

Supplementary material for SILVA, RDP, Greengenes, NCBI and OTT — How do these tax- onomies compare?

Part A: Mapping dissimilarities

	Strict	Loose	Path	Sparklines
RDP to SILVA	0.51	0.06	0.01	█ - -
Greengenes to SILVA	0.21	0.12	0.03	- - -
NCBI to SILVA	0.51	0.34	0.08	█ █ -
OTT to SILVA	0.48	0.32	0.06	█ █ -
SILVA to RDP	0.68	0.48	0.02	█ █ -
Greengenes to RDP	0.21	0.15	0.02	- - -
NCBI to RDP	0.97	0.88	0.05	█ █ █
OTT to RDP	0.99	0.82	0.08	█ █ █
SILVA to Greengenes	0.70	0.67	0.03	█ █ █
RDP to Greengenes	0.53	0.50	0.01	█ █ █
NCBI to Greengenes	0.97	0.94	0.07	█ █ █
OTT to Greengenes	0.99	0.96	0.11	█ █ █
SILVA to NCBI	0.40	0.06	0.02	█ - -
RDP to NCBI	0.50	0.02	0.01	█ - -
Greengenes to NCBI	0.19	0.10	0.02	- - -
OTT to NCBI	0.29	0.12	0.00	█ - -
SILVA to OTT	0.42	0.06	0.02	█ - -
RDP to OTT	0.54	0.03	0.01	█ - -
Greengenes to OTT	0.26	0.11	0.02	- - -
NCBI to OTT	0.25	0.02	0.00	█ - -

Table S1: Mapping dissimilarities for all pairs of taxonomies. Sparklines indicate the magnitudes of strict, loose and path mapping dissimilarities for each pair of taxonomies and are provided as a quick overview of the table.

	SILVA	RDP	Greengenes	NCBI	OTT
Root	1	1	1	1	1
Domain	3	8	2	5	3
Phylum	256	87	88	214	213
Class	585	186	248	304	357
Order	1126	403	404	1392	1575
Family	1409	1227	513	8391	15584
Genus	8529	4175	1200	79573	214991

Table S2: Number of nodes at each rank in the five taxonomies.

Part B: Detailed mapping summaries

Below we provide parallel sets plots and heatmaps that give a broader overview of each mapping than a summary score as shown in the main publication.

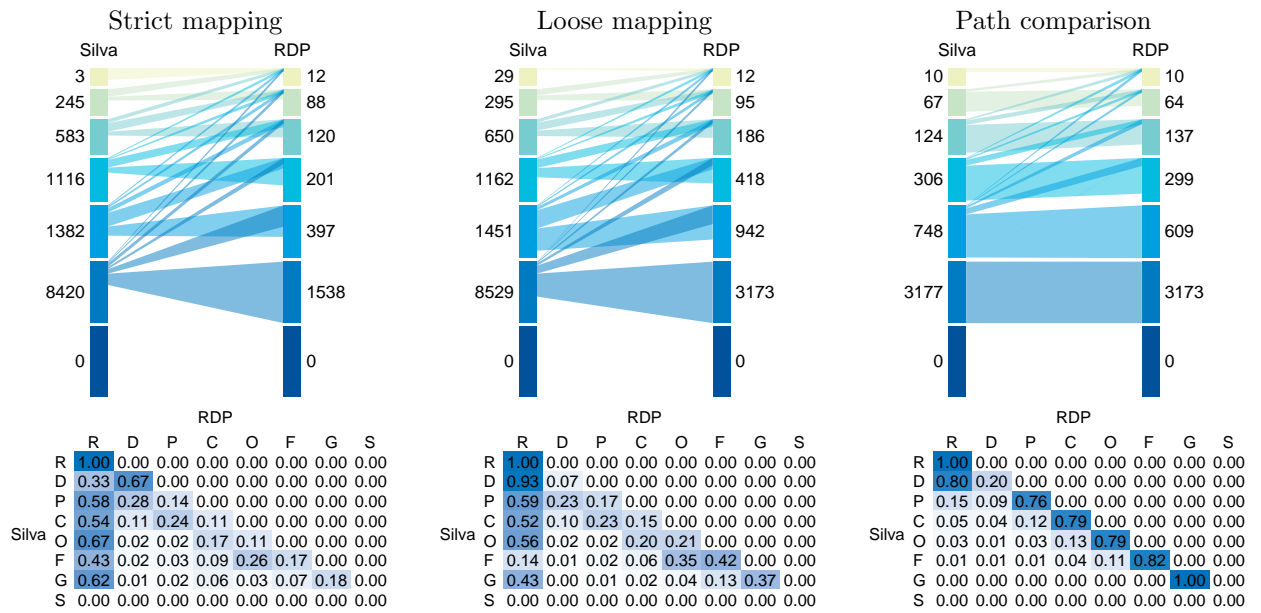
Parallel sets

Bars in the parallel sets plots are of constant size for each rank and the actual number of nodes is indicated by the number on the side. Parallel flow lines indicate perfect mappings whereas lines going upwards show mapping to higher ranks. Ranks are displayed in the same order in all of the plots (from the top): domain, phylum, class, order, family, genus, species.

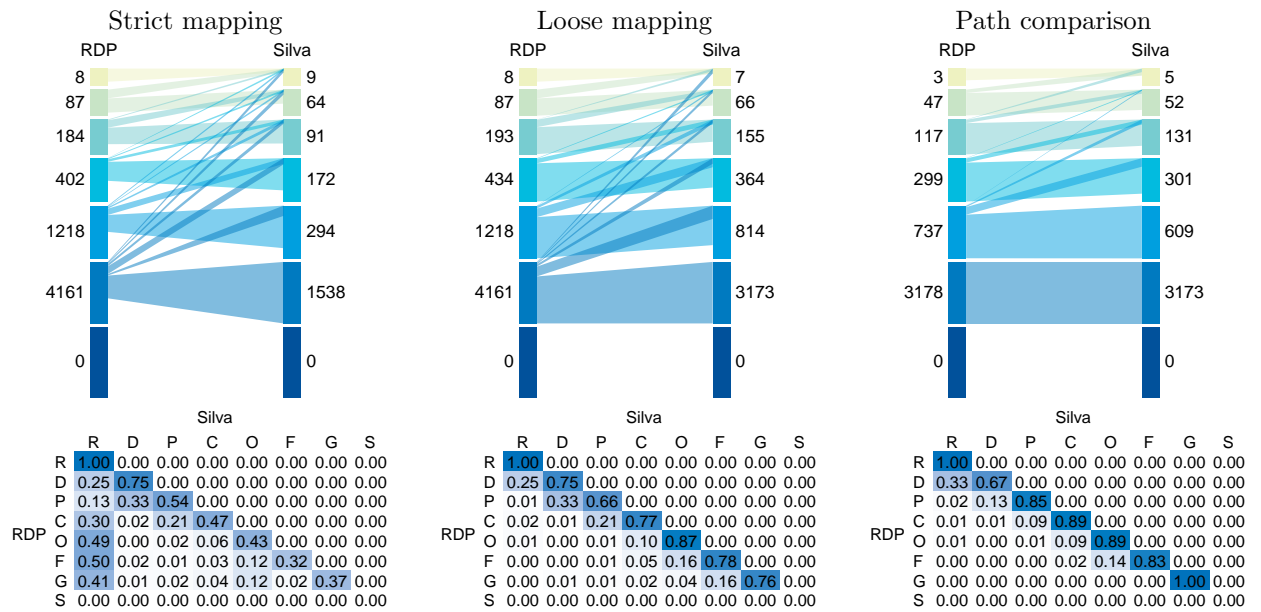
Heatmaps

Values in heatmaps are normalized by each row. Bright diagonal indicates that most of the nodes were mapped to the same rank, whereas a column popping out signifies high incompatibility below the rank associated with the column.

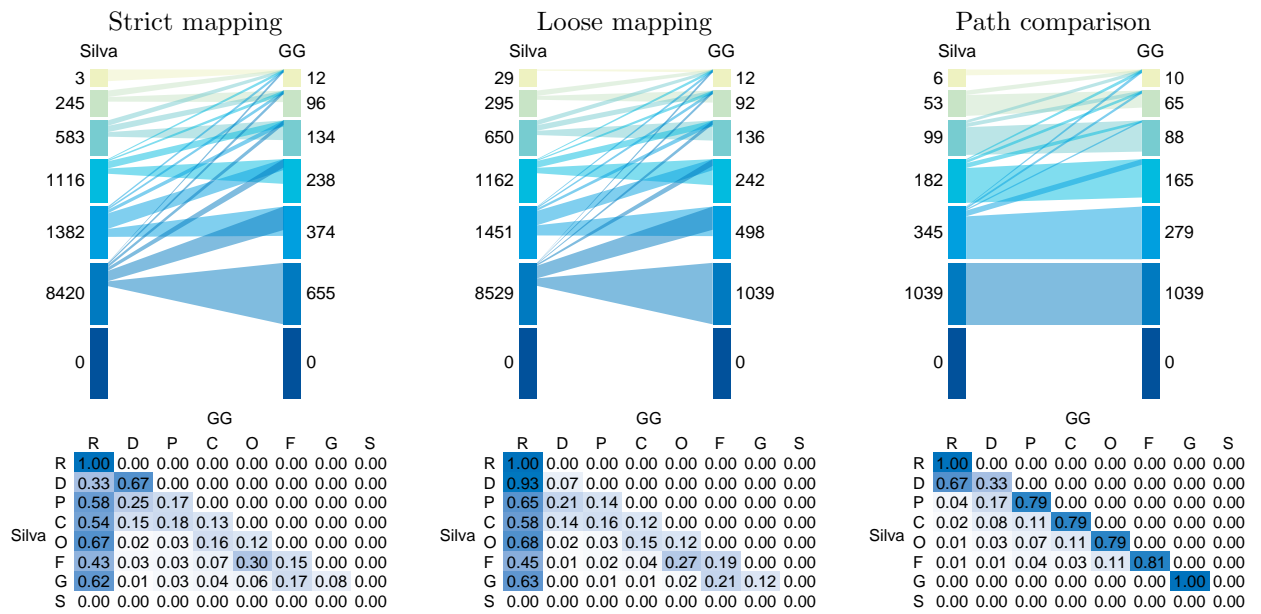
SILVA to RDP



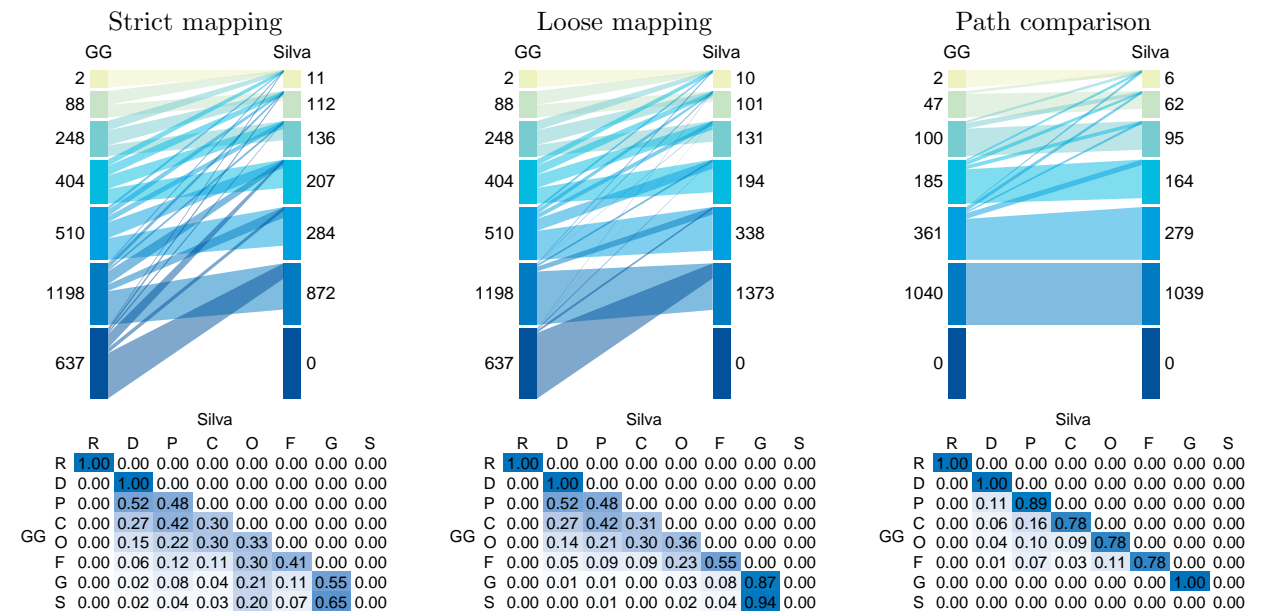
RDP to SILVA



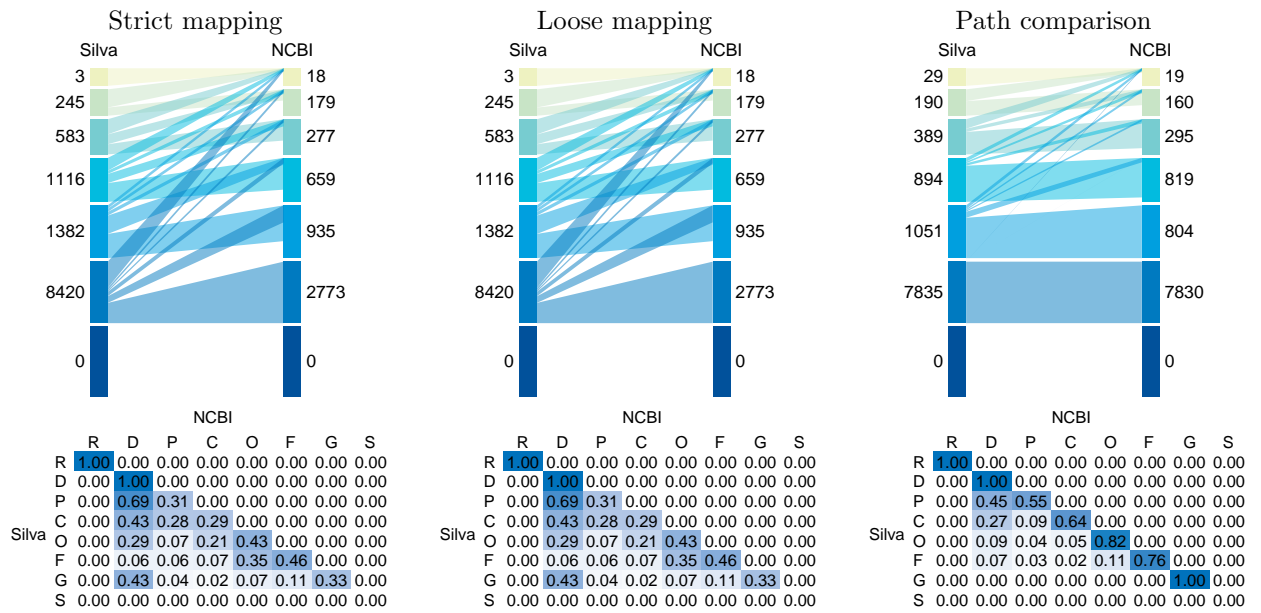
SILVA to Greengenes



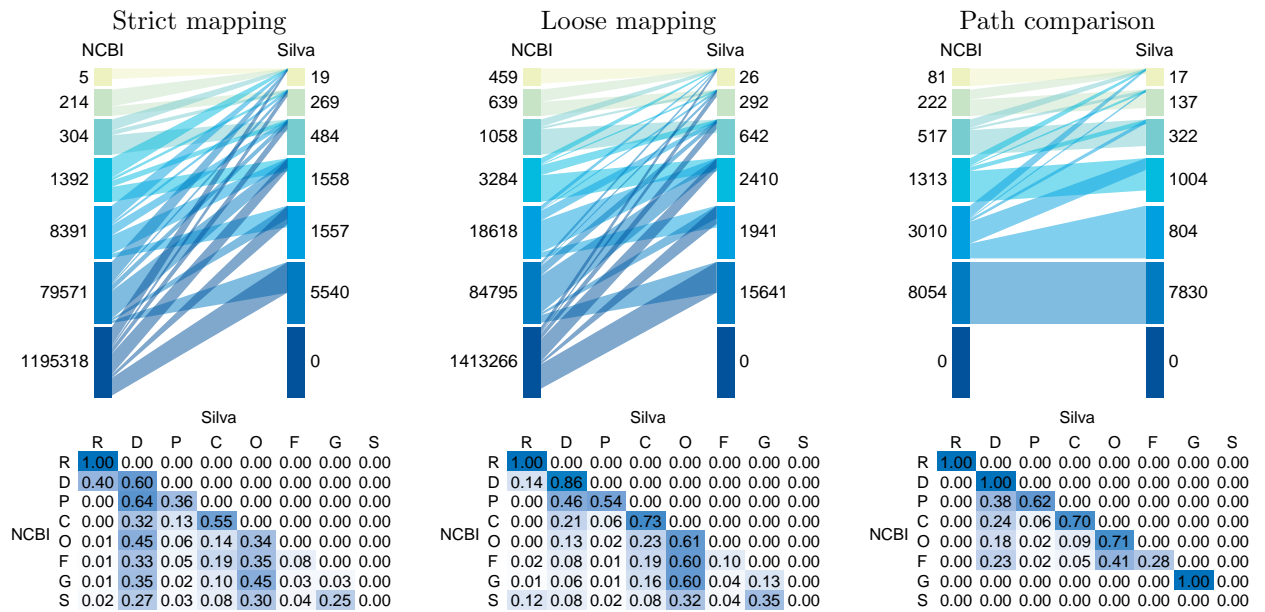
Greengenes to SILVA



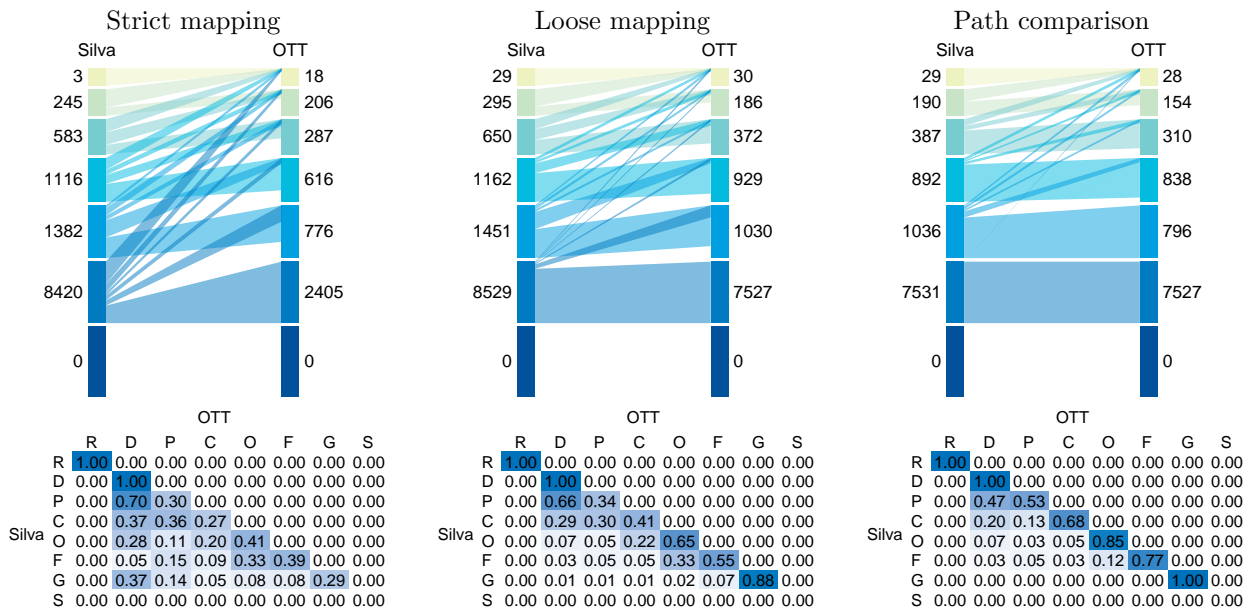
SILVA to NCBI



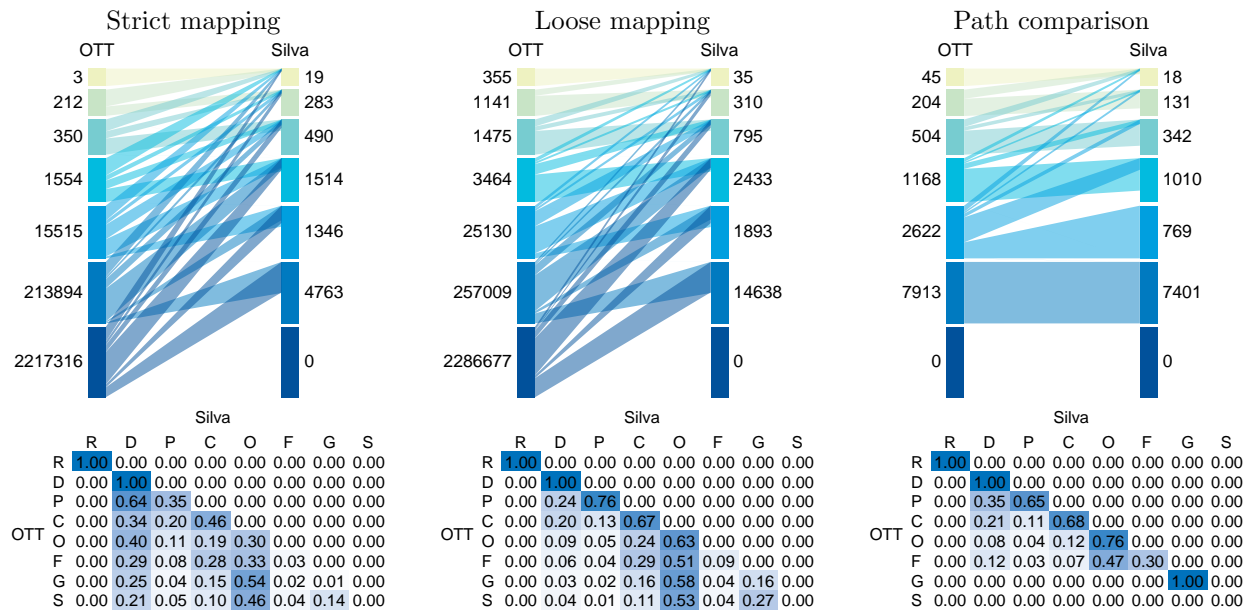
NCBI to SILVA



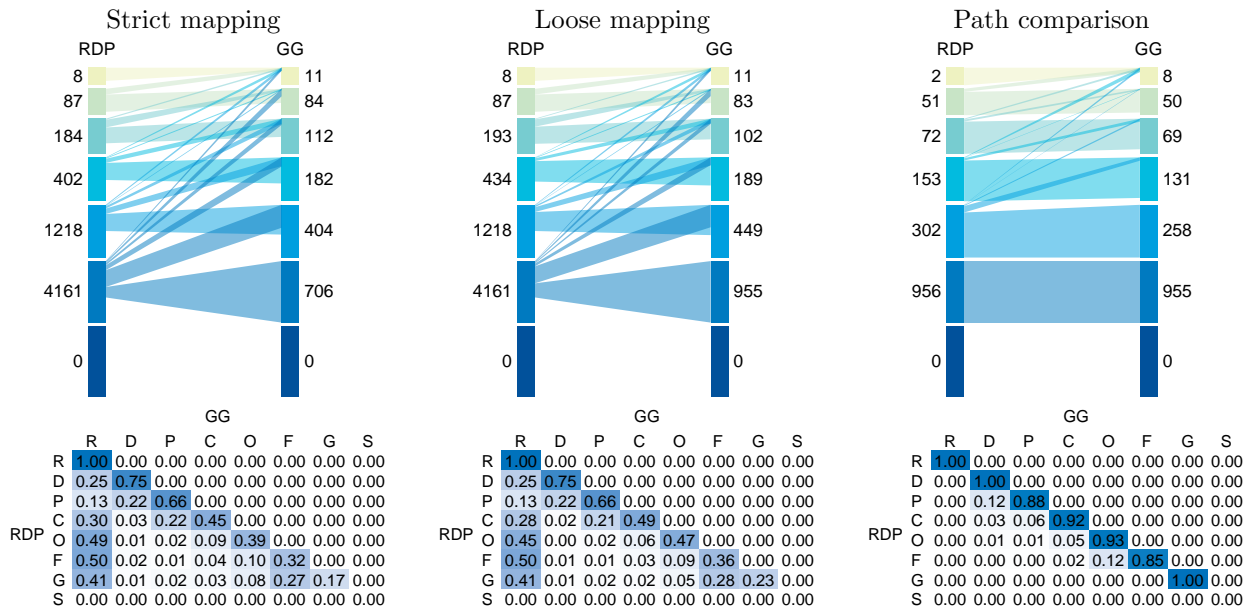
SILVA to OTT



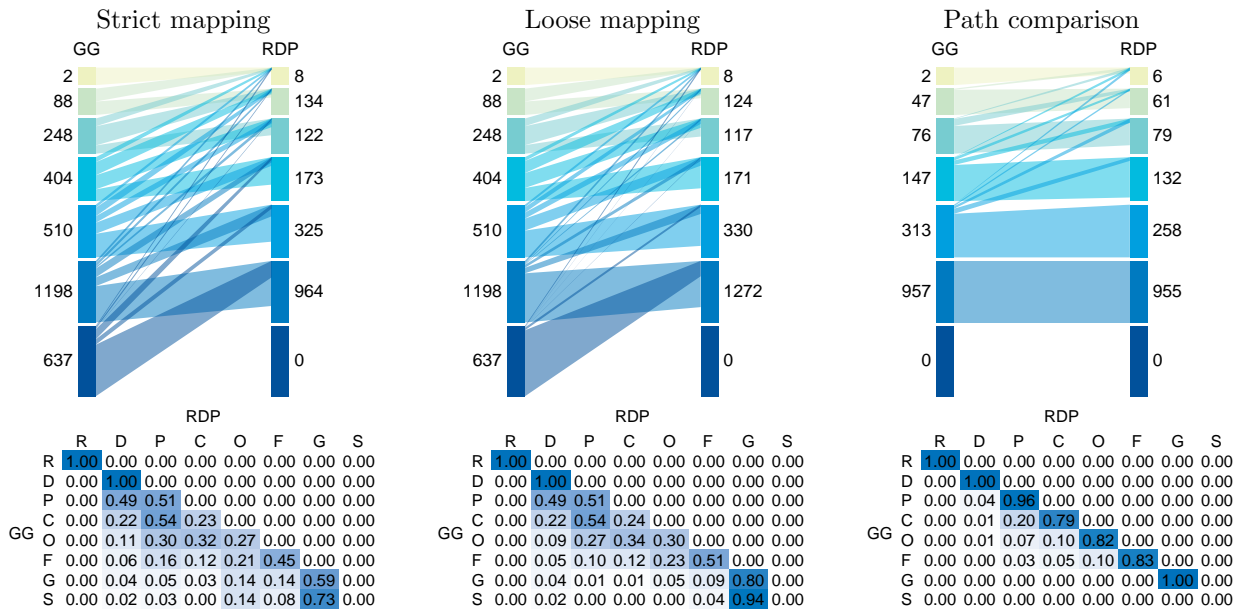
OTT to SILVA



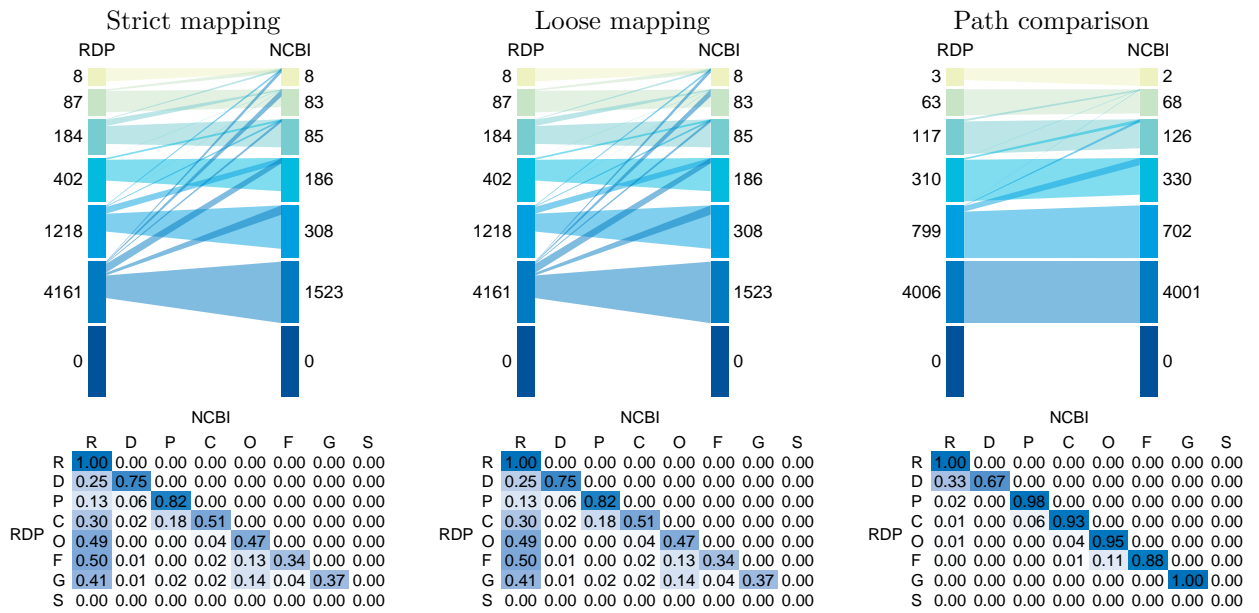
RDP to Greengenes



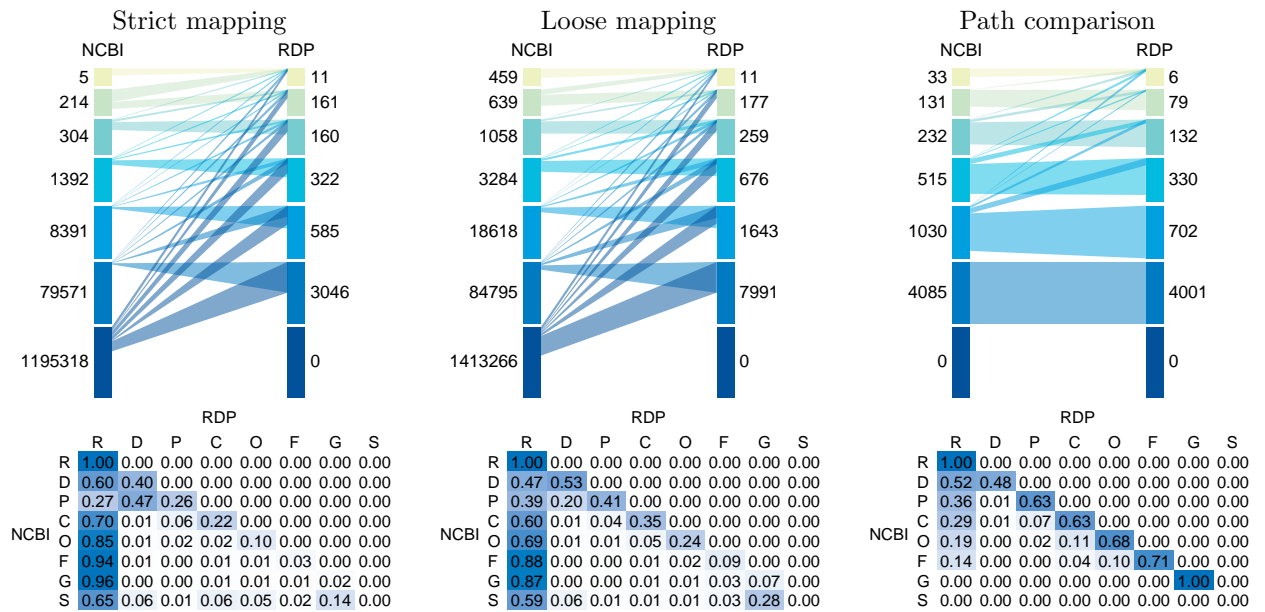
Greengenes to RDP



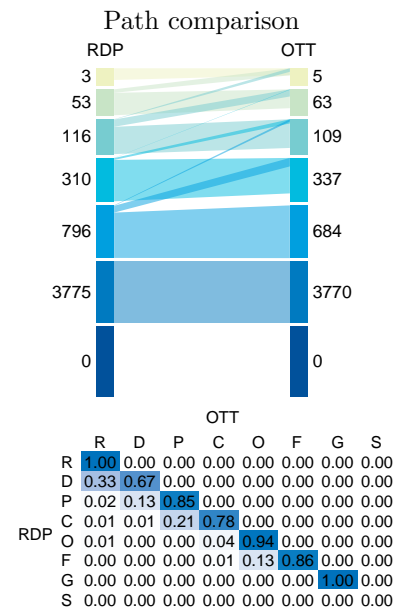
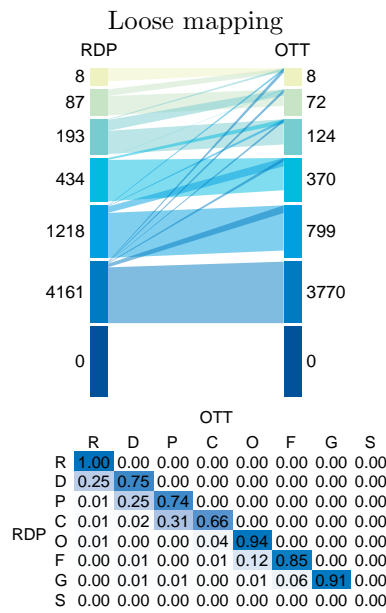
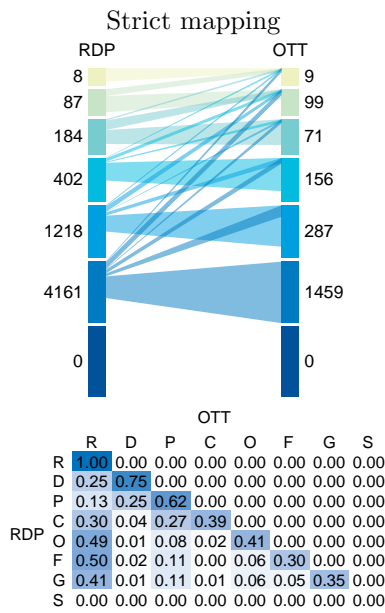
RDP to NCBI



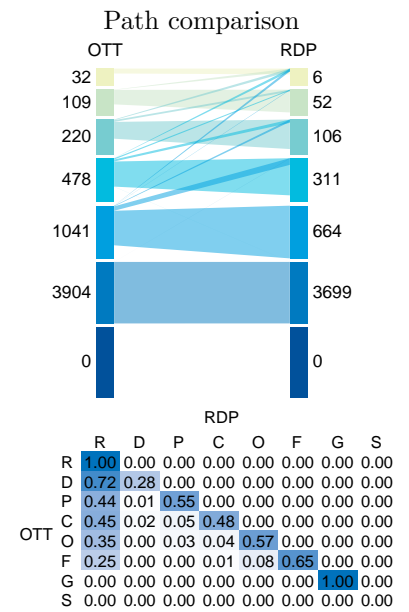
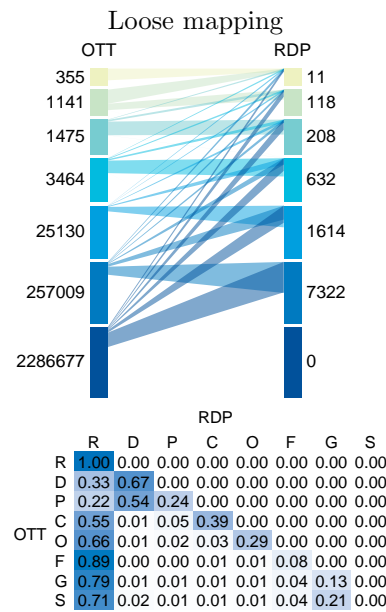
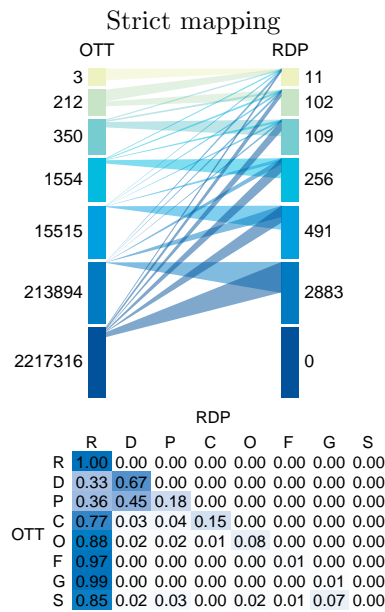
NCBI to RDP



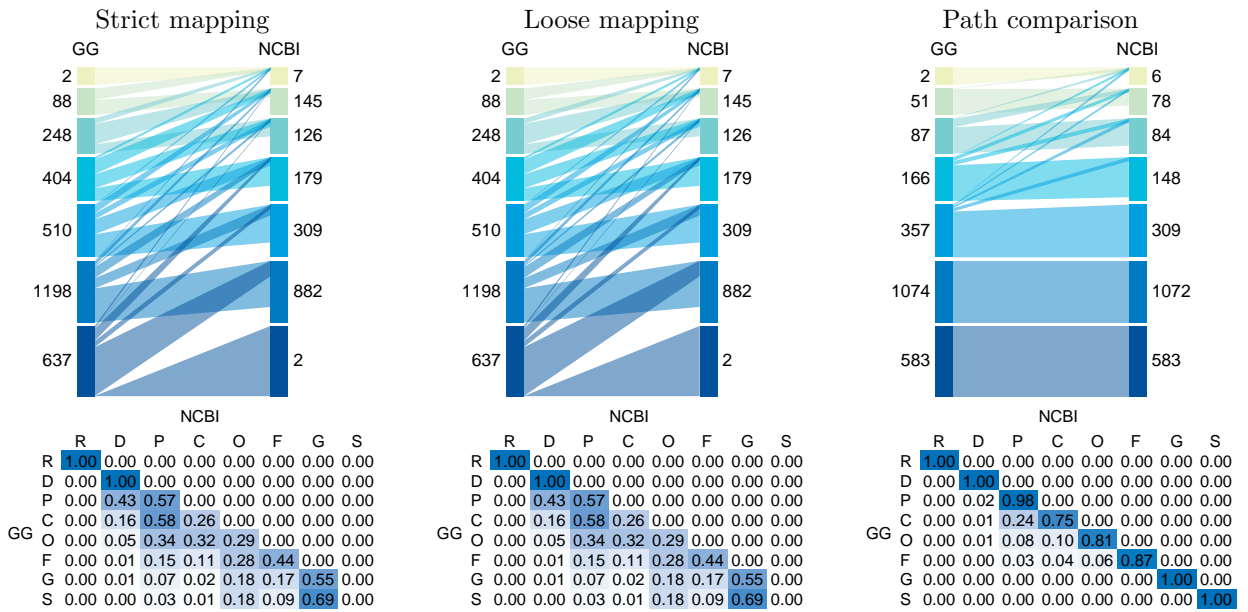
RDP to OTT



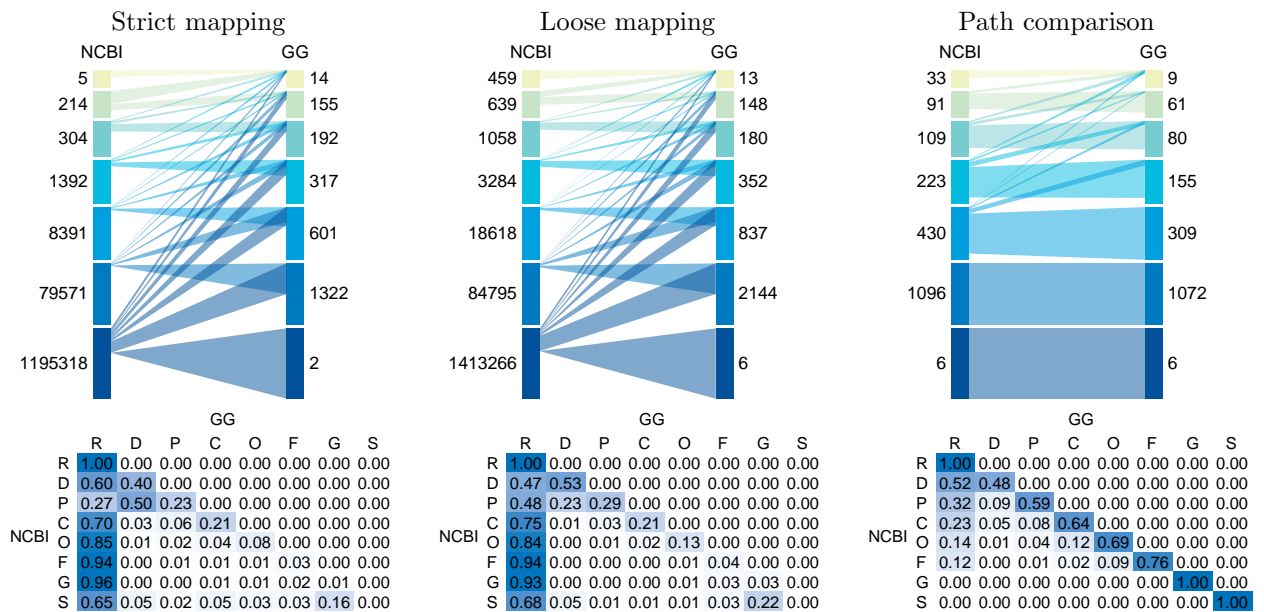
OTT to RDP



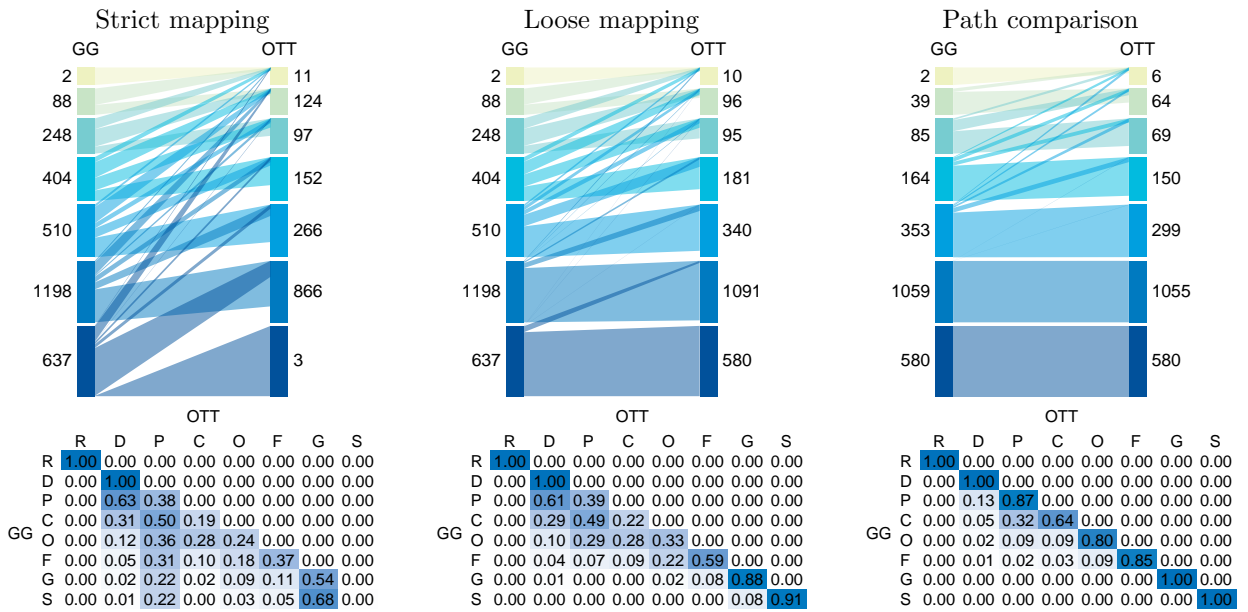
Greengenes to NCBI



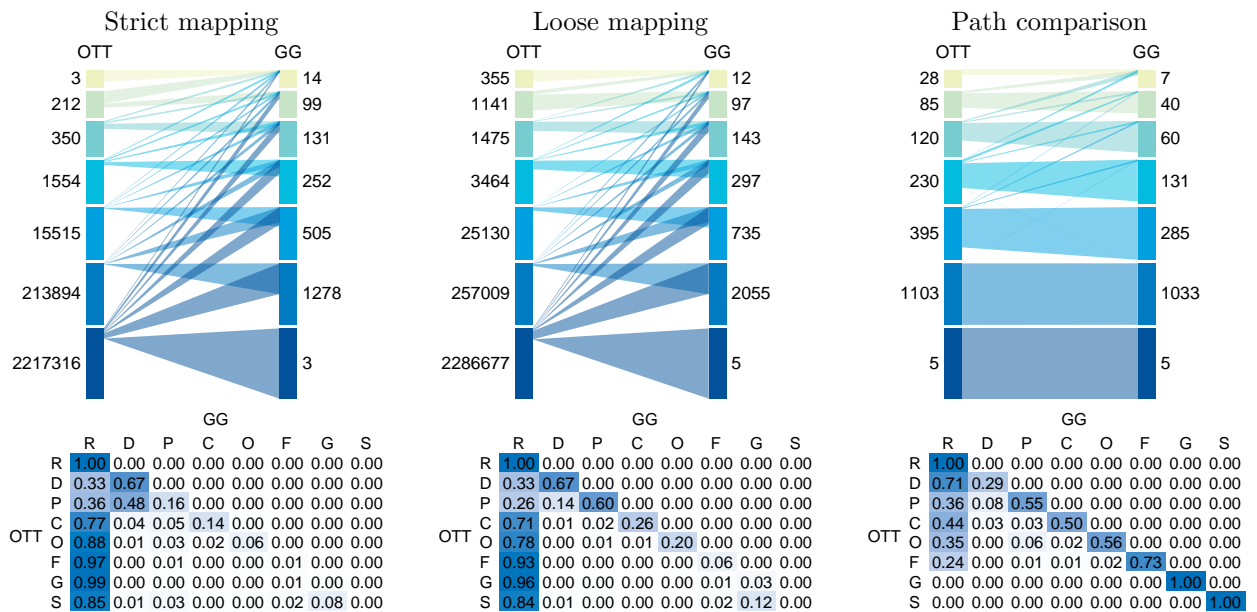
NCBI to Greengenes



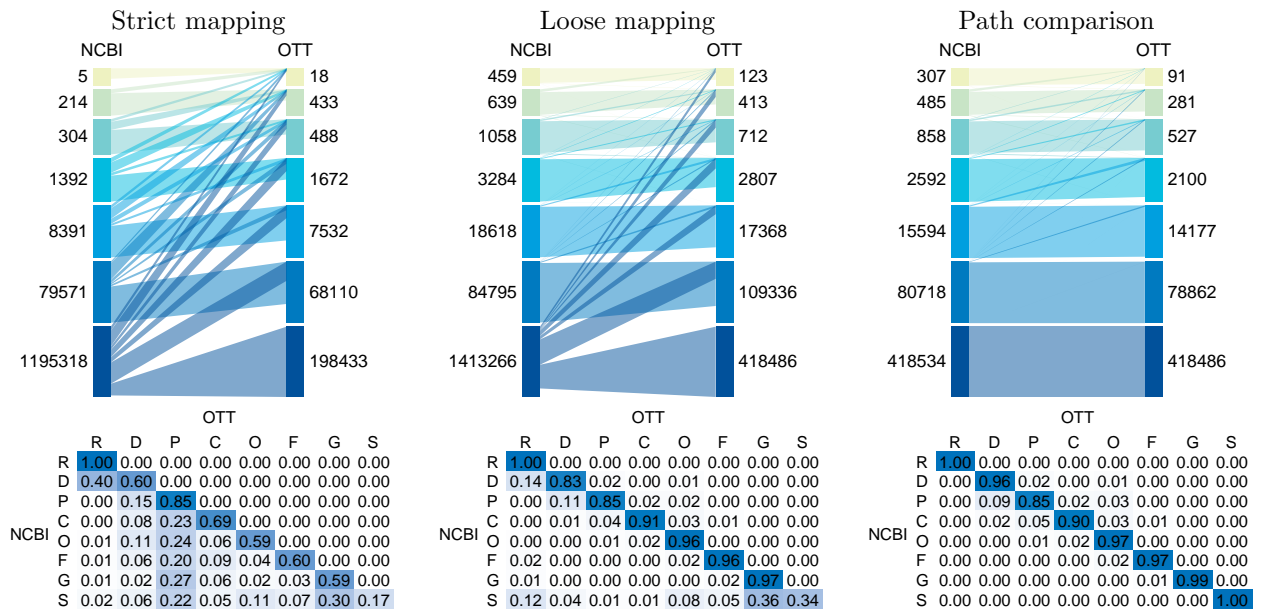
Greengenes to OTT



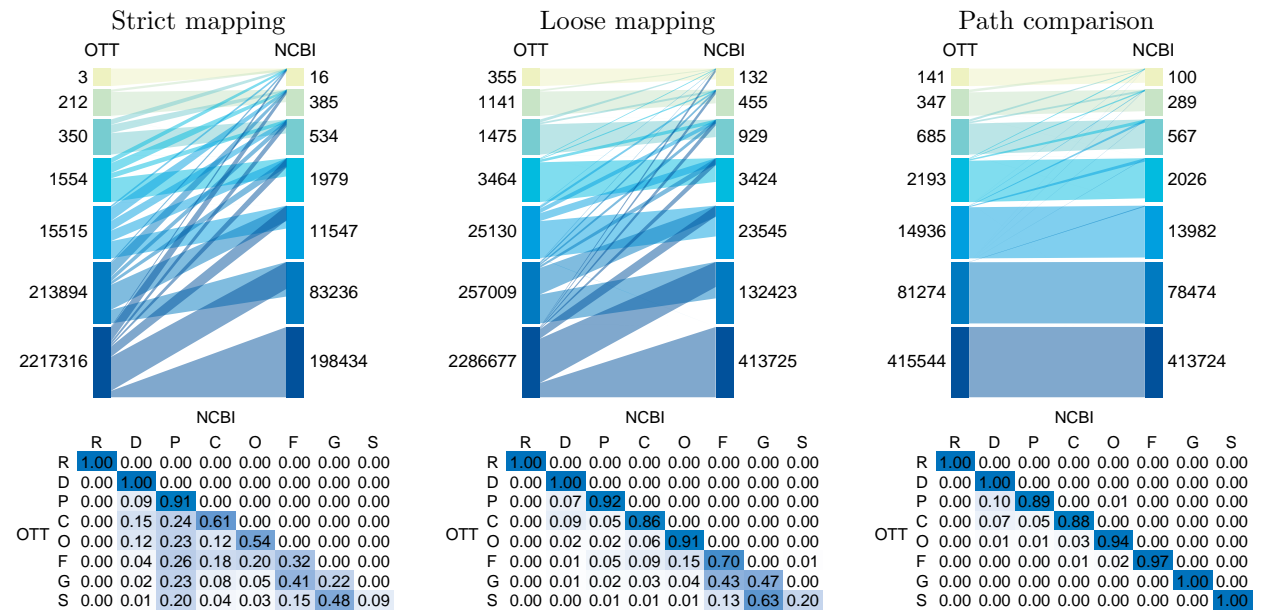
OTT to Greengenes



NCBI to OTT



OTT to NCBI



Part C: Comparisons limited to Bacteria and Archaea domains

	Strict	Loose	Path	Sparklines
RDP to SILVA	0.14	0.03	0.01	- - -
Greengenes to SILVA	0.21	0.12	0.03	- - -
NCBI to SILVA	0.18	0.08	0.02	- - -
OTT to SILVA	0.28	0.13	0.00	■ - -
SILVA to RDP	0.21	0.10	0.01	- - -
Greengenes to RDP	0.21	0.15	0.02	- - -
NCBI to RDP	0.23	0.12	0.01	- - -
OTT to RDP	0.30	0.19	0.00	■ - -
SILVA to Greengenes	0.24	0.16	0.02	- - -
RDP to Greengenes	0.17	0.13	0.01	- - -
NCBI to Greengenes	0.25	0.16	0.02	- - -
OTT to Greengenes	0.32	0.21	0.01	■ - -
SILVA to NCBI	0.15	0.07	0.02	- - -
RDP to NCBI	0.12	0.03	0.00	- - -
Greengenes to NCBI	0.19	0.10	0.02	- - -
OTT to NCBI	0.23	0.08	0.00	■ - -
SILVA to OTT	0.23	0.08	0.02	- - -
RDP to OTT	0.19	0.04	0.01	- - -
Greengenes to OTT	0.26	0.11	0.02	- - -
NCBI to OTT	0.19	0.04	0.02	- - -

Table S3: Mapping dissimilarities for all pairs of taxonomies limited to Bacteria and Archaea domains. Sparklines indicate the magnitudes of strict, loose and path mapping dissimilarities for each pair of taxonomies and are provided as a quick overview of the table.

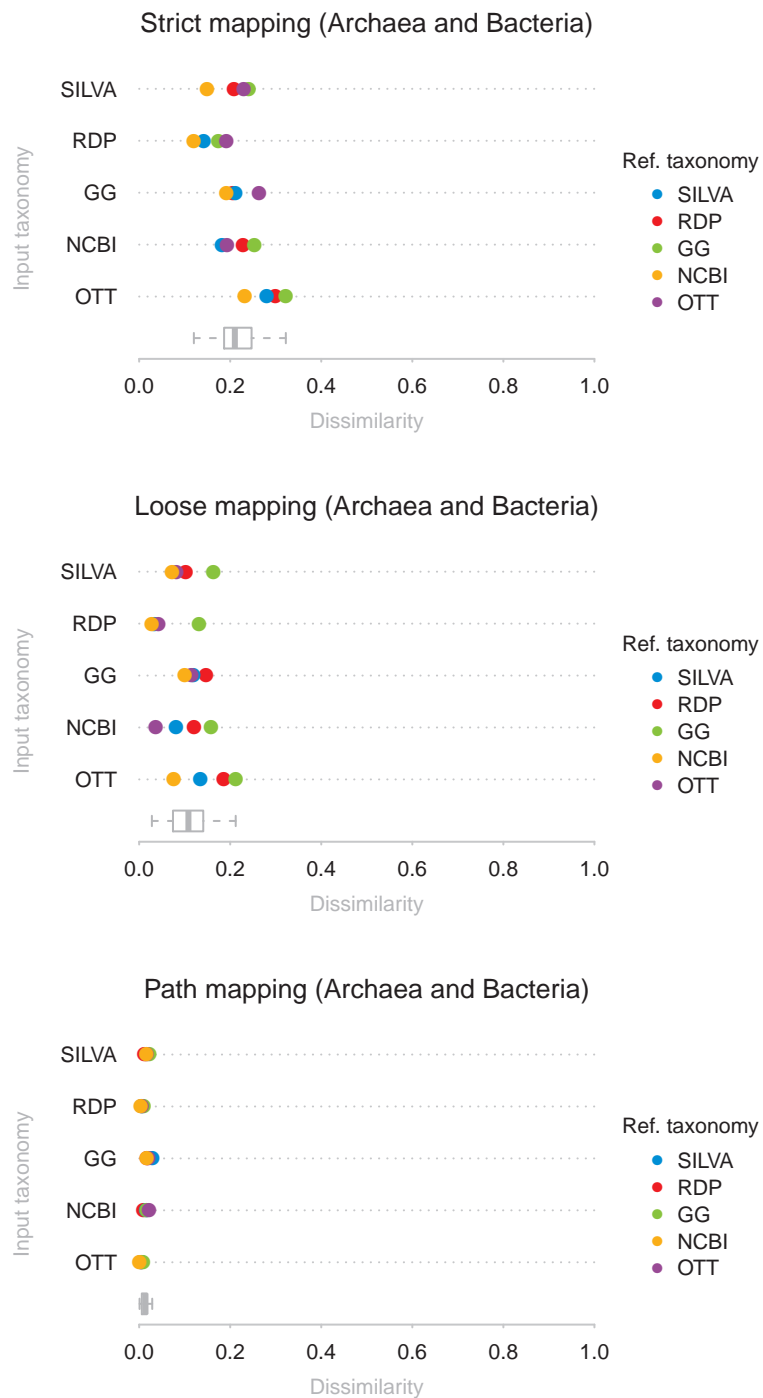


Figure 1: Distributions of mapping dissimilarities.