Advances in Biology Laboratory Education
Publication of the Association for Biology Laboratory Education
Volume 43, Article 17, 2023

Using genome annotation projects to teach eukaryotic gene structure and to engage students in genomics research

Maria S. Santisteban[1], James Godde[2], Anya L. Goodman[3], Charles R. Hauser[4], Don Paetkau[5], Catherine Reinke[6], Wilson Leung[7], Cindy Arrigo[8], Nathan T. Mortimer[9], and Laura K. Reed[10]

[1] University of North Carolina at Pembroke, Department of Biology, PO Box 1510, Pembroke NC 28352 USA
[2] Monmouth College, 700 E Broadway, Monmouth, IL 61462
[3] California Polytechnic State University, Chemistry and Biochemistry Dept., San Luis Obispo CA 93407 USA
[4] St. Edward's University, School of Natural Science, 3001 South Congress, Austin, TX 78704-6489 USA
[5] Saint Mary's College, Notre Dame, IN 46556
[6] Linfield College, Biology Department, 900 SE Baker Street, McMinnville OR 97128 USA
[7] Washington University in St. Louis, Department of Biology, Campus Box 1137, One Brookings Drive, St. Louis MO 63130-4899 USA
[8] New Jersey City University, Department of Biology, 2039 Kennedy Blvd., Jersey City, NJ 07305
[9] Oregon State University, Department of Biochemistry & Biophysics, Linus Pauling Institute, Corvallis, OR 97331 USA
[10] University of Alabama Tuscaloosa, Department of Biological Science, Box 870344, Tuscaloosa, AL 35487-0344 USA
(maria.santisteban@uncp.edu; jgodde@monm.edu; agoodman@calpoly.edu; charlesh@stedwards.edu; dpaetkau@saintmarys.edu; creinke@linfield.edu; wleung@wustl.edu; carrigo@njcu.edu; nathan.mortimer@oregonstate.edu; lreed1@ua.edu)

The Genomics Education Partnership (GEP; https://thegep.org) began as a consortium of 16 faculty in 2006 with a goal of providing students with Course-based Undergraduate Research Experiences (CUREs) in genomics. As of fall 2021, GEP has over 200 faculty from more than 180 institutions and engages more than 3,900 undergraduates in authentic genomics research annually. These faculty joined and continued to participate in the GEP for many reasons, including the collaborative nature of the research, the well-established infrastructure, and the supportive network of like-minded colleagues. Faculty implement GEP materials in diverse settings — ranging from short modules (2–8 weeks) within a course, to a standalone full-semester course, to independent student research. GEP students show significant gains in scientific knowledge and attitudes toward science. In addition to improving their understanding of the research process and how new knowledge is created in the field, GEP students acquire desirable and transferable skills essential for future participation in the workforce, such as problem solving, independence, application of knowledge, teamwork, and collaboration. Students also gain competence in using computational algorithms to analyze large biological datasets — there by preparing students for a

growing need of a workforce trained at applying statistics and computational tools to analyze large datasets. In addition, GEP students and their faculty mentors are eligible to be co-authors on the scientific publications that are based on their work. In this workshop, we provided an overview of the GEP community, a hands-on guided tour of our introductory curriculum aimed to teach gene structure, transcription, translation, and processing, and a step-by-step walkthrough that illustrates the protocol for annotating a protein-coding gene in *Drosophila*. Participants received information on how to join the GEP community and received training and resources to enable their implementations.

Keywords: Genomics Education Partnership (GEP), bioinformatics, undergraduate research, eukaryotic gene structure, gene annotation, genome browser, mRNA processing, transcription, translation

Introduction

GEP's mission and trajectory

The Genomics Education Partnership (GEP; https://thegep.org; Elgin et al. 2017) began as a consortium of 16 faculty in 2006 with a goal of providing students with Course-based Undergraduate Research Experiences (CUREs) in genomics. This CURE teaches students how to analyze data to annotate genes and genomic regions; the knowledge generated goes to support comparative genomics investigations. As of Fall 2021, GEP has over 200 faculty and engages more than 3,900 undergraduates in authentic genomics research annually (Figure 1a; https://thegep.org/about/directory/).

Compared to wet lab research, genomics has several advantages that make it ideal for engaging large numbers of undergraduate students in research during the academic year. Bringing genomics research into the classroom is cost effective because only computers and access to the Internet are required. Large genomics datasets are publicly available, and students can use freely available bioinformatics tools and databases (e.g., genome browsers, NCBI BLAST) to analyze these datasets. Due to the lack of safety issues, restricting access to laboratory installations is not a concern. Students can quickly acquire the skills needed to help each other, which results in peer teaching and enables faculty to support larger numbers of students. Because mistakes are relatively inexpensive, students can iteratively attempt different approaches to solve a challenging research problem, thereby becoming more familiar with the research process and gaining deeper insights into the research problem. Genomics also introduces students to the use of computational algorithms to analyze large biomedical datasets — preparing students for a growing need of a workforce trained at applying statistics (e.g., machine learning) and computational tools to analyze large datasets. Many students present their work in local, regional, or national conferences. Students and their faculty mentors are eligible to be co-authors on the scientific publications that are based on their work.

Faculty joined and continued to participate in the GEP for many reasons, including the collaborative nature of the research, the well-established infrastructure, and the supportive network of like-minded colleagues (Shaffer et al. 2010; Lopatto et al. 2014). Developing and maintaining quality CUREs that challenge and engage students is demanding and requires a great deal of dedication and expertise. The GEP community meets these challenges by hosting online and in-person training workshops for new members to prepare them to teach CUREs, and by leveraging resources among participating institutions, resources that are not available outside R1 institutions (e.g. McDonnell Genome Institute at Washington University), and by providing up-to-date curriculum materials as well as technical and pedagogical support during the academic year. GEP also hosts an annual Alumni Workshop where GEP faculty can come together and share best practices, collaboratively develop new curricula and tools, and work on grants and papers.

Since 2017, the GEP organization has transitioned from a centralized model to distributed leadership. From 2006–2017, the GEP was led by Professor Sarah C.R. Elgin at Washington University in St. Louis. The central

organization hosted the new Faculty/TA Training Workshops and Faculty Alumni Workshops, maintained the GEP curriculum materials, provided technical and teaching support, managed the assessments and evaluations of GEP faculty and students, as well as the creation, claiming, submission, and reconciliation of student projects. The central organization also managed the data analysis and drafting of the scientific (e.g., Leung et al. 2017) and education research (e.g., Lopatto et al. 2014; Shaffer et al. 2014; Lopatto et al. 2020) manuscripts.

Under the new distributed leadership model, Professor Laura K. Reed at The University of Alabama serves as the GEP director, but the different responsibilities of the GEP are distributed among multiple committees led by other GEP faculty members (https://thegep.org/about/leadership/). The current GEP committees include: Steering; Assessment; Curriculum; Diversity, Equity, and Inclusion; Professional Development and Mentoring; and Science/IT. Each GEP faculty is affiliated with at least one committee, and the leadership for each committee (e.g., chair, vice chair, members) rotates every 2–3 years.

The community is also exploring and implementing new ways to recruit and train new members. Since 2020, GEP faculty and staff have organized online training workshops with experienced GEP faculty serving as trainers and mentors for new members (https://thegep.org/community/new-member-trainings/). Most of the new member training workshops were held during the summer — GEP faculty organized six online training workshops in Summer 2020 and three online training workshops in Summer 2021 for new members. During the 2020–2021 academic year, GEP faculty organized three online training workshops and one Regional Node Training workshop to train new members. These online training opportunities enabled faculty to incorporate the web-based GEP curriculum into existing courses, and allowed students to engage in research despite the lack of access to the physical lab environment during the COVID-19 pandemic. The pandemic, despite forcing new member training to be conducted online, rather than the traditional in-person workshops, actually led to a significant increase in new GEP members, as well as interest in teaching using modules that could easily be adapted to an on-line format. Faculty that had previously been hesitant about using in-silico labs were suddenly interested in using the materials that the GEP had already developed.

One outcome of our virtual training and the rapid growth of our community has been a discovery of the need to connect individual faculty members with each other to offer easy access to experienced colleagues. Thus, volunteer mentors have been matched with newly trained faculty based on geography and if possible — based on similar instructional implementation. In addition to the virtual community, we have a network of regional nodes (Figure 1b; https://thegep.org/community/nodes/regional-nodes-map-of-current-members/) that during non-pandemic times hold in-person activities, such as student research symposiums, faculty meetings, and TA trainings. During pandemic times these nodes have held virtual trainings, virtual faculty happy hours, and virtual research symposiums highlighting student work as a way to keep faculty and students connected and supported while the organization grows.

Collaborative Science Projects

GEP has diversified the science projects pursued by the community. In the past, the GEP research project has focused on investigating the properties and evolution of the Muller F element (also known as the dot chromosome) in multiple *Drosophila* species. This domain is unusual in the *Drosophila* genome because it exhibits both heterochromatic and euchromatic characteristics. The distal 1.3 Mb of the *D. melanogaster* F element is primarily heterochromatic by many criteria (e.g., high repeat density, low rates of recombination, enriched in the histone modifications H3K9me2/3 and the chromosomal protein HP1a). While heterochromatin is generally associated with gene silencing, the F element also contains ~80 protein-coding genes that exhibit a similar range of expression compared to genes that reside in the euchromatic portions of the genome. In order to use comparative genomics to characterize the evolution of genes and repeats in this unusual domain, GEP students participated in improving the genome sequence and annotating the small, mostly heterochromatic F element and a comparable portion of the euchromatic D element in different *Drosophila* species. Their work contributed to three publications on the evolution of the F element, with all contributing students as co-authors (Leung et al. 2010; Leung et al. 2015; Leung et al. 2017).

The most recent F element research project engages GEP faculty and students in improving the assembly and annotating F elements from a group of species that shared a common ancestor with *D. melanogaster* about 10–15 million years ago. These species are at the ideal evolutionary distances from *D. melanogaster* for the identification of conserved regulatory motifs via comparative genomics. Each student (or team of 2–3 students) took on the challenge of finishing ~100 kb and/or annotating a ~40 kb project from one of these species, either from the F element or a comparison region on the D element. In addition to the coding regions, GEP students also annotated the promoter regions of these genes to decipher the characteristics of the transcription start sites. Ongoing student annotation reconciliation and collaboration with Dr. Jeremy Buhler from the Washington University in St. Louis Computer Science & Engineering Department will utilize the Magma program (Ihuegbu et al. 2012) to identify conserved regulatory motifs via phylogenetic footprinting. The analysis could tell us much more about this interesting chromosome and the genes that reside there, potentially giving additional insights into the regulatory factors that enable F element genes to be expressed in a heterochromatic environment.

Three new scientific projects have been incorporated to the GEP community: the expanded F element project, the parasitoid wasp venom gene annotation project, and the *Drosophila* pathways project. These new opportunities expand both the range of scientific fields in which faculty and students may find particular interests, as well as the possibilities of pairing annotation projects with wet lab investigations.

The Expanded F Element Project

The expanded F element project aims to identify factors that contribute to changes in chromosome size and to assess the impact of chromosome size on gene characteristics. The assembled portion of the *D. ananassae* F element (~19.1 Mb) is substantially larger than the *D. melanogaster* F element (~1.3 Mb). Previous analysis of a 1.4 Mb region of the *D. ananassae* F element suggests this expansion can primarily be attributed to the increase in transposon density (Leung et al. 2017). Preliminary analyses of the *Drosophila* genomes sequenced by the modENCODE project indicates that the F elements in *D. bipectinata*, *D. kikkawai*, and *D. takahashii* are also larger than the *D. melanogaster* F element. However, the extent of the F element expansion in these three species are unclear due to gaps and misassemblies in the published genome assemblies.

The advent of sequencing technologies that produce long reads (e.g., Pacific Biosciences, Nanopore) provide the opportunity to improve the quality of the F element assemblies for these four *Drosophila* species. The GEP has constructed new genome assemblies for *D. bipectinata*, *D. kikkawai*, and *D. takahashii* based on sequencing data generated by the Pacific Biosciences, Nanopore, and Illumina sequencers. GEP students will analyze the improved genome assemblies and annotate the coding regions on the expanded F elements as well as genes on euchromatic reference regions near the base of the Muller D elements. In collaboration with Dr. Thomas Gingeras at Cold Spring Harbor Laboratory, the GEP has generated RAMPAGE (Batut et al. 2013) data for these four species. GEP students can utilize the RAMPAGE data to annotate the transcription start sites and to characterize the promoter architecture of F element genes in the four species. The gene annotations for these four *Drosophila* species will be used in a comparative analysis to identify the major contributors to the F element expansion and to assess the impact of the expansion on gene characteristics (e.g., promoter shape, gene size, codon bias, etc.). The comparative analysis might also identify factors that enable genes to function in a heterochromatic region with high repeat density. This project is led by Dr. Cindy J. Arrigo at New Jersey City University.

The Parasitoid Wasp Venom Gene Annotation Project

The parasitoid wasp venom gene annotation project aims to provide further insights into the function and evolution of parasitoid venoms. Current research has hinted at putative mechanisms of action, and it is expected that improved annotation of venom genes will help to characterize the putative functions of venom proteins and to identify proteins with potential dominant negative functions. Similarly, understanding the genomic context of venom genes should help resolve questions concerning the evolution of venom genes. Initially, GEP students will focus their annotation efforts on the *Ganaspis sp. 1* venom-encoding genes, and then expand to *Leptopilina boulardi* and

Leptopilina heterotoma. These gene annotations will utilize *Nasonia vitripennis* and *D. melanogaster* as the informant genomes, and leverage experimental data from RNA-Seq and protein mass spectrometry to facilitate the annotation of novel genes. The gene models produced by GEP students will allow us to explore the conservation of evolutionary and functional mechanisms across parasitoid species. This project is led by Dr. Nathan T. Mortimer at Illinois State University.

The Drosophila Pathways Project

The *Drosophila* pathways project uses network analysis approaches to elucidate the evolution and function of biological pathways, initially focusing on the insulin signaling pathway. Past studies have shown that the rates of evolution for the gene's protein-coding regions are correlated with its position within the network, the number of physical interactions, expression levels, and the existence of closely-related paralogs. GEP students will produce gene annotations for the putative orthologs and paralogs of *D. melanogaster* genes involved in the insulin signaling pathway in 27 *Drosophila* species. These improved gene models will be used to ascertain if the rates of evolution for the genes' regulatory regions parallels that of their protein-coding regions. GEP students will also draft microPublication (μPub) articles describing the annotation of each gene (e.g., Lose et al. 2021). This project is led by Dr. Laura K. Reed at The University of Alabama.

Table 1. The table shows how tasks align student and research goals for all three science areas. Adapted from CUREnet (https://serc.carleton.edu/curenet/collection/215335.html)

Research Goals → Student Goals ↓	**Research Goal 1: How did the Muller F element of *Drosophila* evolve, and how does its evolution impact the function of the resident genes?**	**Research Goal 2: In the context of a pathway, how have the genes and the regulation of those genes evolved across species (e.g., in *Drosophila*, parasitoid wasps)?**
Student Goal 1: Describe and conduct comparative genomics analysis	Students will integrate genomic data from other *Drosophila* species with experimental data for the target species (e.g., RNA-Seq) to develop a hypothesis for the gene model in their assigned species.	Students will integrate genomic data from other *Drosophila* and wasp species with experimental data (e.g., RNA-Seq, protein mass spectrometry) to develop a hypothesis for the gene model in their assigned species.
Student Goal 2: Correctly build a eukaryotic gene model	Students will use tools to verify that the proposed gene model satisfies basic biological constraints.	Students will use tools to verify that the proposed gene model satisfies basic biological constraints.
Student Goal 3: Participate productively in the science research process	Students prepare a report to defend their gene model given the available evidence. This report along with their model are used in the compilation and reconciliation of the full data set for the comparative analysis of the Muller F element.	Students prepare a report to defend their gene model given the available evidence. This report along with their model are used in the compilation and reconciliation of the full data set for the network analysis.

Models for Implementation

One of the factors which makes the resources provided by the GEP a valuable and attractive instrument for teaching and research is the potential for flexibility and diversity in the implementation. GEP faculty are from diverse institutions that vary in size, selectivity, and student characteristics. There are currently about 214 active GEP faculty members representing 2-year colleges, 4-year colleges with a primary focus in teaching, institutions

offering masters and PhDs, and Minority-serving institutions, including Historically Black Colleges and Universities (HBCUs) and tribal colleges, etc. (Figure 1; https://thegep.org/about/directory/). The commitment of the GEP to broadening participation has provided access to research for students who otherwise would not be able to access mentored research. GEP has made it a priority to make these students authors in publications. Efforts toward greater inclusivity have also challenged us to improve the curriculum so that they conform to the ADA Standards for Accessible Design (https://www.ada.gov/law-and-regs/design-standards/2010-stds/); we also currently provide some of our introductory curriculum in Spanish (https://www.ada.gov/law-and-regs/design-standards/2010-stds/).

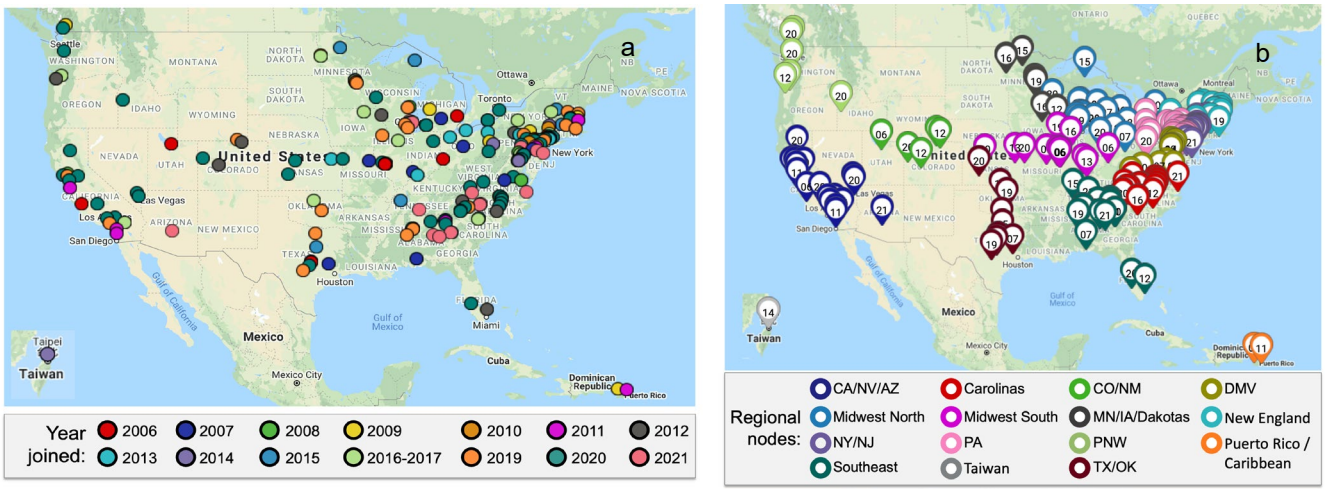

Figure 1. Map of GEP members (a) and regional nodes (b) as of Fall 2021. The current map of GEP members and the current map of regional nodes are available on the GEP website (at https://thegep.org/about/directory/ and https://thegep.org/community/nodes/regional-nodes-map-of-current-members/, respectively).

Faculty have different pedagogical goals which result in differences in implementation. The courses that utilize GEP materials range from Introductory Biology to Genetics, Biotechnology, Molecular Biology, Cell Biology, Genomics and Bioinformatics courses. For some faculty, providing students with a research experience is an essential goal, whereas others are looking to teach students the concepts of genomics, bioinformatics, and/or genetics with a hands-on approach that uses available genomics data organized in a genome browser. Some faculty implement GEP activities as a module within a course, devoting between 2 to 8 weeks to the GEP materials. Some faculty implement GEP as a standalone laboratory course for the entire semester, while others use it for independent research experiences. We have found that students at all institution types receive the same significant benefit from their participation in GEP. The one factor that scales with the degree of benefit is the amount of class time devoted to GEP activities — more instructional time invested in the project leads to better outcomes (Shaffer et al. 2014).

Workflow for GEP Scientific Projects

The GEP uses a “divide and conquer” strategy to facilitate student contributions to the GEP scientific projects (Figure 2). The lead scientists for each scientific project define the list of genomic regions or a collection of genes that require manual curation based on their research objectives. Visualization of the publicly available genome assemblies in conjunction with computational (e.g., protein sequence alignments, gene predictions) and experimental data (e.g., RNA-Seq, RAMPAGE) within a single workspace can assist the lead scientists in identifying the regions of interests and in generating the student projects. To enable faculty with limited technical expertise to construct genome browsers, the GEP has partnered with the Galaxy Project (https://galaxyproject.org) to create G-

OnRamp (https://g-onramp.org), a web-based platform for constructing UCSC Assembly Hubs and JBrowse/Apollo genome browsers for eukaryotic genomes (Liu et al. 2019). G-OnRamp was used to construct genome browsers for four wasp species as part of the GEP parasitoid wasp venom gene annotation project (Sargent et al. 2020).

GEP faculty claim the annotation projects, and each annotation project is completed by at least two students working independently at two different institutions. The completed annotation projects are submitted to the lead scientist and then reconciled by experienced students to create the reconciled gene models set for downstream analyses. The analysis results are reported in scientific publications and the associated datasets are submitted to public databases. GEP students who complete an annotation project and their faculty mentors are eligible to be co-authors on the resulting publication, provided that they read, critique, and approve the manuscript. GEP science publications have many student co-authors. For example, the manuscript analyzing the evolution of the F element in four *Drosophila* species has 1014 co-authors; 940 of whom participated as students (Leung et al. 2015). For the *Drosophila* pathways project, GEP students and faculty also create and submit articles describing the annotation of each gene model to the microPublication Biology journal (Rele et al. 2021; Lose et al. 2021).

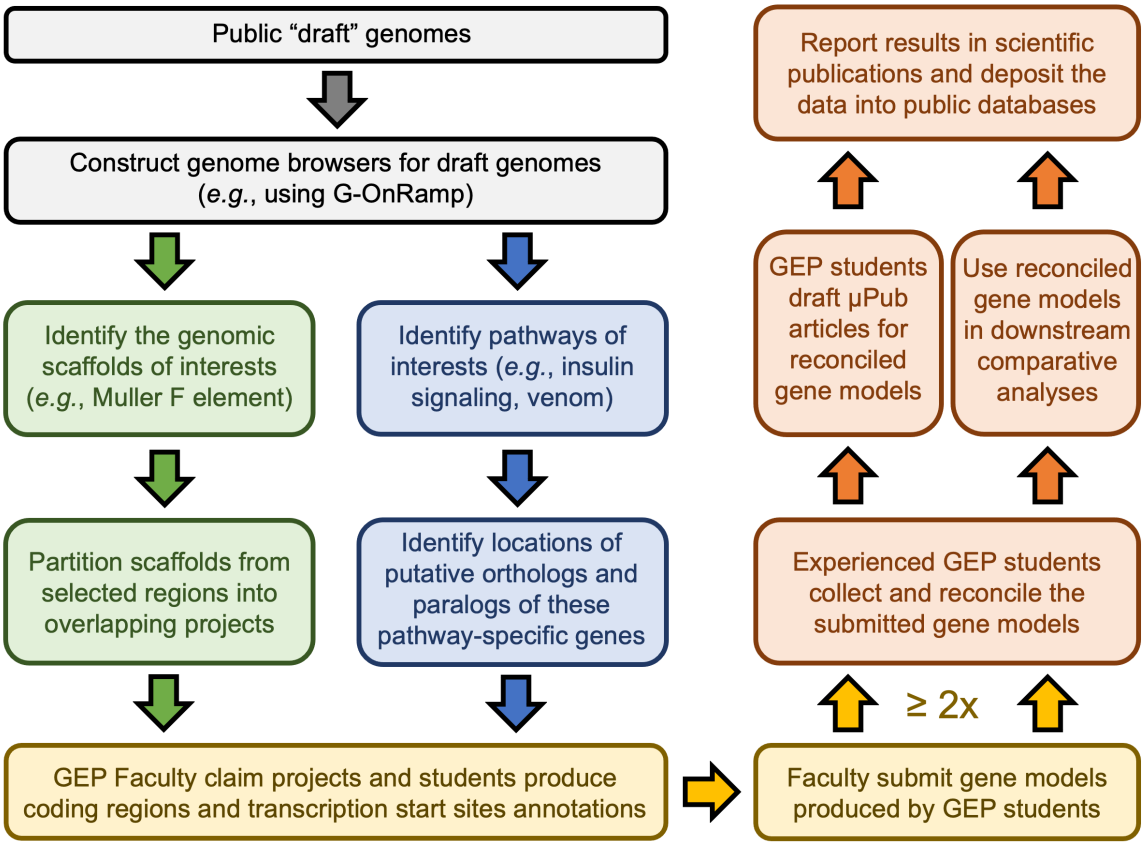

Figure 2. Overview of the annotation workflow for GEP scientific research projects

Curriculum to Prepare Students for Genome Annotation

Many faculty find that it is beneficial for students to review the organization of eukaryotic genes/genomes prior to engaging in the GEP research projects. GEP faculty members have developed and published six "Understanding Eukaryotic Genes" (UEG) modules that provide a review of eukaryotic gene structure, transcription, and translation (Laakso et al. 2017; https:/thegep.org/ueg/)Other introductory curriculum developed by GEP faculty includes four Transcription Start Sites modules that provide an overview of the promoter architecture in *Drosophila*, and the different types of experimental evidence (e.g., CAGE, RAMPAGE data) and computational evidence (e.g., blastn search results, whole genome multiple sequence alignments) that can be used to characterize *Drosophila*

promoters (see: https://thegep.org/tss/). In addition to these introductory materials, GEP faculty have also developed presentations, step-by-step walkthroughs, workflows, and exercises for the GEP research projects. These resources are freely available for faculty and student use through the "Curriculum" search page of the GEP website (https://thegep.org/curriculumsearch).

For this workshop, we introduced a condensed version of some of the UEG modules and completed a walkthrough that illustrates the protocol for annotating a *Drosophila* gene.

Understanding Eukaryotic Genes

The six UEG modules were created to help students understand eukaryotic gene structure and functionality via an active learning approach instead of traditional lectures. These modules leverage the features provided by the genome browser to enable students to visualize and explore eukaryotic genes/genomes at multiple scales (from a single nucleotide to a genomic region with multiple isoforms and genes).

These six modules now, also available in Spanish, https://thegep.org/translate-initiative/, are:

1. Introduction to the Genome Browser: What is a gene?
2. Transcription, Part I: From DNA sequence to transcription unit
3. Transcription, Part II: What happens to the initial transcript made by RNA pol II?
4. Removal of introns from pre-mRNA by splicing
5. Translation: The need for an Open Reading Frame
6. Alternative splicing

The UEG modules can be used in introductory classes by themselves or can be used as a preamble to the annotation process. If time is a concern, instructors can assign the modules as an outside-of-class independent project, in which students work through each Module independently, watching the videos that accompany each Module on the GEP YouTube channel (https://thegep.org/videos), and answering the questions in each Module for submission. Most students can complete the UEG modules in eight to ten 1-hour class periods.

Instructors can use multiple active learning approaches with these modules and deploy them in many different contexts. All or a subset of UEG modules can be used by an instructor, as each module is a stand-alone exercise. In most of our implementations, students work in groups of two or three using their own computers or computers in a computer lab. The instructor gives a brief introduction to the Module, and then students discuss each part of the lesson within their small groups while they explore the genome browser. We also often use large group discussions or peer instruction to address student questions and challenging concepts. Typically, students work together to complete each Module and produce a gene model by the end of Module 6. Students submit their answers to the questions associated with each Module to their instructor. As students discuss the exercises within each Module, they gather evidence, evaluate potential answers to each question, and occasionally resolve contradictions in the evidence. Using a group format has some risks of uneven participation, but generally stimulates useful dialogue.

Annotation Walkthrough

To prepare students to participate in one of science research projects, instructors usually start with a presentation which provides the scientific background of the research project, and describe how the data they are analyzing was generated. Instructors then typically ask their students to complete a step-by-step annotation walkthrough which illustrates the annotation protocol using one or more example genes. For this workshop, the student handouts include a modified version of the annotation walkthrough for the *Drosophila* F element research project (the original version is available through the F Element Project page on the GEP website; https://thegep.org/projects/felement/). The walkthrough for the *Drosophila* pathways project is available on the Pathways Project page on the GEP website (https://thegep.org/projects/pathways/). (The walkthrough for the parasitoid wasp venom gene annotation project is available on the Parasitoid Wasp Genomics page on the GEP website (https://thegep.org/projects/wasps/).

Once the students have completed the annotation training, instructors can claim annotation projects and assign them to students either independently or in groups of two or three. When the students have completed their project, the instructor submits the project back to the GEP where it will be combined with other student projects to produce the datasets for the downstream analyses and eventual publication. Most independent annotation projects take approximately twelve hours of class time.

Assessment of Participating Students and Faculty

To enable the GEP to learn and continue to optimize the student experience, participating students are encouraged to take a pre-course quiz and survey before starting to work with GEP materials (available through the Assessment page on the GEP website; https://thegep.org/community/assessment/). At the end of the course, students will take the post-course quiz and survey. To mitigate test sensitization, there are two versions of the knowledge quiz that cover the same concepts. Students are randomly assigned to one version of the quiz for the pre-course assessment and assigned to the other version of the quiz for the post-course assessment. A separate set of quizzes for the "Understanding Eukaryotic Genes" curriculum has been developed that aims to assess students' knowledge gains.

Student/faculty gains and testimonials

Besides the gain in knowledge about eukaryotic genes and genomes captured in the quizzes, faculty have also observed many positive impacts for their students in the areas of problem solving, independence, application of knowledge, involvement in peer-to-peer teaching, teamwork and collaboration, and developing a sense of ownership (Shaffer et al. 2010). The GEP approach facilitates an introduction to the process of research and helps to develop an understanding of how new knowledge is created in the field. Indeed, students in GEP-associated courses gained many of the skills and attitudes normally associated with a wet lab summer undergraduate research experience (Lopatto et al. 2008).

The two tables below (adapted from Lopatto et al. in preparation) list sample comments from students that reflect their perception of "research emphasis practices" utilized by their faculty.

Table 2. Comments from GEP student focus groups that reflect strategies to emphasize active research participation within the classroom.

Topic from Faculty Log	Example 1	Example 2
Better understood genes,	I feel like the goals were to understand a lot of the like nitty-gritty ways that genes work like start [codons] and stop [codons]	We also learned a lot of tricks and gimmicks as to how to annotate a gene or just how to look at a genome because the

annotations, genomics	and specific things to look for in like a sequence that you know is or possibly could be a gene. But also to understand on like a larger scale how to use the program to do that for you and how to use, like we had like four or five programs that we used on a regular basis to like check with each other and how to know when you needed more information or when you had the information you needed because going into it you don't know.	first one we looked at it was like what is this. And then at the end of the course we were like, oh, I recognize this.
Better understood how science was performed	Science isn't only being in the lab doing really cool things, you might say; science is most importantly the part in which we are analyzing data. All of the models that make up science as we know it now come from that.	It really is because I think one thing I learned is not everything will go according to what's been predicted or planned. It's going to keep on changing, so I think it, I mean, helps you to your mind towards it like - okay, not everything is going to work the way it should. So it… you know, it makes it a bit more adaptive and flexible in trying out different methods or techniques to find out a solution.
Discussed poster, paper, oral presentations	We also did presentations throughout the year so I felt like every week or couple weeks we would go up with our partner and present where we were at which was really helpful because it was like at that point not just working with your partner but having people be like well, you could try this or that next.	And I also wanted to say that we have a presentation at the end of the course, so I think that that presents our results to other teammates and basically from there we can infer different things and have different conclusions about our different contigs that we annotated. So I feel like it was a research team.
No right answer, construct an argument without knowing the right answer.	Well I always really like the feeling, just the… of accomplishment after you solve a particularly complex problem. Especially one that hasn't been addressed before like this project where there's really no… no one knows the right answer yet. And so I think it's… it's a real good… it's a real strong sense of accomplishment. You feel like you've done something useful for the world, you know?	Also with scientific knowledge, it's never really 100% a fact because we're always… it's always one of those things where we're just waiting until something disproves it and/or strengthens it, you know? And it's always a theory, never 100% a fact. And I think with scientific knowledge, it just strengthens… it gets to a point where it strengthens and strengthens and strengthens until something challenges it and then that challenge branches off to either be a part of that already existing theory or completely turns it around and disproves something we've known for so long and we kind of follow a new branch and try to explore that. It's always an explorable knowledge. It's never an indefinite type of thing.
Novel Research Problem (Real Research)	Okay. I think that it was definitely a very different experience than any other lab-based class you could have, since you're actually dealing with novel research. But	I really liked the dynamic of integrating current research and that all the data that we are providing are going to benefit other people and that we are contributing to

	I think that between the integration of like actually doing the novel research and understanding — like learning the principles behind them, it was definitely like a very positive and like much more learning-based experience than previous lab-type classes.	something real, not like the usual lab work that is just for us to observe the application of what we've learned in theory.

Table 3. Example comments from student surveys that reflect the GEP faculty use of strategies to emphasize active research participation within the classroom.

Topic from Faculty Log	Example 1	Example 2
Novel Research Problem (Real Research)	I think the effort of integrating research was a great strength to this course. We have done so many labs where our work was not important and it was getting old. It was special that we felt important and our work was valuable, it was not something that would be just thrown away.	I think it was cool that we got to actually conduct an experiment that was real. It made things seem more important than if we were just doing an experiment to do it.
Better understood genes, annotations, genomics	I thought this was a great class for anyone interested in genomics or even genetics in general. It goes in depth about the structure of genes, and gives students hands-on experience with many of the basic tools used by scientists working in genomics.	I think that despite the difficulty learning and understanding the research this course was valuable in teaching genomics. It is more beneficial to me to learn by doing the work than just by lecture.
Better understood how science was performed	This course was the most challenging course I had ever taken in my life. I have learned so much about how to conduct experiments, analyze data, and write scientific papers.	What was more valuable to me was being able to learn how researching works in general. Considering this is my first time, this gives me an opportunity to understand the difficulty of research a certain topic. I'll be more prepared and know what to expect when I do any future research instead of guessing what researching might be like.
No right answer, construct an argument without knowing the right answer.	Understand that sometimes the answer is not always clear and to try many different approaches to try and figure something out. This was challenging but very useful and our professor was so helpful.	One of the strengths was that [students can] engage their knowledge with real scientific problems, but the fact that nobody knows if they are right was scary.

Student Outline

Understanding Eukaryotic Genes

These modules are designed to facilitate student learning of gene structure, transcription, translation, and splicing using the genome browser. Each Module begins with a lesson plan that describes the objectives, pre-requisites, homework, and class instructions for the Module, followed by short readings to explain key terms and concepts. Each Module also contains multiple questions to help assess student understanding of the biological concepts. This set of curriculum also includes a Glossary which explains the key terms used in the UEG modules.

The modules require you to use the genome browser to visually explore DNA sequence and other genomic features surrounding the *tra* gene in *Drosophila melanogaster*. The genome browser is an important bioinformatics research tool, that is being used in these exercises as a pedagogical tool. The official UCSC Genome Browser (https://genome.ucsc.edu/) and the GEP UCSC Genome Browser mirror (https://gander.wustl.edu/) are publicly available online. There are also other project-specific mirrors of the UCSC Genome Browser and UCSC Assembly Hubs available online [e.g., UCSC Genome Archive (https://hgdownload.soe.ucsc.edu/hubs/), EMBL-EBI Track Hub Registry (https://trackhubregistry.org/)].

Module 1 will provide you with an introduction to the genome browser and its navigation. The rest of the modules will provide further training in the use of this freely available tool to understand complex biological phenomena. It is recommended that you watch the supporting videos that explain concepts such as splicing and phase, RNA-Seq data and the TopHat algorithm, or specific functionalities (e.g., “Short Match”) of the genome browser. The collection of videos is available through the “Understanding Eukaryotic Genes (UEG)” playlist on the GEP YouTube channel.

As determined by your instructor, you will be working either independently or cooperatively to understand eukaryotic gene structure and produce a gene model by the end of Module 6. You will be required to work through the questions embedded in the modules and submit the answers to your instructor. Upon completion of the modules, you will be better equipped to participate in a course-based research experience involving gene annotation.

Understanding Eukaryotic Genes Curriculum https://thegep.org/ueg/

Annotation of a *Drosophila* gene walkthrough

This walkthrough uses the annotation of the *CG31997* gene on the *D. biarmipes* Muller F element to illustrate the comparative annotation protocol for the F element project. This document illustrates how you can investigate a portion of a genomic region in an annotation project using FlyBase, the Gene Record Finder, and the gene prediction and RNA-Seq evidence tracks on the GEP UCSC Genome Browser. The walkthrough starts by identifying a genomic feature of interest in a *D. biarmipes* project based on the gene predictions tracks. FlyBase blastp search is used to compare the predicted protein sequence for this *D. biarmipes* feature with a database of proteins from the informant genome (*D. melanogaster*) to identify the putative ortholog. Once a putative ortholog has been identified the walkthrough demonstrates how to identify the precise coordinates of each coding exon using NCBI BLAST searches and RNA-Seq data. The walkthrough also describes a consideration of splicing and the need for compatible splice donor and acceptor splice sites between adjacent coding exons in order to maintain the open reading frame of annotated splice isoforms. The walkthrough concludes by verifying aspects of the proposed gene model using the Gene Model Checker and it includes a sample F Element Annotation Report demonstrating the submission of an annotated gene model report.

Annotation Workflows

GEP faculty and students have developed workflows that summarize the key analysis steps and program parameters in the GEP annotation protocol. Many students find these workflows useful as a quick reference when

they are working on their own annotation projects. These workflows are also provided as part of the student handouts for your reference:

1. The “GEP Annotation Workflow” provides an overview of the key analysis steps and bioinformatics tools used for annotation in the F element project.

2. The “Identify *D. melanogaster* Ortholog” decision tree illustrates the list of criteria that you can use to identify the putative *D. melanogaster* ortholog of a predicted gene.

3. The “Annotating Splice Sites” workflow describes the logic and the different lines of evidence that could be used to identify the splice donor and acceptor sites of coding exons.

Annotation of Drosophila gene walkthrough
GEP annotation workflow
Identify *D. melanogaster* Ortholog workflow
Annotation of Splice Sites: workflow

Materials

IT and Audio-Visual Support & Classroom Requirements

These exercises require a classroom or computer lab with a computer lectern with easy-to-use controls connected to a high-resolution ceiling-mounted projector and a sound system. Students would also need a fast Ethernet or wireless connection to the Internet. If students use their own laptops or tablets, well-distributed power outlets and large work surfaces are important. Students devices should have access to an Office Suite (e.g., Microsoft Office, Office 365, Google Docs), a simple text editor (e.g., WordPad, TextEdit), and a modern web browser (e.g., Google Chrome, Mozilla Firefox). If students use computers in a computer lab, instructors may need to consult with their local IT administrators to adjust the security settings on the computers (e.g., enable JavaScript, cookies, pop-up windows).

Online Documents

Materials developed for use by GEP members and other educators are freely available from the GEP website (https://thegep.org/; under "Curriculum").

The "Understanding Eukaryotic Genes" curriculum was developed by Margaret Laakso, Carina Howell, Cathy Silver Key, Leocadia Paliulis, Maria Santisteban, Chiyedza Small, Joyce Stamm, and Elena Gracheva (https://thegep.org/ueg/). Members of the GEP assisted by reviewing and revising these materials. All the modules are available with answer keys in a single package that uses the GEP name/password. Dr. Leocadia Paliulis (Bucknell University) has developed multiple videos that accompanied this curriculum and the videos are available on the GEP YouTube channel (https://www.youtube.com/c/GenomicsEducationPartnership).

The Annotation of the *Drosophila* Gene walkthrough was developed by Wilson Leung at Washington University in St. Louis (https://thegep.org/lessons/wleung-walkthrough-annotation_drosophila_gene/)

The three workflows included in the student handouts that provide a quick guide to the annotation protocol for the F Element project are available through the "Workflow" Curriculum Type tag on the GEP website (https://thegep.org/lesson-types/workflow/).

Tools for the GEP annotation projects are available through the "Projects" section of the GEP website (https://thegep.org/projects/). Each GEP science projects require a different set of annotation tools that are listed under the "Resources & Tools" section of the project page on the GEP website. For example, the Annotation of the *Drosophila* Gene walkthrough uses the annotation tools listed on the F Element project page (https://thegep.org/projects/felement/).

Finally, PowerPoint presentations used in this workshop as well as the answer keys can be found in the Appendices.

Annotation Tools

Gene Model Checker (https://thegep.org/checker)

This tool tests whether the proposed model satisfies basic biological constraints (e.g., contains a start codon, stop codon, canonical splice sites), and to visualize a comparison of the proposed gene model against the putative ortholog in *D. melanogaster*. This tool can help students identify and correct errors in the proposed gene models. The Gene Model Checker can also be used as a teaching tool to allow students to propose an initial gene model, receive feedback from their instructors, and then iteratively refine the gene model.

Gene Record Finder (https://thegep.org/finder)

This tool provides the gene structure information for the ortholog within *D. melanogaster* (i.e., the informant genome). The information is presented in graphical and table formats. Students can use this tool to examine the coding exon usage in the different isoforms of a *D. melanogaster* gene, and to retrieve the amino acid sequences for each coding exon to facilitate the mapping of coding exons against the target genome.

Annotations File Merger (https://thegep.org/merger)

This tool combines the annotation files for individual isoforms (produced by the Gene Model Checker) into a single file for project submission.

Notes for the Instructor

The CURE is introduced to students through faculty that have been trained in the tools, curriculum, and scientific background of the project (via in-person or online workshops). In some implementations, teaching assistants (TAs) also facilitate the learning process. TAs can be students who have previously taken the course and did well, or have successfully participated in an independent study project that utilizes GEP materials. Some of the TAs may have also received training from the GEP community at various regional node workshops. The GEP also has a group of virtual TAs that provide students and faculty with real time support in GEP activities (https://thegep.org/taschedule/).

TAs and peers are extremely valuable in CUREs because of the hands-on nature of this teaching. A faculty member might be able to help 5–6 groups of students and have insightful discussions about their project, approaches, assumptions, and conclusions. However, it becomes challenging to provide quality feedback to more groups. TAs can often help students who have questions on how to interpret analysis results (e.g., BLAST searches, Gene Model Checker checklists) and to resolve technical issues (e.g., with web tools, web browsers, operating systems, etc.). Hence TAs can help reduce the amount of time that a student would need to wait for help, and mitigate frustrations caused by technical issues. We have found that TAs are essential in guiding students in large sections with 40 or more students.

However, we caution faculty about providing students with too much help too early in the analysis process, as this can rob them of the aha moments and of the gains that come from struggling through the material. We have observed that students often struggle with annotation, yet find value in that struggle. Analysis of the faculty log, knowledge quizzes, and surveys from over a hundred faculty and their students supports the idea that frustration is "formative" for student learning (Lopatto et al. 2020). As mentioned above, iterations in genomics research can be done quickly and inexpensively. We therefore recommend creating a supportive environment of faculty, TAs, and peers that allows for formative frustration and iteration to enhance student learning.

We suggest communicating to students early on (and with periodic reminders) that they are participating in authentic research. Indeed, we have determined that making students aware that their work contributes to scientific knowledge improves both learning about genes and genomes and learning about the nature of science (Lopatto et al., 2022). Further, in our recent study, faculty used up to nine different "research emphasis practices" in their course implementations. We find that the use of these active learning practices shows positive correlations with the average student benefit reported in the post-course student survey and with the score on the post-course annotation quiz. However, we did not find any single active learning practice that is more effective than others in affecting student performance.

We would also recommend performing frequent progress checks as students work on their research project. This approach breaks the whole project into smaller parts with well-defined goals that would avoid overwhelming the students. Having students turn in results at designated points along the process also helps catch fundamental

problems early and also keeps students on track. For example, the *Drosophila* pathways project developed an Annotation Form to help students keep track of their work (https://thegep.org/lessons/path ways-project-annotation-form/). For annotation projects with multiple genes, it is a good idea to ask students to submit the annotation for one of the genes before they complete the entire project. This would help detect misconceptions or issues with the analysis approach early in the process, and avoid propagating the errors to all the gene models in a project.

As mentioned above, the UEG modules can be implemented by themselves and they are designed for students at the introductory undergraduate level. If teaching a junior/senior level course, students can spend two weeks working through a subset of the UEG modules (e.g., 1, 5, and 6), followed by one week on the Annotation of a *Drosophila* gene walkthrough, and one week on an annotation example. They would then be ready to work on their annotation projects for about 4–6 weeks.

If the implementation is a standalone course, students can be given a deeper introduction to BLAST in the beginning and work through an exercise (https://thegep.org/curriculumsearch; Curriculum Type: "Lesson with Exercises"). The GEP has also developed curriculum materials which focus on the analysis of RNA-Seq data, annotation of transcription start sites, and using *de novo* motif discovery tools (e.g., MEME Suite) for motif finding. To help GEP students better understand the limitations of bioinformatics tools, the GEP has developed curriculum materials on the algorithms used for sequence alignments (e.g., dynamic programming) and computational gene predictions (e.g., hidden Markov models). Depending on the pedagogical goals of the course, students can also perform activities beyond annotation, such as performing wet lab experiments to test hypotheses made during their annotation, and using multiple sequence alignments to examine the evolution of a gene across 27 *Drosophila* species.

Most faculty implementations require students to present their work either to the class at the end of the course or to the larger community, sometimes as a poster within an institution-wide research symposium. Many faculty also require their students to submit a written report that summarizes their analysis. For students participating in the research project, they will also need to complete the GEP Annotation Report form, and to prepare the annotation files for submission.

Cited References

Batut P, Dobin A, Plessy C, Carninci P, Gingeras TR. 2013. High-fidelity promoter profiling reveals widespread alternative promoter usage and transposon-driven developmental gene expression. Genome Res. Jan;23(1):169–80.

Elgin SCR, Hauser C, Holzen TM, Jones C, Kleinschmit A, Leatherman J; Genomics Education Partnership. 2017. The GEP: Crowd-Sourcing Big Data Analysis with Undergraduates. Trends Genet. Feb;33(2):81–85.

Ihuegbu NE, Stormo GD, Buhler J. 2012. Fast, sensitive discovery of conserved genome-wide motifs. J Comput Biol. Feb;19(2):139–47.

Laakso MM, Paliulis LV, Croonquist P, Derr B, Gracheva E, et al. 2017. An undergraduate bioinformatics curriculum that teaches eukaryotic gene structure. CourseSource. https://doi.org/10.24918/cs.2017.13.

Leung W, Shaffer CD, Cordonnier T, Wong J, Itano MS, et al. 2010. Evolution of a distinct genomic domain in *Drosophila*: Comparative analysis of the dot chromosome in *Drosophila melanogaster* and *Drosophila virilis*. Genetics. 185:1519–1534.

Leung W, Shaffer CD, Reed LK, Smith ST, Barshop W, et al. 2015. *Drosophila* Muller F elements maintain a distinct set of genomic properties over 40 million years of evolution. G3. 5:719–740.

Leung W, Shaffer CD, Chen EJ, Quisenberry TJ, Ko K, et al. 2017. Retrotransposons are the major contributors to the expansion of the *Drosophila ananassae* Muller F element. G3. 7:2439–2460.

Liu Y, Sargent L, Leung W, Elgin SCR, Goecks J. 2019. G-OnRamp: A Galaxy-based platform for collaborative annotation of eukaryotic genomes. Bioinformatics. May 9. pii:btz309.

Lopatto D, Alvarez C, Barnard D, Chandrasekaran C, Chung HM, et al. 2008. Education Forum: Genomics Education Partnership. Science. 322:684–685.

Lopatto D, Hauser C, Jones CJ, Paetkau D, Chandrasekaran V, et al. 2014. A central support system can facilitate implementation and sustainability of a classroom-based undergraduate research experience (CURE) in genomics. CBE Life Sci Educ. 13:711–723.

Lopatto D, Rosenwald AG, DiAngelo JR, Hark AT, Skerritt M, et al. 2020. Facilitating growth through frustration: using genomics research in a Course-Based Undergraduate Research Experience. J Microbiol Biol Educ. 21(1):21.1.6.

Lopatto D, Rosenwald AG, Burgess RC, Silver Key C, Van Stry M, et al., 2022. Student Attitudes Contribute to the Effectiveness of a Genomics CURE. J Microbiol Biol Educ. May 16;23(2):e00208-21. doi: 10.1128/jmbe.00208-21.

Lose B, Myers A, Fondse S, Alberts I, Stamm J, Youngblom JJ, Rele CP, Reed LK. 2021. *Drosophila yakuba — Tsc1*. microPublication Biology. 10.17912/micropub.biology.000474.

Reed L, Leung W. (2020, March 2). CUREnet: Genomics Education Partnership. Retrieved from https://serc.carleton.edu/curenet/collection/215335.html.

Rele CP, Williams J, Reed LK, Youngblom JJ, Leung W. 2021. *Drosophila grimshawi — Rheb*. microPublication Biology. 10.17912/micropub.biology.000371.

Sargent L, Liu Y, Leung W, Mortimer NT, Lopatto D, Goecks J, Elgin SCR. 2020. G-OnRamp: Generating genome browsers to facilitate undergraduate-driven collaborative genome annotation. PLoS Comput Biol. Jun 4. 16(6):e1007863.

Shaffer CD, Alvarez C, Bailey C, Barnard D, Bhalla S, et al. 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.

Shaffer CD, Alvarez CJ, Bednarski AE, Dunbar D, Goodman AL, et al. 2014. A course-based research experience: How benefits change with increased investment in instructional time. CBE-Life Sci. Educ. 13:111–130.

Acknowledgments

GEP is currently focusing its recruiting efforts for new members on faculty at Community Colleges and Minority-Serving Institutions (https://thegep.org/contact/).

The Genomics Education Partnership (GEP) is supported by the National Science Foundation (NSF) under Grant No. 1915544 and National Institute of General Medical Sciences (NIGMS) of the National Institutes of Health (NIH) under award number R25GM130517 to Laura K. Reed. The GEP is hosted by the Department of Biological Sciences at The University of Alabama with continuing support from the Department of Biology at Washington University in St. Louis. The F Element Project is supported by the NSF under Grant No. 2114661 to Cindy J. Arrigo. The Parasitoid Wasps Project is supported by NIGMS of the NIH under award number R35GM133760 to Nathan T. Mortimer.

About the Authors

Maria S. Santisteban
Department of Biology, UNC Pembroke, Pembroke, NC 28352

Dr. Maria Santisteban is a Professor of Biology at the University of North Carolina at Pembroke. Her research interests focus on the role that the H2A.Z histone variant plays in gene regulation, which her lab investigates using the model organism *Saccharomyces cerevisiae*. She teaches Cell Biology (BIO 3710), Molecular Biology (BIO 4130), and Microbiology (BIO 3150) and has integrated GEP materials into the first two courses, as well as independent research student projects. Dr. Santisteban joined the GEP in 2012 and implemented GEP

material for the first time in 2013. She served as chair of the GEP Curriculum Committee in 2021-2022 and is the current NC node leader.

James Godde
Department of Biology, Monmouth College, Monmouth, IL 61462

Dr. James Godde is the McGrath Professor of Biology at Monmouth College in Illinois, where he has taught for the past 20 years. His research focuses on repetitive elements of DNA such as trinucleotide repeats in eukaryotes and CRISPRs in prokaryotes. Dr. Godde teaches Bioinformatics (BIOC 300), Genetics (BIOL202), Microbiology (BIOL302), and Molecular Biology (BIOL354). It is in this final course that he has implemented GEP material. Dr. Godde joined the GEP in 2019 and taught students using the Understanding Eukaryotic Genes modules for the first time this same year. He is currently serving as Chair of the GEP Professional Development and Mentoring Committee.

Wilson Leung
Department of Biology, Washington University in St. Louis, St. Louis, MO 63130

Wilson Leung serves as the IT director for the Genomics Education Partnership (GEP). He helps maintain the GEP web framework and infrastructure. He creates the student projects and the evidence tracks for the *Drosophila* genome assemblies available on the GEP UCSC Genome Browser, and is the primary developer of the annotation tools (e.g., Gene Model Checker and the GEP Project Management System) used by GEP faculty and students. He also contributes to the development of GEP curriculum materials, and serves as an instructor for the GEP Faculty Training Workshops. He is the Program Manager for the G-OnRamp project, and a member of the Genomics Education Alliance (GEA) IT/Infrastructure Committee. His research interests include chromatin packaging and the evolution of the *Drosophila* Muller F element. He is a lab instructor for the Bio 4342 Research Explorations in Genomics course at Washington University in St. Louis — an upper-level laboratory course that implements GEP materials.

Catherine Reinke
Department of Biology, Linfield University, McMinnville, OR 97128

Dr. Catherine Reinke is an Associate Professor of Biology at Linfield University in McMinnville, OR, where she has taught for the past 10 years. Her research interests focus on the biogenesis and regulation of microRNAs, which her lab investigates using the model organism *Drosophila melanogaster*. She teaches Principles of Biology (BIOL 210), Genetics (BIOL 270), Molecular and Cellular Biology (BIOL 400), and Animal Development (BIOL 420) and has used GEP materials to teach sophomores, juniors and seniors in Genetics, Molecular Cell Biology, and Animal Development, as well as independent research student projects. Dr. Reinke joined the GEP in 2012 and implemented GEP material for the first time in 2012. Currently she serves in the GEP Professional Development and Mentorship Committee as the Director of New Member Training.

Mission, Review Process & Disclaimer

The Association for Biology Laboratory Education (ABLE) was founded in 1979 to promote information exchange among university and college educators actively concerned with teaching biology in a laboratory setting. The focus of ABLE is to improve the undergraduate biology laboratory experience by promoting the development and dissemination of interesting, innovative, and reliable laboratory exercises. For more information about ABLE, please visit http://www.ableweb.org/.

Papers published in *Advances in Biology Laboratory Education: Peer-Reviewed Publication of the Conference of the Association for Biology Laboratory Education* are evaluated and selected by a committee prior to presentation at the conference, peer-reviewed by participants at the conference, and edited by members of the ABLE Editorial Board.

Citing This Article

Santisteban MS, Godde J, Goodman AL, Hauser CR, Paetkau D, Reinke C, Leung W, Arrigo C, Mortimer NT, and Reed LK. 2023. Using genome annotation projects to teach eukaryotic gene structure and to engage students in genomics research. Article #17 In: Boone E and Thuecks S, eds. *Advances in biology laboratory education*. Volume 43. Publication of the 43nd Conference of the Association for Biology Laboratory Education (ABLE).
https://doi.org/10.37590/able.v43.art17