Undergraduate Virtual Engagement in Community Genome Annotation Provides Flexibility to Overcome Course Disruptions

Surya Saha,a Teresa D. Shippy,b Susan J. Brown,b Joshua B. Benoit,c and Tom D’Eliad

aBoyce Thompson Institute, Ithaca, NY 14853, USA, and Animal and Comparative Biomedical Sciences, University of Arizona, Tucson, AZ 85721, USA
bDivision of Biology, Kansas State University, Manhattan, Kansas, 66506, USA
cDepartment of Biological Sciences, University of Cincinnati, Cincinnati, Ohio, 45221, USA
dBiology Department, Indian River State College, Fort Pierce, Florida, 34981, USA

INTRODUCTION

Disruptions due to natural disasters and disease outbreaks have created new challenges for synchronous and in-person learning at all levels of teaching. Students and instructors struggle with a transition to remote learning due to technological issues, lack of personal interaction, and the inability to continue scientific projects. We have implemented community genome annotation as an online and asynchronous learning model that increases interactivity among students and between students and instructors, through scientific research. Moreover, this approach reinforces general biology concepts, provides computational literacy, and enables undergraduate students to directly contribute to active and publishable research projects. Our multi-institution project (Appendix 1) (1) was based on a hybrid remote and in-person model until early 2020, when we successfully transitioned to fully remote instruction. We continued online instruction throughout 2020 (summer and fall).

We present community genome annotation as a model of instruction that can be integrated into the curriculum as a module in a biology course, as an entire course, or as a directed research project. Genome annotation projects focus on manually improving the structural and functional annotation of genes. Biology curricula that use genome annotation have been shown to increase student comprehension of foundational concepts related to genetics (2–4). Large-scale programs that use genome annotation as part of the course content have proven successful (5, 6), contributing to the evidence that annotation increases understanding of genetics (5, 7–9). Course-based implementation of genome annotation also aligns the course curriculum with recommended guidelines for bioinformatics and evolutionary biology (10, 11). Here, we present tips for developing an instructional platform for genome annotation that can seamlessly transition to fully online instruction when necessary. We also provide an overview of tools that facilitate instruction and educational outcomes that can be expected.

PROCEDURE

Organization and technological tools

The onset of remote learning presented many challenges to our annotation workflow. The following guide provides tips to overcoming these issues (Fig. 1). We use predominantly open access tools for our community annotation instruction model. Apollo (12) is a widely used technological tool for gene curation, which has been implemented at Citrusgreening.org (13), along with BLAST (14) for sequence comparisons. In case a web server is not available, there are a number of public installations of Apollo for manual curation (15). Apollo provides access to genomes, with associated evidence tracks for curating gene models. Videoconferencing (e.g., Zoom or Skype) is used for virtual meetings. Free resources such as Google Drive are used for sharing data, project documentation, and training materials.

The following steps outline the organizational process we follow in our annotation workflow. The details of the activities for meetings and outcomes are described in Appendix 2.

1. Annotation Community: Entire annotation community meets biweekly on videoconferences (research scientists, instructional faculty, students).
2. Local Annotation Team: Local annotation groups meet weekly via videoconference (faculty, student coordinator, student annotators).
3. Orientation and Training: If possible, local
videoconferences are supplemented with in-person meetings for orientation of a new cohort of students and troubleshooting challenging annotation problems. Students attend i5k training webinars (https://i5k.nal.usda.gov/news) to get additional exposure to complicated curation issues.

4. Annotation: Students use aforementioned web-based tools to annotate gene models based on transcript and ortholog evidence.

5. Reporting: Each student writes a detailed gene report describing the identified gene models once the genes are curated.

This annotation workflow reinforces fundamental concepts, as students apply principles learned in their courses to understanding the genomic architecture and evolution of a gene family.

**Teaching tools and learning goals**

As students begin the annotation process, they complete an online instructional worksheet (Appendix 3) that demonstrates they understand key concepts (central dogma). Students who complete the worksheets should be able to compare protein sequences from different organisms, evaluate RNA-Seq data relative to exon structure, interpret BLAST results and assess the quality of their gene model based on their data. A list of online resources frequently used during annotation (Appendix 4) is provided to students.

Student annotation efforts are supported and monitored closely by the organizers. Progress reports (Appendix 5) provide “real-time” updates to the instructor. Self-reflection by the students is used to assess their progress and report their strengths and weaknesses. Students prepare presentation updates for local and community virtual meetings. These updates inform the instructor of support that may be required, either through the online course, through virtual peer mentoring, or from community members.

These assessments help students determine whether modifications to gene models are correct, evaluate scientific output in general, and develop professional competencies.

As students finalize their annotations, they prepare a number of reports, including posters (Appendix 6), gene reports (supplementary notes in 16), and oral presentations, and contribute to formal reports for publication (16–22). Students also submit abstracts to virtual conferences, which, during the current pandemic, have reduced or removed registration fees and require no travel. Student final reports are assessed for their ability to synthesize conclusions from acquired data and to demonstrate presentation skills, both of which contribute to their professional development.

**CONCLUSION**

Our community genome annotation program presents an instructional model that is inherently adaptive for transition between in-person and online instruction. Student and faculty participants were introduced to online tools to annotate genes, share data, and attend meetings within and between institutions. When these tools became the primary means for remote instruction following campus closures, both the faculty and students were prepared to use these virtual instructional technologies. As a result, after the onset of the COVID-19 pandemic, students were able to seamlessly continue their annotation projects, interact with the research team, and prepare abstracts and posters for virtual conferences. Furthermore, this model allows transition to online instruction that will not impose any additional workload on students, and the peer-mentoring activities help in onboarding new students. Online gene annotation also provides student research opportunities at a time when many labs are unable to accommodate undergraduate researchers due to COVID-19–related restrictions. The

<table>
<thead>
<tr>
<th>Challenges</th>
<th>Solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hard to recruit students when they are off campus</td>
<td>Apollo, NCBI, OrthoDB, Student access to internet and a computer, Peer mentoring</td>
</tr>
<tr>
<td>Access to computer labs limited</td>
<td>Google Drive, Online Lab notebook, Virtual lab partners for peer review</td>
</tr>
<tr>
<td>Software and local data storage on student computers limits collaborative annotation</td>
<td>One-on-one online meetings with faculty, Video conferences, PowerPoint Presentations, Local and Community level reporting, Prepare gene reports, posters, preprints and manuscripts, Attend virtual scientific conferences</td>
</tr>
</tbody>
</table>

FIG 1. Workflow for students to learn and complete genome annotation during remote learning. Challenges presented during the COVID-19 pandemic are presented, along with solutions to these problems to continue community annotation.
The community aspect of the project provided a virtual network among student peers and faculty providing social–emotional support and normalcy during this disruptive time. The virtual network developed during the pandemic can translate into internship and in-person research opportunities for undergraduate students in the future.

SUPPLEMENTAL MATERIALS

Appendix 1: Table of participating institutions and courses
Appendix 2: Meeting details and outcomes
Appendix 3: Genome annotation instructional guide worksheet for students
Appendix 4: List of annotation resources and links
Appendix 5: Student progress report
Appendix 6: Example student genome annotation poster

ACKNOWLEDGMENTS

We thank Will Tank and participating students for assistance with supplemental files and Kascha Johnson for the organization of community annotation. This work is supported by Hispanic-Serving Institutions Education Grants Program grant #2020-38422-32252 from the USDA National Institute of Food and Agriculture (NIFA) awarded to D’Elia and Saha. The student annotators and postdocs involved in the community annotation project were also funded through USDA-NIFA grant 2015-70016-23028. We declare there are no conflicts of interest.

REFERENCES


