

Site Search

Learning objectives:

- Use keywords in site search.
- Explore site search results.
- Filter site search results by categories.
- Filter site search results by organisms.
- Filter site search results by category fields.
- Export results to a search strategy.
- Find a specific gene using its ID in site search.

The site search is located in the header of the site and is available from every page. The site search queries the database for a term (e.g., text) or ID and returns a list of pages and documents that contain the query term.

Site search: text, term or gene id.

1. Enter the word kinase in the site search window (at the top centre of the page). Click on the "enter" key on your keyboard or on the search icon as shown in the screenshot below.



2. How many results with the word kinase did you get? Are all of these records genes?
3. Explore the filter panel on the left side of the page. Filter the results to view gene results only (hint: click on the word *Genes* in the *Filter results* section):

This screenshot shows the search results for 'kinase' on the FungiDB website. At the top, it says 'All results matching kinase' and '1 - 20 of 325,297'. Below this is a 'Filter results' sidebar with various categories like 'Genome', 'Genes' (which has an orange arrow pointing to it), 'Population biology', 'Metabolism', 'Data access', etc. The main area lists results with their counts: 316,869 for Genes, 7,869 for Population biology, 352 for Metabolism, 9 for Data access, etc. Each result entry includes a link to a detailed page with gene information like 'Gene - A9K55_006619 MAP kinase kinase kinase'.

Notice that clicking on the “Genes” category reveals additional filtering options.

4. Select and apply the *Product descriptions* filter.

Note: The applied filter can be easily cleared by clicking on “Clear filter” option.

5. In the “Filter organisms” section, select to filter gene results by *Malassezia restricta* KCTC 27527. How many genes contain “kinase” in the product description field in this organism?

6. Export the results to a search strategy.

To achieve this, click on the blue button called “Export as a search strategy...” at the top right-hand side of the results page. Notice that before the Genes category was selected this button was inactive. This is because the search strategy can be deployed on a single category only (e.g. Genes or Data sets, but not both).

7. Try running the same search but this time use a wild card (*) (e.g., kinase*).

When the wild card is combined with a word (kinase * or *kinase), the search will retrieve compound words ending or beginning with the word kinase (e.g. phosphofructokinase). The wild card (*) can be used alone to retrieve all records available to the site search (see screenshot below).

All results matching *

1 - 20 of 4,901,548

Hide zero counts

Filter results

Genome	1,885,291
Genes	162,441
Genomic sequences	
Organism	186
Organisms	
Transcriptomics	
ESTs	1,709,817
Population biology	
Popset isolate sequences	1,077,920
Metabolism	
Metabolic pathways	3,045
Compounds	61,998
Data access	
Data sets	381
Searches	435
Instructional Tutorials	15
Workshop exercises	1
About	2
News	
General info pages	16

Filter fields
Select a result filter above

Filter organisms
select all | clear all | expand all | collapse all

Type a taxonomic name

Hide zero counts

Export as a Search Strategy to download or mine your results

Compound - CHEBI:10000 Vismone D
Compound - CHEBI:10001 Visanadin
Compound - CHEBI:10002 Visnagin
Compound - CHEBI:10003 ribostamycin sulfate
Definition: An aminoglycoside sulfate salt resulting from the reaction of ribostamycin with sulfuric acid.
Compound - CHEBI:10014 nalidixic acid
Definition: A monocarboxylic acid comprising 1,8-naphthyridin-4-one substituted by carboxylic acid, ethyl and methyl groups at positions 3, 1, and 7, respectively.
Compound - CHEBI:10015 vobasine
Definition: An indole alkaloid that is vobasan in which the bridgehead methyl group is substituted by a methoxycarbonyl group and an additional oxo substituent is present in the 3-position.
Compound - CHEBI:10016 vobtusine
Compound - CHEBI:10017 volmentol
Definition: A heptol that is heptane-1,2,3,4,5,6,7-heptol that has R-configuration at positions 2, 3, 5 and 6.
Compound - CHEBI:10018 volkenin
Definition: A cyanogenic glycoside that is (4R)-4-hydroxycyclopent-2-ene-1-carbonitrile attached to a beta-D-glucopyranosyloxy at position 1.
Compound - CHEBI:10019 Vornicine
Compound - CHEBI:10022 Vomitoxin
Compound - CHEBI:10024 voriconazole
Definition: A triazole-based antifungal agent used for the treatment of esophageal candidiasis, invasive pulmonary aspergillosis, and serious fungal infections caused by *Scedosporium apiospermum* and *Fusarium* spp. It is an inhibitor of cytochrome P450 2C9 (CYP2C9) and CYP3A4.
Compound - CHEBI:100241 ciprofloxacin
Definition: A quinolone that is quinolin-4(1H)-one bearing cyclopropyl, carboxylic acid, fluoro and piperazin-1-yl substituents at positions 1, 3, 6 and 7, respectively.

COMMUNITY CHAT

8. The site search also works with gene ids. Run a site search for the following gene id: Afu2g13260

The gene id search will return the gene record card for [Afu2g13260](#) (see screenshot below). Click on the gene link in blue to navigate to the gene record page for this gene.

Genes matching Afu2g13260

1 - 1 of 1

Hide zero counts

Filter results

Genome	1
Genes	1

Filter Gene fields
select all | clear all
 External links
 Gene ID
 Names, IDs, and aliases
 User comments

Filter organisms
select all | clear all | expand all | collapse all
Type a taxonomic name Reference only
 Fungi
 Ascomycota
1

Gene - Afu2g13260 Developmental regulator medA, putative
Gene name or symbol: medA
Gene type: protein coding gene
Organism: *Aspergillus fumigatus* Af293
▶ Fields matched: External links; Gene ID; Names, IDs, and aliases; User comments

Gene - Afu2g13260 Developmental regulator medA, putative
Gene name or symbol: medA
Gene type: protein coding gene
Organism: *Aspergillus fumigatus* Af293
▶ Fields matched: External links; Gene ID; Names, IDs, and aliases; User comments

1 - 1 of 1

Export as a Search Strategy to download or mine your results

Note: a single gene id can be also exported as a search strategy. This may be useful if you are interested in cross-referencing different types of data for one gene.

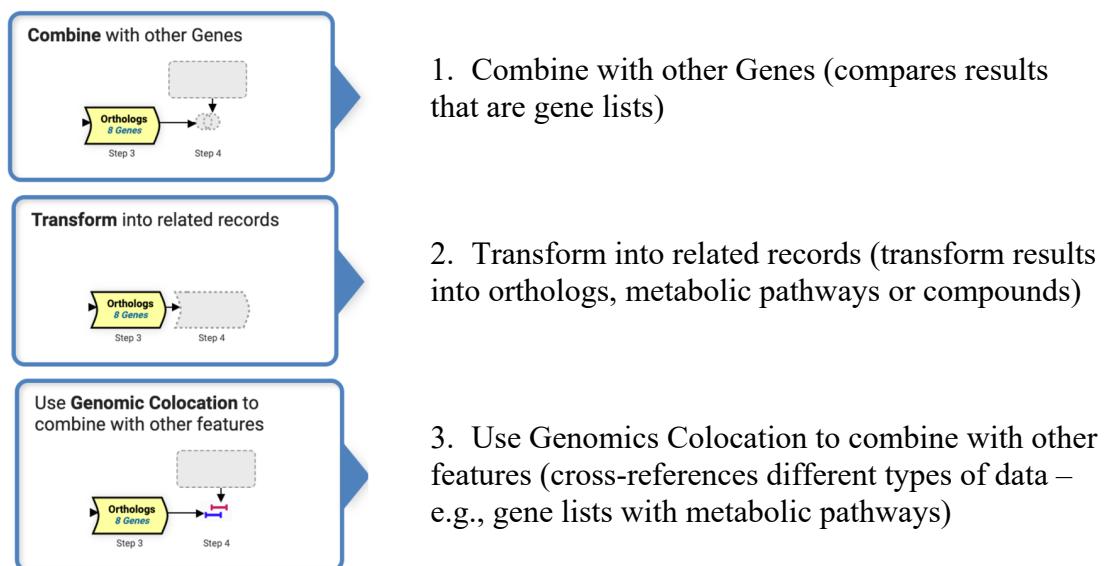
Advanced Search Strategies

Learning objectives:

- Use sites search and other types of searches to create a multi-step query across different types of records and genomes.

The strategy system offers over 100 structured searches that can be combined to produce multi-step strategies. Searches can be deployed from the site search, or ‘Search For...’ menu on the home page and from the ‘Searches’ dropdown menu in the header of every page. Searches listed under Genes will return a list of gene IDs, while searches listed under ‘SNPs’ or ‘Metabolic Pathways’ will return record IDs representing SNPs, or metabolic pathways, respectively, etc.

The searches can be combined via three major approaches:



Strategy steps are connected via the Boolean operators that can intersect, unite, or subtract similar records (e.g., gene lists) and cross-references different types of data via the genomic colocation option. Steps can be masked off from the strategy with the help of “ignore step” Boolean operators.

Revise as a boolean operation	
<input checked="" type="radio"/> 1 INTERSECT 2	<input type="radio"/> 1 UNION 2
<input type="radio"/> 1 MINUS 2	<input type="radio"/> 2 MINUS 1
Revise as a span operation	
<input type="radio"/> 1 RELATIVE TO 2, using genomic colocation	
Ignore one of the inputs	
<input type="radio"/> IGNORE 2	<input type="radio"/> IGNORE 1
<input type="button" value="Revise"/>	

M. restricta can cause skin disorders and is one of the most common fungal species found on human skin. *Malassezia* cannot produce fatty acids and relies on fatty acid uptake from external sources. Secreted lipases are thought to contribute to *Malassezia* pathogenicity. In this strategy we will identify secreted lipases in *M. restricta* KCTC 27527, cross-reference annotation with InterPro domain annotations and find orthologs of *M. restricta* genes in another *Malassezia* strain and also *Candida albicans* (REF).

To build this strategy, use the following approach:

- **Use site search** to identify genes that have “lipase” annotation in *Malassezia restricta* KCTC 27527. This search identifies genes that have “lipase” annotation in several evidence fields.
- **Identify Genes by Signal peptide prediction.** This search returns genes predicted to have signal peptide.
- **Identify Genes based on InterPro domain.** This search identifies genes with specific domain signature – secreted lipase (LIP).
- **Transform by Orthology into another organism.** FungiDB integrates ortholog profiles from OrthoMCL. The OrthoMCL algorithm clusters proteins into ortholog groups based on BLAST similarity across at 150 genomes that span the tree of life. The transform we perform here will convert a list of genes in one organism to their orthologs in a different organism. In this case, we will transform a list of *M. restricta* KCTC 27527 genes into their orthologs in *Malassezia restricta* CBS 7877 and *Candida albicans* SC5314.

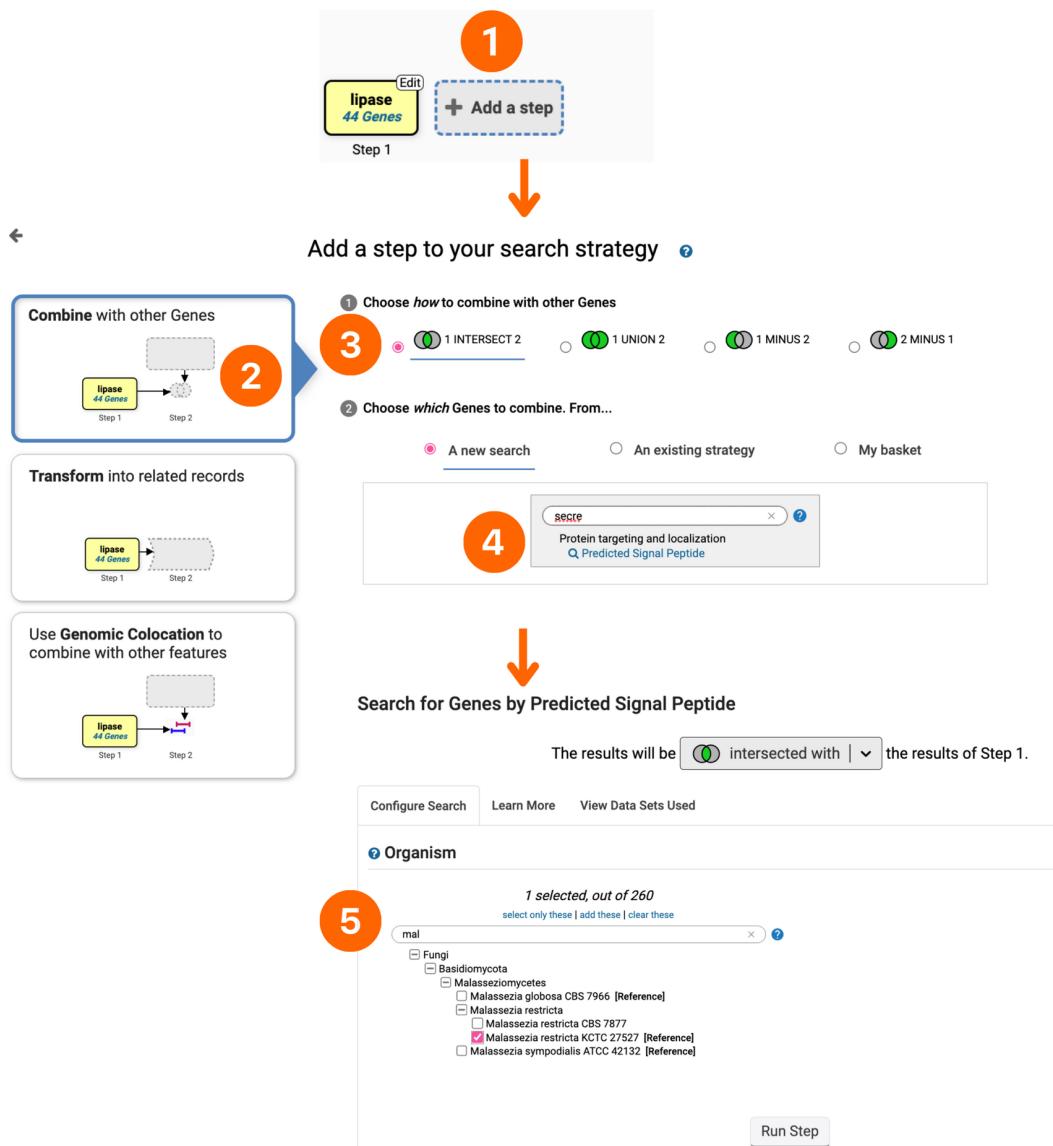
- Use site search to identify genes that have “lipase” annotation in *Malassezia restricta* KCTC 27527

1. Run site search for genes annotated with “lipase” and filter on Genes.
2. Use Gene fields to filter your results as shown.
3. Restrict your search to *M. restricta* KCTC 27527 genes.
4. Export results as a search strategy.

The screenshot shows a search interface with the following components:

- Search Bar:** Contains the text "lipase".
- Header:** Includes links for "My Strategies", "Searches", "Tools", "My Workspace", "Data", "About", "Help", and "Contact Us".
- Section 1 (Filter results):** Shows a list of filters applied: "Genome Genes" (selected), "Hide zero counts" (unchecked), and a "Clear filter" button (labeled 44).
- Section 2 (Filter Gene fields):** Shows a list of selected filters: "EC descriptions and numbers" (19), "GO terms" (4), "InterPro domains" (21), "Notes from annotators" (0), "Orthologs" (37), "Product descriptions" (14), "Phenotype" (0), "Preferred product description" (17), "Product descriptions" (17), and "User comments" (0). A "Clear filter" button is also present.
- Section 3 (Filter organisms):** Shows a tree view of organisms: "mala" (selected), "Fungi", "Basidiomycota", "Malasseziomycetes", "Malassezia", "Malassezia globosa CBS 7966 [Ref] (181), "Malassezia restricta" (87), "Malassezia restricta CBS 7977 (43), "Malassezia restricta KCTC 27527 [Ref] (44), and "Malassezia sympodialis ATCC 42132 [Ref] (47). A "Clear filter" button is also present.
- Results List:** Displays a list of genes matching the criteria, each with details like name, type, organism, and fields matched.
- Export Button:** Located in the top right corner, labeled "Export as a Search Strategy" with a link to "to download or mine your results".

- **Identify Genes by Signal peptide prediction.** This step will identify lipases that may be secreted.
 1. Click on the “Add step” button.
 2. Choose “Combine with other genes” search.
 3. Choose to “intersect” your results with the previous step.
 4. Filter the available searches to deploy the “Predicated Signal Peptide” search.
 5. Restrict the search to *M. restricta* KCTC 27527 and click on the “Run Step” button.

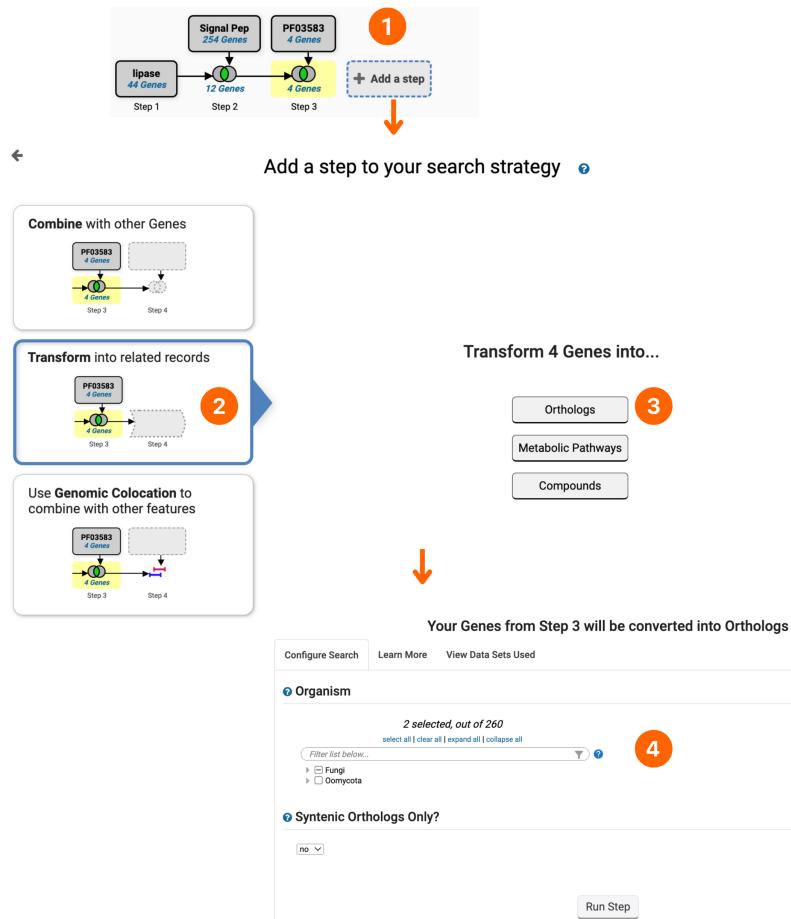


- **Identify Genes based on InterPro domain.** This search identifies genes with specific domain signature – secreted lipase (LIP).
 1. Click on the “Add step” button.
 2. Choose “Combine with other genes” search.
 3. Choose to “intersect” your results with the previous step.
 4. Filter the available searches to deploy the “InterPro domain” search.
 5. Restrict to *M. restricta* KCTC 27527, select “Secretory lipase” domain (PF03583 : LIP Secretory lipase), and click on the “Run Step” button.



- **Transform by Orthology into another organism/s.** This search is particularly useful if you are working with a poorly annotated genome and want to take advantage of annotations from another, better annotated, genome. In this exercise, we will practice finding orthologs in *Malassezia globosa* CBS 7966 and *Candida albicans* SC5314.

1. Click on the “Add step” button.
2. Choose “Transform into related records” search.
3. Choose to deploy the “Orthologs” search.
4. Restrict the orthologs search to *M. globosa* and *Candida albicans* SC5314 and click on the “Run Step” button.



Examine your results. Do they make sense?

The screenshot shows the FungiDB ortholog search interface. At the top, a search history panel displays the following steps:

- Step 1: lipase (44 Genes)
- Step 2: Signal Pep (254 Genes) - resulting in 12 Genes
- Step 3: PF03583 (4 Genes) - resulting in 4 Genes
- Step 4: Orthologs (16 Genes)

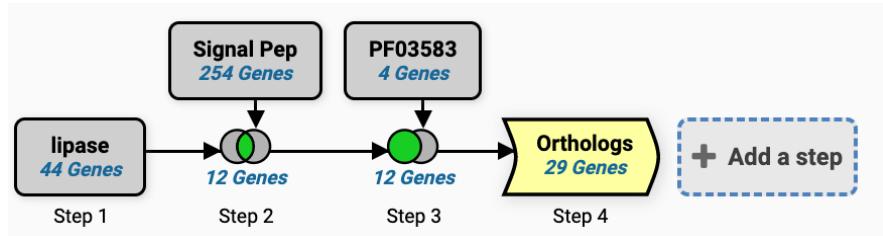
A '+ Add a step' button is available for further refinement.

The main area shows a table of 16 ortholog groups with the following columns:

Gene ID	Transcript ID	Organism	Product Description	Input Ortholog(s)	Ortholog Group	Paralog count
C1_09420W_A	C1_09420W_A-T	<i>Candida albicans</i> SC5314	Triacylglycerol lipase [Source:UniProtKB/TremBL;Acc:Q5APG1]	MRET_0930,MRET_1179,MRET_4098,MRET_4099 OG6_130000	9	
C1_09580C_A	C1_09580C_A-T	<i>Candida albicans</i> SC5314	Triacylglycerol lipase [Source:UniProtKB/TremBL;Acc:A0A1DBPER6]	MRET_0930,MRET_1179,MRET_4098,MRET_4099 OG6_130000	9	
C1_09590C_A	C1_09590C_A-T	<i>Candida albicans</i> SC5314	Lipase 10 [Source:UniProtKB/Swiss-Prot;Acc:Q9P4E5]	MRET_0930,MRET_1179,MRET_4098,MRET_4099 OG6_130000	9	
C1_09600C_A	C1_09600C_A-T	<i>Candida albicans</i> SC5314	Lipase 6 [Source:UniProtKB/Swiss-Prot;Acc:Q9P4E8]	MRET_0930,MRET_1179,MRET_4098,MRET_4099 OG6_130000	9	
C1_09900W_A	C1_09900W_A-T	<i>Candida albicans</i> SC5314	Triacylglycerol lipase [Source:UniProtKB/TremBL;Acc:A0A1DBPE03]	MRET_0930,MRET_1179,MRET_4098,MRET_4099 OG6_130000	9	
C6_04490W_A	C6_04490W_A-T	<i>Candida albicans</i> SC5314	Lipase 4 [Source:UniProtKB/Swiss-Prot;Acc:Q9PBW1]	MRET_0930,MRET_1179,MRET_4098,MRET_4099 OG6_130000	9	
C7_02830C_A	C7_02830C_A-T	<i>Candida albicans</i> SC5314	Lipase 5 [Source:UniProtKB/Swiss-Prot;Acc:Q9PBW0]	MRET_0930,MRET_1179,MRET_4098,MRET_4099 OG6_130000	9	
C7_02880C_A	C7_02880C_A-T	<i>Candida albicans</i> SC5314	Lipase 9 [Source:UniProtKB/Swiss-Prot;Acc:Q9P4E6]	MRET_0930,MRET_1179,MRET_4098,MRET_4099 OG6_130000	9	
C7_03300C_A	C7_03300C_A-T	<i>Candida albicans</i> SC5314	Lipase 8 [Source:UniProtKB/Swiss-Prot;Acc:Q9P8V9]	MRET_0930,MRET_1179,MRET_4098,MRET_4099 OG6_130000	9	
CR_09220C_A	CR_09220C_A-T	<i>Candida albicans</i> SC5314	Lipase 7 [Source:UniProtKB/Swiss-Prot;Acc:Q9P4E7]	MRET_0930,MRET_1179,MRET_4098,MRET_4099 OG6_130000	9	
MGL_1311	MGL_1311-126_1	<i>Malassezia globosa</i> CBS 7966	hypothetical protein	MRET_0930,MRET_1179,MRET_4098,MRET_4099 OG6_130000	5	
MGL_3507	MGL_3507-126_1	<i>Malassezia globosa</i> CBS 7966	hypothetical protein	MRET_0930,MRET_1179,MRET_4098,MRET_4099 OG6_130000	5	
MGL_4051	MGL_4051-126_1	<i>Malassezia globosa</i> CBS 7966	hypothetical secretory lipase (family LIP)	MRET_0930,MRET_1179,MRET_4098,MRET_4099 OG6_130000	5	

At the bottom right of the table is a 'COMMUNITY CHAT' button.

How can you lower the stringency of the search by removing the third step from the search without deleting it? (Hint: you will need to use a certain Boolean operator).



Strategy URL:

<https://fungidb.org/fungidb/app/workspace/strategies/import/d3a431b32ee7b32f>

References:

Park et al. J. Microbiol. Biotechnol. 2021; 31(5): 637-644 doi:10.4014/jmb.2012.12048