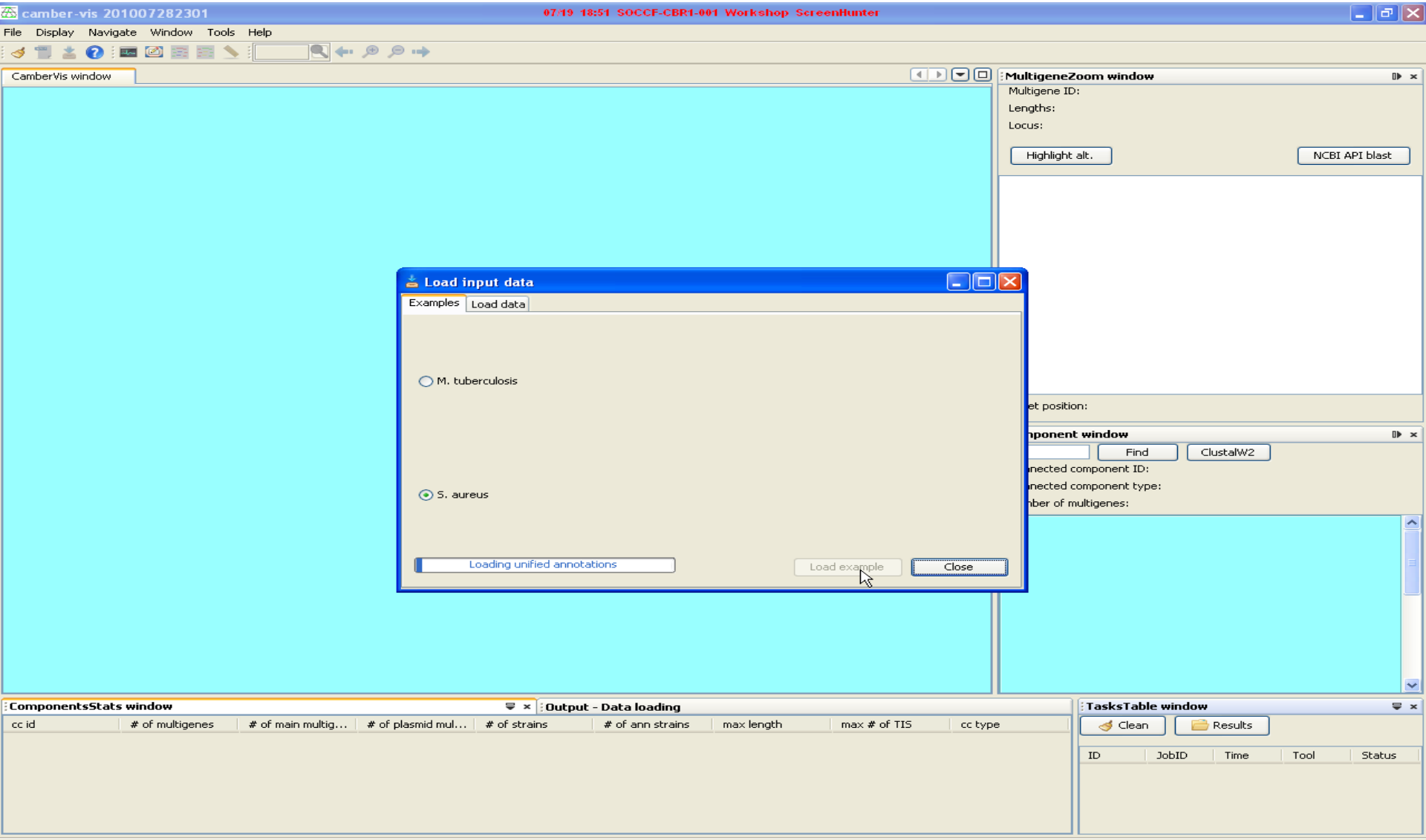
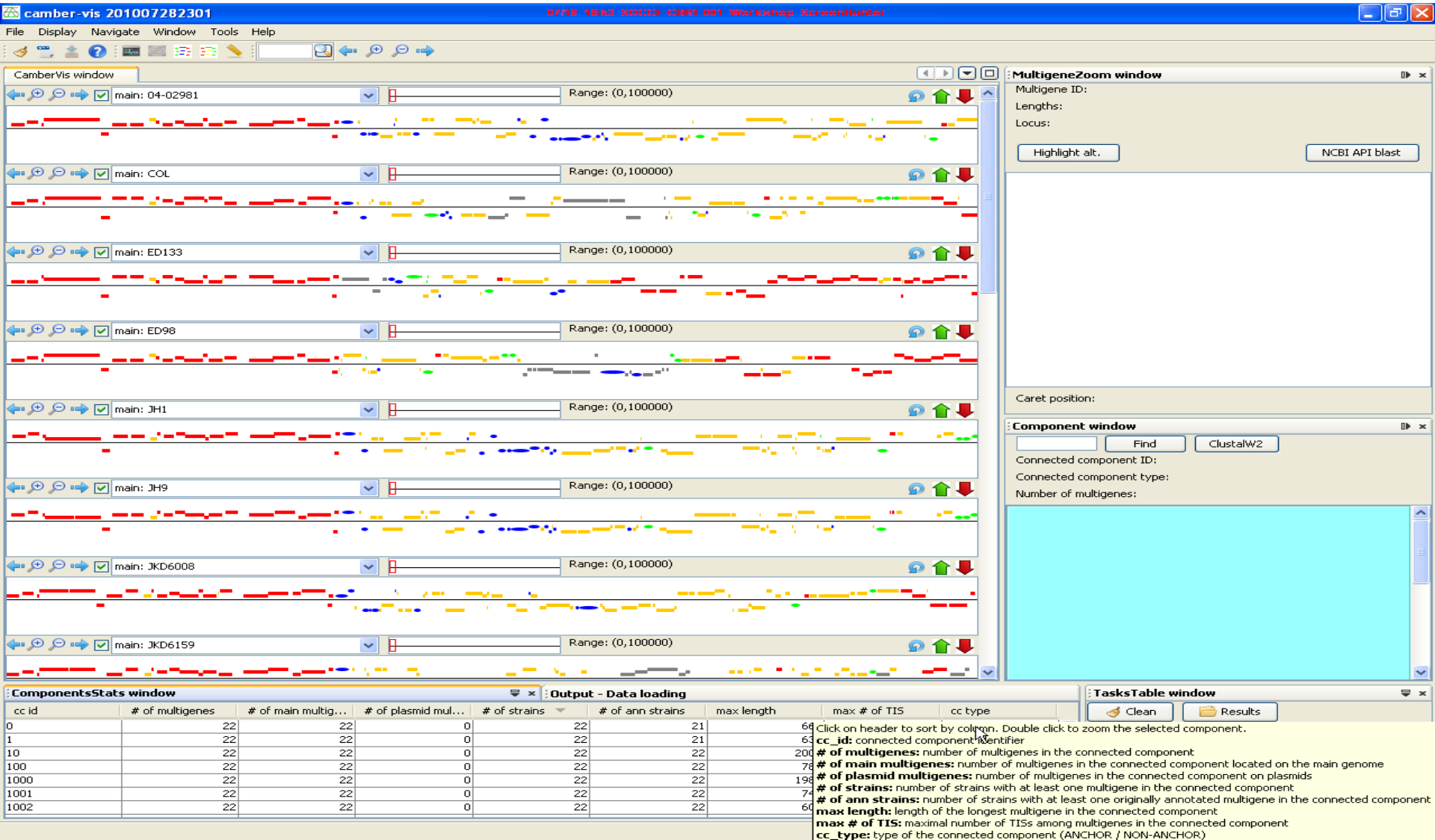
An example of CAMBerVis usage to identify connected components (gene families) that are highly conserved but annotated in very few strains.

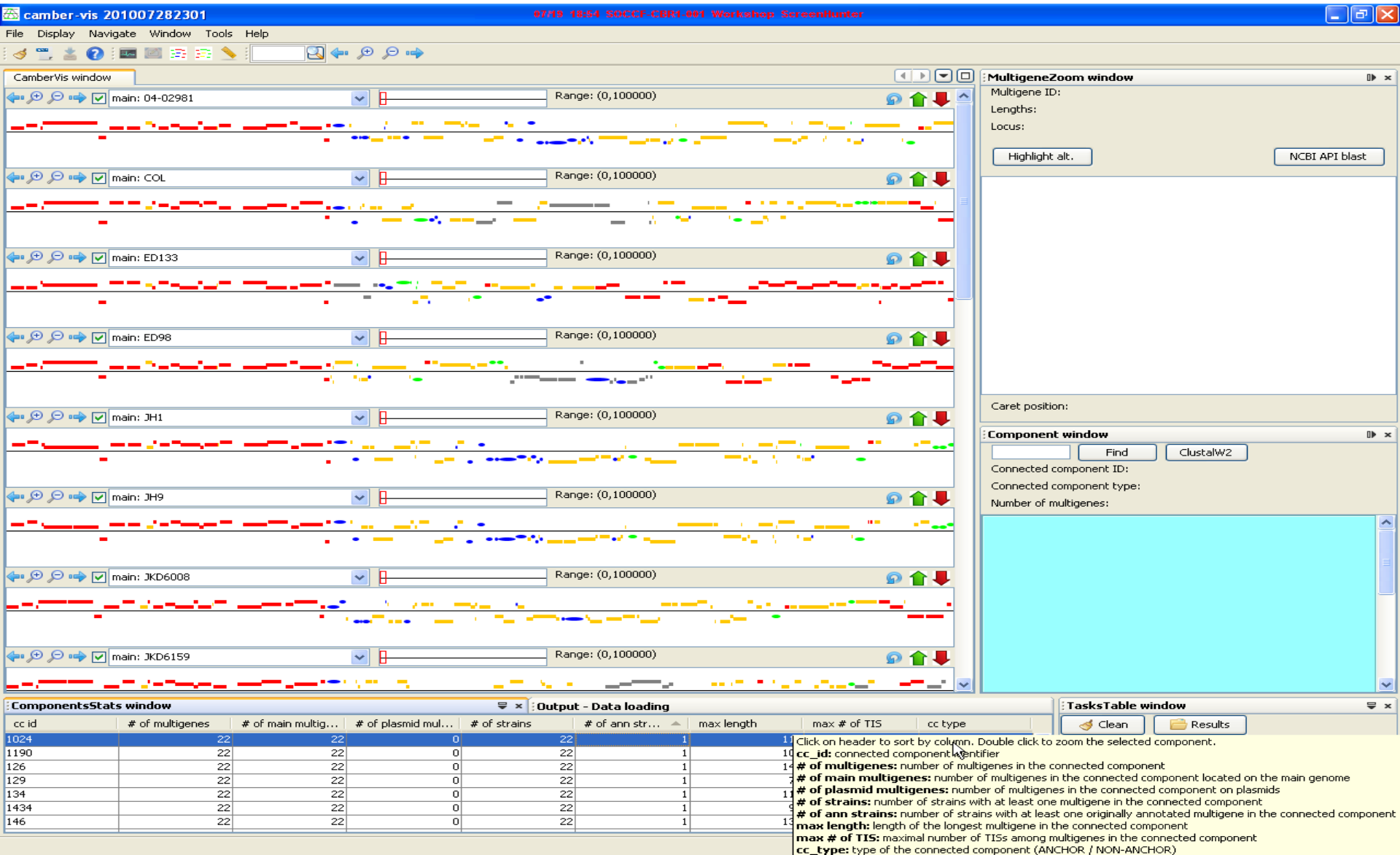
We load example data for *S. aureus*



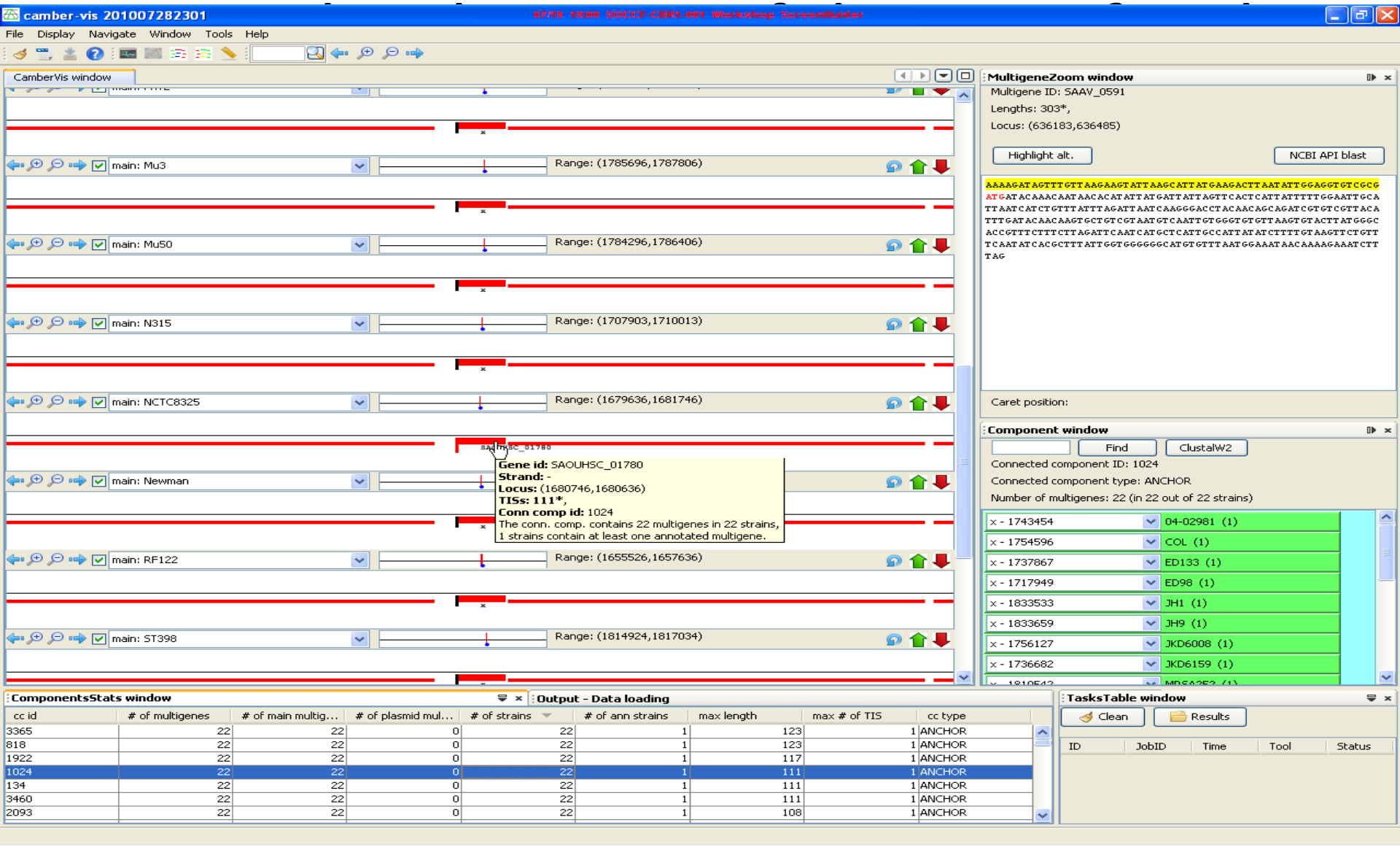
We sort gene families by the column „# of strains” in order to identify highly conserved gene families (with genes present in all strains)



Then we sort the same table by the column „# of ann strains” means the number of strains with at least one annoteted element of ta given gene family. There are examples of gene families with elements predicted in all strains, but annotated in only one strain.

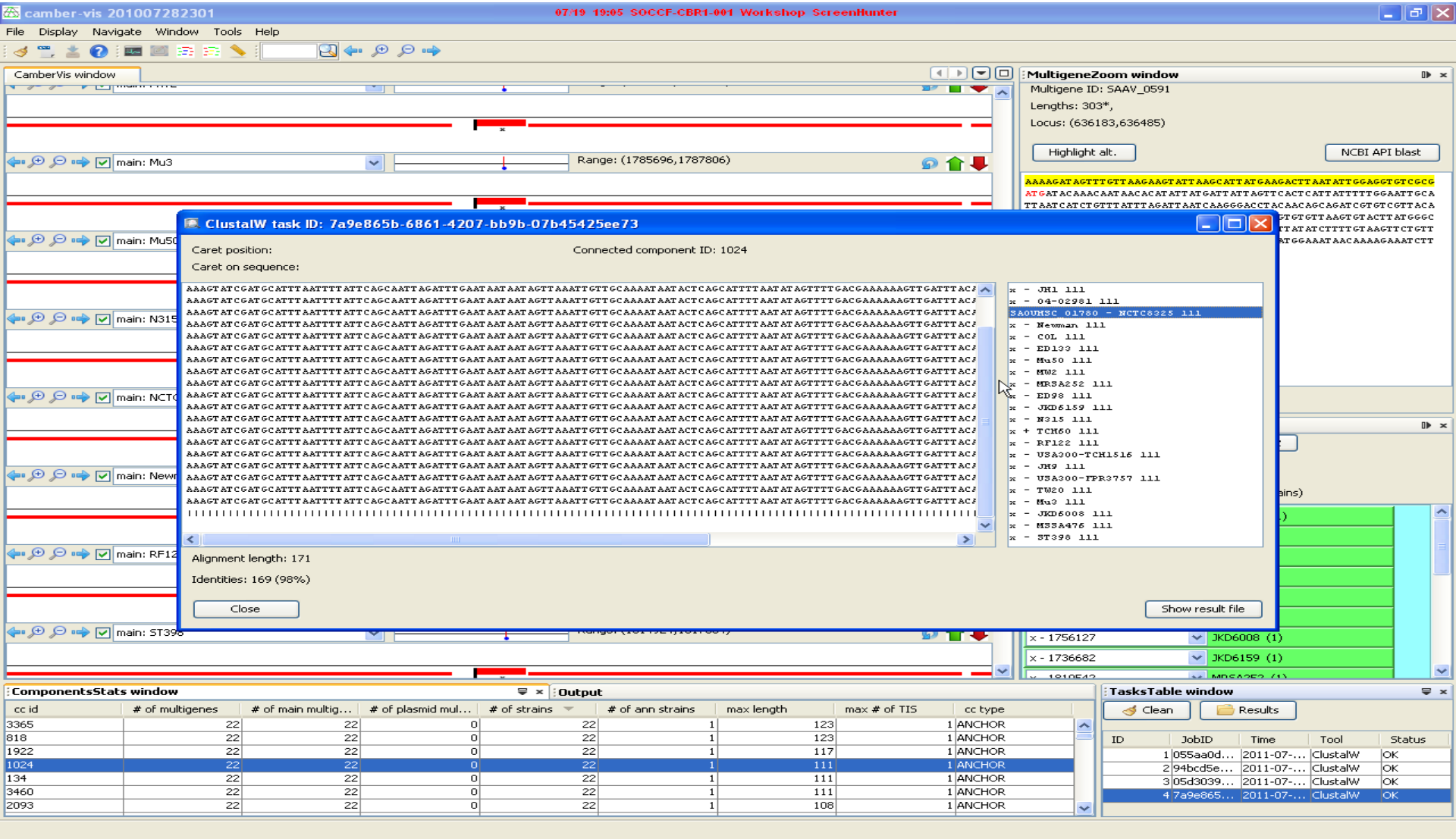


families are usually short. The gene family number 1-24 is a typical

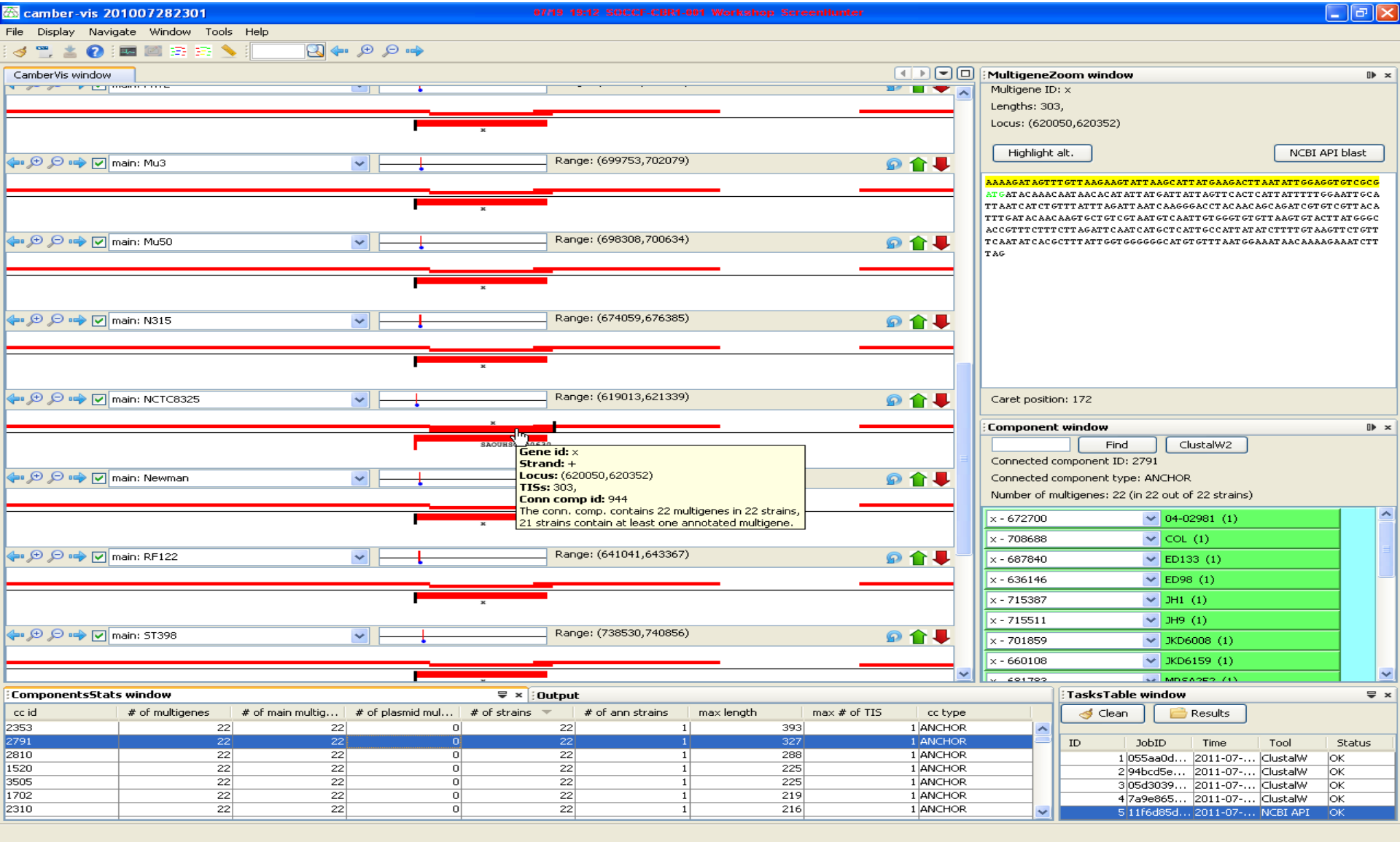


cc id	# of multigenes	# of main multig...	# of plasmid mul...	# of strains	# of ann strains	max length	max # of TIS	cc type
3365	22	22	0	22	1	123	1	ANCHOR
818	22	22	0	22	1	123	1	ANCHOR
1922	22	22	0	22	1	117	1	ANCHOR
1024	22	22	0	22	1	111	1	ANCHOR
134	22	22	0	22	1	111	1	ANCHOR
3460	22	22	0	22	1	111	1	ANCHOR
2093	22	22	0	22	1	108	1	ANCHOR

Multiple alignments of the gene family reveals that the sequences are highly conserved. Promotors of length 60bp. Are exactly identical. This suggest two possibilities: the annotation for NCTC8325 is wrong and was propagated by CAMBer or the annotation is missing in other strains. The possible reason for differences in annotations is the untypical start codon TTG.



The problem also can be found in gene families with pretty long genes like. Connected component number 2791 have elements with length 327bp. We zoom the genome browser on in by double click. Here the spourius annotation might by caused by another highly overlapping gene family: 944, which is annotated in all strains exept NCTC8325. It may be a case of overlapping genes or wrong annotations.



cc id	# of multigenes	# of main multig...	# of plasmid mul...	# of strains	# of ann strains	max length	max # of TIS	cc type
2353	22	22	0	22	1	393	1	ANCHOR
2791	22	22	0	22	1	327	1	ANCHOR
2810	22	22	0	22	1	288	1	ANCHOR
1520	22	22	0	22	1	225	1	ANCHOR
3505	22	22	0	22	1	225	1	ANCHOR
1702	22	22	0	22	1	219	1	ANCHOR
2310	22	22	0	22	1	216	1	ANCHOR

ID	JobID	Time	Tool	Status
1	055aa0d...	2011-07-...	ClustalW	OK
2	94bcd5e...	2011-07-...	ClustalW	OK
3	05d3039...	2011-07-...	ClustalW	OK
4	7a9e865...	2011-07-...	ClustalW	OK
5	11f6d85d...	2011-07-...	NCBI API	OK