

Overseas Conference Attendance Journal

EMBO Practical Course on Computational Molecular Evolution

by Yuki Minegishi (Leiden University)

From 3rd to 13th May, I was in Crete for the EMBO Practical Course on Computational Molecular Evolution. The course was organised by Alexandros Stamatakis (developer of RAxML), Giorgos Kotoulas of the Hellenic Center for Marine Research, Nick Goldman of EBI, Ziheng Yang (developer of PAML) and Aidan Budd of EMBL, and sponsored by the European Molecular Biology Organization.

I am currently a post-grad student at Leiden University. Before moving to the Netherlands, I was studying molecular phylogenetics and molecular groups in eels. My current research topic is fish physiology, with a focus on development, breeding and egg production.

The theme of this summer course was “The theory and practice of molecular evolution analysis”. This doesn't have any direct bearing on my current research, but as I am still currently analysing data, and I had never been to any lecture on the theory of evolutionary lineage analysis, and particularly as the lecturers were all the current leaders in the field, I decided that I could not miss this opportunity and applied for the course. I realised only later that out of 280 applicants, only 40 would be chosen to join the course, a mere 15%. I found out that I was accepted on the course around the middle of February, which made me very happy and very nervous at the same time. Since I had been in the Netherlands, I had moved away from the study of molecular phylogenetics or population genetics. The programme announced turned out to be more dense than I had thought and I went to Crete on 2nd May with some trepidation.

Unlike Holland, where it was still coat weather, Crete was brilliantly sunny and warm, with temperatures well exceeding 25 degrees. While Athens was experiencing riots and unrest, Crete remained a peaceful haven of white beaches and blue sea as the course started, on 3rd May.

As this was the first year this particular course took place, and indeed the first time that any course of this type was held in this location, its continuation in future would depend on its success, giving the whole event a certain feeling of novelty and tension.

The location was the Institute of Marine Biology and Genetics of Hellenic Centre for Marine Research, which was about a 20-minute walk from the hotel I was staying in. The programme on a typical day was: individual study (9:00 to 10:00), two lectures (10:00 to 11:00 and 12:00 to 13:00), lunch (13:00 to 15:00), two more lectures (15:00 to 16:00 and 17:00 to 18:00), with a keynote lecture from 19:00 to 20:00. Excluding the excursion day, the course lasted 11 days (see picture 1; caption: “A typical lecture. A lot of time was devoted to practical work on laptops.”).

With 40 participants, and a staff of 20 lecturers and teaching assistants, the event involved a full 60 people. As planned by the organisers, the participants came from a variety of different backgrounds. The majority were from the UK, Germany and Greece, but there were also people from France, Italy, Switzerland, Spain, Portugal, Denmark, Croatia, Belgium, Bulgaria, Turkey, India, South Africa, Kenya, Costa Rica and Canada. Around half the participants were, like myself, the only representatives of their country, so it seemed hopeless to try to count exactly how many different nationalities were present.

There was also diversity in the participants' area of research and level of experience, with everything from fresh university graduates and PhD students to senior researchers and

technicians. There were even some, particularly among the students, who had only just begun studying evolution and biology, or who were new to bioinformatics, and had little knowledge of the themes of this course.

The participants were divided equally into researchers in bioinformatics and in evolutionary biology. Among the biologists were experts on all sorts of groups, including arthropods (fossils), nematodes, insects, reptiles, birds, fish and mammals. It's not often that people from so many different backgrounds come together, and we had some very interesting discussions about our work during breaks and mealtimes.

The course was broadly divided into five themes: Interpreting Trees, Resources & Alignment, Building Trees, Coalescent Theory and Hypothesis Testing.

After a lecture on the basics of using Unix and Linux, which must have been boring for those who were already familiar with them, the course started in earnest.

The first lecture was led by Aidan Budd on Interpreting Trees. It was an introduction to phylogenetics, giving a basic overview of molecular evolution and molecular phylogeny, and defining some of the terminology, such as “node”, “branch” and “clade”. The content was quite basic and common knowledge to anyone who has studied evolution, but even at my level, I was glad of the chance to consolidate my own understanding of the topic.

Next came a lecture by Bill Pearson and Javier Herrero on Resources & Alignment. I personally found Bill Pearson's lecture to be very complicated, but the discussion of “similarity” and “homology” was very interesting. It emphasised the ambiguity inherent in the term “homology”, discussing what BLAST and other search software is really doing, how to measure sequence similarity and how to infer homology from sequence similarity.

The next lecture was about Building Trees and was given by Olivier Gascuel and John Huelsenbeck. By this time, it was clear that some participants were getting tired – they would skip breakfast and arrive only just in time for the lecture at 10:00. This lesson focused on arithmetic and statistics. Around half the participants normally used computer software and didn't have a particularly good grounding in the theory of arithmetic or statistics. To be honest, I am not very good at it myself.

Those who had specialised in bioinformatics were able to follow without difficulty, but the rest of us were really struggling not to be left behind. This was definitely the lecture which produced most questions like “Did you understand all that?” and “Explain that to me again” during our break (see picture 2; caption: “Dr Nick Goldman explaining likelihood-ratio tests”).

The next lecture topic, Coalescent Theory, was one of the themes on this course that I was most familiar with. Although I was getting very tired at this point, I was really looking forward to this lecture. It was given by Mikkel Schierup and Rasmus Nielsen.

Coalescent Theory is one of the classic theories of population genetics, but due to a recent breakthrough in computer analysis, it has become very commonly cited in articles about population genetics in the last few years. I myself have been running computer analyses for a few years, so I was very fortunate to be able to learn from the basics of “What is coalescence?”.

The last topic on the course was on Hypothesis Testing, which was taught by Nick Goldman and Ziheng Yang. The core of the lecture was on choice of models and testing of trees, and they explained what to compare with what, what it is possible to compare with what, what

constitutes a null hypothesis, the risks of running a likelihood-ratio test and the basics of statistical examination.

Due to the disruption caused by the volcano in Iceland, Jeff Thorne (developer of multivtime), who had been invited to give a keynote lecture on Estimation of Divergence Times was not able to attend the course. So on the 10th day, Ziheng Yang was forced to hurriedly prepare and give a lecture on Estimation of Divergence Times. Time was limited so this could only cover the basics, but I had been looking forward to this lecture, so I was very grateful that it was not cancelled altogether. During a quiet moment, I even asked Dr Yang to have a look at my data and give me his advice.

When doing analyses nowadays, we enter data into the program, click a few times, enter a few lines of command and out come the results, whether we understand them or not. We never consider doing the calculations ourselves, and the scale of the data is so huge nowadays that it becomes detached from reality. In many of the lectures on this course, we took relatively simple data, drew trees by hand, worked out probability by hand, and then used the program to perform simple analyses.

For instance, in John Huelsenbeck's lecture, we were all given 20-sided dice and told to calculate the probability of base substitution, based on a certain transition probability and stationary distribution. Nick Goldman had us doing likelihood ratio tests on a chi-square test distribution table.

This is the sort of thing we are all supposed to understand and most of the participants had a good theoretical grasp of it. However, by doing these calculations by hand we gained a better understanding of what the software is doing and where the difficulties lie, which will be very useful in future analyses.

When talking to one of the organisers, I heard that one of the secondary objectives of the course was to bring bioinformatics and evolutionary biology closer together. Researchers in bioinformatics and statistics are studying evolution and genealogy without ever coming into contact with living things. They analyse data from public databases and use a computer to put our (evolutionary biologists') analytical techniques into practice.

On the other end, we evolutionary biologists analyse data from our own samples and experiments. With limited data, we try to make as unbiased and accurate an estimation as possible. However, sometimes the data is not sufficient, or the calculations can become biased, or there can be too much variance in the data, making such a precise analysis impossible (at least, at this time), which is something we find hard to admit.

In contrast, bioinformatics specialists probably find it hard to understand why we don't (or can't) use ideal data and insist on using inconclusive data. In that sense, I felt that there was a gap between these two different perspectives, which was evident among the participants on the course, from our conversations about our research.

Of course, each side has different research topics and goals, so it doesn't seem absolutely necessary to me that we understand each other completely. However, having the chance to hear about the problems encountered by the other side and elucidate some points of our own through discussion in the group was extremely worthwhile.

Another remarkable thing about this course was how close the lecturers felt to the participants, which created a very pleasant atmosphere. They taught us every day, stayed in the same hotel and ate all their meals with us, so it may seem natural that we felt comfortable with each other, but I heard from someone who had taken part in other courses before that the atmosphere isn't always as nice.

Although we were all working hard from morning to evening, the lecturers and participants got together nearly every night for a drink after dinner. I have good memories of people playing billiards with John Huelsenbeck and Rasmus Nielsen, and of going out for drinks with Alexandros Stamatakis and Nick Goldman. At those times, we rarely talked about work, but would mostly just chat about our countries and the differences between them, our hobbies and other random topics.

Of course, the main focus of seminars and courses is to learn new things and study. However, I also value the social aspect of these events – meeting and talking to different people. This summer course, with its participants coming from such varied backgrounds, was especially rewarding in that respect.

On the last day, after all the lectures had finished, Dr Yang and Dr Goldman said:

Dr Yang: “If you did not understand half of the course, that's all right. We expected that.”

Dr Goldman: “...if you have understood between the half and two third of the course, we offer you a job. If you understand three quarters you get our jobs...”

Dr Yang: “In the beginning of our work we were confused. Now we are confused on a higher level.”

Although we had 11 days of lectures given by excellent teachers, I couldn't possibly claim I understood everything perfectly. But as Dr Yang and Dr Goldman said, the way forward is to try and further our understanding through getting confused, then learning a bit more, then getting confused again, learning again and so on.

This was a very intensive and informative summer course for me, which I am very glad to have taken part in. I would like to take this opportunity to thank the organisers of the course, the lecturers, who gave us such excellent teaching, my fellow participants, with whom I spent so much time studying and enjoying myself, and finally Masaki Miya, of the Natural History Museum and Institute of Chiba Prefecture, who first recommended the course to me.