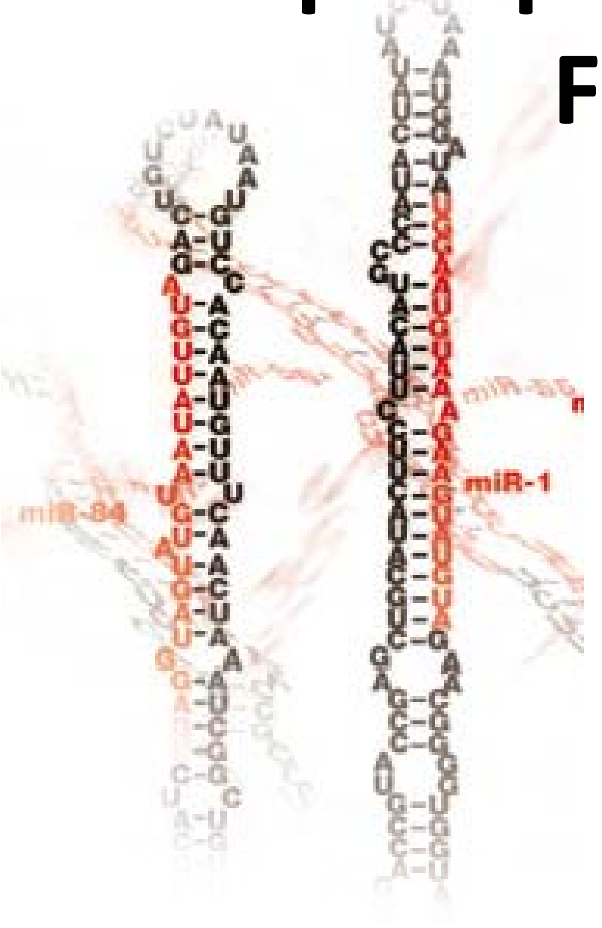
Deep-sequencing Small RNA analysis in Flagellates & Barley



Po-Jung Huang, Ph.D.

Chang Gung Bioinformatics Center,
Chang Gung Molecular Medicine
Research Center

Discovery of microRNA

C. elegans

lin-4 (*Lee R et al. 1993*)
let-7 (*Reihhart et al. 2000*)

Victor Ambros

First miRNA - *C. elegans* lin-4 in 1993 by Victor Ambros & Lee

lin-4 causes temporal decreasing in LIN-14 protein expressed in the first larval stage (L1).

"The Caenorhabditis elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14". CELL (1993)

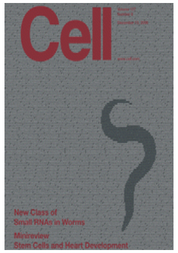
Frank Slack

Second miRNA - *C. elegans* let-7 in 2000 by Frank Slack

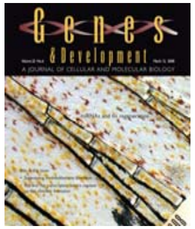
miRNA might be conservative in organic evolution and plays an important role in regulation of organic development.

"The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans" . NATURE (2000)

miRNAs host genes and transcription

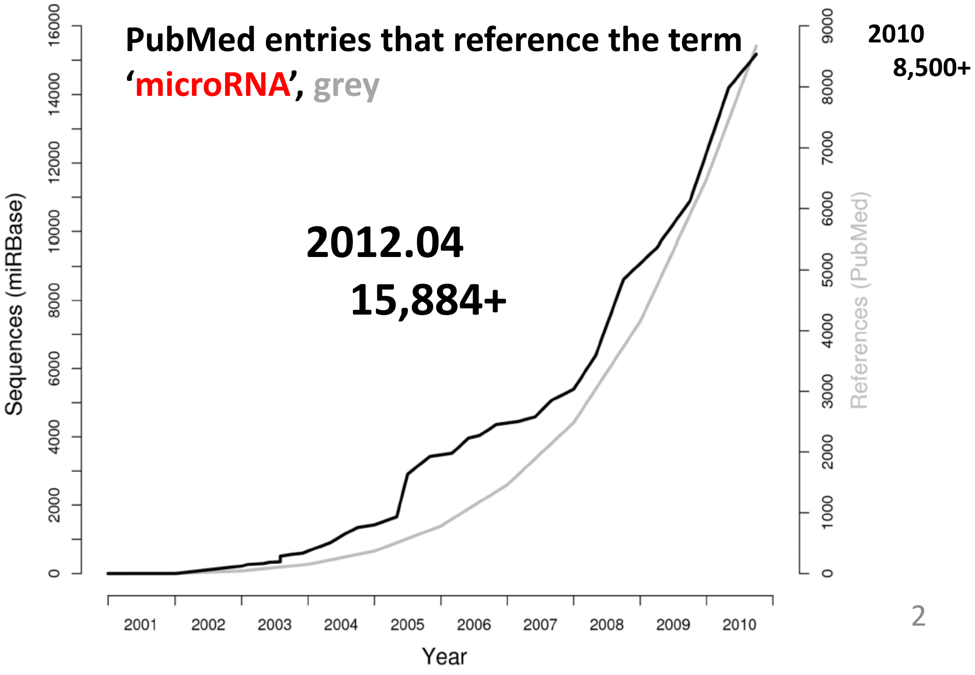


New class of small RNAs in worms



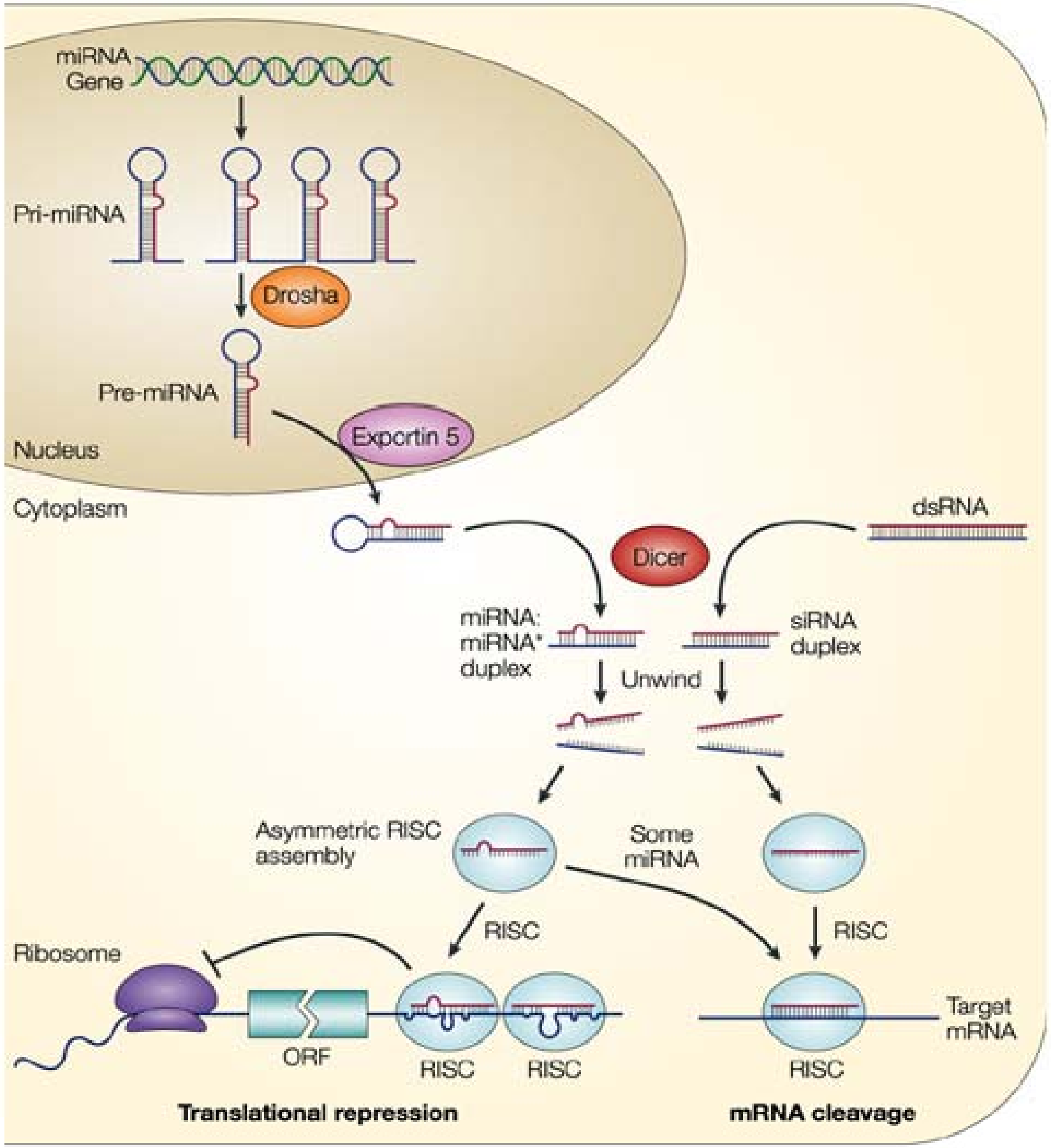
miRNAs and fin development

Stem cell division



PubMed entries that reference the term 'microRNA', grey

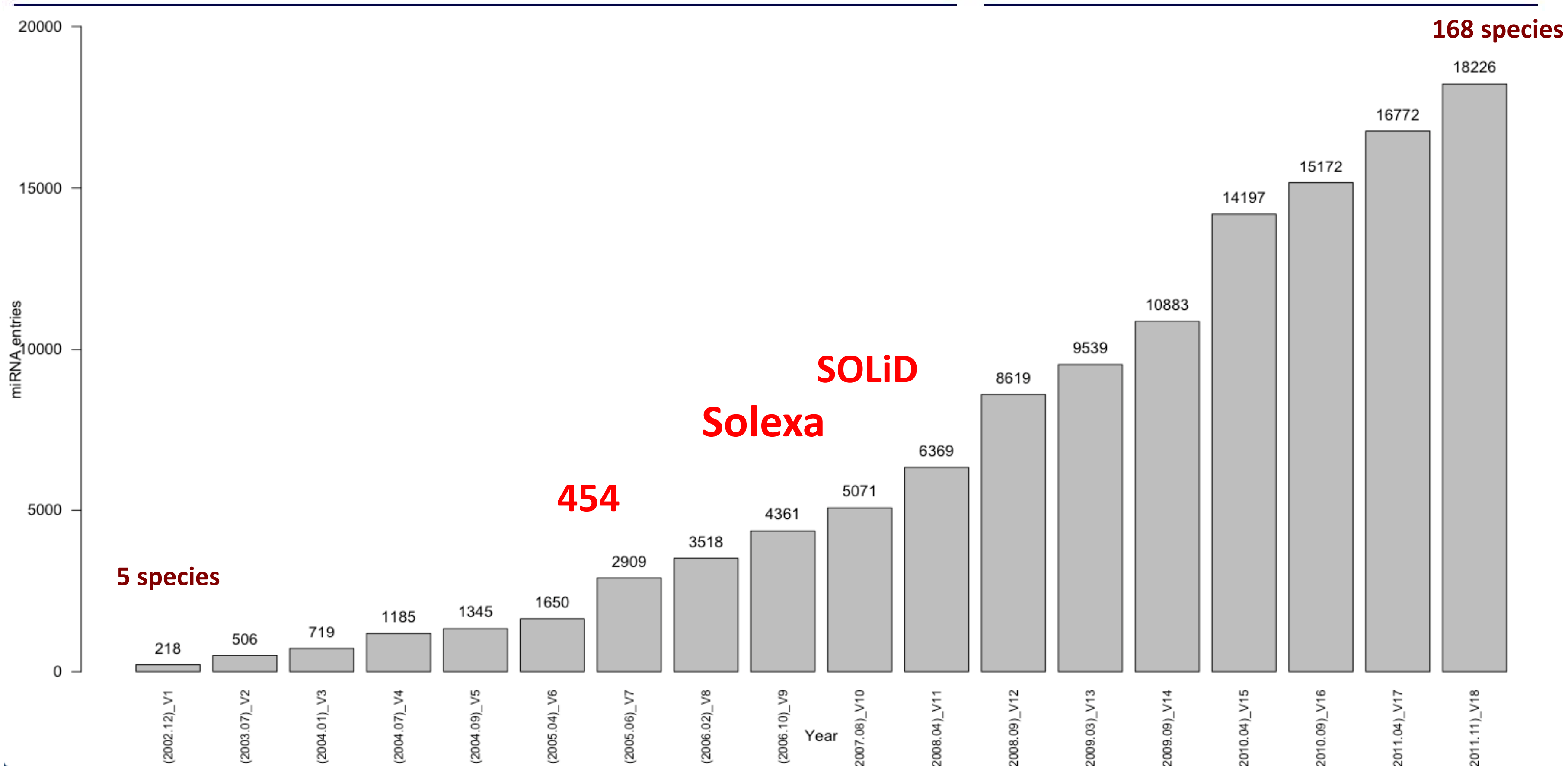
Mechanism of miRNA Regulation



Nature Reviews | Genetics

Current Methods for studying miRNAs

	Northern	miRAGE	qRT-PCR	Microarrays	NGS
Starting material	5-25 µg	1mg	1-10ng	~5µg	1-10µg
Sensitivity	Low	Low	High	Moderate	High
Specificity	Low	Low	High	Moderate	High
Dynamic Range	~2 logs	presence/ absence	7 logs	< 4 logs	Broad range
Time to results	Several days	Several days	Hours	Several days	3 days
High throughput	No	No	No	**Yes**	**Yes**
Novel miRNA identification	Yes	Yes	Yes	No	**Yes**
Reliably distinguishes mature miRNA from precursor	Yes	No	Yes	No	Yes illumina AB applied biosystems



168 species

5 species

454

Solexa

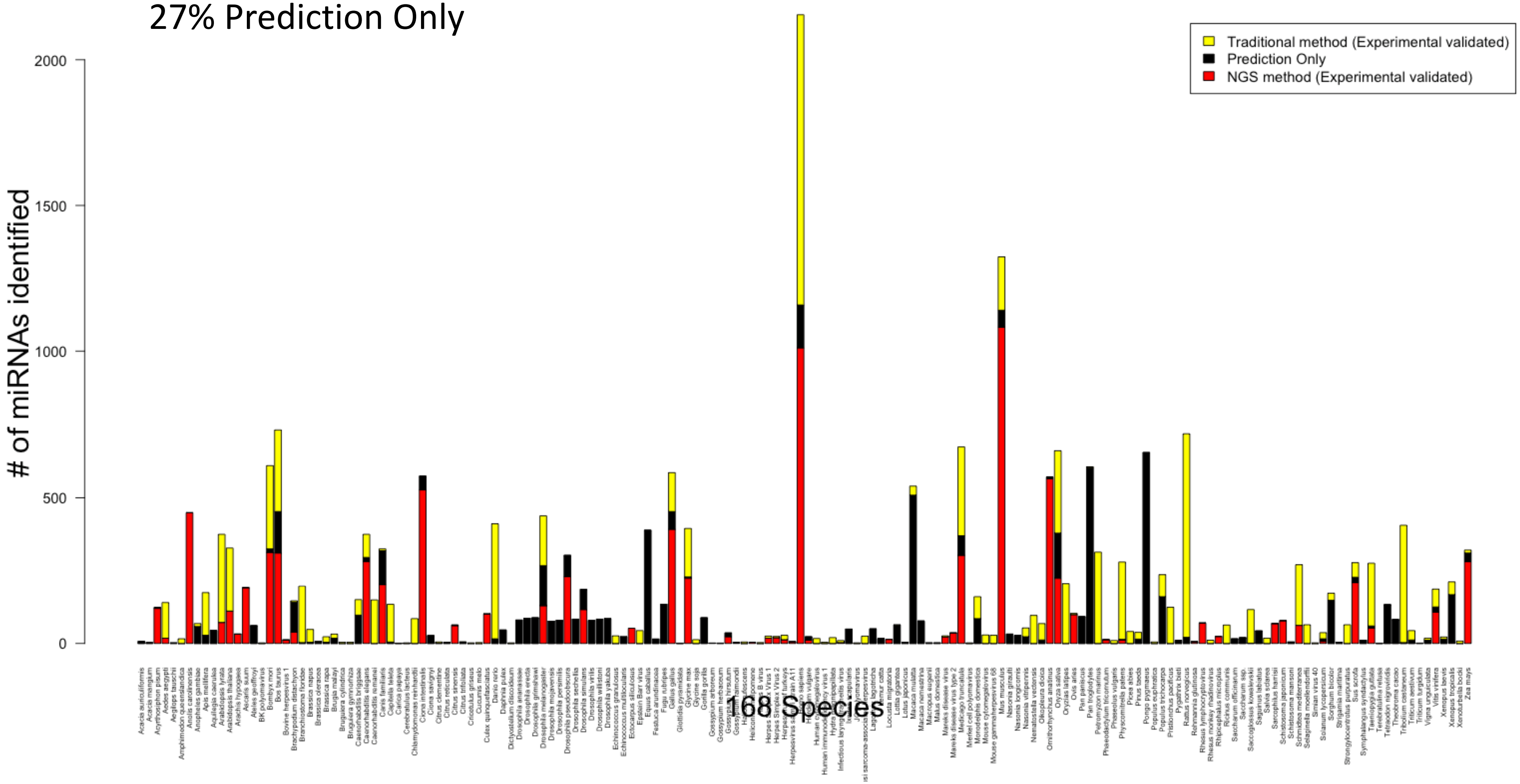
SOLiD

2002 2005 2006 2007 2011

NGS Validated miRNAs

73% Experimental Validated (**~50% NGS**)

27% Prediction Only



miRBase

Latest miRBase blog posts

miRBase website "at risk", Thu 10th to Fri 18th Nov By sam (November 8, 2011)

Due to server room refurbishment, the miRBase website may experience some instability between Thu 10th and Fri 18th November 2011. The plan is for just 30 minutes or so down time at either end of that period, but the website should be considered "at risk" throughout. Apologies for any inconvenience.

miRBase 18 released By sam (November 3, 2011)

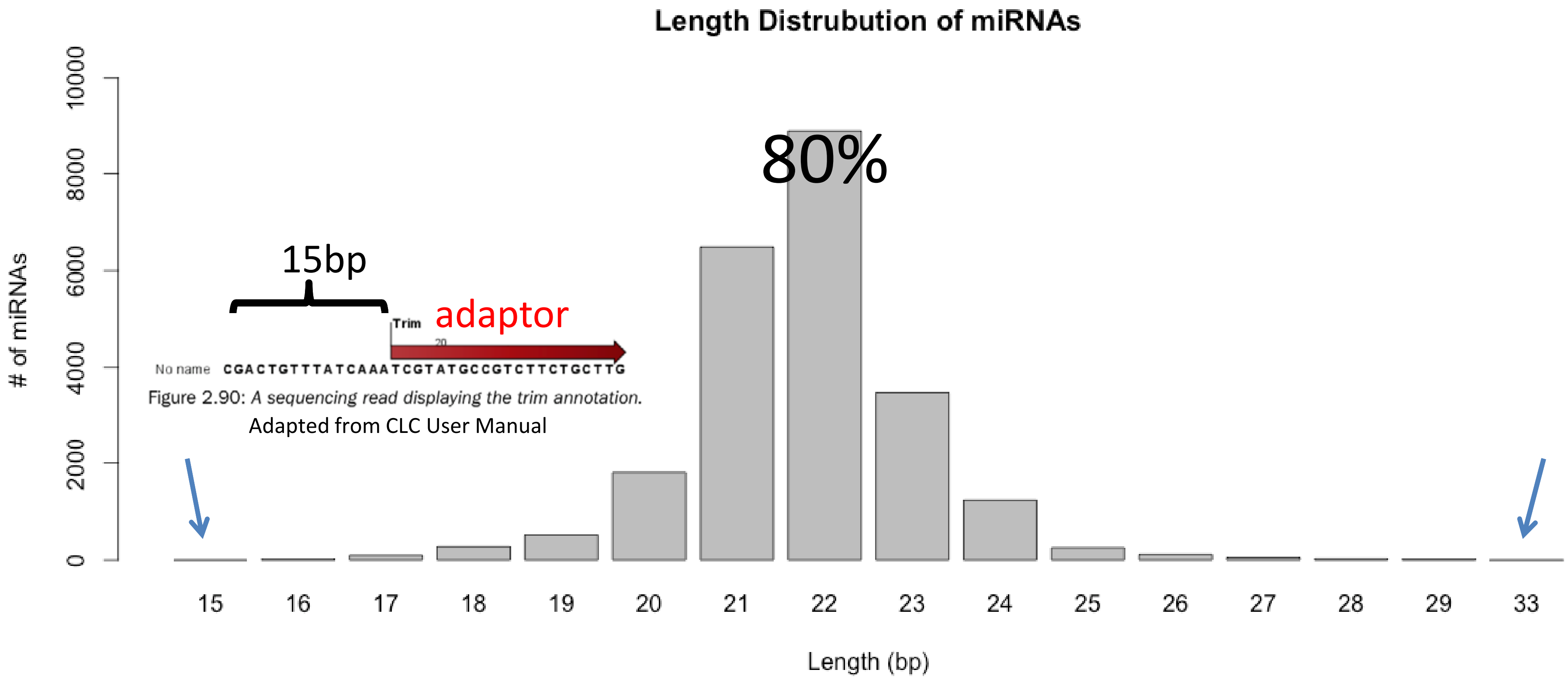
After a little more pain than usual, miRBase 18 is finally released. The database contains 18226 entries representing hairpin precursor miRNAs, expressing 21643 mature miRNA products, in 168 species. That represents 1488 new hairpin

miRNA count: 18226 entries

Release 18: November 2011

Search by miRNA name or keyword

Download published miRNA data



Length Distrubution of miRNAs

Figure 2.90: *A sequencing read displaying the trim annotation.*

Adapted from CLC User Manual

Analysis Tools for small-RNA Deep Sequencing

LINUX/UNIX BASED TOOLS

miRDeep Discovering microRNAs from deep sequencing data
Nature Biotechnology (2008) 26 (4), pp. 407-415

MIRExpress: Analyzing high-throughput sequencing data for profiling microRNA expression
BMC Bioinformatics (2009) 10, art. no. 1471, pp. 328 .

SeqBuster, a bioinformatic tool for the processing and analysis of small RNAs datasets, reveals ubiquitous miRNA modifications in human embryonic cells.
Nucleic acids research (2010) 38 (5), pp. e34

MIReNA: Finding microRNAs with high accuracy and no learning at genome scale and from deep sequencing data
Bioinformatics (2010) 26 (18), art. no. btq329, pp. 2226-2234

MiRNAkey: A software pipeline for the analysis of microRNA Deep Sequencing data.
Bioinformatics (2010) Aug 27

WEB-BASED TOOLS

miRCat: A toolkit for analysing large-scale plant small RNA datasets.
Bioinformatics (2008) 24 (19): 2252-2253 Application Note

miRanalyzer: A microRNA detection and analysis tool for next-generation sequencing experiments
Nucleic Acids Research (2009) 37 (SUPPL. 2), pp. W68-W76

DSAP: deep-sequencing small RNA analysis pipeline
Nucleic Acids Research. (2010) 38(Web Server issue): W385–W391.

mirTools: microRNA profiling and discovery based on high-throughput sequencing
Nucleic Acids Research. (2010) 38(Web Server issue): W392–W397

miRanalyzer2: an update on the detection and analysis of microRNAs in high-throughput sequencing experiments
Nucleic Acids Research. (2011) 39 (11)

Commercial TOOLS

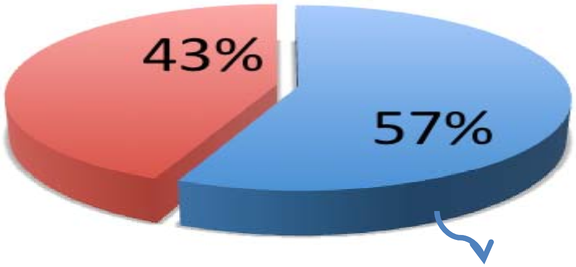
CLC Genomics Workbench, Genomatix, InteRNA, Partek, and Avadis NGS

Comparison of small RNA Analysis Tools

Tools	miRanalyzer	miRExpress	miRDeep	DSAP	CLC	Avadis NGS
Version	v1,v2		v1, v2	v1, v2	5.1	
No-limit on Data size		✔	✔	✔	✔	✔
No-limit on Organisms	11, 36 species	✔	Genomic seqs required	✔	✔	✔
Classification of miRNA	✔	✔	✔	✔	✔	✔
Classification of small RNA	✔		✖✔	✔	✔	✔
Phylogenetic distribution of miRNAs				✔		
Quantification (DGE)	✖✔	✔	✖✔	✔	✔	✔
Novel Small RNA Discovery	✔		✔	✖✔	✔	✔
miRNA Target Analysis				✖✔		✔
Go Analysis				✖✔	✔	✔

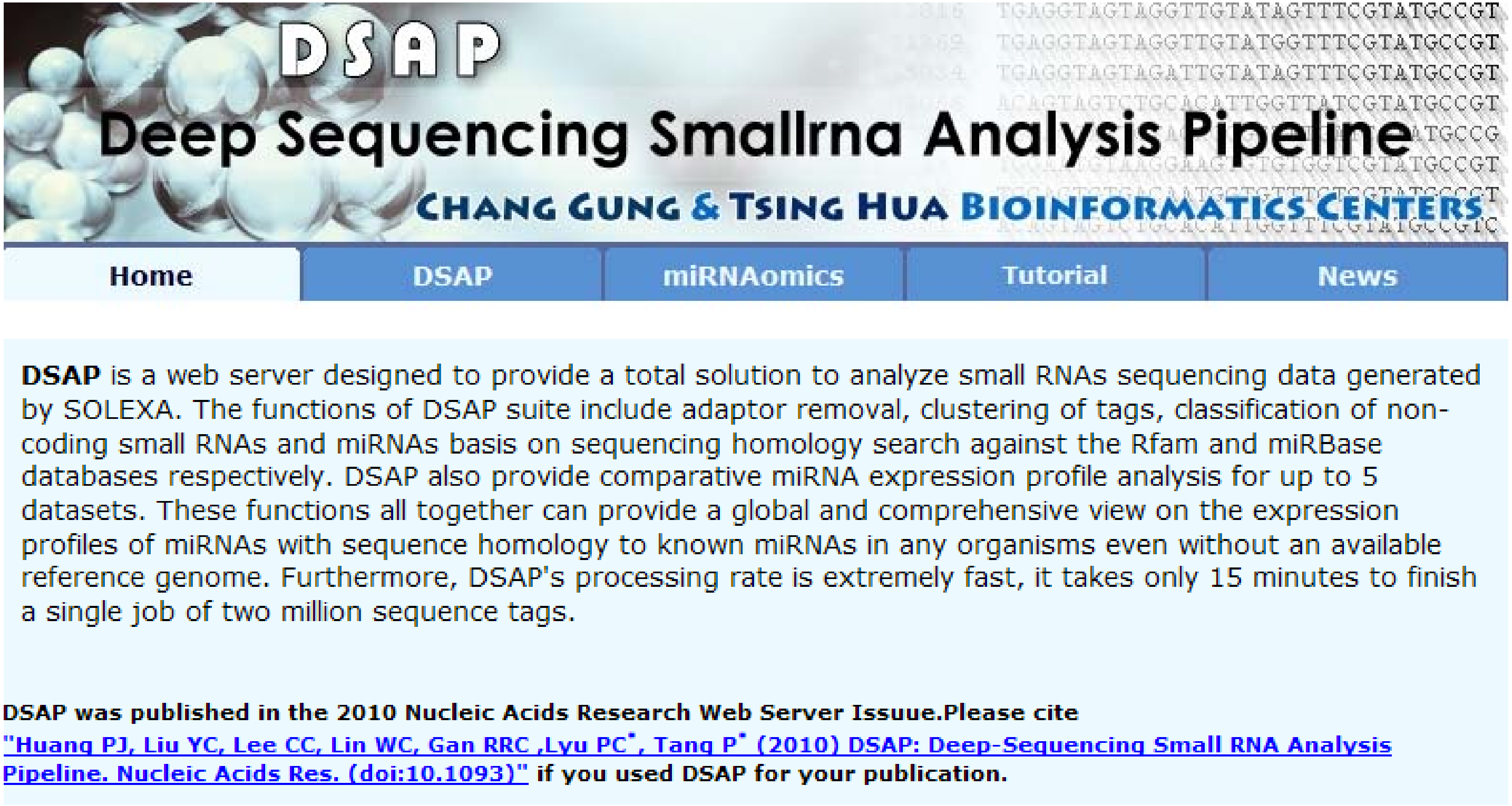
Release 18.0 (11 Nov 2011)

168 species



WITHOUT
a Reference
genome

http://dsap.cgu.edu.tw



DSAP is a web server designed to provide a total solution to analyze small RNAs sequencing data generated by SOLEXA. The functions of DSAP suite include adaptor removal, clustering of tags, classification of non-coding small RNAs and miRNAs basis on sequencing homology search against the Rfam and miRBase databases respectively. DSAP also provide comparative miRNA expression profile analysis for up to 5 datasets. These functions all together can provide a global and comprehensive view on the expression profiles of miRNAs with sequence homology to known miRNAs in any organisms even without an available reference genome. Furthermore, DSAP's processing rate is extremely fast, it takes only 15 minutes to finish a single job of two million sequence tags.

DSAP was published in the 2010 Nucleic Acids Research Web Server Issuue.Please cite "Huang PJ, Liu YC, Lee CC, Lin WC, Gan RRC ,Lyu PC*, Tang P* (2010) DSAP: Deep-Sequencing Small RNA Analysis Pipeline. Nucleic Acids Res. (doi:10.1093)" if you used DSAP for your publication.

Workflow

Clean up

↓

Clustering

↓

ncRNA matching

(Rfam V10.1)

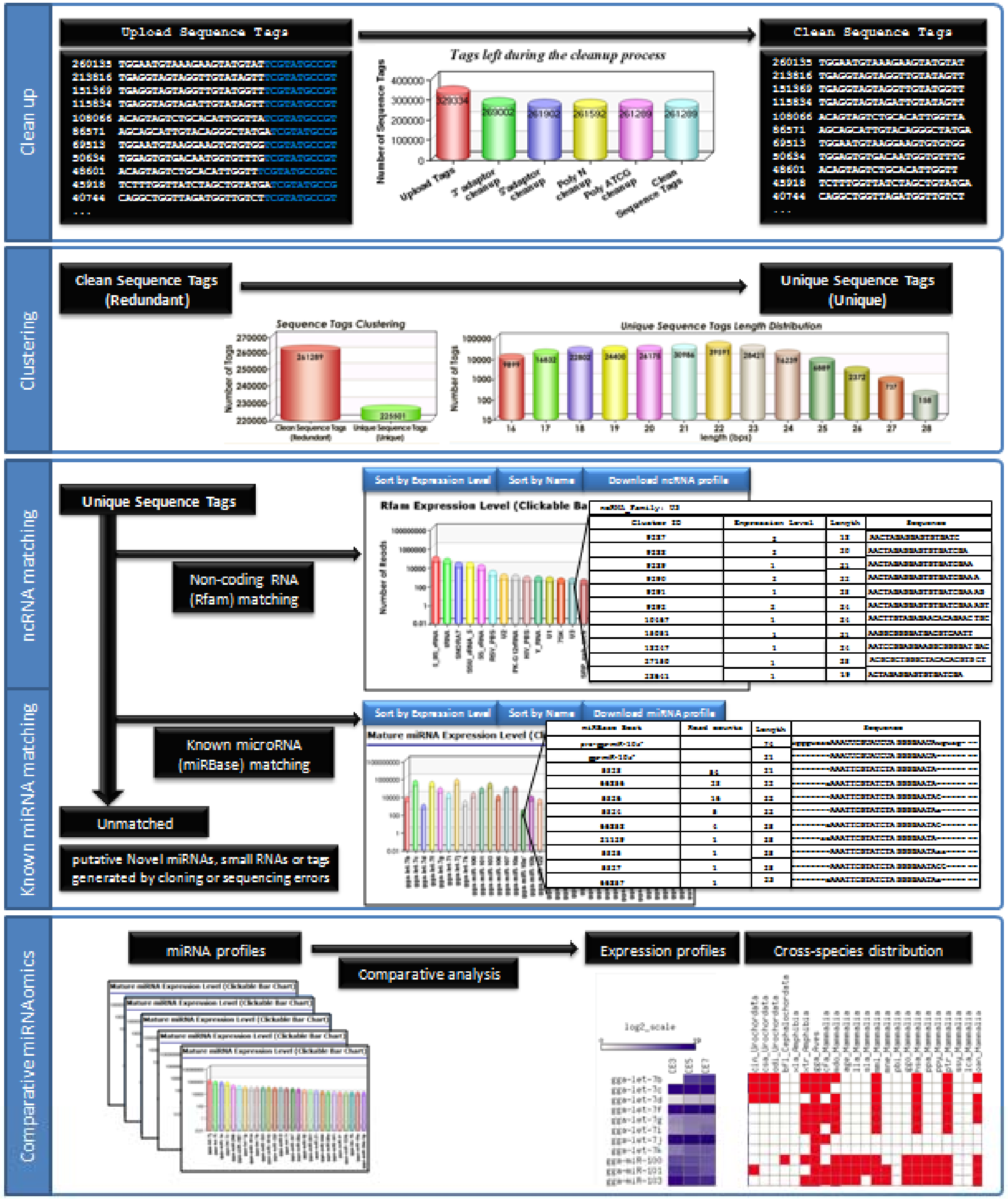
↓

Known miRNA Matching

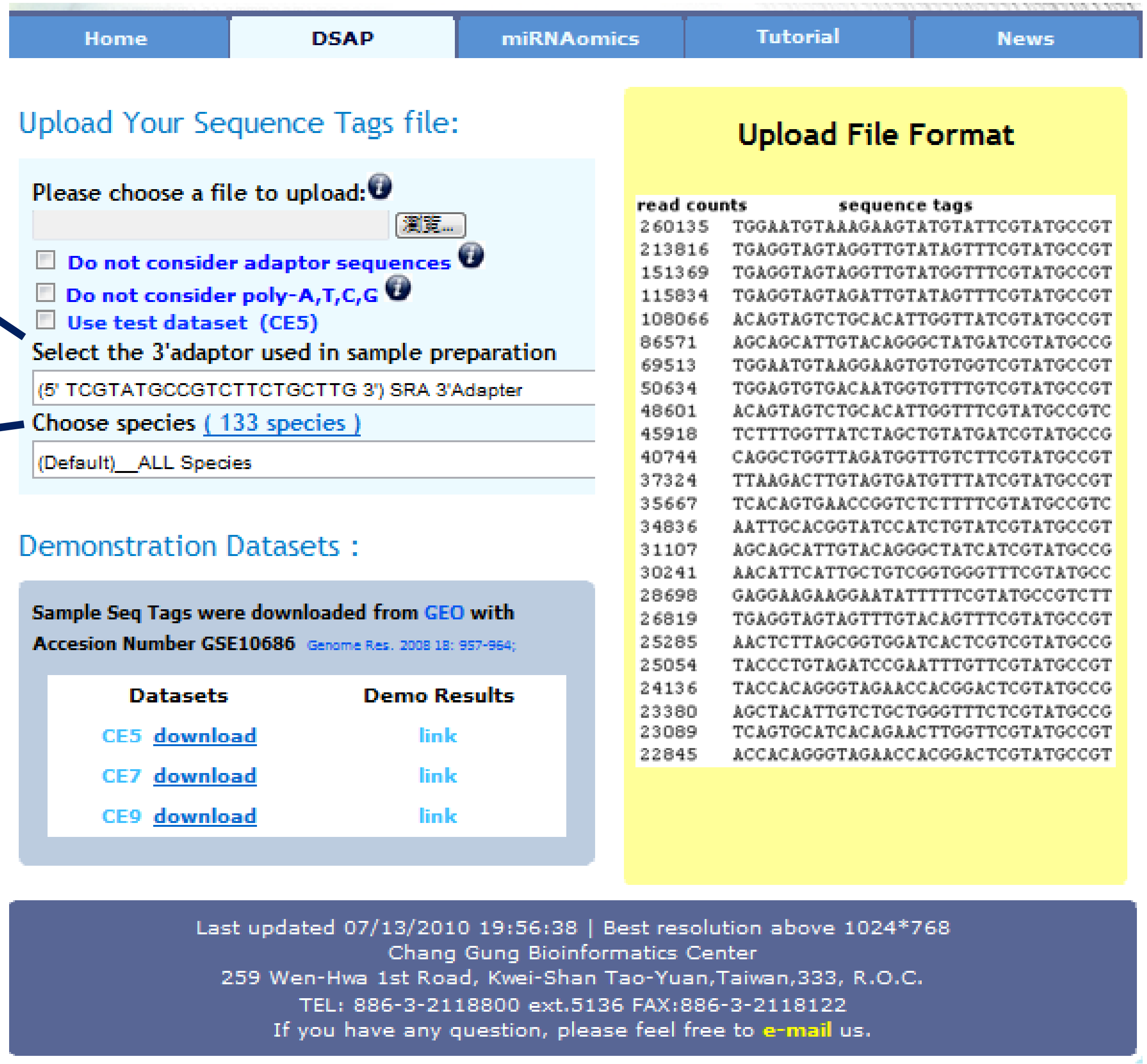
(miRBase V18)

↓

Comparative miRNAomics



DSAP Data Input Page



Upload Your Sequence Tags file:

Please choose a file to upload:

☐ Do not consider adaptor sequences

☐ Do not consider poly-A,T,C,G

☐ Use test dataset (CE5)

Select the 3'adaptor used in sample preparation

(5' TCGTATGCCGTCTTCTGCTTG 3') SRA 3'Adapter

Choose species (133 species)

(Default)__ALL Species

Demonstration Datasets :

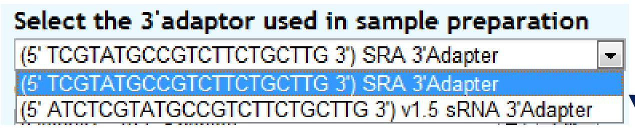
Sample Seq Tags were downloaded from GEO with Accesion Number GSE10686 Genome Res. 2008 18: 957-964;

Datasets	Demo Results
CE5 download	link
CE7 download	link
CE9 download	link

Upload File Format

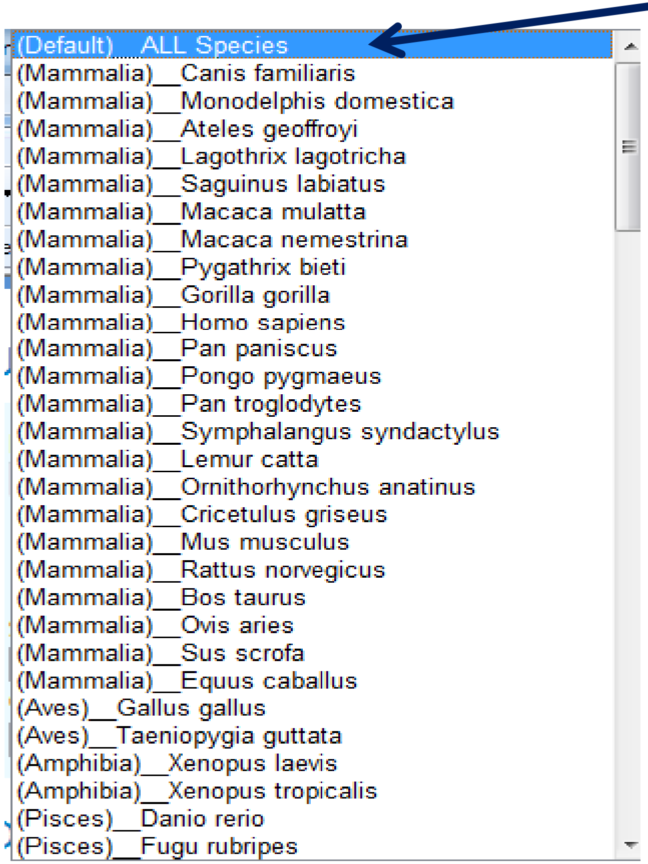
read counts	sequence tags
260135	TGGAATGTAAAGAAGTATGTATTCGTATGCCGT
213816	TGAGGTAGTAGGTTGTATAGTTTCGTATGCCGT
151369	TGAGGTAGTAGGTTGTATGGTTTCGTATGCCGT
115834	TGAGGTAGTAGATTGTATAGTTTCGTATGCCGT
108066	ACAGTAGTCTGCACATTGGTTATCGTATGCCGT
86571	AGCAGCATTGTACAGGGCTATGATCGTATGCCG
69513	TGGAATGTAAGGAAGTGTGTGGTCGTATGCCGT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

Last updated 07/13/2010 19:56:38 | 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.



Select the 3'adaptor used in sample preparation

(5' TCGTATGCCGTCTTCTGCTTG 3') SRA 3'Adapter
(5' TCGTATGCCGTCTTCTGCTTG 3') SRA 3'Adapter
(5' ATCTCGTATGCCGTCTTCTGCTTG 3') v1.5 sRNA 3'Adapter



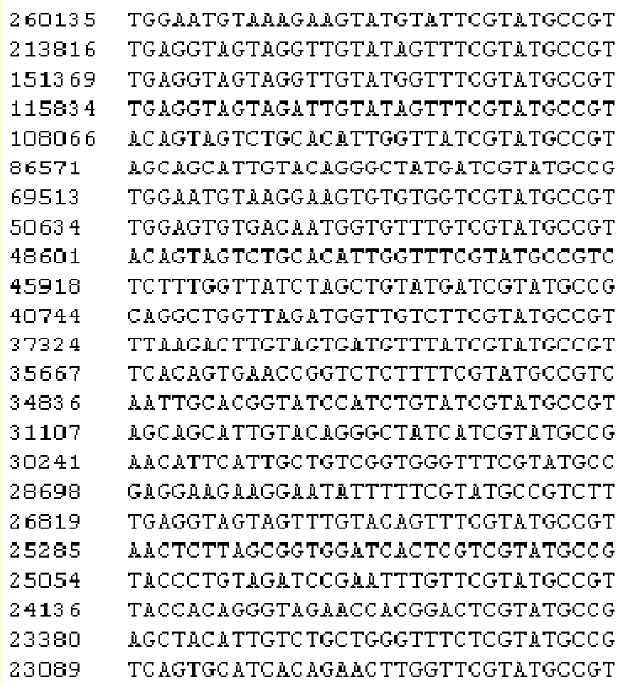
(Default)__ALL Species
(Mammalia)__Canis familiaris
(Mammalia)__Monodelphis domestica
(Mammalia)__Ateles geoffroyi
(Mammalia)__Lagothrix lagotricha
(Mammalia)__Saguinus labiatus
(Mammalia)__Macaca mulatta
(Mammalia)__Macaca nemestrina
(Mammalia)__Pygathrix bieti
(Mammalia)__Gorilla gorilla
(Mammalia)__Homo sapiens
(Mammalia)__Pan paniscus
(Mammalia)__Pongo pygmaeus
(Mammalia)__Pan troglodytes
(Mammalia)__Symphalangus syndactylus
(Mammalia)__Lemur catta
(Mammalia)__Ornithorhynchus anatinus
(Mammalia)__Cricetulus griseus
(Mammalia)__Mus musculus
(Mammalia)__Rattus norvegicus
(Mammalia)__Bos taurus
(Mammalia)__Ovis aries
(Mammalia)__Sus scrofa
(Mammalia)__Equus caballus
(Aves)__Gallus gallus
(Aves)__Taeniopygia guttata
(Amphibia)__Xenopus laevis
(Amphibia)__Xenopus tropicalis
(Pisces)__Danio rerio
(Pisces)__Fugu rubripes

Implementation

DSAP runs on a Linux CentOS 64-bit server housing two quad-core Intel® Xeon® 5300 Series Processors and 16 GB RAM.

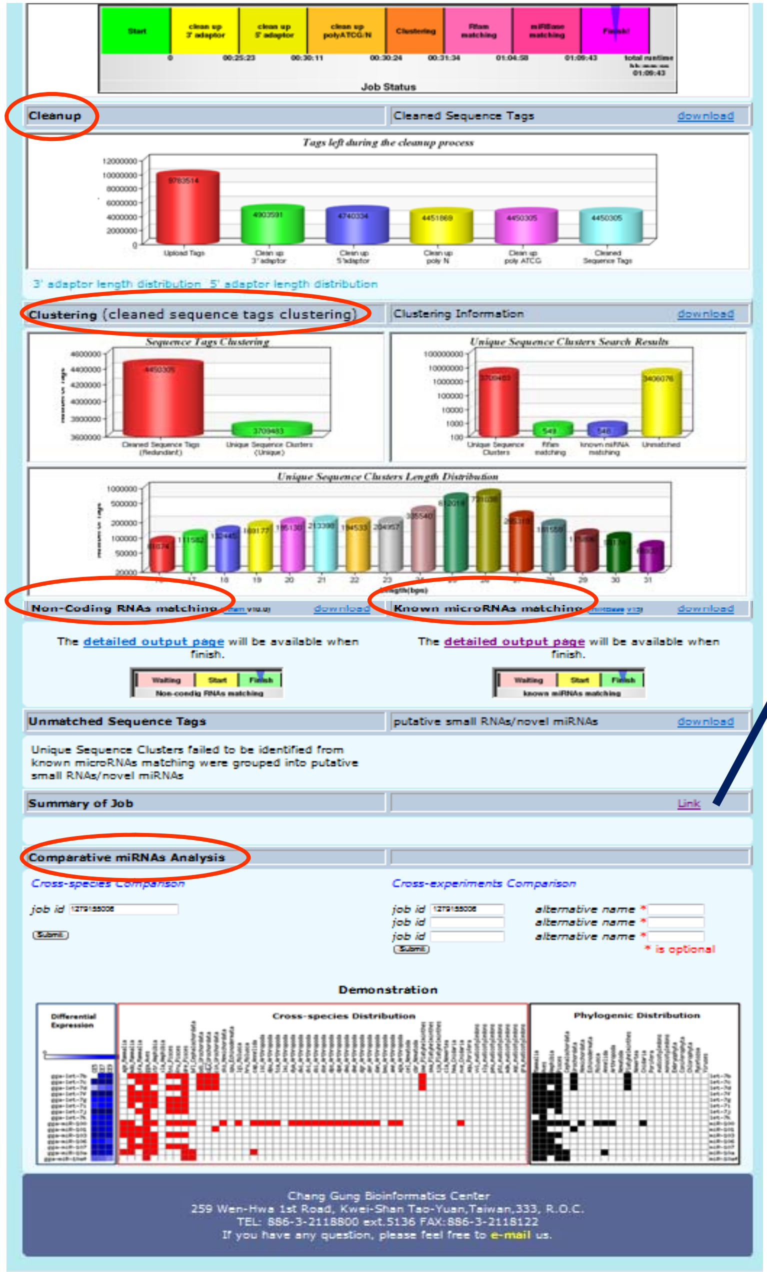
Datasets	Number of Sequence Tags(reads)	Processing time (hh:mm:ss)
Chicken 1	153,406	00:01:20
Chicken 2	220,166	00:01:58
Protist 1	395,939	00:02:50
Protist 2	697,983	00:03:56
Chicken 3	329,334	00:03:42
Protist 3	754,059	00:04:38
Protist 4	736,939	00:05:02
Plant 1	1,643,030	00:11:34
Plant 2	2,090,730	00:13:53
Fish	2,121,561(9,460,897)	00:44:29
Human	9,783,514(38,346,007)	01:09:43

Sequence Tags File Format



```
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
```

DSAP Data Output



Cleanup | Cleaned Sequence Tags | download

3' adaptor length distribution 5' adaptor length distribution

Clustering (cleaned sequence tags clustering) | Clustering Information | download

Non-Coding RNAs matching | download

Known microRNAs matching | download

The detailed output page will be available when finish.

The detailed output page will be available when finish.

Unmatched Sequence Tags | putative small RNAs/novel miRNAs | download

Unique Sequence Clusters failed to be identified from known microRNAs matching were grouped into putative small RNAs/novel miRNAs

Summary of Job | Link

Comparative miRNAs Analysis

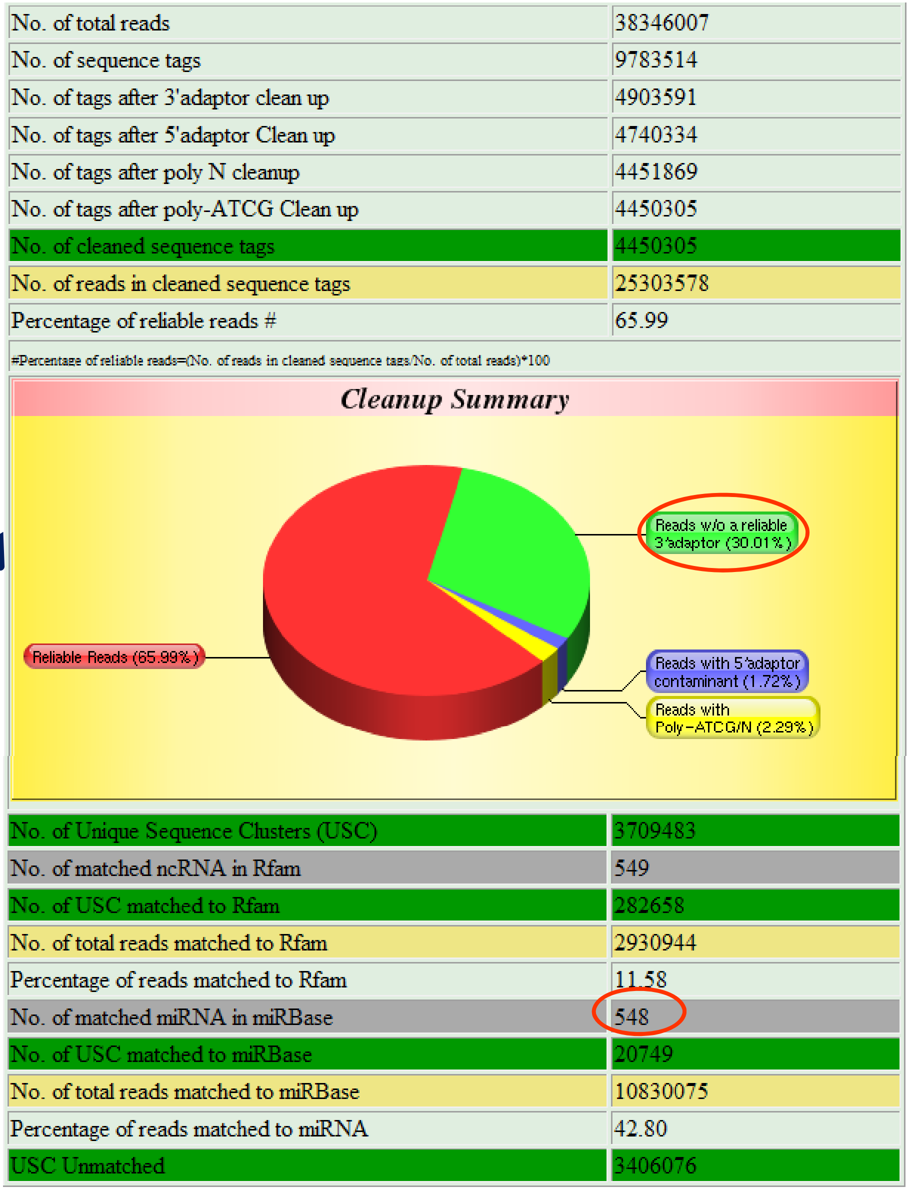
Cross-species Comparison

job id 1279155006

Cross-experiments Comparison

job id 1279155006 *alternative name* *
job id *alternative name* *
job id *alternative name* *
* is optional

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.



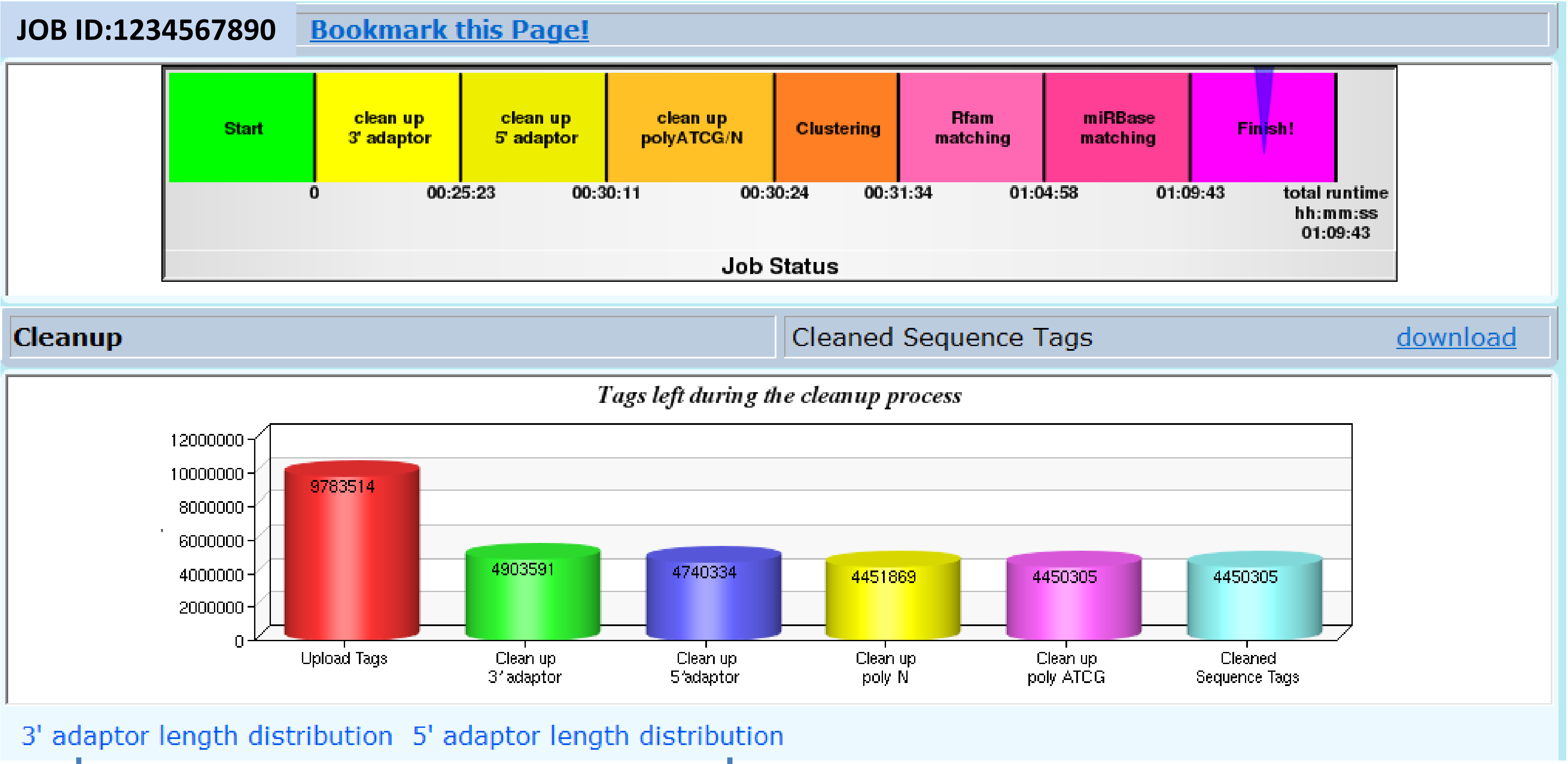
No. of total reads	38346007
No. of sequence tags	9783514
No. of tags after 3'adaptor clean up	4903591
No. of tags after 5'adaptor Clean up	4740334
No. of tags after poly N cleanup	4451869
No. of tags after poly-ATCG Clean up	4450305
No. of cleaned sequence tags	4450305
No. of reads in cleaned sequence tags	25303578
Percentage of reliable reads #	65.99

#Percentage of reliable reads=(No. of reads in cleaned sequence tags/No. of total reads)*100

Cleanup Summary

No. of Unique Sequence Clusters (USC)	3709483
No. of matched ncRNA in Rfam	549
No. of USC matched to Rfam	282658
No. of total reads matched to Rfam	2930944
Percentage of reads matched to Rfam	11.58
No. of matched miRNA in miRBase	548
No. of USC matched to miRBase	20749
No. of total reads matched to miRBase	10830075
Percentage of reads matched to miRNA	42.80
USC Unmatched	3406076

Number of reads: 38,346,007 Number of Tags: 9,783,514

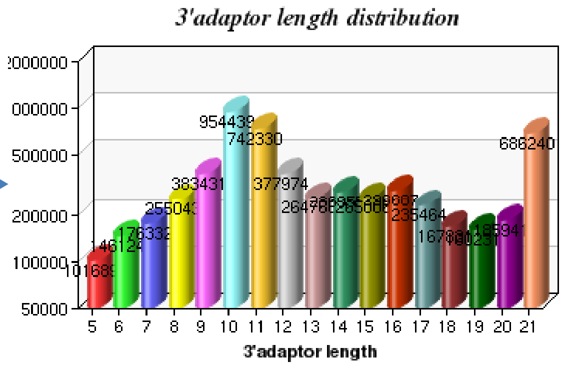


JOB ID:1234567890 Bookmark this Page!

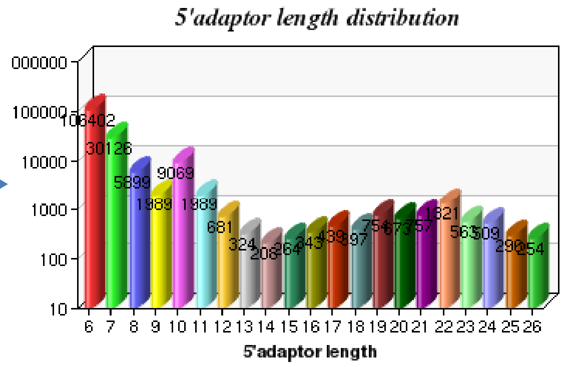
Cleanup Cleaned Sequence Tags download

Tags left during the cleanup process

3' adaptor length distribution 5' adaptor length distribution



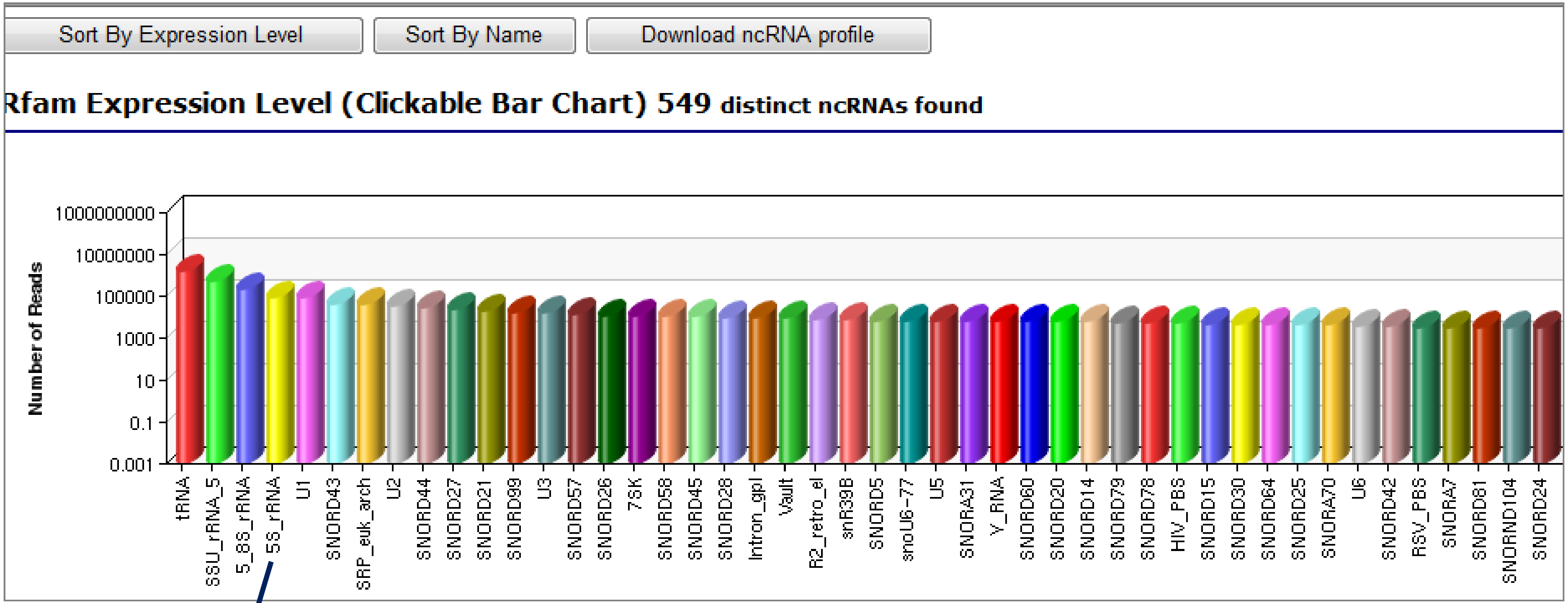
3'adaptor length distribution



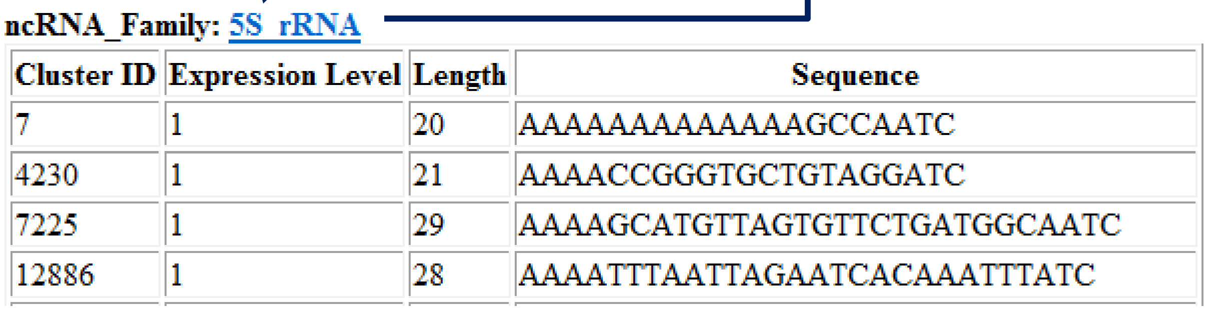
5'adaptor length distribution

Non-Coding RNAs Searching

keyword search Go



Rfam Expression Level (Clickable Bar Chart) 549 distinct ncRNAs found

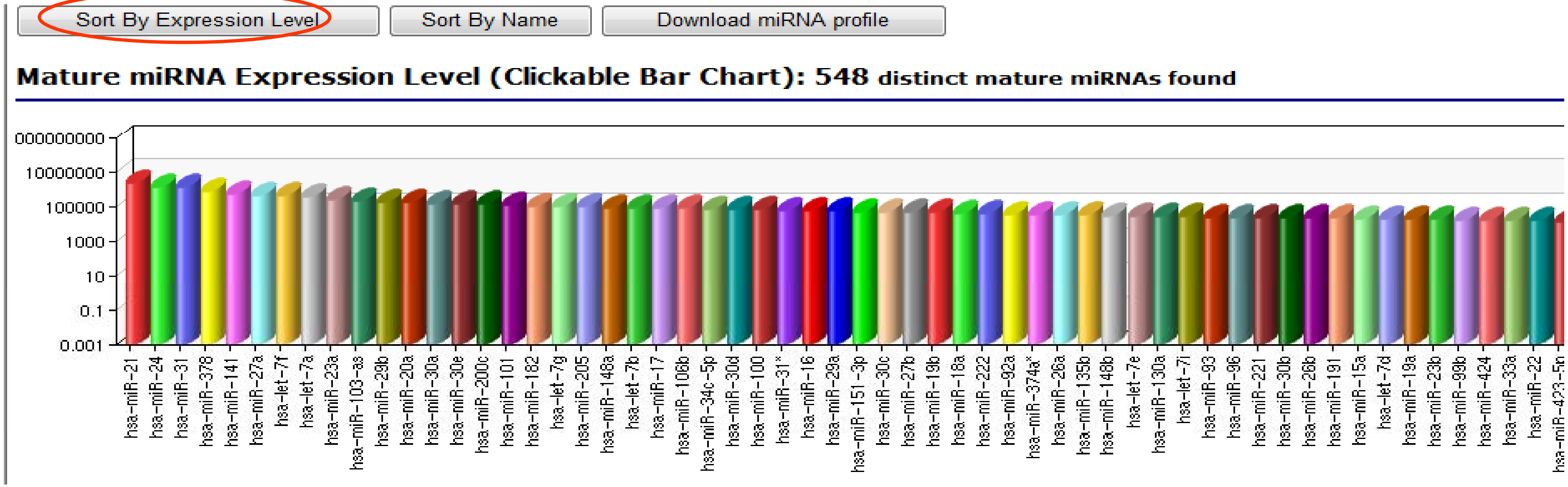


ncRNA_Family: 5S_rRNA

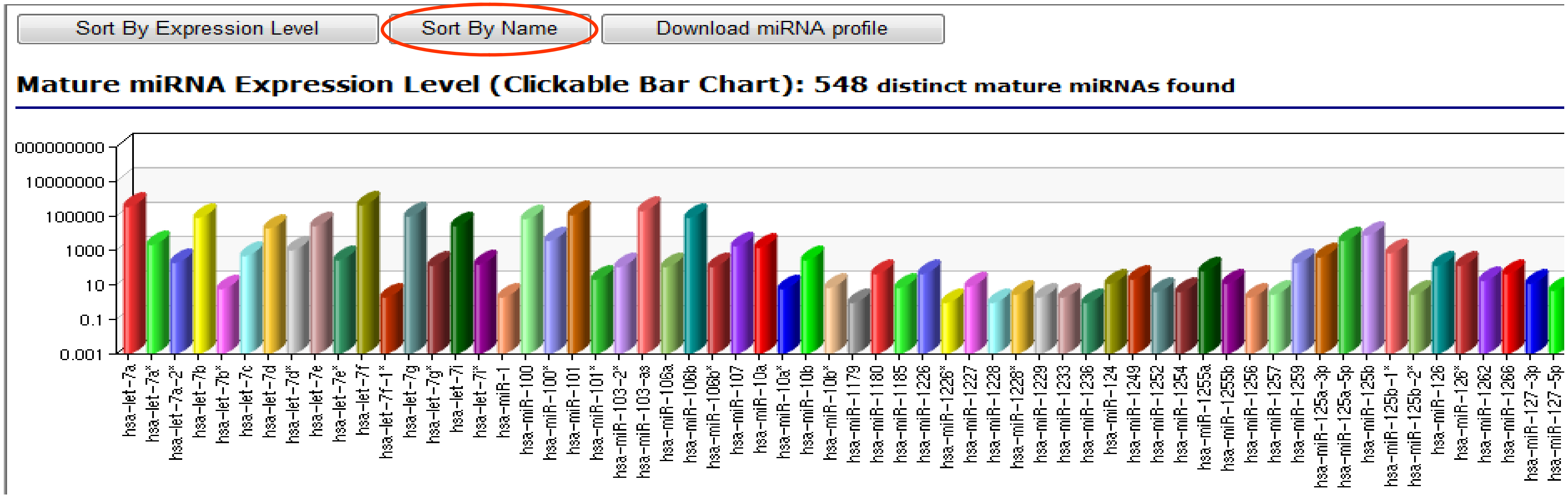
Cluster ID	Expression Level	Length	Sequence
7	1	20	AAAAAAAAAAAAAAGCCAATC
4230	1	21	AAAACCGGGTGCTGTAGGATC
7225	1	29	AAAAGCATGTTAGTGTTCTGATGGCAATC
12886	1	28	AAAATTTAATTAGAATCACAAATTTATC

Family: *5S_rRNA* (RF00001)
Description: *5S ribosomal RNA*

miRNAs Searching

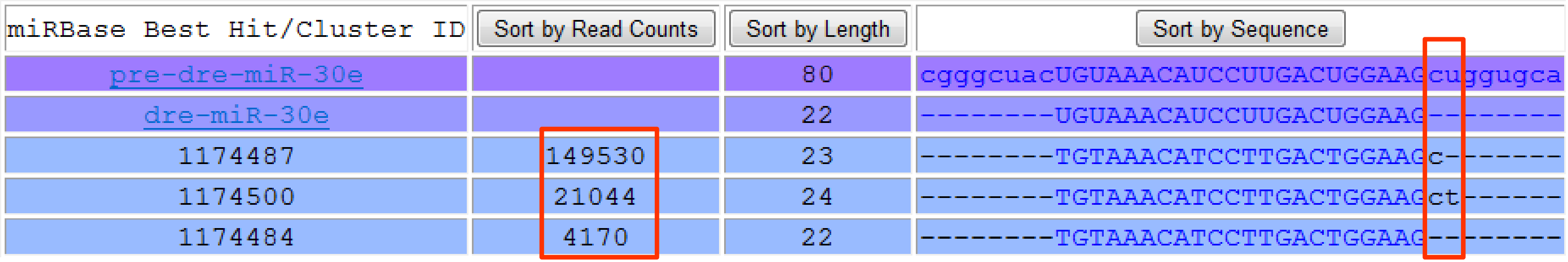


Mature miRNA Expression Level (Clickable Bar Chart): 548 distinct mature miRNAs found

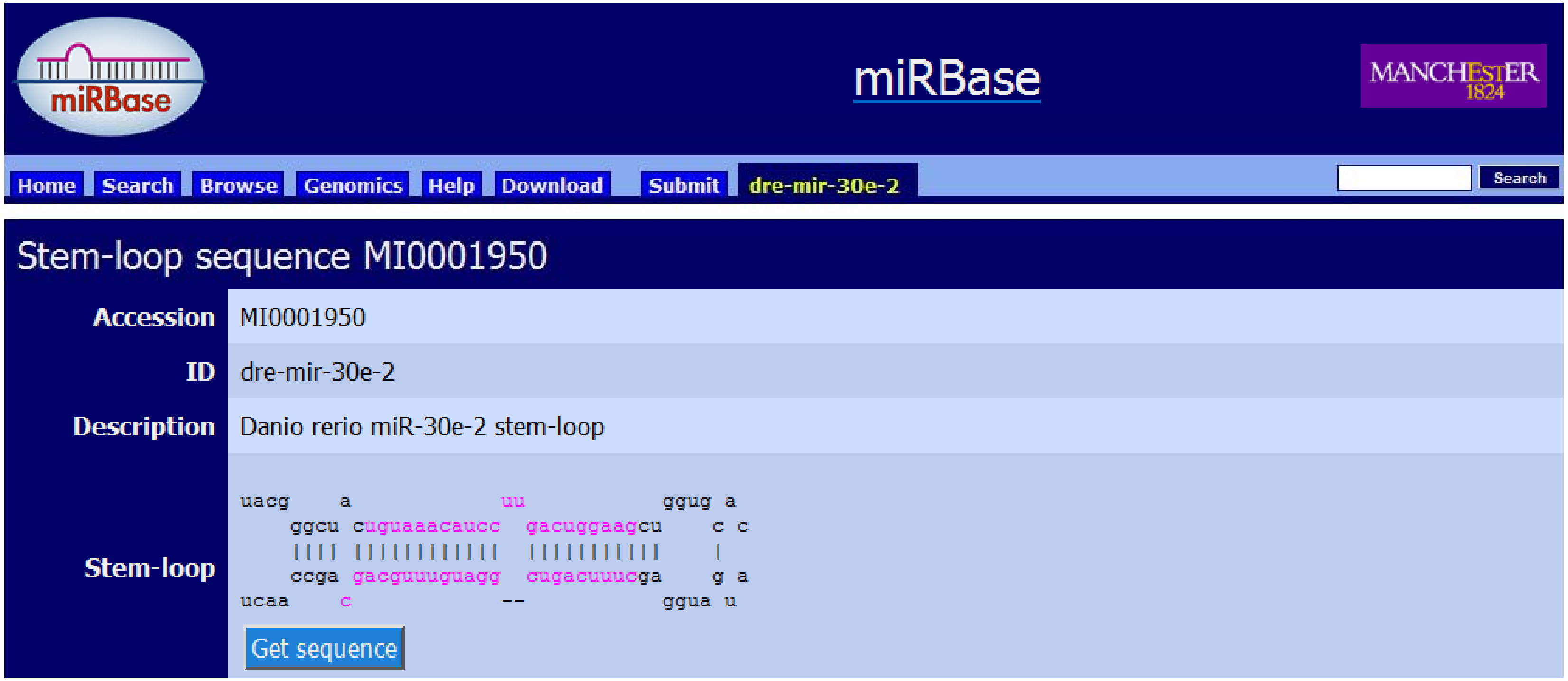


Mature miRNA Expression Level (Clickable Bar Chart): 548 distinct mature miRNAs found

Iso-miR



miRBase Best Hit/Cluster ID	Sort by Read Counts	Sort by Length	Sort by Sequence
pre-dre-miR-30e		80	cgggcuacUGUAAACAUCCUUGACUGGAAGcuggugca
dre-miR-30e		22	--------UGUAAACAUCCUUGACUGGAAG--------
1174487	149530	23	--------TGTAAACATCCTTGACTGGAAGc-------
1174500	21044	24	--------TGTAAACATCCTTGACTGGAAGct------
1174484	4170	22	--------TGTAAACATCCTTGACTGGAAG--------



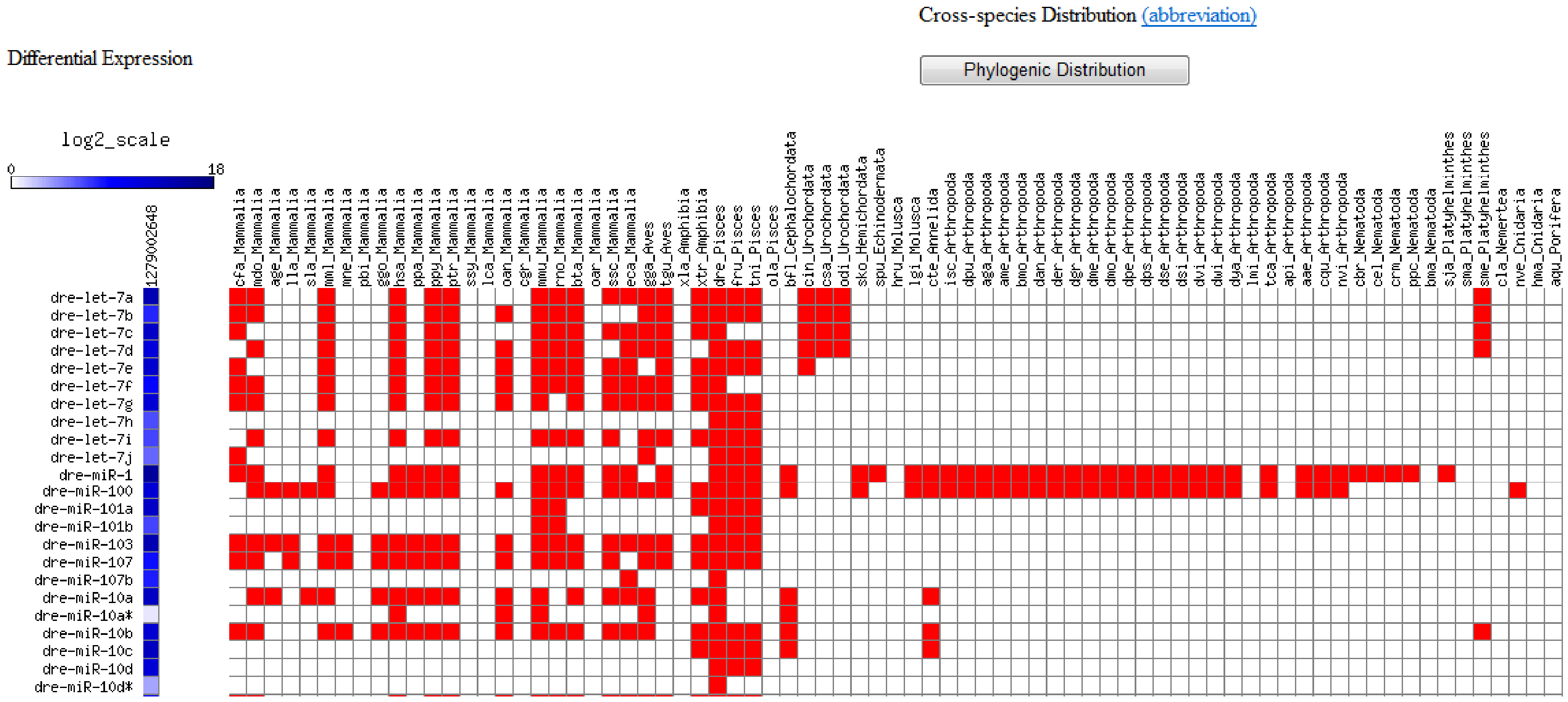
Stem-loop sequence MI0001950

Accession	MI0001950
ID	dre-mir-30e-2
Description	Danio rerio miR-30e-2 stem-loop

Stem-loop

```
uacg    a            uu           ggug a
    ggcu cuguaaacaucc  gacuggaagcu    c c
    |||| ||||||||||||  |||||||||||    |
    ccga gacguuuguagg  cugacuuucga    g a
ucaa    c            --           ggua u
```

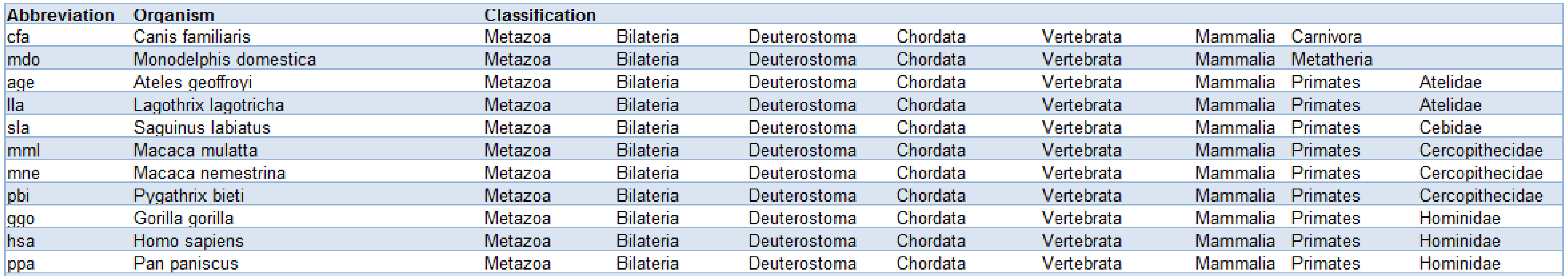
Cross-species Distribution of Identified miRNAs



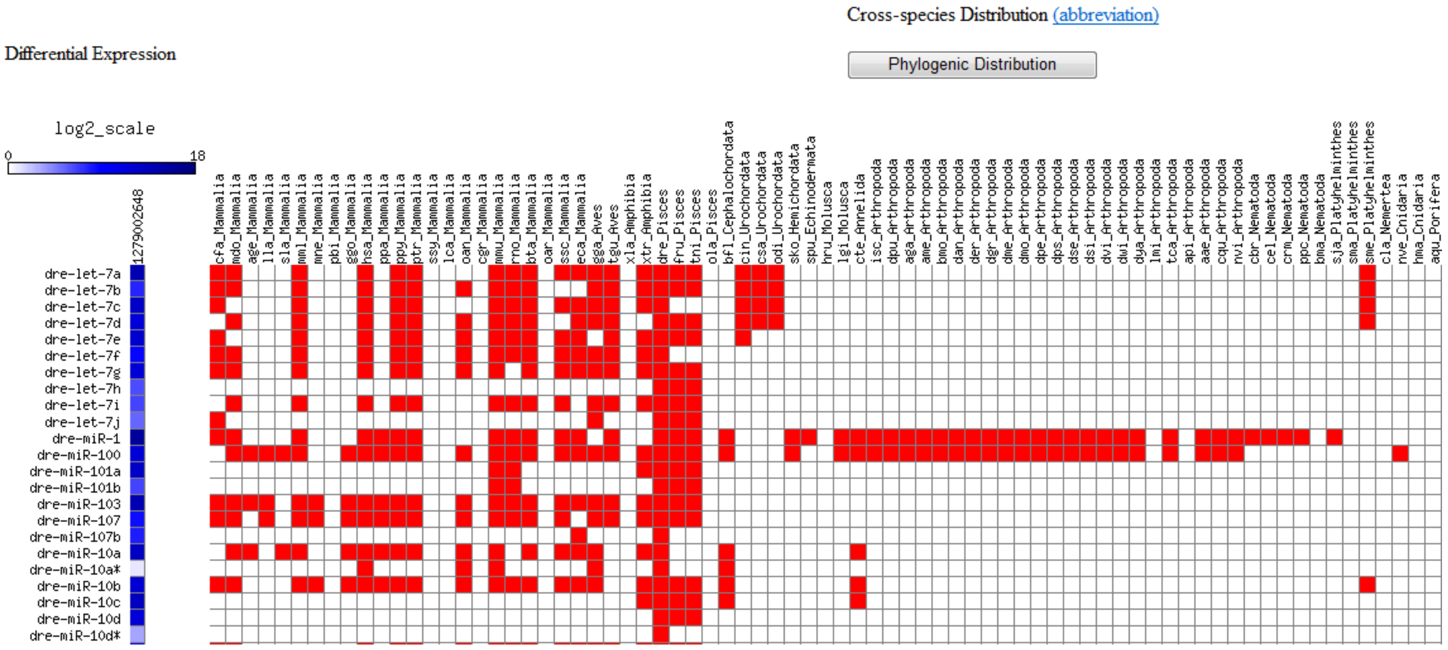
Differential Expression

Cross-species Distribution (abbreviation)

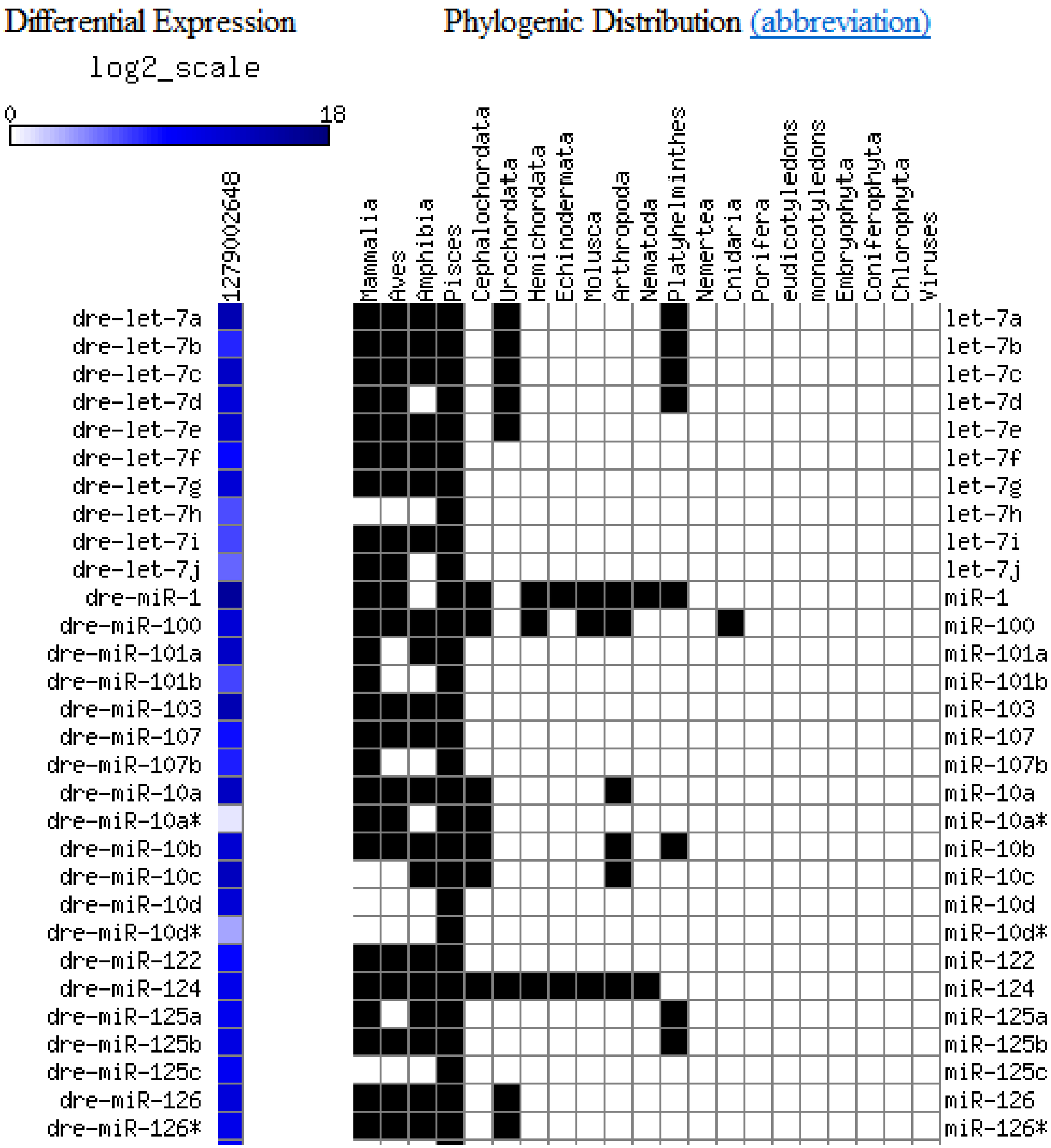
Cross-species Distribution of Identified miRNAs



Abbreviation	Organism	Classification							
cfa	Canis familiaris	Metazoa	Bilateria	Deuterostoma	Chordata	Vertebrata	Mammalia	Carnivora	
mdo	Monodelphis domestica	Metazoa	Bilateria	Deuterostoma	Chordata	Vertebrata	Mammalia	Metatheria	
age	Ateles geoffroyi	Metazoa	Bilateria	Deuterostoma	Chordata	Vertebrata	Mammalia	Primates	Atelidae
lla	Lagothrix lagotricha	Metazoa	Bilateria	Deuterostoma	Chordata	Vertebrata	Mammalia	Primates	Atelidae
sla	Saguinus labiatus	Metazoa	Bilateria	Deuterostoma	Chordata	Vertebrata	Mammalia	Primates	Cebidae
mml	Macaca mulatta	Metazoa	Bilateria	Deuterostoma	Chordata	Vertebrata	Mammalia	Primates	Cercopithecidae
mne	Macaca nemestrina	Metazoa	Bilateria	Deuterostoma	Chordata	Vertebrata	Mammalia	Primates	Cercopithecidae
pbi	Pygathrix bieti	Metazoa	Bilateria	Deuterostoma	Chordata	Vertebrata	Mammalia	Primates	Cercopithecidae
ggo	Gorilla gorilla	Metazoa	Bilateria	Deuterostoma	Chordata	Vertebrata	Mammalia	Primates	Hominidae
hsa	Homo sapiens	Metazoa	Bilateria	Deuterostoma	Chordata	Vertebrata	Mammalia	Primates	Hominidae
ppa	Pan paniscus	Metazoa	Bilateria	Deuterostoma	Chordata	Vertebrata	Mammalia	Primates	Hominidae



Phylogenic Distribution of Identified miRNAs



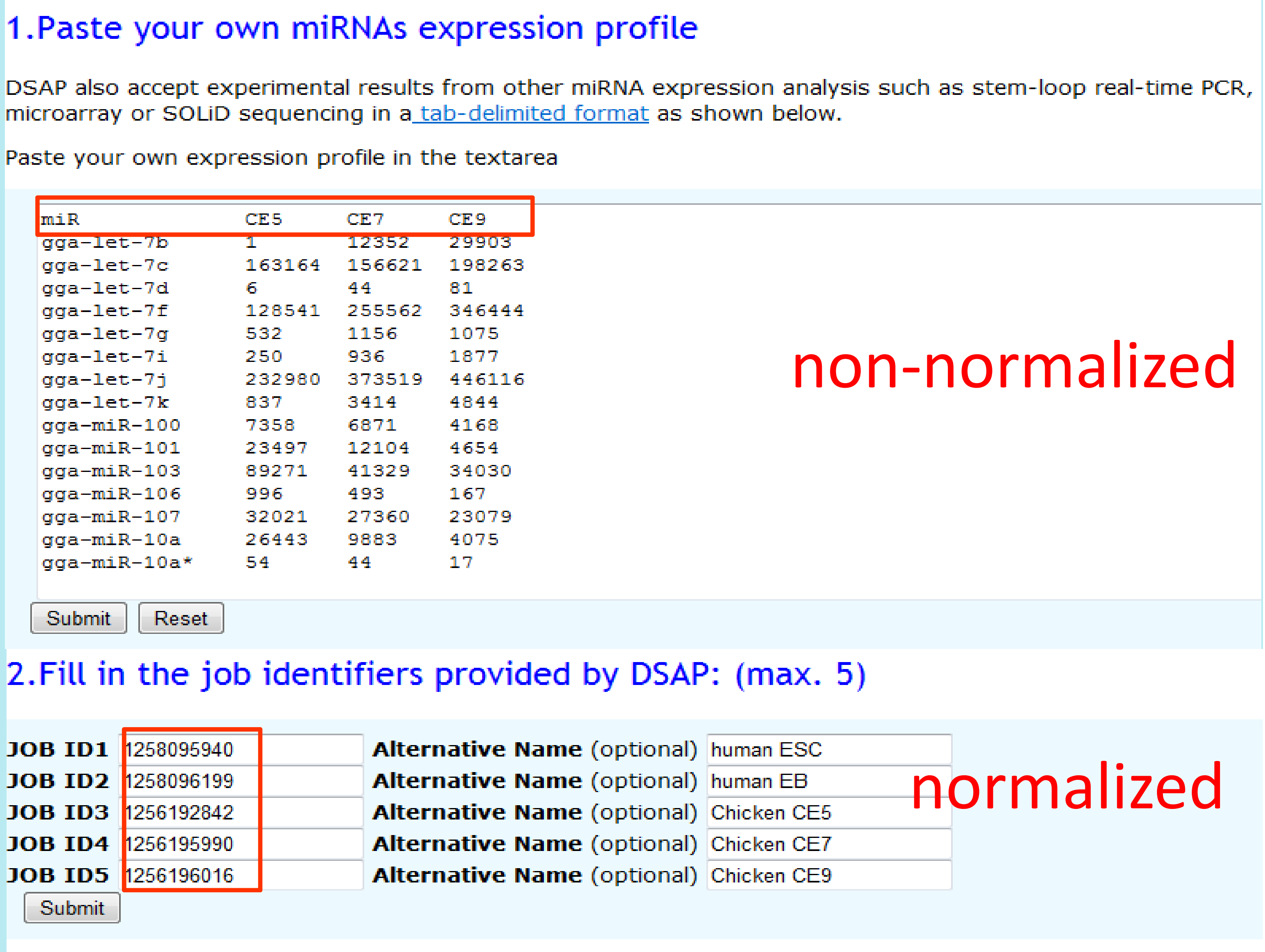
Differential Expression

log2_scale

Phylogenic Distribution (abbreviation)

	1279002648	Mammalia	Aves	Amphibia	Pisces	Cephalochordata	Urochordata	Hemichordata	Echinodermata	Molusca	Arthropoda	Nematoda	Platyhelminthes	Nemertea	Cnidaria	Porifera	eudicotyledons	monocotyledons	Embryophyta	Coniferophyta	Chlorophyta	Viruses	
dre-let-7a	🟦	■	■	■	■		■						■										let-7a
dre-let-7b	🟦	■	■	■	■		■						■										let-7b
dre-let-7c	🟦	■	■	■	■		■						■										let-7c
dre-let-7d	🟦	■	■		■		■						■										let-7d
dre-let-7e	🟦	■	■	■	■		■																let-7e
dre-let-7f	🟦	■	■	■	■																		let-7f
dre-let-7g	🟦	■	■	■	■																		let-7g
dre-let-7h	🟦				■																		let-7h
dre-let-7i	🟦	■	■	■	■																		let-7i
dre-let-7j	🟦	■	■		■																		let-7j
dre-miR-1	🟦	■	■		■	■		■	■	■	■	■	■										miR-1
dre-miR-100	🟦	■	■	■	■	■		■		■	■				■								miR-100
dre-miR-101a	🟦	■		■	■																		miR-101a
dre-miR-101b	🟦	■			■																		miR-101b
dre-miR-103	🟦	■	■	■	■																		miR-103
dre-miR-107	🟦	■	■	■	■																		miR-107
dre-miR-107b	🟦	■			■																		miR-107b
dre-miR-10a	🟦	■	■	■	■	■					■												miR-10a
dre-miR-10a*	🟦	■	■		■	■																	miR-10a*
dre-miR-10b	🟦	■	■	■	■	■					■		■										miR-10b
dre-miR-10c	🟦			■	■	■					■												miR-10c
dre-miR-10d	🟦				■																		miR-10d
dre-miR-10d*	🟦				■																		miR-10d*
dre-miR-122	🟦	■	■	■	■																		miR-122
dre-miR-124	🟦	■	■	■	■	■	■	■	■	■	■	■											miR-124
dre-miR-125a	🟦	■		■	■								■										miR-125a
dre-miR-125b	🟦	■	■	■	■								■										miR-125b
dre-miR-125c	🟦				■																		miR-125c
dre-miR-126	🟦	■	■	■	■		■																miR-126
dre-miR-126*	🟦	■	■	■	■		■																miR-126*

Comparative miRNAomics



1.Paste your own miRNAs expression profile

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 as shown below.

Paste your own expression profile in the textarea

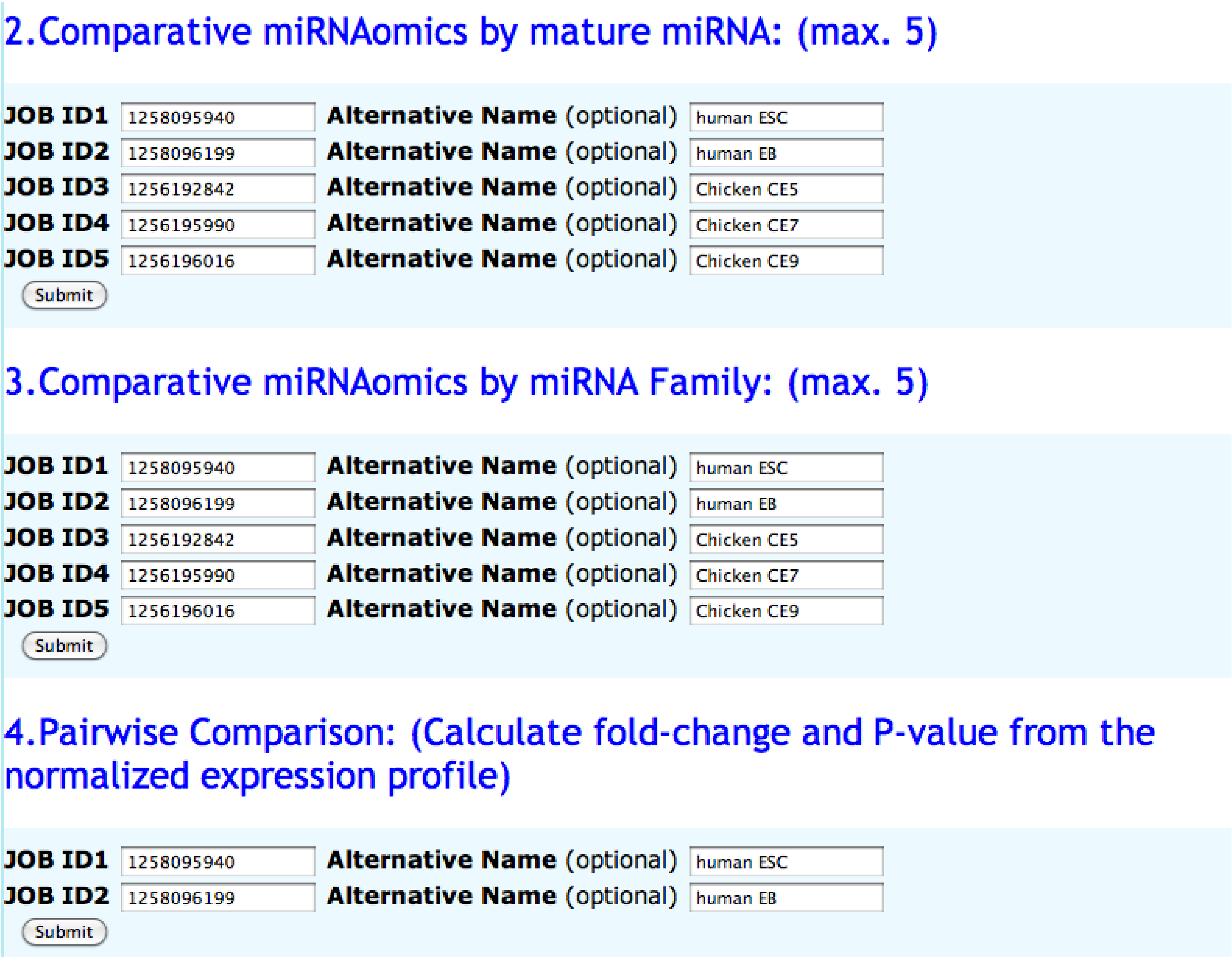
```
miR             CE5     CE7     CE9
gga-let-7b      1       12352   29903
gga-let-7c      163164  156621  198263
gga-let-7d      6       44      81
gga-let-7f      128541  255562  346444
gga-let-7g      532     1156    1075
gga-let-7i      250     936     1877
gga-let-7j      232980  373519  446116
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
```

non-normalized

2.Fill in the job identifiers 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

normalized



2.Comparative miRNAomics by mature miRNA: (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

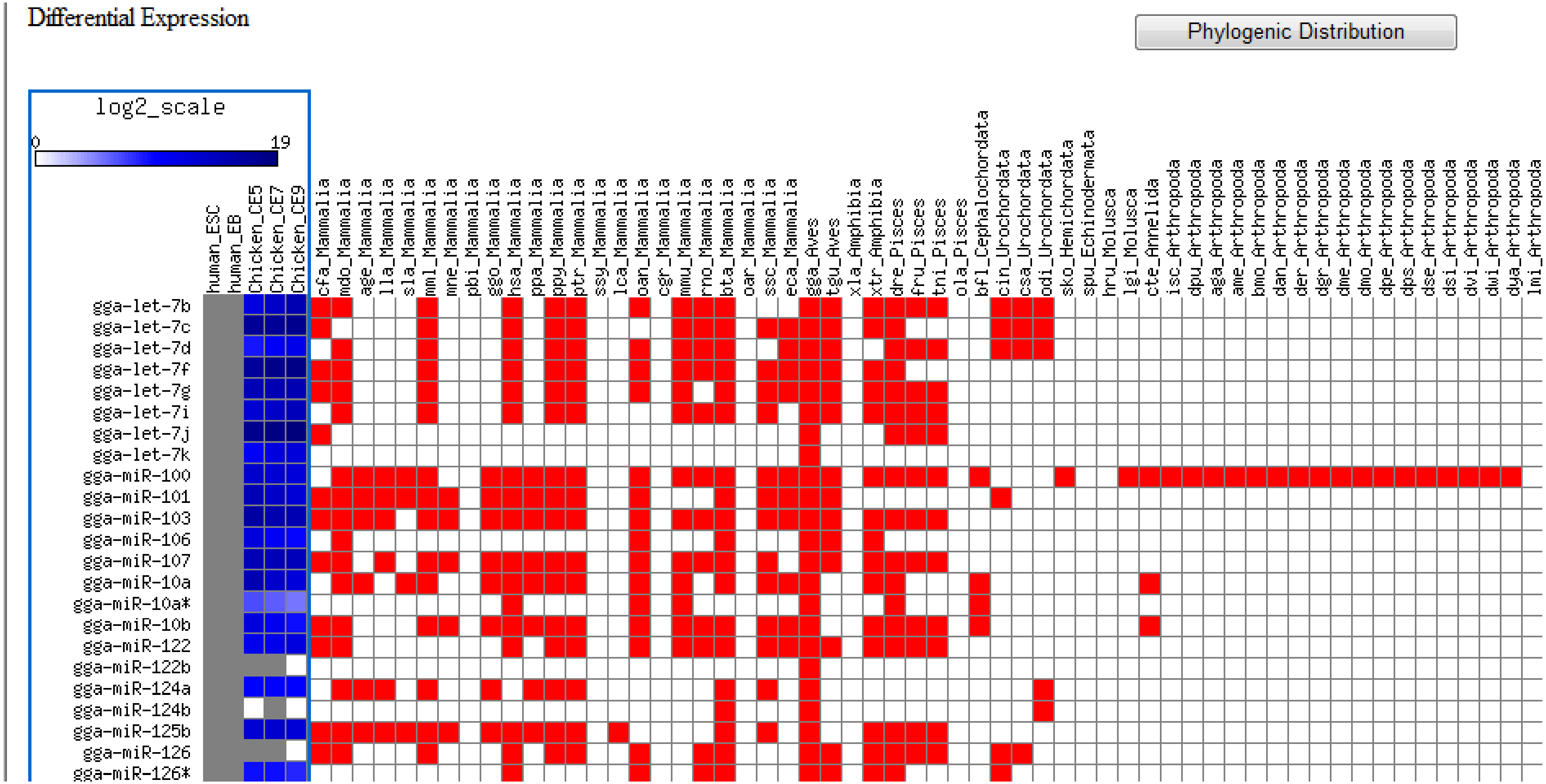
3.Comparative miRNAomics by miRNA Family: (max. 5)

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

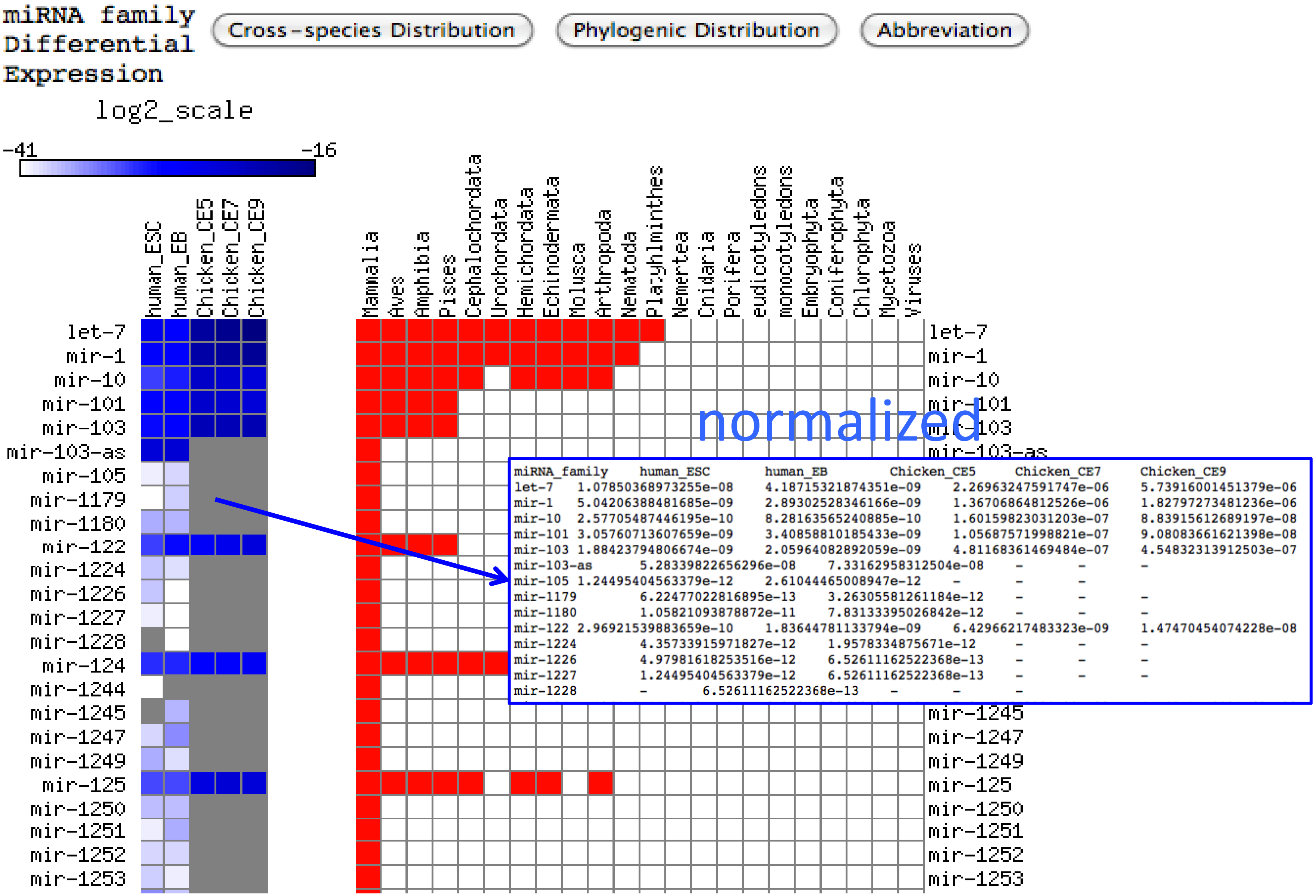
4.Pairwise Comparison: (Calculate fold-change and P-value from the normalized expression profile)

JOB ID1	1258095940	**Alternative Name** (optional)	human ESC
JOB ID2	1258096199	**Alternative Name** (optional)	human EB

Comparative miRNAomics by mature miRNA



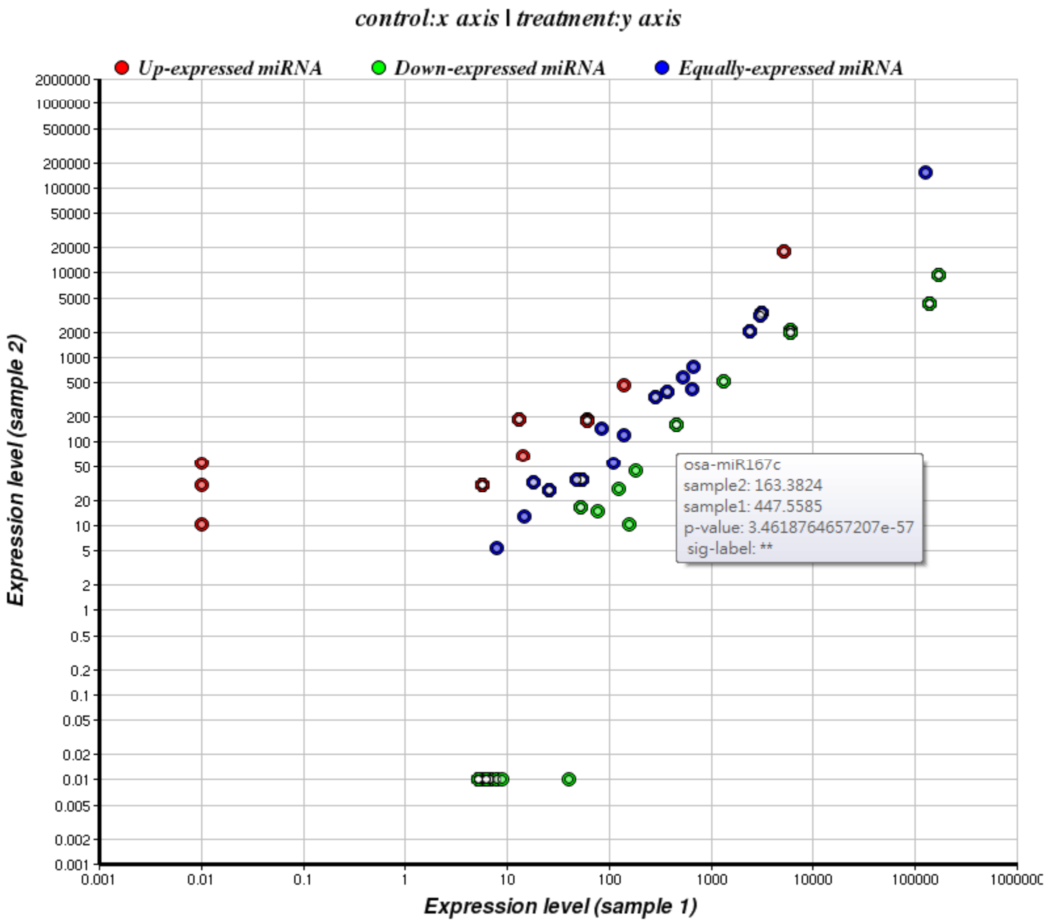
Comparative miRNAomics by miRNA family



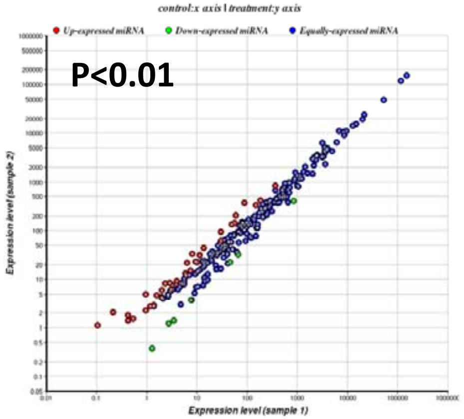
```
miRNA_family     human_ESC          human_EB           Chicken_CE5        Chicken_CE7        Chicken_CE9
let-7    1.07850368973255e-08    4.18715321874351e-09    2.26963247591747e-06    5.73916001451379e-06
mir-1    5.04206388481685e-09    2.89302528346166e-09    1.36706864812526e-06    1.82797273481236e-06
mir-10   2.57705487446195e-10    8.28163565240885e-10    1.60159823031203e-07    8.83915612689197e-08
mir-101  3.05760713607659e-09    3.40858810185433e-09    1.05687571998821e-07    9.08083661621398e-08
mir-103  1.88423794806674e-09    2.05964082892059e-09    4.81168361469484e-07    4.54832313912503e-07
mir-103-as       5.28339822656296e-08    7.33162958312504e-08    -       -       -
mir-105  1.24495404563379e-12    2.61044465008947e-12    -       -       -
mir-1179         6.22477022816895e-13    3.26305581261184e-12    -       -       -
mir-1180         1.05821093878872e-11    7.83133395026842e-12    -       -       -
mir-122  2.96921539883659e-10    1.83644781133794e-09    6.42966217483323e-09    1.47470454074228e-08
mir-1224         4.35733915971827e-12    1.9578334875671e-12     -       -       -
mir-1226         4.97981618253516e-12    6.52611162522368e-13    -       -       -
mir-1227         1.24495404563379e-12    6.52611162522368e-13    -       -       -
mir-1228         -       6.52611162522368e-13    -       -       -
```

Pairwise Comparsion

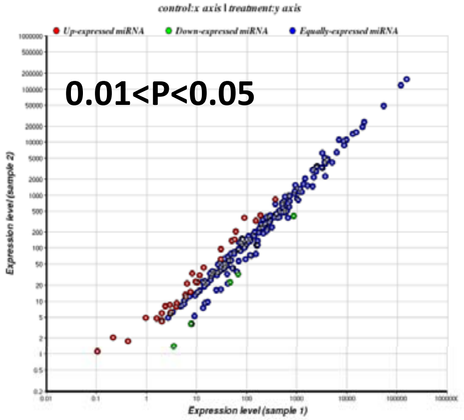
No limit
P<0.01
P<0.05



osa-miR167c
sample2: 163.3824
sample1: 447.5585
p-value: 3.4618764657207e-57
sig-label: **



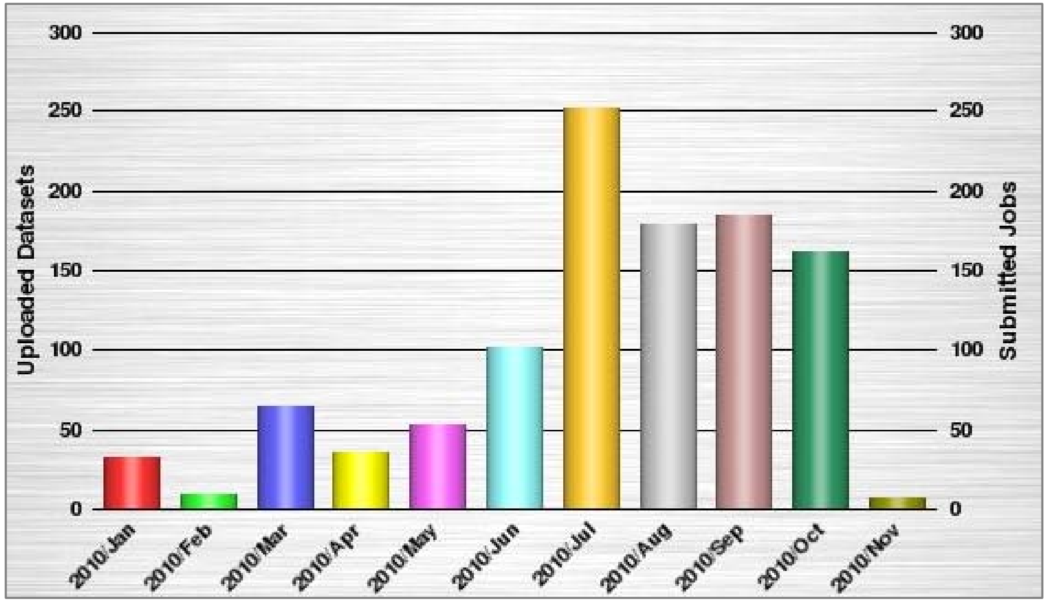
P<0.01



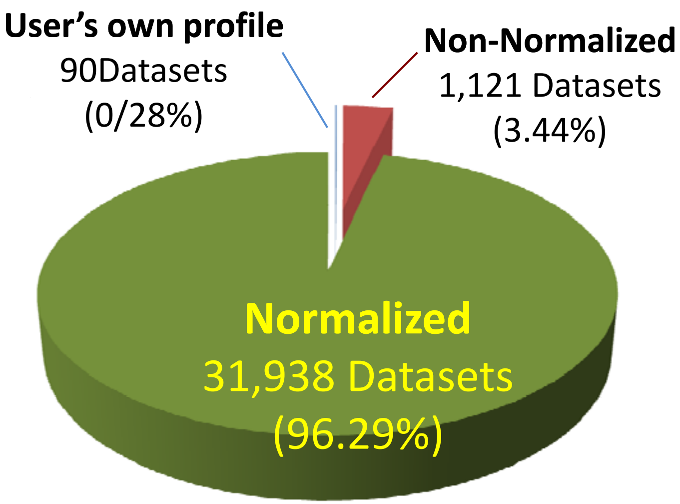
0.01<P<0.05

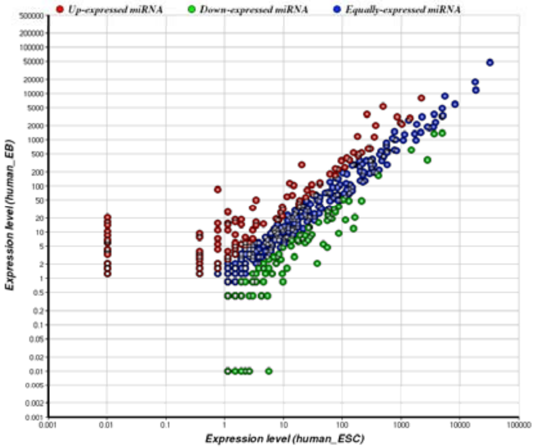
Job Statistics of Comparative miRNAomics

Just passing the 1,000 jobs mark in October, 2010

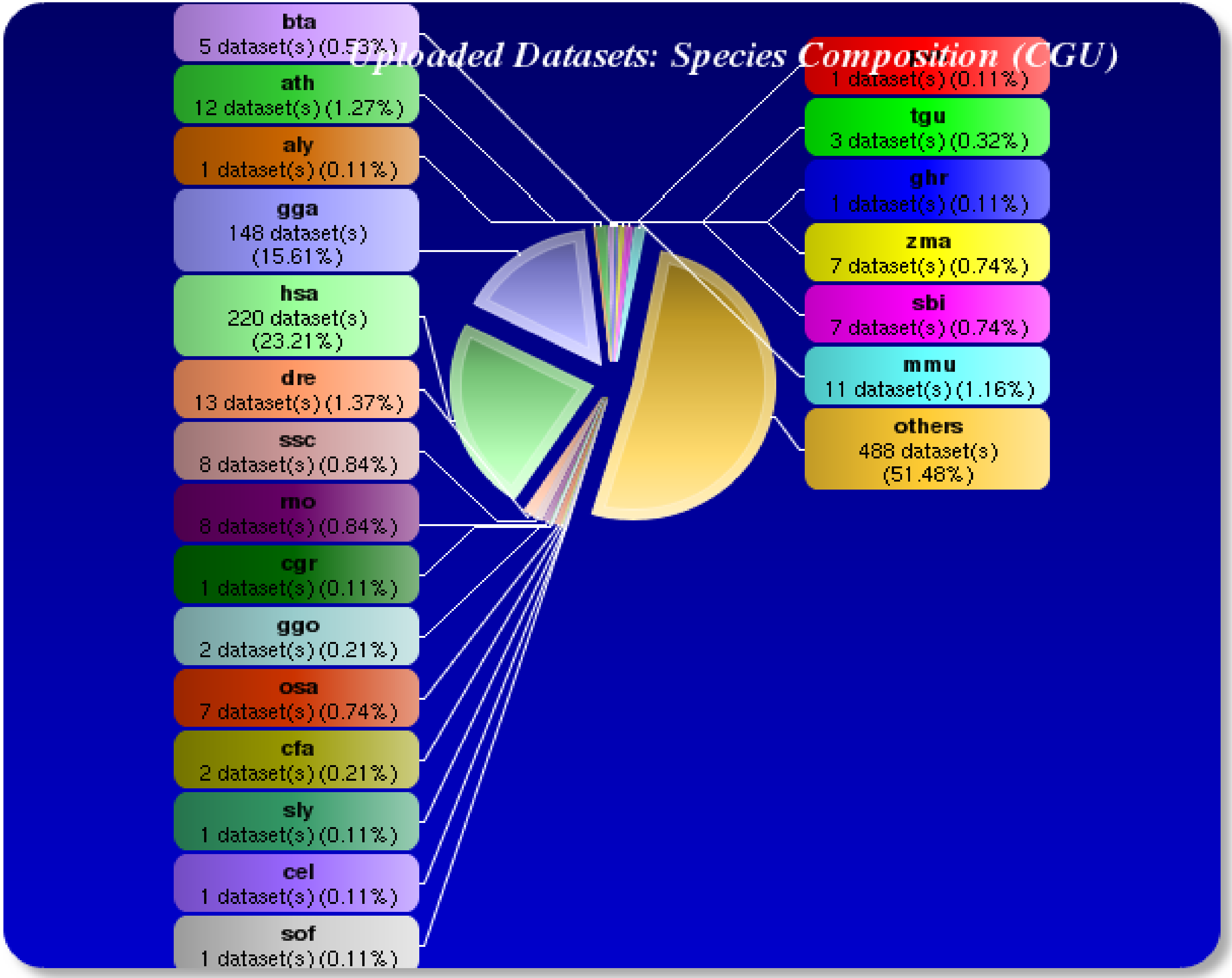


Over 32,549 requests for Comparative miRNAomics





Job Statistics by Organisms



DSAP^2

Map2Genome

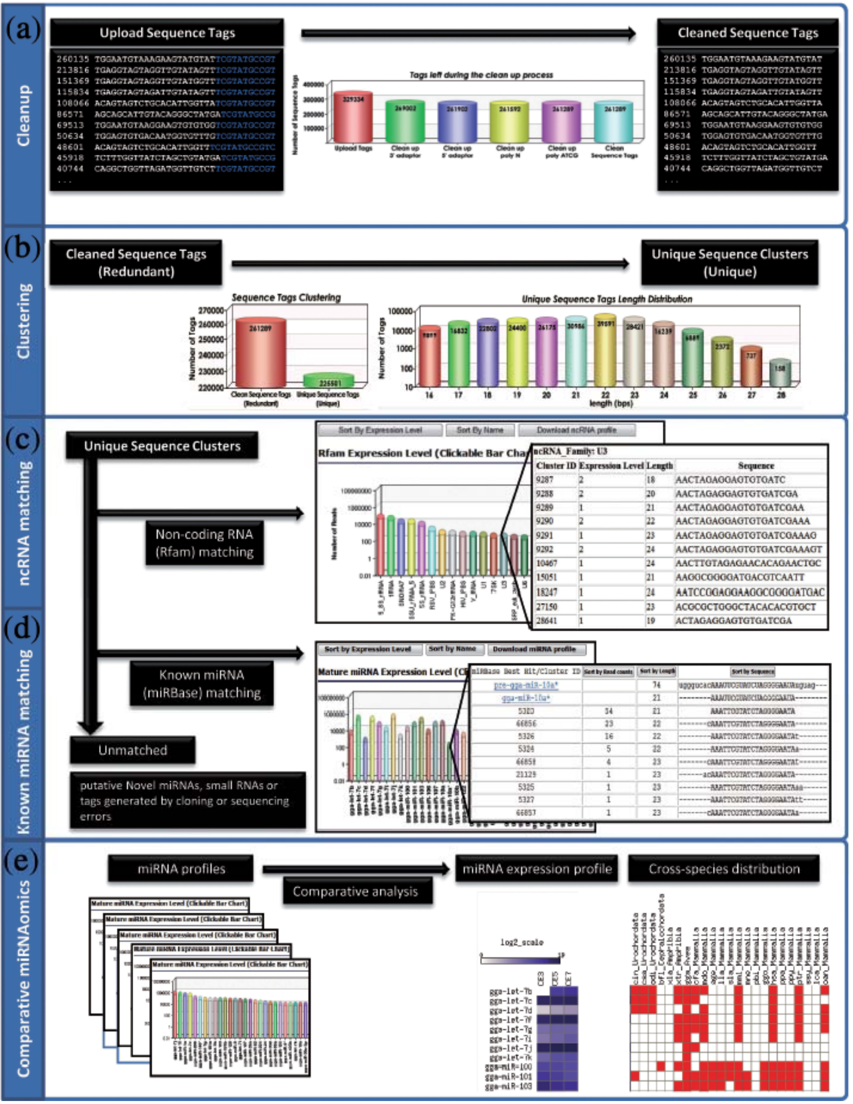
SOAPaligner/soap2 SOAPaligner/soap2 is a program for faster and efficient alignment for short oligonucleotide onto reference sequences. SOAPaligner/soap2 is compatible with numerous applications, including single-read or pair-end resequencing.

Novel miRNA prediction

isomiR & miRNA editing

>hsa-let-7a-1					
TGGGATGAGGTAGTAGGTTGTATAGTTTTAGGGTCACACCCACCACTGGGAGATAACTATACAATCTACTGTCTTTCCTA	hsa-let-7a-1	80			
(((((.(((..((((((((((((((((...(((.....))).((....))....))))))))))))))))..))))))))	structure	-35.6			
`*****TGAGGTAGTAGGTTGTATAGTT*****************************************************`	hsa-let-7a	22		isomiRs	
`*****************************************************CTATACAATCTACTGTCTTTC***`	hsa-let-7a*	21		5'	3'
..GGATGAGGTAGTAGGTTGTATAG...	t0227878	23	1	-3	-2
....ATGAGGTAGTAGGTTGTATAGTT...	t0006845	23	42	-1	0
....ATGAGGTAGTAGGTTGTATAGT..	t0021457	22	9	-1	-1
....ATGAGGTAGTAGGTTGTATAG...	t0039011	21	4	-1	-2
....ATGAGGTAGTAGGTTGTATAGTTT..	t0165920	24	1	-1	1
.....TGAGGTAGTAGGTTGTATAGTT...	t0000002	22	1043926	0	0
.....TGAGGTAGTAGGTTGTATAGT..	t0000015	21	96011	0	-1
.....TGAGGTAGTAGGTTGTATAGTTT..	t0000044	23	36942	0	1
.....TGAGGTAGTAGGTTGTATAG...	t0000082	20	13528	0	-2

Differential Expression

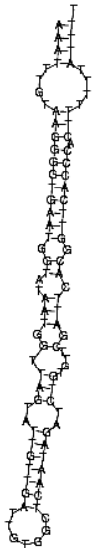


Novel miRNA Prediction

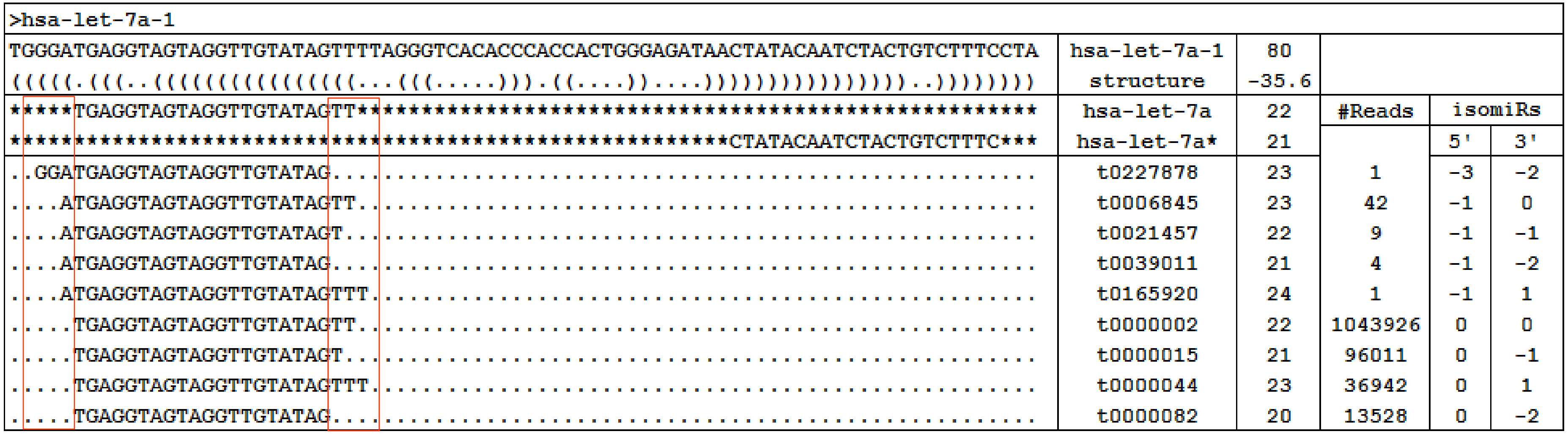
Soap2genome → RNAfold → Filter → Vienna

Trichomonas Vaginalis

```
TvE-m0002 Trichomonas_vaginalis_G3|DS113908|2007-01-11|ds-DNA|Trichomonas_vaginalis_Carlton:29154:29248:- 95(nt) -22.00(kcal/mol)
AAATTTGTAAGGGGTGAATGGTATAATGGTTAGTATTGTTGATTGTGGCTCAATAGATCTGTGTTCGATTCACGGTTCACCCACTTTTTTATTTT TvE-m0002 337985
((((....((((((((((.((..(((.(.(((..(((((((.......)))))))..)))....).))).)).)))))))).)))....)))).
**********GGGGTGAATGGTATAATGGTTA*************************************************************** TvE-m0002-5p 337865
*************************************************************TGTTCGATTCACGGTTCACCCACTT********* TvE-m0002-3p 29
------GTAAGGGGTGAATGGTATAATGGTT---------------------------------------------------------------- Cluster_116016 1
--------AAGGGGTGAATGGTATAATGGTTA--------------------------------------------------------------- Cluster_2927 22
--------AAGGGGTGAATGGTATAATGGTTAG-----------                                            ------- Cluster_2928 8
---------AGGGGTGAATGGTATAATGGTTA------------                                            ------- Cluster_14163 4
---------AGGGGTGAATGGTATAATGGTTAG-----------                                            ------- Cluster_14164 2
---------AGGGGTGAATGGTATAATGGTTAGT----------                                            ------- Cluster_14165 1
----------GGGGTGAATGGTATAATG----------------                                            ------- Cluster_102092 447
----------GGGGTGAATGGTATAATGG---------------                                            ------- Cluster_102117 1093
----------GGGGTGAATGGTATAATGGT--------------                                            ------- Cluster_102195 1483
----------GGGGTGAATGGTATAATGGTT-------------                                            ------- Cluster_102237 2981
----------GGGGTGAATGGTATAATGGTTA------------                                            ------- Cluster_102238 192067
----------GGGGTGAATGGTATAATGGTTAG-----------                                            ------- Cluster_102269 91570
----------GGGGTGAATGGTATAATGGTTAGT----------                                            ------- Cluster_102342 9375
----------GGGGTGAATGGTATAATGGTTAGTA---------                                            ------- Cluster_102343 18014
----------GGGGTGAATGGTATAATGGTTAGTAT--------                                            ------- Cluster_102368 9145
-----------GGGTGAATGGTATAATGG---------------                                            ------- Cluster_107222 36
-----------GGGTGAATGGTATAATGGT--------------                                            ------- Cluster_107235 59
```

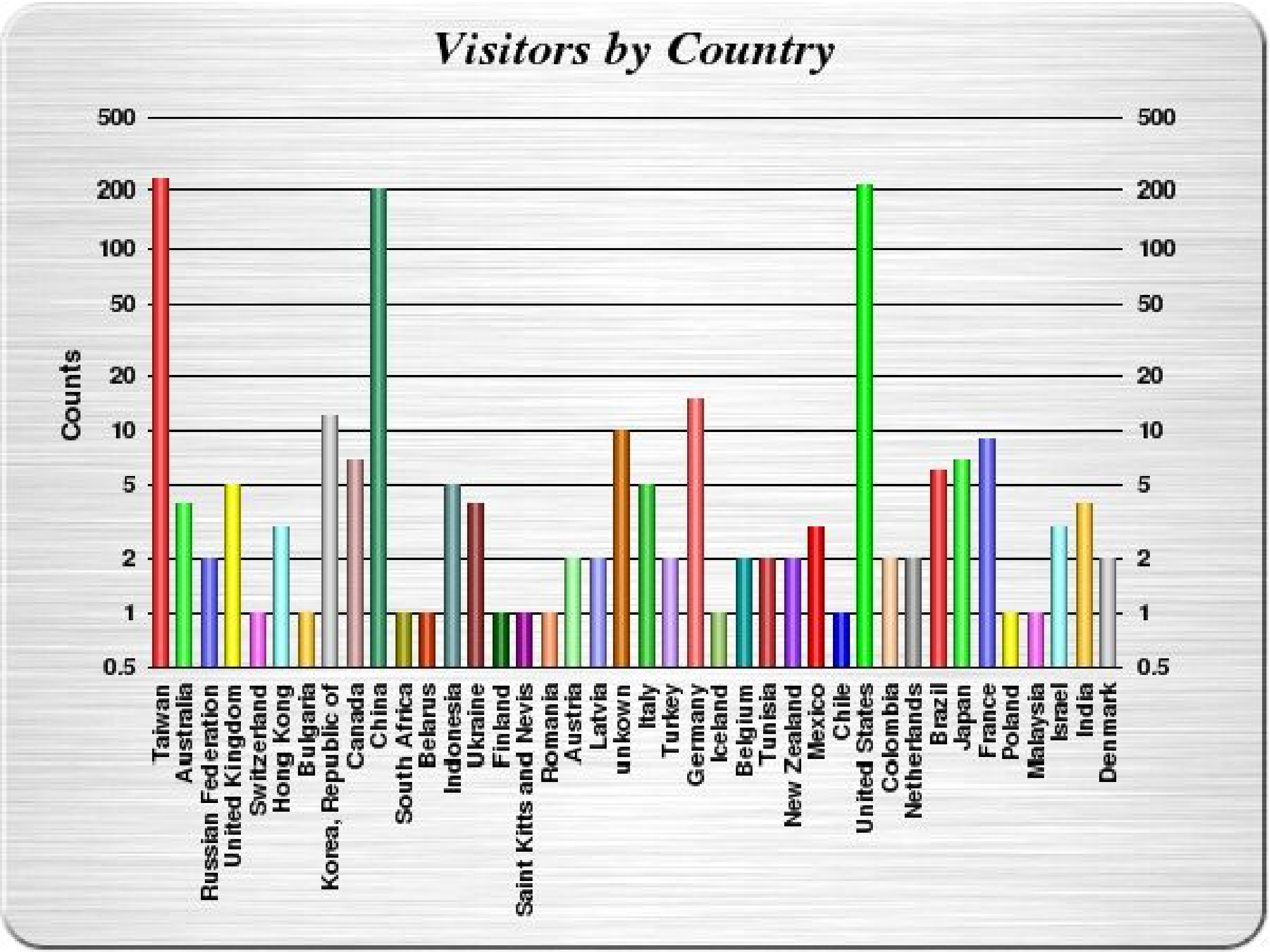


IsomiRs & miRNA Editing



>hsa-let-7a-1					
TGGGATGAGGTAGTAGGTTGTATAGTTTTAGGGTCACACCCACCACTGGGAGATAACTATACAATCTACTGTCTTTCCTA	hsa-let-7a-1	80			
(((((.(((..((((((((((((((((...(((.....))).((....))....))))))))))))))))..))))))))	structure	-35.6			
*****TGAGGTAGTAGGTTGTATAGTT***	hsa-let-7a	22	#Reads	isomiRs	
CTATACAATCTACTGTCTTTC*	hsa-let-7a*	21		5'	3'
..GGATGAGGTAGTAGGTTGTATAG...	t0227878	23	1	-3	-2
....ATGAGGTAGTAGGTTGTATAGTT...	t0006845	23	42	-1	0
....ATGAGGTAGTAGGTTGTATAGT..	t0021457	22	9	-1	-1
....ATGAGGTAGTAGGTTGTATAG...	t0039011	21	4	-1	-2
....ATGAGGTAGTAGGTTGTATAGTTT..	t0165920	24	1	-1	1
.....TGAGGTAGTAGGTTGTATAGTT...	t0000002	22	1043926	0	0
.....TGAGGTAGTAGGTTGTATAGT..	t0000015	21	96011	0	-1
.....TGAGGTAGTAGGTTGTATAGTTT..	t0000044	23	36942	0	1
.....TGAGGTAGTAGGTTGTATAG...	t0000082	20	13528	0	-2

Job Statistics by Country



DSAP Mirror sites

DSAP @ Chang Gung University

http://dsap.cgu.edu.tw

Mirror site @ Tsing Hua University

http://dsap.life.nthu.edu.tw

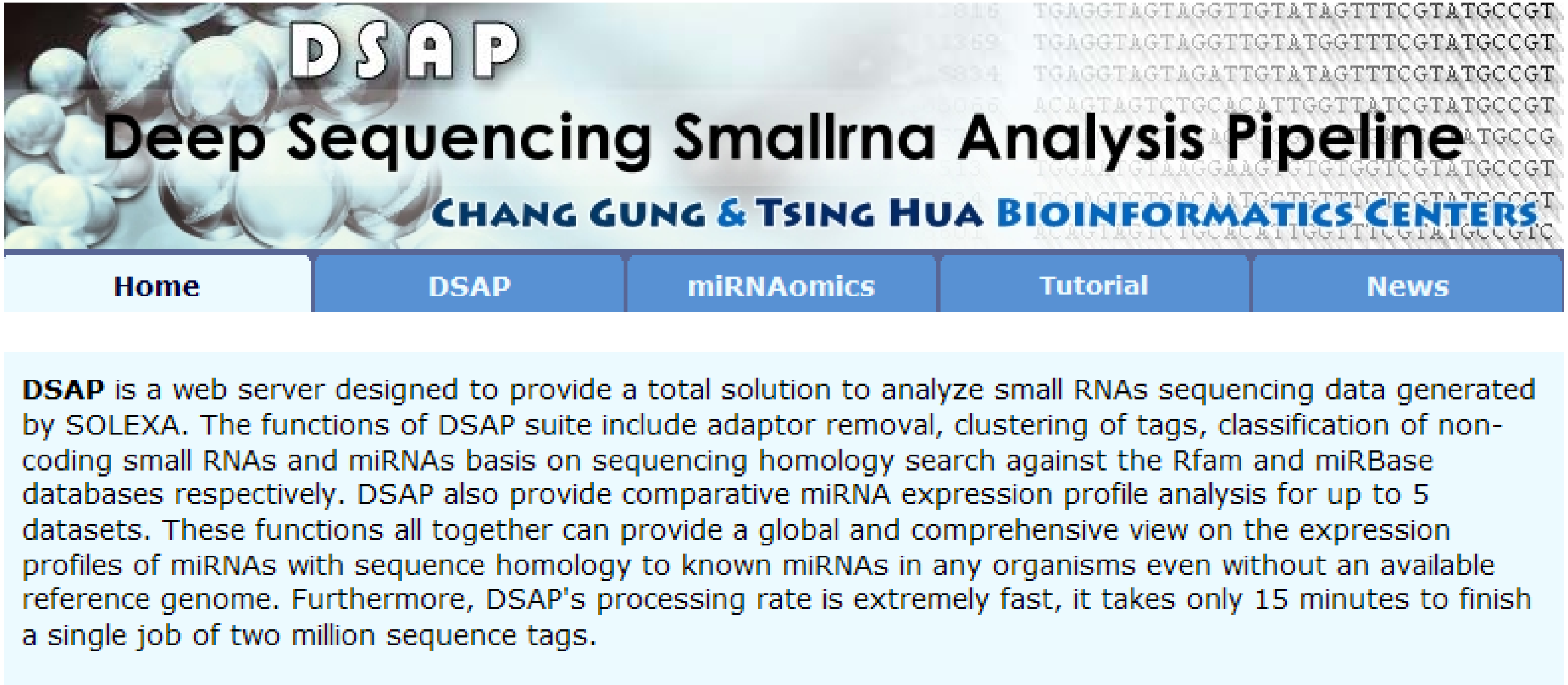
Mirror site @ Beijing Institute of Genomics , Chinese Academy of Sciences.

http://dsap.big.ac.cn

Mirror site @ Hong Kong Polytechnic University

http://dsap.polyu.edu.hk

http://dsap.cgu.edu.tw



DSAP is a web server designed to provide a total solution to analyze small RNAs sequencing data generated by SOLEXA. The functions of DSAP suite include adaptor removal, clustering of tags, classification of non-coding small RNAs and miRNAs basis on sequencing homology search against the Rfam and miRBase databases respectively. DSAP also provide comparative miRNA expression profile analysis for up to 5 datasets. These functions all together can provide a global and comprehensive view on the expression profiles of miRNAs with sequence homology to known miRNAs in any organisms even without an available reference genome. Furthermore, DSAP's processing rate is extremely fast, it takes only 15 minutes to finish a single job of two million sequence tags.