

# Ten Simple Rules on How to Organise a Bioinformatics Hackathon

Susanne Hollmann<sup>1,2</sup>, Babette Regierer<sup>2</sup>, Teresa K Attwood<sup>3</sup>, Andreas Gisel<sup>4,5</sup>, Jacques Van Helden<sup>6</sup>, Gregoire Rossier<sup>7</sup>, Paul J Kersey<sup>8</sup>, Eija Korpelainen<sup>9</sup>, Gert Vriend<sup>10</sup>, Erik Bongcam-Rudloff<sup>11</sup>✉

<sup>1</sup>Focus Area Plant Genomics and Systems Biology, Institute of Biochemistry and Biology, Potsdam University, Potsdam, Germany

<sup>2</sup>SB Science Management UG (haftungsbeschränkt), Berlin, Germany

<sup>3</sup>The University of Manchester, Manchester, United Kingdom

<sup>4</sup>CNR Institute for Biomedical Technologies, Bari, Italy

<sup>5</sup>International Institute for Tropical Agriculture, Ibadan, Nigeria

<sup>6</sup>Department of Theory and Approaches to Genome, Institut Français de Bioinformatique, Évry, France

<sup>7</sup>Swiss Institute of Bioinformatics, Lausanne, Switzerland

<sup>8</sup>EMBL-European Bioinformatics Institute, Hinxton, and Bioinformatics and Genomics, Royal Botanic Gardens, Surrey, United Kingdom

<sup>9</sup>IT Center for Science, Espoo, Finland

<sup>10</sup>Centre for Molecular and Biomolecular Informatics, Radboudumc, Nijmegen, Netherlands

<sup>11</sup>Swedish University of Agricultural Sciences, Uppsala, Sweden

Competing interests: SH none; BR none; TKA none; AG none; JVH none; GR none; PJK none; EK none; GV none; EBR none

## Abstract

The completion of the human genome sequence triggered worldwide efforts to unravel the secrets hidden in its deceptively simple code. Numerous bioinformatics projects were undertaken to hunt for genes, predict their protein products, function and post-translational modifications, analyse protein-protein interactions, etc. Many novel analytic and predictive computer programmes fully optimised for manipulating human genome sequence data have been developed, whereas considerably less effort has been invested in exploring the many thousands of other available genomes, from unicellular organisms to plants and non-human animals. Nevertheless, a detailed understanding of these organisms can have a significant impact on human health and well-being.

New advances in genome sequencing technologies, bioinformatics, automation, artificial intelligence, etc., enable us to extend the reach of genomic research to all organisms. To this aim gather, develop and implement new bioinformatics solutions (usually in the form of software) is pivotal. A helpful model, often used by the bioinformatics community, is the so-called hackathon. These are events when all stakeholders beyond their disciplines work together creatively to solve a problem. During its runtime, the consortium of the EU-funded project AllBio - Broadening the Bioinformatics Infrastructure to cellular, animal and plant science - conducted many successful hackathons with researchers from different Life Science areas. Based on this experience, in the following, the authors present a step-by-step and standardised workflow explaining how to organise a bioinformatics hackathon to develop software solutions to biological problems.

## Introduction

The vast advances in technologies of the past decade enabled researchers to reach genomic research to quit all organisms. Among other large-scale sequencing initiatives for plants, microbes or animals like fish, the **Earth Biogenome Project**<sup>1</sup> (EBP) was launched

<sup>1</sup>[www.earthbiogenome.org](http://www.earthbiogenome.org)

to sequence the DNA of all life on Earth within the next decade. This long-term project will lead both to a greater understanding of Earth's biodiversity and responsible stewardship of its resources, tackling the new millennium's most crucial scientific and social challenges. While the focus of the EBP is on collecting genomic data, other initiatives have centred on data

## Article history

Received: 02 December 2020

Accepted: 12 May 2021

Published: 19 October 2021

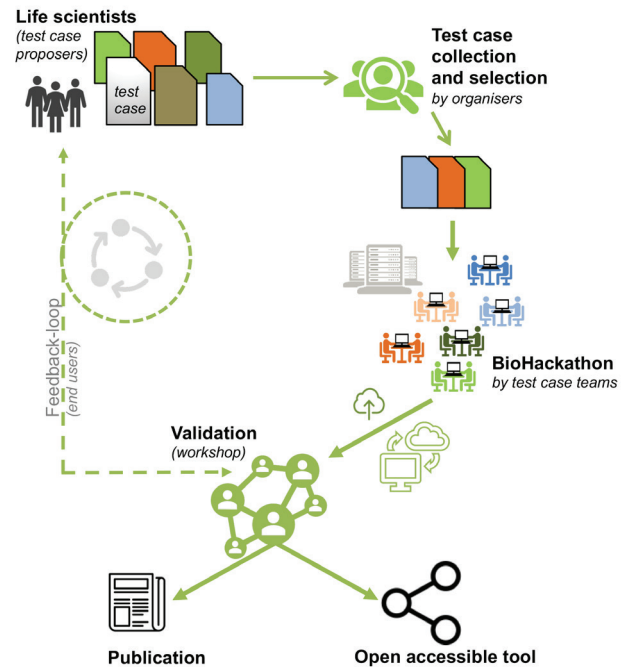
© 2021 Hollmann *et al.*; the authors have retained copyright and granted the Journal right of first publication; the work has been simultaneously released under a Creative Commons Attribution Licence, which allows others to share the work, while acknowledging the original authorship and initial publication in this Journal. The full licence notice is available at <http://journal.embnet.org>.

analysis. For example, the EU-funded project AllBio<sup>2</sup> - Broadening the Bioinformatics Infrastructure to unicellular, animal and plant science (FP7 GA 289452) - concentrated on non-human genomes, applying human-genome-derived computational solutions to non-human organisms. This project involved collecting a range of biological problems, the so-called test cases, and the relevant computational solutions. Some of these test cases were worked out in detail during hackathons. A hackathon is a short (1-day to 1-week) event where stakeholders with diverse skills and backgrounds gather to develop and implement solutions (usually in the form of software) to relevant problems. The term hackathon is a composite of the words 'hack' (meaning exploratory programming) and 'marathon' (a common metaphor for long and intensive events). Hackathons are typical in informatics communities but still relatively new to the life sciences. In part, this may be because there are still considerable communication gaps between life- and computational-science researchers. Bioinformatics hackathons or bio-hackathons aim to address such gaps by bringing IT professionals (and interested amateurs) and life science scientists together to communicate and exchange ideas around practical research questions. These type of events can indeed be highly productive for interdisciplinary teams to solve well-defined problems or to accelerate solution provision in a particular area (hackseq Organising Committee 2016 (2017); Friedberg *et al.*, 2015; Poncette *et al.*, 2020; Braune *et al.*, 2021), generate innovations (Lyndon *et al.*, 2018) or serve as educational tools (Silver *et al.*, 2016; Wang *et al.*, 2018). Here, the co-development principle involving the problem providers and developers in the entire process will ensure that a suitable solution is created. Online events might work for the technical solution creation but will most likely lack the lively interdisciplinary interaction.

The short time usually available for bio-hackathons generally allows for the design and implementation of prototype solutions. For the outputs to be helpful, the developed code must have the potential to undergo subsequent development by interested parties. Therefore, all results should be made available via openly accessible platforms to allow researchers to improve the product after the event is terminated.

The philosophy of the AllBio project was to solicit life science scientists to identify topic challenges directly. Around 60 of such test cases were collected via questionnaires and interviews, out of which 15 (encompassing unicellular organisms, plants, and farm animals) were deemed solvable with adaptations to software or workflows initially designed for human-genome data (Bongcam-Rudloff *et al.*, 2019). Eight were subsequently addressed in bio-hackathons (Amar *et al.*, 2014; Gomez-Cabrero *et al.*, 2014; Leung *et al.*, 2015). A problem was considered solvable when:

- a generic question relating to the analysis of a unicellular, animal or plant genome had been well defined;



**Figure 1.** Identification of test cases. AllBio workflow illustrating the fate of test-cases proposed by life science scientists. After initial interviews, the test cases were collected and assessed for their tractability. The bio-hackathon teams comprised the proposer (life science scientist), a leader (bio-informatician), hackers (programmers) and, usually, a local organiser. Where a tool or meta-tool arose from work, it was proposed for testing during a validation workshop. Ultimately, the Team prepared an open-source tool and published or otherwise disseminated the results.

- a community of domain-expert bioscientists and bio-informaticians had been formed; and
- scientific meetings (in vivo or in-silico) had already taken place, and collaborations had begun.

The workflow for AllBio bio-hackathons involved collecting and selecting the test cases, preparing and organising the events, and finally - in the chance of success - publish the results (Fig. 1).

During the AllBio project, a rigorous regime of evaluating past events allowed each bio-hackathon to build on lessons learned from previous ones. This iterative process demonstrated that the events must be well-prepared and long in advance for bio-hackathons to be successful. The biological problems they set out to tackle must be tractable. They must have access to requisite computational infrastructure and sufficient time to complete the necessary tasks. Other essential pre-requisites are efficient leadership, an appropriate mix of skills/expertise, and effective communication strategies. A preparatory phase should precede bio-hackathons to check feasibility and practicability, e.g., can the data be moved around and read? There should also be commitments afterwards to finalise any tools (or outputs), to test and validate them with end-users, and to disseminate the results. Based on the experience gained in AllBio, we present ten rules that we believe are

<sup>2</sup>[www.allbioinformatics.eu](http://www.allbioinformatics.eu)

crucial when organising bioinformatics hackathons, or bio-hackathons: these fall into four main categories - the Problem, the Team, the bio-hackathon and the Answer, which are described in detail below. There will, of course, be other important considerations (funding, *etc.*), but we focus here on the practicalities of organising successful bio-hackathons. The presentation of the described process is following the scheme of the [Ten Simple Rules series of PLOS](#)<sup>3</sup>.

## The Ten Rules

### The Problem

#### Rule 1: Understand the biological Problem (s) and select the theme

It might seem self-evident to state that a good starting point is to understand the Problem before trying to address it. But, solving biological problems via hackathons requires a spectrum of understanding that encompasses the biology of the Problem (including *in vivo* aspects), the nature of the data available, computational requirements and expected output(s), and how all of these can be brought together to implement a viable solution. One of the keys to success is that those responsible for implementing the technical solution(s) must appreciate, at least at some level, the underlying biology. Ultimately, this requires some investment of time to allow them to begin to understand the language of those whose biological problems they are trying to solve. One way to help achieve this, even for small events, might be to run a small cycle of webinars before the event to give participants more information about the theme. This is likely to facilitate team building and may also provide opportunities to come up with new ideas for possible approaches and solutions.

#### Rule 2: Ensure that the Problem is tractable

Bio-hackathons are driven by practical research questions, but not all biological problems are amenable to solution by hackathons. An early step in setting up any such event should therefore be to estimate whether the size of the Problem is compatible with the hackathon format. For example, while *de novo* software design is generally not the goal of hackathons (design of new algorithms tends to require more than just a few days), proof-of-concept implementations can fit the format quite well. Ideally, therefore, the necessary software components must already exist so that bio-hackathon sessions can readily combine them into bespoke workflows. Ideally, workflows should not contain any single point of failure. Notably, both the biological datasets and the software components must be available without restrictions.

### The Team

#### Rule 3: Put together the right Team with carefully assigned roles

Start building the Team as soon as possible. Ideally, establish the core group two months before the event. Think about life- and computational-science colleagues and students who have the requisite skills and knowledge in the Problem area. Generate a checklist with the minimal requirements needed to ensure that the complete project can be implemented during the event. This will form the basis for participant selection. If necessary, promote the bio-hackathon widely (*e.g.*, using social media), providing as much information about the event as possible (including when, where, what, how, fees - if needed - and registration forms). Some incentives might be helpful to engage bio-hackathon participants, such as cooperation with university groups that might be willing to give credit points for participation or formulating problems whose solutions are suitable for academic publication and crediting those participants as authors.

Biohackathon teams are generally most effective when they comprise no more than eight to ten participants. In general, they should include a proposer or biological Problem owner, typically a life science scientist, whose needs will drive the event. A leader, usually a bioinformatician. The hackers, bioinformaticians and computer scientists and, ideally, an overall organiser/coordinator. Those with computational skills should include at least one IT professional or bioinformatician and programmers with experience in scripting, workflow design, use of ontologies, evaluation of data quality, and so on. These professionals must be able to communicate effectively with the leader and remain focused on the primary objective.

The bio-hackathon leader is responsible for monitoring and guiding the workflow during the event. The organiser must take responsibility for the overall coordination of the event, maintaining good communication within the Team (rule 4), orchestrating the validation (rule 9) and dissemination (rule 10) activities. The organiser must be local to the venue of the bio-hackathon and will be responsible for many mundane practical tasks: reserving the venue, testing bandwidth in the meeting room before the actual hackathon, providing travel instructions, communicating with the compute provider, selecting the participants and dealing with subsistence/refreshment issues, *etc.* One person may assume several roles, but it is vital that each partner knows his/her role and that all roles are maintained before, during and after the hackathon itself. To facilitate discussion and assignment of tasks as the project progresses, we suggest adopting a convenient communication platform, *e.g.*, [Trello](#)<sup>4</sup>, [Slack](#)<sup>5</sup> or

<sup>3</sup><https://collections.plos.org/collection/ten-simple-rules/>

<sup>4</sup><https://trello.com>

<sup>5</sup><https://slack.com>



comparable platform such as [ownCloud](https://owncloud.org)<sup>6</sup>, [GoogleDrive](https://drive.google.com)<sup>7</sup> or [Dropbox](https://www.dropbox.com)<sup>8</sup>.

#### Rule 4: Communicate effectively and establish the ground rules

Communication – before, during and after hackathons – is key. The value of good communication, and the impact of not getting it right, is hard to over-emphasise. Bio-hackathons include partners from different disciplines who tend to speak very different languages. If a bio-hackathon is to be maximally productive, it is critical to take time, early on, to identify and resolve potential language barriers. Frequent conversations before the bio-hackathon (in person if possible, or electronically if not) are essential to understand, define and refine the biological question, identify and shape the overall analytical approach, and thence to build ownership of the tasks. As the technical partners assimilate the nature of the biological Problem and the biological partners begin to appreciate the heart of the technical challenges, the Team's purpose, focus and cohesiveness will mature.

If multiple projects are being tackled in one bio-hackathon, ensure that all requirements have been established beforehand, including the process of team-building, the time-frame available for each Problem (equal conditions for every Team, so that each has the same relative chance of success), and the rules for allowing participants to move between teams.

#### Rule 5: Prepare the ground-work well in advance

Bio-hackathons are generally time-limited; good preparatory work is therefore essential. A crucial part of the preparation is to test the necessary software and hardware before the event to prevent problems that could reduce the time available for hands-on work. Any heavy computational tasks should be pre-computed to allow participants to hit the ground running with real data. Bio-hackathon leaders must, therefore, comprehensively understand all the components in advance, arrange to have them tested in good time, and ensure that both software tools and hardware facilities are adequate for the tasks at hand. For example, CPU-intensive tasks might require massive pre-calculations or specialised equipment (such as all-against-all [BLAST](https://blast.ncbi.nlm.nih.gov/Blast.cgi)<sup>9</sup>) computations on datasets with millions of sequences, or the assembly of large genomes). Just as important is verification of the quality of any datasets to be used during the event, as poor-quality datasets are likely to jeopardise the success of bio-hackathon sessions. To not waste valuable time, any task that can be tackled by a participant in isolation (without requiring the insight of the entire Team) should be completed in advance. It is vital to test all software and hardware before the event. Work with the hackers to establish the hardware requirements. Ensure that

hardware equipment/components can be provided or temporarily replaced if need be.

Prepare a budget forecast for the event. The budget will be dedicated to the rental of premises, IT requirements and subsistence. Gather options of suitable venues and their prices. Look at the premises and find out what the rental includes. Fix the premises for the scheduled date.

Decide the total amount you can spend on subsistence. We recommend creating a spreadsheet of all costs. If you have no funds available, you will need to set a fee (which will ultimately be determined by the number of participants, including lecturers, organisers, and so on). If you do have to set fees, you should also be aware of the potential fiscal risks. Involve your administration in the process to ensure that you do not run into trouble: they will know best how to treat fee income. If feasible, search for potential sponsors – e.g., companies with an interest in your bio-hackathon theme.

The accuracy in silico simulation datasets is of great importance for benchmarking bioinformatics tools as well as for experimental design. For that reason, it is a good recommendation to create simulated sequencing data (mock data). For this purpose, there are now several freely available software packages to simulate mock data. Two examples are ART (Huang *et al.*, 2012) and InSilicoSeq (Gourlé *et al.*, 2019). When selecting the bio-hackathon venue, the proper mock data can be chosen accordingly. If the hackathon is organised in an academic environment with high computational capability, the mock data could be of substantial size. The data simulated can be the minimal required to perform the proper testings if, for travel logistics, a hotel close to an airport is chosen.

We recommend creating a checklist for all tasks to be done before, during and after the event. Spread responsibility between the organisers, but ensure that they do their job seriously. Discuss and agree on the rules and procedures, and take care that rules are followed strictly. Figure 2 collates the organisational workflow for a complete bio-hackathon cycle, including the preparatory, implementation and follow-up phase.

## The bio-hackathon

### Rule 6: Choose a convenient location

Bio-hackathons should take place at convenient locations for the registered number of participants, and locations have to fulfil all scientific/computing and non-scientific (housing, food, *etc.*) needs. University/national computing centres are likely to offer excellent computational facilities but may have restrictive opening hours. Hotels, on the other hand, while often very convenient in many aspects, may overestimate the bandwidth they can provide, so this needs to be tested extensively upfront.

Specific requirements to consider include:

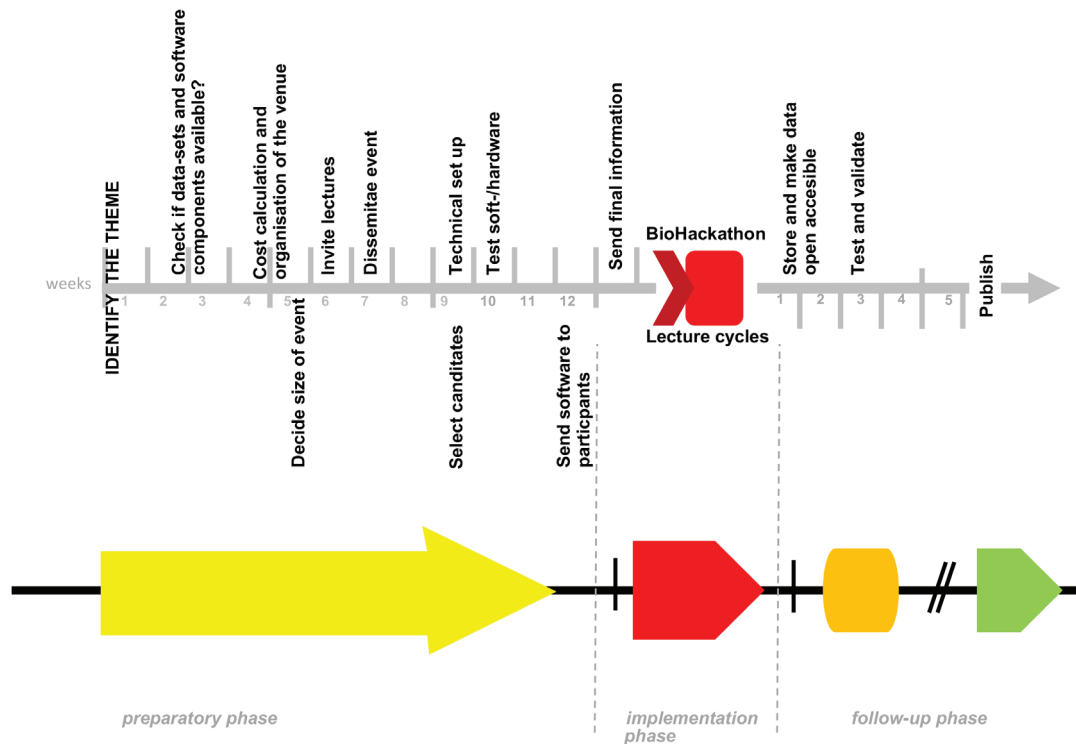
- location convenient for participants to reach (minimise travel time and cost);

<sup>6</sup><https://owncloud.org>

<sup>7</sup><https://drive.google.com>

<sup>8</sup><https://www.dropbox.com>

<sup>9</sup><https://blast.ncbi.nlm.nih.gov/Blast.cgi>



**Figure 2.** Workflow. The scheme demonstrates an optimal workflow for bio-hackathons, including the preparatory, implementation and follow-up phase for a complete cycle. Each phase is subdivided into different consecutive steps: in particular, the preparatory phase comprises a broad spectrum of tasks, including the selection of challenges, recruiting of participants, organisation of the venue and technical set-up, as well as the creation of webinars to prepare participants for the event.

- short distance between accommodation and meeting venue (if the venue is not the hotel);
- venue technically well equipped (beamer, screen, *etc.*), with liberal opening hours (often, much work is done outside regular working hours, and it is essential to facilitate this);
- venue has sufficient and stable bandwidth;
- food and drink are either available at the venue or allowed to be brought in. Often, many productive discussions occur informally over dinner, so arrangements that encourage the participants to keep together while eating are strongly preferred.

Of course, these events can also be conducted remotely if, for reasons such as the current pandemic situation, a face-to-face event is not possible. If so, Rule 6 is omitted and needs to be replaced by the organisation and establishment of a suitable online meeting tool (*e.g.* Zoom, [Big Blue Button](https://www.bigbluebutton-hosting.de/)<sup>10</sup>, [Microsoft Teams](https://www.microsoft.com/de-de/microsoft-teams/group-chat-software)<sup>11</sup>, Slack). However, since bio-hackathons are based on an intensive and iterative exchange between biologists and computer experts, it is always preferable to hold them offline.

#### Rule 7: Ensure appropriate computer access

All bio-hackathons are not equal: some will have greater computational requirements than others. Some analyses might run efficiently on participants' laptops; some

<sup>10</sup><https://www.bigbluebutton-hosting.de/>

<sup>11</sup><https://www.microsoft.com/de-de/microsoft-teams/group-chat-software>

might require access to large clusters, supercomputers, dedicated hardware, or the cloud, which universities or national computer centres may be willing to provide. Regardless, the pre-requisites are i) fast internet connection at the hackathon venue, and ii) possibility for remote login to the computes facilities before and after the event. This last point is essential to prepare the ground-work beforehand, whereas any remaining work can be completed later. The local organiser should ensure (and check) that logins are available for all participants and ideally perform a test run before the bio-hackathon.

Similarly, if participants use their laptops, the requisite software should be installed before the event. It is recommended to create a Virtual Machine to provide a common computing environment for participants. To gain an overview of the software and hardware that will be needed during the bio-hackathon, we recommend gathering information about technical requirements via the registration form. Share this information with the hackers at the latest ten days before the event.

#### Rule 8: Ensure the duration is sufficient to obtain valuable outputs

Bio-hackathons are short, intensive working sessions, typically spanning a few days. Several considerations determine the duration of these events: the complexity of the workflow, how much computer work is envisaged (and how much can be done in advance), the funds

available, how much time participants can commit, and whether writing documentation or article outlines are also intended to be part of the exercise. The expected outputs must therefore be clearly defined early on, and the duration of the event adjusted accordingly. It generally works well to organise hackathons over a weekend, as this affords participants greater flexibility with their schedules.

To kick off the event, plan to run a series of short lectures to better inform participants about the theme of the bio-hackathon and introduce its biological and computational components. Ensure the availability of suitably qualified lecturers. Disseminate information about these lectures to the participants and a broader audience at the latest two weeks before the hackathon. This may stimulate greater interest in the event and gain visibility within the community.

The lecture hall and workspaces might be at different locations. Ensure that you provide sufficient and detailed information about where and when to go to each place. If there is insufficient space to accommodate additional lecture series participants comfortably, focus on briefing the Team. This can also be done in the form of webinars before the event.

## The Answer

### Rule 9: Validate the results

Bio-hackathons aim to address particular biological problems. The events may focus on prototyping ideas, or they may lead to the production of tools or meta-tools that will ultimately be made available to the community. Before public release, validation events should be organised, in which participants are given opportunities to test the tool(s) with a variety of different datasets. Even though validation is normally done after hackathons, it should nevertheless be part of the initial planning to ensure that validation data exist, and that the software set-up is sufficiently generic to allow its use in validation. In an ideal case, most (if not all) of the original bio-hackathon Team should be present or (remotely) available during validation sessions.

### Rule 10: Disseminate the results

Peer-reviewed publications are still the primary vehicles for disseminating scientific results, and reusable outputs from bio-hackathons are a good stimulus for article publication. However, public accessibility of all workflows must also be part of the dissemination strategy. Therefore, only open access publication platforms such as [F1000research](https://f1000research.com/)<sup>12</sup> should be used for publications. To maximise the outcomes impact, all workflows should also be properly documented and licensed, and inputs and outputs should be appropriately described following the FAIR principles (Wilkinson *et al.*, 2016) and using standardisation measures (Hollmann *et al.*, 2021). Ideally, alongside any publicly accessible documentation or article, small datasets that the workflows can use

<sup>12</sup><https://f1000research.com/>

should also be included, inclusive of its corresponding Standard Operating Protocols (Hollmann *et al.*, 2020).

Optionally, Virtual Machine images to run workflows might also be provided. Results should be made available through openly accessible platforms such as [SEEK](https://seek4science.org/)<sup>13</sup>, [OpenAIRE](https://explore.openaire.eu/participate/deposit-publications)<sup>14</sup>, [zenodo](https://zenodo.org)<sup>15</sup> or [GitHub](https://github.com/)<sup>16</sup> that can guarantee longevity, as good workflows that answer biological questions often remain valuable for several years.

## Potential pitfalls

The experience of the AllBio bio-hackathons provided an inside view of potential pitfalls that might limit the success of such events. A primary challenge is the careful selection of appropriate Problems; not all are suitable for inclusion in a bio-hackathon. It requires expert knowledge from both the biology and bioinformatics fields to evaluate the challenges and avoid frustration for the participants.

A specific function that bio-hackathons can perform is to enable interdisciplinary collaboration between the participants from the different expert fields. Sufficient time needs to be dedicated to training participants and finding a common language to discuss the challenges and develop efficient solutions.

Other more practical aspects may limit the success of events: *e.g.*, some early AllBio bio-hackathons struggled to deliver concrete outputs because:

- their teams were too small ( $\leq 5$  people);
- the Team had no real leadership;
- the datasets on which they were obliged to work were too large to be processed fruitfully within the given time frame;
- the opening hours of computing centres limited the time available for productive work;
- the distance between hackathon venues and participants' hotels posed time and cost constraints.

A barrier to success may also occur if the meeting organiser/leader is no longer available after the event. The validation and follow-up phase is essential for summarising the results and ensuring the quality of solutions that have been developed. Moreover, publication of the results, whether via a journal article or upload to a repository, needs to be completed after the bio-hackathon. Costs associated with the dissemination of results need to be considered in the overall budget plan.

## Conclusions

Bio-hackathons were powerful tools in the AllBio project for articulating and solving problems in the scientific community. They highlighted the need to consider the different disciplinary backgrounds of all participants, hence the vital role of the preparatory phase for ensuring

<sup>13</sup><https://seek4science.org/>

<sup>14</sup><https://explore.openaire.eu/participate/deposit-publications>

<sup>15</sup><https://zenodo.org>

<sup>16</sup><https://github.com/>

the success of events. They also provided excellent opportunities, especially for young researchers, to learn new skills at the interface between disciplines, participate in advancing their field of research, and gain unique hands-on training with real challenges.

Some of the rules listed here may seem obvious, trivial, or even superfluous; nevertheless, all proved crucial in real-life scenarios. The ten rules provide practical guidelines for future bio-hackathon organisers, including preparations before, during and after the event itself.

### Key Points

- New advances in sequencing technologies, bioinformatics, automation, artificial intelligence, *etc.*, to tackle this there is a need for continuous development of new bioinformatics solutions.
- Bio-hackathons are a helpful model to create bioinformatics solutions, often used by the bioinformatics community,
- Based on the work from the ALLBIO project this article present a step-by-step and standardised workflow explaining how to organise a bioinformatics hackathon.

## Acknowledgments

The AllBio project was funded by the European Framework Programm GA 289452. The hackathon team members gratefully acknowledge SARA for providing computer facilities and hosting one hackathon. We thank NBIC's BioAssist programme for helping us find excellent programmers and allow them to spend adequate time on the projects.

## References

1. Amar, D., Frades, I., Danek, A. *et al.* Evaluation and integration of functional annotation pipelines for newly sequenced organisms: the potato genome as a test case. *BMC Plant Biol* **14**, 329 (2014). <http://dx.doi.org/10.1186/s12870-014-0329-9>
2. Braune K, Rojas PD, Hofferbert J, Valera Sosa A, Lebedev A *et al.* (2021) Interdisciplinary Online Hackathons as an Approach to Combat the COVID-19 Pandemic: Case Study. *J Med Internet Res.* **23**(2):e25283. <https://www.doi.org/10.2196/25283>
3. Bongcam-Rudloff E *et al.* (2019) Broadening the bioinformatics infrastructure to unicellular, animal, and plant science D5.3 - Final summary report on solved test cases. Zenodo, <http://dx.doi.org/10.5281/zenodo.3525052>
4. Friedberg I, Wass MN, Mooney SD, Radivojac P (2015) Ten simple rules for a community computational challenge. *PLoS Comput Biol.* **11**(4):e1004150. <http://dx.doi.org/10.1371/journal.pcbi.1004150>
5. Gomez-Cabrero D, Abugessaisa I, Maier D, Teschendorff A, Merckenschlager M *et al.* (2014) Data integration in the era of omics: current and future challenges. *BMC Syst Biol.* **8**(Suppl 2):I1. <http://dx.doi.org/10.1186/1752-0509-8-S2-I1>
6. Gourlé H, Karlsson-Lindsjö O, Hayer J, Bongcam-Rudloff E (2019) Simulating Illumina metagenomic data with InSilicoSeq. *Bioinformatics* **35**(3),521-522. <http://dx.doi.org/10.1093/bioinformatics/bty630>
7. Hackseq Organizing Committee 2016 (2017) hackseq: Catalysing collaboration between biological and computational scientists via hackathon. *F1000Research* **28**;6:197. <https://f1000research.com/articles/6-197/v2>
8. Hollmann S, Frohme M, Endrullat C, Kremer A, D'Elia D *et al.* (2020) Ten simple rules on how to write a standard operating procedure. *PLoS Comput Biol* **16**(9):e1008095. <http://dx.doi.org/10.1371/journal.pcbi.1008095>
9. Hollmann S, Kremer A, Baebler S, *et al.* (2021) The need for standardisation in life science research - an approach to excellence and trust. *F1000Research* **9**:1398. <http://dx.doi.org/10.12688/f1000research.27500.2>
10. Huang W, Li L, Myers JR, Marth GT (2012) ART: a next-generation sequencing read simulator. *Bioinformatics* **28**(4),593-4. <http://dx.doi.org/10.1093/bioinformatics/btr708>
11. Leung W Y, Marschall T, Paudel Y, Falquet L, Mei H *et al.* (2015) SV-AUTOPILOT: optimised, automated construction of structural variation discovery and benchmarking pipelines. *BMC genomics* **16**(1), 238. <http://dx.doi.org/10.1186/s12864-015-1376-9>
12. Lyndon MP, Cassidy MP, Celi LA, Hendrik L, Kim YJ *et al.* (2018) Hacking Hackathons: Preparing the next generation for the multidisciplinary world of healthcare technology. *Int J Med Inform.* **112**, 1-5. <http://dx.doi.org/10.1016/j.ijmedinf.2017.12.020>
13. Poncette AS, Rojas PD, Hofferbert J, Valera Sosa A, Balzer F *et al.* (2020) Hackathons as Stepping Stones in Health Care Innovation: Case Study With Systematic Recommendations. *J Med Internet Res.* **22**(3):e17004. <https://www.doi.org/10.2196/17004>
14. Silver JK, Binder DS, Zubcevik N, Zafonte RD(2016) Healthcare Hackathons Provide Educational and Innovation Opportunities: A Case Study and Best Practice Recommendations. *J Med Syst.* **40**(7):177. <https://www.doi.org/10.1007/s10916-016-0532-3>
15. Wang JK, Pamnani RD, Capasso R, Chang RT (2018) An Extended Hackathon Model for Collaborative Education in Medical Innovation. *J Med Syst.* **42**(12):239. <http://dx.doi.org/10.1007/s10916-018-1098-z>
16. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M *et al.* (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**:160018. <http://dx.doi.org/10.1038/sdata.2016.18>