Data Sets

We have gathered several data sets for you to use in the course, and in this manual we use different data sets as examples.

The main thing you need to know is whether a data set is a 16S amplicon library or a random community metagenome.

Coral and Algae

This is coral, algae, CCA, and water (control) samples from Kevin Walsh in Liz Dinsdale's lab.

This is a random community metagenome data set

Read more about the coral and algae data sets

Ground Water

This data comes from SRA project SRP075429

Hernsdorf AW, Amano Y, Miyakawa K, Ise K, Suzuki Y, Anantharaman K, Probst A, Burstein D, Thomas BC, Banfield JF. 2017. Potential for microbial H2 and metal transformations associated with novel bacteria and archaea in deep terrestrial subsurface sediments. ISME J 11:1915–1929

This is a random community metagenome data set

Read more about the ground water datasets

Gut

This data comes from SRA project SRP074153

This is a random community data set

Brooks B, Olm MR, Firek BA, Baker R, Thomas BC, Morowitz MJ, Banfield JF. 2017. Strain-resolved analysis of hospital rooms and infants reveals overlap between the human and room microbiome. Nat Commun 8:1814

Read more about the gut datasets

Drinking Water

A drinking water study from the University of Adelaide, Australia

This is 16S amplicon dataset with SRP ID SRP059994

Read more about the drinking water datasets