Hackathons as essential tools in an interdisciplinary biological training - report from trainings for Sub-saharan students

Marvin van Aalst[1], Tobias Pfennig[1,2], Faraimunashe Chirove[3], Marilyn Ronoh[4], Anna Matuszyńska[1,2*],

[1] Computational Life Science, Department of Biology, RWTH Aachen University, Aachen, Germany
[2] Cluster of Excellence on Plant Sciences (CEPLAS), Heinrich-Heine University Düsseldorf, Germany
[3] Department of Mathematics and Applied Mathematics, University of Johannesburg, South Africa
4 Department of Mathematics and Statistics, University of Embu, Kenya

* Correspondence: anna.matuszynska@cpbl.rwth-aachen.de

Abstract

Hackathons are collaborative, fast-paced events where participants from various fields work together to solve real-world problems. They are increasingly used in higher education to foster collaboration, problem-solving and applied computational skills, yet their role in interdisciplinary biological training remains under-documented. We report on the design and implementation of two computational biology summer schools in Kenya (2022, 2023), each culminating in a hackathon that integrated biological problems, quantitative methods, and coding. Both events targeted early-career researchers from multiple sub-Saharan countries and combined intensive teaching in programming and modelling with a time-bound group challenge using authentic marine conservation and synthetic epidemiological datasets. We describe the educational design, including its grounding in project-based learning, authentic learning, and Self-Determination Theory, and we document how performance-based assessment and structured participant feedback were used to evaluate learning outcomes. We present a critical reflective account of what worked, our teaching philosophy, and how hackathons can be embedded responsibly within biological curricula. We argue that, when embedded in sustained training and supported by appropriate mentoring, hackathons provide a practical and effective way to help biologists build computational skills, communicate across disciplines, and gain confidence in shaping their own research.

Keywords: authentic learning, project-based learning, datathon, modelling

1 Introduction

1.1 Hackathons as a response to biology's interdisciplinary data challenge

Modern biological research is constrained less by data availability than by the capacity to analyse and interpret complex datasets. High-throughput sequencing, imaging and environmental monitoring generate volumes of data [1]. To visualize the scale of this challenge, consider that 402.74 million terabytes of data are generated globally every

day [2]; equivalent to handing out a smartphone, completely full of apps, pictures, messages, files, and video, to every third person on the planet, every day[1]. In biology specifically, technological advances in omics, imaging, and high-throughput screening have created an unprecedented data deluge. The Eli and Edythe L. Broad Institute of MIT and Harvard alone produces nearly 20 terabytes of genomic data daily [3].

However, making sense out of these data and answering relevant, disruptive questions that push the boundaries of biological knowledge requires more than computational power: it requires collaboration between experts in life sciences, mathematics, statistics and data science [4, 5]. Yet bringing these experts together is not simply about putting people in the same room and hoping they will identify both the problem and the solution. To effectively work together, we need to teach and cultivate a shared understanding and a common language, fostering smooth communication and collaboration at the intersection of diverse disciplines—an essential 21st-century skill emphasized by educational frameworks and workforce reports [6, 7]. This creates a pedagogical challenge for biology education: how do we train the next generation of biologists to work effectively across disciplinary boundaries? Traditional biology curricula, often structured around domain-specific knowledge delivered through lectures and laboratory practicals, may not adequately prepare students for collaborative, data-intensive problem-solving in authentic research contexts [8, 9].

1.2 Why hackathons for biological education?

Hackathons, intensive, collaborative events where participants work in teams to solve real-world problems within a constrained timeframe, offer a promising pedagogical approach for addressing this gap. The term combines "hack" (referring to creative, exploratory problem-solving, originating from the MIT Tech Model Railroad Club's culture of clever technical solutions) and "marathon" (symbolizing the intensive, time-bound nature of the activity). While hackathons originated in software development, with the first recorded event organized by OpenBSD in 1999 [10], they have since expanded across disciplines [11–15], demonstrating particular value in contexts requiring rapid innovation and cross-disciplinary collaboration.

In biological education, hackathons hold specific promise because they mirror the authentic work of modern interdisciplinary research teams: they require participants to integrate diverse forms of expertise, communicate across disciplinary boundaries, work with real or realistic datasets, make decisions under time constraints, and produce tangible outputs [13]. Emerging evidence from engineering, software development, and health sciences demonstrates that hackathons can enhance problem-solving skills, teamwork, communication, and engagement [13, 16–18]. Recent systematic reviews confirm hackathons' educational value across STEM disciplines [19, 20], though rigorous controlled studies remain limited [21], with few reports detailing how to design, implement, and assess hackathons as integral components of biology training programs [22].

1.3 Hackathons as capstone experiences in computational biology training

This report extends the evidence base for hackathons in biological education by presenting detailed accounts of two computational biology summer schools in Kenya (2022 and 2023), where hackathons served as capstone experiences for over 50 early-career researchers from five sub-Saharan countries. Both programs addressed a critical need: training biologists with little to no prior computational experience to work

[1] assuming current population size to be 8.2 Billion and storage of a typical smartphone to be 128 GB

confidently with data and mathematical models, using authentic biological problems relevant to their regional contexts (marine conservation, HIV epidemiology, and harmful algal blooms).

Over the years, we, the authors, have first participated in, and later conducted, multiple hackathons in diverse settings. Our primary goal has been to enable students to use computational skills to address real-world biological problems. By fostering an environment that encourages open exchange of ideas and providing practical examples, we consistently observed a strong sense of achievement as students recognised their ability to tackle authentic challenges. While hackathons offer substantial opportunities for innovation and learning [23], their rapid pace and participant diversity can produce communication barriers, with differences in terminology, methods and disciplinary perspectives hindering collaboration. To mitigate these issues, we integrated hackathons not as isolated showcase events, but as the culminating components of structured training programs (summer school, workshop). Hackathons and workshops share a hands-on, active-learning foundation [24, 25], where hackathons emphasise rapid problem solving and knowledge integration [23], workshops offer structured, guided skill building. In combination, depth from workshops and the creative, solution-focused nature of hackathons reinforce each other, helping participants consolidate expertise while navigating interdisciplinary challenges in today's data-driven biology.

This design allowed participants to consolidate newly acquired skills through authentic collaborative workflows while demonstrating competence in addressing real biological challenges. Hackathons are particularly well-suited to this capstone role because modern biology increasingly requires: (i) data literacy (importing, cleaning, visualizing, and interpreting complex datasets); (ii) computational thinking (translating biological questions into algorithmic or mathematical frameworks); (iii) interdisciplinary communication (bridging conceptual gaps between biology, mathematics, and computer science); (iv) applied problem-solving (working with incomplete information and real-world constraints); and (v) collaborative skills (integrating diverse expertise within teams). We present this work as a descriptive, analytically grounded experience report examining how pedagogical design choices shaped implementation and outcomes. Our goal is to contribute practical, theoretically informed guidance for biology educators seeking to incorporate authentic, collaborative problem-solving experiences into their programs, particularly in contexts where computational skills are underdeveloped and where bridging the gap between biological domain knowledge and data science capabilities is urgent.

2 Materials and Methods

2.1 Pedagogical Framework and Design Rationale

Our hackathon design was grounded in three complementary educational frameworks that shaped our implementation choices:

- **Project-Based Learning (PBL)**. We adopted PBL as our core instructional approach [26], where participants engage in hands-on discovery and problem-solving to create tangible outputs rather than passively receiving information [17]. Specifically, our hackathons required teams to: (i) formulate biologically meaningful questions from authentic and synthetic datasets, (ii) select and apply appropriate analytical methods independently, (iii) iterate their approaches based on feedback, and (iv) communicate findings to specific audiences. This structure positions learners as active knowledge constructors rather than consumers, which research shows enhances retention and transfer of skills [24, 27].

- **Authentic Learning.** We deliberately centered both training programs on socially relevant, real-world problems using genuine datasets from local contexts—sea turtle bycatch data collected by fishermen (Case Study 1) and epidemiological challenges affecting East African communities (Case Study 2). Authentic learning theory posits that learning is most effective when situated in contexts that mirror the complexity and social embeddedness of professional practice [28]. In our case, this meant participants worked with actual conservation data that would inform local decision-making, rather than sanitized textbook examples. The requirement to present findings in Kiswahili to fishermen (Case Study 1) further emphasized the authentic nature of the communication challenge facing interdisciplinary researchers.

- **Self-Determination Theory (SDT)**. Our design choices directly addressed the three psychological needs identified by SDT as essential for intrinsic motivation: **competence**, **autonomy**, and **relatedness** [29]. We fostered competence through progressive skill-building during the training weeks, followed by immediate application during the hackathon where participants could demonstrate mastery. The intensive mentoring support (approximately 1:10 trainer-to-participant ratio, see below) provided structure that allowed participants to succeed at challenging tasks without becoming overwhelmed. We supported autonomy by allowing teams to choose their own analytical approaches, decide how to divide responsibilities, and select visualization methods without prescriptive instructions. Relatedness was cultivated through deliberate team formation between disciplines and institutions, collaborative problem-solving structures, and social activities designed to build trust and community (poster sessions, field trips, informal discussions). This combination has been shown to enhance both learning outcomes and persistence in STEM fields [29–31].

These frameworks are interconnected: PBL provides the instructional structure, authentic learning ensures biological and social relevance, and SDT explains why this combination enhances motivation and engagement. Our "No One Left Behind" teaching philosophy (detailed below) operationalized these principles through specific design features.

2.2 Context, Participants, and Selection

Case Study 1: Summer School 2022, Watamu, Kenya

The first event was a two-week intensive program (September 5-16, 2022) titled "Computational Biology for Sustainable Environmental Research," co-organized by RWTH Aachen University (Germany) and Mawazo Institute (Kenya). The program was held in Watamu, Kenya, a coastal town adjacent to one of the world's oldest marine protected areas, providing immediate access to conservation sites and local environmental organizations.

Environmental Hackathon with real data

Recruitment and selection. We received 98 applications from 12 Sub-Saharan countries (Benin, Cameroon, Congo DRC, Kenya, Eswatini, Ethiopia, Ghana, Senegal, Somalia, South Africa, Tanzania, and Zimbabwe). The application required: (i) a 90-second video explaining motivation and research interests, (ii) a résumé, and (iii) written responses to three questions addressing current research, expected benefits from the training, and views on sustainable research. Five trainers independently evaluated applications using a rubric assessing research alignment, motivation, potential to benefit from and contribute to the program, and commitment to environmental sustainability, resulting in a selection of 25 participants.

Participant characteristics. The final cohort included 25 participants (72% women, 28% men; Fig. 1B) from six countries, with the majority (56%) residing in Kenya and significant representation from Ghana (32%), Uganda (8%), and Nigeria (4%). Educational levels ranged from undergraduate (8%) to post-doctoral researchers (4%), with the majority being Master's students (64%) or doctoral candidates (20%) (Fig. 1C). Disciplines represented included marine biology, food security, environmental science, and policy. Critically, most participants reported little to no prior programming experience, though all possessed foundational knowledge in their biological domains. English proficiency was confirmed during video review; all instruction and materials were delivered in English, though participants presented hackathon findings in Kiswahili to local stakeholders.

Funding and accessibility. Participation was fully funded, covering accommodation, meals, all training materials, and travel. We provided travel grants to all participants who requested support, ensuring financial barriers did not limit access. Recognizing that 12% of applicants indicated childcare needs, we allocated budget for on-site childcare support; two participants used this service, bringing young children to the program.

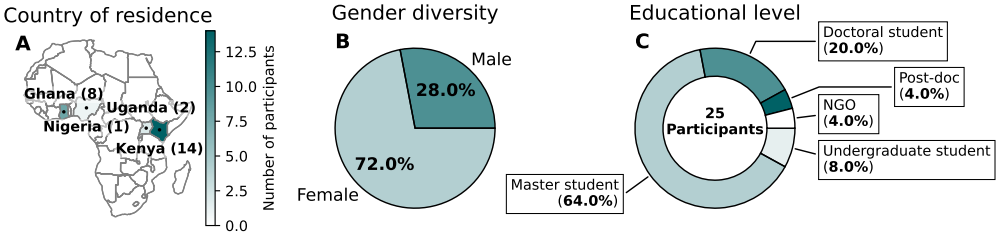

Fig 1. Participant distribution by country, gender, and education level in Case Study 1. A) Countries of residency of the 25 Sub-Saharan participants are marked on the map. The countries' colors reflect the number of participants. B) The participants' gender distribution. C) The educational level of the participants at the time of the hackathon. Participants stemmed from African universities or Non-Governmental organizations (NGOs).

Case Study 2: Workshop 2023, Embu, Kenya

The second event was a condensed five-day workshop (dates in September 2023) at the University of Embu, Kenya, focusing on computational modeling for epidemiology and environmental science.

Epidemiological Hackathon with generic data

Recruitment and selection. We distributed an open call through African university networks and social media, receiving applications from students across the continent. The selection process followed a similar rubric to Case Study 1, prioritizing applicants who would directly benefit from computational training for their ongoing research, resulting in a selection of 22 participants.

Participant characteristics. The cohort included 22 participants (63.6% women, 36.4% men; Fig. 2B) predominantly from Kenyan universities (91%), with one participant from Zimbabwe (4.5%) and Uganda (4.5%) (Fig. 2A). Educational levels were concentrated among Master's students (72.7%) and doctoral candidates (22.7%), and undergraduate (4.5%) (Fig. 2C). Disciplines included mathematics, statistics, public health, and ecology. As in Case Study 1, most participants had limited programming experience but strong quantitative and biological foundations.

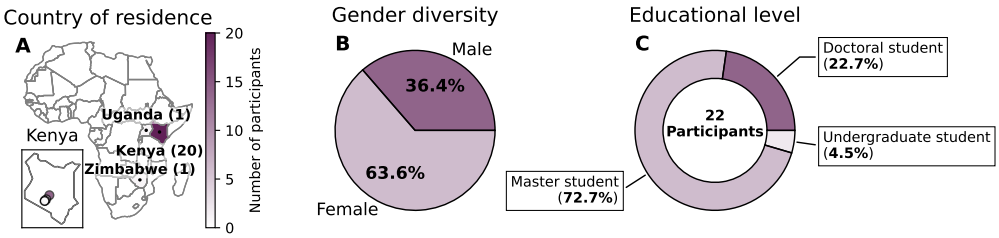

Fig 2. Participant distribution by country, gender, and education level in Case Study 2. A) Countries of residency of the 23 participants are marked on the map. The countries' colors reflect the number of participants. The inset plot shows the location of the participants' Kenyian universities. B) The participants' gender distribution. C) The educational level of the participants at the time of the hackathon.

Limitations of sampling approach. We acknowledge several important limitations. First, both cohorts were self-selected from applicants motivated enough to complete a competitive application, potentially inflating apparent engagement and satisfaction relative to compulsory course settings. Second, the competitive selection process identified participants already primed for success—motivated, academically strong, and institutionally supported—limiting generalizability to broader student populations. Third, prior skill heterogeneity was substantial: while most had minimal coding experience, some had completed introductory programming courses or used computational tools in their research, creating variation in starting points that we accommodated through intensive mentoring but could not fully control. Fourth, language may have presented barriers for some participants, as technical instruction in English required fluency that, while verified, may have disadvantaged non-native speakers. Finally, our intensive support model (described below) is resource-intensive and may not be replicable in standard institutional settings. These factors mean our results reflect "best-case" conditions for hackathon implementation rather than typical classroom constraints.

2.3 Hackathon Structure and Implementation

Both hackathons followed a five-component structure (Fig. 3), though with different durations and data sources:

1. **Introduction and Context** (1 hour). Trainers introduced the dataset structure, biological/social context, and challenge objectives. In Case Study 1, Local Ocean

Conservation (LOC) staff presented their turtle bycatch monitoring program and conservation goals. In Case Study 2, we framed the HIV and algal bloom challenges using recent literature and regional health/environmental statistics.

2. **Individual Group Work** (Case Study 1: 16 hours over 2 days; Case Study 2: 11 hours over 2 days). Participants worked in assigned teams (4 groups of 5-6 members each) on data exploration, analysis, and model development. Teams were deliberately heterogeneous, balancing: (i) disciplinary background (e.g., pairing marine biologists with statisticians), (ii) career stage (mixing undergraduates, Master's, and PhD students), (iii) institutional affiliation, and (iv) geographic origin to maximize diverse perspectives and minimize pre-existing collaboration patterns. Teams had autonomy to choose their analytical approaches, division of labor, and presentation strategies.

3. **Mentoring Check-ins** (every 60 minutes, 5-10 minutes each). Trainers circulated continuously and conducted brief check-ins with each team every hour to: (i) assess progress and identify bottlenecks, (ii) provide technical troubleshooting (e.g., debugging code, clarifying statistical methods), (iii) offer conceptual guidance (e.g., suggesting alternative visualizations, prompting biological interpretation), and (iv) ensure equitable participation within teams. Mentors explicitly avoided prescribing solutions, instead using Socratic questioning to guide teams toward their own insights.

4. **Plenary Peer Exchange Sessions** (Case Study 1: 3 sessions; Case Study 2: 1 session). Teams briefly presented interim findings to the full cohort, received peer feedback, and coordinated to avoid duplicated efforts. This structure encouraged comparison of approaches and integration of complementary insights.

5. **Final Presentation Preparation and Dissemination** (Case Study 1: 5 hours; Case Study 2: 3 hours). Teams synthesized analyses into coherent narratives, designed visualizations, and prepared presentations tailored to their audiences. In Case Study 1, presentations were delivered in Kiswahili to local fishermen at Mida Creek, without electronic aids, requiring teams to create flipchart summaries accessible to non-technical audiences. In Case Study 2, teams presented to university leadership and peers using slides and live code demonstrations.

Component	Description	Teaching objective	Case Study 1 25 hours	Case Study 2 14 horus
1. Introduction & Background	Introduction to the data structure, description of the content, setting the goal.	Participants will ... **identify** key data features and **explain** the context and objectives of the task.	60 minutes 8 - 9 Day 1	60 minutes 14 - 15 Day 1
2. Individual work	Brainstorming, analysis, and applying core skills to explore the dataset.	... **apply** analysis techniques to **explore** the dataset and document their findings for discussion.	**4 sessions** 9 - 12 Day 1 13 - 18 Day 1 19 - *open end* 10 - 13 Day 2	**2 sessions** 15 - 18 Day 1 8 - 14 Day 2
3. Check-up sessions	Mentoring and support by instructors.	... **analyze** their progress and **evaluate** their approach with instructors' guidance to refine their methods.	Every hour 5 - 10 minutes	Every hour 5 - 10 minutes
4. Plenary sessions	Exchange of information between groups using peer review method.	... **evaluate** peer contributions, **compare** approaches, and **integrate** feedback into their projects.	**three sessions** 12 - 13 Day 1 18 - 19 Day 1 9 - 10 Day 2	**one session** 18- 19 Day 1
5. Presentation	Creation and design of the presentation catered to the target audience.	... **synthesize** their analysis into a coherent narrative, **design** visuals, and **deliver** an engaging presentation.	**5 hours** 2 h preparation + 3 h dissemination	**3 hours** 2 h preparation + 1 h disseminatin

Fig 3. Structure of a Hackathon. Recommendation for organizing and holding a successful hackathon for education. Ensure that five critical components are present, regardless of the length of the planned event, starting from providing the context for the challenge and finishing with a proper dissemination of the final results to the most relevant stakeholders. The proposed teaching outcomes of each component are outlined according to Bloom's Taxonomy of Educational Objectives for Knowledge-Based Goals [32] and the exemplary allocation of time is provided based on the experience from the two presented here case studies.

"No One Left Behind" Teaching Philosophy

Throughout both programs, we implemented an intensive support model to ensure accessibility for participants with limited computational backgrounds.

- **Extensive hands-on practice:** 75% of instructional time allocated to applied problem-solving rather than passive listening.
- **High trainer-to-participant ratio:** approximately 1 trainer per 10 participants, with at least 2 trainers present at all times during teaching blocks.
- **Frequent checkpoints:** trainers confirmed that all participants could reproduce examples and complete exercises independently on their own computers before proceeding.
- **Peer mentoring:** we encouraged faster learners to assist peers, fostering collaborative rather than competitive dynamics. **Technical troubleshooting support:** junior trainers (van Aalst, Pfennig) provided continuous one-on-one debugging assistance.

This model required substantial staffing (6 trainers for 25 participants in Case Study 1; 4 trainers for 22 participants in Case Study 2) but was essential for ensuring participants with no prior coding experience could fully engage.

2.4 Evaluation and Data Sources

To evaluate learning and the educational value of the hackathons, we used three complementary data sources aligned with the project-based design: (i) performance-based assessment of team outputs, (ii) structured participant feedback, and (iii) instructor observations. We did not use formal pre/post knowledge tests; instead, we treated the hackathon projects themselves as the primary demonstration of achieved

outcomes.

First, each team's project was assessed using a concise, pre-defined rubric applied to their learning artifacts/materials (Git repositories, notebooks, slides, outreach materials):

1. **Completion and coherence** – delivery of a functioning, self-contained project, including executable code or notebooks and a final presentation or summary.

2. **Use of taught methods** – appropriate application of core skills covered in the training (e.g. data import, cleaning, visualisation, basic modelling or statistical exploration) without major methodological errors.

3. **Biological relevance and interpretation** – formulation of a biologically meaningful question (e.g. turtle bycatch patterns, infection or bloom dynamics) and interpretation of outputs in a way that is consistent with the underlying biology and dataset limitations.

4. **Communication and audience alignment** – clarity and structure of the narrative for the intended audience (peers, local stakeholders, or university leadership), including where applicable the ability to translate technical results into accessible language.

We mapped these criteria to Bloom's Taxonomy's cognitive levels [32]:

- Criterion 1 (Completion) reflects *Application* (applying learned procedures)

- Criterion 2 (Methods use) spans *Application* and *Analysis* (selecting and implementing appropriate techniques)

- Criterion 3 (Biological interpretation) requires *Analysis* and *Evaluation* (interpreting data within biological context, judging validity)

- Criterion 4 (Communication) demands *Synthesis/Create* (reorganizing findings into coherent narratives for specific audiences).

The teaching team jointly reviewed all group outputs against these criteria to determine whether each project met the expected standard for successful completion. This performance-based evaluation served as our primary indicator of whether participants were able to integrate the computational tools and biological concepts addressed in the schools.

Second, immediately after each event, participants completed an anonymous online survey including Likert-type items on organisation, support, perceived learning gains, confidence in applying computational methods to biological problems, and open-ended questions on strengths and weaknesses of the format (see Fig. 4). The instrument was developed specifically for these events and is not a validated scale; we use these self-reports only as supportive, descriptive evidence and recognise their susceptibility to social desirability and novelty effects.

Third, instructors maintained brief field notes documenting observable patterns during the hackathons (engagement, collaboration, difficulties, reliance on mentoring, inclusion of less experienced participants). These notes were used to contextualise the performance and survey data and to identify recurring challenges and design implications. All analyses are descriptive; no inferential statistics are reported, and we do not claim causal proof of effectiveness.

Data and materials availability.

All teaching materials, participant projects, and documentation are publicly archived to support reproducibility and adoption by other educators:

Case Study 1 Summer School 2022: Teaching materials, schedules, and anonymized outputs are available at:
`https://git.rwth-aachen.de/computational-life-science/summer-school`

Case Study 2 Workshop 2023: Complete teaching materials, Jupyter notebooks, participant project repositories, and workshop documentation are available at:
`https://github.com/Computational-Biology-Aachen/HackathonEmbu2023`

These repositories include setup instructions (environment configuration, required software, installation guides), teaching notebooks with worked examples, and datasets used during training. We provide these resources to facilitate replication and adaptation by other institutions.

3 Results

3.1 Performance-based outcomes (primary evidence)

Using the pre-specified rubric (Methods), all teams in both case studies delivered a coherent, runnable project that:

1. **Completed a self-contained workflow** (executable notebooks/code and a final presentation).
2. **Applied taught methods** (data import/cleaning, visualisation, and, where relevant, basic modelling) without major methodological errors.
3. **Addressed a biologically meaningful question** and provided interpretations consistent with data limitations.
4. **Communicated findings** to the intended audience (stakeholders in Watamu; university peers/leadership in Embu).

In **Case Study 1**, teams identified spatial and gear-specific patterns in turtle bycatch and produced low-tech outreach artifacts for local fishers. In **Case Study 2**, teams formulated simple dynamical models (e.g., compartmental or growth models) for HIV-infected T cells and phytoplankton, related parameters to mechanisms, and contrasted dynamical behaviours across the two contexts. Taken together, artifacts indicate progression beyond *apply* towards *analyse/evaluate* and, in several cases, *create* levels on the revised Bloom taxonomy.

3.2 Participant perceptions (supporting evidence)

Post-event surveys (anonymous; non-validated, descriptive) indicated strong perceived value of the hackathon format. In **Case Study 2** (Fig. 4):

- Overall ratings for content relevance, concept clarity, and example usefulness were predominantly at the top of the scale.
- Binary items (engagement, recommendation, intention to apply/share skills) were uniformly positive.

- Free-text responses most frequently highlighted practical application, data analysis, programming, and teamwork as the most valuable topics; most comments supported including hackathons in the regular curriculum (Table S2).

We interpret these results cautiously (potential social desirability, novelty, and response biases) and use them to contextualise, not replace, the performance-based outcomes.

3.3 Process observations and inclusion

Instructor field notes documented high engagement across teams, frequent peer mentoring, and effective use of scheduled checkpoints to surface bottlenecks (e.g., debugging, choice of visualisations, framing biological questions). Heterogeneity in prior coding experience required sustained mentor presence (trainer:participant ∼1:4 to 1:10) to keep all participants progressing ("no one left behind"). This support intensity was integral to feasibility in both settings but is resource demanding.
:contentReference[oaicite:3]index=3

3.4 Limitations (result-level)

Findings reflect best-case conditions: motivated, self-selected participants; short, intensive formats; and high mentor availability. There were no standardised pre/post tests or longitudinal follow-up; thus, we do not claim causal learning gains or skill retention. The evidence base for effectiveness in typical compulsory courses remains limited and warrants targeted evaluation in future iterations.

3.5 Summary

Across two implementations, all teams produced reproducible analyses addressing authentic biological questions and communicated results to appropriate audiences, meeting the primary performance criteria. Perception data and observations converged with artifact quality, suggesting that embedded, mentored hackathons function as feasible capstones for interdisciplinary biological training, while highlighting the need for lightweight, prospective assessment additions in future runs.

4 Discussion

The outcomes of both hackathons suggest that, when embedded within structured training programs and supported by close mentoring, hackathons can function as effective project-based capstones for interdisciplinary biological education. All teams delivered coherent, runnable projects that applied core methods to authentic biological questions, reaching at least the "analyse–create" levels of Bloom's taxonomy. However, our evaluation approach (performance-based judgement of artifacts supplemented by self-reports and instructor observations) remains methodologically modest and is constrained by strong self-selection, heterogeneous prior skills, and intensive staff support. We frame our contribution as a descriptive, analytically grounded experience report that documents implementation under specific conditions (intensive, residential, internationally funded programs for highly motivated Sub-Saharan African early-career researchers). This positions our work transparently within the emerging evidence base while acknowledging what we cannot claim about generalisability or causal effectiveness.

We therefore recommend that future implementations retain the hackathon as an authentic, product-focused assessment, but strengthen their evidential basis through minimal, targeted additions: (i) publishing a short, transparent rubric for project evaluation (covering correctness, biological relevance, reproducibility and

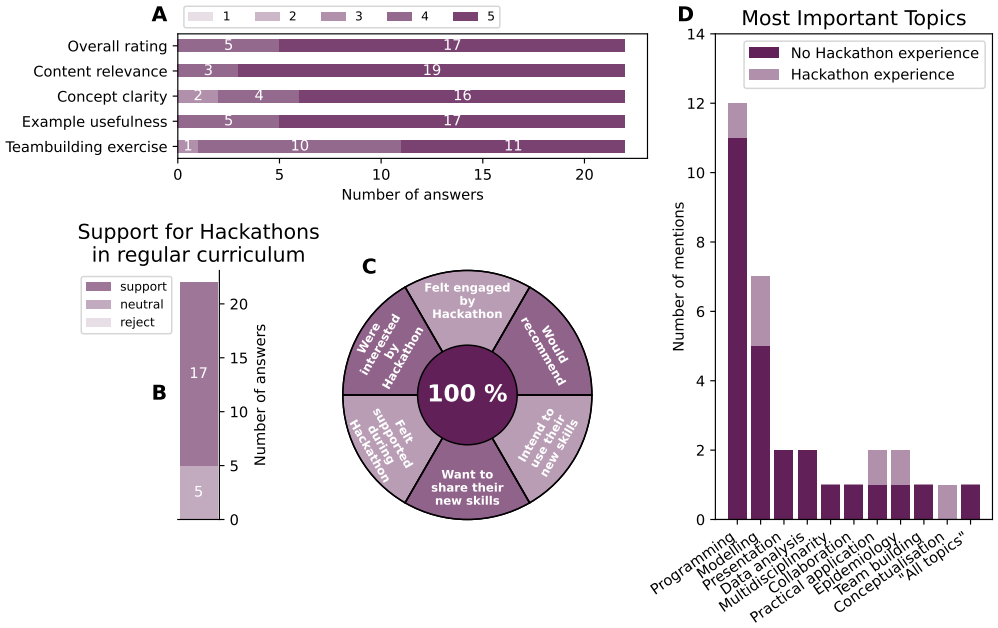

Fig 4. Participants' evaluation results of Case Study 2. A) Participants' rating of the hackathon-style workshop on a scale of 1 (worst) to 5 (best). B) Participants' response to having Hackathons included in their regular curriculum. Answers with no clear position were counted as "neutral". C) Summary of binary questions to the participants that were uniformly answered with "Yes." D) The most participants' important topics in the workshop, separated by whether they had previously participated in a hackathon. The participants' free text answers were assigned all common topics explicitly mentioned. Figure D shows the number of mentions of each topic across participants. Results in B and D stem from free text (see Tables S2.

communication) and applying it systematically across teams; (ii) integrating a brief pre/post micro-task aligned with key learning goals (e.g. selecting appropriate visualisations, identifying data quality issues, formulating biologically meaningful questions) to capture shifts in applied reasoning without overburdening participants; and (iii) optionally collecting a short follow-up report on whether and how participants used the acquired skills in subsequent coursework or research. Such measures are feasible within the constraints of intensive schools, respect the principles of project-based and authentic learning, and would move hackathon reports beyond enthusiasm and satisfaction towards a more robust, comparable evidence base for their role in modern biology education.

4.1 Positioning Within the Hackathon Education Evidence Base

Our descriptive approach reflects the field's current state. Systematic reviews show no standardized hackathon setup has been exhaustively investigated, with most studies relying on perceived rather than assessed learning [20]. Scoping reviews identify lack of long-term impact research as critical gaps [33, 34]. We share these constraints but advance the field through: (i) detailed implementation accounts reviews identify as lacking [21, 35], (ii) performance-based assessment addressing over-reliance on satisfaction surveys [36], and (iii) extension to biology education where evidence lags engineering contexts [19].

4.2 Recommendations for Educators

Based on our experience implementing hackathons in two computational biology training programs, we offer the following evidence-based guidance for educators considering similar approaches. These recommendations aim to maximize learning while acknowledging resource constraints that many institutions face.

Design and Structure. (1) Embed hackathons as capstone experiences following structured skill-building (1-2 weeks or semester-long courses), not as standalone events [17]. (2) Prioritize authentic, locally relevant biological problems through partnerships with conservation organizations, health agencies, or research institutes—authentic contexts enhance motivation and transfer [28]. (3) Design for heterogeneous skills using tiered challenges, frequent checkpoints, and peer mentoring to accommodate varied prior experience [20].

Support Infrastructure. (4) Maintain adequate mentoring capacity (approximately 1:10 trainer-to-participant ratio) through teaching assistants, peer facilitators, or reduced cohort sizes—immediate technical support prevents compounding frustration. (5) Balance intensity with well-being by scheduling breaks, normalizing error, and limiting duration (14-25 hours over 2 days worked well) to prevent burnout while maintaining productive challenge [23].

Implementation Details. (6) Compose teams intentionally across disciplines, career stages, and institutions to maximize diverse perspectives and prevent dominance by subgroups [37]. (7) Prepare technical contingencies including offline documentation, pre-installed software, spare hardware, and platform alternatives for inevitable technical failures. (8) Establish clear data use agreements addressing ownership, anonymization, and ethical responsibilities when using partner data.

Assessment and Adaptation. (9) Integrate assessment from the start using transparent rubrics, brief pre/post tasks, and structured reflection prompts rather than retroactive evaluation [38]. (10) Adapt formats to institutional contexts by considering mini-hackathons in existing courses, extendeding timelines with milestones, or switching to virtual platforms when residential intensive programs are infeasible [39].

4.3 Broader Implications for Biology Education

Biology's computational transformation demands graduates who can translate between biological questions and quantitative methods, collaborate across disciplines, communicate to diverse stakeholders, and navigate real-world data complexity [9]. Traditional lecture-dominated, discipline-siloed curricula inadequately prepare students for this reality [40]. Hackathons offer one pedagogical approach for developing these interdisciplinary competencies, particularly when embedded as capstone experiences consolidating prior learning through authentic application. They complement rather than replace foundational coursework, laboratories, and research apprenticeships—occupying a valuable niche for synthesis, collaboration, and communication skills that resist traditional assessment [7]. Our Kenya case studies demonstrate feasibility in resource-constrained Global South settings, where computational training can build local capacity to address regional biological challenges (infectious diseases, biodiversity loss, climate impacts). The competence demonstrated by participants (many programming for the first time) challenges deficit narratives, illustrating that structured opportunity and quality instruction enable success [29]. However, pedagogical innovation requires rigorous evaluation. The field needs transparent reporting, performance-based assessment, longitudinal follow-up, comparative studies, and validated instruments enabling cross-study

synthesis [20, 25, 41, 42]. As biology education evolves, we must couple experimentation with systematic evidence generation, building a robust knowledge base guiding future practice for data-intensive, interdisciplinary biological training.

5 Acknowledgments

This work was supported by the Volkswagen Foundation, UNESCO, the International Development Research Centre (Canada), and the Deutsche Forschungsgemeinschaft (DFG, EXC-2048/1 under Germany's Excellence Strategy). We would like to thank the Local Ocean Conservation (LOC) in Watamu, Kenya, in particular Nicky Parazzi, who trusted us with their twenty-year-old data collection. The authors thank Julia Lorke and Tim Nies for critically reading the first draft of this article. Finally, we extend our gratitude to Otho Mantegazza, with whom we have created the educational content, for his involvement, ideas, and heart he put into educational programs, and to the Mawazo Institute, in particular Fiona Moejes, who was an essential co-organizer of the discussed above events.

Disclaimer: In compliance with the Taylor & Francis AI Policy, we, the authors, disclose the use of ChatGPT (version 4) to enhance the clarity and coherence of our manuscript. This tool helped refine language and structure to effectively communicate complex ideas. We have thoroughly reviewed and take full responsibility for the content generated with ChatGPT's assistance, ensuring its accuracy and integrity.

References

1. Zheng S, Donahue TR. Systems Biology: Generating and Understanding Big Data. In: Hunter CJ, Dudeja V, editors. Success in Academic Surgery: Basic Science. Cham: Springer Nature Switzerland; 2025. p. 347–358. Available from: `https://link.springer.com/10.1007/978-3-031-83034-1_19`.

2. Data growth worldwide 2010-2028;. Available from: `https://www.statista.com/statistics/871513/worldwide-data-created/`.

3. Production Data Processing;. Available from: `https://www.broadinstitute.org/data-sciences-platform/production-data-processing`.

4. Park M, Leahey E, Funk RJ. Papers and patents are becoming less disruptive over time. Nature. 2023;613(7942):138–144. doi:10.1038/s41586-022-05543-x.

5. Cowen T, Southwood B. Is the Rate of Scientific Progress Slowing Down? SSRN Electronic Journal. 2019;doi:10.2139/ssrn.3822691.

6. Cropley A. Creativity-focused Technology Education in the Age of Industry 4.0. Creativity Research Journal. 2020;32(2):184–191. doi:10.1080/10400419.2020.1751546.

7. Binkley M, Erstad O, Herman J, Raizen S, Ripley M, Miller-Ricci M, et al. Defining Twenty-First Century Skills. In: Griffin P, McGaw B, Care E, editors. Assessment and Teaching of 21st Century Skills. Dordrecht: Springer Netherlands; 2012. p. 17–66. Available from: `https://doi.org/10.1007/978-94-007-2324-5_2`.

8. Singer SR, Nielsen NR, Schweingruber HA. Biology Education Research: Lessons and Future Directions. CBE—Life Sciences Education. 2013;12(2):129–132. doi:10.1187/cbe.13-03-0058.

9. Williams JJ, Drew JC, Galindo-Gonzalez S, Robic S, Dinsdale E, Morgan WR, et al. 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. 2019;14(11):e0224288. doi:10.1371/journal.pone.0224288.

10. Hackathons;. Available from: `https://www.openbsd.org/hackathons.html`.

11. Aboab J, Celi LA, Charlton P, Feng M, Ghassemi M, Marshall DC, et al. A "datathon" model to support cross-disciplinary collaboration. Science Translational Medicine. 2016;8(333):333ps8–333ps8. doi:10.1126/scitranslmed.aad9072.

12. Celi LA, Lokhandwala S, Montgomery R, Moses C, Naumann T, Pollard T, et al. Datathons and Software to Promote Reproducible Research. Journal of Medical Internet Research. 2016;18(8):e230. doi:10.2196/jmir.6365.

13. Zimolzak AJ, Davila JA, Punugoti V, Balasubramanyam A, Klotman PE, Petersen LA, et al. Lessons learned from an enterprise-wide clinical datathon. Journal of Clinical and Translational Science. 2022;6(1):e125. doi:10.1017/cts.2022.450.

14. Szymańska I, Sesti T, Motley H, Puia G. The Effects of Hackathons on the Entrepreneurial Skillset and Perceived Self-Efficacy as Factors Shaping Entrepreneurial Intentions. Administrative Sciences. 2020;10(3). doi:10.3390/admsci10030073.

15. Vilariño F, Serra-Ruiz J. Hackathons as a teaching methodology for challenge-based learning. In: EDULEARN21 Proceedings. 13th International Conference on Education and New Learning Technologies. IATED; 2021. p. 8123–8130.

16. Uffreduzi M. Hackathon as Emerging Innovation Practice. Politechnico Milano 1863. Milano; 2016. Available from: `https://www.politesi.polimi.it/bitstream/10589/137237/5/Hackathon%20as%20Emerging%20Innovation%20Practice.pdf`.

17. Almulla MA. The Effectiveness of the Project-Based Learning (PBL) Approach as a Way to Engage Students in Learning. Sage Open. 2020;10(3):2158244020938702. doi:10.1177/2158244020938702.

18. Tandon J, Pakpour N, Gumina M. Hack Days as a Method to Increase Interdisciplinary Knowledge and Collaboration Skills of College Students in STEM Fields. Journal of Interdisciplinary Studies in Education. 2021;10(2):81–92.

19. Chau CW, Gerber EM. On Hackathons: A Multidisciplinary Literature Review. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. Hamburg Germany: ACM; 2023. p. 1–21. Available from: `https://dl.acm.org/doi/10.1145/3544548.3581234`.

20. Schulten C, Chounta IA. How do we learn in and from Hackathons? A systematic literature review. Education and Information Technologies. 2024;29(15):20103–20134. doi:10.1007/s10639-024-12668-1.

21. Medina Angarita MA, Nolte A. What do we know about hackathon outcomes and how to support them?—A systematic literature review. In: Collaboration Technologies and Social Computing. Springer; 2020. p. 50–64.

22. Hollmann S, Regierer B, Attwood TK, Gisel A, Van Helden J, Rossier G, et al. Ten Simples Rules on How to Organise a Bioinformatics Hackathon. EMBnetjournal. 2021;26:e983. doi:10.14806/ej.26.0.983.

23. Garcia MB. Fostering an Innovation Culture in the Education Sector: A Scoping Review and Bibliometric Analysis of Hackathon Research. Innovative Higher Education. 2023;48(4):739–762. doi:10.1007/s10755-023-09651-y.

24. Prince M. Does Active Learning Work? A Review of the Research. Journal of Engineering Education. 2004;93(3):223–231. doi:10.1002/j.2168-9830.2004.tb00809.x.

25. Freeman S, Eddy SL, McDonough M, et al. Active learning increases student performance in science, engineering, and mathematics. Proceedings of the National Academy of Sciences. 2014;111(23):8410–8415. doi:10.1073/pnas.1319030111.

26. Krajcik JS, Blumenfeld PC. Project-based learning. In: Sawyer RK, editor. The Cambridge Handbook of the Learning Sciences. Cambridge: Cambridge University Press; 2006. p. 317–334.

27. Kokotsaki D, Menzies V, Wiggins A. Project-based learning: A review of the literature. Improving Schools. 2016;19(3):267–277. doi:10.1177/1365480216659733.

28. Yildirim R, Ortak c. The Effects of Authentic Learning Approach on Academic Achievement and Attitudes in Social Studies Course. International Journal of Psychology and Educational Studies. 2022;8(3):193–208. doi:10.52380/ijpes.2021.8.3.508.

29. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist. 2000;55(1):68–78. doi:10.1037/0003-066X.55.1.68.

30. Deci EL, Ryan RM. Self-determination theory: A macrotheory of human motivation, development, and health. Canadian Psychology/Psychologie Canadienne. 2008;49(3):182–185. doi:10.1037/a0012801.

31. Estrada M, Burnett M, Campbell AG, et al. Improving underrepresented minority student persistence in STEM. CBE—Life Sciences Education. 2016;15(3):es5. doi:10.1187/cbe.16-01-0038.

32. Bloom BS, Engelhart MD, Furst EJ, Hill WH, Krathwohl DR. Taxonomy of educational objectives. vol. 2. Longmans, Green New York; 1964.

33. Silver JK, Binder DS, Zubcevik N, Zafonte RD. Healthcare Hackathons Provide Educational and Innovation Opportunities: A Case Study and Best Practice Recommendations. Journal of Medical Systems. 2016;40(7):177. doi:10.1007/s10916-016-0532-3.

34. Rooholamini A, Salajegheh M. Health profession education hackathons: a scoping review of current trends and best practices. BMC Medical Education. 2024;24(1):554. doi:10.1186/s12909-024-05519-7.

35. Kienzler H, Fontanesi C. Learning through inquiry: A global health hackathon. Teaching in Higher Education. 2017;22(2):129–142. doi:10.1080/13562517.2016.1221805.

36. Oyetade K, Zuva T, Harmse A. Evaluation of the impact of hackathons in education. Cogent Education. 2024;11(1):2392420. doi:10.1080/2331186X.2024.2392420.

37. Wang G, Kasraian D, Valk C, Kersten A, Hurst W, Lu Y, et al. Interdisciplinary hackathons for community-based co-design with older adults: a case study. CoDesign. 2024; p. 1–16. doi:10.1080/15710882.2024.2378933.

38. Afshar Y, Bahrehvar M, Moshirpour M, Behjat L. Hackathon as an Effective Learning and Assessment Tool: An Analysis of Student Proficiency against Bloom's Taxonomy. Proceedings of the Canadian Engineering Education Association (CEEA). 2022;doi:10.24908/pceea.vi.15925.

39. Sotaquirá-Gutiérrez R, Beltran LM, and JPGR. Hackathons as experiential learning platforms for engineering design skills. Cogent Education. 2025;12(1):2442187. doi:10.1080/2331186X.2024.2442187.

40. Handelsman J, Ebert-May D, Beichner R, et al. Scientific Teaching. Science. 2004;304(5670):521–522. doi:10.1126/science.1096022.

41. Hora MT, Ferrare JJ. Instructional systems of practice: A multidimensional analysis of math and science undergraduate course planning and classroom teaching. Journal of the Learning Sciences. 2013;22(2):212–257. doi:10.1080/10508406.2012.729767.

42. Dolan EL. Biology education research 2.0. CBE—Life Sciences Education. 2015;14(4):ed1. doi:10.1187/cbe.15-11-0229.

KNOWLEDGE FOR TOMORROW –
COOPERATIVE RESEARCH PROJECTS IN SUB-SAHARAN AFRICA

Kenya 2022
Summer School
in Computational Biology
for Sustainable Environment

Program

Summer School in "Computational Biology for Sustainable Environmental Research"

WEEK 1

		Mo 5.09	Tue 6.09	Wed 7.09	Thu 8.09	Fr 9.09	Sa 10.09	Su 11.09
Morning Sessions	8:00 - 09:30	Arrival	Different facades of computational Biology	Intro to Python: numpy & pandas	Working with data sets using pandas	Intro to modelbase Building mathematical models	Field trip to the Conservation Sites	Day Off
	Coffee break							
	10:00 - 11:30		Intro to Python: syntax & semantic	Intro to Python: numpy & pandas	Working with data sets using pandas	Modelling population dynamics		
Mid daybreak 11:30 - 14:30								
Afternoon practical	14:30 - 16:00	**Opening session** Mission & Short-term goals **Keynote**	Intro to Python: syntax & semantic 1	Intro to Python: numpy & pandas	Introduction to mathematical modelling (ODE)	Zoology and Fishery		
	Coffee break							
	16:30 - 18:00/19:00	**Kick-Off Social Event*** with Poster Session** Get to know for participants. Everyone is presenting their science to each other	Intro to Python: syntax & semantic 2	Discussion: What is sustainability?		Modelling population dynamics		

WEEK 2

		Mon 12.09	Tue 13.09	Wed 14.09	Thu 15.09	Fr 16.09
Morning Sessions	8:00 - 09:30	Intro to the R Project for Statistical Computing	Data Opportunities in Local Organisations	Hackathon with real biological and environmental data provided by the Local Ocean Conservation Group	Hackathon	Closing remarks combined with the field class at the Snake Farm
	Coffee break					
	10:00 - 11:30	Load, Clean and Manipulate with Data in R	Explorative Data Analysis and Statistical Models in R		Preparing final presentation	
Mid daybreak 11:30 - 14:30						
Afternoon practical	14:30 - 16:00	Explore your Data with Visualizations in R	Explorative Data Analysis and Statistical Models in R		Final Outreach Activity Presentation to the local Community	
	Coffee break					
	16:30 - 18:00/19:00	Explorative Data Analysis and Statistical Models in R	What is a Hackathon? Preparation for the event			

LECTURES INTERACTIVE COMPUTATIONAL SESSIONS INTERACTIVE SESSIONS

Fig S1. Case Study 1: Comprehensive Program Schedule for the Summer School in 2022, detailing the sequence of sessions, topics covered, and key activities designed to enhance participants' learning experiences, including a 25-hours hackathon.

2023 Hackathon Workshop on Computational Modeling to Unleash the Computational Power of Young African Researchers						
		MONDAY 21/08/2023	TUESDAY 22/08/2023	WEDNESDAY 23/08/2023	THURSDAY 24/08/2023	FRIDAY 25/08/2023
Morning Sessions	8:00 - 9:00	Registration and Welcome	**Foundations of dynamic modeling**	**Better Graphs** Practical tips to improve the data visualizations you are working on	Registration to the Public Event	**Open Hackathon Session**
	9:00 - 10:30	**Opening session** Mission & Short-term goals **Keynote**	**Intro to Data Visualization**	**Intro to building mathematical models** and efficient use of on-line resources	**PUBLIC LECTURE** **Collaborative Research to Maximize the Potential of Data-Driven Approaches in Interdisciplinary Breakthroughs.**	
	Coffee Break					
	11:00 - 13:00	**Public Health, Epidemiology, and Models**	**Data Visualization with Pandas**	**Modelling Population Dynamics**		
Afternoon practical	13:00 - 14:00	Lunch Break				
	14:00 - 16:00	**Introduction to Computational Biology and Modelling**	**Working with real-world data sets**	**Modelling Population Dynamics**	**Open Hackathon Session**	**Discussion Future of Computational Biology: Data Science and AI**
	Tea Break					Tea Break
	16:30 - 18:00	**Introduction to programming with Python**	Social Event	**Model Fitting to Data**		Closing dinner

LECTURES INTERACTIVE COMPUTATIONAL SESSIONS INTERACTIVE SESSIONS

Fig S2. Case Study 2: Comprehensive Program Schedule for the Hackathon Workshop in 2023, detailing the sequence of sessions, topics covered, and key activities designed to enhance participants' learning experiences, including a 14-hours hackathon.

Table S2. Summary of the open answer of the survey distributed among participants of the Case Study 2. Participants' answers to the question "*After participating in this hackathon-style workshop, what are your opinions on incorporating the hackathon teaching approach into your undergraduate or postgraduate curriculum? This approach involves multiple instructors in a classroom setting, with hands-on practical activities following the introduction of each concept, including programming activities and group challenges to test the understanding of concepts and computational skills.*" The right column shows the interpreted support shown in Fig 4B. If an answer contained no explicit support for the incorporation of hackathons, it was counted as neutral.

Answer	Support
Incorporating multiple instructors	Neutral
I think that would be amazing. Most often than not we don't get this kind of training setup in our educational systems. It would be a different experience for students and would encourage practical learning. I wish I had this kind of exposure earlier on in my career.	Support
The strategy should be implemented since the student will learn many skills.	Support
It brings out the best in individuals.	Support
I wish we stop just talking and applying it. It is the best way. I am using it now. I wish I went through such a system	Support
I guess incorporating this teaching approach would likely require careful planning and resources, though it definitely has the potential to significantly enhance the quality of education for students, particularly in fields that involve computational and practical skills and this would better prepare students for the demands of the modern workforce. It aligns well with the goal of producing graduates who are not just knowledgeable but also skilled and adaptable problem solvers. So, I believe.	Support
Yes will incorporate coding	Support
THE STYLE ENGAGES EVERYONE UNLIKE THE NORMAL CLASS SET UP	Support
The skills and experience gained am going to apply them in my research since they are relevant and useful	Neutral
Grouping learners is very important Knowledge in programming was very useful	Support
Undergraduates and postgraduates To be taught these skills	Neutral
I believe that incorporating the hackathon teaching approach into our curriculum would not only enhance our academic experience but also prepare us better for real-world challenges. It promotes active learning, critical thinking, and practical skills development, making it a valuable addition to our education.	Support
It is a good concept as people are more likely not to forget what they learn with a friend in a friendly setting	Support
Am up to it. It was a good experience.	Support
Bioinformatics and computational skills should be introduced in undergraduate curriculum	Neutral
yes it very effective	Support
It's a good teaching and learning approach	Support
I think the hackathon teaching approach can be a valuable addition to undergraduate or postgraduate curriculums. It can help students to learn in a more hands-on and collaborative way, and it can help them to develop the skills and knowledge they need to succeed in the workplace.	Support
I love the idea	Support
I think it's the best idea since the instructors will focus more on a learner as an individual	Support
This will definitely be a game changer. In our out of workshop discussions with my colleagues we all agreed that what we gained is what a normal student will gain in a whole semester. So imagine if the approach was used in postgraduate curriculum, then within a year we will definitely have a cure fir cancer, HIV and Covid-19. With do doubt I highly support that.	Support
Scientific python makes data analysis easier as compared to other mathematical softwares	Neutral