

## Standardized(Canonical) 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.

Metagenomes with cryptic and esoteric names are ambiguous to the general reader and make it impossible to build a database of metagenomic genes from comparable environments. Some examples of uninformative metagenome names entered by users are as follows:

- **Saliva\_contig300**
- **HK Metagenome T1**
- **Hkbic1**
- **US Sludge**

Such names, may make sense to an individual researcher who is submitting the metagenome, but does not include any information describing the environment of the sample, making it meaningless for researchers who will be looking at these individually or as part of comparative analysis. To avoid the shortcomings of cryptic names like above, GOLD follows a standardized naming system for metagenomes.

A GOLD metagenome name consists of four parts:

a) **Habitat**, which describes the environment from which the sample came from such as soil, marine sediment, human fecal, activated sludge etc.

b) **Community**, which identifies the type of organisms, which are present in the sample such as bacterial, viral, fungal or eukaryotic

c) **Location**, which provides information about the geographic location of the sample

d) **Identifier** that uniquely identifies the particular sample and distinguishes it from other closely related samples.

**Example Q1:** How does one name a metagenome that examines viruses from the waters of the Black Sea that are acidified?

**Answer:** Marine viral communities from Black Sea – Acidified-1

**Habitat** = Marine  
**Community** = viral communities  
**Location** = from Black Sea  
**Identifier** = Acidified-1

**Example Q2:** How does one name a metagenome examining bacterial communities from sewage samples from a medical facility in Germany?

**Answer:** Wastewater bacterial communities from medical facility sewage samples near Freiburg, Germany – A1

**Habitat** = Wastewater  
**Community** = bacterial communities  
**Location** = medical facility sewage samples near Freiburg, Germany  
**Identifier** = A1

This naming convention described above for a GOLD metagenomic sample also applies to Studies, Sequencing Projects and Analysis Projects. 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 facilitate comparative analysis of related samples.