# Historical *Arabidopsis thaliana genomes* from Germany

Sergio M. Latorre[1,2], Patricia L. M. Lang[2,3], Hernán A. Burbano[1,2,*]

[1]Centre for Life's Origins and Evolution, University College London, London, UK
[2]Max Planck Institute for Biology, Tuebingen, Germany
[3]Current address: Department of Biology, Stanford University, Stanford, CA 94305, USA
*Corresponding author: h.burbano@ucl.ac.uk

We report short-read Illumina sequencing of 35 *Arabidopsis thaliana* herbarium specimens collected in Southern Germany from 1817 to 1957. The historical genomes were sequenced to an average depth of ~7X. The herbarium-derived sequences showed patterns of degradation typical of ancient DNA with average read lengths of 78 bp, endogenous DNA content between 26%-93% (mean 72%; median 74%) and enrichment of cytosine to thymine substitutions at the read's termini. We provide the sequences to the public with open access.

## Introduction

The plant *Arabidopsis thaliana* occurs naturally across a wide range of environments in Eurasia and North Africa (Weigel and Mott 2009) and has successfully colonized other regions of the world such as North America (Platt et al. 2010; Exposito-Alonso et al. 2018; Shirsekar et al. 2021). Thus, this species harnesses genetic diversity that has permitted its adaptation to a wide range of biotic and abiotic selective pressures. To date, more than 1000 *A. thaliana* genomes have been sequenced (1001 Genomes Consortium. 2016), which can be analyzed in conjunction with extensive phenotypic, phenological and ecological data (e.g. (Exposito-Alonso et al. 2019)). These resources and the fact that *A. thaliana* is the most used plant in basic plant research make this species ideal to study its population and adaptive history. To give a temporal perspective to these type of studies, it is now possible to sequence genomes from historical herbarium specimens (Lang et al. 2019). The joint analysis of present-day and historical *A. thaliana* herbarium-derived genomes have already contributed to refining our understanding of the plant's colonization of North America (Exposito-Alonso et al. 2018) and to characterizing the genetic diversity of its African lineages (Durvasula et al. 2017). Here, we present 35 *A. thaliana* genomes derived from herbarium specimens collected between 1817 and 1957. We deposit these genomes in public databases for open access.

## Results

We sequenced 35 *A. thaliana* genomes from herbarium specimens collected in Southern Germany between 1817-1957 (Figure 1A-B and Table S1). The percentage of endogenous *A. thaliana*-derived reads varied between 26%-93% (mean 72%; median 74%) (Table S1). As expected from herbarium-derived ancient DNA (Weiß et al. 2016), the DNA was highly degraded with an average read length per library of 78 bp (Figure 2C). All samples showed patterns of nucleotide misincorporation typical of aDNA with an excess of cytosine to thymine substitutions at the reads' termini (Weiß et al. 2016) (Fig 2D). All sequencing reads were deposited in the GenBank under the BioProject number PRJNA887392.

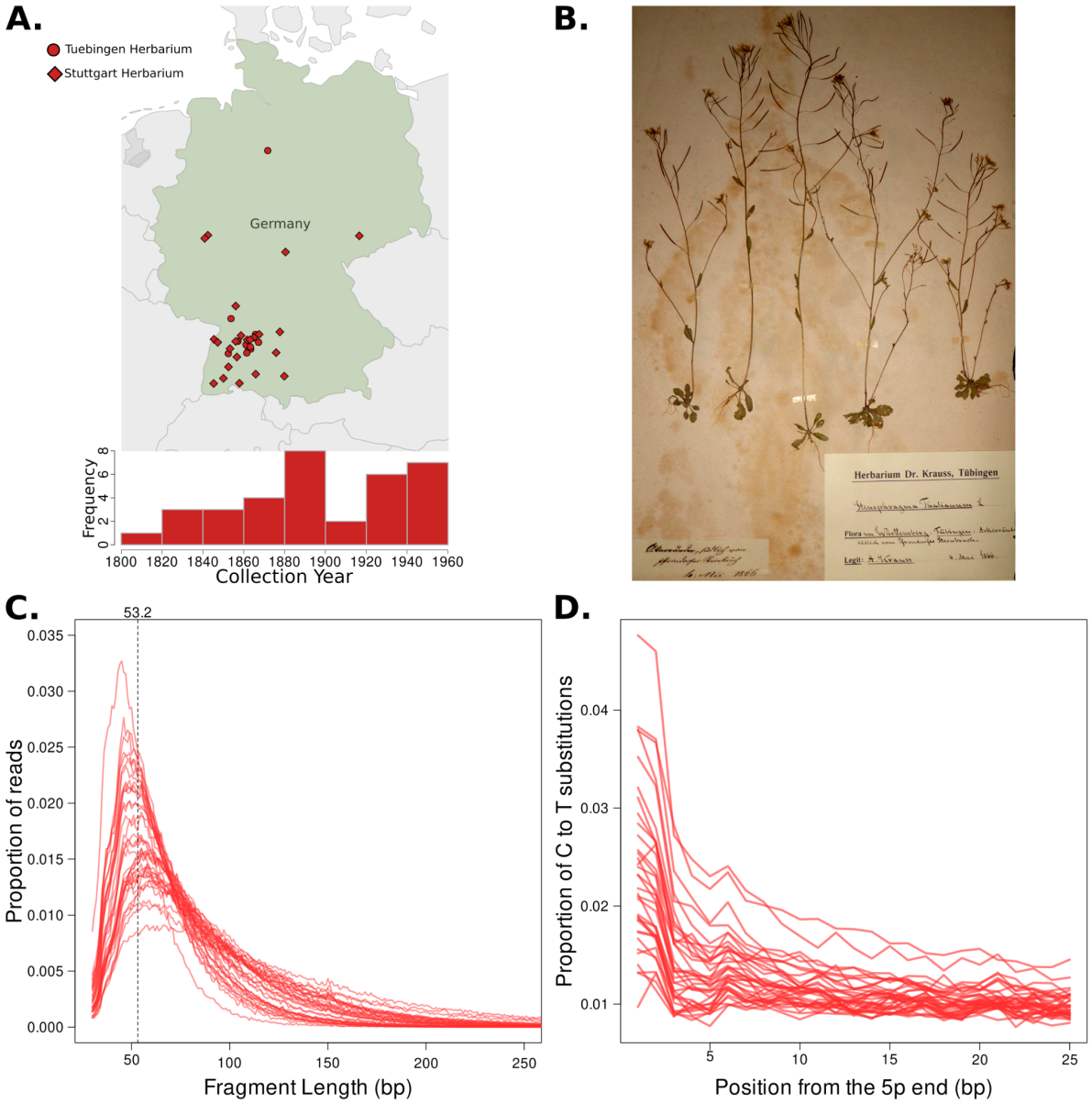


**Figure 1. Provenance and characteristics of *Arabidopsis thaliana* herbarium samples. A.** The map shows the approximate geographic origin of the herbarium specimens and the histogram represents the temporal distribution of the collection dates. **B.** Herbarium specimen HB0729 sampled at the Herbarium Tubingense at the University of Tuebingen, Germany. **C.** Length distributions of merged reads. **D.** Cytosine to Thymine (C-to-T) substitutions at the 5' end of sequenced DNA reads.

## Conclusions

We generated a set of 35 *A. thaliana* herbarium genomes, which are deposited in the GenBank to be used by the community (Table S1).

## Material and Methods

Sample collection: Herbarium specimens of *A. thaliana* were sampled at the herbaria of Stuttgart (Staatliches Museum für Naturkunde Stuttgart - STU) and Tübingen (Herbarium Tubingense - TUB), Germany (Table S1). We conducted sampling as minimally destructive as possible, collecting a maximum of ~1 $cm^2$ of leaf tissue as previously described (Lang et al. 2020).

DNA extraction and library preparation: DNA from herbarium specimens was extracted and sequencing libraries were prepared in a clean room facility at the University of Tübingen, Tübingen, Germany. The DNA was extracted following Basic Protocol 1, and libraries were prepared following Basic Protocol 2, as previously described (Latorre et al. 2020).

Bioinformatic processing and mapping of reads to the *A. thaliana* genome: We used a previously described pipeline for basic screening of aDNA-derived libraries (Latorre et al. 2020). Briefly, the pipeline assesses the quality of the sequencing run, the magnitude of DNA degradation, and the percentage of endogenous DNA of each sample, and authenticates the historical nature of the libraries by quantifying aDNA-associated misincorporation patterns.

Data availability: Illumina raw sequences have been deposited in the GenBank under the BioProject number PRJNA887392.

## Acknowledgements

We are grateful to Cornelia Krause and Anette Rosenbauer from the herbaria in Tübingen and Stuttgart, respectively, for their introduction to and help in the herbaria, and the kind permission to sample specimens; Verena Schuenemann and Ella Reiter for access to clean-room facilities and technical support; Detlef Weigel for supporting the Research Group for Ancient Genomics and Evolution, and the Max Planck Society for funding.

## Citations

1001 Genomes Consortium. Electronic address: magnus.nordborg@gmi.oeaw.ac.at, and 1001 Genomes Consortium. 2016. “1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis Thaliana.” *Cell* 166 (2): 481–91. https://doi.org/10.1016/j.cell.2016.05.063.

Durvasula, Arun, Andrea Fulgione, Rafal M. Gutaker, Selen Irez Alacakaptan, Pádraic J. Flood, Célia Neto, Takashi Tsuchimatsu, et al. 2017. “African Genomes Illuminate the Early History and Transition to Selfing in Arabidopsis Thaliana.” *Proceedings of the National Academy of Sciences of the United States of America*, May. https://doi.org/10.1073/pnas.1616736114.

Exposito-Alonso, Moises, 500 Genomes Field Experiment Team, Hernán A. Burbano, Oliver Bossdorf, Rasmus Nielsen, and Detlef Weigel. 2019. “Natural Selection on the Arabidopsis Thaliana Genome in Present and Future Climates.” *Nature* 573 (7772): 126–29. https://doi.org/10.1038/s41586-019-1520-9.

Exposito-Alonso, Moises, Claude Becker, Verena J. Schuenemann, Ella Reiter, Claudia Setzer, Radka Slovak, Benjamin Brachi, et al. 2018. “The Rate and Potential Relevance of New Mutations in a Colonizing Plant Lineage.” *PLoS Genetics* 14 (2): e1007155. https://doi.org/10.1371/journal.pgen.1007155.

Lang, Patricia L. M., Clemens L. Weiß, Sonja Kersten, Sergio M. Latorre, Sarah Nagel, Birgit Nickel, Matthias Meyer, and Hernán A. Burbano. 2020. “Hybridization ddRAD-Sequencing for Population Genomics of Non-Model Plants Using Highly Degraded Historical Specimen DNA.” *Molecular Ecology Resources*, April. https://doi.org/10.1111/1755-0998.13168.

Lang, Patricia L. M., Franziska M. Willems, J. F. Scheepens, Hernán A. Burbano, and Oliver Bossdorf. 2019. “Using Herbaria to Study Global Environmental Change.” *The New Phytologist* 221 (1): 110–22. https://doi.org/10.1111/nph.15401.

Latorre, Sergio M., Patricia L. M. Lang, Hernán A. Burbano, and Rafal M. Gutaker. 2020. “Isolation, Library Preparation, and Bioinformatic Analysis of Historical and Ancient Plant DNA.” *Current Protocols in Plant Biology* 5 (4): e20121. https://doi.org/10.1002/cppb.20121.

Platt, Alexander, Matthew Horton, Yu S. Huang, Yan Li, Alison E. Anastasio, Ni Wayan Mulyati, Jon Agren, et al. 2010. “The Scale of Population Structure in Arabidopsis Thaliana.” *PLoS Genetics* 6 (2): e1000843. https://doi.org/10.1371/journal.pgen.1000843.

Shirsekar, Gautam, Jane Devos, Sergio M. Latorre, Andreas Blaha, Maique Queiroz Dias, Alba González Hernando, Derek S. Lundberg, Hernán A. Burbano, Charles B. Fenster, and Detlef Weigel. 2021. “Multiple Sources of Introduction of North American Arabidopsis Thaliana from across Eurasia.” *Molecular Biology and Evolution* 38 (12): 5328–44. https://doi.org/10.1093/molbev/msab268.

Weigel, Detlef, and Richard Mott. 2009. “The 1001 Genomes Project for Arabidopsis Thaliana.” *Genome Biology* 10 (5): 107. https://doi.org/10.1186/gb-2009-10-5-107.

Weiß, Clemens L., Verena J. Schuenemann, Jane Devos, Gautam Shirsekar, Ella Reiter, Billie A. Gould, John R. Stinchcombe, Johannes Krause, and Hernán A. Burbano. 2016. “Temporal Patterns of Damage and Decay Kinetics of DNA Retrieved from Plant Herbarium Specimens.” *Royal Society Open Science* 3 (6): 160239. https://doi.org/10.1098/rsos.160239.