

Polymorphism-aware phylogenetic models

Workshop, MIC-Phy 2021

Dominik Schrempf

February 16, 2021



Eötvös Loránd
University

Introduction

This workshop is [available on GitHub](#).

Our goal is to understand polymorphism-aware phylogenetic models (PoMo) and learn how to apply PoMo to phylogenetic data.

In the course of this workshop, we will infer a phylogenetic tree from a test data set using IQ-TREE2.

If you need help, please interrupt me anytime!

Preparation - Command line shell

Basic knowledge of the command line shell of your choice is assumed.

- If you do not know basic commands such as `cd`, `ls`, or `less`, just follow the presentation.
- Otherwise, try to complete the workshop yourself.

In case you are lost:

- Have a look at the manual pages, if they exist.

```
1 man less
```

- Read how commands are used.

```
1 less --help
```

Preparation - Download workshop and data

Option 1: If you have git installed, use it.

```
1 git clone https://github.com/pomo-dev/micphy-workshop.git
2 cd micphy-workshop
```

Option 2: Manually download the archive.

```
1 wget https://github.com/pomo-dev/micphy-workshop/archive/master.zip
2 unzip master.zip
3 cd micphy-workshop-master
```

The advantage of Option 1 is that you can:

- update your working tree using `git pull` (if I have to change something during the workshop);
- reset to the initial state in case you mess something up with `git reset --hard HEAD` (be careful, this erases all uncommitted changes made by you).

Preparation - IQ-TREE2: Installation

Option 1: Repository of your Linux distribution. For example, using Arch Linux and the Arch User Repository.

```
1 yay -S iqtree
2 aura -A iqtree
```

Option 2: Download the binary executable from the [IQ-TREE2 homepage](#). Make sure that you have permission to execute the file (`chmod +x`), and that the executable is either in your `PATH`, or that you give the path during execution.

```
1 iqtree2 --version
2 /path/to/iqtree2 --version
3 ./relative/path/to/iqtree2 --version
```

Preparation - IQ-TREE2: Installation

Option 3: Compile yourself (not shown).

Option 4 for script kiddies or experienced users: Use `nix-shell` and the `shell.nix` expression provided in the base directory of the repository.

```
1 nix-shell
```

Welcome to the MIC-Phy PoMo workshop.

The following version of IQ-TREE2 is available:

IQ-TREE multicore version 2.1.2 COVID-edition for Linux 64-bit built Jan 1 1980
Developed by Bui Quang Minh, James Barbetti, Nguyen Lam Tung,
Olga Chernomor, Heiko Schmidt, Dominik Schrempf, Michael Woodhams.

Preparation - IQ-TREE2: Test version

Depending on how you installed IQ-TREE2, please check that the version agrees with the one I am using.

```
1 iqtree2 --version
2 # /path/to/iqtree2 --version
3 # ./relative/path/to/iqtree2 --version
```

IQ-TREE multicore version 2.1.2 COVID-edition for Linux 64-bit built Jan 1 1980
Developed by Bui Quang Minh, James Barbetti, Nguyen Lam Tung,
Olga Chernomor, Heiko Schmidt, Dominik Schrempf, Michael Woodhams.

Exercise - IQ-TREE2: Access help

- How to get help (IQ-TREE).

Exercise - Run a DNA substitution model

- Run normal model.

Fruit fly data

Data from [PopFly](#)¹. 9 populations with an average number of samples per population of approximately 19, and an estimated heterozygosity of 0.0109:

NTH Netherlands

EG Egypt

FR France

GA Gabon

GU Guinea

EF Ethiopia

KN Kenya

SB South Africa (Barkly East)

SP South Africa (Phalaborwa)

Why is it important to check the heterozygosity?

¹Hervas et al. ([2017](#)).

Exercise - Run PoMo

```
1 iqtree2 -nt 4 -redo -s data/fruit_flies_1000.cf -m HKY+P+N9 -pre fruit_flies_1000.cf.N9
```

- Run PoMo.
- Find best N.
- Compare different DNA substitution models.
- Use gamma rate heterogeneity.
- Probably perform model test?
- Bootstrapping.
- Compare branch lengths for different N values.
- Compare to using normal DNA substitution models.
- Probably provide results for the 10k alignment (different N, different G, with bootstrapping?).

Literature





PoMo De Maio et al. (2015).

Reversible PoMo Schrempf et al. (2016) and Schrempf and Hobolth (2017).



Advanced models with PoMo Schrempf et al. (2019).

IQ-TREE2 Minh et al. (2020).

Bibliography I

-  De Maio, Nicola, Dominik Schrempf, and Carolin Kosiol (2015). “PoMo: An Allele Frequency-Based Approach for Species Tree Estimation.” In: *Systematic Biology* 64.6, pp. 1018–1031. DOI: [10.1093/sysbio/syv048](https://doi.org/10.1093/sysbio/syv048).
-  Hervas, Sergi, Esteve Sanz, Sònia Casillas, John E Pool, and Antonio Barbadilla (2017). “Popfly: the Drosophila Population Genomics Browser.” In: *Bioinformatics* 33.17, pp. 2779–2780. DOI: [10.1093/bioinformatics/btx301](https://doi.org/10.1093/bioinformatics/btx301).
-  Minh, Bui Quang, Heiko A Schmidt, Olga Chernomor, Dominik Schrempf, Michael D Woodhams, Arndt von Haeseler, and Robert Lanfear (2020). “IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era.” In: *Molecular Biology and Evolution* 37.5. Ed. by Emma Teeling, pp. 1530–1534. DOI: [10.1093/molbev/msaa015](https://doi.org/10.1093/molbev/msaa015).
-  Schrempf, Dominik, Bui Quang Minh, Nicola De Maio, Arndt von Haeseler, and Carolin Kosiol (2016). “Reversible polymorphism-aware phylogenetic models and their application to tree inference.” In: *Journal of Theoretical Biology* 407, pp. 362–370. DOI: [10.1016/j.jtbi.2016.07.042](https://doi.org/10.1016/j.jtbi.2016.07.042).

Bibliography II

-  Schrempf, Dominik and Asger Hobolth (2017). “An alternative derivation of the stationary distribution of the multivariate neutral Wright–Fisher model for low mutation rates with a view to mutation rate estimation from site frequency data.” In: *Theoretical Population Biology* 114, pp. 88–94. DOI: 10.1016/j.tpb.2016.12.001.
-  Schrempf, Dominik, Bui Quang Minh, Arndt von Haeseler, and Carolin Kosiol (2019). “Polymorphism-Aware Species Trees with Advanced Mutation Models, Bootstrap, and Rate Heterogeneity.” In: *Molecular Biology and Evolution* 36.6. Ed. by Naruya Saitou, pp. 1294–1301. DOI: 10.1093/molbev/msz043.