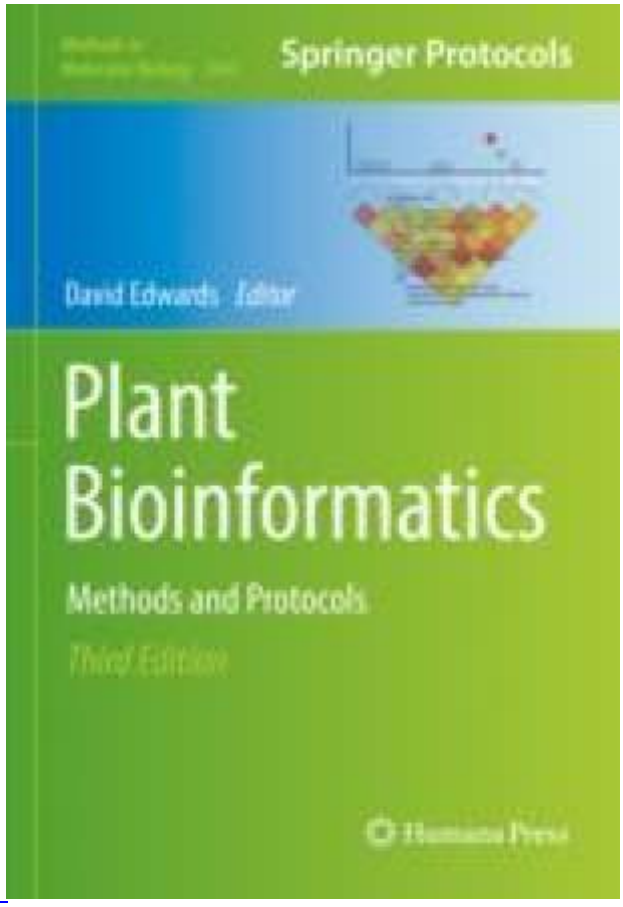

Plant Bioinformatics

Plant Bioinformatics pp 211-232 | Cite as

Trimming and Validation of Illumina Short Reads Using Trimmomatic, Trinity Assembly, and Assessment of RNA-Seq Data

- Authors

- Steven O. Sewe
- Gonçalo Silva
- Paulo Sicat
- Susan E. Seal

- Paul Visendi

Protocol
First Online: 01 January 2022
Part of the Methods in Molecular Biology book series (MIMB, volume 2443)

Abstract

Next-generation sequencing (NGS) technologies can generate billions of reads in a single sequencing run. However, with such high-throughput comes quality issues which have to be addressed before undertaking downstream analysis. Quality control on short reads is usually performed at default settings due to a lack of in-depth understanding of a particular software's parameters and their effect if changed on the output. Here we demonstrate how to optimize read trimming using Trimmomatic. We highlight the benefits of trimming by comparing the quality of transcripts assembled using trimmed and untrimmed reads.

Key words

Quality control Illumina adapters Transcriptome assembly Trimmomatic Trinity Busco Fastqc Quast

References

1. 1.

   Goodwin S, McPherson JD, McCombie WR (2016) Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet 17:333–351. https://doi.org/10.1038/nrg.2016.49CrossRefPubMedGoogle Scholar

2. 2.

   Payá-Milans M, Olmstead JW, Nunez G et al (2018) Comprehensive evaluation of RNA-seq analysis pipelines in diploid and polyploid species. Gigascience 7:giy132. https://doi.org/10.1093/gigascience/giy132CrossRefPubMedCentralGoogle Scholar

3. 3.

   Turner FS (2014) Assessment of insert sizes and adapter content in fastq data from NexteraXT libraries. Front Genet 5:5. https://doi.org/10.3389/fgene.2014.00005CrossRefPubMedPubMedCentralGoogle Scholar

4. 4.

   Yang S-F, Lu C-W, Yao C-T, Hung C-M (2019) To trim or not to trim: effects of read trimming on the de novo genome assembly of a widespread east Asian Passerine, the Rufous-Capped Babbler (Cyanoderma ruficeps

Blyth). Genes
10:737. https://doi.org/10.3390/genes10100737CrossRefPubMedCentralGoogle Scholar

5. 5.

Pereira R, Oliveira J, Sousa M (2020) Bioinformatics and computational tools for next-generation sequencing analysis in clinical genetics. J Clin Med
9:132. https://doi.org/10.3390/jcm9010132CrossRefPubMedCentralGoogle Scholar

6. 6.

Pfeiffer F, Gröber C, Blank M et al (2018) Systematic evaluation of error rates and causes in short samples in next-generation sequencing. Sci Rep
8:10950. https://doi.org/10.1038/s41598-018-29325-6CrossRefPubMedPubMedCentralGoogle Scholar

7. 7.

Abnizova I, te Boekhorst R, Orlov YL (2017) Computational errors and biases in short read next generation sequencing. J Proteomics Bioinform
10:1–17. https://doi.org/10.4172/jpb.1000420CrossRefGoogle Scholar

8. 8.

Fabbro CD, Scalabrin S, Morgante M, Giorgi FM (2013) An extensive evaluation of read trimming effects on Illumina NGS data analysis. PLoS One
8:e85024. https://doi.org/10.1371/journal.pone.0085024CrossRefPubMedPubMedCentralGoogle Scholar

9. 9.

Bansal V (2017) A computational method for estimating the PCR duplication rate in DNA and RNA-seq experiments. BMC Bioinformatics
18:43. https://doi.org/10.1186/s12859-017-1471-9CrossRefPubMedPubMedCentralGoogle Scholar

10. 10.

Visendi P, Berkman PJ, Hayashi S et al (2016) An efficient approach to BAC based assembly of complex genomes. Plant Methods
12:778. https://doi.org/10.1186/s13007-016-0107-9CrossRefGoogle Scholar

11. 11.

Heydari M, Miclotte G, Demeester P et al (2017) Evaluation of the impact of Illumina error correction tools on de novo genome assembly. BMC Bioinformatics 18:374. https://doi.org/10.1186/s12859-017-1784-8CrossRefPubMedPubMedCentralGoogle Scholar

12. 12.

Tan G, Opitz L, Schlapbach R, Rehrauer H (2019) Long fragments achieve lower base quality in Illumina paired-end sequencing. Sci Rep 9:2856. https://doi.org/10.1038/s41598-019-39076-7CrossRefPubMedPubMedCentralGoogle Scholar

13. 13.

van Gurp TP, McIntyre LM, Verhoeven KJF (2013) Consistent errors in first strand cDNA due to random hexamer mispriming. PLoS One 8:e85583. https://doi.org/10.1371/journal.pone.0085583CrossRefPubMedPubMedCentralGoogle Scholar

14. 14.

Mbandi SK, Hesse U, Rees DJG, Christoffels A (2014) A glance at quality score: implication for de novo transcriptome reconstruction of Illumina reads. Front Genet 5:17. https://doi.org/10.3389/fgene.2014.00017CrossRefPubMedPubMedCentralGoogle Scholar

Copyright information

Authors and Affiliations

- Steven O. Sewe
  - 1
- Gonçalo Silva
  - 1
- Paulo Sicat
  - 1
- Susan E. Seal
  - 1
- Paul Visendi

1. 1.Natural Resources Institute University of Greenwich Kent UK
2. 2.Centre for Agriculture and the Bioeconomy Queensl and University of Technology Brisbane Australia

About this protocol

Cite this protocol as:

Sewe S.O., Silva G., Sicat P., Seal S.E., Visendi P. (2022) Trimming and Validation of Illumina Short Reads Using Trimmomatic, Trinity Assembly, and Assessment of RNA-Seq Data. In: Edwards D. (eds) Plant Bioinformatics. Methods in Molecular Biology, vol 2443. Humana, New York, NY. https://doi.org/10.1007/978-1-0716-2067-0_11

- First Online 01 January 2022
- DOI https://doi.org/10.1007/978-1-0716-2067-0_11
- Publisher Name Humana, New York, NY
- Print ISBN 978-1-0716-2066-3
- Online ISBN 978-1-0716-2067-0
- eBook Packages Springer Protocols