

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).

```
AACAGTCCTAGAAAATGGAGCTGCTTGGCATGTAGTCTTATTGCCCTGCTAAGTTTCATGCTGGG  
GGTCAGACTGGGAGTCTGATAGAAATTCCATTCCACCGCTGGCCTCTAACCAATGACTTGCTGCACAA  
CCTGAGTGGTCTCCTGGGAGACCAGAGTTCTAACTTGTAGCAGGGGACAAAGACATGTATTTGTCAC  
CAGCCACTGCCACTTCCTGCCAGAATACTTCAGCAGTCTCCATGCCAGTCAGATCACCCATTATAAGG  
TATTCTGTCATGGGCACAGCTCCTCCCAGCAGGAAGCACCCAGAATCCAGACGAGAAAACAGTCAGTG
```

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:

The screenshot shows a search result for "The genetic history of Ice Age Europe". A modal window for author "Damien Flas" is displayed, containing their author profile, ORCID ID (0000-0002-9159-8036), affiliation (1. TRACES - UMR 5608, Université Toulouse Jean Jaurès, Maison de la Recherche, 31058 Toulouse Cedex 9, France), and a list of publications. The ORCID ID is highlighted with a red circle.

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)

Jump to: 1000 Genomes Project Phase 3 (32) gnomAD exomes r2.1.1 (9) NCBI ALFA (12) TOPMed (1)				
1000 Genomes Project Phase 3 (32)				
Population	Allele: frequency (count)	Genotype: frequency (count)		Genotypes
ALL	C: 0.9998003194888 T: 0.0001996805111 18 (5007) 82109 (1)	C C: 0.999600638977636 (2503)	C T: 0.0003993610223642 17 (1)	Show
AFR	C: 0.999 (1321) T: 0.001 (1)	C C: 0.998 (660)	C T: 0.002 (1)	Show
ACB	C: 1.000 (192)	C C: 1.000 (96)		Show
ASW	C: 1.000 (122)	C C: 1.000 (61)		Show
ESN	C: 1.000 (198)	C C: 1.000 (99)		Show
GWD	C: 0.996 (225) T: 0.004 (1)	C C: 0.991 (112) C T: 0.009 (1)		Hide
LWK	C: 1.000 (198)	C C: 1.000 (99)		Show
MSL	C: 1.000 (170)	C C: 1.000 (85)		Show
YRI	C: 1.000 (216)	C C: 1.000 (108)		Show
AMR	C: 1.000 (694)	C C: 1.000 (347)		Show
CLM	C: 1.000 (188)	C C: 1.000 (94)		Show
MXL	C: 1.000 (128)	C C: 1.000 (64)		Show
PEL	C: 1.000 (170)	C C: 1.000 (85)		Show
PUR	C: 1.000 (208)	C C: 1.000 (104)		Show
EAS	C: 1.000 (1008)	C C: 1.000 (504)		Show
CDX	C: 1.000 (186)	C C: 1.000 (93)		Show
CHB	C: 1.000 (206)	C C: 1.000 (103)		Show
CHS	C: 1.000 (210)	C C: 1.000 (105)		Show
JPT	C: 1.000 (208)	C C: 1.000 (104)		Show
KHV	C: 1.000 (198)	C C: 1.000 (99)		Show
EUR	C: 1.000 (1006)	C C: 1.000 (503)		Show
CEU	C: 1.000 (198)	C C: 1.000 (99)		Show

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	Genotype (forward strand)	Population(s)	Father Mother
Sample (Male/Female/Unknown)			
HG02568 (F)	C/T	AFR, ALL, GWD	- -
HG02491 (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-GEOID-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 Antônio, Luca Pagani, Alexia Cardona
¹ University of Cambridge

Accession E-GEOID-55175
Study type comparative genomic hybridization by array , genotyping by array 
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 Wichå 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 Wichå living <500 m from the Province of Salta in Argentina was genotyped

Protocols 

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

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

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:

Expression Atlas

Gene expression across species and biological conditions

Home | Browse experiments | Download | Release notes | FAQ | Help | Licence | About | Support | To Single Cell Expression Atlas ▶

Transcription profiling by array of Arabidopsis whole plants and discrete root, hypocotyl and shoot responses to spaceflight

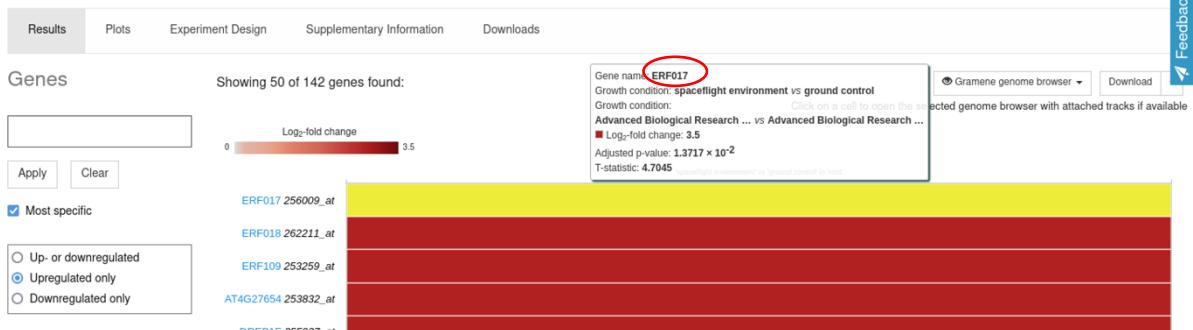
Microarray 1-colour mRNA

Organism: *Arabidopsis thaliana*

Array Design(s): Affymetrix Genechip Arabidopsis Genome [ATH1]

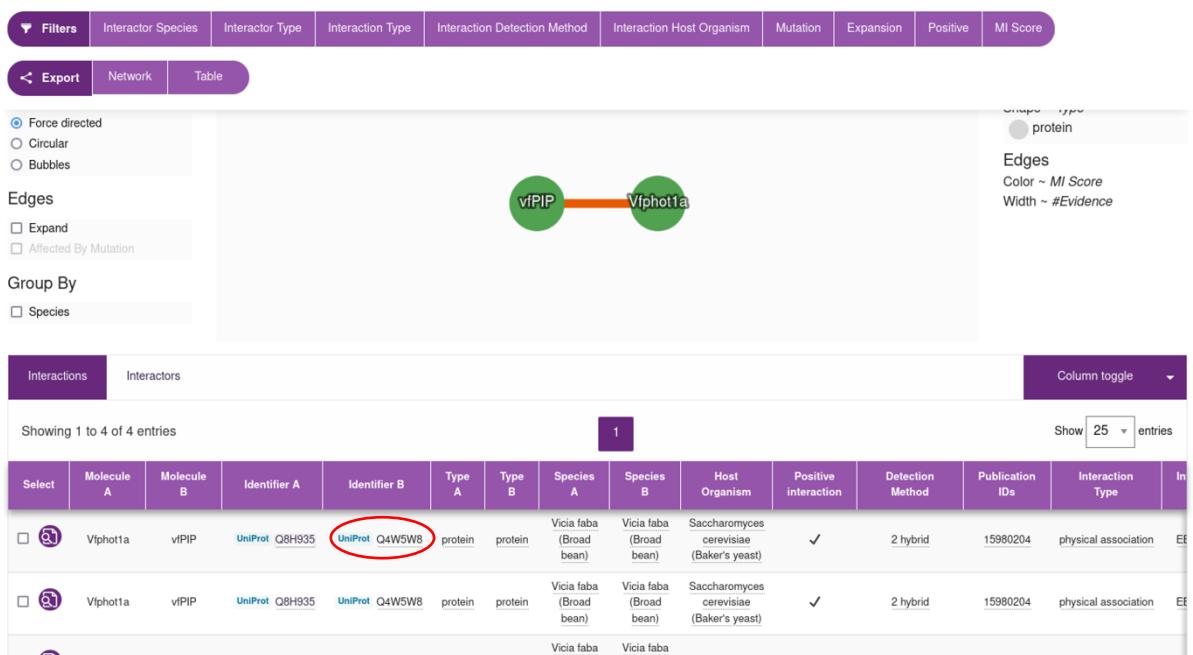
Publication:

• Paul AL, Zupanska AK, Schultz ER, Ferl RJ. (2013) *Organ-specific remodeling of the Arabidopsis transcriptome in response to spaceflight.*



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.

ChEMBL | compounds with bioactivity against Phytophthora infestans

19 Compounds
0 Selected - Select All
Browse Activities

Table Cards Graph Heatmap

CSV TSV SDF

Filters

Records per page: 20

Showing 1-19 out of 19 records

ChEMBL ID	Search Hit	Name	Synonyms	Type	Max Phase	Molecular Weight	Targets	Bioactivities
CHEMBL3991035		SODIUM STIBOGLUCONATE		Small molecule	4	745.74	No Data	No Data
CHEMBL1788392		GLYCOBIARSOL	Bismuth glycolylarsanilate, Glycobiarsol, Milibis, NSC-221709	Small molecule	2	499.06	1 By Type: 	1 By Std. T
CHEMBL1788396		DIFETARSONE	Anhydrous difetarsone sodium, Bemarsal, Difetarsone, Difetarsone sodium, Difetarsone sodium anhydrous, Diphetarsone, N,n'-ethylenediarsonilic acid, RP-4763, RP 4763 [AS SODIUM SALT]	Small molecule	2	460.15	2 By Type: 	19 By Std. T

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:

Complex Portal Examples: GO:0016491Ndc80Q05471PCNACPX-2158nuclear poreQ15554, P54274, Q96AP0

Home About Documentation Basket Support

CLOCK-BMAL1 transcription complex

ComplexAc: CPX-3229

Homo sapiens; 9606

Download Basket Evidence by physical interaction evidence

Participants

Legend	Description	Stoichiometry
	protein - BMAL1 (unspecified role) O00327 ⚡ Basic helix-loop-helix ARNT-like protein 1	1
	protein - CLOCK (unspecified role) O15516 ⚡ Circadian locomoter output cycles protein kaput	1

This website requires cookies, and the limited processing of your personal data in order to function. By using the site you agree, dismiss this banner

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 AOA024B7W1, señalado a continuación en rojo:

Protein Data Bank in Europe

Examples: hemoglobin, BRCA1_HUMAN Advanced search Feedback

PDBE / SEARCH

Text : virus Zika

AND Experimental method : El...

AND Journal : Science

Filter by :

Entry Information

- Entry status (1)
- REL (3)
- + Experimental methods (1)
- + Authors (32)
- + Homo / hetero assembly (2)
- + Assembly composition (2)
- + Assembly polymer count (2)
- + Resolution distribution
- + Release year distribution
- + Journal (1)

Macromolecules

Advanced search

Entries Macromolecules Compounds Protein families

Entries 1 to 3 of 3 Sort by 10 /page ▾

Select all entries on this page

Sire The cryo-EM structure of Zika Virus
Sirohi D, Chen Z, Sun L, Klose T, Pierson T, Rossmann M, Kuhn R
Science (2016) [PMID: 27033547]

Source organism: *Zika virus*

Assembly composition: protein/protein complex

Carbohydrate polymer components:
Molecule 1 - NAGI2

Assembly name: RNA-directed RNA polymerase NS5 (Preferred) [search this complex](#)

PDBe complex ID: PDB-CPX-100068 (Preferred) [search this ID](#)

PDBe-KB: AOA024B7W1

[3D Visualisation](#) [Download files](#)

Electron Microscopy
3.8 Å resolution
Released: 30 Mar 2016
DOI: 10.2210/pdb5ire
Model geometry Fit model/data Data not analysed

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

```

Terminal
Archivo Editar Ver Buscar Terminal Ayuda
(base) [UNIVERSIDADVIU\msevillanogonzalez@a-1keohkce3uhb4 ~]$ conda create --name envAP1
Retrieving notices: ...working... done
Channels:
- conda-forge
- bioconda
- defaults
Platform: linux-64
Collecting package metadata (repodata.json): done
Solving environment: done

## Package Plan ##

environment location: /home/msevillanogonzalez/miniconda3/envs/envAP1

Proceed ([y]/n)? y

Preparing transaction: done
Verifying transaction: done
Executing transaction: done
#
# To activate this environment, use
#
#     $ conda activate envAP1
#
# To deactivate an active environment, use
#
#     $ conda deactivate

(base) [UNIVERSIDADVIU\msevillanogonzalez@a-1keohkce3uhb4 ~]$ conda activate envAP1
(envAP1) [UNIVERSIDADVIU\msevillanogonzalez@a-1keohkce3uhb4 ~]$

```

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

```

Terminal
Archivo Editar Ver Buscar Terminal Ayuda
fastq-dump : 3.1.0
(envAP1) [UNIVERSIDADVIU\msevillanogonzalez@a-1keohkce3uhb4 ~]$ fasterq-dump -V
fasterq-dump : 3.1.0

(envAP1) [UNIVERSIDADVIU\msevillanogonzalez@a-1keohkce3uhb4 ~]$ xargs -n1 fastq-dump < SRR_Acc_List_PRJNA298959.txt
bash: SRR_Acc_List_PRJNA298959.txt: No such file or directory
(envAP1) [UNIVERSIDADVIU\msevillanogonzalez@a-1keohkce3uhb4 ~]$ cd Descargas/
Descargas/ Desktop/
(envAP1) [UNIVERSIDADVIU\msevillanogonzalez@a-1keohkce3uhb4 ~]$ cd Descargas/
(envAP1) [UNIVERSIDADVIU\msevillanogonzalez@a-1keohkce3uhb4 Descargas]$ xargs -n1 fastq-dump < SRR_Acc_List_PRJNA298959.txt
Read 4000000 spots for SRR2666947
Written 4000000 spots for SRR2666947
Read 4000000 spots for SRR2666956
Written 4000000 spots for SRR2666956
Read 4000000 spots for SRR2666958
Written 4000000 spots for SRR2666958
Read 4000000 spots for SRR2666961
Written 4000000 spots for SRR2666961
Read 4000000 spots for SRR2666964
Written 4000000 spots for SRR2666964
Read 4000000 spots for SRR2666985
Written 4000000 spots for SRR2666985
Read 4000000 spots for SRR2666986
Written 4000000 spots for SRR2666986
Read 4000000 spots for SRR2666987
Written 4000000 spots for SRR2666987
Read 4000000 spots for SRR2666988
Written 4000000 spots for SRR2666988
Read 1792093 spots for SRR2666990
Written 1792093 spots for SRR2666990
Read 12873548 spots for SRR2666994
Written 12873548 spots for SRR2666994
Read 12879585 spots for SRR2666997
Written 12879585 spots for SRR2666997
(envAP1) [UNIVERSIDADVIU\msevillanogonzalez@a-1keohkce3uhb4 Descargas]$ 

```

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

SRR_Acc_List_PRJNA298959.txt - LibreOffice Calc																
File Edit View Insert Format Sheet Data Tools Window Help																
Liberation Sans 10 Aa																
A1 SRR2666947																
1	SRR2666947															
2	SRR2666956															
3	SRR2666958															
4	SRR2666961															
5	SRR2666964															
6	SRR2666985															
7	SRR2666986															
8	SRR2666987															
9	SRR2666988															
10	SRR2666990															
11	SRR2666994															
12	SRR2666997															
13																
14																
15																
16																
17																
18																
19																
20																
21																

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

```
Archivo Editar Ver Buscar Terminal Ayuda
@SRR2666997.1 1 length=60
AAACAAACAGGTGGTGTGATCTACTGACGACATGGTTCTACATACCAACATCTGGT
+SRR2666997.1 1 length=60
>>1>AA-11B1B1FAF1A3FAFDFFDF1BB?CE0AFCHD1A221121A11//011D2BB
@SRR2666997.2 2 length=60
CGATGTGATGCCACGAGGTCTCTGACGGCAAAGAACGTCAGCTGCAGGTGCG
+SRR2666997.2 2 length=60
>A1A1FB3DFFF3FE111EBFEGEBF11A000000000BA0//A/B0///F/A010BB/
@SRR2666997.3 3 length=60
CCGTCGATGCCACGAGGTCTCTGACAGTTAAACCTGACGTACGCTGCAGGTGCG
+SRR2666997.3 3 length=60
1>1AAAD1AAFFG3AE000EEFFG3121101DDFF011AABA//A/A/A/AF/A010BB/
@SRR2666997.4 4 length=60
CTTGACGGTGTGGCTCTGAGTGTGACGATGTTACCATATTGTTGATTCTAGC
+SRR2666997.4 4 length=60
1AA7A3B1AFAAFA11EEAAF0A0320000/A/B1DB22A1D12BA00B0/2DB22221
@SRR2666997.5 5 length=60
CTTGACGGTGTGGCTCTGAGTGTGCTAGACACTCGCGCTATTGATGAATTCTAGC
+SRR2666997.5 5 length=60
1AA7A3B1>CAAF1EEEED0F00BA111BF111A1AA0///A/AD12D221DDF222BD
@SRR2666997.6 6 length=60
GCCAATGATGCCACGAGGTCTATGAGCCAGGTACCTTGTACGCTGCTGGTCGAC
+SRR2666997.6 6 length=60
1>>A1FB1FFFFFGCE111EBFFG3D3311A00011B1DAB111B21//B/111B1//
@SRR2666997.7 7 length=60
CTTGACGGTGTGGCTCTGAGTGTGACGATGTTGACGCTGCAGGTGCG
+SRR2666997.7 7 length=60
AAAAAABBBBBFGFGE111EEFFG0B0001A11AA1121B1111/B/A/AF/A010BB/
@SRR2666997.8 8 length=60
CAGATCCGGTGTGGCTCTGAGCTGTTAAGCAAGAGGTTGAATCGATGAATTGATCT
+SRR2666997.8 8 length=60
>>11>CFF1DDAFFE?FFGD0A000AB0BB22110000/1B002DA00B211DDE/22BB
@SRR2666997.9 9 length=60
CCGTCGATGCCACGAGGTCTCTGGCGCTCAAGATTTATGACGTACGCTGCAGGTGCG
+SRR2666997.9 9 length=60
>A1AAAD?AAFFGDGG000EFFFD0000AA//111B1DD2D221/B/A/AA/A010BB/
:
```