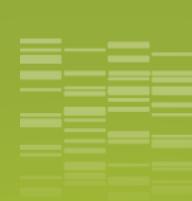


# Déposer et récupérer des données de séquençages depuis SRA (Sequence Read Archive)

# Dépôt des données sur SRA

PCR with 12 cycles. The resulting PCR products were purified and loaded onto the Illumina MiSeq cartridge (Illumina, San Diego, CA, USA) according to the manufacturer's instructions. Each pair-end sequence was assigned to its sample with the help of the previously integrated index. Sequencing reads were deposited in the National Center for Biotechnology Information Sequence Read Archive (NCBI SRA; SRP139897).

Metagenomique 16S MiSeq



<https://www.ncbi.nlm.nih.gov/>

NCBI Resources How To

All Databases Search

NCBI National Center for Biotechnology Information

NCBI Home Resource List (A-Z)

All Resources Chemicals & Bioassays Data & Software DNA & RNA Domains & Structures Genes & Expression Genetics & Medicine Genomes & Maps Homology Literature Proteins Sequence Analysis Taxonomy Training & Tutorials Variation

Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

[About the NCBI](#) | [Mission](#) | [Organization](#) | [NCBI News & Blog](#)

**Submit** Deposit data or manuscripts into NCBI databases 

**Download** Transfer NCBI data to your computer 

**Learn** Find help documents, attend a class or watch a tutorial 

**Develop** Use NCBI APIs and code libraries to build applications 

**Analyze** Identify an NCBI tool for your data analysis task 

**Research** Explore NCBI research and collaborative projects 

**Popular Resources**

PubMed  
Bookshelf  
PubMed Central  
BLAST  
Nucleotide  
Genome  
SNP  
Gene  
Protein  
PubChem

**NCBI News & Blog**

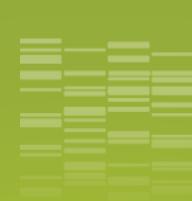
Prokaryotic Genome Annotation Pipeline (PGAP) now produces results suitable for submission to GenBank 13 May 2019  
We are happy to announce that you can

May 15, 2019 Webinar: Using taxonomic information and other improvements in standalone BLAST+ (2.9.0) and the v5 databases 08 May 2019  
Computational Medicine in the Cloud Hackathon: June 10-11, 2019 08 May 2019  
We are pleased to announce the first ever Computational Medicine in the

**National Center for Biotechnology Information**

Dépôt de données dans SRA

L. Cauquil NEM, 21-22 mai 2019



<https://www.ncbi.nlm.nih.gov/>

The screenshot shows the NCBI homepage with a sidebar on the left containing a 'Resource List (A-Z)' with links to All Resources, Chemicals & Bioassays, Data & Software, DNA & RNA, Domains & Structures, Genes & Expression, Genetics & Medicine, Genomes & Maps, Homology, Literature, Proteins, Sequence Analysis, Taxonomy, Training & Tutorials, and Variation. The main content area features a 'Welcome to NCBI' section with links to About the NCBI, Mission, Organization, and NCBI News & Blog. Below this are six main sections: 'Submit' (Deposit data or manuscripts into NCBI databases, highlighted with a red box), 'Download' (Transfer NCBI data to your computer), 'Learn' (Find help documents, attend a class or watch a tutorial), 'Develop' (Use NCBI APIs and code libraries to build applications), 'Analyze' (Identify an NCBI tool for your data analysis task), and 'Research' (Explore NCBI research and collaborative projects). The right sidebar lists 'Popular Resources' including PubMed, Bookshelf, PubMed Central, BLAST, Nucleotide, Genome, SNP, Gene, Protein, and PubChem. The 'NCBI News & Blog' section lists recent articles: 'Prokaryotic Genome Annotation Pipeline (PGAP) now produces results suitable for submission to GenBank' (13 May 2019), 'May 15, 2019 Webinar: Using taxonomic information and other improvements in standalone BLAST+ (2.9.0) and the v5 databases' (08 May 2019), and 'Computational Medicine in the Cloud Hackathon: June 10-11, 2019' (08 May 2019).

**National Center for Biotechnology Information**

Dépôt de données dans SRA

L. Cauquil NEM, 21-22 mai 2019

NIH U.S. National Library of Medicine NCBI National Center for Biotechnology Information

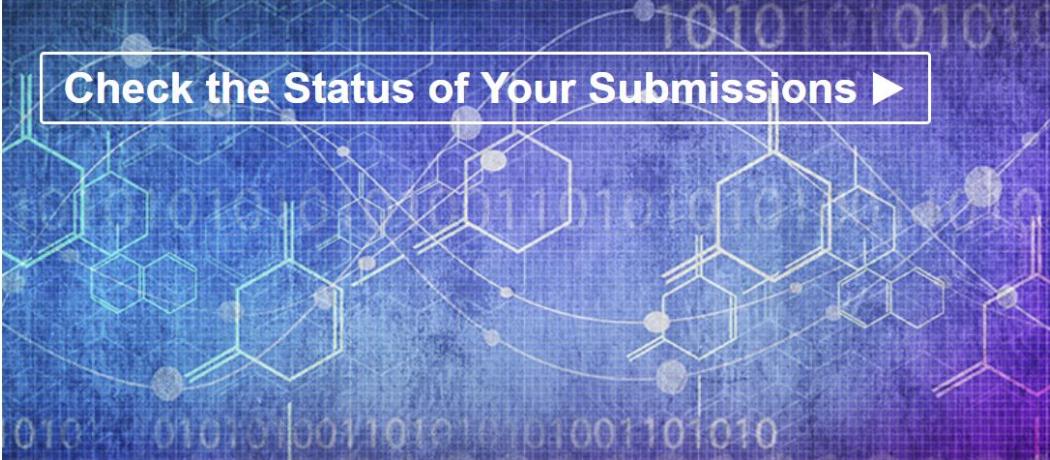
thgi My NCBI Sign Out

NCBI HOME LITERATURE HEALTH GENOMES GENES PROTEINS CHEMICALS POPULAR RESOURCES ▾

All Databases Search NCBI Search

## Submit

NCBI collects submissions of data for the world's largest public repository of biological and scientific information.



**Check the Status of Your Submissions ►**

**QuickStart**  
You know where you want to go. Select it now!

Sequence Read Archive (SRA) ▾ GO

**Submission Wizard**  
Need help figuring out where to start? Try this!

+

<https://www.ncbi.nlm.nih.gov/home/submit-wizard/>

NCBI to phase out support for non-human organism data in dbSNP and dbVar Tuesday, May 9, 2017

Eleven eukaryotic annotations added to RefSeq in April 2017 Monday, May 8, 2017

GenBank release 219.0 is available via FTP Thursday, May 4, 2017

RefSeq release 81 now public Tuesday, March 14, 2017

GenBank release 218.0 is now available Tuesday, March 14, 2017

New NCBI Insights post: Converting Lots of GI Numbers to Accession.version Friday, December 23, 2016

GI numbers will be removed from sequence record presentations Monday, October 17, 2016

Genomes-announce listserv reactivated Wednesday, September 28, 2016



NIH U.S. National Library of Medicine  
National Center for Biotechnology Information

Log in

## Submission Portal

Home Templates

### Sequence Read Archive (SRA)

The SRA accepts genetic data and the associated quality scores produced by next generation sequencing technologies. Please refer to the [File Format Guide](#).

- Files can be compressed using **gzip** or **bzip2**, and may be submitted in a tar archive but archiving and/or compressing your files is not required. **Do not use zip!**
- All file names must be **unique**.
- **Each file must be listed** in the SRA metadata table. If you are uploading a tar archive, list each file name, not the archive name.
- Use the **preload** option if you are uploading files over **10 GB** or more than files. All files for a submission must be uploaded into a **single folder**.

Before submitting, read the [SRA Submission Wizard Help](#).

Contact [sra@ncbi.nlm.nih.gov](mailto:sra@ncbi.nlm.nih.gov) with any question or concern about your data or submission.

Please [login](#) to create a new submission or to see your existing submissions.

### National Center for Biotechnology Information

8600 Rockville Pike  
Bethesda MD, 20894 USA

[About us](#) [Contact us](#) [Policies](#) [FOIA](#)

NCBI Support Center

U.S. National Library of Medicine  
National Institutes of Health  
Department of Health and Human Services  
USA.gov

Last revision: 1.86.1

NIH U.S. National Library of Medicine  
National Center for Biotechnology Information

**Submission Portal**

Home Templates

**Log in**

## Sequence Read Archive (SRA)

The SRA accepts genetic data and the associated quality scores produced by next generation sequencing technologies. Please refer to the [File Format Guide](#).

- Files can be compressed using **gzip** or **bzip2**, and may be submitted in a tar archive but archiving and/or compressing your files is not required. **Do not use zip!**
- All file names must be **unique**.
- **Each file must be listed** in the SRA meta name, not the archive name.
- Use the **preload** option if you are upload must be uploaded into a **single folder**.

Before submitting, read the [SRA Submission Guide](#)

Contact [sra@ncbi.nlm.nih.gov](mailto:sra@ncbi.nlm.nih.gov) with any quest

Please [login](#) to create a new submission or to see your existing submissions.

Please [login](#) to create a new submission or to see your existing submissions.

### National Center for Biotechnology Information

8600 Rockville Pike  
Bethesda MD, 20894 USA

[About us](#) [Contact us](#) [Policies](#) [FOIA](#)

 NCBI Support Center

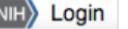
U.S. National Library of Medicine  
National Institutes of Health  
Department of Health and Human Services  
USA.gov

Last revision: 1.86.1

NCBI Resources How To Sign in to NCBI

### Sign in to NCBI

Sign in with

 Google  NIH Login  eRA Commons

[See more 3rd party sign in options](#)

---

OR

#### Sign in directly to NCBI

NCBI Username

Password

Keep me signed in

**Sign In**

[Forgot NCBI username or password?](#)

[Register for an NCBI account](#)

My NCBI retains user information and database preferences to provide customized services for many NCBI databases.

 [My NCBI Overview](#)

My NCBI features include:

- Save searches & automatic e-mail alerts
- Display format preferences
- Filter options
- My Bibliography & NIH public access policy compliance
- [SciENcv](#): a researcher biosketch profile service
- Highlighting search terms
- Recent activity searches & records for 6 months
- LinkOut, document delivery service & outside tool selections

**NIH funded investigator?**

Extramural NIH-funded investigators looking for NIH Public Access Compliance tools can sign in with either "eRA Commons" or "NIH Login". Use your eRA Commons credentials on the subsequent sign in page. Once signed in, navigate to the My Bibliography section.

Documentation for using these features is located in the [Managing Compliance to the NIH Public Access Policy](#) section of the NCBI Help Manual.

Information about the NIH Public Access Policy is located at <https://publicaccess.nih.gov>.

**Account Troubleshooting FAQ**

[Expired email confirmation link message](#)  
[Multiple My NCBI accounts](#)  
[Link eRA Commons, University, or other account to your NCBI account](#)



NCBI

NCBI Resources ▾ How To ▾

Sign in to NCBI

My NCBI » Register

help

### Register for an NCBI Account

\* required information

#### Select a username and password

Username: \*

Password: \*

Repeat password: \*

#### Contact information

E-mail: \*

#### In case you forget your password

Please provide a question and answer that you can use to unlock your account:

Question:

Answer:

Please type the following characters: \*



Create account

Sign in with an existing account

NIH U.S. National Library of Medicine  
National Center for Biotechnology Information

thgi

Submission Portal

Sequence Read Archive (SRA) [New submission](#)

Short description and brief instructions [+](#)

Options to preload data:

- Aspera browser plugin upload [+](#)
- Aspera command line and FTP upload options [+](#)
- Amazon S3 instructions [+](#)

10 submissions

Submission	Title	Status	Updated
SUB5416748	Internal 16S standard, Apr 05 '19	<ul style="list-style-type: none"> <li>✓ BioProject: <b>Processed</b> PRJNA531076 : Internal 16S standard</li> <li>✓ BioSample: <b>Processed</b> Successfully loaded (10 objects) Download attributes file with BioSample accessions</li> <li>✓ SRA: <b>Processed</b> (10 objects) Download metadata file with SRA accessions View and manage my SRA submission data</li> </ul>	Apr 05

**Etat des dernières soumissions**

Filter / Search

From date  To date  Status  Sort by   desc

Data archives [+](#)

Query [?](#)  [Search](#) [Clear](#)

U.S. National Library of Medicine  
National Center for Biotechnology Information

NIH

 thgi

Submission Portal
Home
My submissions
Manage data
Templates
My profile

Sequence Read Archive (SRA)
New submission

Short description and brief instructions
+

Options to preload data:

Aspera browser plugin upload
+

Aspera command line and FTP upload options
+

Amazon S3 instructions
+

10 submissions

Submission
Title
Status
Updated

SUB5416748
Internal 16S standard, Apr 05 '19

✓ BioProject: Processed  
 PRJNA531076 : Internal 16S standard
 
Apr 05

**Etat des dernières soumissions**

Filter / Search
+

From date
To date
Status
Sort by

Not deleted
desc

Data archives
+

Query
Search
Clear

NIH U.S. National Library of Medicine  
National Center for Biotechnology Information

thgi

Submission Portal

Home My submissions Manage data Templates My profile

**Sequence Read Archive (SRA) submission: SUB5620720**

New [Delete submission](#)

1 SUBMITTER > 2 GENERAL > 3 METADATA > 4 FILES > 5 REVIEW & SUBMIT

**Submitter**

Required fields are marked with **\*** asterisk

\* First (given) name Middle name \* Last (family) name

\* Email (primary) Email (secondary)  
  At least one email should be from the organization's domain.

Group for this submission  
 (affiliation from my personal profile)

[Create group for shared submissions](#) Allow selected collaborators to read, modify, submit and delete your submissions

\* Submitting organization Submitting organization URL \* Department

Phone ? Fax ?

\* Street \* City State/Province \* Postal code \* Country

[Continue](#)  Update my contact information in profile

**National Center for Biotechnology Information**

8600 Rockville Pike  
Bethesda MD 20894 USA

NCBI Support Center  
U.S. National Library of Medicine  
National Institutes of Health

NIH U.S. National Library of Medicine  
National Center for Biotechnology Information

Submission Portal

Sequence Submission

New

1 SUBMITTER

2 GENERAL

3 METADATA

4 FILES

5 REVIEW & SUBMIT

My profile

Submission

with ★ asterisk

First (given) name Middle name Last (family) name  
laurent cauquil

Email (primary) Email (secondary)  
Laurent.Cauquil@toulouse.inra.fr Laurent.Cauquil@inra.fr At least one email should be from the organization's domain.

Group for this submission  
No group (affiliation from my personal profile)

Create group for shared submissions Allow selected collaborators to read, modify, submit and delete your submissions

Submitting organization URL Department  
INRA UMR1388 GenPhySE Agriculture

Phone Fax  
[ ] [ ]

Street City State/Province Postal code Country  
Chemin de Borde Rouge BP 52627 Castanet Tolosan C 31326 France

Continue  Update my contact information in profile

National Center for Biotechnology Information  
8600 Rockville Pike  
Bethesda MD 20894 USA

NCBI Support Center  
U.S. National Library of Medicine  
National Institutes of Health

## Submission Portal

Home [My submissions](#) Manage data Templates My profile

### Sequence Read Archive (SRA) submission: SUB5620720

New

[Delete submission](#)

1 SUBMITTER > 2 GENERAL > 3 METADATA > 4 FILES > 5 REVIEW & SUBMIT

#### Submitter

Required fields are marked with \* asterisk

\* First (given) name  Middle name  \* Last (family) name

\* Email (primary)  Email (secondary)

 At least one email should be from the organization's domain.

#### Group for this submission

No group (affiliation from my personal profile)

[Create group for shared submissions](#)

 Allow selected collaborators to read, modify, submit and delete your submissions

Pré-remplis à partir des info du login

#### \* Submitting organization

#### Submitting organization URL

#### \* Department

#### Phone

#### Fax

#### \* Street

#### \* City

#### State/Province

#### \* Postal code

#### \* Country

[Continue](#)

Update my contact information in profile

#### National Center for Biotechnology Information

8600 Rockville Pike  
Bethesda MD 20894 USA

 NCBI Support Center

U.S. National Library of Medicine  
National Institutes of Health

## Submission Portal

Home My submissions Manage data Templates My profile

thgi

Sequence Read Archive (SRA) submission: SUB5620720

Delete submission

New

1 SUBMITTER 2 GENERAL 3 METADATA 4 FILES 5 REVIEW &amp; SUBMIT

## General Information

\* Required fields are marked with \* asterisk

## BioProject

 BioProject describes the goal of your research effort.

Did you already register a BioProject for this research, e.g. for the submission of the reads to SRA and/or of the genome to GenBank?

Yes  No

\* Existing BioProject 



## BioSample

 The BioSample records the detailed biological and physical properties of the sample that was sequenced. A BioSample can be used in more than one BioProject since it should be used for all the data that were obtained from that sample. Usually SRA data sets are generated from more than one sample.

Did you already register a BioSample for this sample, e.g. for the submission of the reads to SRA and/or of the genome to GenBank?

Yes  No



## Release date

 Note: Release of BioProject or BioSample is also triggered by the release of linked data.

\* When should this submission be released to the public?

Release immediately following processing  
 Release on specified date or upon publication, whichever is first



Continue

NIH U.S. National Library of Medicine  
National Center for Biotechnology Information

thgi

Submission Portal

Sequence Read Archive (SRA) submission: SUB5620720

New

1 SUBMITTER 2 GENERAL 3 METADATA 4 FILES 5 REVIEW & SUBMIT

Home My submissions Manage data Templates My profile

Delete submission

**BioProject**

**General Information**

**BioProject**

**Did you already register a BioProject for this research, e.g. for the submission of the reads to SRA and/or of the genome to GenBank?**

Yes  No

**Existing BioProject**

**Release date**

**Release date**

Release date

**Note:** Release of BioProject or BioSample is also triggered by the release of linked data.

**When should this submission be released to the public?**

Release immediately following processing  
 Release on specified date or upon publication, whichever is first

Continue

NIH U.S. National Library of Medicine  
National Center for Biotechnology Information

thgi

Submission Portal

Home My submissions Manage data Templates My profile

**Sequence Read Archive (SRA) submission: SUB5620720**

New

**Delete submission**

1 SUBMITTER 2 GENERAL 3 METADATA 4 FILES 5 REVIEW & SUBMIT

**General Information**

Required fields are marked with **\*** asterisk

**BioProject**

**Info:** BioProject describes the goal of your research effort.

Did you already register a BioProject for this research, e.g. for the submission of the reads to SRA and/or of the genome to GenBank?

Yes  No

**BioSample**

**Info:** The BioSample records the detailed biological and physical properties of the sample that was sequenced. A BioSample can be used in more than one BioProject since it should be used for all the data that were obtained from that sample. Usually SRA data sets are generated from more than one sample.

Did you already register a BioSample for this sample, e.g. for the submission of the reads to SRA and/or of the genome to GenBank?

Yes  No

**Release date**

**Info:** Note: Release of BioProject or BioSample is also triggered by the release of linked data.

**★ When should this submission be released to the public?**

Release immediately following processing  
 Release on specified date or upon publication, whichever is first

**★ Projected release date**  

**Info:** You will be able to change release date later (please refer to the [SRA Update Guide](#)). The data must be released upon publication.

**Continue**

NIH U.S. National Library of Medicine  
National Center for Biotechnology Information

thgi

Submission Portal

Home My submissions Manage data Templates My profile

Sequence Read Archive (SRA) submission: SUB5620720

New

1 SUBMITTER 2 GENERAL 3 METADATA 4 FILES 5 REVIEW & SUBMIT

General Information

Required fields are marked with **\*** asterisk

**BioProject**

**Did you already publish or plan to publish this BioProject and/or of the data it contains?**

Yes  No

**BioSample**

**Did you already publish or plan to publish this BioSample and/or of the data it contains?**

Yes  No

**Release date**

**★ When should this submission be released to the public?**

Release immediately following processing  
 Release on specified date or upon publication, whichever is first

Date format: YYYY-MM-DD, whichever is first  
Example: 2014-10-23

**★ Projected release date**

**Note:** Release of BioProject or BioSample is also triggered by the release of linked data.

**★ When should this submission be released to the public?**

Release immediately following processing  
 Release on specified date or upon publication, whichever is first

**★ Projected release date**

**Note:** You will be able to change release date later (please refer to the [SRA Update Guide](#)).  
The data must be released upon publication.

Continue

U.S. National Library of Medicine  
National Center for Biotechnology Information

thgi

Home My submissions Manage data Templates My profile

Submission Portal

Sequence Read Archive (SRA) submission: SUB5620720

New

1 SUBMITTER 2 GENERAL INFO 3 PROJECT INFO 4 BIOSAMPLE TYPE 5 BIOSAMPLE ATTRIBUTES 6 SRA METADATA 7 FILES 8 REVIEW & SUBMIT

Project Info

Required fields are marked with \* asterisk

\* Project title [?](#)

\* Public description [?](#)

Relevance [?](#)

\* Is your project part of a larger initiative which is already registered with NCBI?  
 No  Yes (not very common)

External Links

Link description [?](#) URL [?](#) Delete [?](#)  
  [Delete](#)

[Add another link](#)

Select your grants

Use this tool to look up grants from many subscribed governmental funding agencies (eg NIH, CDC, FDA and VA) and some non-governmental funding sources (eg HHMI and Autism Speaks). You can search by grant number, title or grantee name. If your grant is not included, you can select the "Add grants manually" option within this tool to add your grant.

[Add grants](#)

Consortium name [?](#) Consortium URL [?](#)

Data provider [?](#) Data provider URL [?](#) Delete [?](#)  
  [Delete](#)

[Add another data provider](#)

[Continue](#)

National Center for Biotechnology Information  
8600 Rockville Pike  
Bethesda MD 20894 USA

NCBI Support Center  
[?](#)  
U.S. National Library of Medicine  
National Institutes of Health

## 3 PROJECT INFO

## 4 BIOSAMPLE TYPE

## 5 BIOSAMPLE ATTRIBUTES

## 6 SRA METADATA

## 7 FILES

## 8 REVIEW & SUBMIT

**Public description**  ⓘ

**Relevance**  ⓘ

**Is your project part of a larger initiative which is already registered with NCBI?**  
 No  Yes (not very common)

**External Links**

<b>Link description</b> <small> ⓘ</small>	<b>URL</b> <small> ⓘ</small>	<b>Delete</b>
<input type="text"/>	<input type="text"/>	

**Add another link**

**Select your grants**

ⓘ Use this tool to look up grants from many subscribed governmental funding agencies (eg NIH, CDC, FDA and VA) and some non-governmental funding sources (eg HHMI and Autism Speaks). You can search by grant number, title or grantee name. If your grant is not included, you can select the "Add grants manually" option within this tool to add your grant.

**Add grants**

<b>Consortium name</b> <small> ⓘ</small>	<b>Consortium URL</b> <small> ⓘ</small>
<input type="text"/>	<input type="text"/>

<b>Data provider</b> <small> ⓘ</small>	<b>Data provider URL</b> <small> ⓘ</small>	<b>Delete</b>
<input type="text"/>	<input type="text"/>	

**Add another data provider**

**Continue**

**National Center for Biotechnology Information**

8600 Rockville Pike  
Bethesda MD 20894 USA

**NCBI Support Center**

U.S. National Library of Medicine  
National Institutes of Health

U.S. National Library of Medicine  
National Center for Biotechnology Information

thgi

Submission Portal

Sequence Read Archive (SRA) submission: SUB5620720

New

1 SUBMITTER 2 GENERAL INFO 3 PROJECT INFO 4 BIOSAMPLE TYPE 5 BIOSAMPLE ATTRIBUTES 6 SRA METADATA 7 FILES 8 REVIEW & SUBMIT

Project Info

Required fields are marked with \* asterisk

**Title**

**Description**

\* Project title [?](#)

\* Public description [?](#)

Relevance [?](#)

\* Is your project part of a larger initiative which is already registered with NCBI?  
 No  Yes (not very common)

External Links

Link description [?](#) URL [?](#) Delete [?](#)  
  [Delete](#)

[Add another link](#)

Select your grants

Use this tool to look up grants from many subscribed governmental funding agencies (eg NIH, CDC, FDA and VA) and some non-governmental funding sources (eg HHMI and Autism Speaks). You can search by grant number, title or grantee name. If your grant is not included, you can select the "Add grants manually" option within this tool to add your grant.

[Add grants](#)

Consortium name [?](#) Consortium URL [?](#)

Data provider [?](#) Data provider URL [?](#) Delete [?](#)  
  [Delete](#)

[Add another data provider](#)

**Continue**

National Center for Biotechnology Information  
8600 Rockville Pike  
Bethesda MD 20894 USA

NCBI Support Center  
[?](#)  
U.S. National Library of Medicine  
National Institutes of Health

## Sample Type

Required fields are marked with \* asterisk

\* Select the package that best describes your samples:

**Pathogen affecting public health**

Use for pathogen samples that are relevant to public health. Required attributes include those considered useful for the rapid analysis and trace back of pathogens.

**Microbe**

Use for bacteria or other unicellular microbes when it is not appropriate or advantageous to use MiS, Pathogen or Virus packages.

**Model organism or animal sample**

Use for multicellular samples or cell lines derived from common laboratory model organisms, e.g., mouse, rat, Drosophila, worm, fish, frog, or large mammals including zoo and farm animals.

**Metagenome or environmental sample**

Use for metagenomic and environmental samples when it is not appropriate or advantageous to use MiS packages.

**Invertebrate**

Use for any invertebrate sample.

**Human sample**

WARNING: Only use for human samples or cell lines that have no privacy concerns. For all studies involving human subjects, it is the submitter's responsibility to ensure that the information supplied protects participant privacy in accordance with all applicable laws, regulations and institutional policies. Make sure to remove any direct personal identifiers from your submission. If there are patient privacy concerns regarding making data fully public, please submit samples and data to NCBI's dbGaP database. [dbGaP](#) has controlled access mechanisms and is an appropriate resource for hosting sensitive patient data. For samples isolated from humans use the Pathogen, Microbe or appropriate MiS package.

**Plant sample**

Use for any plant sample or cell line.

**Virus sample**

Use for all virus samples not directly associated with disease. Viral pathogens should be submitted using the Pathogen: Clinical or host-associated pathogen package.

**Genome, metagenome or marker sequences (MiS compliant)**

Use for genomes, metagenomes, and marker sequences. These samples include specific attributes that have been defined by the Genome Standards Consortium (GSC) to formally describe and standardize sample metadata for genomes, metagenomes, and marker sequences. The samples are validated for compliance based on the presence of the required core attributes as described in [MiS](#).

**Metagenome**

### Sequence Read Archive (SRA) submission: SUB5620720

Projet test, May 15 '19

[Delete submission](#)

1 SUBMITTER &gt; 2 GENERAL INFO &gt; 3 PROJECT INFO &gt; 4 BIOSAMPLE TYPE &gt; 5 BIOSAMPLE ATTRIBUTES &gt; 6 SRA METADATA &gt; 7 FILES &gt; 8 REVIEW &amp; SUBMIT

#### Attributes

Required fields are marked with **\*** asterisk.  
At least one of the fields marked with **\*\***, **††** or **‡‡** is required.

**\*** How do you want to provide your BioSample attributes?

- Use built-in table editor
- Upload a file using Excel or text format (tab-delimited) that includes the attributes for each of your BioSamples

[Continue](#)


Metadata associées à chaque échantillon



#### National Center for Biotechnology Information

 8600 Rockville Pike  
Bethesda MD, 20894 USA

[About us](#) [Contact us](#) [Policies](#) [FOIA](#)
[NCBI Support Center](#)

 U.S. National Library of Medicine  
National Institutes of Health  
Department of Health and Human Services  
USA.gov

Last revision: 1.86.1

**Sequence Read Archive (SRA) submission: SUB5620720**

Projet test, May 15 '19

[Delete submission](#)[1 SUBMITTER](#) > [2 GENERAL INFO](#) > [3 PROJECT INFO](#) > [4 BIOSAMPLE TYPE](#) > **5 BIOSAMPLE ATTRIBUTES** > [6 SRA METADATA](#) > [7 FILES](#) > [8 REVIEW & SUBMIT](#)**Attributes** Required fields are marked with **\*** asterisk.At least one of the fields marked with **\*\***, **††** or **‡‡** is required.**★ How do you want to provide your BioSample attributes?**

- Use built-in table editor
- Upload a file using Excel or text format (tab-delimited) that includes the attributes for each of your BioSamples

[Continue](#)**National Center for Biotechnology Information**8600 Rockville Pike  
Bethesda MD, 20894 USA[About us](#) [Contact us](#) [Policies](#) [FOIA](#) NCBI Support CenterU.S. National Library of Medicine  
National Institutes of Health  
Department of Health and Human Services  
USA.gov

Last revision: 1.86.1

NIH U.S. National Library of Medicine  
National Center for Biotechnology Information

thgi

## Submission Portal

Home My submissions Manage data Templates My profile

### Sequence Read Archive (SRA) submission: SUB5620720

Projet test, May 15 '19

Delete submission

1 SUBMITTER 2 GENERAL INFO 3 PROJECT INFO 4 BIOSAMPLE TYPE 5 BIOSAMPLE ATTRIBUTES 6 SRA METADATA 7 FILES 8 REVIEW & SUBMIT

#### Attributes

Required fields are marked with \* asterisk.  
At least one of the fields marked with \*\*, †† or # is required.

\* How do you want to name your samples? Sample Name is a name that you choose for the sample. It can have any format, but we suggest that you make it concise, unique and consistent within your lab, and as informative as possible. Every Sample Name from a single Submitter must be unique.

Use built-in template  
Upload a file template

Right-click on the table if you want to edit rows, columns or find and replace text.

	Sample Name	Sample title	BioProject accession	Organism	host	isolation source	collection date	geographic location	latitude and longitude	reference for biomaterial	relationship to oxygen	sample
1												
2												
3												
4												
5												
6												
7												
8												
9												
10												
11												
12												
13												
14												
15												
16												
17												
18												
19												
20												

NIH U.S. National Library of Medicine  
National Center for Biotechnology Information

thgi

Submission Portal

Sample name

Sequence Read Archive (SRA) submission: SRR5520720

Projet test, May 15 '19

1 SUBMITTER 2 GENERAL INFO 3

Organisme

Attributes

Host

Isolation source

Collection date

Geographic location

Latitude and longitude

Required fields are marked with \* asterisk.  
At least one of the fields marked with \*\*, †† or # is required.

Right-click on the table if you want to edit rows, columns or find and replace text.

Sample Name is a name that you choose for the sample. It can have any format, but we suggest using a name that is descriptive of the sample within your lab, and a unique name for each sample. You can upload a file containing sample names from a single Submitter or from multiple Submitters.

Sample title

isolation source collection date geographic location latitude and longitude reference for biomaterial relationship to oxygen sample

Sample Name	Sample title	isolation source	collection date	geographic location	latitude and longitude	reference for biomaterial	relationship to oxygen	sample
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								
11								
12								
13								
14								
15								
16								
17								
18								
19								
20								



NIH U.S. National Library of Medicine  
National Center for Biotechnology Information

thgi

Submission Portal

Home My submissions Manage data Templates My profile

**Sequence Read Archive (SRA) submission: SUB5620720**

Projet test, May 15 '19

**Delete submission**

1 SUBMITTER 2 GENERAL INFO 3 PROJECT INFO 4 BIOSAMPLE TYPE 5 BIOSAMPLE ATTRIBUTES \* 6 SRA METADATA 7 FILES 8 REVIEW & SUBMIT

**Attributes**

Required fields are marked with **\*** asterisk.  
At least one of the fields marked with **\*\***, **††** or **‡‡** is required.

**Sample name**

ech\_1  
ech\_2

**Unique attribute too!**

**!**

**\* How do you want to provide your BioSample attributes?**

Use built-in table editor

Upload a file using Excel or text format (tab-delimited) that includes the attributes for each of your BioSamples

Right-click on the table if you want to edit rows, columns or find and replace text.

Add column:

- ID
- tag

## Submission Portal

[Home](#) [My submissions](#) [Manage data](#) [Templates](#) [My profile](#)Sequence Read Archive (SRA) submission: SUB5620720

Projet test, May 15 '19

[Delete submission](#)

1 SUBMITTER > 2 GENERAL > 3 BP INFO > 4 SAMPLE TYPE > 5 ATTRIBUTES > 6 METADATA > 7 FILES > 8 REVIEW & SUBMIT >

## SRA metadata

 Required fields are marked with **\*** asterisk

 For more detailed help with SRA submission please read the [SRA Submission Wizard Help](#).

**★ How do you want to provide your metadata?**

- Use built-in table editor
- Upload a file using Excel or text format (tab-delimited)

[Continue](#)

## ★ Drop-down menu

ⓘ For more detailed help with SRA submission please read the [SRA Submission Wizard Help](#).

★ How do you want to provide your metadata?

Use built-in table editor

Upload a file using Excel or text format (tab-delimited)

ⓘ Right-click on the table if you want to edit Exact file name, including the extension.

★ Sample name	★ Library ID ⓘ	★ Title ⓘ	★ Library strategy	★ Library source	★ Library selection	★ Library layout	★ Platform	★ Instrument model	★ Design description ⓘ	★ Filetype	Reference FASTA File ⓘ	Reference assembly ⓘ	★ Filename ⓘ	filename2	filename3
1 ech_1	ech_1_ATGC_R	test	AMPLICON	METAGENOMIC	PCR	paired	ILLUMINA	Illumina MiSeq	PCR of V3-V4 DNA16S region	fastq			ech_1_ATGC_R.fastq.gz		
2 ech_2	ech_2_TGAC_R	test	AMPLICON	METAGENOMIC	PCR	paired	ILLUMINA	Illumina MiSeq	PCR of V3-V4 DNA16S region	fastq			ech_2_ATGC_R.fastq.gz		
3															
4															
5															
6															
7															
8															
9															
10															
11															
12															
13															
14															
15															
16															
17															
18															
19															
20															

For more detailed help with SRA submission please read the [SRA Submission Wizard Help](#).

\* How do you want to provide your metadata?

Use built-in table editor

Upload a file using Excel® text format (tab-delimited)

Right-click on the table if you want to edit **Exact file name, including the extension.**

★ Sample name	★ Library ID	★ Title	★ Library strategy	★ Library source	★ Library selection	★ Library layout	★ Platform	★ Instrument model	★ Design description	★ Filetype	Reference FASTA File	Reference assembly	★ Filename	filename2	filename3
1 ech_1	ech_1_ATGC_R	test	AMPLICON	METAGENOMIC	PCR	paired	ILLUMINA	Illumina MiSeq	PCR of V3-V4 DNA16S region	fastq			ech_1_ATGC_R.fastq.gz		
2 ech_2	ech_2_TGAC_R	test	AMPLICON	METAGENOMIC	PCR	paired	ILLUMINA	Illumina MiSeq	PCR of V3-V4 DNA16S region	fastq			ech_2_ATGC_R.fastq.gz		
3															
4															
5															
6															
7															
8															
9															
10															
11															
12															
13															
14															
15															
16															
17															
18															
19															
20															

★ **Library ID: Filename sans extension**

★ **Filename: Filename avec extension**

**MiSeq paired: 2 fichiers par échantillon (R1, R2)**

Filename -> ech\_1\_R1.fastq.gz

filename2-> ech\_1\_R2.fastq.gz

Right-click on the table if you want to edit rows, columns or cells.

n	Library layout	Platform	Instrument model	Design description	Filetype	Reference FASTA File	Reference assembly	Filename	filename2	file
▼	paired	ILLUMINA	Illumina MiSeq	PCR V3-V4 DNA 16S region	fastq	▼	▼	▼	ech_1_ATGC_R1.fastq.gz	ech_1_ATGC_R2.fastq.gz
▼	paired	ILLUMINA	Illumina MiSeq	PCR V3-V4 DNA 16S region	fastq	▼	▼	▼	ech_2_ATGC_R1.fastq.gz	ech_2_ATGC_R2.fastq.gz
▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼
▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼

## Sequence Read Archive (SRA) submission: SUB5620720

Projet test, May 15 '19

[Delete submission](#)

1 SUBMITTER > 2 GENERAL > 3 BP INFO > 4 SAMPLE TYPE > 5 ATTRIBUTES > 6 METADATA > 7 FILES > 8 REVIEW & SUBMIT

### SRA metadata

 Required fields are marked with **\*** asterisk



**Warning:** Please be aware that for metagenomic and metatranscriptomic data sets, we expect sequence data to be split by each sample barcode for individual data files.



**Warning:** It appears there are paired-end files in the spreadsheet, but are on different rows. Paired-end files should appear on the same row of the spreadsheet for a given sample. Please edit your spreadsheet if your files are paired-end. [Show pairs](#)



**Warning:** Paired-end data in fastq format usually comes in 2 files, please make sure to enter both matching file names on the same row of the spreadsheet using the extra 'filename' columns provided.

 For more detailed help with SRA submission please read the [SRA Submission Wizard Help](#).

**\*** How do you want to provide your metadata?

Use built-in table editor

Upload a file using Excel or text format (tab-delimited)

 Right-click on the table if you want to edit rows, columns or find and replace text.

<b>*</b> Sample name	<b>*</b> Library ID 	<b>*</b> Title 	<b>*</b> Library strategy	<b>*</b> Library source	<b>*</b> Library selection	<b>*</b> Library layout	<b>*</b> Platform	<b>*</b> Instrument model	<b>*</b> Design description 	<b>*</b> Filetype	<b>*</b> Filename 	filename2	filename3	filename4	filename5	filename6	filename7
1 ech_1	ech_1_ATGC_R	test	AMPLICON	METAGENOMIC	PCR	paired	ILLUMINA	Illumina MiSeq	PCR of V3-V4 DNA16S region	fastq	ech_1_ATGC_R.fastq.gz						
2 ech_2	ech_2_TGAC_R	test	AMPLICON	METAGENOMIC	PCR	paired	ILLUMINA	Illumina MiSeq	PCR of V3-V4 DNA16S region	fastq	ech_2_ATGC_R.fastq.gz						
3																	
4																	
5																	
6																	
7																	
8																	
9																	
10																	
11																	
12																	
13																	
14																	
15																	
16																	

## Submission Portal

[Home](#) [My submissions](#) [Manage data](#) [Templates](#) [My profile](#)

### Sequence Read Archive (SRA) submission: SUB5620720

Projet test, May 15 '19

[Delete submission](#)

 1 SUBMITTER **2 GENERAL** 3 BP INFO 4 SAMPLE TYPE 5 ATTRIBUTES 6 METADATA 7 FILES 8 REVIEW & SUBMIT

#### Files

 Required fields are marked with **\*** asterisk

- Each file must be listed in the [SRA metadata table you uploaded](#). If you are uploading a **tar** archive, list each file name, not the archive name.
- **Unique file names** should be used for all files.
- Files can be compressed using **gzip** or **bzip2**, and may be submitted in a **tar** archive, but archiving or compressing your files is not required. **Do not use zip!**

**\* How do you want to provide files for this submission?**

- Web browser upload via HTTP or Aspera Connect plugin  
Do not use web browser HTTP upload if you are uploading files over 10 GB or more than 300 files.
- FTP or Aspera Command Line file preload  
All files for a submission must be uploaded into a single folder.
- Amazon S3 bucket

 Current versions of browsers Firefox, Chrome, Safari or Internet Explorer are recommended.  
To upload large files (larger than 2 GB), please use [Aspera Connect plugin](#).

**Files**


Aucun fichier choisi

 **Autofinish submission** 
[Continue](#)
**National Center for Biotechnology Information**

8600 Rockville Pike

 NCBI Support Center

[U.S. National Library of Medicine](#)

- I have all files preloaded for this submission  
 I will upload all the files now via HTTP/Aspera  
 I have all files in the cloud for this submission

 Current versions of browsers Firefox, Chrome, Safari or Internet Explorer are recommended.  
To upload large files (larger than 2 GB), please use [Aspera Connect plugin](#).

**Files**

Parcourir... Aucun fichier sélectionné.

Name	Size	Created	Delete
SC160201_AATAAT_L001_R.extendedFrags.fastq.gz	5.9 MB	<div style="width: 100%;"><div style="width: 100%;"></div></div> 4/13/2018 09:57	
SC160202_GGCTGC_L001_R.extendedFrags.fastq.gz	4.0 MB	4/13/2018 09:57	
SC160203_TTCCGT_L001_R.extendedFrags.fastq.gz	4.2 MB	<div style="width: 100%;"><div style="width: 100%;"></div></div> 4/13/2018 09:57	
SC160204_AGTGTC_L001_R.extendedFrags.fastq.gz	3.7 MB	4/13/2018 09:57	
SC160205_CCGTAG_L001_R.extendedFrags.fastq.gz	4.5 MB	4/13/2018 09:57	
SC160206_CATGAG_L001_R.extendedFrags.fastq.gz	3.1 MB	4/13/2018 09:57	
SC160207_TGAATA_L001_R.extendedFrags.fastq.gz	4.2 MB	4/13/2018 09:57	
SC160208_CGTGCG_L001_R.extendedFrags.fastq.gz	3.2 MB	<div style="width: 100%;"><div style="width: 100%;"></div></div> 4/13/2018 09:57	
SC160209_GCCCGA_L001_R.extendedFrags.fastq.gz	3.7 MB	4/13/2018 09:57	

## Informations à collecter

Submitter

Create account

General info

Release date

Project info

Title, Description

Biosample attributes

Sample name, Organisme, Host, Isolation source,  
Collection date, Geographic location, Latitude and longitude



Metadata

Sample name, Library ID,  
Design description (library info, PCR, V3-V4, Illumina....)  
Filename

## Validations des étapes par mail

SRA submission SUB5416748, "Internal 16S s... ven. 05/04/2019 22:20  
Your submission has been published ven. 05/04/2019 10:57  
BioSample accessions SAMN11349279, SAM... ven. 05/04/2019 10:10  
BioProject ID PRJNA531076 ven. 05/04/2019 10:09  
17 Ko  
19 Ko  
24 Ko  
19 Ko

 Répondre  Répondre à tous  Transférer  MI  
ven. 05/04/2019 22:20  
 sra@ncbi.nlm.nih.gov  
SRA submission SUB5416748, "Internal 16S standard, Apr 05 '19"  
À Laurent Cauquil

Dear laurent cauquil,

This is an automatic acknowledgment that your recent submission to SRA database has been successfully processed and will be released on the date specified.

Please reference PRJNA531076 in your publication. This BioProject accession number is provided above in lieu of SRP and should be used in your publication as it will allow better searching in Entrez.

SRA accession: PRJNA531076

Temporary Submission ID: SUB5416748

Release date: 2019-04-05

Your SRA records will be accessible with the following link after the indicated release date:

<https://www.ncbi.nlm.nih.gov/sra/PRJNA531076>

DONE!

Send questions and update requests to [sra@ncbi.nlm.nih.gov](mailto:sra@ncbi.nlm.nih.gov); include the SRA accession PRJNA531076 in any correspondence.

Regards,

NCBI SRA Submissions Staff  
Bethesda, Maryland USA

# Extraction de données depuis SRA



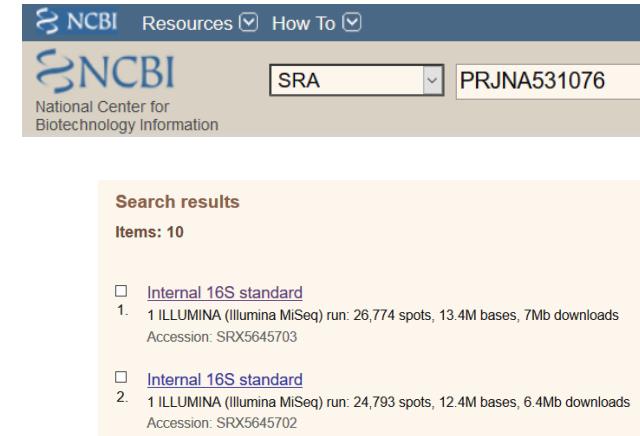


# Différentes étapes

SRA accession number: SRP- , PRJ-

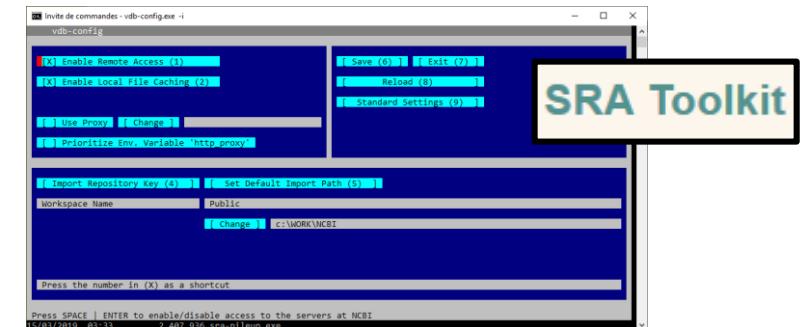
Liste des BioSample à extraire

Téléchargement des fichiers



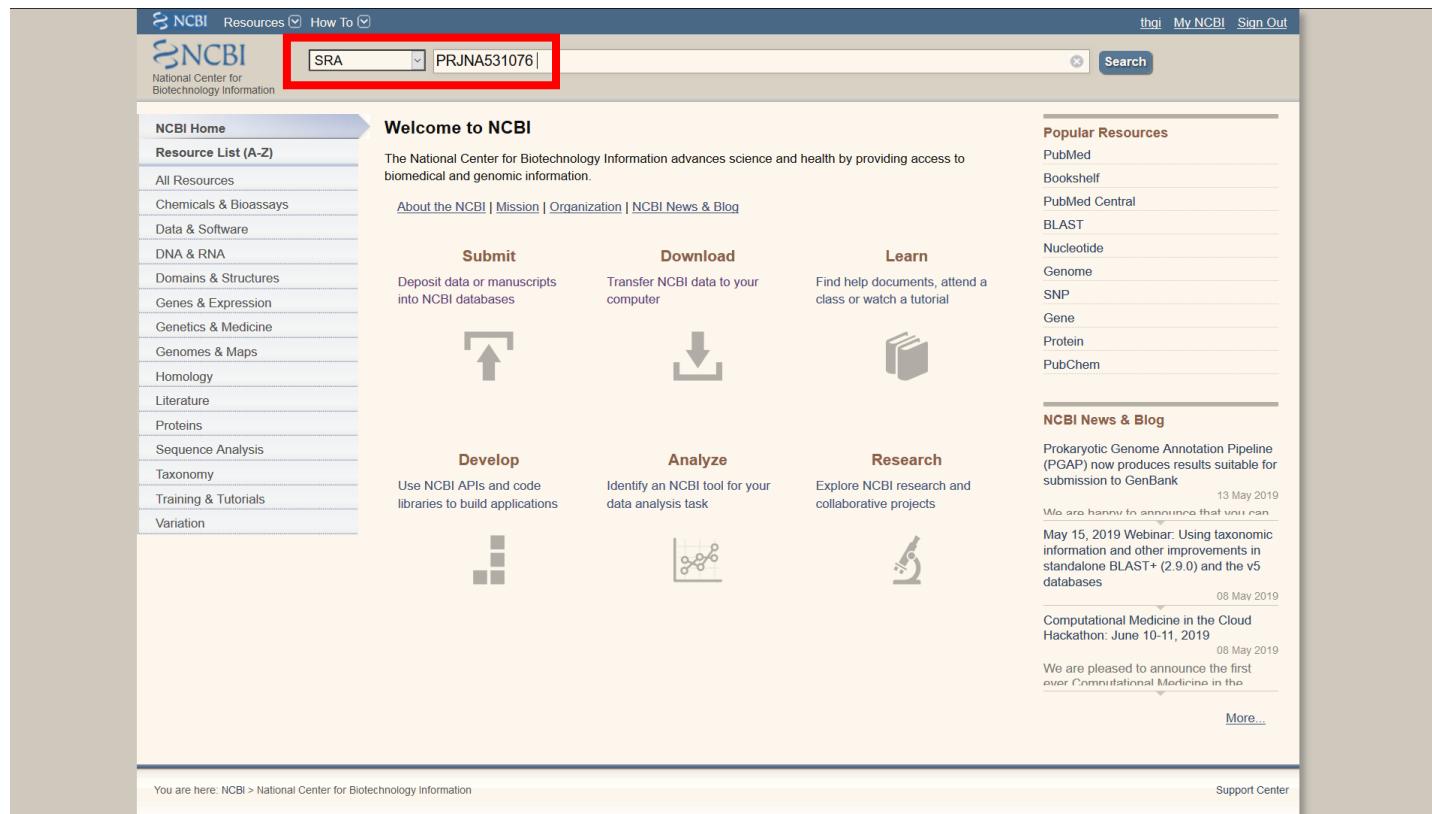
Search results  
Items: 10

- Internal 16S standard  
1. 1 ILLUMINA (Illumina MiSeq) run: 26,774 spots, 13.4M bases, 7Mb downloads  
Accession: SRX5645703
- Internal 16S standard  
2. 1 ILLUMINA (Illumina MiSeq) run: 24,793 spots, 12.4M bases, 6.4Mb downloads  
Accession: SRX5645702



# Accession number

SRA accession number: SRP- , PRJ-



The screenshot shows the NCBI homepage. The search bar at the top has 'SRA' in the dropdown and 'PRJNA531076' in the text field, with a red box highlighting this area. The page title 'Welcome to NCBI' is visible, along with a brief description of the center's mission. Below the search bar are several navigation links and sections: 'Submit', 'Download', 'Learn', 'Develop', 'Analyze', and 'Research'. The right sidebar lists 'Popular Resources' including PubMed, Bookshelf, PubMed Central, BLAST, Nucleotide, Genome, SNP, Gene, Protein, and PubChem. The 'NCBI News & Blog' section contains several recent articles with dates like 13 May 2019 and 08 May 2019.



# BioSample

NCBI Resources How To Sign in to NCBI

SRA SRA PRJNA531076 Search Create alert Advanced Help

Access Public (10) Summary 20 per page Send to: Filters: Manage Filters

Source DNA (10) Internal 16S standard - BioProject

Description of an internal standard compatible with 16S sequencing of the V4 region  
Raw sequence reads project  
Accession: PRJNA531076

Clear all Show additional filters

Send results to Blast

Search in related databases

Database	Access		
	public	controlled	all
BioSample			
BioProject	1		1
dbGaP			
GEO Datasets			

Find related data Database: Select

Find items

Search details PRJNA531076 [All Fields]

Search See more...

Recent activity Turn Off Clear

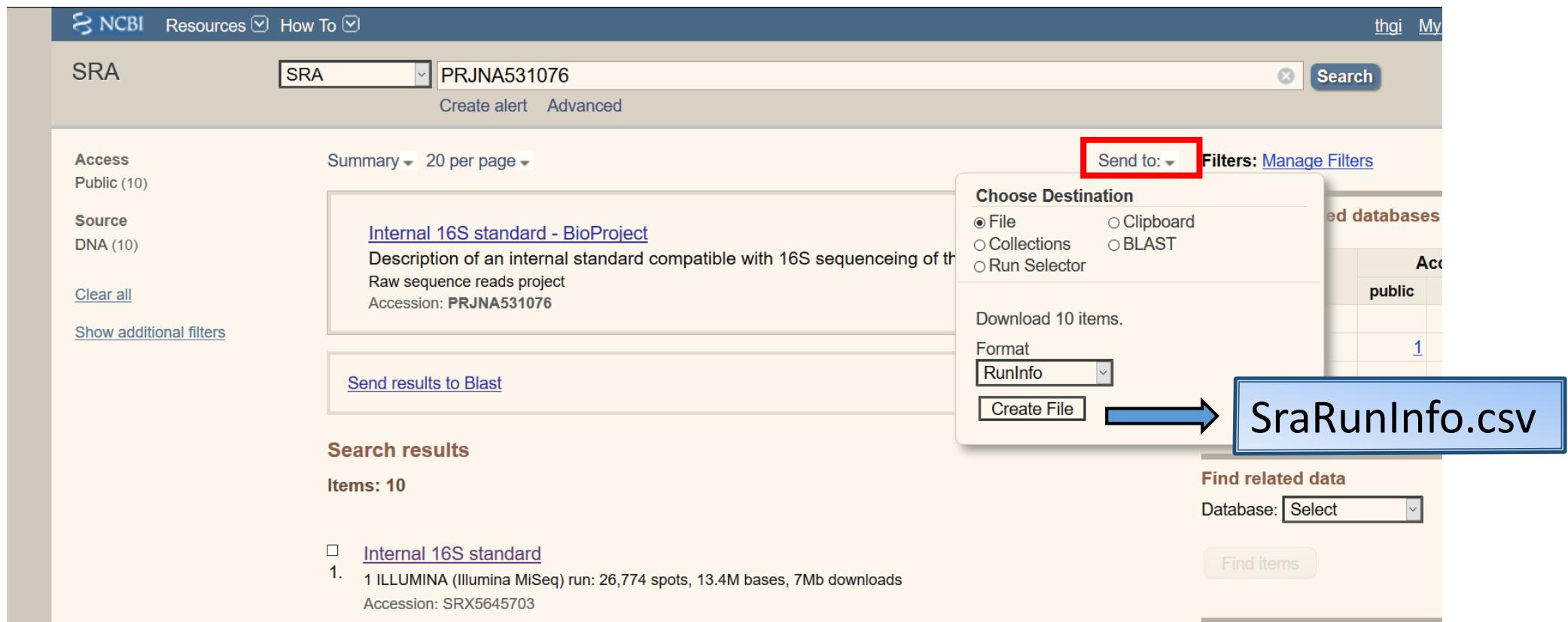
PRJNA531076 (10)

## Liste des BioSample

Search results Items: 10

- Internal 16S standard 1. 1 ILLUMINA (Illumina MiSeq) run: 26,774 spots, 13.4M bases, 7Mb downloads Accession: SRX5645703
- Internal 16S standard 2. 1 ILLUMINA (Illumina MiSeq) run: 24,793 spots, 12.4M bases, 6.4Mb downloads Accession: SRX5645702
- Internal 16S standard 3. 1 ILLUMINA (Illumina MiSeq) run: 30,031 spots, 15M bases, 7.9Mb downloads Accession: SRX5645701
- Internal 16S standard 4. 1 ILLUMINA (Illumina MiSeq) run: 37,141 spots, 18.6M bases, 9.3Mb downloads Accession: SRX5645700
- Internal 16S standard

## Liste des BioSample à extraire



NCBI Resources How To

SRA SRA PRJNA531076 Search

Create alert Advanced

Access Public (10)

Source DNA (10)

Clear all

Show additional filters

Summary 20 per page

[Internal 16S standard - BioProject](#)  
Description of an internal standard compatible with 16S sequencing of the Raw sequence reads project  
Accession: PRJNA531076

[Send results to Blast](#)

Choose Destination

File  Clipboard  
 Collections  BLAST  
 Run Selector

Download 10 items.

Format [RunInfo](#) [Create File](#)

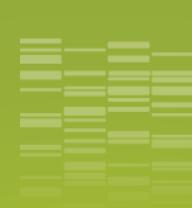
Send to: Filters: Manage Filters

Find related data

Database: Select

Find items

**SraRunInfo.csv**



# BioSample

## SraRunInfo.csv

	A	B	C	D	E	F	G	H	I	J
1	Run	ReleaseDate	LoadDate	spots	bases	spots_with_meanAvgLength	size_MB	AssemblyName	download_path	
2	SRR8858147	05/04/2019 04:19	05/04/2019 04:12	28008	1E+07	28008	500	7		<a href="https://sra-download.ncbi.nlm.nih.gov/traces/sra26/SRR/008650/SRR8858147">https://sra-download.ncbi.nlm.nih.gov/traces/sra26/SRR/008650/SRR8858147</a>
3	SRR8858144	05/04/2019 04:19	05/04/2019 04:12	24297	1E+07	24297	500	6		<a href="https://sra-download.ncbi.nlm.nih.gov/traces/sra43/SRR/008650/SRR8858144">https://sra-download.ncbi.nlm.nih.gov/traces/sra43/SRR/008650/SRR8858144</a>
4	SRR8858143	05/04/2019 04:19	05/04/2019 04:12	37141	2E+07	37141	500	9		<a href="https://sra-download.ncbi.nlm.nih.gov/traces/sra8/SRR/008650/SRR8858143">https://sra-download.ncbi.nlm.nih.gov/traces/sra8/SRR/008650/SRR8858143</a>
5	SRR8858142	05/04/2019 04:19	05/04/2019 04:12	30031	2E+07	30031	500	7		<a href="https://sra-download.ncbi.nlm.nih.gov/traces/sra57/SRR/008650/SRR8858142">https://sra-download.ncbi.nlm.nih.gov/traces/sra57/SRR/008650/SRR8858142</a>
6	SRR8858141	05/04/2019 04:19	05/04/2019 04:12	24793	1E+07	24793	500	6		<a href="https://sra-download.ncbi.nlm.nih.gov/traces/sra5/SRR/008650/SRR8858141">https://sra-download.ncbi.nlm.nih.gov/traces/sra5/SRR/008650/SRR8858141</a>
7	SRR8858140	05/04/2019 04:19	05/04/2019 04:12	26774	1E+07	26774	500	6		<a href="https://sra-download.ncbi.nlm.nih.gov/traces/sra14/SRR/008650/SRR8858140">https://sra-download.ncbi.nlm.nih.gov/traces/sra14/SRR/008650/SRR8858140</a>
8	SRR8858149	05/04/2019 04:19	05/04/2019 04:12	29510	1E+07	29510	500	7		<a href="https://sra-download.ncbi.nlm.nih.gov/traces/sra6/SRR/008650/SRR8858149">https://sra-download.ncbi.nlm.nih.gov/traces/sra6/SRR/008650/SRR8858149</a>
9	SRR8858148	05/04/2019 04:19	05/04/2019 04:12	28146	1E+07	28146	500	7		<a href="https://sra-download.ncbi.nlm.nih.gov/traces/sra37/SRR/008650/SRR8858148">https://sra-download.ncbi.nlm.nih.gov/traces/sra37/SRR/008650/SRR8858148</a>
10	SRR8858146	05/04/2019 04:19	05/04/2019 04:12	22948	1E+07	22948	500	5		<a href="https://sra-download.ncbi.nlm.nih.gov/traces/sra45/SRR/008650/SRR8858146">https://sra-download.ncbi.nlm.nih.gov/traces/sra45/SRR/008650/SRR8858146</a>
11	SRR8858145	05/04/2019 04:19	05/04/2019 04:12	25820	1E+07	25820	500	6		<a href="https://sra-download.ncbi.nlm.nih.gov/traces/sra24/SRR/008650/SRR8858145">https://sra-download.ncbi.nlm.nih.gov/traces/sra24/SRR/008650/SRR8858145</a>



## SraRunInfo.csv

Liste des BioSamples:

- ID unique
- Release date
- ...
- Run
- ...

## Biosamples

SRR8858147  
SRR8858144  
SRR8858143  
SRR8858142  
SRR8858141  
SRR8858140  
SRR8858149  
SRR8858148  
SRR8858146  
SRR8858145

## SraRunInfo.csv

### Liste des BioSamples:

- ID unique
- Release date
- ...
- [Path to download](#)
- ...

### Path to download

<https://sra-download.ncbi.nlm.nih.gov/traces/sra26/SRR/008650/SRR8858147>  
<https://sra-download.ncbi.nlm.nih.gov/traces/sra43/SRR/008650/SRR8858144>  
<https://sra-download.ncbi.nlm.nih.gov/traces/sra8/SRR/008650/SRR8858143>  
<https://sra-download.ncbi.nlm.nih.gov/traces/sra57/SRR/008650/SRR8858142>  
<https://sra-download.ncbi.nlm.nih.gov/traces/sra5/SRR/008650/SRR8858141>  
<https://sra-download.ncbi.nlm.nih.gov/traces/sra14/SRR/008650/SRR8858140>  
<https://sra-download.ncbi.nlm.nih.gov/traces/sra6/SRR/008650/SRR8858149>  
<https://sra-download.ncbi.nlm.nih.gov/traces/sra37/SRR/008650/SRR8858148>  
<https://sra-download.ncbi.nlm.nih.gov/traces/sra45/SRR/008650/SRR8858146>  
<https://sra-download.ncbi.nlm.nih.gov/traces/sra24/SRR/008650/SRR8858145>



**SRA Toolkit**

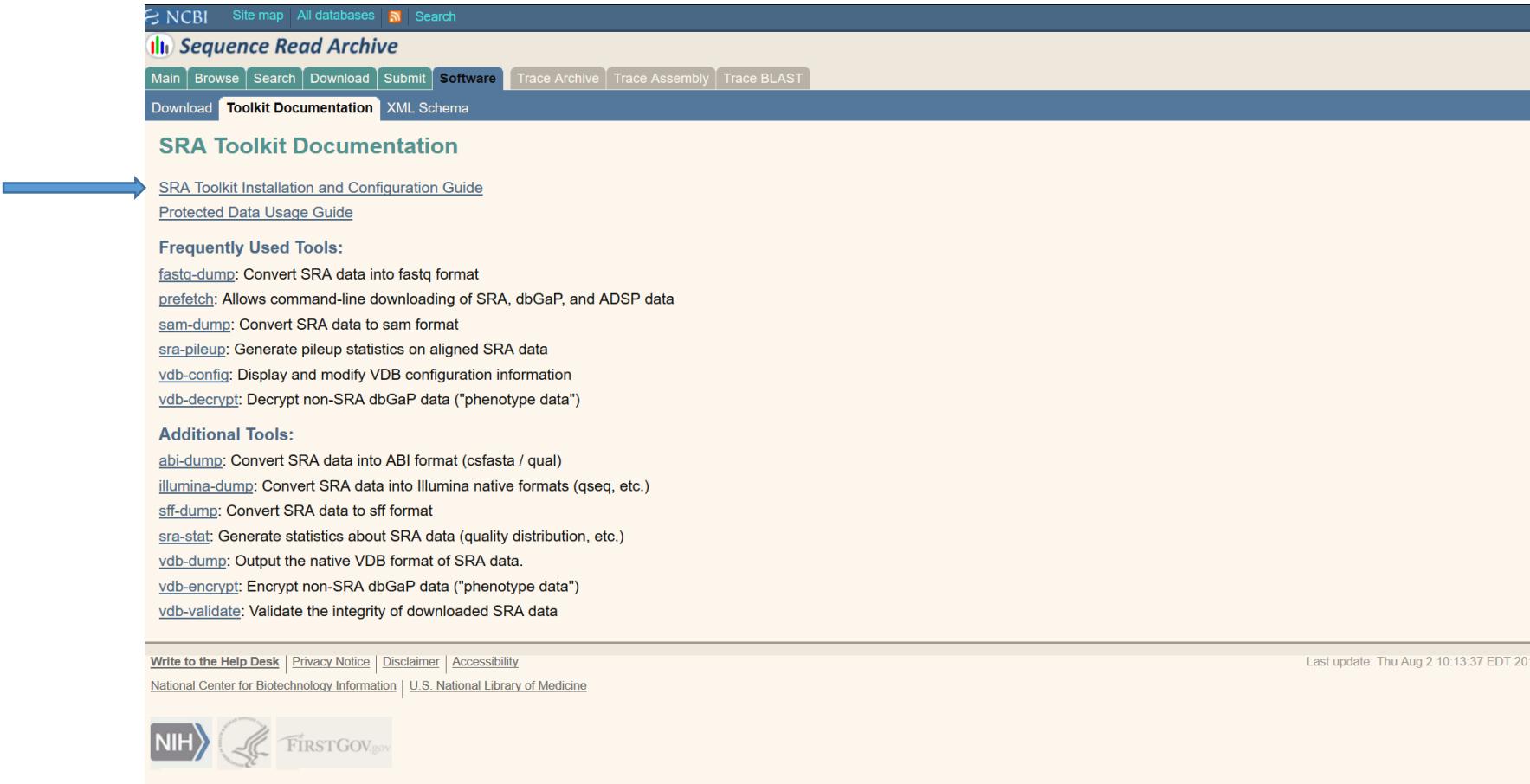


## SRA Toolkit

- Petit programme à copier sur son PC (pas d'installation)
- Ligne de commande
- Interface de configuration "old school"
- Mises à jour fréquentes



[https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=toolkit\\_doc](https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=toolkit_doc)



NCBI Site map All databases Search

Sequence Read Archive

Main Browse Search Download Submit Software Trace Archive Trace Assembly Trace BLAST

Download Toolkit Documentation XML Schema

## SRA Toolkit Documentation

[SRA Toolkit Installation and Configuration Guide](#)  
[Protected Data Usage Guide](#)

**Frequently Used Tools:**

- [fastq-dump](#): Convert SRA data into fastq format
- [prefetch](#): Allows command-line downloading of SRA, dbGaP, and ADSP data
- [sam-dump](#): Convert SRA data to sam format
- [sra-pileup](#): Generate pileup statistics on aligned SRA data
- [vdb-config](#): Display and modify VDB configuration information
- [vdb-decrypt](#): Decrypt non-SRA dbGaP data ("phenotype data")

**Additional Tools:**

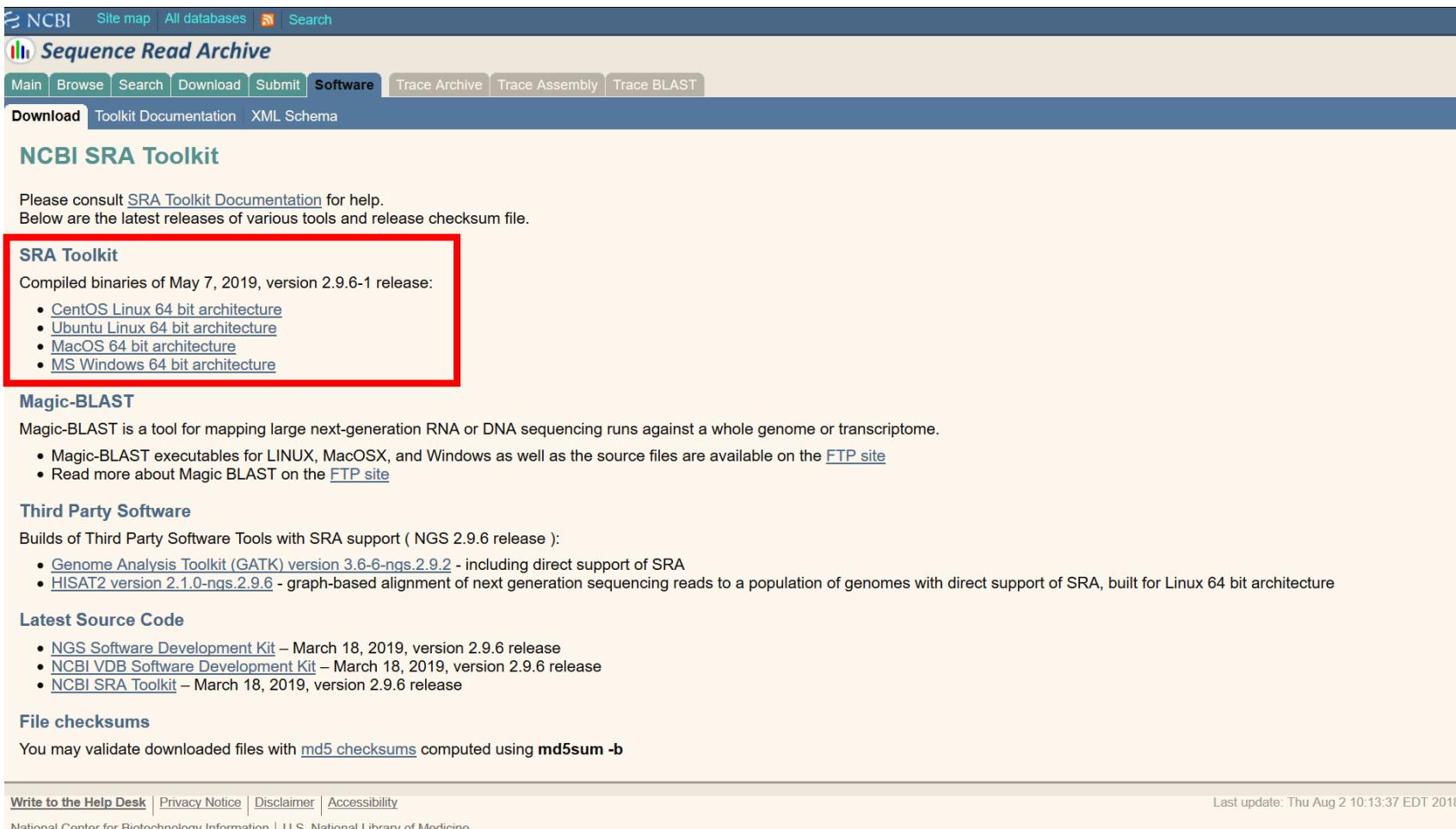
- [abi-dump](#): Convert SRA data into ABI format (csfasta / qual)
- [illumina-dump](#): Convert SRA data into Illumina native formats (qseq, etc.)
- [sff-dump](#): Convert SRA data to sff format
- [sra-stat](#): Generate statistics about SRA data (quality distribution, etc.)
- [vdb-dump](#): Output the native VDB format of SRA data.
- [vdb-encrypt](#): Encrypt non-SRA dbGaP data ("phenotype data")
- [vdb-validate](#): Validate the integrity of downloaded SRA data

Write to the Help Desk | Privacy Notice | Disclaimer | Accessibility  
National Center for Biotechnology Information | U.S. National Library of Medicine

Last update: Thu Aug 2 10:13:37 EDT 2018

NIH FIRSTGOV.gov

<https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=software>



NCBI Site map All databases Search

**Sequence Read Archive**

Main Browse Search Download Submit Software Trace Archive Trace Assembly Trace BLAST

Download Toolkit Documentation XML Schema

**NCBI SRA Toolkit**

Please consult [SRA Toolkit Documentation](#) for help.  
Below are the latest releases of various tools and release checksum file.

**SRA Toolkit**

Compiled binaries of May 7, 2019, version 2.9.6-1 release:

- [CentOS Linux 64 bit architecture](#)
- [Ubuntu Linux 64 bit architecture](#)
- [MacOS 64 bit architecture](#)
- [MS Windows 64 bit architecture](#)

**Magic-BLAST**

Magic-BLAST is a tool for mapping large next-generation RNA or DNA sequencing runs against a whole genome or transcriptome.

- Magic-BLAST executables for LINUX, MacOSX, and Windows as well as the source files are available on the [FTP site](#)
- Read more about Magic BLAST on the [FTP site](#)

**Third Party Software**

Builds of Third Party Software Tools with SRA support ( NGS 2.9.6 release ):

- [Genome Analysis Toolkit \(GATK\) version 3.6-6-nqs.2.9.2](#) - including direct support of SRA
- [HISAT2 version 2.1.0-nqs.2.9.6](#) - graph-based alignment of next generation sequencing reads to a population of genomes with direct support of SRA, built for Linux 64 bit architecture

**Latest Source Code**

- [NGS Software Development Kit](#) – March 18, 2019, version 2.9.6 release
- [NCBI VDB Software Development Kit](#) – March 18, 2019, version 2.9.6 release
- [NCBI SRA Toolkit](#) – March 18, 2019, version 2.9.6 release

**File checksums**

You may validate downloaded files with [md5 checksums](#) computed using **md5sum -b**

[Write to the Help Desk](#) | [Privacy Notice](#) | [Disclaimer](#) | [Accessibility](#) Last update: Thu Aug 2 10:13:37 EDT 2018

National Center for Biotechnology Information | U.S. National Library of Medicine



NCBI Site map All databases Search

Sequence Read Archive

Main Browse Search Download Submit Software Trace Archive Trace Assembly Trace BLAST

Download Toolkit Documentation XML Schema

## NCBI SRA Toolkit

Please consult [SRA Toolkit Documentation](#) for help.

Below are the latest releases of various tools available for download:

### SRA Toolkit

Compiled binaries of May 7, 2019, version 2.9.6

- CentOS Linux 64 bit architecture
- Ubuntu Linux 64 bit architecture
- MacOS 64 bit architecture
- MS Windows 64 bit architecture

### Magic-BLAST

Magic-BLAST is a tool for mapping large next-generation sequencing (NGS) reads against a database of reference sequences.

- Magic-BLAST executables for LINUX, Mac, and Windows
- Read more about Magic BLAST on the [FTP site](#)

### Third Party Software

Builds of Third Party Software Tools with SRA support:

- Genome Analysis Toolkit (GATK) version 3.8.0
- HISAT2 version 2.1.0-ngs.2.9.6 - graph-based aligner

### Latest Source Code

- NGS Software Development Kit – March 18, 2019
- NCBI VDB Software Development Kit – March 18, 2019
- NCBI SRA Toolkit – March 18, 2019, version 2.9.6

### File checksums

You may validate downloaded files with [md5 checksums](#) computed using `md5sum -b`

Ouverture de sratoolkit.2.9.6-1-win64.zip

Vous avez choisi d'ouvrir :

**sratoolkit.2.9.6-1-win64.zip**  
qui est un fichier de type : ZIP archive (50,4 Mo)  
à partir de : <https://ftp-trace.ncbi.nlm.nih.gov>

Que doit faire Firefox avec ce fichier ?

Ouvrir avec IZArc Archiver

Enregistrer le fichier

Toujours effectuer cette action pour ce type de fichier.

OK Annuler

Report of SRA, built for Linux 64 bit architecture

Write to the Help Desk | Privacy Notice | Disclaimer | Accessibility

National Center for Biotechnology Information | U.S. National Library of Medicine

Last update: Thu Aug 2 10:13:37 EDT 2018



## SRA Toolkit

- Décompression anywhere (pas d'installation)

**C : /SRA**



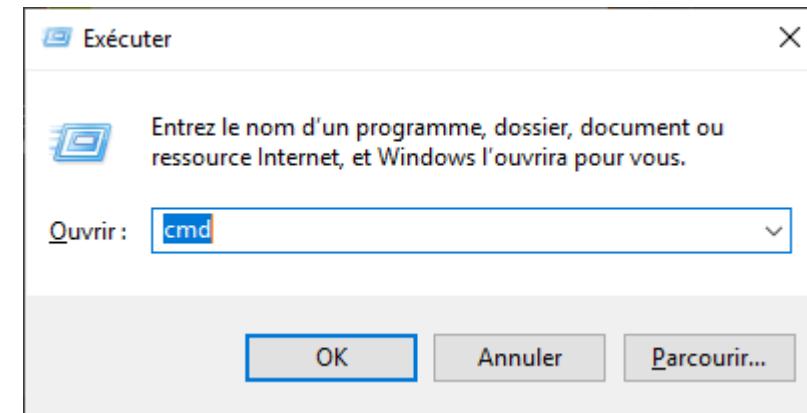
## SRA Toolkit

- Décompression anywhere (pas d'installation)
- Configuration du répertoire de récupération des données

C : / **SRA**

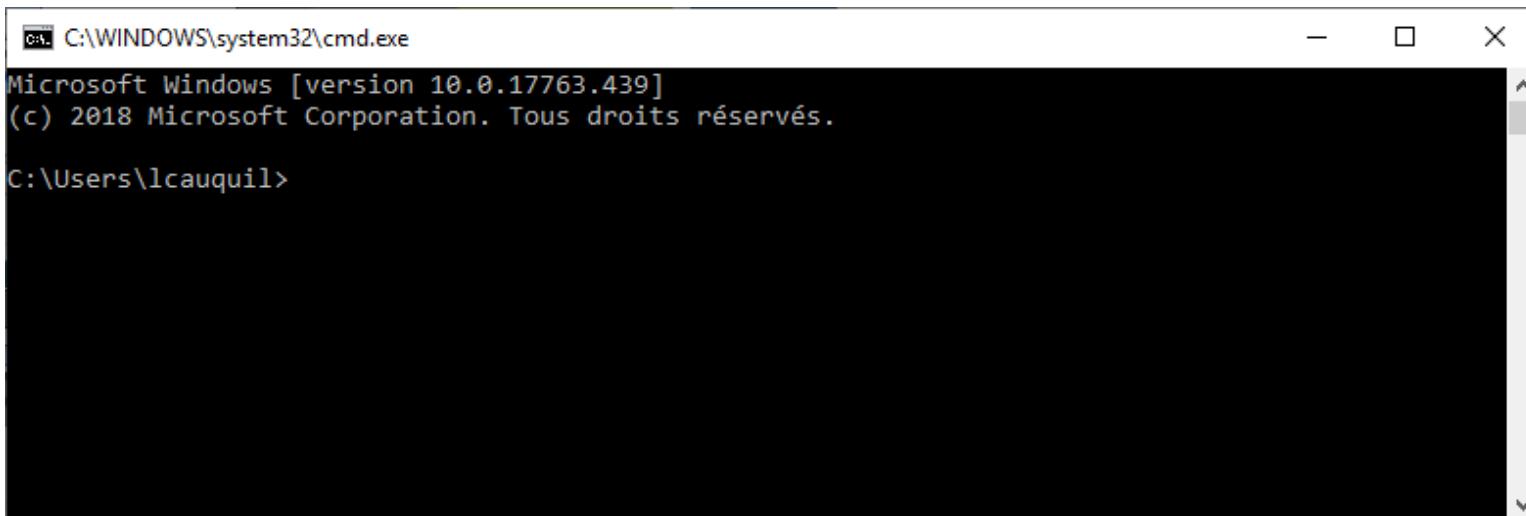
### Mode console (Windows):

-  + R
- cmd



- Décompression anywhere (pas d'installation)
- Configuration du répertoire de récupération des données

**C : /SRA**

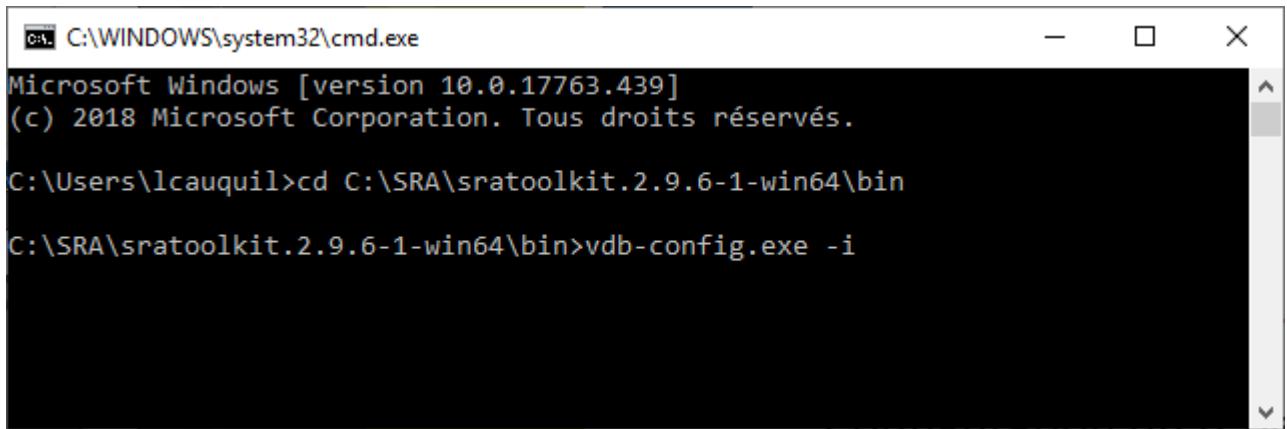


- Se déplace dans le répertoire de l'application (C:\SRA)

```
cd C:\SRA\sratoolkit.2.9.6-1-win64\bin
```

- Lance la fenêtre de configuration

```
vdb-config.exe -i
```

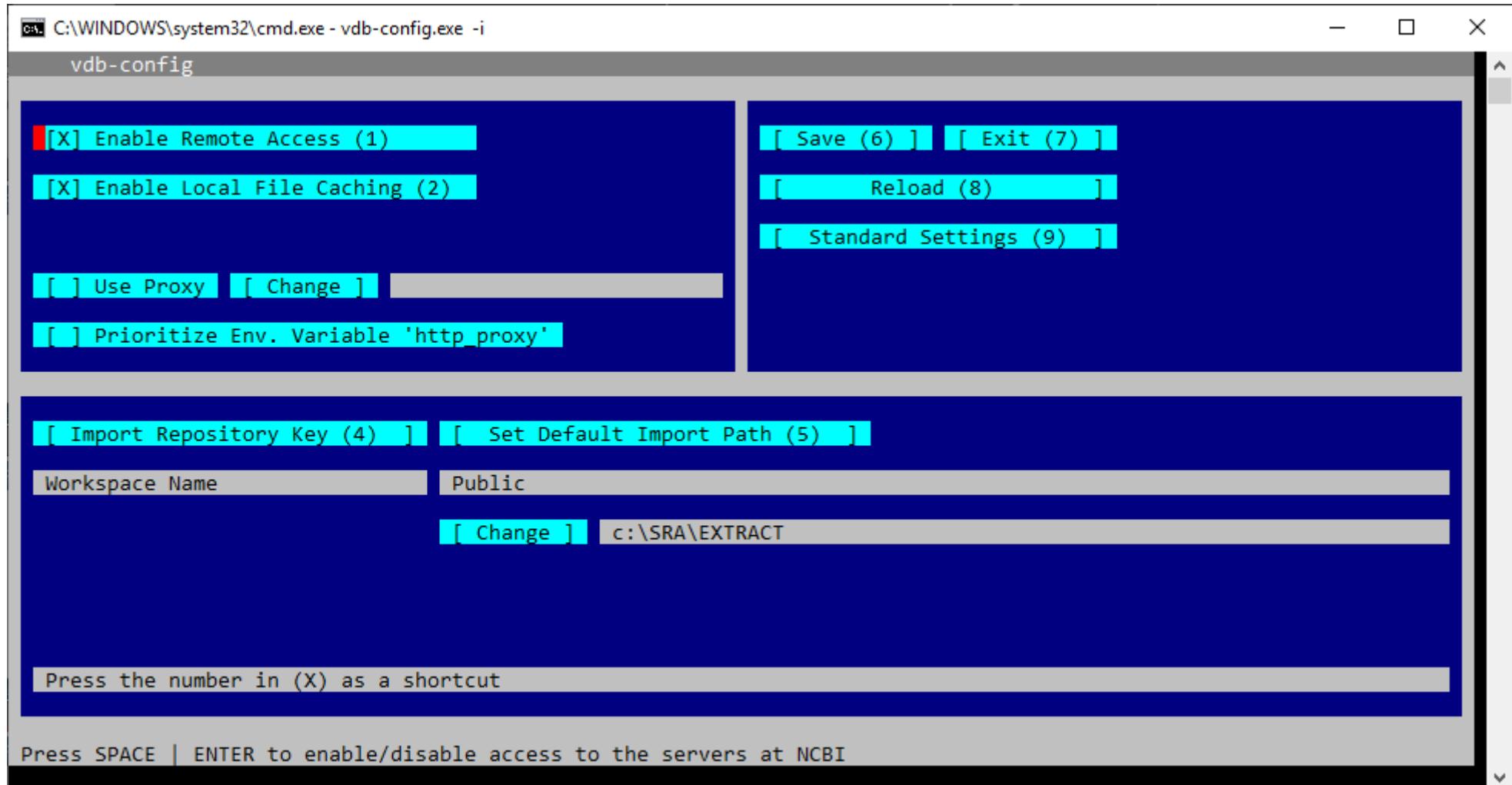


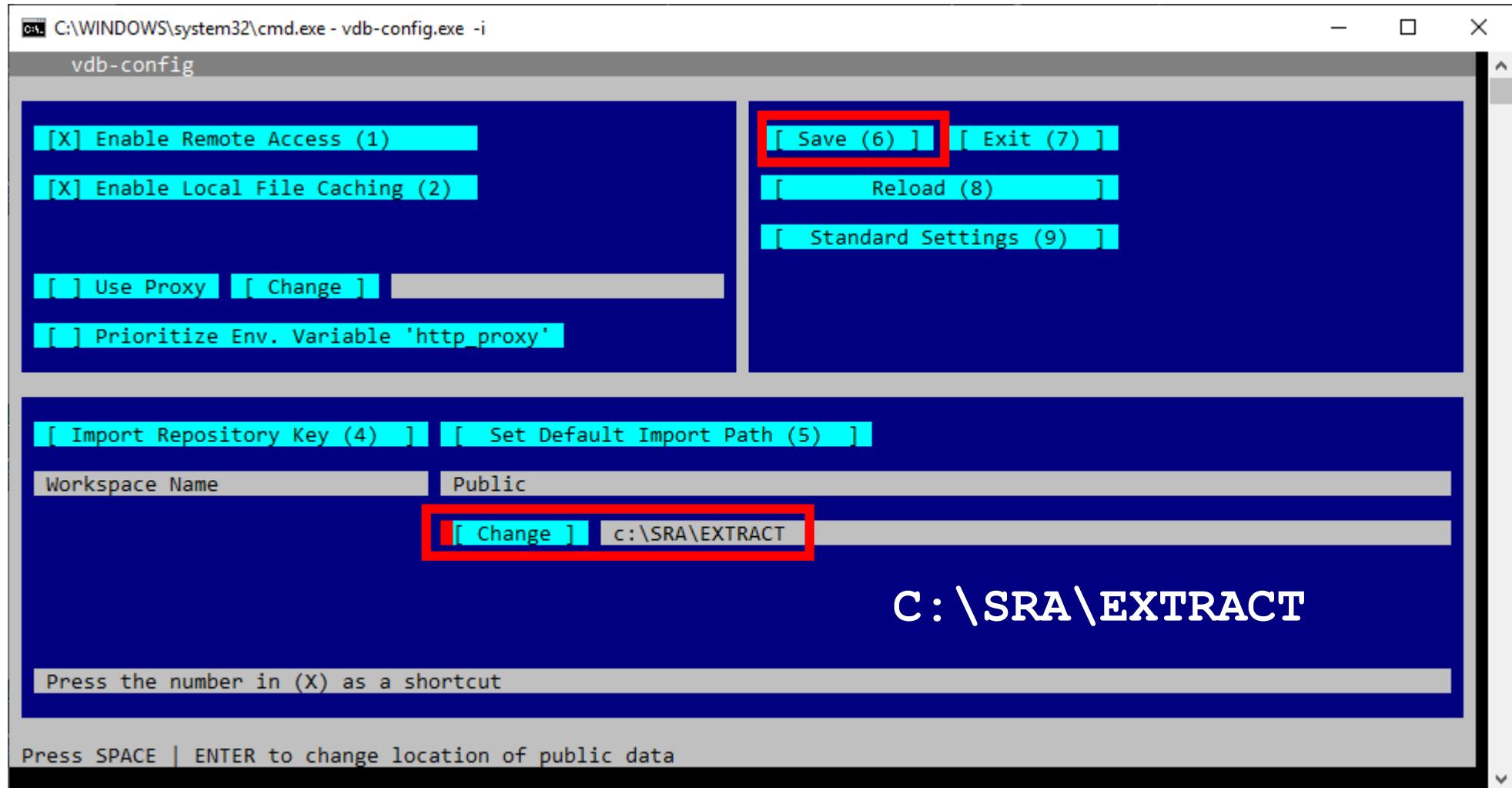
The screenshot shows a Windows Command Prompt window with the title 'C:\WINDOWS\system32\cmd.exe'. The window displays the following text:

```
Microsoft Windows [version 10.0.17763.439]
(c) 2018 Microsoft Corporation. Tous droits réservés.

C:\Users\lcauquil>cd C:\SRA\sratoolkit.2.9.6-1-win64\bin

C:\SRA\sratoolkit.2.9.6-1-win64\bin>vdb-config.exe -i
```







```
fastq-dump.exe -O C:\SRA\EXTRACT\sra
```



```
fastq-dump.exe -O C:\SRA\EXTRACT\sra
```

## Récupérer les Biosamples

- **SRR8858147**
- **SRR8858148**

```
fastq-dump.exe -O C:\SRA\EXTRACT\sra SRR8858147 SRR8858148
```

## SRA Toolkit

```
C:\WINDOWS\system32\cmd.exe
Microsoft Windows [version 10.0.17763.439]
(c) 2018 Microsoft Corporation. Tous droits réservés.

C:\Users\lcauquil>cd C:\SRA\sratoolkit.2.9.6-1-win64\bin

C:\SRA\sratoolkit.2.9.6-1-win64\bin>fastq-dump.exe -O C:\SRA\EXTRACT\sra SRR8858147 SRR8858148
Read 28008 spots for SRR8858147
Written 28008 spots for SRR8858147
Read 28146 spots for SRR8858148
Written 28146 spots for SRR8858148
Read 56154 spots total
Written 56154 spots total

C:\SRA\sratoolkit.2.9.6-1-win64\bin>
```



```
C:\WINDOWS\system32\cmd.exe
Microsoft Windows [version 10.0.17763.439]
(c) 2018 Microsoft Corporation. Tous droits réservés.

C:\Users\lcauquil>cd C:\SRA\sratoolkit.2.9.6-1-win64\bin

C:\SRA\sratoolkit.2.9.6-1-win64\bin>fastq-dump.exe -O C:\SRA\EXTRACT\sra SRR8858147 SRR8858148
Read 28008 spots for SRR8858147
Written 2
Read 2814
Written 2
Read 5615
Written 5

OS (C:) > SRA > EXTRACT > sra
Rechercher dans : sra

Nom                         Modifié le   Type   Taille
SRR8858147.fastq          17/05/2019 17:23 Fichier FASTQ 29 278 Ko
SRR8858147.sra              17/05/2019 17:23 Fichier SRA 7 404 Ko
SRR8858148.fastq          17/05/2019 17:23 Fichier FASTQ 29 422 Ko
SRR8858148.sra              17/05/2019 17:23 Fichier SRA 7 349 Ko
```

```
sri<-read.csv("C:/SRA/SraRunInfo.csv", stringsAsFactors=FALSE)
files<-basename(sri$download_path)
for(i in 1:length(files))
  download.file(sri$download_path[i], files[i])
```



ATTENTION: enlever l'extension ".1" qui peut s'ajouter aux noms dans l'objet files

```
# Assure that all the files has been downloaded successfully
# Remember, the R object files has been created in the previous code chunk
stopifnot( all(file.exists(files)) )
for(f in files) {
  cmd = paste("C:/SRA/srato toolkit.2.9.6-1-win64/bin/fastq-dump -O C:/SRA/EXTRACT/sra", f)
  cat(cmd, "\n") #print the current command
  system(cmd) # invoke command
}
```

<http://www.sthda.com/english/wiki/wiki.php?title=from-sra-to-fastq-file>

# SRA Toolkit



DONE!

Nom	Modifié le	Type	Taille
 SRR8858140.fastq	20/05/2019 09:14	Fichier FASTQ	27 986 Ko
 SRR8858140.sra	20/05/2019 09:14	Fichier SRA	7 132 Ko
 SRR8858141.fastq	20/05/2019 09:13	Fichier FASTQ	25 912 Ko
 SRR8858141.sra	20/05/2019 09:13	Fichier SRA	6 559 Ko
 SRR8858142.fastq	20/05/2019 09:13	Fichier FASTQ	31 396 Ko
 SRR8858142.sra	20/05/2019 09:13	Fichier SRA	8 064 Ko
 SRR8858143.fastq	20/05/2019 09:12	Fichier FASTQ	38 839 Ko
 SRR8858143.sra	20/05/2019 09:12	Fichier SRA	9 529 Ko
 SRR8858144.fastq	20/05/2019 09:11	Fichier FASTQ	25 393 Ko
 SRR8858144.sra	20/05/2019 09:11	Fichier SRA	6 398 Ko
 SRR8858145.fastq	20/05/2019 09:16	Fichier FASTQ	26 987 Ko
 SRR8858145.sra	20/05/2019 09:16	Fichier SRA	6 892 Ko
 SRR8858146.fastq	20/05/2019 09:15	Fichier FASTQ	23 981 Ko
 SRR8858146.sra	20/05/2019 09:15	Fichier SRA	6 004 Ko
 SRR8858147.fastq	20/05/2019 09:11	Fichier FASTQ	29 278 Ko
 SRR8858147.sra	20/05/2019 09:11	Fichier SRA	7 404 Ko
 SRR8858148.fastq	20/05/2019 09:15	Fichier FASTQ	29 422 Ko
 SRR8858148.sra	20/05/2019 09:15	Fichier SRA	7 349 Ko
 SRR8858149.fastq	20/05/2019 09:14	Fichier FASTQ	30 850 Ko
 SRR8858149.sra	20/05/2019 09:14	Fichier SRA	7 924 Ko

The NCBI SRA Software Development Kit  
Contact: [sra-tools@ncbi.nlm.nih.gov](mailto:sra-tools@ncbi.nlm.nih.gov)

NCBI Release: SRA Toolkit 2.9.6-1

May 7, 2019

prefetch: rolled back to version 2.9.3 to fix regression that prevented download of vdbcache files

NCBI Release: SRA Toolkit 2.9.6

March 18, 2019

prefetch, vfs: fixed regression that prevented re-download of incomplete files

NCBI Release: SRA Toolkit 2.9.5

March 6, 2019

prefetch: fixed regression that caused download of incomplete files

NCBI Release: SRA Toolkit 2.9.4-2

March 6, 2019

prefetch: rolled back to version 2.9.3 to fix regression that caused download of incomplete files

## Package SRAdb

- Parse all NCBI SRA metadata
- Query
- Download

# SRA Toolkit

[Home](#) [Install](#) [Help](#) [Developers](#) [About](#)

**Bioconductor**  
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

Home » Bioconductor 3.9 » Software Packages » SRAdb

## SRAdb

platforms: all rank: unknown posts: 1 / 0 / 0 / 0 in Bioc: 9 years  
build: ok updated: before release

DOI: [10.18129/B9.bioc.SRAdb](https://doi.org/10.18129/B9.bioc.SRAdb) [f](#) [t](#)

A compilation of metadata from NCBI SRA and tools

Bioconductor version: Release (3.9)

The Sequence Read Archive (SRA) is the largest public repository of sequencing data from the next generation of sequencing platforms including Roche 454 GS System, Illumina Genome Analyzer, Applied Biosystems SOLiD System, Helicos Heliscope, and others. However, finding data of interest can be challenging using current tools. SRAdb is an attempt to make access to the metadata associated with submission, study, sample, experiment and run much more feasible. This is accomplished by parsing all the NCBI SRA metadata into a SQLite database that can be stored and queried locally. Fulltext search in the package make querying metadata very flexible and powerful. fastq and sra files can be downloaded for doing alignment locally. Beside ftp protocol, the SRAdb has funcitons supporting fastp protocol (ascp from Aspera Connect) for faster downloading large data files over long distance. The SQLite database is updated regularly as new data is added to SRA and can be downloaded at will for the most up-to-date metadata.

Author: Jack Zhu and Sean Davis

Maintainer: Jack Zhu <[zhujack@mail.nih.gov](mailto:zhujack@mail.nih.gov)>

Citation (from within R, enter `citation("SRAdb")`):

Zhu Y, Stephens RM, Meltzer PS, Davis SR (2013). "SRAdb: query and use public next-generation sequencing data from within R." *BMC bioinformatics*, **14**(1), 19.

**Documentation »**

**Bioconductor**

- Package [vignettes](#) and manuals.
- [Workflows](#) for learning and use.
- [Course and conference](#) material.
- [Videos](#).
- Community [resources](#) and [tutorials](#).

[R / CRAN packages and documentation](#)

**Support »**

Please read the [posting guide](#). Post questions about Bioconductor to one of the following locations:

- [Support site](#) - for questions about Bioconductor packages
- [Bioc-devel](#) mailing list - for package developers



# Questions ?

