

Construction and Annotation on Non-model organisms- PAGE: Penaeus Genome Database



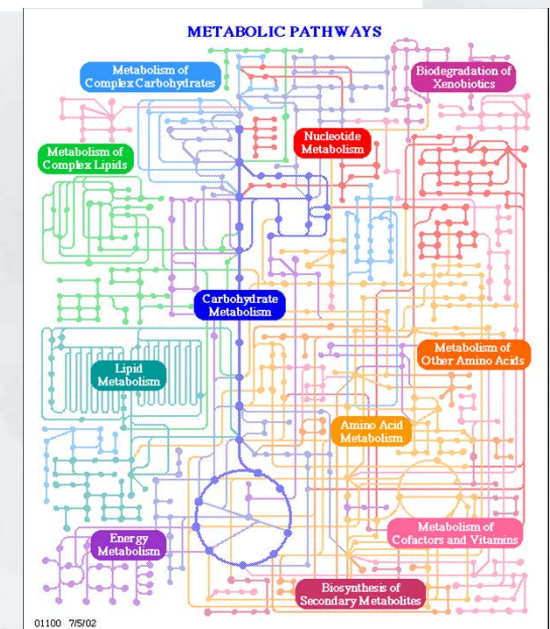
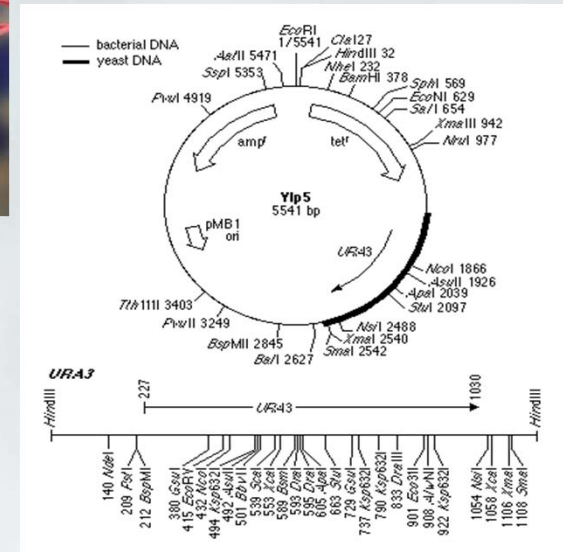
Reporter: Yu Bin Wang (王育彬)

7 Oct 2010



Non-model v.s. Genomic model

- Genome project
- Gene function
- Signal transduction pathway data completeness



Why use Non-model Organism

- *Apply to other industry*



- *Specificity in non-model organism*



How to Start

- Data collecting
- Association linkage
- Your own database construction





What's PAGE?

Penaeus Genome Database (PAGE) is a genome database with integrated analysis tools which is Penaeus genome oriented for over 200,000 Expressed Sequence Tags (ESTs). In PAGE, we provide sequences and tentative functional annotations for each assemble contigs and ESTs. Users can conduct search easily in keywords or sequences to those ESTs and contigs in specific or across species.

Search in PAGE

Contig

PAGE include sequence data of





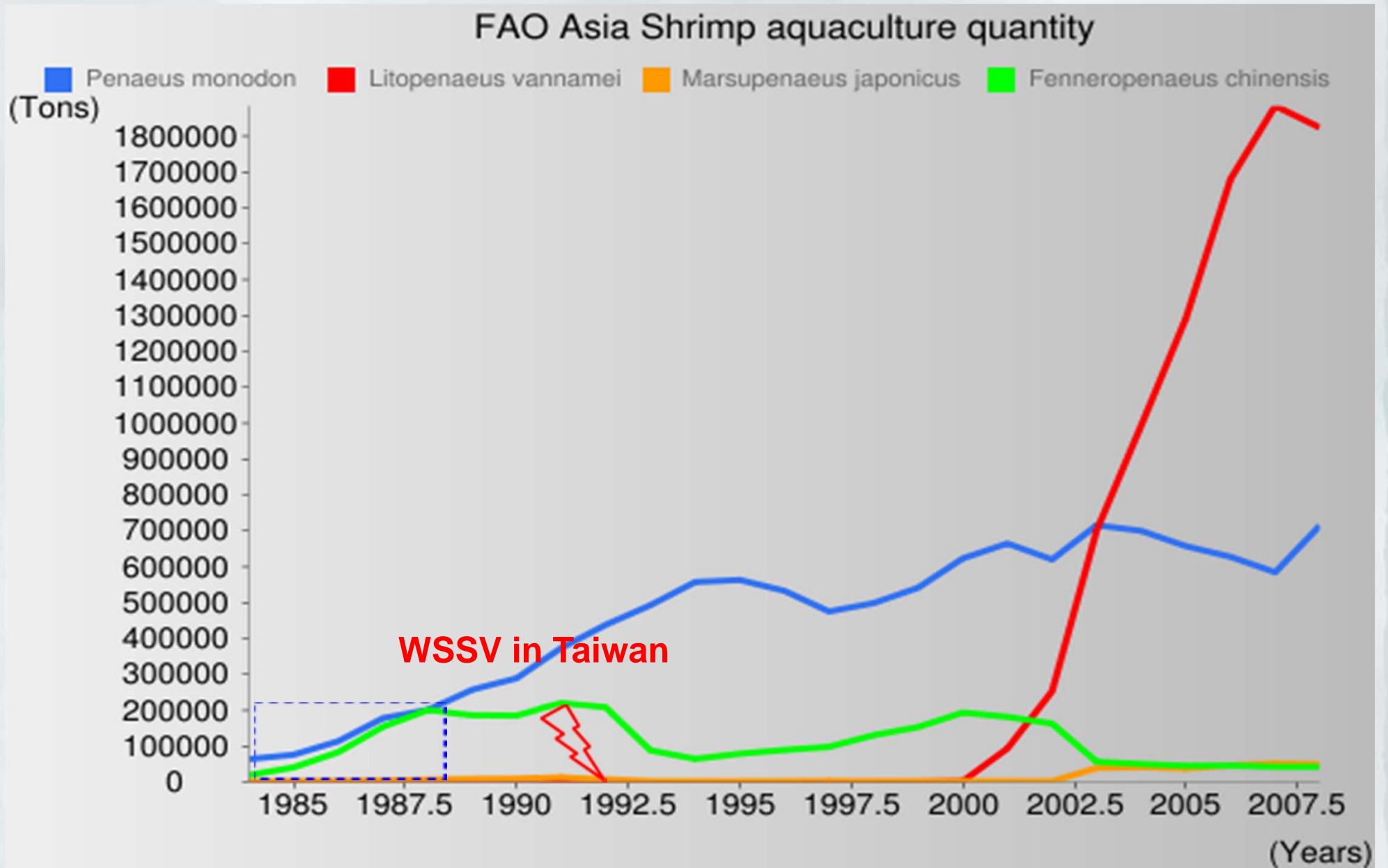
Penaeus monodon	Litopenaeus vannamei	Marsupenaeus japonicus	Fenneropenaeus chinensis
			

photo from [the Fish Database of Taiwan](#)

New function

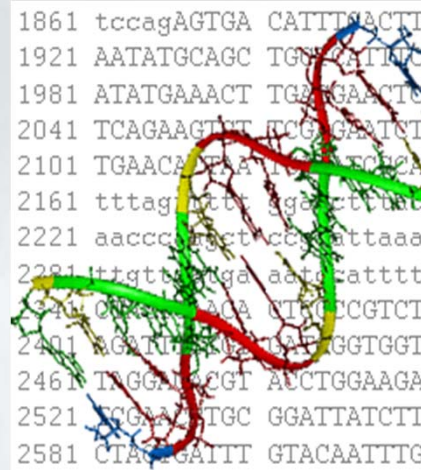
- [find shrimp sequence homologous to fly immunity proteins](#)

Shrimp Aquaculture in Asia



Data Collecting

- material: EST from NCBI

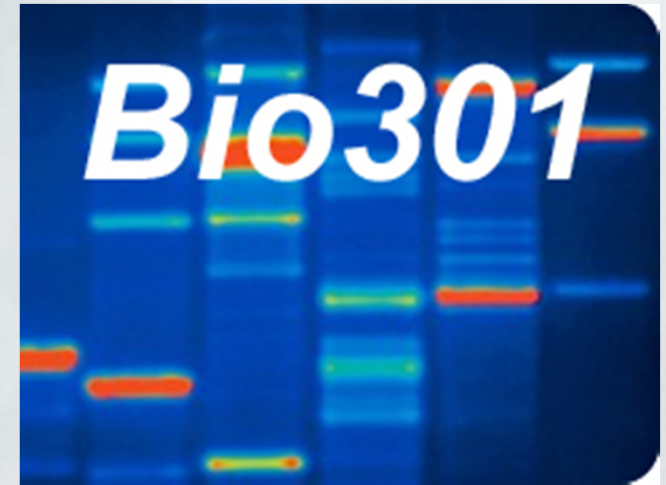


- species: Fenneropenaeus chinensis, Litopenaeus vannamei, Marsupenaeus japonicus and Penaeus monodon

- add extra EST from Prof. Lo's group



Association Linkage



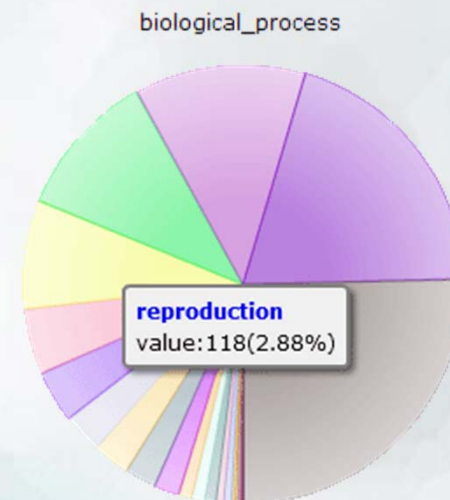
We use "[Bio301](#)" pipeline to get sequence association

- Assembly ESTs to Contigs



- Use Contigs Blast to Uniprot database to get annotation

- Get GeneOntology by uniprot path



Bio301 Pipeline

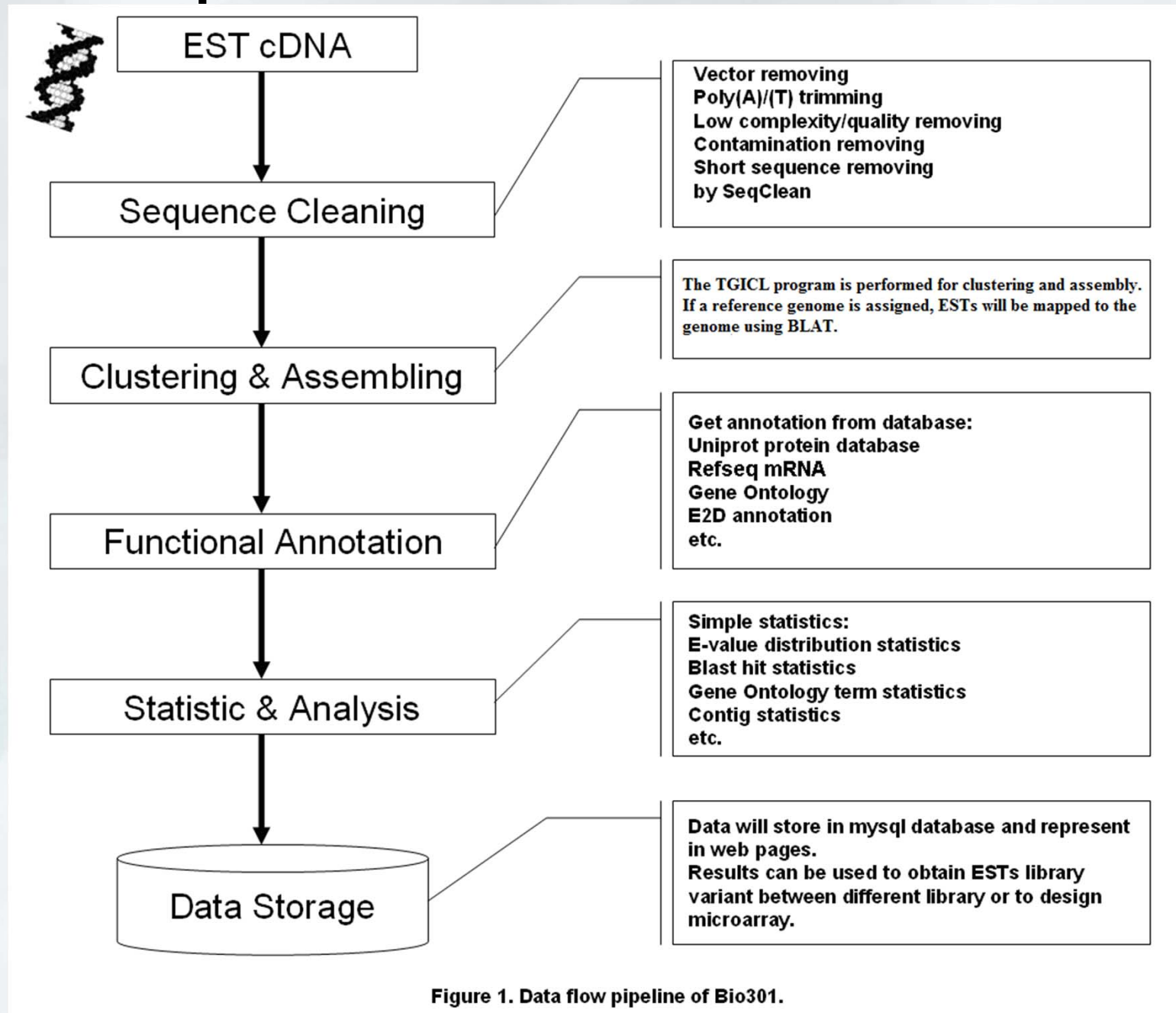


Figure 1. Data flow pipeline of Bio301.

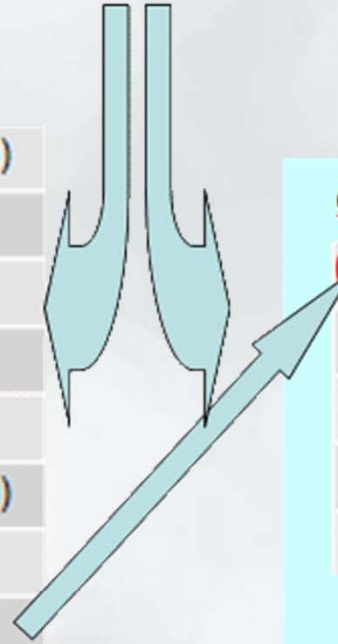
Association Linkage

species_table

taxid	species_name
6687	Penaeus monodon
6689	Litopenaeus vannamei
27405	Marsupenaeus japonicus
139456	Fenneropenaeus chinensis

est_info

<u>est_id</u>	mediumint(8)
gi	int(10)
gb	varchar(10)
des	tinytext
seq	text
taxid	mediumint(8)
inhouse	varchar(20)
contig_id	smallint(5)



contig_info

<u>contig_id</u>	smallint(5)
taxid	mediumint(8)
con_type	char(3)
con_ac	smallint(5)
con_seq	text

blast_info

<u>contig_id</u>	smallint(5)
des	tinytext
uniprot_id	varchar(10)
score	float
e_value	float
identity	tinyint(3)

contig_go

<u>contig_id</u>	smallint(5)
GO_ID	smallint(5)

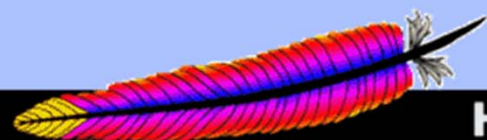
Your own database construction

taxid	species_name
6687	Penaeus monodon
6689	Litopenaeus vannamei
27405	Marsupenaeus japonicus
139456	Fenneropenaeus chinensis

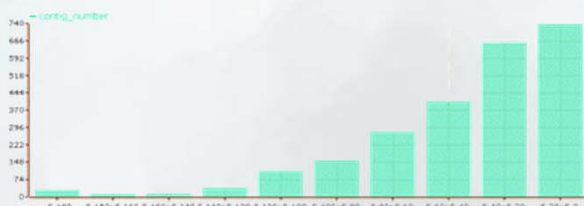
est_id	mediumint(8)
gi	int(10)
gb	varchar(10)
des	tinytext
seq	text
taxid	mediumint(8)
inhouse	varchar(20)
contig_id	smallint(5)

contig_id	smallint(5)
taxid	mediumint(8)
con_type	char(3)
con_ac	smallint(5)
con_seq	text
contig_id	smallint(5)
GO_ID	smallint(5)

contig_id	smallint(5)
des	tinytext
uniprot_id	varchar(10)
score	float
e_value	float
identity	tinyint(3)



Apache
HTTP SERVER PROJECT



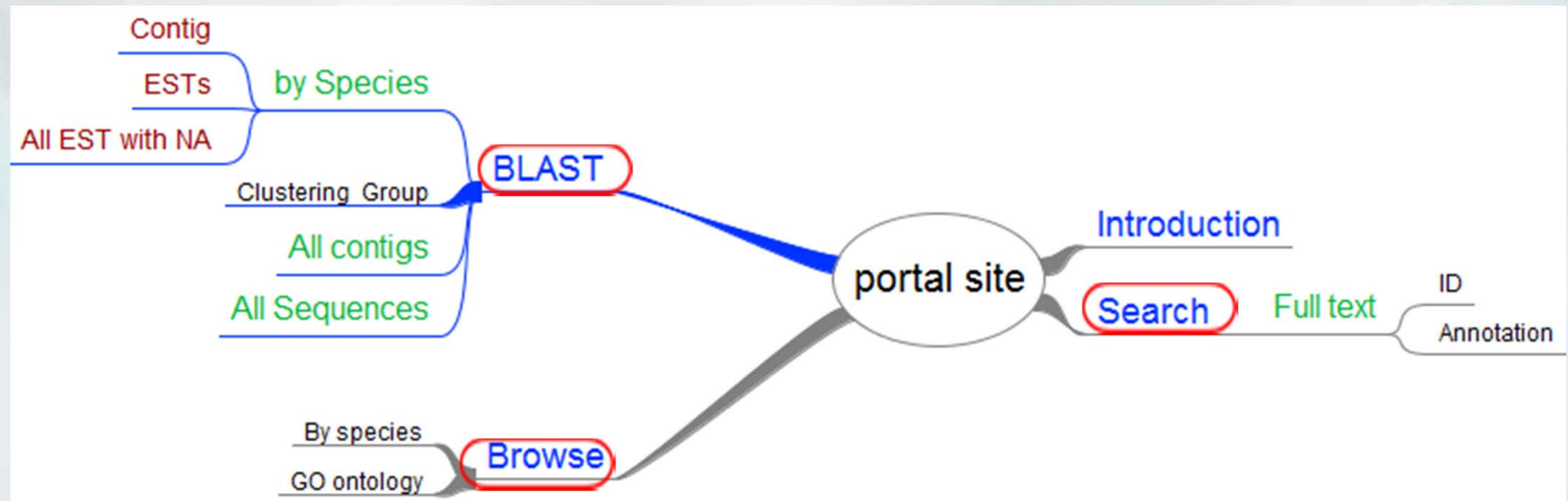
filtered: 7533

Type	Total ESTs	Number of groups	Average sequence length
Contig(TUC)	12403	2326	811
Contig(TUS)	5725	5725	507

[download the Contig fasta format](#)

Your own database construction

Web Schema





What's PAGE?

Penaeus Genome Database (PAGE) is a genome database with integrated analysis tools which is Penaeus genome oriented for over 200,000 Expressed Sequence Tags (ESTs). In PAGE, we provide sequences and tentative functional annotations for each assemble contigs and ESTs. Users can conduct search easily in keywords or sequences to those ESTs and contigs in specific or across species.

Search in PAGE

Contig

PAGE include sequence data of





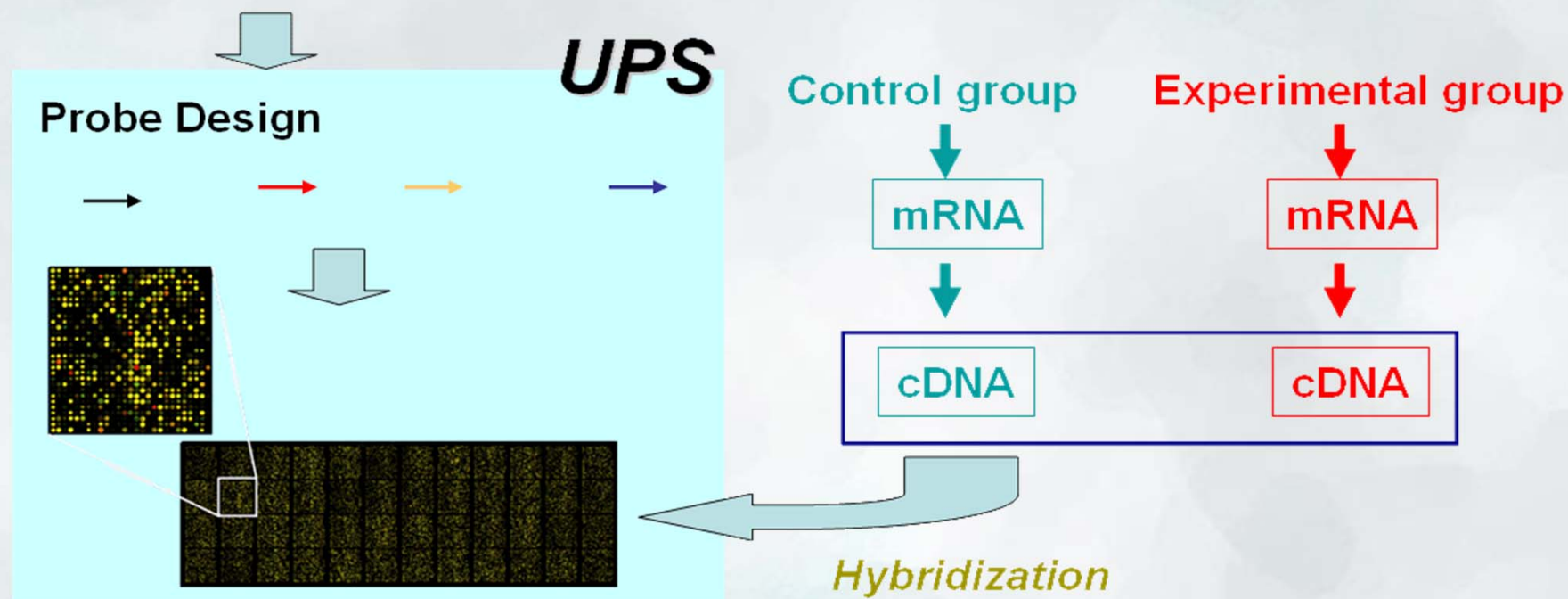
Penaeus monodon	Litopenaeus vannamei	Marsupenaeus japonicus	Fenneropenaeus chinensis
			

photo from [the Fish Database of Taiwan](#)

New function

- [find shrimp sequence homologous to fly immunity proteins](#)

Apply to experiment...



UPS(Unique Probe Selector)

<http://array.iis.sinica.edu.tw/ups>

UPS
Unique Probe Selector

2.0



Home



Demo



Help



Contact

Job Note : PTP family
Type of Probe Uniqueness : Unique Probe within group

Page 1

Output for UPS

Total : 111

Advanced Options filter

Sequence_ID	Rank	CG%	Tm	probe sequence	delta G
ABL	1	57	62	ctgagcagggtgatgacaggggacacctac	-0.401
ARG	1	33	56	cttgctacaataccttctcaattaaatct	1.129
EGFR	1	47	60	gaagagaaagaataccatgcagaaggaggc	0.97
TNK1	1	70	69	tgcgccacggagccttctcgtcgcctcgg	0.859
TXK	1	43	57	ggaatgacaaggtacgttttgatgatgag	1.587
TYK2	1	67	68	cgaaggccacgagtactaccgcgtgcgcga	-0.259
TYRO3	1	50	62	tactatcgtcaaggctgtgcctccaaactg	-1.193
VEGFR1	1	43	58	tttgtgattttggccttgcccgggatattt	0.146
VEGFR2	1	50	61	tctgtgactttggcttgcccgggatattt	0.304
VEGFR3	1	70	68	cgactacgtccgcaagggcagtgcccgct	-0.315

[\[Previous\]](#) [\[Next\]](#)

Output for Download

We provide more information for each probe in following files.

1. Best probes in fasta format without cross-hybridization
2. Best probes in fasta format
3. All probes in fasta format
4. All probes in CSV (with Tm, CG%, deltaG, Best_hit, Max_overlap, Identity)
5. In silico hybridization check for each probe by BlastN



Thanks for your Attention



LAB OF System Biology & Network Biology

中央研究院資訊科學研究所

@iis, Academia Sinica, TAIWAN

系統生物學與網路生物學實驗室

<http://eln.iis.sinica.edu.tw>