European Molecular Biology Organization

Practical Course on

COMPUTATIONAL MOLECULAR EVOLUTION

IMBG-HCMR, Heraklion, Crete, Greece, 3rd - 12th Mary 2010

Organizers

Nick Goldman, European Bioinformatics Institute

Alexandros Stamatakis, Technical University of Munich

Giorgos Kotoulas, Hellenic Center for Marine Research, Heraklion

Ziheng Yang, University College London

Aidan Budd, European Molecular Biology Laboratory, Heidelberg

Local Organising Committee (IMBG)

Giorgos Kotoulas

Antonis Magoulas

Stelios Kastrinakis

Eftichia Mironaki

Pelagia Petraki

Contact

Eftichia Mironaki: e.mironaki@her.hcmr.gr

EMBO Practical Course on Computational Molecular Evolution

Introduction– Stamatakis, Kotoulas

Dear Participants,

Welcome to Crete! We are glad to welcome you to the beautiful island of Crete. We would also like to congratulate you for being accepted at our summer school, since the acceptance rate was only 15%.

Giorgos and I will provide you a quick overview over the course, and ask all students to introduce themselves within 1-2 minutes. Thereafter, we will take a tour of the Cretan Aquarium which is situated right next to the campus of IMBG-HCMR. Should you have organizational questions or require travel tips for Crete, please ask Giorgos, Alexis, or Eftihia (the IMBG secretary).

Regarding scientific questions you should try to squeeze out as much as you can from our lecturers, they are here to interact with you, so please don't be shy.

Overall, we will have a mix of theoretical lectures and practical exercises to learn the basics, as well as evening lectures that focus more on current research topics. For the practicals, there are 4 TAs here to help you: Maria, Natalie, Tim, and Clemens. They are the real soul of this course.

Overall, our days will be packed and intense. On Wednesday afternoon of the first week we will have an optional poster session and a free evening. On Sunday morning, we will make our hiking tour through the Rouvas gorge (approximately 3 hours total) and the rest of the day will be at your disposal.

Please keep in mind that Crete is located at the same latitude as parts of northern Africa (it's further to the south than Tunis) so the sun can be extremely intense and you need to protect yourself against it (I remember a central European getting a really bad sun-burn at the beginning of March). Please make sure to have enough water with you for the hiking trip.

Please let us know if you need assistance with anything or if there is a problem with the course or accommodation.

Welcome again and we hope that you will enjoy the course,

Alexis & Giorgos

Useful telephone numbers

Police: 100   Alexis (Mobile): +30 698 4052547
General Emergency: 112   Giorgos (Mobile): +30 697 6116358
Ambulance: 166   Eftixia: (Mobile) +30 697 2417288,
Fire: 199   (Office) +30 2810 337801
Forest Fire: 191   Hotel: +30 28970 41141
Coast Guard: 108   +30 28970 41142

TEACHING STAFF

FACULTY

Michail Averof - Institute of Molecular Biology and Biotechnology, Crete, Greece
Aidan Budd - European Molecular Biology Laboratory (EMBL), Heidelberg, Germany
Martin Embley - Institute of Cell and Molecular Biosciences, University of Newcastle, UK
Olivier Gascuel - Laboratoire d'Informatique, de Robotique et de Microélectronique (LIRMM), Montpellier, France
Nick Goldman - EMBL-European Bioinformatics Institute, Hinxton, UK
Javier Herrero - EMBL-European Bioinformatics Institute, Hinxton, UK
John Huelsenbeck - Department of Integrative Biology, University of California, Berkeley, USA
Carolin Kosiol - Institut für Populationsgenetik, Veterinärmedizinische Universität Wien, Vienna, Austria
Giorgos Kotoulas - Hellenic Centre for Marine Research, Marine Biology and Genetics Institute, Crete, Greece, and University of Connecticut, USA
Aoife McLysaght - Smurfit Institute of Genetics, Trinity College Dublin, Ireland
Rasmus Nielsen - Department of Integrative Biology, University of California, Berkeley, USA
Bill Pearson - Department of Computer Science, University of Virginia, USA
Antonis Rokas - Vanderbilt University, Department of Biological Sciences, Nashville, Tennessee, USA
Mikkel Schierup - Bioinformatics Research Centre, Aarhus University, Denmark
Alexandros Stamatakis - Department of Computer Science, Technishe Universitaet, Munich, Germany
Jeff Thorne - Bioinformatics Research Center, North Carolina State University, USA
Ken Wolfe - Smurfit Institute of Genetics, Trinity College Dublin, Ireland
Ziheng Yang - Department of Biology, University College London, UK

TEACHING ASSISTANTS

Maria Anisimova, Computational Biochemistry Research Group, ETH Zurich, Switzerland
Research intersts: Detecting positive selection using codon models

Clemens Lakner, School of Computational Science, Florida State University,   USA
Research intersts: Bayesian inference of phylogeny

Tim Massingham, EMBL-European Bioinformatics Institute, Hinxton, UK
Research intersts: Likelihood methods in phylogenetic analysis

Natalie Cusimano, Systematic Botany, University of Munich, Germany

Programme

Day 0 (Sunday May 2nd, 2010)

Day 0 (Sunday May 2nd, 2010)

15:30   For those who arrive early: computer trouble shooting session with Aidan Budd, Tim Massingham, and Sergio Vargas. We will meet at 15:30 at the hotel reception desk and decide ad hoc where to go.

Day 1 (Monday May 3rd, 2010)

Giorgos Kotoulas & Alexandros Stamatakis
09:00-10:15   Introduction and course overview, participants briefly introduce themselves

10:15-11:30   Free tour of the CreteAcquarium
11:30-12:00   Coffe break

Tim Massingham
12:00-13:30   Unix/Windows command line basics and exercises

13:30 -15:00 Lunch at CreteAcquarium

Aidan Budd
topics: Introducing applications of phylogenies; Introduction to phylogenetic terminology, NEWICK/PHYLIP tree format; Rooted/unrooted phylogenies; Branch lengthsGene tree-species tree conflicts (gene duplication, horizontal transfer, lineage sorting); Phylogenetic networks

15:00-18:30   Interpreting molecular phylogenetic trees (I) with coffee break

19:00-20:00   Evening lecture 1: **Ken Wolfe: "Using phylogenetics to study gene and gene family evolution"**

20:30   Dinner at hotel

Day 2 (Tuesday May 4th, 2010)

09:00 – 10:00 Free Computing & Analysis practice with TAs

Aidan Budd
10:00 - 13:30 Interpreting molecular phylogenetic trees (II) with coffee break

13:30 – 15:00 Lunch at CreteAcquarium

Bill Pearson & Javier Herrero
topics: Sequence databases (EMBL nucleotides; UniProt; others) andgenome databases (Ensembl; UCSC genome browser; plants; bacteria; others); Databases of (or at least containing) alignments and/or trees (TreeBASE, Pfam, Pandit, TreeFam; others); Searching

sequence databases (by keywords; by accession numbers; search software e.g. Blast; others);Pairwise alignment; Multiple sequence alignment (concepts, heuristics, softwares, "improvement by eye"); GBLOCKS (and others) to extract conserved blocks of alignments; Obtaining nucleotide alignments given protein alignment and associated nucleotide sequences – RevTrans (others); Struggling with formats — problems with getting your software of choice to accep your alignment of interest

15:00-18:30   Genomics resources and sequence alignment (I) with coffee break

19:00-20:00   Evening lecture 2: **Martin Embley: "The Tree of Life"**

20:30   Dinner at hotel

Day 3 (Wednesday May 5th, 2010)

08:00-14:00   Excusrion & hike through Rouvas gorge, bus will pick us up at 08:00 at the hotel. After the hike, lunch at Zaros

20:30   Dinner at hotel

Day 4 (Thursday May 6th, 2010)

09:00-10:00   Free Computing & Analysis practice with TAs

Bill Pearson & Javier Herrero
10:00-13:30   Genomics resources and sequence alignment (II) with coffee break

13:30 -15:00   Lunch at CreteAcquarium

15:00-16:30   Genomics resources and sequences alignment (II)

16:30-17:00   Coffee break

17:00-18:30   Participants can (no obligation) present research posters and talk about the posters with the lecturers/instructors

20:30   Dinner at hotel

Day 5 (Friday May 7th, 2010)

09:00 -10:00   Free Computing & Analysis practice with TAs

Olivier Gascuel & John Huelsenbeck
topics: Probabilistic models of character change/substitutions; Distance (minimum evolution, NJ, Fitch-Margoliash, etc.) and Parsimony methods; Likelihood and Bayesian methods; Tree search; Sources of error, both systematic and non-systematic, and approaches to address them: Stochastic variation, Model mis-specification (base composition; substitution

matrices; others), Long-branch attraction, Taxon-sampling; others

10:00-13:30   Phylogeny-reconstruction methods (I) with coffee break

13:30-15:00   Lunch at CreteAcquarium

15:00-18:30   Phylogeny-reconstruction methods (II) with coffee break

19:00-20:00   Evening lecture 3:   **Alexies Stamakis: "Short reads, multiple cores, and RAxML"**

20:30   Dinner at hotel

Day 6 (Saturday May 8th, 2010)

09:00-10:00   Free Computing & Analysis practice with TAs

Olivier Gascuel & John Huelsenbeck
10:00-13:30   Phylogeny-reconstruction methods (III) with coffee break

13:30 -15:00   Lunch at CreteAcquarium

15:00-18:30   Phylogeny-reconstruction methods (IV) with coffee break

19:00-20:00   Evening lecture 4: **Nick Goldman: "Alignment and PRANK"**

20:30   Dinner at hotel

Day 7 (Sunday May 9th, 2010)

09:00-10:00   Free Computing & Analysis practice with TAs

Mikkel Schierup & Rasmus Nielsen
topics: Fisher-Wright model and neutral coalescent; Population demography and subdivision; *Ancestral polymorphism and lineage sorting*

10:00-13:30   Coalescent model and inference from population data (I) with coffee break

13:30-15:00   Lunch at CreteAcquarium

15:00-18:30   Coalescent model and inference from population data (II) with coffee break

19:00-20:00   Evening lecture 5: **Michail Averof: "Evolution of development"**

20:30   Dinner at hotel

Day 8 (Monday May 10th, 2010)

09:00-10:00   Free Computing & Analysis practice with Tas

Mikkel Schierup & Rasmus Nielsen
10:00-13:30   Coalescent model and inference from population data (III) with coffee break

13:30-15:00   Lunch at CreteAcquarium

Nick Goldman & Ziheng Yang
topics: Approaches (Likelihood/Bayesian); Tests of Substitution Models (Simple models [e.g. JC vs. HKY; PAM vs. WAG], Rate heterogeneity, Clocks, Selection/adaptation, others); Test of topologies (Bootstrapping [parametric and non-parametric], Comparisons of topologies [KH-test; SOWH-test]); Confidence sets/congruency of topologies (SH-test, Bayesian methods)

15:00-18:30   Hypothesis testing in phylogenetics (I) with coffee break

19:00-20:00   Evening lecture 6: **Antonis Rokas: "Experimental Molecular Evolution"**

20:30   Dinner at hotel

Day 9 (Tuesday May 11th, 2010)

09:00-10:00   Free Computing & Analysis practice with TAs

Nick Goldman & Ziheng Yang

10:00-13:30   Hypothesis testing in phylogenetics (II) with coffee break

13:30-15:00   Lunch at CreteAcquarium

15:00-18:30   Hypothesis testing in phylogenetics (III) with coffee break

19:00-20:00   Evening lecture 7: **Carolin Kosiol: "Advances in codon models"**

20:30   Dinner at hotel

Day 10 (Wednesday May 12th 2010)

09:00-10:00   Free Computing & Analysis practice with TAs

Nick Goldman & Ziheng Yang

10:00-13:30   Hypothesis testing in phylogenetics (IV) with coffee break

13:30-15:00   Lunch at CreteAcquarium

15:00-18:30   Q&A session, students can give oral presentations if they want to, with coffee break

20:30   Dinner at hotel