

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

This manuscript ([permalink](#)) was automatically generated from [apeltzer/eager2-paper@0b87d41](#) on May 29, 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 Max Planck Society

- **Thiseas C. Lamnidis**

 [0000-0003-4485-8570](#) ·  [TCLamnidis](#) ·  [TCLamnidis](#)

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

- **Maxime Borry**

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

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

- **Aida Andrades Valtueña**

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

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

- **Zandra Fagernäs**

 [0000-0003-2667-3556](#) ·  [ZandraFagernas](#) ·  [ZandraSelina](#)

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

- **Stephen Clayton**

 [XXXX-XXXX-XXXX-XXXX](#) ·  [sc13-bioinf](#)

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

- **Maxime U. Garcia**

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

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

- **Judith Neukamm**

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

Palaeogenetics Group, Institute of Evolutionary Medicine, University of Zurich

- **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 broadening 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 genomic 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 [2](#) and [nf-core](#) [3](#). The pipeline aims to address three main points: adaptability to different computing configurations, reproducibility to ensure robust analytical standards in the field, and bring the pipeline up-to-date with the [latest routine and ancient genomic practises](#). This new version of EAGER has been developed within the [nf-core](#) initiative, to ensure high quality software development and maintenance support; contributing to a long-term lifecycle for the pipeline. [nf-core/eager](#) will assist in ensuring that [ancient DNA](#) sequencing data can be utilised by a diverse range of research groups and fields.

## Introduction

---

Ancient DNA has become a widely accepted source of biological data, helping to provide new perspective [for a range](#) of fields including archaeology, [ecology, cultural heritage](#), and palaeontology. The utilisation of next-generation-sequencing has allowed the recovery of ancient DNA from a wide variety of sources, including but not limited to, the skeletal remains of animals ([4](#), [5](#), [6](#), [\[??\]](#)), modern and archaic humans ([7](#), [\[??\]](#)), bacteria ([8](#), [9](#), [10](#)), Viruses ([11](#), [12](#)), plants ([13](#), [14](#)), but also coprolites ([15](#), [16](#)), dental calculus ([17](#), [18](#)), sediments ([19](#), [20](#)), medical slides ([21](#)) parchment ([22](#)), and most recently ancient 'chewing gum' ([23](#), [24](#)). Improvement in laboratory protocols to increase yields of otherwise trace amount of DNA has at the same time led to studies that can range from one to thousands of ancient individuals ([25](#), [\[??\]](#)), spanning single ([26](#)) to thousands of organisms ([17](#)). These differences of disciplines have led to a heterogeneous landscape in terms of types of analyses thus types of computing resources required by different labs - particularly in regions of the world where ancient DNA is still emerging.

Two previously published and commonly used pipelines in the field are PALEOMIX ([27](#)) 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](#) ([28](#)) for ultra-short reads and [mapDamage](#) ([29](#)) for aDNA characteristic evaluation). Yet, neither have had major updates to bring in line with current routinely carried out aDNA analyses. In particular, metagenomic screening of off-target genomic reads for pathogens or microbiomes ([17](#), [18](#)) has becoming particularly common, given it's role in revealing widespread infectious disease and possible epidemics that had previously been undetected in the archaeological record ([30](#), [31](#), [11](#), [12](#)). Without easy access to the latest field-established analytical routines, aDNA studies from groups new to the field risk being published without the necessary quality controls checks to ensure authenticity of their data and without yielding the full range of possibilities from the data.

To address these short comings, we have completely re-implemented the latest version of the EAGER pipeline in Nextflow, a domain-specific-language (DSL) designed for the construction of 'omic analysis pipelines ([2](#)). In addition, the newly named pipeline - [nf-core/eager](#) - has been developed in the context of the [nf-core](#) community framework ([3](#)). [nf-core](#) enforces strict guidelines for best-practises in software development to ensure robust, long-term maintenance and high quality pipelines.

## Results and Discussion

### Scalability, Portability, and Efficiency

---

The reimplementaion of EAGER into Nextflow offers a range of benefits over the original custom pipeline framework.

Firstly, the new framework provides immediate integration of nf-core/eager into various schedulers in POSIX High-Performance-Cluster (HPC) environments as well as local computers. This portability allows both small and big labs to run nf-core/eager regardless of the type of computer or cluster cluster sizes with minimal effort or configuration, facilitating reproducibility and therefore maintenance of standards within the field. This is further assisted by the in-built compatibility with software environments and contains such as conda, docker and singularity. This ensures exact versions of software are used by a user, regardless of the set-up of the cluster environment. Another major change with nf-core/eager is that the GUI input is now replaced with a command-line-interface as the primary user interaction mode. This is more compatible and portable with most HPCs (that may not offer X11 forwarding) and is in line with the vast majority of bioinformatic tools. We therefore believe this will not be a hindrance to new researchers from outside computational biology, however the nf-core initiative is in the processing in the near future of offering multiple alternatives including a command-line wizard, and a web-based input GUI.

Reproducibility is further made easier through the use of 'profiles' that can define configuration parameters for both the computing environment (schedulers, cache locations, maximum resource etc.), which can be centrally managed to ensure all users of a group use the same settings. Pipeline level profiles, specifying default parameters for nf-core/eager itself is therefore more portable between different groups through separation of the parameters from input data. Compared to the original EAGER that utilised per-FASTQ XML files with hardcoded filepaths, researchers can publish alongside their publications the specific profile used in their nf-core/eager runs to ensure other groups can generate the same results. Using of profiles also reduces mistakes caused by insufficient 'prose' based reporting of program settings, which often occurs in papers written by researchers unfamiliar with informatics. The default nf-core/eager profile uses parameters evaluated in different aDNA specific contexts (e.g. [32](#)), and will be updated in each new release as new studies are published.

nf-core/eager provides improved efficiency over the original EAGER pipeline by replacing the sample-by-sample sequential processing with Nextflow's asynchronous parallelisation. This combined with nf-core/eager-defined per-process customisation of resource parameters reduces unnecessary resource allocation that can occur with new users to each step of an NGS data processing pipeline. This is particularly pertinent given the increasing use of centralised HPCs or cloud computing, which often use per-hour cost calculations.

## Latest aDNA practices

---

nf-core/eager follows a similar structural foundation with the original EAGER. Given Illumina short-read FASTQ and/or BAM files, and a reference FASTA file.

This can be split into four main stages:

### ADD CITATIONS

1. Preprocessing
  1. Sequencing quality control (FastQC)
  2. Sequencing artefact clean-up (AdapterRemoval2)
  3. Preprocessing statistics generation
2. Mapping and post- processing
  1. Alignment against reference genome (BWA, CircularMapper)

2. Mapping quality filtering (Samtools)
3. PCR duplicate removal (DeDup, Samtools MarkDuplicates)
4. Mapping statistics generation (PreSeq, Qualimap)
3. aDNA Evaluation and Modification
  1. Damage profiling (DamageProfiler)
  2. aDNA reads selection (PMDTools)
  3. Damage removal (Bamutils)
  4. (Human) contamination estimation (ANGSD)
4. Genotyping and Consensus Sequencing (GATK, VCF2Genome)

In nf-core/eager, all tools also originally used in EAGER have been updated to latest compatible versions, as available on conda to ensure widespread accessibility and stability. **Modules that have not been carried over are Schmutzi and MapDamage because ALEX TO WRITE SOMETHING NICE HERE.** Support for Bowtie2 will be added in the near future, after consultation with the nf-core user community. New tools to the basic workflow include fastp ([33](#)) for the removal of poly-G sequencing artefacts that are common in 2-colour Illumina sequencing machines (such as the increasingly popular NextSeq and NovaSeq platforms). We have also included the FreeBayes genotyper as an alternative to the human-focused GATK tools. We have also maintained the possibility of using the now unsupported GATK UnifiedGenotyper, as the GATK HaplotypeCaller performs *de novo* assembly around possible variants, which may not be suitable for low-coverage aDNA data.

We have further extended functionality of the pipeline, primarily focusing on allowing for ancient metagenomic analysis to be run alongside genomic of the reference genome. We have added the ability to screen all off-target reads from mapping to the supplied reference FASTA with two metagenomic profilers: MALT ([34](#), [35](#)) and Kraken2 ([36](#)). Characterisation of properties of authentic ancient DNA ([37](#)) from MALT alignments is carried out with the HOPS pipeline ([38](#)). The pipeline also integrates a SNP alignment generation tool, MultiVCFAnalyzer (doi:10.1038/nature13591), which allows assessment of level of cross-mapping to a reference genome from different related taxa - a common challenge in ancient bacteria genome reconstruction ([39](#)). Simple coverage statistics of particular annotations (e.g. genes) of an input reference is offered by bedtools ([40](#)).

New tools supplied with the pipeline include a dedicated 'endogenous DNA' calculator (endorS.py, by A.A.V), to give approximate level of on-target DNA yield within a sample. Secondly, ancient metagenomic studies may include comparative samples from living day individuals ([41](#)). To support open data, whilst respecting data privacy, nf-core/eager includes a 'strip\_fastq' script (by M.B.) which creates raw FASTQ files but with reference-genome mapped reads removed. This then allows safe upload of sequencing data to public repositories with identifiable data removed. When using a human reference genome, nf-core/eager also can give estimates of the biological sex of a given individual with Sex.DetEERRmine ([42](#)).

Complex sequencing strategies is also now accounted for with nf-core/eager. Given the large amount of sequencing often required to yield sufficient genome coverage **from ancient DNA data**, palaeogeneticists can sometimes use multiple (differently treated) libraries or sequencing runs. As an alternative to direct paths, the pipeline can also accept a TSV file which includes file paths and additional metadata such as sample name, library name, sequencing lane, colour chemistry and UDG treatment. This allows simultaneous processing and appropriate merging of heterogeneous data of multiple sequencing runs and/or libraries types.

Finally, the original EAGER tabular report format has been replaced with MultiQC ([43](#)). The original EAGER pipeline required users to look through many independent output directories and files to make full assessment of their sequencing data. Aggregation of all log files into a single interactive report will assist users in making fuller assessment of their sequencing and analysis runs. Most tools within nf-core/eager have a corresponding MultiQC module to ensure as complete evaluation as possible.

## Accessibility

---

Alongside the portable new pipeline report, we have written extensive documentation on all parts of running and interpreting the output of the pipeline. Given that a large fraction of aDNA researchers come from fields such as social sciences that have limited computational training (such as J.A.F.Y), we have written documentation that also gives guidance on how to interpret each section of the report, specifically in the context of NGS and aDNA. This includes schematic images (by Z.F.) of best practices or expected output, published under CC-BY licenses to allow for use in other training material. This open-access resource will make the aDNA discipline more accessible to researchers new to the field, by providing practical and 'applied' knowledge to how aDNA characteristics translate to downstream analyses.

The development of nf-core/eager in nextflow and the nf-core initiative will also improve open-source community contributions to the pipeline. While nextflow is written primarily in the language groovy, the nextflow DSL simplifies a lot of concepts to a level intermediate level that bioinformaticians without Java/Groovy experience can easily access (regardless of own programming language experience). Furthermore, Nextflow places ubiquitous command-line interfaces, such as bash, in a prominent position within the code, rather than custom java code and classes. This is particularly relevant for researchers ancient DNA, who often come from social sciences and do not have backgrounds or training in computational biology. We hope this will motivate further bug fixes and feature contributions to keep the updated pipeline with standard practises during a longer life-cycle of this version of the pipeline. This will also be supported by the active and welcoming nf-core community who provide general guidance and advice on developing nextflow and nf-core pipelines.

## Conclusion

---

nf-core/eager is an efficient, portable, and accessible pipeline for processing ancient DNA genomic data. The re-implementation with nextflow and nf-core will improve reproducibility and inclusion of rapidly increasing ancient DNA datasets, for both large and small laboratories. Extensive documentation also allow newcomers to the field get a practical understanding on how to interpret ancient DNA in the context of NGS data processing. Finally nf-core/eager brings in the latest tools and routine screening analysis commonly used in the field, and sets up the pipeline staying at the forefront of palaeogenetic analysis.

## Methods

---

### Installation

nf-core/eager requires java, nextflow and either a environment/container system.

### Running

Further customisation can be made using profiles

Basic input is FASTQ or BAMs, FASTA reference. Complex and/or heterogeneous input can be supplied with a TSV file.

Runs can be monitored via terminal console or tower.nf.

Example data is included.

Output includes both files for close inspection and downstream analysis, as well as the emailable MultiQC report.

## Competing Interests

---

No competing interests are declared.

## Data and software availability

---

All code is available on github at <https://github.com/nf-core/eager> and archived with Zenodo under the DOI [10.5281/zenodo.1465061](https://doi.org/10.5281/zenodo.1465061). All test data is from the ENA public repository.

## Acknowledgements

---

We thank the nf-core community for general support and suggestions during the writing of the pipeline. We also thank Arielle Munters, Hester van Schalkwyk, Irina Velsko, Katerine Eaton, Luc Venturini, Marcel Keller, Pierre Lindenbaum, Pontus Skoglund, Raphael Eisenhofer, Torsten Günter, Kevin Lord for bug reports and feature suggestions. We are grateful to the members of the Department of Archaeogenetics at the Max Planck Institute for the Science of Human History who performed beta testing of the pipeline. We thank the aDNA twitter community for responding to polls regarding design decisions of the pipeline.

We also acknowledge Christina Warinner and the Max Planck Society who provided funds for travel to nf-core events.

## 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/ggqzpk>

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. Nextflow enables reproducible computational workflows**

Paolo Di Tommaso, Maria Chatzou, Evan W Floden, Pablo Prieto Barja, Emilio Palumbo, Cedric Notredame

*Nature Biotechnology* (2017-04-11) <https://doi.org/gfj52z>

DOI: [10.1038/nbt.3820](https://doi.org/10.1038/nbt.3820) · PMID: [28398311](https://pubmed.ncbi.nlm.nih.gov/28398311/)

**3. The nf-core framework for community-curated bioinformatics pipelines**

Philip A. Ewels, Alexander Peltzer, Sven Fillinger, Harshil Patel, Johannes Alneberg, Andreas Wilm, Maxime Ulysse Garcia, Paolo Di Tommaso, Sven Nahnsen

*Nature Biotechnology* (2020-02-13) <https://doi.org/ggk3qh>

DOI: [10.1038/s41587-020-0439-x](https://doi.org/10.1038/s41587-020-0439-x) · PMID: [32055031](https://pubmed.ncbi.nlm.nih.gov/32055031/)

**4. Complete Genomes Reveal Signatures of Demographic and Genetic Declines in the Woolly Mammoth**

Eleftheria Palkopoulou, Swapan Mallick, Pontus Skoglund, Jacob Enk, Nadin Rohland, Heng Li, Ayça Omrak, Sergey Vartanyan, Hendrik Poinar, Anders Götherström, ... Love Dalén

*Current Biology* (2015-05) <https://doi.org/34d>

DOI: [10.1016/j.cub.2015.04.007](https://doi.org/10.1016/j.cub.2015.04.007) · PMID: [25913407](https://pubmed.ncbi.nlm.nih.gov/25913407/) · PMCID: [PMC4439331](https://pubmed.ncbi.nlm.nih.gov/PMC4439331/)

**5. 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/)

**6. 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/)

**7. 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/)

**8. 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/)

**9. 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/)

**10. 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, Dorte 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/)

**11. 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/)

**12. 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/)

**13. Ancient reveals the timing and persistence of organellar genetic bottlenecks over 3,000 years of sunflower domestication and improvement**

Nathan Wales, Melis Akman, Ray H. B. Watson, Fátima Sánchez Barreiro, Bruce D. Smith, Kristen J. Gremillion, M. Thomas P. Gilbert, Benjamin K. Blackman  
*Evolutionary Applications* (2018-02-13) <https://doi.org/gf568v>  
DOI: [10.1111/eva.12594](https://doi.org/10.1111/eva.12594) · PMID: [30622634](https://pubmed.ncbi.nlm.nih.gov/30622634/) · PMCID: [PMC6304678](https://pubmed.ncbi.nlm.nih.gov/PMC6304678/)

**14. The origins and adaptation of European potatoes reconstructed from historical genomes**

Rafal M. Gutaker, Clemens L. Weiß, David Ellis, Noelle L. Anglin, Sandra Knapp, José Luis Fernández-Alonso, Salomé Prat, Hernán A. Burbano  
*Nature Ecology & Evolution* (2019-06-24) <https://doi.org/10.1038/s41559-019-0921-3>  
DOI: [10.1038/s41559-019-0921-3](https://doi.org/10.1038/s41559-019-0921-3) · PMID: [31235927](https://pubmed.ncbi.nlm.nih.gov/31235927/)

**15. 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/)

**16. CoproID predicts the source of coprolites and paleofeces using microbiome composition and host DNA content**

Maxime Borry, Bryan Cordova, Angela Perri, Marsha Wibowo, Tanvi Prasad Honap, Jada Ko, Jie Yu,



Kate Britton, Linus Girdland-Flink, Robert C. Power, ... Christina Warinner  
*PeerJ* (2020-04-17) <https://doi.org/dr8x>  
DOI: [10.7717/peerj.9001](https://doi.org/10.7717/peerj.9001) · PMID: [32337106](https://pubmed.ncbi.nlm.nih.gov/32337106/) · PMCID: [PMC7169968](https://pubmed.ncbi.nlm.nih.gov/PMC7169968/)

**17. 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 Fiddymment, ... 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/)

**18. Neanderthal behaviour, diet, and disease inferred from ancient DNA in dental calculus**

Laura S. Weyrich, Sebastian Duchene, Julien Soubrier, Luis Arriola, Bastien Llamas, James Breen, Alan G. Morris, Kurt W. Alt, David Caramelli, Veit Dresely, ... Alan Cooper  
*Nature* (2017-03-08) <https://doi.org/f9szrm>  
DOI: [10.1038/nature21674](https://doi.org/10.1038/nature21674) · PMID: [28273061](https://pubmed.ncbi.nlm.nih.gov/28273061/)

**19. Fifty thousand years of Arctic vegetation and megafaunal diet**

Eske Willerslev, John Davison, Mari Moora, Martin Zobel, Eric Coissac, Mary E. Edwards, Eline D. Lorenzen, Mette Vestergård, Galina Gussarova, James Haile, ... Pierre Taberlet  
*Nature* (2014-02-05) <https://doi.org/f2zr4s>  
DOI: [10.1038/nature12921](https://doi.org/10.1038/nature12921) · PMID: [24499916](https://pubmed.ncbi.nlm.nih.gov/24499916/)

**20. 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) · PMID: [28450384](https://pubmed.ncbi.nlm.nih.gov/28450384/)

**21. 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/)

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

M. D. Teasdale, N. L. van Doorn, S. Fiddymment, 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/)

**23. 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/)

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

Natalija Kashuba, Emrah Kırdök, 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/)

**25. The Beaker phenomenon and the genomic transformation of northwest Europe**

Iñigo Olalde, Selina Brace, Morten E. Allentoft, Ian Armit, Kristian Kristiansen, Thomas Booth, Nadin Rohland, Swapan Mallick, Anna Szécsényi-Nagy, Alissa Mittnik, ... David Reich

*Nature* (2018-02-21) <https://doi.org/gcx74m>

DOI: [10.1038/nature25738](https://doi.org/10.1038/nature25738) · PMID: [29466337](https://pubmed.ncbi.nlm.nih.gov/29466337/) · PMCID: [PMC5973796](https://pubmed.ncbi.nlm.nih.gov/PMC5973796/)

**26. A draft genome of *Yersinia pestis* from victims of the Black Death**

Kirsten I. Bos, Verena J. Schuenemann, G. Brian Golding, Hernán A. Burbano, Nicholas Waglechner, Brian K. Coombes, Joseph B. McPhee, Sharon N. DeWitte, Matthias Meyer, Sarah Schmedes, ... Johannes Krause

*Nature* (2011-10-12) <https://doi.org/fk87wk>

DOI: [10.1038/nature10549](https://doi.org/10.1038/nature10549) · PMID: [21993626](https://pubmed.ncbi.nlm.nih.gov/21993626/) · PMCID: [PMC3690193](https://pubmed.ncbi.nlm.nih.gov/PMC3690193/)

**27. 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/)

**28. 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/)

**29. mapDamage: testing for damage patterns in ancient DNA sequences**

Aurélien 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/)

**30. Early Divergent Strains of *Yersinia pestis* in Eurasia 5,000 Years Ago**

Simon Rasmussen, Morten Erik Allentoft, Kasper Nielsen, Ludovic Orlando, Martin Sikora, Karl-Göran Sjögren, Anders Gorm Pedersen, Mikkel Schubert, Alex Van Dam, Christian Mollin Outzen Kapel, ... Eske Willerslev

*Cell* (2015-10) <https://doi.org/f3mxqd>

DOI: [10.1016/j.cell.2015.10.009](https://doi.org/10.1016/j.cell.2015.10.009) · PMID: [26496604](https://pubmed.ncbi.nlm.nih.gov/26496604/) · PMCID: [PMC4644222](https://pubmed.ncbi.nlm.nih.gov/PMC4644222/)

**31. The Stone Age Plague and Its Persistence in Eurasia**

Aida Andrades Valtueña, Alissa Mittnik, Felix M. Key, Wolfgang Haak, Raili Allmäe, Andrej Belinskij, Mantas Daubaras, Michal Feldman, Rimantas Jankauskas, Ivor Janković, ... Johannes Krause

*Current Biology* (2017-12) <https://doi.org/cgmv>

DOI: [10.1016/j.cub.2017.10.025](https://doi.org/10.1016/j.cub.2017.10.025) · PMID: [29174893](https://pubmed.ncbi.nlm.nih.gov/29174893/)

**32. Improving ancient DNA read mapping against modern reference genomes**

Mikkel Schubert, Aurélien Ginolhac, Stinus Lindgreen, John F Thompson, Khaled AS AL-Rasheid, Eske Willerslev, Anders Krogh, Ludovic Orlando

*BMC Genomics* (2012) <https://doi.org/gb3ff7>

DOI: [10.1186/1471-2164-13-178](https://doi.org/10.1186/1471-2164-13-178) · PMID: [22574660](https://pubmed.ncbi.nlm.nih.gov/22574660/) · PMCID: [PMC3468387](https://pubmed.ncbi.nlm.nih.gov/PMC3468387/)

33. **fastp: an ultra-fast all-in-one FASTQ preprocessor**  
Shifu Chen, Yanqing Zhou, Yaru Chen, Jia Gu  
*Bioinformatics* (2018-09-01) <https://doi.org/gd9mrb>  
DOI: [10.1093/bioinformatics/bty560](https://doi.org/10.1093/bioinformatics/bty560) · PMID: [30423086](https://pubmed.ncbi.nlm.nih.gov/30423086/) · PMCID: [PMC6129281](https://pubmed.ncbi.nlm.nih.gov/PMC6129281/)
34. **MALT: Fast alignment and analysis of metagenomic DNA sequence data applied to the Tyrolean Iceman**  
Alexander Herbig, Frank Maixner, Kirsten I. Bos, Albert Zink, Johannes Krause, Daniel H. Huson  
*bioRxiv* (2016-04-27) <https://doi.org/ggxkk9>  
DOI: [10.1101/050559](https://doi.org/10.1101/050559)
35. **Salmonella enterica genomes from victims of a major sixteenth-century epidemic in Mexico**  
Åshild J. Vågene, Alexander Herbig, Michael G. Campana, Nelly M. Robles García, Christina Warinner, Susanna Sabin, Maria A. Spyrou, Aida Andrades Valtueña, Daniel Huson, Noreen Tuross, ... Johannes Krause  
*Nature Ecology & Evolution* (2018-01-15) <https://doi.org/ggxkk7>  
DOI: [10.1038/s41559-017-0446-6](https://doi.org/10.1038/s41559-017-0446-6) · PMID: [29335577](https://pubmed.ncbi.nlm.nih.gov/29335577/)
36. **Improved metagenomic analysis with Kraken 2**  
Derrick E. Wood, Jennifer Lu, Ben Langmead  
*Genome Biology* (2019-11-28) <https://doi.org/ggfk55>  
DOI: [10.1186/s13059-019-1891-0](https://doi.org/10.1186/s13059-019-1891-0) · PMID: [31779668](https://pubmed.ncbi.nlm.nih.gov/31779668/) · PMCID: [PMC6883579](https://pubmed.ncbi.nlm.nih.gov/PMC6883579/)
37. **Patterns of damage in genomic DNA sequences from a Neandertal**  
A. W. Briggs, U. Stenzel, P. L. F. Johnson, R. E. Green, J. Kelso, K. Prufer, M. Meyer, J. Krause, M. T. Ronan, M. Lachmann, S. Paabo  
*Proceedings of the National Academy of Sciences* (2007-08-21) <https://doi.org/bs4w7h>  
DOI: [10.1073/pnas.0704665104](https://doi.org/10.1073/pnas.0704665104) · PMID: [17715061](https://pubmed.ncbi.nlm.nih.gov/17715061/) · PMCID: [PMC1976210](https://pubmed.ncbi.nlm.nih.gov/PMC1976210/)
38. **HOPS: automated detection and authentication of pathogen DNA in archaeological remains**  
Ron Hübner, Felix M. Key, Christina Warinner, Kirsten I. Bos, Johannes Krause, Alexander Herbig  
*Genome Biology* (2019-12-16) <https://doi.org/ggxkmb>  
DOI: [10.1186/s13059-019-1903-0](https://doi.org/10.1186/s13059-019-1903-0) · PMID: [31842945](https://pubmed.ncbi.nlm.nih.gov/31842945/) · PMCID: [PMC6913047](https://pubmed.ncbi.nlm.nih.gov/PMC6913047/)
39. **A Robust Framework for Microbial Archaeology**  
Christina Warinner, Alexander Herbig, Allison Mann, James A. Fellows Yates, Clemens L. Weiß, Hernán A. Burbano, Ludovic Orlando, Johannes Krause  
*Annual Review of Genomics and Human Genetics* (2017-08-31) <https://doi.org/gf5wqv>  
DOI: [10.1146/annurev-genom-091416-035526](https://doi.org/10.1146/annurev-genom-091416-035526) · PMID: [28460196](https://pubmed.ncbi.nlm.nih.gov/28460196/) · PMCID: [PMC5581243](https://pubmed.ncbi.nlm.nih.gov/PMC5581243/)
40. **BEDTools: a flexible suite of utilities for comparing genomic features**  
Aaron R. Quinlan, Ira M. Hall  
*Bioinformatics* (2010-03-15) <https://doi.org/cmrms3>  
DOI: [10.1093/bioinformatics/btq033](https://doi.org/10.1093/bioinformatics/btq033) · PMID: [20110278](https://pubmed.ncbi.nlm.nih.gov/20110278/) · PMCID: [PMC2832824](https://pubmed.ncbi.nlm.nih.gov/PMC2832824/)
41. **Microbial differences between dental plaque and historic dental calculus are related to oral biofilm maturation stage**  
Irina M. Velsko, James A. Fellows Yates, Franziska Aron, Richard W. Hagan, Laurent A. F. Frantz, Louise Loe, Juan Bautista Rodriguez Martinez, Eros Chaves, Chris Gosden, Greger Larson, Christina Warinner  
*Microbiome* (2019-07-06) <https://doi.org/ggxkmc>  
DOI: [10.1186/s40168-019-0717-3](https://doi.org/10.1186/s40168-019-0717-3) · PMID: [31279340](https://pubmed.ncbi.nlm.nih.gov/31279340/) · PMCID: [PMC6612086](https://pubmed.ncbi.nlm.nih.gov/PMC6612086/)

42. **Ancient Fennoscandian genomes reveal origin and spread of Siberian ancestry in Europe**

Thiseas C. Lamnidis, Kerttu Majander, Choongwon Jeong, Elina Salmela, Anna Wessman, Vyacheslav Moiseyev, Valery Khartanovich, Oleg Balanovsky, Matthias Ongyerth, Antje Weihmann, ... Stephan Schiffels

*Nature Communications* (2018-11-27) <https://doi.org/ggxkk6>

DOI: [10.1038/s41467-018-07483-5](https://doi.org/10.1038/s41467-018-07483-5) · PMID: [30479341](https://pubmed.ncbi.nlm.nih.gov/30479341/) · PMCID: [PMC6258758](https://pubmed.ncbi.nlm.nih.gov/PMC6258758/)

43. **MultiQC: summarize analysis results for multiple tools and samples in a single report**

Philip Ewels, Måns Magnusson, Sverker Lundin, Max Käller

*Bioinformatics* (2016-10-01) <https://doi.org/f3s996>

DOI: [10.1093/bioinformatics/btw354](https://doi.org/10.1093/bioinformatics/btw354) · PMID: [27312411](https://pubmed.ncbi.nlm.nih.gov/27312411/) · PMCID: [PMC5039924](https://pubmed.ncbi.nlm.nih.gov/PMC5039924/)