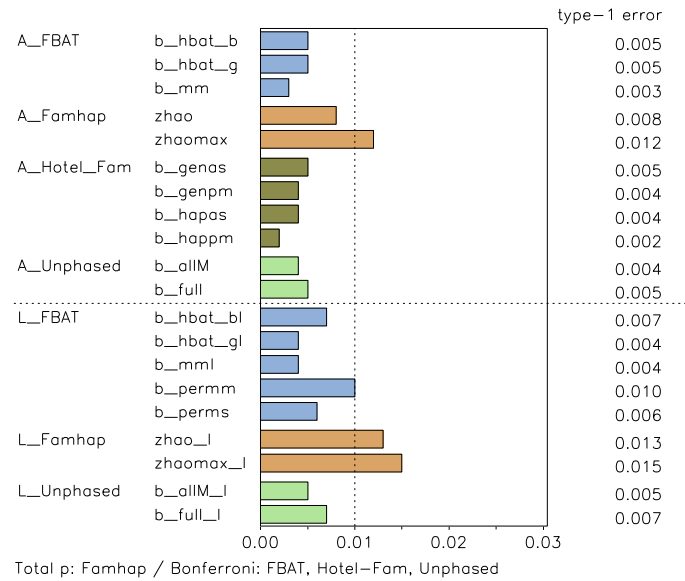
Type 1 Error: alpha=0.01, 200 AAUU-families, NMod1



Type 1 Error: alpha=0.01, 1000 AAU-families, NMod1

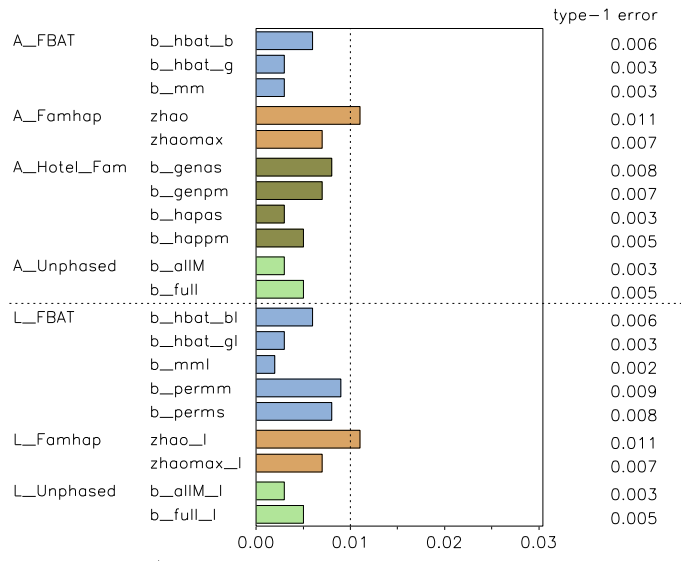


Type 1 Error: alpha=0.01, 2500 AAU-families, NMod1



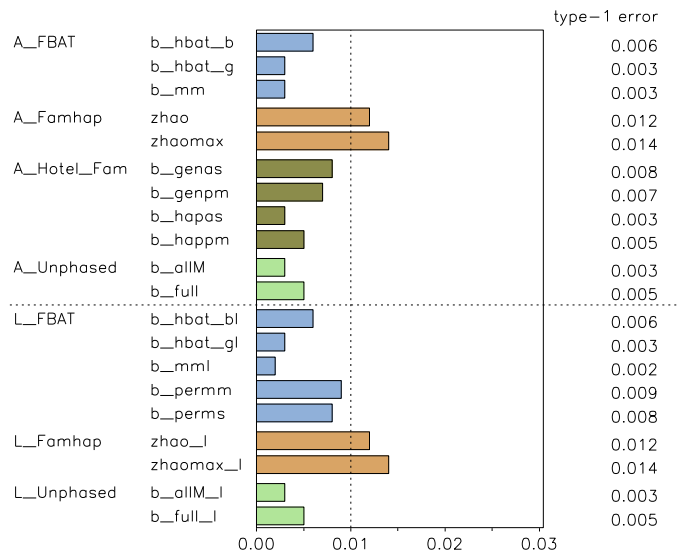
## NMod 2, Trios

Type 1 Error: alpha=0.01, 200 trios, NMod2



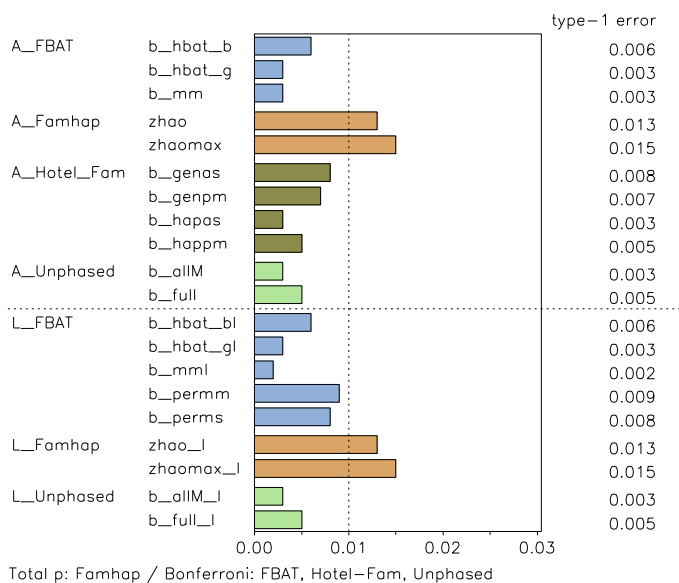
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Type 1 Error: alpha=0.01, 1000 trios, NMod2



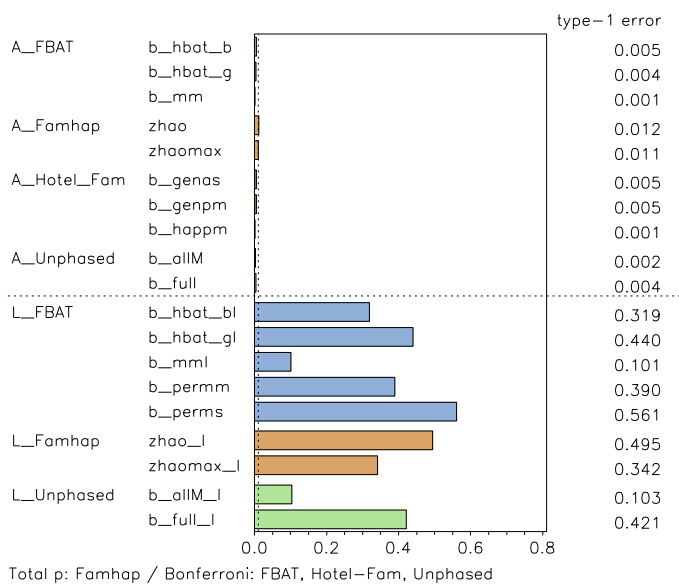
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Type 1 Error: alpha=0.01, 2500 trios, NMod2

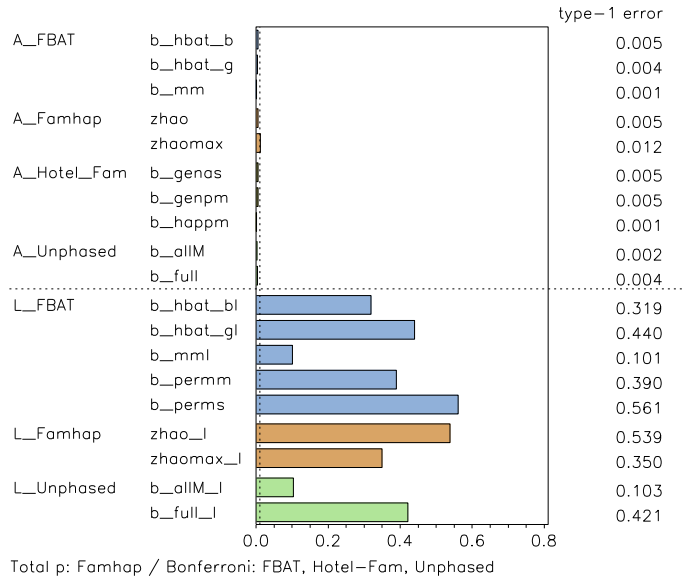


## NMod 2, AA-Familien

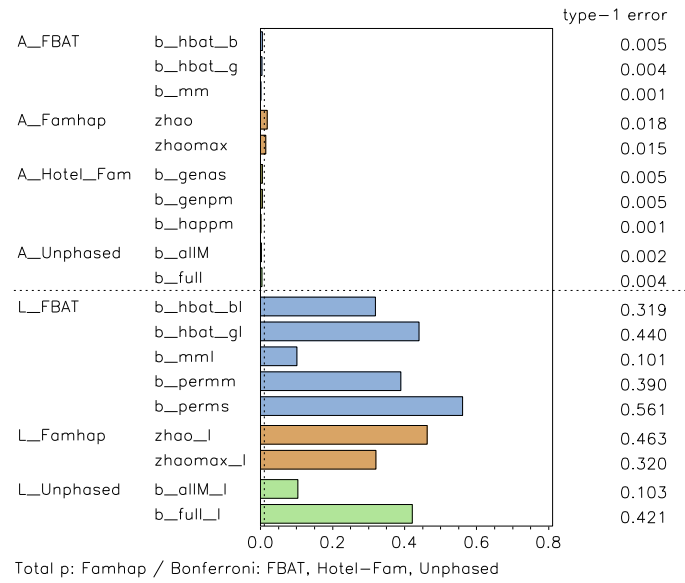
Type 1 Error: alpha=0.01, 200 AA-families, NMod2



Type 1 Error: alpha=0.01, 1000 AA-families, NMod2



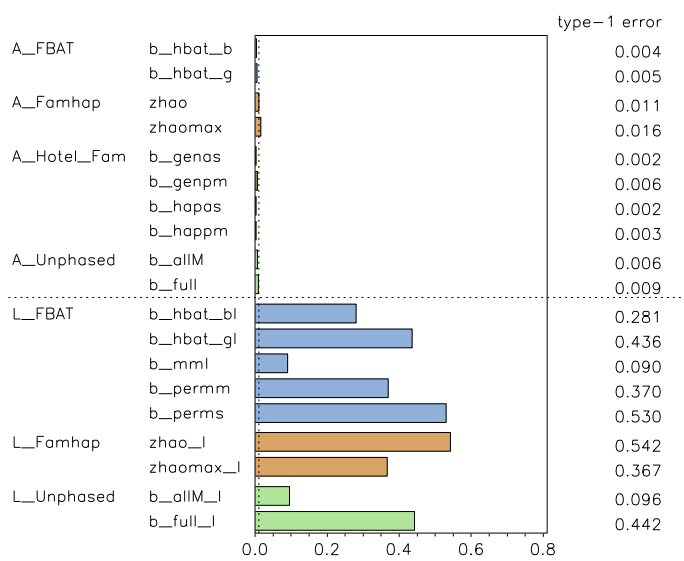
Type 1 Error: alpha=0.01, 2500 AA-families, NMod2





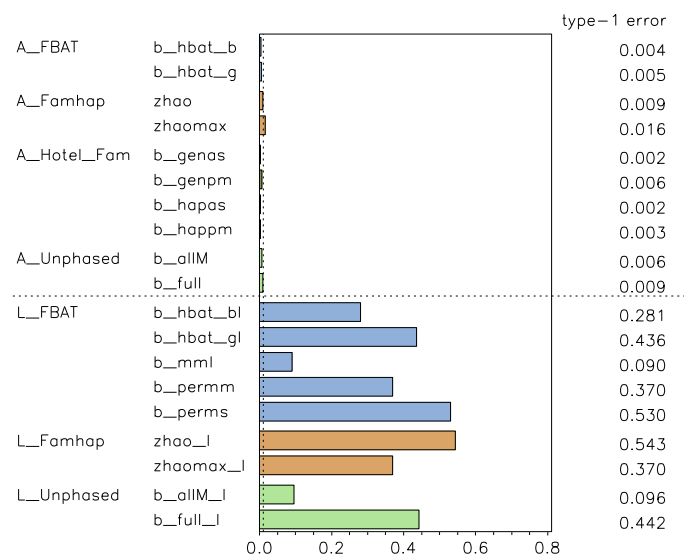
## NMod 2, AAUU-Familien

Type 1 Error: alpha=0.01, 200 AAUU-families, NMod2



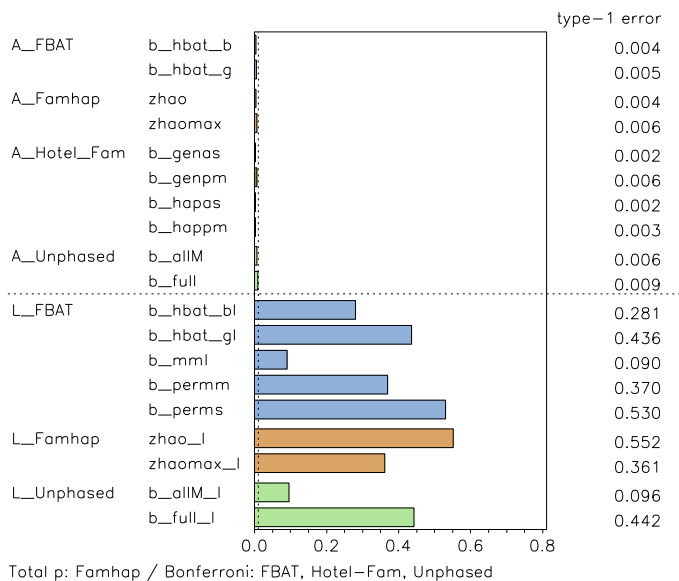
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Type 1 Error: alpha=0.01, 1000 AAUU-families, NMod2



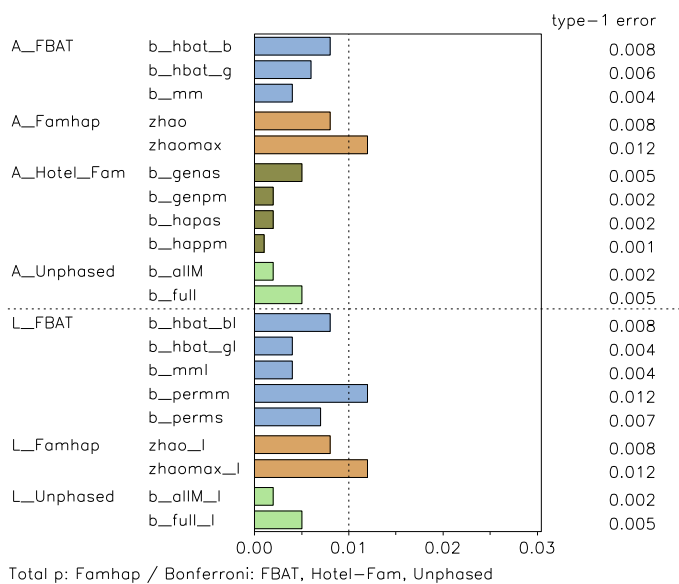
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Type 1 Error: alpha=0.01, 2500 AAU-family, NMod2

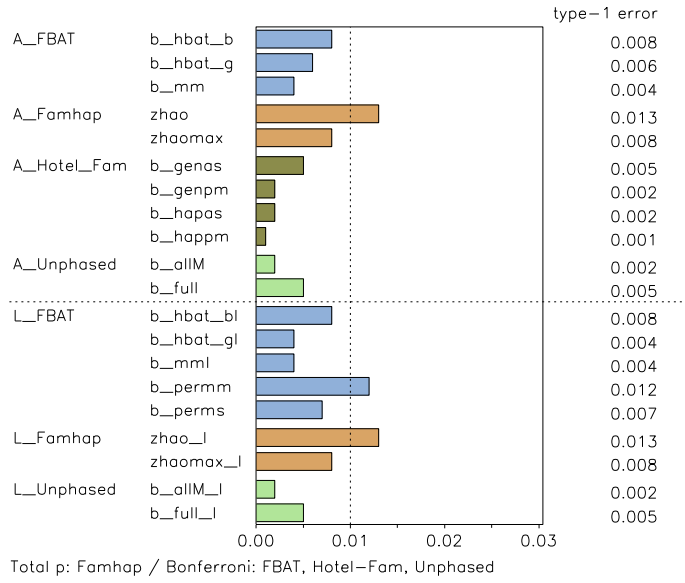


### NMod 3, Trios

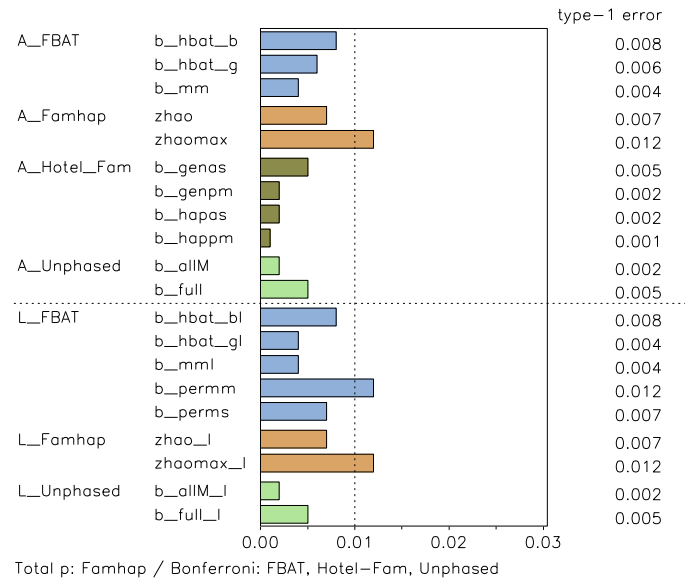
Type 1 Error: alpha=0.01, 200 trios, NMod3



Type 1 Error: alpha=0.01, 1000 trios, NMod3

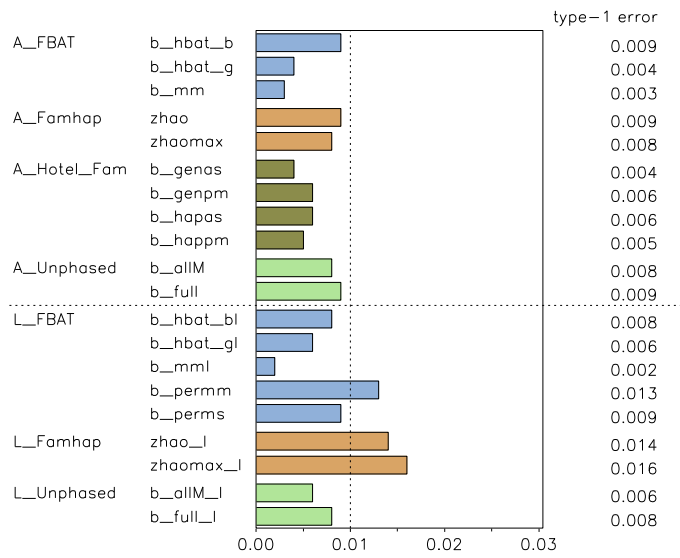


Type 1 Error: alpha=0.01, 2500 trios, NMod3



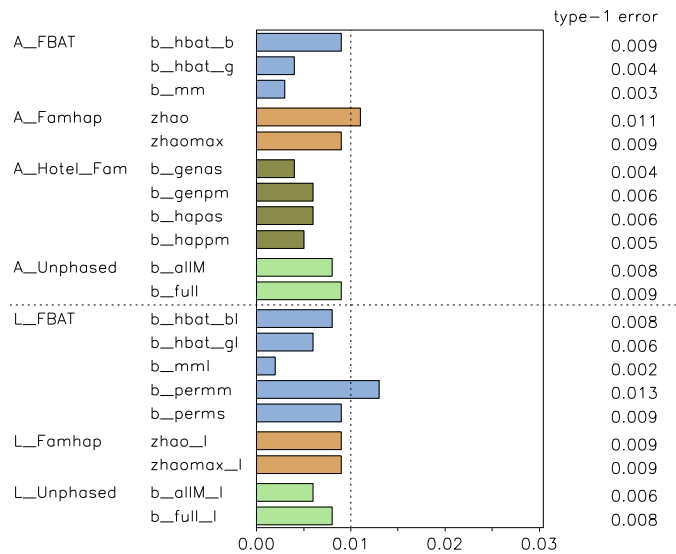
## NMod 3, AA-Familien

Type 1 Error: alpha=0.01, 200 AA-families, NMod3



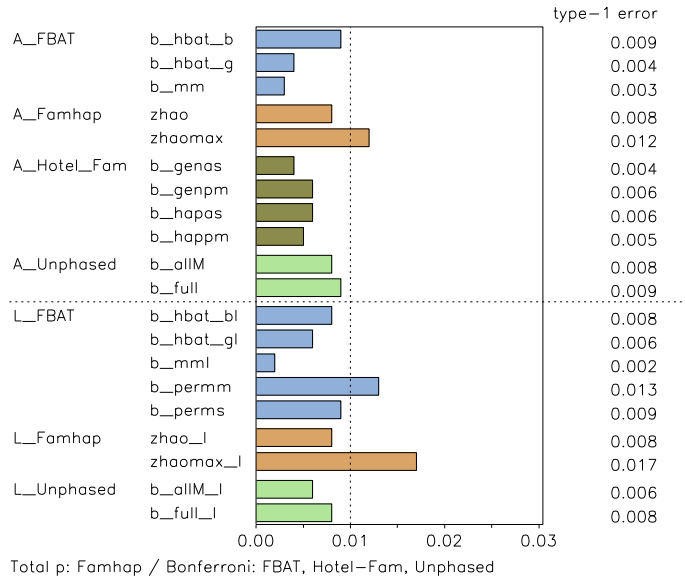
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Type 1 Error: alpha=0.01, 1000 AA-families, NMod3



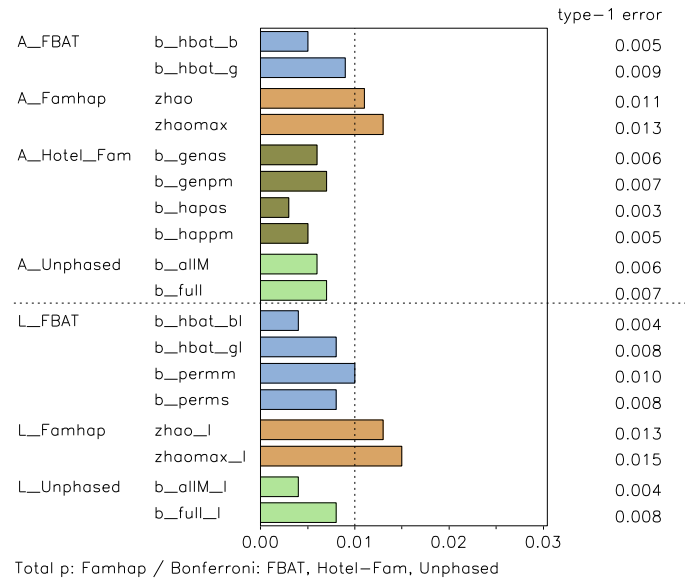
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Type 1 Error: alpha=0.01, 2500 AA-families, NMod3

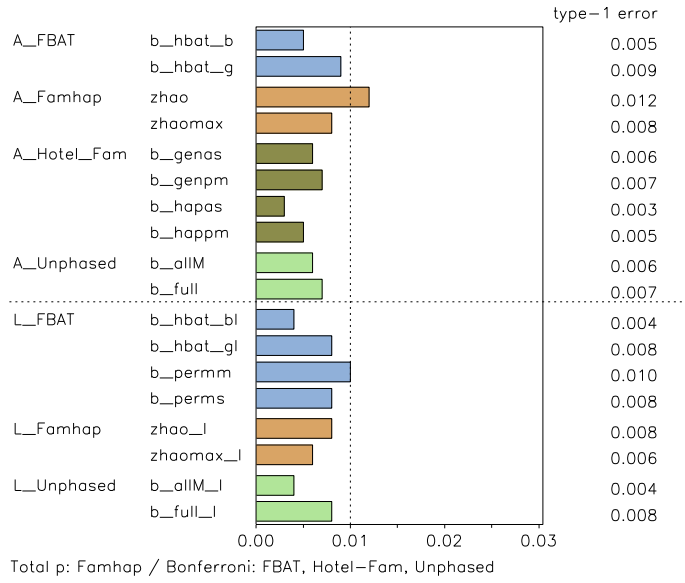


### NMod 3, AAUU-Familien

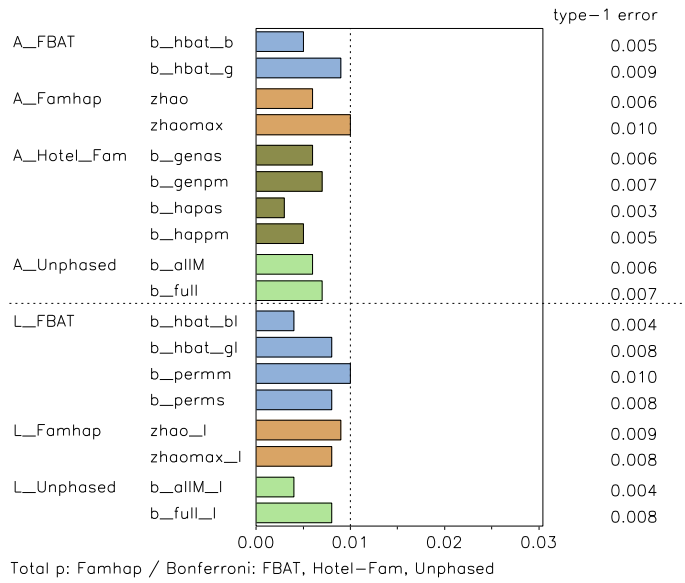
Type 1 Error: alpha=0.01, 200 AAUU-families, NMod3



Type 1 Error: alpha=0.01, 1000 AAU-families, NMod3



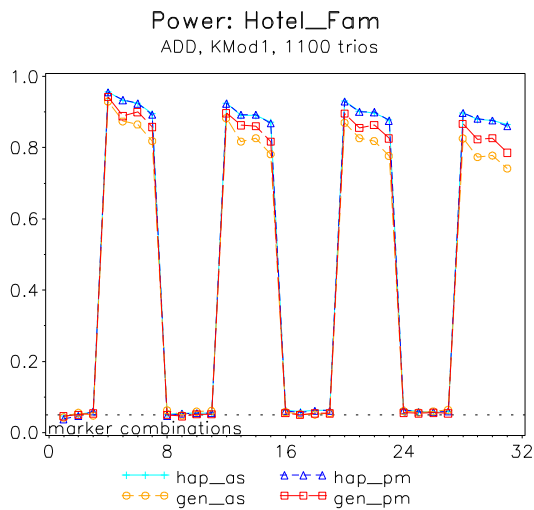
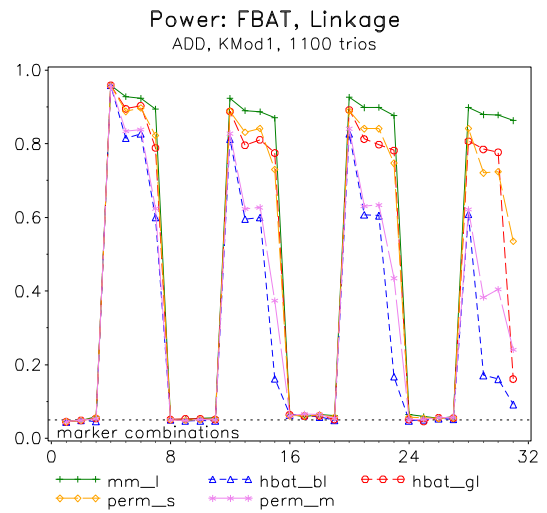
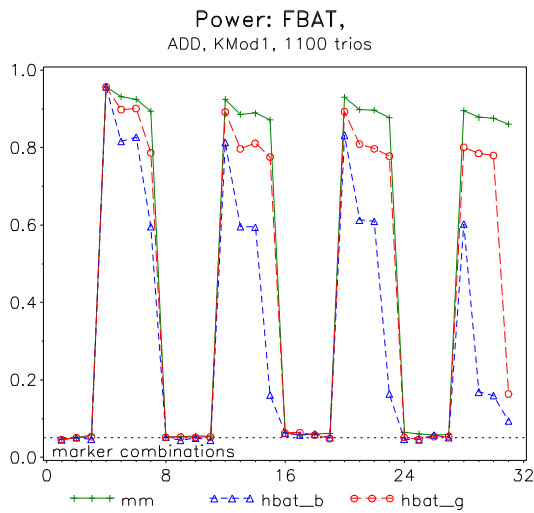
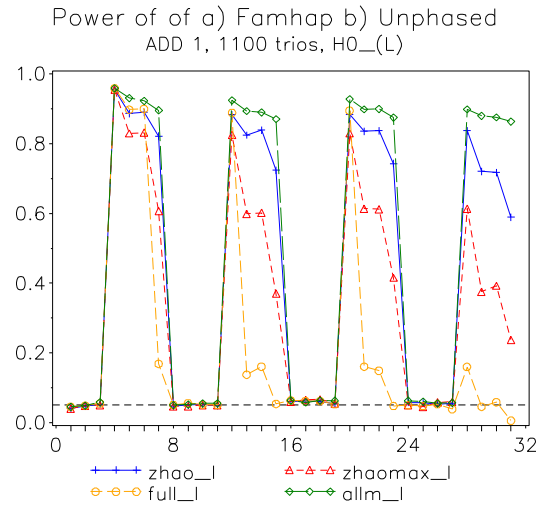
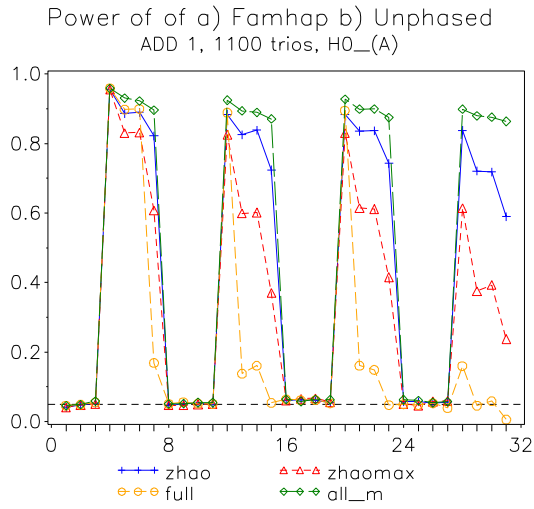
Type 1 Error: alpha=0.01, 2500 AAU-families, NMod3



# Power für alle Markerkombinationen, $\alpha = 0.05$

## KMod 1, Trios

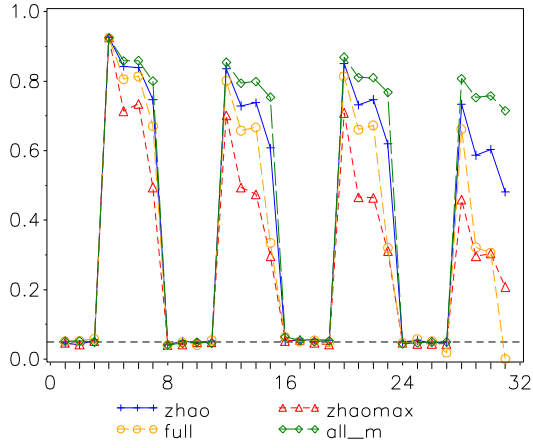
### Additiver Vererbungsmodus



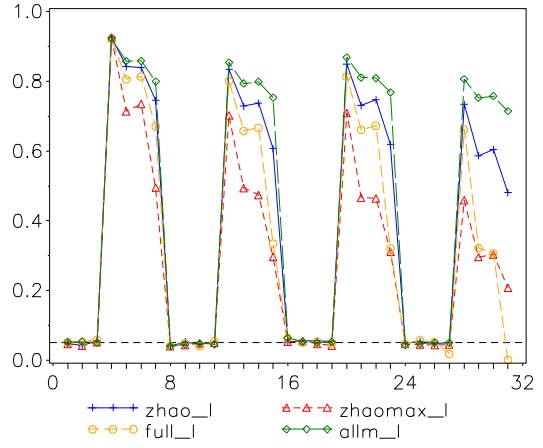
# KMod 1, trios

## Dominanter Vererbungsmodus

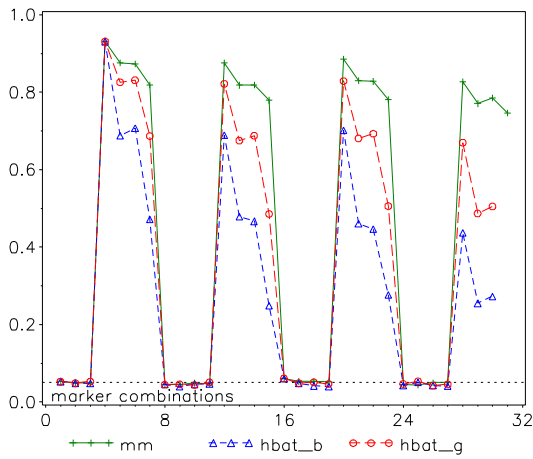
Power of of a) Famhap b) Unphased  
DOM 1, 400 trios, H0\_(A)



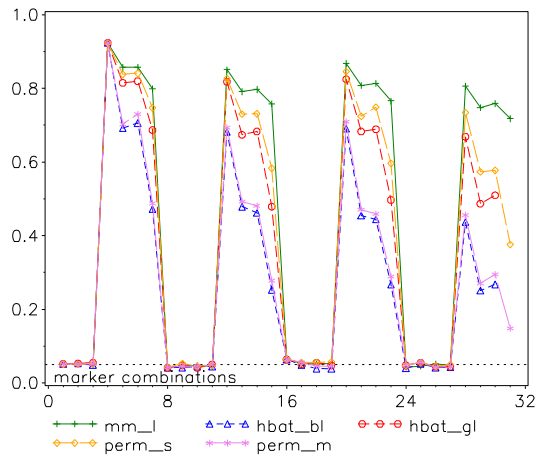
Power of of a) Famhap b) Unphased  
DOM 1, 400 trios, H0\_(L)



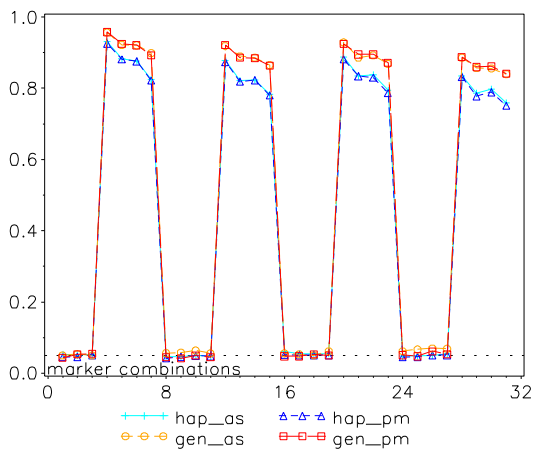
Power: FBAT,  
DOM, KMod1, 400 trios



Power: FBAT, Linkage  
DOM, KMod1, 400 trios



Power: Hotel\_Fam  
DOM, KMod1, 400 trios

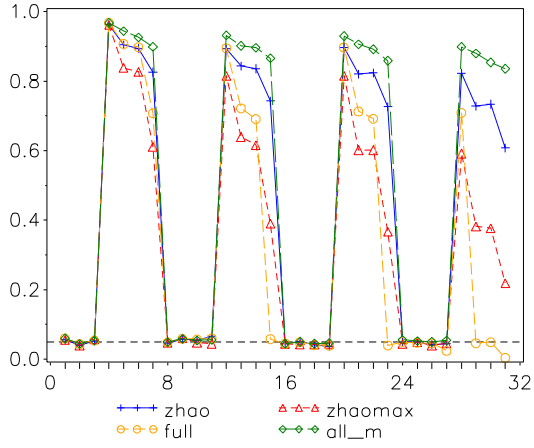




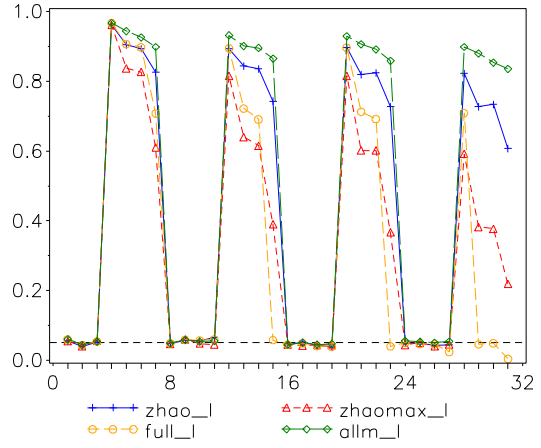
# KMod 1, trios

## Multiplikativer Vererbungsmodus

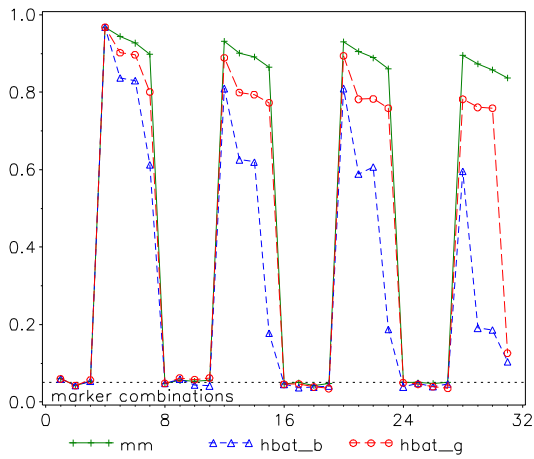
Power of of a) Famhap b) Unphased  
MUL 1, 700 trios, H0\_(A)



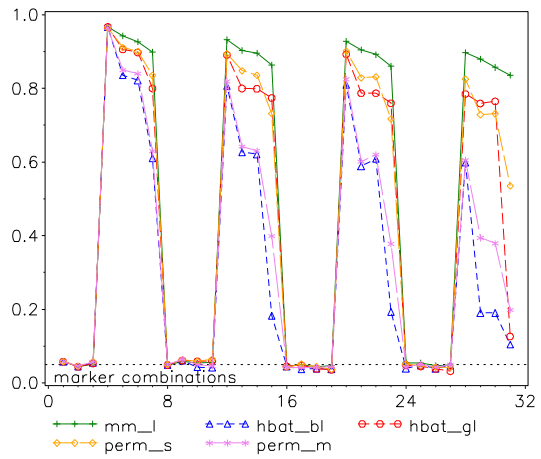
Power of of a) Famhap b) Unphased  
MUL 1, 700 trios, H0\_(L)



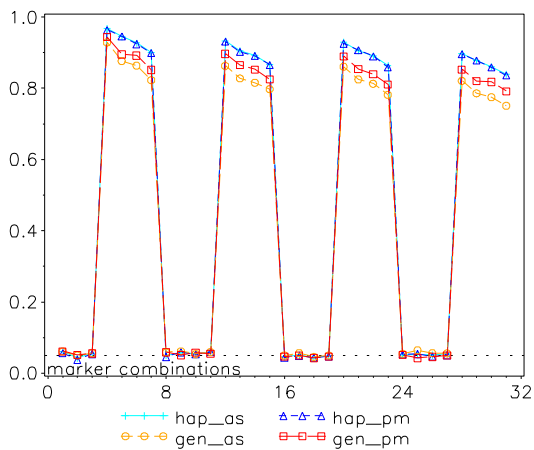
Power: FBAT,  
MUL, KMod1, 700 trios



Power: FBAT, Linkage  
MUL, KMod1, 700 trios



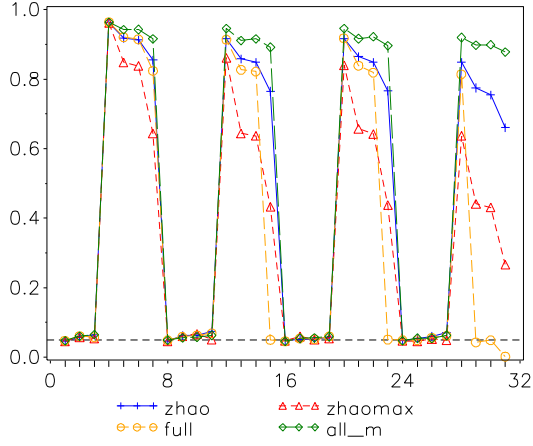
Power: Hotel\_Fam  
MUL, KMod1, 700 trios



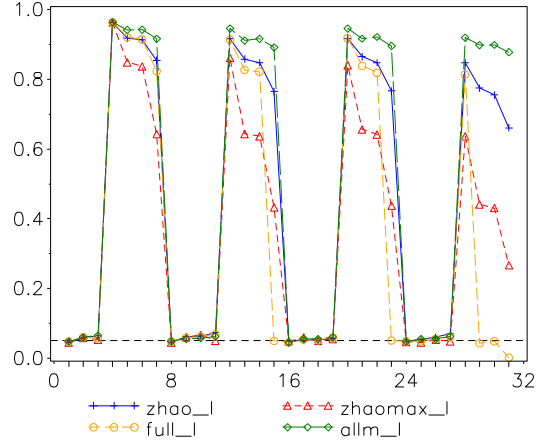
# KMod 1, trios

## Rezessiver Vererbungsmodus

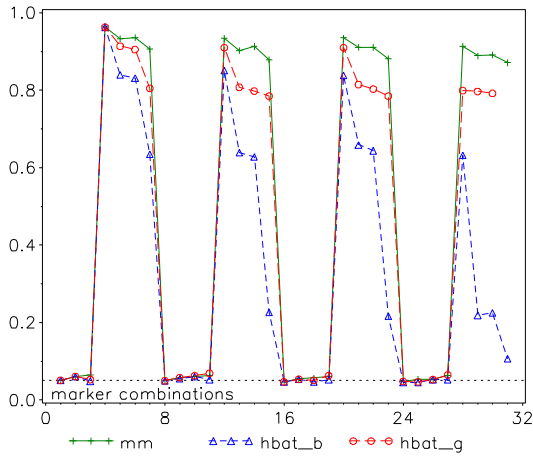
Power of of a) Famhap b) Unphased  
REZ 1, 2500 trios, H0\_(A)



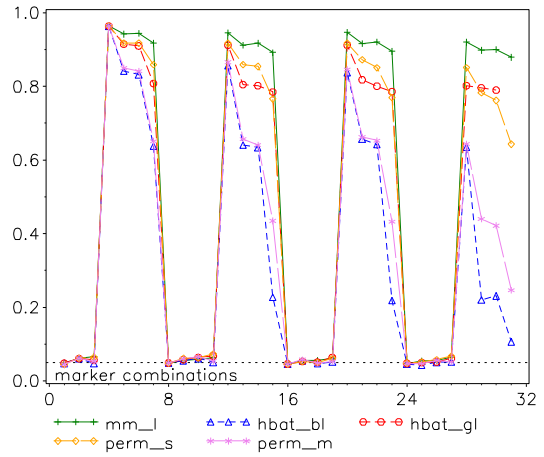
Power of of a) Famhap b) Unphased  
REZ 1, 2500 trios, H0\_(L)



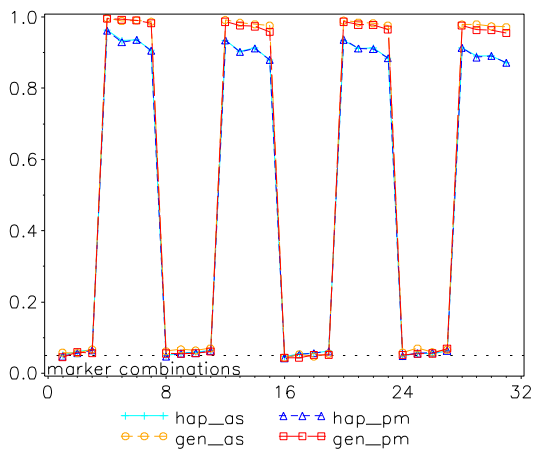
Power: FBAT,  
REZ, KMod1, 2500 trios



Power: FBAT, Linkage  
REZ, KMod1, 2500 trios



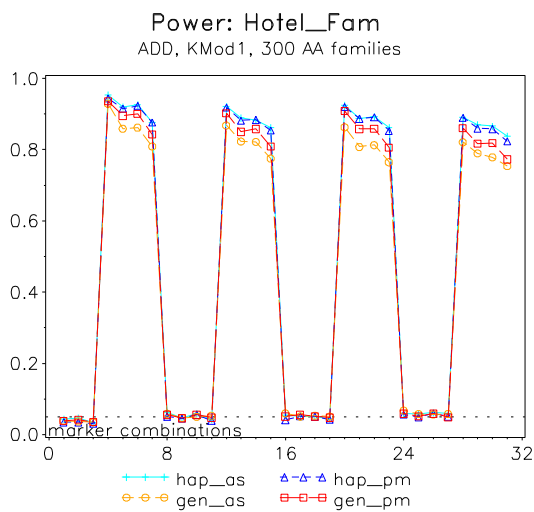
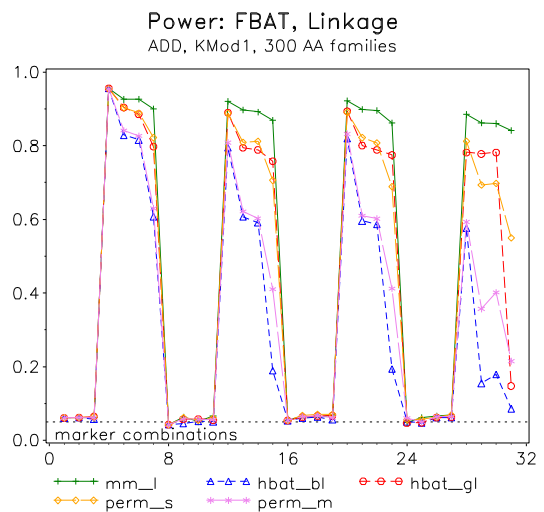
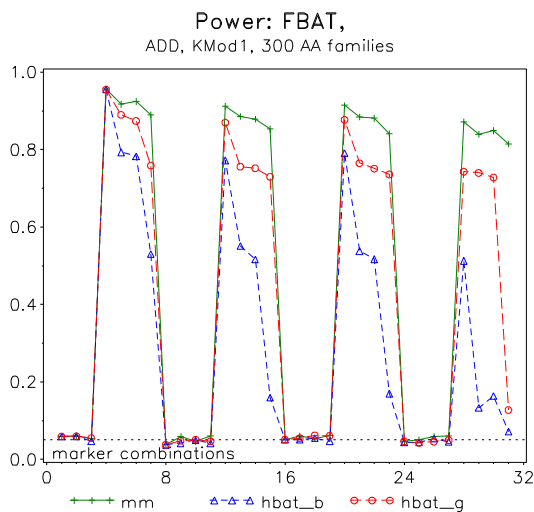
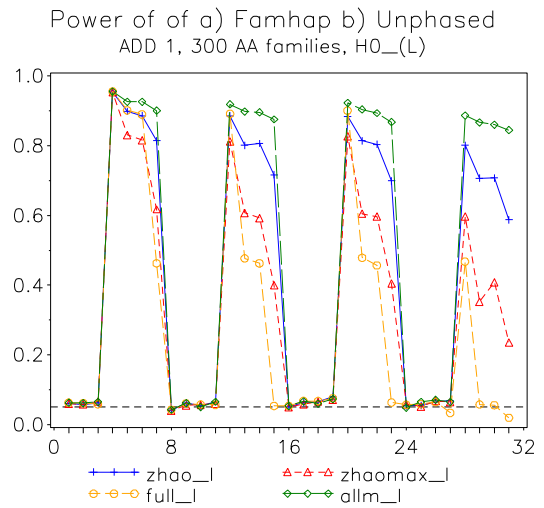
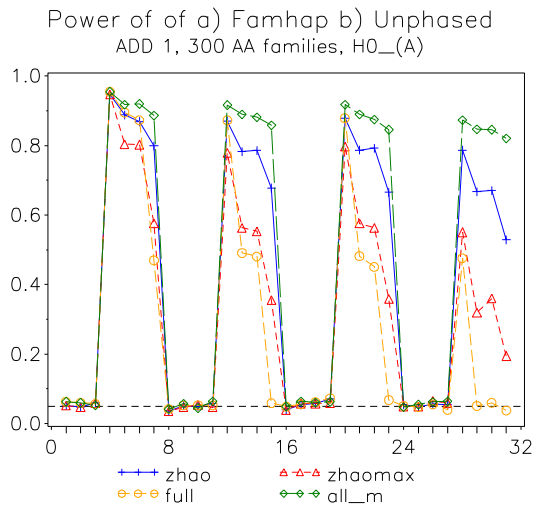
Power: Hotel\_Fam  
REZ, KMod1, 2500 trios



# Power für alle Markerkombinationen, $\alpha = 0.05$

## KMod 1, AA families

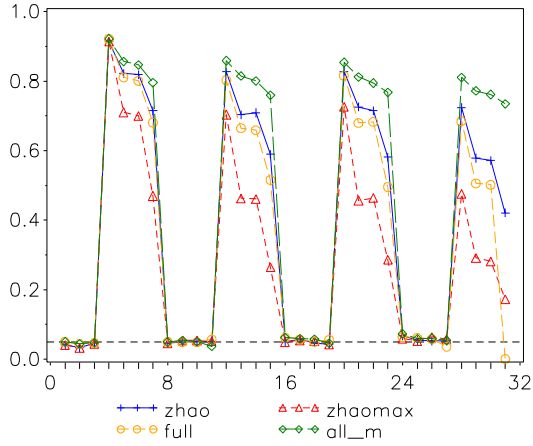
### Additiver Vererbungsmodus



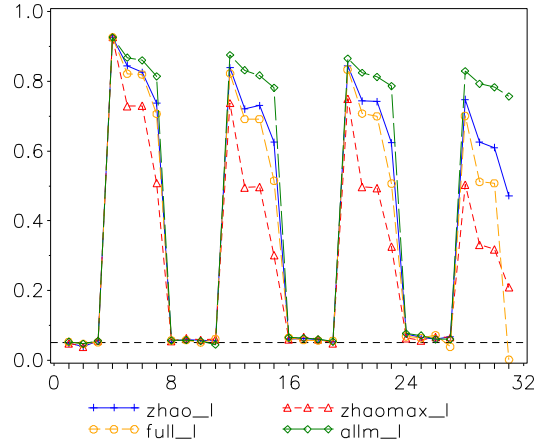
# KMod 1, AA families

## Dominanter Vererbungsmodus

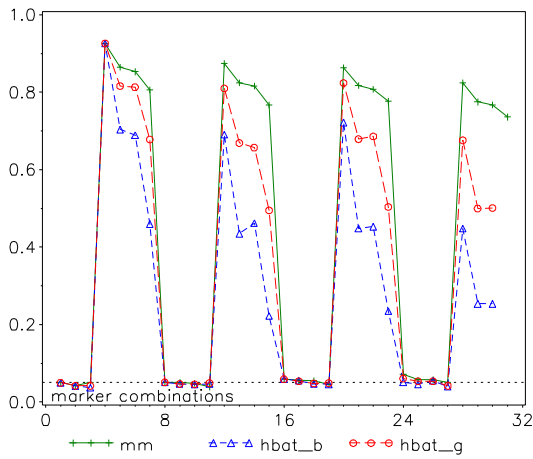
Power of of a) Famhap b) Unphased  
DOM 1, 300 AA families, H0\_(A)



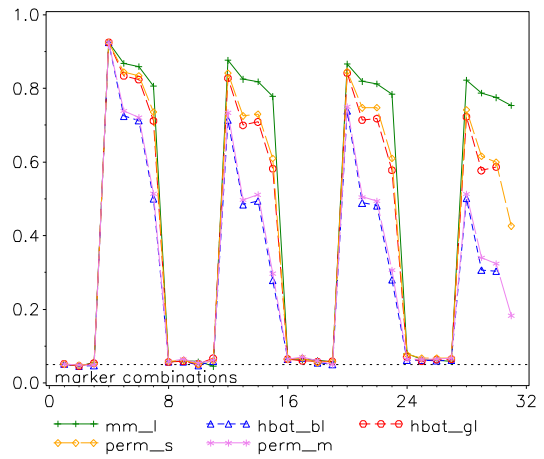
Power of of a) Famhap b) Unphased  
DOM 1, 300 AA families, H0\_(L)



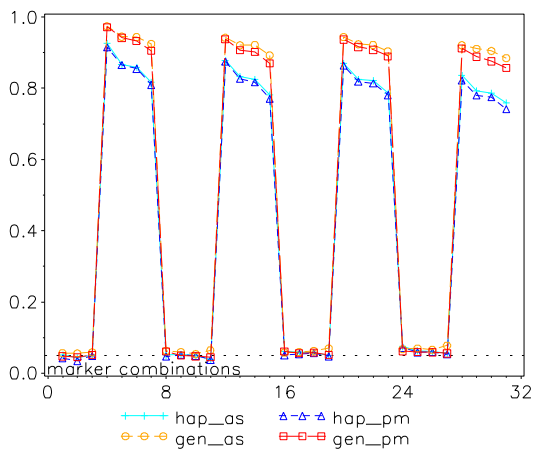
Power: FBAT,  
DOM, KMod1, 300 AA families



Power: FBAT, Linkage  
DOM, KMod1, 300 AA families



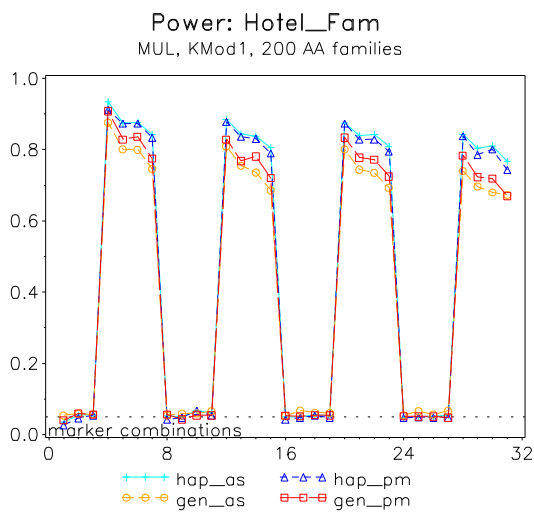
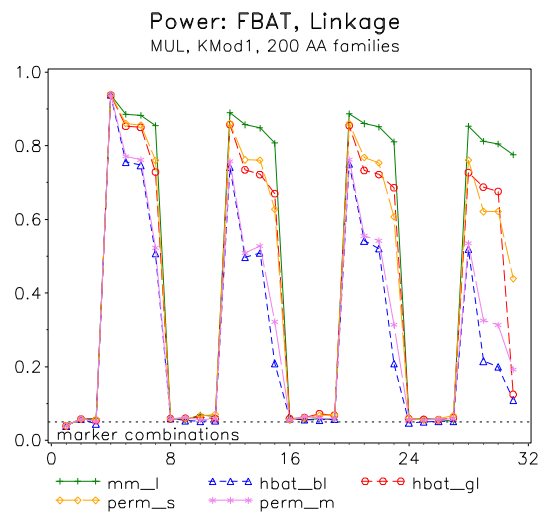
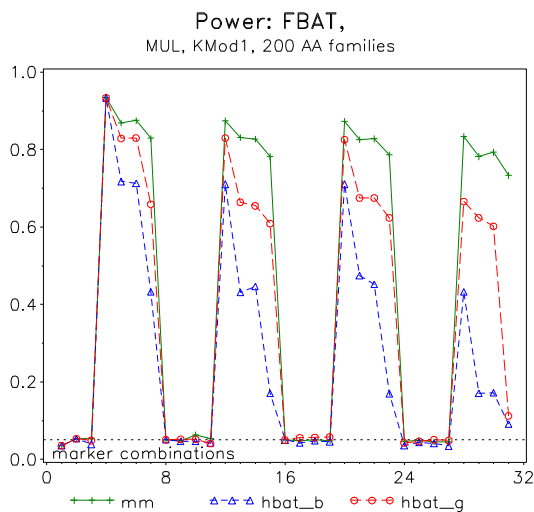
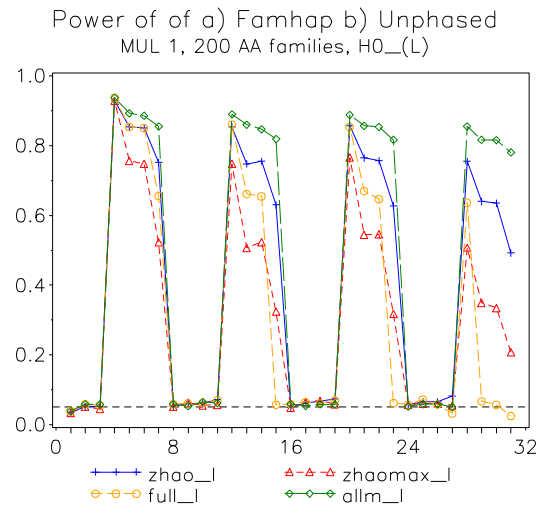
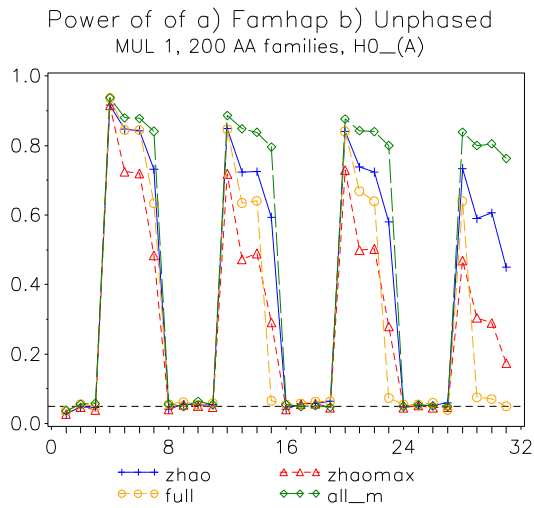
Power: Hotel\_Fam  
DOM, KMod1, 300 AA families



# Power für alle Markerkombinationen, $\alpha = 0.05$

## KMod 1, AA families

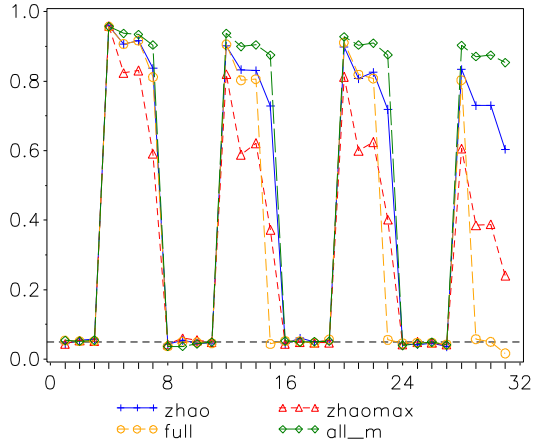
### Multiplikativer Vererbungsmodus



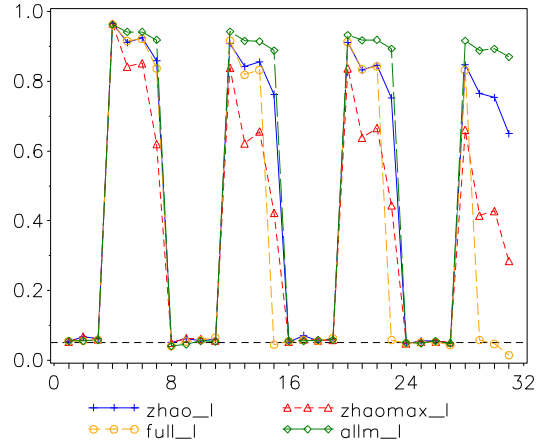
# KMod 1, AA families

## Rezessiver Vererbungsmodus

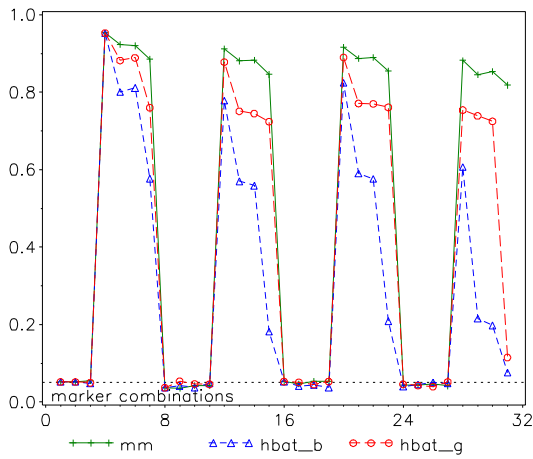
Power of of a) Famhap b) Unphased  
REZ 1, 500 AA families, H0\_(A)



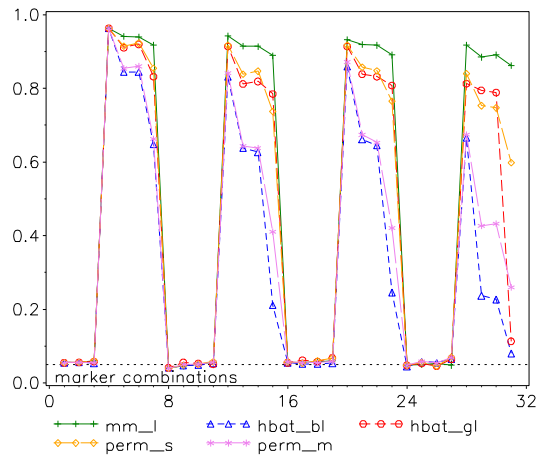
Power of of a) Famhap b) Unphased  
REZ 1, 500 AA families, H0\_(L)



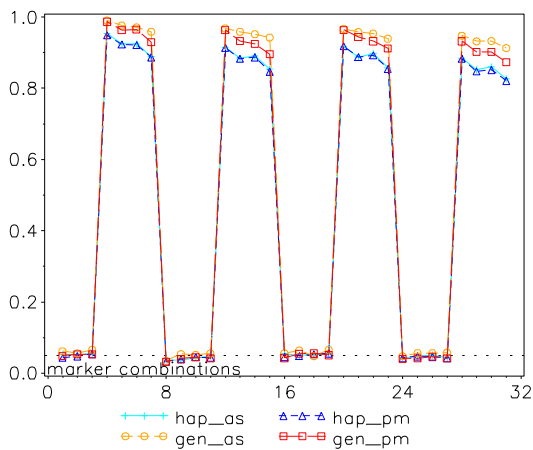
Power: FBAT,  
REZ, KMod1, 500 AA families



Power: FBAT, Linkage  
REZ, KMod1, 500 AA families



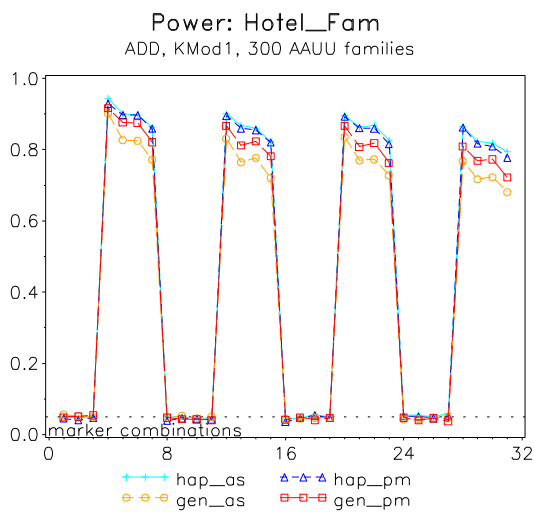
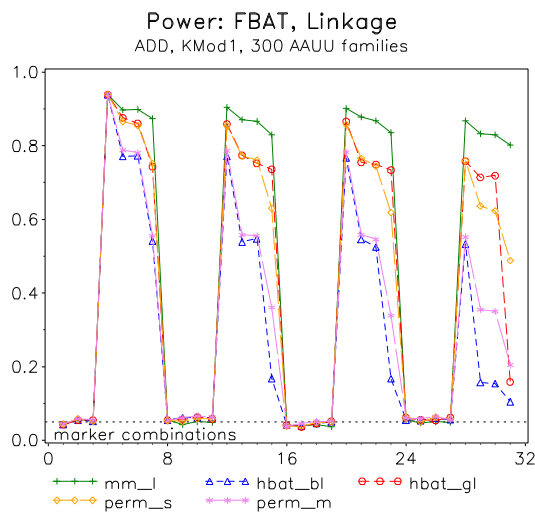
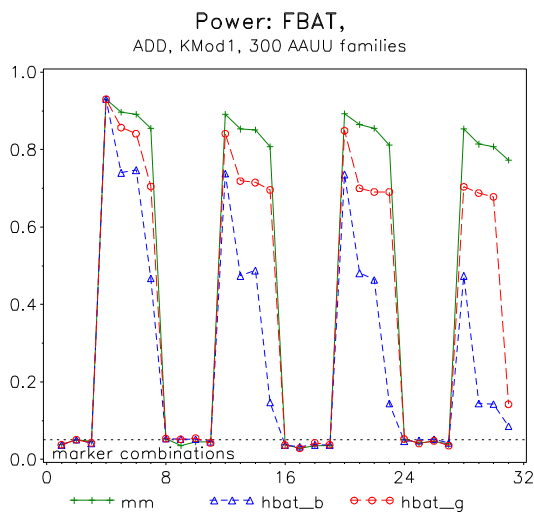
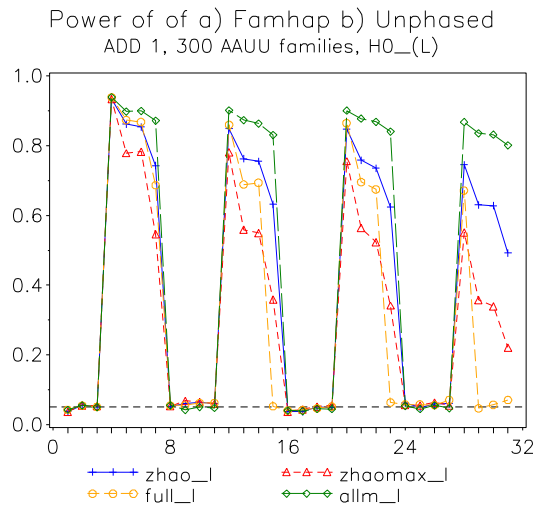
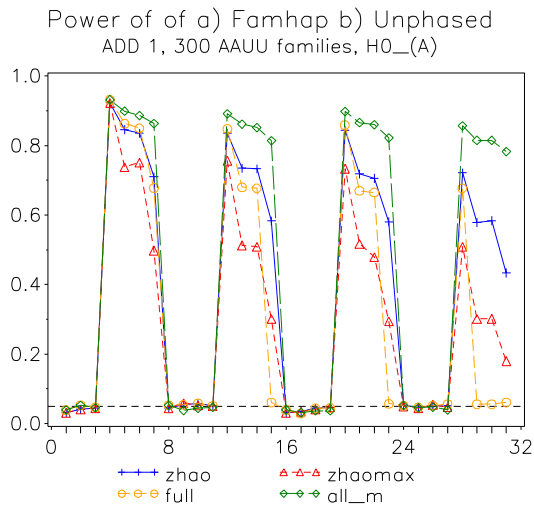
Power: Hotel\_Fam  
REZ, KMod1, 500 AA families



# Power für alle Markerkombinationen, $\alpha = 0.05$

## KMod 1, AAUU families

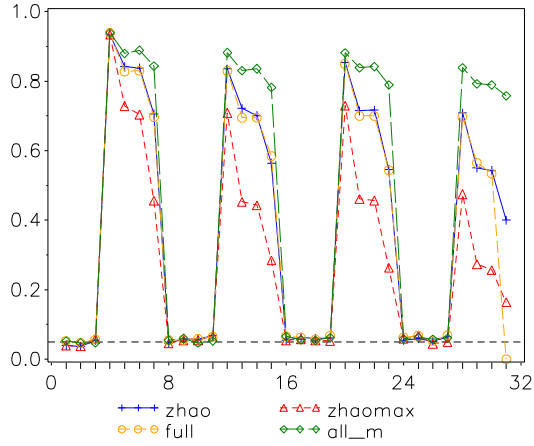
### Additiver Vererbungsmodus



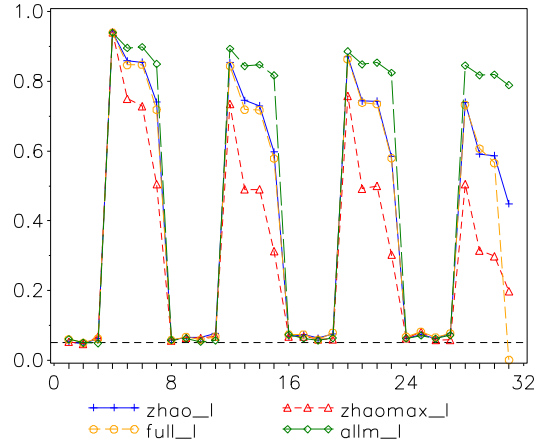
# KMod 1, AAU families

## Dominanter Vererbungsmodus

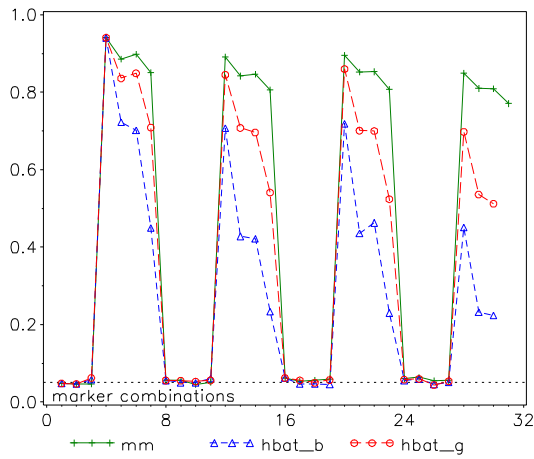
Power of of a) Famhap b) Unphased  
DOM 1, 300 AAU families, H0\_(A)



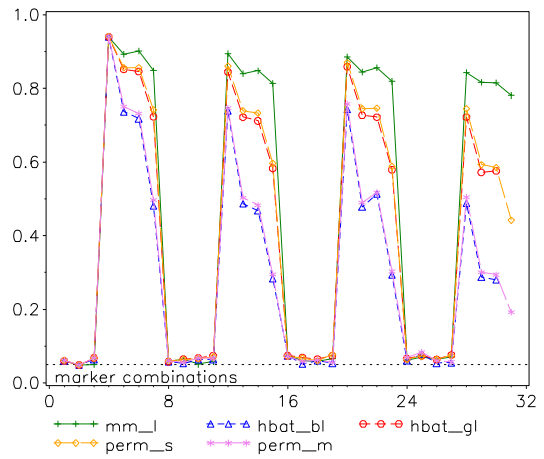
Power of of a) Famhap b) Unphased  
DOM 1, 300 AAU families, H0\_(L)



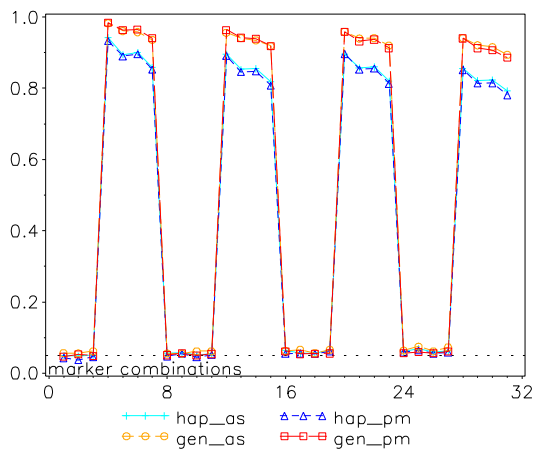
Power: FBAT,  
DOM, KMod1, 300 AAU families



Power: FBAT, Linkage  
DOM, KMod1, 300 AAU families



Power: Hotel\_Fam  
DOM, KMod1, 300 AAU families

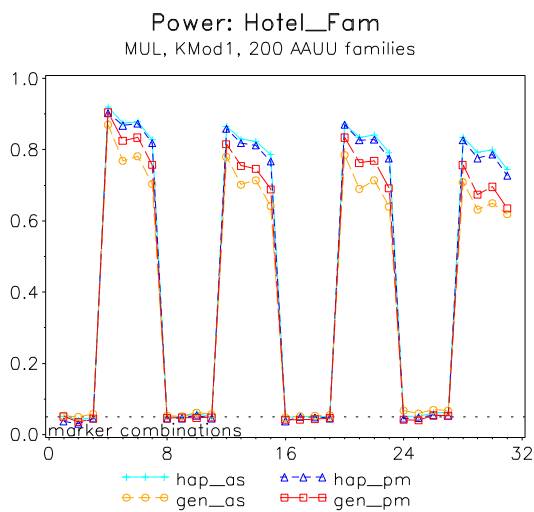
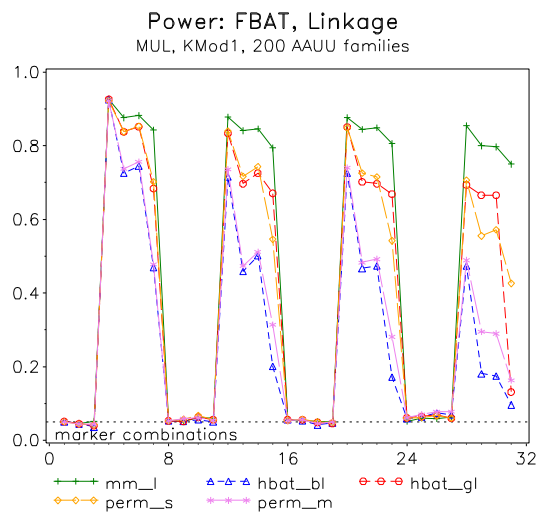
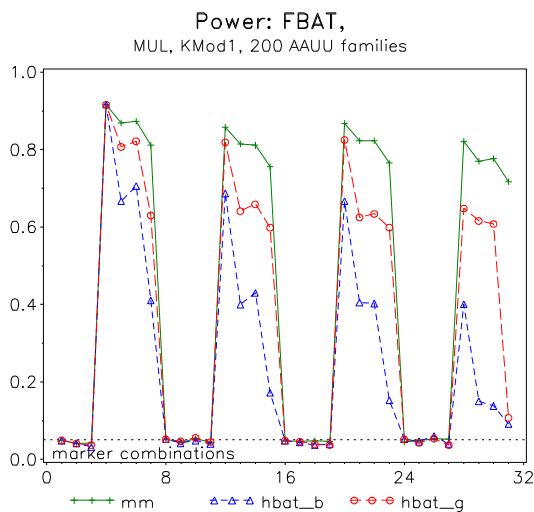
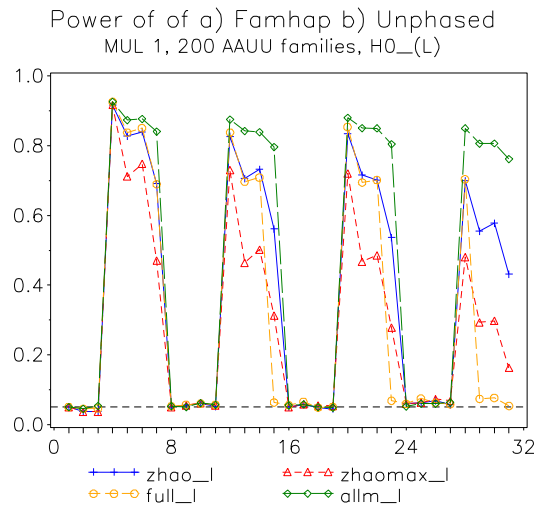
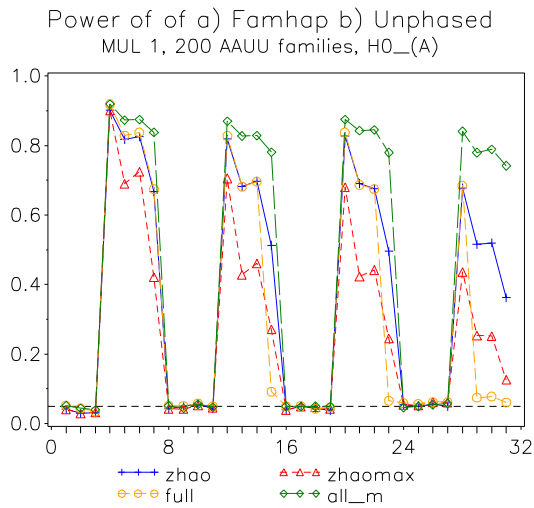




# Power für alle Markerkombinationen, $\alpha = 0.05$

## KMod 1, AAUU families

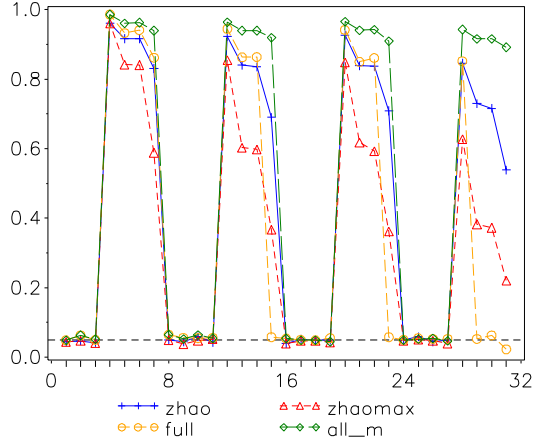
### Multiplikativer Vererbungsmodus



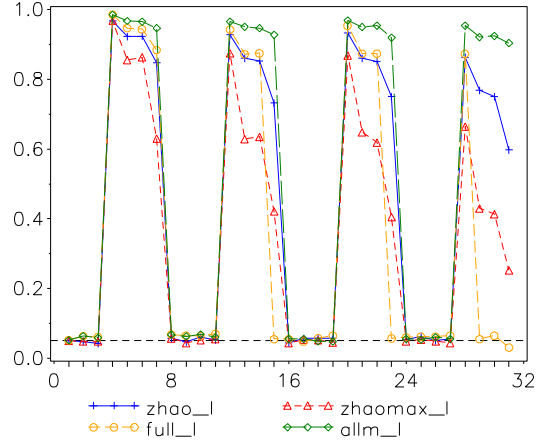
# KMod 1, AAU families

## Rezessiver Vererbungsmodus

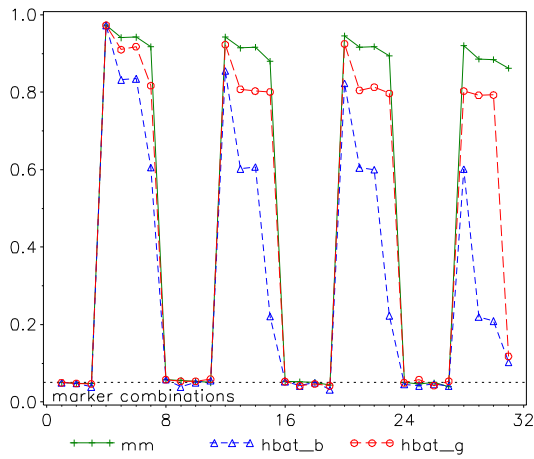
Power of of a) Famhap b) Unphased  
 REZ 1, 600 AAU families, H0\_(A)



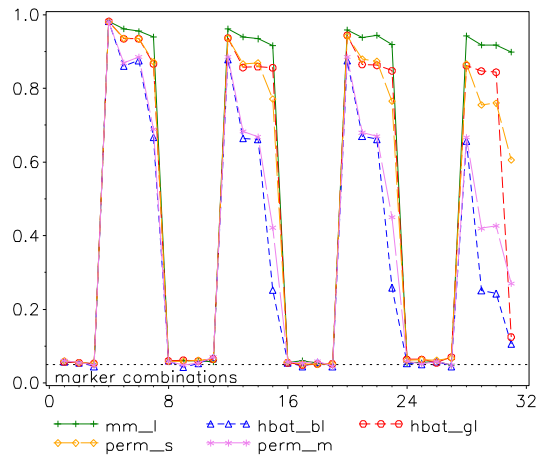
Power of of a) Famhap b) Unphased  
 REZ 1, 600 AAU families, H0\_(L)



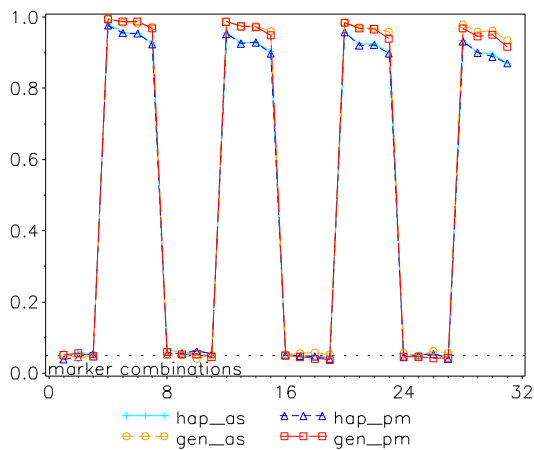
Power: FBAT,  
 REZ, KMod1, 600 AAU families



Power: FBAT, Linkage  
 REZ, KMod1, 600 AAU families



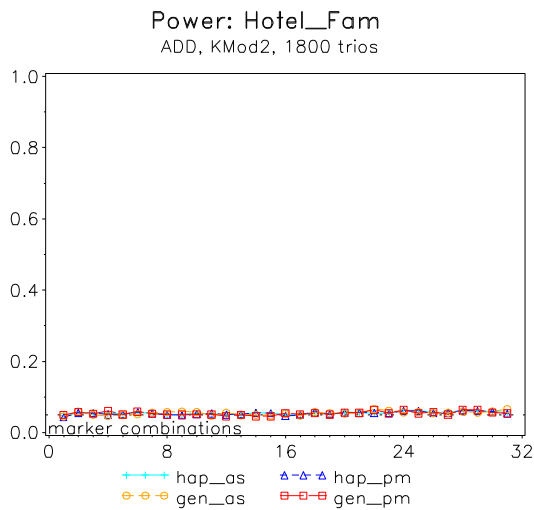
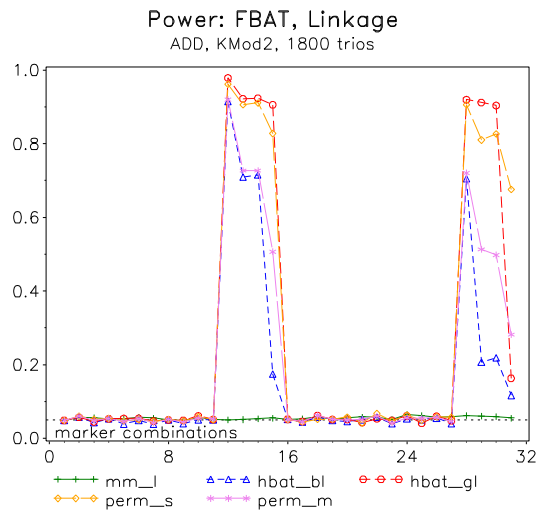
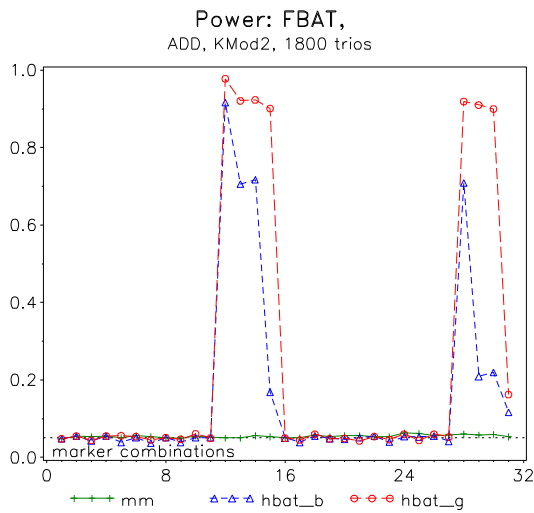
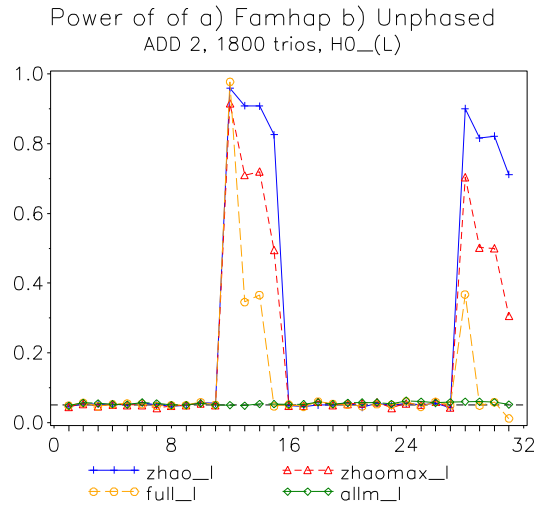
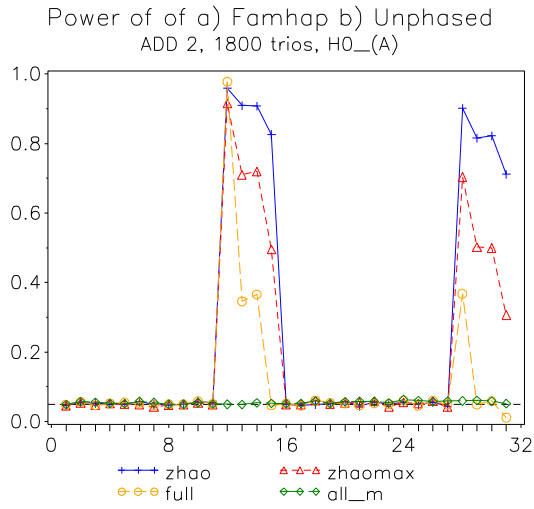
Power: Hotel\_Fam  
 REZ, KMod1, 600 AAU families



# Power für alle Markerkombinationen, $\alpha = 0.05$

## KMod 2

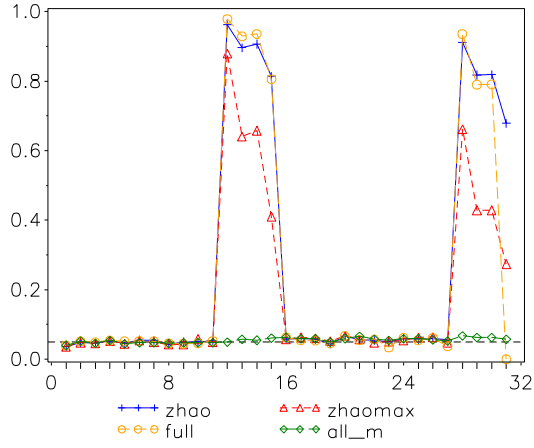
### Additiver Vererbungsmodus



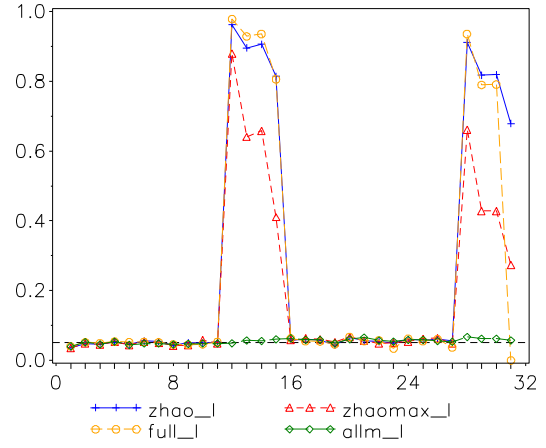
## KMod 2, trios

### Dominanter Vererbungsmodus

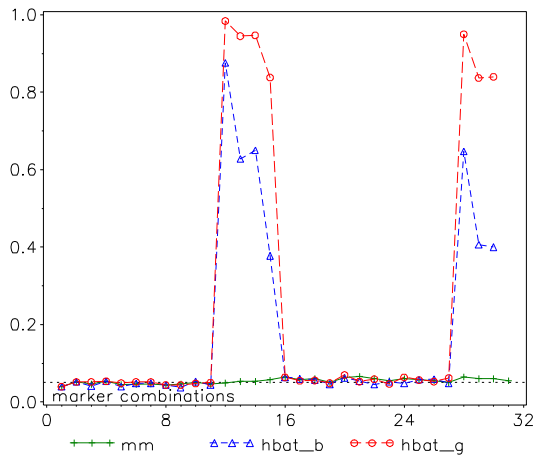
Power of of a) Famhap b) Unphased  
DOM 2, 1000 trios, H0\_(A)



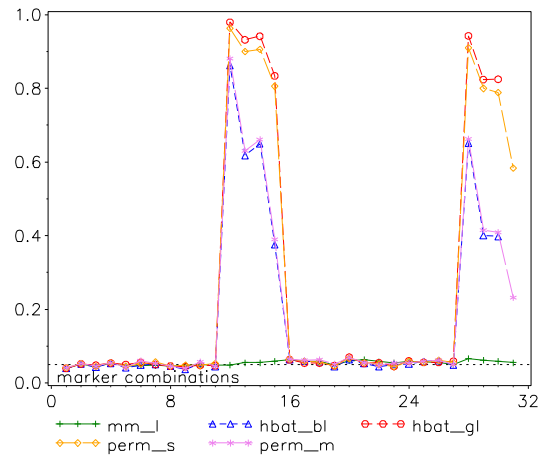
Power of of a) Famhap b) Unphased  
DOM 2, 1000 trios, H0\_(L)



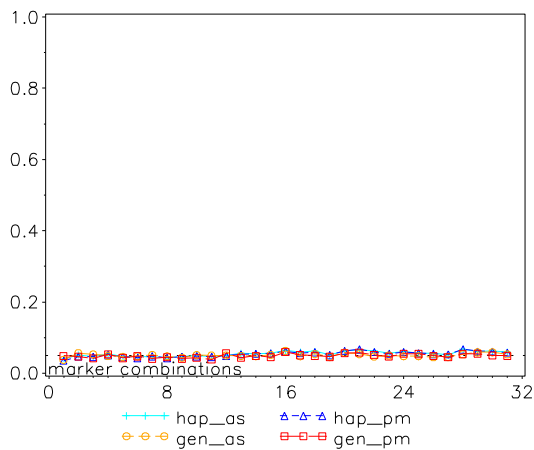
Power: FBAT,  
DOM, KMod2, 1000 trios



Power: FBAT, Linkage  
DOM, KMod2, 1000 trios



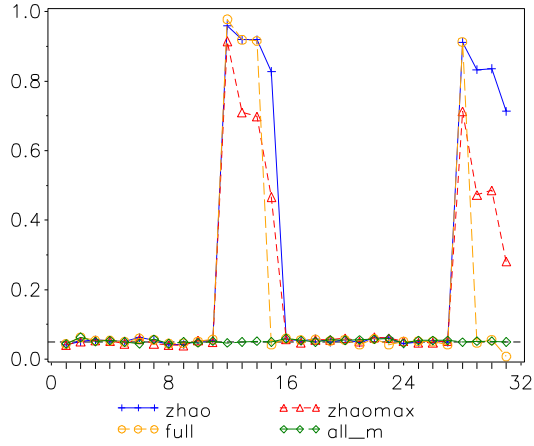
Power: Hotel\_Fam  
DOM, KMod2, 1000 trios



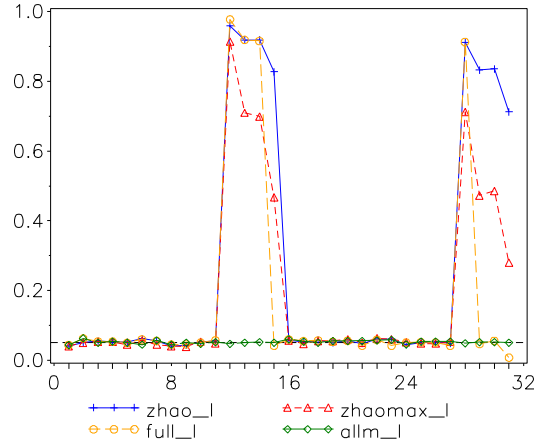
## KMod 2, trios

### Multiplikativer Vererbungsmodus

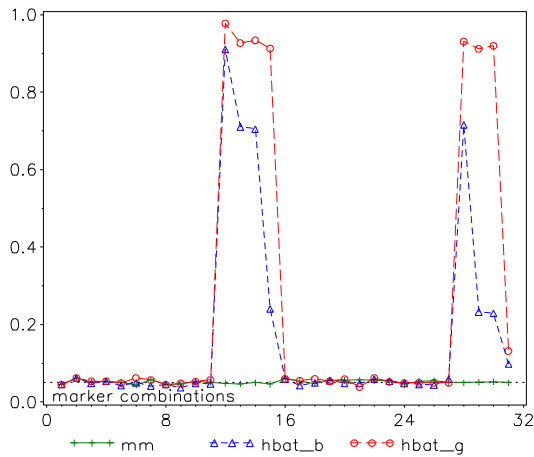
Power of of a) Famhap b) Unphased  
MUL 2, 1200 trios, H0\_(A)



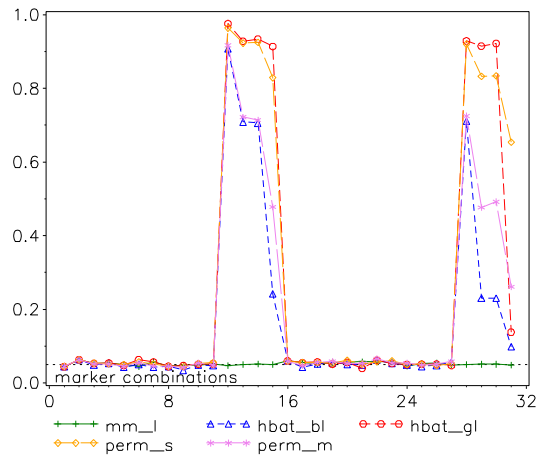
Power of of a) Famhap b) Unphased  
MUL 2, 1200 trios, H0\_(L)



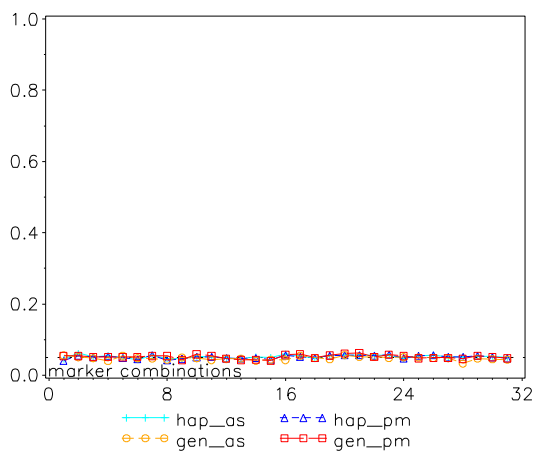
Power: FBAT,  
MUL, KMod2, 1200 trios



Power: FBAT, Linkage  
MUL, KMod2, 1200 trios



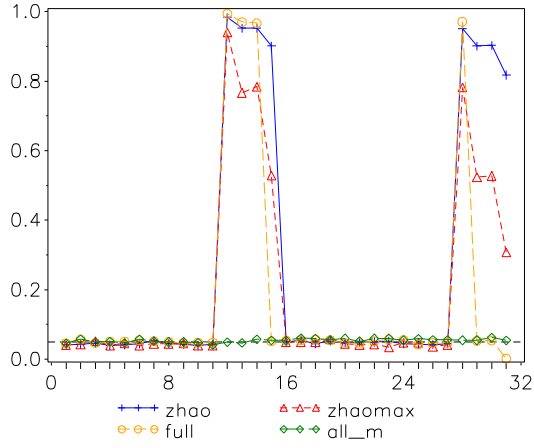
Power: Hotel\_Fam  
MUL, KMod2, 1200 trios



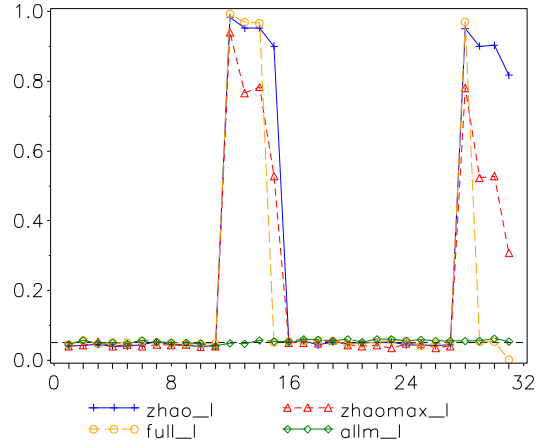
# KMod 2, trios

## Rezessiver Vererbungsmodus

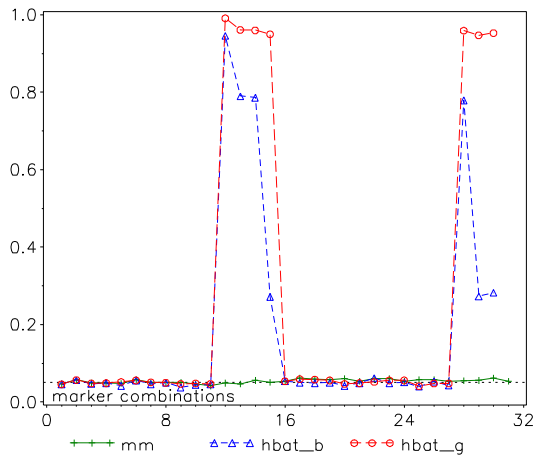
Power of of a) Famhap b) Unphased  
REZ 2, 6000 trios, H0\_(A)



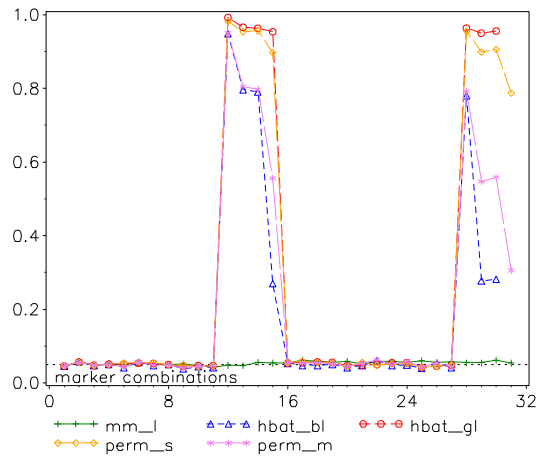
Power of of a) Famhap b) Unphased  
REZ 2, 6000 trios, H0\_(L)



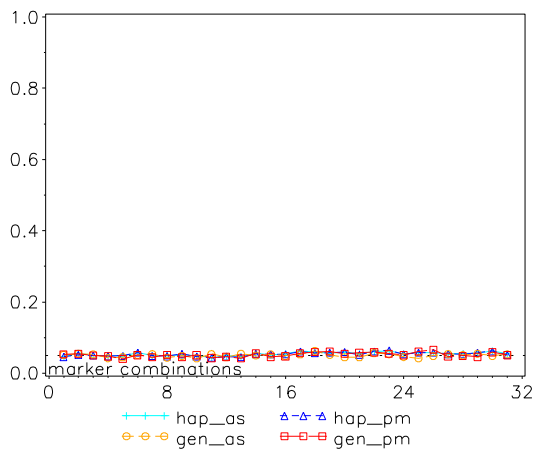
Power: FBAT,  
REZ, KMod2, 6000 trios



Power: FBAT, Linkage  
REZ, KMod2, 6000 trios



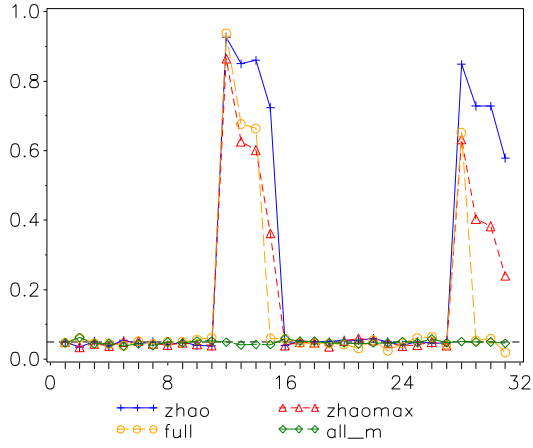
Power: Hotel\_Fam  
REZ, KMod2, 6000 trios



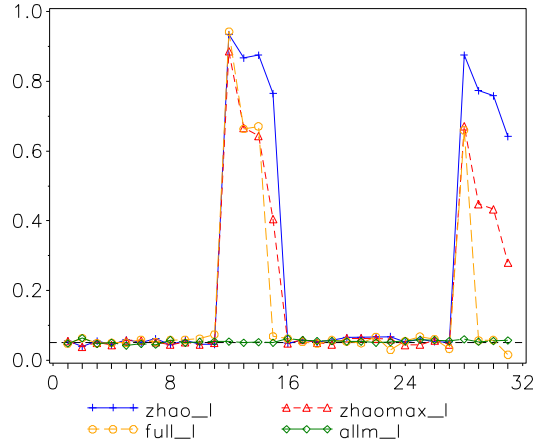
# KMod 2, AA families

## Additiver Vererbungsmodus

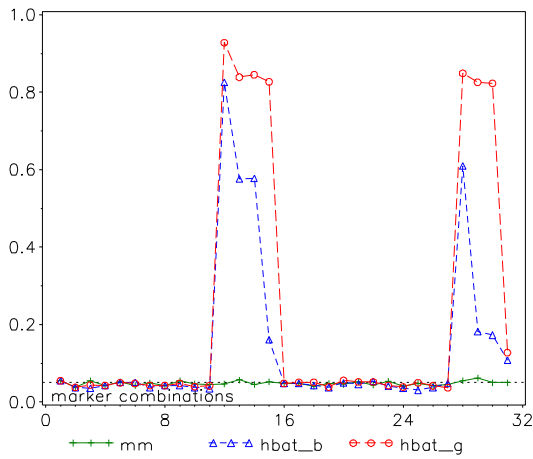
Power of of a) Famhap b) Unphased  
ADD 2, 400 AA families, H0\_(A)



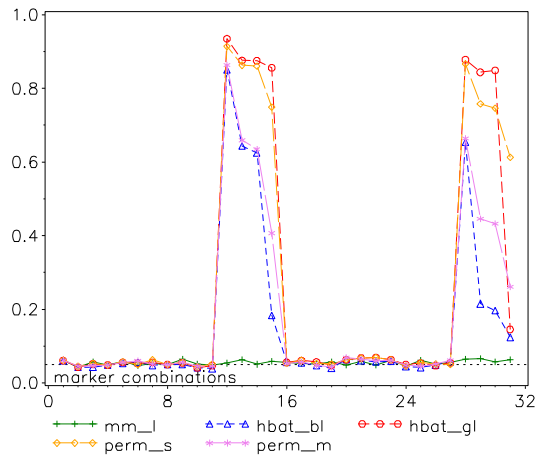
Power of of a) Famhap b) Unphased  
ADD 2, 400 AA families, H0\_(L)



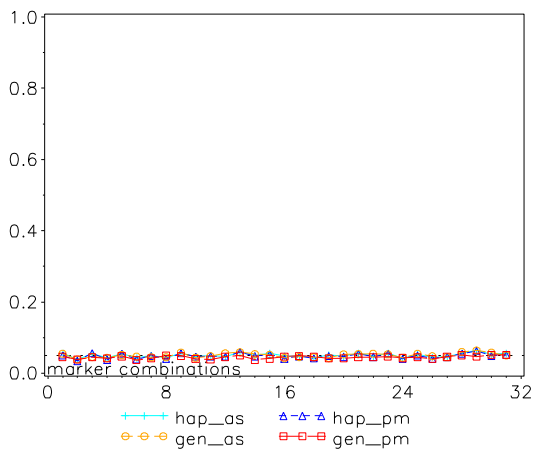
Power: FBAT,  
ADD, KMod2, 400 AA families



Power: FBAT, Linkage  
ADD, KMod2, 400 AA families



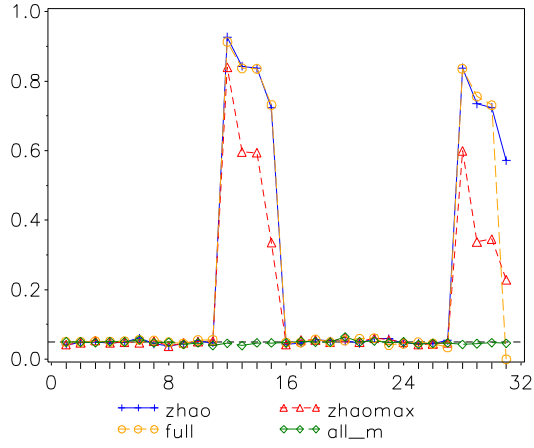
Power: Hotel\_Fam  
ADD, KMod2, 400 AA families



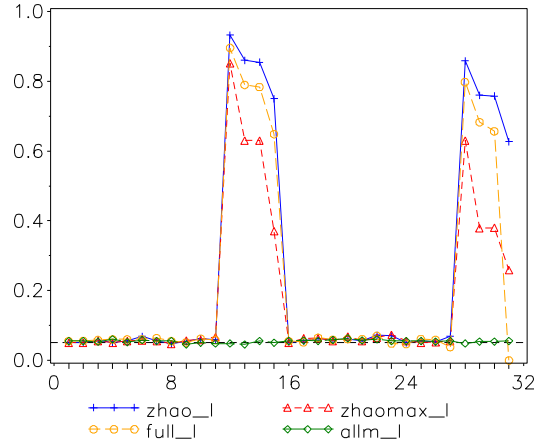
## KMod 2, AA families

### Dominanter Vererbungsmodus

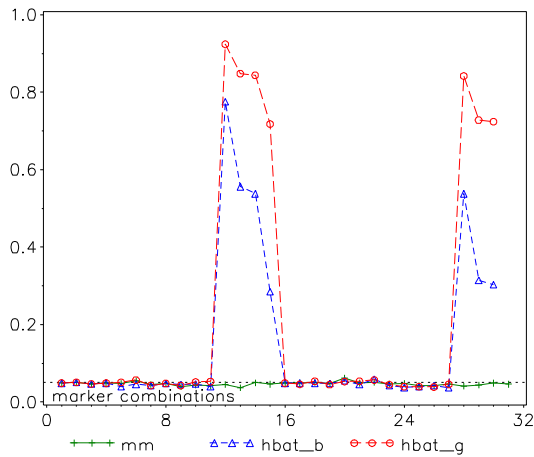
Power of of a) Famhap b) Unphased  
DOM 2, 500 AA families, H0\_(A)



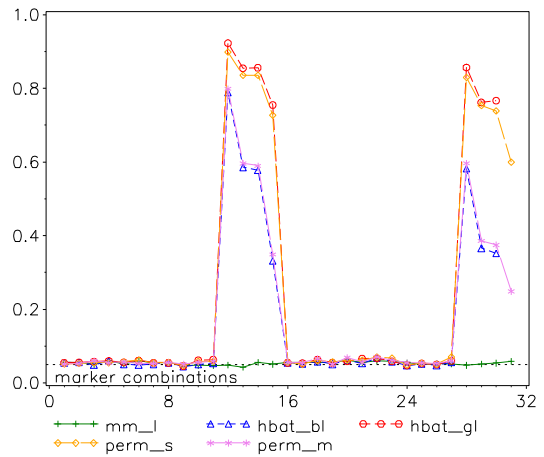
Power of of a) Famhap b) Unphased  
DOM 2, 500 AA families, H0\_(L)



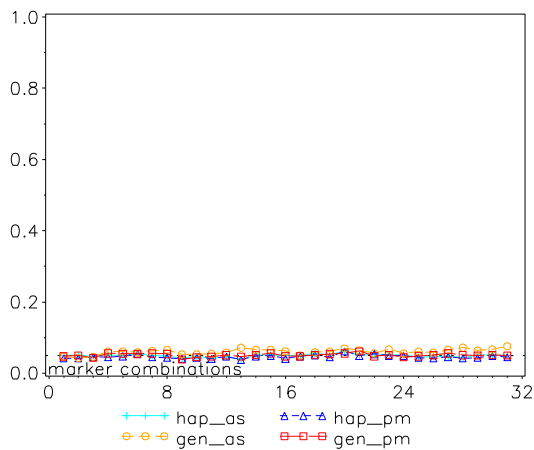
Power: FBAT,  
DOM, KMod2, 500 AA families



Power: FBAT, Linkage  
DOM, KMod2, 500 AA families



Power: Hotel\_Fam  
DOM, KMod2, 500 AA families

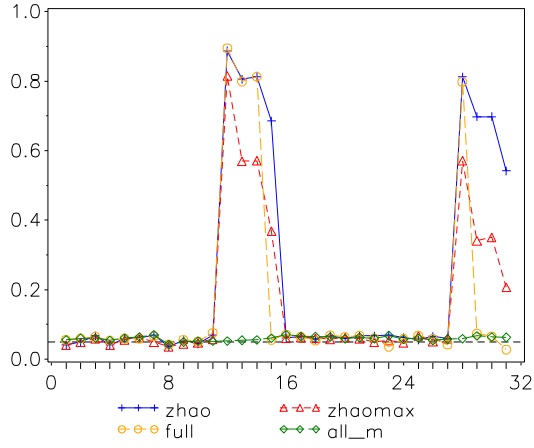




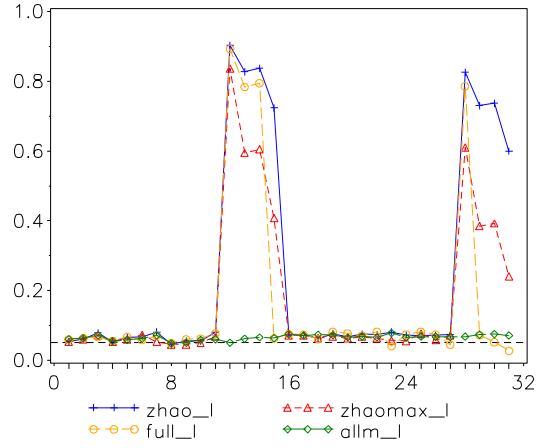
## KMod 2, AA families

### Multiplikativer Vererbungsmodus

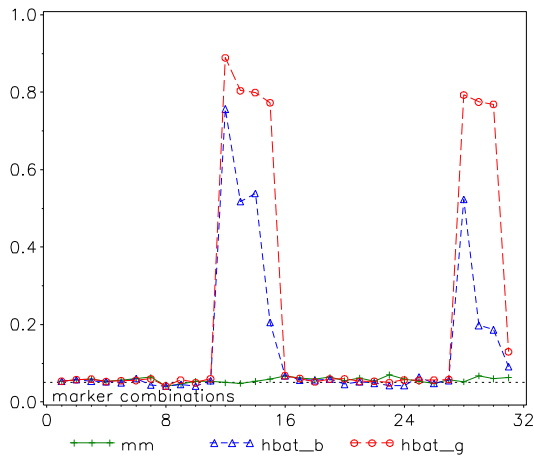
Power of of a) Famhap b) Unphased  
MUL 2, 300 AA families, H0\_(A)



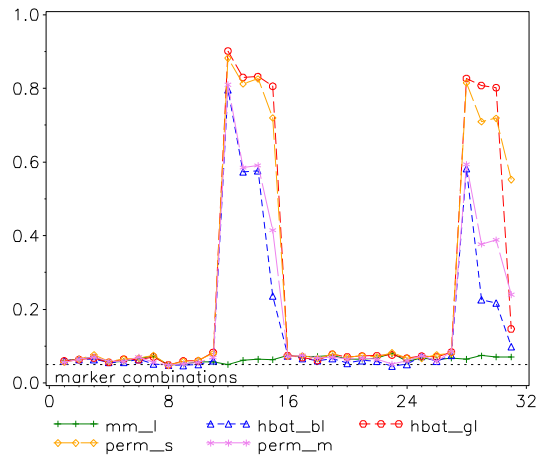
Power of of a) Famhap b) Unphased  
MUL 2, 300 AA families, H0\_(L)



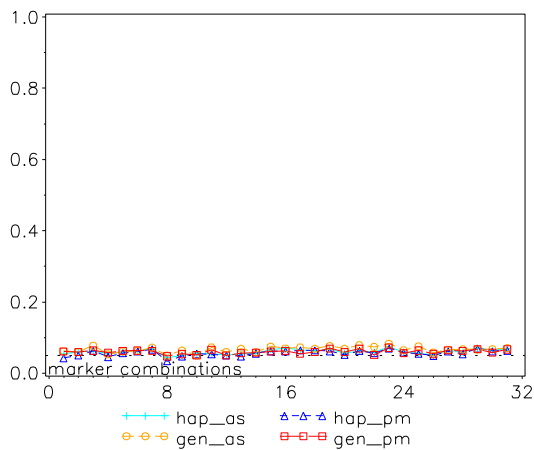
Power: FBAT,  
MUL, KMod2, 300 AA families



Power: FBAT, Linkage  
MUL, KMod2, 300 AA families



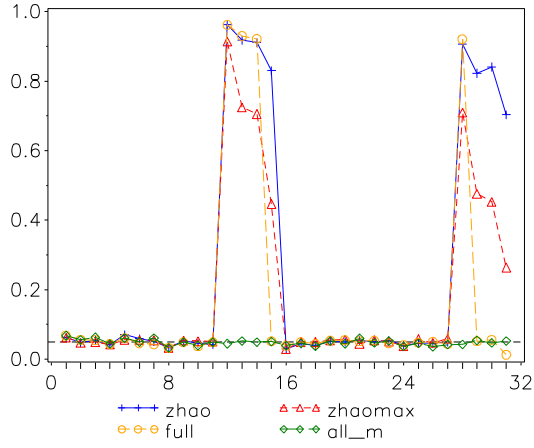
Power: Hotel\_Fam  
MUL, KMod2, 300 AA families



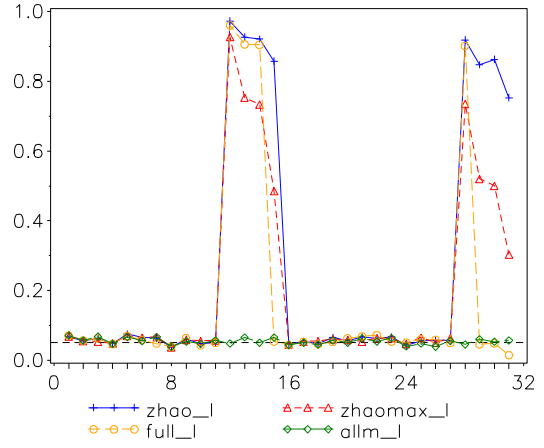
## KMod 2, AA families

### Rezessiver Vererbungsmodus

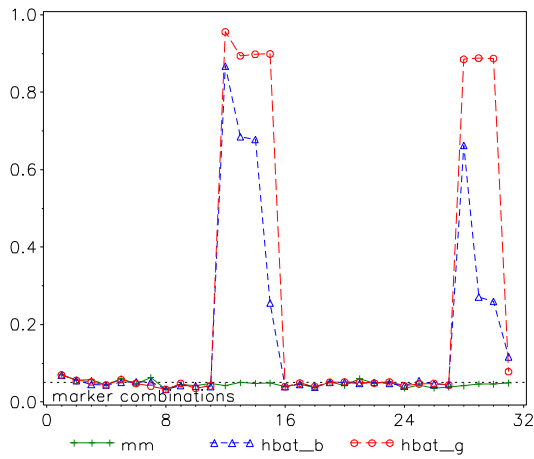
Power of of a) Famhap b) Unphased  
REZ 2, 900 AA families, H0\_(A)



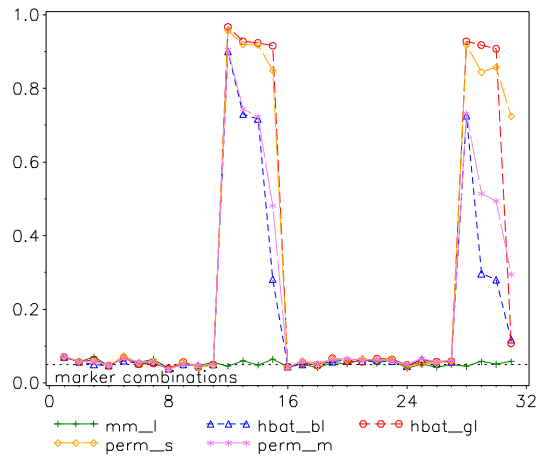
Power of of a) Famhap b) Unphased  
REZ 2, 900 AA families, H0\_(L)



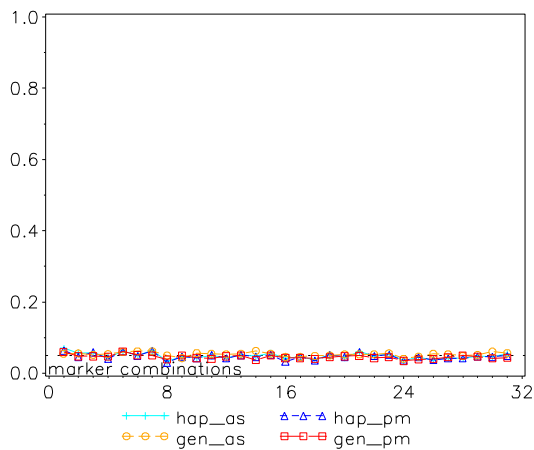
Power: FBAT,  
REZ, KMod2, 900 AA families



Power: FBAT, Linkage  
REZ, KMod2, 900 AA families



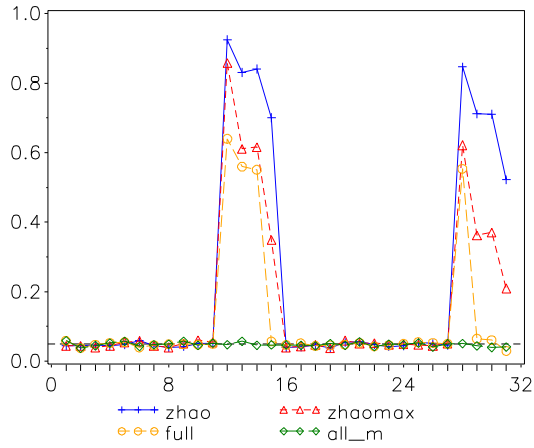
Power: Hotel\_Fam  
REZ, KMod2, 900 AA families



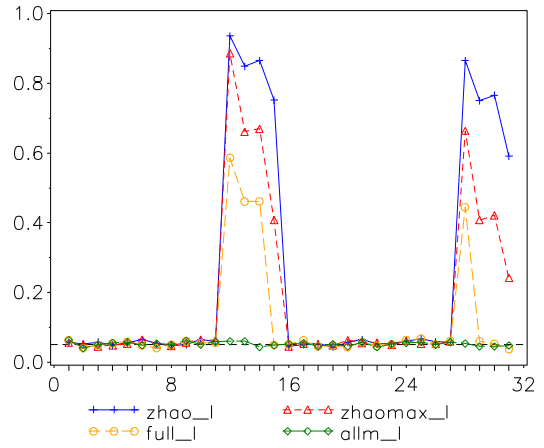
## KMod 2, AAU families

### Additiver Vererbungsmodus

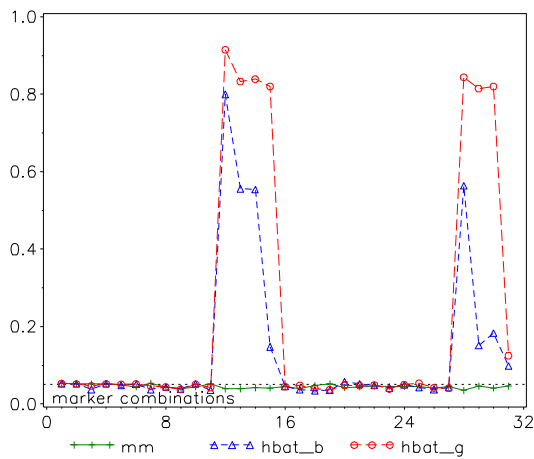
Power of of a) Famhap b) Unphased  
ADD 2, 400 AAU families, HO\_(A)



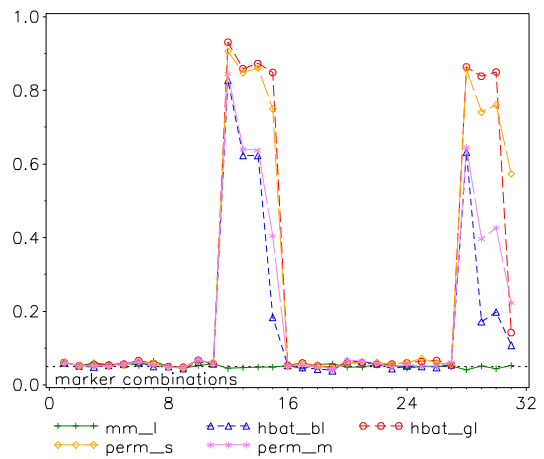
Power of of a) Famhap b) Unphased  
ADD 2, 400 AAU families, HO\_(L)



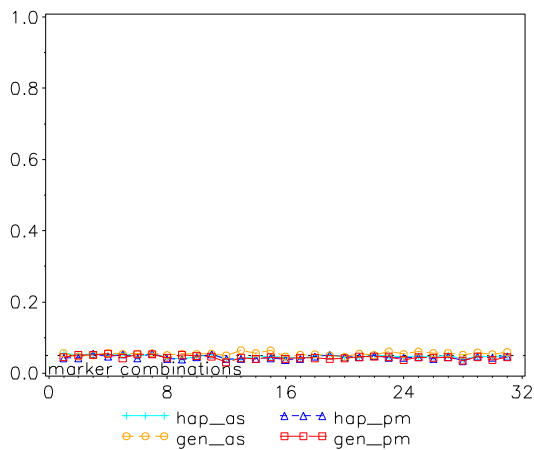
Power: FBAT,  
ADD, KMod2, 400 AAU families



Power: FBAT, Linkage  
ADD, KMod2, 400 AAU families



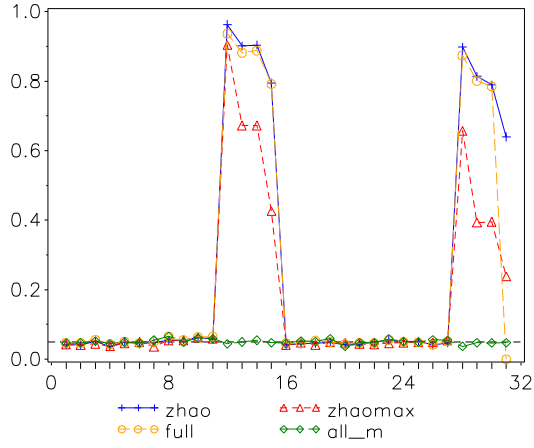
Power: Hotel\_Fam  
ADD, KMod2, 400 AAU families



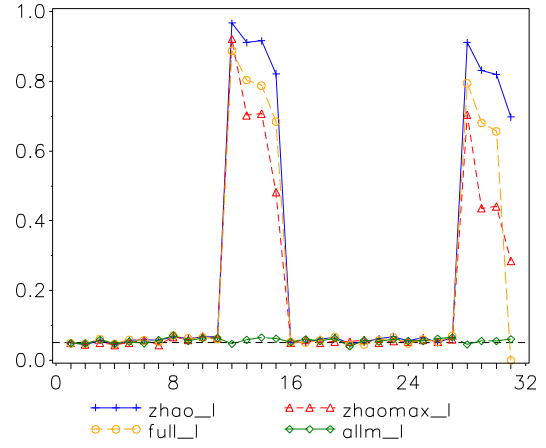
## KMod 2, AAU families

### Dominanter Vererbungsmodus

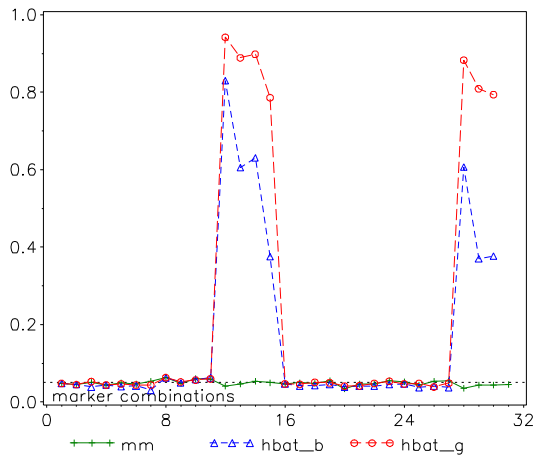
Power of of a) Famhap b) Unphased  
DOM 2, 500 AAU families, H0\_(A)



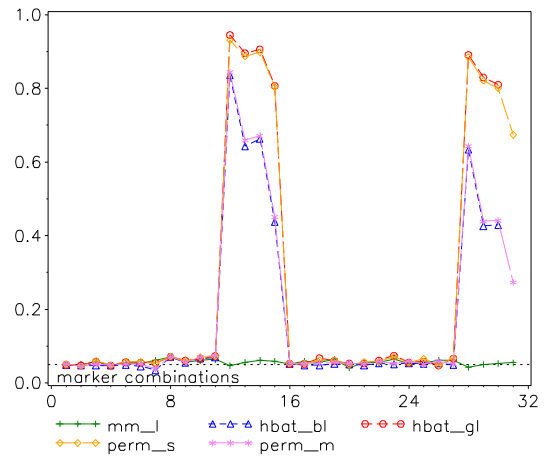
Power of of a) Famhap b) Unphased  
DOM 2, 500 AAU families, H0\_(L)



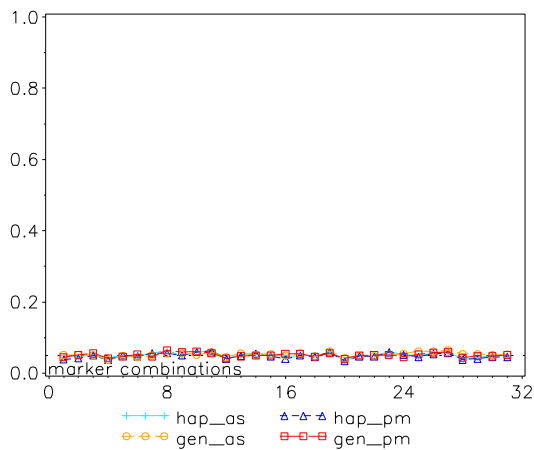
Power: FBAT,  
DOM, KMod2, 500 AAU families



Power: FBAT, Linkage  
DOM, KMod2, 500 AAU families



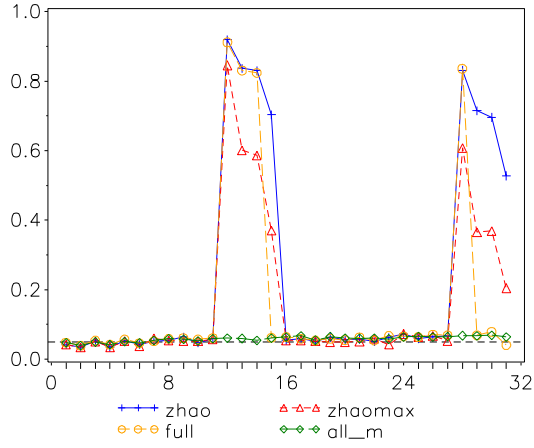
Power: Hotel\_Fam  
DOM, KMod2, 500 AAU families



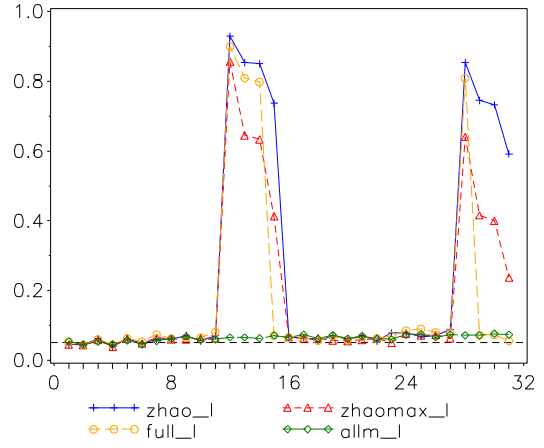
## KMod 2, AAU families

### Multiplikativer Vererbungsmodus

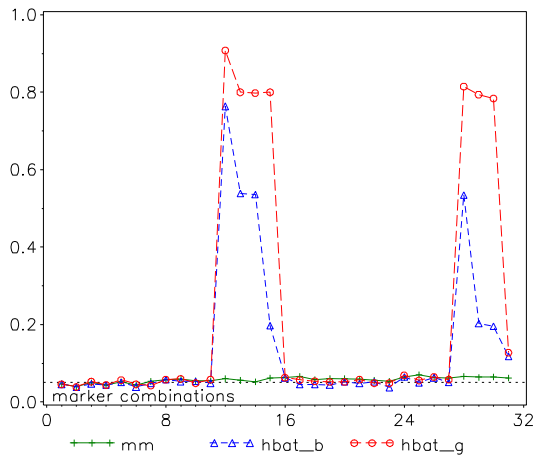
Power of of a) Famhap b) Unphased  
MUL 2, 300 AAU families, H0\_(A)



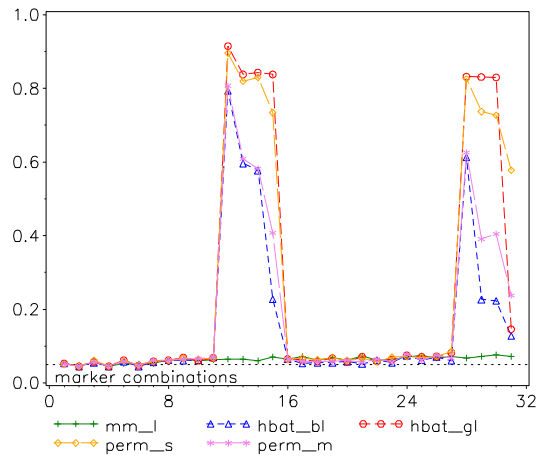
Power of of a) Famhap b) Unphased  
MUL 2, 300 AAU families, H0\_(L)



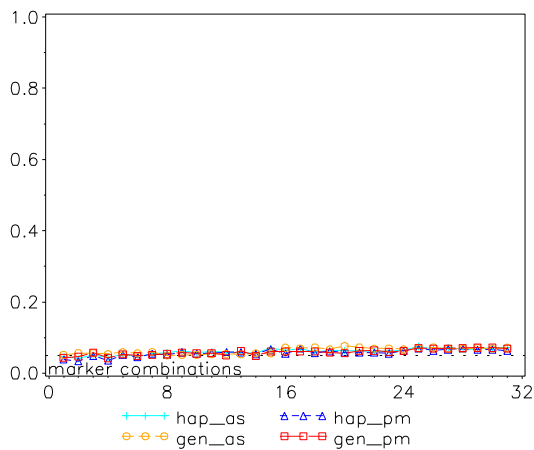
Power: FBAT,  
MUL, KMod2, 300 AAU families



Power: FBAT, Linkage  
MUL, KMod2, 300 AAU families



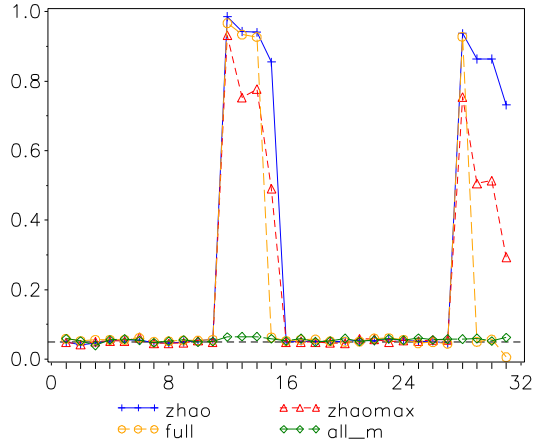
Power: Hotel\_Fam  
MUL, KMod2, 300 AAU families



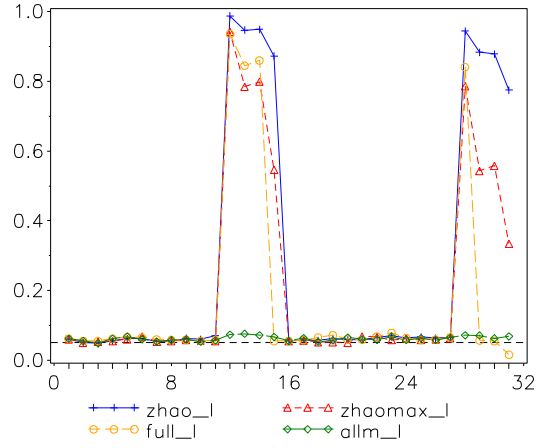
## KMod 2, AAU families

### Rezessiver Vererbungsmodus

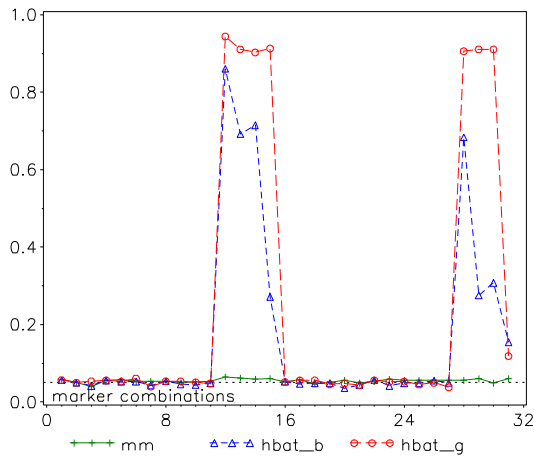
Power of of a) Famhap b) Unphased  
REZ 2, 900 AAU families, H0\_(A)



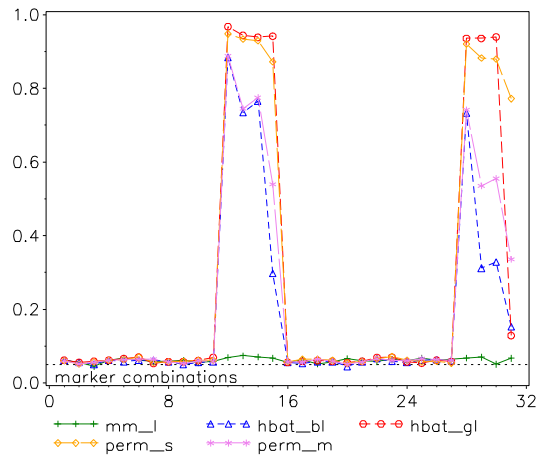
Power of of a) Famhap b) Unphased  
REZ 2, 900 AAU families, H0\_(L)



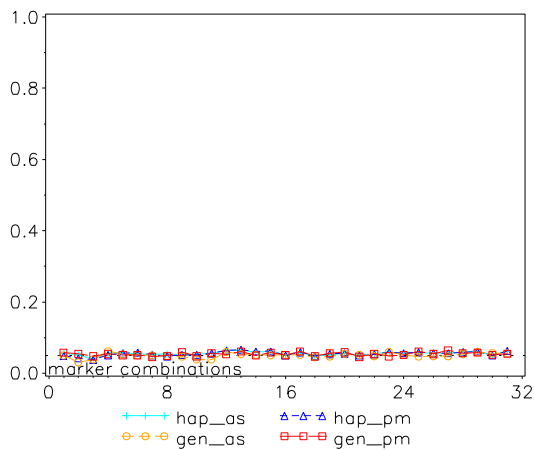
Power: FBAT,  
REZ, KMod2, 900 AAU families



Power: FBAT, Linkage  
REZ, KMod2, 900 AAU families



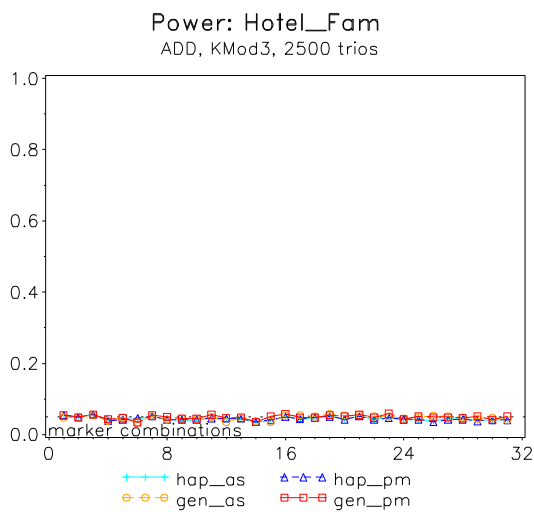
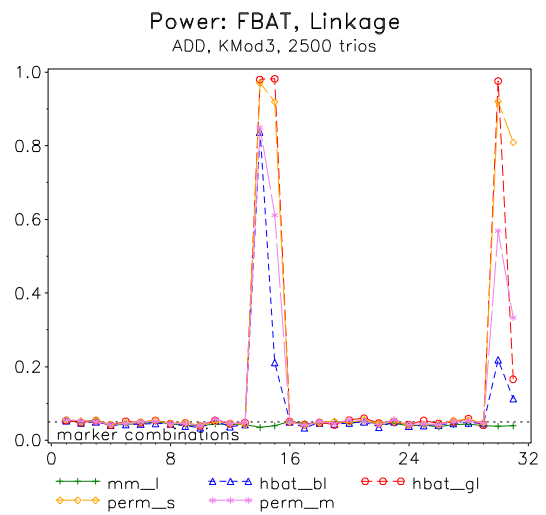
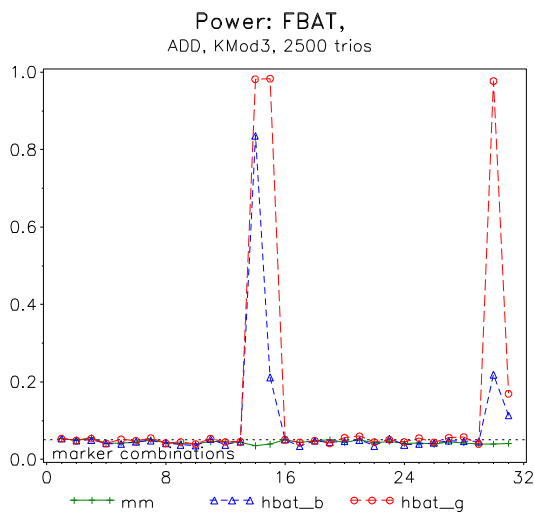
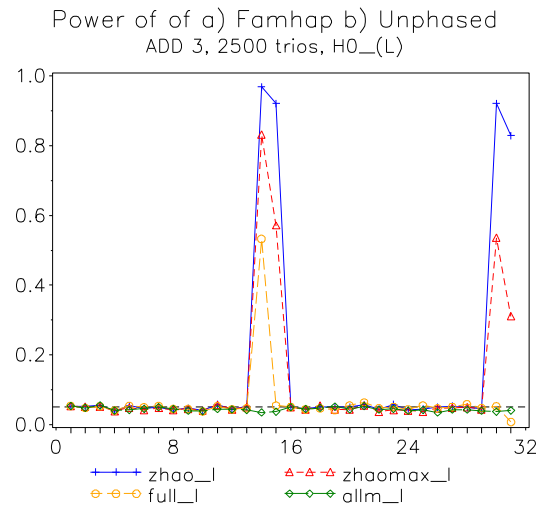
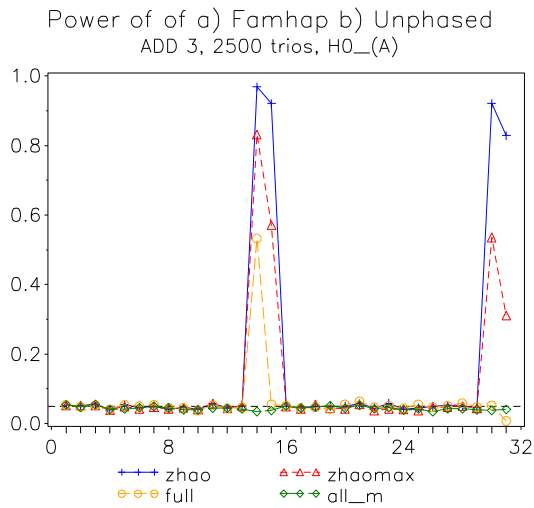
Power: Hotel\_Fam  
REZ, KMod2, 900 AAU families



# Power für alle Markerkombinationen, $\alpha = 0.05$

## KMod 3, Trios

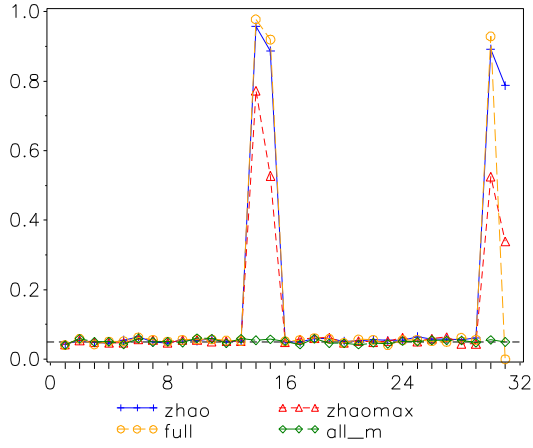
### Additiver Vererbungsmodus



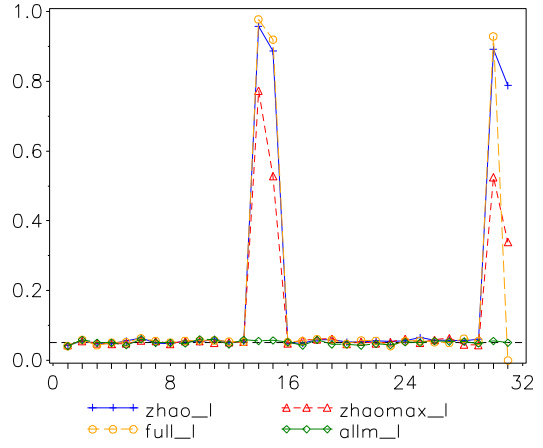
# KMod 3, Trios

## Dominanter Vererbungsmodus

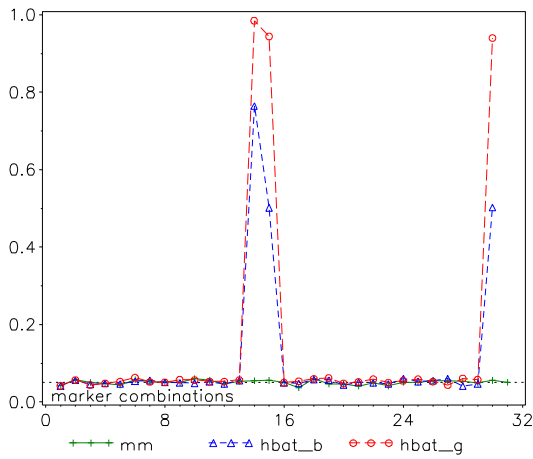
Power of of a) Famhap b) Unphased  
DOM 3, 1500 trios, H0\_(A)



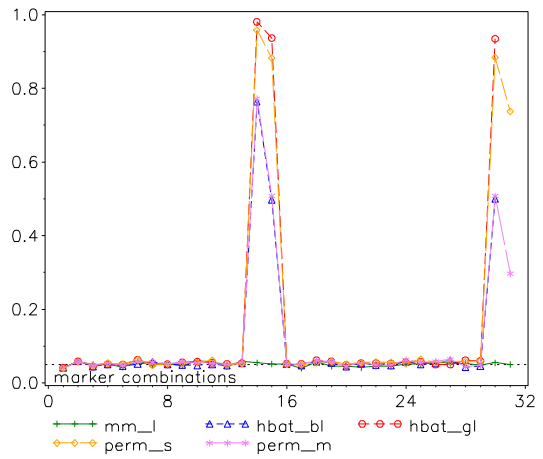
Power of of a) Famhap b) Unphased  
DOM 3, 1500 trios, H0\_(L)



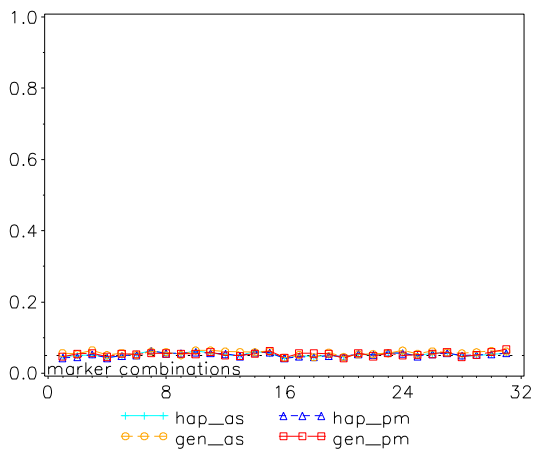
Power: FBAT,  
DOM, KMod3, 1500 trios



Power: FBAT, Linkage  
DOM, KMod3, 1500 trios



Power: Hotel\_Fam  
DOM, KMod3, 1500 trios

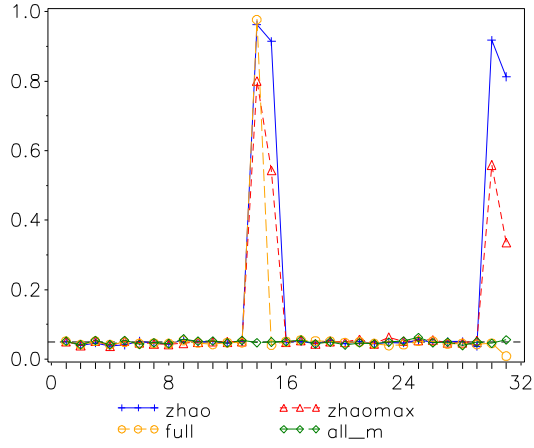




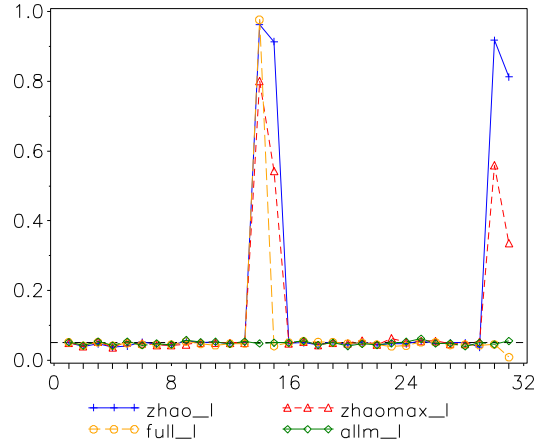
# KMod 3, Trios

## Multiplikativer Vererbungsmodus

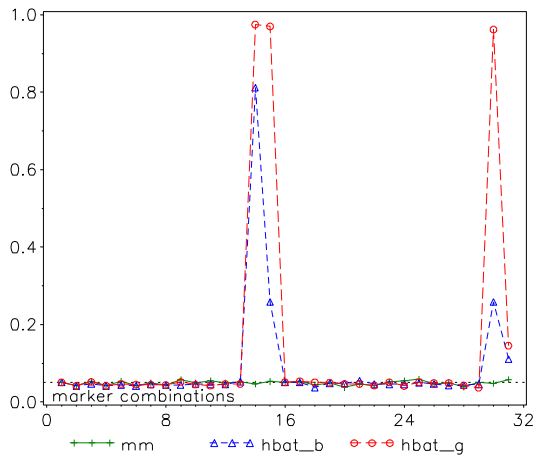
Power of of a) Famhap b) Unphased  
MUL 3, 1700 trios, H0\_(A)



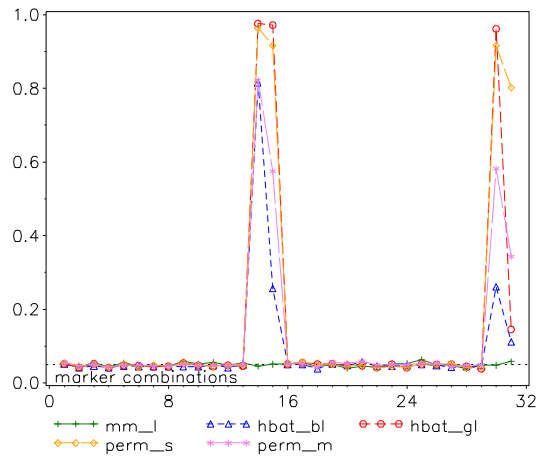
Power of of a) Famhap b) Unphased  
MUL 3, 1700 trios, H0\_(L)



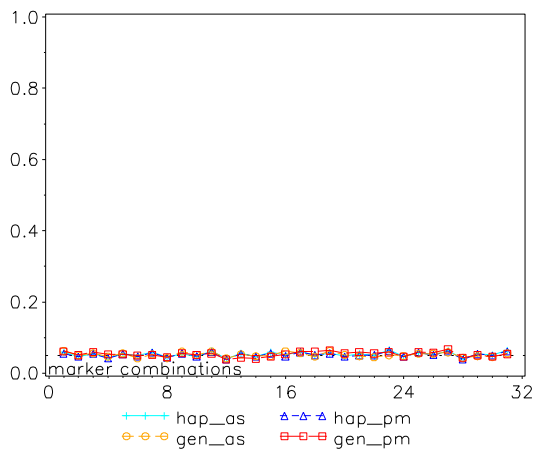
Power: FBAT,  
MUL, KMod3, 1700 trios



Power: FBAT, Linkage  
MUL, KMod3, 1700 trios



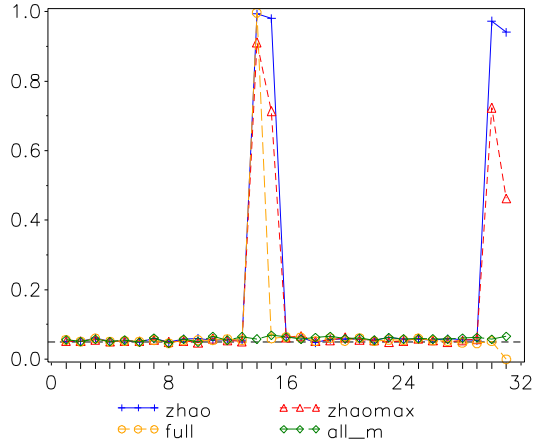
Power: Hotel\_Fam  
MUL, KMod3, 1700 trios



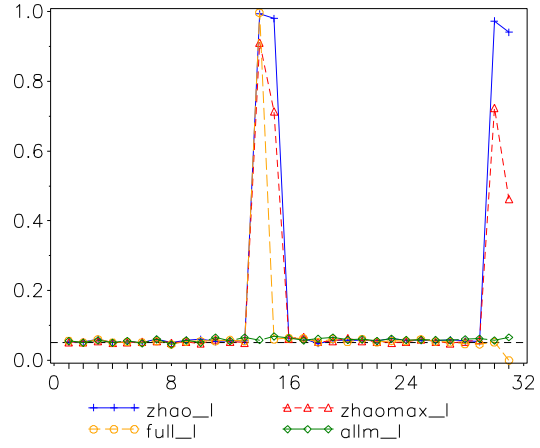
# KMod 3, Trios

## Rezessiver Vererbungsmodus

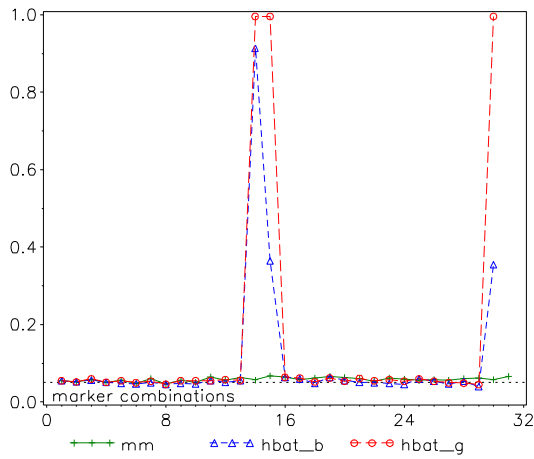
Power of of a) Famhap b) Unphased  
REZ 3, 11000 trios, H0\_(A)



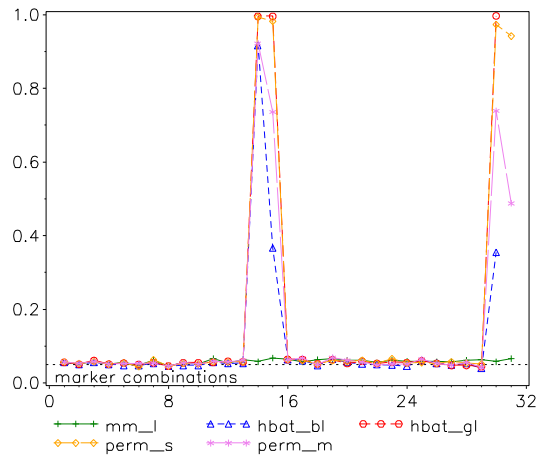
Power of of a) Famhap b) Unphased  
REZ 3, 11000 trios, H0\_(L)



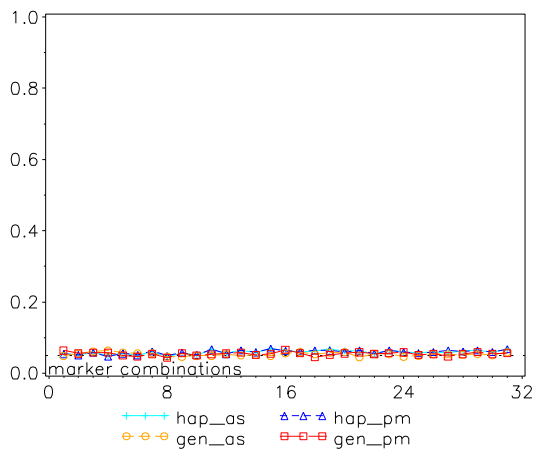
Power: FBAT,  
REZ, KMod3, 11000 trios



Power: FBAT, Linkage  
REZ, KMod3, 11000 trios



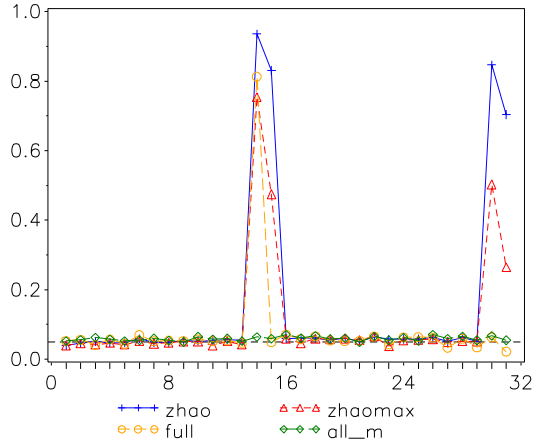
Power: Hotel\_Fam  
REZ, KMod3, 11000 trios



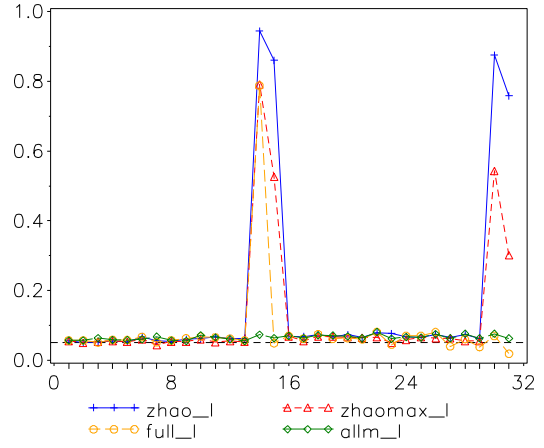
# KMod 3, AA-Familien

## Additiver Vererbungsmodus

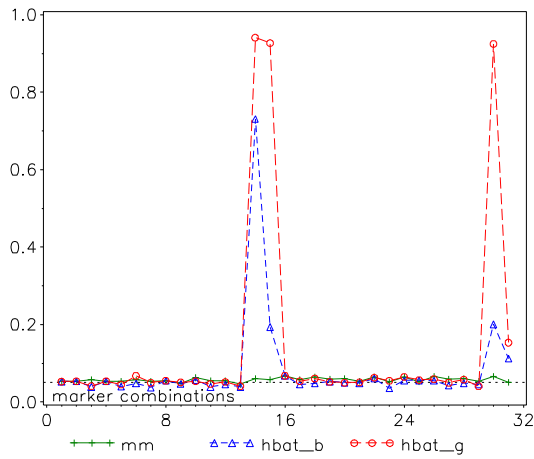
Power of of a) Famhap b) Unphased  
ADD 3, 500 AA families, H0\_(A)



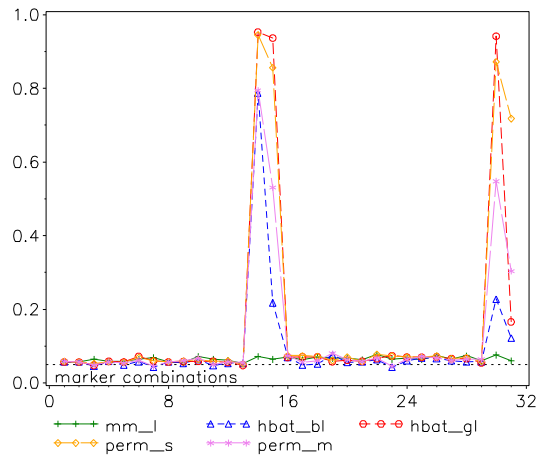
Power of of a) Famhap b) Unphased  
ADD 3, 500 AA families, H0\_(L)



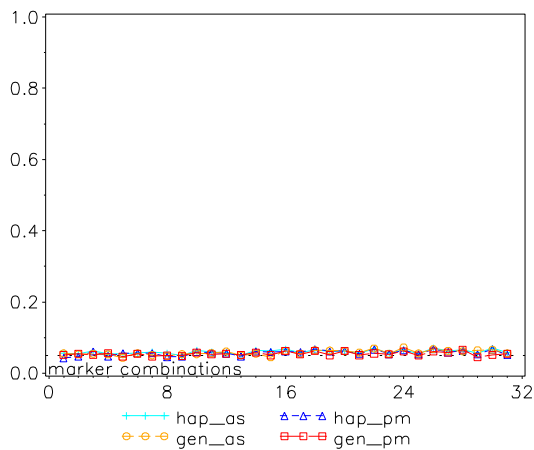
Power: FBAT,  
ADD, KMod3, 500 AA families



Power: FBAT, Linkage  
ADD, KMod3, 500 AA families



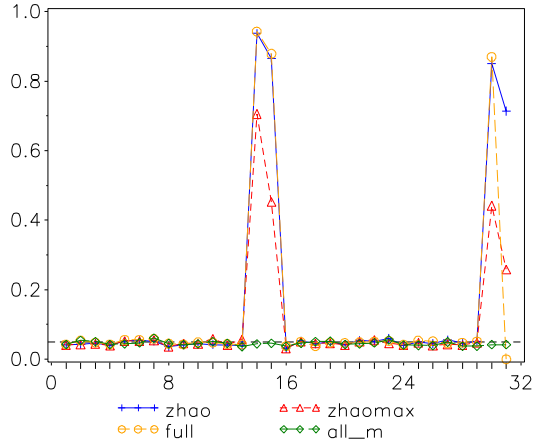
Power: Hotel\_Fam  
ADD, KMod3, 500 AA families



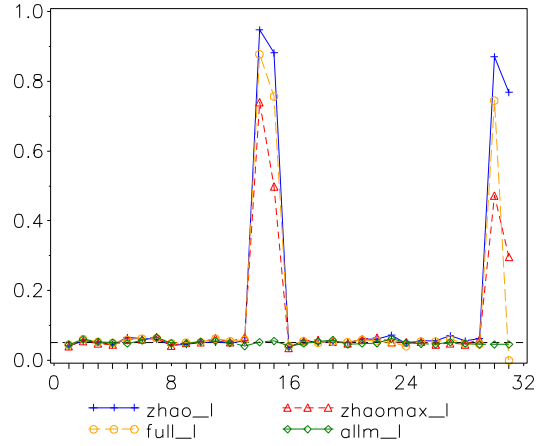
# KMod 3, AA-Familien

## Dominanter Vererbungsmodus

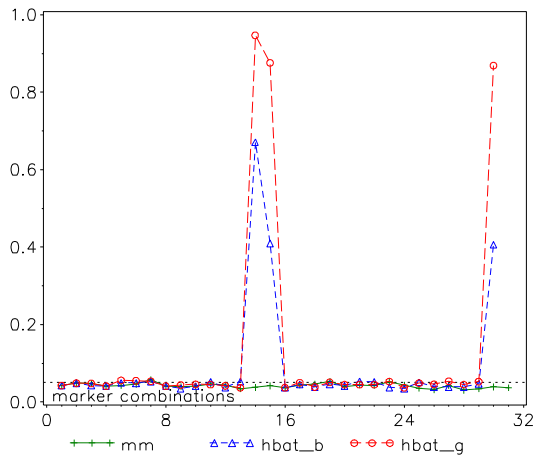
Power of of a) Famhap b) Unphased  
DOM 3, 700 AA families, H0\_(A)



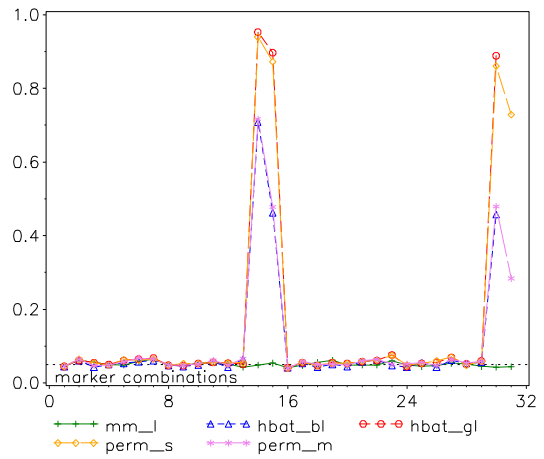
Power of of a) Famhap b) Unphased  
DOM 3, 700 AA families, H0\_(L)



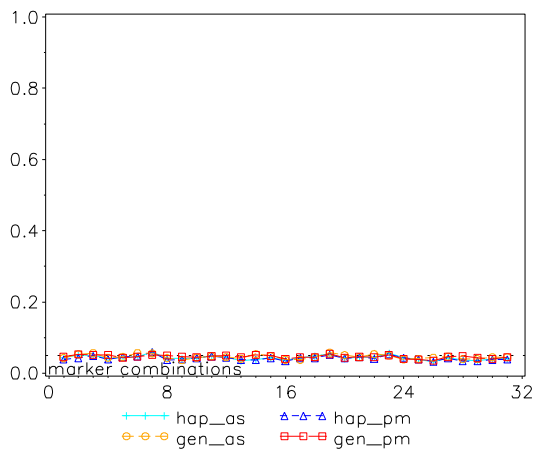
Power: FBAT,  
DOM, KMod3, 700 AA families



Power: FBAT, Linkage  
DOM, KMod3, 700 AA families



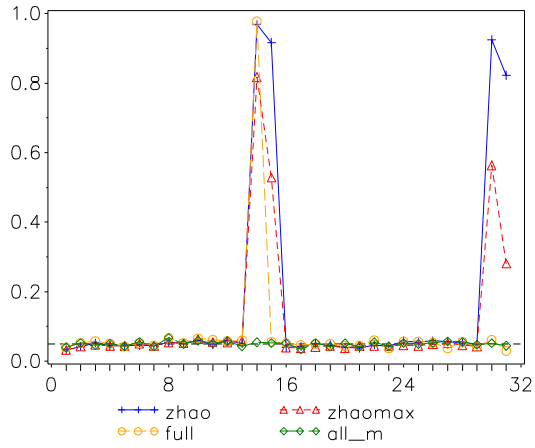
Power: Hotel\_Fam  
DOM, KMod3, 700 AA families



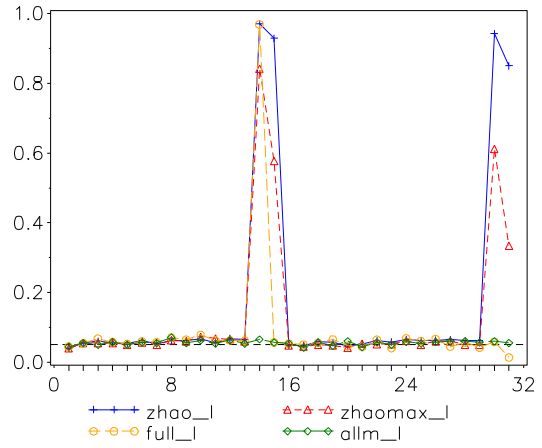
# KMod 3, AA-Familien

## Multiplikativer Vererbungsmodus

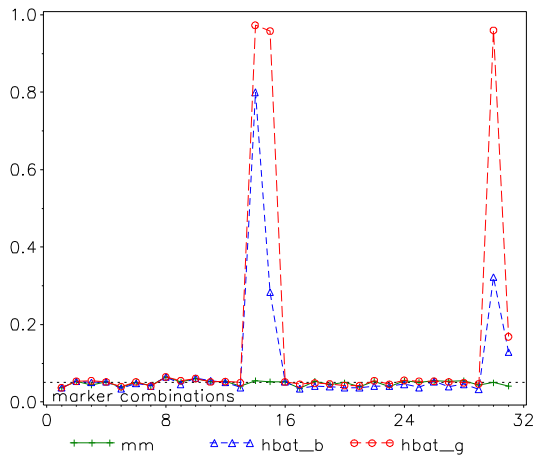
Power of of a) Famhap b) Unphased  
MUL 3, 500 AA families, H0\_(A)



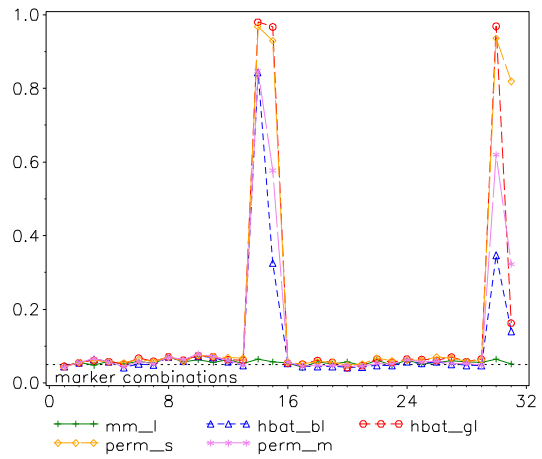
Power of of a) Famhap b) Unphased  
MUL 3, 500 AA families, H0\_(L)



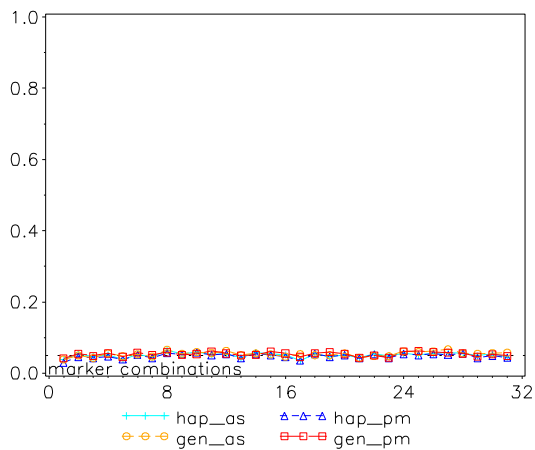
Power: FBAT,  
MUL, KMod3, 500 AA families



Power: FBAT, Linkage  
MUL, KMod3, 500 AA families



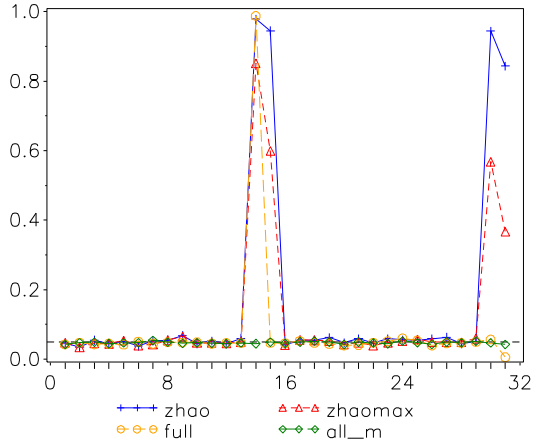
Power: Hotel\_Fam  
MUL, KMod3, 500 AA families



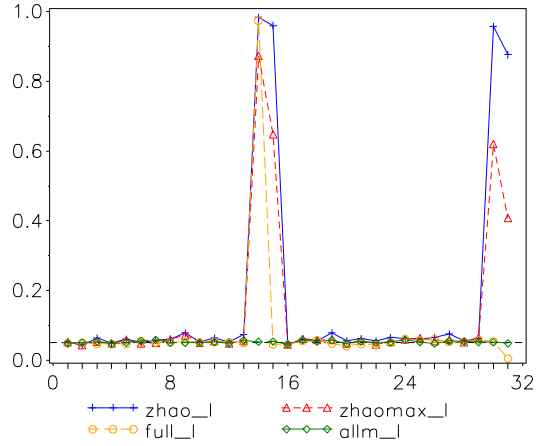
# KMod 3, AA-Familien

## Rezessiver Vererbungsmodus

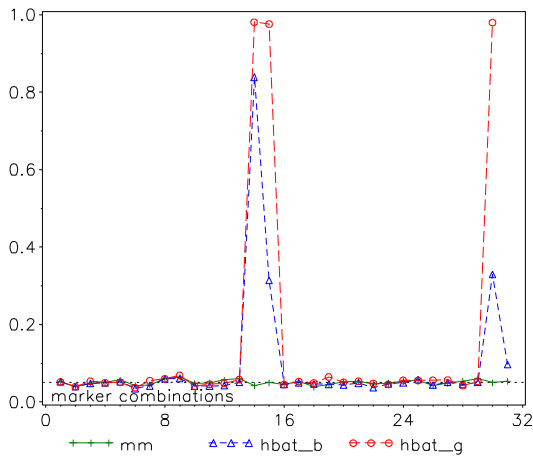
Power of of a) Famhap b) Unphased  
 REZ 3, 1400 AA families, H0\_(A)



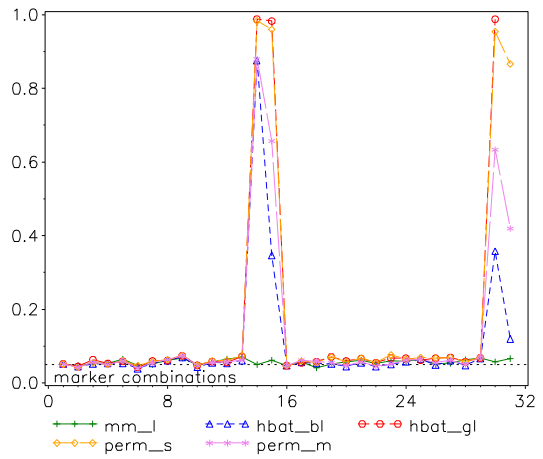
Power of of a) Famhap b) Unphased  
 REZ 3, 1400 AA families, H0\_(L)



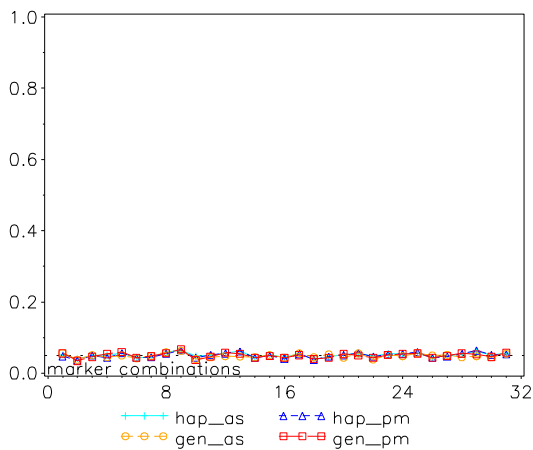
Power: FBAT,  
 REZ, KMod3, 1400 AA families



Power: FBAT, Linkage  
 REZ, KMod3, 1400 AA families



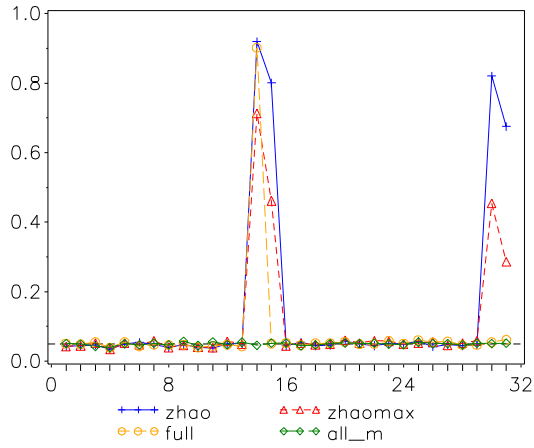
Power: Hotel\_Fam  
 REZ, KMod3, 1400 AA families



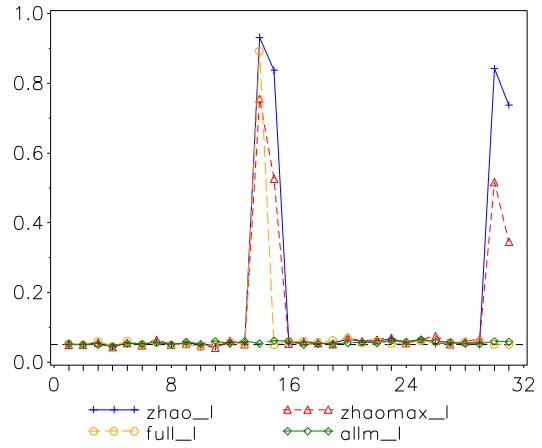
# KMod 3, AAU-Familien

## Additiver Vererbungsmodus

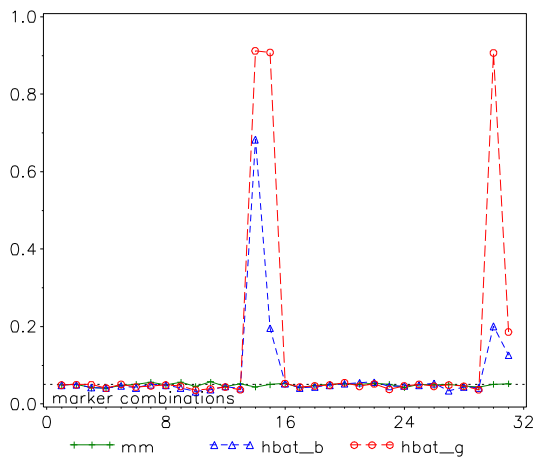
Power of of a) Famhap b) Unphased  
ADD 3, 500 AAU families, H0\_(A)



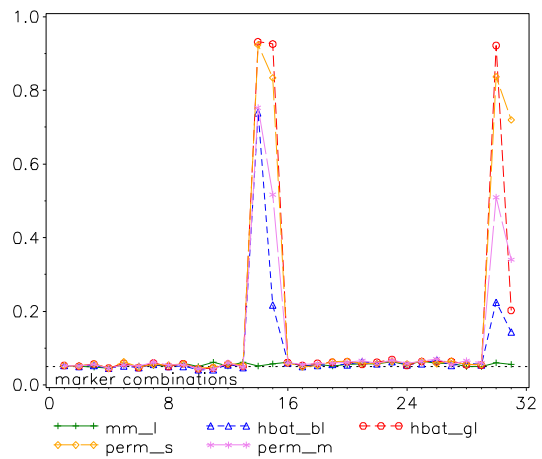
Power of of a) Famhap b) Unphased  
ADD 3, 500 AAU families, H0\_(L)



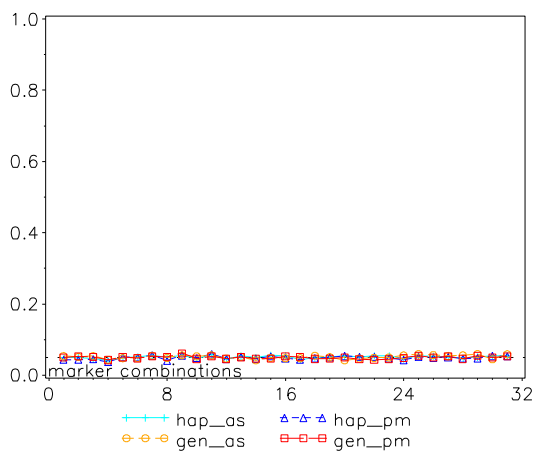
Power: FBAT,  
ADD, KMod3, 500 AAU families



Power: FBAT, Linkage  
ADD, KMod3, 500 AAU families



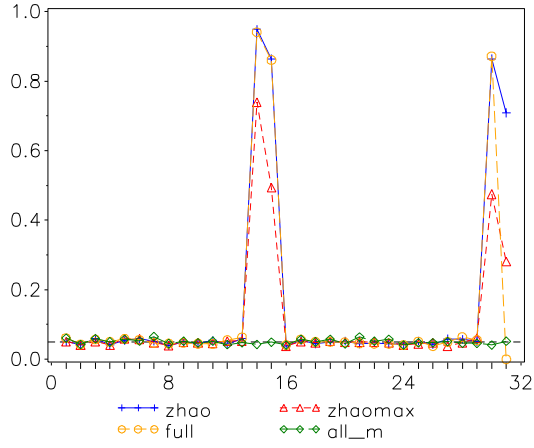
Power: Hotel\_Fam  
ADD, KMod3, 500 AAU families



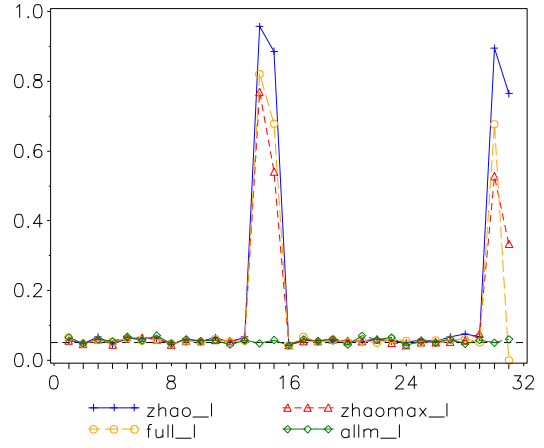
# KMod 3, AAU-Familien

## Dominanter Vererbungsmodus

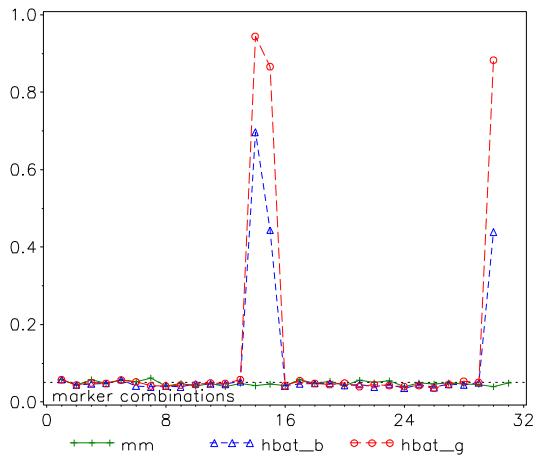
Power of of a) Famhap b) Unphased  
DOM 3, 600 AAU families, H0\_(A)



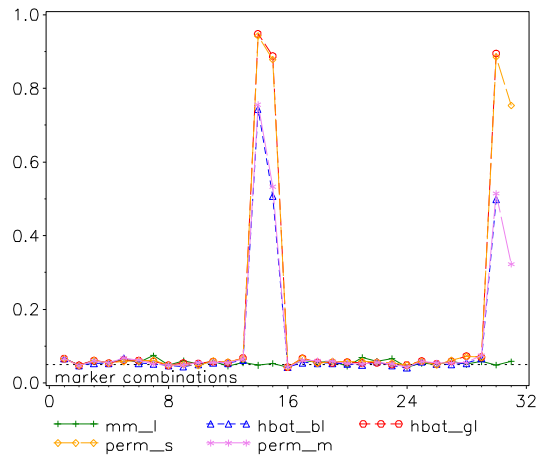
Power of of a) Famhap b) Unphased  
DOM 3, 600 AAU families, H0\_(L)



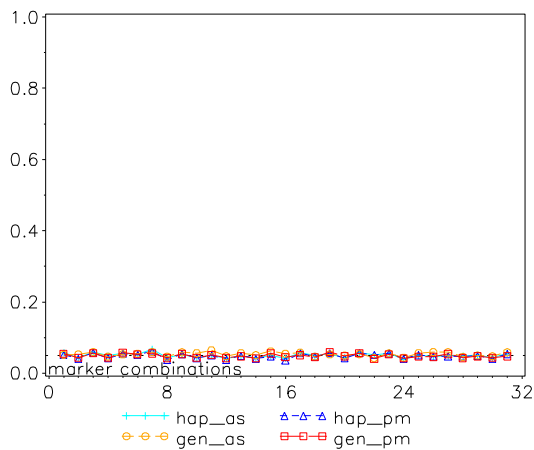
Power: FBAT,  
DOM, KMod3, 600 AAU families



Power: FBAT, Linkage  
DOM, KMod3, 600 AAU families



Power: Hotel\_Fam  
DOM, KMod3, 600 AAU families

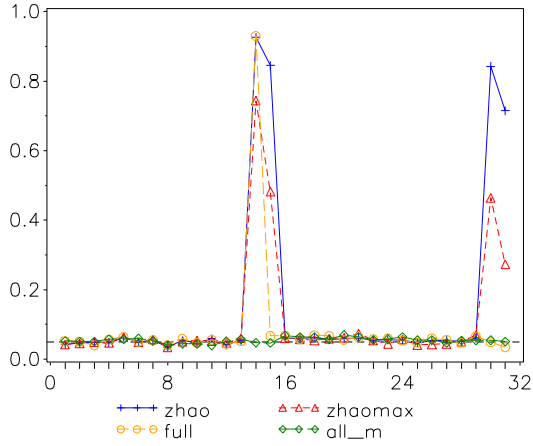




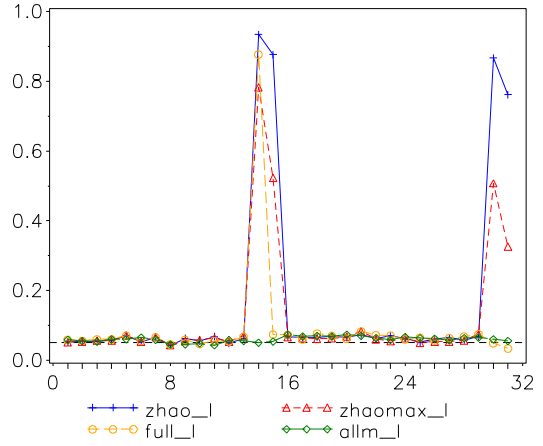
# KMod 3, AAU-Familien

## Multiplikativer Vererbungsmodus

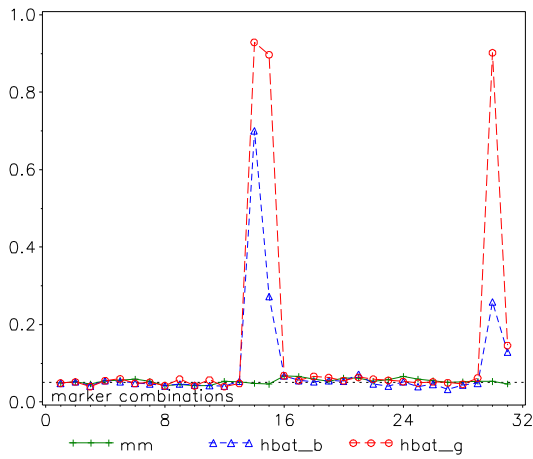
Power of of a) Famhap b) Unphased  
MUL 3, 400 AAU families, H0\_(A)



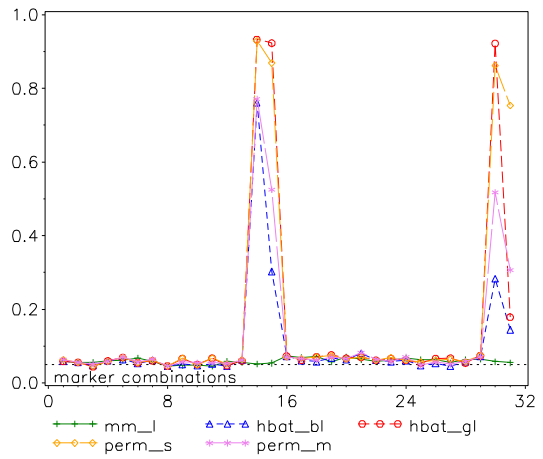
Power of of a) Famhap b) Unphased  
MUL 3, 400 AAU families, H0\_(L)



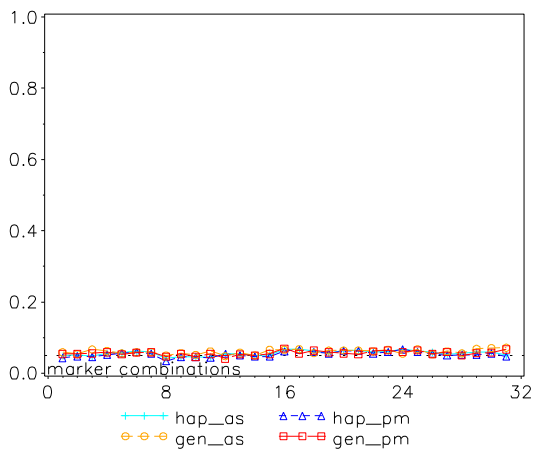
Power: FBAT,  
MUL, KMod3, 400 AAU families



Power: FBAT, Linkage  
MUL, KMod3, 400 AAU families



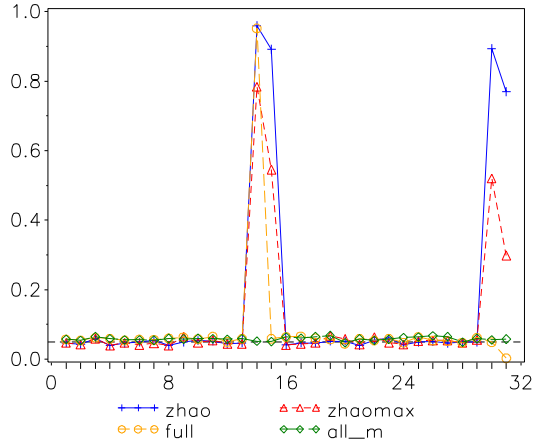
Power: Hotel\_Fam  
MUL, KMod3, 400 AAU families



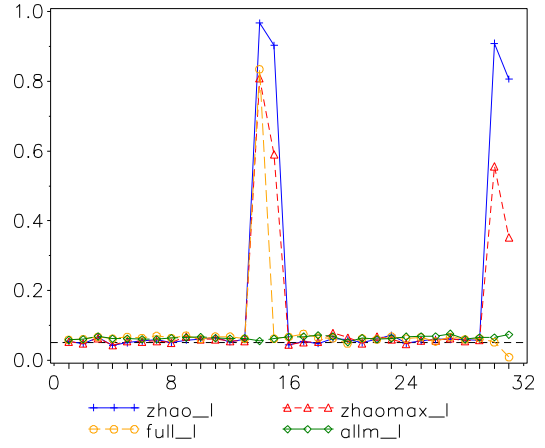
# KMod 3, AAU-Familien

## Rezessiver Vererbungsmodus

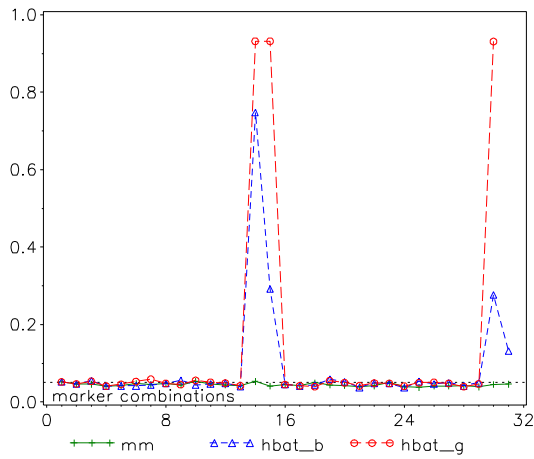
Power of of a) Famhap b) Unphased  
 REZ 3, 1000 AAU families, H0\_(A)



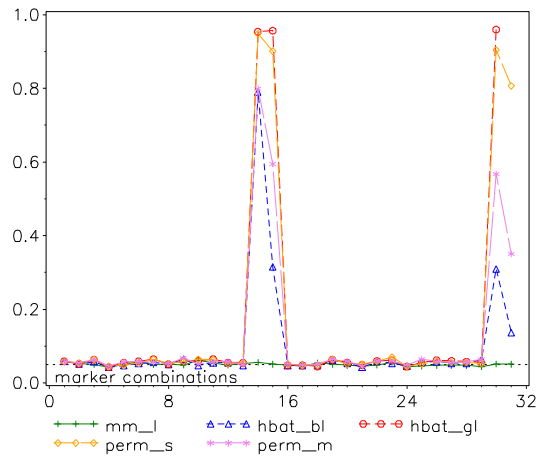
Power of of a) Famhap b) Unphased  
 REZ 3, 1000 AAU families, H0\_(L)



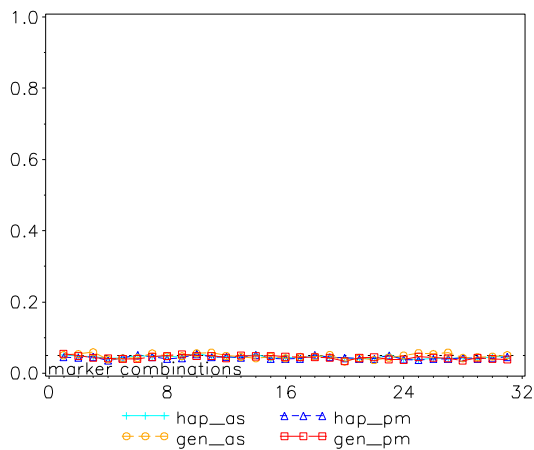
Power: FBAT,  
 REZ, KMod3, 1000 AAU families



Power: FBAT, Linkage  
 REZ, KMod3, 1000 AAU families



Power: Hotel\_Fam  
 REZ, KMod3, 1000 AAU families

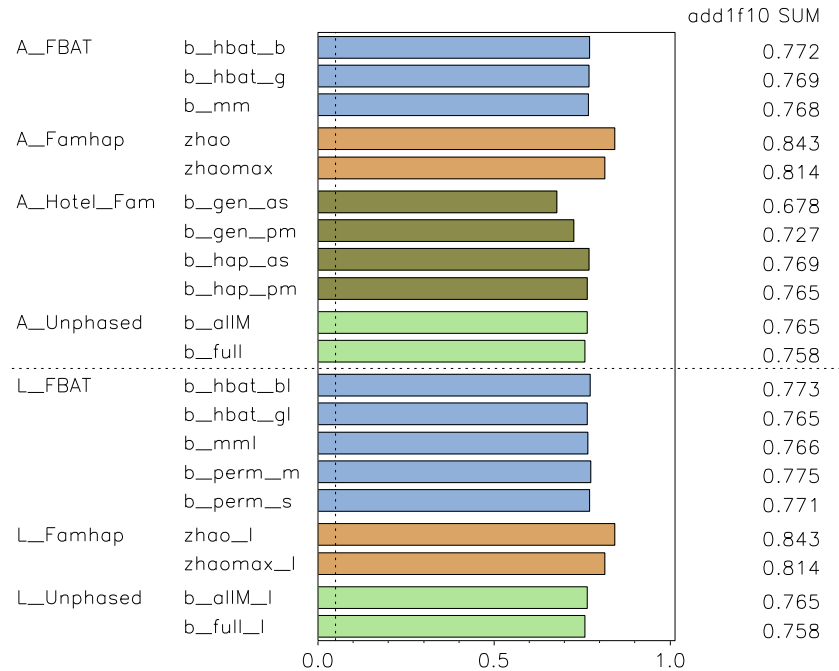


## Overall-Power, $\alpha = 0.05$

**Famhap: totaler p-Wert über alle Markerkombinationen vs. FBAT, Hotel\_Fam, Unphased: Bonferroni-korrigierter minimaler p-Wert; KMod 1**

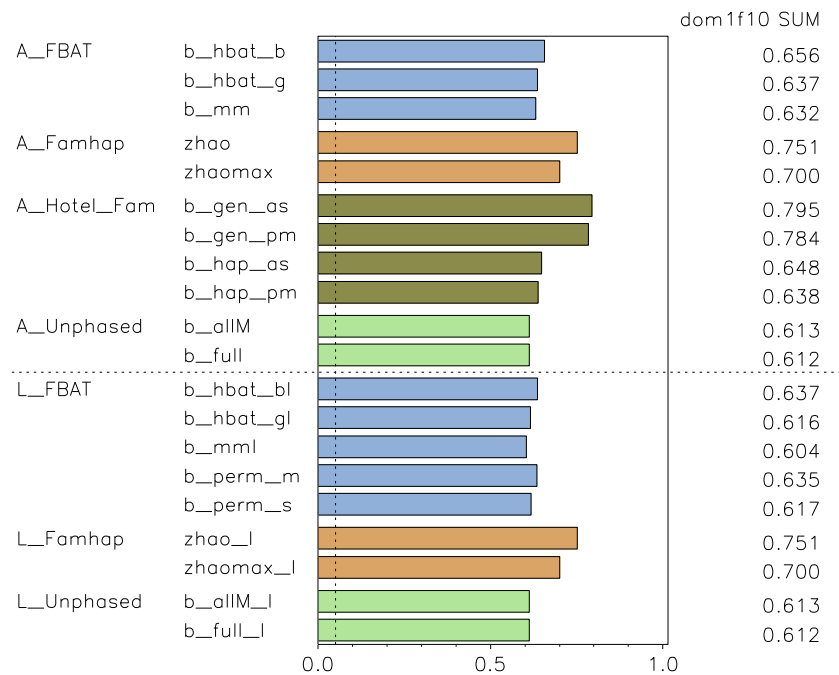
### Trios

Overall power: alpha=0.05, ADD, KMod1, 1100 trios



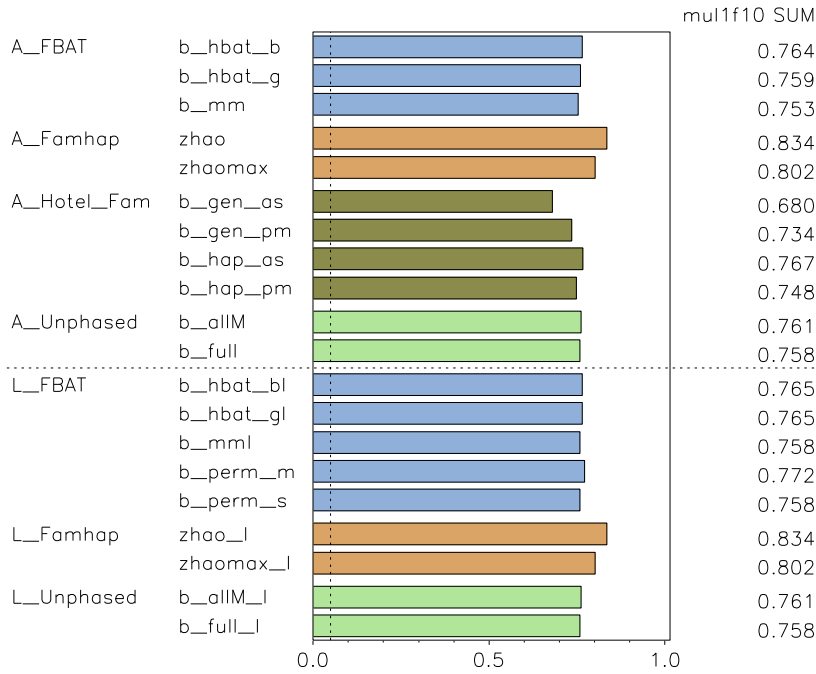
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Overall power: alpha=0.05, DOM, KMod1, 400 trios



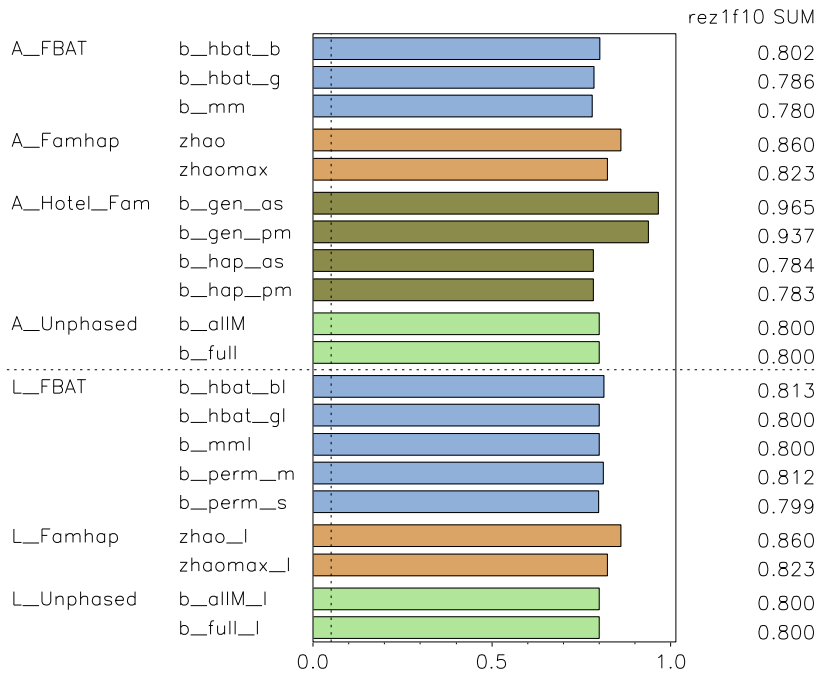
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Overall power: alpha=0.05, MUL, KMod1, 700 trios



Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

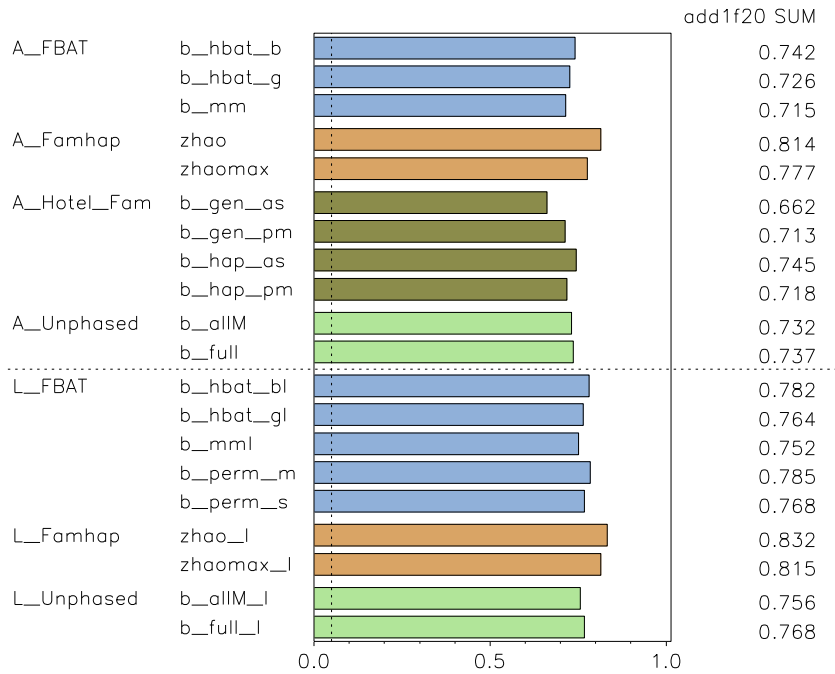
Overall power: alpha=0.05, REZ, KMod1, 2500 trios



Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

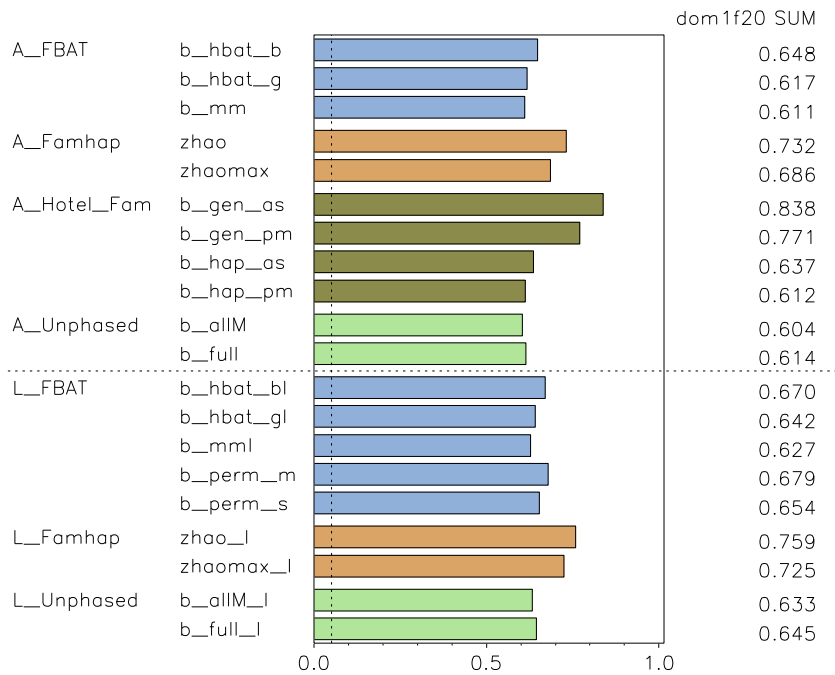
## AA-Familien

Overall power: alpha=0.05, ADD, KMod1, 300 AA-families



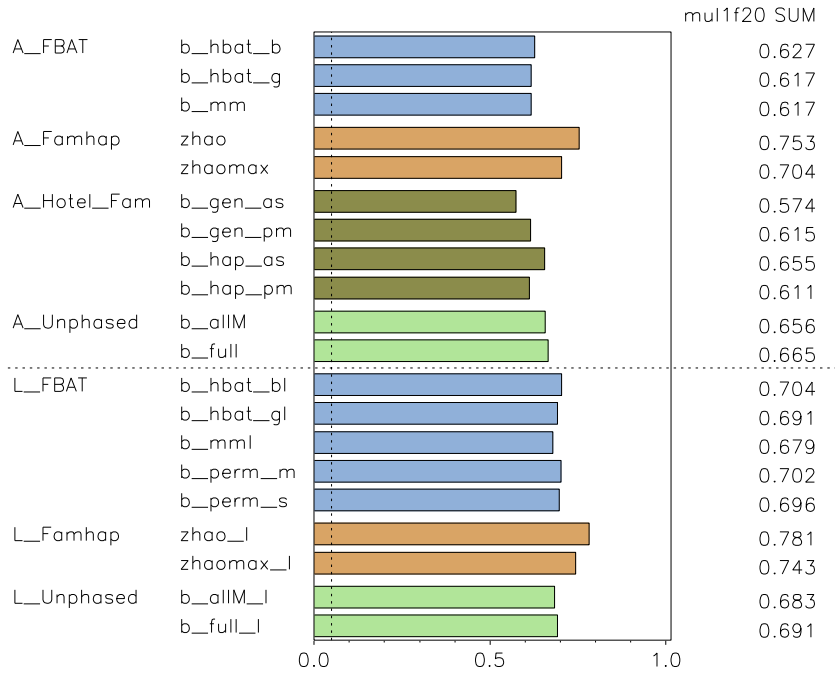
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Overall power: alpha=0.05, DOM, KMod1, 300 AA-families



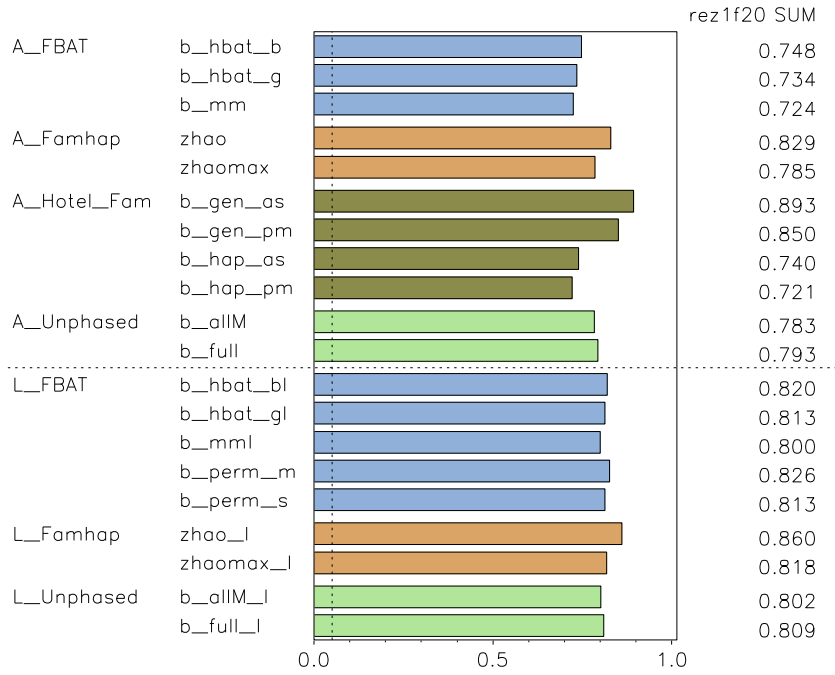
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Overall power: alpha=0.05, MUL, KMod1, 200 AA-families



Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

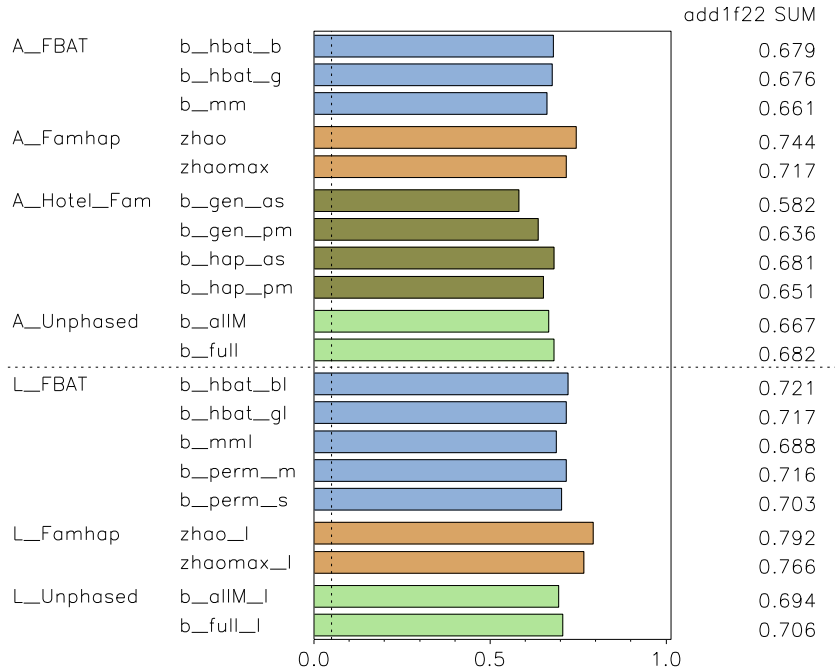
Overall power: alpha=0.05, REZ, KMod1, 500 AA-families



Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

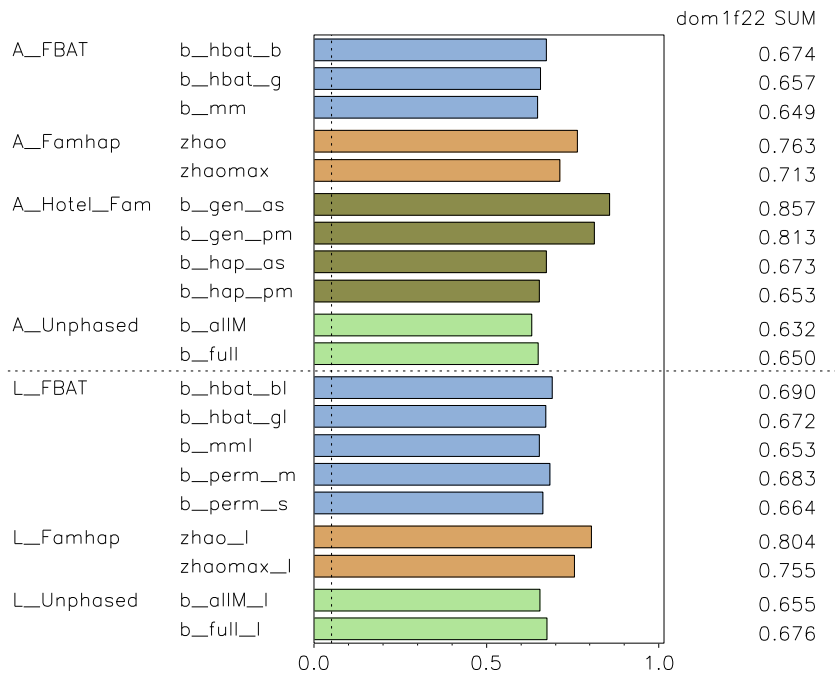
## AAUU-Familien

Overall power: alpha=0.05, ADD, KMod1, 300 AAUU-families



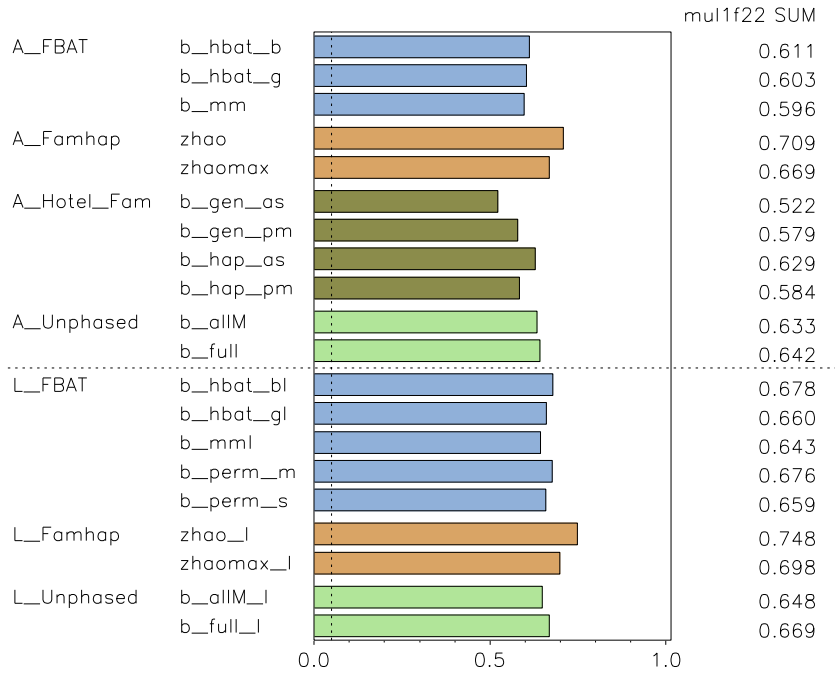
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Overall power: alpha=0.05, DOM, KMod1, 300 AAUU-families



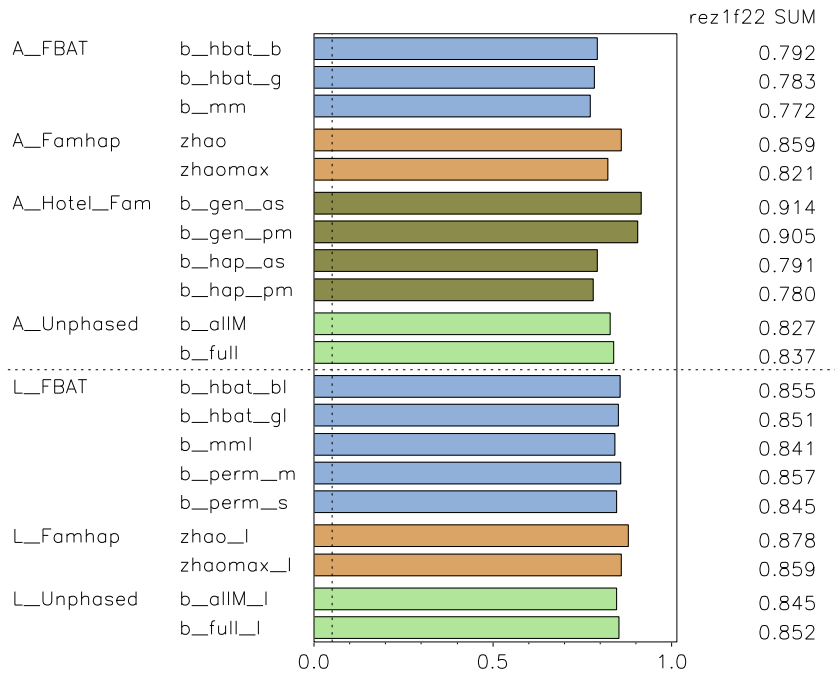
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Overall power: alpha=0.05, MUL, KMod1, 200 AAU-families



Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Overall power: alpha=0.05, REZ, KMod1, 600 AAU-families



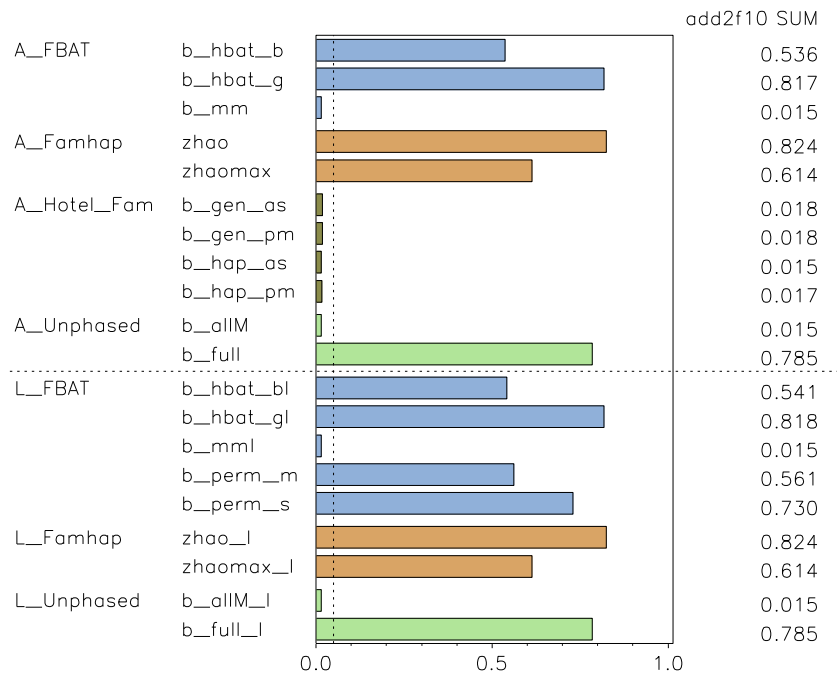
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased



## KMod 2

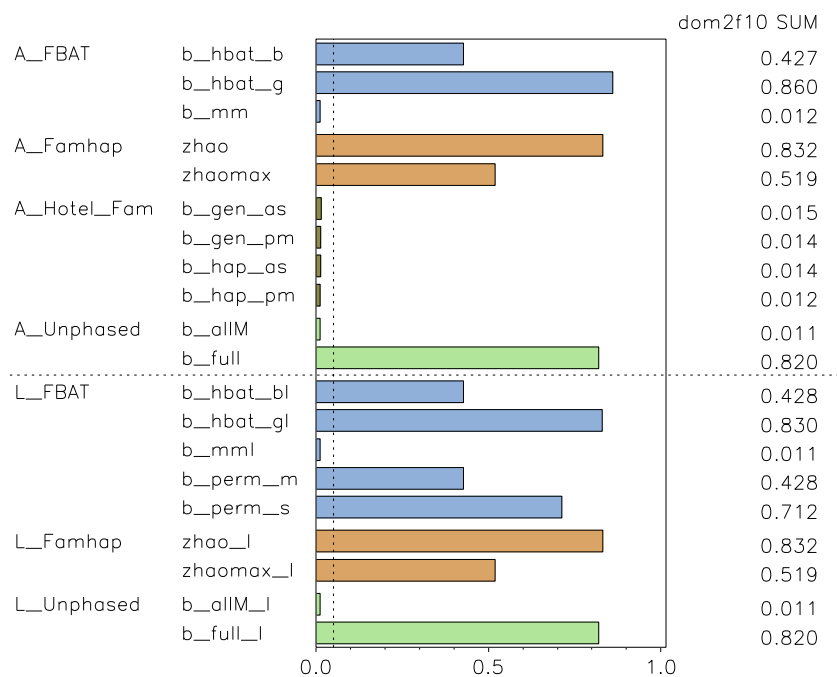
### Trios

Overall power: alpha=0.05, ADD, KMod2, 1800 trios



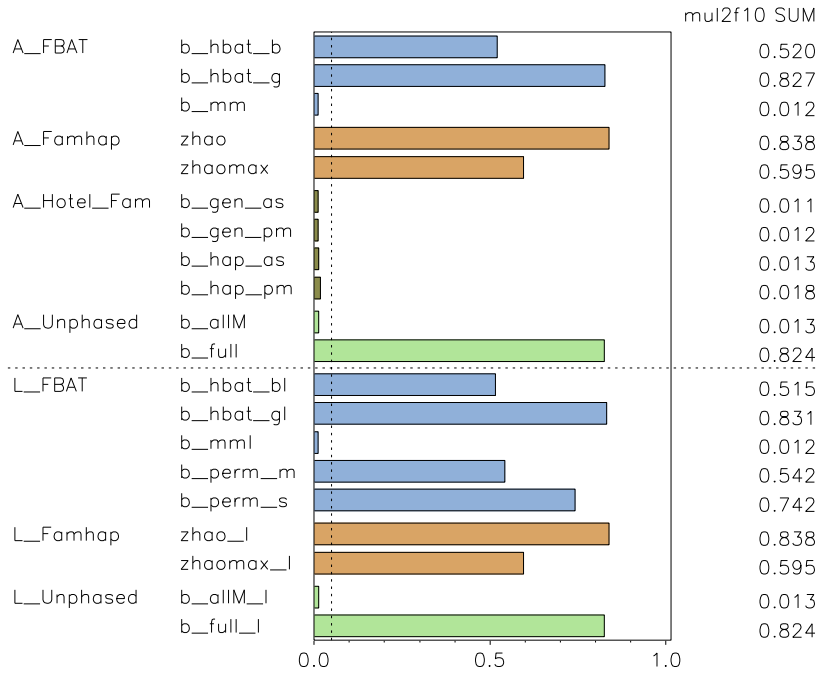
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Overall power: alpha=0.05, DOM, KMod2, 1000 trios



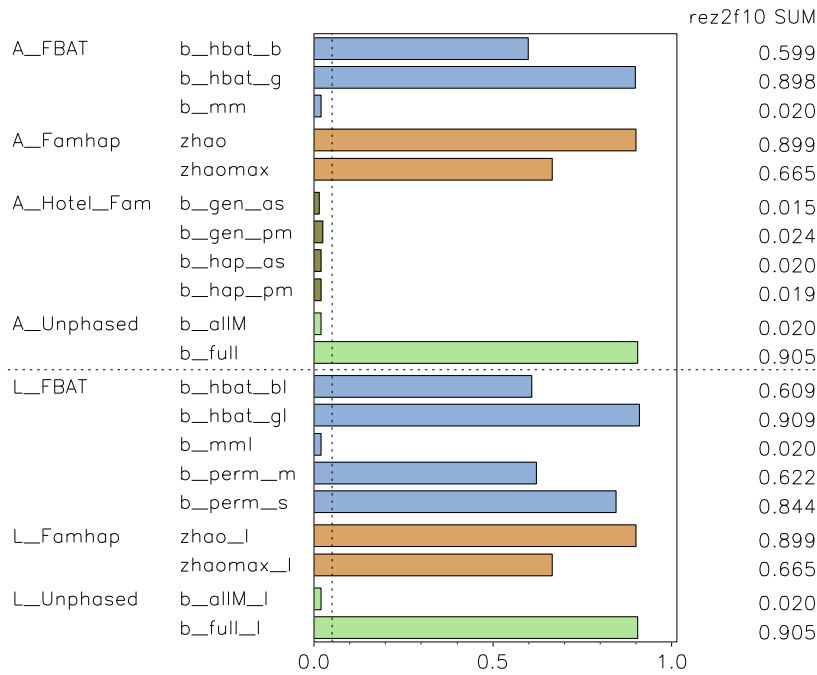
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Overall power: alpha=0.05, MUL, KMod2, 1200 trios



Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

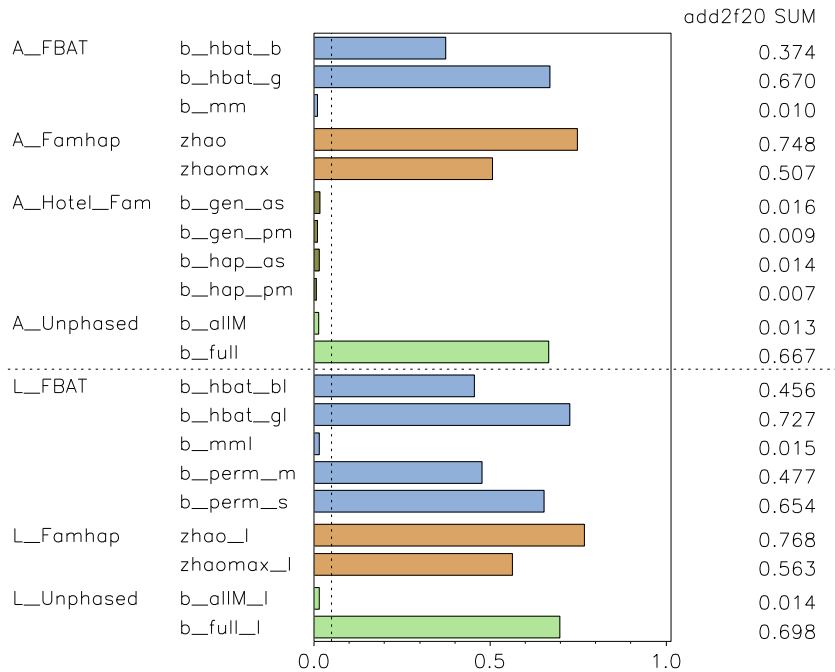
Overall power: alpha=0.05, REZ, KMod2, 6000 trios



Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

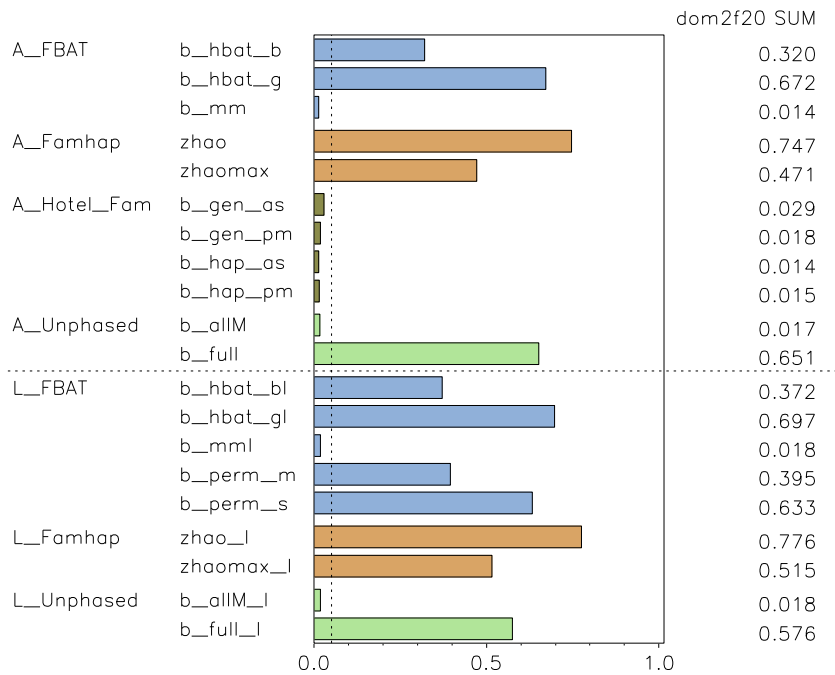
## AA-Familien

Overall power: alpha=0.05, ADD, KMod2, 400 AA-families



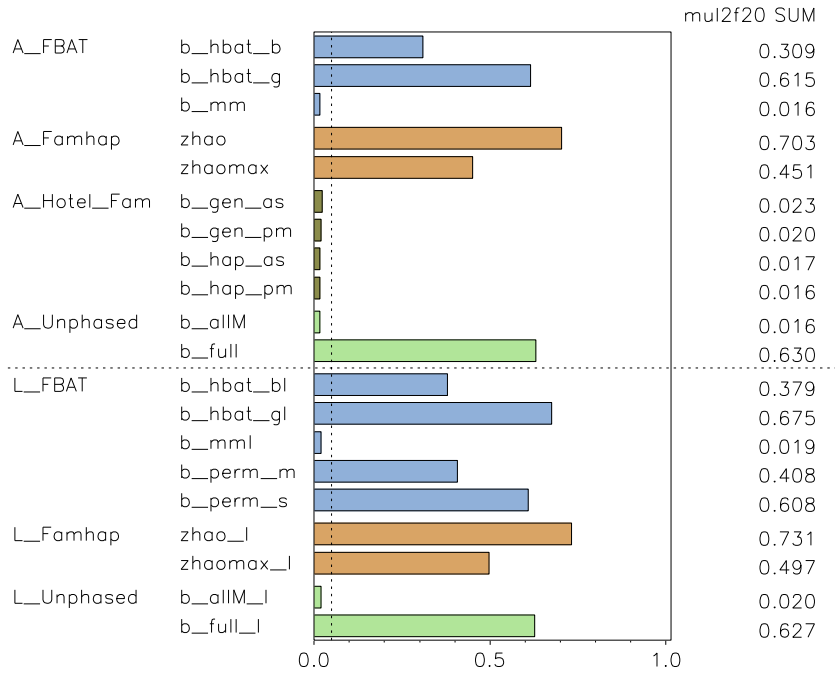
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Overall power: alpha=0.05, DOM, KMod2, 500 AA-families



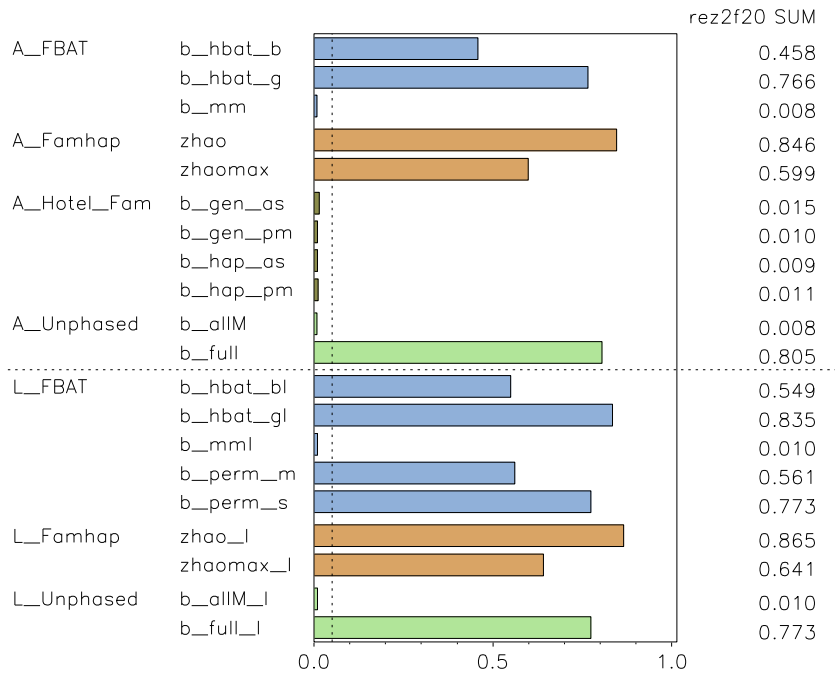
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Overall power: alpha=0.05, MUL, KMod2, 300 AA-families



Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

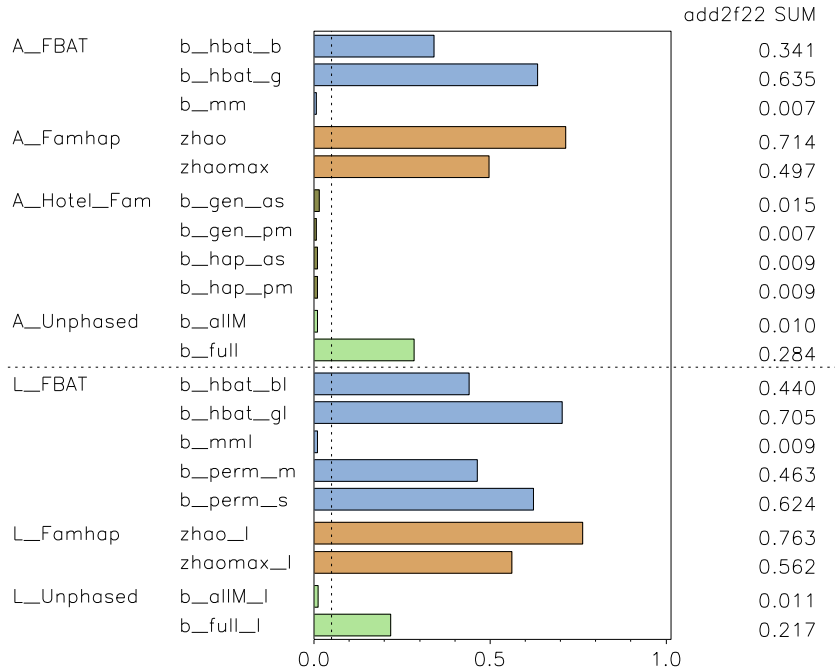
Overall power: alpha=0.05, REZ, KMod2, 900 AA-families



Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

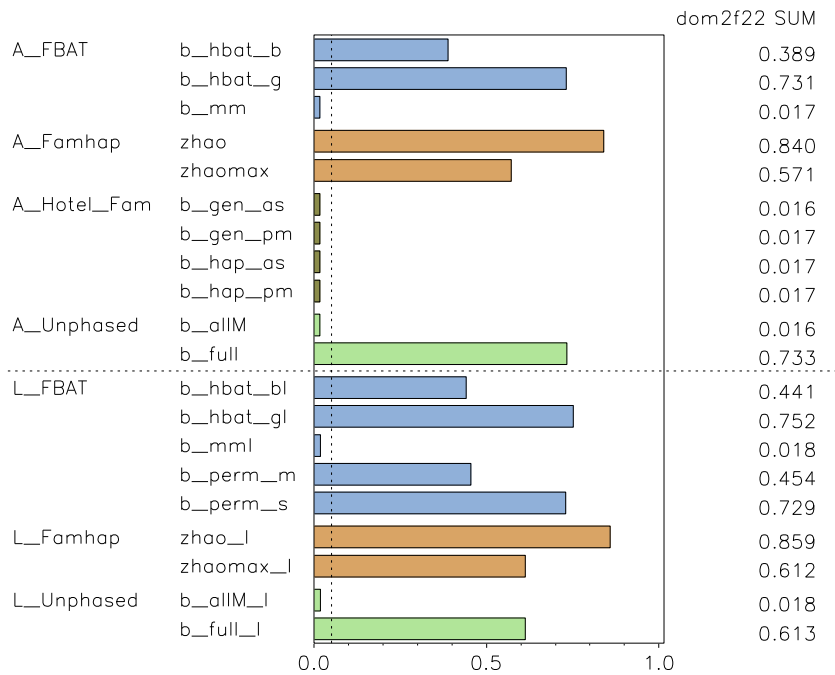
## AAUU-Familien

Overall power: alpha=0.05, ADD, KMod2, 400 AAUU-families



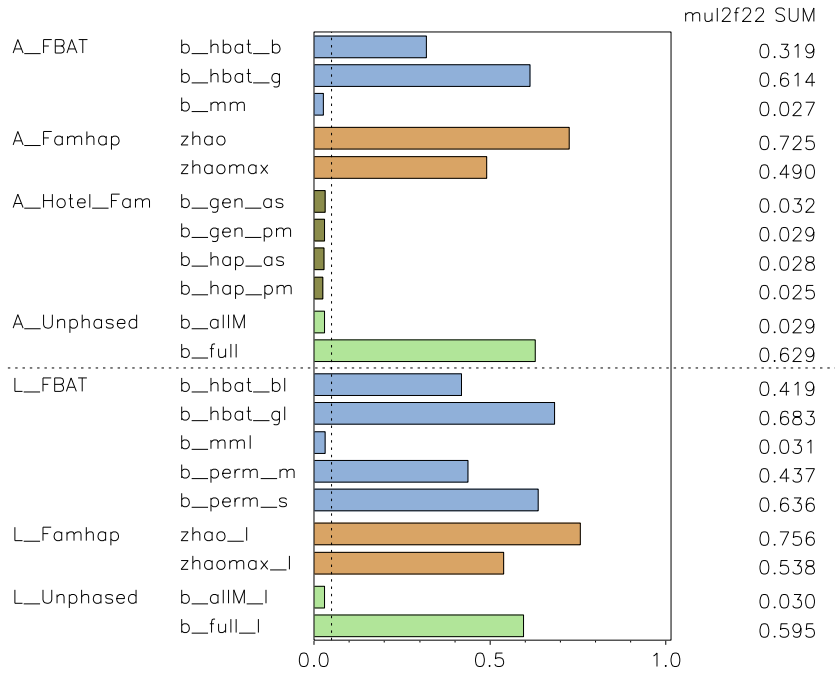
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Overall power: alpha=0.05, DOM, KMod2, 500 AAUU-families



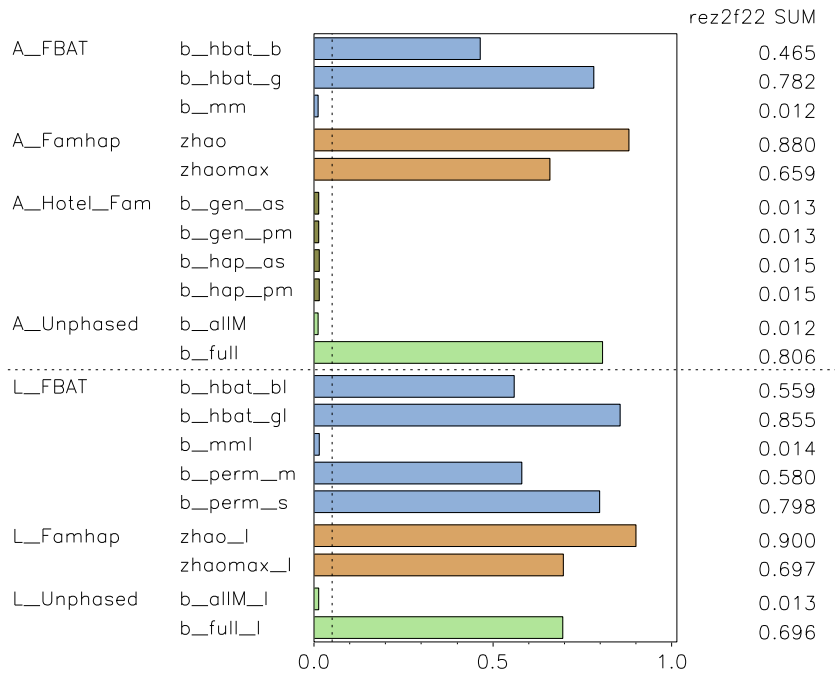
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Overall power: alpha=0.05, MUL, KMod2, 300 AAU-families



Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Overall power: alpha=0.05, REZ, KMod2, 900 AAU-families

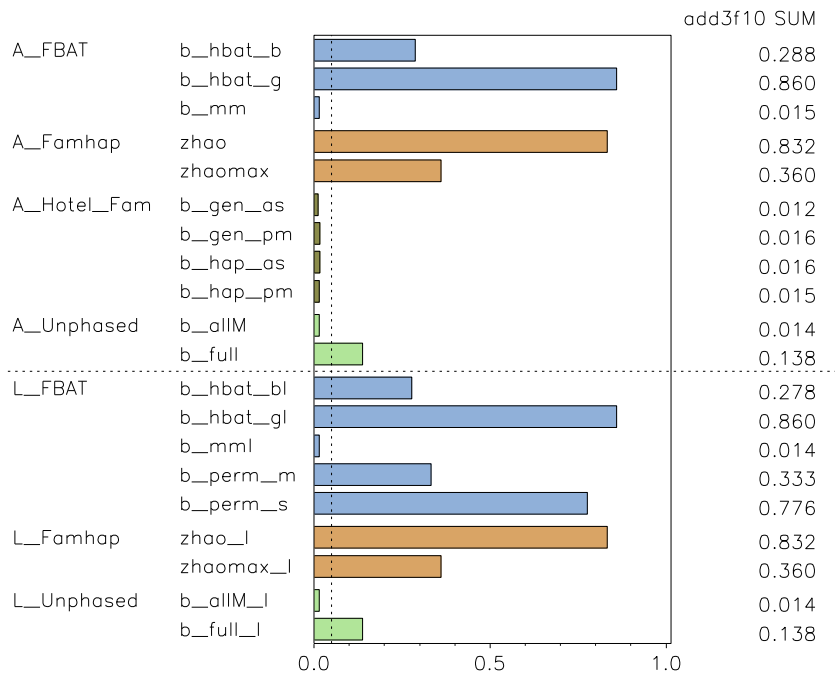


Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

### KMod 3

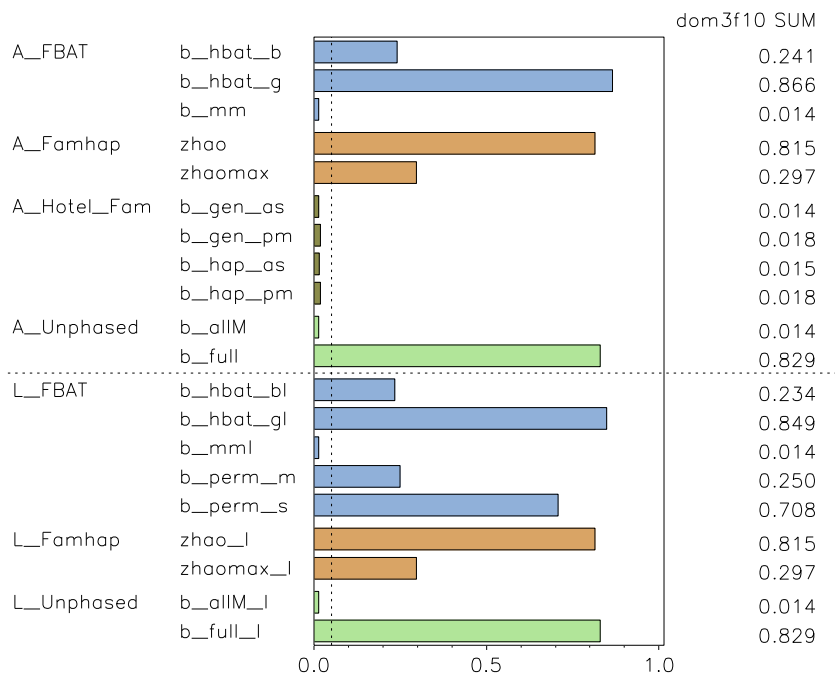
#### Trios

Overall power: alpha=0.05, ADD, KMod3, 2500 trios



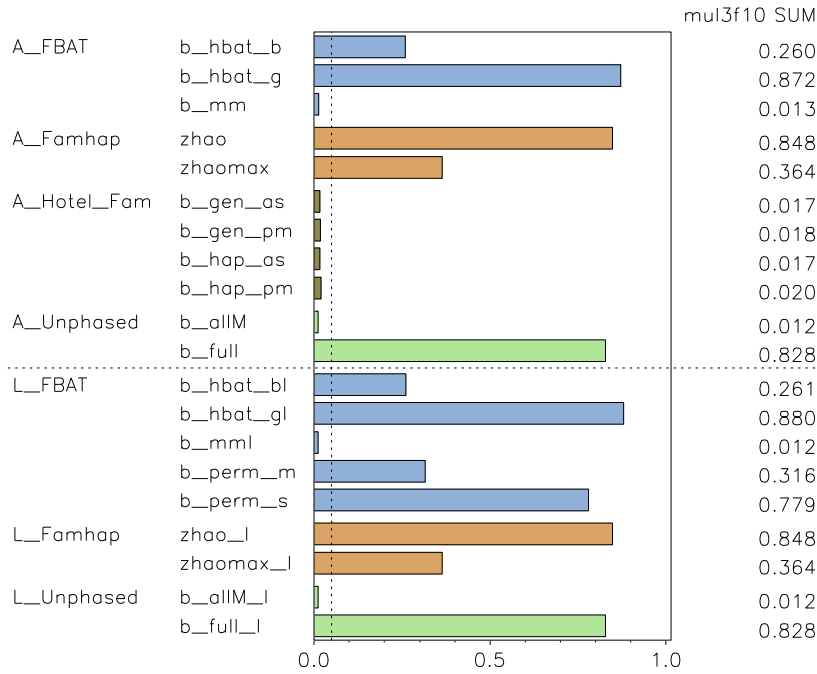
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Overall power: alpha=0.05, DOM, KMod3, 1500 trios



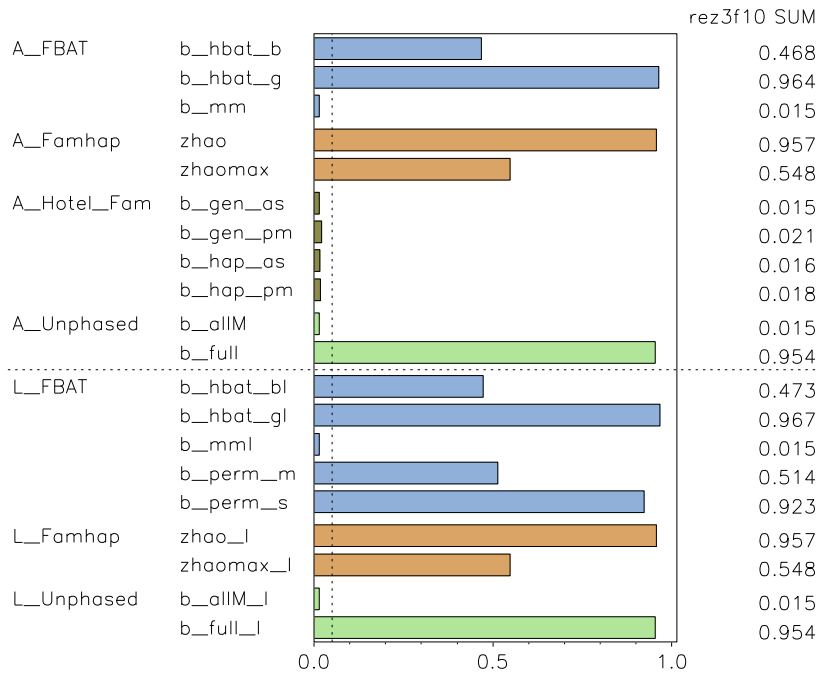
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Overall power: alpha=0.05, MUL, KMod3, 1700 trios



Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Overall power: alpha=0.05, REZ, KMod3, 11000 trios

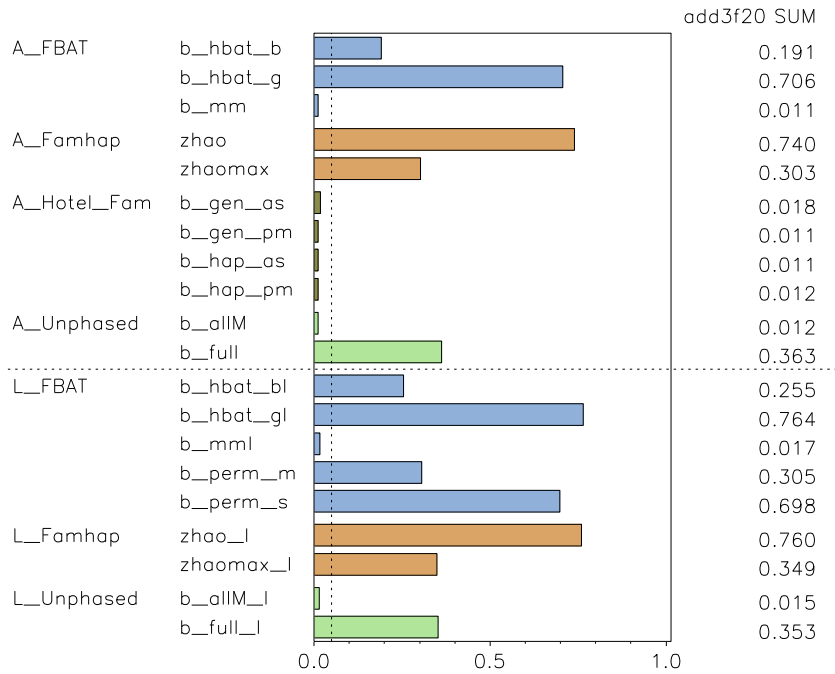


Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased



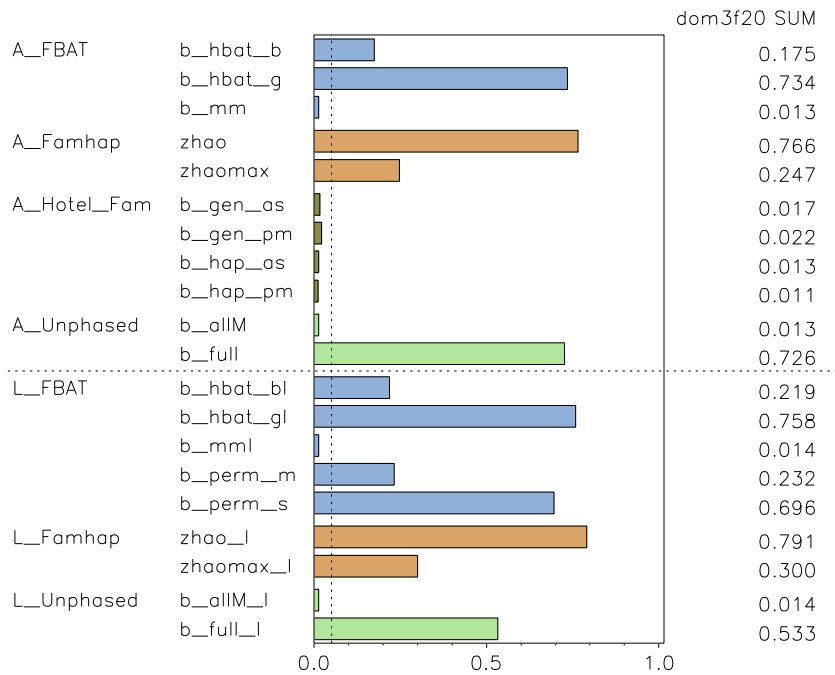
## AA-Familien

Overall power: alpha=0.05, ADD, KMod3, 500 AA-families



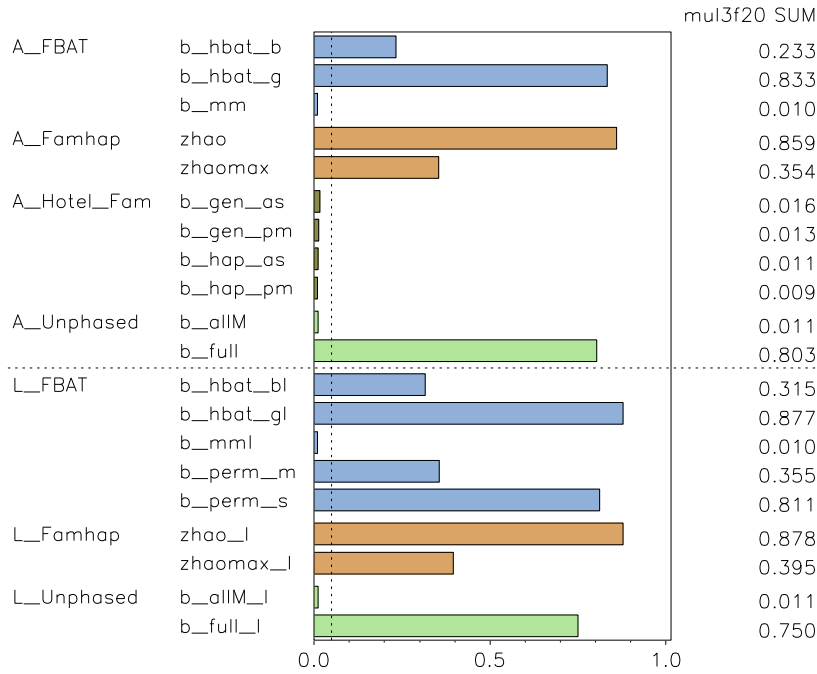
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Overall power: alpha=0.05, DOM, KMod3, 700 AA-families



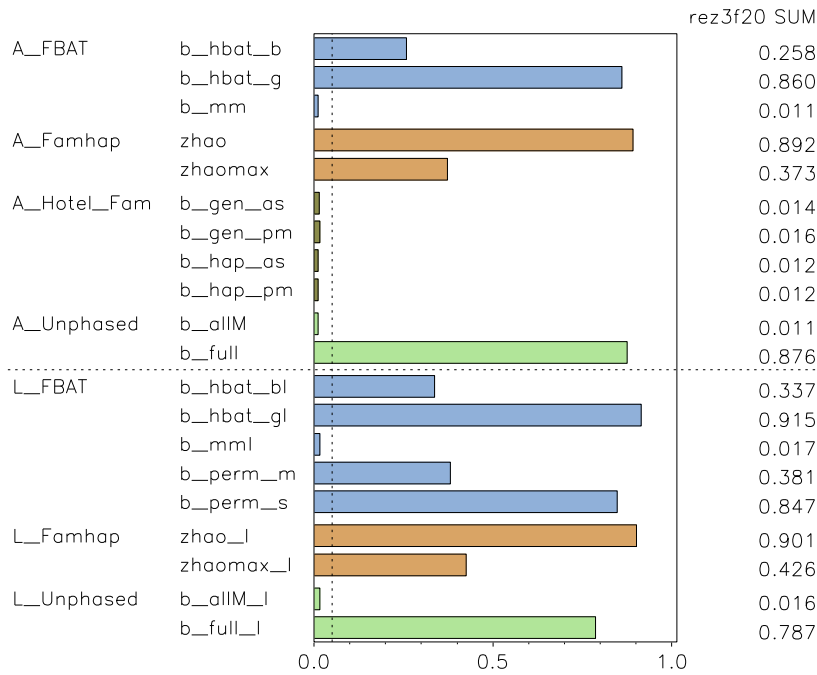
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Overall power: alpha=0.05, MUL, KMod3, 500 AA-families



Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

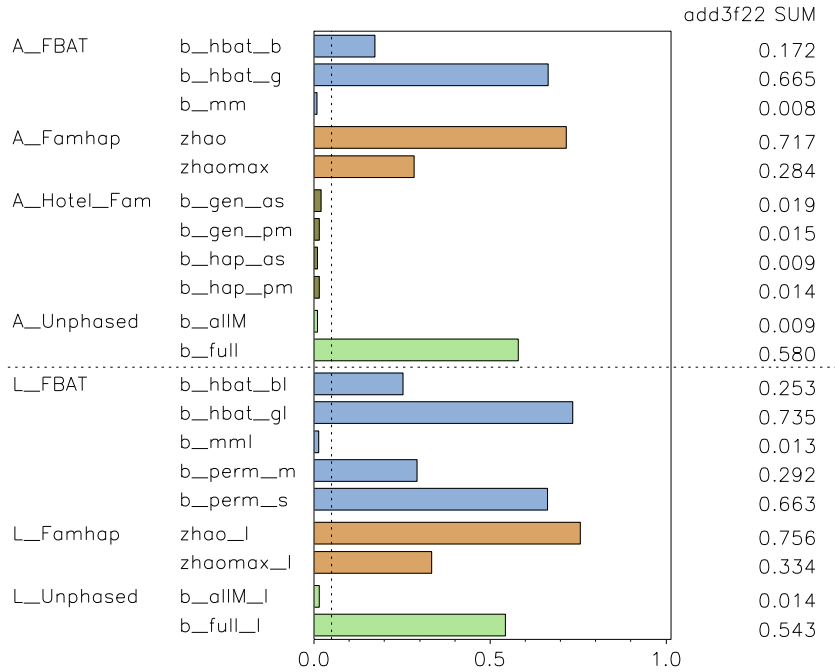
Overall power: alpha=0.05, REZ, KMod3, 1400 AA-families



Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

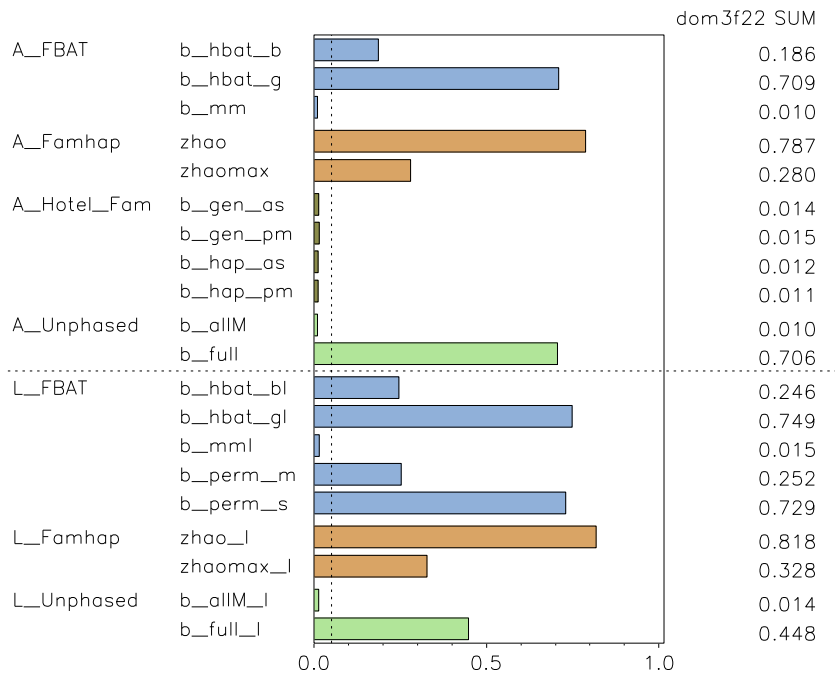
## AAUU-Familien

Overall power: alpha=0.05, ADD, KMod3, 500 AAUU-families



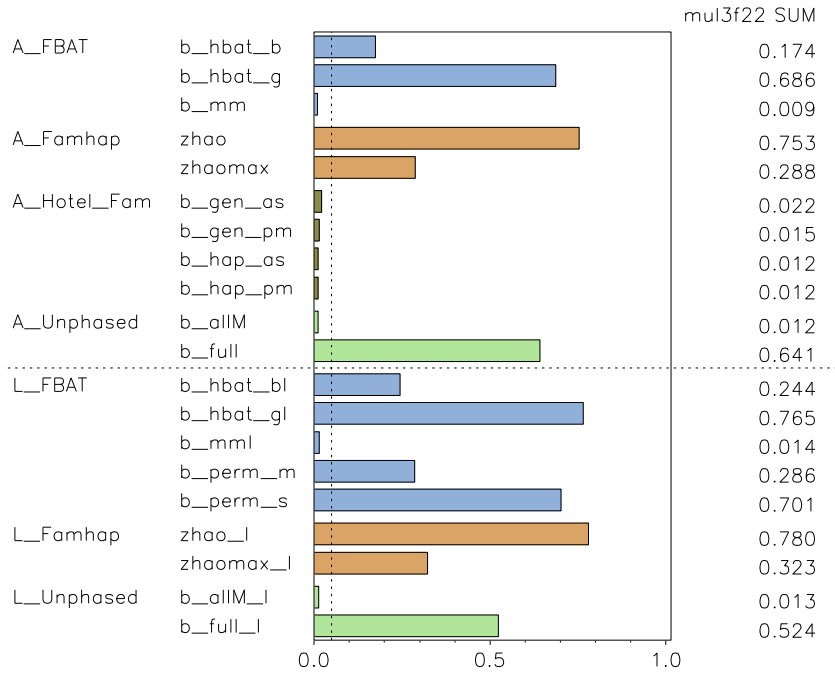
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Overall power: alpha=0.05, DOM, KMod3, 600 AAUU-families



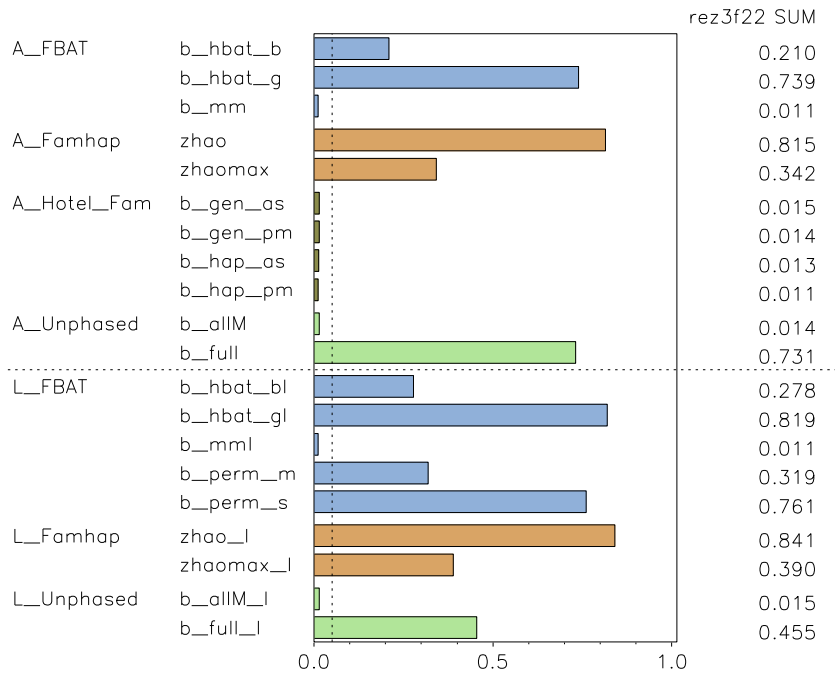
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Overall power: alpha=0.05, MUL, KMod3, 400 AAU-families



Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Overall power: alpha=0.05, REZ, KMod3, 1000 AAU-families



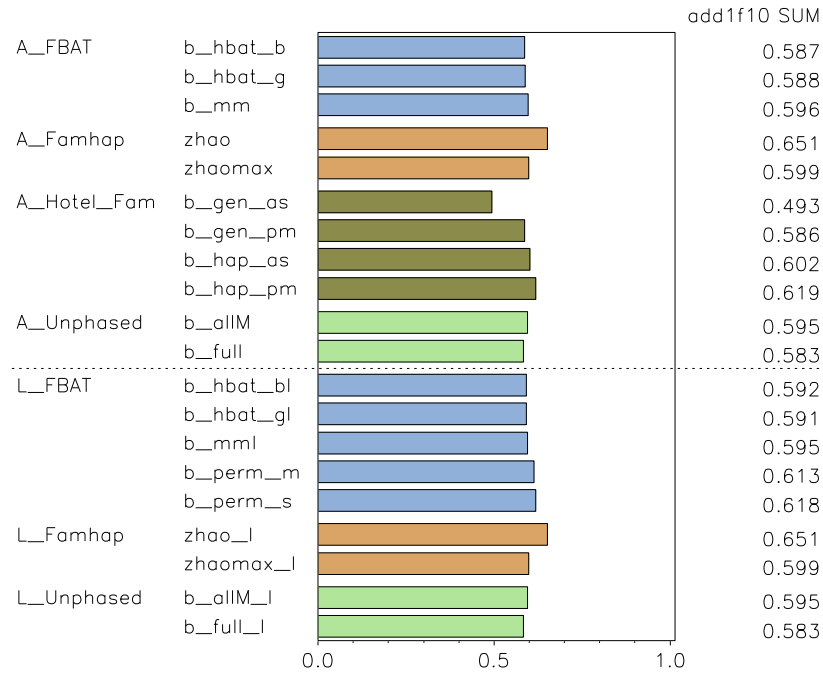
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

## Overall-Power, $\alpha = 0.01$

**Famhap: totaler p-Wert über alle Markerkombinationen vs. FBAT, Hotel\_Fam, Unphased: Bonferroni-korrigierter minimaler p-Wert; KMod 1**

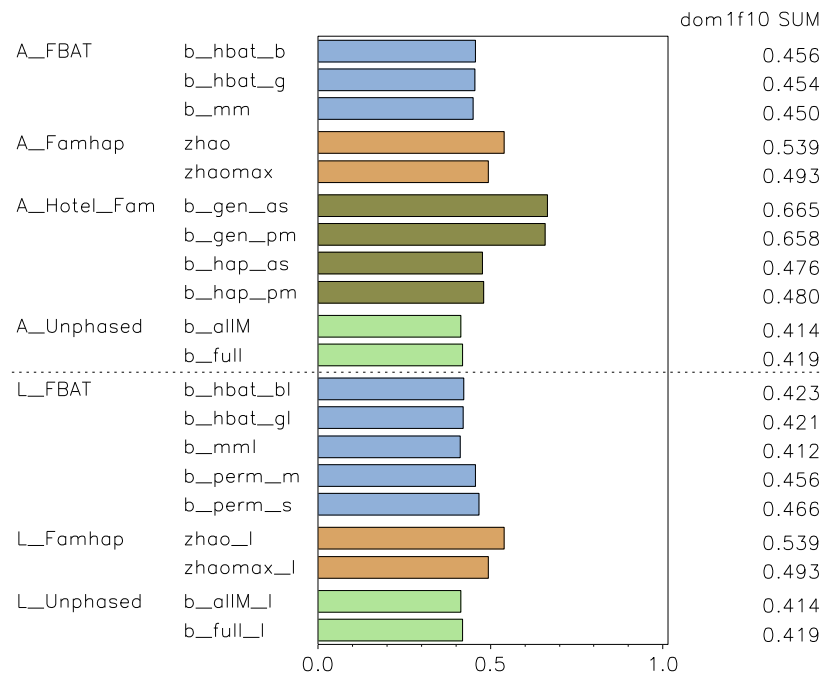
### Trios

Overall power: alpha=0.01, ADD, KMod1, 1100 trios



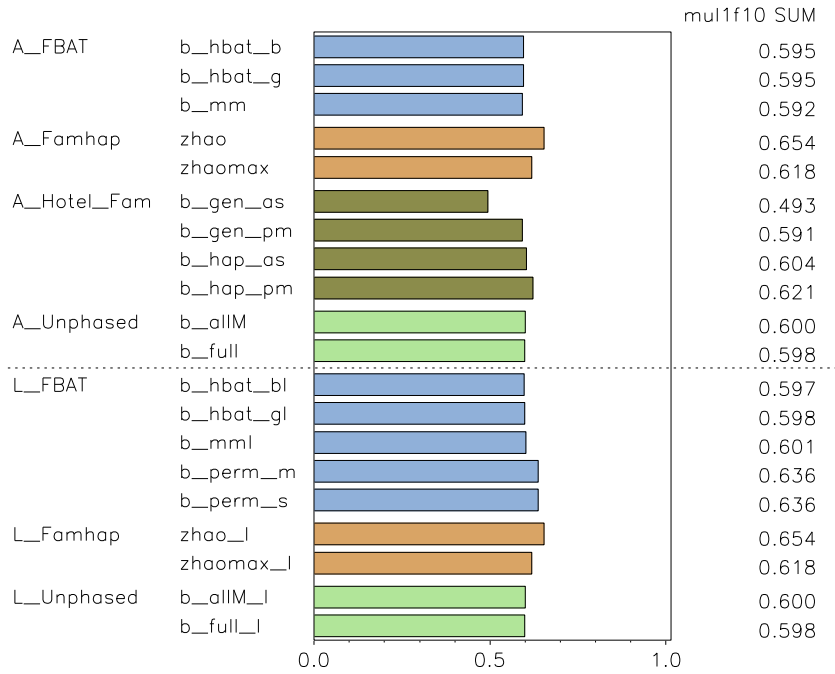
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Overall power: alpha=0.01, DOM, KMod1, 400 trios



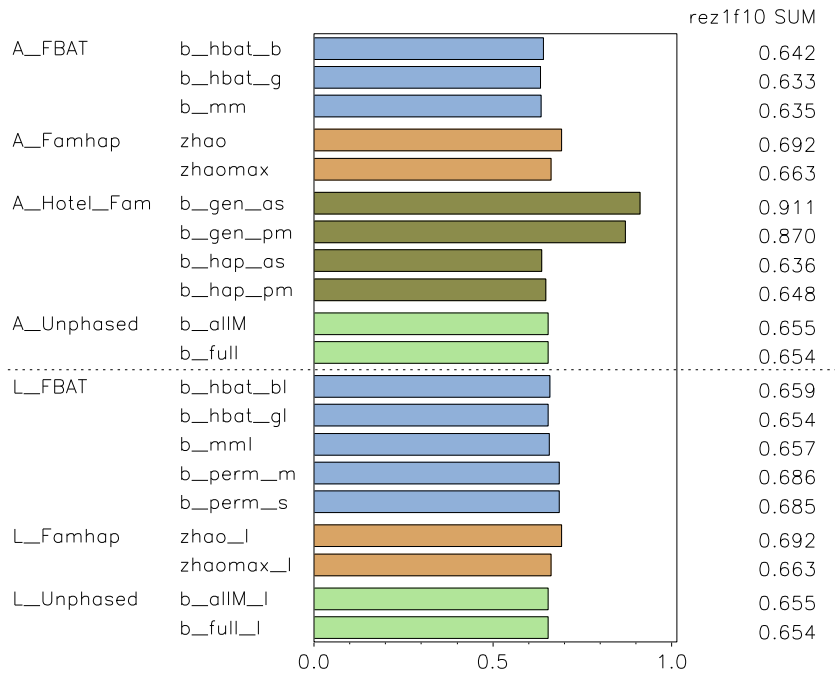
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Overall power: alpha=0.01, MUL, KMod1, 700 trios



Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

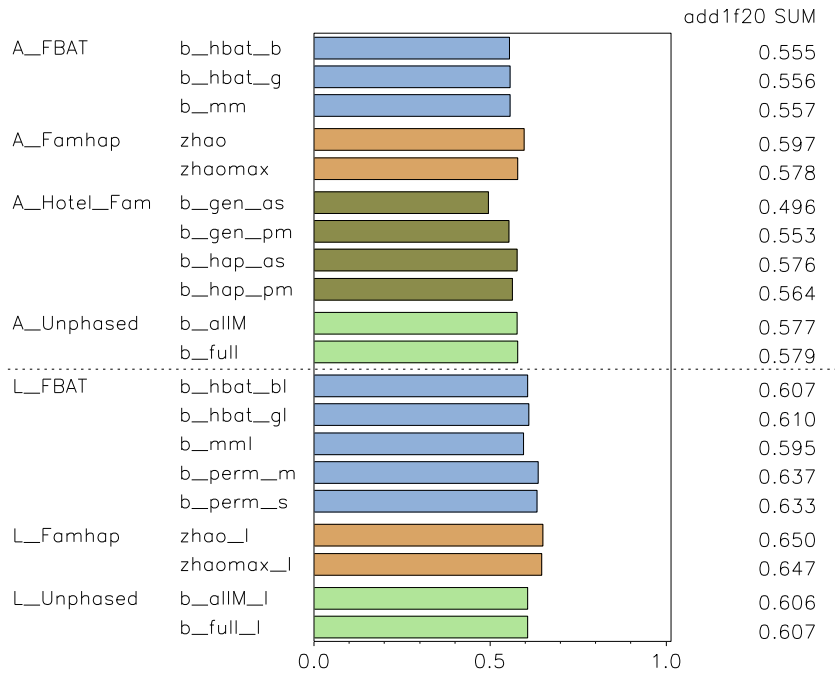
Overall power: alpha=0.01, REZ, KMod1, 2500 trios



Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

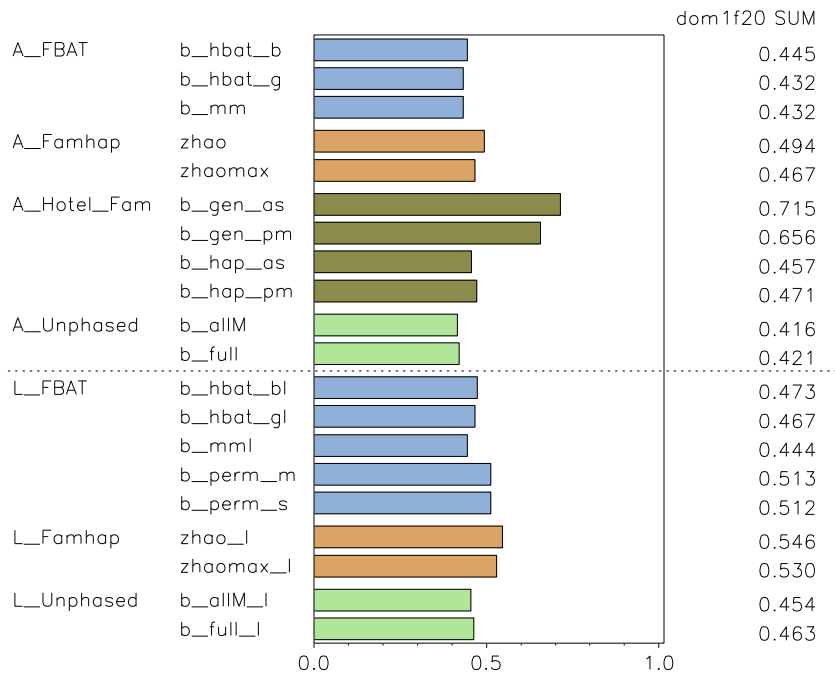
## AA-Familien

Overall power: alpha=0.01, ADD, KMod1, 300 AA-families



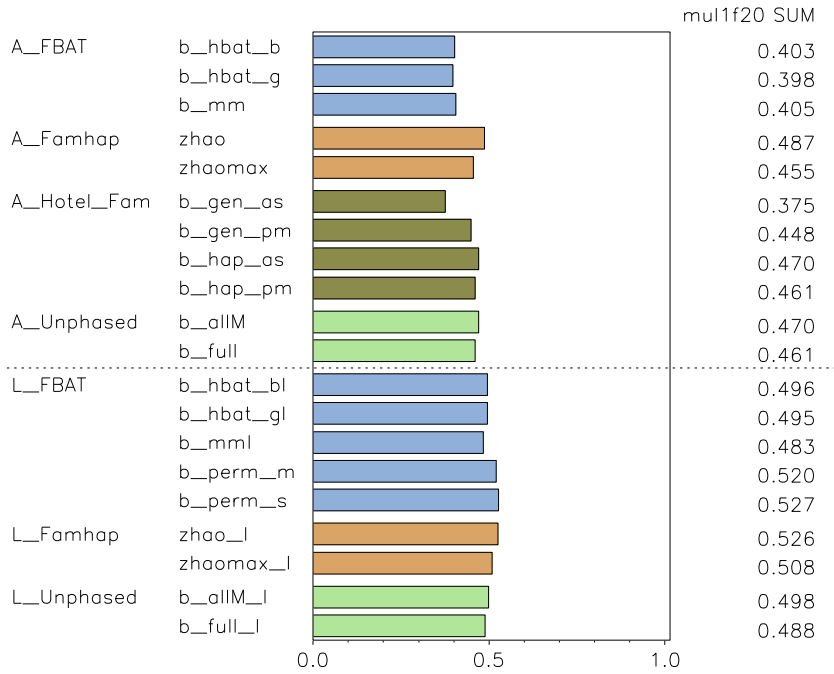
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Overall power: alpha=0.01, DOM, KMod1, 300 AA-families



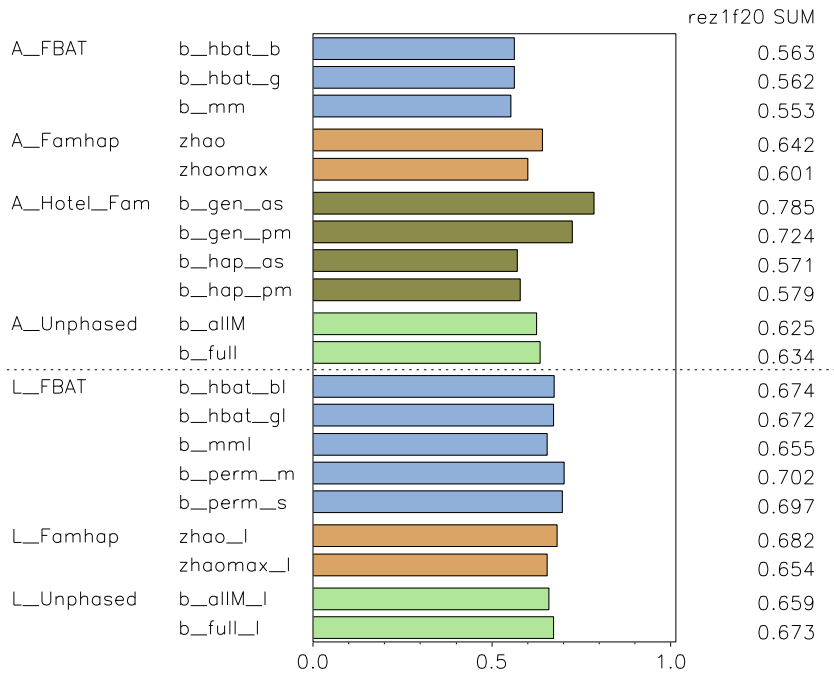
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Overall power: alpha=0.01, MUL, KMod1, 200 AA-families



Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Overall power: alpha=0.01, REZ, KMod1, 500 AA-families

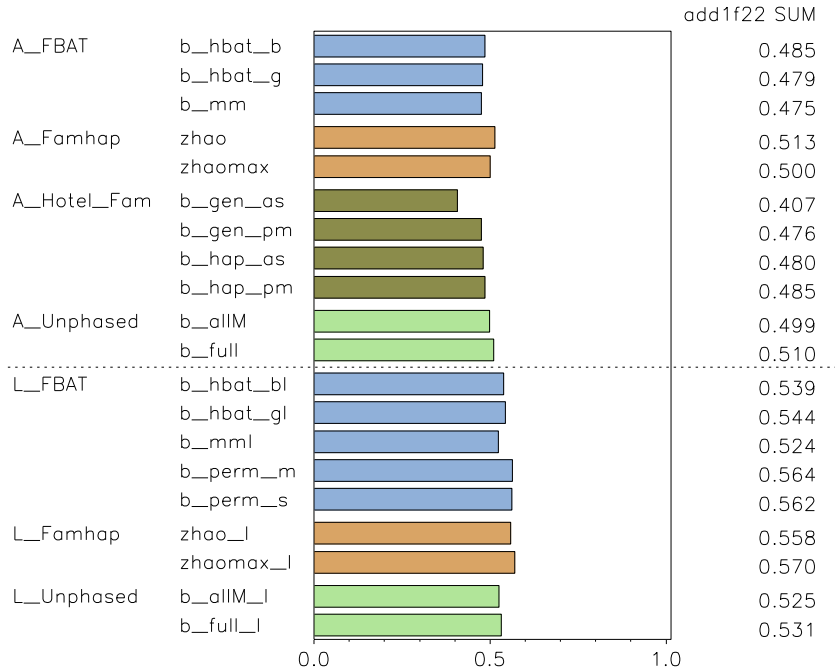


Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased



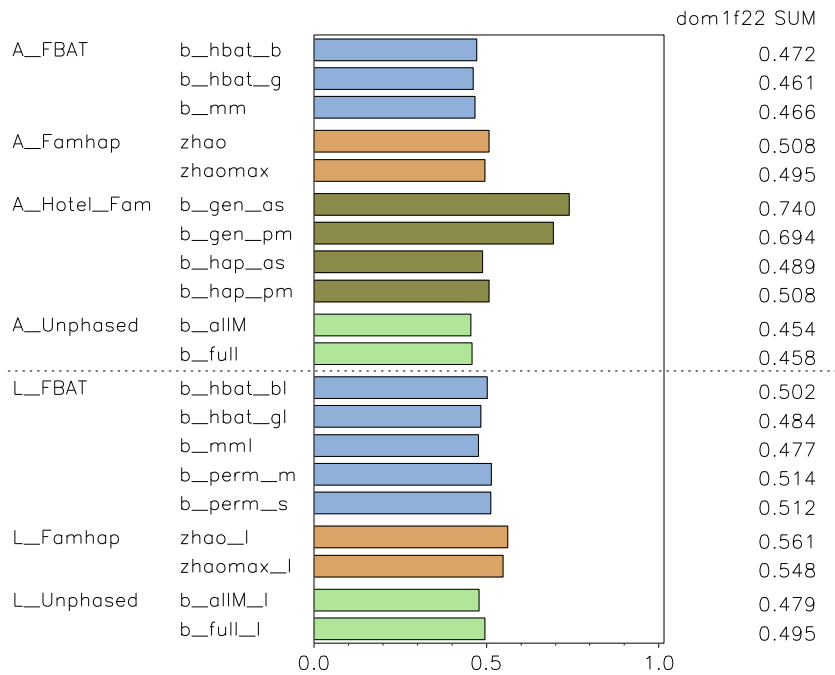
## AAUU-Familien

Overall power: alpha=0.01, ADD, KMod1, 300 AAUU-families



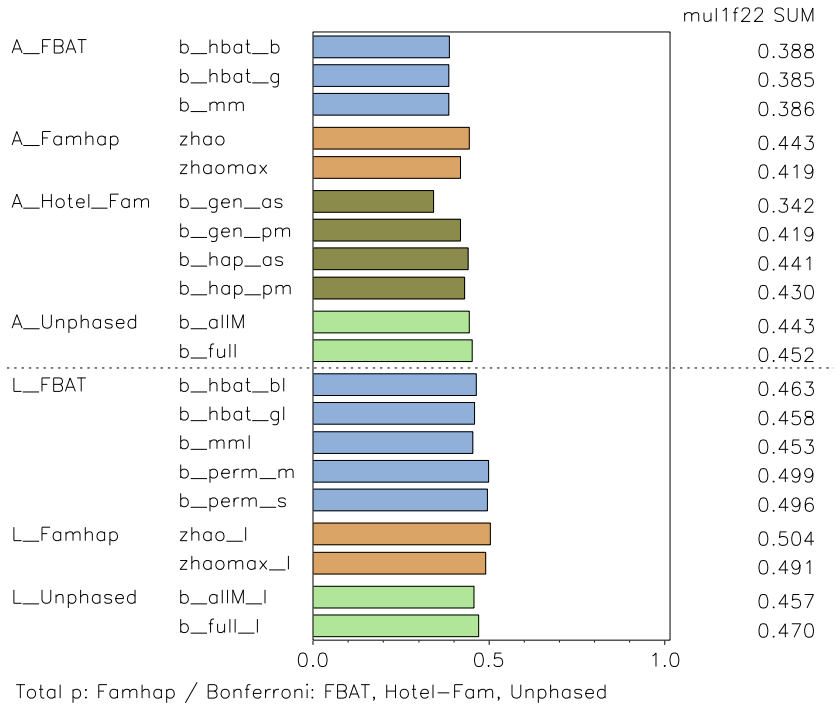
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Overall power: alpha=0.01, DOM, KMod1, 300 AAUU-families

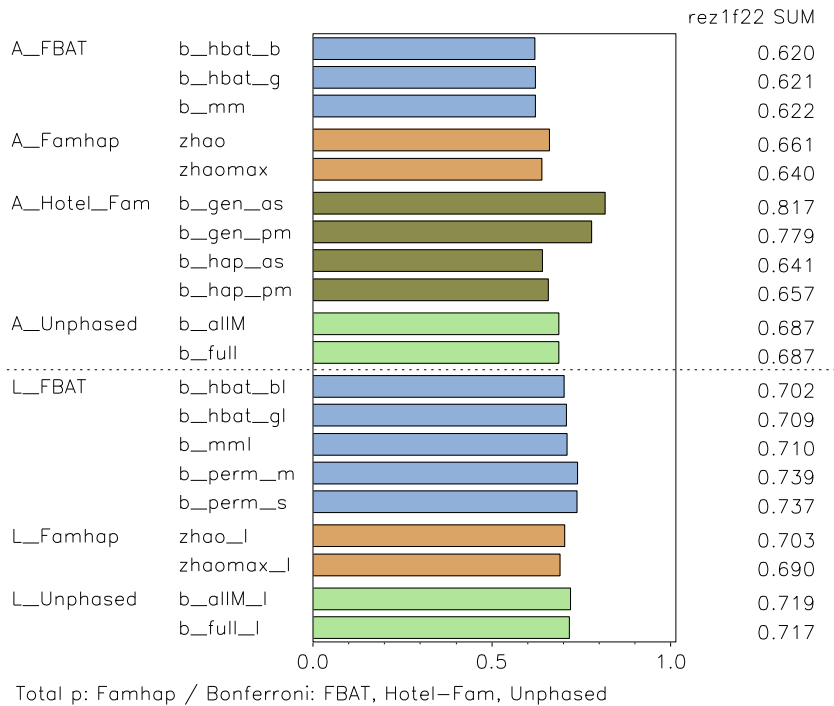


Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Overall power: alpha=0.01, MUL, KMod1, 200 AAU-families



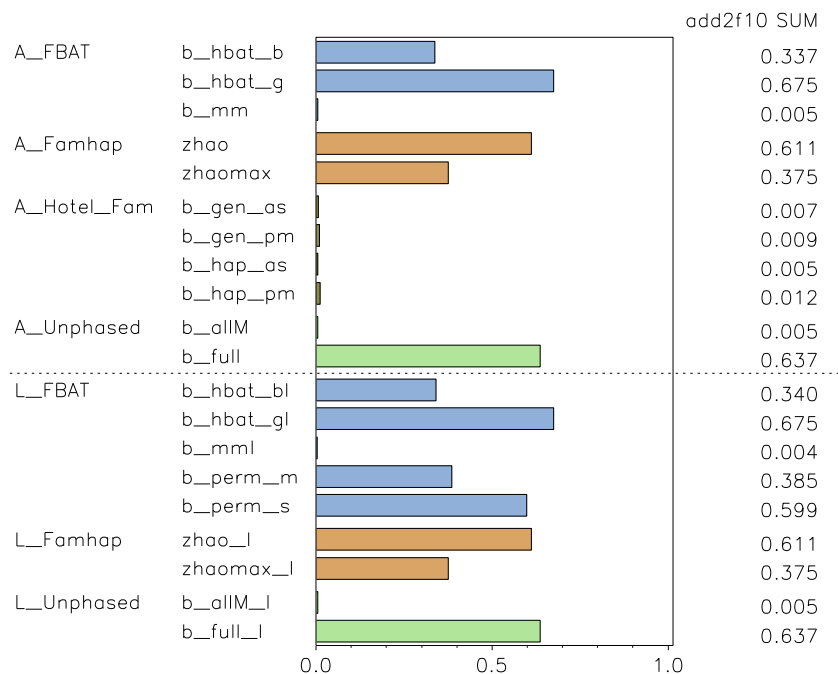
Overall power: alpha=0.01, REZ, KMod1, 600 AAU-families



## KMod 2

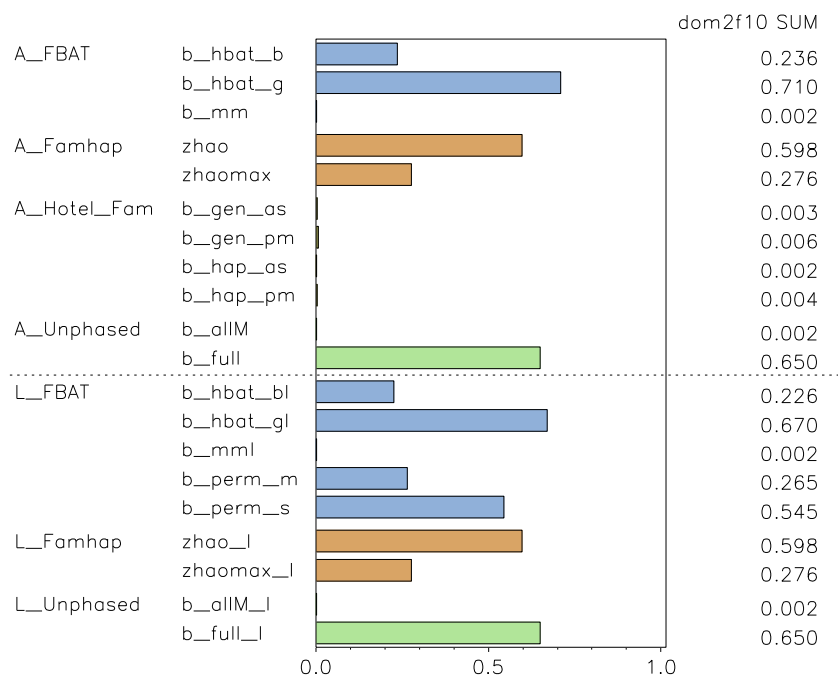
### Trios

Overall power: alpha=0.01, ADD, KMod2, 1800 trios



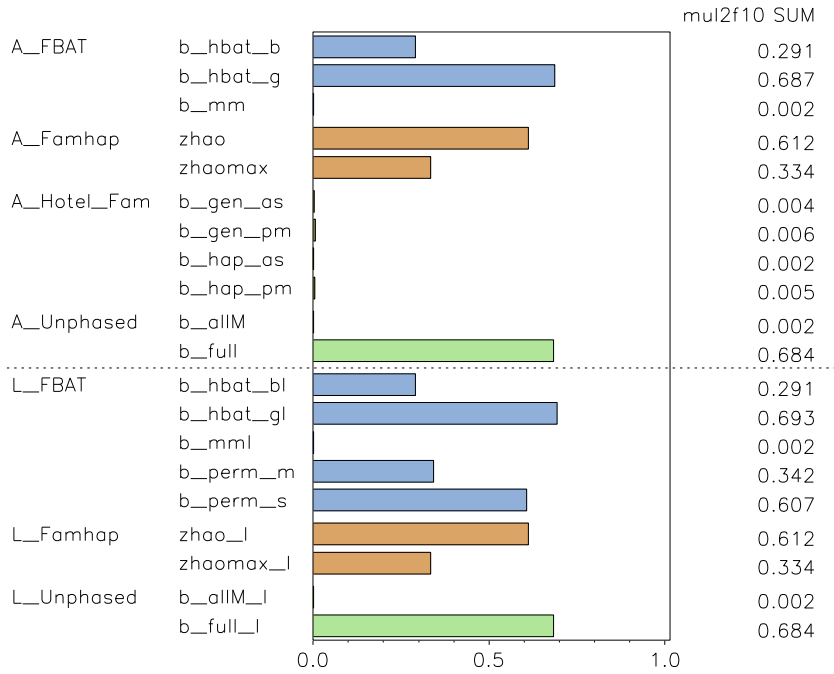
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Overall power: alpha=0.01, DOM, KMod2, 1000 trios



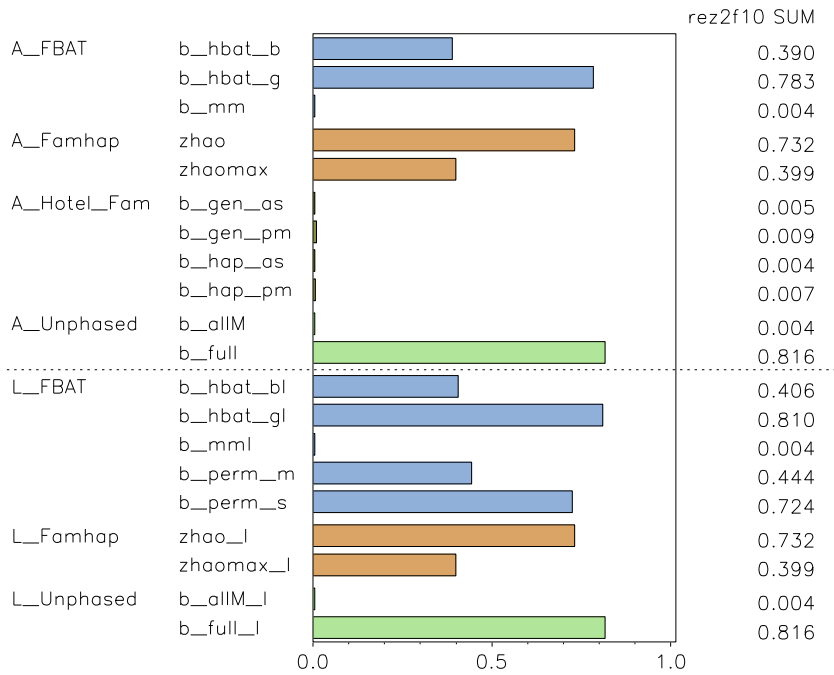
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Overall power: alpha=0.01, MUL, KMod2, 1200 trios



Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

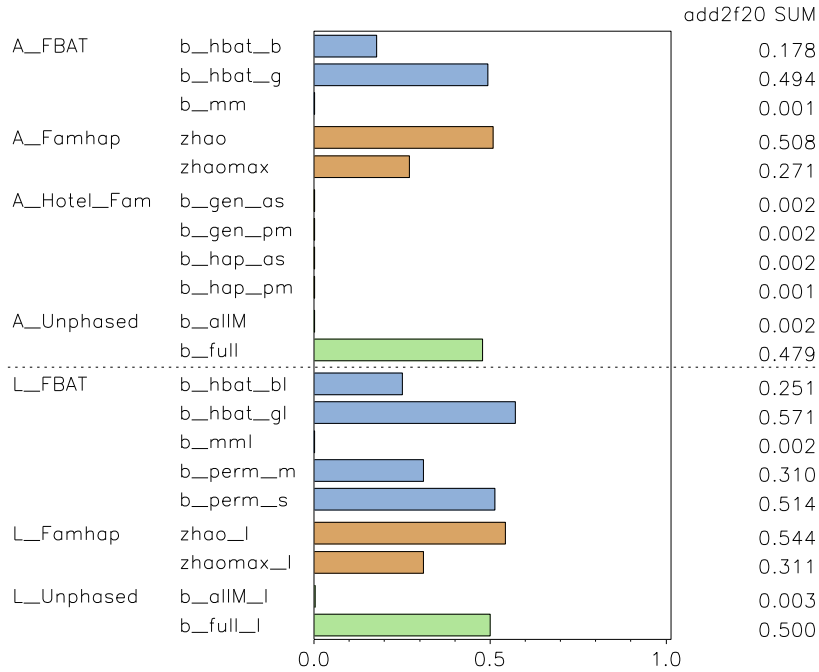
Overall power: alpha=0.01, REZ, KMod2, 6000 trios



Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

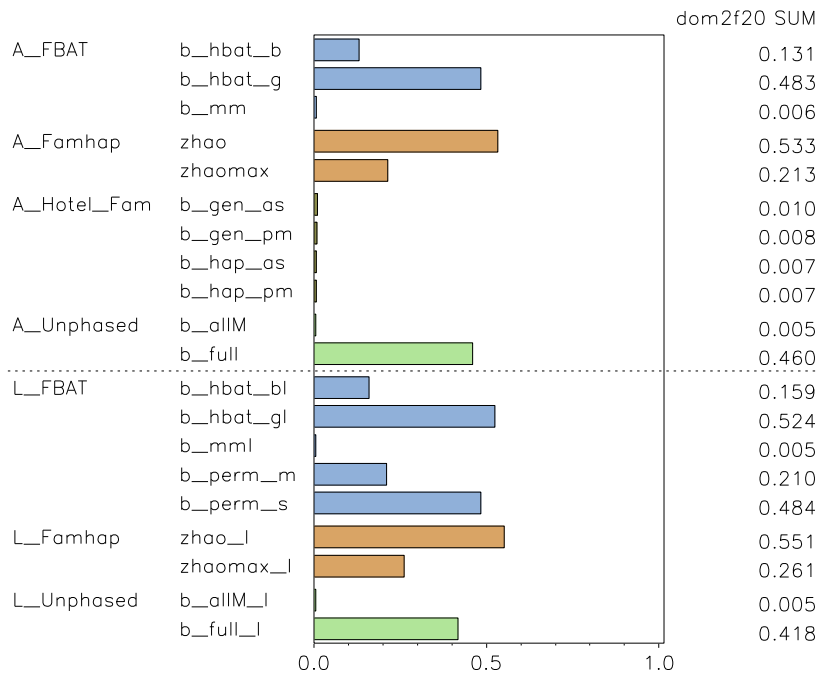
## AA-Familien

Overall power: alpha=0.01, ADD, KMod2, 400 AA-families



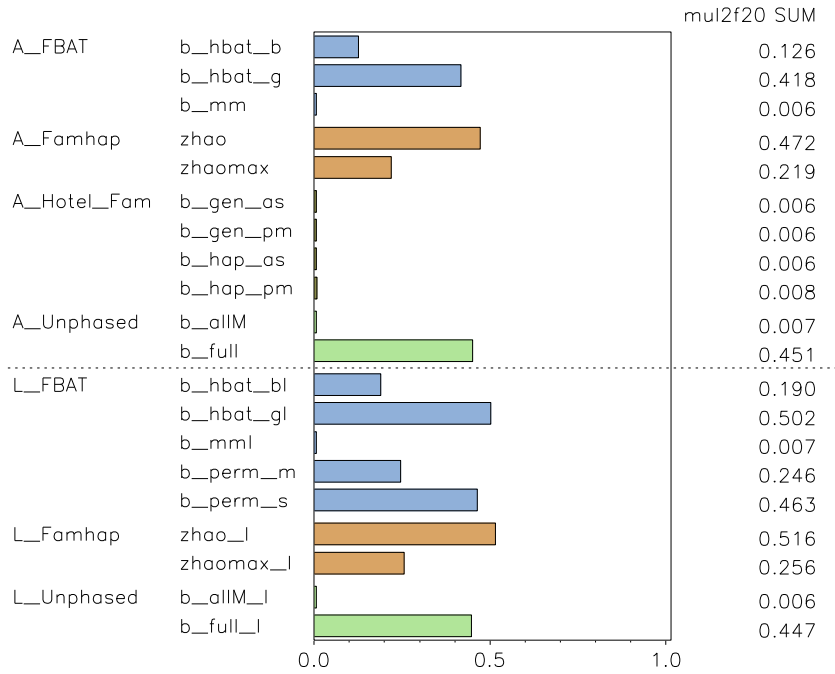
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Overall power: alpha=0.01, DOM, KMod2, 500 AA-families



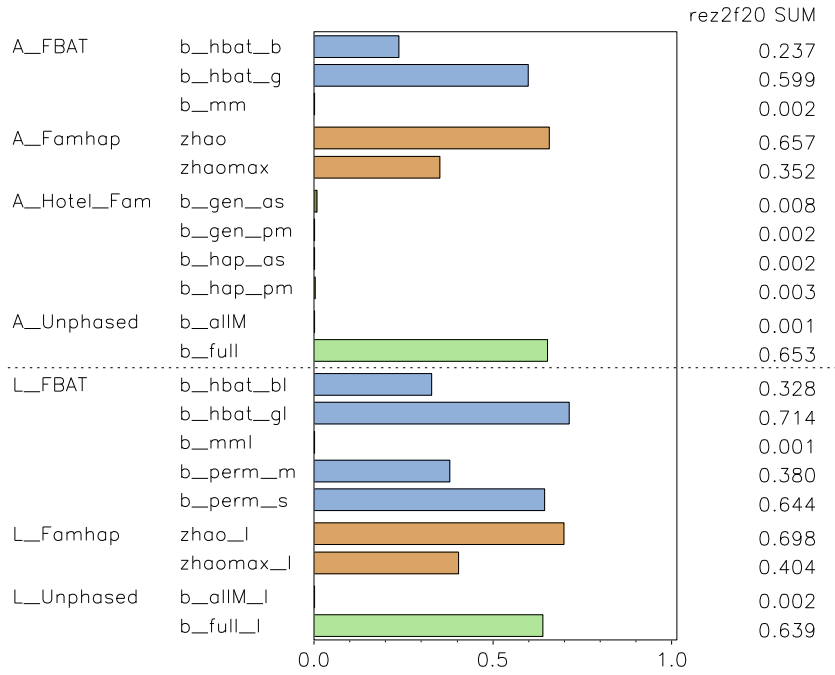
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Overall power: alpha=0.01, MUL, KMod2, 300 AA-families



Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

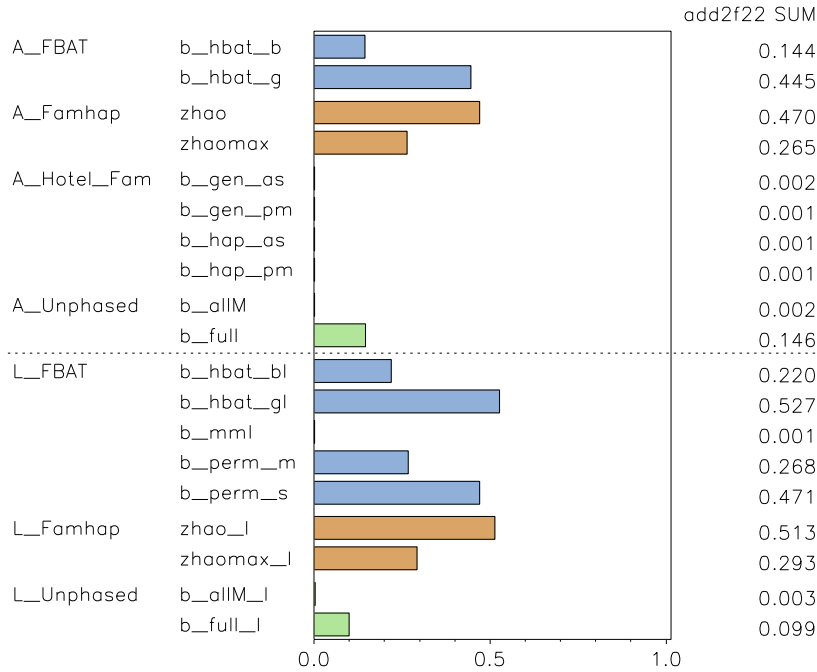
Overall power: alpha=0.01, REZ, KMod2, 900 AA-families



Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

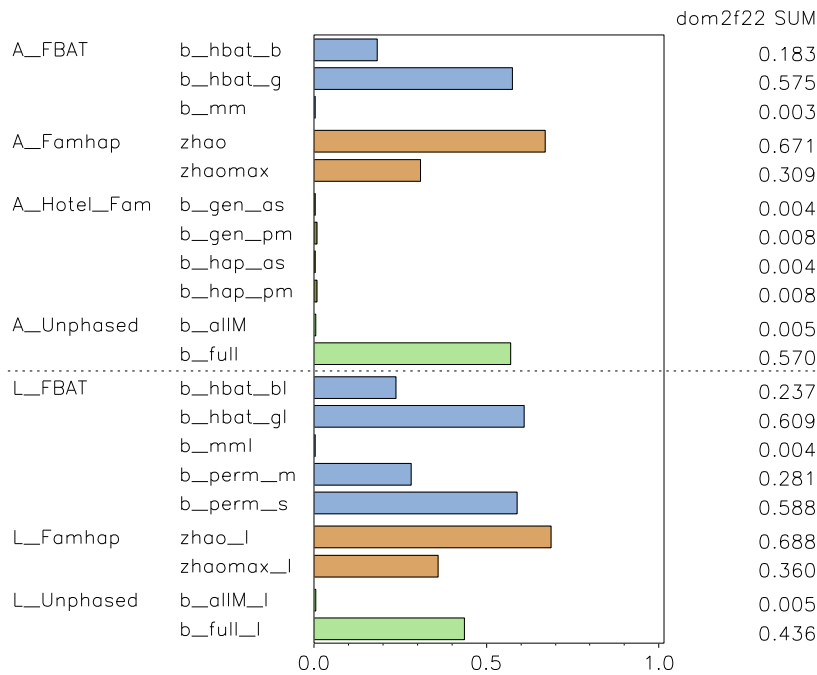
## AAUU-Familien

Overall power: alpha=0.01, ADD, KMod2, 400 AAUU-families



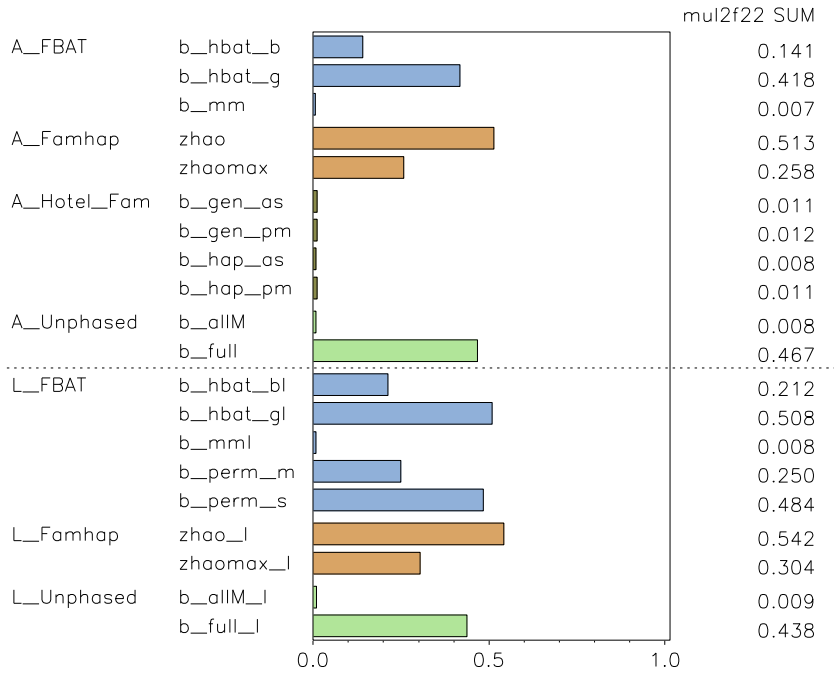
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Overall power: alpha=0.01, DOM, KMod2, 500 AAUU-families



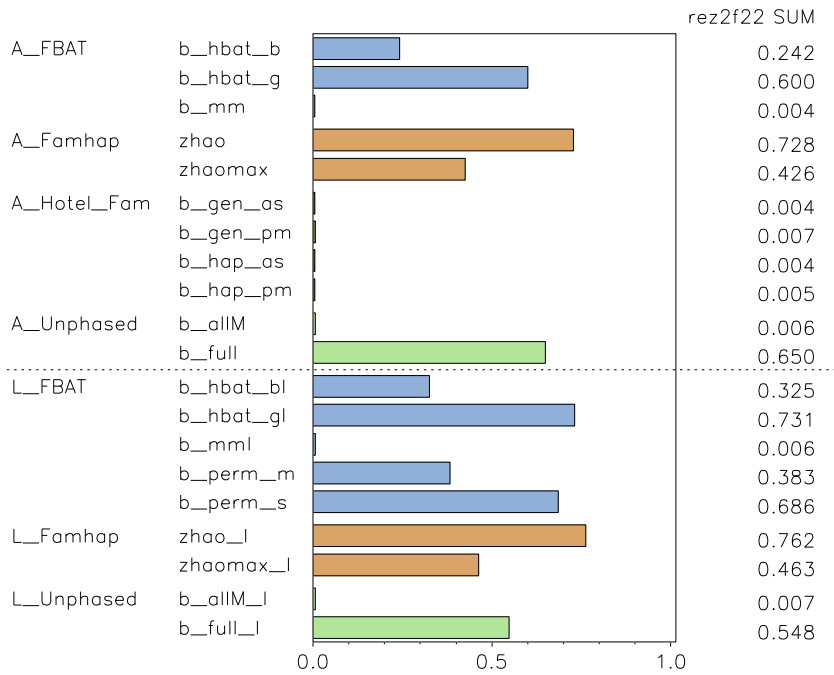
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Overall power: alpha=0.01, MUL, KMod2, 300 AAU-families



Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Overall power: alpha=0.01, REZ, KMod2, 900 AAU-families



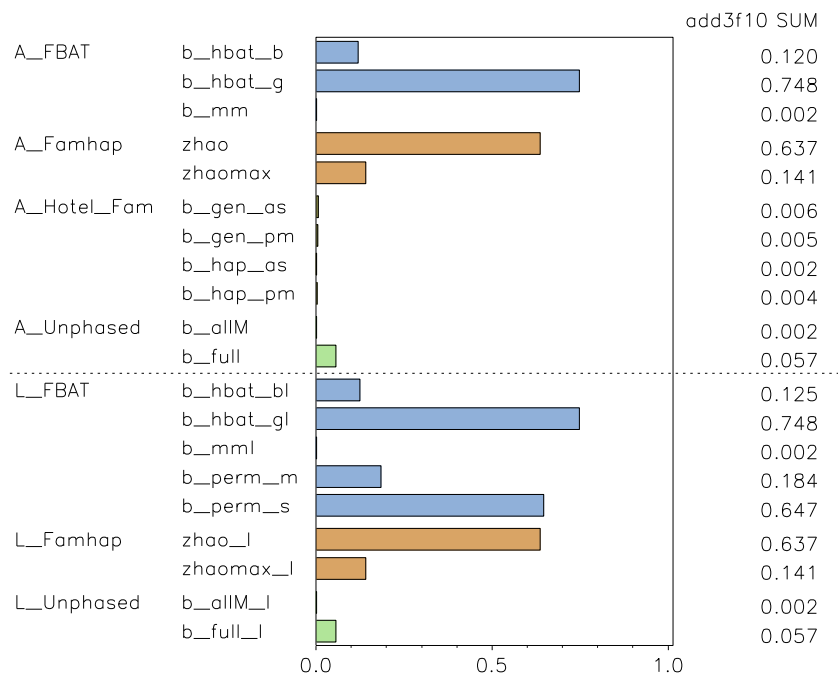
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased



### KMod 3

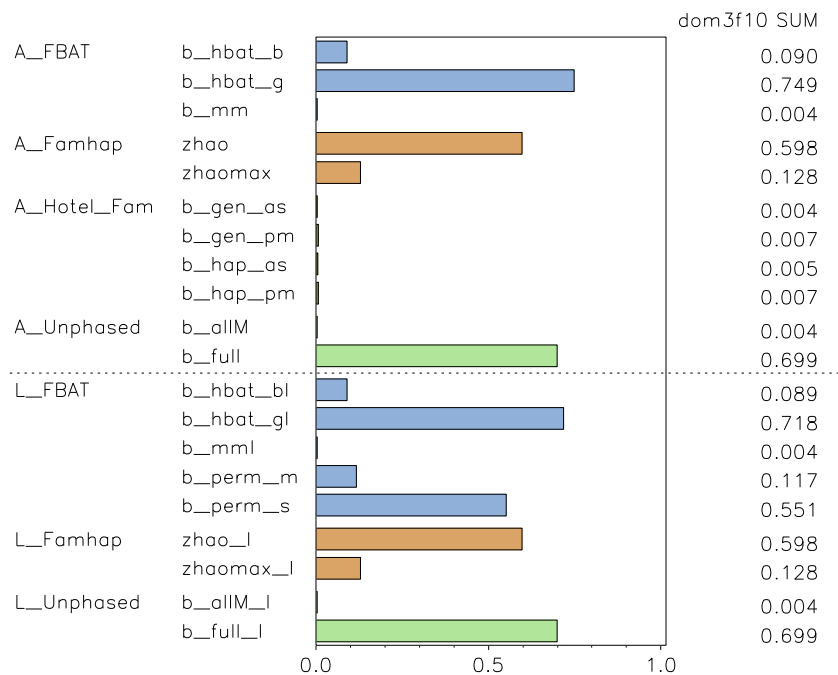
#### Trios

Overall power: alpha=0.01, ADD, KMod3, 2500 trios



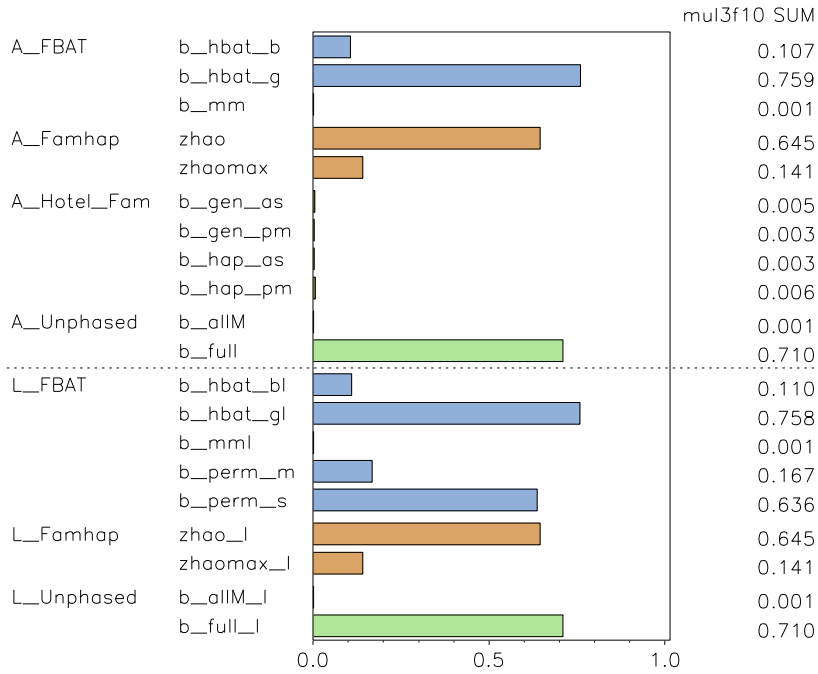
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Overall power: alpha=0.01, DOM, KMod3, 1500 trios



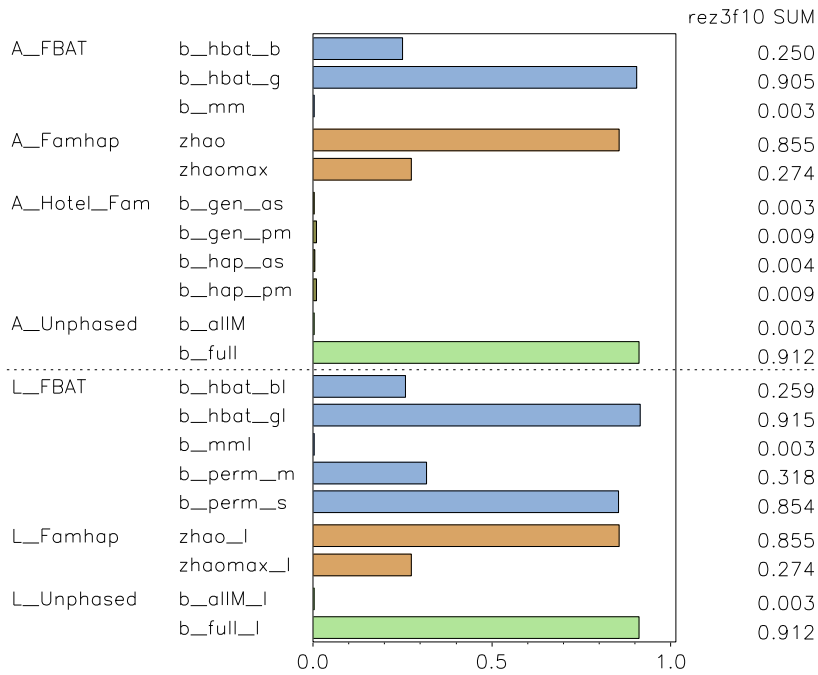
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Overall power: alpha=0.01, MUL, KMod3, 1700 trios



Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

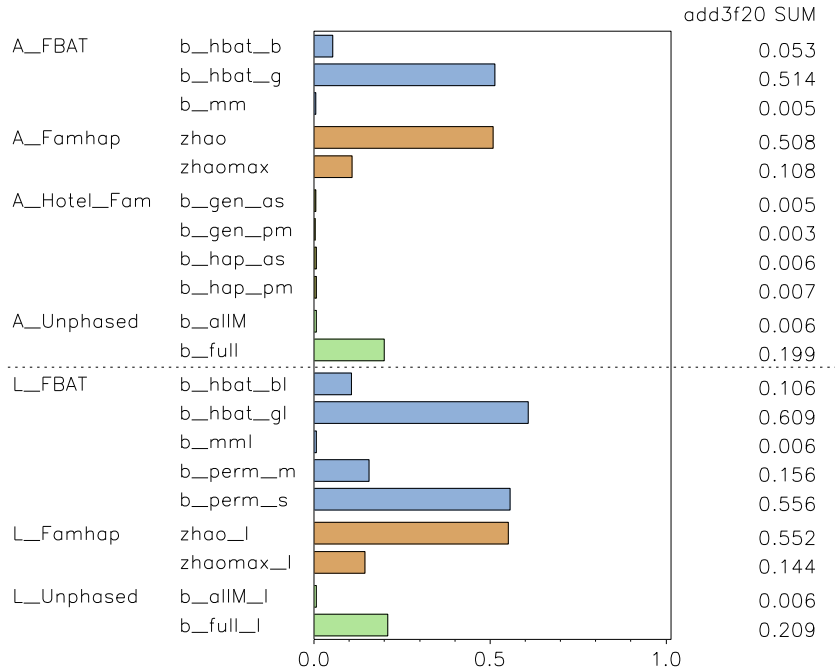
Overall power: alpha=0.01, REZ, KMod3, 11000 trios



Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

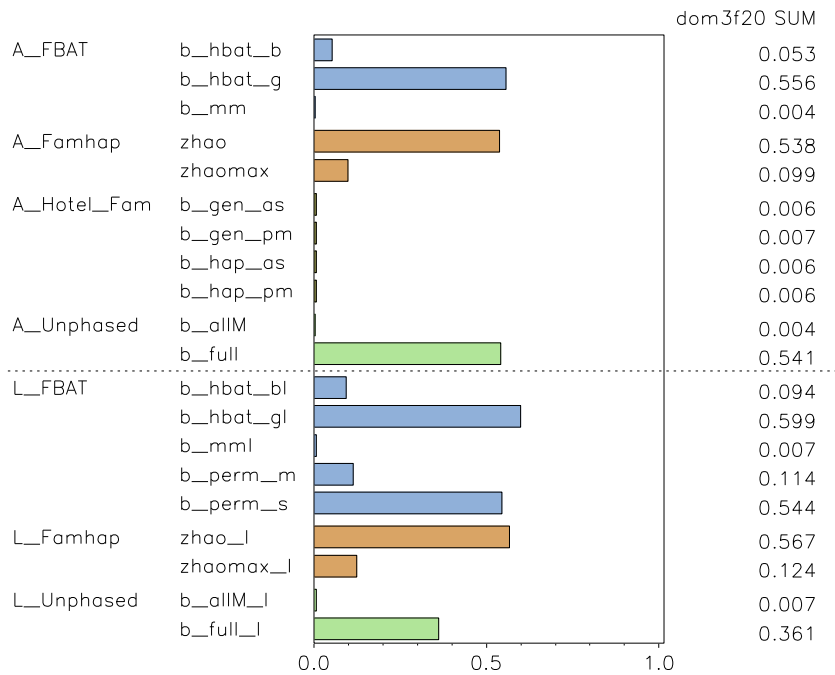
## AA-Familien

Overall power: alpha=0.01, ADD, KMod3, 500 AA-families



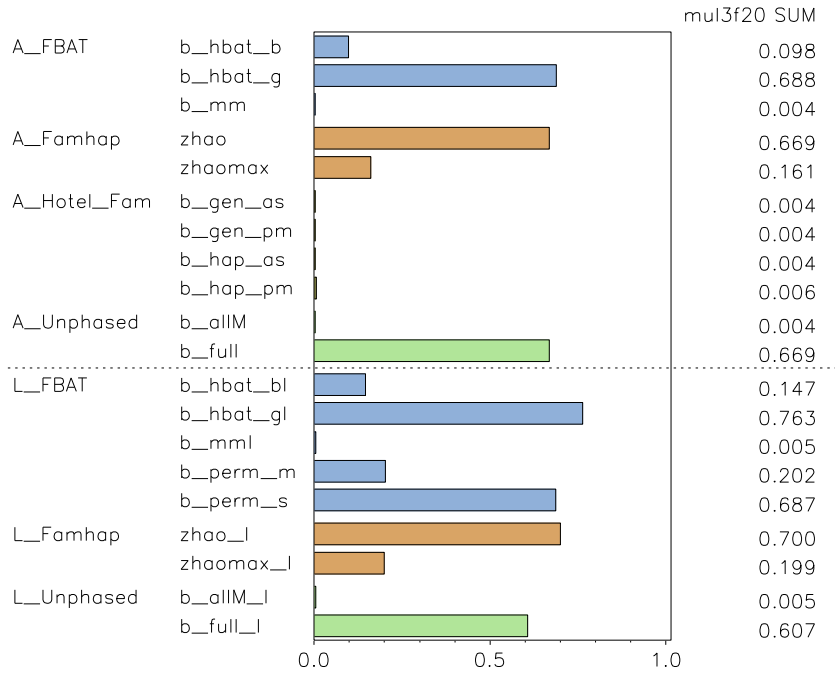
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Overall power: alpha=0.01, DOM, KMod3, 700 AA-families



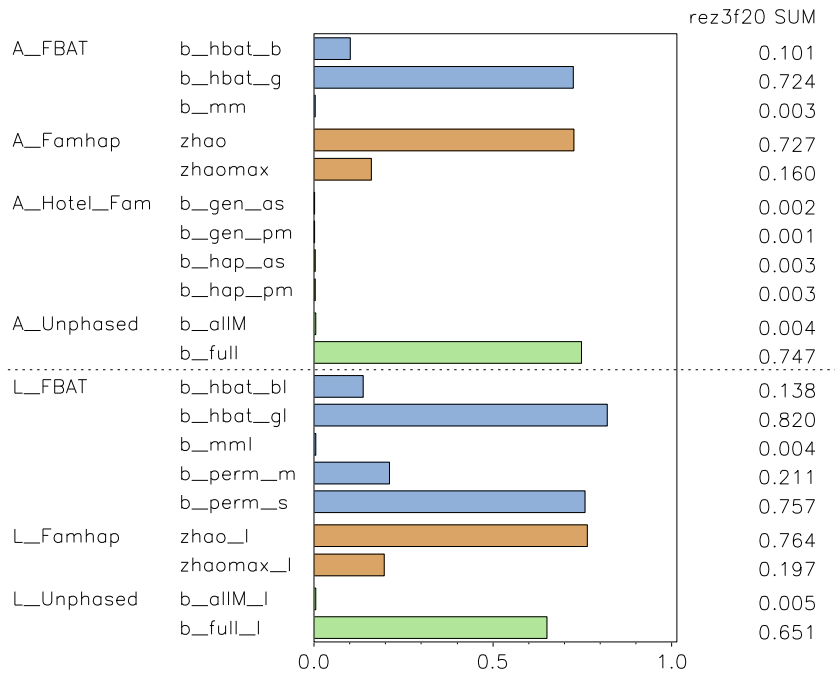
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Overall power: alpha=0.01, MUL, KMod3, 500 AA-families



Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

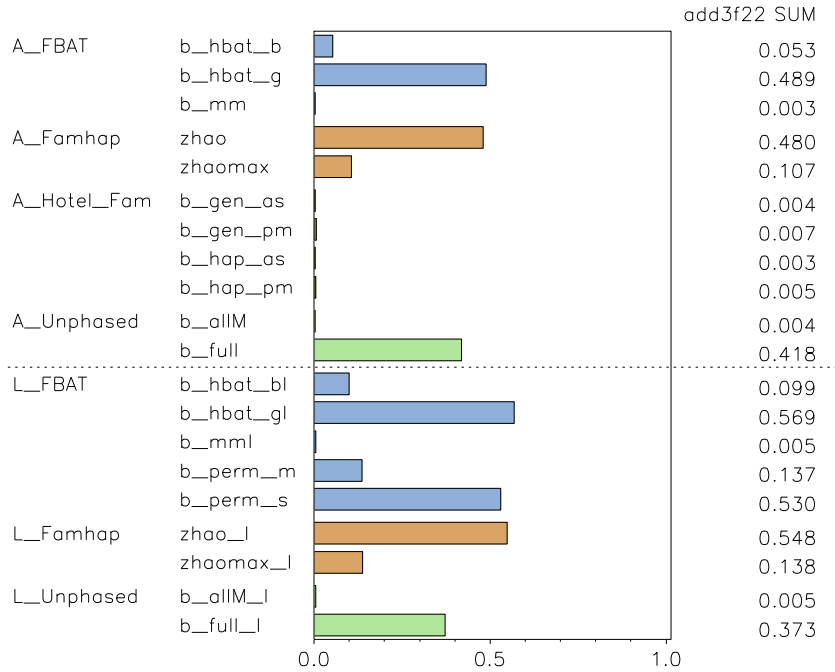
Overall power: alpha=0.01, REZ, KMod3, 1400 AA-families



Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

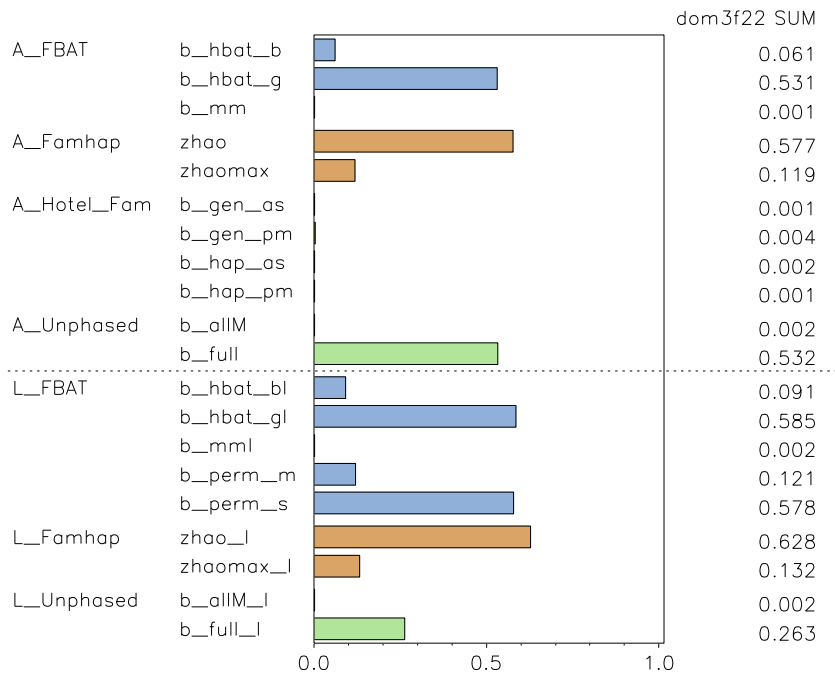
## AAUU-Familien

Overall power: alpha=0.01, ADD, KMod3, 500 AAUU-families



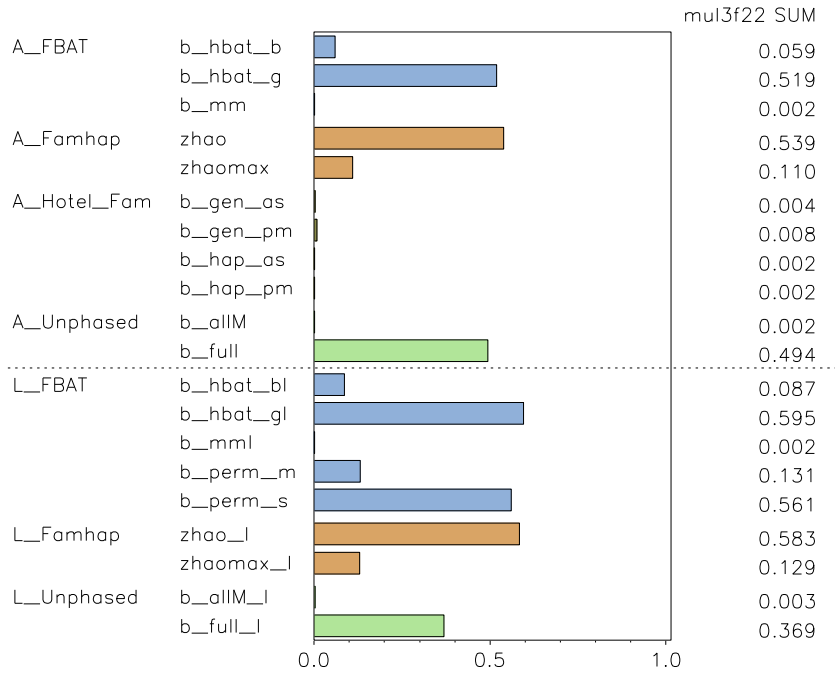
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Overall power: alpha=0.01, DOM, KMod3, 600 AAUU-families



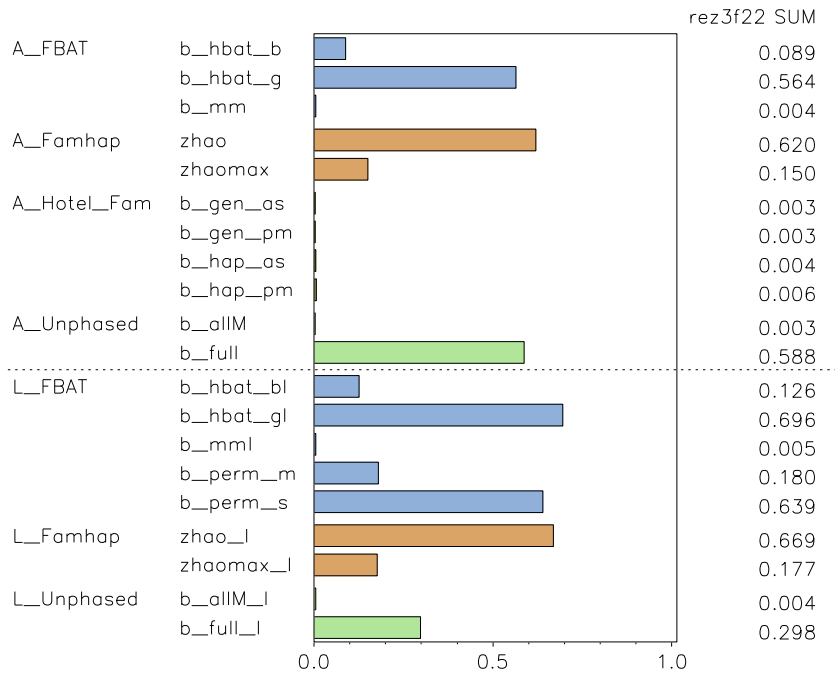
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Overall power: alpha=0.01, MUL, KMod3, 400 AAU-families



Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Overall power: alpha=0.01, REZ, KMod3, 1000 AAU-families



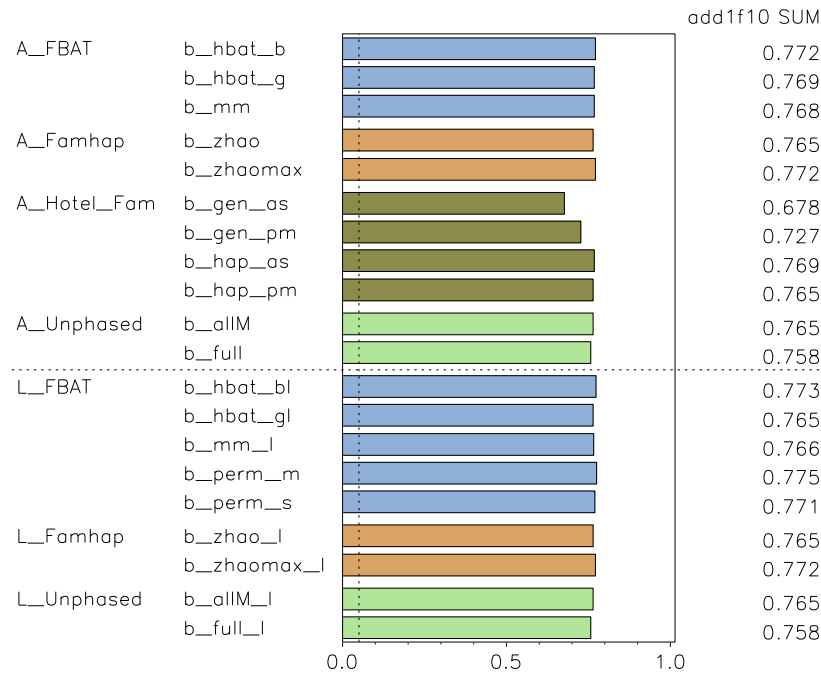
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

## Overall-Power, $\alpha = 0.05$

### Alle Programme: Bonferroni-korrigierte minimale p-Werte über die Mk; KMod 1

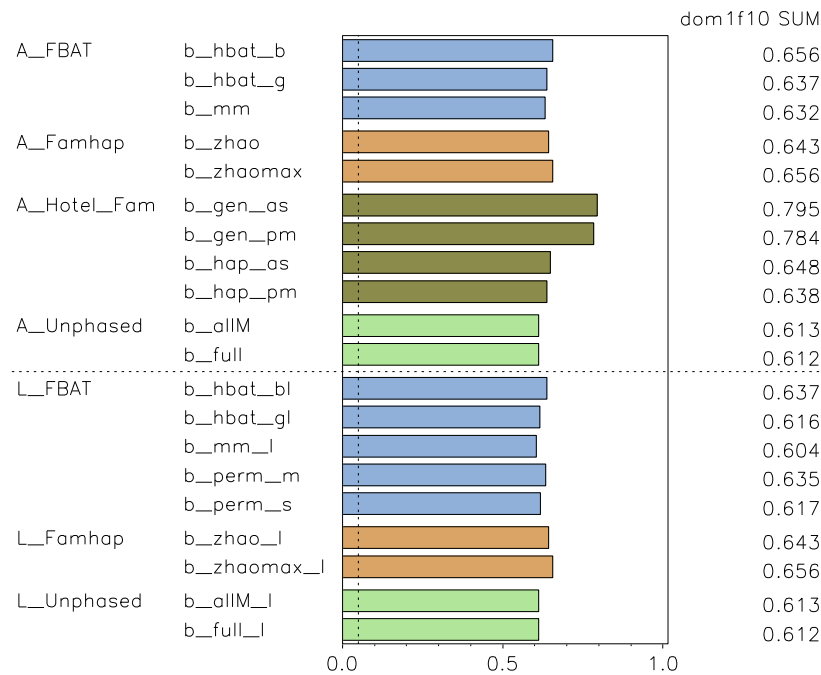
#### Trios

Power: alpha=0.05, ADD, KMod1, 1100 trios



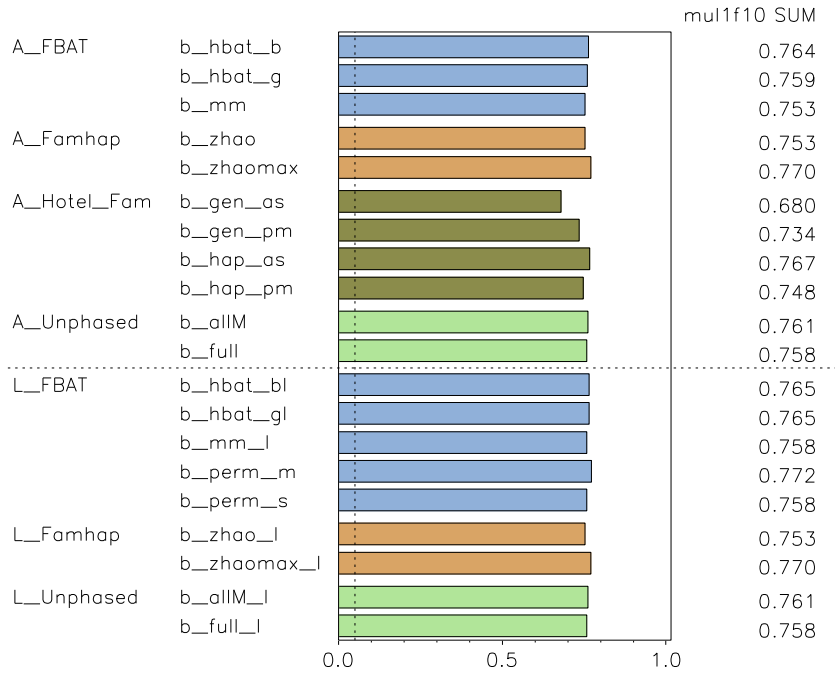
Bonferroni-corrected min\_p: Famhap, FBAT, Hotel-Fam, Unphased

Power: alpha=0.05, DOM, KMod1, 400 trios



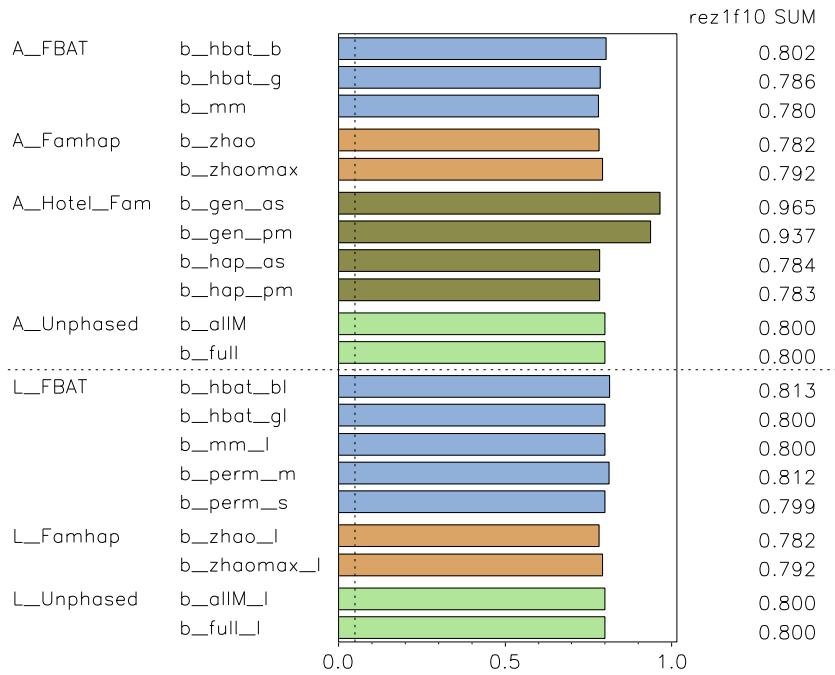
Bonferroni-corrected min\_p: Famhap, FBAT, Hotel-Fam, Unphased

Power: alpha=0.05, MUL, KMod1, 700 trios



Bonferroni-corrected min\_p: Famhap, FBAT, Hotel-Fam, Unphased

Power: alpha=0.05, REZ, KMod1, 2500 trios

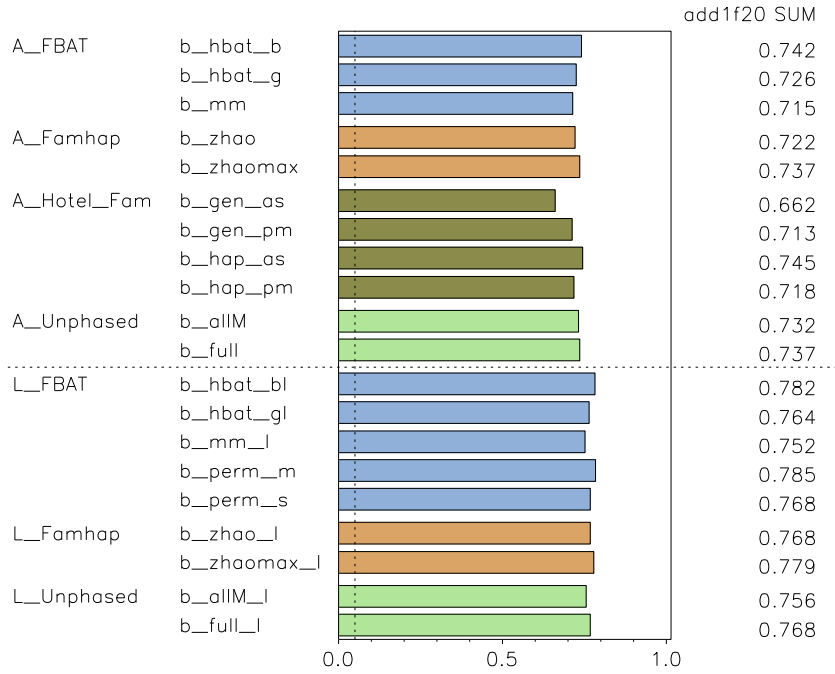


Bonferroni-corrected min\_p: Famhap, FBAT, Hotel-Fam, Unphased



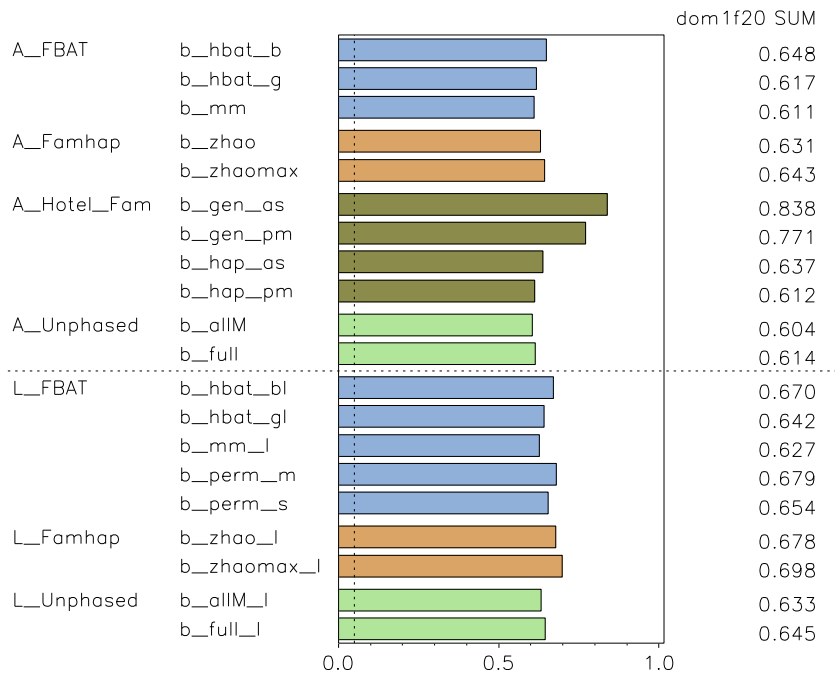
## AA-Familien

Power: alpha=0.05, ADD, KMod1, 300 AA-families



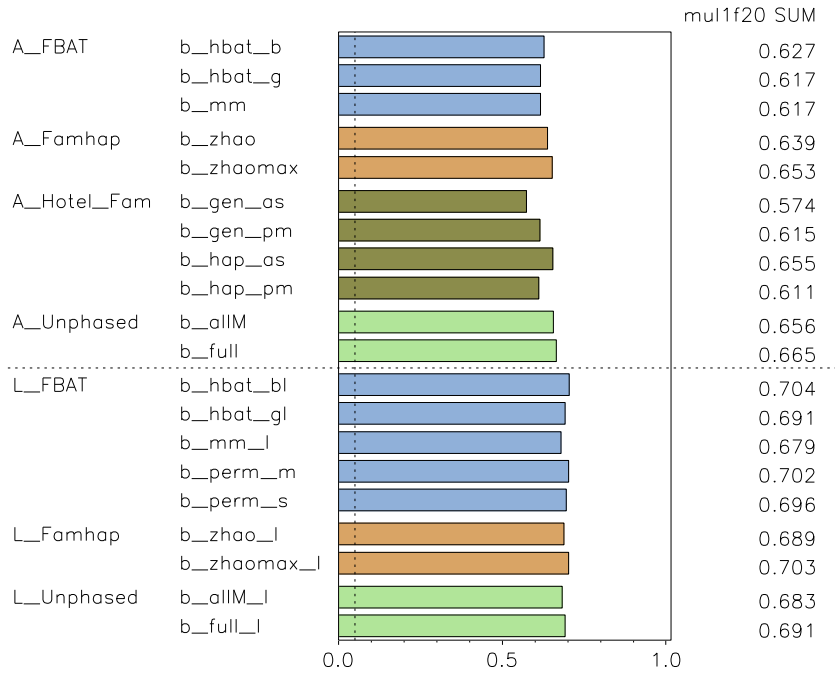
Bonferroni-corrected min<sub>p</sub>: Famhap, FBAT, Hotel-Fam, Unphased

Power: alpha=0.05, DOM, KMod1, 300 AA-families



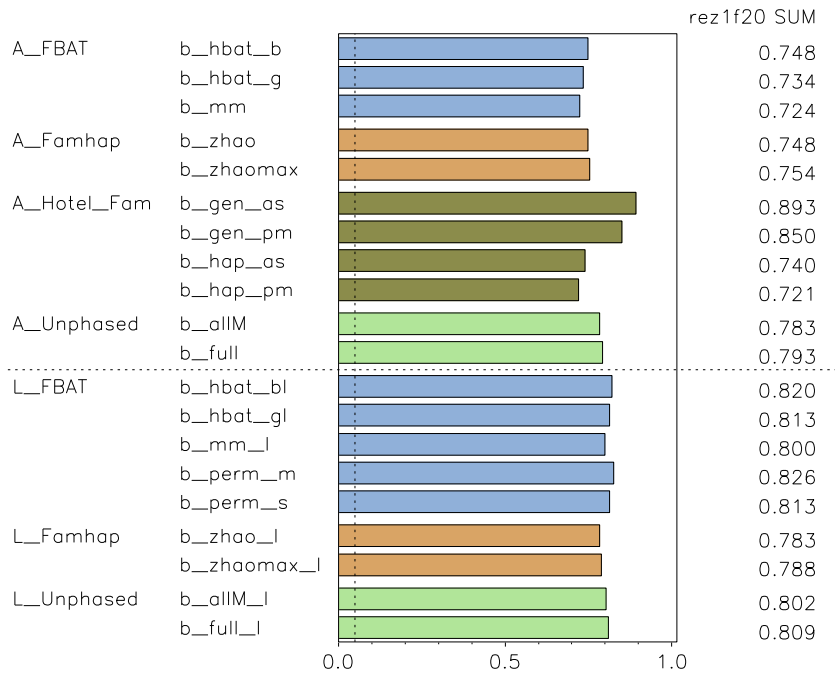
Bonferroni-corrected min<sub>p</sub>: Famhap, FBAT, Hotel-Fam, Unphased

Power: alpha=0.05, MUL, KMod1, 200 AA-families



Bonferroni-corrected min<sub>p</sub>: Famhap, FBAT, Hotel-Fam, Unphased

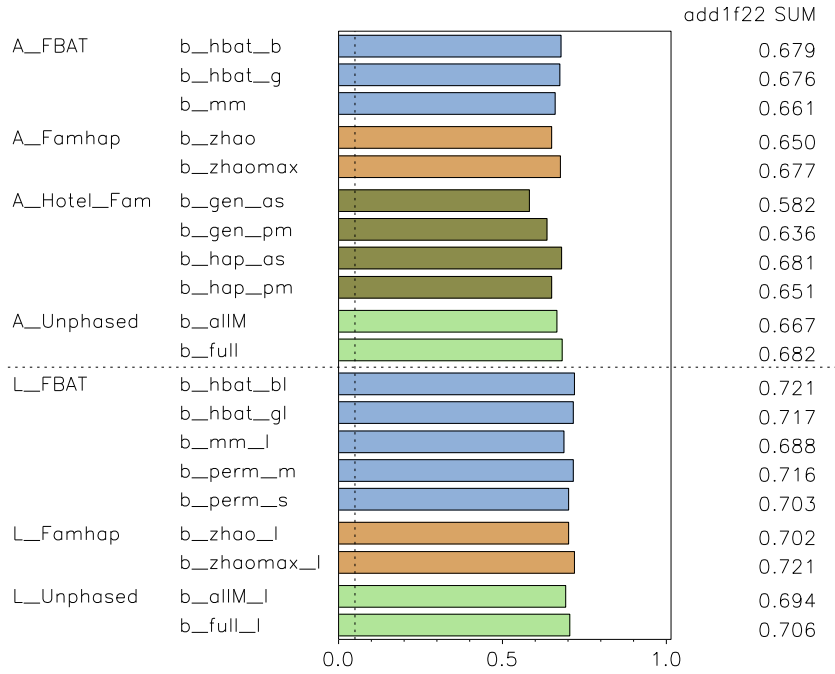
Power: alpha=0.05, REZ, KMod1, 500 AA-families



Bonferroni-corrected min<sub>p</sub>: Famhap, FBAT, Hotel-Fam, Unphased

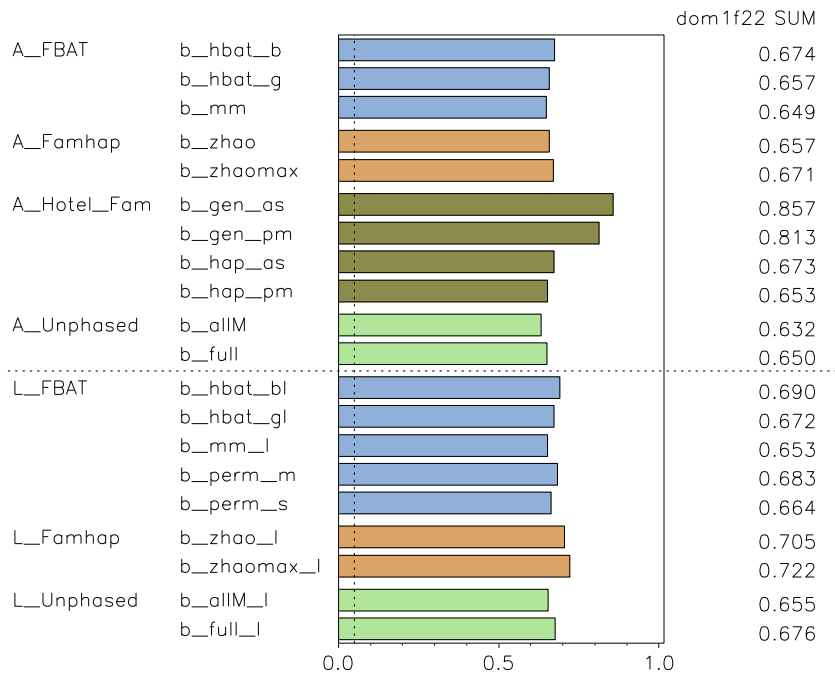
### AAUU-Familien

Power: alpha=0.05, ADD, KMod1, 300 AAUU-families



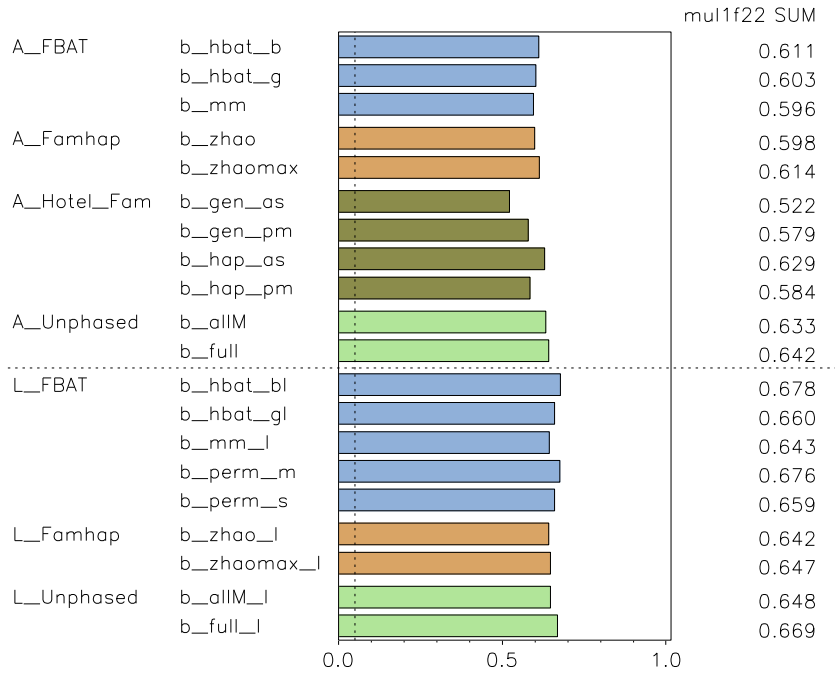
Bonferroni-corrected min\_p: Famhap, FBAT, Hotel-Fam, Unphased

Power: alpha=0.05, DOM, KMod1, 300 AAUU-families



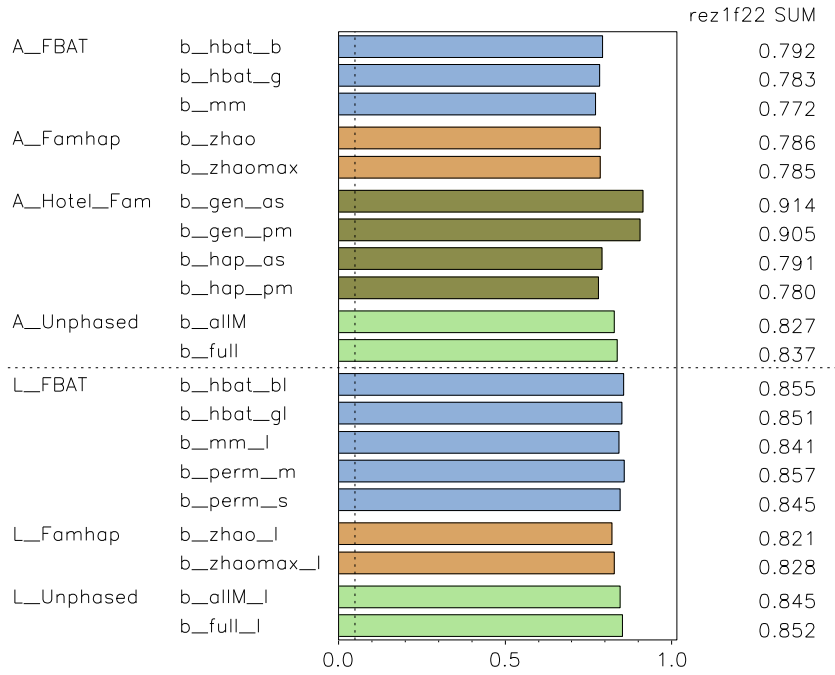
Bonferroni-corrected min\_p: Famhap, FBAT, Hotel-Fam, Unphased

Power: alpha=0.05, MUL, KMod1, 200 AAU-families



Bonferroni-corrected min<sub>p</sub>: Famhap, FBAT, Hotel-Fam, Unphased

Power: alpha=0.05, REZ, KMod1, 600 AAU-families

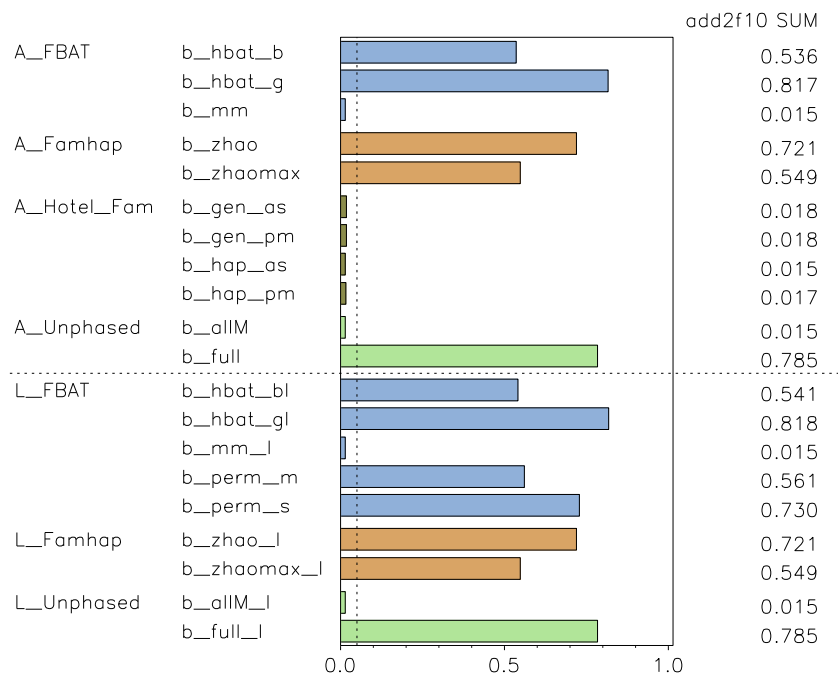


Bonferroni-corrected min<sub>p</sub>: Famhap, FBAT, Hotel-Fam, Unphased

## KMod 2

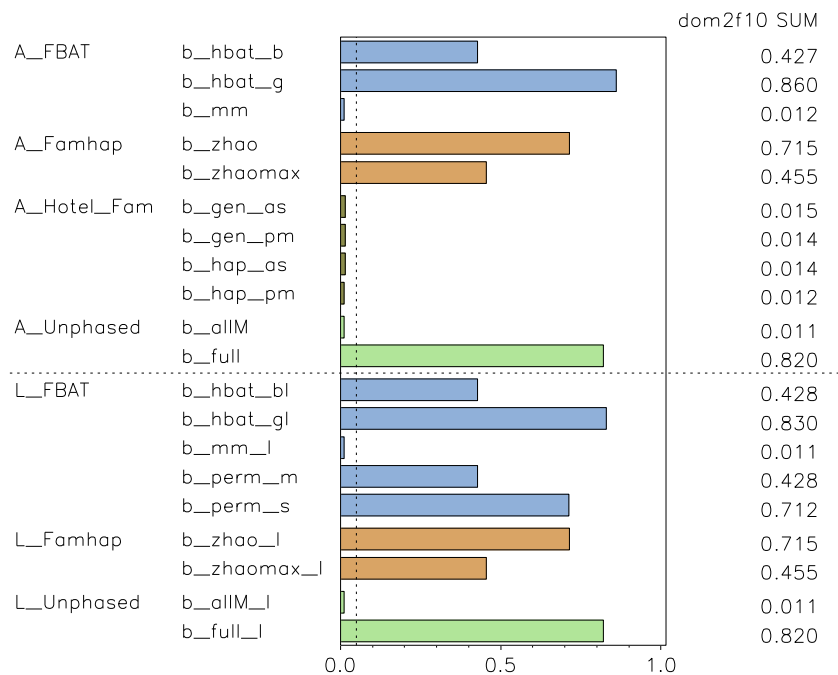
### Trios

Power: alpha=0.05, ADD, KMod2, 1800 trios



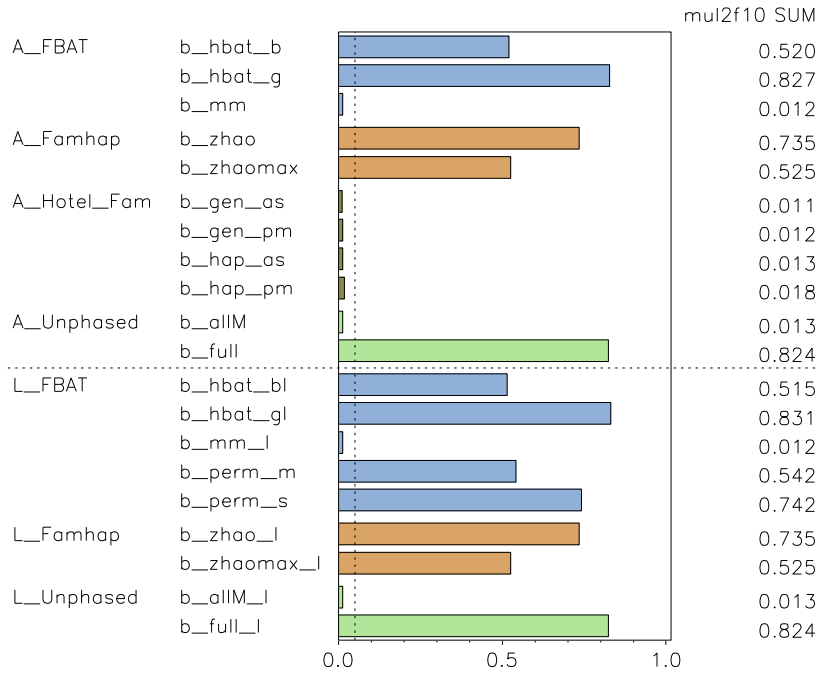
Bonferroni-corrected min\_p: Famhap, FBAT, Hotel-Fam, Unphased

Power: alpha=0.05, DOM, KMod2, 1000 trios



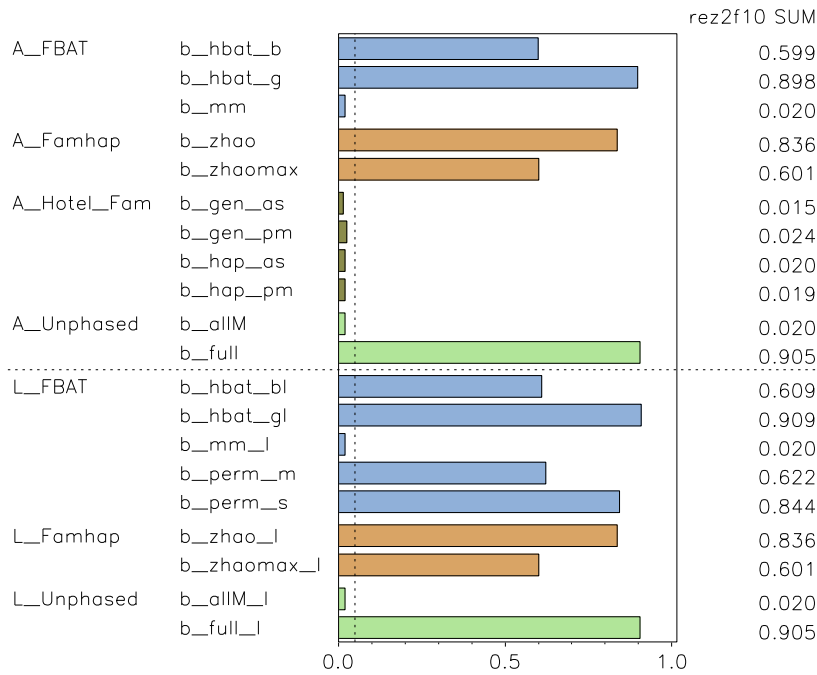
Bonferroni-corrected min\_p: Famhap, FBAT, Hotel-Fam, Unphased

Power: alpha=0.05, MUL, KMod2, 1200 trios



Bonferroni-corrected min\_p: Famhap, FBAT, Hotel-Fam, Unphased

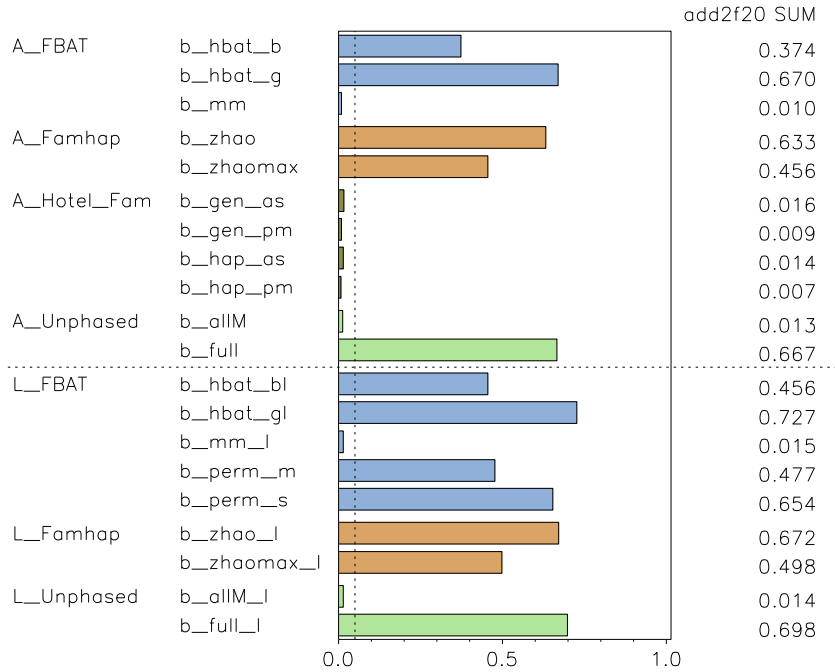
Power: alpha=0.05, REZ, KMod2, 6000 trios



Bonferroni-corrected min\_p: Famhap, FBAT, Hotel-Fam, Unphased

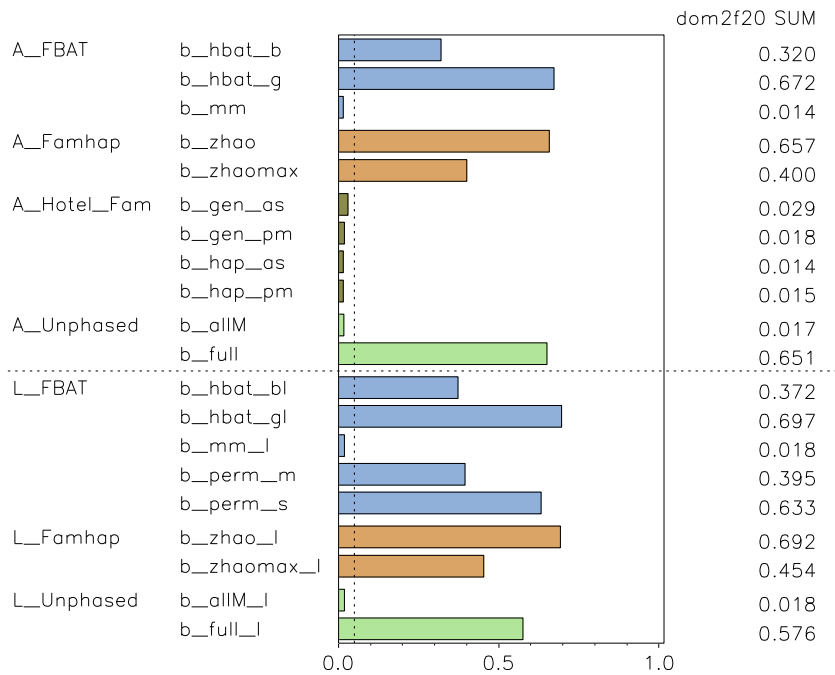
## AA-Familien

Power: alpha=0.05, ADD, KMod2, 400 AA-families



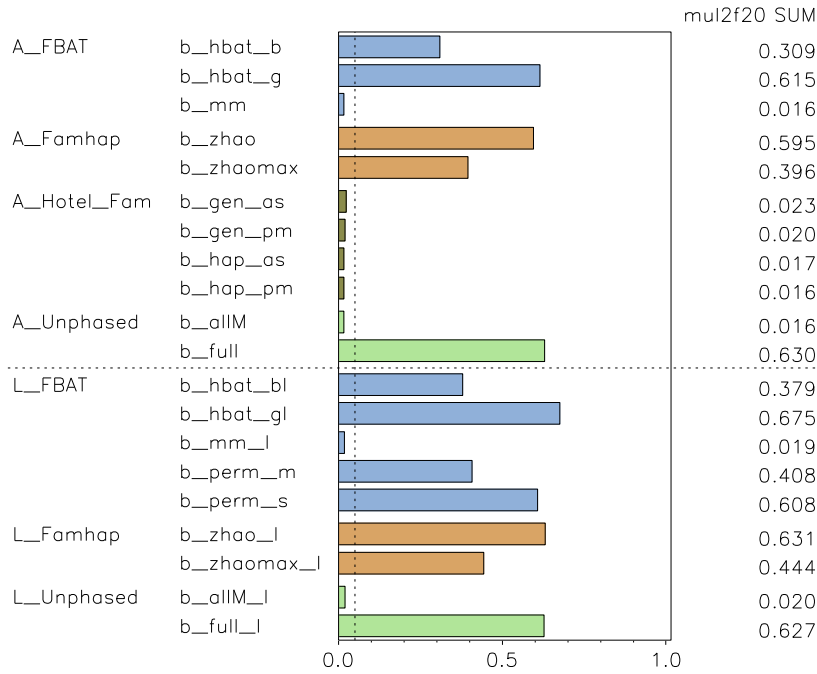
Bonferroni-corrected min\_p: Famhap, FBAT, Hotel-Fam, Unphased

Power: alpha=0.05, DOM, KMod2, 500 AA-families



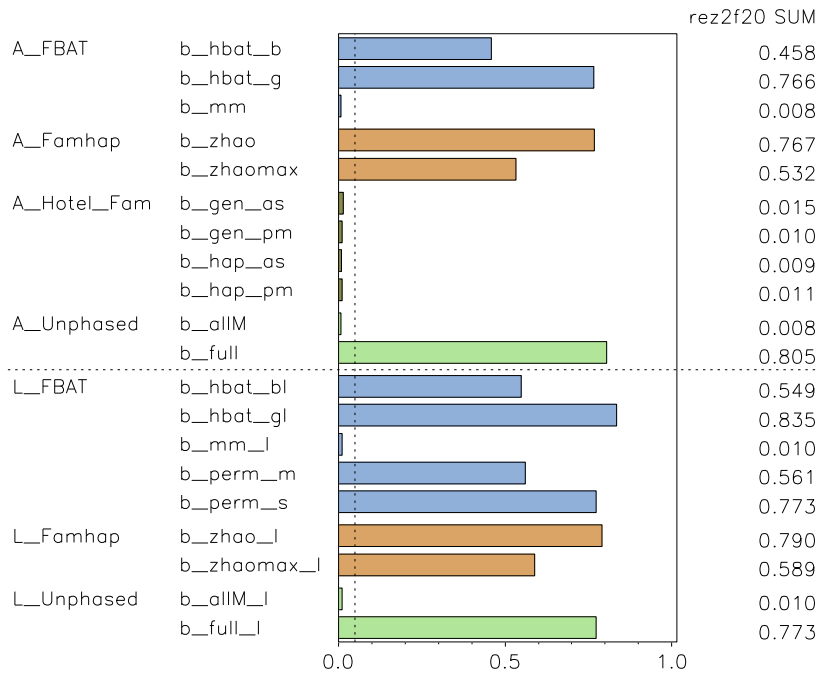
Bonferroni-corrected min\_p: Famhap, FBAT, Hotel-Fam, Unphased

Power: alpha=0.05, MUL, KMod2, 300 AA-families



Bonferroni-corrected min\_p: Famhap, FBAT, Hotel-Fam, Unphased

Power: alpha=0.05, REZ, KMod2, 900 AA-families

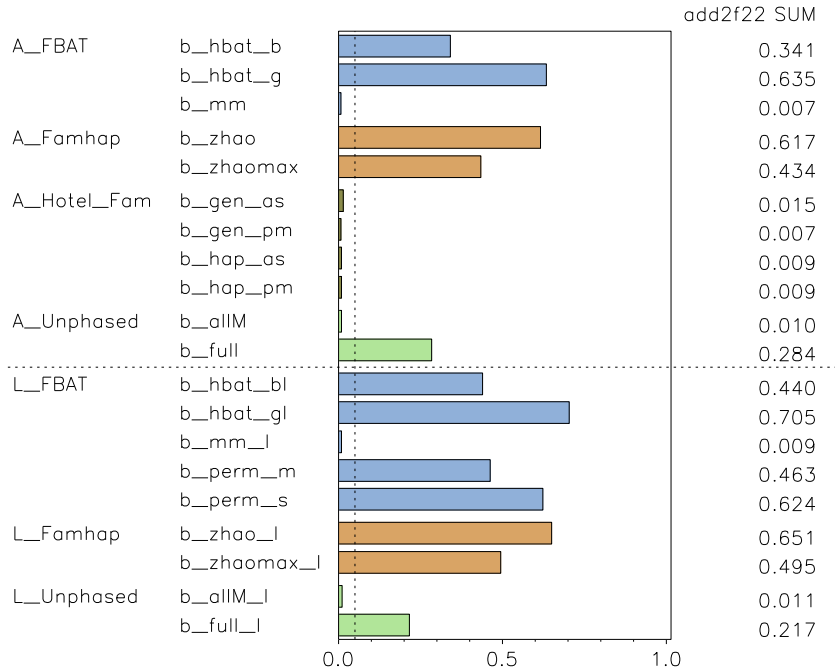


Bonferroni-corrected min\_p: Famhap, FBAT, Hotel-Fam, Unphased



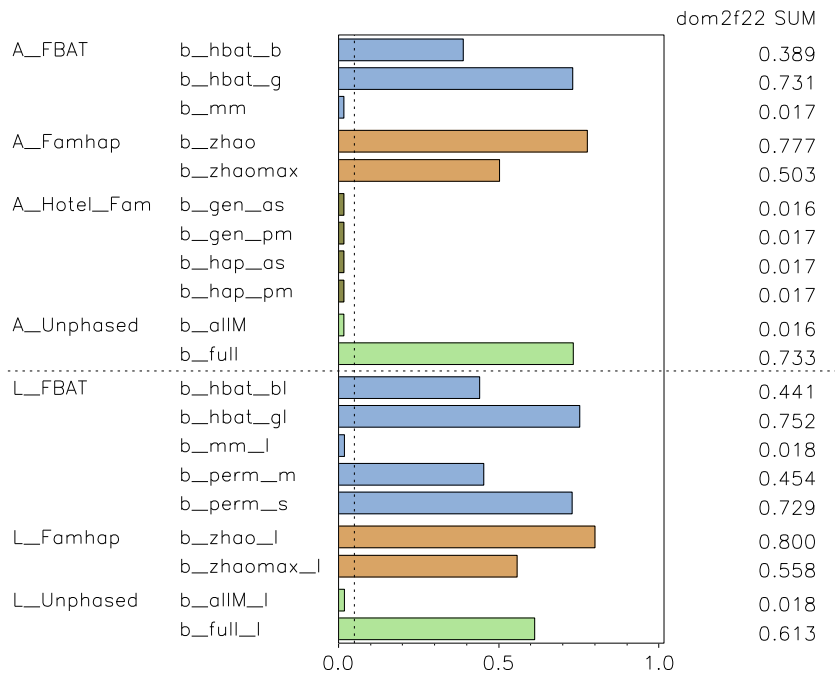
### AAUU-Familien

Power: alpha=0.05, ADD, KMod2, 400 AAUU-familien



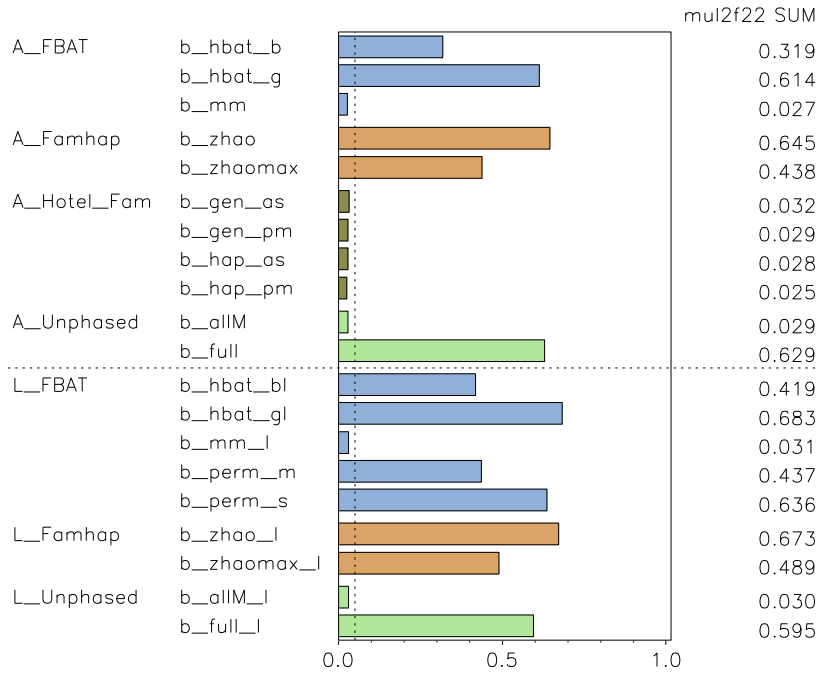
Bonferroni-corrected min\_p: Famhap, FBAT, Hotel-Fam, Unphased

Power: alpha=0.05, DOM, KMod2, 500 AAUU-familien



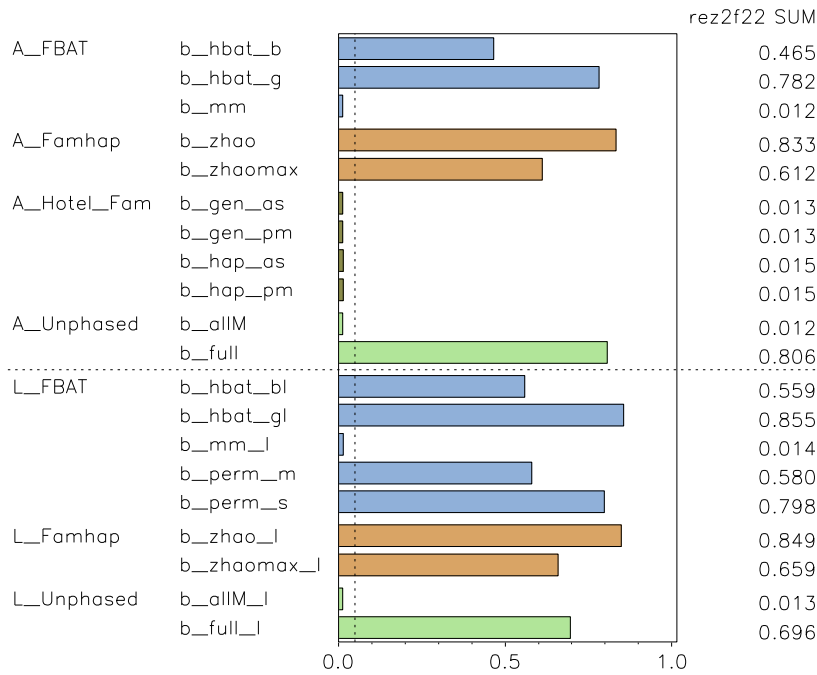
Bonferroni-corrected min\_p: Famhap, FBAT, Hotel-Fam, Unphased

Power: alpha=0.05, MUL, KMod2, 300 AAU-families



Bonferroni-corrected min\_p: Famhap, FBAT, Hotel-Fam, Unphased

Power: alpha=0.05, REZ, KMod2, 900 AAU-families

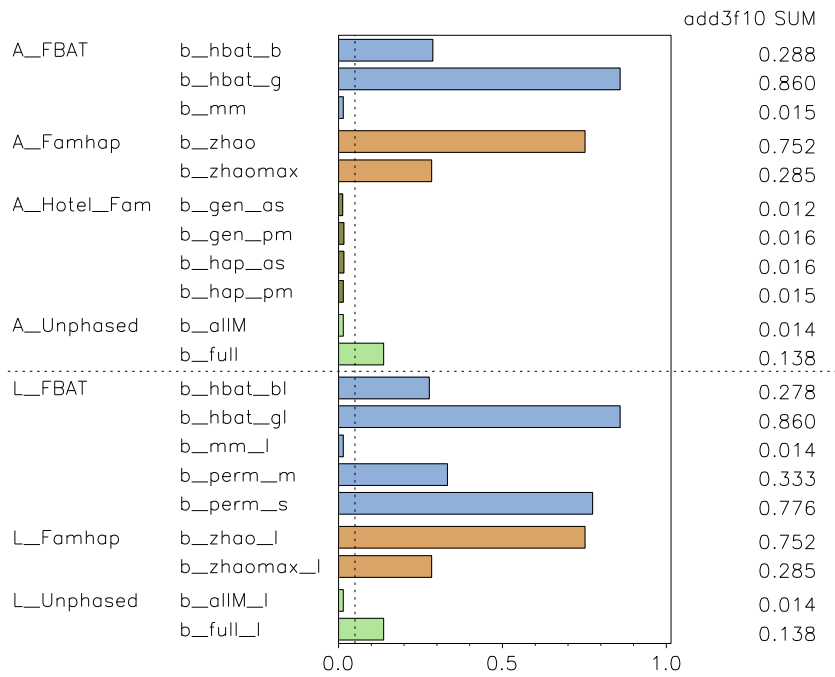


Bonferroni-corrected min\_p: Famhap, FBAT, Hotel-Fam, Unphased

### KMod 3

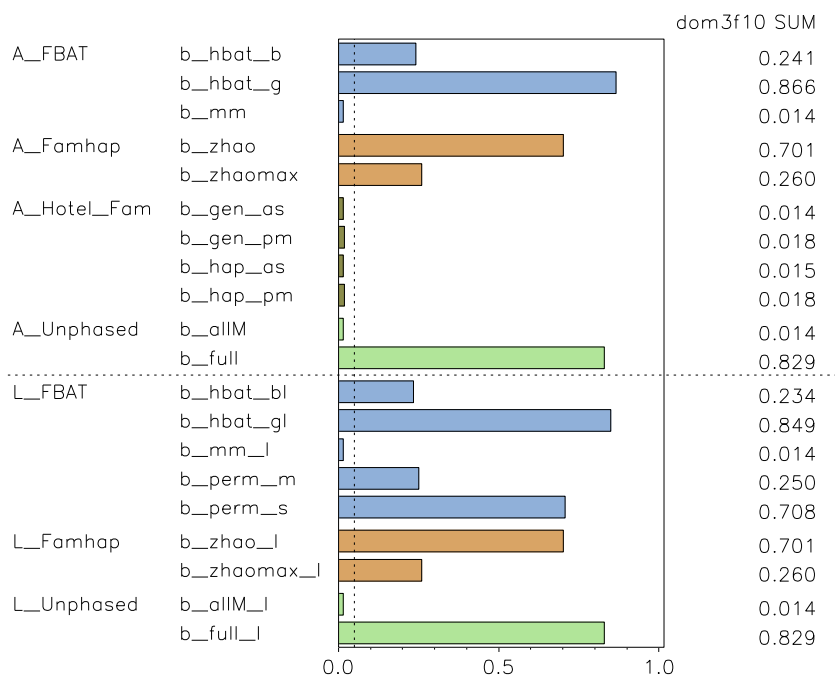
#### Trios

Power: alpha=0.05, ADD, KMod3, 2500 trios



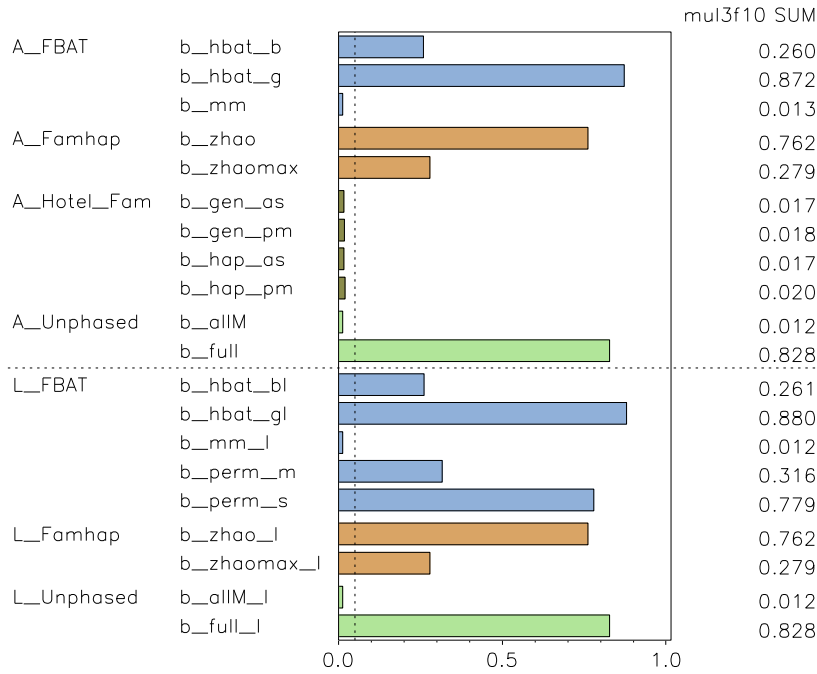
Bonferroni-corrected min\_p: Famhap, FBAT, Hotel-Fam, Unphased

Power: alpha=0.05, DOM, KMod3, 1500 trios



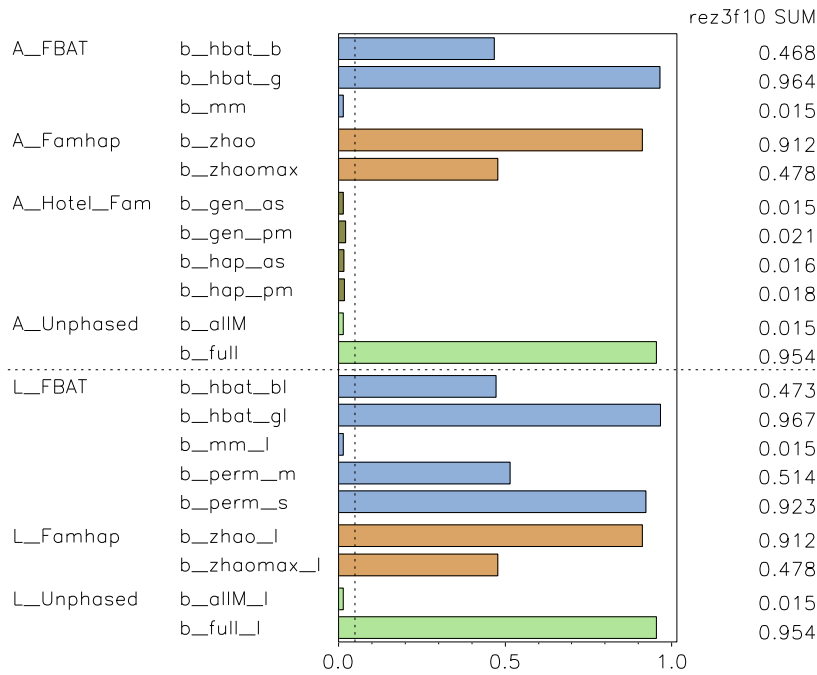
Bonferroni-corrected min\_p: Famhap, FBAT, Hotel-Fam, Unphased

Power: alpha=0.05, MUL, KMod3, 1700 trios



Bonferroni-corrected min\_p: Famhap, FBAT, Hotel-Fam, Unphased

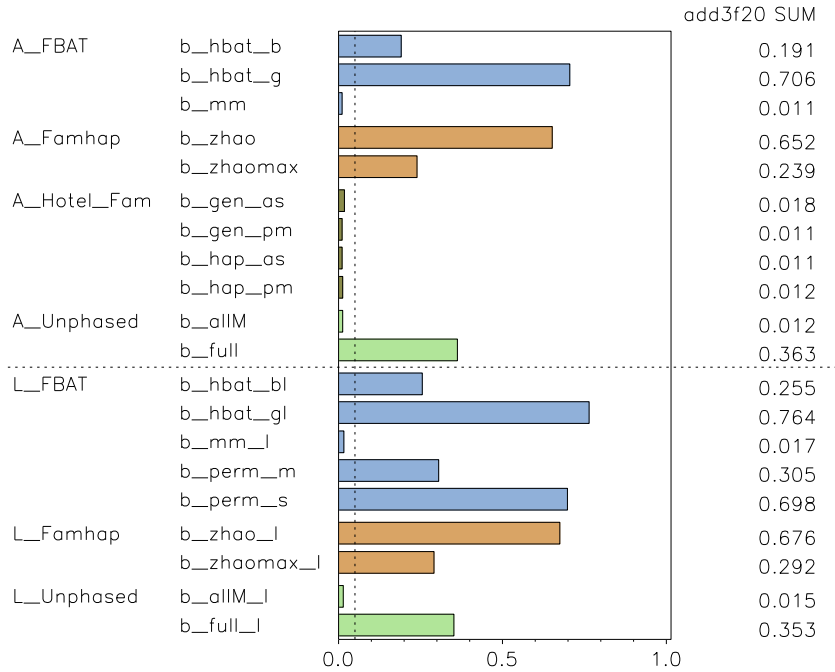
Power: alpha=0.05, REZ, KMod3, 11000 trios



Bonferroni-corrected min\_p: Famhap, FBAT, Hotel-Fam, Unphased

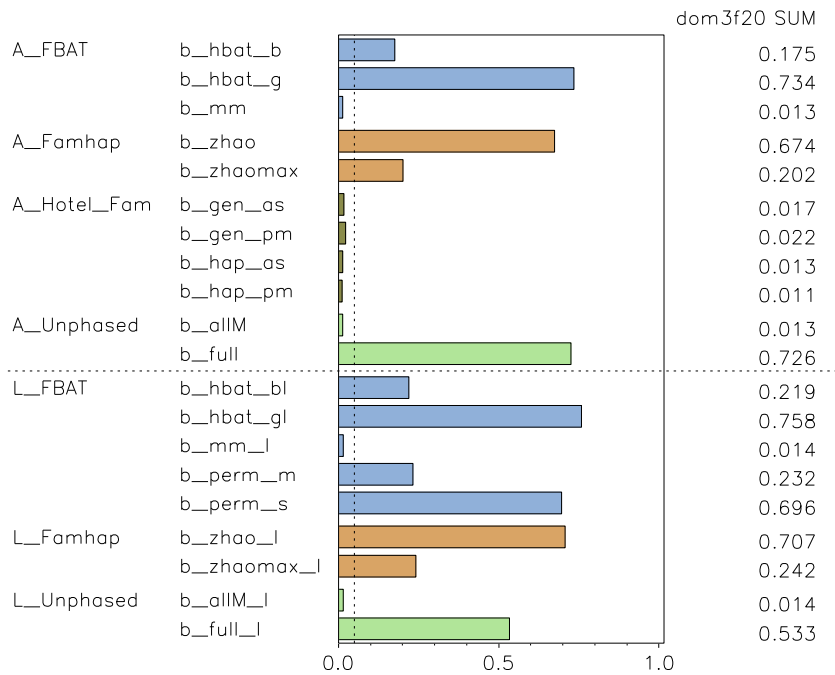
### AA-Familien

Power: alpha=0.05, ADD, KMod3, 500 AA-families



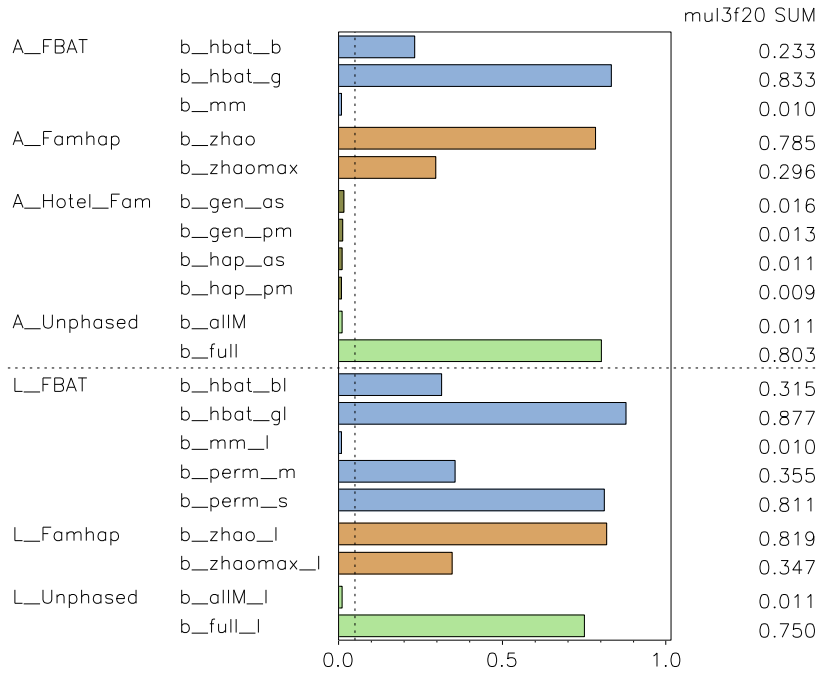
Bonferroni-corrected min\_p: Famhap, FBAT, Hotel-Fam, Unphased

Power: alpha=0.05, DOM, KMod3, 700 AA-families



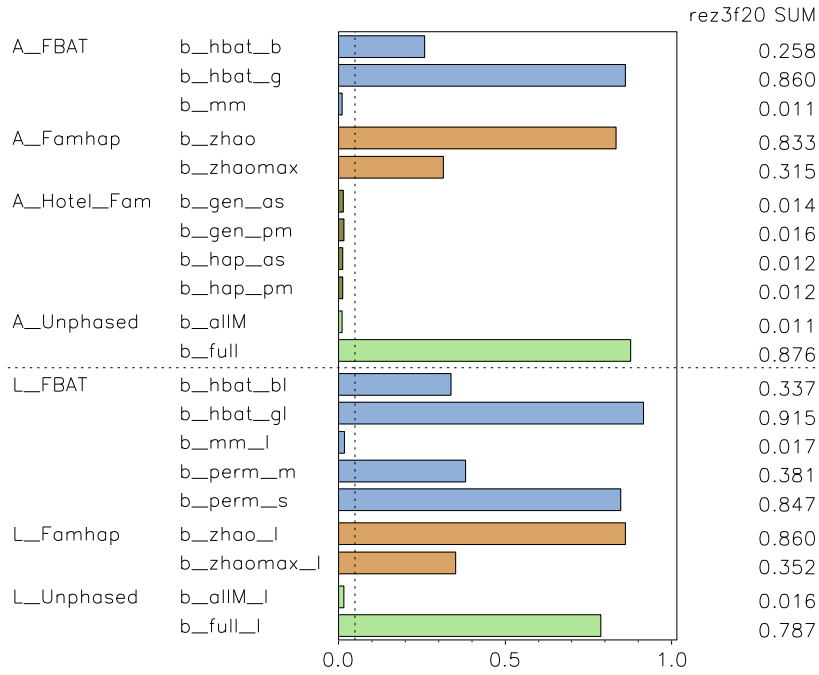
Bonferroni-corrected min\_p: Famhap, FBAT, Hotel-Fam, Unphased

Power: alpha=0.05, MUL, KMod3, 500 AA-families



Bonferroni-corrected min\_p: Famhap, FBAT, Hotel-Fam, Unphased

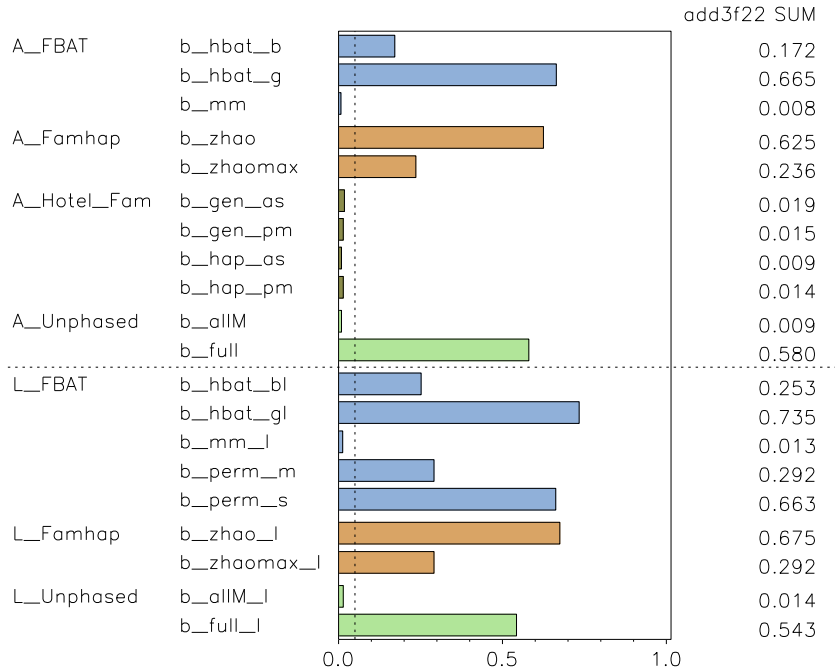
Power: alpha=0.05, REZ, KMod3, 1400 AA-families



Bonferroni-corrected min\_p: Famhap, FBAT, Hotel-Fam, Unphased

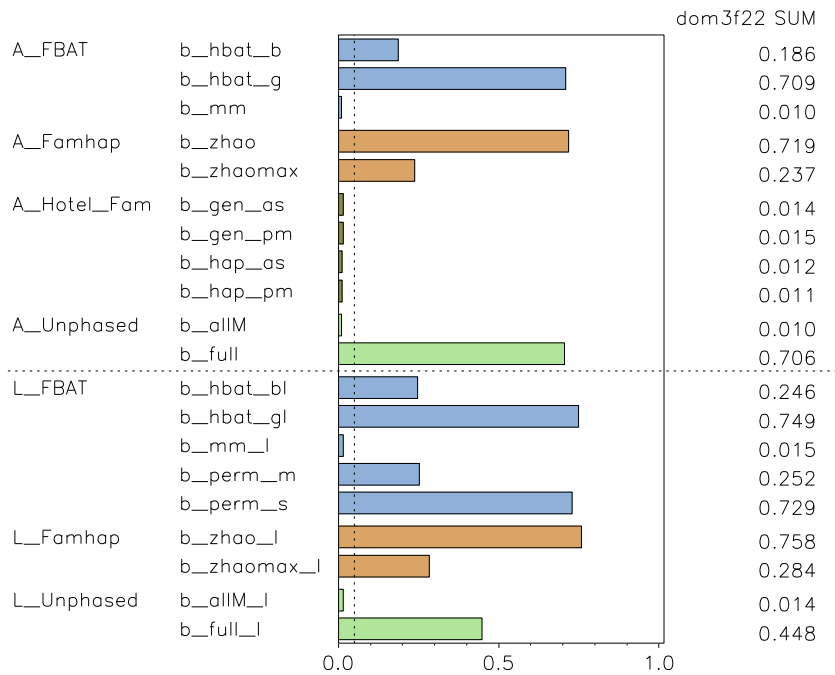
## AAUU-Familien

Power: alpha=0.05, ADD, KMod3, 500 AAUU-families



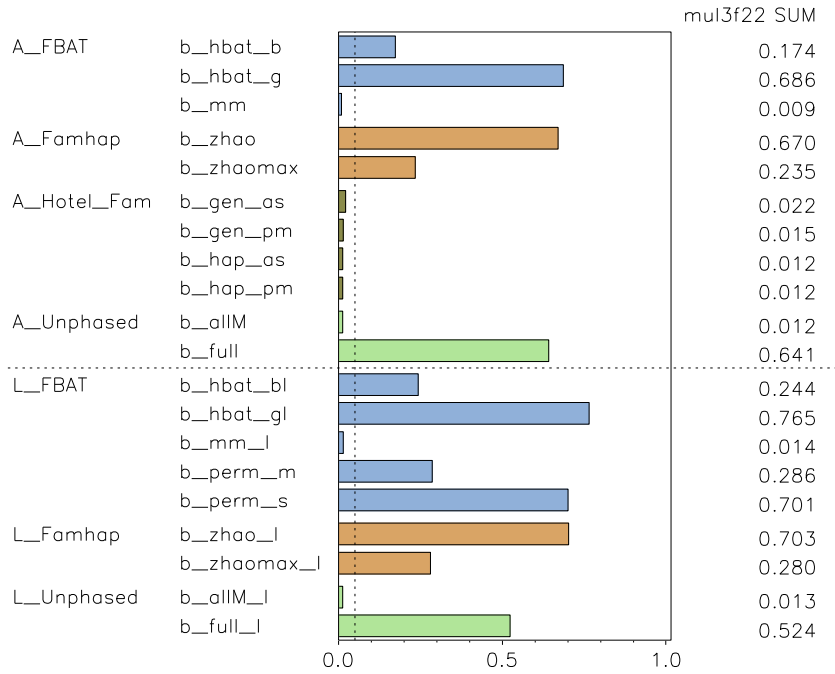
Bonferroni-corrected min\_p: Famhap, FBAT, Hotel-Fam, Unphased

Power: alpha=0.05, DOM, KMod3, 600 AAUU-families



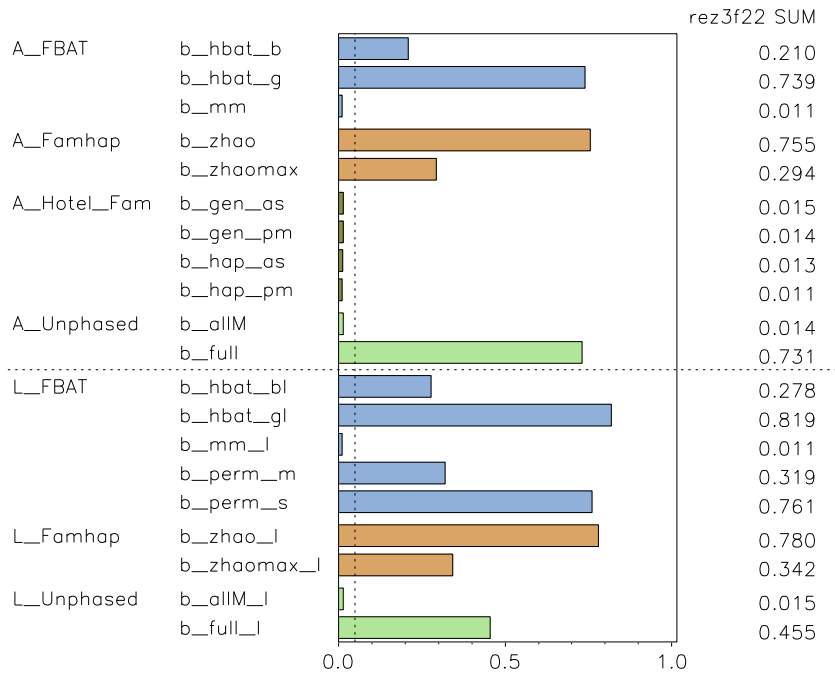
Bonferroni-corrected min\_p: Famhap, FBAT, Hotel-Fam, Unphased

Power: alpha=0.05, MUL, KMod3, 400 AAU-families



Bonferroni-corrected min\_p: Famhap, FBAT, Hotel-Fam, Unphased

Power: alpha=0.05, REZ, KMod3, 1000 AAU-families



Bonferroni-corrected min\_p: Famhap, FBAT, Hotel-Fam, Unphased

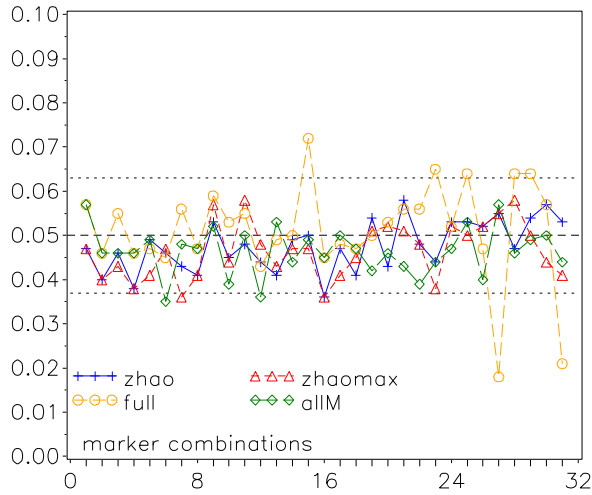


# Analyse mit fehlenden Genotypen: Fehler 1. Art, $\alpha = 0.05$

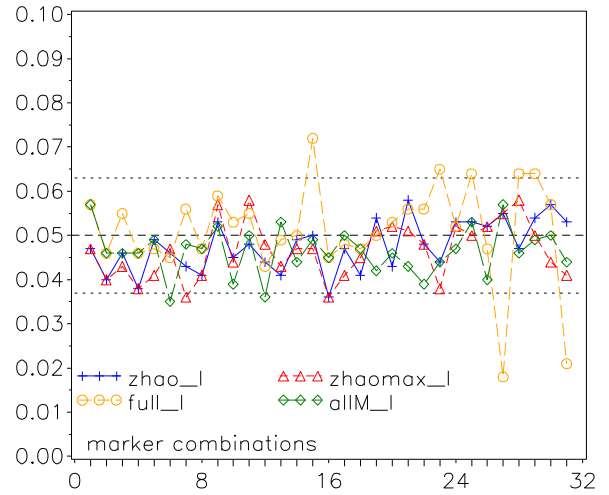
## 200 Familien

### NMod 1

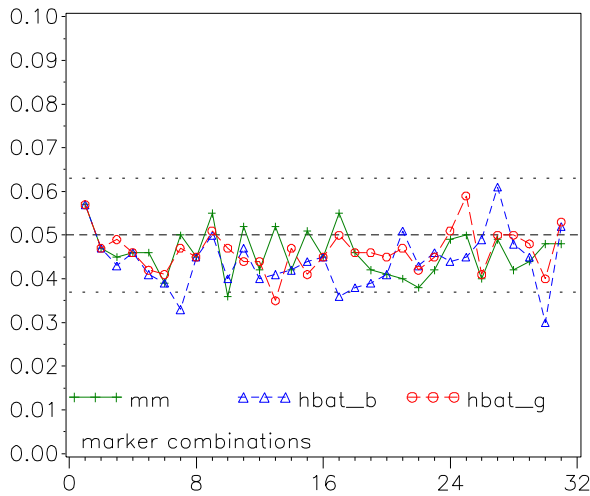
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 200 trios, NMod 1, H0\_(A)



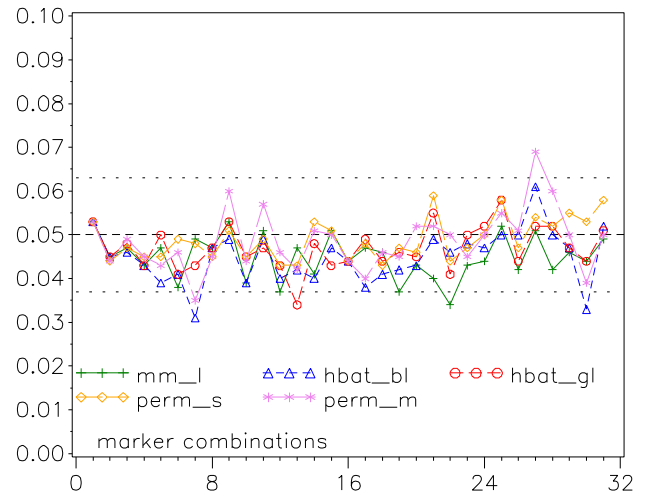
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 200 trios, NMod 1, H0\_(L)



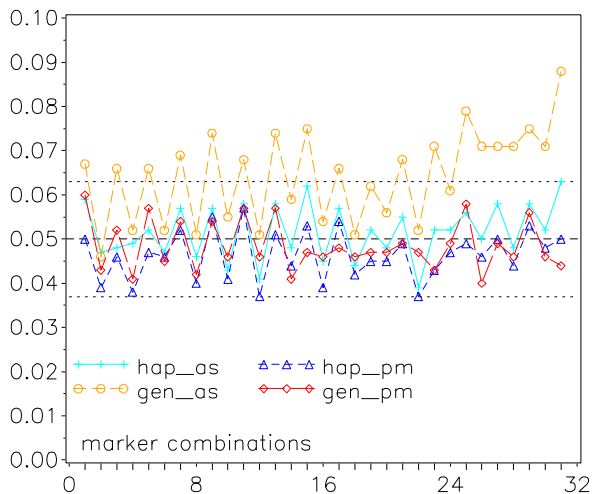
Type I error: FBAT,  
Missing genotypes, 200 trios, Nmod 1, H0\_(A)



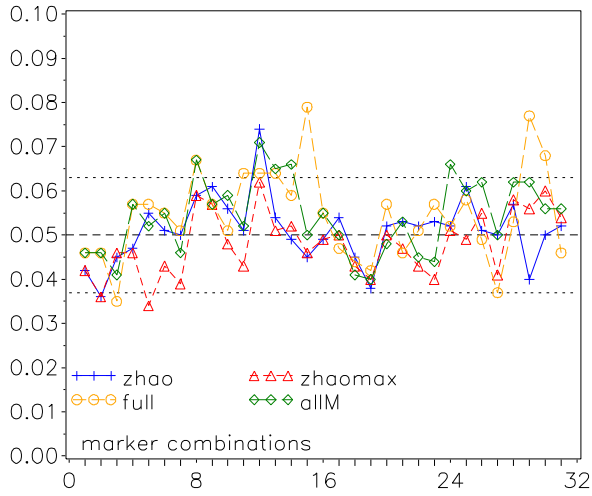
Type I error: FBAT,  
Missing genotypes, 200 trios, Nmod 1, H0\_(L)



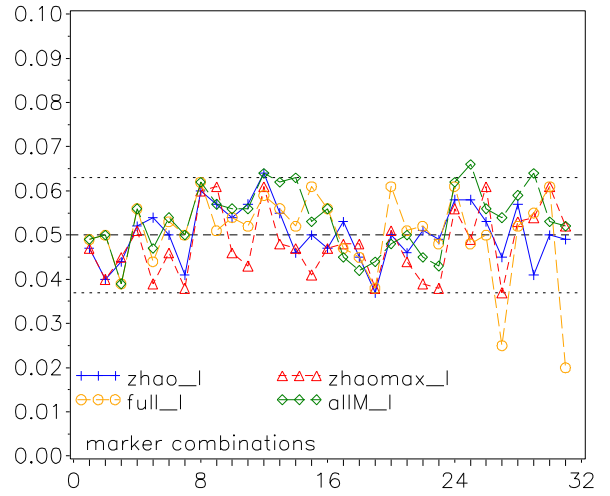
Type I error of Hotel\_Fam  
Missing genotypes, 200 trios, NMod 1



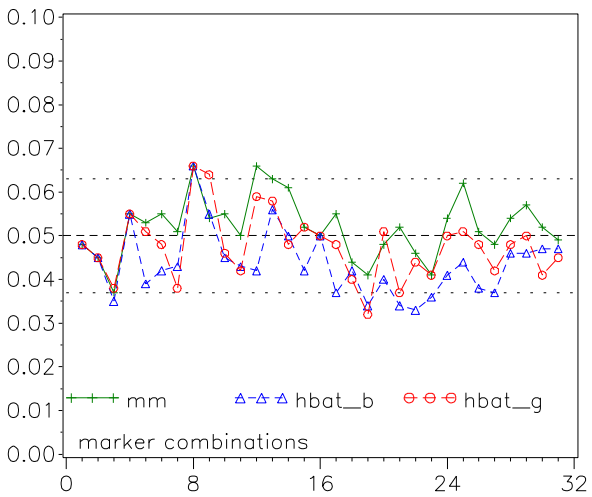
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 200 AA fam., NMod 1, H0\_(A)



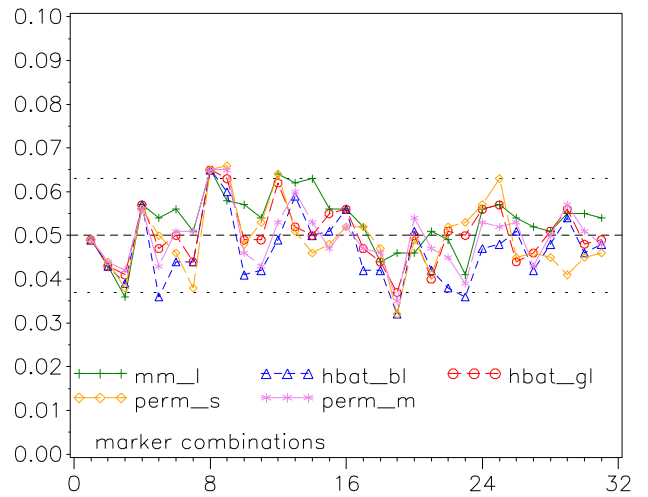
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 200 AA fam., NMod 1, H0\_(L)



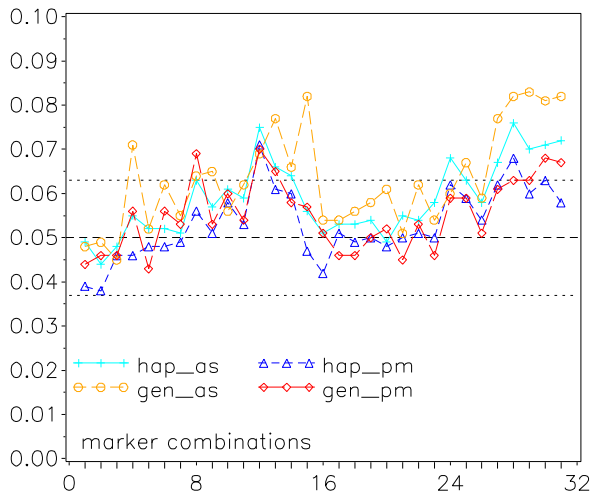
Type I error: FBAT,  
Missing genotypes, 200 AA fam., Nmod 1, H0\_(A)



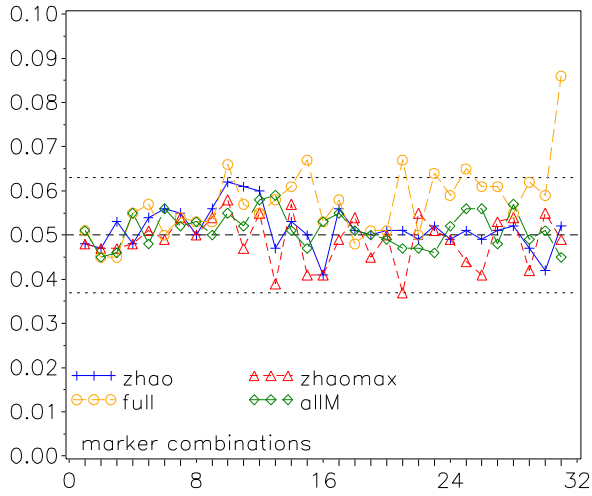
Type I error: FBAT,  
Missing genotypes, 200 AA fam., Nmod 1, H0\_(L)



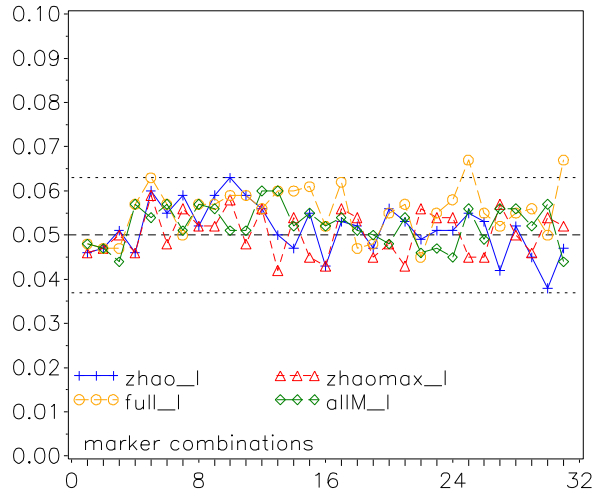
Type I error of Hotel\_Fam  
Missing genotypes, 200 AA families, NMod 1



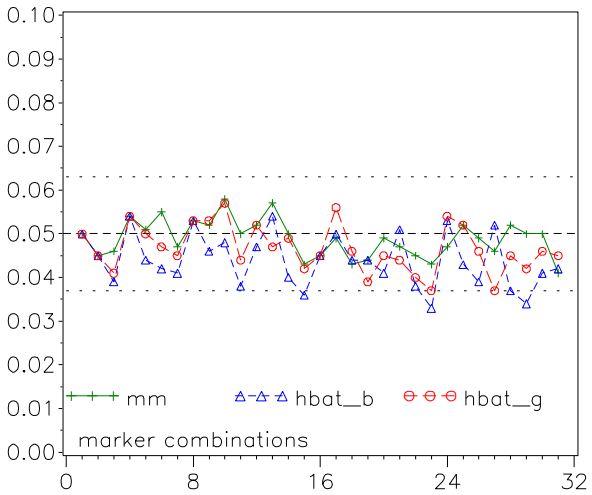
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 200 AAU fam., NMod 1, H0\_(A)



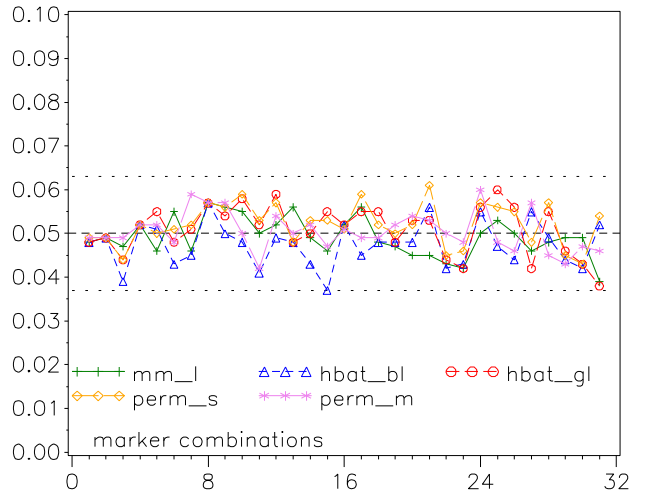
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 200 AAU fam., NMod 1, H0\_(L)



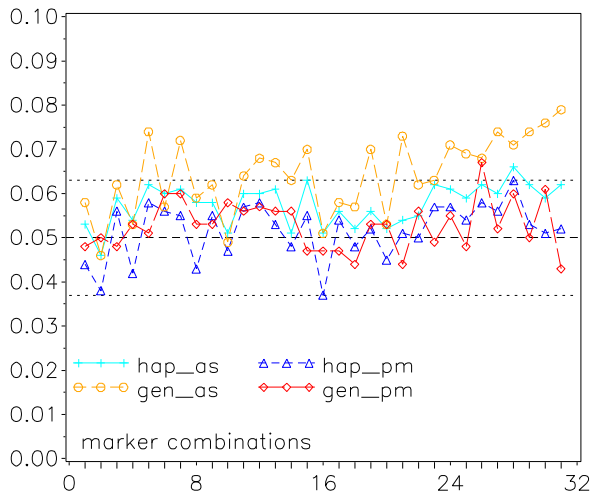
Type I error: FBAT,  
Missing genotypes, 200 AAU fam., Nmod 1, H0\_(A)



Type I error: FBAT,  
Missing genotypes, 200 AAU fam., Nmod 1, H0\_(L)

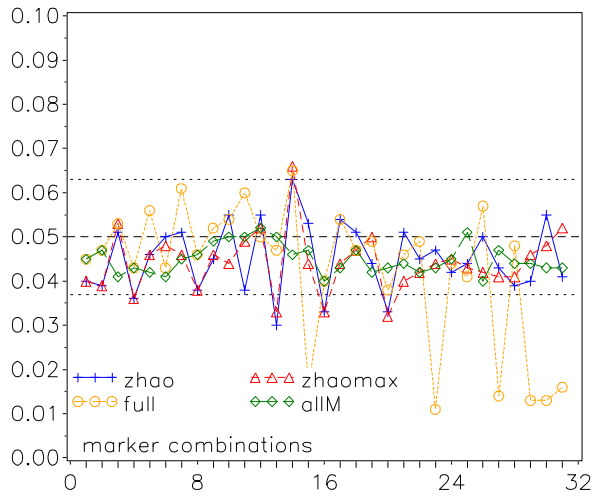


Type I error of Hotel\_Fam  
Missing genotypes, 200 AAU families, NMod 1

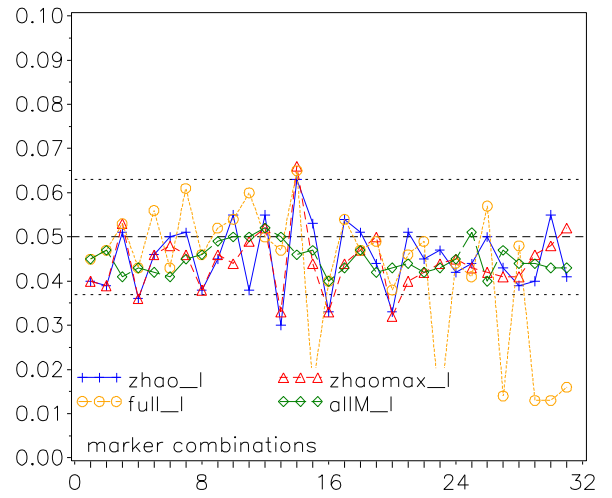


## NMod 2

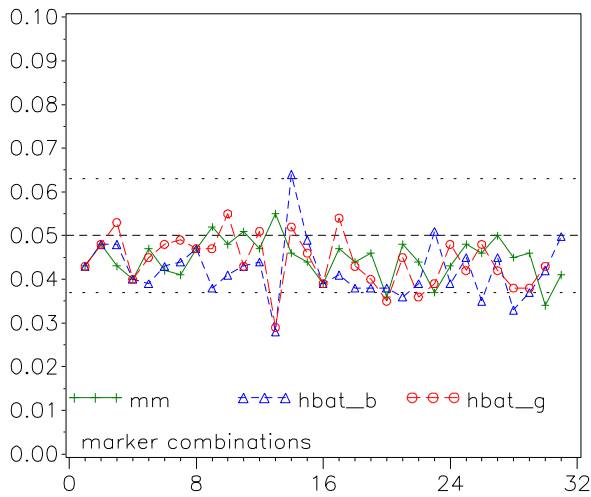
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 200 trios, NMod 2, H0\_(A)



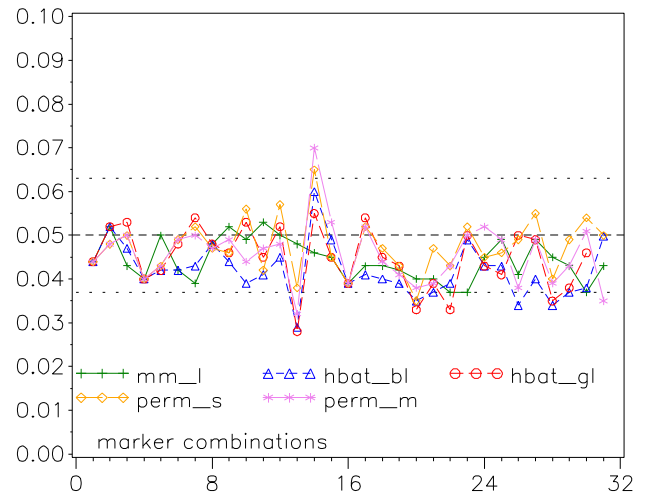
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 200 trios, NMod 2, H0\_(L)



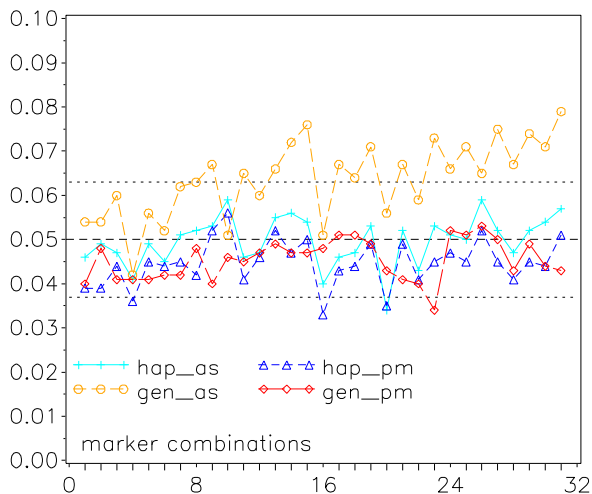
Type I error: FBAT,  
Missing genotypes, 200 trios, Nmod 2, H0\_(A)



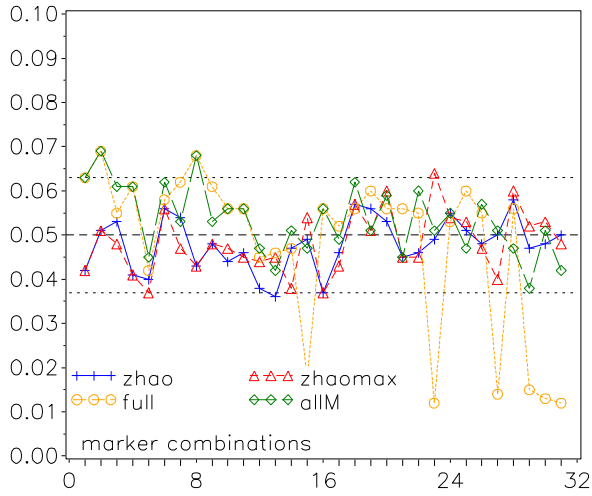
Type I error: FBAT,  
Missing genotypes, 200 trios, Nmod 2, H0\_(L)



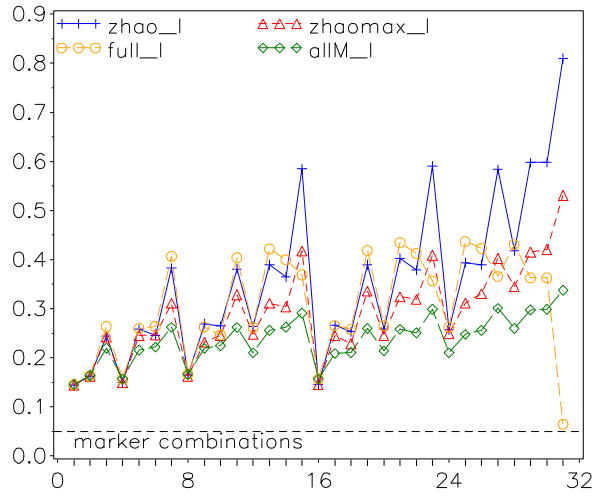
Type I error of Hotel\_Fam  
Missing genotypes, 200 trios, NMod 2



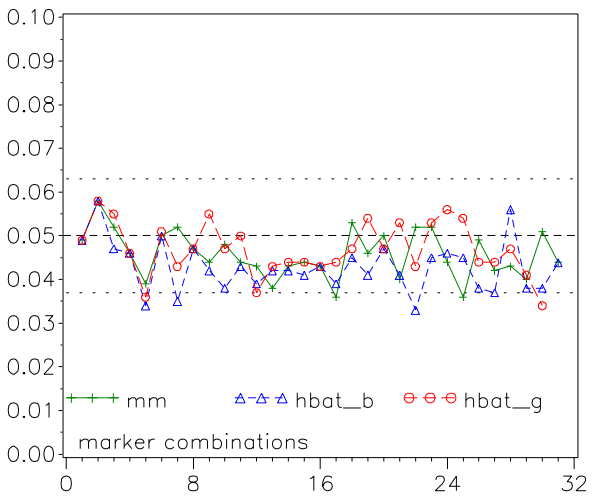
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 200 AA fam., NMod 2, H0\_(A)



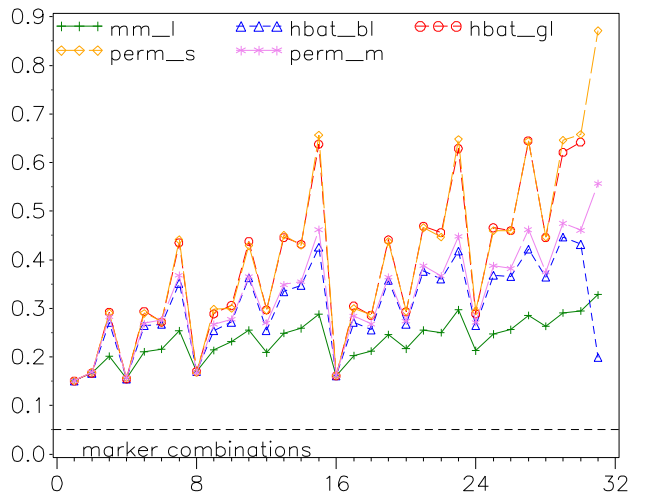
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 200 AA fam., NMod 2, H0\_(L)



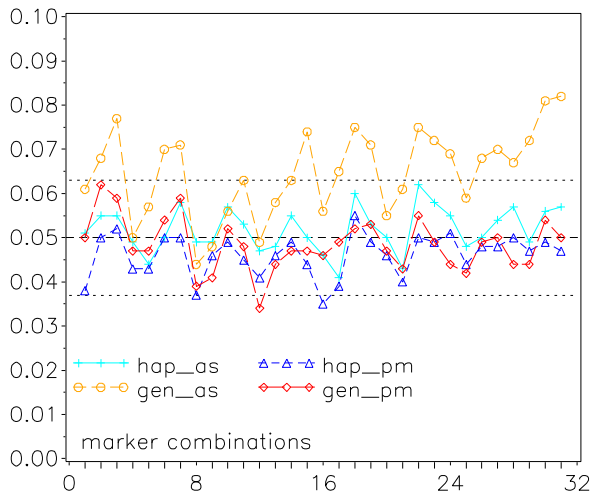
Type I error: FBAT,  
Missing genotypes, 200 AA fam., Nmod 2, H0\_(A)



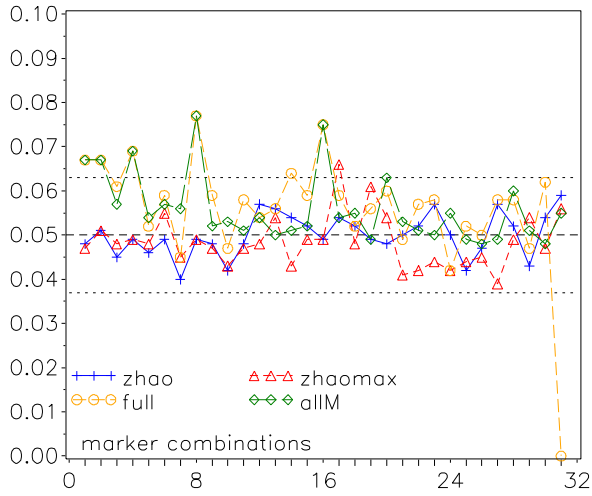
Type I error: FBAT,  
Missing genotypes, 200 AA fam., Nmod 2, H0\_(L)



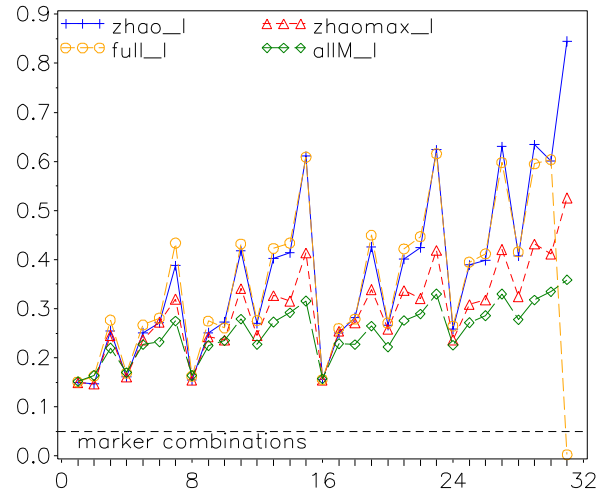
Type I error of Hotel\_Fam  
Missing genotypes, 200 AA families, NMod 2



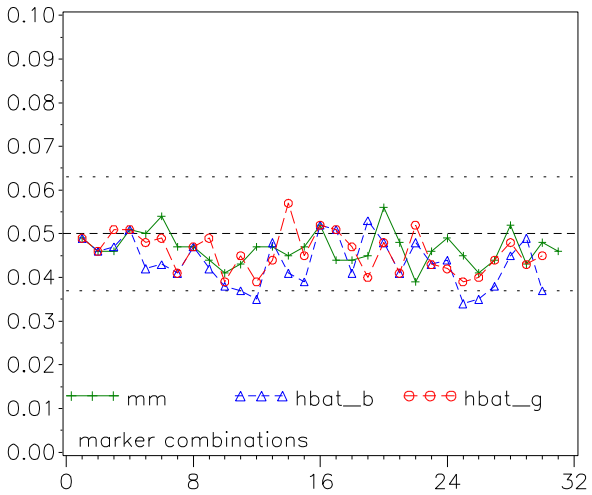
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 200 AAU fam., NMod 2, H0\_(A)



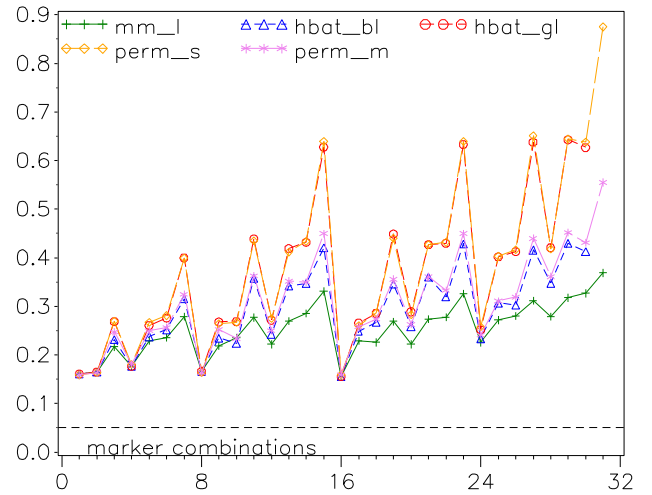
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 200 AAU fam., NMod 2, H0\_(L)



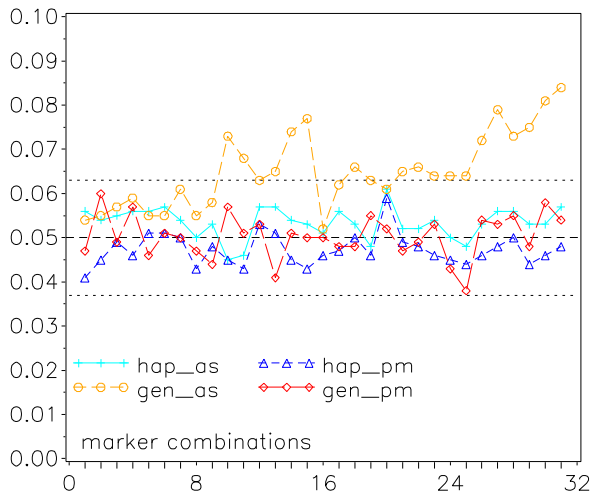
Type I error: FBAT,  
Missing genotypes, 200 AAU fam., Nmod 2, H0\_(A)



Type I error: FBAT,  
Missing genotypes, 200 AAU fam., Nmod 2, H0\_(L)

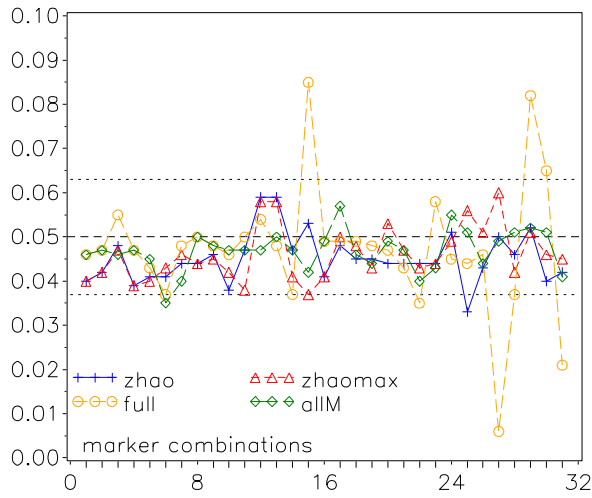


Type I error of Hotel\_Fam  
Missing genotypes, 200 AAU families, NMod 2

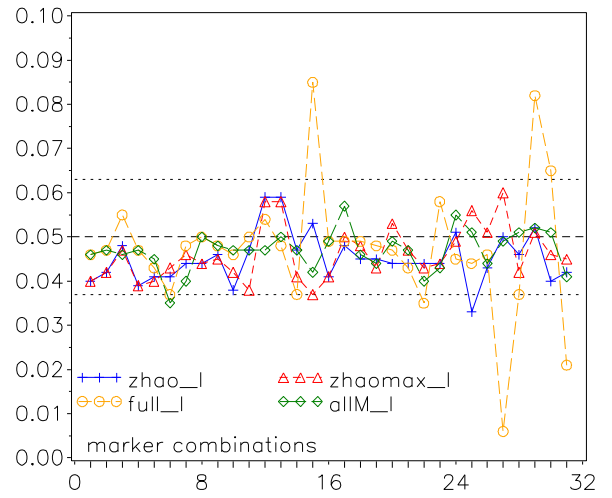


### NMod 3

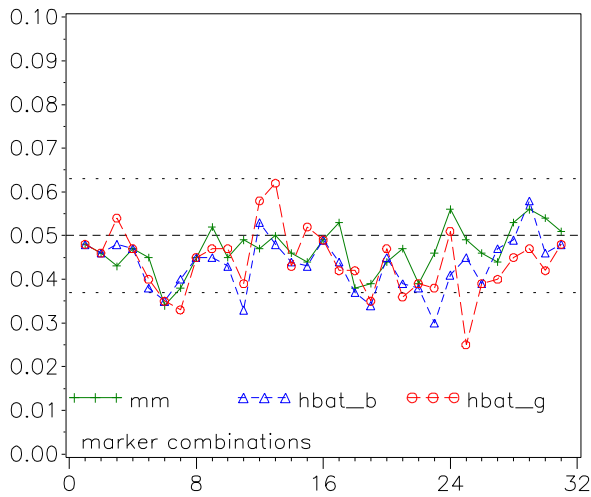
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 200 trios, NMod 3, H0\_(A)



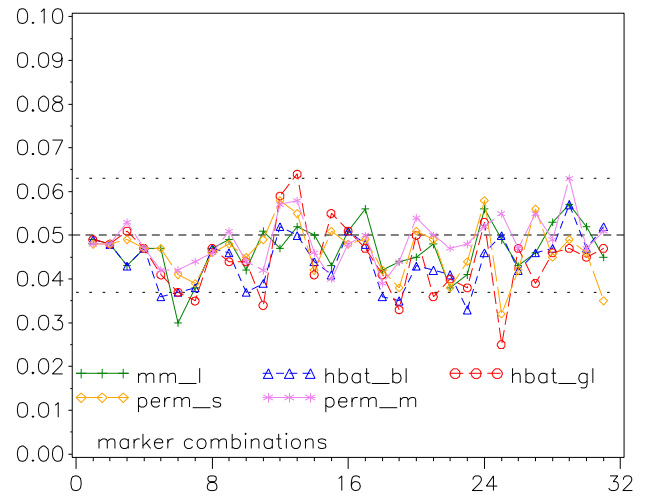
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 200 trios, NMod 3, H0\_(L)



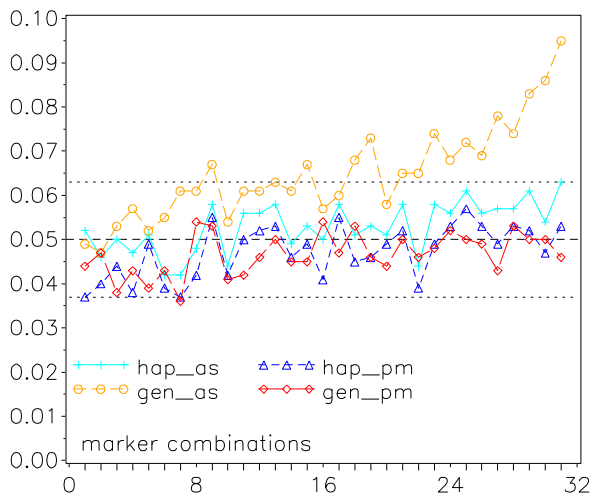
Type I error: FBAT,  
Missing genotypes, 200 trios, Nmod 3, H0\_(A)



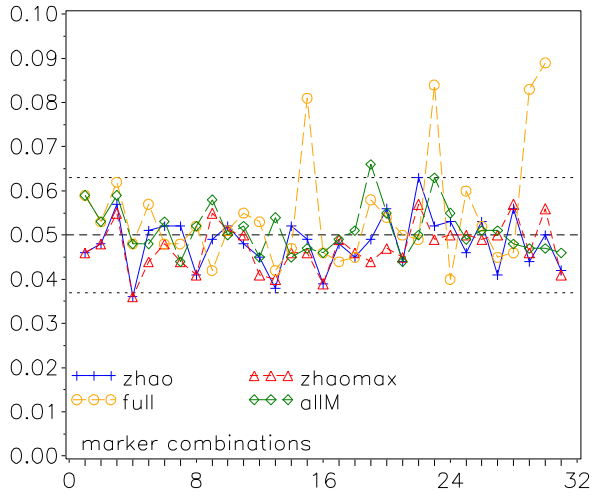
Type I error: FBAT,  
Missing genotypes, 200 trios, Nmod 3, H0\_(L)



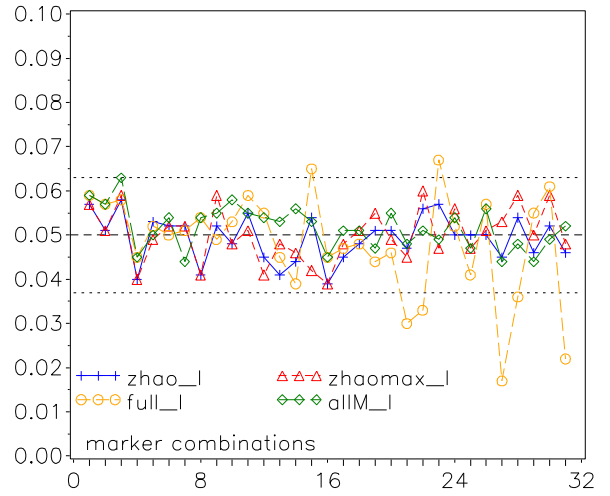
Type I error of Hotel\_Fam  
Missing genotypes, 200 trios, NMod 3



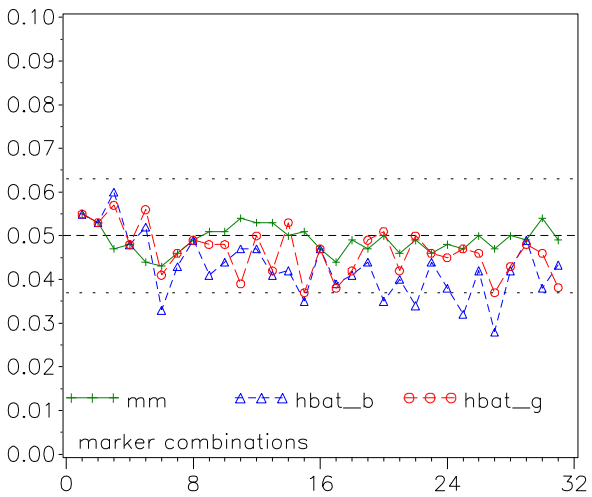
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 200 AA fam., NMod 3, H0\_(A)



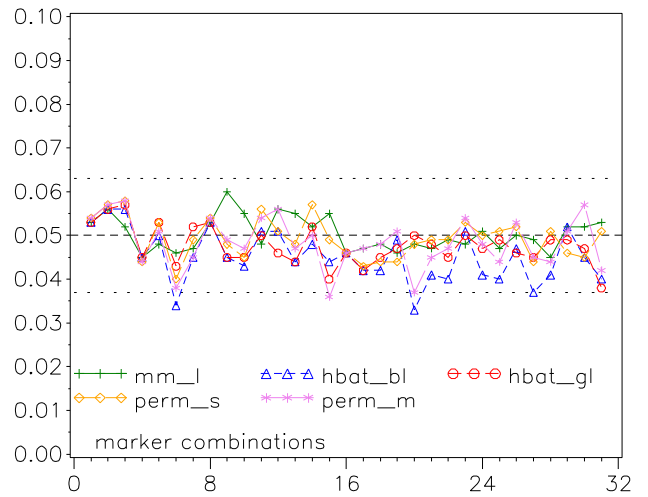
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 200 AA fam., NMod 3, H0\_(L)



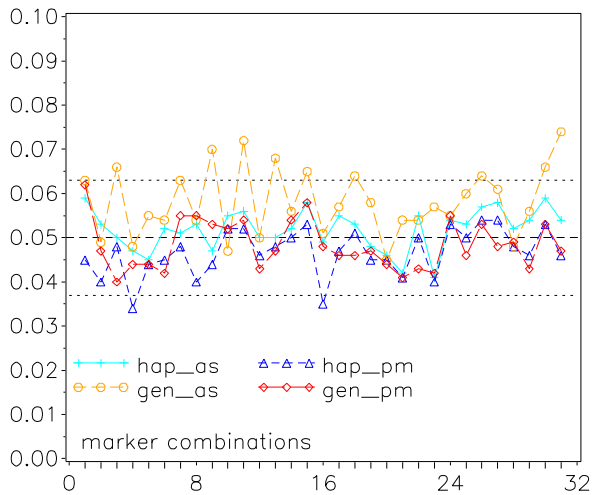
Type I error: FBAT,  
Missing genotypes, 200 AA fam., Nmod 3, H0\_(A)



Type I error: FBAT,  
Missing genotypes, 200 AA fam., Nmod 3, H0\_(L)

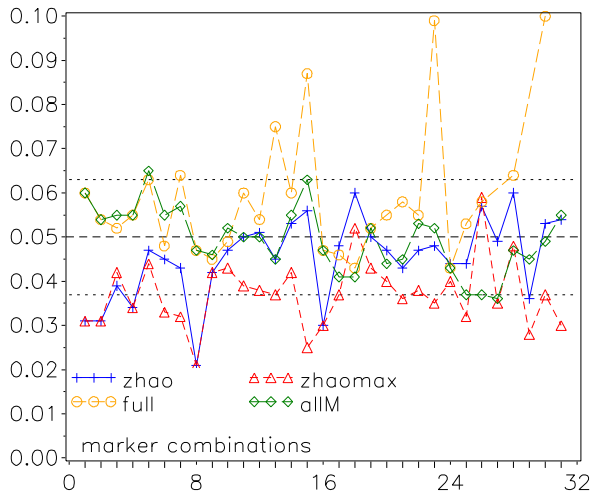


Type I error of Hotel\_Fam  
Missing genotypes, 200 AA families, NMod 3

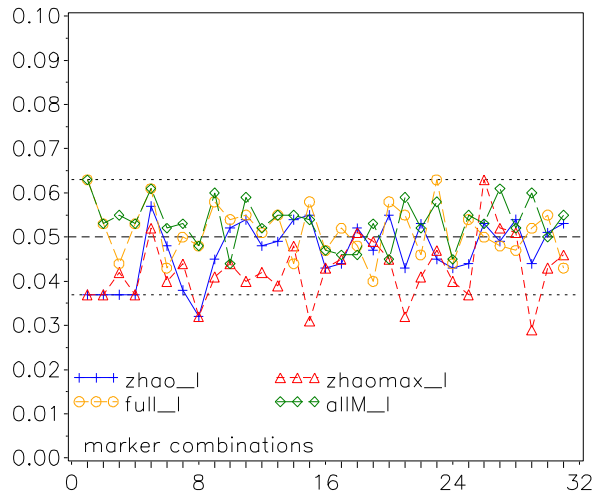




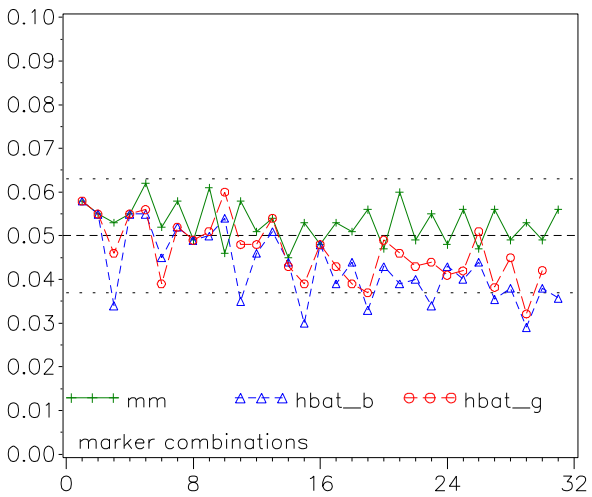
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 200 AAU fam., NMod 3, H0\_(A)



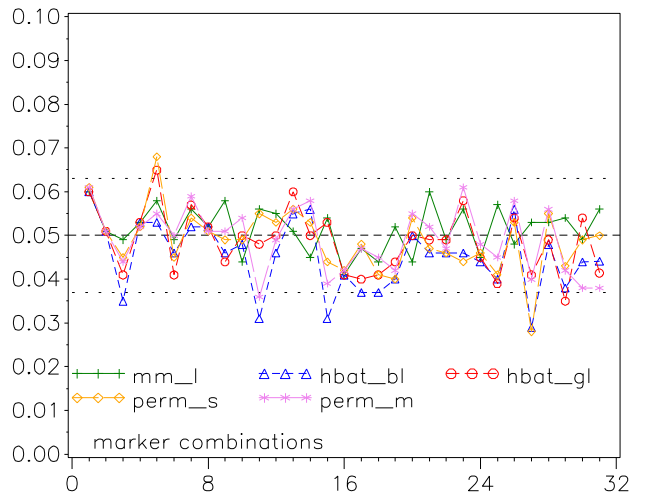
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 200 AAU fam., NMod 3, H0\_(L)



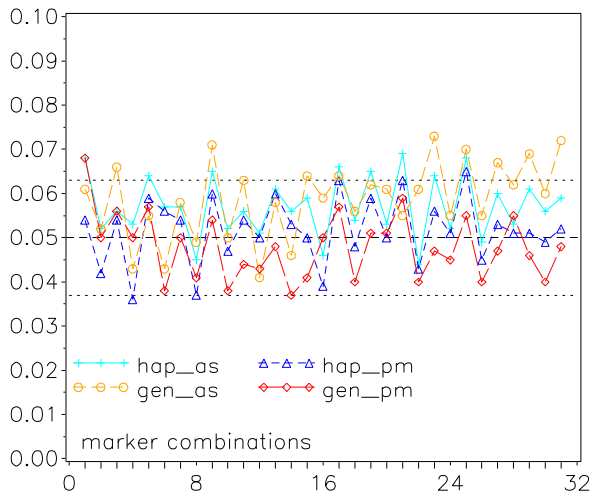
Type I error: FBAT,  
Missing genotypes, 200 AAU fam., Nmod 3, H0\_(A)



Type I error: FBAT,  
Missing genotypes, 200 AAU fam., Nmod 3, H0\_(L)



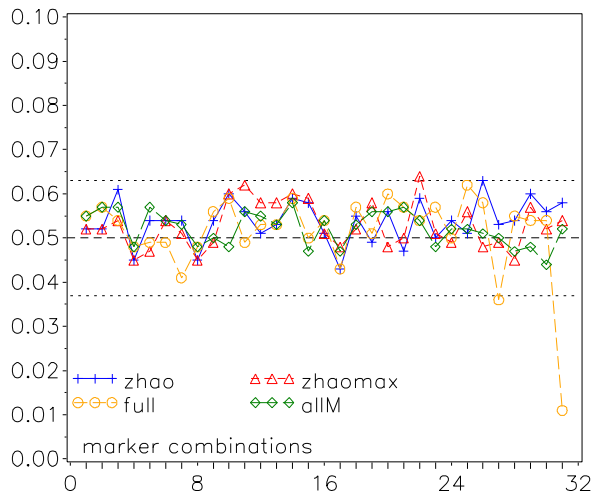
Type I error of Hotel\_Fam  
Missing genotypes, 200 AAU families, NMod 3



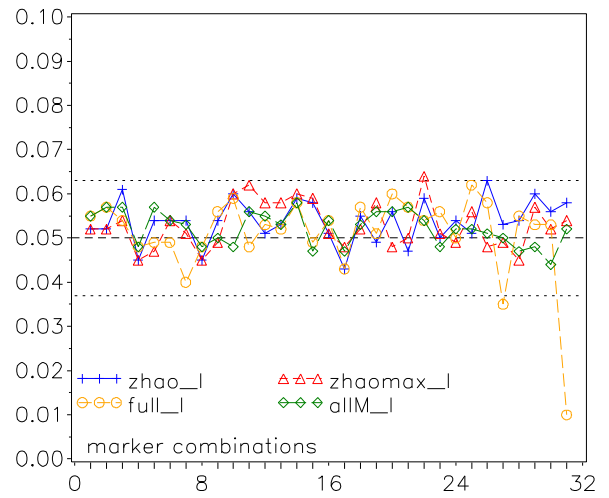
# 1000 Familien

## NMod 1

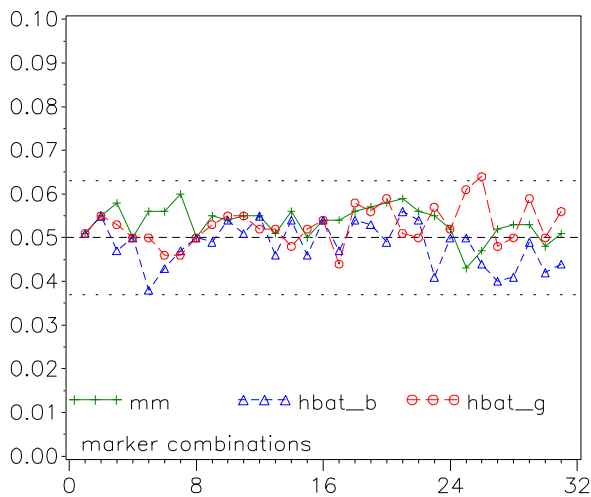
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 1000 trios, NMod 1, H0\_(A)



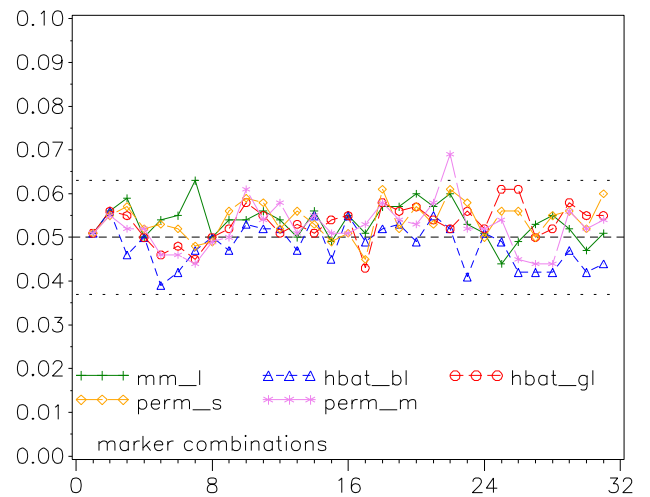
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 1000 trios, NMod 1, H0\_(L)



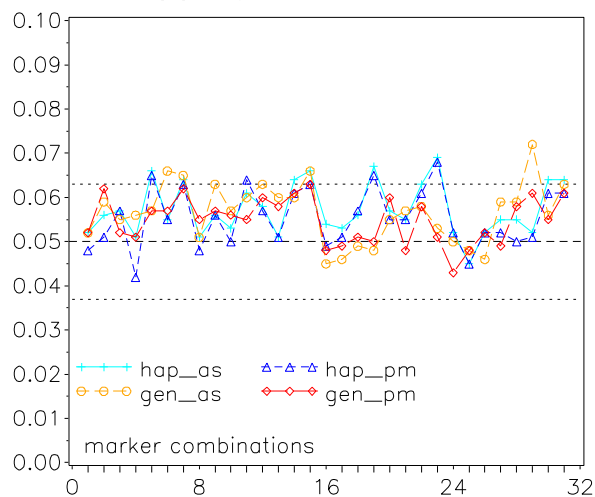
Type I error: FBAT,  
Missing genotypes, 1000 trios, Nmod 1, H0\_(A)



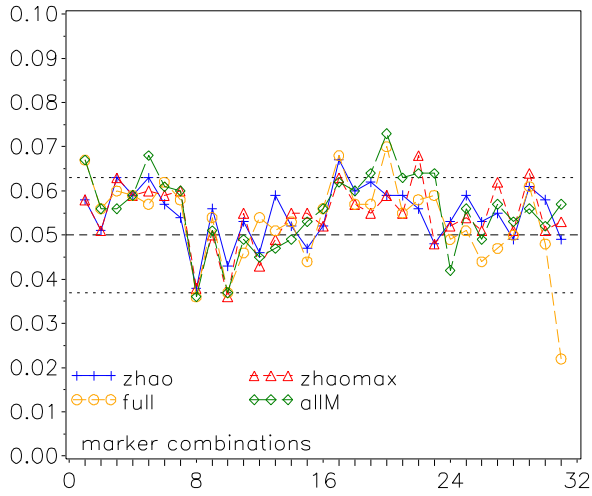
Type I error: FBAT,  
Missing genotypes, 1000 trios, Nmod 1, H0\_(L)



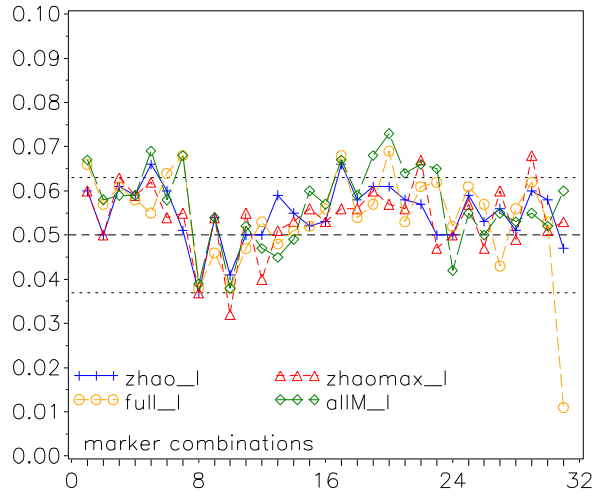
Type I error of Hotel\_Fam  
Missing genotypes, 1000 trios, NMod 1



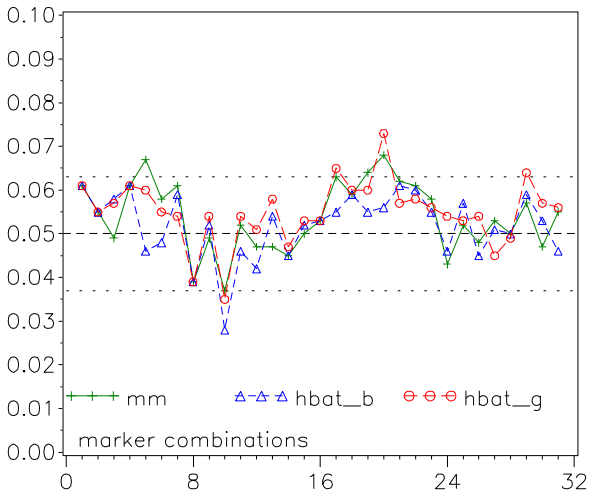
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 1000 AA fam., NMod 1, H0\_(A)



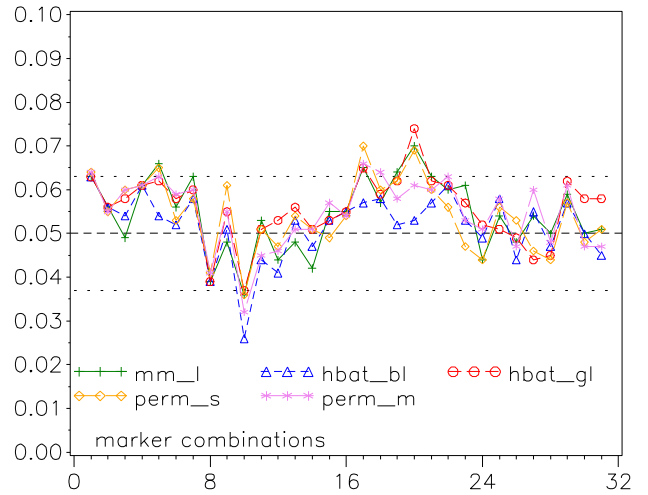
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 1000 AA fam., NMod 1, H0\_(L)



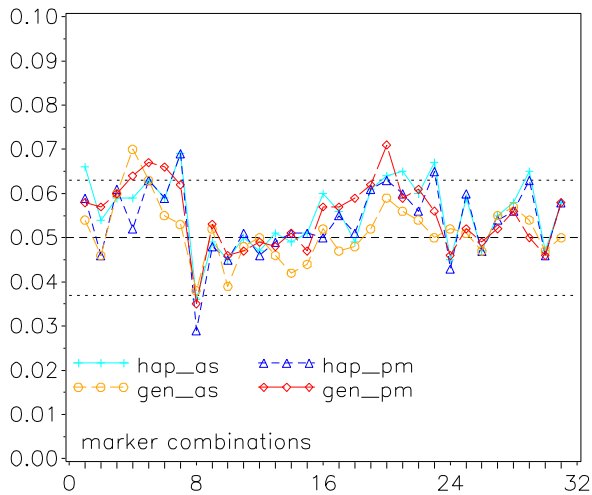
Type I error: FBAT,  
Missing genotypes, 1000 AA fam., Nmod 1, H0\_(A)



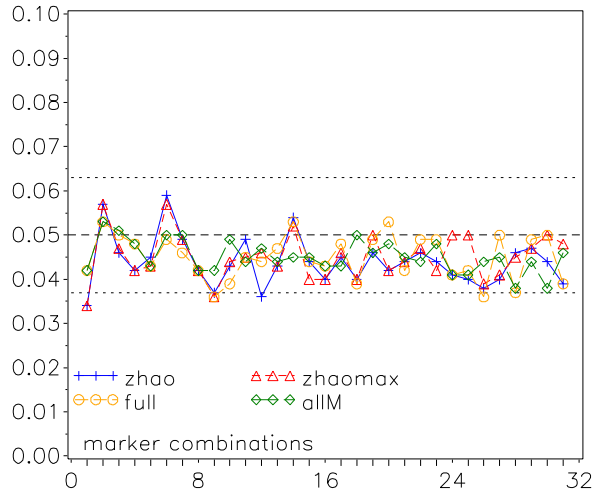
Type I error: FBAT,  
Missing genotypes, 1000 AA fam., Nmod 1, H0\_(L)



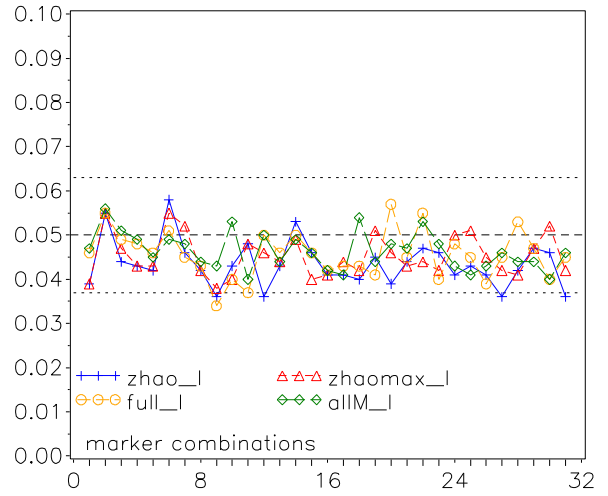
Type I error of Hotel\_Fam  
Missing genotypes, 1000 AA families, NMod 1



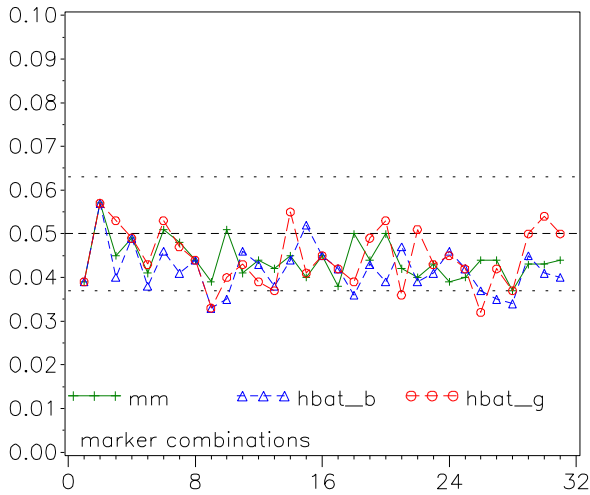
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 1000 AAUU fam., NMod 1, H0\_(A)



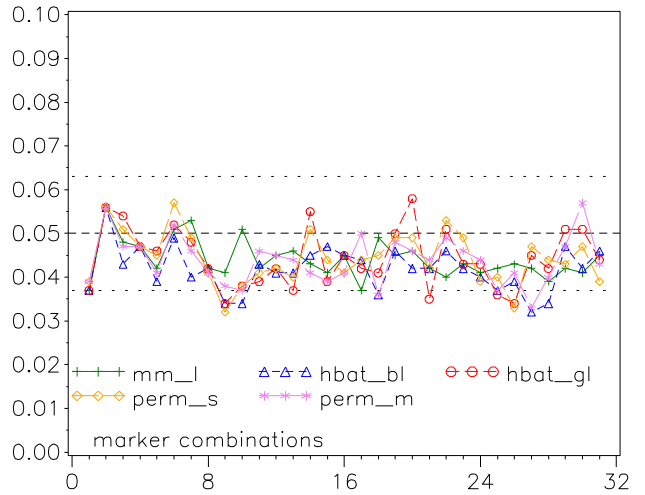
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 1000 AAUU fam., NMod 1, H0\_(L)



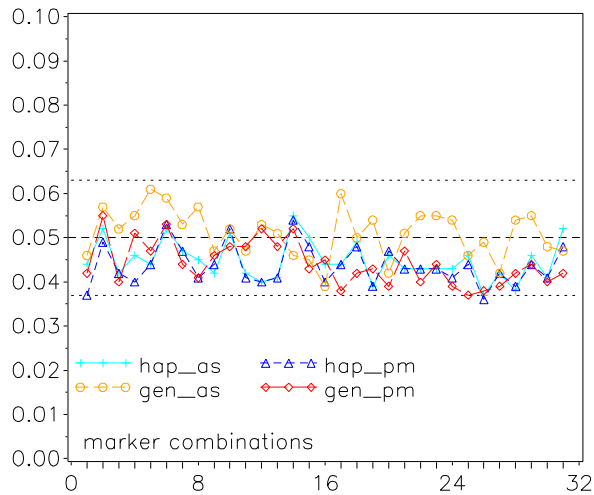
Type I error: FBAT,  
Missing genotypes, 1000 AAUU fam., Nmod 1, H0\_(A)



Type I error: FBAT,  
Missing genotypes, 1000 AAUU fam., Nmod 1, H0\_(L)

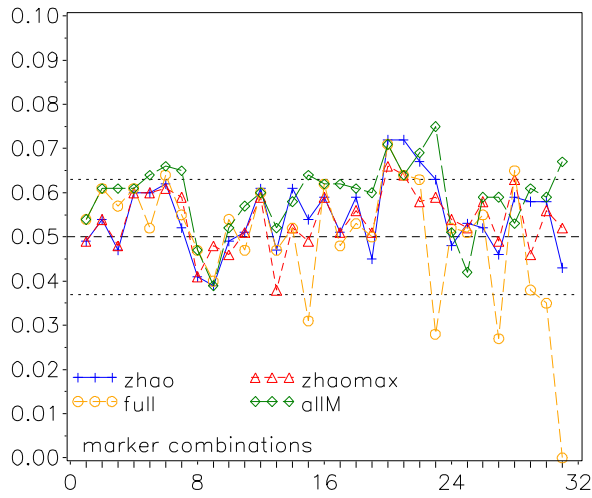


Type I error of Hotel\_Fam  
Missing genotypes, 1000 AAUU families, NMod 1

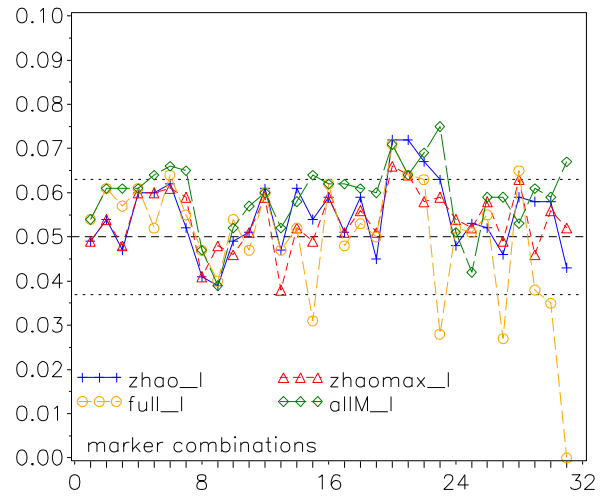


## NMod 2

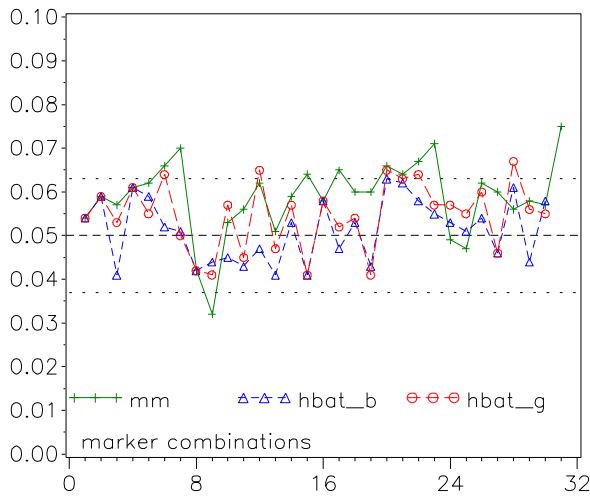
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 1000 trios, NMod 2, H0\_(A)



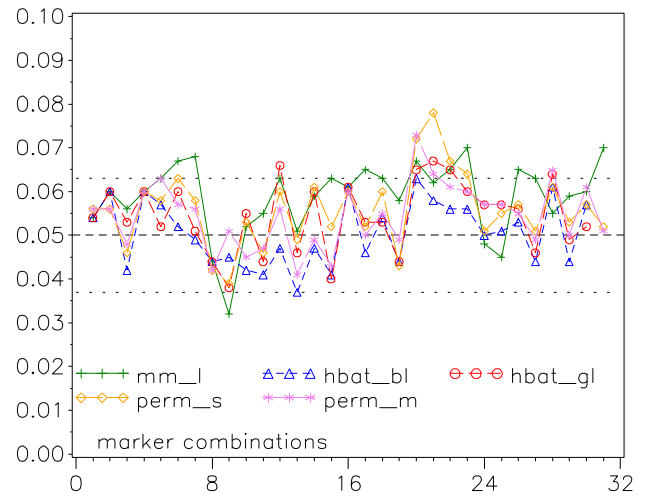
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 1000 trios, NMod 2, H0\_(L)



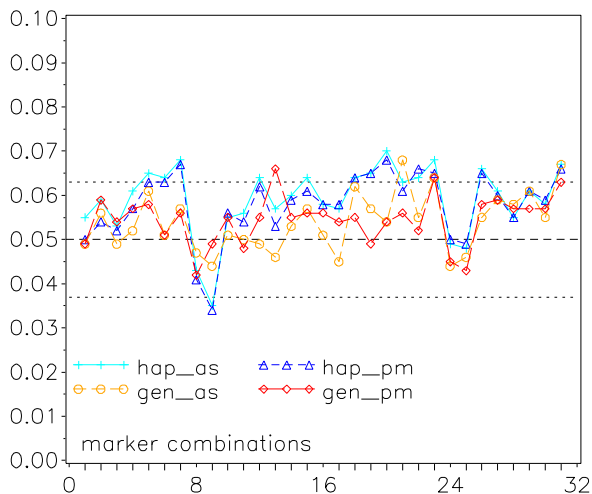
Type I error: FBAT,  
Missing genotypes, 1000 trios, Nmod 2, H0\_(A)



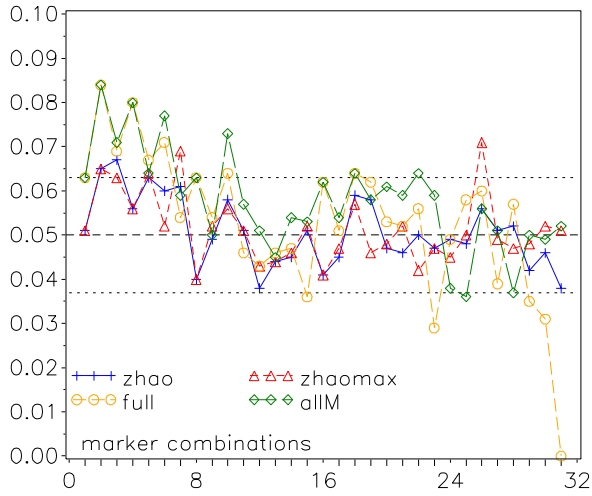
Type I error: FBAT,  
Missing genotypes, 1000 trios, Nmod 2, H0\_(L)



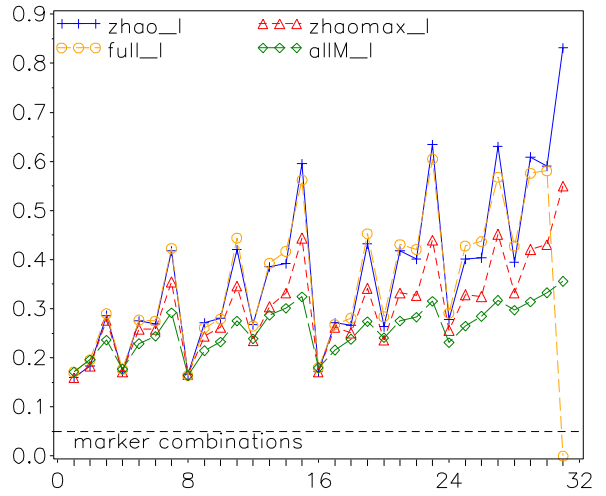
Type I error of Hotel\_Fam  
Missing genotypes, 1000 trios, NMod 2



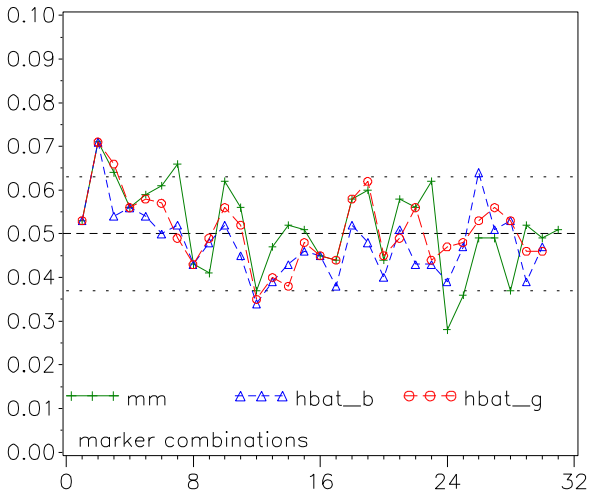
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 1000 AA fam., NMod 2, H0\_(A)



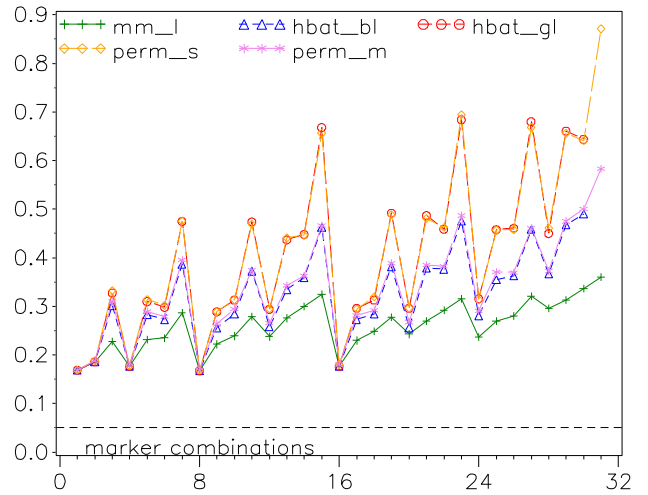
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 1000 AA fam., NMod 2, H0\_(L)



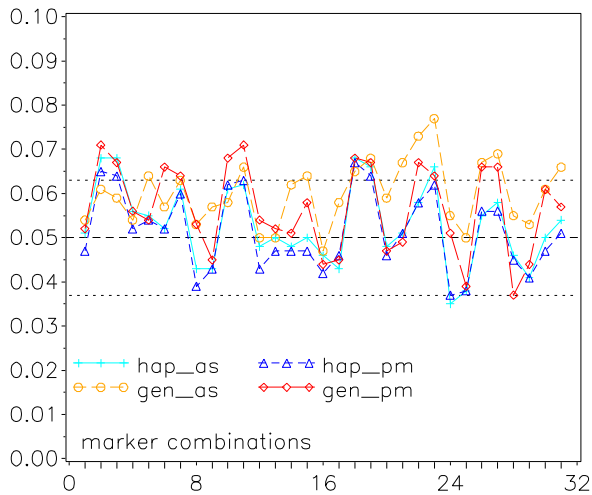
Type I error: FBAT,  
Missing genotypes, 1000 AA fam., Nmod 2, H0\_(A)



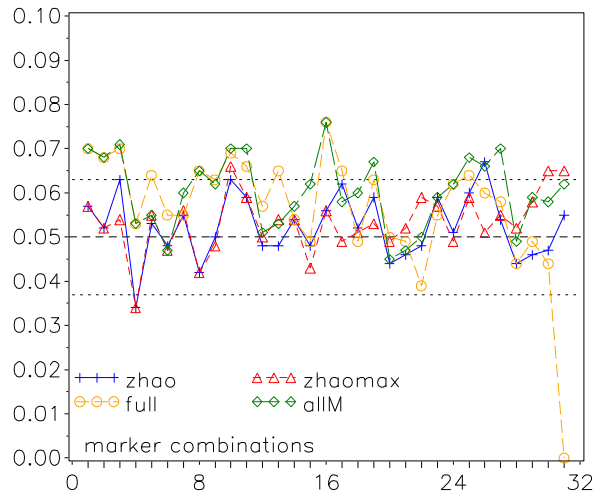
Type I error: FBAT,  
Missing genotypes, 1000 AA fam., Nmod 2, H0\_(L)



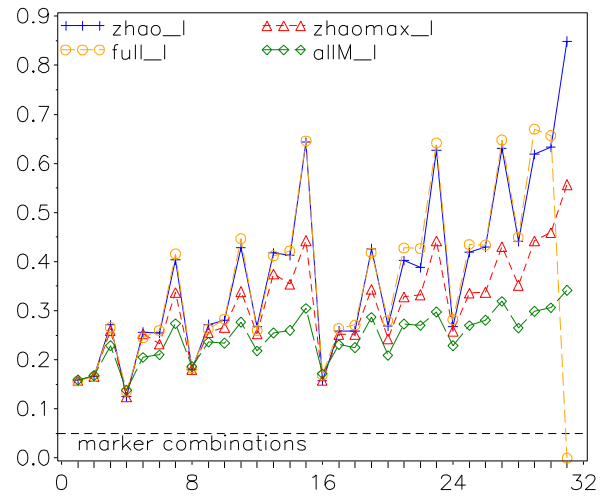
Type I error of Hotel\_Fam  
Missing genotypes, 1000 AA families, NMod 2



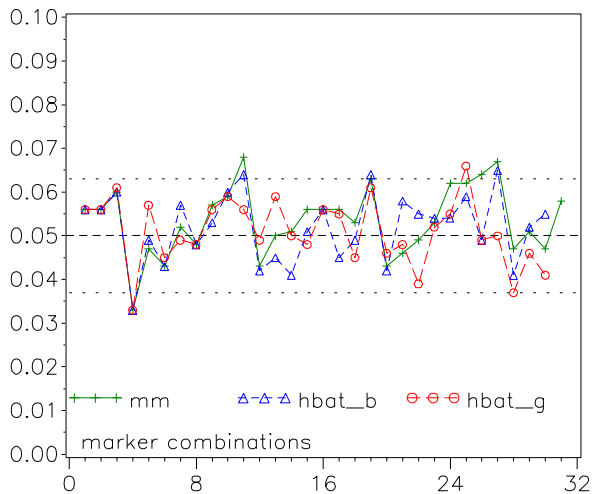
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 1000 AAUU fam., NMod 2, H0\_(A)



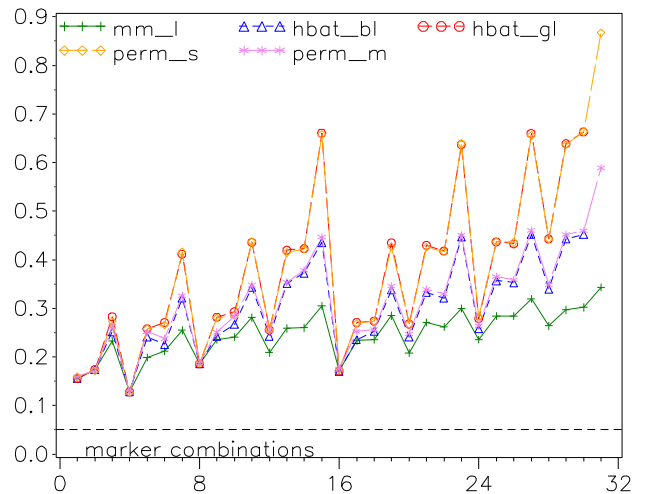
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 1000 AAUU fam., NMod 2, H0\_(L)



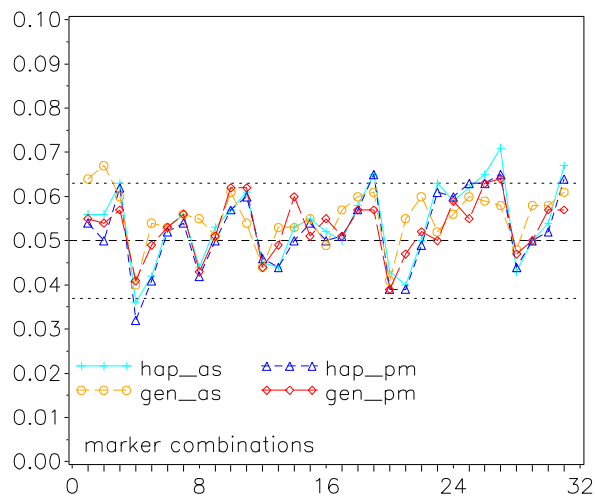
Type I error: FBAT,  
Missing genotypes, 1000 AAUU fam., Nmod 2, H0\_(A)



Type I error: FBAT,  
Missing genotypes, 1000 AAUU fam., Nmod 2, H0\_(L)

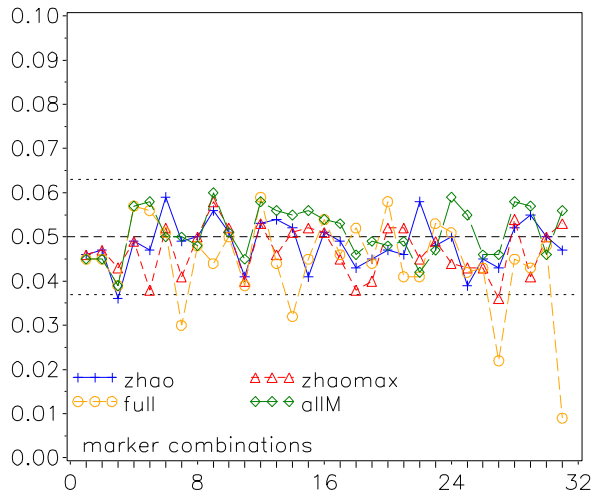


Type I error of Hotel\_Fam  
Missing genotypes, 1000 AAUU families, NMod 2

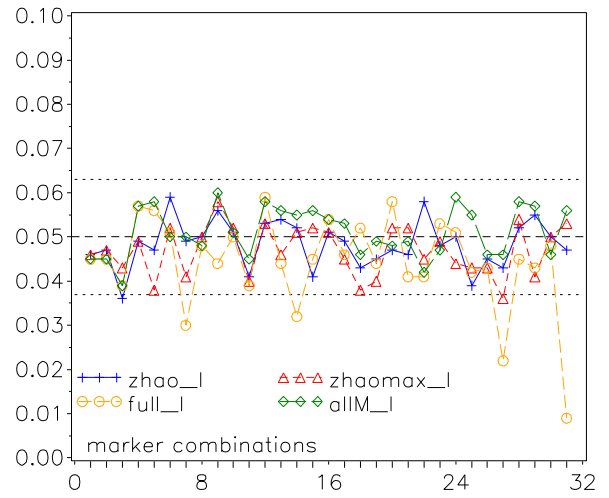


### NMod 3

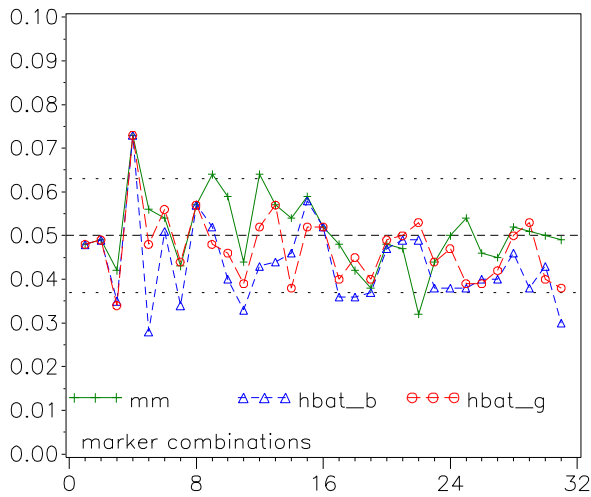
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 1000 trios, NMod 3, H0\_(A)



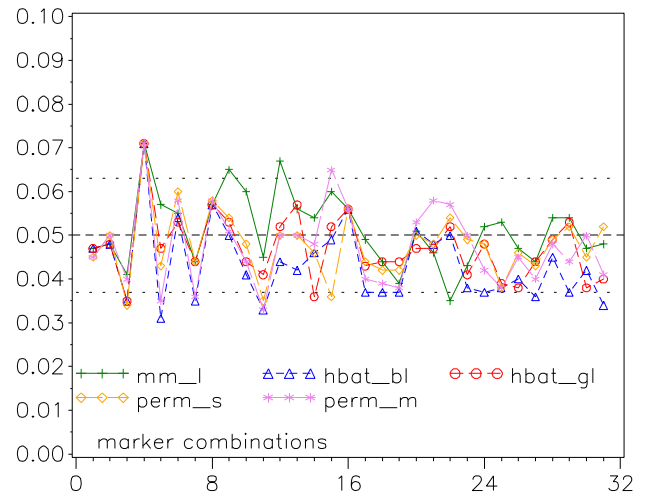
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 1000 trios, NMod 3, H0\_(L)



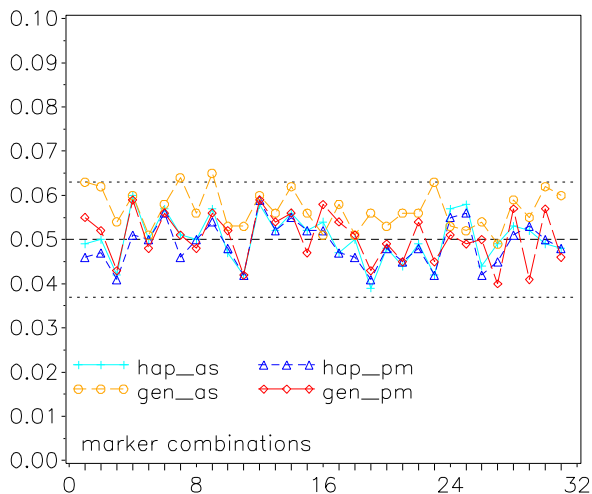
Type I error: FBAT,  
Missing genotypes, 1000 trios, Nmod 3, H0\_(A)



Type I error: FBAT,  
Missing genotypes, 1000 trios, Nmod 3, H0\_(L)

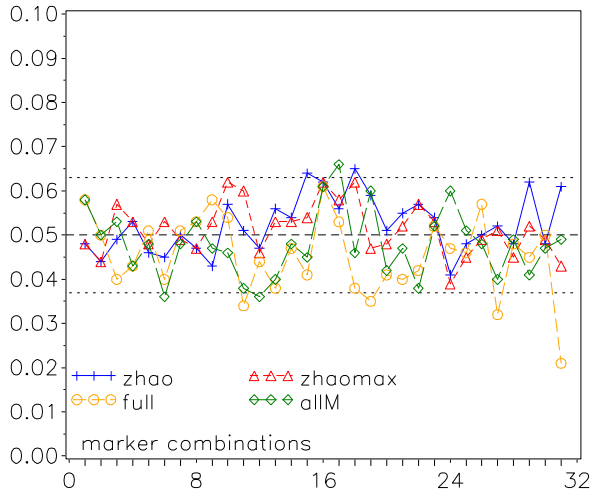


Type I error of Hotel\_Fam  
Missing genotypes, 1000 trios, NMod 3

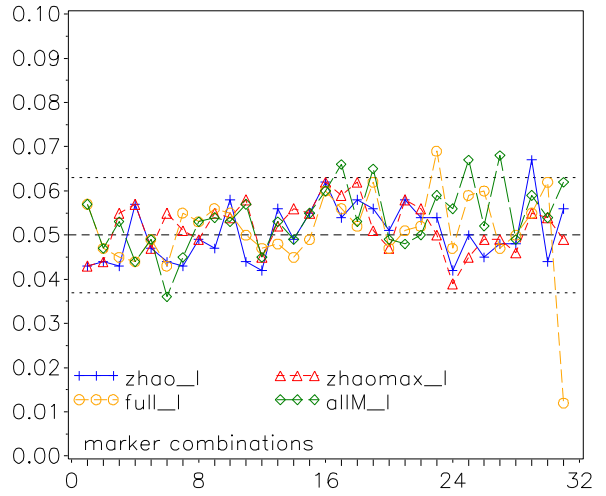




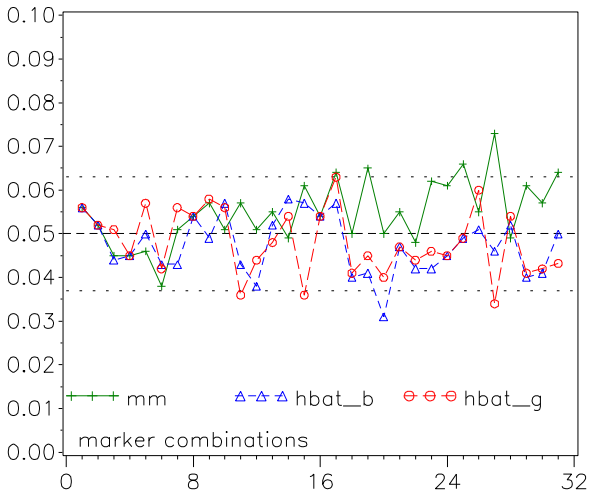
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 1000 AA fam., NMod 3, H0\_(A)



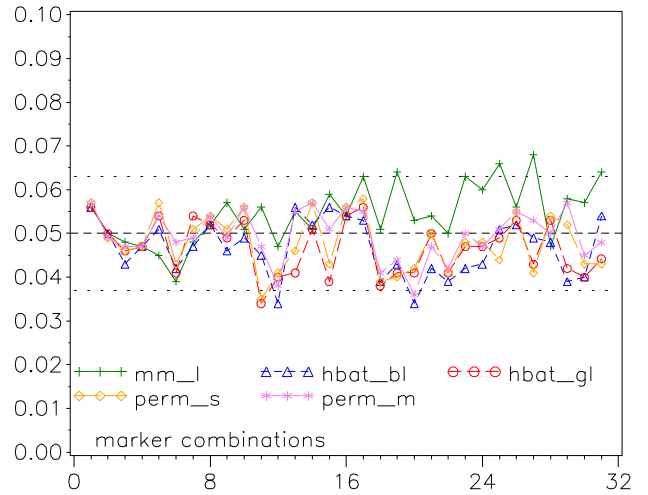
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 1000 AA fam., NMod 3, H0\_(L)



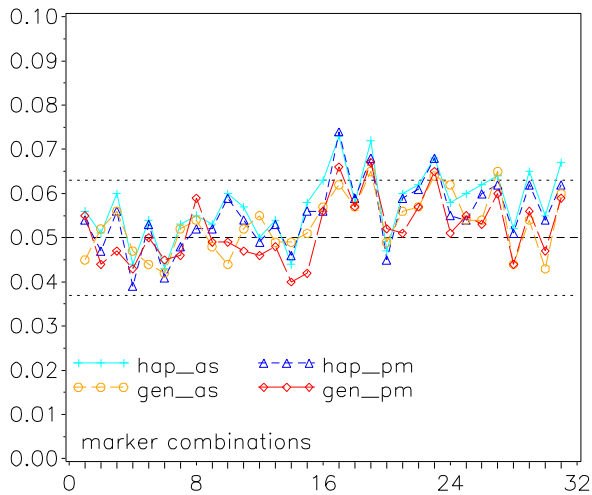
Type I error: FBAT,  
Missing genotypes, 1000 AA fam., Nmod 3, H0\_(A)



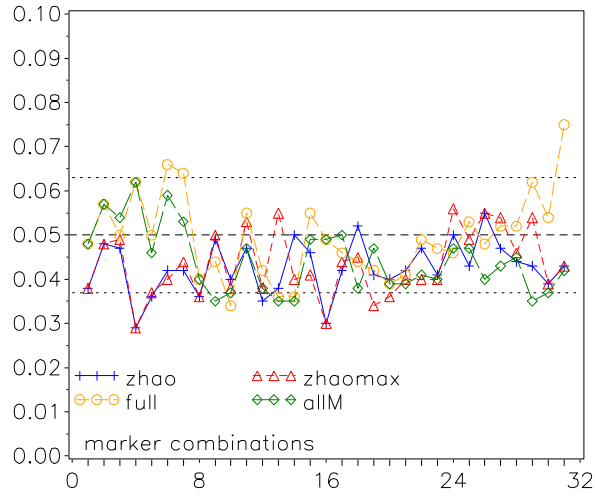
Type I error: FBAT,  
Missing genotypes, 1000 AA fam., Nmod 3, H0\_(L)



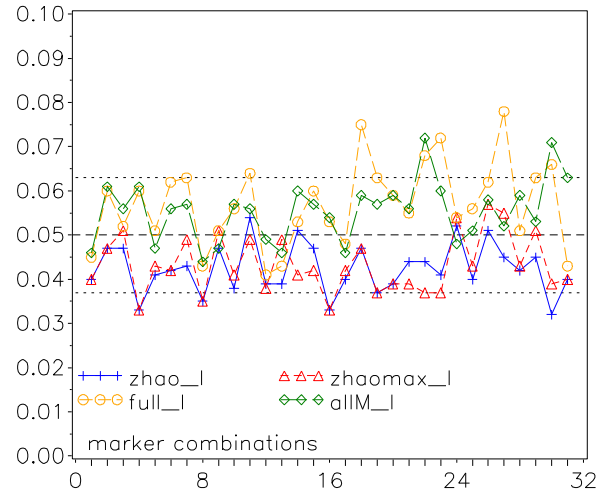
Type I error of Hotel\_Fam  
Missing genotypes, 1000 AA families, NMod 3



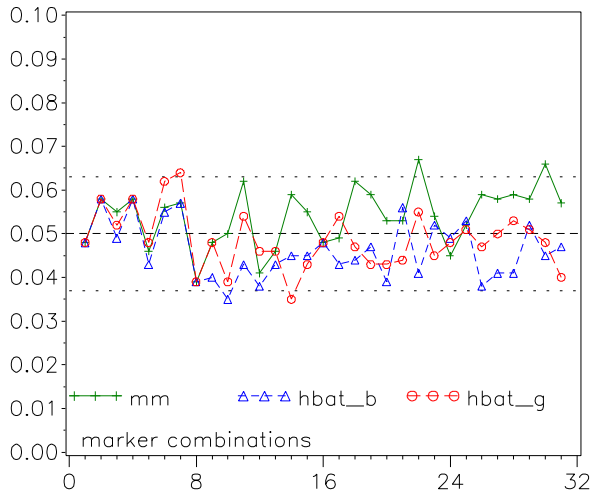
Type I error of a) Famhap b) Unphased, Missing genotypes, 1000 AAU fam., NMod 3, H0\_(A)



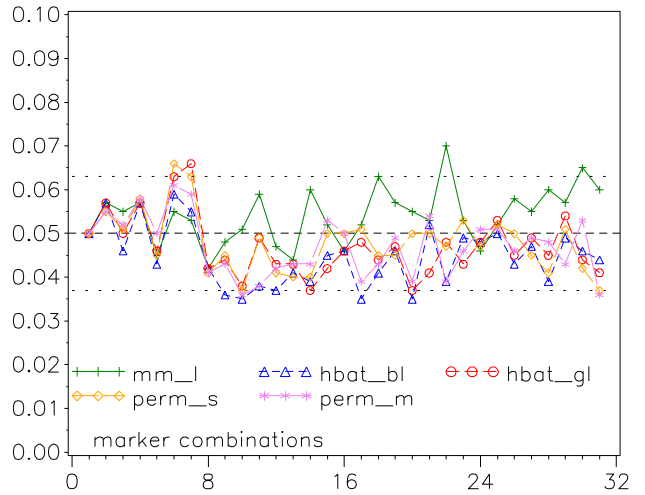
Type I error of a) Famhap b) Unphased, Missing genotypes, 1000 AAU fam., NMod 3, H0\_(L)



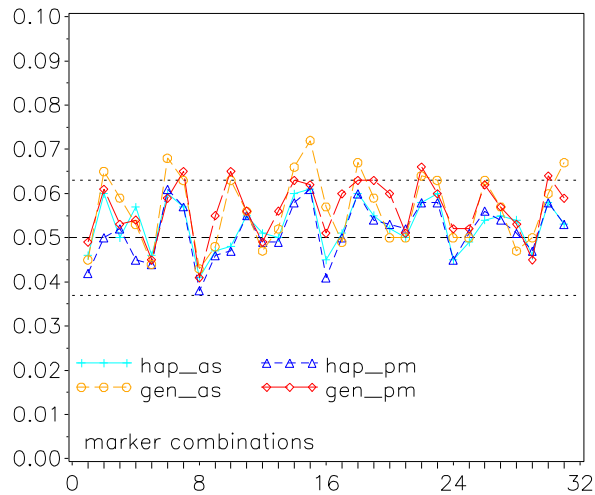
Type I error: FBAT, Missing genotypes, 1000 AAU fam., Nmod 3, H0\_(A)



Type I error: FBAT, Missing genotypes, 1000 AAU fam., Nmod 3, H0\_(L)



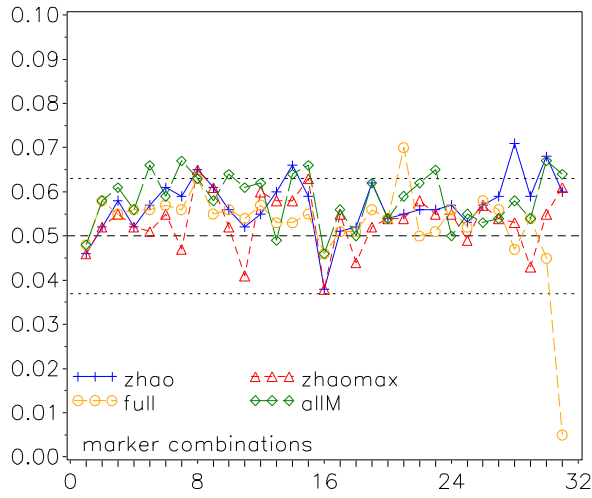
Type I error of Hotel\_Fam Missing genotypes, 1000 AAU families, NMod 3



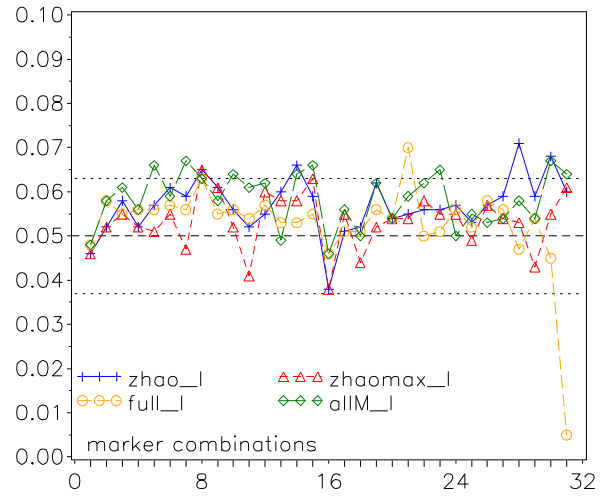
# 2500 Familien

## NMod 1

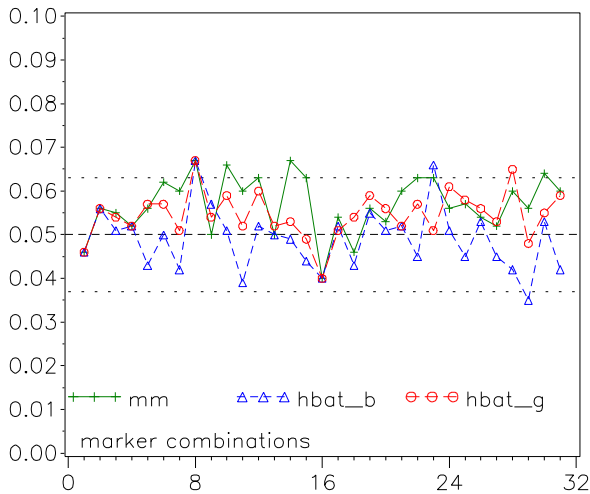
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 2500 trios, NMod 1, H0\_(A)



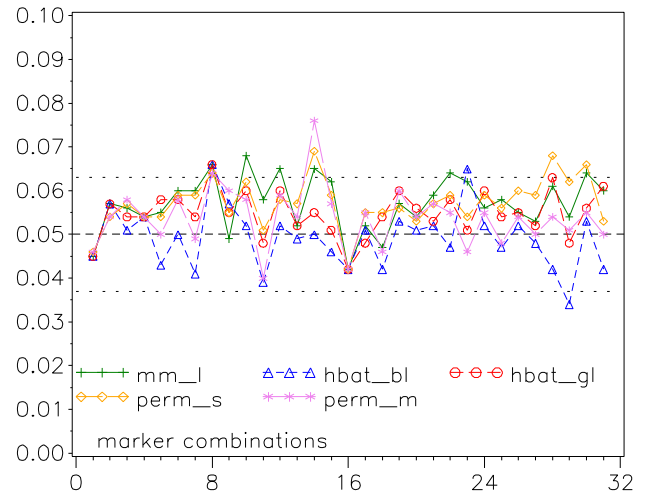
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 2500 trios, NMod 1, H0\_(L)



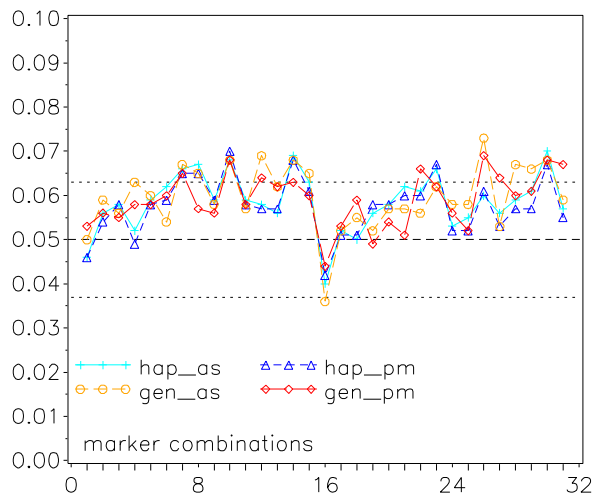
Type I error: FBAT,  
Missing genotypes, 2500 trios, Nmod 1, H0\_(A)



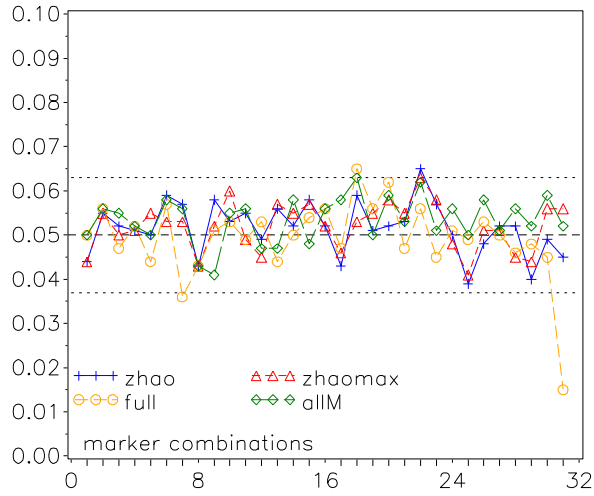
Type I error: FBAT,  
Missing genotypes, 2500 trios, Nmod 1, H0\_(L)



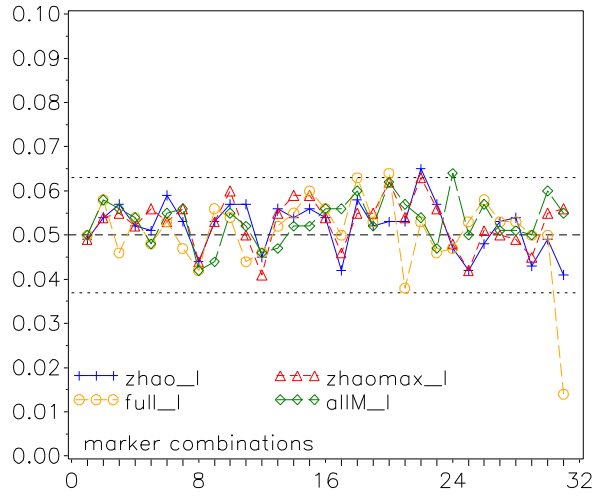
Type I error of Hotel\_Fam  
Missing genotypes, 2500 trios, NMod 1



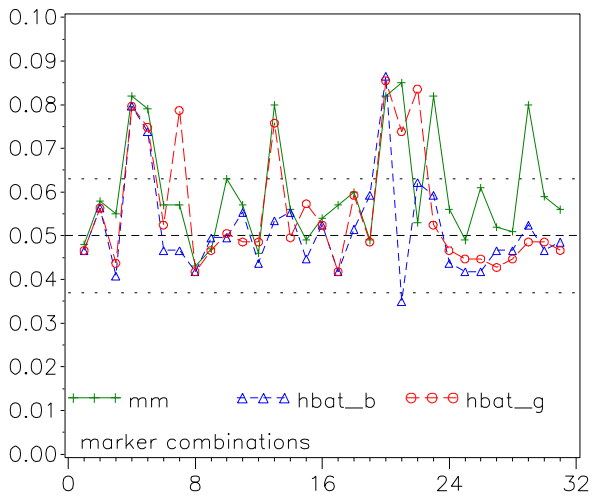
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 2500 AA fam., NMod 1, H0\_(A)



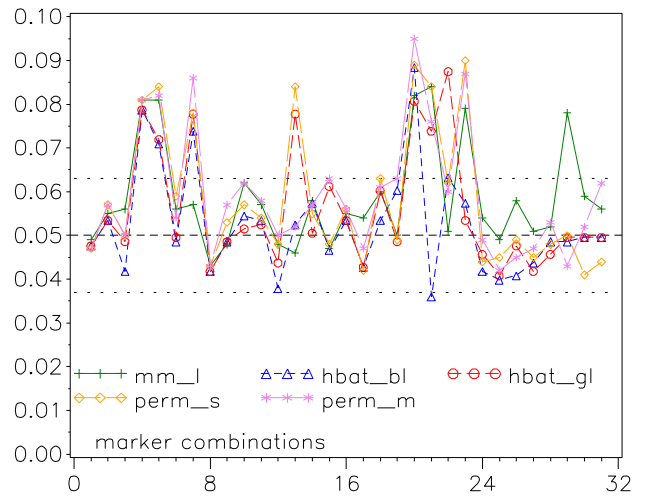
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 2500 AA fam., NMod 1, H0\_(L)



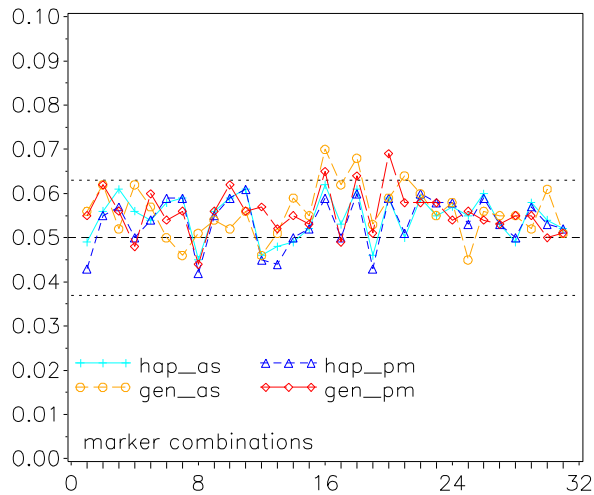
Type I error: FBAT,  
Missing genotypes, 2500 AA fam., Nmod 1, H0\_(A)



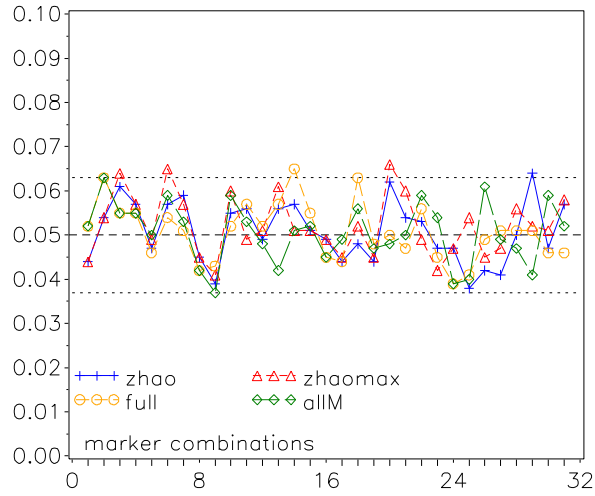
Type I error: FBAT,  
Missing genotypes, 2500 AA fam., Nmod 1, H0\_(L)



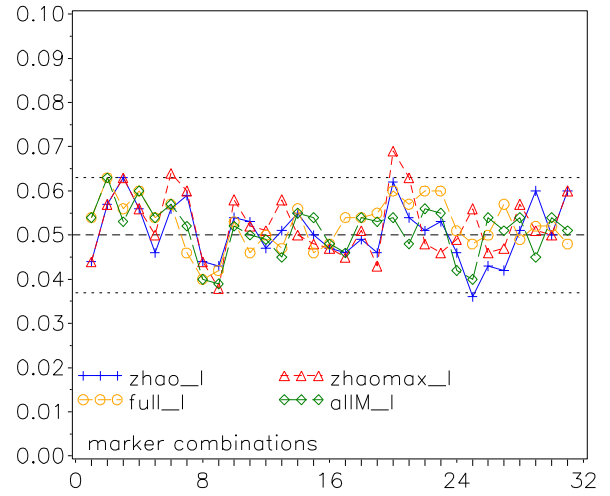
Type I error of Hotel\_Fam  
Missing genotypes, 2500 AA families, NMod 1



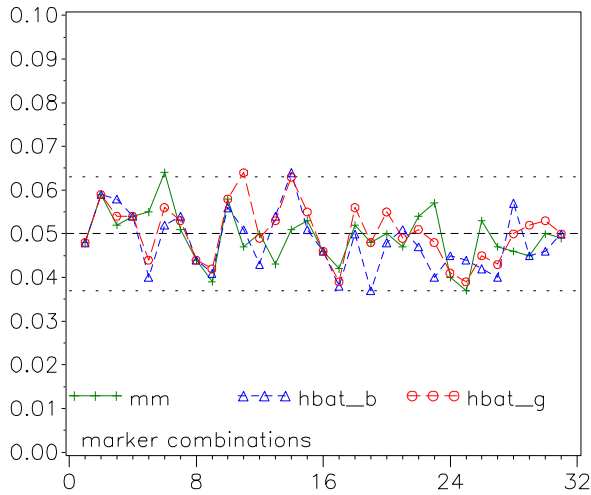
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 2500 AAUU fam., NMod 1, H0\_(A)



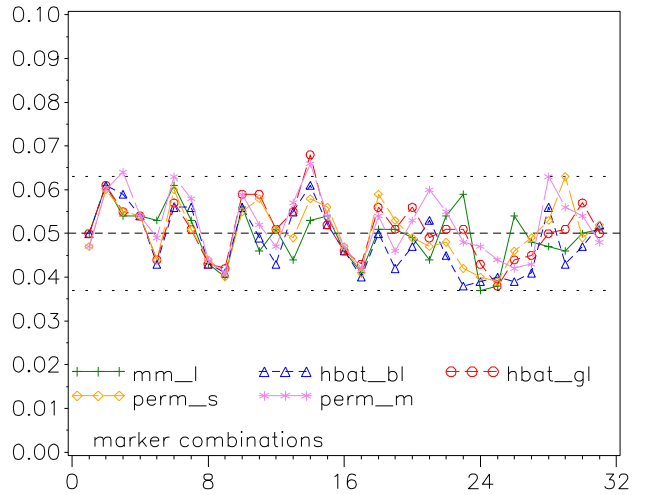
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 2500 AAUU fam., NMod 1, H0\_(L)



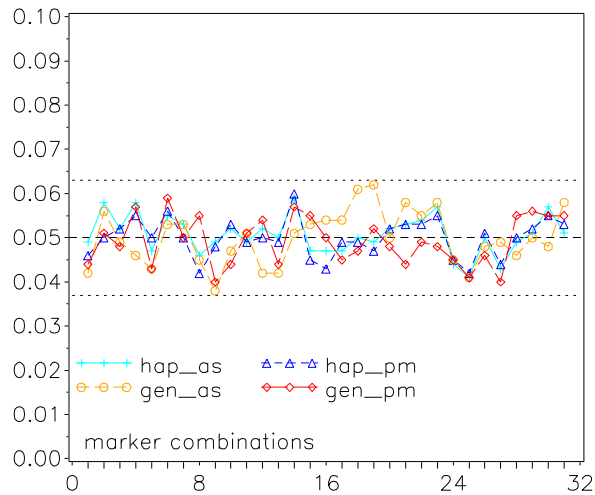
Type I error: FBAT,  
Missing genotypes, 2500 AAUU fam., Nmod 1, H0\_(A)



Type I error: FBAT,  
Missing genotypes, 2500 AAUU fam., Nmod 1, H0\_(L)

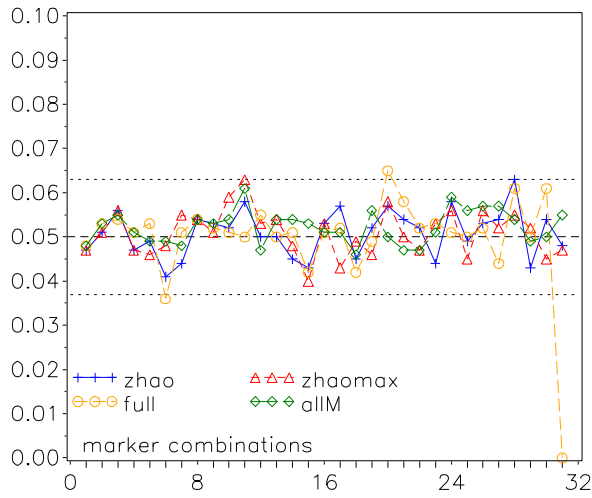


Type I error of Hotel\_Fam  
Missing genotypes, 2500 AAUU families, NMod 1

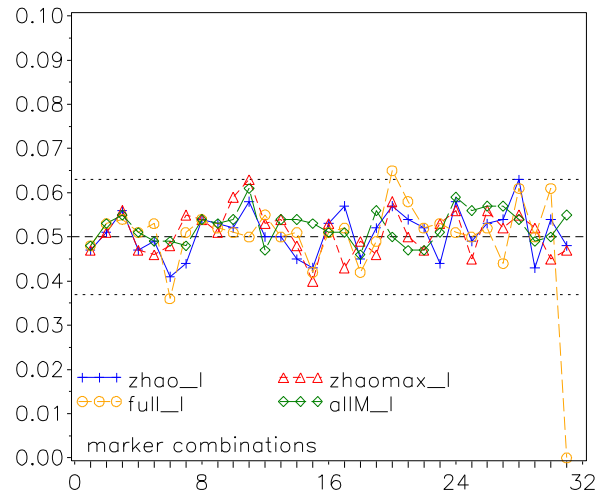


## NMod 2

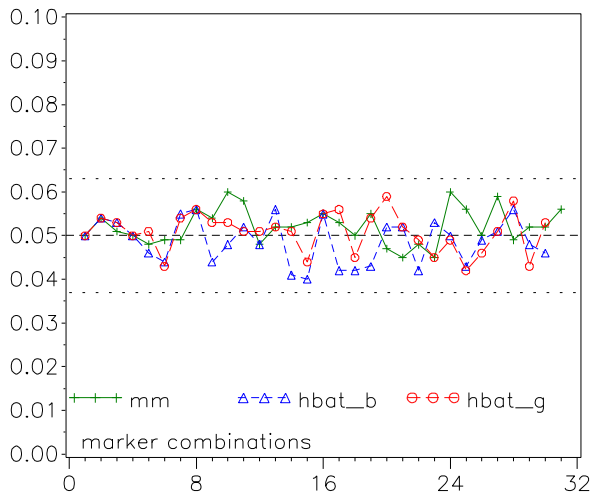
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 2500 trios, NMod 2, H0\_(A)



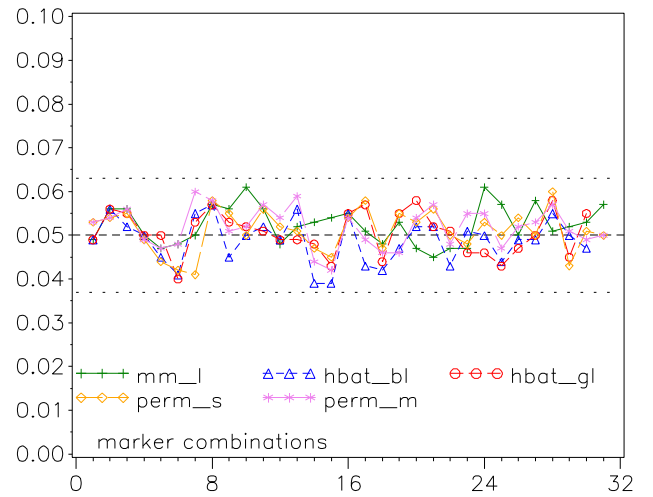
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 2500 trios, NMod 2, H0\_(L)



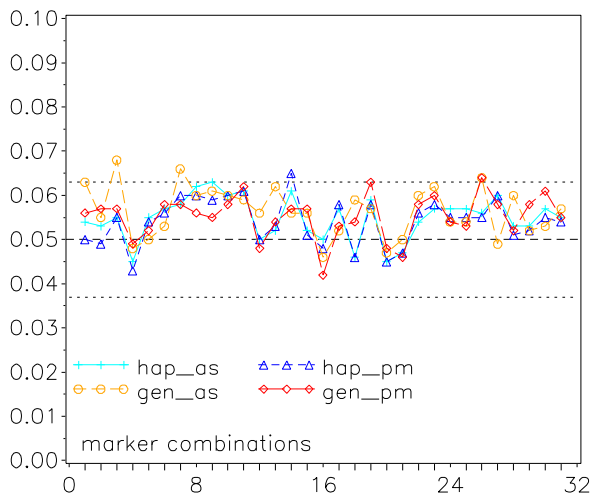
Type I error: FBAT,  
Missing genotypes, 2500 trios, Nmod 2, H0\_(A)



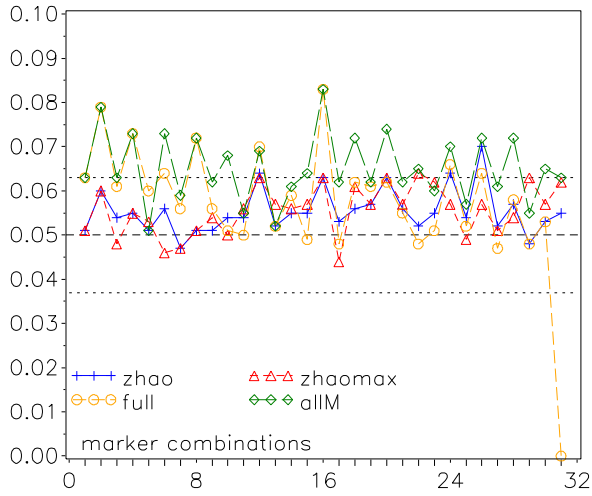
Type I error: FBAT,  
Missing genotypes, 2500 trios, Nmod 2, H0\_(L)



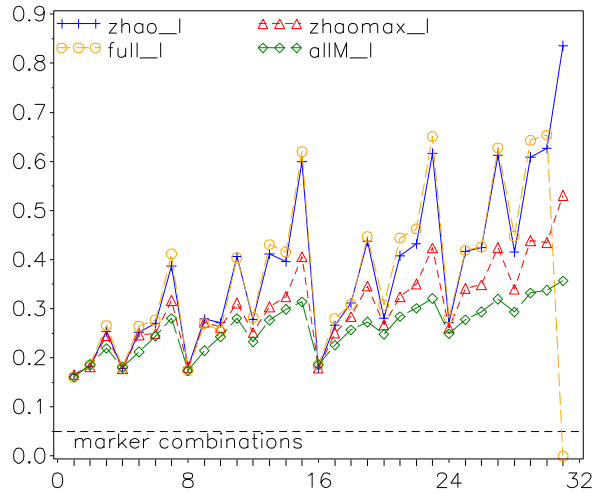
Type I error of Hotel\_Fam  
Missing genotypes, 2500 trios, NMod 2



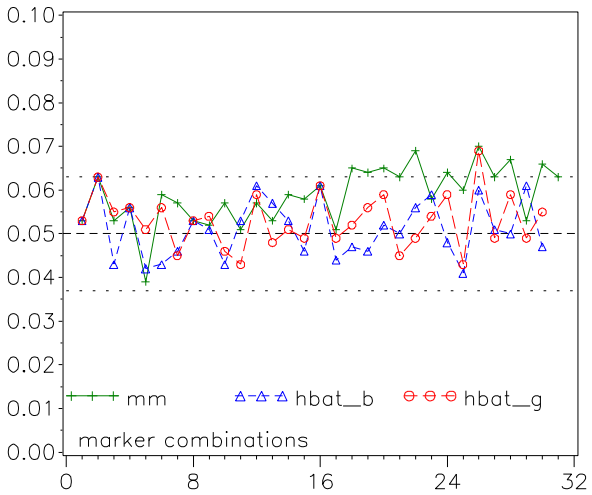
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 2500 AA fam., NMod 2, H0\_(A)



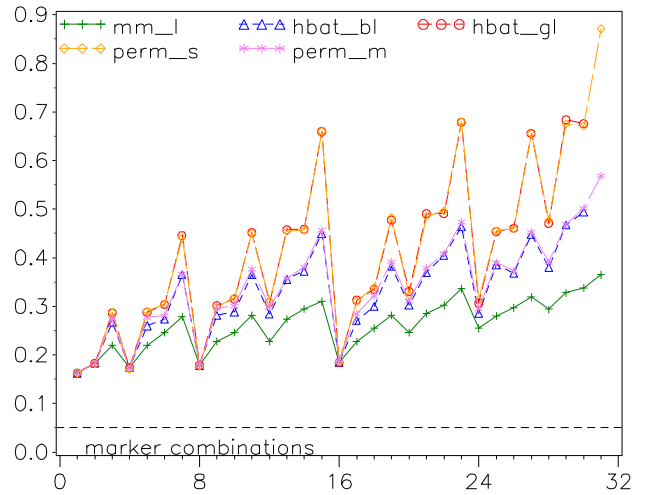
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 2500 AA fam., NMod 2, H0\_(L)



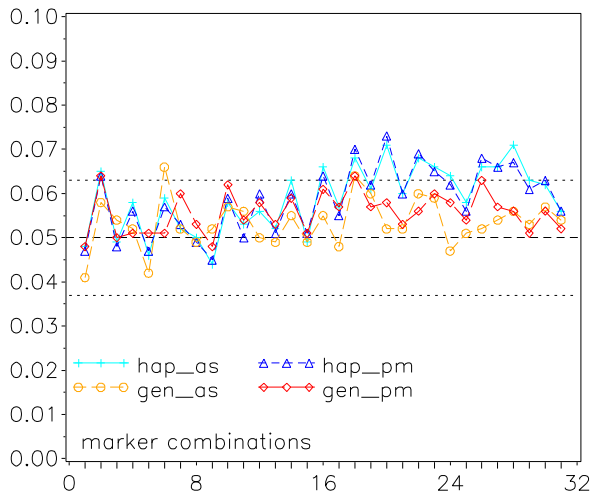
Type I error: FBAT,  
Missing genotypes, 2500 AA fam., Nmod 2, H0\_(A)



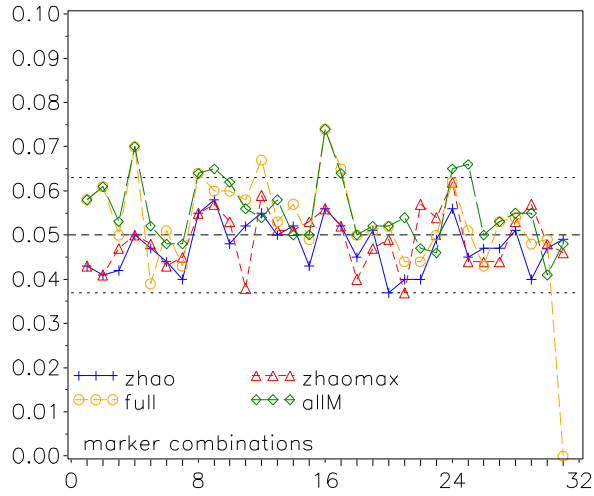
Type I error: FBAT,  
Missing genotypes, 2500 AA fam., Nmod 2, H0\_(L)



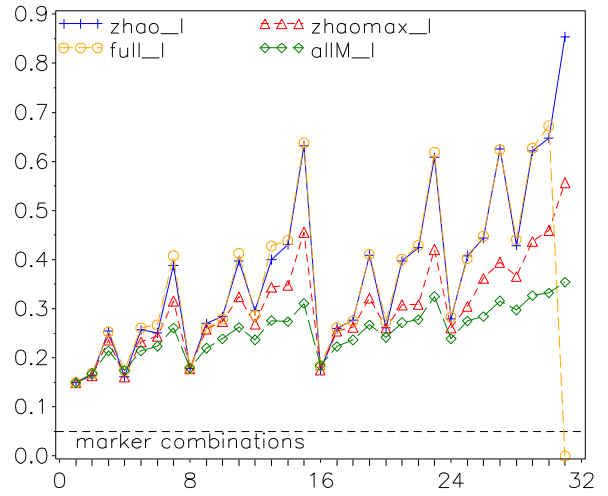
Type I error of Hotel\_Fam  
Missing genotypes, 2500 AA families, NMod 2



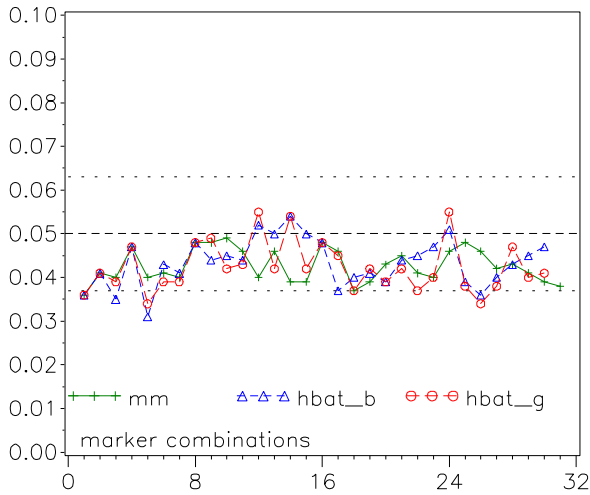
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 2500 AAUU fam., NMod 2, H0\_(A)



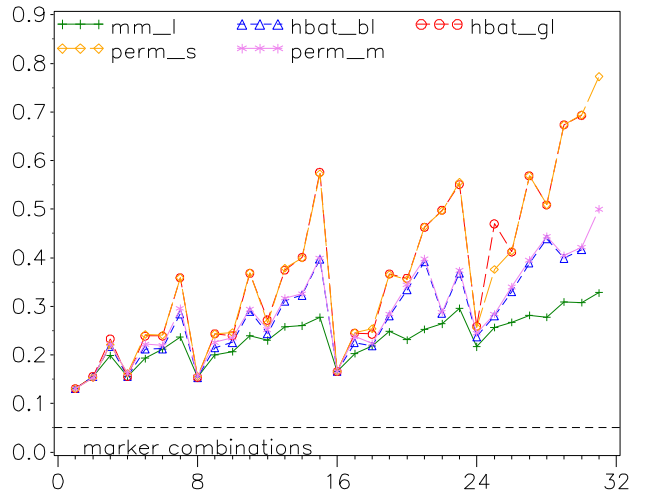
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 2500 AAUU fam., NMod 2, H0\_(L)



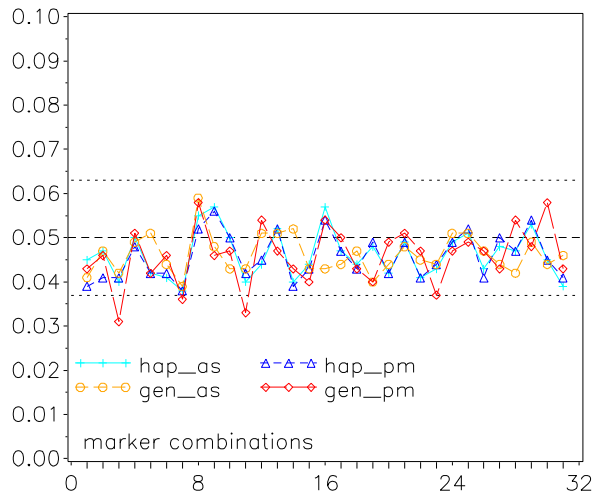
Type I error: FBAT,  
Missing genotypes, 2500 AAUU fam., Nmod 2, H0\_(A)



Type I error: FBAT,  
Missing genotypes, 2500 AAUU fam., Nmod 2, H0\_(L)



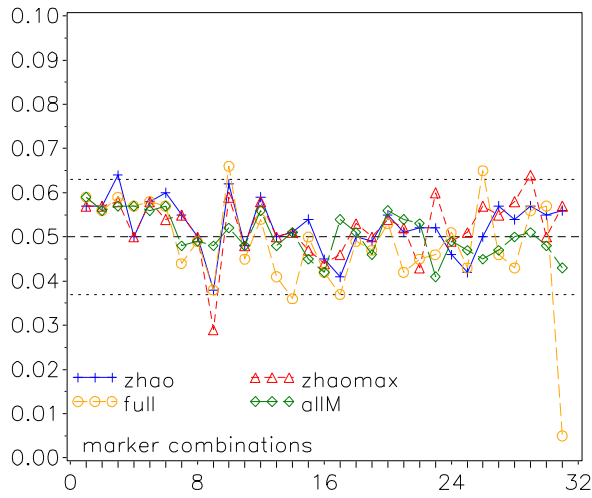
Type I error of Hotel\_Fam  
Missing genotypes, 2500 AAUU families, NMod 2



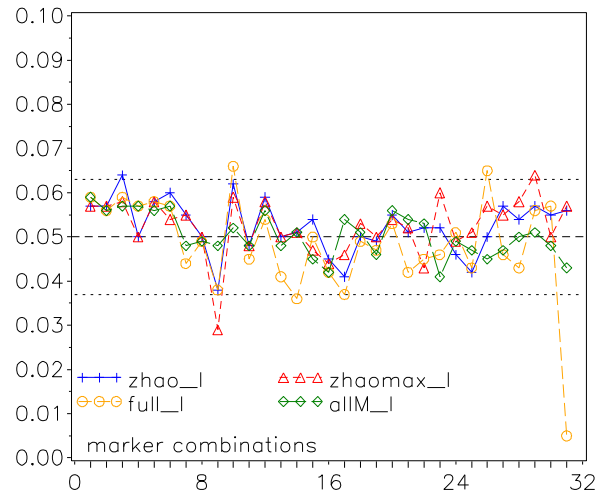


### NMod 3

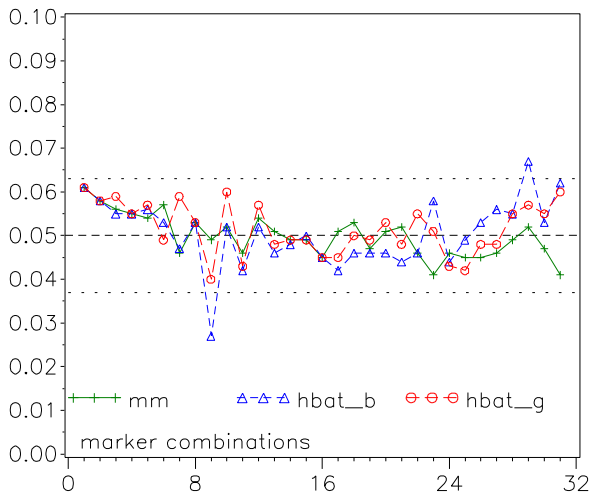
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 2500 trios, NMod 3, H0\_(A)



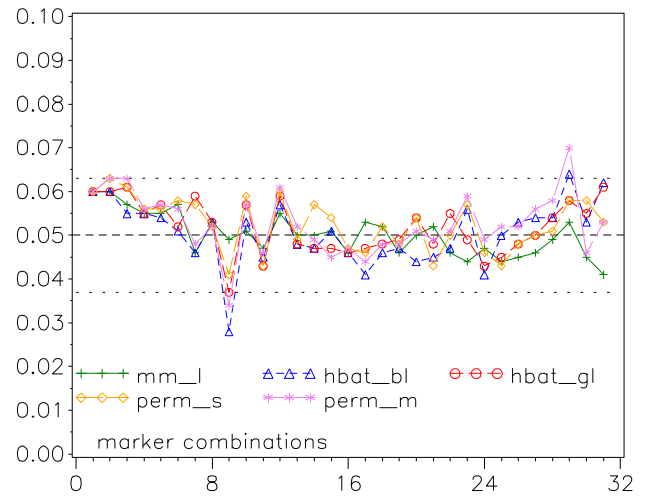
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 2500 trios, NMod 3, H0\_(L)



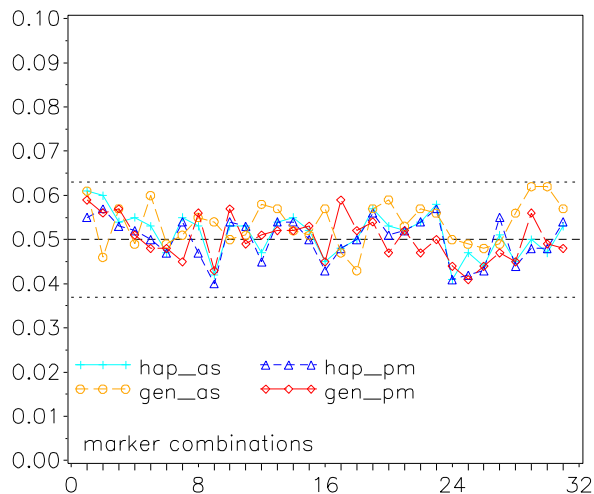
Type I error: FBAT,  
Missing genotypes, 2500 trios, Nmod 3, H0\_(A)



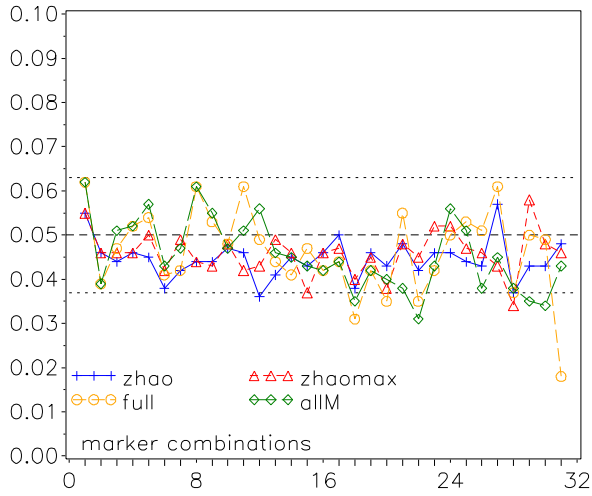
Type I error: FBAT,  
Missing genotypes, 2500 trios, Nmod 3, H0\_(L)



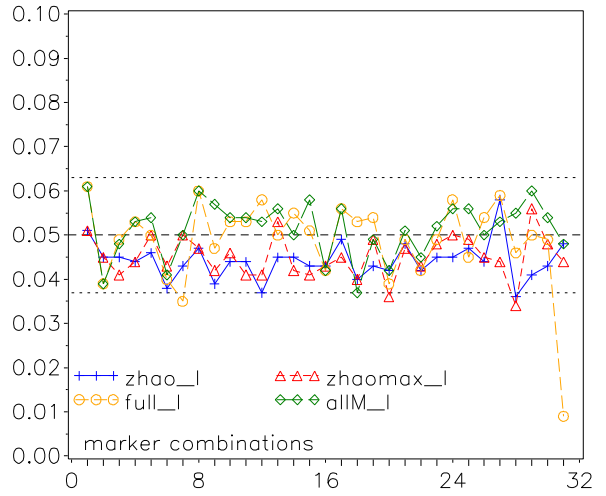
Type I error of Hotel\_Fam  
Missing genotypes, 2500 trios, NMod 3



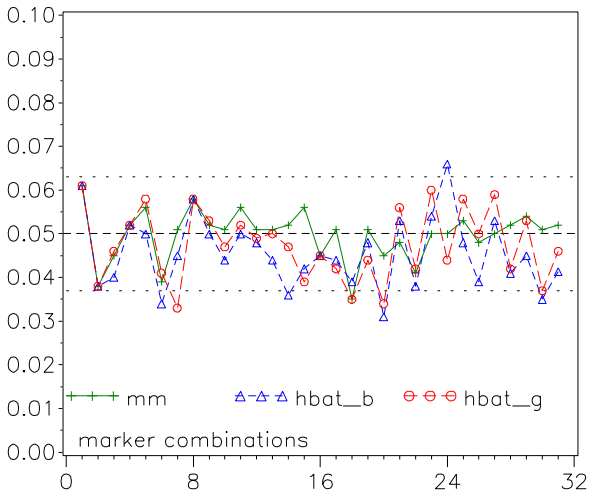
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 2500 AA fam., NMod 3, H0\_(A)



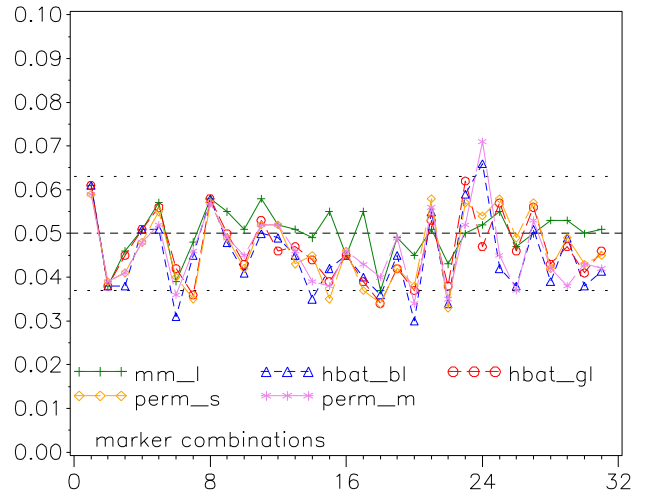
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 2500 AA fam., NMod 3, H0\_(L)



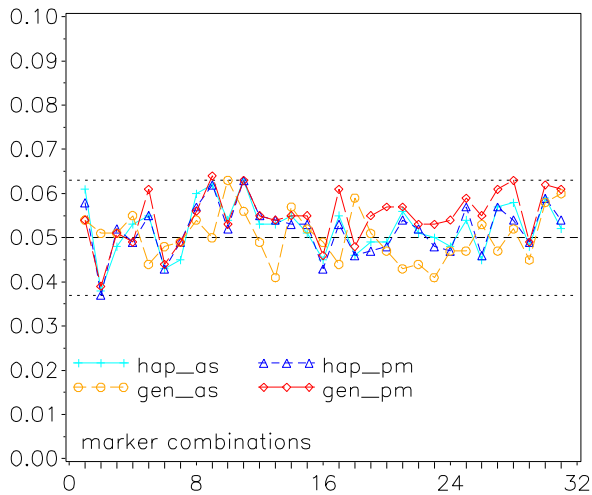
Type I error: FBAT,  
Missing genotypes, 2500 AA fam., Nmod 3, H0\_(A)



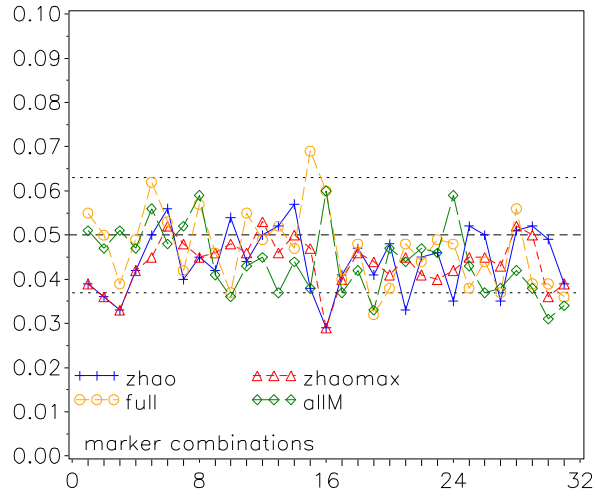
Type I error: FBAT,  
Missing genotypes, 2500 AA fam., Nmod 3, H0\_(L)



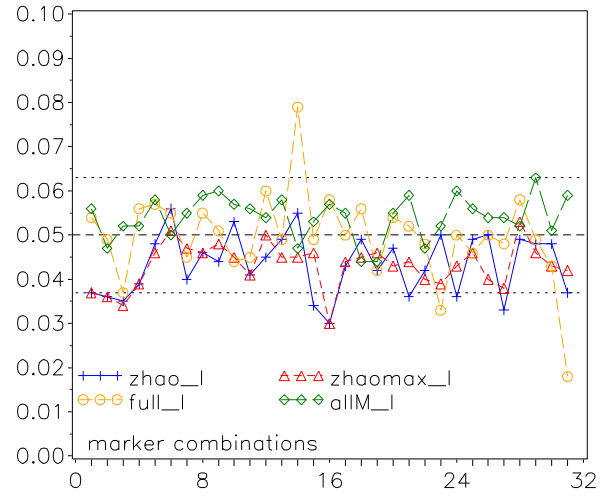
Type I error of Hotel\_Fam  
Missing genotypes, 2500 AA families, NMod 3



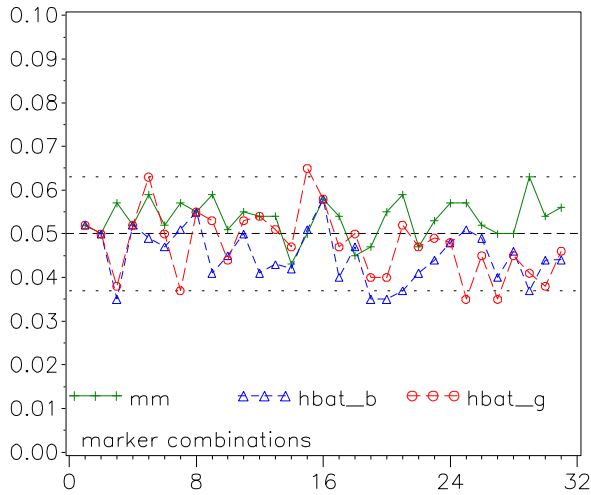
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 2500 AAUU fam., NMod 3, H0\_(A)



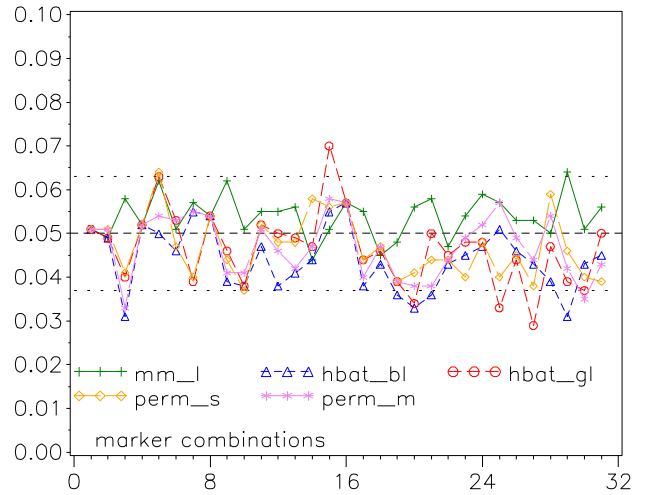
Type I error of a) Famhap b) Unphased,  
Missing genotypes, 2500 AAUU fam., NMod 3, H0\_(L)



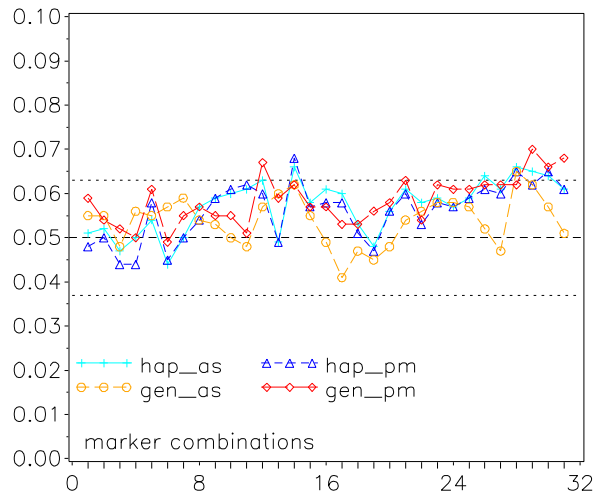
Type I error: FBAT,  
Missing genotypes, 2500 AAUU fam., Nmod 3, H0\_(A)



Type I error: FBAT,  
Missing genotypes, 2500 AAUU fam., Nmod 3, H0\_(L)



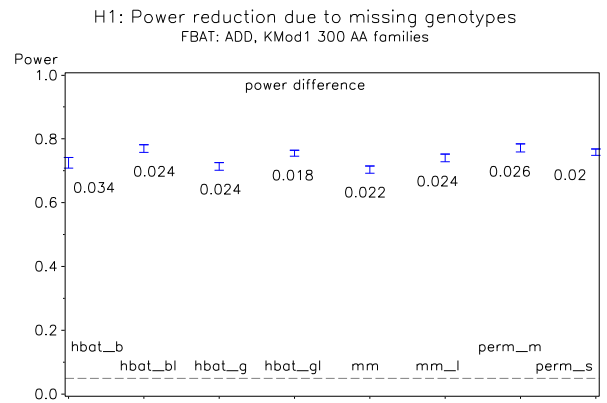
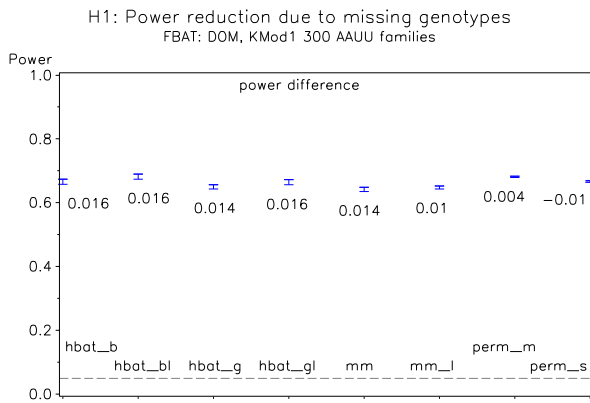
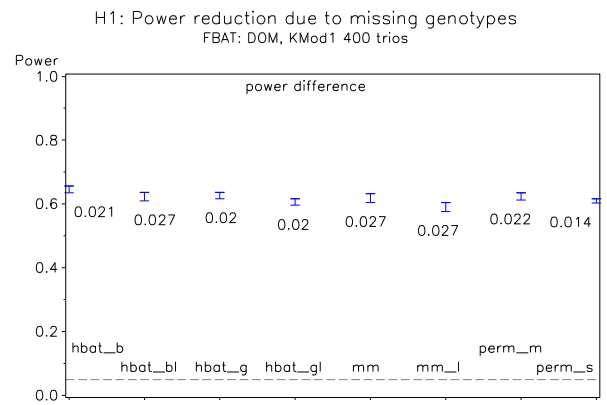
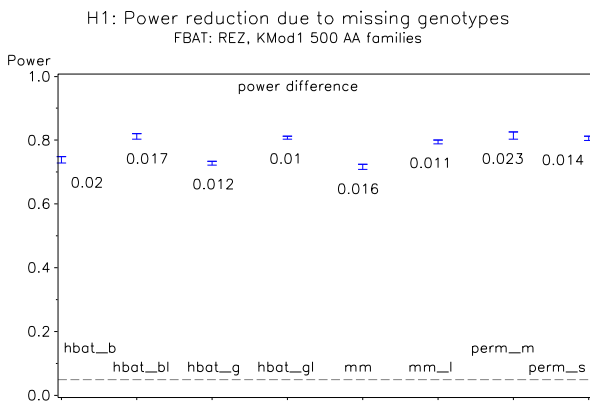
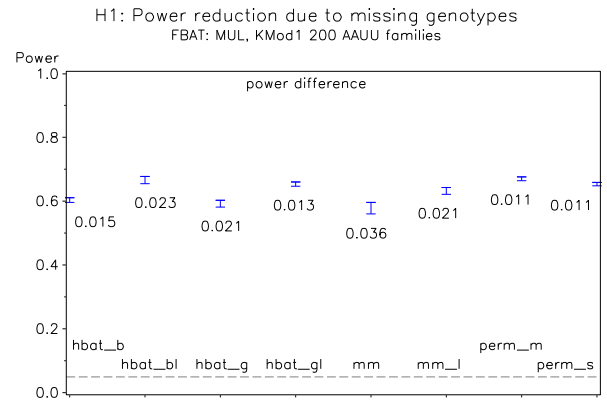
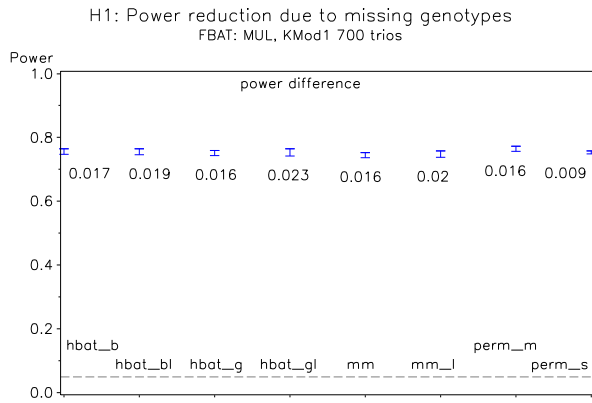
Type I error of Hotel\_Fam  
Missing genotypes, 2500 AAUU families, NMod 3



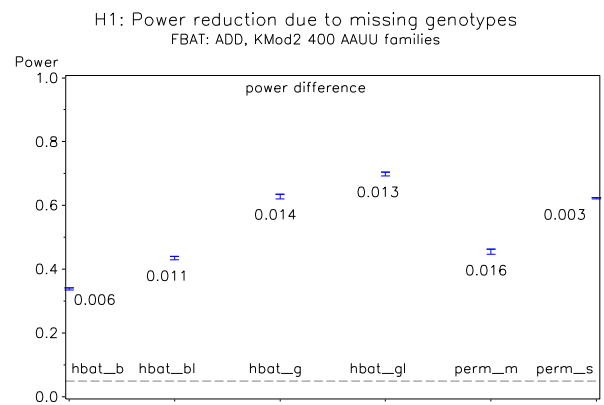
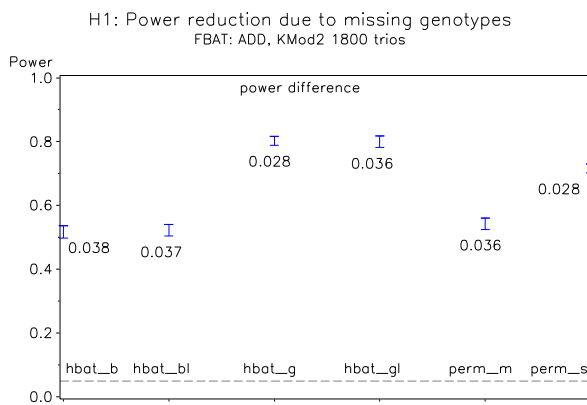
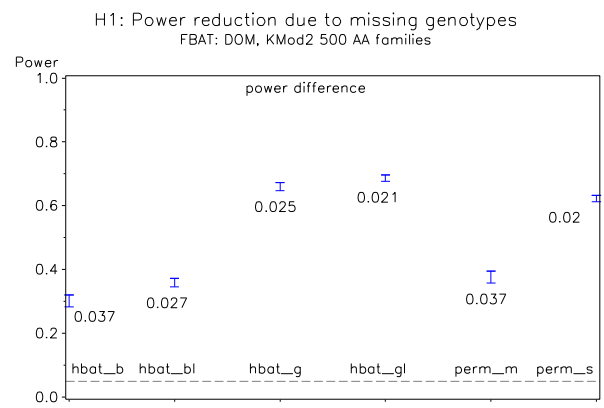
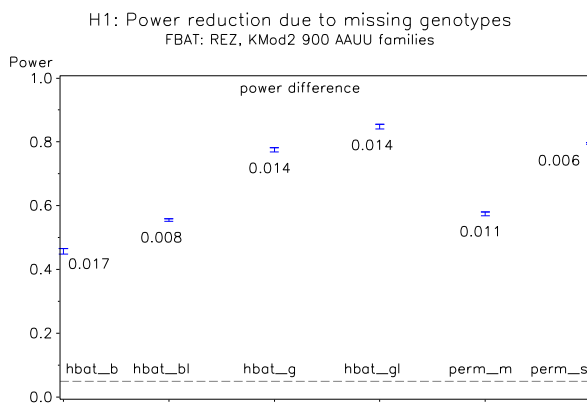
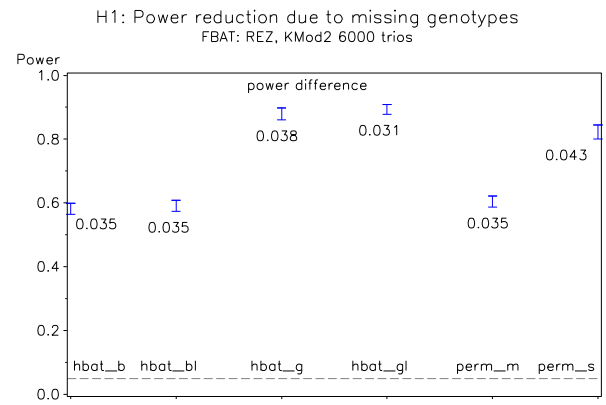
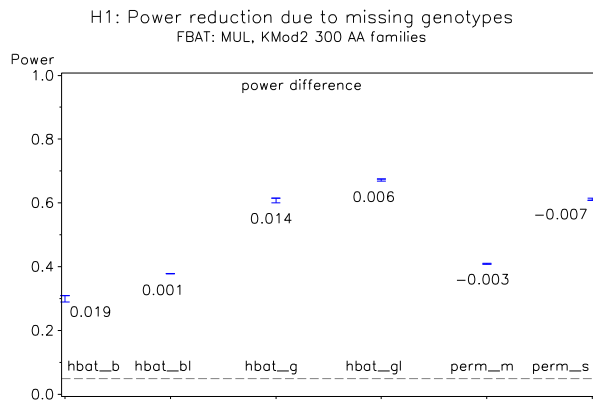
# Analyse mit fehlenden Genotypen: Powerverlust

## FBAT

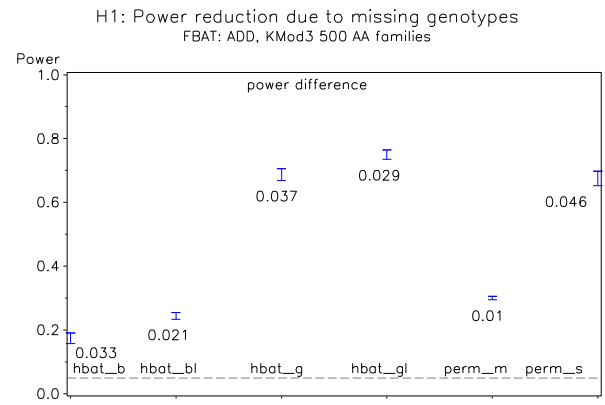
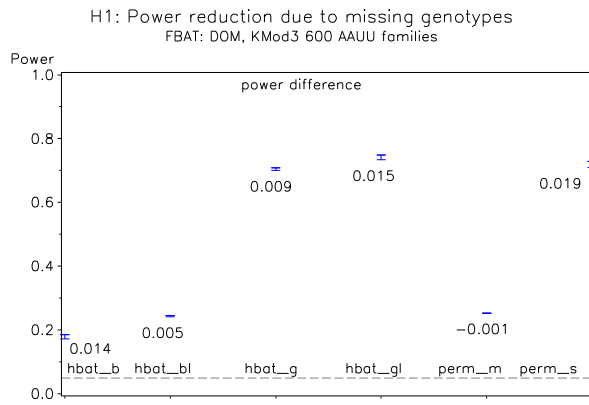
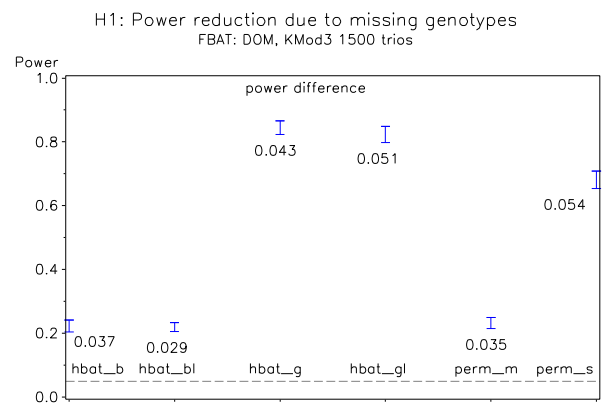
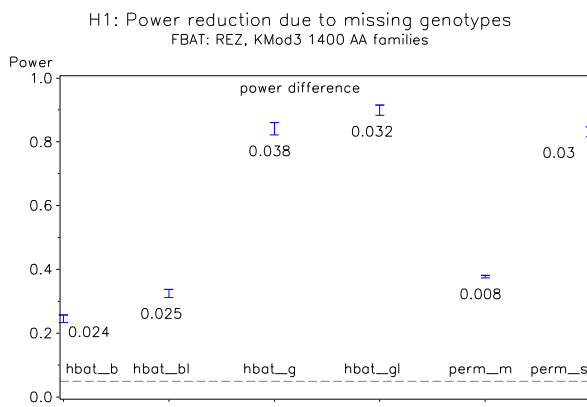
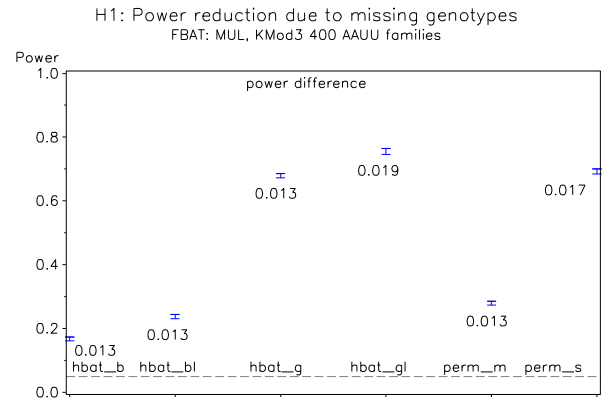
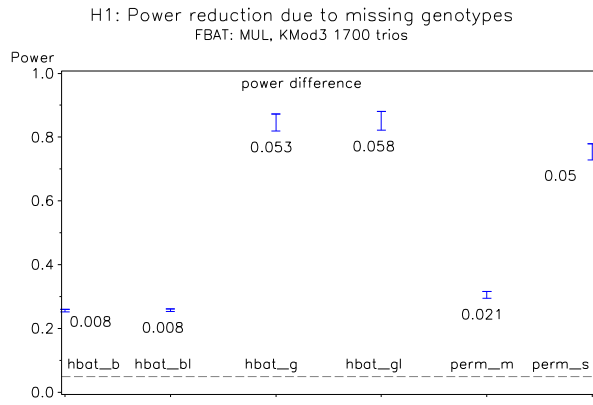
### KMod 1



## KMod 2

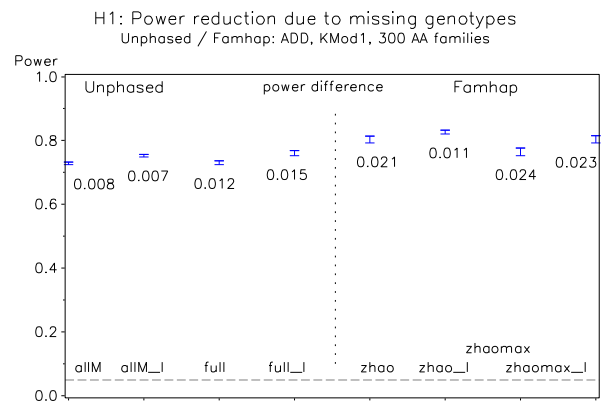
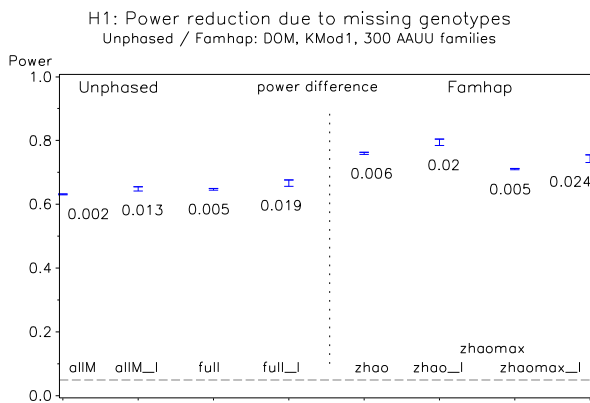
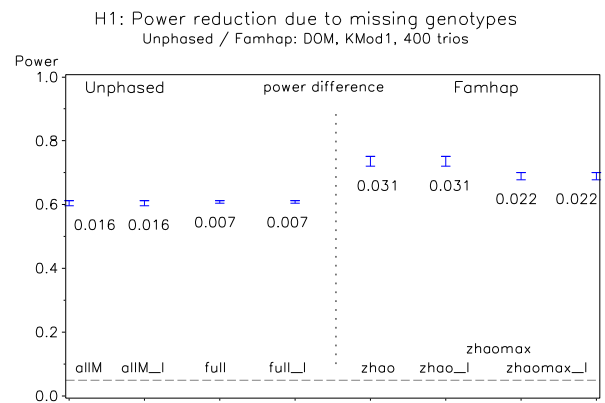
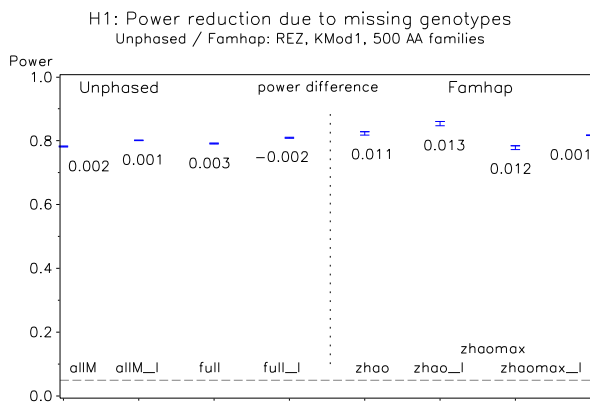
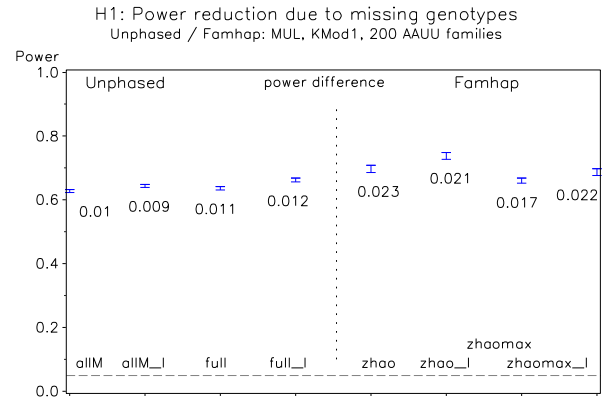
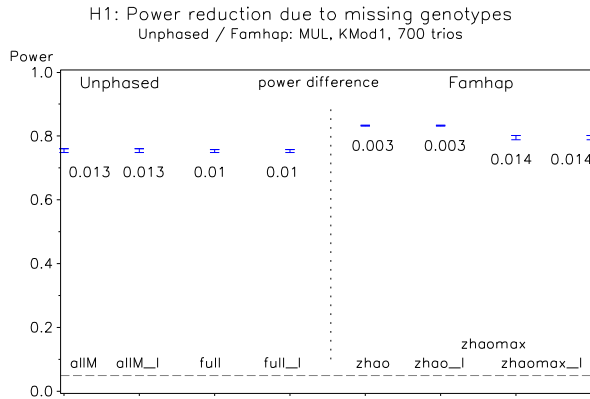


### KMod3

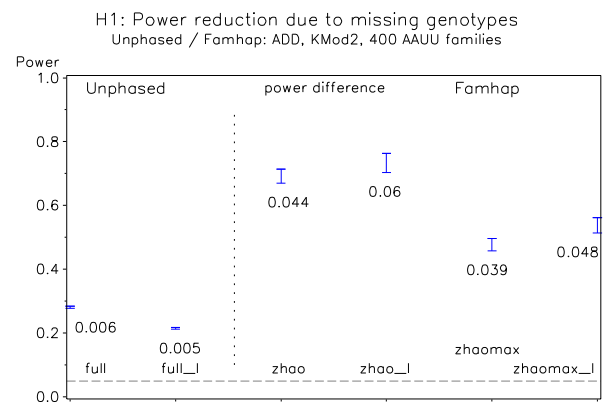
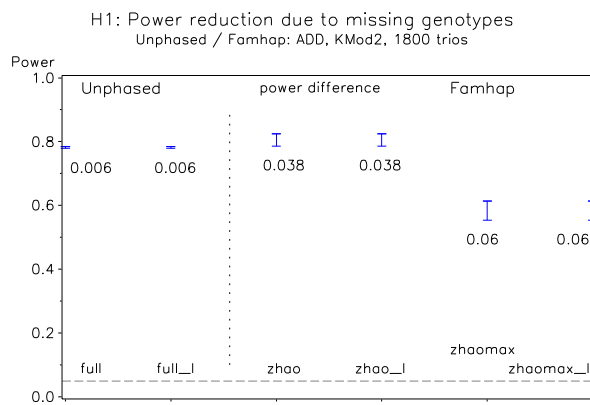
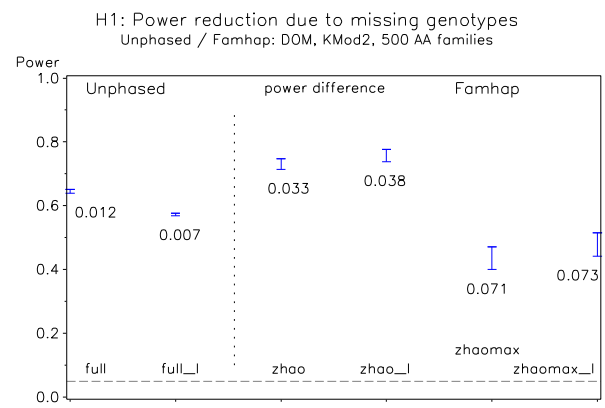
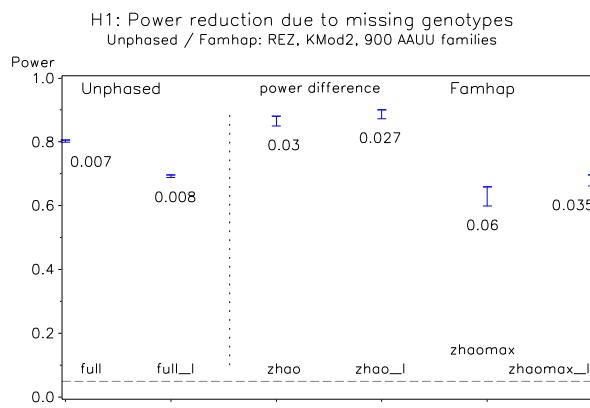
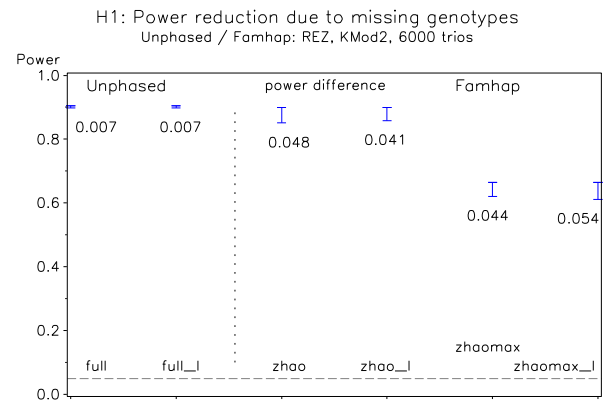
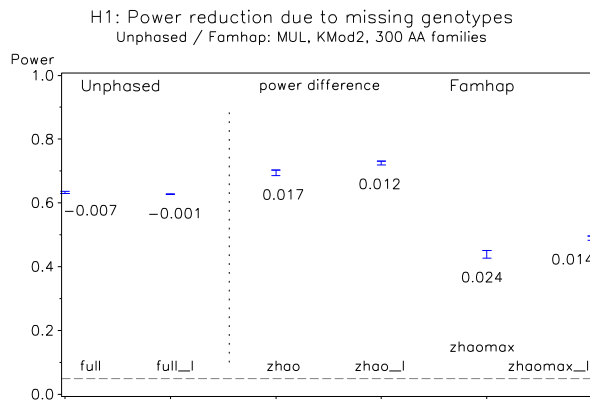


# Famhap und Unphased

## KMod 1

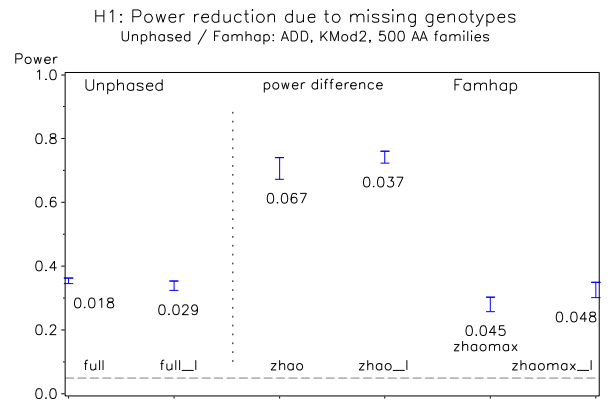
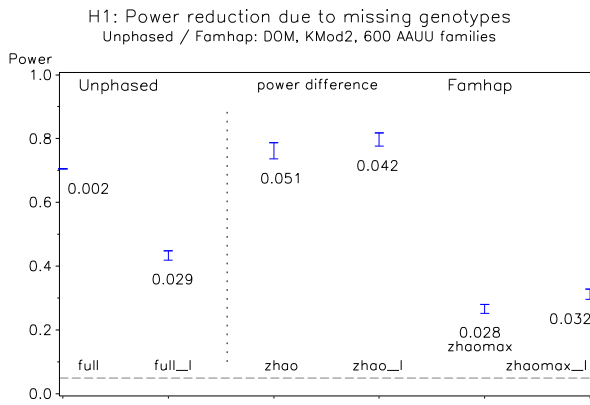
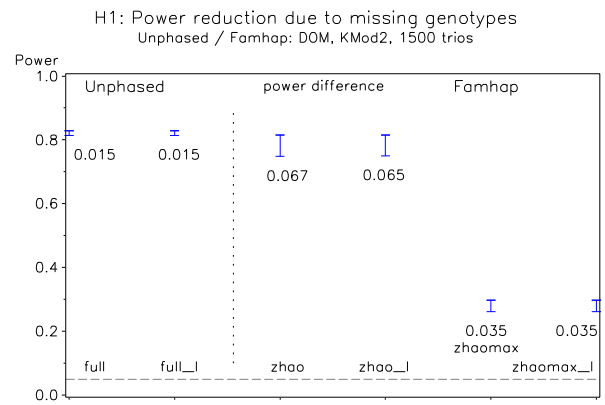
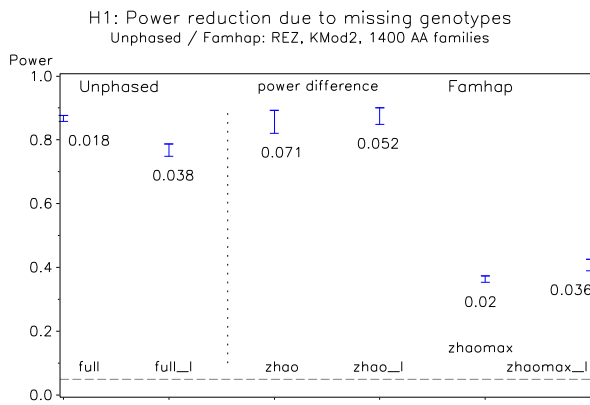
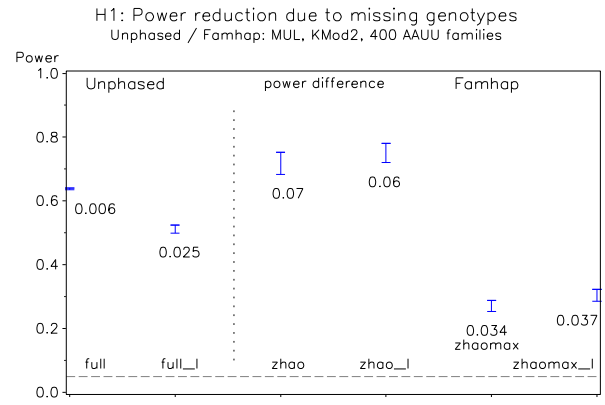
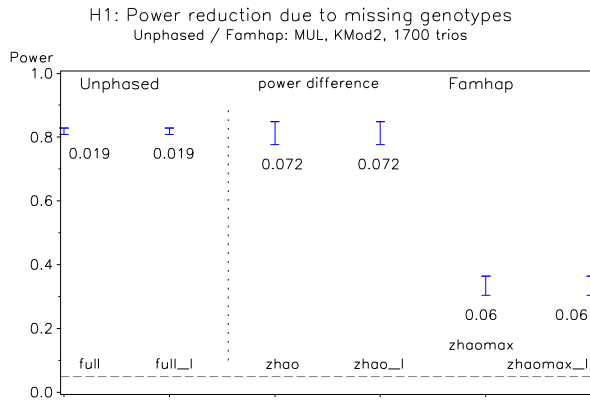


## KMod 2





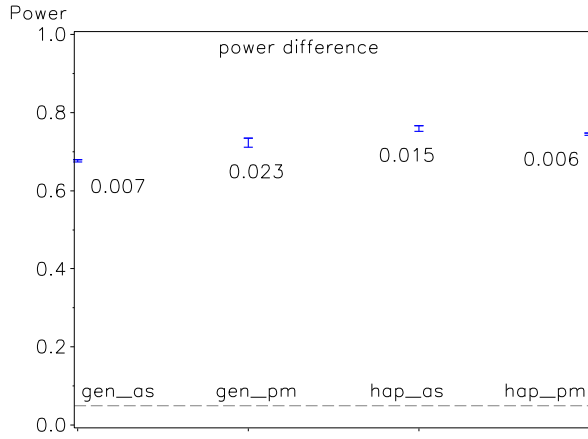
### KMod3



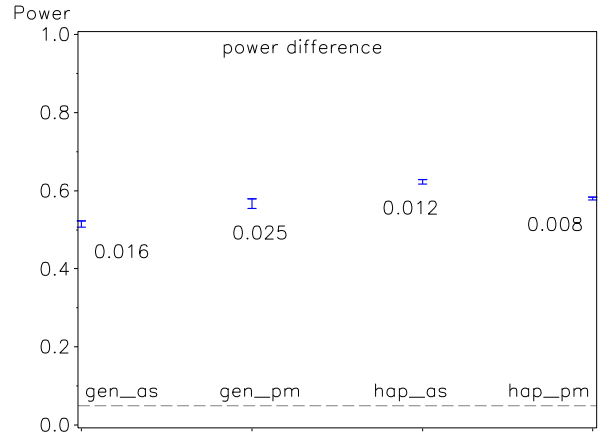
# Hotell\_Fam

## Nur KMod 1

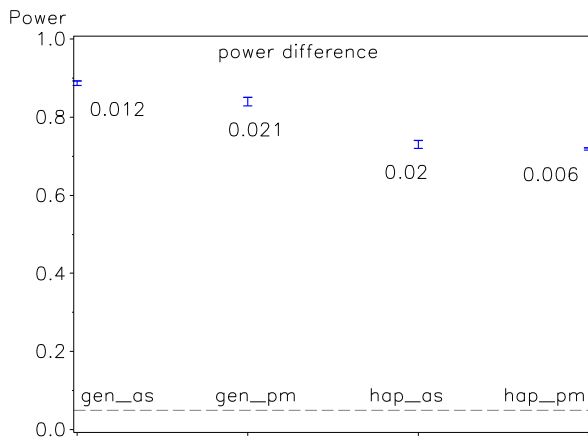
H1: Power reduction due to missing genotypes  
Hotell\_Fam: MUL KMod1, 700 trios



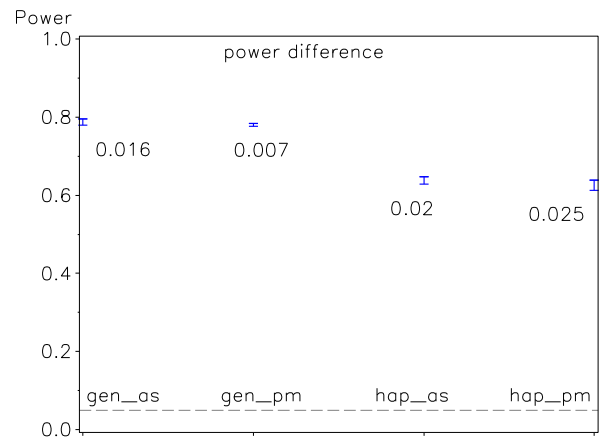
H1: Power reduction due to missing genotypes  
Hotell\_Fam: MUL KMod1, 200 AAU families



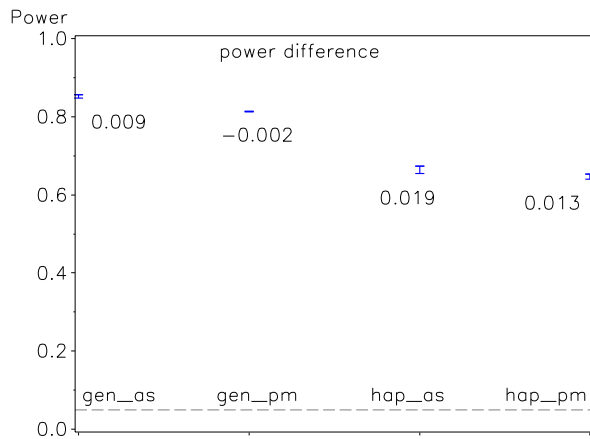
H1: Power reduction due to missing genotypes  
Hotell\_Fam: REZ KMod1, 500 AA families



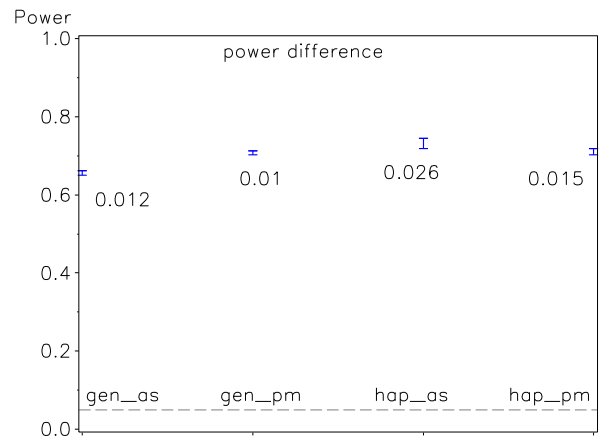
H1: Power reduction due to missing genotypes  
Hotell\_Fam: DOM KMod1, 400 trios



H1: Power reduction due to missing genotypes  
Hotell\_Fam: DOM KMod1, 300 AAU families



H1: Power reduction due to missing genotypes  
Hotell\_Fam: ADD KMod1, 300 AA families



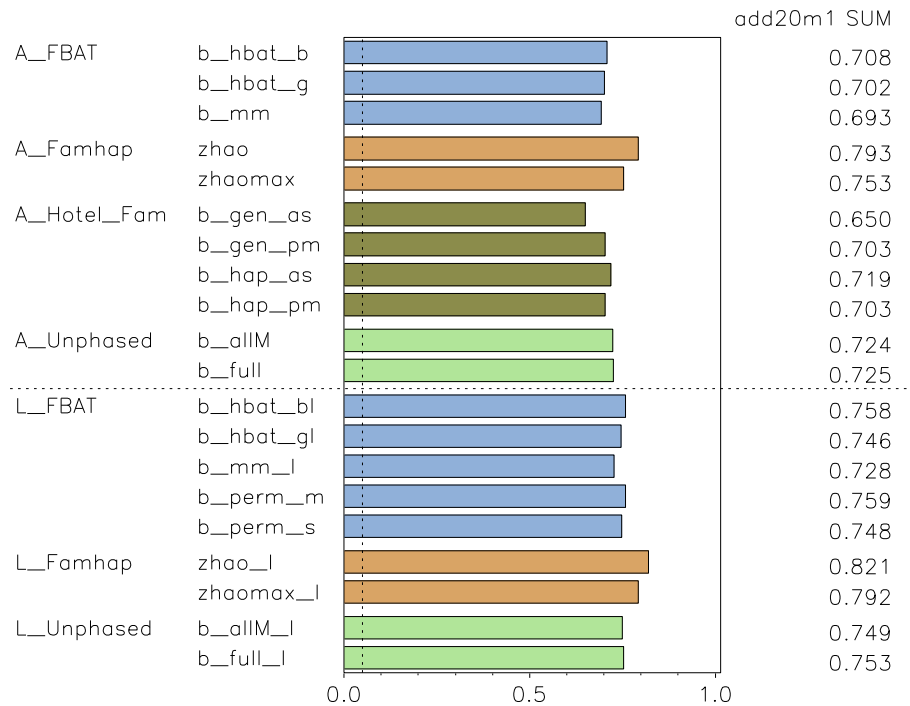
# Analyse mit fehlenden Genotypen: Vergleich der Overall-Power

$\alpha = 0.05$

## KMod 1

Analysis with missing genotypes: alpha=0.05

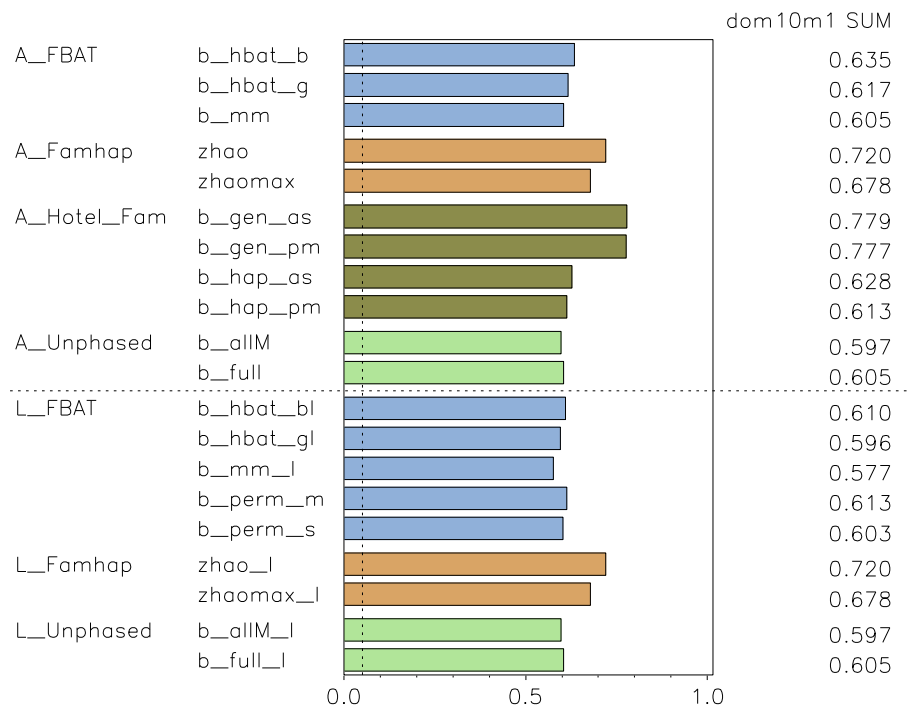
Overall power: ADD, Kmod 1, 300 AA-families



Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

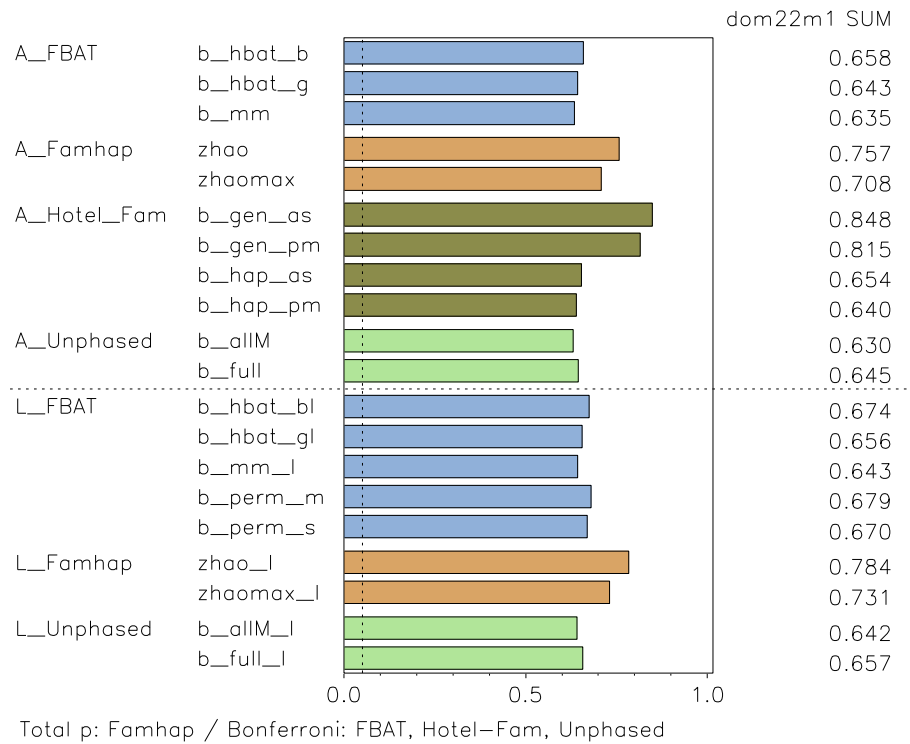
Analysis with missing genotypes: alpha=0.05

Overall power: DOM, Kmod 1, 400 trios

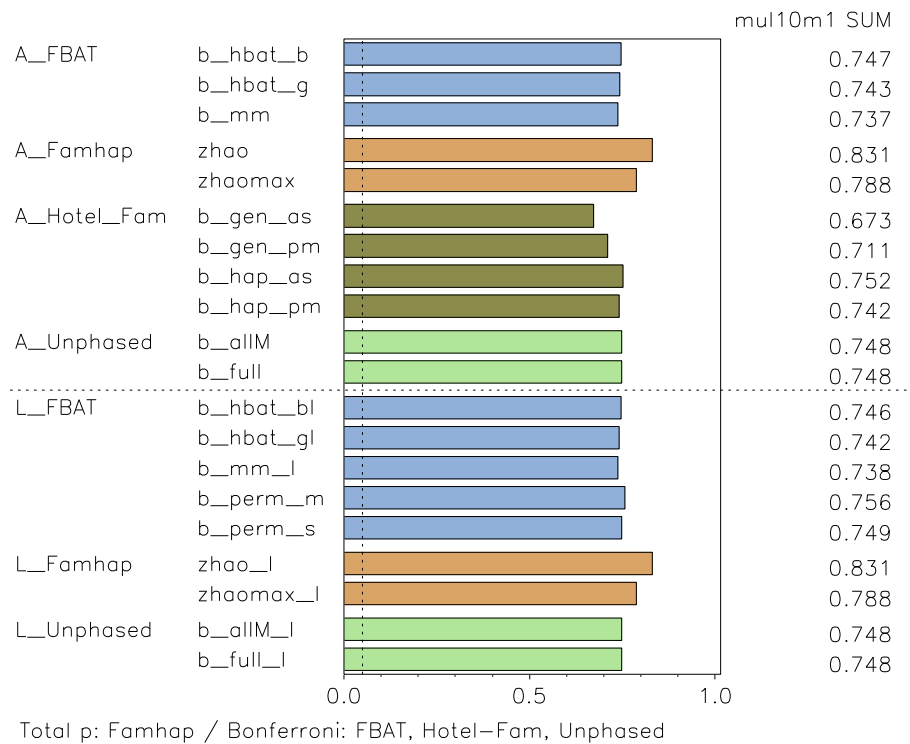


Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

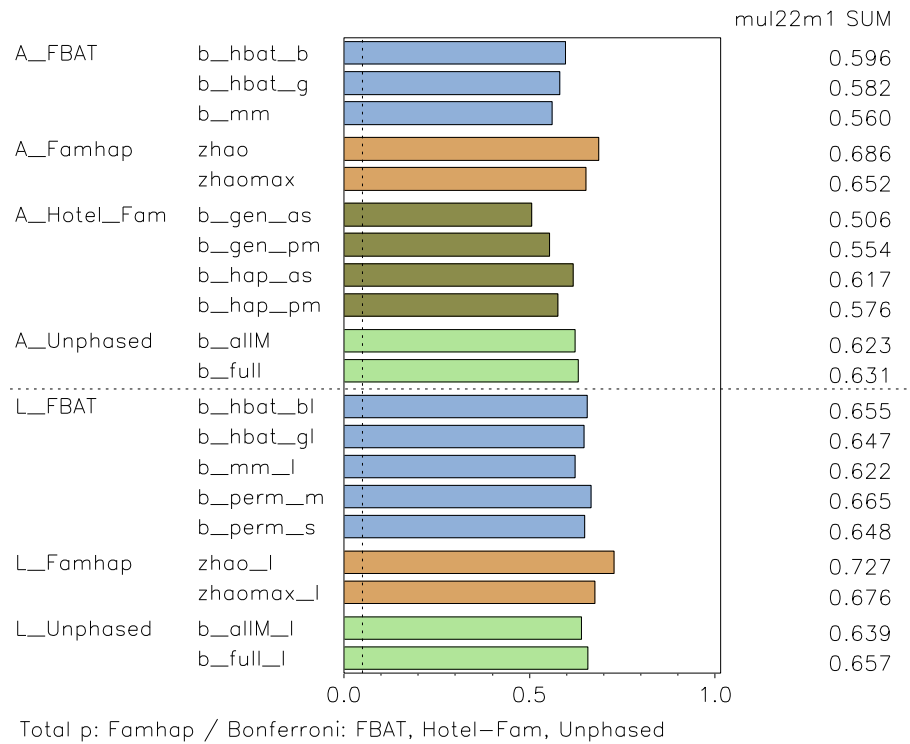
Analysis with missing genotypes: alpha=0.05  
 Overall power: DOM, Kmod 1, 300 AAU-families



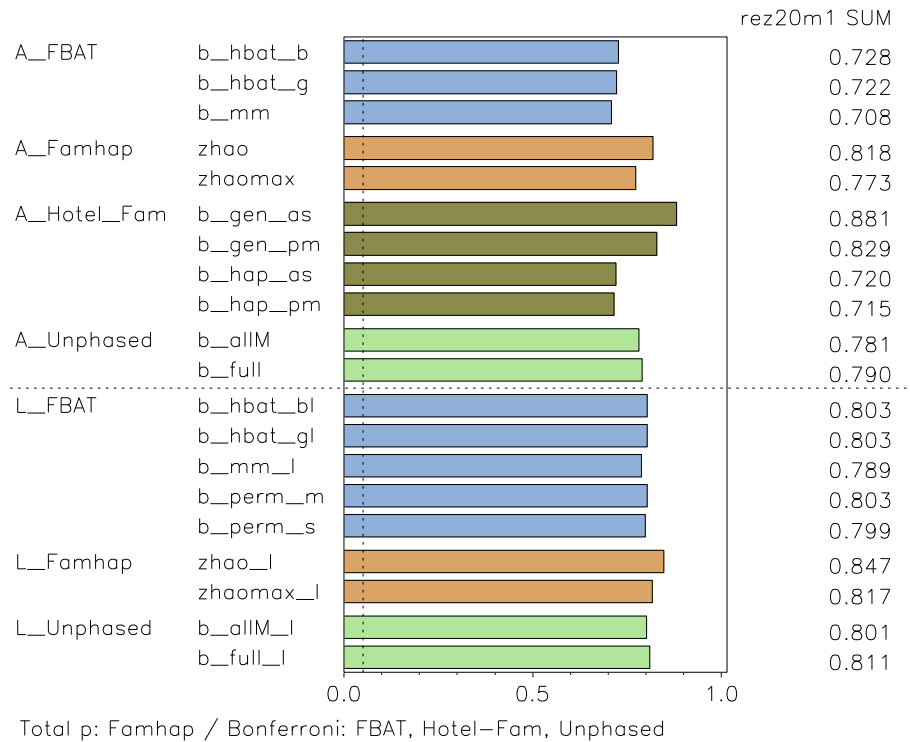
Analysis with missing genotypes: alpha=0.05  
 Overall power: MUL, Kmod 1, 700 trios



Analysis with missing genotypes: alpha=0.05  
 Overall power: MUL, Kmod 1, 200 AAU-families



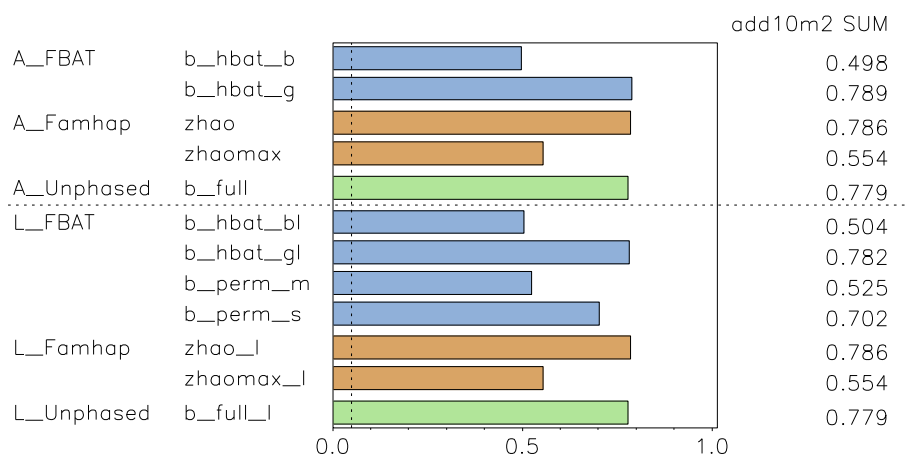
Analysis with missing genotypes: alpha=0.05  
 Overall power: REZ, Kmod 1, 500 AA-families



## KMod 2

Analysis with missing genotypes:  $\alpha=0.05$

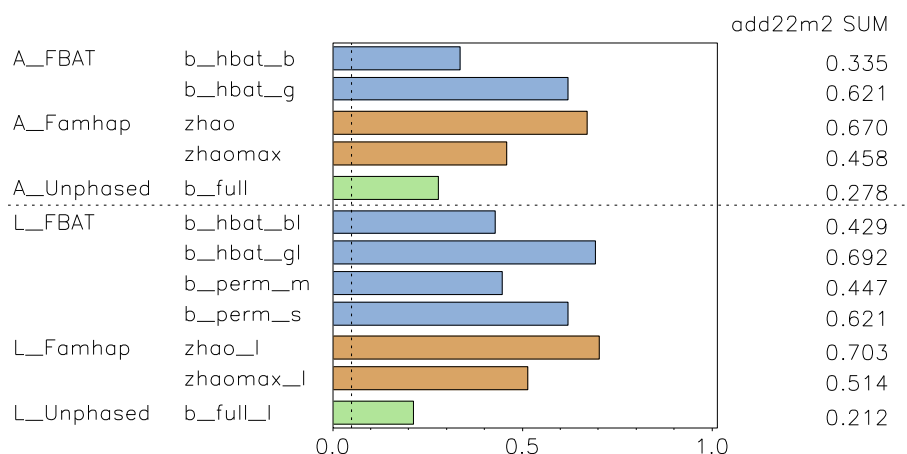
Overall power: ADD, Kmod 2, 1800 trios



Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Analysis with missing genotypes:  $\alpha=0.05$

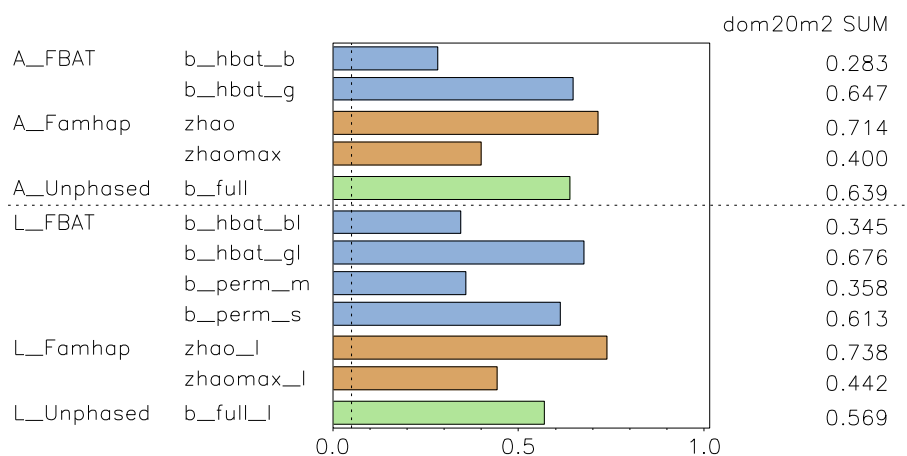
Overall power: ADD, Kmod 2, 400 AAU-families



Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Analysis with missing genotypes:  $\alpha=0.05$

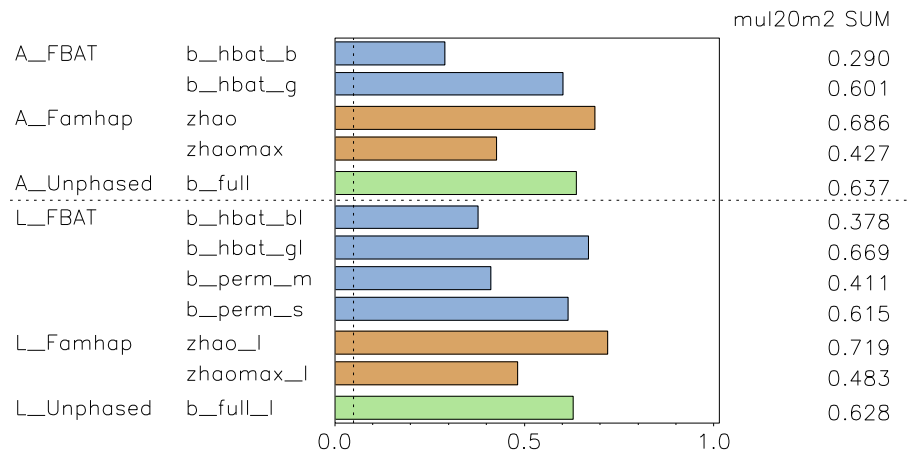
Overall power: DOM, Kmod 2, 500 AA-families



Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Analysis with missing genotypes: alpha=0.05

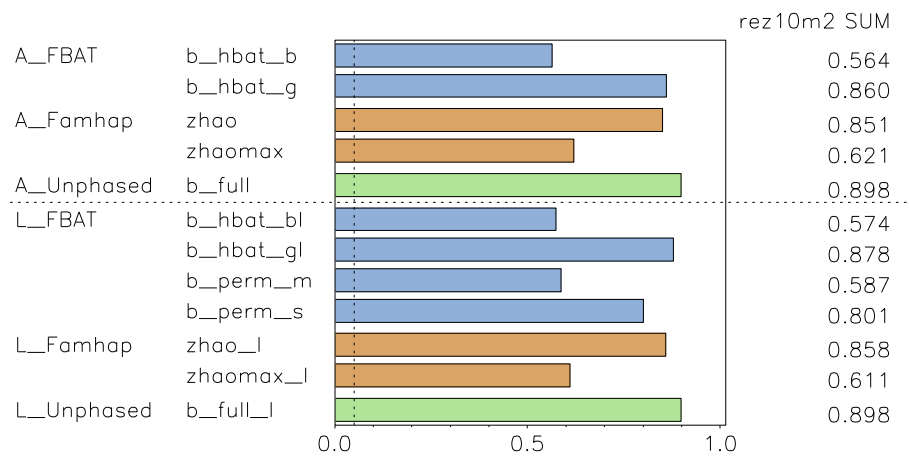
Overall power: MUL, Kmod 2, 300 AA-families



Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Analysis with missing genotypes: alpha=0.05

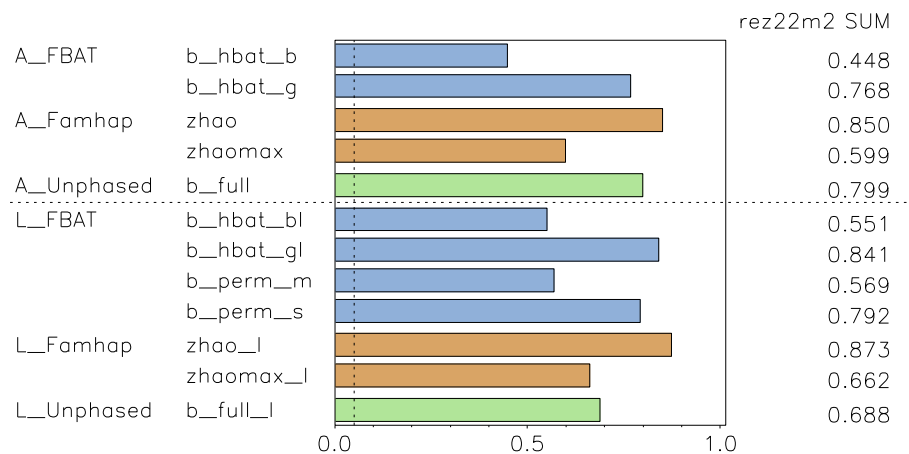
Overall power: REZ, Kmod 2, 6000 trios



Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Analysis with missing genotypes: alpha=0.05

Overall power: REZ, Kmod 2, 900 AAU-families

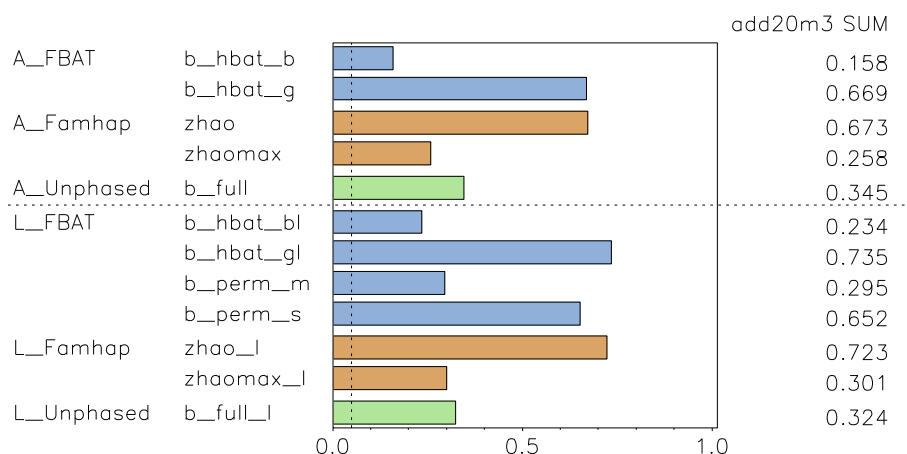


Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

### KMod 3

Analysis with missing genotypes: alpha=0.05

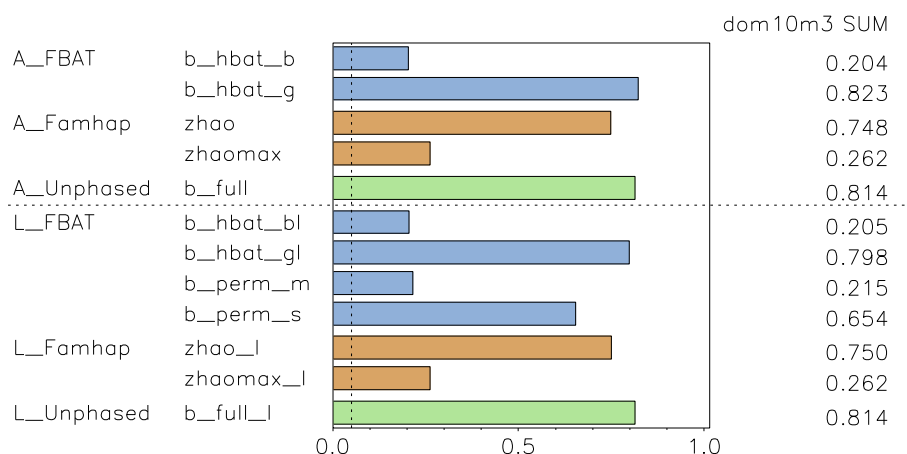
Overall power: ADD, Kmod 3, 500 AA-families



Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Analysis with missing genotypes: alpha=0.05

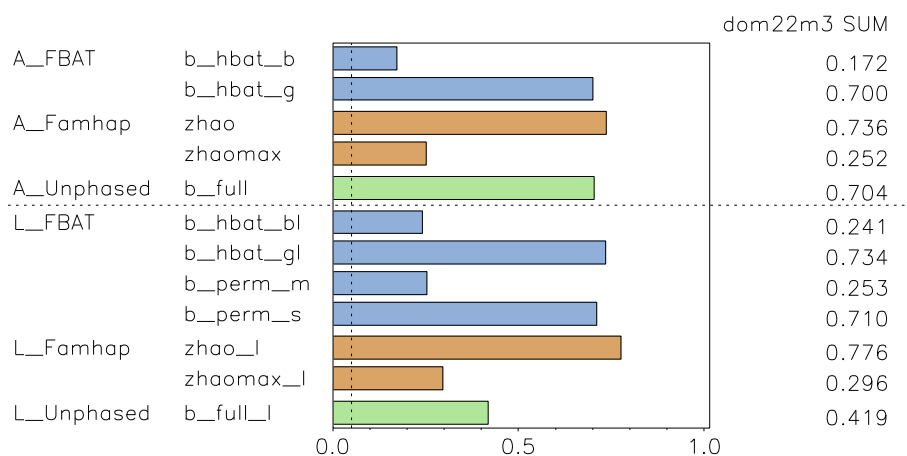
Overall power: DOM, Kmod 3, 1500 trios



Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Analysis with missing genotypes: alpha=0.05

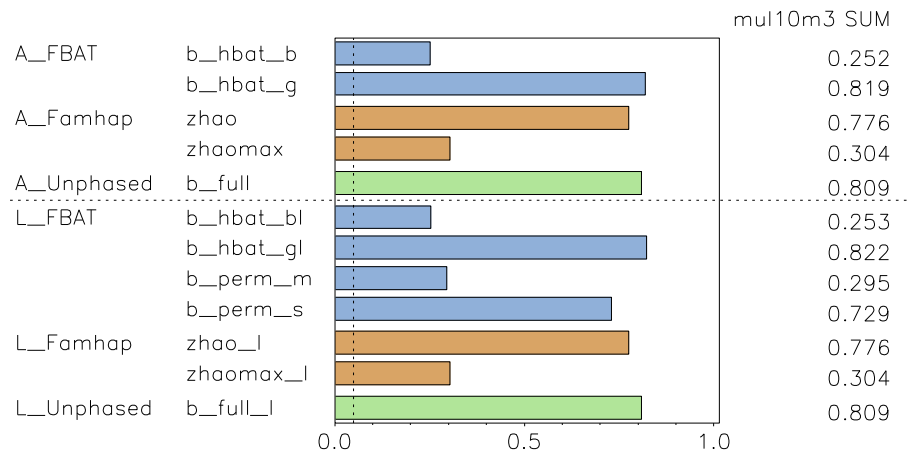
Overall power: DOM, Kmod 3, 600 AAUU-families



Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

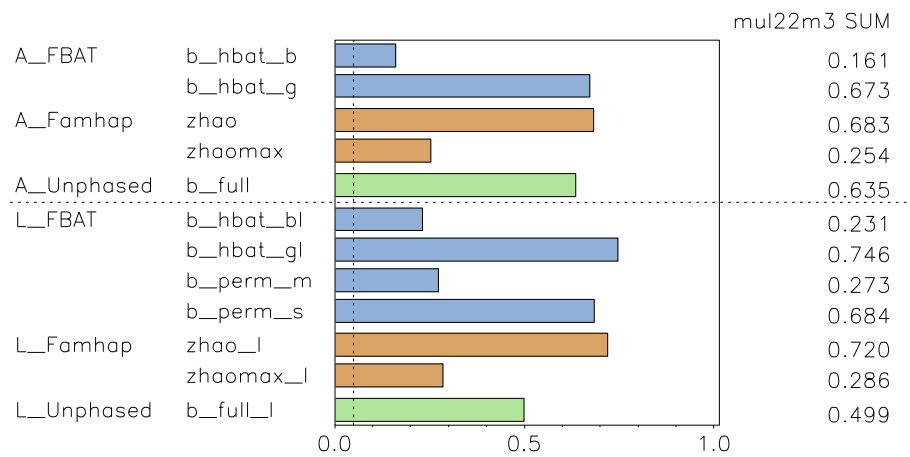


Analysis with missing genotypes: alpha=0.05  
 Overall power: MUL, Kmod 3, 1700 trios



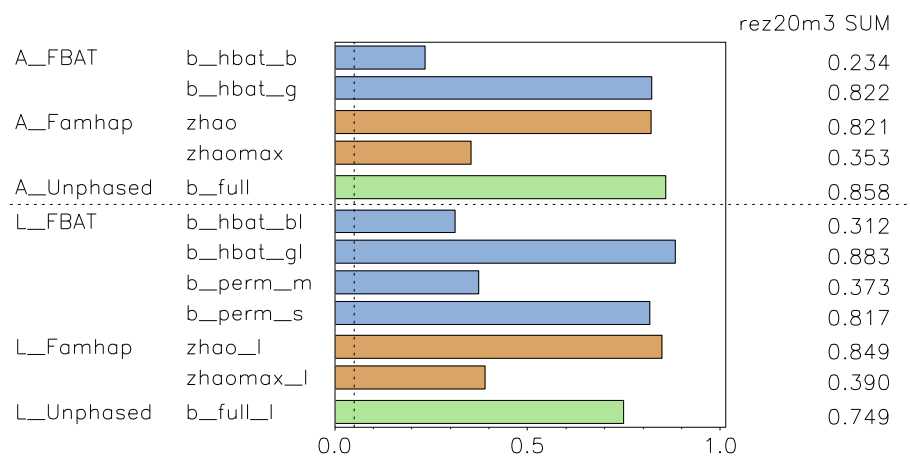
Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Analysis with missing genotypes: alpha=0.05  
 Overall power: MUL, Kmod 3, 400 AAU-families



Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

Analysis with missing genotypes: alpha=0.05  
 Overall power: REZ, Kmod 3, 1400 AA-families

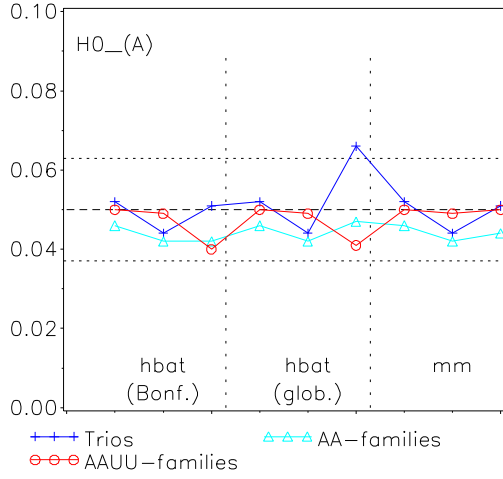


Total p: Famhap / Bonferroni: FBAT, Hotel-Fam, Unphased

# Stratifikation: Fehler 1. Art

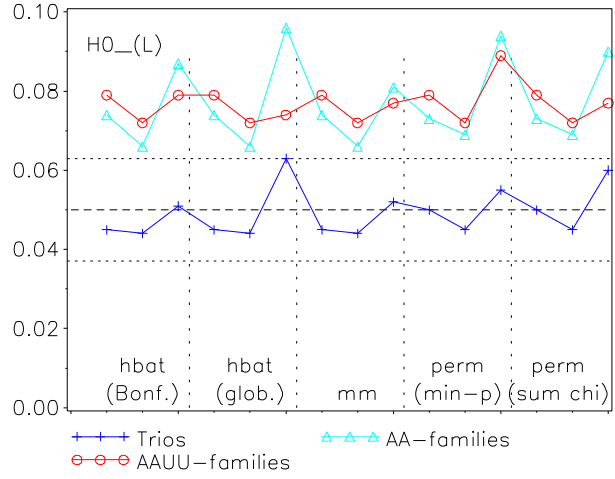
## FBAT, Tests auf $H_0^A$ ,

FBAT: Type I error, mixed population  
Moderate model, 1000 families

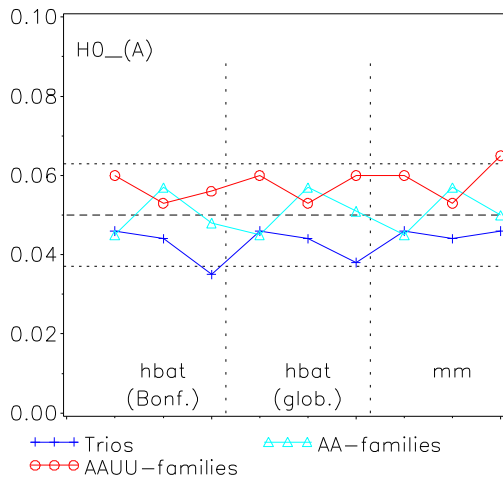


## Tests auf $H_0^L$

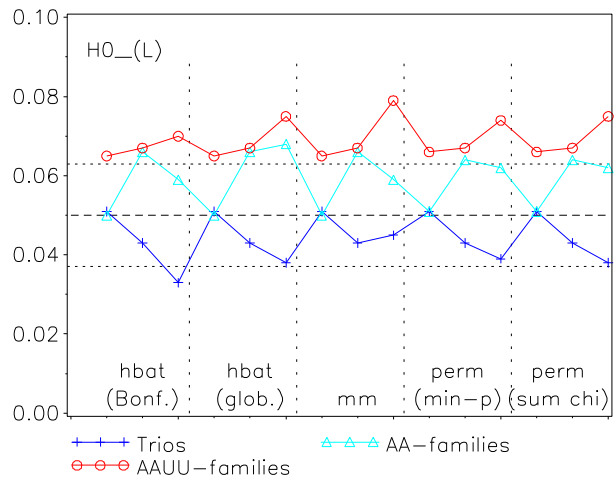
FBAT: Type I error, mixed population  
Moderate model, 1000 families



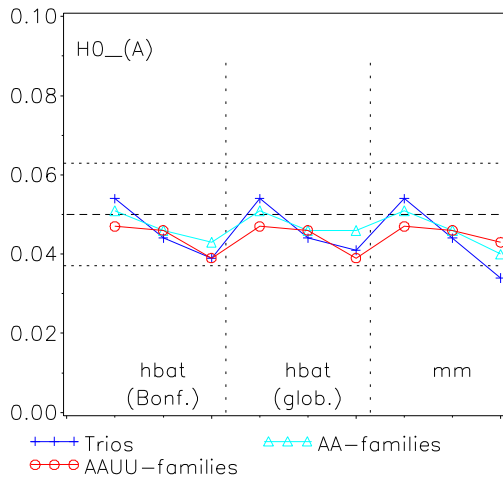
FBAT: Type I error, mixed population  
Extreme model, 500 families



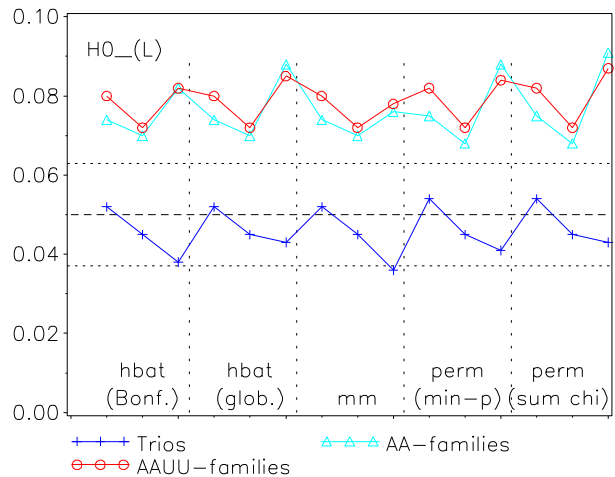
FBAT: Type I error, mixed population  
Extreme model, 500 families



FBAT: Type I error, mixed population  
Moderate model, missing genotypes, 1000 families

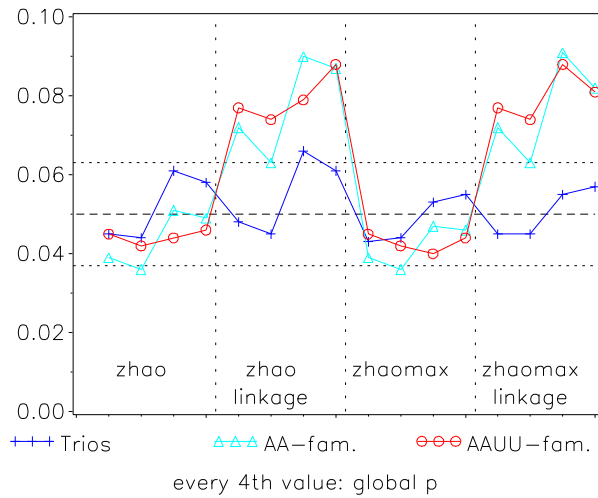


FBAT: Type I error, mixed population  
Moderate model, missing genotypes, 1000 families

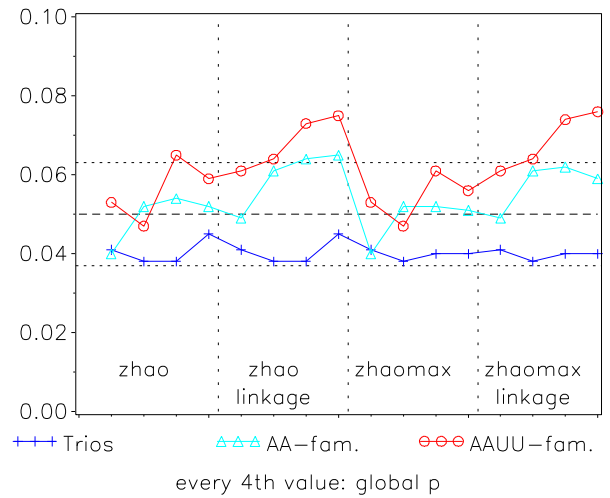


## Famhap: jeder 4. Wert: totaler p-Wert über alle 3 Markerkombinationen

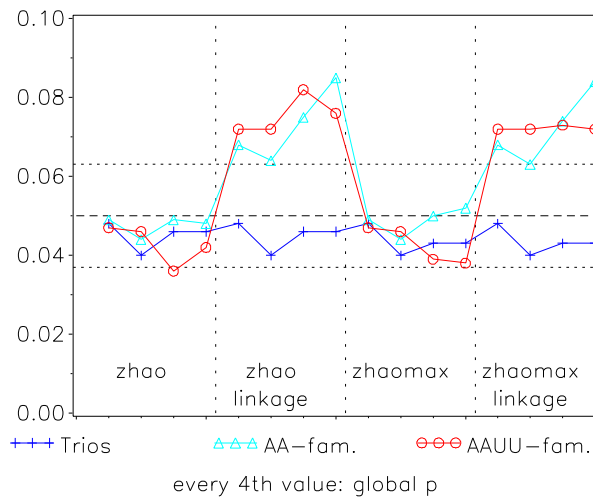
Famhap: Type I error, mixed population, moderate model, 1000 families



Famhap: Type I error Mixed population, extreme model, 500 families

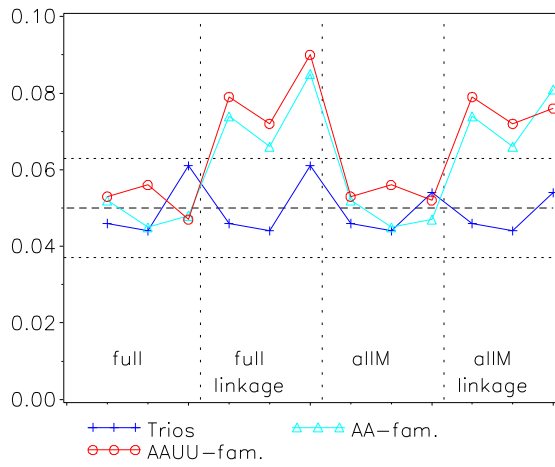


Famhap: Type I error, mixed population, moderate model, 1000 families, missing genotypes

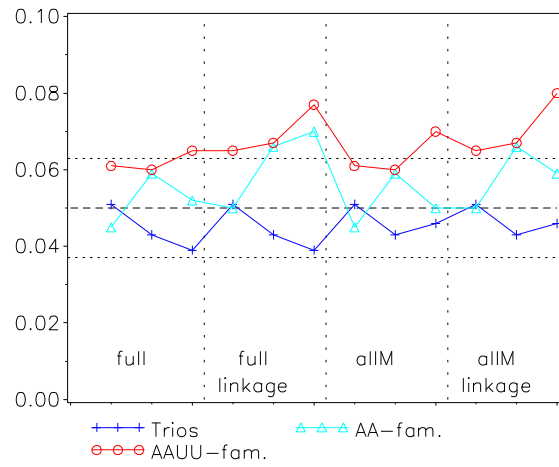


## Unphased

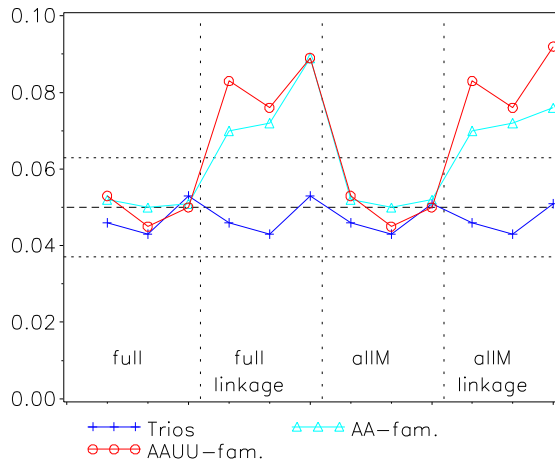
Unphased: Type 1 error, mixed population  
Moderate model, 1000 families



Unphased: Type 1 error, mixed population  
Extreme model, 500 families

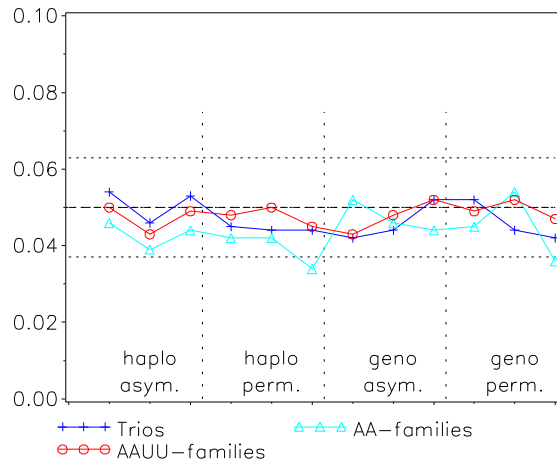


Unphased: Type 1 error, mixed population  
Moderate model, 1000 families, missing genotypes

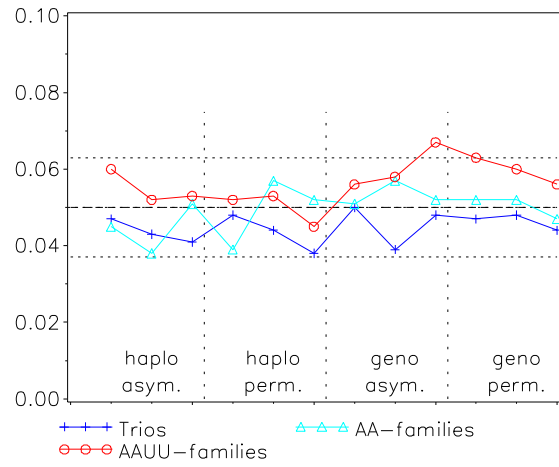


## Hotel\_Fam

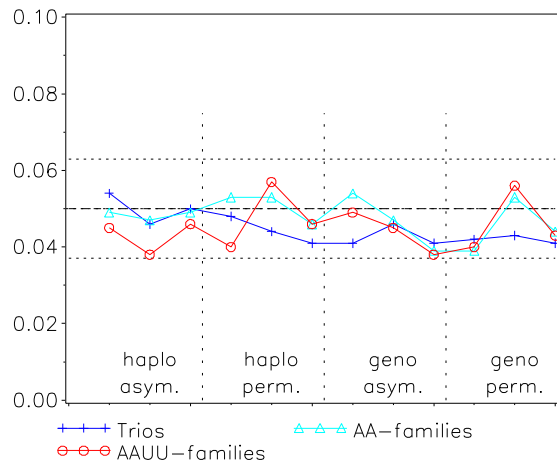
Type I error: Hotel\_Fam, stratified population  
Moderate model, 1000 families



Type I error: Hotel\_Fam, stratified population  
Extreme model, 500 families

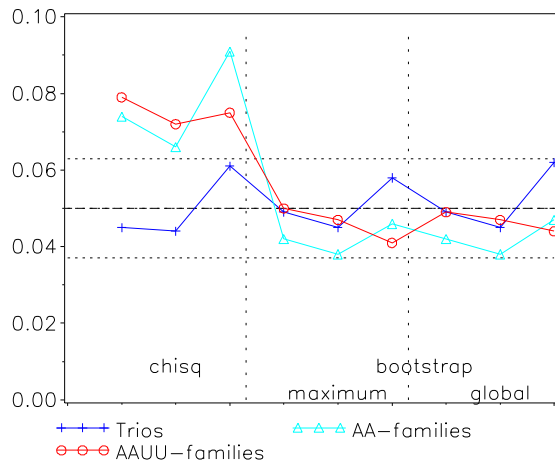


Type I error: Hotel\_Fam, stratified population  
Moderate model, missing genotypes, 1000 families

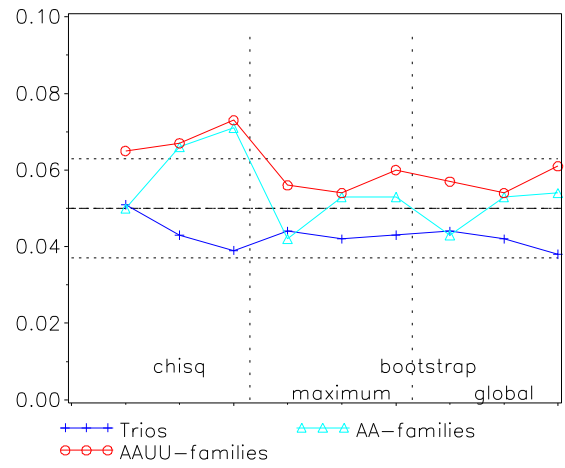


## Transmit

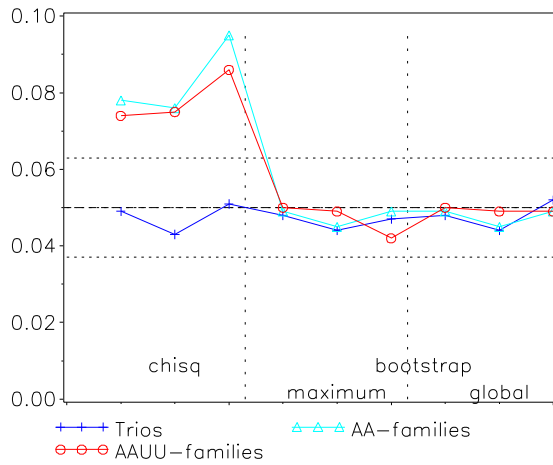
Transmit: Type I error, mixed population, moderate model, 1000 families



Transmit: Type I error, mixed population, extreme model, 1000 families



Transmit: Type I error, mixed population, moderate model, 1000 families, missing genotypes

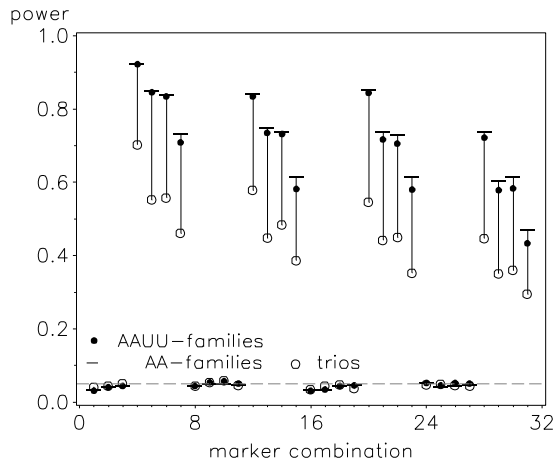


# Power für verschiedene Familienstrukturen

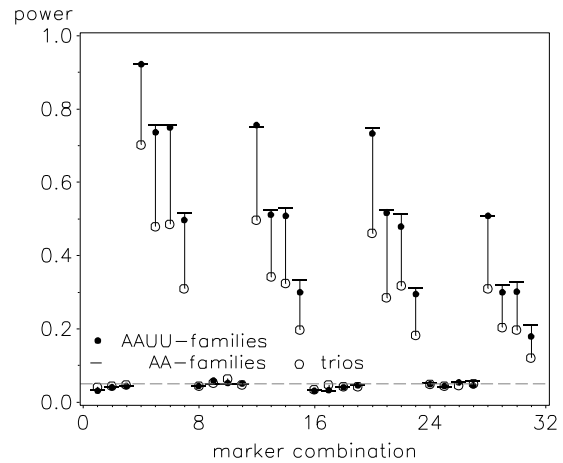
## Famhap

### Additives Modell KMod 1

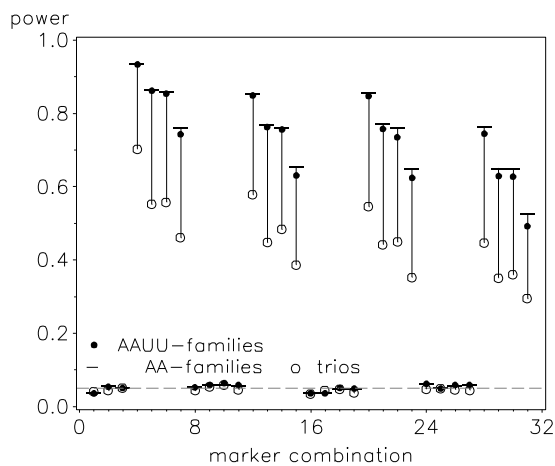
Power reduction F22–F20–F10–families  
H1, Famhap zhao: ADD KMod 1, 300 families



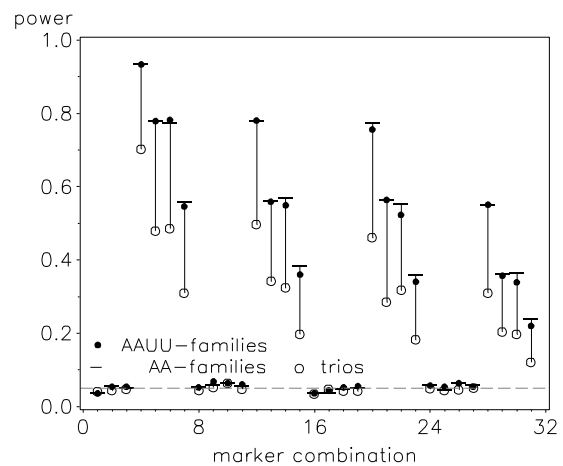
Power reduction F22–F20–F10–families  
H1, Famhap zhaomax: ADD KMod 1, 300 families



Power reduction F22–F20–F10–families  
H1, Famhap zhao\_I: ADD KMod 1, 300 families

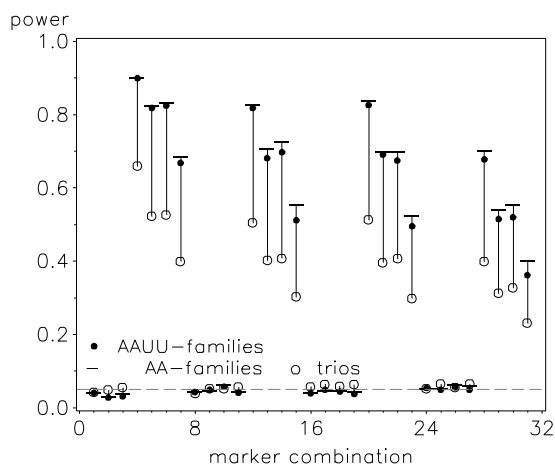


Power reduction F22–F20–F10–families  
H1, Famhap zhaomax\_I: ADD KMod 1, 300 families

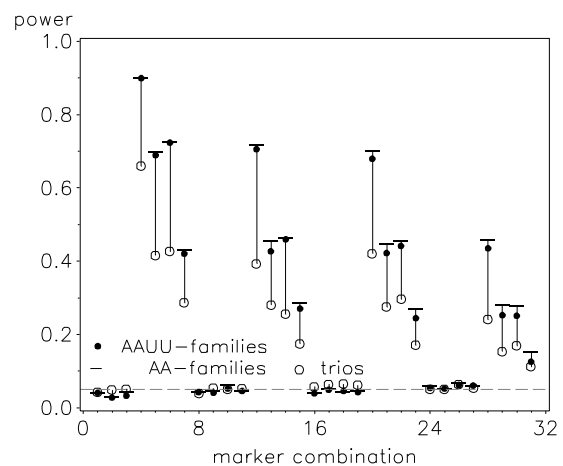


### Multiplikatives Modell KMod 1

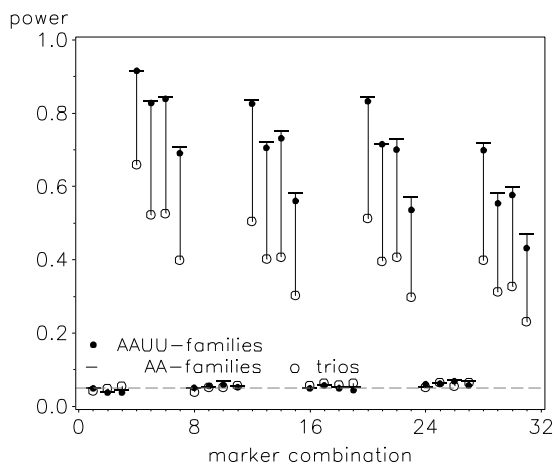
Power reduction F22–F20–F10–families  
H1, Famhap zhao: MUL KMod 1, 200 families



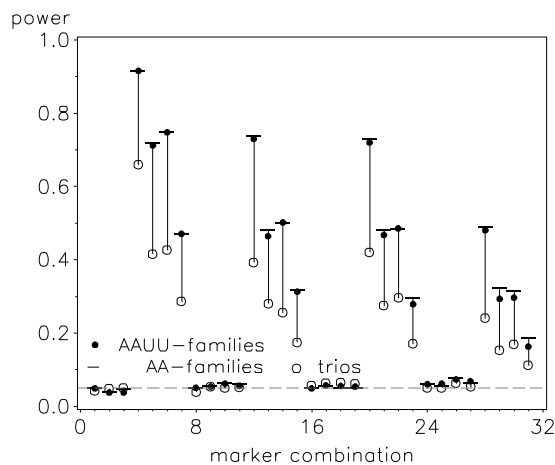
Power reduction F22–F20–F10–families  
H1, Famhap zhaomax: MUL KMod 1, 200 families



Power reduction F22-F20-F10-families  
H1, Famhap zhao\_l: MUL KMod 1, 200 families

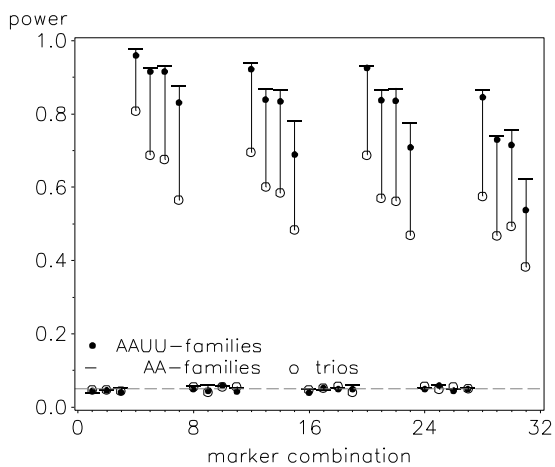


Power reduction F22-F20-F10-families  
H1, Famhap zhaomax\_l: MUL KMod 1, 200 families

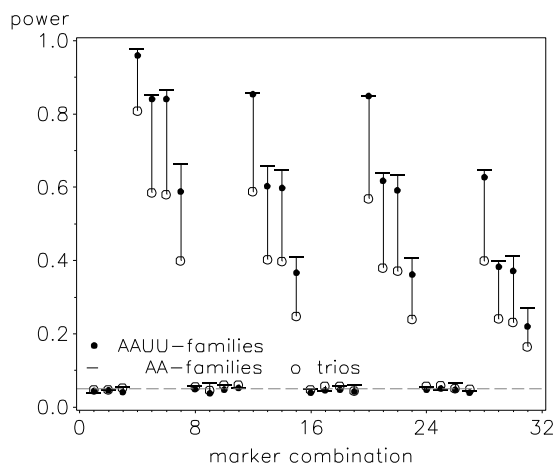


### Rezessives Modell KMod 1

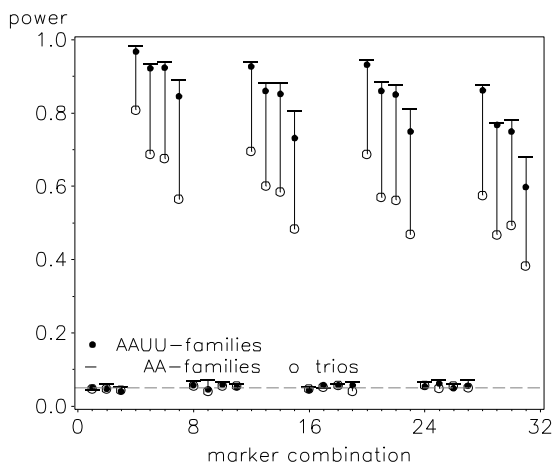
Power reduction F22-F20-F10-families  
H1, Famhap zhao: REZ KMod 1, 600 families



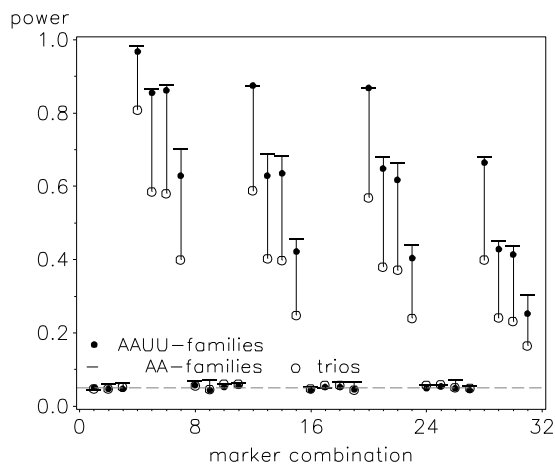
Power reduction F22-F20-F10-families  
H1, Famhap zhaomax: REZ KMod 1, 600 families



Power reduction F22-F20-F10-families  
H1, Famhap zhao\_l: REZ KMod 1, 600 families



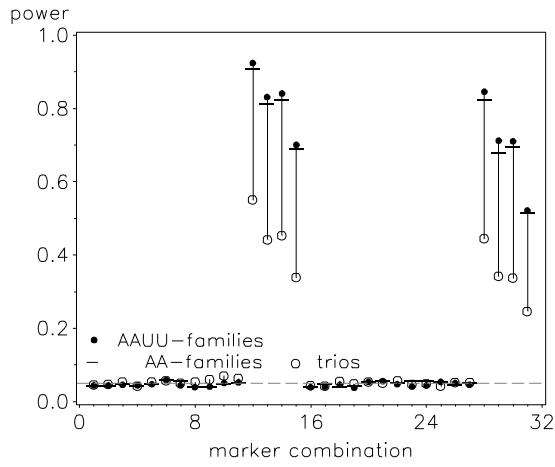
Power reduction F22-F20-F10-families  
H1, Famhap zhaomax\_l: REZ KMod 1, 600 families



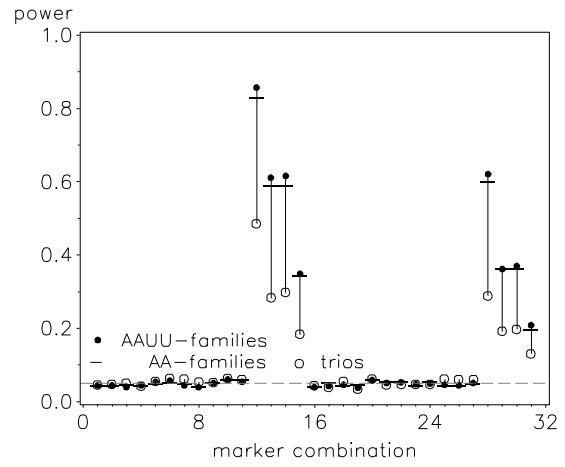


### Additives Modell KMod 2

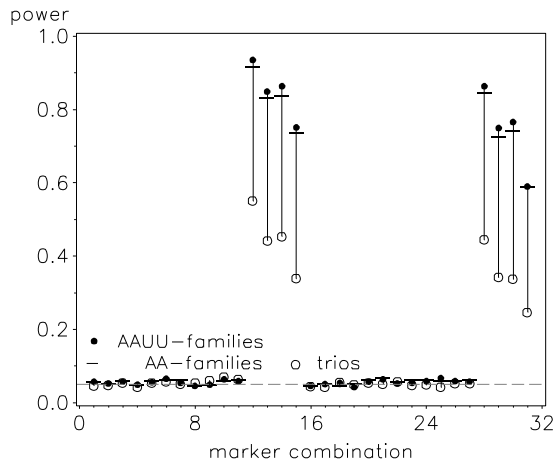
Power reduction F22-F20-F10-families  
H1, Famhap zhao: ADD KMod 2, 400 families



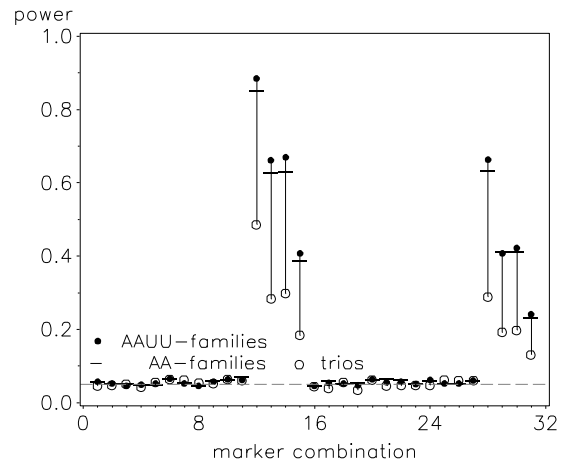
Power reduction F22-F20-F10-families  
H1, Famhap zhaomax: ADD KMod 2, 400 families



Power reduction F22-F20-F10-families  
H1, Famhap zhao\_I: ADD KMod 2, 400 families

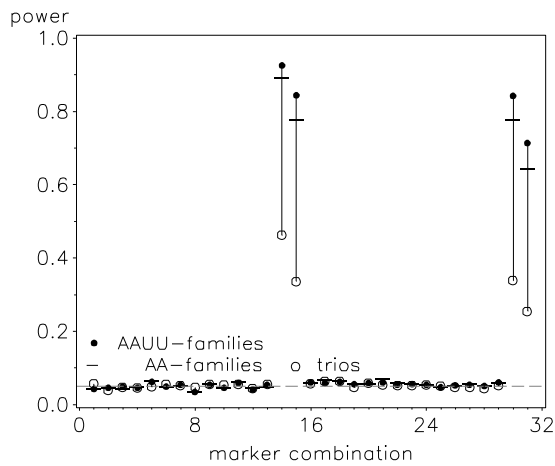


Power reduction F22-F20-F10-families  
H1, Famhap zhaomax\_I: ADD KMod 2, 400 families

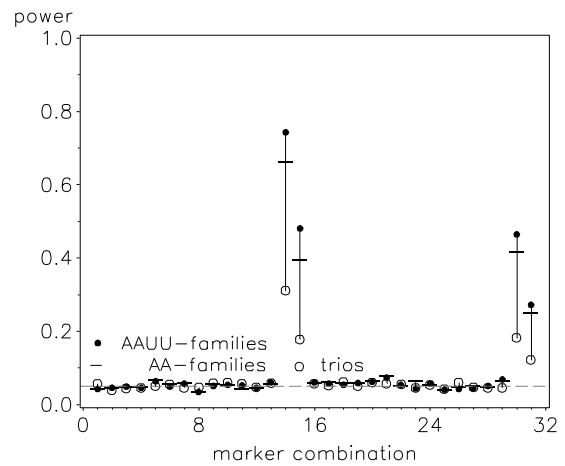


### Multiplikatives Modell KMod 3

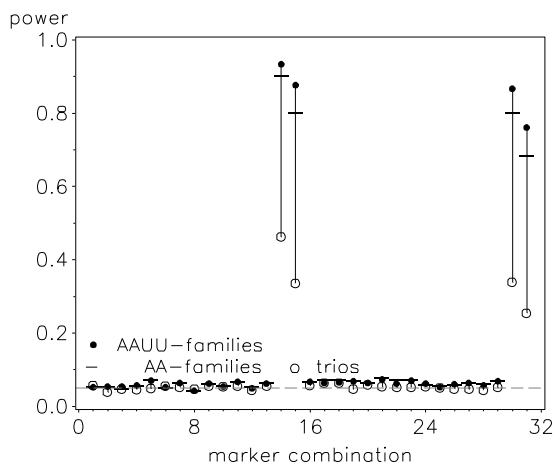
Power reduction F22-F20-F10-families  
H1, Famhap zhao: MUL KMod 3, 400 families



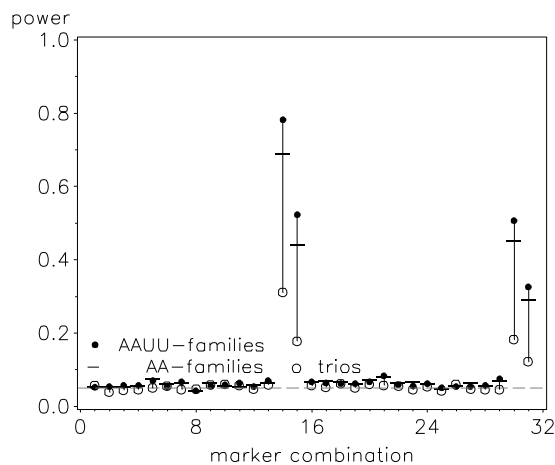
Power reduction F22-F20-F10-families  
H1, Famhap zhaomax: MUL KMod 3, 400 families



Power reduction F22–F20–F10–families  
H1, Famhap zhao\_l: MUL KMod 3, 400 families

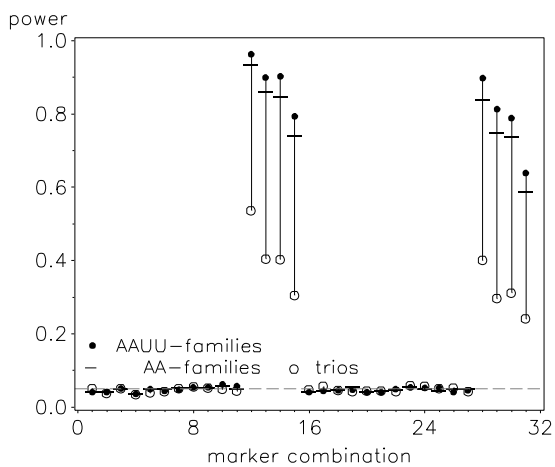


Power reduction F22–F20–F10–families  
H1, Famhap zhaomax\_l: MUL KMod 3, 400 families

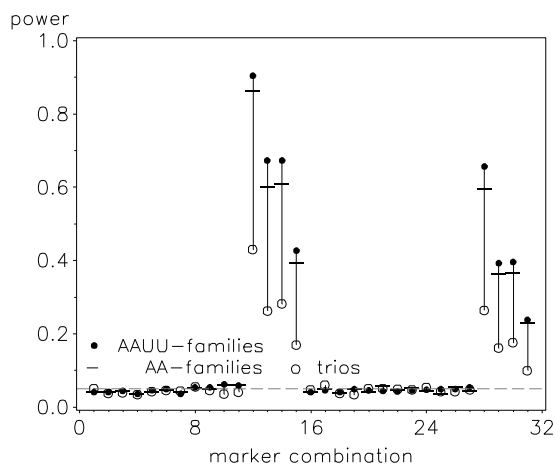


### Dominantes Modell KMod 2

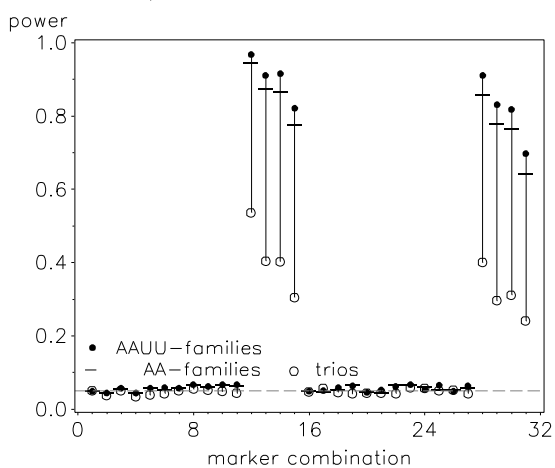
Power reduction F22–F20–F10–families  
H1, Famhap zhao: DOM KMod 2, 500 families



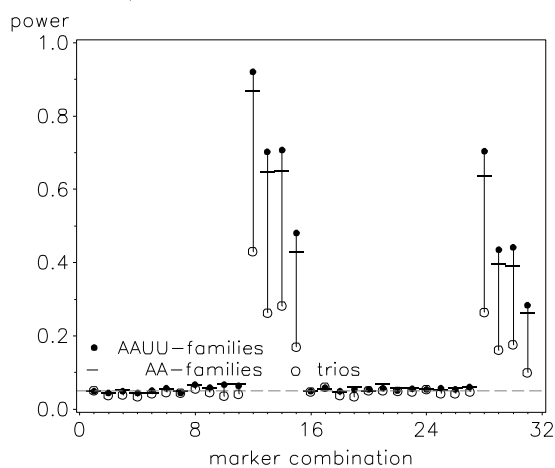
Power reduction F22–F20–F10–families  
H1, Famhap zhaomax: DOM KMod 2, 500 families



Power reduction F22–F20–F10–families  
H1, Famhap zhao\_l: DOM KMod 2, 500 families



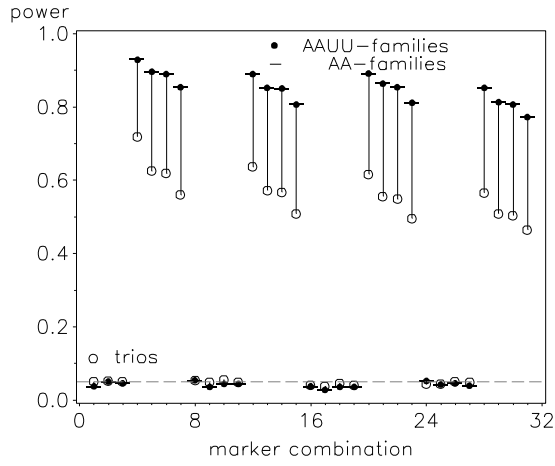
Power reduction F22–F20–F10–families  
H1, Famhap zhaomax\_l: DOM KMod 2, 500 families



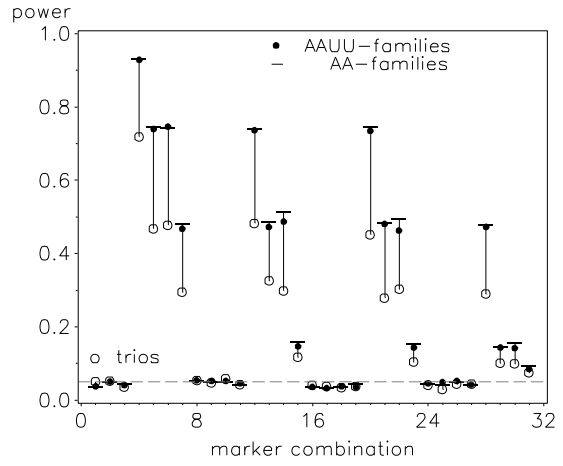
# FBAT

## Additives Modell KMod 1

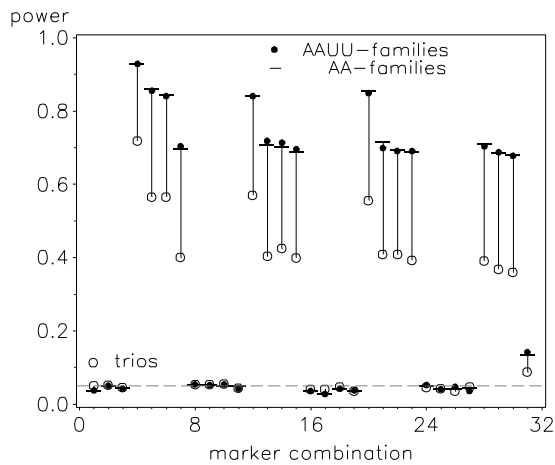
Power reduction F22-F20-F10-families  
H1, FBAT mm: ADD KMod 1, 300 families



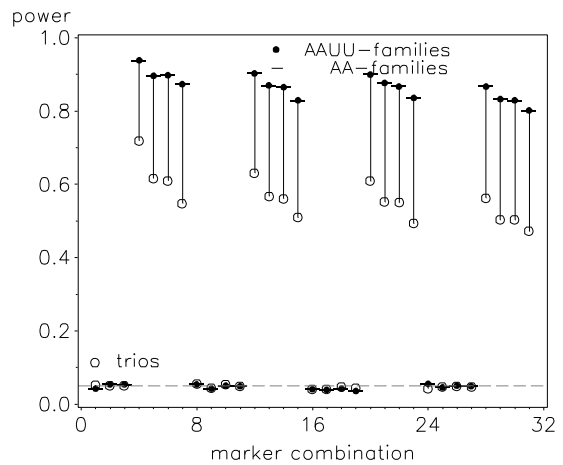
Power reduction F22-F20-F10-families  
H1, FBAT hbat\_b: ADD KMod 1, 300 families



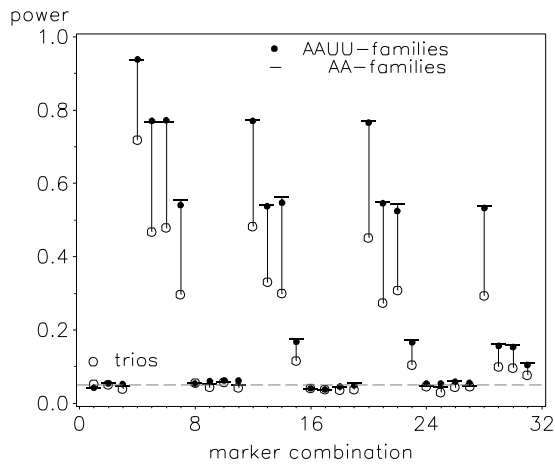
Power reduction F22-F20-F10-families  
H1, FBAT hbat\_g: ADD KMod 1, 300 families



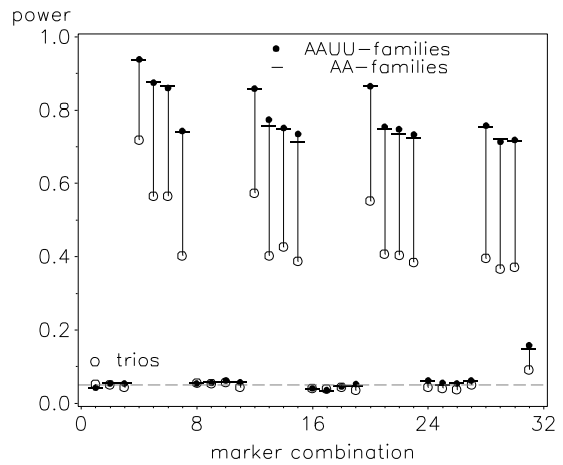
Power reduction F22-F20-F10-families  
H1, FBAT mm\_l: ADD KMod 1, 300 families



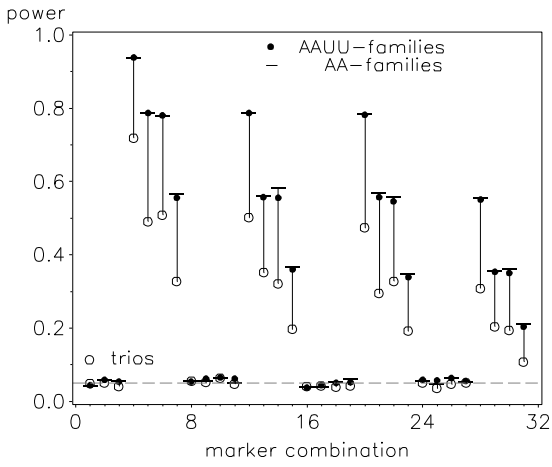
Power reduction F22-F20-F10-families  
H1, FBAT hbat\_bl: ADD KMod 1, 300 families



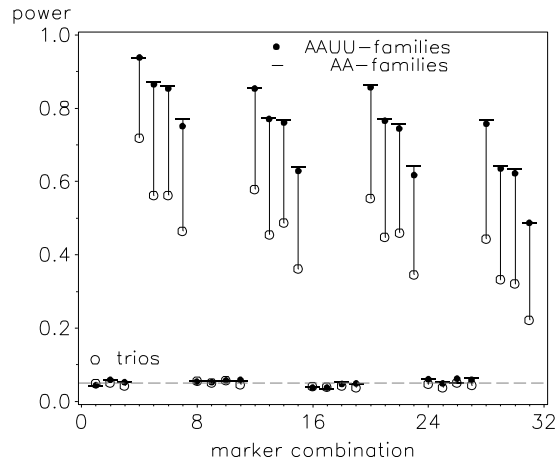
Power reduction F22-F20-F10-families  
H1, FBAT hbat\_gl: ADD KMod 1, 300 families



Power reduction F22–F20–F10–families  
H1, FBAT perm\_m: ADD KMod 1, 300 families

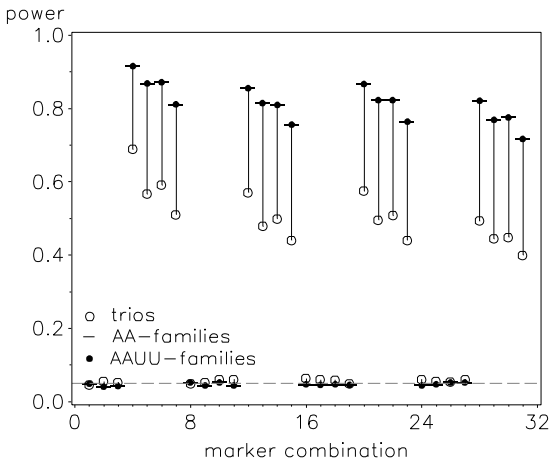


Power reduction F22–F20–F10–families  
H1, FBAT perm\_s: ADD KMod 1, 300 families

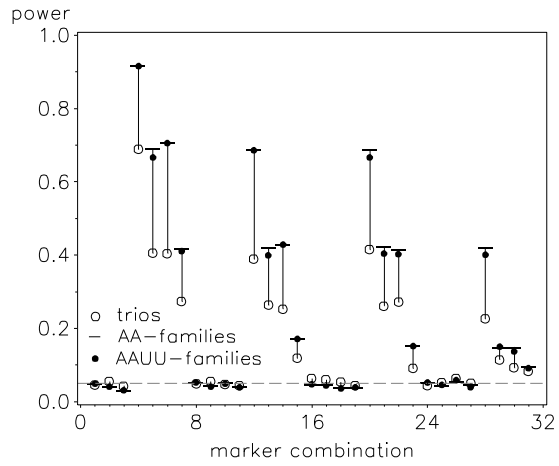


**Multiplikatives Modell KMod 1**

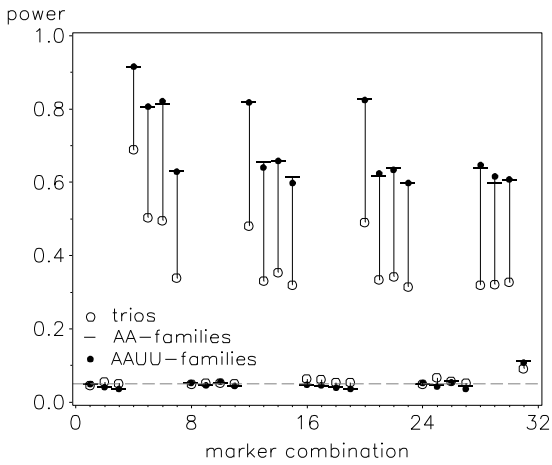
Power reduction F22–F20–F10–families  
H1, FBAT mm: MUL KMod 1, 200 families



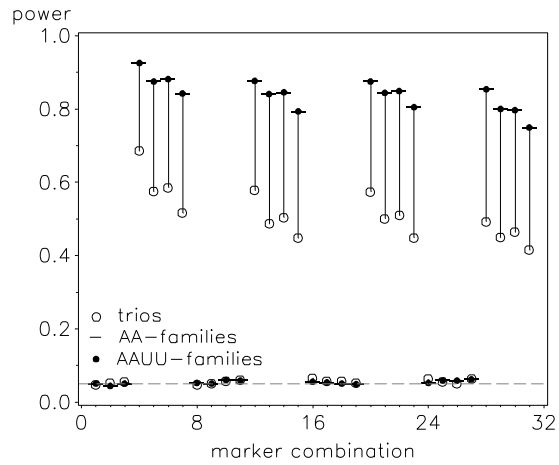
Power reduction F22–F20–F10–families  
H1, FBAT hbat\_b: MUL KMod 1, 200 families



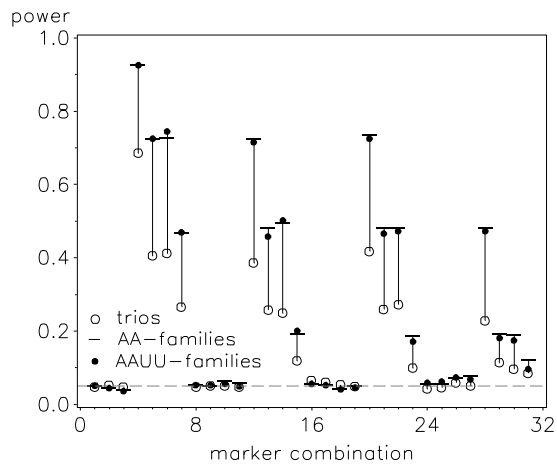
Power reduction F22–F20–F10–families  
H1, FBAT hbat\_g: MUL KMod 1, 200 families



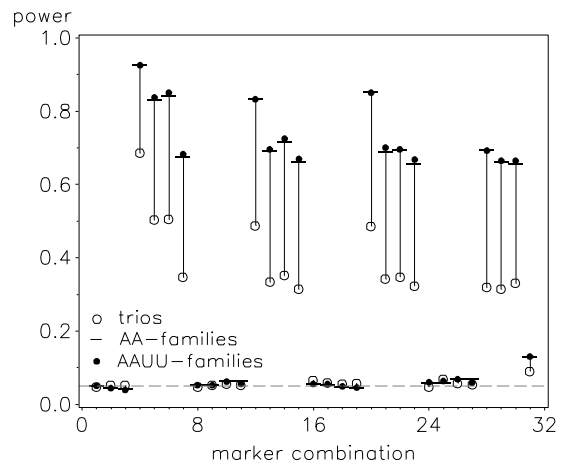
Power reduction F22–F20–F10–families  
H1, FBAT mm\_l: MUL KMod 1, 200 families



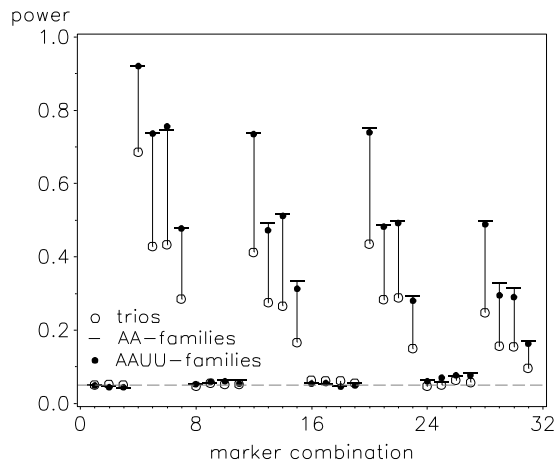
Power reduction F22-F20-F10-families  
H1, FBAT hbat\_bl: MUL KMod 1, 200 families



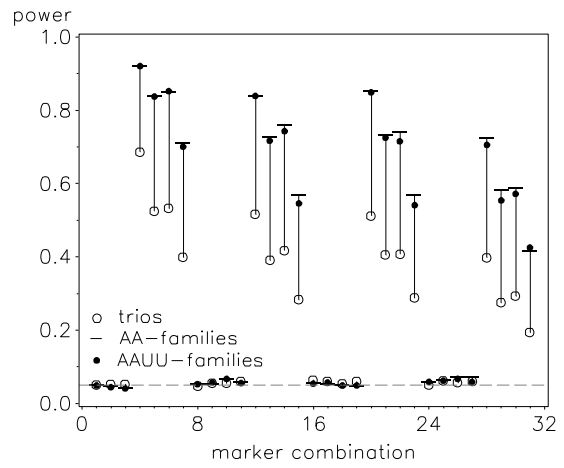
Power reduction F22-F20-F10-families  
H1, FBAT hbat\_gl: MUL KMod 1, 200 families



Power reduction F22-F20-F10-families  
H1, FBAT perm\_m: MUL KMod 1, 200 families

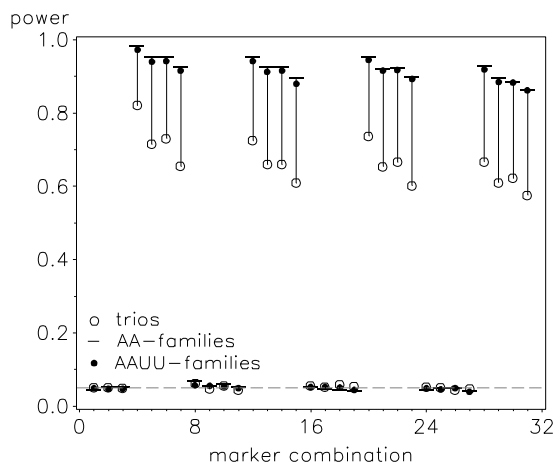


Power reduction F22-F20-F10-families  
H1, FBAT perm\_s: MUL KMod 1, 200 families

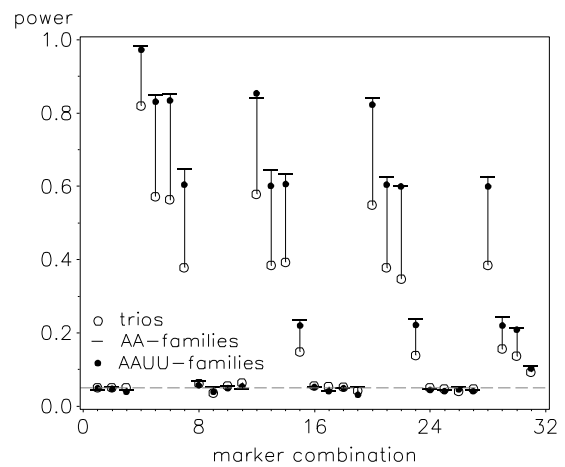


### Rezessives Modell KMod 1

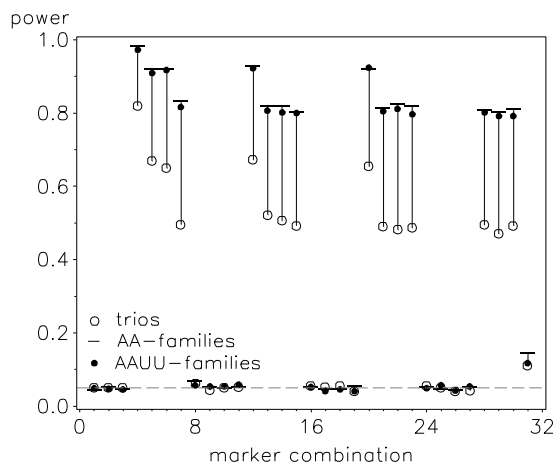
Power reduction F22-F20-F10-families  
H1, FBAT mm: REZ KMod 1, 600 families



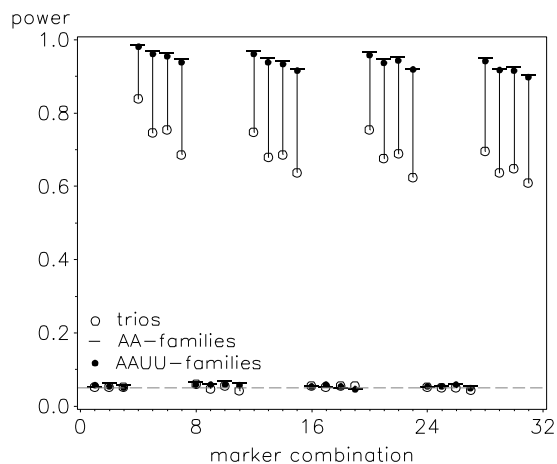
Power reduction F22-F20-F10-families  
H1, FBAT hbat\_b: REZ KMod 1, 600 families



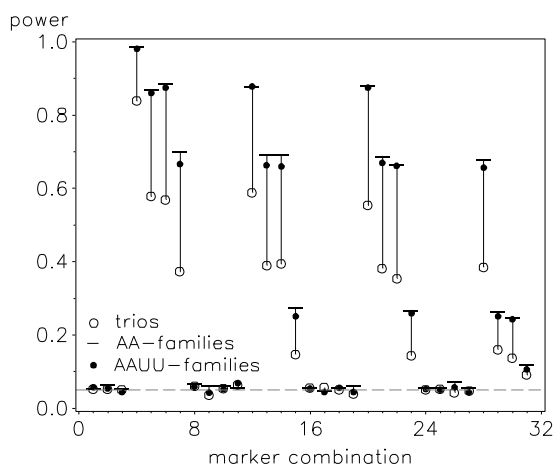
Power reduction F22–F20–F10–families  
H1, FBAT hbat\_g: REZ KMod 1, 600 families



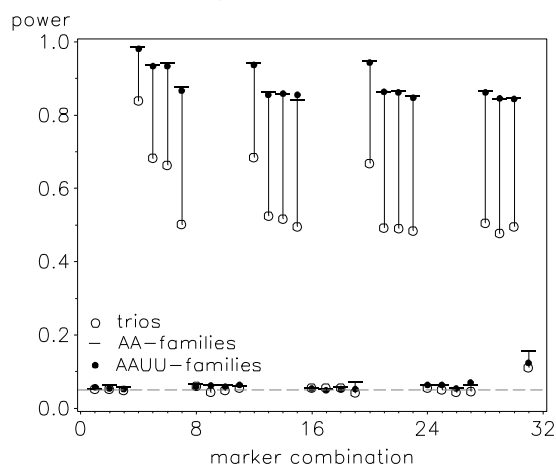
Power reduction F22–F20–F10–families  
H1, FBAT mm\_l: REZ KMod 1, 600 families



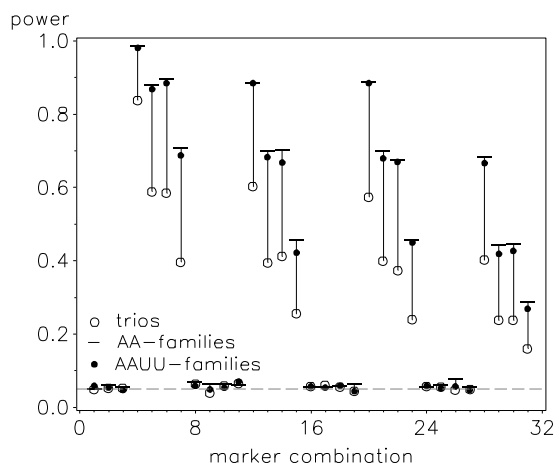
Power reduction F22–F20–F10–families  
H1, FBAT hbat\_bl: REZ KMod 1, 600 families



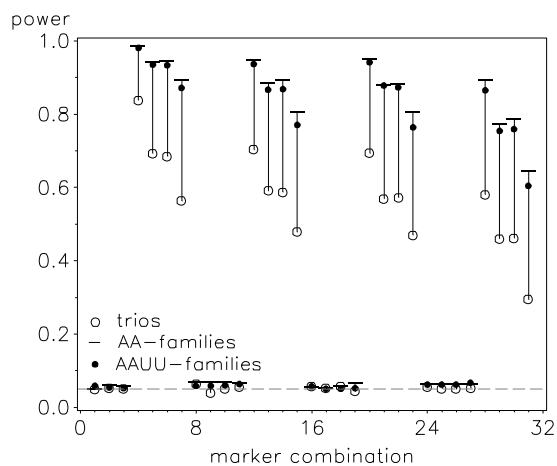
Power reduction F22–F20–F10–families  
H1, FBAT hbat\_g: REZ KMod 1, 600 families



Power reduction F22–F20–F10–families  
H1, FBAT perm\_m: REZ KMod 1, 600 families

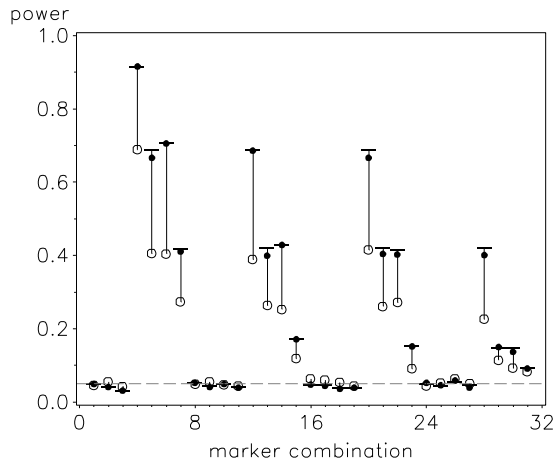


Power reduction F22–F20–F10–families  
H1, FBAT perm\_s: REZ KMod 1, 600 families

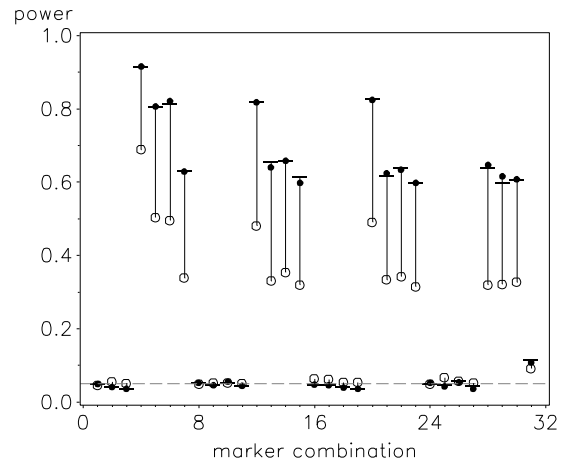


## Additives Modell KMod 2

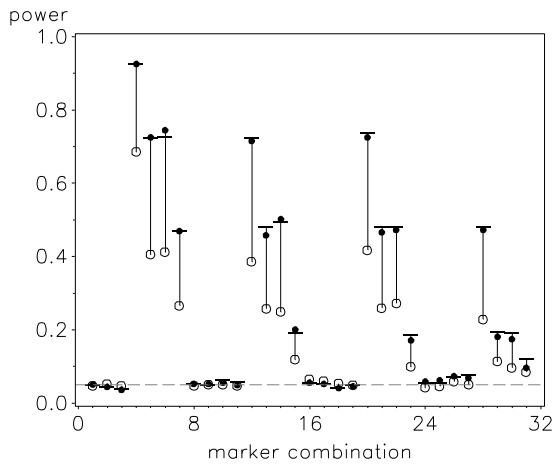
Power reduction F22–F20–F10–families  
H1, FBAT hbat\_b: ADD KMod 2, 400 families



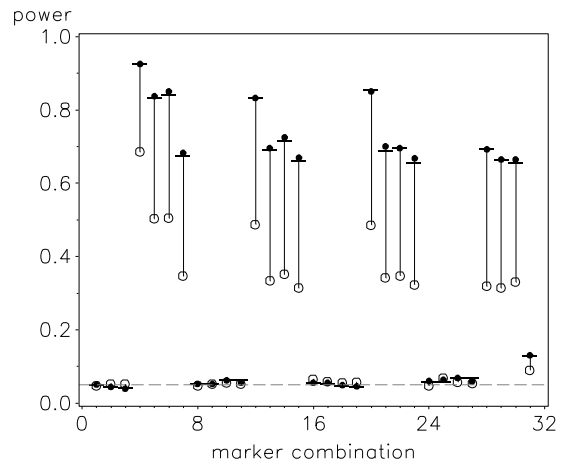
Power reduction F22–F20–F10–families  
H1, FBAT hbat\_g: ADD KMod 2, 400 families



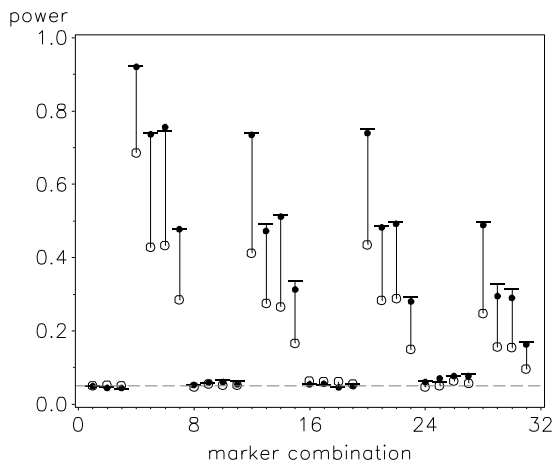
Power reduction F22–F20–F10–families  
H1, FBAT hbat\_bl: ADD KMod 2, 400 families



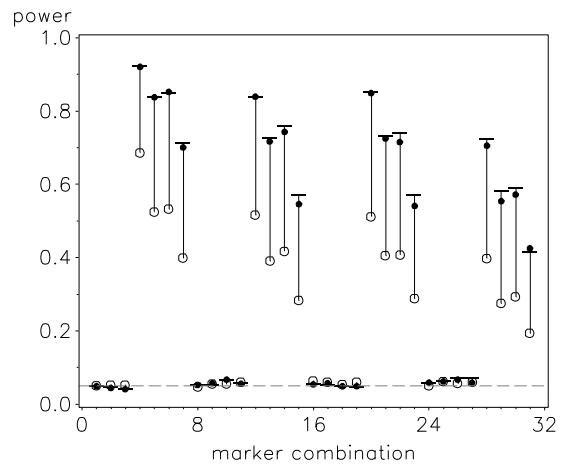
Power reduction F22–F20–F10–families  
H1, FBAT hbat\_gl: ADD KMod 2, 400 families



Power reduction F22–F20–F10–families  
H1, FBAT perm\_m: ADD KMod 2, 400 families

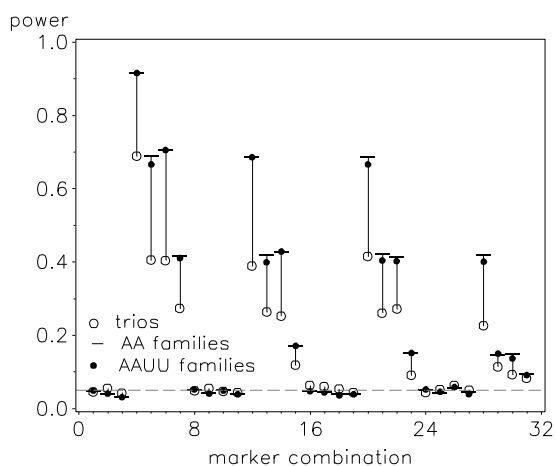


Power reduction F22–F20–F10–families  
H1, FBAT perm\_s: ADD KMod 2, 400 families

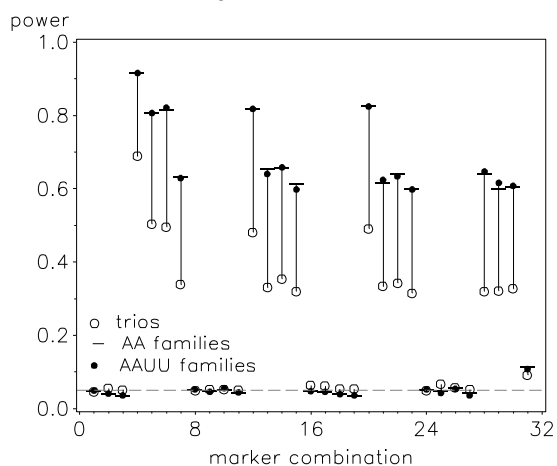


### Multiplicatives Modell KMod 3

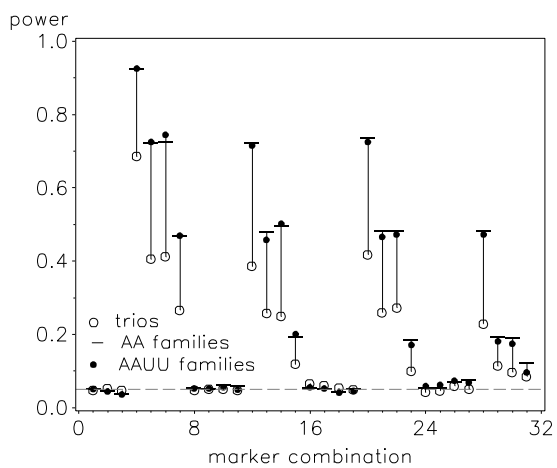
Power reduction F22–F20–F10–families  
H1, FBAT hbat\_b: MUL KMod 3, 400 families



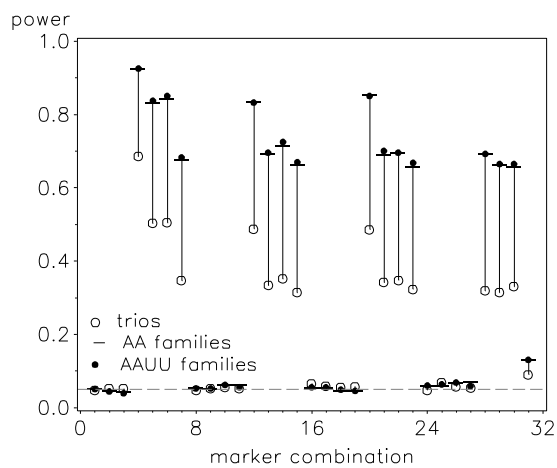
Power reduction F22–F20–F10–families  
H1, FBAT hbat\_g: MUL KMod 3, 400 families



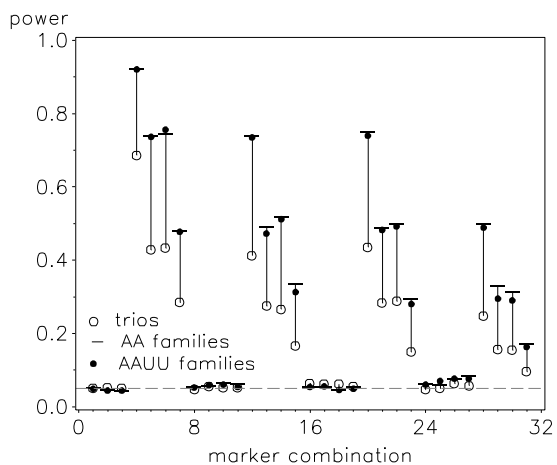
Power reduction F22–F20–F10–families  
H1, FBAT hbat\_bl: MUL KMod 3, 400 families



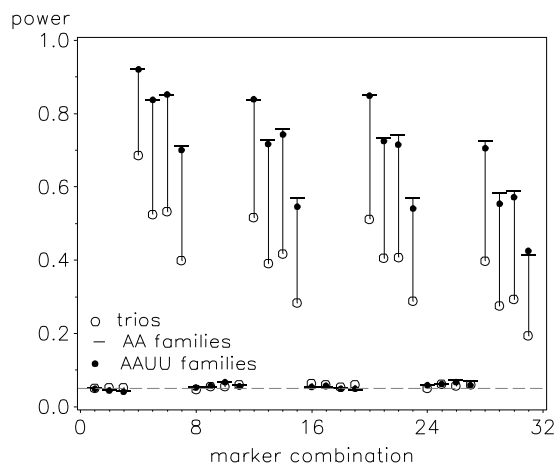
Power reduction F22–F20–F10–families  
H1, FBAT hbat\_gl: MUL KMod 3, 400 families



Power reduction F22–F20–F10–families  
H1, FBAT perm\_m: MUL KMod 3, 400 families



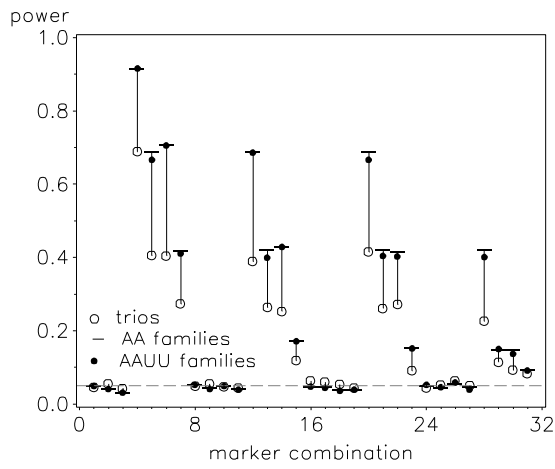
Power reduction F22–F20–F10–families  
H1, FBAT perm\_s: MUL KMod 3, 400 families



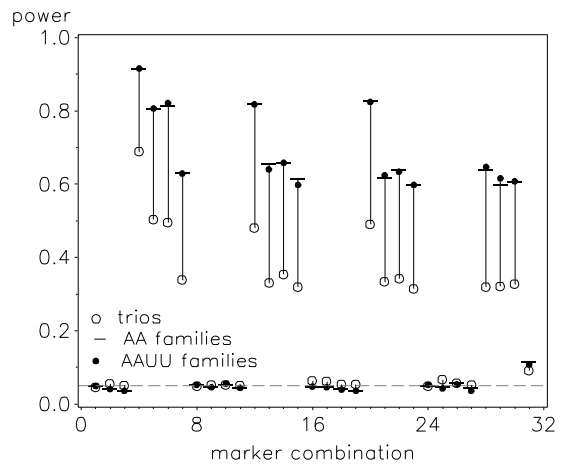


## Dominantes Modell KMod 2

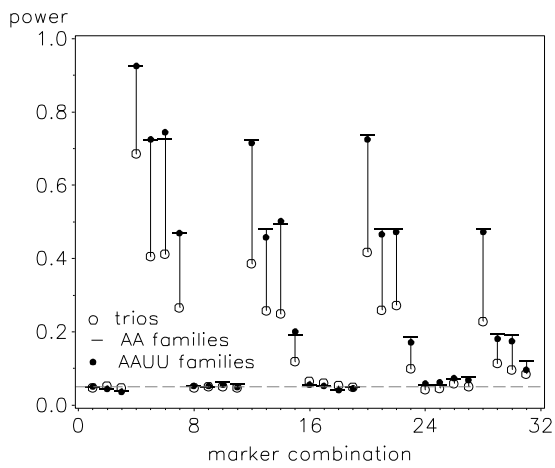
Power reduction F22–F20–F10–families  
H1, FBAT hbat\_b: DOM KMod 2, 500 families



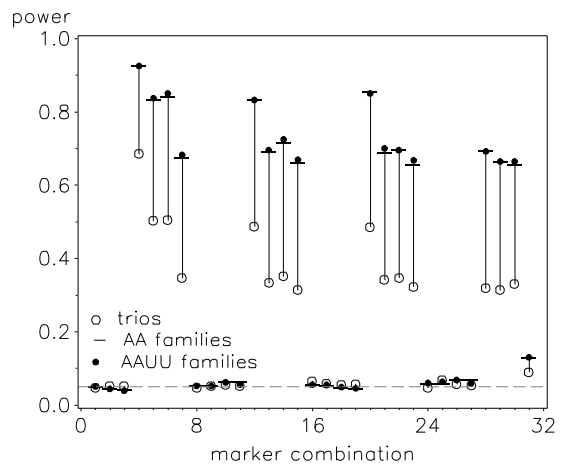
Power reduction F22–F20–F10–families  
H1, FBAT hbat\_g: DOM KMod 2, 500 families



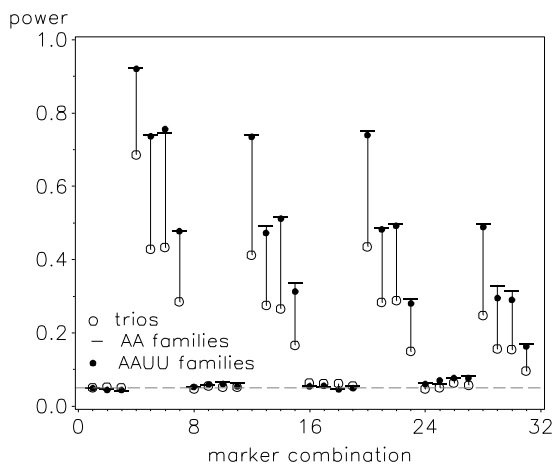
Power reduction F22–F20–F10–families  
H1, FBAT hbat\_bl: DOM KMod 2, 500 families



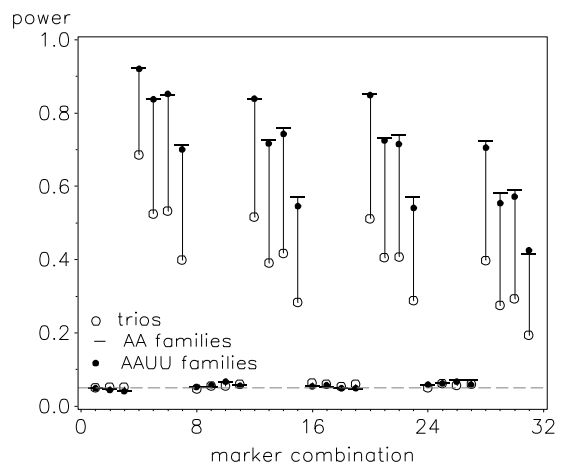
Power reduction F22–F20–F10–families  
H1, FBAT hbat\_gj: DOM KMod 2, 500 families



Power reduction F22–F20–F10–families  
H1, FBAT perm\_m: DOM KMod 2, 500 families



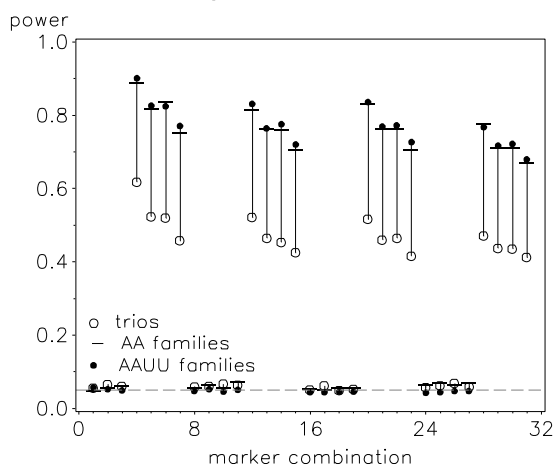
Power reduction F22–F20–F10–families  
H1, FBAT perm\_s: DOM KMod 2, 500 families



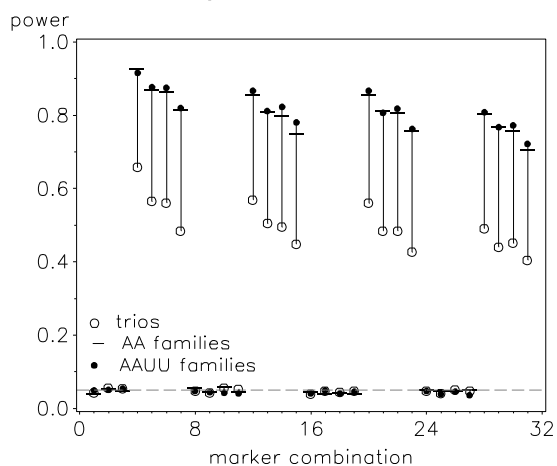
# Hotel\_Fam: nur KMod 1

## Additives Modell

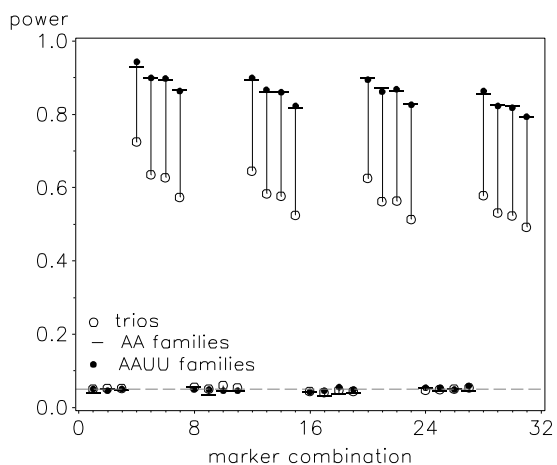
Power reduction F22–F20–F10–families  
H1, Hotel\_Fam gen\_as: KMod 1, 300 families



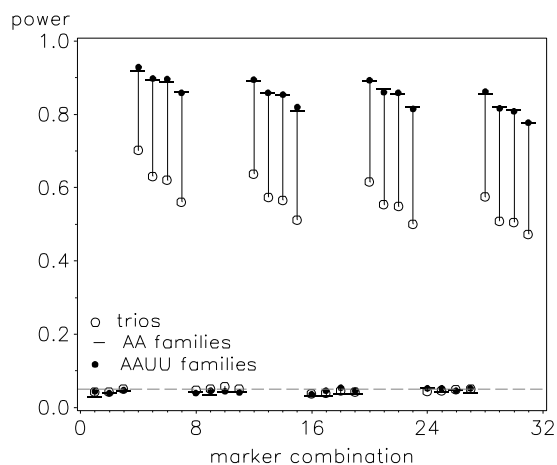
Power reduction F22–F20–F10–families  
H1, Hotel\_Fam gen\_pm: KMod 1, 300 families



Power reduction F22–F20–F10–families  
H1, Hotel\_Fam hap\_as: KMod 1, 300 families

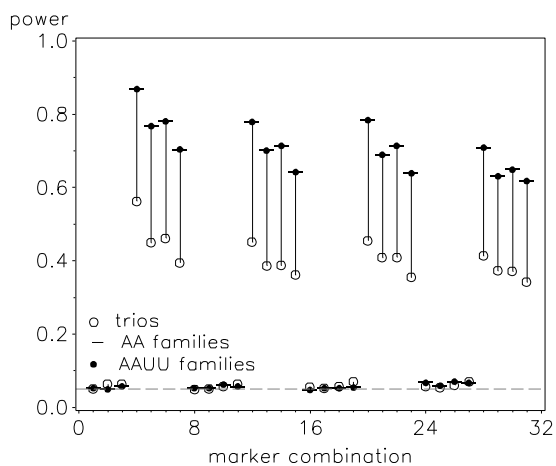


Power reduction F22–F20–F10–families  
H1, Hotel\_Fam hap\_pm: KMod 1, 300 families

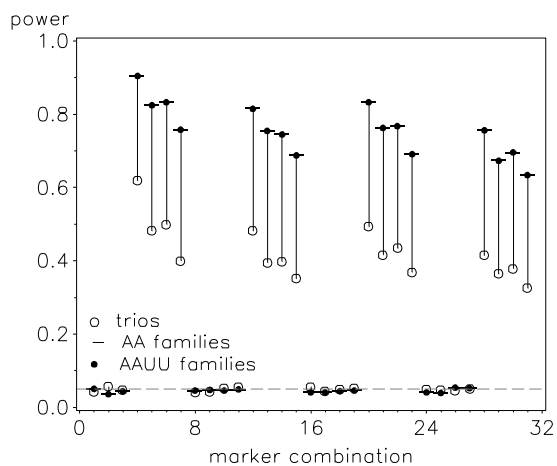


## Multiplikatives Modell

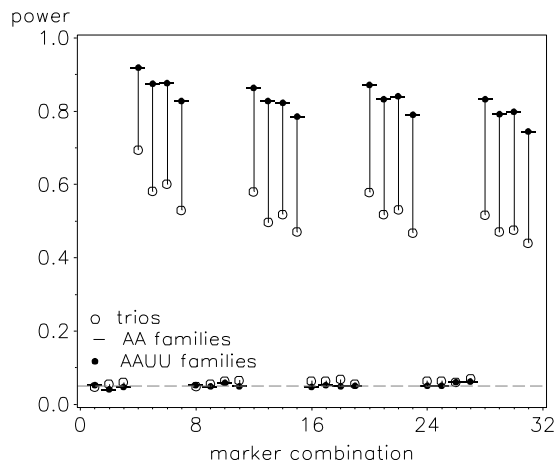
Power reduction F22–F20–F10–families  
H1, Hotel\_Fam gen\_as: KMod 1, 200 families



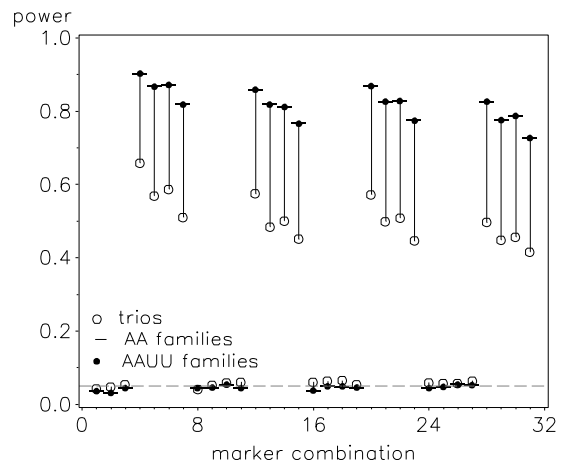
Power reduction F22–F20–F10–families  
H1, Hotel\_Fam gen\_pm: KMod 1, 200 families



Power reduction F22–F20–F10–families  
H1, Hotel\_Fam hap\_as: KMod 1, 200 families

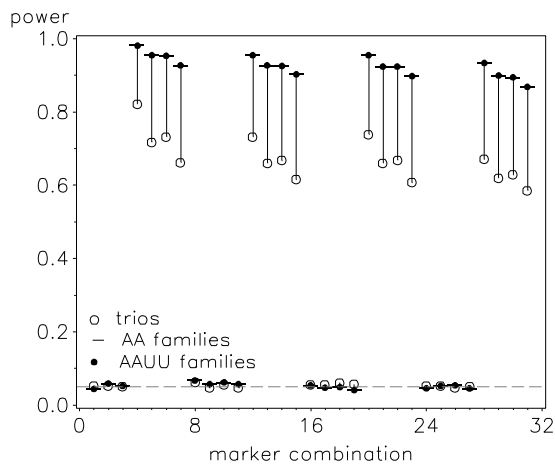


Power reduction F22–F20–F10–families  
H1, Hotel\_Fam hap\_pm: KMod 1, 200 families

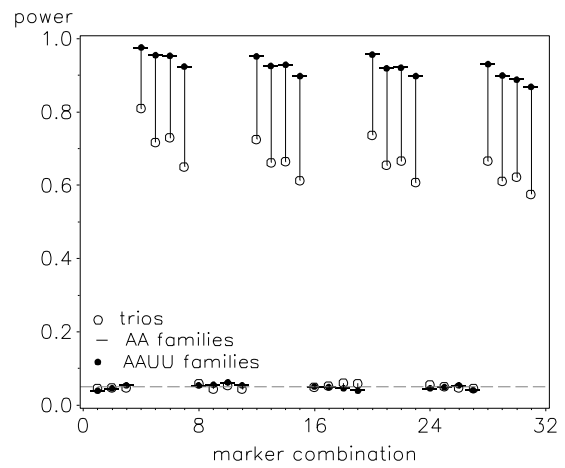


### Rezessives Modell

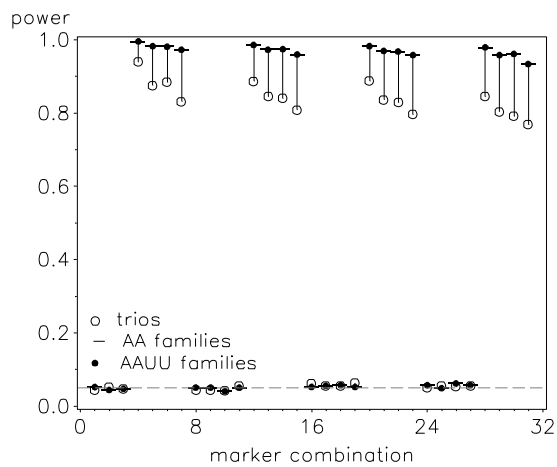
Power reduction F22–F20–F10–families  
H1, Hotel\_Fam hap\_as: KMod 1, 600 families



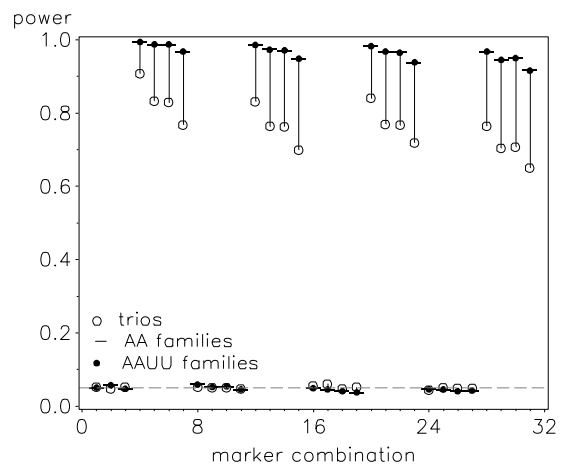
Power reduction F22–F20–F10–families  
H1, Hotel\_Fam hap\_pm: KMod 1, 600 families



Power reduction F22–F20–F10–families  
H1, Hotel\_Fam gen\_as: KMod 1, 600 families



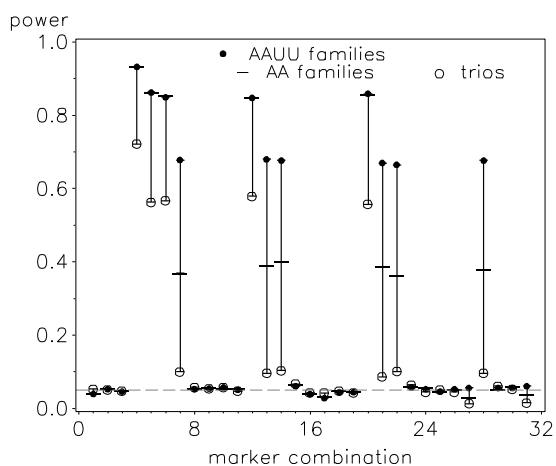
Power reduction F22–F20–F10–families  
H1, Hotel\_Fam gen\_pm: KMod 1, 600 families



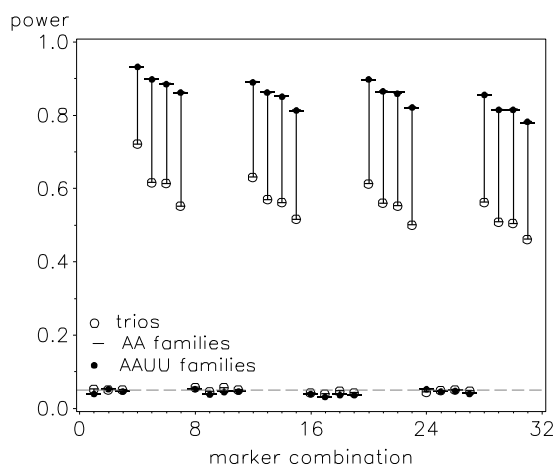
# Unphased

## Additives Modell KMod 1

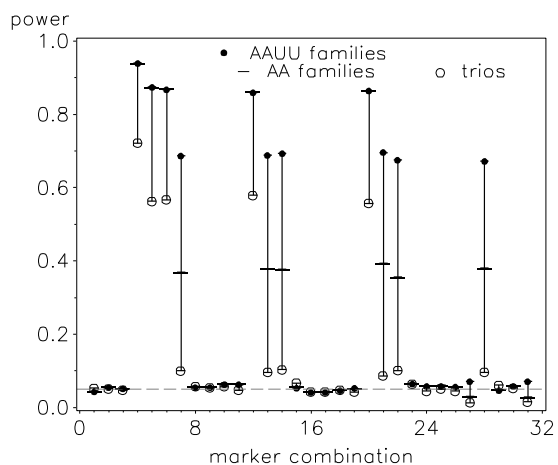
Power reduction F22–F20–F10–families  
H1, Unphased full: ADD, KMod 1 300 families



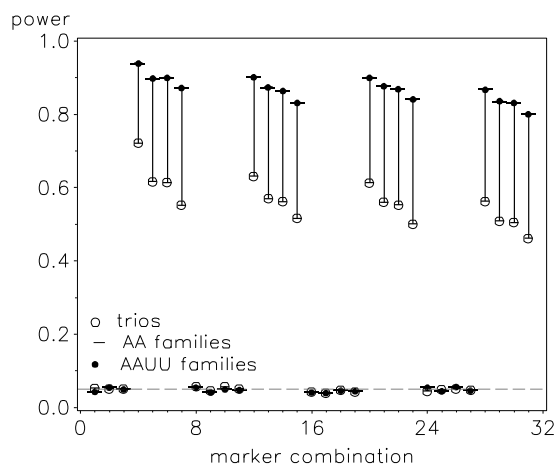
Power reduction F22–F20–F10–families  
H1, Unphased allm: ADD, KMod 1 300 families



Power reduction F22–F20–F10–families  
H1, Unphased full\_: ADD, KMod 1 300 families

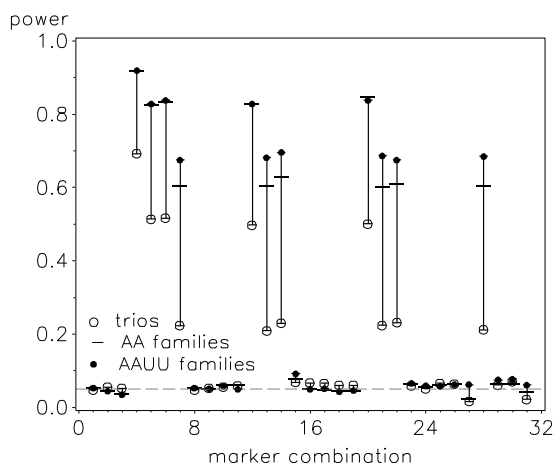


Power reduction F22–F20–F10–families  
H1, Unphased allm\_: ADD, KMod 1 300 families

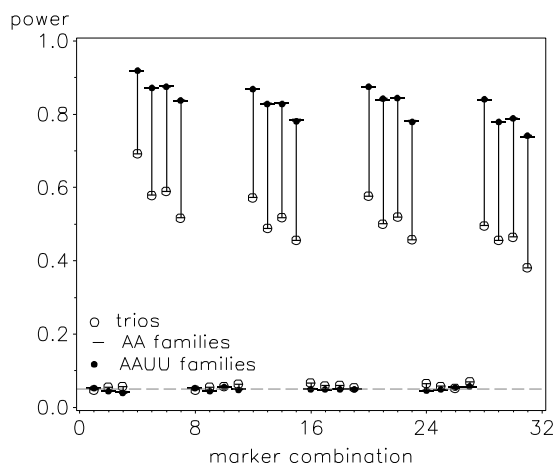


## Multiplikatives Modell KMod 1

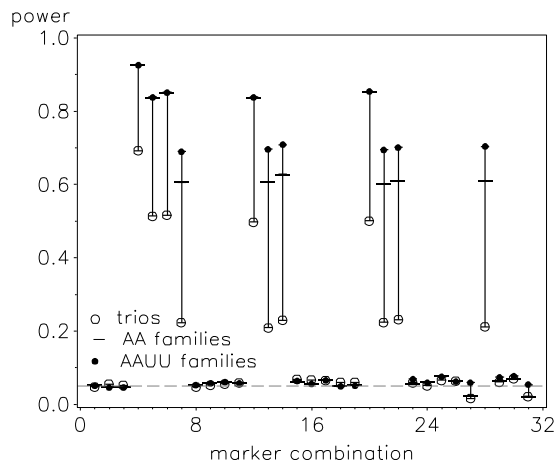
Power reduction F22–F20–F10–families  
H1, Unphased full: MUL, KMod 1 200 families



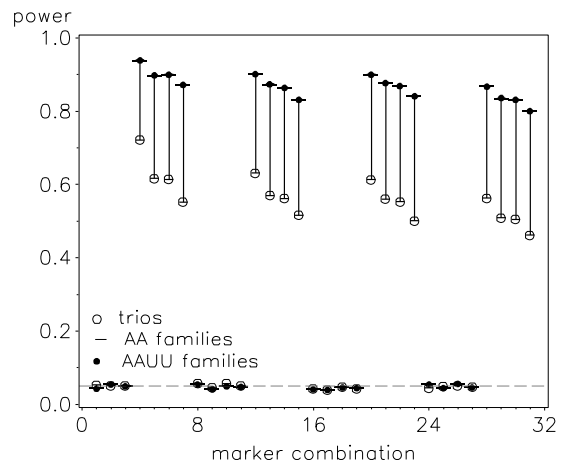
Power reduction F22–F20–F10–families  
H1, Unphased allm: MUL, KMod 1 200 families



Power reduction F22–F20–F10–families  
H1, Unphased full\_I: MUL, KMod 1 200 families

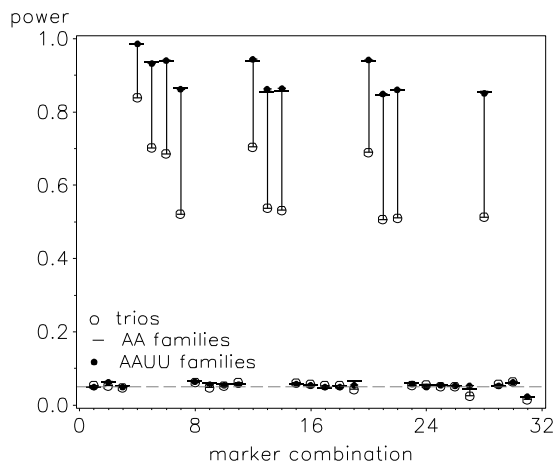


Power reduction F22–F20–F10–families  
H1, Unphased allm\_I: ADD, KMod 1 300 families

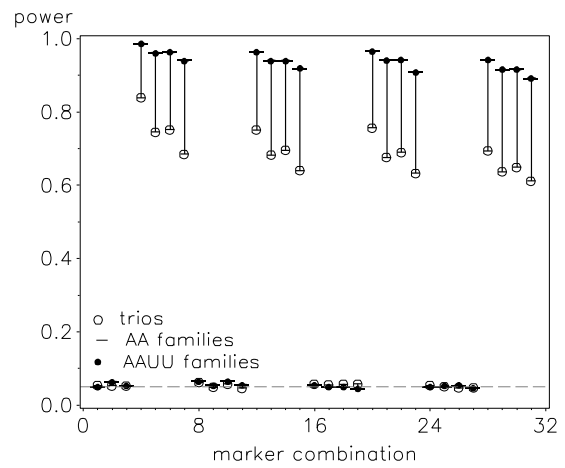


### Rezessives Modell KMod 1

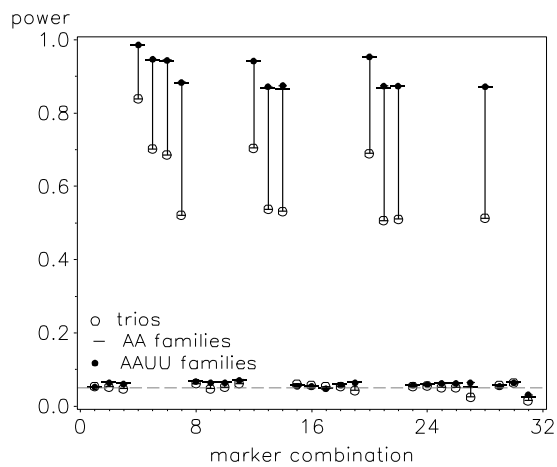
Power reduction F22–F20–F10–families  
H1, Unphased full: REZ, KMod 1 600 families



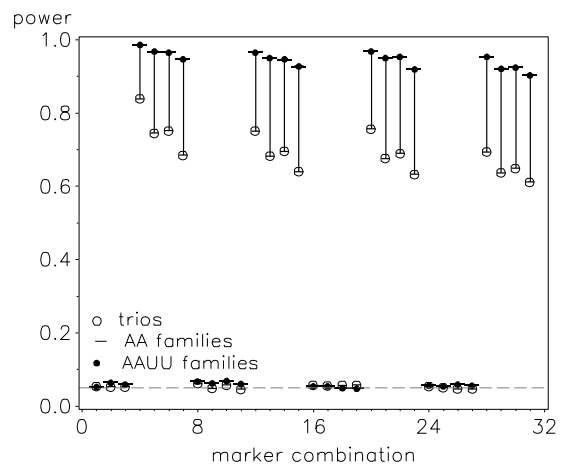
Power reduction F22–F20–F10–families  
H1, Unphased allm: REZ, KMod 1 600 families



Power reduction F22–F20–F10–families  
H1, Unphased full\_I: REZ, KMod 1 600 families

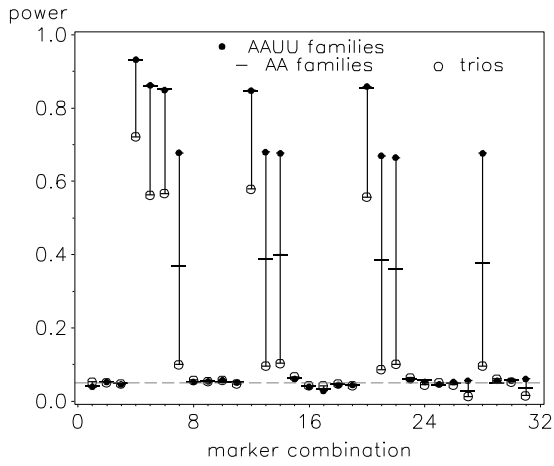


Power reduction F22–F20–F10–families  
H1, Unphased allm\_I: REZ, KMod 1 600 families

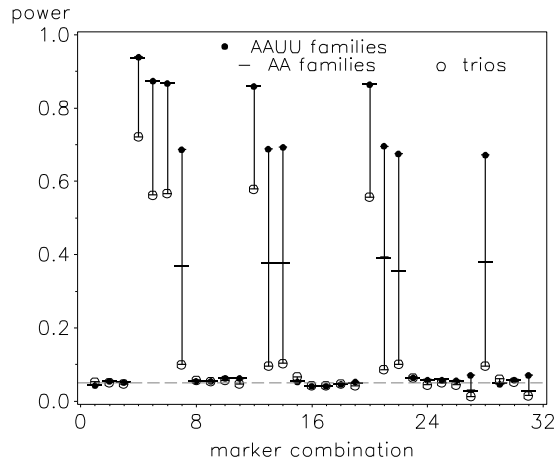


### Additives Modell KMod 2

Power reduction F22–F20–F10–families  
H1, Unphased full: ADD, KMod 1 300 families

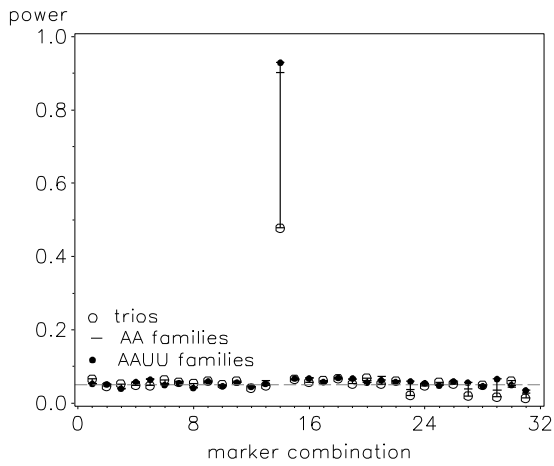


Power reduction F22–F20–F10–families  
H1, Unphased full\_I: ADD, KMod 1 300 families

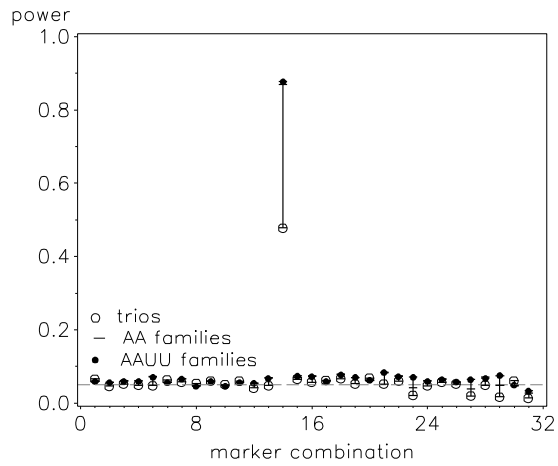


### Multiplikatives Modell KMod 3

Power reduction F22–F20–F10–families  
H1, Unphased, full: MUL, KMod 3 600 families

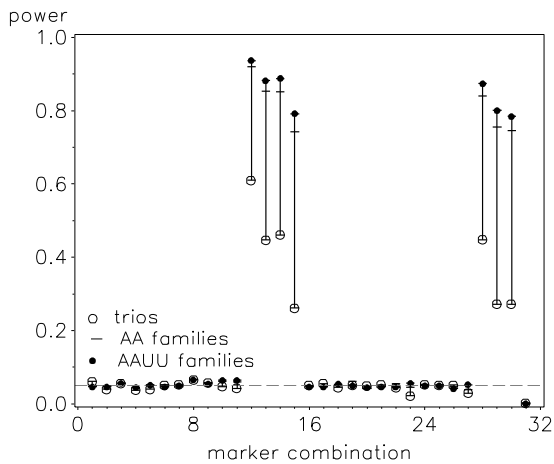


Power reduction F22–F20–F10–families  
H1, Unphased, full\_I: MUL, KMod 3 600 families

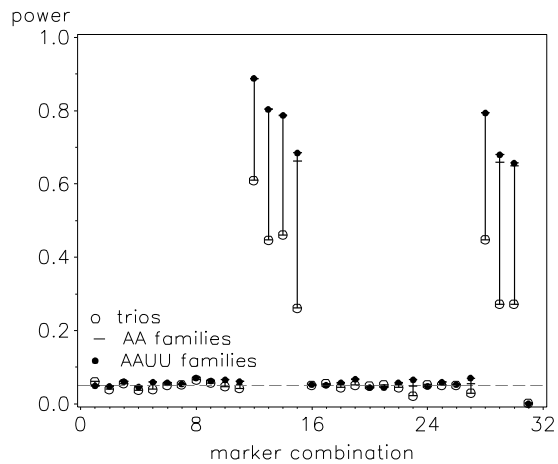


### Dominantes Modell KMod 2

Power reduction F22–F20–F10–families  
H1, Unphased, full: DOM, KMod 2 600 families

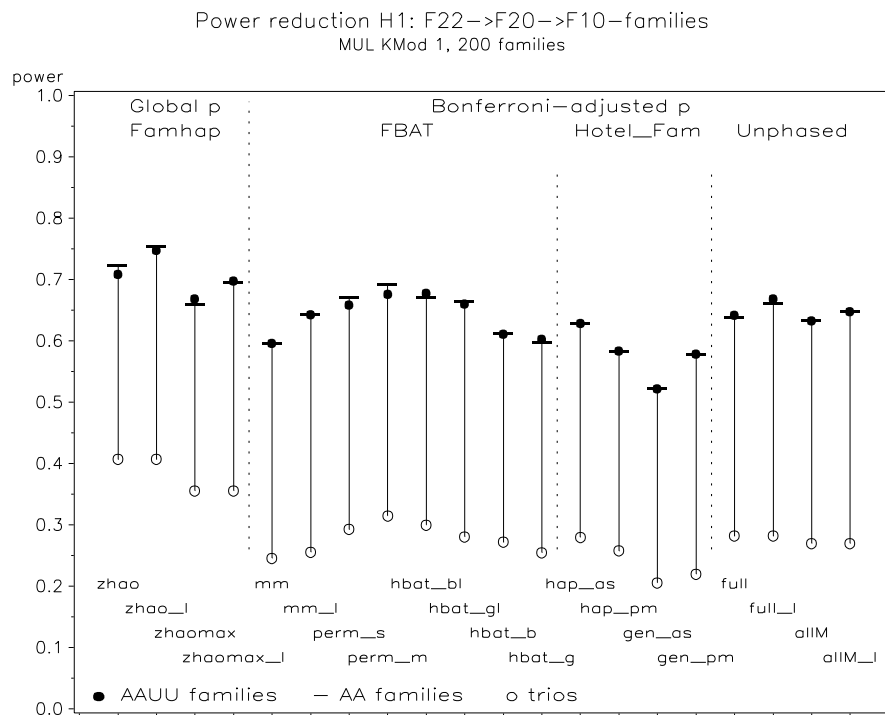
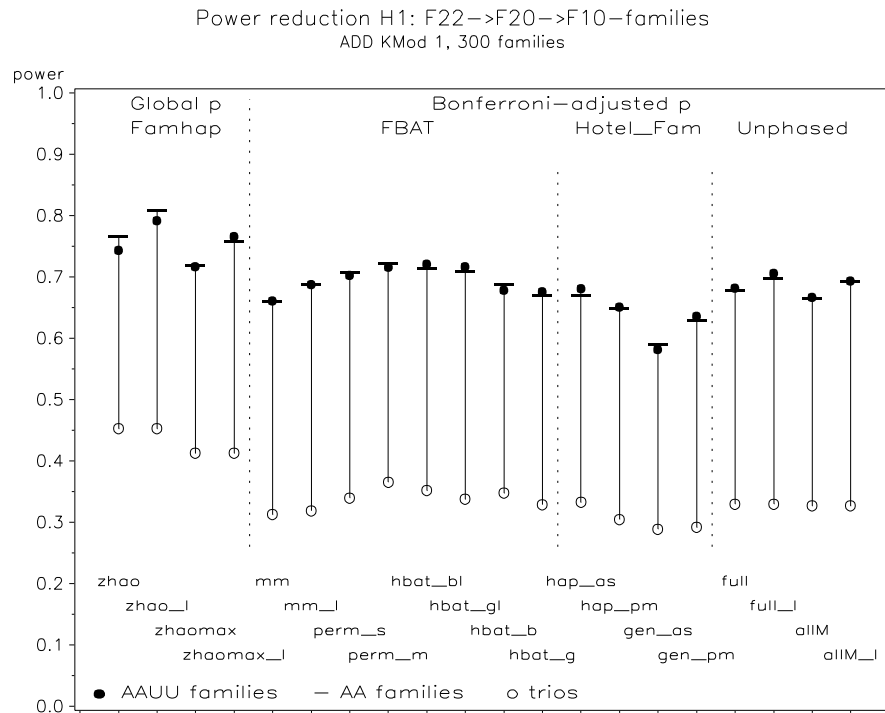


Power reduction F22–F20–F10–families  
H1, Unphased, full\_I: DOM, KMod 2 600 families

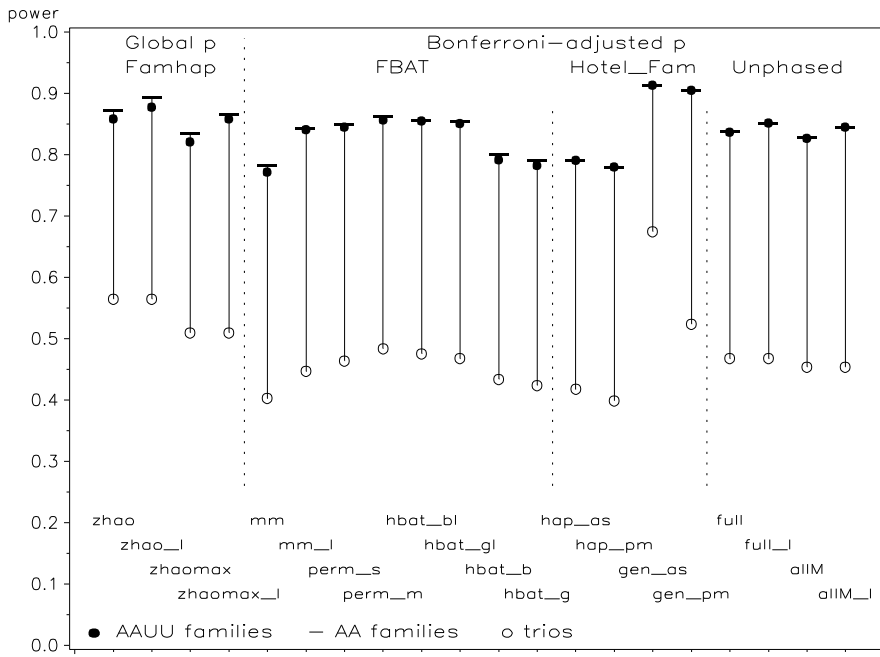


# Vergleich aller Methoden

## Modell KMod 1

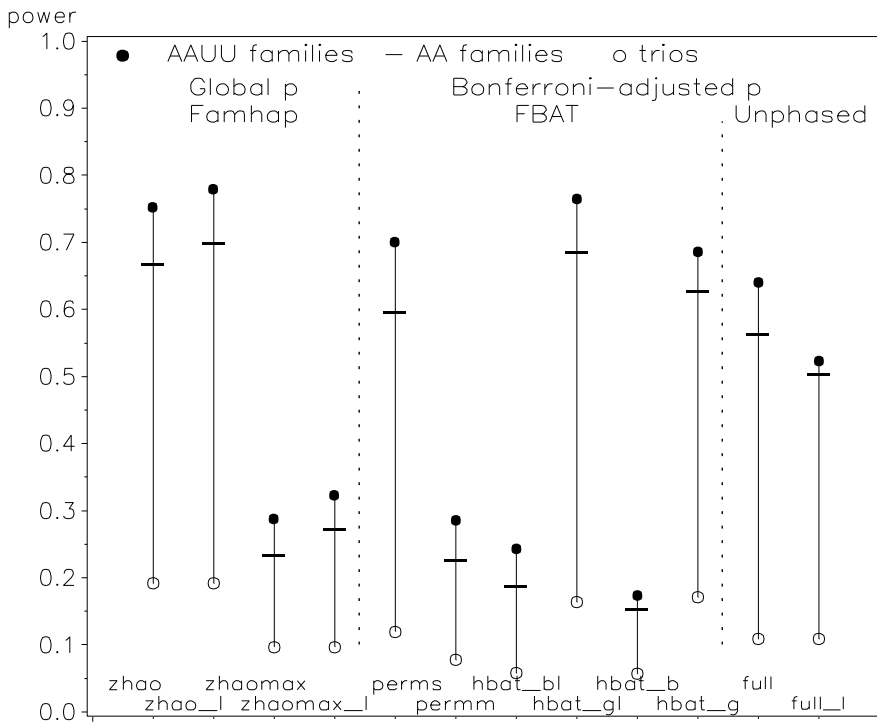


Power reduction H1: F22→F20→F10-families  
 REZ KMod 1, 600 families



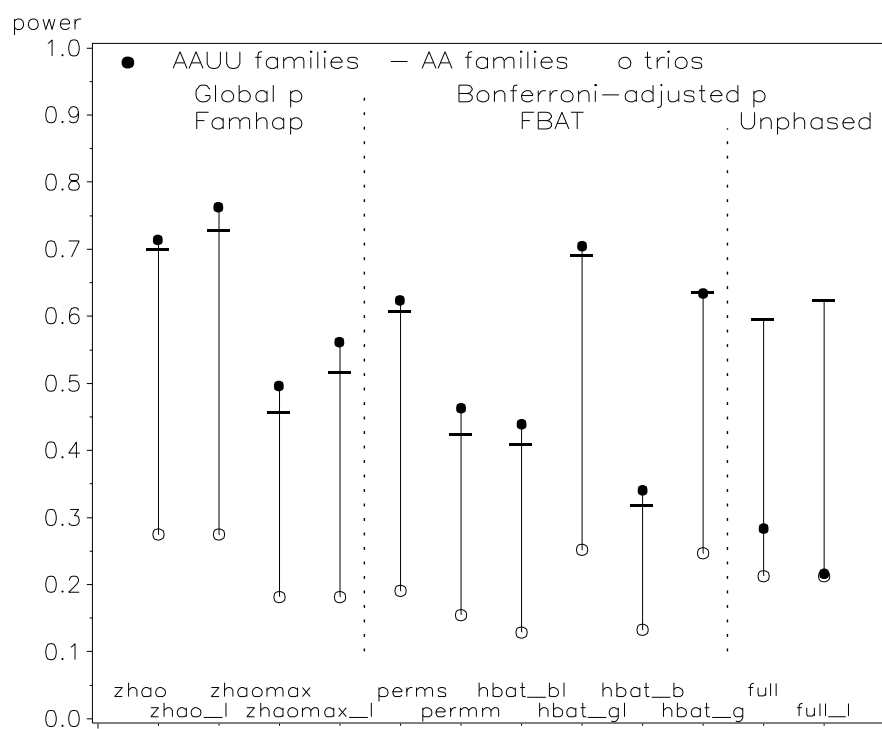
**Modelle KMod 2 und 3**

Power reduction H1: F22→F20→F10-families  
 MUL KMod 3, 400 families

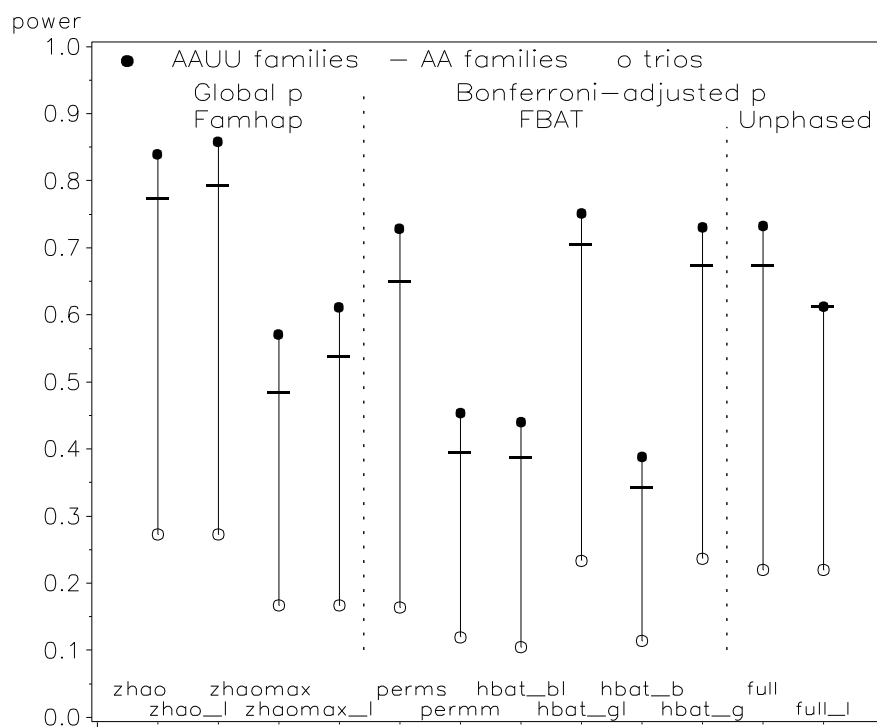




Power reduction H1: F22→F20→F10–families  
ADD KMod 2, 400 families

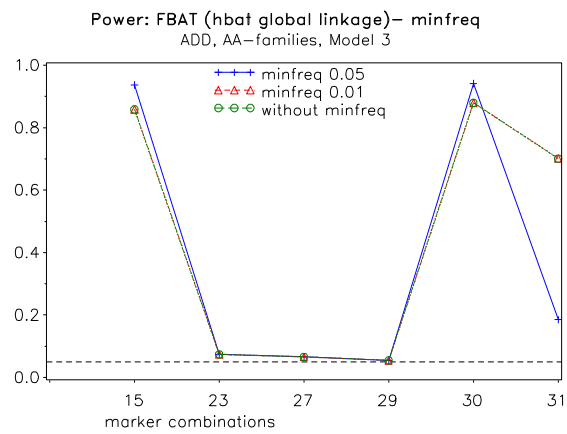
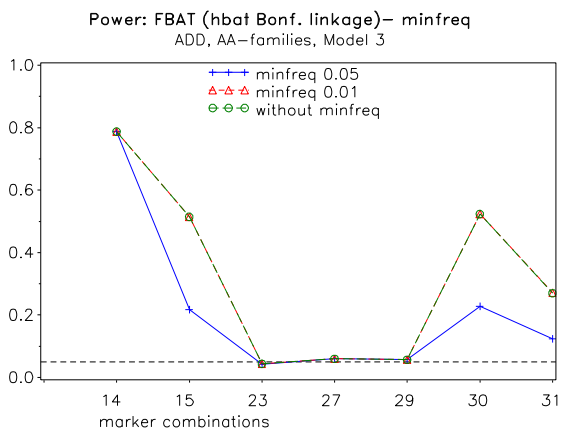
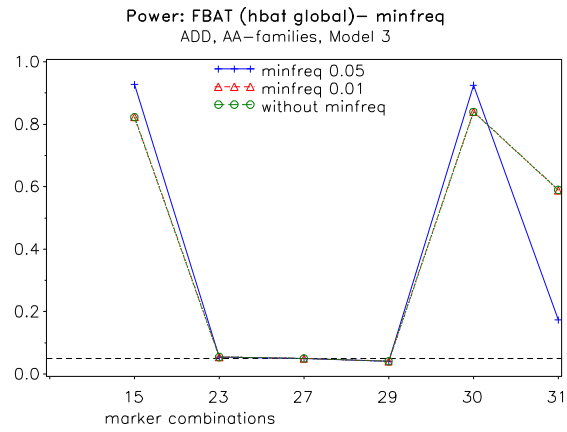
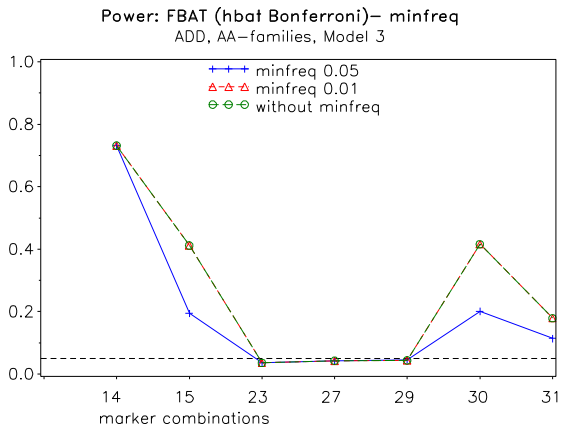


Power reduction H1: F22→F20→F10–families  
DOM KMod 2, 500 families

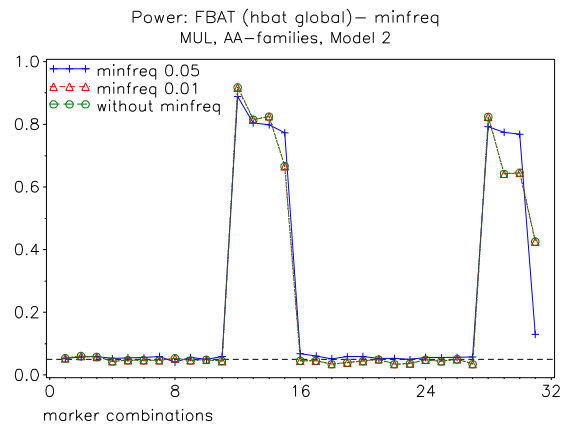
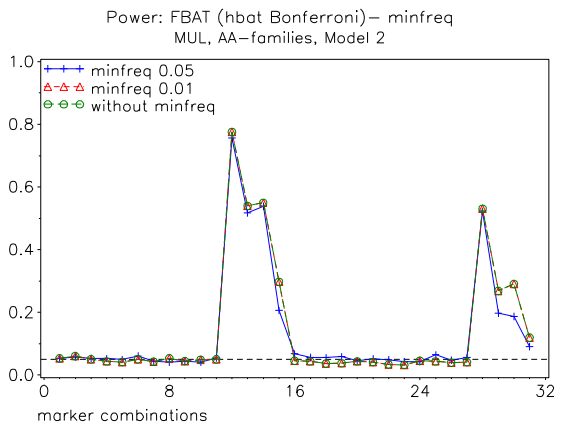


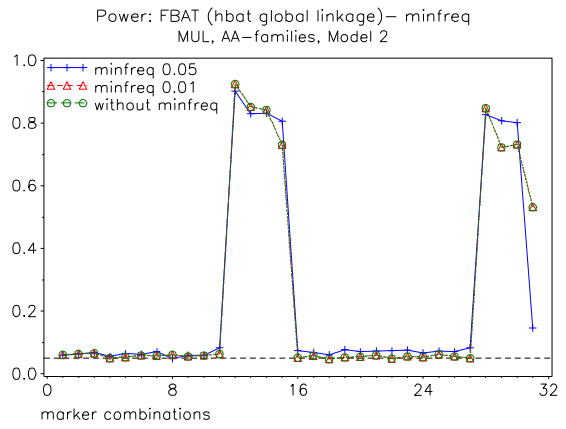
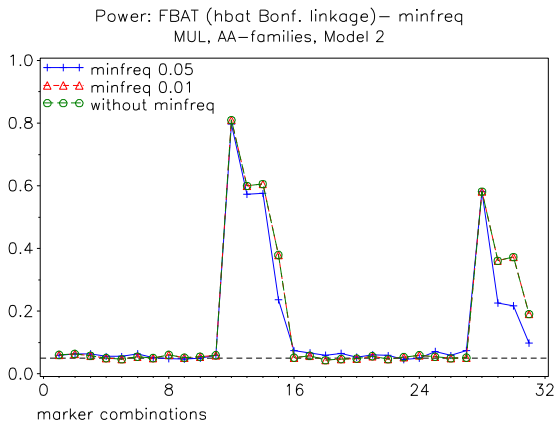
# Einfluss der FBAT-Optionen minfreq 0.05, 0.01, 0

## Additives Modell KMod 3, AA-Familien

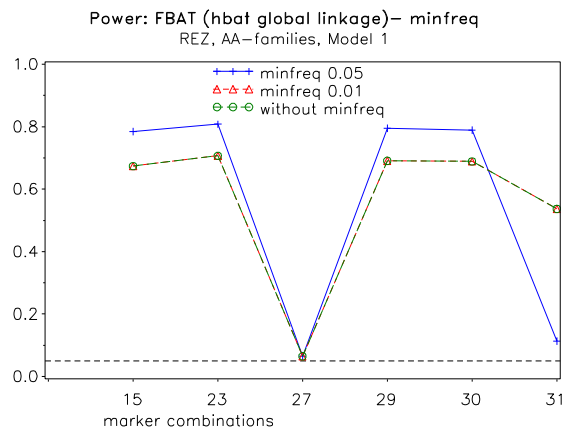
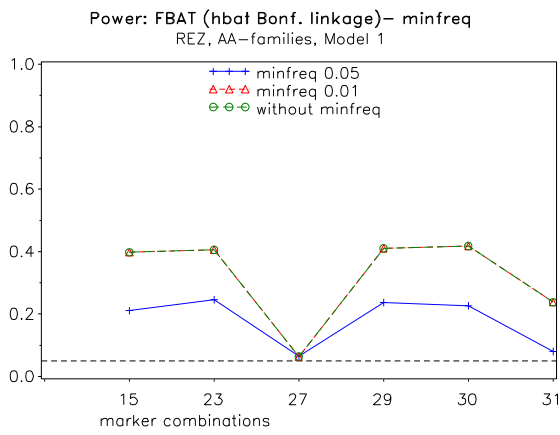
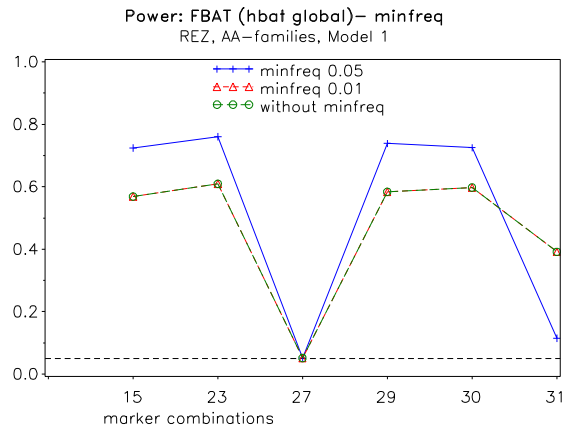
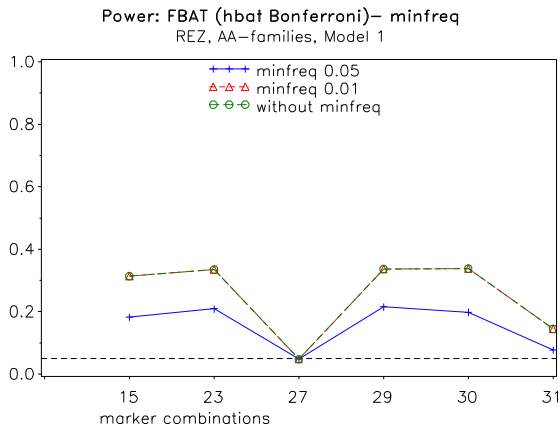


## Multiplikatives Modell KMod 2, AA-Familien



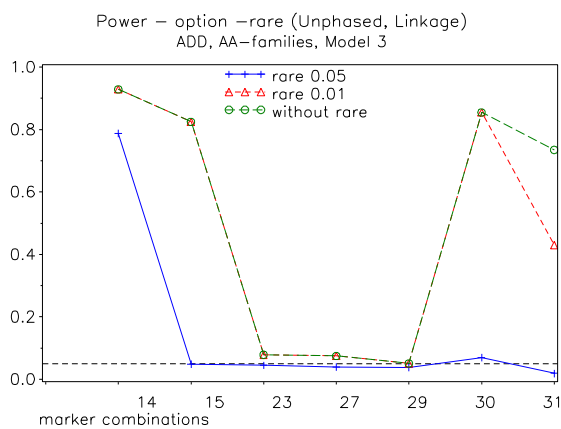
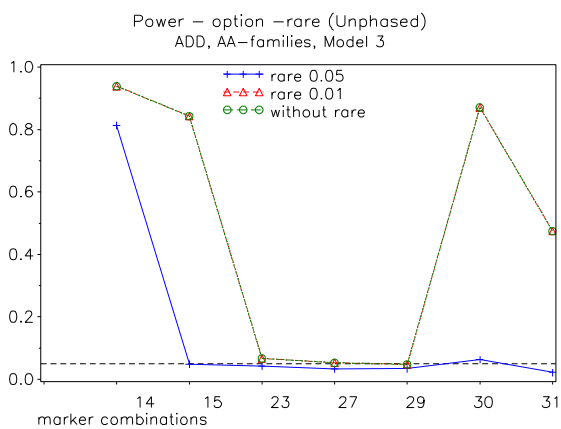


### Rezessives Modell KMod 1, AA-Familien

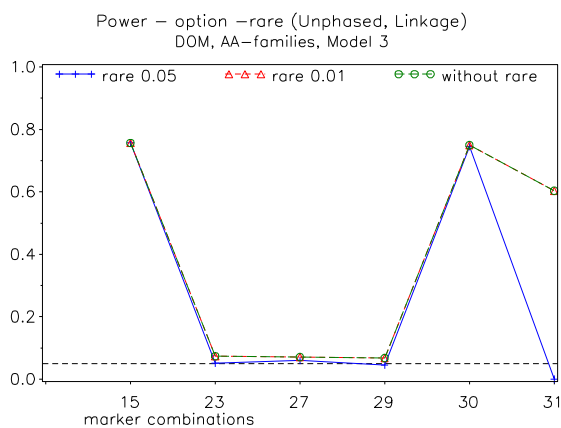
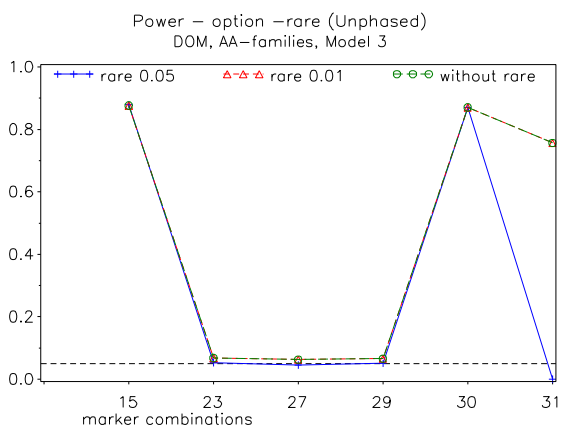


# Einfluss der Unphased-Optionen -rare 0.05, 0.01, 0

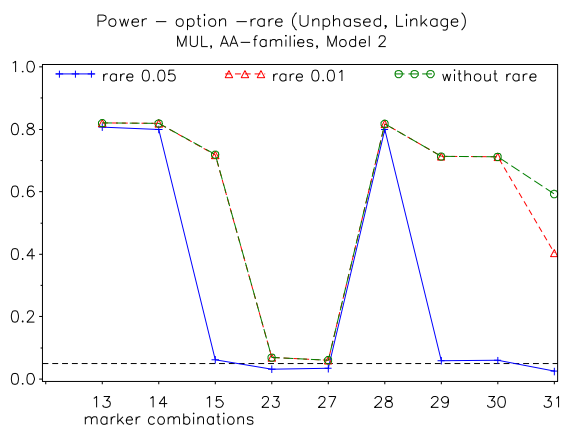
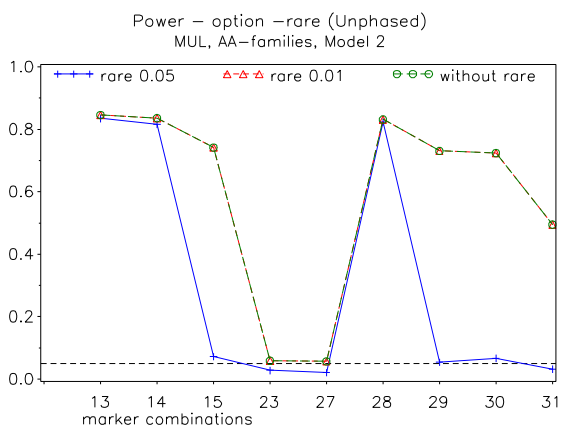
## Additives Modell KMod 3, AA-Familien



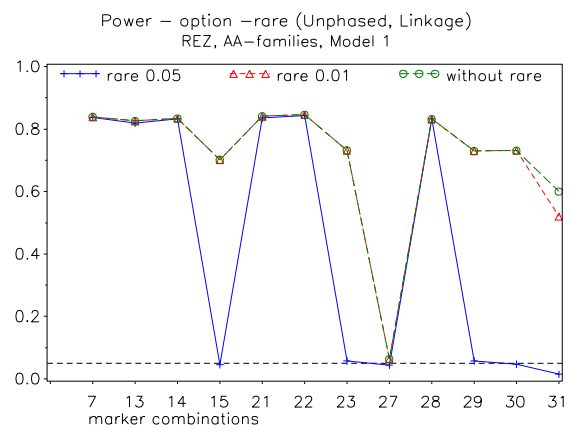
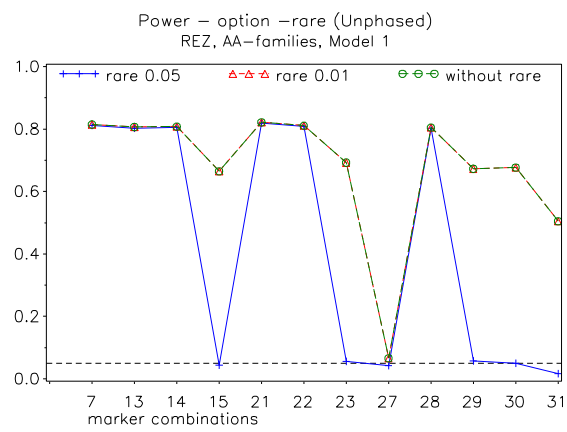
## Dominantes Modell KMod 3, AA-Familien



## Multiplikatives Modell KMod 2, AA-Familien



## Rezessives Modell KMod 1, AA-Familien



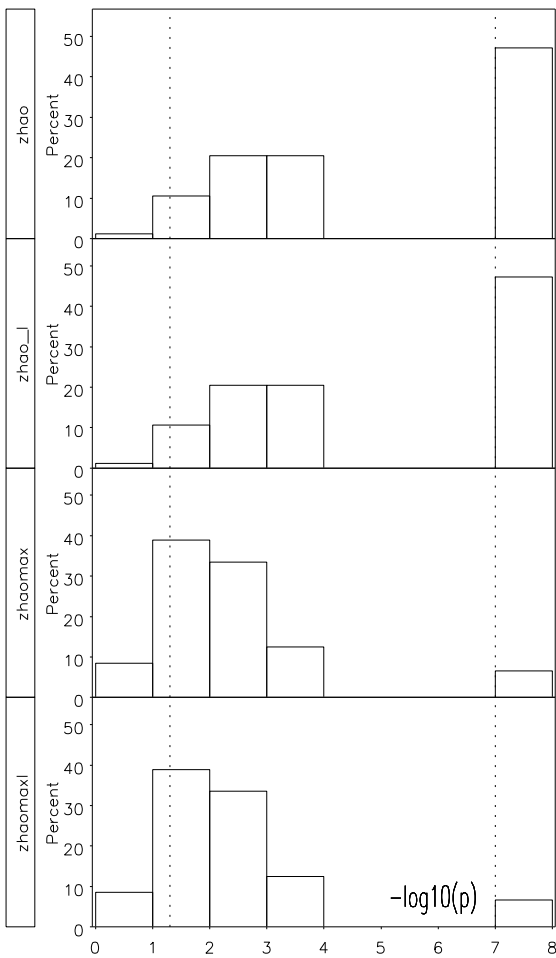
# P-Wert-Verteilung für ausgewählte Statistiken und Markerkombinationen

## Famhap / Unphased

### Additives Modell KMod 3, Trios

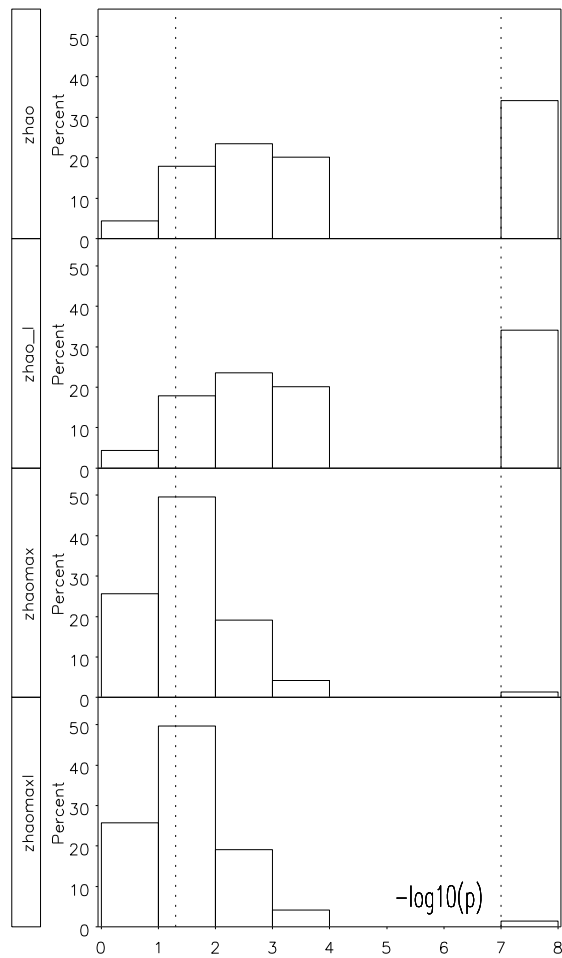
#### Markerkombination 14

Distribution of p-values: Famhap, ADD, KMod 3, 2500 trios, marker combination 14



#### Markerkombination 15

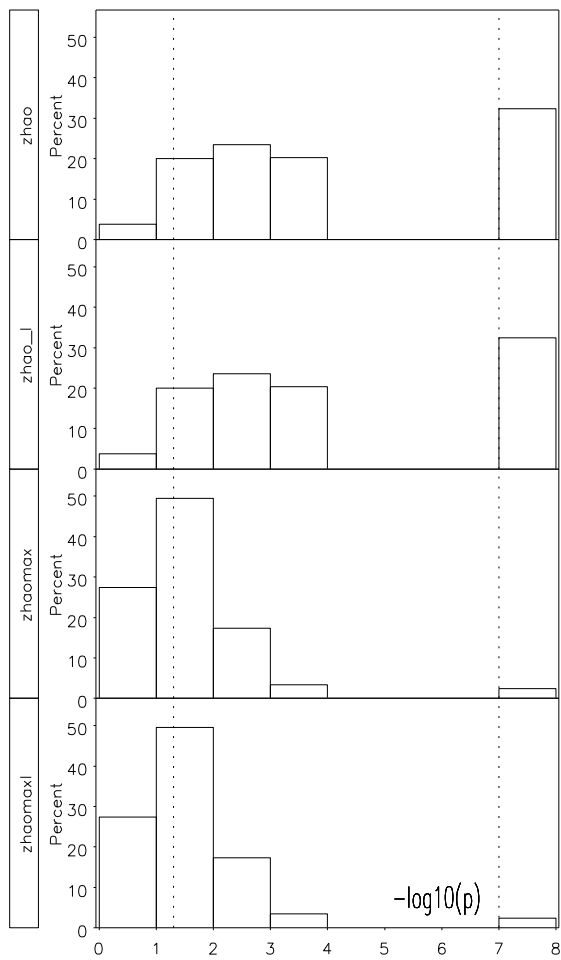
Distribution of p-values: Famhap, ADD, KMod 3, 2500 trios, marker combination 15



## Trios

### Markerkombination 30

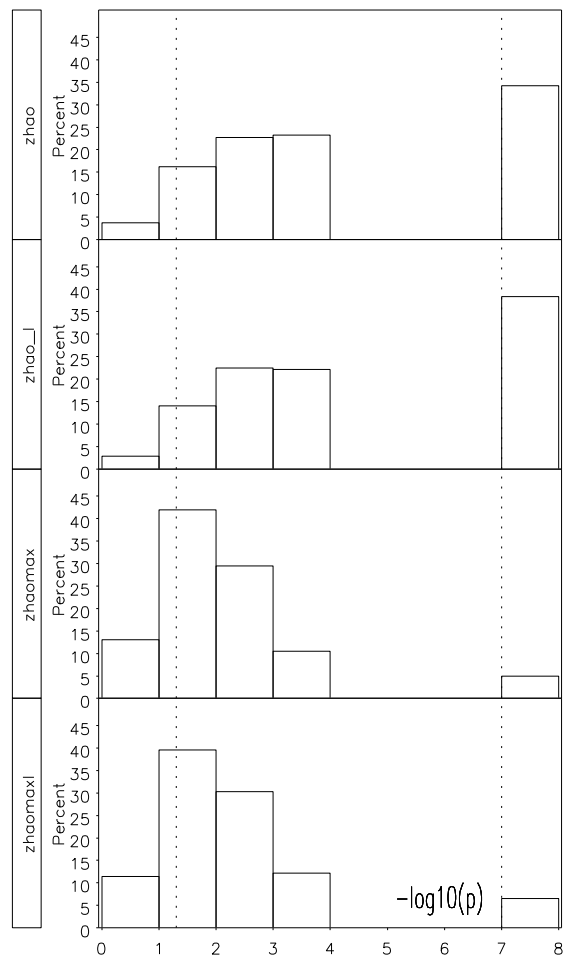
Distribution of p-values: Famhap, ADD,  
KMod 3, 2500 trios, marker combination 30



## AA-Familien

### Markerkombination 14

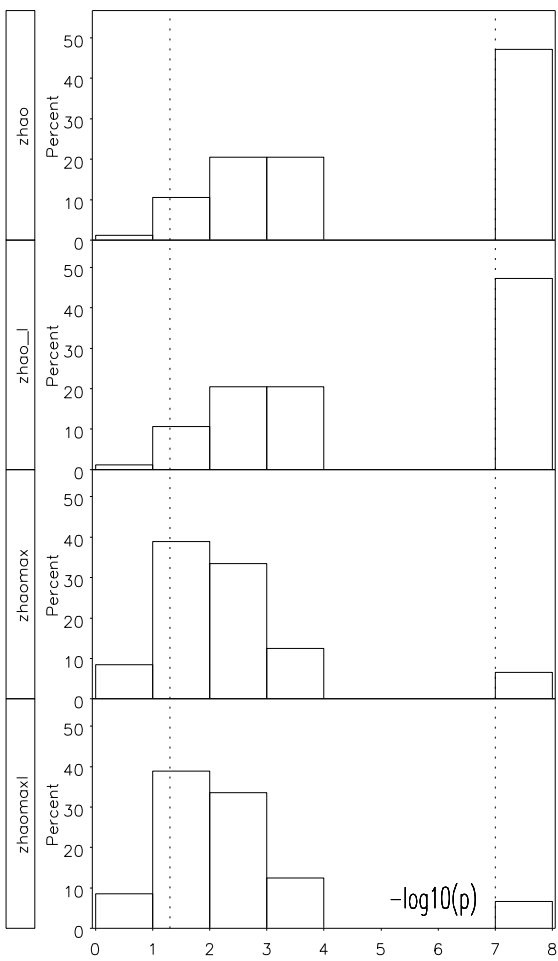
Distribution of p-values: Famhap, ADD,  
KMod 3, 500 AA families, marker combination 14



# AA-Familien

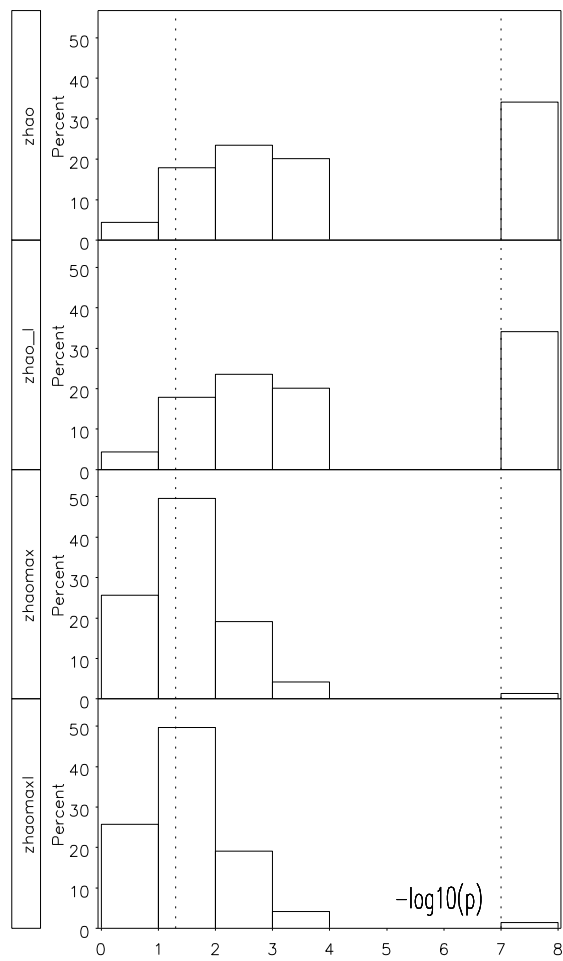
## Markerkombination 15

Distribution of p-values: Famhap, ADD, KMod 3, 2500 trios, marker combination 14



## Markerkombination 30

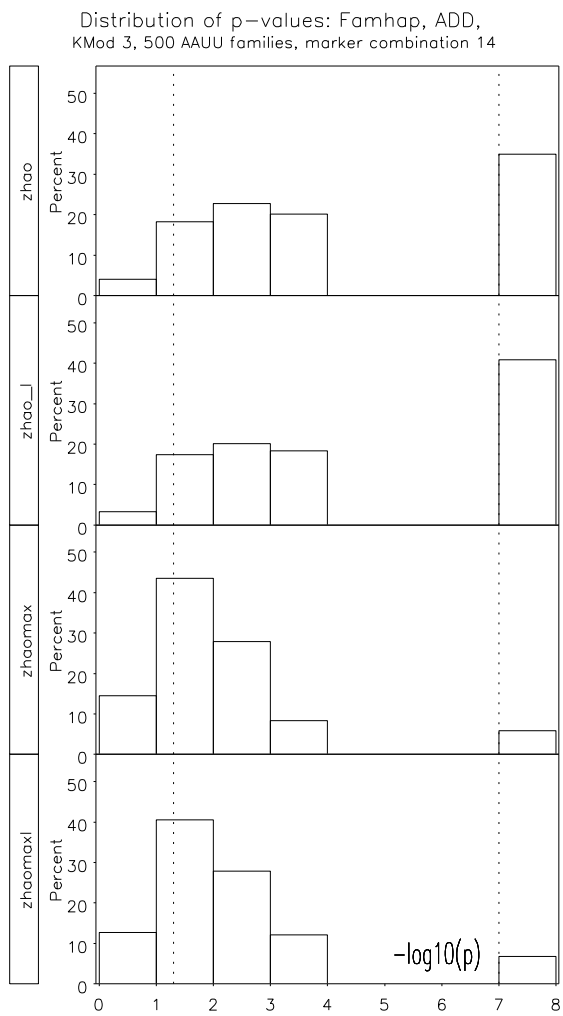
Distribution of p-values: Famhap, ADD, KMod 3, 2500 trios, marker combination 15



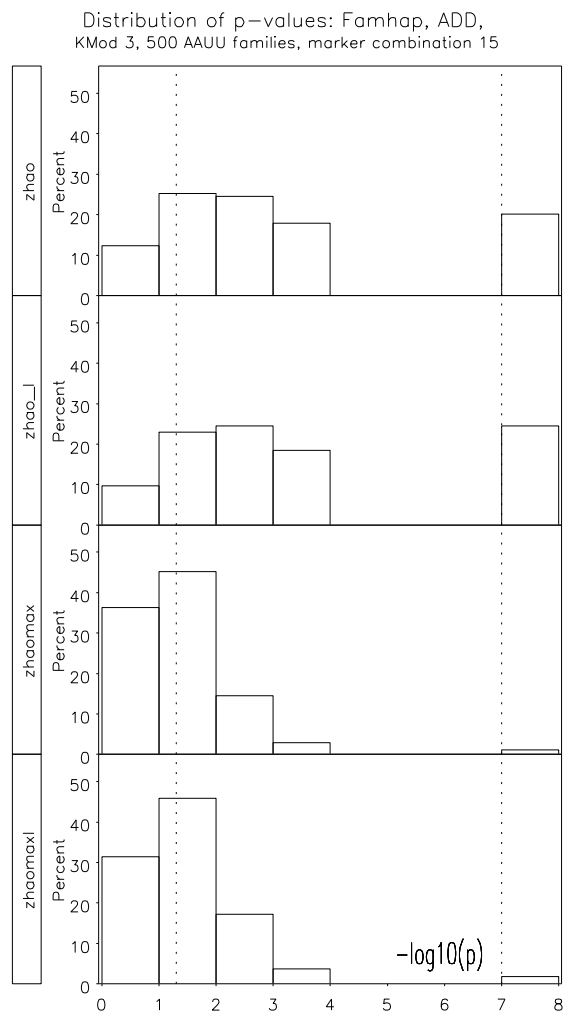


## AAUU-Familien

### Markerkombination 14



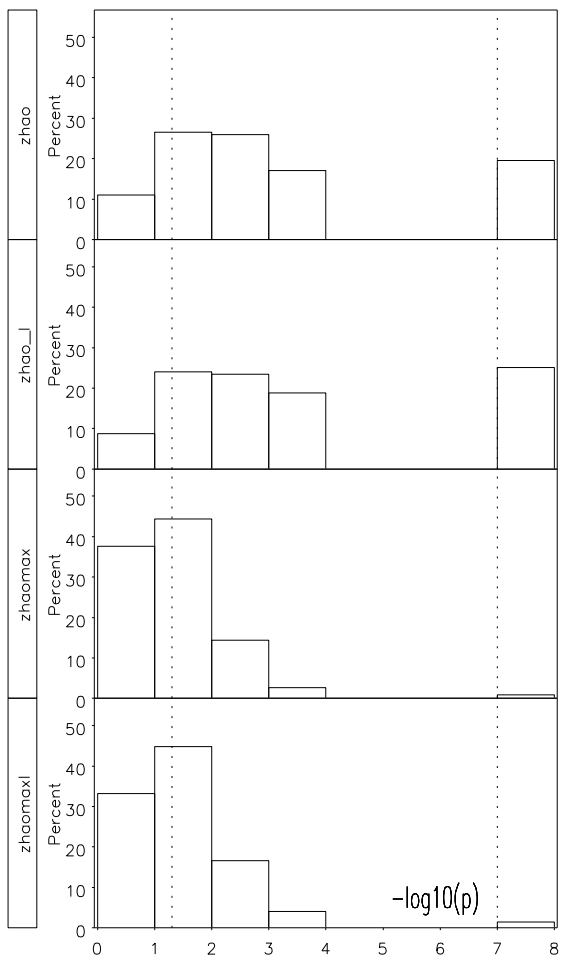
### Markerkombination 15



## AAUU-Familien

### Markerkombination 30

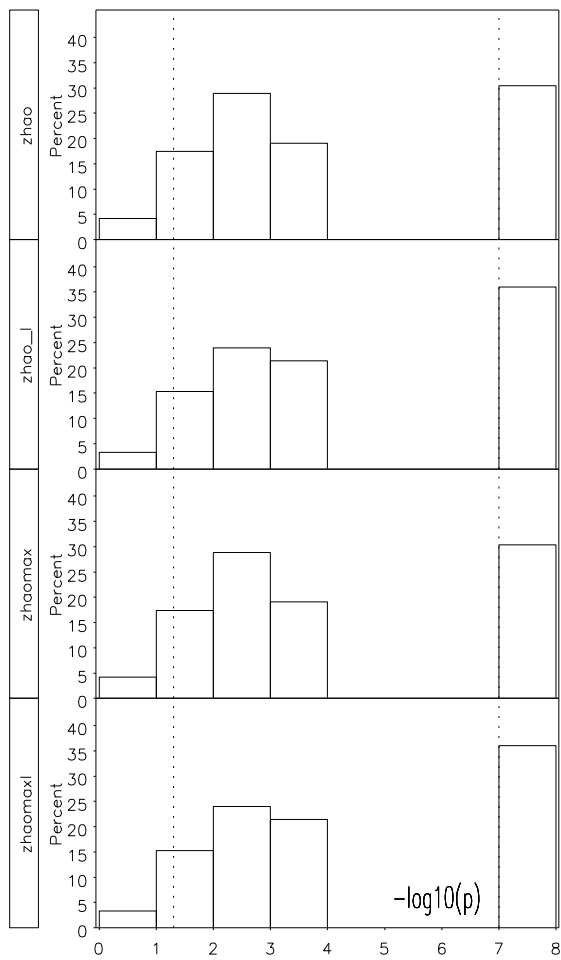
Distribution of p-values: Famhap, ADD,  
KMod 3, 500 AAUU families, marker combination 30



## Multiplikatives Modell KMod 1, AA-Familien

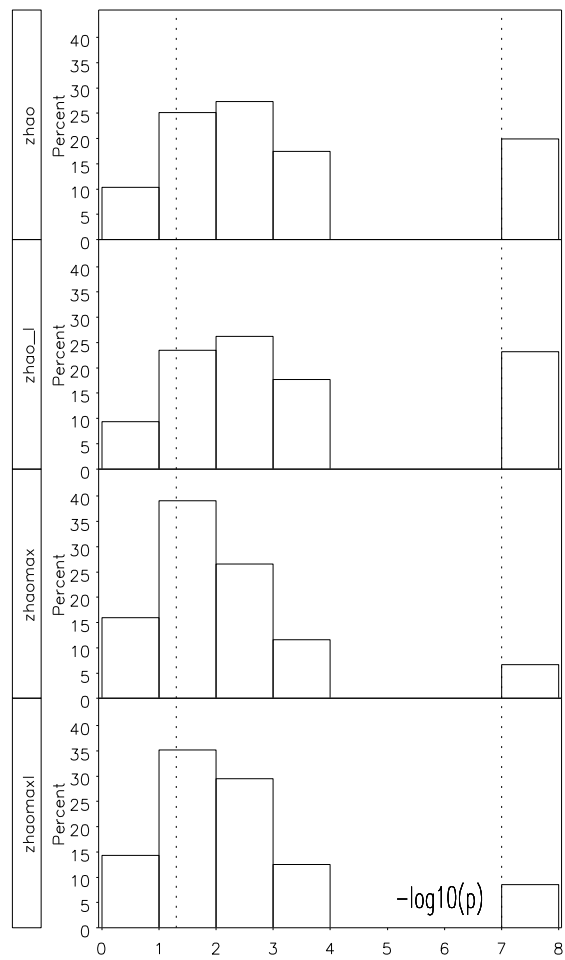
### Markerkombination 4

Distribution of p-values: Famhap, MUL, KMod 1, 200 AA families, marker combination 4



### Markerkombination 20

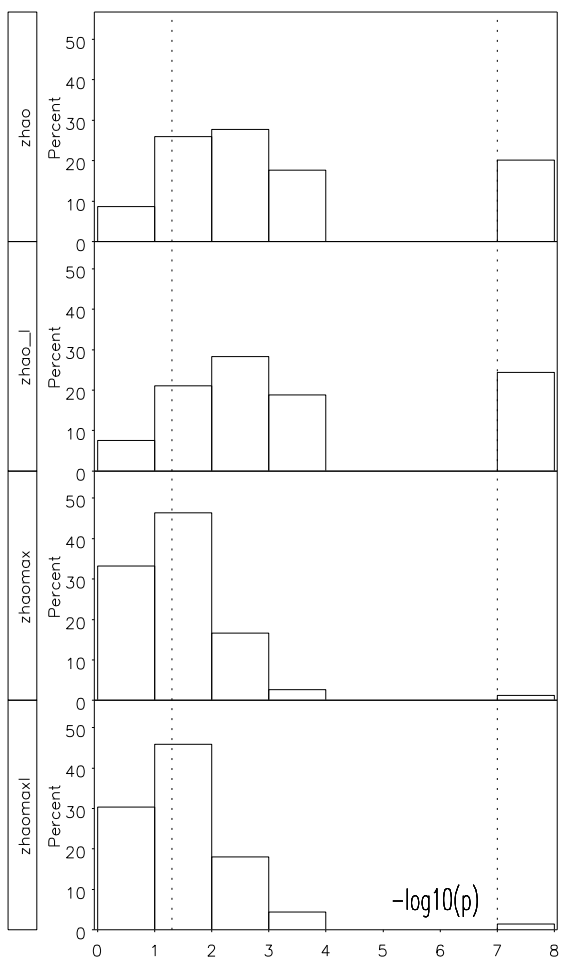
Distribution of p-values: Famhap, MUL, KMod 1, 200 AA families, marker combination 20



## AA-Familien

### Markerkombination 30

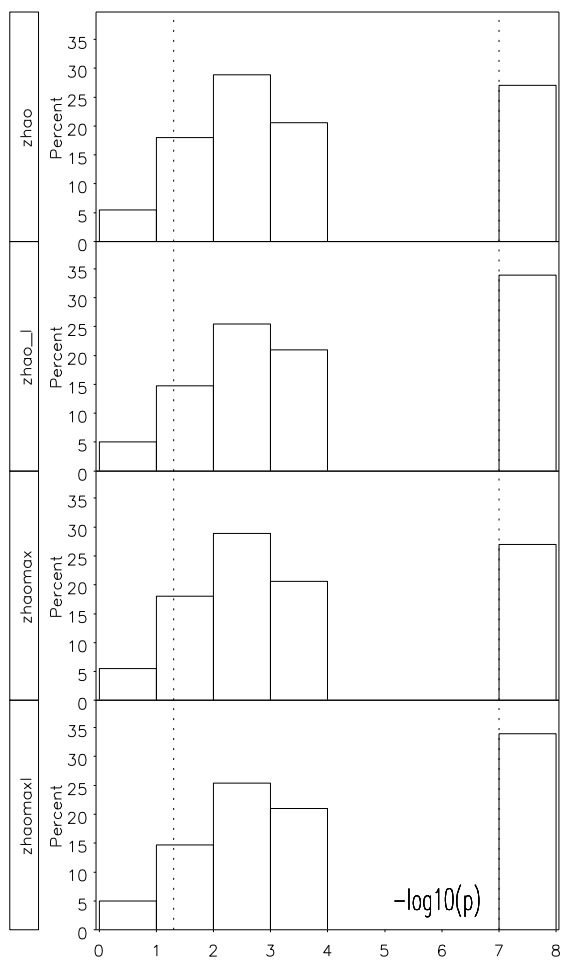
Distribution of p-values: Famhap, ADD,  
KMod 3, 500 AA families, marker combination 30



## AAUU-Familien

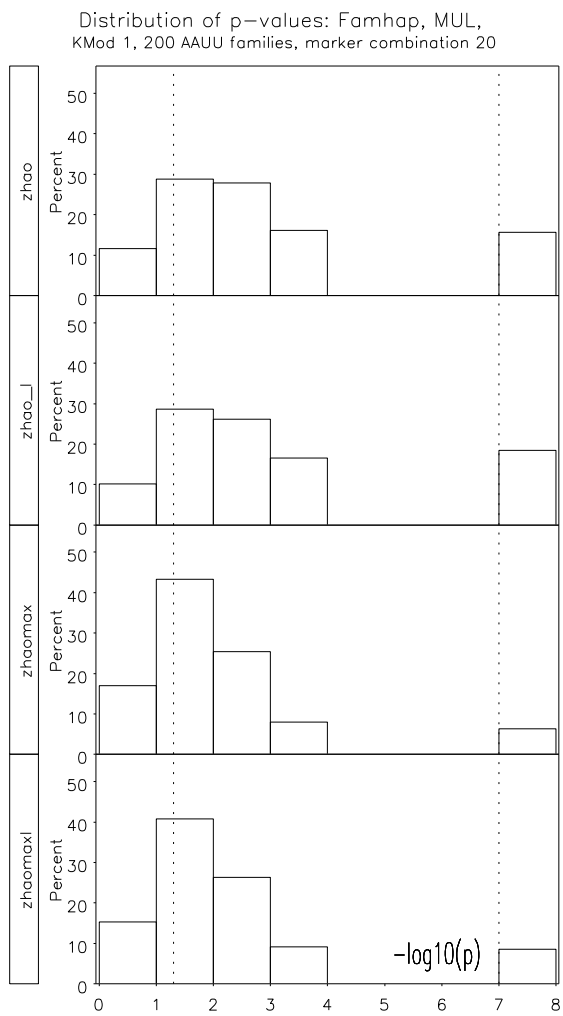
### Markerkombination 4

Distribution of p-values: Famhap, MUL,  
KMod 1, 200 AAUU families, marker combination 4

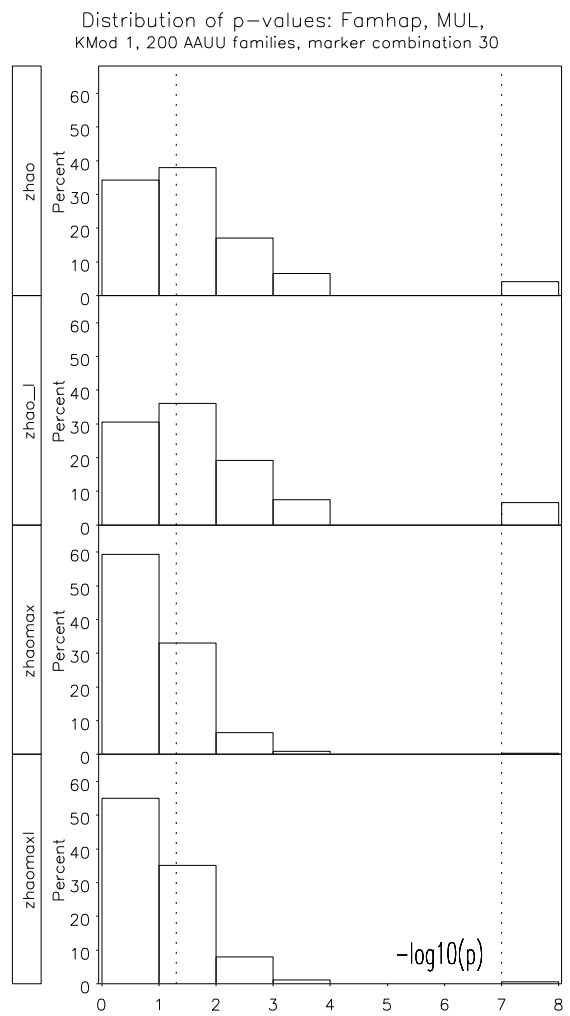


## AAU-Familien

### Markerkombination 20



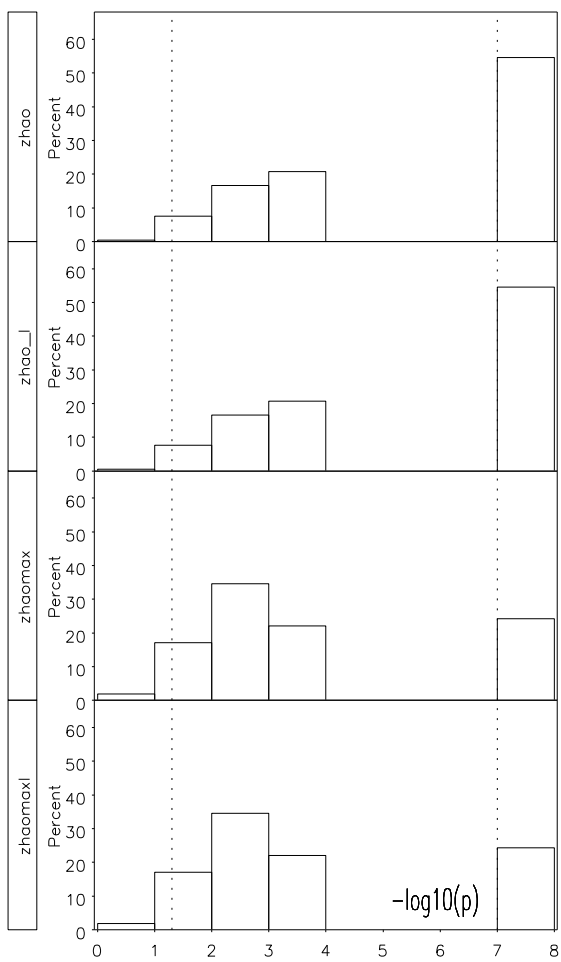
### Markerkombination 30



## Rezessives Modell KMod 2, Trios

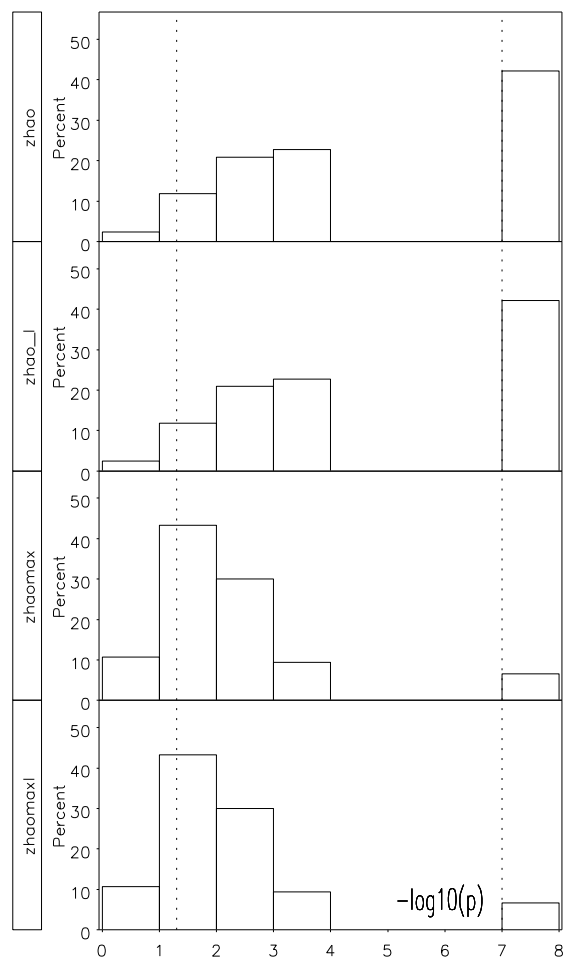
### Markerkombination 12

Distribution of p-values: Famhap, REZ,  
KMod 2, 6000 trios, marker combination 12



### Markerkombination 14

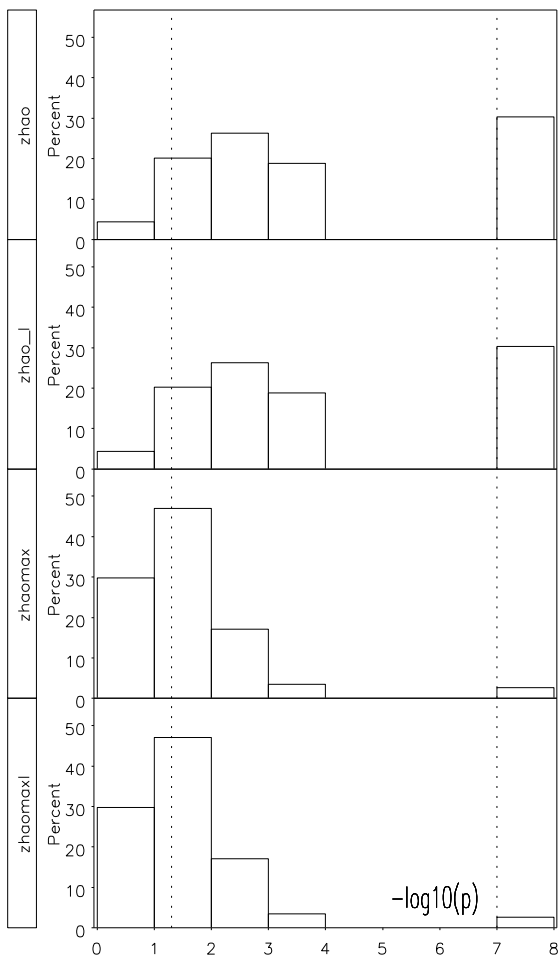
Distribution of p-values: Famhap, REZ,  
KMod 2, 6000 trios, marker combination 14



### Trios

#### Markerkombination 30

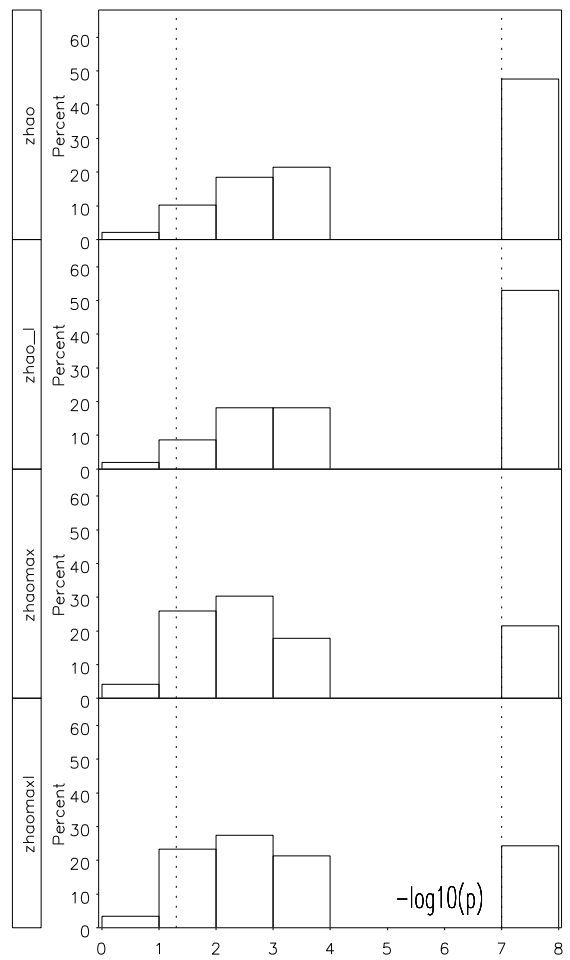
Distribution of p-values: Famhap, REZ, KMod 2, 6000 trios, marker combination 30



### AA-Familien

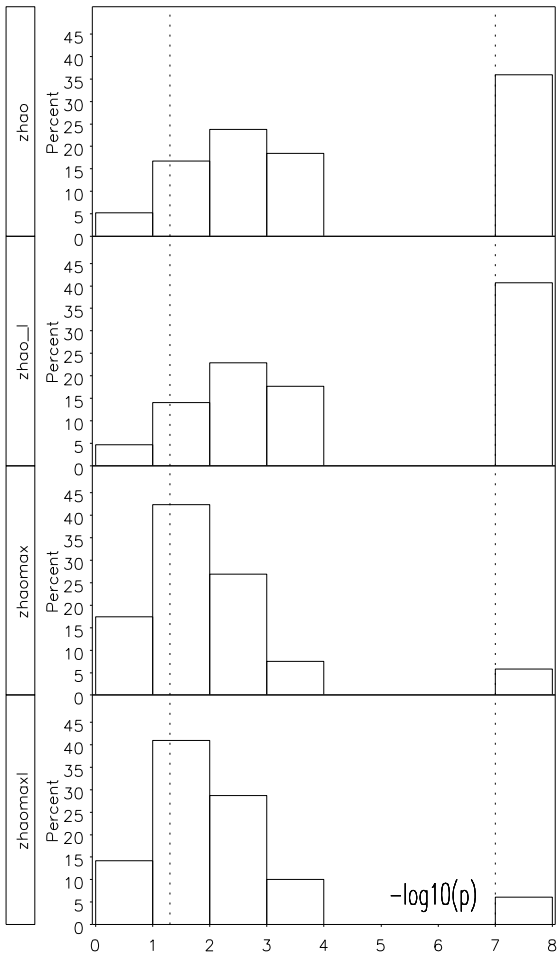
#### Markerkombination 12

Distribution of p-values: Famhap, REZ, KMod 2, 900 AA families, marker combination 12



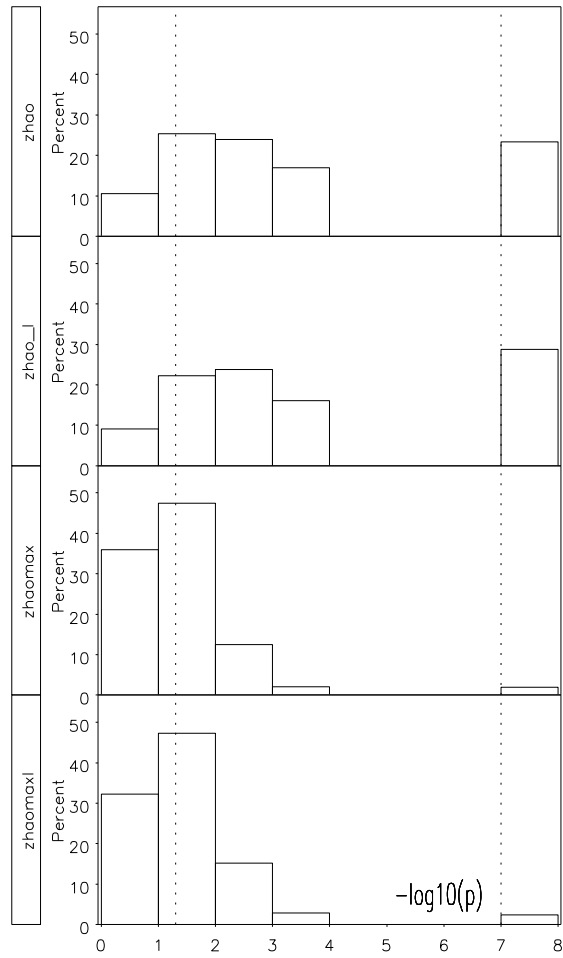
### Markerkombination 14

Distribution of p-values: Famhap, REZ, KMod 2, 900 AA families, marker combination 14



### Markerkombination 30

Distribution of p-values: Famhap, REZ, KMod 2, 900 AA families, marker combination 30

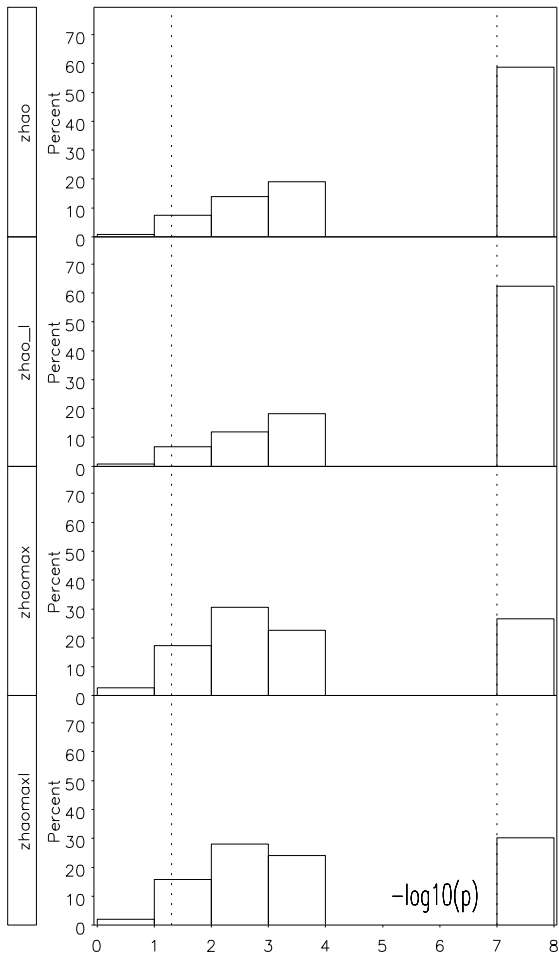




# AAU-Familien

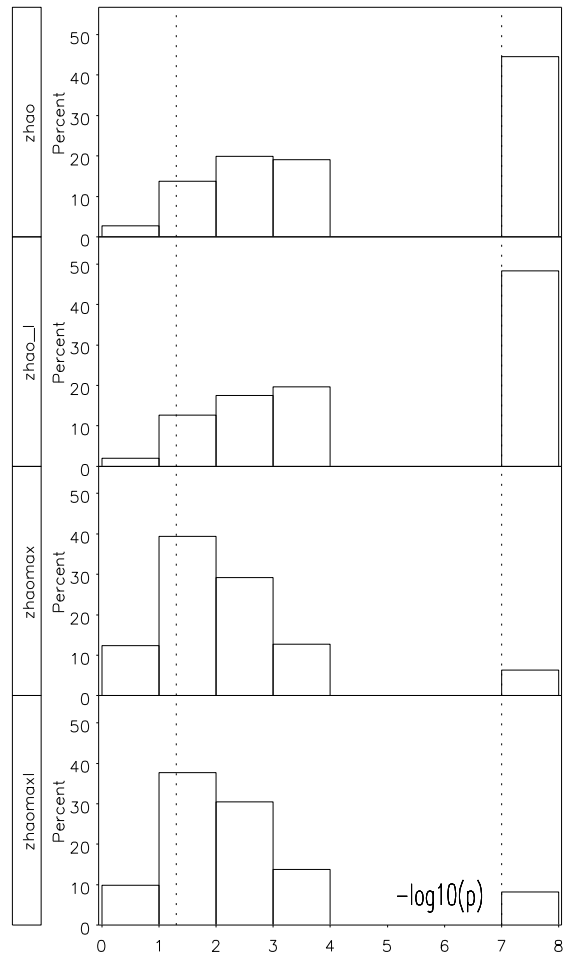
## Markerkombination 12

Distribution of p-values: Famhap, REZ, KMod 2, 900 AAU families, marker combination 12



## Markerkombination 14

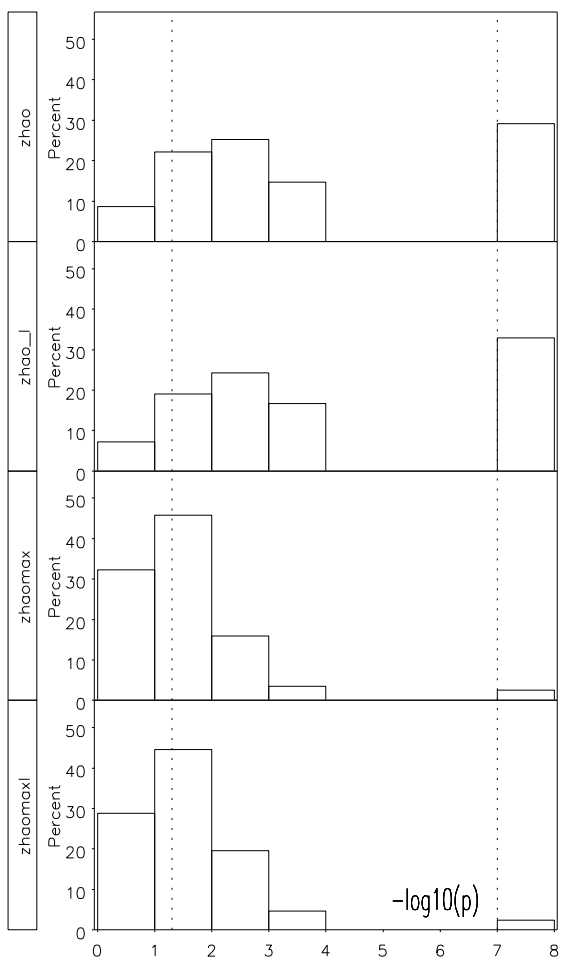
Distribution of p-values: Famhap, REZ, KMod 2, 900 AAU families, marker combination 14



## AAUU-Familien

### Markerkombination 30

Distribution of p-values: Famhap, REZ,  
KMod 2, 900 AAUU families, marker combination 30

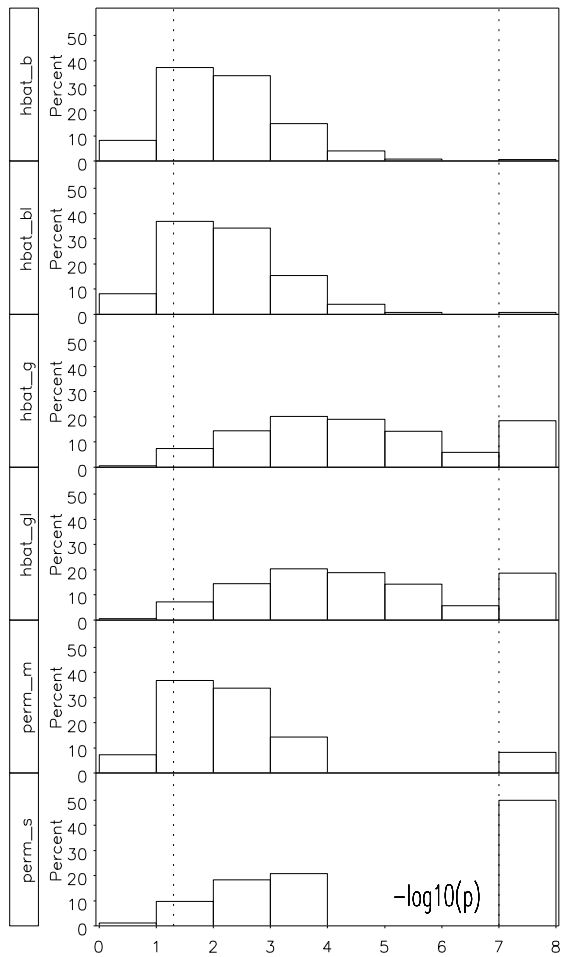


# FBAT

## Additives Modell KMod 3, Trios

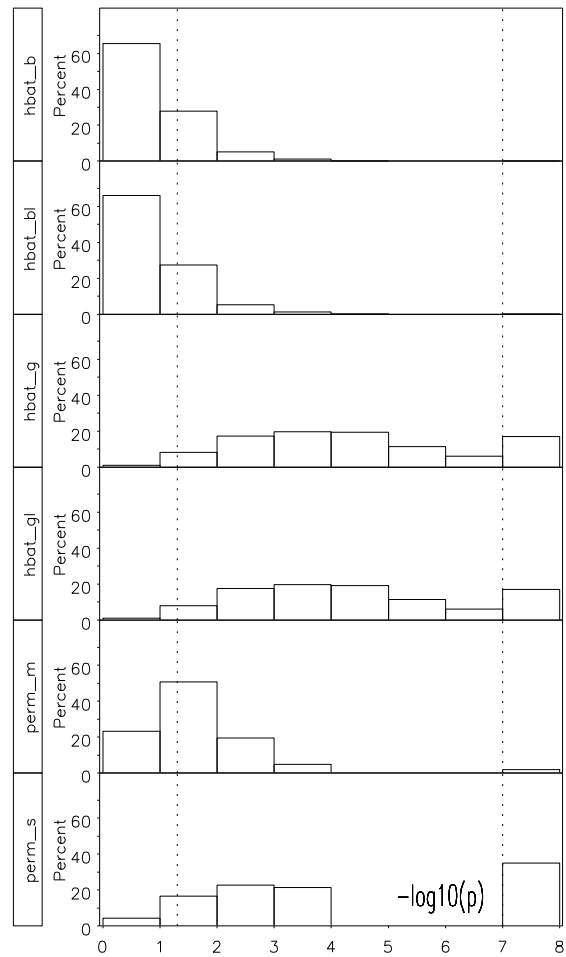
### Markerkombination 14

Distribution of p-values: FBAT, ADD, KMod 3, 2500 trios, marker combination 14



### Markerkombination 15

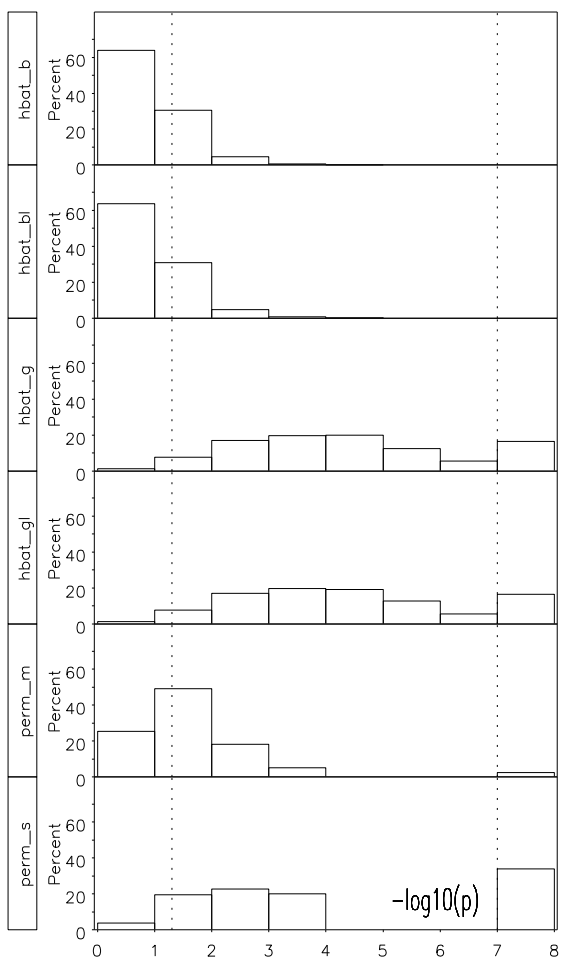
Distribution of p-values: FBAT, ADD, KMod 3, 2500 trios, marker combination 15



## Trios

### Markerkombination 30

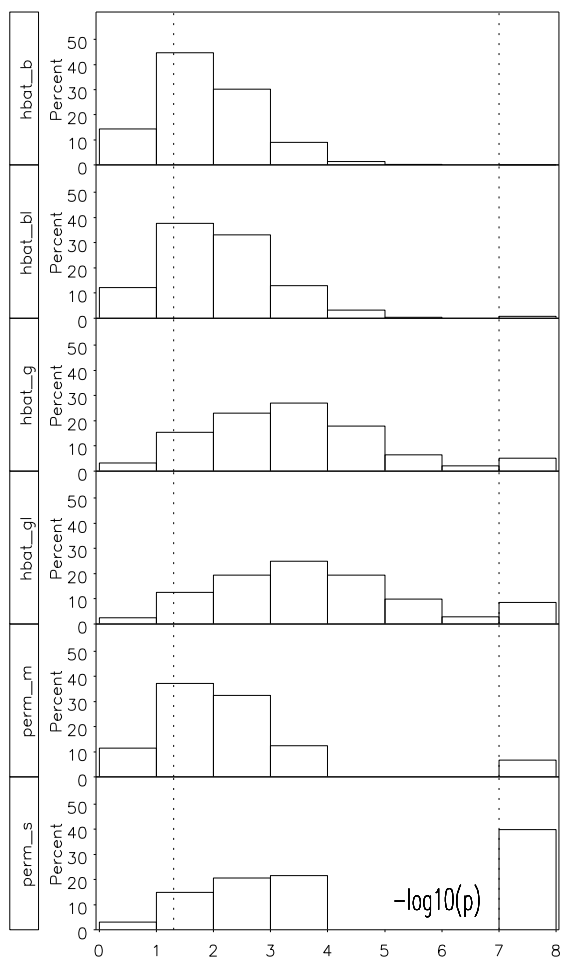
Distribution of p-values: FBAT, ADD, KMod 3, 2500 trios, marker combination 30



## AA-Familien

### Markerkombination 14

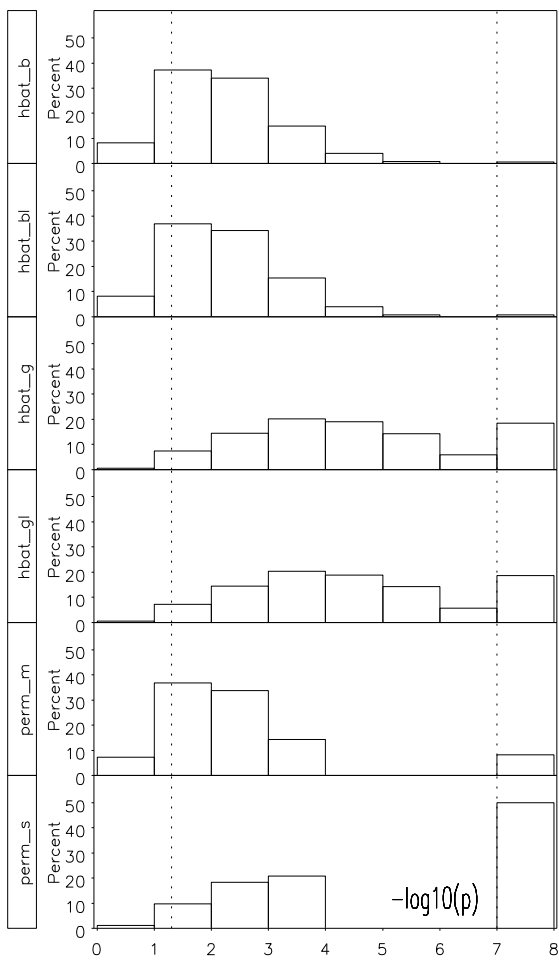
Distribution of p-values: FBAT, ADD, KMod 3, 500 AA families, marker combination 14



## AA-Familien

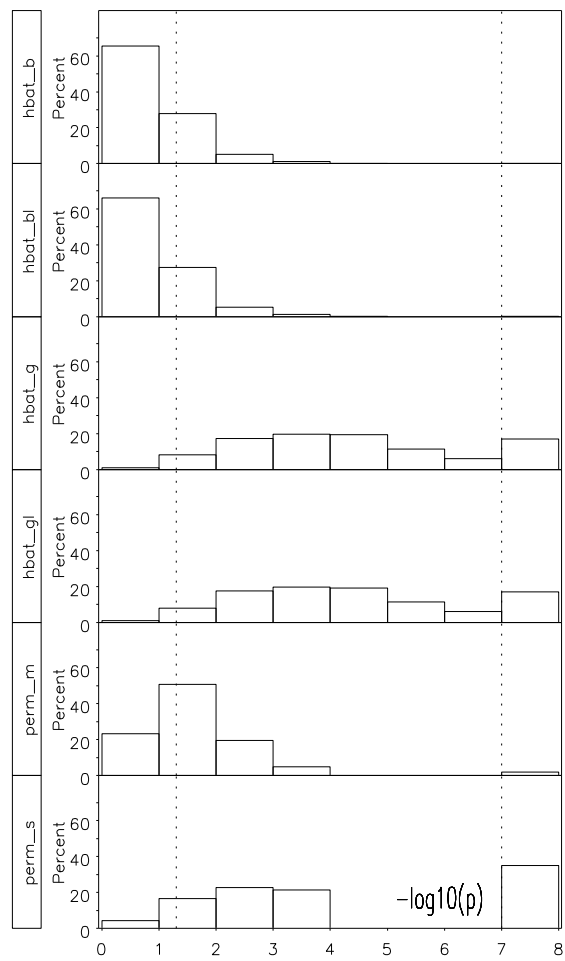
### Markerkombination 15

Distribution of p-values: FBAT, ADD,  
KMod 3, 2500 trios, marker combination 14



### Markerkombination 30

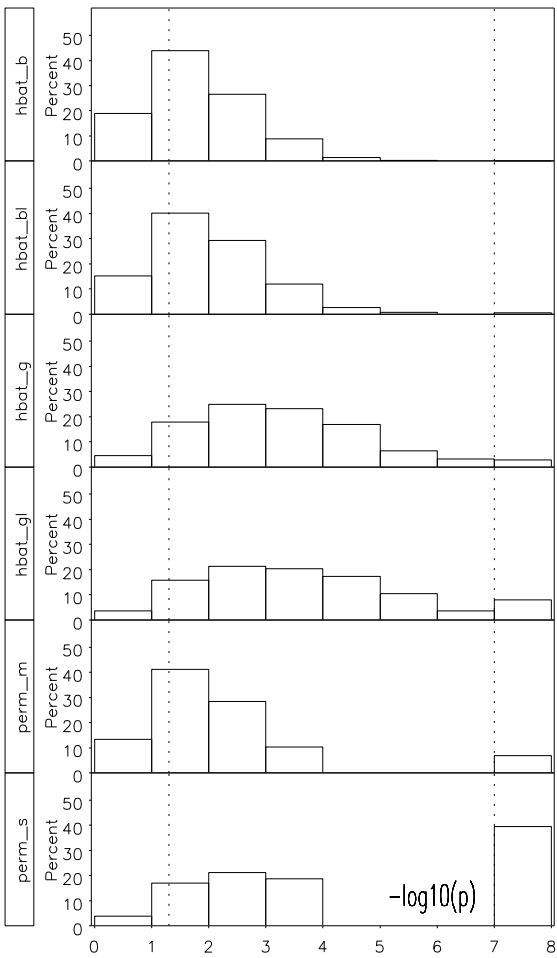
Distribution of p-values: FBAT, ADD,  
KMod 3, 2500 trios, marker combination 15



# AAUU-Familien

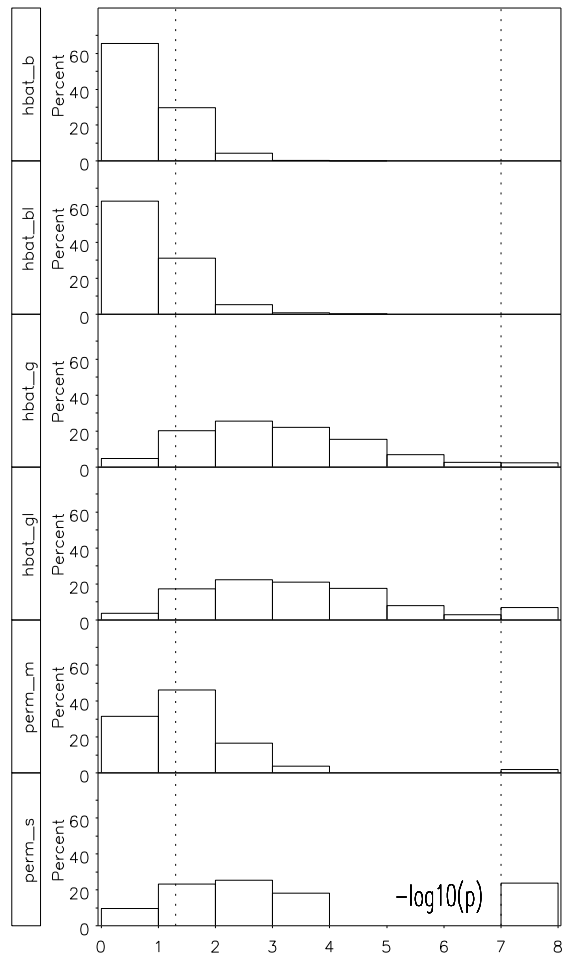
## Markerkombination 14

Distribution of p-values: FBAT, ADD,  
KMod 3, 500 AAUU families, marker combination 14



## Markerkombination 15

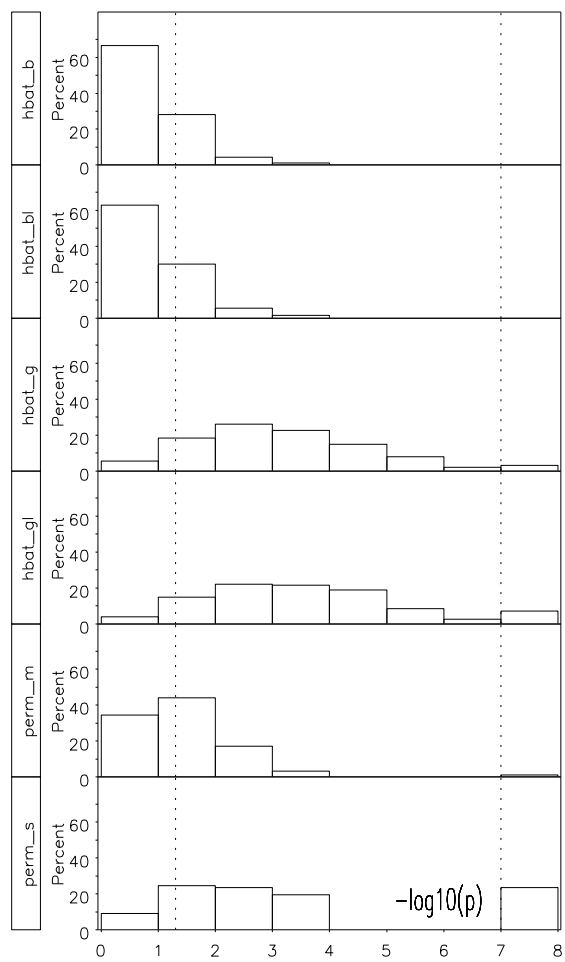
Distribution of p-values: FBAT, ADD,  
KMod 3, 500 AAUU families, marker combination 15



## AAUU-Familien

### Markerkombination 30

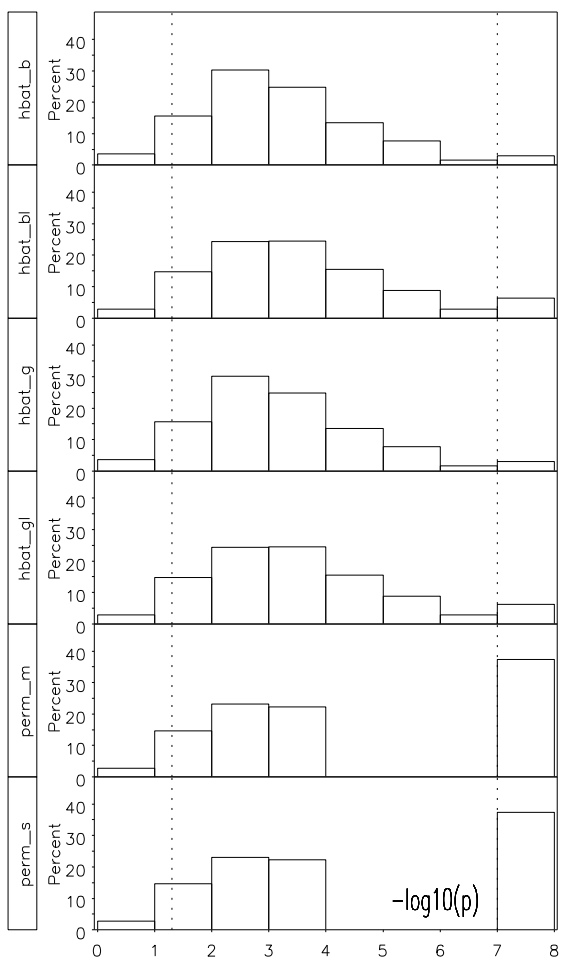
Distribution of p-values: FBAT, ADD,  
KMod 3, 500 AAUU families, marker combination 30



## Multiplikatives Modell KMod 1, AA-Familien

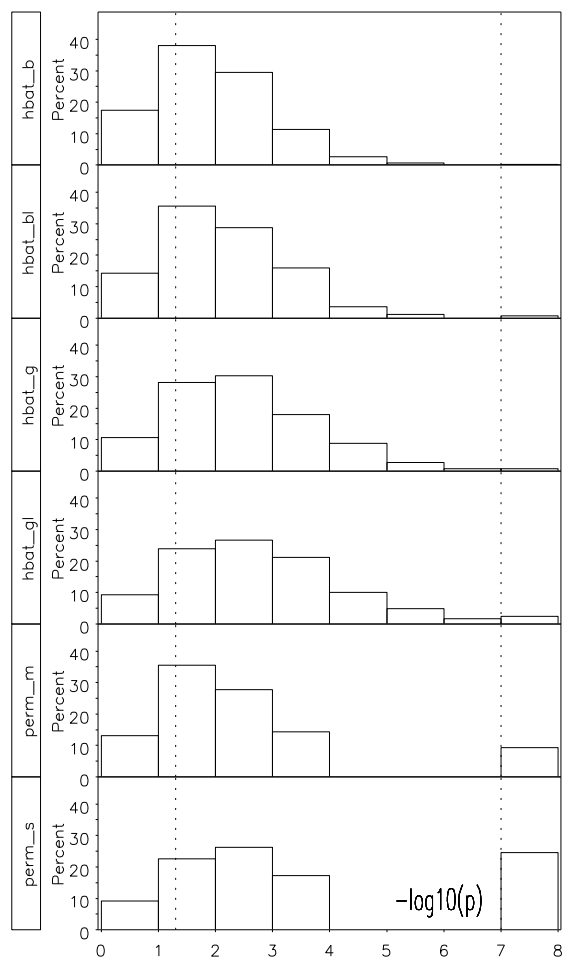
### Markerkombination 4

Distribution of p-values: FBAT, MUL,  
KMod 1, 200 AA families, marker combination 4



### Markerkombination 20

Distribution of p-values: FBAT, MUL,  
KMod 1, 200 AA families, marker combination 20

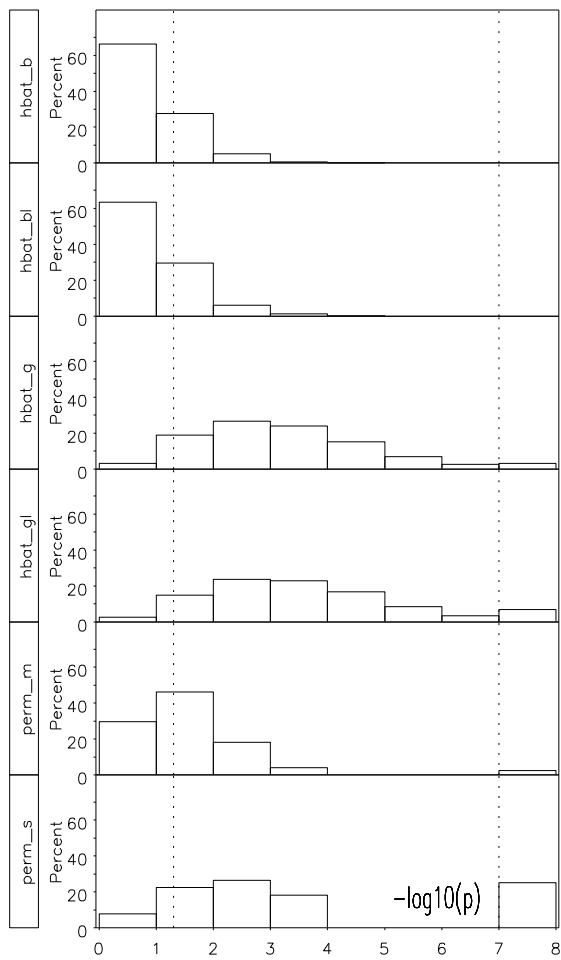




## AA-Familien

### Markerkombination 30

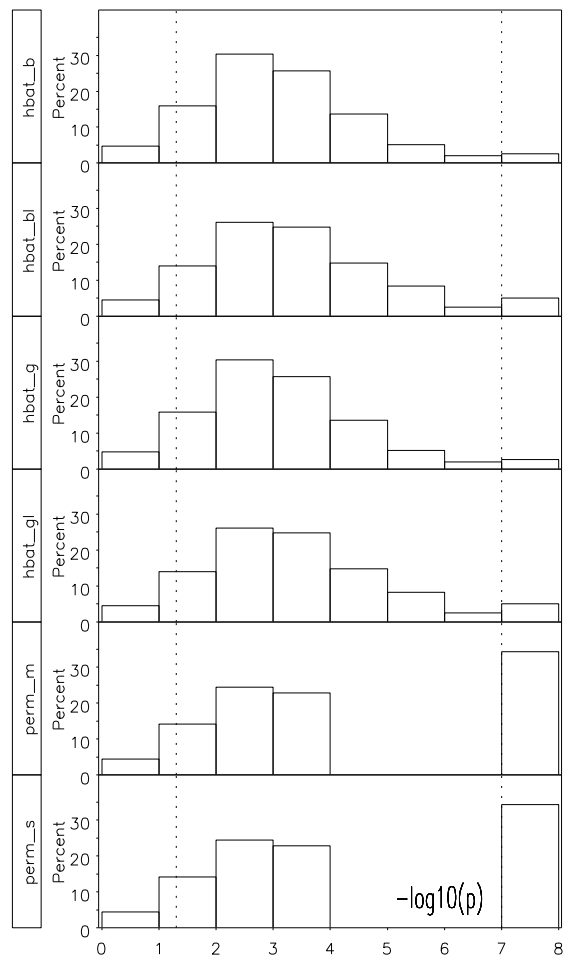
Distribution of p-values: FBAT, ADD,  
KMod 3, 500 AA families, marker combination 30



## AAUU-Familien

### Markerkombination 4

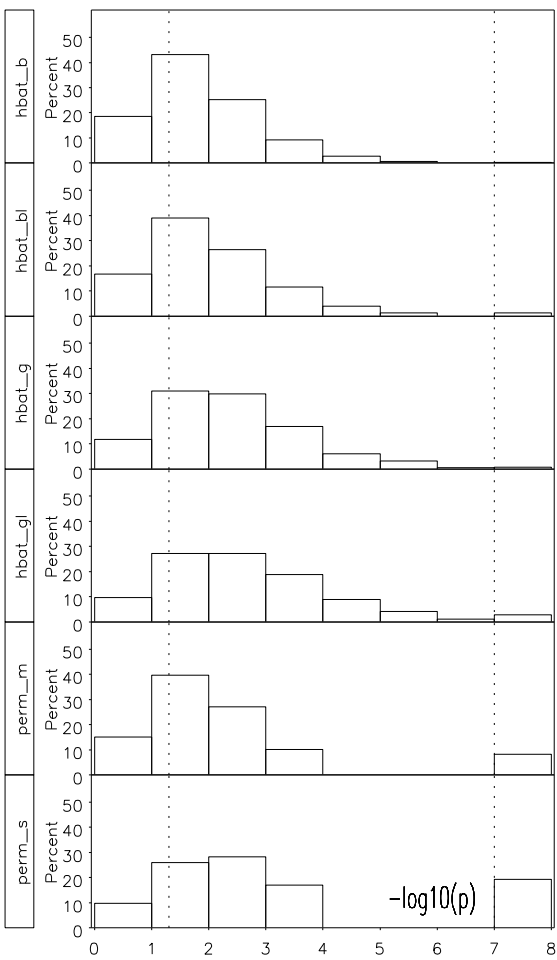
Distribution of p-values: FBAT, MUL,  
KMod 1, 200 AAUU families, marker combination 4



# AAUU-Familien

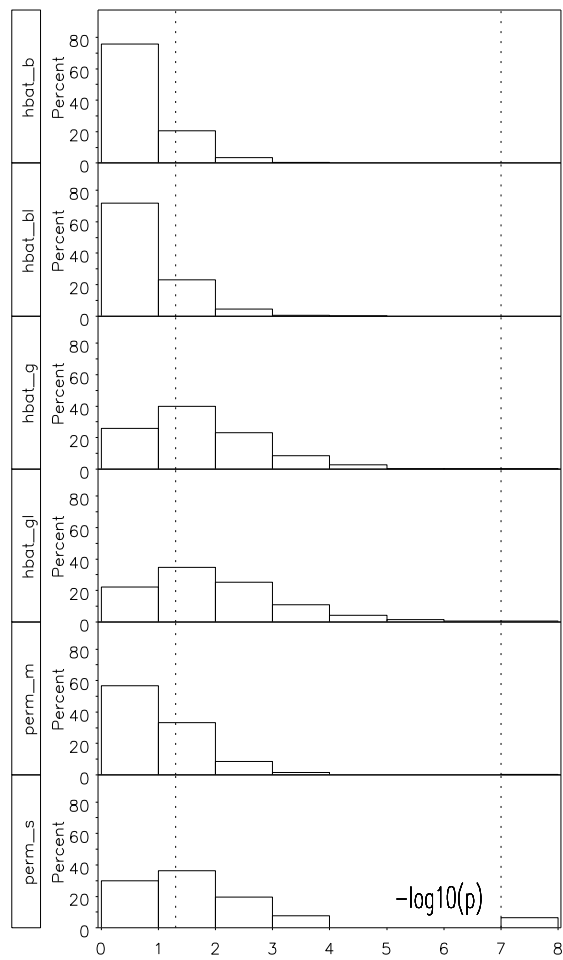
## Markerkombination 20

Distribution of p-values: FBAT, MUL,  
KMod 1, 200 AAUU families, marker combination 20



## Markerkombination 30

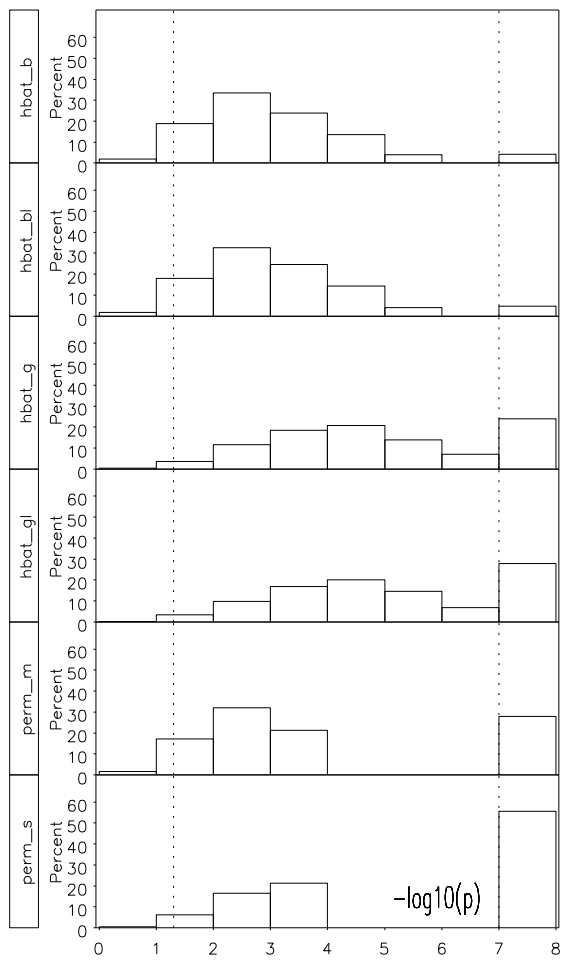
Distribution of p-values: FBAT, MUL,  
KMod 1, 200 AAUU families, marker combination 30



## Rezessives Modell KMod 2, Trios

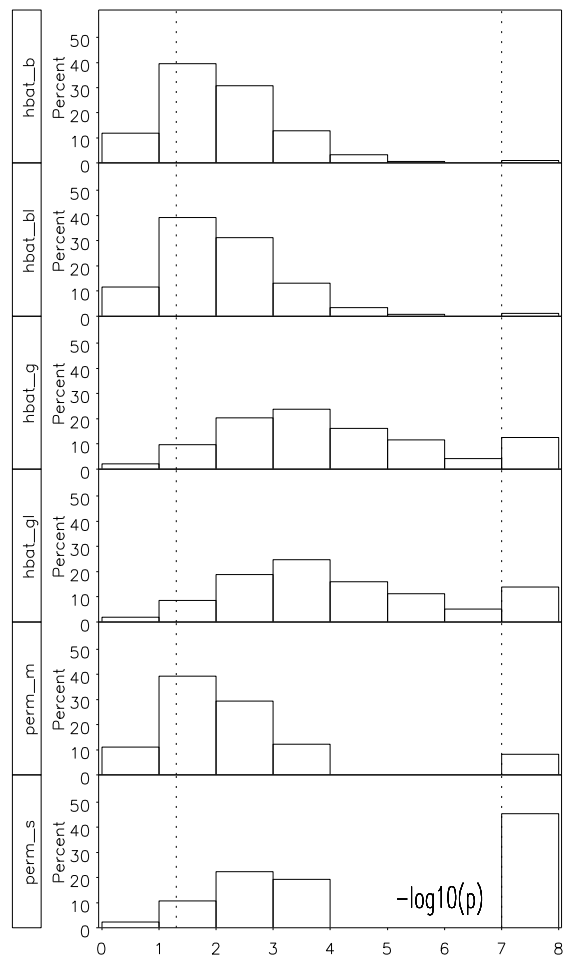
### Markerkombination 12

Distribution of p-values: FBAT, REZ,  
KMod 2, 6000 trios, marker combination 12



### Markerkombination 14

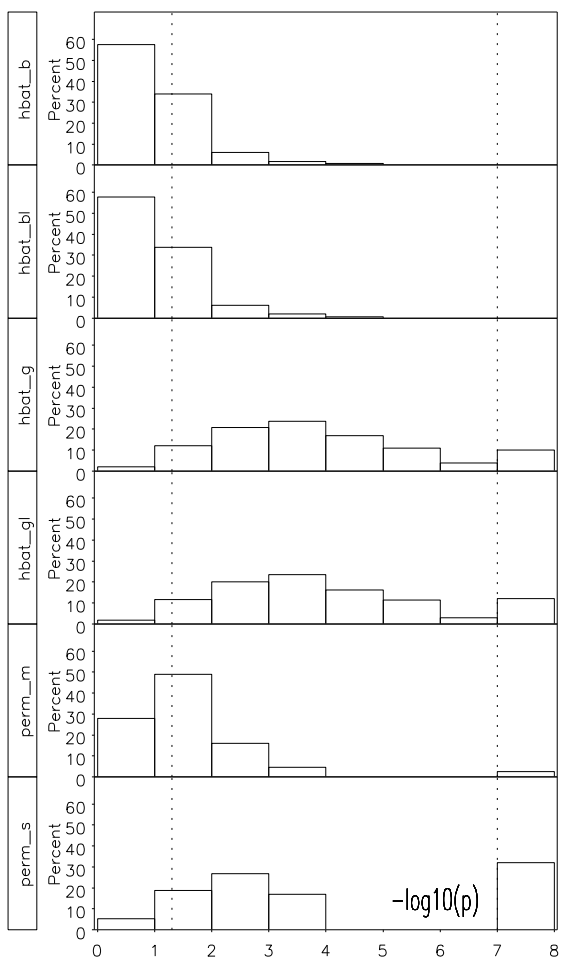
Distribution of p-values: FBAT, REZ,  
KMod 2, 6000 trios, marker combination 14



## Trios

### Markerkombination 30

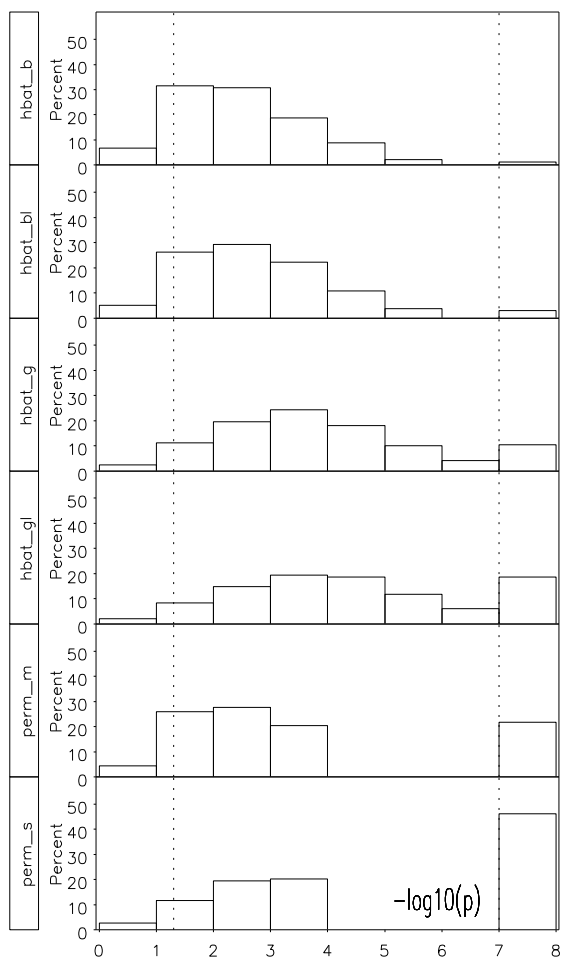
Distribution of p-values: FBAT, REZ,  
KMod 2, 6000 trios, marker combination 30



## AA-Familien

### Markerkombination 12

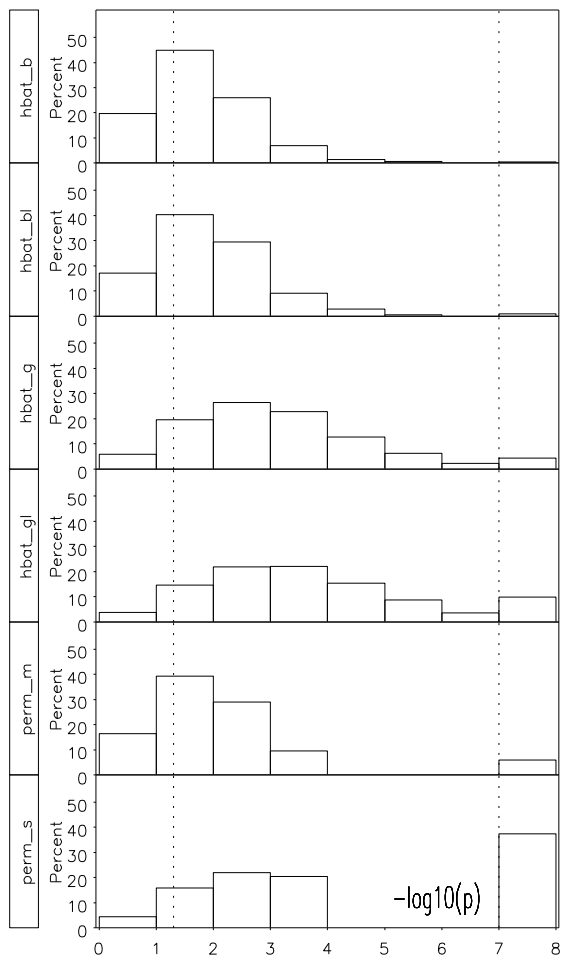
Distribution of p-values: FBAT, REZ,  
KMod 2, 900 AA families, marker combination 12



## AA-Familien

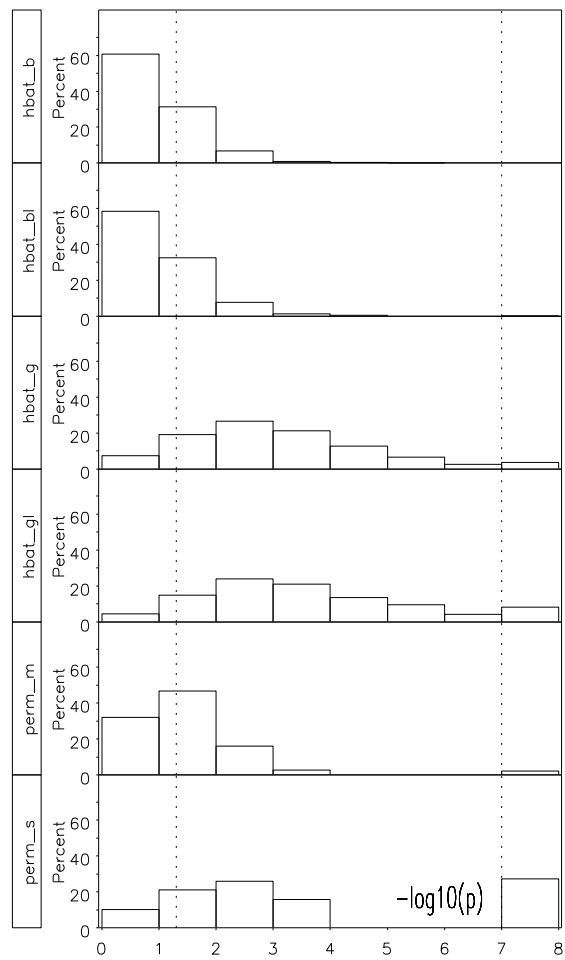
### Markerkombination 14

Distribution of p-values: FBAT, REZ,  
KMod 2, 900 AA families, marker combination 14



### Markerkombination 30

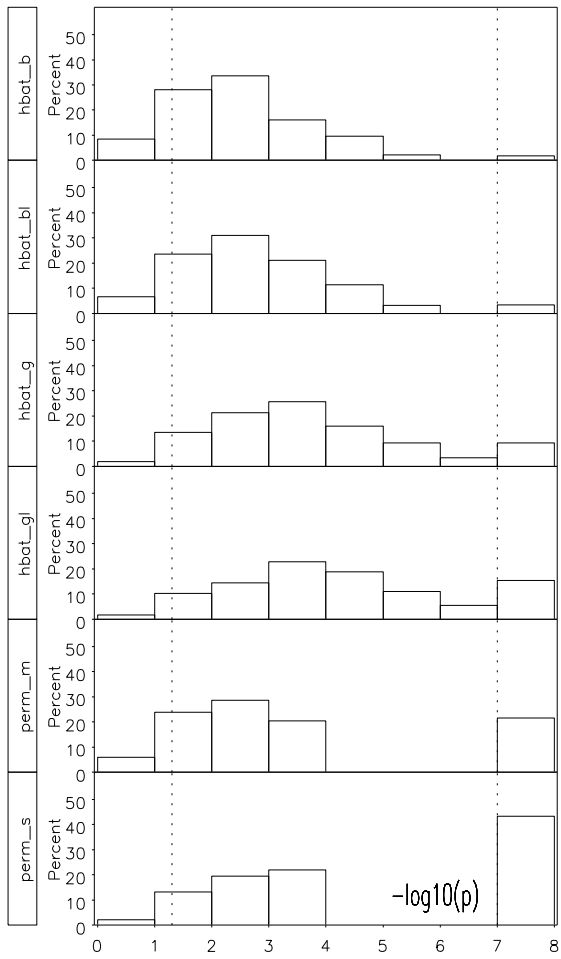
Distribution of p-values: FBAT, REZ,  
KMod 2, 900 AA families, marker combination 30



# AAU-Familien

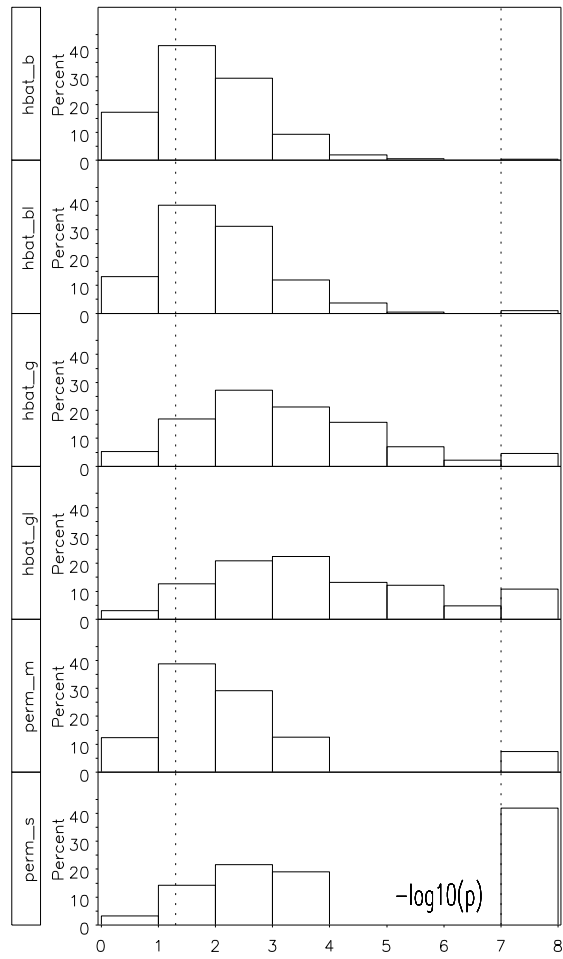
## Markerkombination 12

Distribution of p-values: FBAT, REZ,  
KMod 2, 900 AAU families, marker combination 12



## Markerkombination 14

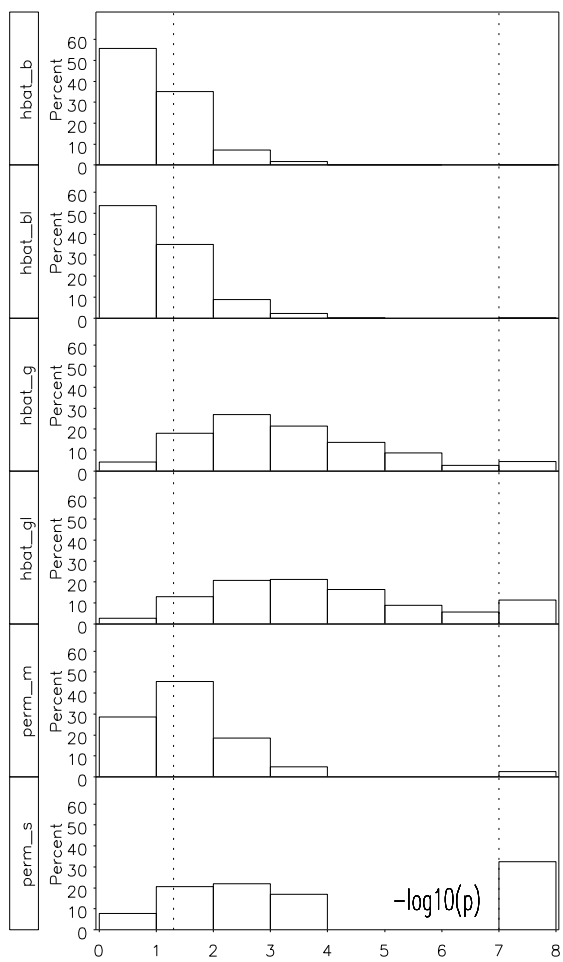
Distribution of p-values: FBAT, REZ,  
KMod 2, 900 AAU families, marker combination 14



## AAUU-Familien

### Markerkombination 30

Distribution of p-values: FBAT, REZ,  
KMod 2, 900 AAUU families, marker combination 30

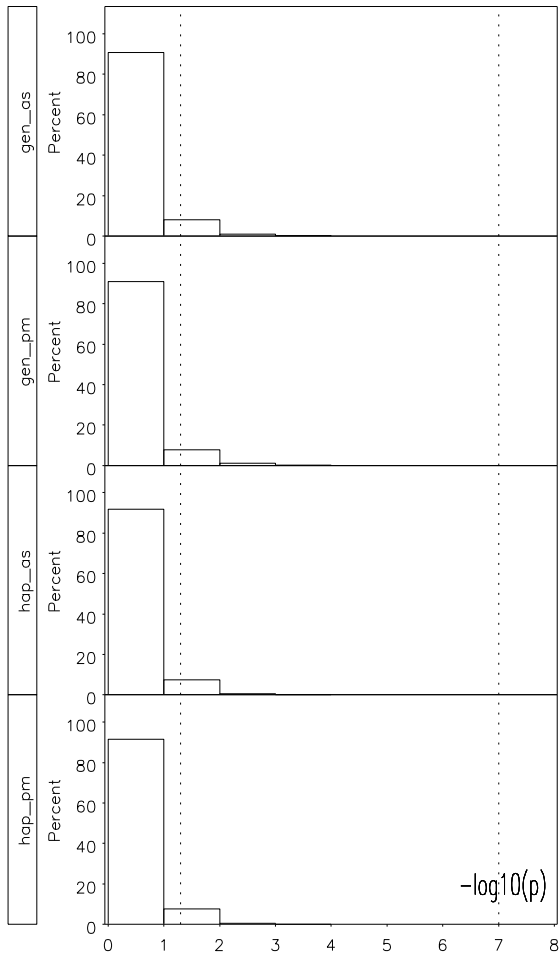


# Hotel\_Fam

## Additives Modell KMod 3, Trios

Markerkombination 30

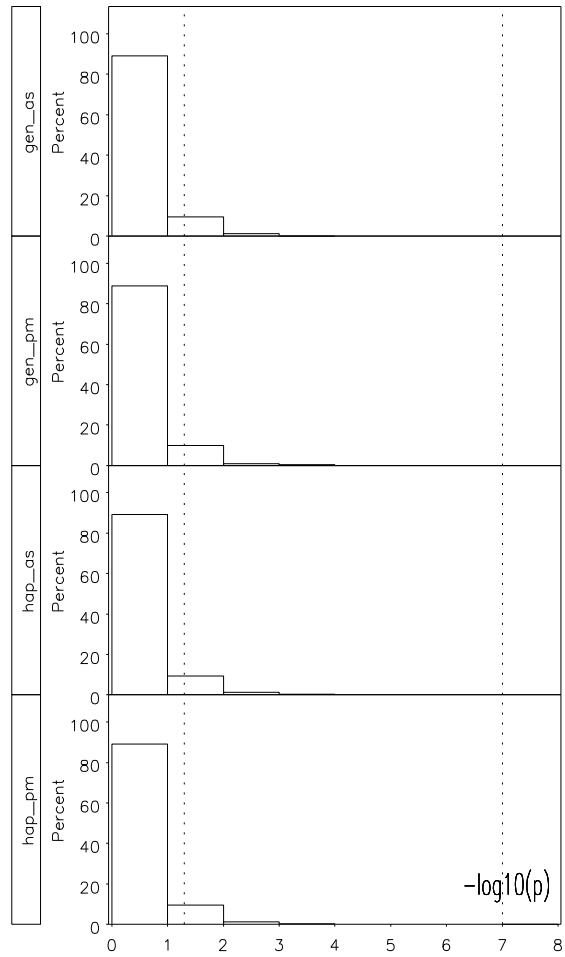
Distribution of p-values: Hotel\_Fam, ADD,  
KMod 3, 2500 trios, marker combination 30



## Rezessives Modell KMod 2, AAUU-Familien

Markerkombination 12

Distribution of p-values: Hotel\_Fam, REZ,  
KMod 2, 900 AAUU families, marker combination 12

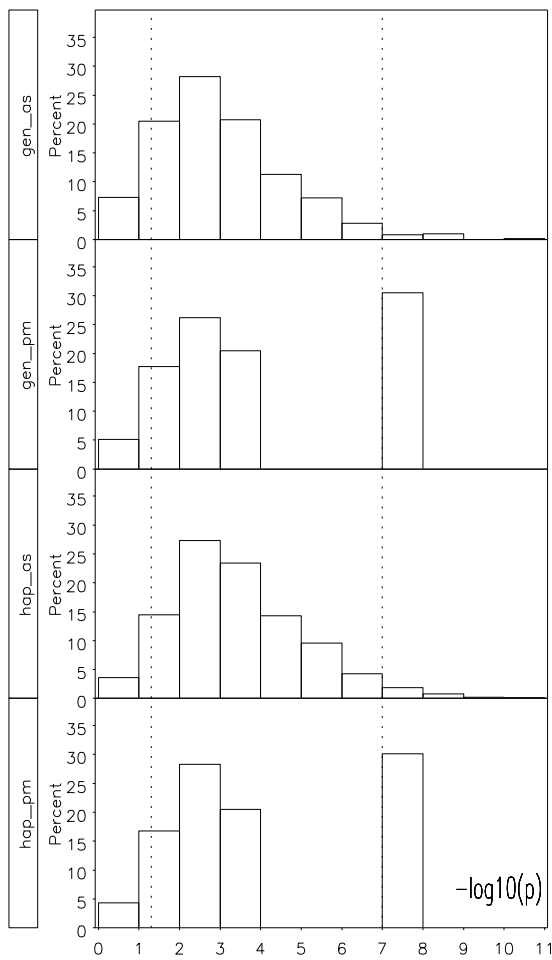




## Multiplikatives Modell KMod 1, AA-Familien

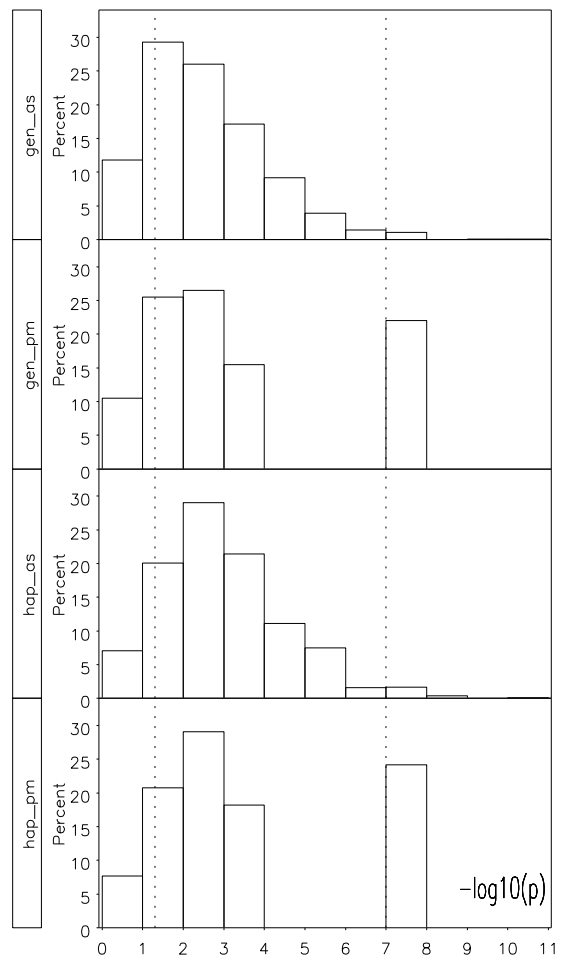
### Markerkombination 4

Distribution of p-values: Hotel\_Fam, MUL, KMod 1, 200 AA families, marker combination 4



### Markerkombination 20

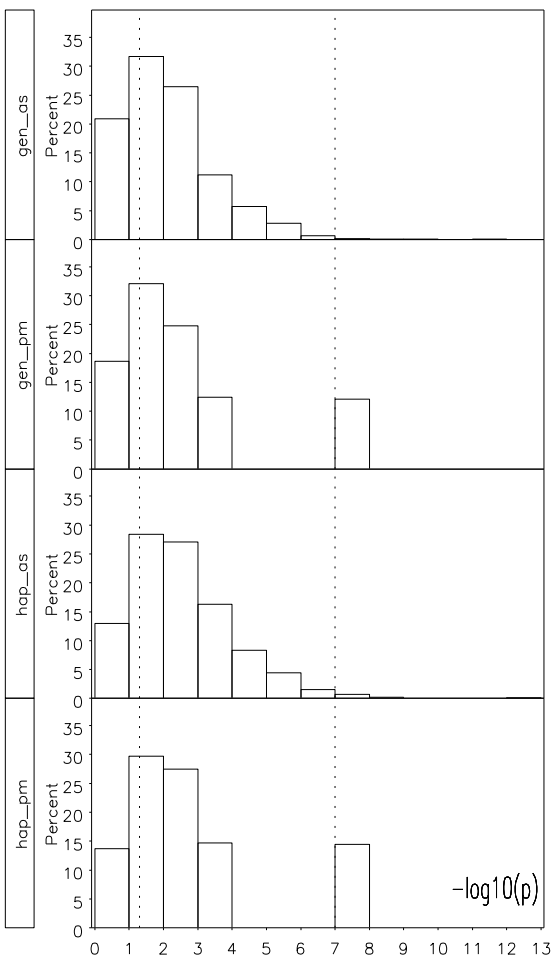
Distribution of p-values: Hotel\_Fam, MUL, KMod 1, 200 AA families, marker combination 20



### AA-Familien

#### Markerkombination 30

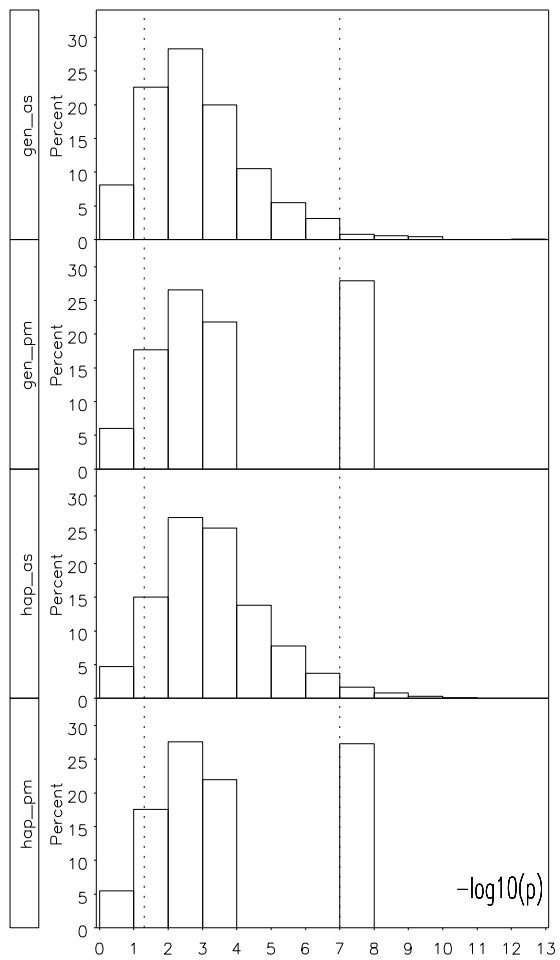
Distribution of p-values: Hotel\_Fam, MUL, KMod 1, 200 AA families, marker combination 30



### AAUU-Familien

#### Markerkombination 4

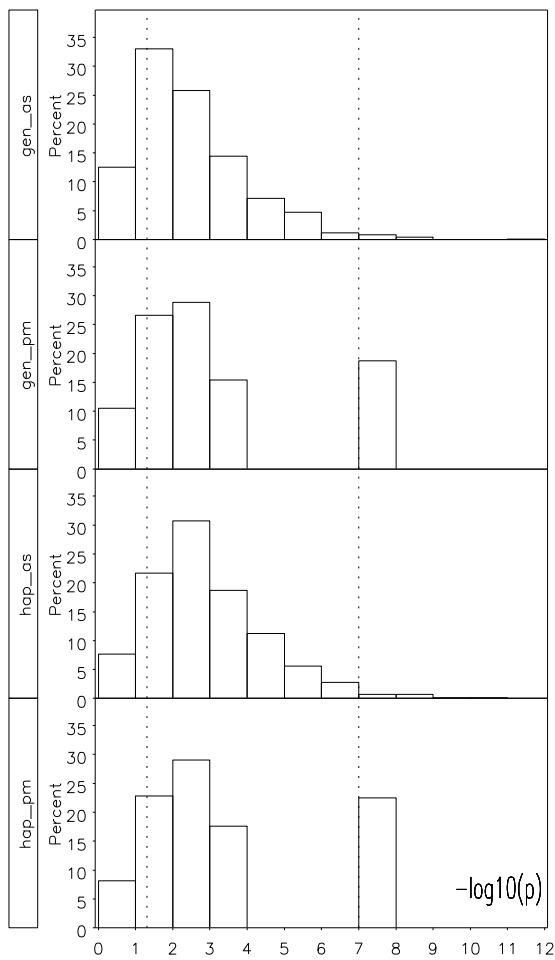
Distribution of p-values: Hotel\_Fam, MUL, KMod 1, 200 AAUU families, marker combination 4



## AAUU-Familien

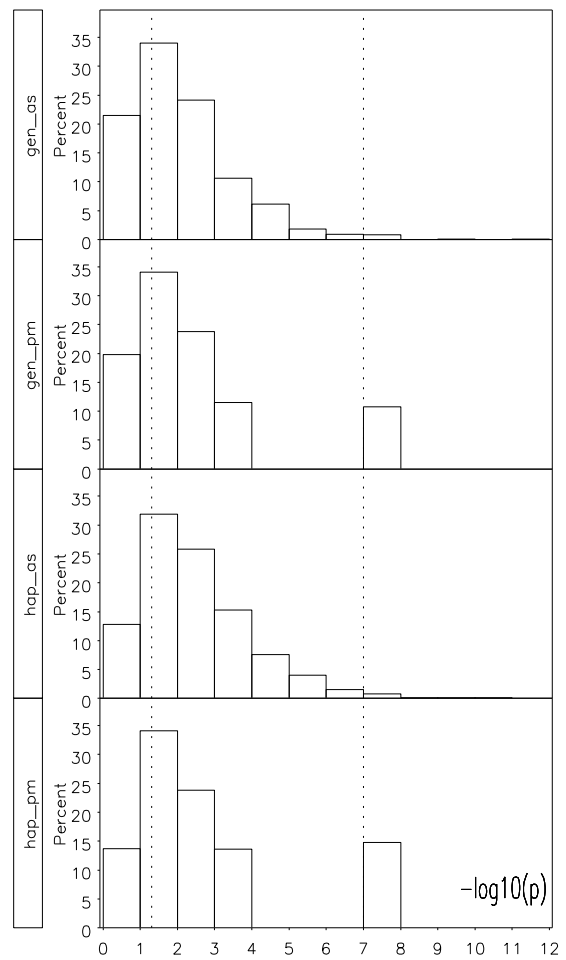
### Markerkombination 20

Distribution of p-values: Hotel\_Fam, MUL, KMod 1, 200 AAUU families, marker combination 20



### Markerkombination 30

Distribution of p-values: Hotel\_Fam, MUL, KMod 1, 200 AAUU families, marker combination 30

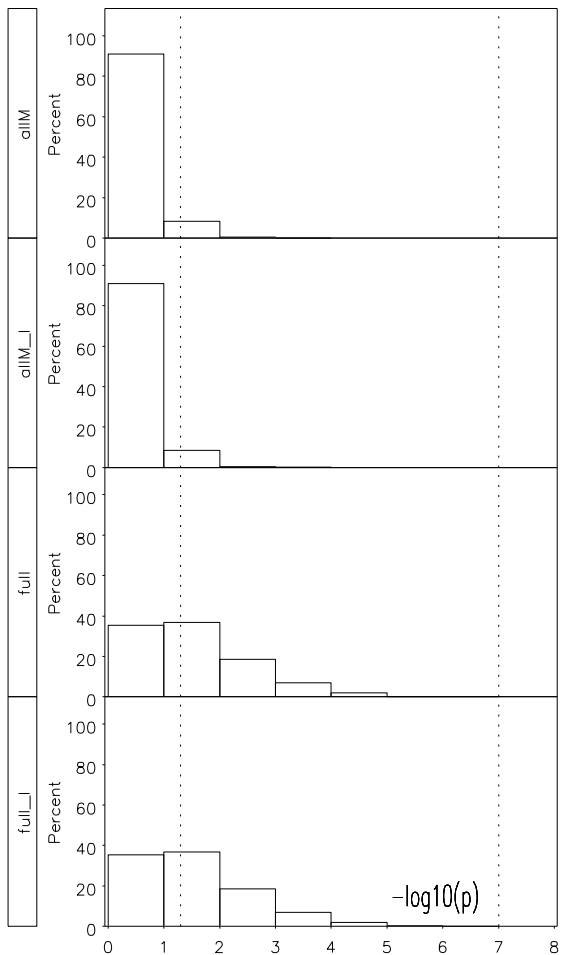


# Unphased

## Additives Modell KMod 3, Trios

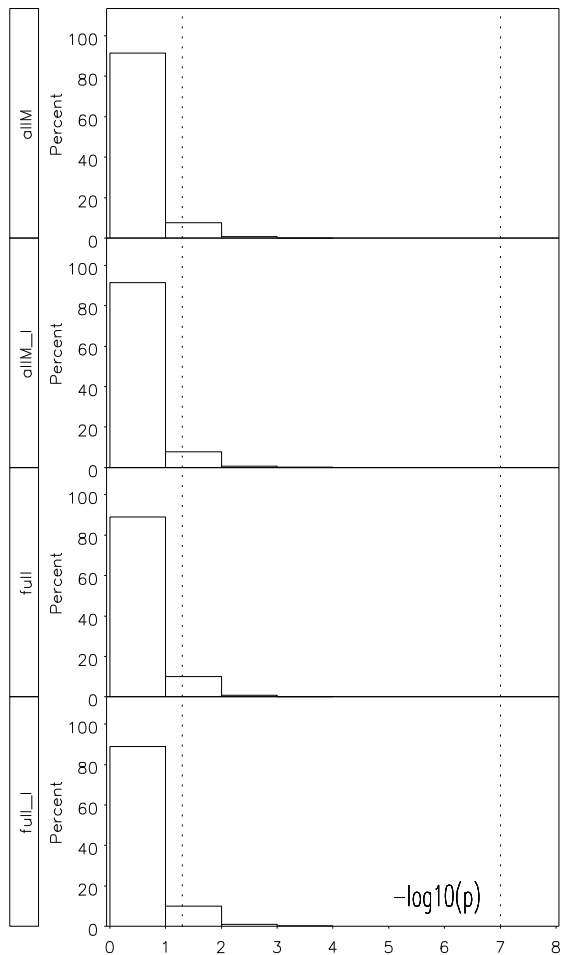
Markerkombination 14

Distribution of p-values: Unphased, ADD, KMod 3, 2500 trios, marker combination 14



Markerkombination 15

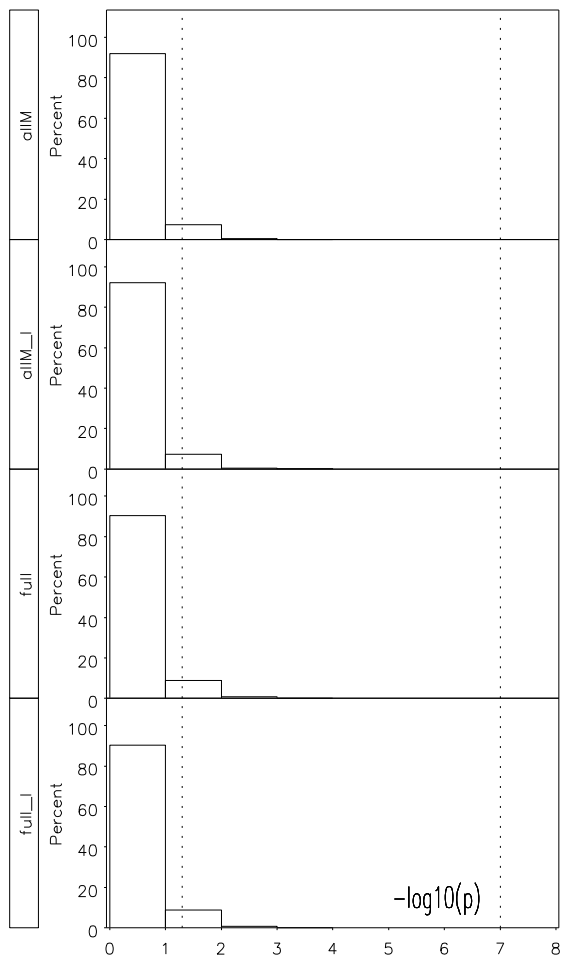
Distribution of p-values: Unphased, ADD, KMod 3, 2500 trios, marker combination 15



### Trios

#### Markerkombination 30

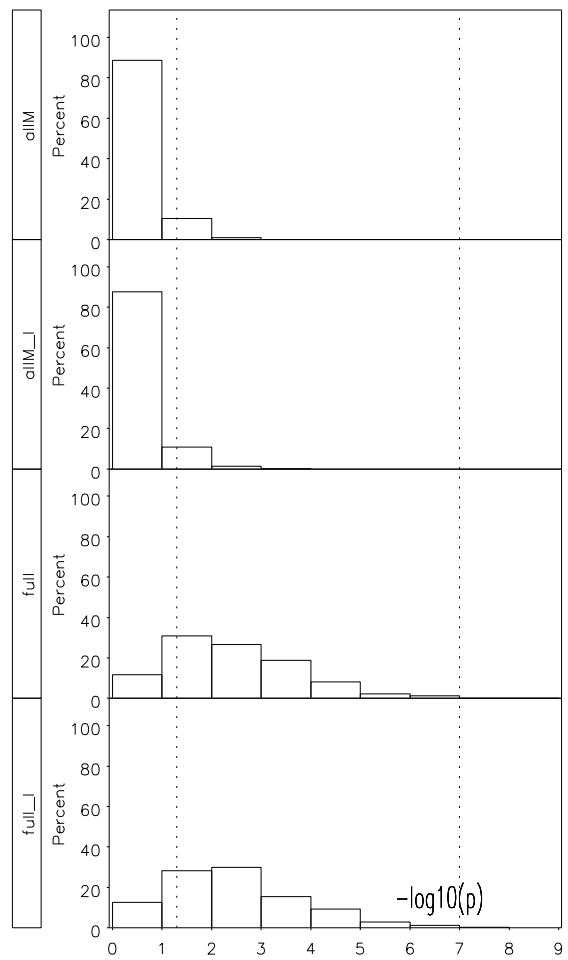
Distribution of p-values: Unphased, ADD, KMod 3, 2500 trios, marker combination 30



### AA-Familien

#### Markerkombination 14

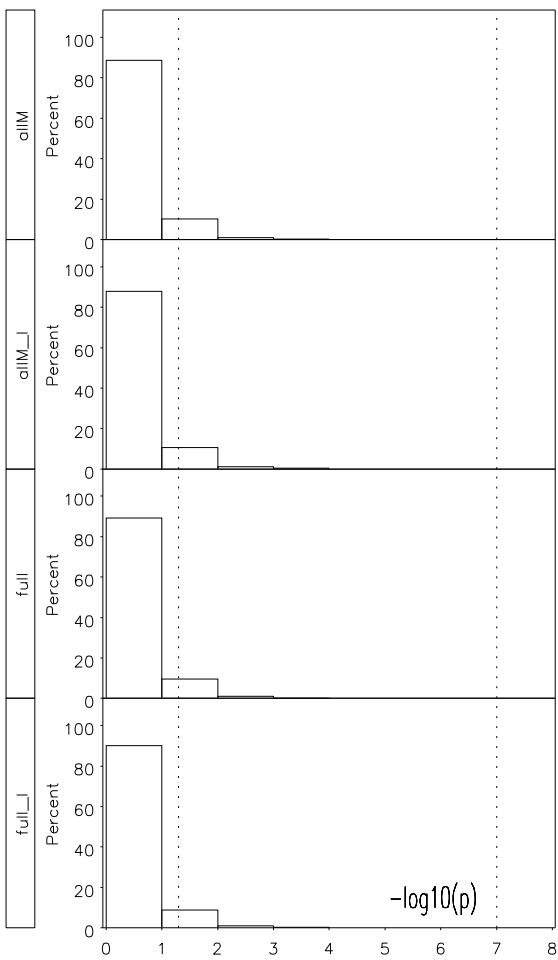
Distribution of p-values: Unphased, ADD, KMod 3, 500 AA families, marker combination 14



## AA-Familien

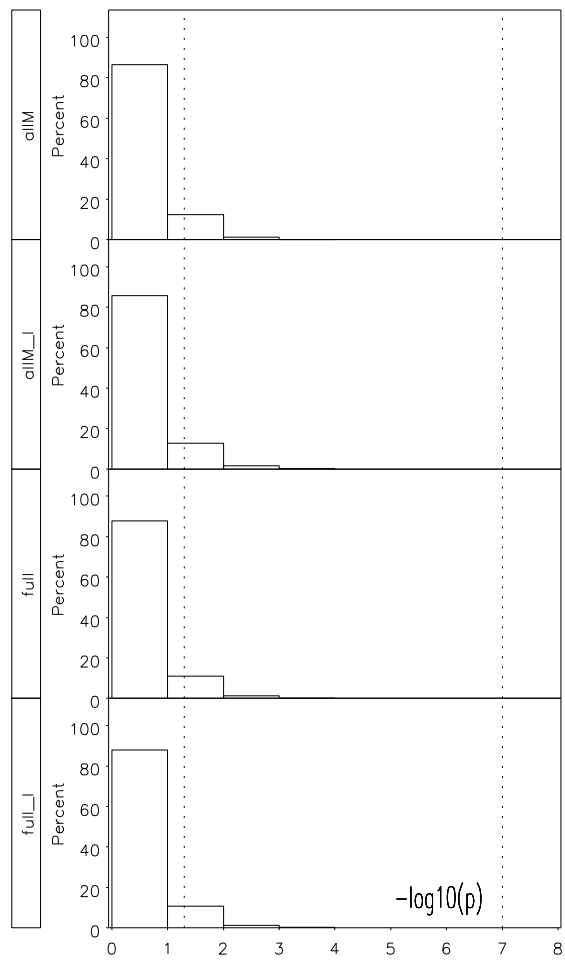
### Markerkombination 15

Distribution of p-values: Unphased, ADD, KMod 3, 500 AA families, marker combination 15



### Markerkombination 30

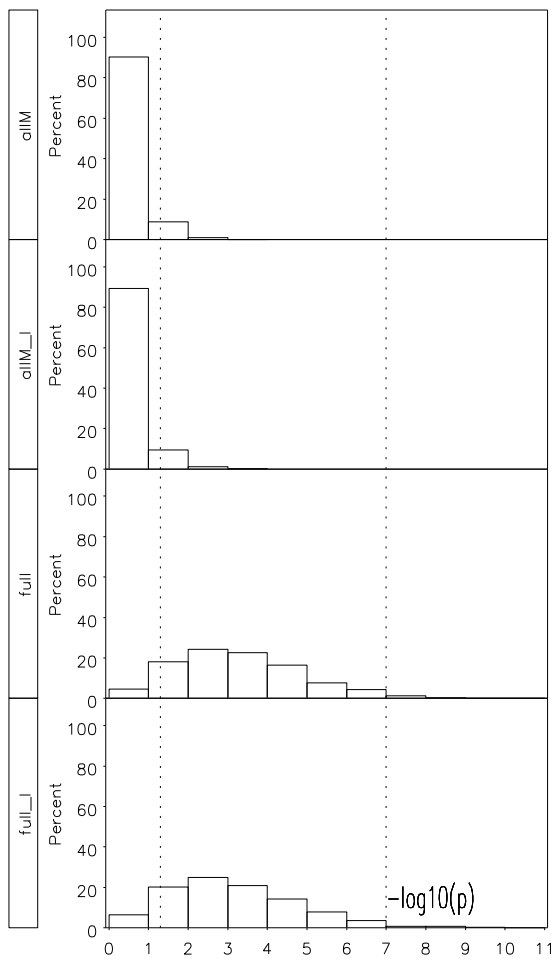
Distribution of p-values: Unphased, ADD, KMod 3, 500 AA families, marker combination 30



## AAUU-Familien

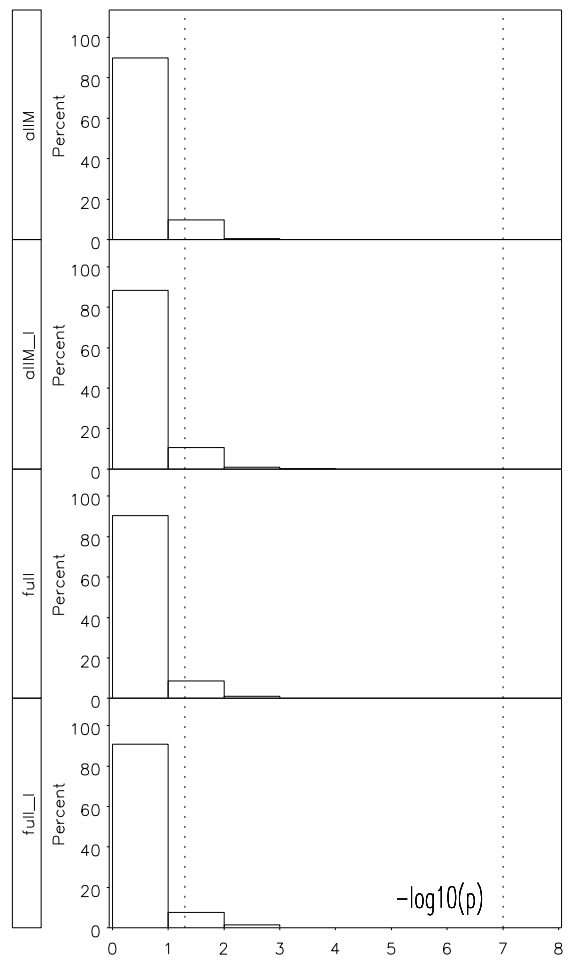
### Markerkombination 14

Distribution of p-values: Unphased, ADD, KMod 3, 500 AAUU families, marker combination 14



### Markerkombination 15

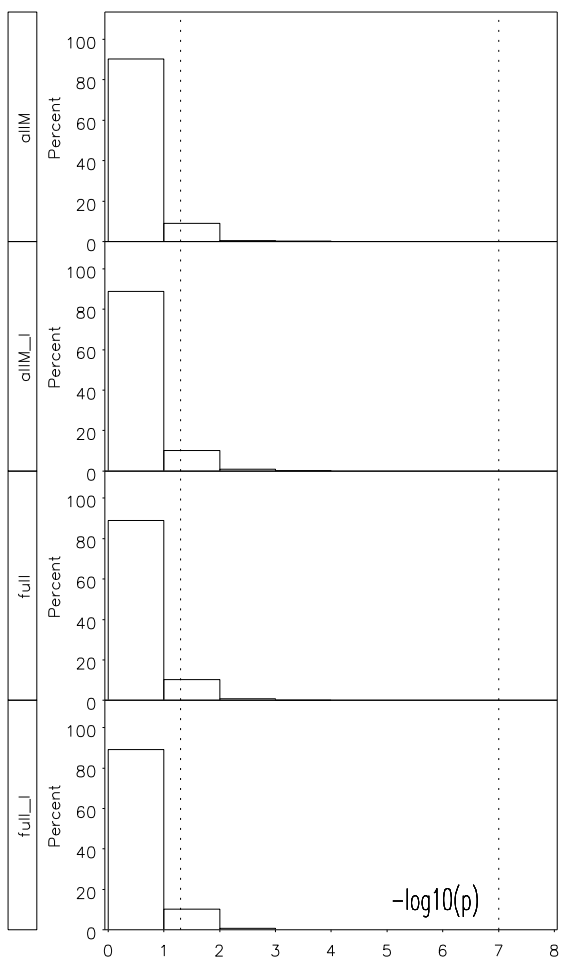
Distribution of p-values: Unphased, ADD, KMod 3, 500 AAUU families, marker combination 15



## AAUU-Familien

### Markerkombination 30

Distribution of p-values: Unphased, ADD, KMod 3, 500 AAUU families, marker combination 30

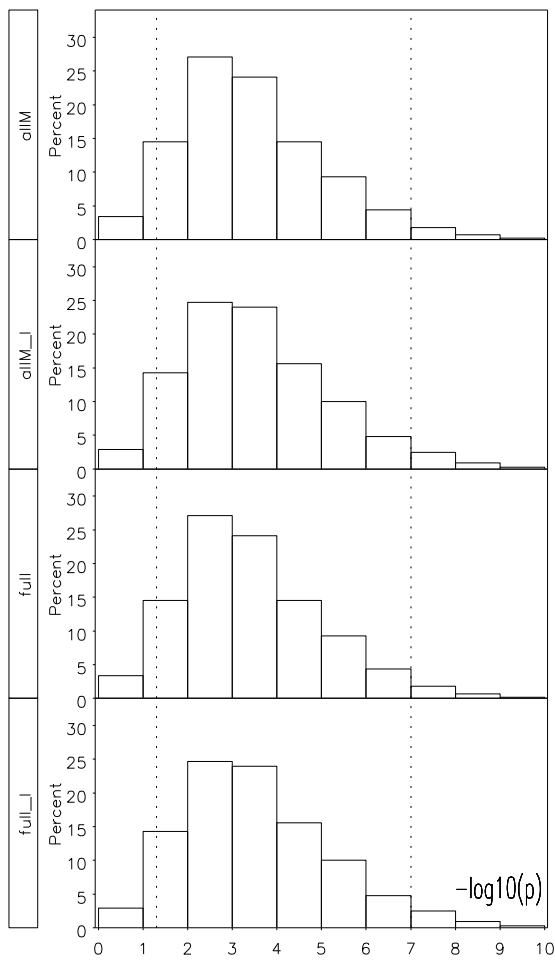




## Multiplikatives Modell KMod 1, AA-Familien

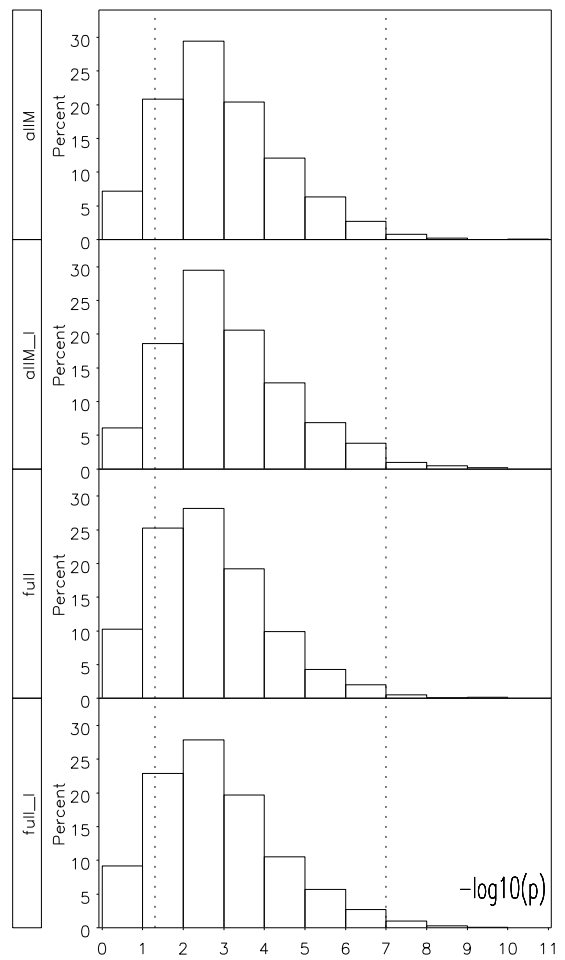
### Markerkombination 4

Distribution of p-values: Unphased, MUL, KMod 1, 200 AA families, marker combination 4



### Markerkombination 20

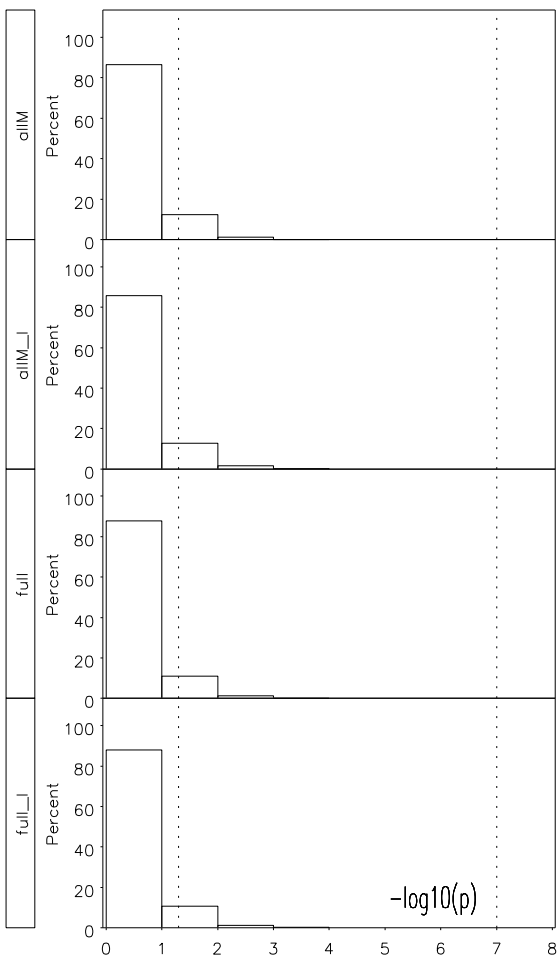
Distribution of p-values: Unphased, MUL, KMod 1, 200 AA families, marker combination 20



## AA-Familien

### Markerkombination 30

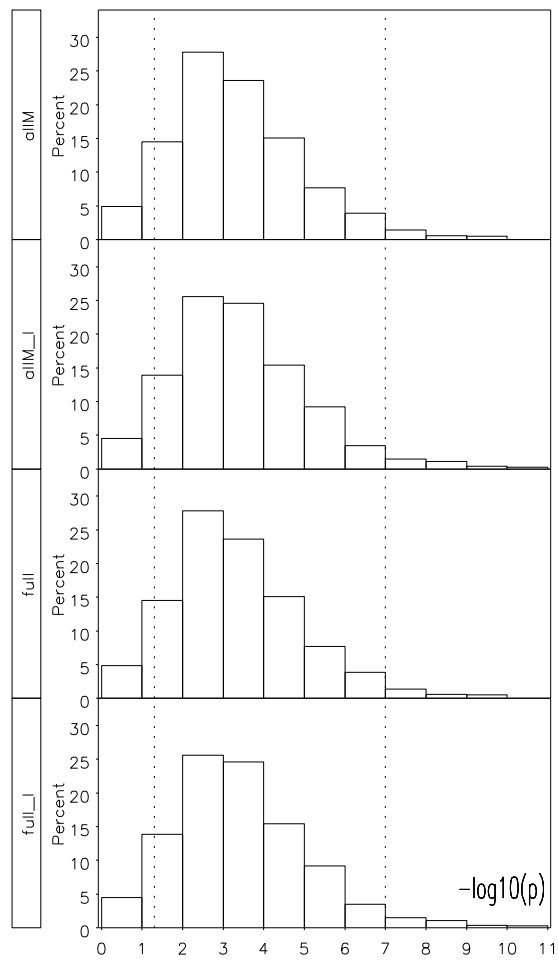
Distribution of p-values: Unphased, ADD, KMod 3, 500 AA families, marker combination 30



## AAUU-Familien

### Markerkombination 4

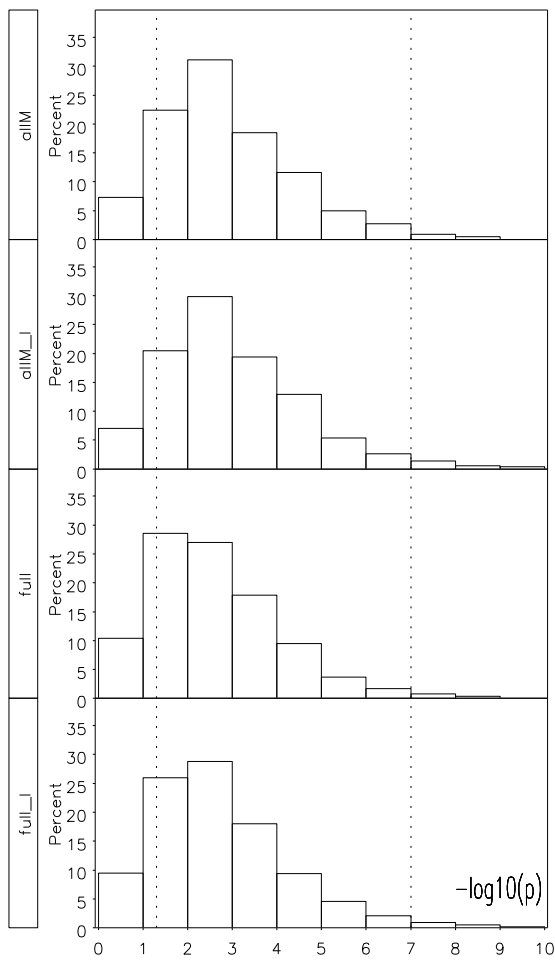
Distribution of p-values: Unphased, MUL, KMod 1, 200 AAUU families, marker combination 4



## AAUU-Familien

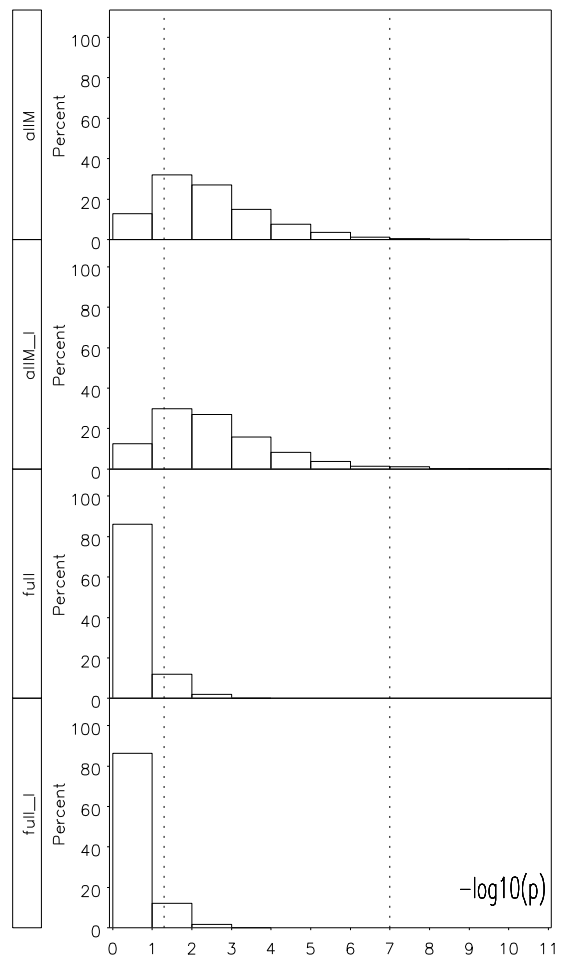
### Markerkombination 20

Distribution of p-values: Unphased, MUL, KMod 1, 200 AAUU families, marker combination 20



### Markerkombination 30

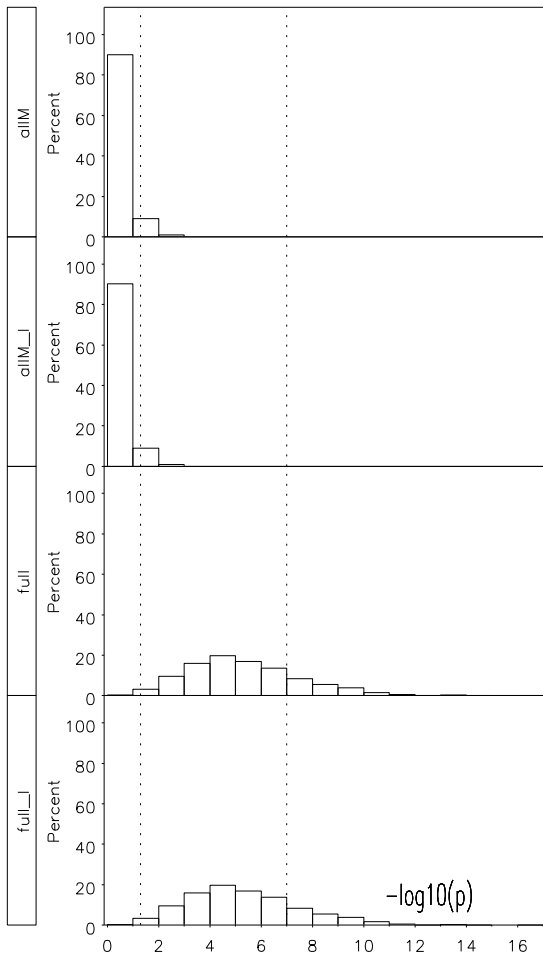
Distribution of p-values: Unphased, MUL, KMod 1, 200 AAUU families, marker combination 30



# Rezessives Modell KMod 2, Trios

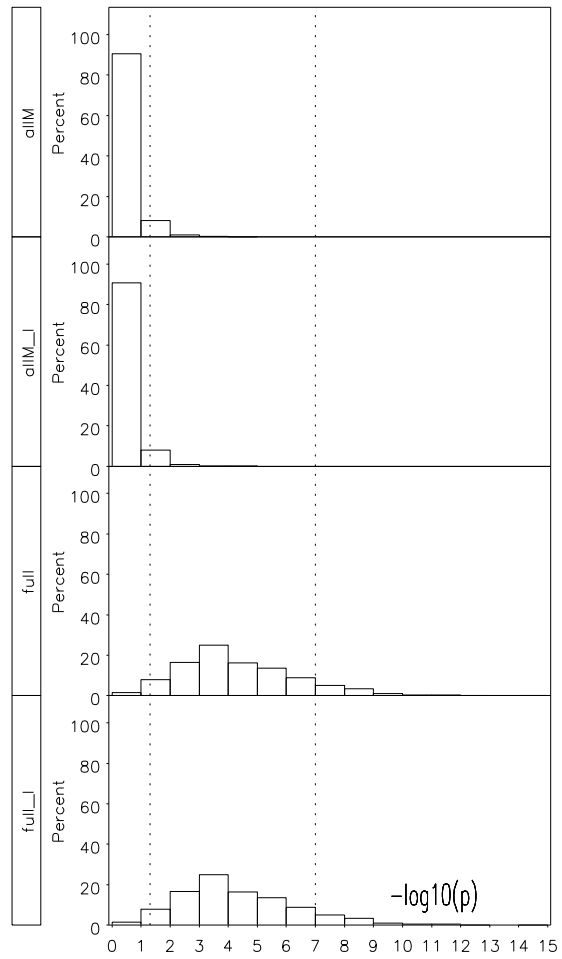
## Markerkombination 12

Distribution of p-values: Unphased, REZ, KMod 2, 6000 trios, marker combination 12



## Markerkombination 14

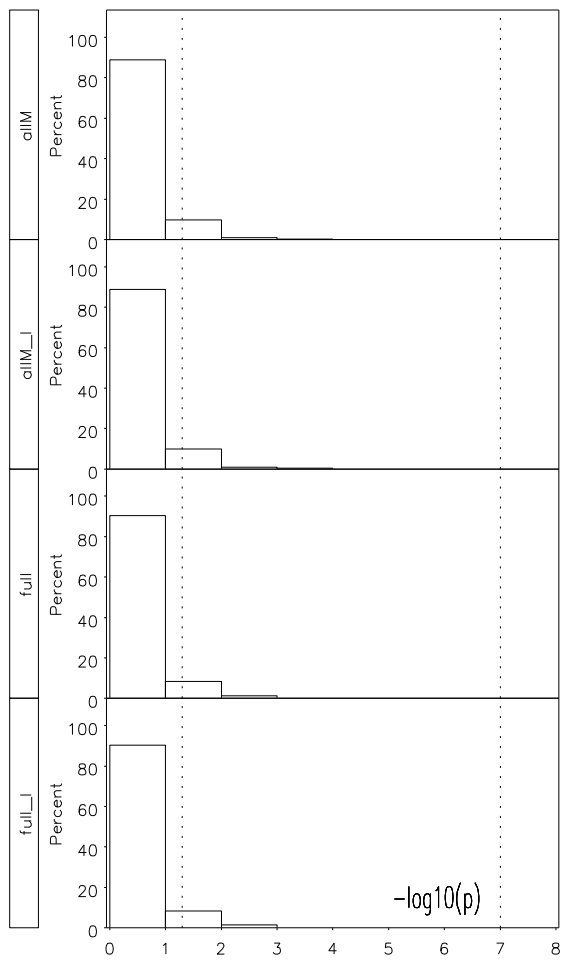
Distribution of p-values: Unphased, REZ, KMod 2, 6000 trios, marker combination 14



### Trios

#### Markerkombination 30

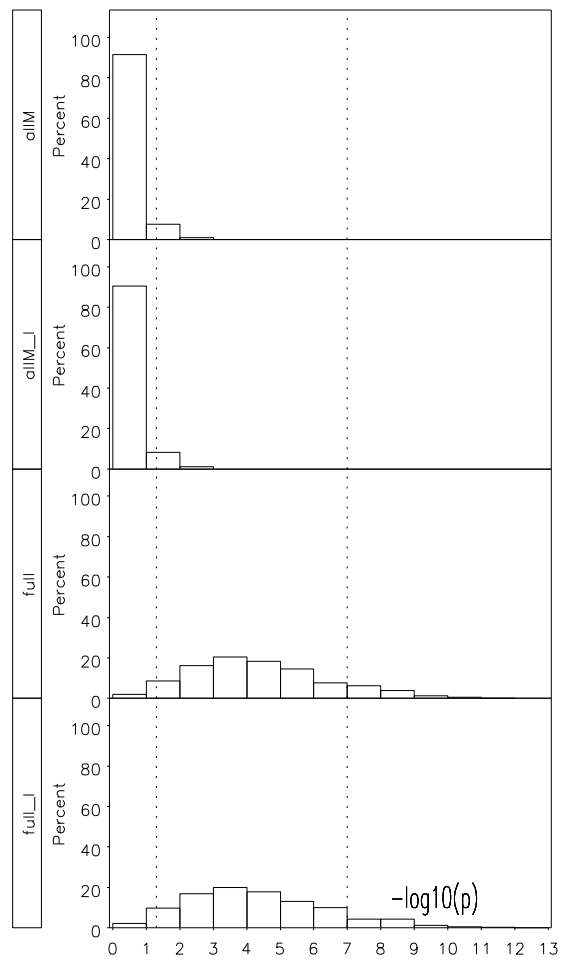
Distribution of p-values: Unphased, REZ, KMod 2, 6000 trios, marker combination 30



### AA-Familien

#### Markerkombination 12

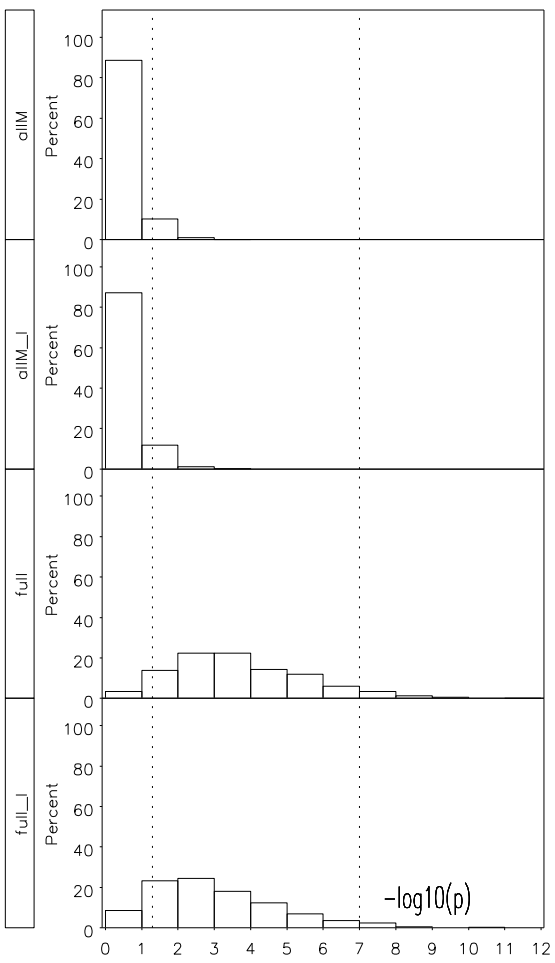
Distribution of p-values: Unphased, REZ, KMod 2, 900 AA families, marker combination 12



## AA-Familien

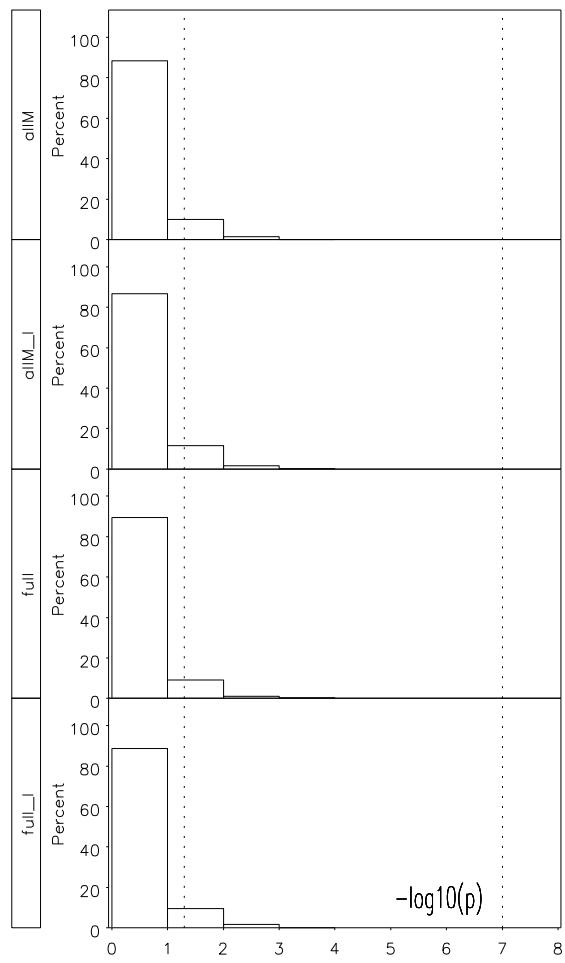
### Markerkombination 14

Distribution of p-values: Unphased, REZ, KMod 2, 900 AAUU families, marker combination 14



### Markerkombination 30

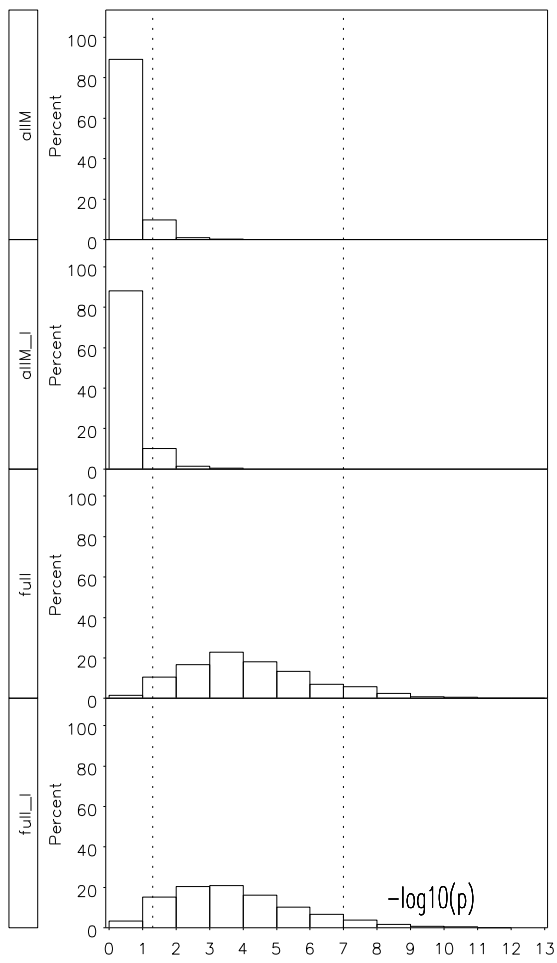
Distribution of p-values: Unphased, REZ, KMod 2, 900 AAUU families, marker combination 30



## AAUU-Familien

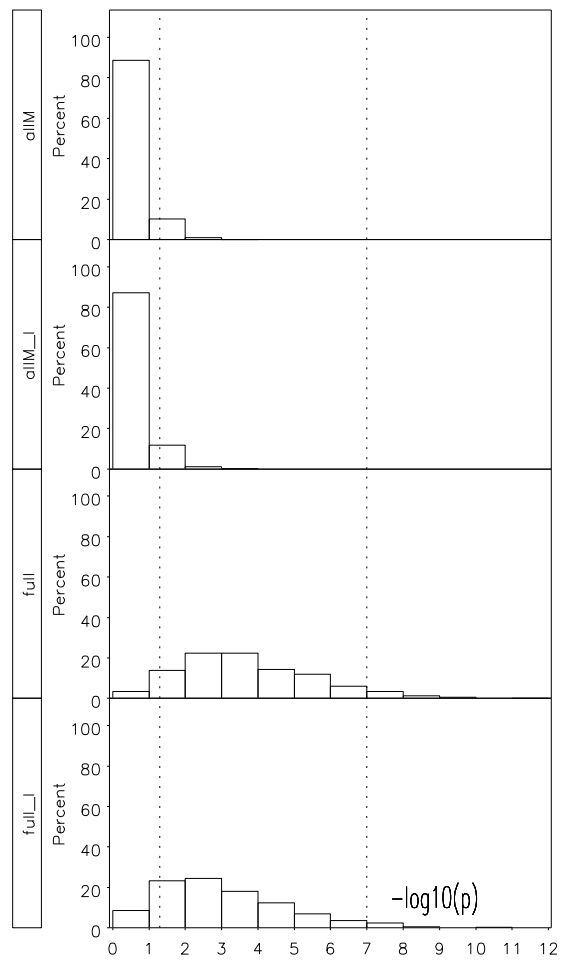
### Markerkombination 12

Distribution of p-values: Unphased, REZ, KMod 2, 900 AAUU families, marker combination 12



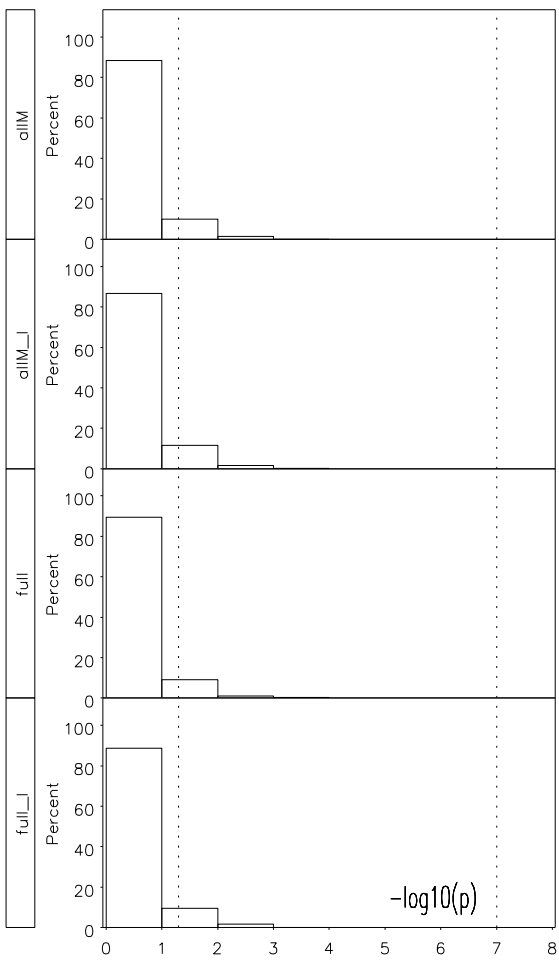
### Markerkombination 14

Distribution of p-values: Unphased, REZ, KMod 2, 900 AAUU families, marker combination 14



## Markerkombination 30

Distribution of p-values: Unphased, REZ,  
KMod 2, 900 AAU families, marker combination 30





## Gegenüberstellung

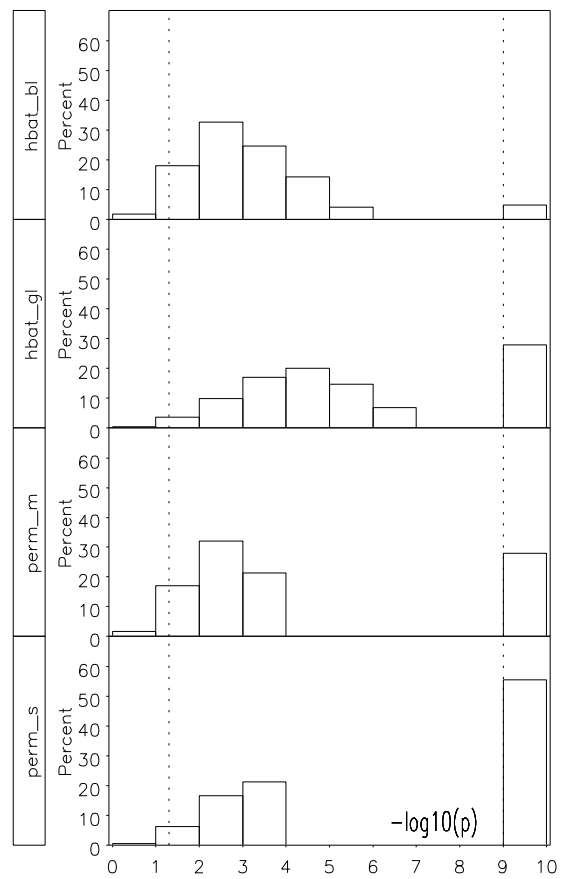
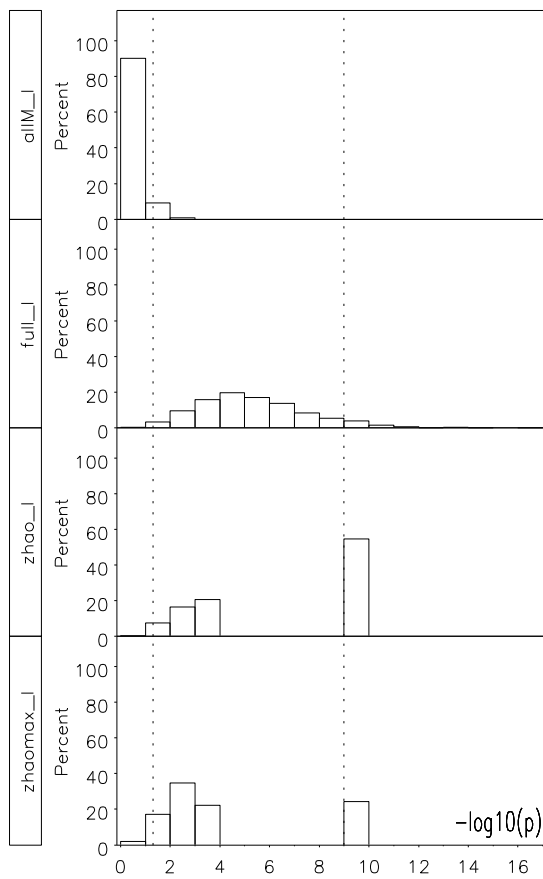
### Rezessives Modell KMod 2, Trios, Markerkombination 12

Famhap / Unphased

FBAT

Distribution of p-values: Famhap / Unphased, REZ,  
KMod 2, 6000 trios, marker combination 12

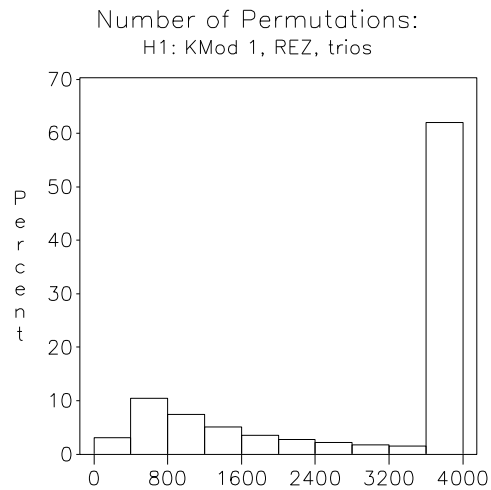
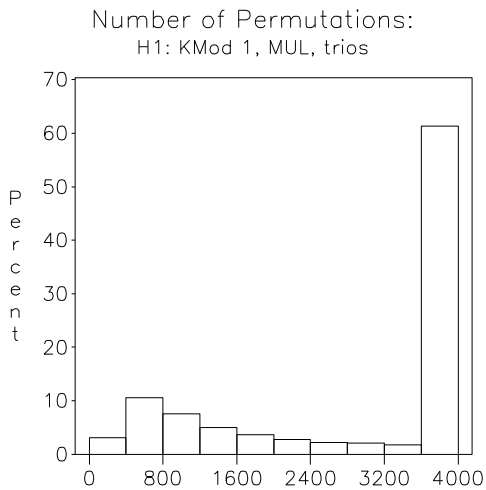
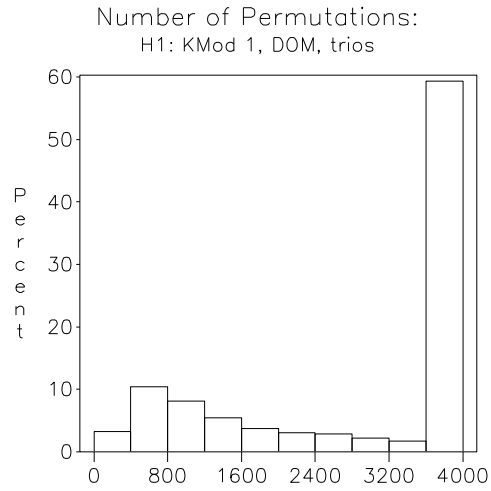
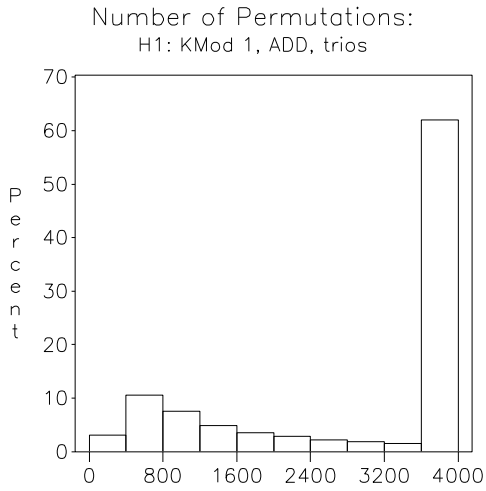
Distribution of p-values: FBAT, REZ,  
KMod 2, 6000 trios, marker combination 12



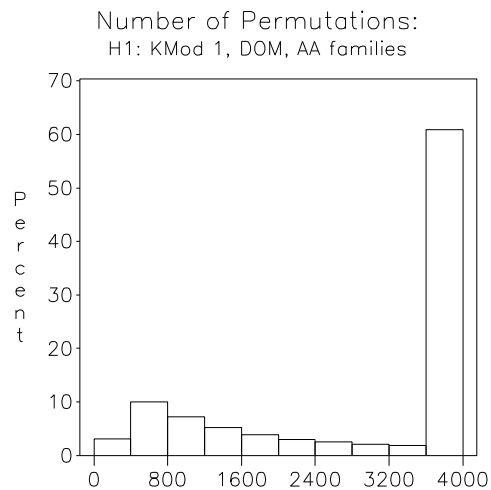
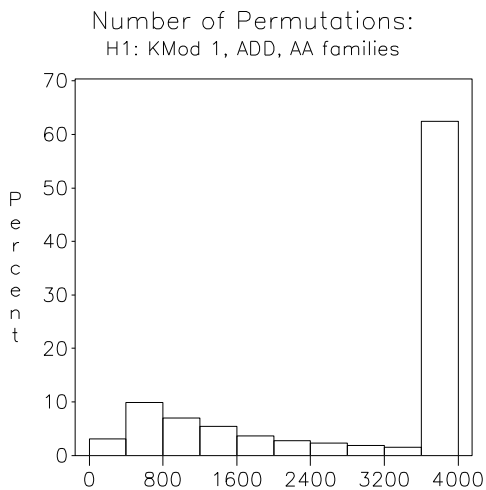
# Anzahl Permutationen: hbat -p4000

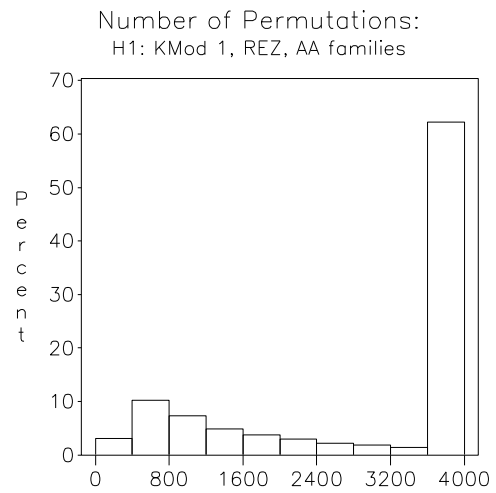
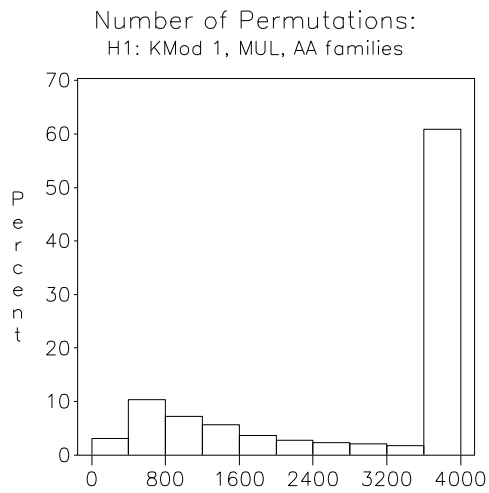
## KMod 1

### Trios

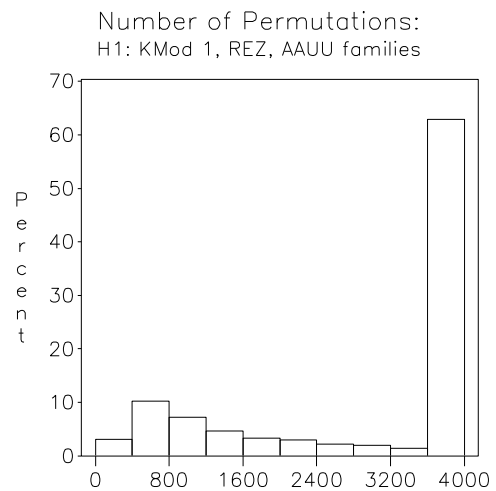
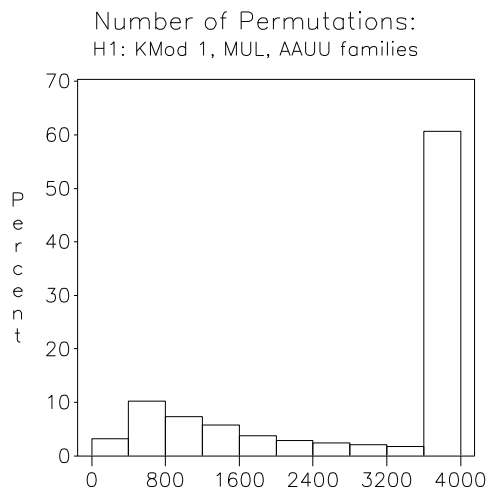
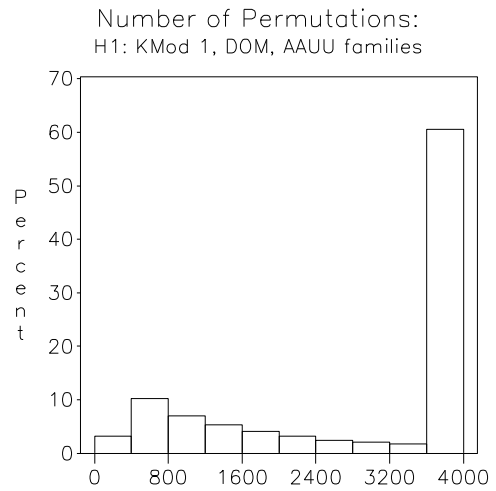
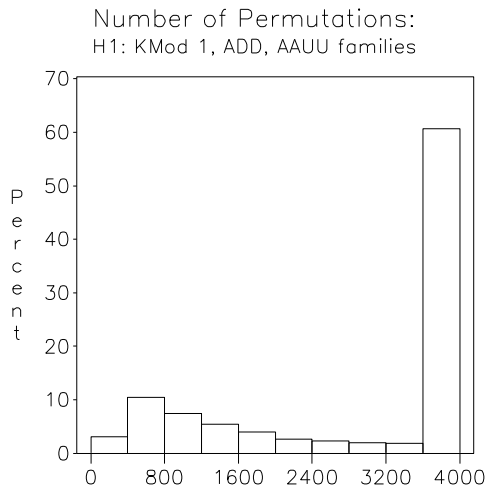


### AA-Familien



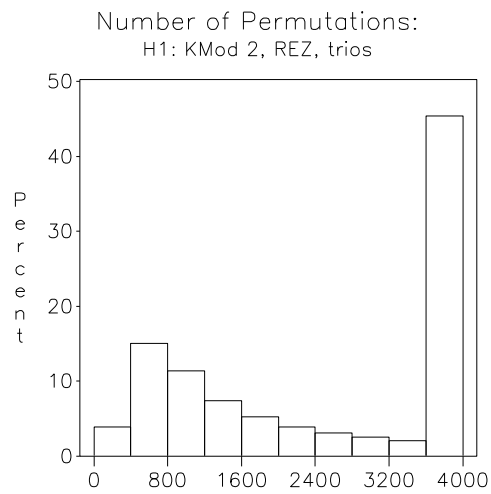
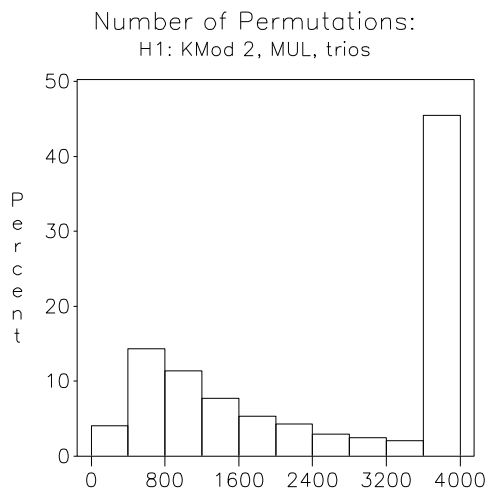
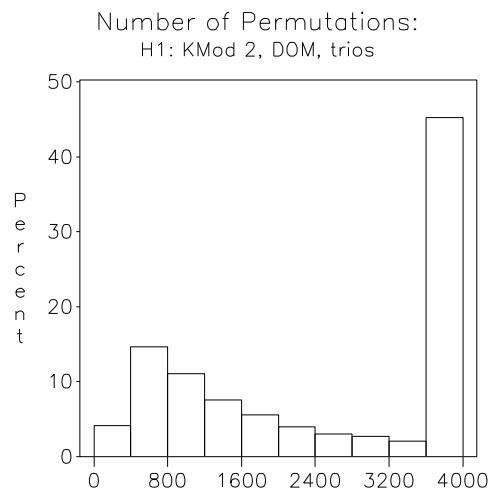
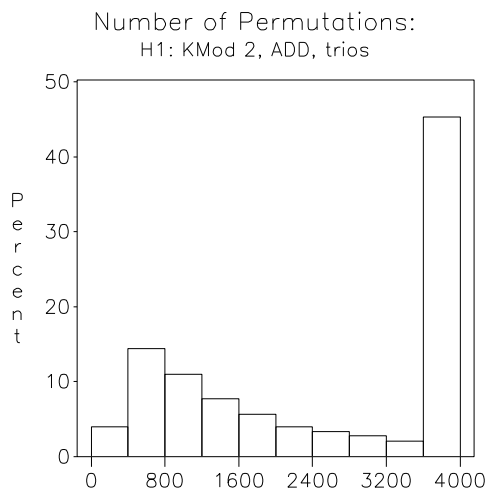


**AAUU-Familien**

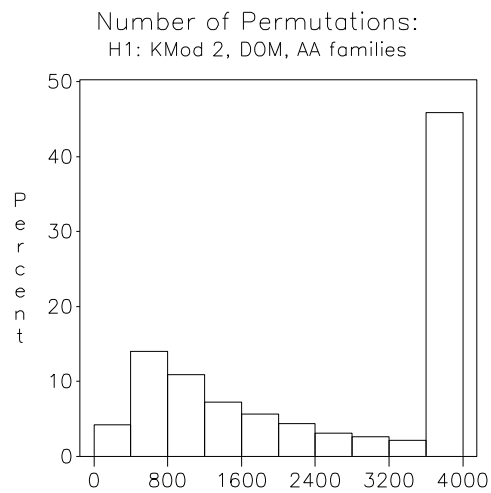
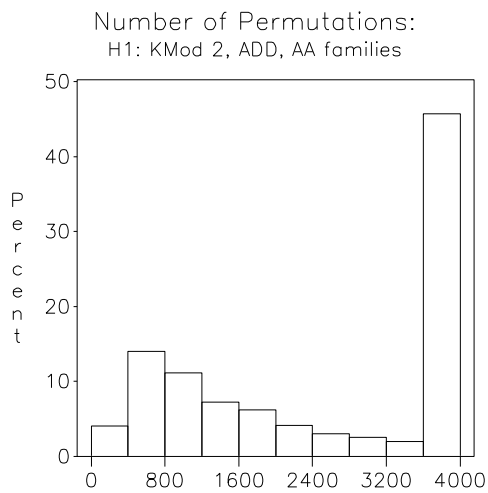


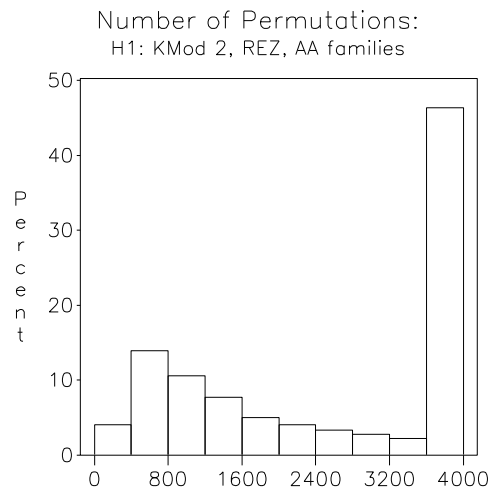
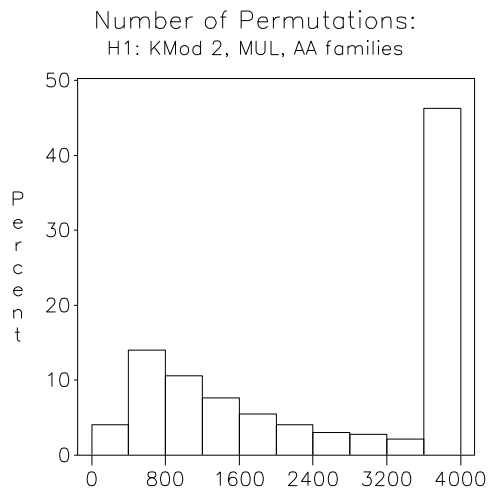
## KMod 2

### Trios

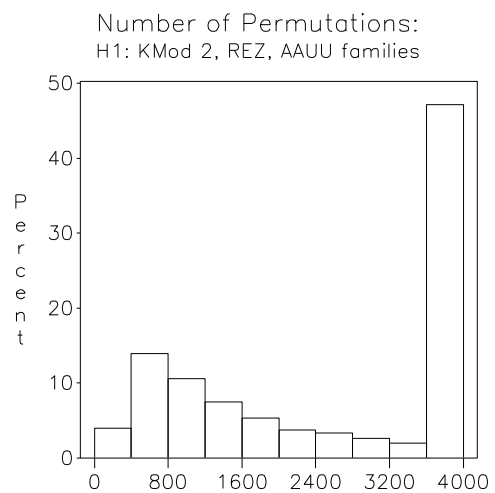
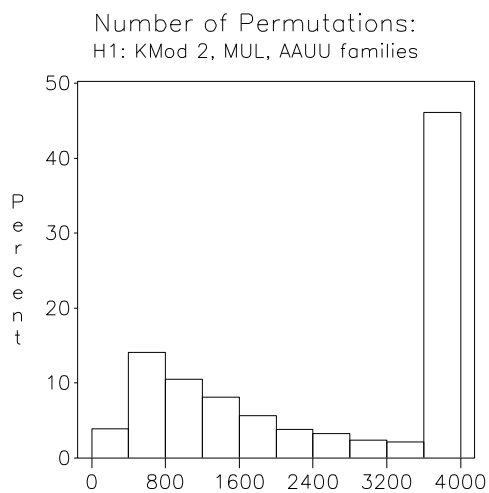
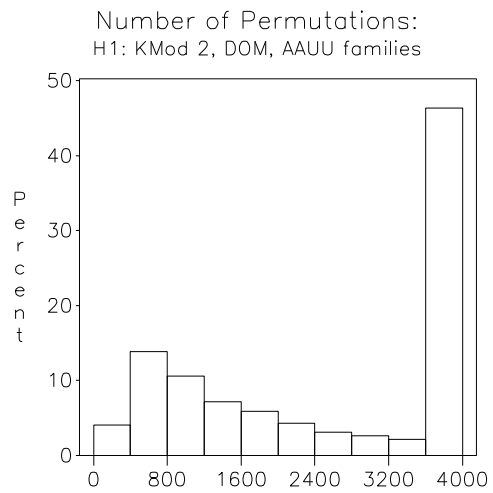
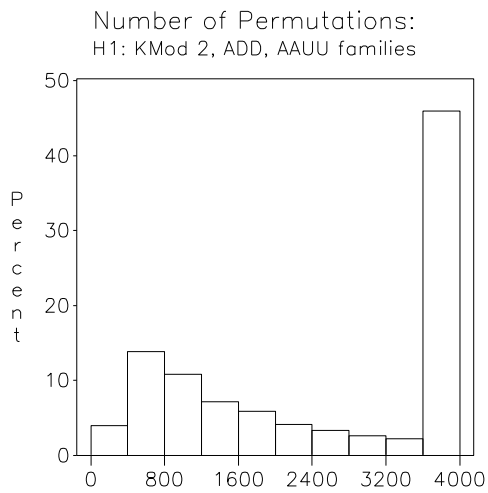


### AA-Familien



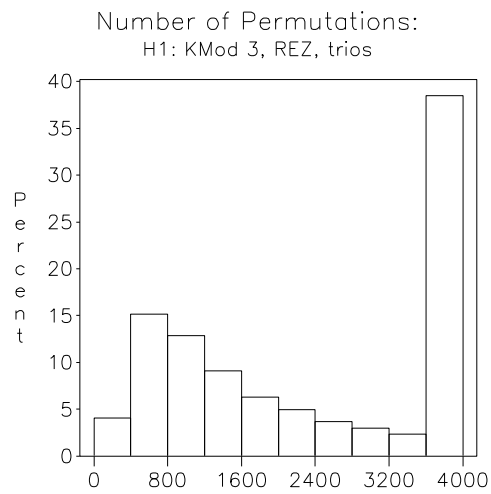
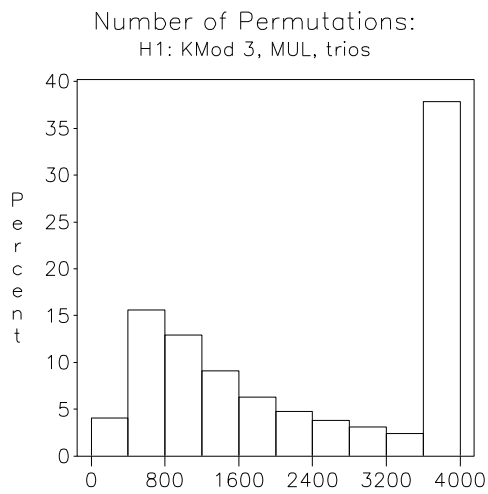
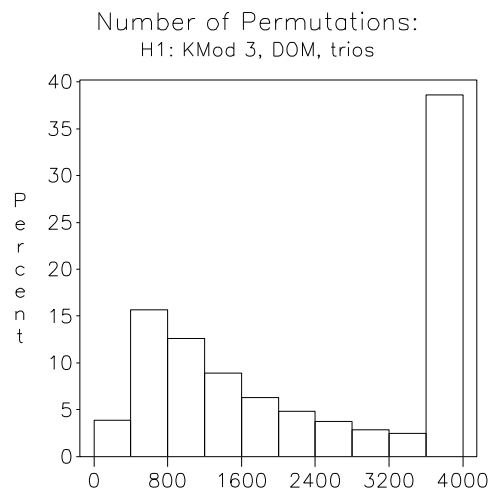
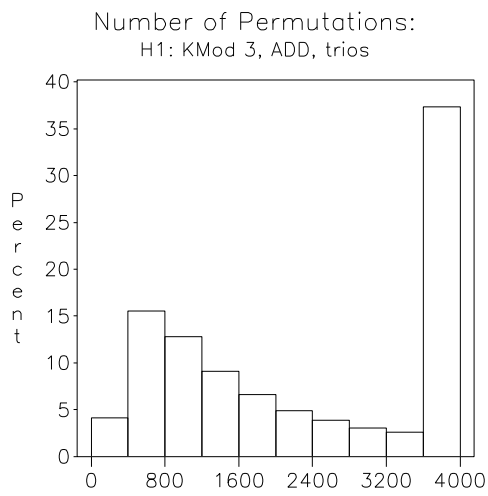


**AAUU-Familien**

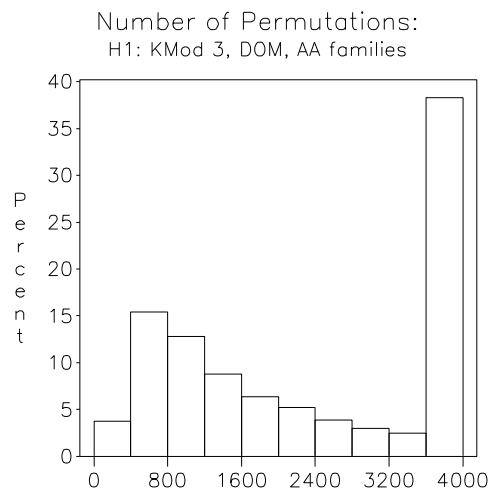
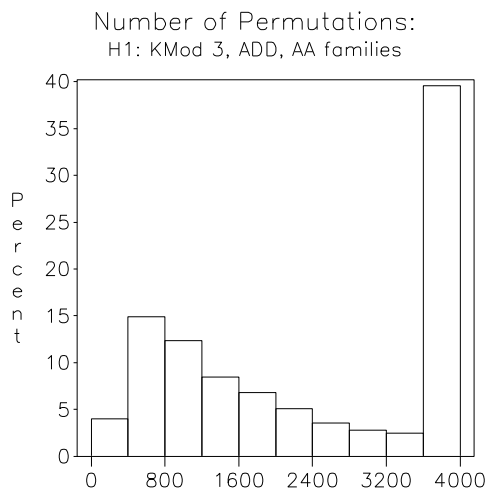


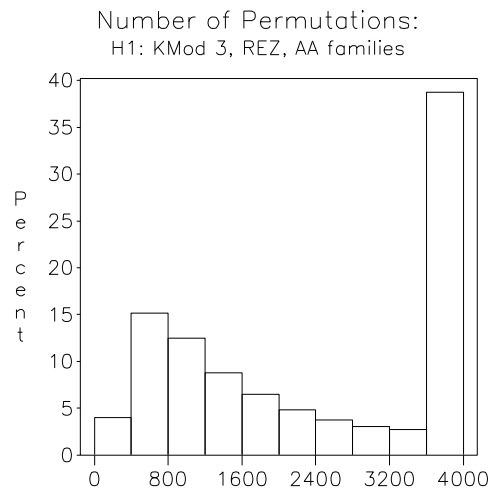
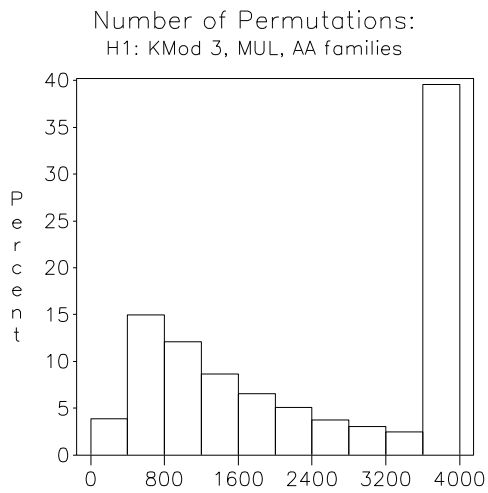
## KMod 3

### Trios



### AA-Familien





**AAUU-Familien**

