



Filtered 16S rDNA amplicons were sorted in decreasing order.

UCLUST  
Identity  $\geq 98\%$   
Coverage  $\geq 90\%$

Sequences of uclust-OTU centroids were defined for the longest sequence in each OTU

Some sequences for the OTU centroids predicted by CHIMERA using reference and *de novo* modes of UCHIME were discarded

Remaining sequences of OTU centroids performed "All-against-all BLASTN"

MCL  
Identity  $\geq 97\%$   
Coverage  $\geq 80\%$

Sequences of OTU centroids were the largest size of uclust-OTU size in each OTU.

Sequences of OTU centroids were annotated into SILVA SSU database using BLASTN. Identity  $\geq 97\%$ , Coverage  $\geq 80\%$

Calculating Canberra distance across samples.

	A1	A2	A3	B1	B2	B3	C1	C2	C3
A1	0	3552	3602	4440	4456	4474	4309	4305	4194
A2	3552	0	3546	4445	4433	4457	4243	4176	4022
A3	3602	3546	0	4410	4427	4445	4143	4079	3888
B1	4440	4445	4410	0	4047	3976	4373	4295	4345
B2	4456	4433	4427	4047	0	3916	4396	4349	4355
B3	4474	4457	4445	3976	3916	0	4395	4312	4339
C1	4309	4243	4143	4373	4396	4395	0	3728	3915
C2	4305	4176	4079	4295	4349	4312	3728	0	3560
C3	4194	4022	3888	4345	4355	4339	3915	3560	0

Selecting the query and reference samples.

		reference data					
		A1	A3	B2	B3	C1	C2
query data	A2	3552	3546	4433	4457	4243	4176
	B1	4440	4410	4047	3976	4373	4295
	C3	4194	3888	4355	4339	3915	3560

Determining the identity of the query.

	$D_A(q,r)$	$D_B(q,r)$	$D_C(q,r)$	$\min D(a,r)$
A2	3549	4445	4209	$D_A(q,r)$
B1	4425	4011	4334	$D_B(q,r)$
C3	4041	4347	3738	$D_C(q,r)$