

Jan 23, 2025

Comparative Evaluation of Individual and Pooled Sequencing for Population Genomics Assessment

DOI

<https://dx.doi.org/10.17504/protocols.io.5jyl8dkr7g2w/v1>

Luciana Gillman¹, Federico Condon², Cesar Petroli³, Mercedes Rivas⁴

¹CURE-UdelaR; ²INIA; ³CIMMYT; ⁴Fagro and CURE-UdelaR



Luciana Gillman

CURE-UdelaR

Create & collaborate more with a free account

Edit and publish protocols, collaborate in communities, share insights through comments, and track progress with run records.

Create free account

OPEN  ACCESS



DOI: <https://dx.doi.org/10.17504/protocols.io.5jyl8dkr7g2w/v1>

Protocol Citation: Luciana Gillman, Federico Condon, Cesar Petroli, Mercedes Rivas 2025. Comparative Evaluation of Individual and Pooled Sequencing for Population Genomics Assessment. **protocols.io**

<https://dx.doi.org/10.17504/protocols.io.5jyl8dkr7g2w/v1>

License: This is an open access protocol distributed under the terms of the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Protocol status: Working

We use this protocol and it's working

Created: January 19, 2025

Last Modified: January 23, 2025

Protocol Integer ID: 118676

Keywords: pooled sequencing for population genomics assessment, population genomics assessment, pooled sequencing, genetic diversity, allele frequency, minor allele frequency, correlation between allele frequency, sequencing depth, sample size, population structure, potential effects of sample size, seq

Funders Acknowledgements:

Luciana Gillman

Grant ID: POS_NAC_2018_1_151772

Luciana Gillman

Grant ID: pasantía 322

Mercedes Rivas

Grant ID: CURE

Federico Condon

Grant ID: INIA

Abstract

This protocol evaluates individual sequencing (ind-seq) and pooled sequencing (pool-seq) datasets, comparing the results obtained with each method. It investigates the correlation between allele frequencies calculated using both approaches to compare their efficacy. Genetic diversity and population structure are also analyzed using both methods. These comparisons explicitly account for the potential effects of sample size, missing data, sequencing depth, and minor allele frequencies.

Attachments



[Fig2.tif](#)

1.4MB



Materials

Plant Material: Seeds from your plant or plant.

- Equipment:

- Seed germination trays.
- NovaDryer-F104 lyophilizer.
- Pipettes, centrifuge, and cold storage facilities.
- Sequencing platform: Illumina Novaseq 6000.
- R-compatible system for bioinformatics analysis.

- Reagents:

- CTAB buffer for DNA extraction.
- Ethanol for DNA precipitation.
- Restriction enzymes (PstI and MseI).
- Barcoded adapters and PCR primers.
- Library preparation kits (DArTseq® protocol).



Seed Germination and Plant Preparation

3w 0d 3h

- 1 Germinate seeds under International Seed Testing Association standards.
- 2 Transplant 60 seedlings per accession into trays.
- 3 Allow seedlings to grow until sufficient leaf tissue develops.

1w

3h

2w

Sample Collection

19h 30m

- 4 **Individual Samples:** Harvest 120 mg of leaf tissue from each of the 60 plants per accession.
- 5 **Pooled Samples:** Combine tissue from 20 randomly selected plants to create a 20-sample pool (6 mg tissue per plant).
- 5.1 Incrementally add 10 seedlings to form 30-, 40-, 50-, and 60-sample pools, maintaining 4 mg tissue per plant.
- 5.2 Record individual identities of plants in each pool for comparison with individual sequencing.

2h

2h

15h



30m

Tissue Processing

3d

- 6 Lyophilize all leaf tissue samples for 72 hours using a NovaDryer-F104.
- 7 Prepare all the sample and paper for shipping.

2d

1d

DNA Extraction and Library Preparation

1d 7h 30m

- 8 Extract DNA using the modified CTAB method.

1d



9 Digest DNA with PstI and MseI restriction enzymes.

2h

10 Ligate barcoded adapters and PCR amplify.

2h

11 PCR

3h

12 Pool PCR products for sequencing. Take into account sequence each pool at depths of 0.9 Mr and 1.5 Mr and individuals to 0.9 Mr.

30m

Sequencing

1d

13 Sequence all samples

1d

SNP Calling

1d

14 Call SNPs using DArTsoft14 software.

1d

Data Analysis

1d 6h

15 Calculate allele frequencies for individual and pooled samples using R scripts.

<https://github.com/LucianaGillman/ind-pool-seq-hexaploid>

3h

16 Compare SNP metrics (e.g., Representativity, CCC) between datasets.

https://github.com/LucianaGillman/ind-pool-seq-hexaploid/blob/main/calculation_CCC_REP_2.3.3.R

3h

16.1 Perform ANOVA to assess effects of sequencing depth, sample size, and minor allele frequency thresholds on SNP metrics.

https://github.com/LucianaGillman/ind-pool-seq-hexaploid/blob/main/ANOVAS_2.7.R

17 Perform population diversity analyses (e.g., expected heterozygosity, genetic distance) using the dartR and BioR packages.

<https://github.com/LucianaGillman/ind-pool-seq-hexaploid>

1d

Protocol references

- Adamack, Aaron T., and Bernd Gruber. 2014. "PopGenReport: Simplifying Basic Population Genetic Analyses in R." *Methods in Ecology and Evolution* 5(4):384–87. doi: 10.1111/2041-210X.12158.
- Anon. 2000. "Corrections." *Biometrics* 56(1):324–25. doi: 10.1111/J.0006-341X.2000.00324.X.
- Brazeiro, Alejandro, Daniel Panario, Alvaro Soutullo, Ofelia Gutiérrez, Angel Segura, and Patricia Mai. 2012. *Clasificación y Delimitación de Las Eco-Regiones de Uruguay. Informe Técnico*. doi: 10.13140/2.1.1328.0328.
- Chambers, John M., Anne E. Freeny, and Richard M. Heiberger. 1992. "Analysis of Variance; Designed Experiments." Pp. 145–93 in *Statistical Models in S*, edited by J. M. Chambers and J. M. Hastie. Wadsworth & Brooks/Cole.
- CIMMYT. 2005. *Laboratory Protocols: CIMMYT Applied Molecular Genetics Laboratory*. Third edition. Mexico, D.F.: CIMMYT.
- Condón, F., M. Jaurena, R. Reyno, C. Otaño, and F. A. Lattanzi. 2017. "Spatial Analysis of Genetic Diversity in a Comprehensive Collection of the Native Grass *Bromus Auleticus* Trinius (Ex Nees) in Uruguay." *Grass and Forage Science* (December 2015). doi: 10.1111/gfs.12275.
- Doyle, J. J., and J. L. Doyle. 1987. "A Rapid DNA Isolation Procedure for Small Quantities of Fresh Leaf Tissue." *Phytochemical Bulletin* 19(1):11–15.
- Fox, John, and Sanford Weisberg. 2019. *An R Companion to Applied Regression*,. Third. Thousand Oaks CA: SAGE Publications Ltd.
- Fracassetti, Marco, Philippa C. Griffin, and Yvonne Willi. 2015. "Validation of Pooled Whole-Genome Re-Sequencing in Arabidopsis Lyrata." *PLoS ONE* 10(10):1–15. doi: 10.1371/journal.pone.0140462.
- Gillman_Luciana. 2025. "LucianaGillman/Ind-Pool-Seq-Hexaploid: Assessment_genomic_diversity_contrasting." International Seed Testing Association (ISTA). 2018. *Handbook on Seedling Evaluation*. Fourth edition. edited by R. Don and S. Ducournau.
- Lin, L. I. 1989. "A Concordance Correlation Coefficient to Evaluate Reproducibility." *Biometrics* 45(1):255–68.
- Mijangos, Jose Luis, Bernd Gruber, Oliver Berry, Carlo Pacioni, and Arthur Georges. 2022. "DartR v2: An Accessible Genetic Analysis Platform for Conservation, Ecology, and Agriculture." *Methods in Ecology and Evolution* 2022(May):1–9. doi: 10.1111/2041-210x.13918.
- Oksanen, J., G. Simpson, F. Blanchet, R. Kindt, P. Legendre, P. Minchin, R. O'Hara, P. Solymos, M. Stevens, E. Szoecs, H. Wagner, M. Barbour, M. Bedward, B. Bolker, D. Borcard, G. Carvalho, M. Chirico, M. De Caceres, S. Durand, H. Evangelista, R. Fitz John, M. Friendly, B. Furneaux, G. Hannigan, M. Hill, L. Lahti, D. McGlinn, M. Ouellette, E. Ribeiro Cunha, T. Smith, A. Stier, C. Ter Braak, and Weedon. J. 2022. "Vegan: Community Ecology Package."
- Pacheco, Ángela, Gregorio Alvarado, Francisco Rodríguez, and Juan Burgueño. 2016. "BIO-R (Biodiversity Analysis with R for Windows) Version 3.2" edited by I. M. and W. I. Center.
- Petroli, Cesar,, and Andrzej Kilian. 2019. "DART-SeqTM High- Throughput Genotyping." *CIMMYT Research Data & Software Repository Network* 1:1–5. doi: hdl:11529/10548358.
- R Core Team. 2018. "R: A Language and Environment for Statistical Computing." Retrieved April 24, 2019 (<https://www.r-project.org/>).
- Sansaloni, Carolina, Jorge Franco, Bruno Santos, Lawrence Percival-Alwyn, Sukhwinder Singh, Cesar Petroli, Jaime Campos, Kate Dreher, Thomas Payne, David Marshall, Benjamin Kilian, Iain Milne, Sebastian Raubach, Paul Shaw, Gordon Stephen, Jason Carling, Carolina Saint Pierre, Juan Burgueño, José Crosa, Hui Hui Li, Carlos Guzman, Zakaria Kehel, Ahmed Amri, Andrzej Kilian, Peter Wenzl, Cristobal Uauy, Marianne Banziger, Mario



- Caccamo, and Kevin Pixley. 2020. "Diversity Analysis of 80,000 Wheat Accessions Reveals Consequences and Opportunities of Selection Footprints." *Nature Communications* 11(1). doi: 10.1038/s41467-020-18404-w.
- Sansaloni, Carolina, Cesar Petroli, Damian Jaccoud, Jason Carling, Frank Detering, Dario Grattapaglia, and Andrzej Kilian. 2011. "Diversity Arrays Technology (DART) and next-Generation Sequencing Combined: Genome-Wide, High Throughput, Highly Informative Genotyping for Molecular Breeding of Eucalyptus." *BMC Proceedings* 5(S7):P54. doi: 10.1186/1753-6561-5-s7-p54.
- Schlötterer, Christian, Raymond Tobler, Robert Kofler, and Viola Nolte. 2014. "Sequencing Pools of Individuals-Mining Genome-Wide Polymorphism Data without Big Funding." *Nature Reviews Genetics* 15(11):749–63. doi: 10.1038/nrg3803.
- Stevenson, Mark, Telmo Nunes, Cord Heuer, Jonathon Marshall, Javier Sanchez, Ron Thorn-, Jeno Reiczigel, Jim Robison-cox, Paola Sebastiani, Peter Solymos, Kazuki Yoshida, Sarah Pirikahu, Simon Firestone, Ryan Kyle, Mathew Jay, Charles Reynard, and Maintainer Mark Stevenson. 2019. "Package 'EpiR'." <http://fvas.unimelb.edu.au/veam>.
- Weisburd, David, Chester Britt, David B. Wilson, and Alese Wooditch. 2020. "Measuring Association for Scaled Data: Pearson's Correlation Coefficient." Pp. 479–530 in *Basic Statistics in Criminology and Criminal Justice*. Springer, Cham.