

Supervisor: David S Roos, E Otis Kendal Professor of Biology – droos@upenn.edu / +1-215-898-2118
... assisted by additional biological curators, annotators, bioinformaticians and biologists from around the world

Class meeting times: flexible

Informational mtgs: Thursday 3 September, from 10-11am & 7-8pm (Philadelphia time)

<https://upenn.zoom.us/j/96128765394>

Course description

The goal of this Group Independent Study Project (Biology 399) is to exploit diverse genomic-scale datasets to determine the complete structural genome annotation (intron/exon gene architecture) for selected eukaryotic species, likely including:

- *Anopheles gambiae*, the primary mosquito species responsible for malaria transmission in Africa
- *Cryptococcus neoformans*, a fungal pathogen responsible for meningitis and other diseases
- *Toxoplasma gondii*, a protozoan parasite causing congenital neurological birth defects, and life-threatening infections associated with AIDS and other immunosuppressed conditions

Developed in order to provide hands-on independent study research experience in the context of disruptions wrought by the ongoing Covid-19 pandemic, this Biology 399 project takes advantage of diverse genomic-scale datasets that have recently been provided by research scientists around the world, allowing students to carry out independent research (not cook-book laboratory exercises) entirely *in silico*. Using a newly-released annotation platform, students will be expected to integrate and critically interpret available:

- DNA sequence data, to assess the quality of genome coverage and assembly
- RNA sequences, to define intron/exon structure and alternative transcripts, for both protein-coding and non-coding genes
- Transcription factor binding sites and epigenetic ‘marks’, to evaluate the regulation of expression
- Comparative genomics information, to determine allelic variation within populations, and cross-species analysis of protein orthologs, in an evolutionary context
- Additional information that may be available for individual species

This project involves three components: **Training**, **Production**, and **Individual Projects**

- During the **Training** phase, hands-on guided instruction (over Zoom) will familiarize students with available datasets, including provisional (draft) structural annotation, and use of the JBrowse genome browser and Apollo annotation platform to define and annotate coding & non-coding regions, splice junctions, and more. Several post-doctoral scientists, professional curators, annotators, and bioinformaticians have been recruited to facilitate training. Depending on the number of participants, we envision multiple sessions each week, allowing students to select one compatible with their schedule.
Optional lectures will provide biological and methodological context for these various datatypes, covering such topics as: biological context of the species under investigation; genome sequencing technologies; bioinformatics tools for gene prediction; RNA-sequencing methods for transcript determination; splicing mechanisms; transcription factors, chromatin marks, and other mechanisms of transcriptional regulation; RNA stability; population-level genetic variation; gene synteny and protein orthology; genome curation; genome informatics resources; etc.
- During the **Production** phase, students will work individually to apply their newfound annotation skills to systematically annotate structural genes for one particular species of interest. Small groups of those working on that species will meet regularly to discuss and resolve challenges in data interpretation, and divide up the work required, with the goal of completing structural annotation of at least one genome, in preparation for publication.
- For their **Individual Projects**, students will define, write-up, and defend a research proposal based on observations they have made during the training and production phases of this project, possibly guided by topics covered in the optional research lectures. There is broad scope for project development, in consultation with course facilitators. Projects might include such topics as:

- functional predictions based on the observation of treatment- or sample-specific gene expression, population-level allelic diversity, etc
- genome-wide analyses, such as the identification of putative translational initiation sites (a hard problem); those with computational skills may want to code up automated methods to automatically define non-coding regions and ncRNAs (another hard problem), regulatory regions, etc.
- identification of novel pathways or therapeutic targets based on cross-species comparisons of protein orthologs, and compilation of phenotypic data available on-line or from the literature

It will likely be possible to take promising proposals forward, working in collaboration with active research groups around the world, to test in the laboratory.

FAQs:

Q: Does this course satisfy departmental and other requirements for independent study research?

A: Yes

Q: What are the prerequisites for this course? Is wet-lab experience required? Is programming experience required?

A: *Permission of the instructor is required. Participants must have a strong grasp of Introductory Biology-level molecular biology concepts. Completion of (or simultaneously enrollment in) a relevant 200-level course, such as Biology 202 or 221, is strongly recommended. No wet-lab or computational experience is required.*

Q: What time commitments are expected?

A: *As with any independent study research, what you get out of this experience is likely to be proportional to the effort you put in. You should plan to devote an absolute minimum of 10 hr/week to your research project.*

Q: Does this independent study course provide face-to-face training and wet-lab experience?

A: *This course will be offered entirely on-line, with no on-campus in-lab activities ... but it will involve extensive face-to-face research training via Zoom, both individually and in small groups.*

Q: When will course training be provided? Can I participate from other time zones?

A: *This course will include multiple supervisors, from various sites in the US and Europe, and we envision multiple sessions at different times of day and night, providing options for participants from around the world. Optional lectures will be recorded, allowing asynchronous participation.*

Q: Is it realistic to envision that this group independent study project will result in a scientific publication?

A: Yes.

Q: Is it possible to extend my research into the Spring semester, enrolling in Biology 499? Will it be possible to add an on-campus 'wet lab' component in the Spring semester, perhaps testing hypotheses developed during the Independent Project phase of this work?

A: *Yes, this is a possibility, depending on your performance in Biology 399. Wet-lab research may also be possible, depending on progress of the Covid pandemic.*

Q: How can I get answers to additional questions about this independent study course?

A: *Information sessions will be held from 10-11am & 7-8 pm on Thursday Sept 3 (Philadelphia time); questions may also be emailed to droos@upenn.edu.*