宏基因组 Galaxy 分析平台 protocol

李蕙秀,丁国春

- Galaxy 分析平台的总体介绍

登录网址为:10.2.5.56

构成版块(如下图),主要包括工具栏、显示栏和操作历史栏三个版块。



工具栏主要包括文件和数据的上传、序列的上传和管理、数据的分类管理和数据分析四大部分。

详细介绍如下图:



- Galaxy 分析平台的详细操作说明

(一) File Manager-----数据的上传

数据上传的方法有两种:upload file from your computer 和 text2file

1、点击 upload file from your computer,出现如下图所示的操作框,操作方法在

图中标注:

<u>Regular</u>	Composite			文件上	传完度	成后, 点击	
				close,	关闭计	该页面	
	(方法一:	将文件拖动	到此空白框内	\sum	文件上传完	成后
						start键被激	活,
			街 Drop fil	les here		点击 start , 上传	开始
	点击,	更改文件类型	🔄 Drop fil 型为tabular	les here		点击start, 上传	开始
	点击,	更改文件类型	🛿 Drop fil 텍为tabular	es here 方法二:点司 从本地文件9	占 choo s	点击start, 上传 se local file, 择文件	开始

- 2、点击 Text2file copy the content in the text area to file , 出现如下图所示的操作
- 框,操作方法在图中标注:

Text2file cop	✓ Options	
Content		
	将需要的数据复制到此框内,点击execute即可	
✓ Execute		

点击输出之后生成的文件需要更改文件类型和名称,具体操作如下图:

Attributes Convert Format Datatype		\checkmark		
Edit Attributes		\cap	Attributes Convert Format Datatype	Permissions
Name: Outfile 改名字 Info: Annotation / Notes:	2059: Outfile 1 line 格式: txt, 数据库: <u>?</u> 巴 O C 네 ?	• / ×	改变数据类型 新类型:文件类型由txt改为 txt tabul 象 dbnsfp.tabular	I tabular 当Galaxy不能正确:
Add an annotation or notes to a dataset; an Database/Build: unspecified (?)			tab⊌lar ▲ 不改文件类型,平 识别该文件	台无法

(二) Meta data manager------对上传的数据进行处理以便后续分析

高 通 量 测 序 数 据 的 原 始 文 件 一 般 为 Report 文 件 , 即 包 含 有

domain、phylum、class、order、family、genus、oturef、样品名称的数据文件。

为方便解释说明,故对下文中出现的一些文件进行编号。

下面以名为"LHXBac.otu.report"的高通量测序数据为例,详述对数据进行重命名、

分组等操作。

按照数据上传的操作步骤将"LHXBac.otu.report"文件上传至平台,编号为1,继

续下列步骤:

1、首先是从所有测序数据中筛选出所有的细菌测序结果:在 meta data manager 模块下,选择 select taxonomic groups from the data frame,如下图中所示,taxonomic OUT report 项目下选择原始数据 LHXBac.otu.report,其他参数如下图设置即可(注意下图中红圈内的两个参数要按照图示的选择),随后得到

名为 data frame selected taxonomic OUT reporter 的文件,编号为 56

	Select Taxonomic groups from the data frame (Galaxy Version 0.1.0)	▼ Options
	Taxonomic OTU report	
	1: LHXBac.otu.report	-
/	Tax Rank	
	1	
	Total Tax Rank	
	7	
	Data type	
	Text	-
	Tax name	
	Bacteria	
~	Seqs Numbers or Relative abundance	
	Seq numbers	-
	✓ Execute	

2、对数据进行重命名

此步骤的目的是:将样品名称改为包含所有实验设计的名称,例如 T1RCon_a,

只在重复前面加下划线,命名规则视实验设计而定。

首先在 TXT 文档中,第一列是原来的样品名称,第二列是对应的更改之后的样

品名称,将该文件命名为 rename samples,上传至平台,改变数据类型(txt 改

为 tabular,操作同(一)-2)。选择 rename samples in a data frame,如下图所

示 data frame 框架中选择上一步得到的 data frame selected taxonomic OTU

reporter 文件, tag: id tabular newid 框架中选择 rename samples, 点击 execute 后

右侧生成名称为 New data frame 的文件,编号为 57,此时可见样品名称已重命

名。New data frame 文件是后续很多分析的源文件。

Rename samples in a data frame (Galaxy Version 0.1.0)	▼ Options
Data frame	
C 4 56: Data frame Selected Taxonomic OTU reporter	•
Tag: id tabular newid	
C 4 C 5: rename samples.txt	•
✓ Execute	

TODO: Fill in help.

3、接下来选择 list samples in a data frame,如下图所示, OTU taxonomic report

框架中选择上一步形成的 new data frame 文件,其他设置不变,点击 execute,

右侧生成名为 samples in the OTU report 的文件,此步目的是为了在系统中列出

所有数据,以便后续分析。

ListSamples in a data frame (Galaxy Version 0.1.0)	•
OTU taxonomic report	
C 57: New Data frame	
Total Tax rank	
7	
✓ Execute	

4、根据实验设计对样品进行分组,初步筛选差异可能存在的处理。在TXT 文

件中列出所有的实验设计(根据样品名称进行分类),例如样品名称为

T1RCon_a、 T2SLow_a、 T3ROrg_b、 T4SCon_c......, 说 明 实 验 包 括

Org、Low、Con 三个处理,T1、T2、T3、T4 四个采样期,R、S 两种生境,因此

在 txt 文件中要反应出这些信息(放在一列)。注意, TXT 文件中的名称要与

数据文件中的样品名称一致,否则系统无法识别。将该 TXT 文件命名为筛选

差异项,上传至平台,并更改数据类型。选择 get one way experiment design,

如下图所示,在 samples listed in the first column.(list samples)框架中选择上一步

生成的 samples in the OUT report 文件,在 one line one factor, levels separated by

tabular 框架中选择筛选差异项.txt 文件,点击 execute,右侧即可生成名为 one

way experiment design 的文件。

GetOneWayExperimentDesign Based on Sample names and Defined levels of Factors. (Galaxy Version 0.1.0)	▼ Options
Samples listed in the first column.(ListSamples)	
1231: Samples in the OTU report	•
One line one factor, levels seperated by tabular.	
● ● ● 8: 筛选差异项.txt	•
Minimum levels to Keep	
2	
✓ Execute	

生成的 one way experiment design 文件的内容为:

1	2	3	4	5
11	T3ROrg	T3SOrg		
23	T1SOrg	T2SOrg	T3SOrg	T4SOrg
4	T4ROrg	T4SOrg		
25	T1SCon	T2SCon	T3SCon	T4SCon
21	T1RLow	T2RLow	T3RLow	T4RLow
20	T3SCon	T3SLow	T3SOrg	
10	T1RLow	T1SLow		
18	T1SCon	T1SLow	T1SOrg	
3	T2ROrg	T2SOrg		
9	T1RCon	T1SCon		
22	T1SLow	T2SLow	T3SLow	T4SLow
14	T2SCon	T2SLow	T2SOrg	
19	T1RCon	T1RLow	T1ROrg	
16	T2RCon	T2RLow	T2ROrg	
24	T1ROrg	T2ROrg	T3ROrg	T4ROrg
8	T2RCon	T2SCon		
15	T4SCon	T4SLow	T4SOrg	
2	T4RLow	T4SLow		
6	T2RLow	T2SLow		
1	T4RCon	T4SCon		
5	T1ROrg	T1SOrg		
1	T1RCon	T2RCon	T3RCon	T4RCon
7	T3RCon	T3SCon		
17	T4RCon	T4RLow	T4ROrg	
13	T3RCon	T3RLow	T3ROrg	
12	T3RLow	T3SLow		

该结果说明每一行出现的两者、 三者或者四者之间有差异 5、根据上述差异筛选的结果对数据按照因子分类,例如想比较不同生境间(R

和 S) 的差异, 可从上面的筛选结果中挑出所有不同生境的分组, 具体操作步

骤如下:首先下载 one way experiment design 文件,从中将所有根围和土体的数

据复制放在 Excel 中, 如图:

H	В	C
1	T1RCon	T1SCon
2	T1RLow	T1SLow
3	T1ROrg	T1SOrg
4	T2RCon	T2SCon
5	T2RLow	T2SLow
6	T2ROrg	T2SOrg
7	T3RCon	T3SCon
8	T3RLow	T3SLow
9	T3ROrg	T3SOrg
10	T4RCon	T4SCon
11	T4RLow	T4SLow
12	T4ROrg	T4SOrg

选择 file manager 版块下的 text 2 file,将上图中的内容复制放在文本框中,如下

图,点击 execute,右侧生成名为 outfile 的文件,可将文件重命名为醒目的名字,

方便操作,此处重命名为 rhizosphere effect。

Text2file copy the content in the text area to file (Galaxy Version 0.1.0)			
Content			
	1	T1RCon	T1SCon
	2	T1RLow	T1SLow
	3	T1ROrg	T1SOrg
	4	T2RCon	T2SCon
	5	T2RLow	T2SLow
	<u> </u>	TODOra	TICOra
	√ E	xecute	

接下来选择 meta data manager 版块下的 select samples of defined factors from

data frame, taxonomic OTU report 框架中选择第2步生成的名为 New data frame

的文件,list of factors to be selected 框架中选择 rhizosphere effect 文件,点击 execute,右侧生成名为 selected OUT report on factors 的数据集,该数据集包括

12个根据上述指定因子分组的数据。注意,此处生成的是一个数据合集,不

是单一的数据。该数据集的分析在分析版块中详述。

Select Samples of defined factors from data frame (Galaxy Version 0.1.0) • Options			
Taxonomic OTU report			
1 1 57: New Data frame	•		
List of factors to be selected			
C 4 D 1145: rhizosphereeffect	-		
✓ Execute			

6、还有一种根据样品名称对数据进行分组的操作,首先将所有分选出来的数据

的名称放在 TXT 文档中(一列),上传至平台,更改数据类型,命名为 different

period bulk soil。之后选择 select samples based on sample list from data frame,

如下图所示,taxonomic OTU report 下选择编号为 57 的 new data frame 文件

List of samples to be selected(in first column)下选择 different period bulk soil,点击

输出,会生成一个 taxonomic OTU report 文件。该文件内就是根据要求分选出来

的相应组别的数据。

Select samples based on sample list from data frame (Galaxy Version 0.1.0)		
Taxonomic OTU report		
57: New Data frame		
Total Tax Ranks		
7		
List of samples to be selected(in frist column)		
1773: different period bulk soil.txt		
✓ Execute		

7、该模块下还有一个工具是对序列的筛选,主要是为后续 BLAST 和 heatmap 的

制作提供源数据。高通量测序除了 OTU 的结果外,还有序列结果,此处序列的 文件名称为 rep_set.fna,将其上传至平台,其他操作和 BLAST 和 heatmap 的制

作流程一起详述。

(三) Meta diversity analyser-----对处理后的数据进行多种分析

1、Count taxnomic numbers based on a a taxnomic otu reporter,点击该工具,生

成下图,OTU Taxonomic reporter 下选择 new data frame 文件, tax rank 选择 6

(属),会生成名为 Bar plot of numbers of tax 的柱形图,主要是注释到的门、

纲、目、科、属的数量,顶部的数字是基因序列的数量。

Count taxnomic numbers based on a a taxnomic otu reporter (Galaxy Version 0.1.0)
OTU Taxonomic reporter
🗅 🙆 🗅 57: New Data frame
Tax Rank
6
✓ Execute

2、Calculate bray curtis based on a data frame,点击该工具,如下图所示,第一

行选择 new data frame 文件,其他设置不变,点击输出,可得到名为 bray curtis

distance 的柯氏距离结果,该文件是后续 UPGMA、NMMD 等分析的源文件。

Calculate bray o	curtis based on a data frame (Galaxy Version 0.1.0)
Data frame	
C 4 C	57: New Data frame
Tax Rank	
7	
MaxTax Rank	
7	
ScaledOrNot	
ByColumn	
✓ Execute	

3、UPGMA cluster analysis based on a dist matrix,该工具主要是基于柯氏距离,

得到聚类图。点击该工具,如下图所示,第一列选择上一步生成的名为 bray

curtis distance 的文件,右侧生成名为 UPGMA cluster 的聚类图,可根据该图反

应不同处理间的微生物群落结构差异情况。

UPGMA cluster analysis based on a dist matrix (Galaxy Version 0.1.0)	
Dist matrix	
B 4 D	2061: Bray Curtis Distance
✓ Execute	

4、Alpha Diversity of each sample,该工具基于 new data frame 文件,计算所比

较的处理间的 α 多样性,主要包括 chao1 丰富度和均匀度指数。如下图,第一

行选择 new data frame 文件,右侧生成名为 Alpha diversity index 的文件,该文件

包含了六个指标的数据,可在此基础上做统计分析,绘制相应的柱状图。

Alpha Diversity of each sample (Galaxy Version 0.1.0)		
	observed Species	
Taxonomic OTU report	Chao1	
57: New Data frame	ACE	
	Shannon	
Number of Taxonomic rank of the report	Inverted Simpson	
7	Pielous Eveness	
✓ Execute		

5、接下来是针对不同因子数据集的分析,选择 one way multicomp using glm

model for binary data, OUT reporter 框架中选择(二)-5 中得到的名为 selected

OUT report on factors 的数据集,此处注意一点,如下图红圈标注文件类型应选

择第三个 dataset collection,此时系统才可显示出数据集资源。可根据分析需

要改变 tax rank to test。点击 execute,右侧生成名为 statistic results (mean + sd)

significant symbols 的数据集,此数据集包括上述分组数据在 glm 分析模型下计

算得到的属水平的微生物的差异,可下载结果查找有显著差异的微生物类群。

One Way Multicomp using glm model for binary data (Galaxy Version 0.1.0)	▼ Options
OTU reporter	
1146: Selected OTU report on factors	•
This is a batch mode input field. Separate jobs will be triggered for each dataset selection.	
Tax Rank to test	
6	
Total Tax rank	
7	
✓ Execute	

6、接下来选择 combine tables of oneway glm tables (one way glm multicomp) into

results tables,如下图所示,在1: statistic results 框架中选择一个上一步中得到

的 statistic results (mean + sd) significant symbols 数据,随后点击红圈内的 insert

statistic results 依次加入所有需要组合的数据。此步的目的是组合所有数据集内

的 statistic results (mean + sd) significant symbols 数据。

Combine tables of oneway glm tables(OneWayGLMMulticomp) into result tables (Galaxy Version 0.1.0)	
Statistic results	
1: Statistic results	
Statistic results(mean+sd) significant symbols	
L L 1213: VIP variables	
+ Insert Statistic results	
✓ Execute	
Combine tables of oneway glm tables(OneWayGLMMulticomp) into result tables (Galaxy Version 0.1.0)	ons
Combine tables of oneway glm tables(OneWayGLMMulticomp) into result tables (Galaxy Version 0.1.0) Control of the second	ons
Combine tables of oneway glm tables(OneWayGLMMulticomp) into result tables (Galaxy Version 0.1.0) Statistic results 1: Statistic results	ons 面
Combine tables of oneway glm tables(OneWayGLMMulticomp) into result tables (Galaxy Version 0.1.0) Option Statistic results Statistic results(mean+sd) significant symbols 	ons ⑪
Combine tables of oneway glm tables(OneWayGLMMulticomp) into result tables (Galaxy Version 0.1.0) Option Statistic results Statistic results(mean+sd) significant symbols I170: Statistic results(mean+sd)significant symbols 	ons ₪
Combine tables of oneway glm tables(OneWayGLMMulticomp) into result tables (Galaxy Version 0.1.0) Option Statistic results Statistic results(mean+sd) significant symbols I170: Statistic results(mean+sd)significant symbols 2: Statistic results 2: Statistic results Combine tables (Galaxy Version 0.1.0) Option Option Statistic results St	ons
Combine tables of oneway glm tables(OneWayGLMMulticomp) into result tables (Galaxy Version 0.1.0) Option Statistic results Statistic results(mean+sd) significant symbols I170: Statistic results(mean+sd)significant symbols Statistic results(mean+sd) significant symbols Statistic results(mean+sd)significant symbols Statistic results(mean+sd) significant symbols Statistic results(mean+sd) significant symbols 	ons ₪ ₪

右侧生成名为 combined statistic table 的文件,下载该文件调整好样品顺序(按

照 T1、T2、T3、T4 采样期排序)。随后将所有数据复制,放入 tex2 file copy the

content in the text area to file 框架内,如下图所示。

Text2file copy the content in the text area to file (Galaxy Version 0.1.0)	🕶 Optio	ns
Content		
T1ROrg T1SOrg T1RLow T1SLow T1RCon T1SCon T2ROrg T2SOrg T2RCon T2SCon T2RLow T2 T3ROrg T3SOrg T3RLow T3SLow T3RCon T3SCon T4ROrg T4SOrg T4RLow T4SLow T4RCon T4SCon	SLow	^
Bacteria Actinobacteria Acidimicrobia Acidimicrobiales Iamiaceae Iamia 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0	~
Bactoria Astinghactoria Astinghactoria Astingmunetales Missohactoriaseaa Astermuses 4.0.1h 0.7.0.7a 0.4.0.1 ✓ Execute	ь	

操作完成后点击 execute,右侧即生成一个新的 outfile,将其重命名为

rhizosphere genus (附件3),意味着是用于分析根围效应的属水平数据。

7、接下来选择 Statistic image mark statistic result in a manner of heatmap style,

将上步的数据转换成更加直观的图形,如下图所示: Statistic result with

significant marker 框架中选择上一步生成的 rhizosphere genus 文件,其他设置不

变,点击 execute,右侧生成名为 statistic heatmap 的图形文件:

Statistic image mark statistic result in a manner of heatmap style. (Galaxy Version 0.1.0)	▼ Options
Statistic result with significant marker	
1174: RhzisphereGenus	-
Column number of Taxa rank(Species labelings)	
6	
How to labeling the species	
Only Last taxa included	-
✓ Execute	

statistic heatmap 图示:可根据该图直观的分析不同处理间在不同分类水平上的

相对丰度的差异。



8、选择 tentative AI analyzer 版块下的 One Way Optimization PCA analyses and

plot,其中 Taxonomic OTU reporter 框架中选择 new data frame 文件,同样可根

据分析需要改变 tax rank to test,点击 execute,右侧生成名为 PCA plot 的图形文

件。

OneWayOptimizationPCA analyses and plot (Galaxy Version 0.1.0)
Taxonomic OTU reporter
C 57: New Data frame
Total Taxonomic rank
7
Taxonomic rank to analyze
7
✓ Execute

9、NMMD based on a dist matrix 和 PCoA based on a dist matrix,这两个工具的结

果反映的都是微生物群落结构,以 NMMD 为例,基于的文件都是 bray Curtis

distance 文件,图形中对不同处理的标记方式有两种,一种是如左图所示的直接

选择数据,另一种是如右图所示的选择自定义的文件,关于该文件所要求的标

记方式如下图小字所示,规定了形状和颜色所代表的数字,根据该要求,在

TXT 文件中设置三列,第一列为样品名称,第二列为形状代码,第三列为颜色

代码,上传至平台, sampling labeling pch or cols 下选择该文件。

	NMMD based on a dist matrix (Galaxy Version 0.1.0)
NMMD based on a dist matrix (Galaxy Version 0.1.0)	Pairwise distance
Pairwise distance	1818: rhizosphere Bray Curtis Distance
1818: rhizosphere Bray Curtis Distance	Points Decoration
Points Decoration data	file Sample labeling pch or cols 1821: different period rhizosphere soil NMMD.txt Execute
sample labels file format: sampleid pchs(r pchs:21->Circular;22->Square;23->Diamond;24->Up TriAng green; 9->green;13->cyan;16->blue;23->Purple;	number) cols(number) le;25->Down TriAngle

10、还有一个工具是 PLSPM analyses and plot,在 Taxonomic OTU reporter 框架中

选择 new data frame 文件,其他设置如下图所示,右侧会生成两个文件,一个

名为 PLSPM Plot(图形文件),另一个名为 VIP variables(数据文件),此步目

的是寻找关键物种。

LSPM analyses and plot (Galaxy Version 0.1.0)
axonomic OTU reporter
🗅 🛍 🗅 57: New Data frame
otal Taxonomic rank
7
axonomic rank to analyze
7
Scaled or not
TRUE
lode of PLSPM
regression
actor Source
data
✓ Execute

11、接下来是根据序列文件制作热图的方法(全部数据的热图绘制)

选择 meta data manager 版块下的 Select Fasta sequences from a file, fasta

sequences 框架中选择 rep_set.fna 文件, seq ids 框架中选择上步中生成的 VIP

variables 文件,此步目的是筛选出潜在物种的序列。

Select Fasta sequences from a file (Galaxy Version 0.1.0)	
Fasta sequences	
1205: rep_set.fna	
Seq ids	
1178: VIP variables	
✓ Execute	

点击 execute,右侧生成名为 selected sequence 的文件,选择 meta diversity

analyzer 版块下的 BLAST analyses locally,如下图所示,query框架中选择

selected sequence 文件,右侧生成名为 BLAST result 的文件,

BLAST analyses locally (Galaxy Version 0.1.0)	▼ Options
blast programe	
BLASTN	•
Database selected	
RDPISO16S	•
query	
1206: Selected Sequences	•
✓ Execute	

BLAST 结果如下图,第一列为 OTU 代码,第二列为与已知物种的相似度。

1	2
509402	92.2% Thiobacter_subterraneus(AB180657)
1084045	100.0% Bacillus_niacini(AB021194)
New.ReferenceOTU4053	89.0% Zavarzinella_formosa(AM162406)
553648	99.7% Pseudomonas_brassicacearum(EU3913
1128107	87.0% Syntrophus_gentianae(X85132)
252249	87.2% Desulfonema_magnum(U45989)
577182	89.0% Blastopirellula_marina(X62912)
564412	93.6% Thauera_linaloolentis(AJ005816)
688327	86.7% Gemmatimonas_aurantiaca(AB072735)
4332123	84.3% Megasphaera_sueciensis(DQ223729)
808847	84.9% Aciditerrimonas_ferrireducens(AB51766
New.ReferenceOTU1252	94.5% Ignavibacterium_album(AB478415)
954381	99.1% Bacillus_infernus(U20385)
563358	80.3% Roseiflexus_castenholzii(CP000804)
142371	87.4% Longilinea_arvoryzae(AB243673)
805777	82.7% Thermacetogenium_phaeum(AB020336
New.ReferenceOTU25720	89.2% Geminicoccus_roseus(AM403172)
New.ReferenceOTU14364	88.2% Bellilinea_caldifistulae(AB243672)
539866	89.0% Ectothiorhodospira_salini(FM244738)
4430429	93.3% Ohtaekwangia_kribbensis(GU117703)
New ReferenceOTI125722	89.8% Cystobacter armeniaca(DO768107)

接下来选择 Select Taxonomic groups from the data frame,如下图所示 Taxonomic

OTU report 框架中选择 new data frame 文件, list of taxs 框架中选择 VIP variables

文件,此步目的是从所有数据中筛选出关键 OTU 的数据,右侧生成名为 data

frame selected taxonomic out reporter 的文件。

Select Taxonom	ic groups from the data frame (Galaxy Version 0.1.0)
Taxonomic OTU	report
0 2 0	57: New Data frame
Tax Rank	
7	C
Total Tax Rank	
7	C
Data type	
File	
List of Taxs	
	1178: VIP variables
Seqs Numbers o	or Relative abundance
Relative abundar	nce
✓ Execute	

接下来选择 Heatmap analysis of selected taxa(species),如下图所示, data frame

for building heatmap 框架中选择上一步生成的 data frame selected taxonomic out

reporter (编号 x)文件, OTU labels: first column OTU id; second column: OTU

description 框架中选择 BLAST result 文件(编号 y),右侧生成名为 heatmap on

data x and data y 的热图。

ata type	
L6S or Similar	
Data Source	
Data frame f	or building heatmap
	1209: Data frame Selected Taxonomic OTU reporter
Total Taxono	mic rank
1	
OTU labels: firs	t column OTU id;second column:OTU description
	1208: BLAST Result
Factor Source	
data	
	 A second s
	<figure></figure>

12、接下来是制作分组数据的热图的方法(分组数据的热图绘制)

选择 PLSPM analyses and plot,如下图所示,Taxonomic OTU reporter 中选择分组

数据,右侧会生成分组数据的 PLSPM Plot (图形文件)和 VIP variables (数据文

件),此处将 VIP variables 命名为 rhizosphere VIP variables。

PLSPM analyses and plot (Galaxy Version 0.1.0)		
Taxonomic OTU reporter		
	1814: rhizosphere Taxonomic OTU report	
Total Taxonomic rank		
7	(
Taxonomic rank	to analyze	
7		
Scaled or not		
TRUE		
Mode of PLSPM		
regression		
Factor Source		
data		
✓ Execute		

接下来选择 Select Fasta sequences from a file , fasta sequences 框架中选择

rep_set.fna 文件, seq ids 框架中选择上步中生成的 rhizosphere VIP variables 文件,

此步目的是筛选出分组数据中可能的关键物种的序列。

Select Fasta sequences from a file (Galaxy Version 0.1.0)		
Fasta sequences		
D 4 D	1205: rep_set.fna	
Seq ids		
D 4 D	2013: rhizosphere VIP variables	
✓ Execute		

点击 execute,右侧生成名为 rhizosphere selected sequence 的文件,选择 meta

diversity analyzer 版块下的 BLAST analyses locally,如下图所示, query 框架中选

择 rhizosphere selected sequence 文件,右侧生成名为 rhizosphere BLAST result 的

文件,

BLAST analyses locally (Galaxy Version 0.1.0)	
blast programe	
BLASTN	
Database selected	
RDPISO16S	
query	
2014: rhizosphere Selected Sequences	
✓ Execute	

接下来选择 Select Taxonomic groups from the data frame,如下图所示 Taxonomic

OTU report 框架中选择分组数据文件,list of taxs 框架中选择 rhizosphere VIP

variables 文件,此步目的是从分组数据中筛选出关键 OTU 的数据,右侧生成名

为 data frame selected taxonomic out reporter 的文件。

Select Taxonomic groups from the data frame (Galaxy Version 0.1.0)	
Taxonomic OTU report	
C 4 C	1814: rhizosphere Taxonomic OTU report
Tax Rank	
7	(
Total Tax Rank	
7	C
Data type	
File	
List of Taxs	
	2013: rhizosphere VIP variables
Seqs Numbers o	or Relative abundance
Relative abundar	nce
✓ Execute	

接下来选择 Heatmap analysis of selected taxa(species),如下图所示, data frame

for building heatmap 框架中选择上一步生成的 rhizosphere data frame selected

taxonomic out reporter (编号 2017)文件, OTU labels: first column OTU id;

second column: OTU description 框架中选择 rhizosphere BLAST result 文件 (编号

2016),右侧生成名为 heatmap on data 2017 and data 2016 的热图,该热图反

应的就是分组数据的 OTU 相对丰度差异。

Heatmap analysis of selected taxa(species) (Galaxy Version 0.1.0)
Data type
16S or Similar
Data Source
Data frame for building heatmap
2017: rhizosphere Data frame Selected Taxonomic OTU reporter
Total Taxonomic rank
1
OTU labels: first column OTU id;second column:OTU description
🗋 省 🗅 2016: rhizosphere BLAST Result
Factor Source
data
✓ Execute

三、一些建议和备注:

全文中加粗、标红的字体和红色圆圈内的内容均需特别注意。 应尽量根据实验设计和实验目的对数据进行分组,有针对性的分 析数据,从而挖掘一些潜在的生物学信息。

文中的举例只是为了更加清晰明了的说明不同分析工具的功能和 操作步骤,大家应该基于自己的数据随机应变,更改相应设置。 本文只详述了目前本人所用到的一些分析工具,有一些工具并没 有提及,欢迎大家丰富扩展。