



## User Guide



support@distributedbio.com

What is AbGenesis? .....	3
Browser Information.....	3
Important Concepts in the System .....	3
General Workflow.....	5
Step 1 – Login in to the system .....	5
<i>Navigation Bar</i> .....	6
Step 2 – Create a Project .....	7
Step 3 - Create an Experiment.....	10
Step 4 - Review Data .....	15
<i>Project Summary Page</i> .....	15
Analysis Apps.....	17
<i>Blast</i> .....	18
<i>Developability</i> .....	19
<i>Fabalign</i> .....	20
Appendix A - Supported Sequence File Formats.....	21
Appendix B – Change Your Password .....	22
Appendix C - File Types.....	24

## What is AbGenesis?

AbGenesis is an analysis platform that allows scientists to manage, analyze and share results from antibody sequencing experiments. AbGenesis provides scientists access to unique Distributed Bio algorithms that will accelerate your antibody research. It is a lightweight highly customizable application that is tailored to your scientists' specific workflows and analysis needs.

abgenesis.distributedbio.org shows a snapshot of how the system could be deployed. For a full demo or discussion of how you could use AbGenesis contact [info@distributedbio.com](mailto:info@distributedbio.com)

## Browser Information

To use AbGenesis you will need a web browser. We recommend either Firefox or Chrome. Both are free and are simple to install on most operating systems. Get Firefox at <http://www.mozilla.org> and Chrome at <http://www.google.com/chrome>

The latest version of Internet Explorer will work with the system, but some data rich pages will load more slowly.

Cookies must be enabled. To enable cookies follow the instructions here:

Firefox	<a href="http://support.mozilla.org/en-US/kb/enable-and-disable-cookies-website-preferences">http://support.mozilla.org/en-US/kb/enable-and-disable-cookies-website-preferences</a>
Chrome	<a href="http://support.google.com/accounts/bin/answer.py?hl=en&amp;answer=61416">http://support.google.com/accounts/bin/answer.py?hl=en&amp;answer=61416</a>

To use the Jalview alignment viewer that is embedded in the system Java must be enabled.

## Important Concepts in the System

User	an individual user of the system, can upload sequences and browse results
Group	a group of users who can all see the same Projects and Experiments, typically all users from the same company are in the same group
Project	a container for multiple experiments that relate to the same goal
Experiment	a set of sequences to be analyzed

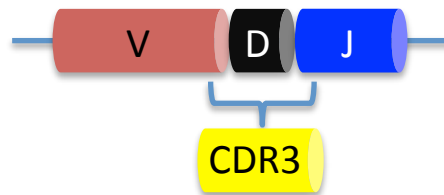
## The Relationship Between Sequences, Clones and Lineages

You give us DNA sequences:

>IgG1

TGACGTGGGAAGACCGATGGGCCCTTGGTGGAGGCT.....

where possible for each sequence we identify a V,D & J gene, CDRs1, 2 & 3



IGHV4-4,IGHD3-22,IGHJ4, GSITNYCWS, GRIYPSGYTNYS, CARVLYDSCGYHFDYW

**CLONE** - we identify a clone as the V & J genes and the CDR3



IGHV4-4\_IGHJ4\_CARVLYDSCGYHFDYW

**Representative Sequence** – if multiple sequences can have the same V/J CDR3 combination we assign 1 sequence to represent this clone. The Representative Sequence will be used in Developability analysis and will be shown in all reports.

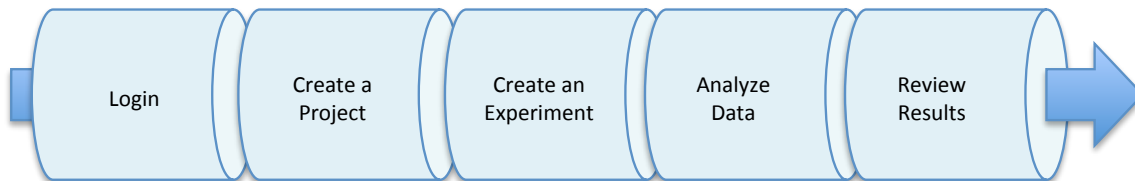
**Lineages** are groups of clones that have the same V & J, and CDR3s that differ by a max of 2 amino acids. A lineage is identified by the first clone in the group, e.g. our clone from above can be found in the lineage below

IGHV4-4\_IGHJ4\_CARVLYDS**R**GYHFDYW

IGHV4-4\_IGHJ4\_CARVLYDS**G**GYHFDYW

## General Workflow

The overall workflow is as follows:

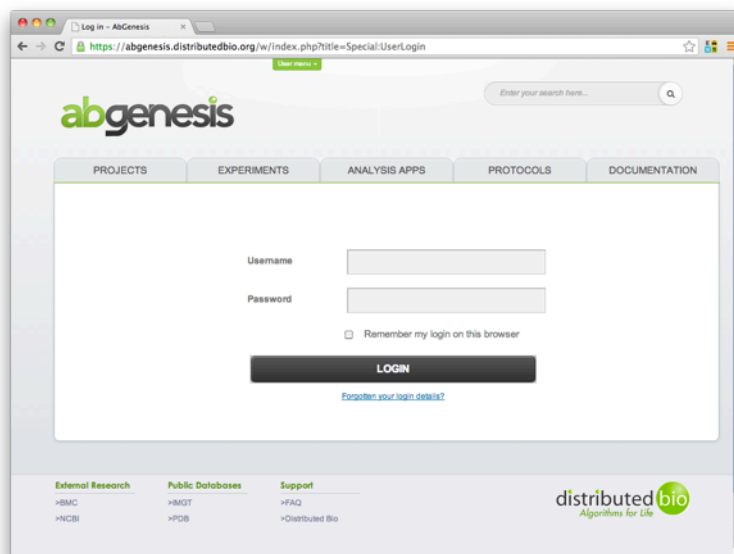


Once you have created a Project you can of course create as many Experiments as you want in it. You should remember that it is best to create Projects where the Experiments or pertain to the same goal. That way some of the analysis AbGenesis presents will be more useful.

## Step 1 – Login in to the system

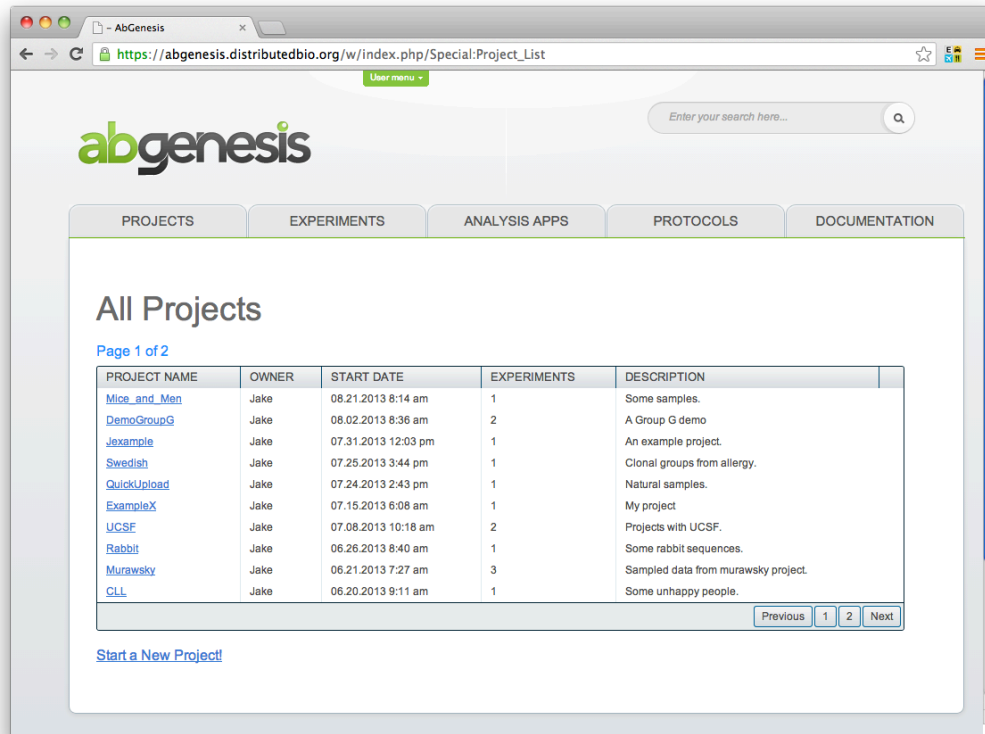
Once you have a working web browser type <http://abgenesis.distributedbio.org> into the URL box.

You will briefly see the landing page and then this page will load



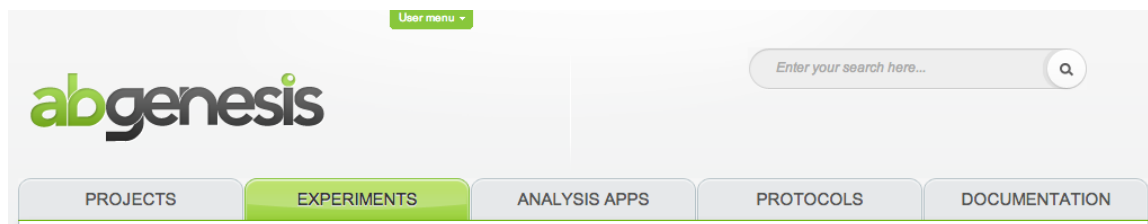
Fill in the username and password you have been provided by Distributed Bio and press the Log In button. To get an account on the system contact your Distributed Bio representative.

When you have logged in this the Projects Home Page will load (below). If no one in your group has created a Project before you will not see any Projects listed.



## Navigation Bar

At the top of all AbGenesis pages is the Navigation Bar. It contains useful links that will help you rapidly navigate the system.



PROJECTS	A list of the Projects your Group have created
EXPERIMENTS	A list of Experiments from all Projects your Group has created
ANALYSIS APPS	A page where you can launch individual analysis runs
PROTOCOLS	A page where you can store information on Protocols your Group uses
DOCUMENTATION	Useful documentation

## Step 2 – Create a Project

The Projects Home Page shown below shows the Project Name, the Owner and Start Date, Number of Experiments in the Project and the Description. At any time if you press the Projects navigation bar link you will come back here.

The screenshot shows the AbGenesis web application interface. The browser address bar displays [https://abgenesis.distributedbio.org/w/index.php/Special:Project\\_List](https://abgenesis.distributedbio.org/w/index.php/Special:Project_List). The page features a navigation bar with the following links: PROJECTS, EXPERIMENTS, ANALYSIS APPS, PROTOCOLS, and DOCUMENTATION. Below the navigation bar, the title 'All Projects' is displayed, followed by 'Page 1 of 2'. A table lists the projects with the following columns: PROJECT NAME, OWNER, START DATE, EXPERIMENTS, and DESCRIPTION. The table contains 10 rows of project data. At the bottom of the table, there is a 'Start a New Project' link.

PROJECT NAME	OWNER	START DATE	EXPERIMENTS	DESCRIPTION
<a href="#">Mice_and_Men</a>	Jake	08.21.2013 8:14 am	1	Some samples.
<a href="#">DemoGroupG</a>	Jake	08.02.2013 8:36 am	2	A Group G demo
<a href="#">Jexample</a>	Jake	07.31.2013 12:03 pm	1	An example project.
<a href="#">Swedish</a>	Jake	07.25.2013 3:44 pm	1	Clonal groups from allergy.
<a href="#">QuickUpload</a>	Jake	07.24.2013 2:43 pm	1	Natural samples.
<a href="#">ExampleX</a>	Jake	07.15.2013 6:08 am	1	My project
<a href="#">UCSF</a>	Jake	07.08.2013 10:18 am	2	Projects with UCSF.
<a href="#">Rabbit</a>	Jake	06.26.2013 8:40 am	1	Some rabbit sequences.
<a href="#">Murawsky</a>	Jake	06.21.2013 7:27 am	3	Sampled data from murawsky project.
<a href="#">CLL</a>	Jake	06.20.2013 9:11 am	1	Some unhappy people.

[Start a New Project!](#)

To create a new Project press the “Start a New Project” link at the bottom of the table.

This page will load:

The screenshot shows a web browser window with the address bar displaying `https://abgenesis.distributedbio.org/w/index.php/Special:AddForm?action=AddProject`. The page features the 'abgenesis' logo and a navigation menu with tabs for PROJECTS, EXPERIMENTS, ANALYSIS APPS, PROTOCOLS, and DOCUMENTATION. A search bar is located in the top right corner. The main content area is titled 'Projects->Add Project' and contains a form with two required fields: 'Project Name:\*' and 'Description:\*'. The 'Project Name' field has a text input box and a note: 'No symbols like (\$\*?) please!'. The 'Description' field has a larger text area. Below the form is a blue 'Add Project!' button. At the bottom of the form, there is a note: 'Alphanumeric input: can be alphanumeric, period, spaces, underscore, and dash' and 'All fields with \* are Required!'.

[Projects->Add Project](#)

Project Name:\*   
No symbols like (\$\*?) please!

Description:\*

[Add Project!](#)

Alphanumeric input: can be alphanumeric, period, spaces, underscore, and dash  
All fields with \* are Required!

Fill in the form and press the "Add Project" button.





support@distributedbio.com

This page will now load showing a summary of what you just entered. The Project number (11 below) is automatically created by the system and is the unique identifier for this Project.

The screenshot shows a web browser window with the address bar displaying `https://abgenesis.distributedbio.org/w/index.php/Project_47`. The page features the 'abgenesis' logo and a navigation menu with tabs for PROJECTS, EXPERIMENTS, ANALYSIS APPS, PROTOCOLS, and DOCUMENTATION. A search bar is located in the top right corner.

## Project 47

[Projects](#) > [Project\\_47](#)

Contents [\[hide\]](#)

- [1 Web\\_Demo](#)
  - [1.1 Experiments](#)
  - [1.2 Analytical Distributions All Experiments](#)
- [2 Comments](#)

### Web\_Demo

- Project ID: 47
- Owner: Giles
- Started: 10.02.2013 12:57 pm

Description: a test project

### Experiments

PROJECT	EXPERIMENT ID	EXPERIMENT NAME	DESCRIPTION	EXPERIMENT DATE	QC STATUS

[Add an Experiment!](#)

Navigation buttons: Previous, Next

<http://abgenesis.distributedbio.org>

### Step 3 - Create an Experiment

Click the “Add an Experiment” button at the bottom of the Experiments Table. This page will load:

AbGenesis

https://abgenesis.distributedbio.org/w/index.php/Special:AddForm?action=AddExperiment&project=47

PROJECTS EXPERIMENTS ANALYSIS APPS PROTOCOLS DOCUMENTATION

[Experiments->Add Experiment](#)

Project:\* Project 47: Web\_Demo

Experiment Name:\*

Experiment Description:\*

Run Lineages? No

Run Developability? none

Maximum Lineages to Display 5

Maximum Clones to Display 5

Upload Sequences:\* Please Upload Sequences

Submit Disabled: Please upload first!

Enter an Experiment Name and Description.

#### Options

Run Lineages	Calculate clone trees based VDJ segments
Run Developability	Calculate bio-physical characteristics that make preferred therapeutic antibodies
Maximum Lineages to Display	If you are running 1000s of sequences you can't display them all in the report – limit the display to this
Maximum Clones to Display	If you are running 1000s of sequences you can't display them all in the report – limit the display to this

Next you need to add some sequence files. You can either drag and drop these onto the “Please Upload Sequences” button, or press the button to activate a file chooser. See Appendix A for acceptable file formats to upload. The Submit button will not be activated until you have chosen some files.

When the files have uploaded a box like this will appear for every file you uploaded:

Uploaded Lane:

B6-IgM.fa

MID:	Species:	Isotype:	Sample Name:	Phenotype:	Time Point:
none	Human	IgM	B6-IgM.fa	Immunized	0

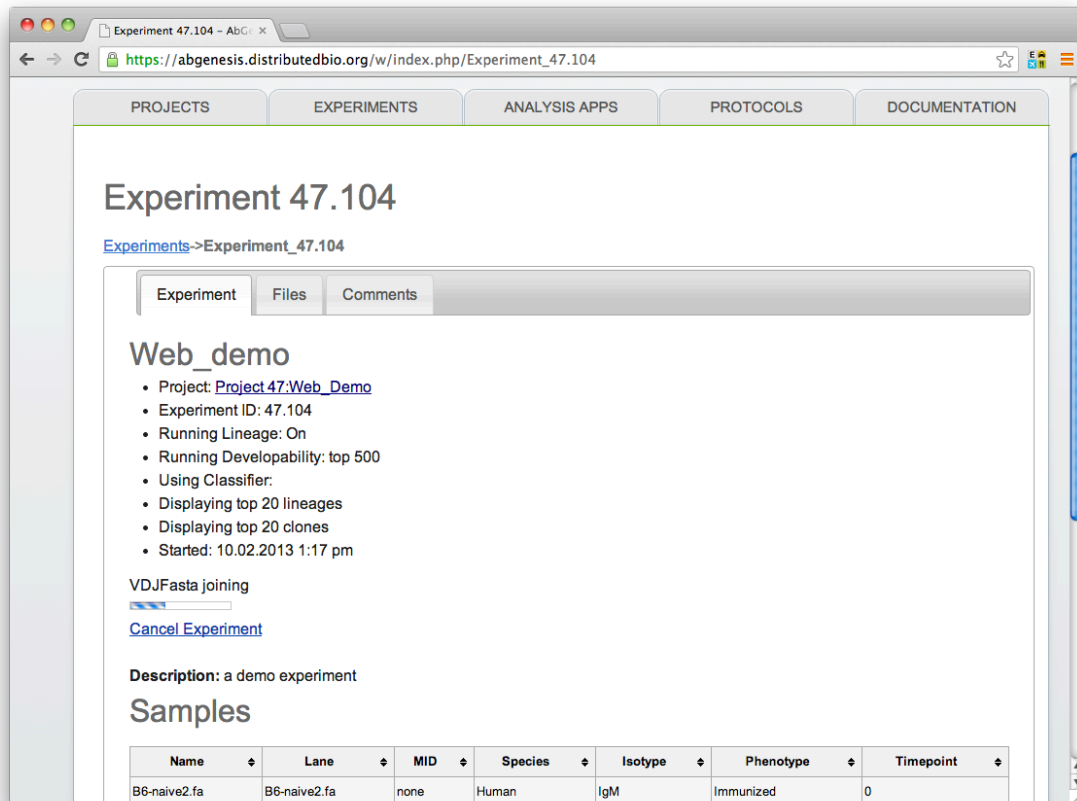
Add New MID

If you uploaded multiple files there will be multiple boxes stacked on top of each other. These boxes allow you to MID and Phenotype details for the sequences contained in that file

Once you have finished adding the extra detail press the “Lets Go” button at the bottom of the screen.

The system will now start analyzing your experiments. This may take a few minutes depending on the number of sequences you uploaded.

You should see a screen that looks like this:



The screenshot shows a web browser window with the URL [https://abgenesis.distributedbio.org/w/index.php/Experiment\\_47.104](https://abgenesis.distributedbio.org/w/index.php/Experiment_47.104). The page has a navigation bar with tabs: PROJECTS, EXPERIMENTS, ANALYSIS APPS, PROTOCOLS, and DOCUMENTATION. The main heading is "Experiment 47.104". Below it is a breadcrumb link "Experiments->Experiment\_47.104". There are three sub-tabs: "Experiment", "Files", and "Comments". The "Experiment" tab is active, showing a "Web\_demo" section with a list of details: Project: [Project 47:Web\\_Demo](#), Experiment ID: 47.104, Running Lineage: On, Running Developability: top 500, Using Classifier, Displaying top 20 lineages, Displaying top 20 clones, and Started: 10.02.2013 1:17 pm. Below this is a "VDJFasta joining" section with a progress bar and a "Cancel Experiment" link. The "Description" is "a demo experiment". The "Samples" section contains a table with the following data:

Name	Lane	MID	Species	Isotype	Phenotype	Timepoint
B6-naive2.fa	B6-naive2.fa	none	Human	IgM	Immunized	0

The Progress Bar seen here in the middle of the screen will update you on the progress of your run. You can leave this page without affecting the completion of your analysis. You can return to the results by clicking through to this Experiment from the Project Home page.

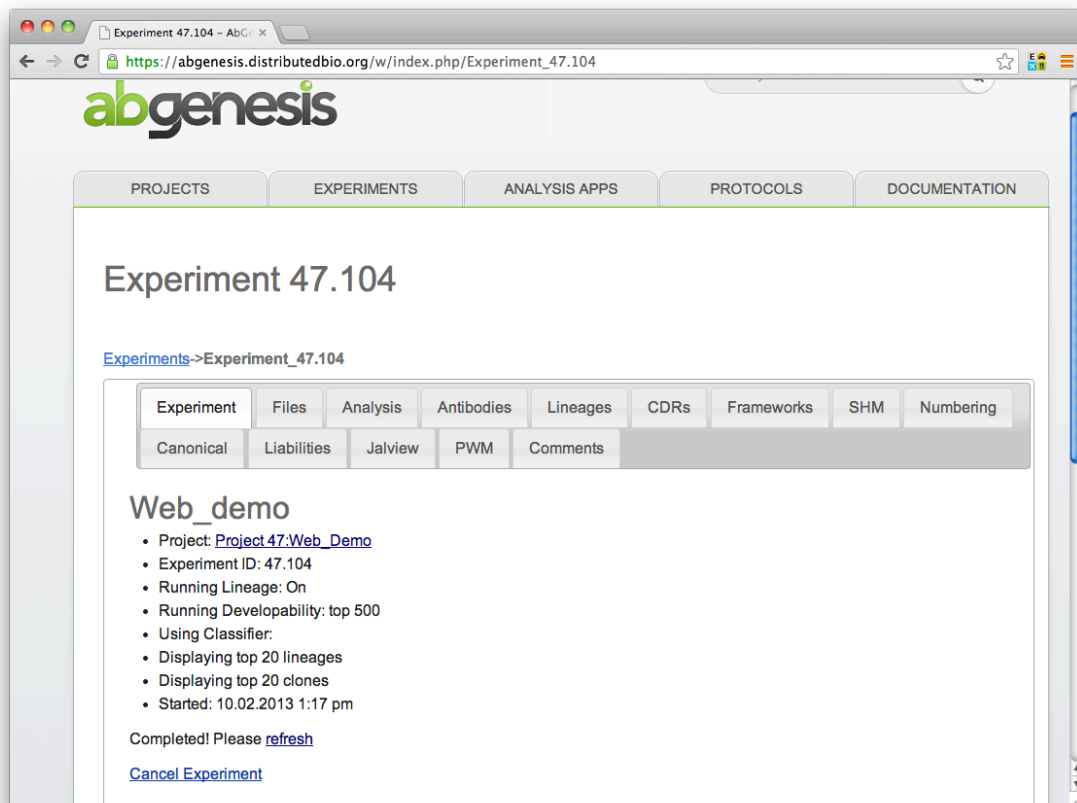
When your analysis has complete you will see a screen that looks like this:

The screenshot shows a web browser window with the URL [https://abgenesis.distributedbio.org/w/index.php/Experiment\\_47.104](https://abgenesis.distributedbio.org/w/index.php/Experiment_47.104). The page has a navigation bar with tabs: PROJECTS, EXPERIMENTS, ANALYSIS APPS, PROTOCOLS, and DOCUMENTATION. The main heading is "Experiment 47.104". Below it is a breadcrumb link "Experiments->Experiment\_47.104". There are three sub-tabs: "Experiment", "Files", and "Comments". The "Experiment" tab is active, showing a "Web\_demo" section with a bulleted list of details: Project: [Project 47: Web\\_Demo](#), Experiment ID: 47.104, Running Lineage: On, Running Developability: top 500, Using Classifier, Displaying top 20 lineages, Displaying top 20 clones, and Started: 10.02.2013 1:17 pm. Below this list, it says "Completed! Please [refresh](#)" and "[Cancel Experiment](#)". A "Description" section follows, stating "a demo experiment". At the bottom, there is a "Samples" section with a table.

Name	Lane	MID	Species	Isotype	Phenotype	Timepoint
B6-naive2.fa	B6-naive2.fa	none	Human	IgM	Immunized	0

The Progress Bar will have changed into a “refresh” link. Click on this to view to your results.

The screen should change to something that looks like this:



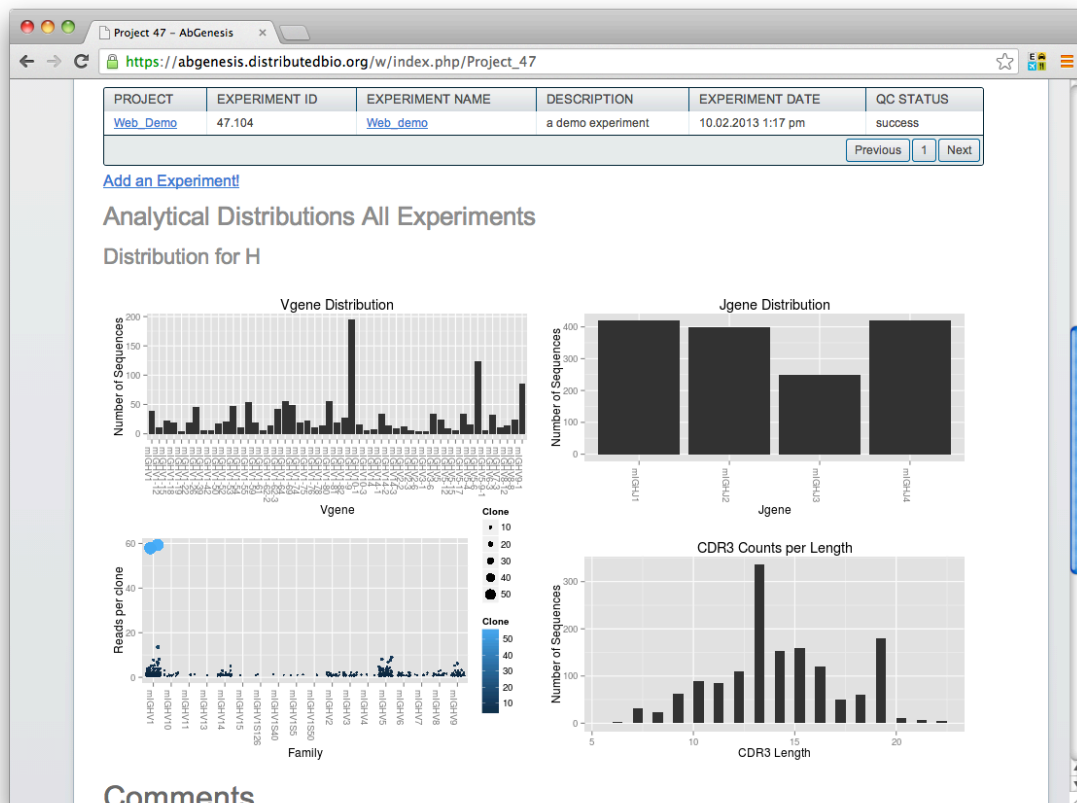
You will now be able to review your results by clicking through the tabs displayed. See Step 4 for details of what all the data is on each tab.

## Step 4 - Review Data

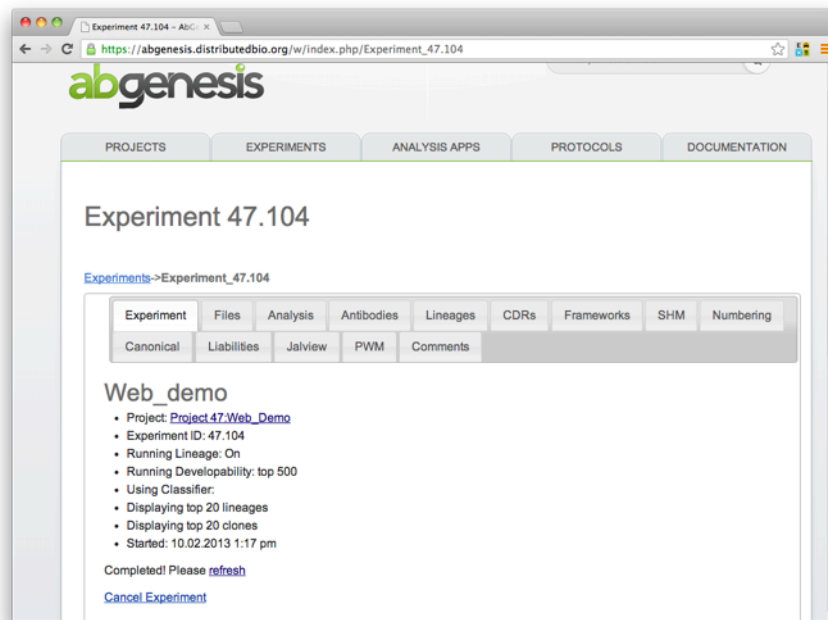
To see the results of the Experiment either click the refresh link on Experiments page that was generated or click through to the Experiment from the Project Summary page.

### Project Summary Page

As shown below will give some summary statistics collated from all Experiments in a Project.



Click on the Experiment Name for the Experiment you are interested in to load the Experiment page as shown below:



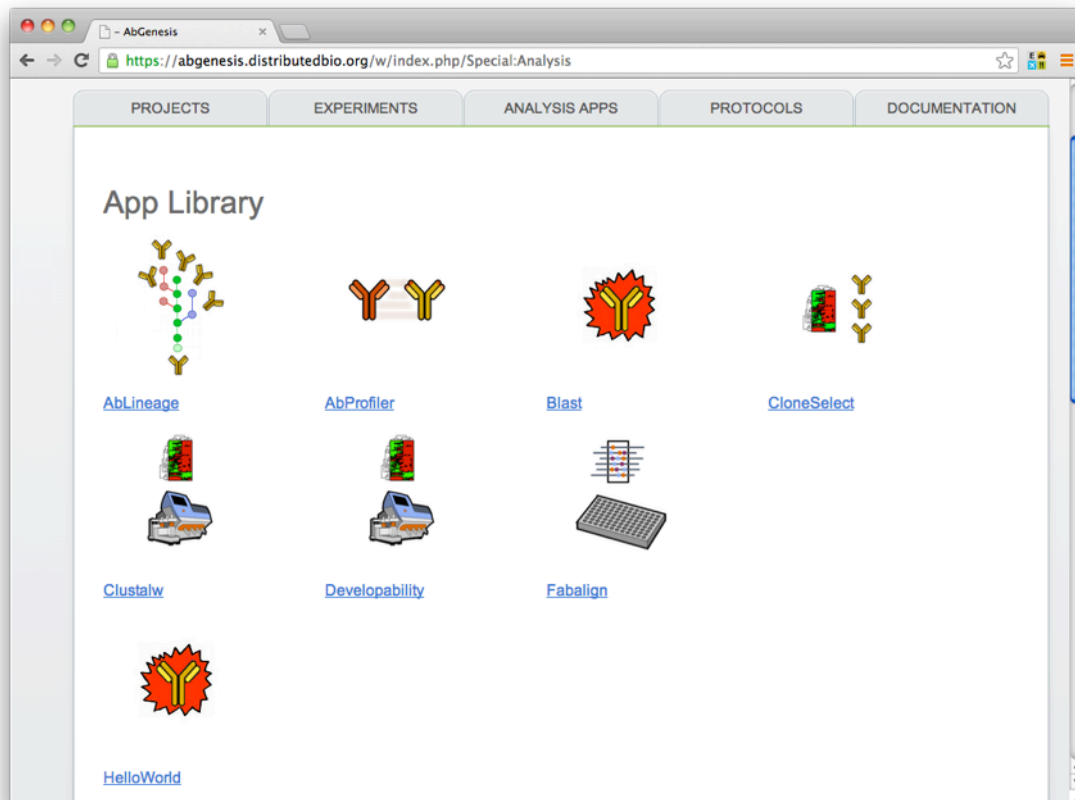
The Experiment page contains the results of the analysis pipeline for the sequences in that file as described below:

Tab	Description
Experiment	Summary of the experiment options chosen
Files	A list of the files uploaded and files generated by the analysis
Analysis	Summary of results across the sequences including: read length; depth; orientation; on-target QC; cross-sample clones & SHM Profiles grouped by input file
Antibodies	Analysis of each sequence for V&J gene distribution
Lineages	Sequences with the same V,J and same composition of CDR3 (allowed to be up to 2 residues different) are the same Lineage. Click on the Lineage to see the sequences that contain it.
CDRs	CDR sequences for each sequence
Frameworks	List of matched frameworks
SHM	Somatic Hypermutations found in the sequences
Numbering	Kabat numbering for each sequence – click on each sequence to see the numbering for that sequence
Canonical	The MLT canonical classification of the CDRs. CDR-3 labeled as kinked or un-kinked depending on whether the salt-bridge is intact.
Liabilities	Biochemical liabilities for the sequence: acid hydrolysis sites; splice sites; cryptic splice sites; N-linked glycosylation sites; methionines.
Jalview	Java applet aligning all the sequences in an interactive window
PWM	Positional Weight Matrices for CDRs in VH, VK & VL
Comments	A tab for Users to add comments on the Experiment



## Analysis Apps

There may be several analysis tools deployed on your system that you can use to run an analysis that you do not want to add to a Project or Experiment. Click on the ANALYSIS APPS tab from the Navigation Bar to see them.

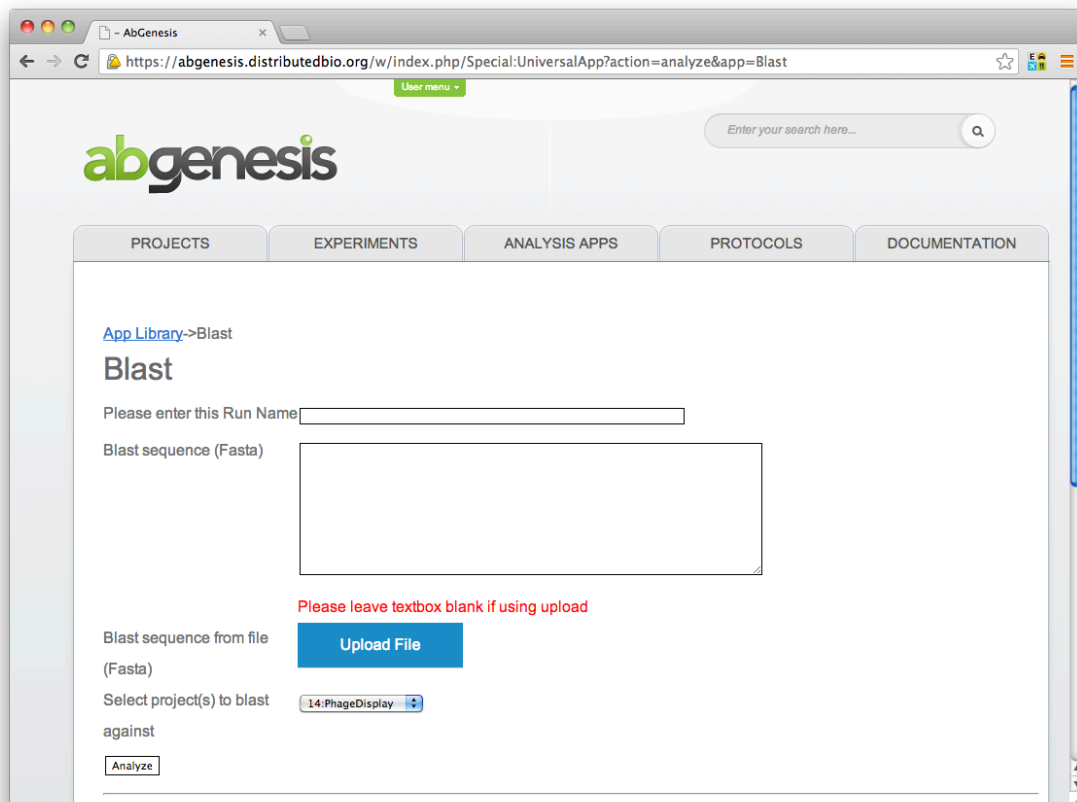


The 4 analysis app icons are shown, plus some information on latest runs and current server activity.

App	Description
Blast	Blast a single sequence against sequences uploaded to the system
Developability	Analyses sequences for various developability features and abnormalities.
Fabalign	Generates a Hidden Markov Model from input fasta amino acid sequences with cdr kabat-restacking.
Humanize	Find the best humanization scaffolds

## Blast

This allows you to perform a single Blast search against any sequence from a Project. Give the run a name and paste in a Fasta (see Appendix A) sequence into the Query box. You can choose an individual Project from the drop down menu, or select All to Blast all sequences.



The screenshot shows a web browser window with the URL <https://abgenesis.distributedbio.org/w/index.php/Special:UniversalApp?action=analyze&app=Blast>. The page features the 'abgenesis' logo and a navigation bar with tabs: PROJECTS, EXPERIMENTS, ANALYSIS APPS, PROTOCOLS, and DOCUMENTATION. A search bar is located in the top right corner. The main content area is titled 'Blast' and includes a link 'App Library->Blast'. Below the title, there is a form with the following elements:

- A text input field for 'Please enter this Run Name'.
- A large text area for 'Blast sequence (Fasta)'.
- A red text warning: 'Please leave textbox blank if using upload'.
- A section for 'Blast sequence from file (Fasta)' with an 'Upload File' button.
- A dropdown menu for 'Select project(s) to blast against', currently showing '14:PhageDisplay'.
- An 'Analyze' button at the bottom.

When you have picked a database click Analyze.

## Developability

This will generate most of the features generated for an Experiment for a group of sequences you do not want to add to a Project or Experiment. Give a name for you run and paste in 1 or more fasta (see Appendix A) nucleotide sequences. Click Analyze.

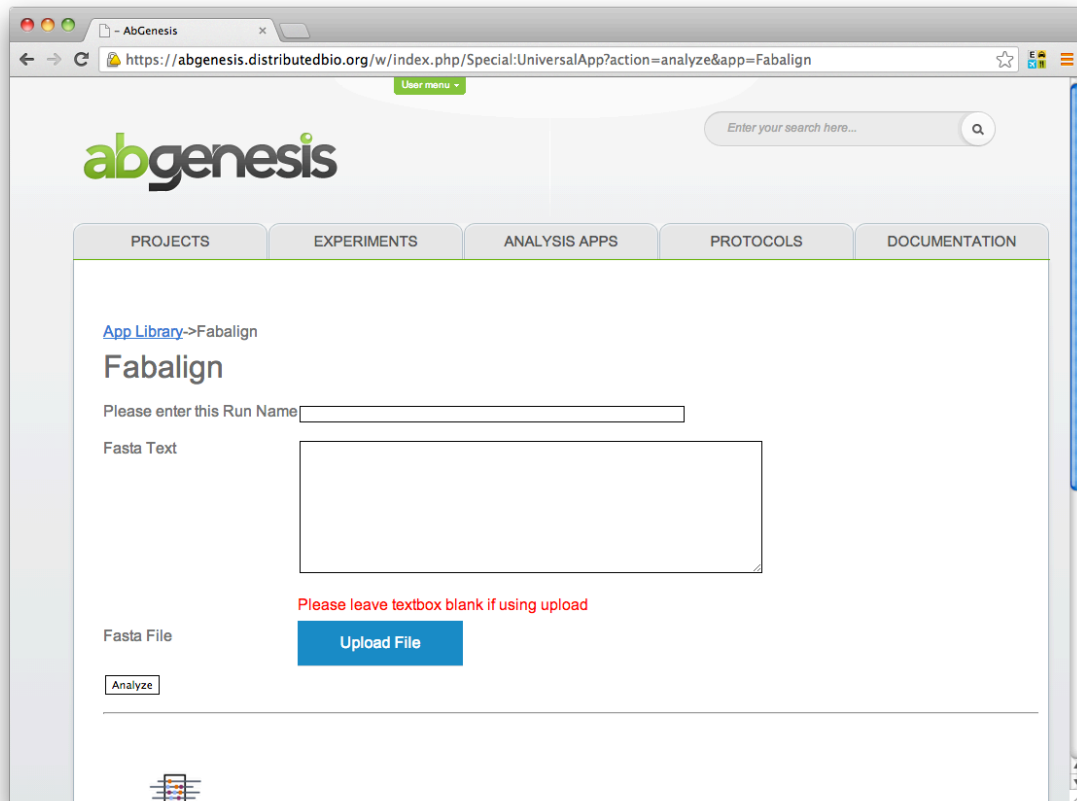
The screenshot shows a web browser window with the URL `https://abgenesis.distributedbio.org/w/index.php/Special:UniversalApp?action=analyze&app=Developability`. The page features the 'abgenesis' logo and a navigation bar with tabs for PROJECTS, EXPERIMENTS, ANALYSIS APPS, PROTOCOLS, and DOCUMENTATION. A search bar is located in the top right corner. The main content area is titled 'Developability' and includes a link to 'App Library->Developability'. Below the title, there is a form with the following elements:

- A text input field labeled 'Please enter this Run Name'.
- A text area labeled 'Fasta Text'.
- A red text instruction: 'Please leave textbox blank if using upload'.
- A label 'Fasta File' next to a blue 'Upload File' button.
- An 'Analyze' button.

A small 3D molecular model is visible in the bottom left corner of the page.

**Fabalign**

Given the amino acid sequence of an antibody heavy or light chain will generate a multiple sequence alignment with CDR Kabat re-stacking. Give your run a name and paste in a fasta query sequence and click Analyze.



The screenshot shows a web browser window with the address bar displaying `https://abgenesis.distributedbio.org/w/index.php/Special:UniversalApp?action=analyze&app=Fabalign`. The page features the 'abgenesis' logo and a navigation menu with tabs for PROJECTS, EXPERIMENTS, ANALYSIS APPS, PROTOCOLS, and DOCUMENTATION. A search bar is located in the top right corner. The main content area is titled 'Fabalign' and includes a link to 'App Library->Fabalign'. Below the title, there is a form with the following elements:

- A text input field labeled 'Please enter this Run Name'.
- A text area labeled 'Fasta Text'.
- A red text instruction: 'Please leave textbox blank if using upload'.
- A label 'Fasta File' next to a blue 'Upload File' button.
- An 'Analyze' button.

A small icon of a document with a red 'X' is visible at the bottom left of the page.

## Appendix A - Supported Sequence File Formats

- Sanger, MiSeq, HiSeq, Ion torrent, 454
- Fasta, fastq, Sanger Zip, fastq.gz
- Allowed file extensions: '.fa', '.fasta', '.fas', '.fna', '.seq', '.zip', '.txt', '.fastq', '.fq', '.gz'

Fasta example:

```
>AF184762
atggagtttgggctgagctgggttttcctgttgctattttaaaagggtgtccactgtgag
gtgcagctggtggagtctggggaggcttagtccagcctgggggatccctgaaactctcc
tgtgcagcctctgggttcaccctcagtggtcaaatgtgcactgggtccgccaggcctcc
gggaaagggtggagtgggttgccgtatcaaaaggaatgctgagtctgacgcgacagca
tatgtctgcgtcgatgagaggcaggctcaccatctccagagatgattcaaagaacacggcg
tttctgcaaatgaacagcctgaaaagcgatgacacggccatgtattattgtgtgatccgg
ggagatgtttacaaccgacagtggggccagggaaccctggtcaccgtctctcagcatcc
ccgaccagcccaaggctcttcccgtgagcctctgcagcaccagccagat
>AF069038
tcctctctccactgcacaaggctctctctccccgggtcatgctgacgcaatcacctcta
tttctgcctccctgggagcctcggtcaacctcacctgcactctgaccagtgggcacagac
gttacgccatcgcatggcatcagcaattgtcagggaaggccctcgtttcttgatgagac
ttaacagtgtggcacttacaccaggggggacgggattcctgatcgcttctccggctcca
cctctgggcctgagcgctacctcaccatctccagcctccagtctgaagatgaggcagatt
attactgtcagacctggggcactggcctttgggttttcggcggagggaccagtctgaccg
tcttaggtcagcccaaggctgccccctcg
```

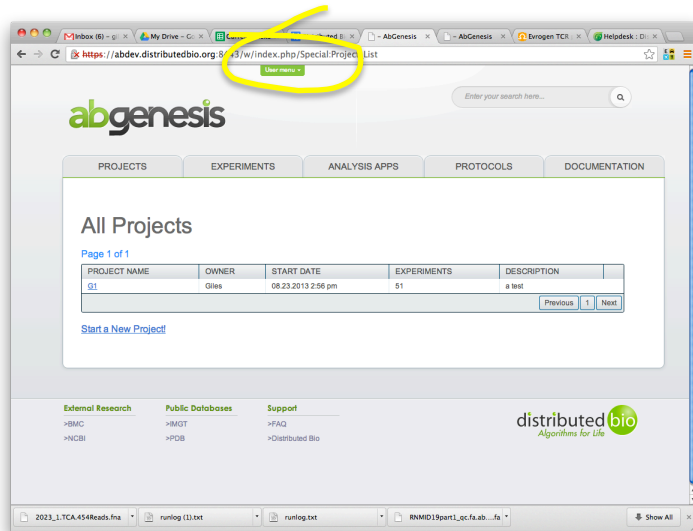
Fastq Example

```
@EAS54_6_R1_2_1_413_324
CCCTTCTTGTCTTCAGCGTTTCTCC
+
;;3;;;;;;;;;;7;;;;;;;;;88
@EAS54_6_R1_2_1_540_792
TTGGCAGGCCAAGGCCGATGGATCA
+
;;;;;;;;;7;;;;-;;3;83
@EAS54_6_R1_2_1_443_348
GTTGCTTCTGGCGTGGGTGGGGGGG
+EAS54_6_R1_2_1_443_348
;;;;;;;;;9;7;;7;393333
```

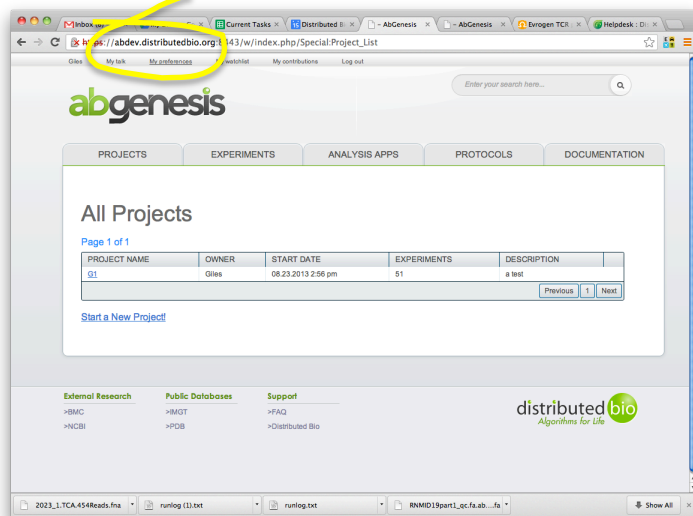
Not Supported: Microsoft Word; Excel.

## Appendix B – Change Your Password

From any screen pull down the User menu from the top of the screen:



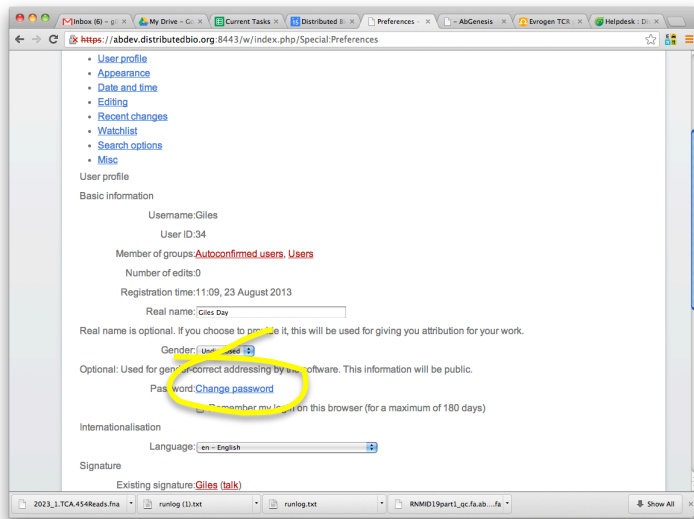
From the options click My preferences



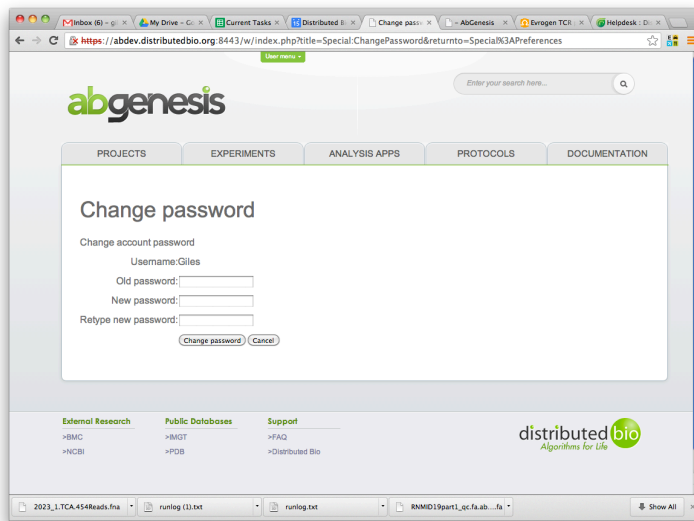


support@distributedbio.com

The Preferences page will load. Scroll down until you Change password, click on that link.



Fill out the form and click the change password button



The next time you log into the system you will need to use this password. If you forget your password contact your system administrator to reset it.

<http://abgenensis.distributedbio.org>

## Appendix C - File Types

file1_qc.fa	Your input file after it has been through the Quality Control process
file1_qc.fa.abgen.aa.VDJ.H3.L3.CH1.dnaH3.a2m	Protein fasta file with extended headers, seq_id and translation frame in position one of header. Aligned protein sequence
file1_qc.fa.abgen.aa.VDJ.H3.L3.CH1.dnaH3.fa	Protein fasta file with extended headers. Same headers as .VDJ.H3.L3.CH1.dnaH3.fa. First field in header is seq_id and translation frame
file1_qc.fa.abgen.developability-annotated.fa	Protein fasta file. Extended headers with developability annotations.
file1_qc.fa.abgen.fa	Original input file with unique identifiers applied
file1_qc.fa.abgen.unclass.fa	All sequences that did not match a V,D or J gene
file1_qc.fa.abgen.VDJ.H3.L3.CH1.dnaH3.fa	Output of VDJFasta. DNA fasta file with extended header with format: >seq_id;vh_field;dh_field;jh_field;h3;l3;isotype;np_boundaries;h1;h2;l1;l2;vl_field;jl_field;dna_h3
file1_qc.fa.abgen.VDJ.H3.L3.CH1.dnaH3.fa.clones.txt	Sequences and clone tab separated for <b>Heavy</b> chain only. All sequences in file and will show NO_HIT if they do not match a Heavy chain clone
file1_qc.fa.abgen-V[H K L].canonical.tab.txt	Tab separated table of every sequence that matched that chain and the CDR canonical classifications
file1_qc.fa.abgen-V[H K L].cdrs.tab.txt	Tab separated table of every sequence that matched that chain and the CDRs
file1_qc.fa.abgen-V[H K L]-clones.fa	DNA fasta file with extended headers, same as .VDJ.H3.L3.CH1.dnaH3.fa. Contains all sequences that match the chain in the filename (VH, VK or VL).
file1_qc.fa.abgen-V[H K L]-clone-table.txt	Summary statistics for each unique clone matching the chain in the file name (VH, VK, VL). Tab separated: Clone frequency, number of matching seqs, Vgene, Dgene, Jgene, Germline %, CDR1, CDR2, CDR3, Matching seq_ids (, separated)
file1_qc.fa.abgen-V[H K L]-clone-table.xls	Full tab separated listing of every sequence and every property calculated by the analysis pipeline. Can be <b>VERY LARGE</b>
file1_qc.fa.abgen-V[H K L].frameworks.tab.txt	Tab separated table of every sequence that matched that chain and the CDRs
file1_qc.fa.abgen-	Tab separated table of every sequence that matched that





support@distributedbio.com

V[H|K|L].liabilities.tab.txt

chain and their liabilities - there is no summary table in this file

file1\_qc.fa.abgen-  
V[H|K|L].numbering.tab.txt

Tab separated table of every sequence that matched that chain and the Kabat Numbering for that sequence

file1\_qc.fa.abgen-  
V[H|K|L].shm.tab.txt

Tab separated table of every sequence that matched that chain and their Somatic Hyper-mutations