DMP du projet "IFB_Training_salivarius"

Plan de gestion de données créé à l'aide de DMP OPIDoR, basé sur le modèle "ANR - Modèle de PGD (français)" fourni par Agence nationale de la recherche (ANR).

Renseignements sur le plan

Titre du plan	DMP du projet "IFB_Training_salivarius"
Langue	fra
Date de création	2020-12-10
Date de dernière modification	2021-01-12
Identifiant	IFB_Training_salivarius

Renseignements sur le projet

Titre du projet	IFB_Training_salivarius
Résumé	*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
Sources de financement	• Agence nationale de la recherche (ANR) :

Produits de recherche :

1. Données de séquençage de 20 génomes de S. salivarius (Jeu de données)

Contributeurs

Nom	Affiliation	Rôles
Hélène Chiapello - https://orcid.org/0000-0001-5102-0632		• Coordinateur du projet • Personne contact pour les données • Responsable du plan

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.

DMP du projet "IFB_Training_salivarius"

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

1a. Comment de nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ?

Question sans réponse.

1b. Quelles données (types, formats et volumes par ex.) seront collectées ou produites ?

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 standart)

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_selection	library_strategy	library_source	Read file 1	Read file 2	read_count
Streptococcus salivarius B35 strain	ILLUMINA	Illumina HiSeq 2000	PAIRED	269	RANDOM	WGS	GENOMIC	B35.f.fastq.gz	B35.r.fastq.gz	7651215
Streptococcus salivarius B50 strain	ILLUMINA	Illumina HiSeq 2000	PAIRED	377	RANDOM	WGS	GENOMIC	B50.f.fastq.gz	B50.r.fastq.gz	16730110
Streptococcus salivarius B57 strain	ILLUMINA	Illumina HiSeq 2000	PAIRED	345	RANDOM	WGS	GENOMIC	B57.f.fastq.gz	B57.r.fastq.gz	12178649
Streptococcus salivarius B63 strain	ILLUMINA	Illumina HiSeq 2000	PAIRED	255	RANDOM	WGS	GENOMIC	B63.f.fastq.gz	B63.r.fastq.gz	5097999
Streptococcus salivarius F1-4 strain	ILLUMINA	Illumina HiSeq 2000	PAIRED	288	RANDOM	WGS	GENOMIC	F14.f.fastq.gz	F14.r.fastq.gz	19730212
Streptococcus salivarius F1-8 strain	ILLUMINA	Illumina HiSeq 2000	PAIRED	303	RANDOM	WGS	GENOMIC	F18.f.fastq.gz	F18.r.fastq.gz	19861928
Streptococcus salivarius F4-2 strain	ILLUMINA	Illumina HiSeq 2000	SINGLE		RANDOM	WGS	GENOMIC	F4-2_CGAGGCTG-TATCCTCT_L008_R1_001.fastq.gz	3618120	
Streptococcus salivarius F4-20 strain	ILLUMINA	Illumina HiSeq 2000	PAIRED	284	RANDOM	WGS	GENOMIC	F420.f.fastq.gz	F420.r.fastq.gz	37084452
Streptococcus salivarius F7-1 strain	ILLUMINA	Illumina HiSeq 2000	SINGLE		RANDOM	WGS	GENOMIC	F7-1_AAGAGGCA-TATCCTCT_L008_R1_001.fastq.gz	6719337	
Streptococcus salivarius L11 strain	ILLUMINA	Illumina HiSeq 2000	PAIRED	229	RANDOM	WGS	GENOMIC	L11.f.fastq.gz	L11.r.fastq.gz	10321795
Streptococcus salivarius L22 strain	ILLUMINA	Illumina HiSeq 2000	PAIRED	256	RANDOM	WGS	GENOMIC	L22.f.fastq.gz	L22.r.fastq.gz	8648792
Streptococcus salivarius L25 strain	ILLUMINA	Illumina HiSeq 2000	PAIRED	409	RANDOM	WGS	GENOMIC	L25.f.fastq.gz	L25.r.fastq.gz	17209755
Streptococcus salivarius L45 strain	ILLUMINA	Illumina HiSeq 2000	PAIRED	251	RANDOM	WGS	GENOMIC	L45.f.fastq.gz	L45.r.fastq.gz	29899773
Streptococcus salivarius L50 strain	ILLUMINA	Illumina HiSeq 2000	PAIRED	283	RANDOM	WGS	GENOMIC	L50.f.fastq.gz	L50.r.fastq.gz	7471393
Streptococcus salivarius L60 strain	ILLUMINA	Illumina HiSeq 2000	PAIRED	326	RANDOM	WGS	GENOMIC	L60.f.fastq.gz	L60.r.fastq.gz	5080937
Streptococcus salivarius L61 strain	ILLUMINA	Illumina HiSeq 2000	PAIRED	286	RANDOM	WGS	GENOMIC	L61.f.fastq.gz	L61.r.fastq.gz	9354977
Streptococcus salivarius L64 strain	ILLUMINA	Illumina HiSeq 2000	PAIRED	274	RANDOM	WGS	GENOMIC	L64.f.fastq.gz	L64.r.fastq.gz	4175017
Streptococcus salivarius N20 strain	ILLUMINA	Illumina HiSeq 2000	PAIRED	387	RANDOM	WGS	GENOMIC	N20.f.fastq.gz	N20.r.fastq.gz	15682551
Streptococcus salivarius N5 strain	ILLUMINA	Illumina HiSeq 2000	PAIRED	390	RANDOM	WGS	GENOMIC	N5.f.fastq.gz	N5.r.fastq.gz	18564793
Streptococcus salivarius T93 strain	ILLUMINA	Illumina HiSeq 2000	PAIRED	289	RANDOM	WGS	GENOMIC	T93.f.fastq.gz	T93.r.fastq.gz	10313757

---

2. Documentation et qualité des données

2a. Quelles métadonnées et quelle documentation (par exemple méthodologie de collecte et mode d'organisation des données) accompagneront les 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
- dépot GitHub ou Gitlab pour le partage du code et de la documentation

---

2b. Quelles mesures de contrôle de la qualité des données seront mises en œuvre ?

- 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

3a. Comment les données et les métadonnées seront-elles stockées et sauvegardées tout au long du 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

3b. Comment la sécurité des données et la protection des données sensibles seront-elles assurées tout au long du processus de recherche ?

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

4a. Si des données à caractère personnel sont traitées, comment le respect des dispositions de la législation sur les données à caractère personnel et sur la sécurité des données sera-t-il assuré ?

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

4b. Comment les autres questions juridiques, comme la titularité ou les droits de propriété intellectuelle sur les données, seront-elles abordées ? Quelle est la législation applicable en la matière ?

Question sans réponse.

4c. Comment les éventuelles questions éthiques seront-elles prises en compte, les codes déontologiques respectés ?

Question sans réponse.

---

5. Partage des données et conservation à long terme

5a. Comment et quand les données seront-elles partagées ? Y-a-t-il des restrictions au partage des données ou des raisons de définir un embargo ?

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.

5b. Comment les données à conserver seront-elles sélectionnées et où seront-elles préservées sur le long terme (par ex. un entrepôt de données ou une archive) ?

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

---

5c. Quelles méthodes ou quels outils logiciels seront nécessaires pour accéder et utiliser les données ?

un terminal ou un navigateur web

5d. Comment l'attribution d'un identifiant unique et pérenne (comme le DOI) sera-t-elle assurée pour chaque jeu de données ?

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

6a. Qui (par exemple rôle, position et institution de rattachement) sera responsable de la gestion des données (c'est-à-dire le gestionnaire des données) ?

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

6b. Quelles seront les ressources (budget et temps alloués) dédiées à la gestion des données permettant de s'assurer que les données seront FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable) ?

Question sans réponse.