Perspective

# Enriching African genome representation through the AGenDA project

https://doi.org/10.1038/s41586-025-09935-7

Received: 23 April 2025

Accepted: 19 November 2025

Published online: 14 January 2026

Check for updates

M. Ramsay[1,48]✉, H. Etheredge[2], F. Tluway[1], M. E. D'Amato[3], Z. Chikwambi[4], Y. Hamdi[5], I. Alhudiri[6], Y. Fakim[7], K. M. Ahmad[8], N. Belguith[9,10], D. Bentley[11], M. Boujemaa[5], N. Calumbuana[12,13], M. Chaouch[14], C. Charfeddine[5,15], G. Chinien[7], N. Dukuze[16,17], M. Eljilani[6], A. Elzagheid[6], N. Ferraz[12,18,19], A. Ghoorah[7], S. Goorah[7], M. Gribaa[20,21], S. Guidara[22,23], M. Guirat[24,25], S. Hazelhurst[1,26], M. Jallul[6], M. Kasu[3], N. Kharrat[27], U. Khumalo[4], Z. Kingsbury[11], I. Kisiangani[28], I. Lopes-Cendes[19,29], P. Lukusa[30,31], P. Makay[30,31], J. Makulo[32], G. Mubungu[30], C. Muhinda[17,33], D. M. Mukhongo[34,35], A. Murwira[36], A. Mustafa[6], J. Ndinkabandi[17], M. Ngole[30,37], Y. Nlandu[32], M. Nyathi[38], L. Pereira[13,39], I. Rejeb[40,41], L. L. Santos[42,43], D. Sengupta[1], A. Shebani[6], N. Smyth[1], A. Souissi[27], M. Trabelsi[44,45], A. Rebai[27], M. M. Chimpolo[12,18], A. Lumaka[30], C. Masimirembwa[4], S. F. Mohamed[28], N. Mulder[46], L. Mutesa[16,17], N. A. Hanchard[47,48] & A. Choudhury[1,48]

African populations remain substantially under-represented in research studies and global genomic databases. As the ancestral home of anatomically modern humans, Africa holds pride of place regarding human genetic diversity, with a deep and complex evolution over hundreds of thousands of years of human migration, admixture, and exposure to climate changes and infectious agents. Yet our present view of genomic diversity in Africa is sparse and poorly captures the rich variation across its more than 2,000 ethnolinguistic groups. To enhance representation, the Assessing Genetic Diversity in Africa (AGenDA) project, under the umbrella of the Human Heredity and Health in Africa (H3Africa) consortium, identified under-represented groups across nine different African countries for human whole-genome sequencing, with a view to enriching global datasets. Here we share our processes, including community engagement, obtaining ethics approvals, navigating legal compliance and developing a common governance framework. AGenDA is a testament to the determination of the scientific community to undertake research in challenging environments. It is led from Africa by African investigators who are the decision-makers in data-sharing processes. AGenDA is a step towards greater African representation in global genomic datasets to advance genomic research towards enabling precision medicine for Africa and the world.

The Human Heredity and Health in Africa (H3Africa) consortium was conceived to leverage the ancestral genomic diversity of Africa in order to provide insights into health and disease both on the continent and globally[1]. A cross-consortium genome-sequencing effort led to the landmark H3Africa population genetics paper, reporting on 432 high-depth whole-genome sequences of individuals from 50 ethnolinguistic groups in 13 countries[2]. It emphasized the extensive genetic variation across the continent by both geography and self-reported ethnicity, and despite the modest sample size (by present-day standards), it identified well over 3 million novel variants with no indication of nearing saturation, even among common variants. Individuals included in that study represented a convenience sampling of appropriately pre-consented individuals, predominantly from the Niger–Congo ethnolinguistic group, the most populous and most-sampled group in Africa. Thus, although the diversity of included genomes was highly illuminating, large geographic swathes of the continent and the majority of the more than 2,000 resident ethnolinguistic groups, especially those from other major ethnolinguistic divisions such as Afro-Asiatic speakers, Nilo-Saharan speakers, southern African hunter–gatherers, central African foragers and also North African populations, were left unexplored.

The Assessing Genetic Diversity in Africa (AGenDA) project is a pan-African genomics research collaboration originating from H3Africa to facilitate the inclusion of previously unstudied African groups into genomic datasets, addressing their historical underrepresentation in genomic research. A key component of current precision-medicine efforts is a comprehensive compilation of human genetic variation; this, for example, allows for more accurate imputation of variants in genome-wide association studies, facilitates better classification of missense variant pathogenicity and improves discrimination of complex alleles (for example, at human leukocyte antigen (HLA) locus) in all persons. The deep genetic tree found among Africans means that African genomes harbour at the human more genetic variation than other continental groups, making the inclusion of additional African

A list of affiliations appears at the end of the paper.

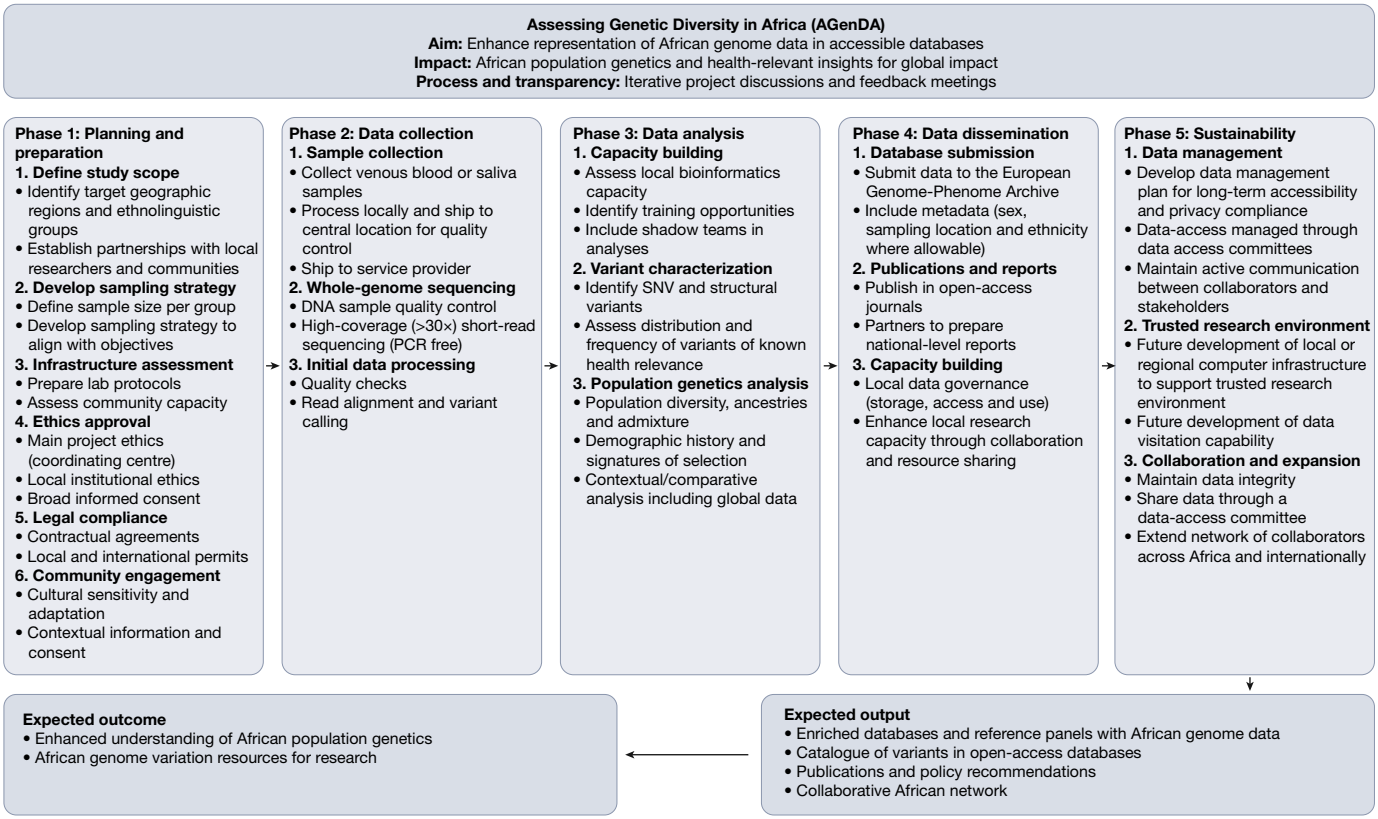



**Fig. 1 | AGenDA framework.** The five phases towards the implementation and execution of the AGenDA project, listing the expected broad outputs and anticipated outcomes.

genomes in variant databases an imperative to advancing human genetics and precision genomics.

AGenDA involves collaboration across nine African countries (samples from eight countries and coordination from South Africa) to generate whole-genome sequence data from at least 1,000 consenting participants. Recognizing that the comprehensive coverage of 2,000 ethnolinguistic groups would require hundreds of thousands of genomes, our strategy was to address the major gaps by enabling representation of each of the major ethnolinguistic divisions and geographic regions in Africa. This sampling strategy not only positions AGenDA to advance inclusion of individuals from Africa but also represents a major step towards promoting greater global representation, which promises to deepen our understanding of disease risk, improve precision health and reduce inequality for all people[3–6].

This Perspective reports the journey towards designing and implementing a multi-national study in Africa while adhering to ethical principles and presenting a regulatory framework for data sharing that is legally compliant and equitable across countries. The AGenDA project is unique in its scope, requiring the navigation of a panoply of laws and regulations in each collaborating country to facilitate the recruitment of participants and the transfer of biological samples and genetic data across borders, both within and outside Africa.

## Enriching African representation

Following an assessment of existing major gaps in genomic data across Africa, AGenDA focused on five broad population categories: (1) populations with substantial hunter–gatherer ancestry that span the contact zone for Khoe–San and rainforest forager groups that originate from the two oldest splits in the human lineage; (2) previously unstudied Niger–Congo-speaking populations (all belonging to the Bantu language family branch) from central and central-southern Africa[7]; (3) Nilo-Saharan-speaking populations; (4) North African- and Afro-Asiatic-speaking populations; and (5) African islands of the Southwest Indian Ocean whose unique demography reflects historical events.

The AGenDA project framework (Fig. 1) highlights the importance of establishing and adhering to guiding principles and the implementation of effective strategies when planning complex, multi-stakeholder initiatives. Doing so for AGenDA allowed the project to overcome challenges, adapt to changes and achieve its initial objectives. This article focuses on the processes and achievements during Phases 1 and 2 and we are currently entering Phase 3. This approach serves as a model for other large-scale genomic research projects on the continent.

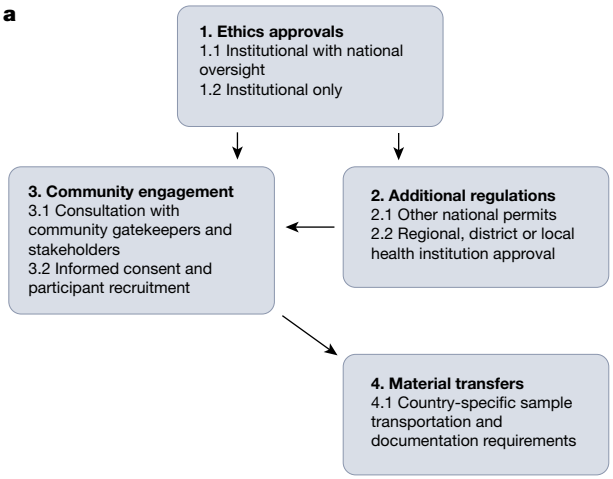
## Strategies for effective collaboration

Collaboration across AGenDA was enabled by five key strategies: cohesiveness, effective communication and trust-building, capacity strengthening, transparency and adaptability. These strategies provided a strong foundation to ensure smooth implementation throughout the initial development period.

### Cohesiveness

The success of AGenDA hinged on robust partnerships among key stakeholders, building an effective team. Bringing together principal investigators, researchers, institutional review boards, national and regional regulators, and local site leadership from the beginning ensured that all parties could contribute their expertise and felt invested in the project. This high-level and early process was pivotal to fostering effective collaboration, reducing overlaps and inefficiencies, and quickly finding solutions to real and potential challenges. Clearly defined project goals and methodologies encouraged openness and sharing.

**a**




**b**

| Fieldwork regulatory workflow | 1. Ethics approvals | | 2. Other regulations | | 3. Commmunity engagement | | 4. Material transfers |
|---|---|---|---|---|---|---|---|
| Country | 1.1 Institutional and national | 1.2 Institutional only | 2.1 Other national permits | 2.2 Regional, district or health institution approval | 3.1 Consultation community gatekeeper | 3.2 Participant recruitment | Country permits |
| Angola | ○ | ● | ○ | ○ | ●[e] | ●[e,f] | ● |
| DRC | ○ | ● | ○ | ○ | ○ | ●[e,f] | ● |
| Kenya | ● | ○ | ●[a] | ●[c] | ● | ●[e] | ● |
| Libya | ● | ○ | ○ | ●[d] | ○ | ●[e,f] | ● |
| Mauritius | ● | ○ | ○ | ○ | ○ | ●[e] | ● |
| Rwanda | ● | ○ | ○ | ○ | ○ | ●[e,f] | ● |
| South Africa | ○ | ● | ○ | ○ | ○ | ○ | ● |
| Tunisia | ○ | ● | ●[b] | ○ | ○ | ●[e,f] | ● |
| Zimbabwe | ○ | ● | ○ | ○ | ● | ● | ● |

**Fig. 2 | Ethical and legal framework processes followed in each country, including ethics approval, national regulations, community engagement and sample transport. a**, General workflow and steps followed by countries. (1) Ethics approval—institutions may fall under national regulation or be autonomous. (2) Additional regulations and procedural approvals. (3) Modes of engagement with participants—prior consent from community leaders (in some countries referred to as community gatekeepers who need to approve entry to the community); individual consent from all participants. (4) Transport of samples outside the country of collection. **b**, Country-specific processes. It is noted that samples were not collected in South Africa. Overall study coordination, sample centralization, and sample and data management were performed by the South African team. [a]National Research Permit. [b]National Authority for Personal Data protection. [c]County officers approvals. [d]Health institution approval. [e]Step collecting informed consent. [f]Step involving hospitals and/or health practicioners. Filled circles, necessary step; empty circles, not necessary or applicable.

As a concrete example, in the first stages of the project, joint monthly meetings together with one-on-one meeting between partners ensured a smooth workflow for sample collection, processing and shipping, as well as data analysis and feedback loops.

### Effective communication and trust-building

Effective communication was the cornerstone of the project's success. Structured communication protocols among researchers included regular meetings, progress updates and shared documentation, ensuring that project stakeholders remained informed of the project's progress. The clear, consistent communication fostered collaboration, improved problem-solving and maintained project momentum. It was important to allow flexibility for adaptations specific to on-site conditions. To engender trust, the processes were documented in meeting minutes, and a co-developed Research Collaboration Agreement (RCA) was signed by all partners. It provided clarity on how data integrity would be maintained, where data would be stored, what would be released into the public domain and how this would be done, how decisions on data use and sharing would be made, and how authorship would be managed. The aim was to provide fair and equitable acknowledgement of efforts while encouraging candidness and ensuring mutual respect.

### Capacity strengthening

The project plan divided the overarching objectives into smaller, clearly articulated and actionable tasks. This streamlining minimized unnecessary work and improved efficiency. To address skill gaps identified during the project, training workshops and reskilling activities were organized, allowing the project to maintain high standards of productivity and quality in its deliverables. The data analysis and publication strategies provide additional opportunities for skills development across the partnership.

### Transparency

AGenDA has a broad public benefit. This was explained as follows: knowledge of genetic variation in a country or community will inform variant interpretation and provide a genetic background for research studies, such that these communities would benefit from precision-medicine initiatives in the future. However, it would have little direct tangible benefit to individual participants. This was clearly articulated when describing to participants what the project entailed, its objectives and the governance framework for the resources that would be generated. Different approaches were used, as appropriate, in different countries to minimize misunderstandings, and by developing a well-structured project plan, we stayed focused on achieving the desired outcomes for AGenDA.

### Adaptability

The AGenDA project began with a small group and gradually expanded to include new partners. A key factor in successfully achieving this was the seamless integration of new members, while ensuring that founding members remained engaged. The project roadmap served as a central reference point, ensuring continuity despite changes in personnel or roles and maintaining momentum to achieve our objectives without substantial disruption.

## Ethical framework and legal compliance

The AGenDA collaboration is complex, and therefore the ethical, legal and regulatory frameworks underpinning a collaboration of this nature were necessarily extensive (Fig. 2 and Table 1). In developing a feasible and implementable project plan, we had to consider local country laws, regulations and standards. This meant meeting the requirements of all parties to satisfy the obligations pertaining to researchers, participants and the funder. It was important to ensure that all parties appreciated the compelling public value of having local population genome sequences to inform future research and possible clinical implementation in those populations. Reaching this common understanding required extensive negotiation, willingness to compromise, consensus building, transparency and patience.

The consent provided by all participants had to be sufficiently detailed to achieve the aims of AGenDA and adhere to ethical principles. We developed a template consent form that was approved by the Wits Human Research Ethics Committee (Medical) and could be adapted for prospective studies. For studies where participants had already been enroled, the template was compared against the consent form that had been used. This was done by a trained bioethicist (H.E.)

**Table 1 | Ethical and legal frameworks in different participating countries**

| Country | Ethics approval(s) | Permits for shipment | Legal framework for data sharing |
|---|---|---|---|
| Angola | Institutional review board | MTA, import/export permit | Yes. Lei Geral de Proteção de Dados Pessoais, 2011, https://apd.ao/fotos/frontend_1/editor2/110617_lei_22-11_de_17_junho-proteccao_dados_pessoais.pdf |
| DRC | Institutional ethics review board or national ethics committee | MTA, import/export permit | No national policy, institutional policies available |
| Kenya | Institutional ethics review board and national research permit (NACOSTI)<br>Accredited institutions: https://nsec.nacosti.go.ke/accredited-isercs/ | MTA, import/export permit | Yes. Data protection act, 2019, https://www.odpc.go.ke/wp-content/uploads/2024/02/TheDataProtectionAct__No24of2019.pdf |
| Libya | Institutional review board and approval from hospital where data was collected<br>These institutional committees all fall under the authority of the Libyan National Committee for Biosafety and Bioethics (LNCBB) | MTA, MoU, import/export permit | No national policy, institutional policies available |
| Mauritius | Institutional, hospital and/or national (Ministry of Health) | MTA, import/export permit | Yes. Data Protection Act, 2017, https://dataprotection.govmu.org/Pages/The%20Law/Data-Protection-Act-2017.aspx |
| Rwanda | Institutional, hospital and national | MTA, import/export permit and MoU | Yes. Data Protection and Privacy Laws, 2021, https://www.risa.gov.rw/data-protection-and-privacy-law |
| South Africa | Institutional Human Research Ethics Committee (HREC) | MTA, import/export permits | Yes. Protection of Personal Information Act (POPIA), 2020, https://popia.co.za/ |
| Tunisia | Institutional ethics committee and the National Authority for Personal Data protection | MTA, RCA, import/export permit | Yes. Recueil de la protection des données personnelles, 2021, https://inpdp.tn/Receuil_INPDP.pdf |
| Zimbabwe | National review board | MTA, import/export permit | No national policy, institutional policies available |

who assessed the documents for equivalence of key clauses necessary for AGenDA, including consent for genetic and genomic studies, and sharing of samples and data with international researchers and the research community. Participants were included in AGenDA only if they had consented to all these key activities.

Where required, consent forms were translated into local languages using contextualized language that would respect cultural norms. Consent included an understanding that individual-level data would be protected and de-identified by removing personally identifiable data (such as names and contact information), access would be restricted, and identities protected. Participation was voluntary, and participants had the right to withdraw from the study at any time without consequence, noting that if data had already been analysed, their data could not be withdrawn, but their samples and data would not be included in prospective studies.

Formal ethical and legal compliance needed approvals from nine different ethics committees and included a partnership agreement between the Wits Health Consortium (University of the Witwatersrand (Wits)) and Illumina Inc. (who partly subsidized and performed whole-genome sequencing) with approval from the funder (National Institutes of Health). At the time the project was initiated, there was no institution in Africa with the capacity to perform 1,000 whole-genome sequences. Fortunately, that is no longer the case. There were eight material transfer agreements (MTAs), export and import permits for shipping to the Sydney Brenner Institute for Molecular Bioscience (SBIMB) at Wits, and one MTA to send the DNA samples to the UK. A data management plan and RCA were signed by all partnering groups, outlining the governance of the data and including the process for data access and use. For the content of the RCA, please refer to Supplementary Note 1.

All partnering groups retain individual custodianship of their DNA samples and decide where to store them in the long term. DNA remaining after sequencing would either be destroyed or returned to the source country, unless otherwise agreed upon in writing.

## Communities and populations of AGenDA

AGenDA explored relationships built through H3Africa and sought partners who worked in regions that are under-represented in African genomic studies. All projects required ethics approval and appropriately consented participants who had agreed to the transfer and sharing of their samples and data internationally. Some samples were provided by biobanks in collaborating countries, others through research projects and some by prospective collection. All country-specific stipulations were followed, and DNA or blood samples were shipped to the SBIMB where DNA was extracted and/or quality assessment and normalization performed before shipping to the service provider and research partner, Illumina Inc. (Cambridge, UK), for whole-genome sequencing. DNA was extracted from venous blood, except for the Kenyan Nilo-Saharan study, where it was extracted from saliva samples. The countries of participant origin and represented groups are shown in Fig. 3 and briefly described below.

A central tenet of the H3Africa consortium was to empower Africa-based genomics researchers and enrich their exposure; this was also at the core of AGenDA's recruitment efforts. Having investigators from local communities leading in-country study activities greatly facilitated both individual and community trust, enabling efficient and robust participation. Community and stakeholder engagement processes across study sites were highly varied and extensive, requiring recruiting teams to consider appropriate local cultures and existing governance structures for genomic research. In general, the process required engaging local leaders and community gatekeepers, who then helped facilitate access to the community and ensured that the research had appropriate approvals at the local and community levels. Engagement with relevant stakeholders and study participants was designed to be both informative and transparent, ensuring that ethics committees and participants understood the study's objectives, potential risks and benefits.

Our inclusion strategy was informed by our aim to bridge some of the gaps in the representation of broad geographic regions and countries and language families in Africa. More specifically, there is a need to include non-Niger–Congo ancestries in public datasets. To address the geographic gaps, we strategically focused on regions of Central, South and North Africa, and the Indian Ocean. Within each region, we then prioritized countries that had no population-scale whole-genome data, which led to the inclusion of Libya, Tunisia, Rwanda, Zimbabwe

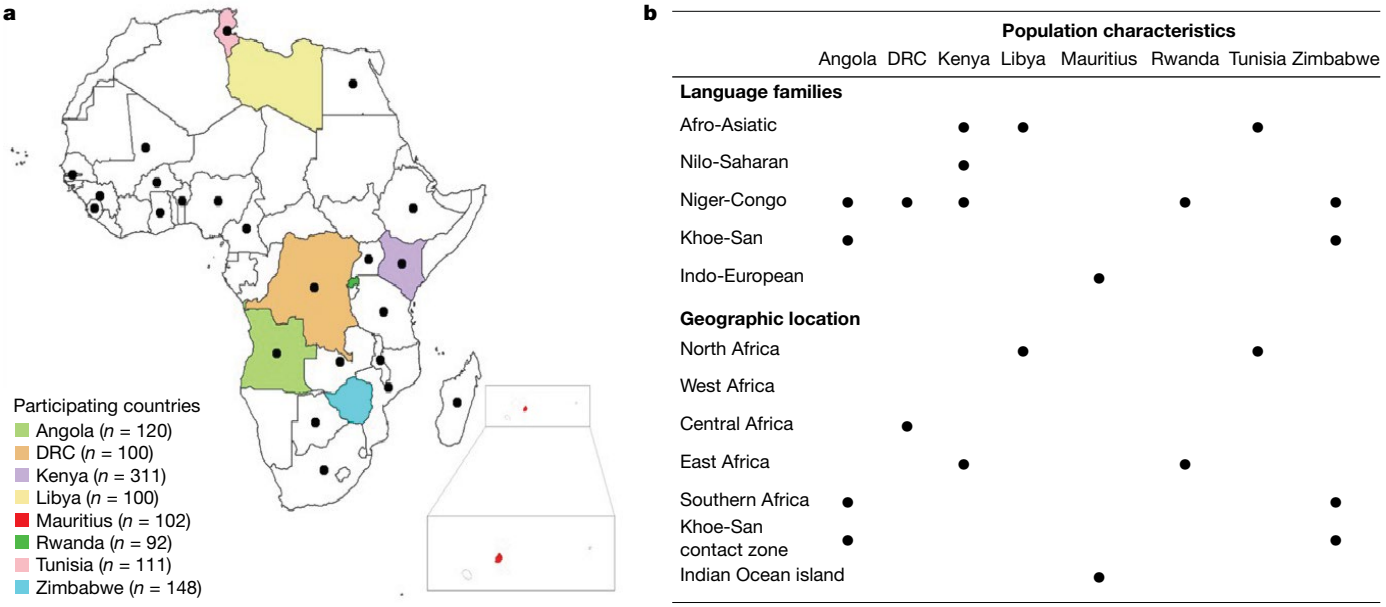



**Fig. 3 | Geographic origin and ethnolinguistic diversity of AGenDA participants. a**, Participating countries with participant representation in AGenDA are shown in different colours and countries with published whole-genome sequence data with dots. It is noted that for AGenDA, we sampled different ethnic groups in those countries that already had data. **b**, Broad linguistic affiliations and geographic regions of participants are shown ($n$ = 1,084 participants). It is noted that these are numbers of samples submitted for whole-genome sequencing and before obtaining the data and applying quality checks. The African and Mascarene island maps were plotted using the R packages tmap, dplyr, rnaturalearth, sf, osmdata, and ggplot2 in R 4.4.2 with RStudio 2024.9.1.

and Mauritius. In addition, Angola, Zimbabwe and the Democratic Republic of Congo (DRC) were selected as they are in the contact zone of hunter–gatherer and Bantu-speaking populations. In Angola and Zimbabwe, we further prioritized subregions inhabited by populations that are expected to have gene flow from Khoe–San populations. Similarly, to bolster the representation of Nilo-Saharan and Afro-Asiatic populations, we included multiple populations from Kenya. Instead of doing a systematic grid-based sampling or ethnolinguistic group-based sampling, we attempted to find a balance and prioritized previously unrepresented countries. Where previous data are available for a country (for example, Kenya), we chose to complement existing data and increase the representation of non-Niger–Congo ancestry. The following paragraphs summarize the sites and the samples included in AGenDA. As we used a mix of prospectively and retrospectively collected samples, and some of the sampling was geography-based whereas others were ethnolinguistic-group-based, the type and depth of information provided differed between sites.

## North Africa

Tunisia and Libya collaborated under a signed Memorandum of Understanding (MoU), whereby samples were shipped from Libya to Tunisia and then jointly sent to South Africa in accordance with national and international ethical and legal frameworks.

**Libya.** Libya is predominantly populated by Arabs, with substantial Amazigh (Berber), Tuareg and Toubou communities. The Amazigh are concentrated in northwestern Libya, whereas the Tuareg and Toubou inhabit the southern regions, where they preserve distinct cultural traditions[8]. Gene flow from the Middle East during the Islamic expansion has largely contributed to genetic diversity across North Africa, including Libya[9]. The Tuareg are Berber-speaking nomads, believed to descend from the Garamantes, who inhabited the Fezzan region during antiquity. The Toubou are also found in Chad and Niger, and they speak a Tebu language, which belongs to the Nilo-Saharan language group. Samples were selected from each major ethnic group according to their estimated population size from various public institutions across northern and southern Libya to reflect the genetic diversity within the country.

**Tunisia.** The Tunisian population has been shaped by its strategic location at the crossroads of Africa, Europe and the Middle East. Although predominantly an Arab population, it includes an indigenous Berber minority and smaller European and Mediterranean components. The ancestors of current Berbers have an ancient sub-Saharan African genomic component dating from the Neolithic era, and this was also observed in other Tunisian groups[10]. Turkish and Andalusian (Spanish) genomic components are represented in minority groups, primarily located in small cities in the northeastern and northwestern regions of the country. This ethnic admixture resulted in a largely homogeneous present-day population with specific genetic backgrounds and cultural diversity. This relative homogeneity has also been influenced by consanguineous marriages and endogamy, which remain common (ranging from 20% to about 40%)[11]. Recently, Tunisia launched a Human Genome Project, Genome Tunisia, to investigate its genetic heritage and to decipher its population structure by sequencing 5% of its adult population[12]. A subset of samples collected under Genome Tunisia was chosen for AGenDA to ensure biological sex balance and representation across regions and ethnic groups across three key regions of Tunisia: Tunis, Sousse and Sfax, representing the northern, central and southern parts of the country, respectively.

## East Africa

**Rwanda.** Study participants were prospectively recruited for AGenDA and included Rwandan participants aged 20–30 years who were students from the University of Rwanda-College of Medicine and Health Sciences in Kigali. Before recruitment started, an outreach programme was conducted to explain the study's scope and objectives, and volunteers provided informed consent before blood sample collection. Participants are all Rwandan and share a common first language (Kinyarwanda) while representing districts nationwide. Owing to historical events, no data on ethnicity were included.

**Kenya.** The AGenDA Kenyan participants are from two different studies and collectively represent three major ethnolinguistic divisions (Niger–Congo (Bantu-speaker branch), Nilo-Saharan and Afro-Asiatic) representing genomic diversity in East Africa.

**Nilo-Saharan.** The Endorois are a Kalenjin-dialect-speaking Indigenous group[13,14] in Kenya, with approximately 45,600 members (Kenyan Bureau of Statistics (KNBS), 2019). Historically, they inhabited the Rift Valley region in the semi-arid lands around Lake Bogoria, Baringo County, which they consider their ancestral lands. They practice pastoralism and wild-honey beekeeping in Lake Bogoria[15]. The Baringo area was historically occupied by Nilotic and Cushitic groups[16,17]. In addition, the presence of Bantu-speaking groups farther south and on the edges in the Rift Valley[18,19], combined with the migrations along the Rift Valley corridor[20], suggest a complex genomic make-up emerging from the interaction between these groups. Research teams in Kenya and South Africa (University of Kibabii and University of the Western Cape) signed an RCA for sample collection, sharing, transport, processing and analysis.

**Niger–Congo and Afro-Asiatic.** Participants were selected from three understudied Kenyan ethnolinguistic groups (Kamba, Kisii and Somali) recruited in the Africa Wits-INDEPTH partnership for Genomic studies (AWI-Gen) study[21]. The Kamba or Akamba are the fifth-largest ethnolinguistic group in Kenya. The Kamba speak Kikamba, a member of the Niger–Congo Bantu language family, and are distributed mostly in the lower-eastern part of the country (Machako, Kitui and Makueni provinces and countries). The Kisii or Gussii, also a Niger–Congo (Bantu-speaking branch) group, are from the western part of the country with a total population of about 2.7 million. The Somali speak a Cushitic branch of the Afro-Asiatic language and are historically from northeastern Kenya, having migrated from Ethiopia over many centuries. Using a combination of participants' self-identified ethnicity and language, parent and grandparent ethnicity and language, and preliminary genetic data from AWI-Gen, participants were selected to represent these groups.

### Central Africa

**DRC.** The DRC is the second largest country in Africa, with over 220 recognized groups, including multiple hunter–gatherer groups. The central location and the presence of nine neighbouring countries offer a unique ethnic meld, with the main groups being Niger–Congo Bantu speakers. The city-province of Kinshasa, with a population of almost 17 million, was the collection site for AGenDA and reflects the broad ethnolinguistic diversity of the country. Participants were recruited from a cohort of unrelated, ethnically diverse adult survivors of COVID-19. Prospective participants were invited to participate by telephone, and those who agreed had their samples taken at either their treating hospital or the Centre for Human Genetics of the University of Kinshasa.

### Southern Africa

**Angola.** Angola, situated on the southwestern African coast, has been highlighted as a crucial intermediate site along the route of Bantu migration, which commenced 4,000–5,000 years ago[2,22,23]. The country is also home to several Khoe–San communities. The participants were sampled from the municipalities of Curoca, Cuvelai, Cuanhama and Namacunde in the southern province of Cunene, within the Kalahari Basin. The research site was selected owing to its historical and cultural richness, underexplored populations and high ethnolinguistic diversity.

**Zimbabwe.** The Zimbabwean population has a rich mixture of Indigenous Khoe–San and migrant groups, reflecting millennia of human migration and cultural exchange in Southern Africa, making it a vital resource for understanding African genomics and personalized medicine[24–26]. The Khoe–San are primarily found in regions such as Tsholotsho and have unique genetic and cultural traits that trace back thousands of years[27]. The major Niger–Congo Bantu-speaking ethnic groups include the Shona, Ndebele and Kalanga, who settled in Zimbabwe during later migration waves[28]. Prospective sampling for AGenDA was done in the Tsholotsho district, which has the highest population of San in Zimbabwe; however, individual members tend to be scattered and reside in difficult to access areas. Recruitment thus necessitated close collaboration with the Chief of the San people as well as the Representative of the San Foundation to facilitate an information session at the chief's homestead, where the study and sample collection processes were performed.

### Southwest Indian Ocean islands

**Mauritius.** The island of Mauritius is part of the Mascarene region, east of Madagascar, which also comprises the La Reunion and Rodrigues islands. It has been inhabited by Arabs, Portuguese, French and British navigators from the seventeenth century onwards. Under French rule, labour was brought in from parts of Africa, whereas in the late-nineteenth century, there were arrivals that shaped the present-day demographics, making Mauritius a multiracial and multicultural society. For AGenDA, the focus was on Indo-Mauritians.

## Data governance and data sharing

The processes followed by the AGenDA project are aligned with the recently published World Health Organization Technical Advisory Group on Genomics' principles on data collection, access, use and sharing[29]. Whole-genome sequence data are high-depth-of-coverage (30×) short-read data from the NovaSeq 6000 platform (Illumina). Data will be submitted to the European Genome–Phenome Archive in line with the funding guidelines and consent provided by participants, and curated datasets (including single nucleotide variants (SNVs) and aggregate allele frequencies) will be openly accessible to the global research community.

Datasets corresponding to the samples from each partner site will be shared with the respective site principal investigator, as outlined in a Data Transfer Agreement. The use of site-specific data will be governed by each site and their institutional ethics review boards and local legislative compliance rules.

Two data-access committees will manage data-access and sharing requests. (1) An internal AGenDA Data Access Committee will manage data-access requests from AGenDA collaborating groups and members of the H3Africa consortium. (2) The H3Africa Data and Biospecimen Access Committee (H3A-DBAC)[30] will manage data requests for the data deposited in the European Genome–Phenome Archive and such requests would typically come from external researchers and may come from industry. Data will be shared in accordance with the agreements, ethics committee requirements and appropriate legal oversight.

The whole-genome-sequencing data will be part of the African H3Africa reference panel and imputation service based at The University of Cape Town and will contribute to genotype imputation for array-based genotyping datasets (usually from African populations). Sharing with industry will be permitted and decided by the H3A-DBAC or the AGenDA Data Access Committee, and will need to meet the same ethical, legal and data management criteria used by AGenDA groups, and follow the same procedures for requesting and obtaining data as any other researcher(s) accessing the data.

## Publication policy

Several core publications are anticipated from the AGenDA project and will be published as open access (when affordable) to ensure accessibility without a paywall. AGenDA publications are likely to fall under the

following categories: (1) AGenDA perspectives, opinions and cohort papers; (2) main population genetics AGenDA paper; (3) site-level publications; and (4) downstream topic-specific publications (Supplementary Fig. 1). Each partner organization may publish site-level papers without notification to AGenDA.

Data and publications are subject to the conditions of the National Human Genome Research Institute (NHGRI)[31] of the National Institutes of Health and the Data Sharing and Access and Release Policy of H3Africa[32] for the protection of data and participants.

## Lessons learnt

This continental African project was led by African scientists, with data sovereignty and decision-making residing in the hands of African investigators and data-access committees. Besides the usual enablers such as funding, infrastructure, skills development and institutional support, key ingredients for success were a team committed to shared goals, effective communication, the development of mutual trust and respect, and a willingness to navigate differences of view and compromise where necessary. We tapped into previous collaborations and trusted colleagues to start the process and then met regularly, with an in-person 2-day meeting in Johannesburg in April 2024 to consolidate the collaboration; this ensured that all parties understood expectations and facilitated planning for the analysis phases of the project. Members' meticulous attention to detail, participating and meaningfully contributing at meetings, and willingness to find feasible solutions were critically important.

An example of trust-building was the co-development of a process for decision-making around data sharing. It was important to get buy-in from each group to combine the data into a single database so that it could be analysed together. There would be centralized decision-making on sharing the full dataset. However, it was clarified that each group would receive a copy of the data from their set of participants and that they would have autonomy over how they used and shared that data.

Putting together this complex partnership was not without challenges. There was limited funding and a tight timeline for deliverables. This required painstaking navigation to manage the extensive ethical and legal documentation and the project activities. The project logistics presented several hurdles, including long delays with ethics approvals and finalizing MTAs and permits, as well as coordinating intercontinental and international shipping of samples. The latter included regulations for biological sample shipment across borders. In some countries, sourcing shipping materials was challenging, and courier company costs were massively inflated when compared with shipping in high-resource countries. DNA extractions were done in different locations and according to different protocols, which necessitated additional quality and processing checks before shipping for sequencing. Recruitment involved navigating the challenges posed by infectious disease outbreaks, including the COVID-19 pandemic and Ebola, as well as political instability, including looming civil unrest and destabilizing local elections. These factors tested the ingenuity of our teams (who have many interesting anecdotes to share), and the hurdles were overcome in the end, largely through sheer force of will and an unerring commitment to the science embodied in AGenDA.

We confirmed an established but under-recognized axiom that projects of this nature require people on the ground who know the local political, regulatory and research infrastructure landscape; are sensitive and aware of cultural nuances that may impact the research; and understand the science and its importance. In some of the study sites, this required engaging with and seeking approval from community gatekeepers (in some cases chiefs or community representatives) and travelling vast distances, sometimes on foot, to access remote and previously unstudied groups. Not all participating institutions had laboratory capacity for DNA extraction and storage, and in these cases, whole blood was shipped. The infrastructure to store genomic data is currently limited. We hope that through further engagement, compute infrastructure development and skills transfer, more groups will be able to use their own data at their own institutions.

The African genomics community is developing ecosystems around data sharing and more sustainable models for data accessibility and integration. These initiatives include the African Bioinformatics Institute (ABI)[33], the Genomic Centres of Excellence (GenCoE) and Africa Centres for Disease Control and Prevention (Africa CDC) initiative on developing a continental 'Roadmap on Human Genomics and Precision Public Health', and would also be supported by national governmental initiatives in countries where there are formalized national genome projects[12,34,35]. As a testament to the development of computational infrastructure, six of the AGenDA sites have the capacity for storing the raw sequence data from their samples. AGenDA is well positioned to collaborate with different population cohorts, including the ambitious plans for building a multi-omics project on a cohort of 250,000 individuals across 10 study sites, as articulated in the African Population Cohorts Consortium blueprint[36].

Careful planning and feasible plans taking into consideration the research culture, resources and skills are essential when developing multi-country collaborations across Africa. Joint planning and developing partnerships that leverage the strengths and experiences of individuals and individual groups are most likely to lead to successful collaborations. When embarking on collaborations, every partner must be committed to the long-term goals of the project and to exercising compassion while holding one another accountable for the success of the project. The African proverb applies: 'If you want to go fast, go alone. If you want to go far, go together'.

### Ethics approvals

**South Africa:** Wits University Health Research Ethics Committee (HREC-Medical), reference number 220414 (AGenDA) and reference number M2210108 (AWI-Gen).
**South Africa/Kenya:** University of Western Cape and Masinde Muliro University IERC, Kenya NACOSTI (Research Permit), reference numbers MMUST/ERC/033/2022, MMUST/MTA/001/2022, NACOSTI/P/22/17195, MMUST/ERC/161/2023 and BM16/3/18.
**Kenya:** AMREF Health Africa, Kenya, reference number AMREF-ESRC P1312/2022.
**Tunisia:** Institut Pasteur de Tunis, Tunisia, reference number 2021/31/I.
**Libya:** Libyan Biotechnology Research Center, reference number BEC-BTRC 12-2022.
**Angola:** Faculdade de Medicina da Universidade A. Neto, Angola, letter of approval.
**Rwanda:** University of Rwanda, Rwanda, reference number 212/CMHS IRB/2024.
**Mauritius:** University of Mauritius, Mauritius, reference number UoMREC/2021/P25.
**Zimbabwe:** Medical Research Council of Zimbabwe (MRCZ), reference number MRC2/A/2979.
**DRC:** Ecole de Santé Publique de Kinshasa, DRC ESP/CE/23B/2022.


1. H3Africa Consortium et al. Research capacity. Enabling the genomic revolution in Africa. *Science* **344**, 1346–1348 (2014).
2. Choudhury, A. High-depth African genomes inform human migration and health. *Nature* **586**, 741–748 (2020); erratum **592**, E26 (2021).
**This seminal publication from the H3Africa consortium identified millions of novel variants, highlighted the complexity of admixture and selection across African populations, and demonstrated the precision-medicine implications of including African genome data.**
3. Madden, E. B. et al. Advancing genomics to improve health equity. *Nat. Genet.* **56**, 752–757 (2024).
4. Corpas, M. et al. Bridging genomics' greatest challenge: the diversity gap. *Cell Genom.* **5**, 100724 (2025).
5. Fatumo, S. et al. A roadmap to increase diversity in genomic studies. *Nat. Med.* **28**, 243–250 (2022).
6. Wonkam, A. et al. Five priorities of African genomics research: the next frontier. *Annu. Rev. Genom. Hum. Genet.* **31**, 499–521 (2022).

7. Fortes-Lima, C. A. et al. The genetic legacy of the expansion of Bantu-speaking peoples in Africa. *Nature* **625**, 540–547 (2024).
8. Yakan, M. (ed.) *Almanac of African Peoples and Nations* (Routledge, 2017).
9. Fadhlaoui-Zid, K. et al. Genome-wide and paternal diversity reveal a recent origin of human populations in North Africa. *PLoS ONE* **8**, e80293 (2013).
10. El Moncer, W. et al. Mixed origin of the current Tunisian population from the analysis of Alu and Alu/STR compound systems. *J. Hum. Genet.* **55**, 827–833 (2010).
11. Elloumi-Zghal H. & Chaabouni Bouhamed H. Genetics and genomic medicine in Tunisia. *Mol. Genet. Genom. Med.* https://doi.org/10.1002/mgg3.392 (2018).
12. Hamdi, Y. et al. Genome Tunisia Project: paving the way for precision medicine in North Africa. *Genom. Med.* **16**, 104 (2024).
13. Ashamu, E. Centre for Minority Rights Development (Kenya) and Minority Rights Group International on behalf of Endorois Welfare Council v Kenya: a landmark decision from the African Commission. *J. Afr. Law* **55**, 300–313 (2011).
14. Kamga, S. A. D. The right to development in the African human rights system: the Endorois case. *De Jure Law J.* **44**, 381–391 (2011).
15. Endorois Peoples' Biocultural Protocol: Sustainable Biodiversity Resource Management for Access and Benefit Sharing and Protection from Threats to Culture http://www.absinitiative.info/fileadmin//media/Knowledge_Center/Pulications/BCPs/Endorois-Peoples-Biocultural-Protocol.pdf (Endorois Welfare Council, 2019).
16. Hobley, C. W. Notes on the geography and people of the Baringo District of the East Africa Protectorate. *Geogr. J.* **28**, 471–481 (1906).
17. Dundas, K. R. Notes on the tribes inhabiting the Baringo District, East Africa Protectorate. *J. R. Anthropol. Inst.* **40**, 49–72 (1910).
18. Munro, J. F. Migrations of Bantu-Speaking peoples of the eastern Kenya Highlands: a reappraisal. *J. Afr. Hist.* **8**, 25–28 (1967).
19. Wanyonyi, P. K. in *The Palgrave Handbook of Kenyan History* (eds Nasong'o, W. S. et al.) 13–21 (Palgrave Macmillan, 2023).
20. Grollemund, R., Schoenbrun, D. & Vansina, J. Moving histories: Bantu language expansions, eclectic economies, and mobilities. *J. Afr. Hist.* **64**, 13–37 (2023).
21. Ramsay, M. et al. H3Africa AWI-Gen Collaborative Centre: a resource to study the interplay between genomic and environmental risk factors for cardiometabolic diseases in four sub-Saharan African countries. *Glob. Health Epidemiol. Genom.* **22**, e20 (2016).
22. Patin, E. et al. Dispersals and genetic adaptation of Bantu-speaking populations in Africa and North America. *Science* **356**, 543–546 (2017).
**On the basis of genotype array data from 35 Bantu-speaking populations, the authors suggest that the Bantu-expansion spread eastwards and southwards after their southwards movement across the equatorial rainforest, and evaluate the ancestral contributions of Bantu-speaking populations to African American populations demonstrating a major ancestry portion originating from western Central Africa**.
23. Tallman, S., Sungo, M. D. D., Saranga, S. & Beleza, S. Whole genomes from Angola and Mozambique inform about the origins and dispersals of major African migrations. *Nat. Commun.* **14**, 7967 (2023).
**Genomes from Angola and Mozambique demonstrate the nuances of the Bantu expansion within southern Africa and were used to generate a reference panel to enhance the imputation accuracy for African-descended populations, especially African Americans and Brazilians, highlighting the importance of incorporating genomic diversity from regions associated with the Trans-Atlantic slave trade**.
24. Collins, F. S. & Varmus, H. A new initiative on precision medicine. *N. Engl. J. Med.* **372**, 793–795 (2019).
25. Montinaro, F. & Cristian Capelli, C. The evolutionary history of Southern Africa. *Curr. Opin. Genet. Dev.* **53**, 157–164 (2018).
26. Wang, K. et al. Ancient genomes reveal complex patterns of population movement, interaction, and replacement in sub-Saharan Africa. *Sci. Adv.* **6**, eaaz0183 (2020).
**These 20 sequenced ancient genomes from DRC, Kenya, Uganda and Botswana illustrate the complexities of interactions and admixture between early foragers and migrating pastoralist and farmer populations, and posit that pastoralism in eastern Africa may have been introduced by multiple, distinct migrating groups indicating a complex, multi-directional process**.
27. Gibbs, R. A. A genomic revolution: how the Human Genome Project changed medicine. *Genet. Med.* **22**, 945–952 (2020).
28. Msindo, E. *Ethnicity in Zimbabwe: Transformations in Kalanga and Ndebele Societies, 1860–1990* (Boydell and Brewer, 2012).
29. *Guidance for Human Genome Data Collection, Access, Use and Sharing* (WHO, 2024).
30. Rebai, A. et al. Responsible governance of genomics data and biospecimens in the context of broad consent: experiences of a pioneering access committee in Africa. *BMJ Glob. Health* **10**, e016026 (2025).
31. Third-party involvement in NHGRI-supported extramural projects. *NHGRI* https://www.genome.gov/about-nhgri/Policies-Guidance/Third-Party-Involvement-in-NHGRI-Supported-Extramural-Projects (2021).
32. *H3Africa Consortium Data Sharing, Access And Release Policy* https://h3africa.org/wp-content/uploads/2020/06/H3Africa-Consortium-Data-Access-Release-Policy-April-2020.pdf (H3Africa, 2020).
33. African Bioinformatics Institute. https://www.bioinformaticsinstitute.africa (accessed 9 December 2025)
34. *A Blueprint For An African Population Cohorts Consortium* https://ce-apcc.org/wp-content/uploads/2024/05/APCC-Blueprint-Document.pdf (Collaboration For the Establishment of the African Population Cohorts Consortium, Africa Population Health Research Centre, 2024).
35. Elmonem, M. A. et al. The Egypt Genome Project. *Nat. Genet.* **56**, 1035–1037 (2024).
36. Gaye, A. et al. Toward building a comprehensive human pan-genome: the SEN-GENOME project. *Am. J. Hum. Genet.* **111**, 2074–2078 (2024).



**Acknowledgements** The project originated from a discussion in the Genome Analysis Working Group of H3Africa and supported by the National Institutes of Health NHGRI and coordinator of the H3Africa consortium, who helped us successfully navigate the intricacies of the project. Each of the participating centres generously contributed time, expertise and samples collected under other projects, and persevered through delays and many difficulties. We are grateful for their perseverance. This study would not have been possible without the hard work of the fieldworkers, managers, scientists and students at each of our institutions, and we are grateful to the participants who agreed to be part of our study. The AWI-Gen Collaborative Center and its supplements were funded by the NHGRI, grant number U54HG006938 as part of the Human Heredity and Health in Africa Consortium. Each partner leveraged funding for data collection, and these will be acknowledged in individual publications. The funders were not involved in the study's design, collection, analysis, and interpretation of data, drafting the report, and did not restrict the publication of the research. A.C. and D.S. were also supported by National Institutes of Health grants (U01HG011717 and U01HL172182). The National Research Foundation of South Africa funded the collection, transport and DNA extractions with the African Origin Platforms grant 136511 to M.E.D. and N.F. Thuthuka grant TK1904084284442 to M.K. Gates Foundation, grant investment ID INV-036801, supported some of the AGenDA work in Zimbabwe.


**Author contributions** The study was conceived and planned by A.C. and N.A.H. as chairs of the H3Africa Data Analysis Working Group who made initial contacts to bring partners on board. M.R. contributed to refining the strategy, securing funding, negotiating agreements, consolidating partnerships and shepherding this publication. H.E. initiated the ethics and governance processes and worked with F.T. to engage with each participating group. F.T. provided overall oversight, arranged meetings and managed the legal processes. N.S. was responsible for sample quality assessments and shipping, including permits. The partner principal investigators are I.A. (Libya), Y.H. (Tunisia), L.M. (Rwanda), I.K. (Kenya), M.E.D. (Kenya), A.L. (DRC), M.C. (Angola), C. Masimirembwa (Zimbabwe) and Y.F. (Mauritius), who led the collection of samples and prioritization of existing samples for this study, community engagement, and obtaining ethics approvals at the site level. The following were writing group members for this paper: M.R., H.E., F.T., M.E.D., Z.C., Y.H., Y.F. and I.A.; and A.C. and N.A.H. reviewed several drafts. Figures were generated by M.E.D. and Z.C. The additional co-authors were involved from the individual study sites and had the opportunity to critically review the paper, and each accepts accountability for the accuracy and integrity of their involvement in the paper, these are: N.B., D.B., M.B., N.C., M.C., C.C., G.C., N.D., M.E., N.F., A.G., M. Gribaa, M. Guitarat, S. Goorah, S. Guidara, S.H., M.J., C. Muhinda, G.M., A. Murwira, A. Mustafa, J.N., M. Ngole, M. Nyathi, Y.N., S.F.M., L.P., I.R., L.L.S., D.S., A. Shebani, A. Souissi, M.T., A.R. and N.M. All co-authors read and approved the paper.

**Competing interests** The authors declare no competing interests.

**Additional information**
**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41586-025-09935-7.
**Correspondence and requests for materials** should be addressed to M. Ramsay.
**Peer review information** *Nature* thanks Jantina DeVries, Segun Fatumo and Daphne Martschenko for their contribution to the peer review of this work.
**Reprints and permissions information** is available at http://www.nature.com/reprints.
**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.




[1]Sydney Brenner Institute for Molecular Bioscience, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa. [2]Steve Biko Centre for Bioethics, Department of Internal Medicine, School of Clinical Medicine, University of the Witwatersrand, Johannesburg, South Africa. [3]Forensic DNA Lab, Biotechnology Department, Faculty of Natural Sciences, University of the Western Cape, Cape Town, South Africa. [4]African Institute of Biomedical Sciences and Technology, Harare, Zimbabwe. [5]Laboratory of Biomedical Genomics and Oncogenetics (LR16IPT05), Institut Pasteur de Tunis, Université Tunis El Manar, Tunis, Tunisia. [6]Genetic Engineering Department, Libyan Biotechnology Research Center, Tripoli, Libya. [7]University of Mauritius, Réduit, Mauritius. [8]Department of Microbiology, Faculty of Medicine, Sebha University, Sebha, Libya. [9]Department of Genetics, Hôpital Charles Nicolle de Tunis, Tunis, Tunisia. [10]Laboratory of Molecular Human Genetics (LR99ES33), Faculté de Médecine de Sfax, Université de Sfax, Sfax, Tunisia. [11]Illumina Centre, Cambridge, UK. [12]Department of Chemistry, Clinical Pathology and Genetics, Medical School, Agostinho Neto University, Luanda, Angola. [13]i3S Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal. [14]Laboratory of Bioinformatics, Biomathematics and Biostatistics (LR 16 IPT 09), Institut Pasteur de Tunis, Tunis, Tunisia. [15]Higher Institute of Biotechnology of Sidi Thabet, BiotechPole of Sidi Thabet, University of Manouba, Ariana, Tunisia. [16]Department of Pediatrics-Clinical Genetics, Rwanda Military Referral and Teaching Hospital, Kigali, Rwanda. [17]Center for Human Genetics, College of Medicine and Health Sciences, University of Rwanda, Kigali, Rwanda. [18]Medical School, Katyavala Bwila University, Benguela, Angola. [19]Department of Medical Genetics, School of Medical Sciences, University of Campinas (Unicamp), Campinas, Brazil. [20]Faculty of Medicine, University of Sousse, Sousse, Tunisia. [21]Department of Genetics, Farhat HACHED University Hospital, Sousse, Tunisia. [22]Department of Human Genetics, Hedi Chaker Hospital, Sfax, Tunisia. [23]Laboratory of Eukaryotes Molecular Biotechnology (LR15CBS02), Centre of Biotechnology of Sfax, University of Sfax, Sfax, Tunisia. [24]Department of Medical Genetics, Hedi Chaker University Hospital, University of Sfax, Sfax, Tunisia. [25]Laboratory of Screening Cellular and Molecular Processes (LR15CBS07), Centre of Biotechnology of Sfax, University of Sfax, Sfax, Tunisia. [26]School of Electrical and Information

Engineering, University of the Witwatersrand, Johannesburg, South Africa. [27]Laboratory of Molecular and Cellular Screening Process (LR15CBS07), Centre of Biotechnology of Sfax, University of Sfax, Sfax, Tunisia. [28]African Population and Health Research Center, Nairobi, Kenya. [29]The Brazilian Institute of Neuroscience and Neurotechnology (BRAINN), Campinas, Brazil. [30]Center for Human Genetics, University of Kinshasa, Kinshasa, Democratic Republic of the Congo. [31]Center for Human Genetics, University Hospitals of Leuven, University of Leuven, Leuven, Belgium. [32]Nephrology Unit, Department of Internal Medicine, University of Kinshasa, Kinshasa, Democratic Republic of the Congo. [33]Department of Clinical Laboratory, Rwanda Military Referral and Teaching Hospital, Kigali, Rwanda. [34]Department of Biological and Environmental Sciences Research, Kabii University, Bungoma, Kenya. [35]Biological and Agricultural Sciences Department, Alupe University, Busia, Kenya. [36]School of Business, Leadership and Governance, Africa University, Mutare, Zimbabwe. [37]Department of Medical Biology, Faculty of Medicine, University of Kinshasa, Kinshasa, Democratic Republic of the Congo. [38]Tsholotsho District Hospital, Tsholotsho, Zimbabwe. [39]Ipatimup Instituto de Patologia e Imunologia Molecular da Universidade do Porto, Porto, Portugal. [40]Department of Genetics, Mongi Slim Hospital, La Marsa, Tunisia. [41]Laboratory "Santé Mère Enfant" (LR22SP01), Tunis, Tunisia. [42]Experimental Pathology and Therapeutics Group, Research Center of IPO Porto (CI-IPOP)/CI-IPOP@RISE (Health Research Network), Porto, Portugal. [43]Surgical Oncology Department, Portuguese Oncology Institute of Porto (IPO Porto)/Porto Comprehensive Cancer Center Raquel Seruca (Porto.CCC Raquel Seruca), Porto, Portugal. [44]Department of Genetics, Charles Nicolle Hospital, Tunis, Tunisia. [45]Human Genetics Laboratory (LR99ES10), Faculty of Medicine, University of Tunis El Manar, Tunis, Tunisia. [46]Computational Biology Division, University of Cape Town, Cape Town, South Africa. [47]Childhood Complex Disease Genomics Section, Center for Precision Health Research, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA. [48]These authors jointly supervised this work: M. Ramsay, N. A. Hanchard, A. Choudhury. ✉e-mail: michele.ramsay@wits.ac.za