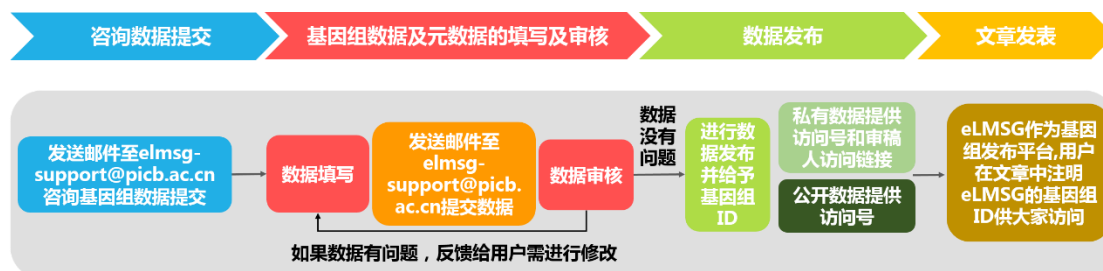


eLMSG 数据提交

微生物系统和基因组学图书馆 (An eLibrary of Microbial Systematics and Genomics, eLMSG, <https://www.biosino.org/elmsg/index>) 是一个提供微生物典型物种 (Type Species) 多相分类鉴定基础数据、基因组序列及高质量注释结果、种群核心与泛基因组信息的数据库平台, 该平台同时提供分离菌株和宏基因组拼接菌株的基因组数据提交。

目前平台提交的流程如下:



一、提交原始数据

目前 eLMSG 数据上传采用线下上传的方式进行, 您可以把您要提交的基因组序列直接邮件给我们 eLMSG 工作组 (邮箱: elmsg-support@picb.ac.cn, 电话: 17710710719)。

二、创建元数据

1、分离菌株的元数据创建

Item	Scientific Name	Strain Name	Type	CV	ANI	Topology	Create date	Genbank 16SrRNA	Etymology	Sample
Species 1	Allenteromonas intesti Z73T		negative		98 Thermaceae agagittgactcctgactct	linear	31-Dec-03	LC066536	en.o.shi.mer.seawater	

元数据的创建目前采用 excel 方式提交, 具体的填写表格参照“eLMSG_v5.0_purified_strain.xlsx”。表格第 14 行是我们 eLMSG 收集的字段, 包括物种基本信息 (Taxonomy information)、采样环境信息 (Ecology)、形态学和理化表型信息 (Morphology、Physiology、Biochemistry、Enzymology) 和基因组信息 (Genomics) 这几大块信息。如果没有测微生物形态学和理化表型信息, 填写物种基本信息、采样环境信息和基因组信息即可。

其中, 红色标记的字段 (R, mandatory fields) 为必填字段, 蓝色标记的字段 (RM, recommended mandatory fields) 表示推荐必填字段, 黄色标记的字段 (C, conditional mandatory fields) 为选填字段。把鼠标放在对应的字段上, 会有该字段的解释和示例数据供大家参考。一个菌株填一行。

General Information								
16SrRNA								
Item	Scientific Name	Strain Name	Type	ANI	Family Name	Sequence	Strandedness	Topology
if CV			CV					
Species 1	Allenteromonas intesti Z73T		negative	98	Thermaceae	agagtttgatcctgactc	single	linear

控制词表一

控制词表二

Species carbon and nitrogen sources enzyme Biome

有些字段有控制词表，我们在第 15 行增加了红色高亮加粗的 CV (controlled vocabulary)，控制词表里边的词少，直接点击填写的空格，可以从下拉菜单中选择（如“控制词表一”所示），当控制词表里的词比较多，下拉菜单无法解决的时候，我们的控制词表会单独列一个清单（如“控制词表二”所示），大家可以去清单找自己需要填写的值。大家在填写的时候可以直接从我们控制词表中选择对应的值，如果您的值不在我们的控制词表里边，请先空着并告知 eLMSG 工作组（邮箱：elmsg-support@picb.ac.cn，电话：17710710719）进行控制词表修改。

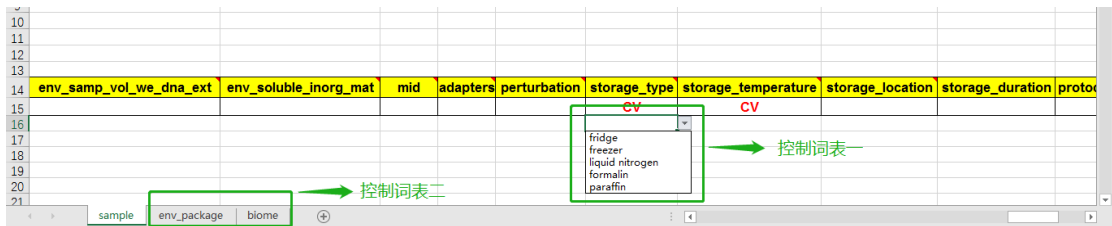
2、MAG (metagenome-assembled genome)的元数据创建

Item	Scientific Name	Strain Name	Type	ANI	Class	Order	Family	Genus	Species	Other note	Etymology	Submitter	Data source
Species 1													
Species 2													
Species 3													

推荐字段 必填字段 字段解释 选填字段

元数据的创建目前采用 excel 方式提交，具体的填写表格参照“eLMSG_v5.0_MAG.xlsx”和样本信息表“eLMSG_v5.0_MAG_Environment_Metagenomic.xlsx”。“eLMSG_v5.0_MAG.xlsx”表格第 14 行是我们 eLMSG 收集的字段，包括物种基本信息 (Taxonomy information)、基因组信息 (Genomics)、样本名称 (Sample name) 和参考文献 (References) 这几大块信息。其中，“eLMSG_v5.0_MAG.xlsx”表的“Sample name”与“eLMSG_v5.0_MAG_Environment_Metagenomic.xlsx”表的“Sample name”对应。

其中，红色标记的字段 (R, mandatory fields) 为必填字段，蓝色标记的字段 (RM, recommended mandatory fields) 表示推荐必填字段，黄色标记的字段 (C, conditional mandatory fields) 为选填字段。把鼠标放在对应的字段上，会有该字段的解释和示例数据供大家参考。一个菌株填一行。



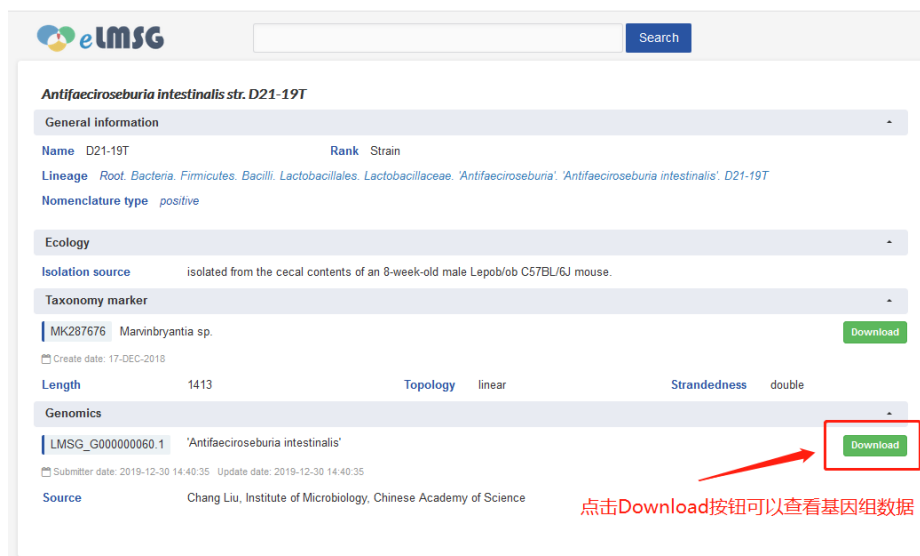
有些字段有控制词表，我们在第 15 行增加了红色高亮加粗的 CV (controlled vocabulary), 控制词表里边的词少，直接点击填写的空格，可以从下拉菜单中选择（如“控制词表一”所示），当控制词表里的词比较多，下拉菜单无法解决的时候，我们的控制词表会单独列一个清单（如“控制词表二”所示），大家可以去清单找自己需要填写的值。大家在填写的时候可以直接从我们控制词表中选择对应的值，如果您的值不在我们的控制词表里边，请先空着并告知 eMSG 工作组（邮箱：elmsg-support@picb.ac.cn，电话：17710710719）进行控制词表修改。

三、数据发布

在保证元数据都准确填写的情况下，我们会给出您提交的数据对应的基因组 ID 和数据链接。私有数据和公开数据会有两种不同的 ID 给出方式：

1、公开数据


直接给您连接，点击对应的“Download”连接查看基因组数据。例如 <https://www.biosino.org/elmsg/record/MSG050834>



2、私有数据

私有数据我们会给您两种链接模式，一种链接是正常的网页连接，您可以通过该连接查看您该基因组的元数据信息，但是基因组序列信息无法查看，例如 <https://www.biosino.org/elmsg/record/MSG051763>；另一种连接是审稿人的专用链接，通过该链接，审稿人可以对私有数据进行查看，点击对应的“Download”连接查看基因组数据。

General information	
Name	YSL
Rank	Strain
Lineage	Root, Archaea, Euryarchaeota, Methanobacteria, Methanobacteriales, Methanobacteriaceae, Methanobacterium, Methanobacterium sp., YSL
Nomenclature type	positive
Ecology	
Isolation source	sediment
Environment	root.Environmental/Aquatic/Freshwater/sediment
Geographic location	Jiehe River, Yantai
Collection date	2013/07
Environment biome	freshwater
Environment feature	harbor
Environment material	sediment
Latitude and longitude	37.08333-37.55000° N 120.13333-120.63333° E
Genomics	
LMSG_G000000181.1	Methanobacterium sp.
Submitter date:	2020/01/02
Source	Yantai Institute of Coastal Zone Research, Chinese Academy of Sciences
Assembly status	complete genome
Assembly class	haploid


Download
 私有数据基因组数据无法查看，需通过特殊链接查看

四、文章发表

在文章发表的时候，审稿人可以通过您文章里边的链接或者基因组 ID 号对您的数据进行审核。

文章中可选择的引用模式如下：

1、公开数据：

***** are available in eLMSG (an eLibrary of Microbial Systematics and Genomics, <https://www.biosino.org/elmsg/index>) under accession numbers *****.

2、私有数据：

1) 逐个菌株描述：

Strain ***** (eLMSG genome ID: LMSG_G*****) can be viewed in eLMSG (<https://www.biosino.org/elmsg/index>) through the URL: https://www.biosino.org/elmsg/record/*****

2) 多个菌株一起描述：

All data can be viewed in eLMSG (<https://www.biosino.org/elmsg/index>) through the URL in document "genome_link.xlsx". Column A refers to strain name. Column B indicates eLMSG genome ID. Column C refers to URL of genome in eLMSG.