AP1 Bases de Datos: Exploración de Recursos y Herramientas de obtención de archivos

AP1.1 Explorar recursos del NCBI: LCT1 (2.5 puntos)

Instrucciones: Contesta en el espacio a la pregunta sin proporcionar captura de pantalla.

1.1.1 Cuales son las 5 primeras líneas de la secuencia (desde el archivo .fasta).

AACAGTTCCTAGAAAATGGAGCTGTCTTGGCATGTAGTCTTTATTGCCCTGCTAAGTTTTTCATGCTGGG

GGTCAGACTGGGAGTCTGATAGAAATTTCATTTCCACCGCTGGTCCTCTAACCAATGACTTGCTGCACAA

CCTGAGTGGTCTCCTGGGAGACCAGAGTTCTAACTTTGTAGCAGGGGACAAAGACATGTATGTTTGTCAC

CAGCCACTGCCCACTTTCCTGCCAGAATACTTCAGCAGTCTCCATGCCAGTCAGATCACCCATTATAAGG

TATTTCTGTCATGGGCACAGCTCCTCCCAGCAGGAAGCACCCAGAATCCAGACGAGAAAACAGTGCAGTG

1.1.2 Cuál es el número de exones que presenta este gen.

17 exones.

1.1.3 Cuales son los dos principales tejidos donde se expresa y su nivel de expresión relativo.

Se expresa principalmente en el intestino delgado, con un nivel de expresión RPKM de 73.8, y sobre todo en el duodeno, con un nivel de expresión RPKM de 121.7.

1.1.4 Nombra los dos genes que se sitúan up-stream y down-stream de la lactasa.

UBXN4 y MCM6

1.1.5 Que número de variantes consideradas patogénicas aparecen en la base de datos del NCBI ClinVar.

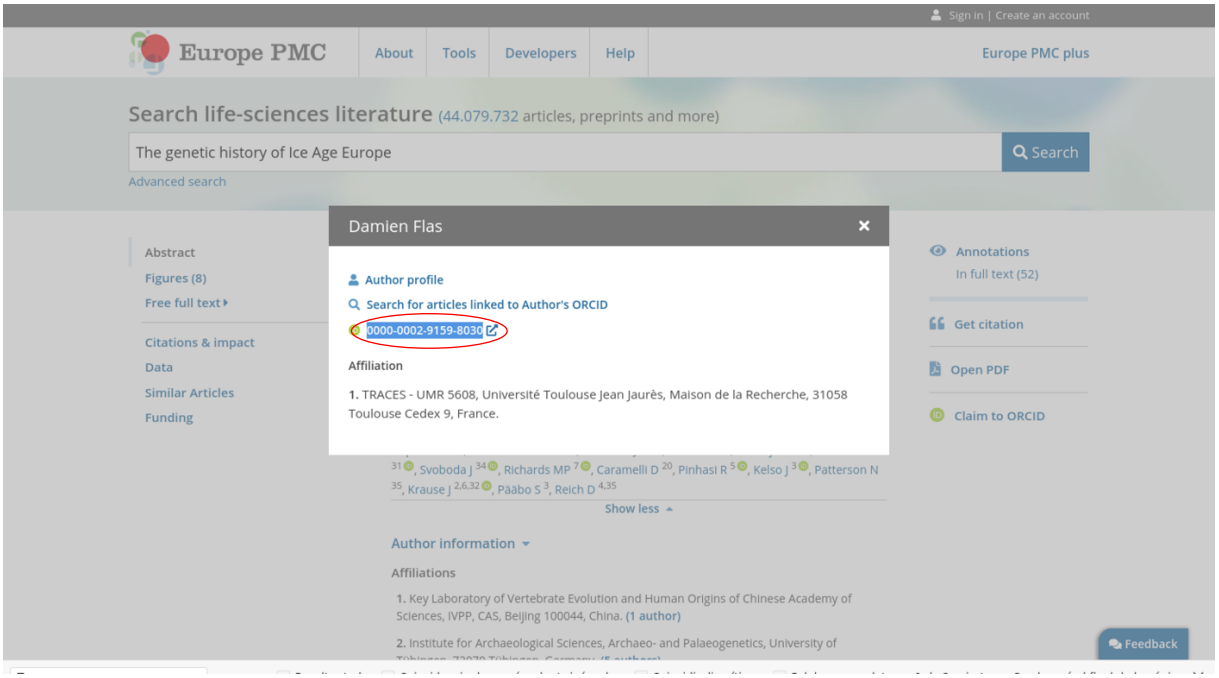
31 variantes patogénicas.

AP1.2 Explorar recursos del EMBL-EBI (4 puntos)

Instrucciones: Contesta en el espacio a la pregunta y pega debajo una captura de pantalla para cada una de las respuestas indicando claramente el lugar donde se encuentra la respuesta.

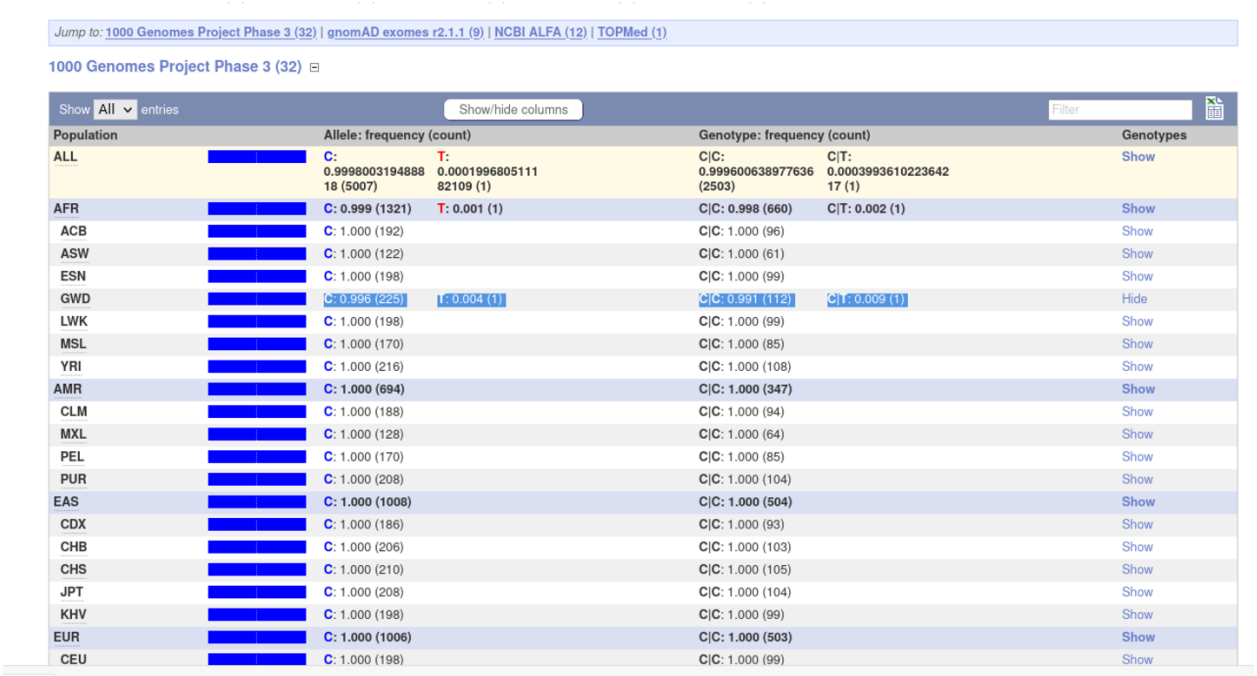
1.2.1 ¿Cuál es el Open Researcher and Contributor ID (ORCID) del coautor que trabaja en la Universidad de Toulouse?

El ORCID del coautor es el 0000-0002-9159-8030, que se encuentra señalado en rojo en la siguiente imagen:



1.2.2 ¿Cuál es el ID de la muestra del individuo con el genotipo C|T?

El ID de la muestra con el genotipo C|T es: HG02568 (F)



En la siguiente imagen se encuentra señalado en rojo el ID del genotipo buscado:

Population	Allele: frequency (count)
TOPMed	C: 0.999992036 T: 7.96400e-06

[back to top]

Genotypes for 1000GENOMES:phase_3:GWD

Show All entries | Show/hide columns | Filter

Sample (Male/Female/Unknown)	Genotype (forward strand)	Population(s)	Father	Mother
HG02568 (F)	C|T	AFR, ALL, GWD	-	-
HG02461 (M)	C|C	AFR, ALL, GWD	-	-
HG02462 (F)	C|C	AFR, ALL, GWD	-	-
HG02464 (M)	C|C	AFR, ALL, GWD	-	-
HG02465 (F)	C|C	AFR, ALL, GWD	-	-
HG02561 (M)	C|C	AFR, ALL, GWD	-	-
HG02562 (F)	C|C	AFR, ALL, GWD	-	-
HG02570 (M)	C|C	AFR, ALL, GWD	-	-
HG02571 (F)	C|C	AFR, ALL, GWD	-	-
HG02573 (M)	C|C	AFR, ALL, GWD	-	-
HG02574 (F)	C|C	AFR, ALL, GWD	-	-
HG02582 (M)	C|C	AFR, ALL, GWD	-	-
HG02583 (F)	C|C	AFR, ALL, GWD	-	-
HG02585 (M)	C|C	AFR, ALL, GWD	-	-
HG02586 (F)	C|C	AFR, ALL, GWD	-	-
HG02588 (M)	C|C	AFR, ALL, GWD	-	-
HG02589 (F)	C|C	AFR, ALL, GWD	-	-
HG02594 (M)	C|C	AFR, ALL, GWD	-	-
HG02595 (F)	C|C	AFR, ALL, GWD	-	-
HG02610 (M)	C|C	AFR, ALL, GWD	-	-
HG02611 (F)	C|C	AFR, ALL, GWD	-	-
HG02613 (M)	C|C	AFR, ALL, GWD	-	-

1.2.3 ¿Cuál es el número de acceso del protocolo de extracción de ácido nucleico utilizado en este estudio?

El número de acceso al protocolo de extracción de ácido nucleico es P-GSE55175-4, señalado en rojo en la siguiente imagen.

ArrayExpress Functional genomics data

BIOSTUDIES / ARRAYEXPRESS / E-GEOD-55175

Release Date: 20 February 2014 • Modified: 7 September 2023 • Views: 962

[Cite] {JSON} →PageTab↓ HTTP FTP Globus

Argentinean high altitude adaptation

Christina Eichstaedt [1], Maru Mormina , Toomas Kivisild , Tiago AntÃEo , Luca Pagani , Alexia Cardona

[1] University of Cambridge

Accession E-GEOD-55175

Study type comparative genomic hybridization by array EFO , genotyping by array EFO

Organism Homo sapiens

Description This study evaluates genetic and phenotypic variation in the high altitude Colla population living in the Argentinean Andes above 3500 m. They were compared to the WichÃ population living in the nearby lowlands of the Gran Chaco region. This study attempts to pinpoint evolutionary mechanisms underlying adaptation to hypobaric hypoxia. We have genotyped 25 individuals from both populations for 730,525 SNPs. DNA from 25 saliva samples from Collas living >3500 m and 25 saliva samples from WichÃ living <500 m from the Province of Salta in Argentina was genotyped

Protocols hide table

Name	Type	Description
P-GSE55175-2	sample treatment protocol	Storage at room temperature for 2 weeks, 1-3 weeks at 4°C, centrifugation
P-GSE55175-3	growth protocol	Saliva mixed with equal volume of lysis buffer (Quinque et al. 2006) to allow storage at room temperature
P-GSE55175-4	nucleic acid extraction protocol	Qiagen DNA Investigator Kit scaled up for 2 ml buffer-saliva mixture
P-GSE55175-1	normalization data transformation protocol	Image data was analyzed using Genomestudio v2011.1 with GT module 1.9.4 (Illumina). ID_REF = VALUE = Genotype: AA, AB, BB,or NC (no call) Score = Theta = R = B Allele Freq = Log R Ratio =
P-GSE55175-5	labelling protocol	Followed Illumina standard protocol

Data files

Show 5 entries Search:

Name	Size	Section
GSM1330779_sample_table.txt	45.1 MB	Process Data
GSM1330770_sample_table.txt	45.1 MB	Process Data
GSM1330799_sample_table.txt	44.6 MB	Process Data
GSM1330798_sample_table.txt	44.7 MB	Process Data
GSM1330782_sample_table.txt	45.3 MB	Process Data

Showing 1 to 5 of 52 entries

Previous 1 2 3 4 5 ... 11 Next

Download all files

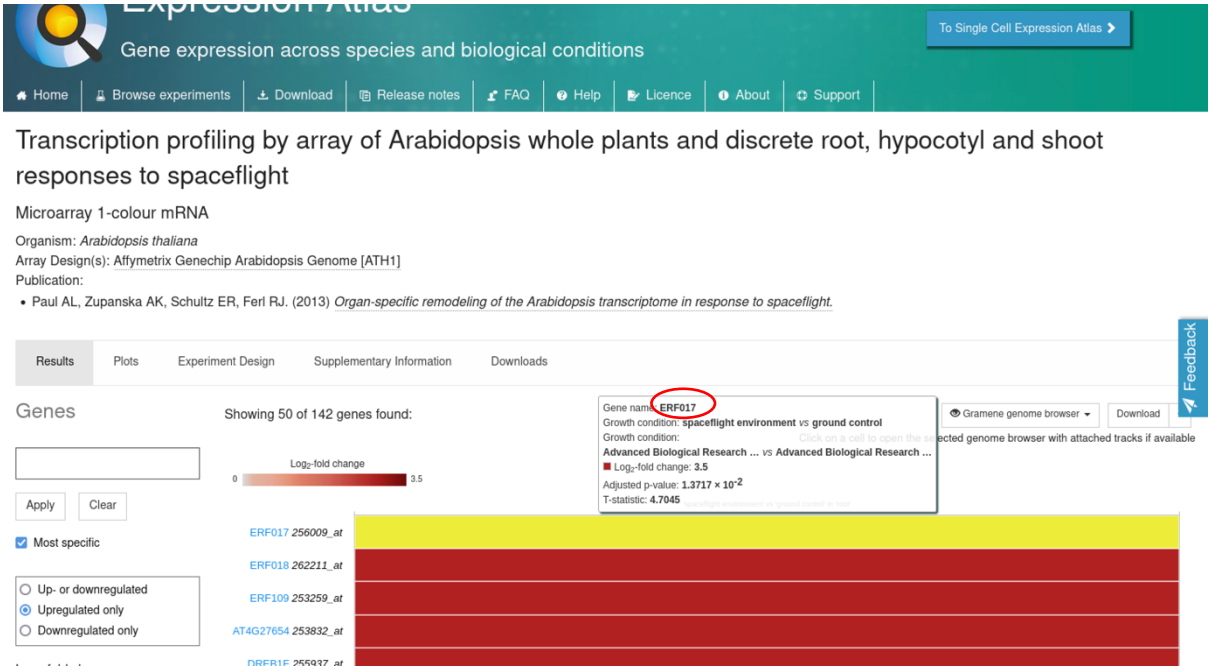
Linked Information

Show 5 entries Search:

Name Type Section

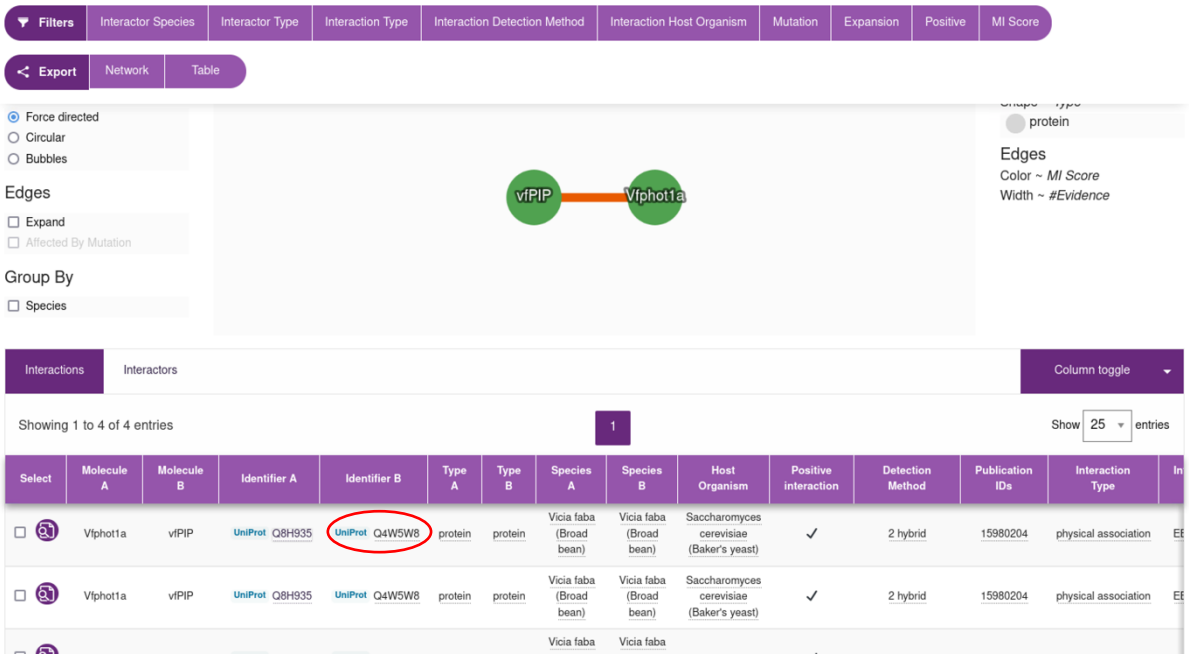
1.2.4 ¿Cuál es el nombre del gen más sobre-expresado en el tejido de la raíz?

El nombre del gen que está más sobre-expresado en la raíz es el ERF017, señalado en rojo a continuación:



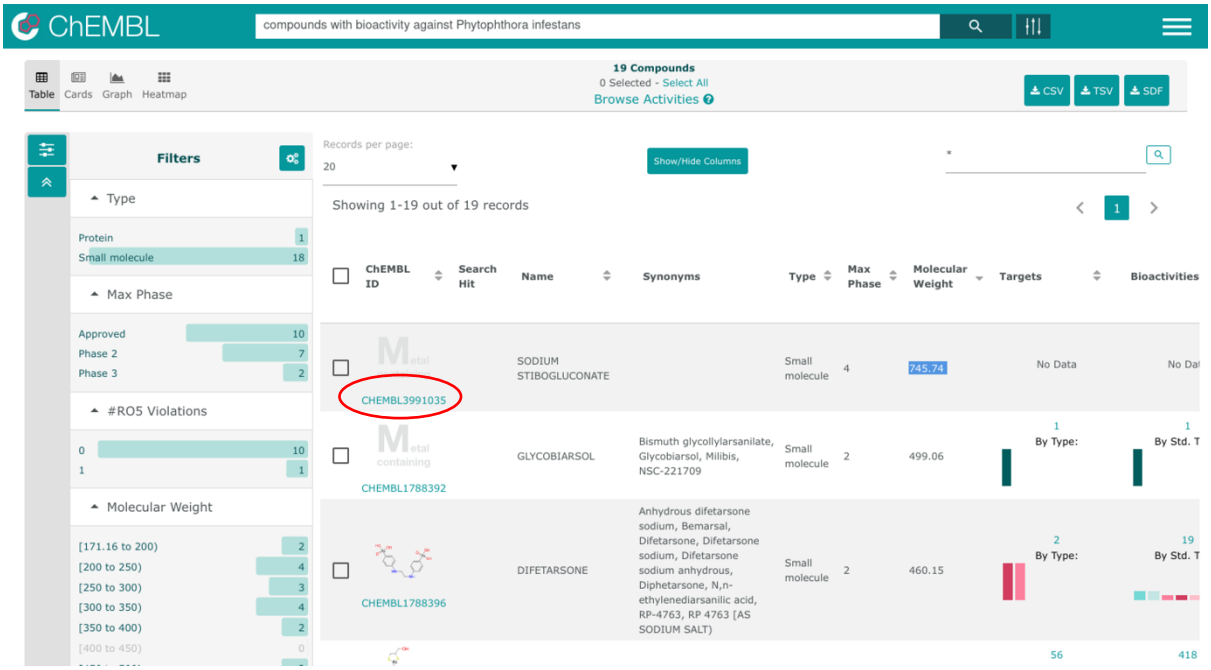
1.2.5 ¿Cuál es el número de acceso UniProt de la proteína con la que interactúa Vfphot1a?

El número de acceso a UniProt de la proteína con la que interactúa Vfphot1a, que es la vfPIP, es el Q4W5W8, rodeado en rojo a continuación:



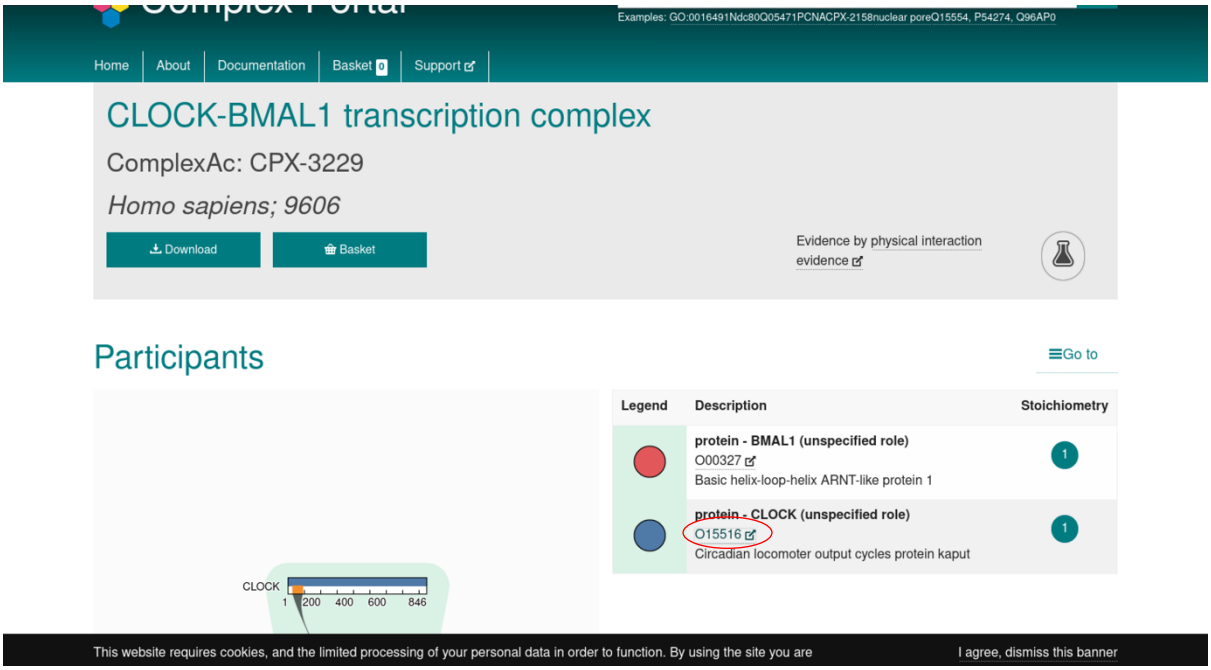
1.2.6 ¿Cuál es el ID ChEMBL del compuesto más grande (por peso molecular)?

Tras haber filtrado en ChEMBL por los compuestos que tienen bioactividad contra Phytophthora infestans y filtrar por peso molecular, el ID del compuesto de mayor peso molecular, y, por tanto, más grande, es el CHEMBL3991035, rodeado en rojo:



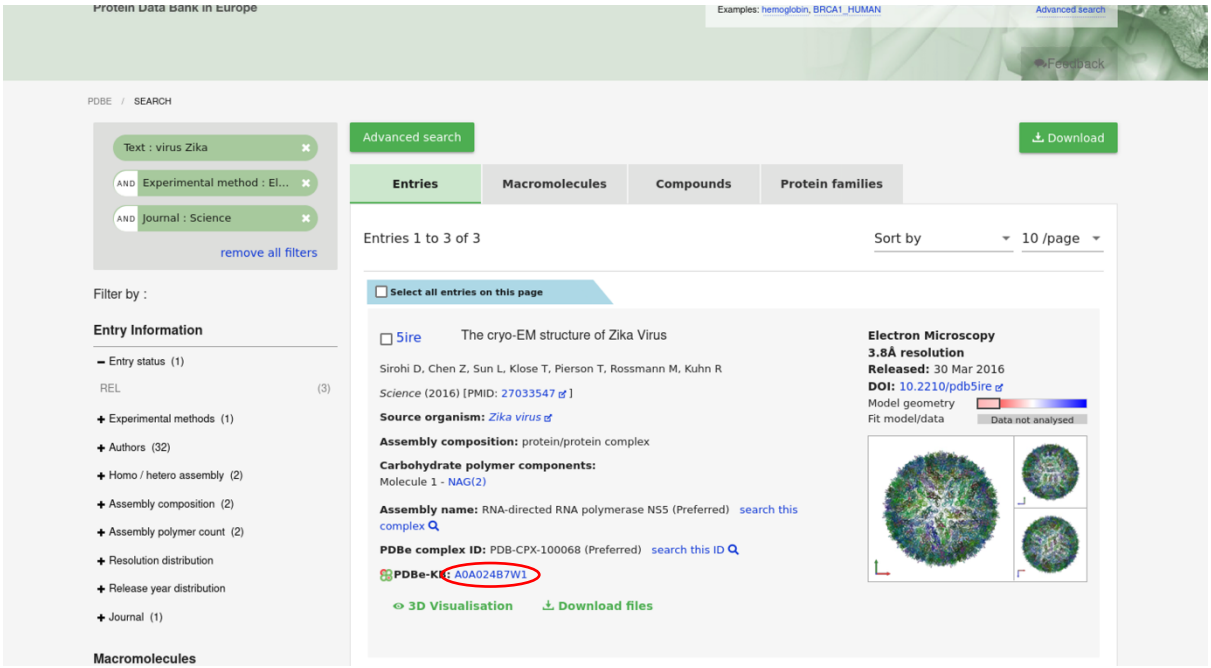
1.2.7 ¿Cuál es el número de acceso UniProt de la proteína ingeniosamente denominada común a estos complejos de transcripción?

El número de acceso UniProt de la proteína común, Clock, de los complejos de transcripción humanos implicados en los ritmos circadianos encontrado en la base de datos Complex Portal, es el O15516, señalado en rojo:



1.2.8 ¿Cuál es el número de acceso de la estructura correspondiente en el Banco de Datos de Microscopía Electrónica (EMDB)?

El número de acceso de la estructura del virus del Zika encontrado en la base de datos PDBe y contrastado en el Banco de Datos de Microscopía Electrónica según la entrada publicada en Science es el A0A024B7W1, señalado a continuación en rojo:



AP1.3 Obtener listado de archivos .fastq mediante terminal (3.5 puntos)

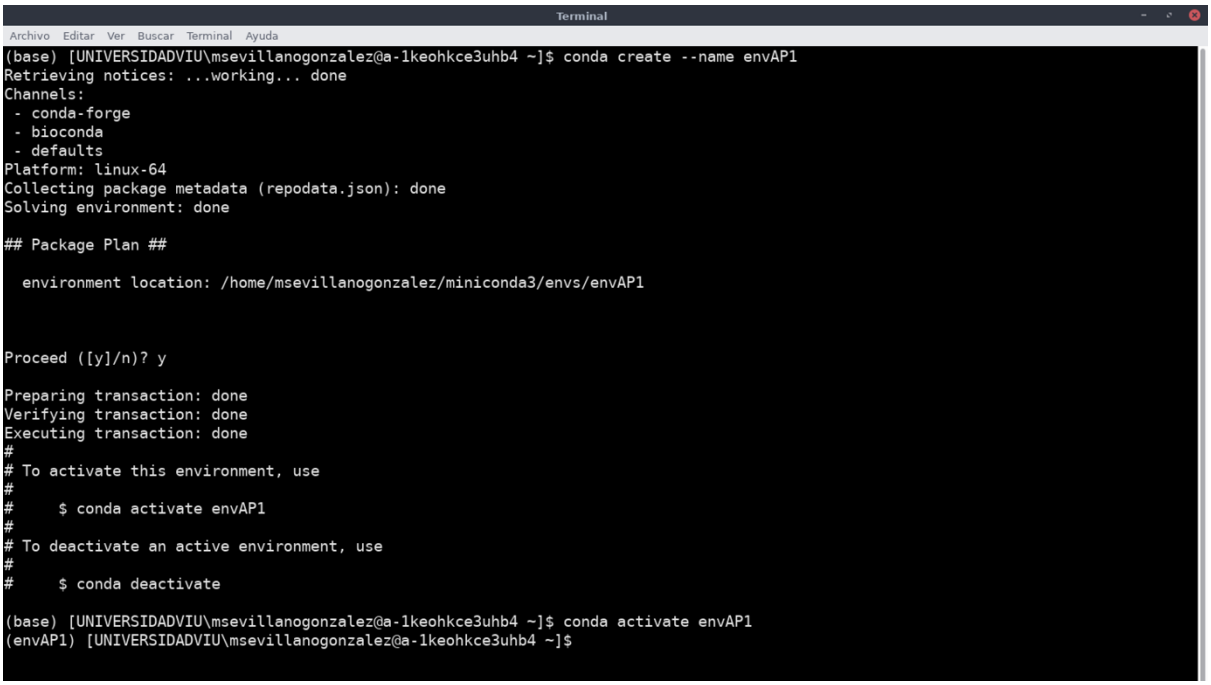
Instrucciones: Contesta en el espacio a la pregunta indicando el/los comando/s solicitados y pega debajo una captura de pantalla para cada una de las respuestas en la que aparezca el terminal con: el prompt, el comando utilizado y el resultado obtenido.

Estás iniciando un nuevo proyecto y necesitas obtener un conjunto de archivos de un trabajo previo. Debes obtener los runs asociados al BioProject con accession number: PRJNA298959

1.3.1 Crea un nuevo environment para tu proyecto, nómbralo “envAP1”.

Para crear un nuevo environment utilizando la herramienta de conda, he usado el comando: conda create –name seguido del nombre del nuevo environment, en este caso llamado envAP1.

Comando: conda create –name envAP1

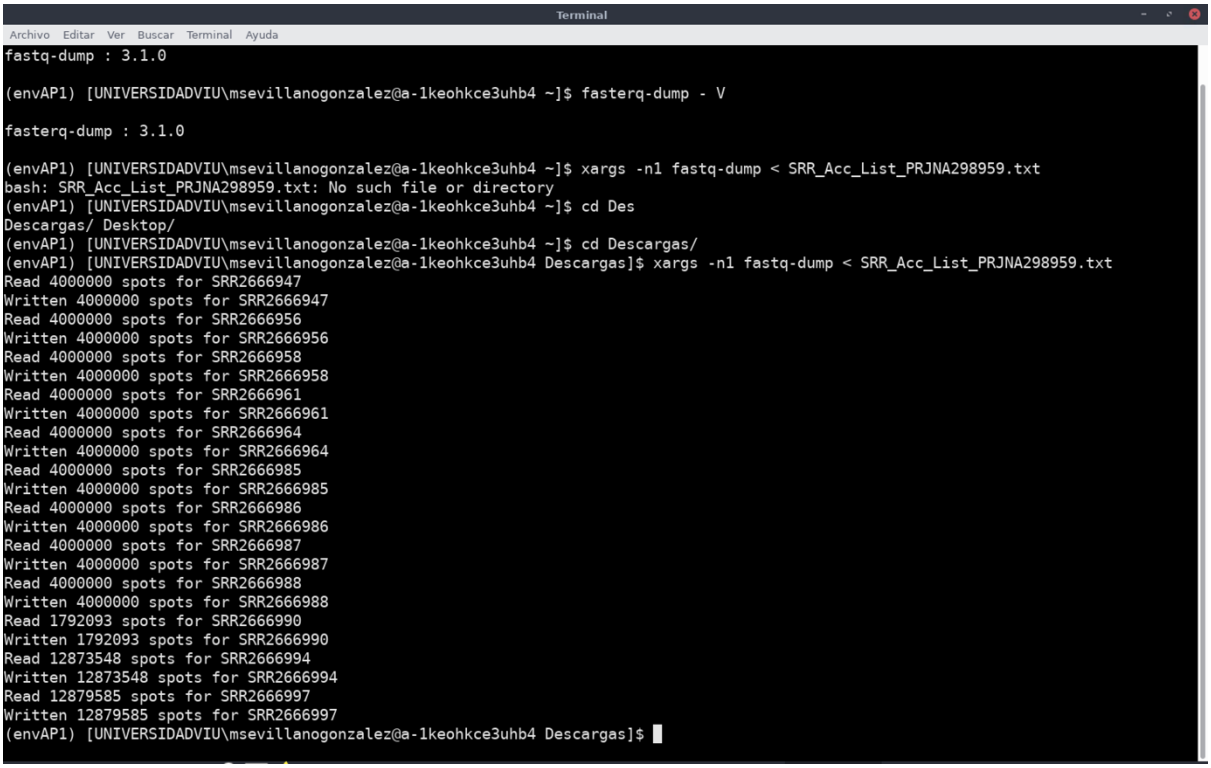


1.3.2 Utiliza la herramienta fastq-dump o fasterq-dump desde este environment y

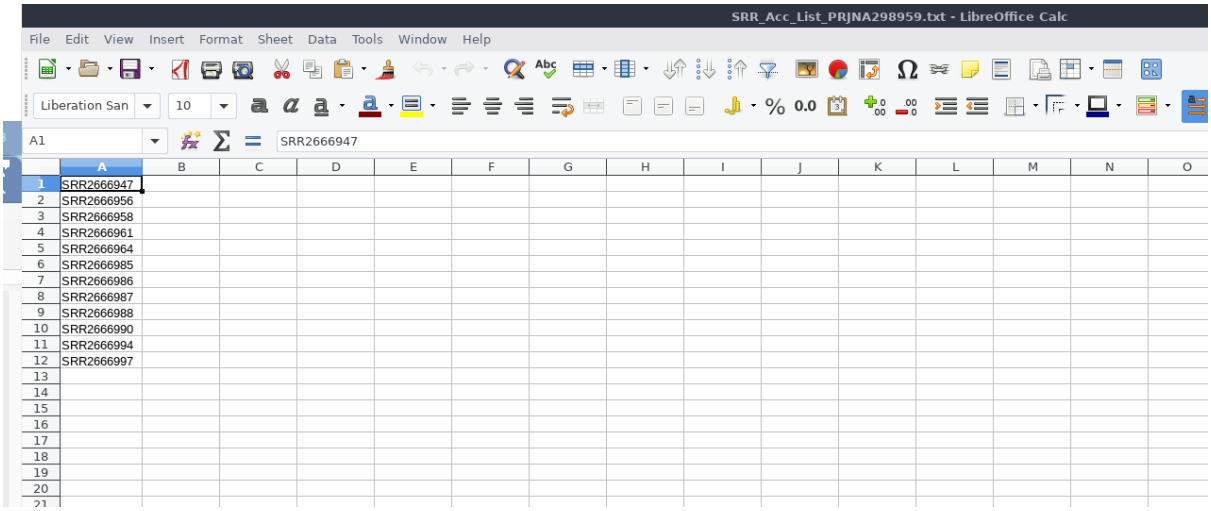
Previamente descargué la lista de los runs asociados al Bioproject PRJNA298959 a través del SRA Run Selector del NCBI, a la cual denominé SRR_Acc_List_PRJNA298959.txt

Utilicé el comando xargs -n1 fastq-dump < seguido del nombre de la lista que contiene los runs de ese Bioproject para poder descargar todos los archivos fastq asociados.

Comando: xargs -n1 fastq-dump < SRR_Acc_List_PRJNA298959.txt



1.3.3 descarga los archivos fastq (all runs) proporcionando un archivo con el listado de los identificadores que necesitas (SRR_Acc_List.txt).



Para comprobar que se había realizado correctamente utilicé el comando less SRR2666997.fastq, para comprobar uno de ellos.