Advances in Graduate Training in Integrative Bioinformatics for Investigating and Engineering Microbiomes (IBIEM)

Dr. Glenda T. Kelly, Duke University

Glenda Kelly is Director of Evaluation and Assessment for the National Science Foundation Engineering Research Center for Precision Microbiome Engineering (PreMiEr), Civil and Environmental Engineering, Duke University. She previously served as Associate Director for Assessment and Outreach for the Center for the Environmental Implications of NanoTechnology (CEINT), Duke. For both Centers she designed the overall evaluation plan and collaborated with leadership to define measurable goals related to training, diversity, inclusion and synergy across partners and research thrusts as well as impacts on institutional infrastructure. Since 2000, Dr. Kelly has worked in the area of STEM program development and evaluation for international research centers, NRTs, PIREs, REUs, GK-12 and faculty advancement. She served as evaluator for the Integrative Bioinformatics for Investigating and Engineering Microbiomes (IBIEM) NSF Research Traineeship (NRT) program (2015-21) and was instrumental in providing feedback to optimize this training model. She has chaired evaluation sessions at NSF NRT and other NSF meetings. She has special interest in developing training to optimally engage students under-represented in STEM, across disciplines and with differing entry level skill sets. Dr. Kelly served as a founding member of the Pratt School of Engineering's Diversity and Inclusion Committee and earned her M.A. and Ph.D. degrees from Duke University with a concentration in development of assessment tools.

Dr. Joshua A. Granek, Duke University

Josh Granek is an Assistant Professor of Biostatistics and Bioinformatics at Duke University. He joined the Duke Faculty in 2012 after postdoctoral positions at Duke and the National Evolutionary Synthesis Center, receiving his Ph.D. from Johns Hopkins University, and his Sc.B. from Brown University. He is Education Co-Director for the National Science Foundation Engineering Research Center for Precision Microbiome Engineering (PreMiEr) and Associate Director for Microbiome and Genomics of the Duke Center for AIDS Research Quantitative Sciences Core. He was previously co-PI and Instructor for IBIEM (Integrative Bioinformatics for Investigating and Engineering Microbiomes), a joint graduate training program between Duke and North Carolina A&T State University (2015-2021). Dr. Granek's research is primarily in Bioinformatics with focuses in microbiome, microbial genomics, cancer genomics and immunoinformatics.

Dr. Claudia K. Gunsch, Duke University

Claudia Gunsch is a Professor of Civil and Environmental Engineering at Duke University and holds secondary appointments in the Nicholas School of the Environment and the Department of Biomedical Engineering. She joined the Duke Faculty in 2004 after obtaining her PhD from the University of Texas at Austin, her MS from Clemson University and her BS from Purdue University. Currently, she serves as the Director for the National Science Foundation Engineering Research Center for Precision Microbiome Engineering (PreMiEr). Previous leadership roles include serving as the Director of IBIEM (Integrative Bioinformatics for Investigating and Engineering Microbiomes), a joint graduate training program between Duke and North Carolina A&T State University (2015-2021). Dr. Gunsch's research bridges environmental engineering and molecular biotechnology. Current research foci include developing microbiome engineering approaches for the built environment and bioremediation, investigating the ecological impacts of emerging contaminants on environmental microbiomes, studying microbial evolution following exposure to anthropogenic contaminants and developing innovative water treatment technologies. Dr. Gunsch was named ASCE Environmental & Water Resources Institute Fellow in 2022. She currently serves as Editor in Chief for Biodegradation and is a member of the Editorial Board for npj Clean Water and Industrial Biotechnology.

Dr. Joseph L. Graves Jr., North Carolina A&T State University

Joseph L Graves, Jr. received his Ph.D. in Environmental, Evolutionary and Systematic Biology from Wayne State University in 1988. In 1994 he was elected a Fellow of the American Association for the

Advancement of Science (AAAS) and in 2012 chosen as one of the ”Sensational Sixty” commemorating 60 years of the NSF Graduate Research Fellowship Award. In 2017, he was listed as an ”Outstanding Graduates” in Biology at Oberlin College; and named ”Innovator of the Year” in US Black Engineer Magazine. He was named ”Senior Researcher of the Year” at North Carolina A&T State University in 2023. He leads programs addressing underrepresentation of minorities in science. His research in the evolutionary genomics of adaptation shapes our understanding of biological aging and bacterial responses to nanomaterials. He is presently Associate Director/co-PI of the Precision Microbiome Engineering (PreMiEr) Engineering Research Center of Excellence (Gen-4 ERC) funded by the National Science Foundation (2022—2027). He also was co-PI for the IBIEM (Integrative Bioinformatics for Investigating and Engineering Microbiomes), a joint graduate training program between Duke and North Carolina A&T State University (2015-2021). His five books include: A Voice in the Wilderness (New York: Basic Books), 2022; with Alan Goodman, Racism, Not Race (Columbia University Press), 2022; Principles and Applications of Antimicrobial Nanomaterials, (Amsterdam NE: Elsevier), 2021; The Emperor’s New Clothes, (Rutgers University Press), 2005 and The Race Myth (Dutton Press), 2005.

Dr. David Singleton, Duke University

David Singleton is an Executive in Residence in the Department of Civil and Environmental Engineering in Duke University’s Pratt School of Engineering. He also hold an adjunct faculty appointment in the Department of Environmental Sciences and Engineering at the University of North Carolina at Chapel Hill in the Gillings School of Global Public Health. He holds a PhD in Microbiology from the University of Georgia. His prior research focused on the bioremediation of contaminated soils and the cultivation and description of novel microorganisms. Currently, he serves as the Administrative Director for the NSF Engineering Research Center for Precision Microbiome Engineering at Duke University.

Advances in Graduate Training in Integrative Bioinformatics for Investigating and Engineering Microbiomes (IBIEM)

Abstract

Innovations by engineers and physical scientists working at the frontiers of microbiome engineering and discovery requires in-depth understanding of microbiome systems with parallel skills in bioinformatics and biostatistics. Despite the importance of integrating bioinformatics and biology into graduate student training in fields outside traditional biological sciences, academic institutions remain challenged with including these disciplines across departmental boundaries. Furthermore, it is critical for students in engineering, bioinformatics, and biostatistics to understand fundamentals behind the biological systems they model, and for biology students to gain competencies in applying bioinformatics and biostatistics to biological questions. To address these needs, we developed the *Integrative Bioinformatics for Investigating and Engineering Microbiomes* (IBIEM*)* graduate training partnership between Duke University and North Carolina Agricultural and Technical State University, which was funded by the National Science Foundation Research Traineeship (NRT) program. IBIEM's goals include training interdisciplinary groups of students to: (a) transform conceptualization and develop skills for application of quantitative biology in microbiome areas; (b) perform cutting edge research requiring interdisciplinary team skills; and to (c) communicate their research across disciplinary barriers and to diverse audiences. The pedagogical framework adapted to foster trainee engagement is learner-centered teaching which emphasizes the importance of self-directed learning with parallel ongoing assessment to optimize student outcomes. Since IBIEM trainees' goals as well as entry-level knowledge and skills across disciplines varied greatly, program implementation was found to be challenging and required rigorous evaluation and refinements for effective training across disciplines and skill levels. A comprehensive program evaluation over five years found that the strongest learning and skills outcomes were linked to several "best practices". Early provision of *depth in fundamentals in R programming and reproducible research* was found to be critical to "jump start" students without programming backgrounds. Addition *of an overview of microbiome experimental design and analysis* added important context as to how and where in the research process informatics fits into design progression and was highly motivating to students. *Course modality was found to impact trainee outcomes* with in-person classes that included hands-on practice and feedback showing greater improvements in training outcomes over hybrid, flipped and virtual course modalities. Furthermore, introduction of *low, medium, and high level "challenges"* along with *in-person tutoring* was found to be impactful in building a common foundation to span expertise levels and for engaging students across entry and advanced levels. Training impacts peaked during year four with cumulative implementation of revised strategies. Innovative training revisions and inclusion of critical elements was strongly linked to program satisfaction and ratings of advances in technical, professional and career skills as well as post-training carry over into trainees' own research and leadership in their labs and careers. Furthermore, this training collaboration and partnership provided the foundation and training model for the newly funded NSF Engineering Research Center for Precision Microbiome Engineering (PreMiEr**)** for work in the critical area of engineering the microbiome in built environments.

Introduction

In the last decade, rapid advances in DNA sequencing technology have transformed the biological sciences [1]. It has become essential for students training in biological disciplines requiring metagenomic and metatranscriptomic analyses to have a working knowledge of bioinformatics and biostatistics. Conversely, it is critical for students training in bioinformatics, biostatistics, and engineering to understand fundamentals behind the biological systems they model and analyze. Despite the importance of integrating bioinformatics and biology, academic institutions are challenged to keep pace with growing training demands in these areas. Current educational models do not effectively promote integrated training across these disciplines. Students in biology often lack the appropriate training in bioinformatics and biostatistics to work in new areas of quantitative biology [2], [3]. Similarly, increasing numbers of engineers and physical scientists are aware that transformative discoveries in the microbiome arena require in-depth understanding of biological phenomena. Existing graduate training tends to target singular disciplinary components, such as programming or molecular biology, rather than developing broader cross disciplinary competencies. Relevant academic units primarily focus on the needs of their within-discipline students, and thus limit development of interdisciplinary competencies essential to addressing society's complex grand challenges, which require the convergence of multiple disciplines and effective two-way engagement with stakeholders and communities [4], [5].

Thus, graduate students with strong quantitative skills, often find it difficult to master the complex information that defines the working knowledge of biologists, and biologists without backgrounds in programing and computer sciences often find entry into bioinformatics and biostatistics department courses challenging. This National Science Foundation Research Traineeship (NRT) program was designed to address this training gap and transcend communication barriers between disciplines while promoting team science through creation of an integrated inter-disciplinary educational model that reflects rapid advances in microbiome research and the need for both interdisciplinary research and professional skills to address these challenges [6]. This paper reports on the evaluation of this project over five years with a focus on challenges identified in training graduate students with different entry level skills and across disciplines. Strategies and training elements implemented to successfully address these challenges were made possible through close collaboration between the evaluation team and project leadership who were highly responsive to evaluation feedback.

Program Overview

The overarching goal for this graduate training program was to develop a novel, replicable interdisciplinary training model that is adaptive to the rapidly changing landscape in microbiome research and to students with varying training backgrounds to develop students who are capable of: (a) performing cutting edge microbiome research and/or engineering and (b) communicating their research across disciplinary barriers and diverse audiences.

The specific training objectives for the IBIEM program were to:

1) Enable trainees to actively engage in microbiome research across biological sciences, engineering, and biostatistics;

2) Cultivate students who are confident at team science and excel in collaborative settings;
3) Develop students who are comfortable communicating their research to diverse audiences;
4) Increase trainees' awareness of non-academic professional opportunities;
5) Provide targeted interactions with practitioners who can provide real-world perspectives; and
6) Increase underrepresented minority and female enrollment in microbiome related fields and promote an academic and social environment where graduate trainees can flourish.

Research Training Opportunities

The research areas of this NRT spanned three interrelated Core Areas (Figure 1): Biological Sciences (CA1), Engineering (CA2) and Biostatistics &Bioinformatics (CA3). As shown below (Table 1) , research bridged core areas through a systems biology approach that ranged from analysis of molecular interactions in simple and defined microbial communities (MI), to systems analysis of complex microbial systems (CMS).

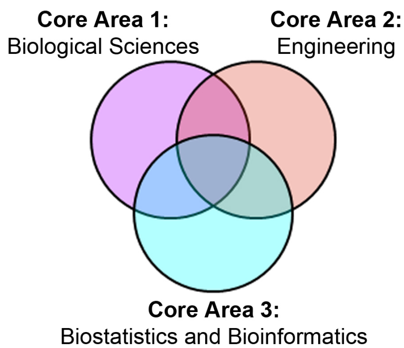


Figure 1. Core research areas

Table 1. Range of NRT research by core areas

	Molecular Interactions (MI) ⟷	Complex Microbial Systems (CMS)
Biological Sciences (CA1)	Identification of key molecular mechanisms of microbial interaction	Study of complex natural communities, and development of model synthetic communities
Engineering (CA2)	Development of strategies for controlling gene expression, activity, and transfer in simple microbial communities	Manipulation of complex microbial community structure and function for a particular environmental or biomedical process
Biostatistics and Bioinformatics (CA3)	Statistical design, analysis and modeling of multi-omic datasets for simple microbial communities	Bioinformatics integration and statistical design, analysis and modeling for complex microbiomes incorporating spatial and dimensional heterogeneity

Trainees were provided opportunities to pursue training across these cores by engaging in interdisciplinary team research projects in these areas:

Core Area 1 (CA1): Biological Sciences -***Exploring the Molecular Mechanisms, Function, and Evolution of Microbial Communities.*** Microbial communities are complex, dynamic, and serve critical roles in animal, plant, and environmental health. The ability to effectively maintain or manipulate microbial communities in these settings is constrained by the limited understanding of mechanisms underlying interactions between members of different microbial lineages. Training within the biological sciences focused on elucidating the genetic, molecular, cellular, and physiological mechanisms that govern interactions between microbes.

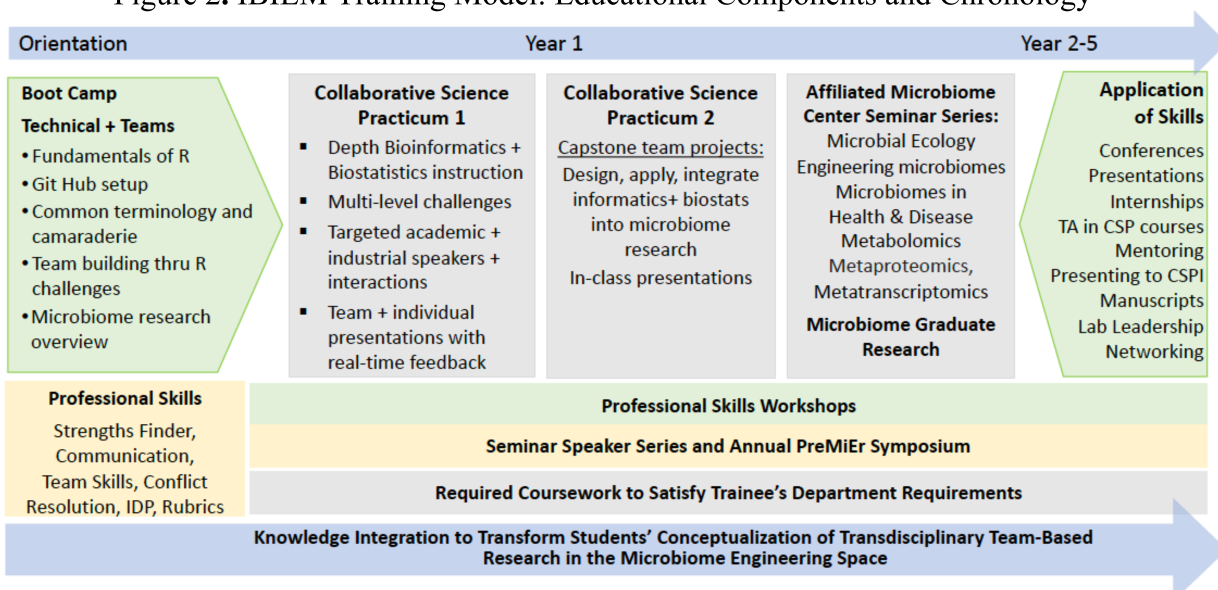
Core Area 2 (CA2): Engineering - ***Manipulating Microbial Communities.*** Research in CA2 built upon findings from CA1 and applied them to engineer microbial communities. Research training within this core was focused on manipulating microbial communities to produce desired functions and behaviors for downstream applications.

Core Area 3 (CA3): Biostatistics and Bioinformatics – ***Novel Tools for Understanding Microbiome Structure and Function.*** Training in biostatistics and bioinformatics were a major focus in the CSP I and II practica (described in section below) and developed trainee skills to work towards developing, optimizing, and implementing novel bioinformatic tools and statistical approaches to define microbial community composition and activity, and identify genetic variants, keystone species, and interaction networks.

Training Model: Components & Pedagogy

The formal IBIEM educational components were comprised of a (a) Boot Camp orientation prior to fall semester; (b) two-semester consecutive Collaborative Science Practica (CSP) and (c) a recurring Speaker Series integrated into the CSP class schedule. These formal training elements were supplemented by professional skill training sessions, and a semi-annual symposium as well as a seminar series (Figure 2 below).

Figure 2**.** IBIEM Training Model: Educational Components and Chronology



Boot Camp. IBIEM'S Boot Camp was primarily held on Duke University's main campus with a subset of activities on the campus of N.C. A&T for the first four years. During the final year it was delivered virtually due to the COVID pandemic. The primary objectives were: 1) to immerse trainees from diverse disciplinary backgrounds (e.g., biological sciences, engineering, computer science and mathematics) into a common fundamental knowledge base and vocabulary spanning the core research areas; and 2) to enable trainees to interact and build a cohort across disciplines as a foundation for entry into the fall collaborative science practica course.

For startup year, Boot Camp lasted 10 days. Technical topics (microbiome and informatics) were taught week one followed by professional skills week two. A major theme in student feedback from the year one Boot Camp evaluation was a request for *more in-depth focus on informatics training with less topical variation in the microbiome related areas*. Accordingly, second year Boot Camp was revised to provide more depth and focus on R programming, along with High-Throughput Sequencing and amplicon sequencing. To balance the second year Boot Camp experience, professional skills training was offered in the afternoons. Due to the *exceptionally strong positive ratings of the professional skills training* components during startup year, we retained the majority of these components for future Boot Camps. Professional skills topics included, Building Dynamic Research Teams, Strengths Finder (identifying/leveraging), Impactful Communication of Science and Mentoring & Aligning Expectations. A second Boot Camp objective was to *develop a team identity and cohesiveness* among trainees. This goal was facilitated by the interactive skill building topics and hands-on exercises on team building and communication in which both trainees and leadership participated.

Collaborative Science Practica (CSP I and II) core courses were primarily taught by a Biostatistics & Bioinformatics faculty member along with integration of the Seminar Speaker Series taught by a range of faculty and partners from the private sector and other universities. Topics covered a breadth of research relevant to trainees' projects and beyond. Each year trainees were assigned to work in teams assembled in such a way that one trainee from each research core area (Biological Sciences, Engineering and Biostatistics & Bioinformatics) was represented. Each program year from three to five interdisciplinary research teams were formed to conduct capstone microbiome research projects as part of this CSPII course. CSP practica were cross listed between Duke and N.C. A&T as for-credit courses.

<u>CSP I</u> was originally *designed to cover breadth of research areas rather than depth*. We redesigned this course year two to *provide more informatics depth and in-class challenges based on evaluation feedback* (see Revisions section below). A secondary goal for CSP I was to expose students to careers outside of academia by bringing external speakers into the classroom.

<u>For the Spring CSP II</u> course, one primary goal was to provide students with practice utilizing skills acquired in CSPI by *immersion in interdisciplinary microbiome related team research projects to allow application and deepening of skills.* Starting year two, in-class time was increased for trainees to work with their TA or faculty mentors and team members on design, data analysis, and to discuss roles and assignments. An array of workshops was also interspersed throughout the course to develop trainees' skills in research presentations (e.g., data visualization, statistical analysis as well as formal and informal presentation skills and practice).

Seminar Speaker Series. This series was designed to provide an overview of how scientists and engineers from an array of career sectors chose their career paths as well as to gain insights into challenges and benefits of working across different settings. Secondly, speakers also discussed their areas of research or engineering thus providing trainees with broader exposure to an array of microbiome connected projects. Speakers' backgrounds ranged from local industries to museum outreach educators. Seminar topics were related to research project themes and time was allotted for speakers' Q&A with trainees.

Pedagogy. Learner-centered teaching was the pedagogy woven into this training model since it has been found to promote engagement and shared ownership in the learning process [7] and to foster "construction" of new conceptual models via interactions between learner's experiences and adaptation of internal conceptual models as described in constructivist theory [8]. This model differs from more traditional instruction in which instructors assemble learning materials and provide one-way lectures with questions directed towards "correct answers". In a learner centered model, *instructors provide a variety of interactive learning experiences, and skill application opportunities which are essential to mastering and integrating material* [9]. This model provided students with opportunities for active engagement which contributed to the learning process through opportunities to apply skills and concepts which has been found to promote retention and transfer and was particularly suited to training in informatics tools and analyses which requires iterative application, revision, and re-application for mastery. This model also promoted collaboration as students formed teams to consider, reflect, and apply new information and skills**.** Literature on the efficacy of "learner-centered teaching" for student training emphasizes the importance of incorporating learner-centered assessment at multiple time points during training, which was built into this NRT and critical to identifying the challenge areas and needed revisions describe in this paper [10]-[13].

Evaluation Questions

The current paper focuses on a subset of formative and summative evaluation questions that includes: 1) identification of revisions in training elements necessary to improve outcomes in three key areas: technical, career path development, team science skills (collaboration and communication); and 2) satisfaction with mentoring and support from core faculty and TAs. IBIEM's overall evaluation is described below with questions addressed in this current paper underlined below.

IBIEM's comprehensive evaluation included both ***formative*** and ***summative*** elements. The ***formative*** evaluation assessed the extent and ways in which IBIEM:

- Developed an integrated interdisciplinary training program blending engineering, biological sciences, biostatistics & bioinformatics;
- Developed program elements that were rated as high quality linked with training objectives;
- Provided multiple opportunities for developing communication skills across audiences;
- Recruited and retained trainees, including those from underrepresented groups;
- Faculty and mentors supported trainees in ways that helped them develop technical and professional skills and confidence; and
- Activities promoted collaboration and team science and were implemented as proposed.

The IBIEM ***summative*** evaluation addressed the ways and extent to which:

- Training led to advances in trainees' conceptualization and application of project's interdisciplinary methodological skills;
- Trainees demonstrated necessary skills to engage in interdisciplinary research design and analysis incorporating biostatistics and bioinformatics with focus on microbiome areas;
- Trainees demonstrated improved team collaboration and communication skills;
- Faculty incorporated training elements into their courses;
- Faculty collaborated on/received new research grants with other faculty across disciplines;
- Participants disseminated findings across disciplines; and
- Institutions integrated and sustained project elements over time.

Data sources for formative and/or summative components were comprised of:

- Project records and database to include products and dissemination
- Student and faculty informal interviews
- Direct observations of Boot Camp, classes, and student presentations

- Peer and faculty-rated rubrics to assess trainees' informal communication training assignments
- Student self-rated *Individual Learning Goal Rubrics* designed to assess: trainee progression along a series of self- identified goals that include communication and collaboration in teams; integration of informatics and biostatistics knowledge and skills application into research; and satisfaction with progress towards their own professional goals.
- Four student surveys:
  - Post- Boot Camp Survey
  - IBIEM Fall Student Survey
  - Annual End-of Year Student Survey
  - Trainee Follow-up Surveys to track trainees after their first year of core courses.
- Annual Faculty Survey

Training Challenges and Revisions: based on formative evaluation over five years

The extent to which IBIEM developed program elements that are high quality and contributed to trainees technical and professional skills development was assessed each year along with satisfaction with the learning environment. Data sources included informal trainee interviews, and three of the post training student surveys (previously described) to provide feedback on satisfaction and degree to which elements contributed to professional and/or technical gains along with recommendations for program improvements**.** Each year revisions were made in Boot Camp as well as the Collaborative Science Practica course content and modalities, and research training based on triangulation of data from these sources, input from faculty surveys, and direct class observations by the internal evaluator. Below is a summary of changes made over the five years of the program which were linked to increases in trainee outcomes that are described in the next section.

Overall revisions were made in four main areas: 1) start-up year challenges; 2) challenges inherent in training students across disciplines with differing entry level skill sets; 3) research training to provide sufficient background and mentoring; and 3) optimal course modality for training across disciplines and cohorts spanning two universities.

Startup year revisions for CSP practica course based on mid-year evaluation.
Not surprising, the strongest need for training adjustments were found year one. Mid-year student interviews by the evaluator as well as mid-year trainee completion of rubrics rating individual learning goal progressions revealed a *wide range in levels of understanding and satisfaction across students with their ability to use and understand the currently used software platforms (Jupyter)* and" *to code data" to run visualization and statistical analysis programs*. While most students had learned some components of programming to run data visualization software, several were confused about exactly what the programs meant and how to innovate and change code on their own. Several also indicated a lack of understanding of the *"big picture" data pipeline* and *"what was under the hood"* of the platform. This feedback was given to the PI and instructors. Mid-course corrections were made for the Spring CSPII Practica to provide more basic instruction and in-class mentoring by instructors to allow for a more satisfactory learning experience, particularly for students with less informatics backgrounds. In addition, a **"**broad picture" overview of where bioinformatics fits into research design and the data pipeline as well

as provision of a syllabus with detailed learning objectives and timeline beginning Spring 2017 resulted in more positive ratings of the Spring course. Other revisions from the 1st two years are described below.

Additional revisions to address challenges (years one and two).

In year two Boot Camp, the breadth of technical topics was narrowed with increased focus on introduction to R, RStudio (the new platform), and reproducible research methods, along with more hands-on practice. This revision was critical to "jump start" students without programming backgrounds so they could move into fall CSPI informatics training with abilities to use the platform. A platform change from Jupyter to R-studio was made year two. Jupyter was found to be challenging for trainees with no computer science background to master in a limited time.

Training across different entry level skills.

i. Addition of *"informatics challenges*" in teams early in boot camp as well as rotating *"challenge chairs"* after each challenge enhanced team building across the whole cohort and facilitated role and task rotations across members spanning different disciplines.

ii. Year two Boot Camp evaluation found that students gained greater perspectives on solving different challenges by working with students from different discipline backgrounds and that the "*challenge rotations*" encouraged students to interact across the whole cohort of trainees which provided a foundation to share skills and interests more quickly prior to fall course.

iii. In-class and homework low, medium, and high "challenge" levels helped span the range of expertise to engage students across levels.

iv. Alternating in-class brief *lecture with iterative hands-on practice* along with *instructor and TA support*, and *in and out of class "office hours*" worked best *across skill levels and disciplines*. Year four, both entry level and more advanced students stayed after class, particularly during Spring when data analysis for research projects was underway. The "real time" opportunity to ask questions stimulated by in-class applications accelerated the pace of trainees' abilities to work independently.

v. TAs for the course were increased to four by year five which allowed more one-on-one mentoring. TAs were hired from trainees who had already completed core IBIEM courses.

vi. Poll-Ev pop-ups were added to in class "lectures" for real-time anonymous feedback on understanding across all class members so the instructor could adjust focus to areas needing more clarification.

vii. By year two, the majority of class instruction materials and topics were available online so students could *access training materials in advance* which accelerated the learning pace for students with more advanced skills and provided students with entry level skills scaffolding to understand the "big picture" of training. This early online access to training material was especially critical in year five during all online instruction due to COVID instruction policy.

Optimal course modality for training across disciplines and spanning two universities.

Year one, the CSP Practica met twice per week at Duke, with N.C. A&T students accessing the 2nd class virtually. Issues with internet connectivity for students accessing remotely were reported with requests for more in-person time with instructor and TAs. Thus, for the remaining years, this course was held once weekly with a longer class. Lunch after class was provided during the Speaker Series to minimize travel for N.C. A.&T students and maximize access for all trainees for in-person networking across career sectors (except for year five, when instruction for this course was changed to completely remote modality due to COVID restrictions).

Year three, flipped classroom modality was introduced in fall based on prior year requests. For students without coding backgrounds, feedback on flipped classroom and online homework indicated that "google was not their friend" for learning R and bioinformatics. Students with no coding experience reported spending excess amounts of time on assignments without sufficient progress. More in-class walk-thru application and guidance was added back into training during spring and was found to be critical to building confidence and perseverance.

Additions to enhance research training.
Early provision of an overview seminar on *microbiome analysis and experimental design considerations* was requested by trainees and added to Boot Camp, providing important context as to how and where in the research process informatics fit and the progression of microbiome research which was highly motivating to students. This seminar continued in the fall semester for greater depth as requested by trainees. The faculty who taught this seminar was also available to give feedback during preliminary and year-end research presentations.

Year one, trainees requested more time with "expert" faculty mentors on research projects rather than primarily relying on team members for guidance in formulating key areas of research selection and later for critical feedback as research progressed. So increased time was allocated for "expert" mentors to work with students.

Addition of these collective refinement strategies was strongly linked to program satisfaction and increased ratings of trainees' technical, professional and career skill development as well as application to trainees' own research and leadership in their home labs and careers, subsequent to participation. A subset of trainee outcomes over five years are described below.

Findings

Results related to trainee outcomes.
Trainee's self-rated *outcomes were strongest in year four* with outcomes linked to revisions in delivery modalities and training components designed to better address different entry levels and research training backgrounds as described in previous section. Cumulative revisions of this training model over four years contributed to these stronger outcomes and include additions of: 1) seminars on microbiome research design progression and considerations, 2) immediate after-class office hours with instructor and TA mentorship; 3) more hands-on applications opportunities along with multi-level challenges; 4) full access to online training material from the outset; and 5) increased faculty and peer support opportunities.

Training modality was found to be important. For years two and four in which training was in-person, stronger gains were found across Technical, Team Science and Career Path areas as well as perceptions of increased peer and faculty support than for year three (flipped class modality), and year one (hybrid modalities) and also than for year five (when all online instruction was used due to COVID restrictions).

When trainee ratings of Technical Skills, Team Science and Career Path_advancements for year four were compared with years one and three ratings (Table 2 below), 100% of year four ratings showed higher absolute values than for years one and three. Of these 87.5% of trainee ratings were significantly higher for year four ($p \leq 0.05$ two-tailed) than startup year ratings, with another 64.7% of ratings significantly higher than year three outcomes.

Year five of the IBIEM program (2020-21), had a major switch in instruction modality to ALL *online instruction*, which was necessary to comply with university safety policies during the early part of the COVID pandemic. Trainees reported missing in-person peer interactions, with *usefulness of peer feedback for understanding research* rated significantly lower (3.0 on 1 to 4 scale, $p \leq 0.05$) during this all-online year than for year four (in-person model (3.69 , $p \leq 0.05$). However, *ratings of receiving adequate IBIEM faculty support for learning goals during Covid conditions* (3.69 on 1-4 scale) was the *highest rated training component that year* and identical to high ratings of faculty support for year four (in-person modality). On questions that allowed open ended feedback, trainees cited outstanding support from both the informatics instructor and the TAs who provided out of class tutoring that was linked to advances in learning and skills application during this challenging all virtual socially distanced year.

When trainee ratings in overall gains in Technical Skills year five were compared with year four gains (highest rated year), absolute value ratings were slightly but not significantly lower except for *understanding statistical application complexities in microbiome areas* which was significantly lower (3.31 on 1 to 4 scale, $p \leq 0.05$) than highest rated year four (3.75).

Table 2. Trainee's Self-Rated Technical, Career, Team Science and Perceived Support Outcomes Across 2016-21 Training Revisions

	TRAINEE OUTCOMES Significant p ≤ 0.05 two-tailed t-test* *(Scale: 1=Strongly Disagree, 2=Disagree, 3=Agree, 4=Strongly Agree)*	**Hybrid Subset** YR 1 16-17 N=16	**In person** YR 2 17-18 N=10	**Flipped** YR 3 18-19 N=15	**In person** YR 4 19-20 N=13	**Online (Covid)** YR 5 20-21 N=17
TECHNICAL SKILLS	Increased ability to structure, organize, visualize microbiome related data. *(Yr2, 4 and 5 all > Yr1)	3.13	3.80	3.53	**3.85**	3.63
	Greater understanding of statistical application complexities of microbiome research projects. *(Yr2, 3, 4, 5 all > Yr1; Yr4> Yr5)	2.69	3.50	3.33	**3.75**	3.31
	More confidence in methods/ skills in microbiome research. *(Yr. 2, 3, **4** and 5 all > Yr1)	2.60	3.50	3.27	**3.58**	3.38
	Can select/use new statistical tools for analysis in microbiome research. *(Yr4>Yr1; Yr5>Yr1; Yr4>Yr3)	2.53	3.20	3.00	**3.67**	3.25
CAREER PATH	Exposed to fields outside of academia. *(Yr2 and Yr4>Yr5)	3.31	3.70	3.27	**3.62**	3.13
	Increased professional network across disciplines. *(Yr2> Yr1 and 3)	2.93	3.50	3.00	**3.38**	3.06
	Increased professional network outside of academia. *(Yr4> Yr1)	2.47	3.10	3.00	**3.15**	2.81
	Increased awareness of skills needed for career outside of academia. *(Yr2 >Yr1; Yr4 > Yr1, 3 and 5)	2.88	3.50	2.93	**3.77**	3.19
TEAM SKILLS	Greater awareness what I contribute to team science culture. *(Yr4>Yr1)	2.93	3.10	3.13	**3.62**	3.25
	Improved communication skills in the microbiome research area with other professionals in my field. *(Yr2, 4 and 5 all >Yr1; Yr4>Yr3; Yr4>Yr5)	2.93	3.40	3.07	**3.69**	3.31
	Improved communication skills with professionals across disciplinary boundaries. *(Yr4 and Yr5>Yr1; Yr4>Yr3)	2.80	3.20	3.00	**3.67**	3.38

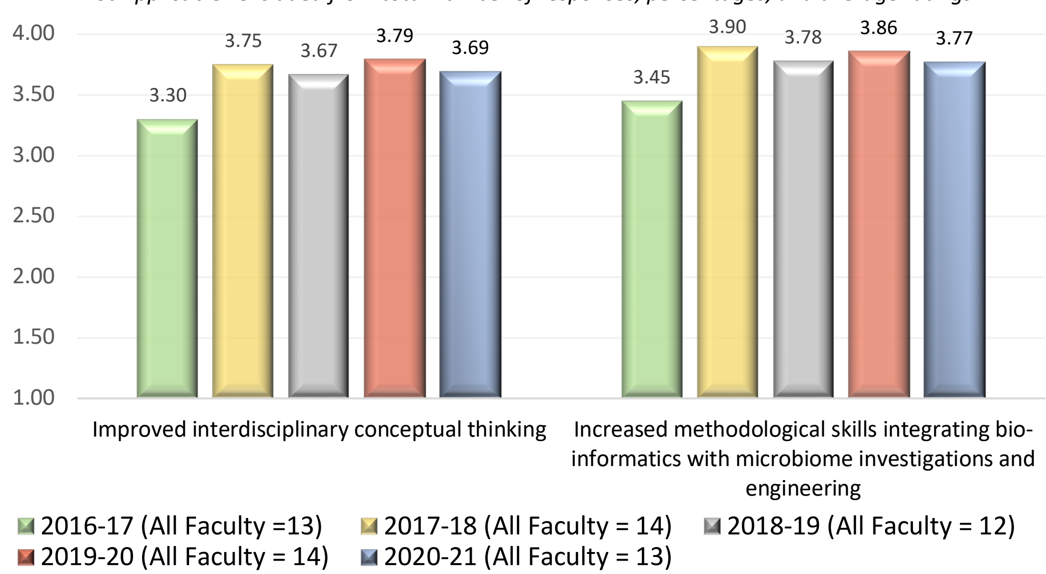
TRAINEE OUTCOMES Significant p ≤ 0.05 two-tailed t-test* *(Scale: 1=Strongly Disagree, 2=Disagree, 3=Agree, 4=Strongly Agree)*		Hybrid Subset YR 1 16-17 N=16	In person YR 2 17-18 N=10	Flipped YR 3 18-19 N=15	In person YR 4 19-20 N=13	Online (Covid) YR 5 20-21 N=17
	Enhanced collaboration skills with scientists across different disciplines. *(Yr2, 4 and 5 all >Yr1; Yr4 > Yr3)	2.94	3.40	2.93	**3.55**	3.33
FACULTY & PEER SUPPORT	Received adequate IBIEM faculty support for learning goals. *(Yr2, 3, 4 and 5 all >Yr1; Yr4 and Yr5 >Yr3)	2.50	3.60	3.13	**3.69**	3.69
	IBIEM mentor(s)/peers actively involved in my training/learning experiences. *(Yr2, 4 and 5 all > Yr1; Yr2, 4 and 5 all > Yr3)	2.63	3.60	2.93	**3.85**	3.31
	Received adequate support from my team for learning goals and research. *(Yr2 and 4> Yr1; Yr4 >Yr3)	2.87	3.40	2.93	**3.75**	3.38
	Peer feedback useful and informative to understand integrate new ways of thinking about my research. *(**Yr4 > Yr1, 3 and 5**)	2.86	3.40	2.86	**3.69**	3.00

Advances in trainees' conceptualization and application of interdisciplinary informatics, statistical and methodological skills in the microbiome research space.

Trainees' self-ratings combined with faculty ratings of trainees offer strong evidence to support advances in *conceptualization and application of project's interdisciplinary methods and skills* as a result of participation. As shown in Table 2, trainees' average ratings for year four (strongest year) ranged from range from 3.67 (1 to 4 scale) *for their ability to select and use new statistical tools* for data analysis to 3.85 *for improved understanding of how to structure, organize, and visualize microbiome related data.*

Consistent with highest self-ratings by trainees for technical skills (Table 2 above) for years two and four, faculty also gave highest absolute value ratings for these years (Figure 3 below) for trainees *improved conceptual* thinking and *increased methodological skills integrating bioinformatics into microbiome investigations*.

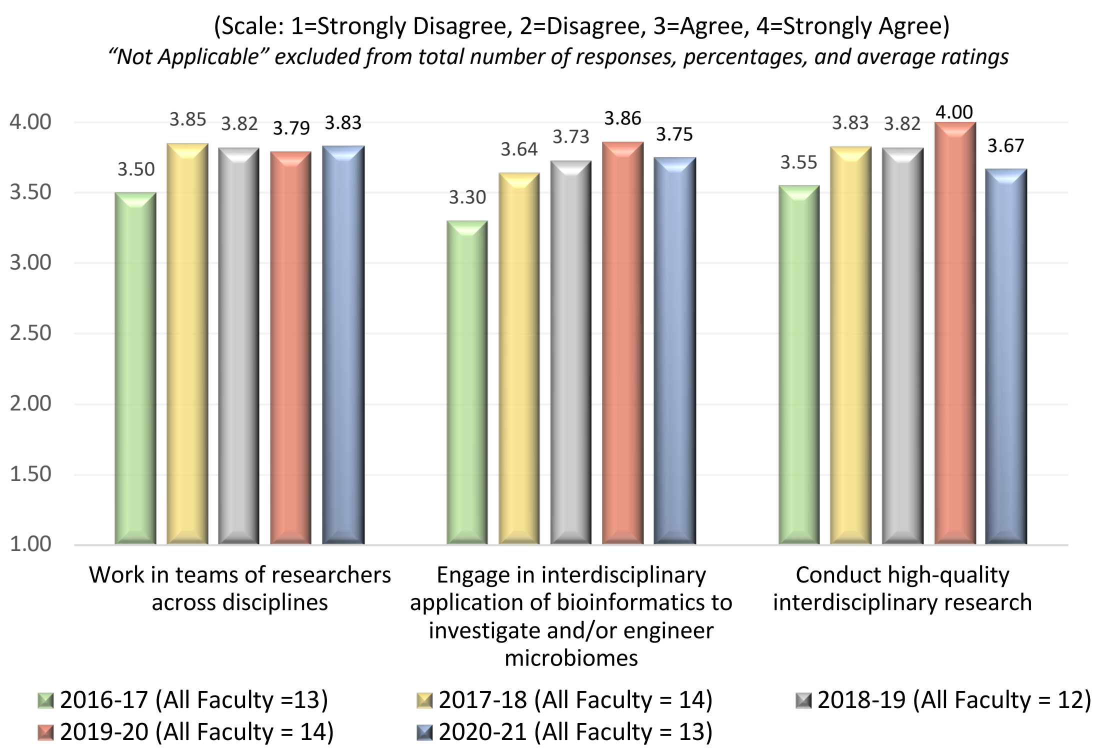
Figure 3. Faculty Ratings of Trainees' Conceptual Thinking and Methodological Skills- 2016-21
(Scale: 1=Strongly Disagree, 2=Disagree, 3=Agree, 4=Strongly Agree)
"Not Applicable" excluded from total number of responses, percentages, and average ratings.



Advances in trainees' abilities to work in interdisciplinary teams with application of bioinformatics to research or engineering in microbiome areas.

IBIEM Faculty gave highest ratings (Figure 4) for the last four years of the program for trainees' abilities to work in teams of researchers across disciplines, to conduct high quality interdisciplinary research, and apply bioinformatics to microbiome engineering or research. Faculty's open-ended feedback indicated that as a result of increased competence by trainees in these methodological and bioinformatics skill areas, their *labs leveraged trainees' expertise* to *initiate new research and training directions in these microbiome areas* and with trainees now serving as TAs in new courses beyond this NRT. In addition, faculty reported *adopting tools used in IBIEM courses, such as computational notebooks, for their own teaching.*

Figure 4. Comparison of 2016-21 Faculty Ratings of Trainees' Preparation to Engage in Interdisciplinary Microbiome Research



Summary and Discussion.

This paper reports on the five-year evaluation of a novel, interdisciplinary National Science Foundation Research Traineeship (NRT) funded graduate training program in microbiome research and engineering, which focused on challenges that were identified in training graduate students with differing entry level skills and across disciplines. Strategies and training elements implemented to successfully address these challenges were made possible through close collaboration between the evaluation team and project leadership who were highly responsive to evaluation feedback.

This IBIEM graduate training program was successful in creating an integrated interdisciplinary program that equipped trainees with skills to work across disciplines on complex challenges in areas of microbiome research and engineering. Through partnership between Duke University and North Carolina Agricultural and Technical State University, a diverse group of 85 scientists and engineers spanning 23 disciplines were trained, with 27.1% underrepresented minority and 52.9% female students.

A learner-centered pedagogy was woven through training and found to promote ownership in the learning process as well as provision of multiple opportunities to apply skills. This model was particularly suited to training in informatics tools and analyses which require iterative application and revision for mastery.

Core training elements were refined annually based on evaluation feedback and included:
1) Boot Camp orientation; 2) Collaborative Science Practica course series; 3) Speaker Seminar Series to introduce students to an array of professionals and careers across sectors; 4) expert mentoring on interdisciplinary team research projects; and 5) financial and mentored support for trainees to present their research at conferences. A key component of research training was provision of opportunities for students to apply their new bioinformatics and statistics research and collaborative skills to real-world research projects working in teams across disciplines and campuses. These team research experiences were among the most highly rated training components and linked to gains across technical, communication, and collaboration skills. Example team research projects are listed below (related publications listed in reference section): *Exploring the relationship between host phylogeny and gut microbiome structure across lemur species* [14]*; Impact of Copper Nanopesticides on Microbial Communities in Wetland and Terrestrial Ecosystems* [15]*; Effects of starvation and re-feeding on the microbiomes of the zebrafish intestine and environment* [16]; and *Microbiomes of oysters in oyster reefs and oysters in floating aquaculture* [17].

A comprehensive program evaluation identified training challenges that resulted in innovative revisions in training elements and modalities that improved strategies over four years and were linked to advances in trainee outcomes. The highest ratings were found in year four which benefited from cumulative refinements over the previous years and preceded the shift to all-virtual training in year five due to COVID-pandemic safety policies. Advances in trainee impacts were linked to revisions in delivery modalities and addition of strategies designed to address trainees' different entry levels skillsets and disciplinary backgrounds.

Evaluation findings indicated that IBIEM was effective in meeting program objectives of:
1) equipping trainees with skills to engage in interdisciplinary microbiome research across biological sciences, engineering, and biostatistics;
2) developing trainees' confidence in their abilities to work in collaborative teams and communicate their research to diverse audiences;
3) increasing trainees' awareness of non-academic career pathways; and
4) increasing participation of underrepresented minority groups in microbiome research areas while providing an environment in which students felt supported.

Training outcomes were highest for years two and four with the majority of trainees reporting: *increased confidence in methodological skills in microbiome research areas,* and *greater understanding of statistical application complexities and abilities to select and apply new statistical tools in microbiome research*. Faculty ratings of trainees' competencies in these areas were consistent with trainee self-ratings with highest ratings also found in years two to five for *improved interdisciplinary thinking*, increased abilities *to apply bioinformatics to microbiome research and engineering* and *work in teams across disciplines conducting high-quality research.*

Timing was rated as optimal by the majority of trainees who reported that entry into this program early in their graduate training allowed application of skills to their own research and to independently apply bioinformatic and biostatistic skills to their dissertations and to develop deeper understanding over time. Faculty mentors reported that trainees had changed the research direction of their labs by providing leadership in bioinformatics expertise. Gains in trainees' research and communication skills were further evidenced by the high numbers of interdisciplinary microbiome research products generated by trainees which included 57 peer reviewed publications, 13 published conference proceedings, and 56 conference and/or poster presentations. In addition, trainees and graduates gave 26 invited talks and participated in 46 outreach events contributing to broad public dissemination.

Faculty reported using more interdisciplinary approaches to teaching as a result of participation and incorporating IBIEM training approaches and online materials into courses that included consideration of students' backgrounds for selection of course materials and strategies to optimize student engagement and build a stronger foundation for success.

Institutional impacts include integration of IBIEM's training activities into courses across both campuses. IBIEM's core informatics courses in microbiome areas provided a basic entry to advanced level training across disciplines not previously available at either Duke or N.C. A&T universities. Faculty reported integrating the online bioinformatics tools and materials into a course that will continue post NRT funding at N.C. A&T.

Conclusion

The IBIEM training program made significant advances achieving its expected outcomes for trainees, faculty, and the participating universities over the five years. Evaluation findings were used by leadership to create innovative training strategies and new components that resulted in increased trainee outcomes over the last four years of the program. Faculty reported integration of interdisciplinary training approaches into their courses and development of new research directions in microbiome areas as a result of participation. Furthermore, this training partnership provided the foundation and training model for a new NSF Engineering Research Center (ERC) for Precision Microbiome Engineering (PreMiEr) https://premier-microbiome.org/ for work in the critical area of engineering the microbiome in built environments. We plan to incorporate effective elements from this model into training in this new ERC and will be working on adapting this model to the challenge of effective online training across five PreMiEr university partners sites.

Acknowledgement

This material is based on work supported by the National Science Foundation under Grant Number DGE 1545220. This work was also supported in part by the Engineering Research Centers Program of the National Science Foundation under NSF Cooperative Agreement No. EEC-2133504. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF.

References

[1] W. Bialek, "Introductory Science and Mathematics Education for 21st-Century Biologists," *Science*, vol. 303, no. 5659, pp. 788–790, Feb. 2004, https://doi.org/10.1126/science.1095480.

[2] P. Pevzner and R. Shamir, "Computing Has Changed Biology--Biology Education Must Catch Up," *Science,* vol. 325, no. 5940, pp. 541–542, Jul. 2009, https://doi.org/10.1126/science.1173876.

[3] P. Pevzner and R. Shamir, Eds., *Bioinformatics for Biologists*. Cambridge: Cambridge University Press, 2011. Accessed: Feb. 28, 2023. [Online]. Available: https://www.cambridge.org/core/books/bioinformatics-for-biologists/9DE144FFAC6A4D826977E32FCDD21E53

[4] J. C. Ford, "'Jack of All Trades and Master of One: Training for Interdisciplinary Research," *Journal of Analytical & Bioanalytical Techniques*, vol. 03, no. 07, 2012, https://doi.org/10.4172/2155-9872.1000e110.

[5] N. R. Council, *Convergence: Facilitating Transdisciplinary Integration of Life Sciences, Physical Sciences, Engineering, and Beyond*. 2014. Accessed: Feb. 28, 2023. [Online]. Available: https://nap.nationalacademies.org/catalog/18722/convergence-facilitating-transdisciplinary-integration-of-life-sciences-physical-sciences-engineering

[6] E. Pecukonis, O. Doyle, and D. L. Bliss, "Reducing barriers to interprofessional training: Promoting interprofessional cultural competence," *Journal of Interprofessional Care*, vol. 22, no. 4, pp. 417–428, Jan. 2008, https://doi.org/10.1080/13561820802190442.

[7] M. E. Huba and J. E. *Freed, Learner-Centered Assessment on College Campuses: Shifting the Focus from Teaching to Learning*. Allyn & Bacon, 160 Gould St, 2000. Available: https://eric.ed.gov/?id=ED438867

[8] L. Cassuto, "Student-Centered Graduate Teaching," *The Chronicle of Higher Education*, Nov. 04, 2013. https://www.chronicle.com/article/student-centered-graduate-teaching/ (accessed Feb. 28, 2023).

[9] M. Weimer, "Five Characteristics of Learner Centered Teaching," in *Faculty Focus : Focused on Today's Higher Education Professional, ed.* Miracosta College, 2012. https://home.miracosta.edu/rfrench/documents/effective-teaching-strateg.pdf :

[10] P. Blumberg, "Beginning journey toward a culture of learning centered teaching," *Journal of Student Centered Learning,* vol. 2, no. 1, pp. 68-80, 2004.

[11] M. Weimer, "Learner-Centered Teaching and Transformative Learning," in *The Handbook of Transformative Learning: Theory, Research, and Practice,* E. W. Taylor and P. Cranton, Eds.: John Wiley & Sons, 2012, pp. 439-454.

[12] T. Doyle, *Learner-centered teaching : putting the research on learning into practice*, 1st ed. ed. Sterling, Va: Stylus Pub., 2011.

[13] S. G. Henry, E. S. Holmboe, and R. M. Frankel, "Evidence-based competencies for improving communication skills in graduate medical education: A review with suggestions for implementation," *Medical Teacher*, vol. 35, no. 5, pp. 395–403, Feb. 2013, https://doi.org/10.3109/0142159x.2013.769677.

[14] S. L. Bornbusch *et al.*, "A comparative study of gut microbiomes in captive nocturnal strepsirrhines," *American Journal of Primatology,* vol. 81, no. 10-11, p. e22986, 2019, https://doi.org/10.1002/ajp.22986.

[15] L. N. Carley *et al.*, "Long-Term Effects of Copper Nanopesticides on Soil and Sediment Community Diversity in Two Outdoor Mesocosm Experiments," *Environmental Science & Technology,* vol. 54, no. 14, pp. 8878-8889, 2020/07/21 2020, doi: 10.1021/acs.est.0c00510

[16]M. Fuirst, C. S. Ward, C. Schwaner, Z. Diana, T. F. Schultz, and D. Rittschof, "Compositional and Functional Microbiome Variation Between Tubes of an Intertidal Polychaete and Surrounding Marine Sediment," (in English), *Frontiers in Marine Science,* Original Research vol. 8, 2021-May-31 2021, https://doi.org/10.3389/fmars.2021.656506.

[17] J. Jawahar *et al.*, "Starvation causes changes in the intestinal transcriptome and microbiome that are reversed upon refeeding," *BMC Genomics,* vol. 23, no. 1, p. 225, 2022/03/22 2022, https://doi.org/10.1186/s12864-022-08447-2

.