

Reproducible, portable, and efficient ancient genome reconstruction with nf-core/eager

This manuscript ([permalink](#)) was automatically generated from [apeltzer/eager2-paper@9addcce](#) on April 27, 2020.

Authors

- **James A. Fellows Yates**

 [0000-0001-5585-6277](#) ·  [jfy133](#) ·  [jfy133](#)

Department of Archaeogenetics, Max-Planck-Institute for the Science of Human History · Funded by Grant XXXXXXXX

- **Thiseas**

 [XXXX-XXXX-XXXX-XXXX](#) ·  [TCLamnidis](#) ·  [placeholder](#)

Department of Something, University of Whatever · Funded by Grant XXXXXXXX

- **Maxime Borry**

 [0000-0001-9140-7559](#) ·  [maxibor](#) ·  [notmaxib](#)

Department of Archaeogenetics, Max-Planck-Institute for the Science of Human History · Funded by Grant XXXXXXXX

- **Aida**

 [XXXX-XXXX-XXXX-XXXX](#) ·  [placeholder](#) ·  [placeholder](#)

Department of Something, University of Whatever · Funded by Grant XXXXXXXX

- **Zandra**

 [XXXX-XXXX-XXXX-XXXX](#) ·  [placeholder](#) ·  [placeholder](#)

Department of Something, University of Whatever · Funded by Grant XXXXXXXX

- **Stephen**

 [XXXX-XXXX-XXXX-XXXX](#) ·  [placeholder](#) ·  [placeholder](#)

Department of Something, University of Whatever · Funded by Grant XXXXXXXX

- **Maxime U. Garcia**

 [0000-0003-2827-9261](#) ·  [MaxUlysse](#) ·  [gau](#)

Department of Oncology-Pathology, Karolinska Institutet · Funded by Barncancerfonden

- **Judith**

 [XXXX-XXXX-XXXX-XXXX](#) ·  [placeholder](#) ·  [placeholder](#)

Department of Something, University of Whatever · Funded by Grant XXXXXXXX

- **Alexander Peltzer**

 [0000-0002-6503-2180](#) ·  [apeltzer](#) ·  [alex_peltzer](#)

Translational Medicine & Clinical Pharmacology, Boehringer Ingelheim Pharma GmbH & CO KG, Biberach an der Riss

Abstract

The increasing utilization of ancient DNA (aDNA) to address anthropological, palaeontological and biological questions is resulting in a rising diversity in the size of laboratories and scale of analyses being performed. In the context of this heterogeneous landscape, we present [nf-core/eager](#), an advanced and entirely redesigned pipeline for the analysis of ancient DNA molecular data. nf-core/eager builds on existing ideas and concepts introduced in the original EAGER pipeline [1](#), and improves various aspects of the analysis procedure by building on computational frameworks such as Nextflow [\[???\]](#) and nf-core [\[???\]](#). The pipeline aims to address three main points: adaptability to different project sizes, reproducibility to ensure robust standards, and extensibility to ensure that new standard analyses (including metagenomic screening) are accessible to research groups of all sizes. The new version of EAGER has been developed within the nf-core initiative, to ensure high quality software development and maintenance support - which will contribute to a long-term lifecycle for the pipeline. nf-core/eager will assist in ensuring that the power of Next Generation Sequencing (NGS) ancient DNA analysis expand to many different applications in the field of ancient DNA research.

Introduction

Despite difficult beginnings, ancient DNA (aDNA) is now becoming a widely accepted source of short and long term evolutionary and anthropological data and questions. The development of next-generation-sequencing, the most common method used in palaeogenetics, has allowed the recovery of ancient DNA from a wide variety of sources including but not limited to the skeletal remains of animals ([\[???\]](#), [2](#), [3](#), [\[???\]](#)), modern and archaic humans ([4](#), doi:0.1038/s41586-018-0455-x), bacteria ([5](#), [6](#), [7](#)), Viruses ([8](#), [9](#)), plants (cITE), but also coprolites ([10](#)), calculus ([11](#), [\[???\]](#)), sediments ([\[???\]](#), [12](#)), medical slides ([13](#)) parchment ([14](#)), and most recently ancient 'chewing gum' ([15](#), [16](#)), amongst others. The scales of these studies can range from one to thousands of ancient individuals ([\[???\]](#), [\[???\]](#)), and can span single ([\[???\]](#)) to thousands of taxa ([11](#)). These differences of projects have led to a heterogeneous landscape in terms of types of analyses and types of computing resources available to different labs - particularly as countries less developed in archaeogenetics start entering the field.

Two commonly used pipelines in the field have been previously published - PALEOMIX ([17](#)) and EAGER ([1](#)). These two pipelines take a similar approach to link together standard tools used for Illumina NGS data processing (quality control, adapter removal/and or merging, mapping, genotyping etc.), but with a specific focus on tools that are designed for or well-suited for ancient DNA (such as bwa aln ([18](#)) for ultra-short reads and mapDamage ([19](#)) for aDNA characteristic evaluation). While popular, both tools have suffered from a common occurrence in academic bioinformatics which is development stagnation (CITE). Neither tool has had major updates to the pipeline, meaning versions remain out of date and new 'standard' analyses are not being included. The ability to reproduce analysis becomes more challenging as dependencies become more difficult to install, when they become deprecated, as well as increasing 'bug fixing' in hidden internal-only versions of tools in different labs. The quality of aDNA studies then suffer, as new labs trying to enter the field without experience and without easy access to field-established routine analyses results in studies being published without the necessary quality controls to ensure authenticity of their data (Cano/Eisenhofer discussions).

To address these challenges, we have developed nf-core/eager, the second version of the EAGER pipeline. This has been written with Nextflow, a domain-specific-language (DSL) designed for the construction of 'omic analysis pipelines ([\[???\]](#)). Nextflow offers two main benefits over the custom pipeline framework used in EAGER, which is a) immediate integration into various schedulers in High-Performance-Cluster (HPC) environments as well as local computers, and b) widely-adopted but simplified programming language with extensive documentation that most intermediate level bioinformaticians can easily access. These two factors allow most labs to run any pipeline efficiently on their own cluster with minimal configuration, and secondly to make it easier for a range of

contributors to fix bugs and add functionality (rather than having to learn a fully fledged language or a single pipeline's-specific framework). In addition, nf-core/eager has been constructed in the context of the nf-core community framework ([???]). nf-core enforces strict guidelines for best-practises in software development, including continuous integration tests, minimum documentation requirement and code peer-review, while facilitating the adherence to these guidelines by providing templates and active community support. The open community allows support for new contributors and thus a greater source of long-term maintenance of the pipeline.

[. Given the challenges and untapped potential of ancient DNA reproducibility is of...]

References

1. **EAGER: efficient ancient genome reconstruction**

Alexander Peltzer, Günter Jäger, Alexander Herbig, Alexander Seitz, Christian Kniep, Johannes Krause, Kay Nieselt

Genome Biology (2016-03-31) <https://doi.org/gggzpk>

DOI: [10.1186/s13059-016-0918-z](https://doi.org/10.1186/s13059-016-0918-z) · PMID: [27036623](https://pubmed.ncbi.nlm.nih.gov/27036623/) · PMCID: [PMC4815194](https://pubmed.ncbi.nlm.nih.gov/PMC4815194/)

2. **Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse**

Ludovic Orlando, Aurélien Ginolhac, Guojie Zhang, Duane Froese, Anders Albrechtsen, Mathias Stiller, Mikkel Schubert, Enrico Cappellini, Bent Petersen, Ida Moltke, ... Eske Willerslev

Nature (2013-06-26) <https://doi.org/q7n>

DOI: [10.1038/nature12323](https://doi.org/10.1038/nature12323) · PMID: [23803765](https://pubmed.ncbi.nlm.nih.gov/23803765/)

3. **Ancient pigs reveal a near-complete genomic turnover following their introduction to Europe**

Laurent A. F. Frantz, James Haile, Audrey T. Lin, Amelie Scheu, Christina Geörg, Norbert Benecke, Michelle Alexander, Anna Linderholm, Victoria E. Mullin, Kevin G. Daly, ... Greger Larson

Proceedings of the National Academy of Sciences (2019-08-27) <https://doi.org/gf9hnf>

DOI: [10.1073/pnas.1901169116](https://doi.org/10.1073/pnas.1901169116) · PMID: [31405970](https://pubmed.ncbi.nlm.nih.gov/31405970/) · PMCID: [PMC6717267](https://pubmed.ncbi.nlm.nih.gov/PMC6717267/)

4. **137 ancient human genomes from across the Eurasian steppes**

Peter de Barros Damgaard, Nina Marchi, Simon Rasmussen, Michaël Peyrot, Gabriel Renaud, Thorfinn Korneliussen, J. Víctor Moreno-Mayar, Mikkel Winther Pedersen, Amy Goldberg, Emma Usmanova, ... Eske Willerslev

Nature (2018-05-09) <https://doi.org/gd8hs5>

DOI: [10.1038/s41586-018-0094-2](https://doi.org/10.1038/s41586-018-0094-2) · PMID: [29743675](https://pubmed.ncbi.nlm.nih.gov/29743675/)

5. **Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis**

Kirsten I. Bos, Kelly M. Harkins, Alexander Herbig, Mireia Coscolla, Nico Weber, Iñaki Comas, Stephen A. Forrest, Josephine M. Bryant, Simon R. Harris, Verena J. Schuenemann, ... Johannes Krause

Nature (2014-08-20) <https://doi.org/f6nk4g>

DOI: [10.1038/nature13591](https://doi.org/10.1038/nature13591) · PMID: [25141181](https://pubmed.ncbi.nlm.nih.gov/25141181/) · PMCID: [PMC4550673](https://pubmed.ncbi.nlm.nih.gov/PMC4550673/)

6. **Integrative approach using *Yersinia pestis* genomes to revisit the historical landscape of plague during the Medieval Period**

Amine Namouchi, Meriam Guellil, Oliver Kersten, Stephanie Hänsch, Claudio Ottoni, Boris V. Schmid, Elsa Pacciani, Luisa Quaglia, Marco Vermunt, Egil L. Bauer, ... Barbara Bramanti

Proceedings of the National Academy of Sciences (2018-12-11) <https://doi.org/ggfn3h>

DOI: [10.1073/pnas.1812865115](https://doi.org/10.1073/pnas.1812865115) · PMID: [30478041](https://pubmed.ncbi.nlm.nih.gov/30478041/) · PMCID: [PMC6294933](https://pubmed.ncbi.nlm.nih.gov/PMC6294933/)

7. **Ancient genomes reveal a high diversity of *Mycobacterium leprae* in medieval Europe**

Verena J. Schuenemann, Charlotte Avanzi, Ben Krause-Kyora, Alexander Seitz, Alexander Herbig, Sarah Inskip, Marion Bonazzi, Ella Reiter, Christian Urban, Dorthe Dangvard Pedersen, ... Johannes Krause

PLOS Pathogens (2018-05-10) <https://doi.org/gdrj4v>

DOI: [10.1371/journal.ppat.1006997](https://doi.org/10.1371/journal.ppat.1006997) · PMID: [29746563](https://pubmed.ncbi.nlm.nih.gov/29746563/) · PMCID: [PMC5944922](https://pubmed.ncbi.nlm.nih.gov/PMC5944922/)

8. **Ancient hepatitis B viruses from the Bronze Age to the Medieval period**

Barbara Mühlemann, Terry C. Jones, Peter de Barros Damgaard, Morten E. Allentoft, Irina Shevnina, Andrey Logvin, Emma Usmanova, Irina P. Panyushkina, Bazartseren Boldgiv, Tsevel

Bazartseren, ... Eske Willerslev
Nature (2018-05-09) <https://doi.org/gddxvj>
DOI: [10.1038/s41586-018-0097-z](https://doi.org/10.1038/s41586-018-0097-z) · PMID: [29743673](https://pubmed.ncbi.nlm.nih.gov/29743673/)

9. Neolithic and medieval virus genomes reveal complex evolution of hepatitis B

Ben Krause-Kyora, Julian Susat, Felix M Key, Denise Kühnert, Esther Bosse, Alexander Immel, Christoph Rinne, Sabin-Christin Kornell, Diego Yepes, Sören Franzenburg, ... Johannes Krause
eLife (2018-05-10) <https://doi.org/gdhck2>
DOI: [10.7554/elife.36666](https://doi.org/10.7554/elife.36666) · PMID: [29745896](https://pubmed.ncbi.nlm.nih.gov/29745896/) · PMCID: [PMC6008052](https://pubmed.ncbi.nlm.nih.gov/PMC6008052/)

10. The *Prevotella copri* Complex Comprises Four Distinct Clades Underrepresented in Westernized Populations

Adrian Tett, Kun D. Huang, Francesco Asnicar, Hannah Fehlner-Peach, Edoardo Pasolli, Nicolai Karcher, Federica Armanini, Paolo Manghi, Kevin Bonham, Moreno Zolfo, ... Nicola Segata
Cell Host & Microbe (2019-11) <https://doi.org/ggc9dc>
DOI: [10.1016/j.chom.2019.08.018](https://doi.org/10.1016/j.chom.2019.08.018) · PMID: [31607556](https://pubmed.ncbi.nlm.nih.gov/31607556/) · PMCID: [PMC6854460](https://pubmed.ncbi.nlm.nih.gov/PMC6854460/)

11. Pathogens and host immunity in the ancient human oral cavity

Christina Warinner, João F Matias Rodrigues, Rounak Vyas, Christian Trachsel, Natallia Shved, Jonas Grossmann, Anita Radini, Y Hancock, Raul Y Tito, Sarah Fiddyment, ... Enrico Cappellini
Nature Genetics (2014-02-23) <https://doi.org/r4n>
DOI: [10.1038/ng.2906](https://doi.org/10.1038/ng.2906) · PMID: [24562188](https://pubmed.ncbi.nlm.nih.gov/24562188/) · PMCID: [PMC3969750](https://pubmed.ncbi.nlm.nih.gov/PMC3969750/)

12. Neandertal and Denisovan DNA from Pleistocene sediments

Viviane Slon, Charlotte Hopfe, Clemens L. Weiß, Fabrizio Mafessoni, Marco de la Rasilla, Carles Lalueza-Fox, Antonio Rosas, Marie Soressi, Monika V. Knul, Rebecca Miller, ... Matthias Meyer
Science (2017-05-12) <https://doi.org/b6jd>
DOI: [10.1126/science.aam9695](https://doi.org/10.1126/science.aam9695)

13. *Plasmodium vivax* Malaria Viewed through the Lens of an Eradicated European Strain

Lucy van Dorp, Pere Gelabert, Adrien Rieux, Marc de Manuel, Toni de-Dios, Shyam Gopalakrishnan, Christian Carøe, Marcela Sandoval-Velasco, Rosa Fregel, Iñigo Olalde, ... Carles Lalueza-Fox
Molecular Biology and Evolution (2020-03) <https://doi.org/ggqzq2>
DOI: [10.1093/molbev/msz264](https://doi.org/10.1093/molbev/msz264) · PMID: [31697387](https://pubmed.ncbi.nlm.nih.gov/31697387/) · PMCID: [PMC7038659](https://pubmed.ncbi.nlm.nih.gov/PMC7038659/)

14. Paging through history: parchment as a reservoir of ancient DNA for next generation sequencing

M. D. Teasdale, N. L. van Doorn, S. Fiddyment, C. C. Webb, T. O'Connor, M. Hofreiter, M. J. Collins, D. G. Bradley
Philosophical Transactions of the Royal Society B: Biological Sciences (2015-01-19)
<https://doi.org/ggqzq3>
DOI: [10.1098/rstb.2013.0379](https://doi.org/10.1098/rstb.2013.0379) · PMID: [25487331](https://pubmed.ncbi.nlm.nih.gov/25487331/) · PMCID: [PMC4275887](https://pubmed.ncbi.nlm.nih.gov/PMC4275887/)

15. A 5700 year-old human genome and oral microbiome from chewed birch pitch

Theis Z. T. Jensen, Jonas Niemann, Katrine Højholt Iversen, Anna K. Fotakis, Shyam Gopalakrishnan, Åshild J. Vågene, Mikkel Winther Pedersen, Mikkel-Holger S. Sinding, Martin R. Ellegaard, Morten E. Allentoft, ... Hannes Schroeder
Nature Communications (2019-12-17) <https://doi.org/ggfm6x>
DOI: [10.1038/s41467-019-13549-9](https://doi.org/10.1038/s41467-019-13549-9) · PMID: [31848342](https://pubmed.ncbi.nlm.nih.gov/31848342/) · PMCID: [PMC6917805](https://pubmed.ncbi.nlm.nih.gov/PMC6917805/)

16. Ancient DNA from mastics solidifies connection between material culture and genetics of mesolithic hunter-gatherers in Scandinavia

Natalija Kashuba, Emrah Kirdök, Hege Damlien, Mikael A. Manninen, Bengt Nordqvist, Per Persson,

Anders Götherström

Communications Biology (2019-05-15) <https://doi.org/ggqzqz>

DOI: [10.1038/s42003-019-0399-1](https://doi.org/10.1038/s42003-019-0399-1) · PMID: [31123709](https://pubmed.ncbi.nlm.nih.gov/31123709/) · PMCID: [PMC6520363](https://pubmed.ncbi.nlm.nih.gov/PMC6520363/)

17. Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX

Mikkel Schubert, Luca Ermini, Clio Der Sarkissian, Hákon Jónsson, Aurélien Ginolhac, Robert Schaefer, Michael D Martin, Ruth Fernández, Martin Kircher, Molly McCue, ... Ludovic Orlando

Nature Protocols (2014-04-10) <https://doi.org/f5x3qm>

DOI: [10.1038/nprot.2014.063](https://doi.org/10.1038/nprot.2014.063) · PMID: [24722405](https://pubmed.ncbi.nlm.nih.gov/24722405/)

18. Fast and accurate short read alignment with Burrows-Wheeler transform

H. Li, R. Durbin

Bioinformatics (2009-05-18) <https://doi.org/dqt59j>

DOI: [10.1093/bioinformatics/btp324](https://doi.org/10.1093/bioinformatics/btp324) · PMID: [19451168](https://pubmed.ncbi.nlm.nih.gov/19451168/) · PMCID: [PMC2705234](https://pubmed.ncbi.nlm.nih.gov/PMC2705234/)

19. mapDamage: testing for damage patterns in ancient DNA sequences

Aurelien Ginolhac, Morten Rasmussen, M. Thomas P. Gilbert, Eske Willerslev, Ludovic Orlando

Bioinformatics (2011-08-01) <https://doi.org/cn45v7>

DOI: [10.1093/bioinformatics/btr347](https://doi.org/10.1093/bioinformatics/btr347) · PMID: [21659319](https://pubmed.ncbi.nlm.nih.gov/21659319/)