蛋白组原始数据上传步骤

及因（上海）生物科技有限公司

Tgene Biotech (Shanghai) Co.,Ltd.

ProteomeXchange(简称 PX)，是应用较为广泛的蛋白质组学质谱数据存储平台（http://www.proteomexchange.org），其旗下包括了 PRIDE Archive, MassIVE, PeptideAtlas, jPOST，iProx，Panorama Public 质谱数据存储平台，研究人员可以通过 iProx 等网站进行数据上传，也可以通过 PX 客户端直接上传数据。接下来以 PX 为例来分享下蛋白组数据上传的具体步骤。

准备工作

1、PX 客户端工具下载（此处 v2.7.2 版本）

2、Java 下载安装

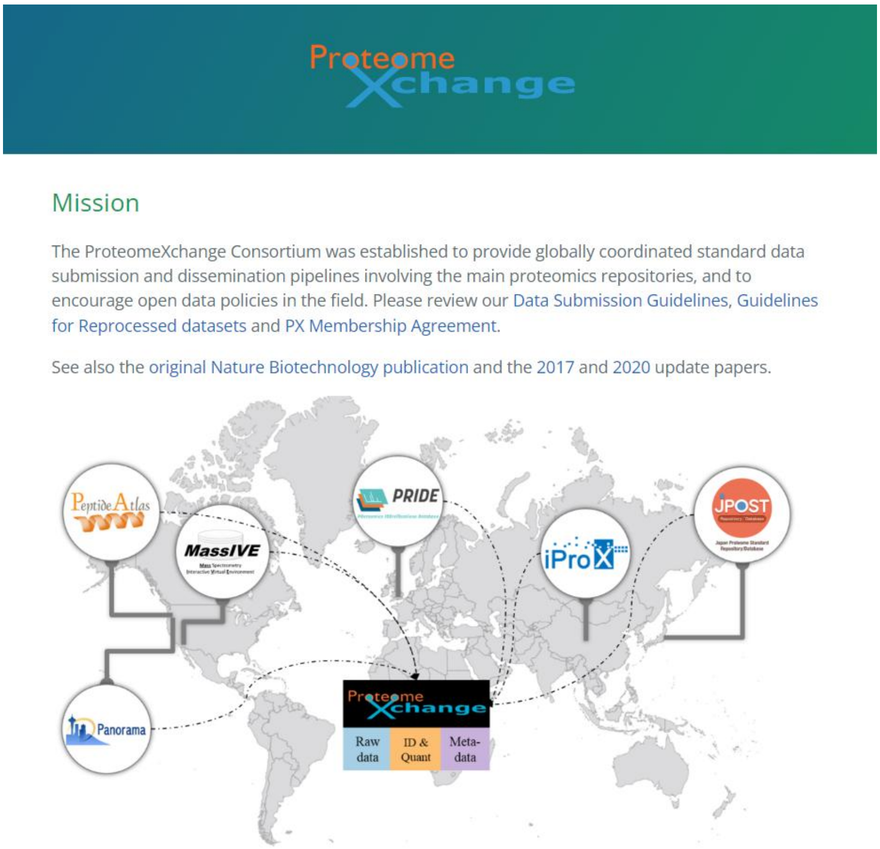
（https://www.oracle.com/java/technologies/downloads/ 自行登录网址下载安装，此处不再介绍）

3、PX 账号注册

4、提供相关文件（包括原始数据文件、搜库文库、上传信息单）

上传工具（PX）下载（已下载可忽略）

1、登入网址：http://www.proteomexchange.org，进入 Proteomexchange 网站界面。

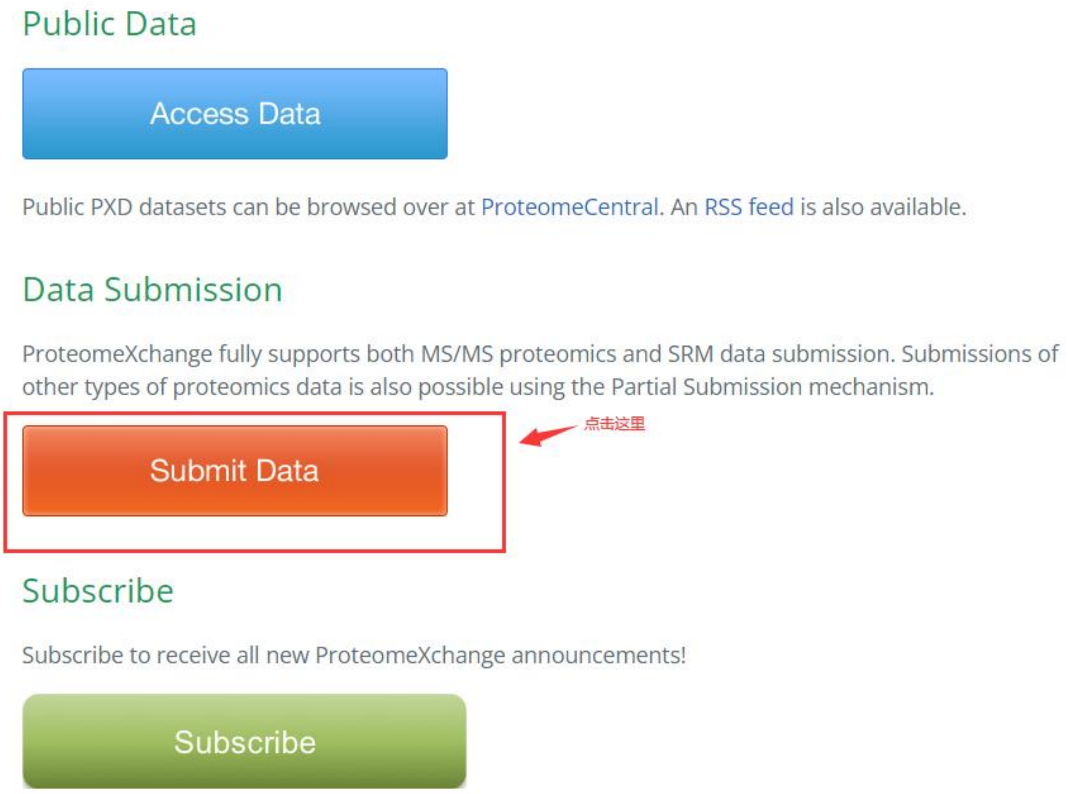


Mission

The ProteomeXchange Consortium was established to provide globally coordinated standard data submission and dissemination pipelines involving the main proteomics repositories, and to encourage open data policies in the field. Please review our Data Submission Guidelines, Guidelines for Reprocessed datasets and PX Membership Agreement.

See also the original Nature Biotechnology publication and the 2017 and 2020 update papers.

2、点击 Submit Data 进入数据上传页面。



Public Data

Public PXD datasets can be browsed over at ProteomeCentral. An RSS feed is also available.

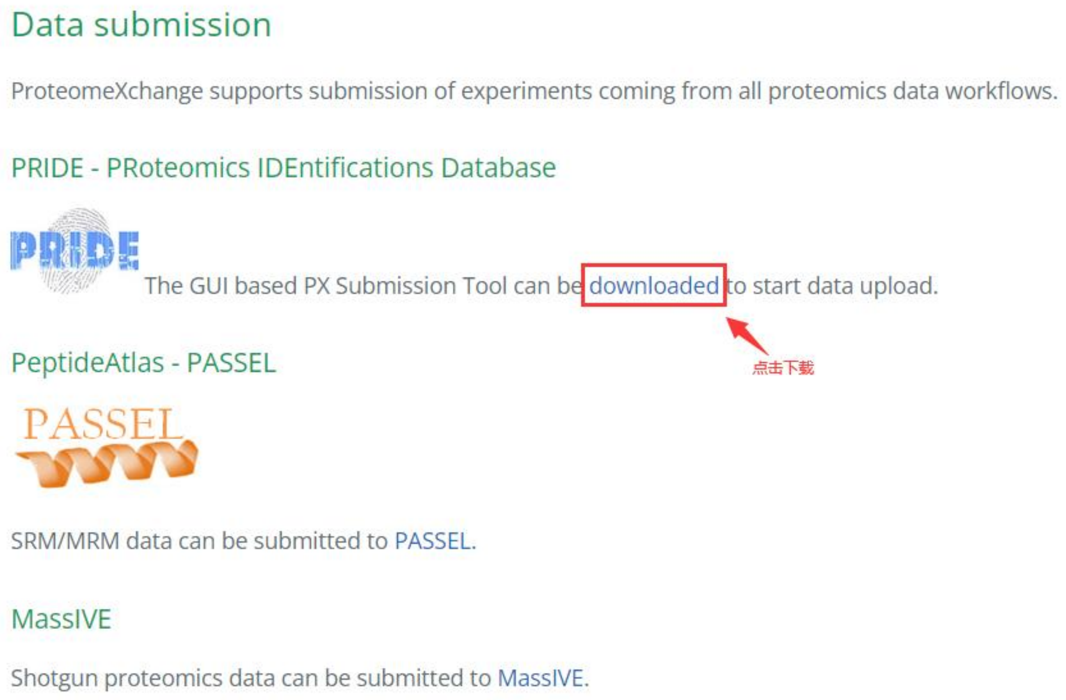
Data Submission

ProteomeXchange fully supports both MS/MS proteomics and SRM data submission. Submissions of other types of proteomics data is also possible using the Partial Submission mechanism.

Subscribe

Subscribe to receive all new ProteomeXchange announcements!

3、点击 downloaded 下载数据上传工具。



Data submission

ProteomeXchange supports submission of experiments coming from all proteomics data workflows.

PRIDE - PRoteomics IDEntifications Database

The GUI based PX Submission Tool can be downloaded to start data upload.

PeptideAtlas - PASSEL

SRM/MRM data can be submitted to PASSEL.

MassIVE

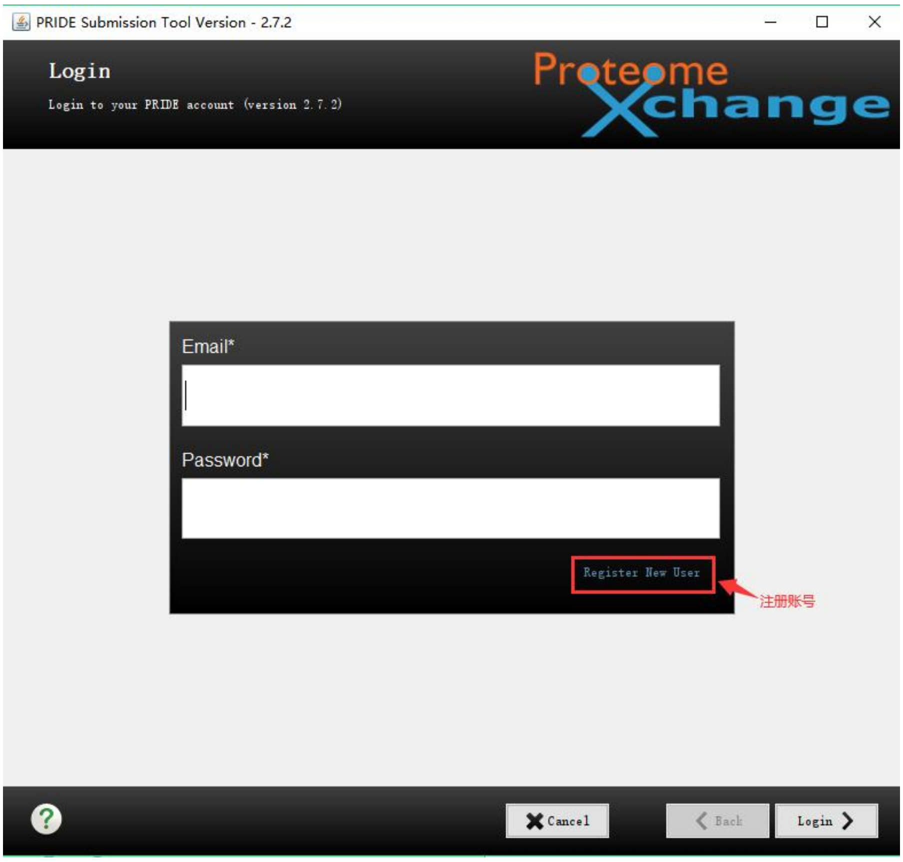
Shotgun proteomics data can be submitted to MassIVE.

4、下载后的数据上传工具是一个压缩包（px-submission-tool.zip），解压缩工具包，双击 px-submission-tool-2.7.2.jar 打开数据上传工具。

注：使用数据上传工具，电脑上必须安装 java（此处不再介绍）才可运行。

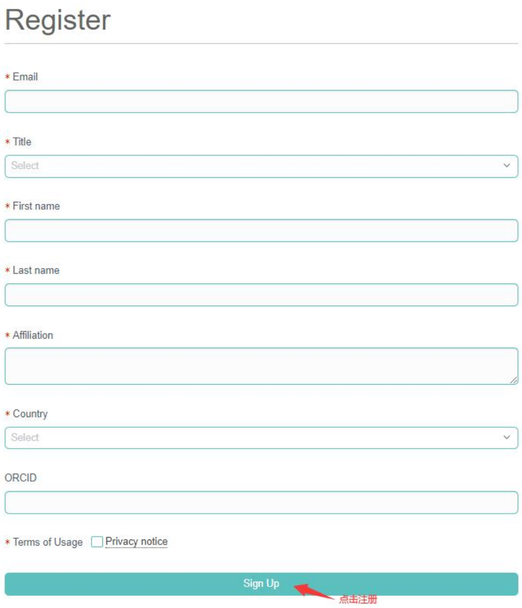
账号注册（已注册可忽略）

1、双击打开数据上传工具(px-submission-tool-2.7.2.jar)，点击 Register New Users。



注册账号

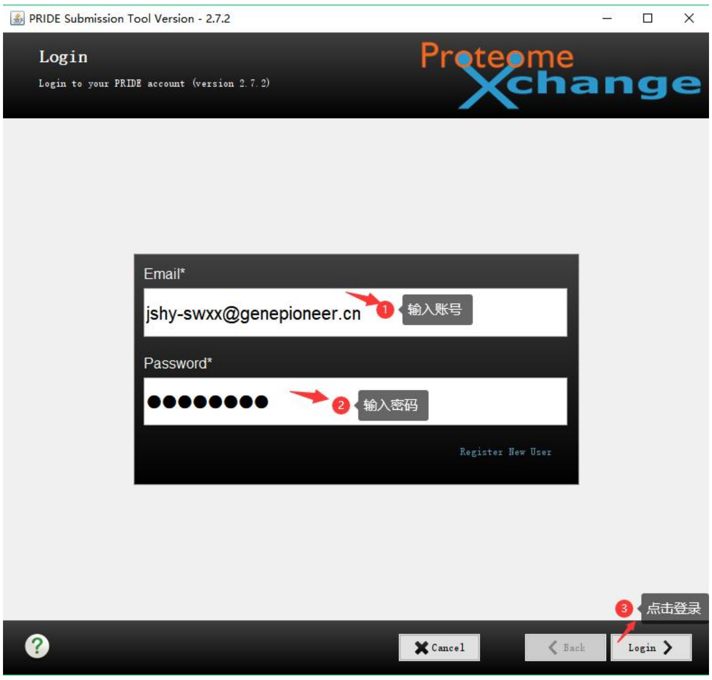
2、进入注册页面 https://www.ebi.ac.uk/pride/register，输入基本信息（电子邮箱、姓名、国家等）进行注册，带*为必填。注册完成后，账号、密码会发送到注册的邮箱中。



点击注册

账号登录

双击打开数据上传工具（px-submission-tool-2.7.2.jar），输入登录账号、密码，点击 Login 登录。



PRIDE Submission Tool Version - 2.7.2

Login

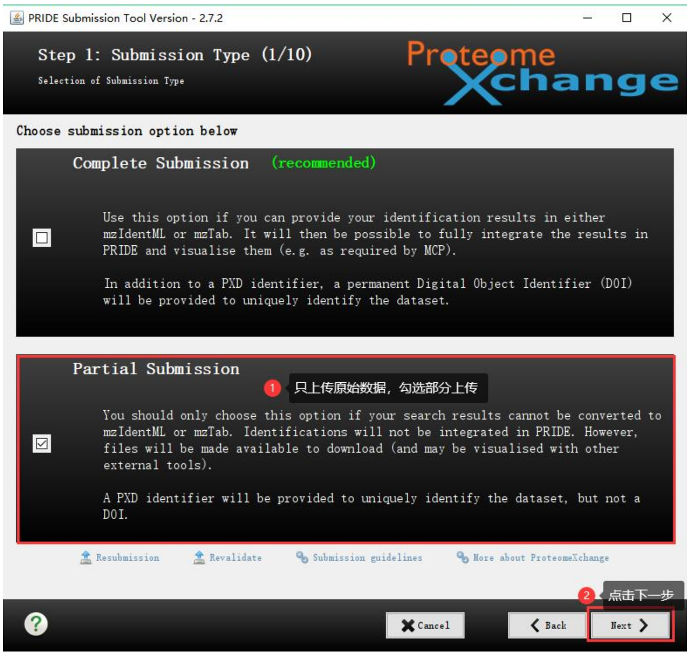
Login to your PRIDE account (version 2.7.2)

Email*: jshy-swxx@genepioneer.cn

Password*: ●●●●●●●●

上传步骤

1、登陆成功之后，首先需要选择上传模式，完全上传（Complete Submission）还是部分（Partial Submission）上传，这里我们只上传原始数据，所以选择 Partial Submission 上传。



PRIDE Submission Tool Version - 2.7.2

Step 1: Submission Type (1/10)

Selection of Submission Type

Choose submission option below

☐ Complete Submission (recommended)

Use this option if you can provide your identification results in either mzIdentML or mzTab. It will then be possible to fully integrate the results in PRIDE and visualise them (e.g. as required by MCP).

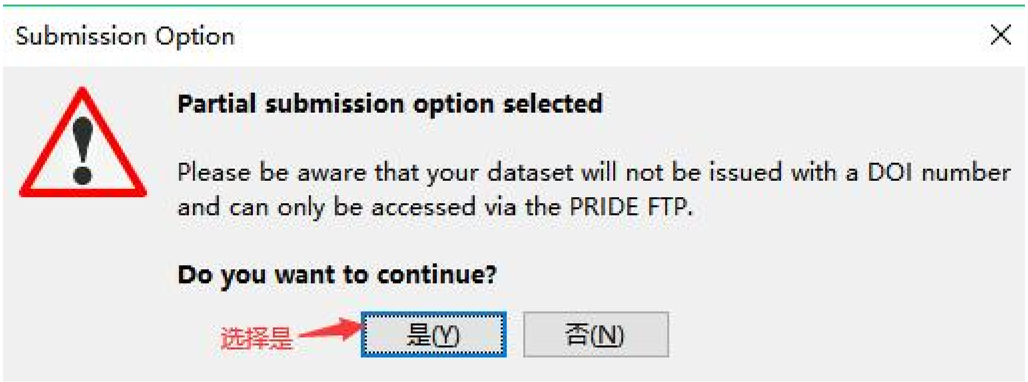
In addition to a PXD identifier, a permanent Digital Object Identifier (DOI) will be provided to uniquely identify the dataset.

☑ Partial Submission

You should only choose this option if your search results cannot be converted to mzIdentML or mzTab. Identifications will not be integrated in PRIDE. However, files will be made available to download (and may be visualised with other external tools).

A PXD identifier will be provided to uniquely identify the dataset, but not a DOI.

点击 Next，弹出对话框，选择是，跳转到下一页。



Submission Option

Partial submission option selected

Please be aware that your dataset will not be issued with a DOI number and can only be accessed via the PRIDE FTP.

Do you want to continue?

2、这一页提示我们需要准备以下信息才能完成数据的上传，无需任何操作，点击 NEXT 即可。

注：需要准备搜库结果文件、原始数据文件、登录账号、实验细节、项目负责人信息。

PRIDE Submission Tool Version - 2.7.2

Prerequisites

Before we begin, please check you have the following information

Search Engine Results
Protein/Peptide Identifications

Raw Data
MS instrument raw output

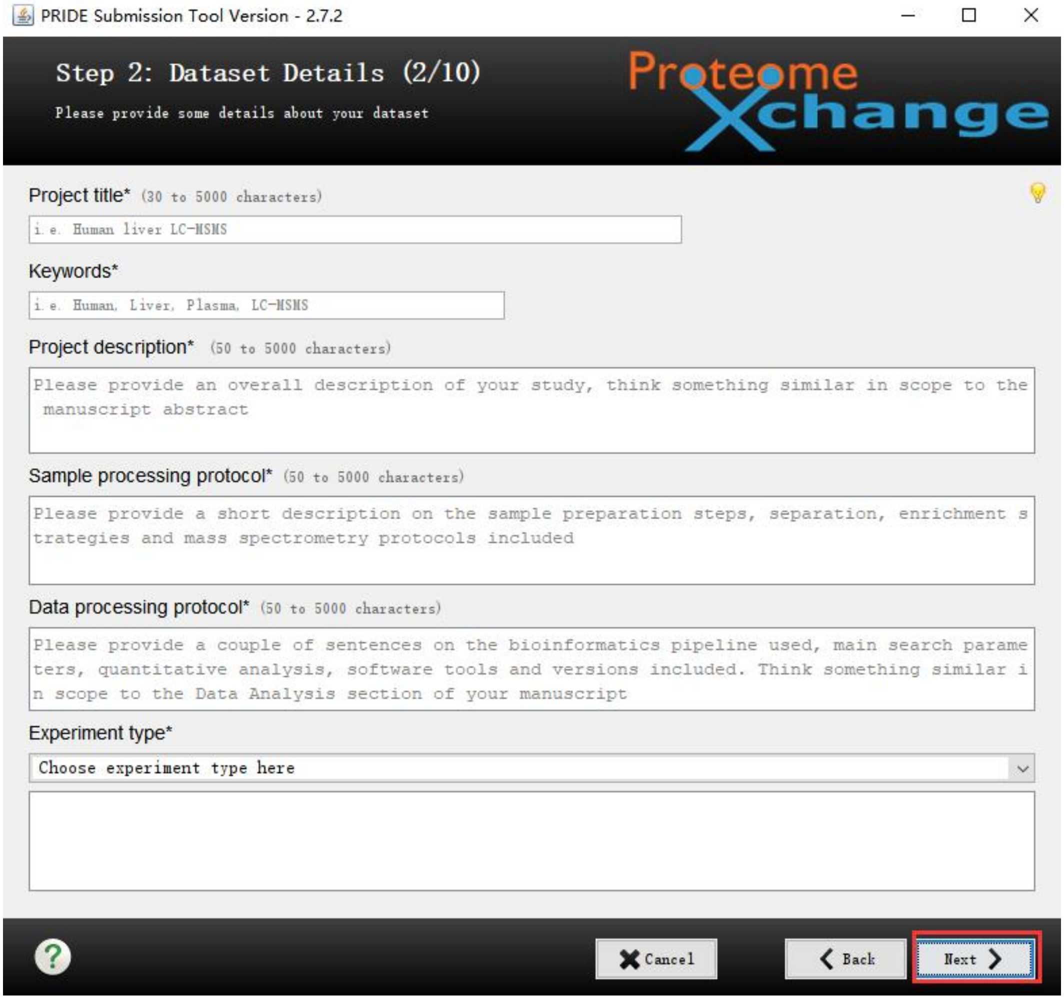
PRIDE Login
PRIDE user credentials Register

Experiment Details
Sample, protocols, species, tissues

Lab Head
Name, email, and affiliation

3、之后需要填写蛋白组数据的基本信息，包括发表文章的标题（30~5000 字符）、关键词、项目描述（50~5000 字符）、样品准备方法（50~5000 字符）、数据处理方法（50~5000 字符）和实验类型（具体如下图，根据实际情况勾选）。

注：可将文章的标题、关键词、摘要和材料与方法对应填入。需要注意字数限制，带*为必填项。



Project title* (30 to 5000 characters)

i.e. Human liver LC-MSMS

Keywords*

i.e. Human, Liver, Plasma, LC-MSMS

Project description* (50 to 5000 characters)

Please provide an overall description of your study, think something similar in scope to the manuscript abstract

Sample processing protocol* (50 to 5000 characters)

Please provide a short description on the sample preparation steps, separation, enrichment strategies and mass spectrometry protocols included

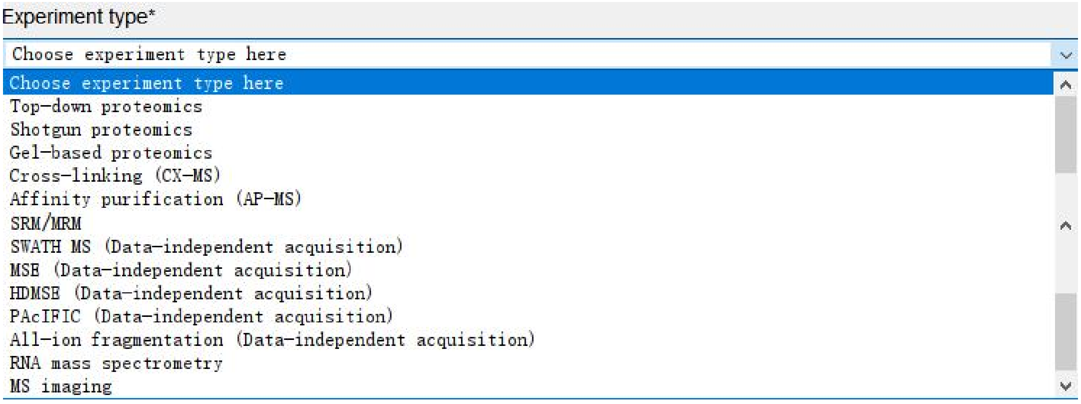
Data processing protocol* (50 to 5000 characters)

Please provide a couple of sentences on the bioinformatics pipeline used, main search parameters, quantitative analysis, software tools and versions included. Think something similar in scope to the Data Analysis section of your manuscript

Experiment type*

Choose experiment type here

具体实验类型如下：



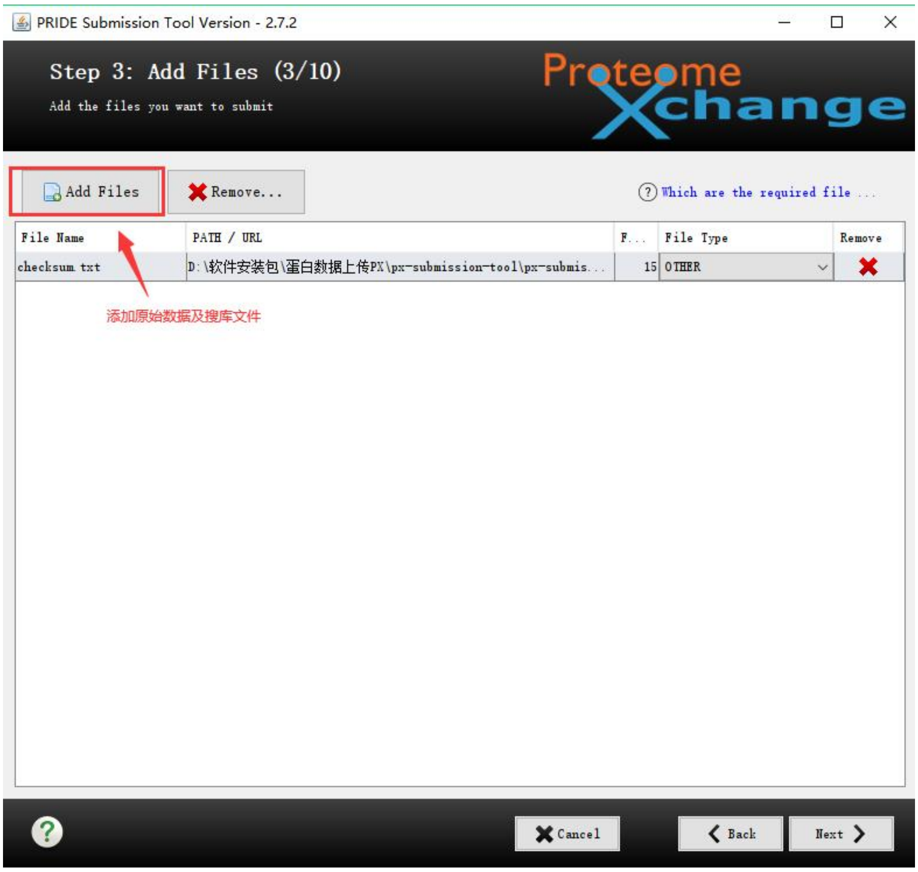
Experiment type*

Choose experiment type here

- Choose experiment type here
- Top-down proteomics
- Shotgun proteomics
- Gel-based proteomics
- Cross-linking (CX-MS)
- Affinity purification (AP-MS)
- SRM/MRM
- SWATH MS (Data-independent acquisition)
- MSE (Data-independent acquisition)
- HDMSE (Data-independent acquisition)
- PAcIFIC (Data-independent acquisition)
- All-ion fragmentation (Data-independent acquisition)
- RNA mass spectrometry
- MS imaging

4、之后添加需要提交的文件，选择.txt 搜库文件，并添加文件为 search 类型；选择原始数据（*.raw/*.scan/*.mzMXL/*.wiff 等后缀），并添加文件为 raw 类型。

注：上传工具下自带的 checksum.txt 文件不要删除，后面校验会出错。



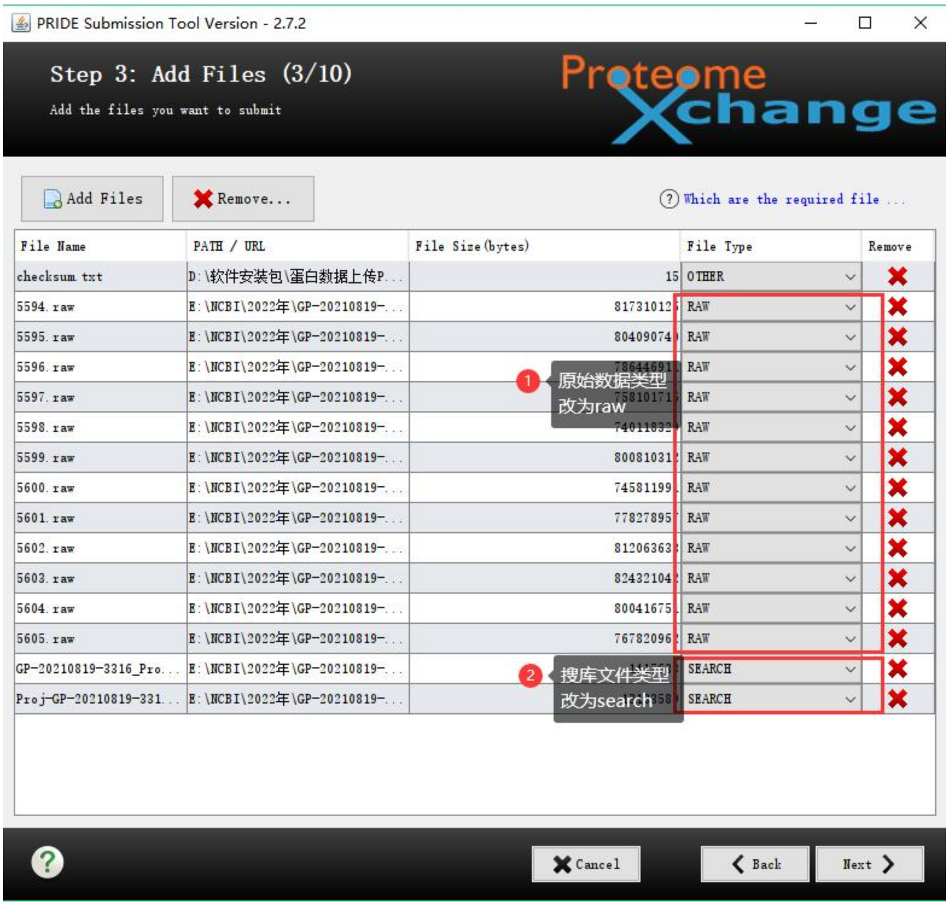
PRIDE Submission Tool Version - 2.7.2

Step 3: Add Files (3/10)

Add the files you want to submit

File Name	PATH / URL	F...	File Type	Remove
checksum txt	D:\软件安装包\蛋白数据上传PX\px-submission-tool\px-submis...	15	OTHER	✖

添加原始数据及搜库文件

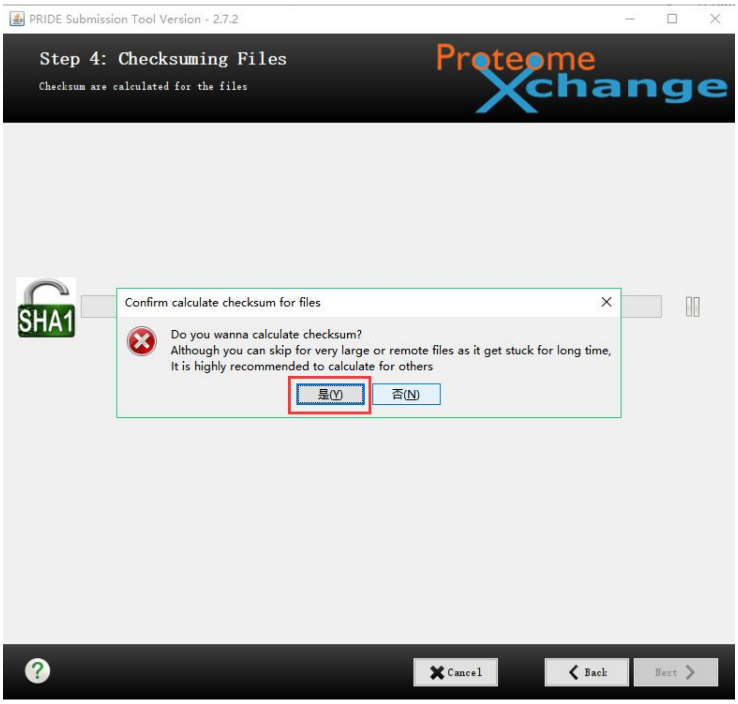


File Name	PATH / URL	File Size (bytes)	File Type	Remove
checksum.txt	D:\软件安装包\蛋白数据上传P...	15	OTHER	✖
5594.raw	E:\NCBI\2022年\GP-20210819-...	81731012	RAW	✖
5595.raw	E:\NCBI\2022年\GP-20210819-...	80409074	RAW	✖
5596.raw	E:\NCBI\2022年\GP-20210819-...	78644691	RAW	✖
5597.raw	E:\NCBI\2022年\GP-20210819-...	75810171	RAW	✖
5598.raw	E:\NCBI\2022年\GP-20210819-...	74011832	RAW	✖
5599.raw	E:\NCBI\2022年\GP-20210819-...	80081031	RAW	✖
5600.raw	E:\NCBI\2022年\GP-20210819-...	74581199	RAW	✖
5601.raw	E:\NCBI\2022年\GP-20210819-...	77827895	RAW	✖
5602.raw	E:\NCBI\2022年\GP-20210819-...	81206363	RAW	✖
5603.raw	E:\NCBI\2022年\GP-20210819-...	82432104	RAW	✖
5604.raw	E:\NCBI\2022年\GP-20210819-...	80041675	RAW	✖
5605.raw	E:\NCBI\2022年\GP-20210819-...	76782096	RAW	✖
GP-20210819-3316_Pro...	E:\NCBI\2022年\GP-20210819-...	[illegible]	SEARCH	✖
Proj-GP-20210819-331...	E:\NCBI\2022年\GP-20210819-...	[illegible]	SEARCH	✖

① 原始数据类型改为raw

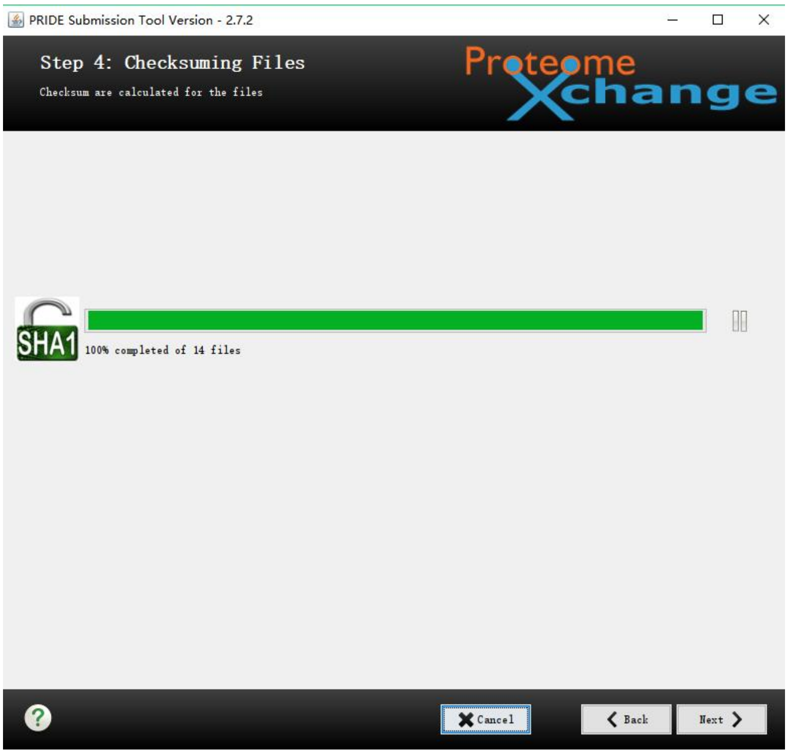
② 搜库文件类型改为search

5、所有文件添加之后，下一步，弹出对话框，选择“是”，系统进行校验，等待完成。



Do you wanna calculate checksum?
Although you can skip for very large or remote files as it get stuck for long time,
It is highly recommended to calculate for others

校验完成，点击“Next”，进入下一步。



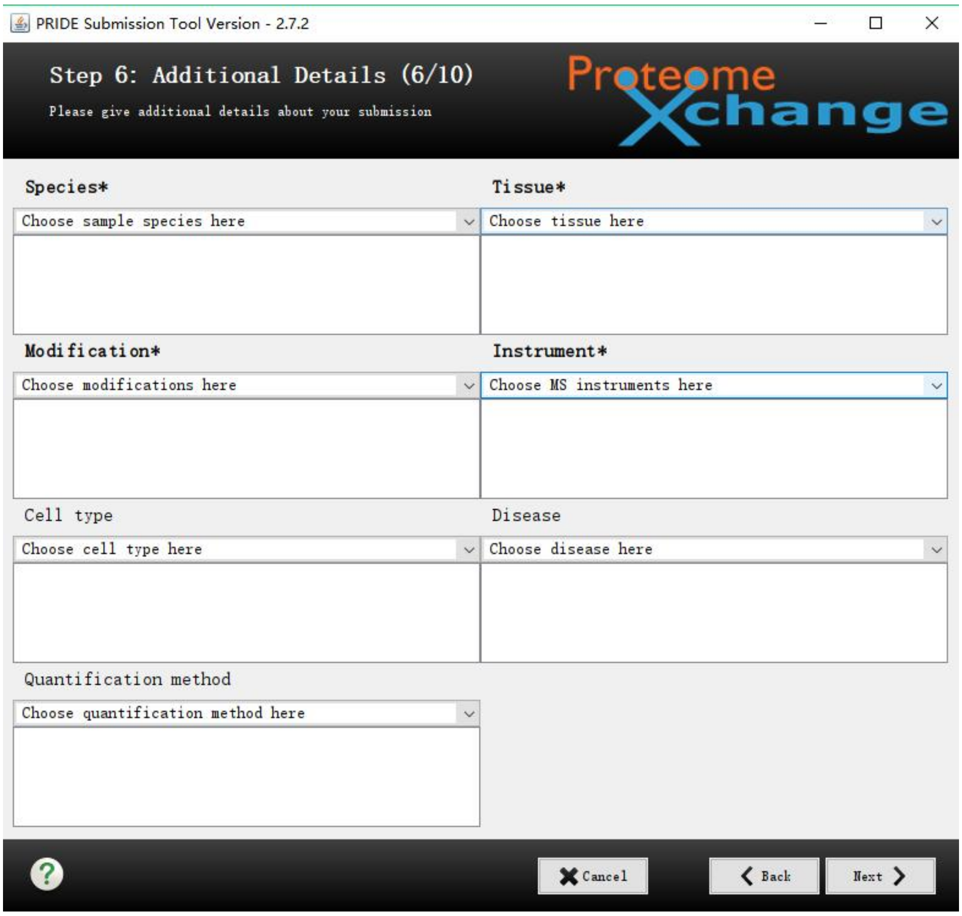
PRIDE Submission Tool Version - 2.7.2

Step 4: Checksuming Files

Checksum are calculated for the files

100% completed of 14 files

6、填入一些项目的具体信息，比如物种、样本类型、修饰类型、仪器型号等。



PRIDE Submission Tool Version - 2.7.2

Step 6: Additional Details (6/10)

Please give additional details about your submission

Species*: Choose sample species here

Tissue*: Choose tissue here

Modification*: Choose modifications here

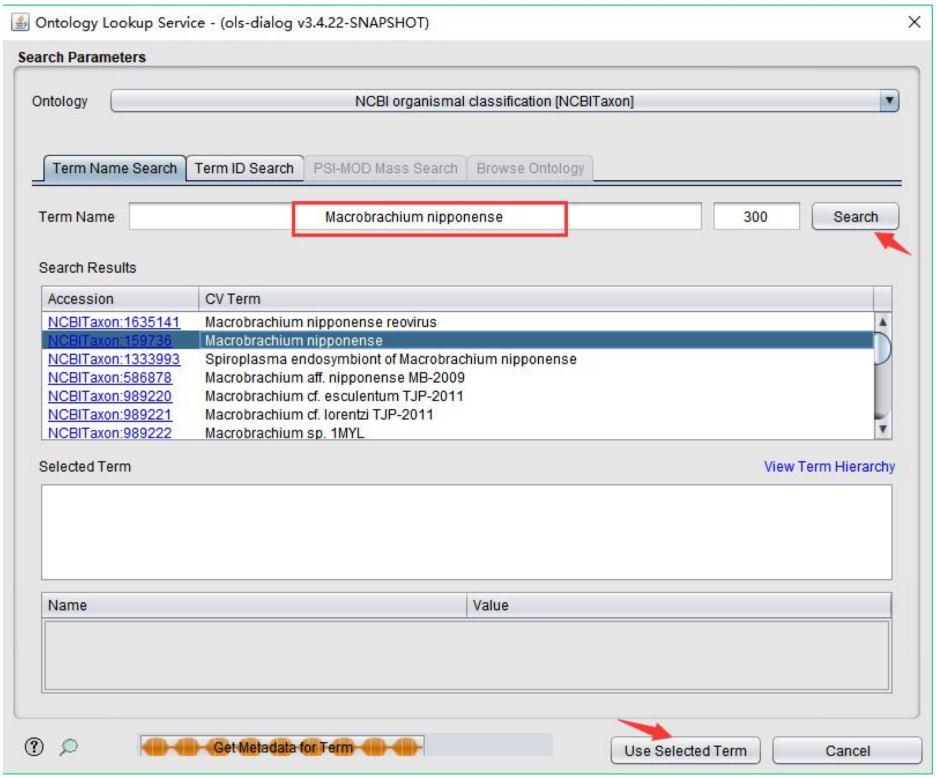
Instrument*: Choose MS instruments here

Cell type: Choose cell type here

Disease: Choose disease here

Quantification method: Choose quantification method here

如果需要填的物种信息在下拉列表里没有，则选择最下方的 Others，弹出对话框里搜索，找到结果后，选中，点击 Use Selected Term，注意搜索页面打开后会自行进行连接，需要等待几十秒钟才能进行搜索。

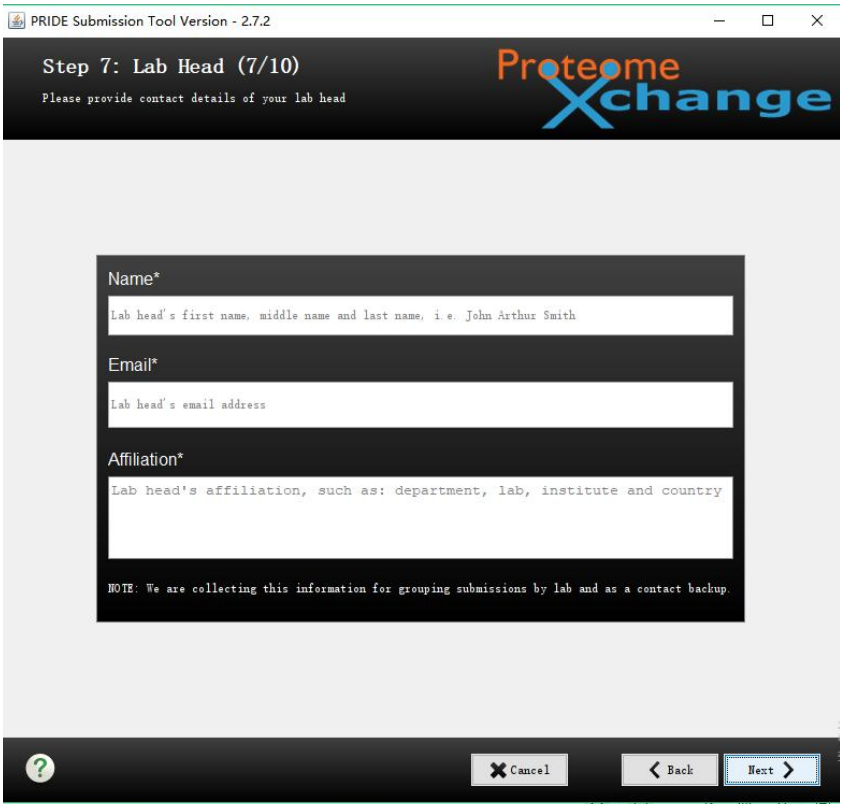


Ontology: NCBI organismal classification [NCBITaxon]

Term Name: Macrobrachium nipponense 300

Accession	CV Term
NCBITaxon:1635141	Macrobrachium nipponense reovirus
NCBITaxon:159736	Macrobrachium nipponense
NCBITaxon:1333993	Spiroplasma endosymbiont of Macrobrachium nipponense
NCBITaxon:586878	Macrobrachium aff. nipponense MB-2009
NCBITaxon:989220	Macrobrachium cf. esculentum TJP-2011
NCBITaxon:989221	Macrobrachium cf. lorentzi TJP-2011
NCBITaxon:989222	Macrobrachium sp. 1MYL

7、填入项目负责人的姓名、邮箱和单位地址，地址要包含国家。



Step 7: Lab Head (7/10)

Please provide contact details of your lab head

Name*

Lab head's first name, middle name and last name, i.e. John Arthur Smith

Email*

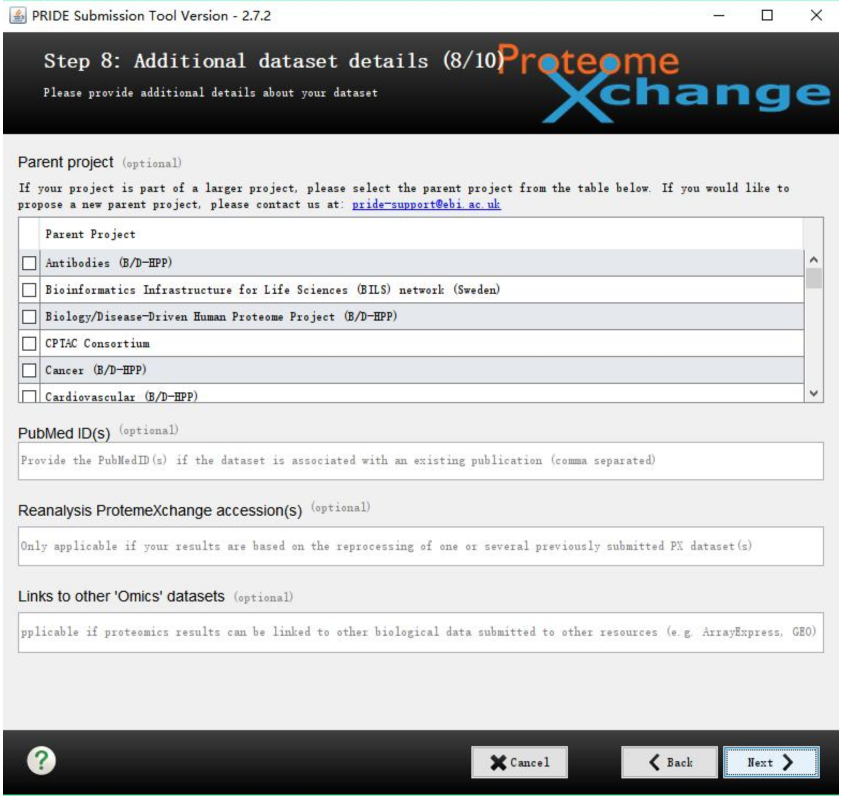
Lab head's email address

Affiliation*

Lab head's affiliation, such as: department, lab, institute and country

NOTE: We are collecting this information for grouping submissions by lab and as a contact backup.

8、填写母项目信息，如果上传数据不属于某一个大项目的子项目，则无需填写，直接 Next 下一步。



PRIDE Submission Tool Version - 2.7.2

Step 8: Additional dataset details (8/10)

Please provide additional details about your dataset

Parent project (optional)

If your project is part of a larger project, please select the parent project from the table below. If you would like to propose a new parent project, please contact us at: pride-support@ebi.ac.uk

	Parent Project
☐	Antibodies (B/D-HPP)
☐	Bioinformatics Infrastructure for Life Sciences (BILS) network (Sweden)
☐	Biology/Disease-Driven Human Proteome Project (B/D-HPP)
☐	CPTAC Consortium
☐	Cancer (B/D-HPP)
☐	Cardiovascular (B/D-HPP)

PubMed ID(s) (optional)

Provide the PubMedID(s) if the dataset is associated with an existing publication (comma separated)

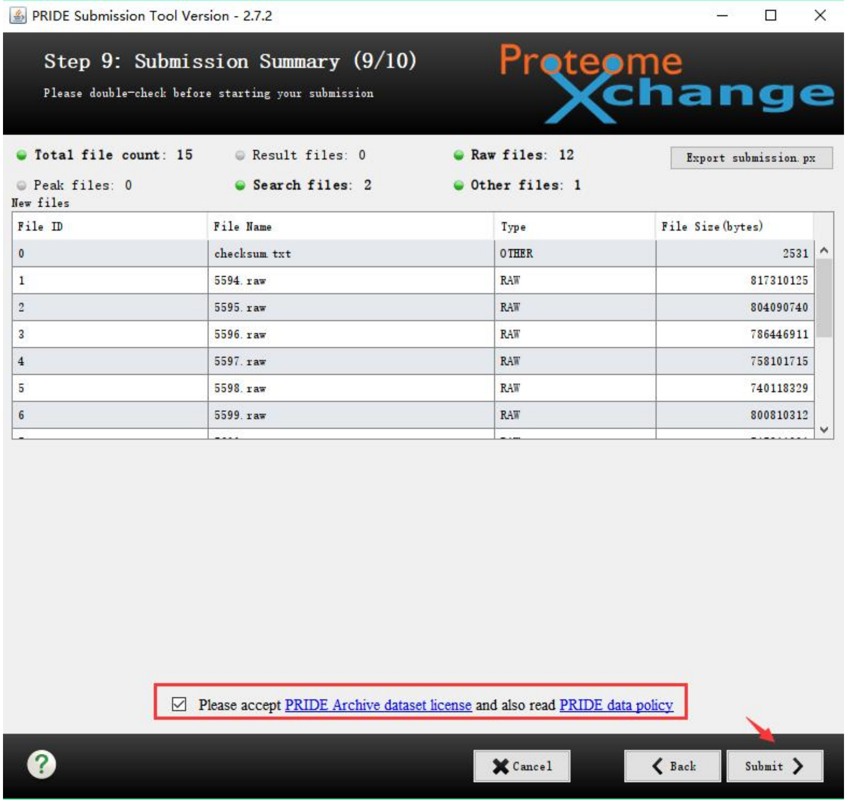
Reanalysis ProtemeXchange accession(s) (optional)

Only applicable if your results are based on the reprocessing of one or several previously submitted PX dataset(s)

Links to other 'Omics' datasets (optional)

pplicable if proteomics results can be linked to other biological data submitted to other resources (e.g. ArrayExpress, GEO)

9、确认文件信息是否有误，确认无误勾选最下方的 Please accept ……，点击 Submit 提交，弹出保存对话框，选择“保存”，进入数据上传界面。



PRIDE Submission Tool Version - 2.7.2

Step 9: Submission Summary (9/10)

Please double-check before starting your submission

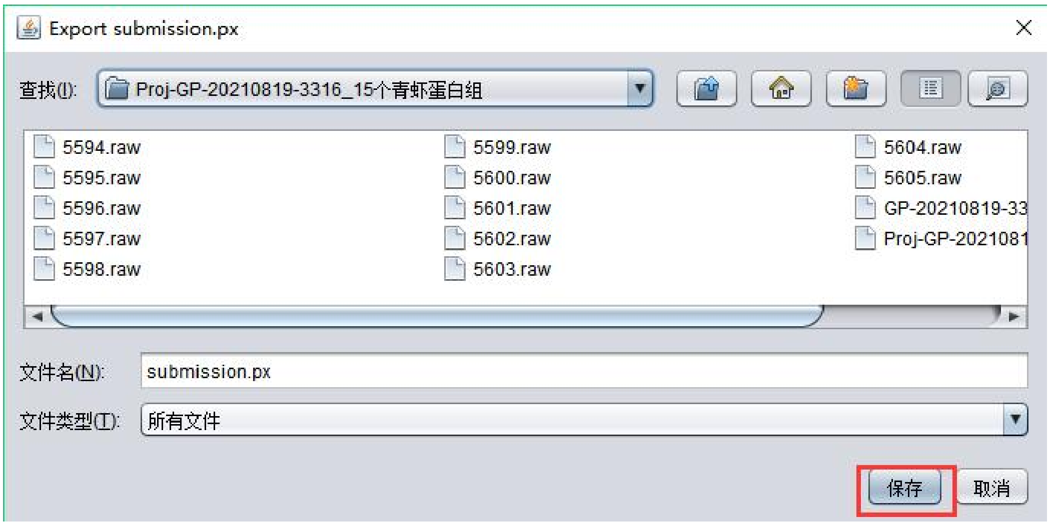
● **Total file count: 15** ○ Result files: 0 ● **Raw files: 12**

○ Peak files: 0 ● **Search files: 2** ● **Other files: 1**

New files

File ID	File Name	Type	File Size (bytes)
0	checksum.txt	OTHER	2531
1	5594.raw	RAW	817310125
2	5595.raw	RAW	804090740
3	5596.raw	RAW	786446911
4	5597.raw	RAW	758101715
5	5598.raw	RAW	740118329
6	5599.raw	RAW	800810312

☑ Please accept PRIDE Archive dataset license and also read PRIDE data policy



Export submission.px

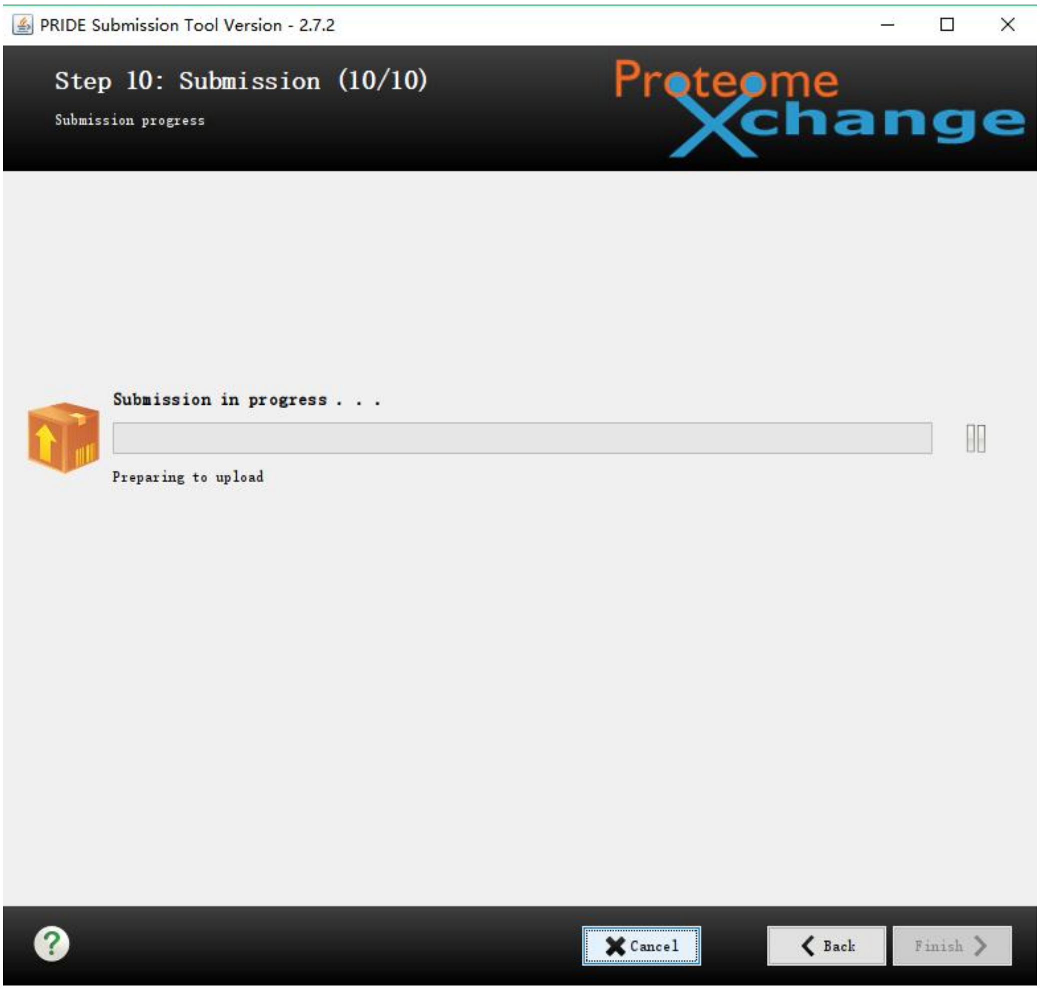
查找(I): Proj-GP-20210819-3316_15个青虾蛋白组

5594.raw
5595.raw
5596.raw
5597.raw
5598.raw
5599.raw
5600.raw
5601.raw
5602.raw
5603.raw
5604.raw
5605.raw
GP-20210819-33
Proj-GP-2021081

文件名(N): submission.px

文件类型(T): 所有文件

10、显示 Submission in progress 后耐心等待数据提交。一般上传过程中避免网络波动，如果出现报错，可以尝试重试，或者关闭防火墙以及杀毒软件等程序，再进行上传。



PRIDE Submission Tool Version - 2.7.2

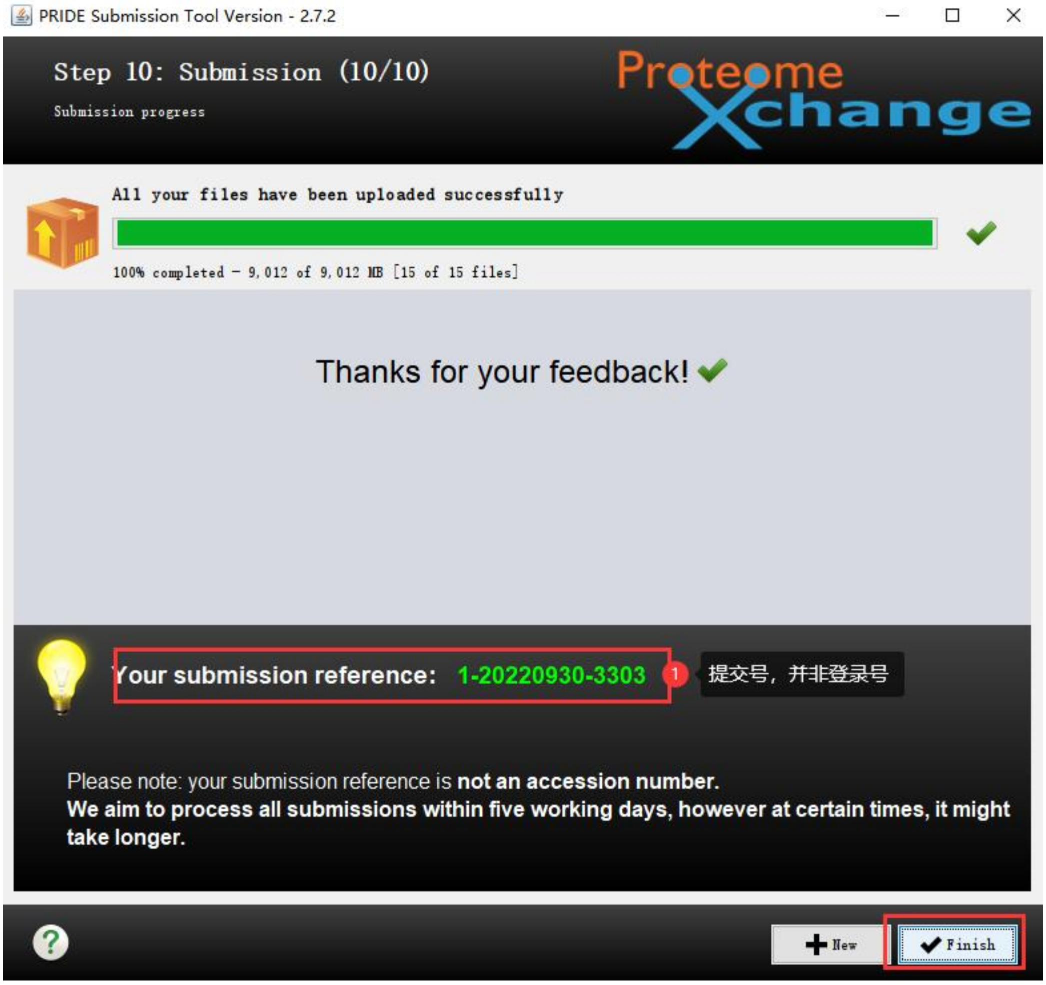
Step 10: Submission (10/10)

Submission progress

Submission in progress . . .

Preparing to upload

11、上传数据成功后，系统会提供一个序列号，但是这个序列号是无法在文章中使用的，且将进入数据审核状态。



Step 10: Submission (10/10)

Submission progress

All your files have been uploaded successfully

100% completed - 9,012 of 9,012 MB [15 of 15 files]

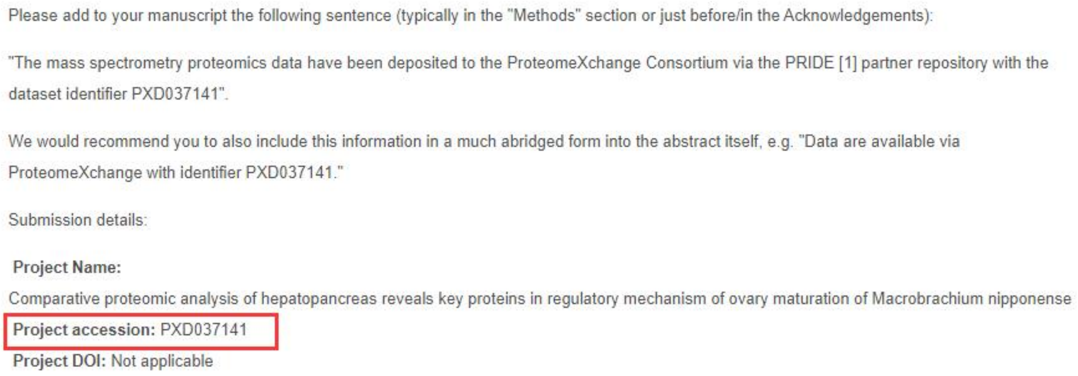
Thanks for your feedback!

Your submission reference: 1-20220930-3303

① 提交号，并非登录号

Please note: your submission reference is **not an accession number.**
We aim to process all submissions within five working days, however at certain times, it might take longer.

12、待后台人工审核完成会收到一封如下邮件，包含项目名称、登录号等，在文章中提供该邮件中提供的 Project accession 号就可以了。



Please add to your manuscript the following sentence (typically in the "Methods" section or just before/in the Acknowledgements):

"The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [1] partner repository with the dataset identifier PXD037141".

We would recommend you to also include this information in a much abridged form into the abstract itself, e.g. "Data are available via ProteomeXchange with identifier PXD037141."

Submission details:

Project Name:
Comparative proteomic analysis of hepatopancreas reveals key proteins in regulatory mechanism of ovary maturation of Macrobrachium nipponense
Project accession: PXD037141
Project DOI: Not applicable