

HMSN204

TD Banques de données biologiques

Interrogation des banques via GQuery

1. Interrogation des banques via GQuery

Donner les requêtes permettant de répondre aux questions des exercices suivants en utilisant la syntaxe de GQuery.

1.1 Exercice 1 : Séquences

Les requêtes portent sur la consultation de la rubrique Nucleotide et traitent du gène SRY qui est impliqué dans le déterminisme génétique sexuel notamment chez les vertébrés. Vous trouverez en annexe des exemples de champs sur lesquels la consultation est possible pour la rubrique Nucleotide.

1. Donner les séquences nucléotidiques associées au gène SRY
2. Donner les séquences associées au gène SRY qui sont soit humaine (Human ou Homo sapiens), soit de souris (Mouse ou Mus musculus)
3. Donner les séquences humaines ou de souris, ayant une taille comprise entre 500 et 1000 pb et qui sont associées au gène SRY
4. Donner les séquences de chat (cat ou Felix catus) qui ne sont pas associées au gène SRY
5. Donner les séquences d'ARNm de mammifères qui sont associées au gène SRY
6. Donner les séquences de référence (banque REFSEQ) d'ARNm humaines référencées qui sont associées au gène SRY et qui sont décrites par une région CDS et une région génique (gene)

1.2 Exercice 2 : Publications

Les requêtes portent sur la consultation de la rubrique PubMed et donc de Medline. Vous trouverez en annexe des exemples de champs sur lesquels la consultation est possible pour la rubrique PubMed.

1. Donner les références bibliographiques parues dans les journaux Nature et Science
2. Donner les références bibliographiques ayant SF Altschul comme auteur
3. Donner les références bibliographiques publiées dans Nature entre 2000 et 2010 et ayant le terme SRY et le terme Mammalian dans leur corps de texte
4. Donner les références bibliographiques ayant SRY dans leur titre, rédigées en Français et qui sont disponibles en accès libre

1.3 Exercice 3 : Gènes

Les requêtes portent sur la consultation de la rubrique Gene. Vous trouverez en annexe des exemples de champs sur lesquels la consultation est possible pour la rubrique Gene.

1. Restituer les informations concernant les gènes humain ayant un symbole qui commence par ABC
2. Restituer tous les gènes présents sur le chromosome 9 humain qui ont une activité ATP-binding
3. Restituer tous les gènes présents sur le chromosome Y qui sont référencés dans la banque de données OMIM et qui sont donc impliqués dans des maladies liées à l'hérédité

2. Annexes

2.1 Rubriques Sequence

Quelques champs sur lesquels la consultation est permise pour les rubriques Sequence (Nucleotide et Protein) sont listés dans le tableau ci-dessous.

Nom du champ	Abréviation	Définition	Exemples
Accession	ACCN	numéro d'accèsion NCBI	AF123456[ACCN]
Feature Key	FKEY	caractéristique biologique décrite au sein de la "Feature Table"	CDS[FKEY]
Gene Name	GENE	noms des gènes décrits par les séquences	BRCA1[GENE]
Organism	ORGN	noms (binomiaux et communs) pour l'ensemble de la taxonomie	human[ORGN]
Properties	PROP	différentes propriétés : type moléculaire, B.D origine	biomol_mrna[PROP]
Sequence Length	SLEN	taille totale de la séquence	100 :1000[SLEN]
Title	TI	élément textuel présent dans la définition	binding [TI]

2.2 Rubrique PubMed

PubMed est l'interface de consultation publique de la banque de références bibliographiques MEDLINE. Une sélection de champs à partir desquels des requêtes peuvent être construites sont listés ci-dessous.

Nom du champ	Abréviation	Définition	Exemples
Author	AU	Auteur de la publication	"Kornberg RD" [AU]
Publication Data	PDAT	Date de publication	2000[PDAT] ou 2000 :2010 [PDAT]
Journal	TA	Journal	Nature [JOURNAL]
Language	LA	Langue de publication	french [LA]
Title	TI	Titre de la publication	DNA [TI]
Text words	TW	Terme dans le corps du texte	SRY [TW]
Filter	SB	différents types de filtres	free full text [SB]
Publication Type	PT	Review, Biography, Letter, Clinical Trial, ...	review [PT]
All Fields	[ALL]	tous les champs	SRY [ALL]

2.3 Rubrique Gene

La rubrique Gene est porteuse de très nombreuses informations et va notamment renseigner sur la structure des gènes et le répertoire génique de nombreux organismes vivants.