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*

🖀 camber-vis 201007282301	07/19 18:51 SOCCF-CBR1-001 Workshop ScreenHunter	
File Display Navigate Window Tools	Help	
🤞 📲 🛓 🕜 💷 🖾 📰 为		
CamberVis window		ltigeneZoom window 🕪 🗴
	Load input data	Highlight alt. NCBI API blast
	<ul> <li>M. tuberculosis</li> <li>● S. aureus</li> </ul>	et position:
	Loading unified annotations Load example Close	
ComponentsStats window	₹ × j:Output - Data loading	TasksTable window 🗸 ×
cc id # of multigenes	# of main multig   # of plasmid mul   # of strains   # of ann strains   max length   max # of TIS   cc type	ID JobID Time Tool Status

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

🚟 camber-vis 20	01007282301		07/19 18:53 SOCCF-CBR1-001 Works	hop ScreenHunter			_ 7 🛛
File Display Navig	gate Window Tools Help						
🤞 🖺 🛓 🔞	🔤 🖾 🚘 🛸 i 🔛	🔁 🗢 🗩 🗩					
CamberVis window				💶 🔍 💶 🕅 Multigene	Zoom window		□▶ ×
💠 🗩 🗩 🔽 r	main: 04-02981	▶ 8	Range: (0,100000)	😡 🎓 📕 🔼 Multigene :	D:		
				Lengths:			
				Locus:			
				Highligh	it alt.		NCBI API blast
🦛 🗩 🗩 🛶 🔽 r	main: COL	✓ I	Range: (0,100000)			_	
	main: ED133	× 1	Range: (0,100000)				
	nain: FD98		Range: (0.100000)				
	Hain. 2090			<u>≥</u>			
	a a tha sa ta ta ta ta ta						
-							
			Rappe: (0, 100000)	Caret posi	ion:		
	main: JHI				ot window		
	and the second second		•	•• •• •• •• •• •• •• •• •• •• •• •	Eind	ClustalW/2	
		· · · · · · · · ·		Connected	component ID:	Clastaniz	
				Connected	component type:		
🗢 🕫 🗢 🔽 r	main: JH9		Range: (0,100000)	🔊 👚 🖊 🛛 Number of	multigenes:		
	and the second second second	•••••••	•				^
-		· · · - · ·					
							а на
🗢 🗩 📿 🗣 🔽 🗗	main: JKD6008		Range: (0,100000)	🖸 🖉 🗣			
	للمستار المستحد والمراجع	<u> </u>		· · · · · · · · · · · · · · · · · · ·			
		1 and 1 and 1 and					
n 👽 🗢 🔍 🕂	main: JKD6159	✓ 8	Range: (0,100000)	🗩 🎓 😡			
				····			
ComponentsEtat	s window		R w Output, Data landian		TacksTable winds		
componentsstat	# of multigenes # of ma	in multin # of plasmid mul	t of strains # of ann strains may lengt	b may # of TIS co type		- Peculto	• *
0	22	22 0	22 21	64 Click on header to sort by colimn. Double click t	zoom the selected com	ponent.	
1	22	22 0	22 21	6 cc_id: connected component identifier			
10	22	22 0	22 22	200 <b># of multigenes:</b> number of multigenes in the 78 <b># of main multigenes:</b> number of multigenes	connected component in the connected compo	onent located on the r	nain genome
1000	22	22 0	22 22	198 # of plasmid multigenes: number of multige	nes in the connected co	mponent on plasmids	-
1001	22	22 0	22 22	# of ann strains: number of strains with at least of # of ann strains: number of strains with at least	e multigene in the conn ast one originally annota	ecceu component ted multigene in the c	connected component
1002	22	22 0	22 22	<sup>60</sup> max length: length of the longest multigene in	the connected component	ent sted component	
				I HOA # UL 113: HIGAIHGH HUHDER OF 1135 among	maiagenes in the conne	CLEA COMPONENC	

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

🚟 camber-vis 201	007282301	_	07/19 19:00	SOCCF-CBR1-001 Workshop	ScreenHunter		-	
File Display Naviga	ite Window Tools	Help						
i 🤞 🚆 🛓 🔞 🗉	u 🔟 🕫 🕾 📏	, 🕂 🥠 🔁	) 🌳					
CamberVis window						MultigeneZ	oom window	D ×
			•	· ·	a 🖉 🖬 🗮 🖉	Multigene ID	: SAAV_0591	
						Lengths: 303	3*,	
						Locus: (6361	.83,636485)	
	aio: Mu3	**	Range; (	1785696,1787806)		Highlight	alt.	NCBI API blast
			•		<u>≥</u> ∕ ∎ →			
						ATGATACAAAC	AAT AACACAT ATT AT GATT ATT AGTT CA	TCATTATTTTTGGAATTGCA
			×			TTAATCATCTG	TTT ATTT AGATT AAT CAAGGGACCT ACA	CAGCAGATCGTGTCGTTACA
						ACCGTTTCTTT	AAGT GCT GT C GT AAT GT C AATT GT GG GG ( CTT AG ATT C AAT C AT GCT C ATT GCC ATT 2	TGTTAAGTGTACTTATGGGC ATATCTTTTGTAAGTTCTGTT
🔶 🗩 🗩 🗩 🖛	ain: Mu50	✓	Range: (	1784296,1786406)	🖸 🌓	TCAATATCACG	CTTT ATT GGT GGG GG GG AT GT GTTT AAT (	GAAATAACAAAAGAAATCTT
						TAG		
			×					
			Dep est /	1707000 1710010				
	ain: N315	× –	Ralige: (	1707903,1710013)	🖊 🚹 😡			
			, , , , , , , , , , , , , , , , , , ,					
🖕 🕀 💬 📫 🔽 ma	ain: NCTC8325		Range: (	1679636,1681746)	o 🛧 📕	Caret positio	ID:	
			•		27 🖬 🔶			
						Component	window	D × €0
			SA 175C_01780				Find ClustalW2	]
			Gene id: SAOUHSO	_01780		Connected c	omponent ID: 1024	
🔶 🗩 🗩 🖛 🔁 🖛	ain: Newman	►	Strand: -	680636)	s 🔁 📕	Connected c	omponent type: ANCHOR	
			TI5s: 111*,			Number of m	ultigenes: 22 (in 22 out of 22 strain:	;)
			Conn comp id: 10	24 phains 22 multigenes in 22 strains		x - 1743454	▼ 04-02981 (1)	<u>^</u>
			1 strains contain at	least one annotated multigene.		1754506		
			Pappe: (	1655526 1657636)		x - 1754596	COL (1)	
	an: RF122	× –	Kanger	1000020,10070000	💶 🗣 🖬 😡	× - 1737867	✓ ED133 (1)	
						× - 1717949	🐱 ED98 (1)	
			x			× - 1833533	✓ JH1 (1)	
						x - 1833659	→ JH9 (1)	
🖕 🗩 🗩 🛶 🔽 ma	ain: ST398	× –	Range: (	1814924,1817034)	o 🏫 📕	× 1756127		
			•			x - 1756127	→ SKD6008 (1)	
						× - 1736682	✓ JKD6159 (1)	v
ComponentsEtats	window			sta laadiaa		1010E42	TasksTable window	
componentsstats	# of multigenes	# of main multin # of ol			may # of TIS			¥ ×
3365	* or mulagenes	22 * or main mulog   * or pi			123 1 ANCHOR			
818	22	22	0 22	1	123 1 ANCHOR		ID JobID Time	Tool Status
1922	22	22	0 22	1	117 1 ANCHOR		1.1.10	
1024	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.

🛱 camber-vis 20100728	32301	07/19 19:05 SOCCF-CBR1-	001 Workshop ScreenHunter		<b>.</b> . <b>. .</b>
File Display Navigate Win	dow Tools Help				
i 🤞 🖺 🛓 🔞 i 🔤 🔟	🚌 🚌 📏 i 🔤 🛃 🦛 🔎	, 💬 🛶			
CamberVis window				MultigeneZoom window	<b>□</b> ×
	-	L	× • •	Multigene ID: SAAV_0591	
				Lengths: 303*,	
		x		Locus: (636183,636485)	
<del>4•</del> 🗩 🗩 🛶 🔽 main: Mu3	~	Range: (1785696,178780	06) 💿 👚 📕		NCBI API blast
				AAAAGAT AGTTT GTT AAGAAGT ATT AAGC ATT AT	AAGACTTAATATTGGAGGTGTCGCG
				AT GAT AC AAAC AAT AAC AC AT ATT AT GATT ATT AG TT 3 STC STCTCTTT STTT 3 C STT 3 STC 3 SCCC3 C	TTCACTCATTATTTTTGGAATTGCA
	ClustalW task ID: 7a9a865	b-6861-4207-bb9b-07b45425ee73			FTGTGTTAAGTGTACTTATGGGC
🖕 🙂 🔎 🛶 🔽 main: Mu50	task ib. 747000	5-0001-1201-8676-018434236613			TT AT AT CTTTT GT AAGTT CT GTT AT GG AAAT AAC AAAAG AAAT CTT
	Caret position:	Connected component ID:	1024		
	Caret on sequence:				
	AAAGTATCGATGCATTTAATTTTATTCAGC	AATT AGATTT GAAT AAT AAT AGTT AAATT GTT G	GCATTTTAATATAGTTTTGACGAAAAAGTTGATTTAC	с - JHI III	
	AAAGTATCGATGCATTTAATTTTATTCAGC AAAGTATCGATGCATTTAATTTTATTCAGC	A ATT AG ATTT GAAT AAT AAT AGTT AAATT GTT G	SCATTTTAATATAGTTTTGACGAAAAAAGTTGATTTAC	: - 04-02981 111 3A0UH3C 01780 - NCTC8325 111	
🗢 🗩 🎾 🖙 🔽 main: N31:	AAAGTATCGATGCATTTAATTTTATTCAGC	AATT AGATTT GAAT AAT AAT AGTT AAATT GTT G	GCATTTTAATATAGTTTTGACGAAAAAGTTGATTTAC	: - Newman 111	
	AAAGTATCGATGCATTTAATTTATTCAGC	AATT AGATTT GAATAATAATAGTT AAATT GTT GCAAAATAATACTCA AATT AGATTT GAATAATAATAGTT AAATT GTT GCAAAATAATACTCA	GCATTTTAATATATAGTTTTGACGAAAAAAGTTGATTTACS	- COL 111 - ED133 111	
	AAAGTATCGATGCATTTAATTTTATTCAGC	AATT AGATTT GAAT AAT AAT AGTT AAATT GTT G	CATTTTAATAT AGTTTTGACGAAAAAAGTTGATTTAC	s - Mu50 111	
	AAAGTATCGATGCATTTAATTTATTCAGC	AATT AGATTT GAAT AAT AAT AGTT AAATT GTT G	GCATTTTAATATAGTTTTGACGAAAAAAGTTGATTTAC	- MW2 111 : - MR3A252 111	
💠 🗩 🔎 🛶 🔽 main: NCT	AAAGTATCGATGCATTTAATTTTATTCAGC	AATT AGATTT GAAT AAT AAT AGTT AAATT GTT G	SCATTTTAATATAGTTTTGACGAAAAAAGTTGATTTAC	- ED98 111	
	AAAGTATCGATGCATTTAATTTTATTCAGC	AATT AGATTT GAAT AAT AAT AGTT AAATT GTT G	SCATTTTAATATAGTTTTGACGAAAAAAGTTGATTTAC	- N315 111	
	AAAGTATCGATGCATTTAATTTTATTCAGC AAAGTATCGATGCATTTAATTTTATTCAGC	AATT AGATTT GAAT AAT AAT AGTT AAATT GTT G	SCATTTTAATATAGTTTTGACGAAAAAAGTTGATTTAC / x	: + TCH50 111	
	AAAGTATCGATGCATTTAATTTTATTCAGC	AATT AGATTT GAAT AAT AAT AGTT AAATT GTT G	GCATTTTAATATAGTTTTGACGAAAAAAGTTGATTTAC	- USA300-TCH1516 111	
👍 🗩 🔎 🛶 🔽 main: New	AAAGTATCGATGCATTTAATTTTATTCAGC AAAGTATCGATGCATTTAATTTTATTCAGC	AATT AG ATTT GAAT AAT AAT AGTT AAATT GTT G	GCATTTTAATATAGTTTTGACGAAAAAAGTTGATTTAC / x GCATTTTAATATAGTTTTGACGAAAAAAGTTGATTTAC / x	: - JH9 111 : - USA300-FPR3757 111	
	AAAGTATCGATGCATTTAATTTTATTCAGC	AATT AGATTT GAAT AAT AAT AGTT AAATT GTT G	GCATTTTAATATAGTTTTGACGAAAAAGTTGATTTAC	- TW20 111	ains)
				: - Mu3 111 : - JKD5008 111	
			× ×	: - M33A476 111	· · · · · · · · · · · · · · · · · · ·
	<u>&lt;</u>			: - 31398 111	
	Alignment length: 171				
	Identities: 169 (98%)				
	Close			Show result file	
				Show result his	
🗢 🗩 🎔 main: ST39	/o	• Kangor (1011)2 ()101100	S 🖬 🖬 🖊	×-1756127 V JKD6008	(1)
				× - 1736682 🛛 😽 JKD6159	(1)
				1010E42	- M 🛛 💆
ComponentsStats window		₹ × Output		TasksTable window	<b>₹</b> ×
cc id # of mu 3365	Itigenes # of main multig # o	r plasmid mul # of strains # # of ann strains	max length max # of TIS cc type	Clean E Res	ults
818	22 22	0 22 1	123 1 ANCHOR	ID JobID Ti	me Tool Status
1922	22 22	0 22 1	117 1 ANCHOR	1 055aa0d 201	1-07 ClustalW OK
1024	22 22 22 22 22		111 1 ANCHOR 111 1 ANCHOR	2 94bcd5e 201	1-07 ClustalW OK
3460	22 22	0 22 1	111 1 ANCHOR	4 7a9e865 201	1-07 ClustalW OK
2093	22 22	0 22 1	108 1 ANCHOR	✓	

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.

🚟 camber-vis 20	01007282301			07/19 19:12 SOCCF-CBR1-	001 Workshop ScreenHunte	r			- 7 🛛
File Display Navi	gate Window Tools	Help							
🤞 🚆 🛓 🕜 🛛	🔤 🖾 🕾 🕾 🔦	•• 🗵 🗣	,⊕ 🖵 🕂						
CamberVis window							MultigeneZ	oom window	
				- · · · · ·		27 🖬 🔫 🔼	Multigene ID	: ×	
							Lengths: 30	3,	
				×			Locus: (6200	)50,620352)	
							<b></b>		
🖕 🗩 🗩 🛶 🔽 🗤	main: Mu3		✓	Range: (699753,702079)	l i i i i i i i i i i i i i i i i i i i	o 🔒 📕	Highlight	alt.	NCBI API blast
			•				AAAAGATAGTI	T GTT AAGAAGT ATT AAGC ATT AT GAAG	ACTTAATATTGGAGGTGTCGCG
							AT GAT AC AAAO	AAT AAC AC AT ATT AT GATT ATT AGTT C	ACT CATT ATTTTT GGA ATT GCA
				×			TTAATCATCTG	TTT ATTT AGATT AAT CAAGGGACCT AC AAGT GCT GT CGT AAT GT CAATT GT GGG	AACAGCAGATCGTGTCGTTACA /TGTGTTAAGTGTACTTATGGGC
				Banan (600000 700604)			ACCGTTTCTT	CTTAGATTCAATCATGCTCATTGCCAT	TATATCTTTTGTAAGTTCTGTT
	main: Mu50		×	Range: (698308,700634,		🔄 😰 🎦 🖳 📔	TCAATATCACG	CTTTATTGGTGGGGGGCATGTGTTTAA	T GGAAAT AACAAAAGAAAT CTT
				×					
🖕 🗩 🗩 🛶 🔽 🗤	main: N315		✓	Range: (674059,676385)	l i i i i i i i i i i i i i i i i i i i	o 🔒 📕			
			•						
				×					
				Barney (610010 (01000)					
	main: NCTC8325		× .	Range: (619013,621339,		🔊 👚 🖊 📋	Caret positio	n: 172	
				×			Component	t window	D ×
				SAOUHS				Find ClustalW2	
				Gene Id: × Strand: +			Connected o	omponent ID: 2791	_
🖕 🕀 🗩 🔽 🗤	main: Newman		✓	Locus: (620050,620352)		o 🔒 📕	Connected o	omponent type: ANCHOR	
				TISs: 303, Conn comn id: 944			Number of m	ultigenes: 22 (in 22 out of 22 stra	ins)
				The conn. comp. contains 22 mu	Itigenes in 22 strains,		× - 672700	v 04-02981 /1	
				<ul> <li>21 strains contain at least one a</li> </ul>	nnotated multigene.		x = 072700		——————————————————————————————————————
				Pappe: (641041-643367)			x - 708688	✓ COL (1)	
	main: RF122		×	Kange: (041041,043307,		🖸 🗣 🔁 🔍	× - 687840	✓ ED133 (1)	
<u> </u>							× - 636146	🛩 ED98 (1)	
				×			× - 715387	✓ JH1 (1)	
							× - 715511	✓ JH9 (1)	
<rr>Image: Control (Image: Control (Image:</rr>	main: ST398		✓	Range: (738530,740856)		🖸 🏚 😡	x - 701859	✓ 1KD6008 (1)	
							× 660109	1KD61E0 (1)	
						~	x - 860108	✓ JKD0159 (1)	v
ComponentsStat	s window		•				201705	TasksTable window	
cc id	# of multigenes	# of main multig	# of plasmid mul	. # of strains	max length max # of	TIS cc type			7
2353	22	22		0 22 1	393	1 ANCHOR			
2791	22	22		0 22 1	327	1 ANCHOR		ID JobID Time	Tool Status
2810	22	22		0 22 1	288	1 ANCHOR		1 055aa0d 2011-0	7 ClustalW OK
3505	22	22		0 22 1	225	1 ANCHOR		2 94bcd5e 2011-0	7 ClustalW OK
1702	22	22		0 22 1	219	1 ANCHOR		4 7a9e865 2011-0	7 ClustalW OK
2310	22	22		0 22 1	216	1 ANCHOR	~	5 11f6d85d 2011-0	7 NCBI API OK