Microbiome Research Data Toolkit v1.0

Toolkit

Purpose

A microbiome metadata collection and phenotype harmonisation toolkit created by H3Africa Microbiome Working Group (https://h3africa.org/index.php/microbiome-working-group/) to promote standardised microbiome research data collection and reporting in Africa.

Toolkit Description

The Microbiome Research Data Toolkit is a multi-purpose toolkit which can be used to annotate microbiome research studies, for curation, proliferation and sharing, as well as prospectively standardise and retrospectively harmonise research participant-related information enrolled in microbiome studies and collaborations.

To achieve this goal, as illustrated in **Figure 1**, the toolkit, consists of two arms:

1. Arm 1: Once-off project (Per project) metadata reporting/annotation
2. Arm 2: Recurring (Per sample/participant) data collection

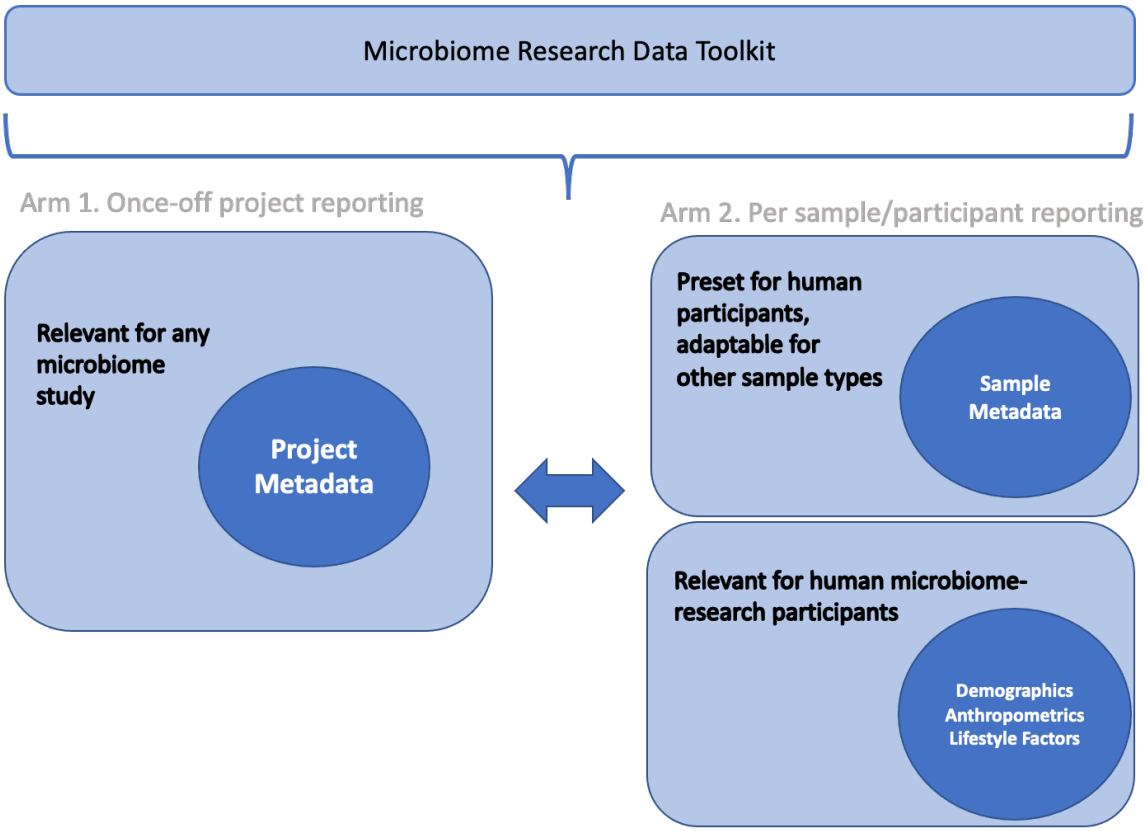

Figure 1. Structure of microbiome research data toolkit.

The purpose of the following guideline is to provide a detailed overview of the contents of toolkit, as well as considerations to be made in its implementation. It will be separated into 2 sections, reflective the separate arms previously highlighted.

Metadata Annotation Form

Guidelines

Project Metadata

The form enables the annotation of broad project-level metadata relevant for curation and sharing purposes. Metadata provided in this form are expected to be uniform across the set of samples (participants) within the project.

Fields	Project name: Study site(s): Project study design: **Response Options**: Case-Control; Clinical Trial; Cohort (Longitudinal); Cross-sectional; Other (If Other) Specify study design: Host of interest: **Response Options**: Homo sampien (Human) Mus musculus (Mouse) Rattus (Rat) Oryctolagus cuniculus (Rabbit) Other (If Other) Specify host of interest:
Notes	- **Projects** should be added on a sub-project basis, if samples within a larger project are subjected to varying experimental methodologies. - **Study site(s)** requires a description of where the study is being conducted e.g., city, country etc. - **Study design** refers to the framework, or the set of methods and procedures used to collect and analyze data on variables specified in a particular research problem. - **Host of interest** refers to the study subject of the project (and methodology) at hand.
Fields	Sample type: **Response Options**: Air; Agriculture; Host body product; Food; Water; Wastewater/Sludge; Skin; Soil; Other (If Other) Specify sample type: Host body product: **Response Options**: Blood; Milk; Saliva; Stool; Urine; Tissue; Other (If Other) Specify host body product: Host body site: **Response Options**: Gut; Oral; Nasopharyngeal; Vaginal; Skin; Urogenital; Other

	(If Other) Specify host body site: Sample collection method: **Response Options**: OMNIgene (DNA Genotek) Kit; Isohelix Kit; Norgen Kit; BioVision Kit; BIOME-preserve (Anaerobe Systems) Kit; Other (If Other) Specify sample collection method: Sample storage method: Sample storage buffer: **Response Options**: OMNIgene; Norgen; Canvas; Investigator-developed; Other (If Other) Specify sample storage buffer: Sample processing method:
Notes	- **Host body product** refers to the substance from which DNA/RNA is extracted, whereas **Host body site** refers to the human site from which the host body product was retrieved. - If 'Other' **Sample collection method** is reported, it is recommended that a link to the standard operating procedure followed for this process. - For **Sample storage method**, it is recommended that links to standard operating procedures are provided if possible. - **Sample processing method** requires brief description of any processing applied to the sample during or after retrieving the sample from environment, or a link to the relevant protocol(s) performed.
Fields	DNA extraction method: **Response Options**: QIAamp (QIAGEN) Kit; Macherel Nagel Kit; Omega Bio-tek Kit; Norgen Kit; PureLink (Thermo Fisher) Kit; BioVision Kit; Isohelix Kit; Zymo Kit; Other (If Other) Specify DNA extraction method: DNA amplification method: Sequencing platform: **Response Options**: Illumina (e.g. MiSeq, HiSeq, NovaSeq, etc.); PacBio (e.g. RS, RS II, Sequel, Sequel II); Oxford Nanopore Technologies (e.g. MinION); Roche 454 Pyrosequencing; Other (If Other) Specify sequencing platform: Assay type: **Response Options**: Amplicon; Shotgun; Other (If Other) Specify assay type: Amplicon region: **Response Options**: V1-9; ITS region; 16S Full-length; 18S Full-length; 18S Partial; Whole Genome; Other

	(If Other) Specify amplicon region: Outcome of interest: Sample size:
Notes	- If 'Other' **Nucleic acid extraction method** is reported, it is recommended that a link to the standard operating procedure followed for this process. - **Nucleic acid amplification method** requires a link to a literature reference, electronic resource or a standard operating procedure (SOP), that describes the enzymatic amplification (PCR, TMA, NASBA) of specific nucleic acids. - **Amplicon region** refers to the DNA or RNA fragments that are the source and/or product of amplification or replication events. - **Sample size** refers to the number of participants enrolled in the project. - **Outcome of interest** refers to the primary outcome investigated by the project.

Participant Data Collection Protocols

Overview

As listed below, Arm 2 consist of 9 Protocols, labelled Protocols A to I:

Protocol	**Phenotypes**	**Protocol**	**Phenotypes**
A	Sample Details	F	Medication Log
B	Demographics	G	Physical Activity
C	Anthropometrics	H	Diet
D	Smoking Status	I	Retrospective Lifestyle Factors
E	Alcohol Consumption		

Important Notes

1. The toolkit employs branching logic, therefore, we recommend that it is used in order, as some variables may or may not appear OR accept input based on the input of previously listed variables. Any addition of variables may also affect branching logic so new variables should be carefully positioned so as not to interrupt branching logic conditions with related variables.

2. Although not highlighted below, each instrument requires a collection date, which can be collected either manually or automatically.

3. Consistent codes are recommended for the identification of missing data; these are specified below:

Code	Value Label
-991	No information
-992	Asked but unknown
-993	Temporarily unavailable
-994	Not asked
-995	Refused
-998	Not applicable

Guidelines

Protocol A: Sample Details

The protocol enables the collection of essential sample information related to the participant.

Questions	Sample ID: Sample collection date: DNA extraction date: Sample quantity: Quantity unit: **Response Options** Volume (mL); Mass (g); Area (m2); Other (If Other) Specify quantity unit: Library layout: **Response Options**: Single; Paired; Other (If Other) Specify library layout: (If Stool) Bristol Stool Type: **Response Options**: Type 1 to 7
Notes	- **Library layout** refers to the configuration of reads in a sequencing library. - **Bristol Stool Chart** is widely used as a research tool to evaluate the effectiveness of treatments for various diseases of the bowel. The chart is used to describe the shapes and types of stools. - **Sample collection and DNA extraction dates** should be collected in following format - DD-MM-YYYY

Protocol B: Demographics

The protocol enables the collection of essential participant demographics such as age, gender, and language.

Questions	What is the participant's date of birth? Age: What was is the participant's biological sex at birth? **Response Options**: Male; Female; Other In which country was the participant born? In which country does the participant reside? In which city/town does the participant reside? What is the participant's residence's urbanisation level? **Response Options**: Rural; Suburban; Urban; Informal; Latitude: Longitude:

Notes	- **Date of birth** should be collected in following format - DD-MM-YYYY
	- Whenever possible participant date or birth should be captured and verified with official documentation.
	- **Age** can be automatically calculated based on the date of birth.
	- If date of birth is unavailable, estimated age may be provided.
	- The **sex** field collects the biological sex of a participant and should not be confused with gender identification. The option "Other" is used to identify Intersex participants who may be sensitive due to stigma about their biological sexual status.
	- **Urbanisation level descriptions:**
	  - A **Rural** region is a geographic area that is located outside towns and cities. Typical rural areas have a low population density and small settlements.
	  - **Suburban** areas are lower density areas that separate residential and commercial areas from one another. They are either part of a city or urban area, or exist as a separate residential community within commuting distance of a city.
	  - **Urban** areas are locations with high population density. Urban areas are in cities and towns. An urban area is often the main area of employment.
	  - **Informal settlements** are residential areas that do not comply with local authority requirements for conventional (formal) townships. They are, typically, unauthorised and are invariably located upon land that has not been proclaimed for residential use.
	- **Longitude** and **Latitude** is a range of African-specific language selections are provided.

Questions	What is the participant's native language? (If Other) Specify other native language: What is the participant's ethnic or tribal affiliation? (If Other) Specify other ethnic or tribal affiliation:
Notes	- The participant's **home/native language** that they were raised with should be completed in the language field. If the participant has multiple native languages, complete the field with the most commonly used native tongue – the language they consider their home language. - A range of African-specific language selections are provided. - The participant's original **ethnic or tribal affiliation** should be collected in the ethnic affiliation field. If the participant identifies with multiple ethnic tribes, document the primary one or collect the one the participant first identified with growing up. - A range of African-specific ethnic affiliations are provided.

Protocol C: Anthropometrics

The protocol enables the collection of anthropometric data, including height and weight, as well as waist and head circumference.

Questions	Height: Weight:

Notes	- **Height** is the distance from the top of the participant's head to the heels of his or her feet (i.e., the vertical length).
	- Height is measured in cm using a stadiometer, if possible.
	- Detailed protocols for measuring height can be found in PhenX:
	  - **Standing Height Protocol:** https://www.phenxtoolkit.org/index.php?pageLink=browse.protocoldetails&id=20703
	  - **Measuring height in seated position** for participants unable to stand: https://www.phenxtoolkit.org/index.php?pageLink=browse.protocoldetails&id=20701
	- **Weight** is measured in kg using a using a floor scale. The instrument should be calibrated daily using standardized weights, and a log of calibration results should be maintained.

Questions	Body-Mass Index (BMI): Middle Upper Arm Circumference (MUAC): Waist Circumference: Hip Circumference: Waist-Hip Ratio:
Notes	- **BMI** is defined as the body mass divided by the square of the body height. - **MUAC,** measured in cm using a flexible, non-stretch measuring tape, is the circumference of the right upper arm measured at the midpoint between the tip of the shoulder and the tip of the elbow. - MUAC is recommended for participants ranging from Infancy to Childhood Life Stages (see **Life Stage**) - **Waist circumference**, measured in cm using a flexible, non-stretch measuring tape, is a measurement taken around the abdomen at the level of the belly button. - Waist circumference should only be collected for participants ranging from the Childhood to Adult Life Stages (see **Life Stage**) - **Hip Circumference,** measured in cm using a flexible, non-stretch measuring tape, is a measurement taken around the widest part of the hips. - **Waist-Hip Ratio** looks at the proportion of fat stored on your body around your waist and hip, calculated by dividing your waist measurement by your hip measurement**.**
Questions	**Dual Energy X-ray Absorptiometry (DEXA)** Total Body Fat Percentage: Total Body Fat Mass: Total Fat-Free Mass:
Notes	- A **DEXA scan** is an imaging test that measures bone density (strength). This test can also measure your body composition, such as body fat and muscle mass

Protocol D: Smoking Status

The protocol enables the self-reported collection of smoking use in participants in the Adolescence and Adult life stages.

Questions	Has the participant smoked at least 1/100 cigarette(s) in their entire life?* **Response Options**: Yes; No (If YES) How old was the participant when they first started smoking cigarettes? How old was the participant when he/she first started smoking tobacco cigarettes? What type of smoker is the participant? **Response Options**: An EVERY day smoker; A FAIRLY REGULAR (some days) smoker; A FORMER smoker; Don't Know; Refused Has the participant EVER smoked tobacco cigarettes EVERY DAY for at least 6 months? **Response Options**: Yes; No
Notes	- This protocol is strictly applicable for the Adolescence and Adult Life Stages. - Interviewers need to be sensitive to the participant's culture and religion and be aware that some participants may be reluctant to answer these questions truthfully. - Participants should be reassured that their answers will be kept confidential. - For adolescents, the first question should be specified as 1 cigarette. - For adults, the first question should be specified as 100 cigarettes. - For adults, it is assumed that people who have smoked less than 100 cigarettes in their lifetime do not have a significant smoking status to investigate.
Questions	On the days that they smoke, on average, how many tobacco cigarettes does the participant smoke? **OR** If a former smoker, on the days that they smoked, on average, how many tobacco cigarettes did the participant smoke? Over the past 30 days, on how many <u>days</u> did the participant smoke? **OR** If former smoker, on average, on how many <u>days</u> did the participant smoke in a month?
Questions	FOR Former Smokers): About how long has it been since the participant COMPLETELY quit smoking cigarettes? Length of time since quitting measured in: **Response Options**: Years; Months; Weeks; Days

Protocol E: Alcohol Use

The protocol enables the self-reported collection of alcohol use in participants in the Adolescence and Adult life stages.

Questions	In their entire life, has the participant had at least 1 drink of any kind of alcohol? **Response Options**: Yes; No

	(if YES) How old was the participant when they first started drinking alcohol?
Notes	- This protocol is strictly applicable for the Adolescence and Adult Life Stages. - Interviewers need to be sensitive to the participant's culture and religion and be aware that some participants may be reluctant to answer these questions truthfully. - Participants should be reassured that their answers will be kept confidential. - Studies should have a defined quantity or way to measure drinks as examples for participants. One "standard" drink (or one alcoholic drink equivalent) contains roughly 14 grams of pure alcohol. In practice, the total volume differs between types of alcohol. For more information, see: https://www.phenxtoolkit.org/protocols/view/30301
Questions	During the past 30 days, on how many days did the participant drink one or more drinks of an alcoholic beverage? On the days that the participant drank during the past 30 days, how many drinks did they usually have each day? What was the LARGEST number of drinks that they ever drank in a single day? What is the participant's preferred alcoholic beverage? **Response Options**: Beer; Brandy; Cider; Gin; Rum; Tequila; Vodka; Whiskey; Wine; Other (If Other) Specify preferred alcoholic beverage:

Protocol F: Medication Log

The protocol enables the collection of information related to the participant's ingestion of concomitant medication as well as medication received during hospitalisation.

Questions	Medication name: Medication Coded Name: Reason for medication: Start date: Ongoing? **Response Options**: Yes; No Stop date: Dose amount: Dose unit: **Response Options**: mg; ml; spray or puff; tablet; pill; softgel; capsule; application Dose frequency: **Response Options**: once per day (QD); twice a day (BID); three times a day (TID); four times a day (QID); nightly (NOCT); as needed (PRN) Route of administration:

	Response Options: Orally; Per rectum; Intravenous; Per vaginal; Inhaled; Intramuscularly; Nasogastric; Subcutaneously; Sublingually; Topical
Notes	- **Start and stop dates** should be collected in following format - DD-MM-YYYY - Participants should be asked to bring all their current medications with them at the time of their interview/appointment. - For over-the-counter or self-prescribed medications (including vitamins and supplements), the details for medications consumed in the previous 2 weeks.

Protocol G: Physical Activity

The protocol enables the self-report collection of information related to a research participant's physical activity at home and at work.

Questions	**Physical Activity at Work** Does the participant's work involve vigorous-intensity activity that causes large increases in breathing or heart rate like for at least 10 minutes continuously? **Response Options**: Yes; No (If YES) In a typical week, on how many days does the participant do vigorous-intensity activities as part of the work? How much time does the participant spend doing vigorous-intensity activities at work on a typical day? Does the participant's work involve moderate-intensity activity that causes small increases in breathing or heart rate for at least 10 minutes continuously? **Response Options**: Yes; No (IF YES) In a typical week, on how many days does the participant do moderate-intensity activities as part of their work? How much time does the participant spend doing moderate-intensity activities at work on a typical day?
Notes	- **Examples** of vigorous-intensity activity: carrying or lifting heavy loads, digging or construction work - **Examples** of moderate-intensity activity: brisk walking, carrying light loads
Questions	**Physical Activity During Travel** Does the participant walk or use a bicycle (pedal cycle) for at least 10 minutes continuously to get to and from places? **Response Options**: Yes; No (If YES) In a typical week, on how many days does the participant walk or bicycle for at least 10 minutes continuously to get to and from places? How much time does the participant spend walking or bicycling for travel on a typical day?
Questions	**Recreational Physical Activity** Does the participant do any vigorous-intensity sports, fitness or recreational

	(leisure) activities that cause large increases in breathing or heart rate for at least 10 minutes continuously? **Response Options**: Yes; No (If YES) In a typical week, on how many days does the participant do do vigorous-intensity sports, fitness or recreational (leisure) activities? How much time does the participant spend doing vigorous-intensity sports, fitness or recreational activities on a typical day? Does the participant do any moderate-intensity sports, fitness or recreational (leisure) activities that cause a small increase in breathing or heart rate for at least 10 minutes continuously? **Response Options**: Yes; No (If YES) In a typical week, on how many days does the participant do moderate-intensity sports, fitness or recreational (leisure) activities? How much time does the participant spend doing moderate-intensity sports, fitness or recreational (leisure) activities on a typical day?
Notes	- **Examples** of vigorous-intensity sports: running or football - **Examples** of moderate-intensity activity: brisk walking, cycling, swimming, volleyball
Questions	**Sedentary Lifestyle** How much time does the participant usually spend sitting or reclining on a typical day?

Protocol H: Diet

The protocol enables the self-report collection of information related to a research participant's dietary intake.

Questions	During the past month, how often did the participant... - eat hot or cold cereals? - have dairy milk (to drink or in cereal)? - drink 100% pure fruit juice? - drink sweetened fruit drinks, sports or energy drinks? - eat fruit? Include fresh, frozen, dried or canned fruit. - eat a green leafy or lettuce salad, with or without other vegetables? - eat any kind of fried potatoes? - eat any other kind of potatoes? - eat refried beans, baked beans, beans in soup, pork and beans or any other type of cooked dried beans? - eat other vegetables? - eat any kind of cheese? - eat whole grain brain? - eat white bread?

	- eat brown rice or other cooked whole grains, such as bulgur, cracked wheat, or millet? - eat white rice? - eat pasta dishes, such as spaghetti, macaroni or noodles? - eat red meat, such as beef, pork, ham, or sausage? - eat processed meat, such as bacon, and deli meat, or hot dogs? - eat chicken or other poultry? - eat seafood? - eat spicy food, containing peppers or spicy seasoning? - eat food high in sodium? **Response Options**: Regularly; Usually, Sometimes, Rarely; Never How often does the participant buy from a *vendor or take-away or restaurant? **Response Options**: Regularly; Usually, Sometimes, Rarely; Never
Notes	- **Frequency Descriptions**: - Usually (>4 days a week); - Sometimes (1-4 days a week); - Rarely (<3 days a month)
Questions	What type of oil or fat is most often used for food preparation in the participant's household? **Response Options**: Vegetable oil; Animal fat; Butter or ghee; Margarine; Any of the Above; None used Has a doctor, nurse, or other healthcare worker ever advised the participant to change their diet? **Response Options**: Yes; No Has a doctor, nurse, or other healthcare worker ever advised the participant to lose weight? **Response Options**: Yes; No

Protocol I: Retrospective Lifestyle Factors

The protocol enables the self-report collection of information related to a research participant's dietary intake.

Questions	**CHRONIC MEDICATION** Does the participant use chronic medication? **Response Options**: Yes; No (If Yes) Which classes of chronic medication does the participant use? **Response Options**: Anti-convulsants Anti-depressants Anti-anxiety Antibiotics

<table>
<tr><td></td><td>Anti-hypertensives
Anti-microbials
Anti-psychotics
Anti-retrovirals
Anti-thyroid drugs
Anti-inflammatory
Statins
Other
(If Other) Specify class of chronic medication:
Has the participant used any antibiotics in the past 3 months?
Response Options: Yes; No</td></tr>
<tr><td>Notes</td><td>- Chronic medication
- Medication classes:
<table>
<tr><td>Anti-convulsants</td><td>A class of medications that act through multiple different mechanisms to control seizures, also known as antiepileptics.</td></tr>
<tr><td>Anti-depressants</td><td>A class of medications used to treat major depressive disorder, anxiety disorders, chronic pain, and addiction.</td></tr>
<tr><td>Anti-anxiety</td><td>Also called anxiolytics, a class of medications used to prevent or treat anxiety symptoms or disorders.</td></tr>
<tr><td>Antibiotics</td><td>Medications that destroy or slow down the growth of bacteria. Doctors prescribe them to treat bacterial infections.</td></tr>
<tr><td>Anti-hypertensives</td><td>A class of drugs that are used to treat hypertension.</td></tr>
<tr><td>Anti-fungals</td><td>Medicines that kill or stop the growth of fungi (the plural of fungus) that cause infections.</td></tr>
<tr><td>Anti-psychotics</td><td>Also called neuroleptics, a class of psychotropic medication primarily used to manage psychosis, principally in schizophrenia but also in a range of other psychotic disorders.</td></tr>
<tr><td>Anti-retrovirals</td><td>A drug used to prevent a retrovirus, such as HIV, from replicating.</td></tr>
<tr><td>Anti-thyroid drugs</td><td>Antithyroid drugs (also called thionamides) are most often used to treat an overactive thyroid (hyperthyroidism) caused by Graves' disease. These drugs block the formation of thyroid hormone by the thyroid gland.</td></tr>
<tr><td>Anti-inflammatory</td><td>A class of drugs which reduces pain, decreases inflammation, decreases fever, and prevents blood clots.</td></tr>
<tr><td>Statins</td><td>Statins are drugs that can lower your cholesterol. They work by blocking a substance your body needs to make cholesterol.</td></tr>
</table></td></tr>
<tr><td>Questions</td><td>DIET
Which diet type best describes the participant's diet?
Response Options:
Low-Calorie Diet
Low Carb Diet</td></tr>
</table>

Low Fat Diet
Blood Type Diet
Mediterranean Diet
Vegetarianism
Veganism
High Protein Diet
High Fibre Diet
Not Specified
Other (If Other) Specify diet type:

- Diet descriptions:

Low-Calorie Diet	An eating plan used to help people lose weight. It involves limiting the overall number of calories you eat or drink in a day.
Low Carb Diet	A diet that limits carbohydrates, primarily found in sugary foods, pasta, grains, legumes, fruits, breads, sweets and starchy vegetables.
Low Fat Diet	A diet which involves eating mostly plant foods (such as vegetables, fruits, and whole grains) and a moderate amount of lean and low-fat, animal-based food (meat and dairy products)
Blood Type Diet	A system of eating that categorizes foods as beneficial, neutral, or harmful based on a person's blood type.
Mediterranean Diet	A diet which emphasizes plant-based foods and healthy fats, eating mostly veggies, fruits and whole grains.
Vegetarianism	The practice of abstaining from the consumption of meat. It may also include abstaining from eating all by-products of animal slaughter.
Veganism	The practice of abstaining from the consumption and use of animal products.
High Protein Diet	A high-protein diet is a diet in which 20% or more of the total daily calories comes from protein. Most high protein diets are high in saturated fat and severely restrict intake of carbohydrates.
High Fibre Diet	A diet rich in fibre with adequate fluids can help shorten the time food takes to move through the intestines and increase stool weight to prevent constipation.

ALCOHOL CONSUMPTION

Does the participant drink alcohol?

Response Options: Yes; No

How frequently does the participant drink alcohol?

Response Options: Usually; Sometimes; Rarely; Occasionally

- **Frequency Descriptions**:
  - Usually (>4 days a week);
  - Sometimes (1-4 days a week);

	- Rarely (<3 days a month) - Occasionally (Monthly or Yearly)
	SMOKING STATUS Has the participant ever smoked a nicotine cigarette? **Response Options**: Yes; No What is the participant's smoking status? **Response Options**: An EVERY day smoker; A FAIRLY REGULAR (some days) smoker; A FORMER smoker; NON-smoker
	- **Nicotine** is a substance found in all tobacco products and some e-cigarette liquids.
	PHYSICAL ACTIVITY [Moderate Physical Activity Frequency - Work Moderate Physical Activity Frequency - Recreational High Intensity Physical Activity Frequency - Work High Intensity Physical Activity Frequency - Recreational Physical Activity Frequency - During Travel/Commute] **Response Options**: Usually; Sometimes; Rarely How much time does the participant usually spend sitting or reclining on a typical day?

	- **Examples** of high-intensity work activities: carrying or lifting heavy loads, digging or construction work
	- **Examples** of moderate-intensity work activities: brisk walking, carrying light loads
	- **Examples** of high-intensity recreational activities: competitive training, running or football
	- **Examples** of moderate-intensity recreational activity: brisk walking, cycling, swimming, volleyball
	- **Frequency Descriptions**:
	  - Usually (>4 days a week);
	  - Sometimes (1-4 days a week);
	  - Rarely (<3 days a month)

Abbreviations

BMI: Body-Mass Index
DEXA: Dual-energy X-ray absorptiometry
DNA: Deoxyribonucleic acid
ID: Identification
MUAC: Middle Upper Arm Circumference

Protocols Administration

Mode of Administration

	Protocols								
	A	B	C	D	E	F	G	H	I
Interview OR Self-administered questionnaire		X	X	X	X	X	X	X	X
Retrospective Recording	X	X	X						X
Clinical assessment									
Bioassay/Lab-based assessment									

Life Stage

	Protocols								
	A	B	C	D	E	F	G	H	I
Infancy (0 - 12 months)	X	X	X	X					
Toddler (13 - 24 months)	X	X	X	X					
Childhood (2-11 years)	X	X	X	X			X	X	X
Adolescence (12 - 18 years)	X	X	X	X	X	X	X	X	X
Adult (18 and older)	X	X	X	X	X	X	X	X	X

Personnel and Training Required

Protocols B to I may be implemented as either self-reported questionnaires or interviewer-administered questionnaires. If interviewer-administered, interviews should be conducted by

trained or study coordinators or data collectors who speak the native/local language of the target population. **Protocols A and I** are retrospectively recorded from previous data records.

References

The Microbiome Research Data Toolkit is based on and aligned with several existing standards, to facilitate data annotation and harmonisation. These resources are listed below:

1. Minimal Information about any Sequence Standard (MIxS)
2. MixS: MIMS (Metagenome or Environmental)
3. African Microbiome Portal Template
4. ENA Metadata Checklist
5. SRA Metadata Template

Contributors

H3Africa Microbiome WG
H3Africa Phenotype Harmonisation WG
Ovokeraye Oduaran
Adetola F. Louis-Jacques
Imane Allali
Julius Mulindwa
Lamech Mwapagha
Lyndon Zass
Mariem Hanachi

Contact Us

For queries related to this standard and guideline, users can log a ticket to the Phenotypes Standards queue in the H3ABioNet Helpdesk. User feedback and improvements on the current module are welcome and encouraged. These can also be submitted through the Helpdesk, or on our GitHub Issues page.