Tommi Mäklin

+358449800114

Education

University of Helsinki, Helsinki, Finland

PhD, Computer Science, 2022
Probabilistic methods for high-resolution metagenomics

University of Helsinki, Helsinki, Finland

MSc, Statistical machine learning, 2017
Major studies in Statistics

University of Helsinki, Helsinki, Finland

BSc, Statistics, 2016
Minor studies in Mathematics and Computer Science

Research experience

University of Helsinki, Helsinki, Finland
Postdoctoral researcher, 2023—ongoing

I work on developing and applying statistical methods for studying the link between ecology and epidemiology of opportunistic pathogens that are also a part of the healthy human gut microbiome.

European Bioinformatics Institute, Hinxton, Cambridgeshire, UK
Visiting postdoctoral fellow, Jan 2023—July 2023

I visited the Pathogen Informatics and Modelling group led by Dr John Lees through the Theory@EMBL program.

University of Helsinki, Helsinki, Finland
Doctoral researcher, 2018—2022

Developed probabilistic models for determining the composition of a DNA sequencing data sample. Produced efficient open source implementations of the said models and distributed them. Published results in scientific journals and presented at international conferences.

Helsinki Institute for Information Technology, Helsinki, Finland
Research assistant, 2016—2017

Designed and implemented models for identifying bacterial strains from DNA sequencing data.

Other employment

Heldata oy, Helsinki, Finland
Part-time entrepreneur, 2018—2022

Planned, organised, and hosted a 250-person data science networking and recruiting event in 2019 and 2021. Ran the day-to-day affairs of the business. Worked with customers to find bespoke solutions.

Stokastiko oy, Helsinki, Finland
Independent consultant, 2016—2018

Designed and taught a 3-day statistics introductory course for the tax avoidance unit of the Finnish Tax Administration. Consulted small businesses and organizations in automating data analysis pipelines. Founded and operated a one-man company.

Publications

Preprints

Enhanced metagenomics-enabled transmission inference with TRACS
G Tonkin-Hill, Y Shao, AE Zarebski, S Mallawaarachchi, Ouli Xie, **T Mäklin**, HA Thorpe, MR Davies, SD Bentley, TD Lawley, and J Corander
preprint (2024), available at *bioRxiv*: [10.1101/2024.08.19.608527](#)
software in GitHub: [TRACS](#)

Basic reproduction number for pandemic *Escherichia coli* clones is comparable to typical pandemic viruses
FA Ojala, H Pesonen, RA Gladstone, **T Mäklin**, G Tonkin-Hill, P Marttinen, and J Corander
preprint (2024), available at *bioRxiv*: [10.1101/2024.05.08.593267](#)

In-press

Geographical variation in colorectal and urinary tract linked cancer incidence is associated with population exposure to colibactin-producing *Escherichia coli*
T Mäklin, A Taira, S Arredondo-Alonso, Y Shao, MR Stratton, TD Lawley, LA Aaltonen, and J Corander
The Lancet Microbe (2024).
Main findings in Zenodo: [10.5281/zenodo.11275910](#).

Peer-reviewed journal articles

Seamless, rapid, and accurate analyses of outbreak genomic data using split *k*-mer analysis
R Derelle, J von Wachsmann, **T Mäklin**, J Hellewell, T Russell, A Lalvani, L Chindelevitch, NJ Croucher, SR Harris, and JA Lees
Genome Research (2024), doi: [10.1101/gr.279449.124](#)
software in GitHub: [SKA2](#)

Pan-pathogen deep sequencing of nosocomial bacterial pathogens in Italy in Spring 2020: a prospective cohort study
HA Thorpe, M Pesonen, M Corbella, H Pesonen, S Gaiarsa, CJ Boinett, G Tonkin-Hill, **T Mäklin**, AK Pöntinen, N MacAlasdair, RA Gladstone, S Arredondo-Alonso, T Kallonen, D Jamrozy, SW Lo, C Chaguza, GA Blackwell, A Honkela, AC Schürch, RJL Willems, C Merla, G Petazzoni, EJ Feil, P Cambieri, NR Thomson, SD Bentley, D Sassera, and J Corander
The Lancet Microbe (2024), doi: [10.1016/S2666-5247(24)00113-7](#)

Deep sequencing of *Escherichia coli* exposes colonisation diversity and impact of antibiotics in Punjab, Pakistan
T Khawaja°, **T Mäklin**°, T Kallonen°, RA Gladstone, AK Pöntinen, S Mero, HA Thorpe, Ø Samuelsen, J Parkhill, M Izhar, WM Akhtar, J Corander, and A Kantele
° joint first authors
Nature Communications (2024), doi: [10.1038/s41467-024-49591-5](#)

Exploring the paediatric nasopharyngeal bacterial microbiota with culture-based MALDI-TOF mass spectrometry and targeted metagenomic sequencing
P Sreymom, T Kallonen, **T Mäklin**, P Sar, J Hopkins, S Soeng, T Miliya, CL Ling, SD Bentley, J Corander, and P Turner
mBio (2024), doi: [10.1128/mbio.00784-24](#)

A novel prophage-like insertion element within yabG triggers early entry into sporulation in *Clostridium botulinum*
FP Douillard, I Martins Portinha, Y Derman, C Woudstra, **T Mäklin**, MB Dorner, H Korkeala, AO Henriques, and M Lindström
Viruses (2023), doi: [10.3390/v15122431](#)

Themisto: a scalable colored *k*-mer index for sensitive pseudoalignment against hundreds of thousands of bacterial genomes
JN Alanko, J Vuohtoniemi, **T Mäklin**, and SJ Puglisi
Bioinformatics (2023), doi: [10.1093/bioinformatics/btad233](#)
software in GitHub: [Themisto](#)

Strong pathogen competition in neonatal gut colonisation
T Mäklin, HA Thorpe, AK Pöntinen, RA Gladstone, Y Shao, M Pesonen, A McNally, PJ Johnsen,

Ø Samuelsen, TD Lawley, A Honkela, and J Corander
Nature communications (2022), doi: [10.1038/s41467-022-35178-5](#)

Genomic and phenotypic characterization of Clostridium botulinum isolates from an infant botulism case suggests adaptation signatures to the gut
FP Douillard, Y Derman, C Woudstra, K Selby, **T Mäklin**, MB Dorner, H Saxén, BG Dorner, H Korkeala, and M Lindström
mBio (2022), doi: [10.1128/mbio.02384-21](#)

Bacterial genomic epidemiology with mixed samples
T Mäklin, T Kallonen, JN Alanko, Ø Samuelsen, K Hegstad, V Mäkinen, J Corander, E Heinz, and A Honkela
Microbial genomics (2021), doi: [10.1099/mgen.0.000691](#)
software in GitHub: [mGEMS](#)

High-resolution sweep metagenomics using ultrafast read mapping and inference
T Mäklin, T Kallonen, S David, CJ Boinett, B Pascoe, G Méric, DM Aanensen, EJ Feil, S Baker, J Parkhill, SK Sheppard, J Corander, and A Honkela
Wellcome open research (2021), doi: [10.12688/wellcomeopenres.15639.2](#)
software in GitHub: [mSWEEP](#)

Closing *Clostridium botulinum* Group III genomes using long-read sequencing
C Woudstra, **T Mäklin**, Y Derman, L Bano, H Skarin, C Mazuet, A Honkela, and M Lindström
Microbiology resource announcements (2021), doi: [10.1128/MRA.01364-20](#)

Book chapters

Identifying bacterial strains from sequencing data
T Mäklin, J Corander, and A Honkela
Data mining for systems biology (2018), doi: [10.1007/978-1-4939-8561-6_1](#)

Grants

Theory@EMBL visitor fellowship
For visit costs, approx. £5 000, 2023

DONASCI thesis finalization grant
Covered six months salary, approx. €20 000, 2021

Research visits

Institute of Basic Medical Sciences, University of Oslo, Oslo, Norway
Visitor, May 2024

European Bioinformatics Institute, Hinxton, Cambridge, UK
Visiting postdoctoral fellow, Jan 2023—July 2023

External affiliations

Wellcome Sanger Institute, Hinxton, Cambridge, UK
Visiting worker, Pathogens and microbes programme, 2022-2024

Positions of trust

University of Helsinki, Helsinki, Finland
Union representative (shop steward), 2023—2024

Volunteer experience

Finnish union of university researchers and teachers

Board member, 2024 ∘ Deputy board member, 2023

Helsinki university researchers and teachers association
Vice chair, 2023—2024 ∘ Board member, 2022—2025 ∘ Deputy board member 2020—2021

University of Helsinki PhD students association
Chair, 2020—2021 ∘ Secretary 2022

Finnish statistical society
Secretary, 2017—2022 ∘ Deputy board member 2022—2023

Service

Boards and committees

Doctoral school in natural sciences, University of Helsinki
Board member, 2020—2021

Doctoral programme in Computer Science, University of Helsinki
Board member, 2020—2021

Peer review

2024 ECML PKDD, *Journal of Open Source Software, Microbiome*
2023 *Genome biology, Nature communications, Microbiome, Scientific reports*
2022 *BMC microbiology*
2021 International conference on machine learning (ICML)
2019 International conference on machine learning (ICML)

Research software

mGEMS
Main contributor
Taxonomic binner with lineage-level resolution, PROBIC/mGEMS

mSWEEP
Main contributor
Taxonomic profiler with lineage-level resolution, PROBIC/mSWEEP

Themisto
With Jarno N Alanko & others
Pseudoaligner and *k*-mer index for short-read sequencing data, algbio/Themisto

SKA2
With John A Lees & others
Split *k*-mer analysis of closely related genomic sequences, bacpop/ska.rust

cpprate
Main contributor
Variable selection in Bayesian regression models with relative centrality, tmaklin/cpprate

---

Media coverage

Newspapers with national reach

Sairaalabakteeri ei tartu terveisiin vastasyntyneisiin, kertoo tutkimus
Helsingin Sanomat, Helsinki, Finland, 8 February 2023
https://www.hs.fi/helsinki/art-2000009380665.html

Presentations and conference activity

Session chair

Probabilistic modelling in microbial ecology workshop
Kellogg College, University of Oxford, Oxford, UK, March 2023

Invited talks

Computational methods in AMR research
University of Turku, Turku, Finland, October 2023

Lawley group meeting
Wellcome Sanger Institute, Hinxton, Cambridge, UK, June 2023

Theory@EMBL seminar
European Molecular Biology Laboratory EMBL, Hinxton, Cambridge, UK, June 2023

Heinz group meeting
Liverpool school of tropical medicine, Liverpool, UK, February 2023

Lahti group meeting
University of Turku, Turku, Finland, January 2023

AMR-bridge annual meeting
AMR-bridge consortium, Tromsø, Norway, June 2022

Thomson group meeting
Wellcome Sanger Institute, Hinxton, Cambridge, UK, June 2020

CompLifeSci 4th annual meeting
Åbo Akademi, Turku, Finland, November 2019

Contributed talks

SMBE Satellite Workshop on Pathogen Pangenome Evolution Between and Within their Hosts Tokyo, Japan, November 2024

Microbial (gen)omics meeting
Wellcome Sanger Institute, Hinxton, Cambridge, UK, July 2023

EBI research retreat
European Bioinformatics Institute, Hinxton, Cambridge, UK, July 2023

ISMB/ECCB 2019
Basel, Switzerland, July 2019

Bioinformatics research and education workshop
Max Planck Institute for Molecular Genetics, Berlin, Germany, May 2019

Mathematical and statistical aspects of molecular biology workshop
European Bioinformatics Institute, Hinxton, Cambridge, UK, April 2019

Probabilistic modelling in genomics
Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, US, November 2018

Regional workshop of European young researchers in Statistics
Henri Poincaré Institute, Paris, France, November 2018

Mathematical and statistical aspects of molecular biology workshop
University of St Andrews, St Andrews, Scotland, UK, March 2018