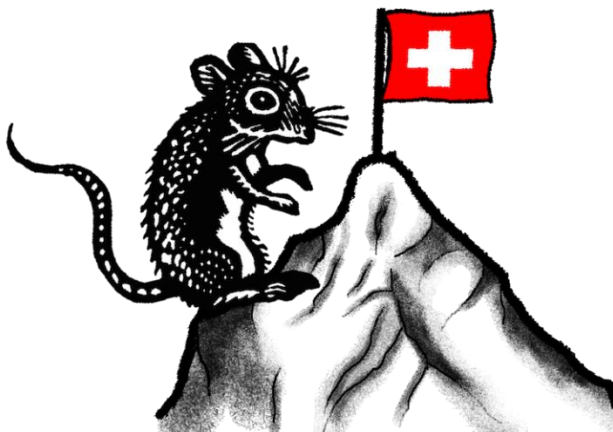




Taming the BEAST

A Host's Workshop Guide



Contents

Introduction	3
Previous workshops	4
The programme	4
Funding	5
Registration and assessing applicants	6
Evaluation	6
Timeline	7
Things to keep in mind	8
Sample timetable and required software	10

Introduction

Phylogenetics and phylodynamics are central topics in modern biology. Phylogenetic inferences reconstruct the evolutionary relationships between organisms, whereas phylodynamic inferences reveal the dynamics that lead to the observed relationships. These two fields have many practical applications in disciplines such as epidemiology, developmental biology, paleontology, ecology and even linguistics. However phylogenetics and phylodynamics are complex and fast-evolving fields. As such, inference tools are not easily accessible to researchers who are not from a computational background.

[BEAST 2](#) is a cross-platform program for Bayesian phylogenetic analysis of molecular sequences. It estimates rooted, time-measured phylogenies using strict or relaxed molecular clock models. It can be used as a method of reconstructing phylogenies but is also a framework for testing evolutionary hypotheses without conditioning on a single tree topology. BEAST 2 uses Markov chain Monte Carlo (MCMC) to average over tree space, so that each tree is weighted proportional to its posterior probability. BEAST 2 includes a graphical user-interface for setting up standard analyses and a suit of programs for analysing the results. A core aspect of the BEAST 2 design philosophy is to allow researchers to extend the core program without the involvement of the BEAST 2 developers. New methods are implemented as standalone packages that are developed and maintained separately of BEAST 2 itself.

BEAST 2 is a complex program and it is challenging to learn BEAST without the help of an expert. Workshops, summer schools and lectures are a great resource, but can only reach a handful of the people interested in performing BEAST analyses. **Taming the BEAST** is a platform for collecting a comprehensive set of BEAST 2 tutorials in one location. By providing a set of well-curated and up-to-date tutorials **Taming the BEAST** aims to fill this gap and provide researchers with the resources necessary to learn how to perform analyses in BEAST 2.

In keeping with the BEAST 2 design, it is possible for anyone to extend or correct the set of tutorials on the website. Thus, Taming the BEAST relies on the community to curate tutorials and raise issues to ensure that all of the tutorials are up-to-date. Furthermore, it is possible for developers to upload tutorials to newly implemented packages. Finally, we also encourage anyone else with useful BEAST 2 teaching materials to upload them. Find out how you can [contribute](#).

Taming the BEAST team

Alexei Drummond, Tanja Stadler, Tim Vaughan and Louis du Plessis

Previous workshops

[Taming the BEAST](#) has had successful workshops in various parts of the world and typically runs in Europe every year during the summer season. Tanja Stadler's group at cEvo ([computational Evolution](#)), ETH Zürich organised the first workshop in 2016 in Engelberg, Switzerland. Taming the BEAST has also been hosted on Waiheke Island, New Zealand in the South Pacific, London in the United Kingdom and again in Europe in Oberägeri, Switzerland. Here we have collected together some notes on what makes a successful workshop.

To begin with, the common denominator between each of these venues is that they're all very stimulating or picturesque locations. A beautiful location engages the participants, and teaching and tutoring staff so much more. Whilst being beautiful, all the locations have also been accessible in terms of proximity to accommodation, transport options and amenities. The last practical aspect of a successful location is it must be able to comfortably accommodate and cater to at least 40 participants (in addition to teaching/tutoring staff) by way of desks, presentation and audio equipment, air conditioning and most importantly, super-fast wireless internet.

Once the venue is secured (or at least decided), the host needs to then decide on the focus for their workshop. In the past we've had general phylogenetics, multispecies coalescent, total-evidence approaches and infectious disease modelling, to name a few. Typically there has not been two consecutive workshops with the same focus. Naturally, the focal point will depend on the hosts' background and research interests, but it is recommended to refer to the [website](#) and to speak to a member of the Taming the BEAST team to see what the current topic of interest is.

With the venue and appropriate focus decided, the host will need to secure the most relevant teaching staff well ahead of time. With 35 - 40 (50 participants maximum) you will ideally want at least 10 - 15 members of staff on rotation (3:1 Students:Tutors) made up of teachers and teaching assistants drawn from local and international locations. This group of staff often includes members from Tanja Stadler's cEvo group and members of the Centre for Computational Evolution.

The programme

Taming the BEAST workshops are one week long. This allows enough time to properly teach the desired focus as well as provide networking opportunities and deliver a truly international experience. Therefore, it is recommended to have a welcome event the afternoon/evening before the workshop begins. There will be the expected morning/afternoon and lunch breaks. However, a charm of Taming the BEAST workshops is that there is also an activities afternoon situated in the middle of the week and one to two dinners to attend throughout the week. The dinners and activities can be organised by the hosts or participants may organise themselves. Previous workshop programmes can be found [here](#).

Funding

With running an international standard workshop, comes associated costs that need to be anticipated to make Taming the BEAST accessible to all learners. It is understood that in order to host a successful workshop, the host/s need to be comfortable and able to afford most of the below costs.

Venue:

The venue, which is assumed to be in a stimulating and accessible location will carry most of the cost. Venues will have their preferred caterers and usually offer day delegate packages to cover meal breaks. Room and equipment may either be included in the event, but it may be an additional cost you incur.

Teaching staff:

Previously members from Tanja Stadler's cEvo group and members of the Centre for Computational Evolution have been able to secure their own funding to allow them to travel and teach at BEAST workshops. But for all other teaching and tutoring staff, there is the cost of their travel and accommodation and whatever else you allow for them. This and the venue are the bare minimum that you must account for.

Activities:

As previously mentioned, the programme should have time allocated for social and networking opportunities. Typically Taming the BEAST workshops have a cocktail style welcome event the night before the workshop starts. There are also one to two dinners across the week. As a host, anything beyond one dinner, you may then either absorb the cost or you could put it back on the participants. Depending on the resources available, the unscheduled afternoon mid-week may provide another opportunity to have an organised activity, such as an excursion to a local attraction. There should be both adventurous and accessible options.

Fee waivers:

Depending on the level of funding or sponsorship a host has access to, fee waivers have provided opportunities to learners that otherwise might have missed out. This is an important consideration as it makes your Taming the BEAST workshop accessible to not only international participants, but also participants who may struggle financially. Fee waivers are a great drawcard and can be done partially or in full. However, they are important and tend to have positive effects.

Fee structures:

There have previously been two types of fee structures that hosts have chosen between.

Fee Structure One: Registration and meals < 300€

Fee Structure Two: Registration, accommodation and meals <1000€

Hosts are encouraged to offer Fee Structure Two for logistical reasons in more remote locations with limited accommodation options and to create a more immersive learning environment. However, Fee Structure One may be preferred in larger cities where there is an abundance of accommodation options.

Registration and assessing applicants

The registrations should be open at least six months before the commencement of the workshop and be open for a period of about one month. To maximise the registration time, start advertising via email to previous participants and relevant groups. Registrations are usually managed in-house at the hosts' institution and are expected to be accessible for online completion. For registration, the applicant is required to provide their CV as it provides useful context including how much computational experience they have and at what stage in their research they are. It is also important that participants provide a link to their homepage or group website. Along with the CV, a relevant abstract (max 500 words) outlining why they should be chosen for the workshop and what they would gain from it in relation to their research interests.

Ideally, if you have assigned 40 spaces for the workshop, you should aim for about 60 applicants, so you can select the most appropriate applicants and also allow for cancellations. Whilst the recommended application period is one month, you can adjust this accordingly to suit your desired number of minimum applicants. Also keep in mind that you may need to extend the deadline to maximise the number of applications you have. However, an unnecessarily long application period is to be avoided since applicants need to put a lot of work into writing an application and conversely organisers have to invest a lot of effort to select the most appropriate applicants.

When registrations close, you want to allow one to two weeks to assess the applications and issue the offer and decline letters. Applications should be ranked in order to maximise the participants' benefits from the workshop. Workshops hosted by the cEvo group tended to prefer applicants who made it clear that they collect data for which they need phylodynamic analysis or would move towards the field of phylodynamics, for example. Secondly, it is a worthy goal to provide an opportunity to applicants who are not in well-connected groups (i.e. not already in a leading phylodynamics group or a group that is well connected to the BEAST network) – applicants who would otherwise struggle to get help with their analyses. This goal tends to also select applicants who are interested in using BEAST on their own datasets, and not theoreticians interested in implementing their own methods. Finally, another consideration when choosing your participant pool is to ensure diversity of representation.

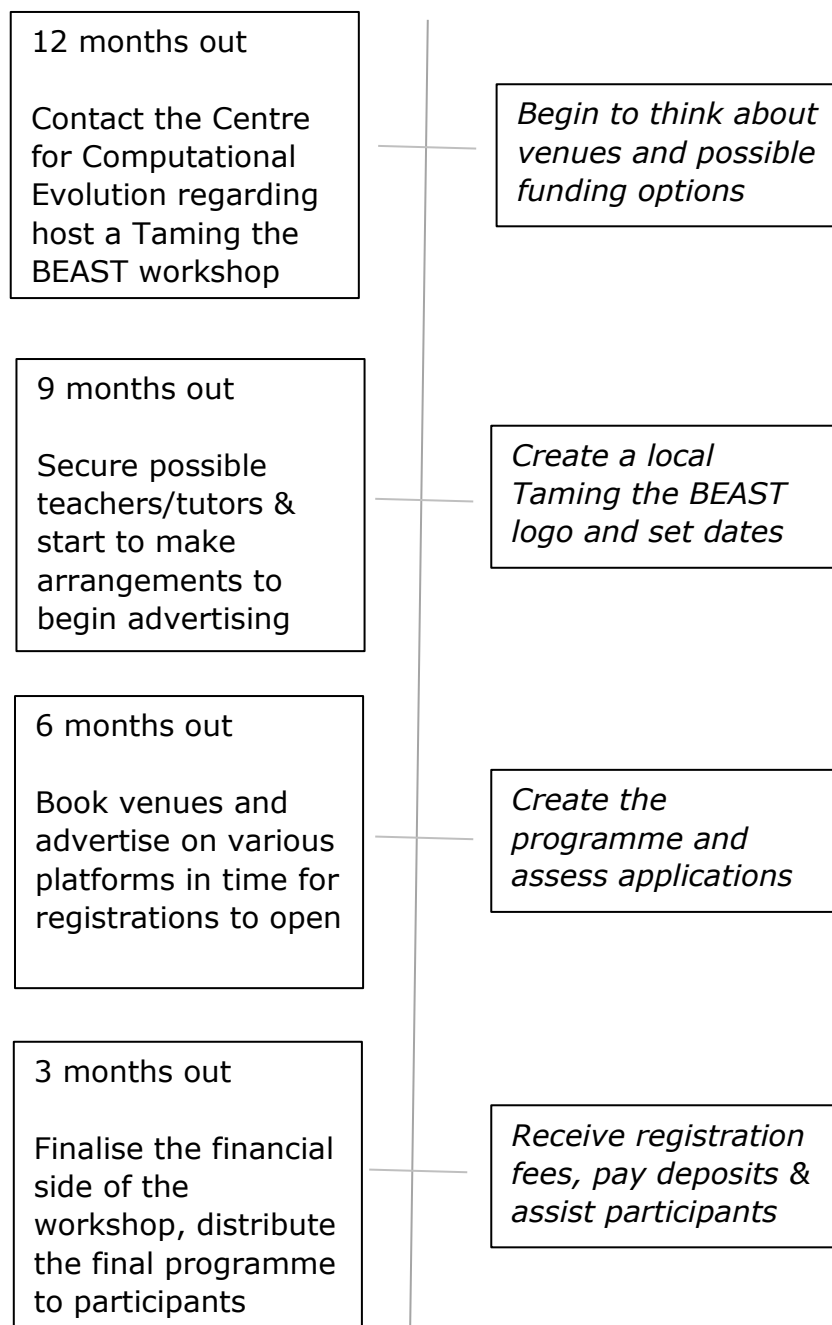
Once the participants have been decided upon, if there have been fee waivers allocated then you again rank those that would most benefit from a fee waiver (full or partial). Generally the further a participant has to travel, the greater the benefit of a fee waiver or if there are participants from a disproportionately underrepresented group and/or disadvantaged background. In Sydney for example, full waivers were offered to those travelling internationally and half waivers were offered to those traveling domestically, but from another state.

Evaluation

Previous workshops have requested participants to evaluate their experience via Google Forms, below is a draft. Since every workshop will vary, this evaluation can merely be edited to suit any particular workshop and made into CSV or PDF files for distribution or to store as the host of a workshop.

<https://docs.google.com/forms/d/1HLPKzY8C4fdyRmMCeriejaS7O8cGUGMUpMz2VdcyFs/edit#responses>

Timeline



Things to keep in mind

Workshop schedule:

To make sure that people do not lose motivation on the 1st day, at the very least make sure that everyone finishes the 1st tutorial. In addition, if you are offering free after-dinner sessions, if your participants are overworked on the first day, this will not be a positive exercise.

Daily schedule suggestions:

- Set clear start and end times for all talks and tutorials and adhere to them with timekeepers.
- Invited speaker talks are usually long (2 hours). Consider this when allocating tutorial time.
- Allow enough time for moderated panel discussions (Q&A session) with all invited speakers.

Academic program suggestions:

- Add a primer on probability theory to bring people on the same page and set the frame either as a pre-workshop or as the first thing in the workshop. However, this is generally covered in one of the first invited speaker talks. There are also two probability primers in the recommended pre-workshop reading list.
- Be detailed on introductory slides during the introduction on the first evening (model, research question, etc.), so people get an idea who they should talk to.
- Define all technical terms (e.g. Yule model), so all participants can be part of conversation.
- Always try to emphasize the similarities between macroevolution and epidemiology, since students are usually only in one discipline and transferring concepts may not come naturally to them.
- More about post-processing of results (e.g. building figures)

Materials suggestions:

Workshop organisers likely won't have to specifically write new tutorials (there is by now a substantial list of tutorials online and new tutorials are better left in the hands of package developers or the BEAST2 team). It is suggested that all tutorials in the workshop programme should be carefully checked against the latest version of BEAST2 (and updated if needed). In the past, workshops have almost always coincided with the release of a new version of BEAST2 which made checks and updates necessary.

Tutorial suggestions:

- TAs should have assigned tutorials where they can navigate the room to offer assistance and answer questions when asked. All TAs should be physically available during the first tutorial to ensure that everyone finishes it and to sort out initial teething problems – connecting to wifi, accessing tutorials, starting up BEAST2 etc.
- The room setup should be in such a way that there is easy access to every student in the tutorial room so that all students feel they can access assistance.
- It is useful to do a short introductory talk before each tutorial to introduce the problem and dataset and motivate the choices for models used. It is also useful to periodically stop the tutorial to check the progress of students and address common problems.
- A feedback session at the end of the tutorial explaining the results is also necessary. Feedback sessions should address the how and why, not just the what – instead of simply scrolling through results in Tracer, TAs should explain how to interpret the results, motivations for why specific priors were set, what the results mean etc. Several tutorials have discussion points interspersed throughout, TAs should follow up on these points during the feedback session.
- When teaching and during tutorials, what AND why needs to be explicit. Reasons for choosing priors or setting parameter values should be explicit.
- All the tutorials on the website follow roughly the same format, but since they have different authors they are not completely standardised. The most important part to standardise is the division of content between boxes and paragraphs. Boxes contain specific instructions, paragraphs contain explanations. All results in the tutorial should be replicable by just reading the boxes and following the instructions.

Free tutorials:

- Where possible, free tutorials should be available such that people can work on their own data. Be mindful to keep this away (in space and time) from social events including ones outside the workshop.
- The tutorial sessions could happen in parallel in two rooms, one room for the scheduled content tutorial and the other for a free tutorial with TAs in both rooms. This way people who are not interested in epidemiology or species evolution can pick which tutorials are most useful for them and then use the rest of the time to work on their own data before the last day.
- TAs should also be specifically assigned to these free tutorials. It may also be useful to convince some of the invited speakers to be available during some of the free sessions.

Daily announcements and feedback:

- Collect all announcements and organizational details for the day by breakfast at the latest, so they can be announced by one person before the first talk.
- TAs and speakers may find it useful to have meetings in the evening to discuss the organization and possible changes for the next day.

Sample timetable

Taming the BEAST 2018

The number and length of time slots for lectures / tutorials should be maintained accordingly. However, organizers are welcome to choose and invite the set of lecturers (time slots 9:00-11:00) and teaching assistants (time slots from 11:00). Further, with more advanced tutorials, these can in particular replace the tutorials of Thursday and Friday with other topics (see the Taming the BEAST website for the available tutorials). In the example timetable below, talks from invited speakers are in blue, standard topics and tutorials are in green and advanced topics in red.

	Sunday 2018/06/17	Monday 2018/06/18	Tuesday 2018/06/19	Wednesday 2018/06/20	Thursday 2018/06/21	Friday 2018/06/22
09:00		Alexei Drummond Introduction to Bayesian Phylogenetics (part 1)	Sebastián Duchêne Clock models, site models, model comparison	Tanja Stadler Structured Birth-death Lecture and Tutorial	Alexandra Gavryushkina Modelling Fossilization	Nicola Müller Multi-species models Tim Vaughan Modelling recombination
10:45	Morning Tea	Morning Tea	Morning Tea	Morning Tea	Morning Tea	Morning Tea
11:15		Louis Du Plessis First Steps in BEAST (lecture , tutorial)	Carsten Magnus Model Averaging and bModelTest (lecture , tutorial)	Jérémie Sciré Structured Birth-death Tutorial	Nicola Müller The Structured Coalescent Nicola Müller MASCOT Tutorial	Tim Vaughan Bacter tutorial
12:45	Lunch	Lunch	Lunch	Lunch	Lunch	Lunch
14:00		Vladimir Minin Introduction to Bayesian Phylogenetics (part 2)	Tim Vaughan Bayesian Skyline Plot and Birth-death Skyline Tutorial	Activities (e.g. Hiking, swimming, etc.)	Rachel Warnock FBD Model Tutorial	Nicola Müller StarBEAST2 Tutorial
15:45	Afternoon Tea	Afternoon Tea	Afternoon Tea		Afternoon Tea	Afternoon Tea
16:15		Veronika Boskova Setting Priors	Jūlija Pečerska Trouble-shooting		Tim Vaughan Understanding BEAST 2 XML	BEAST Clinic
16:45	Welcome Aperol	Veronika Boskova Setting Priors Tutorial	Jūlija Pečerska Trouble-shooting Tutorial		Tim Vaughan XML Hacking Tutorial	Departure
19:00	Dinner	Dinner	Dinner	Dinner	Dinner	
20:00	Introduction of Participants	BEAST Clinic (informal)	BEAST Clinic (informal)	BEAST Clinic (informal)	BEAST Clinic (informal)	

Required software

Participants are required to bring their own laptop to the workshop. Power strips are available for each participant to keep their devices charged during the workshop. The software below **must** be installed prior to arrival: *

- **Java 8 or higher:** This is required by BEAST 2 ([download](#)).
- **BEAST 2:** BEAST 2 is a cross-platform program for Bayesian phylogenetic analysis of molecular sequences. We will be using this software to perform analyses on the data brought to the workshop. We will be using **BEAST v2.5** for the workshop ([download](#)).
- **FigTree:** FigTree is a graphical viewer of phylogenetic trees, specifically designed to display summarised and annotated trees produced by BEAST ([download](#)).
- **Tracer:** Tracer is a program for analysing the trace files generated by Bayesian MCMC runs. It can be used to check mixing, ESS, parameter correlations, etc. ([download](#)).
- **BEAGLE (optional):** BEAGLE is a high-performance library that can perform the core calculations in BEAST2 (and many other phylogenetics packages) much faster. BEAGLE isn't necessary for the workshop, but it will speed up your analyses ([download](#)).
- **R:** R is a free software environment that provides a wide variety of statistical (linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, clustering) and graphical techniques, and is highly extensible ([download](#)).
- **RStudio (optional):** RStudio provides a friendly front-end for R ([download](#)).

* Email participants about the above requirement 1 – 2 weeks before commencement of workshop

Suggested reading

First chapters of the BEAST2 book are a great way to get started for both hosts and participants alike.

If it isn't found at your institution's library, then you can find it on [Google books](#).