Overview:

- **DSAP** is an useful tool which can cope with the large small-RNAs sequencing data generated by solexa sequencer through the web and return a user friendly analysis report quickly. DASP also provide a mature miRNAs comparison function between individual jobs. It provides a very convenient way to analyze up/down regulation of miRNAs at different develop stages or evolutional conservation of mature miRNAs cross-species using a color scaling matrix. Furthermore, DSAP also accept experimental results from other miRNA expression analysis such as stem-loop real-time PCR, microarray or SOLiD sequencing in a tab-delimited format.

How to use DSAP:

(i) Input file

A single Solexa sequencing run produces two kinds of data:

1. **Raw data** (a FASTQ file contains an identifier, sequence reads and quality values for each base). The sizes of FASTQ files are usually in the gigabytes, which is not suitable for sending over the web.
2. **Sequence tags** (a tab-delimited file which holds only the unique sequence read (tag) and its corresponding number of copies).

A Biopieces' script is available to transform the FASTQ file into unique sequence tags.

Biopieces' command**:** (read_fastq -i **INPUT.fastq** | uniq_seq -c | sort_records -r -k SEQ_COUNTn | write_tab -k SEQ_COUNT,SEQ -xo **OUTPUT.tag**)

DSAP takes a sequence tag file under 300Mb as input material. (Must be same format as Figure1.)

read counts	sequence tags
260135	TGGAATGTAAAGAAGTATGTATTCGTATGCCGT
213816	TGAGGTAGTAGGTTGTATAGTTTCGTATGCCGT
151369	TGAGGTAGTAGGTTGTATGGTTTCGTATGCCGT
115834	TGAGGTAGTAGATTGTATAGTTTCGTATGCCGT
108066	ACAGTAGTCTGCACATTGGTTATCGTATGCCGT
86571	AGCAGCATTGTACAGGGCTATGATCGTATGCCG
69513	TGGAATGTAAGGAAGTGTGTGGTCGTATGCCGT
50634	TGGAGTGTGACAATGGTGTTTGTCGTATGCCGT
48601	ACAGTAGTCTGCACATTGGTTTCGTATGCCGTC
45918	TCTTTGGTTATCTAGCTGTATGATCGTATGCCG
40744	CAGGCTGGTTAGATGGTTGTCTTCGTATGCCGT
37324	TTAAGACTTGTAGTGATGTTTATCGTATGCCGT
35667	TCACAGTGAACCGGTCTCTTTTCGTATGCCGTC
34836	AATTGCACGGTATCCATCTGTATCGTATGCCGT
31107	AGCAGCATTGTACAGGGCTATCATCGTATGCCG
30241	AACATTCATTGCTGTCGGTGGGTTTCGTATGCC
28698	GAGGAAGAAGGAATATTTTTCGTATGCCGTCTT
26819	TGAGGTAGTAGTTTGTACAGTTTCGTATGCCGT
25285	AACTCTTAGCGGTGGATCACTCGTCGTATGCCG
25054	TACCCTGTAGATCCGAATTTGTTCGTATGCCGT
24136	TACCACAGGGTAGAACCACGGACTCGTATGCCG
23380	AGCTACATTGTCTGCTGGGTTTCTCGTATGCCG
23089	TCAGTGCATCACAGAACTTGGTTCGTATGCCGT
22845	ACCACAGGGTAGAACCACGGACTCGTATGCCGT

Figure 1: Input file format

(ii) parameters

- **Choose species:** The user can upload a sequence tag file **under 300 Mb** and then choose from among 115 species, or the use the default of all species if the organism is not listed.(as shown in **Figure 2.**)
- **Do not consider adaptor sequences:** Remember to click the checkbox if you have an adaptor-removed sequence tags file or DSAP will consider it as an adaptor-contained file, discard tags without a reliable 3-adaptor and skip tags <16nt after the removal of 3-adaptor. Solexa use standard 3- and 5-adaptor sequences in their small RNA library preparation kit so it is not necessary for the user to upload the adaptor sequences.
- **Do not consider poly-A, T, C, G:** The user can choose whether to remove continous poly-A/T/C/G reads in the cleanup step since lots of poly-A, T, C, G or N sequences came from sequencing errors.
- **Use test dataset:** We provided a sequence tag file with 329,334 tags as a test dataset.

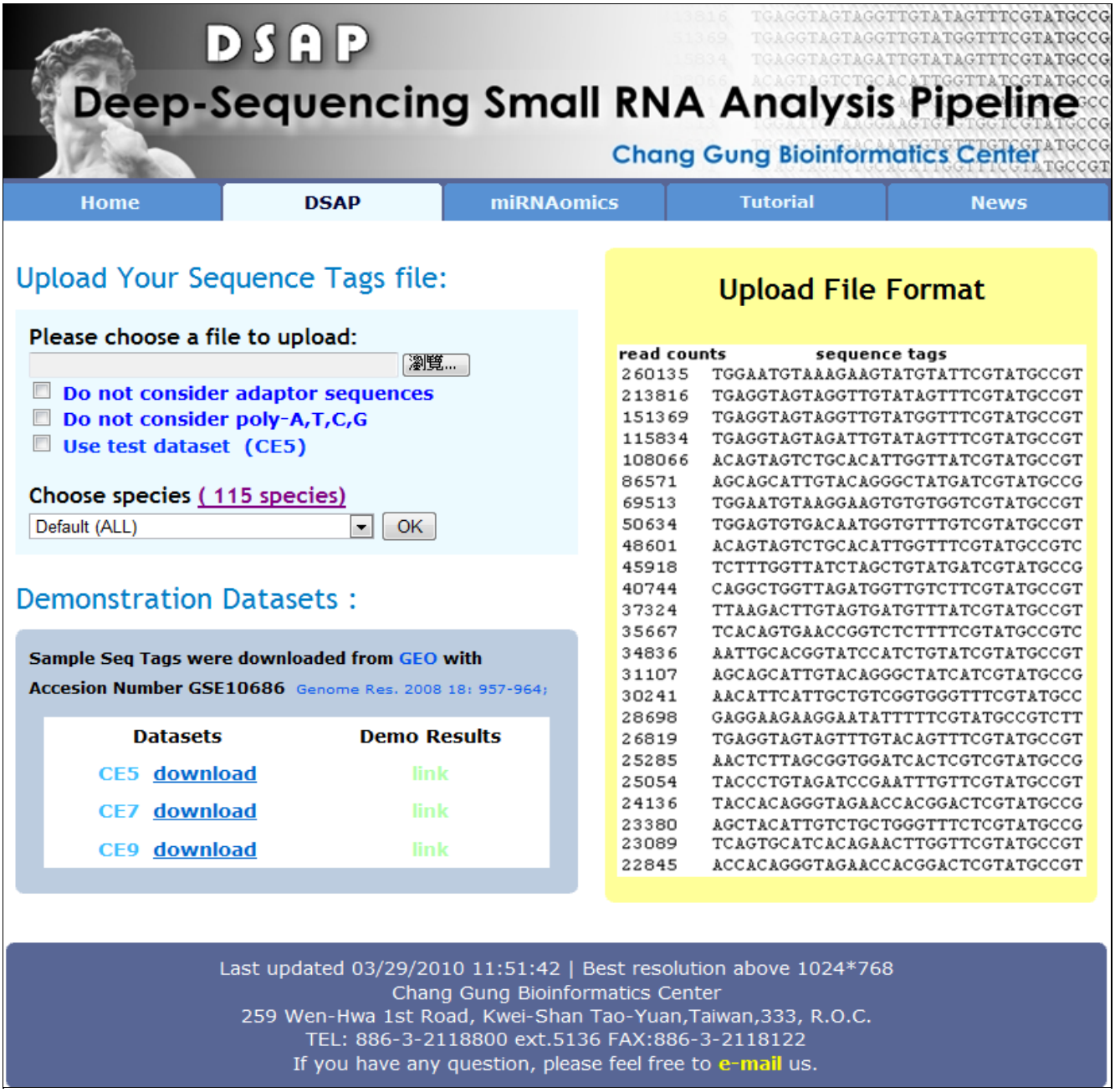


Figure 2: DSAP input page

Description of Output:

- The output page is composed of several blocks which represents the analysis workflow of DSAP

  (i) job status

  (ii) cleanup

  (iii) clustering

  (iv) non-coding RNA matching

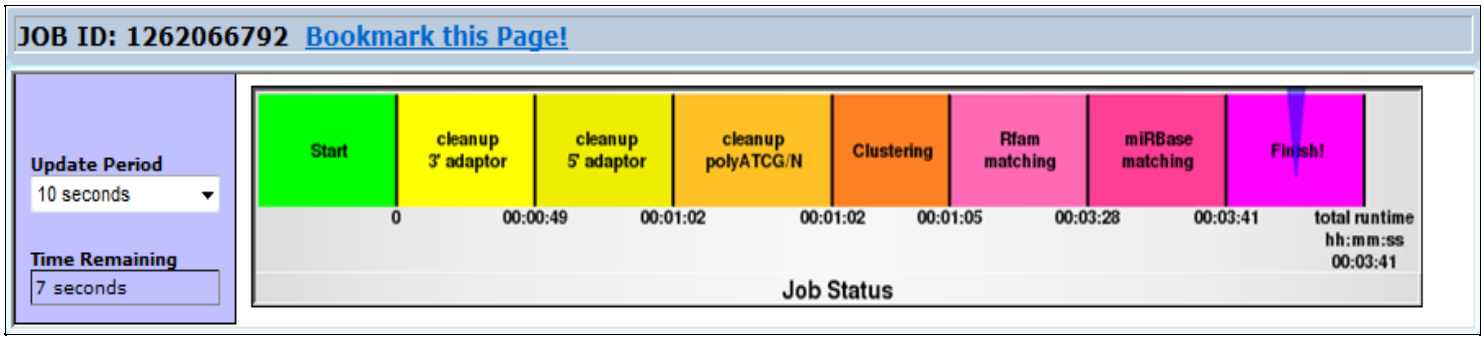
  (v) known microRNA matching

  (vi) summary of job

  (vii) comparative microRNAs analysis

(i) job status

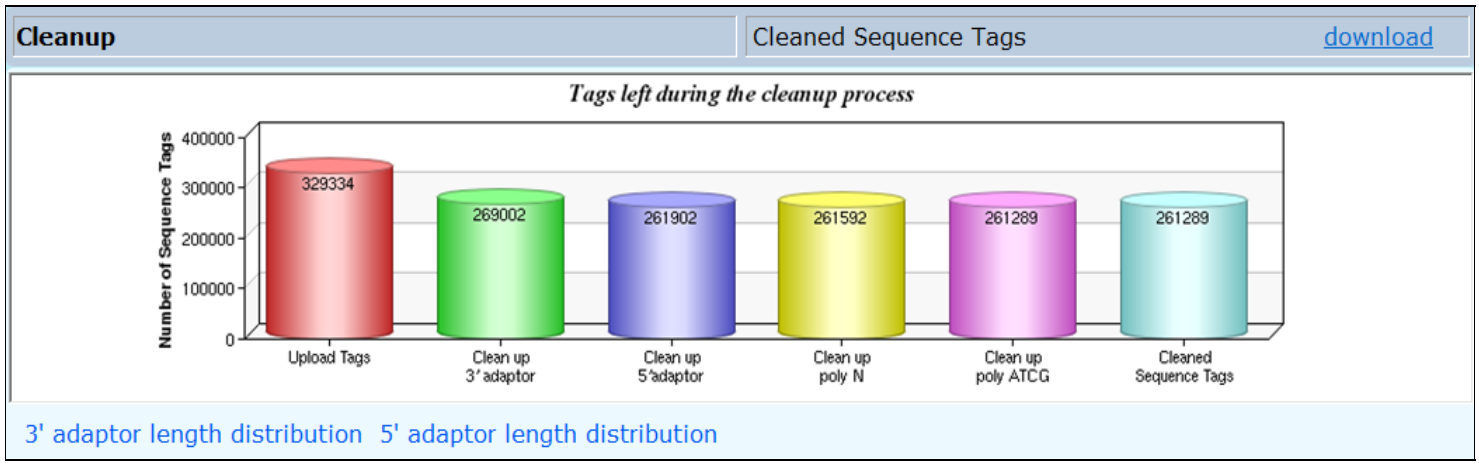
After successful upload, the web server will return a page using timestamp as identifier (JOB ID). Job status can be monitored by a real-time meter graph which contains exact run time of each step. Besides, users can bookmark this web page for future reference.

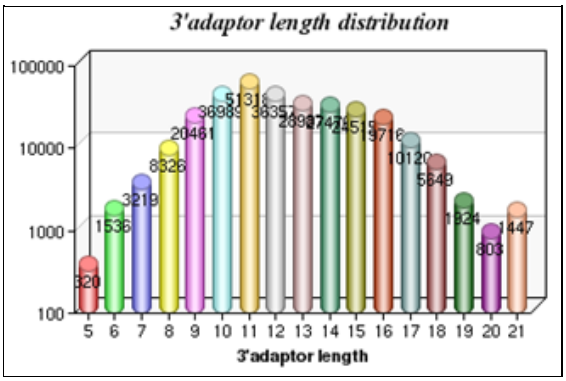


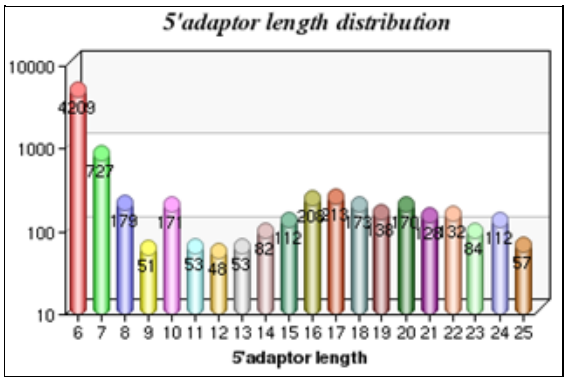
(ii) cleanup

A bar chart dynamically recording the number of sequence tags during the cleanup process. It also provides a link to detailed information about the length distribution of attached adaptors.

Cleaned Sequence Tags (in FASTA format) are available through the download link.

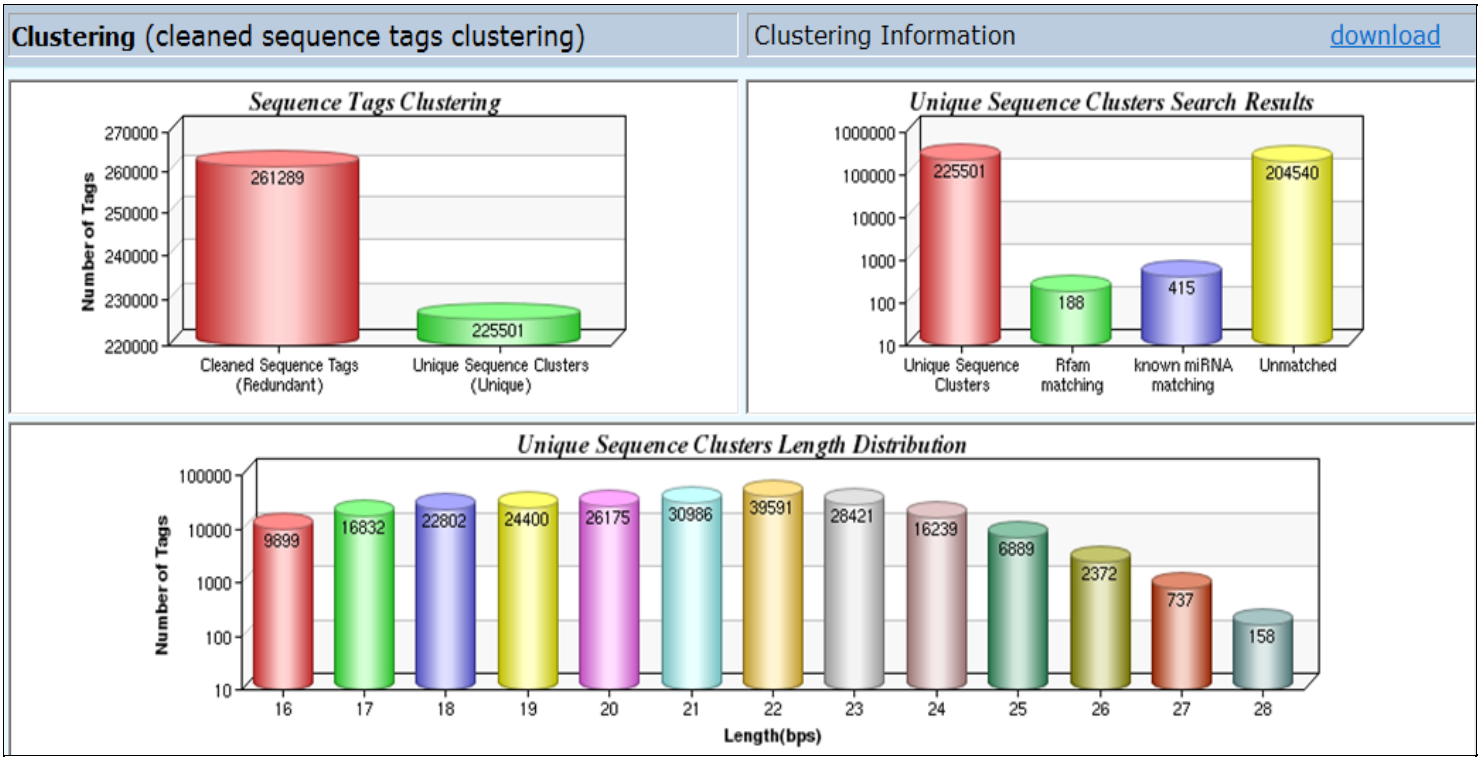






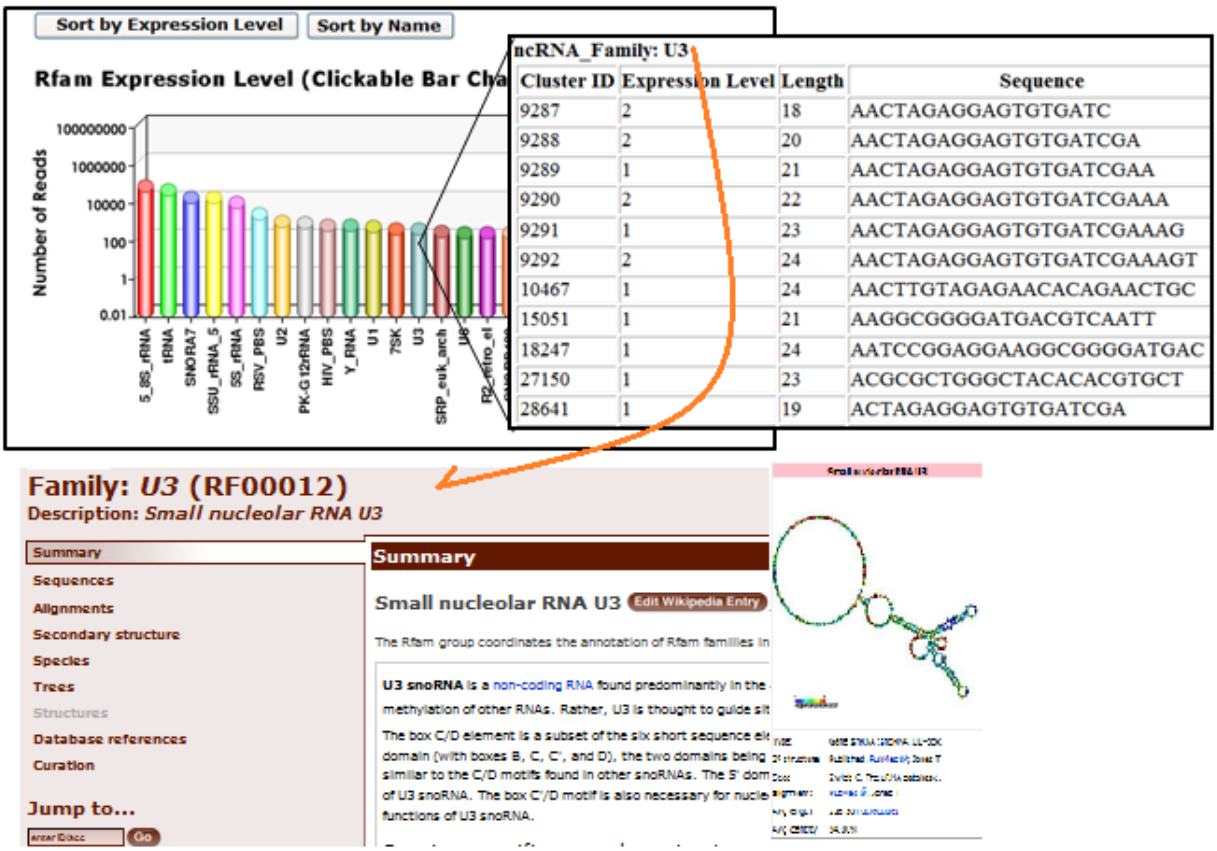
(iii) clustering

This block shows the clustering state of the cleaned sequence tags and provides each unique sequence cluster and its member information in a tab-delimited file

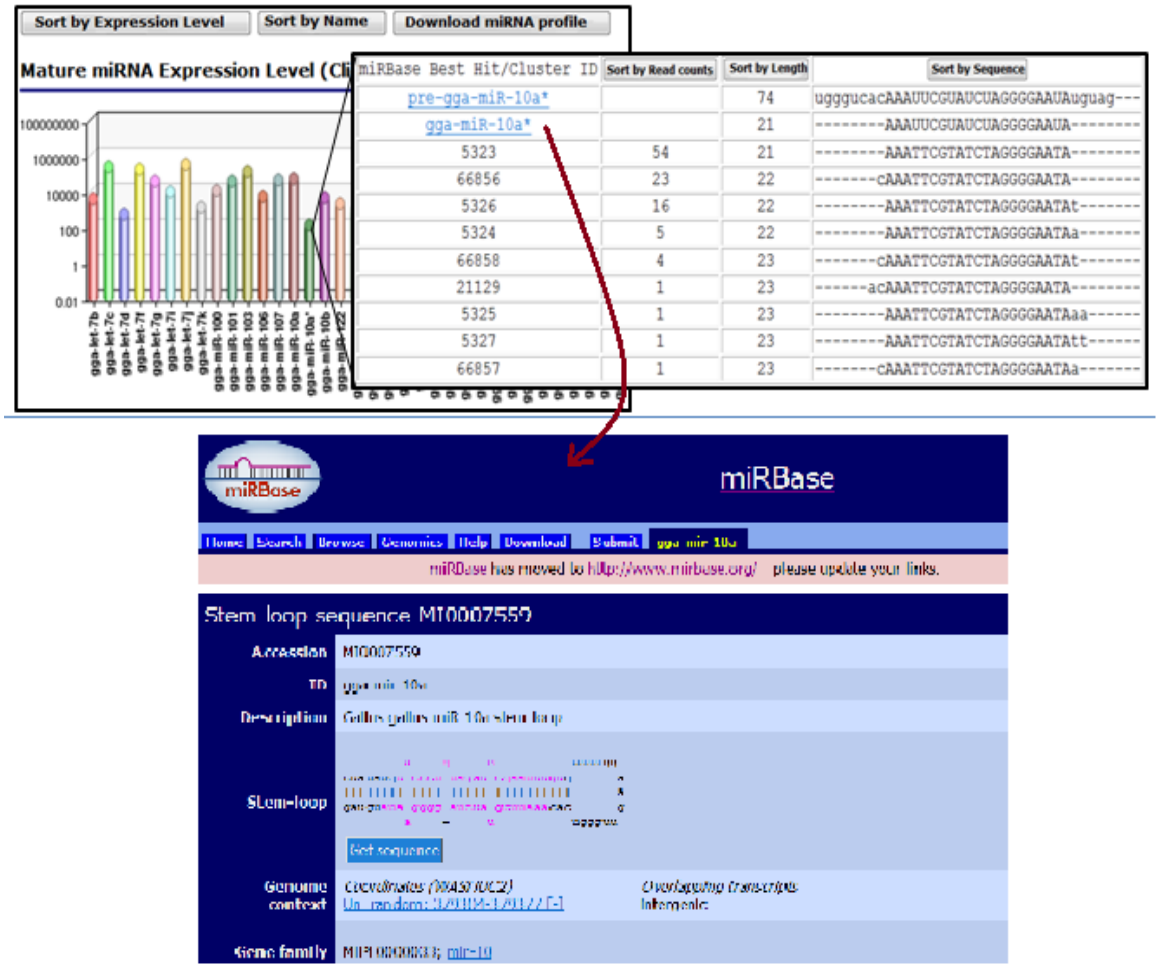


(iv)non-coding RNA matching, (v) known microRNA matching

The fourth and fifth blocks summarized the results of the unique tag clusters matched to Rfam and miRBase respectively. Each matched RNA family and its related expression level was summarized in a multi-colour clickable bar chart which linked to external database such as miRBase for further detail information. All the results were downloadable from the website in a tab-delimited text file. Representative sequence tags failed to be identified from the known microRNAs matching step can be downloaded for the identification of putative novel miRNAs



#The alignment of unique sequence clusters with the corresponding miRNA hairpin is optimized for the observation of isomiRs.



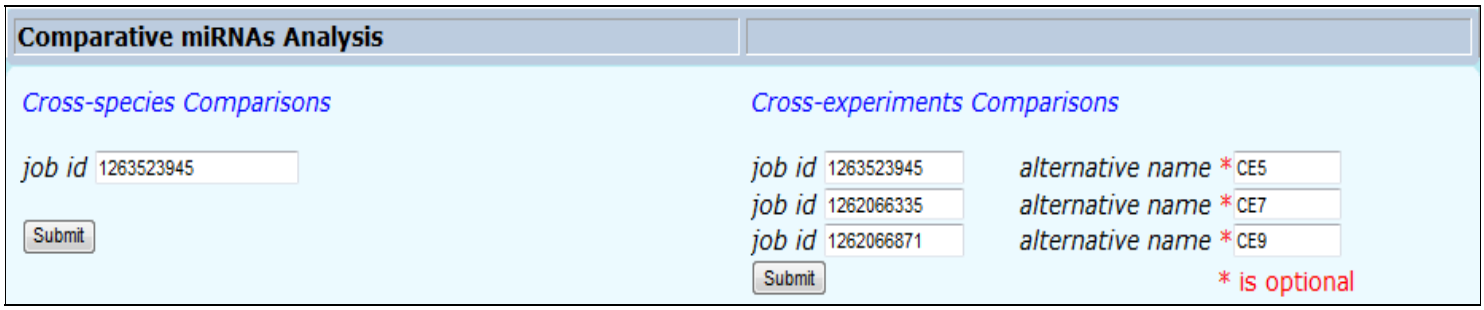
Unmatched Sequence Tags	putative small RNAs/novel miRNAs	download

(vi) summary of job

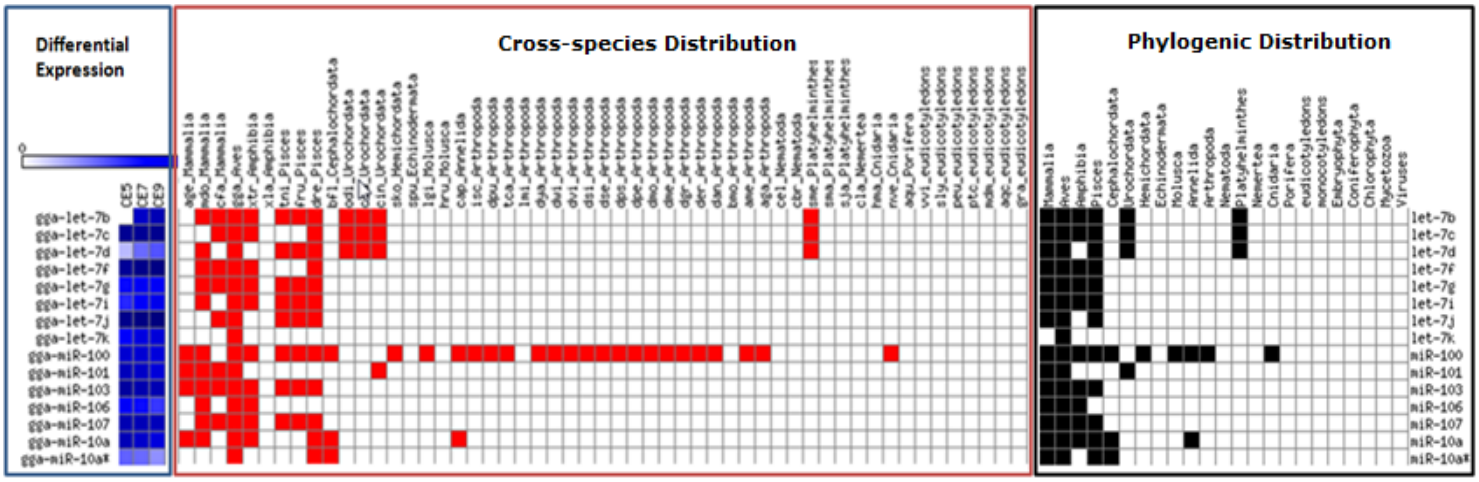
JOB_ID	1269781273
No. of total reads	3257227
No. of sequence tags	329334
No. of tags after 3'adaptor clean up	269002
No. of tags after 5'adaptor Clean up	261902
No. of tags after poly N cleanup	261592
No. of tags after poly-ATCG Clean up	261289
No. of cleaned sequence tags	261289
No. of reads in cleaned sequence tags	3076497
Percentage of reliable reads #	94.45
#Percentage of reliable reads=(No. of reads in cleaned sequence tags/No. of total reads)*100	
No. of Unique Sequence Clusters (USC)	225501
No. of matched ncRNA in Rfam	188
No. of USC matched to Rfam	18406
No. of total reads matched to Rfam	95979
Percentage of reads matched to Rfam	3.12
No. of matched miRNA in miRBase	415
No. of USC matched to miRBase	2555
No. of total reads matched to miRBase	1687442
Percentage of reads matched to miRNA	54.85
USC Unmatched	204540

(vii) comparative microRNAs analysis

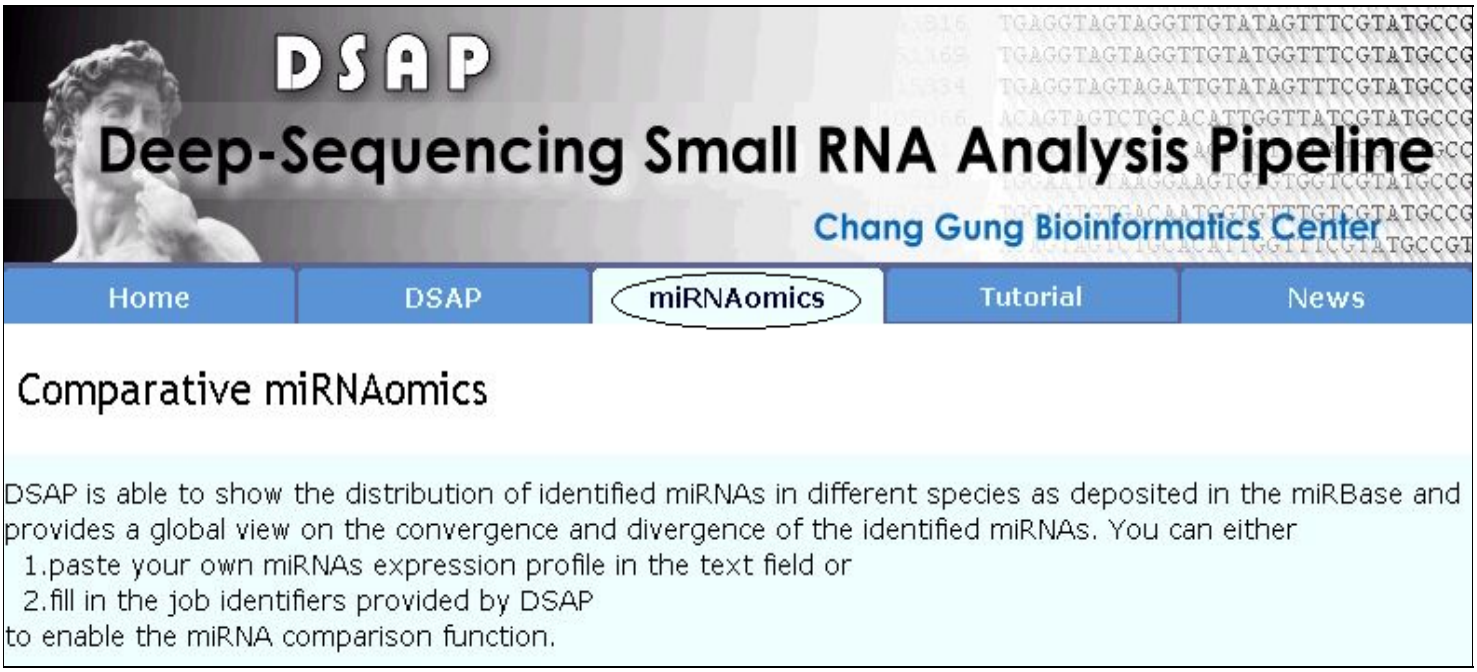
DSAP is capable of displaying different microRNAs expression levels from different jobs using a log2-transformed color matrix. Furthermore, a cross-species comparative function is also provided to show the distribution of identified miRNAs in different species as deposited in the miRBase.



Demonstration output



miRNAomics Page:



1.Paste your own miRNA expression profiles

miR	CE5	CE7	CE9
gga-let-7b	1	12352	29903
gga-let-7c	163164	156621	198263
gga-let-7d	6	44	81
gga-let-7f	128541	255562	346444
gga-let-7g	532	1156	1075
gga-let-7i	250	936	1877
gga-let-7j	232980	373519	446116
gga-let-7k	837	3414	4844
gga-miR-100	7358	6871	4168
gga-miR-101	23497	12104	4654
gga-miR-103	89271	41329	34030
gga-miR-106	996	493	167
gga-miR-107	32021	27360	23079
gga-miR-10a	26443	9883	4075
gga-miR-10a*	54	44	17

Submit Reset

DSAP **accept** users' own miRNA profiles from **different experiment methods** such as **stem-loop real-time PCR, microarray or SOLiD sequencing** , but should avoid file format problems with cautions.

- The input files are **space** or **tab-delimited**.
- Zero or Missing values should be reaplaced with a dash "**-**". Zero will not work because we use a log2 transformation.
- You can have miRNA descriptors as leftmost column. All other columns must be your numeric data (read counts) and need headings to label individual experiments.
- The parser uses the header to figure out how many experiments you have, so never skip the upper left corner string or it will appear that you have extra data, resulting in an error message.

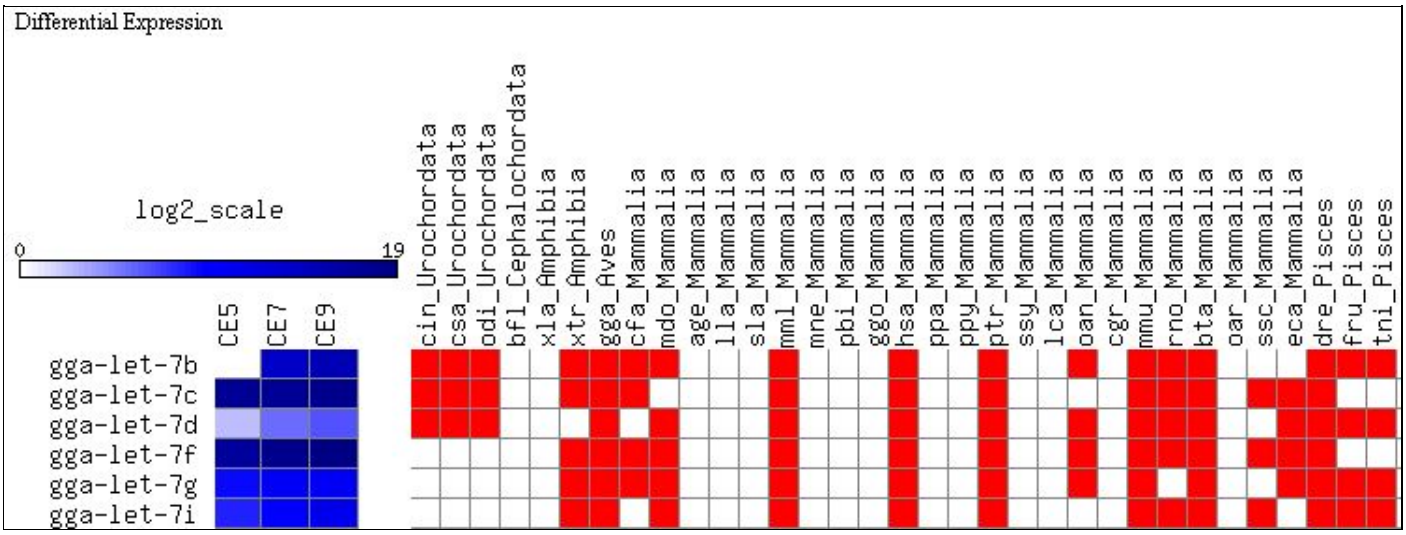
anything other than blank	Experiment_1	Experiment_2	Experiment_3	Experiment_4	...
gga-let-7a	-	-	-	...	...
gga-let-7a*	-	-	-	...	...
gga-let-7b	3016	13464	13464	...	...
gga-let-7c	171231	162866	162866	...	...
gga-let-7d	378	1195	1195	...	...
gga-let-7d*	-	-	-	...	...
gga-let-7e	-	-	-	...	...
gga-let-7f	133033	263496	263496	...	...
gga-let-7g	30778	35464	35464	...	...
gga-let-7i	7164	12892	12892	...	...
gga-let-7j	246460	392646	392646	...	...
gga-let-7k	970	3736	3736	...	...
gga-miR-1	-	-	-	...	...
gga-miR-100	8508	7623	7623	...	...
gga-miR-101	27615	14015	14015	...	...
gga-miR-101*	-	-	-	...	...
gga-miR-103	93461	42635	42635	...	...
gga-miR-103-2*	-	-	-	...	...
gga-miR-103-as	-	-	-	...	...
gga-miR-105	-	-	-	...	...
gga-miR-106	4097	1550	1550	...	...
gga-miR-106a	-	-	-	...	...
gga-miR-106a*	-	-	-	...	...
gga-miR-106b	-	-	-	...	...
gga-miR-106b*	-	-	-	...	...
gga-miR-107	32263	27562	27562	...	...

heading

miRNA descriptor

Read counts

Demonstration output:

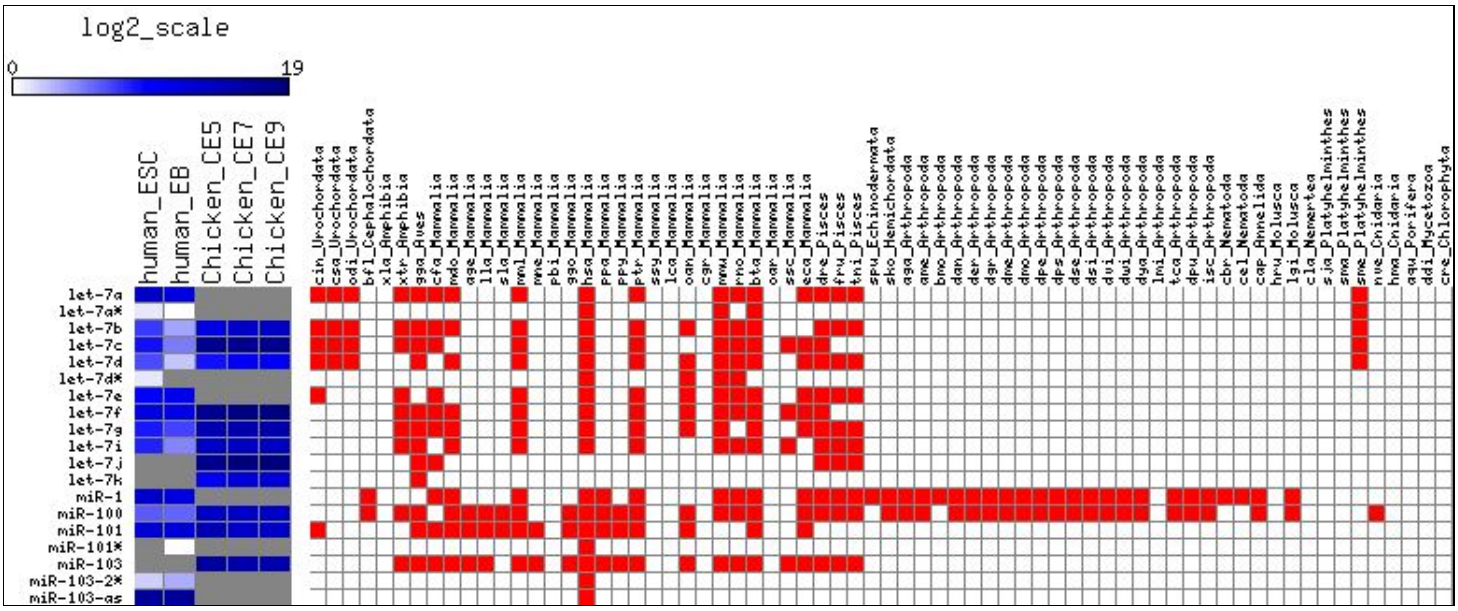


2.Fill in the job ids provided by DSAP (max. 5)

JOB ID1	1258095940	**Alternative Name** (optional)	human ESC
JOB ID2	1258096199	**Alternative Name** (optional)	human EB
JOB ID3	1256192842	**Alternative Name** (optional)	Chicken CE5
JOB ID4	1256195990	**Alternative Name** (optional)	Chicken CE7
JOB ID5	1256196016	**Alternative Name** (optional)	Chicken CE9

Submit

Demonstration output:



Last updated Wed, 08 Dec 2010 16:11:19 GMT | Best resolution above 1024*768

Chang Gung Bioinformatics Center
259 Wen-Hwa 1st Road, Kwei-Shan Tao-Yuan,Taiwan,333, R.O.C.
TEL: 886-3-2118800 ext.5136 FAX:886-3-2118122
If you have any question, please feel free to e-mail us.