MGnify Services and annotation pipelines

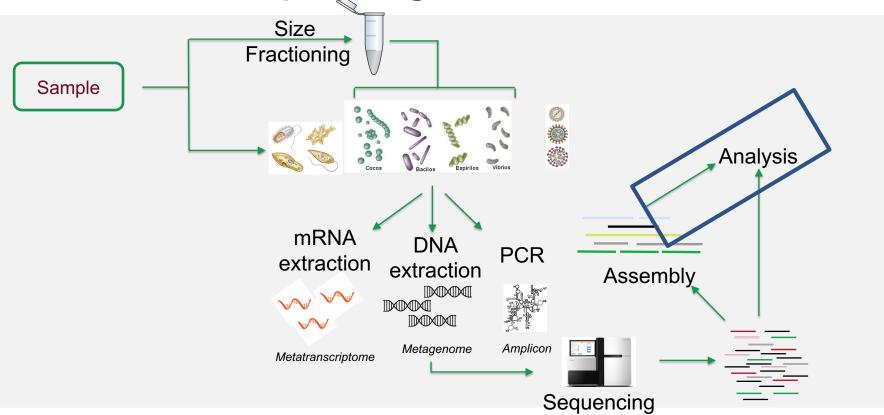


Varsha Kale

Bioinformatician



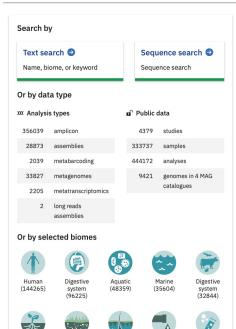
Microbiome Sequencing







Overview Submit data Text search 🖫 Sequence search Browse data 🖫 API About Help Login



Skin

(10861)

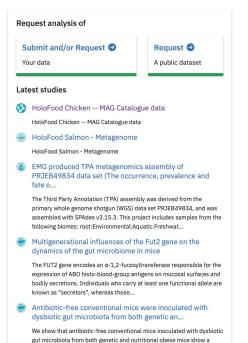
production

View all biomes

Plants

(28577)

(23805)

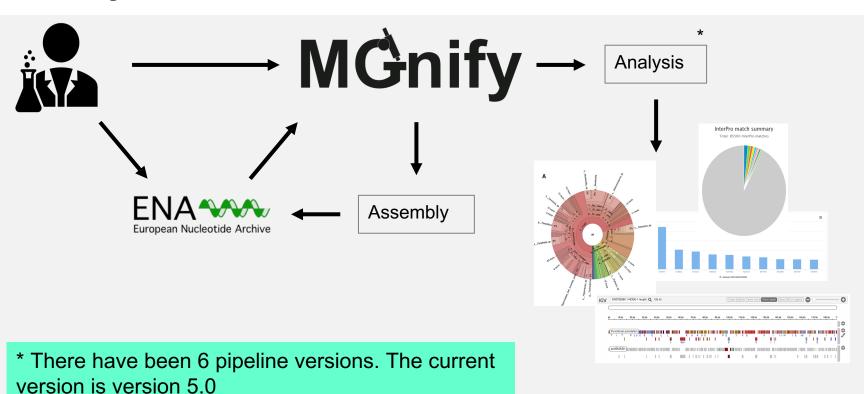


beneficial impact on the liver and the adipose tissue. The transfer

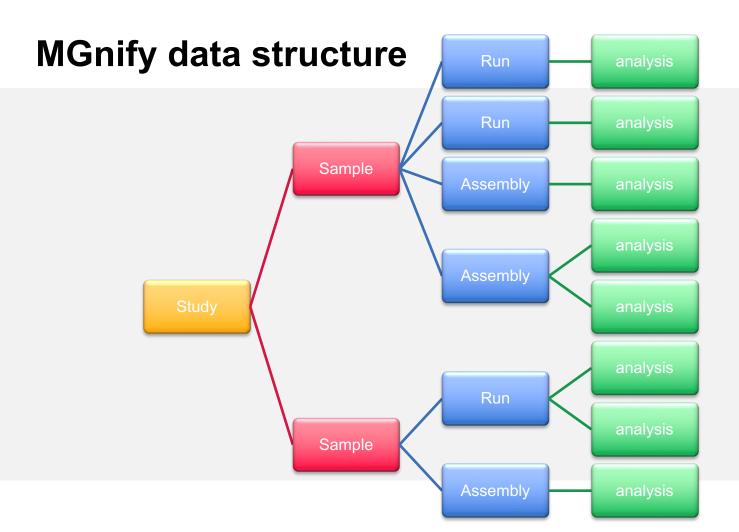
changes both the gut microbiota and the mi...

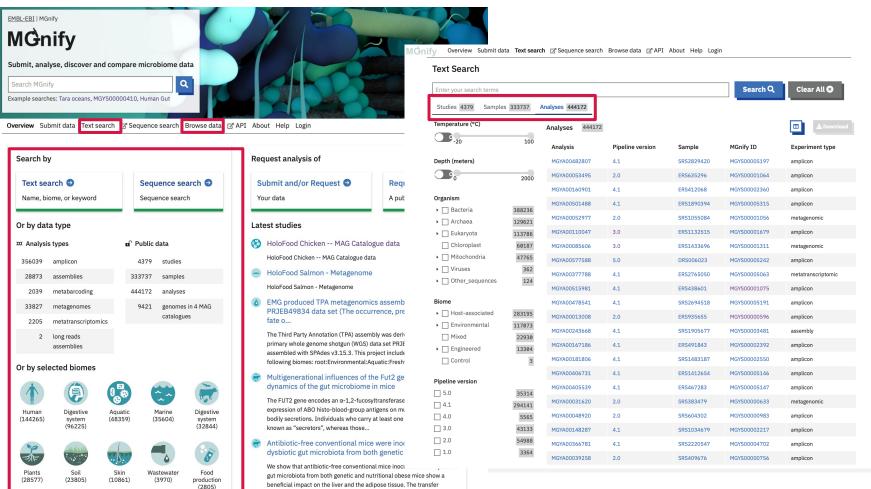
- Assembly, analysis and archiving of microbiome data
- Datatypes:
 - Metabarcoding/Amplicon
 - Metagenome
 - Metatranscriptome
 - Assembly
- High performance computing and standardised analysis pipelines.

MGnify data flow









changes both the gut microbiota and the mi...

View all biomes



A word about metadata

For data to have longevity and be useful to the scientific community, sequences need to be archived with contextual metadata

- How was sampled?
- How was it extracted?
- How was it stored?
- What sequencing platform was used?
- Where did it come from?
- What were the lat/long, depth, pH, salinity, temperature, etc.
- What were the clinical observations?









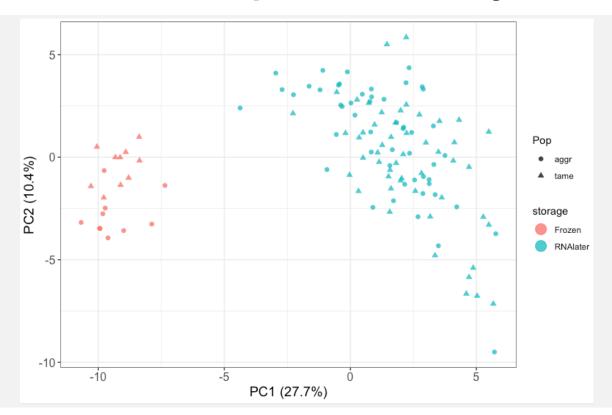








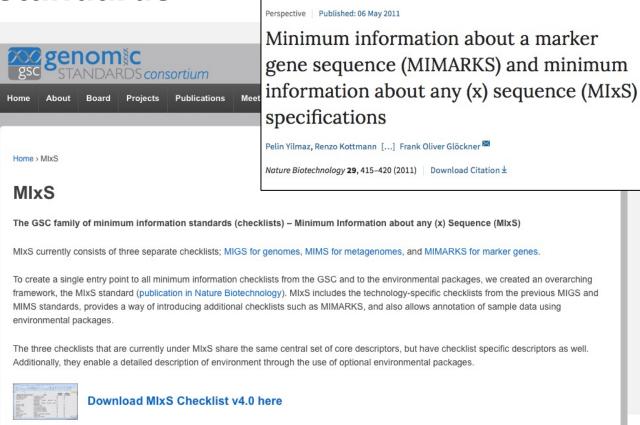
Metadata underpins data analysis



Courtesy of Lara C Puetz, University of Copenhagen



Data standards



nature

biotechnology

Metadata in MGnify

Sample overview (ERS3792294)

Sample RH0001-DID-DW-100-AUT

Last updated: Wed Mar 09 2022

Description

REHAB

External links

- ■ ENA website (ERS3792294)
- ■ EBI biosample (SAMEA5989477)

Classification



Sample metadata



investigation type: metagenome project name: REHAB geographic location (longitude): -1.251716 geographic location (country and/or sea,region): United Kingdom collection date: 2017-01-01 environment (biome): Sediment environment (feature): Sediment environment (material): Sediment sediment environmental package: sequencing method: Illumina elevation: 0.0

geographic location (latitude):

ENA checklist:

51.619392

ERC000021

EMBL-EBI

Metadata in MGnify

Sample overview (ERS3792294)

Sample RH0001-DID-DW-100-AUT

Last updated: Wed Mar 09 2022

Description

REHAB

External links

- ■ ENA website (ERS3792294)
- ☑ EBI biosample (SAMEA5989477)

Classification



Sample metadata Associated studies Analy

investigation type: metagenome REHAB project name: geographic location (longitude): -1.251716 geographic location (country and/or sea,region): United Kingdom collection date: 2017-01-01 environment (biome): Sediment environment (feature): Sediment environment (material): Sediment environmental package: sediment sequencing method: Illumina 0.0 elevation: geographic location (latitude): 51.619392

ENA checklist:

ERC000021



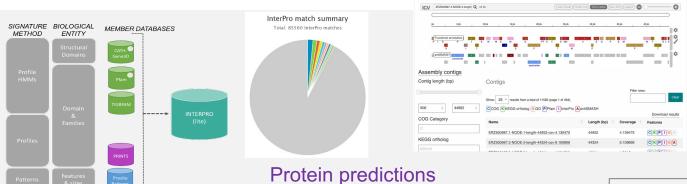
1 The organism where the microbiome is found

primates' frontal

× **Annotations from Europe PMC** Powered by Europe PMC Through automated text-mining, many publications may have additional metadata available in the form of management and attack mining, many publications may have additional metadata available in the form of management and attack mining many publications may have additional metadata available in the form of management management management may be additional metadata available in the form of management manag Sample processing Library strategy Other State **Engineered environment Body site** 1 Host body region/structure where microbiome is found frontal lobes Gene 1 Target gene(s) (e.g. hypervariable regions of 16s/18s rRNA gene) mas5 Host

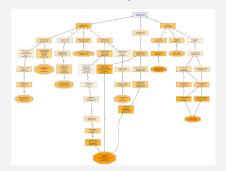


MGnify pipeline



Contig Browser

Functional protein and pathway annotations





Marker gene taxonomy

Analysis

Quality control

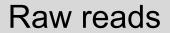
Taxonomic annotation



MGnify pipeline

- Pipeline implemented in CWL and run using Toil workflow engine
- Updates to existing tools and databases
- New tools and databases introduced



















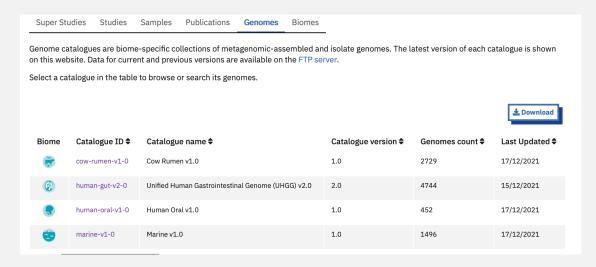






What else is available?

- R and python notebook servers for fetching metadata and comparative analysis
- Genomes



Protein database



Acknowledgements

MGnify

Rob Finn

Lorna Richardson

Alex Almeida

Ekaterina Sakharova

Martin Beracochea

Sandy Rogers

Tanya Gurbich

Ekaterina Sakharova

Ales Escobar Zepeda

Juan Caballero



- Josie Burgin
- Guy Cochrane



- Isuru Liyanage
- Dipayan Gupta







