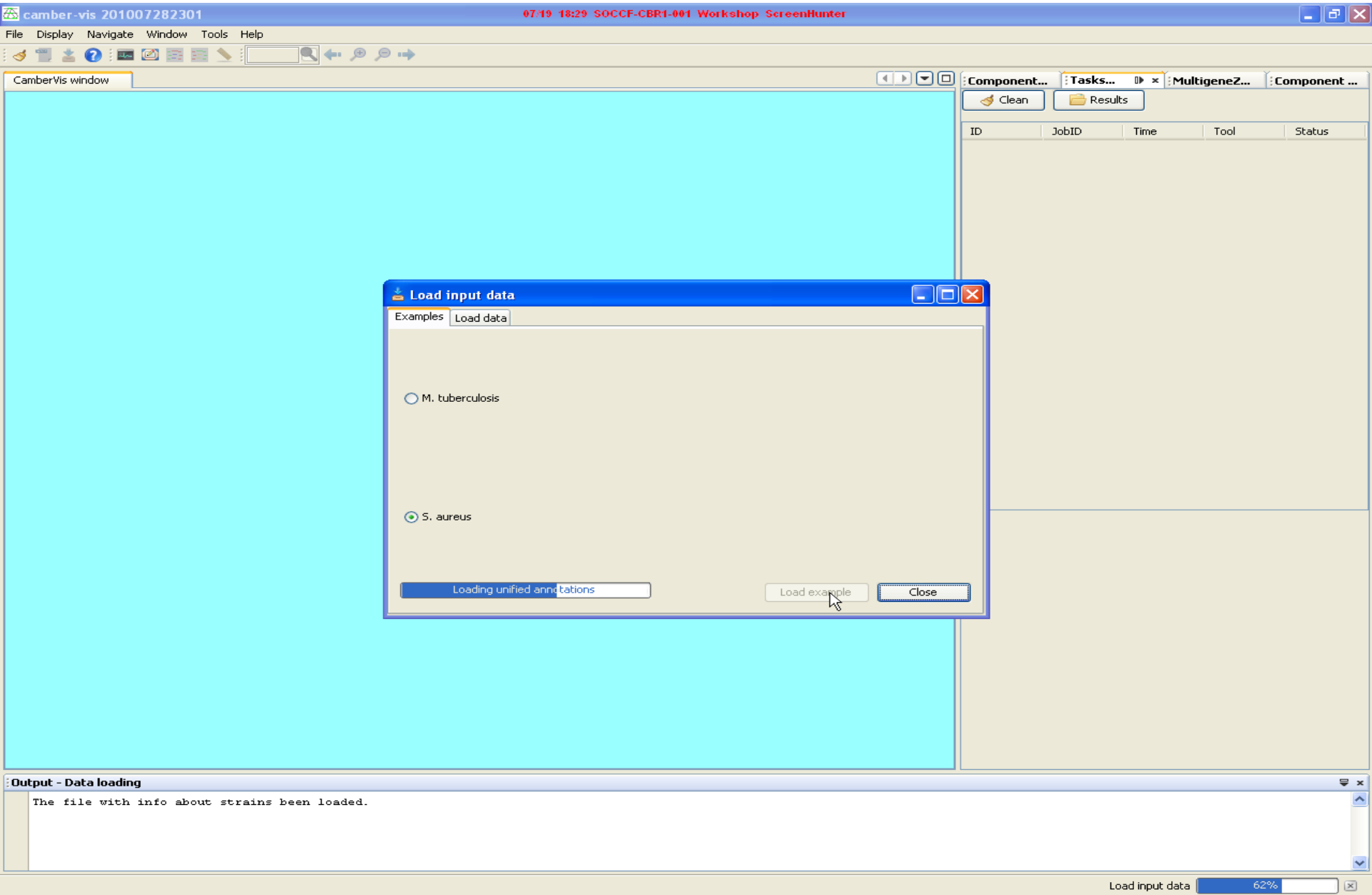
An example of CAMBerVis usage to identify gene families with highly inconsistent annotations of TISs among strains

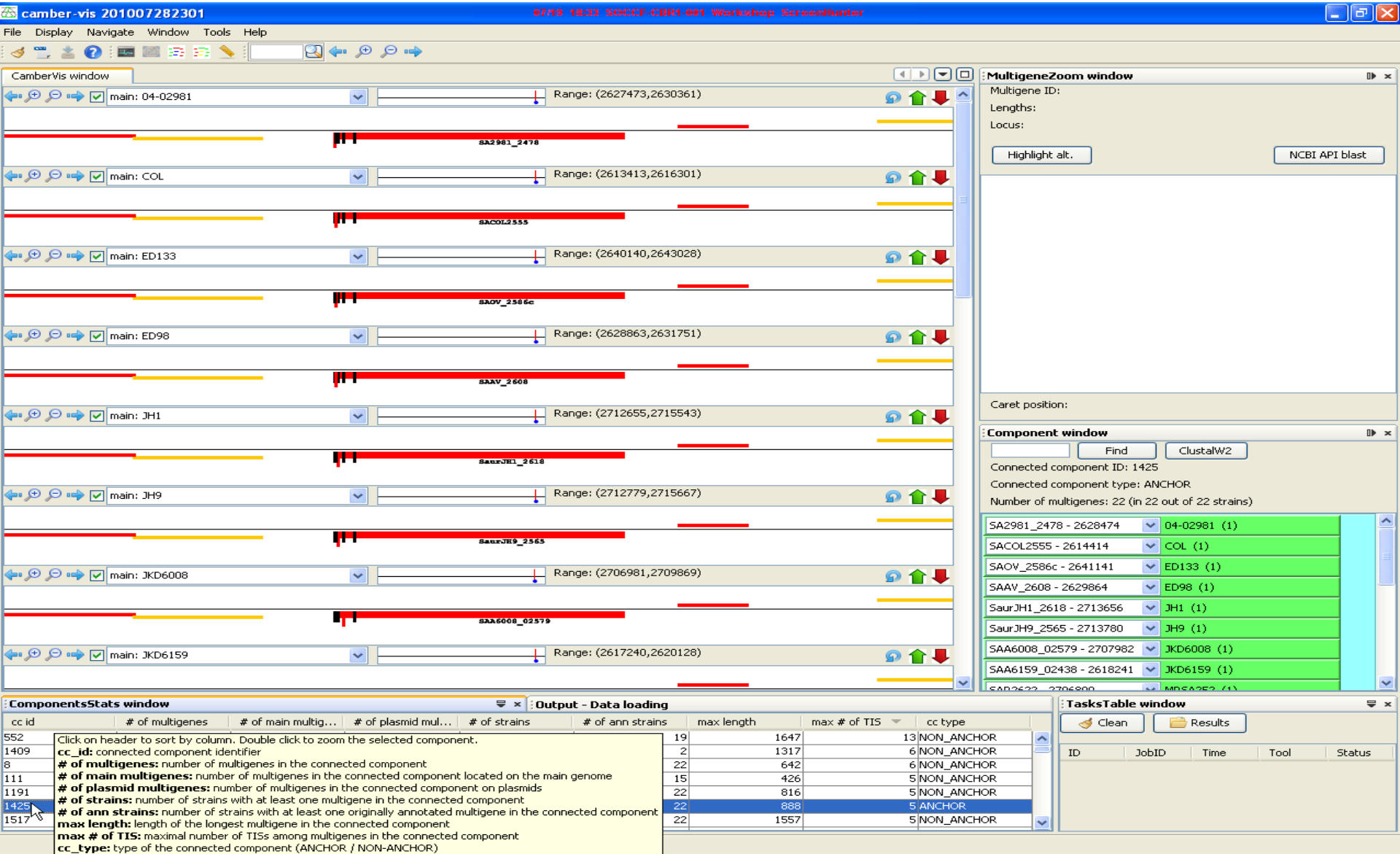
We load example input data for *S. aureus*



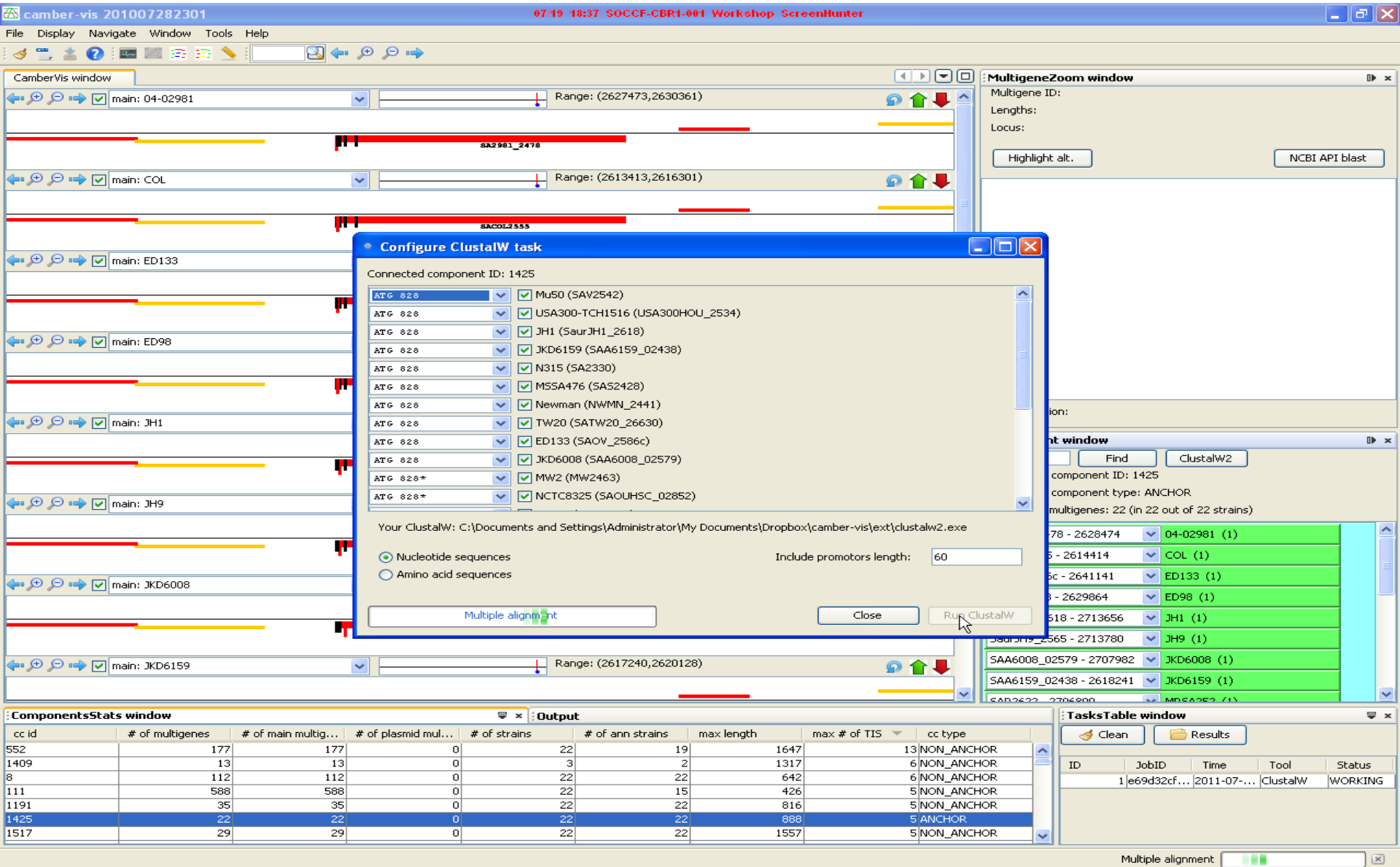
We can personalize the panels localization (OPTIONAL)

cc id	# of multigenes	# of main multig...	# of plasmid mul...	# of strains	# of ann strains	max length	max # of TIS	cc type
0	22			22	21	663	1	ANCHOR
1	22	22	0	22	21	636	1	ANCHOR
10	22	22	0	22	22	2007	1	ANCHOR
100	22	22	0	22	22	789	1	ANCHOR
1000	22	22	0	22	22	1986	2	ANCHOR
1001	22	22	0	22	22	744	1	ANCHOR
1002	22	22	0	22	22	600	1	ANCHOR

We sort the list of connected components (gene families) by the number of TISs. In the example we pick the gene family with ID:1425. It is an ANCHOR and has 5 different TISs annotated. Double click zooms in the genome browser.



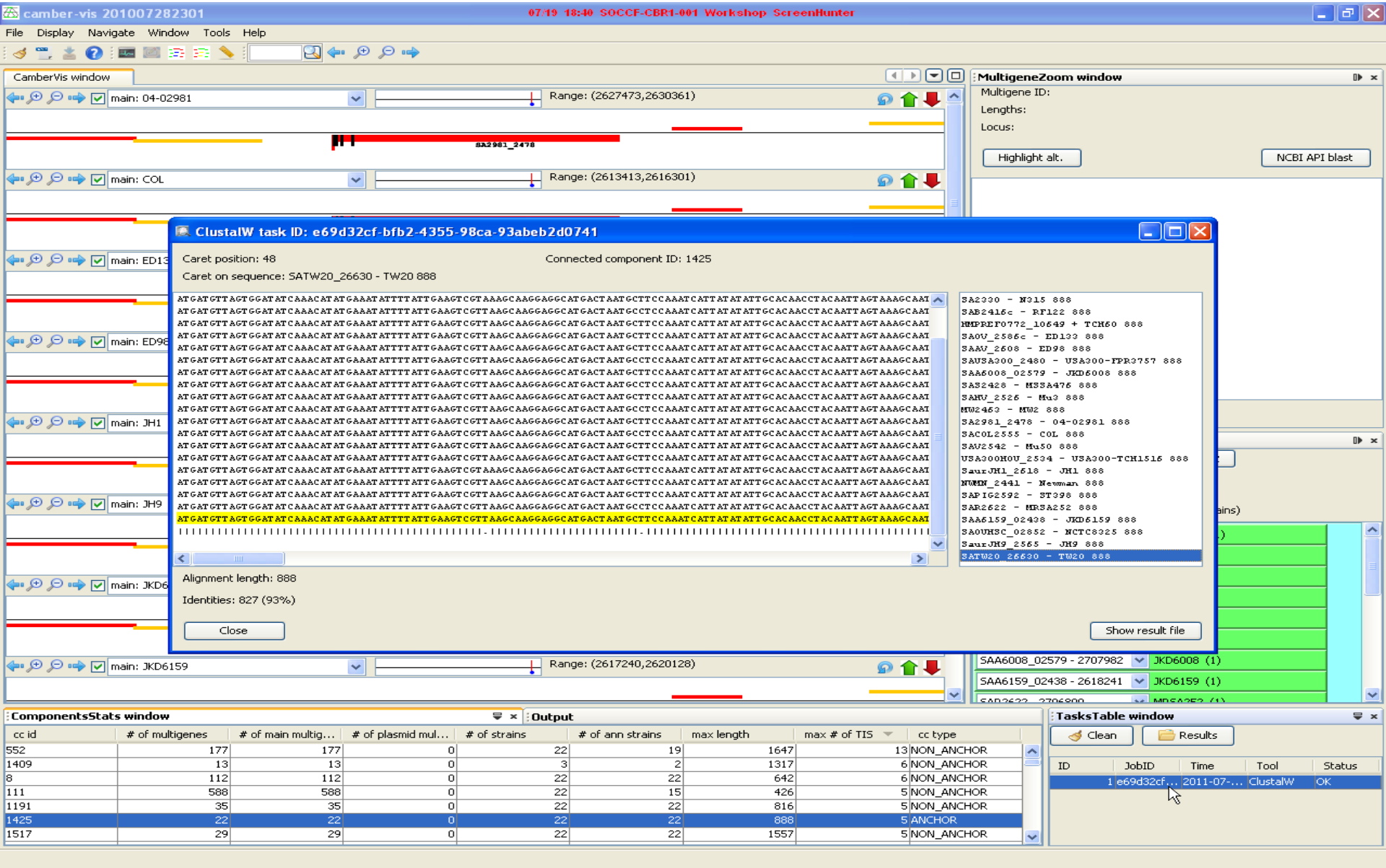
We run on-the-fly analysis by CLUSTALW to check how conserved is the gene family (we included promotors of length 60bp).



cc id	# of multigenes	# of main multig...	# of plasmid mul...	# of strains	# of ann strains	max length	max # of TIS	cc type
552	177	177	0	22	19	1647	13	NON_ANCHOR
1409	13	13	0	3	2	1317	6	NON_ANCHOR
8	112	112	0	22	22	642	6	NON_ANCHOR
111	588	588	0	22	15	426	5	NON_ANCHOR
1191	35	35	0	22	22	816	5	NON_ANCHOR
1425	22	22	0	22	22	888	5	ANCHOR
1517	29	29	0	22	22	1557	5	NON_ANCHOR

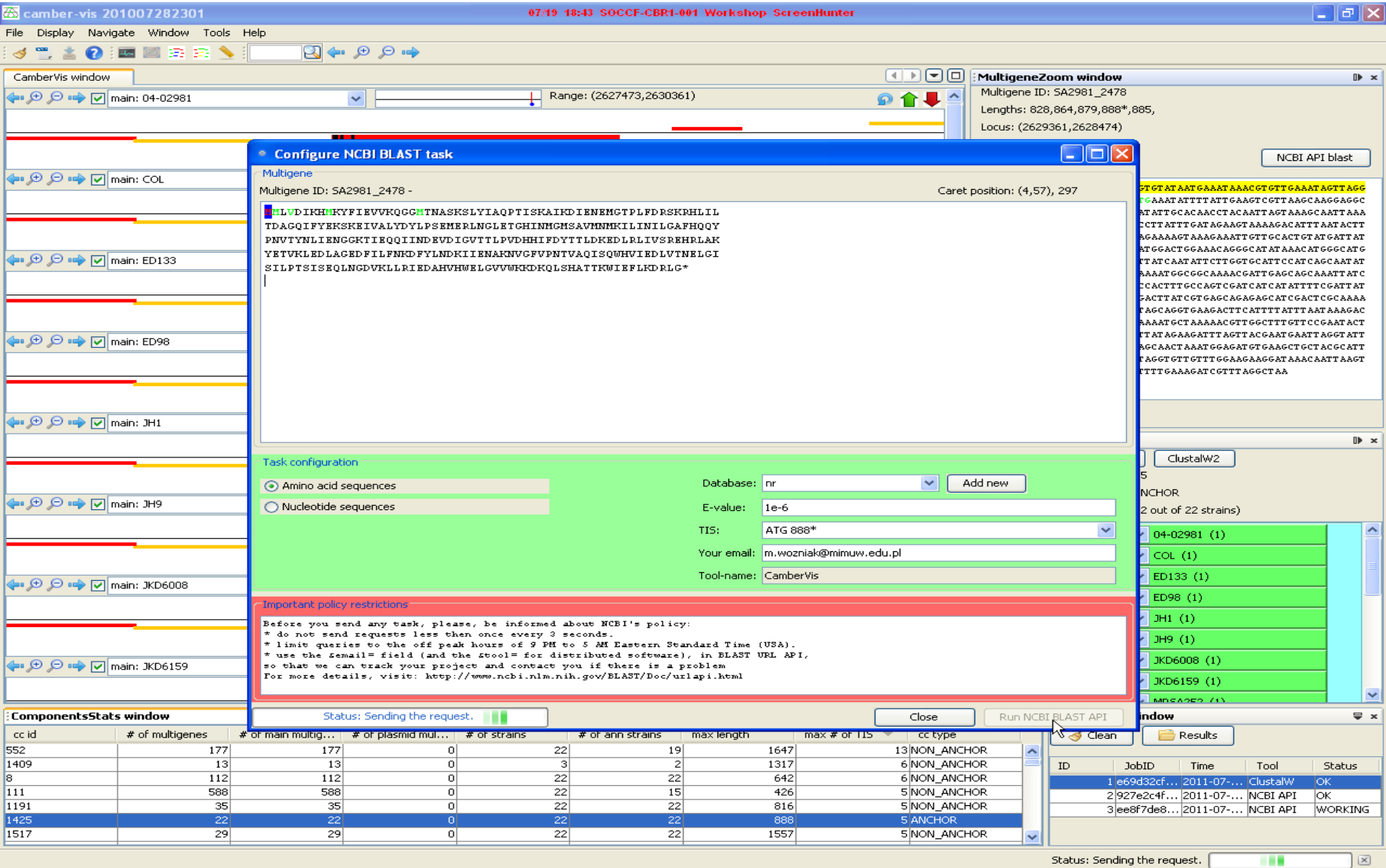
ID	JobID	Time	Tool	Status
1	e69d32cf...	2011-07-...	ClustalW	WORKING

We have computed basic statistcs of the multiple alignment, there are only 2 mutation points. This suggest that the differences in annotations are caused by different annotation tools.

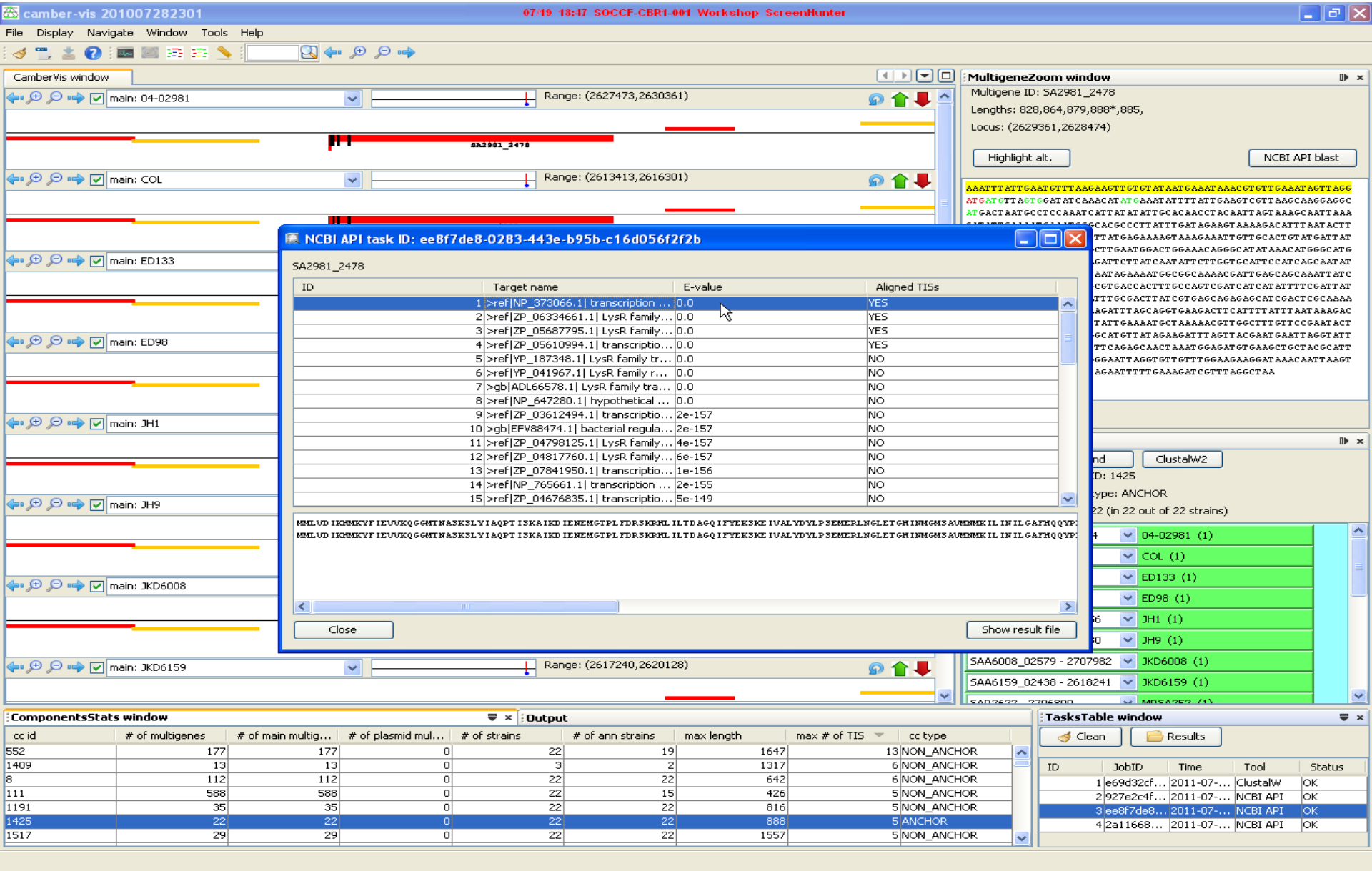


cc id	# of multigenes	# of main multig...	# of plasmid mul...	# of strains	# of ann strains	max length	max # of TIS	cc type
552	177	177	0	22	19	1647	13	NON_ANCHOR
1409	13	13	0	3	2	1317	6	NON_ANCHOR
8	112	112	0	22	22	642	6	NON_ANCHOR
111	588	588	0	22	15	426	5	NON_ANCHOR
1191	35	35	0	22	22	816	5	NON_ANCHOR
1425	22	22	0	22	22	888	5	ANCHOR
1517	29	29	0	22	22	1557	5	NON_ANCHOR

By queries to NCBI database (nr) we can check which TIS is most often annotated.
We run NCBI queries for different TISs.



The results show that the TIS which correspond to the longest gene is most often annotated in other strains. It suggest that this is the correct one.



cc id	# of multigenes	# of main multig...	# of plasmid mul...	# of strains	# of ann strains	max length	max # of TIS	cc type
552	177	177	0	22	19	1647	13	NON_ANCHOR
1409	13	13	0	3	2	1317	6	NON_ANCHOR
8	112	112	0	22	22	642	6	NON_ANCHOR
111	588	588	0	22	15	426	5	NON_ANCHOR
1191	35	35	0	22	22	816	5	NON_ANCHOR
1425	22	22	0	22	22	888	5	ANCHOR
1517	29	29	0	22	22	1557	5	NON_ANCHOR

ID	JobID	Time	Tool	Status
1	e69d32cf...	2011-07-...	ClustalW	OK
2	927e2c4f...	2011-07-...	NCBI API	OK
3	ee8f7de8...	2011-07-...	NCBI API	OK
4	2a11668...	2011-07-...	NCBI API	OK