

Determinació de les proteïnes multifuncionals de *Mycoplasma genitalium*

Neus Paredes Felipe

Àrea 2

Màster en Bioinformàtica i Bioestadística

Luis Franco Serrano

Carles Ventura Royo



Aquesta obra està subjecta a una llicència de
Reconeixement

<https://creativecommons.org/licenses/by-nc/3.0/es/>

FITXA DEL TREBALL FINAL

Títol del treball:	Determinació de les proteïnes multifuncionals de <i>Mycoplasma genitalium</i>
Nom de l'autor:	<i>Neus Paredes Felipe</i>
Nom del consultor/a:	<i>Luis Franco Serrano</i>
Nom del PRA:	<i>Carles Ventura Royo</i>
Data de lliurement:	2 de juny del 2022
Titulació:	<i>Màster en Bioinformàtica i Bioestadística</i>
Àrea:	<i>Àrea 2</i>
Idioma del treball:	Català
Nombre de crèdits:	15
Paraules clau:	proteïnes moonlighting, proteïnes multifuncionals, <i>Mycoplasma genitalium</i>

Resum

Les proteïnes moonlighting són proteïnes amb més d'una funció. Moltes estan implicades en malalties humanes i en són dianes de fàrmacs actuals. També s'ha vist que estan implicades en la virulència de microorganismes patògens. En aquest treball s'ha estudiat les proteïnes moonlighting d'un microorganisme patògen causant de malalties de transmissió sexual com la uretritis en homes i la cervicitis en dones, i associada a altres: *Mycoplasma genitalium*. L'anàlisi s'ha realitzat mitjançant eines bioinformàtiques, combinant cerques d'homologia remota (PSI-BLAST) amb bases de dades d'interactòmica (APID) juntament amb la recerca a altres bases de dades de proteïnes com UniProt i InterPro, ja que s'ha vist que era el mètode més eficaç en aquest cas. S'ha analitzat un 24.8% del proteoma conegut d'aquest bacteri i un 9.17% d'aquestes proteïnes analitzades han resultat ser moonlighting. Una d'elles és un factor de virulència conegut, fet que pot ser d'especial interès per a crear una vacuna contra aquest bacteri. En un futur, si s'acaba analitzant tot el proteoma de *M. genitalium* és molt probable que es descobreixin més proteïnes moonlighting que estiguin implicades en la virulència d'aquest patògen. Ara bé, per facilitar aquesta tasca es podria millorar les eines bioinformàtiques que hi ha actualment per a realitzar aquest tipus d'anàlisi i fer-lo més eficaç.

Abstract

Moonlighting proteins are proteins with more than one function. Many of them are involved in human disease and are targeted by current drugs. They are also involved in the virulence of pathogenic microorganisms. In this project it has been studied the moonlighting proteins of a pathogenic microorganism that causes sexually transmitted diseases such as urethritis in men and cervicitis in women, and associated with others: *Mycoplasma genitalium*. The analysis was performed using bioinformatics tools, combining remote homology searches (PSI-BLAST) with interactomics databases (APID) along with searching other protein databases such as UniProt and InterPro, as it has been shown to be the most effective method in this case. 24.8% of the known proteome of this bacterium has been analyzed and 9.17% of these proteins analyzed have been moonlighting. One of them is a known virulence factor, which may be of particular interest in creating a vaccine against this bacterium. In the future, if the entire *M. genitalium* proteome is to be analyzed, it is very likely that more moonlighting proteins will be discovered that are involved in the virulence of this pathogen. However, to facilitate this task, it would be useful to improve the current bioinformatics tools to carry out this type of analysis and make it more effective.

ÍNDEX

Resum

Abstract

1. Introducció	2
2. Estat de l'art	7
3. Metodologia	8
4. Resultats	17
5. Discussió	37
6. Conclusions	39
7. Glossari	42
8. Bibliografia	43
9. Annexos	46

Llista de figures

Figura 1. Diagrama de Gantt

Figura 2. Exemple de l'estructura de les carpetes per a la configuració de la base de dades.

Figura 3. Exemple de com accedir a les variables d'entorn per a configurar-les

Figura 4. Exemple de com configurar la nostra base de dades com a nova variable d'entorn.

Figura 5. Exemple de base de dades creada amb Microsoft SQL Server Management Studio 18 a partir dels resultats de PSI-BLAST d'una seqüència de proteïna.

Llista de taules

Taula 1. Resultats PSI-BLAST.

Taula 2. Possibles funcions moonlighting.

Taula 3. "Verdaders positius", proteïnes moonlighting.

Taula 4. Llista de proteïnes de *Mycoplasma genitalium*.

1 Introducció

1.1 Context i justificació del treball

Les proteïnes les acostumem a relacionar amb una sola funció, però s'ha vist que hi ha proteïnes que poden tenir dos o més funcions en una mateixa cadena polipeptídica: les proteïnes moonlighting o multitasking. Actualment se'n tenen catalogades unes 700 i es calcula que un 78% d'aquestes tenen una implicació en malalties humanes i un 48% en són dianes de fàrmacs actuals [1]. A més, s'ha vist que un 25% d'aquestes proteïnes estan implicades en la virulència de microorganismes patògens, gran part en el mecanisme d'adhesió a l'hoste. En aquest treball, es vol estudiar les proteïnes multifuncionals d'un microorganisme patògen causant de malalties de transmissió sexual com la uretritis en homes i la cervicitis en dones, i associada a altres: *Mycoplasma genitalium* [2]. Identificar aquestes proteïnes seria un avenç en l'àmbit de la biomedicina, ja que en permetria conèixer millor el mecanisme d'acció d'aquest patògen i, per tant, millorar-ne el tractament per controlar-lo. A més, permetria ampliar el llistat de proteïnes multifuncionals identificades fins l'actualitat, que és mínim comparat amb el que s'espera descobrir en un futur [3].

M. genitalium té un genoma molt petit, ja que també té molt pocs gens implicats en la biosíntesi [2], fet que el converteix en un bacteri exigent. Al ser tant petit significa que és més senzill conèixer tot el seu genoma i proteoma. Això pot facilitar el descobriment de factors de virulència i l'estudi de la seva patogenicitat, que també pot presentar certa similitud a la de molts altres patògens. Aquest bacteri té molts factors de virulència responsables de la seva patogenicitat dels quals se'n coneixen 20 [4], tot i que encara n'hi ha moltes incògnites [5]. Per tal de poder millorar el tractament contra aquest patògen és necessari entendre el funcionament i implicació d'aquests factors. Està clar que les proteïnes multifuncionals contribueixen en la virulència d'un rang important de patògens, incloent *M. genitalium* [6]. És per això que, en aquest treball, es vol identificar aquest tipus

de proteïnes d'aquest microorganisme. Per tal d'aconseguir-ho, s'espera trobar i usar un mètode eficaç.

1.2 Objectius del treball

1.2.1 Objectius generals:

- A. Identificar les proteïnes multifuncionals de *M. genitalium*.

1.2.2 Objectius específics:

- B. Aconseguir un llistat de totes les proteïnes conegudes de la soca escollida de *M. genitalium*, juntament amb la seva funció canònica.
- C. Obtenir un llistat de les possibles segones funcions de cada proteïna de *M. genitalium*.
- D. Obtenir un llistat de “verdaders positius” de les proteïnes multifuncionals, i les seves respectives funcions, de *M. genitalium*.

1.3 Enfoc i mètode seguit

La majoria de proteïnes multifuncionals, fins ara, s'han descobert experimentalment per “accident”. Ara bé, utilitzant la bioinformàtica hi ha altres mètodes que ens poden ser d'utilitat. Es poden determinar noves proteïnes multifuncionals mitjançant cerques d'homologia remota amb altres proteïnes amb funcions diferents i conegudes, amb la detecció de motius o dominis funcionals, cerques en bases de dades d'interactòmica o també estudiant la relació co-evolutiva entre aminoàcids amb anàlisis de correlació de mutacions.

L'anàlisi d'homologia remota es pot fer utilitzant el programa no-redundant PSI-BLAST [7]. Aquest programa és especialment adequat per revelar proteïnes multifuncionals perquè pot identificar trams de residus d'aminoàcids conservats de diferents dominis

[8, 9]. No obstant, aquest ordena els resultats d'acord amb les seves puntuacions matemàtiques i no les biològiques, llavors els "verdaders" resultats no es troben a les posicions més altes, sinó a les més baixes. Per a solucionar aquest problema es recomanava utilitzar el programa ByPass [10] introduint-hi els resultats de PSI-BLAST [3], ara bé, s'ha vist que aquest programa ja no és accessible, per tant, s'ha hagut de realitzar l'anàlisi dels resultats de PSI-BLAST manualment. Tanmateix, segons [11], ByPass tampoc et garantia que sortissin les funcions canòniques i moonlighting a les primeres posicions i, per tant, no era el millor mètode.

L'anàlisi de motius/dominis es podia fer utilitzant InterPro [12] o Blocks [13]. Actualment Blocks no funciona. De totes formes, en general, es recomanava utilitzar InterPro, ja que Blocks no havia sigut actualitzat des del 2006. Ara bé, Blocks era una base de dades no curada, per tant, podies trobar indicis de més d'una funció (característica de les proteïnes multifuncionals), fet que amb altres programes més actuals com InterPro no passa perquè estan programats per trobar el motiu/domini principal.

Mitjançant un servidor amb dades d'interaccions proteïna-proteïna (PPI) com APID [14, 15] també es poden descobrir proteïnes multifuncionals [11]. Aquest servidor col·lecciona interactomes de proteïnes, 400 d'ells basats en interaccions físiques proteïna-proteïna validades experimentalment. A més, integra bases de dades d'interactòmica com IntAct, Mint, Human Protein Reference Database, BioGRID, BIND, DIP; estructures 3D resoltes experimentalment (PDB); anotacions de Gene Ontology, Reactome, Pfam i InterPro; entre altres [14].

Realitzar un anàlisi de correlació de mutacions, juntament amb informació estructural ens pot ajudar a localitzar zones funcionals i com ha arribat a sorgir evolutivament una segona funció d'una proteïna en concret. Ara bé, per fer un anàlisi de correlació de mutacions es necessiten unes condicions molt específiques; requereix l'existència d'una família multialineada de seqüències de proteïnes.

Alguns autors [3,11] afirmen que el mètode més efectiu actual per a trobar proteïnes multifuncionals és combinant les cerques d'homologia remota (PSI-BLAST i ByPass) amb la utilització de bases de dades de PPI, però tenint en compte que ja no es pot utilitzar ByPass, en aquest treball s'ha realitzat l'anàlisi dels resultats de PSI-BLAST manualment.

1.4 Planificació del treball

Per a realitzar el treball ha sigut necessari més d'un ordinador, un disc dur extern d'un terabit i accés a internet. Les tasques realitzades són les següents:

1.4.1 Tasques:

Paquet de treball 1 (Objectiu específic 1):

1. Triar la soca de *M. genitalium* en què es basa el treball.
2. Buscar i anotar el nom i funció canònica de totes les proteïnes conegudes de la soca escollida de *M. genitalium* mitjançant UniProt [16].

Paquet de treball 2 (Objectiu específic 2):

3. Buscar les seqüències de les proteïnes de la soca escollida de *M. genitalium* mitjançant UniProt.
4. Fer una cerca d'homologia remota mitjançant PSI-BLAST de totes les seqüències de la soca escollida de *M. genitalium*.
5. Analitzar les funcions predites per PSI-BLAST i crear una llista amb les possibles segones funcions de cada proteïna juntament amb la búsqueda d'interaccions amb APID.

Paquet de treball 3 (Objectiu específic 3):

6. Deliberar quines de les proteïnes tenen més d'una funció i quines són aquestes funcions.

1.4.2 Calendari:

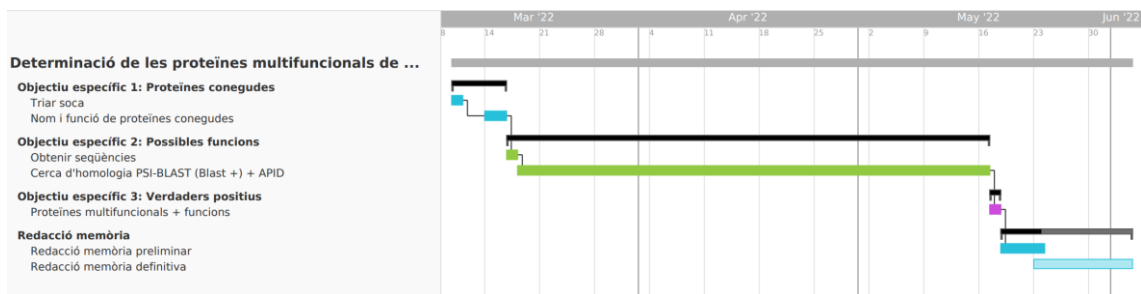


Figura 1. Diagrama de Gantt. Realitzat a <https://www.teamgantt.com/?fpr=phil81>.

1.5 Breu sumari dels productes obtinguts

S'ha obtingut un llistat de totes les proteïnes conegudes de *M. genitalium*, juntament amb les seves seqüències d'aminoàcids, un altre llistat de possibles funcions moonlighting d'aquestes proteïnes i, finalment, una llista definitiva de proteïnes multifuncionals d'aquest organisme.

1.6 Breu descripció dels altres capítols de la memòria

Seguidament s'explica l'estat de l'art de les proteïnes multifuncionals en general i els mètodes bioinformàtics que hi ha per predir-les, és a dir, el que es coneix en l'actualitat sobre el tema del treball. Després es procedeix a explicar els mètodes aplicats i els resultats obtinguts. Tot seguit hi ha la discussió, les conclusions, que és la part més rellevant del treball. Finalment, hi ha el glossari i la bibliografia consultada per a aquest estudi.

2 Estat de l'art

Avui dia s'està "començant" a descobrir la importància de les proteïnes multifuncionals, ja que en general s'acostumen a focalitzar els estudis en la funció canònica de les proteïnes. Ara bé, s'ha vist que un 25% d'aquestes proteïnes multifuncionals catalogades estan implicades en la virulència de microorganismes patògens. Es calcula que un 78% d'aquestes proteïnes catalogades estan involucrades en malalties humanes i un 48% en són dianes de fàrmacs actuals. Tot i així, aquests percentatges són molt orientatius perquè realment encara queda molt per a descobrir. Això indica que fins i tot podrien tenir més importància del que es pensa actualment.

Aquest treball es focalitza en trobar les proteïnes multifuncionals del bacteri patògen *Mycoplasma genitalium*, ja que, a part de provocar una malaltia de transmissió sexual (uretritis en homes i cervicitis en dones), és el segon organisme amb el genoma més petit i seqüenciat i, per tant, suposa un exemple de genoma mínim a analitzar. Així doncs, amb aquest treball s'espera aportar nous coneixements (proteïnes multifuncionals) per a millorar l'estudi de la virulència d'aquest bacteri i un bon mètode per trobar les seves proteïnes multifuncionals amb eines bioinformàtiques.

3 Metodologia

3.1 Tria de la soca

Es coneixen 5 soques de *Mycoplasma genitalium*, segons el NCBI (National Center for Biotechnology Information): G-37, M2288, M2321, M6282 i M6320. La soca G-37, també anomenada ATCC 33530 per la American Type Culture Collection, és la soca de referència d'aquesta espècie, la més ben descrita i de la que se'n coneix més proteïnes. Aquest fet es reflecteix en les bases de dades de proteïnes com UniProt: per la soca G-37 hi ha 518 sortides, per la M2288 n'hi ha 5, per la M6282 n'hi ha 4 i per la M6320 només 1. El seu genoma es va seqüenciar sencer el 1995 [17] i se n'han predit 20 factors de virulència, que són candidats competents pel desenvolupament d'una vacuna per a prevenir la infecció de *M. genitalium* [4].

3.2 Anotació del nom i funció canònica de les proteïnes i obtenció de seqüències

Per a poder fer els anàlisis, s'han obtingut els noms i funcions canòniques de les proteïnes conegudes de *M. genitalium* des de la pàgina web de UniProt (<https://www.uniprot.org/>) en format Excel com a document comprimit. S'ha realitzat el mateix procediment per obtenir les seqüències d'aminoàcids de totes aquestes proteïnes però en format FASTA "(canonical)" també com a document comprimit.

3.3 Cerca d'homologia remota

Per a la cerca d'homologia remota es volia utilitzar el programa PSI-BLAST en línia que es troba a la web de National Library of Medicine [7], però per no saturar el servidor s'ha utilitzat el programa BLAST+ [18]. És una eina basada en comandes que serveix per fer les cerques des del teu propi servidor sense restriccions de mida, volum, ni de bases

de dades. És a dir, et permet crear la teva pròpia base de dades i fer les cerques sense necessitat d'estar en-línia. Una eina molt útil per analitzar grans volums de seqüències, com és el cas.

3.3.1 Descàrrega de BLAST+ i de la base de dades

Per tal d'obtenir el programa s'ha descarregat l'última versió ("LATEST") de BLAST+ que es troba en l'apartat dels executables de la pàgina web <https://ftp.ncbi.nlm.nih.gov/blast/>.

La base de dades necessària també s'ha obtingut a partir de la mateixa pàgina web, dins l'apartat de "db". Els documents descarregats tenen el nom de nr.##.tar.gz, on "##" fa referència al número del document. Aquests corresponen a una col·lecció de seqüències de proteïnes amb entrades de GenPept, Swissprot, PDB, PRF, PIR i el projecte NCBI Reference Sequence (RefSeq) [19]. Un cop descarregats s'han descomprimit.

3.3.2 Configuració BLAST+ i variables d'entorn

Tot seguit, per a que el programa detectés la base de dades com a tal s'ha hagut de crear una carpeta amb el nom de "db" dins de la carpeta del programa anomenada "blast-BLAST_VERSION+". Dins la carpeta "db" s'hi ha creat una altra carpeta anomenada "nr" que agrupa tots els documents descarregats per la base de dades (Figura 2).

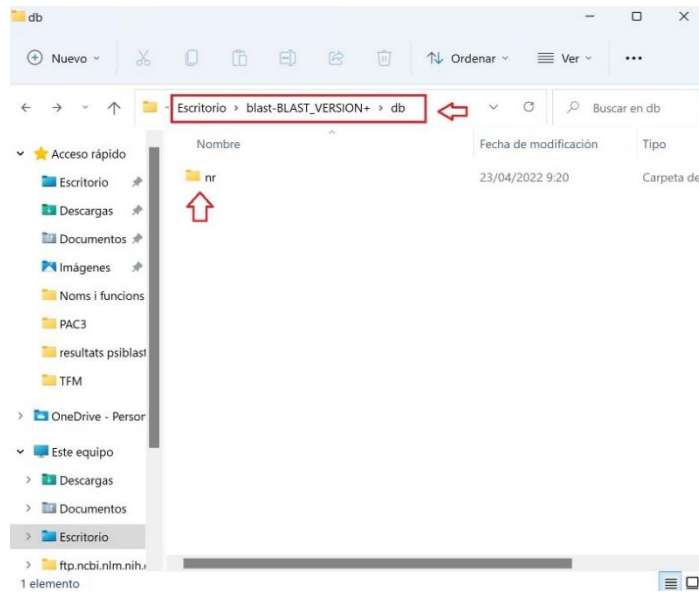


Figura 2. Exemple de l'estructura de les carpetes per a la configuració de la base de dades. Font: creació pròpia.

Després s'ha configurat les variables d'entorn. S'ha buscat "variables d'entorn" en el buscador de Windows 11 i al obrir-se una pestanya anomenada "Propietats del sistema" s'ha clicat a "Variables d'entorn" (Figura 3).

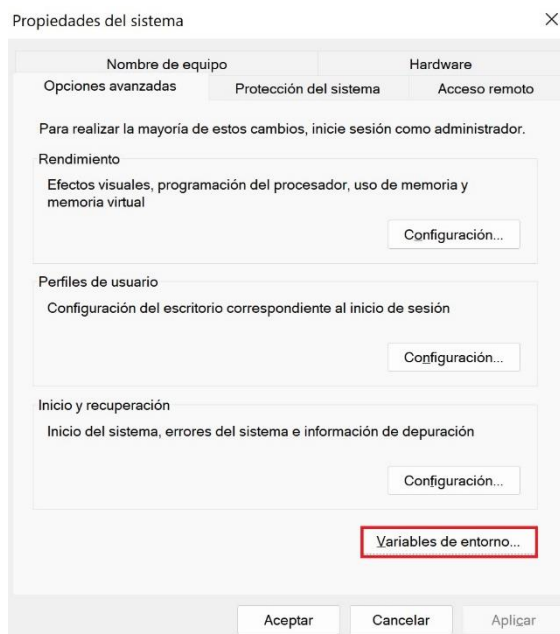


Figura 3. Exemple de com accedir a les variables d'entorn per a configurar-les. Font: creació pròpia.

Tot seguit s'ha creat una nova variable d'usuari anomenada "BLASTDB" amb el directori de la carpeta "nr" i s'ha clicat "Aceptar", tal i com es mostra a la Figura 4.

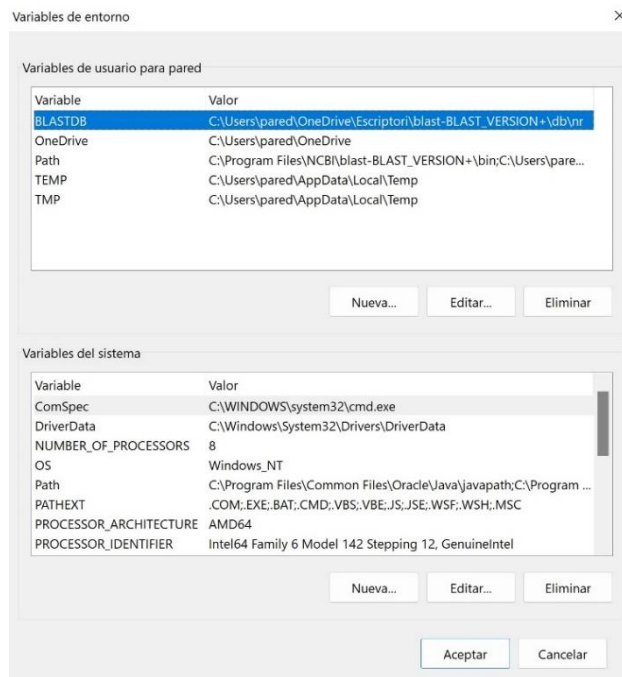


Figura 4. Exemple de com configurar la nostra base de dades com a nova variable d'entorn. Font: creació pròpia.

3.3.3 Preparació de seqüències

Per a analitzar les seqüències amb PSI-BLAST és necessari que cada una es trobi en un document en format FASTA per separat. Així que s'ha utilitzat el programa IDLE (Python 3.10) [20] per a la creació de documents amb extensió “.py” amb scripts en el llenguatge de programació Python per crear automàticament una carpeta per cada seqüència de d'aminoàcids en format FASTA. L'script utilitzat ha sigut el següent:

```

import sys
infile = open("file.fa")
outfile = []

num = 0
for line in infile:
    if line.startswith(">"):
        if (outfile != []): outfile.close()

        num = num + 1
        num = str(num)
        filename = "input" + num + ".fa"
        num = int(num)
        outfile = open(filename,'w')
        outfile.write(line)
    else:
        outfile.write(line)
outfile.close()

```

Un cop creat el document amb extensió “.py” s’ha executat des de la finestra de sol·licitud d’ordres de Windows (“command prompt” en anglès) amb la següent comanda, sempre tenint en compte que ens trobàvem al mateix directori que la carpeta amb extensió “.py”, en aquest cas “Separar.py”:

```
python Separar.py file.fa
```

3.3.4 Automatització de la creació de comandes

L’eina BLAST+ s’utilitza mitjançant comandes, per tant, per a poder analitzar cada proteïna amb PSI-BLAST es necessita una comanda diferent. S’ha creat un document executable per a cada comanda de forma automatitzada per no haver de fer-ho manualment. L’script utilitzat és el següent:

```

import os
# Primer de tot s'ha indicat el directori que ens interessa
os.chdir('/Users/pared/OneDrive/Escriptori/blast-BLAST_VERSION+')
number= 0
for number in range(9,20):
    number = number +1
    number = str(number)
    comanda = "psiblast -query input"+number+".fa -gapextend 1 -gapopen 11 -num_iterations 5 -evaluate 10 -out
output"+number+".txt -db nr"
    arxiu="comanda" + number + ".cmd"
    number = int(number)
    print(comanda, file=open(arxiu, "a"))

```

Un exemple de comanda resultant és la següent:

```

C:\Users\pared\OneDrive\Escriptori\blast-BLAST_VERSION+>psiblast -query input100.fa -gapextend 1 -gapopen 11 -
num_iterations 5 -evaluate 10 -out output100.txt -db nr

```

On “input100.fa” és el document on es troba la seqüència número 100 en format FASTA, i el document “output100.txt” on s’hi guarden els resultats de PSI-BLAST d’aquesta seqüència. S’han utilitzat els mateixos paràmetres que Hernández et al. (2015): -gapextend: 1, -gapopen: 11, -num_iterations: 5 i -evaluate 10. “gapextend” representa el cost per estendre un buit, “gapopen” el cost per obrir un buit , “num_iterations” és el nombre d’iteracions desitjades fins a la convergència i l’“evaluate” és el valor esperat (E) [21].

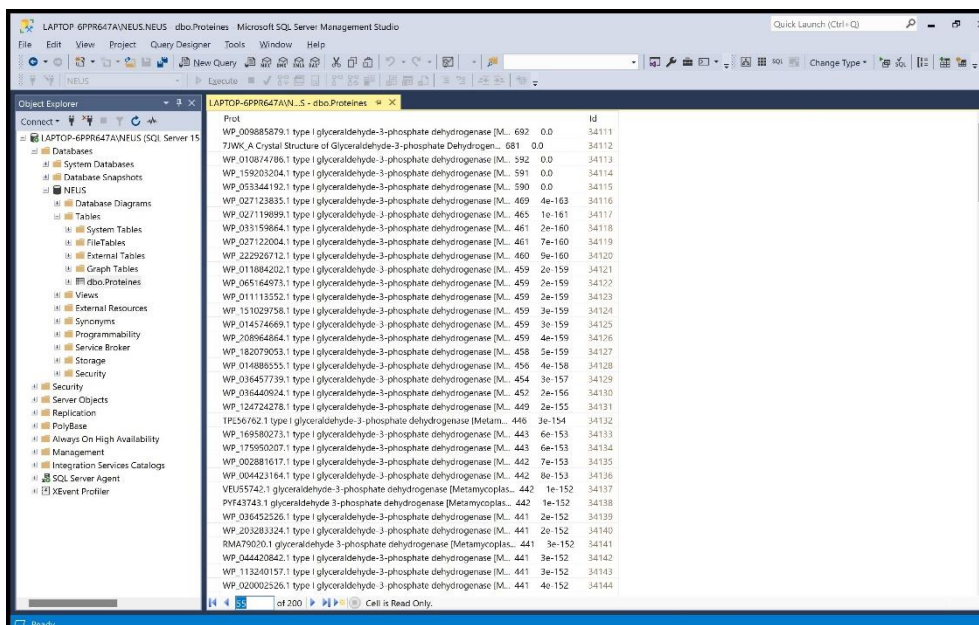
3.3.5 Recursos per l’optimització del temps

Per tal de no haver d’estar pendent tota l’estona de si una seqüència ja havia finalitzat o no per començar amb la següent, s’ha obert aproximadament 10 documents executables simultàniament. A més, per agilitzar el procés, s’han configurat un total de 3 ordinadors per fer l’anàlisi de les diferents proteïnes simultàniament. Com que la base de dades necessària és molt pesada, s’ha trobat que algun ordinador no tenia prou

emmagatzematge disponible, per tant, s'ha utilitzat un disc dur extern d'un terabit per a configurar-hi la base de dades i variables d'entorn.

3.3.6 Creació de consultes als resultats de PSI-BLAST

Els resultats de PSI-BLAST solen ser molt repetitius i feixucs d'analitzar. Amb el llenguatge SQL es pot crear un filtre per eliminar les sortides repetides i fer l'anàlisi més senzill. Per a aquesta tasca s'ha utilitzat Microsoft SQL Server Management Studio 18 [22]. Primer de tot s'ha creat una base de dades per a cada resultat de PSI-BLAST analitzat.



The screenshot shows the Microsoft SQL Server Management Studio interface. The Object Explorer on the left shows a server instance named 'LAPTOP-6P9R647A\NEUS' with a database named 'dbo.Proteines'. The main window displays a table with the following columns: 'Prot', 'Id', and 'Id'. The table contains multiple rows of protein sequence identifiers and accession numbers.

Prot	Id	Id	
WP_009885879.1 type I glyceraldhyde-3-phosphate dehydrogenase [M...	692	0.0	34111
7JWK_A Crystal Structure of Glyceraldhyde-3-phosphate Dehydrogen...	681	0.0	34112
WP_010814789.1 type I glyceraldhyde-3-phosphate dehydrogenase [M...	592	0.0	34113
WP_150203204.1 type I glyceraldhyde-3-phosphate dehydrogenase [M...	591	0.0	34114
WP_053344192.1 type I glyceraldhyde-3-phosphate dehydrogenase [M...	590	0.0	34115
WP_027123835.1 type I glyceraldhyde-3-phosphate dehydrogenase [M...	469	4e-163	34116
WP_027119899.1 type I glyceraldhyde-3-phosphate dehydrogenase [M...	465	1e-161	34117
WP_033159864.1 type I glyceraldhyde-3-phosphate dehydrogenase [M...	461	2e-160	34118
WP_027122004.1 type I glyceraldhyde-3-phosphate dehydrogenase [M...	461	7e-160	34119
WP_22929712.1 type I glyceraldhyde-3-phosphate dehydrogenase [M...	460	9e-160	34120
WP_011084202.1 type I glyceraldhyde-3-phosphate dehydrogenase [M...	459	2e-159	34121
WP_065164873.1 type I glyceraldhyde-3-phosphate dehydrogenase [M...	459	2e-159	34122
WP_011113552.1 type I glyceraldhyde-3-phosphate dehydrogenase [M...	459	2e-159	34123
WP_151029758.1 type I glyceraldhyde-3-phosphate dehydrogenase [M...	459	3e-159	34124
WP_014574669.1 type I glyceraldhyde-3-phosphate dehydrogenase [M...	459	3e-159	34125
WP_208864864.1 type I glyceraldhyde-3-phosphate dehydrogenase [M...	459	4e-159	34126
WP_182079053.1 type I glyceraldhyde-3-phosphate dehydrogenase [M...	458	5e-159	34127
WP_014880555.1 type I glyceraldhyde-3-phosphate dehydrogenase [M...	456	4e-158	34128
WP_036457739.1 type I glyceraldhyde-3-phosphate dehydrogenase [M...	454	3e-157	34129
WP_036440694.1 type I glyceraldhyde-3-phosphate dehydrogenase [M...	452	2e-156	34130
WP_124724278.1 type I glyceraldhyde-3-phosphate dehydrogenase [M...	449	2e-155	34131
IPL36762.1 type I glyceraldhyde-3-phosphate dehydrogenase [Metam...	446	3e-154	34132
WP_169580273.1 type I glyceraldhyde-3-phosphate dehydrogenase [M...	443	6e-153	34133
WP_179592077.1 type I glyceraldhyde-3-phosphate dehydrogenase [M...	443	6e-153	34134
WP_002881617.1 type I glyceraldhyde-3-phosphate dehydrogenase [M...	442	7e-153	34135
WP_004423184.1 type I glyceraldhyde-3-phosphate dehydrogenase [M...	442	8e-153	34136
VFU55742.1 glyceraldhyde-3-phosphate dehydrogenase [Metamycoplas...	442	1e-152	34137
PYF43743.1 glyceraldhyde-3-phosphate dehydrogenase [Metamycoplas...	442	1e-152	34138
WP_036452526.1 type I glyceraldhyde-3-phosphate dehydrogenase [M...	441	2e-152	34139
WP_203283324.1 type I glyceraldhyde-3-phosphate dehydrogenase [M...	441	2e-152	34140
RMA79020.1 glyceraldhyde-3-phosphate dehydrogenase [Metamycoplas...	441	3e-152	34141
WP_04420342.1 type I glyceraldhyde-3-phosphate dehydrogenase [M...	441	3e-152	34142
WP_113240157.1 type I glyceraldhyde-3-phosphate dehydrogenase [M...	441	3e-152	34143
WP_020002528.1 type I glyceraldhyde-3-phosphate dehydrogenase [M...	441	4e-152	34144

Figura 5. Exemple de base de dades creada amb Microsoft SQL Server Management Studio 18 a partir dels resultats de PSI-BLAST d'una seqüència de proteïna. Font: creació pròpia.

Després s'ha creat l'script per a cada proteïna en concret. A continuació se'n mostra un exemple:

```

select * from [dbo].[Proteines] where

Prot not like '%restriction endonuclease subunit S%'
and
Prot not like '%Glyceraldehyde-3-phosphate dehydrogenase%'
and
Prot not like '%type I glyceraldehyde-3-phosphate dehydrogenase%'
and
Prot not like '%unknown%'
and
Prot not like '%hypothetical protein%'
and
Prot not like '%unnamed protein%'

```

3.3.7 Creació d'una taula amb resultats de PSI-BLAST

S'ha filtrat 120 dels 518 resultats de PSI-BLAST i s'han anotat a en una taula Excel. Aquestes proteïnes filtrades corresponen a les 119 primeres de la llista del proteoma (Taula 4) juntament amb una de les dues proteïnes multifuncionals conegudes de *M. genitalium* (*DNA binding subunit from restriction system I*), que no es troba entre les primeres de la Taula 4. No s'hi ha registrat tots els resultats estrictament diferents, ja que n'hi ha que només en canvia la sintaxi i s'ha considerat innecessari. Cal destacar que dins de les 518 seqüències d'aminoàcids totals conegudes de *M. genitalium* n'hi ha unes 35 que són fragments de proteïnes. Aquestes no s'han analitzat, ja que s'ha prioritzat l'anàlisi de proteïnes senceres.

3.4 Anàlisi de resultats de PSI-BLAST

Tot i que els resultats de PSI-BLAST s'hagin filtrat, s'ha trobat que segueixen havent-hi noms diferents que fan referència a la mateixa funció. Per diferenciar les possibles funcions moonlighting s'ha analitzat manualment cada sortida diferent de PSI-BLAST

amb UniProt (accessible a <https://www.uniprot.org/>) i InterPro (accessible a <https://www.ebi.ac.uk/interpro/>).

3.5 Elecció de “verdaders positius”

Un cop obtingut el llistat de possibles funcions moonlighting, per a trobar els verdaders positius s’ha estudiat si aquestes proteïnes tenien alguna interacció registrada a la base de dades APID (accessible a <http://cicblade.dep.usal.es:8080/APID/init.action>) que coincideixi amb el procés biològic o funció de la seva possible funció moonlighting. Ara bé, també s’ha considerat com a verdaders positius les proteïnes que segons UniProt són multifuncionals i les que als resultats de PSI-BLAST han mostrat homologia amb proteïnes amb les paraules “bifunctional” o “multifunctional” al seu nom. Aquestes últimes han hagut de presentar la mateixa funció canònica i una segona funció diferent. La funció diferent és la que s’ha considerat com a moonlighting.

Finalment s’ha buscat les proteïnes multifuncionals de *M. genitalium* ja existents a la base de dades MultitaskProtDB accessible a <http://wallace.uab.es/multitask/> [1] i s’ha comprovat que aquestes també presentin interaccions amb la seva funció moonlighting.

4 Resultats

Un cop s'ha descarregat la llista de proteïnes disponibles a UniProt de *Mycoplasma genitalium*, s'ha estudiat quines d'aquestes són els 20 factors de virulència coneguts. A la Taula 4 dels Annexos s'hi pot trobar la taula resultant. Hi ha un total de 483 proteïnes i moltes encara estan sense caracteritzar.

Tot seguit s'han obtingut les seqüències de totes les proteïnes conegudes de *M. genitalium*. Es troben adjuntes al següent [enllaç](#). Hi ha un total de 518 seqüències, 35 dels quals corresponen a fragments de proteïnes, per aquesta raó no es troben anotades a la Taula 4.

Un cop s'han passat les seqüències pel programa PSI-BLAST, s'ha realitzat el filtratge mitjançant consultes amb SQL per tal d'eliminar les funcions repetides. A la Taula 1 s'hi troben els resultats:

Taula 1. Resultats PSI-BLAST. *S'ha omès algun resultat de PSI-BLAST que representava el mateix que d'altres, però en sintaxi diferent.

Entrada (ID)	Nom d'entrada a UniProt	Nom proteïna/funció	Nom/funció PSI-BLAST			
Q9ZB73	PBDGT_MYCGE	Processive diacylglycerol beta-glycosyltransferase	glycosyltransferase family 2 protein			
			diacylglycerol beta-glycosyltransferase			
			PGL/p-HBAD biosynthesis glycosyltransferase			
			UDP-glucose 4-epimerase GalE			
			Beta-(1,4)-galactosyltransferase			
			CDP-glycerol glycerophosphotransferase family prot			
			glycosyltransferase/CDP-glycerol:poly			
			Chondroitin polymerase/sintasa			
			bifunctional glycosyltransferase family 2 protein/ CDP-glycerol:glycerophosphate glycerophosphotransferase			
			putative glycosyltransferase WchA			
P47543	G3P_MYCGE	Glyceraldehyde-3-phosphate dehydrogenase	exopolysaccharide biosynthesis protein sugar transfera			
			glycosyl transferase GT2 family			
			glycosyltransferase involved in cell wall biosynthesis			
			type I glyceraldehyde-3-phosphate dehydrogenase			
			P47304	KPRS_MYCGE	Ribose-phosphate pyrophosphokinase	Ribose-phosphate pyrophosphokinase
						ribose-phosphate diphosphokinase
						phosphoribosylpyrophosphate synthetase
			P47632	CH60_MYCGE	Chaperonin GroEL	Chaperonin GroEL
						heat shock protein 60
			P47696	HGPRT_MYCGE	Hypoxanthine-guanine phosphoribosyltransferase	Hypoxanthine-guanine phosphoribosyltransferase
hypoxanthine phosphoribosyltransferase						
bifunctional protein TilS/HprT						
P47457	PFKA_MYCGE	Probable ATP-dependent 6-phosphofructokinase	Probable ATP-dependent 6-phosphofructokinase			
			diphosphate--fructose-6-phosphate 1-phosphotransferase			
			PfkA			
P47634	GLYA_MYCGE	Serine hydroxymethyltransferase	Serine hydroxymethyltransferase			
			glycine hydroxymethyltransferase			
			aminotransferase class I/ II-fold pyridoxal phosphate-dependent enzyme			

P47417	KAD_MYCGE	Adenylate kinase	Adenylate kinase nucleoside monophosphate kinase
P47695	FTSH_MYCGE	ATP-dependent zinc metalloprotease FtsH	ATP-dependent zinc metalloprotease FtsH cell division protein FtsH AAA family ATPase ATP-dependent metalloprotease FtsH/Yme1/Tma family cell-division protein and general stress protein
P47368	TOP1_MYCGE	DNA topoisomerase 1	Type 1 DNA topoisomerase
P47331	HPRK_MYCGE	HPr kinase/phosphorylase	HPr(Ser) kinase/phosphatase serine kinase
P47669	GPMI_MYCGE	2,3-bisphosphoglycerate-independent phosphoglycerate mutase	2,3-bisphosphoglycerate-independent phosphoglycerate mutase phosphoglycerate mutase
P55825	FPG_MYCGE	Formamidopyrimidine-DNA glycosylase	DNA-formamidopyrimidine glycosylase bifunctional DNA-formamidopyrimidine glycosylase/ DNA-(apurinic or apyrimidinic site) lyase
P47259	FOLD_MYCGE	Bifunctional protein Fold	bifunctional 5,10-methylenetetrahydrofolate dehydrogenase/ 5,10-methenyltetrahydrofolate cyclohydrolase Methylenetetrahydrofolate dehydrogenase (NADP+) / Methenyltetrahydrofolate cyclohydrolase
P47482	NNAAH_MYCGE	Probable nicotinate-nucleotide adenylyltransferase/ Ap4A hydrolase	nicotinate-nucleotide adenylyltransferase nicotinate (nicotinamide) nucleotide adenylyltransferase nicotinic acid mononucleotide adenylyltransferase bis(5'-nucleosyl)-tetraphosphatase (symmetrical) bidomainal protein ribosome silencing factor [Bacterium]
P47458	KPYK_MYCGE	Pyruvate kinase	pyruvate kinase
P47612	THII_MYCGE	Probable tRNA sulfurtransferase	tRNA 4-thiouridine(8) synthase Thil thiamine biosynthesis protein Thil thiamine biosynthesis/tRNA modification protein Thil thiamine biosynthesis protein

P47542	PGK_MYCGE	Phosphoglycerate kinase	Phosphoglycerate kinase
P47587	SYI_MYCGE	Isoleucine--tRNA ligase	Isoleucine--tRNA ligase isoleucyl-tRNA synthetase
P47599	ACKA_MYCGE	Acetate kinase	Acetate kinase acetate/propionate family kinase
P47293	METK_MYCGE	S-adenosylmethionine synthase	methionine adenosyltransferase Inter-dimeric interface controls function and stability of S-methionine MetK2
P47658	DPO3X_MYCGE	DNA polymerase III subunit gamma/tau	DNA polymerase III subunit gamma/tau DNA polymerase III subunits gamma and tau DNA polymerase III subunit tau dNA-directed DNA polymerase III gamma/tau subunits Holliday junction ATP-dependent DNA helicase RuvB
Q49408	NAOX_MYCGE	Probable NADH oxidase	FAD-dependent oxidoreductase NADH oxidase NADPH-dependent 2,4-dienoyl-CoA reductase, sulfur reductase NAD(FAD)-dependent dehydrogenase NADPH-dependent L-lysine N(6)-monooxygenase
P47351	DACB_MYCGE	Diadenylate cyclase	Diadenylate cyclase DisA domain-containing protein DNA integrity scanning protein DisA nucleotide-bin diadenylate cyclase CdaM DisA bacterial checkpoint controller nucleotide-bindin diadenylate cyclase Cda TIGR00159 family protein conserved hypothetical containing domain DUF147 YbbP sporulation-specific diadenylate cyclase CdaS cyclic di-AMP synthase CdaA membrane protein ybbP
P47445	PARE_MYCGE	DNA topoisomerase 4 subunit B	DNA topoisomerase IV subunit B topoisomerase II subunit B

			DNA gyrase subunit B
			DNA gyrase (subunit B) GyrB
			type IIA DNA topoisomerase subunit B
P47647	ENO_MYCGE	Enolase	phosphopyruvate hydratase
			Chain A, Enolase
			2-phospho-D-glycerate hydro-lyase
			Enolase 2
P47429	PEPF_MYCGE	Oligoendopeptidase F homolog	Oligoendopeptidase F homolog
			oligoendopeptidase F
			M3 family metallopeptidase
			Oligoendopeptidase F%2C
			oligopeptidase F. Metallo peptidase. MEROPS family M03B
			peptidase
Q49427	PLSX_MYCGE	Phosphate acyltransferase	phosphate acyltransferase PlsX
			glycerol-3-phosphate acyltransferase
			fatty acid-phospholipid synthesis protein
			Fatty acid/phospholipid synthesis protein
P47456	SCPB_MYCGE	Segregation and condensation protein B	SMC-Scp complex subunit ScpB
			segregation and condensation protein B
			condensin subunit ScpB
P47576	SYV_MYCGE	Valine--tRNA ligase	Valine--tRNA ligase
			valyl-tRNA synthetase
P47276	UPP_MYCGE	Uracil phosphoribosyltransferase	Uracil phosphoribosyltransferase
P47448	UVRC_MYCGE	UvrABC system protein C	excinuclease ABC subunit UvrC
			excinuclease ABC subunit C
			Heat-inducible transcription repressor hrcA
P47486	UVRD_MYCGE	Probable DNA helicase II homolog	ATP-dependent helicase
			UvrD-helicase domain-containing protein
			DNA helicase PcrA
			aTP-dependent DNA helicase PcrA
			AAA family ATPase
			DNA helicase-2/ATP-dependent DNA helicase PcrA
P47319	UVRB_MYCGE	UvrABC system protein B	excinuclease ABC subunit B

			excinuclease ABC subunit UvrB
P47622	URK_MYCGE	Uridine kinase	Uridine kinase
P47660	UVRA_MYCGE	UvrABC system protein A	UvrABC system protein A
P47515	ODPB_MYCGE	Pyruvate dehydrogenase E1 component subunit beta	alpha-ketoacid dehydrogenase subunit beta
			2-oxoisovalerate dehydrogenase
			Pyruvate dehydrogenase (lipoamide), beta chain
			pdhB
Q9ZB82	RL332_MYCGE	50S ribosomal protein L33 2	50S ribosomal protein L33
			large subunit ribosomal protein L33
P47703	RNPA_MYCGE	Ribonuclease P protein component	Ribonuclease P protein component
			protein C5 component of RNase P
			PA5569
P47615	SYT_MYCGE	Threonine--tRNA ligase	Threonine--tRNA ligase
			threonyl-tRNA synthetase
P47372	SYW_MYCGE	Tryptophan--tRNA ligase	Tryptophan--tRNA ligase
			Tryptophanyl-tRNA synthetase
			tryptophanyl-tRNA synthetase TrpS
P47525	SYP_MYCGE	Proline--tRNA ligase	Proline--tRNA ligase
			prolyl-tRNA synthetase
P47251	SYS_MYCGE	Serine--tRNA ligase	Serine--tRNA ligase
			seryl-tRNA synthetase
P47267	SYM_MYCGE	Methionine--tRNA ligase	Methionine--tRNA ligase
			methionyl-tRNA synthetase
			class I tRNA ligase family protein
			metG
P47500	RF1_MYCGE	Peptide chain release factor 1	Peptide chain release factor 1
			peptide chain release factor RF-1
P47318	SECA_MYCGE	Protein translocase subunit SecA	preprotein translocase subunit SecA
			Protein translocase subunit SecA
			protein export cytoplasm protein SecA
			preprotein translocase seca subunit
P47480	TIG_MYCGE	Trigger factor	Trigger factor
P47315	PTG3C_MYCGE	PTS system glucose-specific EIICBA component	PTS system glucose-specific EIICBA component

			PTS transporter subunit IIABC
			glucose PTS transporter subunit IIA
			PTS glucose transporter subunit IIB
			PTS transporter subunit EIIC
			glucose-specific PTS transporter subunit IIBC
			EIICBA-Glc
			PTS system D-glucosamine-specific IIC component
			Glucose-specific phosphotransferase enzyme IIA compone
Q49425	RUVB_MYCGE	Holliday junction ATP-dependent DNA helicase RuvB	Holliday junction branch migration DNA helicase RuvB AAA family ATPase
P47356	RSGA_MYCGE	Small ribosomal subunit biogenesis GTPase RsgA	Small ribosomal subunit biogenesis GTPase RsgA ribosome small subunit-dependent GTPase A conserved hypothetical ATP/GTP-binding protein ribosome biogenesis GTPase putative ATP/GTP-binding protein
P47464	RSMH_MYCGE	Ribosomal RNA small subunit methyltransferase H	Ribosomal RNA small subunit methyltransferase H 16S rRNA (cytosine(1402)-N(4))-methyltransferase RsmH S-adenosyl-methyltransferase MraW
P47701	RSMA_MYCGE	Ribosomal RNA small subunit methyltransferase A	Ribosomal RNA small subunit methyltransferase A 16S rRNA (adenine(1518)-N(6)/ adenine(1519)-N(6))-dimethyltransferase dimethyladenosine transferase
P47302	RSMI_MYCGE	Ribosomal RNA small subunit methyltransferase I	16S rRNA (cytidine(1402)-2'-O)-methyltransferase Tetrapyrrole (Corrin-Porphyrin) methylase family protein tetrapyrrole (corrin/porphyrin) methyltransferase
P47620	RSMG_MYCGE	Ribosomal RNA small subunit methyltransferase G	16S rRNA (guanine(527)-N(7))-methyltransferase Rsm glucose inhibited division protein B methyltransferase GidB
Q49424	RUVA_MYCGE	Holliday junction ATP-dependent DNA helicase RuvA	Holliday junction ATP-dependent DNA helicase RuvA Holliday junction branch migration protein RuvA helix-hairpin-helix domain-containing protein
P47672	PYRH_MYCGE	Uridylate kinase	Uridylate kinase UMP kinase

P47355	PKNS_MYCGE	Putative serine/threonine-protein kinase	serine/threonine protein kinase Stk1 family PASTA domain-containing Ser/Thr kinas serine/threonine-protein kinase PrkC
P47520	SPOT_MYCGE	Probable guanosine-3'5'-bis(diphosphate) 3'-pyrophosphohydrolase	RelA/SpoT family protein Probable guanosine-3'5'-bis(diphosphate) 3'-pyrophosphohydrolase bifunctional (p)ppGpp synthetase/ guanosine-3',5'-bis(diphosphate) ppGpp synthetase GTP diphosphokinase
P47339	RL9_MYCGE	50S ribosomal protein L9	50S ribosomal protein L9 rplI
P47376	RNY_MYCGE	Ribonuclease Y	Ribonuclease Y HDIG domain-containing protein Rnase Y domain-containing protein rny 2,3 cyclic-nucleotide 2-phosphodiesterase phosphodiesterase RNA binding metal dependent phosphohydrolase YmdA/YtgF family protein
P47396	RS10_MYCGE	30S ribosomal protein S10	30S ribosomal protein S10 rpsJ BS13
P47455	SCPA_MYCGE	Segregation and condensation protein A	Segregation and condensation protein A segregation/condensation protein A ScpA/B protein condensin subunit ScpA chromosome segregation protein ScpA chromosome condensation and partitioning factor
P47670	TPIS_MYCGE	Triosephosphate isomerase	triose-phosphate isomerase bifunctional PGK/TIM
P47330	TILS_MYCGE	tRNA(Ile)-lysidine synthase	tRNA(Ile)-lysidine synthase

			tRNA lysidine(34) synthetase TlIS
			PP-loop superfamily ATPase/cell cycle protein
			cell cycle protein
			TPA: hypoxanthine phosphoribosyltransferase
P47663	RS15_MYCGE	30S ribosomal protein S15	30S ribosomal protein S15
			ribosomal protein S15
			small subunit ribosomal protein S15
			SSU ribosomal protein S15p (S13e)
P58061	SECG_MYCGE	Probable protein-export membrane protein SecG	Probable protein-export membrane protein SecG
			preprotein translocase subunit SecG
P47412	RL6_MYCGE	50S ribosomal protein L6	50S ribosomal protein L6
			Structure of the large subunit of the chloro-ribosome
			Ribosomal protein L6 alpha-beta domain superfamily
			Ribosomal protein L6 family
			50S ribosomal protein chloroplastic-like
			60S ribosomal protein L6
			50S ribosomal L6, chloroplastic
P47421	RS13_MYCGE	30S ribosomal protein S13	30S ribosomal protein S13
P47332	LGT_MYCGE	Phosphatidylglycerol--prolipoprotein diacylglyceryl transferase	Phosphatidylglycerol--prolipoprotein diacylglyceryl transferase
P47418	MAP1_MYCGE	Methionine aminopeptidase	Methionine aminopeptidase
			type I methionyl aminopeptidase
P47299	MANB_MYCGE	Phosphomannomutase	Phosphomannomutase
			phospho-sugar mutase
			phosphoglucomutase
			ribonuclease HII
P47269	ALF_MYCGE	Fructose-bisphosphate aldolase	Fructose-bisphosphate aldolase
			class II fructose-1,6-bisphosphate aldolase
Q49396	K1PF_MYCGE	Putative 1-phosphofructokinase	1-phosphofructokinase
			PfkB family carbohydrate kinase
P47280	KITH_MYCGE	Thymidine kinase	Thymidine kinase
P47364	GALE_MYCGE	UDP-glucose 4-epimerase	UDP-glucose 4-epimerase
			UDP-glucose 4-epimerase GalE
			GDP-mannose 4,6-dehydratase

			Homology to UDP-galactose-4-epimerase X06226
Q49405	DPO3A_MYCGE	DNA polymerase III subunit alpha	DNA polymerase III subunit alpha
			DNA-directed DNA polymerase
P47666	HPRR_MYCGE	Hydroperoxide reductase	Hydroperoxide reductase
			OsmC family protein
			peroxiredoxin
P47335	EFG_MYCGE	Elongation factor G	Elongation factor G
			fus
			translation elongation factors (GTPase)-like protein
P47539	FTSY_MYCGE	Signal recognition particle receptor FtsY	Signal recognition particle receptor FtsY
			signal recognition particle-docking protein FtsY
			signal recognition particle GTPase FtsY
			Cell division protein FtsY
P47524	GREA_MYCGE	Transcription elongation factor GreA	Transcription elongation factor GreA
			GreA/GreB family elongation factor
Q49413	HMW1_MYCGE	Cytadherence high molecular weight protein 1	terminal organelle protein HMW1
			EAGR box-containing protein
			terminal organelle assembly protein TopJ
			Fibrillin-1
			molecular chaperone DnaJ
			terminal organelle protein P200
P47566	AMPP_MYCGE	Putative Xaa-Pro aminopeptidase	Putative Xaa-Pro aminopeptidase
			aminopeptidase P family protein
			Xaa-Pro dipeptidase
			family metallopeptidase
			metallopeptidase family M24 aminopeptidase
			Aminopeptidase YpdF
			M24B family metallopeptidase
P47691	GALU_MYCGE	UTP--glucose-1-phosphate uridylyltransferase	UTP--glucose-1-phosphate uridylyltransferase
			UDP-glucose pyrophosphorylase
P47641	ATPA_MYCGE	ATP synthase subunit alpha	ATP synthase subunit alpha
			adenosinetriphosphatase
			ATP synthase F1 sector alpha subunit

			F-type H ⁺ -transporting ATPase subunit alpha
P47384	LEPA_MYCGE	Elongation factor 4	Elongation factor 4
			ribosomal elongation factor, GTPase
			GTP-binding protein LepA
			translation elongation factor LepA
P47252	KTHY_MYCGE	Thymidylate kinase	Thymidylate kinase
			dTMP kinase
P47481	LON_MYCGE	Lon protease	Lon protease
			endopeptidase La
			ATP-dependent protease La
			mf-Lon with ssrA degradation tag LVA
			class III heat-shock ATP-dependent LonA protease
P47254	MNME_MYCGE	tRNA modification GTPase MnmE	tRNA modification GTPase MnmE
			tRNA uridine-5-carboxymethylaminomethyl(34)
			tRNA modification GTPase TrmE
			tRNA modification GTPase
			tRNA uridine(34) 5-carboxymethylaminomethyl synthesis GTPase MnmE
P47512	LPLA_MYCGE	Probable lipoate-protein ligase A	Probable lipoate-protein ligase A
			lipoate-protein ligase A
P47623	NADE_MYCGE	NH(3)-dependent NAD(+) synthetase	NH(3)-dependent NAD(+) synthetase
			NAD(+) synthase
			NH3-dependent NAD+ synthetase
P47447	HRCa_MYCGE	Heat-inducible transcription repressor HrcA	Heat-inducible transcription repressor HrcA
			HrcA family transcriptional regulator
P47324	OPPC_MYCGE	Oligopeptide transport system permease protein OppC	Oligopeptide transport system permease protein OppC
			ABC transporter permease
			oligopeptide transport system permease protein
			peptide/nickel transport system permease protein
			Glutathione transport system permease protein gsiD
P47624	OBG_MYCGE	GTPase Obg	GTPase Obg
			GTPase ObgE

			GTP-binding protein Obg/CgtA family protein
			GTPase CgtA
			Obg family GTPase CgtA
			small GTP-binding protein
P47645	ATP6_MYCGE	ATP synthase subunit a	ATP synthase subunit a
			ATP synthase FO sector subunit a
			ATP synthase FO subunit A
			adenosinetriphosphatase
			ATP synthase FO subcomplex A subunit
			F-type H ⁺ -transporting ATPase subunit a
Q49421	ESL3_MYCGE	Putative esterase/lipase 3	Putative esterase/lipase 3
			alpha/beta hydrolase
			lipase/esterase family protein
			alpha/beta fold hydrolase
			succinyl-6-hydroxy-2, 4-cyclohexadiene-1-carboxylate synthase
			triacylglycerol lipase
			Sigma factor SigB regulation protein RsbQ
			Pseudogene of Triacylglycerol lipase(N terminal part)...
P47571	DER_MYCGE	GTPase Der	ribosome biogenesis GTPase Der
			GTP-binding protein Der
			ribosome-associated GTPase EngA
P47513	DLDH_MYCGE	Dihydrolipoyl dehydrogenase	Dihydrolipoyl dehydrogenase
			Dihydrolipoamide dehydrogenase (E3)
			component of pyruvate dehydrogenase complex
			dihydrolipoamide dehydrogenase
			pyruvate dehydrogenase E2 component
Q49433	CDSA_MYCGE	Putative phosphatidate cytidyltransferase	phosphatidate cytidyltransferase
			CDSA protein
			EOG090X04NQ
			CDP-diacylglycerol synthase
			Cytidyltransferase family
			CDP-archaeol synthase
			GSCOCG00003264001-RA-CDS

			FYVE domain-containing protein 3 Autophagy-linked phosphatidate photoreceptor-specific
			DUF92 domain-containing protein
			E2.7.7.41
Q49406	EX53_MYCGE	5'-3' exonuclease	5'-3' exonuclease
			DNA polymerase I
			polymerase protein
			dNA polymerase
			flap endonuclease
P47374	NADK_MYCGE	NAD kinase	NAD kinase
			NAD(+) kinase
			NAD(+)/NADH kinase
			inorganic polyphosphate kinase
			putative sugar kinase
P47492	DNAG_MYCGE	DNA primase	DNA primase
			homology to DnaE primase M10040
P47593	IPYR_MYCGE	Inorganic pyrophosphatase	Inorganic pyrophosphatase
			inorganic diphosphatase
			ppa
P47285	GLPO_MYCGE	Glycerol 3-phosphate oxidase	Glycerol 3-phosphate oxidase
			FAD-dependent oxidoreductase
			NAD(P)/FAD-dependent oxidoreductase
			FAD-dependent oxidoreductase
			type 2 glycerol-3-phosphate oxidase
			FAD/NAD(P)-binding oxidoreductase
			glycerol-3-phosphate dehydrogenase
			FAD dependent oxidoreductase
			Oxidized L-alpha-Glycerophosphate Oxidase
			L-2-hydroxyglutarate oxidase LhgO
			GAB DTP gene cluster repressor
			glycerophosphate oxidase
			glycerol oxydase
			glycerol-3-phospate oxidase

			putative dehydrogenase
P47353	KGUA_MYCGE	Guanylate kinase	Guanylate kinase gmk
P13927	EFTU_MYCGE	Elongation factor Tu	Elongation factor Tu Elongation factor translation elongation factor Crystal structure of an ancient sequence-reconstructed Elo
P47642	ATPD_MYCGE	ATP synthase subunit delta	ATP synthase subunit delta adenosinetriphosphatase ATP synthase CF1 subunit delta ATP synthase, delta (OSCP) subunit ATP synthase FOF1 subunit delta F-type H ⁺ -transporting ATPase subunit delta F-type ATPase subunit delta
P47265	DNAJ_MYCGE	Chaperone protein DnaJ	Chaperone protein DnaJ molecular chaperone DnaJ J domain-containing protein Homologous to Swiss-Prot Accession Number P28616, heat shock protein DNAJ, partial dnaj
P47477	END4_MYCGE	Probable endonuclease 4	endonuclease 4 deoxyribonuclease IV endonuclease IV apurinic endonuclease (APN1)
P47284	GLPK_MYCGE	Glycerol kinase	Glycerol kinase
Q49412	ESL1_MYCGE	Putative esterase/lipase 1	esterase/lipase 1 alpha/beta hydrolase triacylglycerol lipase alpha/beta fold family hydrolase 2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase Sigma factor SigB regulation protein RsbQ Pimeloyl-ACP methyl ester carboxylesterase acetoin dehydrogenase dihydrolipoyllysine-residue acet...

			proline-specific peptidase
P47346	GATB_MYCGE	Aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit B	Aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit B Aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit B aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subu Asp-tRNA(Asn)/Glu-tRNA(Gln) amidotransferase Gat sp-tRNA(Asn)/ Glu-tRNA(Gln) amidotransferase GatCAB subunit B Glu-tRNA(Gln) amidotransferase, subunit B Glu-tRNA(Gln) amidotransferase, subunit B
P47279	GLPF_MYCGE	Probable glycerol uptake facilitator protein	glycerol uptake facilitator protein aquaporin family protein aquaporin MIP family channel proteins MIP family major intrinsic protein channel protein glycerol uptake facilitator related permease
P47466	FTSZ_MYCGE	Cell division protein FtsZ	Cell division protein FtsZ cell division FtsZ family protein Ftsz
P47296	DEOC_MYCGE	Deoxyribose-phosphate aldolase	Deoxyribose-phosphate aldolase 2-deoxyribose-5-phosphate aldolase
P47352	DEF_MYCGE	Peptide deformylase	Peptide deformylase formylmethionine deformylase homolog formylmethionine deformylase A
P47644	ATPL_MYCGE	ATP synthase subunit c	ATP synthase subunit c ATP synthase F0 subunit C ATP synthase FOF1 subunit C F-type H ⁺ -transporting ATPase subunit c Lipid-binding protein Lipid-binding protein
P47463	MRAZ_MYCGE	Transcriptional regulator MraZ	Transcriptional regulator MraZ cell division/cell wall cluster transcriptional repressor MraZ Protein MraZ Homology to X52063

			Crystal Structure Of A Cell Division And Cell Wall Biosynthesis Protein Upf0040
P47648	MSRA_MYCGE	Peptide methionine sulfoxide reductase MsrA	Peptide methionine sulfoxide reductase MsrA peptide-methionine (R)-S-oxide reductase MsrB peptide-methionine (S)-S-oxide reductase MsrA methionine-S-sulfoxide reductase peptide-methionine (S)-S-oxide reductase peptide-methionine (R)-S-oxide reductase peptide methionine sulfoxide reductase methionine-R-sulfoxide reductase bifunctional peptide-methionine (S)-S-oxide reductase MsrA/ peptide-methionine (R)-S-oxide reductase MsrB bifunctional methionine sulfoxide reductase A/B
Q49434	T1SX_MYCGE	Putative type I specificity subunit S.MgeORF438P	type I restriction enzyme specificity protein restriction endonuclease subunit S Type I restriction modification DNA specificity domain type I R-M system specificity subunit Crystal structure of Type-I restriction-modification system S subunit type I R-M system S protein
Q49417	P32_MYCGE	P32 adhesin	P32 adhesin 32kDa protein revertant adhesin-related protein P3 cytadhesin protein MGC2 cytoadhesin cytadherence-related surface protein Mgc2 Cytadhesin P30/P32 liver stage-specific antigen-1 FHA domain-containing protein FecR domain-containing protein sporulation protein translation initiation factor IF-2

	kinase-like domain-containing protein
	putative chitinase 10 like protein
	zinc finger protein 683
	mediator of RNA polymerase II transcription subunit
	protein piccolo-like
	toxin glutamine deamidase domain-containing protein
	glutenin, high molecular weight subunit PW212-like
	RNA-binding protein 33 isoform X2
	zonadhesin
	tryptophan-rich antigen
	lepA 1
	leucine-rich repeat protein
	translation initiation factor IF-2
	fibronectin
	circumsporozoite protein-like
	snRNA-activating protein complex subunit 4-like
	FHA domain-containing protein
	DUF805 domain-containing protein

S'hi observen sortides molt similars per una mateixa proteïna i algunes que a simple vista semblen funcions diferents, però que tracten del mateix, ja que fa referència a l'estructura química, família, gen que codifica per aquesta proteïna, entre d'altres.

Un cop analitzades totes les funcions predites per PSI-BLAST s'han establert les possibles funcions moonlighting i s'ha estudiat si hi havia les corresponents interaccions entre la funció canònica i la possible funció moonlighting amb APID (Taula 2).

Taula 2. Possibles funcions moonlighting

Entrada (ID)	Funció canònica	Possible funció moonlighting	Parelles PPI (només es mostren algunes sortides)
Q9ZB73	Diacylglycerol beta-glycosyltransferase	CDP-glycerol glycerophosphotransferase activity	-
P47696	Hypoxanthine-guanine phosphoribosyltransferase	tRNA(Ile)-lysidine synthase	-
P55825	DNA-formamidopyrimidine glycosylase	DNA-(apurinic or apyrimidinic site) lyase activity	-
P47259	Methylenetetrahydrofolate dehydrogenase (NADP+) activity	methenyltetrahydrofolate cyclohydrolase activity	-
P47482	Nicotinate-nucleotide adenyltransferase	bis(5'-nucleosyl)-tetrphosphatase (symmetrical) = Ap4A hydrolase	-
		ribosome silencing factor	-
P47658	DNA polymerase III subunit gamma/tau	Holliday junction ATP-dependent DNA helicase RuvB	GO:ATP-dependent chromatin remodeling (ATP-dependent DNA helicase Q1)
P47448	Excinuclease ABC activity subunit UvrC	Heat-inducible transcription repressor hrcA	-
P47520	Guanosine-3',5'-bis(diphosphate) 3'-pyrophosphohydrolase	ppGpp synthetase	-
P47670	Triosephosphate isomerase	Phosphoglycerate kinase	-
P47330	tRNA(Ile)-lysidine synthase	TPA: hypoxanthine phosphoribosyltransferase	-
P47299	Phosphomannomutase	phosphoglucomutase	-
		ribonuclease HII	-
P47364	UDP-glucose 4-epimerase	GDP-mannose 4,6-dehydratase	-
Q49413	Cytoskeleton-like protein	molecular chaperone DnaJ	-

Q49421	Putative esterase/lipase 3	succinyl-6-hydroxy-2, 4-cyclohexadiene-1-carboxylate synthase	-
Q49433	Putative phosphatidate cytidyltransferase	FYVE domain-containing protein 3 Autophagy-linked	-
P47285	Glycerol 3-phosphate oxidase	L-2-hydroxyglutarate oxidase LhgO GAB DTP gene cluster repressor	- -
Q49412	Putative esterase/lipase 1	2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase Acetoin dehydrogenase dihydrolipoyllysine-residue acetyltransferase subunit proline-specific peptidase	- - - -
P47648	Peptide-methionine (S)-S-oxide reductase (MsrA)	peptide-methionine (R)-S-oxide reductase (MsrB)	-

Hi ha hagut el cas d'una proteïna que, segons UniProt, és bifuncional, però que el programa PSI-BLAST també hi ha trobat homologia amb una altra proteïna que, fent recerca, s'ha vist que també és multifuncional, però que només comparteixen una funció [16]. És el cas de la proteïna *Probable nicotinate-nucleotide adenyltransferase/Ap4A hydrolase*. En aquesta proteïna se li prediu la bifuncionalitat que indica al seu nom, i la funció de *ribosome silencing factor [Bacterium]*, que també es considera multifuncional amb la funció *nicotinate-nucleotide adenyltransferase*. Tot i així, s'ha seguit el mètode establert per predir proteïnes multifuncionals i s'ha considerat que no ho és.

Un altre cas particular ha sigut el de la *P32 adhesin*, que al tenir tantes sortides diferents en el PSI-BLAST, s'ha procedit directament a mirar si es coneixia alguna interacció per aquesta proteïna a APID abans d'estudiar cada una de les sortides. No se n'ha trobat cap, així que s'ha descartat com a possible proteïna moonlighting.

Respecte a les PPI només se n'ha trobat a una de les candidates: DNA polymerase III subunit gamma/tau, que interacciona amb una helicasa dependent d'ATP. Ara bé, després de buscar les proteïnes multifuncionals conegudes d'aquest organisme a la base

de dades MultitaskProtDB s'ha vist que la *DNA binding subunit from restriction system I* no ha sigut una proteïna candidata a ser multifuncional segons els resultats de PSI-BLAST, però sí segons MultitaskProteinDB. A més, a l'altre proteïna multifuncional coneguda (*Glyceraldehyde-3-phosphate dehydrogenase*) no se li ha trobat interacció amb la seva funció moonlighting.

Finalment, s'han definit els "verdaders positius", és a dir, les proteïnes multifuncionals de *M. genitalium* resultants de tot aquest estudi. S'ha trobat un total de 11 proteïnes moonlighting. A Taula 3 hi ha el llistat de proteïnes multifuncionals actualment catalogades.

Taula 3. "Verdaders positius", proteïnes moonlighting. *Segons Multitasking Protein Database

Entrada (ID)	Funció canònica	Funció moonlighting
P47543	Glyceraldehyde-3-phosphate dehydrogenase	Mucin binding*
Q49434	DNA binding subunit from restriction system I	Cytoskeleton protein*
Q9ZB73	diacylglycerol beta-glycosyltransferase	CDP-glycerol glycerophosphotransferase activity
P47696	Hypoxanthine-guanine phosphoribosyltransferase	tRNA(Ile)-lysidine synthase
P55825	DNA-formamidopyrimidine glycosylase	DNA-(apurinic or apyrimidinic site) lyase activity
P47259	methylenetetrahydrofolate dehydrogenase (NADP+) activity	methenyltetrahydrofolate cyclohydrolase activity
P47482	nicotinate-nucleotide adenylyltransferase	bis(5'-nucleosyl)-tetrphosphatase (symmetrical) = Ap4A hydrolase
P47658	DNA polymerase III activity subunit gamma/tau	ATP-dependent DNA helicase activity
P47520	guanosine-3',5'-bis(diphosphate) 3'-pyrophosphohydrolase	ppGpp synthetase
P47670	Triosephosphate isomerase	Phosphoglycerate kinase
P47648	peptide-methionine (S)-S-oxide reductase (MsrA)	peptide-methionine (R)-S-oxide reductase (MsrB)

Cal destacar que la proteïna *DNA polymerase III activity subunit gamma/tau* és la única que no estava establerta com a proteïna bifuncional o multifuncional prèviament per cap de les bases de dades MultitaskProtDB o UniProt. Ara bé, sí que s'ha trobat interacció a APID.

5 Discussió

Les proteïnes que segons UniProt són bifuncionals curiosament no tenen interacció amb la segona funció a APID. Ara bé, en aquesta base de dades d'interactòmica no hi ha registrada la interacció de totes les proteïnes que es coneixen de tots els individus, així que una de les raons podria ser que la interacció amb la segona funció encara no es coneix i no pas que no existeixi. Una altre raó podria ser que les proteïnes no s'haguessin cercat correctament en aquesta base de dades, ja que hi ha molts noms similars per una sola proteïna. Aquesta hipòtesi també es podria aplicar a la proteïna *Glyceraldehyde-3-phosphate dehydrogenase*, que també estava establerta com a multifuncional prèviament per MultitaskProtDB i no s'ha detectat cap interacció a APID.

També cal destacar que, com s'ha vist a la Taula 4, hi ha moltes proteïnes de *M. genitalium* que encara estan per caracteritzar, així que encara queda molt per a descobrir.

6 Conclusions

6.1 Conclusions

Es conclou que la manera més eficaç de trobar proteïnes multifuncionals de *Mycoplasma genitalium* actualment és combinant els resultats de PSI-BLAST amb la base de dades PPI APID i amb les prediccions de multifuncionalitat que fa PSI-BLAST juntament amb la informació que es pot trobar a UniProt sobre aquesta multifuncionalitat. Mitjançant aquest criteri s'ha trobat que un 9.17% de les proteïnes analitzades són multifuncionals. Ara bé, s'ha de tenir en compte que hi ha tantes variables que pot ser molt complicat trobar una eina única i útil per a tots els estudis encarats a trobar funcions moonlighting.

No s'han aconseguit tots els objectius inicials a causa dels imprevistos (no accessibilitat a ByPass, aprenentatge en la utilització d'un nou programa, etc.) i, a falta d'eines bioinformàtiques per automatitzar tots els processos, s'ha requerit més temps del que s'esperava.

6.2 Línies de futur

En un futur es podria acabar d'analitzar tot el proteoma de *M. genitalium*, ja que hi ha altes probabilitats de trobar més proteïnes multifuncionals. Després faltaria comprovar aquesta multifuncionalitat experimentalment.

A més, algunes d'aquestes proteïnes moonlighting també són factors de virulència, com per exemple la *peptide-methionine (S)-S-oxide reductase (MsrA)*, i pot facilitar molt més l'estudi del seu funcionament i la creació de vacunes.

També seria útil crear programes que englobin tot aquest anàlisi de forma automàtica i que sigui més eficaç. Ja que hi ha proteïnes que en aquest treball se'ls ha trobat possibles segones funcions, però s'han establert com a falsos positius d'acord amb el criteri establert quan potser són veritables positius. Per exemple: *Phosphatidylglycerol--prolipoprotein diacylglycerol transferase*, segons UniProt, és multifuncional en altres organismes com *Rathayibacter tanacetii*, tenint com a segona funció *Tryptophan synthase alpha chain*. En els resultats obtinguts de PSI-BLAST no s'ha intuït cap segona funció. Ara bé, seria interessant un estudi amb més exhaustiu d'aquest tipus de proteïnes, ja que aquesta en concret està implicada en la virulència de molts patògens [23].

6.3 Seguiment de la planificació

No s'ha seguit la planificació i metodologia ja que no es tenia previst:

- Realitzar una cerca de programes alternatius per la recerca de funcions de proteïnes i d'interaccions, però no se n'ha trobat de més adequats, ja que UniProt i APID són recursos molt complets .
- La instal·lació del programa BLAST+, recerca d'informació de com funciona aquest programa de comandes, de quines bases de dades s'han de descarregar per a realitzar la cerca d'homologia remota corresponent i de com descarregar-les. A més a més de l'aprenentatge de la configuració del programa.
- La utilització del programa IDLE (Python 3.10) per la creació de scripts en Python per a:

- Crear un document amb extensió en “.fa” (format FASTA) per a cada seqüència a partir d’un sol document on s’hi trobaven totes les seqüències juntes.
 - Automatitzar la creació de documents executables amb la comanda necessària incorporada per analitzar cada seqüència amb PSI-BLAST.
-
- La utilització de Microsoft SQL Server Management Studio 18 per a crear un script en SQL i poder fer consultes per facilitar la creació d’un llistat de funcions predites per PSI-BLAST.

 - La utilització de la pàgina web d’InterPro per obtenir informació per a poder conèixer millor la classificació de les proteïnes analitzades i poder distingir amb més facilitat les diferents funcions.

7 Glossari

Funció moonlightning: funció d'una proteïna que és diferent de la canònica.

Multifuncional: que té més d'una funció.

PDB: Protein Data Bank

PPI: interacció proteïna-proteïna

Proteïna moonlightning: proteïna que té més d'una funció

8 Bibliografia

- [1] Franco-Serrano, L., Hernández, S., Calvo, A., Severi, M. A., Ferragut, G., Pérez-Pons, J. A., Pinõl, J., Pich, Ò., Mozo-Villarias, Á., Amela, I., Querol, E., & Cedano, J. (2018). MultitaskProtDB-II: An update of a database of multitasking/moonlighting proteins. *Nucleic Acids Research*, 46(D1), D645–D648. <https://doi.org/10.1093/nar/gkx1066>
- [2] Párraga-Niño, N., Colomé-Calls, N., Canals, F., Querol, E., & Ferrer-Navarro, M. (2012). A comprehensive proteome of *Mycoplasma genitalium*. In *Journal of Proteome Research* (Vol. 11, Issue 6, pp. 3305–3316). <https://doi.org/10.1021/pr300084c>
- [3] Hernández, S., Calvo, A., Ferragut, G., Franco, L., Hermoso, A., Amela, I., Gómez, A., Querol, E., & Cedano, J. (2014). Can bioinformatics help in the identification of moonlighting proteins? *Biochemical Society Transactions*, 42(6), 1692–1697. <https://doi.org/https://doi.org/10.1042/BST20140241>
- [4] Yang, Z., Hou, J., Mu, M., & Wu, S. Y. (2020). Subtractive proteomics and systems biology analysis revealed novel drug targets in *Mycoplasma genitalium* strain G37. *Microbial Pathogenesis*, 145, 104231.
- [5] MC le Roux, & AA Hoosen. (2010). *Mycoplasma genitalium*: a brief review. In *South Afr J Epidemiol Infect* (Vol. 25, Issue 4).
- [6] Henderson, B., & Martin, A. (2011). Bacterial virulence in the moonlight: Multitasking bacterial moonlighting proteins are virulence determinants in infectious disease. In *Infection and Immunity* (Vol. 79, Issue 9, pp. 3476–3491). <https://doi.org/10.1128/IAI.00179-11>
- [7] Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. In *Nucleic Acids Research* (Vol. 25, Issue 17). Oxford University Press. <https://academic.oup.com/nar/article/25/17/3389/1061651>

- [8]Gomez, A., Domedel, N., Cedano, J., Pinol, J., and Querol, E. (2003). Do current sequence analysis algorithms disclose multifunctional (moonlighting) proteins? *Bioinformatics* 19, 895–896. doi:10.1093/bioinformatics/btg111
- [9]Khan, I., Chitale, M., Rayon, C., and Kihara, D. (2012). Evaluation of function predictions by PFP, ESG, and PSI-BLAST for moonlighting proteins. *BMC Proc.* 6(Suppl. 7):S5. doi:10.1186/1753-6561-6-S7-S51753-6561-6-S7-S5
- [10]Gomez, A., Cedano, J., Espadaler, J., Hermoso, A., Pinol, J., and Querol, E. (2008). Prediction of protein function improving sequence remote alignment search by a fuzzy logic algorithm. *Protein J.* 27, 130–139. doi:10.1007/s10930-007-9116-x
- [11]Hernández, S., Franco, L., Calvo, A., Ferragut, G., Hermoso, A., Amela, I., Gómez, A., Querol, E., & Cedano, J. (2015). Bioinformatics and moonlighting proteins. *Frontiers in Bioengineering and Biotechnology*, 3(JUN). <https://doi.org/10.3389/fbioe.2015.00090>
- [12]Hunter, S., Jones, P., Mitchell, A., Apweiler, R., Attwood, T. K., Bateman, A., et al. (2012). InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Res.* 40, D306–D312. doi:10.1093/nar/gkr948gkr948
- [13]Henikoff, S., Henikoff, J. G., and Pietrokovski, S. (1999). Blocks+: a non-redundant database of protein alignment blocks derived from multiple compilations. *Bioinformatics* 15, 471–479. doi:10.1093/bioinformatics/15.6.471
- [14]Alonso-López, D., Gutiérrez, M. A., Lopes, K. P., Prieto, C., Santamaría, R., & de Las Rivas, J. (2016). APID interactomes: Providing proteome-based interactomes with controlled quality for multiple species and derived networks. *Nucleic Acids Research*, 44(W1), W529–W535. <https://doi.org/10.1093/nar/gkw363>
- [15]Alonso-López, Di., Campos-Laborie, F. J., Gutiérrez, M. A., Lambourne, L., Calderwood, M. A., Vidal, M., & de Las Rivas, J. (2019). APID database: Redefining protein-protein interaction experimental evidences and binary interactomes. *Database*, 2019. <https://doi.org/10.1093/database/baz005>
- [16]The UniProt Consortium
UniProt: the universal protein knowledgebase in 2021
Nucleic Acids Res. 49:D1 (2021)

- [17]Ueno PM, Timenetsky J, Centonze VE, et al. Interaction of *Mycoplasma genitalium* with host cells: evidence for nuclear localization. *Microbiol* 2008; 154: 3033-3041
- [18]Camacho C., Coulouris G., Avagyan V., Ma N., Papadopoulos J., Bealer K., & Madden T.L. (2008) "BLAST+: architecture and applications." *BMC Bioinformatics* 10:421. PubMed
- [19]National Center for Biotechnology Information. (2020). *Index of /blast/documents*. <https://ftp.ncbi.nlm.nih.gov/blast/documents/blastdb.html>
- [20]Python. (2022). *IDLE Shell* (3.10.4) [programari]
- [21]BLAST® Command Line Applications User Manual [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2008-. Appendices. 2008 Jun 23 [Updated 2021 Mar 14]. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK279684/>
- [22]Microsoft. (2022). - *Microsoft SQL Server Management Studio 18* (v18.11.1) [programari]
- [23]Nogueira, W. G., Jaiswal, A. K., Tiwari, S., Ramos, R. T. J., Ghosh, P., Barh, D., Azevedo, V., & Soares, S. C. (2021). Computational identification of putative common genomic drug and vaccine targets in *Mycoplasma genitalium*. *Genomics*, 113(4), 2730–2743. <https://doi.org/10.1016/j.ygeno.2021.06.011>

9 Annexos

9.1 Llistat de proteïnes de *M. genitalium*

Taula 4. Llista de proteïnes de *Mycoplasma genitalium*. Remarcades en groc i negreta les que són factors de virulència coneguts [4].

Entrada (ID)	Nom d'entrada a UniProt	Nom proteïna / Funció canònica
Q9ZB73	PBDGT_MYCGE	Processive diacylglycerol beta-glycosyltransferase
P47543	G3P_MYCGE	Glyceraldehyde-3-phosphate dehydrogenase
P47304	KPRS_MYCGE	Ribose-phosphate pyrophosphokinase
P47632	CH60_MYCGE	Chaperonin GroEL
P47696	HGPRT_MYCGE	Hypoxanthine-guanine phosphoribosyltransferase
P47457	PFKA_MYCGE	Probable ATP-dependent 6-phosphofructokinase
P47634	GLYA_MYCGE	Serine hydroxymethyltransferase
P47417	KAD_MYCGE	Adenylate kinase
P47695	FTSH_MYCGE	ATP-dependent zinc metalloprotease FtsH
P47368	TOP1_MYCGE	DNA topoisomerase 1
P47331	HPRK_MYCGE	HPr kinase/phosphorylase
P47669	GPMI_MYCGE	2,3-bisphosphoglycerate-independent phosphoglycerate mutase
P55825	FPG_MYCGE	Formamidopyrimidine-DNA glycosylase
P47259	FOLD_MYCGE	Bifunctional protein FOLD
P47482	NNAAH_MYCGE	Probable nicotinate-nucleotide adenyltransferase/Ap4A hydrolase
P47458	KPYK_MYCGE	Pyruvate kinase
P47612	THII_MYCGE	Probable tRNA sulfurtransferase
P47542	PGK_MYCGE	Phosphoglycerate kinase
P47587	SYI_MYCGE	Isoleucine--tRNA ligase
P47599	ACKA_MYCGE	Acetate kinase
P47293	METK_MYCGE	S-adenosylmethionine synthase
P47658	DPO3X_MYCGE	DNA polymerase III subunit gamma/tau
Q49408	NAOX_MYCGE	Probable NADH oxidase
P47351	DACB_MYCGE	Diadenylate cyclase
P47445	PARE_MYCGE	DNA topoisomerase 4 subunit B
P47647	ENO_MYCGE	Enolase
P47429	PEPF_MYCGE	Oligoendopeptidase F homolog
Q49427	PLSX_MYCGE	Phosphate acyltransferase
P47456	SCPB_MYCGE	Segregation and condensation protein B
P47576	SYV_MYCGE	Valine--tRNA ligase
P47276	UPP_MYCGE	Uracil phosphoribosyltransferase
P47448	UVRC_MYCGE	UvrABC system protein C
P47486	UVRD_MYCGE	Probable DNA helicase II homolog
P47319	UVRB_MYCGE	UvrABC system protein B

P47622	URK_MYCGE	Uridine kinase
P47660	UVRA_MYCGE	UvrABC system protein A
P47515	ODPB_MYCGE	Pyruvate dehydrogenase E1 component subunit beta
Q49417	P32_MYCGE	P32 adhesin
Q9ZB82	RL332_MYCGE	50S ribosomal protein L33 2
P47703	RNPA_MYCGE	Ribonuclease P protein component
P47615	SYT_MYCGE	Threonine--tRNA ligase
P47372	SYW_MYCGE	Tryptophan--tRNA ligase
P47525	SYP_MYCGE	Proline--tRNA ligase
P47251	SYS_MYCGE	Serine--tRNA ligase
P47267	SYM_MYCGE	Methionine--tRNA ligase
P47500	RF1_MYCGE	Peptide chain release factor 1
P47318	SECA_MYCGE	Protein translocase subunit SecA
P47480	TIG_MYCGE	Trigger factor
P47315	PTG3C_MYCGE	PTS system glucose-specific EIICBA component
Q49425	RUVB_MYCGE	Holliday junction ATP-dependent DNA helicase RuvB
P47356	RSGA_MYCGE	Small ribosomal subunit biogenesis GTPase RsgA
P47464	RSMH_MYCGE	Ribosomal RNA small subunit methyltransferase H
P47701	RSMA_MYCGE	Ribosomal RNA small subunit methyltransferase A
P47302	RSMI_MYCGE	Ribosomal RNA small subunit methyltransferase I
P47620	RSMG_MYCGE	Ribosomal RNA small subunit methyltransferase G
Q49424	RUVA_MYCGE	Holliday junction ATP-dependent DNA helicase RuvA
P47672	PYRH_MYCGE	Uridylate kinase
P47355	PKNS_MYCGE	Putative serine/threonine-protein kinase
P47520	SPOT_MYCGE	Probable guanosine-3'5'-bis(diphosphate) 3'-pyrophosphohydrolase
P47339	RL9_MYCGE	50S ribosomal protein L9
P47376	RNY_MYCGE	Ribonuclease Y
P47396	RS10_MYCGE	30S ribosomal protein S10
P47455	SCPA_MYCGE	Segregation and condensation protein A
P47670	TPIS_MYCGE	Triosephosphate isomerase
P47330	TILS_MYCGE	tRNA(Ile)-lysidine synthase
P47663	RS15_MYCGE	30S ribosomal protein S15
P58061	SECG_MYCGE	Probable protein-export membrane protein SecG
P47412	RL6_MYCGE	50S ribosomal protein L6
P47421	RS13_MYCGE	30S ribosomal protein S13
P47332	LGT_MYCGE	Phosphatidylglycerol--prolipoprotein diacylglyceryl transferase
P47418	MAP1_MYCGE	Methionine aminopeptidase
P47299	MANB_MYCGE	Phosphomannomutase
P47269	ALF_MYCGE	Fructose-bisphosphate aldolase
Q49396	K1PF_MYCGE	Putative 1-phosphofructokinase
P47280	KITH_MYCGE	Thymidine kinase
P47364	GALE_MYCGE	UDP-glucose 4-epimerase
Q49405	DPO3A_MYCGE	DNA polymerase III subunit alpha
P47666	HPRR_MYCGE	Hydroperoxide reductase
P47335	EFG_MYCGE	Elongation factor G
P47539	FTSY_MYCGE	Signal recognition particle receptor FtsY

P47524	GREA_MYCGE	Transcription elongation factor GreA
Q49413	HMW1_MYCGE	Cytadherence high molecular weight protein 1
P47566	AMPP_MYCGE	Putative Xaa-Pro aminopeptidase
P47691	GALU_MYCGE	UTP--glucose-1-phosphate uridylyltransferase
P47641	ATPA_MYCGE	ATP synthase subunit alpha
P47384	LEPA_MYCGE	Elongation factor 4
P47252	KTHY_MYCGE	Thymidylate kinase
P47481	LON_MYCGE	Lon protease
P47254	MNME_MYCGE	tRNA modification GTPase MnmE
P47512	LPLA_MYCGE	Probable lipoate-protein ligase A
P47623	NADE_MYCGE	NH(3)-dependent NAD(+) synthetase
P47447	HRCA_MYCGE	Heat-inducible transcription repressor HrcA
P47324	OPPC_MYCGE	Oligopeptide transport system permease protein OppC
P47624	OBG_MYCGE	GTPase Obg
P47645	ATP6_MYCGE	ATP synthase subunit a
Q49421	ESL3_MYCGE	Putative esterase/lipase 3
P47571	DER_MYCGE	GTPase Der
P47513	DLDH_MYCGE	Dihydrolipoyl dehydrogenase
Q49433	CDSA_MYCGE	Putative phosphatidate cytidylyltransferase
Q49406	EX53_MYCGE	5'-3' exonuclease
P47374	NADK_MYCGE	NAD kinase
P47492	DNAG_MYCGE	DNA primase
P47593	IPYR_MYCGE	Inorganic pyrophosphatase
P47285	GLPO_MYCGE	Glycerol 3-phosphate oxidase
P47353	KGUA_MYCGE	Guanylate kinase
P13927	EFTU_MYCGE	Elongation factor Tu
P47642	ATPD_MYCGE	ATP synthase subunit delta
P47265	DNAJ_MYCGE	Chaperone protein DnaJ
P47477	END4_MYCGE	Probable endonuclease 4
P47284	GLPK_MYCGE	Glycerol kinase
Q49412	ESL1_MYCGE	Putative esterase/lipase 1
P47346	GATB_MYCGE	Aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit B
P47279	GLPF_MYCGE	Probable glycerol uptake facilitator protein
P47466	FTSZ_MYCGE	Cell division protein FtsZ
P47296	DEOC_MYCGE	Deoxyribose-phosphate aldolase
P47352	DEF_MYCGE	Peptide deformylase
P47644	ATPL_MYCGE	ATP synthase subunit c
P47463	MRAZ_MYCGE	Transcriptional regulator MraZ
P47648	MSRA_MYCGE	Peptide methionine sulfoxide reductase MsrA
P47326	OPPF_MYCGE	Oligopeptide transport ATP-binding protein OppF
P47300	NUSG_MYCGE	Transcription termination/antitermination protein NusG
P47489	PLSY_MYCGE	Glycerol-3-phosphate acyltransferase
P47290	POTC_MYCGE	Spermidine/putrescine transport system permease protein PotC homolog
P47289	POTB_MYCGE	Spermidine/putrescine transport system permease protein PotB homolog
P47487	MTHFS_MYCGE	5-formyltetrahydrofolate cyclo-ligase
P47627	ERA_MYCGE	GTPase Era

P47640	ATPG_MYCGE	ATP synthase gamma chain
Q49420	CSD_MYCGE	Probable cysteine desulfurase
P47577	ENGB_MYCGE	Probable GTP-binding protein EngB
P47443	GRPE_MYCGE	Protein GrpE
P47633	CH10_MYCGE	Co-chaperonin GroES
Q49418	ESL2_MYCGE	Putative esterase/lipase 2
P47541	PTAS_MYCGE	Phosphate acetyltransferase
P47472	NRDI_MYCGE	Protein NrdI
P47514	ODP2_MYCGE	Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex
P47287	PTHP_MYCGE	Phosphocarrier protein HPr
P47572	KCY_MYCGE	Cytidylate kinase
Q49422	RECU_MYCGE	Holliday junction resolvase RecU
P47439	RL35_MYCGE	50S ribosomal protein L35
P47420	RL36_MYCGE	50S ribosomal protein L36
P47397	RL3_MYCGE	50S ribosomal protein L3
P47398	RL4_MYCGE	50S ribosomal protein L4
P47303	RNM5_MYCGE	Ribonuclease M5
P47619	MNMG_MYCGE	tRNA uridine 5-carboxymethylaminomethyl modification enzyme MnmG
Q49400	MTM1_MYCGE	Type II methyltransferase M.MgeORF184P
P47400	RL2_MYCGE	50S ribosomal protein L2
P47441	RNH3_MYCGE	Ribonuclease HIII
P47325	OPPD_MYCGE	Oligopeptide transport ATP-binding protein OppD
P47387	NUSA_MYCGE	Transcription termination/antitermination protein NusA
P47354	PPH_MYCGE	Putative protein phosphatase
P47474	RL21_MYCGE	50S ribosomal protein L21
P47402	RL22_MYCGE	50S ribosomal protein L22
P47704	RL34_MYCGE	50S ribosomal protein L34
P47358	RPE_MYCGE	Ribulose-phosphate 3-epimerase
P47618	SYR_MYCGE	Arginine--tRNA ligase
P47508	SYL_MYCGE	Leucine--tRNA ligase
P47532	P29_MYCGE	Probable ABC transporter ATP-binding protein p29
P47682	RL19_MYCGE	50S ribosomal protein L19
P47607	RNC_MYCGE	Ribonuclease 3
P47491	SIGA_MYCGE	RNA polymerase sigma factor SigA
P47656	RS9_MYCGE	30S ribosomal protein S9
P47411	RS8_MYCGE	30S ribosomal protein S8
P57085	RS21_MYCGE	30S ribosomal protein S21
Q49434	T1SX_MYCGE	Putative type I specificity subunit S.MgeORF438P
P22746	NRNA_MYCGE	Bifunctional oligoribonuclease and PAP phosphatase NrnA
P47401	RS19_MYCGE	30S ribosomal protein S19
P47415	RL15_MYCGE	50S ribosomal protein L15
P47700	SYE_MYCGE	Glutamate--tRNA ligase
P47328	RL1_MYCGE	50S ribosomal protein L1
P47405	RL29_MYCGE	50S ribosomal protein L29
P47422	RS11_MYCGE	30S ribosomal protein S11
P47414	RS5_MYCGE	30S ribosomal protein S5

P47553	RS4_MYCGE	30S ribosomal protein S4
P47336	RS6_MYCGE	30S ribosomal protein S6
P47537	MNMA_MYCGE	tRNA-specific 2-thiouridylase MnmA
P47408	RL24_MYCGE	50S ribosomal protein L24
P47673	RRF_MYCGE	Ribosome-recycling factor
P47423	RPOA_MYCGE	DNA-directed RNA polymerase subunit alpha
P47327	RL11_MYCGE	50S ribosomal protein L11
P47603	RL32_MYCGE	50S ribosomal protein L32
P47446	PARC_MYCGE	DNA topoisomerase 4 subunit A
P47706	PARA_MYCGE	ParA family protein MG470
P47651	PSTA_MYCGE	Phosphate transport system permease protein PstA homolog
P47268	RPOE_MYCGE	Probable DNA-directed RNA polymerase subunit delta
P47406	RS17_MYCGE	30S ribosomal protein S17
P55750	RS20_MYCGE	30S ribosomal protein S20
P47473	RIR1_MYCGE	Ribonucleoside-diphosphate reductase subunit alpha
P47413	RL18_MYCGE	50S ribosomal protein L18
P47583	RPOB_MYCGE	DNA-directed RNA polymerase subunit beta
P47294	SRP54_MYCGE	Signal recognition particle protein
P47437	SYFB_MYCGE	Phenylalanine--tRNA ligase beta subunit
P47693	SYT_MYCGE	Tyrosine--tRNA ligase
P47410	RS14Z_MYCGE	30S ribosomal protein S14 type Z
P47567	RL331_MYCGE	50S ribosomal protein L33 1
P47440	RL20_MYCGE	50S ribosomal protein L20
P47409	RL5_MYCGE	50S ribosomal protein L5
P47350	RNR_MYCGE	Ribonuclease R
P47684	RS16_MYCGE	30S ribosomal protein S16
P47333	RS12_MYCGE	30S ribosomal protein S12
P47399	RL23_MYCGE	50S ribosomal protein L23
P47540	SMC_MYCGE	Chromosome partition protein Smc
P47499	RL31_MYCGE	50S ribosomal protein L31
P36255	RL7_MYCGE	50S ribosomal protein L7/L12
P47338	RS18_MYCGE	30S ribosomal protein S18
P47495	SYC_MYCGE	Cysteine--tRNA ligase
P47391	RIBF_MYCGE	Bifunctional riboflavin kinase/FMN adenylyltransferase
P47636	RPIB_MYCGE	Probable ribose-5-phosphate isomerase B
P47305	SSRP_MYCGE	SsrA-binding protein
Q9ZB80	Y149A_MYCGE	Uncharacterized protein MG149.1
P47395	Y149_MYCGE	Uncharacterized lipoprotein MG149
P47306	Y060_MYCGE	Uncharacterized glycosyltransferase MG060
P47359	SYN_MYCGE	Asparagine--tRNA ligase
P47416	SECY_MYCGE	Protein translocase subunit SecY
P47337	SSB_MYCGE	Single-stranded DNA-binding protein
P47471	RIR2_MYCGE	Ribonucleoside-diphosphate reductase subunit beta
P47348	TRXB_MYCGE	Thioredoxin reductase
P47282	SYD_MYCGE	Aspartate--tRNA ligase
P36263	RL10_MYCGE	50S ribosomal protein L10

P47424	RL17_MYCGE	50S ribosomal protein L17
P47407	RL14_MYCGE	50S ribosomal protein L14
P47476	RL27_MYCGE	50S ribosomal protein L27
P47665	RL28_MYCGE	50S ribosomal protein L28
P47436	SYFA_MYCGE	Phenylalanine--tRNA ligase alpha subunit
P47534	SYA_MYCGE	Alanine--tRNA ligase
P47379	Y133_MYCGE	Uncharacterized protein MG133
P47278	Y032_MYCGE	Uncharacterized protein MG032
P47288	POTA_MYCGE	Spermidine/putrescine import ATP-binding protein PotA
P47702	YIDC_MYCGE	Membrane protein insertase YidC
P47516	ODPA_MYCGE	Pyruvate dehydrogenase E1 component subunit alpha
P47323	OPPB_MYCGE	Oligopeptide transport system permease protein OppB
P47249	GYRB_MYCGE	DNA gyrase subunit B
P47385	RNJ_MYCGE	Ribonuclease J
P47403	RS3_MYCGE	30S ribosomal protein S3
P47370	THIO_MYCGE	Thioredoxin
P47253	Y007_MYCGE	Uncharacterized protein MG007
P47637	Y397_MYCGE	Uncharacterized protein MG397
P47606	Y366_MYCGE	Uncharacterized protein MG366
P47522	Y280_MYCGE	Uncharacterized protein MG280
P47609	Y369_MYCGE	Uncharacterized protein MG369
Q49403	Y220_MYCGE	Uncharacterized protein MG220
P22747	MGP3_MYCGE	Mgp-operon protein 3
Q49411	Y294_MYCGE	Uncharacterized protein MG294
P47334	RS7_MYCGE	30S ribosomal protein S7
P47677	Y439_MYCGE	Uncharacterized lipoprotein MG439
P47404	RL16_MYCGE	50S ribosomal protein L16
Q49329	Y269A_MYCGE	Uncharacterized protein MG269.1
P47621	Y381_MYCGE	Uncharacterized protein MG381
P47592	Y350_MYCGE	Uncharacterized protein MG350
P47459	P65H_MYCGE	Proline-rich P65 protein homolog
P47533	P69_MYCGE	ABC transport system permease protein p69
P47360	PGSA_MYCGE	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase
P47281	SYH_MYCGE	Histidine--tRNA ligase
P47373	Y127_MYCGE	Uncharacterized protein MG127
P47494	Y252_MYCGE	Uncharacterized tRNA/rRNA methyltransferase MG252
P47657	RL13_MYCGE	50S ribosomal protein L13
P47382	SYK_MYCGE	Lysine--tRNA ligase
P47582	RPOC_MYCGE	DNA-directed RNA polymerase subunit beta'
P47316	RS2_MYCGE	30S ribosomal protein S2
Q2MHT0	Y141A_MYCGE	Uncharacterized protein MG141.1
P47584	Y342_MYCGE	Uncharacterized protein MG342
P47434	Y188_MYCGE	Probable ABC transporter permease protein MG188
Q49402	PLSC_MYCGE	Probable 1-acyl-sn-glycerol-3-phosphate acyltransferase
P47386	Y140_MYCGE	Uncharacterized ATP-dependent helicase MG140
P47256	Y010_MYCGE	Uncharacterized protein MG010

P47549	Y307_MYCGE	Uncharacterized lipoprotein MG307
P47274	Y028_MYCGE	Uncharacterized protein MG028
P47628	YBEY_MYCGE	Endoribonuclease YbeY
P47427	Y181_MYCGE	Uncharacterized protein MG181
P47283	Y037_MYCGE	Uncharacterized protein MG037
P47380	Y134_MYCGE	Uncharacterized protein MG134
Q49430	Y390_MYCGE	Putative ABC transporter ATP-binding MG390
P47568	Y326_MYCGE	DegV domain-containing protein MG326
P47478	Y236_MYCGE	Uncharacterized protein MG236
P47585	Y343_MYCGE	Uncharacterized protein MG343
P47505	Y263_MYCGE	Putative phosphatase MG263
P47451	Y209_MYCGE	Uncharacterized RNA pseudouridine synthase MG209
Q9ZB74	Y323A_MYCGE	Uncharacterized protein MG323.1
Q49423	Y356_MYCGE	Uncharacterized protein MG356
P47705	Y467_MYCGE	Putative ABC transporter ATP-binding protein MG467
P47616	Y376_MYCGE	Uncharacterized protein MG376
P47652	Y412_MYCGE	Uncharacterized lipoprotein MG412
P47483	Y241_MYCGE	Uncharacterized protein MG241
P47367	Y121_MYCGE	Uncharacterized protein MG121
P47545	Y303_MYCGE	Putative ABC transporter ATP-binding protein MG303
P47528	Y286_MYCGE	Uncharacterized protein MG286
P47596	Y354_MYCGE	Uncharacterized protein MG354
P47347	Y101_MYCGE	Uncharacterized HTH-type transcriptional regulator MG101
P47313	Y067_MYCGE	Uncharacterized lipoprotein MG067
P47650	PSTB_MYCGE	Phosphate import ATP-binding protein PstB
P47588	TRML_MYCGE	Putative tRNA (cytidine(34)-2'-O)-methyltransferase
P47261	Y015_MYCGE	Putative ABC transporter ATP-binding protein MG015
P47310	Y064_MYCGE	Uncharacterized ABC transporter permease MG064
P47479	Y237_MYCGE	Uncharacterized protein MG237
P47311	Y065_MYCGE	Putative ABC transporter ATP-binding protein MG065
P47435	Y189_MYCGE	Probable ABC transporter permease protein MG189
P47590	Y348_MYCGE	Uncharacterized lipoprotein MG348
P47257	Y011_MYCGE	Uncharacterized protein MG011
P47490	Y248_MYCGE	Putative tRNA methyltransferase MG248
P47493	SYG_MYCGE	Glycine--tRNA ligase
P47362	Y116_MYCGE	Uncharacterized protein MG116
P47341	Y095_MYCGE	Uncharacterized lipoprotein MG095
P47329	PTH_MYCGE	Peptidyl-tRNA hydrolase
P47485	Y243_MYCGE	Uncharacterized protein MG243
P47258	Y012_MYCGE	Uncharacterized protein MG012
P47449	Y207_MYCGE	Putative metallophosphoesterase MG207
P47393	Y147_MYCGE	Uncharacterized protein MG147
Q49415	Y314_MYCGE	Uncharacterized protein MG314
P47699	Y461_MYCGE	Uncharacterized protein MG461
P47685	Y447_MYCGE	Uncharacterized protein MG447
P47273	Y027_MYCGE	Uncharacterized protein MG027

P47653	Y414_MYCGE	Uncharacterized protein MG414
P47655	Y415_MYCGE	Uncharacterized protein MG415
P47681	Y443_MYCGE	Uncharacterized protein MG443
P47270	YCHF_MYCGE	Ribosome-binding ATPase YchF
P47394	Y148_MYCGE	Uncharacterized protein MG148
P47431	Y185_MYCGE	Uncharacterized lipoprotein MG185
P47488	Y246_MYCGE	Putative phosphatase/phosphodiesterase MG246
P47649	Y409_MYCGE	Uncharacterized protein MG409
P47343	UNG_MYCGE	Uracil-DNA glycosylase
P47264	Y018_MYCGE	Uncharacterized ATP-dependent helicase MG018
P47369	Y123_MYCGE	Uncharacterized protein MG123
P47589	TRMB_MYCGE	tRNA (guanine-N(7)-)-methyltransferase
P47260	Y014_MYCGE	Putative ABC transporter ATP-binding protein MG014
P47255	Y009_MYCGE	Uncharacterized metal-dependent hydrolase MG009
P47548	Y306_MYCGE	Uncharacterized protein MG306
P47444	Y202_MYCGE	Uncharacterized protein MG202
Q49428	Y371_MYCGE	Uncharacterized protein MG371
Q9ZB72	Y350A_MYCGE	Uncharacterized protein MG350.1
Q9ZB71	Y384A_MYCGE	Uncharacterized protein MG384.1
P47565	Y323_MYCGE	Uncharacterized protein MG323
P47563	Y321_MYCGE	Uncharacterized lipoprotein MG321
P47530	Y288_MYCGE	Uncharacterized protein MG288
P47683	TRMD_MYCGE	tRNA (guanine-N(1)-)-methyltransferase
P47344	Y098_MYCGE	Uncharacterized protein MG098
P56723	Y210A_MYCGE	Uncharacterized protein MG210.1
Q49426	Y360_MYCGE	Uncharacterized protein MG360
P47307	Y061_MYCGE	Major facilitator superfamily transporter MG061
P52271	Y308_MYCGE	Probable ATP-dependent RNA helicase MG308
Q49310	Y335A_MYCGE	UPF0154 protein MG335.1
P47453	Y211_MYCGE	Uncharacterized protein MG211
P47266	PIP_MYCGE	Putative proline iminopeptidase
P47320	Y074_MYCGE	Uncharacterized protein MG074
P47465	Y223_MYCGE	Uncharacterized protein MG223
P47469	TYSY_MYCGE	Thymidylate synthase
P47349	WHIA_MYCGE	Probable cell division protein WhiA
P47286	Y040_MYCGE	Uncharacterized lipoprotein MG040
P47321	Y075_MYCGE	Uncharacterized protein MG075
P47573	Y331_MYCGE	Uncharacterized protein MG331
Q49409	Y277_MYCGE	Uncharacterized protein MG277
P47558	Y316_MYCGE	Uncharacterized protein MG316
P47432	Y186_MYCGE	Uncharacterized lipoprotein MG186
P47661	Y422_MYCGE	Uncharacterized protein MG422
P47390	Y144_MYCGE	Uncharacterized protein MG144
P47546	Y304_MYCGE	Putative ABC transporter ATP-binding protein MG304
Q49416	Y315_MYCGE	Uncharacterized protein MG315
P47604	Y364_MYCGE	Uncharacterized protein MG364

P47564	Y322_MYCGE	Uncharacterized cation transporter MG322
P47617	Y377_MYCGE	Uncharacterized protein MG377
P47363	Y117_MYCGE	Uncharacterized protein MG117
P47314	Y068_MYCGE	Uncharacterized lipoprotein MG068
P47580	Y338_MYCGE	Uncharacterized lipoprotein MG338
P47510	Y268_MYCGE	Uncharacterized protein MG268
P47535	Y293_MYCGE	Uncharacterized protein MG293
P47610	Y370_MYCGE	Uncharacterized RNA pseudouridine synthase MG370
P47561	Y319_MYCGE	Uncharacterized protein MG319
Q49407	Y269_MYCGE	Uncharacterized protein MG269
P47591	Y349_MYCGE	Uncharacterized protein MG349
P47475	Y233_MYCGE	Uncharacterized protein MG233
P47664	Y425_MYCGE	Probable ATP-dependent RNA helicase MG425
P47378	YHIT_MYCGE	Uncharacterized HIT-like protein MG132
P47377	Y131_MYCGE	Uncharacterized protein MG131
P47342	Y096_MYCGE	Uncharacterized protein MG096
P47595	Y353_MYCGE	Uncharacterized protein MG353
P47484	Y242_MYCGE	Uncharacterized protein MG242
P47468	Y226_MYCGE	Uncharacterized protein MG226
P47527	Y285_MYCGE	Uncharacterized protein MG285
P47509	Y267_MYCGE	Uncharacterized protein MG267
P47433	Y187_MYCGE	Putative ABC transporter ATP-binding protein MG187
P47625	Y385_MYCGE	Uncharacterized protein MG385
P47507	Y265_MYCGE	Putative phosphatase MG265
P47678	Y440_MYCGE	Uncharacterized protein MG440
P47365	Y119_MYCGE	Putative carbohydrate transport ATP-binding protein MG119
P47301	Y055_MYCGE	Uncharacterized protein MG055
P47366	Y120_MYCGE	Uncharacterized protein MG120
Q9ZB76	YQGF_MYCGE	Putative pre-16S rRNA nuclease
P47692	OHRL_MYCGE	Organic hydroperoxide resistance protein-like
Q49410	P37_MYCGE	High affinity transport system protein p37
Q49429	P200_MYCGE	Protein P200
P47668	PT1_MYCGE	Phosphoenolpyruvate-protein phosphotransferase
P47574	Y332_MYCGE	Probable transcriptional regulatory protein MG332
P47450	Y208_MYCGE	Uncharacterized protein MG208
P47428	TRUA_MYCGE	tRNA pseudouridine synthase A
P47271	Y025_MYCGE	Uncharacterized glycosyltransferase MG025
P47502	Y260_MYCGE	Uncharacterized lipoprotein MG260
P47579	Y337_MYCGE	Uncharacterized protein MG337
Q49399	Y146_MYCGE	UPF0053 protein MG146
P47679	Y441_MYCGE	Uncharacterized protein MG441
Q49397	Y129_MYCGE	Putative phosphotransferase enzyme IIB component MG129
P47551	Y309_MYCGE	Uncharacterized lipoprotein MG309
P47544	Y302_MYCGE	Uncharacterized protein MG302
P47526	Y284_MYCGE	Uncharacterized protein MG284
P47521	Y279_MYCGE	Uncharacterized protein MG279

P47538	Y296_MYCGE	Uncharacterized protein MG296
P47694	Y456_MYCGE	Uncharacterized protein MG456
Q49414	Y313_MYCGE	Uncharacterized protein MG313
P47275	Y029_MYCGE	Uncharacterized protein MG029
P47381	Y135_MYCGE	Uncharacterized protein MG135
Q49419	Y328_MYCGE	Uncharacterized protein MG328
Q9ZB77	Y255A_MYCGE	Uncharacterized protein MG255.1
P47629	Y389_MYCGE	Uncharacterized protein MG389
P47635	Y395_MYCGE	Uncharacterized lipoprotein MG395
P47562	Y320_MYCGE	Uncharacterized protein MG320
Q49436	Y459_MYCGE	Uncharacterized protein MG459
P47688	Y450_MYCGE	DegV domain-containing protein MG450
Q49432	Y432_MYCGE	Uncharacterized protein MG432
P47461	Y219_MYCGE	Uncharacterized protein MG219
Q49460	Y468_MYCGE	Uncharacterized ABC transporter permease MG468
P47292	TSAD_MYCGE	tRNA N6-adenosine threonylcarbamoyltransferase
P47297	TYPH_MYCGE	Thymidine phosphorylase
Q49431	Y406_MYCGE	Uncharacterized protein MG406
Q9ZB70	Y468A_MYCGE	Putative ABC transporter ATP-binding protein MG468.1
P47361	Y115_MYCGE	Protein MG115
Q9ZB81	Y055B_MYCGE	Uncharacterized protein MG055.2
P47523	Y281_MYCGE	Uncharacterized protein MG281
P47613	Y373_MYCGE	Uncharacterized protein MG373
P47662	Y423_MYCGE	Uncharacterized protein MG423
P47667	Y428_MYCGE	Uncharacterized protein MG428
P47371	Y125_MYCGE	Putative phosphatase MG125
P47312	TKT_MYCGE	Transketolase
P47322	Y076_MYCGE	Uncharacterized protein MG076
P47291	Y045_MYCGE	Uncharacterized lipoprotein MG045
P47687	Y449_MYCGE	Uncharacterized protein MG449
P47614	Y374_MYCGE	Uncharacterized protein MG374
P47497	Y255_MYCGE	Uncharacterized protein MG255
P47498	Y256_MYCGE	Uncharacterized protein MG256
Q9ZB78	Y218A_MYCGE	Uncharacterized protein MG218.1
P47690	Y452_MYCGE	Uncharacterized protein MG452
P47308	PTF3A_MYCGE	PTS system fructose-specific EIIABC component
P47467	Y225_MYCGE	Uncharacterized protein MG225
Q49404	Y259_MYCGE	Uncharacterized protein MG259
P47686	MSRB_MYCGE	Peptide methionine sulfoxide reductase MsrB
P47389	RBFA_MYCGE	Ribosome-binding factor A
Q49435	RBGA_MYCGE	Probable ribosome biogenesis GTPase A
P47581	RECA_MYCGE	Protein RecA
P47317	ATCL_MYCGE	Probable cation-transporting P-type ATPase
P47298	CDD_MYCGE	Cytidine deaminase
P47248	DNAJL_MYCGE	DnaJ-like protein MG002
P47442	DNAJM_MYCGE	DnaJ-like protein MG200

P47357	G6PI_MYCGE	Glucose-6-phosphate isomerase
P47438	IF3_MYCGE	Translation initiation factor IF-3
P47388	IF2_MYCGE	Translation initiation factor IF-2
P47643	ATPF_MYCGE	ATP synthase subunit b
P47639	ATPB_MYCGE	ATP synthase subunit beta
P35888	DNAA_MYCGE	Chromosomal replication initiator protein DnaA
P47250	GYRA_MYCGE	DNA gyrase subunit A
Q57081	HMW3_MYCGE	Cytadherence high molecular weight protein 3
P47460	HMW2_MYCGE	Cytadherence high molecular weight protein 2
P47345	GATA_MYCGE	Glutamyl-tRNA(Gln) amidotransferase subunit A
P47698	LDH_MYCGE	L-lactate dehydrogenase
Q49401	LSPA_MYCGE	Lipoprotein signal peptidase
P47295	DEOD_MYCGE	Purine nucleoside phosphorylase DeoD-type
P47246	EFTS_MYCGE	Elongation factor Ts
P47419	IF1_MYCGE	Translation initiation factor IF-1
P47340	DNAB_MYCGE	Replicative DNA helicase
P47277	DPO3_MYCGE	DNA polymerase III PolC-type
P47470	DYR_MYCGE	Dihydrofolate reductase
P47496	DNLJ_MYCGE	DNA ligase
P47425	ECFA1_MYCGE	Energy-coupling factor transporter ATP-binding protein EcfA1
P47426	ECFA2_MYCGE	Energy-coupling factor transporter ATP-binding protein EcfA2
Q9ZB79	ACPS_MYCGE	Holo-[acyl-carrier-protein] synthase
P47518	APT_MYCGE	Adenine phosphoribosyltransferase
P47597	CLPB_MYCGE	Chaperone protein ClpB
P47506	COAE_MYCGE	Dephospho-CoA kinase
Q49398	GLF_MYCGE	UDP-galactopyranose mutase
P47605	FMT_MYCGE	Methionyl-tRNA formyltransferase
P47247	DPO3B_MYCGE	Beta sliding clamp
P47575	AZOR_MYCGE	FMN-dependent NADH:quinone oxidoreductase
P47547	DNAK_MYCGE	Chaperone protein DnaK
P47631	AMPA_MYCGE	Probable cytosol aminopeptidase
P20796	ADP1_MYCGE	Adhesin P1
P47638	ATPE_MYCGE	ATP synthase epsilon chain
P47272	EFP_MYCGE	Elongation factor P
P47529	ACPH_MYCGE	Acyl carrier protein homolog