UNIVERSITÄT ZU LÜBECK
INSTITUT FÜR
NEURO- UND BIOINFORMATIK

Masterarbeit im Rahmen des Studiengangs Medizinische Informatik

Der Genetische Code und seine Optimierung statistisch untersucht

The genetic code and its optimization inspected with statistic methods

Seves Arne Martenstein
Matr.-Nr.: 628521, Medizinische Informatik

Betreut und ausgegeben von:
PD Dr. rer. nat. Amir Madany Mamlouk

Lübeck, den 30. April 2019

Ich versichere an Eides statt, die vorliegende Arbeit selbstständig und nur unter Benutzung der angegebenen Quellen und Hilfsmittel angefertigt zu haben.

Lübeck, den 30. April 2019

Inhaltsverzeichnis

1 Abstract

Der genetische Code ist auf unserem Planeten fast universell. Die genauen Gründe dafür werden bisher nur vermutet, es konnte noch nicht nachgewiesen werden, dass es keinen Code gibt, der dem Selektionsdruck der Evolution besser gewachsen ist. Die eigentliche Frage ist daher, welche Faktoren es sind, nach denen der genetische Code optimiert ist.
Vorherige Arbeiten haben gezeigt, dass die Polarität der codierten Aminosäuren sehr gut durch den Code konserviert wird. Jedoch gibt es auch dort Codes, die dies weitaus besser können. Zudem wurde gezeigt, dass der genetische Code auch durch eine geschickte Verteilung der Nukleotide den aus Mutationen resultierenden Fehler minimiert, jedoch nicht in den codierenden Sequenzen.
Diese Arbeit reproduziert und verifiziert die bisherigen Ergebnisse und untersucht genauer, wie der genetische Code auf andere Faktoren wie Hydrophobizität oder isoelektrischen Punkt der Aminosäuren optimiert ist. Zudem werden in fast allen bisherigen Arbeiten die Stoppcodons nicht betrachtet. Diese Arbeit untersucht daher genauer, welche Rolle Stoppcodons in der Evolution des genetischen Codes gespielt haben könnten.

The genetic code is almost universal on our planet. The exact reasons for this are not yet known exactly. It was not yet possible to prove that there is no code which is fitter in terms of evolution than the natural genetic code. The real question therefore is, which are the optimized points of the genetic code.
Previous publications have shown that the polar requirement of the coded amino acids conserved very well by the genetic code. But there are codes which are even better in conserving the polar requirements of the amino acids. Also it has been shown that the real sequence of the human genome supports the conservation of the polar requirements by preferring codons and nucleotides which reduce the error when a mutation occurs. But this effect is not present on the actual coding sequences.
This work reproduces and verifies already known results and takes a stronger look at other characteristics of the amino acids like hydropathy or isoelectric point of the amino acids. Also almost all prevoius publications do not handle stop codons. This work also does some research on their possible role in the evolution of the genetic code.

2 Einleitung

Im Jahr 1953 entdeckten J. Watson und F. Crick die Struktur der Desoxyribonukleinsäure, kurz DNA[2]. Sie beschrieben die Doppelhelix als Molekülstruktur mit den vier Nukleotiden als Bestandteile. Die Zuordnung der Tripletts zu den 20 essentiellen Aminosäuren wurde 1966 von Crick vorgestellt [1]. Die sogenannte Codesonne in Abbildung 1 zeigt welche Kombinationen aus Aminosäuren der Ribonukleinsäure (RNA) für welche Aminosäuren codieren. Die Leserichtung ist dabei vom 5'- zum 3'-Ende der Nukleotidkette, entsprechend ist die Codesonne von innen nach außen zu lesen. Die Im Rahmen dieser Arbeit verwendeten Sequenzen stammen aus dem Genom und bestehen dementsprechend aus DNA. Da DNA anders als RNA jedoch nicht das Nukleotid Uracil (U) sondern Thymin (T) enthält, wird die Codesonne in dieser Arbeit dementsprechend übersetzt verwendet. Dabei wird einfach das U durch T ersetzt.

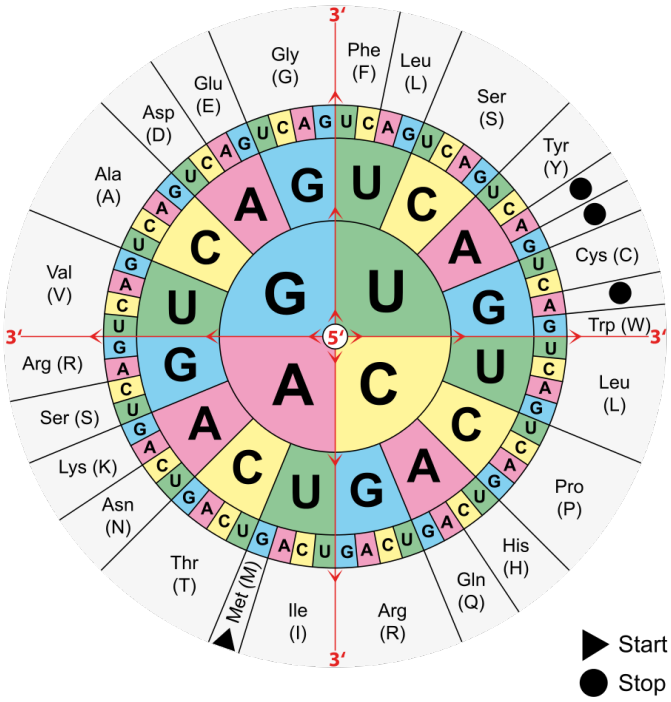

Abbildung 1: Die Zuordnung der Aminosäuren zu den 64 Codons der RNA [21]

Dass dieser genetische Code, also die Zuordnung der Nukleotid-Tripletts zu den 20 Aminosäuren, in fast allen Lebewesen gleich ist, ist weithin bekannt. Nur wenige Lebensformen, wie zum Beispiel Mitochondrien in vielen Lebewesen, haben diesen Code verändert. Neben dem bekannten genetischen Code sind derzeit lediglich 32 weitere Varianten bekannt [25]. Die Veränderungen in diesen alternativen Codes betreffen jedoch nur wenige Codons die dann für eine andere Aminosäure codieren. Daher stellt sich nach wie vor die Frage, warum dieser Code denn überall verwendet wird. Die naheliegendste Erklärung ist, dass er dem Selektionsdruck der Evolution am besten standhalten kann und somit der beste Code ist und jede Mutation einen

schlechteren Code hervorbringen würde. Um zu prüfen, inwiefern diese Hypothese zutrifft, kann man nach Faktoren suchen, die den natürlichen genetischen Code besonders machen. Daher wird in dieser Arbeit nach Eigenschaften gesucht, die im genetischen Code „optimiert“ sind. Der Begriff „optimiert“ beschreibt hier jedoch Faktoren in denen der natürliche genetische Code außerordentlich gut abschneidet. Dabei wird der natürliche Code stets mit einer Menge an zufälligen genetischen Codes verglichen. Solche „optimierten“ Faktoren können eine Rolle in der Evolution gespielt haben und somit der Grund dafür sein, warum sich dieser genetische Code gegen alle anderen durchgesetzt hat.
In der 1966 erschienenen Arbeit „The molecular basis for the genetic code“[7] beschreiben Woese et al. verschiedene molekulare Charakteristika der Aminosäuren. Unter anderen wurden Hydhrophobie, Polaritäten und Molekulargewicht gemessen. Basierend auf den von Woese et al. gemessenen Polaritäten veröffentlichten David Haig, und Laurence D. Hurst 1991 den Artikel „A Quantitative Measure of Error Minimization in the Genetic Code“[8]. Darin vergleichen Sie den natürlichen genetischen Code mit 10.000 zufällig generierten Codes indem sie anhand der aminosäurenspezifischen Werte Scores definieren, die eine Größenordnung für die Polaritätsveränderung durch Aminosäurenaustausch angeben. Diese Scores bilden über alle möglichen Punktmutationen den Mittelwert der Veränderung eines Parameters durch diese Mutation. Sie kommen dort zu dem Ergebnis, dass der genetische Code besonders effektiv darin ist, die Veränderung der Polaritäten der Aminosäuren im Falle einer Mutation zu minimieren. Sie berechnen über die Menge ihrer Stichprobe (10.000), wie viele der Zufallscodes die jeweiligen Faktoren besser als der natürliche Code konservieren. Aber auch die Hydrophobizität wird durch den genetischen Code außerordentlich gut konserviert. Tabelle 1 zeigt die Ergebnisse dieser Arbeit.

Tabelle 1: Von Haig und Hurst berechnete Wahrscheinlichkeit p dass ein zufälliger Code aus den 10.000 Zufallscodes das entsprechende Merkmal besser konserviert. [8]

Parameter	Position 1	Position 2	Position 3	Gesamt (1-3)
Polarität	0,0037	0,0214	0,0002	0,0002
Hydrophobizität	0,0003	0,9100	0,0142	0,0089
Molekularvolumen	0,3812	0,3763	0,3503	0,3003
Isoelektrischer Punkt	0,9828	0,7487	0,3452	0,9281

1998 veröffentlichten Stephan J. Freeland und Laurence D. Hurst ihre Arbeit „The Genetic Code Is One in a Million“[9]. In dieser Arbeit vergleichen Sie den natürlichen genetischen Code mit einer Million Zufallscodes und bestätigen das Ergebnis von Haig und Hurst. Zusätzlich führen sie das Transition/Transversion Bias ein, welches Polaritätsveränderungen aus Transitionen (Austausch A↔G oder C↔T) höher gewichtet als Transversionen (alle anderen Basenaustausche). Trotz des Faktes dass es vier mögliche Transversionen und nur zwei Transitionen gibt, treten in der natürlichen Umgebung etwa doppelt so viele Transitionen als Transversionen auf. Dies ist durch verschiedene biologische Mechanismen begründet, welche in dem Artikel „Estimate of the Mutation Rate per Nucleotide in Humans“[19] präsentiert werden. Freeland und Hurst zeigten, dass der genetische Code gegen die häufiger auftretenden Transitionen deutlich robuster ist als gegen Transversionen. Durch das Anwenden des Bias

mit dem Faktor 2 (Transitionsfehler werden doppelt so hoch bewertet wie Transversionsfehler) konnten sie zeigen, dass nur noch ein Zufallscode aus einer Million die Polarität der Aminosäuren besser konserviert als der natürliche genetische Code.

R. Geyer erweiterte die Berechnungen von Freeland und Hurst in ihrer Bachelorarbeit 2014 um die Berücksichtigung von leserasterverschiebenden Mutationen (Frameshift) [6]. Sie stellte fest, dass der genetische Code auch bei Frameshifts optimiert ist und auch die Wahrscheinlichkeit p einen Zufallscode zu finden, der robuster gegen Frameshiftmutationen ist als der natürliche Code, bei lediglich 0,000267 liegt. Alle diese bisher vorgestellten Arbeiten gehen implizit von einer Sequenz aus, in der alle 4 Nukleotide gleich häufig verteilt sind und die Wahrscheinlichkeiten bei Shiftmutationen für aufeinanderfolgende Nukleotide ebenfalls gleich sind. Doch sowohl in den Auftretenswahrscheinlichkeiten der Nukleotide und Tripletts (Punktmutationen) als auch in den Übergangswahrscheinlichkeiten für aufeinander folgende Nukleotide und Tripletts ist keine bekannte Sequenz genau gleich verteilt.

In meiner Bachelorarbeit (2016)[5] wurden alle bisherigen Scores in einem globalen GMS-Score gebündelt und die Mutationen mit Gewichtungen aus realen Sequenzen versehen. Dabei wurde der Ansatz den Stephan J. Freeland und Laurence D. Hurst mit dem Transition/Transversion Bias begannen, Daten aus der „Realität“ mit einfließen zu lassen, weiter verfolgt und aus realen Sequenzen A-priori- und Übergangswahrscheinlichkeiten für Nukleotide und Tripletts berechnet. So konnte ich zeigen, dass die Anzahl der konservativeren Codes unter der Zuhilfenahme dieser Gewichtungen auf bis zu einem in zwei Millionen sinkt. Außerdem begann ich mit einem Greedy-Algorithmus gezielt nach noch konservativeren Codes zu suchen. Dabei konnte ich zeigen, dass es Codes gibt, die den Fehler um ein vielfaches besser minimieren als der natürliche Code.
B. Klaucke berechnete in ihrer Bachelorarbeit[4] daraufhin getrennte Statistiken für codierende Sequenzen und gesamte Chromosomen. Die codierenden Sequenzen wurden im korrekten Leseraster abgelesen und die Berechnungen aus [5] damit wiederholt.
Sie konnte dann zeigen, dass gerade auf codierenden Sequenzen durch die Anwendung der Gewichtungen wieder mehr zufällige Codes „besser“ sind als der natürliche Code. Eine Erklärung für diesen Effekt wurde dort jedoch nicht gefunden.

Da die in den vorherigen Arbeiten berechneten Scores auf Veränderungen der Polaritäten der resultierenden Aminosäuren basieren, wurden jegliche Mutationen, die aus Stoppcodons entstehen oder zu Stoppcodons führen, aus der betrachteten Menge der Mutationen entfernt. Dies ist ein logisches Vorgehen, da bei einem Abbruch keine Polarität bestimmt werden kann und somit kein Wert für die weitere Berechnung der Scores zur Verfügung steht. In dieser Arbeit wird gezielt untersucht, ob es Faktoren in der Verteilung der Stoppcodons gibt, welche den natürlichen Code in seiner konservierenden Funktion unterstützen oder hemmen. Dabei wird unter anderem untersucht, ob Mutationen in den codierenden Sequenzen des menschlichen Genoms häufiger oder seltener zu neuen Stoppcodons führen und ob die Verteilung der Stoppcodons einen schnellen Abbruch der Transkription nach einer Mutation

fördert.

Zudem wird in dieser Arbeit der Ansatz von Haig und Hurst wieder aufgegriffen, neben den Polaritäten auch noch andere Charakteristika der Aminosäuren zu vergleichen. Ein besonderer Schwerpunkt wird dabei darauf gelegt, ob bestimmte Faktoren, auch wenn sie insgesamt durch den Code weder besonders gut noch besonders schlecht konserviert werden, gerade durch die Gewichtungen der CCDS besser oder schlechter abschneiden als andere zufällige Codes.

3 Daten und Methoden

3.1 Der genetische Code

Der genetische Code wird von allen bisher bekannten Lebensformen verwendet. Es gibt nur sehr wenige Lebensformen oder Zellorganellen wie Mitochondrien, die eine leicht veränderte Variante des genetischen Codes verwenden [25]. In dieser Arbeit wird die Bezeichung „genetischer Code“ als Synonym für den „Standard Genetic Code“ verwendet. Er beschreibt, wie Ketten der Basen Adenin (A), Thymin (T), Guanin (G) und Cytosin(G) als Sequenz gelesen und über die Zwischenschritte der Transkription in RNA und der Translation an den Ribosomen in Polypeptidketten übersetzt werden können. Die vier Basen liegen als Kette mitsamt eines komplementären Gegenstranges in einer Doppelhelix in der DNA vor. Je drei Moleküle (auch Triplett oder Codon genannt) codieren für eine der 20 essentiellen Aminosäuren. Die Transkription startet immer bei dem Startcodon „ATG“ und bricht bei einem der Stoppcodons „TGA“, „TAA“ oder „TAG“ ab. Der Code ist degeneriert, das bedeutet, dass mehrere Codons für die gleiche Aminosäure codieren können. Die genaue Zuordnung der Codes zu den Aminosäuren ist in Abbildung 1 ersichtlich.

3.2 Charakteristika der Aminosäuren

Alle bereits genannten Arbeiten verwenden für ihre Berechnungen die Polaritäten der Aminosäuren. Die in Tabelle 2 dargestellten Werte wurden von Woese et al.[7] im Jahr 1966 veröffentlicht. Sie geben an, wie polar die Umgebung sein muss, damit eine Aminosäure mit anderen Molekülen wechselwirken kann. Der Begriff Polarität ist daher nicht ganz korrekt, wird jedoch in anderen Arbeiten genau so verwendet. Die korrekte Übersetzung wäre „polare Anforderung“. Die Polaritäten der Aminosäuren sind sowohl für die Faltung als auch für die Funktion eines Proteins enorm wichtig, denn nur wenn die Reaktionen oder Bindungen an den Aminosäuren genau dann stattfinden, wenn es geplant ist, führt das Protein seine geplante Funktion korrekt aus.
Neben der Polarität werden in dieser Arbeit analog zu Haig und Hurst [8] verschiedene Charakteristika der Aminosäuren verwendet. Die gezeigten Hydrophobizitätswerte wurden 1982 von J. Kyte und RF. Doolittle in ihrer Arbeit „A simple method for displaying the hydropathic character of a protein“[12] vorgestellt.
Die Daten zum Molekularvolumen stammen aus „Amino acid difference formula to help explain protein evolution“ von Grantham[15]. Häufig wird das Molekülvolumen auch als Van-Der-Waals-Volumen angegeben, zur besseren Vergleichbarkeit mit [8] werden hier jedoch die Werte von Grantham verwendet. Die Daten zum Isoelekrtischen Punkt in [8] sind zum Großteil identisch mit den in Tabelle 3 gezeigten Werten zu pI. In dieser Artbeit wurden die in Tabelle 3 gezeigten Werte zu pI benutzt, da die Messungen der dort verwendeten Quelle aktueller sind.

Tabelle 2: Polarität, Hydrophobitizät und Molekularvolumen der essentiellen Aminosäuren. (entspricht Tabelle 1 in [8])

Aminosäure	Polarität	Hydrophobizität	Molekularvolumen
Ala	7,0	1,8	31
Arg	9,1	-4,5	124
Asn	10,0	-3,5	54
Asp	13,0	-3,5	56
Cys	4,8	2,5	55
Glu	12,5	-3,5	83
Gln	8,6	-3,5	85
Gly	7,9	-0,4	3
His	8,4	-3,2	96
Ile	4,9	4,5	111
Leu	4,9	3,8	111
Lys	10,1	-3,9	119
Met	5,3	1,9	105
Phe	5,0	2,8	132
Pro	6,6	-1,6	32,5
Ser	7,5	-0,8	32
Thr	6,6	-0,7	61
Trp	5,2	-0,9	170
Tyr	5,4	-1,3	136
Val	5,6	4,2	84

Zusätzlich zu der Hydrophobizität und der Polarität wurden aus dem „CRC Handbook of Chemistry“[17] noch Daten zu folgenden Charakteristika verwendet:

- M_r: Molekulargewicht
- pK_a: Negativer Logarithmus der Säuredissoziationskonstanten der COOH-Gruppen
- pK_b: Negativer Logarithmus der Säuredissoziationskonstanten der NH_2-Gruppen
- pI: pH-Wert am isoelektrischen Punkt

Die Werte dieser Eigenschaften sind Tabelle 3 zu entnehmen.

Tabelle 3: Molekulare Eigenschaften der essentiellen Aminosäuren

Aminosäure	M_r	pK_a	pK_b	pI
Ala	89,09	2,33	9,71	6,00
Arg	174,20	2,03	9,00	10,76
Asn	132,12	2,16	8,73	5,41
Asp	133,10	1,95	9,66	2,77
Cys	121,16	1,91	10,28	5,07
Glu	147,13	2,16	9,58	3,22
Gln	146,15	2,18	9,00	5,65
Gly	75,07	2,34	9,58	5,97
His	155,16	1,70	9,09	7,59
Ile	131,17	2,26	9,60	6,02
Leu	131,17	2,32	9,58	5,98
Lys	146,19	2,15	9,16	9,74
Met	149,21	2,16	9,08	5,74
Phe	165,19	2,18	9,09	5,48
Pro	115,13	1,95	10,47	6,30
Ser	105,09	2,13	9,05	5,68
Thr	119,12	2,20	8,96	5,60
Trp	204,23	2,38	9,34	5,89
Tyr	181,19	2,24	9,04	5,66
Val	117,15	2,27	9,52	5,96

3.3 Generierung der Zufallscodes

Im Rahmen meiner Bachelorarbeit wurde ein Codegenerator entwickelt, welcher zuverlässig auf allen Codonpositionen gleich verteilte Zufallscodes generiert. Da in diese Arbeit jedoch den genetischen Code nur mit jeweils einer Million Zufallscodes vergleicht, wird zur Vergleichbarkeit das Code-Set von R. Geyer[6] verwendet. Dieses Code-Set enthält Codes, welche blockweise die Zuordnung der Aminosäuren zu den Tripletts permutieren. Das bedeutet, dass alle Tripletts, die vorher für eine Aminosäure codieren, auch in jedem der Zufallscodes für die gleiche Aminosäure codieren. Diese blockweise Randomisierung ermöglicht es 20! ($\geq 2,4 \times 10^{18}$) verschiedene Codes zu erzeugen. Das bedeutet, dass durchschnittlich lediglich einer von $2,4329 \times 10^{12}$ Codes in dieser Stichprobe enthalten ist. Die gesamte Menge aller möglichen Code-Variationen ist mit aktuellen Computern nicht berechenbar.
Würden anstelle der Blöcke alle 64 Tripletts einzeln permutiert, wären 42! verschiedene Codes möglich. Dies wurde jedoch nicht als sinnvolle Erweiterung gesehen, da aktuell bereits nur mit einer sehr kleinen Stichprobe der Gesamtmenge gearbeitet werden kann.

3.4 Implementierung der Berechnungen

Grundlage aller Berechnungen ist das Java-Framework, welches bereits in meiner Bachelorarbeit zum Einsatz kam [5][20][23]. Dieses Framework wurde jedoch von

Grund auf umstrukturiert. Unter anderem wurden die Gewichtungen als eigene Objekte erstellt und die Main-Methode übersichtlicher gemacht. Die Berechnungen, die den natürlichen Code mit dem Zufallscodeset vergleichen, wurden weiterhin über mehrere Threads durchgeführt um die Laufzeit überschaubar zu halten. Alle anderen Berechnungen wurden nativ im Single-Thread Modus durchgeführt. Einige von B. Klaucke vorgenommene Änderungen und Fehlerbehebungen wurden auch in dieses Framework übernommen. Der Code ist öffentlich zugänglich, Details sind im Anhang zu finden.

3.5 Z-Score

Der Z-Score ist ein Verfahren, mit welchem ein Wert in einer Population abhängig vom Mittelwert $\bar{x}$ und der Standardabweichung σ übersichtlich angegeben werden kann. [3]

$$z = \frac{x - \bar{x}}{\sigma} \tag{1}$$

Der Betrag von z stellt die Distanz zwischen Mittelwert der Wertemenge und dem aktuellen Wert dar. Die Maßeinheit ist dabei die Standardabweichung. Positive Z-Scores liegen demnach über dem Mittelwert, negative darunter.
Die Berechnung der empirischen Standardabweichung wird mit der Formel

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{2}$$

durchgeführt.

3.6 Erzeugung einer Zufallssequenz

Um die Berechnungen dieser Arbeit zu validieren wurden mit zufälligen, gleichverteilten Daten getestet. Dafür wurde eine pseudozufällige Nukleotidsequenz generiert. Diese Sequenz hat eine Länge von 250 Millionen Basen und wurde mit dem Xorshift Generator [10] generiert. Korrekt implementierte Berechnungen sollten mit Gewichtungen aus der mit Xorshift generierten Sequenz etwa die gleichen Ergebnisse liefern wie eine Berechnung ohne Gewichtungen.
Der Xorshift Algorithmus wurde gewählt, da er schneller als die Java Math.Random -Funktion ist und zudem auf allen Bits gleich verteilt ist. Die generierte Sequenz hat daher nahezu eine Gleichverteilung der Nukleotide und eine zufällige Reihenfolge. Xorshift zählt zu den Pseudozufallsgeneratoren, das bedeutet, dass mit dem gleichen Startwert (Seed) stets die gleiche Zahlenfolge erzeugt wird. Die in dieser Arbeit verwendete Implementierung von Xorshift hat eine Zykluslänge von 2^{64}. Als Seed wird dabei die aktuelle Uhrzeit über den Aufruf `System.nanoTime()` verwendet. Die in [10] vorgestellte Implementierung wurde im Rahmen dieser Arbeit etwas modifiziert um analog zu Java Math.Random eine Zufallszahl in einem bestimmten Bereich erzeugen zu können.

```
public int getInt(int bound) {
```

```
    seed ^= (seed << 21);
    seed ^= (seed >>> 35);
    seed ^= (seed << 4);
    cnt += 1234567891234567 89L;
    int result = (int) ((seed + cnt) % bound);
    return (result < 0) ? -result : result;
}
```

Die Werte 21, 35 und 4 wurden gewählt, da diese dafür sorgen, dass der Zyklus des Algorithmus die volle Länge von 2^{64} hat[11]. Es gibt neben dem hier verwendeten Werteset jedoch noch weitere Kombinationen, die eine volle Zykluslänge garantieren. Da der Algorithmus nur bei sehr wenigen Seed-Werten 0 zurückgeben kann ist zusätzlich die Addition einer ungeraden Zahl nötig. Über den Modulo Operator wird dann die 64 Bit Ergebnis-Zahl auf den Suchbereich von 0 (inklusive) und bound (exklusive) projiziert.
Die Wahl dieses Generators ist lediglich eine Vorsichtsmaßnahme, da die Zufallssequenz keinerlei Informationen enthalten darf und der Java-Zufallsgenerator einige statistische Tests zur Zufälligkeit nicht besteht. Auch der Xorshift-Algorithmus besteht nicht alle diese Tests, jedoch einige mehr als der Java Math.Random Zufallsgenerator. Vermutlich ist diese Maßnahme jedoch überflüssig.

3.7 Gewichtungen

Aus realen Nukleotidsequenzen können Statistiken berechnet werden. Diese Statistiken beschreiben bestimmte Eigenschaften der untersuchten Sequenz. Normalisiert man diese Statistiken auf einen Mittelwert von 1, so können sie als Gewichtung für weitere Berechnungen verwendet werden. Zur besseren Übersicht werden die Tabellen mit Z-Scores dargestellt. Die für den Z-Score benötige Varianz wird dabei aus allen in der jeweiligen Matrix vorhandenen Elementen berechnet. Alle Statistiken werden nur auf dem Vorwärtsstrang erfasst.
Folgende bereits in [5] eingeführte Statistiken werden in dieser Arbeit verwendet:

- **Nukleotid a priori (NA)** Enthält die Auftretenswahrscheinlichkeit für jedes der vier Nukleotide T, C, A und G in der untersuchten Sequenz.
- **Triplett a priori (TA)** Enthält die Auftretenswahrscheinlichkeit für jedes der 64 Tripletts in der untersuchten Sequenz.
- **Nukleotid-Übergang (NT)** Enthält die Wahrscheinlichkeit für jede mögliche Kombination aus zwei Nukleotiden X und Y, dass die Abfolge XY in der untersuchten Sequenz auftritt.
- **Triplett-Übergang (TT und TT2)** Enthält die Übergangswahrscheinlichkeiten für Tripletts. Die TT-Gewichtung enthält für jedes der 64 möglichen Tripletts die Wahrscheinlichkeiten, dass ein bestimmtes Nukleotid davor und danach auftritt. Die hier neu eingeführte TT2-Gewichtung enthält für alle Kombinationen aus je 2 Tripletts die Wahrscheinlichkeit, dass diese Tripletts in der Sequenz direkt aufeinander folgen.

3.8 Scores zur Bewertung eines genetischen Codes

Zur Bewertung, wie konservativ ein genetischer Code ist, wurden die in den vorherigen Arbeiten definierten Scores MS1, MS2, MS3, MS0 [8], rMS, lMS und fMS [6] verwendet und wie in [5] mit den Gewichtungen ergänzt. Zudem wird der kombinierte GMS-Score [5] verwendet.
Diese Scores berechnen einen Wert für verschiedene Mutationstypen, um deren Auswirkungen auf die Veränderung der Polaritäten der Aminosäuren zu beschreiben. Für die Wertungen wird folgende Notation verwendet:

- $P(c_i)$ ist die Polarität der Aminosäure, die durch das Codon c_i codiert wird.
- $P(M^j(c_i))$ ist die Polarität der Aminosäure, die durch die j-Mutation des Codons c_i codiert wird.
- m_i ist die Anzahl der möglichen Mutationen, welche das Codon c_i nicht zu einem Stoppcodon machen. Für Punktmutationen gibt es an der ersten Position 174 und an den beiden anderen 176 mögliche Mutationen. Für Shiftmutationen können je Richtung 232 verschiedene Mutationen vorkommen.
- W sind die Gewichtungen zur Mutation

Das in [9] verwendete Transition/Transversion Bias wird in dieser Arbeit nicht verwendet und wird daher zur besseren Verständlichkeit weggelassen. Der Grund dafür ist, dass gezielt nach Mechanismen gesucht wird, die durch den natürlichen Code besonders gut oder besonders schlecht konserviert werden und als Basis daher eine gleiche Häufigkeit der Mutationen angenommen wird.
Die Standardabweichung bei Mutationen wird wie folgt berechnet:

$$D_x = \sum_{i=1}^{61} \sum_{j=1}^{m_i} (P(c_i) - P(M^j(c_i))^2 W) \tag{3}$$

Die Scores MS1, MS2 und MS3 repräsentieren den quadrierten Mittelwert der Abweichung durch eine Mutation, MS0 den Mittelwert über alle drei Codonpositionen.

$$MS1 = \frac{D_1}{m_1} \ , \ MS2 = \frac{D_2}{m_2} \ , \ MS3 = \frac{D_3}{m_3} \ , \ MS0 = \frac{D_1 + D_2 + D_3}{m_1 + m_2 + m_3} \tag{4}$$

R. Geyer führte die Scores rMS, lMS und fMS ein, welche die Abweichung der Polaritäten der Aminosäuren nach einem Verschieben des Leserasters bewerten.

$$rMS = \frac{D_r}{m_r} \ , \ lMS = \frac{D_l}{m_l} \ , \ fMS = \frac{D_r + D_l}{m_r + m_l} \tag{5}$$

Um einen Score zu haben, der sowohl über die Punkt- als auch über die Shiftmutationen summiert, wird der GMS verwendet:

$$GMS = \frac{D_1 + D_2 + D_3 + D_r + D_l}{m_1 + m_2 + m_3 + m_r + m_l} \tag{6}$$

3.9 Auswahl und Vorverarbeitung der Sequenzen

Grundlagen für die Daten in dieser Arbeit sind die Sequenzen der CCDS (Consens Coding Sequence)[24] des Menschen sowie das komplette menschliche Chromosom 1 (GenBank NC_000001.11). B. Klaucke zeigte, dass die Statistiken aus Chromosom 1, dem größten Chromosom des Menschen, für das gesamte Genom repräsentativ sind. Aus diesem Grund ist es möglich, dass die Berechnungen jeweils nur mit Chromosom 1 und nicht mit dem gesamten Genom durchgeführt werden.
Um die Sequenzen besser auswerten zu können wurden sie gefiltert bzw. vorverarbeitet. Das frisch aus der Genbank geladene Chromosom 1 ist im IUPAC Sequence Format[26] codiert. BioJava kann jedoch nur mit den Nukleotiden C, T, A und G in einer DNA-Sequenz umgehen. Um die Sequenz verwenden zu können wurden daher folgende Zeichen aus der Sequenz entfernt:

- **U**: Uracil: Kann in seltenen Fällen auch in der DNA vorkommen.
- **R**: G oder A (Purin)
- **Y**: T oder C (Pyrimidin)
- **K**: G oder T (Keto)
- **M**: A oder C (Amino)
- **S**: G oder C
- **W**: A oder T
- **B**: G, T oder C
- **D**: G, A oder T
- **H**: A, C, oder T
- **V**: G, C oder A
- **N**: A, G, C oder T (jede)

Diese Zeichen bedeuten, dass dort eine Unsicherheit bei der Sequenzierung besteht. Für diese Basen eines der vier Nukleotide einzusetzen würde zu einer systematischen Verzerrung führen. Da die gesamte Sequenz betrachtet wird, und nur ein sehr geringer Teil davon tatsächlich im Leseraster abgelesen wird, wurden diese Zeichen ersatzlos aus der Sequenz entfernt. Die resultierende Sequenz ist um ca 18,5 Millionen Basen kürzer als die originale Sequenz (249 Millionen Basen).
Die von B. Klaucke verwendete Datei für die CCDS enthält diese Zeichen nicht. Eine solche Vorverarbeitung ist dort nicht nötig. Die dort enthaltenen Sequenzen wurden trotzdem noch einem Plausibilitätsfilter unterzogen. Alle 33384 in der Datei enthaltenen Sequenzen wurden nacheinander nach folgenden Kriterien gefiltert:

- 1. Die Länge der Sequenz ist durch 3 teilbar (2 Sequenzen herausgefiltert)
- 2. Beginnt mit dem Startcodon „ATG“ (41 Sequenzen herausgefiltert)

- 3. Die Sequenz endet mit einem der Stoppcodons (keine Sequenzen herausgefiltert)

Nach der Filterung blieben 33.331 Sequenzen mit einer durchschnittlichen Länge von 1703 Nukleotiden übrig, die für die weiteren Berechnungen verwendet wurden. B. Klaucke prüfte in ihrer Arbeit nicht, ob in den Sequenzen noch vor dem Ende ein Stoppcodon im Leseraster auftritt. In einer weiteren Prüfung wurden dann 55 Stoppcodons im Leseraster der Sequenzen gefunden. Bei diesen Sequenzen würde die Transkription dementsprechend bereits vor dem Ende abbrechen. Aufgrund der geringen Anzahl und um die Vergleichbarkeit mit den Ergebnissen von B. Klaucke zu bewahren wurden diese Sequenzen nicht herausgefiltert sondern weiterhin verwendet.

3.10 Verteilungs- und Übergangsmatrizen

Die Zählung der Häufigkeiten erfolgte lediglich entlang des Vorwärtsstranges der genannten Sequenzen. Bei Chromosom 1 wurde der komplementäre Strang sowie die Rückrichtung beider Stränge nicht beachtet. In den CCDS wurde die Sequenz jeweils in Leserichtung ausgewertet.
Tabelle 4 zeigt die relative Häufigkeit der 4 Nukleotide, jeweils in der codierenden Sequenz und auf dem kompletten Chromosom 1 (NA-Gewichtung).

Tabelle 4: A-priori-Wahrscheinlichkeiten der Nukleotide in der CCDS und im menschlichen Chromosom 1

Sequenz	T	C	A	G
Homo Sapiens CCDS	0,2187	0,2573	0,2620	0,2620
Homo Sapiens Chromosom 1	0,2918	0,2085	0,2910	0,2087

Die Tabellen 5 und 6 zeigen die Z-Scores der Triplett A-priori-Gewichtungen. Die diesen Tabellen zugrunde liegenden Daten haben beide ein arithmetisches Mittel von 1. Die empirische Standardabweichung ist bei Tabelle 5 $\sigma = 0,5459$ und bei Tabelle 6 $\sigma = 0,4450$.

Tabelle 5: Z-Scores zur relativen Tripletthäufigkeit in der CCDS

TTT	0,1830	TTC	0,3932	TTA	-0,8928	TTG	-0,3102
TCT	0,0044	TCC	0,2152	TCA	-0,3167	TCG	-1,3150
TAT	-0,4002	TAC	-0,1300	TAA	-1,8317	TAG	-1,8317
TGT	-0,6148	TGC	-0,4639	TGA	-1,8314	TGG	-0,4308
CTT	-0,2545	CTC	0,3399	CTA	-0,9921	CTG	2,6614
CCT	0,3028	CCC	0,4637	CCA	0,2537	CCG	-1,0294
CAT	-0,5207	CAC	-0,0987	CAA	-0,3107	CAG	2,2516
CGT	-1,2982	CGC	-0,6675	CGA	-1,0879	CGG	-0,4966
ATT	0,0810	ATC	0,5178	ATA	-0,9330	ATG	0,7080
ACT	-0,2269	ACC	0,3082	ACA	-0,0027	ACG	-1,1364
AAT	0,2371	AAC	0,3805	AAA	1,2121	AAG	1,9586
AGT	-0,3139	AGC	0,4812	AGA	-0,3812	AGG	-0,4606
GTT	-0,4959	GTC	-0,1896	GTA	-0,9653	GTG	1,3674
GCT	0,3504	GCC	1,3664	GCA	0,0963	GCG	-1,0018
GAT	0,8835	GAC	1,1257	GAA	1,8899	GAG	2,9205
GGT	-0,5735	GGC	0,6883	GGA	0,1391	GGG	0,0254

Tabelle 6: Z-Scores zur relativen Tripletthäufigkeit in Chromosom 1 der menschlichen DNA

TTT	3,1089	TTC	0,5913	TTA	0,6055	TTG	0,4639
TCT	0,9536	TCC	0,0348	TCA	0,5749	TCG	-1,9130
TAT	0,5440	TAC	-0,6604	TAA	0,6095	TAG	-0,3996
TGT	0,6133	TGC	-0,1522	TGA	0,5525	TGG	0,4787
CTT	0,6449	CTC	0,2847	CTA	-0,4118	CTG	0,7550
CCT	0,4194	CCC	-0,2548	CCA	0,4553	CCG	-1,8294
CAT	0,3243	CAC	-0,0592	CAA	0,4289	CAG	0,7807
CGT	-1,8745	CGC	-1,8858	CGA	-1,9204	CGG	-1,8263
ATT	1,1852	ATC	-0,3527	ATA	0,5453	ATG	0,3176
ACT	0,0822	ACC	-0,5466	ACA	0,5944	ACG	-1,8814
AAT	1,1667	AAC	-0,1678	AAA	3,0673	AAG	0,6118
AGT	0,0739	AGC	-0,1772	AGA	0,9669	AGG	0,4117
GTT	-0,1693	GTC	-0,8730	GTA	-0,6456	GTG	-0,0443
GCT	-0,1823	GCC	-0,4430	GCA	-0,1498	GCG	-1,8831
GAT	-0,3397	GAC	-0,8642	GAA	0,5724	GAG	0,2823
GGT	-0,5449	GGC	-0,4430	GGA	0,0519	GGG	-0,2583

Um eine systematische Struktur (z.B. ein Gen) in den Gewichtungen festhalten zu können wurden Übergangswahrscheinlichkeiten ermittelt. Aus diesen Werten lässt sich ablesen, wie wahrscheinlich das Vorkommen von zwei aufeinanderfolgenden Nukleotiden in der untersuchten Sequenz ist. Die Tabellen 7 und 8 dürfen nicht als konventionelle Übergangsmatrix gesehen werden, da sie nicht pro Zeile sondern als Ganzes normalisiert wurden. Sie enthalten somit die relative Häufigkeit einer Nukleotidabfolge verglichen mit allen Nukleotid-Übergangshäufigkeiten (NT-Gewichtung). In den Tabellen ist vertikal das erste Nukleotid und horizontal das zweite Nukleotid ablesbar.

Tabelle 7: Übergangswahrscheinlichkeiten zwischen den Nukleotiden in der CCDS

Base	T	C	A	G
T	-0,69376	0,45399	-0,59236	-1,04140
C	-0,31372	0,82832	-0,43379	0,36670
A	-1,86095	1,05512	0,63544	0,86491
G	0,95906	-1,88992	1,12107	0,54166

Tabelle 8: Übergangswahrscheinlichkeiten zwischen den Nukleotiden in Chromosom 1

Base	T	C	A	G
T	1,63414	0,43610	0,58085	-0,59348
C	-0,11980	-0,41454	-0,59995	-0,91070
A	0,03203	0,50540	1,60273	-0,11950
G	0,51125	-2,57195	0,43702	-0,40931

Tabellen 33 und 34 (im Anhang) zeigen die TT2-Gewichtungen aus der CCDS und aus Chromosom 1. Da die TT-Gewichtung lediglich ein Mittelwert über einzelne Wertegruppen der TT2-Gewichtung ist, wird diese hier nicht aufgeführt.

4 Untersuchung der Fehlertoleranz des genetischen Codes

J. Freeland und Laurence D. Hurst stellten in ihrer Arbeit „The Genetic Code Is One in a Million“.[9] die Hypothese auf, dass der genetische Code optimiert ist, um die Polarität der codierten Aminosäuren möglichst gut zu konservieren. Sie zeigten mit ihren Berechnungen, dass die Wahrscheinlichkeit, durch Zufall einen besser konservierenden Code zu finden, bei 1 zu 1.000.000 liegt. Dies würde die Tatsache erklären, dass es so gut wie keine Lebewesen gibt, die nicht den gleichen genetischen Code verwenden. Mit den Ergebnissen von Freeland und Hurst ist dies durch den Selektionsdruck der Evolution erklärbar. Ein möglichst konservativer genetischer Code bietet möglicherweise den Vorteil, dass die Polaritäten der codierten Aminosäuren, also das Produkt aus der Erbinformation, robuster gegenüber Veränderungen der Genomsequenz ist. Somit haben Lebewesen, die diesen Code verwenden, einen Vorteil gegenüber denen, die andere Codevarianten verwenden.
Doch trotz des Evolutionsdurckes hat die Natur in allen Bereichen unzählige verschiedene Varianten jeder Funktion hervorgebracht. Da diese Funktionen jedoch nicht für alle Lebewesen gleich wichtig sind, sind auch nicht so optimale Varianten noch oft zu finden (z.B. schlechte Augen beim Maulwurf). Der genetische Code ist für alle Lebewesen existenziell und der natürliche genetische Code ist insgesamt der Beste. Dies konnte jedoch bisher noch nicht bewiesen werden. Diese Arbeit zeigt jedoch ein paar Faktoren, die bei der Optimierung des Codes eine Rolle spielen könnten.
Auch die Existenz der alternativen genetischen Codes lässt sich mit der Theorie über den optimalen Code in Einklang bringen. Viele der bekannten alternativen genetischen Codes finden sich in Mitochondrien, den Organellen, die für die Regenerierung des energiereichen Moleküls Adenosintriphosphat zuständig sind. Mitochondrien befinden sich im Zellplasma, besitzen ihre eigene DNA (mtDNA) und teilen sich selbstständig. Dass sie einen leicht veränderten genetischen Code verwenden lässt sich auch mit der Endosymbiontenhypothese [13] erklären. Die Endosymbiontenhypothese besagt, dass Mitochondrien früher selbstständige Lebewesen waren, dann jedoch als Symbionten in höher entwickelten Zellen aufgenommen wurden. Geschützt durch den Wirt haben Mitochondrien jedoch keinen so starken Selektionsdruck. Dadurch konnten sie ihren alternativen genetischen Code weiterverwenden und starben nicht aus, obwohl dieser nicht so optimal ist.
Die Arbeit von R. Geyer [6] stützt die These, dass der natürliche genetische Code der optimale Code ist, indem sie zeigte, dass der natürliche genetische Code auch für Frameshift-Mutationen sehr konservierend wirkt. In [5] wurde diese These nochmals gestärkt, indem gezeigt werden konnte, dass die Anwendung von Gewichtungen aus realen genetischen Sequenzen die Anzahl der potentiell besseren Codes weiter reduzierte. Zudem wurde gezeigt, dass es zwar zufällig generierte Codes gibt, die im globalen Score GMS einen geringeren Fehlerwert erreichen, aber keine Codes, die in allen Scores (MS1, MS2, MS3, MS0, rMS, lMS, fMS) einen besseren Wert als der natürliche Code erreichen. Diese Arbeit verwendete jedoch die Gewichtungen aus dem kompletten menschlichen Chromosom 1, davon sind jedoch nahezu 99% nicht

codierend, nur ca. 1% des menschlichen Genoms machen die Exons aus [14].
B. Klaucke zeigte in ihrer Arbeit nun jedoch, dass der konservierende Effekt der Nukleotidverteilung auf die Polarität der Aminosäuren gerade in den codierenden Sequenzen nicht gilt. Dies zeigte sich, da bei Anwendung von Gewichtungen aus den codierenden Sequenzen die Wahrscheinlichkeit höher war, einen konservativeren Code zu finden als ohne die Gewichtungen.
Um die Validität dieser Resultate zu prüfen wurden diese Berechnungen im Rahmen dieser Arbeit wiederholt.

4.1 Reproduktion der Ergebnisse von B. Klaucke

B. Klaucke zeigte in ihrer Arbeit nun dass die Anwendung der Gewichtungen aus den codierenden Sequenzen dafür sorgt, dass mehr Zufallscodes einen kleineren GMS-Score besitzen als der natürliche Code. Dafür wurden die Berechnungen aus meiner Bachelorarbeit wiederholt, jedoch unter Verwendung der verschiedenen Übergangs- und A-priori-Wahrscheinlichkeiten aus den Sequenzen der CCDS. Diese Sequenzen wurden anders als Chromosom 1 vorher im Leseraster abgelesen, sodass auch nur Tripletts gezählt wurden, die auch im Leseraster zu finden sind. Mit diesen Gewichtungen versehen wurde der natürlich Code dann mit dem Random Code Set mit einer Million Zufallscodes verglichen. Tabelle 9 zeigt die Reproduktion der Ergebnisse von B. Klaucke. Die Zahlen unterscheiden sich wahrscheinlich dadurch, dass andere Implementierungen gewählt wurden. Die generelle Tendenz ist jedoch identisch.

Tabelle 9: Wahrscheinlichkeit p dass ein zufälliger Code aus dem Codeset von R. Geyer[6] einen geringeren GMS-Score besitzt als der natürliche Code. Reproduktion der Ergebnisse von B. Klaucke.

Gewichtung	Chromosom 1	CCDS	Chr1 (Klaucke)	CCDS (Klaucke)
keine	0,000037	0,000037	0,000037	0,000037
NA	0,000040	0,000041	0,000058	0,000035
TA	0,000004	0,000437	0,000010	0,000130
TT	0,000005	0,000068	0,000004	0,000072
NA+TA	0,000007	0,000489	0,000017	0,000113
NA+TT	0,000011	0,000078	0,000011	0,000063
TA+TT	0,000002	0,001731	0,000003	0,000281
NA+TA+TT	0,000003	0,001935	0,000007	0,000257

Zusätzlich zu den Berechnungen aus dem menschlichen Chromosom 1 wurde diese Berechnung noch einmal auf den kompletten Chromosomsequenzen und der CCDS von Ciona intestinalis (Schlauchseescheide) und E. Coli durchgeführt. Dies diente zur Untersuchung der Hypothese, ob dieser Effekt eventuell mit dem alternativen Splicing in der eukaryontischen Zelle zusammenhängen könnte. Das alternative Splicing wurde 1977 zuerst bei Adenoviren beobachtet [16] und widerlegt die bis dahin verfolgte These „Ein Gen ergibt ein Protein“. Es beschreibt einen Prozess der Nachverarbeitung der abgelesenen mRNA. Dabei werden verschiedene Gensegmente

(Exons) miteinander kombiniert, um so mit einem Gen mehrere verschiedene Proteine synthetisieren zu können. Anders als bei Prokaryonten, bei denen zumeist kein Splicing stattfindet, sind die Gene bei Eukaryonten fast immer in mehrere Exons aufgeteilt. Diese liegen jedoch auch nicht geordnet direkt hintereinander auf der Genomsequenz, sondern werden durch die nicht codierenden Introns voneinander getrennt. Introns wie Exons können in ihrer Länge stark variieren. Die in Tabelle 10 aufgeführten Ergebnisse zeigen, dass auch in Prokaryontischen Zellen (E. Coli) der genetische Code unter Verwendung der Gewichtungen weniger konservierend ist, also Introns oder Splicing keinen Einfluss haben.

Tabelle 10: Neuberechnung der Daten aus Tabelle 9 mit Daten aus Ciona intestinalis (Schlauchseescheide) und E. Coli. Angabe der Wahrscheinlichkeit p dass ein zufälliger Code konservativer ist als der natürliche Code.

Gewichtung	Chr1 C.Int.	CCDS C.Int.	Chr1 E. Coli	CCDS E. Coli
keine	0,000037	0,000037	0,000037	0,000037
NA	0,000049	0,000053	0,000036	0,000039
TA	0,000019	0,000683	0,000016	0,000488
TT	0,000007	0,000066	0,000022	0,000131
NA+TA	0,000058	0,000927	0,000016	0,000523
NA+TT	0,000021	0,000096	0,000022	0,000152
TA+TT	0,000033	0,002938	0,000012	0,001665
NA+TA+TT	0,000145	0,003939	0,000013	0,001921

Die Tabellen 9 und 10 bestätigen die Ergebnisse von B. Klaucke.

4.2 Untersuchung des Histogramms der polaritätsverändernden Mutationen

Um die Gründe der schlechteren Werte des natürlichen Codes zu finden, wurde ein Histogramm über alle möglichen Mutationen und deren Veränderung der Polarität (Quadrat der Differenz) erstellt. Abbildung 2 zeigt, wie viele Mutationen jeweils mit und ohne Gewichtungen zu welchen Veränderungen der Polarität führen. Dabei wurden die Werte aus Gewichtungen und die mit NA+TA+TT gewichteten Werte verglichen. Im direkten Vergleich (unterer Teil) ist zu sehen, dass sowohl mit als auch ohne Anwendung der Gewichtungen ähnliche Mutationsfehler entstehen. Bei beiden Varianten gibt es besonders viele Mutationen mit keiner oder einer sehr geringen Veränderung der Polarität der codierten Aminosäure. Dies zeigt, dass viele Mutationen bei Verwendung des natürlichen genetischen Codes still ablaufen (kein Austausch der Aminosäure) oder die durch die Mutation codierte Aminosäure eine ähnliche Polarität besitzt. Sowohl mit als auch ohne Gewichtungen treten sehr wenige Mutationen mit einer höheren gewichteten Polaritätsveränderung auf. In der Differenz der Histogramme ($Histogramm_{Gewichtungen} - Histogramm_{Ohne}$, oberer Teil) lässt sich jedoch erkennen, dass bei Mutationen mit hohen Veränderungen der Polarität (99+) eine Tendenz zu erkennen ist. Diese Mutationen fallen nur bei Anwendung der Gewichtungen so stark durch ihre Wertung auf.

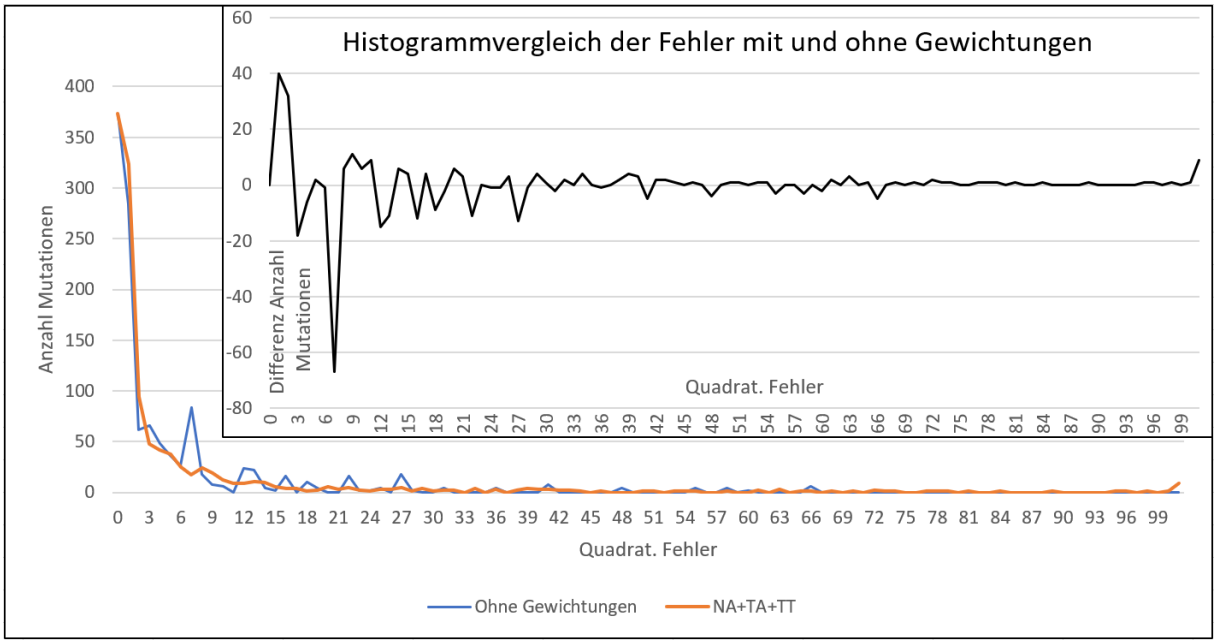

Abbildung 2: Histogramm der Fehler mit den Gewichtungsoptionen NA+TA+TT aus Tabelle 9

Betrachtet man die größten der einzelnen Summanden (Tabellen 11 und 12) die bei der Berechnung des GMS mit einfließen, so ist die in Abbildung 2 zu erkennende Erhöhung von Mutationen mit sehr hohen gewichteten Fehlerwerten deutlich erkennbar. Während die auf der gesamten Sequenz des Chromosom 1 erzeugte Gewichtungen eine maximale gewichtete Veränderung von 65,61 (Shiftmutation Asp $\leftrightarrow$ Ile) erzeugen, so sind es auf der CCDS 262,1 (Asp $\rightarrow$ Met) oder 222,28 (Glu $\rightarrow$ Gly). Dies bedeutet, dass gerade die Triplets oder Nukleotide, die für die Mutationen, bei denen höhere Veränderungen der Polarität entstehen, in der CCDS häufiger vorkommen. Dadurch werden die Fehler mit der relativen Häufigkeit multipliziert und erreichen diese höheren Werte. Zudem ist der Anteil der Shiftmutationen bei den Mutationen mit den höchsten Veränderungen der Polarität erhöht. Dies kann bedeuten, dass der genetische Code besser für Punktmutationen optimiert ist als für Shiftmutationen. Im hier verwendeten Vergleichskriterium werden alle möglichen Mutationen gleich hoch gewichtet. Eine Verwendung von realen Mutationsstatistiken als Gewichtung könnte weitere Klarheit schaffen.
Insgesamt treten mit den Gewichtungen aus der CCDS 25 Werte auf, die größer sind als das Maximum der Differenz bei Berechnung mit Gewichtungen aus Chromosom 1. Von diesen 25 „schlechten“ Mutationen sind 7 Punktmutationen und 18 Shiftmutationen. 12 gehen auf eine Mutation eines für Asp codierenden Tripletts zurück, 10 auf Glu und 3 auf Ile, Met und Val. Bis auf zwei Ausnahmen entstehen alle dieser 25 Mutationen aus einem Triplett welches mit den Nukleotiden GA beginnt. Eben diese Tripletts weisen, wie aus den Tabellen 5 und 6 zu entnehmen ist, eine erhöhte Häufigkeit in der codierenden Sequenz auf.

Tabelle 11: Top 15 Mutationen mit den höchsten gewichteten Veränderungen der Polarität, Gewichtungsdaten aus der CCDS, Gewichtungen NA+TA+TT, Δ^2 in Einheiten der Polarität

Δ^2	Mutationstyp	Mutation	Aminosäuren
262,0958	SHIFT	GAT→ATG	Asp→Met
222,2769	SHIFT	GAG→GGA	Glu→Gly
163,1373	SHIFT	GAG→AGC	Glu→Ser
149,7889	SHIFT	GAC→ACA	Asp→Thr
118,7772	SHIFT	GAC→ACC	Asp→Thr
118,3388	SHIFT	ATG→GAT	Met→Asp
106,4101	SHIFT	GAG→AGG	Glu→Arg
102,9841	SNP	GAG→GTG	Glu→Val
102,9006	SHIFT	GAT→ATA	Asp→Ile
99,5634	SHIFT	ATC→GAT	Ile→Asp
97,1776	SHIFT	GAA→GGA	Glu→Gly
95,6187	SHIFT	GAC→GGA	Asp→Gly
94,108	SHIFT	GAT→ATC	Asp→Ile
88,9365	SHIFT	GAG→AGA	Glu→Arg
83,0498	SNP	GTG→GAG	Val→Glu

Tabelle 12: Top 15 Mutationen mit den höchsten gewichteten Veränderungen der Polarität, keine Gewichtungen, Δ^2 in Einheiten der Polarität

Δ^2	Mutationstyp	Mutation	Aminosäuren
65,61	SHIFT	GAT→ATT	Asp→Ile
65,61	SHIFT	GAT→ATC	Asp→Ile
65,61	SHIFT	GAT→ATA	Asp→Ile
65,61	SHIFT	ATT→GAT	Ile→Asp
65,61	SHIFT	ATC→GAT	Ile→Asp
65,61	SHIFT	ATA→GAT	Ile→Asp
59,29	SHIFT	GAT→ATG	Asp→Met
59,29	SHIFT	ATG→GAT	Met→Asp
57,76	SNP	TAT→GAT	Tyr→Asp
57,76	SNP	TAC→GAC	Tyr→Asp
57,76	SNP	GAT→TAT	Asp→Tyr
57,76	SNP	GAC→TAC	Asp→Tyr
54,76	SNP	GTT→GAT	Val→Asp
54,76	SNP	GTC→GAC	Val→Asp
54,76	SNP	GAT→GTT	Asp→Val

Um zu prüfen, ob diese extrem erhöhten Veränderungen der Polarität der Auslöser für die schlechtere Bewertung sind, wurde die Berechnung noch einmal durchgeführt, diesmal wurden jedoch alle Werte von Δ^2, die über 66 lagen, auf den Wert 0 gesetzt. Tabelle 13 ist zu entnehmen, dass diese Berechnung zeigt, dass diese sehr hohen Werte nicht die Ursache des schlechteren Abschneidens bei Berechnungen mit Gewichtungen aus der CCDS sind.

Tabelle 13: Wahrscheinlichkeit p dass ein zufälliger Code aus dem Codeset von R. Geyer[6] einen geringeren GMS-Score besitzt als der natürliche Code. Wiederholung der Berechnungen für Tabelle 9, alle Werte der quadrierten Polaritätsveränderungen über 66 wurden ignoriert.

Gewichtung	CCDS
keine	0,000073
NA	0,000078
TA	0,000705
TT	0,000124
NA+TA	0,000769
NA+TT	0,000137
TA+TT	0,002646
NA+TA+TT	0,002915

4.3 Fazit zur Fehlertoleranz des genetischen Codes

Es scheint so, als wenn die Polaritätsveränderung der Aminosäuren bei Mutationen dem natürlichen genetischen Code einen einmaligen Selektionsvorteil gebracht hat. Im gesamten menschlichen Genom konnte jedoch beobachtet werden, dass die Häufigkeiten und Abfolgen von Nukleotiden und Tripletts diese Konservierung optimieren.
Betrachtet man jedoch nur proteincodierende Sequenzen, so ist dieser Effekt nicht mehr beobachtbar, die Wahrscheinlichkeit effizientere Codes zu finden steigt dann wieder. Diese von B. Klaucke beschriebene Beobachtung konnte im Rahmen dieser Arbeit verifiziert werden. Die für Tabelle 10 durchgeführten Berechnungen zeigten zudem, dass dieser Effekt nicht nur im menschlichen Genom zu beobachten ist. Auch bei primitiveren Eukaryonten und dem Bakterium E. Coli wurde dieser Effekt festgestellt. Dies widerlegt die Hypothese, dass die schlechtere Konservierung in codierenden Sequenzen ein Endeffekt des alternativen Splicings in eukaryontischen Zellen ist.
Eine Untersuchung der durch die Gewichtungen besonders hoch bewerteten Mutationen konnte keine besonderen Muster zeigen. Die höheren Abweichungen entstehen dadurch, dass die Mutationen mit sehr hohen Veränderungen der Polarität öfter auftreten können, somit werden die höheren Differenzen auch höher gewichtet. Eine Berechnung, bei der alle quadrierten Differenzen über 66 ignoriert wurden, senkte jedoch die Wahrscheinlichkeit nicht, zufällige konservativere Codes zu finden. Damit wurde die Hypothese widerlegt, dass es lediglich die höher gewichteten Mutationen mit hohen Veränderungen der Polarität sind, die dafür sorgen, dass der natürliche Code die Polaritäten der Aminosäuren mit Gewichtungen aus der CCDS nicht so gut konserviert.

5 Auftretenswahrscheinlichkeiten von Stoppcodons

Die bisherigen Berechnungen berücksichtigten keine Stoppcodons.
B. Klaucke legte mit ihren Berechnungen den Grundstein zu der Hypothese, dass die Verteilung von Stoppcodons, abseits ihrer Funktion ein regulär abgelesenes Transkript zu beenden, eine regulierende Funktion hat, um so z.B. nach Frameshift schnell einen Abbruch der Transkription zu erzeugen. Ein solcher Mechanismus könnte zum Beispiel dafür sorgen, dass eine Transkription bei einem Frameshift in den codierenden Sequenzen schnell in ein Stoppcodon läuft und somit abbricht und der Organismus in diesem Fehlerfall Energie spart. Eine andere Theorie ist, dass außerhalb der codierenden Sequenzen Stoppcodons häufiger auftreten und so die (nutzlose) Polypeptidkette eines dort abgelesenen Sequenzschnipsels kürzer ist.

5.1 Sequenzlänge bis zum Auftreten eines Stoppcodons

Basierend auf den statistischen Daten der TA, NT und TT2-Gewichtungen können Modelle entwickelt werden, um für jedes Triplett die durchschnittliche Distanz zum nächsten Stoppcodon oder zu einem speziellen Triplett zu berechnen. Ob und wie sich diese Distanz bei der Anwendung von Gewichtungen verändert, kann dann Aufschluss darüber geben, ob die Verteilung von Stoppcodons in einer Sequenz einem Muster folgt.

Als Modell wurde ein Baum gewählt, der von der Wurzel (ein beliebiges Triplett) ausgehend alle möglichen Kombinationen von Triplettabfolgen enthält. Die Wahl des Baumes als Struktur ist vielversprechend, da sie alle potentiellen Triplettabfolgen enthält.Die Länge der Verbindungen zwischen einem Knoten und allen seinen Unterknoten wird dann als anhand der NT und/oder TT2-Gewichtungen W definiert als $1\frac{1}{W}$. Dies bedeutet, dass die Kanten zwischen Triplettabfolgen, die häufiger in der Referenz-Sequenz vorkommen, kürzer sind und somit dort auftretende Stoppcodons höher gewichtet werden. Eine Skizze dieses Baumes ist in Abbildung 3 dargestellt.
Ein solcher Baum wurde für alle 64 Tripletts generiert und die erhaltenen Scores basierend auf der A-priori-Wahrscheinlichkeit des Wurzeltripletts (TA) in einen Gesamtscore gemittelt. Da die Abfolge von Tripletts potentiell unendlich lang sein kann, bis eines der gesuchten Codons auftritt, muss der Baum in der Tiefe limitiert werden. Die Limitierung wurde umgesetzt, indem beim Erreichen der maximalen Tiefe zurückgegeben wurde, dass das gesuchte Codon hier gefunden wurde. Diese Limitierung sorgt jedoch für Ungenauigkeiten, die Berechnungen haben jedoch gezeigt, dass der Fehler bei einer ausreichend hohen Suchtiefe gegen 0 konvergiert. Eine alternative zu dieser Begrenzung wäre, nur für die Blätter an denen tatsächlich ein Stoppcodon auftritt einen Wert zurückzugeben und in der Zusammenführung der Wahrscheinlichkeiten zu berücksichtigen, wie viele der unter dem aktuellen Knoten liegenden Blätter einen Wert zurückgegeben haben. Für die erstmalige Prüfung welche Ergebnisse interessant sein können genügt in diesem Fall jedoch die einfache Implementierung.

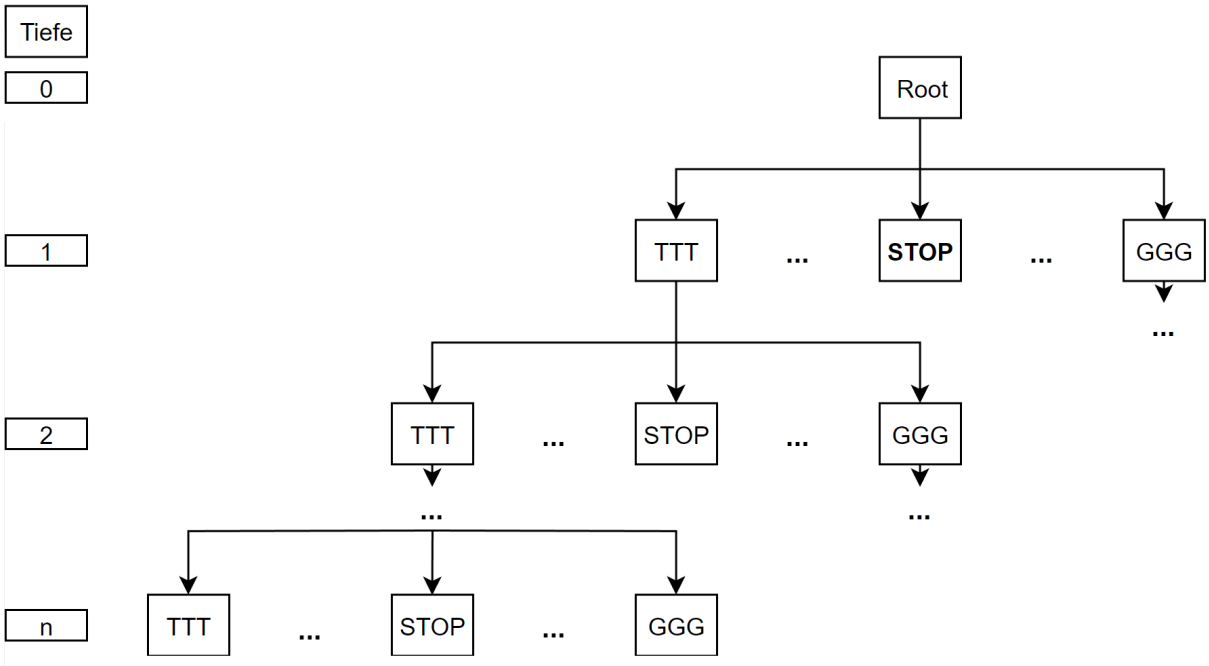

Abbildung 3: Schematische Darstellung des Baumes zur Ermittlung der mittleren Distanz zum Stoppcodon.

Der Score eines jeden Baumes wird mit der folgenden Methode rekursiv berechnet. Die Startparameter sind dabei das Codon welches die Wurzel bildet sowie die Tiefe 0.

```
private double getDistToStoppcodonRecursive(int depth,
   Codon codon){
     if(depth >= maxdepth) return maxdepth;
     if(isStoppcodonOrTarget(codon)) return depth;
     double distsum;
     for(Codon codon2 : Allcodons){
          double weight = 1/64 * getWeight(codon, codon2);
          distsum += weight *
             getDistToStoppcodonRec(depth+1, codon2);
     }
     return distsum;
}
```

Die Scores der einzelnen Bäume werden anschließend aufsummiert und jeweils mit der A-priori-Wahrscheinlichkeit des jeweiligen Tripletts gewichtet: $\frac{1}{64} \sum B(n)A(n)$ Dabei ist $A(n)$ die A-priori-Gewichtung und $B(n)$ der Score aus dem Baum für das entsprechende Triplett.

5.1.1 Begrenzung der Laufzeit

Die vorgestellte Implementation hat durch die enorme Breite des Baumes (in jeder Ebene 64x breiter) eine Laufzeit von $\mathcal{O}(n^{64})$. Dies ist bereits bei Suchtiefen von ca. 20 Tripletts ein Problem. Um die Suchtiefe zu begrenzen wurde die Tatsache genutzt,

dass viele der Sub-Bäume identisch sind. Genau genommen ist aufgrund des Aufbaus des Baumes der Effekt nutzbar, dass ein Sub-Baum mit dem gleichen Wurzeltriplett und der gleichen Start-Tiefe auch den gleichen Score besitzt. Somit kann das Ergebnis für jede Rückgabe des rekursiven Aufrufs anhand seines Codons und seiner Tiefe für die spätere Verwendung im Cache gespeichert werden. Bei jedem Aufruf der Methode kann dann geprüft werden, ob der Wert eventuell schon berechnet wurde und dieser dann verwendet werden. Eine Erweiterung der Implementierung mit diesem Ansatz reduzierte die Laufzeit auf $\mathcal{O}(n)$. Somit sind Suchtiefen von 20.000 und mehr Ebenen in wenigen Minuten berechenbar. Dafür muss bei Bedarf der Java Stack angepasst werden da jede Ebene ein Element auf den Stack legt und es so bei großen Suchtiefen zum Stack Overflow kommen kann. Die Berechnung wurde daher wie folgt an-

```
        private double getDistToStoppcodonRecursive(int depth,
           Codon codon){
            if(cache.contains(codon, depth)) return
               cache.get(codon,depth);
            if(depth >= maxdepth) return maxdepth;
            if(isStoppcodonOrTarget(codon)) return depth;
            double distsum;
gepasst:    for(Codon codon2 : Allcodons){
                double weight = 1/64 * getWeight(codon, codon2);
                distsum += weight *
                   getDistToStoppcodonRec(depth+1, codon2);
            }
            cache.add(codon,depth,distsum);
            return distsum;
        }
```

5.1.2 Alternative Begrenzung der Suchtiefe

Um sicher zu stellen, dass die Rückgabe des Tiefenwertes beim Erreichen der maximalen Tiefe das Ergebnis nicht verfälscht, wurde eine weitere Implementierungsvariante erstellt, die statt der Suchtiefe dort den Wert 0 zurück gibt. Ein direkter Vergleich zeigt jedoch, dass mit größerer Suchtiefe kein Unterschied mehr zwischen diesen Beiden Varianten besteht, da die Werte mit zunehmender Distanz zur maximalen Tiefe weniger ins Gewicht fallen. Abbildung 4 zeigt die Werte bei Berechnung des Baumes für die CCDS im Leseraster mit TT2-Gewichtungen. Klar zu erkennen ist, dass mit einer höheren Suchtiefe beide Implementationen zum gleichen Wert konvergieren. Die Variante, die die Tiefe zurück gibt, liefert besonders bei geringeren Suchtiefen bereits genauere Ergebnisse.

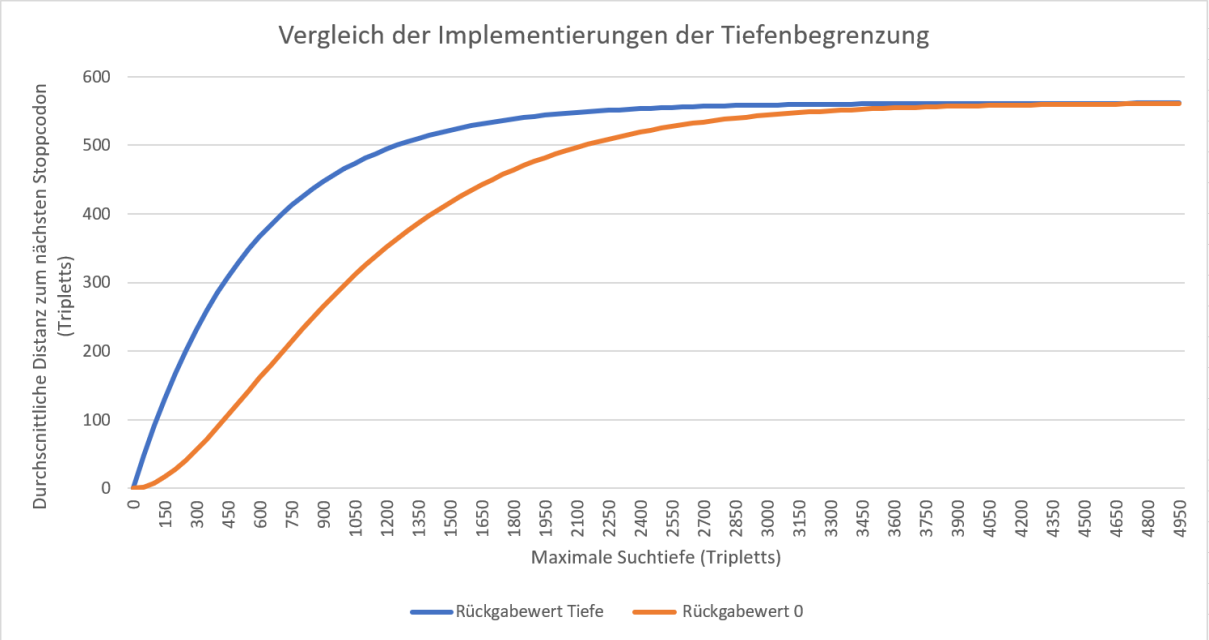

Abbildung 4: Vergleich der Implementationen zur Begrenzung der Suchtiefe im Bezug auf ihre Rückgabewerte bei verschiedenen Suchtiefen

5.2 Untersuchung der Distanz zum nächsten Stoppcodon

Bei der Berechnung der durchschnittlichen Sequenzlänge bis zu einem Stoppcodon wurde ebenfalls die Zufallssequenz verwendet. Mit dieser kann geprüft werden, ob die Ergebnisse plausibel sind. Auf einer zufälligen und exakt gleichverteilten Sequenz wären dementsprechend alle Gewichtungen gleich. Im Falle der Distanz zum Stoppcodon kann daher unter Annahme einer Gleichverteilung eine erwartete durchschnittliche Sequenzlänge berechnet werden. Dabei ist p die Wahrscheinlichkeit, dass das Triplett auftauchen soll (in diesem Fall 1), i ist die Anzahl der gesuchten Elemente (3 Stoppcodons) und n ist die Anzahl aller möglichen Elemente (hier 64 Tripletts). Die Subtraktion der Konstanten 1 gleicht die Indexverschiebung durch die Implementation des Baumes aus. Diese beginnt bei einer minimalen Distanz von 0.

$$D = \frac{p}{\frac{i}{n}} - 1 = \frac{1}{\frac{3}{64}} - 1 = 20,3333 \tag{7}$$

Das Ergebnis der Berechnungen bis zu einer maximalen Tiefe von 20.000 Ebenen ist in Tabelle 14 dargestellt.

Tabelle 14: Gewichtete Sequenzlänge in Tripletts bis zum nächsten Stoppcodon, maximale Suchtiefe von 20.000 Ebenen

Sequenz	Gewichtung	Frame 0	Frame 1	Frame 2
CCDS	NT	26,0120	25,0262	25,0623
CCDS	TT2	561,4617	22,4415	17,0909
Chromosom 1	NT	14,1391	14,1393	14,1394
Chromosom 1	TT2	18,0582	18,0520	18,0619
Random	NT	20,2838	20,2838	20,2847
Random	TT2	20,2845	20,2788	20,3054

Die in Tabelle 14 berechneten Werte der Zufallssequenz sind sehr nah an den errechneten 20,3333 Tripletts Durchschnittslänge, somit kann davon ausgegangen werden, dass die Ergebnisse valide sind.
Zwischen den Werten der CCDs und Chromosom 1 ist bereits eine deutliche Tendenz absehbar. Besonders bei Verwendung der TT2-Gewichtung fällt die Länge der im korrekten Leseraster abgelesenen CCDS auf. Der Unterschied zwischen den beiden Gewichtungsmöglichkeiten lässt sich in der Feinheit der berechneten Werte erkennen. Währen TT2-Gewichtungen in der CCDS zu sehr verschiedenen Werten für die verschiedenen Leseraster führen, ist bei der NT-Gewichtung zwar der Trend erkennbar, die Differenz jedoch nicht so stark ausgeprägt. Hier sind die NT-Gewichtungen zu feingranular um die triplettbasierte Struktur gut abbilden zu können.
Auf den ersten Blick überraschend ist der extrem große Wert von 561,4617 Tripletts, die in der CCDS im Leseraster im Schnitt vor einem Stoppcodon sind. Bei einer durchschnittlichen Länge von 1703 Nukleotiden (567,67 Tripletts) bedeutet das, dass die Abfolge der Nukleotide einem so stringenten Muster folgt, dass die in den TT2 und TA-Gewichtungen enthaltenen Informationen einem nahezu perfekt gelernten Muster entsprechen.
Bei genauerer Betrachtung der Berechnungsweise der TT2-Gewichtungen fällt jedoch auf, dass dieser Wert erwartbar ist und keine neuen Erkenntnisse birgt. Die TT2-Gewichtungen enthalten durch ihre Berechnungsweise neben den Wahrscheinlichkeiten für aufeinander folgende Tripletts auch die A-priori-Häufigkeiten der Tripletts. Bei einer durchschnittlichen Sequenzlänge von 567 Tripletts treten somit die Stoppcodons nur sehr selten auf. Die Wahrscheinlichkeit, dass auf ein beliebiges Codon ein Stoppcodon folgt, ist somit auch sehr gering. Dies folgt direkt aus der Art und Weise wie die CCDS-Gewichtungen gewonnen wurden, denn dort ist im Leseraster je Sequenz nur ein Stoppcodon zu finden.

5.2.1 Untersuchung der Sequenzlänge vor allen anderen Codons

Analog zu der Berechnung der durchschnittlichen Sequenzlänge vor Stoppcodons lässt sich der Wert auch für jedes einzelne der 64 Codons berechnen. Für die Tabellen 15 und 16 wurde die durchschnittliche Länge bis zu jedem Codon mit einer Suchtiefe von 20.000 berechnet. Die Werte in diesen Tabellen sind als durchschnittliche Anzahl anderer Tripletts vor dem jeweiligen Triplett zu verstehen.
Eine solche Berechnung wurde ebenfalls für die Zufallsequenz durchgeführt. Die

Werte dort waren sehr homogen mit einem Mittelwert von 63,0007 und einer Standardabweichung von 0,09.

Tabelle 15: Durchschnittliche Sequenzlänge (in Tripletts) vor einem Triplett in der CCDS (Mittelwert: 169,7344 σ: 428,4374)

Codon	Länge	Codon	Länge	Codon	Länge	Codon	Länge
TTT	60,1949	TTC	55,6268	TTA	126,9896	TTG	80,0682
TCT	66,4816	TCC	60,1573	TCA	80,4331	TCG	**236,9741**
TAT	85,3199	TAC	72,3396	TAA	**2078,1906**	TAG	**2689,7234**
TGT	98,8793	TGC	89,6878	TGA	**1155,2436**	TGG	86,6250
CTT	77,5256	CTC	56,5069	CTA	144,5872	CTG	27,8435
CCT	57,0050	CCC	52,6132	CCA	58,6636	CCG	**156,8200**
CAT	92,9075	CAC	70,9442	CAA	79,9668	CAG	30,2650
CGT	**227,9352**	CGC	106,1565	CGA	**162,3836**	CGG	92,8936
ATT	63,5783	ATC	50,8486	ATA	135,9331	ATG	52,5431
ACT	75,1963	ACC	57,7922	ACA	66,3035	ACG	**175,0604**
AAT	58,7200	AAC	55,6402	AAA	39,6264	AAG	32,8560
AGT	79,3936	AGC	53,8638	AGA	83,8308	AGG	89,0500
GTT	90,1657	GTC	73,6697	GTA	141,9891	GTG	38,7648
GCT	56,8178	GCC	37,7724	GCA	64,0910	GCG	**153,4811**
GAT	46,0429	GAC	40,7847	GAA	33,9094	GAG	26,6203
GGT	96,7793	GGC	47,6584	GGA	61,8628	GGG	64,4051

Tabelle 16: Durchschnittliche Sequenzlänge (in Tripletts) vor einem Triplett in Chromosom 1 (Mittelwert: 101,3659 σ: 110,5759)

Codon	Länge	Codon	Länge	Codon	Länge	Codon	Länge
TTT	27,7094	TTC	49,9049	TTA	49,8941	TTG	51,8015
TCT	44,1239	TCC	62,4974	TCA	49,6949	TCG	**424,1063**
TAT	51,0385	TAC	89,0090	TAA	49,6245	TAG	76,6602
TGT	49,2055	TGC	67,7064	TGA	50,1194	TGG	51,9638
CTT	48,7347	CTC	55,8182	CTA	77,5147	CTG	46,8289
CCT	53,3832	CCC	70,4921	CCA	52,4308	CCG	**342,3480**
CAT	54,5721	CAC	65,2612	CAA	52,3836	CAG	46,3153
CGT	**380,6846**	CGC	**397,3986**	CGA	**433,5621**	CGG	**339,7060**
ATT	41,0308	ATC	74,6224	ATA	51,2224	ATG	54,8721
ACT	60,1247	ACC	83,7851	ACA	49,4732	ACG	**387,8327**
AAT	41,4038	AAC	68,1090	AAA	28,0370	AAG	49,4523
AGT	60,4396	AGC	68,7352	AGA	44,0575	AGG	53,6146
GTT	68,0170	GTC	101,7981	GTA	88,5785	GTG	64,9880
GCT	68,8063	GCC	77,6811	GCA	67,6756	GCG	**394,2731**
GAT	74,3263	GAC	101,3857	GAA	50,4663	GAG	56,0707
GGT	83,6892	GGC	77,6669	GGA	62,0286	GGG	70,6569

Den Daten aus Tabelle 15 und 16 ist zu entnehmen, dass es einige Tripletts gibt, bei denen die durchschnittliche Sequenzlänge zwischen diesen Tripletts erheblich höher oder niedriger ist als der statistisch erwartete Mittelwert von 63. Zur besseren Übersichtlichkeit wurden alle Werte > 150 fett markiert. Betrachtet man die Tabellen im direkten Vergleich zu den TA-Häufigkeiten (Tabellen 5 und 6), so fällt auf, dass gerade die Tripletts zu den in Tabelle 15 und 16 markierten Werten eine niedrige Auftretenswahrscheinlichkeit haben. Die Z-Scores der TA-Wahrscheinlichkeiten für alle markierten Tripletts/Werte liegen alle deutlich im negativen Bereich. Auch ist ersichtlich, dass Tripletts mit einem sehr geringen Wert in Tabelle 15 oder 16 einen hohen Z-Score in der entsprechenden Tabelle zu den TA-Häufigkeiten haben (z.B das Triplett GAG im Bezug auf die CCDS).
Jedoch sind auch diese Ergebnisse lediglich auf die Auftretenswahrscheinlichkeiten der einzelnen Tripletts zurückzuführen. Im direkten Vergleich mit den Tabellen 5 und 6 ist eine deutliche Korrelation trivial erkennbar.
Um die Verfälschung der Daten durch die absolute Häufigkeit eines Tripletts zu bereinigen, wurden alle Werte der TT2-Gewichtung zusätzlich mit den TA-Häufigkeiten der beiden jeweiligen Tripletts multipliziert. Die TA-Gewichtung an der Wurzel des Baumes wurde nicht verändert. Diese „saubere“ TT2-Gewichtung wurde dann verwendet, um eine erneute Berechnung durchzuführen. Das Ergebnis dieser Berechnungen für die CCDS ist in den Tabellen 17 und 18 zu finden. Die Ergebnisse für die Zufallssequenz und Chromosom 1 befinden sich im Anhang.

Aus den Daten in Tabelle 18 ist ersichtlich, dass die Abfolge von Tripletts nicht dafür optimiert ist, dass ein Stoppcodon häufiger oder seltener auftritt. Die vorher sehr eindeutig aussehenden Werte sind lediglich auf die A-priori-Wahrscheinlichkeit der jeweiligen Tripletts zurückzuführen. In den bereinigten Werten sind die Sequenzlängen

Tabelle 17: Durchschnittliche Sequenzlänge (in Tripletts) vor einem Triplett in der CCDS (TA-bereinigt)

Codon	Länge	Codon	Länge	Codon	Länge	Codon	Länge
TTT	68,00	TTC	73,12	TTA	42,97	TTG	69,09
TCT	66,75	TCC	69,17	TCA	68,47	TCG	73,01
TAT	69,45	TAC	71,85	TAA	66,09	TAG	68,22
TGT	71,51	TGC	72,36	TGA	70,41	TGG	71,74
CTT	68,81	CTC	71,50	CTA	64,12	CTG	72,64
CCT	62,57	CCC	68,56	CCA	67,82	CCG	73,40
CAT	67,56	CAC	70,88	CAA	67,20	CAG	69,76
CGT	65,48	CGC	69,04	CGA	68,01	CGG	73,52
ATT	70,33	ATC	35,58	ATA	67,95	ATG	74,59
ACT	68,96	ACC	75,25	ACA	70,62	ACG	76,10
AAT	69,60	AAC	75,00	AAA	68,89	AAG	74,48
AGT	71,29	AGC	75,27	AGA	70,91	AGG	74,07
GTT	47,36	GTC	66,38	GTA	65,67	GTG	67,70
GCT	67,20	GCC	66,80	GCA	66,63	GCG	73,07
GAT	67,93	GAC	67,91	GAA	67,22	GAG	71,59
GGT	55,81	GGC	57,74	GGA	62,86	GGG	60,04

vor Stoppcodons nicht bedeutend länger oder kürzer als vor anderen Codons.

Tabelle 18: Z-Scores zur durchschnittlichen Sequenzlänge (in Tripletts) vor einem Triplett in der CCDS (TA-bereinigt), Mittelwert: 67,8731 σ: 7,0836

Codon	Länge	Codon	Länge	Codon	Länge	Codon	Länge
TTT	0,02	TTC	0,74	TTA	-3,52	TTG	0,17
TCT	-0,16	TCC	0,18	TCA	0,08	TCG	0,72
TAT	0,22	TAC	0,56	TAA	-0,25	TAG	0,05
TGT	0,51	TGC	0,63	TGA	0,36	TGG	0,55
CTT	0,13	CTC	0,51	CTA	-0,53	CTG	0,67
CCT	-0,75	CCC	0,10	CCA	-0,01	CCG	0,78
CAT	-0,04	CAC	0,42	CAA	-0,10	CAG	0,27
CGT	-0,34	CGC	0,17	CGA	0,02	CGG	0,80
ATT	0,35	ATC	-4,56	ATA	0,01	ATG	0,95
ACT	0,15	ACC	1,04	ACA	0,39	ACG	1,16
AAT	0,24	AAC	1,01	AAA	0,14	AAG	0,93
AGT	0,48	AGC	1,04	AGA	0,43	AGG	0,87
GTT	-2,90	GTC	-0,21	GTA	-0,31	GTG	-0,02
GCT	-0,09	GCC	-0,15	GCA	-0,18	GCG	0,73
GAT	0,01	GAC	0,01	GAA	-0,09	GAG	0,52
GGT	-1,70	GGC	-1,43	GGA	-0,71	GGG	-1,11

5.3 Stoppcodon-Mutationswahrscheinlichkeit

Bei Mutationen können codierende Tripletts in ein Stoppcodon mutiert werden. Dies kann jedoch nur bei wenigen Mutationen passieren. Über alle möglichen Punktmutationen gesammelt resultieren 9 Mutationen an der ersten Codonposition und jeweils 7 Mutationen an der 2. und 3. Codonposition in einem der Stoppcodons. Über alle möglichen Frameshift-Mutationen können dabei je Richtung 12 Stoppcodons entstehen.

Die Anzahl der durch Punktmutationen resultierenden Stoppcodons kann auch mit Gewichtungen versehen werden. Die TA und NA - Gewichtungen sind dafür geeignet, da sie beschreiben wie oft das Nukleotid welches ausgetauscht wird, beziehungsweise das Triplett welches mutiert wird, in der Sequenz vorkommt. Analog dazu können die Gewichtungen NT und TT auf die Leserasterverschiebung angewendet werden.

Tabelle 19: Gewichtete Häufigkeit von Punktmutationen, die das mutierte Codon zu einem Stoppcodon machen. Die erste Zeile enthält die ungewichtete Anzahl möglicher Mutationen.

Gewichtungsquelle	Gewichtung	1. Base	2. Base	3. Base
—	—	9	7	7
CCDS	NA	9.37572	6.76030	6.76030
CCDS	TA	13.04897	4.34338	5.33463
CCDS	NA+TA	13.61326	4.21098	5.17658
Chromosom 1	NA	8.49893	6.83803	6.83803
Chromosom 1	TA	11.21688	7.66128	7.35544
Chromosom 1	NA+TA	11.03982	7.64493	7.39379

Tabelle 20: Gewichtete Häufigkeit von Frameshiftmutationen, die das mutierte Codon zu einem Stoppcodon machen. Die erste Zeile enthält die ungewichtete Anzahl möglicher Mutationen.

Gewichtungsquelle	Gewichtung	Links	Rechts
—	—	12	12
CCDS	NT	9.10071	14.75080
CCDS	TT	14.41665	10.90651
CCDS	NT+TT	14.17319	13.41347
Chromosom 1	NT	12.74718	14.49374
Chromosom 1	TT	13.35700	13.35700
Chromosom 1	NT+TT	14.27132	16.27622

B. Klaucke untersuchte in ihrer Arbeit ebenfalls die Wahrscheinlichkeit, dass Mutationen zu einem neuen Stoppcodon führen. Sie führte die Berechnungen nur für Punktmutationen durch. Für diese Arbeit wurden nun auch die Frameshift-Mutationen betrachtet. Die hier berechneten Daten decken sich jedoch nicht mit Klauckes Ergebnissen. Eine Ursache dafür konnte bisher nicht gefunden werden, da eine Einsicht in die Berechnungsweise ihres Algorithmus nicht möglich war.
Tabelle 19 zeigt, dass die A-priori-Wahrscheinlichkeiten der CCDS Mutationen an der ersten Codonposition höher „bestrafen“ als die Wahrscheinlichkeiten aus Chromosom 1. An der zweiten Codonposition ist jedoch die Wahrscheinlichkeit in der CCDS geringer, dass eine Mutation ein Stoppcodon erzeugt. Auch an der dritten Codonposition treten mit den Statistiken der CCDS weniger neue Stoppcodons auf, jedoch etwas mehr als an Codonposition 2.
Tabelle 20 zeigt die gewichtete Anzahl der entstehenden Stoppcodons bei leserastervschiebenden Mutationen. Sowohl bei Gewichtungen aus der CCDS, als auch bei denen aus Chromosom 1 erhöht die Anwendung der Gewichtung die Anzahl entstehender Stoppcodons. Dabei wird die Anzahl bei Shifts in beide Richtungen erhöht. Besonders die Anwendung beider Übergangswahrscheinlichkeiten sorgt für eine höhere Anzahl Stoppcodons. Dieser Effekt ist jedoch bei Anwendung der Gewichtungen aus Chromosom 1 stärker ausgeprägt als bei denen aus der CCDS.

5.4 Absolute Anzahl von Stoppcodons

Als Grundlage für weitere Untersuchungen zu den unterschiedlichen Verteilungen von Stoppcodons in der gesamten Sequenz des menschlichen Chromosom 1 und der CCDS wurde die Anzahl der Stoppcodons in Relation zur Anzahl aller betrachteten Codons ermittelt. Diese Berechnung fasst die Werte der A-priori-Wahrscheinlichkeit zusammen. Die Ergebnisse dieser Berechnungen sind in Tabelle 21 aufgeführt.

Tabelle 21: Durchschnittliche Anzahl der Stoppcodons auf 100 betrachtete Codons

Sequenz	Im Leseraster gelesen	Verschiebung	Anzahl
Chromosom 1	ja	0	5,21787
Chromosom 1	ja	1	5,21906
Chromosom 1	ja	2	5,21580
Chromosom 1	nein	-	5,21758
CCDS	ja	0	0,17647
CCDS	ja	1	4,26036
CCDS	ja	2	5,62157
CCDS	nein	-	3,35093

Im Chromosom 1 macht es erwartungsgemäß keinen Unterschied, in welchem Leseraster gezählt wird. Die Werte liegen relativ dicht an der bei einer gleichverteilten Sequenz erwarteten Menge von 4,6875 $((3/64) * 100)$. Auf der CCDS finden sich im korrekten Leseraster - ebenfalls vorhersehbar - nur sehr wenige Stoppcodons. De Jong und Ryden stellten in ihrer Arbeit „Causes of more frequent deletions than insertions in mutations and protein evolution.“[18] vor, warum häufiger Deletionen als Insertionen auftreten. Die Ergebnisse aus Tabelle 21 lassen sich mit dieser Häufigkeitsinformation durchaus erklären. Bei den häufigeren Deletionen (Verschiebung -1, in dieser Arbeit äquivalent zu +2) ist die Wahrscheinlichkeit höher, dass die Transkription schneller abbricht als bei Insertionen. Somit könnte die Zelle Energie sparen.

5.5 Fazit zur Rolle von Stoppcodons im genetischen Code

Die Berechnung der durchschnittlichen Sequenzlänge bis zum Auftreten eines Stoppcodons brachte zwar auf den ersten Blick interessante Ergebnisse, bei genauerer Betrachtung wurde jedoch ersichtlich, dass die Werte lediglich ein Resultat der ungleich verteilten Sequenzen waren. Die vermeintlich langen Sequenzlängen vor einem Stoppcodon ist bei Verwendung von Gewichtungen aus der CCDS waren ebenfalls auf die extrem geringe Häufigkeit von Stoppcodons in den Sequenzen der CCDS zurückzuführen.
Eine Bereinigung der Daten von den absoluten Häufigkeiten der Tripletts (TA) zeigte dann, dass nach Eliminierung dieser Verzerrung die Ergebnisse keine eindeutigen Tendenzen mehr besitzen. Die Varianz der Ergebnisse sank ebenfalls deutlich. Betrachtet man die Z-Scores dieser Ergebnisse, finden sich jedoch ein paar Details, die ggf. doch noch Informationen bergen können. Die Z-Werte liegen bei den meisten Tripletts im Bereich zwischen -1 und 1, Ausreißer nach oben sind nicht zu beobachten, jedoch gibt es einige Tripletts, die einen sehr geringen Z-Score erreichen (z.B. ATC, GTT und TTA). Diese Tripletts treten nach den Daten in Tabelle 5 weder besonders häufig noch besonders selten auf. Eine Untersuchung hier könnte noch weitere Ergebnisse bringen.
Bei der Wahrscheinlichkeit, dass Mutationen ein Stoppcodon entstehen lassen, gab es ebenfalls keine deutlichen Ergebnisse. An bestimmten Positionen sinkt sie, an anderen steigt sie, eine Tendenz lässt sich nicht ohne Weiteres erkennen. Auch bei

Leserasterverschiebungen waren die Ergebnisse nicht klar. Die Ergebnisse decken sich nicht mit denen von B. Klaucke. Eine Ursache dafür konnte mangels Code-Einsicht nicht gefunden werden.
Die absolute Anzahl der Stoppcodons wurde ebenfalls untersucht, hier wurde erwartungsgemäß festgestellt, dass auf der CCDS im Leseraster gelesen so gut wie keine Stoppcodons existieren. Bei einen Rechtsshift wurden weniger, bei einem Linksshift mehr Stoppcodons als erwartet gefunden. Eine mögliche Erklärung dafür ist, dass die Zelle bei häufiger auftretenden Deletionen durch ein schnelles Abbrechen der Transkription Energie spart.

6 Erweiterung der Vergleichsparameter

Das Ergebnis der in dieser Arbeit verifizierten Berechnungen von B. Klaucke steht im Widerspruch zu der bisher verfolgten Hypothese, dass die genetische Sequenz zusammen mit dem genetischen Code dafür optimiert ist, dass bei Mutationen die Veränderung der Polaritäten der codierten Aminosäuren minimiert wird.
Dies wirft dann die Frage auf, welche Mechanismen denn dafür sorgen, dass es gerade in der codierenden Sequenz eine höhere Fehleranfälligkeit gibt.
David Haig und Laurence D. Hurst verwendeten in ihrer Arbeit „A Quantitative Measure of Error Minimization in the Genetic Code“[8] neben den Polaritäten auch weitere Faktoren. Sie untersuchten, welche der dort verwendeten Faktoren durch den natürlichen Code besonders konserviert werden. Sie fanden heraus, dass gerade die Polarität besonders gut konserviert wird. Doch auch die Hydrophobizität der Aminosäuren wird im Vergleich zu 10.000 Zufallscodes vergleichsweise gut konserviert, nur 89 Codes waren konservierender als der natürliche Code. Sie betrachteten die einzelnen Faktoren getrennt, bedachten jedoch nicht die Möglichkeit, dass der genetische Code auch optimiert dafür sein könnte, mehrere Faktoren zu optimieren. Um dies zu prüfen wurden nun die Daten der Hydrophobizität [12] aus Tabelle 2 verwendet und mit den Polaritätsdaten kombiniert.
Da die Daten der Polarität eine andere Varianz als die der Hydrophobizität haben, wurden für diese Berechnungen die Abweichungen als ein Vielfaches der Varianz ausgegeben. Somit werden die beiden Faktoren gleich stark ins Endergebnis mit einbezogen. Die Berechnung der Scores basiert dafür statt auf der quadratischen Differenz der Polaritäten ($\delta = (Polar_{old} - Polar_{new})^2$) auf der folgenden Formel:

$$\delta = (\frac{Polar_{old} - Polar_{new}}{\sigma_{Polar}})^2 + (\frac{Hydro_{old} - Hydro_{new}}{\sigma_{Hydror}})^2 \tag{8}$$

Die Ergebnisse der Berechnungen sind in Tabelle 22 aufgeführt. Ihnen ist zu entnehmen, dass der Anteil der besseren Codes mit den Gewichtungen aus der CCDS nicht ganz so hoch ist wie bei den polaritätsbasierten Berechnungen, jedoch nicht geringer als die Gewichtung aus dem gesamten Chromosom 1. Dies deutet darauf hin, dass die Konservierung der Hydrophobizität als Ergänzung zur Konservierung der Polarität der codierten Aminosäuren in den codierenden Sequenzen durchaus eine evolutionäre Relevanz haben könnte. Die beiden Charakteristika wurden zusammen untersucht, um zu prüfen, ob eventuell die Verwendung von beiden Faktoren dem natürlichen Code einen Vorteil verschafft. Dies könnte der Fall sein, wenn z.B. die Codes, die bezüglich der Polarität konservativer sind als der natürliche Code, bei Betrachtung der Hydrophobizität weniger konservativ sind.

Tabelle 22: Untersuchung der Konservierung des genetischen Codes im Bezug auf Polarität und Hydrophobizität der codierten Aminosäuren. Angabe der Wahrscheinlichkeit p dass ein zufälliger Code konservativer ist, als der natürliche Code.

Gewichtung	Chromosom 1	CCDS
keine	0,000405	0,000405
NA	0,000472	0,000371
TA	0,000038	0,000273
TT	0,000038	0,000095
NA+TA	0,000054	0,000250
NA+TT	0,000070	0,000085
TA+TT	0,000021	0,000246
NA+TA+TT	0,000037	0,000246

Als weitere Untersuchung wurde daher die Berechnung nur mit der Hydrophobizität durchgeführt. Das Ergebnis dazu ist in Tabelle 23 dargestellt.

Tabelle 23: Untersuchung der Konservierung des genetischen Codes im Bezug auf die Hydrophobizität der codierten Aminosäuren. Angabe der Wahrscheinlichkeit p dass ein zufälliger Code konservativer ist, als der natürliche Code.

Gewichtung	Chromosom 1	CCDS
keine	0,033357	0,033357
NA	0,037333	0,030332
TA	0,007598	0,006819
TT	0,007637	0,005791
NA+TA	0,010307	0,006099
NA+TT	0,010618	0,005132
TA+TT	0,005103	0,002093
NA+TA+TT	0,007856	0,001949

Die Werte in den Tabellen 22 und 23 bestätigen die Ergebnisse von Haig und Hurst [8]. Dort wurde eine Wahrscheinlichkeit von 0,0089 für das Finden von konservativeren Codes berechnet, jedoch nur über die Punktmutationen. Auch bei der Erweiterung der Berechnungen auf Frameshift-Mutationen sind ca 97% der Zufallscodes weniger konservativ als der natürliche Code im Bezug auf die Hydrophobizität. Zudem ist ersichtlich, dass die Verteilungsstatistiken aus Chromosom 1 wie bei der Polarität dazu führen, dass weniger Codes konservativer sind.
Der Effekt, dass die codierenden Sequenzen nicht dafür optimiert sind ist in Bezug auf die Hydrophobizität nicht beobachtbar. Im Gegenteil: die Statistiken der codierenden Sequenzen sorgen dafür, dass weniger Zufallscodes konservativer sind.
Aufgrund dieser Ergebnisse wurden die Berechnungen noch auf die in Tabelle 3 aufgelisteten Charakteristika erweitert. Tabelle 24 zeigt, wie der genetische Code und die Verteilungen der Nukleotide im menschlichen Genom die verschiedenen Faktoren konservieren. Die Werte geben die Wahrscheinlichkeit p an, dass ein zufälliger Code aus dem Random Code Set einen kleineren GMS besitzt.

Tabelle 24: Untersuchung der Konservierung des genetischen Codes im Bezug auf verschiedene Charakteristika der codierten Aminosäuren. Für die Sequenzen von Chromosom 1 und CCDS wurde die Gewichtungskombination NA+TA+TT verwendet. Angabe der Wahrscheinlichkeit p dass ein zufälliger Code konservativer ist, als der natürliche Code.

Merkmal	Ohne Gewichtungen	Chr. 1	CCDS
Mol. Volumen	0,4243	0,1616	0,1304
M_r	0,5310	0,1402	0,1588
pK_a	0,4034	0,2002	0,2270
pK_b	0,5526	0,3506	0,3294
pI	0,7717	0,5530	0,9418

6.1 Fazit zu den erweiterten Vergleichsparametern

Betrachtet man anstelle der Polaritäten der Aminosäure ihre Hydrophobizität, so ist der von B. Klaucke gezeigte Effekt, dass die Konservierung bei Anwendung der Gewichtungen aus der CCDS schwächer ist, nicht zu beobachten. Eine Anwendung der Gewichtungen aus Chromosom 1 sorgt dafür, dass sich die Anzahl der konservativeren Codes verringert. Wenn Gewichtungen aus der CCDS verwendet werden ist diese Reduktion sogar noch etwas stärker. Daraus lässt sich die Hypothese herleiten, dass der genetische Code nicht nur für die Konservierung der Polarität der Aminosäuren optimiert ist, sondern auch für andere Faktoren wie die Hydrophobizität der Aminosäuren.
Tabelle 24 zeigt diesbezüglich noch einige neue Erkenntnisse. Für die Merkmale Molekularvolumen, M_r, pK_a und pK_b ist klar erkennbar, dass die Anwendung der Gewichtungen zu einer deutlichen Reduktion der konservativeren Codes führt. Ob jedoch Gewichtungen aus Chromosom 1 oder der CCDS verwendet werden, macht dort keinen großen Unterschied. Anders sieht es bei dem pH-Wert am isoelektrischen Punkt (pI) aus. Hier reduzieren die Gewichtungen von Chromosom 1 die Anzahl der konservativeren Codes, während die Gewichtungen aus der CCDS diese Anzahl erhöhen. Dies deutet dort auf einen umgekehrten Mechanismus hin, der Codes bevorzugt, die im Fehlerfall hohe Veränderungen erzeugen.
Vermutlich gibt es aber noch eine Reihe weiterer Faktoren, die durch den genetischen Code optimal konserviert werden. Erst wenn man alle dieser Faktoren identifizieren und gemäß ihrer biologischen Wichtigkeit gewichten könnte, müsste es möglich sein, nachzuweisen, dass sich mit diesen Möglichkeiten keine konservativeren Codes mehr finden lassen. Durch diesen Nachweis hätte man dann die für die Universalität des natürlichen genetischen Codes verantwortlichen Selektionskriterien identifiziert.

7 Zusammenfassung

Die Sequenz der codierenden DNA des Menschen ist nicht dafür optimiert, um mithilfe des genetischen Codes die Robustheit gegen Mutationen im Hinblick auf die Polaritäten der Aminosäuren zu optimieren. Im Vergleich mit einer Million Zufallscodes steigt die Anzahl der Zufallscodes, die einen geringeren GMS-Score erreichen, wenn man statt der Gewichtungen aus der Sequenz des Chromosoms 1 die CCDS verwendet.
Damit konnte ein wesentlicher Teil der Arbeit von B. Klaucke reproduziert werden. Die Ergebnisse dazu sind zudem auch bei den Genomen von E. Coli und Ciona intestinalis reproduzierbar. Somit konnte ausgeschlossen werden, dass der für diese Ergebnisse verantwortliche Effekt nur in höher entwickelten Lebewesen auftritt. Bei einer genaueren Betrachtung, welche Mutationen bei Anwendung der Gewichtungen aus der CCDS besonders hoch gewichtet wurden, konnte jedoch kein erklärbares Muster gefunden werden.

Die auf den ersten Blick vielversprechenden Ergebnisse bei der Untersuchung der Auftretenshäufigkeiten von Stoppcodons bargen bei genauerer Betrachtung keine neuen Informationen. Lediglich die Betrachtung der Mutationen die Stoppcodons erzeugen zeigt einige interessante Resultate. So ist es bei Punktmutationen von der Position im Codon abhängig, ob die Verteilungen der CCDS dafür sorgen, dass die Mutationen die ein Stoppcodon erzeugen häufiger auftreten können. An der ersten Codonposition steigt die Wahrscheinlichkeit, Stoppcodons zu erzeugen, an der 2. und 3. Position sinkt sie im Vergleich zu einer gleich verteilten Sequenz. Mit den Gewichtungen aus Chromosom 1 steigt die Stoppcodon-Mutationswahrscheinlichkeit an der ersten Position ebenfalls, an den anderen beiden verändert sie sich fast nicht. Bei den leserasterverschiebenden Mutationen erhöht sich bei der Anwendung der Gewichtungen die Anzahl der entstehenden Stoppcodons in beiden Richtungen bei beiden Sequenzen als Gewichtungsquelle. Bei der Betrachtung der absoluten Anzahl der Stoppcodons wurde wie zu erwarten festgestellt, dass im Leseraster der CCDS nur sehr wenige Stoppcodons existieren, jedoch bei einem nach Rechts verschobenen Leseraster wurden nicht so viele Stoppcodons wie bei einem nach links verschobenen Leseraster gefunden.

Die Betrachtung der erweiterten Vergleichsparameter wie Hydrophobizität, Molekularvolumen oder pH-Wert am isoelektrischen Punkt zeigten, dass der in Kapitel 4 aufgeführte Effekt, dass der genetische Code die Polaritäten der Aminosäuren in den codierenden Sequenzen schlechter konserviert, nicht für einige der neuen Parameter gilt. Verwendet man Gewichtungen aus der CCDS, so sinkt die Wahrscheinlichkeit, einen konservativeren Code zu finden im Vergleich zu den Gewichtungen aus Chromosom 1 für das Molekularvolumen und die Hydrophobizität. Bei dem pH-Wert am isoelektrischen Punkt konnte ein sehr starker Trend in die entgegengesetzte Richtung festgestellt werden. Mit Chromosom 1 waren ca 55% der Zufallscodes konservativer, bei der CCDS waren es über 94%.

8 Ausblick

Die bisherigen Berechnungen stützen sich stark auf die proteincodierende Funktion der DNA. Doch diese Funktion wird nur von etwa einem Prozent der Sequenz ausgeübt. Andere Sequenztypen wie Transposons, virale Fragmente, Introns der codierenden Regionen, Centromere, Telomere, repetetive Sequenzen und einige Andere wurden bisher nicht berücksichtigt. Eine Neuberechnung mit Statistiken aus solchen Sequenzen könnte weitere interessante Ergebnisse bringen.
Auch CpG-Inseln wurden bei diesen Berechnungen bisher nicht verwendet, jedoch ist durch zahlreiche Publikationen eine Anhäufung dieser Sequenzelemente vor codierenden Sequenzen gefunden worden. Ob der genetische Code CpG-Inseln ebenfalls konserviert, bleibt noch zu prüfen.
Auch betrachten die meisten Arbeiten bisher nur die Polarität der Aminosäuren. Diese Arbeit konnte zeigen, dass gerade die Gewichtungen aus der CCDS die Hydrophobizität und einige andere Charakteristika der Aminosäuren im natürlichen Code besser konservieren. Eine Erweiterung der Berechnungen auf andere Charaktereigenschaften der Aminosäuren bietet eine Vielzahl an interessanten Ansätzen.
Anders als vorherige Arbeiten wurden Transitionen und Transversionen als gleich verteilt angenommen. Dies begrenzt zwar die erzeugte Datenmenge und ermöglicht eine bessere Identifikation der Faktoren die durch den natürlichen genetischen Code optimiert werden, führt jedoch zu Ergebnissen, die weniger nah an der realen Biologie sind. Weitere Berechnungen mit Daten zu Transition/Transversion Bias dürften viele Trends zeigen. In der Arbeit „Estimate of the Mutation Rate per Nucleotide in Humans“[19] präsentieren die Autoren Daten, die für eine solche Berechnung geeignet sein könnten.
Die Betrachtung der Rolle der Stoppcodons bezieht sich in dieser Arbeit stets nur auf den natürlichen genetischen Code und seine Optimierung in realen Sequenzen. Ein Vergleich, wie die in Kapitel 5 gezeigten Berechnungen mit anderen möglichen Stoppcodons oder genetischen Codes aussehen, klingt ebenfalls interessant.
Bis auf die Kombination der Hydrophobizität mit der Polarität wurden alle Charakteristika der Aminosäuren getrennt betrachtet. Hier wäre eine Untersuchung interessant, die prüft, ob die Zufallscodes die im Bezug auf die Polarität konservativer sind als der natürliche Code, auch bei Betrachtung der anderen Charakteristika diese besser konserviert.

9 Anhang

Der gesamte Code dieser Arbeit ist öffentlich auf GitHub verfügbar. Zu finden ist er unter `https://github.com/Phreag/DNA_Distribution_Analysis`. Das Repository enthält zudem eine PDF-Version dieser Arbeit.
Zur besseren Reproduzierbarkeit sind in Tabelle 25 die für diese Arbeit verwendeten Methodenaufrufe dokumentiert. Alle dort genannten Methoden befinden sich in der Klasse `MainClass.java` und sind statisch sodass sie direkt aus der main-Methode aufgerufen werden können.

Um längere Sequenzen wie das menschliche Chromosom 1 verarbeiten zu können, ist es nötig den Java Heapspace zu vergrößern. Für tiefe Rekursionsaufrufe ist außerdem eine Erhöhung des Stacks nötig.
In dieser Arbeit wurden Java dafür die Programmparameter `-Xmx8G -Xss8m` mitgegeben.

Tabelle 25: Verwendete Codeaufrufe zur Berechnung der Daten

Methodenaufruf	Daten/Verwendung
`compareNA_CCDS_CHR1()`	Tabelle 4
`nonsenseMutationCount()`	Tabellen 19 und 20
`compareRandomCodesEColi()`	Tabelle 10
`compareRandomCodesCionaI()`	Tabelle 10
`compareNT_CCDS_CHR1()`	Tabellen 7 und 8
`calculateErrorHistogram()`	Abbildung 2
`StoppcodonMarkovChain()`	Tabelle 14
`getTA_ZScores()`	Tabellen 5 und 6
`StoppcodonMarkovChainCompareImpl()`	Abbildung 4
`countStoppcodonsInSequences()`	Tabelle 21
`getAverageCCDSSequenceLength()`	Durschnittl. Länge der CCDS Sequenzen
`generateRandomChromosome()`	Generierung der Zufallssequenz
`getAverageDistToEachCodon()`	Tabellen zur Sequenzlänge zu jedem Codon
`cleanTT2Weightings()`	Befreiung der TT2-Gewichtungen von TA
`getAverageToEachCodonTA_Cleared()`	Anwendung der TA-befreiten TT2-Gewichtungen
`millionHydropathyAndPolar()`	Tabelle 22
`millionHydropathyOnly()`	Tabelle 23
`millionCutOffHighDeltas()`	Tabelle 13
`millionOtherAminoAcidProperties()`	Tabelle 24

Tabelle 26: Verwendete Nucleotidsequenzen aus der GenBank

Name	ID/Suchname
Chromosom 1 des Menschen	NC_000001.11
E. Coli CCDS	Escherichia_coli.HUSEC2011CHR1.cdna.all
E. Coli Chromosom 1	Escherichia_coli.HUSEC2011CHR1.dna.chromosome
C. Intestinalis CCDS	Ciona_intestinalis_CCDS
C. Intestinalis Chromosom 1	NC_020166.2

Tabelle 27: Sequenzlänge (in Tripletts) zwischen zwei gleichen Tripletts (TA-Bereinigt, CCDS)

TTT	68,00	TTC	73,12	TTA	42,97	TTG	69,09
TCT	66,75	TCC	69,17	TCA	68,47	TCG	73,01
TAT	69,45	TAC	71,85	TAA	66,09	TAG	68,22
TGT	71,51	TGC	72,36	TGA	70,41	TGG	71,74
CTT	68,81	CTC	71,50	CTA	64,12	CTG	72,64
CCT	62,57	CCC	68,56	CCA	67,82	CCG	73,40
CAT	67,56	CAC	70,88	CAA	67,20	CAG	69,76
CGT	65,48	CGC	69,04	CGA	68,01	CGG	73,52
ATT	70,33	ATC	35,58	ATA	67,95	ATG	74,59
ACT	68,96	ACC	75,25	ACA	70,62	ACG	76,10
AAT	69,60	AAC	75,00	AAA	68,89	AAG	74,48
AGT	71,29	AGC	75,27	AGA	70,91	AGG	74,07
GTT	47,36	GTC	66,38	GTA	65,67	GTG	67,70
GCT	67,20	GCC	66,80	GCA	66,63	GCG	73,07
GAT	67,93	GAC	67,91	GAA	67,22	GAG	71,59
GGT	55,81	GGC	57,74	GGA	62,86	GGG	60,04

Tabelle 28: Z-Scores zur Sequenzlänge zwischen zwei gleichen Tripletts (TA-Bereinigt, CCDS), Mittelwert: 67,8731 Sigma: 7,0836

TTT	0,02	TTC	0,74	TTA	-3,52	TTG	0,17
TCT	-0,16	TCC	0,18	TCA	0,08	TCG	0,72
TAT	0,22	TAC	0,56	TAA	-0,25	TAG	0,05
TGT	0,51	TGC	0,63	TGA	0,36	TGG	0,55
CTT	0,13	CTC	0,51	CTA	-0,53	CTG	0,67
CCT	-0,75	CCC	0,10	CCA	-0,01	CCG	0,78
CAT	-0,04	CAC	0,42	CAA	-0,10	CAG	0,27
CGT	-0,34	CGC	0,17	CGA	0,02	CGG	0,80
ATT	0,35	ATC	-4,56	ATA	0,01	ATG	0,95
ACT	0,15	ACC	1,04	ACA	0,39	ACG	1,16
AAT	0,24	AAC	1,01	AAA	0,14	AAG	0,93
AGT	0,48	AGC	1,04	AGA	0,43	AGG	0,87
GTT	-2,90	GTC	-0,21	GTA	-0,31	GTG	-0,02
GCT	-0,09	GCC	-0,15	GCA	-0,18	GCG	0,73
GAT	0,01	GAC	0,01	GAA	-0,09	GAG	0,52
GGT	-1,70	GGC	-1,43	GGA	-0,71	GGG	-1,11

Tabelle 29: Sequenzlänge (in Tripletts) zwischen zwei gleichen Tripletts (TA-Bereinigt, Chr1)

TTT	70,24	TTC	64,45	TTA	70,15	TTG	64,16
TCT	63,00	TCC	63,24	TCA	63,62	TCG	58,45
TAT	69,30	TAC	67,10	TAA	69,97	TAG	66,80
TGT	63,74	TGC	61,64	TGA	62,24	TGG	59,61
CTT	64,36	CTC	59,58	CTA	66,86	CTG	60,99
CCT	60,72	CCC	58,90	CCA	61,01	CCG	51,70
CAT	64,95	CAC	62,88	CAA	66,41	CAG	62,32
CGT	60,67	CGC	53,40	CGA	59,22	CGG	55,54
ATT	66,84	ATC	62,84	ATA	70,27	ATG	64,41
ACT	63,93	ACC	63,22	ACA	65,07	ACG	59,62
AAT	69,27	AAC	66,27	AAA	72,30	AAG	66,09
AGT	62,51	AGC	62,43	AGA	62,25	AGG	59,63
GTT	68,34	GTC	66,00	GTA	69,97	GTG	64,57
GCT	64,42	GCC	59,55	GCA	66,71	GCG	54,28
GAT	64,92	GAC	63,20	GAA	70,18	GAG	63,23
GGT	60,28	GGC	59,20	GGA	64,00	GGG	60,13

Tabelle 30: Z-Scores zur Sequenzlänge zwischen zwei gleichen Tripletts (TA-Bereinigt, CHR1), Mittelwert: 63,4869 Sigma: 4,2541

TTT	1,59	TTC	0,23	TTA	1,57	TTG	0,16
TCT	-0,11	TCC	-0,06	TCA	0,03	TCG	-1,18
TAT	1,37	TAC	0,85	TAA	1,52	TAG	0,78
TGT	0,06	TGC	-0,43	TGA	-0,29	TGG	-0,91
CTT	0,21	CTC	-0,92	CTA	0,79	CTG	-0,59
CCT	-0,65	CCC	-1,08	CCA	-0,58	CCG	-2,77
CAT	0,34	CAC	-0,14	CAA	0,69	CAG	-0,27
CGT	-0,66	CGC	-2,37	CGA	-1,00	CGG	-1,87
ATT	0,79	ATC	-0,15	ATA	1,59	ATG	0,22
ACT	0,10	ACC	-0,06	ACA	0,37	ACG	-0,91
AAT	1,36	AAC	0,66	AAA	2,07	AAG	0,61
AGT	-0,23	AGC	-0,25	AGA	-0,29	AGG	-0,91
GTT	1,14	GTC	0,59	GTA	1,52	GTG	0,25
GCT	0,22	GCC	-0,92	GCA	0,76	GCG	-2,16
GAT	0,34	GAC	-0,07	GAA	1,57	GAG	-0,06
GGT	-0,75	GGC	-1,01	GGA	0,12	GGG	-0,79

Tabelle 31: Sequenzlänge (in Tripletts) zwischen zwei gleichen Tripletts (TA-Bereinigt, Random)

TTT	62,96	TTC	62,95	TTA	63,21	TTG	63,07
TCT	62,95	TCC	62,96	TCA	62,95	TCG	62,94
TAT	62,96	TAC	62,95	TAA	62,95	TAG	62,95
TGT	62,95	TGC	62,95	TGA	62,95	TGG	62,96
CTT	62,95	CTC	62,96	CTA	63,21	CTG	63,08
CCT	62,95	CCC	62,95	CCA	62,95	CCG	62,95
CAT	62,95	CAC	62,94	CAA	62,94	CAG	62,95
CGT	62,95	CGC	62,95	CGA	62,95	CGG	62,95
ATT	62,95	ATC	62,95	ATA	63,22	ATG	63,08
ACT	62,95	ACC	62,95	ACA	62,95	ACG	62,95
AAT	63,08	AAC	63,08	AAA	63,09	AAG	63,07
AGT	63,08	AGC	63,09	AGA	63,09	AGG	63,08
GTT	62,95	GTC	62,94	GTA	63,22	GTG	63,09
GCT	62,94	GCC	62,96	GCA	62,96	GCG	62,95
GAT	63,08	GAC	63,09	GAA	63,08	GAG	63,08
GGT	62,96	GGC	62,95	GGA	62,95	GGG	62,96

Tabelle 32: Z-Scores zur Sequenzlänge zwischen zwei gleichen Tripletts (TA-Bereinigt, Random), Mittelwert: 63,0007 Sigma: 0,0797

TTT	-0,47	TTC	-0,58	TTA	2,66	TTG	0,93
TCT	-0,60	TCC	-0,55	TCA	-0,66	TCG	-0,75
TAT	-0,52	TAC	-0,61	TAA	-0,62	TAG	-0,61
TGT	-0,70	TGC	-0,67	TGA	-0,60	TGG	-0,50
CTT	-0,62	CTC	-0,55	CTA	2,61	CTG	1,06
CCT	-0,63	CCC	-0,60	CCA	-0,63	CCG	-0,60
CAT	-0,63	CAC	-0,70	CAA	-0,70	CAG	-0,58
CGT	-0,70	CGC	-0,62	CGA	-0,59	CGG	-0,66
ATT	-0,62	ATC	-0,64	ATA	2,74	ATG	0,98
ACT	-0,63	ACC	-0,63	ACA	-0,65	ACG	-0,59
AAT	1,01	AAC	1,02	AAA	1,09	AAG	0,93
AGT	1,03	AGC	1,10	AGA	1,09	AGG	0,99
GTT	-0,66	GTC	-0,71	GTA	2,74	GTG	1,12
GCT	-0,73	GCC	-0,53	GCA	-0,55	GCG	-0,65
GAT	1,05	GAC	1,06	GAA	1,03	GAG	0,99
GGT	-0,57	GGC	-0,68	GGA	-0,62	GGG	-0,55

Tabelle 33: Z-Scores der TT2-Gewichtungen der CCDS

TTT	T	0,04	0,03	-0,28	-0,04	0,40	-0,03	0,06	-0,88	0,08	-0,29	-0,97	-1,00	-0,39	-0,57	-0,98	-0,44
	C	0,18	-0,00	-0,41	0,70	0,34	-0,19	0,22	-0,85	-0,07	-0,34	-0,00	0,75	-0,76	-0,74	-0,57	-0,63
	A	0,26	-0,19	-0,46	-0,09	-0,10	-0,20	-0,17	-0,89	0,21	-0,06	0,56	0,36	-0,32	-0,41	-0,43	-0,64
	G	0,42	0,41	-0,05	1,77	1,06	1,35	0,75	-0,70	1,98	1,61	2,58	2,35	0,37	0,84	1,22	0,80
TTC	T	0,64	1,73	-0,39	0,19	0,48	1,30	0,10	-0,57	0,13	0,87	-0,98	-0,98	0,15	0,58	-0,90	0,45
	C	0,13	1,27	-0,34	3,51	0,25	0,59	0,24	-0,42	0,04	0,69	0,01	2,75	-0,49	0,31	-0,24	0,29
	A	0,61	1,79	-0,47	0,97	0,34	1,38	0,34	-0,32	0,30	1,09	0,83	1,69	0,49	1,47	0,05	0,02
	G	-0,82	-0,47	-0,89	0,27	-0,75	-0,15	-0,81	-0,75	-0,57	-0,13	-0,69	0,09	-0,82	-0,23	-0,74	-0,40
TTA	T	-0,24	-0,68	-0,54	-0,54	-0,35	-0,62	-0,46	-0,95	-0,43	-0,70	-0,99	-1,00	-0,68	-0,82	-0,99	-0,74
	C	-0,43	-0,65	-0,66	-0,27	-0,31	-0,66	-0,35	-0,93	-0,47	-0,66	-0,23	0,20	-0,88	-0,90	-0,77	-0,80
	A	-0,30	-0,64	-0,56	-0,32	-0,42	-0,63	-0,32	-0,94	0,11	-0,45	0,51	0,18	-0,52	-0,69	-0,40	-0,63
	G	-0,61	-0,78	-0,68	-0,50	-0,34	-0,48	-0,28	-0,91	0,12	-0,46	0,87	0,10	-0,69	-0,72	-0,39	-0,73
TTG	T	-0,14	-0,52	-0,57	-0,31	-0,19	-0,29	-0,32	-0,86	-0,42	-0,49	-0,99	-1,00	-0,55	-0,64	-0,97	-0,56
	C	-0,17	-0,25	-0,49	1,05	0,02	-0,20	0,04	-0,78	-0,43	-0,39	-0,16	0,81	-0,84	-0,72	-0,77	-0,61
	A	-0,23	-0,43	-0,57	-0,07	-0,29	-0,36	-0,19	-0,81	0,15	-0,22	0,85	0,83	-0,42	-0,39	-0,31	-0,41
	G	-0,23	-0,29	-0,46	0,58	0,51	0,51	0,29	-0,69	1,07	0,47	1,70	1,66	-0,36	-0,08	0,02	-0,25
TCT	T	0,01	-0,08	-0,28	0,14	0,69	0,52	0,59	-0,74	-0,42	-0,36	-0,97	-1,00	-0,51	-0,60	-0,98	-0,39
	C	0,01	-0,03	-0,48	0,95	0,72	0,21	0,85	-0,63	-0,20	-0,33	-0,05	0,98	-0,74	-0,66	-0,54	-0,42
	A	-0,36	-0,49	-0,66	-0,39	-0,20	-0,29	-0,12	-0,85	-0,28	-0,49	0,07	-0,09	-0,45	-0,48	-0,55	-0,69
	G	0,26	0,32	-0,22	1,61	0,85	1,00	0,73	-0,60	1,04	0,70	1,74	1,80	0,04	0,49	0,89	0,62
TCC	T	0,38	1,33	-0,42	0,37	0,59	1,47	0,51	-0,30	0,03	0,63	-0,97	-0,96	-0,05	0,46	-0,91	0,50
	C	-0,17	0,53	-0,55	2,37	0,29	0,15	0,74	-0,12	-0,23	0,26	-0,14	2,09	-0,72	-0,17	-0,49	0,05
	A	0,26	1,46	-0,48	1,13	0,13	0,99	0,43	-0,20	0,12	0,76	1,01	1,76	0,59	1,69	0,08	0,26
	G	-0,82	-0,55	-0,89	0,05	-0,77	-0,38	-0,76	-0,76	-0,65	-0,31	-0,63	0,07	-0,82	-0,42	-0,71	-0,46
TCA	T	-0,01	-0,45	-0,51	-0,39	0,28	-0,03	0,07	-0,81	-0,49	-0,63	-0,99	-1,00	-0,68	-0,71	-0,99	-0,56
	C	-0,37	-0,42	-0,68	0,30	0,04	-0,10	0,16	-0,78	-0,47	-0,45	-0,27	0,31	-0,86	-0,80	-0,82	-0,73
	A	-0,48	-0,68	-0,65	-0,40	-0,22	-0,43	-0,08	-0,86	0,09	-0,31	0,34	0,19	-0,19	-0,36	-0,20	-0,49
	G	-0,12	-0,25	-0,35	0,61	0,52	0,86	0,52	-0,68	1,12	0,66	1,84	1,72	-0,22	0,18	0,60	0,20
TCG	T	-0,83	-0,85	-0,95	-0,86	-0,77	-0,57	-0,84	-0,84	-0,92	-0,87	-1,02	-1,02	-0,90	-0,85	-1,02	-0,79
	C	-0,87	-0,63	-0,94	0,07	-0,72	-0,38	-0,69	-0,74	-0,93	-0,77	-0,92	-0,45	-0,98	-0,81	-0,97	-0,80
	A	-0,92	-0,92	-0,97	-0,83	-0,88	-0,83	-0,89	-0,94	-0,89	-0,89	-0,84	-0,74	-0,89	-0,78	-0,90	-0,84
	G	-0,86	-0,69	-0,91	-0,06	-0,58	0,01	-0,68	-0,65	-0,57	-0,29	-0,52	0,17	-0,80	-0,27	-0,65	-0,44
TAT	T	0,13	-0,09	-0,47	-0,28	-0,16	-0,31	-0,32	-0,90	-0,26	-0,28	-0,99	-1,01	-0,67	-0,69	-1,00	-0,64
	C	-0,37	-0,43	-0,73	-0,11	-0,29	-0,52	-0,23	-0,90	-0,47	-0,47	-0,34	0,23	-0,80	-0,74	-0,69	-0,70
	A	-0,10	-0,32	-0,65	-0,39	-0,41	-0,45	-0,41	-0,89	-0,10	-0,34	0,28	-0,02	-0,56	-0,55	-0,53	-0,71
	G	-0,05	-0,00	-0,37	0,79	0,22	0,47	0,23	-0,76	1,08	0,97	2,07	1,59	-0,11	0,43	0,59	0,11
TAC	T	0,15	0,80	-0,60	-0,13	-0,37	0,08	-0,33	-0,59	-0,05	0,42	-0,99	-1,00	-0,17	0,23	-0,93	0,14
	C	-0,24	0,32	-0,55	2,28	-0,19	-0,01	-0,04	-0,56	-0,25	0,23	-0,22	1,61	-0,68	0,05	-0,51	-0,11
	A	0,20	1,10	-0,56	0,64	-0,23	0,46	0,15	-0,33	0,15	0,76	0,80	1,35	0,03	0,93	0,07	-0,12
	G	-0,81	-0,50	-0,88	0,17	-0,78	-0,30	-0,75	-0,72	-0,57	-0,19	-0,55	0,22	-0,79	-0,30	-0,69	-0,46
TAA	T	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03
	C	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03
	A	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03
	G	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03
TAG	T	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03
	C	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03
	A	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03
	G	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03
TGT	T	-0,42	-0,47	-0,64	-0,42	-0,37	-0,45	-0,49	-0,93	-0,63	-0,60	-1,00	-1,01	-0,74	-0,77	-1,00	-0,74
	C	-0,49	-0,47	-0,76	-0,12	-0,29	-0,34	-0,37	-0,88	-0,57	-0,53	-0,51	-0,02	-0,86	-0,77	-0,77	-0,75
	A	-0,49	-0,58	-0,75	-0,55	-0,56	-0,53	-0,56	-0,94	-0,22	-0,41	0,02	-0,15	-0,64	-0,62	-0,70	-0,81
	G	-0,23	-0,05	-0,54	0,72	0,11	0,58	-0,10	-0,77	0,53	0,83	1,24	1,17	-0,13	0,85	0,55	0,83
TGC	T	-0,13	0,39	-0,72	-0,35	-0,27	0,27	-0,41	-0,75	-0,40	0,02	-1,00	-1,00	-0,17	0,37	-0,94	-0,10
	C	-0,26	0,30	-0,70	1,39	-0,18	0,18	-0,12	-0,55	-0,35	0,15	-0,31	1,29	-0,74	-0,09	-0,64	-0,23
	A	-0,17	0,44	-0,71	0,02	-0,41	0,14	-0,23	-0,60	-0,15	0,17	0,31	0,74	-0,05	0,66	-0,33	-0,22
	G	-0,91	-0,66	-0,95	-0,24	-0,84	-0,37	-0,87	-0,78	-0,78	-0,45	-0,78	-0,16	-0,85	-0,31	-0,81	-0,45
TGA	T	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03
	C	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03
	A	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03
	G	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,03	-1,02	-1,02	-1,03	-1,03
TGG	T	-0,07	-0,07	-0,73	-0,56	-0,45	-0,34	-0,53	-0,85	-0,42	-0,21	-1,00	-1,00	-0,56	-0,46	-0,97	-0,19
	C	-0,31	-0,06	-0,60	1,17	-0,44	-0,30	-0,41	-0,77	-0,47	-0,32	-0,42	0,60	-0,81	-0,59	-0,78	-0,52
	A	-0,17	0,00	-0,61	0,21	-0,36	-0,11	-0,15	-0,68	0,11	0,14	0,43	0,96	-0,34	0,05	-0,22	-0,18
	G	-0,52	-0,32	-0,68	0,30	-0,15	0,28	-0,22	-0,69	0,20	0,44	0,45	0,95	-0,48	0,09	-0,26	-0,45
CTT	T	0,07	0,02	-0,32	0,05	0,42	-0,01	-0,06	-0,86	-0,13	-0,47	-0,99	-1,01	-0,38	-0,53	-0,98	-0,45
	C	0,15	0,21	-0,40	1,08	0,67	0,05	0,37	-0,77	0,15	-0,19	0,23	1,47	-0,58	-0,59	-0,25	-0,25
	A	-0,06	-0,38	-0,62	-0,35	-0,18	-0,44	-0,37	-0,89	-0,19	-0,49	0,08	-0,17	-0,52	-0,58	-0,61	-0,73
	G	-0,01	-0,16	-0,43	0,39	0,29	0,17	0,05	-0,84	0,42	0,04	0,92	0,42	-0,14	0,09	0,46	-0,02
CTC	T	0,81	2,48	-0,46	0,10	0,44	1,18	0,14	-0,48	0,27	1,17	-0,99	-0,98	0,15	0,83	-0,88	0,38
	C	0,03	1,05	-0,45	3,06	0,15	0,26	0,14	-0,40	-0,17	0,21	-0,11	2,05	-0,62	-0,06	-0,33	0,19
	A	0,68	2,36	-0,49	1,12	0,37	1,64	0,33	-0,22	0,38	1,44	0,55	2,22	0,41	1,81	-0,16	0,17
	G	-0,82	-0,51	-0,91	-0,13	-0,71	-0,27	-0,79	-0,84	-0,67	-0,52	-0,74	-0,33	-0,82	-0,34	-0,75	-0,46
CTA	T	-0,54	-0,77	-0,80	-0,71	-0,70	-0,72	-0,71	-0,96	-0,70	-0,76	-1,00	-1,01	-0,79	-0,84	-1,00	-0,77
	C	-0,55	-0,60	-0,74	-0,18	-0,44	-0,55	-0,53	-0,91	-0,58	-0,60	-0,31	0,49	-0,87	-0,79	-0,80	-0,71
	A	-0,51	-0,70	-0,74	-0,42	-0,56	-0,65	-0,49	-0,91	-0,21	-0,50	0,02	0,17	-0,57	-0,60	-0,54	-0,63
	G	-0,70	-0,74	-0,75	-0,33	-0,42	-0,41	-0,40	-0,90	-0,14	-0,28	0,57	0,44	-0,69	-0,51	-0,41	-0,60
CTG	T	0,54	0,68	-0,48	0,36	0,59	1,50	0,23	-0,30	0,17	0,79	-0,98	-0,97	0,26	0,76	-0,91	0,71
	C	0,73	2,82	-0,04	8,35	1,41	3,16	0,85	0,20	0,30	1,87	0,50	6,98	-0,29	1,29	-0,26	1,92
	A	0,41	0,82	-0,39	1,47	0,59	1,80	0,70	-0,05	0,88	1,35	1,96	4,64	0,53	2,08	0,41	1,05
	G	0,37	1,37	-0,10	5,00	2,34	5,51	1,70	0,76	2,85	4,62	3,32	9,03	0,55	3,97	1,18	1,96
CCT	T	0,05	0,08	-0,35	0,14	0,62	0,46	0,36	-0,73	-0,04	-0,15	-0,99	-1,00	-0,49	-0,51	-0,99	-0,45
	C	0,06	0,10	-0,53	1,10	1,58	0,88	1,59	-0,35	-0,11	-0,13	-0,05	1,39	-0,67	-0,55	-0,46	-0,21
	A	-0,31	-0,48	-0,67	-0,37	-0,28	-0,30	-0,23	-0,84	-0,26	-0,50	0,06	-0,09	-0,51	-0,51	-0,61	-0,71
	G	0,41	0,54	-0,15	2,00	1,11	1,72	0,85	-0,40	1,15	1,04	1,97	2,96	0,44	1,44	1,40	1,37
CCC	T	0,17	1,19	-0,63	0,15	0,22	0,86	0,15	-0,23	0,14	0,90	-0,99	-0,99	-0,20	0,45	-0,90	0,38
	C	-0,28	0,38	-0,66	2,08	0,73	0,26	1,30	0,30	-0,30	0,36	-0,18	2,54	-0,69	-0,02	-0,40	0,39
	A	0,16	1,57	-0,57	1,14	0,25	1,57	0,56	-0,04	0,23	1,27	1,17	2,55	0,59	2,55	0,01	0,61
	G	-0,82	-0,43	-0,90	0,33	-0,60	0,35	-0,56	-0,28	-0,57	0,14	-0,45	1,16	-0,73	0,30	-0,49	0,08
CCA	T	0,17	-0,29	-0,56	-0,34	0,30	0,18	0,09	-0,76	-0,20	-0,41	-1,00	-1,00	-0,55	-0,60	-0,98	-0,44
	C	-0,24	-0,15	-0,63	0,83	1,27	1,08	1,32	-0,31	-0,31	-0,27	-0,23	1,10	-0,78	-0,64	-0,68	-0,48
	A	-0,22	-0,48	-0,57	-0,11	-0,04	-0,13	0,14	-0,77	0,33	-0,04	0,39	0,81	-0,09	-0,04	-0,06	-0,25

Tabelle 33: Z-Scores der TT2-Gewichtungen der CCDS

T1	T1 B1 Base	TT	TC	TA	TG	CT	CC	CA	CG	AT	AC	AA	AG	GT	GC	GA	GG
	G	0,19	0,05	-0,24	1,13	1,06	1,74	0,96	-0,45	1,38	1,14	2,37	2,89	0,24	1,20	1,21	0,88
CCG	T	-0,79	-0,70	-0,93	-0,81	-0,75	-0,45	-0,80	-0,78	-0,78	-0,72	-1,02	-1,02	-0,85	-0,72	-1,00	-0,69
	C	-0,80	-0,38	-0,90	0,67	-0,30	0,42	-0,31	0,11	-0,86	-0,52	-0,86	0,14	-0,93	-0,51	-0,90	-0,45
	A	-0,88	-0,85	-0,95	-0,73	-0,83	-0,68	-0,82	-0,85	-0,84	-0,81	-0,76	-0,51	-0,80	-0,56	-0,84	-0,67
	G	-0,82	-0,61	-0,91	-0,05	-0,49	0,37	-0,59	-0,45	-0,52	-0,18	-0,43	0,52	-0,67	0,14	-0,51	-0,16
CAT	T	-0,02	-0,23	-0,44	-0,20	-0,10	-0,29	-0,25	-0,88	-0,31	-0,42	-0,99	-1,01	-0,64	-0,67	-0,99	-0,72
	C	-0,19	-0,22	-0,64	0,20	-0,04	-0,31	-0,13	-0,84	-0,27	-0,28	-0,26	0,78	-0,74	-0,70	-0,67	-0,62
	A	-0,29	-0,52	-0,68	-0,44	-0,07	-0,56	-0,45	-0,90	-0,35	-0,55	-0,03	-0,08	-0,58	-0,61	-0,64	-0,77
	G	-0,13	-0,14	-0,51	0,47	0,13	0,29	-0,06	-0,79	0,29	0,24	1,05	0,85	-0,19	0,22	0,27	-0,04
CAC	T	0,03	0,60	-0,57	-0,08	-0,29	0,06	-0,23	-0,55	-0,26	0,27	-1,00	-1,00	-0,12	0,34	-0,92	0,12
	C	-0,09	0,51	-0,59	2,28	0,02	0,36	0,06	-0,34	-0,22	0,57	-0,31	1,59	-0,70	-0,08	-0,57	-0,09
	A	0,08	0,92	-0,52	0,66	0,18	0,38	0,28	-0,21	-0,03	0,47	0,33	1,13	0,14	1,12	-0,04	0,05
	G	-0,77	-0,41	-0,86	0,33	-0,71	-0,16	-0,68	-0,67	-0,64	-0,31	-0,58	0,19	-0,78	-0,13	-0,67	-0,39
CAA	T	-0,36	-0,57	-0,62	-0,63	-0,56	-0,67	-0,54	-0,96	-0,38	-0,59	-0,99	-1,01	-0,54	-0,73	-0,99	-0,63
	C	-0,12	-0,34	-0,57	0,18	-0,26	-0,40	-0,22	-0,88	-0,21	-0,45	0,13	0,86	-0,84	-0,80	-0,76	-0,76
	A	-0,06	-0,28	-0,35	0,13	-0,19	-0,32	-0,07	-0,83	0,38	-0,05	0,53	0,61	-0,20	-0,15	0,11	-0,24
	G	-0,03	-0,18	-0,31	0,52	0,35	0,45	0,38	-0,79	0,98	0,44	2,24	1,42	-0,14	0,20	0,58	-0,07
CAG	T	1,26	1,28	-0,13	0,46	0,61	0,74	0,38	-0,50	0,63	1,00	-0,97	-0,98	0,51	0,68	-0,85	0,85
	C	1,03	2,30	0,10	5,87	1,57	2,40	1,44	0,19	0,59	1,47	1,39	6,61	-0,34	0,66	-0,21	0,90
	A	1,01	1,76	0,02	2,15	0,55	1,37	0,91	-0,11	1,51	1,83	2,75	4,48	0,70	1,94	1,18	1,46
	G	0,27	0,70	-0,30	2,82	1,56	3,24	1,22	0,24	1,66	2,66	2,91	6,13	0,22	1,96	0,56	0,45
CGT	T	-0,63	-0,60	-0,83	-0,72	-0,63	-0,65	-0,76	-0,96	-0,77	-0,74	-1,02	-1,02	-0,84	-0,84	-1,02	-0,81
	C	-0,72	-0,66	-0,87	-0,36	-0,68	-0,73	-0,73	-0,94	-0,75	-0,77	-0,82	-0,47	-0,90	-0,80	-0,87	-0,76
	A	-0,79	-0,77	-0,93	-0,85	-0,85	-0,83	-0,87	-0,98	-0,83	-0,85	-0,72	-0,75	-0,88	-0,83	-0,86	-0,90
	G	-0,68	-0,55	-0,80	-0,08	-0,61	-0,34	-0,70	-0,90	-0,55	-0,45	-0,46	-0,20	-0,67	-0,33	-0,58	-0,51
CGC	T	-0,32	0,67	-0,84	-0,59	-0,50	0,19	-0,56	-0,63	-0,38	0,28	-1,01	-1,01	-0,35	0,38	-0,99	0,11
	C	-0,55	0,26	-0,74	1,47	-0,57	-0,30	-0,54	-0,64	-0,63	-0,07	-0,68	0,67	-0,76	0,09	-0,73	-0,13
	A	-0,53	0,52	-0,88	-0,00	-0,69	-0,11	-0,51	-0,57	-0,65	-0,23	-0,38	0,51	-0,56	0,49	-0,59	-0,21
	G	-0,93	-0,62	-0,96	-0,14	-0,85	-0,18	-0,86	-0,63	-0,85	-0,51	-0,89	-0,10	-0,91	-0,26	-0,88	-0,55
CGA	T	-0,59	-0,68	-0,83	-0,80	-0,73	-0,76	-0,79	-0,97	-0,73	-0,70	-1,01	-1,02	-0,80	-0,81	-1,01	-0,77
	C	-0,68	-0,64	-0,83	-0,40	-0,69	-0,68	-0,67	-0,92	-0,70	-0,70	-0,72	-0,33	-0,89	-0,81	-0,87	-0,79
	A	-0,55	-0,60	-0,78	-0,50	-0,69	-0,70	-0,65	-0,94	-0,52	-0,60	-0,36	-0,31	-0,60	-0,52	-0,48	-0,59
	G	-0,65	-0,52	-0,72	-0,16	-0,42	-0,14	-0,45	-0,84	-0,19	-0,08	0,06	0,33	-0,63	-0,17	-0,35	-0,37
CGG	T	-0,63	-0,51	-0,86	-0,72	-0,72	-0,53	-0,73	-0,90	-0,77	-0,60	-1,02	-1,02	-0,80	-0,68	-1,00	-0,62
	C	-0,53	0,17	-0,73	1,70	-0,26	0,53	-0,31	-0,41	-0,61	-0,14	-0,58	1,21	-0,75	0,04	-0,68	0,25
	A	-0,59	-0,22	-0,77	-0,25	-0,62	-0,25	-0,50	-0,71	-0,49	-0,10	-0,05	0,80	-0,57	0,00	-0,41	-0,07
	G	-0,72	-0,33	-0,78	0,22	-0,25	0,85	-0,14	-0,40	-0,15	0,70	0,20	1,94	-0,56	0,48	-0,45	-0,42
ATT	T	0,20	0,01	-0,02	0,28	0,41	-0,02	0,09	-0,88	0,09	-0,25	-0,98	-1,00	-0,36	-0,56	-0,99	-0,48
	C	0,27	0,03	-0,45	0,64	0,61	-0,03	0,42	-0,81	0,44	-0,01	0,13	1,26	-0,62	-0,61	-0,39	-0,55
	A	0,18	-0,27	-0,48	-0,21	-0,24	-0,42	-0,12	-0,90	0,12	-0,26	0,78	0,26	-0,50	-0,58	-0,50	-0,74
	G	0,35	0,25	-0,16	1,10	0,99	0,91	0,58	-0,76	1,44	1,10	2,22	1,57	0,08	0,42	0,69	0,15
ATC	T	0,51	1,59	-0,41	0,22	0,27	1,06	0,01	-0,61	0,16	0,84	-0,99	-0,99	0,09	0,63	-0,94	0,20
	C	0,12	1,13	-0,39	3,07	0,32	0,73	0,35	-0,45	0,01	0,75	-0,07	2,73	-0,53	0,36	-0,26	0,26
	A	0,84	2,33	-0,32	1,37	0,52	1,58	0,71	-0,28	0,63	1,49	1,13	2,23	0,28	1,20	-0,04	-0,14
	G	-0,72	-0,12	-0,83	0,53	-0,54	0,34	-0,58	-0,68	-0,39	0,18	-0,47	0,62	-0,78	-0,22	-0,70	-0,36
ATA	T	0,03	-0,52	-0,59	-0,60	-0,43	-0,63	-0,49	-0,95	-0,51	-0,73	-1,00	-1,02	-0,61	-0,78	-1,00	-0,68
	C	-0,59	-0,77	-0,80	-0,49	-0,45	-0,66	-0,43	-0,95	-0,56	-0,70	-0,47	-0,17	-0,91	-0,93	-0,85	-0,85
	A	-0,22	-0,61	-0,62	-0,34	-0,41	-0,60	-0,38	-0,92	-0,06	-0,41	0,33	0,11	-0,59	-0,65	-0,49	-0,70
	G	-0,52	-0,69	-0,65	-0,37	-0,30	-0,37	-0,20	-0,91	0,07	-0,19	0,70	0,07	-0,71	-0,64	-0,39	-0,71
ATG	T	0,57	0,43	-0,37	0,07	0,32	0,31	-0,00	-0,60	0,04	0,28	-0,97	-0,99	-0,22	-0,22	-0,94	-0,01
	C	0,04	0,23	-0,39	2,53	0,37	0,41	0,33	-0,53	-0,16	0,10	0,00	1,84	-0,72	-0,44	-0,66	-0,19
	A	0,32	0,49	-0,39	1,20	0,25	0,60	0,32	-0,53	0,82	0,92	1,47	2,54	-0,03	0,41	0,13	0,13
	G	0,04	0,19	-0,34	1,79	1,17	1,83	0,94	0,22	1,59	1,66	2,32	3,55	-0,12	0,82	0,39	0,23
ACT	T	0,03	-0,05	-0,31	0,04	0,22	-0,11	0,05	-0,82	-0,50	-0,28	-1,00	-1,01	-0,45	-0,58	-1,00	-0,47
	C	-0,20	-0,23	-0,52	0,67	0,34	-0,04	0,46	-0,73	-0,25	-0,21	-0,18	0,63	-0,78	-0,70	-0,63	-0,59
	A	-0,23	-0,49	-0,65	-0,42	-0,37	-0,43	-0,24	-0,87	-0,40	-0,55	-0,00	-0,33	-0,61	-0,62	-0,62	-0,79
	G	0,21	0,30	-0,18	1,49	0,57	0,67	0,48	-0,70	0,67	0,46	1,41	1,34	0,02	0,41	1,28	0,28
ACC	T	0,36	1,60	-0,47	0,24	0,22	0,70	0,15	-0,48	-0,05	0,84	-1,00	-1,00	0,06	0,69	-0,96	0,29
	C	-0,11	0,67	-0,57	2,39	0,35	0,33	0,59	-0,30	-0,31	0,28	-0,23	1,85	-0,72	-0,23	-0,63	-0,13
	A	0,64	2,32	-0,38	1,39	0,34	1,73	0,71	-0,15	0,36	1,29	1,14	2,59	0,28	1,65	-0,12	0,09
	G	-0,75	-0,30	-0,85	0,37	-0,71	-0,16	-0,70	-0,73	-0,63	-0,17	-0,57	0,25	-0,83	-0,31	-0,71	-0,45
ACA	T	0,26	-0,24	-0,36	-0,33	0,16	-0,11	0,04	-0,81	-0,34	-0,40	-1,00	-0,99	-0,48	-0,59	-0,98	-0,39
	C	-0,25	-0,27	-0,56	0,70	0,28	0,09	0,40	-0,72	-0,30	-0,23	-0,21	0,59	-0,83	-0,73	-0,78	-0,62
	A	-0,10	-0,41	-0,45	-0,08	-0,04	-0,16	0,13	-0,80	0,24	-0,15	0,51	0,43	-0,21	-0,24	-0,18	-0,39
	G	0,16	0,12	-0,20	1,28	0,84	1,25	0,82	-0,58	1,39	1,11	2,16	2,08	0,01	0,44	0,72	0,23
ACG	T	-0,66	-0,64	-0,87	-0,73	-0,72	-0,56	-0,76	-0,84	-0,80	-0,68	-1,02	-1,01	-0,76	-0,69	-1,01	-0,59
	C	-0,80	-0,46	-0,84	0,75	-0,62	-0,22	-0,61	-0,68	-0,88	-0,61	-0,86	-0,19	-0,96	-0,74	-0,94	-0,68
	A	-0,81	-0,77	-0,92	-0,67	-0,81	-0,71	-0,78	-0,88	-0,84	-0,85	-0,75	-0,61	-0,85	-0,77	-0,85	-0,78
	G	-0,79	-0,52	-0,87	0,25	-0,59	0,11	-0,56	-0,64	-0,51	-0,24	-0,42	0,35	-0,75	-0,15	-0,65	-0,47
AAT	T	0,29	-0,04	-0,11	0,09	0,32	-0,06	0,17	-0,87	-0,01	-0,19	-0,98	-1,00	-0,51	-0,63	-1,00	-0,55
	C	0,16	-0,08	-0,48	0,38	0,30	-0,12	0,30	-0,83	-0,17	-0,29	0,09	0,60	-0,74	-0,70	-0,64	-0,69
	A	0,53	0,04	-0,27	0,14	-0,14	-0,30	0,01	-0,84	0,37	-0,00	1,10	0,44	-0,32	-0,31	-0,37	-0,66
	G	0,49	0,57	-0,04	1,35	0,99	1,31	0,85	-0,70	1,55	1,44	3,24	2,29	0,33	1,11	1,53	0,73
AAC	T	0,57	1,24	-0,37	0,34	0,09	0,51	-0,02	-0,60	0,17	0,73	-0,98	-0,99	0,06	0,45	-0,93	0,40
	C	0,17	1,12	-0,39	2,63	0,40	0,86	0,45	-0,48	-0,12	0,49	0,12	2,01	-0,65	-0,03	-0,46	-0,05
	A	0,73	1,94	-0,24	1,33	0,20	0,79	0,59	-0,33	0,55	1,38	1,41	2,20	0,43	1,54	0,33	0,06
	G	-0,73	-0,27	-0,87	0,46	-0,66	-0,01	-0,65	-0,70	-0,55	-0,14	-0,47	0,28	-0,75	-0,13	-0,59	-0,25
AAA	T	0,77	0,24	0,20	0,18	0,64	0,30	0,50	-0,77	0,68	0,56	-0,93	-0,98	0,67	-0,10	-0,93	0,05
	C	0,73	0,52	-0,02	1,51	0,91	0,69	1,00	-0,74	0,38	0,21	0,86	1,54	-0,52	-0,43	-0,40	-0,37
	A	1,24	0,81	0,50	1,31	0,70	0,66	1,12	-0,61	1,50	0,78	1,73	2,43	0,36	0,31	0,82	0,27
	G	0,72	0,55	0,28	1,56	1,50	1,73	1,47	-0,61	2,62	1,65	4,92	3,02	0,32	0,87	1,55	0,43
AAG	T	1,02	1,10	-0,14	0,26	0,51	0,71	0,50	-0,55	0,35	0,93	-0,94	-0,96	0,23	0,31	-0,88	0,44
	C	0,53	1,21	-0,08	4,25	1,01	1,62	1,32	-0,22	0,35	0,90	0,88	3,53	-0,49	0,28	-0,27	0,53
	A	1,10	1,89	0,10	2,24	0,65	1,40	1,09	-0,09	1,64	2,01	4,89	6,56	0,43	1,18	0,99	1,21
	G	0,27	0,66	-0,28	2,54	1,26	2,61	1,06	-0,08	2,07	2,76	3,80	5,80	0,05	1,31	0,41	0,45
AGT	T	-0,07	-0,24	-0,38	-0,17	0,10	-0,17	-0,11	-0,89	-0,46	-0,38	-1,00	-1,01	-0,62	-0,74	-1,00	-0,66
	C	-0,29	-0,36	-0,70	0,01	0,13	-0,13	-0,14	-0,86	-0,39	-0,40	-0,24	0,26	-0,76	-0,68	-0,71	-0,66
	A	-0,26	-0,48	-0,65	-0,44	-0,41	-0,45	-0,38	-0,91	-0,20	-0,27	0,20	-0,25	-0,38	-0,34	-0,61	-0,77
	G	-0,00	0,08	-0,48	0,65	0,46	0,89	0,17	-0,72	1,22	1,37	1,80	1,61	0,14	0,80	0,66	0,44
AGC	T	0,26	1,07	-0,56	-0,00	0,44	1,27	0,18	-0,49	-0,14	0,51	-0,98	-0,99	0,03	0,65	-0,92	0,29

Tabelle 33: Z-Scores der TT2-Gewichtungen der CCDS

T1	T1 B1 Base	TT	TC	TA	TG	CT	CC	CA	CG	AT	AC	AA	AG	GT	GC	GA	GG
	C	0,11	1,17	-0,52	2,76	0,73	1,61	0,53	-0,23	-0,10	0,70	0,05	2,40	-0,56	0,40	-0,42	0,23
	A	0,35	1,41	-0,45	0,86	0,27	1,30	0,62	-0,37	0,41	1,06	1,15	1,85	1,00	3,10	0,09	0,24
	G	-0,77	-0,32	-0,91	0,18	-0,61	0,25	-0,69	-0,61	-0,48	0,13	-0,55	0,60	-0,73	0,24	-0,65	-0,20
AGA	T	-0,14	-0,51	-0,54	-0,52	-0,38	-0,54	-0,39	-0,93	-0,43	-0,34	-0,99	-1,00	-0,52	-0,66	-0,97	-0,57
	C	-0,31	-0,41	-0,63	-0,00	-0,14	-0,31	-0,21	-0,88	-0,28	-0,32	-0,14	0,19	-0,81	-0,75	-0,69	-0,71
	A	0,32	-0,28	-0,33	-0,06	-0,12	-0,39	-0,04	-0,83	0,56	0,12	1,03	0,65	-0,08	-0,08	0,37	-0,20
	G	-0,18	-0,34	-0,49	0,00	0,10	0,19	0,09	-0,81	0,71	0,37	1,44	0,99	-0,37	-0,08	0,29	-0,35
AGG	T	-0,32	-0,38	-0,70	-0,59	-0,48	-0,45	-0,53	-0,87	-0,61	-0,49	-1,00	-1,00	-0,65	-0,63	-0,99	-0,54
	C	-0,38	-0,15	-0,65	0,92	-0,08	0,03	-0,18	-0,67	-0,47	-0,28	-0,31	0,72	-0,82	-0,53	-0,69	-0,38
	A	-0,24	-0,08	-0,55	0,10	-0,25	-0,02	-0,12	-0,67	0,05	0,25	1,09	1,67	-0,41	0,08	0,02	0,07
	G	-0,64	-0,35	-0,76	-0,15	-0,29	0,16	-0,30	-0,74	-0,03	0,34	0,37	0,93	-0,61	-0,16	-0,48	-0,65
GTT	T	-0,04	-0,15	-0,26	-0,04	0,24	-0,11	-0,14	-0,90	-0,19	-0,43	-1,00	-1,01	-0,50	-0,59	-1,00	-0,58
	C	0,01	-0,15	-0,56	0,27	0,32	-0,18	0,10	-0,85	-0,17	-0,27	-0,14	0,49	-0,70	-0,69	-0,52	-0,66
	A	-0,07	-0,43	-0,63	-0,45	-0,31	-0,43	-0,37	-0,91	-0,28	-0,53	0,04	-0,36	-0,65	-0,70	-0,69	-0,83
	G	-0,01	-0,16	-0,37	0,27	0,26	0,01	-0,03	-0,88	0,38	-0,02	0,71	0,10	-0,11	0,00	0,27	-0,18
GTC	T	0,24	1,17	-0,55	-0,10	0,03	0,60	-0,26	-0,74	-0,19	0,30	-0,99	-1,00	-0,20	0,12	-0,95	-0,10
	C	-0,06	0,64	-0,52	1,95	0,08	0,23	-0,05	-0,60	-0,29	0,13	-0,43	1,13	-0,71	-0,32	-0,58	-0,28
	A	0,56	1,92	-0,47	0,73	0,38	1,46	0,27	-0,48	0,19	0,77	0,33	1,16	0,02	0,87	-0,31	-0,31
	G	-0,86	-0,59	-0,91	-0,36	-0,79	-0,51	-0,85	-0,89	-0,78	-0,67	-0,82	-0,62	-0,87	-0,65	-0,82	-0,63
GTA	T	-0,16	-0,65	-0,64	-0,59	-0,37	-0,59	-0,55	-0,95	-0,52	-0,70	-1,01	-1,01	-0,69	-0,81	-0,99	-0,71
	C	-0,60	-0,73	-0,79	-0,40	-0,43	-0,58	-0,44	-0,94	-0,57	-0,68	-0,41	-0,02	-0,89	-0,88	-0,83	-0,80
	A	-0,37	-0,71	-0,66	-0,48	-0,39	-0,56	-0,34	-0,91	-0,14	-0,50	0,14	-0,07	-0,75	-0,74	-0,64	-0,77
	G	-0,56	-0,72	-0,64	-0,37	-0,21	-0,28	-0,27	-0,91	0,07	-0,28	0,68	0,14	-0,70	-0,66	-0,43	-0,65
GTG	T	0,73	0,66	-0,48	0,03	0,57	1,04	-0,02	-0,57	-0,04	0,34	-0,99	-0,98	0,06	0,34	-0,92	0,25
	C	0,04	0,94	-0,33	4,25	0,73	1,62	0,34	-0,35	-0,23	0,53	-0,17	2,68	-0,63	0,24	-0,57	0,32
	A	0,50	0,75	-0,39	1,14	0,55	1,27	0,65	-0,32	0,74	0,93	1,53	3,14	-0,21	0,53	-0,14	0,08
	G	0,45	1,35	-0,04	3,82	1,87	3,88	1,06	-0,00	2,18	3,19	2,57	5,32	0,03	2,04	0,41	0,72
GCT	T	0,45	0,23	-0,23	0,45	0,42	0,34	0,11	-0,76	-0,19	-0,44	-1,00	-1,01	-0,42	-0,52	-0,99	-0,46
	C	0,10	0,09	-0,46	1,51	0,45	0,25	0,39	-0,62	-0,20	-0,19	-0,10	1,37	-0,70	-0,54	-0,51	-0,34
	A	-0,03	-0,37	-0,60	-0,16	-0,21	-0,27	-0,15	-0,85	-0,26	-0,52	0,14	-0,04	-0,58	-0,63	-0,61	-0,76
	G	0,72	0,85	-0,02	2,98	2,07	2,81	1,43	-0,20	1,17	1,33	2,05	3,10	0,40	1,14	1,11	1,07
GCC	T	1,13	2,85	-0,33	1,03	0,80	1,94	0,41	0,01	0,29	1,09	-0,99	-0,98	0,29	1,06	-0,90	1,00
	C	0,25	1,41	-0,37	5,05	0,80	1,13	0,86	0,21	-0,02	0,82	0,06	4,19	-0,51	0,58	-0,25	1,12
	A	1,05	3,13	-0,30	2,58	0,77	2,57	1,23	0,30	0,69	1,91	1,67	4,35	0,71	2,53	0,22	0,75
	G	-0,73	-0,07	-0,83	1,16	-0,29	1,58	-0,45	-0,02	-0,52	0,40	-0,40	1,69	-0,72	0,46	-0,60	0,13
GCA	T	0,31	-0,33	-0,44	-0,26	0,23	-0,03	-0,16	-0,78	-0,33	-0,58	-1,00	-1,01	-0,55	-0,70	-0,99	-0,53
	C	-0,20	-0,26	-0,59	0,86	0,12	0,13	0,04	-0,71	-0,40	-0,32	-0,19	0,90	-0,82	-0,71	-0,72	-0,52
	A	-0,06	-0,48	-0,50	0,03	-0,15	-0,29	0,03	-0,80	0,27	-0,15	0,71	0,73	-0,26	-0,36	-0,08	-0,38
	G	0,15	-0,02	-0,22	1,38	1,61	2,28	1,32	-0,30	1,33	0,90	2,38	2,99	-0,01	0,67	0,70	0,53
GCG	T	-0,73	-0,70	-0,93	-0,77	-0,77	-0,47	-0,84	-0,80	-0,86	-0,79	-1,02	-1,02	-0,86	-0,76	-1,01	-0,73
	C	-0,78	-0,16	-0,86	1,28	-0,61	0,10	-0,71	-0,45	-0,89	-0,53	-0,88	0,09	-0,92	-0,46	-0,92	-0,40
	A	-0,83	-0,80	-0,94	-0,68	-0,84	-0,71	-0,85	-0,87	-0,87	-0,85	-0,78	-0,59	-0,87	-0,70	-0,88	-0,78
	G	-0,80	-0,59	-0,88	0,25	-0,26	1,02	-0,41	0,37	-0,57	-0,17	-0,46	0,76	-0,70	0,27	-0,60	-0,15
GAT	T	0,87	0,45	0,01	0,51	0,73	0,23	0,25	-0,80	0,41	0,03	-0,98	-1,00	-0,37	-0,57	-0,98	-0,45
	C	0,39	0,21	-0,41	0,82	0,69	0,27	0,38	-0,78	-0,10	-0,18	0,07	0,70	-0,64	-0,61	-0,50	-0,55
	A	0,86	0,23	-0,17	0,27	-0,03	-0,10	0,03	-0,82	0,43	0,05	1,22	0,48	-0,24	-0,31	-0,29	-0,67
	G	1,17	1,16	0,26	2,78	1,94	2,32	1,24	-0,47	3,53	3,14	4,84	4,36	1,02	1,92	1,92	1,47
GAC	T	1,24	2,14	-0,18	0,80	0,61	1,13	0,39	-0,24	0,51	1,09	-0,98	-0,98	0,47	0,78	-0,89	0,99
	C	0,53	1,76	-0,21	4,36	1,04	1,85	0,80	-0,16	0,10	0,82	0,23	2,56	-0,48	0,34	-0,25	0,36
	A	1,28	2,59	-0,15	2,00	0,57	1,56	1,03	-0,00	0,80	1,62	1,70	2,99	1,12	2,51	0,65	0,54
	G	-0,66	-0,10	-0,81	0,97	-0,50	0,39	-0,55	-0,45	-0,24	0,52	-0,26	1,35	-0,63	0,50	-0,56	0,03
GAA	T	1,35	0,45	0,29	0,35	0,64	0,13	0,27	-0,79	0,99	0,26	-0,96	-0,99	1,35	0,03	-0,96	0,18
	C	0,95	0,47	0,03	1,76	0,50	0,15	0,58	-0,72	0,44	0,12	1,07	1,92	-0,52	-0,44	-0,33	-0,25
	A	1,79	1,08	0,81	1,96	0,85	0,63	1,28	-0,56	3,10	1,80	3,52	3,53	0,85	0,65	1,30	0,59
	G	1,19	0,67	0,56	1,96	1,99	2,13	1,83	-0,49	4,69	2,89	7,77	5,95	0,70	1,43	1,95	1,04
GAG	T	1,19	1,48	-0,22	0,48	0,58	0,92	0,29	-0,39	0,48	1,01	-0,98	-0,99	0,57	0,58	-0,92	0,72
	C	0,73	2,21	0,17	7,53	1,36	2,42	1,32	0,25	0,35	1,34	0,87	5,29	-0,31	0,90	-0,14	1,39
	A	1,20	2,33	-0,05	2,74	0,69	1,99	1,04	0,18	1,73	2,71	4,57	7,69	0,84	2,14	1,13	1,41
	G	0,31	1,00	-0,14	3,87	2,02	4,39	1,55	0,72	2,80	4,42	5,18	11,39	0,31	2,49	0,54	1,04
GGT	T	-0,14	-0,20	-0,55	-0,31	-0,10	-0,24	-0,43	-0,91	-0,28	-0,44	-1,01	-1,01	-0,63	-0,75	-1,01	-0,68
	C	-0,33	-0,32	-0,76	0,06	0,01	-0,05	-0,29	-0,84	-0,45	-0,44	-0,47	0,02	-0,76	-0,72	-0,72	-0,73
	A	-0,43	-0,52	-0,77	-0,65	-0,54	-0,47	-0,56	-0,95	-0,58	-0,59	-0,23	-0,73	-0,66	-0,60	-0,73	-0,88
	G	0,03	0,10	-0,49	0,71	0,37	0,61	-0,04	-0,74	0,47	0,58	0,80	0,63	0,20	0,94	0,43	0,19
GGC	T	0,69	2,07	-0,50	0,14	0,67	1,97	0,23	-0,38	0,32	1,32	-1,00	-1,00	0,23	1,03	-0,96	0,97
	C	0,12	1,36	-0,52	3,02	0,60	1,38	0,40	-0,21	-0,10	0,90	-0,15	2,30	-0,47	0,76	-0,34	0,48
	A	0,40	1,80	-0,52	1,12	0,23	1,54	0,33	-0,31	0,15	0,92	1,10	2,06	0,63	2,52	-0,05	0,08
	G	-0,74	-0,17	-0,90	0,37	-0,57	0,51	-0,66	-0,27	-0,56	-0,06	-0,64	0,84	-0,55	1,37	-0,60	0,01
GGA	T	0,45	-0,20	-0,43	-0,45	-0,16	-0,26	-0,29	-0,88	-0,21	-0,46	-1,01	-1,01	-0,60	-0,68	-1,00	-0,48
	C	-0,07	-0,14	-0,59	0,26	0,05	0,15	0,09	-0,80	-0,22	-0,19	-0,04	0,79	-0,74	-0,61	-0,65	-0,59
	A	0,31	0,01	-0,28	0,38	0,03	0,01	0,27	-0,76	0,88	0,47	1,61	1,07	0,17	0,22	0,31	-0,25
	G	0,03	0,00	-0,32	0,40	0,63	1,04	0,48	-0,64	1,20	1,33	2,48	2,53	0,16	0,93	1,00	0,28
GGG	T	-0,35	-0,44	-0,75	-0,57	-0,27	-0,14	-0,47	-0,80	-0,64	-0,49	-1,01	-1,02	-0,74	-0,70	-1,00	-0,64
	C	-0,23	0,30	-0,64	2,03	0,36	1,07	0,12	-0,34	-0,51	-0,16	-0,39	1,43	-0,76	-0,28	-0,68	-0,27
	A	-0,25	0,04	-0,66	0,21	-0,14	0,43	0,08	-0,56	0,02	0,49	1,36	2,10	-0,32	0,41	-0,26	-0,25
	G	-0,41	0,12	-0,68	0,48	0,51	1,60	0,09	-0,29	0,42	1,59	0,72	2,57	-0,22	1,36	-0,20	-0,48

Tabelle 34: Z-Scores der TT2-Gewichtungen aus Chromosom 1

TTT	T	16,15	3,38	5,04	4,03	4,69	1,92	2,64	-0,97	3,88	0,88	4,42	1,71	3,66	0,92	2,36	1,29
	C	4,77	2,13	1,66	2,16	2,42	1,04	1,64	-1,02	2,16	1,09	1,76	1,41	-0,94	-1,03	-1,04	-1,05
	A	4,58	0,69	2,12	1,04	0,92	-0,00	1,23	-1,05	2,24	0,49	4,35	1,03	1,02	-0,07	0,82	0,20
	G	3,08	0,62	1,97	1,29	1,26	0,36	1,09	-1,07	0,93	0,09	1,90	1,79	0,73	0,18	1,06	1,41
TTC	T	4,43	1,77	1,07	0,92	2,23	1,57	1,69	-0,99	0,83	0,11	0,95	0,44	1,66	0,80	1,13	0,84
	C	2,31	1,07	0,52	1,27	1,13	0,46	1,16	-0,97	1,09	0,27	0,88	1,01	-1,00	-1,08	-1,08	-1,05
	A	2,09	0,46	0,89	0,67	0,77	0,36	0,86	-0,91	0,63	0,48	1,82	1,24	0,78	0,41	0,99	0,54
	G	-0,75	-1,10	-0,99	-1,01	-1,04	-1,06	-1,10	-1,21	-1,05	-1,14	-1,04	-0,87	-1,08	-1,09	-1,09	-1,05
TTA	T	4,75	0,77	1,70	0,77	0,79	-0,10	0,42	-1,11	1,56	-0,05	1,43	0,34	0,69	-0,23	0,36	-0,04
	C	0,76	-0,19	0,04	0,25	0,11	-0,41	0,03	-1,12	0,77	-0,18	0,59	1,34	-1,01	-1,14	-1,10	-1,09
	A	1,87	0,07	1,23	0,51	0,39	-0,25	0,54	-1,08	2,18	0,37	4,41	0,97	0,42	-0,27	0,68	-0,00
	G	0,35	-0,40	0,42	-0,17	0,21	0,10	0,06	-1,13	0,09	-0,40	0,96	0,17	-0,15	-0,50	0,27	-0,28
TTG	T	2,96	0,37	0,55	0,72	0,69	-0,14	0,34	-1,10	1,20	-0,29	0,74	0,20	0,93	-0,09	0,61	0,36
	C	1,24	0,24	0,08	0,65	0,51	0,24	0,30	-1,08	0,45	-0,16	0,41	0,83	-1,08	-1,10	-1,11	-1,07
	A	0,83	-0,25	0,17	0,19	0,08	-0,33	0,09	-1,11	0,98	0,77	1,77	0,52	0,18	0,04	1,40	0,59
	G	0,50	-0,23	-0,06	0,18	0,58	0,40	0,17	-1,09	0,09	-0,39	1,10	0,42	0,05	-0,26	1,29	0,26
TCT	T	3,86	1,81	1,25	1,33	1,75	1,11	1,12	-0,93	0,50	-0,13	0,68	-0,00	0,52	0,29	0,64	0,57
	C	1,70	1,55	1,12	1,48	1,73	0,80	1,03	-0,99	0,88	0,58	1,38	1,46	-1,00	-0,90	-0,78	-0,76
	A	0,83	-0,12	0,17	-0,04	0,57	-0,22	0,19	-1,10	0,31	-0,35	0,81	-0,05	-0,16	-0,38	0,15	-0,15
	G	1,52	1,00	0,58	1,04	0,87	0,98	0,81	-1,00	0,42	0,02	0,92	0,72	0,09	0,14	0,61	0,74
TCC	T	1,87	1,03	0,18	0,61	1,01	1,02	0,78	-0,99	0,00	-0,35	0,18	-0,05	0,52	0,88	1,29	1,30
	C	0,79	0,63	-0,12	0,82	0,43	-0,07	0,87	-0,92	0,43	0,48	1,30	2,41	-1,01	-0,96	-0,83	-0,77
	A	0,84	0,41	-0,05	0,40	0,20	0,42	0,34	-0,97	-0,02	-0,20	0,96	0,21	0,25	1,04	0,59	0,69
	G	-1,03	-0,92	-1,11	-0,99	-1,05	-0,69	-1,08	-1,18	-1,13	-1,14	-1,09	-1,04	-1,12	-1,06	-1,06	-0,99
TCA	T	2,32	0,78	0,54	0,34	0,83	0,02	0,53	-1,05	0,57	-0,38	0,36	0,03	0,43	0,15	0,41	0,16
	C	1,03	0,35	-0,08	1,27	0,54	0,08	0,71	-0,95	0,57	0,31	0,46	0,92	-0,94	-0,68	-0,93	-0,99
	A	0,48	-0,34	0,42	0,12	0,34	-0,35	0,43	-1,08	1,09	0,29	2,32	0,70	0,46	0,21	0,63	0,15
	G	0,57	-0,11	0,15	0,44	0,66	1,17	0,58	-1,02	0,35	-0,09	1,24	0,78	0,23	0,14	1,30	0,26
TCG	T	-0,90	-1,04	-1,09	-0,92	-1,01	-1,10	-1,07	-1,20	-1,02	-1,16	-1,10	-1,10	-1,03	-1,01	-0,92	-1,00
	C	-0,85	-0,91	-1,13	-0,99	-1,02	-0,93	-1,00	-1,18	-1,11	-1,08	-1,11	-1,05	-1,21	-1,13	-1,21	-1,18
	A	-1,06	-1,06	-1,13	-1,06	-1,12	-1,13	-1,12	-1,22	-1,02	-0,95	-1,05	-1,08	-1,09	-1,11	-0,78	-1,01
	G	-1,08	-1,12	-1,13	-1,05	-0,80	-0,79	-1,08	-1,19	-1,12	-1,15	-1,08	-1,05	-1,08	-1,07	-0,82	-0,99
TAT	T	5,57	1,59	2,51	1,31	1,15	0,16	1,01	-1,09	1,73	-0,07	1,22	0,09	0,53	-0,17	0,41	-0,13
	C	0,86	0,17	0,15	0,20	0,04	-0,38	0,09	-1,14	0,31	-0,22	0,15	-0,07	-1,09	-1,14	-1,13	-1,16
	A	1,90	0,02	3,15	0,66	0,07	-0,38	0,73	-1,08	1,27	-0,19	2,19	0,06	0,07	-0,40	0,20	-0,30
	G	0,83	-0,21	0,68	0,52	-0,06	-0,41	0,10	-1,13	0,18	-0,41	0,69	-0,09	-0,14	-0,40	-0,05	-0,38
TAC	T	1,04	-0,07	0,11	0,07	-0,01	-0,43	0,14	-0,94	0,08	-0,52	0,42	-0,52	0,08	-0,35	0,06	-0,23
	C	-0,02	-0,27	-0,24	-0,05	-0,35	-0,64	-0,05	-1,11	0,07	-0,30	-0,08	-0,26	-1,11	-1,15	-1,14	-1,15
	A	0,86	-0,38	0,69	0,23	-0,18	-0,55	0,52	-1,06	0,29	-0,36	1,82	-0,09	0,20	-0,28	0,55	1,04
	G	-1,02	-1,12	-1,03	-0,99	-1,12	-1,13	-1,10	-1,22	-1,10	-1,17	-1,06	-1,12	-1,10	-1,15	-1,11	-1,11
TAA	T	3,18	0,36	1,04	0,22	0,31	0,83	0,22	-1,12	1,21	-0,06	1,72	0,09	0,41	-0,23	0,52	-0,03
	C	0,54	-0,25	-0,13	-0,02	-0,13	-0,58	-0,12	-1,14	0,50	-0,19	0,55	0,03	-1,03	-1,15	-1,09	-1,13
	A	1,92	0,21	2,51	1,42	0,50	-0,24	1,08	-1,03	3,50	1,02	5,08	1,36	0,72	-0,01	1,24	0,25
	G	0,21	-0,38	0,11	0,11	-0,23	-0,35	0,10	-1,12	0,16	-0,28	1,05	0,09	-0,26	-0,40	0,17	-0,38
TAG	T	0,85	-0,23	-0,15	-0,33	-0,19	-0,31	-0,27	-1,16	-0,05	-0,64	0,04	0,11	-0,29	-0,59	-0,08	-0,30
	C	-0,06	-0,35	-0,20	0,75	-0,20	-0,61	0,24	-0,98	-0,03	-0,40	0,06	-0,12	-1,12	-1,15	-1,13	-1,13
	A	0,23	-0,52	0,15	-0,03	-0,29	-0,63	0,30	-1,11	0,67	-0,23	1,72	0,29	-0,16	-0,34	1,13	-0,10
	G	-0,18	-0,61	-0,25	-0,25	-0,27	-0,59	-0,06	-1,09	-0,14	-0,57	0,60	-0,01	-0,44	-0,63	-0,12	-0,53
TGT	T	3,41	0,98	1,07	1,23	0,99	0,17	0,62	-1,06	0,54	-0,19	0,55	0,00	0,72	0,23	0,43	0,60
	C	1,15	0,99	-0,02	0,59	0,44	0,00	0,17	-1,09	0,40	0,18	0,09	0,17	-1,08	-0,98	-1,11	-1,10
	A	1,67	-0,14	0,73	0,20	-0,08	-0,40	0,14	-1,11	1,49	-0,21	1,27	0,10	0,27	-0,37	0,23	-0,25
	G	1,21	0,36	0,54	2,65	0,40	0,37	0,48	-0,92	0,74	-0,07	0,86	0,64	0,66	0,18	0,46	0,40
TGC	T	1,47	0,45	0,10	0,39	0,72	0,03	0,38	-1,06	0,19	-0,43	0,06	-0,21	0,67	0,37	0,55	1,22
	C	0,76	1,28	-0,05	1,46	0,31	-0,07	1,04	-0,79	0,43	0,47	0,12	0,39	-1,03	-1,03	-1,08	-1,02
	A	0,72	-0,19	0,11	0,23	0,52	-0,09	0,42	-0,94	0,39	0,02	0,94	0,26	1,50	0,35	0,67	0,36
	G	-0,93	-1,09	-1,10	-0,93	-1,06	-0,91	-1,08	-1,17	-1,00	-1,14	-1,11	-1,09	-0,97	-1,04	-1,08	-1,01
TGA	T	1,61	0,35	0,21	0,06	0,56	0,11	0,03	-1,09	0,33	-0,39	0,42	-0,17	0,39	-0,22	0,56	0,37
	C	0,57	-0,08	-0,28	0,10	0,56	-0,31	0,03	-1,09	0,29	-0,20	0,34	0,48	-1,03	-1,11	-1,07	-1,07
	A	1,10	0,01	0,84	0,78	0,59	0,23	0,57	-0,96	1,75	1,03	2,47	0,90	0,49	0,11	1,02	0,45
	G	0,51	-0,13	0,12	0,23	0,51	1,29	0,41	-1,04	1,08	0,93	1,48	0,55	0,76	1,13	0,78	0,21
TGG	T	1,05	-0,08	-0,11	-0,10	0,77	-0,40	-0,00	-1,12	0,04	-0,46	0,02	-0,08	0,28	-0,01	0,70	1,31
	C	0,68	0,96	0,24	0,78	0,71	-0,09	1,15	-0,92	0,38	0,19	0,31	0,60	-0,91	-0,89	-1,00	-0,88
	A	0,40	-0,18	0,25	0,25	-0,07	-0,50	0,17	-1,06	0,93	-0,02	1,87	0,88	0,98	0,07	1,13	0,84
	G	0,39	-0,26	-0,05	0,75	0,46	-0,07	1,03	-0,67	1,19	0,13	1,07	2,12	0,39	0,08	0,86	0,36
CTT	T	3,58	1,76	1,35	1,12	2,38	0,97	0,95	-1,03	1,11	0,02	1,03	0,04	1,17	0,37	0,86	1,23
	C	1,49	1,01	0,28	1,07	1,88	0,88	0,86	-0,99	0,83	0,14	0,50	0,69	-0,77	-1,09	-1,07	-1,04
	A	0,70	-0,30	0,05	-0,23	-0,02	-0,43	0,03	-1,08	0,27	-0,33	0,99	-0,13	-0,30	-0,37	0,03	-0,29
	G	0,45	-0,12	-0,08	0,23	0,56	-0,01	0,06	-1,10	0,01	-0,32	1,21	0,56	0,02	0,43	0,22	0,42
CTC	T	1,77	0,80	0,10	0,46	1,80	0,74	0,60	-1,00	0,06	0,34	0,11	-0,15	1,79	0,98	0,74	0,69
	C	0,96	0,79	-0,01	2,46	1,03	0,50	2,14	-0,48	0,61	0,35	0,41	1,46	-0,84	-0,79	-1,05	-0,97
	A	0,84	0,16	-0,07	0,44	1,38	0,16	0,87	-0,73	0,07	-0,23	1,63	0,55	0,51	1,59	0,91	1,05
	G	-1,03	-1,05	-1,12	-0,89	-0,84	-0,99	-1,08	-1,17	-1,01	-1,10	-0,86	-1,02	-1,02	-0,50	-1,04	-0,76
CTA	T	1,14	-0,03	0,10	-0,19	0,01	-0,51	-0,19	-1,13	0,11	-0,54	0,34	-0,35	0,12	-0,48	-0,07	-0,42
	C	0,26	0,03	0,11	-0,16	-0,12	-0,50	-0,13	-1,12	-0,09	-0,50	0,18	0,25	-1,06	-1,13	-1,11	-1,12
	A	0,96	-0,46	0,08	-0,14	-0,25	-0,61	0,00	-1,11	0,50	-0,29	1,71	0,04	-0,22	-0,47	-0,01	-0,33
	G	-0,22	-0,63	-0,52	-0,52	-0,29	-0,49	-0,29	-1,15	-0,25	-0,30	0,44	-0,15	-0,37	-0,39	-0,04	-0,46
CTG	T	1,43	0,26	0,04	0,38	1,30	0,07	0,44	-0,95	0,22	-0,31	1,38	0,26	0,93	0,31	0,70	0,67
	C	0,79	0,40	-0,06	0,90	2,07	0,62	0,57	-0,95	0,34	0,65	0,92	1,12	-0,95	-0,97	-1,03	-0,95
	A	0,36	-0,32	-0,07	0,20	0,21	0,33	0,23	-1,05	0,59	-0,10	1,33	0,68	0,51	0,34	1,18	1,74
	G	0,14	0,24	-0,25	0,20	1,02	0,87	0,39	-0,99	0,25	-0,13	1,03	1,49	0,98	1,44	2,35	0,73
CCT	T	1,32	0,82	0,31	0,66	0,99	0,99	0,45	-0,91	-0,10	-0,49	-0,04	-0,29	0,26	-0,09	0,17	0,39
	C	0,77	0,74	-0,14	1,04	1,58	2,63	0,83	-0,79	0,34	0,19	0,43	1,76	-0,90	-0,99	-1,02	-0,68
	A	-0,04	-0,54	-0,26	-0,36	-0,27	-0,40	-0,26	-1,10	-0,15	-0,50	0,20	-0,30	-0,49	-0,45	-0,15	-0,15
	G	0,21	0,33	1,05	0,52	0,53	1,41	0,37	-0,97	-0,09	0,41	0,53	1,04	0,09	1,40	0,72	2,18
CCC	T	0,57	0,37	-0,35	-0,07	0,52	0,74	0,22	-0,99	-0,46	-0,53	-0,28	-0,45	0,41	0,41	0,26	0,53
	C	0,11	0,21	-0,53	0,15	-0,22	-0,53	0,37	-0,90	0,33	0,38	0,25	0,74	-0,79	-0,83	-1,00	-0,90
	A	0,33	0,34	-0,38	0,00	0,29	0,88	0,44	-0,91	-0,31	-0,17	1,39	0,43	0,36	2,27	0,75	2,20
	G	-1,06	-0,80	-1,11	-0,98	-0,98	-0,51	-1,00	-1,12	-1,11	-1,04	-1,05	-0,76	-1,05	-0,53	-1,00	-0,53
CCA	T	1,27	0,35	-0,02	0,05	0,98	0,24	0,35	-1,04	0,04	-0,56	-0,04	-0,42	0,72	0,17	0,15	0,09
	C	0,35	-0,01	-0,31	1,20	0,92	0,52	1,29	-0,75	0,38	0,31	0,26	0,65	-0,88	-0,74	-1,01	-0,93
	A	-0,02	-0,43	-0,13	-0,26	-0,11	-0,26	0,58	-1,05	0,73	0,04	1,28	1,22	0,26	-0,13	0,57	0,40

Tabelle 34: Z-Scores der TT2-Gewichtungen aus Chromosom 1

T1	T1 B1	TT	TC	TA	TG	CT	CC	CA	CG	AT	AC	AA	AG	GT	GC	GA	GG
	G	0,19	-0,09	-0,16	0,25	1,47	1,90	1,20	-0,95	0,08	-0,05	0,84	0,68	0,55	2,02	1,29	0,53
CCG	T	-0,97	-1,08	-1,13	-1,10	-0,64	-1,03	-1,07	-1,19	-1,10	-1,16	-1,09	-1,12	-0,92	-0,97	-1,01	-0,93
	C	-0,97	-1,02	-1,13	-0,96	-0,49	-0,72	-0,87	-1,08	-1,07	-1,01	-1,07	-0,95	-1,17	-1,03	-1,18	-1,11
	A	-1,11	-1,15	-1,15	-1,11	-1,09	-1,07	-1,09	-1,19	-1,08	-1,14	-1,05	-1,04	-0,86	-1,02	-0,75	-0,69
	G	-1,11	-1,11	-1,15	-1,05	-0,80	-0,74	-1,02	-1,13	-1,11	-1,10	-1,06	-0,97	-0,80	-0,64	-0,77	-0,90
CAT	T	3,05	1,20	1,46	0,99	1,22	0,29	0,89	-1,03	0,59	-0,28	0,53	-0,14	0,41	-0,00	0,05	-0,23
	C	0,73	0,85	-0,02	0,60	0,47	0,01	0,23	-1,07	0,42	0,11	0,19	0,20	-1,03	-1,08	-1,07	-1,10
	A	0,84	-0,26	0,67	0,25	-0,21	-0,50	0,20	-1,07	0,36	-0,43	1,05	-0,23	-0,09	-0,38	0,07	-0,35
	G	0,90	-0,05	0,24	0,61	0,23	0,65	0,24	-1,03	0,35	-0,30	0,62	0,44	0,66	0,08	0,40	0,01
CAC	T	1,46	0,20	0,13	0,39	0,75	0,57	0,24	-1,05	-0,11	-0,48	-0,16	-0,52	0,72	1,49	0,43	0,24
	C	0,20	0,58	-0,25	1,08	0,08	0,06	0,78	-0,83	0,94	0,98	0,16	0,20	-0,86	-0,83	-1,05	-1,06
	A	0,72	0,03	0,53	0,60	0,26	0,50	2,44	-0,85	0,38	-0,25	1,06	0,05	0,54	0,49	1,35	0,51
	G	-0,93	-1,01	-0,99	-0,73	-0,97	-0,36	-0,95	-1,15	-0,98	-1,10	-1,03	-0,89	-0,88	-0,98	-0,98	-0,98
CAA	T	1,03	0,14	0,20	-0,31	0,11	-0,30	0,05	-1,12	0,44	-0,40	0,77	-0,19	0,13	-0,32	0,33	0,11
	C	0,14	0,12	-0,33	-0,11	0,42	-0,31	-0,10	-1,12	0,82	-0,09	0,69	0,37	-1,01	-1,12	-1,07	-1,09
	A	1,06	0,25	1,28	0,99	0,69	-0,01	1,08	-1,00	2,04	1,02	3,86	1,09	1,18	0,29	1,24	0,61
	G	0,12	-0,20	0,07	0,62	0,02	-0,20	0,38	-0,80	0,37	0,26	0,90	0,45	0,09	0,09	0,46	-0,07
CAG	T	1,08	0,06	0,07	-0,12	0,41	-0,22	0,09	-1,10	-0,01	-0,51	0,39	-0,14	0,42	-0,02	1,26	1,21
	C	0,53	0,54	0,74	0,75	3,03	0,42	0,77	-0,84	0,49	0,79	0,64	0,91	-0,98	-1,00	-0,99	-0,95
	A	0,47	-0,08	0,19	0,56	0,23	-0,16	0,60	-0,97	1,06	0,19	2,33	1,21	1,06	0,85	1,47	1,04
	G	0,26	-0,25	-0,05	1,11	1,72	0,28	1,45	-0,36	0,32	-0,07	1,24	2,46	0,36	0,27	0,81	0,15
CGT	T	-0,85	-0,93	-1,03	-0,99	-1,00	-1,02	-0,92	-1,20	-1,07	-1,14	-1,08	-1,11	-1,03	-1,09	-1,09	-0,91
	C	-0,99	-0,64	-1,11	-0,96	-1,03	-1,03	-1,06	-1,17	-1,05	-1,06	-1,11	-1,05	-1,20	-1,19	-1,21	-1,19
	A	-1,02	-1,11	-1,02	-1,08	-1,14	-1,15	-1,11	-1,22	-1,08	-1,14	-1,04	-1,08	-1,09	-1,13	-1,08	-1,11
	G	-0,92	-1,02	-1,06	-0,86	-1,00	-0,81	-0,98	-1,17	-0,82	-1,06	-0,92	-0,73	-0,73	-0,95	-0,93	-0,91
CGC	T	-0,85	-1,03	-1,12	-0,87	-0,88	-0,99	-1,04	-1,17	-1,12	-1,14	-1,13	-1,15	-0,99	-0,95	-1,03	-0,95
	C	-0,98	-0,60	-1,09	-0,37	-0,99	-0,88	-0,67	-0,87	-0,88	-0,71	-1,09	-0,99	-1,17	-1,07	-1,19	-1,13
	A	-1,09	-1,10	-1,12	-1,03	-1,09	-0,94	-1,02	-1,16	-1,07	-1,12	-1,07	-1,10	-0,99	-0,98	-1,02	-0,97
	G	-1,20	-1,18	-1,22	-1,16	-1,17	-0,96	-1,17	-1,14	-1,14	-1,20	-1,21	-1,18	-1,09	-1,12	-1,18	-1,11
CGA	T	-1,01	-0,92	-1,12	-1,12	-0,82	-1,04	-1,10	-1,21	-1,11	-1,16	-1,11	-1,13	-1,05	-1,10	-1,06	-1,04
	C	-1,10	-1,10	-1,16	-1,10	-1,05	-1,11	-1,12	-1,21	-1,11	-1,11	-1,11	-0,95	-1,20	-1,20	-1,21	-1,19
	A	-1,07	-1,12	-1,09	-1,02	-0,93	-1,14	-1,09	-1,20	-1,03	-1,02	-0,99	-1,05	-1,06	-1,08	-1,04	-1,04
	G	-1,05	-1,09	-0,94	-1,05	-1,06	-1,04	-1,05	-1,17	-0,80	-0,65	-0,99	-0,99	-0,84	-0,78	-0,98	-1,00
CGG	T	-1,01	-1,09	-1,14	-1,12	-1,06	-1,10	-1,10	-1,20	-1,14	-1,16	-1,11	-1,12	-1,05	-1,06	-0,94	-0,74
	C	-1,02	-0,76	-0,97	-0,91	-0,66	-0,94	-0,92	-1,07	-1,08	-1,05	-1,08	-0,94	-1,18	-1,12	-1,18	-1,08
	A	-1,10	-1,00	-1,14	-1,07	-1,10	-1,12	-1,09	-1,18	-1,09	-1,12	-1,03	-0,99	-0,92	-0,95	-0,99	-0,79
	G	-0,86	-1,09	-1,11	-0,84	-0,97	-0,98	-0,80	-0,87	-0,96	-1,07	-0,96	-0,47	-0,81	-0,88	-0,92	-0,90
ATT	T	6,53	2,19	3,49	2,09	2,65	0,90	1,89	-1,03	2,88	0,52	2,17	0,51	1,39	0,51	1,19	0,76
	C	1,47	1,29	0,61	1,02	0,79	0,14	0,79	-1,08	1,03	0,13	1,12	0,72	-1,05	-1,13	-1,04	-1,08
	A	2,33	0,16	1,29	0,38	0,27	-0,31	1,49	-1,08	1,16	0,00	2,23	0,27	0,01	0,27	0,31	-0,15
	G	0,77	-0,18	0,30	0,38	0,55	-0,25	0,37	-1,07	0,28	-0,38	0,61	0,09	-0,14	-0,34	-0,01	-0,31
ATC	T	1,32	0,46	0,16	0,40	1,04	0,25	1,08	-0,81	0,28	-0,29	0,10	-0,25	0,52	0,14	0,37	0,09
	C	0,43	0,23	-0,15	0,32	0,10	-0,30	1,18	-1,05	0,26	-0,03	0,07	0,25	-1,09	-0,99	-1,12	-1,11
	A	0,89	0,08	0,17	0,34	0,55	0,03	0,61	-0,84	0,27	-0,28	0,89	-0,00	0,04	-0,18	0,41	-0,08
	G	-1,05	-1,10	-1,10	-0,97	-0,88	-1,08	-1,03	-1,14	-1,09	-1,16	-1,05	-1,01	-1,13	-1,14	-1,13	-1,11
ATA	T	3,44	0,68	1,24	0,45	0,72	-0,14	0,32	-1,11	3,57	0,47	1,56	0,11	0,97	-0,09	0,48	0,04
	C	0,39	-0,40	-0,04	-0,01	-0,06	-0,50	0,03	-1,14	0,90	0,03	1,20	0,23	-0,98	-1,13	-1,08	-1,09
	A	1,68	0,18	1,70	0,59	0,18	-0,42	0,54	-1,07	2,87	0,56	3,94	1,12	0,07	-0,24	0,49	-0,09
	G	0,25	-0,46	0,07	-0,10	-0,06	-0,49	0,18	-1,12	0,28	-0,34	0,87	0,16	-0,31	-0,44	0,00	-0,46
ATG	T	1,54	0,22	0,49	0,83	0,48	-0,11	0,26	-1,11	0,90	-0,21	0,79	-0,11	0,97	0,11	0,56	0,38
	C	0,41	-0,20	-0,02	0,49	0,69	-0,11	0,35	-1,08	0,26	-0,20	0,41	0,33	-1,07	-1,09	-1,11	-1,06
	A	0,79	-0,04	0,31	0,44	0,12	-0,38	0,39	-1,05	1,02	0,06	1,96	0,72	0,12	0,03	0,86	0,35
	G	0,13	-0,33	0,13	0,93	0,29	-0,22	0,46	-0,88	0,26	-0,35	1,13	0,64	-0,17	-0,31	0,43	0,34
ACT	T	1,70	0,48	0,62	1,20	0,70	0,33	0,48	-1,06	0,08	-0,26	0,42	-0,17	-0,04	-0,27	0,50	0,25
	C	0,46	-0,01	-0,15	0,92	0,73	-0,11	0,96	-0,93	0,12	0,17	0,18	0,52	-1,08	-1,11	-1,08	-0,86
	A	0,25	-0,48	0,05	-0,09	-0,27	-0,58	0,27	-1,10	0,00	-0,52	1,02	-0,30	-0,56	-0,65	-0,01	-0,50
	G	0,29	-0,16	0,20	0,54	0,27	-0,11	1,49	-0,96	0,04	-0,28	0,76	0,43	-0,23	-0,19	0,26	0,29
ACC	T	0,37	-0,06	-0,30	0,09	0,49	0,55	0,70	-0,83	-0,31	-0,47	-0,15	-0,37	0,29	0,00	0,22	0,52
	C	-0,20	-0,19	-0,44	0,35	-0,18	-0,36	0,40	-0,81	-0,18	-0,02	0,05	1,01	-1,09	-0,98	-1,08	-0,80
	A	0,27	0,04	-0,14	0,65	0,24	0,08	0,67	-0,74	-0,15	-0,23	0,71	0,02	-0,24	0,37	0,08	0,09
	G	-1,09	-1,10	-1,10	-0,88	-1,08	-1,07	-0,97	-1,09	-1,13	-1,15	-1,08	-1,03	-1,14	-1,10	-1,12	-1,05
ACA	T	2,21	0,32	0,40	0,15	0,37	-0,21	0,26	-1,08	0,97	0,05	0,68	0,12	0,53	0,02	0,39	0,72
	C	0,34	0,06	-0,31	0,43	0,23	-0,20	0,28	-1,05	0,83	1,86	0,64	1,00	-0,92	-0,99	-1,04	-0,92
	A	0,53	-0,21	0,50	0,41	-0,02	-0,33	0,56	-1,06	1,35	0,62	3,51	1,04	-0,05	-0,27	0,49	0,12
	G	0,28	-0,18	0,09	0,69	0,28	0,08	0,68	-1,00	0,52	0,15	1,91	2,01	0,29	0,88	0,50	0,40
ACG	T	-0,88	-1,02	-1,04	-1,00	-1,00	-1,07	-1,03	-1,20	-0,98	-1,11	-1,01	-1,06	-0,92	-0,96	-0,94	-0,83
	C	-1,04	-1,08	-1,12	-0,98	-0,69	-0,91	-0,91	-1,18	-1,07	-1,05	-1,08	-0,99	-1,20	-1,18	-1,21	-1,17
	A	-1,06	-1,00	-1,10	-1,03	-1,11	-1,14	-1,08	-1,20	-1,05	-1,12	-0,96	-1,02	-1,08	-1,11	-1,01	-0,90
	G	-1,09	-1,12	-1,10	-0,86	-1,05	-1,07	-1,02	-1,18	-1,09	-1,12	-1,00	-0,84	-1,03	-1,07	-0,91	-0,79
AAT	T	4,21	1,14	1,90	1,20	1,45	0,41	1,09	-1,06	1,67	0,28	1,87	0,95	0,55	0,04	0,47	0,00
	C	0,70	0,48	0,22	0,41	0,46	0,82	0,40	-1,10	0,79	0,24	0,80	0,37	-1,06	-0,94	-1,06	-1,11
	A	2,54	0,54	1,89	0,75	0,49	-0,13	1,57	-1,03	2,34	0,40	4,56	0,70	0,25	-0,02	0,80	-0,02
	G	1,24	0,23	0,83	0,72	0,61	0,01	0,67	-1,09	0,88	0,16	1,99	0,84	0,32	0,23	0,89	0,32
AAC	T	1,16	0,42	0,21	0,12	0,49	0,65	0,74	-1,05	0,23	-0,24	0,36	-0,07	0,27	-0,01	0,57	0,19
	C	0,13	0,42	-0,19	0,26	0,04	0,08	0,38	-0,92	0,13	-0,01	0,50	0,12	-1,09	-1,11	-1,08	-1,11
	A	1,27	0,10	0,81	0,90	0,66	-0,20	1,00	-0,93	0,76	0,28	3,03	0,44	0,29	0,11	1,35	0,20
	G	-0,93	-1,07	-1,02	-0,93	-1,05	-1,10	-1,06	-1,20	-1,00	-1,12	-0,96	-1,03	-1,09	-1,09	-1,03	-0,90
AAA	T	3,03	1,14	3,16	1,00	1,23	0,37	1,59	-1,01	3,55	1,64	4,75	1,13	2,19	0,90	2,27	1,30
	C	1,06	0,75	1,00	1,10	0,50	0,62	1,03	-1,03	1,53	1,03	2,92	1,29	-0,89	-1,04	-0,95	-0,81
	A	4,19	1,48	5,55	3,06	1,88	0,90	3,34	-0,85	6,52	2,32	16,03	3,44	1,66	0,94	3,56	1,44
	G	1,07	0,46	1,03	1,42	0,64	0,27	1,43	-1,00	1,28	0,74	4,22	1,58	0,29	0,31	1,85	0,56
AAG	T	1,03	0,13	0,36	0,13	0,45	-0,15	0,58	-1,10	0,38	-0,27	0,77	0,27	0,34	0,46	1,03	0,53
	C	0,20	-0,04	-0,06	0,53	0,20	-0,14	0,69	-1,03	0,43	-0,05	1,21	0,79	-1,05	-1,02	-0,85	-1,03
	A	0,68	0,05	0,69	0,70	0,48	0,10	1,16	-1,00	1,58	0,42	4,54	1,43	0,35	0,18	1,69	0,71
	G	0,04	-0,17	-0,01	0,19	0,32	-0,08	0,78	-0,99	0,37	-0,01	2,27	0,86	-0,20	-0,17	0,73	0,11
AGT	T	1,76	0,68	0,63	0,71	0,63	-0,07	0,57	-0,94	0,16	-0,30	0,32	-0,16	-0,02	-0,38	0,19	-0,10
	C	0,48	0,42	-0,30	0,30	0,04	0,05	-0,08	-1,11	0,12	-0,11	0,08	0,21	-1,11	-1,13	-1,12	-1,09
	A	0,49	-0,42	0,07	-0,23	-0,29	-0,60	-0,09	-1,13	0,28	-0,19	0,93	-0,01	-0,27	0,23	0,58	-0,27
	G	0,40	-0,23	-0,16	0,25	0,76	-0,22	0,50	-1,09	0,55	-0,13	0,77	1,38	0,20	0,50	0,43	0,29
AGC	T	0,65	0,20	-0,24	0,01	0,42	0,17	0,51	-1,06	-0,04	0,28	0,21	-0,30	0,29	0,14	0,65	1,49

Tabelle 34: Z-Scores der TT2-Gewichtungen aus Chromosom 1

T1	T1 B1	TT	TC	TA	TG	CT	CC	CA	CG	AT	AC	AA	AG	GT	GC	GA	GG
	C	0,28	1,29	-0,27	1,74	0,17	0,00	0,45	-0,97	0,30	0,77	0,58	1,02	-1,05	-0,94	-0,80	-0,81
	A	0,63	-0,05	-0,08	0,23	0,76	-0,16	0,38	-1,00	0,54	-0,17	1,24	0,55	0,39	0,38	0,93	0,54
	G	-1,05	-1,08	-1,13	-0,97	-0,97	-1,03	-1,06	-1,17	-0,88	-1,11	-1,04	-0,84	-1,07	-1,04	-1,05	-0,97
AGA	T	1,45	0,12	0,30	0,12	0,10	-0,33	0,56	-0,82	0,54	-0,13	0,79	0,01	0,33	-0,05	0,83	0,97
	C	0,40	0,22	-0,18	0,39	0,02	-0,11	0,77	-1,06	0,47	0,24	0,70	1,64	-1,01	-1,06	-1,01	-0,64
	A	1,48	0,60	1,08	1,23	0,60	-0,15	0,98	-1,00	2,54	1,16	4,75	2,26	0,68	0,43	1,86	0,95
	G	0,72	-0,02	0,00	0,49	0,43	0,07	0,94	-0,88	1,05	0,78	2,21	1,84	0,44	0,42	1,01	0,51
AGG	T	0,50	-0,18	-0,20	0,41	-0,03	-0,43	0,64	-1,05	-0,06	-0,51	0,10	-0,11	0,23	-0,09	0,57	0,92
	C	0,21	0,00	-0,14	3,01	0,40	-0,10	0,69	-0,66	0,68	0,19	0,48	2,10	-0,69	-0,85	-1,01	-0,49
	A	0,48	-0,18	0,04	0,45	0,04	-0,37	0,45	-1,03	0,78	0,07	2,64	1,68	0,66	0,15	1,74	1,55
	G	0,03	-0,23	-0,36	0,08	0,16	-0,16	0,30	-1,00	0,12	-0,13	1,08	1,00	-0,20	-0,09	0,42	-0,22
GTT	T	2,33	0,95	1,03	0,99	1,01	0,29	0,93	-1,02	0,56	-0,33	0,50	-0,30	0,74	-0,17	0,30	0,10
	C	0,45	0,07	-0,22	0,18	0,08	-0,28	-0,04	-1,12	0,09	-0,29	0,80	-0,10	-1,09	-1,17	-0,94	-1,14
	A	0,38	-0,54	-0,20	-0,42	-0,30	-0,65	-0,22	-1,13	0,00	-0,58	0,36	-0,33	-0,52	-0,59	-0,35	-0,41
	G	0,30	-0,53	-0,36	-0,09	-0,17	-0,14	0,01	-1,12	-0,28	-0,65	0,34	-0,22	-0,22	-0,22	-0,05	-0,18
GTC	T	0,74	0,03	-0,27	0,26	0,79	0,02	0,92	-0,65	-0,34	-0,71	-0,44	-0,59	0,18	-0,26	-0,17	-0,04
	C	-0,01	-0,26	-0,58	-0,06	-0,13	-0,37	0,13	-1,07	-0,35	-0,59	-0,41	-0,15	-1,11	-1,14	-1,15	-1,09
	A	0,17	-0,37	-0,40	-0,30	-0,12	-0,15	-0,08	-1,06	-0,30	-0,64	0,05	-0,34	-0,29	-0,43	-0,04	0,41
	G	-1,12	-1,14	-1,16	-1,10	-1,11	-0,99	-1,14	-1,20	-1,16	-1,18	-1,14	-1,10	-1,15	-1,13	-1,14	-1,04
GTA	T	1,66	-0,20	-0,07	-0,39	-0,13	-0,65	-0,39	-1,16	0,48	-0,54	0,00	-0,54	0,04	-0,70	-0,39	-0,55
	C	-0,26	-0,68	-0,64	-0,52	-0,51	-0,73	-0,48	-1,16	-0,19	-0,64	-0,30	-0,31	-1,10	-1,18	-1,16	-1,16
	A	0,28	0,44	-0,06	-0,28	-0,18	-0,68	-0,20	-1,14	0,51	-0,32	0,89	0,07	-0,22	-0,63	-0,14	-0,44
	G	-0,24	-0,44	-0,51	-0,48	0,27	-0,66	-0,42	-1,14	-0,28	-0,69	0,19	0,35	-0,46	-0,69	-0,35	-0,53
GTG	T	0,76	-0,14	-0,19	0,05	0,23	-0,29	-0,21	-1,12	0,13	-0,63	-0,26	-0,47	1,93	-0,01	0,31	0,28
	C	-0,04	-0,37	-0,39	0,77	0,18	-0,26	0,19	-1,05	-0,17	-0,39	-0,12	0,59	-0,95	-1,07	-1,08	-1,01
	A	0,25	0,21	-0,22	0,26	-0,11	-0,44	0,19	-1,06	0,14	-0,31	1,09	0,14	-0,08	0,77	0,59	0,20
	G	-0,00	-0,44	-0,31	0,98	0,76	-0,20	0,47	-0,70	0,11	-0,58	0,40	0,47	-0,04	-0,12	0,45	0,37
GCT	T	0,93	0,15	0,02	0,44	0,41	0,20	0,11	-1,07	-0,26	-0,63	-0,26	-0,47	-0,26	-0,42	0,14	-0,12
	C	0,20	-0,15	-0,46	0,74	0,16	-0,14	0,01	-0,95	0,03	0,75	0,04	0,33	-1,10	-1,09	-1,11	-1,03
	A	-0,01	-0,68	-0,39	-0,36	0,22	-0,67	-0,37	-1,13	0,28	-0,59	-0,08	-0,38	-0,67	-0,74	-0,25	-0,45
	G	0,09	-0,30	-0,29	0,50	0,37	0,04	0,35	-0,96	-0,19	-0,43	0,30	1,58	0,39	-0,06	1,05	2,29
GCC	T	0,29	0,18	-0,45	0,10	0,37	1,93	1,02	-0,77	-0,43	-0,68	-0,50	-0,38	0,89	0,05	0,12	2,05
	C	-0,16	-0,16	-0,63	0,29	-0,09	-0,31	0,07	-0,88	-0,30	-0,13	-0,26	1,42	-1,07	-0,84	-1,08	-0,63
	A	0,21	-0,26	-0,41	0,08	0,49	0,60	0,15	-0,95	-0,36	-0,23	0,21	0,43	-0,19	-0,06	0,13	1,41
	G	-1,10	-1,10	-1,15	-0,98	-0,97	-0,90	-1,05	-1,11	-1,13	-1,13	-1,09	-0,49	-1,10	-1,00	-1,06	-0,54
GCA	T	0,90	-0,08	-0,22	-0,31	-0,02	-0,41	-0,22	-1,10	-0,08	-0,69	-0,26	-0,48	0,02	-0,25	0,15	0,18
	C	0,47	0,11	-0,59	-0,02	-0,08	-0,32	-0,01	-1,06	0,11	-0,13	-0,13	0,31	-0,92	-1,01	-1,02	-0,97
	A	0,08	-0,28	-0,18	-0,02	-0,38	-0,17	0,22	-1,10	0,41	-0,19	0,90	0,35	-0,06	-0,41	0,30	-0,11
	G	0,06	-0,42	-0,35	1,48	0,15	0,22	0,36	-0,95	0,19	-0,33	0,80	0,96	-0,00	0,06	0,89	0,40
GCG	T	-1,04	-1,01	-1,14	-1,08	-1,05	-1,09	-1,11	-1,20	-1,13	-1,18	-1,14	-1,13	-1,00	-1,02	-0,67	-0,74
	C	-1,02	-1,07	-1,15	-1,00	-0,85	-0,85	-0,89	-1,12	-1,09	-1,07	-1,10	-0,96	-1,18	-1,10	-1,13	-1,03
	A	-0,94	-0,97	-1,14	-1,05	-1,13	-1,14	-0,98	-1,19	-1,14	-1,16	-1,03	-1,09	-1,11	-1,08	-0,90	-0,99
	G	-1,10	-1,12	-1,15	-0,82	-1,00	-0,99	-1,04	-1,07	-0,98	-1,13	-1,08	-0,79	-0,98	-0,84	-0,96	-0,83
GAT	T	1,47	0,48	0,21	0,35	0,58	-0,22	0,00	-1,12	0,18	0,44	0,05	-0,45	-0,20	-0,27	-0,32	-0,44
	C	0,05	0,07	-0,52	-0,09	-0,14	-0,52	-0,17	-1,00	-0,04	0,21	-0,25	-0,32	-1,04	-0,98	-1,05	-1,14
	A	0,54	-0,52	0,02	-0,26	-0,42	-0,67	-0,14	-1,11	0,19	-0,54	0,69	-0,31	-0,47	-0,68	-0,10	-0,54
	G	0,10	-0,45	-0,37	0,17	-0,09	-0,40	-0,21	-1,10	0,10	-0,37	0,46	0,16	0,03	-0,26	0,44	0,34
GAC	T	0,34	-0,09	-0,37	-0,14	-0,01	-0,15	-0,13	-1,09	-0,47	-0,44	-0,41	-0,61	-0,20	-0,43	-0,11	-0,08
	C	-0,21	0,19	-0,61	-0,25	-0,25	-0,45	-0,26	-1,09	-0,33	-0,41	-0,24	0,25	-1,11	-1,12	-1,12	-1,11
	A	0,24	-0,41	-0,22	-0,07	-0,22	-0,48	0,37	-1,03	-0,20	-0,52	0,61	-0,12	-0,15	-0,27	1,34	0,33
	G	-1,07	-1,12	-1,13	-1,01	-1,08	-1,09	-1,09	-1,18	-1,11	-1,14	-1,09	-1,05	-1,10	-1,10	-0,92	-0,81
GAA	T	1,22	0,24	0,37	0,15	0,16	-0,31	0,35	-0,92	0,53	-0,28	0,77	-0,02	0,30	-0,16	0,74	0,37
	C	0,17	0,22	-0,25	0,04	-0,17	-0,15	-0,08	-1,10	0,23	-0,20	0,35	0,26	-1,01	-1,10	-1,02	-1,08
	A	1,16	0,24	1,74	1,20	0,84	0,46	1,28	-0,81	2,20	0,91	3,34	1,71	0,42	0,13	1,78	0,71
	G	0,26	-0,09	-0,06	0,21	0,20	0,01	0,47	-1,03	0,30	0,02	1,78	0,81	-0,08	0,04	1,01	0,36
GAG	T	0,82	0,22	-0,26	-0,00	0,23	-0,43	-0,10	-1,10	-0,41	-0,68	-0,19	0,03	-0,21	0,11	0,36	-0,02
	C	-0,04	-0,12	-0,34	0,55	0,01	-0,02	0,96	-0,75	-0,16	-0,36	0,24	0,66	-1,08	-1,06	-0,91	-1,02
	A	0,61	0,07	0,18	0,86	0,41	0,31	0,99	-0,64	1,08	0,05	2,13	0,98	-0,06	-0,13	1,58	0,72
	G	0,41	0,17	-0,26	0,58	1,30	0,72	1,27	-0,60	0,23	-0,26	1,05	0,79	-0,20	-0,16	0,64	0,19
GGT	T	0,97	0,47	-0,24	0,01	-0,16	-0,43	0,23	-1,14	-0,41	-0,68	-0,33	-0,61	-0,34	-0,17	-0,36	-0,26
	C	0,11	0,31	-0,64	-0,21	-0,36	-0,46	-0,52	-1,13	-0,37	-0,43	-0,28	0,33	-1,14	-1,14	-1,13	-1,08
	A	-0,13	-0,67	-0,37	-0,49	-0,60	-0,73	-0,39	-1,14	-0,31	-0,66	-0,01	-0,43	-0,58	-0,68	-0,16	-0,40
	G	-0,19	-0,48	-0,55	0,47	-0,19	-0,29	-0,09	-0,94	0,03	-0,16	0,36	0,18	0,09	0,59	0,45	0,88
GGC	T	0,27	0,01	-0,36	-0,19	0,07	-0,13	1,29	-1,04	-0,48	-0,65	0,10	-0,45	0,05	0,24	1,10	1,90
	C	0,06	0,58	-0,57	0,29	-0,15	-0,26	-0,08	-0,99	-0,22	-0,25	0,44	0,88	-1,06	-0,99	-0,79	-0,74
	A	0,01	-0,40	-0,41	0,67	-0,22	-0,28	0,37	-0,81	-0,26	-0,16	0,36	0,00	-0,13	0,04	1,05	1,41
	G	-1,10	-1,10	-1,13	-0,36	-1,04	-0,80	-0,91	-0,96	-1,08	-0,99	-1,06	-0,99	-1,07	-0,91	-0,69	-0,52
GGA	T	0,47	-0,31	0,83	-0,30	-0,33	-0,57	0,11	-1,09	0,01	-0,64	-0,07	-0,51	-0,21	-0,42	0,02	0,25
	C	-0,14	-0,42	-0,31	-0,22	-0,42	-0,51	-0,40	-1,10	-0,11	-0,29	-0,15	0,07	-1,07	-1,08	-1,10	-1,04
	A	0,46	-0,22	0,16	0,26	-0,07	-0,44	0,17	-1,06	1,07	0,60	1,93	0,94	0,21	0,19	1,08	0,96
	G	0,60	-0,15	-0,43	0,57	0,17	-0,12	0,07	-0,98	0,38	0,01	1,56	0,72	0,54	1,93	1,01	0,74
GGG	T	0,67	-0,15	-0,57	-0,33	-0,11	-0,52	-0,30	-1,11	-0,50	-0,74	-0,42	-0,51	-0,22	-0,35	0,09	0,51
	C	-0,12	-0,01	-0,61	0,41	-0,10	-0,34	-0,08	-0,93	-0,11	-0,28	0,23	0,61	-0,91	-0,85	-0,92	-0,74
	A	0,91	-0,52	-0,36	0,01	0,04	-0,45	-0,00	-1,03	0,05	-0,31	1,06	0,87	-0,13	-0,13	0,80	2,61
	G	0,07	-0,46	-0,64	0,04	0,00	-0,24	-0,07	-0,89	-0,31	-0,38	0,47	0,50	-0,37	-0,31	-0,08	-0,51

Tabellenverzeichnis

Abbildungsverzeichnis

Literatur

[1] F.H.C. Crick, *The origin of the genetic code*, Journal of Molecular Biology,Volume 38, Issue 3,1968, Pages 367-379,

[2] J. D. Watson, F. H. C. Crick, *A Structure for Deoxyribose Nucleic Acid* Nature, 1953

[3] E. Kreyszig, *Advanced Engineering Mathematics (Fourth ed.)*, p. 880, eq. 5. ISBN 0-471-02140-7, 1979

[4] B. Klaucke, *Der Effekt von nicht zufälligen Verteilungen in codierenden Daten auf die Mutationsstabilität des genetischen Codes*, Universität zu Lübeck, 2017. Bachelorarbeit

[5] Keser, S. *The DNA is More than One in a Million.* Universität zu Lübeck, 2016. Bachelorarbeit.

[6] Geyer, R. *Frameshift Mutations of the Genetic Code and Their Impact on the Polarity Conservation of Amino Acids.* Universität zu Lübeck, 2014. Bachelorarbeit.

[7] Woese, C. R., Dugre, D. H., Saxinger, W. C., and Dugre, S. A. .*The Molecular Basis for the Genetic Code.* Proc Natl Acad Sci USA 55(4) (1996), 966–974.

[8] Haig, D. und Hurst, L. D. *A Quantitative Measure of Error Minimization in the Genetic Code.* Journal of Molecular Evolution 33 (1991), 412–417.

[9] Freeland, S. J. und Hurst, L. D. *The Genetic Code Is One in a Million.* Journal of Molecular Evolution 47 (1998), 238–248.

[10] *Xorshift RNGs*, George Marsaglia, The Florida State University, 2003

[11] William H. Press, Saul A. Teukolsky, William T. Vetterling, Brian P Flannery, *Numerical Recipes, The Art of Scientific Computing, Third Edition*,Cambridge, Page 346f., 2007

[12] J. Kyte, RF. Doolittle, *A simple method for displaying the hydropathic character of a protein.*, J Mol Biol 157:105-132m, 1982

[13] L. Sagan, *On the origin of mitosing cells*, Journal of theoretical biology, 1967 Mar;14(3):255-74.

[14] International Human Genome Sequencing Consortium, *Finishing the euchromatic sequence of the human genome*, Nature 431, 931–945, 2004

[15] R. Grantham *Amino acid difference formula to help explain protein evolution.* Science 185:862-864, 1974

[16] Louise T. Chow, Richard E. Gelinas, Thomas R. Broker, Richard J. Roberts, *An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA*, Cell. 12 (1): 1–8., 1977

[17] D.R. Lide, *Handbook of Chemistry and Physics*, 85th Edition, CRC Press, 2004, Section 7, Page 1.

[18] W W. De Jong, L. Ryden, *Causes of more frequent deletions than insertions in mutations and protein evolution.* Nature. 1981 Mar 12;290(5802):157-9.

[19] M W. Nachman, S L. Crowell, *Estimate of the Mutation Rate per Nucleotide in Humans* GENETICS September 1, 2000 vol. 156 no. 1 297-304

[20] *BioJava 4.1.0 Open Source Bibliothek*, Website: http://biojava.org Quellcode: https://github.com/biojava/biojava/

[21] *Wikipedia: Code-Sonne*, gemeinfreies Bild, Quelle: https://de.wikipedia.org/wiki/Code-Sonne#/media/File:Aminoacids_table.svg

[22] *NCBI GenBank*, Website: https://www.ncbi.nlm.nih.gov/genbank/

[23] *Apache commons-io 1.3.2 Open Source Bibliothek*, Website: https://commons.apache.org/proper/commons-io/

[24] NCBI *The Consensus CDS (CCDS) project*, Jan. 2018. https://www.ncbi.nlm.nih.gov/projects/CCDS/CcdsBrowse.cgi

[25] NCBI, *The Genetic Codes*, Webseite: https://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi?chapter=tgencodes#SG2

[26] UIPAC, *DNA Sequence Format*, Webseite: https://iupac.org/, Codes bezogen über: https://www.genomatix.de/online_help/help/sequence_formats.html#IUPAC