SEA-PHAGES CACAO Q&A

What is CACAO?
Jim Hu (Texas A&M University) has been hosting CACAO for a number of years as part of his GONUTS project. Through the GONUTS wiki, Jim has set up a competition in which students try to generate GO annotations. The competition is divided in innings and students get to make annotations, challenge those made by rivals, revise them, etc. This is all judged by experts and students get credit for correct annotations and challenges.

Wait, what is GO and what are GO annotations?
GO is the Gene Ontology, an ontology devoted to the annotation of gene product molecular functions, roles in biological processes and locations in/outside the cell.

GO provides the barebones structure to make statements about gene products. It defines terms like GO:0034335 (DNA supercoiling activity). A GO annotation is a statement made on a gene product using the GO formalism. For instance, Mycobacterium smegmatis strain ATCC 700084 DNA gyrase subunit B (A0QNE0) has GO:0034335 DNA supercoiling activity. Crucially, a GO annotation needs a reference for this statement. In this case PMID:8878580. And it also needs to state what sort of evidence did the authors of PMID:8878580 bring forward to make that claim. In this case, it was “Inferred from Direct Assay” (IDA).

How would my students benefit from performing GO annotations in the context of CACAO?
Well, in many ways. Participating in CACAO forces students to read papers and interpret them, extract the relevant information and put it formally in the proper context in terms of the generated knowledge. These are all essential skills in today’s biology and in confronting higher-level paper-based courses. This is all done in a wiki environment, so it is public and that makes them take responsibility. Furthermore, they get challenged by other students, not by their instructors, which helps to motivate them. And, of course, if they make successful annotations these will be submitted to GO, so they are directly contributing to the advancement of science (albeit in a less spectacular fashion than sequencing a phage, granted).

What can be done in SEA-PHAGES CACAO?
GO annotations require identifiable gene products, typically in the form of UniProt protein identifiers. That means that the phage you are currently annotating cannot be used in this year’s CACAO competition.

That said, you can still annotate your own genomes from past years (which are in GenBank and hence have UniProt accession numbers), or you can have students annotate any phage proteins they like. Again, this is a course on phage genome annotation, and GO annotation is an increasingly important part of the annotation process. Exposing your students to the GO annotation process (regardless of the phage used) is formative in itself.
How can I annotate my own phage genome (from last year), if GO annotations need a PubMed reference as evidence for the annotation and nobody has done (yet) any experimental work on my phage?

Well, there is always a way. GO annotations accept other references beyond PubMed identifiers. In particular, they accept “GO references” (GO_REFs). GO_REFs are methodological descriptions showing the process by which curators arrive at specific conclusions about a gene that allow them to make a sound annotation. For instance, the GOA team at EBI/EMBL set up a well-defined protocol to automatically detect orthologs and decided to use it conservatively to transfer experimental GO annotations from the original annotated protein to its orthologs. GO annotations made in this manner by the GOA team use GO_REF:0000019 as reference, together with an “Inferred from Electronic Annotation” (IEA) evidence code.

To enable GO annotations of SEA-PHAGES GO genomes, I requested the creation of a GO_REF (GO_REF:0000100, “Gene Ontology annotation by SEA-PHAGES biocurators”) that essentially subsumes the criteria for the in silico methods established in the SEA-PHAGES annotation manual. Our students can therefore use this GO_REF to transfer GO annotations from experimentally annotated proteins to proteins in your phage of interest.

What should I have students do: annotate my own phage genome or just phage proteins at large?

Well, that depends on how much overhead you want from this activity. The easiest way to go is to have students annotate phage proteins from any phage for which protein sequences are available in UniProt. That will facilitate their work, especially if they decide for phage lambda or T4, where there is the largest abundance of published literature (which still needs to be transferred to GO!).

Annotating one of your past phage genomes will likely entice your students. It will, however, also make things quite a bit more complex. To fully understand why, please read the attached primer on GO annotations (Phage_Hunters_GO_annotation_guide.pdf). The bottom line is that “transfer annotations” (where you use GO_REF: 0000100 to state that your phage gene product has some function/role/location) require that you first establish orthology between your gene product with a gene product that (a) already has a “regular” GO annotation or (b) you can make a “regular” GO annotation for (i.e. there are available literature sources). This often amounts to an elaborate dance through BLAST/HHPred results, trying to quickly assess whether any result above significance level has a GO annotation or might have enough supporting literature to make one. Some students excel at it, others don’t.

Based on our experience at UMBC last year, my advice would be to allow (and encourage if you will) “own phage” annotations, but not to require them.

What should I expect?
Trouble, of course. Where would the fun be otherwise? Many students have trouble digesting the formalism of GO. Expect repeated questions about what an evidence code is, what a reference is, why do we need this and that, etc. Once students get the gist, they are good to GO, but it may take a while for some of them.

This is not a time-free activity. Students may take between 15 to 60 min to complete one GO annotation (maybe two if they are performing a regular+transfer annotation). So this can take a substantial amount of time of their computer lab work. We will schedule the CACAO to start mid-semester, so that most of the raw genome annotation on your phage has already been done by that time. In any case, consider what a realistic target is for your students and adapt your syllabus accordingly (e.g. require 2 annotations + 1 challenge per team).

This is not a time-free activity on your end either. You will need to contribute verifying your and other student annotations, coaching your students, helping them with questions, etc. No free lunch!

**What is the minimal setup?**
You will need to provide some background on GO (and the concept of ontology, and its usefulness), and possibly go through the attached Phage_Hunters.GO_annotation_guide.pdf with your students (accompanying slides and further reference material will be provided). You will also need to let us know of your intention to participate and the particulars of your institution setup.

**How do I fit this in my syllabus?**
This is really where CACAO excels. CACAO was created as an inter-campus competition, but has the ability to split participating campuses into teams. Yet the annotations made by a student are recorded with their ID (and assigned to the student’s team). This means that you can evaluate CACAO at any level you want and deploy it as a class activity in any form you want (although your campus reputation might resent from some!). Here are some suggestions:

- **Vanilla CACAO**
  If you are using team-based learning in your class (or any other group-based teaching methodology), just use the same teams for CACAO and then integrate the CACAO score into your team grading scheme.

- **Single shot CACAO**
  If you are not using a group-based learning approach in your class, CACAO might provide the opportunity to test the waters. You can assign groups in class and then evaluate individually, or add in an extra-credit factor for the team score. Again, a team activity already set up and ideal for experimenting.

- **Extra sugar CACAO**
  Not sure if CACAO will work for you? Use it as an opt-in (grade enhancer) solo activity. Register your whole class as a single team and track, at the end of the semester, which
students have done what. Use this to decide if you give partial or total extra credit to each student. The downside of this approach is that you still have to provide some background and training and do so in a more personalized manner.