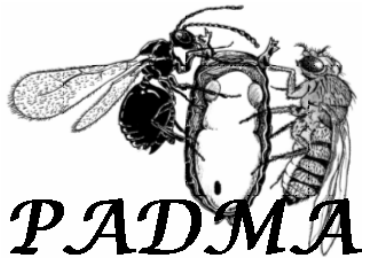


USER MANUAL

PADMA Database

Beta Release Version

Version: Beta Release v0.1
Release Date 11/1/2010

User Manual Content:

Section 1.0 – Welcome to PADMA Database

Pathogen Associated Drosophila MicroArray (PADMA) is a new database designed for easy retrieval and comparison of microarray results from experiments in *Drosophila* infected with natural pathogens or certain mutant backgrounds. We designed PADMA so users like you can access and search microarray data of various conditions (i.e. after parasitoid infections of whole larvae) to assess expression profile of target genes due the infection or mutation. The user can further drill and analyze pertinent data by refining and exporting the query results to an external data analysis application. As more pathogen-related microarray experiments become publicly available and uploaded to PADMA, the utility of this database will grow. We anticipate that PADMA will become an integral tool in designing and implementing your research.

This User Manual will provide all the necessary information and details for you to use PADMA Database. The manual is organized as follows: **Section 2** covers general navigation and introduction to features; **Section 3** covers PADMA Database data warehouse principles; **Section 4** Query Procedures and Examples; **Section 5** covers how a user can upload data (for private, restricted access only to the user); and **Section 6** covers copyright guidelines;

We suggest that you start by reading Section 2, "Getting Started." This section gives you the overview of all the features, and defines any terminology you may encounter during your use of PADMA Database. Should you have any questions, need further assistance, or have general comments, please don't hesitate to contact us, as described in **Section 2.4** or on the "Contact Us" tab on our website.

We are very happy to have you onboard, and we hope that you find PADMA Database an indispensable companion to your research efforts.

Section 2.0 – Getting Started

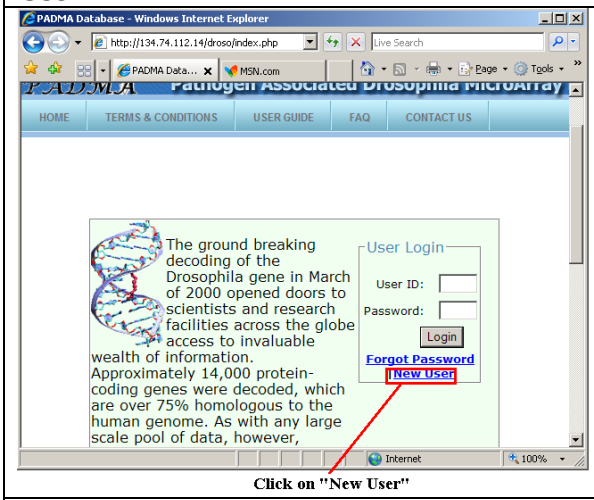
This section provides an introduction and overview of PADMA Database.

Section 2.1 – Registration (User ID)

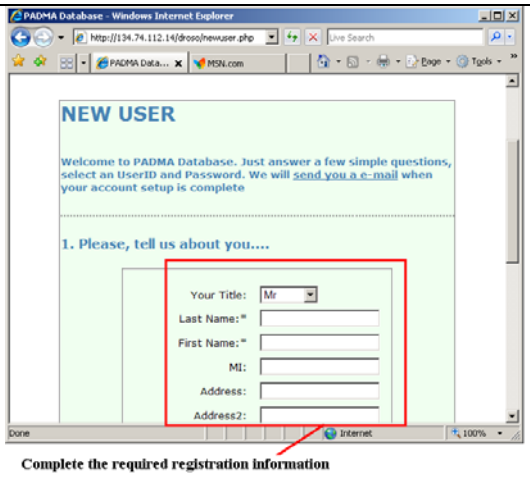
There are two types of PADAM users: Visitor and Registered User. Individuals who simply visit by landing on PADMA website have limited access to PADMA website contents, graphics, and demo/manual. We don't require Visitors to register or submit any information for use of our contents, graphics, or demo/manual on PADMA Website (please refer to ***Terms of Use*** for specific terms and conditions for accessing PADMA Website).

Individuals who wish to gain access to contents and materials beyond those permitted to Visitors must register with PADMA. We require registration to ensure security and data integration. Interested individuals can follow the following steps in obtaining User ID:

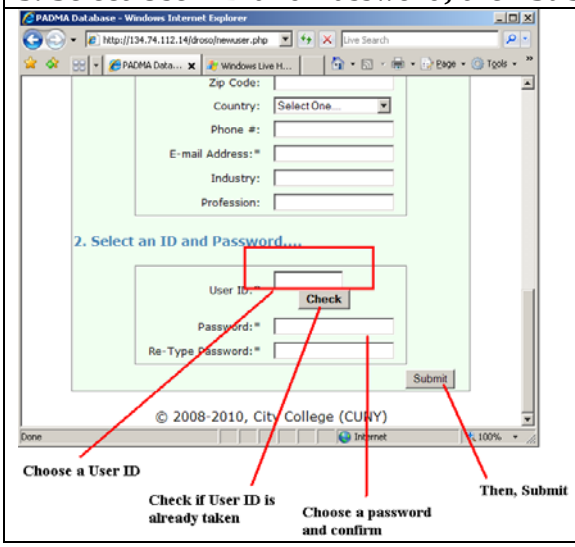
1. Log onto the website and click on "New User"



2. Complete the registration information

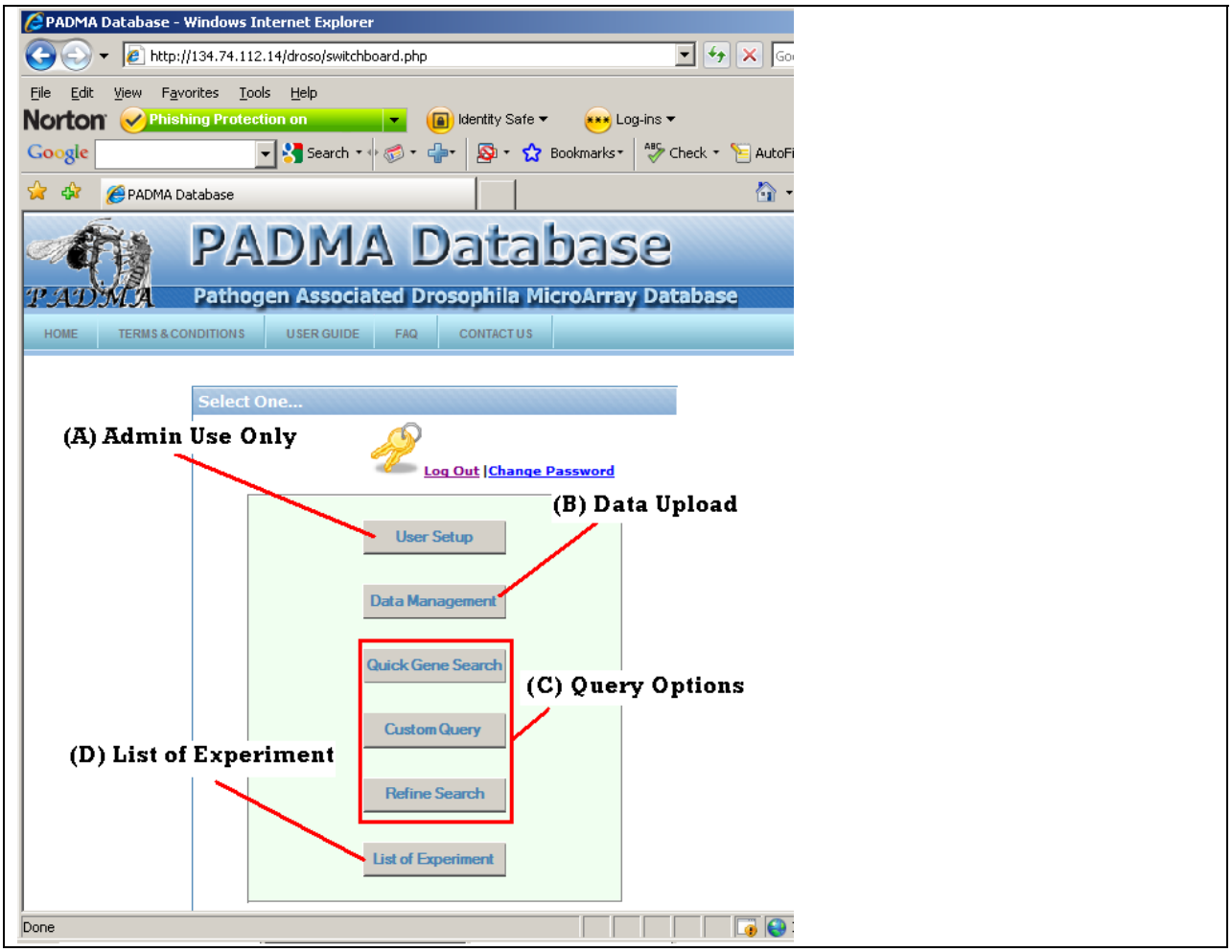


3. Select User ID and Password, then Submit



Once you successfully submit your registration information, the PADMA Database Administrator will get back to you via the email address you provide with further instructions on your access to registration-required portion of the website.

Section 2.2 – Navigating PADMA Website



(A) Administrator Use Only: Access to Administrator functions.

(B) Data Upload: Access to upload microarray data by individual users. Data from uploads by individual will only be displayed to the individual that uploaded the data. Thus, the data will remain confidential and private. For detail, refer to **Section 4.0**, below.

(C) Query Options: There are three ways of querying the data warehouse. For detail, refer to **Section 3.0**, below.

(D) List of Experiment: Summaries of individual datasets in the data warehouse. This alleviates for users to read the full journal to understand the experimental conditions.

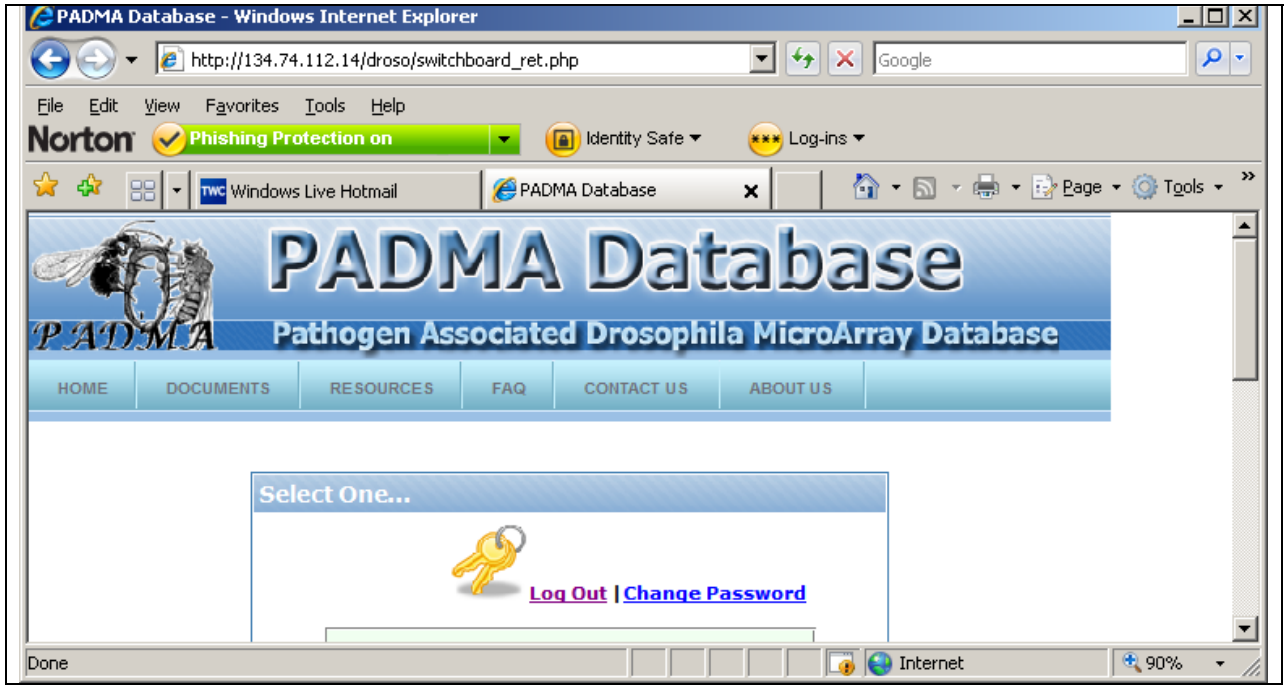
Section 2.3 – PADMA Definitions

While most of the terms used throughout PADMA website, PADMA Database, and all the contents and documents associated are straightforward, there may be some words that may carry special meaning. Thus, we created a list of possible terms that may need a little explanation:

Active Category	In PADMA, there are three categories: Parasitoid, Microbial, and Mutant. Thus, category refers to either a mutant or pathogen, and if pathogen, what type of infection (parasitic or microbial).
Active Species	This refers to a specific pathogen (like *A. tabida* infection) or a specific mutant (like $Toll^{10b}$) that is altering the gene expression compared to the control group (wild-type, uninfected, etc.).
Data Type: Metadata, Reference data, Experiment data	PADMA makes reference to three types of data: Metadata, Reference Data, and Experiment Data. Metadata is commonly known as non numeric data that is not analyzed, like class variables. Reference Data contains gene/probe related data we obtained from various sources and compiled in PADMA. These are universal references to any dataset. Experiment Data contains experiment specific data, including Fold Induction. It is this data type Users prepare and upload.
Dataset	A collection of variables and associated values are termed dataset. PADMA interchangeably uses this term to mean processed microarray in PADMA Format, as well as loose raw microarray replicate results.
Experiment Name	This is an arbitrary name that PADMA assigns to a dataset. The nomenclature is as follows: 6 letters of the first author's last name (if shorter than 6, we fill in with an X per each letter short), 5 letter for pathogen or mutation, and 2 digit hour followed by HR. If it's not an infection, then we place XX in lieu of 2 digit hour.
Experiment Subject	The specimen that the RNA was extracted to perform the microarray, which could be whole body larvae, specific tissue, or even adult flies.
Fold Induction	It is simply Experimental/Control of the average signal value obtained from the raw microarray dataset deposited by the publication authors. Since experimental conditions are usually not the same across publications, we advise our Users to take caution in making comparisons.
Regulation Value	While Fold Induction is the raw value induction averaged over multiple replicates, Regulation Value is a derived value consisting of -1, 0, 1. This provides a quick reference to select potentially differentially expressed genes (probe set). The calculation in deriving these three values is done through logical formatting with specific cutoff values. Generally, -1 means that the specific gene or probe set is likely downregulated, while 1 means it is likely upregulated. A value of 0 is indicative that the specific gene or probe set is likely unchanged compared to the control group. As in Fold Induction, since experimental conditions are usually not the same across publications, we advise our Users to take caution in making comparisons/conclusions.

Section 2.4 – Feedback and Contacting PADMA

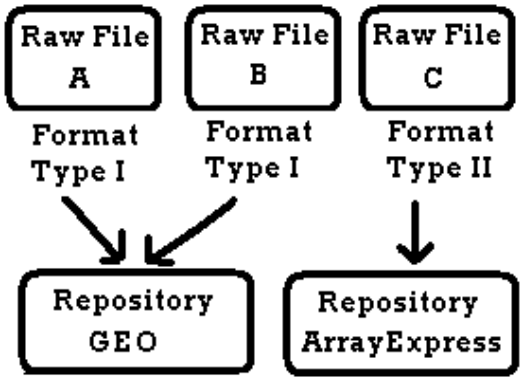
To contact PADMA, please visit “Contact Us” on our website:



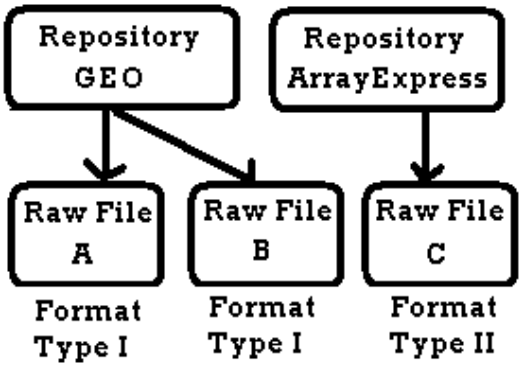
We appreciate any comments or feedbacks you may have. Also, should you have any questions, do not hesitate to contact us. We will try our best to get back to you as soon as possible.

Section 3.0 – About PADMA's Data Warehouse

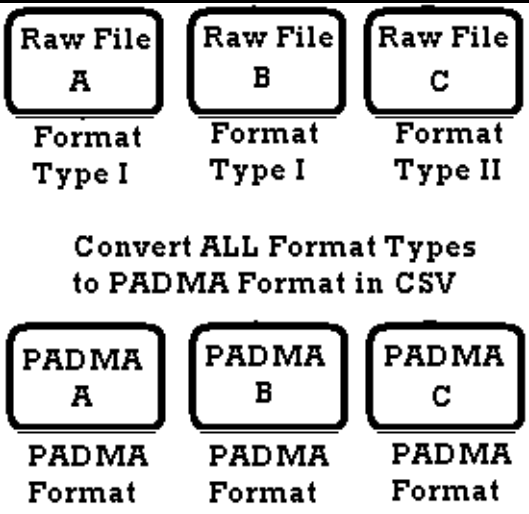
Authors deposit microarray datasets they publish into one of many different repositories around the world. These files are usually raw data files, formatted according to different requirements outlined by each of the repositories.



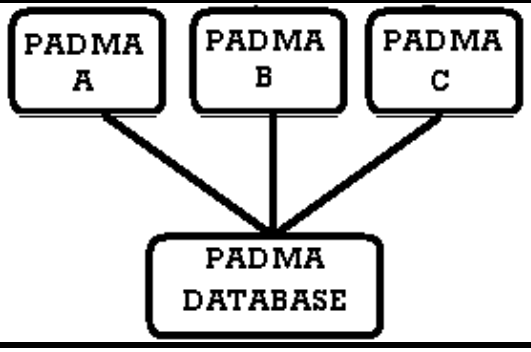
Based on the supplemental information authors provide in their publications, PADMA accesses different repositories and downloads raw data file.



PADMA examines raw files downloaded from repositories and converts them into unified PADMA Format. For more information on specific PADMA Format requirements, please refer to **Section 5.0**, below.



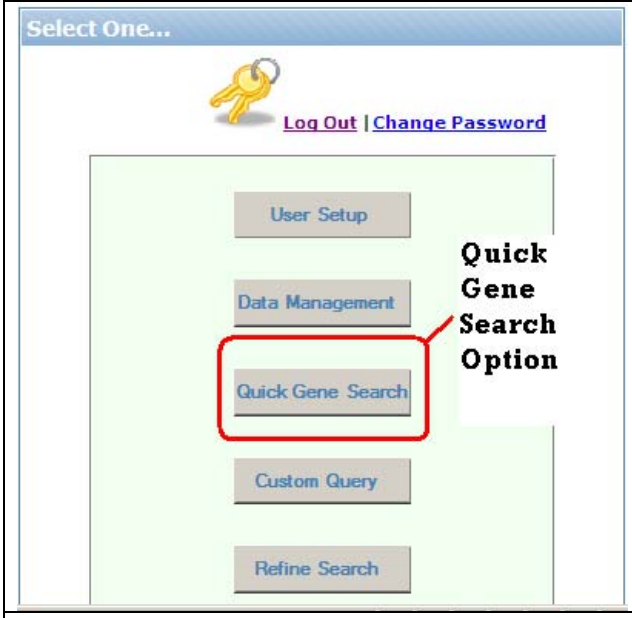
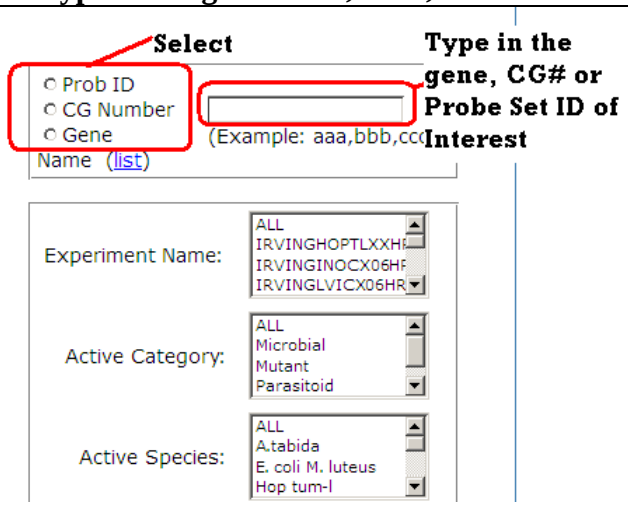
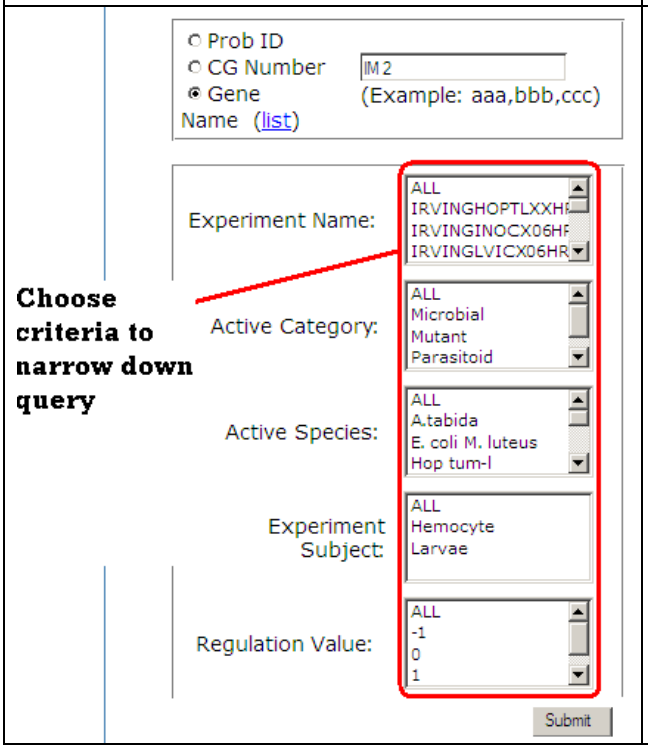
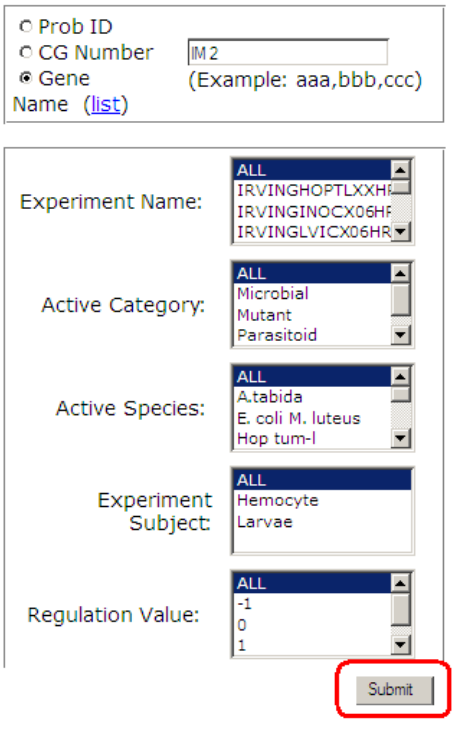
The PADMA formatted raw files are then uploaded into PADMA Data Warehouse, which is then accessible through PADMA Database.



Section 4.0 – Querying PADMA

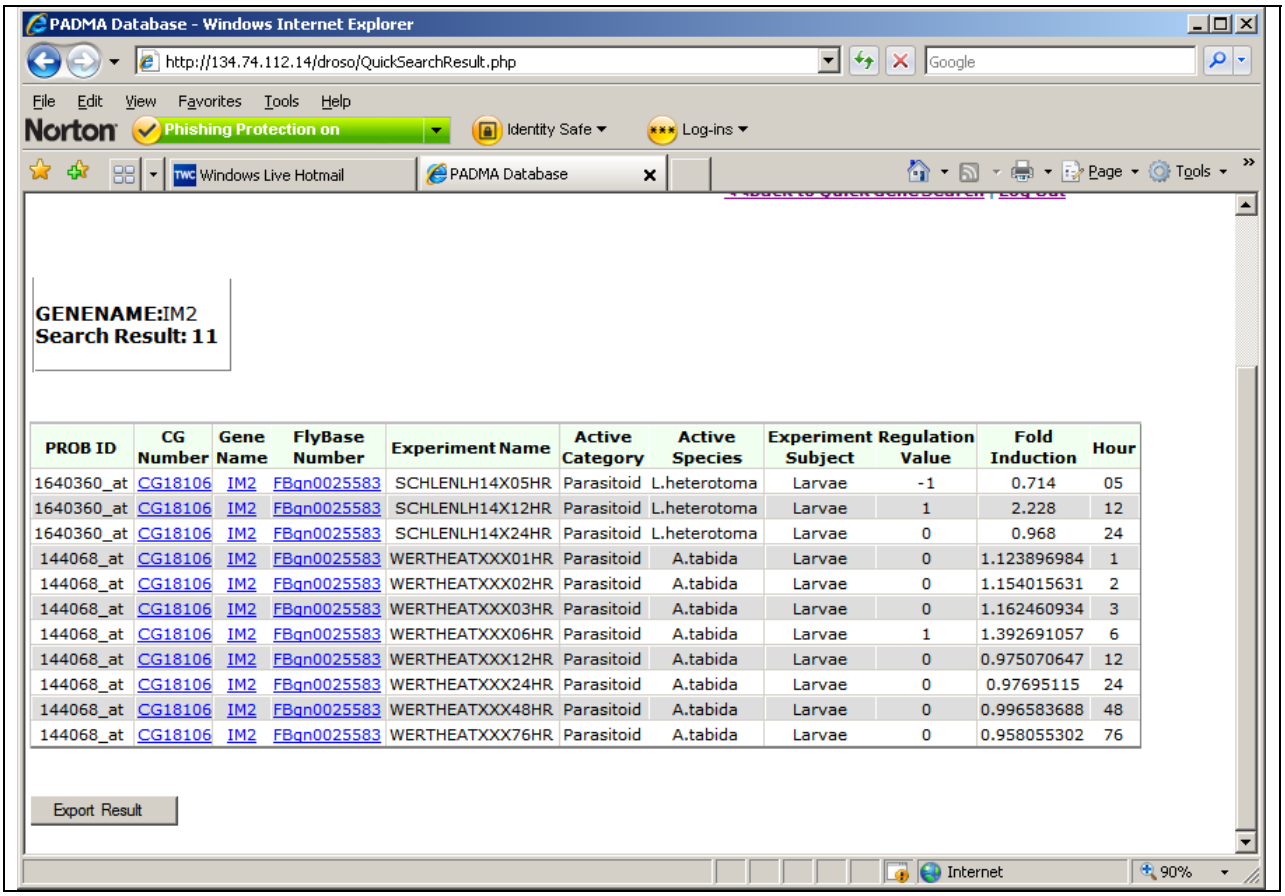
There are three query options available on PADMA. Each option is uniquely designed for a specific query purpose, which are detailed in subsequent sub-section. However, all options will lead to a query result output screen with the option of exporting the results.

Section 4.1 – Quick Search

<table>
<tr><th>1. Select “Quick Gene Search”</th><th>2. Type in the gene name, CG#, or Probe ID</th></tr>
<tr><td> 
</td><td> 
</td></tr>
<tr><th>3. Choose different criteria</th><th>4. Submit</th></tr>
<tr><td> 
</td><td> 
</td></tr>
</table>

Result from Quick Gene Search:

Following the four steps for Quick Search as detailed in the prior page will results in this query output table:



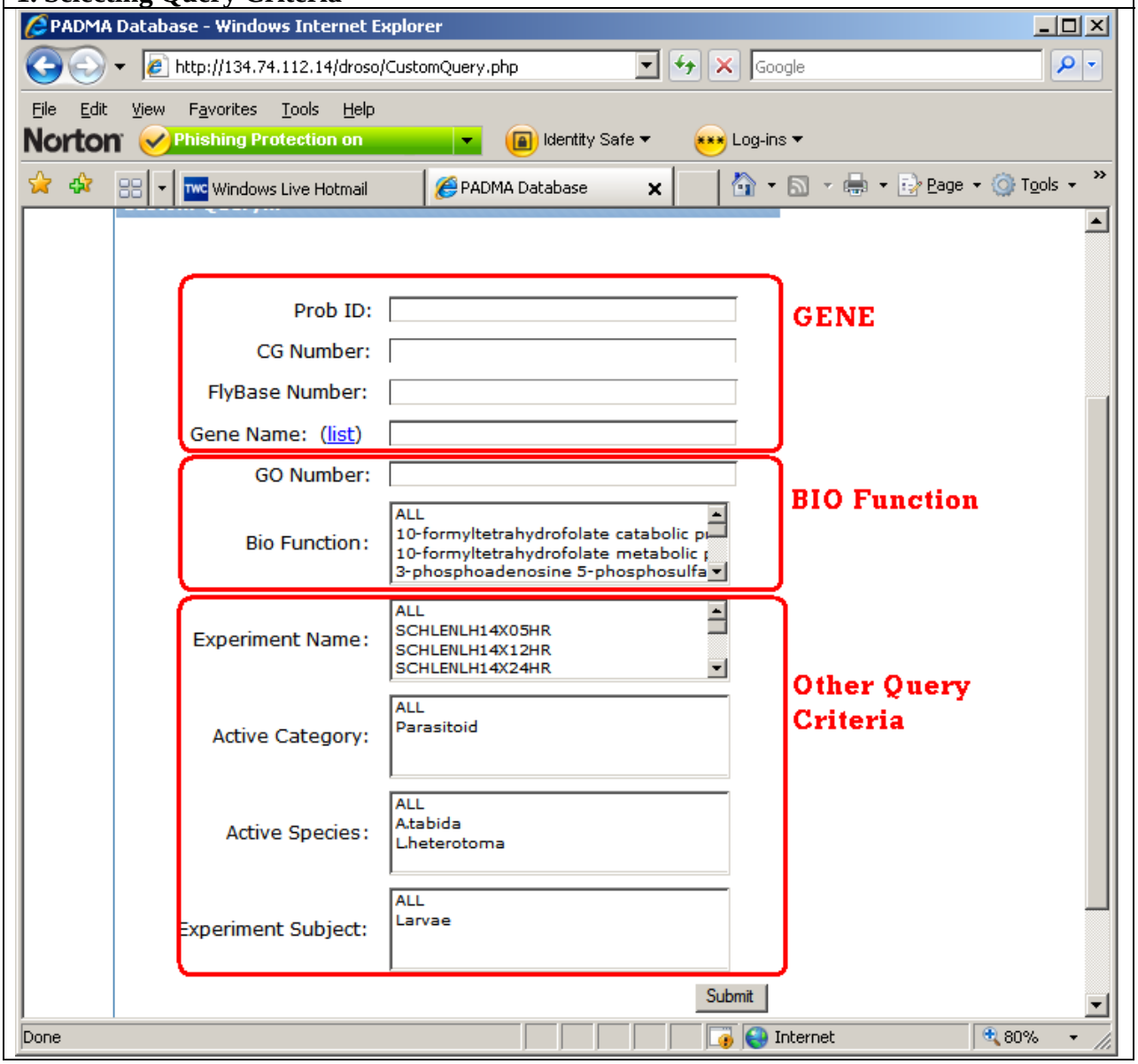
PROB ID	CG Number	Gene Name	FlyBase Number	Experiment Name	Active Category	Active Species	Experiment Subject	Regulation Value	Fold Induction	Hour
1640360_at	CG18106	IM2	FBgn0025583	SCHLENLH14X05HR	Parasitoid	L.heterotoma	Larvae	-1	0.714	05
1640360_at	CG18106	IM2	FBgn0025583	SCHLENLH14X12HR	Parasitoid	L.heterotoma	Larvae	1	2.228	12
1640360_at	CG18106	IM2	FBgn0025583	SCHLENLH14X24HR	Parasitoid	L.heterotoma	Larvae	0	0.968	24
144068_at	CG18106	IM2	FBgn0025583	WERTHEATXXX01HR	Parasitoid	A.tabida	Larvae	0	1.123896984	1
144068_at	CG18106	IM2	FBgn0025583	WERTHEATXXX02HR	Parasitoid	A.tabida	Larvae	0	1.154015631	2
144068_at	CG18106	IM2	FBgn0025583	WERTHEATXXX03HR	Parasitoid	A.tabida	Larvae	0	1.162460934	3
144068_at	CG18106	IM2	FBgn0025583	WERTHEATXXX06HR	Parasitoid	A.tabida	Larvae	1	1.392691057	6
144068_at	CG18106	IM2	FBgn0025583	WERTHEATXXX12HR	Parasitoid	A.tabida	Larvae	0	0.975070647	12
144068_at	CG18106	IM2	FBgn0025583	WERTHEATXXX24HR	Parasitoid	A.tabida	Larvae	0	0.97695115	24
144068_at	CG18106	IM2	FBgn0025583	WERTHEATXXX48HR	Parasitoid	A.tabida	Larvae	0	0.996583688	48
144068_at	CG18106	IM2	FBgn0025583	WERTHEATXXX76HR	Parasitoid	A.tabida	Larvae	0	0.958055302	76

You can compare the gene expression profiles on the table, access hyperlink to FlyBase to obtain further gene related information, or export the table in csv format by selecting “Export Result” on the bottom of the page. For details of Export, please refer to **Section 4.4** below.

Please note that the Probe ID column may show multiple probe ID for the same gene. This could be due to a combination of different GeneChip used by the authors of different publications (i.e. Affymetrix Drosophila Genome Version 1 or 2), and two or more probe sets mapping to the same gene.

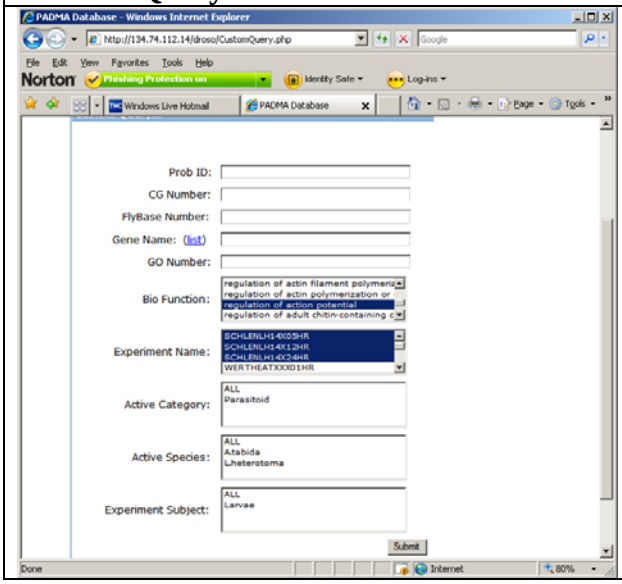
Section 4.2 – Custom Search

1. Selecting Query Criteria

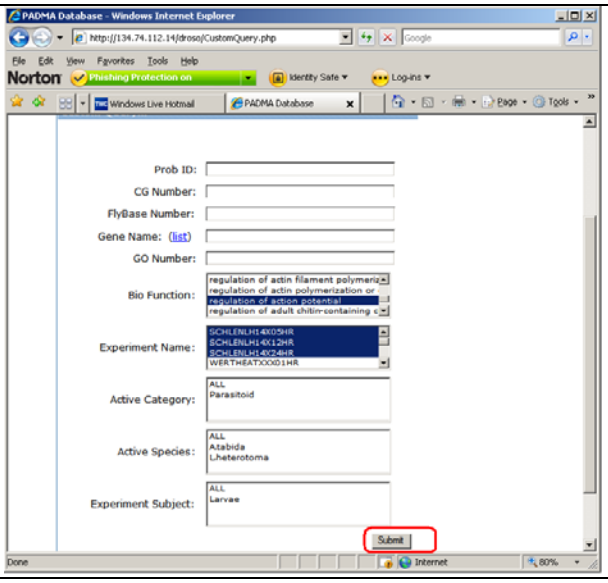


If you are interested in a specific Bio Function and want to search for genes that have been associated with such function by Gene Ontology, you can select that a specific Bio Function (or if multiple, press the "control" key on your key board and select more). You can narrow your search by selecting other query criteria.

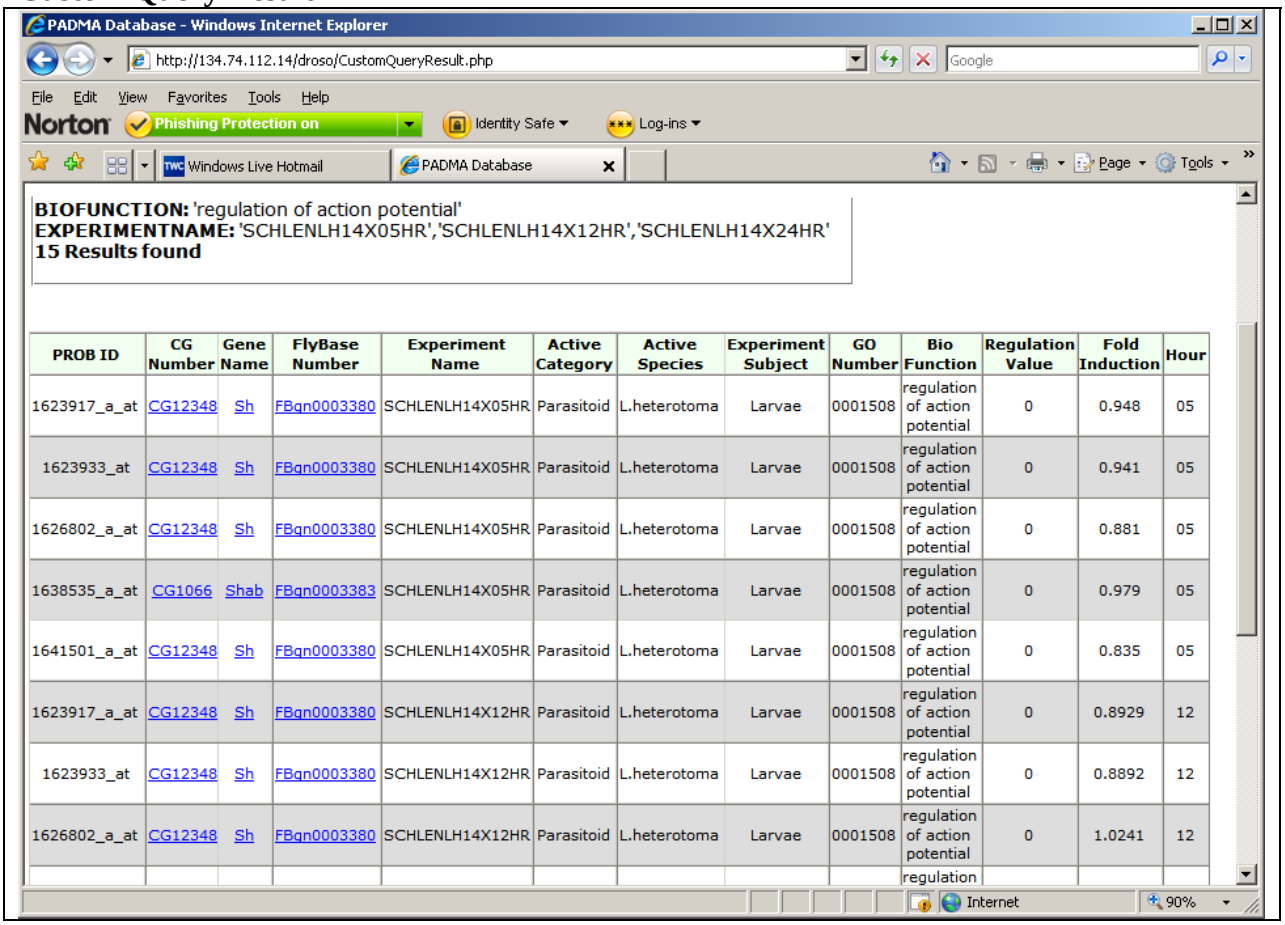
2. Select Query Criteria



3. Submit



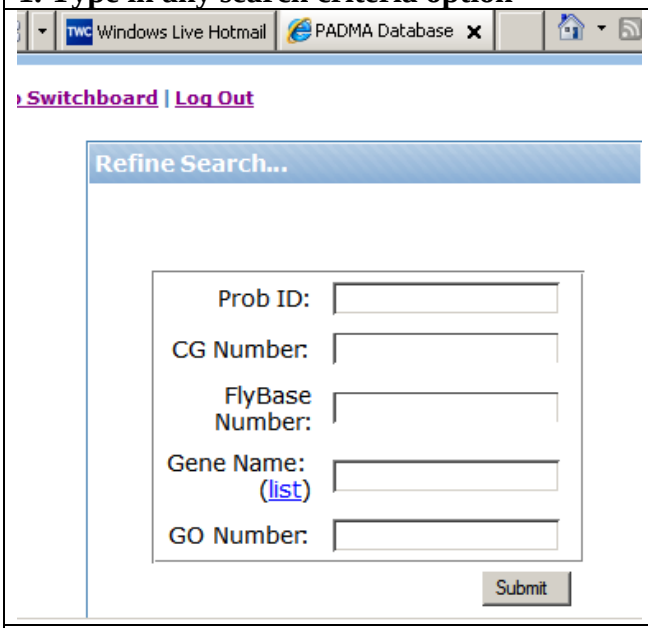
Custom Query Result



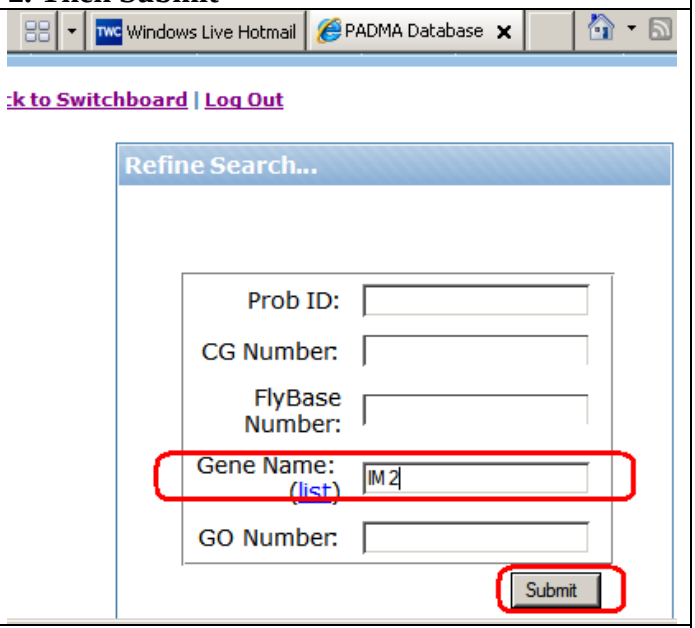
PROB ID	CG Number	Gene Name	FlyBase Number	Experiment Name	Active Category	Active Species	Experiment Subject	GO Number	Bio Function	Regulation Value	Fold Induction	Hour
1623917_a_at	CG12348	Sh	FBgn0003380	SCHLENLH14X05HR	Parasitoid	L.heterotoma	Larvae	0001508	regulation of action potential	0	0.948	05
1623933_at	CG12348	Sh	FBgn0003380	SCHLENLH14X05HR	Parasitoid	L.heterotoma	Larvae	0001508	regulation of action potential	0	0.941	05
1626802_a_at	CG12348	Sh	FBgn0003380	SCHLENLH14X05HR	Parasitoid	L.heterotoma	Larvae	0001508	regulation of action potential	0	0.881	05
1638535_a_at	CG1066	Shab	FBgn0003383	SCHLENLH14X05HR	Parasitoid	L.heterotoma	Larvae	0001508	regulation of action potential	0	0.979	05
1641501_a_at	CG12348	Sh	FBgn0003380	SCHLENLH14X05HR	Parasitoid	L.heterotoma	Larvae	0001508	regulation of action potential	0	0.835	05
1623917_a_at	CG12348	Sh	FBgn0003380	SCHLENLH14X12HR	Parasitoid	L.heterotoma	Larvae	0001508	regulation of action potential	0	0.8929	12
1623933_at	CG12348	Sh	FBgn0003380	SCHLENLH14X12HR	Parasitoid	L.heterotoma	Larvae	0001508	regulation of action potential	0	0.8892	12
1626802_a_at	CG12348	Sh	FBgn0003380	SCHLENLH14X12HR	Parasitoid	L.heterotoma	Larvae	0001508	regulation of action potential	0	1.0241	12
									regulation			

Section 4.3 – Refine Search

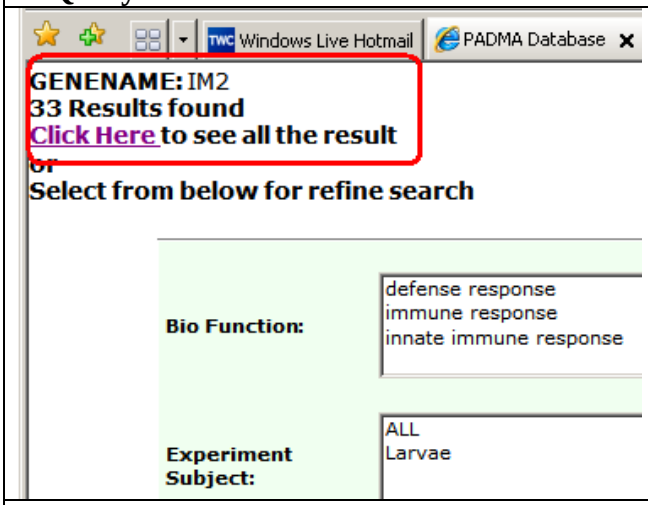
1. Type in any search criteria option



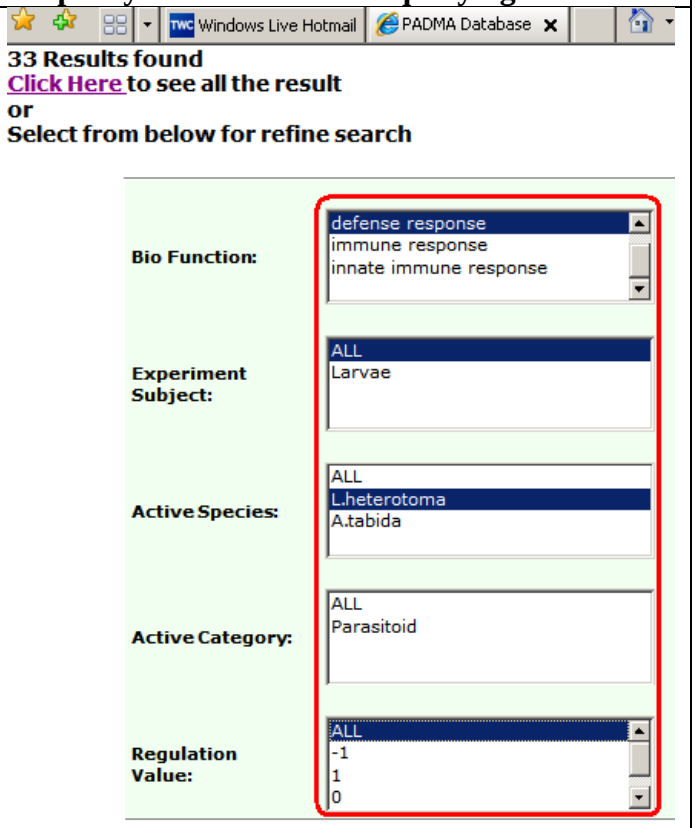
2. Then Submit



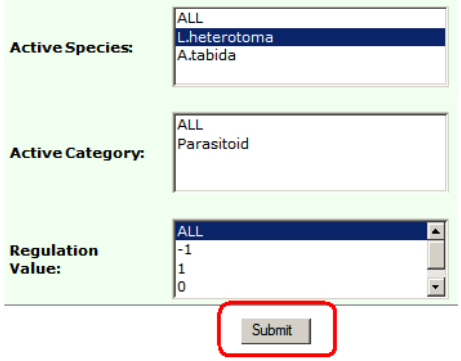
3. Query Results Found 33 Records



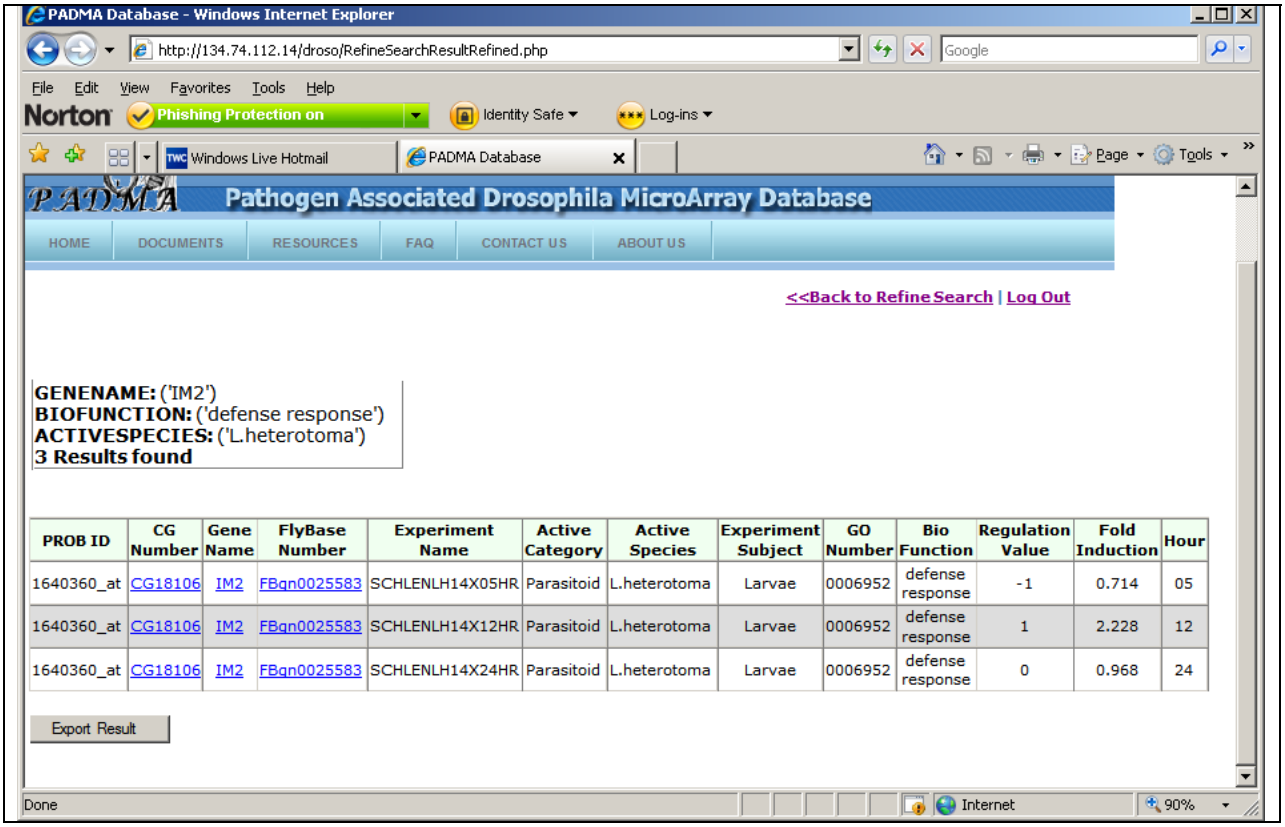
4. Specify criteria for refine querying



5. Submit



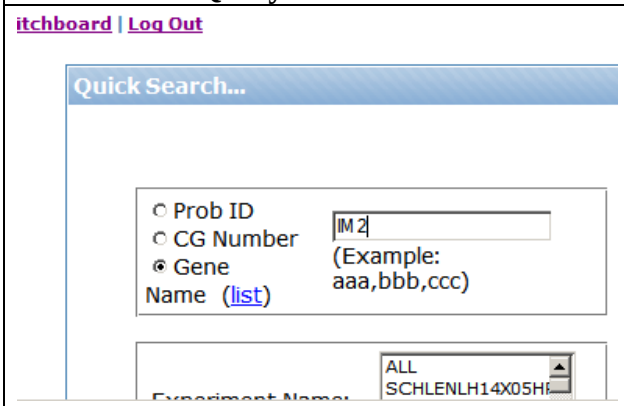
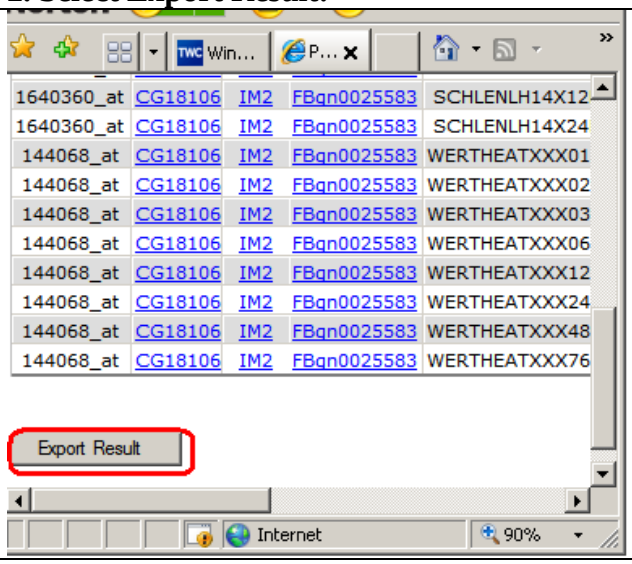
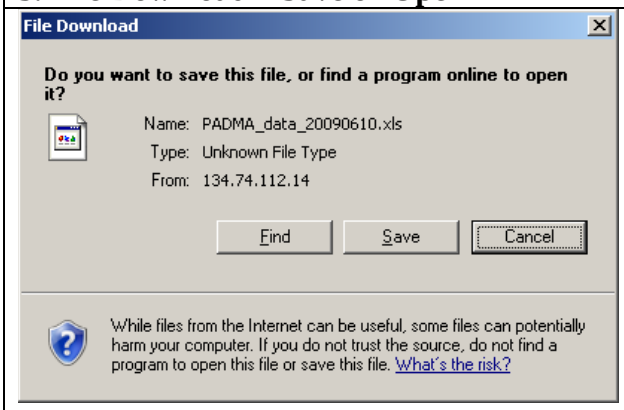
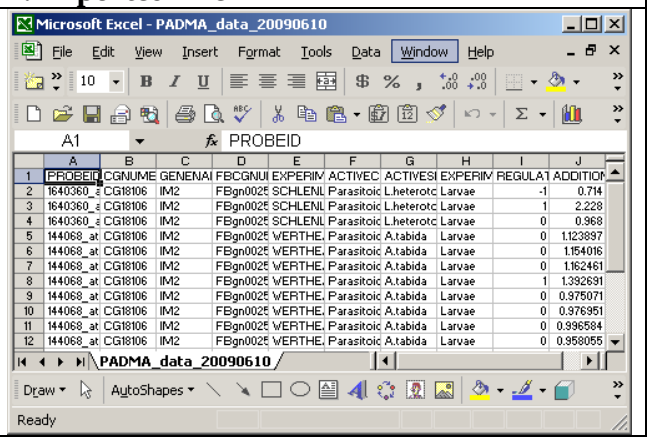
Refine Query Results



PROB ID	CG Number	Gene Name	FlyBase Number	Experiment Name	Active Category	Active Species	Experiment Subject	GO Number	Bio Function	Regulation Value	Fold Induction	Hour
1640360_at	CG18106	IM2	FBgn0025583	SCHLENLH14X05HR	Parasitoid	L.heterotoma	Larvae	0006952	defense response	-1	0.714	05
1640360_at	CG18106	IM2	FBgn0025583	SCHLENLH14X12HR	Parasitoid	L.heterotoma	Larvae	0006952	defense response	1	2.228	12
1640360_at	CG18106	IM2	FBgn0025583	SCHLENLH14X24HR	Parasitoid	L.heterotoma	Larvae	0006952	defense response	0	0.968	24

Section 4.4 – Exporting Files

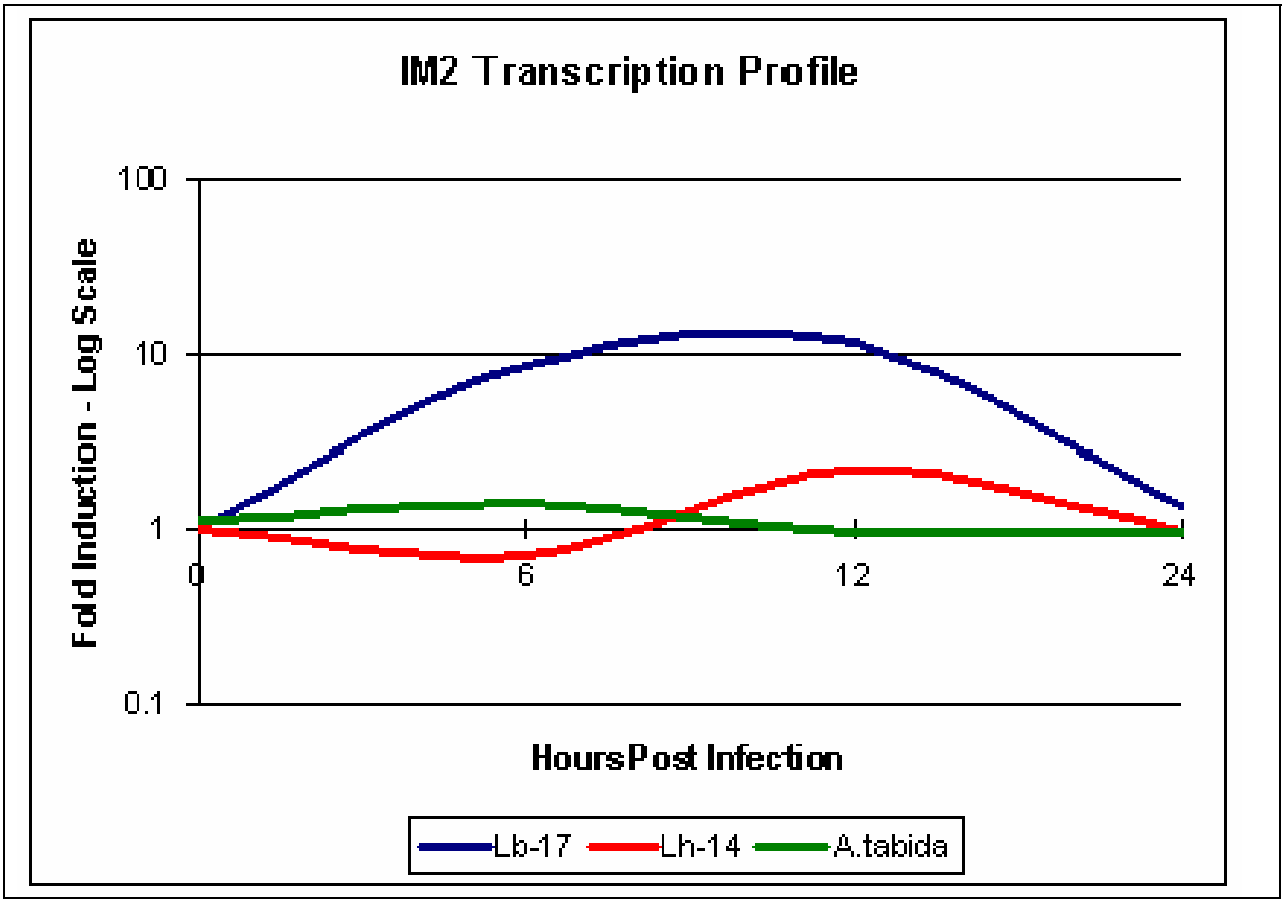
In addition to displaying a query result table, PADMA also provide Users with the option of exporting query results into an MS Excel workbook for further comparisons and analysis. This option is available to all Query Types (Quick Gene Search, Custom Query, and Refine Query).

1. Perform a Query	2. Select Export Result:
	
3. File Download – Save or Open	**4. Exported File**
	

The exported file will have the same columns as the query result table. Therefore, depending on your Query Type, you will have different columns exported.

Graphing Exported Data

One of the powerful use of PADMA is the ability of graphing gene expression profile for a visual screening. While there are many conditions and assumptions made in comparing dataset to dataset (which are obtained from different publications, and thus, different experimental conditions), it is of great utility to have all these information in one graph.



Section 5.0 – User Upload

In PADMA, users are allowed to upload their own microarray (self-microarray) dataset directly onto the data warehouse without making it public. This allows users to easily compare and contrast self-microarray results with microarray results in the PADMA Data Warehouse (public-microarray). Among many, one important utility of user upload is the ability for users to confidentially upload self-microarray results and take advantage of PADMA's powerful metadata referencing capabilities. This not only saves time and effort, but it also ensures that self-microarray results are not made public while users advance their research efforts.

Section 5.1 – Upload File Format

Generally, Affymetrix or other microarray service providers will send the raw microarray results in text files, among other files (like graphics, charts, documentation). The User needs to follow the steps outlined below in converting these raw files into PADMA Format Files, which is saved as a Comma Separated Value (csv) file.

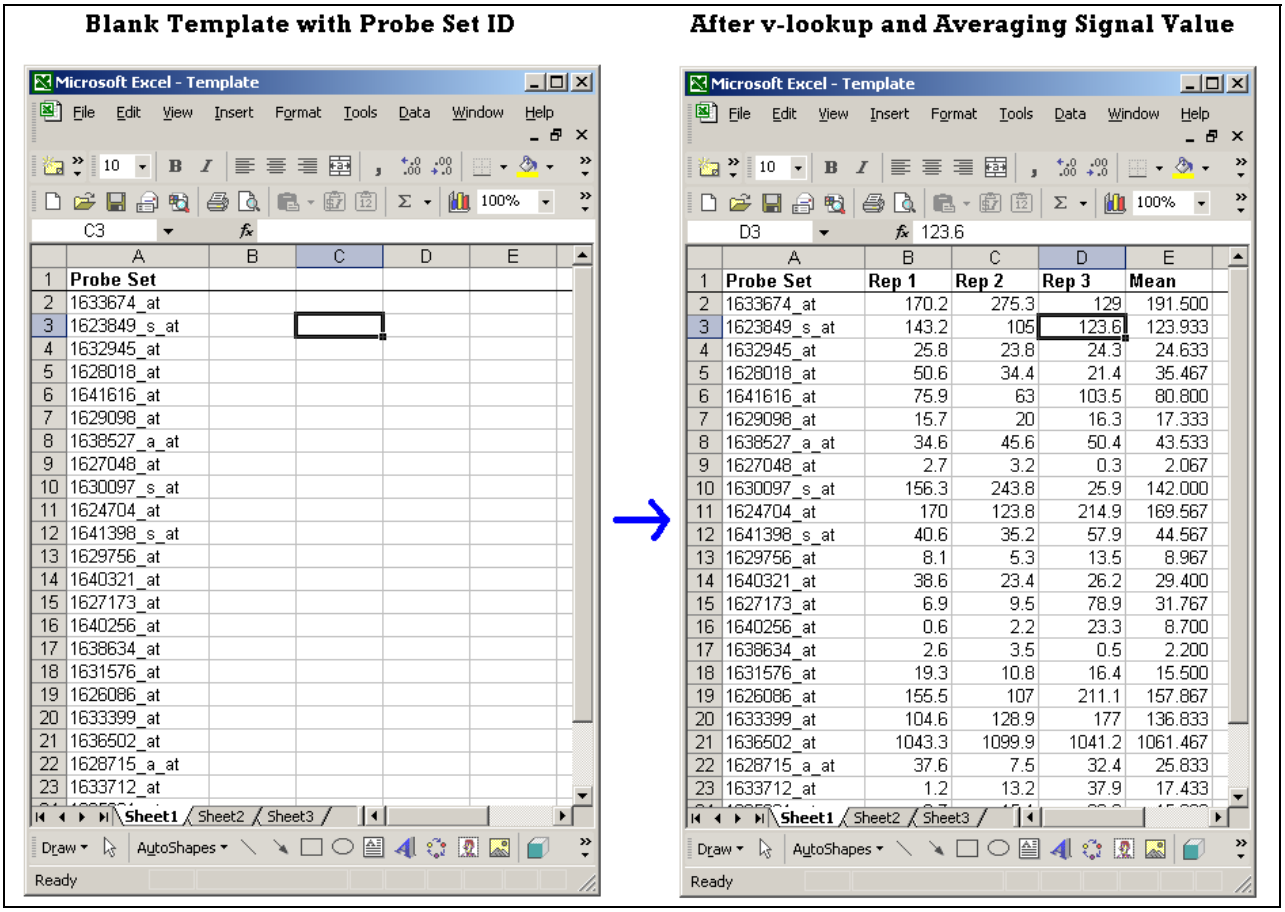
The easiest way to create a csv file is by using MS Excel. In MS Excel, each column is separated by a comma. Therefore, after scrubbing the data in Excel, the User can save the file as csv. This is easy and convenient.

Step 1. Combine Replicate Files. Combine all the replicates into one file by taking the mean (average) of the signal value. Generally, since experiments are conducted in replicates, the User is advised to average across all replicates. Some signal values may or may not carry significance. While Users are advised to pay attention to p and q values provided in the raw file, due to the volume of data, it may be practical to include all values.

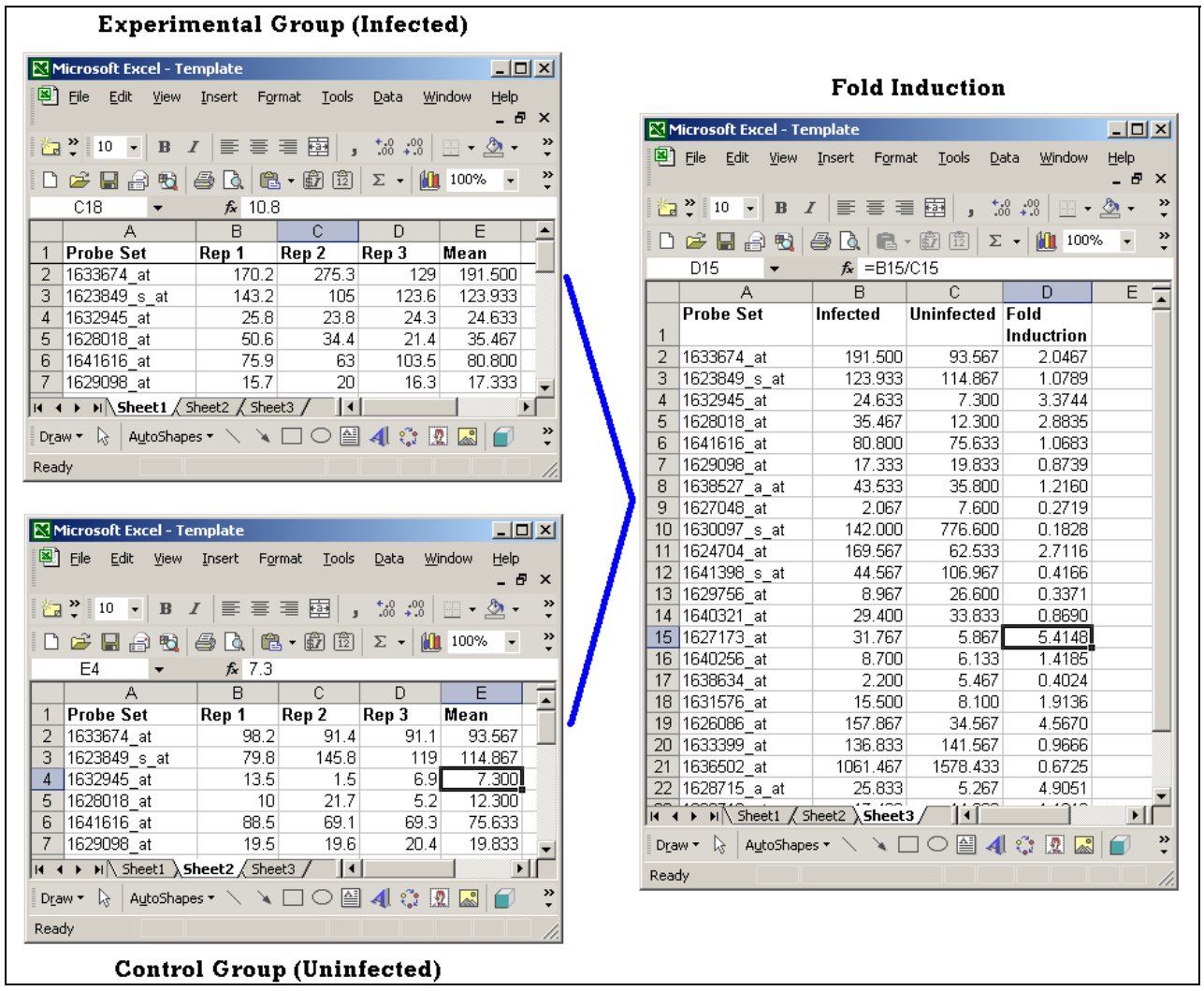
The easiest way is to create a template in MS Excel (or use the one provided by PADMA on the website under "Resources") and perform a vlookup against the replicate files provided by Affymetrix or other service providers by Probe Set ID.

For instructions on how to perform a vlookup, please refer to "VLOOKUP in MS Excel" on the website under "Resources"). This must be done for both the Experimental and Control Groups.

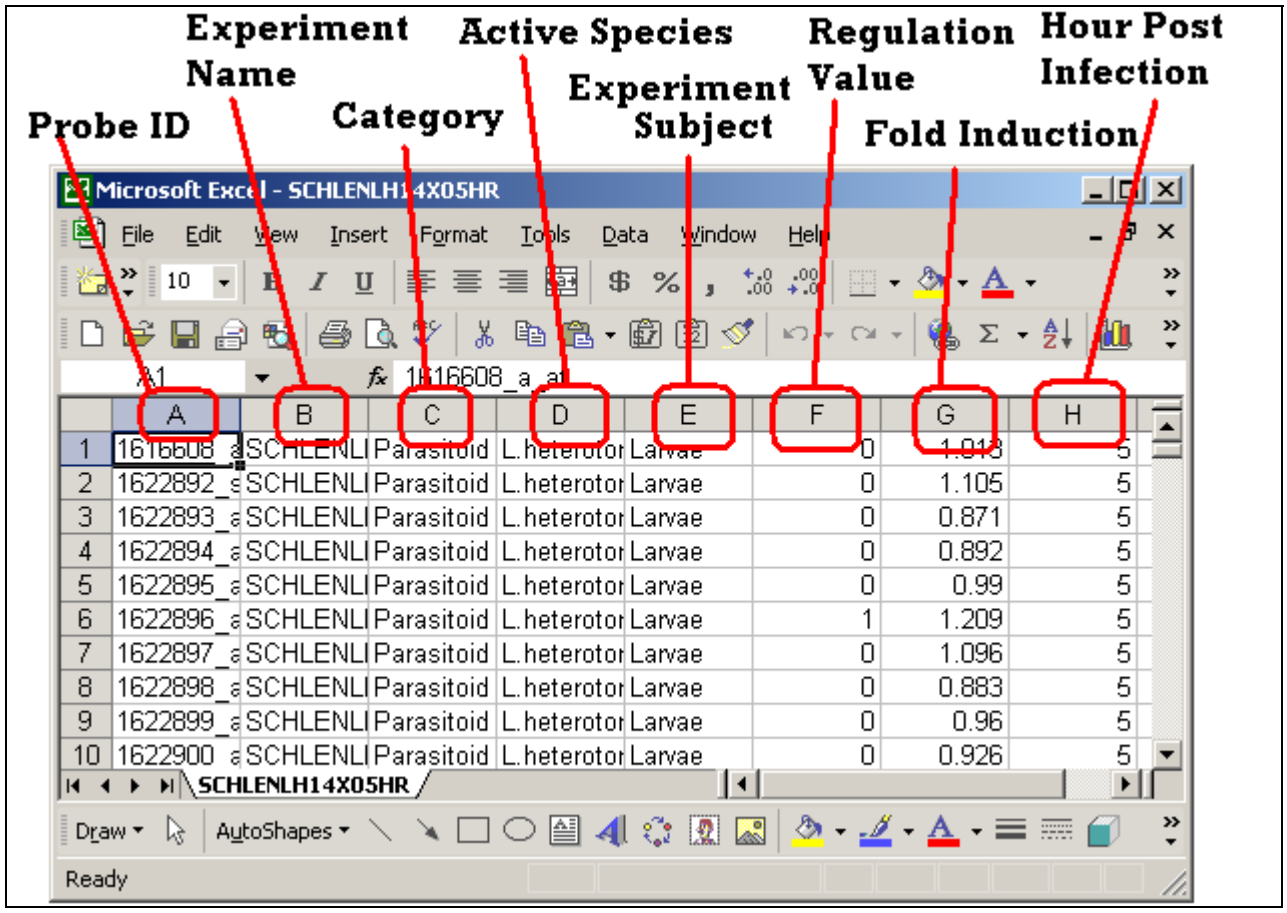
Blank Template with Probe Set ID **After v-lookup and Averaging Signal Value**



Step 2. Calculating Fold Induction. Once the User has generated the Average Signal Value files for both Experimental and Control Groups, the next step is to generate the Fold Induction file. This is done by dividing Experimental/Control.



Step 3. Adding Specific Information. Each experiment has specific criteria that need to be uploaded onto PADMA Data Warehouse for proper query function. **Please note that the order of these criteria (variables) is crucial.** Inconsistent order will result in failing the validation process during upload.



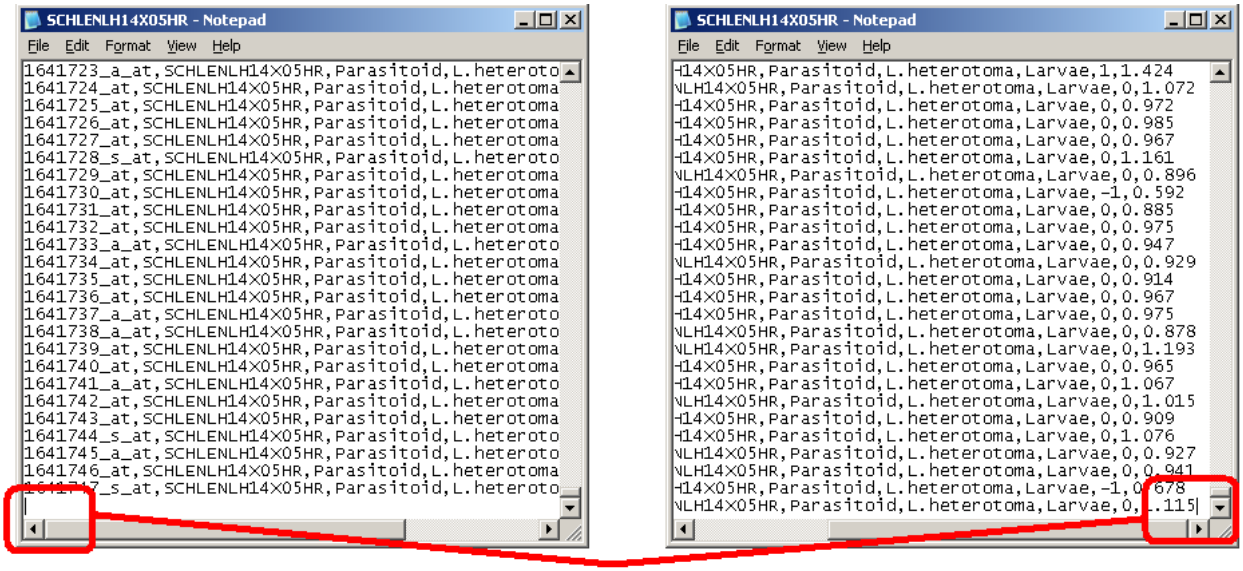
A	B	C	D	E	F	G	H
1	1616608_a	SCHLENLI	Parasitoid	L.heterotor	Larvae	0	1.013	5
2	1622892_s	SCHLENLI	Parasitoid	L.heterotor	Larvae	0	1.105	5
3	1622893_a	SCHLENLI	Parasitoid	L.heterotor	Larvae	0	0.871	5
4	1622894_a	SCHLENLI	Parasitoid	L.heterotor	Larvae	0	0.892	5
5	1622895_a	SCHLENLI	Parasitoid	L.heterotor	Larvae	0	0.99	5
6	1622896_a	SCHLENLI	Parasitoid	L.heterotor	Larvae	1	1.209	5
7	1622897_a	SCHLENLI	Parasitoid	L.heterotor	Larvae	0	1.096	5
8	1622898_a	SCHLENLI	Parasitoid	L.heterotor	Larvae	0	0.883	5
9	1622899_a	SCHLENLI	Parasitoid	L.heterotor	Larvae	0	0.96	5
10	1622900_a	SCHLENLI	Parasitoid	L.heterotor	Larvae	0	0.926	5

Column	Description
Probe ID	Contained in the raw file provided by Affymetrix or other provider
Experiment Name	Any name you want to give to your dataset for easy identification; to follow PADMA nomenclature, please refer to **Section 2.3** above
Category	Indicates whether the experimental group is mutant or infection (if infection, indicate what type: parasitoid, microbial)
Active Species	Indicates pathogen or mutant genotype of the experimental group
Experiment Subject	Indicates where the RNA was extracted from: larvae, adult, or specific tissue (i.e. hemocytes, gut, wing discs, etc…)
Reg Value	Arbitrary calculated value that indicates if a Probe Set is up, down, or no change; for PADMA cut-off and reasoning, please refer to **Section 2.3.** You can put "0" for values in this column, or come put your own thresholds.
Fold Induction	Calculate differential induction of the average Experiment Group signal value over average Control Group signal value
Hour	If infection, indicate time after infection in hour; if not an infection, then put XX in the fields

Step 4. Saving the file. Once you updated your template by calculating the Fold Induction and filling-in all the criteria (i.e. Category, Experiment Name, etc.), you can save the file as csv format from Excel. To do that, simply go to "File" and select "Save As." The system will give you various formats to save the file as. Choose csv (Comma Separated Value).

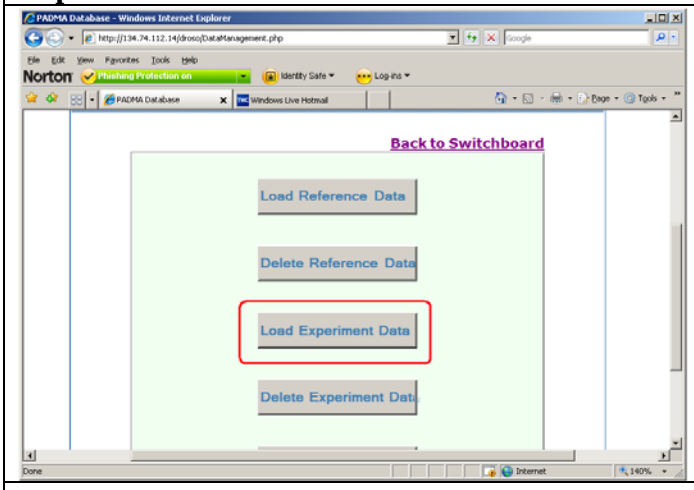
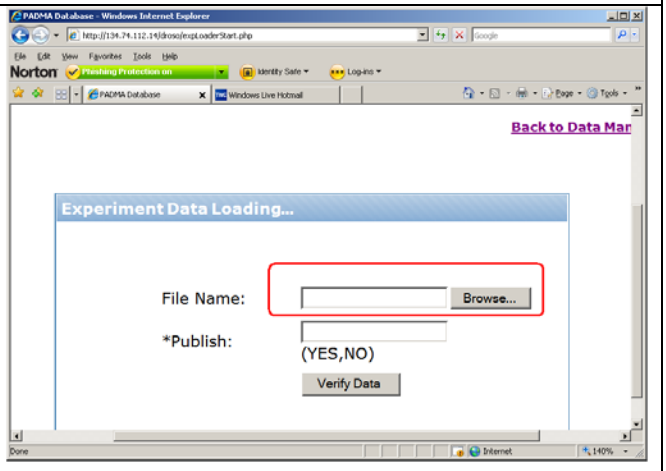
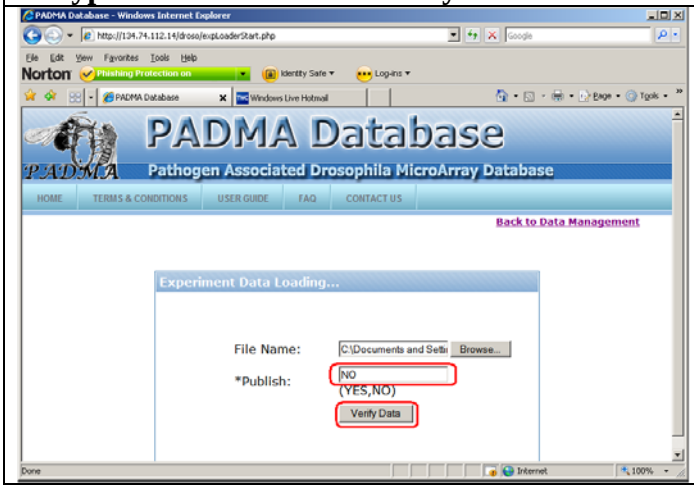
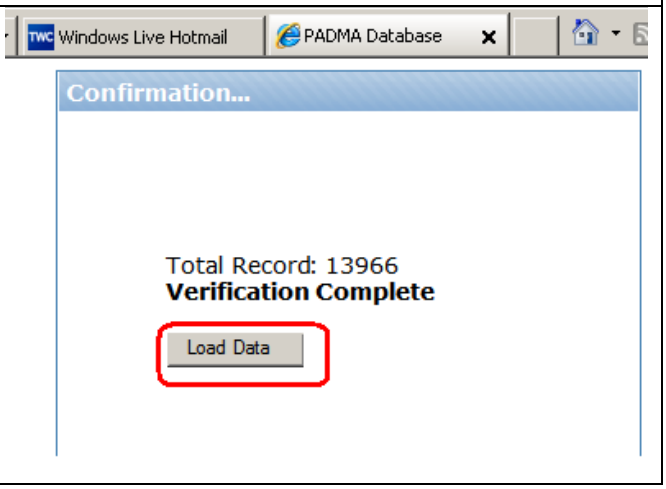
One item to note is, once the User prepares a csv file that conforms to all PADMA Format requirements, the User has to open the file in a text editor like "Notepad" and delete the last, hanging empty line. Because of software incompatibility, some operating system may have a service pack that allows a hanging empty space at the end of a csv file while others don't.

In order to ensure that all files conform to the same standards, during the verification process, PADMA will specifically reject any User load file that has a hanging space. **So please, delete the last hanging, empty space in the csv file by opening it in a text editor like "Notepad".**



Delete the last hanging space on the CSV file by opening it in a text editor like "Notepad"

Step 4. Uploading.

1. Go to Data Management and select Load Experiment Data	2. Hit "Browse" and locate your file
	
3. Type in "NO" and hit "Verify Data"	**4. Select "Load Data"**
	

The Publish field is reserved for datasets that are bound to be uploaded and made available to the public. Since User Uploads involved unpublished data, you must type in "NO" to preserve the confidential nature of your data. By default, User level access is restricted to upload data confidentially, thus it will never publish User data for public access.

Once you press Load Data, it will take a minute or two to load. If successful, you will receive that message that your experiment was successfully inserted.

Section 6.0 – Copyright Guidelines

PADMA integrates microarray data of publication supplemental from various sources. These datasets belong to the respective owners (authors, publishers, etc…) and PADMA does not claim ownership for the data contents. Therefore, in addition to acknowledging PADMA, we advise our users to reference/provide citation to any publication in which the user intends to present, reference, or significantly derive an aspect of his/her research from PADMA query results.

For example, you ran a query using PADMA on Gene XYZ for microbial infection and obtained microarray results from six different publications (A through F). Your interest is in publications A and C only, and intend to run RT-PCR based on microarray results from A and C. In this case, you should cite publications A and C, as well as, PADMA.

Please refer to the Terms of Use for specific terms & conditions regarding copyright and ownership on PADMA website under "Documents".

Section 6.1 – Referencing PADMA

Kawaguchi, A., Mondal, A., Montesdeoca, N., Govind, S., Lee, M.J. (2009) PADMA Database: Pathogen Associated Drosophila MicroArray Database. In *The International Conference on Computing in Engineering, Science and Informatics (ICC2009), Fullerton, California, April 2-4, 2009.*

Section 6.2 – Referencing Other Authors

Please follow the same guidelines you would use to provide citation of other publications you intend to reference or cite.