sRNAtoolbox Manual

sRNAtoolbox aims to provide small RNA researchers with several useful tools including sRNA expression profiling from deep sequencing experiments and several downstream analysis tools. The centrepiece of sRNAtoolbox is sRNAbench (the successor program of <u>miRanalyzer</u> which was originally published in 2009), which allows for expression profiling from deep sequencing experiments. Either the other tools can be launched on sRNAbench results, or independently using the appropriate file formats.

Currently the toolbox comprises the following methods:

- sRNAbench
- sRNAde
- sRNAblast
- sRNAcons
- miRNAconsTarget: animalconstarget and plantconstarget

sRNAbench

Profiling of small RNAs and their sequence variants in single or multi-species high-throughput experiments

Description

Several commonly used input formats like fastq, sra, fasta or read/count are supported. Input files can be provided via upload, URL, SRA accession, Google Drive or Dropbox. sRNAbench performs expression profiling of known microRNAs and other types of small RNAs, detection and analysis of isomiRs, outputs several graphical summaries and predicts novel microRNAs in both plants and animals. <u>Result example</u>

Sequencing data input page

SRNADench:	
Profiling of small RNAs and its sequence variants in single or multi-species high-throughput	experiments
Data to upload 🛛 😧	
Add sample(s)	
	NEYT
	NEAT
Add samples to be analyzed	×
Add samples to be analyzed	
Upload file(s) SRA accession(s) Link to file(s) 🛆 Choose from Drive 🕏 Choose from Dropbox	
There are 5 ways you can provide read files:	
Upload a file (typically fastq or fastq.gz)	
 Provide a link/URL with the data Provide an accession for a SRA run (they start with SRR,ERR or DRR e.g. SRR1563062) 	
You can also use files from your Dropbox or Google Drive accounts	
	Close

The sequencing data can be provided in 5 different ways:

- Upload one or several files (typically fastq or fastq.gz)
- Provide one or several link/URL(s) with the data (it will be downloaded and then profiled)
- Provide one/several accession(s) for SRA runs (they start with SRR, ERR or DRR e.g. <u>SRR1563062</u>)
- Choose files from your Google Drive account
- Choose files from your Dropbox account

To add samples, simply click on the desired button and then add the files using the prompt or add the SRA accessions/URLs (one per line) and click on the "ADD" button.

Once all your read files have been provided, you can click "Next" to move on to the parameter specification.

Traditionally, we have supported an unlimited number of samples to be uploaded but we may revisit this policy if some users start engaging in abuse of these conditions.

Example on how to use SRA data

The <u>following link</u> will take us to the SRA project page which contains all information provided by the authors (see below)

12 F	Runs found	b									
010	↓ Run	BioSample	Sample name	DATASTORE provider	DATASTORE location	AvgSpotLen	Experiment	LoadDate	MBases	MBytes	disease
	SRR1563064	SAMN03015564	RN sRNA exosome fraction	ncbi		101	SRX691181	2014-10-03	765	534	
	SRR1563063	SAMN03015563	RN sRNA cell fraction	ncbi		101	SRX691180	2014-10-03	951	666	
	SRR1563062	SAMN03015562	IM-1 sRNA exosome fraction	sra-sos	sra-sos.public	100	SRX691179	2014-09-05	184	109	
	SRR1563061	SAMN03015561	IM-1 sRNA cell fraction	sra-sos	sra-sos.public	100	SRX691178	2014-09-05	773	455	
	SRR1563060	SAMN03015320	IK140508 sRNA exosome fraction	ncbi		101	SRX691177	2014-10-03	597	417	
	SRR1563059	SAMN03015319	IK140508 sRNA cell fraction	ncbi		101	SRX691176	2014-10-03	572	397	
	SRR1563058	SAMN03015312	MUTU I clone 9 sRNA exosome fraction	sra-sos	sra-sos.public	100	SRX691175	2014-09-05	799	464	Burkitt Lymphoma 9
	SRR1563057	SAMN03015311	MUTU I clone 9 sRNA cell fraction	sra-sos	sra-sos.public	100	SRX691173	2014-09-05	472	261	Burkitt Lymphoma 9
	SRR1563056	SAMN03015312	MUTU I clone 9 sRNA exosome fraction	ncbi		100	SRX691171	2014-10-03	1,139	697	Burkitt Lymphoma 9
	SRR1563018	SAMN03015308	MUTU I clone 3 sRNA cell fraction	sra-sos	sra-sos.public	100	SRX691160	2014-09-05	714	399	Burkitt Lymphoma 3
	SRR1563017	SAMN03015315	BJAB sRNA exosome fraction	ncbi		100	SRX691158	2014-10-03	584	301	Diffuse Large B-cell Lymphoma
	SRR1563015	SAMN03015313	BJAB sRNA cell fraction	ncbi		101	SRX691157	2014-10-03	1,284	873	Diffuse Large B-cell Lymphoma

This table can be downloaded and the SRR identifiers can be extracted and pasted into the corresponding textbox (see below).

Input parameters page

Select species annotation:

Choose miRNA annotation reference database	Choose short name from MiRGeneDB
MirGeneDB 2.1 ~	Homo sapiens (Hsa) -
Species (Genome assembly)	
Do not map to genome (Library mode)	
\square Do not profile other ncRNAs (you are interested in known microRNAs only!) 🤪	
Predict New miRNAs	

Alignment of small RNA reads <u>requires</u> at least a miRNA reference annotation or genomic annotations. The left side dropdown selector allows you to choose the miRNA reference database (MirGeneDB or PmiREN are recommended if available), the right side selector allows you to choose one or several species of interest. Then the remaining selector will let you choose genome assembly and annotations from Ensembl for one or several species of interest.

You can decide not to map to the genome (slightly faster but miRNA/sRNA loci cannot always be distinguished), not to profile to other ncRNAs or if you want to predict novel miRNAs from your sequencing data (which requires a genome assembly). This prediction will be performed after all miRNA and other ncRNA reads have been assigned.

Reads preprocessing:

<u>Reads preprocessing</u>: we have implemented five different library preparation protocols for straightforward adapter removal. If your experiment was not designed using these five protocols, custom settings can still be selected. If reads are already trimmed, it can also be specified here.

Optional Parameters

- <u>Quality Control</u>: There are two methods to filter out reads of low quality: *mean-based*, a read is accepted if the mean Phred score is above a given threshold; and *minimum per nucleotide*, a read is eliminated if a single position of the read has a Phred score below the given threshold. The Phred threshold can be defined by the user.
- <u>Parameters</u>: other parameters that can be adjusted like bowtie mapping parameters and minimum read counts or maximum and minimum read lengths.

Status page

?≣ Jobs Progress							
List of Job IDs							
8IF5HBXLAHLK9B3,8LH41GL5XS021NW,5	KZP11G6ZMJJCYR,BSI	0UX60BWW04XG,XFDIR50K723HDMQ,KB7XB8308TTY	2V7,XLN0XBGKMD7YS44,YHLIXRDI5A902NB				
Show 10 × entries			Search:				
job ID	▲ Status ≑	Started \$	Finished \$	Input \$	Name \$	Group ‡	\$
5KZP11G6ZMJJCYR	Finished	08:49:08, 21 Mar 2022	09:45, 21 Mar 2022	SRR3174962	DROSHA_1_1	Not annotated	
8IF5HBXLAHLK9B3	Finished	08:49:06, 21 Mar 2022	09:43, 21 Mar 2022	SRR3174968	DICER_2	DICER	
8LH41GL5XSO21NW	Finished	08:49:08, 21 Mar 2022	09:38, 21 Mar 2022	SRR3174961	Parental_1_2	WT	
BSI0UX6OBWW04XG	Finished	08:49:09, 21 Mar 2022	09:41, 21 Mar 2022	SRR3174960	Parental_1_1	WT	
KB7XB8308TTY2V7	Finished	08:49:09, 21 Mar 2022	09:42, 21 Mar 2022	SRR3174967	DICER_1	DICER	
XFDIR50K723HDMQ	Finished	08:49:09, 21 Mar 2022	09:42, 21 Mar 2022	SRR3174964	Parental_2	WT	
XLNOXBGKMD7YS44	Finished	08:49:07, 21 Mar 2022	09:41, 21 Mar 2022	SRR3174965	XPO5	XPO5	
YHLIXRDI5A902NB	Finished	08:49:07, 21 Mar 2022	09:42, 21 Mar 2022	SRR3174963	DROSHA_1_2	Not annotated	
Showing 1 to 8 of 8 entries Bulk download results Relaunch Job	s Annotate Sample	Previo	te metrix				

Samples launched will be listed in the status page. In this table users can check information about the status of their job, start and finish time, what input was used and name or group information (if provided) as well as a link to navigate to the individual result of each sample. To provide sample names and groups for differential expression (once all jobs are done running) you can click on the button "Annotate Samples". This will take you to a different page where samples can be annotated using an excel file. Other functions include bulk zipped download of all results, job relaunching (sample samples with different parameters), Differential Expression (Launch sRNAde) and Matrix Generation.

 Launch jobs on sRNAde: Once all sRNAbench jobs are finished aligning, the user can launch an sRNAde job to perform differential expression. Users can navigate to a new page where they can manually provide groups/conditions for their samples or they can use annotations previously provided using the "Annotate Samples" button.

Annotate samples:

?≣ Uploaded files		
Uploaded Files	Sample name	Group
SRR3174962	DROSHA_1_1	Not annotated
SRR3174968	DICER_2	DICER
SRR3174965	XP05	XPO5
SRR3174963	DROSHA_1_2	Not annotated
SRR3174960	Parental_1_1	WT
SRR3174961	Parental_1_2	WT
SRR3174967	DICER_1	DICER
SRR3174964	Parental_2	WT
Provide annotation file Choose file template.xisx	l	B Download template sheet

Users can easily annotate their samples using this feature which provides a pregenerated template (Download template file) with their samples and sample names or sample groups if they had previously been annotated. If a sample is not assigned to a group (i.e. left blank) it will not be taken into consideration for the differential expression analysis.

• Generate matrix:

Generate matrix from sRNAbenc	h jobs ×
You can generate a matrix to summarise dif selectors and buttons below to generate a	ferent files from sRNAbench results. Use the natrix of your choice.
Annotation file 💡	File type 🛛 😮
Read Length ~	microRNA ~
Units	
Read Count Percentage ~	Generate matrix
	Close

By clicking on the "Generate matrix" button users can access this menu where different files from sRNAbench can be summarized into a matrix. For some matrices, a visualization is also available.

Individual job output page

By clicking on the link of each sRNAbench job, users can access a report for each individual sample. There are five sections in the sRNAbench report:

- <u>Parameters</u>: a display of all the parameters used to run the job including sequencing protocol, genome/species and input file.
- <u>Summary</u>: a series of preprocessing and mapping statistics/plots that can provide an overview of the quality of the sequencing library and of the mapping.
- <u>Genome mapping</u>: a summary of the reads mapping to genome and unique reads.
- <u>microRNA summary</u>: a summary of reads mapping to miRNAs (selected reference and genomes) including plots and links to tables to individually explore each mature miRNA sequence, hairpin align files and isomiR distribution.
- <u>sRNA summary</u>: similarly to the previous section, abundance tables are provided for each sRNA category both using single and multiple assignment.
- <u>Other features</u>: this tab is designed to contain current and future functionalities that do not fall within any previous section. Right now, it contains anticodon statistics of tRNAs present in the profile.

Besides these sections, users can download a zip file containing all the result files of the sRNAbench command line tool (check the tool manual for more details). Feel free to contact us if you cannot find a file you were expecting in there or if there is some information you would like to see included.

sRNAde

Consensus differential expression (DESeq, edgeR and NOISeq), visual cluster analysis and sequencing statistics.

sRNAde:

Consensus differential expression (DESeq, DEseq2, t-test, edgeR and NOISeq), visual cluster analysis and sequencing statistics.

Input data	Help		
Use Job IDs Use Group String (advanced)	Link to Web Manual Link to Test Data Results		
Use a list of sRNAbench IDs (comma separated): 💡			
e.g: id1,id2,id3	Test Data Description		
Sample description (provided names will replace jobIDs in analysis, optional) 🕢	Exosome vs Cell fraction from a panel of		
e.g: Normal_1:Normal_2:Tumorl_1:Tumorl_2:Tumorll_2:Tumorll_3	human B cells. Link To Publication		
Sample groups (hash separated, required):			
e.g: Normal#Tumorl#TumorlI	Test Data Parameters		
SUBMIT	Load test Data		
	jobIDs group String		

Description

sRNAde is a module for the detection of differentially expressed sRNA. The input is either a number of sRNAbench output folders or a user-given expression matrix. For the detection of differentially expressed RNAs it applies three widely used methods: edgeR, DESeq and NOISeq. Apart of the output from the individual methods, it provides a consensus differential expression file. Additionally, this tool can perform cluster analysis (heat maps), isomiR analysis and it gives a summary on the sequencing statistic of all used samples (if the input has been from sRNAbench). Result example

Input

Users should provide a list of sRNAbench jobIDs on which differential expression analysis of miRNAs will be performed. There are two ways to do this:

 <u>A list of jobIDS and group names</u>: On the "Use Job IDs" tab the user can provide a list of sRNAbench job IDs (comma separated) and group names (comma separated, as many as the user wants to include in the analysis). In a second step, each sample can be assigned to the group it belongs using a dropdown selector (see image below)

job ID	Input \$	Sample Name \$	Group \$
51UQ20HBJ3XUWPJ	SRR1563064	-	✓ Normal tumor
5DZDVDGE0AZ5GQ0	SRR1563059	-	Normal \$
6IJISG8JIDJC0CA	SRR1563062	-	Normal \$

a. <u>A list of colon and hash jobIDs using # to separate groups (GroupString)</u>: job IDs should be provided in the following way:

 $f1_1:f2_1#f1_2:f2_2$ being $f1_1$ the ID of the first sample of the first group (controls in a case/control study), $f2_1$ the second sample of the first group, $f1_2$ the first sample of the second group, etc. That means that groups are separated by hashes (#), while samples are separated by colon (:).

Note, the ID is assigned by sRNAbench and can be found either in the URL or in the output page

For example, the test data run can also be launched using this grpString:

D42PPR3ZVTVQBK8:EYHC83BJX80ATIL:UWCUSRT3IZO3ME2:XXEYZ4EZ55BHKUG:YH5LQ JZVNP9W3NX:5DZDVDGE0AZ5GQ0#51UQ20HBJ3XUWPJ:6IJISG8JIDJC0CA:FKZ2D88FZN WZCS5:OE1ZNRX2RSJ0OV1:OUSQYGT9V0V6ZA5:ZBM4WYEO1C80DTU

and the following group description string: exosome#cell

The test data comprises therefore 2 experimental groups (exosome and cell). sRNAde calculates the differential expression between all possible combinations (only exosome vs cell in this case).

The user can also provide a description of the samples separated by colons. If no description is given, the sRNAbench job name is used by default.

Finally, the user can provide differential expression thresholds for NOISeq, DESeq, DEseq2, ttest and edgeR.

Please, note they are all adjusted p-value.

Additionally the user can enable isomiR analysis, checking the box "isomiR Analysis".

Results

ou use the sRNAde	please check How to C	be removed on: Tue 15 Ja ite.	an 2019).		
arameters					
esults summary	Preprocessing/QC	Mapping statistics	miRNA and isomiR statistics	Differential expression	
Mapping statist	ics				
Number of m	anned or assigned re	ape			
Number of In	apped of assigned re	445			
		N	umber of mapped or as	ssigned reads	
	DM	N	umber of mapped or as	ssigned reads	 SRR1563018 grp1_XXEY SRR1563056 grp2_ZBM4
1	om	N	umber of mapped or as	ssigned reads	 SRR1563018 grp1_XXEY SRR1563056 grp2_ZBM4 SRR1563064 grp2_51UQ SRR1563061 grp1_EYHC
1 seds	0M 8M	T	umber of mapped or as	ssigned reads	 SRR1563018 grp1_XXEY SRR1563056 grp2_ZBM4 SRR1563064 grp2_51UQ SRR1563050 grp1_EVHC SRR1563059 grp1_5DZD SRR1563063 grp1_VH5L SRR1563063 grp1_int LWCL

b. **Results summary**: Mapping statistics of reads.

- a. Number of mapped or assigned reads
- b. Percentage of mapped or assigned reads
- c. Percentage per detected RNA category
- d. Detected number of miRNAs and precursor sequences
- c. **Preprocessing/QC**: Statistics of reads preprocessing and quality control.
 - a. Preprocessing statistics
 - b. Read length distribution (full, analysis and genome mapped)
- d. Mapping statistics: The user can visualize the genome mapping distribution.
- e. miRNA and isomiR statistics: The user can visualize the fraction of different isomiR classes

per sample.

- f. Differential expression: We provide the result of differential expression of the 5 methods for each comparison between groups defined by user. In this section, only significantly differentially expressed miRNAs are shown (the cut-off value can be provided by the user). The result description of each method is widely explained in their manuals:
 - a. <u>NOISeq</u>
 - b. <u>DESeq</u>
 - c. <u>DESeq2</u>
 - d. EdgeR

sRNAblast

Profiling of small RNAs of unknown origin

Description

This tool is intended for the analysis of reads that could not be mapped using sRNAbench or other profiling tools. The results could point towards either contamination sources or biological meaningful information like the presence of unexpected viral or bacterial RNA molecules. <u>Result</u> <u>example</u>

rofiling of		ofunknow	n origin		
ronning of	Small RNA	S OF UNKNOW	n ongin		
Input data					
Upload the rea	ds (fastq.gz, fa.gz	or rc.gz) 🔞			
Seleccionar ar	chivo Ningún archi	vo seleccionado			
Or provide a U	RL for big files (red	commended!)			
Or provide a sl	RNAbench jobID:				
Number of uni	que unmapped rea	ds to blast			
1					
Database					
nr					
Evalue Maximu	m threshold				

Input

The datasets must be provided uploading a file from a local computer or by means of an *URL*. It accepts the same input formats as sRNAbench. In general, all formats can be compressed with gzip. Additionally a sRNAbench ID can be used as input - in this case, only the unmapped reads

are used, i.e. to determine the origin of those unmapped reads.

Adapter Removal

sRNAblast can perform the adapter trimming. The web-server version will by default search for the first 10 bases of the adapter allowing a maximum of one mismatches. It is recommended to provide the adapter sequence or select one of the options given by the application, which are the most common adapters used on microRNA analysis: Illumina RA3, Illumina (alternative) or SOLiD (SREK). If the adapter is not known, although it is not recommended, guess the adapter sequence option should be activated. Then, sRNAblast will align the first 250,000 reads to the genome using the bowtie seed functionality (the adapters will not count for the mismatches).

Out of all aligned reads, the adapter sequence is defined as the most frequent 10-mer starting at the first mismatch. Lastly, when the adapter is sequenced at the very end of the read, sometimes its length is shorter than the length threshold, so it must be search in a recursively way without taking into account the minimum length.



	sma californicu	m			5657.260869565218	0.5189300159758515
Mycoplas	sma californicu	m HAZ160_1			11088.594202898552	1.017136119321666
Showing 1	to 10 of 27 er	tries			Previous 1	2 3 Next
llast Resu	ilts					
10 🔻	records per	page			Search:	
Query Id	Read Count d	Subject Id	¢ Evalu	e P † Identity†	Specie Name 🗢	Specie Title
6	90588	gi[14089695]emb[AL445564.1]	0.018	100.00	Mycoplasma pulmonis	Mycoplasma pulmonis (strain UAB CTIP) complete genome: segment 2/3
9	53176	gi]14089695[emb]AL445564.1]	0.24	95.83	Mycoplasma pulmonis	Mycoplasma pulmonis (strain UAB CTIP) complete genome; segment 2/3
10	45740	gi[656322258 db][AP013353.1]	0.018	100.00	Mycoplasma californicum HAZ160_1	Mycoplasma californicum HAZ160_1 DNA, complete genome
10	45740	gij640844484jgbjCP005933.1j	0.018	100.00	Mycoplasma bovis CQ-W70	Mycopiasma bovis CQ-W70, complete genome
10	45740	gij640837134 gb CP007521.1	0.018	100.00	Mycoplasma californicum	Mycoplasma californicum strain ST-6, complete genome
10	45740	gij635210191 tpejLK013876.1	0.018	100.00	Mycoplasma agalactiae PG2	TPA: Mycoplasma agalactiae PG2 tRNA Phe-GAA-1-1 gene
10	45740	gij576910268 gb CP007154_1	0.018	100 00	Mycoplasma bovoculi M165/69	Mycoplasma bovoculi M165/69, complete genome
10	45740	gij557878761 gb CP006849.1	0.018	100.00	Mycoplasma hyorhinis DBS 1050	Mycoplasma hyorhinis DBS 1050, complete genome
10	45740	gij523582824 gb CP003802.1	0.018	100.00	Mycoplasma hyopheumoniae 7422	Mycoplasma hyopneumoniae 7422, complete genome
10	45740	gi[506957411]gb[CP003131.1]	0.018	100.00	Mycoplasma hyopneumoniae 168-L	Mycoplasma hyopneumoniae 168-L, complete genome

Results

- 1. This table provide the number of reads detected in each taxonomic group:
 - a. Taxonomy: Taxonomy group Name
 - b. Read Count: Number of reads detected in this taxonomic group
 - c. Percentage Read Count: Percentage of reads detected in this taxonomic group
- 2. This table provide the number of reads detected in each specie.
 - a. Species: Species Name
 - b. Read Count: Number of reads detected for this species
 - c. Percentage Read Count: Percentage of reads detected for this species
- 3. In this section, the full result can be downloaded as zip file.
- 4. This table contain the number of reads for each blast result:
 - a. Query id: Query id
 - b. Read count: Number of reads in this subject sequence
 - c. Subject id: Subject id
 - d. Evalue: Blast e-value
 - e. Percentage of Identity: Blast percentage of identity value
 - f. Specie Name: Subject specie
 - g. Specie Title: NCBI Subject Title

miRNAconsTarget: animal- and plantconstarget

Predict microRNA targets on user defined sets of microRNAs and 3'UTRs

Description

Most microRNA target prediction tools are implemented as webservers on a limited number of species, or do allow only the prediction of a limited number of microRNAs and/or mRNAs. miRNAconsTargets uses three microRNA target prediction programs for animals (<u>PITA</u>, <u>miRanda</u> and <u>TargetSpy</u>) and 2 for plants (<u>psRobots</u> and <u>TAPIR</u>) and applies them to the user supplied sets of microRNAs and 3'UTRs. It reports both, the individual predictions and a consensus prediction. <u>Result example</u>

miRNAconsTarget:
Consensus microRNA target prediction
Input data
Choose miRNA input
Upload miRNAs file
Seleccionar archivo Ningún archivo seleccionado
Or paste your miRNAs here
Choose target input
Upload targets file
Seleccionar archivo Ningún archivo seleccionado
Or choose UTR from the list
None selected V Or choose cDNA from Ensembl
None selected v
Or paste your targets here

Input

The input files (miRNAs file and utr files) must be provided uploading a file in fasta format. If the user wants to detect plant microRNA target genes, the 'Is Plant Analysis' checkbox must be activated. Three options can be chosen for animals and three for plants. Note that TAPIR can be

used in RNAhybrid mode and fasta (alignment) mode.

10 • records per page Search:				
microRNA	◆ mRNA ⇔	Number Of Program ¢	Detected By	
mo-miR-1-3p	NM_031518	2	[ts, miranda]	
rno-miR-1-3p	NM_031506	2	[ts, miranda]	
rno-mi <mark>R-1</mark> -3p	NM_153296	2	[ts, miranda]	
rno-miR- <mark>1</mark> -3p	NM_017236	2	[ts, miranda]	
rno-miR-1-3p	NM_001107092	2	[ts, miranda]	
rno-miR-1-3p	NM_001037792	2	[ts, miranda]	
rno-mi <mark>R-1</mark> -3p	NM_001109262	2	[ts, miranda]	
rno-miR-1-3p	NM_001107086	2	[ts, miranda]	
rno-miR-1-3p	NM_001107075	2	[ts, miranda]	
rno-miR-1-3p	NM_001014233	2	[ts, miranda]	
howing 1 to 10 of 1,462 entries Download TargetSpy Resu Download Miranda Result		Previous 1 2 3 4 5 147 N	iext	

Results

- 1. In this table consensus result are shown:
 - a. miRNA: miRNA name
 - b. mRNA: transcript Target name
 - c. Number of Programs: Number of programs that predicted this mRNA as target
 - d. Detect by: Names of programs that predicted this mRNA as target
- 2. Link to download TargetSpy results
- 3. Link to download miRanda results
- 4. Link to download Consensus results

sRNAcons

Estimates the conservation depth of a set of small RNAs

Description

User provided small RNA sequences (fasta format) are mapped against all assemblies available in sRNAtoolbox. The tool has two different outputs, i) The conservation depth for all small RNA input sequences, i.e the percentage of genomes in which the sequence was found, and ii)the percentage of mapped input sequences per genome.

sRNAcons:

Mapping of small RNA sequences to all genomes in sRNAtoolbox

nput data
Choose your input 😦
Upload URL/link
Upload input file(Fasta file)
Seleccionar archivo Ninguno archivo selec.
Parameters
Number of mistmatches*
1
Kingdom
Animal
SURMIT

Input

The input file should be a fasta file (provided updating the file of by means of an *URL*) with the small RNA sequences of which you want to estimate conservation depth (for example a set of mature miRNAs). You can also choose the number of mismatches you want to allow in the sequences alignment, and the kingdom (animal or plants) against which you want to align your

sequences.



Results

- 1. Graphic summary:
 - a. Top 20 conserved miRNAs
 - b. Top 20 species with the highest number of conserved sRNAs
- 2. **Conservation depth per sRNA :** The user can visualize/download a table with the conservation depth per sRNA sequence, including the frequency and a list of species in which the sequence is conserved per sRNA.
- **3. Conservation depth per species:** The user can visualize/download a table with the frequency of the input sRNAs per species.

Helper Tools

sRNAtoolbox implements several 'helper tools' which are intended to aid the generation of a local sRNAbench database or/and which might be useful for the preparation of input data of some of

the tools. Below, the six different helper tools are described in more detail.

- Ensembl Parser
- NCBI Parser
- RNA central parser
- Genomic tRNA database parser
- Remove duplicates from a fasta file
- Extract Sequences from a fasta file

Helper Tools These tools are intended to help the user to either setting up a local sRNAbench database (Ensembl Parser, NCBI Parser, RNA Central Parser, genomic tRNA Parser) or to prepare input data for other sRNAtoolbox tools. If you use any of the generated data in your publication, please cite the papers given at the data retrieval page (NCBI, ENSEMBL, genomic tRNA and RNA central databases). Tools: Ensembl Parser NCBI Parser RNA central parser Genomic tRNA database parser Remove duplicates from a fasta file Extract Sequences from a fasta file

Ensembl Parser

Ensembl Parser:	
sRNAbench can read specially pre reads a Ensembl fasta annotation	pared annotations, i.e. the transcript name and the classification separated by ':'. The 'Ensembl Parser' file and generates the sRNAbench format. In can use cDNA and ncRNA fasta files from this page .
Link to test data	
If you use this data please cite: Pa	ul Flicek et al. (2012). Ensembl 2013. Nucleic Acids Res.
Input data	
Upload input file(Ensembl file):	
Browse No file selected.	
Or provide a URL for big files (re	commended!)
D.L. I. Draw	
Submit Reset	

NCBI Parser

NCBI Parser:	
sRNAbench can read reads a NCBI fasta ar	specially prepared annotations, i.e. the transcript name and the classification separated by ':'. The 'NCBI Parser' notation file and generates the sRNAbench format. In can use *.rna.fna.gz files from the following page .
Link to test data	
If you use this data pl genome annotation p	ase cite: Kim D. Pruitt et al. (2012). NCBI Reference Sequences (RefSeq); current status, new features and licy. Nucleic Acids Res. PMCID: PMC3245008.
Input data	
Upload input file(NC	31 file):
Browse No file s	elected.
Or provide a URL for	big files (recommended!)
Submit Reset	

RNA central parser

RNA central parser	
This parser extracts the non-codin 2015) (http://rnacentral.org/) and	g sequences of a given species or taxonomy level from the RNA central database version 2 (Februar prepares the sRNAbench format libraries.
Test with: Hordeum vulgare	
If you use this data please cite: Th 10.1093/nar/gku991	e RNAcentral Consortium (2014) RNAcentral: an international database of ncRNA sequences. doi:
Input data	
Input data Provide a Species Name (Must b	e Scientific Name):
Input data Provide a Species Name (Must b Dr provide a Taxonomy Name: [e Scientific Name):

Genomic tRNA database parser



Remove duplicates from a fasta file

emove Duplicates from a Fasta File :	
he UCSC table browser allows to obtain 3' UTR sequences which are needed when searching for microRNA target genes. How e output files have the following format:	vever
>hg19_refGene_NM_001184906 range=chr17:37408897-37417712 5'pad=0 3'pad=0 strand=- repeatMasking=non CAATGGAGGTGGTCAACCTTGGCGAACTGAGTATTTAATGACACTTCTAG AGCTACCGTGGAGTCTCTCCAGTGGAAGCAACCCCAGTGTTCTGAGCAAG	e
he name of the sequence ' hg19_refGene_NM_001184906' would not be recognized by downstream analysis programs (funct nrichment analysis). This parser allows to substitute certain parts of the sequence name in order to recover the name of the tra NM_001184906 in this case).	tional anscript
becifying 'hg19_refGene_' in the textbox would make the parser to remove this string from the sequence name leaving only the transcript. This parser removes also duplicated IDs and gives the additional possibility to remove duplicated sequence.	e name
put data	
oload input file(Fasta file):	
Browse No file selected.	
provide a URL for big files (recommended!)	
ovide a string of characters to be drop out from the sequence names:	
Remove also duplicate SEQUENCES	
Submit Reset	

Extract Sequences from a fasta file

Extract Sequences from a fasta file:			
This parser salows to specify a search criterion for fasta sequence names. For example, microRNA sequences from miRBase start with the species name like 'has' (Homo sapiens). Providing a miRBase fasta file specifying 'has' would make the program to return only humar (hsa) sequences from the file.			
Input data			
Upload input file(Fasta file):			
Browse No file selected.			
Or provide a URL for big files (recommended!)]		
Provide a string the characters used to select the sequences:			
Submit Reset			