Enabling Africa's implementation of the Kunming-Montreal Global Biodiversity Framework through the African digital sequence information data bank

Achraf El Allali[1↓], Julien Alban Nguinkal[2↓], Sally Mueni Katee[3↓], Girish Beedessee[4], Bouabid Badaoui[5], Marietjie Botes[6], Semir Bechir Suheil Gaouar[7], Ichrak Hayah[5], Justin Eze Ideozu[8], Sadik Muzemil[9], Denye Nathaniel Ogeh[10], Abdoallah Sharaf[11,12], Kassahun Tesfaye[13,14], Anne WT Muigai[15,16], ThankGod Echezona Ebenezer[17]

[1]Bioinformatics Laboratory, College of Computing, Mohammed VI Polytechnic University, Benguerir, Morocco, [2]Bernhard-Nocht Institute of Tropical Medicine, Hamburg, Germany, [3]International Livestock Research Institute, Nairobi, Kenya, [4]Northumbria University, Newcastle-upon-Tyne, United Kingdom, [5]Mohammed V University in Rabat, Rabat, Morocco, [6]University of KwaZulu Natal, Durban, South Africa, [7]Applied genetics in agriculture, ecology and public health laboratory, University of Tlemcen, Tlemcen, Algeria, [8]MyAfroDNA, Port Harcourt, Nigeria, [9]Rothamsted Research, Harpenden, United Kingdom, [10]Eurekan Biotechnologies, Florida, United States, [11]SequAna Core Facility, Department of Biology, University of Konstanz, Konstanz, Germany, [12]Genetic Department, Faculty of Agriculture, Ain Shams University, Cairo, Egypt, [13]Bio and Emerging Technology Institute (BETin), Addis Ababa, Ethiopia, [14]Addis Ababa University, Addis Ababa, Ethiopia, [15]National Defence University-Kenya, Nakuru, Kenya, [16]Jomo Kenyatta University of Agriculture and Technology [17]African BioGenome Project (AfricaBP), Nairobi, Kenya

[↓]These authors contributed equally
***Correspondence:** achraf.elallali@um6p.ma and thankgod1980@yahoo.co.uk

Keywords: Digital Sequence Information, DSI, databases, data analysis platforms, data visualization platforms, data sharing platforms, ELSI, KMGBF, biodiversity, agriculture, genomics, bioinformatics, infrastructure, Africa

Abstract

The African BioGenome Project (AfricaBP) is a Pan-African effort aimed at sequencing the genomes of 105,000 African endemic and indigenous species to support food systems, conservation, and ensure data-sharing and equitable benefits. This effort aligns with the Kunming-Montreal Global Biodiversity Framework (KMGBF), which aims to prevent or mitigate biodiversity loss while facilitating equitable access and benefit-sharing from genetic resources and Digital Sequence Information (DSI) and securing adequate technical and scientific cooperations. The AfricaBP Open Institute for Genomics and Bioinformatics (AfricaBP Open Institute) is the knowledge exchange programme of the AfricaBP which aims to overcome infrastructural barriers through the development of technology and infrastructure. A key component of AfricaBP Open Institute's vision is the establishment of the African Digital Sequence Information Data Bank for Biodiversity and Agriculture (African DSI Data Bank), a federated platform for storing, analyzing, visualizing and sharing DSI data across the African continent. The African DSI Data Bank will address the current fragmentation of DSI across African institutions by linking existing databases and resources while ensuring compliance with regional and global standards. It will use a federated model, leveraging existing (and new) infrastructures across Africa, that allow institutions and countries to retain data sovereignty while adhering to national, regional, and international access and benefit-sharing regulations. Through a proposed Global Access Point (GAP), researchers will be able to gain equitable access to sequence data and metadata via a decentralized network. Furthermore, to understand the current landscape of biodiversity and agricultural DSI databases, analyses, visualization, and data sharing platforms, AfricaBP Open Institute conducted a survey across Africa, and recorded 161 responses. Although the majority of these participants shared common challenges such as limited infrastructure, funding, and capacity building, the overwhelming indication was that they support an African-based DSI platform through an inclusive governance model. Consequently, we describe the proposed roadmap for the creation of an African DSI Data Bank that includes African DSI federated database, visualization, analysis, and sharing platforms, as well as the ethical, legal, social, KMGBF, and sustainability considerations associated with such an infrastructure.

Background

The African BioGenome Project (AfricaBP) (https://africanbiogenome.org/) is a coordinated Pan-African effort established in 2021 to locally sequence the genomes of 105,000 endemic biological species (plants, animals, fungi, protists and other eukaryotes) to improve food systems, conservation, and data sharing and benefits (Ebenezer, *et al.,* 2022). This will be an estimated 203,000 gigabases of DNA sequence (Sharaf, *et al.,* 2023).

The Kunming-Montreal Global Biodiversity Framework (KMGBF) aims to halt and reverse biodiversity loss, put nature on a path to recovery for the benefit of people and the planet by conserving and sustainably use biodiversity and by ensuring the fair and equitable sharing of benefits from the use of genetic resources, while providing the necessary means of implementation (UNEP Convention on Biological Diversity, 2022). The AfricaBP goals and activities closely align with the objectives of the KMGBF, including facilitating appropriate access to (and benefit from) genetic resources and Digital Sequence Information (DSI), and securing adequate technical and scientific cooperation to fully implement the KMGBF (UNEP Convention on Biological Diversity, 2022, Katee, *et al.,* 2024). DSI is a placeholder terminology under the CBD and refers to genetic and/or biochemical composition of genetic resources such as nucleic acids and macromolecules as well as observational data that provides contextual information such as taxonomy, gene expression, ecological relationships, and modalities of use (UNEP Convention on Biological Diversity, 2016).

The AfricaBP Open Institute for Genomics and Bioinformatics (AfricaBP Open Institute) is the knowledge exchange programme of the AfricaBP which aims to overcome infrastructural barriers through the development of technology and infrastructure. Globally, infrastructures and databases supporting biodiversity and agriculture have expanded significantly (Wilkinson, *et al.,* 2016, Antognoli, *et al.,* 2017, Gullotta, *et al.,* 2023, Gioti, et al., 2024, Mc Cartney, *et al.,* 2024). Despite these advancements, gaps in data coverage and standardization remain, particularly for under-represented species and geographical regions such as Africa. In Africa, efforts to build databases and infrastructure for biodiversity and agriculture are growing, but face challenges such as limited funding, political and regulatory barriers, and fragmented infrastructures, further hampering cross-border collaborations and hindering comprehensive biodiversity data integration (Kumwenda, *et al.,* 2017, Chapman, *et al.,* 2022). These challenges underscore the need to harness the established Pan-African DSI initiative, AfricaBP Open Institute, which fosters continental and regional collaboration and knowledge exchange. Such efforts are crucial for integrating African biodiversity and agricultural DSI into the global research ecosystem (Ebenezer, *et al.,* 2022).

One of the key aims of the AfricaBP Open Institute is the creation of the African Digital Sequence Information Data Bank for Biodiversity and Agriculture (African DSI Data Bank), a federated platform that connects networks of African institutions to facilitate storing, analyzing, visualizing, and sharing DSI data in line with their national framework while promoting responsible and equitable open science (Cengiz, *et al.,* 2023, Katee, *et al.,* 2024; Sharaf, *et al.,* 2024). DSI across Africa are often dispersed among universities, research centers, and government agencies (Text Box 1), creating inefficiencies in coordinated collaborations, partnerships and knowledge-sharing. AfricaBP Open Institute seeks to consolidate the fragmented DSI data currently spread across multiple African institutions by facilitating access through a unified platform called the Global Access Point (GAP). This is also inline with the African Union's (AU) Science, Technology, and Strategy for Africa 2024, whereby building or upgrading research infrastructures is identified as one of the four mutually reinforcing pillars of the strategy (African Union, 2024).

African biodiversity remains largely underrepresented in global genomic databases, limiting scientific research and innovation in biodiversity conservation and sustainable agricultural improvement (Ebenezer, *et al*., 2022). The African DSI Data Bank aims to fill gaps such as limited data of African plants and animal genomes in databases (Sharaf, et al., 2023, 2024). The development of this federated infrastructure could be co-funded between African institutions and governments, private sectors, and international partners (Ebenezer, *et al.,* 2022), ensuring that underrepresented taxa and geographical regions of the world in genomics could better work on their priorities (Fernandez-Prada, *et al.,* 2024).

Here, we delineate a roadmap for an African DSI infrastructure, highlighting four key actions (federated database, analysis, visualization, data sharing platforms) Africa must undertake to maximize benefits from the KMGBF and advance global DSI efforts. We present our ongoing efforts in assembling a network of African institutions with existing DSI-related infrastructure capabilities for biodiversity and agriculture as well as provide an initial steer for African institutions with potential capabilities but without DSI infrastructures. We define a proposed implementation roadmap, informed by our experiences in the AfricaBP and its Open Institute as well as insights from our survey of 161 African researchers (Figure S1 - S7), for an African DSI Data Bank as well as its significance in the era of KMGBF such as solving DSI data fragmentations and enabling appropriate access in line with national frameworks. Finally, we propose the adoption of this roadmap, or its adjusted/optimized version, into the permanent structures of African Union science agencies for biodiversity and agriculture as well as recommendations to the African Group on the implementation of the KMGBF across Africa.

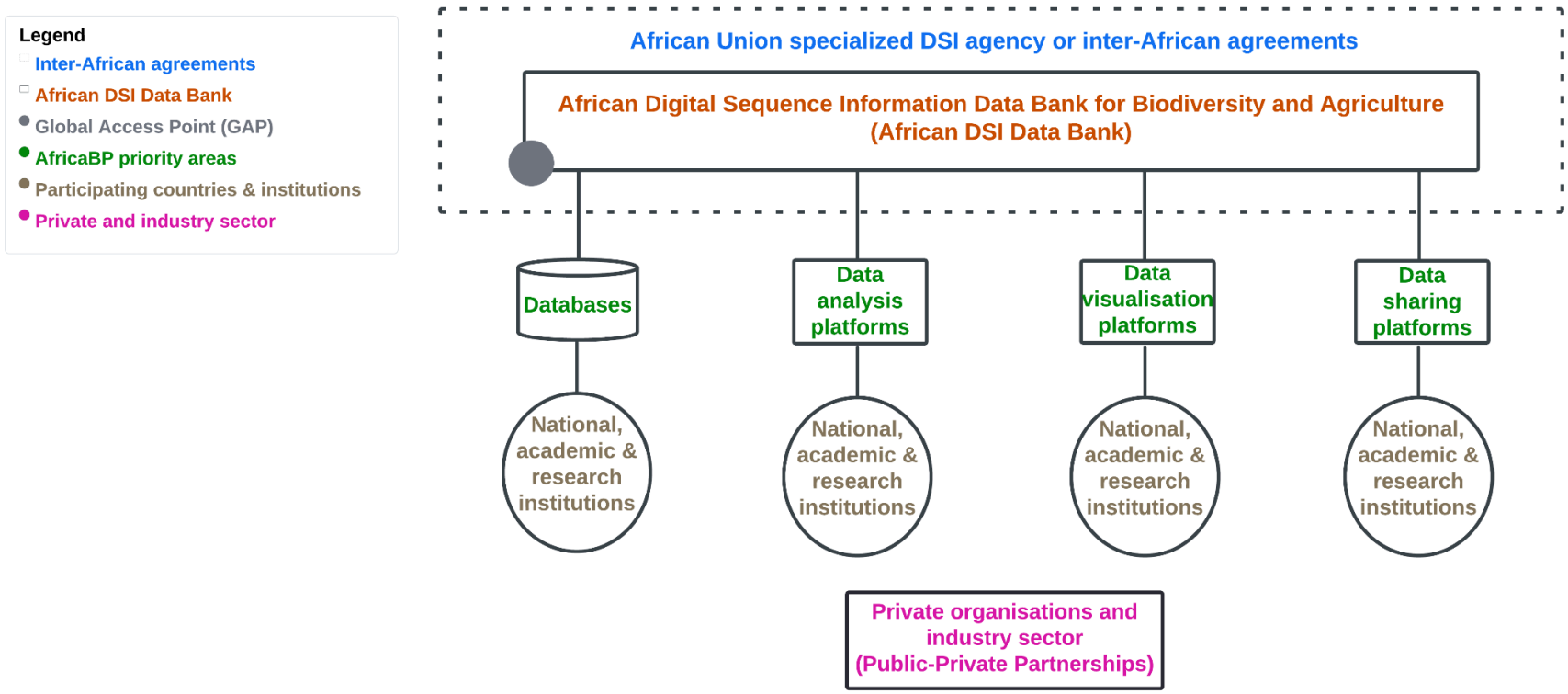
Legend
Inter-African agreements
African DSI Data Bank
Global Access Point (GAP)
AfricaBP priority areas
Participating countries & institutions
Private and industry sector
African Union specialized DSI agency or inter-African agreements
African Digital Sequence Information Data Bank for Biodiversity and Agriculture (African DSI Data Bank)
Databases
Data analysis platforms
Data visualisation platforms
Data sharing platforms
National, academic & research institutions
National, academic & research institutions
National, academic & research institutions
National, academic & research institutions
Private organisations and industry sector (Public-Private Partnerships)

Figure 1: Five priorities of the African BioGenome Project to enable Africa implement the Kunming-Montreal Global Biodiversity Framework (KMGBF) through the African Digital Sequence Information Data Bank for Biodiversity and Agriculture (African DSI Data Bank). A hub and spoke model showing proposed configuration and delivery of the African DSI Data Bank through inter-African agreements and current (or new) DSI infrastructure across Africa. Legends: Dotted line signifies DSI specialized institution under the African Union or any inter-African governmental agreements. Text in magenta color within the rectangular box depicts the central coordinating hub that facilitates DSI data storage, analysis, visualization and sharing through the Global Access Point in circular steel color in the bottom left of the rectangle. This is placed within the dotted lines to show data deposition and sharing in line with national frameworks such as national biodiversity strategies and action plans. Green texts depict databases, data analysis, visualization and sharing platforms that contribute to the African DSI Data Bank. Brown texts in circular shape show African national, academic and research institutes DSI databases, data analysis, visualization and sharing platforms as spokes in this distributed model. Pink texts in rectangular shape show public-private partnerships in the delivery of the African DSI Data Bank's objectives.

Five priorities

The primary aim of the African DSI Data Bank is to harmonize and centralize access to the rapidly increasing number of DSI by creating a network of African institutions with existing databases, data analysis, visualization and sharing platforms focused on African plants, animals, and other non-human eukaryotes, while also guiding and supporting those with the potential to develop these capabilities. This will be done through the development and implementation of GAP, a unified and central point for accessing distributed African databases, data analysis, visualization and sharing platforms. By doing so, and aligning with the KMGBF, the African DSI Data Bank will directly address the critical data gap hindering the advancements of biodiversity genomics efforts across the continent.

While recognizing the challenges of establishing a comprehensive platform, we are embarking on a phased approach, focusing initially on creating a network of African institutions with existing infrastructural resources and gradually expanding capabilities such as equipping and establishing, through legislative instruments, a physical infrastructure as the initiative evolves. To address the critical needs across Africa, the African DSI Data Bank has five priorities which involve creating a network of:

1. *African DSI database platforms:* The first priority of the African DSI Data Bank is to identify, link, and integrate existing (or new) African DSI database platforms, and prioritize on the principle that it will be impossible to aggregate all the fragmented DSI data types in Africa (Text Box 1, Figure S2) (Balaji & Terry, 2015) into a central repository due to multiple reasons, including regulatory. Hence, the need for a federated database configuration that could be mainstreamed into a central coordinating database established through legislative instruments under the African Union or any of its specialized scientific agencies to ease regulatory hurdles, becomes necessary (Figure 1) (Sharaf, *et al.,* 2024). This will enable the storage of diverse data types from different species, reducing fragmentation and ensuring that African data standards are compatible with international platforms such as those hosted by major public databases (Arita, *et al.,* 2021, INSDC, 2024, Vecchia, 2024) for data submission, curation, scalability, archiving, enabling easy DSI data exchange and collaborations.

2. *African data analysis platforms:* The power and scale of the African DSI Data Bank will come from its ability to provide a cost-free, integrated and easily accessible federated analysis platform for African researchers on the African continent as the sequencing of new African species produces vast amounts of genomic data. Access to advanced computing resources in Africa is often concentrated in a few well-funded institutions, leaving many researchers without

the necessary tools to perform cutting-edge research (Krauss & Loebbecke, 2021). By creating a unified portal to help researchers access available resources, African researchers can overcome these hurdles and realize the full potential of Africa's vast genetic and biodiversity data (Haas, 2015). For example, researchers in some African countries have less access to computing platforms for research and scientific purposes, and as such, increased access to analysis platforms will directly result in increased collaborations and DSI data depositions (Langmead & Nellore, 2018), especially through incorporation of essential bioinformatics tools and customized data analysis pipelines tailored to their specific needs.

Therefore, the African DSI Data Bank aims to identify, link, and integrate existing (or new) African DSI data analysis platforms, through a distributed user-friendly, cloud- and super- computing infrastructure to support data analysis accessible through GAP. This will provide African researchers with direct access to computational resources such as cloud-based solutions or shared HPC infrastructures (Pinthong, *et al.,* 2016, Rowe, *et al.,* 2019, Langmead & Nellore, 2018), fostering a more independent and collaborative research environment while empowering African researchers to conduct their own data analysis and contribute to the goals of AfricaBP Open Institute. Through the GAP interface, African researchers can request computational resources, resource type, usage, and duration, and this request will be processed, granting or denying access based on availability and other criteria, and ultimately, managing resource allocation and usage monitoring.

3. *African DSI data visualization platforms:* To make complex genomic information accessible, the African DSI Data Bank will be designed to include interactive visualization platforms that provide clear and interpretable representations of genomic data. This platform will allow African researchers to explore genetic diversity, biodiversity patterns, and sequence data through an integrated sequence viewer and customizable visualizations**.** Tools for visualizing genomic data, comparative analysis including phylogenetic tree viewers, DSI browsers, and network graphs (see Figures S1 - S5, Text Box 1), will be incorporated and accessible via the GAP interface.

4. *African DSI data sharing platforms:* DSI data produced by one researcher or organization are relevant when made available and utilized by end-users (Wood-Charlson, 2023), and this needs to be shared in line with national frameworks to promote responsible and equitable open science (Wright, *et al.,* 2015, Byrd, *et al.,* 2020, Hudson, *et al.,* 2020). With the growing scale of large genomic sequencing efforts globally, and particularly, the AfricaBP initiative to

sequence 105,000 endemic African species (Ebenezer, *et al.,* 2022), it is very critical to ensure an accessible and reliable system that provides appropriate access to DSI in line with national regulations. When scientists can appropriately access African genetic resource DSI data (UNEP Convention on Biological Diversity, 2022), it enables collaboration that can lead to discoveries and innovations in medicine, food security and conservation efforts (Wright, *et al.,* 2015, Adler Miserendino, *et al.,* 2022). A coordinated African DSI data-sharing platform, for example, facilitated through GAP will enable collaboration, maximize data sharing (Sharaf, *et al.,* 2023, Katee, *et al.,* 2024) and support large-scale data-intensive research, management, and policy (Abarenkov, *et al.,* 2023). However, while we encourage institutions to share DSI data, the ultimate decision to do so rests with the institutions, their national frameworks and domestic DSI provisions in line with KMGBF.

5. *Engage African countries and policy makers for the establishment of a specialized institution on DSI under the African Union or its parastatal:* The AfricaBP Open Institute will promote and drive discussions for the establishment of a permanent, specialized institution on DSI under the African Union (or its parastatal). In the long-term, this specialized institution, established through new or existing legislative instruments, could coordinate, manage, and create new large-scale DSI data storage, analysis, visualization and sharing for African member states, especially for underrepresented African countries without DSI management and use capabilities.

The African landscape of DSI databases, sharing, analysis and visualization platforms

Extensive insights on the landscape of global DSI databases, tools for analysis and visualizations, data-sharing platforms for biodiversity and agriculture, have sufficiently been reported (Agrawal, *et al.,* 2012, Ambrosino, *et al.,* 2017, Roumpeka, *et al.,* 2017, Ambrosino, *et al.,* 2019, Krassowski, *et al.,* 2020, Gadelha, *et al.,* 2021, Chelliah, *et al.,* 2022, Lewin, *et al.*, 2022, Waterhouse, *et al.,* 2021, Wysel, *et al.,* 2021, 2023, Gioti, *et al.,* 2024). However, little is known about the landscape of fragmented African DSI databases, data analysis, visualization and data sharing platforms. Here, we summarize these efforts using African institutions and projects as case studies (also see Text Box 1 and Figures S1 - S5):

1. *University Mohammed VI Polytechnic (UM6P), Morocco:* The Bioinformatics Laboratory at UM6P in Morocco is establishing a platform that includes plant

biological databases and analysis web servers to advance the understanding and conservation of biodiversity and accelerate plant breeding programs (https://bioinformatics.um6p.ma/services/). By integrating genomics and bioinformatics, this platform provides comprehensive insights into the genetic makeup of plant species, enabling more effective conservation strategies and the sustainable management of genetic resources.

The platform includes specialized databases developed for different aspects of plant genomics relevant to biodiversity and plant breeding. These databases include: PlantLAI (Mokhtar, *et al.,* 2023) which assesses the quality of genome assembly for all available plant genome sequences, providing a key metrics for downstream analysis of plant genomes, PltRNAdb (Mokhtar & El Allali, 2022) focuses on transfer RNAs (tRNAs) in different species, sheds light on the functional roles of tRNAs that are crucial for the study of evolutionary biology and adaptation, and PlantLTRdb (Mokhtar, *et al.,* 2023b) and CicerSpTEdb (Mokhtar, *et al.,* 2021) which provide detailed information on long terminal repeat retrotransposons (LTR-RTs) and transposable elements (TEs) that are important for understanding genome structure and evolutionary processes.

The analysis web servers that complement the platform offer real-time access to tools such as MegaSSR (Mokhtar, *et al.,* 2023a) and MegaLTR (Mokhtar & El Allali, 2023) for analyzing SSRs and LTR-RTs. These tools include user-friendly interfaces to support large-scale genomic studies. The platform enables the rapid identification of genetic traits associated with species resilience and adaptation, providing critical data to support decision making on conservation strategies and breeding programs (Hassan, *et al.,* 2024) (see https://bioinformatics.um6p.ma/statistics/ for statistics).

2. *Bio and Emerging Technology Institute (BETin), Ethiopia*: In 2016, the Ethiopian government established the Bio and Emerging Technology Institute (BETin), formerly known as the Ethiopian Biotechnology Institute (EBTi), under Regulation No. 388 (Federal Negarit Gazette, 2016). BETin has a primary mandate of leading, coordinating, and conducting national biotechnological research and development endeavors across all sectors, especially in areas of omics (genomics, metagenomics, transcriptomics, proteomics, metabolomics, etc.) and bioinformatics (Federal Negarit Gazette, 2016). Its responsibilities include providing bioinformatics services and infrastructures such as data storage, supercomputing, servers, analysis platforms, maintaining and developing databases and algorithms, and mainstreaming bioinformatics in higher education (BETin, 2024).

3. *South African national biodiversity and computational institutions:* The South African National Biodiversity Institute (SANBI), as a knowledge management institute, has a mandate to collect, generate, process, coordinate and disseminate information about biodiversity and sustainable use of indigenous biological resources, and maintain biodiversity databases (Parker-Allie, *et al.,* 2023), for example, a Service Oriented Architecture that integrates distributed, separately deployed and maintained software components that may be controlled by various owners (Daly & Ranwashe, 2023). Furthermore, the South African National Bioinformatics Institute (SANBI), was founded in 1996 under the University of the Western Cape with a primary focus for the development and implementation of computational methodologies which allow researchers to fast-track their genomics data analyses, including leading the conceptualization of the data archive for disease outbreak datasets in partnership with the Africa CDC, a continental data platform called Agari (SANBI, 2023). Additional institutions include the Council for Scientific and Industrial Research (CSIR) National Integrated Cyberinfrastructure System (NICIS) which manages the HPC Ecosystems Project by providing HPC access to local research institutes in South Africa and within partner countries (Beukman, *et al.,* 2024, Johnston, *et al.,* 2024)

4. *African traditional medicinal plants and natural products databases:* Three databases contain Natural Products (NPs) from the African flora and fauna (Sorokina & Steinbeck, 2020, Sorokina, *et al.,* 2021). The Northern African Natural Products Database contains over 4500 NPs from plants, endophytes, fungi and bacteria (Ntie-Kang, *et al.,* 2017), the South African Natural Compound Database contains over 1012 NPs isolated from South African biodiversity (Diallo, *et al.,* 2021), while the Mitishamba database contains 1100 NPs isolated from Kenyan plants (Derese, *et al.,* 2015). Additionally, African Traditional Medicine (ATM) is a rich and developed traditional medicine, and they also contain a number of databases focusing on NPs from plants used in traditional medicines on the African continent (Sorokina & Steinbeck, 2020). These include, amongst others (Sorokina & Steinbeck, 2020, Sorokina, *et al.,* 2021), Afrotryp (Ibezim, *et al.,* 2017), the Central African Medicinal Plants database (Ntie-Kang *et al.,* 2014), and the Ethiopian Traditional Medicine Database (Bultum, *et al.,* 2019). However, not all African countries host ATM databases. For example, scientists in Nigeria have proposed the establishment of a national or regional repository for medicinal plant species, extracts, and natural compounds to improve the value of small-scale efforts (Ungogo, *et al.,* 2020). It is also noteworthy that some African NP databases, such as Mitishamba (http://mitishamba.uonbi.ac.ke), could not expand beyond its national borders

due to funding (Woldeamanuel, 2014) and the database is not currently maintained or accessible, leading to a significant loss of data on NPs (Sorokina & Steinbeck, 2020) - however, this is not peculiar to Africa (Fernandez-Prada, *et al.,* 2024). Hence, supporting AfricaBP Open Institute's proposition to integrate all African DSI-related databases to a continental coordinating platform for sustainability and continuity (Figure 1).

ELSI and KMGBF Considerations

Here, we discuss eleven actionable recommendations that will maximize the benefit of the African DSI Data Bank to the African people (and globally) and create measurable impacts across Africa:

1. *Consent, recognitions, and intellectual property:* Indigenous and local communities often possess valuable traditional knowledge and genetic resources essential for biodiversity conservation (Anand, *et al.,* 2023). To respect their rights and contributions, the African DSI Data Bank must include mechanisms for obtaining and documenting consent (Fredriksson, 2021). This involves creating protocols for engaging with communities, explaining how their knowledge and resources will be used, and ensuring that intellectual property rights are acknowledged and protected (Laltaika, 2022). Additionally, benefit-sharing agreements must be in place to ensure fair compensation and recognition (Humphries, *et al*., 2022). Similar models have been successfully used, such as travel passports (or identification documents) for research mobility, (Katee, *et al.,* 2024), Biodiversity Passport for biodiversity access monitoring (Daly & Ranwashe, 2023) and GA4GH Passports and the Crypt4GH file container in the human genomics community to protect minority communities and offer culturally acceptable processes to encourage data contributions from underrepresented communities (Oliva, *et al.,* 2024).

2. *Data privacy, access and security:* The African DSI Data Bank will handle a vast amount of sensitive data, including genetic sequences and traditional knowledge, and would aim to provide appropriate access in line with the objectives of the KMGBF and prevent misuse (UNEP Convention on Biological Diversity, 2022, Bonomi, *et al.,* 2020). Implementing robust data security measures such as secure access controls, is crucial (Wan, *et al.,* 2022, Shih, *et al.,* 2023). Privacy policies and benefit-sharing mechanisms should be developed and implemented to ensure that data is handled in compliance with international standards and that

personal and sensitive information is anonymized or aggregated to prevent misuse and misappropriation (Bhat & Taiwo, 2015).

3. *Compliance with international agreements:* The African DSI Data Bank must comply with international agreements such as Nagoya Protocol and the Convention on Biological Diversity (CBD) that govern access to genetic resources and benefit-sharing (Knight, *et al.,* 2023). This involves integrating provisions of the Nagoya Protocol, which requires obtaining prior informed consent and ensuring fair and equitable sharing of benefits derived from genetic resources (Ivey, *et al.,* 2023). Ensuring that all international obligations are met by facilitating the documentation of consent and benefit-sharing agreements (Lawson, *et al.,* 2024), will be a priority for the data bank.

4. *Harmonization of legal frameworks:* Africa is home to a range of legal frameworks concerning biodiversity and genetic resources, which can vary significantly between countries (Zerbe, 2005). To create a functional and cohesive data bank, efforts must be made to harmonize these legal frameworks (Masehela & Barros, 2023). This includes standardizing policies and procedures for data management and benefit-sharing, and ensuring that the data bank's operations align with national laws while supporting cross-border co-operations (Gbadegesin & Gbadamosi, 2024).

5. *Policy for bilateral benefit-sharing:* Policies for bilateral benefit-sharing should be designed to ensure that benefits derived from the data bank are distributed fairly among all stakeholders. This includes creating transparent procedures for determining and distributing bilateral benefits, and involving all relevant parties in decision-making processes to ensure equity and fairness (Wynberg, 2023).

6. *Multilateral benefit-sharing:* Multilateral benefit-sharing involves distributing benefits from genetic resources among multiple stakeholders, including countries, communities, and organizations, and discussions are currently ongoing at the CBD (Kreiken & Arts, 2024). If a multilateral benefit-sharing mechanism is agreed, the African DSI Data Bank must establish clear frameworks and agreements for managing these benefits. This includes developing protocols for equitable distribution, facilitating collaboration between stakeholders, and ensuring that benefits are shared fairly according to the contributions and needs of all parties involved (Ebert, *et al.,* 2023).

7. *Equitable access and inclusion:* To maximize the utility of the African DSI Data Bank, it must be accessible to a diverse range of stakeholders and appropriate access must be facilitated and granted in line with the objectives of the KMGBF.

This includes providing access to local communities, researchers, policymakers, and global stakeholders (Kennedy, *et al.*, 2022, Mbelebele, *et al.,* 2024). Policies should promote inclusivity by removing barriers to access, such as language, regulations and technological limitations (Yanou, *et al.,* 2023), and ensuring that all stakeholders can contribute to and benefit from the African DSI Data Bank.

8. *Capacity building and education:* Large-scale effective use of the African DSI Data Bank requires large-scale training and capacity building for local communities, researchers, and policymakers (Sharaf, *et al.,* 2023, 2024). Investing in educational programs and workshops, as being deployed by the AfricaBP Open Institute, can enhance the skills needed to manage and utilize the data bank effectively (Sharaf, *et al.,* 2023, 2024). This includes training on data interpretation, conservation strategies, and the integration of database information into decision-making processes (Browne, *et al.,* 2024).

9. *Integration with national policies:* The African DSI Data Bank will complement and support national biodiversity policies and strategies, especially since NBSAPs is extremely important in the implementation and achievement of the objectives of KMGBF (Katee, *et al.,* 2024). This involves aligning the database's goals with national priorities, ensuring that it contributes to existing conservation efforts, and facilitating coordination between national and regional initiatives (Sharaf, *et al.,* 2024).

10. *Sustainability and funding:* It is important that the African DSI Data Bank shows diversity and equitability in its funding mechanisms for its sub-projects and services, and should include scientific funding agencies to enable scientific impacts through international partnership commitments, human rights approach and national funding mechanisms in line with the objectives of KMGBF for indigenous peoples and local communities, and finally, public-private partnerships for long-term sustainability (Figure 1) (Katee, *et al.,* 2024, UNEP Convention on Biological Diversity, 2022).

    This type of multi-partner arrangements will prevent resources for underrepresented regions, such as Africa, from going 'dark' when one funding source pulls out as seen with VEUPATHDB, a genomic resource for neglected tropical diseases (Fernandez-Prada, *et al.,* 2024). Some percentages of the KMGBF multilateral funds received by each African country could be channeled into establishing and funding the federated African DSI data bank (Katee, *et al.,* 2024). Such platforms could then be connected with other global platforms (e.g. INSDC and the UNEP ABS Clearing house) in line with the national frameworks, thereby maximizing responsible and equitable open science. The percentage

contributions from African countries could also count towards the 1% GDP agreed by African member states for contributions to research and development (Adepoju, 2022, African Union, 2024).

11. *Legal Framework for Data Sharing:* Cross-border data sharing requires clear legal agreements and protocols to manage data transfer and use between countries (Cengiz, *et al.,* 2024). The African DSI Data Bank hub will include agreements that address legal and logistical issues, ensuring that data sharing complies with national and international regulations.

Text Box 1

AfricaBP Open Institute ran a survey to better understand the current landscape of biodiversity and agricultural DSI databases, analysis, visualization and sharing platforms, and this received 161 responses (Figure S1 - S7). Here, we summarize the responses from this survey:

1. *Metadata of respondents:* The survey received substantial engagement from African institutions, with the highest participation being from Nigeria (22.4%), followed by Ethiopia (19.3%), Kenya (9.3%), and Cameroon and South Africa (5.0%) respectively (Figure S1). The majority of respondents were from universities (62.1%), with 28.0% representing research institutions, underscoring the central role of academic and research entities in advancing DSI initiatives in Africa. Among these respondents, researchers made up 61.5%, with smaller proportions being principal investigators (6.2%) and bioinformaticians (6.2%) . Funding for their research efforts predominantly came from academic institutions (28.8%) and international grants (15.8%), while government funding accounted for 28.4% (Figure S1). Private sector funding and contributions from African science or policy agencies were lower at 8.4% respectively..

2. *DSI database platforms in Africa:* Plants (16.7%) and animal (16%) species are primarily handled by the respondents databases. The reliance on institutional data centers (32.6%) and local servers (31.1%) for storage, alongside the growing adoption of hybrid solutions (14.7%) that merge local and cloud storage (Figure S2), reflects the evolving nature of data storage infrastructure across the continent. Eleven respondents (7.6%) indicated that they use African-specific databases and 50% of the respondents indicated that the file formats they interact with are FASTA/FASTQ (Figure S2).

3. *DSI data sharing platforms in Africa:* In terms of accessibility, the majority of institutions (43.7%) indicated that their databases are accessible through open access, fostering collaboration and transparency, while 23.0% operate under restricted access due to ethical or legal constraints, and 16.4% have formal data-sharing policies in place (Figure S3). Consortium-based access is used by 10.9%, while 5.5% of respondents do not share their data (Figure S3). Data-sharing platforms play a critical role, with 60.7% of respondents utilizing institutional repositories outside Africa, and 27.9% using African-specific platforms (Figure S3). However, significant barriers to data sharing persist, including infrastructure limitations, co-development of training, ethical and legal constraints, and restrictive institutional policies. These challenges highlight an urgent need for capacity-strengthening initiatives and policy reforms to improve data sharing across the continent.

4. *DSI data analysis platforms in Africa:* Respondents indicated that sequence alignments, phylogenetics and population genomics are their most regular analysis (14.3%) while their most used analysis tool is R (18.6 %) and Python (6.2%) (Figure S4), suggesting that efforts to advance the analysis of genomics in Africa needs to incorporate these two programming languages. Local servers (27.5%) and cloud (16.7%) are major data analysis platforms used in Africa, and respondents (19.4%) indicated that regular workshops, in-house and external training programs would be advantageous to them (Figure S4).

5. *DSI data visualization platforms in Africa:* The respondents indicated that the most commonly used data visualization tools are genomic maps, integrated genomics viewer, Ensembl browser and phylogenetics viewers, and these tools are used to generate visualizations for genomic maps and phylogenetic trees (Figure S5). There is a high amount of pathway tools used for metabolomics visualization followed by metabolomics atlas (Figure S5).

6. *Needs and regulatory compliance for African scientists in biodiversity and agriculture:* More than 70 % of the respondents are somewhat familiar with the concept of a digital sequence information data bank for biodiversity and agriculture and there was no doubt that such a repository would be extremely valuable (Figure S6).

   The sentiment analysis further underscores the challenges in developing and maintaining DSI platforms (Figure S6). An impressive consensus of 97.5% of respondents consider such a repository important (Figure S6). This demonstrates the critical role that a local repository could play in safeguarding Africa's biodiversity and agricultural research. Similarly, 75.2% of respondents view data

sovereignty as "very important," and 23.0% rate it as "important" (Figure S6). Concerns about data ownership further highlight the importance of national governance and oversights over DSI platforms. In terms of implementation priorities by the AfricaBP Open Institute, 38.5% of respondents called for an integrated platform that supports data storage, analysis, visualization, and sharing, while 24.2% emphasized the need for strong data analysis tools (Figure S6).

Respondents also expressed optimism about the platform's potential to address pressing local challenges such as climate change, biodiversity conservation, and food security (Figure S6). The survey further underscores the importance of inclusive governance in developing and managing the DSI platform. Key stakeholders mentioned by the respondents include researchers, government agencies, and African institutions, with strong support for a multi-stakeholder approach that also involves local communities, policymakers, and private sector partners. This inclusive governance model is seen as essential for the African DSI Data Bank's long-term sustainability and effectiveness.

First steps

UM6P, home to Toubkal, the fastest supercomputer in Africa (DELL Technologies, 2022, UM6P, 2021), has confirmed interests to partner with AfricaBP Open Institute and offer African researchers working on plants, animal and microbial genomics in Africa the opportunity to apply for access to up to 10,000 CPU hours per project, including support for the design and implementation of GAP. BETin in Ethiopia, on the other hand, has confirmed interest to partner with AfricaBP Open Institute on the African DSI Data Bank as part of its existing national bioinformatics mandate for Ethiopian researchers only.

The first, and immediate, step in the creation of the African DSI data bank is the establishment of a collaboration agreement between the AfricaBP Open Institute and its confirmed members such as UM6P, BETin, other interested African and international institutions and partners, and the African Union (or any of its DSI-related scientific parastatal).

We anticipate that additional African and international institutions and organizations will join this effort by connecting their Africa-based databases, offering access to their HPC or cloud computing, based on their own sharing model. The GAP central resource will then be set up to provide researchers with easy access to the most up-to-date list of participating institutions. Interested parties can express their interest by contacting the corresponding authors with clear SMART (specific, measurable, achievable, relevant,

and time-bound) contributions aligned with the goals of the African DSI Data Bank. AfricaBP Open Institute will prioritize the establishment of federated database and analysis platforms as an initial effort as well as part of an outcome from the survey (see Text Box 1, Figure S2 and S7).

Ultimately, the success of the African DSI Data Bank depends on establishing it through legislative frameworks across Africa to effectively implement KMGBF (Figure 1). So, in parallel, the AfricaBP Open Institute will engage in discussions for the creation of a permanent, specialized institution within the African Union (or related scientific agency) to manage coordination of the decentralized databases / storage of DSI from African member states.

Conclusion

Overall, the creation of an African DSI Data Bank that integrates existing plants, animals, fungi, protists, and eukaryotic microbial databases, analysis, visualization and sharing platforms in Africa is a critical advancement for biodiversity conservation. By harnessing the power of genomics and fostering collaborations, these tools provide a pathway to a sustainable future in which the rich genetic diversity of African biodiversity is conserved and responsibly utilized for the benefit of ecosystems and human communities (Lewin, *et al.,* 2022, Ebenezer, *et al.,* 2022). However, this model requires resourcing and funding from participating countries, institutions, and organizations. To ensure the long-term success of these efforts, securing sustainable funding for infrastructure maintenance, data curation and ongoing system improvement is critical. In addition to funding, a governance framework needs to monitor progress towards actualization of the KMGBF. As a moral responsibility, this body will also ensure that indigenous knowledge and contributions of local communities are respected and their rights to their genetic resources are protected. The tasks of this body could also include monitoring and ensuring compliance with national frameworks (Katee, *et al.,* 2024).

Acknowledgements

This work has benefited from the support and resources provided by the African BioGenome Project (AfricaBP). We would like to thank Pandam Salifu, Kwame Nkrumah University of Science and Technology (KNUST), Ghana, for making initial input in the earlier discussions for this paper.

Competing interests

JEI is Principal Investigator at MyAfroDNA, DNO is a Consultant Bioinformatician at Eurekan Biotechnologies.

Authors contributions

S/N	Author name	Conceived idea	Literature review	Survey development and collection	Survey analysis	Production of manuscript diagram	Manuscript drafting	Manuscript revision	Supervision - oversight and leadership responsibilities	Journal correspondence	Manuscript review
1	Achraf El Allali	✔					✔				✔
2	ThankGod Echezona Ebenezer	✔	✔	✔		✔		✔	✔	✔	✔
3	Julien Alban Nguinkal		✔	✔							✔
4	Semir Bechir Suheil Gaouar	✔	✔				✔				✔
5	Ichrak Hayah		✔				✔				✔
6	Anne WT Muigai		✔				✔				✔

7	Denye Nathaniel Ogeh		✔				✔				✔
8	Abdoallah Sharaf		✔				✔				✔
9	Sadik Muzemil		✔				✔				✔
10	Kassahun Tesfaye		✔				✔				✔
11	Bouabid Badaoui						✔				✔
12	Marietjie Botes		✔								✔
13	Justin Eze Ideozu		✔				✔				✔
14	Sally Mueni Katee						✔				✔
15	Girish Beedessee				✔	✔					✔

References

Abarenkov, K. Anderson, A.F., Bissett, A., Fossøy, F., Grosjean, M., Hope, M., Kõljalg, U.,a Lundin, D., Nilsson, H., Prager, M., Jeppesen, T. S., Svenningsen, C., Schigel, D., Abarenkov, K., Provoost, P., Suominen, S., & Frøslev, T. G. (2023). *Publishing DNA-derived data through biodiversity data platforms*. https://doi.org/10.35035/DOC-VF1A-NR22

Acemoglu, D., & Akcigit, U. (2012). INTELLECTUAL PROPERTY RIGHTS POLICY, COMPETITION AND INNOVATION. *Journal of the European Economic Association*, *10*(1), 1–42. https://doi.org/10.1111/j.1542-4774.2011.01053.x

Adepoju, P. (2022). Africa's future depends on government-funded R&D. *Nature Africa*, d44148-022-00134–4. https://doi.org/10.1038/d44148-022-00134-4

Adler Miserendino, R. A., Meyer, R. S., Zimkus, B. M., Bates, J., Silvestri, L., Taylor, C., Blumenfield, T., Srigyan, M., & Pandey, J. L. (2022). The Case for Community Self-Governance on Access and Benefit Sharing of Digital Sequence Information. *BioScience*, *72*(5), 405–408. https://doi.org/10.1093/biosci/biac019

African Union (2024) Science Technology and Innovation Strategy for Africa 2024. Retrieved on October 1, 2024 from https://au.int/sites/default/files/newsevents/workingdocuments/33178-wd-stisa-english_-_final.pdf

Agrawal, R. C., Archak, S., & Tyagi, R. K. (2012). An overview of biodiversity informatics with special reference to plant genetic resources. *Computers and Electronics in Agriculture*, *84*, 92–99. https://doi.org/10.1016/j.compag.2011.12.003

Ambrosino, L., Colantuono, C., Monticolo, F., & Chiusano, M. L. (2018). Bioinformatics Resources for Plant Genomics: Opportunities and Bottlenecks in The -omics

Era. *Current Issues in Molecular Biology*, 71–88. https://doi.org/10.21775/cimb.027.071

Ambrosino, L., Tangherlini, M., Colantuono, C., Esposito, A., Sangiovanni, M., Miralto, M., Sansone, C., & Chiusano, M. L. (2019). Bioinformatics for Marine Products: An Overview of Resources, Bottlenecks, and Perspectives. *Marine Drugs*, *17*(10), 576. https://doi.org/10.3390/md17100576

Anand, A. V., Sreedevi, M. J., & Swapna, T. S. (2023). Plant Conservation Associated with Traditional Knowledge: Past and Future. In S. T. Sukumaran & K. T R (Eds.), *Conservation and Sustainable Utilization of Bioresources* (Vol. 30, pp. 261–290). Springer Nature Singapore. https://doi.org/10.1007/978-981-19-5841-0_11

Antognoli, E., Sears, J., & Parr, C. (2017). *Inventory of online public databases and repositories holding agricultural data in 2017* [Dataset]. Ag Data Commons. https://doi.org/10.15482/USDA.ADC/1389839

Arita, M., Karsch-Mizrachi, I., & Cochrane, G. (2021). The international nucleotide sequence database collaboration. *Nucleic Acids Research*, *49*(D1), D121–D124. https://doi.org/10.1093/nar/gkaa967

Balaji, D., & Terry, S. F. (2015). Benefits and Risks of Sharing Genomic Information. *Genetic Testing and Molecular Biomarkers*, *19*(12), 648–649. https://doi.org/10.1089/gtmb.2015.29008.sjt

BETin (2024). Genomics and Bioinformatics Research – Bio and Emerging Technology Institute- BETin. Reterieved from https://www.betin.gov.et/research-and-development/biotechnology-center/genomics-and-bioinformatics-research/ on October, 1, 2024

Beukman, M., Jumoorty, S., Faller, J. L., De Beste, E., & Johnston, B. (2024). Second Life: Enabling a Student Special Interest Group through Decommissioned

HPC Systems. *Practice and Experience in Advanced Research Computing 2024: Human Powered Computing*, 1–4. https://doi.org/10.1145/3626203.3670573

Bhat, J. A., & Taiwo, S. (2015). Plant Genetic Resources and Traditional Knowledge in Light of Recent Policy Developments. In R. K. Salgotra & B. B. Gupta (Eds.), *Plant Genetic Resources and Traditional Knowledge for Food Security* (pp. 183–198). Springer Singapore. https://doi.org/10.1007/978-981-10-0060-7_10

Kreiken (Bob), B. E., & (Bas) Arts, B. J. M. (2024). Disruptive data: How access and benefit-sharing discourses structured ideas and decisions during the Convention on Biological Diversity negotiations over digital sequence information from 2016 to 2022. *Global Environmental Change*, *87*, 102892. https://doi.org/10.1016/j.gloenvcha.2024.102892

Bonomi, L., Huang, Y., & Ohno-Machado, L. (2020). Privacy challenges and research opportunities for genomic data sharing. *Nature Genetics*, *52*(7), 646–654. https://doi.org/10.1038/s41588-020-0651-0

Browne, R. K., Luo, Q., Wang, P., Mansour, N., Kaurova, S. A., Gakhova, E. N., Shishova, N. V., Uteshev, V. K., Kramarova, L. I., Venu, G., Vaissi, S., Taheri-Khas, Z., Heshmatzad, P., Bagaturov, M. F., Janzen, P., Naranjo, R. E., Swegen, A., Strand, J., McGinnity, D., & Dunce, I. (2024). Ecological Civilisation and Amphibian Sustainability through Reproduction Biotechnologies, Biobanking, and Conservation Breeding Programs (RBCs). *Animals*, *14*(10), 1455. https://doi.org/10.3390/ani14101455

Bultum, L. E., Woyessa, A. M., & Lee, D. (2019). ETM-DB: Integrated Ethiopian traditional herbal medicine and phytochemicals database. *BMC Complementary and Alternative Medicine*, *19*(1), 212. https://doi.org/10.1186/s12906-019-2634-1

Byrd, J. B., Greene, A. C., Prasad, D. V., Jiang, X., & Greene, C. S. (2020). Responsible, practical genomic data sharing that accelerates research. *Nature Reviews Genetics*, *21*(10), 615–629. https://doi.org/10.1038/s41576-020-0257-5

Cengiz, N., Kabanda, S. M., Esterhuizen, T. M., & Moodley, K. (2023). Exploring perspectives of research ethics committee members on the governance of big data in sub-Saharan Africa. *South African Journal of Science*, *119*(5/6). https://doi.org/10.17159/sajs.2023/14905

Cengiz, N., Kabanda, S. M., & Moodley, K. (2024). Cross-border data sharing through the lens of research ethics committee members in sub-Saharan Africa. *PLOS ONE*, *19*(5), e0303828. https://doi.org/10.1371/journal.pone.0303828

Chapman, C. A., Abernathy, K., Chapman, L. J., Downs, C., Effiom, E. O., Gogarten, J. F., Golooba, M., Kalbitzer, U., Lawes, M. J., Mekonnen, A., Omeja, P., Razafindratsima, O., Sheil, D., Tabor, G. M., Tumwesigye, C., & Sarkar, D. (2022). The future of sub-Saharan Africa's biodiversity in the face of climate and societal change. *Frontiers in Ecology and Evolution*, *10*, 790552. https://doi.org/10.3389/fevo.2022.790552

Chelliah, R., Banan-MwineDaliri, E., Khan, I., Wei, S., Elahi, F., Yeon, S.-J., Selvakumar, V., Ofosu, F. K., Rubab, M., Ju, H. H., Rallabandi, H. R., Madar, I. H., Sultan, G., & Oh, D. H. (2022). A review on the application of bioinformatics tools in food microbiome studies. *Briefings in Bioinformatics*, *23*(2), bbac007. https://doi.org/10.1093/bib/bbac007

Chorev, N., & Ball, A. C. (2022). The Knowledge-Based Economy and the Global South. *Annual Review of Sociology*, *48*(1), 171–191. https://doi.org/10.1146/annurev-soc-080321-071214

Daly, B., & Ranwashe, F. (2023). South Africa's initiative toward an integrated biodiversity data portal. *Frontiers in Ecology and Evolution*, *11*, 1124928. https://doi.org/10.3389/fevo.2023.1124928

DELL Technologies (2024). Advanced technology that is pioneering African Research. Retrieved September 21, 2024 from https://www.delltechnologies.com/asset/ja-jp/solutions/infrastructure-solutions/customer-stories-case-studies/university-of-cambridge-all-access-africa-customer-story.pdf

Derese, Solomon, Ndakala, Albert, Rogo, Michael, Maynim, Scholastica, & Oyim, James. (2015, September 31). *Mitishamba database: A web based in silico database of natural products from Kenya plants*. 16th Symposium of the Natural Products Reseach Network for Eastern and Central Africa (NAPRECA), Arusha Tanzania.

Diallo, B. N., Glenister, M., Musyoka, T. M., Lobb, K., & Tastan Bishop, Ö. (2021). SANCDB: An update on South African natural compounds and their readily available analogs. *Journal of Cheminformatics*, *13*(1), 37. https://doi.org/10.1186/s13321-021-00514-2

Ebenezer, T. E., Muigai, A. W. T., Nouala, S., Badaoui, B., Blaxter, M., Buddie, A. G., Jarvis, E. D., Korlach, J., Kuja, J. O., Lewin, H. A., Majewska, R., Mapholi, N., Maslamoney, S., Mbo'o-Tchouawou, M., Osuji, J. O., Seehausen, O., Shorinola, O., Tiambo, C. K., Mulder, N., … Djikeng, A. (2022). Africa: Sequence 100,000 species to safeguard biodiversity. *Nature*, *603*(7901), 388–392. https://doi.org/10.1038/d41586-022-00712-4

Ebert, A. W., Engels, J. M. M., Schafleitner, R., Hintum, T. van, & Mwila, G. (2023). Critical Review of the Increasing Complexity of Access and Benefit-Sharing Policies of Genetic Resources for Genebank Curators and Plant

Breeders–A Public and Private Sector Perspective. *Plants 2023, Vol. 12, Page 2992*, *12*(16), 2992. https://doi.org/10.3390/PLANTS12162992

Federal Negarit Gazette (2016). Ethiopian Biotechnology Council and Biotechnology Institute Establishment Council of Ministers Regulation No., 388/2016 – Ethiopian Legal Brief. Retrieved October 1, 2014. From Federal Negarit Gazette of the Federal Democratic Republic of Ethiopia website https://chilot.wordpress.com/2021/03/08/ethiopian-biotechnology-council-and-biotechnology-institute-establishment-council-of-ministers-regulation-no-388-2016/ and https://www.metaappz.com/References/ethiopian_laws/federal/re_388_2016/en/pdf

Food and Agriculture Organisation of the United Nations (FAO). Information and Communication Technologies (ICT). Retrieved September 21, 2024, from https://aims.fao.org/information-and-communication-technologies-ict

Fernandez-Prada, C., Moretti, N. S., & Do Monte-Neto, R. L. (2024). Critical loss: The effects of VEuPathDB defunding on global health. *The Lancet Microbe*, 100980. https://doi.org/10.1016/j.lanmic.2024.100980

Fredriksson, M. (2021). Dilemmas of protection: Decolonising the regulation of genetic resources as cultural heritage. *International Journal of Heritage Studies*, *27*(7), 720–733. https://doi.org/10.1080/13527258.2020.1852295

Gadelha, L. M. R., De Siracusa, P. C., Dalcin, E. C., Da Silva, L. A. E., Augusto, D. A., Krempser, E., Affe, H. M., Costa, R. L., Mondelli, M. L., Meirelles, P. M., Thompson, F., Chame, M., Ziviani, A., & De Siqueira, M. F. (2021). A survey of biodiversity informatics: Concepts, practices, and challenges. *WIREs Data Mining and Knowledge Discovery*, *11*(1), e1394. https://doi.org/10.1002/widm.1394

Gbadegesin, Opeyemi. A., & Gbadamosi San, O. A. (2024). Indigenous Ecological Knowledge: A Transformative Approach to Biodiversity Legislation in Nigeria. *Environmental Management*, *74*(2), 317–331. https://doi.org/10.1007/s00267-024-01994-2

Gioti, A., Theodosopoulou, D., Bravakos, P., Magoulas, A., & Kotoulas, G. (2024). The bioinformatics landscape in environmental omics: Lessons from a national ELIXIR survey. *iScience*, *27*(6), 110062. https://doi.org/10.1016/j.isci.2024.110062

Gullotta, G., Engels, J. M. M., & Halewood, M. (2023). What Plant Genetic Resources for Food and Agriculture Are Available under the Plant Treaty and Where Is This Information? *Plants*, *12*(23), 3944. https://doi.org/10.3390/plants12233944

Haas, Fabian. (2015). *Biodiversity knowledge commons and sharing of research results with providers in East Africa* (1st ed.). Routledge. https://doi.org/10.4324/9781315717838

Hassan, A. H., Mokhtar, M. M., & El Allali, A. (2024). Transposable elements: Multifunctional players in the plant genome. *Frontiers in Plant Science*, *14*, 1330127. https://doi.org/10.3389/fpls.2023.1330127

Hudson, M., Garrison, N. A., Sterling, R., Caron, N. R., Fox, K., Yracheta, J., Anderson, J., Wilcox, P., Arbour, L., Brown, A., Taualii, M., Kukutai, T., Haring, R., Te Aika, B., Baynam, G. S., Dearden, P. K., Chagné, D., Malhi, R. S., Garba, I., … Carroll, S. R. (2020). Rights, interests and expectations: Indigenous perspectives on unrestricted access to genomic data. *Nature Reviews Genetics*, *21*(6), 377–384. https://doi.org/10.1038/s41576-020-0228-x

Humphries, F., Lawson, C., Benzie, J. A. H., & Morrison, C. (2023). African Aquaculture: Genetic Resource and Traditional Knowledge Access and Benefit Sharing

Measures. *Reviews in Fisheries Science & Aquaculture*, *31*(2), 180–196. https://doi.org/10.1080/23308249.2022.2027866

Ibezim, A., Debnath, B., Ntie-Kang, F., Mbah, C. J., & Nwodo, N. J. (2017). Binding of anti-Trypanosoma natural products from African flora against selected drug targets: A docking study. *Medicinal Chemistry Research*, *26*(3), 562–579. https://doi.org/10.1007/s00044-016-1764-y

INSDC. (2024). INSDC publishes guidance for prospective new members (20-05-2024). Retrieved September 25, 2024, from International Nucleotide Sequence Database Collaboration (INSDC) website: https://www.insdc.org/news/insdc-publishes-guidance-for-prospective-new-members/

Isard, M., Budiu, M., Yu, Y., Birrell, A., & Fetterly, D. (2007). Dryad: Distributed data-parallel programs from sequential building blocks. *Proceedings of the 2nd ACM SIGOPS/EuroSys European Conference on Computer Systems 2007*, 59–72. https://doi.org/10.1145/1272996.1273005

Ivey, P. J., Hill, M. P., Kenfack Voukeng, S. N., & Weaver, K. N. (2023). Nagoya Protocol and Africa's willingness to share biological control agents, are we deterred by barriers instead of using opportunities to work together? *BioControl*, *68*(3), 253–259. https://doi.org/10.1007/s10526-023-10184-8

Johnston, B., Timm, L., Macleod, D., & Poole, J. (2024). Ten Years of the HPC Ecosystems Project—Transforming HPC in Africa for the past decade. *Practice and Experience in Advanced Research Computing 2024: Human Powered Computing*, 1–8. https://doi.org/10.1145/3626203.3670537

Katee, S. M., Botes, M. W., Shetty, V., Kwon-Ndung, E. H., Beedessee, G., Paez, S., Ndlovu, S. I., Badaoui, B., Muigai, A. W., Ideozu, J. E., & Ebenezer, T. E. (2024). *African biodiversity genomics in the era of the Kunming-Montreal*

Global Biodiversity Framework: Ethical, Legal, and Social Perspectives. https://doi.org/10.31219/osf.io/2xq5e

Kennedy, G., Wang, Z., Maundu, P., & Hunter, D. (2022). The role of traditional knowledge and food biodiversity to transform modern food systems. *Trends in Food Science & Technology*, *130*, 32–41. https://doi.org/10.1016/j.tifs.2022.09.011

Knight, J., Flack-Davison, E., Engelbrecht, S., Visagie, R. G., Beukes, W., & Mwale, M. (2023). Compliance with the Nagoya protocol in the Southern African Development Community. *South African Journal of Botany*, *159*, 302–318. https://doi.org/10.1016/j.sajb.2023.06.015

Krassowski, M., Das, V., Sahu, S. K., & Misra, B. B. (2020). State of the Field in Multi-Omics Research: From Computational Needs to Data Mining and Sharing. *Frontiers in Genetics*, *11*, 610798. https://doi.org/10.3389/fgene.2020.610798

Krauss, K., & Loebbecke, C. (2021). Association for Information Systems Association for Information Systems AIS Electronic Library (AISeL) AIS Electronic Library (AISeL) Cloud Computing Diffusion in South Africa, Kenya, and Rwanda Cloud Computing Diffusion in South Africa, Kenya, and Rwanda. *AIS Electronic Library (AISeL)*. Retrieved from https://aisel.aisnet.org/menacis2021

Kumwenda, S., Niang, E. H. A., Orondo, P. W., William, P., Oyinlola, L., Bongo, G. N., & Chiwona, B. (2017). Challenges facing young African scientists in their research careers: A qualitative exploratory study. *Malawi Medical Journal*, *29*(1), 1. https://doi.org/10.4314/mmj.v29i1.1

Laltaika, E. (2022). Chapter 24: Western Intellectual Property Rights Regimes and Traditional Knowledge Protection Systems in Africa. In O. C. Ruppel & K. Ruppel-Schlichting (Eds.), *Environmental Law and Policy in Namibia* (pp.

643–660). Nomos Verlagsgesellschaft mbH & Co. KG. https://doi.org/10.5771/9783748933564-643

Lane, M. A. & Edwards, J. L. (2007). The global biodiversity information facility (GBIF). *Systematics Association Special Volume*, *73*(1).

Langmead, B., & Nellore, A. (2018). Cloud computing for genomic data analysis and collaboration. *Nature Reviews Genetics*, *19*(4), 208–219. https://doi.org/10.1038/nrg.2017.113

Lawson, C., Humphries, F., & Rourke, M. (2024). Challenging the existing order of knowledge sharing governance with digital sequence information on genetic resources. *Journal of Intellectual Property Law and Practice*, *19*(4), 337–357. https://doi.org/10.1093/jiplp/jpad129

Lewin, H. A., Richards, S., Lieberman Aiden, E., Allende, M. L., Archibald, J. M., Bálint, M., Barker, K. B., Baumgartner, B., Belov, K., Bertorelle, G., Blaxter, M. L., Cai, J., Caperello, N. D., Carlson, K., Castilla-Rubio, J. C., Chaw, S.-M., Chen, L., Childers, A. K., Coddington, J. A., … Zhang, G. (2022). The Earth BioGenome Project 2020: Starting the clock. *Proceedings of the National Academy of Sciences*, *119*(4), e2115635118. https://doi.org/10.1073/pnas.2115635118

Masehela, T. S., & Barros, E. (2023). The African continent should consider a harmonized consultative and collaborative effort towards coordinated policy and regulatory guidelines across the fields of biotechnology. *Frontiers in Bioengineering and Biotechnology*, *11*, 1211789. https://doi.org/10.3389/fbioe.2023.1211789

Mbelebele, Z., Mdoda, L., Ntlanga, S. S., Nontu, Y., & Gidi, L. S. (2024). Harmonizing Traditional Knowledge with Environmental Preservation: Sustainable Strategies for the Conservation of Indigenous Medicinal Plants (IMPs) and

Their Implications for Economic Well-Being. *Sustainability*, *16*(14), 5841. https://doi.org/10.3390/su16145841

Mc Cartney, A. M., Formenti, G., Mouton, A., De Panis, D., Marins, L. S., Leitão, H. G., Diedericks, G., Kirangwa, J., Morselli, M., Salces-Ortiz, J., Escudero, N., Iannucci, A., Natali, C., Svardal, H., Fernández, R., De Pooter, T., Joris, G., Strazisar, M., Wood, J. M. D., … Mazzoni, C. J. (2024). The European Reference Genome Atlas: Piloting a decentralised approach to equitable biodiversity genomics. *Npj Biodiversity*, *3*(1), 28. https://doi.org/10.1038/s44185-024-00054-6

Mokhtar, M. M., Abd-Elhalim, H. M., & El Allali, A. (2023). A large-scale assessment of the quality of plant genome assemblies using the LTR assembly index. *AoB PLANTS*, *15*(3), plad015. https://doi.org/10.1093/aobpla/plad015

Mokhtar, M. M., Alsamman, A. M., Abd-Elhalim, H. M., & El Allali, A. (2021). CicerSpTEdb: A web-based database for high-resolution genome-wide identification of transposable elements in Cicer species. *PLOS ONE*, *16*(11), e0259540. https://doi.org/10.1371/journal.pone.0259540

Mokhtar, M. M., Alsamman, A. M., & El Allali, A. (2023a). MegaSSR: A web server for large scale microsatellite identification, classification, and marker development. *Frontiers in Plant Science*, *14*, 1219055. https://doi.org/10.3389/fpls.2023.1219055

Mokhtar, M. M., Alsamman, A. M., & El Allali, A. (2023b). PlantLTRdb: An interactive database for 195 plant species LTR-retrotransposons. *Frontiers in Plant Science*, *14*, 1134627. https://doi.org/10.3389/fpls.2023.1134627

Mokhtar, M. M., & El Allali, A. (2022). PltRNAdb: Plant transfer RNA database. *PLOS ONE*, *17*(5), e0268904. https://doi.org/10.1371/journal.pone.0268904

Mokhtar, M. M., & El Allali, A. (2023). MegaLTR: A web server and standalone pipeline for detecting and annotating LTR-retrotransposons in plant genomes. *Frontiers in Plant Science*, *14*, 1237426. https://doi.org/10.3389/fpls.2023.1237426

Ntie-Kang, F., Onguéné, P. A., Scharfe, M., Owono Owono, L. C., Megnassan, E., Mbaze, L. M., Sippl, W., & Efange, S. M. N. (2014). ConMedNP: A natural product library from Central African medicinal plants for drug discovery. *RSC Adv.*, *4*(1), 409–419. https://doi.org/10.1039/C3RA43754J

Ntie-Kang, F., Telukunta, K. K., Döring, K., Simoben, C. V., A. Moumbock, A. F., Malange, Y. I., Njume, L. E., Yong, J. N., Sippl, W., & Günther, S. (2017). NANPDB: A Resource for Natural Products from Northern African Sources. *Journal of Natural Products*, *80*(7), 2067–2076. https://doi.org/10.1021/acs.jnatprod.7b00283

Oliva, A., Kaphle, A., Reguant, R., Sng, L. M. F., Twine, N. A., Malakar, Y., Wickramarachchi, A., Keller, M., Ranbaduge, T., Chan, E. K. F., Breen, J., Buckberry, S., Guennewig, B., Haas, M., Brown, A., Cowley, M. J., Thorne, N., Jain, Y., & Bauer, D. C. (2024). Future-proofing genomic data and consent management: A comprehensive review of technology innovations. *GigaScience*, *13*, giae021. https://doi.org/10.1093/gigascience/giae021

Parker-Allie, F., Gibbons, M. J., & Harebottle, D. M. (2023). A conceptual approach to developing biodiversity informatics as a field of science in South Africa. *Frontiers in Ecology and Evolution*, *11*, 1107212. https://doi.org/10.3389/fevo.2023.1107212

Parr, C. S., Wilson, N., Leary, P., Schulz, K., Lans, K., Walley, L., Hammock, J., Goddard, A., Rice, J., Studer, M., Holmes, J., & Corrigan, Jr., R. (2014). The Encyclopedia of Life v2: Providing Global Access to Knowledge About Life

on Earth. *Biodiversity Data Journal*, *2*, e1079. https://doi.org/10.3897/BDJ.2.e1079

Phale, K., Li, F., Adjei Mensah, I., Omari-Sasu, A. Y., & Musah, M. (2021). Knowledge-Based Economy Capacity Building for Developing Countries: A Panel Analysis in Southern African Development Community. *Sustainability*, *13*(5), 2890. https://doi.org/10.3390/su13052890

Pinthong, W., Muangruen, P., Suriyaphol, P., & Mairiang, D. (2016). A simple grid implementation with Berkeley Open Infrastructure for Network Computing using BLAST as a model. *PeerJ*, *4*, e2248. https://doi.org/10.7717/peerj.2248

Ratnasingham, S., & Hebert, P. D. N. (2007). BOLD: The Barcode of Life Data System (http://www.barcodinglife.org). *Molecular Ecology Notes*, *7*(3), 355–364. https://doi.org/10.1111/j.1471-8286.2007.01678.x

Rezny, L., White, J. B., & Maresova, P. (2019). The knowledge economy: Key to sustainable development? *Structural Change and Economic Dynamics*, *51*, 291–300. https://doi.org/10.1016/j.strueco.2019.02.003

Roumpeka, D. D., Wallace, R. J., Escalettes, F., Fotheringham, I., & Watson, M. (2017). A Review of Bioinformatics Tools for Bio-Prospecting from Metagenomic Sequence Data. *Frontiers in Genetics*, *8*. https://doi.org/10.3389/fgene.2017.00023

Rowe, J., Drou, N., Youssef, A., & Gunsalus, K. C. (2019). *BioSAILs: Versatile workflow management for high-throughput data analysis*. https://doi.org/10.1101/509455

SANBI (2023). Annual Report 2023. In South African National Bioinformatics Institute (SANBI). Retrieved on 21 September 2024 from SANBI website https://www.sanbi.org/about/annual-reports/sanbi-annual-reports-overview.

Sharaf, A., Ndiribe, C. C., Omotoriogun, T. C., Abueg, L., Badaoui, B., Badiane Markey, F. J., Beedessee, G., Diouf, D., Duru, V. C., Ebuzome, C., Eziuzor, S. C., Jaufeerally Fakim, Y., Formenti, G., Ghanmi, N., Guerfali, F. Z., Houaga, I., Ideozu, J. E., Katee, S. M., Khayi, S., … Ebenezer, T. E. (2023). Bridging the gap in African biodiversity genomics and bioinformatics. *Nature Biotechnology*, *41*(9), 1348–1354. https://doi.org/10.1038/s41587-023-01933-2

Sharaf, A., Nesengani, L. T., Hayah, I., Kuja, J. O., Mdyogolo, S., Omotoriogun, T. C., Odogwu, B. A., Beedessee, G., Smith, R. M., Barakat, A., Moila, A. M., El Hamouchi, A., Benkahla, A., Boukteb, A., Elmouhtadi, A., Mafwila, A. L., Abushady, A. M., Elsherif, A. K., Ahmed, B., … Ebenezer, T. E. (2024). Establishing African genomics and bioinformatics programs through annual regional workshops. *Nature Genetics*, *56*(8), 1556–1565. https://doi.org/10.1038/s41588-024-01807-6

Shih, C. C., Chen, J., Lee, A. S., Bertin, N., Hebrard, M., Khor, C. C., Li, Z., Juan Tan, J. H., Meah, W. Y., Peh, S. Q., Mok, S. Q., Sim, K. S., Liu, J., Wang, L., Wong, E., Li, J., Tin, A., Cheng, C.-Y., Heng, C.-K., … Dorajoo, R. (2023). A five-safes approach to a secure and scalable genomics data repository. *iScience*, *26*(4), 106546. https://doi.org/10.1016/j.isci.2023.106546

Sorokina, M., Merseburger, P., Rajan, K., Yirik, M. A., & Steinbeck, C. (2021). COCONUT online: Collection of Open Natural Products database. *Journal of Cheminformatics*, *13*(1), 2. https://doi.org/10.1186/s13321-020-00478-9

Sorokina, M., & Steinbeck, C. (2020). Review on natural products databases: Where to find data in 2020. *Journal of Cheminformatics*, *12*(1), 20. https://doi.org/10.1186/s13321-020-00424-9

UNEP Convention on Biological Diversity. (2016). *REPORT OF THE AD HOC TECHNICAL EXPERT GROUP ON DIGITAL SEQUENCE INFORMATION*

ON GENETIC RESOURCES. Retrieved from https://www.cbd.int/doc/c/4493/56b5/4ec917358b3cb37c7f7061f5/dsi-ahteg-2018-01-04-en.pdf

UNEP Convention on Biological Diversity. (2022). Conference of the Parties to the Adopted Decision 15/4, Kunming-Montreal Global Biodiversity Framework. Fifteenth meeting – Part II. Montreal, Canada, 7–19 (2022).

UM6P (2021). UM6P launches the Data Center and Africa's most powerful SuperCalculator for scientific research and innovation. Retrieved September 29, 2024 from https://www.um6p.ma/en/node/713

Ungogo, M. A., Ebiloma, G. U., Ichoron, N., Igoli, J. O., De Koning, H. P., & Balogun, E. O. (2020). A Review of the Antimalarial, Antitrypanosomal, and Antileishmanial Activities of Natural Compounds Isolated From Nigerian Flora. *Frontiers in Chemistry*, *8*, 617448. https://doi.org/10.3389/fchem.2020.617448

Vecchia, E. D. (2024). Pathoplexus: Towards fair and transparent sequence sharing. *The Lancet Microbe*, 100995. https://doi.org/10.1016/j.lanmic.2024.100995

Wan, Z., Hazel, J. W., Clayton, E. W., Vorobeychik, Y., Kantarcioglu, M., & Malin, B. A. (2022). Sociotechnical safeguards for genomic data privacy. *Nature Reviews Genetics*, *23*(7), 429–445. https://doi.org/10.1038/s41576-022-00455-y

Waterhouse, R. M., Adam-Blondon, A.-F., Agosti, D., Baldrian, P., Balech, B., Corre, E., Davey, R. P., Lantz, H., Pesole, G., Quast, C., Glöckner, F. O., Raes, N., Sandionigi, A., Santamaria, M., Addink, W., Vohradsky, J., Nunes-Jorge, A., Willassen, N. P., & Lanfear, J. (2021). Recommendations for connecting molecular sequence and biodiversity research infrastructures through ELIXIR. *F1000Research*, *10*, 1238. https://doi.org/10.12688/f1000research.73825.1

Waterhouse, R. M., Adam-Blondon, A.-F., Balech, B., Barta, E., Ying Shi Chua, P., Di Cola, V., Heil, K. F., Hughes, G. M., Jermiin, L. S., Kalaš, M., Lanfear, J., Pafilis, E., Palagi, P. M., Papageorgiou, A. C., Paupério, J., Psomopoulos, F., Raes, N., Burgin, J., & Gabaldón, T. (2024). The ELIXIR Biodiversity Community: Understanding short- and long-term changes in biodiversity. *F1000Research*, *12*, 499. https://doi.org/10.12688/f1000research.133724.2

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018. https://doi.org/10.1038/sdata.2016.18

Woldeamanuel, S. (2014). Mitishamba database of natural products from Kenyan plants for design of antimalarial drugs. *Grand Challenges Canada*. Retrieved from the Grand Challenges Canada Website September 21, 2024. https://www.grandchallenges.ca/grantee-stars/0260-01/

Wood-Charlson, E. M. (2023). The Importance of Sharing Data in Systems Biology. *Metabolites*, *13*(1), 99. https://doi.org/10.3390/metabo13010099

Wright, C. F., Hurles, M. E., & Firth, H. V. (2016). Principle of proportionality in genomic data sharing. *Nature Reviews Genetics*, *17*(1), 1–2. https://doi.org/10.1038/nrg.2015.5

Wynberg, R. (2023). Biopiracy: Crying wolf or a lever for equity and conservation? *Research Policy*, *52*(2), 104674. https://doi.org/10.1016/j.respol.2022.104674

Wysel, M., Baker, D., & Billingsley, W. (2021). Data sharing platforms: How value is created from agricultural data. *Agricultural Systems*, *193*, 103241. https://doi.org/10.1016/j.agsy.2021.103241

Yanou, M. P., Ros-Tonen, M. A. F., Reed, J., Moombe, K., & Sunderland, T. (2023). Integrating local and scientific knowledge: The need for decolonising knowledge for conservation and natural resource management. *Heliyon*, *9*(11), e21785. https://doi.org/10.1016/j.heliyon.2023.e21785

Zerbe, N. (2005). Biodiversity, ownership, and indigenous knowledge: Exploring legal frameworks for community, farmers, and intellectual property rights in Africa. *Ecological Economics*, *53*(4), 493–506. https://doi.org/10.1016/j.ecolecon.2004.10.015