

Description of seqenv's main internal data structures

Lucas Sinclair, lucas.sinclair@me.com

Last modified:

Tue Mar 15 2016

seq_to_gis	<i>One to many</i>
C0 -> [343206452, 343204350, 199999999]	
C1 -> [234523451, 343204354]	
C2 -> []	

unique_gis	<i>Set</i>
343206452, 343204350, 199999999, 234523451	

source_db	<i>One to three</i>		
343206452 -> "the deepest sea mud of the Mariana Trench"	pb421341233	ENV012, ENV034, ENV0534	
343204354 -> "isolated from terrestorial samples near ..."	pb421341233	ENV099999	
343204350 -> "the deepest sea mud of the Mariana Trench"	pb534533443	ENV0155	
... etc etc			

gis_with_envo	<i>Set</i>
343206452, 199999999	

seq_to_counts	<i>One to many to one</i>
C0 -> (ENV0:00002007 -> 0.667, ENV0:00002041 -> 0.333)	
C1 -> (ENV0:00002007 -> 0.123, ENV0:00002041 -> 0.821)	
C2 -> ()	

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