Metabarcoding Pipeline Building

An introduction with hands on exercises

CUSO – DPEE Activity

Gerhard Thallinger, PhD, Graz University of Technology

Rachel Korn, PhD Student, Université de Fribourg

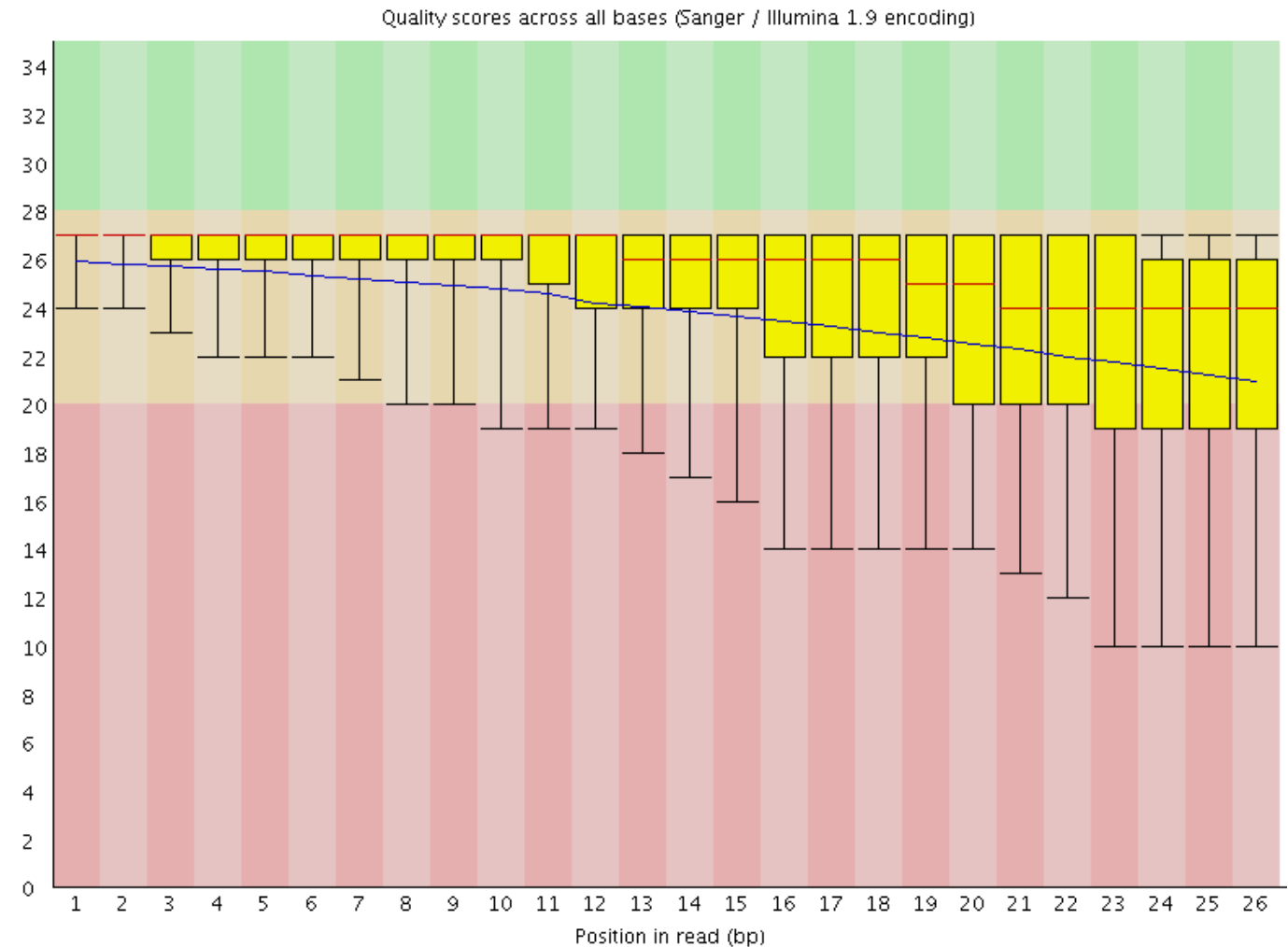Magdalena Steiner, PhD, Agroscope Wädenswil

Gerhard Thallinger, PhD, Graz University of Technology
Rachel Korn, PhD, Université de Fribourg
Magdalena Steiner, PhD, Agroscope Wädenswil

Remove primers from the reads

Au et al. 2016

Metabarcoding Pipeline Building

Gerhard Thallinger, PhD, Graz University of Technology
Rachel Korn, PhD, Université de Fribourg
Magdalena Steiner, PhD, Agroscope Wädenswil

July 2021 | 2

Quality filtering

- Usually a decrease in quality towards the end of the reads (performance drops during a run)
- Forward reads sometimes of higher quality than reverse reads

Quality scores across all bases (Sanger / Illumina 1.9 encoding)

Position in read (bp)

Metabarcoding Pipeline Building

Gerhard Thallinger, PhD, Graz University of Technology
Rachel Korn, PhD, Université de Fribourg
Magdalena Steiner, PhD, Agroscope Wädenswil

July 2021 | 3

Phred quality score

- Logarithmically related to the base-calling error probabilities P

  - $Q = -10 \log_{10} P \Leftrightarrow P = 10^{-10/Q}$

Phred Quality Score	Probability of incorrect base call	Base call accuracy
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1000	99.9%
40	1 in 10,000	99.99%
50	1 in 100,000	99.999%
60	1 in 1,000,000	99.9999%

Wikipedia Phred quality score 2021

Metabarcoding Pipeline Building

Gerhard Thallinger, PhD, Graz University of Technology
Rachel Korn, PhD, Université de Fribourg
Magdalena Steiner, PhD, Agroscope Wädenswil

July 2021 | 4

maxEE

- Sum of the base-calling error probabilities *P* → dependent on read length

  - $E = \sum_i p_i = \sum_i 10^{-Q_i/10}$

Edgar et al. 2015

Gerhard Thallinger, PhD, Graz University of Technology
Rachel Korn, PhD, Université de Fribourg
Magdalena Steiner, PhD, Agroscope Wädenswil

Rarefaction curves

- Plateauing indicates sufficient sequencing depth

- Necessary sequencing depth is proportional to α-diversity

Gerhard Thallinger, PhD, Graz University of Technology
Rachel Korn, PhD, Université de Fribourg
Magdalena Steiner, PhD, Agroscope Wädenswil