

GIPSy: Genomic Island Prediction Software Version 1.1.2

Summary

I. What is GIPSy?
II. Importance of genomic islands
III. Method summary
IV. GIPSy input formats
IV.1 EMBL format (.embl)
IV.2 Genbank format (.genbank, .gb, .gbk)4
1. Installation process:5
2. Running GIPSy
2.1 Step 1
2.2 Going further from one Step to another9
2.3 Step 2
2.4 Step 3 4
2.5 Step 4 5
2.6 Step 56
2.6.1 Visualizing the amino acid composition and blast result for each gene (Step 5)
2.7 Step 6
2.8 Step 7 8
2.9 Step 8
3. Saving results and exporting additional analyses10
4. TroubleShooting
4.1 GIPSy breaks at step 5 on Linux11
4.2 GIPSy breaks at steps 4 and 7 on Linux11
4.3 GIPSy breaks at steps 3 to 7 on Linux11
5. Acknowledgments

I. What is GIPSy?

GIPSy is a software for accurate prediction of genomic islands into the classes: pathogenicity islands (PAIs), resistance Islands (RIs), metabolic Islands (MIs) and symbiotic Islands (SIs).

II. Importance of genomic islands

Bacteria are highly diverse organisms that are able to adapt to a broad range of environments and hosts due to their high genomic plasticity. Horizontal gene transfer plays a pivotal role in this genome plasticity and evolution by leaps through the incorporation of large blocks of genome sequences, ordinarily known as genomic islands (GEIs). GEIs may harbor genes encoding virulence, metabolism, antibiotic resistance and symbiosis-related functions, namely pathogenicity islands (PAIs), metabolic islands (MIs), resistance islands (RIs) and symbiotic islands (SIs).

III. Method summary

GIPSy predicts GEIs based on commonly shared features: genomic signature deviation (G+C content and codon usage); presence of tranposase genes; virulence, metabolism, antibiotic resistance, or symbiosis factors; flanking tRNA genes; and absence in other organisms of the same genus or closely related species. Eight steps are necessary to evaluate the presence of these genomic features in GIPSy.

IV. GIPSy input formats

GIPSy accepts complete genomes only of bacteria in embl or genbank formats.

IV.1 EMBL format (.embl)

The embl format has the following structure, where the regions highlighted with red boxes are important for GIPSy analyses.

FT	source	1330	9401				
FT	Dealer	/organ	ism="Corvne	ebacterium (glutamicum /	ATCC 13032"	
FT		/strain	n="ATCC 130	032"			
FT		/mol t	vne="genom"	ic DNA"			
FT		/db xr	ef="taxon:	196627"			
FT	CDS	1157	5				
FT		/codon	start=1				
FT		/trans	1 table=11				
FT		/locus	tag="Cg100	001"			
FT		/produ	ct="ATPase	involved in	n DNA replic	ation initiat	ion"
FT		/trans	lation="MS(QNSSSLLETWR	QVVADLTTLSQ	ADSGFDPLTPTQR	AYLNLTK
FT		PIAIVD	GYAVLSTPNAM	MAKNVIENDLG	DALTRVLSLRM	RSFSLAVSVEPEQ	EIPETPA
FT		QQEFKY	QPDAPVISSN	KAPKQYEVGGR	SEASTSDGWER7	THSAPAPEPHPAPI	ADPEPEL
FT		ATPORI	PRETPAHNPN	REVSLNPKYTF	ESEVIGPENRE	ANAAAVAVAESPAK	AFNPLFI
FT		SGGSGL	GKTHLLHAVG	NYAQELQPGLR	IKYVSSEEFTNI	YINSVRDDRQETF	KRRYRNL
FT		DILMVD	DIQFLAGKEG	TQEEFFHTFNA	LHQADKQIILS:	DRPPKQLTTLEDR	LRTRFEG
FT		GLITDI	OPPDLETRIA	ILMKKAQTDGT	HVDREVLELIAS	SRFESSIRELEGAL	IRVSAYS
FT		SLINQP	IDKEMAIVAL	RDILPEPEDME	ITAPVIMEVTAR	EYFEISVDTLRGAG	KTRAVAH
FT		ARQLAM	YLCRELTDMS	LPKIGDVFGGK	DHTTVMYADRK	RQEMTEKRDTYDE	IQQLTQL
FT		IKSRGR	N"				
XX							
SQ	Sequence 33	309401 BP; 7	64350 A; 89	94542 C; 88	6255 G; 7642	254 T; 0 other	;
	gtgagccaga	actcatcttc	tttgctcgaa	acctggcgcc	aagttgttgc	cgatctcaca	60
	actttgagcc	agcaagcgga	cagtggattc	gacccattga	cgccaactca	acgtgcatat	120
	ttgaacctga	cgaagccgat 1	tgccatcgtc	gatggctacg	ccgtgctgtc	cacacccaac	180
	gcgatggcaa	aaaatgtcat	tgaaaacgat	ttgggcgatg	ctttgacccg	tgtgttgtcg	240
	ctgcgcatgg	gccgatcatt	cagettgget	gtcagtgtgg	agcctgagca	ggaaattcca	300
	gaaaccccag	ctcagcagga	gtttaaatat	cagectgacg	cacctgtgat	ctcttccaac	360
	aaggcgccaa	agcagtatga	agttggtggt	cggggagagg	cgtcgacaag	cgacggctgg	420
	gaacgtaccc	actctgcacc	ggctcccgag	ccgcacccgg	cacctatcgc	cgatcctgag	480
	ccagagetgg	ccaccccgca	gcgcattccg	cgcgaaaccc	cagetcacaa	ccctaatcgg	540
	gaagtgtccc	tcaacccgaa	atacactttt	gaaagcttcg	tgatcgggcc	gttcaaccgt	600
	ttcgccaatg	cageegeagt :	tactataaca	gaaagcccag	cgaaagcttt	caacccgctg	660

IV.2 Genbank format (.genbank, .gb, .gbk)

Although presenting similar information, the genbank file presents a different structure, as shown below. Likewise the embl format example, the important features are highlighted by red boxes in the figure.

FEATURES		Location/Qualifiers
sourc	e	15231428
		/organism="Escherichia coli CFT073"
		/mol type="genomic DNA"
		/strain="CFT073"
		/db xref="taxon:199310"
gene		190255
		/gene="thrL"
		/locus tag="c5491"
CDS		190255
		/gene="thrL"
		/locus_tag="c5491"
		/function="leader; Amino acid biosynthesis: Threonine"
		/note="Thr operon attenuator; Escherichia coli K-12
		ortholog: b0001; Escherichia coli 0157:H7 ortholog: z0001
		/codon_start=1
		/transl_table=11
		/product="Thr operon leader peptide"
		/protein_id="AAN78501.1"
		/protein_id="E.coli:c5491"
		/db_xref="GI:26106315"
		/translation="MKRISTTITTITITITGNGAG"
DRIGIN		
1	agcttttcat	tctgactgca acgggcaata tgtctctgtg tggattaaaa aaagagtgtc
61	tgatagcagc	ttctgaactg gttacctgcc gtgagtaaat taaaatttta ttgacttagg
121	tcactaaata	ctttaaccaa tataggcata gcgcacagac agataaaaat tacagagtac
181	acaacatcca	tgaaacgcat tagcaccacc attaccacca ccatcaccat taccacaggt
241	aacggtgcgg	gctgacgcgt acaggaaaca cagaaaaaag cccgcacctg acagtgcggg
301	CTTTTTTTT	tgcacagaaa acccccagct aggctggggg ttccggaaag ctttcagctt
361	tgagccagtt	attaaaaccc cttttgattt gttaaaacac cttgcggtct ggcaactgca
421	agtgtcaaac	aagaaatcaa aagggggtcc caatggggaa cgaaaagagc ttagcgcaca
481	cccgatggaa	ctgtaaatat cacatagtat ttgcgccaaa ataccgaaga caggtgttct
E 4 1		acategeace steadotate tittaegees actatataeg tageseeata

1. Installation process:

GIPSy may be used both in Windows- and Linux-based platforms. It only requires a Java Virtual Machine 1.7.0_51-b13 or higher to work. However, we strongly advise the use of openjdk instead of the Oracle version of java virtual machine when working in linux-based machines as the Oracle version may result in some exceptions during the analyses. The download of all databases and dependencies as well as the environment setup are performed by GIPSy.Installer.jar.

GIPSy.Installer.jar may be run in Windows by simply double clicking, whereas in Linux it only requires a simple command line, as follow:

java –jar GIPSy.Installer.jar



a) choose the appropriate folder for installation:

b) click install:



c) the installer will automatically download the databases, dependencies, GIPSy.jar and set up all the environment:

<u>\$</u>	GIPSy Installer	- 🗆 🗙
	ika ta inatall (TROWD
	ike to install G	stpsy?
C:\Users\Siomar\GIPSy		Choose folder
Install		
50% Dov	wnloading database	s
		6
		Close
If you have any doubt,	please contact the	e developer:
Siomar Soares ciomars/	agmail.com	
Sional Soares sionars	-	
	CIDCulastallar	- D ×
د د د د د د د د د د د د د د د د د د د	GIPSy Installer	- • ×
ی است میں	GIPSy Installer	×
Where would you i C:\Users\Siomar\GIPSy	GIPSy Installer	- ×
Where would you i C:\Users\Siomar\GIPSy	GIPSy Installer	- Choose folder
Where would you I C:\Users\Siomar\GIPSy Install Dot: Dot	GIPSy Installer ike to install G	- Choose folder
Where would you I C:\Users\Siomar\GIPSy Instal 100% Dov 4%	GIPSy Installer ike to install G wnloading database wnloading depender	- Choose folder
Where would you i C:\Users\Siomar\GIPSy Install 100% Dov 4%	GIPSy Installer ike to install G wnloading database wnloading depender	- Choose folder
Where would you i C:\Users\Siomar\GIPSy Instal 100% Dov 4%	GIPSy Installer ike to install G wnloading database wnloading depender	- Choose folder
Where would you i C:\Users\Siomar\GIPSy Instal 00% Dov 4%	GIPSy Installer ike to install G wnloading database wnloading depender	_ Choose folder
Where would you i C:\Users\Siomar\GIPSy Instal 100% Dov 4%	GIPSy Installer ike to install G wnloading database wnloading depender	- Choose folder
Where would you I C:\Users\Siomar\GIPSy install 100% Dov 4%	GIPSy Installer ike to install G wnloading database wnloading depender	Lose
Where would you I C:\Users\Siomar\GIPSy Install 100% Dov 4% If you have any doubt, Siomar Soares siomars6	GIPSy Installer ike to install G wnloading database wnloading depender please contact the agmail.com	Lose e developer:

d) one shortcut will be created on Desktop.

OBS.: If it is not created, a prompt will appear warning that the Desktop folder was not found. In this case, simply go to the installation folder defined on the beginning of the installation process and execute GIPSy.jar from within this folder or create a Desktop shortcut.



2. Running GIPSy

Similarly to the GIPSy.Installer.jar, GIPSy.jar may be run in Windows by simply double clicking, whereas in Linux-based environments, it requires a simple command line, as follow:

java – jar GIPSy.jar

2.1 Step 1

After the software is open, click the button "Open Query" and choose the file for the pathogenic organism "Escherichia_coli_CFT073.gbk" (downloadable as part of the dataset.zip example files at http://www.bioinformatics.org/downloads/index.php?file_id=587) and click the buttom "Create files".

A) Creating files for the query genome.



b) Creating files for the subject genome

Click the button "Open Subject" and choose the file for the non-pathogenic closely-related organism "Escherichia_coli-K-12_MG1655.gbk" (downloadable as part of the dataset.zip example files at <u>http://www.bioinformatics.org/downloads/index.php?file_id=587</u>) and click the buttom "Create files".



2.2 Going further from one Step to another.

After finishing Step 1, you may click Step 2 or next in order to go further.

Obs.: All steps are dependent on Step 1. After step 1 is finished, you may run Steps 2-7 concomitantly. Step 8 is dependent on all previous Steps. Be sure to finish all of them before going to Step 8.



2.3 Step 2

< Previous

On Step 2, you may choose the cutoff value, in standard deviations (SD), for anomalous G+C analyses. The standard value is 1.5 SDs from the mean. After choosing the cutoff for both query and subject, run the analyses and click visualize in order to see the results.

	CIPCon Company's Island Designation Conference	_ _ ×
Ella Dapandancias	GIPSy: Genomic Island Prediction Software	
Step 1 Step 2 Step 3 Step 4 Step 5 Step 6 Step	p 7 Step 8	
Perform G+C content analyses		
Calculate G+C deviation on query genome	Visualize	
Calculate G+C deviations on subject genome 1.5 Kg Run	Visualize	
< Previous		Next >
		-
\$	GIPSy: Genomic Island Prediction Software	- • ×
File Dependencies		
Step 1 Step 2 Step 3 Step 4 Step 5 Step 6 Step	p 7 Step 8	
Perform G+C content analyses		
	1 March 199	
Calculate G+C deviations on subject genome	Visualize	
	VISUAIIZE	
< Previous		Next >
< Previous		Next >
< Previous		Next >
< Previous		Next>
< Previous	GIPSy: Genomic Island Prediction Software	Next >
< Previous File Dependencies	GIPSy: Genomic Island Prediction Software	Next>
<pre></pre>	GIPSy: Genomic Island Prediction Software	Next>
File Dependencies Step 1 Step 2 Step 3 Step 4 Step 5 Step 6 Step Perform G+C content analyses	GIPSy: Genomic Island Prediction Software	Next >
File Dependencies Step 1 Step 2 Step 3 Step 4 Step 5 Step 6 Step Perform G+C content analyses	GIPSy: Genomic Island Prediction Software	Next>
File Dependencies File Dependencies Step 1 Step 2 Step 3 Step 4 Step 5 Step 6 Step Perform G+C content analyses Calculate G+C deviation on query genome 15 Pun Televiste Graditation con the step of approxeme 15 Pun	GIPSy: Genomic Island Prediction Software p7 Step 8 Visualize Donel Visualize Donel	Next >
File Dependencies Step 1 Step 2 Step 3 Step 5 Step 6 Step 7 Perform G+C content analyses Calculate G+C deviation on query genome 1.5 Run Calculate G+C deviations on subject genome 1.5 Run	GIPSy: Genomic Island Prediction Software p7 Step 8 Visualize Done! Done! Done!	Next>
File Dependencies Step 1 Step 2 Step 3 Step 4 Step 5 Step 6 Step 7 Perform G+C content analyses Calculate G+C deviation on query genome 1.5 Run Calculate G+C deviations on subject genome 1.5 Run	GIPSy: Genomic Island Prediction Software p7 Step 8 Visualize Done! Visualize Done!	Next>
	GIPSy: Genomic Island Prediction Software p7 Step 8 Visualize Done! Visualize Done!	Next >
File Dependencies File Dependencies Step 1 Step 2 Step 3 Step 4 Step 5 Step 6 Step Perform G+C content analyses Calculate G+C deviation on query genome 1.5 Run Calculate G+C deviations on subject genome 1.5 Run #Total G+C content 2640743.0 #GC percent 50.479	GIPSy: Genomic Island Prediction Software p7 Step 8 Visualize Done! Done!	Next >
File Dependencies File Dependencies Step 1 Step 2 Step 3 Step 4 Step 5 Step 6 Ste Perform G+C content analyses Calculate G+C deviation on query genome 15 Run #Total G+C content 2640743.0 #Gomme Ster 321428.0 #Total G+C content 50.479 #The reference values using 1.5 SD are : lower limit (42.13301251 #Total reference values using 1.5 SD are : lower limit (42.13301251) #Total reference values using 1.5 SD are : lower limit (42.13301251) #Total reference values using 1.5 SD are : lower limit (42.13301251) #Total reference values using 1.5 SD are : lower limit (42.13301251) #Total reference values using 1.5 SD are : lower limit (42.13301251) #Total reference values using 1.5 SD are : lower limit (42.13301251) #Total reference values using 1.5 SD are : lower limit (42.13301251) #Total reference values using 1.5 SD are : lower limit (42.13301251) #Total reference values using 1.5 SD are : lower limit (42.13301251) #Total reference values using 1.5 SD are : lower limit (42.13301251) #Total reference values using 1.5 SD are : lower limit (42.13301251) #Total reference values using 1.5 SD are : lower limit (42.13301251) #Total reference values using 1.5 SD are : lower limit (42.13301251) #Total reference values using 1.5 SD are : lower limit (42.13301251) #Total reference values using 1.5 SD are : lower limit (42.13301251) #Total reference values using 1.5 SD are : lower limit (42.13301251) #Total reference values using 1.5 SD are : lower limit (42.13301251) #Total reference values using 1.5 SD are : lower limit (42.13301251) #Total reference values using 1.5 SD are : lower limit (42.13301251) #Total reference values using 1.5 SD are : lower limit (42.13301251) #Total reference values using 1.5 SD are : lower limit (42.13301251) #Total reference values using 1.5 SD are : lower limit (42.13301251) #Total reference values using 1.5 SD are : lower limit (42.13301251) #Total reference values using 1.5 SD are : lower limit (42.13301251) #Total reference values using 1.5 SD are : lower limit (42.13301251) #Total reference va	GIPSy: Genomic Island Prediction Software p7 Step 8 Visualize Done! Usualize Done! 14249174) and upper limit (58.824987485750825)	Next >
File Dependencies Step 1 Step 2 Step 3 Step 4 Step 5 Step 6 Step Perform G+C content analyses Calculate G+C deviation on query genome 1.5 Run Elouidate G+C deviations on subject genome 1.5 Run #Total G+C content: 2640743.0 #Genome Size: 5231428.0 #GO percent: 50.479 #The reference values using 1.5 SD are: lower limit (42:13301251 #The reference values using 1.5 SD are: lower limit (42:13301251) #The reference values using 1.5 SD are: lower limit (42:1351515151) NORWAL 0.51515151515151	GIPSy: Genomic Island Prediction Software p7 Step 8 Visualize Done! Done! 14249174) and upper limit (58.824987485750825)	Next>
Image: Step 2 Step 3 Step 4 Step 5 Step 6 Step 7 Step 1 Step 2 Step 3 Step 4 Step 5 Step 6 Step 7 Perform G+C content analyses Image: Step 3 Step 4 Step 5 Step 6 Step 7 Calculate G+C deviations on subject genome Image: Step 3 Step 7 Run If Total G+C content: 2640743.0 Image: Step 3 Run If Total G+C content: 2640743.0 Image: Step 3 Run If Total G+C content: 2640743.0 Image: Step 3 Run If Total G+C content: 2640743.0 Image: Step 3 Run If Total G+C content: 2640743.0 Image: Step 3 Run If Total G+C content: 2640743.0 Image: Step 3 Run If Total G+C content: 2640743.0 Image: Step 3 Run If Total G+C content: 2640743.0 Image: Step 3 Run If Total G+C content: 2640743.0 Image: Step 3 Run If Total G+C content: 2640743.0 Image: Step 3 Run If Total G+C content: 2640743.0 Image: Step 3 Run If Total G+C content: 2640743.0 Image: Step 3 Run	GIPSy: Genomic Island Prediction Software p7 Step 8 Visualize Done! Done! 14249174) and upper limit (58.824987485750825)	Next >
Previous File Dependencies Step 1 Step 2 Step 3 Step 4 Step 5 Step 6 Ste Perform G+C content analyses Calculate G+C deviation on query genome 15 Run Calculate G+C deviations on subject genome 15 Run protal G+C content 2840743.0 #Compone Size \$221428.0 #Compone Size \$221428.0 #The reference values using 1.5 SD are : lower limit (42.13301251 #The reference values using 1.5 SD are : lower limit (42.13301251 #The reference values using 1.5 SD are : lower limit (42.13301251 #The reference values using 1.5 SD are : lower limit (42.13301251 #The reference values Using 1.5 SD are : lower limit (42.13301251 #The reference values Using 1.5 SD are : lower limit (42.13301251 #The reference values Using 1.5 SD are : lower limit (42.13301251 #The reference values Using 1.5 SD are : lower limit (42.13301251 #The reference values Using 1.5 SD are : lower limit (42.13301251 #The reference values Using 1.5 SD are : lower limit (42.13301251 #The reference values Using 1.5 SD are : lower limit (42.13301251 #The reference values Using 1.5 SD are : lower limit (42.13301251 #The reference values Using 1.5 SD are : lower limit (42.13301251 #The reference values Using 1.5 SD are : lower limit (42.13301251 #The reference values Using 1.5 SD are : lower limit (42.13301251 #The reference values Using 1.5 SD are : lower limit (42.13301251 #The reference values Using 1.5 SD are : lower limit (42.13301251 #The reference values Using 1.5 SD are : lower limit (42.13301251 #The reference values Using 1.5 SD are : lower limit (42.13301251 #The reference values Using 1.5 SD are : lower limit (42.13301251 #The reference values Using 1.5 SD are : lower limit (42.13301251 #The reference values Using 1.5 SD are : lower limit (42.13301251 #The reference values Using 1.5 SD are : lower limit (42.13301251 #The reference values Using 1.5 SD are : lower limit (42.13301251 #The reference v	GIPSy: Genomic Island Prediction Software p7 Step 8 Visualize Done! 14249174) and upper limit (58.824987485750825)	Next>
Previous File Dependencies Step 1 Step 2 Step 3 Step 4 Step 5 Step 6 Step Perform G+C content analyses Calculate G+C deviation on query genome 15 Run Pfolal G+C content 2640743.0 #GC percent 50.479 #The teference values using 1.5 SD are: lower limit (42.1330.1261 # Code Value	GIPSy: Genomic Island Prediction Software p7 Step 8 Visualize Done! Visualize Done! 14249174) and upper limit (56.824987485750825)	Next >
	GIPSy: Genomic Island Prediction Software	
Previous File Dependencies Step 1 Step 2 Step 3 Step 4 Step 5 Step 6 Ste Perform G+C content analyses Calculate G+C deviation on query genome 15 Run #Total G+C content 26:40743.0 #GC percent: 50:473 #The reference values using 1.5 SD are: lower limit (42:13301251 # #For percent: 50:473 #The reference values using 1.5 SD are: lower limit (42:13301251 # # Code and the set of the	GIPSy: Genomic Island Prediction Software p7 Step 8 Visualize Done! Usualize Done! Jone! 14249174) and upper limit (58.824987485750825)	×
Image: system in the image is a system in the im	GIPSy: Genomic Island Prediction Software p7 Step 8 Visualize Done! 14249174) and upper limit (58.824987485750825)	Next >
Image: system in the image is a system in the im	GIPSy: Genomic Island Prediction Software p7 Step 8 Visualize Done! Done! 14249174) and upper limit (58.824987485750825)	- □ ×

Next >

2.4 Step 3

On Step 3, you may choose the sensitivity, from 0.5 to 0.95, for codon usage deviation analyses in Colombo/SigiHMM. The standard value is 0.95. After choosing the cutoff for both query and subject, run the analyses and click visualize in order to see the results.

📓 GIPSy: Genomic Island Prediction Software – 🗖	×
File Dependencies	
Step 1 Step 2 Step 3 Step 4 Step 5 Step 6 Step 7 Step 8	
Perform Codon Usage Analyses Cilleraris Emandamical Colombo 2 Blok Similum clans	
Cluder stanman our ay dependencies Contribu-Lator ayunnin class	
Define sensitivity to calculate codon usage deviation on query genome:	
Define sensitivity to calculate codon usage deviation on subject genome: 0.95 reaction Run Visualize	
	_
< Previous Ne:	(t >

<u>\$</u>	GIPSy: Genomic Island Prediction Software	_ 🗆 🗙
File Dependencies		
Step 1 Step 2 Step 3 Step 4 Step	5 Step 6 Step 7 Step 8	
Perform Codon Usage Analyses C:\Users\Siomar\GIPSy\Dependencies\Colomb	bo_3.8/ok SigiHMM.class	
Define sensitivity to calculate codon usage de	eviation on query genome: 0.95 Run Visualize Running! Please wait!	
Define sensitivity to calculate codon usage de	eviation on subject genome: 0.95 Run Visualize Running! Please wait!	
< Previous		Next >
e.	GIDS:r Genomic Island Prediction Software	- D ×
ے۔ File Dependencies	GIPSy: Genomic Island Prediction Software	- • ×
file Dependencies Sten 1 \Sten 2 \Sten 3 \Sten 4 \Sten	GIPSy: Genomic Island Prediction Software	- • ×
File Dependencies File Dependencies Step 1 Step 2 Step 3 Step 4 Step Perform Codon Usage Analyses C:Users\SiomariGIPSylDependencies\Colomb	GIPSy: Genomic Island Prediction Software	×
File Dependencies File Dependencies Step 1 Step 2 Step 3 Step 4 Step Perform Codon Usage Analyses C:/Users!Siomar/GIPSy/Dependencies/Colomb Define sensitivity to calculate codon usage de Define sensitivity to calculate codon usage de	GIPSy: Genomic Island Prediction Software 5 Step 6 Step 7 Step 8 so_3.8iok SigiHMM.class eviation on query genome: 0.95 Run Visualize Donel eviation on subject genome: 0.95 Run Visualize	×
File Dependencies File Dependencies Step 1 Step 2 Step 3 Step 4 Step Perform Codon Usage Analyses Culsers/Siomar/GIPSy/Dependencies/Colomb Define sensitivity to calculate codon usage de Define sensitivity to calculate codon usage de # # Core-specific parameters: # Menu parameters: # Menu parameters: # PHX masking "on" # Mask Outliers "off" # Johing distance "1" # Johing distance "1" # Advanced options: # Sensitivity .0.95 # Sensity .0.95 # Sensitivity .0.95 # Sensitivity .0.95 # Sensitivity .0.95	GIPSy: Genomic Island Prediction Software	
File Dependencies Step 1 Step 2 Step 3 Step 4 Step Perform Codon Usage Analyses C:UsersISiomar/GIP Sy/Dependencies/Colomb Define sensitivity to calculate codon usage de Define sensitivity to calculate codon usage de P# # Core-specific parameters: # Menu parameters: # Menu parameters: # Menu parameters: # Menu parameters: # Advanced options: Sensitivity. 0.95 Colombo/SigiHMM c0001 Colombo/SigiHMM c0004 Colombo/SigiHMM c0005 Colombo/SigiHMM c0006 Colombo/SigiHMM c0009 c0009 Colombo/SigiHMM c0009 c0009 Colombo/SigiHMM c0009 c0009 Colombo/SigiHMM c0009 c009	SilPSy: Genomic Island Prediction Software 5 Step 6 Step 7 Step 8 so_3.8iok SigiHMM.class sviation on guery genome: 0.95 Run Visualize Donet viviation on subject genome: 0.95 Run Visualize Donet NORMAL NORMAL NORMAL NORMAL NORMAL NORMAL NORMAL NORMAL NORMAL NORMAL	

2.5 Step 4

On Step 4, you may choose the e-value for transposase prediction using HMMer. The standard value is 0.0001. After choosing the e-value, run the analyses and click visualize in order to see the results.

<u><u><u></u></u></u>		
File Dependencies	GIPSy: Genomic Island Prediction Software	
Sten 1 Sten 2 Sten 3 Sten 4 Sten	5 Stan 6 Stan 7 Stan 8	
Search for transposase genes		
C:\Users\Siomar\GIPSy\Dependencies\hmme	r\hmmsearch.exe	
Search for transposase genes on query gen	ome 0.0001 T Run Visualize	
< Previous		Next >
<u>s</u>	GIDSv: Genomic Island Prediction Software	- 🗆 🗙
File Dependencies	dirsy, denomicisiand rediction software	
Step 1 Step 2 Step 3 Step 4 Step	5 Sten 6 Sten 7 Sten 8	
Search for transposase genes		
C:\Users\Siomar\GIPSy\Dependencies\hmme	r/hmmsearch.exe	
Search for transposase genes on query gen	ome 0.0001 Run Visualize Running! Please vait!	
< Previous		Next >
< Previous		Next>
< Previous	GIPSur Genomic Island Prediction Software	Next>
< Previous	GIPSy: Genomic Island Prediction Software	Next>
< Previous File Dependencies Stant J Stant Fata	GIPSy: Genomic Island Prediction Software	Next>
< Previous File Dependencies Step 1 Step 2 Step 3 Step 4 Step	GIPSy: Genomic Island Prediction Software	Next>
Previous File Dependencies Step 1 Step 2 Step 3 Step 4 Step 5 Scarch for transposase genes Scalusers/Grand/DS/yDependencies/hmme	GIPSy: Genomic Island Prediction Software	×
<pre>< Previous File Dependencies Step 1 Step 2 Step 3 Step 4 Step Search for transposase genes C:UlservilSiomariGIPSy/Dependencies/Imme Face bet for transposase genes Face bet for transposase gene</pre>	GIPSy: Genomic Island Prediction Software	Next>
Previous File Dependencies Step 1 Step 2 Step 3 Step 4 Step 5 Search for transposase genes C:Userai SlomariGIPSyDependencies/Imme Search for transposase genes on query gene Parased File	GIPSy: Genomic Island Prediction Software	Next>
File Dependencies Step 1 Step 2 Step 4 Step 5 Step 1 Step 2 Step 4 Step 4 Step 4 Step 1 Step 2 Step 3 Step 4 Step 4 Step 4 Step 1 Step 2 Step 3 Step 4 <	GIPSy: Genomic Island Prediction Software	Next>
File Dependencies Step 1 Step 2 Step 4 Step 5 Search for transposase genes C:Usersi Siomari GiP SyDependencies Imme Step 7 Step 4 Step 4 Parsed off transposase genes C:Usersi Siomari GiP SyDependencies Imme Step 1 Step 2 Step 4 Step 4 # Mimms and: Search for transposase genes C:Usersi Siomari GiP SyDependencies Imme Step 4	GIPSy: Genomic Island Prediction Software	Next>
File Dependencies Step 1 Step 2 Step 3 Step 4 Step 4 Step 1 Step 2 Step 3 Step 4 Step 4 Search for transposase genes C:UsersiSiomariGiPSyDependenciesihmme Step 7 Step 4 Step 7 Parased for transposase genes HMIRE 30: Alder: AD101, Minhamme org Alder: AD101, Minhamme org Polyopht (C) 2010 Howard Hughes Medica # Freely distingt dingt free Child lasers/Siomard/G Somard Hull Bite Child lasers/Siomard/G	GIPSy: Genomic Island Prediction Software	Next>
File Dependencies Step 1 Step 2 Step 3 Step 4 Step 4 Step 1 Step 2 Step 3 Step 4 Step 4 Step 1 Step 2 Step 3 Step 4 Step 4 Step 1 Step 2 Step 3 Step 4 Step 4 Search for transposase geness C:UsersiSiomarGiPSyDependencies/hmme org Step 4 Mider: A010,	GIPSy: Genomic Island Prediction Software 25 Step 6 Step 7 Step 8 r/Immsearch.exe ome 0.0001 Run Visualize quence database Institute. Dontel Institute. Step 6 (GPL/3).	Next>
File Dependencies File Dependencies Step 1 Step 2 Step 3 Step 4 Step 4 Step 1 Step 2 Step 3 Step 4 Step 4 Step 1 Step 2 Step 3 Step 4 Step 4 Step 1 Step 2 Step 3 Step 4 Step 4 Search for transposase genes C:UsersisionariGiP SyDependencies/hmme org Pravad Hudre 2010; Minhamme org Pravad Hudre 2010; Minhamme org Propright (C) 2010 Howard Hudres Medica USE (Number 60); Provide Hudres 10: C:UsersiSionariG Pravad Step 5 Pravad Cole 5 C:UsersiSionariG # ready distingt sequence datases: C:UsersiSionariG Propright (C) 2010 Howard Hudres Medica USE (Number 60); Provide Step 5 Proversionarid Cole 5 Prov	GIPSy: Genomic Island Prediction Software 25 Step 6 Step 7 Step 8 rthmmsearch.exe Dontel	Next>
File Dependencies Step 1 Step 2 Step 4 Step 4 Step 2 Step 3 Step 4 Step 5 Search for transposase genes C:Users/SiomarG/BPSyUbpendencies/Immmes org/ Step 4 Parsed File PHMLRER 30 (March 2010) http://mmmes org/ Pireview Mathbuled under the CNU Genaral P P comprisht C: 2010 Howard Hupes Medical Pireview Stematical Culters/SiomarG Pireview Stematical Cuters/SiomarG # comprisht C: 2010 Howard Hupes Medical Pireview Stematical Cuters/SiomarG Pireview Stematical Cuters/SiomarG # comprisht C: 2010 Howard Hupes Medical Pireview Stematical Cuters/SiomarG Pireview Stematical Cuters/SiomarG # comprisht Pirevical Cuters/SiomarG Cuters/SiomarG Pireview Stematical Cuters/SiomarG # comprisht Pirevical Cuters/SiomarG Cuters/SiomarG Pireview Stematical Cuters/SiomarG # comprisht Pirevical Cuters/SiomarG Cuters/SiomarG Pireview Stematical Cuters/SiomarG	GIPSy: Genomic Island Prediction Software 15 Step 6 Step 7 Step 8 rhtmmsearch.exe	
Previous File Dependencies File Dependencies Step 1 Step 2 Step 3 Step 4 Step Search for transposase genes C:Users/Siomar/GiP Sy/Dependencies/Imme Search for transposase genes on query gen # Hmmsearch : search profile(s) against a se # HMMER 30 (March 2010), http://mmmsr.org/ # gequenci reprofile the SNUL General P # Guery HMM file: C:Users/Siomar/ # gequenci reprofile the SNUL Centers/Siomar/ # gequenci reprofile the SNUL Centers/Siomar/ # gequenci reprofile the SNUL Centers/Siomar/ # gequenci reprofile the SNUL Centers/Siomar/ # gequenci reprofile the SNUL Centers/Siomar/ # gequenci reprofile the SNUL Centers/Siomar/ # gequenci reprofile the SNUL Centers/Siomar/ # gequenci reprofile the SNUL Centers/Siomar/ # gequenci reprofile the SNUL Centers/Siomar/ # Guery State the SNUL Centers/Siomar/ # gequenci reprofile the SNUL Centers/ # gequenci reprofile the SNUL Centers/ # g	GIPSy: Genomic Island Prediction Software	Next>
Previous File Dependencies File Dependencies Step 1 Step 2 Step 3 Step 4 Step Search for transposase genes Search for transposase genes Search for transposase genes Search for transposase genes HMMER 3.0 (March 2010), htp://mmms.org/ #Target sequence respondencies/hmme #GoupyInt(2:0010 Howard Huben Medical # Freidy distributed under the CAU General # GoupyInt(2:0010 Howard Huben Medical # Freidy distributed under the CAU General # GoupyInt(2:0010 Howard Huben Medical # Freidy distributed under the CAU General # GoupyInt(2:0010 Howard Huben Medical # Freidy distributed under the CAU General # GoupyInt(2:0010 Howard Huben Medical # GoupyInt(2:0010 Howard Huben Medical # GoupyInt(2:0010 Huben Huben Medical # Freidy distributed under the CAU General # GoupyInt(2:0010 Huben Huben Medical # Freidy distributed under the CAU General # GoupyInt(2:0010 Huben Huben Medical # Freidy distributed under the CAU General # GoupyInt(2:0010 Huben Huben Medical # Freidy distributed under the CAU General # GoupyInt(2:0010 Huben Huben Medical # GoupyInt(2:0010 Huben Huben Medical # Freidy distributed under the CAU General # GoupyInt(2:0010 Huben Huben Medical # Freidy distributed under the CAU General # GoupyInt(2:0010 Huben Huben Medical # Freidy distributed under the CAU General # Freidy distributed under the CAU General # GoupyInt(2:0010 Huben Huben Medical # Freidy distributed under the CAU General # GoupyInt(2:0010 Huben Huben Huben Huben Huben # GoupyInt(2:0010 Huben Huben Huben # GoupyInt(2:0010 Huben Huben Huben # GoupyInt(2:0010 Huben Huben # Freidy distributed # Freidy distributen # Freidy distributen # Freidy distributen # Fre	GIPSy: Genomic Island Prediction Software	Next>
Previous File Dependencies Step 1 Step 2 Step 3 Step 4 Step Search for transposase genes Culsers/Sioman/GIPSyUbpendencies/mmme Parsed File PhmRer3 0 (March 2010), http://mmme.org/ Popright (C) 2010 Howard Hughes Medical PhMRer3 0 (March 2010), http://mmme.org/ P Copright (C) 2010 Howard Hughes Medical File Courses/Sioman/OIPSyUbpendencies/Simme Previous/Simmer catabase: C.Users/Sioman/OIPSyUbpendencies/Simmer # sequence reporting threshold. E-value <= # # # # # Sequence reporting threshold. E-value <= #	GIPSy: Genomic Island Prediction Software	Next>
Previous File Dependencies Step 1 Step 2 Step 3 Step 4 Step Search for transposase genes Culteers/Siomar(GPSyDependencies/Immer org) Parsed File # Immearch: search profile(s) against as e # HMIER 30 (March 2010), Min/mmer org) # Copright (C) 2010 Howard Hughes Medica # Transposase genes on query gen # Previous Miteach 2010, Min/mmer org # copright (C) 2010 Howard Hughes Medica # copright (C) 2010 His Found	GIPSy: Genomic Island Prediction Software	Next>
Previous File Dependencies Step 1 Step 2 Step 3 Step 4 Step Search for transposase genes C:Users/SiomarGPSyDependencies/Imme org/ Parsa 61is PhMIRE 3 (Uarch 2010), http://mme.org/ Porgint (C) 2010 Howard Hughes Medica PhMIRE 30 (Uarch 2010), http://mme.org/ Porgint (C) 2010 Howard Hughes Medica Transposase genes on query gen # aurop HMM file: C.Users/SiomarG # aurop 4 MM file: C.Users/Siomar	GIPSy: Genomic Island Prediction Software	
Previous File Dependencies Step 1 Step 2 Step 3 Step 4 Step Search for transposase genes C:Users\SiomariGPSyDependencies\Imme ParsedFile Mommsarch.search profile(s) against as e HMIRER 30 (March 2010), http://mme.org/ ParsedFile Mommsarch.search profile(s) against as e HMIRER 30 (March 2010), http://mme.org/ ParsedFile Mommsarch.search profile(s) against as e HMIRER 30 (March 2010), http://mme.org/ ParsedFile Mommsarch.search profile(s) against as e HMIRER 30 (March 2010), http://mme.org/ ParsedFile Mommsarch.search profile(s) against as e Freely distributed under the CNU General P # Joudput director to hite. C/Users/SiomariG Flarget sequence reporting threshold. E-value <=1 # Mommsarch.search profile(s) against as e doubt and the Found Codol No Hits Found codol No Hits Found codol No Hits Found	GIPSy: Genomic Island Prediction Software	
Previous File Dependencies Step 1 Step 2 Step 3 Step 4 Step Search for transposace genes C:Users\SiomariGPSyDependencies\Immer PrarsedFile Momsaarch: search profile(s) agains1 as # Momsaarch: search profile(s) agains1 as # HMLER 30 (Warch 200), thir/mmer.org/ # Copyright (C) 2010 Howard Hughes Medica # HMLER 30 (Warch 200), thir/mmer.org/ # copyright (C) 2010 Howard Hughes Medica # devery HML file: C:Users\SiomariG # Guene: reporting threshold: E-value <= 1 #	GIPSy: Genomic Island Prediction Software	Next>

2.6 Step 5

On Step 5, you may choose the e-value for the prediction of specific factors using blastp. The standard value is 0.000001. Moreover, if you are striving to predict Pathogenicity Islands, run the prediction for virulence factors. If you are searching for Resistance Islands, run the analyses for antibiotic resistance genes, and so on. After choosing the e-value, run the analyses for "Virulence Factors" and click visualize in order to see the results.

•	CIPE: Conomic Iel	and Droc	liction C	oftwara						×
Eile Dependensis	GIPSy: Genomic Isla	ind Prec	netion S	onware						
The Dependencie										
Step 1 Step 2	Step 3 Step 4 Step 5 Step 6 Step 7 Step 8									
Search for sp	ecific factors									
C:\Users\Sioma	r\GIPSy\Dependencies\blast\bin\blastall.exe				e-val	ue				
Virulence Fac	tors (Pathogenicity Islands)				0.000	001	Run	Visu	ualize	
Antibiotic Re	sistance (Resistance Islands)				0.000		Dun	Vie	ualizo	
					0.000		Due	Min	alize	
Metabolism (Metabolic Islands)				0.000	001	Run	Vist	Jalize	
Symbiosis (S	ymbiotic Islands)				0.000	001	Run	Visu	Jalize	
Query ID	Subject ID	%ID	Alianm	Mismat	Gap O	Q Start	Q End	S Star	t SEnd E-valu	e bit score
										^
		_	_							
										=
										-
< Previous										Next >
		_	_	_	_		_	_		
<u>\$</u> ,	GIPSy: Genomic Isla	and Prec	liction S	oftware						- 🗆 🗙
File Dependencie	S									
Sten 1 Stor 2	Sten 3 Sten 4 Sten 5 Sten 6 Sten 7 Sten 9									
step i step z	steh a steh a steh a steh t steh t									
Search for sp	ecific factors									
C:\Users\Sioma	\GIPSy\Dependencies\blast\bin\blastall.exe		e	-value					14%	
Virulence Fac	tors (Pathogenicity Islands)		C	.000001	R	up I	Visualize			
Antibiotic Re	sistance (Resistance Islands)		l	0.000001	R		Visualize			
				000001			Vieualizo	-		
Metabolism (Metabolic Islands)		0				visualize			
Symbiosis (S	ymbiotic Islands)		0	0.000001	R	un \	Visualize			
Query ID	Subject ID	96ID	Alianm	Miema	Gan O	O Start	O End	S Start	S End E-value	hit score
Guery in	ougeris	7010	/ arginna	morriu	oup o	G.Oturt	GLLING .	0.otun	O.End E Valde	
										-
										_
										-
< Previous										Next >
<u>\$</u>	GIPSy: Genomic Isla	and Prec	liction S	oftware						- 🗆 🗙
File Dependencie	S									
Stan 4 (Stan 2	Chan 2 Chan 4 Chan 5 Chan 6 Chan 7 Chan 0									
Step 1 Step 2	Step 3 Step 4 Step 5 Step 6 Step 7 Step 8									
Search for sp	ecific factors									
C:\Users\Sioma	lGIPSy\Dependencies\blast\bin\blastall.exe		e	-value					Done!	
Virulence Fac	tors (Pathogenicity Islands)		C	.000001	R	un 🚺	Visualize			
Antibiotic Re	sistance (Resistance Islands)		Ī	0.000001	R	un	Visualize	13		
			5	000004			Visualizo			
metabolism (metabolic 15lands)		1				. Journal Co			
Symbiosis (S	ymbiotic Islands)		0	0.000001	R	un V	visualize			
Query ID	Subject ID	%ID	Alianm	Misma	Gan O	Q.Start	Q.End	S.Start	S.End E-value	bit score
c5491	No Hits Found		, angerer ta						- vuide	-
c0002 c0003	16516/vfid 37622/vsiid 58459 ssid hypothetical_protein_c1256_[Esche	100%	125	0		1 1	25 4	5	169 4e-092	263 📃
c0004	No Hits Found									
c0005 c0006	No Hits Found No Hits Found	-								
c0007	No Hits Found									
c0008 c0009	No Hits Found	-	-							
c0010	No Hits Found									
c0011 c0012	20539/vtd/45666/vsiid/66503/ssid/putative_transcriptional_regulato	30%	432	282	18	12 4	30 1	2 4	438 7e-044	160
c0013	No Hits Found	1	1							•
< Previous										Next>
										most -

2.6.1 Visualizing the amino acid composition and blast result for each gene (Step 5)

The first and second columns of the table in Step 5 are clickable. Choose some gene and click in the first column in order to see its amino acid composition.

e Dependenci	es	GIPSy: Genomic Island Prediction Software			
tep 1 Step 2	Step 3 Step	4 Step 5 Step 7 Step 8			
~					
search for s	pecific factors	2	x		
C:\Users\Sioma	ar\GIP Sy\Depende	-		Done!	
Virulence Fa	ctors (Pathod	>c0002 E 534:911 Hypothetical protein			
	(5	VEYREKRRAIGCILRKLCEWKSVRILEAECCADHIHMLVEIPPKMSVSGEMGYLKGKSSL			
Antibiotic Re	esistance (Res	MPYEQFGDLKFKYRNREFWCRGYYVDTVGKNTAKIQDYIKHQLEEDKMGEQLSIPYPGSP			
		FTGRK			
Metabolism	(Metabolic Isl				
Symbiosis (S	Symbiotic Isla				
Query ID				S.End E-value	bit score
c5491	No Hits Found			400 4 000	A
c0002	16516[VII0]376			169 46-092	263 =
0003_7	No Hits Found				
c0005	No Hits Found				
c0006	No Hits Found				
c0007	No Hits Found				
c0008	No Hits Found				
c0009	No Hits Found				
c0010	No Hits Found		-		
c0011	20539 vfid 456	1	-	438 7e-044	160
c0012	No Hits Found				
c0013	No Hits Found				•
	1				
< Previous					Next >
	-				

If you want to see the blast result of the given gene against the specific factor database, click the second column and the alignment will appear.

tep 1 Step 2	2 Step 3 Step 4 Step 5 Step 6 Step	7 Step 8
Search for s	specific factors	<u>&</u>
C:\Users\Siom	ar\GIPSv\Dependencies\blast\bin\blastall.exe	
		BLASTP 2.2.26 [Sep-21-2011]
Virulence Fa	ictors (Pathogenicity Islands)	
Antibiotic Re Metabolism	esistance (Resistance Islands) (Metabolic Islands)	Reference: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), 'Gapped BLAST and PSI-BLAST: a new generation of protein database search programs', Nucleic Acide Res. 25:3389-3402
Symbiosis (S	Symbiotic Islands)	Reference for compositional score matrix adjustment. Altschul, Stephen F., John C. Wontton F. Michael Gertz Richa Angrwata Aleksandr Morrulis
Query ID	Subject ID	Alejandro A. Schaffer, and Yi-Kuo Yu (2005) "Protein database searches
c5491	No Hits Found	using compositionally adjusted substitution matrices", FEBS J. 272:5101-5109.
c0002	16516 vfid 37622 vsiid 58459 ssid hypothetical_r	
	No Hits Found	Query= c0002 E 534:911 Hypothetical protein
c0003	No Filito Found —	
c0003 c0004	No Hits Found	(125 letters)
c0003 c0004 c0005	No Hits Found No Hits Found	(125 letters)
c0003 c0004 c0005 c0006	No Hits Found No Hits Found No Hits Found	(125 letters) Database: /home/jsmith/virulenceDB/BlastDB//rulenceDB.protein.fasta
c0003 c0004 c0005 c0006 c0007	No Hits Found No Hits Found No Hits Found No Hits Found No Hits Found	(125 letters) Database: /home/jsmith/virulenceDB/BlastDB/virulenceDB.protein.fasta 55,193 sequences; 10,514,445 total letters
c0003 c0004 c0005 c0006 c0007 c0008	No Hits Found No Hits Found No Hits Found No Hits Found No Hits Found No Hits Found	(125 letters) Database: /home/jsmith/virulenceDB/BlastDB/virulenceDB.protein.fasta 55,193 sequences; 10,514,445 total letters
c0003 c0004 c0005 c0006 c0007 c0008 c0009	No Hits Found No Hits Found	(125 letters) Database: /homei/smith/irulenceDB/BlastDB/irulenceDB protein.fasta 55,193 sequences; 10,514,445 total letters Searching
c0003 c0004 c0005 c0006 c0007 c0008 c0009 c0010	No Hits Found No Hits Found	(125 letters) Database: /home/jsmith/virulenceDB/BlastDBMrulenceDB.protein.fasta 55,193 sequences; 10,514,445 total letters Searching
c0003 c0004 c0005 c0006 c0007 c0008 c0009 c0010 c0011	No Hits Found No Hits Found 20539Md/46666[vslid]66503[ssid]putallve_trans	(125 letters) Database: /home/jsmith/VirulenceDB/BlastDB/virulenceDB, protein.fasta 55,193 sequences; 10,514,445 total letters Searching
c0003 c0004 c0005 c0006 c0007 c0008 c0009 c0009 c0010 c0011 c0012	No Hits Found No Hits Found	C(125 letters) Database: /home/jsmith/irulenceDB/BlastDB/irulenceDB.protein.fasta 55,193 sequences; 10,514,445 total letters Searching

2.7 Step 6

On Step 6, you may choose the e-value for reciprocal blast analyses between the query and subject genomes. The standard value is 0.000001. After choosing the e-value, run the analyses and click visualize in order to see the results.

Dependencie				GIPSy: (Genomic Isla	nd Predi	ction So	ftware						-	
- spondonoid	8S														
ep 1 Step 2	Step 3	Step 4 Step 5	Step 6	Step 7 Step 8	;										
erform recip	rocal Blasts	between ger	nomes												
Users\Siomar\	\GIP Sy\Depend	encies\blast\bir	\blastall.exe	•											
erform reciproc	cal BLASTp an	alyses between	CDSs of que	ery and subject ge	nomes 0.00	001 R	un	Visualize							
Query ID			Subject	ID		%ID	Alignm.	Mismat	Gap O.	Q.Start	Q.End	S.Start	S.End	E-valu	e bit so
						_									
						_									
						_									
						_									
						_							-		
						_	-								
						_									
						_									
< Previous															Ne
- Tronouo															
				GIPSy: (Genomic Isla	nd Predi	ction So	ftware						_	
Dependencie	es			GIPSy: (Genomic Isla	nd Predi	ction So	ftware						-	
Dependencie	es Step 3 1	Step 4 / Step 5	Step 6	GIPSy: (Step 7 / Step 8	Genomic Isla	nd Predi	ction So	ftware						-	
Dependencie ep 1 Step 2	es Step 3 !	Step 4 Y Step 5	Step 6	GIPSy: (Step 7 Step 8	Genomic Isla	nd Predi	ction So	ftware						-	
Dependencie ep 1 / Step 2 erform recipi Users\Siomar\	es Step 3 Step 5	Step 4 / Step 5 between ger lencies\blast\bir	Step 6 Tomes	GIPSy: (Step 7 Step 8	Genomic Isla	nd Predi	ction So	ftware						-	
Dependencie ep 1 Step 2 erform recipi Users\Siomar\ inform recipirod	es Step 3 5 rocal Blasts (GIP Sy\Depend cal BLASTp an	Step 4 / Step 5 between ger enciesüblastübir alyses between	Step 6 Nomes Nblastall.exe	GIPSy: (Step 7 Step 8 e ery and subject ge	Genomic Isla	nd Predi	ction So	ftware Visualize				Runn	ng! Ple	- ase wai	. -
Dependencie ep 1 Step 2 erform recipi Users\Siomar\ erform reciproc Query ID	es Step 3 1 rocal Blasts (GIP SylDepend cal BLASTp an	Step 4 / Step 5 between ger lencies\blast\bin alyses between	Step 6 Tomes Vblastall.exe CDSs of que Subject II	GIPSy; (Step 7 / Step 8 e ery and subject ge D	Genomic Isla	nd Predi	ction So un	ftware Visualize Mismat	Gap O	Q.Start	Q.End	Runn S.Start	ng! Ples	ase wai E-value	t! bit scor
Dependencie ep 1 Step 2 erform recipi Users\Siomar\ rform recipiroo Query ID	es Step 3 1 rocal Blasts VGIP Sy\Depend cal BLASTp an	Step 4 Step 5 between ger enciesüblastübir alyses between	Step 6 Tomes Ublastall.exe CDSs of que Subject II	GIPSy: (Genomic Isla	nd Predi 001 R	un Alignim	ftware Visualize Mismat.	Gap O	Q.Start	Q.End	Runn S.Start	ng! Ple S.End	ase wai E-value	t! bit scor
Dependencie 2p 1 Step 2 2rform recipi Users\Siomar\ rform reciproc Query ID	es Step 3 ? rocal Blasts VGIP SylDepend cal BLASTp an	Step 4 Step 5 between ger denciesüblastübir alyses between	i Step 6 iomes iblastall.exe CDSs of que Subject II	GIPSy: (Genomic Isla	nd Predi	un Alignm	ftware Visualize	Gap O	Q.Start	Q.End	Runn S.Start	ng! Plex S.End	ase wait	t! bit scor
Dependencie ap 1 Step 2 erform recipionari rform recipionari Query ID	es Step 3 1 Step 3 1 Step 3 1 Step 3 1 Step 3 Step 3 1 Step	Step 4 Step 5 between ger ienciesüblastübir alyses between	i Step 6 nomes Vblastall.exe CDSs of que Subject II	GIPSy: (Step 7 Step 8 ery and subject ge D	nomes 0.00	nd Predi	un Alignm	ftware Visualize Mismat	Gap O	Q.Start	Q.End	Runn S.Start	ng! Ple S.End	ase wai E-value	t! bit scor
Dependencie sp 1 Step 2 erform recipio Users\Siomar' rform recipioc Query ID	es Step 3 1 1 rocal Blasts KGIP SylDepend cal BLASTp an	Step 4 Step 5 between ger enciesüblastibir alyses between	Step 6 Nomes Waastall.exe CDSs of que Subject II	GIPSy: (Step 7 Step 8 ery and subject ge D	nomes 0.00	nd Predi 001 R %ID	un Alignm	ftware Visualize	Gap O	Q.Start	Q.End	Runn S.Start	ng! Plex	ase wai	t! bit scor
Dependencie sp 1 Step 2 erform recipio Users\Siomar' rform recipioc Query ID	es Step 3 1 1 rocal Blasts KGIP SylDepend cal BLASTp an	Step 4 / Step 5 between ger encies/blastbir alyses between	Step 6 Omes Wastall.exe CDSs of que Subject II	GIPSy: (Step 7 Step 8 ery and subject ge D	Genomic Isla	001 R 961D	un Alignm	ftware Visualize Mismat	Gap O	Q.Start	Q.End	Runn S.Start	ng! Ple S.End	ase wai	ti bit scor
Dependenck ep 1 Step 2 erform recipio ulsers\Siomari urform recipio Query ID	es Step 3 1 9 rocal Blasts VGIP SylDepend cal BLASTp an	Step 4 / Step 5 between ger encies/blast/bir alyses between	Step 6 Nomes Waastall.exe CDSs of que Subject II	GIPSy: (Step 7 Step 8 ery and subject ge D	Genomic Isla	nd Predi	un Alignm	ftware Visualize Mismat	Gap O	Q.Start	Q.End	Runn S.Start	ng! Ple S.End	ase wai	t! bit scor
Dependencic ep 1 / Step 2 erform recipro erform reciproc Query ID	es Step 3 1 1 rocal Blasts GIP SylDepend cal BLASTp an	Step 4 / Step 5 between ger enciesblastive alyses between	Step 6 Nomes Watastall.exe CDSs of que Subject II	GIPSy: (Genomic Isla	nd Predi	un Alignm	ftware Visualize Mismat.	Gap O	Q.Start 	Q.End	Runn S Start	ng! Plex S.End	ase wai	t!
Dependencia ep 1 / Step 2 erform recipto Guery ID	es Step 3 1 rocal Blasts (GP SylOpenda al BLASTp an	Step 4 Step 5 between ger encies/blast/bir alyses between	Step 6 homes Vblastall.exe CDSs of que Subject II	GIPSy: (Step 7 Step 6 ery and subject ge D	nomes 0.00	nd Predi	un Alignm	ftware Visualize	Gap O	Q.Start	Q.End	S Start	ng! Ple	E-value	ti bit scor
Dependencia ep 1 Step 2 erform recipion Ulsers\Siomar\ Ulsers\Siomar\ Cuery ID	es Step 3 1 1 rocal Blasts GIP SylDepend cal BLASTp an	Step 4 ⁷ Step 5 between ger encies/blast/bir alyses between	Step 6 Iomes Iblastall.exe CDSs of que Subject II	GIPSy: (Step 7 Step 8 erry and subject ge D	Senomic Isla	nd Predi	un Alignm	Visualize	Gap 0	Q.Start 	Q.End	Runn S Start	ng! Ple S.End	ase wai	tt bit scor
Dependencie ep 1 Step 2 erform recipi Users\Siomari utform reciproc Query ID	es Step 3 7 3 rocal Blasts KGIP SyDepend cal BLASTP an	Step 4 Step 5 between ger between seine solution alyses between	Step 6 Domes Viblastall.exe CDSs of que Subject li	GIPSy: ([^] Step 7 [^] Step 8 erry and subject ge D	Genomic Isla	nd Predi	un Alignim	Visualize	Gap 0	Q.Start	Q.End	Runn S Start	ng! Ple S.End	ase waii	tt bit scor
Dependencié ep 1 Step 2 erform recipion Users\Siomar' utsers\Siomar' Query ID	es Step 3 / 1 rocal Blasts KGIP SylDepend cal BLASTp an	Step 4 / Step 5 between ger enciesübastibu alyses between	Step 6 Nomes Ublastall.exe CDSs of que Subject II	GIPSy: (Step 7 Step 8 ery and subject ge D	Genomic Isla	001 R 960	un Alignim	ftware Visualize	Gap O	Q.Start	Q.End	Runn S.Statt	ingi Ple	ase wal	tt bit scor
Dependencié sp 1 Step 2 erform recipion utsers(Slomartsers) rform recipion Query ID	es Step 3 1 2 rocal Blasts GGP SyDepend cal BLASTp an	Step 4 / Step 5 between ger enciesblastbild alyses between	i Step 6 homes Ublastall.exe CDSs of que Subject II	GIPSy: (Step 7 Step 8 ery and subject ge D	Genomic Isla	001 R 991D	un Alignim	Visualize Mismat.	Gap O	Q.Start	Q.End	Runn S Start	ing! Ples	E-value	tt bit scor
Dependencia p 1 Step 2 erform reciproc Query ID	es Step 3 [*] rocal Blasts (GIP Syl0epend cal BLASTp an	Step 4 / Step 5 between ger encies/blast/bit alyses between	i Step 6 Iomes Walastall.exe CDSs of que Subject II	GIPSy: (nomes 0.00	nd Predi	un Alignm.	ftware Visualize	Gap O	Q.Start	Q.End	Runn S Start	ngt Ple S.End	======================================	ti bit score

	GIPSy: Genomic Isla	nd Pred	iction S	oftware								>
le Dependenc	ies											
Step 1 Step 3	2 Step 3 Step 4 Step 5 Step 6 Step 7 Step 8											
Perform reci	procal Blasts between genomes											
C:\Users\Sioma	r\GIPSv\Dependencies\blast\bin\blastall.exe											
					_		_					
Perform recipro	ocal BLASTp analyses between CDSs of query and subject genomes 0.000	001	Run	Visualize	•				Don	e!		
Query ID	Subject ID	%ID	Alignm.	Mismat	Gap O	Q.Start	Q.End	S.Start	S.End	E-va	bit scor	re
b0135	c0166 R 160235:161524 Hypothetical fimbrial-like protein vadC	61%	422	151	11	1	412	22	429	e-180 W	505	٦.
b0171	c0207 F 202030:202755 Uridylate kinase	100%	241	0	7	1	241	22	241	e-180	490	
b0213	c0250 F 246421:247161 Hypothetical protein vafS	99%	240	2	1	1	240	22	246	e-180	492	Т
b0851	c0984 F 945559:946281 Oxygen-insensitive NADPH nitroreductase	98%	240	4	1	1	240	22	240	e-180	491	1
01120	c1395 F 1323747:1324568 CobB protein	99%	236	1	9	1	236	22	273	e-180	492	1
01606	c1998 F 1841671:1842393 Hypothetical oxidoreductase ydgB	98%	240	4	55	1	240	22	240	e-180	491	1
b1950	c2367 F 2174662:2175447 Flagellar biosynthetic protein fliR	98%	261	5	1	1	261	22	261	e-180	494	1
b2144	c2677 F 2524573:2525292 SanA protein	100%	239	0	1	1	239	22	239	e-180	491	1
b2232	c2774 F 2626420:2627142 3-demethylubiquinone-9 3-methyltransferase	98%	240	3	8	1	240	22	240	e-180	492	1
b2945	c3531 F 3386299:3387009 Endonuclease I precursor	98%	235	3	18	1	235	22	236	e-180	490	1
b2958	c3544 R 3396763:3397482 Hypothetical protein yggN	100%	239	0	18	1	239	22	239	e-180	492	Т
b3141	c3898 F 3725892:3726659 Putative galactosamine-6-phosphate	95%	250	11	3	1	250	22	250	e-180	492	1
03208	c3968 R 3782573:3783301 Monofunctional biosynthetic peptidoglycan	98%	242	3	18	1	242	22	242	e-180	491	Т
b3324	c4095 F 3886957:3887772 Probable general secretion pathway protein	96%	271	10	1	1	271	22	271	e-180	494	٦
b3718	c4640 R 4399800:4400522 Hypothetical protein yieK	98%	240	3	2	1	240	22	240	e-180	491	Т
b3724	c4648 R 4407978:4408721 Phosphate transport system protein phoU	99%	241	1	1	1	241	22	247	e-180	492	
04334	c5417 R 5157844:5158617 Hypothetical protein yjiL	95%	254	12	7	1	254	22	256	e-180	492	
b0306	c0422 F 406102:406821 Hypothetical protein ykgE	99%	239	2	2	1	239	22	239	e-179	489	
0524	c0639 R 622276:622998 UDP-2,3-diacylglucosamine hydrolase	98%	240	4	1	1	240	22	240	e-179	488	
b0652	c0736 R 716589:717314 Glutamate/aspartate transport ATP-binding	98%	241	4	13	1	241	22	241	e-179	489	
p1194	c1644 R 1480891:1481625 Hypothetical protein ycgR	96%	244	8	3	1	244	22	244	e-179	488	
												_
< Previous											Next	>

2.8 Step 7

On Step 7, you may choose the e-value for tRNA prediction using HMMer. The standard value is 0.0001. After choosing the e-value, run the analyses and click visualize in order to see the results.



6				GIPSy: Ger	nomic Island Prediction Software	- 🗆 🗙
ile Depend	tencies					
Step 1 S	tep 2 Step 3	Step 4 St	ep 5 Step 6	Step 7 Step 8		
Search fo	or tRNA genes					
C:\Users\Si	omar\GIP Sy\Depen	dencies\hmr	ner\hmmsearch	.exe		
0 C					Densel	
Search for	tKNA genes on que	ery genome	0.0001	un visualize	Done:	
#Parsed Fil	le			15		·
1	tRNA	236817	236889	Forward		
2	tRNA	236911	236976	Forward		
3	tRNA	240329	240405	Forward		
4	tRNA	248554	248630	Forward		
5	tRNA	370224	370299	Forward		-
6	tRNA	627071	627147	Forward		
7	Broken-tRNA	630476	630521	Forward		
8	tRNA	803225	803300	Forward		
9	tRNA	803336	803411	Forward		
10	tRNA	803414	803489	Forward		
11	tRNA	803541	803616	Forward		
12	tRNA	803620	803695	Forward		
13	tRNA	803744	803819	Forward		
14	tRNA	803853	803928	Forward		
15	tRNA	1211153	1211228	Forward		
16	tRNA	1211232	1211304	Forward		
17	tRNA	1342304	1342379	Forward		
18	tRNA	1342388	1342463	Forward		
19	tRNA	1342477	1342553	Forward		
20	tRNA	1902573	1902649	Forward		_
21	tRNA	1902654	1902730	Forward		
< Previo	us					Next >
						L

2.9 Step 8

Finally, on Step 8, all the data from the previous steps are used to predict the chosen genomic islands of the given organism. Run the analyses for "Pathogenicity Islands" and click visualize in order to see the results.

				GIPSy:	Genomic	sland Prediction Software					
ile Dependencies											
Step 1 Step 2 Step 3	Step 4 St	en 5 Ster	n 6 Sten	7 Step	R						
Predict Genomic Islands based on previously identified features											
Pathogenicity Islands	Run	Visualize		Runnin	g <mark>! Plea</mark> se	wait!					
Resistance Islands Run Visualize											
Metabolic Islands Run Visualize											
FIELEDOIIC ISIAITUS	Dun	Minualiza	-								
Symbiotic Islands	KUII	visualize	_								
Putative Genomic Is	land	G+C Devi	Codon U	Specific	Hypotheti	Gene Composition	Position	Predictio			
1											
le Dependencies				GIPSy:	Genomic	sland Prediction Software					
le Dependencies Step 1 Step 2 Step 3 Decedict Concernic Labord	Step 4 St	ep 5 Stej	p6 Step	GIPSy:	Genomic B	sland Prediction Software					
 e Dependencies Step 1 Step 2 Step 3 Predict Genomic Islands	Step 4 St	ep 5 Stej	p6 ∕Step ridentifie	GIPSy: 7 Step d feature	Genomic B	sland Prediction Software					
ie Dependencies Step 1 Step 2 Step 3 Predict Genomic Islands Pathogenicity Islands	Step 4 St s based on p Run	ep 5 Ster previously Visualize	p 6 Step v identifie	GIPSy: 7 Step : d feature	Genomic ⁸ 5 Done!	sland Prediction Software					
le Dependencies Step 1 Step 2 Step 3 Predict Genomic Islands Pathogenicity Islands Resistance Islands	Step 4 St based on Run Run	ep 5 Stej previously Visualize Visualize	p 6 Step v identifie	GIPSy: 7 Step f	Genomic 8 5 Done!	sland Prediction Software					
ie Dependencies Step 1 Step 2 Step 3 Predict Genomic Islands Pathogenicity Islands Resistance Islands	Step 4 St based on p Run (Run (Run (ep 5 / Step previously Visualize Visualize	p 6 Step v identified	GIPSy: 7 _ Step : d feature	Genomic ⁸ 5 Done!	sland Prediction Software		- 0			
e Dependencies tep 1 Step 2 Step 3 Predict Genomic Islands Pathogenicity Islands Resistance Islands Aetabolic Islands	Step 4 St based on Run Run Run	ep 5 / Step previously Visualize Visualize Visualize	p 6 Step videntifier	GIPSy: 7 Step a d feature	Genomic ⁸ s Done!	sland Prediction Software					
e Dependencies tep 1 Step 2 Step 3 Predict Genomic Islands Pathogenicity Islands Resistance Islands Metabolic Islands Symbiotic Islands	Step 4 St based on [Run [Run [Run [Run [ep 5 / Step previously Visualize Visualize Visualize Visualize	p 6 Step r identifier	GIPSy: 7 Step : d feature	Genomic 8 5 Done!	sland Prediction Software					
Dependencies tep 1 Step 2 Step 3 redict Genomic Islands rathogenicity Islands Resistance Islands Metabolic Islands Symbiotic Islands Putalive Genomic Islands Putalive Genomic Islands	Step 4 St based on 1 Run [Run [Run [Run [iland	ep 5 Step previously Visualize Visualize Visualize Visualize G+C Devi.	p 6 Step r identifier	GIPSy: 7 Step : d feature	Genomic B S Done! Hypotheti	sland Prediction Software	Position	- D			
e Dependencies tep 1 Step 2 Step 3 Predict Genomic Islands Pathogenicity Islands Resistance Islands Metabolic Islands Symbiotic Islands Putative Genomic Island Senome Putative Genomic Island 1	Step 4 St based on 1 Run [Run [Run [Run] Run]	ep 5 Step previously Visualize Visualize Visualize Visualize G+C Devi 10% 66%	p 6 Step v identifies Codon U 24%	GIPSy: 7 Step d feature	Genomic B T Done! Hypotheti 41% 66%	sland Prediction Software Gene Composition NA c0133-c0139	Position NA 126688.132406	Predictio.			
e Dependencies tep 1 Step 2 Step 3 Predict Genomic Islands Pathogenicity Islands Resistance Islands Metabolic Islands Putative Genomic Island 1 Putative Genomic Island 1 Putat	Step 4 St Stased on 1 Run [Run [Run [Run [Iand	ep 5 Step previously Visualize Visualize Visualize Visualize G+C Devi 10% 66% 21%	p 6 Step y identifier Codon U 24% 100% 66%	GIPSy: 7 Step 1 d feature Virulence 44%	Genomic 8 5 Done! Hypotheti 41% 66% 76%	Sland Prediction Software Gene Composition NA c0133-c0139 c0253-c0386	Position NA 126688.132406 248421.348625	Predictio NA NA Strong			
Dependencies Tep 1 Step 2 Step 3 Predict Genomic Islands Pathogenicity Islands Resistance Islands Atetabolic Islands Putative Banome Putative Genomic Island 1 Putative Pathogenicity	Step 4 St based on 1 Run [Run [Run] iland	ep 5 Step Visualize Visualize Visualize Visualize Visualize Visualize 21% 66% 21% 37% 60%	p 6 Step y identifier codon U 24% 100% 66% 37%	GIPSy: 7 Step : d feature: 44% 33% 64% 75%	Genomic 8 5 Done! Hypotheti 41% 66% 76% 50% 232	Sland Prediction Software	Position NA 126688.132406 248421.348625 370109.378117 296072.96714	Predictio NA NA Strong Normal			
	Step 4 St Step 4 St Standard Standard	ep 5 Step previously Visualize Visualize Visualize Visualize G+C Devi 10% 66% 21% 37% 0%	p 6 Step r identifier Codon U 24% 100% 66% 37% 66% 37%	GIPSy: 7 Step 1 d feature: 44% 33% 64% 75% 100%	Genomic 8 5 Done! Hypotheti 41% 66% 76% 50% 33% 41%	Gene Composition NA c0133-c0139 c0253-c0368 c0291-c0398 c0409-c0414 c0402-c0414 c0402-c0414 c0402-c0414 c0402-c0414	Position NA 126688.132406 248421.348625 370109.378117 388978.395711 908853 342509	Predictio NA NA Strong Normal Normal			
	Step 4 St Step 4 St Standard St Standard St Standard St St St St St St St St St St	ep 5 Step previously Visualize Visualize Visualize Visualize Visualize S+C Devi 10% 66% 21% 37% 0% 20% 13%	p 6 Step v identifier v iden	GIPSy: 7 Step 1 d feature: Virulence 44% 75% 100% 77% 100%	Genomic B Hypotheti Hypotheti Hypotheti S Done! Hypotheti S S S S S S S S S S S S	Sland Prediction Software Gene Composition N4 c0133-c0139 c0253-c0386 c0391-c0396 c0391-c039 c0391-c03 c0391-c039 c039 c0391-c039	Position NA 126688.132406 248421.346625 370109.378117 388978.395711 908853.942509 1127423.1241384	Predictio NA Strong Normal Normal Strone			
Dependencies tep 1 Step 2 Step 3 redict Genomic Islands athogenicity Islands esistance Islands putative Genomic Islands Putative Genomic Island Putative Genomic Island 1 utative Pathogenicity Island 1	Step 4 St based on 1 Run (Run (Run (iland	ep 5 Step previously Visualize Visualize Visualize G+C Devi 10% 66% 21% 37% 0% 20% 13%	p 6 Step r identified Codon U 24% 100% 66% 37% 66% 37% 66% 63% 37%	GIPSy: 7 Step 1 d feature: 44% 33% 5% 100% 77% 100% 67%	Genomic 8 Done! Hypotheti 41% 66% 76% 50% 33% 41% 59% 57%	Sland Prediction Software Gene Composition NA co133-c0139 co253-c0388 co391-c0388 co409-co414 co392-c0392 c1485-c1293 c1400-c1475	Position NA 126688.132406 248421.348625 370109.378117 388978.395711 908853.342509 1127423.1241384 1327847.1373052	Predictio NA NA Strong Normal Normal Normal Normal Strong Strong			
Dependencies tep 1 Step 2 Step 3 redict Genomic Islands athogenicity Islands tesistance Islands tetabolic Islands putative Genomic Island Tutative Genomic Island tutative Genomic Island tutative Pathogenicity Island	Step 4 St sbased on p Run [Run [Run [Run [Run] I I I I I I I I I I I I I	ep 5 Step previously Visualize Visualize Visualize Visualize Visualize Visualize G+C Devi 10% 66% 21% 21% 21% 21% 13% 21%	p 6 Step r identifier Codon U 24% 100% 66% 37% 66% 37% 66% 72% 59%	GIPSy: 7 Step 3 d feature 44% 55% 100% 77% 100% 67%	Genomic 8 5 Done! Hypotheti 41% 66% 50% 33% 41% 59% 57% 57%	Gene Composition NA 00133-00139 00253-00388 00391-00398 00392-00398 00409-00414 00392-00479 0165-01293 01409-01475 01409-01475	Position NA 126688.132406 248421.348625 370109.378117 288978.395711 908853.942509 1127423.1241384 1327847.1373052	Predictio NA NA Strong Strong Strong Strong Strong NA			
	Step 4 St Step 4 St Run [Run] Run [Run] St St St St St St St St St St St St St	ep 5 Step previously Visualize Visualize Visualize Visualize Visualize Visualize (+C Devi 10% 66% 21% 21% 13% 20% 13% 21%	p 6 Step r identified 24% 100% 66% 37% 66% 37% 59% 59% 58%	GIPSy: 7 Step 1 d feature 44% 33% 75% 100% 67% 14% 67%	Genomic 8 5 Done! Hypotheti 41% 66% 76% 50% 33% 41% 59% 57% 39%	Sland Prediction Software Gene Composition NA 00133-c0139 00253-c0386 00391-c0388 00391-00388 00391-00388 00409-c0414 00322-20879 c1405-c1293 c1400-c1475 c1485-c1507 c1451-c1507 c155 c155 c155 c155 c155 c155 c155 c15	Position NA 126638.132406 248421.348625 370109.378117 388978.395711 990853.942509 1127423.1241384 1327847.1373052 1377690.1382209 1397769.1452493	Predictio NA NA Strong Normal Normal Strong Strong Strong NA NA			
	Step 4 St based on p Run (Run (Run) iland	ep 5 Step previously Visualize Visualize Visualize Visualize Visualize G+C Devi 10% 66% 21% 13% 21% 13% 21% 53%	p 6 Step r identified Codon U 24% 100% 66% 37% 66% 37% 66% 37% 66% 59% 59% 59% 100%	GIPSy: 7 Step 1 d feature: 44% 33% 64% 75% 100% 77% 100% 77% 100% 15% 15%	Genomic B Done! Hypotheti 41% 66% 50% 33% 41% 59% 57% 57% 59% 100% 100%	Gene Composition NA 0133-0139 00253-0388 00391-00388 00391-00388 00391-00388 00392-0079 01165-01293 00392-0079 01165-01293 01400-01475 01400-01475 01400-01475 01400-01475 01400-01875 01400-01875 01400-01875 01400-01875 01400-01875 01400-01875 01400-0187 000 000 000 000 000 000 000 000 000 0	Position NA 12688.132406 244241.348625 2370109.378117 388078.395711 908853.342509 1127423.1241384 1327847.13052 1377690.1389209 1395726.1452493 11375525.1728047	Predictio NA NA Normal Normal Normal Normal Strong Strong Strong NA NA NA			
e Dependencies tep 1 Step 2 Step 3 redict Genomic Islands Pathogenicity Islands Resistance Islands Metabolic Islands Putalive Genomic Island Putalive Genomic Island Putalive Pathogenicity Island Putalive Pathogenicy Island Putalive Pathogenicy Island Putalive Pathogenicy Island Putalive Pathogenicy Island	Step 4 St sbased on p Run [Run] Run [Run] land 1 2 3 4 5 5 7 3	ep 5 Step previously Visualize Visualize Visualize Visualize Visualize C+C Devi 10% 66% 21% 21% 20% 13% 20% 13% 21% 13% 20% 13% 21% 33%	p 6 Step y identified Codon U 24% 66% 37% 66% 37% 66% 37% 68% 72% 58% 100% 58%	GIPSy: 7 Step : d feature 44% 5% 100% 77% 100% 67% 100% 67% 100% 77% 100% 77% 100% 77%	Genomic 8 5 Done! Hypotheti 41% 66% 76% 33% 41% 59% 57% 74% 39% 100% 44% 0%	Gene Composition NA OD33-00139 OD33-00139 OD39-00398 OD409-00414 OD392-00398 Od409-00414 C0392-00398 C1400-01475 C1405-C1293 C1400-01475 C1405-01475 C1405-0145 C1405-015 C	Position NA 126588.132406 24421.348625 370109.378117 388978.395711 990835.342509 1127423.241384 1327847.1373052 1377690.1241384 1327847.1373052 1397526.1452493 11715525.128047 11715525.128047	Predictio. NA NA Strong Strong Strong Strong Strong NA Normal NA Normal NA			
le Dependencies Step 1 Step 2 Step 3 Predict Genomic Islands Pathogenicity Islands Resistance Islands Metabolic Islands Putalive Genomic Island 1 Putalive Cenomic Island 1 Putalive Pathogenicity Island 2 Putalive Pathogenicity Island 1 Putalive Pathogenicity Island 2 Putalive Pathogenicity Island 2 Putalive Pathogenicity Island 2 Putalive Pathogenicity Island 2 Putalive Pathogenicity Island 3 Putalive Pathogenicity Island 3 Putalive Pathogenicity Island 3 Putalive Pathogenicity Island 3 Putalive Pathogenicity Island 4 Putalive Genomic Island 5 Putalive Genomic Island 4 Putalive Genomic Island 5 Putalive Genom	Step 4 St Step 4 St Standard Standard	ep 5 Step previously Visualize Visualize Visualize Visualize Visualize S+C Devi 10% 66% 21% 37% 0% 20% 13% 53% 0% 53% 0% 53% 0% 53%	p 6 Step r identifier Codon U 24% 100% 66% 37% 66% 37% 66% 100% 58% 100% 58% 100% 58%	GIPSy: 7 Step 1 d feature 44% 33% 64% 75% 100% 77% 100% 67% 15% 77% 16% 91%	Genomic 8 5 Done! Hypotheti 41% 66% 66% 66% 66% 59% 57% 59% 57% 74% 39% 100% 44% 0% 63%	Gene Composition N4 C0132-C0139 C0253-C036 C0391-C0398 C0291-C0398 C0291-C1475 C1480-C1475 C1481-C1507 C1515-C1802 C1685-C1803 C1935-C1643 C1935-C164 C1935-C1643 C1935-C164 C1955-C164 C1955-C164 C1955-C164 C1955-C164 C1955-C164 C1955-C164 C1955-C164 C1955-C164 C1955	Position NA 12668.122406 248421.348625 370109.378117 388978.395711 990853.942509 1127423.241384 1327847.1373052 1127423.241384 132769.1452493 1715525.1728047 1715525.1728047 17177183.1787317 1796112.1802390 2193645.255680	Predictio NA NA Strong Normal Strong Strong Strong Normal Normal Normal Normal Na Normal Na Na Na Na Na Strong Strong			

3. Saving results and exporting additional analyses

After all the analyses have been performed, the results may be saved by clicking "File \rightarrow Save Results \rightarrow PAI Results".

le Dependencies									
ave Results + PAI Results		- 5 V 01-1		7 01					
xport 🕨 RI Results	step 4 Ste	p 5 Step	56 Step	/ Step a	s				
lose MI Results	bacad on n	roviouch	idantifia	1 foaturo					
Si Realt									
Sincound									
Pathogenicity Islands	Run	Visualize Done!							
Resistance Islands	Run	Visualize							
			_						
Metabolic Islands	Run	Visualize							
Symbiotic Islands	Run	visualize							
Putative Genomic Is	sland	G+C Devi	Codon U	Virulence	Hypotheti	Gene Composition	Position	Predictio	
Conomo		10%	24%	44%	41%	NA	NA	NA	
Genome									
Putative Genomic Island 1		66%	100%	33%	66%	c0133-c0139	126688132406	NA	
Putative Genomic Island 1 Putative Pathogenicity Island	1	66% 21%	100% 66%	33% 64%	66% 76%	c0133-c0139 c0253-c0368	126688132406 248421348625	NA Strong	
Putative Genomic Island 1 Putative Pathogenicity Island Putative Pathogenicity Island	1 2	66% 21% 37%	100% 66% 37%	33% 64% 75%	66% 76% 50%	c0133-c0139 c0253-c0368 c0391-c0398	126688132406 248421348625 370109378117	NA Strong Normal	
Putative Genomic Island 1 Putative Pathogenicity Island Putative Pathogenicity Island Putative Pathogenicity Island	1 2 3	66% 21% 37% 0%	100% 66% 37% 66%	33% 64% 75% 100%	66% 76% 50% 33%	c0133-c0139 c0253-c0368 c0391-c0398 c0409-c0414	126688.132406 248421.348625 370109.378117 388978.395711	NA Strong Normal Normal	
Putative Genomic Island 1 Putative Pathogenicity Island Putative Pathogenicity Island Putative Pathogenicity Island Putative Pathogenicity Island	1 2 3 4	66% 21% 37% 0% 20%	100% 66% 37% 66% 37%	33% 64% 75% 100% 77%	66% 76% 50% 33% 41%	c0133-c0139 c0253-c0368 c0391-c0398 c0409-c0414 c0932-c0979	126688132406 248421348625 370109378117 388978395711 908853942509	NA Strong Normal Normal Normal	
Putative Genomic Island 1 Putative Pathogenicity Island Putative Pathogenicity Island Putative Pathogenicity Island Putative Pathogenicity Island	1 2 3 4 5	66% 21% 37% 0% 20% 13%	100% 66% 37% 66% 37% 68%	33% 64% 75% 100% 77% 100%	66% 76% 50% 33% 41% 59%	c0133-c0139 c0253-c0368 c0391-c0398 c0409-c0414 c0932-c0979 c1165-c1293	126688.132406 248421.348625 370109.378117 388978.395711 908853.942509 1127423.1241384	NA Strong Normal Normal Strong	
Putative Genomic Island 1 Putative Pathogenicity Island Putative Pathogenicity Island Putative Pathogenicity Island Putative Pathogenicity Island Putative Pathogenicity Island	1 2 3 4 5 6	66% 21% 37% 0% 20% 13% 21%	100% 66% 37% 66% 37% 68% 72%	33% 64% 75% 100% 77% 100% 67%	66% 76% 50% 33% 41% 59% 57%	c0133-c0139 c0253-c0388 c0391-c0398 c0409-c0414 c0932-c0979 c1165-c1293 c1400-c1475	126688.132406 248421.348625 370109.378117 388978.395711 908853.942509 1127423.1241384 1327847.1373052	NA Strong Normal Normal Strong Strong	
Vatative Genomic Island 1 Putative Pathogenicity Island Putative Pathogenicity Island Putative Pathogenicity Island Putative Pathogenicity Island Putative Pathogenicity Island Putative Genomic Island 2	1 2 3 4 5 6	66% 21% 37% 0% 20% 13% 21% 14%	100% 66% 37% 66% 37% 68% 72% 59%	33% 64% 75% 100% 77% 100% 67% 14%	66% 76% 50% 33% 41% 59% 57% 74%	c0133-c0139 c0253-c0368 c0391-c0398 c0409-c0414 c0932-c0979 c1165-c1293 c1400-c1475 c1481-c1507	126988.132406 244421.348625 370109.378117 388978.395711 908853.942509 1127423.1241384 1327847.1373052 1377690.1389209	NA Strong Normal Normal Strong Strong NA	
Vatative Genomic Island 1 Putative Pathogenicity Island Putative Pathogenicity Island Putative Pathogenicity Island Putative Pathogenicity Island Putative Pathogenicity Island Putative Genomic Island 2 Putative Genomic Island 2	1 2 3 4 5 6 7	66% 21% 37% 0% 20% 13% 21% 14% 14%	100% 66% 37% 66% 37% 68% 72% 59% 58%	33% 64% 75% 100% 77% 100% 67% 14% 62%	66% 76% 50% 33% 41% 59% 57% 74% 39%	0/133-0/139 0/233-0/388 0/331-0/398 0/409-0/114 0/409-0/114 0/409-0/114 0/1400-0/1475 0/1400-0/1475 0/140-0/1475	126688, 132406 248421, 348625 370109, 378117 388978, 395711 90683, 342509 1127423, 1241384 1327694, 1373062 1377690, 1389209 1397780, 1452483	NA Strong Normal Normal Strong Strong NA Normal	
Varlative Genomic Island 1 Putative Pathogenicity Island Putative Pathogenicity Island Putative Pathogenicity Island Putative Pathogenicity Island Putative Pathogenicity Island Putative Genomic Island 2 Putative Pathogenicity Island Putative Cenomic Island 3	1 2 3 4 5 6 7	66% 21% 37% 0% 20% 13% 21% 14% 17% 53%	100% 66% 37% 66% 37% 68% 72% 59% 58% 100%	33% 64% 75% 100% 77% 100% 67% 14% 62% 15%	66% 76% 50% 33% 41% 59% 57% 74% 39% 100%	cb133-cb139 cb233-cb388 cb391-cb398 cb409-cb144 cb932-cb79 c1185-c1293 c1400-c1475 c1401-c1507 c1481-c1507 c1515-c1602 c1581-c1893	126688.122406 224421.346825 370109.378117 308078.396711 90883.942509 1127423.1241384 1127429.1241384 11377800.1308209 11395728.1452493 1175528.1452493	NA Strong Normal Normal Strong Strong NA Normal NA	
Variative Genomic Island 1 Putative Pathogenicity Island Putative Pathogenicity Island Putative Pathogenicity Island Putative Pathogenicity Island Putative Pathogenicity Island Putative Pathogenicity Island Putative Genomic Island 2 Putative Genomic Island 3 Putative Genomic Island 3	1 2 3 4 5 6 7 8	66% 21% 37% 0% 20% 13% 21% 14% 17% 53% 0%	100% 66% 37% 66% 37% 68% 72% 59% 59% 58% 100% 77%	33% 64% 75% 100% 77% 100% 67% 14% 62% 15% 77%	66% 76% 50% 33% 41% 59% 57% 74% 39% 100% 44%	0133-0139 0023-01398 0023-01398 0039-00414 0032-00799 01485-01293 01495-01475 0149-01475 0149-01475 01516-01602 01516-01602 01516-01893 0133-01943	126688, 132406 248421, 348625 370109, 378117 308078, 35711 908853, 942509 1127423, 1241384 13274847, 1373052 1377690, 1389209 1377690, 1389209 1375728, 1452483 1775528, 1728047 1777183, 1787317	NA Strong Normal Normal Strong Strong NA Normal NA Normal	
Valative Genomic Island 1 Putative Pathogenicity Island Putative Pathogenicity Island Putative Pathogenicity Island Putative Pathogenicity Island Putative Pathogenicity Island Putative Genomic Island 2 Putative Genomic Island 2 Putative Pathogenicity Island Putative Pathogenicity Island Putative Pathogenicity Island	1 2 3 4 5 6 7 8	66% 21% 37% 0% 20% 13% 21% 14% 53% 0% 33%	100% 66% 37% 66% 37% 68% 72% 59% 58% 100% 77% 100%	33% 64% 75% 100% 77% 100% 67% 14% 62% 15% 77% 16%	66% 76% 50% 33% 41% 59% 57% 74% 39% 100% 44% 0%	0133-0139 00233-01398 00391-00398 00409-00414 00409-00414 00409-00414 01405-01293 01400-01475 01408-01475 01408-01475 01481-01693 0135-01943 0135-01943	126688.122406 224421.346825 370109.378117 308078.396711 908853.942509 1127423.1241384 1137780.1380209 11395728.1452493 1135728.1452493 1175528.1728047 1175528.1728047 1175528.1783.1787317	NA Strong Normal Normal Strong Strong NA Normal NA Normal NA	

Also, the results for each intermediary step may also be exported in "File \rightarrow Export".

4	2						GIPSy:	Genomic	Island Prediction Software		- 🗆 🗙			
F	ile Dependen	cies												
	Save Results >													
	Stop 1 b Oungr Lines Step 5 Step 6 Step 7 Step 8													
	xport /	Step 1 P	Quer	yries v		F (1)								
	lose	e Step 2 Subject riles Subject . Jaa rile ied features												
		Step 3 🕨			Subjectfna	ubjectfna File								
	Pathogenic	Step 4 ►	ls l	Run	Subjectffn	ubjectffn File Done!								
	rutilogenie	Step 5 +	۲. T		Subjectem	Subjectembl File								
	Resistance	Step 6 F		Run	Subject - abl									
		Stop 7 h	· ۱		oubject ign									
	Metabolic Is	dande		Run	Visualize									
	Pictubolic 12	aunus				_								
	Symbiotic Islands Run Visualize													
	Puta	land	G+C Devi.	Codon U	Virulence	Hypotheti	Gene Composition	Position	Predictio					
	Genome		10%	24%	44%	41%	NA	NA	NA 🔺					
	Putative Geno	mic Island	1		66%	100%	33%	66%	c0133-c0139	126688132406	NA			
	Putative Patho	genicity Isl	land 1		21%	66%	64%	76%	c0253-c0368	248421348625	Strong			
	Putative Patho	genicity Isl	land 2		37%	37%	75%	50%	c0391-c0398	370109378117	Normal			
	Putative Patho	genicity Isl	land 3		0%	66%	100%	33%	c0409-c0414	388978395711	Normal			
	Putative Patho	genicity Isl	land 4		20%	37%	77%	41%	c0932-c0979	908853942509	Normal			
Putative Pathogenicity Island 5					13%	68%	100%	59%	c1165-c1293	11274231241384	Strong			
Putative Pathogenicity Island 6					21%	72%	67%	57%	c1400-c1475	13278471373052	Strong			
Putative Genomic Island 2					14%	59%	14%	74%	c1481-c1507	13776901389209	NA			
Putative Pathogenicity Island 7					17%	58%	62%	39%	c1515-c1602	13957261452493	Normal			
	Putative Geno	mic Island	3		53%	100%	15%	100%	c1881-c1893	17155251728047	NA			
	Putative Patho	genicity Isl	land 8		0%	77%	77%	44%	c1935-c1943	17771831787317	Normal			
Putative Genomic Island 4				33%	100%	16%	0%	c1955-c1960	17961121802390	NA				
Putative Pathonenicity Island 9					51%	61%	91%	63%	r2392-r2438	2193646 2250680	Strong			
	< Previous													

4. TroubleShooting

4.1 GIPSy breaks at step 5 on Linux

If GIPSy returns an error message at step 5 on Linux, there is probably a limitation on the number of files submitted to blastp. In order to circumvent this problem, please try increasing the ulimit. In Ubuntu, you must modify /etc/security/limits.conf. In a terminal, type:

sudo gedit /etc/security/limits.conf

Then, before the end of file, add the following lines:

* soft nofile 40000 * hard nofile 40000 # End of file

Also, you need to modify the following just before the end of file:

sudo gedit /etc/pam.d/common-session*

session required pam_limits.so

end of pam-auth-update config

Finally, restart your machine and test your new ulimit typing on the terminal:

ulimit -n

4.2 GIPSy breaks at steps 4 and 7 on Linux

In this version of GIPSy, we used a 32 bits version of hmmer, which prevents it from working properly on some Linux distributions. We are currently working to update the dependencies package; however, this problem may also be circumvented by installing some 32 bits libraries. On Ubuntu, open a terminal and type:

sudo apt-get install libgtk2.0-0:i386 libnss3-1d:i386 libnspr4-0d:i386 lib32nss-mdns libxml2:i386 libstlc++6:i386

4.3 GIPSy breaks at steps 3 to 7 on Linux

Finally, when the files or their paths have spaces in their names, the third-party software may not work properly. We tried to circumvent this the most as possible. However, the command formatdb from blast is still not working fine. So, in order to circumvent this problem, avoid any space in the names of files.

Wrong: Corynebacterium pseudotuberculosis 1002.gbk

Right: Corynebacterium_pseudotuberculosis_1002.gbk

Also, be sure that the complete path does not have any spaces in names.

Wrong: /home/siomar/gipsy folder/Corynebacterium.gbk

Right: /home/siomar/gipsy_folder/Corynebacterium.gbk

5. Acknowledgments

We would like to kindly thank Taryn Takebayashi for sharing her experience with the software and also for providing a valuable feedback on the problems from sections 4.1 and 4.2.