

Comment



SHANE CROSS/NATURE PICTURE LIBRARY

The red mangrove tree is indigenous to Africa and is being sequenced as part of the African BioGenome pilot project.

Africa: sequence 100,000 species to safeguard biodiversity

ThankGod Echezona Ebenezer, Anne W. T. Muigai, Simplicie Nouala, Bouabid Badaoui, Mark Blaxter, Alan G. Buddie, Erich D. Jarvis, Jonas Korch, Josiah O. Kuja, Harris A. Lewin, Rokšana Majewska, Ntanganedzeni Mapholi, Suresh Maslamoney, Michèle Mbo'o-Tchouawou, Julian O. Osuji, Ole Seehausen, Oluwaseyi Shorinola, Christian Keambou Tiambo, Nicola Mulder, Cathrine Ziyomo & Appolinaire Djikeng

Build a major genomics resource on the continent to help breeders and conservationists.

Sleeper fish (*Bostrychus africanus*) are a staple food in West Africa. Harvesting them provides an important source of income for hundreds of communities across the Gulf of Guinea in the Atlantic Ocean. Yet little is known about the genetics of this fish – information that is crucial to safeguarding its genetic diversity, and to

enhancing its resilience in the face of climate change and other pressures.

This situation is all too familiar across Africa. Consider orphan crops, which have a crucial role in regional food security, even though they are not typically traded internationally. More than 50% of these have not had their genomes sequenced – from the

fluted pumpkin (*Telfairia occidentalis*) to the marama bean (*Tylosema esculentum*). The same is true of more than 95% of the continent's known endangered species (see 'Africa's neglected genomes').

What's more, by our estimate, around 70% of the 35 or so projects that have focused on studying, conserving or improving biological diversity in Africa over the past 15 years have been led from outside the continent. In fact, among the plant genomes sequenced globally over the past 20 years, almost all of the African species were sequenced elsewhere – mainly in the United States, China and Europe¹. This offshoring slows down the much-needed building of expertise and resources in genomics and bioinformatics in Africa (see 'Africa left out of global genomics efforts').

The African BioGenome Project (AfricaBP) is an effort to sequence the genomes of 105,000 endemic species: plants, animals, fungi, protists and other eukaryotes. It currently involves 109 African scientists (87 of whom work in Africa) and 22 African organizations.

This store of reference genomes – built in Africa, for Africa – will help plant and animal breeders to produce resilient and sustainable food systems. It will inform biodiversity conservation across the continent. And it will strengthen Africa's ability to deliver on the goals of the post-2020 global biodiversity framework of the Convention on Biodiversity (CBD). These goals, one of which is to maintain at least 90% of genetic diversity for all known species by 2030, are to be agreed on next month at a meeting in Kunming, China.

An AfricaBP pilot project was launched in June 2021. In this, researchers are sequencing 2,500 indigenous African species, including the Boyle's beaked blind snake (*Rhinotyphlops boylei*) from southern Africa and the red mangrove tree (*Rhizophora mangle*) from Nigeria. They are also mapping out the ethical, legal and social issues raised by a major biodiversity sequencing project – because of cultural sensitivities around certain species, or questions around who has access to the data and who benefits from any resulting discoveries.

For AfricaBP to be scaled up and sustained over the next decade, agencies and organizations need to allocate long-term investments to the project. Such groups include the African Union Commission, national and regional scientific agencies (such as the African Academy of Sciences), and international partners and organizations, including the US National

Africa left out of global genomics efforts

Most projects that aim to study, conserve or improve biological diversity in Africa have been led by researchers outside the continent.

Projects to sequence biodiversity rarely meet the needs of people in Africa or align with its countries' science agendas^{4,14–16} (such as on agricultural technologies¹⁵). Take the Human Genome Project. Less than 2% of genomes analysed in the two decades since the project began are from African individuals, even though Africa harbours more human genetic diversity than any other continent.

African researchers who contribute to data collection in such projects are not always credited for their work. A 2021 study¹⁷ revealed that about 15% of 32,061 articles on global health research conducted in sub-Saharan Africa had no authors based in the country in which the research took place.

Currently, the International Nucleotide Sequence Database Collaboration, the core infrastructure for the collection and sharing of the world's nucleotide sequence data and metadata, names only those who have submitted samples or sequence data, not the primary owners or custodians of the sample. In practice, this means that if an African scientist collects samples from Bioko squeaker frogs (*Athroleptis bioko*) in Equatorial Guinea, for example, and

sends them to a colleague in Canada who then submits a sequence to the database, only the Canadian researcher will receive recognition for the data. Recent efforts by the consortium and others¹⁸ will help to address some of these gaps. By December this year, the consortium will make it mandatory for those submitting sequence data to declare the country or region in which the sample was collected. But it is still unclear whether the credit given to sample custodians will be similar to that of sample submitters.

Besides the lack of recognition, African researchers rarely retain access to the data they help to collect, nor do they receive related benefits – either from royalties resulting from specific discoveries in genetics, or those stemming from technological advances and growth in scientific capability that such projects can bring.

For instance, during the 2014–16 Ebola epidemic in West Africa, around 269,000 blood samples were obtained from patients for diagnosis. Thousands of those samples were shipped overseas, including to Europe and North America. None of the genomics researchers working in Africa knows where these samples are now housed¹⁹ and, as far as the African human-genetics community knows, the sample providers never received the results of their blood collections.

Science Foundation and the UK research funder Wellcome. By our calculations, this will require at least US\$100 million per year for the next 10 years (see 'AfricaBP: structure and costs').

Some might argue that \$1 billion would be better spent on combating malnutrition and disease in impoverished communities across Africa. Yet consider the Human Genome Project, which cost around \$3 billion in 2003.

“Thousands of African species have been ignored by the global genomics community.”

By 2019, the human genetics and genomics sector alone was contributing \$265 billion annually to the US economy². Likewise, the World Bank invested millions of dollars in outbreak preparedness from 2017, some of which was used to fund the African Centre of Excellence for Genomics of Infectious Diseases in Ede, Nigeria. This investment meant that Africa was much better equipped to meet the challenges presented by the COVID-19 pandemic.

Species sidelined

Thousands of African species have been ignored by the global genomics community. Only 20 of the 798 plant genomes sequenced globally over the past 20 years are native to Africa¹, for example. Yet sub-Saharan Africa

Comment

alone, which is home to at least 45,000 plant species³, is the second-largest contributor to global plant diversity after South America. Last year, researchers reported that 60% of these species are endemic, and that many could have potential applications in agriculture or drug development⁴. Evidence suggests, for instance, that African ginger (*Siphonochilus aethiopicus*) could be used to treat asthma and influenza, among other conditions^{5,6}.

Most of the genomics and bioinformatics expertise that does exist across Africa, including the sequencing facilities, is concentrated in private and non-governmental organizations, such as Inqaba Biotechnical Industries in Pretoria, South Africa, and Redeemer's University in Nigeria. This means that, although the national research institutes are given the responsibility of setting the country's scientific agenda, the tools needed to actually improve public health, agriculture and conservation are outside their control⁷.

AfricaBP will focus on endemic African species that have economic, scientific and cultural significance for African communities.

Sustained government investment in genomics – including the creation of permanent university positions – will help to ensure that African scientists who have received training through African-coordinated genomics projects stay in Africa.

National and regional expansion of tissue-sample collection, taxonomic identification, biobanking of samples and cataloguing of metadata will make it much easier for researchers to monitor species – and ultimately to protect them. Species discovered as a result of the genomics project could be added to the CBD 2030 targets.

Lastly, if the African Union Commission includes AfricaBP in the suite of schemes it is currently backing, the project could enable the commission to achieve at least three of the development goals encapsulated in the African Union Agenda 2063: The Africa We Want. These are: the use of modern techniques and technology to increase agricultural productivity sustainably; the sustainable use of ocean resources to drive economic growth; and the development of environmentally sustainable and climate-resilient economies. (Agenda 2063 is the blueprint for the continent's transformation into a global powerhouse, as laid out by leaders of the 55 African Union member states in 2013.)

Key priorities

AfricaBP will bring together national and regional institutions, countries and corporations, including already recognized genomics infrastructures, such as the National Institute for Biomedical Research in Kinshasa in the Democratic Republic of the Congo. The project has three main goals.

Improve food systems. The first goal is to provide a resource that enables plant and animal breeders to use various approaches (from conventional breeding to gene editing) to build resilient and sustainable food systems. A 2021 genome analysis⁸ of 245 Ethiopian indigenous chickens, for instance, revealed the genetic basis of various adaptations that enable the chickens to tolerate harsh environmental conditions (from cold temperatures to water scarcity) – crucial information for poultry producers worldwide. To help achieve this goal, AfricaBP

“To achieve such a massive sequencing feat, African researchers need state-of-the-art genome technologies.”

will partner with the African Plant Breeding Academy and the African Animal Breeding Network, both of which were established in the past decade to improve African breeders' training and research practices.

Improve conservation. The second goal is to make it easier for researchers to identify species and populations that are at risk of extinction, and to design and implement effective conservation strategies. A 2020 study⁹ on the genetic structure of African savannah elephant populations, for example, revealed that the long-term survival of the elephants

requires establishing at least 14 wildlife corridors between 16 of the protected areas in Tanzania. Similarly, a genome study¹⁰ of 13 individuals representing 2 subspecies of eastern gorilla showed that inbreeding has led to the purging of severely harmful recessive mutations from one of the subspecies (*Gorilla beringei beringei*, or mountain gorillas). The accumulation of such damaging mutations in eastern gorillas over the past 100,000 years has reduced their resilience to environmental change and pathogen evolution.

Improve sharing of data and benefits. The third goal is to kick-start a process in which existing multilateral agreements around data sharing are improved and harmonized across the continent – to ensure that the benefits derived from genetic resources are shared equitably across Africa.

In 2010, nations adopted the Nagoya Protocol on Access and Benefits Sharing to ensure that the benefits arising from the use of biological resources are shared fairly. Certainly, any benefit derived from the genetic resources obtained through AfricaBP should be shared by the people of Africa – whether it be a superior strain of drought-resistant sugar beet (*Beta macrocarpa* Guss) or a new drug derived from the rooibos plant (*Aspalathus linearis*).

As written, however, the Nagoya Protocol has gaps when it comes to Africa. It fails to take into account the customs and practices of the diverse ethnic groups across the



A technician checks cassava plants in a research laboratory near Abidjan, Côte d'Ivoire.

continent. These might not be documented or written into law, but have shaped how people interact with certain plants or animals for hundreds – sometimes thousands – of years. In West Africa, for example, some communities forbid the cutting down or harming of iroko trees, which are thought to have supernatural powers.

There are also inconsistencies in how the Nagoya Protocol is applied in different countries. The African Union guidelines for the implementation of the Nagoya Protocol in Africa states that those countries that are not parties to the Nagoya Protocol should be refused access to the genetic resources of other African member states. But only some countries follow this; South Africa grants non-parties access to the nation's genetic resources, whereas Ethiopia does not.

Likewise, not all countries require researchers wanting to extract genetic resources to consult community protocols. These include the rules and standards around the handling of biological specimens – as laid out by communities under the guidance of the custodians of customary laws (local chiefs and community heads). These custodians, in turn, work closely with state and national governments; sometimes, community protocols will refer to state, national or international laws. In Benin, for example, such protocols state that researchers cannot enter Gbévozoun forest or take any specimens from it because it houses the deity Gbévo, which protects the community.

Ultimately, it is the responsibility of the African Union Commission to improve and harmonize the treaties and guidelines around data and benefit sharing. Doing this would make it easier for AfricaBP researchers to obtain sampling permits, in accordance with the Nagoya Protocol and material transfer agreements (the legal documents required to send biological materials from one organization to another, or from one country to another).

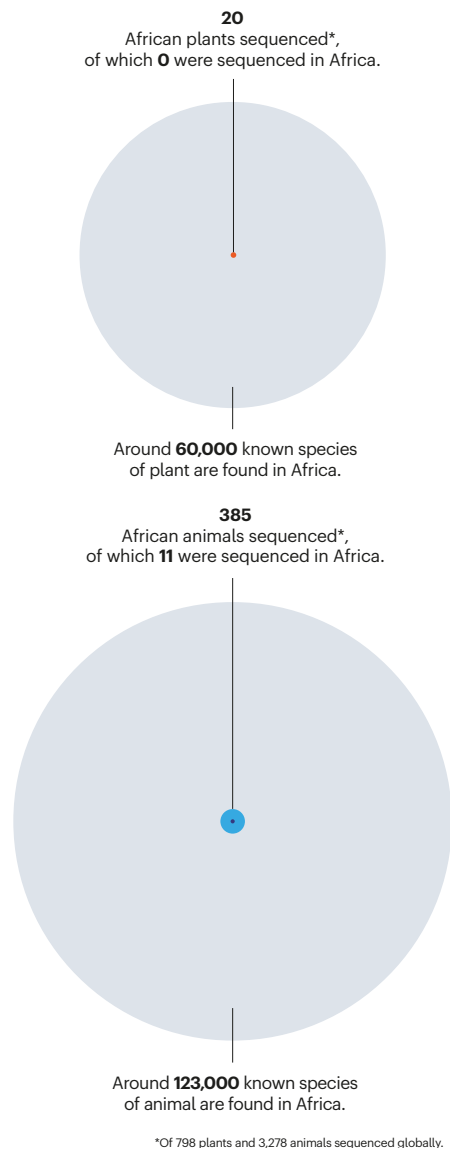
But AfricaBP will enable the African Union, the CBD and other African agencies, such as the African Academy of Sciences, to integrate genomic information into their policymaking around biological diversity across Africa. This in itself will raise awareness about the Nagoya Protocol, and so encourage greater harmonization in its use.

Furthermore, the 109 scientists championing AfricaBP will coordinate with the African Group of Negotiators on Biodiversity (researchers, policymakers and other stakeholders who represent the continent in CBD negotiations) to ensure that sequencing information is specifically included in the post-2020 global biodiversity framework.

Currently, the Nagoya Protocol specifies that 'biological samples' can be exchanged for scientific training or technology transfer. The

AFRICA'S NEGLECTED GENOMES

Around 95% of Africa's known endangered species are currently unsequenced.



inclusion of sequencing information would mean that early-career researchers who are members of an Indigenous community, such as the Amhara people in Ethiopia, could negotiate to receive training in genome sequencing and analysis if researchers from South Africa, say, wanted to collect tissue samples from their country.

Lastly, everyone involved in the AfricaBP project – now and over the next decade – will engage local chiefs and other custodians of traditional knowledge in the project from the outset. One way for researchers to engage with local communities or Indigenous peoples is through monthly meetings with government officials involved in Africa's Access and Benefit Sharing National Focal Points. These individuals are specifically tasked with guiding compliance between the producers of biological resources, such as the Bedouin community in Egypt, and the users of those

resources, such as researchers at the Pasteur Institute of Tunis in Tunisia. Another way this could be achieved is through AfricaBP ethics committees surveying thousands of people in a particular community – such as through town-hall meetings, electronic messages or telecommunications.

Making it happen

Since 2009, \$22 million has been spent on building bioinformatics capacity across Africa through the Pan African Bioinformatics Network for H3Africa (H3ABioNet) project – including through training 150 researchers in core bioinformatics approaches and technologies. But around 10–15% of the trainees in this Africa-led project have relocated to North America or Europe, and there is no guarantee that they will return. What's more, H3ABioNet funding winds down this year, and there are few permanent positions for trained bioinformatics personnel in African institutions. Because of this, up to 50% of the researchers who have received training through H3ABioNet could leave Africa.

In the case of AfricaBP, around 600 eligible early-career African researchers (those pursuing PhDs or postdocs) will be granted 3-year fellowships over the next 10 years. They will be able to work with AfricaBP's global partners¹¹, such as the Wellcome Sanger Institute in Hinxton, UK, through exchange programmes. But they will be based mainly in national and regional AfricaBP facilities, to ensure that any skills they acquire are fed back into the continent.

Cloud-based computing and data storage will need to be coordinated to meet regional needs. Exchange programmes involving AfricaBP partners could help those regions or countries that lack resources; there are currently 87 genomic infrastructures in southern Africa, but only 8 in Central Africa⁷, for instance. These would be similar to the Newton International Fellowships, which enable early-career researchers from overseas to work for two years at a UK institution.

The 374 state-of-the-art Pacific Biosciences HiFi genome-sequencing machines that currently exist worldwide (as of 31 December 2021) can produce high-quality sequence data for more than 350 species per day¹². But although the city of Cambridge, UK, alone has 12 of these machines, there are only 2 in the entire African continent. Building genomics capacity on the ground is a huge challenge in Africa because of the difficulty of transporting intact samples in countries that have poor transport infrastructure and hot climates, and because of Africa's expensive and low-quality Internet service.

To achieve such a massive sequencing feat, African researchers need state-of-the-art genome technologies. They also need mobile (albeit less accurate¹³) sequencing

Comment

technologies that are less reliant on electricity and Internet connectivity, such as the Oxford Nanopore Technologies MinION machine. These are easily transportable and can be used in remote areas; they are roughly the size of a mobile phone¹³, whereas the Pacific Biosciences HiFi machines are about the size of a household refrigerator.

The 109 scientists spearheading AfricaBP are currently in discussion with leading institutions about the development of mobile sequencing platforms and integrated mobile laboratories. Encouragingly, portable, low-cost computing platforms, such as Raspberry Pi and eBioKit, are already being used in Africa, for instance at Makerere University in Kampala, Uganda, in bioinformatics training programmes.

We ask all African life-science agencies to join AfricaBP. We also ask the African Union Commission and the African Academy of Sciences to provide the core funds – US\$100 million per year for the next 10 years. In our view, this investment will be dwarfed by the economic and other pay-offs that will stem from AfricaBP-enabled innovations and discoveries.

The authors

ThankGod Echezona Ebenezer is a bioinformatician at the European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Hinxton, UK. **Anne W. T. Muigai** is a professor of genetics at the Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya. **Simplice Nouala** is head of the Division of Agriculture and Food Security at the African Union Commission, Addis Ababa, Ethiopia. **Bouabid Badaoui** is a professor at Mohammed V University in Rabat, Morocco. **Mark Blaxter** is head of the Tree of Life Programme at the Wellcome Sanger Institute, Cambridge, UK. **Alan G. Buddie** is research lead for microbiology and molecular biology at the Centre for Agriculture and Bioscience International (CABI), Egham, UK. **Erich D. Jarvis** is a professor of neurogenetics of language at the Rockefeller University, New York City, and an investigator of the Howard Hughes Medical Institute, Chevy Chase, Maryland, USA. **Jonas Korlach** is chief scientific officer at Pacific Biosciences, Menlo Park, California, USA. **Josiah O. Kuja** is a researcher at the Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya. **Harris A. Lewin** is a distinguished professor of evolution and ecology at the University of California, Davis, USA. **Roksana Majewska** is an extraordinary senior lecturer at North-West University, Potchefstroom, South Africa. **Ntanganedzeni Mapholi** is a researcher at the University of South Africa,

AfricaBP: structure and costs

The African BioGenome Project (AfricaBP) will involve researchers and organizations from all economic regions in the African Union, and will cost US\$100 million per year.

AfricaBP will convene 55 African researchers and policymakers from genomics, bioinformatics, biodiversity and agriculture – 11 for each of the 5 African Union geographical regions (northern, eastern, southern, central and western Africa). Another 165 people will be involved in the project (33 for each geographical region), including academic and industrial researchers, policymakers, and staff from governmental organizations, such as the National Institute of Agricultural Research of Morocco.

Ultimately, these people will feed genome sequences into various national or regional facilities. These include the National Gene Bank of Tunisia, which is using genetics to promote the conservation and sustainable use of Africa's plants, animals, fungi and protists, and the International Center for Research and Development on Livestock

Pretoria, South Africa. **Suresh Maslamoney** is a systems administrator at the University of Cape Town, South Africa. **Michèle Mbo'o-Tchouawou** is deputy director at African Women in Agricultural Research and Development, Nairobi, Kenya. **Julian O. Osuji** is head of the Molecular Genetics and Cytotaxonomy Unit at the University of Port Harcourt, Nigeria. **Ole Seehausen** is a professor of ecology and evolution at the Institute of Ecology and Evolution, University of Bern, Switzerland; and at Eawag, Swiss Federal Institute of Aquatic Science and Technology, Kastanienbaum, Switzerland. **Oluwaseyi Shorinola** is a research fellow at the International Livestock Research Institute, Nairobi, Kenya. **Christian Keambou Tiambo** is a scientist, co-lead of the Reproductive Technologies Programme, and Access and Benefits Sharing officer at the Centre for Tropical Livestock Genetics and Health–International Livestock Research Institute, Nairobi, Kenya. **Nicola Mulder** is head of the Computational Biology Division, University of Cape Town, South Africa. **Cathrine Ziyomo** is a senior scientist at the International Livestock Research Institute, Nairobi, Kenya. **Appolinaire Djikeng** is director of the Centre for Tropical Livestock Genetics and Health, Roslin Institute, University of Edinburgh, UK.

in the Subhumid Zone in Bobo-Dioulasso, Burkina Faso, which was established in 1994 to reduce poverty by improving food and nutritional security.

We estimate that producing high-quality reference genomes for around 105,000 endemic African species will cost around \$850 million to sequence, and around \$20 million to store, download, transfer and process the data (using high-performance computing and a mix of cloud platforms).

We reach this sum using our estimate of average genome sizes for plants and animals – 2.5 and 1.5 gigabases, respectively – and because the average cost per species per gigabase is \$4,200 (taking into account the price differences between North America and Africa for consumables, shipments and other overheads). We estimate the costs of sample collection, including permits, consultations and workshops, at \$41 million. Lastly, using the Newton International Fellowship as a benchmark, AfricaBP's early-career research fellowships will cost roughly \$90 million over a 10-year period.

e-mails: thankgod1980@yahoo.co.uk; awmuigai@fsc.jkuat.ac.ke; appolinaire.djikeng@ctlgh.org

A full list of co-signatories accompanies this Comment online (see go.nature.com/3t1654b).

1. Marks, R. A., Hotaling, S., Frandsen, P. B. & VanBuren, R. *Nature Plants* **7**, 1571–1578 (2021).
2. Tripp, S. & Grueber, M. *The Economic Impact and Functional Applications of Human Genetics and Genomics* (American Society of Human Genetics & TEconomy Partners, 2021).
3. Linder, H. P. *Front. Ecol. Evol.* <https://doi.org/10.3389/fevo.2014.00038> (2014).
4. Ghazal, H. et al. *Plant J.* **107**, 21–36 (2021).
5. Street, R. A. & Prinsloo, G. J. *Chem.* **2013**, 205048 (2013).
6. Gericke, N. S. *Afr. J. Bot.* **77**, 850–856 (2011).
7. Inzaule, S. C., Tessema, S. K., Kebede, Y., Ogwel Ouma, A. E. & Nkengasong, J. N. *Lancet Infect. Dis.* **21**, E281–E289 (2021).
8. Gheyas, A. A. et al. *Mol. Biol. Evol.* **38**, 4268–4285 (2021).
9. Lohay, G. G., Weathers, T. C., Estes, A. B., McGrath, B. C. & Cavener, D. R. *Ecol. Evol.* **10**, 11069–11089 (2020).
10. Xue, Y. et al. *Science* **348**, 242–245 (2015).
11. Lewin, H. A. et al. *Proc. Natl Acad. Sci. USA* **119**, e2115635118 (2022).
12. Rhie, A. et al. *Nature* **592**, 737–746 (2021).
13. Tedersoo, L., Albertsen, M., Anslan, S. & Callahan, B. *Appl. Environ. Microbiol.* **87**, e0062621 (2021).
14. Dandara, C. et al. *AAS Open Res.* **2**, 19 (2019).
15. Marshall, K. et al. *Front. Genet.* **10**, 297 (2019).
16. Mulder, N. et al. *Annu. Rev. Biomed. Data Sci.* **4**, 57–81 (2021).
17. Rees, C. A. et al. *BMJ Glob. Health* **6**, e006982 (2021).
18. Lange, M. et al. *GigaScience* **10**, giab084 (2021).
19. Yakubu, A., Munung, N. S. & De Vries, J. *AMA J. Ethics* **22**, E156–E163 (2020).

J.K. declares competing interests; see go.nature.com/3t1654b for details.