INSACOG data submission guidelines at IBDC

Version 1.2

IBDC has developed a separate track for the submission of Covid-19 sequencing data at the **Indian Nucleotide Data Archive - Controlled Access (INDA-CA)**.

STEP1: User registration

In order to submit COVID19 data sets, the user must register at INDA-CA(https://inda.rcb.ac.in/indasecure).

1. Click on the 'Register' button on the top panel on the INDA-CA homepage.
2. Provide all the relevant details.
3. All new user accounts needs to be activated after verification by IBDC staff. An email would be sent confirming the account activation at the registered email.

STEP 2: Select data submission track

1. Visit IBDC website (www.ibdc.rcb.res.in) and click on **'Submit Data'** tab *or* directly go to **INDA-CA** home page (skip to step 3).
2. Select **'INDA-CA'** on the next page.
3. **Select 'Template-based submission (INSACOG)' track within the box** .

   Important points:

   IBDC provides a special track for submission of the INSACOG COVID 19 sequencing data:

   a. **Template-based submission:** Metadata is provided in an 'MS Excel' file and sequence data is uploaded via FTP.
4. Use login credentials provided by IBDC staff to login via any of the above submission tracks.

STEP 3: Submit data using the 'Template-based submission' track

1. Log in to the **'Template Based submission' track** using the login credentials received from IBDC.
2. Download the **template 'MS Excel'** by clicking on the **'MetaData_INSACOG_template'** link provided on the dashboard (Left panel). Fill in the metadata for multiple samples in the template file.
3. **Register 'Study':** Register a study by providing relevant details.
4. **Upload Metadata:** Click the link **"Upload metadata"** on the left panel of the dashboard.
   a. Select an appropriate 'Study' to which the data has to be submitted.
   b. Subsequently, select the appropriate 'metadata file'.
   c. Click on the **'Upload and Save button'**.

If fields of the metadata file are as per format a success message will be displayed on the page.
If any discrepancy is found an appropriate message will be displayed and the user has to correct the file and re-upload.

Important points:

- *The metadata template downloaded should be filled in according to the headers.*
- *Some fields are mandatory which have to be filled and it is advised not to edit the first two lines of the metadata template.*

- *Following fields are mandatory:* ***'Virus name', 'Collection Date','State','Sequencing Technology', 'Sample id given by the sample provider', 'Filename1' ,'md5sum1' and 'Strategy'.***
    - ***Virus name** should be unique, duplicate entries will not be taken.*
    - *Please enter the date in a uniform format of* ***"dd/mm/yyyy" or "dd-mm-yyyy"*** *in* ***Collection date and Last vaccinated fields*** *.*
    - ***Filename1** (Raw data filename) should be given with the extension for example "Filename.fastq.gz".*
    - *Please do not enter text or any other characters in date fields.*
    - *Please check for special character/tab/space before the dates in Collection date or in Last vaccinated column.*
- *The filename given in the metadata details* ***should be exactly the same*** *as the file which has to be uploaded in the next step.*
- *A valid* ***Md5sum*** *value for each file has to be given.*
- ***Strategy** field should be given either as **single** or **paired**. Values apart from this will not be considered.*

5. **Raw Data Upload:** Once the metadata has been uploaded successfully user can use the FTP details after clicking on the **'Raw Data upload'** tab on the left panel on the dashboard. It is recommended to create a new folder in the **FTP** directory and upload the **fastq files** in the created folder. The filename should be the same as given in the metadata file uploaded.

6. **Map Uploaded files**

- In this step user has to map the uploaded 'raw data files' with the 'metadata'.
- User has to select the folder of the uploaded files and click on the button **"Click here to Map All Uploaded Files'**. All the files with matching metadata will be mapped, a count of mapped and not mapped files will be shown on the page. IBDC accession will be provided when the files are mapped correctly for each uploaded sample.

Important points:

- *The files will be mapped only if the filename given in the metadata matches the uploaded filenames.*
- *INDA-CA will verify the 'md5sum' value provided in the meta-data file with that of the data file. If they don't match user will be notified via email and the correct file has to be re-uploaded in the same folder.*

Support: For any queries please write to us at 'indasupport@ibdc.rcb.res.in'