PLOS | COMPUTATIONAL BIOLOGY

EDUCATION

The eBioKit, a stand-alone educational platform for bioinformatics

Rafael Hernández-de-Diego[1], Etienne P. de Villiers[2,3,4], Tomas Klingström[1], Hadrien Gourlé[1], Ana Conesa[5,6], Erik Bongcam-Rudloff[1]*

1 SLU-Global Bioinformatics Centre, Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden, **2** International Livestock Research Institute (ILRI), Nairobi, Kenya, **3** KEMRI-Wellcome Trust Research Programme, Kilifi, Kenya, **4** Centre for Tropical Medicine and Global Health, University of Oxford, Oxford, United Kingdom, **5** Genomics of Gene Expression Lab, Centro de Investigación Príncipe Felipe, Valencia, Spain, **6** Microbiology and Cell Science Department, Institute for Food and Agricultural Sciences, University of Florida, Gainesville, United States of America

* erik.bongcam@slu.se

Citation: Hernández-de-Diego R, de Villiers EP, Klingström T, Gourlé H, Conesa A, Bongcam-Rudloff E (2017) The eBioKit, a stand-alone educational platform for bioinformatics. PLoS Comput Biol 13(9): e1005616. https://doi.org/10.1371/journal.pcbi.1005616

Editor: Francis Ouellette, Genome Quebec, CANADA

Published: September 14, 2017

Funding: The work was supported by H3ABioNet with a grant by the National Institutes of Health Common Fund under grant number U41HG006941 and also supported by ILRI/BECA with a Swedish Ministry for Foreign Affairs through Sida fund. 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.

Abstract

Bioinformatics skills have become essential for many research areas; however, the availability of qualified researchers is usually lower than the demand and training to increase the number of able bioinformaticians is an important task for the bioinformatics community. When conducting training or hands-on tutorials, the lack of control over the analysis tools and repositories often results in undesirable situations during training, as unavailable online tools or version conflicts may delay, complicate, or even prevent the successful completion of a training event. The eBioKit is a stand-alone educational platform that hosts numerous tools and databases for bioinformatics research and allows training to take place in a controlled environment. A key advantage of the eBioKit over other existing teaching solutions is that all the required software and databases are locally installed on the system, significantly reducing the dependence on the internet. Furthermore, the architecture of the eBioKit has demonstrated itself to be an excellent balance between portability and performance, not only making the eBioKit an exceptional educational tool but also providing small research groups with a platform to incorporate bioinformatics analysis in their research. As a result, the eBioKit has formed an integral part of training and research performed by a wide variety of universities and organizations such as the Pan African Bioinformatics Network (H3ABioNet) as part of the initiative Human Heredity and Health in Africa (H3Africa), the Southern Africa Network for Biosciences (SAnBio) initiative, the Biosciences eastern and central Africa (BecA) hub, and the International Glossina Genome Initiative.

This is a *PLOS Computational Biology* Education paper.

Introduction

High throughput technologies and next generation sequencing require the development of new methods to manage the data generated by researchers. It is therefore imperative that

training in bioinformatics is available to educate experts as well as other researchers in order to allow them to plan their research and properly assert the true cost and effort to complete a project [1].

In developing countries, bioinformatics has been a strategic investment for many countries due to its positive contributions to other fields of life science as well as the comparatively low costs of the discipline. Equipping and running a bioinformatics teaching laboratory cost less than equipping and running a biology laboratory [2] and many developing or formerly socialist countries have access to trained professionals in advanced mathematics and/or computational science [3], which form the basis of the field when combined with biology. From a political perspective, an enhanced capacity in bioinformatics allows researchers to conduct advanced analysis inside the country to ensure that the immaterial property rights are retained within the country and can support the development of a national life science industry [3, 4, 5].

Initial efforts in developing countries have generated numerous hubs of excellence located in the bigger or more affluent countries, but smaller countries are following suit [6]. Furthermore, international networks such as H3ABioNet [7] are developing a network of expert hubs for bioinformaticians collaborating with each other to strengthen international collaboration in developing countries [8]. Extensive training is, however, necessary to provide to the research professionals necessary to populate these networks and analyze the virtual mountains of data generated by modern research [9, 10].

The major challenges towards the creation or expansion of viable communities of bioinformaticians vary across the world based on the available resources and priorities within the education system. In the Western world, the chief concern regarding bioinformatics is a lack of trained professionals within the field who can conduct research and/or maintain infrastructures [11]. In the Asia-Pacific region, recruitment to the field is regarded as less of a problem, as young researchers perceive the field as an attractive career choice. Instead, chief concerns relate to the comparative lack of infrastructure in many countries [12]. These differences are also evident when comparing key development indicators for communications technology and research. Several of the most highly developed countries in East Asia are competitive with European nations in regards to the number of researchers per million people and student-to-teacher ratios in higher education. But only Japan and Singapore rank above the European median regarding the number of secure servers and fixed broadband subscriptions per capita (Fig 1). Furthermore, several of the nations in the region may place low on population-adjusted metrics but can still provide a high-quality infrastructure for universities and the growing middle class.

Such clusters of high capacity are, however, significantly rarer in South America and Africa [13] (Fig 1), which makes capacity building significantly more challenging, as logistics become a significant challenge when planning training sessions. Unreliable internet access, few local teachers, and a lack of suitable students are common issues and it is therefore important that training sessions are not delayed or disrupted, as the number of training opportunities involving international experts is limited. This makes it important that bioinformatics training in Africa is carefully planned and that measures are taken to ensure access to infrastructure suitable for bioinformatics [10]. As a result, African networks such as the H3ABioNet need to rely on using creative approaches to overcome these issues by seeking low latency alternatives and using portable devices that host data and tools and run independently of the network [14]. Key performance indicators provided by the World Bank DataBank [15] (Fig 1) and other resources indicate that internet connectivity [16, 17] as well as internet infrastructure are improving at a rapid rate in developing countries. Access to trained personnel in the form of researchers, technicians, and teachers is, however, increasing at a lower rate, indicating that

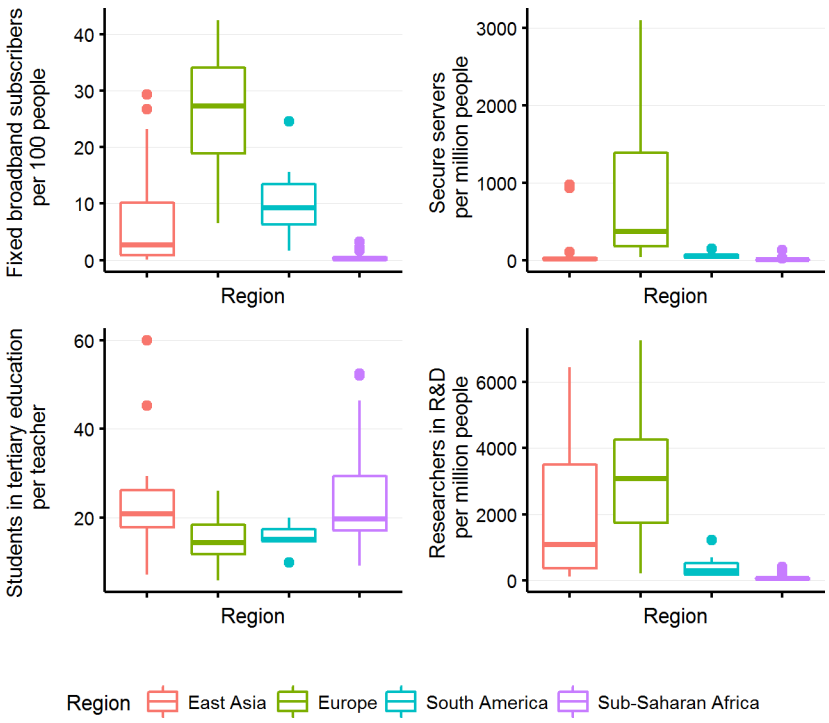

Fig 1. Box plots displaying access to technology and expertise in East Asia, Europe, South America, and sub-Saharan Africa. All data is calculated from the World Bank DataBank [15] (see S1 Table for for full data). The vertical lines extend from the 25th and 75th percentiles to the lowest/highest value that is within 1.5 times the distance between the 25th and 75th percentiles (the interquartile range). Data beyond the end of the lines are outliers and plotted as points. Data for fixed broadband internet subscribers (per 100 people) are per 2014; the number of secure servers by 2014 and other data is from a range of years from 2010 to 2015, depending on data availability (see S2 Table for summarized data per country).

https://doi.org/10.1371/journal.pcbi.1005616.g001

even as internet connectivity technology improves, international support in the form of education and training will remain important.

The eBioKit was first developed in 2007 as a response to the lack of reliable and sufficient internet connections and the short time available to visiting researchers for conducting hands-on training at workshop or short courses. In many cases, utilization of large public databases of biomolecular data by the course software is required and valuable time is lost configuring locally provided computers to participants of the course. Furthermore, unforeseen delays are common even when a sufficient internet connection is available, as remote servers might be suddenly down or new software versions have been released that make on-site exercises fail or give different results than expected. This is an important burden, especially in these short courses provided by external lecturers, who have limited time to go over the teaching material.

Having experienced the difficulties for running the on-site courses in many world locations, it became clear that a solution was needed that would facilitate teaching without dependence on the internet and instabilities of the software. This has motivated the development of the eBioKit, a portable device for bioinformatics training (example installation: http://www.ebiokit.eu). In this paper, we present this teaching resource; describe the architecture, contents, and utilization of the system; and illustrate several projects in which the tool has been successfully used for teaching as well as research.

Materials and methods

Content of the platform

The eBioKit is a portable bioinformatics educational platform, the main purpose of which is to significantly reduce dependence on the internet by offering locally a wide range of services and repositories widely used in genomic research as well as documentation and material for training in their use (Fig 2). Local availability and portability are key elements in making the eBioKit an excellent educational tool in places with limited infrastructure.

There are 3 basic types of content in the eBioKit: databases, software, and tutorials. Research for both human and nonhuman model organisms is supported by the inclusion of Ensembl Mammals and Ensembl Genomes [18, 19], as both biomedical and environmental research is frequently relevant at targeted teaching locations. Protein functional analysis is supported by the inclusion of databases such as UniProtKB/Swiss-Prot [20], UniProtKB/TrEMBL [20], protein family database (Pfam) [21], and the Protein Data Bank [22]. In addition to databases, tools for sequencing homology search, protein structure prediction, next-generation sequencing (NGS) data analysis, functional annotation, and genome-wide association studies (GWASs) among others are included to give support to any bioinformatics discipline. Some of the services installed on the latest version of the eBioKit are listed in Table 1, and a more complete description can be found at http://www.ebiokit.eu/information. Moreover, many other popular bioinformatics tools and software utilities are also distributed on the eBioKit as downloadable resources or as part of comprehensive collections of analysis tools such as the Chipster platform [23] or generic model organism database (GMOD) in a Box [24], both available as virtual machine images.

User interface

The eBioKit is usually installed as a centralized service on the local network. Students can connect to the system by accessing to the local internet protocol (IP) address or a known uniform resource locator (URL) assigned to the eBioKit (Fig 2). The eBioKit has 2 basic access modes. The most common way is using a web browser of choice present in the student's computer. As depicted in Fig 3, the eBioKit website is divided into 2 main parts, the working area and the applications menu. Using this lateral menu, the students can switch between the installed tools and databases, and the content of the working area will be adapted to the selected service.

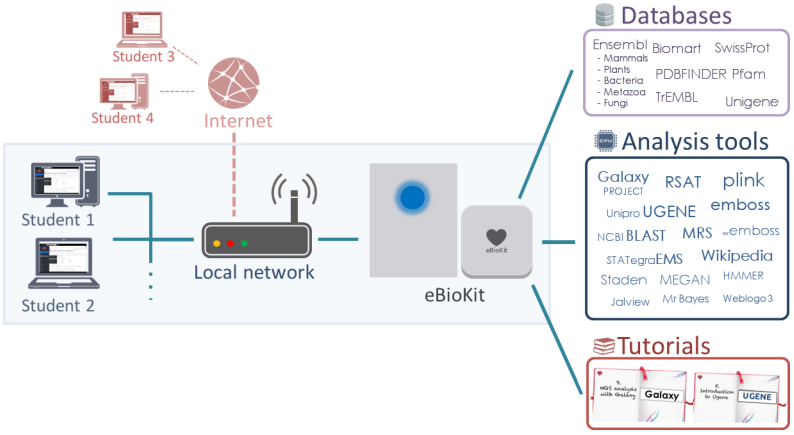

Fig 2. A typical installation of the eBioKit on a local network. Students and researchers can access the tools, databases, and tutorials installed in the eBioKit using the assigned local internet protocol (IP) address or a known uniform resource locator (URL). The eBioKit includes some administration tools that simplify the setting up of the platform on new networks. Additionally, if network configuration allows, the eBioKit services can also be accessed from external networks.

https://doi.org/10.1371/journal.pcbi.1005616.g002

Table 1. A selection of services installed on the eBioKit.

Service name	Type	Access				Version
		CL	DW	WB	GT	
Chipster [23]	Image		×			-
EMBOSS [25]	Tool	×			×	6.6.0
Ensembl[1] [18, 19]	Database			×		Releases 75 and 22
BioMart [26]	Database			×		0.7
Galaxy [27]	Tool			×		July 2014
GMOD in a Box [24]	Image		×			2.05
Jalview [28]	Tool	×	×	×		2.8.1
MRS [29]	Database	×		×		6.0.3
NCBI Blast[2] [30]	Tool	×		×	×	2.2.26 and 2.2.29+
PLINK [31]	Tool		×	×		v1.07
RSAT [32]	Tool	×	×	×		October 2014
STATegra EMS [33]	Tool			×		0.6r1
WebApollo [34]	Tool			×		v1.0.3
wEMBOSS [35]	Tool			×		v2.2.1

[1]Ensembl Mammals (release 75), Bacteria (release 22 [r22]), Fungi (r22), Metazoa (r22), Plants (r22) and Protists (r22).

[2]NCBI Blast and NCBI Blast+.

Abbreviations: CL, command-line; DW, downloadable resource; EMBOSS, the European Molecular Biology Open Software Suite; EMS, Experiment Management System; GMOD, Generic Model Organism Database; GT, Galaxy tool; MRS, Maarten's Retrieval System; NCBI, National Center for Biotechnology Information; RSAT, Regulatory Sequence Analysis Tools; WB, web-based; wEMBOSS, web-interface to EMBOSS

https://doi.org/10.1371/journal.pcbi.1005616.t001

Alternatively, students can connect via command-line interface using Secure Shell (ssh) on a terminal. This also gives the opportunity to train on command-line analysis tools that are not available with a graphical user interface and allows for flexibility in the definition of course contents and competences by the instructor, who can choose to include programming modules in the course material or simply teach web-/tool-based bioinformatics.

Teaching material

Tutorials are a fundamental part of the eBioKit and are hosted on an e-learning platform in a unified environment, ensuring a cohesive learning experience (Fig 4). The included tutorials range from basic bioinformatics concepts to advanced topics such as high throughput sequencing analysis or GWAS. Tutorials are organized in courses, which are divided into lessons that usually correspond to a specific task for the student to accomplish such as building a reference genome or manipulating a dataset (Fig 5). All the required software and databases for each lesson are locally available, and data is often adapted to allow students to perform their analysis in a timely manner.

Most of the tutorials included in the eBioKit have been written and refined by our instructors over the years, but a special effort is being made to acquire new content from the community and adapt it for inclusion in the eBioKit. Tutorials are written in Markdown, a lightweight markup language that allows creating styled documentation, and most of them are available on the web-based Git repository GitHub for anyone to modify or reuse [36].

System administration

An important aspect for the reliable operation of the eBioKit is the administration of the system. As usual in this field, the administrators for the eBioKit must ensure the proper

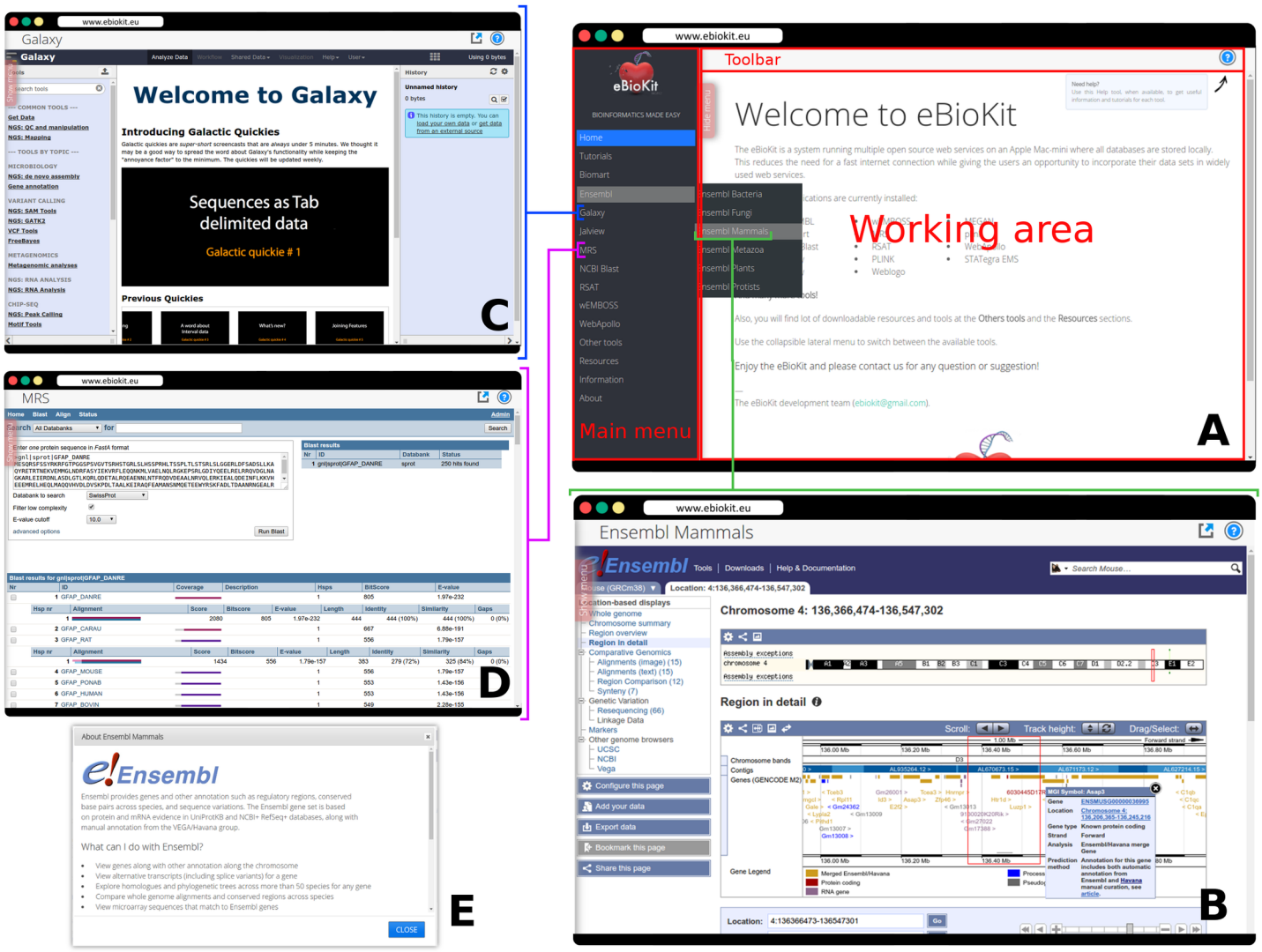

Fig 3. Web interface for the eBioKit. (A) The arrangement of the components that compose the web interface. Users can easily switch between the installed services on the eBioKit using the lateral menu. When users choose an option on the menu, the working area is replaced by the corresponding service and the menu is hidden, allowing users to fully interact with the service. (B), (C), and (D) show the familiar web interface users see when working with Ensembl Mammals, Galaxy, and MRS, respectively, in eBioKit. With the upper toolbar, users can open the services on a secondary window and, more importantly, can get a description as well as download documentation and tutorials for the selected service (E).

https://doi.org/10.1371/journal.pcbi.1005616.g003

functioning of the installed services as well as provide users with support and keep the system updated and secure. For an easier administration, the eBioKit includes several tools that simplify some usual tasks in the management of services and users. These administration tools, which can be individually executed as command-line programs, are compiled in a Java application named "eBioKit Control Panel" that provides a user-friendly interface both as a desktop-based application and as a command-line application (Fig 6A and 6B). Moreover, an online help desk portal is maintained, where the administrators can get support directly from the developers as well as share their experiences or suggestions and find documentation, news, and other useful information related with the administration of the eBioKit (Fig 6C).

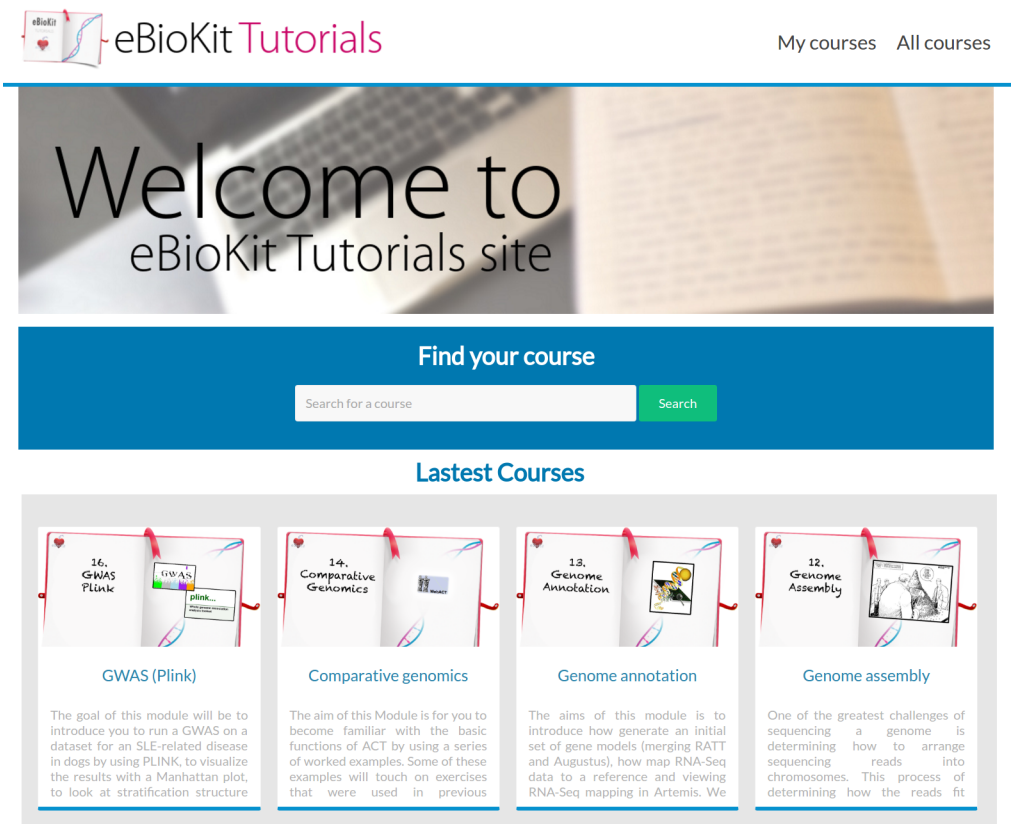

Fig 4. Entry page for the training portal in the eBioKit.

https://doi.org/10.1371/journal.pcbi.1005616.g004

Selection of hardware

Both computational and space requirements of many key bioinformatics tools are heavy and this turns portability into a complex objective to achieve. To address this issue, the eBioKit system has been historically built on Apple Mac Mini machines, which accomplish a brilliant balance of portability (the dimensions of the latest model are 197 x 197 x 36 mm and 1.2 kg of weight), computational capacity (up to 3.0 GHz dual-core Intel Core i7 and 16 gigabytes [GB] of random access memory [RAM]in the latest models) [37], and quality and reliable hardware.

In addition to the Mac Mini version, an alternative architecture using Mac Pro machines is available, slightly reducing the portability of the system (251 mm height, 167 mm diameter, and 5 kg of weight) but dramatically increasing the computational power (up to 3.5 GHz six-core Intel Xeon E5 and 64 GB in the latest models) [38].

Storage supposes an added difficulty for portability and performance. The sizes of the biological resources installed on the eBioKit, such as the Ensembl Mammals databases, are in the range of several terabytes and increase with each new release. Nowadays, it is becoming easier to find on the market external storages in the multiple-terabytes range, most of them based on universal serial bus (USB) v3.0. For the eBioKit, the chosen storage solution was the LaCie 5big Thunderbolt (10 terabytes [TB] RAID0, 7,200 rpm, 173 x 220 x 196 mm and 9.9 kg), which takes advantage of the Thunderbolt port available on Mac machines (both in Mac Mini and Mac Pro versions), achieving a transfer rate of up to 700 megabytes [MBs] for read and write operations, independently [39] (Fig 7).

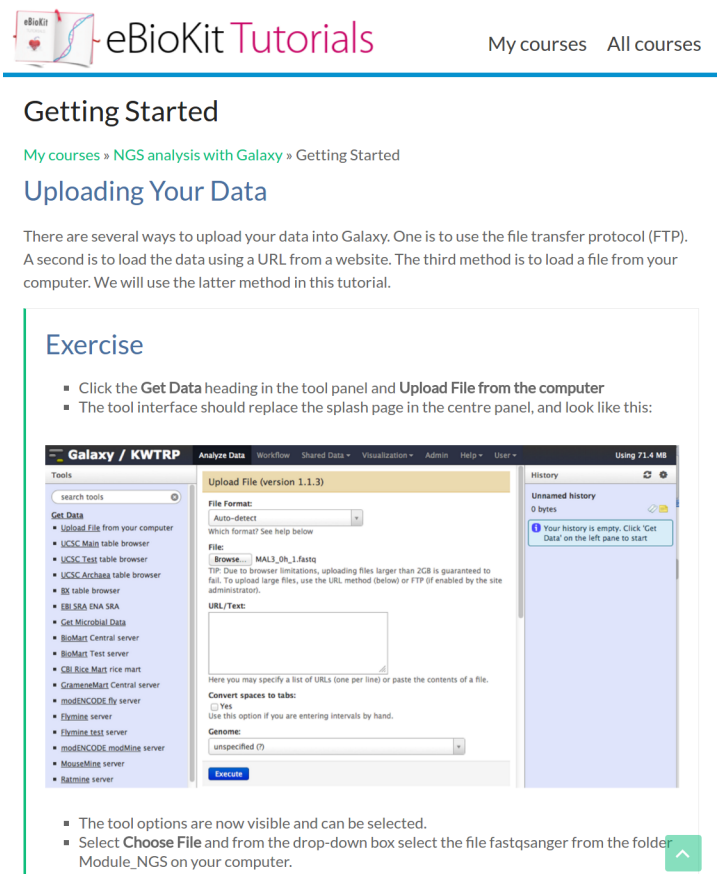

Fig 5. An example of a lesson in the eBioKit training portal. The image displays an extract of the "Getting started" lesson for the tutorial "next-generation sequencing (NGS) analysis with Galaxy." During a tutorial, the students will find multiple exercises that allow them to put into practice the content learned.

https://doi.org/10.1371/journal.pcbi.1005616.g005

Results

The eBioKit is distributed as an affordable and self-contained computing platform and database system containing up to 6 terabytes of biological data and software tools of relevance to bioinformatics researching, including the Ensembl database systems [18, 19], the European Molecular Biology Open Software Suite (EMBOSS) [25], Galaxy [27], National Center for Biotechnology Information (NCBI) Blast [30], and PLINK [31], which are made locally available through a unified web-based user interface.

From a teaching perspective, almost each tool or database installed in the platform includes a tutorial that introduces to its use. A total of 13 courses are currently available in the eBioKit. Courses encompass a wide range of bioinformatics disciplines ranging from basic bioinformatics tasks, the UNIX environment, and programming, to more advanced topics such as GWAS, RNA sequencing (RNA-Seq) analysis, genome assembly and annotation, and comparative genomics. S3 Table summarizes the content and the structure for the included courses.

A total of 24 training activities have been successfully organized during the last years with the help of the eBioKit in different research centres and universities across Europe, Africa, Asia, and South America in collaboration with international organisms such as the Pan African Bioinformatics Network (H3ABioNet) as part of the H3Africa initiative [15], the Southern Africa Network for Biosciences (SAnBio) [40], the Biosciences eastern and central Africa-

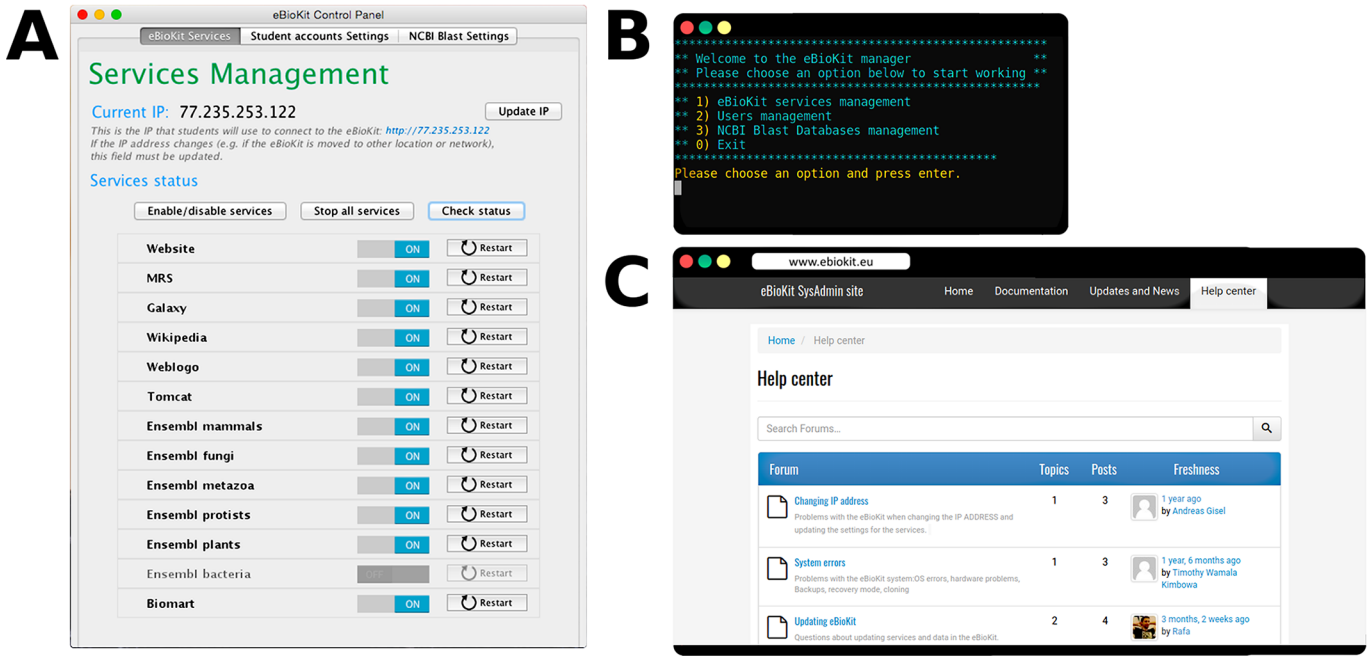

Fig 6. Tools for eBioKit administrators. (A) The eBioKit Control Panel desktop application. At the image, the section for services management displays the state for the installed services and includes some options for their configuration. Other sections provide access to user management and for NCBI Blast databases manipulation. This control panel is also available as a command-line utility, ideal for remote administration (B). Finally, the online help desk (C) contains news and updates for the eBioKit and introduces a communication channel between administrators and the eBioKit developers.

https://doi.org/10.1371/journal.pcbi.1005616.g006

International Livestock Research Institute (BecA-ILRI) hub [41], and the International Glossina Genome Initiative [42]. Moreover, the system has been acquired by many of those institutions as part of their computing facilities (Fig 8), allowing researchers to conduct bioinformatics-based research without having access to a reliable internet connection.

Concerning system performance, the version of the eBioKit built on Apple Mac Mini machines with 16 GB of RAM has been successfully used for courses with up to 25 students

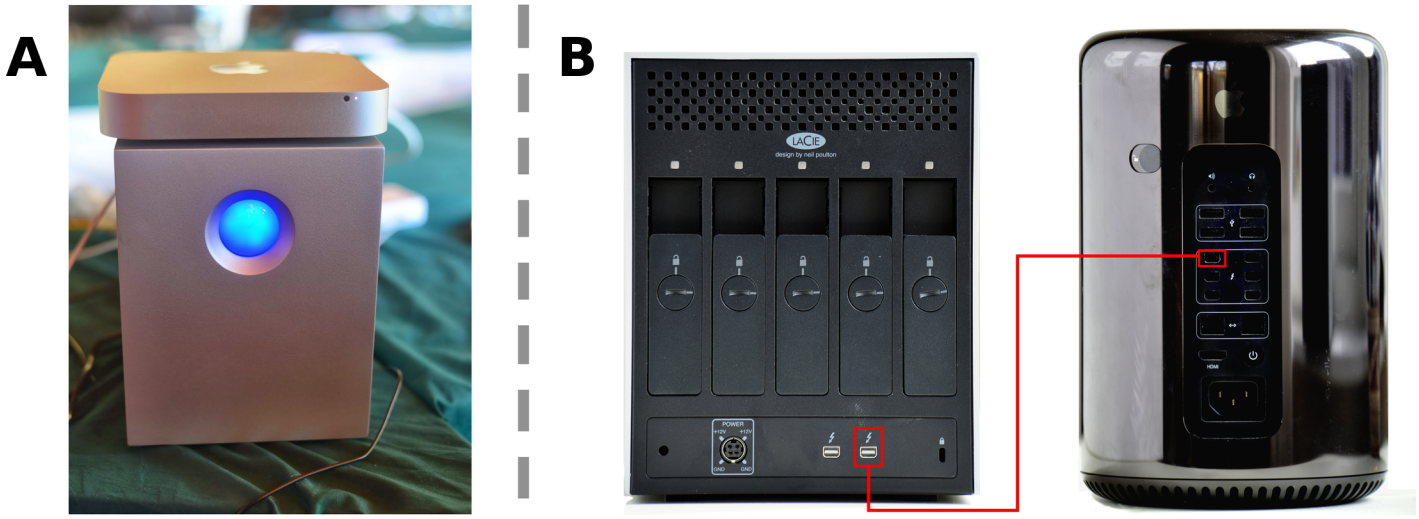

Fig 7. The 2 variants for the architecture of the eBioKit. (A) An eBioKit built on a Mac Mini with a 10-terabyte (TB) LaCie 5big Thunderbolt hard disk. (B) Detail for the Thunderbolt connection for the eBioKit Mac Pro version.

https://doi.org/10.1371/journal.pcbi.1005616.g007

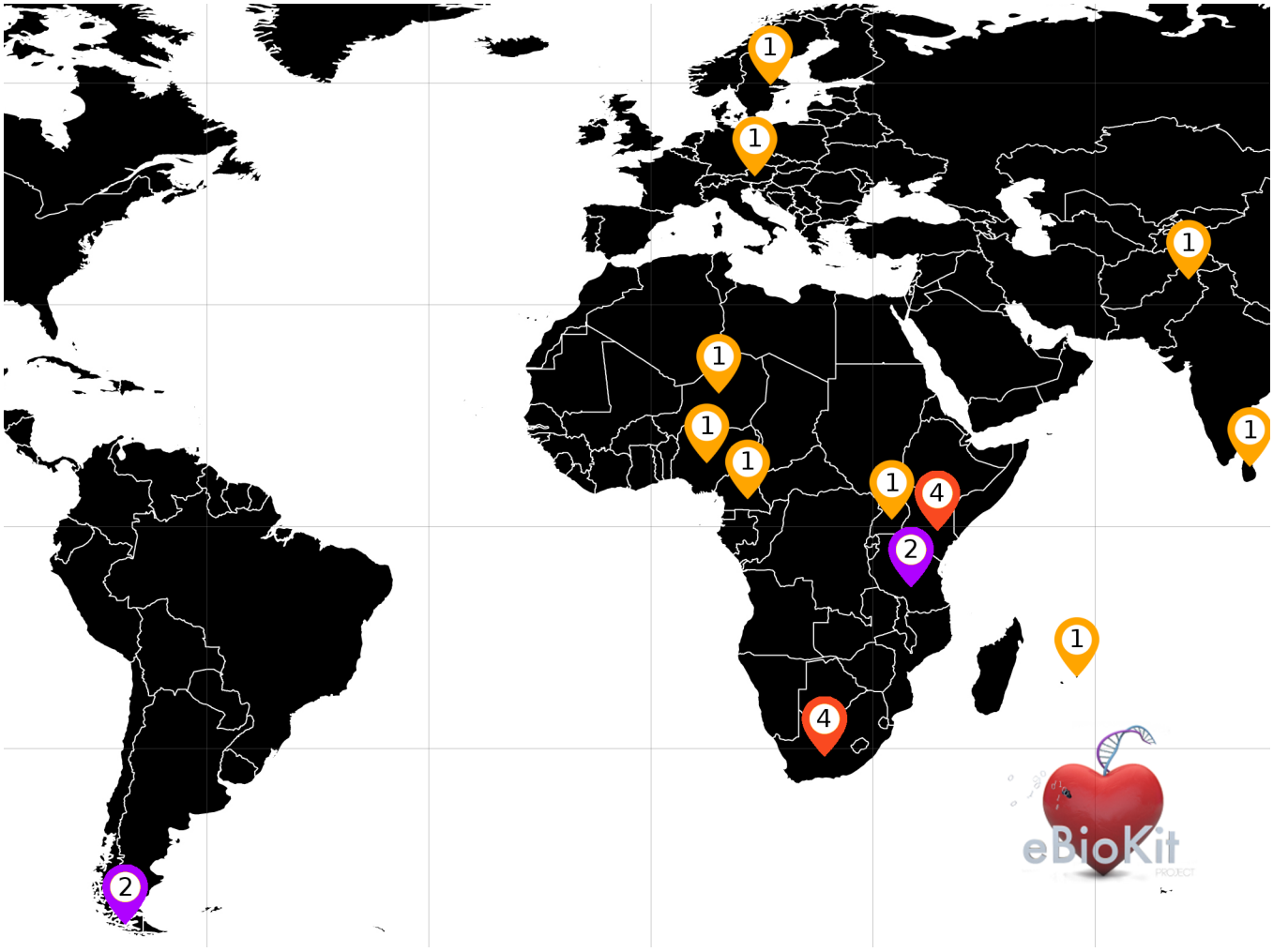

Fig 8. Global distribution for the eBioKit project. Numbers indicate the number of eBioKits deployed in each country. European institutions: Swedish University of Agricultural Sciences (SLU) (Sweden) and Medizinische Universität (Austria). South America: Instituto Antártico Chileno (INACH) and Universidad de Magallanes (Chile). Asian institutions: University of Colombo (Sri Lanka) and COMSATS (Pakistan). Africa: Centre de Recherche Medicale et Sanitaire (Niger), The International Institute of Tropical Agriculture (IITA) (Nigeria), University of Buea (Cameroon), Uganda Virus Research Institute (Uganda), BECA/ILRI (Kenya), Pwani University (Kenya), International Centre of Insect Physiology and Ecology (Kenya), Technical University of Kenya (Kenya), University of Dar es Salaam (Tanzania), Mikocheni Agricultural Research Institute (Tanzania), University of Mauritius (Mauritius), University of the Witwatersrand (South Africa), University of Pretoria (South Africa), University of Cape Town (South Africa), and SANBI South African National Bioinformatics Institute (South Africa).

https://doi.org/10.1371/journal.pcbi.1005616.g008

working in parallel on the courses included in the platform. This version may not be recommended for large NGS analysis work. On the other hand, the Mac Pro–based version with 32 GB RAM allows up to 40 students to attend to a course and work simultaneously. In spite of the system having not been benchmarked for intensive work yet, it is known that the system is routinely used as an analytical resource by some of the project partners.

Discussion

The key benefit of the eBioKit is that it provides a controlled and reliable environment for bioinformatics. Tools and databases are constantly updated as research in the field progresses and results as well as user interfaces may change as new versions are released. For a researcher conducting training, this may present major issues as students are confused or critical components

fail, which prevent the students from properly completing their tasks. Unfortunately, the costs of failure in training are the highest in the areas that can least afford to pay them.

In a developed region with access to local teachers and high-quality infrastructure, replacements and extra trainers can in most cases be brought in to complete the training goals. In a resource-constrained setting, teachers are often brought in from afar, which puts strict time limits on training as tickets are booked in advance, repairs of infrastructure often take longer, and access to local experts who can quickly solve issues or help complete the training is not always available.

The eBioKit is based on standardized and compact hardware, which makes it easy for trainers to prepare in advance. As all software is either open source or at least free for academic use, a trainer can, when necessary, purchase the necessary hardware and clone the eBioKit content, copying all necessary tools and data to the new server. The server is then brought along by the trainer to the training location and the students access the eBioKit through the local area network and work directly on the server, which avoids installation issues, unforeseen updates of web services, or failures in the local infrastructures. Upon completion of the training, the eBioKit can then either be brought back for future training sessions or left behind for use by local researchers or trainers.

Conclusions

Bioinformatics has gradually established itself as an essential discipline for many life sciences and the consequent demand of qualified researchers has boosted the emergence of new educational approaches. Providing training in bioinformatics is challenging from many perspectives. The growth in the volume of biological data, the multidisciplinary skills required for students, and the necessary computing infrastructure as well as the constant development of methods and tools are some of the hurdles that must be tackled when setting up and maintaining an effective teaching infrastructure.

Several solutions have been developed in the last years for educational purposes in bioinformatics, but many of them demonstrate a lack of sustainability and a strong dependence on the internet, computing capacity, and third-party services, which usually lead to outdated tools and frustrating errors. We have developed a stand-alone and portable educational platform that allows the deployment of new educational resources together with the bioinformatics tools and databases needed for sustainable reuse of the teaching materials.

The eBioKit has been conceived to be a robust, user-friendly, and easy-to-manage teaching tool for courses and workshops and has demonstrated itself to be a valuable resource for institutions, universities, research centers, or schools that would like to start teaching bioinformatics or even provide bioinformatics capabilities for their groups. The platform is based on open source and open access licenses that ensure its availability and distribution and can be ordered directly to the developer team at no cost except those derived from the purchase of the necessary hardware to run the system (i.e., the Mac Mini or Mac Pro machines) and transportation. The advantage of the eBioKit as a training platform is the fact that it has self-contained courses and tutorials, teaching both basic and advanced bioinformatics using software and databases installed locally on the platform.

The eBioKit is a live project in constant development, providing a responsive support for users and administrators as well as inspiring other projects [43, 44]. Each iteration of the project is, however, functioning as a stable stand-alone platform, allowing researchers to teach and use the platform without compatibility issues. This allows researchers to conduct projects and training sessions without spending valuable time or resources on recreating a functioning environment each time a new course or project is initiated. More information about how to

order an eBioKit and how to contribute to the project as well as other frequently asked questions and tools for contacting the eBioKit team can be found at http://www.ebiokit.eu.

Supporting information

S1 Table. Worldwide distribution for internet access, access to secure internet servers, pupil-teacher ratio, researchers in research and development (R&D), and technicians in R&D by country. Source: The World Bank databank and others [13, 16, 17].
(XLSX)

S2 Table. Summarized worldwide distribution for internet access, access to secure internet servers, pupil-teacher ratio, researchers in research and development (R&D), and technicians in R&D by country. Source: The World Bank databank and others [13, 16, 17].
(XLSX)

S3 Table. Overview of the included courses in the eBioKit. Each course in the eBioKit comprises several lessons, which cover popular topics in bioinformatics analysis and introduce the students to the usage of the software and databases locally installed.
(DOCX)

Acknowledgments

We thank the EMBnet, H3Abionet, SANBio, and BECA/ILRI communities and members for valuable discussions and for providing logistics and resources. This paper is published with the permission of the Director of KEMRI.

References

1. Sboner A, Mu XJ, Greenbaum D, Auerbach RK, Gerstein MB. The real cost of sequencing: higher than you think! Genome Biol. 2011; 12(8):125. https://doi.org/10.1186/gb-2011-12-8-125 PMID: 21867570
2. Counsell D. A review of bioinformatics education in the UK. Brief Bioinform. 2003 Mar; 4(1):7–21. Review. PMID: 12715830.
3. Bujnicki JM, Tiuryn J. Bioinformatics and computational biology in Poland. PLoS Comput Biol. 2013; 9 (5):e1003048. https://doi.org/10.1371/journal.pcbi.1003048 PMID: 23658507
4. Kulkarni-Kale U, Sawant S, Chavan V. Bioinformatics education in India. Brief Bioinform. 2010 Nov; 11 (6):616–25. https://doi.org/10.1093/bib/bbq027 PMID: 20705754
5. Motari M, Quach U, Thorsteinsdóttir H, Martin DK, Daar AS, Singer PA. South Africa—blazing a trail for African biotechnology. Nat Biotechnol. 2004 Dec; 22 Suppl:DC37–41.
6. Karikari TK. Bioinformatics in Africa: The Rise of Ghana? PLoS Comput Biol. 2015 Sep 17; 11(9): e1004308. https://doi.org/10.1371/journal.pcbi.1004308 PMID: 26378921
7. Mulder NJ, Adebiyi E, Alami R, Benkahla A, Brandful J, Doumbia S, et al. H3ABioNet Consortium. H3ABioNet, a sustainable pan-African bioinformatics network for human heredity and health in Africa. Genome Res. 2016 Feb; 26(2):271–7. https://doi.org/10.1101/gr.196295.115 PMID: 26627985
8. Adoga MP, Fatumo SA, Agwale SM. H3Africa: a tipping point for a revolution in bioinformatics, genomics and health research in Africa. Source Code Biol Med. 2014 May 8; 9:10. https://doi.org/10.1186/1751-0473-9-10 PMID: 24829612
9. Schneider MV, Watson J, Attwood T, Rother K, Budd A, McDowall J, et al. Bioinformatics training: a review of challenges, actions and support requirements. Brief Bioinform. 2010 Nov; 11(6):544–51. https://doi.org/10.1093/bib/bbq021 PMID: 20562256
10. Via A, Blicher T, Bongcam-Rudloff E, Brazas MD, Brooksbank C, Budd A, et al. Best practices in bioinformatics training for life scientists. Brief Bioinform. 2013 Sep; 14(5):528–37. https://doi.org/10.1093/bib/bbt043 PMID: 23803301
11. Chang J. Core services: Reward bioinformaticians. Nature. 2015 Apr 9; 520(7546):151–2. https://doi.org/10.1038/520151a PMID: 25855439

12. Ranganathan S, Schönbach C, Nakai K, Tan TW. Challenges of the next decade for the Asia Pacific region: 2010 International Conference in Bioinformatics (InCoB 2010). BMC Genomics. 2010 Dec 2; 11 Suppl 4:S1. https://doi.org/10.1186/1471-2164-11-S4-S1 PMID: 21143792

13. Chimusa ER, Mbiyavanga M, Masilela V, Kumuthini J. "Broadband" Bioinformatics Skills Transfer with the Knowledge Transfer Programme (KTP): Educational Model for Upliftment and Sustainable Development. PLoS Comput Biol. 2015 Nov 19; 11(11):e1004512. https://doi.org/10.1371/journal.pcbi.1004512 PMID: 26583922

14. H3Africa Consortium, Rotimi C, Abayomi A, Abimiku A, Adabayeri VM, Adebamowo C, et al. Enabling the genomic revolution in Africa. Science. 2014 Jun 20; 344(6190):1346–8. https://doi.org/10.1126/science.1251546 PMID: 24948725

15. United Nations Educational, Scientific, and Cultural Organization (UNESCO) Institute for Statistics [cited 27 April 2017]. In: Researchers in R&D (per million people) [Internet]. Available from: http://data.worldbank.org/indicator/SP.POP.SCIE.RD.P6

16. The International Telecommunication Union [cited 27 April 2017]. In: The ITU ICT Facts and Figures 2016 [Internet]. Available from: http://www.itu.int/en/ITU-D/Statistics/Pages/facts/default.aspx

17. Akamai Technologies [cited 27 April 2017]. In: State of the Internet / Connectivity Trends Report for Q4 2016 [Internet]. Available from: https://content.akamai.com/pg8231-q4-2016-soti-connectivity-report-uk.html

18. Flicek P, Amode MR, Barrell D, Beal K, Billis K, Brent S, et al. Ensembl 2014. Nucleic Acids Res. 2014 Jan; 42(Database issue):D749–55. https://doi.org/10.1093/nar/gkt1196 PMID: 24316576

19. Kersey PJ, Allen JE, Armean I, Boddu S, Bolt BJ, Carvalho-Silva D, et al. Ensembl Genomes 2016: more genomes, more complexity. Nucleic Acids Res. 2016 Jan 4; 44(D1):D574–80. https://doi.org/10.1093/nar/gkv1209 PMID: 26578574

20. UniProt Consortium. UniProt: a hub for protein information. Nucleic Acids Res. 2015 Jan; 43 (Database issue):D204–12. https://doi.org/10.1093/nar/gku989 PMID: 25348405

21. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 2016 Jan 4; 44(D1):D279–85. https://doi.org/10.1093/nar/gkv1344 PMID: 26673716

22. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res. 2000 Jan 1; 28(1):235–42. PMID: 10592235

23. Kallio MA, Tuimala JT, Hupponen T, Klemelä P, Gentile M, Scheinin I, et al. Chipster: user-friendly analysis software for microarray and other high-throughput data. BMC Genomics. 2011 Oct 14; 12:507. https://doi.org/10.1186/1471-2164-12-507 PMID: 21999641

24. The GMOD project [cited 27 April 2017]. In: The Generic Model Organism Database, overview [Internet]. Available from: http://gmod.org/wiki/Overview

25. Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. 2000 Jun; 16(6):276–7. PMID: 10827456

26. Smedley D, Haider S, Durinck S, Pandini L, Provero P, Allen J, et al. The BioMart community portal: an innovative alternative to large, centralized data repositories. Nucleic Acids Res. 2015 Jul 1; 43(W1): W589–98. https://doi.org/10.1093/nar/gkv350 PMID: 25897122

27. Afgan E, Baker D, van den Beek M, Blankenberg D, Bouvier D, Čech M, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res. 2016 Jul 8; 44(W1):W3–W10. https://doi.org/10.1093/nar/gkw343 PMID: 27137889

28. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009 May 1; 25(9):1189–91. https://doi.org/10.1093/bioinformatics/btp033 PMID: 19151095

29. Hekkelman ML, Vriend G. MRS: a fast and compact retrieval system for biological data. Nucleic Acids Res. 2005 Jul 1; 33(Web Server issue):W766–9. https://doi.org/10.1093/nar/gki422 PMID: 15980580

30. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990 Oct 5; 215(3):403–10. https://doi.org/10.1016/S0022-2836(05)80360-2 PMID: 2231712

31. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007 Sep; 81(3):559–75. https://doi.org/10.1086/519795 PMID: 17701901

32. Medina-Rivera A, Defrance M, Sand O, Herrmann C, Castro-Mondragon JA, Delerce J, et al. RSAT 2015: Regulatory Sequence Analysis Tools. Nucleic Acids Res. 2015 Jul 1; 43(W1):W50–6. https://doi.org/10.1093/nar/gkv362 PMID: 25904632

33. Hernández-de-Diego R, Boix-Chova N, Gómez-Cabrero D, Tegner J, Abugessaisa I, Conesa A. STATegra EMS: an Experiment Management System for complex next-generation omics experiments. BMC Syst Biol. 2014; 8 Suppl 2:S9. https://doi.org/10.1186/1752-0509-8-S2-S9 PMID: 25033091

34. Lee E, Helt GA, Reese JT, Munoz-Torres MC, Childers CP, Buels RM, Stein L, Holmes IH, Elsik CG, Lewis SE. Web Apollo: a web-based genomic annotation editing platform. Genome Biol. 2013 Aug 30; 14(8):R93. https://doi.org/10.1186/gb-2013-14-8-r93 PMID: 24000942

35. Sarachu M, Colet M. wEMBOSS: a web interface for EMBOSS. Bioinformatics. 2005 Feb 15; 21 (4):540–1. https://doi.org/10.1093/bioinformatics/bti031 PMID: 15388516

36. Gourlé H, Karlsson O, Ohlsson J, Bongcam-Rudloff E. [cited 27 April 2017]. Repository for eBioKit tutorials. Available from https://github.com/eBioKit/tutorials. https://doi.org/10.5281/zenodo.545765

37. Apple Inc. [cited 27 April 2017]. In: Mac Mini—Technical Specifications [Internet]. Available from: https://www.apple.com/mac-mini/specs/

38. Apple Inc. [cited 27 April 2017]. In: Mac Pro—Technical Specifications [Internet]. Available: https://www.apple.com/mac-pro/specs/

39. LaCie Inc. [cited 27 April 2017]. In: LaCie 5big Thunderbolt—Technical Specifications [Internet]. Available from: http://www.lacie.com/products/thunderbolt/5big-thunderbolt-2/

40. Fuxelius HH, Bongcam-Rudloff E, Jaufeerally Y. The contribution of the eBioKit to Bioinformatics Education in Southern Africa. EMBnet.Journal. 2010; 16(1):29–30. https://doi.org/https://doi.org/10.14806/ej.16.1.173

41. de Villiers EP, Bongcam-Rudloff E. eBioKit bioinformatics workshops in Dar es Salaam, Tanzania. EMBnet.Journal- 2014; 20:e755. https://doi.org/https://doi.org/10.14806/ej.20.0.755

42. Christoffels A, Masiga D, Berriman M, Lehane M, Touré Y, Aksoy S. International glossina genome initiative 2004–2014: a driver for post-genomic era research on the African continent. PLoS Negl Trop Dis. 2014 Aug 21; 8(8):e3024. https://doi.org/10.1371/journal.pntd.0003024 PMID: 25144472

43. Klingström T, Mendy M, Meunier D, Berger A, Reichel J, Christoffels A, et al. Supporting the development of biobanks in low and medium income countries. 2016 IST-Africa Week Conference, Durban. 2016:1–10. https://doi.org/10.1109/ISTAFRICA.2016.7530672

44. Tastan Bishop Ö, Adebiyi EF, Alzohairy AM, Everett D, Ghedira K, Ghouila A, et al. Bioinformatics education—perspectives and challenges out of Africa. Brief Bioinform. 2015 Mar; 16(2):355–64. https://doi.org/10.1093/bib/bbu022 PMID: 24990350