Description of seqenv internal pipeline structure

Lucas Sinclair, lucas.sinclair@me.com

Last modified:

Thu Jul 23 2015

User input
(There is only one sequence here)

Search result against NT database
(max 10 hits so max 10 G-ids)

Every GI has 0 or 1 isolation source

Every isolation source has 0, 1 or more ENVO terms

All numbers are fake
This document just describes a plan for processing data.

"seq_to_gis"

OTU_2
also called C1
AtCGCGTCAGTCATGCAT
GCATCGTAATCGtCGCGT
CAGTCATGCATGCATCGT
AATCGAtCGCGTCAGTCA
TGCATGCATCGTAATCGA
tCGCGTCAGTCATGCATG
CATCGTAATCG

G-id (global identifier)
343206452

Isolation source
the deepest sea mud of the Mariana Trench

ENVO term
lake
ENVO:00000020

ENVO term
ground water
ENVO:00002041

G-id (global identifier)
324498760

Isolation source
isolated from terrestorial samples near Terra Nova Bay, Antarctica.

ENVO term
river bank
ENVO:00000143

Count it 1

ENVO term
ground water
ENVO:00002041

Count it 1

ENVO term
wetland
ENVO:00000043

Count it 1

G-id (global identifier)
324499999

G-id (global identifier)
324498738

Isolation source
My bath

None

G-id (global identifier)
324498738

None