BioNT - Network for Training

Deliverable 1.1 | Bioinformatics Introduction

Version 1.0

28 September 2023

Grant Agreement number	101100604 - DIGITAL-2022-TRAINING-02
Action Acronym	BioNT
Action Title	Bio Network for Training
Deliverable number and title	1.1 Bioinformatics Introduction
Work package number and title	1 Training design and development
Dissemination level	Public
Authors	B. Batut, S. Di Giorgio, T. Müller, L. Paladin, I. Paredes Cisneros, S. Razick, P. Samarakoon, approved by all BioNT partners
Delivery date	2023-09-28

Project partners

1. EUROPEAN MOLECULAR BIOLOGY LABORATORY (EMBL)

2. BIOBYTE SOLUTIONS GMBH (BIOBYTE)

3. HPC NOW CONSULTING SL (HPCNOW)

4. UNIVERSITETET I OSLO (UO)

5. UNIVERSITAT DE BARCELONA (UB)

6. INFORMATION CENTRE FOR LIFE SCIENCE (ZBMED)

7. SIMULA CONSULTING (SIMULA)

8. ALBERT-LUDWIGS-UNIVERSITAET FREIBURG (ALU-FR)

9. ECOLE POLYTECHNIQUE FEDERALE DE LAUSANNE (EPFL)

Table of Contents

Project overview

The BioNT consortium is dedicated to providing a comprehensive training program and fostering a community for digital skills relevant to the biotechnology industry and biomedical sector. With a curriculum tailored for both beginners and advanced professionals, BioNT aims to equip individuals with the necessary expertise in handling, processing, and visualising biological data, as well as utilising computational biology tools. Leveraging the consortium's strong background in digital literacy training and extensive network of collaborations, BioNT is poised to professionalise life sciences data management, processing, and analysis skills.

First training workshop report in summary

BioNT delivered its first training workshop in September, 2023. This document reports about this workshop regarding its organisation, the applicants and participants, as well as their feedback about it. The report also touches on the different advertising channels used to reach the project's target audience and on the methods developed exclusively to create and deliver BioNT's training workshops.

Bioinformatics Introduction workshop

BioNT delivered its first training workshop, a Bioinformatics Introduction workshop from 4th to 8th of September, 2023. It was fully online and free of costs.

Within this practical introduction to bioinformatics, key concepts of sequencing data analysis were explained and a broader overview on bioinformatics data analysis was given. The focus was on RNA-sequencing. Indeed, RNA-sequencing revolutionised genomics and health care research, by being a high-throughput and comparatively low-cost method. Together with the development of computational tools to enable their analysis, it has fundamentally improved our understanding of RNA biology.

The Galaxy platform was selected as the training infrastructure to support this workshop. Galaxy is a worldwide open source project. The [European Galaxy server](#) is the biggest instance in Europe with more than 70,000 users. It offers free access to databases and more than 3,200 bioinformatics [tools](#) which can be used through a graphical user interface.

Within this workshop, the 30+ participants learnt how to use Galaxy and how to perform a full RNA-seq data analysis. In bioinformatics, RNA-seq data is one of the most commonly performed data analyses. However, there are many different kinds of data and potential ways to analyse them. Therefore, the last workshop day provided an overview of the wide range of bioinformatics databases and data formats.

Organisation of the workshop

The workshop "A practical introduction to bioinformatics and RNA-seq using Galaxy" ran for 5 days, from 4th to 8th of September 8, 2023. It was fully online and free of costs. Each day, a 4-hour session was delivered from 9:00 to 13:00 CEST.

Since the workshop was the first in the BioNT basic curriculum, many organisational and best practice steps needed to be defined and implemented. For an efficient execution of the next workshops, a concise standard operating procedure (SOP) was developed, describing every step needed to successfully develop and deliver a BioNT training workshop.

Webpage and registrations

The CECAM event management platform, provided by the EPFL, was used to create a dedicated webpage (accessible at https://www.cecam.org/workshop-details/1260) for the workshop, which included the workshop description, learning objectives, requirements, program (Figure 1), and any further information relevant to potential participants.

Monday September 4th 2023 - Day 1

Galaxy introduction

- 09:00 to 09:20 - Introduction + Ice breaker
- 09:20 to 09:50 - Galaxy introduction
- 09:50 to 10:00 - Break
- 10:00 to 11:15 - Tutorial: From peaks to genes (I)
- 11:15 to 11:30 - Break
- 11:30 to 12:50 - Tutorial: From peaks to gene (II)
- 12:50 to 13:00 - Summary

Tuesday September 5th 2023 - Day 2

Mapping and quality control

- 09:00 to 09:15 - Welcome + recap
- 09:15 to 09:35 - Introduction to quality control
- 09:35 to 11:00 - Tutorial: Quality control (I)
- 11:00 to 11:10 - Break
- 11:10 to 11:30 - Introduction to mapping
- 11:30 to 12:30 - Tutorial: Mapping
- 12:30 to 12:40 - Break
- 12:40 to 12:50 - Genome browser demonstration
- 12:50 to 13:00 - Summary

Wednesday September 6th 2023 - Day 3

RNA-seq data analysis (first half)

- 09:00 to 09:10 - Introduction + recap
- 09:10 to 09:40 - Transcriptomic introduction
- 09:40 to 09:50 - Break
- 09:50 to 11:20 - Tutorial: Reference-based RNA-seq - From reads to counts (I)
- 11:20 to 11:30 - Break
- 11:30 to 12:50 - Tutorial: Reference-based RNA-seq - From reads to counts (II)
- 12:50 to 13:00 - Summary

Thursday September 7th 2023 - Day 4

RNA-seq data analysis (second half)

- 09:00 to 09:10 - Introduction + recap
- 09:10 to 10:10 - Tutorial: Reference-based RNA-seq - From counts to Differentially Expressed Genes and interpretation (I)
- 10:10 to 10:20 - Break
- 10:20 to 11:30 - Tutorial: Reference-based RNA-seq - From counts to Differentially Expressed Genes and interpretation (II)
- 11:30 to 11:40 - Break
- 11:40 to 12:40 - Tutorial: Reference-based RNA-seq - From counts to Differentially Expressed Genes and interpretation (III)
- 12:40 to 13:00 - Summary

Friday September 8th 2023 - Day 5

Data types and databases

- 09:00 to 09:10 - Introduction + ice breaker
- 09:10 to 10:10 - Bioinformatics Data Types and Databases
- 10:10 to 10:20 - Break
- 10:20 to 11:30 - Tutorial: Learning about one gene across biological resources and formats
- 11:30 to 11:40 - Break
- 11:40 to 12:40 - Tutorial: One protein along the UniProt page
- 12:40 to 13:00 - Summary and closing remarks

Figure 1 - Workshop program as displayed on the event page on CECAM.

For registration, the CECAM platform was used to manage the applicant's information and communication. In parallel, the EMBL servers were used to collect pre- and post-workshop information through pseudo-anonymised surveys. The survey data was linked to the applicant's data only via a unique identifier, provided in the CECAM registration process, as well as in the EMBL-based survey. This ensures that only the workshop organisers access the applicants' personal data while still collecting information relevant to the workshop separately. To register, applicants had to: (i) register on the CECAM platform, (ii) complete and submit the pre-workshop survey, and finally (iii) complete the application on the CECAM platform using the unique identifier generated at the end of the pre-workshop survey.

Applications were reviewed based on answers in the pre-workshop survey (containing no personal information). Applicants working in small and medium enterprises (SMEs) or that identified themselves as job seekers would have been prioritised, but after a thorough assessment of technical and personnel capacity, all 60 applicants where accepted. The communication of the application outcome to all participants, as well as any additional communication, was performed via the CECAM platform.

Advertisement campaign

The workshop was advertised via social media (Figure 2 A-B), several websites (Figure X C), mailing lists or Slack spaces of networks and communities (ELIXIR, Research Data Alliance (RDA), Bioconductor, de.NBI, NFDI4Microbiota, GOBLET, LifeSciTrainer, OLS, BioRN, CASTIEL2, EuroHPC, Oslo Bioinformatics).
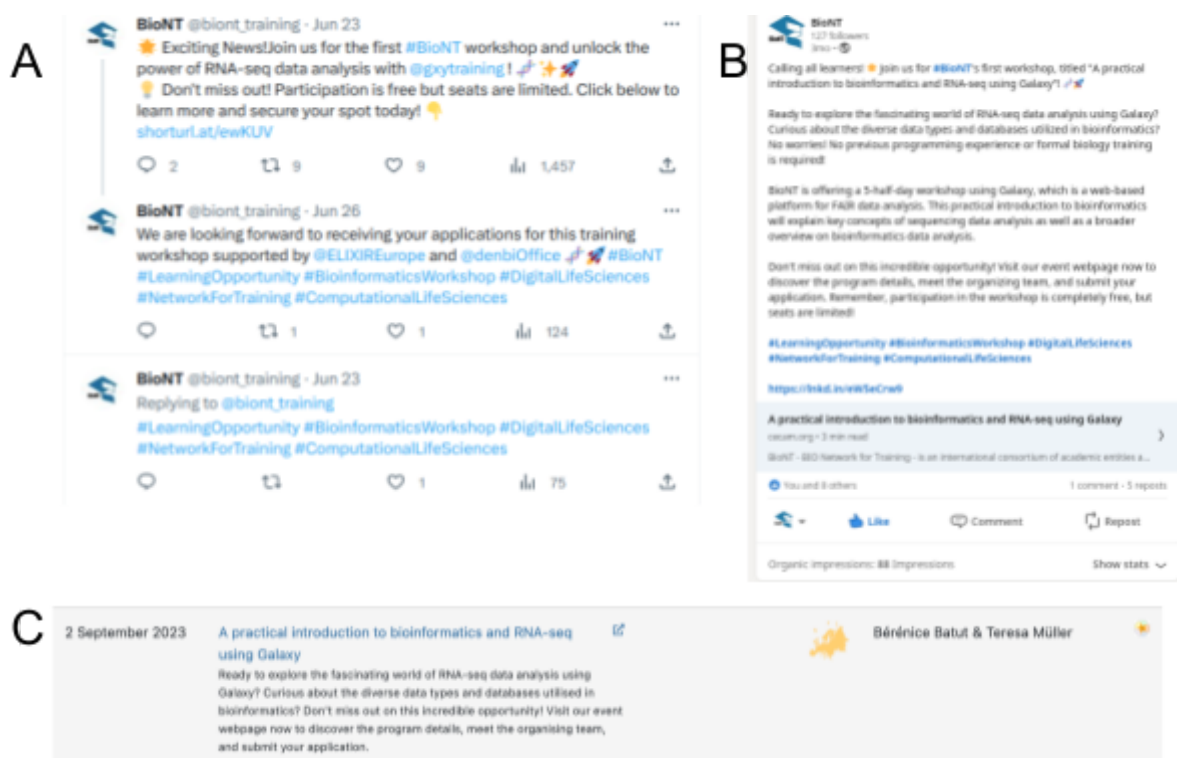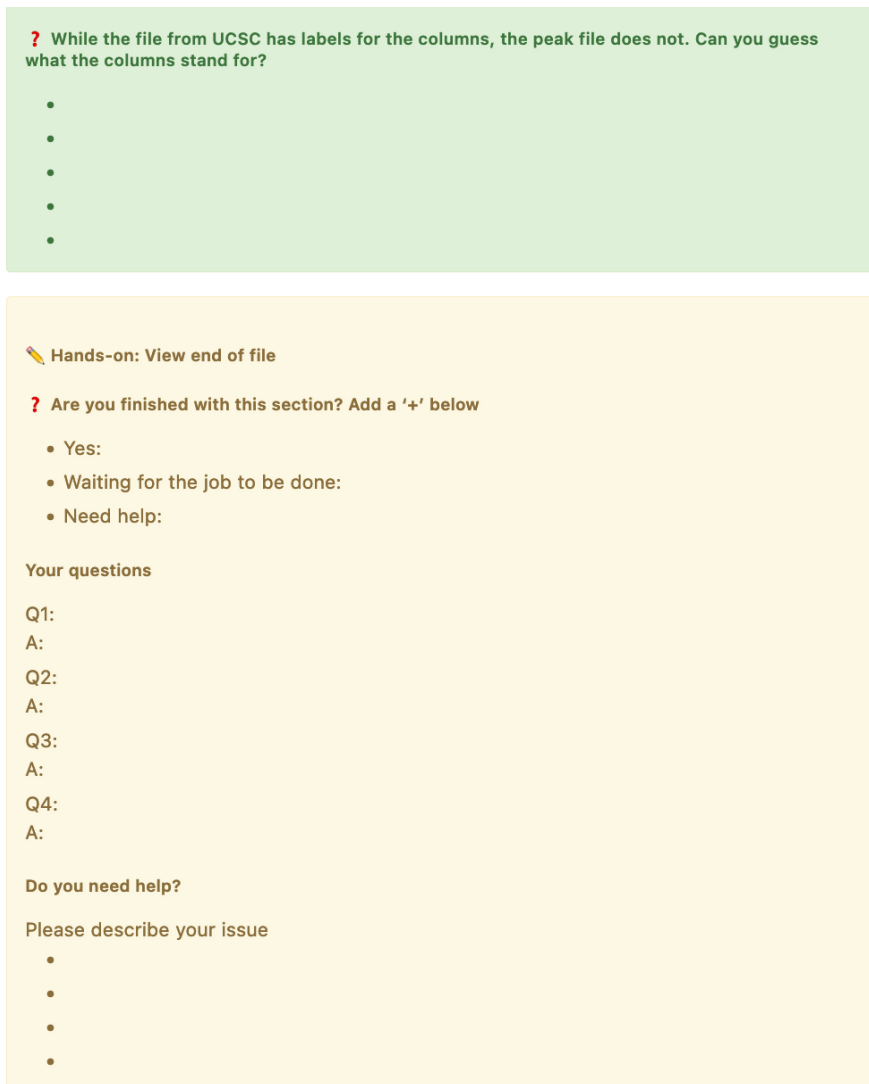

Figure 2 - Screenshots of the examples of advertisement through social media (**A** - Twitter, **B** - LinkedIn) and websites (**C** - Galaxy Hub at https://galaxyproject.org/events/).

Infrastructure for the workshop

The workshop was delivered in a Zoom webinar format, with participants' visibility disabled to preserve their privacy. Participants were able to see and learn from the trainers. However, direct interactions (e.g. chat or voice) were not possible within Zoom during the sessions. All direct interaction took place in written form via a collaborative document, where anonymous participation was ensured. To serve the collaborative documents, a HedgeDoc collaborative space was set up by BIOBYTE, and is hosted on their server.

A **Main** collaborative document, setup by the instructors and organisers, was shared with the participants before the session. Each section of the workshop had dedicated *Hands-on* boxes to report on the task status, ask questions or raise issues. Helpers engaged and helped participants by answering the questions and issues directly in this document. This Main document was filled in live during the workshop. Separate boxes to answer questions were used to improve the participant engagement and as an indirect learning assessment (Figure 3).

Figure 3 - Example of question and hands-on boxes.

To help with the organisation, four HedgeDoc documents were used: (i) a **Template** with all instructions and boxes for hands-on, questions, etc; (ii) the **Main** document with the information for the participants during the workshop, filled with boxes related to the section covered by the instructor to help with the navigation and cleaned during each break to avoid an overcrowded document; (iii) a **History** document with everything that needed to be kept and shared with participants after the workshop. In addition, (iv) a document for Helpers & Instructors was created providing the workshop setup, interactions and explaining tasks of the helpers and instructors.

Galaxy, an open source platform for accessible, reproducible, and transparent computational research, via the European Galaxy server, was used during the workshop because it enables participants to access 3,000+ tools, high-performance compute resources, through a standard web browser, without requiring informatics expertise. To support this workshop, a dedicated job queue and resources was provided via the Training Infrastructure as a Service (TIaaS - https://usegalaxy.eu/tiaas/). This service also provided instructors with a dashboard (Figure 4), enabling them to monitor the progress of the participants.
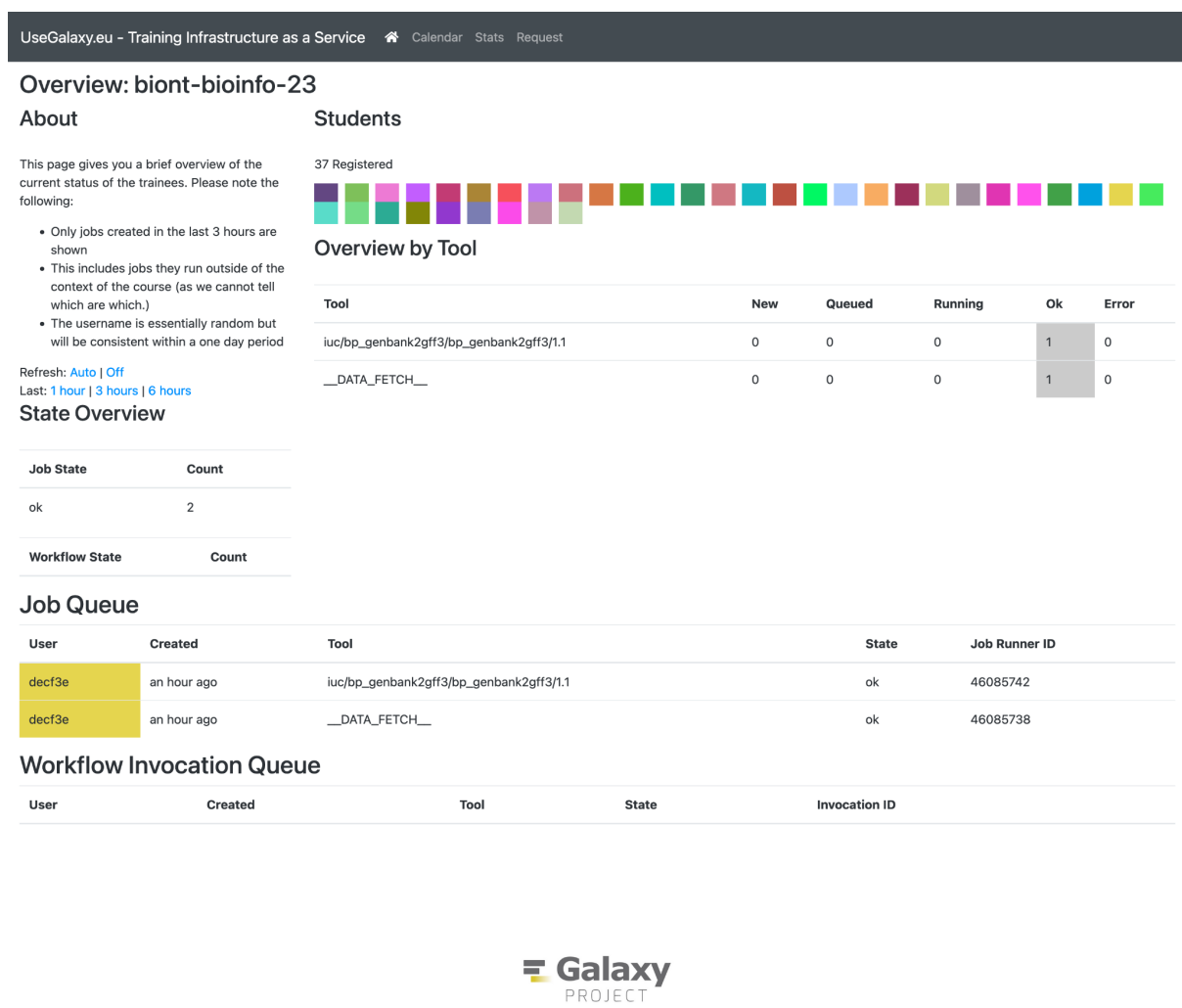

Figure 4 - Screenshot of the dashboard for TiaaS provided for instructors.

The training material used during this workshop (Table 1), i.e. slides and tutorials, are all stored and publicly available on the Galaxy Training Network website (https://training.galaxyproject.org). For the first four days, existing tutorials and slides were used. For the last day, dedicated tutorials and slides were developed specifically for this workshop and contributed to the Galaxy Training Network.

Day	Topic	Material
1	Galaxy Introduction	Slides + Tutorial
2	Quality Control	Slides + Tutorial
	Mapping	Slides + Tutorial
3	Introduction to Transcriptomics	Slides
4	Reference-based RNA-seq	Tutorial
5	Learning about one gene across biological resources and formats	Slides + Tutorial
	One protein along the UniProt page	Tutorial

Table 1 - Program and training material per day.

Certificates

Certificates (Figure 5) were provided to those participants who explicitly required them after checking for their participation using two criteria: (i) they joined at least one session on Zoom or notified us that they could not attend live and (ii) they provided links to their Galaxy histories corresponding to the taught tutorials.
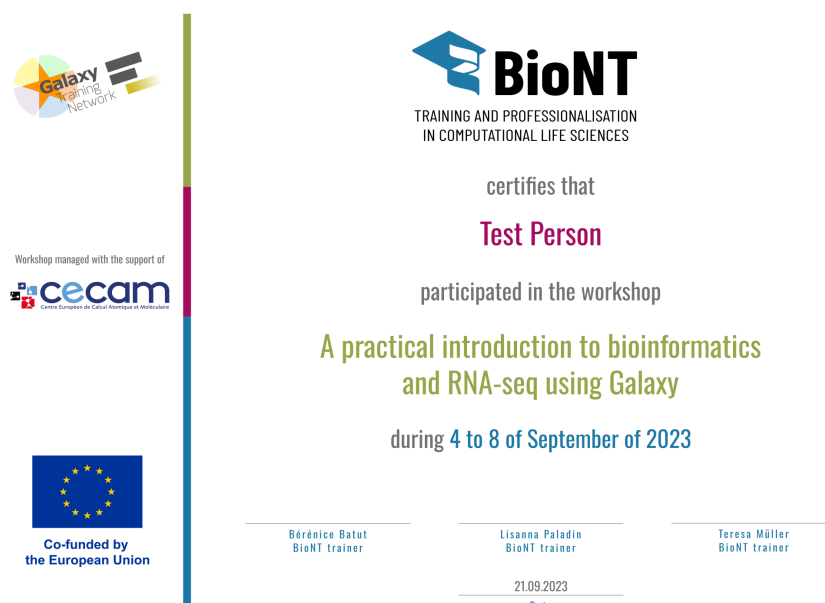

Figure 5 - Template used to generate certificates.

Outcomes of the workshop

Applications and pre-workshop survey

73 applicants completed the pre-workshop survey and 60 submitted their application form via the CECAM platform. The pre-workshop survey consisted of 13 questions covering skills, demographics, and miscellaneous topics. The majority of questions were not mandatory for the successful completion and submission of the survey.

General information

The 73 applicants, almost at a balanced gender representation with 43.84% female and 46.58% male, were from and were working in 35 countries (Table 2): 16 countries in Europe, 10 in Asia, 5 in Africa, 3 in South America, and 1 in North America. 38.36% described themselves as White, 27.4% as Asian, 9.59% as Black or African American, 8.22% Hispanic or Latino(a).

Country	Nationality	Employment
Argentina	3	2
Austria	2	1
Bulgaria	2	2
Cameroon	1	1
China	1	
Colombia	2	
Czechia	1	1
France		1
Germany	5	15
Greece	2	1
India	11	11
Ireland		1
Israel	1	1
Italy	2	2
Kenya	2	2
Malaysia	1	1
Moldova	1	1

Montenegro	1	1
Nigeria	4	3
Norway	1	3
Pakistan	1	
Philippines	1	1
Poland	2	2
Portugal	1	1
Romania	2	2
Sint Maarten		1
Spain	3	3
Sri Lanka	1	1
Sweden	1	2
Tunisia	4	3
Turkey	9	3
United States of America		3
Vietnam	1	
Zimbabwe	1	1

Table 2 - Workshop applicants' nationality and country of employment from the pre-workshop survey.

Most applicants were from the Genetics, Genomics, or Bioinformatics field (Figure 5 - B). Most applicants (63%) were academic employees (Figure 5 - A) as graduate students (31%) or research staff (27%).

Figure 5 - Definition (**A**) and Fields (**B**) of applicants from in the pre-workshop survey.

Related to the original target audience of this workshop, 6 applicants were industry employees and 21 were job seekers (with 12 in Academia and 3 in Industry). Regarding the connections with SMEs, 13 mentioned working in an SME, 12 collaborating with SME(s) and 22 aiming at working in an SME.

Applicants found out about the workshop in different ways (Figure 6), mostly through social media.
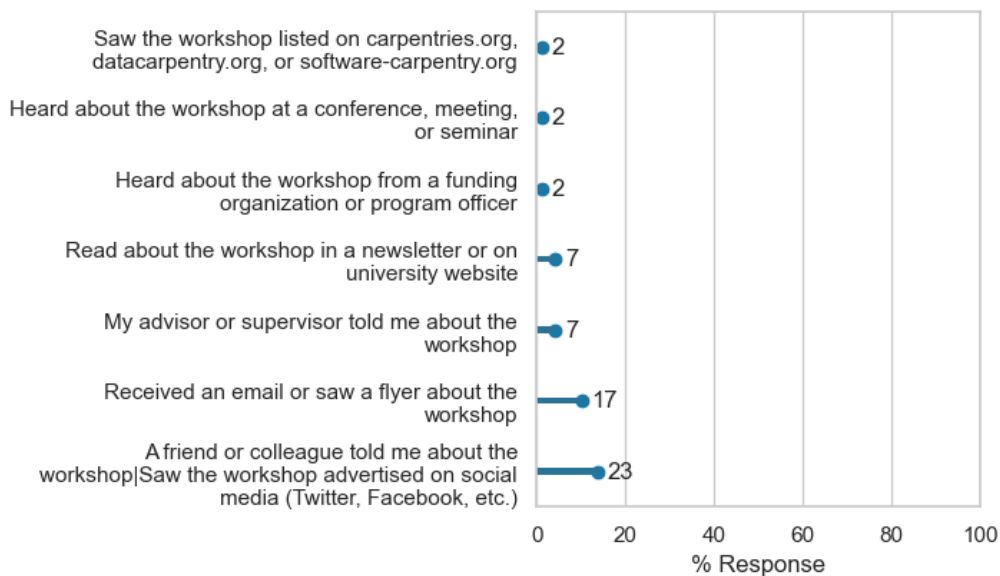
Figure 6 - Answers in the pre-workshop survey to the question: How did you find out about this workshop?"

63% of the applicants planned to use Windows for the workshop, 20% macOS and 12% Linux. Finally, two applicants had visual accessibility requirements. These two applicants were contacted in advance to enquire about their accessibility requirements. The organisers afterwards provided the full learning materials and an offer for further assistance, if needed.

Background information

The aim of this workshop was to give an introduction to bioinformatics and RNA-seq using Galaxy. When asked about their knowledge and use of Galaxy, the answers are quite diverse, from "never" to "daily" (Figure 7).
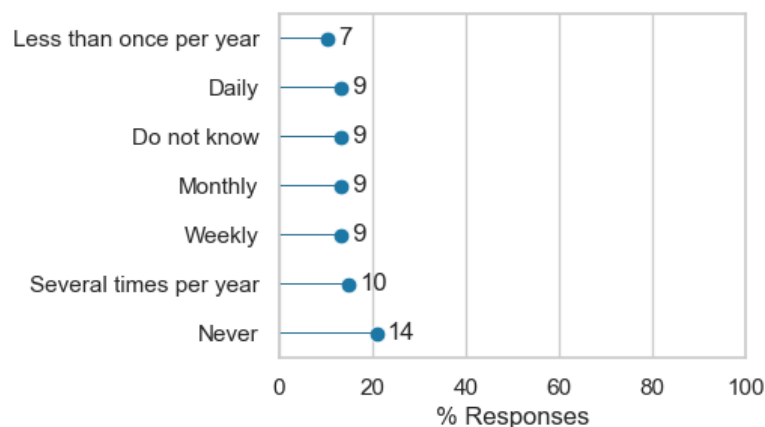

Figure 7 - Answers in the pre-workshop survey to the question: You know what Galaxy is and you use it?

There was an interest in learning how to analyse sequencing data, given the frequency with which they analyse data: 68% analyse sequencing data at least several times per year

(Figure 8 - A), but the vast majority never builds bioinformatics workflows (Figure 8 - B). They use programming languages and/or command line tools for that (Figure 8 - C-D).
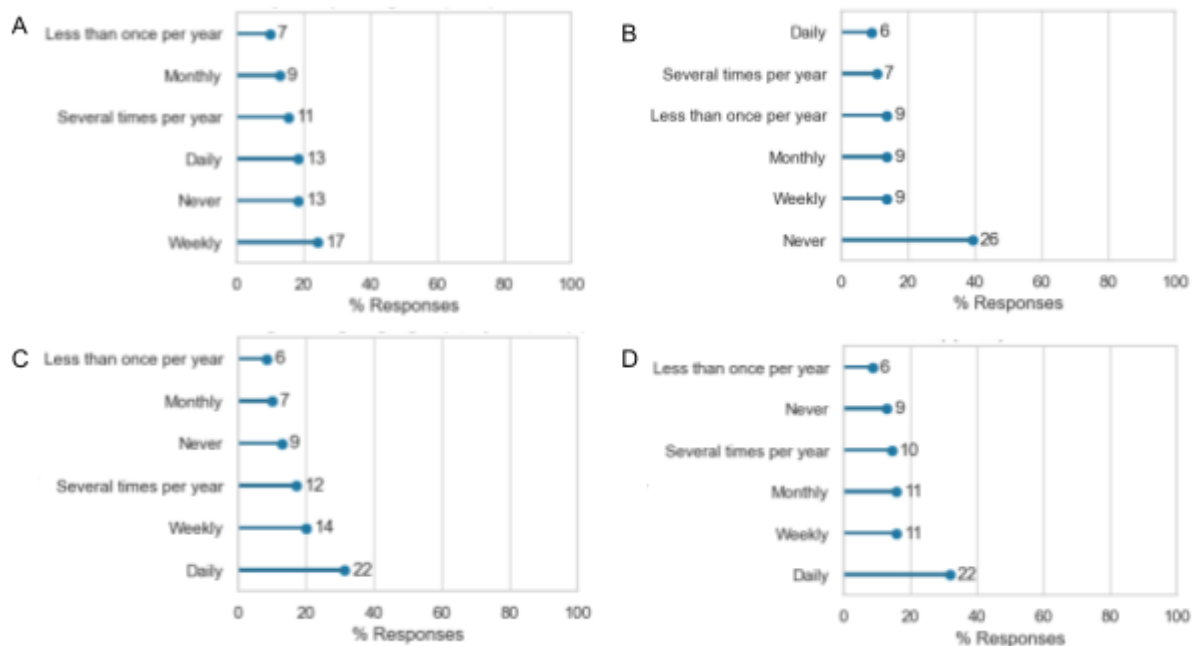

Figure 8 - Answers in the pre-workshop survey to questions: **A** - How often do you currently analyse sequencing data?; **B** - How often do you currently build bioinformatics workflows?; **C** - How often do you currently use Programming languages (R, Python, etc.)?; **D** - How often do you currently use a command shell (usually accessed through Terminal on macOS or PowerShell on Windows)?

Most applicants (85%) agree with the importance of having access to the original, raw data to be able to repeat an analysis (data not shown). The majority agree that using a programming language can make their analyses easier to reproduce, but 37% of applicants were satisfied with their current data management and analysis workflow (Figure 9).
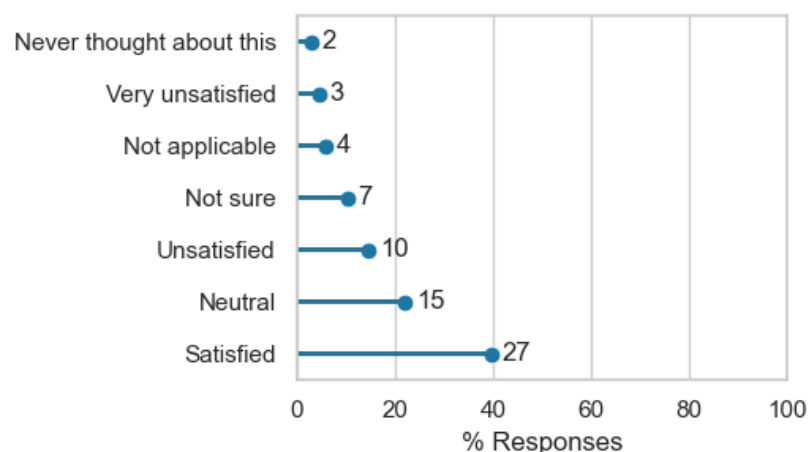

Figure 9 - Answers in the pre-workshop survey to the question: Please rate your level of satisfaction with your current data management and analysis workflow (i.e. how you collect, organise, store and analyse your data).

Expectations

The expectations of the applicants regarding the workshop were quite diverse. They were mostly interested in learning (Figure 10): How to perform basic analyses, in particular RNA-seq; how to build pipelines and workflows; how to use Galaxy and cloud based bioinformatics infrastructure without complex coding; new tools; how to process (massive) biological data.
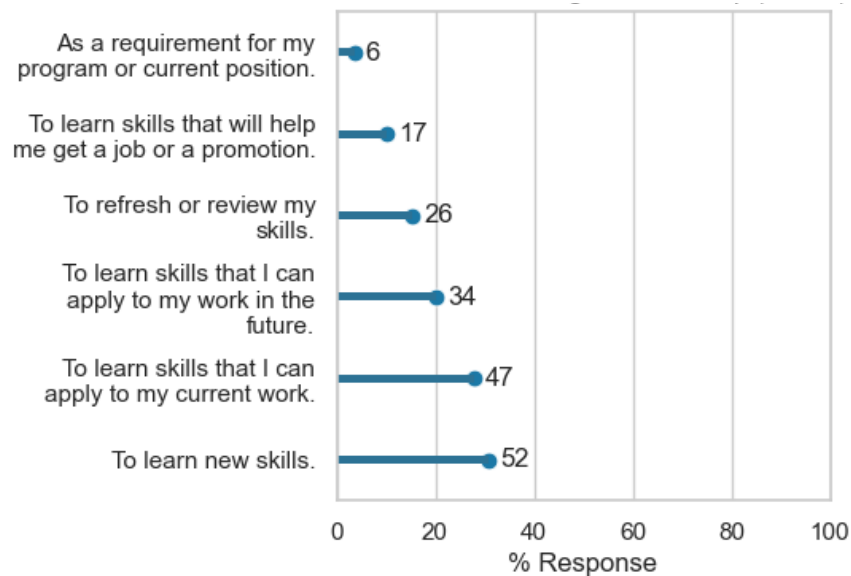

Figure 10 - Answers in the pre-workshop survey to the question: Why are you participating in this workshop?

Participation

All applicants that submitted their complete application were selected to participate in the workshop, but not all of them joined for the complete workshop (Table 3).

Day	Participants	Instructors	Helpers
1	38	1	4
2	35	1	3
3	32	1	3
4	33	1	3
5	32	1	3

Table 3 - Number of participants, instructors and helpers per day.

After the workshop, 18 participants requested a certificate. The certificate was provided to those 16 participants that completed the post-workshop survey and submitted their Galaxy history.

Daily feedback

At the end of each day, participants were asked for feedback on the following three points:

- Please share one thing that was good about today
- Please share one thing that that could be improved about today
- Do you have any other comments?

The daily feedback is summarised in the following Table 4.

Day	Topic	Good about today	To improve	Any other comments?
1	Galaxy Introduction	Nice tutorial, easy to follow and perform steps on Galaxy Nice speed to follow all the steps Clear explanations of performed steps Real-time interactions :-) Good idea to use HedgeDoc	Some steps require more time than for the instructor Sometimes, it is not clear why a specific tool or step is needed Maybe some information on how to choose the tool like would there be alternatives or is there only the one we used	
2	Quality Control Mapping	The speed of the class was perfect The repetitions after loading the data or each analysis were helpful We covered a lot of material and the explanations were good Everything was excellent The class was interactive and very understandable The class was very practical. The class was well packed with hands on skills	Reduce the info or increase time because it was a little bit overwhelming with the amount of information Maybe it would it be better to modify the tutorial guides, or rather not to follow them completely, for us to cover all the agenda for the meeting	Thank you and continue with the great workshops! Thank you for this world class workshop. I got new information today. Thank you for this workshop! Let's keep going! You are doing a great job! You are doing a great job! Thank you so much :) Good job
3	Introduction to	Great job! Thank you both!	Share a history that people can keep on	

	Transcriptomics Reference-based RNA-seq (Part I)	Really great how the complex content was explained See the troubles with running the jobs and how to fix mistakes	working in case someone gets stuck because it doesn't run or after a long waiting time failed Fetch the needed datasets from the data libraries (if they are available) before the workshop if possible. Install IGV locally the previous day to have it ready Extend the RNA seq analysis day because of the demanding calculations on Galaxy	
4	Reference-based RNA-seq (Part II)	The speed of the class, it was really easy to follow, and the graph explanation. Clear explanations of steps and selected options, as well as of obtained results New Galaxy tools and protocols Big help with answering also general questions	A little bit longer	Great job, great explanation! Thank you! Excellent work to explain as much as it was possible and repeatedly for us
5	Learning about one gene across biological resources and formats One protein along the UniProt page	Great job! The pace was good The atmosphere - relaxed and excellent for work Really detailed explanations, really interesting and informative Nice informative introduction to the accessibility of gene and protein related information Overview of different data formats and also other biodata platforms (refresher including new features)	Perhaps more illustrative images aside the plain text/data in the initial presentation (from the platforms etc), how they are related	Thanks for the whole workshop, great job. Thanks a lot for all your time and patience! It was really great and helpful! -Thanks!! That was a really informative workshop. Thank you for such an informative workshop. I learnt a lot.

				Thank you all. Thank you for the workshop Thank you for this workshop! Y'all did great! We learnt a lot because of your efforts <3

Table 4 - Summary of the daily feedback collected in the HedgeDoc document.

Postworkshop survey

At the end of the workshop, participants were asked to fill a post-workshop survey consisting of 20 questions. In total, 28 participants completed this survey.

Regarding the workshop environment, the answers were overall positive. Participants felt comfortable learning in this workshop environment and interacting with the instructors (Figure 11 A-B). They felt that the instructors were enthusiastic about the workshop and knowledgeable about the material being taught (Figure 11 C-D). Participants could get clear answers to their questions from the instructors (Figure 11 E) and felt they could immediately apply what they learned at the workshop (Figure 11 F).

Only one accessibility issue was reported: Unstable internet connection on Wednesday that affected their ability to participate in this workshop.

Participants also shared examples of how an instructor or helper affected their learning experience: One participant liked that the instructors motivated participants to self-explore the additional available training material, as well as to finish parts of the tutorial by themselves after the course. Participants felt that it gives them a boost of inspiration and motivation to continue learning on their own, specially with the excellently prepared tutorials. Participants appreciated that instructors and helpers answered detailed questions about all the aspects of tutorials, but were also able to answer general questions that helped them improve their work even more. They appreciated the help when experiencing problems by sharing their history with helpers. Finally, it was appreciated that instructors made mistakes on their own and showed afterwards how to solve the problem.
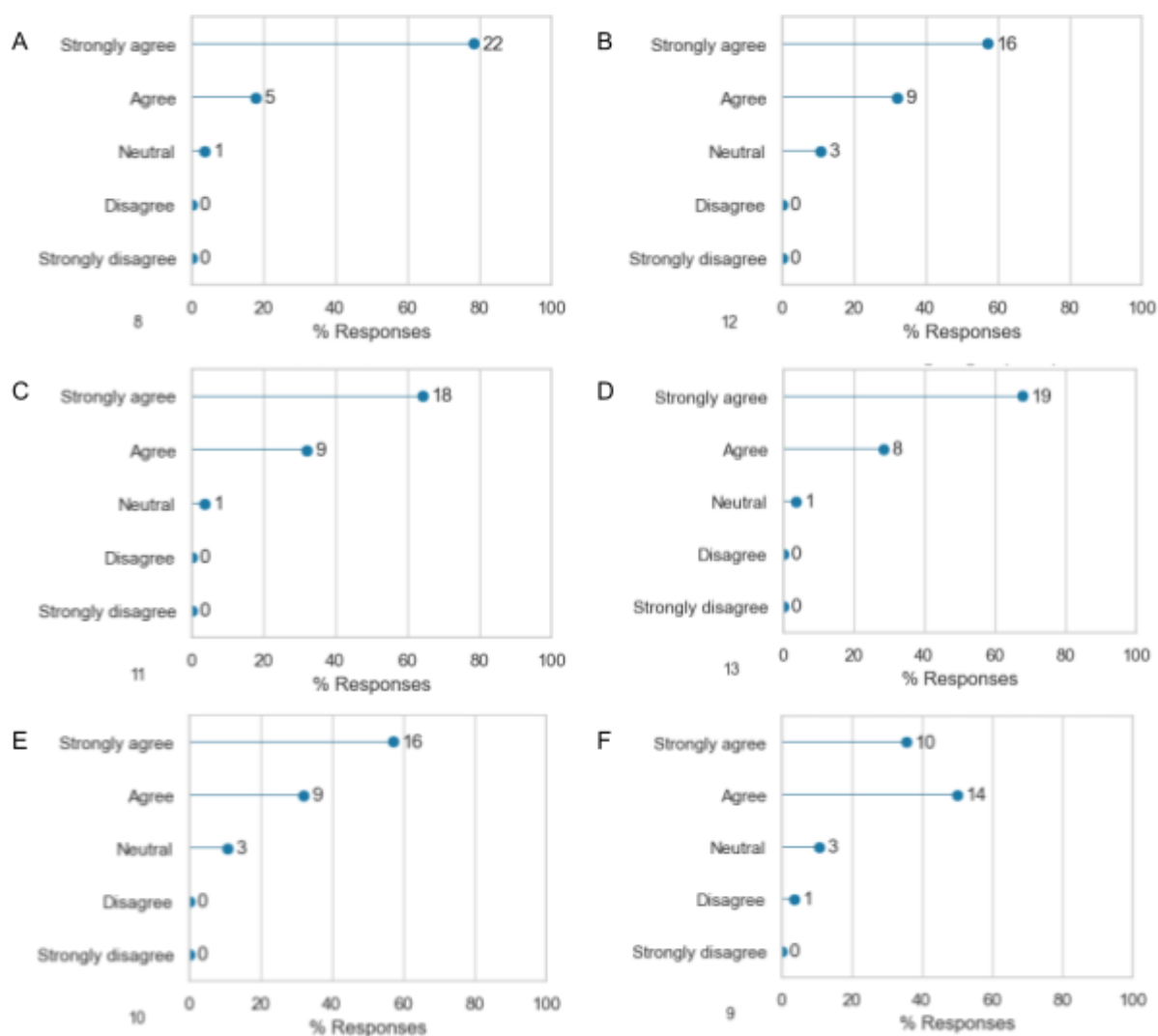
Figure 11 - Rating of participant's agreement with the following statements in the post-workshop survey: **A** - I felt comfortable learning in this workshop environment; **B** - I felt comfortable interacting with the instructors; **C** - The instructors were enthusiastic about the workshop; **D** - The instructors were knowledgeable about the material being taught; **E** - I was able to get clear answers to my questions from the instructors; **F** - I can immediately apply what I learned at this workshop.

Participants were also asked about strengths and ways to improve the workshop. All answers are summarised in the following Table 5.

Major strengths of this workshop	Ways the workshop could be improved
- Speed and way of explanations was great - Easy to follow also if you have not used Galaxy before - Good overview of the important steps of RNA-seq analysis. - Nice interaction via the HedgeDoc document - Links to additional information and/or other tutorials are provided within the hands-on sections - That questions were answered "in the background" by additional helpers	More time for some parts of the tutorials so we would be able to finish it all
Highly interesting and needed topic; hand-ons with step-by-step clear tutorials and highly empathic and patient instructors; highly interactive and fun and relaxed atmosphere	Increase a bit the time for RNA seq hand-ons (wednesday and thursday), info about downloading needed free software day(s) in advance, sharing the instructor's history from Galaxy in case the server is highly loaded with other computations
Step by step explanations of the tutorials, access to the instructors, anonymity that made it even more easier to ask questions	Some of the parts could last a little bit longer so we could cover some more details
Good introduction. Needs second stage	More time, and more exercises with examples
- Relevance - Easy to use tools - Step-by-step guide	I think the last session about formats and databases will be nice to include at the workshop's beginning
- The materials are informative and can be independently used by the learner - The instructors have the expertise about the topics - The instructors are interacting well with the learners despite only using a HedgeDoc - The workshop is well-paced given it runs for a whole week	- The contents of the workshop can be modified to cover all the objectives for the day. Some topics can be skipped while the important or tricky ones can be given emphasis - Certificate is suggested to be provided at the end of the workshop. This will serve as proof that the participants gained the skills/knowledge offered by the workshop. It will also serve as some sort of reward for them. However, if certificates will be given, there must be established conditions on who are qualified to receive them
Nicely explained tutorial and set by step hands-on	More time for RNA sequence analysis part

19

● Good insight in sequencing analysis pipeline ● Overall good explanations	● Maybe explain more in detail why a specific tool is used ● Calculate more time for the workshop
● Hands on engagements ● Knowledgeable facilitators	Make it 10 days
● HedgeDoc is a quite useful platform that enables interaction with instructors while the personal privacy of the participants is still protected ● Hands-on ● Good introduction ● Good practice materials	Since the steps are sequential, if there is an interruption in between i.e. due to technical reasons (in my case my internet connection was not stable for 2 hours during RNA-seq analysis Part-1) it is hard to follow the next steps during the hands-on part. My recommendation is to share the history link with the participants so that everyone can follow you even if there is an interruption. Maybe, it would also be nice to share recordings and complete HedgeDoc file or some other related documents at the end of the corresponding day
Hand-on sessions, good explanation of used tools and why we set some parameters, interaction with instructors	
● Accessible ● Understandable ● Keeping participants anonymous is a good idea	
● Teaching skills evoked ● Display of a variety of different tools ● User-friendly but also technical working digital environment	
It has a lot of practice in the workshop, not only theoretical sessions. It helps me to understand better how the program works.	I think the workshop it's fine and I don't know how to improve more it
Excellent hands-on and detailed answers to questions	Making use of current literatures
● The organisation ● Availability of materials ● Clear Explanations	
Material was well explained but online videos from previous courses were available.	Sometimes the allocated time for one session was too short and the last tasks were just rushed through

Workshop was really user friendly which means that anyone who is not aware about Galaxy can do and learn something from the workshop. Also, the environment was great. Everyone gets their answers for questions	I think time management could be fixed because some days' workshops took much more time than explained before

Table 5 - Strengths and possible improvements for this workshop, as suggested by participants in the post-workshop survey.

The feedback on the post-workshop survey was positive and participants were likely to recommend this workshop to a friend or colleague (Figure 12).
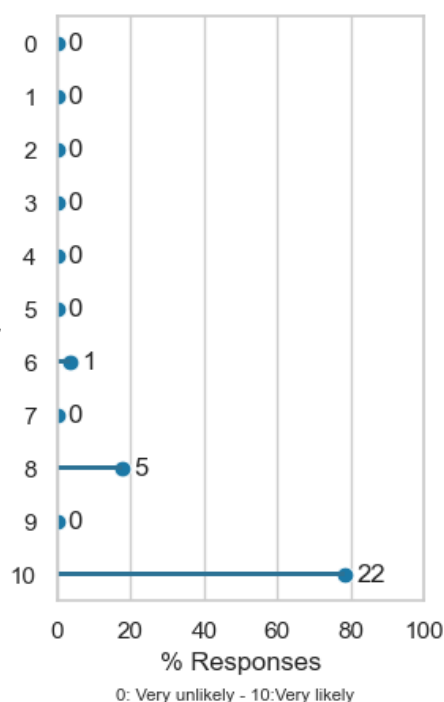

Figure 12 - Answers in the post-workshop survey to the question: How likely would participants recommend this workshop to a friend or colleague?

Conclusion

The first BioNT workshop, a Bioinformatics Introduction workshop, was successfully held from 4th to 8th of September 8, 2023. It was fully online and free of costs.

As it was the first workshop, the infrastructure for communication, registration, collecting surveys, and teaching had to be developed and implemented. Everything has been documented and will be reused for the second workshop that will be run in November 2023.

60 persons applied for this workshop and were accepted, 41 participated and 28 filled the post-workshop survey. The feedback was positive. Some of the proposed improvements, like sharing the history document and the video directly after each workshop day, could be implemented for the second workshop. Overall, the BioNT consortium concludes that the workshop successfully achieved its goals.