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

Surya Saha[1], Teresa D. Shippy[2], Susan J. Brown[2], Joshua B. Benoit[3], and Tom D'Elia[4]*

[1]*Boyce Thompson Institute, Ithaca, NY 14853, and Animal and Comparative Biomedical Sciences, University of Arizona, Tucson, AZ 85721;*
[2]*Division of Biology, Kansas State University, Manhattan, KS 66506;*
[3]*Department of Biological Sciences, University of Cincinnati, Cincinnati, OH 45221;*
[4]*Biology Department, Indian River State College, Fort Pierce, FL 34981*

Recently, students and faculty have been forced to deal with unprecedented disruptions to their courses and broader uncertainties that have presented serious challenges to quality instruction. We present a flexible, team-based approach to teaching and learning that can transition seamlessly between face-to-face, hybrid, and fully online instruction when disruptions occur. We have built a community genome annotation program that can be implemented as a module in a biology course, as an entire course, or as directed research projects. This approach maintains an engaging and supportive educational environment and provides students the opportunity to learn and contribute to science with undergraduate research. Students are provided guidance through multiple interactions with faculty and peer mentors to support their progress and encourage learning. Integration of the developed instructional tools with available technology ensures that students can contribute remotely. Through this process, students seamlessly continue their annotation coursework, participate in undergraduate research, and prepare abstracts and posters for virtual conferences. Importantly, this strategy does not impose any additional burden or workload on students, who may already be overwhelmed with the additional work associated with the transition to remote learning. Here, we present tips for implementing this instructional platform, provide an overview of tools that facilitate instruction, and discuss expected educational outcomes.

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 stu-

dents 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
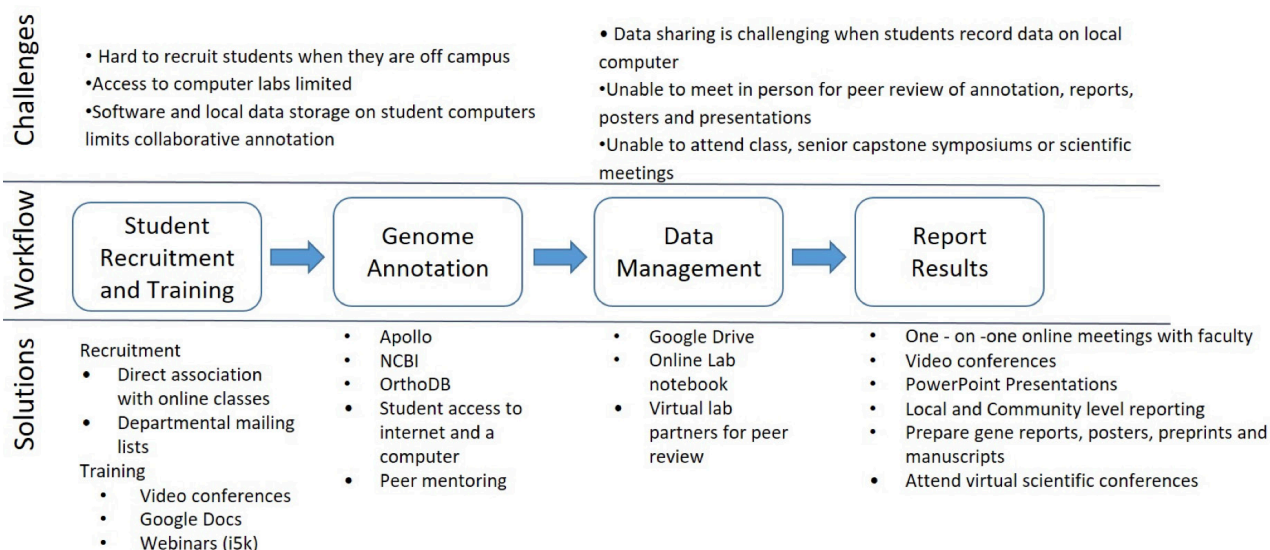

FIGURE 1. Workflow for students to learn and complete genome annotation during remote learning. Challenges presented during the COVID-19 pandemic are presented, along with solutions to these problems to continue community annotation.

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 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

REFERENCES

1. Hosmani PS, Shippy T, Miller S, Benoit JB, Munoz-Torres M, Flores-Gonzalez M, Mueller LA, Wiersma-Koch H, D'Elia T, Brown SJ, Saha S. 2019. A quick guide for student-driven community genome annotation. PLOS Comput Biol 15(4):e1006682.

2. Beagley CT. 2013. Genome annotation in a community college cell biology lab. Biochem Mol Biol Educ 41:44–49.

3. Laakso MM, Paliulis LV, Croonquist P, Derr B, Gracheva E, Hauser C, Howell C, Jones CJ, Kagey JD, Kennell J, Silver Key SC, Mistry H, Robic S, Sanford J, Santisteban M, Small C, Spokony R, Stamm J, Van Str, M, Leung W, Elgin SCR. 2017. An undergraduate bioinformatics curriculum that teaches eukaryotic gene structure. CourseSource. doi: 10.24918/cs.2017.13.

4. Pogoda CS, Keepers KG, Stanley JT, Kane NC. 2019. A CURE-based approach to teaching genomics using mitochondrial genomes. CourseSource. doi:10.24918/cs.2019.33.

5. Shaffer CD, Alvarez C, Bailey C, Barnard D, Bhalla S, Chandrasekaran C, Chandrasekaran V, Chung HM, Dorer DR, Du C, Eckdahl TT, Poet JL, Frohlich D, Goodman AL, Gosser Y, Hauser C, Hoopes LL, Johnson D, Jones CJ, Kaehler M, Kokan N, Kopp OR, Kuleck GA, McNeil G, Moss R, Myka JL, Nagengast A, Morris R, Overvoorde PJ, Shoop E, Parrish S, Reed K, Regisford EG, Revie D, Rosenwald AG, Saville K, Schroeder S, Shaw M, Skuse G, Smith C, Smith M, Spana EP, Spratt M, Stamm J, Thompson JS, Wawersik M, Wilson BA, Youngblom J, Leung W, Buhler J, Mardis ER, Lopatto D, Elgin SC. 2010. The genomics education partnership: successful integration of research into laboratory classes at a diverse group of undergraduate institutions. CBE Life Sci Educ 9(1):55–69.

6. Jordan TC, Burnett SH, Carson S, Caruso SM, Clase K, DeJong RJ, Dennehy JJ, Denver DR, Dunbar D, Elgin SCR, Findley AM, Gissendanner CR, Golebiewska UP, Guild N, Hartzog GA, Grillo WH, Hollowell GP, Hughes LE, Johnson A, King RA, Lewis LO, Li W, Rosenzweig F, Rubin MR, Saha MS, Sandoz J, Shaffer CD, Taylor B, Temple L, Vazquez E, Ware VC, Barker LP, Bradley KW, Jacobs-Sera D, Pope WH, Russell DA, Cresawn SG, Lopatto D, Bailey CP, Hatfull GF. 2014. A broadly implementable research course in phage discovery and genomics for first-year undergraduate students. MBio 5:e01051–13. doi:10.1128/mBio.01051-13.

7. Buckner B, Beck J, Browning K, Fritz A, Grantham L, Hoxha E, Kamvar Z, Lough A, Nikolova O, Schnable PS, Scanlon MJ, Janick-Buckner D. 2007. Involving undergraduates in the annotation and analysis of global gene expression studies: creation of a maize shoot apical meristem expression database. Genetics 176(2):741–747.

8. Mitchell CS, Cates A, Kim RB, Hollinger SK. 2015. Undergraduate biocuration: developing tomorrow's researchers while mining today's data. J Undergrad Neurosci Educ 14:A56–65.

9. Pope WH, Bowman CA, Russell DA, Jacobs-Sera D, Asai DJ, Cresawn SG, Jacobs WR, Hendrix RW, Lawrence JG, Hatfull GF; Science Education Alliance Phage Hunters Advancing Genomics and Evolutionary Science, Phage Hunters Integrating Research and Education; Mycobacterial Genetics Course. 2015. Whole genome comparison of a large collection of mycobacteriophages reveals a continuum of phage genetic diversity. Elife 4:e06416.

10. Merkel S, Reynolds J, Hung K, Siegesmund A, Smith A, Smith H. 2012. ASM curriculum guidelines for undergraduate microbiology education. http://www.asm.org/index.php/guidelines/curriculum-guidelines.

11. Welch L, Lewitter F, Schwartz R, Brooksbank C, Radivojac P, Gaeta B, Schneider MV. 2014. Bioinformatics curriculum guidelines: toward a definition of core competencies. PLOS Comput Biol 10(3):e1003496.

12. Lee E, Helt GA, Reese JT, Munoz-Torres MC, Childers CP, Buels RM, Stein L, Holmes IH, Elsik CG, Lewis SE. 2013. Web Apollo: a web-based genomic annotation editing platform. Genome Biol 14(8):R93.

13. Flores-Gonzalez M, Hosmani PS, Fernandez-Pozo N, Mann M, Humann JL, Main D, Heck M, Brown S, Mueller LA, Saha S. 2019. Citrusgreening.org: an open access and integrated systems biology portal for the Huanglongbing (HLB) disease complex. bioRxiv 868364. doi: 10.1101/868364

14. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215(3):403–410.

15. Dunn NA, Unni DR, Diesh C, Munoz-Torres M, Harris NL, Yao E, Rasche H, Holmes IH, Elsik CG, Lewis SE. 2019. Apollo: democratizing genome annotation. PLOS Comput Bio 15(2):e1006790.

16. Saha S, Hosmani PS, Villalobos-Ayala K, Miller S, Shippy T, Flores M, Rosendale A, Cordola C, Bell T, Mann H, DeAvila G, DeAvila D, Moore Z, Buller K, Ciolkevich K, Nandyal S, Mahoney R, Van Voorhis J, Dunlevy M, Farrow D, Hunter D, Morgan T, Shore K, Guzman V, Izsak A, Dixon DE, Cridge A, Cano L, Cao X, Jiang H, Leng N, Johnson S, Cantarel BL, Richards S, English A, Shatters RG, Childers C, Chen MJ, Hunter W, Cilia M, Mueller LA, Munoz-Torres M, Nelson D, Poelchau MF, Benoit JB, Wiersma-Koch H, D'Elia T, Brown SJ. 2017. Improved annotation of the insect vector of citrus greening disease: biocuration by a diverse genomics community. Database bax032. doi:10.1093/database/bax032.

17. Hosmani PS, Flores-Gonzalez M, Shippy T, Vosburg C, Massimino C, Tank W, Reynolds M, Tamayo B, Miller S, Norus J, Kercher K, Grace B, Jernigan M, Harper D, Adkins S, DeLaFlor Y, Paris T, Vandervoort S, Adams R, Norman S, Ventura J, Perry M, Weirauch M, Benoit J, Hunter WB, Wiersma-Koch H, D'Elia T, Brown S, Mueller LA, Saha S. 2019. Chromosomal length reference assembly for *Diaphorina citri* using single-molecule sequencing and Hi-C proximity ligation with manually curated genes in developmental, structural and immune pathways. bioRxiv 869685. doi:10.1101/869685

18. Vosburg C, Reynolds M, Noel R, Shippy T, Hosmani PS, Flores-Gonzalez M, Mueller LA, Hunter WB, Brown SJ, D'Elia T, Saha S. 2020. Characterization of Wnt signaling genes in *Diaphorina citri*. 2020. bioRxiv 2020.09.21.306100. doi:10.1101/2020.09.21.306100

19. Miller S, Tamayo B, Shippy TD, Hosmani PS, Flores-Gonzalez M, Mueller LA, Hunter WB, Brown SJ, D'Elia T, Saha S. 2020. Chitin biosynthesis genes in *Diaphorina citri*. bioRxiv 2020.09.22.30921. doi:10.1101/2020.09.22.309211

20. Massimino C, Vosburg C, Shippy T, Hosmani PS, Flores-Gonzalez M, Mueller LA, Hunter WB, Benoit JB, Brown SJ, D'Elia, Saha S. 2020. Annotation and analysis of yellow genes in *Diaphorina citri*, vector for the Huanglongbing disease. bioRxiv doi:10.1101/2020.12.22.422960

21. Miller S, Shippy T, Tamayo B, Hosmani PS, Flores-Gonzalez M, Mueller LA, Hunter WB, Benoit JB, Brown SJ, D'Elia, Saha S. 2020. Characterization of chitin deacetylase genes in the *Diaphorina citri* genome. bioRxiv doi:10.1101/2020.12.22.424074

22. Miller S, Shippy T, Hosmani PS, Flores-Gonzalez M, Mueller LA, Hunter WB, Benoit JB, Brown SJ, D'Elia T, Saha S. 2020. Segmentation pathway genes in the Asian citrus psyllid, *Diaphorina citri*. bioRxiv doi:10.1101/2020.12.24.424320