

GenoFig: A user-friendly application for the vizualisation and comparison of genomic regions

Maxime Branger, Sébastien Leclercq

▶ To cite this version:

Maxime Branger, Sébastien Leclercq. GenoFig: A user-friendly application for the vizualisation and comparison of genomic regions. Journées Ouvertes en Biologie, Informatique et Mathématiques - JOBIM 2023, Jun 2023, Tours, France. , 2023, JOBIM 2023 Proceedings. hal-04170376

HAL Id: hal-04170376 https://hal.inrae.fr/hal-04170376v1

Submitted on 25 Jul 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



GenoFig: A user-friendly application for the visualization and comparison of genomic regions

INRAQ

Maxime Branger and <u>Sébastien Leclercq</u>



UMR 1282 ISP, INRAE – université de Tours, Nouzilly, France

Visual comparison of genomic regions from related species or strains is crucial to understand the molecular evolutionary history of species, particularly in complex mobile regions [1]. GenoFig is a stand-alone graphical application dedicated to the generation of high quality figures for comparative genomics. It is intended to be as easy to use as possible and flexible enough to adapt to a variety of needs. The application works on Linux, MacOS and Windows platforms.

How to simply produce a figure with GenoFig?

1. Add sequences

Image width: 7500			300		ba	ckgrou	nd color 🛛	2																
Draw figure in: 📔 Figure_Poster.png						470 margins 20 enhance graph O																		
										ennan	ce graph		0											
							200																	
equences Features	Homol	ogies Leg	gends Decora	tion																				
sequence name	infos	active	position	min	max	reverse	height	space below	line	width	color	seq. label	type	position	offset	color	size	BI	feat. label	type	positior	n color	size	rot.
General		I	left				100	250		10			locus+description	left	60		38			product	тор		30	0
K D86934	Ō		-	4500	max		-	-		20	-		organism	-	-	-	-			-	-	-	-	-
• D00334	õ		best blast	1	max		-	-		-	-		locus+size	bottom/right	-	-	-	□ 🗹		-	-	-	-	-
			best blast	4267	24067		-	290		-	-		-	top/left	-	-	-			-	-	-	-	-
X03216	Ō		-	1	max		-	0		-	-		description	bottom	-		-			-	-	-	-	-
X03216NZ_UHDL01000001	Q Q	\checkmark						550		-	-		-	top/right	130	-	·			-	-	-	-	-
X03216	-		x = 3200	2192	5 40925	\checkmark	-	550						coprigite										

- (multi-)GenBank/FASTA accepted

- possibility to set up multiple parameters sequence by sequence or globally (General)

- a 'best blast' option for sequence positionning, for optimal alignment of homologous regions.

- various information about sequences can be displayed as sequence labels

2. Add features

Þ	selected	in legend	type	filter	in field	strand	shape	height ratio	fill	color line	width	colo	r hatching	label	type	position	BIC	olor s	size
General						none	arrow	1			1		none						
🕻 🖷 other CDS			CDS		any	-	signal	0.7	<	-	-	-	-		gene	Bottom	- 0	-	
🕻 🖷 hypothetical proteins			CDS	hypothetical protein	product	lead up	-	-	\checkmark		-	-	-		-	Bottom	- 0	-	
🕻 🖷 Other transposases			CDS	transposase tnp[ABC]	any	-	-	-			2	-	-		product	-		-	
🕻 🖷 IS431			CDS	IS431	product	-	-	-			2		diagonal		product	-		-	
🕻 🖷 SCCmec			misc_feature	SCCmec	note	-	range	4.0		-	3	-	-	\checkmark	note	Тор	2	4	10
🕻 🖷 Mobile elements			mobile_element	:	any	-	range	1.5			4		-	\checkmark	mobile_element_type	Тор	2	-	
🕻 🖷 D-xylose metabolism	\checkmark		CDS	xvloselRS00850	anv	-	-	-			-	-	-		-	-	- 0	-	
🕻 🖷 sulfide metabolism			CDS	sulfite sulfide DsrE	product	-	-	-			-	-	-	\checkmark	gene	-	- 💟 –	-	
🕻 🖷 Other resistance genes			CDS	ermA spc bleO aadD1 ant\(kan	any	-	-	-			2	-	-		-	-	- 0	-	
🕻 🖷 ccR locus			CDS	ссг	gene	-	-	-			4	-	-	\checkmark	gene	Opposite	- 💟 –	-	
🕻 🖷 mec locus			CDS	mec	gene	-	-	-			4	-	-	\checkmark	gene	-	- 💟 –	-	
🕻 🖷 orfX			CDS	orfX protein X	product	-	-	-			2	-	-	\checkmark	product	Тор	- 0	-	

- displayed features are selected according to the detection of regular expressions ('filter' column) in userdefined fields of GenBank annotations.

- bottom-to-top priority: features will be displayed with parameters of the last matching filter (*i.e.* transposases) of the **IS431** family will be displayed with hatching)

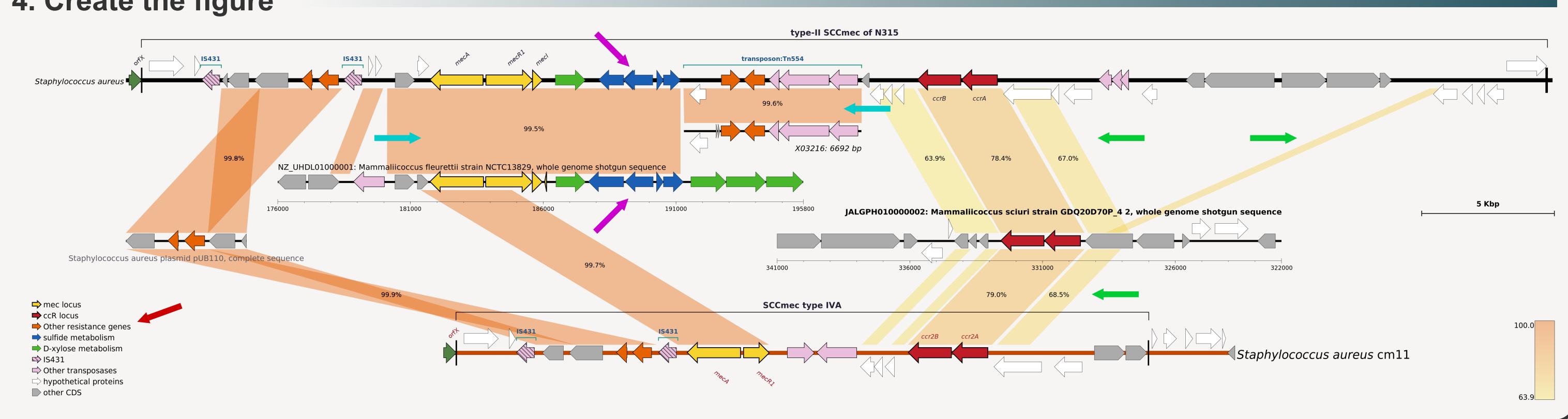
- feature names are user-defined and can be displayed in the legend.

3. Add homologies



- homologies are based on blastN or tblastX. Blast hits (matches) are stored in a file and are not computed every time.
- define which and how matches will be displayed

- select between which pair of sequences to display homologies with the 'Custom' matches selection.



4. Create the figure

Technical implementation

GenoFig is developed in python3, using biopython and cairosvg modules. The GUI (graphical user interface) is implemented with the GTK3+ API. The drawing engine is based on the open-source code of **Easyfig 2** [2], with some major reimplementation.

[1] Partridge, S.R., Kwong, S.M., Firth, N., Jensen, S.O., 2018. Mobile Genetic Elements Associated with Antimicrobial Resistance. Clin Microbiol Rev 31, e00088-17. [2] Sullivan, M.J., Petty, N.K., Beatson, S.A., 2011. Easyfig: a genome comparison visualizer. Bioinformatics 27, 1009–1010.