Open Institute of the African BioGenome Project: Bridging the gap in African biodiversity genomics and bioinformatics

Citation for published version:
Sharaf, A, Ndiribe, CC, Omotoriogun, TC, Abueg, L, Badaou, B, Markey, FJB, Beedessee, G, Diouf, D, Duru, VC, Ebuzome, C, Eziuzo, SC, Jaufeerally Fakim, Y, Formenti, G, Ghanm, N, Guerfali, FZ, Houaga, I, Ideozu, JE, Katee, SM, Khayi, S, Kuja, JO, Kwon-Ndung, EH, Marks, RA, Moila, AM, Mungloo-Dilmohamud, Z, Muzemi, S, Negussie, H, Osuji, JO, Ras, V, Tchiechoua, YH, Zoclanclounon, YAB, Tolley, KA, Ziyomo, C, Mapholi, N, Muigai, A & Djikeng, A 2023 'Open Institute of the African BioGenome Project: Bridging the gap in African biodiversity genomics and bioinformatics' OSF Preprints, pp. 1-18.
https://doi.org/10.31219/osf.io/adwh4

Digital Object Identifier (DOI):
10.31219/osf.io/adwh4

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Open Institute of the African BioGenome Project: Bridging the gap in African biodiversity genomics and bioinformatics

Abdoallah Sharaf[1,2], Charlotte C. Ndiribe[3], Taiwo Crossby Omotoriogun[4,5], Linelle Abueg[6], Bouabid Badaoui[7,8], Fatu J. Badiane Markey[9], Girish Beedessee[10], Diaga Diouf[11], Vincent C. Duru[12], Chukwuike Ebuzome[13], Samuel C. Eziuzor[14], Yasmina Jaufeerally Fakim[15], Giulio Formenti[6], Nidhal Ghanmi[16], Fatma Zahra Guerfali[17,18], Isidore Houaga[19], Justin Eze Ideozu[20], Sally Mueni Katee[21], Slimane Khayi[22], Josiah O. Kuja[23], Emmanuel Hala Kwon-Ndung[24], Rose A. Marks[25,26], Acclaim M. Moila[27], Zahra Mungloo-Dilmohamud[28], Sadik Muzemil[29], Helen Negussie[30], Julian O. Osuji[31], Verena Ras[32,33], Yves H. Tchiechoua[34], Yedomon Ange Bovys Zoclanclounon[35], Krystal A. Tolley[36,37], Cathrine Ziyomo[21], Ntanganedzeni Mapholi[38], Anne Muigai[39]*, Appolinaire Djikeng[19,38]*, ThankGod Echezona Ebenezer[40]*

[1]Department of Biology, University of Konstanz, 78457 Konstanz, Germany, [2]Genetic Department, Faculty of Agriculture, Ain Shams University, Cairo 11241, Egypt, [3]Department of Cell Biology and Genetics, University of Lagos, Lagos, Nigeria, [4]Biotechnology Unit, Department of Biological Sciences, Elizade University, Ilara-Mokin P.M.B. 002, Ondo State, Nigeria, [5]A. P. Leventis Ornithological Research Institute, University of Jos, Jos, Nigeria, [6]Vertebrate Genome Lab, The Rockefeller University, New York, NY 10021, USA, [7]Mohammed V University in Rabat, Rabat 10101, Morocco, [8]African Sustainable Agriculture Research Institute (ASARI), Mohammed VI Polytechnic University (UM6P), Laâyoune, Morocco, [9]Rutgers University, School of Graduate Studies, Newark, NJ 7039, USA, [10]Department of Biochemistry, University of Cambridge, Cambridge CB2 1QW, United Kingdom, [11]Laboratoire Campus de Biotechnologies Végétales, Département de Biologie Végétale, Faculté des Sciences et Techniques, Université Cheikh Anta Diop, Code postal 10700 Dakar-Fann, Dakar, Sénégal, [12]Department of Parasitology and Entomology, Nnamdi Azikiwe University, Awka, Nigeria, [13]Finima Nature Park, Port Harcourt 503101, Rivers State, Nigeria, [14]Department of Isotope Biogeochemistry, Helmholtz Center for Environmental Research-UFZ, Leipzig 4318, Germany, [15]Department of Agriculture University of Mauritius, Mauritius, [16]Bioinformatics Lab , Pasteur Institute of Tunis, Tunis 1002, Tunisia, [17] Laboratory of Transmission, Control and Immunobiology of Infections, Institut Pasteur de Tunis, Tunisia 13 Place Pasteur, BP 74, Tunis-Belvédère 1002, Tunisia, [18]University of Tunis El Manar, University Campus Farhat Hached, Romana-Tunis 1068, Tunisia, [19]Centre for Tropical Livestock Genetics and Health (CTLGH), Roslin Institute, University of Edinburgh, Midlothian, Edinburgh EH25 9RG, United Kingdom, [20]Genomic Medicine, Genomics Research Center, AbbVie, North Chicago, USA, [21]International Livestock Research Institute, Nairobi P.O BOX 30709-00200, Kenya, [22]Biotechnology research unit, CRRA-Rabat, National Institute of Agricultural Research, Rabat 10101, Morocco, [23]Bioinformatics Center, University of Copenhagen, Copenhagen 2200 København N., Denmark, [24]Department of Plant Science and Biotechnology, Federal University of Lafia, PMB 146, Lafia, Nigeria, [25]Department of Horticulture, Michigan State University, East Lansing, MI 48824, USA, [26]Department of Molecular and Cell Biology, University of Cape Town, Rondebosch 7701, South Africa, [27]Inqaba Biotec, Private Bag X12, Menlo Park, 0102, Pretoria, South Africa, [28]Digital Technologies Department, University of Mauritius, Reduit 80837, Mauritius, [29]School of Life Science, University of Warwick, Coventry CV4 7AL, United Kingdom, [30]Department of Microbial, Cellular and Molecular Biology, Addis Ababa University, Addis Ababa P.O.Box 1176, Ethiopia, [31]University of Port Harcourt, Port Harcourt P.M.B. 5323, Rivers State, Nigeria, [32]Computational Biology Division, Department of Integrative Biomedical Sciences, IDM, CIDRI Africa Wellcome Trust Centre, University of Cape Town, Cape Town, South Africa, [33]Department of Biodiversity and Conservation Biology, University of the Western Cape, Bellville, South Africa, [34]Pan African University Institute for Basic Sciences Technology and Innovation, Nairobi, Kenya, [35]Department of Crop Sciences and Biotechnology, Jeonbuk National University, South Korea, [36]South African National Biodiversity Institute, Private Bag X7 Claremont, Cape Town, South Africa, [37]Centre for Ecological Genomics and Wildlife Conservation, University of Johannesburg, Auckland Park, 2006 Johannesburg, South Africa, [38]Department of Agriculture and Animal Health, University of South Africa, Private Bag X6, Florida, 1710, Pretoria, South Africa, [39]Jomo Kenyatta University of Agriculture and Technology (JKUAT), P. O. Box 62000-00200, Nairobi, Kenya, [40]European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Cambridge CB10 1SD, United Kingdom

*Corresponding authors: thankgod1980@yahoo.co.uk, appolinaire.djikeng@ed.ac.uk, awmuigai@yahoo.co.uk

Africa, a continent of 1.3 billion people, had 326 researchers per one million people in 2018 (Schneegans, 2021; UNESCO, 2022), despite the global average for the number of researchers per million people being 1368 (Schneegans, 2021; UNESCO, 2022). Nevertheless, a strong research community is a requirement to advance scientific knowledge and innovation and drive economic growth (Agnew, et al., 2020; Sianes, et al., 2022). This low number of researchers extends to scientific research across Africa and finds resonance with genomic projects such as the African BioGenome Project (Ebenezer, et al., 2022).

The African BioGenome project (AfricaBP) plans to sequence 100,000 endemic African species in 10 years (Ebenezer, et al., 2022) with an estimated 203,000 gigabases of DNA sequence. AfricaBP aims to generate these genomes on-the-ground in Africa. However, for AfricaBP to achieve its goals of on-the-ground sequencing and data analysis, there is a need to empower African scientists and institutions to obtain the required skill sets, capacity and infrastructure to generate, analyse, and utilise these sequenced genomes in-country.

The Open Institute is the genomics and bioinformatics knowledge exchange programme for the AfricaBP (Figures 1 & 2). It consists of 10 participating institutions including the University of South Africa in South Africa and National Institute of Agricultural Research in Morocco. It aims to: develop biodiversity genomics and bioinformatics curricula targeted at African scientists, promote and develop genomics and bioinformatics tools that will address critical needs relevant to the African terrain such as limited internet access, and advance grassroot knowledge exchange through outreach and public engagement such as quarterly training and workshops.

The AfricaBP Open Institute is designed to close infrastructural gaps that exist in the biodiversity genomics space and build a critical mass of researchers across Africa. For instance, the sequencing of the genome of Hyacinth bean (*Lablab purpureus*), the first chromosome-scale plant genome assembly locally sequenced in Africa, benefited from in-depth bioinformatics training of four African scientists in African yam bean (*Sphenostylis stenocarpa*) over a period of 8 months (Njaci et al., 2022). Using this model, assuming other factors such as funding remain constant, if all the 100,000 African endemic species are sequenced and analysed through the AfricaBP Open Institute over the next 10 years, and two genomes where to be analysed by four African scientists, we estimate that this could train 200,000 African scientists in genomics and bioinformatics.

To deliver these goals, the AfricaBP Open Institute has established openly accessible workshops. The AfricaBP Open Institute workshop on endemic African species was held in May 2022 with nearly 300 registered attendees from across 20 African institutions and 29 countries. This workshop showcased the sequencing and assembly of two African endemic species in partnership with Inqaba Biotechnical Industries and the Vertebrate Genome Project. The AfricaBP Open Institute workshop on biodiversity genomic technologies and infrastructures was held in September 2022 with more than 400 registered participants from across 28 African countries. This workshop showcased the cutting-edge technologies shaping the biodiversity genomics field, including current understanding on global genomic databases, tools, and resources.

Here, we discuss these workshops and demonstrate how the AfricaBP Open Institute efforts could be further developed through a distributed hub and spoke model (Figures 1 & 2).

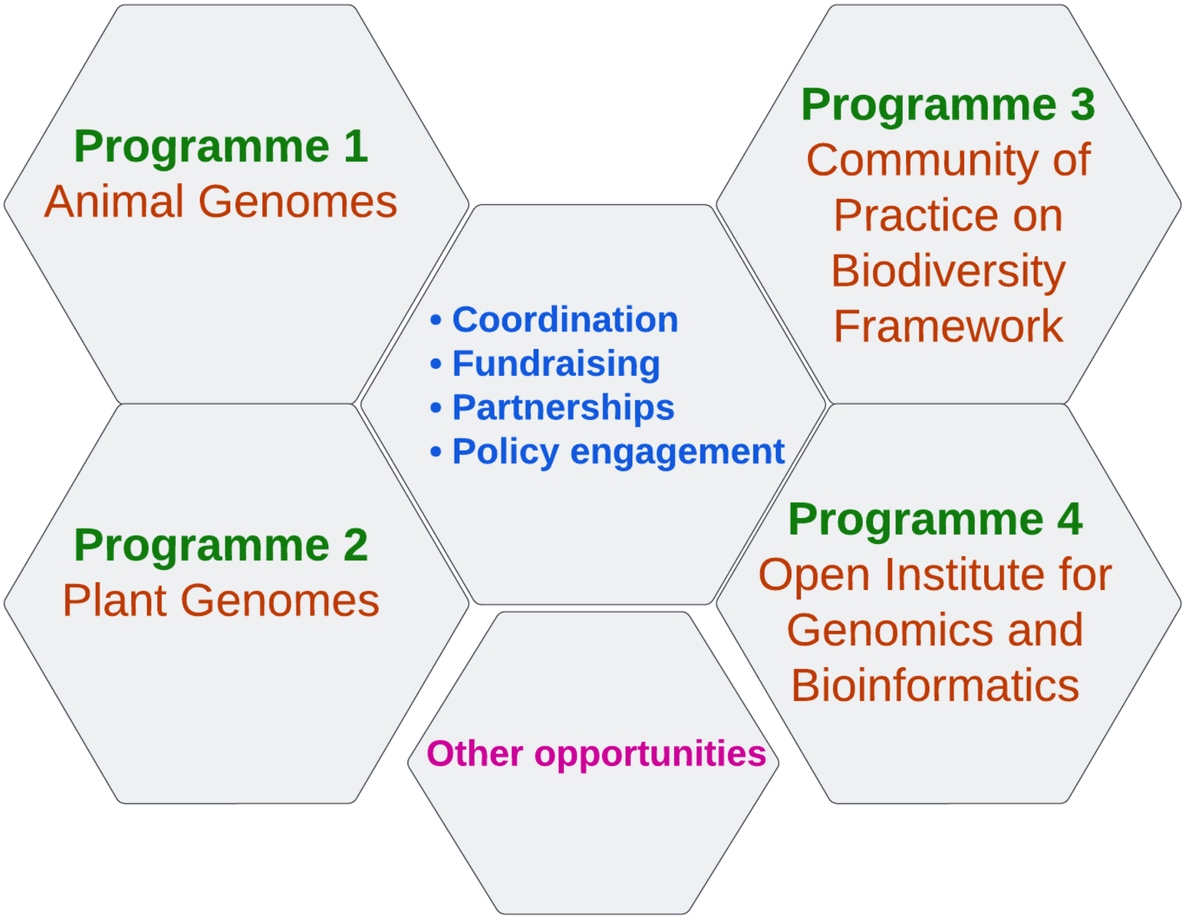

Figure 1: The African BioGenome Project will leverage on four programmes to achieve its goals. The AfricaBP Open Institute is the genomic and bioinformatics knowledge exchange programme for the African BioGenome Project (AfricaBP). *Programme 1:* The animal genomes programme of the AfricaBP focusing on animal species. This programme is also called the Nelson Mandela Genomes Initiative for

Conservation of Nature. *Programme 2:* The plant genomes programme of the AfricaBP focusing on plant species. *Programme 3:* Community of Practice on Biodiversity Framework. This programme focuses on issues around Access and Benefit Sharing of the Nagoya Protocol and the Post-2020 Global Biodiversity Framework. *Programme 4:* AfricaBP Open Institute for Genomics and Bioinformatics.

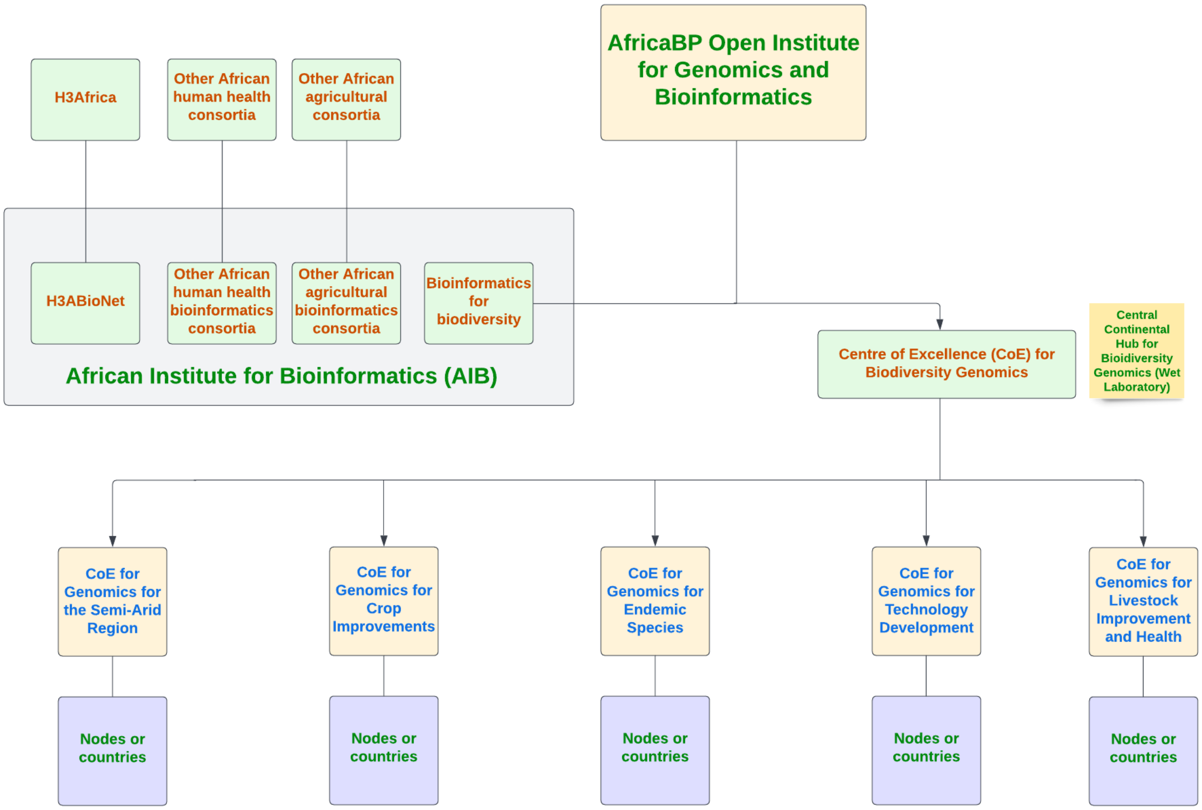
H3Africa
Other African human health consortia
Other African agricultural consortia
AfricaBP Open Institute for Genomics and Bioinformatics
H3ABioNet
Other African human health bioinformatics consortia
Other African agricultural bioinformatics consortia
Bioinformatics for biodiversity
African Institute for Bioinformatics (AIB)
Centre of Excellence (CoE) for Biodiversity Genomics
Central Continental Hub for Bioidiversity Genomics (Wet Laboratory)
CoE for Genomics for the Semi-Arid Region
CoE for Genomics for Crop Improvements
CoE for Genomics for Endemic Species
CoE for Genomics for Technology Development
CoE for Genomics for Livestock Improvement and Health
Nodes or countries
Nodes or countries
Nodes or countries
Nodes or countries
Nodes or countries

Figure 2: The AfricaBP Open Institute will operate a distributed model. The bioinformatics arm of the AfricaBP Open Institute (Bioinformatics for biodiversity) will link up with other bioinformatics consortia such as the H3ABioNet and the Bioinformatics Community of Practice across Africa to set up the African Institute for Bioinformatics (AIB). Bottom boxes depict regional hubs for specific areas of expertises for the Centre of Excellence (CoE) linked to by participating African countries or nodes. H3ABioNet is the bioinformatics network for H3Africa. Other African human health and agricultural bioinformatics consortia are independent communities and projects that are not part of the AfricaBP Open Institute.

Five priorities

The Open Institute of the AfricaBP aims to lower some of the barriers that prevent the advancement of biodiversity genomics and bioinformatics knowledge exchange across Africa. It has five critical priority areas:

1. Curriculum development.

One of the barriers in the development of training materials in Africa are training that is given in languages in which some participants may not fully understand (Ras, et al., 2021, Moore, et al., 2021). For example, genomic and bioinformatics courses are held in English for participants in Egypt where most people speak Arabic (Almarri, et al., 2021; EL-Attar, et al., 2022). The AfricaBP Open Institute aims to ensure that biodiversity genomic data and resources adhere to the FAIR principles (Findable, Accessible, Interoperable and Reusable). However, with the current systems of training, genomics data cannot adhere to the FAIR principles if the data is only findable in English or French. For example, an Egyptian scientist who understands Arabic better than English will not be able to maximise uptake and assimilation of training materials if such material is only available in English or French (Abdelhafiz, et al., 2021).

Learning in local languages, for example Swahili, would glue knowledge learned in the participant's native environment, enable a deeper understanding of concepts and help link biodiversity genomic practices to local knowledge (Bowden, et al., 2013; Tavares, 2015; Woolston and Osório, 2019). For instance, in a focused group discussion among 115 participants from Yoruba populations in West Africa, it was evident that words that describe the heritability of characters, traits and diseases such as horses inheriting the ability to race (*ere sisa la fi bi eshin*) and sickle cell disease (*arunmolegun*) exist in the Yoruba language - this can be used to improve the understanding of prior informed consent in genomics research (Taiwo, et al., 2020).

To address this challenge and advance the FAIR principles, the AfricaBP Open Institute will promote the incorporation of local languages in prior informed consents and employ the use of machine learning and natural language processing tools for the translation of training materials into selected widely-spoken African languages such as Kikongo, Swahili, isiZulu, Wolof, Arabic, and Amharic. This will be done by partnering with organisations such as *Masakhane*, an African organisation whose mission is to strengthen natural language processing research in African languages (Nekoto, *et al.,* 2020).

Another barrier to curriculum development in Africa is the use of non-African species as training datasets. Participants will benefit from training exercises where such is carried out with species endemic to the country where the training is taking place. For instance, in West Africa, a sequenced genome of African locust bean (*Parkia biglobosa*) and bush mango (*Cordyla pinnata*) could be used for training students in structural genomics to understand genome organisation and easily link gene content, adaptability and socio-economic relevance. Secondly, the classical organisms used for bioinformatics trainings in Africa, for example *Escherichia coli* or *Homo sapiens* (Mlotshwa et al., 2017; Akindolire, et al., 2019), do not provide a broad repertoire of challenges such as difficulty in sequencing or high genetic diversity that exists in non-model, endemic species. For instance, non-model vertebrates such as African lions and leopards will require unconventional sample materials and sequencing approaches (da Fonseca et al., 2016; Dures et al., 2019; Curry et al., 2021; Lehocká et al., 2021; Pečnerová et al., 2021; Rhie, et al., 2021).

2. Technology development and infrastructure

To enable knowledge exchange between the human genetics and biodiversity genomics communities across Africa, the AfricaBP Open Institute will work with communities such as the bioinformatics network for H3Africa (H3ABioNet) and other African human health and agricultural consortia to advance conversations for a centralised bioinformatics infrastructure in Africa (Figure 2) - this could be similar to the European Molecular Biology Laboratory's European Bioinformatics Institute (EMBL-EBI) for the European continent. Such infrastructure will provide expertise, services, and coordination for genomics and bioinformatics in data generation, storage, sharing and access to genomic data across Africa. Such infrastructure will also increase collaborations and maximise resources between the African human genetics community and the African biodiversity genomics community (Figure 2). For example, a research group within the African human genetics community who develops a bioinformatics tool could easily share this tool with another research group in the African biodiversity genomics community. The genomics arm of the AfricaBP Open Institute will form specialised African Centres of Excellence for Biodiversity Genomics, employing a public-private partnership, distributed hub and spoke model (Figure 2) - this could be similar to the African Centres of Excellence in infectious diseases in Mali, Nigeria, and Uganda (Folarin, et al., 2014; Glovanni, et al., 2022).

Similarly, one of the challenges experienced so far in the AfricaBP pilot project is the logistics of transporting samples from field locations to sequencing centres. The AfricaBP Open Institute aims to develop technology and tools to coordinate logistics of sample collections and transport from sampling locations to sequencing centres by using modular and lightweight supply chain technology for the transport of samples and specimens.

3. Promote grassroot knowledge exchange and equitable partnerships

A major goal of the AfricaBP Open Institute is to build grassroot capacity on Digital Sequence Information (DSI), Nagoya Protocol procedures, and the Post 2020 biodiversity framework (Ebenezer, et al., 2022). Bioinformatics and genomics capacity and infrastructure are not equally distributed across Africa, which could result in differences in the ability of countries to exploit the resulting sequencing data and resources. Under-resourced groups across Africa are unlikely to have all of these genomic analytical skills, hence, will require support to ensure they are able to gain from benefits derived (Ebenezer, et al., 2022). The AfricaBP Open Institute aims to support African research and academic institutions with fewer resources by supporting data analysis and building open tools and software to support their research needs. For instance, the AfricaBP has already engaged with communities such as Galaxy Africa to collaborate on ensuring these resources are available to the AfricaBP community.

The AfricaBP Open Institute also aims to support the formation of equitable partnerships through co-creation of project proposals, joint mobilisation of resources between African scientists and their partners, delivery of short- and longer- term courses, workshops, wet-lab trainings, and online and in-person residential capacity building programmes to facilitate knowledge exchange.

4. Maximise data ownership and sovereignty

Three entities across Africa will benefit from AfricaBP. This includes Africa's early career researchers and established scientists, agricultural and biodiversity conservation centres, and policy makers such as Africa's National Focal Points for the Global Environment Facility and Access and Benefit Sharing of the Nagoya Protocol within the structures of the Convention on Biological Diversity CBD).

In 2022 the CBD, the United Nations arm responsible for issues concerning biodiversity, adopted the Post-2020 Global Biodiversity Framework. This is a set of principles aimed to safeguard at least 90% of global biodiversity by 2030. At the moment, most African countries and entities described above are limited by instruments and capacities to maximise the benefits that the Post-2020 Global Biodiversity Framework presents as well as concerns around data ownership and governance.

The AfricaBP Open Institute presents the opportunity for African countries to build the required capacities to benefit from the Post-2020 Global Biodiversity Framework. For instance, an AfricaBP Open Institute workshop is planned for mid-2023 to create awareness on issues around DSI, the Post-2020 Biodiversity Framework and the Nagoya Protocol. Subsequent workshops will include training and equipping Africa's policy makers such as the National Focal Points of the CBD to be able to benefit from this.

5. Scientific entrepreneurship and industry

In 2019 the African Continental Free Trade Area (AfCFTA) agreement, the largest single market in the world by number of participating countries, came into force to connect 1.3 billion people across 55 countries with a combined gross domestic product of US$3.4 trillion (Gathii, 2019). Amongst AfCFTA's many benefits, it will enable and increase intra-Africa trade and consumption of African bioeconomy such as agricultural products and services (Pasara & Diko, 2020; Fusacchia, et al., 2022).

The AfricaBP Open Institute will promote scientific entrepreneurship to support Africa's bioeconomy. For instance, Africa experiences higher operating costs for science laboratories because most consumables such as reagents and equipment are imported into Africa from Europe, North America or Asia. The added shipping costs plus the levies and duties imposed by the importing African countries drive the costs up to become non-competitive with research laboratories in developed or industrialised countries (Helmy, et al., 2016; Schwarze et al., 2019). The AfricaBP Open Institute will engage conversations with African policy makers and governments to provide exemptions from duties on certified scientific consumables and equipment and will support scientists and entrepreneurs who will leverage on AfCFTA to lower laboratory operational costs, for example, by setting up franchises for local manufacturing of reagents and equipment, laboratory sharing and hackathons (Dolgin, 2018; Webb, 2019).

Similarly, insufficient capacity to translate genomic research into commercial products is one of the major challenges facing Africa's research ecosystem. For example, the diversity of some South African endophytes, bacterial or fungal species that live within a plant and have an endosymbiotic relationship, has been established. However, the biotechnology potential has not been fully explored (Abdalla & McGaw, 2018). The AfricaBP Open Institute will promote the use and translation of findings from basic genomic sequencing efforts into societal benefits. For example, an AfricaBP scientist who identifies a bioactive metabolite from the yew tree (*Taxus brevifolia*) in South Africa could partner with an AfricaBP industry partner to commercialise it in compliance with the South African laws on Access and Benefit Sharing (Kamau, 2022).

Moving forward with purpose

The AfricaBP Open Institute kick-started the process of implementing its priorities. In May 2022 it organised the *First AfricaBP Open Institute workshop on endemic African species*, while in September 2022 it organised the *AfricaBP Open Institute workshop on biodiversity genomic technologies and infrastructure*.

The 1st AfricaBP Open Institute workshop on endemic African species

The *1st AfricaBP Open Institute workshop on endemic African species* recorded a total of 306 applications, and 292 of these applications were from 29 African countries including Nigeria, Morocco, and Uganda (Figure 3). The applicants were affiliated with 193 African organisations and had various educational backgrounds, ranging from graduate students to full professors. The two-day workshop involved theoretical presentations and demonstrations of the journey undertaken during the sequencing of the first two AfricaBP genomes (speckled mousebird - *Colius striatus*, and beaked blind snake, *Rhinotyphlops lalandei*) in collaboration with the Inqaba Biotechnical Industries, and the Vertebrate Genome Project (VGP), and includes sample collections and processing, sample permits acquisition, ethical considerations, library preparations and sequencing, quality control, assembly, and assembly reproduction.

In addition, apart from theoretical presentations, the workshop also undertook a four-week remote residential practical exercise focusing on genome assembly reproduction (via the Slack platform) to train selected African scientists on genome assembly using the Galaxy Europe instance (https://assembly.usegalaxy.eu) in collaboration with the VGP (Table 1). Step-by-step practical demonstrations and hands-on tutorials were provided on the VGP pipeline to assemble sample data from the yeast (*Saccharomyces cerevisiae* S288C) genome. Afterwards, the genome assembly of the speckled mousebird (*Colius striatus*) was reproduced using the VGP pipeline (Lariviere, et al., 2022).

The aim of the four-week residential post-workshop exercise was to reproduce the assembly of the *Colius striatus* genome, ensuring that the results are comparable to those generated by the VGP team (Table 1). The training began with the acquisition of the most recent VGP workflow from the official GitHub repository (https://tinyurl.com/vgpassemblyv2). The workflow files were imported to the Galaxy Europe platform as described on the official VGP GitHub repository. The VGP pipeline consists of five main workflows: Meryldb creation, Hifiasm-HiC-assembly, Purged-assembly, Bionano scaffolding, and Hi-C scaffolding (Lariviere, et al., 2022). The Purged-assembly and Bionano scaffolding workflows are optional. The purged assembly workflow is used to purge duplications and is necessary only when duplications are identified by QC after the assembly workflow (Lariviere, et al., 2022).

Table1: The assembly of the *Colius striatus* genome statistics produced by the VGP assembly team versus the assembly reproduced by AfricaBP Open Institute selected attendees.

Evaluating metrics	***Scaffolds**	**Assembly category**	**VGP assembly**	**AfricaBP assembly**
Total length (bp)	Post-Scaffolding	Scaffold assembly	1159734284	1159722284
Number of contigs/scsffolds			227	219
**BUSCO score			C:3009[S:2987,D:22], F:42,M:303,n:3354	C:3008[S:2987, D:21],F:42,M:304,:n:3354
Scaffold N50			58796297	58796297

** Complete (C) [Single-copy (S), Duplicated (D)], Fragmented (F), Missing (M)
*Only post-scaffolding is reported

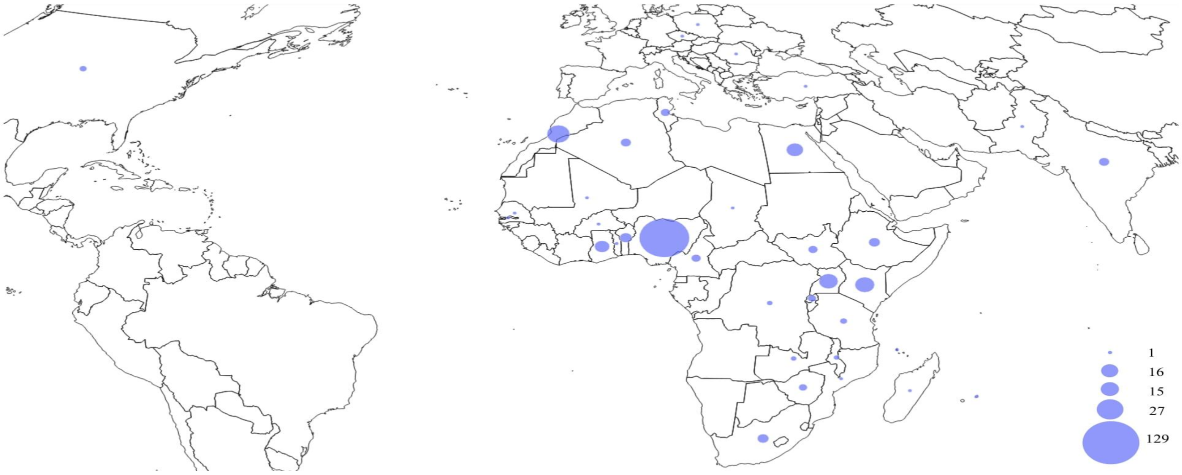

Figure 3: The 1st AfricaBP Open Institute workshop on endemic African species has participants representations from all geographic regions of Africa, and shows the diversity of the workshop participants outside Africa. The size of the coupled circles indicate the number of participants from each country.

The AfricaBP Open Institute workshop on biodiversity genomics technologies and infrastructure

Over 400 candidates from 28 African countries registered to attend the AfricaBP Open Institute workshop on biodiversity genomic technologies and infrastructure. More than 70% of the registered candidates were early career researchers with ongoing or upcoming projects in diverse areas of genomics (such as plants, vertebrates, microbial genomics). Apparently, 56% of the applicants were very familiar with genomics resources hosted at the National Center for Biotechnology Information (NCBI) and European Bioinformatics Institute (EMBL-EBI). About 45% of the applicants have limited experiences with genome assembly and annotation of genomes while 91% had prior experience with global genomic databases (Figure 4).

Feedback from participants (n=165) showed that at least 95% of participants described the workshop as interesting and that it presented the right amount of information (Figure 5). Participants also found that the workshop provided sufficient time for discussion, met their expectations and was of the right length of time. Importantly, 98% of participants expressed interest in attending future workshops of similar nature that AfricaBP will organise (Figure 5).

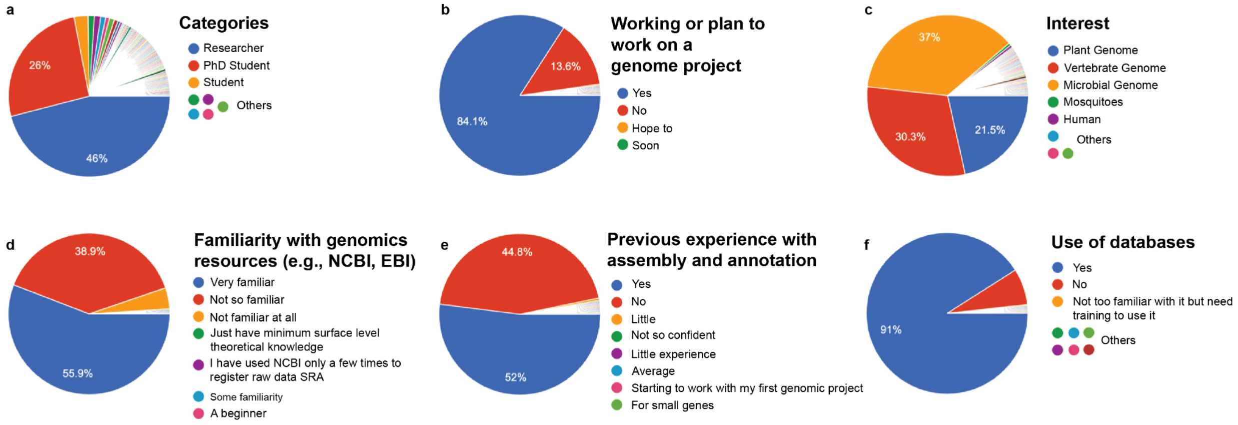

Figure 4. Response of pre-workshop survey of the AfricaBP workshop on Genomic Technologies and Infrastructures to identify skill needs of African researchers and genomic gaps in Africa. Participants' responses with respect to familiarity with genomic resources/databases and past experience in genome assembly and annotation. Number of respondents = 404.

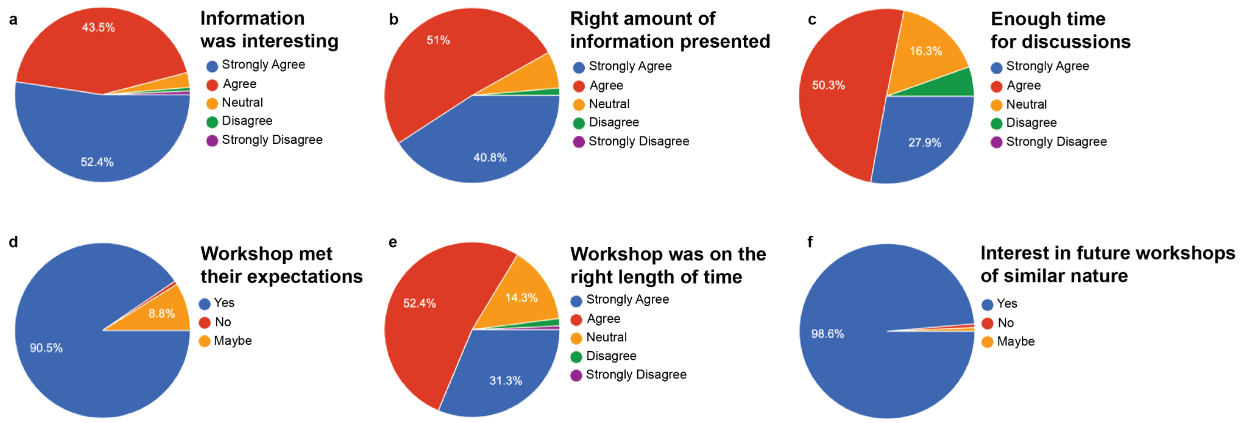

Figure 5. Responses of post-workshop survey to evaluate impact and outcome of the AfricaBP workshop on Genomic Technologies and Infrastructures. Participants' responses pertaining to quality of the workshop and future interest in similar workshops. Number of respondents = 165.

Conclusion and next steps

The organisation of the *AfricaBP Open Institute workshop on endemic African species* and *biodiversity genomics technologies and infrastructure* present clear indicators for training African scientists in genomic procedures, technologies, infrastructures as well as in ethical, legal and social issues that accompany genomics practices. This is reflected in the increasing number of candidates (Figure 3, 4, and 5) participating in AfricaBP Open Institute workshops in the last few months, involvement of African scientists in diverse genomic projects, satisfaction with material and content delivery and interests to participate in future and upcoming workshops. For instance, the VGP genome assembly pipeline on Galaxy Europe makes it intuitive for a biologist with minimal computational skills to assemble genomes and generate genome assembly assessment and visualisations of results. It is particularly relevant to students and early career researchers in Africa where limited computational facilities are available or accessible to non-specialists. In the future, the AfricaBP Open Institute will work with the VGP to make this assembly workflow available through the Galaxy Africa instance.

It is noteworthy that the majority of the participants were from countries with active genomics research such as Nigeria, Kenya, and Morocco (Figure 3). There are large geographical areas of the continent, for example Angola, Namibia, and Sudan, that were not represented. This could be that the outreach of the AfricaBP Open Institute in these areas was not effective or simply because minimal genomics activity happens in these areas.

The AfricaBP Open Institute workshops are currently being led and delivered by African scientists. This, in turn, is helping to deliver the goal of building leadership in genomics and bioinformatics across Africa by mentoring early career scientists, establishing key contacts, and building networks. For instance, the AfricaBP has secured a platform within the African Galaxy instance and a licence to host its own Research Electronic Data Capture (REDCap, https://redcap.africanbiogenome.org) as a secure survey and data capture platform. The Galaxy platform will assist scientists with limited resources or computational skills, as it is user-friendly, flexible, and provides several implemented bioinformatics tools and supports hands-on workshops. Currently, the Galaxy Africa instance is supported by one of the computational clusters hosted at Institut Pasteur de Tunis, Tunisia. Efforts are underway within AfricaBP to secure cloud-based support from cloud providers. This will help with storage and increase processing capacity as the AfricaBP project scales.

The AfricaBP Open Institute will build on the successes of the *workshops on endemic African species* and *biodiversity genomics technologies and infrastructure.* It already has 9 workshops planned in 2023. This includes four online workshops and five regional hybrid workshops. For instance, one of these workshops is an online workshop which will focus on science communication and grant writing, and it will involve a residential grant writing exercise to apply for an identified grant. Another is a regional hybrid workshop to be hosted by the University of Port Harcourt in Nigeria. This workshop will focus on the importance of biodiversity genomics in conservation of species. It will include sessions tailored to emphasise the value of taxonomy in ensuring integrity of species of interest and proceed to steps involved in collection and processing of samples to laboratory protocols involved in resolving genome sequences.

The information and perspectives given to African researchers on biodiversity genomic technologies and infrastructure resources in Africa can broaden coursework applicability and collaborative perspectives. Post-workshop feedback (Figure 5) will guide future events content and organisation.

Acknowledgements

We would like to thank the following scientists for helping to review and provide feedback on the initial draft of this manuscript:

1. Nicola Mulder, Prof., Computational Biology Division, Faculty of Health Sciences, University of Cape Town, South Africa
2. Mark Blaxter, Prof., Tree of Life, Wellcome Sanger Institute, Cambridge UK

Competing interest

Acclaim Moila is a member of staff of Inqaba Biotechnical Industries (Pty) Ltd.
Justin Eze Ideozu is a member of staff of Abbvie.

References

Abdalla, M. A., & McGaw, L. J. (2018). Bioprospecting of South African Plants as a Unique Resource for Bioactive Endophytic Microbes. *Frontiers in Pharmacology*, *9*(MAY), 456. https://doi.org/10.3389/FPHAR.2018.00456

Agnew, A., Francescon, D., Martin, R., Rhannam, M., & Schemm, Y. (2020). The Power of Data to Advance the SDGs Mapping research for the Sustainable Development Goals. *Elsevier*, 1–51. Retrieved from https://www.elsevier.com/__data/assets/pdf_file/0004/1058179/Elsevier-SDG-Report-2020.pdf

Akindolire, M. A., Aremu, B. R., & Ateba, C. N. (2019). Complete Genome Sequence of Escherichia coli O157:H7 Phage PhiG17. *Microbiology Resource Announcements*, *8*(3). https://doi.org/10.1128/MRA.01296-18

Cheng, H., Concepcion, G. T., Feng, X., Zhang, H., & Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods 2021 18:2*, *18*(2), 170–175. https://doi.org/10.1038/s41592-020-01056-5

Curry, C. J., Davis, B. W., Bertola, L. D., White, P. A., Murphy, W. J., & Derr, J. N. (2021). Spatiotemporal Genetic Diversity of Lions Reveals the Influence of Habitat Fragmentation across Africa. *Molecular Biology and Evolution*, *38*(1), 48–57. https://doi.org/10.1093/MOLBEV/MSAA174

da Fonseca, R. R., Albrechtsen, A., Themudo, G. E., Ramos-Madrigal, J., Sibbesen, J. A., Maretty, L., … Pereira, R. J. (2016). Next-generation biology: Sequencing and data analysis approaches for non-model organisms. *Marine Genomics*, *30*, 3–13. https://doi.org/10.1016/J.MARGEN.2016.04.012

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., … Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature 2012 485:7398*, *485*(7398), 376–380. https://doi.org/10.1038/nature11082

Dolgin, E. (2018). How to start a lab when funds are tight career-feature. *Nature*, *559*(7713), 291–293. https://doi.org/10.1038/D41586-018-05655-3

Dures, S. G., Carbone, C., Loveridge, A. J., Maude, G., Midlane, N., Aschenborn, O., & Gottelli, D. (2019). A century of decline: Loss of genetic diversity in a southern African lion-conservation stronghold. *Diversity and Distributions*, *25*(6), 870–879. https://doi.org/10.1111/DDI.12905

Ebenezer, T. E., Muigai, A. W. T., Nouala, S., Badaoui, B., Blaxter, M., Buddie, A. G., … Djikeng, A. (2022). Africa: sequence 100,000 species to safeguard biodiversity. *Nature 2022 603:7901*, *603*(7901), 388–392. https://doi.org/10.1038/d41586-022-00712-4

Folarin, O. A., Happi, A. N., & Happi, C. T. (2014). Empowering African genomics for infectious disease control. *Genome Biology*, *15*(11), 515. https://doi.org/10.1186/S13059-014-0515-Y/METRICS

Formenti, G., Abueg, L., Brajuka, A., Brajuka, N., bal Gallardo-Alba, C., Giani, A., … Appel Alzheimer Disease, R. (2022). Gfastats: conversion, evaluation and manipulation of genome sequences using assembly graphs. *Bioinformatics*, *38*(17), 4214–4216. https://doi.org/10.1093/BIOINFORMATICS/BTAC460

Fusacchia, I., Balié, J., & Salvatici, L. (2022). The AfCFTA impact on agricultural and food trade: a value added perspective. *European Review of Agricultural Economics*, *49*(1), 237–284. https://doi.org/10.1093/ERAE/JBAB046

Gathii, J. T. (2019). Agreement Establishing The African Continental Free Trade Area. *International Legal Materials*, *58*(5), 1028–1083. https://doi.org/10.1017/ILM.2019.41

Ghurye, J., Rhie, A., Walenz, B. P., Schmitt, A., Selvaraj, S., Pop, M., … Koren, S. (2019). Integrating Hi-C links with assembly graphs for chromosome-scale assembly. *PLOS Computational Biology*, *15*(8), e1007273. https://doi.org/10.1371/JOURNAL.PCBI.1007273

Giovanni, M. Y. (2022). African Centers of Excellence in Bioinformatics and Data Intensive Science: Building Capacity for Enhancing Data Intensive Infectious Diseases Research in Africa. *Journal of Infectious Diseases & Microbiology*. https://doi.org/10.37191/MAPSCI-JIDM-1(2)-006

Helmy, M., Awad, M., & Mosa, K. A. (2016). Limited resources of genome sequencing in developing countries: Challenges and solutions. *Applied & Translational Genomics*, *9*, 15–19. https://doi.org/10.1016/J.ATG.2016.03.003

Kamau, E.C. (2022). The South African ABS Regime: New Wine in Old Wine Skins?. In: Chege Kamau, E. (eds) Global Transformations in the Use of Biodiversity for Research and Development. Ius Gentium: Comparative Perspectives on Law and Justice, vol 95. Springer, Cham. https://doi.org/10.1007/978-3-030-88711-7_6

Lariviere, D., Ostrovsky, A., Gallardo, C., Syme, A., Abueg, L., Pickett, B., Formenti, G. and Sozzoni, M. (2022). VGP assembly pipeline (Galaxy Training Materials). https://training.galaxyproject.org/archive/2022-03-01/topics/assembly/tutorials/vgp_genome_assembly/tutorial.html Online; accessed Jul 14 2022

Lehocká, K., Black, S. A., Harland, A., Kadlečík, O., Kasarda, R., & Moravčíková, N. (2021). Genetic diversity, viability and conservation value of the global captive population of the Moroccan Royal lions. *PLOS ONE*, *16*(12), e0258714. https://doi.org/10.1371/JOURNAL.PONE.0258714

Lieberman-Aiden, E., Van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., … Dekker, J. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science (New York, N.Y.)*, *326*(5950), 289–293. https://doi.org/10.1126/SCIENCE.1181369

Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., & Zdobnov, E. M. (2021). BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Molecular Biology and Evolution*, *38*(10), 4647–4654. https://doi.org/10.1093/MOLBEV/MSAB199

Mapunda, G., & Gibson, H. (2022). On the suitability of Swahili for early schooling in remote rural Tanzania: do policy and practice align? *Journal of the British Academy*, *10s4*, 141–168. https://doi.org/10.5871/JBA/010S4.141

Mlotshwa, B. C., Mwesigwa, S., Mboowa, G., Williams, L., Retshabile, G., Kekitiinwa, A., … Mpoloka, S. W. (2017). The collaborative African genomics network training program: a trainee perspective on training the next generation of African scientists. *Genetics in Medicine*, *19*(7), 826–833. https://doi.org/10.1038/GIM.2016.177

Nekoto, W., Marivate, V., Matsila, T., Fasubaa, T., Kolawole, T., Fagbohungbe, T., … Bashir, A. (2020). Participatory Research for Low-resourced Machine Translation: A

Case Study in African Languages. *Findings of the Association for Computational Linguistics Findings of ACL: EMNLP 2020*, *9*, 2144–2160. https://doi.org/10.18653/V1/2020.FINDINGS-EMNLP.195

Njaci, I., Waweru, B., Kamal, N., Muktar, M. S., Fisher, D., Gundlach, H., … Jones, C. S. (2022). Chromosome-scale assembly of the lablab genome - A model for inclusive orphan crop genomics. *BioRxiv*, 2022.05.08.491073. https://doi.org/10.1101/2022.05.08.491073

Pasara, M. T., & Diko, N. (2020). The Effects of AfCFTA on Food Security Sustainability: An Analysis of the Cereals Trade in the SADC Region. *Sustainability 2020, Vol. 12, Page 1419*, *12*(4), 1419. https://doi.org/10.3390/SU12041419

Pečnerová, P., Garcia-Erill, G., Liu, X., Nursyifa, C., Waples, R. K., Santander, C. G., … Hanghøj, K. (2021). High genetic diversity and low differentiation reflect the ecological versatility of the African leopard. *Current Biology*, *31*(9), 1862-1871.e5. https://doi.org/10.1016/J.CUB.2021.01.064

Rhie, A., Walenz, B. P., Koren, S., & Phillippy, A. M. (2020). Merqury: Reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biology*, *21*(1), 1–27. https://doi.org/10.1186/S13059-020-02134-9/FIGURES/6

Savara, J., Novosád, T., Gajdoš, P., & Kriegová, E. (2021). Comparison of structural variants detected by optical mapping with long-read next-generation sequencing. *Bioinformatics (Oxford, England)*, *37*(20), 3398–3404. https://doi.org/10.1093/BIOINFORMATICS/BTAB359

Schneegans, S., Lewis, J., & Straza, T. (2021). The race against time for smarter development. *UNESCO Science Report: The Race against Time for Smarter Development*, 30–77. Retrieved from https://unesdoc.unesco.org/ark:/48223/pf0000377433.locale=en

Schwarze, K., Buchanan, J., Fermont, J. M., Dreau, H., Tilley, M. W., Taylor, J. M., … Wordsworth, S. (2019). The complete costs of genome sequencing: a microcosting study in cancer and rare diseases from a single center in the United Kingdom. *Genetics in Medicine 2019 22:1*, *22*(1), 85–94. https://doi.org/10.1038/s41436-019-0618-7

Sianes, A., Vega-Muñoz, A., Tirado-Valencia, P., & Ariza-Montes, A. (2022). Impact of the Sustainable Development Goals on the academic research agenda. A scientometric analysis. *PLOS ONE*, *17*(3), e0265409. https://doi.org/10.1371/JOURNAL.PONE.0265409

Tavares, N.J. (2015). How strategic use of L1 in an L2-medium mathemathics classroom facilitates L2 interaction and comprehension. *International Journal of Bilingual Education and Bilingualism* (Special Issue on Multilingual and Multimodal CLIL). DOI: http://dx.doi.org/10.1080/13670050.2014.988115

Tibategeza, E. R., & Du Plessis, T. (2018). The prospects of kiswahili as a medium of instruction in the Tanzanian education and training policy. *Journal of Language and Education*, *4*(3), 88–98. https://doi.org/10.17323/2411-7390-2018-4-3-88-98

UNESCO. (2022). UIS Statistics - science, technology and innovation. Retrieved January 15, 2023, from http://data.uis.unesco.org/

Vurture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H., Gurtowski, J., & Schatz, M. C. (2017). GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics*, *33*(14), 2202–2204. https://doi.org/10.1093/BIOINFORMATICS/BTX153

Webb, H., Bezuidenhout, L., Nurse, J. R. C., & Jirotka, M. (2019). Lab hackathons to overcome laboratory equipment shortages in Africa: Opportunities and challenges. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3290607.3299063

Yuan, Y., Chung, C. Y. L., & Chan, T. F. (2020). Advances in optical mapping for genomic research. *Computational and Structural Biotechnology Journal*, *18*, 2051–2062. https://doi.org/10.1016/J.CSBJ.2020.07.018