

Sara Kemmler 5760949

Robin Bonkaß 5769588

1	2	3	4	$\Sigma$

## Übungsblatt Nr. 08

(Abgabetermin 30.06.22)

### Aufgabe 1

Oxford Nanopore Sequenzierung:

- Nutzt flow cells, welche ein Array von kleinen “Löchern” sogenannten Nanoporen besitzen. Diese sind eingebettet in eine elektroresistente Membran.
- Oxford Nanopore ist die einzige Sequenzieretechnologie, welche Echtzeitanalyse in allen möglichen Formaten (von Geräten im Taschenformat bis hin zu Geräten, welche ganze Populationen assemblieren können) zur Verfügung stellt. Vorteile hiervon sind unter anderem der schnelle Zugriff auf zeitkritische Informationen und mehr Kontrolle über das Sequenzierexperiment
- Die Sequenzierung ist nur durch die Länge des DNA/RNA-Fragments begrenzt, welches der Pore repräsentiert wird, weshalb repetitive Regionen überspannt und strukturelle Varianten aufgelöst werden können (möglich durch ultra-long reads)

Quelle: Technologies o. D.

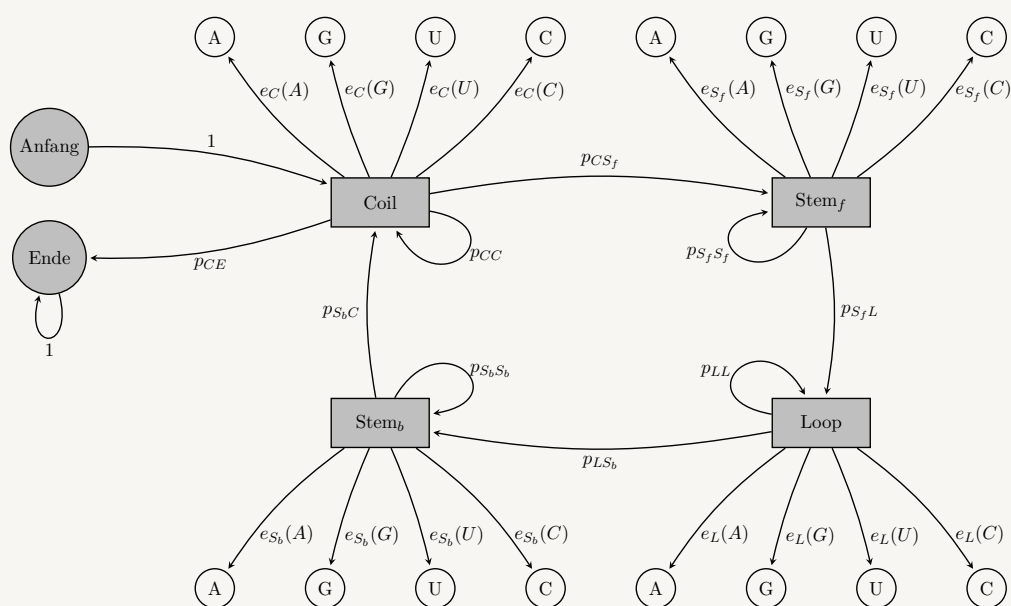
4

### Aufgabe 2

Hidden States: Coil, Stem, Loop

Emission states: Nukleotide (A, G, U, C)

Anzahl an Parametern: 27



3

## Aufgabe 3

Observed Frequencies:

$c_{st}$	b	A	T	C	G	e
b	0	0	1	3	2	0
A	0	1	1	3	3	1
T	0	4	6	4	7	1
C	0	2	5	3	5	1
G	0	2	9	3	5	3
e	0	0	0	0	0	0

Probabilities (transition matrix):

$p_{st}$	b	A	T	C	G	e
b	0	0	0.17	0.50	0.33	0
A	0	0.11	0.11	0.33	0.33	0.11
T	0	0.18	0.27	0.18	0.32	0.05
C	0	0.13	0.31	0.19	0.31	0.06
G	0	0.09	0.41	0.14	0.23	0.14
e	0	0	0	0	0	0

2

## Aufgabe 4

a) - e)

Der Code, der sich im file `Sara_Kemmler_Robin_Bonkass_A8.py` befindet, kann mit folgendem Befehl ausgeführt werden:

```
python3 Robin_Bonkass_Sara_Kemmler_A8.py -cds ./material/cds_set_new.fasta -notcds
./material/notcds_set_new.fasta -contig ./material/contig.fasta -namecds cds_matrix
-namenotcds notcds_matri
```

Zum Copy-Pasten:

```
python3 Robin_Bonkass_Sara_Kemmler_A8.py -cds ./material/cds_set_new.fasta -notcds ./material/notcds_set_new.fasta -contig ./material/contig.fasta -namecds cds_matrix -namenotcds notcds_matrix
```

Das Programm speichert zunächst zwei berechnete Transition Matizes aus den übergebenen files `cds_set_new.fasta` und `notcds_set_new.fasta`. Diese werden als `cds_matrix.txt` und `notcds_matrix.txt` gespeichert. Dann werden diese Matrizen benutzt um die Wahrscheinlichkeit der Sequenz in der Datei `contig.fasta` bezüglich der beiden Modelle (Matrizen) zu berechnen. Diese werden in der Konsole ausgegeben. Zudem wird die log-odds ratio berechnet und ebenfalls in der Konsole ausgegeben.

**f)**

Die berechnete log-odds ratio der Sequenz  $x$  bezüglich dem  $\text{model}^+ := \text{Matrix des cds\_set}$  und  $\text{model}^- := \text{Matrix des notcds\_set}$  ist  $\approx 10.39$ . Daraus folgt:

$$\begin{aligned} 10.39 > 0 &\implies \mathbb{P}(x \mid \text{model}^+) > \mathbb{P}(x \mid \text{model}^-) \\ &\implies \text{sequence } x \text{ ist wahrscheinlicher nicht ein CDS} \end{aligned}$$

11

## Literatur

[Tec] Oxford Nanopore Technologies. How nanopore sequencing works. URL: <https://nanoporetech.com/how-it-works>.