

Karel Břinda

INRIA: French National Institute for DigitSci & Tech

Inria/Irisa, GenScale
Campus de Beaulieu
35042 Rennes, France
karel.brinda@inria.fr
www.brinda.eu

Updated on April 22, 2024

Research Interests

Bioinformatics / Computational Biology: Algorithms, Genomics, Rapid Diagnostics, Pathogens, Antibiotic Resistance

Academic Appointments

- 2022–present **Inria – French National Institute for Research in Digital Science and Technology**
Inria Starting Faculty (a permanent PI position), GenScale team, Inria Rennes, France
- 2022–present **Irisa – Research Institute for Computer Science and Random Systems (UMR 6074)**
Associate Member, Irisa Rennes, France
- 2017–2021 (5y) **Harvard Medical School & Harvard TH Chan School of Public Health**
 - 2019–2021 *Research Associate*, Department of Biomedical Informatics
 - 2018–2019 *Research Associate*, Department of Epidemiology, Center for Communicable Disease Dynamics
 - 2017–2018 *Postdoctoral Research Fellow*, Department of Epidemiology, Center for Communicable Disease Dynamics.
Advisors: Michael Baym & William P. Hanage

Education

- 2013–2016 **PhD**, Computer Science, LIGM University of Paris-Est (Paris, France).
Advisors: Gregory Kucherov (CNRS) and Valentina Boeva (Curie Institute & Cochin Institute)
Dissertation: *Novel computational techniques for mapping and classification of Next-Generation Sequencing data*
- 2011–2013 **MSc (Ing)**, Mathematical Computer Science (with honors), Czech Technical University in Prague, Faculty of Nuclear Sciences and Physical Engineering (Prague, Czech Republic).
- 2008–2011 **BSc**, Mathematical Computer Science, Czech Technical University in Prague, Faculty of Nuclear Sciences and Physical Engineering (Prague, Czech Republic).

Funding and Awards

- 2023 **Défis scientifiques 2023 (PI)**, University of Rennes' funding for interdisciplinary research initiatives, 2023 (1y).
- 2022 **Inria Exploratory Action (PI)**, Inria funding for high-risk high-reward exploratory research, 2023–2026 (3y).
- 2022 **Campus France PHC Procope 2023 (PI)**, Mobility funding for collaborative French-German research initiatives, 2023–2024 (2y).
- 2019 **Best poster**, Marvin Zelen Data Science Symposium, Dana-Farber Cancer Institute, Boston, USA.
- 2018 **Most innovative poster**, Lake Arrowhead Microbial Genomics Conference (LAMG), Lake Arrowhead, USA.
- 2015 **Travel fellowship of the MSTIC doctoral school**, University of Paris-Est, France.
- 2012 **Fellowship of the RuFiDim organization committee**.
- 2012 **Travel fellowship of the Hlávka Foundation**.

Publications

Complete list: https://scholar.google.com/citations?user=TTaAd_MAAAAJ >1400 citations, h-index 13

Five selected publications are marked by *

Preprints

- [1] O. Sladký, P. Veselý, and K. Břinda, “Function-assigned masked superstrings as a versatile and compact data type for k-mer sets,” *bioRxiv* 2024.03.06.583483, 2024. doi:10.1101/2024.03.06.583483.

- [2] * **K. Břinda**, L. Lima, S. Pignotti, N. Quinones-Olvera, K. Salikhov, R. Chikhi, G. Kucherov, Z. Iqbal, and M. Baym, “Efficient and robust search of microbial genomes via phylogenetic compression,” *bioRxiv* 2023.04.15.536996, 2023. doi:10.1101/2023.04.15.536996.
- [3] * O. Sladký, P. Veselý, and **K. Břinda**, “Masked superstrings as a unified framework for textual k -mer set representations,” *bioRxiv* 2023.02.01.526717, *RECOMB-Seq 2023*, 2023. doi:10.1101/2023.02.01.526717.

Journal Articles

- [4] N. B. Wikle, T. N.-A. Tran, B. Gentilese, S. Leighow, J. Albert, E. R. Strong, **K. Břinda**, H. Inam, F. Yang, S. Hossain, P. Chan, W. P. Hanage, M. Messick, J. Pritchard, E. Hanks, and M. F. Boni, “SARS-CoV-2 epidemic after social and economic reopening in three U.S. states reveals shifts in age structure and clinical characteristics,” *Science Advances* **8**(4), 2022. doi:10.1126/sciadv.abf9868.
- [5] Y. Che, X. Xu, Y. Yang, **K. Břinda**, W. P. Hanage, C. Yang, and T. Zhang, “High-resolution genomic surveillance elucidates a multilayered hierarchical transfer of resistance between WWTP-and human/animal-associated bacteria,” *Microbiome* **10**(1), 2022. doi:10.1186/s40168-021-01192-w.
- [6] * **K. Břinda**, M. Baym, and G. Kucherov, “Simplitigs as an efficient and scalable representation of de Bruijn graphs,” *Genome Biology* **22**(96), 2021. doi:10.1186/s13059-021-02297-z.
- [7] Y. Che, Y. Yang, X. Xu, **K. Břinda**, M. F. Polz, W. P. Hanage, and T. Zhang, “Conjugative plasmids interact with insertion sequences to shape the horizontal transfer of antimicrobial resistance genes,” *Proceedings of the National Academy of Sciences* **118**(6), 2021. doi:10.1073/pnas.2008731118.
- [8] T. N.-A. Tran, N. Wikle, E. Albert, H. Inam, E. R. Strong, **K. Břinda**, S. M. Leighow, F. Yang, S. Hossain, J. R. Pritchard, P. Chan, W. P. Hanage, E. M. Hanks, and M. F. Boni, “Optimal SARS-CoV-2 vaccine allocation using real-time attack-rate estimates in Rhode Island and Massachusetts,” *BMC Medicine* **19**(162), 2021. doi:10.1186/s12916-021-02038-w.
- [9] * **K. Břinda**, A. Callendrello, K. C. Ma, D. R. MacFadden, T. Charalampous, R. S. Lee, L. Cowley, C. B. Wadsworth, Y. H. Grad, G. Kucherov, J. O’Grady, M. Baym, and W. P. Hanage, “Rapid inference of antibiotic resistance and susceptibility by Genomic Neighbour Typing,” *Nature Microbiology* **5**, pp. 455–464, 2020. doi:10.1038/s41564-019-0656-6.
- [10] D. R. MacFadden, B. Coburn, **K. Břinda**, A. Corbeil, N. Daneman, D. Fisman, R. Lee, M. Lipsitch, A. McGeer, R. Melano, S. Mubareka, and W. P. Hanage, “Using genetic distance from archived samples for the prediction of antibiotic resistance in *Escherichia coli*,” *Antimicrobial Agents and Chemotherapy* **64**(5), pp. 455–464, 2020. doi:10.1128/aac.02417-19.
- [11] B. Grüning, R. Dale, A. Sjödin, B. A. Chapman, J. Rowe, C. H. Tomkins-Tinch, R. Valieris, J. Köster, and The Bioconda Team (incl. **K. Břinda**), “Bioconda: sustainable and comprehensive software distribution for the life sciences,” *Nature Methods* **15**(7), pp. 475–476, 2018. doi:10.1038/s41592-018-0046-7.
- [12] **K. Břinda**, V. Boeva, and G. Kucherov, “RNF: a general framework to evaluate NGS read mappers,” *Bioinformatics* **32**(1), pp. 136–139, 2016. doi:10.1093/bioinformatics/btv524.
- [13] * **K. Břinda**, M. Sykulski, and G. Kucherov, “Spaced seeds improve k -mer-based metagenomic classification,” *Bioinformatics* **31**(22), pp. 3584–3592, 2015. doi:10.1093/bioinformatics/btv419.
- [14] **K. Břinda**, E. Pelantová, and O. Turek, “Balances of m -bonacci words,” *Fundamenta Informaticae* **132**(1), pp. 33–61, 2014. doi:10.3233/FI-2014-1031.
- [15] L. Balková, **K. Břinda**, and O. Turek, “Abelian complexity of infinite words associated with quadratic parry numbers,” *Theoretical Computer Science* **412**(45), pp. 6252–6260, 2011. doi:10.1016/j.tcs.2011.08.016.

Peer-Reviewed Conference Proceedings

- [16] P. Červenka, **K. Břinda**, M. Hanousková, P. Hofman, and R. Seifert, “Blind friendly maps: tactile maps for the blind as a part of the public map portal (mapy.cz),” in *Computers Helping People with Special Needs, ICCHP 2016, 2016, Proceedings, Part II*, pp. 131–138, 2016, doi:10.1007/978-3-319-41267-2_18.
- [17] **K. Břinda**, “Languages of lossless seeds,” in *Proceedings 14th International Conference on Automata and Formal Languages (AFL 2014), Electronic Proceedings in Theoretical Computer Science*, **151**, pp. 139–150, 2014, doi:10.4204/EPTCS.151.9.

Technical Reports

- [18] **K. Břinda**, V. Boeva, and G. Kucherov, “Ococo: an online variant and consensus caller,” *arXiv* 1712.01146 [q-bio.GN], 2018. doi:10.48550/arXiv.1712.01146.
- [19] **K. Břinda**, V. Boeva, and G. Kucherov, “Dynamic read mapping and online consensus calling for better variant detection,” *arXiv* 1605.09070 [q-bio.GN], 2016. doi:10.48550/arXiv.1605.09070.

Patents

- [20] W. P. Hanage, **K. Břinda**, and M. Baym, “Rapid identification of strains from sequence data,” *United States Patent Application* 17/251,343, US 2021/0246502 A1, 2021.

Theses

- [21] **K. Břinda**, “Novel computational techniques for mapping and classification of Next-Generation Sequencing data.” PhD thesis, *University of Paris-Est*, 2016.
- [22] **K. Břinda**, “Lossless seeds for approximate string matching.” MSc thesis, *Czech Technical University*, 2013.
- [23] **K. Břinda**, “Abelian complexity of infinite words.” BSc thesis, *Czech Technical University*, 2011.

Selected Software

Complete list: <http://github.com/karel-brinda>

>3.4M downloads*

- [SW1] **MiniPhy** – <http://github.com/karel-brinda/miniphy> — Python and Snakemake.
Phylogenetic compression of extremely large genome collections
- [SW2] **Phylign** – <http://github.com/karel-brinda/phylign> — Python and Snakemake.
BLAST-like alignment across all pre-2019 bacterial genomes on personal computers
- [SW3] **RASE** – <http://github.com/c2-d2/rase-pipeline> — Python, Snakemake, Make, R, and Bash.
A proof-of-concept framework for Genomic Neighbor Typing for real-time predictions of antibiotic resistance using nanopore sequencing. Pipeline, tool, library, two species databases (*S. pneumoniae* and *N. gonorrhoeae*), skeleton for novel databases, and demonstrations of within-minutes diagnostic from isolates and metagenomes.
- [SW4] **ProPhyle** – <http://prophyle.github.io> — Python, C, and C++.
An accurate, resource-frugal, and deterministic metagenomic classifier, based on k -mer propagation, simplitigs, and k -mer indexing using the Burrows-Wheeler Transform. [2.1M downloads]
- [SW5] **SAMsift** – <http://github.com/karel-brinda/samsift> — Python.
Advanced filtering and tagging of SAM/BAM alignments using Python expressions. [445k downloads]
- [SW6] **RNFtools** – <http://rnftools.github.io> — Python.
A generic format for naming simulated sequencing reads using arbitrary tools and the associated toolkit and pipeline for read simulation and read mapper evaluation. [456k downloads]
- [SW7] **Ococo** – <http://github.com/karel-brinda/ococo> — C++.
Online variant and consensus caller, based on streaming and maintaining variant statistics in small counters per individual genomic positions, which enables real-time analyzes of unsorted SAM/BAM data. [122k downloads]
- [SW8] **ProphAsm** – <http://github.com/prophyle/prophasm> — C++.
A tool for a rapid and memory-efficient computation of simplitigs and set operations with k -mer sets. [37k downloads]
- [SW9] **ProPhex** – <http://github.com/prophyle/prophex> — C.
An efficient k -mer index based on the Burrows-Wheeler Transform and a rolling window. Co-developed with Kamil Salikhov. [47k downloads]

* Cumulative sum of the statistics from Bioconda and Biocontainers

Teaching and Mentoring

Courses

École normale supérieure de Rennes (ENS Rennes)

- 2023–2024 **BOX Experimental Bioinformatics** – Lecturer — 1 semester, MSc level
 ◦ An introductory course of Bioinformatics for ENS masters’ students.

Harvard Medical School – Department of Biomedical Informatics

- 2019–2020 **BMIF01 Concepts in Genome Analysis** – Teaching Assistant — 1 semester, PhD level
 ◦ An introductory course of Computational Biology for Harvard PhD students.

- Recitations, grading, mentoring the students, supervision of the final projects.
- A close collaboration with the four faculties responsible for the course: Michael Baym, Heng Li, Cheng-Zhong Zhang, and Shamil Sunyaev.

Czech Technical University in Prague – Faculty of Nuclear Sciences and Physical Engineering

- 2009–2011 **01SWP Software Project 1,2** – Supervisor — 4 semesters, master's level
 2010–2013 **01ASTE Assistive Technology** – Instructor, co-developed the course — 3 semesters, master's level

Postdoctoral Fellows and Research Engineers

- 2023–present **Loren Dejoies**, *Inria*, Postdoctoral Researcher
 2023 **Gaëtan Benoit**, *Inria*, Software Engineer

PhD Students

- 2023–present **Francesca Brunetti**, *Sapienza University of Rome*, visiting PhD student
 2023–2024 **Arya Kaul**, *Harvard Medical School*, visiting PhD student (Chateaubriand fellow)

Master's Students

- 2023–present **Veronika Hendrychová**, *Czech Technical University in Prague*, visiting MSc student (Erasmus)
 2022–2023 **Léo Laffeach**, *École normale supérieure de Rennes*, MSc research project, co-supervised with P. Peterlongo and G. Gourdel
 2017–2018 **Simone Pignotti**, *University of Paris-Est*, MSc research project at *Harvard Medical School*, co-supervised with M. Baym — Now **Bioinformatics Team Leader @Eligo Bioscience**
 2015–2016 **Mikhail Dubov**, *University of Paris-Est*, MSc research project, co-supervised with G. Kucherov — Now **Head of Data @Revolut**

Bachelor Students

- 2022–present **Ondřej Sladký**, *Charles University in Prague*, BSc research project, co-supervised with P. Veselý
 2022–2023 **Veronika Hendrychová**, *Czech Technical University in Prague*, BSc thesis, co-supervised with L. Dvořáková. — **2nd-place award at Rektorys Competition for best theses in mathematics**
 2022–2023 **Ekaterina Milyutina**, *Charles University in Prague*, BSc thesis, co-supervised with P. Veselý

Related Work Experience

- 2005–2012 **Software Engineer / Assistive Technology Specialist**, Czech Technical University in Prague.
Major achievement: Co-developed Hapticke.Mapy.cz, automatically generated tactile maps for blind users. Time to create a tactile map for any part of the world *reduced from several weeks to 5 mins*. **Industrial collaboration with Seznam.cz, the biggest Czech Internet company.**

Selected Media Coverage

- 2020 **BBC World Service – Science in Action** — [Link](#) .
 An interview about our paper on rapid diagnostics of antibiotic resistance by Genomic Neighbor Typing.
 2020 **The Bioinformatics Chat – Simplitigs** — [Link](#) .
 An interview about our paper on simplitigs.
 2020 **Massive Science – Rapid DNA sequencing of unknown bacteria helps doctors choose which antibiotics to treat it with** — [Link](#) .
 A popularization article on how our method of rapid diagnostics could help doctors prescribe antibiotics better.
 2020 **The Emory EMERGENT blog – 7 significant bacterial genomics papers from the 2010s** — [Link](#) .
 Our paper on rapid diagnostics of antibiotic resistance was listed among selected top bacterial papers of the decade.

Talks (last 10 year)

Invited Talks

- 2023 **Charles University in Prague**, Bioinformatics seminar. *Efficient search of microbial genomes via phylogenetic compression*. Prague, Czech Republic.

- 2023 **Sorbonne University**, Graphs and Bioinformatics 2023. *K-mers and their graphs in computational biology*. Paris, France.
- 2023 **Alpaca/Pangaia**, joint seminar of the two international pan-genomic consortia. *Efficient and robust search of microbial genomes via phylogenetic compression*. Online.
- 2023 **Institute of Medical Microbiology, University Hospital of Düsseldorf**, Dilthey group seminar. *Rapid inference of antibiotic resistance and susceptibility by Genomic Neighbor Typing*. Düsseldorf, Germany.
- 2023 **École normale supérieure**, CPM 2023 Summer School. *Under the hood: The role of string algorithms in contemporary biological research*. Paris, France.
- 2022 **Irset University of Rennes / EHESP French School of Public Health**, Irset Seminar. *Rapid inference of antibiotic resistance and susceptibility by Genomic Neighbor Typing*. Rennes, France.
- 2022 **CRIStal Lille**, Bioinformatics Seminar. *Efficient and robust search of microbial genomes via phylogenetic compression*. Lille, France.
- 2022 **JOBIM 2022**, The indexing minisymposium. *The tree of life enables efficient and robust compression and search of microbes*. Rennes, France.
- 2021 **ICCMG 5**, Fifth International Conference on Clinical Metagenomics. *Rapid inference of antibiotic resistance and susceptibility by Genomic Neighbor Typing*. Geneva, Switzerland (remote).
- 2021 **LIRMM Montpellier**, Methods and Algorithms for Bioinformatics Seminar. *Rapid inference of antibiotic resistance and susceptibility by Genomic Neighbor Typing*. Montpellier, France (remote).
- 2020 **BAARN 2020**, Boston Area Antimicrobial Research Network Symposium. *Rapid inference of antibiotic resistance and susceptibility by Genomic Neighbor Typing*. Boston, USA (remote).
- 2019 **Harvard MSI**, Microbiology Chalk Talk Seminar. *Rapid inference of antibiotic resistance and susceptibility by Genomic Neighbor Typing*. Boston, USA.
- 2019 **MIT CSAIL/MIT Department of Mathematics**, Bioinformatics Seminar. *Lineage calling can identify antibiotic resistant clones within minutes*. Boston, USA.
- 2016 **Inria/Irisa Rennes, France**, Bioinformatics Seminar. *BWT-based indexing structure for metagenomic classification*. Rennes, France.
- 2015 **Institut Curie**, Systems Biology Seminar. *Spaced seeds improve metagenomic classification*. Paris, France.

Other Talks

- 2023 **Pan-genomics - metagenomics workshop (PHC Procope)**. *Towards comprehensive sequence data search engines*. Heinrich Heine University Düsseldorf, Germany
- 2022 **Data Structures in Bioinformatics 2022**. *Efficient and robust search of microbial genomes via phylogenetic compression*. Düsseldorf, Germany
- 2021 **Data Structures in Bioinformatics 2021**. *All microbes on a flash drive*. Milano, Italy (remote).
- 2021 **Human Technopole**. *Genome search to combat antibiotic resistance*. Milano, Italy (remote).
- 2021 **Helmholtz Centre for Infection Research**. *Algorithms to combat antibiotic resistance*. Braunschweig, Germany (remote).
- 2020 **Human and Biotechnology at TU Graz**. *Algorithms to combat antibiotic resistance*. Graz, Austria (remote).
- 2018 **Jonathan Freeman Symposium on the Epidemiology of Infectious Disease 2020**. *Lineage calling can identify antibiotic resistant clones within minutes*. Boston, USA.
- 2017 **Permafrost workshop on bacterial evolution 2017**. *Efficient phylogeny-based metagenomic classification*. Bormio, Italy.
- 2016 **SeqBio 2016**. *ProPhyle: a memory efficient BWT-based metagenomic classifier using k-mer propagation*. Nantes, France.
- 2016 **Quantitative Genomics 2016**. *BWT-based indexing structure for metagenomic classification*. University College London, UK.
- 2016 **Data Structures in Bioinformatics 2016**. *Indexing structures for metagenomic classification*. Bielefeld, Germany.
- 2015 **Algorithmics, Bioinformatics and Statistics for NGS data analysis**. *RNF: a general framework to evaluate NGS read mappers*. Paris, France.

- 2015 **SeqBio 2015**. *Dynamic mappers of NGS reads*. Montpellier, France.
 2014 **Automata and Formal Languages**. *Languages of lossless seeds*. Szeged, Hungary.

Service and Organization

Program and Organizing Committees

- 2023 SeqBIM 2023 Workshop Program Committee
 2023 ISMB/ECCB 2023 Conference Program Committee
 2023 Inria Rennes Intercultural Workshop Organizing Committee
 2022 SeqBIM 2022 Workshop Program Committee
 2022 JOBIM 2022 Minisymposium Organizing Committee
 2021 ISMB/ECCB 2021 Conference Program Committee

PhD Committees

- 2022–present Khodor Hannoush, doctoral advisory committee, University of Rennes, France
 2022 Yoshihiro Shibuya, dissertation committee, Gustave Eiffel University, France

Referee for Journal Articles

- 2014–present Nature Biotechnology
 Nature Communications
 Genome Research
 Genome Biology
 Bioinformatics
 ASM mSystems
 Microbial Genomics
 NAR Genomics and Bioinformatics
 PeerJ
 Journal of Bioinformatics and Computational Biology

Referee for Conference Proceedings

- 2014–present RECOMB, Research in Computational Molecular Biology
 ISMB, Intelligent Systems for Molecular Biology
 ECCB, European Conference on Computational Biology
 HITSEQ, High Throughput Sequencing
 WABI, Workshop on Algorithms in Bioinformatics
 SPIRE, String Processing and Information Retrieval
 ICCHP, International Conference on Computers Helping People with Special Needs
 WALCOM, International Conference and Workshop on Algorithms and Computation

Others

- 2014–2016 Czech Students and Young Professionals Organization in France.
 Organized events for incoming scholars, in collaboration with the *Embassy of the Czech Republic in Paris*.
 2010–2013 Academic Senate of the Faculty of Nuclear Sciences and Physical Engineering.
 Elected student representative.
 Elected vice-chair of the senate.
 2010 Ministry of the Interior of the Czech Republic.
 Served as a member of the Working Group for Accessible Documents.