

EDUCATION

Bringing bioinformatics to schools with the 4273pi project

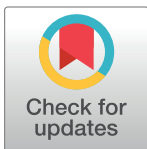
Stevie A. Bain^{1#a*}, Heleen Plaisier^{1#b*}, Felicity Anderson², Nicola Cook³, Kathryn Crouch⁴, Thomas R. Meagher³, Michael G. Ritchie³, Edward W. J. Wallace², Daniel Barker^{1*}

1 Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, United Kingdom, **2** Institute for Cell Biology and SynthSys, School of Biological Sciences, University of Edinburgh, Edinburgh, United Kingdom, **3** Centre for Biological Diversity, School of Biology, University of St Andrews, St Andrews, United Kingdom, **4** Wellcome Centre for Integrative Parasitology, University of Glasgow, Glasgow, United Kingdom

^{#a} Current address: School of Life Sciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, United Kingdom

^{#b} Current address: Royal Botanic Garden Edinburgh, Edinburgh, United Kingdom

* Stevie.Bain@glasgow.ac.uk (SAB); HPlaisier@rbge.org.uk (HP); Daniel.Barker@ed.ac.uk (DB)



OPEN ACCESS

Citation: Bain SA, Plaisier H, Anderson F, Cook N, Crouch K, Meagher TR, et al. (2022) Bringing bioinformatics to schools with the 4273pi project. *PLoS Comput Biol* 18(1): e1009705. <https://doi.org/10.1371/journal.pcbi.1009705>

Editor: Francis Ouellette, McGill University, CANADA

Published: January 20, 2022

Copyright: © 2022 Bain et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The work was supported by the Science and Technology Facilities Council (STFC) under Grants STFC ST/R000328/1 (including salary to S.A.B., D.B., H.P., T.R.M. and non-salary costs) and STFC ST/T000872/1 (including salary to S.A.B., D.B., K.C., T.R.M. and non-salary costs), the Darwin Trust of Edinburgh (<https://darwintrust.bio.ed.ac.uk>; including salary to S.A.B. and H.P. and non-salary costs), the Wellcome Trust-University of Edinburgh Institutional Strategic Support Fund under Wellcome Trust Grant number 204804/Z/16/Z (salary to H.P.), a Public Engagement with Genetics Tier 2 Grant from the Genetics Society (<https://genetics.org.uk>; non-salary costs), the Natural Environment Research Council (NERC) under Grant NE/P000592/1 (including salary to N.

Abstract

Over the last few decades, the nature of life sciences research has changed enormously, generating a need for a workforce with a variety of computational skills such as those required to store, manage, and analyse the large biological datasets produced by next-generation sequencing. Those with such expertise are increasingly in demand for employment in both research and industry. Despite this, bioinformatics education has failed to keep pace with advances in research. At secondary school level, computing is often taught in isolation from other sciences, and its importance in biological research is not fully realised, leaving pupils unprepared for the computational component of Higher Education and, subsequently, research in the life sciences. The 4273pi Bioinformatics at School project (<https://4273pi.org>) aims to address this issue by designing and delivering curriculum-linked, hands-on bioinformatics workshops for secondary school biology pupils, with an emphasis on equitable access. So far, we have reached over 180 schools across Scotland through visits or teacher events, and our open education resources are used internationally. Here, we describe our project, our aims and motivations, and the practical lessons we have learned from implementing a successful bioinformatics education project over the last 5 years.

Introduction

Rapid advances in technology over recent decades have changed the practice of research across all areas of the life sciences. In particular, with the advent of high-throughput, next-generation sequencing technologies, biological research has become more data driven and digital [1]. Bioinformatics—a discipline combining aspects of biology, computing, mathematics, and statistics—is now an essential element of modern biology. Increasingly, over the last few decades, researchers have been required to understand and utilise computational tools to store, manage,

C. and M.G.R. and non-salary costs), the Biotechnology and Biological Sciences Research Council (BBSRC) under Grant BB/S018506/1 (including salary to F.A. and non-salary costs), the School of Biological Sciences at the University of Edinburgh (<https://www.ed.ac.uk/biology>; including salary to S.A.B. and H.P. and non-salary costs) and its Institute of Evolutionary Biology (<https://www.ed.ac.uk/biology/evolutionary-biology>; non-salary costs), the Access for Rural Communities project (ARC) at University of St Andrews (<https://www.st-andrews.ac.uk/study/access/projects/arc>; non-salary costs) and the Engineering and Physical Sciences Research Council (EPSRC) under Grant EP/V52038X/1 (including salary to S.A.B. and non-salary costs). E.W.J.W. is supported by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society [208779/Z/17/Z] (including salary to E.W.J.W.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

and analyse biological data [2,3]. Therefore, biology graduates with bioinformatics training and experience have increased employment opportunities [4] and are better prepared for post-graduate research in life sciences fields [5].

Despite this, bioinformatics education has failed to keep pace with advances in life sciences research [5–7]. In general, much bioinformatics instruction takes place at postgraduate level, e.g., MSc courses and doctoral training programmes [8], meaning that many life sciences students can earn their degrees with little or no bioinformatics training [5,9]. In the last few years, there has been a push to incorporate bioinformatics into undergraduate curricula [10–13]; however, a recent study found that one of the key barriers to this integration is a lack of student preparation [9].

Introducing bioinformatics at secondary school level may resolve this issue. It has been suggested that the earlier computational skills are embedded, the more likely students are to retain this knowledge and their confidence in the subject will increase [14]. Indeed, a number of school-level bioinformatics education programmes and initiatives have been designed with this goal in mind (e.g., [15–17]). In Scotland, unlike in the rest of the UK, bioinformatics is included in the biology curricula for both Higher level Biology and Human Biology [18,19]. These are senior-level qualifications open to pupils usually aged 15 to 18 years old and are required for entry to Higher Education. However, the mention of bioinformatics in the curricula and in the course-recommended textbooks [20,21] is brief. Therefore, while pupils are expected to know the term “bioinformatics” and its relationship to DNA sequencing, there is no formal requirement for practical experience and contextualisation.

To address the disparity between future bioinformatics workforce needs and the level of bioinformatics training at secondary school level, we designed two curricula-linked workshops that introduce pupils (and teachers) to the field of bioinformatics and also reinforce other key topics in the biology curricula such as mutation. In this paper, we describe our educational resources, our approach in targeting our audience, and summarise our successes and limitations over the duration of the project. We also discuss the practical lessons learned from running a successful bioinformatics education project for schools in Scotland.

4273pi bioinformatics education project

Our workshops

A key reason for the success of our project is our strong link to the Scottish Qualifications Authority (SQA) Biology curricula. As well as bringing practical bioinformatics to the classroom, we incorporate other key topics from the curricula to reinforce learning, e.g., mutation and nutrition. We have designed 2 different workshops to fit in with the Scottish biology curricula. Our workshop for pupils in lower secondary school stages—the National level workshop—was designed with the SQA National 4 and National 5 level curricula in mind [22]. These qualifications are open to pupils usually aged 13 to 16 years old, and they are often required for progression to the SQA Higher level qualifications. Our workshop for senior secondary school pupils and students in Further Education (FE)—the Higher level workshop—was designed with the SQA Higher Biology and Human Biology curricula in mind. [Fig 1](#) shows the ages of individuals reached through each of these workshops.

Both workshops use freely available online research resources: the National Center for Biotechnology Information (NCBI) database and its BLAST tool [23,24] and are built on case studies involving DNA sequences. The Higher level workshop additionally introduces the Linux command-line using Raspberry Pi computers. These are small computers designed initially with the aim of promoting the study of computer science in school. They are inexpensive and easily transportable, making them ideal for school visits. We preinstall each Raspberry Pi

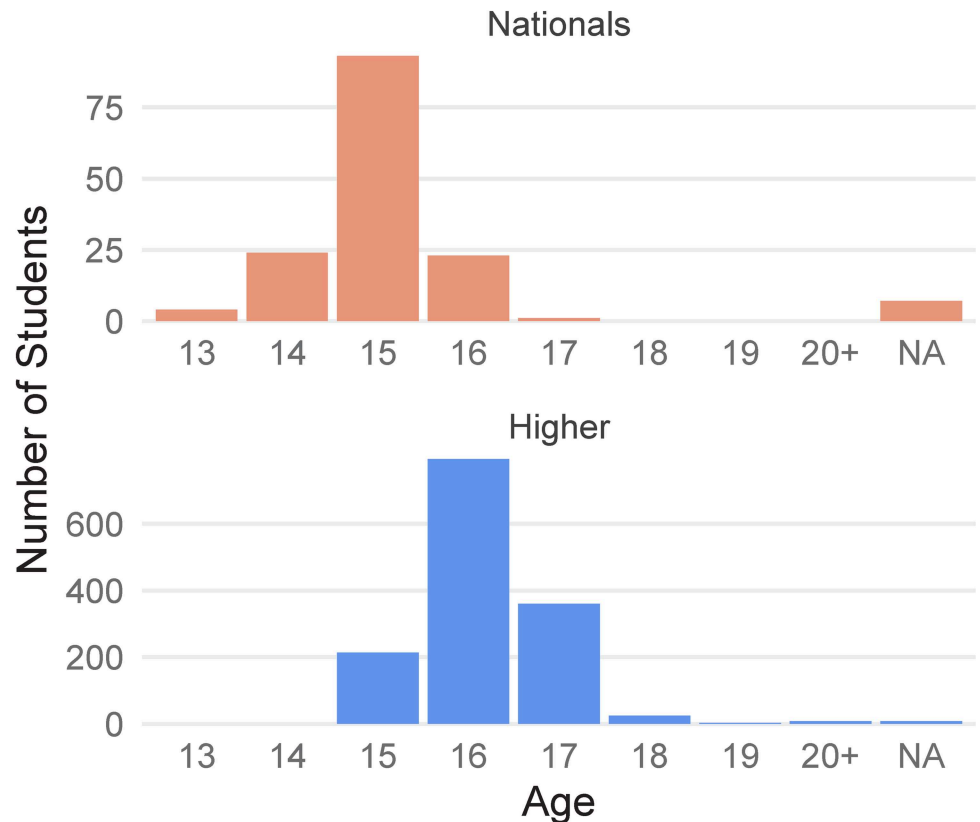


Fig 1. Age range of individuals taking part in the bioinformatics workshops. Nationals graph (top, orange) shows the age range of 152 pupils that have taken part in our SQA National 4/5 level workshop, Bioinformatics: Food Detective. Higher graph (bottom, blue) shows the age range of 1,412 pupils that have taken part in our SQA Higher level workshop, Bioinformatics: The Power of Computers in Biology. This is not the total number of individuals that have participated in the Higher workshop but a subset that filled out evaluation forms. These data represent workshops led by our project staff or volunteers.

<https://doi.org/10.1371/journal.pcbi.1009705.g001>

with the 4273pi variant of the Raspbian operating system (named after a former undergraduate module at the University of St Andrews) [12], and all required software and files. Internet or school network access is not required on these machines during the workshop, as such, their use keeps the focus on the advantages of command-line computing and away from potential IT problems. The Raspberry Pi is important for our workshops but is also ideally suited to bringing a range of computational biology activities to school [25]. Furthermore, these computers are also used for research, meaning that pupils can gain hands-on experience of an authentic computational research environment in their own classrooms.

Bioinformatics: The power of computers in biology

We designed our Higher level workshop based on materials originally developed by DB for first-year undergraduate biology students at the University of St Andrews. This workshop has 2 major aims. The first is educational: We want to help young people understand the role of bioinformatics and the interdisciplinary links between computing, mathematics, and biology by exploring these themes in the context of the SQA Biology and Human Biology curricula, including mutations, genomics, and evolution. The second aim is public engagement: We use bioinformatics as an introduction to techniques in computational science. A full description of this workshop is available in [26]. Feedback from both teachers and pupils suggests that this

workshop was beneficial and supports our hypothesis that practical bioinformatics activities can be conducted by school pupils.

Bioinformatics: Food detective

Our more recently developed workshop is designed for pupils studying SQA National 4 and National 5 level Biology, providing an opportunity to influence the study and career choices of young people at an earlier stage of their education [22]. The workshop is based on real DNA sequences extracted from a handmade pork sausage as a case study in metabarcoding. Pupils use publicly available tools and data to discover that the sausage contains pig DNA but also traces of sheep, chicken, cattle, and human DNA. This informs pupils about a range of subjects including data reliability, contamination, and food fraud. As with our Higher workshop, feedback from both teachers and pupils suggests that this activity is useful and enjoyable. A full description of the workshop and evaluation is reported in [22].

Active learning is known to increase student performance in science, technology, engineering, and mathematics (STEM) subjects [27]. Therefore, both of our workshops aim to maximise pupil participation by asking them to undertake tasks in pairs or small groups, completing the accompanying worksheet as they go along. We also come together as a class between tasks to discuss pupils' findings and to consider the topics covered in more detail. For example, "What happens if humans do not get enough Vitamin C?" and "Why might human DNA be found in the sausage sample?" Overall, the aim of our project is to provide pupils and teachers with hands-on experience of computing in biology and increase understanding of why it is essential in many areas of life sciences research.

While we provide lesson plans for each of our workshops based on how we run our school visits, we also encourage teachers to use our resources in whichever way best suits their needs. This is beneficial for teachers as they can fit certain activities into preexisting lessons, for example, teachers have reported that they used our Food Detective workshop alongside a DNA extraction practical. They felt that our materials complemented the existing practical lesson and reinforced curriculum learning objectives. We believe that creating resources, which can be used flexibly, either fully or in part, can increase teacher ownership of the lesson. Thus, our resources feel less "outsourced" and more embedded in the delivery of the course material.

Throughout the development of this project, we have made use of the advice provided by others involved in school-level bioinformatics education such as [28]. We have created a list of lessons learned and tips for others interested in developing their own school projects in [Box 1](#).

Action research

Action research is a solution-oriented methodological approach commonly used in the social sciences. It is a reflective and often collaborative process where researchers engage in a cycle of posing questions, gathering data, critically reflecting on findings and experiences, and implementing practical changes based on reflection [29]. We use our experiences in the classroom with pupils and teachers, and formal feedback to influence the development and delivery of our resources. We found that strong partnerships with schools and teachers are key to successfully implementing action research. Teachers have a thorough understanding of school curricula, the level at which activities should be pitched, and what activities pupils will learn from and enjoy. Therefore, incorporating teacher feedback throughout the workshop design and delivery process has been key to the success of our project.

Action research was used in the development of both of our workshops. We piloted the Food Detective workshop with junior pupils in several Scottish schools. Initially, we were concerned that the workshop may be too challenging for pupils; however, we found that pupils

Box 1. Lessons learned and tips for others

- Strong links to the curriculum are vital. Develop interesting, real-life case studies that pupils can relate to.
- Engage with teachers while creating and delivering workshops as they have a good understanding of what content will work for their pupils.
- Participate in action research—use experience and feedback to develop resources and make changes when needed.
- Provide training and resources for teachers. This will increase project reach and legacy.
- On school visits, work with teachers. Make sure a teacher is present to cover matters such as pupil well-being or requests to leave the classroom.
- Embed computing activities into the school biology curriculum to help address gender gaps and improve access and exposure to this traditionally male-biased subject.
- Have a diverse delivery team. For example, a mixed-gender team demonstrates the important points that computing is not “just for boys” and biology is not “just for girls.” Other dimensions of diversity are just as important.

performed well and enjoyed the activities. They also managed to complete the workshop more quickly than we had anticipated, and in their feedback, both teachers and pupils requested more activities. We added a second task to this workshop, which focused on e-values and the reliability of results. In subsequent workshops, we found that pupils were more challenged but still engaged. The feedback showed that they enjoyed both tasks and teachers felt that the activities were pitched at the correct level for this age group. In this case, action research helped us to refine the content of our workshop based on the interests and abilities of pupils.

Delivery of workshops and dissemination of resources

The main challenges to incorporating bioinformatics into the secondary school curriculum are infrastructure, logistics, and teacher expertise. Our project aims to address these issues through a number of actions. The Nationals workshop and two-thirds of the Higher workshop (i.e., the non-Raspberry Pi-based sections) can be carried out on any device that has access to the internet, including school desktop computers, tablets, or smartphones. Many local authorities across Scotland are now providing pupils with tablet devices for use in the classroom, which means that most of our activities can be carried out without the need to book a computer suite or library, reducing the organisational burden for teachers and schools.

We learned that to effectively reach large numbers of pupils without potential audience bias and exclusion that may occur through self-selection for extracurricular activities, whole-class visits during the school day are best. In the beginning, our workshops extended across 4 periods in the school day (approximately 3 hours 20 minutes); we later compressed this to 3 periods. However, we became aware of the disruption this caused to the school timetable and decided to adjust the duration to fit within a timetabled double period of biology (approximately 1 hour 40 minutes). We now avoid negatively impacting on other subjects and have eased the logistical burden teachers face when inviting us to their schools. We also visit all

classes for any given qualification. For example, if a school has 5 Higher Biology classes, we arrange to deliver the workshop to all these classes. This approach ensures that we reach the maximum number of pupils possible.

In addition to our school workshops, our programme of free Continuing Professional Development (CPD) workshops for teachers provides an excellent opportunity to expand the reach of our project and strengthen our partnerships with schools. Teachers often appreciate the opportunity to gain a better understanding of bioinformatics and our resources with their colleagues before using them with their classes. We also contribute to initial teacher training for students enrolled in Professional Graduate Diploma in Education courses for STEM subjects. We provide teachers and student teachers with training that will enable them to deliver our workshops to their classes, which helps to increase the reach of our project. We also provide additional training on how bioinformatics can be implemented in other areas of the biology curricula, e.g., how to use the NCBI database to find the amino acid sequence of an enzyme discussed in class. Our school workshops and teacher training are delivered by a core team of staff and trained volunteers from several universities across the UK.

When coordinating events for teachers, collaborating with educational institutions (in our case, Scottish Schools Education Research Centre (SSERC) or the Science Skills Academy of Highlands and Islands Enterprise (SSA)) and local authorities is beneficial. These institutions can provide support such as venues, access to school mailing lists, and details of in-service and teacher training days.

The importance of online and open education resources

Open education resources (OERs) have a key role in ensuring inclusive and equitable quality education opportunities for all (United Nations Sustainable Development Goal 4) [30]. The Coronavirus Disease 2019 (COVID-19) pandemic has further highlighted their importance at both a local and global level. Our workshops are free to attend, and our materials are available as OERs, removing financial burdens from teachers and schools. To further extend our reach, we have also published Practical Guides for each of our workshops with GOBLET, the Global Organisation for Bioinformatics Learning, Education, and Training [26,31]. GOBLET provides high-quality, open-source resources to support bioinformatics learning, education, and training across the world [2]. There are resources designed to cover all levels of bioinformatics training from schools to research laboratories, and thus this partnership allows us to increase our audience reach.

In response to the COVID-19 pandemic, we developed several additional online resources that made it possible for us to continue to support and encourage teachers to incorporate bioinformatics into biology lessons. In addition to their usefulness throughout the pandemic, these resources have demonstrated the potential to increase the reach and scalability of our project in the future. We created a teacher-oriented video series designed to support teacher CPD (<https://4273pi.org/videos>). These short videos (5 to 10 minutes) run through the key aspects of our workshops and also provide information about using the NCBI database in other areas of biology curricula, e.g., using the NCBI 3D protein viewer to examine the structure of enzymes. All videos have subtitles and accompanying transcripts to increase accessibility.

Feedback from pupils, teachers, and STEM education professionals shows that our workshops and OERs are useful and effective learning tools. As a testament to this, our workshops and videos have been added to a list of online OERs compiled by Education Scotland to support teachers with online learning.

We also adapted our in-person CPD event for teachers to an online delivery model. Discussions with teachers and colleagues from Education Scotland indicated that twilight sessions

(between 4.30 to 6.30 PM) would maximise teacher participation without disrupting the school day. In addition, we offer CPD during teacher in-service days. In advance of each CPD event, we send teachers a resource kit containing a Raspberry Pi computer complete with SD card image customised for bioinformatics, plus a case, power supply, and monitor cables; an introductory bioinformatics textbook [32]; and a hard copy of our handouts (these are also available electronically). As with our in-person CPD, these events and resource kits are free, and our aim is to provide teachers with the equipment, knowledge, and confidence required to integrate bioinformatics into their classroom teaching. At some CPD events, we also host guest speakers who present case studies of their own computational biology research. Even when we resume in-person CPD, we will also continue to provide a programme of online CPD events, as these allow us to reach a larger audience more easily, particularly in remote areas such as the Highlands and Islands region of Scotland. Our online CPD events also eliminate any travel costs associated with attending, which is important as many schools have limited budgets and costs can be substantial for teachers, especially those in remote areas.

Project reach and prioritising socially disadvantaged areas

Every education project should have a clearly defined target audience, one that will benefit from the resources produced and that is logistically feasible to reach. Our primary focus is Scotland, for several reasons. The first reason is that bioinformatics is included in the Scottish Biology curriculum, meaning that we can create resources that link directly to course content in schools. Secondly, Scotland provides a tractable target audience that our project can easily reach without the need for satellite groups. By focusing our efforts on a specific geographical area, we can achieve fuller coverage and build relationships with schools. Additionally, there is only one Scottish exam board, the SQA, and all Scottish schools are supported by the same Government agencies, e.g., Education Scotland.

We have directly reached over 180 Scottish secondary schools since 2016 (Fig 2), including almost 50% of state secondary schools. We advertise our project across Scotland via a biology teacher e-mail list and ask any interested teachers to get in touch with us. We also have strong working relationships with organisations such as Education Scotland, SSERC, and SSA that have supported us by facilitating our in-person CPD events (SSERC and SSA) and contributing to the coordination of our online CPD events. Although our primary focus is Scotland, our OERs are used internationally (e.g., elsewhere in the UK, Mexico, Argentina, India, the USA, Finland, Brazil, China, the Netherlands, and Saudi Arabia).

Science and technology can be inspiring to all; however, we know that the playing field is not level [33]. One set of systemic biases prominent in STEM sectors is discrimination against socioeconomically disadvantaged groups [34]. These inequalities are also seen at the school level where socioeconomically disadvantaged pupils, despite being interested in science, have fewer opportunities to engage with science both inside and outside of school [35].

We often have a waiting list for our visits and, initially, we used to target specific geographical areas each year to maximise our coverage of Scotland. However, we know that there are areas in the country where pupils do not have the same opportunities to explore and engage with science and technology as others. Therefore, we now use the Scottish Index of Multiple Deprivation (SIMD) [36] to prioritise our school visits. The SIMD is a relative measure of deprivation across Scotland. In this context, “deprived” is not synonymous with poor, and the SIMD is used to describe areas, not people. However, those who live in the “most deprived” areas are more likely have limited access to Higher Education Institutes and public engagement opportunities. The index looks at 7 domains relevant to resources and opportunities (income, employment, education, health, access to services, crime, and housing) within a

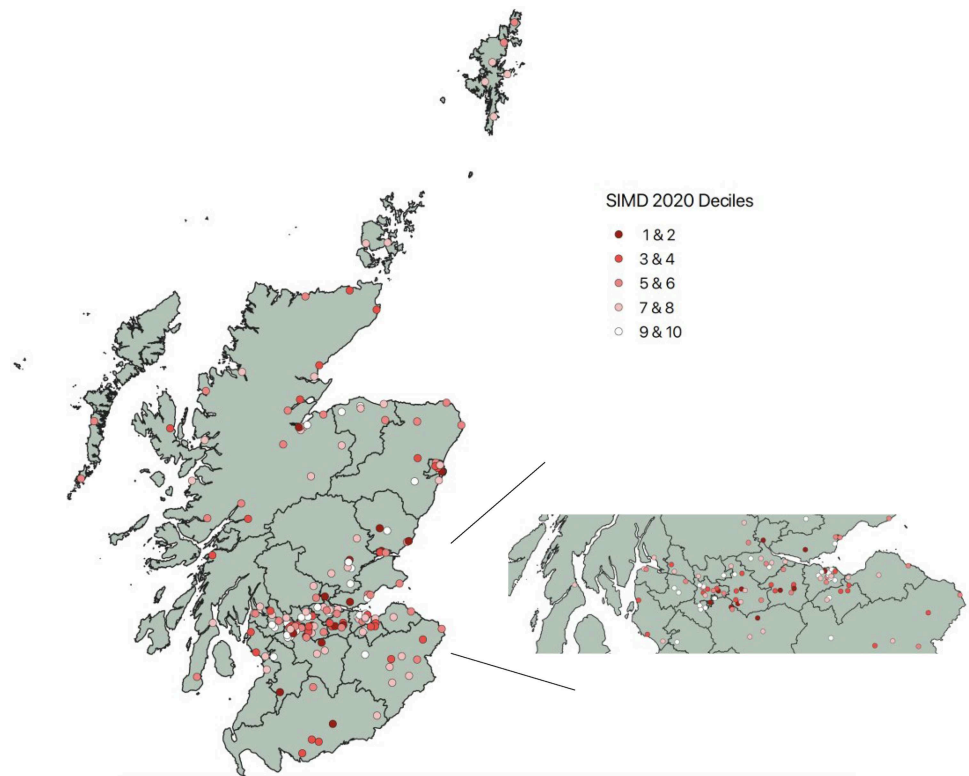


Fig 2. A map of Scotland showing the schools our project has reached since 2016. Each school reached is represented by a coloured circle on the map. Some schools have been reached more than once. Circles are coloured according to Scottish Index of Multiple Deprivation 2020 (SIMD) decile (1—most deprived to 10—least deprived) [36]. Dark red circles represent schools in the most deprived areas of Scotland according to SIMD, and the white circles represent schools in the least deprived areas. A gradient colour scheme represents those areas in between. The zoomed-in section shows the densely populated central belt region with more clarity. (Map shape file downloaded from <https://www.nrscotland.gov.uk/statistics-and-data/geography/our-products/census-datasets/2011-census/2011-boundaries>. Contains NRS data Copyright Crown copyright and database right 2021. Contains Ordnance Survey data Copyright Crown copyright and database right 2021.

<https://doi.org/10.1371/journal.pcbi.1009705.g002>

certain geographical area, called a datazone. The SIMD is used by the Scottish Government and funding bodies to effectively target policies and funding to areas where deprivation is high. We target low SIMD areas (i.e., those most deprived on the SIMD scale), which often include postindustrial towns and cities. Initially, our project used the SIMD of the school itself to target visits, by using the school's postcode. However, we now recognise that this is not a perfect proxy for targeting pupils that come from low SIMD areas. For example, some pupils in a school will almost certainly live in areas that have a different SIMD score than the score of their school, particularly if the school is in a city centre. To address this, we have recently focused our efforts into targeting schools that have a high proportion of pupils who actually live in low SIMD areas. Publicly available data from the Scottish Government (available from: <https://www2.gov.scot/Topics/Statistics/Browse/School-Education/Datasets/contactdetails>) showed 42 secondary schools in Scotland where >50% of pupils attending lived in SIMD deciles 1 and 2 (most deprived areas in Scotland). From these data, we have identified areas and schools that we will target for future visits. By targeting low SIMD schools for our teacher CPD and school visits, we can promote our OERs and support schools to access and use them. Equitable access to OERs is essential to ensure that these resources will overcome existing inequalities rather than entrench them [37]. It is important to note that our workshops are not

designed specifically for or delivered only to low SIMD areas, but we do believe it is important to prioritise schools where public engagement and university contact is less likely to happen.

As well as the SIMD, we also take into consideration the location of schools. Approximately 17% of Scotland's population live in areas defined as rural, and approximately 20% of Scotland's secondary schools are rural [34,35]. Although not necessarily classified as "most deprived" datazones in overall SIMD analyses, rural areas often have reduced access to public engagement and education events hosted by Higher Education institutes compared to non-rural areas [36,37]. The commission on the Delivery of Rural Education 2013 [38] recommended to the Scottish Government that Further and Higher Education institutions, local authorities, and schools should work together to provide the widest possible range of opportunities to young people and adults in rural areas. Another recommendation was to provide access to CPD to teachers in rural areas, reducing isolation and sustaining skills and development. At an early stage of the project, the strong demand for CPD in locations outside the Scottish central belt became apparent to us through direct contact with teachers. Therefore, each year we run events that specifically target schools and teachers in the Highlands and Islands regions. Our online CPD events allow us to reach these teachers and schools more effectively.

Addressing the gender bias in computing

In the UK and some other countries, computer science education is associated with a startling gender inequality (Table 1) [38]. Ironically, computer programmers were initially predominantly female, partly due to the accelerated development of computer technology during World War II [39,40]. Later, computing became male dominated, due to misguided efforts of government and industry in the 1960s to 1980s to bring the topic to a managerial (i.e., male) level [41]. In the UK, only 18% of those studying a computing science at university level are female [38]. Importantly, it should be noted that this bias is not common to all countries, for example, in India, women represent 45% of total computer science enrolment in universities [42]. Early introduction of computing sciences at primary and secondary school level is

Table 1. Percentage of female pupils at different levels of Biology and Computing across Scotland.

Subject	SQA level	SCQF level	Percentage of pupils that are female
Biology	National 4	4	63%
	National 5	5	67%
	Higher	6	66%
Computing	National 4	4	16%
	National 5	5	15%
	Higher	6	16%

These figures are taken from the SQA Statistical Report 2018 [44]. The figures clearly show that Biology is a female-biased subject and that Computing is the opposite with a striking male bias. Therefore, we believe that by targeting Biology classes, we are bringing computation to an audience that would not necessarily receive this education. SCQF levels indicate the difficulty of particular subjects in the Scottish education system. These are comparable to the RQF and CQFW in England and Wales, and the EQF. SCQF levels 4, 5, and 6 are comparable to RQF and CQFW levels 1, 2, and 3, respectively, and EQF levels 2, 3, and 4, respectively (more detail available at <https://www.sqa.org.uk/sqa/64561.html>).

CQFW, Credit and Qualifications Framework; EQF, European Qualifications Framework; RQF, Regulated Qualifications Framework; SCQF, Scottish Credit and Qualifications Framework; SQA, Scottish Qualifications Authority.

<https://doi.org/10.1371/journal.pcbi.1009705.t001>

suggested to have a significant impact on gender representation at university level [42]. In Malaysia, gender ratios in computer science education are not merely balanced; sometimes they are dominated by women [43]. Again, early computer science education is noted as a factor influencing these patterns [43]. This highlights the fact that more can and needs to be done to remedy the strong gender bias that exists in Scotland, the UK, and in many other countries. There is strong evidence to suggest that the earlier computer science education is incorporated into learning, the better this bias is addressed [14]. We also believe that school engagement and educational programmes must look beyond the obvious audience of pupils already engaged with computing education (male biased; see Table 1) and even beyond the audience of young people with any self-declared interest in computing.

Our project specifically targets biology classes; approximately 66% of pupils enrolled in SQA Higher Biology are female [44]. Furthermore, 63% and 67% of pupils enrolled in SQA National 4 and 5 level Biology, respectively, are female [44]. In our own data, we see that 69% of our pupil audience identifies as female. This is a reversal of the strong gender bias associated with computing classes. SQA National 4 and 5 Computing classes are only 16% and 15% female, respectively. SQA Higher Computing classes are 16% female. Our project has the potential to inform and influence the future career goals of these pupils, thus helping to address the extreme gender gap at later career stages in computer sciences. Furthermore, feedback from pupils shows that there are no significant differences in enjoyment and usefulness of our workshops between the genders [22]. At educational levels where pupils have a choice of subjects, targeting biology classes—rather than computing classes—will have advantages for reducing gender bias in bioinformatics.

Project impact and plans

Looking to the future, we aim to increase our reach and to focus particularly on schools with a high percentage of pupils from low SIMD areas, as we realise that these often miss out on interactions with Higher Education Institutes. We also hope the project can play a role in increasing the participation of students from underrepresented backgrounds (e.g., social class, first-generation students, and ethnic minorities, to name a few) in Higher Education. In particular, we aim to increase female pupil exposure to computational biology, highlighting the field as a potential career avenue.

In addition, we plan to develop a workshop for pupils studying computing and target this workshop at early secondary school pupils, when computing is still compulsory. Our aim is to highlight that computation is far from an insular subject. It has many diverse applications, and the storage and analysis of big data in biology provide an excellent example of this. As computing is increasingly central to many areas of life, bioinformatics is increasingly central to life sciences. We are proud to be part of the effort to bring bioinformatics to young people.

Acknowledgments

The authors would like to thank all participants in our activities and the volunteers and staff who helped deliver the workshops.

References

1. Marx V. The big challenges of big data. *Nature*. 2013; 498:255–60. <https://doi.org/10.1038/498255a> PMID: 23765498
2. Attwood TK, Bongcam-Rudloff E, Brazas ME, Corpas M, Gaudet P, Lewitter F, et al. GOBLET: The Global Organisation for Bioinformatics Learning, Education and Training. Welch L, editor. *PLoS Comput Biol*. 2015; 11: e1004143. <https://doi.org/10.1371/journal.pcbi.1004143> PMID: 25856076

3. Pevzner P, Shamir R. Computing Has Changed Biology—Biology Education Must Catch Up. *Science*. 2009; 325:541–2. <https://doi.org/10.1126/science.1173876> PMID: 19644094
4. Levine AG. An Explosion of Bioinformatics Careers. *Science*. 2014. <https://doi.org/10.1126/science.opms.r1400143>
5. Hack C, Kendall G. Bioinformatics: Current practice and future challenges for life science education. *Biochem Mol Biol Educ*. 2006; 33:82–5. <https://doi.org/10.1002/bmb.2005.494033022424> PMID: 21638550
6. Cummings MP, Temple GG. Broader incorporation of bioinformatics in education: opportunities and challenges. *Brief Bioinform*. 2010; 11:537–43. <https://doi.org/10.1093/bib/bbq058> PMID: 20798182
7. Porter SG, Smith TM. Bioinformatics for the Masses: The Need for Practical Data Science in Undergraduate Biology. *OMICS J Integr Biol*. 2019; 23:297–9. <https://doi.org/10.1089/omi.2019.0080> PMID: 31107144
8. Magana AJ, Taleyarkhan M, Alvarado DR, Kane M, Springer J, Clase K. A Survey of Scholarly Literature Describing the Field of Bioinformatics Education and Bioinformatics Educational Research. *CBE—Life Sci Educ*. 2014; 13:17. <https://doi.org/10.1187/cbe.13-10-0193> PMID: 25452484
9. Williams JJ, Drew JC, Galindo-Gonzalez S, Robic S, Dinsdale E, Morgan WR, et al. Barriers to integration of bioinformatics into undergraduate life sciences education: A national study of US life sciences faculty uncover significant barriers to integrating bioinformatics into undergraduate instruction. *PLoS ONE*. 2019; 14:e0224288. <https://doi.org/10.1371/journal.pone.0224288> PMID: 31738797
10. Gatherer D. Reflections on integrating bioinformatics into the undergraduate curriculum: The Lancaster experience. *Biochem Mol Biol Educ*. 2020; 48:118–27. <https://doi.org/10.1002/bmb.21320> PMID: 31793726
11. Centeno NB, Villà-Freixa J, Oliva B. Teaching structural bioinformatics at the undergraduate level. *Biochem Mol Biol Educ*. 2003; 31:386–91. <https://doi.org/10.1002/bmb.2003.494031060287>
12. Barker D, Ferrier DE, Holland PW, Mitchell JB, Plaisier H, Ritchie MG, et al. 4273π: Bioinformatics education on low cost ARM hardware. *BMC Bioinformatics*. 2013; 14:243. <https://doi.org/10.1186/1471-2105-14-243> PMID: 23937194
13. Oke M, Agbalajobi R, Osifeso M, Muhammad B, Lawal H, Mai M, et al. Design and implementation of structural bioinformatics projects for biological sciences undergraduate students: Structural Bioinformatics Project for Undergraduate Students. *Biochem Mol Biol Educ*. 2018; 46:547–54. <https://doi.org/10.1002/bmb.21169> PMID: 30369034
14. Attwood TK, Blackford S, Brazas MD, Davies A, Schneider MV. A global perspective on evolving bioinformatics and data science training needs. *Brief Bioinform*. 2019; 20:398–404. <https://doi.org/10.1093/bib/bbx100> PMID: 28968751
15. Machluf Y, Yarden A. Integrating bioinformatics into senior high school: design principles and implications. *Brief Bioinform*. 2013; 14:648–60. <https://doi.org/10.1093/bib/bbt030> PMID: 23665511
16. Martins A, Fonseca MJ, Tavares F. Mining the Genome: Using Bioinformatics Tools in the Classroom to Support Student Discovery of Genes. *Am Biol Teach*. 2018; 80:619–24. <https://doi.org/10.1525/abt.2018.80.8.619>
17. Martins A, Fonseca MJ, Lemos M, Lencastre L, Tavares F. Bioinformatics-Based Activities in High School: Fostering Students' Literacy, Interest, and Attitudes on Gene Regulation, Genomics, and Evolution. *Front Microbiol*. 2020; 11:578099. <https://doi.org/10.3389/fmicb.2020.578099> PMID: 33162959
18. Scottish Qualifications Authority. Higher Course Specification: Higher Biology. SQA; 2019. Available from: https://www.sqa.org.uk/files_ccc/HigherCourseSpecBiology.pdf
19. Scottish Qualifications Authority. Higher Course Specification: Higher Human Biology. SQA; 2020. Available from: https://www.sqa.org.uk/files_ccc/HigherCourseSpecHumanBiology.pdf
20. Marsh C, Simms J, Stevenson C, Torrance J, Fullarton J. Higher biology. 2nd ed. Hodder Gibson; 2019.
21. Torrance J. Higher human biology. 2nd ed. Hodder Gibson; 2019. Available from: <https://public.ebookcentral.proquest.com/choice/publicfullrecord.aspx?p=5802579>
22. Bain SA, Meagher TR, Barker D. Design, delivery and evaluation of a bioinformatics education workshop for 13-16-year-olds. *J Biol Educ*. 2020; 1–11. <https://doi.org/10.1080/00219266.2020.1858932>
23. Resource NCBI. Coordinators. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*. 2016; 44:D7–D19. <https://doi.org/10.1093/nar/gkv1290> PMID: 26615191
24. Altschul S, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997; 25:3389–402. <https://doi.org/10.1093/nar/25.17.3389> PMID: 9254694

25. Fletcher AC, Mura C. Ten quick tips for using a Raspberry Pi. *PLoS Comput Biol*. 2019; 15:e1006959. <https://doi.org/10.1371/journal.pcbi.1006959> PMID: 31048834
26. Barker D, Plaisier H, Bain SA, Attwood TK. *Bioinformatics: The Power of Computers in Biology—a Practical Guide*. F1000Res. 2019. <https://doi.org/10.7490/F1000RESEARCH.1116842.1>
27. Freeman S, Eddy SL, McDonough M, Smith MK, Okoroafor N, Jordt H, et al. Active learning increases student performance in science, engineering, and mathematics. *Proc Natl Acad Sci*. 2014; 111:8410–5. <https://doi.org/10.1073/pnas.1319030111> PMID: 24821756
28. Form D, Lewitter F. Ten Simple Rules for Teaching Bioinformatics at the High School Level. *PLoS Comput Biol*. 2011; 7:e1002243. <https://doi.org/10.1371/journal.pcbi.1002243> PMID: 22046117
29. Kemmis S, McTaggart R, Nixon R. *The Action Research Planner*. Singapore: Springer Singapore; 2014. <https://doi.org/10.1007/978-981-4560-67-2>
30. United Nations Statistics Division. *The Sustainable Development Goals Report*. United Nations; 2021. Available from: <https://unstats.un.org/sdgs/report/2021/#sdg-goals>
31. Bain SA, Barker D, Attwood TK. *Bioinformatics: Food Detective—a Practical Guide*. F1000Res. 2020. <https://doi.org/10.7490/F1000RESEARCH.1118007.1>
32. Attwood TK, Pettifer SR, Thorne D. *Bioinformatics challenges at the interface of biology and computer science: mind the gap*. 1st ed. John Wiley & Sons Inc; 2016.
33. Science and Technology Facilities Council. *The Wonder initiative*. In: *Public Engagement: The Wonder initiative* [Internet]. 16 Apr 2019 [cited 2021 Aug 25]. Available from: <https://stfc.ukri.org/public-engagement/wonder/>
34. Campaign for Science and Engineering. *Improving diversity in STEM*. London: Campaign for Science and Engineering; 2014. Available from: <https://www.sciencecampaign.org.uk/static/uploaded/50c4b928-d252-4ce8-825065f92d8deca3.pdf>
35. Archer MO. Schools of all backgrounds can do physics research—on the accessibility and equity of the Physics Research in School Environments (PRISE) approach to independent research projects. *Geosci Commun*. 2021; 4:189–208. <https://doi.org/10.5194/gc-4-189-2021>
36. Scottish Government. *The Scottish Index of Multiple Deprivation 2020*. 2020 [cited 2020 Apr 15]. Available from: <https://www.gov.scot/collections/scottish-index-of-multiple-deprivation-2020/#keydocuments>
37. Campbell L, Wilson J. *Open Scotland*. In: *Open Educational Resources: An equitable future for education in Scotland* [Internet]. 2021 Mar 19. Available from: <https://openscot.net/further-education/open-educational-resources-an-equitable-future-for-education-in-scotland/>
38. HESA. *Higher Education Student Statistics: UK, 2018/19—Subjects studied*. 2020. Available from: <https://www.hesa.ac.uk/news/16-01-2020/sb255-higher-education-student-statistics/subjects>
39. Frink. *Researcher reveals how “Computer Geeks” replaced “Computer Girls”*. *Gender News Stanford University*. 2015. Available from: <https://web.archive.org/web/20150312130252/http://gender.stanford.edu/news/2011/researcher-reveals-how-“computer-geeks”-replaced-“computergirls”>
40. Smith EE. *Recognizing a Collective Inheritance through the History of Women in Computing*. *CLCWeb Comp Lit Cult*. 2013; 15. <https://doi.org/10.7771/1481-4374.1972>
41. Hicks M. *Programmed inequality: how Britain discarded women technologists and lost its edge in computing*. Cambridge, MA: MIT Press; 2017.
42. Thakkar D, Sambasivan N, Kulkarni P, Kalenahalli Sudarshan P, Toyama K. *The Unexpected Entry and Exodus of Women in Computing and HCI in India*. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems—CHI '18*. Montreal QC, Canada: ACM Press; 2018. p. 1–12. <https://doi.org/10.1145/3173574.3173926>
43. Lagesen VA. *A Cyberfeminist Utopia?: Perceptions of Gender and Computer Science among Malaysian Women Computer Science Students and Faculty*. *Sci Technol Hum Values*. 2008; 33:5–27. <https://doi.org/10.1177/0162243907306192>
44. Scottish Qualifications Authority. *Annual Statistical Report 2018*. Scotland: Scottish Qualifications Authority; 2018. Available from: <https://www.sqa.org.uk/sqa/63001.html>