

**Genomische Abbildungen
-
Grundlagen der
n-Tupelmethode**

Uwe Kraeft

2017

Berichte aus der Medizinischen Informatik und Bioinformatik

Uwe Kraeft

Genomische Abbildungen
—
Grundlagen der n-Tupelmethode

Shaker Verlag
Aachen 2017

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 2017

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

Printed in Germany.

ISBN 978-3-8440-5156-8

ISSN 1432-4385

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

Vorwort

Dieser Text ist eine Ergänzung [KrVIIIb] zu Teil III der „Elemente der angewandten Zahlentheorie und Approximationen“ [KrVIIIa] des „Lehrgangs der Mathematik“. Er ist elementar geschrieben und bereits mit den grundlegenden Kenntnissen von [KrI] auch für interessierte Schüler der höheren Gymnasialklassen einfach verständlich.

Ausgangspunkt war unter anderem die Beschäftigung mit der Bioinformatik und den Sequenzen des mitochondrialen Genoms in den Bänden 2 bis 4 der „Bivalvia-Bivalves-Muscheln“ von Uwe Kraeft & Michael Kraeft [KK2], [KK3] und [KK4] (siehe Seite vi), aus denen Teile entnommen wurden und in denen zahlreiche Anwendungsbeispiele zu finden sind. Organismen sind bekanntlich Abbildungen des Genoms. In vergleichbarer Weise sind die hier gebildeten Tripel- oder Quadrupelhäufigkeiten (n-Tupelverfahren siehe nächste Seite) ein Bild des Genoms und damit vergleichbar mit den Organismen. Dabei ist zumindest anzunehmen, dass identische oder ähnliche Sequenzen gleiche oder vergleichbare n-Tupelhäufigkeiten zeigen.

Hier werden in 6 Kapiteln die Abbildungen von diskreten Elementen, die Rekonstruktion als Abbildung, die Abbildung von alphanumerischen Sequenzen durch sich überlappende n-Tupel, intermittierendes Scannen und Zusammenfügung von sich überlappenden n-Tupeln, die mitochondrialen Gene der Hominiden sowie der Bivalvia in elementarer Weise dargestellt. Eine Literaturliste und nahezu alle Programme sind beigefügt. In einem Anhang werden die Spiralen kurz behandelt. Die Literaturzitate betreffen wie in den vorangehenden Bänden nicht nur die Übernahme von Inhalten, sondern sind auch ein Hinweis für interessierte Leser, um mehr über ein Thema zu erfahren.

Das Buch stellt die Meinung des Autors nach dem Studium der Literatur und dessen Kenntnissen dar. Der Inhalt wurde sorgfältig auf Fehler geprüft, die aber nicht gänzlich ausgeschlossen werden können. Eine Gewährleistung oder Garantie für die Richtigkeit des Textes kann nicht übernommen werden. Ich bin für entsprechende Hinweise oder Verbesserungsvorschläge dankbar.

Auswahl von Symbolen (weitere Symbole siehe Text)

\forall	für alle
\exists	es gibt
$\Rightarrow, \Leftarrow, \Leftrightarrow$	hieraus folgt (in den angegebenen Richtungen)
\in	ist Element von (ist enthalten in)
$A=\{a,b,c\}$	Beispiel einer Menge A mit Elementen a, b und c
$\{\{...\}$	Ansammlung
\emptyset	die leere Menge
\mathbb{N}	natürliche Zahlen 1, 2, 3, ...
\mathbb{Z}	ganze Zahlen ..., -2, -1, 0, 1, 2, ...
$\varphi(a)$	Abbildung, Funktion
$=$	mathematisch gleich
$:=$	gleich per definitionem
bp	Basen, Basenpaare (Symbolik der Tabellen siehe Seite 34)

n-Tupelverfahren

Die bestimmten endlichen und möglichst vollständigen Sequenzen aus den Basen zum Beispiel von Mitochondrien oder allgemein aus bestimmten Zahlen beziehungsweise Buchstaben werden intermittierend um eine Position fortschreitend in Tripeln, Quadrupeln oder anderen n-Tupeln gescannt, wobei gleiche n-Tupel gezählt und in Prozent von der Gesamtzahl der n-Tupel umgerechnet werden. Die bestimmten Anteile können in Abhängigkeit von den durchnummerierten verschiedenartigen n-Tupeln grafisch dargestellt und dabei mit verschiedenen Arten oder allgemein Objekten verglichen werden. Dieser Vergleich kann auch quantitativ mit Bildung der Summen der absoluten linearen beziehungsweise quadratischen Differenzen oder Abweichungen von den n-Tupelhäufigkeiten der Vergleichsarten oder -objekte durchgeführt werden; dies entspricht den Flächen beziehungsweise Volumina der absoluten Integraldifferenzen von stetigen Kurven. Falls wie bei den Basen der Lebewesen inverse n-Tupel existieren, können diese generell oder bei Zweifeln an den Bestimmungen der Basensequenzen jeweils mit den zugehörigen n-Tupeln addiert und beiden jeweils zum Beispiel zur Hälfte als Häufigkeiten der Basenpaare zugeordnet werden.

Inhalt

	Seite
1. Abbildungen von diskreten Elementen - - - -	1
2. Rekonstruktion als Abbildung - - - -	5
3. Die Abbildung von alphanumerischen Sequenzen durch sich überlappende n-Tupel - - - -	7
3.1 Ganze Zahlen aus vier Ziffern - - - -	7
3.2 n-Tupel von Genomen aus vier Basen - - - -	9
3.3 Sequenzermittlung aus n-Tupeln - - - -	12
3.4 Relative Minima von Basenpaartripeln - - - -	18
3.5 Statistische Betrachtung - - - -	20
4. Intermittierendes Scannen und Zusammenfügung von sich überlappenden n-Tupeln - - - -	25
5. Die mitochondrialen Gene der Hominiden - - - -	33
6. Die mitochondrialen Gene der Bivalvia - - - -	53
 Literaturauswahl - - - -	 67
 Lehrgang der Mathematik- - - -	 69
 Programme - - - -	 75
GENOM3 - - - -	76
GENOM4 - - - -	81
INVERS - - - -	90
ROTATION - - - -	95
POSITIONA - - - -	99
LOTTO - - - -	-103
 Anhang: Spiralen und Schrauben - - - -	 -105