

Die Reihe

Biometrie und Medizinische Informatik Greifswalder Seminarberichte

wird herausgegeben von:

*Universitätsprofessor Dr. rer. nat. habil. Karl-Ernst Biebler
und Dr. rer. nat. Bernd Jäger*

Institut für Biometrie und Medizinische Informatik
Ernst-Moritz-Arndt-Universität Greifswald
Walther-Rathenau-Straße 48
17475 Greifswald

Anliegen der Greifswalder Seminarberichte Biometrie und Medizinische Informatik ist es, die Zusammenarbeit im Berührungsfeld von Medizin, Biowissenschaften, Mathematik und Informatik zu fördern. In diesem Sinne steht die Mitarbeit an den Publikationen allen denen offen, die sich einer solchen Vermittlung zwischen den Wissenschaften verpflichtet fühlen und die dazu einen Beitrag leisten wollen.

Biometrische Methoden der Genomanalyse

Karl-Ernst Biebler,
Bernd Paul Jäger,
Michael Wodny
(Editoren)

Biometrie und medizinische Informatik
Greifswalder Seminarberichte

**Karl-Ernst Biebler,
Bernd Jäger,
Michael Wodny**

Biometrische Methoden der Genomanalyse

Shaker Verlag
Aachen 2013

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Copyright Shaker Verlag 2013

Alle Rechte, auch das des auszugsweisen Nachdruckes, der auszugsweisen oder vollständigen Wiedergabe, der Speicherung in Datenverarbeitungsanlagen und der Übersetzung, vorbehalten.

Printed in Germany.

ISBN 978-3-8440-1591-1

ISSN 1439-5320

Shaker Verlag GmbH • Postfach 101818 • 52018 Aachen

Telefon: 02407 / 95 96 - 0 • Telefax: 02407 / 95 96 - 9

Internet: www.shaker.de • E-Mail: info@shaker.de

Inhaltsverzeichnis

1	Die Anfänge der Populationsgenetik	1
	<i>von Charlotte Jana und Karl-Ernst Biebler</i>	
1.1	Themenkreis und Aufgabenstellung	1
1.2	Gedankengut vor Mendel	1
1.3	Gregor Mendel	3
1.4	Mathematisierung	4
1.5	Einfluss der Populationsgenetik auf die Evolutionsforschung	6
2	Datenanalyse in der molekularen Genetik	11
	<i>von Karl-Ernst Biebler</i>	
2.1	Einleitung	11
2.2	Die allgemeine Problemstellung	12
2.3	Datengewinnung	13
2.4	Auswertemethoden	13
2.5	Eine Assoziationsanalyse	19
2.6	Abschließende Bemerkungen	19
3	Sparse Statistical Modelling	23
	<i>von Carolin Malsch</i>	
3.1	Hauptkomponentenanalyse	24
3.2	Sparse Hauptkomponentenanalyse	27
3.3	Anwendung der beschriebenen Methoden auf den Beispieldatensatz	33
4	Konfidenzschätzungen von Allelwahrscheinlichkeiten	49
	<i>von Michael Wodny</i>	
4.1	Ein Genort und zwei Allele	50
4.2	Zweidimensionale asymptotische Konfidenzbereiche . .	60

4.3	Exakte zweidimensionale Konfidenzbereiche im VM_1 ? .	66
5	Methoden zur Analyse von Genkopplung	73
	<i>von Manuela Barth</i>	
5.1	Genetische Marker und ihr Informationsgehalt	73
	Genetische Marker	73
	Heterozygotie	74
	Polymorphie und Kopplung	75
	Vergleich am Beispiel	77
5.2	Genetische Distanzen	78
	Physikalische Distanz	78
	Gendistanzen	78
	Die Haldane-Distanz	79
	Gendistanzen im Vergleich	80
	Zusammenhang genetischer und physikalischer Distanz	81
5.3	Modellbasierte Kopplungsanalyse	82
	Kopplungsungleichgewicht	82
	Analyse bei bekannten Ahnendaten	83
	Test auf Kopplung	85
	Anwendung bei mehreren Familien	87
6	Familienbasierte Assoziationsanalyse	91
	<i>von Julia Bischof</i>	
6.1	Problemstellung und biologische Grundlagen	91
6.2	Herleitung der Nullhypothese für die betrachteten Tests	95
6.3	Transmission/Disequilibrium Test (TDT)	98
6.4	Sib-Transmission/Disequilibrium Test (S-TDT)	102
6.5	Sibship Disequilibrium Test (SDT)	108
7	Fall-Kontroll-Studien bei Betrachtung eines Genlocus	115
	<i>von Lars Romoth</i>	
7.1	Fall-Kontroll-Studien und Allele	115
7.2	Berechnung der Testgrößen	117
	Odds Ratio Test	118
	Chi-Quadrat Test	121

Fishers exakter Test	121
Armitage Trend Test	122
7.3 Testvergleiche mittels Simulation in SAS	122
8 Einfluss mehrerer Loci bei Fall-Kontroll-Studien	131
<i>von Bernd Paul Jäger</i>	
8.1 Einführendes Beispiel	131
8.2 Tests für geschichtete Vierfeldertafeln	136
Der MANTEL-HAENSZEL-Test	136
Definition und Durchführung des BRESLOW-DAY-Tests und seine Verbesserung durch TARONE	141
Konfidenzschätzung für OR bei Schichtung	144
9 Varianten des Wright-Fisher-Modells	149
<i>von Ulrike Horn</i>	
9.1 Die Bedeutung des Wright-Fisher-Modells	149
9.2 Das Grundmodell	150
Annahmen des Grundmodells	150
Das Wright-Fisher-Modell als Markovkette	151
Das Wright-Fisher-Modell als Martingal	152
9.3 Mutationseffekte	154
Mutation innerhalb des Genpools	154
Das „infinite alleles“ Modell und die Neutrale Theorie	157
Erweiterungen und Folgerungen des „infinite alleles“ Modells	160
Das „infinite sites“ Modell	162
9.4 Selektionseffekte	163
9.5 Ausblick - Das Koaleszenzmodell	167
10 Berechnung phylogenetischer Stammbäume	171
<i>von Kristina Plate</i>	
10.1 Abstract	171
10.2 Einleitung	171
10.3 Molekularbiologische Grundlagen	173
Speicher der Erbinformation	173

	Vom Gen zum Merkmal	173
	Mutationstypen	175
	Rolle der DNA in der Phylogenetik	176
10.4	Phylogenetische Grundbegriffe	177
	Grundbegriffe	177
	Stammbäume und Topologien	178
10.5	Mathematische Methoden (nach Knoop & Müller, 2009)	179
	Baumkonstruktion	180
	Most Parsimony	181
	Distanz-Verfahren	185
	Maximum-Likelihood und Bayesianische Verfahren . . .	190
	Zusammenfassung	193
10.6	Phylogenetische Programme	195

Vorwort

Im Sommersemester 2012 wurde das Seminar Biomathematik für Studierende der Studiengänge BSc Biomathematik und MSc Mathematik und Informatik der Ernst-Moritz-Arndt-Universität als Blockveranstaltung durchgeführt. Diese fand vom 29.05. bis 1.6. in der Biologischen Station auf Hiddensee statt.



Seminarteilnehmer (v. l. o.):

Michael Wodny, Lars Romoth, Karl-Ernst Biebler, Julia Bischof, Carolin Malsch, Christina Plate, Manuela Barth, Bernd Paul Jäger, Charlotte Jana

Zum Themenkreis „Mathematische Methoden in der Genetik“ bearbeiteten die Seminarteilnehmer Fragestellungen nach eigener Wahl oder nach Vorschlag des fachlichen Betreuers. Diese Beiträge wurden in einem Vortrag von etwa 90 Minuten dargestellt und ausführlich diskutiert.

Die schriftlichen Ausarbeitungen dazu liegen in diesem Heft 21 der Reihe „Biometrie und Medizinische Informatik - Greifswalder Seminarberichte“ vor. Sie waren für die Studierenden die erste Gelegenheit eine wissenschaftliche Arbeit während des Studiums eigenständig anzufertigen, mündlich sowie schriftlich darüber zu berichten und in eine Publikationsform zu bringen. Dem entsprechend wird man keine eigenen wissenschaftlichen Resultate erwarten. Es ist aber bemerkenswert, mit welchem Einsatz hier weitgehend selbständig an den Themen gearbeitet wurde. Insbesondere die rechentechnischen Umsetzungen waren mit einem hohen Arbeitsaufwand verbunden, der im Nachhinein oft kaum noch sichtbar ist.

Frau PD Dr. I. Blindow, Biologische Station Hiddensee, Frau Zenke und Herrn Prof.Dr. Gliesche, (AG Mikrobielle Ökologie, Hiddensee) danken die Teilnehmer sehr herzlich für die gewährte Unterstützung.

Weiterhin danken wir Carolin Malsch für die Hilfe bei der Organisation des Seminares und Charlotte Jana für die Schlussbearbeitung dieses Heftes.

Karl-Ernst Biebler, Bernd Jäger und Michael Wodny