Standardized metagenome naming system in GOLD

Genetic material that is collected directly from the environment is called a metagenome. While the name of an individual organism is determined by a standardized nomenclature system, there is no well-defined rule to name a metagenome. Different aspects of the environment such as the habitat, geographic location, collection site are important descriptors to accurately define a metagenomic sample.

It is logical that the name of the metagenome should reflect those different aspects to make it possible to build a database of metagenomic genes from comparable environments. Based on that, GOLD developed its own naming convention for metagenomes.

Examples of unconventional and unacceptable names:

Saliva_contig300
HK Metagenome T1
Hkbic1
US Sludge

According to the GOLD naming convention, the standardized name of metagenome must contain:

[Habitat] [Type of communities] [Location, including the country/ocean] – [Identifier]

Habitat describes the environment from which the sample came from such as soil, marine sediment, human feces, activated sludge etc.

Community identifies the type of organisms, which are present in the sample; it can be broad – microbial, or more specific - bacterial, archaeal, viral, fungal, or eukaryotic.

Location provides information about the geographic location of the sample and must include the name of the country or the ocean.

Identifier uniquely identifies the particular sample and distinguishes it from other closely related samples.

The GOLD naming convention applies to all four levels of metagenomic project classification: Biosample, Sequencing Project, Analysis Project and StudY. It should be kept in mind that the standardized name for a metagenome Study does not have a unique identifier and can be more general, to account for a wide variety of individual samples within a study.

The canonical naming system described above ensures that every environmental sample has a standardized name, which accurately describes its source environment and facilitates comparative analysis of related samples.

Select examples of Biosample /Sequencing Project /Analysis Project names for metagenomes from different environments:

Example 1. How does one name a metagenome examining microbial communities in water sample E0112 from western Lake Erie?

Freshwater microbial communities from western Lake Erie, Ohio, USA - E0112

Habitat: freshwater

Communities: microbial communities

Location: western Lake Erie, Ohio, USA

Identifier: E0112

Sample Collection Site: lake water

Example 2. How does one name a metagenome examining microbial communities in the forest soil sample from Jyndevad Skov Nat. Forest that were enriched in the laboratory?

Lab enriched soil microbial communities from Jyndevad Skov Nat. Forest in Herning, Denmark - Sample098

Habitat: soil (or forest soil)

Communities: microbial communities

Location: Jyndevad Skov Nat. forest in Herning, Denmark

Identifier: Sample098

Sample Collection Site: forest topsoil

Example 3. How does one name a metagenome examining archaeal communities in wastewater from WWTP in Schifflange?

Wastewater archaeal communities from WWTP in Schifflange, Luxembourg - D49

Habitat: wastewater

Communities: archaeal communities

Location: WWTP in Schifflange, Luxembourg

Identifier: D49

Sample Collection Site: sludge from aeration tank

Example 4. How does one name a metagenome examining viral communities in a sample of activated sludge from EBPR bioreactor in Queensland?

Activated sludge viral communities from EBPR bioreactor in Queensland, Australia - V92

Habitat: activated sludge

Communities: viral communities

Location: EBPR bioreactor in Queensland, Australia

Identifier: V92

Sample Collection Site: sludge from bioreactor

If a sample is **host-associated**, the information about the host should be included into the habitat (see examples below).

Example 5. How does one name a metagenome examining bacterial communities in a sample of feces from acute cholecystitis patient in NIID?

Human fecal bacterial communities from acute cholecystitis patient, NIID, Tokyo, Japan - AC-P6

Habitat: Human fecal

Communities: bacterial communities

Location: NIID, Tokyo, Japan

Identifier: AC-P6

Sample Collection Site: human feces from acute cholecystitis patient

Example 6. How does one name a metagenome examining microbial communities in deep-sea sponge from Gakkel Ridge?

Deep-sea sponge associated microbial communities from Gakkel Ridge, Arctic Ocean - PS101

Habitat: deep-sea sponge

Communities: microbial communities

Location: Gakkel Ridge, Arctic Ocean

Identifier: PS101

Sample Collection Site: sponge tissue

Example 7. How does one name a metagenome examining microbial communities in a rumen from dairy cattle in Bangalore?

Bovine rumen microbial community from dairy cattle in Bangalore, India - RS1

Habitat: bovine rumen

Communities: microbial community

Location: Bangalore, India

Identifier: RS1

Sample Collection Site: rumen from a cow

Common Errors to avoid:

1. **a random capitalization:**

INCORRECT:

Wastewater **Bacterial Communities** from **Medical Facility** sewage samples near Freiburg, Germany - A1

CORRECT:

Wastewater bacterial communities from medical facility sewage samples near Freiburg, Germany - A1

2. **misspelled words:**

INCORRECT:

Fecal bacterial communities from **acuet cholecistitis pateint**, NIID, Tokyo, Japan - P6

CORRECT:

Fecal bacterial communities from acute cholecystitis patient, NIID, Tokyo, Japan - P6

3. **a hyphen instead of a dash between the main part of the name and an identifier:**

INCORRECT:

Seawater microbial communities from coral reef in Red Sea, Saudi **Arabia-E2D**

CORRECT:

Seawater microbial communities from coral reef in Red Sea, Saudi Arabia - E2D