Check for updates

OPINION ARTICLE

REVISED *hackseq*: Catalyzing collaboration between biological and computational scientists via hackathon [version 2; peer review: 2 approved]

hackseq Organizing Committee 2016

v2 First published: 28 Feb 2017, 6:197
<https://doi.org/10.12688/f1000research.10964.1>
Latest published: 10 Apr 2017, 6:197
<https://doi.org/10.12688/f1000research.10964.2>

Abstract

hackseq (<http://www.hackseq.com>) was a genomics hackathon with the aim of bringing together a diverse set of biological and computational scientists to work on collaborative bioinformatics projects. In October 2016, 66 participants from nine nations came together for three days for *hackseq* and collaborated on nine projects ranging from data visualization to algorithm development. The response from participants was overwhelmingly positive with 100% (n = 54) of survey respondents saying they would like to participate in future hackathons. We detail key steps for others interested in organizing a successful hackathon and report excerpts from each project.

Keywords

Hackathon, Genomics, Bioinformatics, Open Science, Diversity in Science

This article is included in the **Bioinformatics Education and Training Collection** collection.

This article is included in the **Hackathons** collection.

Open Peer Review**Approval Status** ✓ ✓

	1	2
version 2 (revision) 10 Apr 2017	✓ view	✓ view
	↑	↑
version 1 28 Feb 2017	✓ view	✓ view

1. **Kate L. Hertweck** , The University of Texas at Tyler, Tyler, USA
2. **Jiarong Guo**, Michigan State University, East Lansing, USA

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author:

Competing interests: No competing interests were disclosed.

Grant information: The author(s) declared that no grants were involved in supporting this work.

Copyright: © 2017 hackseq Organizing Committee 2016. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the [Creative Commons Zero "No rights reserved" data waiver](#) (CC0 1.0 Public domain dedication).

How to cite this article: hackseq Organizing Committee 2016. *hackseq: Catalyzing collaboration between biological and computational scientists via hackathon [version 2; peer review: 2 approved]* F1000Research 2017, 6:197 <https://doi.org/10.12688/f1000research.10964.2>

First published: 28 Feb 2017, 6:197 <https://doi.org/10.12688/f1000research.10964.1>

REVISED Amendments from Version 1

We added a brief description of the project leaders and participants role during the hackathon to the research project summaries. We also expanded the discussion of female participation, as suggested by the referees. We also corrected some typos."

See referee reports

Introduction

Technological advances in the biological sciences have led to an increasing availability of so-called ‘-omic’ datasets, allowing fundamental questions in biology to be answered at an unprecedented rate¹. However, these datasets are complex, requiring novel and specialized informatics tools for proper analysis and to overcome the computational bottleneck in research. Open-source bioinformatics tools and pipeline development accelerates the rate of research by allowing the community to both reuse and thoroughly assess such methods. Thus, by solving biological problems in an open and collaborative manner, the field can progress at a faster rate than if code remains unavailable to the larger community².

Hackathons offer a solution to catalyze tool and pipeline development for biological data science, as well as foster interdisciplinary collaborations³. These events aim to solve defined computational problems over a period of days by bringing together small teams of individuals with different and diverse skillsets. Although frequently valuable for the outputs they generate, hackathons have faced criticism due to low levels of diversity amongst participants⁴. We therefore established *hackseq*, a genomics hackathon collective (<http://www.hackseq.com>) that aims to promote open science, collaboration and diversity. We placed special emphasis in promoting leadership amongst women, minorities and early-career scientists. The inaugural *hackseq* event took place over three days in October 2016 in Vancouver (British Columbia, Canada) and was a satellite event to the annual American Society of Human Genetics (ASHG) meeting. Here we report a summary of this hackathon in the hopes of promoting similar events in the future.

hackseq format

hackseq was the first genomics hackathon in Vancouver and was based on the NCBI hackathon format³. Some hackathons can be perceived as high-pressure events exclusive to technically inclined and experienced individuals. We thus took measures to ensure that people of all skill levels and backgrounds were encouraged to apply. We structured *hackseq* as a three-day event that runs primarily from 8AM – 5PM on the Saturday/Sunday/Monday prior to the 2016 ASHG meeting. The *hackseq* itinerary is accessible on the *hackseq* github repository (https://github.com/hackseq/October_2016/blob/master/hackseq_2016_schedule.md).

First, we opened a call for ‘team leaders’ to propose a project and lead a small team at *hackseq* through social media, such as Twitter, making announcements at the Vancouver Bioinformatics User Group (VanBUG) and bioinformatics.ca, as well as direct email contact with potential leaders. We screened the projects

to confirm that their aims and scope would be appropriate for a 72 hour hackathon. For the ten accepted projects we used GitHub as a discussion board, creating issue threads (https://github.com/hackseq/hackseq_projects_2016/issues) for each project, allowing prospective participants to view and discuss the details about each project before applying to join a particular team.

Once we established the ten *hackseq* projects, we opened the call for participants. Our main goal in recruiting participants was to reach out to a diverse group of individuals and to promote participation of women, minorities and early-career scientists. To this end, we partnered with organizations, such as Society for Canadian Women in Science and Technology (SCWIST) and VanBUG, to attract local participants. To encourage early-career scientist involvement, we contacted undergraduate and graduate-level computational sciences and bioinformatics programs at regional universities. To reach the global scientific audience, we contacted several human genetics societies around the world asking them to email participant application information to their respective mailing lists.

To promote economic diversity and lower the barrier to entry for international participants, we partnered with ASHG to create travel awards based on financial need and/or minority status. *hackseq* had no registration fee. Lastly, we made announcements on Twitter, Galaxy Project’s events calendar, and various international conferences, such as Bioinformatics Open Source Conference 2016 and the 13th International Congress of Human Genetics, leading up to the hackathon.

In the participant application form, prospective participants ranked the top four projects on which they would like to work. Participants could apply for travel awards by ASHG and request child care, covered by our budget, to promote participation amongst parents. The organizing committee and the team leaders evaluated the applications based on not only their skill levels, but also their interests and passion for genomics. To ensure well-rounded teams, we considered both the project preferences and skill levels during the team assignment phase, ensuring a balance between novices and expert coders, and biological and computational expertise. All forms developed for *hackseq* are available online (https://github.com/hackseq/October_2016/blob/master/Forms.md).

By defining the projects and teams beforehand, participants got to know their team members and prepare technical infrastructure. Teams hit the ground running on the first day, beginning work unprompted by the organizers at 8AM of the first day.

hackseq had 66 participants in attendance from nine nations, divided into nine teams ranging from 3 to 10 individuals. Of the accepted ten projects, two team leaders withdrew prior to the hackathon for personal reasons, and one popular project split into two teams, resulting in nine teams. The mode age-category was 30–34 years old (62.5%) for team leaders, and 25–29 years old (58%) for participants (Figure 1A). Graduate students made up the largest fraction of participants with 48.2%, followed by academic staff (15.5%), industrial scientists (13.8%), undergraduates (10.3%), postdoctoral fellows (6.9%) and academic faculty (5.2%). Notably, the team

Figure 1. Participant diversity at hackseq 16. To measure diversity of hackseq participants, we asked team leaders and participants to self-report their (A) age, (B) current occupation, (C) ethnicity and (D) gender. Data is shown for team leaders (yellow) and participants (blue).

leaders were more likely to be industry scientists (44.4%) or young faculty (22.2%) (Figure 1B). In total, 22 of 62 (35.5%) participants identified as female and 40 as male. A total of 41% self-identified as Caucasian, 40% as Asian or Pacific Islander, and 19% as Arab, Latin American or unspecified (Figures 1C and D).

Dataset 1. hackseq demographics

<http://dx.doi.org/10.5256/f1000research.10964.d152802>

De-identified demographic data from *hackseq* participants in the pre-meeting survey/confirmation of attendance.

Dataset 2. Post-hackseq survey responses

<http://dx.doi.org/10.5256/f1000research.10964.d152803>

De-identified post-hackseq survey response data for the figures.

Technical and logistical requisites

Hackathons have little essential resource requirements. In this section, we outline the core logistics and technical infrastructure

we employed. While these requisites could be stripped down, our experience was that attention and planning for these details maximized the efficiency of our teams to focus on coding and development.

Core logistics

To encourage participation, *hackseq* had no entry-cost for participants. To ensure teams could focus on the hackathon and not technical or logistical issues, we secured funding for the venue, technical infrastructure, food, transportation and stationery by partnering with different organizations.

A sponsorship package was created to approach different academic, non-profit and industry organizations. Besides asking for financial support, we also made communication and marketing requests given that one of *hackseq*'s goals was to recruit a diverse pool of participants. A strong emphasis was placed on women's groups in science and technology.

In November 2015, we contacted ASHG to ask if we could be a satellite event for their meeting. Given that ASHG 2016 conference was planned to be held in Vancouver, *hackseq* gained

exposure from the ASHG's communication strategy. The ASHG also provided three travel grants to participants based on financial need and diversity.

These partnerships allowed *hackseq* to take place in a large, bright atrium at the University of British Columbia. This allowed all the teams to be in a single-space and interact with one another. Food was provided to minimize distraction and two social events were hosted, one the first night and one on the last night to encourage collaboration and networking amongst participants.

Technical infrastructure

Reliable technical infrastructure is necessary for organizing a successful hackathon; primarily, electrical power, Internet access and computing resources. We ensured the venue had adequate electrical outlets for the participants' laptops and organized a dedicated Wi-Fi network connection be established for the event through the university's information technology office.

Unlike many hackathons, *hackseq* was not restricted to coding. It also included genomic data analyses, which required additional computational resources. To promote reproducibility and collaboration, all the projects were based on pre-organized GitHub teams and repositories (see *hackseq* organization on Github; <https://github.com/hackseq>). To provide teams with reliable and powerful computation, we secured in-kind donation of cloud computing from Amazon Web Services Elastic Compute Cloud (AWS-EC2), and Canada's Michael Smith Genome Science Centre genOmics Research Container Architecture (ORCA). We used Linuxbrew, a cross-platform package management tool, to install bioinformatics software on ORCA⁵.

There was an equal usage of AWS-EC2 and ORCA amongst the participants (43%, not mutually exclusive) and an additional 12% using high-performance computing resources from their resident institutions. Users showed a preference for resources they were previously experienced with, and reported that it was not feasible to learn to use a new computing resource in the given time. Allowing team leaders and participants access to computing resources ahead of time in the future to 'experiment' and familiarize themselves with the different resources is advisable.

Each team chose which programming language and software they used. The majority of participants relied on Python (82.6%) and R (53.8%) programming languages and also used specialized software that related to their particular projects (Figure 2).

In summary, the infrastructure requisites for running a successful hackathon are minimal and many can be acquired as in-kind donations from related organizations. In highlighting the essentials and key lessons, we hope to encourage the motivated reader to run a local scientific hackathon.

Research project summaries

The projects undertaken during *hackseq* were from diverse fields within bioinformatics, ranging from human genomic variation analysis, microbial ecology and transcriptomics, to bioinformatic algorithm development. The projects were proposed by the team leaders, who defined the scope of the work, with the idea that at the end of the 72 hours there will have developed a working prototype. At *hackseq*, the teams organized organically, with team leaders defining the problem and teaching the participants the necessary background while participants offered their expertise on how

Figure 2. Software usage during hackseq 16. At the conclusion of *hackseq*, we asked participants to complete a survey on their experience at *hackseq*. There were 52 responses to the question, "Which programming languages and tools did you and your team use during the course of *hackseq*? (Comma delimited please)." These responses were parsed and the number of unique instances is reported. Languages or software listed <2 is reported as 'Other'.

to implement a solution. In this way a collaborative project goal could be explicitly defined (and re-defined), and the teams would work towards completing that goal together.

Here we provide brief summaries from the projects. Scientific abstracts, videos of final presentations and updated information on each project can be found at www.hackseq.com/projects16.

VASCO: Visualization app for single cell exploration (led by Grace X.Y. Zheng)

Modern transcriptomics analysis tools have limited capacity for analyzing thousands of single-cell RNA-sequencing data (scRNA-seq). VASCO is an intuitive user-interface to visualize gene-cell expression and cell clustering data to explore the relationship between populations of cells and gene expression, including cell cluster of differentiation markers (CD-markers). This project was awarded the “People’s Choice” for the most outstanding project developed at *hackseq*.

XYalign: Hacking sex chromosome variation (led by Melissa A. Wilson Sayres)

Human sex chromosomes violate typical ploidy assumptions made for NGS autosome copy number and variant measurement, which is further confounded by mis-alignment between the X and Y chromosomes. XYalign was developed to measure sex chromosome ploidy and remap reads based on the inferred sex for downstream analysis.

ParetoParrot: A tool to optimize the parameters of command line software (led by Shaun Jackman)

Many bioinformatics software, such as genome sequence alignment and assembly, requires optimization of several input parameters to maximize a target metric. ParetoParrot measured the performance of several ‘black-box’ optimization algorithms to improve the performance of genome sequence assembly software.

BaklavaWGS: Pseudo-WGS variant calling for common cell types aggregating ChIP-seq, RNA-seq and DHS from ENCODE and Roadmap Epigenomics data (led by Luca Pinello)

There is a wealth of sequencing datasets for cell types that have helped to understand and prioritize non-coding variants. Unfortunately, for many of those cell types we still don’t have complete genotype information. BaklavaWGS recovers genotype data from cell lines aggregating sequencing data to aid downstream allele specific analysis. A preliminary analysis is available at <http://www.baklavawgs.com/>.

Evaluating epigenetic modifications in ChIP-seq and methylation data across cell types and states (led by Manuel Belmadani)

A variety of datasets and approaches were investigated for analyzing cell type and state-specific genome regulation. The outcome of the experimental work in exploring differentially methylated

regions from different epigenomic data and public databases, such as ENCODE ChIP-seq, IHEC and JASPAR, is presented.

Selection of tag SNPs for an African SNP array by LD and haplotype based methods (led by Tommy Cartensen)

Commercial SNP arrays fail to capture the diversity of African populations and limit the capacity to conduct large-scale medical genetic studies. Using African whole genome sequencing (WGS) data, an algorithm was developed to quickly identify SNP tags for this population. This will be used to improve upon SNP arrays for this richly diverse continent.

Somatic mutation from separated haplotypes (SMUSH) (led by Patrick Marks)

Calling somatic mutations relies on matched tumour and normal DNA sequencing, but a matched normal sample is often not available. The SMUSH algorithm was developed to differentiate wild type, germline and somatic mutations from linked-read DNA sequencing libraries.

MetaGenius (led by Michael Schnall-Levin)

Analysis of shotgun metagenomic sequencing data is limited in its capacity to assemble over homologous sequences. MetaGenius uses linked-read DNA sequencing to improve the assembly of a mixture of five bacterial species.

mICP: Metagenomic indicator contig predictor (led by Ben Busby)

Metagenomic sequencing has largely focused on 16S rRNA amplicons. This mICP strategy uses a mixture of long PacBio and short Illumina reads to identify contigs from environmental sequencing samples, which predict the environmental state from which they were found.

Discussion

The overarching themes of *hackseq* were inclusivity, open science and collaboration. To gauge the extent to which we were successful in delivering on these themes, we performed a final survey at the conclusion of *hackseq*. Participants overwhelmingly described their experience as positive (Figure 3), with 100% (n = 54) of the survey respondents indicating that they would participate in an event like *hackseq* again and a further 80% indicating that they would like to take on an organizational/ leadership role in future *hackseq* events. Participants specifically highlighted that *hackseq* created ample recruitment, employment and collaborative opportunities, while also exposing participants to different datasets and analyses. We believe this reflects the underlying desire amongst young scientists to share, collaborate and learn from one another. They only need be given the opportunity to do so.

By organizing *hackseq* as a satellite meeting of an international conference like ASHG, we were able to attract team leaders and participants from around the world, including a large proportion of

remote participation options to reduce proximity/cost restrictions. Further improvements could include educational resources to address common technical issues, the provision of an overnight area for participants who would like to continue to work after hours and additional activities to encourage interaction with members from different teams.

Conclusion

The nature of biological sciences has shifted to an increasing emphasis on computational analysis. Collaborative events, such as *hackseq*, offer an exciting platform to bring together a wide spectrum of scientists to work together and innovate. We present demographic information about the first *hackseq* hackathon and encourage future organizers to do likewise, to quantify social inequalities that may be present in such events, and strive to achieve equal representation in the sciences. It's our hope that the information presented here will aid and encourage others in organizing genomics hackathons.

Data availability

Dataset 1: *hackseq* demographics: De-identified demographic data from *hackseq* participants in the pre-meeting survey/confirmation of attendance. doi, [10.5256/f1000research.10964.d152802](https://doi.org/10.5256/f1000research.10964.d152802)⁶

Dataset 2: Post-*hackseq* survey responses: De-identified post-*hackseq* survey response data for the figures. doi, [10.5256/f1000research.10964.d152803](https://doi.org/10.5256/f1000research.10964.d152803)⁷

Author contributions

All members of the *hackseq* Organizing Committee 2016 contributed equally to *hackseq* and participated in the discussions expressed in this manuscript:

Artem Babaian, *Terry Fox Laboratory, BC Cancer Agency, Vancouver, BC, Canada*

Britt Drögemöller, *Faculty of Pharmaceutical Sciences, University of British Columbia*

Bruno M Grande, *Department of Molecular Biology and Biochemistry, University of British Columbia*

Shaun D Jackman, *BC Cancer Agency Genome Sciences Centre*

Amy Huei-Yi Lee, *Department of Microbiology and Immunology, University of British Columbia*

Santina Lin, *Bioinformatics Training Program, University of British Columbia*

Catrina Loucks, *Department of Molecular Biology and Biochemistry, Simon Fraser University*

Adriana Suarez-Gonzalez, *Department of Botany, University of British Columbia*

Tiffany Timbers, *Masters of Data Science & Department of Statistics, University of British Columbia*

Galen Wright, *Centre for Molecular Medicine and Therapeutics, BC Children's Hospital Research Institute, University of British Columbia*

AB, BD, BG, AL, SL, ASG and GW wrote the first draft of the manuscript. All authors were involved in the revision of the manuscript and agreed to the final version.

Competing interests

No competing interests were disclosed.

Grant information

The author(s) declared that no grants were involved in supporting this work.

Acknowledgements

We would first and foremost thank the *hackseq* participants without which this event would not have happened. By team we'd like to thank, Jean-Christophe Berube, Ogan Mancarci, Erin Marshall, Edward Mason, Celia Siu, Ben Weisburd, Shing Hei Zhan, **Grace X.Y. Zheng**; Madeline Couse, Bruno Grande, Eric Karlins, Tanya Phung, Phillip Richmond, Timothy H. Webster, Whitney Whitford, **Melissa A. Wilson Sayres**; Craig Glastonbury, Daisie Huang, Hamid Younesy, Jasleen Grewal, Laura Gutierrez Funderburk, Lisa Bang, **Shaun Jackman**, Veera Manikandan Rajagopal, Y. Brian Lee; Carolyn Ch'ng, David Brazel, Karthigayini Sivaprakasam, Jill Moore, Shobhana Sekar, Stephen Kan, Jing Yun Alice Zhu, Ka-Kyung Kim, **Luca Pinello**; Fotis Tsetos, Kieran O'Neill, Shreejoy Tripathy, **Manuel Belmadani**; Ayton Meintjes, Scott Hazelhurst, Vincent Montoya, Marcia MacDonald, Jocelyn Lee, Dan Fornika, Brian Lee, Austin Reynolds, **Tommy Carstensen**; Amanjeet Sethi, Eric Zhao, Hua Ling, **Patrick Marks**, Peng Zhang, Samantha Kohli; Erik Gafni, Dan Kvitek, Jake Lever and **Michael Schnall-Levin**; **Ben Busby**, Justin Chu, Jessica Hardwicke, Sean La and Feng Xu.

We would like to thank our sponsorship partners 10X Genomics, ECOSCOPE, Amazon AWS, American Society of Human Genetics, Vancouver Tourism, Genome British Columbia, Association for Computing Machinery – Women, Affymetrix, bioinformatics.ca, and GitHub. We also partnered with local organizations for logistical support: Society for Canadian Women in Science and Technology, BC Cancer Agency Graduate Student and Post-Doctoral Fellow Society, National Center for Biotechnology Information and the Vancouver Bioinformatics User Group. Sponsorship partners had no role in data collection and analysis or preparation of the manuscript. Thanks to the reviewers for their time and suggestions.

References

1. Stephens ZD, Lee SY, Faghri F, *et al.*: **Big Data: Astronomical or Genomical?** *PLoS Biol.* Public Library of Science; 2015; **13**(7): e1002195.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
2. Prins P, de Ligt J, Tarasov A, *et al.*: **Toward effective software solutions for big biology.** *Nat Biotechnol.* 2015; **33**(7): 686–687.
[PubMed Abstract](#) | [Publisher Full Text](#)
3. Busby B, Lesko M; August 2015 and January 2016 Hackathon participants, *et al.*: **Closing gaps between open software and public data in a hackathon setting: User-centered software prototyping [version 2; referees: not peer reviewed].** *F1000Res.* 2016; **5**: 672.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
4. Richard GT, Kafai YB, Adleberg B, *et al.*: **StitchFest: Diversifying a College Hackathon to Broaden Participation and Perceptions in Computing.** *Proceedings of the 46th ACM Technical Symposium on Computer Science Education - SIGCSE '15.* New York, USA: ACM Press; 2015; 114–119.
[Publisher Full Text](#)
5. Jackman S, Birol I: **Linuxbrew and Homebrew for cross-platform package management [version 1; not peer reviewed].** *F1000Res.* 2016; **5**(ISCB Comm J): 1795 (poster).
[Publisher Full Text](#)
6. *hackseq* Organising Committee (2016): **Dataset 1 in: *hackseq*: Catalyzing collaboration between biological and computational scientists via hackathon.** *F1000Research.* 2017.
[Data Source](#)
7. *hackseq* Organising Committee (2016): **Dataset 2 in: *hackseq*: Catalyzing collaboration between biological and computational scientists via hackathon.** *F1000Research.* 2017.
[Data Source](#)

Open Peer Review

Current Peer Review Status:

Version 2

Reviewer Report 08 May 2017

<https://doi.org/10.5256/f1000research.12239.r21700>

© 2017 Hertweck K. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Kate L. Hertweck

Department of Biology, The University of Texas at Tyler, Tyler, TX, USA

My comments have been accommodated.

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 13 April 2017

<https://doi.org/10.5256/f1000research.12239.r21699>

© 2017 Guo J. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Jiarong Guo

Center for Microbial Ecology, Michigan State University, East Lansing, MI, USA

My previous comments have been addressed by authors in the new version and also in comment section. I have no further comments to make.

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 31 March 2017

<https://doi.org/10.5256/f1000research.11818.r20964>

© 2017 Guo J. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Jiarong Guo

Center for Microbial Ecology, Michigan State University, East Lansing, MI, USA

The authors reported a detailed summary of their genomic hackathon, hackseq, to help those interested in organizing similar hackathons in future. The hackseq brought together 66 biological and computational scientists with diverse demographic background to collaborate on nine projects on genomics ranging from data visualization to algorithm development. All the participants had positive responses in post assessment and showed interests in future hackathons.

The background is clearly articulated. There are detailed descriptions on hackseq format and technical and logistical requisites, which are useful for future hackathons. Brief research project summaries are also described with more information available on GitHub. The data for reproducing the figures are made available on F1000 and schedules and application forms are available on GitHub.

Major comments:

Strengths:

Overall, sharing details and experiences of the hackseq such as recruiting project leaders and participants, assigning teams, logistical and technical requisites, and post assessment is valuable for the open science community to organize future hackathons.

1. It is a great idea to organize the hackathon as a satellite event of bigger events. The bigger events can help on travel cost of participants and more importantly promote hackathon participations.
2. This hackseq is very successful at recruiting diverse participants, because it has a representative committee that encourages female, minority and early career scientist participations and also good promotion strategies that it partners with organizations such as Society for Canadian Women in Science.

Weakness:

The team leaders seem to have a critical role in each project, but their roles and responsibilities during the hackathon are not clearly mentioned in the manuscripts.

Minor comments:

Last paragraph in "Hackseq format" section: it states that participants ranked top three projects on which they would like to work, but there are four projects of choice in the application form.

First paragraph in "Core logistics" section: stationary -> stationery.

Second paragraph in "Discussion": female percentage in biology is significantly higher than in engineering. Thus the direct comparison of female participation rate of hackseq to engineering (NASA and Spotify) hackathons is not meaningful.

Discussion: assessment is difficult with participants from diverse background. Some discussion on the current assessment and possible improvement would be useful.

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reader Comment 31 Mar 2017

Shaun Jackman, BC Cancer Agency Genome Sciences Centre, Canada

Thank you for your review, Jiarong.

> The team leaders seem to have a critical role in each project, but their roles and responsibilities during the hackathon are not clearly mentioned in the manuscripts.

I helped organize Hackseq and was also a project leader. Key responsibilities of the team leaders are:

Prior to the event

1. Describing the proposed project.
2. Describing to the organizers:
 1. desired number of participants and skill set
 2. required compute resources
 3. required software
 4. required data
 5. logistical requirements
3. Discussing the project with interested participants.
4. Discussing the suitability of the participants assigned to the project with the organizers.
5. Confirming that the required software is installed and works.
6. Downloading the required data.
7. Planning the scope and strategy of the project.
8. Dividing the project into separable components.

During the event

1. Introducing the participants to each other.
2. Introducing the project and necessary background information to the participants.
3. Describing the components of the project to the participants.
4. Assigning those components to participants based on their interest.
5. Periodically discussing progress with the participants.
6. Troubleshooting technical issues with the help of organizers when needed.
7. Organizing the final report and presentation.

Competing Interests: No competing interests were declared.

Reviewer Report 31 March 2017

<https://doi.org/10.5256/f1000research.11818.r20626>

© 2017 Hertweck K. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Kate L. Hertweck

Department of Biology, The University of Texas at Tyler, Tyler, TX, USA

Thank you to the authors for writing a summary of what seems to be a very successful collaborative coding event, with this manuscript in particular focused on preparation for the event, managing logistic concerns during the event, and an overview of the projects supported. The manuscript is quite well written, and I have no concerns about the content presented therein.

I especially appreciate the recommendations for how to solicit diverse leaders/participants, engage with partner organizations, and carefully craft a sense of community among attendees. Moreover, the authors include suggestions on how to improve similar events in the future. The data reported here provide an important context for comparison for events which continue to encourage participation from underrepresented groups.

Although not highlighted in the paper, the itinerary for the three-day meeting described here includes a number of additional details which would be useful to other coding event organizers. For example, while the majority of meeting time was dedicated to team work, extra workshops and talks were offered (e.g., introduction to git) that would be encourage skills development for students or other participants new to the field. I'll be interested to see whether this model for hosting a hackathon continues at ASHG or other meetings.

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research