

Bioinformatics training needs of Australian researchers: 2021/22 survey

Melissa L. Burke^{1,2,3}, Ann Backhaus⁴, Mariana Barnes⁵, Michael Charleston⁶, Tyrone Chen^{7,8}, Tracy Chew⁹, Jeffrey H. Christiansen^{1,2,3}, Mark Crowe³, Deanna Deveson¹⁰, Victoria Perreau¹¹, Jingbo Wang¹², Nathan Watson-Haigh¹³, Christina R. Hall^{1,11}

1. Australian BioCommons, Australia, 2. Research Computing Centre, The University of Queensland, Queensland, Australia, 3. Queensland Cyber Infrastructure Foundation, Queensland, Australia, 4. Pawsey Supercomputing Research Centre, Western Australia, Australia, 5. Menzies School of Health Research, Darwin, Northern Territory, Australia, 6. University of Tasmania, Tasmania, Australia, 7. COMBINE - Australian Computational Biology and Bioinformatics Student Society, Australia, 8. Monash University, Victoria, Australia, 9. Sydney Informatics Hub, University of Sydney, New South Wales, Australia, 10. Monash Bioinformatics Platform, Biomedicine Discovery Institute, Faculty of Medicine, Nursing and Health Sciences, Monash University, Victoria, Australia, 11. Melbourne Bioinformatics, University of Melbourne, Victoria, Australia, 12. National Computational Infrastructure, Australian National University, Australian Capital Territory, Australia, 13. South Australian Genomics Centre, South Australia, Australia.

On behalf of the National Bioinformatics Training Cooperative.



Table of contents

| | |
|---|----|
| Introduction | 2 |
| Survey design and roll out | 2 |
| Data and analysis | 3 |
| Results | 4 |
| Who responded to the survey? | 4 |
| Current training | 9 |
| Preferred training format | 11 |
| Desired topics | 12 |
| Next steps | 14 |
| Summary | 15 |
| Contributors | 15 |
| Licence | 15 |
| Supplementary files in this Zenodo record | 16 |
| References | 16 |

Introduction

Bioinformatics has become a core part of life science research. Life scientists must continually gain and refresh their bioinformatics skills in order to take advantage of new technologies, move between research topics and progress their careers.

To ensure training remains relevant, life science trainers must keep pace with the rapidly developing field and evolve their training offerings to respond to demand and changes in scientific thinking.

In 2016 EMBL-ABR (a precursor of Australian BioCommons) surveyed life scientists and medical researchers to gauge bioinformatics and computational biology needs in Australia. More than 90% of respondents considered training to be important, with demand for immediate training on all suggested topics: basic computing (Linux) and scripting (Python,R); data management and metadata; integrating multiple data types; and scaling analysis to cloud ¹. Similar results were obtained from surveys in European and US settings ².

Much has changed in the world of bioinformatics since 2016. New technologies and methodologies have been developed (e.g. scRNAseq) and certain tools and platforms have become more accessible and popular among life scientists (e.g Galaxy, RStudio). The pandemic has also changed the way that we work and collaborate. Given these changes, a re-evaluation of training needs and preferred modes of learning was necessary.

Here we summarise the results of the 2021/22 Australian Bioinformatics Training Needs Survey.

Survey design and roll out

The survey was developed by the [National Bioinformatics Training Cooperative](#) convened by Australian BioCommons. It is modelled on a 2020 survey of the Melbourne Biomedical Precinct conducted by the Parkville Bioinformatics Training Group (personal communication Dr Victoria Perreau, Melbourne Bioinformatics). The questions included in our survey are provided as a supplementary file.

The survey was designed to address four key questions:

- What training do people already access?
- Are their training needs being met?
- What format of training is preferred?
- What topics need to be covered?



The survey was administered using Google Forms and was open for responses between December 2021 and February 2022. It was distributed to life science communities via a number of channels including:

- Australian BioCommons and National Bioinformatics Training Cooperative social media accounts
- Australian BioCommons and partner newsletters
- National communities of practice and forums
- Australian BioCommons workshops and webinars

Data and analysis

Anonymised raw data is available as a supplementary file.

Data was collected and visualised using Google Forms and Google Sheets. Free text responses were lightly curated to aid interpretation and identification of themes. Specifically:

- State/territory is inferred from the postcode provided by respondents.
- Similar responses to free text questions for example 'None', 'NONE' and 'none' were consolidated.
- Organisation names were curated for consistent spelling and results are shown at the organisation level (e.g. Diamantina Institute University of Queensland is shown as 'University of Queensland').
- Responses where dual appointments are indicated are counted under both organisations e.g. Garvan, UNSW = Garvan AND UNSW.

Results

Who responded to the survey?

We received 126 responses from individuals in all Australian states and territories except the Northern Territory (Figure 1).

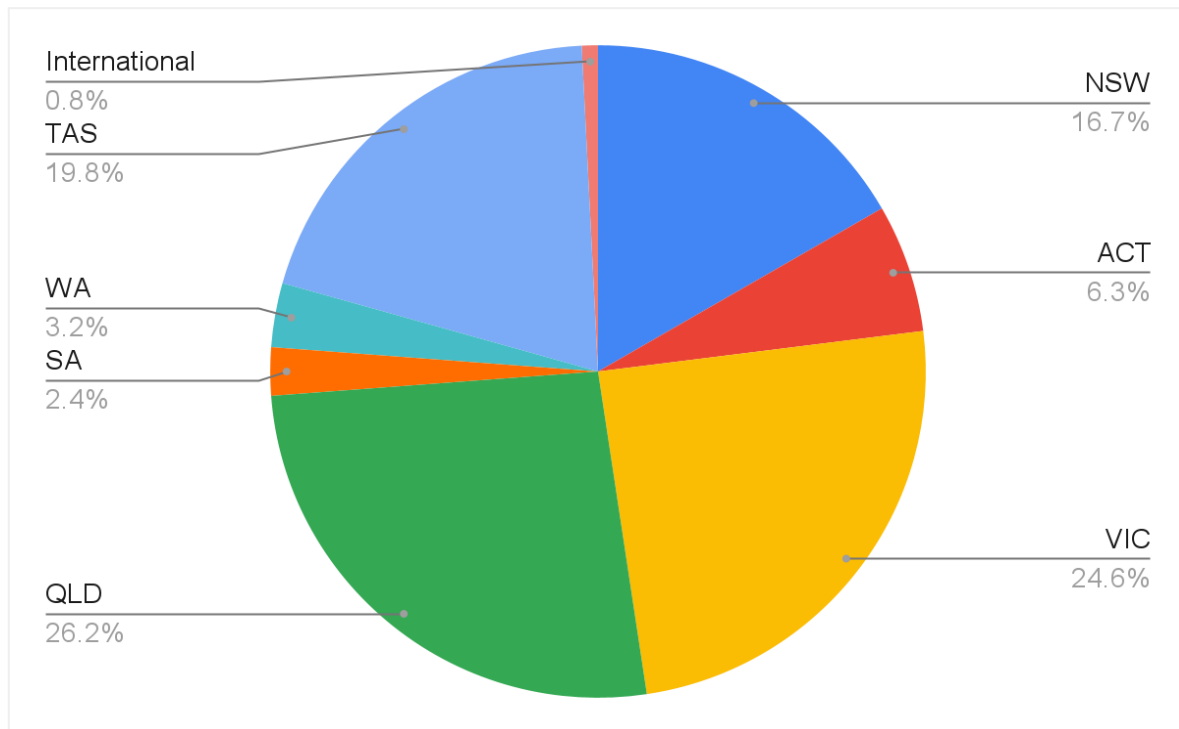


Figure 1: Distribution of responses by states and territories.

Forty-eight organisations were represented amongst the responses with the largest number of respondents from CSIRO, Monash University, The University of Queensland and The University of Tasmania (Table 1).

Table 1: Organisations that respondents are affiliated with.

| Organisation | Number of responses |
|--|---------------------|
| CSIRO | 24 |
| University of Tasmania | 16 |
| Monash University | 15 |
| University of Queensland | 15 |
| UNSW | 4 |
| University of Sydney | 4 |
| Children's Cancer Institute | 3 |
| Queensland University of Technology | 3 |
| University of Queensland | 3 |
| Not provided | 3 |
| Curtin University | 2 |
| Garvan Institute of Medical Research | 2 |
| Hudson Institute of Medical Research | 2 |
| James Cook University | 2 |
| Menzies Institute for Medical Research | 2 |
| Telethon Kids Institute | 2 |
| University of Adelaide | 2 |
| University of Melbourne | 2 |
| University of Southern Queensland | 2 |
| University of Wollongong | 2 |
| Wicking Dementia Research and Education Centre | 2 |
| Central Queensland University | 1 |
| Deakin University | 1 |
| Florey Institute of Neuroscience and Mental Health | 1 |
| Monash Biomedical Discovery Institute | 1 |
| Sydney Institute of Marine Science | 1 |
| University of Sargodha | 1 |
| University of the Sunshine Coast | 1 |
| Victoria University | 1 |
| Victorian Infectious Diseases Reference Laboratory | 1 |
| Western Sydney University | 1 |
| Westmead Institute for Medical research | 1 |
| Total responses | 123 |

Most respondents identified as PhD students (44 respondents, 34.9%), Postdocs (32 respondents, 25.4%) or Senior Scientists/Principal Investigators (27 respondents, 21.4%) (Figure 2) and work across a wide variety of life science topics (Figure 3).

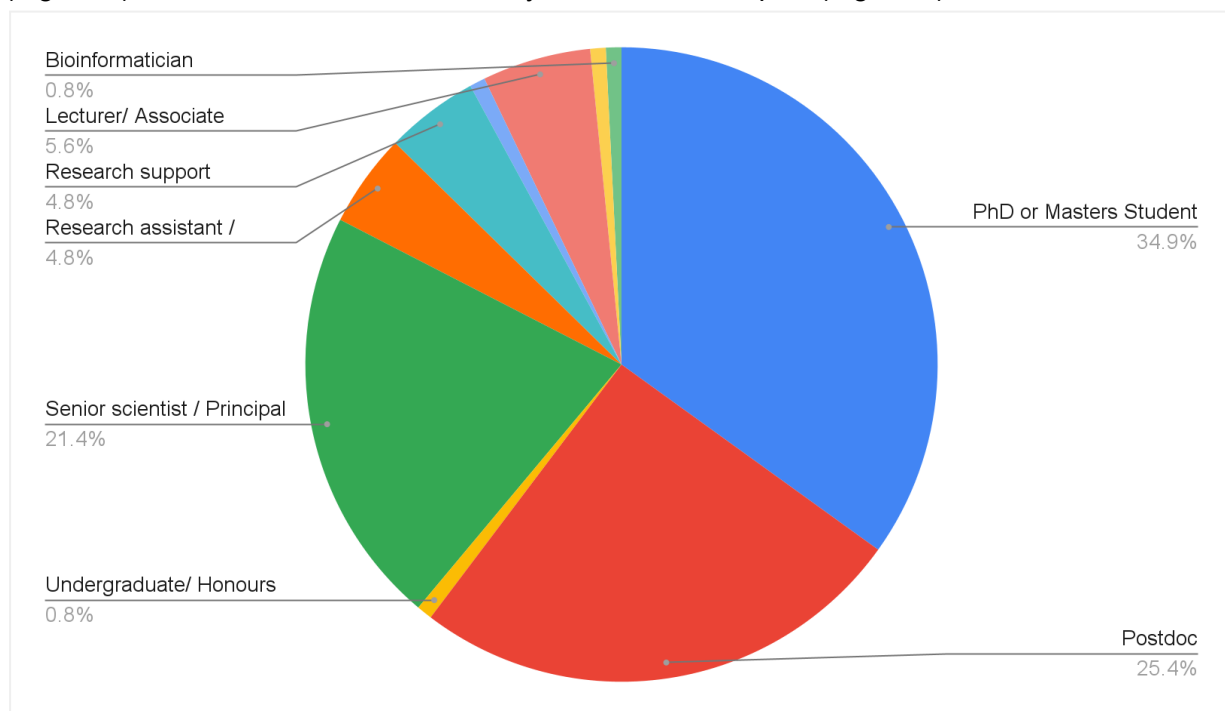


Figure 2: Distribution of responses by primary role.

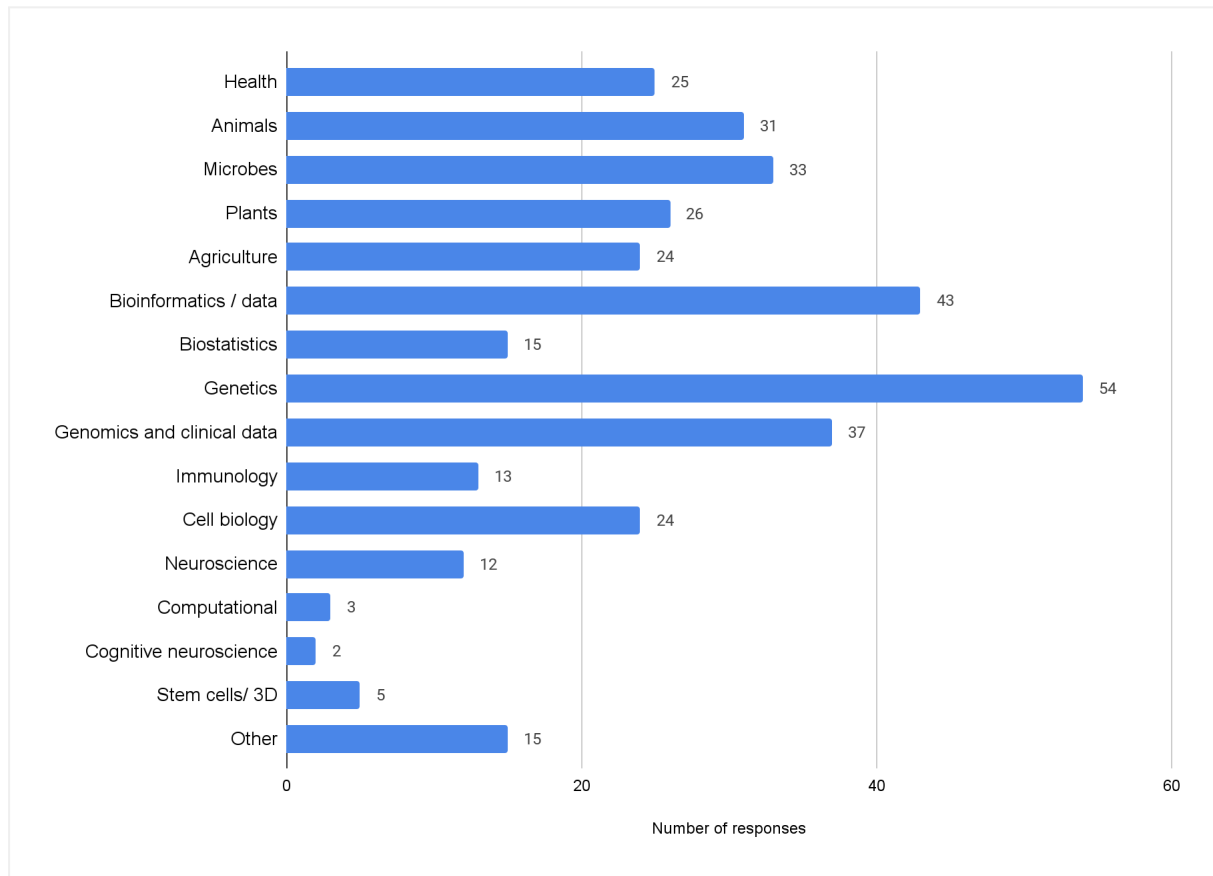


Figure 3: Fields of research in which respondents work. Multiple options could be selected. Topics included under 'Other' were: proteomics, venoms, mass spec omics data, epigenomics, conservation, metagenome, environment, ecology, molecular nutrition, computational biochemistry, transcriptomics, wastewater treatment, epigenetics, microbiology, food science

When asked to rate their level of competency in bioinformatics most considered themselves to be at a beginner (48 respondents, 38.1%) or intermediate level (44 respondents, 34.9%) (Figure 4).

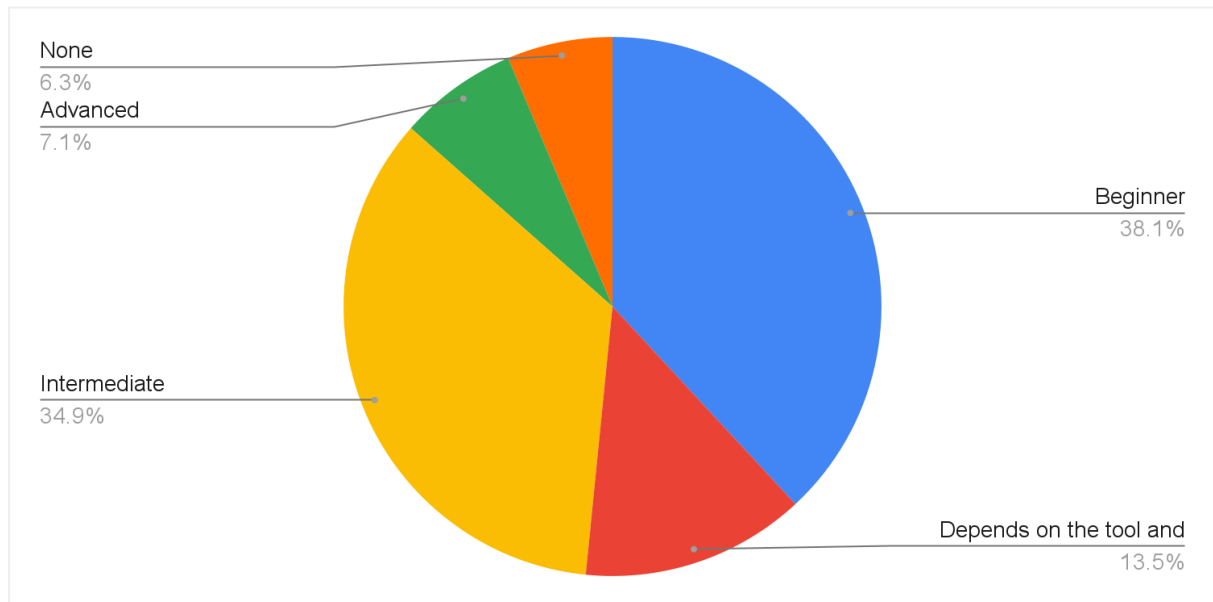


Figure 4: Distribution of respondents by competency in bioinformatics and computational biology. The level of competency is self-assessed and reported.

Current training

Most respondents have accessed training offered by one or more of the members of the National Bioinformatics Training Cooperative (Figure 5).

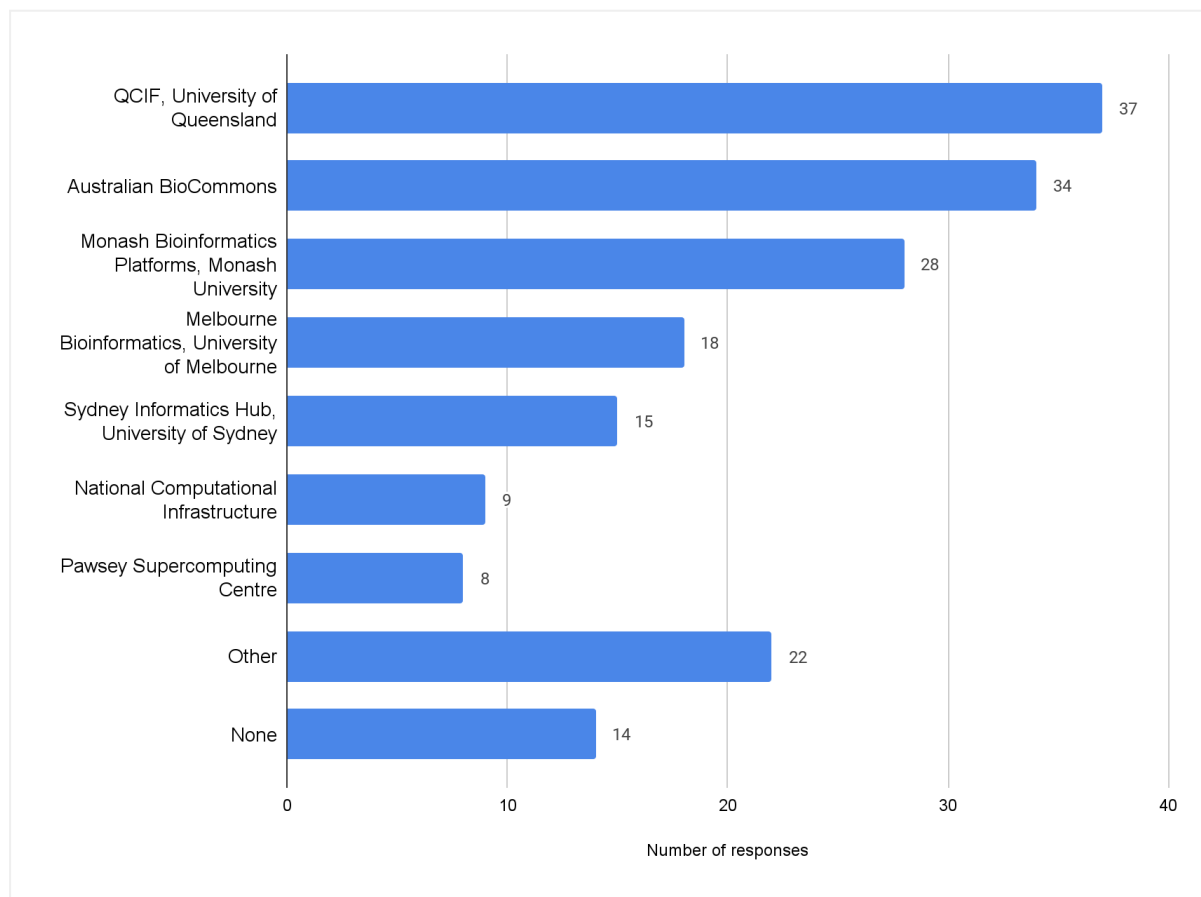


Figure 5: Number of respondents accessing training offered by the National Bioinformatics Training Cooperative members. Multiple options could be selected.

When the data was stratified by location and providers we saw that higher proportions of respondents reported accessing training by their local providers. For example, 27 of 33 respondents from Queensland had accessed QCIF training, 5 of 6 respondents from Sydney had accessed SIH training and 20 of 30 respondents from Victoria had accessed Monash training. This result is likely to be influenced by the routes by which the survey was promoted (i.e by the training providers and mailing lists of past participants).

Other training providers named by respondents included the [Centre for Biodiversity Analysis \(ANU\)](#), [Codecademy](#), [COMBINE](#), CSIRO, [DataCamp](#), [DataTas](#), [Galaxy events](#), [Galaxy Training Materials](#), Harvard University (Trinity and Tuxedo), [Intersect](#), Messer lab (SLiM), [Ramaciotti Centre Training](#), [RCC UQ](#), [RMIT Bioinformatics Group](#), University of Adelaide Bioinformatics Hub (now defunct), University of Tasmania, tertiary education courses in New Zealand (not specified), online courses (not specified).

A small proportion (14 respondents, 11%) responded 'None', 'Nil' or 'NA' to this question suggesting that they are unaware of or have not accessed training by any of the providers listed. Additionally, more than two thirds of respondents were not sure, or felt that their training needs would not be met by their local institutions (Figure 6). Together with anecdotal reports from training providers that life scientists are unaware of sources of training, this suggests that awareness of national bioinformatics training opportunities could be improved.

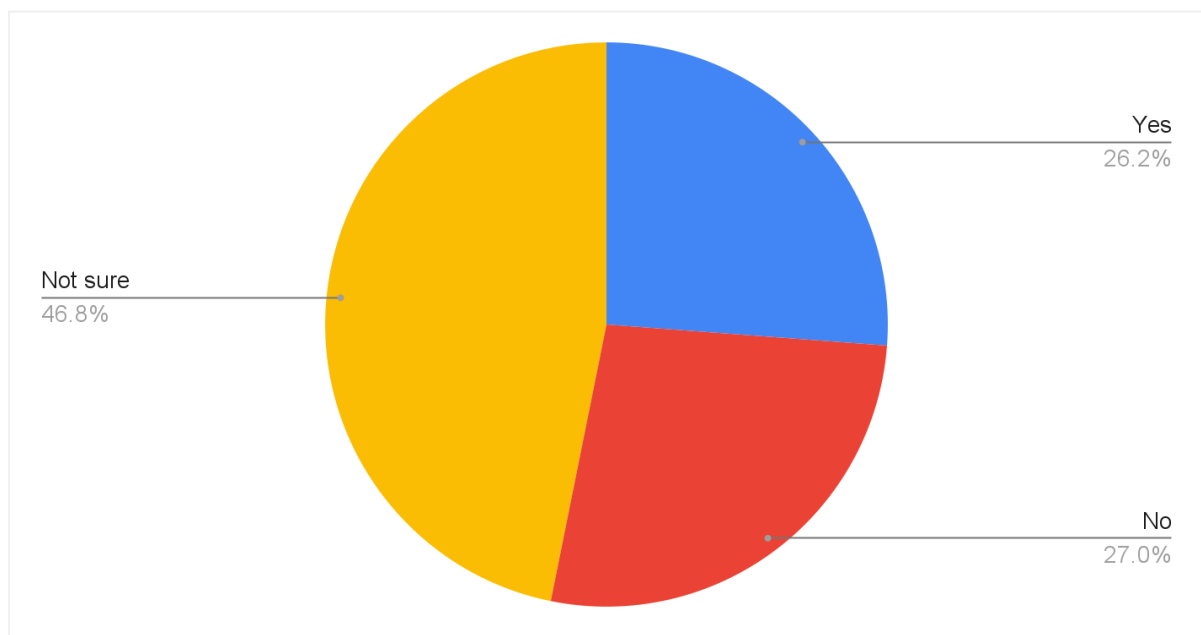


Figure 6: Confidence that training needs will be met by local institutions.

Preferred training format

Similar preferences were expressed for online and in person instructor led training where online options were slightly more popular. Self-paced training or working through examples were also popular formats (Figure 7). There was a clear preference for shorter 3-4 hour sessions and multiple spread out sessions (Figures 8 and 9).

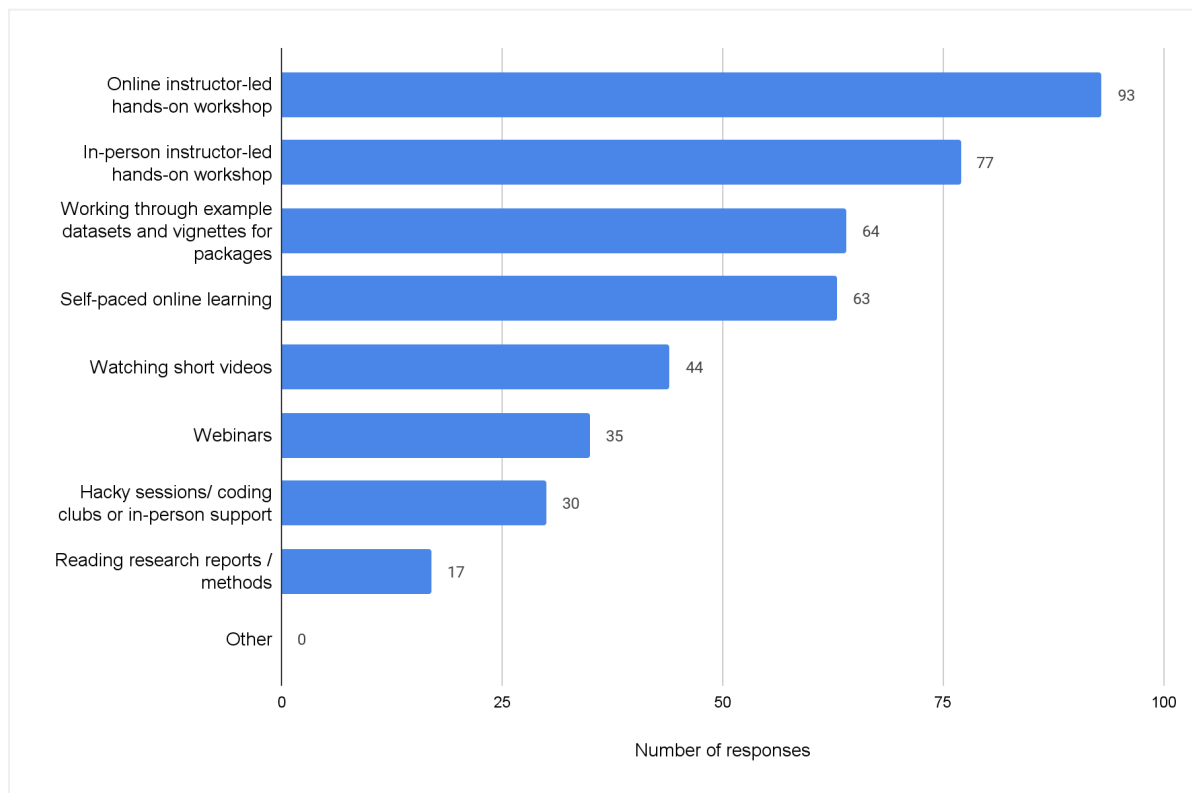


Figure 7: Preferred modes of learning. Multiple options could be selected.

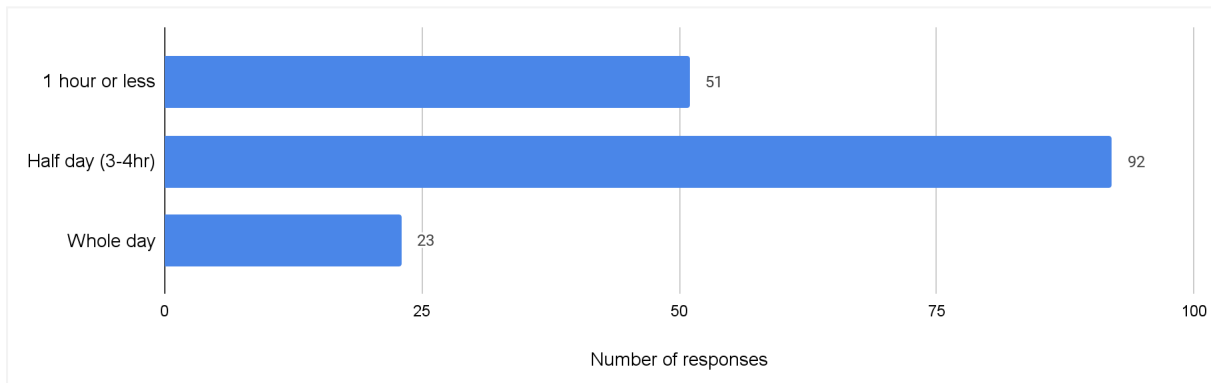


Figure 8: Preferred length of training activity. Multiple options could be selected.

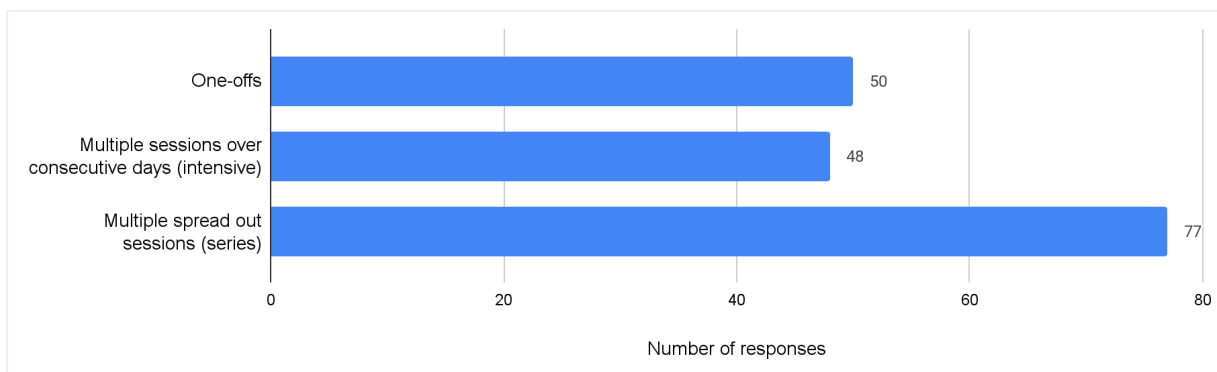


Figure 9: Preferred frequency of training events. Multiple options could be selected.

Desired topics

There is demand for training on all topics and skills suggested by the survey. Higher demand was seen for widely used or applicable techniques such as genome assembly, network analysis and RNAseq. Machine learning was also an in demand topic (Figure 10). Coding languages such as R and Python as well as statistics remain popular (Figure 11). Additional topics suggested by respondents (n=1 per suggestion) include analysis of specific data types (e.g. metagenomics, GWAS, nanopore data), use of particular software (e.g. MATLAB), statistics, ethics, workflows and automation, best practice in coding and software.

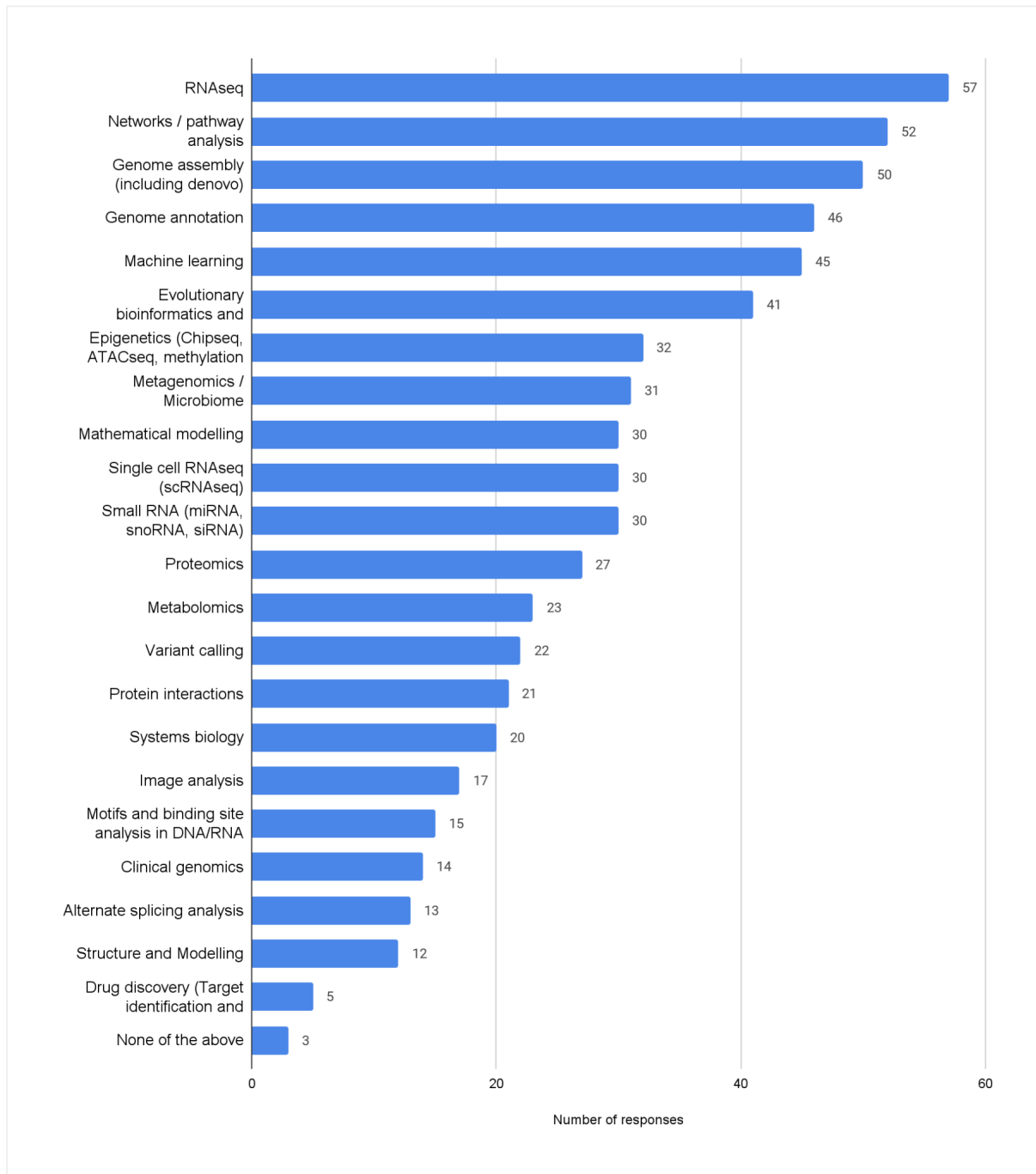


Figure 10: Desired training by topic. Respondents could select multiple options.

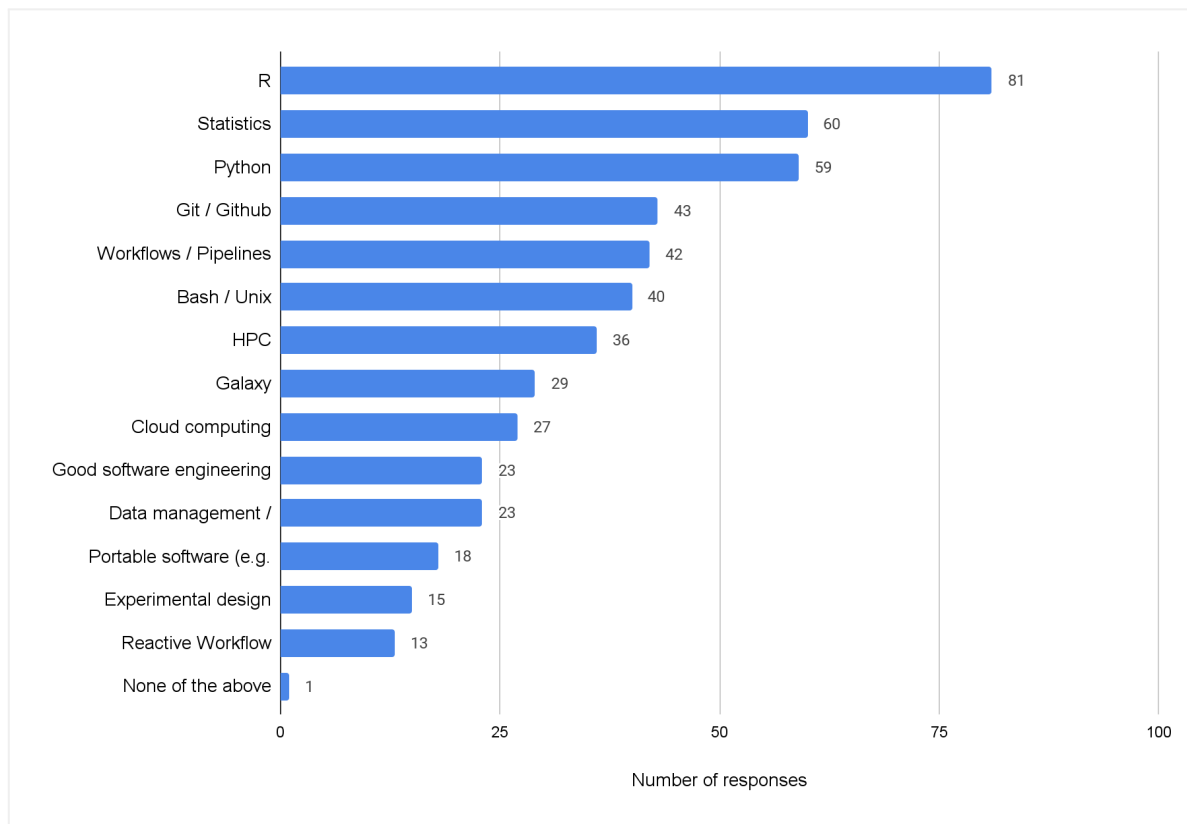


Figure 10: Desired training by skill. Respondents could select multiple options.

Next steps

There are an estimated 30,000 life science researchers in Australia³. The results of this survey offer a glimpse of the bioinformatics training needs of a small subset of this group and provide a platform for the discussion and prioritisation of bioinformatics training in Australia. We plan to repeat this survey on a regular basis to gauge changes in training needs over time and to reach a larger subset of the life sciences community.

Currently, the National Bioinformatics Training Cooperative is using the information gathered in this survey to identify gaps in training offered, prioritise training topics and inform the development of our training program. In view of the results of the survey and current COVID restrictions, we continue to prioritise live online training with the intention of shifting to a mixture of hybrid (as per Australian BioCommons' [previously published model](#)⁴) and fully online events in the future.

The survey indicated that there is some need to raise awareness of existing bioinformatics training in Australia. As a first step to address this we have compiled a list of [Australian bioinformatics training providers](#) and made this publicly available. The Australian BioCommons is also actively involved in promoting the use of training registries such as [DReSA](#)⁵ and [ELIXIR's TeSS](#)⁶ to improve findability of digital skills and bioinformatics



training. Australian BioCommons training events and their associated materials are listed in these registries.

Summary

This survey provides a snapshot of bioinformatics training needs in the life science communities in Australia in late 2021.

It indicates that there is continued demand for training across a wide variety of bioinformatics topics (including omics analysis and programming languages) with new topics emerging since the last survey was conducted in 2016 (e.g. scRNAseq and machine learning). There has also been a shift in attitudes concerning the format of training with online training now more appealing than it was in the past ². Many respondents were unsure if their training needs will be met locally. The cooperative is well placed to address this challenge of access and visibility by enabling training to be delivered online, to a national audience regardless of location.

The information gathered in this survey is being used by the National Bioinformatics Training Cooperative, Australian BioCommons and partners to inform the development of our training programs. Future surveys will enable us to track changes in training needs over time and respond to shifts in demand and scientific thinking.

Contributors

This survey was developed, run and analysed by the [National Bioinformatics Training Cooperative](#), convened by Australian BioCommons.

Licence

This report, the survey questions and anonymised results are provided under [Creative Commons Attribution 4.0 International \(CC BY 4.0\) License](#)

Supplementary files in this Zenodo record

- Questions asked in the survey (PDF):
Australian_bioinformatics_training_needs_survey_2021_2022 .pdf
- Anonymised responses to the survey (names and email address have been redacted):
Australian_bioinformatics_training_needs_survey_2021_2022_anon_results.csv

References

1. Schneider, M. V., Flannery, M. & Griffin, P. Survey of Bioinformatics and Computational Needs in Australia 2016.pdf. (2016) doi:10.6084/m9.figshare.4307768.v1.
2. Attwood, T. K., Blackford, S., Brazas, M. D., Davies, A. & Schneider, M. V. A global perspective on evolving bioinformatics and data science training needs. *Brief. Bioinform.* **20**, 398–404 (2017).
3. Lonie, A. J. What is the Australian BioCommons?
4. Hall, C. R., Griffin, P. C., Lonie, A. J. & Christiansen, J. H. Application of a bioinformatics training delivery method for reaching dispersed and distant trainees. *PLOS Comput. Biol.* **17**, e1008715 (2021).
5. Unsworth, Kathryn *et al.* DReSA: Project team reflections. (2021)
doi:10.5281/ZENODO.5712128.
6. Beard, N. *et al.* TeSS: a platform for discovering life-science training opportunities. *Bioinformatics* **36**, 3290–3291 (2020).