

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

Maxime Branger, Sébastien Leclercq

▶ To cite this version:

Maxime Branger, Sébastien Leclercq. GenoFig: a user-friendly application for vizualisation and comparison of genomic regions. 9. Symposium on Antimicrobial Resistance in Animals and the Environment - ARAE, Jul 2023, Tours (France), France. 2023, Abstracts book - 9. Symposium on Antimicrobial Resistance in Animals and the Environment - ARAE. hal-04170283

HAL Id: hal-04170283 https://hal.inrae.fr/hal-04170283

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

										GenoFi	g 1.01 F	igure_Poster.genofig												• • • •
Application																								
Image width: 7500 Draw figure in: 朣 Figure) px _Poste	r.png				470 n	300 nargins	20	backgroo enhar	und color nce graph (CREATE FIG	URE											
	Iomolo	gies L	egends Decora	tion			200																	
sequence name i	infos	active	position	min	max	reverse	height	space lir	e width	color	seq.	hipo	position	offect	color	cizo	D I	feat.	tupo	position	color	cizo	rot	
sequence name i	infos	active	position	min	max	reverse	height	space below lir	width	color	seq. label	type	position	offset	color	size	BI	feat. label	type	position	color	size	rot.	
sequences reactives in sequence name i General B86934	infos	active	position left	min	max	reverse	height	space below lin	width	color	seq. label	type locus+description organism	position left	offset 60	color	size 38	B I	feat. label	type product	position Top -	color	size 30	rot. 0	
 sequences reactires in sequence name i General D86934 X03216 	infos Q Q	active	position left	min 4500 1	max max max	reverse	height 100 -	space below lir 250 - -	width	color	seq. label	type locus+description organism locus+size	position left - bottom/right	offset 60 -	color -	size 38 -	B I	feat. label	type product -	position Top -	color -	size 30 -	rot. 0 -	
 sequences reactires in sequence name i General D86934 X03216 NZ UHDL01000001 	infos Q Q Q	active	position left best blast best blast	min 4500 1 4267	max max max 24067	reverse	height 100 - -	space below lin 250 - - 290	e width 10 20 -	color	seq. label	type locus+description organism locus+size -	position left - bottom/right top/left	offset 60 - -	color - -	size 38 - -	B I	feat. label	type product - -	position Top - -	color 	size 30 -	rot. 0 - -	
 sequences reactires in sequence name i General D86934 X03216 NZ_UHDL01000001 NC_001384 	infos Q Q Q Q	active	position left best blast best blast	min 4500 1 4267 1	max max max 24067 max	reverse	height 100 - - -	space below lin 250 - - 290 0	e width 10 20 - -	color	seq. label	type locus+description organism locus+size - description	position left - bottom/right top/left bottom	offset 60 - -	color - -	size 38 - - -	B I	feat. label	type product - - -	position Top - - -	color 	size 30 - -	rot. 0 - - -	
 sequences reactires rea	infos Q Q Q Q Q Q	active	position left best blast best blast - x = 3200	min 4500 1 4267 1 2192	max max max 24067 max 5 40925	reverse	height 100 - - - -	space below lin 250 - - 290 0 550	e width 10 20 - - - -	color - - - -	seq. label	type locus+description organism locus+size - description -	position left - bottom/right top/left bottom top/right	offset 60 - - - 130	color - - -	size 38 - - -	B I	feat. label	type product - - - -	position Top - - -	color :	size 30 - - -	rot. 0 - - -	

- (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

1	selected	in legend	type	filter	in field	strand	shape	height ratio	Fill	color	line	width	color	hatching	label	type	position	BI	color s	size
General						none	arrow	1			1	I		none						
💼 other CDS			CDS		any	-	signal	0.7	S	-	-		-	-		gene	Bottom			ſ
💼 hypothetical proteins	\checkmark		CDS	hypothetical protein	product	lead up	-	-	<		-		-	-		-	Bottom			1
💼 Other transposases			CDS	transposase tnp[ABC]	any	-	-	-			2	2	-	-		product	-		-	1
🗐 IS431	\checkmark		CDS	IS431	product	-	-	-			2	2		diagonal		product	-		-	1
SCCmec			misc_feature	SCCmec	note	-	range	4.0		-	3	3	-	-		note	Тор	S 🗆	4	IO
📋 Mobile elements			mobile_element	:	any	-	range	1.5			4	1		-		mobile_element_type	Тор	S 🗌	-	f
🖶 D-xylose metabolism			CDS	xvloselRS00850	anv	-	-	-			-		-	-		-	-			1
🗐 sulfide metabolism			CDS	sulfite sulfide DsrE	product	-	-	-			-		-	-		gene	-	🗆 🗹		
🖶 Other resistance genes			CDS	ermA spc bleO aadD1 ant\(kan	any	-	-	-			2	2	-	-		-	-			4
ccR locus	≤		CDS	ссг	gene	-	-	-			4	1	-	-	\checkmark	gene	Opposite	🗆 🗹		f
💼 mec locus			CDS	mec	gene	-	-	-			4	1	-	-		gene	-	🗆 🗹		4
🗐 orfX	≤		CDS	orfX protein X	product	-	-	-			2	2	-	-	\checkmark	product	Тор			l.

- 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.