



Quality Control Report

March, 02 2020

Contact us: info@taugc.com

About us: www.taugc.com

DISCLAIMER

This report is intended for general guidance and information purposes only. This report is under no circumstances intended to be used or considered directly for any kind of scientific or academic publication and clinical diagnosis. Please note that this is reference material for the original authors of the study. The report is to be considered only as a tool for hypotheses and data interpretation by the authors along with all data provided.



DISCLAIMER

The data used for constructing this material in the report is obtained directly from the customer (customer). We have taken reasonable care to ensure that, and to the best of our knowledge, material information contained herein is in accordance with information and guidelines provided by the customer. All data presented are prepared by TauGC Bioinformatics in accordance with best practices from third parties, as part of the service included in the report. The estimates are subject to uncertainties, misinterpretations and other factors that may cause actual events to differ materially from any anticipated hypotheses. Please note that we make no assurance that the underlying statements are free from errors. Neither the authors nor TauGC Bioinformatics is making any representation or warranty, express or implied, as to the accuracy or completeness of data presented in this report and none of the authors or TauGC Bioinformatics will have any liability towards any other person resulting in incorrect use this report. TauGC do not intend, and do not assume any obligation to update or correct the information included in this report.

The information contained herein may be subject to interpretation by authors. TauGC does not accept any form of liability, neither legally nor financially, for loss (direct or indirect) caused by the understanding and/or use of this report or its content.

This report is only intended for the recipients, and should not be copied or otherwise distributed, in whole or in part, to any other person. This report is subject to Brazilian law, and any dispute arising in respect of this report is subject to the exclusive jurisdiction of Brazilian courts.

CONSIDERATIONS ABOUT AUTHORSHIP & LIABILITY

TauGC Bioinformatics is not to be considered as author or co-author of any kind publication solely based on direct use of this report. Any consideration for authorship by TauGC must be discussed further by authors and TauGC Bioinformatics.

TauGC and its members follow the Nature Journals Policy for Authorship (<https://www.nature.com/authors/policies/authorship.html>) adapted from McNutt et al. 2018 (DOI: 10.1073/pnas.1715374115).

TauGC Bioinformatics - Rua Apiacas 886, São Paulo - SP

This report is only intended for the recipients, and should not be copied or otherwise distributed, in whole or in part, to any other person.



Controle de Qualidade

Tab. 1- Informações gerais sobre os dados.

Amostra	Sequenciamento	No. reads	No. reads após trimming
Ecoli	Pair-End	1641235	1514422 (92.27%)

Fig. 1- Distribuição do tamanho das reads (R1 e R2).

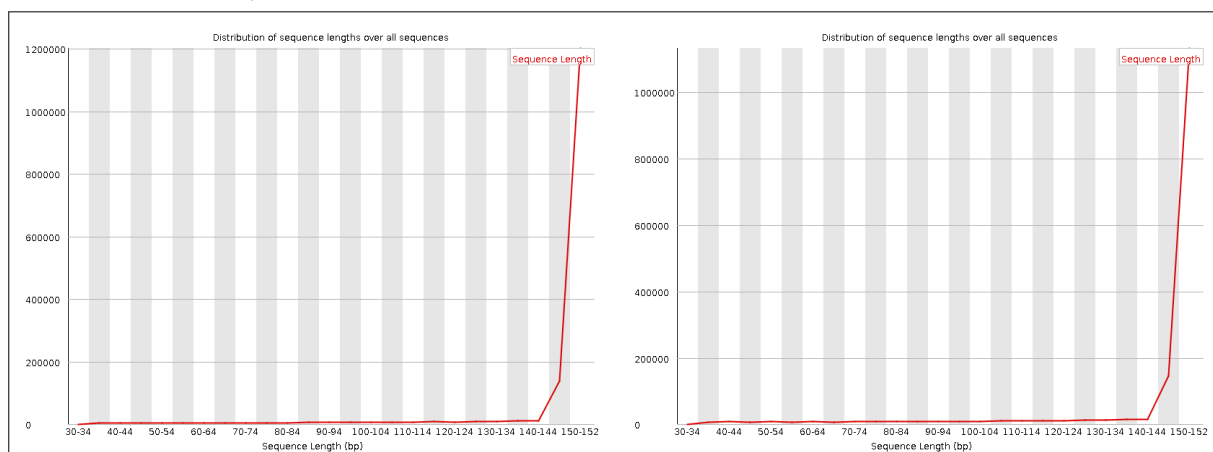
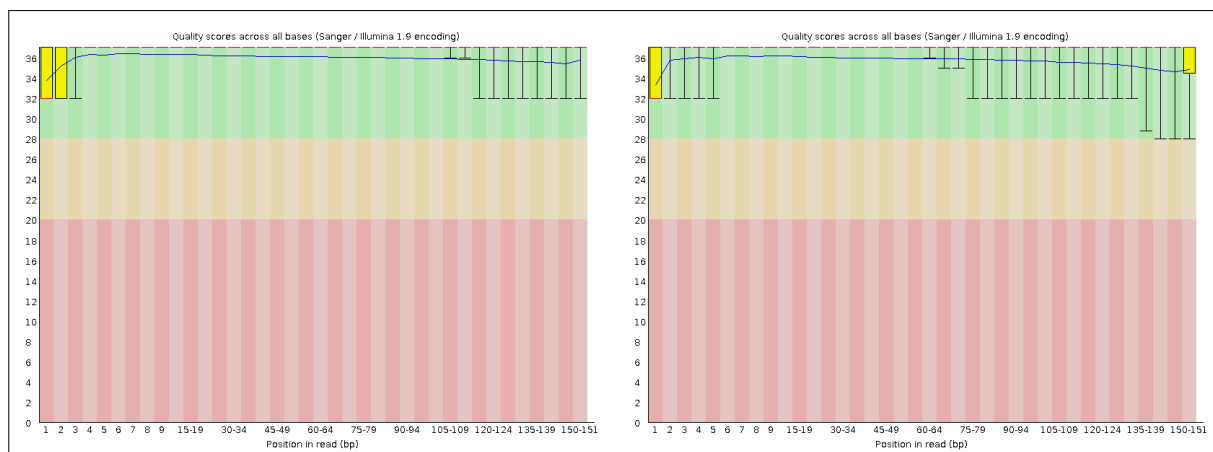


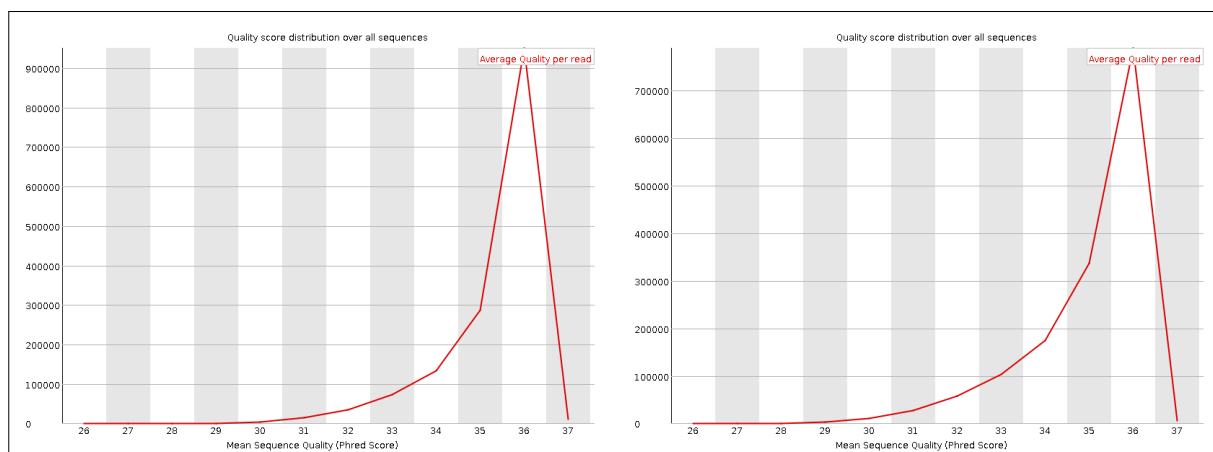


Fig. 2- Qualidade por base das sequências (R1 e R2).



No eixo X está a posição de cada base na 'read'. O eixo Y mostra o 'score' de qualidade. Quanto maior o 'score', melhor. O plano de fundo do gráfico divide o eixo Y em 'scores' bons (verde), razoáveis (laranja) e de baixa qualidade (vermelho).

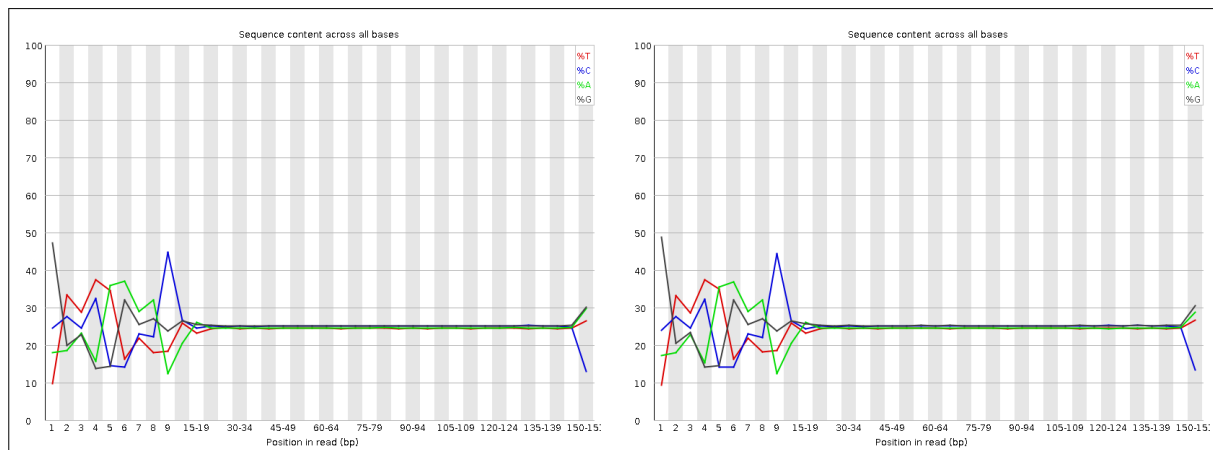
Fig. 3- Média de qualidade por sequência (R1 e R2).



No eixo X temos a média de qualidade ao longo das sequências de todas as leituras (reads) e no eixo Y o número total de 'reads'. É importante notar que a qualidade média das reads deve apresentar um pico na faixa superior do gráfico.

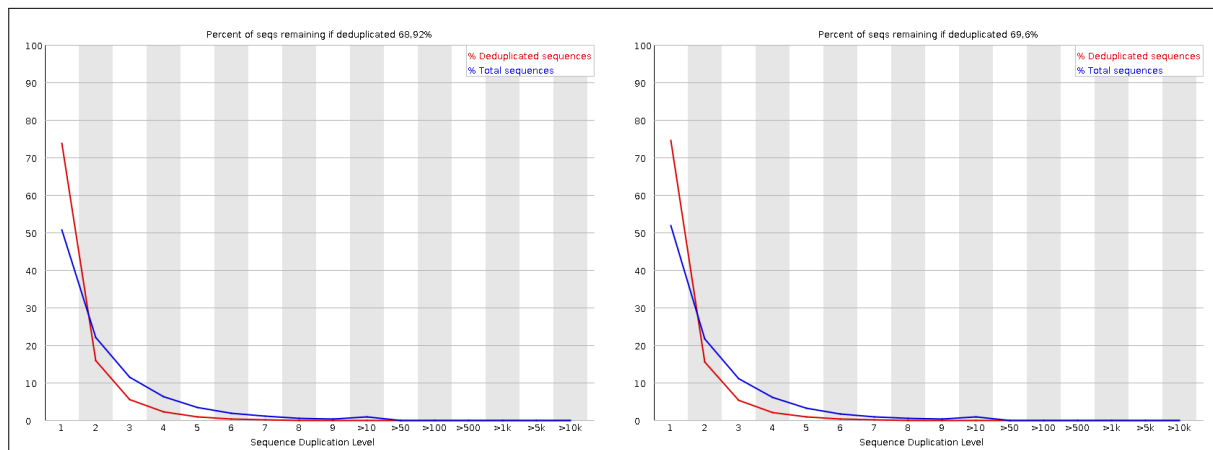


Fig. 4- Conteúdo das reads por base (R1 e R2).



Porcentagem de cada um dos quatro nucleotídeos (T, C, A, G) em cada posição em todas as reads. O esperado é que haja pouca ou nenhuma diferença entre as quatro bases. A proporção de cada uma deve ser relativamente constante ao longo do comprimento da read com $\%A = \%T$ e $\%G = \%C$, sendo que as linhas neste gráfico devem ser paralelas entre si.

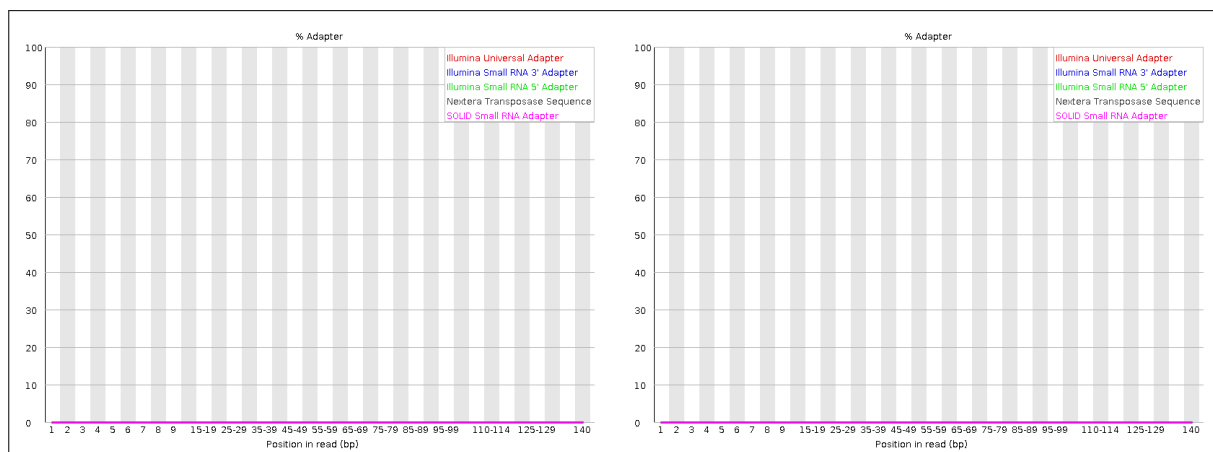
Fig. 5- Nível de duplicação das sequências (R1 e R2).



Espera-se que em uma biblioteca diversificada, a maioria das sequências ocorra apenas uma vez no conjunto final. Um baixo nível de duplicação pode indicar um nível alto de cobertura da sequência alvo, mas um alto nível de duplicação provavelmente indica algum viés de enriquecimento. A linha azul mostra a distribuição dos níveis de duplicação para o conjunto completo de sequências, já a linha vermelha apresenta a distribuição para as sequências deduplicadas. A maioria das sequências deve cair na extremidade esquerda do gráfico nas linhas vermelha e azul. Isso indica uma biblioteca altamente diversificada que não foi sequenciada em excesso.



Fig. 6- Adaptadores nas sequências analisadas (R1 e R2).



No eixo X está a posição nas sequências em pares de base e no eixo Y a porcentagem de adaptador.

References

Batut et al., 2018. Community-Driven Data Analysis Training for Biology. *Cell Systems*, 6(6):752-758.

Bérénice Batut, 2019. Quality Control. </training-material/topics/sequence-analysis/tutorials/quality-control/tutorial.html>. Online; accessed 2020-02-20.