

Tutorial of LIRBase

LIRBase is a web server for comprehensive analysis of **siRNAs** (small interfering RNA) derived from **long inverted repeat** in eukaryotic genomes.

Source code: <https://github.com/venyao/LIRBase>

Online use: <http://venyao.xyz/lirbase/>

Contact: yaowen@henau.edu.cn

Content

1. Browse LIRBase for long inverted repeats identified in 424 eukaryotic genomes	2
2. Search LIRBase for LIRs in a specific genome by genomic regions	4
3. Search LIRBase for LIRs in a specific genome by the identifiers of LIRs	5
4. Search LIRBase by sequence similarity using BLAST	6
5. Detect and annotate long inverted repeats in user-uploaded DNA sequences	8
6. Annotate and quantify the expression level of LIRs using sRNA sequencing data	10
7. Differential expression analysis of long inverted repeats and small RNAs	12
8. Download LIRs of 424 eukaryotic genomes, the BLAST database and the Bowtie index database	13
9. Information of 424 genomes collected in LIRBase	14

Using IRF (<https://tandem.bu.edu/irf/irf.download.html>), we identified a total of 6,619,473 long inverted repeats in the whole genomes of 424 eukaryotes, including 297,317 LIRs in 77 metazoa genomes, 1,731,978 LIRs in 139 plant genomes and 4,590,178 LIRs in 208 vertebrate genomes. LIRBase is deployed at <http://venyao.xyz/lirbase/> for online use.

The homepage of **LIRBase** displays the main functionalities of LIRBase (Figure 1). The definition of long inverted repeat, the biogenesis pathway of siRNAs from long inverted repeat and the biological roles of siRNAs generated in this pathway are elaborated in the homepage of LIRBase. These results implied that **a platform for comprehensive annotation and analysis of siRNAs derived from long inverted repeat is in urgent need.**

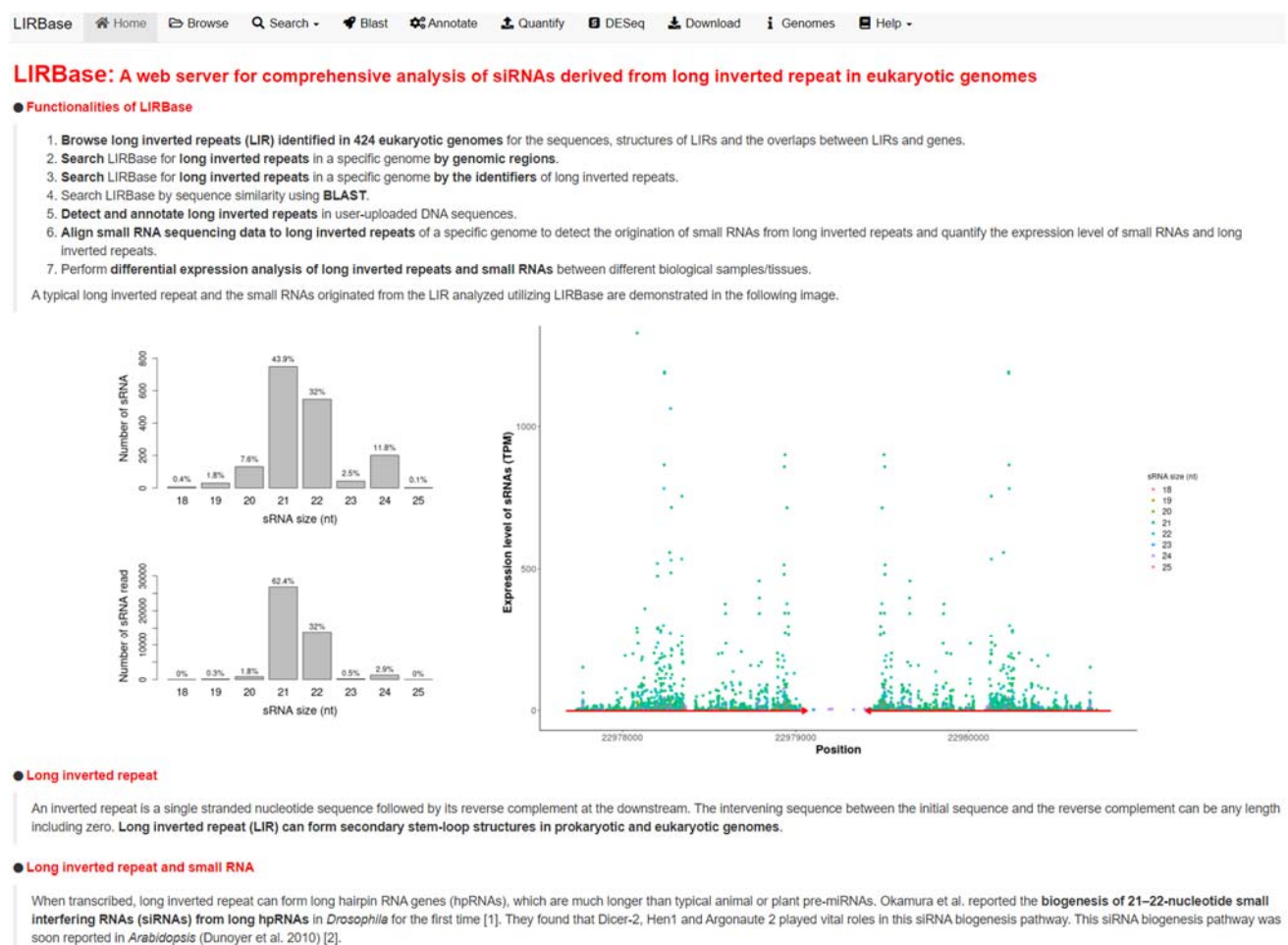


Figure 1. The homepage of LIRBase.

1. Browse LIRBase for long inverted repeats identified in 424 eukaryotic genomes

The images and the species names of 424 eukaryotes are listed in the “Species” panel of the “Browse” menu of LIRBase (Figure 2). Click of the image or the species name of any genome would take you

to the “LIRs annotated by IRF” panel of the “Browse” menu, which displays all the LIRs identified in the selected genome (Figure 3). A brief summary of all the LIRs of the selected genome and a table of all the LIRs showing the structure of each LIR is demonstrated in the “LIRs annotated by IRF” panel. Click of the ID of any LIR in the table of all LIRs would take you to the “Details of the LIR selected” panel of the “Browse” menu, which displays the sequence, structure of the selected LIR and the overlaps between the selected LIRs and gene (Figure 3 and 4).

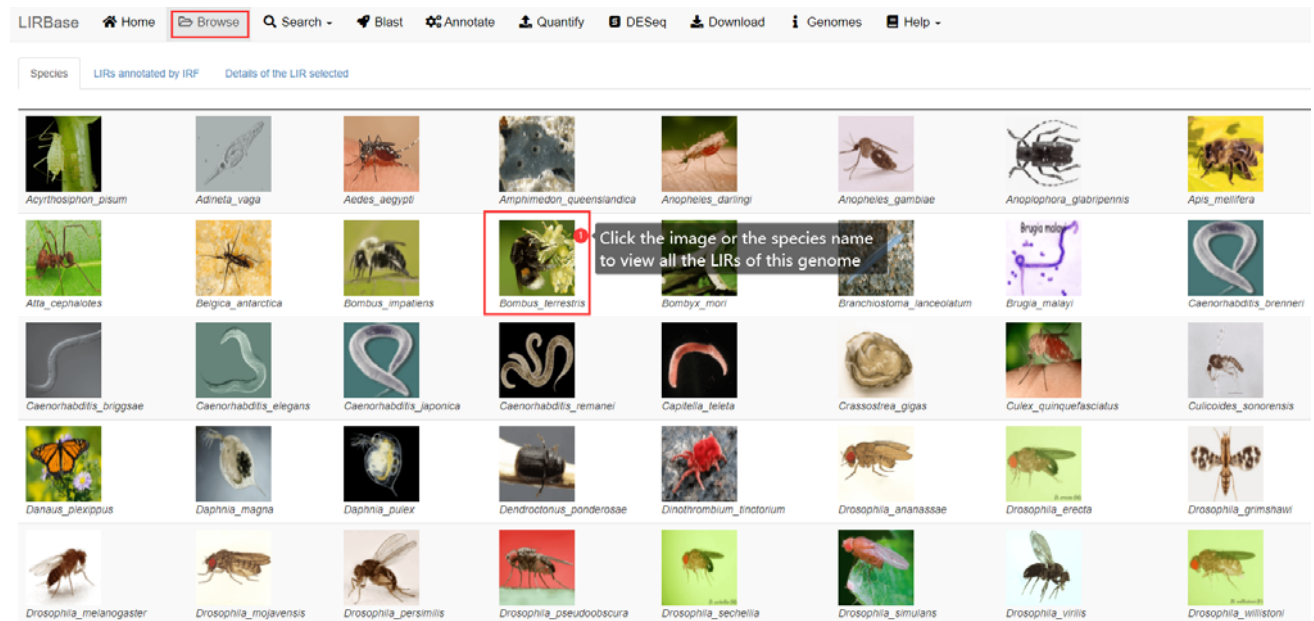


Figure 2. Species name and images of 424 eukaryotic genomes listed in the “Species” panel of the “Browse” menu.

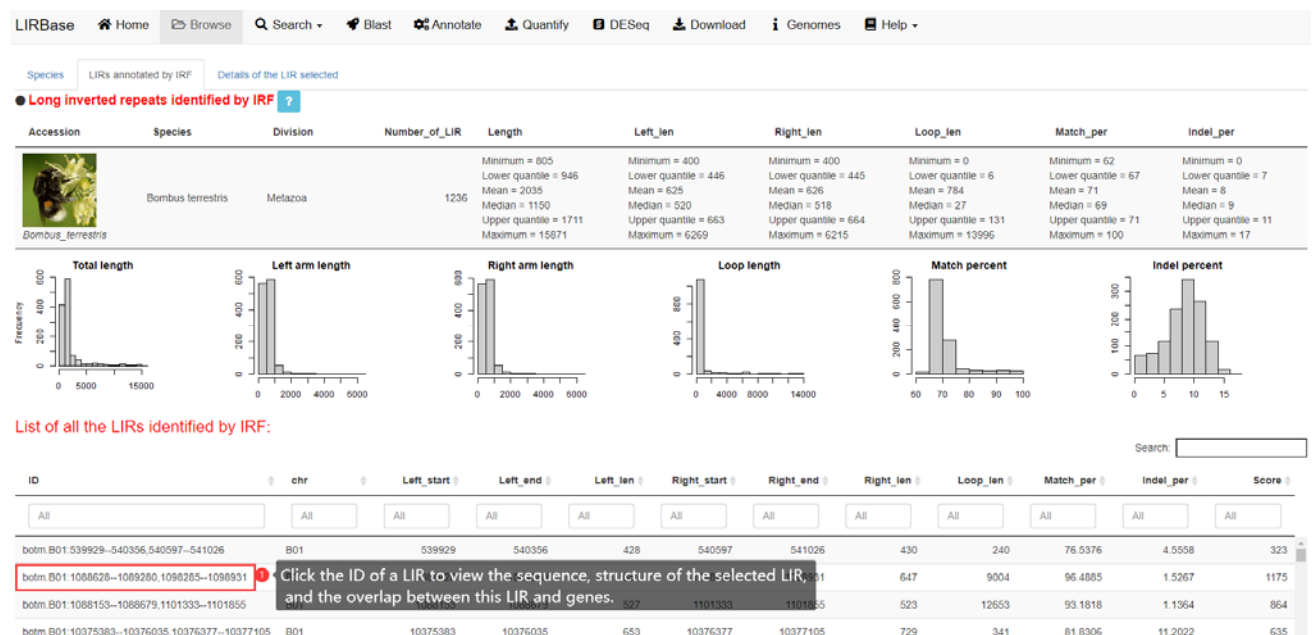



Figure 3. List of all the LIRs identified by IRF for a selected genome.

[illegible]

2. Search LIRBase for LIRs in a specific genome by genomic regions

LIRBase allows searching for LIRs of any of the 424 eukaryotic genomes by genomic regions (Figure 5). The detailed steps are shown in Figure 6.



LIRBase Home Browse Search Blast Annotate Quantify DESeq Download Genomes Help

Search by genomic region ?

Choose genome

Search by genomic location

Search by LIR identifier

Search...

Choose from:

Acyrtosiphon_pisum

Adineta_vaga

Aedes_aegypti

Amphimedon_queenstandica

Anopheles_darlingi

You have selected:

Choose chromosome

HG380764

Search! ?

Download the structure of LIRs in the search result

Download the sequence of LIRs in the search result

Figure 5. The “Search by genomic region” submenu of the “Search” menu.

[illegible]

3. Search LIRBase for LIRs in a specific genome by the identifiers of LIRs

LIRBase allows searching for LIRs of any of the 424 eukaryotic genomes by the identifiers (IDs) of long inverted repeats (Figure 7). The detailed steps are shown in Figure 8. After clicking the “Search” button in the “Input” panel shown in Figure 8, the results would be displayed in the “Output” panel (Figure 9).

LIRBase
Home
Browse
Search
Blast
Annotate
Quantify
DESeq
Download
Genomes
Help

Search by LIR identifier
?

Input LIR identifiers

One item in one row

Choose genome

Search!
?
Reset
Load example

Figure 7. The “Search by LIR identifier” submenu of the “Search” menu.

LIRBase
Home
Browse
Search
Blast
Annotate
Quantify
DESeq
Download
Genomes
Help

● Search by LIR identifier ?

Input
Output

▶ Input LIR identifiers
1 Input LIR identifiers in separated rows

drmm.2L:983450--984514,996778--997842
drmm.2L:1776370--1781782,1795872--1801274
drmm.2R:1523505--1525931,1527463--1529891
drmm.2L:6431564--6436015,6440661--6445112
drmm.2L:12731893--12735820,12739226--12743092
drmm.3R:56--7751,8029--15736
drmm.3R:14248--20462,20495--26665

▶ Choose genome
2 Choose a genome to search

Drosophila_melanogaster

Search! ? Reset Load example

3 Click the Search button to start the search

Figure 8. Steps to search LIRBase by LIR identifiers.

● Search by LIR identifier ?

Input
Output

Download structure of LIRs in the search result
Download sequence of LIRs in the search result

Show 10 entries
1 Click to download the structures of all LIRs in the search result
2 Click to download the sequences of all LIRs in the search result

ID	chr	Left_start	Left_end	Left_len	Right_start	Right_end	Right_len	Loop_len	Match_per	Indel_per	Score
drmm.2L:983450--984514,996778--997842	2L	983450	984514	1065	996778	997842	1065	12263	100	0	2130
drmm.2L:1776370--1781782,1795872--1801274	2L	1776370	1781782	5512	1795872	1801274	5403	14089	97.5161	0.9536	10067
drmm.2L:6431564--6436015,6440661--6445112	2L	6431564	6436015	4452	6440661	6445112	4452	4645	99.9775	0	8899
drmm.2L:12731893--12735820,12739226--12743092	2L	12731893	12735820	3928	12739226	12743092	3867	3405	67.4261	4.8322	1097
drmm.2R:1523505--1525931,1527463--1529891	2R	1523505	1525931	2427	1527463	1529891	2429	1531	99.506	0.0823	4794
drmm.3R:56--7751,8029--15736	3R	56	7751	7696	8029	15736	7708	277	64.6446	6.631	809
drmm.3R:14248--20462,20495--26665	3R	14248	20462	6215	20495	26665	6171	32	63.8638	7.4915	279

Showing 1 to 7 of 7 entries

Previous 1 Next

Overlaps between the selected LIR and genes:

ID	chr	Left_start	Left_end	Left_len	Right_start	Right_end	Right_len	Loop_len	gene.start	gene.end	gene.id	gene.biotype
drmm.2L:983450--984514,996778--997842	2L	983450	984514	1065	996778	997842	1065	12263	987807	991300	FBgn0266036	ncRNA

Sequence of the LIR:

```

>drmm.2L:983450--984514,996778--997842
ttaaaggatggaattgtcatcttcgaagatgcctttggaagatggaagatgcataaggggaatgatgaaggt
<ggagggctaaagtaagagtcggggaggtatatacagcccttaacaaatttgctcttggaggaataaggt
<tgttgcacaaatgaggctggcagagtcacaaagagtcataacattcagtcctaaagttatattatcaaca
CCTTATATAGAGAGAGTATTGTTGGATCCGCGAAGATTTTGGATCCGCGAAGATTTTAAATCATACTAGAGG
ATTGGAACGATTGAAATGAAATTAATTTATGAAGAGAGACCATCAAAATTTGAAGATTGAAATTTCCATCATATT

```

Alignment of the left arm against the right arm:

```

983440 >> (L) CAAAAGACG
997852 << (R) ATCAACGGGT

983450 >> CCATACATTCACTCCGAAAGTGAAGATTATTATCAACACCTTATATAGAGAGAGTATTGTTGGATCCGCGAAGAT >> 983524
997842 << *****

```

Figure 9. The “Output” panel of the “Search by LIR identifier” submenu.

4. Search LIRBase by sequence similarity using BLAST

Users can choose to search LIRBase by sequence similarity utilizing BLAST (Figure 10). A graphical interface was implemented in LIRBase for users to perform BLAST alignment through the NCBI BLAST+ program. BLASTN databases were constructed for all the LIRs identified in each of the 424 eukaryotic genomes. Users can choose to BLAST against any one or more genomes. The detailed steps to perform BLAST in LIRBase is shown in Figure 10.

LIRBase Home Browse Search **Blast** Annotate Quantify DESeq Download Genomes Help

● Search LIRBase by sequence similarity using BLAST ?

Input Output

▶ Paste or upload input data? ? ❶ Choose a method to input the query sequence of BLAST

Paste input data

Input sequence

```
>test1
TTCGACGATTGCTTTGCTGTGTCGAGACATGAACATATTGTCTGAACGCCATGCTATTTCTGTAAACG
ATGATTCAAGCAGTAACATACTTGGTGT
ATCCTTTGTACTCTTCTATAATACGTTGTCTTGCCTCTTCAGTAAATGCATTACATTCTCTATTTCTGTG
CAGTTCAACTAGTAGAGCTTGTCTT
TCGTGAACGTTGTACTGATTGGCCAGCACATGACCTTAATAAATTTGGTCTAAGTTAGTCTTGGTACCAAT
TACCAATTATAAACATCACAGCTTGATA
TAAAGAAATGATGATTGCGAGTGTCAAAACTACAGCTCAGTCTGATGACGTTTGGTCTGCTGAAVAG
CAAAATTTAATCTGGGCATTAAAGACAGCC
AACTGAACAGTGTACACATTTATTTGACTAGTGTGCTTATGTACATATGCTCTTTTATCACAATCTGGAGC
TAGTTGCAATTAATCATCTTGTGTTTT
GCTGCTTTACCATTTATTTGGTGTAGTGGCTACGGAGTCAAATGATGCAGAAAGCAAGCAAGAGAGC
GTAATGCAAAAGAAACAGTGAATGTAATG
CAAGAGATTCCAGGTAAAGCACAATTGAACAAGTTATCAGCAATTTGGTCTTTTATCATCTGGTGTGATGAAT
TTTTGTGCTAAAGCTTCATAGAAGAAAT
AATAAATTAATGGGTATTTAATTTATATGAGAATATTATTGTAGCACAATTTACTGAATATATACGAGAGCT
GTCTATATGACATGTGGTCCAGAA
ATTTTAAATATCAGGTGTCCGATATATTCGATATTGGACTCGTTTGTGTGATAATTTGAATAGAAAGTGAAT
GGTTCAATGCATTTTAATCCGATA
TATGATTCAAAAGATTGCAACTATTAAACCAAAAGAAATACGCAAGTCTATCTTACTACTGAGCATGCA
```

❷ Choose BLAST databases ?

Search... Choose the BLAST databases

Choose from:

- Acyrtosiphon_pisum
- Adineta_vaga
- Aedes_aegypti
- Amphimedon_queenslandica
- Anopheles_darlingi

You have selected:

- Oryza_sativa.MH53
- Oryza_sativa.Nipponbare

❸ Set other parameters for BLAST

E-value cutoff ?

Maximum no. of hits ?

10

10

BLAST! ? Reset Load example

❹ Click the BLAST button to start the BLAST

Figure 10. Steps to BLAST in LIRBase.

Once the BLAST alignment is finished, you would be taken to the “Output” panel of the “Blast” menu, which displays the BLAST result in details (Figure 11). You can view and download the whole BLAST results, which was shown as a table. By clicking a row of this table, you can view the detailed information of a BLAST hit, including the alignment of a query sequence and a subject LIR sequence in the BLAST database represented by this BLAST hit. The structure, sequence of the LIR in this BLAST hit and the overlaps between this LIR and genes in the corresponding genome was also shown in the “Output” panel after clicking a row of the BLAST result table (Figure 11).

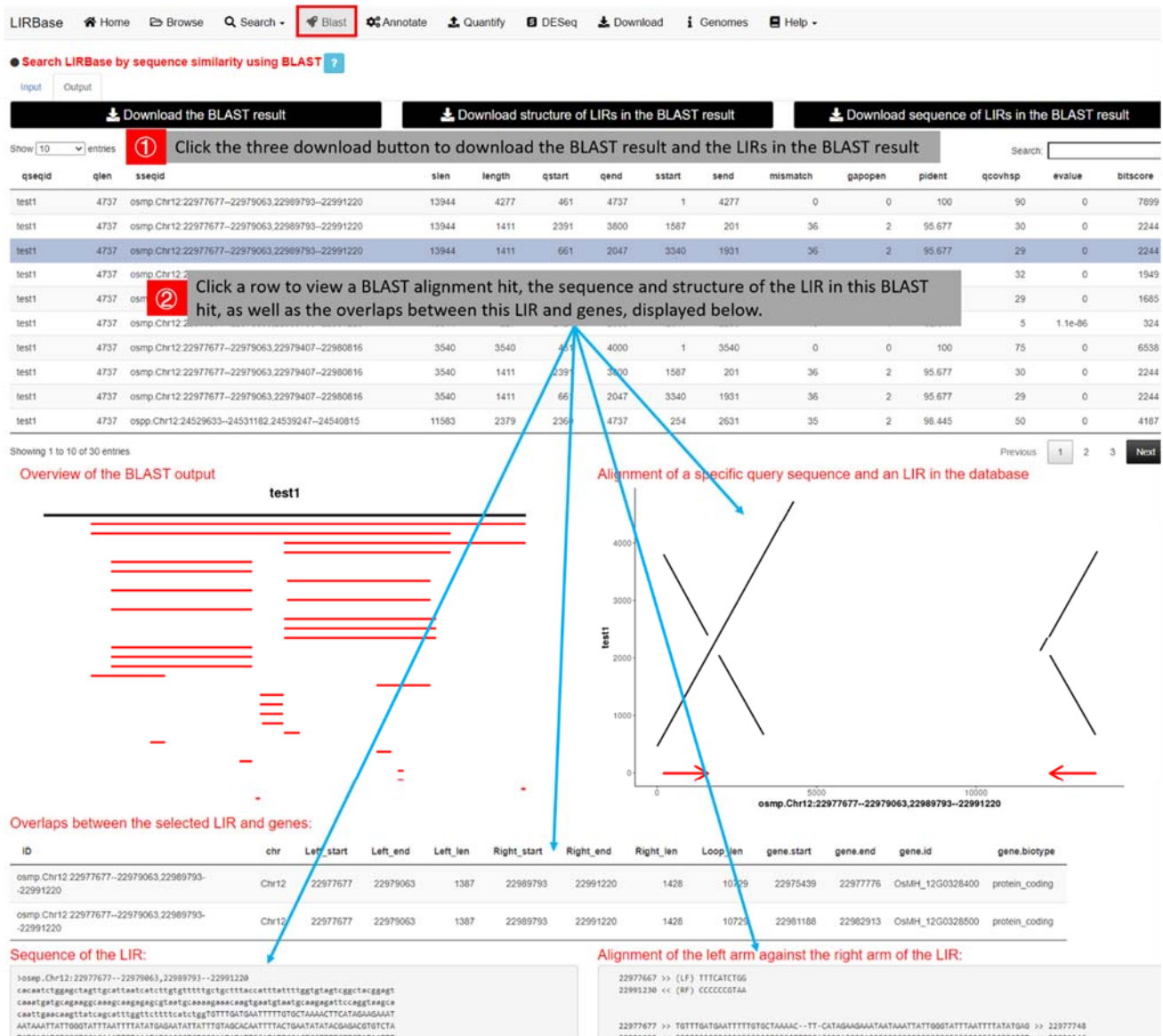


Figure 11. The “Output” panel of the “Blast” menu.

5. Detect and annotate long inverted repeats in user-uploaded DNA sequences

The software IRF (<https://tandem.bu.edu/irf/irf.download.html>) was utilized to identify long inverted repeats in the 424 eukaryotic genomes collected in LIRBase. IRF can only be used in the command line. We implemented a graphical interface for users to annotate long inverted repeats in user-uploaded DNA sequences by IRF (Figure 12). The detailed steps to annotate LIRs in user-uploaded DNA sequences are shown in Figure 12. The input DNA sequences for IRF can be pasted in a text area provided or be uploaded from a local text file. **The input data must be DNA sequence in fasta format. Each sequence should have a unique ID start with “>”.**

LIRBase Home Browse Search Blast **Annotate** Quantify DESeq Download Genomes Help

Annotate LIRs in user-uploaded DNA sequences

Paste or upload input data?

Paste input data

The input DNA sequence can be pasted in the textarea or be uploaded from a local file

```

CATTACATTCTCTATTTCTGTGCGAGTTCAACTAGTAGAGCTTGTCAIT
TCGTGAACGTTGTACTGATTTGCCAGCACATGACCTTAATAAATTTGGTC
CTAAGTTAGTCTTGGTACCAATTACCAATTATAACATCACAGCTTGATA
TAAAGAAAATGATGATTCAGTGTCAAAACTACAGCTCAGTCTGATG
ACCGTTTGTCTGCTGAAAAGCAAAATTAATACTGGGCATTAAGACAG
CC
AACTGAACTAGTTCACAGCTTATTGACTAGTGTGCTTATGTACATATGC
TTCTTTTATCACAACTCGGAGTAGTTGCATTAATCATCTTGTGTTTT
OCTGCTTTACCAATTATTTGGTGTAGTCGGCTACGGAGTCAAATGATG
CAGAAAGCAAGCAAGAGAGCGTATGCAAAAGAAACAGTGAATGTA
ATG

```

Submit Clear Load example

Download structure of predicted LIRs

Download sequence of predicted LIRs

Inverted Repeats Finder Program written by: Gary Benson
Department of Biomathematical Sciences
Mount Sinai School of Medicine
Version 3.05

Click to download the results of IRF

Multiple Sequence Summary

Only sequences containing repeats are shown!

Click on sequence description to view repeat table.

Sequence Index	Sequence Description	Number of Repeats
1	test1	1
2	test2	2

Click to view the result of IRF in HTML pages

Click the Submit button to start the annotation of LIRs by IRF

Matching weight

Set other parameters for IRF

Mismatching penalty

Indel penalty

Figure 12. The “Annotate” menu of LIRBase to annotate LIRs in user-uploaded DNA sequences.

The sequences and structures of LIRs identified by IRF can be downloaded as text files (Figure 12). The result of IRF can also be view in HTML pages (Figrue12 and 13).

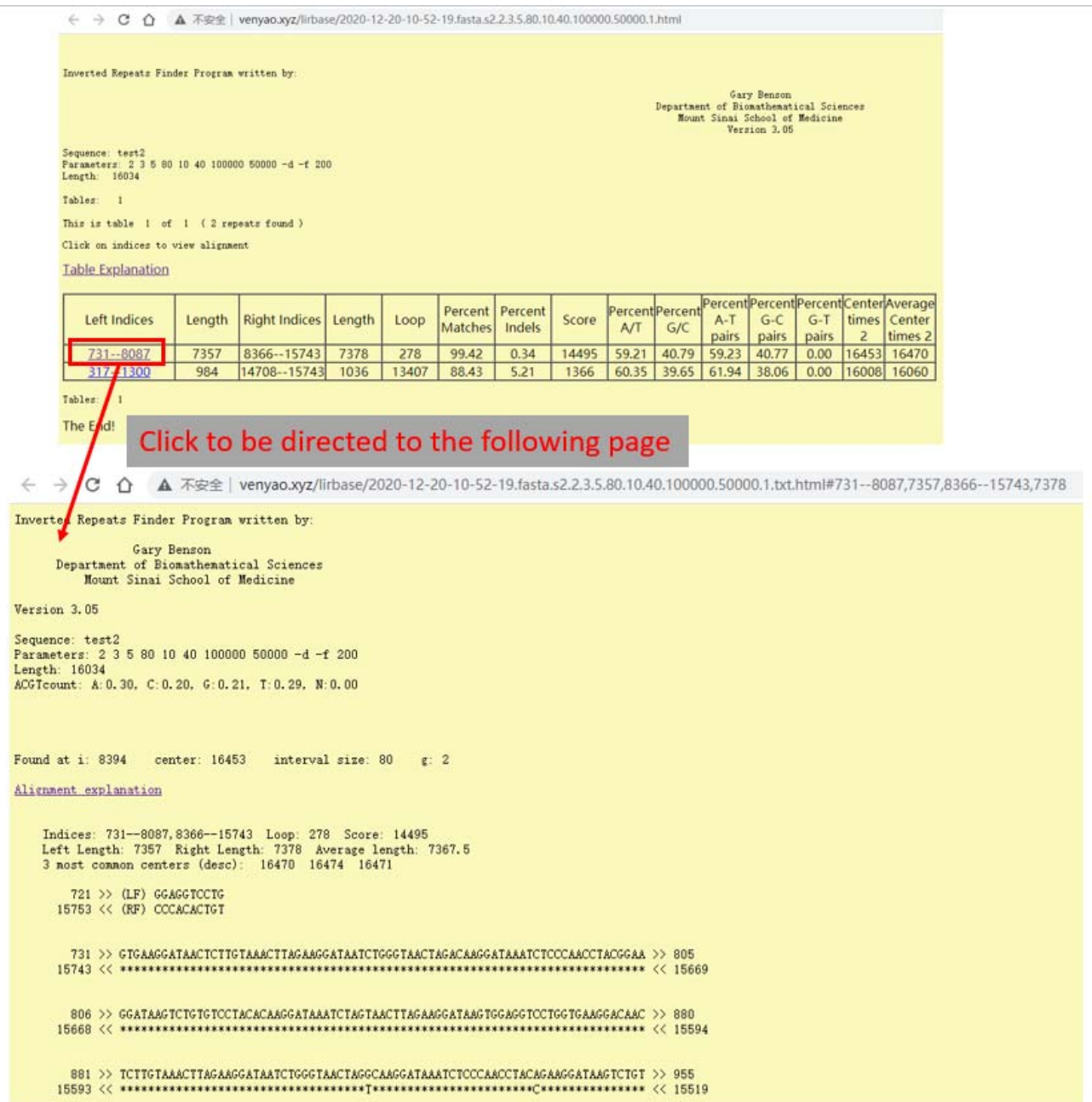


Figure 13. The LIRs identified by IRF viewed in HTML pages.

6. Annotate and quantify the expression level of LIRs using sRNA sequencing data

When transcribed, long inverted repeat can form long hairpin RNA genes (hpRNAs), which are much longer than typical animal or plant pre-miRNAs. Okamura et al. (2008) reported the biogenesis of 21–22-nucleotide small interfering RNAs (siRNAs) from long hpRNAs in *Drosophila* for the first time. This siRNA biogenesis pathway was soon reported in *Arabidopsis* (Dunoyer et al. 2010).

To facilitate the annotation of small RNAs derived from LIRs archived in LIRBase, we

implemented a functionality in LIRBase allowing alignment of user-uploaded small RNA sequencing data to all the identified LIRs of a genome (Figure 14). The input data should be read count of small RNAs rather than the raw small sequencing data as shown in Figure 14. The input small RNA read count data can be pasted in a text area provided or be uploaded from a local text file.

After clicking the “Align!” button, the alignment would be performed. The alignment results would be displayed in the “Output” panel of the “Quantify” menu (Figure 15). The detailed alignment result, the summary of alignment and the sRNA read count of aligned LIRs can be downloaded. What’s more, the summary of alignment result and the sRNA read count of aligned LIRs can be viewed as data tables in the HTML page. By clicking on a single row of the table of sRNA read count of LIRs, the size distributions of sRNAs and the alignment of sRNAs to the LIR would be plotted in figures. The detailed information of the chosen LIR would be displayed in the bottom of the “Output” panel.

LIRBase Home Browse Search Blast Annotate **Quantify** DESeq Download Genomes Help

● Annotate and quantify the expression level of LIRs using sRNA sequencing data ?

Input Output

► Input sRNA read count data ?

Paste input data Paste or upload small RNA read count data

Paste sRNA read count

sRNA	readCount
TGGACACTGTCACTTTTCGGCG	4764
AGAAGATAGCAATCTCTGGA	3412
TCBAGCAACAGTABAGAA/BGA	3590
TCAACTTCGTGCGAGGTGGCT	2926
TATGACTGACATTGATGACGAG	2671
AGAAGCTGCGCGTAGTCTCAATTA	2511
ACCAACCGGGACTAAAGATGGCTC	2428
TCTAGACGAGCGAGGACTACC	2360
TCAAGACAGGAGATAGACCA/G	2275
CGAGCAACAGTAGAGAGGAG	2196
TCAAGGACAGGAGATAGACCA	2169
TGGAGACTGGCAGAAAACTC	1986
AGGAATGAAGTAGATTTCGGTA	1959
TTGGACACTGTCACTTTTCGGC	1912
TAATTACAA/GAGAGGACGCA	1895
ACACCAACCGGGACTAAAGATGGC	1893
GTTGAATTGGAACATGTGACGG	1869
AAGGACAGGAGATAGACCA/G	1817
AGGACAGGAGATAGACCA/G	1693

► Choose an LIR database to align the sRNA data ?

Oryza_sativa.MH53 Choose a LIR database to align the small RNA read count data using Bowtie

► Max number of alignment hits of sRNA ?

1 11 21 31 41 50 61 71 81 91 100

Set the max number of alignment hits for each small RNA

Align ? Reset Load example

4 Click the Align button to start the alignment

Figure 14. The “Quantify” menu of LIRBase to align small RNA sequencing data to a LIR database.

are required as input data for the differential expression analysis. The sample in the count matrix and the sample in the information table must be in the same order. Check the example data provided by LIRBase for the format of a sample information table.

The results of DESeq2 can be downloaded as a plain text file or can be viewed in a data table in the HTML page (Figure 16). In addition, the MA-plot and the volcano plot showing the identified differentially expressed LIRs/sRNAs are also generated. A heatmap displaying the sample-to-sample distance is shown in the bottom of the “DESeq” menu.

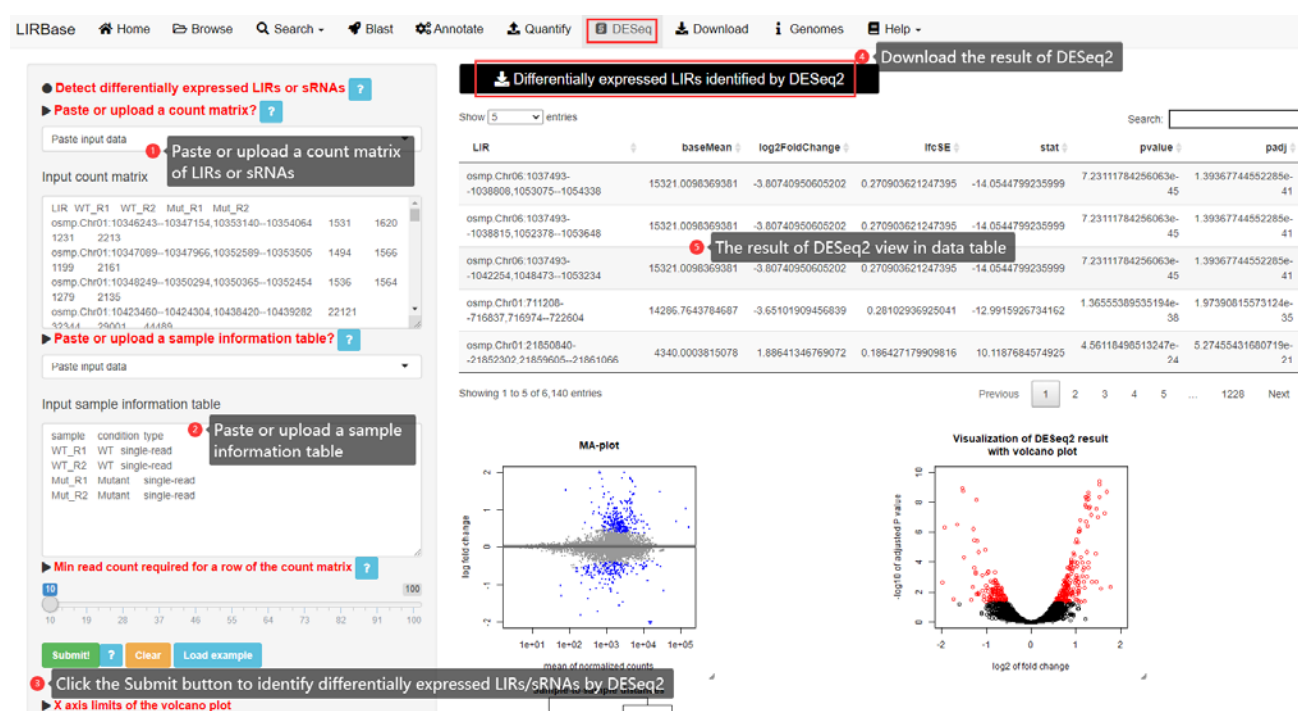


Figure 16. The “DESeq” menu of LIRBase to perform differential expression analysis of LIRs/sRNAs.

8. Download LIRs of 424 eukaryotic genomes, the BLAST database and the Bowtie index database

In addition to be used online at <http://venyao.xyz/lirbase/>, LIRBase can be deployed on a personal local or web Linux server. Deployment of LIRBase is platform independent, i.e., LIRBase can be deployed on any platform with the R environment available. The detailed steps are described in the “Installation” submenu of the “Help” menu of LIRBase (Figure 17). The source code of LIRBase is deposited in GitHub (<https://github.com/venyao/LIRBase>). As the file size of identified LIRs and the corresponding BLAST/Bowtie databases of the 424 eukaryotic genomes are too large, these datasets were not uploaded to GitHub. Instead, these data can be downloaded from <http://venyao.xyz/lirbase/>

through the “Download” menu (Figure 18).

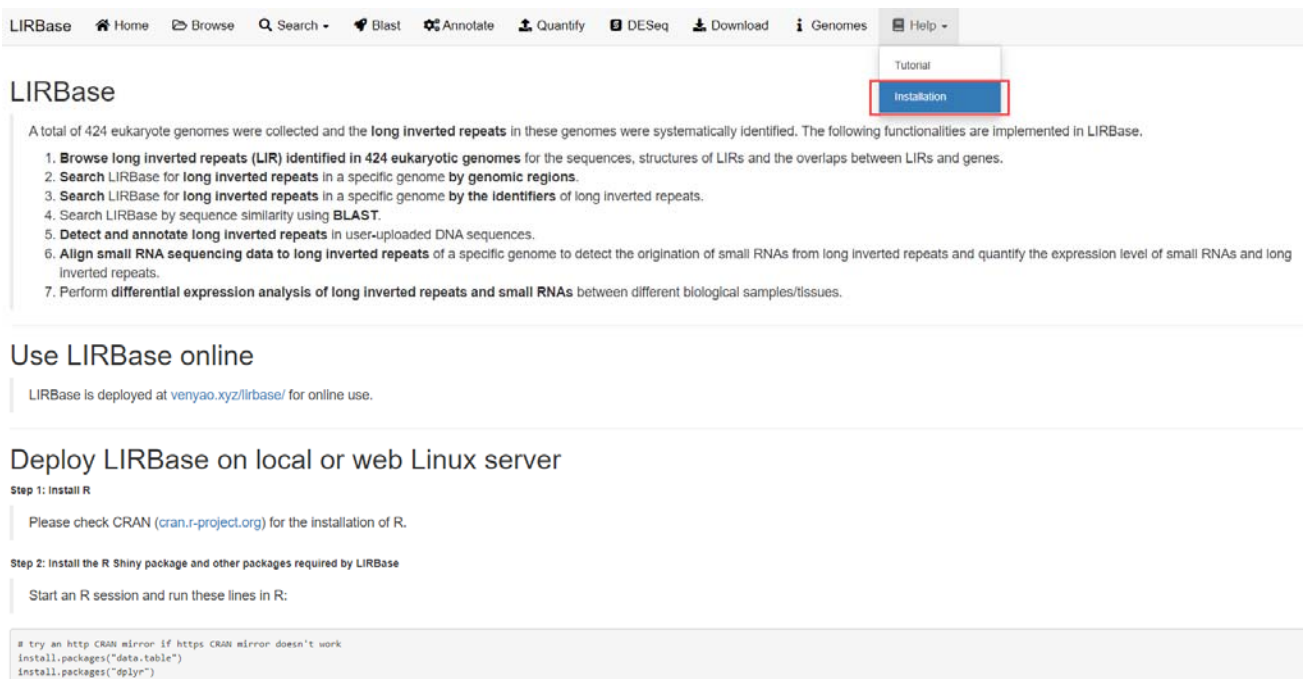


Figure 17. The “Installation” submenu of the “Help” menu of LIRBase.

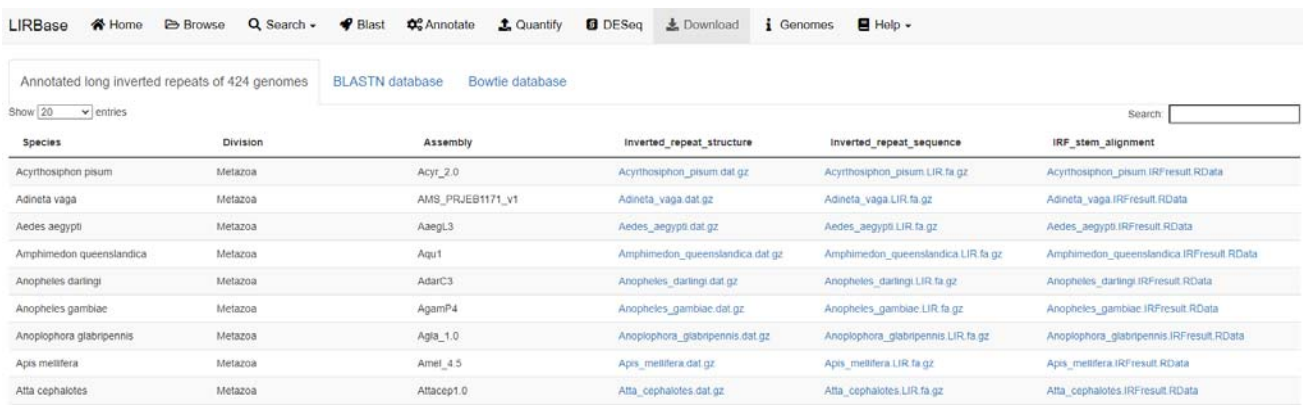


Figure 18. The “Download” menu of LIRBase.

9. Information of 424 genomes collected in LIRBase

The information of 424 genomes collected in LIRBase is displayed in the “Genomes” menu of LIRBase (Figure 19).

LIRBase	Home	Browse	Search	Blast	Annotate	Quantify	DESeq	Download	Genomes	Help
---------	------	--------	--------	-------	----------	----------	-------	----------	---------	------

Information of 424 genomes collected in this database.

Show entries

Search:

Species	Division	Assembly	Source	Publication
<i>Acyrthosiphon pisum</i>	Metazoa	Acyr_2.0	http://metazoa.ensembl.org/Acyrtosiphon_pisum/info/Index	10.1371/journal.pbio.1000313
<i>Adineta vaga</i>	Metazoa	AMS_PRJEB1171_v1	http://metazoa.ensembl.org/Adineta_vaga/info/Index	10.1038/nature12326
<i>Aedes aegypti</i>	Metazoa	AaegL3	http://metazoa.ensembl.org/Aedes_aegypti/info/Index	10.1126/science.1136878
<i>Amphimedon queenslandica</i>	Metazoa	Aqu1	http://metazoa.ensembl.org/Amphimedon_queenslandica/info/Index	10.1038/nature09201
<i>Anopheles darlingi</i>	Metazoa	AdarC3	http://metazoa.ensembl.org/Anopheles_darlingi/info/Index	10.1093/nar/gk1484
<i>Anopheles gambiae</i>	Metazoa	AgamP4	http://metazoa.ensembl.org/Anopheles_gambiae/info/Index	10.1126/science.1076181
<i>Anoplophora glabripennis</i>	Metazoa	Agla_1.0	http://metazoa.ensembl.org/Anoplophora_glabripennis/info/Index	10.1186/s13059-016-1006-8
<i>Apis mellifera</i>	Metazoa	Amel_4.5	http://metazoa.ensembl.org/Apis_mellifera/info/Index	10.1038/nature05260
<i>Atta cephalotes</i>	Metazoa	Attacep1.0	http://metazoa.ensembl.org/Atta_cephalotes/info/Index	10.1371/journal.pgen.1002007
<i>Belgica antarctica</i>	Metazoa	ASM77530v1	http://metazoa.ensembl.org/Belgica_antarctica/info/Index	10.1038/ncomms5611

Figure 19. The “Genomes” menu of LIRBase.