Workshop on computational genomics

Time: 6-9 November 2025
Venue: Beijing Normal University (Main Campus)

Organizers

Professor Ziheng Yang, University College London
Professor Dayong Zhang, Beijing Normal University
Professor Jie Han, Beijing Normal University
Professor Weining Bai, Beijing Normal University
Professor Yi Hu, Beijing Normal University
Associate Professor, Yufeng Zhang, Beijing Normal University

Instructors

Sandra Álvarez-Carretero (UCL),
Thomas Flouris (UCL),
Xiyun Jiao (SusTech),
Yuttapong Thawornwattana (UCL),
Tianqi Zhu (AMSS, Beijing),
Xiao-Xu Pang (Lanzhou University),
Ziheng Yang (UCL)

Apply at
https://www.wjx.cn/vm/OP20tiG.aspx#
or scan the QR code above
Closing date for application: **31 August 2025**

Please note: Registration is free. Participants are responsible for their own travel, accommodation, and meals.

Overview

This workshop will focus on Bayesian estimation of species divergence times incorporating fossil calibrations and Bayesian analyses of genomic data from closely related species under the multispecies coalescent model. The workshop consists of a mixture of lectures on the theory and methods and software practice (mcmctree and bpp). We will discuss compilation of multi-locus sequence data, give an introduction to Bayesian methods and Markov chain Monte Carlo algorithms, and focus on Bayesian relaxed-clock dating and Bayesian inference under the multispecies coalescent model, including estimation of species trees and species divergence times, estimation of the rate of interspecific gene flow, and species delimitation.

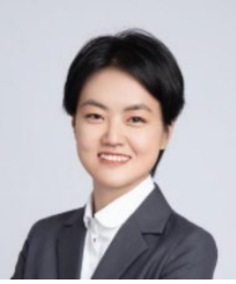
The workshop will be suitable for PhD students, postdocs and researchers who work in the areas of phylogenomics and population genomics, in particular researchers who already have experience with phylogenetic programs and who have genomic data to analyze. Bioinformaticians may apply as well. We expect 30-40 participants. There will be a selection process.

Selected participants will be asked to do some reading and get familiar with linux command line. They shall bring their own laptops and must install the software before the workshop. We will send instructions and provide tutorials on Linux command line which the participants can go through before the workshop. Familiarity with a scripting language such as Python or R will also be helpful. We will set up a WeChat group for announcements, questions and discussions.

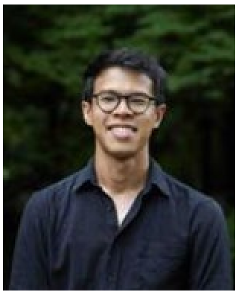
Instructor blurbs

Sandra Álvarez-Carretero, Computational Biologist and research fellow at University College London. Sandra's work focuses on Bayesian timetree inference and the development of reproducible methods for evolutionary studies. She specializes in MCMC algorithms and empirical datasets, investigating divergences of mammals, vertebrate terrestrialisation, and the eukaryotic Tree of Life.

Thomas Flouris, Senior Research Fellow at University College London. Thomas develops statistical methods for analysing phylogenomic datasets and is a key contributor to tools like RAxML and BPP. His current work focuses on species tree inference and gene flow analysis using the BPP software for multilocus sequence data.



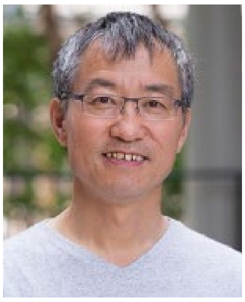
Xiyun Jiao, Assistant Professor, Southern University of Science and Technology. Her research focuses on computational statistics, Bayesian statistics, Markov Chain Monte Carlo algorithms, and statistical methods in population genetics.



Yuttapong Thawornwattana, Research Fellow at University College London. Yuttapong specializes in phylogenomics and inference of gene Flow using genomic data. With a background in Biological Sciences (UCL) and Computational Statistics (UCL), his PhD at Harvard focused on introgression in butterflies and mosquitoes, alongside epidemiological studies in malaria and global health.

Tianqi Zhu, Associate Professor, Academy of Mathematical and Systems Sciences, Chinese Academy of Sciences. Tianqi works in molecular evolution and phylogeneics, and computational biology. She develops maximum likelihood and Bayesian methods of inference in Computational genomics.

Xiao-Xu Pang, Postdoctoral Researcher, Lanzhou University. Xiao-Xu works in phylogenomics with an emphasis on phylogenetic network inference. She has an BSc in Applied Mathematics (BSc, Lanzhou University) and MSc and PhD in Biological Sciences (Beijing Normal University). Her doctoral work involved studying ghost introgression, focusing on its influence on species tree reconstruction and methodological detection strategies.



Ziheng Yang, RA Fisher Professor at University College London. He works in molecular evolution, systematics, and computational biology. He develops statistical models and algorithms to analyse genetic data for evolutionary inference. Yang maintains two widely-used programs, PAML and BPP. He was elected a Fellow of the Royal Society in 2006.

Draft programme

Date		Speaker
6 November	**Introduction & Overview**	
10:00-10:10	Welcome address	Dayong Zhang
	Intro to Bayesian statistics & Markov chains	
10:10-11:40	**Lecture**: Bayesian statistics and Markov chain Monte Carlo	TF/XJ/YT/SAC
11:40-12:30	Practical on MCMC	TF/XJ/YT/SAC
12:30-14:00	**Lunch**	
14:00-15:00	Practical on MCMC	SAC/TF/XJ/YT
15:00-15:30	Genomic data processing	YT/TF
15:30-15:50	Break	
15:50-17:30	Genomic data processing	YT/TF
7 November	**Bayesian relaxed clock dating of species divergences**	
09:00-10:30	Lecture: Markov chain models of DNA evolution, phylogenetic likelihood	TF/YT/SAC/TZ
10:30-10:50	Break	
10:50-12:30	Practical on phylogeny reconstruction with IQ-TREE	YT/TF/SAC/TZ
12:30-14:00	**Lunch**	
14:00-15:30	Lecture: Bayesian molecular clock dating of species divergences	SAC/ZY
15:30-15:50	Break	
15:50-17:30	Practical: MCMCtree	SAC/TF/YT
8 November	**Multispecies coalescent & species tree estimation**	
09:00-10:30	Lecture: Multispecies coalescent model	ZY/XJ
10:30-10:50	Break	
10:50-12:30	Practical: Multispecies coalescent model	ZY/XJ
12:30-14:00	**Lunch**	
14:00-15:30	Lecture/Practical: Species-tree Inference	ZY/TF
15:30-15:50	Break	
15:50-17:30	Practical: BPP Species-tree Inference	TF/ZY
9 November	**Gene flow and species delimitation**	
09:00-10:30	Lecture: Models of gene flow (introgression and migration)	ZY/TF
10:30-10:50	Break	
10:50-12:30	Practical: BPP Introgression models (msci-create, test of gene flow)	YT/TF/XP/ZY
12:30-14:00	**Lunch**	
14:00-15:30	Practical: BPP migration Models (including threads, checkpointing)	YT/TF/XP/ZY
15:30-15:50	Break	
15:50-16:20	Lecture: Species delimitation	ZY/TF
16:20-17:30	Practical: Species delimitation using BPP/HHSD	TF/ZY
17:30-17:45	**Discussion, Q&A, summary, & feedback**	ZY/TF/SAC/XJ
	Departure	

SAC: Sandra, TF: Thomas, XJ: Xiyun, YT: Yuttapong, TZ: Tianqi, XP: Xiaoxu, ZY: Ziheng

Please install the following software programs before the workshop

R (https://www.r-project.org/)
RStudio (https://posit.co/download/rstudio-desktop/)
(Please install R before installing RStudio. Windows users may need to install RTools. Please make sure that you get the version that is compatible with the R version you are downloading).

paml/mcmctree (https://github.com/abacus-gene/paml)
The current version is 4.10.7.

Tracer (https://beast.community/tracer)

FigTree (https://beast.community/figtree)

bpp (https://github.com/bpp/bpp)
The current version is 4.8. Please install the most recent release.

IQ-TREE (http://www.iqtree.org/#download)
Please install the latest release (at the time of writing, IQ-TREE v2.4.0).

Linux exercises before the workshop

If you have the time, you are most welcome to read some of the review papers before the workshop. See list at the end of this document.

We will use the linux command line during our computer practice, even though in theory you can run the windows or mac versions of mcmctree and bpp as well. We assume that you have basic computer skills and know common linux commands (cp, rm, ls, cat, nice, bg, etc.) and basic uses of linux programs (tar, grep, sed, awk). If you do not, please go through the short tutorials we provide below before the workshop. There are also many tutorials you can follow on the Internet. You can also ask your friend or student to teach you.

We will set up a WeChat group where you can post questions about linux commands or about issues you may have when trying to install the software. Others can help answering questions.

MS Windows command line:
http://abacus.gene.ucl.ac.uk/software/CommandLine.Windows.pdf
Linux/mac OSX command line:
http://abacus.gene.ucl.ac.uk/software/CommandLine.MACosx.pdf
Linux command line (a much more comprehensive tutorial):
http://abacus.gene.ucl.ac.uk/software/CommandLine.Unix.pdf

Workshop Reading List

Bayesian methods and MCMC

- Nascimento, F.F., dos Reis, M., and Yang, Z. A biologist's guide to Bayesian phylogenetic analysis. *Nat. Ecol. Evol.* 1, 1446–1454 (2017). https://doi.org/10.1038/s41559-017-0280-x.
- Chen, M.H., Kuo L., and Lewis P.O. (2014) Bayesian phylogenetics: Methods, algorithms, and applications. CRC Press.
- Yang, Z. 2014. Molecular Evolution: A Statistical Approach. Oxford University Press, Oxford, England, Chapters 6-9.

Reviews and tutorials

- Yang, Z., and Rannala, B. 2012. Molecular phylogenetics: principles and practice. *Nat. Rev. Genet.* 13: 303-314.
- dos Reis, M., Donoghue, P.C.J., Yang, Z. 2016. Bayesian molecular clock dating of species divergences in the genomics era. *Nat. Rev. Genet.* 17:71-80.
- Xu, B., and Yang, Z. 2016. Challenges in species tree estimation under the multispecies coalescent model. *Genetics*. 204: 1353-1368.
- Kubatko, L. 2019. The multispecies coalescent. In: Balding D, Moltke I, Marioni J, editors. Handbook of Statistical Genomics. New York: Wiley. p. 219-245.
- Jiao, X., Flouri, T., and Yang, Z. 2021. Multispecies coalescent and its applications to infer species phylogenies and cross-species gene flow. *Nat. Sci. Rev.* 8: DOI: 10.1093/nsr/nwab1127.
- Hibbins, M.S., and Hahn, M.W. 2022. Phylogenomic approaches to detecting and characterizing introgression. *Genetics*. 220:10.1093/genetics/iyab1173.
- Flouri, T., Rannala, B., and Yang, Z. 2020. A tutorial on the use of BPP for species tree estimation and species delimitation. Pp. 5.6.1-16 in Scornavacca, C., Delsuc, F., and Galtier, N., eds. Phylogenetics in the Genomic Era.
- dos Reis, M., and Yang, Z. 2019. Bayesian molecular clock dating using genome-scale datasets. In: Anisimova, M. (ed.) Evolutionary Genomics. Methods in Molecular Biology, vol 1910. Humana, New York, NY. https://doi.org/10.1007/978-1-4939-9074-0_10.

Clock, relaxed clock, and Bayesian dating of divergences

- dos Reis, M. and Yang, Z., 2011. Approximate likelihood calculation on a phylogeny for Bayesian estimation of divergence times, *Mol. Biol. Evol*, 28: 2161–2172, https://doi.org/10.1093/molbev/msr045.
- Lartillot, N. PhyloBayes: Bayesian phylogenetics using site-heterogeneous models. Scornavacca, Celine; Delsuc, Frédéric; Galtier, Nicolas. Phylogenetics in the Genomic Era, No commercial publisher | Authors open access book, pp.1.5:1–1.5:16, 2020. ffhal-02535342.
- Holder, M., and Lewis, P. Phylogeny estimation: traditional and Bayesian approaches. *Nat*. *Rev*. *Genet*. 4, 275-284 (2003). https://doi.org/10.1038/nrg1044.
- Yang, Z. 2014. Molecular Evolution: A Statistical Approach. Oxford University Press, Oxford, England, Chapter 10.
- Álvarez-Carretero, S., and dos Reis, M. (2021) Bayesian Phylogenomic Dating. In: Ho, S.Y.W. (eds) The Molecular Evolutionary Clock. Springer, Cham. https://doi.org/10.1007/978-3-030-60181-2_13.

Species tree estimation

- Zhu, T, and Yang, Z. 2021. Complexity of the simplest species tree problem. *Mol. Biol. Evol.* 39:3993–4009. 10.1093/molbev/msab009.
- Flouri, T., Jiao, X., Rannala, B., and Yang, Z. 2018. Species tree inference with BPP using genomic sequences and the multispecies coalescent. *Mol*. *Biol*. *Evol*. 35:2585-2593.

Gene flow (MSC-I and MSC-M)

- Flouri, T., Jiao, X., Rannala, B., and Yang, Z. 2020. A Bayesian implementation of the multispecies coalescent model with introgression for phylogenomic analysis. *Mol. Biol. Evol.* 37:1211-1223.
- Flour,i T., Jiao, X., Huang, J., Rannala, B., and Yang, Z. 2023. Efficient Bayesian inference under the multispecies coalescent with migration. *PNAS* 120 (44): e2310708120.
- Jiao, X., Flouri, T., Rannala, B., and Yang, Z. 2020. The impact of cross-species gene flow on species tree estimation. *Syst. Biol.* 69:830-847.
- Jiao, X., and Yang, Z. 2021. Defining species when there is gene flow. *Syst. Biol.* 70:108–119.
- Ji, J., Jackson, D.J., Leache, A.D., and Yang, Z. 2023. Power of Bayesian and heuristic tests to detect cross-species introgression with reference to gene flow in the Tamias quadrivittatus group of North American chipmunks. *Syst. Biol.* 72:446-465.
- Pang, X.X., and Zhang, D.Y. 2024. Detection of ghost introgression requires exploiting topological and branch length information. *Syst. Biol.* 73:207-222.

Species delimitation

- Yang, Z., and Rannala, B. 2010. Bayesian species delimitation using multilocus sequence data. *Proc. Natl. Acad. Sci. U.S.A.* 107:9264-9269.
- Leaché, A.D., Zhu, T., Rannala, B., and Yang, Z. 2019. The spectre of too many species. *Syst. Biol.* 68:168-181. 10.1093/sysbio/syy051.
- Kornai, D., Jiao, X., Ji, J., Flouri, T., and Yang, Z. 2024. Hierarchical heuristic species delimitation under the multispecies coalescent model with migration. *Syst. Biol*. 10.1093/sysbio/syae050.
- Chambers EA, Lara-Tufino JD, Campillo-Garcia G, Cisneros-Bernal AY, Dudek DJ, Jr., Leon-Regagnon V, Townsend JH, Flores-Villela O, Hillis DM. 2025. Distinguishing species boundaries from geographic variation. *Proc. Natl. Acad. Sci. U.S.A.* 122:e2423688122.

Applications

- Thawornwattana, Y., Dalquen, D.A., and Yang, Z. 2018. Coalescent analysis of phylogenomic data confidently resolves the species relationships in the *Anopheles gambiae* species complex. *Mol. Biol. Evol.* 35:2512-2527.
- Thawornwattana, Y., Seixas, F.A., Mallet, J., and Yang, Z. 2022. Full-likelihood genomic analysis clarifies a complex history of species divergence and introgression: the example of the erato-sara group of Heliconius butterflies. *Syst. Biol.* 71:1159-1177.
- Thawornwattana, Y., Seixas, F.A., Yang, Z., and Mallet, J. 2023. Major patterns in the introgression history of *Heliconius* butterflies. *eLife* 12: RP90656, DOI:90610.97554/eLife.90656.
- Álvarez-Carretero, S., Tamuri, A.U., Battini, M. *et al.* 2022. A species-level timeline of mammal evolution integrating phylogenomic data. *Nature* 602, 263–267. https://doi.org/10.1038/s41586-021-04341-1.
- Moody, E.R.R., Álvarez-Carretero, S., Mahendrarajah, T.A. *et al.* 2024. The nature of the last universal common ancestor and its impact on the early Earth system. *Nat*. *Ecol*. *Evol*. 8, 1654–1666. https://doi.org/10.1038/s41559-024-02461-1.