

---

## IFB\_Training\_salivarius

Plan de gestion de données créé à l'aide de DMP OPIDoR

**Créateur du PGD :** Hélène Chiapello

**Affiliation du créateur principal :** INRAE - Institut national de recherche pour l'agriculture l'alimentation et l'environnement

**Modèle du PGD :** ANR - Modèle de PGD (français)

**Dernière modification du PGD :** 05/01/2021

**Financier :** Agence nationale de la recherche (ANR)

### Résumé du projet :

*Streptococcus salivarius* est une bactérie de la division des Firmicutes qui colonise l'être humain juste après la naissance et est présente dans différents tissus (bouche, peau, pharynx, tube digestif, voies génito-urinaires, sang,...)

Elle est considérée comme un pathogène opportuniste : certaines souches peuvent provoquer des méningites, des abcès pancréatiques, de l'impétigo, des péritonites, de la sinusite,...

Il existe une grande diversité génomique dans l'espèce majoritairement due aux Éléments Génétiques Mobiles

**Chercheur Principal :** Hélène Chiapello

**Identifiant ORCID :** 0000-0001-5102-0632

**Contact pour les Données :** Hélène Chiapello

### Droits d'auteur

Le(s) créateur(s) de ce plan accepte(nt) que tout ou partie de texte de ce plan soit réutilisé et personnalisé si nécessaire pour un autre plan. Vous n'avez pas besoin de citer le(s) créateur(s) en tant que source. L'utilisation de toute partie de texte de ce plan n'implique pas que le(s) créateur(s) soutien(nen)t ou aient une quelconque relation avec votre projet ou votre soumission.

# IFB\_Training\_salivarius

## 1. Description des données et collecte ou réutilisation de données existantes

Question sans réponse.

Origine : échantillons humains de différents types (infant oral cavity, blood, feces, cantal, vaginal, skin, milk,...)

Type : lectures (données brutes de séquençage) au format texte

Plateforme : Illumina, Instrument : HiSeq2000

Format : fastq (format standard)

Volume : XX Go répartis sur environ 40 fichiers

**Tableau des échantillons**

tax_id	scientific_name	Strain	Origin	Isolation Source	Sampling year	Geographic Location
1304	Streptococcus salivarius	B35	lot_labo	Human blood	2010	France: Besançon
1304	Streptococcus salivarius	B50	lot_labo	Human blood	2009	France: Besançon
1304	Streptococcus salivarius	B57	lot_labo	Human blood	2009	France: Besançon
1304	Streptococcus salivarius	B63	lot_labo	Human blood	2011	France: Besançon
1304	Streptococcus salivarius	F1-4	lot_labo	Human feces	2011	France: Vandoeuvre-les-Nancy
1304	Streptococcus salivarius	F1-8	lot_labo	Human feces	2011	France: Vandoeuvre-les-Nancy
1304	Streptococcus salivarius	F4-2	lot_labo	Human feces	2011	France: Vandoeuvre-les-Nancy
1304	Streptococcus salivarius	F4-20	lot_labo	Human feces	2011	France: Vandoeuvre-les-Nancy
1304	Streptococcus salivarius	F7-1	lot_labo	feces	2011	France: Vandoeuvre-les-Nancy
1304	Streptococcus salivarius	L11	lot_labo	Human blood	2003	France: Limoges
1304	Streptococcus salivarius	L22	lot_labo	Human blood	2004	France: Limoges
1304	Streptococcus salivarius	L25	lot_labo	Human blood	2005	France: Limoges
1304	Streptococcus salivarius	L45	lot_labo	Human blood	2007	France: Limoges
1304	Streptococcus salivarius	L50	lot_labo	Human blood	2009	France: Limoges
1304	Streptococcus salivarius	L60	lot_labo	Human blood	2010	France: Limoges
1304	Streptococcus salivarius	L61	lot_labo	Human blood	2010	France: Limoges
1304	Streptococcus salivarius	L64	lot_labo	Human pleural fluid	2010	France: Limoges
1304	Streptococcus salivarius	N5	lot_labo	Human cerebral abces	2011	France: Nancy
1304	Streptococcus salivarius	N20	lot_labo	Human abdominal fluid	2010	France: Nancy
1304	Streptococcus salivarius	T93	lot_labo	Human blood	2003	France: Tours

**Tableau des données brutes de séquençage**

sample_title	instrument_platform	instrument_model	library_layout	insert_size	library_strategy	library_source	Read file 1	Read file 2	read_count	fastq_r
Streptococcus salivarius B35 strain	ILLUMINA	Illumina HiSeq 2000	PAIRED	269	WGS	GENOMIC	B35.f.fastq.gz	B35.r.fastq.gz	7651215	529a2f
Streptococcus salivarius B50 strain	ILLUMINA	Illumina HiSeq 2000	PAIRED	377	WGS	GENOMIC	B50.f.fastq.gz	B50.r.fastq.gz	16730110	5f061c
Streptococcus salivarius B57 strain	ILLUMINA	Illumina HiSeq 2000	PAIRED	345	WGS	GENOMIC	B57.f.fastq.gz	B57.r.fastq.gz	12178649	773637
Streptococcus salivarius B63 strain	ILLUMINA	Illumina HiSeq 2000	PAIRED	255	WGS	GENOMIC	B63.f.fastq.gz	B63.r.fastq.gz	5097999	3b624f
Streptococcus salivarius F1-4 strain	ILLUMINA	Illumina HiSeq 2000	PAIRED	288	WGS	GENOMIC	F14.f.fastq.gz	F14.r.fastq.gz	19730212	70c00e
Streptococcus salivarius F1-8 strain	ILLUMINA	Illumina HiSeq 2000	PAIRED	303	WGS	GENOMIC	F18.f.fastq.gz	F18.r.fastq.gz	19861928	2450a1
Streptococcus salivarius F4-2 strain	ILLUMINA	Illumina HiSeq 2000	SINGLE		WGS	GENOMIC	F4-2_CGAGGCTG-TATCCTCT_L008_R1_001.fastq.gz	3618120	ca97ec8aae211f899a2a4e02507d2f31	33040f
Streptococcus salivarius F4-20 strain	ILLUMINA	Illumina HiSeq 2000	PAIRED	284	WGS	GENOMIC	F420.f.fastq.gz	F420.r.fastq.gz	37084452	de0d8c
Streptococcus salivarius F7-1 strain	ILLUMINA	Illumina HiSeq 2000	SINGLE		WGS	GENOMIC	F7-1_AAGAGGCA-TATCCTCT_L008_R1_001.fastq.gz	6719337	04311bf4a9af62abff060294dc942e17	58019f
Streptococcus salivarius L11 strain	ILLUMINA	Illumina HiSeq 2000	PAIRED	229	WGS	GENOMIC	L11.f.fastq.gz	L11.r.fastq.gz	10321795	8f2932
Streptococcus salivarius L22 strain	ILLUMINA	Illumina HiSeq 2000	PAIRED	256	WGS	GENOMIC	L22.f.fastq.gz	L22.r.fastq.gz	8648792	36755f
Streptococcus salivarius L25 strain	ILLUMINA	Illumina HiSeq 2000	PAIRED	409	WGS	GENOMIC	L25.f.fastq.gz	L25.r.fastq.gz	17209755	9618c7
Streptococcus salivarius L45 strain	ILLUMINA	Illumina HiSeq 2000	PAIRED	251	WGS	GENOMIC	L45.f.fastq.gz	L45.r.fastq.gz	29899773	86840f
Streptococcus salivarius L50 strain	ILLUMINA	Illumina HiSeq 2000	PAIRED	283	WGS	GENOMIC	L50.f.fastq.gz	L50.r.fastq.gz	7471393	deffd71
Streptococcus salivarius L60 strain	ILLUMINA	Illumina HiSeq 2000	PAIRED	326	WGS	GENOMIC	L60.f.fastq.gz	L60.r.fastq.gz	5080937	6975e1
Streptococcus salivarius L61 strain	ILLUMINA	Illumina HiSeq 2000	PAIRED	286	WGS	GENOMIC	L61.f.fastq.gz	L61.r.fastq.gz	9354977	093a4f
Streptococcus salivarius L64 strain	ILLUMINA	Illumina HiSeq 2000	PAIRED	274	WGS	GENOMIC	L64.f.fastq.gz	L64.r.fastq.gz	4175017	bb05ef
Streptococcus salivarius N20 strain	ILLUMINA	Illumina HiSeq 2000	PAIRED	387	WGS	GENOMIC	N20.f.fastq.gz	N20.r.fastq.gz	15682551	631a6e
Streptococcus salivarius N5 strain	ILLUMINA	Illumina HiSeq 2000	PAIRED	390	WGS	GENOMIC	N5.f.fastq.gz	<a href="#">N5.r.fastq.gz</a>	18564793	7136c1
Streptococcus salivarius T93 strain	ILLUMINA	Illumina HiSeq 2000	PAIRED	289	WGS	GENOMIC	T93.f.fastq.gz	T93.r.fastq.gz	10313757	a9af16

## 2. Documentation et qualité des données

Source des métadonnées : Checklist ENA <https://www.ebi.ac.uk/ena/browser/checklists>

Trouver la mieux adaptée au jeu de données

Format XML

Organisation :

- espace projet créé sur l'infrastructure IFB avec l'organisation habituelle /shared/projects/mon\_projet
- dépôt GitHub ou Gitlab pour le partage du code et de la documentation

- contrôle de la qualité des données brutes avec fastqc
- génération d'un rapport de qualité avec multiqc

```
$ conda install -c bioconda fastqc=0.11.9
```

```
$ conda install -c bioconda multiqc=1.9
```

## 3. Stockage et sauvegarde pendant le processus de recherche

- Données brutes stockées dans un espace projet dédié sauvegardé et accessible en lecture seule de l'infrastructure IFB /shared/projects/mon\_projet
- Cet espace est sauvegardé fréquemment

Création d'un compte personnel avec une adresse institutionnelle

Validation du compte par les administrateurs

Accès par une connexion sécurisée ssh avec mot de passe

#### **4. Exigences légales et éthiques, codes de conduite**

Utilisation de données publiées (voir la publication pour plus d'infos :

Question sans réponse.

Question sans réponse.

#### **5. Partage des données et conservation à long terme**

Habituellement les données sont rendues publiques dans une banque internationale comme l'ENA au moment de la publication. Un identifiant de projet unique permet ensuite à la communauté d'accéder aux données.

C'est la banque internationale (ici l'ENA) qui se charge du stockage à long terme

un terminal ou un navigateur web

La banque internationale (ici l'ENA) se charge de la création d'un identifiant pérenne et unique

#### **6. Responsabilités et ressources en matière de gestion des données**

Le gestionnaire des données est Hélène Chiapello (data broker, Ingénieure de Recherches INRAE/IFB)

Question sans réponse.