1. **What is agriGO?**

The agriGO is designed to automate the job for experimental biologists to identify enriched Gene Ontology (GO) terms in a list of microarray probe sets or gene identifiers (with or without expression information) and it is also a GO-related database. The agriGO specially focus on agricultural species.

2. **Why use agriGO?**

The agriGO provides heavy support to agricultural species. Not only limited to SEA analysis, GSEA which is achieved using PAGE method is also available. Furthermore we have BLAST4ID tool for ID transfer or annotation. And search as well as download function is accessible. The agriGO can give out rich outputs like graphical result, bar chart result and hierical tree which composing a comprehensive understanding of biological meaning of user’s input data.

3. **What data argiGO contains?**

We currently support on 35 species including 280 datatypes. Please check the data statistics page for detail information. We will continue adding more species and datatypes.

4. **How agriGO prepare its data?**

Raw GO annotation data is generated using BLAST, Pfam, InterproScan by agriGO or obtained from B2G-FAR center or from Gene Ontology. Arabidopsis genome data is from TAIR. Rice TIGR genome data is from Rice Genome Annotation Project. Rice KOME data is from KOME database. Rice Gramene data is from Gramene center. Populus genome data is collected from JGI. Soybean and Sorghum genome data is compiled from phytozome. Grape genome data is compiled from Genoscope. Medicago genome data is from Medicago truncatula sequencing resources. Maize genome data is from MaizeSequence.org. Castor bean genome data is from Castor Bean Genome Database. Brachypodium distachyon genome data is from Ensembl. Bovien genome data is from Bovine Genome Database. Silkworm genome data is from SilkDB. M. grisea genome data is from Magnaporthe grisea Database. affymetrixmetrix CSV files and array sequences are from NetAffx.
5. How to use tools provided by agriGO?

Quick introduction to analysis procedure

1. Choose tool and set parameters
You should choose one tool to go forward. At the right side, several frames containing annotation text are interactive. The content will change depending on exact parameters you chose. You can make the help frames show or hidden by using HELP buttons at top-right of the page.

2. Submit your job and perform analysis
After submitting your job, the agriGO will pre-check the validity of your upload data. If your job is submitted successfully, a job ID will be given. Since the analysis process could take a while, you may close the waiting page and use the job ID to check the work later. Please note that results of your jobs will be stored on our server for THREE DAYS. After 3 days all information of the job will be deleted. If you want elongation contact me.

3. Explore results
The agriGO provides different ways to browse results of different tools. Some of them are flexible but you may need some specific setting to make them to castor to your own demands. And detailed introduction to these tools in the manual in the following will help you to achieve it.
A. How to use Singular Enrichment Analysis (SEA) analysis?

SEA is a traditional and widely used method. It is simple to use and simple to understand. User only needs to prepare a list of gene/probe names, and enrichment GO terms will be found out after statistical test from pre-calculated background or customized one.

**STEP 1:**

To use SEA analysis, you should firstly select the type of your query list, either single names or names with GO accession. If you choose using supported species in agriGO, you only need to provide a list of sequence identifiers. It should be noted that you would better select species and check all allowed ID types of corresponding species, then submit your IDs. Only allowed IDs are suitable to be analysis in this type mode. And you can mix your IDs from different types. Just ensure they are allowed IDs.

If you choose customized mode, you are no long limited by agriGO-owned species any more. You can use any IDs you have, but only be noticed IDs should attach with GO accession!

OK, theoretically you can just click “Submit” to perform analysis now and simply skip following steps. Nevertheless, if you want set more advanced parameters, then keep on reading this manual.
STEP 2:
Now you can set the background or reference. There are three types: suggested backgrounds, customized reference and customized annotated reference.

The default parameter is using suggested backgrounds. For each species, agriGO will give all possible the background types. To those species without a relatively completed profile, backgrounds from neighbored organisms are suggested. Users can select based on their practical need, otherwise use customized reference.

In the case that you do not want any of suggested pre-computed background, you can use customized reference instead. NOTE: IDs in reference list should from the same species that one selected above for query list.

Also you can use any IDs if you choose customized annotated reference mode, however, the price is to attach with GO accession to obtain such freedom.

You can paste direct or upload your file, for latter, please make sure the file no bigger than 4MB.
STEP 3:
The advanced options are optional but quite important. These options are default hidden, and need to one click to make them visible. In SEA analysis, there are three statistical test methods: hypergeometric, chi-square and fisher test. When the input/query list is compared with the previously computed background, or is a subset of reference list, choose hypergeometric or fisher. When both of your query list number and reference list number are quite small, you may better choose fisher test. When the input/query list has few or no intersections with the reference list, the Chi-square tests are more appropriate. Next you can choose method to do the multi-test adjustment. Seven adjustment methods are available here, including: Yekutieli (FDR under dependency), Bonferroni, Hochberg, Hochberg (FDR), Hommel, Holm, False Discovery Rate. Though I would suggest perform adjustment test, you truly can turn off it and use no adjust. While you choose no adjust, then you may set significant level below higher. Terms under the cutoff of the significant level will be highlighted, and emphasized in analysis results, and it will affect your test output. Minimum number of mapping entries means that GO annotations that do not appear in at least the selected number of entries will not be shown. In other word, higher you set the number, more entries needed to make one GO term appear in the analysis result. Gene ontology type: Plant GO slim is a cut-down version of the GO ontologies containing a subset of the terms in the whole GO for plant. Last, if you provide a mail address, a notification will be send when the analysis is completed with the link to the results. Providing a email address is optional to SEA analysis, because it is very fast.

Greeting! You can now click submit to perform the analysis. You can always get interactive help from the right help frames, and a detailed tutorial in this manual, if you still have any question then contact me directly.
In the following we will discuss the outputs of the SEA analysis.
Singular Enrichment Analysis (SEA) Results

Part 1:
A brief summary of your job will be given. The job ID is useful within 3 days. A file containing all entities in the query list that can be annotated by GO associated with descriptions is able to download.

<table>
<thead>
<tr>
<th>Analysis Brief Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Job ID: 952249782 [Useful within 3 days]</td>
</tr>
<tr>
<td>Species: Arabidopsis thaliana</td>
</tr>
<tr>
<td>Background/Reference: Affymetrix ATH1 Genome Array (SPL198)</td>
</tr>
<tr>
<td>Annotated number in query list: 168 [ Download ]</td>
</tr>
<tr>
<td>Annotated number in background/reference: 22479</td>
</tr>
<tr>
<td>Significant GO terms: 23 [ Details ]</td>
</tr>
</tbody>
</table>

Part 2:
In this part, you can browse the hierarchical graph result. Note that the graphical result was generated as separate graphs for each of the three GO categories, namely biological process, molecular function, and cellular component. After select the category, users can specified their favorite output format, graph rank direction and font size. The result format means which output format you preferred. The rank direction is used to define the direction in your output, for instance the direction in the example image is ‘top to bottom’. And the font size is self-evident that user can set smaller size if there are many nodes in their result.

Click the ‘generate image’ button after you set all parameters well. The graphical result will be presented according to your own settings. The graphical result is a GO hierarchical image containing all statistically significant terms. These nodes in the image are classified into ten levels which are associated with corresponding specific colors. The smaller of the term’s q-value, the more significant statistically, and the node’s color is darker and redder (Note: q-value here means that the value of the multiple-test adjusted p-value). Inside the box of the significant terms, the information includes: GO term, q-value, GO description, item number mapping the GO in the query list and background, and total number of query list and background. But when those term whose q-value is higher than the cutoff set by the user, only GO information will be given in the box.

To better understand the graphical result, investigation of the annotation diagram is suggested. If user chooses PNG or JPG or GIF result format, linkage to the term’s detail is available by clicking those blocks.
Part 3:
The terms selected here are children terms of root one (or called secondary level terms) or significant terms of secondary level terms. Thus, the bar chart gives user a brief portray since the GO terms are relatively general description. Similar to the procedure of graphical result, user should specified their parameters before create the GO abundance chart. User can try these setting to obtain favorite view of the chart bar. Note the setting you used will be recorded in your cookie and these settings will be default ones in your future jobs. In other word, you may try several times and make your last attempt as your own features.

Here the bar chart is using glass bar style, default colors, GO annotation as X legend, 14px font and 300 for X legend rotation. Here are some tips: 1. Have a glance of all four bar styles and select one you like, 2. Use HEX format to define colors and there is a website we already suggested, 3. If you prefer GO annotation as X legend content, you may use smaller font once there are too many words, 4. 270 to 315 is suggested for X legend rotation, in which 270 means vertical, and 315 means 45 degree slope, and you can try other number which may satisfy your taste but seems somehow strange to me 😊.

The bar chart is created based on scripts from Open Flash Chart. It is powerful. You can drag borders to resize and adjust the image size and ratio. And bars are accessible to term’s detailed information. A ‘Save as Image’ bottom is existed but only useful when you are using FireFox browser, and if you can also use your ‘Print Screen’ bottom on your keyboard or other tools to download this image.
Part 4:
In this part, detailed information is given. All GO significant terms will presented in the following table. And you can browse the GO terms using tree traversing mode (we will discussed it later), or can browse all GO terms in the similar type table, or just the data.
User can select terms to draw graphical result or create bar chart. Please note that the parameters used in graph or chart generating is fetched from your cookie, and your cookie will be set or changed when you generate graphical results or GO abundant chart which has been mentioned in part 2 and 3. While it will make you a bit trouble if you would like adjust the images created here to redo the part 2 or 3 work once more to change the settings.
Click the checkbox left to ‘GO term’ can select all GO terms at one time.

You can click the GO name to collapse/extend ontology terms in tree traversing mode. A bottom that can make all significant selected or not is available and those selected terms can be used in drawing graphical results or to create bar chart. Please note that at least one significant term should be included in graph generation, otherwise the graphical result will be some kind blank and meaningless. Click on the number will lead you to term’s detail information.
The term’s detail page is as following. The agriGO will give all entries can be annotated to the term besides a brief summary. And for each entry the annotation includes: GO terms, GO source, description.
B. How to use Parametric Analysis of Gene Set Enrichment (PAGE)?

PAGE method is argued by Kim [BMC Bioinformatics 2005, 6:144]. Using Central Limit Theorem in statistics, this method is simple and efficient. Different to SEA, it takes expression level into account, and can deal with a long list of genes/probesets.

STEP 1:
Firstly, you should choose the species for your query data. Please make sure that identifiers in your input should be one of datatypes inside the right information table. If your identifiers are not stored in agriGO, there is another two ways: one is provided your own GO annotation file, the other is to use our BLAST4ID service.

![Select the species](image)

STEP 2:
In PAGE analysis, user should pay more attention to input data. As presented in the following image, as least two rows must be provided. The first row is sequence identifiers, and followings are numerical value. The numerical value is fold change (FC) or log2-transformed FC value (latter preferred) of the identifiers' expression under different condition. If you do not have expression data, then SEA may be the alternative choice.

In agriGO’s example, there are 3 rows in this example. First row is ATH1 probeset name, the second row is expression fold change (FC) value of cold treatment to CK(cold/CK) after half hour. Third row is expression FC of cold/CK after 24 hour cold treatment. Only 600 probesets are in the quick example for the fast load of the HTML page. To obtain a full view of PAGE method, you can download the full example file and explore the following analysis procedure.

![Submit your data](image)
**STEP 3:**
Next you can choose method to do the multi-test adjustment. Seven adjustment methods are available here, including: Yekutieli (FDR under dependency), Bonferroni, Hochberg, Hochberg (FDR), Hommel, Holm, False Discovery Rate. Though I would suggest perform adjustment test, you truly can turn off it and use no adjust. While you choose no adjust, then you may set significant level below higher. Terms under the cutoff of the significant level will be highlighted, and emphasized in analysis results, and it will affect your test output.
Minimum number of mapping entries means that GO annotations that do not appear in at least the selected number of entries will not be shown. In other word, higher you set the number, more entries needed to make one GO term appear in the analysis result.
Gene ontology type: Plant GO slim is a cut-down version of the GO ontologies containing a subset of the terms in the whole GO for plant.
If you can also upload your own customized GO annotation file once your identifiers are not accepted directly by agriGO. The file’s size is limited to 4MB.

OK, now you can click submit to start analysis now. You may explore output of analysis results in the following part of manual.
Parametric Analysis of Gene Set Enrichment (PAGE) Result

Results generated by PAGE analysis have many similar points to SEA analysis, thus it is suggested to browse SEA result introduction part firstly. And only unique features to PAGE results will be explained. Since PAGE tool can analysis several rows at one time, and terms in each row will be calculated, each row has its significant GO terms. Number of significant GO terms for each row is listed in the brief summary part.

The number of terms is determined by the row you selected which is colored by red. A simple colorful model named CM for short is available. The color used in the CM is same to the color used in graphical result in which red color system means up regulated and blue means down regulated. And each block present the term’s Z-score for the row. You can select row(s) and term(s) to generate further images.

The term’s detailed information is generated if you click the number. This page may be a bit simple because it is quite possible that there are too many entries mapping to the GO.

In graphical result part, user can choose one or two rows to draw the image. If two rows are selected, a third color system (purple colors) will be used in demonstrating those terms have different regulation direction in two rows.
The following example presents two rows in one graph. You can check the annotation diagram below the result. There are three color systems: red means up regulated terms, blue means down regulated and purple presents the term is regulated in different direction in tow rows. And if the term has same regulated direction in both rows, it will have double borders. In the box ‘r1=1e-10’ means the q-value of the term in row1 is 1e-10, and ‘zs’ presents Z-score.

In bar chart generation part, Z-score is the statistical value in PAGE calculation, mean value is the mean of the value of all entries in the row. Mean change is mean minus standard deviation which presents the change of expression when comparing to the whole row background. While user can set two color values for up-regulation terms and down-regulation terms.

As mentioned before, Z-score which is bigger than 0 or smaller than 0 will be presented using different colors which set by user.
But if you choose mean value, they are in the same color since all mean is bigger than 0.
C. How to use BLAST4ID tool?

The BLAST4ID tool is not an analysis tool, but an associated one used mainly for two purposes: 1. Transfer your IDs which are not available to agriGO to available ones, 2. use blast search to annotate your sequences with GO.

To use BLAST4ID, user should set target database at first, and then E-value cutoff. The program should be correctly selected based on sequence types of user’s input and target database. Generally speaking, all array sequences are nucleotide and other genome sequences are protein. The process may take a long while thus the E-mail address should be given for the Email notification.

2. Set parameters:

<table>
<thead>
<tr>
<th>Target database</th>
<th>Arabidopsis TAIR pep [eg:AT3G34250.1]</th>
</tr>
</thead>
<tbody>
<tr>
<td>E-value cutoff</td>
<td>1e-3</td>
</tr>
<tr>
<td>Choose a program</td>
<td>BLASTP, BLASTN, BLASTX, TBLASTN, TBLASTX</td>
</tr>
<tr>
<td>Email address (required)</td>
<td></td>
</tr>
</tbody>
</table>

The result interface is a bit simple, but enough for usage.

There are 88 entities of total 129 matched. Download

Downloadable text result:

<table>
<thead>
<tr>
<th>query_name</th>
<th>target_name</th>
<th>evalue</th>
<th>GO_accessions</th>
</tr>
</thead>
<tbody>
<tr>
<td>IRG4L4101000.1</td>
<td>AT5G25190.1</td>
<td>0.0</td>
<td>00:0019825 00:0042761 00:005783 00:010394</td>
</tr>
<tr>
<td>IRG4L4101001.1</td>
<td>AT3G26700.1</td>
<td>0.0</td>
<td>00:005075 00:003674 00:0038159</td>
</tr>
<tr>
<td>IRG4L4101002.1</td>
<td>AT5G37860.1</td>
<td>3e-34</td>
<td>00:0006578 00:000674</td>
</tr>
<tr>
<td>IRG4L4101003.1</td>
<td>AT4G38650.1</td>
<td>2e-24</td>
<td>00:0009733 00:0009409 00:003674 00:0005881</td>
</tr>
<tr>
<td>IRG4L4101004.1</td>
<td>AT5G0005.1</td>
<td>1e-06</td>
<td>00:000575 00:0008159 00:003676</td>
</tr>
<tr>
<td>IRG4L4101005.1</td>
<td>AT2G10700.1</td>
<td>4e-14</td>
<td>00:0004074 00:0004060 00:012805 00:000521</td>
</tr>
<tr>
<td>IRG4L4101006.1</td>
<td>AT1G76200.1</td>
<td>1e-16</td>
<td>00:0003676 00:000321 00:0038160</td>
</tr>
<tr>
<td>IRG4L4101007.1</td>
<td>AT3G26520.1</td>
<td>5e-45</td>
<td>00:0005534 00:0005974 00:008478 00:0009371 00:003676 00:000784</td>
</tr>
<tr>
<td>IRG4L4101008.1</td>
<td>AT4G10130.1</td>
<td>1e-35</td>
<td>00:0005076 00:0031072 00:0035487</td>
</tr>
<tr>
<td>IRG4L4101009.1</td>
<td>AT4G28500.1</td>
<td>3e-62</td>
<td>00:0006972 00:010256</td>
</tr>
<tr>
<td>IRG4L4101010.1</td>
<td>AT3G18520.1</td>
<td>2e-110</td>
<td>00:0008276</td>
</tr>
<tr>
<td>IRG4L4101011.1</td>
<td>AT4G36840.1</td>
<td>2e-25</td>
<td>00:0009733 00:0009409 00:003674 00:0005881</td>
</tr>
</tbody>
</table>
D. How to use search tool?
Search tools in agriGO are easy to understand and use. Here are some tips:

1. Unless you contact me and ask for elongation, the job ID is available within 3 days.

2. There is a short-cut at top-right corner for job searching.

3. In advance search, you have to define the species firstly.

4. You can either search single one or a list (no more than 100) of sequence identifiers.

5. Input IDs are case insensitive, but will be agriGO’s format in the output.

![Search analysis result](image)

![Advance search](image)
6. FAQ

What is agriGO?
The agriGO is designed to automate the job for experimental biologists to identify enriched Gene Ontology (GO) terms in a list of microarray probe sets or gene identifiers (with or without expression information) and it is also a GO-related database. The agriGO specially focus on agricultural species.

What is GO?
"The Gene Ontology (GO) project provides a controlled vocabulary to describe gene and gene product attributes in any organism. The GO project is a collaborative effort to address the need for consistent descriptions of gene products in different databases. The GO collaborators are developing three structured, controlled vocabularies (ontologies) that describe gene products in terms of their associated biological processes, cellular components and molecular functions in a species-independent manner. There are three separate aspects to this effort: first, we write and maintain the ontologies themselves; second, we make cross-links between the ontologies and the genes and gene products in the collaborating databases, and third, we develop tools that facilitate the creation, maintainence and use of ontologies." Definition from http://www.geneontology.org/

What is updated in agriGO compare with EasyGO?
The agriGO is a successor of EasyGO, and it go further. 1. We create new website interface. The database structure and scripts of agriGO are redesigned. Both page loading speed and analysis speed of agriGO now are improved because of the change. 2. The agriGO service is especially focus on agricultural species. It supports species is extended to 35 including as much as 280 datatypes. 3. We added new analysis tools for new agriGO, such as PAGE analysis and BLAST4ID tool. 4. The result output and information are richer compare with EasyGO. 5. The agriGO could also work as a GO database with search and download service.

What are the unique features of agriGO compare with other GO webserver/database?
The agriGO provides heavy support to agricultural species. Not only limited to SEA analysis, GSEA which is achieved using PAGE method is also available. Furthermore we have BLAST4ID tool for ID transfer or annotation. And search as well as download function is accessible. The agriGO can give out rich outputs like graphical result, bar chart result and hierical tree which composing a comprehensive understanding of biological meaning of user's input data.
What is SEA analysis?
SEA analysis means Singular enrichment analysis which is traditional but widely used. SEA analysis is designed to identify enriched Gene Ontology (GO) terms in a list of microarray probe sets or gene identifiers. Finding enriched GO terms corresponds to finding enriched biological facts, and term enrichment level is judged by comparing query list to a background population from which the query list is derived.

Which statistics method should I choose in SEA tool?
When the input list is compared with the previously computed background, or is a subset of reference list, choose hypergeometric or fisher, for latter only when your query number is quite small. When the input list has few or no intersections with the reference list, the Chi-square tests are more appropriate.

What is PAGE analysis?
PAGE is Parametric Analysis of Gene Set Enrichment [Kim et. 2005 BMC Bioinfomatics]. PAGE method is using Central Limit Theorem in statistics, this method is simple and efficient. Different to SEA, it takes expression level into account, and can deal with a long list of genes/probesets. PAGE use a two-tailed test to count Z score, and the calculation of p-value will be:
if Z score >= 0: p-value is 2 * (1 - x)
if Z score < 0: p-value is 2 * x

What is BLAST4ID?
The BLAST4ID tool is not an analysis tool, but an associated one used mainly for two purposes: 1. Transfer your IDs which are not available to agriGO to available ones, 2. use blast search to annotate your sequences with GO.

Which tool should I choose?
It will depend on what data you have. If you only have a list of identifiers or only interested about them, SEA will be your choice. And if you like take expression data into count and would like compare several dateset then you may try PAGE. The BLAST4ID is only an associated tool, use it if you really need it.

Why graphical/chart image does not display on my PC?
The bar chart result need flash player to browse correctly. And you may need different tool to display different format graphical result, for example: Adobe reader, SVG browser. Contact me if you install related tool but still can not see the results.
How many datatypes are supported by agriGO?
We currently support on 35 species including 280 datatypes. Please check the data statistics page for detail information. We will continue adding more species and datatypes.

How agriGO obtains its data source?
Raw GO annotation data is generated using BLAST, Pfam, InterproScan by agriGO or obtained from B2G-FAR center or from Gene Ontology. Arabidopsis genome data is from TAIR. Rice TIGR genome data is from Rice Genome Annotation Project. Rice KOME data is from KOME database. Rice Gramene data is from Gramene center. Populus genome data is collected from JGI. Soybean and Sorghum genome data is compiled from phytozome. Grape genome data is compiled from Genoscope. Medicago genome data is from Medicago truncatula sequencing resources. Maize genome data is from MaizeSequence.org. Castor bean genome data is from Castor Bean Genome Database. Brachypodium distachyon genome data is from Ensembl. Bovien genome data is from Bovine Genome Database. Silkworm genome data is from SilkDB. M. grisea genome data is from Magnaporthe grisea Database. affymetrixmetrix CSV files and array sequences are from NetAffx.

How often does agriGO update?
Normally we will update our database every 3 months, but if we will update agriGO if some important data source is newly available. Improvement and updating to agriGO tools are irregulated.

Can I check result from old version by new agriGO?
Sorry, but no. Because we reconstructed the database and redesigned the website organization, analysis result from EasyGO is not supported in agriGO.

How to make agriGO add new customized datatype?
User can contact the agriGO administrator by email (adugduzhou@gmail.com) to discuss more details.