

Improving knowledge of metagenome-assembled genomes (MAGs) through bioinformatics and article annotation

Jason M. Whitham,^{1,2} Carlos C. Goller³

AUTHOR AFFILIATIONS See affiliation list on p. 11.

ABSTRACT DNA from microbial communities can be sequenced and assembled to gain insight into the microbes that may be present in unique environments. Powerful computational tools, combined with more accessible sequencing technologies, have enabled metagenome-assembled genome (MAG) analysis in course-based settings. However, the computational methods and assumptions surrounding the creation of MAGs, as well as their application in understanding microbes in biomes, are often complex and intimidating to new users. The metagenomics course we designed enrolls undergraduate and graduate students in a half-semester lab experience. We hypothesized that collaborative annotation of specific bioinformatics research articles, paired with student application of tools using guided case studies on the powerful KBase bioinformatics web platform, would enhance learning of key MAG concepts. Student learning outcomes on conceptual quizzes, as well as learner perceptions of the assignment and confidence in using KBase and other bioinformatics tools, were analyzed. Assessments and surveys of student perceptions were collected over several semesters using consistent assignments, readings, and KBase narratives. Learning gains were identified for specific MAG analysis concepts and data interpretation. Nevertheless, misconceptions continue, and confidence in bioinformatics approaches varies. Additional exploration of qualitative data may suggest concepts to reinforce and resources to support learners. Combining KBase, collaborative annotation of primary literature, guided case studies, and aligned assessments effectively promotes students' conceptual and mechanistic understanding of MAGs and the assumptions underlying their creation and use.

KEYWORDS case studies, bioinformatics, metagenomics, binning, microbiome, discussion forums, annotation, KBase, primary literature, critical reading

Thanks to advances in sequencing technologies and the availability of core facilities and commercial services, sequencing of nucleic acids from microbial communities has become a nearly routine laboratory technique. Analysis of the resulting metagenomic data can be challenging; however, it is computationally intensive and often requires working with large and complex data sets. Fortunately, graphical user interfaces (GUIs) and web-based tools have simplified analysis, enabling graduate and undergraduate students to perform complex analyses that were previously limited to high-capacity labs and core facilities. For example, platforms such as KBase (1) provide students with access to powerful bioinformatics tools in a user-friendly environment, supporting education and independent exploration.

As students begin to explore metagenomic workflows, developing an understanding of the algorithms behind bioinformatics tools becomes essential. This experience not only deepens a learner's grasp of molecular biology and computational principles but also equips them with in-demand interdisciplinary skills. National surveys and educational initiatives have underscored the importance of integrating bioinformatics into

Editor David R. Wessner, Davidson College, Davidson, North Carolina, USA

Address correspondence to Carlos C. Goller, ccgoller@ncsu.edu.

The authors declare no conflict of interest.

Received 26 September 2025

Accepted 30 January 2026

Published 26 February 2026

Copyright © 2026 Whitham and Goller. This is an open-access article distributed under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International license](#).

life sciences curricula, while also highlighting persistent barriers to its adoption (2–4). These barriers—ranging from data complexity to a lack of instructional time or student preparation—can deter instructors from incorporating bioinformatics into their courses (3, 5).

One area of bioinformatics that holds tremendous educational value, despite its complexity, is the study of metagenome-assembled genomes (MAGs). MAGs are draft genomes reconstructed directly from metagenomic sequencing data, thereby bypassing the need to isolate or culture individual organisms (6). By assembling contiguous sequences (contigs) from sequencing reads and grouping them based on properties such as abundance patterns and nucleotide composition, researchers can separate the genomes of individual species present in diverse environmental samples (7). To visualize this, one may think of MAGs as components of a de-homogenized smoothie—quickly and cleverly sorted and reconstructed into nearly whole or partial berries, seeds, greens, and other pre-blended ingredients (Fig. 1A). MAGs have revolutionized microbiology by enabling the study of uncultivated and previously unknown microorganisms—often referred to as microbial “dark matter”—thereby significantly expanding our understanding of microbial diversity, ecology, and function (8). Although often incomplete, fragmented, or contaminated with sequences from other organisms, MAGs offer critical insights into individual members’ metabolic potential and phylogeny (9).

Recognizing both the promise and the challenges of teaching MAG concepts, we developed a series of case studies and reading assignments to progressively build students’ competence in metagenomic analysis. These activities specifically guide students through the computational steps of assembling metagenomes and binning sequences into MAGs, as reviewed by Meziti et al. (10) and Mirete et al. (11). We hypothesized that collaborative annotation of bioinformatics literature (12), when combined with practical, guided analyses using the KBase platform (13), would enhance student understanding. To evaluate this, we collected data across several semesters using consistent assessments, KBase narratives, and reading assignments. The results suggested learning gains in MAG-specific concepts and data interpretation skills. Nevertheless, some misconceptions persisted, and student confidence varied—often influenced by factors such as academic background, interest in computing, or access to peer support. These findings highlight the utility of integrating interactive case studies, social annotation, and accessible bioinformatics platforms to enhance student learning. Moreover, they underscore the importance of explicitly teaching the underlying assumptions and mechanisms of widely used computational tools. As metagenomic analyses continue to evolve, integrating and refining these pedagogical strategies will be critical to preparing the next generation of life scientists.

Intended audience

The metagenomics course was designed for undergraduate juniors and seniors, as well as graduate students from diverse majors and academic plans, and was previously described (14). Students from a variety of majors and graduate programs are eligible to enroll after completing a semester-long prerequisite molecular biology course (15). The intended audience comprises students in the life sciences, engineering, environmental sciences, chemistry, materials science, and other related fields. We believe this lesson can be adapted for various additional courses and settings, including microbiology and 200- and 300-level genetics experiences (refer to the “Possible adaptations and modifications” section).

Learning time

The course met twice per week: one 5 h lab and one 110-min in-person active-learning session. Homework assignments were designed to be completed in ~6–8 h of out-of-class time throughout the half-semester course. This lesson can be adapted to a variety of formats and lengths, including a completely asynchronous online format, since the

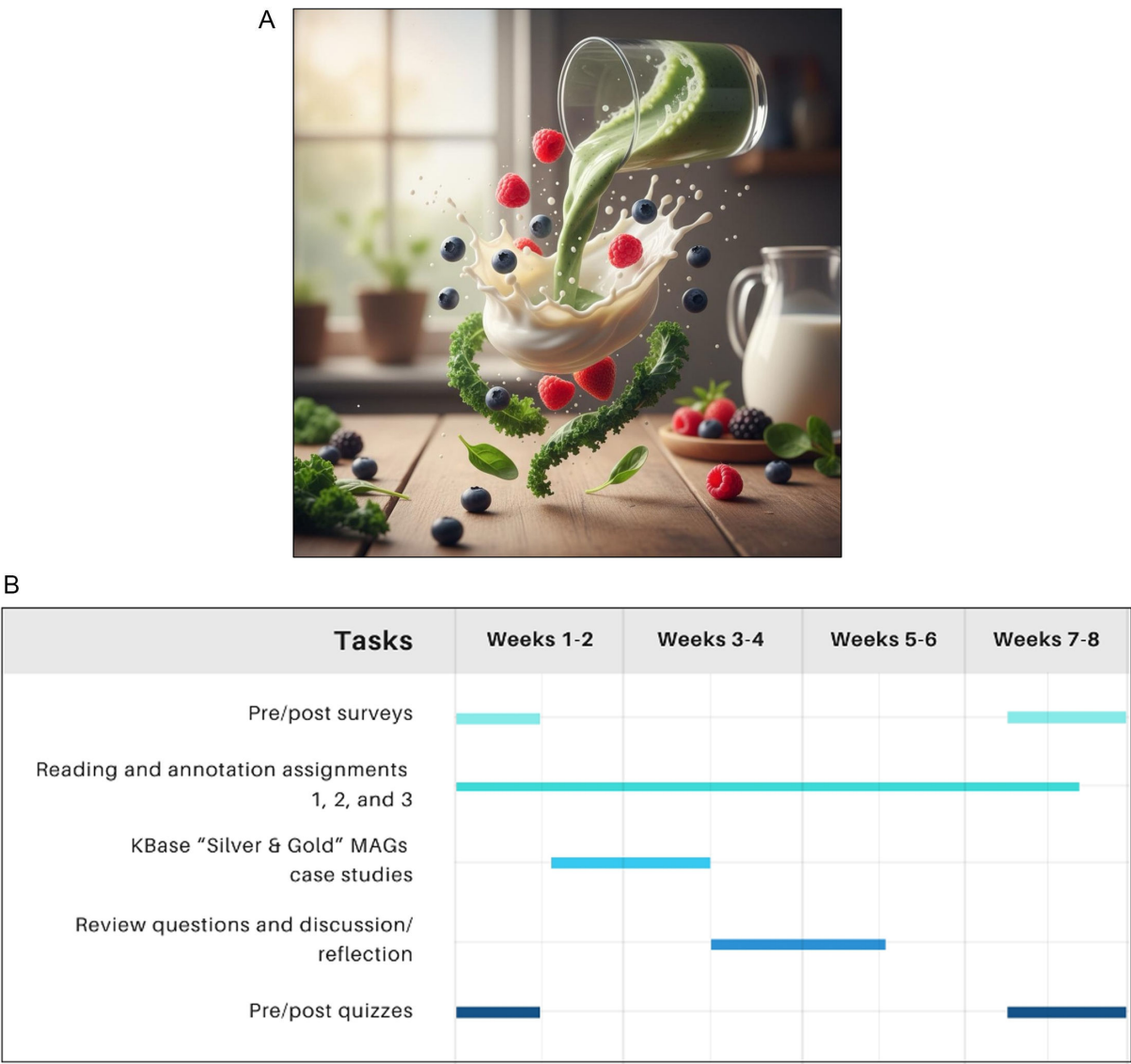


FIG 1 Timing of surveys, assessments, and MAGs case studies and lessons. KBase training and case studies are part of the second module of the half-semester course. (A) Part of the KBase training is generating MAGs. A smoothie, deconstructed into its berries, vegetables, and milk, is a visual analogy for how metagenomic data can be reconstructed into MAGs. The image was created with OpenArt. (B) Reading and annotation of three relevant articles on the techniques used are embedded throughout the semester to engage students in practicing the reading and analysis of algorithms and assumptions in sequence analysis tools. The image was created with Canva.com.

tools used for creating and analyzing MAGs (KBase) and annotating literature (Perussall or Hypothes.is) can be used online.

Prerequisite student knowledge

The prerequisite for the metagenomics course is the completion of a semester-long molecular biology course for undergraduate juniors and seniors, as well as graduate students, described previously (15). This course provides familiarity with biotechnology, lab experience, and knowledge of specific cloning and protein expression techniques.

Instructional team

The instructional team consisted of two instructors and one graduate teaching assistant (GTA). The team has experience in microbiology, biotechnology, and molecular biology.

The GTA assisted with facilitating lab sessions and grading, including evaluating lab notebook entries and lab write-ups.

Learning outcomes

The course was designed to help students gain confidence in using metagenomics approaches to explore diverse sample types. Participants engage with a variety of bioinformatics tools and lab-based sample preparation techniques. A central objective of the course is to promote critical thinking and responsible innovation by providing a framework for designing a sequencing-based microbiome study. For most enrolled students, this is their first exposure to metagenomics laboratory procedures. The course goals focus on scientific practices needed to prepare, sequence, and analyze metagenomic samples successfully. The metagenomics course has the course outcomes (COs) listed below that can be adapted for a variety of courses, disciplines, and academic levels.

- CO 1. Demonstrate laboratory skills required of a modern-day molecular biologist in the era of high-throughput sequencing.
- CO 2. Evaluate how bioinformatics methods were employed by the authors to explore a particular hypothesis (from the *CourseSource* framework).
- CO 3. Given a scientific question, develop a hypothesis and computational approaches that could be used to explore the hypothesis (from *CourseSource* framework).
- CO 4. Use pre-existing tools to analyze a metagenomic data set to determine the set of organisms present in a metagenomic sample (e.g., 16S rRNA, Greengenes, mothur, etc.) (from *CourseSource* framework).
- CO 5. Interpret data and identify limitations related to metagenomic surveys.
- CO 6. Design a critical thinking scenario and explain analyses for hypothesis testing of metagenomic data.

A course map graphically depicting the alignment of outcomes, instructional materials, and assessments is provided in Material S3.

PROCEDURE

The course consists of eight modules, each lasting 1 or 2 weeks. Each module includes online activities, such as interactive presentations (H5P slides with videos and links to external resources), accompanied by knowledge-check questions (not graded). During active learning sessions, participants review material from the presentation, prepare for lab sessions, and work in groups to explore central concepts in metagenomics, bacterial diversity metrics, experimental design, and science communication. During later sessions, groups work on bioinformatics tasks using data generated in the lab. The experimentation modules include bioinformatic and “wet lab” sessions culminating in the presentation of unique group projects based on the class data. Table 1 describes how the MAGs lesson is sequenced, and articles that are collaboratively annotated are read throughout the semester.

The second module in the course focuses on MAGs. Table 2 lists the module outcomes, activities, and aligned assessments. Figure 1B depicts the timeline of activities related to the MAG module and the pre/post quizzes and surveys.

KBase introduction and training

Participants are introduced to the research project through a combination of in-person and online resources. Time was provided in class to introduce KBase, including playing the overview and interactive narrative tutorials available on the KBase YouTube channel. The two case studies were introduced in class (~30 min). Next, groups of 3–5 students were encouraged to work collaboratively on the first KBase narrative, discussing the steps, data, and results.

TABLE 1 Weekly activities, including reading/annotation assignments, for the MAGs lesson

Week	Instructional materials and learning activities
Module 1: What is metagenomics? Introduction to topic and methods.	• Pre-quiz and pre-survey • KBase orientation videos (KBase YouTube channel) • Reading assignment 1: (16)
Module 2: What are MAGs? Whole metagenome sequencing and analysis.	• Module 2. Video overview and activity • KBase introduction assignment videos • Assignment of KBase Silver narrative • Assignment of KBase Gold narrative
Module 3: What is diversity? Different metrics.	• Module 3. Video overview and activity • KBase Silver and Gold wrap-up discussion. Lab entry with results due • Reading assignment 2: MetaBAT (17)
Module 4: What pipelines are available? Pipelines and computational resources.	• Module 4. Video overview and activity
Module 5: What are OTUs and ASVs? Introduction to the cloud environment and workflows.	• Module 5. Video overview and activity • Reading assignment 3: Potential applications of MAGs (18)
Module 6: How can we plot this? How certain are we? Data visualization and statistics.	• Module 6. Video overview and activity
Module 7: What did we discover? Project presentations and podcasts.	• Module 7. Video overview and activity • Post-quiz and post-survey

Collaborative annotation of primary literature

Throughout the course, three primary literature articles were assigned and discussed during lab sessions. The online collaborative annotation tools Hypothes.is and Perusall were used to annotate and asynchronously discuss articles, with Hypothes.is being used in the first year of the study, then transitioning to Perusall, integrated with the Moodle Learning Management System (LMS) we used. Hypothes.is and Perusall are described as “social annotation” tools that allow participants to collaboratively read and comment on digital texts, images, and videos, supporting asynchronous online discussions (12, 19, 20). We previously used social annotation tools in this course to encourage peer support and careful analysis of often complex primary literature articles (12).

Students were encouraged to annotate each article with at least 10 meaningful annotations per reading assignment. Full credit was awarded if students completed the minimum number of annotations. Students were encouraged to revisit the articles and post additional connections within 2 weeks of the in-person discussion date. Three sets of annotations, one for each article, were weighted equally and averaged to account for 10% of the course grade. Reading assignments and annotation guidelines are available in Appendices S1 and S2, respectively.

Materials

Access to all materials was provided through the LMS. Reading lists, guidelines, and links to the KBase tutorials and case study narratives were shared as Google Docs. Reminders of the due dates for completion of the case studies and reading assignments were shared via announcements and weekly “Reminders” summary slides.

The KBase narratives we developed for this activity are publicly available at the links below.

TABLE 2 Outcomes, activities, and assessments for “Module 2: What are MAGs?”

Module outcomes (MO)	Activities	Assessments
1. MO 2.1. Perform quality control of read libraries (CO 3, CO 4, CO 5)	1. Module 2. Video overview and activity (MO 2.1, 2.2)	1. Complete the review/reflection questions from the narratives for in-class discussion
2. MO 2.2. Predict the taxonomic population structure of environmental shotgun reads (CO 3, CO 4, CO 5)	2. KBase Silver narrative (MO 2.1-2.6)	2. Complete a lab entry summarizing results
3. MO 2.3. Identify high-quality bins to extract and annotate (CO 3, CO 4, CO 5)	3. KBase Gold narrative (MO 2.1-2.5)	
4. MO 2.4. Annotate genes of extracted MAGs (CO 3, CO 4, CO 5)	4. Lab session 2: Library prep and quantification	
5. MO 2.5. Place MAGs into a species tree (CO 3, CO 4, CO 5)		
6. MO 2.6. Build metabolic models (CO 3, CO 4, CO 5)		

- KBase Silver Narrative: Determining Media Formulation Requirements for Isolation of Microbiome Constituents - <https://kbase.us/n/68579/143/>
- KBase Gold Narrative: Can you find *Delftia*? - <https://kbase.us/n/67335/259/>

The KBase Silver narrative guides the user through the process of extracting and annotating high-quality genomes from metagenomic data, performing phylogenomic analysis, building metabolic models, and using these to predict nutrient requirements for growth and isolate the corresponding microbes. Students use this narrative to identify growth factors for a myxobacteria (“slime bacteria”) yet to be isolated from the rhizosphere of *Miscanthus giganteus* (a hybrid of “Silvergrass”), cultivated at the Kellogg Biological Station in Michigan.

In the KBase Gold narrative, students mine metagenomic data for *Delftia* genomes. *Delftia* is a genus of bacteria that includes strains that can produce gold nanoparticles from gold ions in their environment. These bacteria have been found living in biofilms with *Cupriavidis metallidurans* on gold nuggets, in soil, sinks, and rhizospheres of different plants, where *Delftia* promotes their growth (21, 22). Students import metagenome data sets from a variety of environments to search through in the Gold narrative. Unfortunately, as is the case in real life, sometimes no *Delftia* MAGs are found. This manages students’ expectations about scientific discovery; they gain experience through their own microbial hunts and specifically try out a variety of tools, including four different assembly apps/tools.

Student instructions

Students begin the course with an introduction to KBase skills through the KBase Overview video available on the KBase YouTube channel. Participants complete an online, self-paced presentation on MAGs created by JW before attending an in-person presentation and review of the MAGs. Instructors emphasize that the reading assignments are sequenced to introduce the use of MAGs, a relevant bioinformatics tool, and finally, an additional study that may benefit from MAG analyses. Students complete individual annotations of reading assignments and are encouraged to work in teams of 3–5 members on KBase Silver and the more open-ended Gold case study narratives (13).

Faculty instructions

Reading assignments (7, 16, 18) are graded based on the completion of the set number of annotations by the deadline for in-class discussion. Over several years, we learned that KBase training should start early in the semester. Thus, we introduce KBase in module one as a powerful tool for metagenome analysis and share links to KBase videos and tutorials. It is important to encourage participants to create free KBase accounts before starting the module on MAGs, so that they can copy the narratives, join the

group/class “Org,” investigate apps/tools, and explore open-access databases. The KBase case studies are not graded; instead, the workflow and tools used are discussed in class, and participants complete a short reflection. Most students complete both case studies, and working in groups facilitates the exchange of ideas and comparison of results. At the beginning and end of the half-semester course (weeks 1–2, 6–7), participants are invited to complete pre- and post-quizzes, which are shared as quizzes on the learning management system (LMS). The quizzes we used are available in Appendix S8. We also survey students at the beginning and the end of the course to learn about their self-reported confidence with bioinformatics tools (Appendix S4 to S7).

Suggestions for determining student learning

To assess student learning, we opted for optional pre- and post-quizzes on MAGs and metagenomic analyses. Alternatively, both pre- and post-quizzes, as well as KBase case studies, can be used as knowledge checks/mastery quizzes for a grade. KBase narratives can be “submitted” for review by the instructor and graded based on accuracy and completion (“Silver Narrative”) or to learn what groups found in the data sets they decided to analyze with the “Gold Narrative.” In-class knowledge-check quizzes using polling software (PollEverywhere, Menti, or Wooclap, for example) can provide feedback to participants.

Safety issues

There are no laboratory safety issues to consider for these computer-based activities.

DISCUSSION

Field testing

This activity was implemented in an undergraduate and graduate-level metagenomics half-semester course at North Carolina State University, offered in the fall semesters from 2021 to 2024. Students provided informed consent via an online form to use their data in this study. From fall 2021 to fall 2024, the course enrolled 54 students, with 30 graduate students and 24 undergraduates. Enrollment fluctuated between 11 and 15 per semester. Twenty-nine undergraduates were classified as seniors, and one was an international exchange student. Most undergraduate participants were majoring in chemical engineering, biological sciences, microbiology, and genetics. Graduate students included 10 non-thesis master’s students in microbial biotechnology (5) and physiology programs (5), and 20 doctoral students in programs such as chemical engineering, civil engineering, and plant pathology.

Evidence of student learning

Pre/post quizzes

Students were encouraged to complete an optional quiz before and after the KBase MAGs module (Appendix S8), comprising 16 questions. From fall 2021 to fall 2024, the average grade on this multiple-choice assignment was 71.34 on the pre-quiz ($n = 53$, $SD = 18.24$, range = 0–100) and 81.37 on the post-quiz ($n = 46$, $SD = 15.79$, range = 21–100). A two-sample t -test assuming equal variances (homoscedastic) returned a P -value of 0.0046, indicating a statistically significant difference between pre- and post-quiz grades. A Cohen’s d value of 0.597 was calculated, which indicates the standardized difference between two means. This represents a medium effect size.

Surveys

At the beginning and last week of the half-semester course, participants were encouraged to complete an optional survey about their confidence and perceptions of bioinformatics and annotation tools. Example surveys and recruitment materials are

available in Appendices S4 through S7. The pre-survey was distributed as a hard copy on the first lab session; the post-survey was distributed electronically using Qualtrics during the last week of class. Respondents were asked to rate their confidence in using bioinformatics tools to address future challenges in the pre-survey. The pre-survey questions were based on the North Carolina Community College Virtual Learning online readiness tool, available [online](#). The post-survey is based on questions from the Madlung *et al.* survey (23) for evaluating perceptions of bioinformatics and several free-response questions about article annotation and KBase use.

Survey analysis

Students were surveyed at the end of the semester to assess their use of bioinformatics tools and their level of confidence in creating new KBase narratives (post-survey) after completing the assigned readings, annotations, and KBase narratives. Table 3 summarizes responses to open-ended questions from several semesters.

Findings

For learners new to bioinformatics tools and platforms, understanding the purpose and assumptions of each tool used provides a skill set that can be applied to various computational biology questions. Through the social annotation of a combination of microbial ecology and bioinformatics publications, students learn fundamental approaches and concepts. We used this background to support learners as they worked on two case studies built on the KBase platform. These interactive narratives provide authentic data sets and tools, challenging students to explore open-access data sets. After implementing this approach over several semesters, our findings from pre- and post-quizzes and student surveys suggest that students' conceptual understanding increased after completing the module on MAGs using KBase and annotating relevant bioinformatics articles, as measured by quiz scores.

Bioinformatics reading assignments and annotation

Participants approached reading articles and case studies critically, asking their peers questions and searching for external sources of complementary information. For example, participants noted that, although challenging, the reading and annotation assignments helped them learn new bioinformatics concepts and apply this knowledge to using KBase for metagenomic data analysis (Table 3).

Participants gained confidence in reading about and using bioinformatics tools

Survey respondents reported confidence in reading primary literature about computational tools. However, confidence in the use of (KBase) bioinformatics apps varied (Table 3).

Conceptual understanding of MAGs increased in participants

Students participating in the study demonstrated improved conceptual understanding of metagenomics and MAG concepts, as evidenced by a statistically significant increase in pre- and post-quiz scores (Fig. 2).

Supporting student confidence in narrative design is needed

When asked what they would do differently, several participants over the years reported wanting to learn more about KBase and adapt it to their data sets. For example, one participant stated, "I would spend more time applying my own data sets to them rather than just reading tutorials." Additionally, some expressed interest in a third exercise without the scaffolding provided by a pre-computed narrative (Silver) or a partially computed one (Gold): "If done again, I would likely try to make my own narrative instead of focusing on using premade ones." However, participants reported varying levels of comfort in starting to design a new KBase MAG narrative on their own: while some were

TABLE 3 Representative quotes from student survey responses^a

Survey question	Common themes (% of responses, <i>n</i>)	Representative quotes
How did you decide to analyze the reading assignments?	• Re-read text (45%, 10) • Reviewed methods (36%, 8) • Reviewed results and figure (27%, 6) • Defined keywords (18%, 4)	• I liked to read through them while making annotations to understand the text. Sometimes, it helped to think of my own summaries after reading each section. • I would look through their methods after identifying their core claims in the abstract. Then I would look at other people's comments or questions. • I like to re-write things to prove that I actually understand what I just read.
Will you continue using this approach beyond the course? Why or why not?	• Yes, I would continue with this approach (86%, 19) • The current approach would likely not work (5%, 1) • Change. I will try to annotate more in the future (5%, 1) • I will not annotate; I have another approach (5%, 1)	• I definitely will. It is an approach that I used before this course, and I will continue to use it! • It seems like some other "genres" of paper are especially dense and wordy, so this method probably wouldn't work as well. I like that biotechnology papers are usually pretty concise. • In the future I will try to annotate more of the text myself in the hopes that I will learn more information.
What was your approach for learning how to use KBase? What would you do differently?	• Followed tutorials (61%, 14) • Trial and error. (26%, 6) • Reviewed narratives again (17%, 4)	• I followed the tutorials on YouTube [...] • Trial and error, with the help of the tutorial. • I walked through the example narratives that were given to us and noted which tools they used and how. I replicated this workflow for my own analysis.
Do you think you can design your own narratives? Why and how?	• Yes (52%, 12) • No (22%, 5) • Maybe, but would require more time (22%, 5)	• I was able to construct my own narrative for the final project by examining the apps that were needed for the data I desired [...] • I think I could with more practice and watching YouTube tutorials of others doing it. • I think to some extent I can but not sure.
Do you feel more confident in using KBase and creating your own narratives for metagenomic analyses?	• Yes (70%, 16) • Not sure (12%, 3) • Not yet (9%, 2) • No change in confidence (4%, 1)	• Yes, I did not know what Kbase was before this class so I am definitely more confident in using it now. • For sure, as before I had no clue about any of these bioinformatics tools. • Not particularly, though I think I could take pieces from different tutorials to create my own.

^aQualitative analysis of student perceptions of bioinformatics case study scenarios reveals that participants learned to use the bioinformatics platform (KBase) and to analyze articles; however, they are not yet confident in designing their own workflows to address their research questions. Participant responses to four open-ended questions were coded using MAXQDA Analytics Pro Release 26.0 by CCG and reviewed by JW to revise codes and identify themes. Representative quotes and the percentage of responses to each question, along with the three or four most common codes, are listed below.

hesitant (Table 3), others noted, "Yes! I think the data analysis gave me confidence that I could give it a shot." Using MAXQDA, we conducted sentiment analysis of the open-ended self-reported confidence survey questions. The results indicated that four responses had positive sentiment (confidence), four responses were slightly positive, 10 responses were neutral, three responses were slightly negative, and two responses were negative. Analysis of student comments, particularly those still unsure about whether they can adapt KBase narratives to address their research questions or data sets, revealed that time, effort, and additional support were needed. This is an area of current and future research beyond the KBase portal: how can bioinformatics assignments provide authenticity and challenge while building confidence for the transfer of skills and tools to workforce development?

Possible applications and modifications

The sequence of activities in this course can be adapted to incorporate components of other microbiome analysis workflows, bioinformatics tools, and KBase narratives into various classes. Importantly, this approach combines critical reading with the application of bioinformatics tools to enhance relevance and foster learner confidence. Instructors

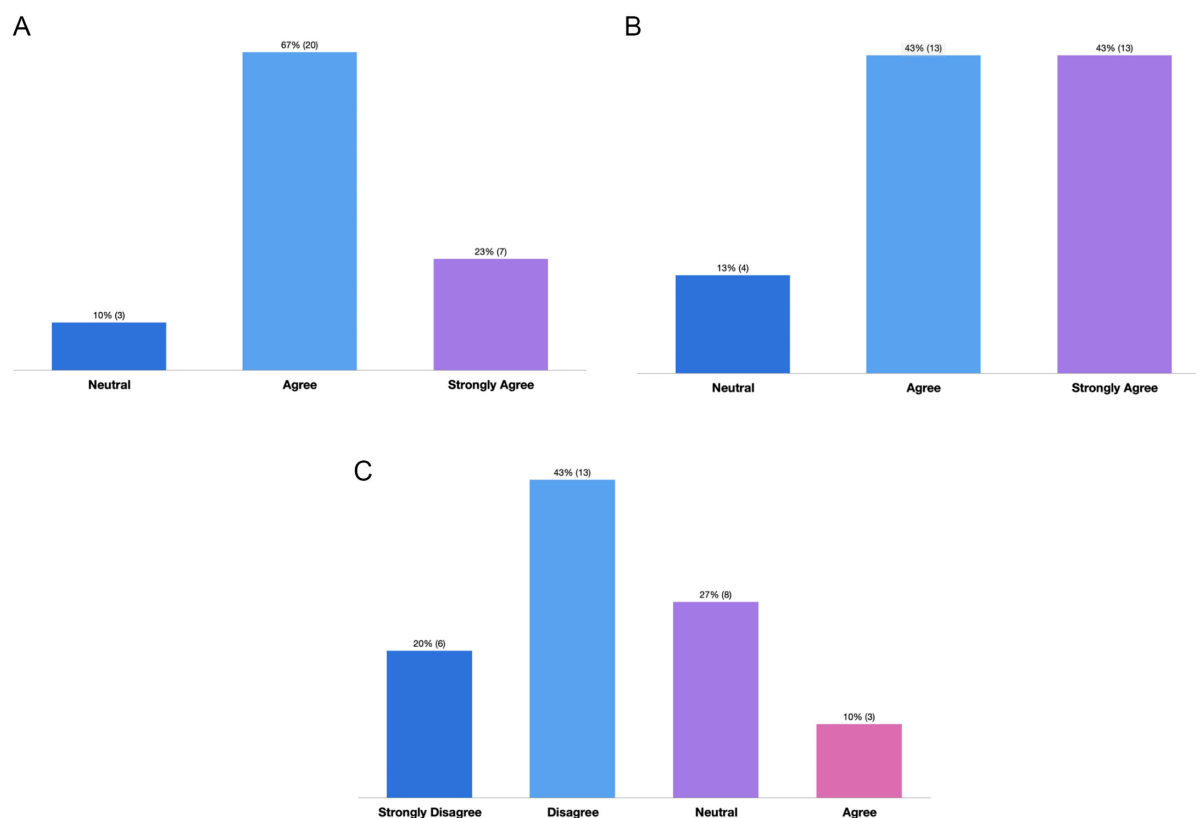


FIG 2 Analysis of post-survey responses suggests that participants value computational tools for addressing biological questions; however, they are hesitant about their ability to overcome challenges that require bioinformatics or computational biology approaches. Participants report they value the reasoning skills to understand bioinformatics (A) and the time devoted to studying concepts in computer science that can be applied to biological data (B). However, respondents report mixed self-confidence levels in their ability to overcome challenges with biology questions that involve the use of bioinformatics or computational biology (C). (A) Participant responses ($n = 30$) to the prompt "Reasoning skills used to understand bioinformatics can be helpful to my everyday life." (B) Participant responses ($n = 30$) to the prompt "It is a valuable use of my time to study some fundamental concepts in computer science that I may apply to computational analysis of biological data in the future." (C) Survey responses ($n = 30$) to the prompt: "If I get stuck on a biology question that involves the use of bioinformatics or computational biology, there is no chance I'll figure it out on my own."

may request that students submit their KBase narrative progress or responses to narrative-embedded critical thinking questions for credit. Class time can be devoted each week to discussing case studies and reading assignments, accompanied by quizzes to reinforce central concepts and prepare students for a summative assessment similar to the post-survey. We also encourage the use of reflections to prompt learners to share topics of interest or challenges as they utilize new computational tools. One area for future research is developing a module that teaches students to compare the binning methods used in the KBase narratives and reading assignments with more recent bioinformatics tools or sequencing approaches, such as long-read sequencing technologies.

ACKNOWLEDGMENTS

We are grateful for the time and engagement demonstrated by BIT 477/577 Metagenomics students over the last several semesters.

The Biotechnology Program at NC State University provided financial support for the materials used in this course. C.C.G. was supported by the Innovative Programs to Enhance Research Training (IPERT) funded by the National Institutes of Health (1R25GM130528-01A1). The National Science Foundation (NSF) Research Coordination Network for Undergraduate Biology Education (RCN-UBE): Microbiomes in Computational Research Opportunities Network (MICROnet) [#2418285] Adam Arkin (PI), Ellen G.

Dow (co-PI), and Elisha Wood-Charlson (co-PI) provided financial and technical support. C.C.G. was supported by the North Carolina State University Biotechnology Teaching Program (BIT). The design, implementation, and dissemination of educational case studies to learn about high-throughput technologies is part of the NSF-RCN HITS (award #1730317).

We are grateful to the students for their patience and commitment to learning and analyzing complex data sets. We thank the KBase Education Community, especially Ellen Dow, Elisha WC, and the summer 2020 education group, for helpful tips and productive conversations.

AUTHOR AFFILIATIONS

¹Department of Plant and Microbial Biology, North Carolina State University, Raleigh, North Carolina, USA

²Biotechnology Program, North Carolina State University, Raleigh, North Carolina, USA

³Department of Biological Sciences, North Carolina State University, Raleigh, North Carolina, USA

PRESENT ADDRESS

Jason M. Whitham, Takara Bio, San Jose, California, USA

AUTHOR ORCIDs

Jason M. Whitham  <http://orcid.org/0000-0003-2623-6292>

Carlos C. Goller  <http://orcid.org/0000-0002-2013-0334>

AUTHOR CONTRIBUTIONS

Jason M. Whitham, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review and editing | Carlos C. Goller, Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review and editing

ETHICS APPROVAL

The study was approved by the NC State University Institutional Review Board (IRB) protocol #24385.

ADDITIONAL FILES

The following material is available [online](#).

Supplemental Material

Supplemental material (jmbe00226-25-S0001.pdf). Appendices S1 to S8.

REFERENCES

1. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, Dehal P, Ware D, Perez F, Canon S, et al. 2018. KBase: the united states department of energy systems biology knowledgebase. *Nat Biotechnol* 36:566–569. <https://doi.org/10.1038/nbt.4163>
2. Wilson Sayres MA, Hauser C, Sierk M, Robic S, Rosenwald AG, Smith TM, Triplett EW, Williams JJ, Dinsdale E, Morgan WR, et al. 2018. Bioinformatics core competencies for undergraduate life sciences education. *PLoS One* 13:e0196878. <https://doi.org/10.1371/journal.pone.0196878>
3. Williams JJ, Drew JC, Galindo-Gonzalez S, Robic S, Dinsdale E, Morgan WR, Triplett EW, Burnette JM, Donovan SS, Fowlks ER, et al. 2019. Barriers to integration of bioinformatics into undergraduate life sciences education: a national study of US life sciences faculty uncover significant barriers to integrating bioinformatics into undergraduate instruction. *PLoS One* 14:e0224288. <https://doi.org/10.1371/journal.pone.0224288>
4. Goller CC, Srougi MC, Chen SH, Schenkman LR, Kelly RM. 2021. Integrating bioinformatics tools into inquiry-based molecular biology laboratory education modules. *Front Educ* 6:711403. <https://doi.org/10.3389/educ.2021.711403>
5. Drew J, Morgan W, Galindo S, Kleinschmit AJ, McWilliams M, Pauley M, Triplett EW, Williams J, Murdoch B, Rosenwald A. 2023. Revisiting barriers

- to implementation of bioinformatics into life sciences education. *Front Educ* 8:1317191. <https://doi.org/10.3389/feduc.2023.1317191>
6. Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, Solovyev VV, Rubin EM, Rokhsar DS, Banfield JF. 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* 428:37–43. <https://doi.org/10.1038/nature02340>
 7. Kang DD, Froula J, Egan R, Wang Z. 2015. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* 3:e1165. <https://doi.org/10.7717/peerj.1165>
 8. Parks DH, Rinke C, Chuvochina M, Chaumeil P-A, Woodcroft BJ, Evans PN, Hugenholtz P, Tyson GW. 2017. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol* 2:1533–1542. <https://doi.org/10.1038/s41564-017-0012-7>
 9. Almeida A, Mitchell AL, Boland M, Forster SC, Gloor GB, Tarkowska A, Lawley TD, Finn RD. 2019. A new genomic blueprint of the human gut microbiota. *Nature* 568:499–504. <https://doi.org/10.1038/s41586-019-0965-1>
 10. Meziti A, Rodriguez-R LM, Hatt JK, Peña-Gonzalez A, Levy K, Konstantinidis KT. 2021. The reliability of metagenome-assembled genomes (MAGs) in representing natural populations: insights from comparing MAGs against isolate genomes derived from the same fecal sample. *Appl Environ Microbiol* 87:e02593-20. <https://doi.org/10.1128/AEM.02593-20>
 11. Mirete S, Sánchez-Costa M, Díaz-Rullo J, González de Figueras C, Martínez-Rodríguez P, González-Pastor JE. 2025. Metagenome-assembled genomes (MAGs): advances, challenges, and ecological insights. *Microorganisms* 13:985. <https://doi.org/10.3390/microorganism13050985>
 12. Goller CC, Vandegrift M, Cross W, Smyth DS. 2021. Sharing notes is encouraged: annotating and cocreating with hypothes.is and google docs. *J Microbiol Biol Educ* 22:2135. <https://doi.org/10.1128/jmbe.v22i1.2135>
 13. Dow EG, Wood-Charlson EM, Biller SJ, Paustian T, Schirmer A, Sheik CS, Whitham JM, Krebs R, Goller CC, Allen B, Crockett Z, Arkin AP. 2021. Bioinformatic teaching resources – for educators, by educators – using KBase, a free, user-friendly, open source platform. *Front Educ* 6:711535. <https://doi.org/10.3389/feduc.2021.711535>
 14. Goller CC, Ott LE. 2020. Evolution of an 8-week upper-division metagenomics course: diagramming a learning path from observational to quantitative microbiome analysis. *Biochem Mol Biol Educ* 48:391–403. <https://doi.org/10.1002/bmb.21349>
 15. Garcia CB, Chapman IF, Chen SH, Lazear E, Lentz TB, Williams C, Dums JT, Goller CC, Robertson SD. 2021. Integrating research into a molecular cloning course to address the evolving biotechnology landscape. *Biochem Mol Biol Educ* 49:115–128. <https://doi.org/10.1002/bmb.21402>
 16. Liang X, Whitham JM, Holwerda EK, Shao X, Tian L, Wu Y-W, Lombard V, Henrissat B, Klingeman DM, Yang ZK, Podar M, Richard TL, Elkins JG, Brown SD, Lynd LR. 2018. Development and characterization of stable anaerobic thermophilic methanogenic microbiomes fermenting switchgrass at decreasing residence times. *Biotechnol Biofuels* 11:243. <https://doi.org/10.1186/s13068-018-1238-1>
 17. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z. 2019. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* 7:e7359. <https://doi.org/10.7717/peerj.7359>
 18. Mason OU, Scott NM, Gonzalez A, Robbins-Pianka A, Bælum J, Kimbrel J, Bouskill NJ, Prestat E, Borglin S, Joyner DC, Fortney JL, Jurelevicius D, Stringfellow WT, Alvarez-Cohen L, Hazen TC, Knight R, Gilbert JA, Jansson JK. 2014. Metagenomics reveals sediment microbial community response to Deepwater Horizon oil spill. *ISME J* 8:1464–1475. <https://doi.org/10.1038/ismej.2013.254>
 19. Porter GW. 2022. Collaborative online annotation: pedagogy, assessment and platform comparisons. *Front Educ* 7. <https://doi.org/10.3389/feduc.2022.852849>
 20. Lemanski MR, Venneman S. 2023. Student satisfaction with the perusal social annotation platform. *J Educ Online* 20. <https://doi.org/10.9743/JEO.2023.20.1.18>
 21. Spence PL, Goller C, Riley N, Norman D, Leggett Z. 2025. Where is Delftia? Implementing citizen science into the biotechnology and environmental science curriculum. *J Microbiol Biol Educ* 26:e0008824. <https://doi.org/10.1128/jmbe.00088-24>
 22. Riley NG, Goller CC, Leggett ZH, Lewis DM, Ciccone K, Dunn RR. 2020. Catalyzing rapid discovery of gold-precipitating bacterial lineages with university students. *PeerJ* 8:e8925. <https://doi.org/10.7717/peerj.8925>
 23. Madlung A. 2018. Assessing an effective undergraduate module teaching applied bioinformatics to biology students. *PLoS Comput Biol* 14:e1005872. <https://doi.org/10.1371/journal.pcbi.1005872>