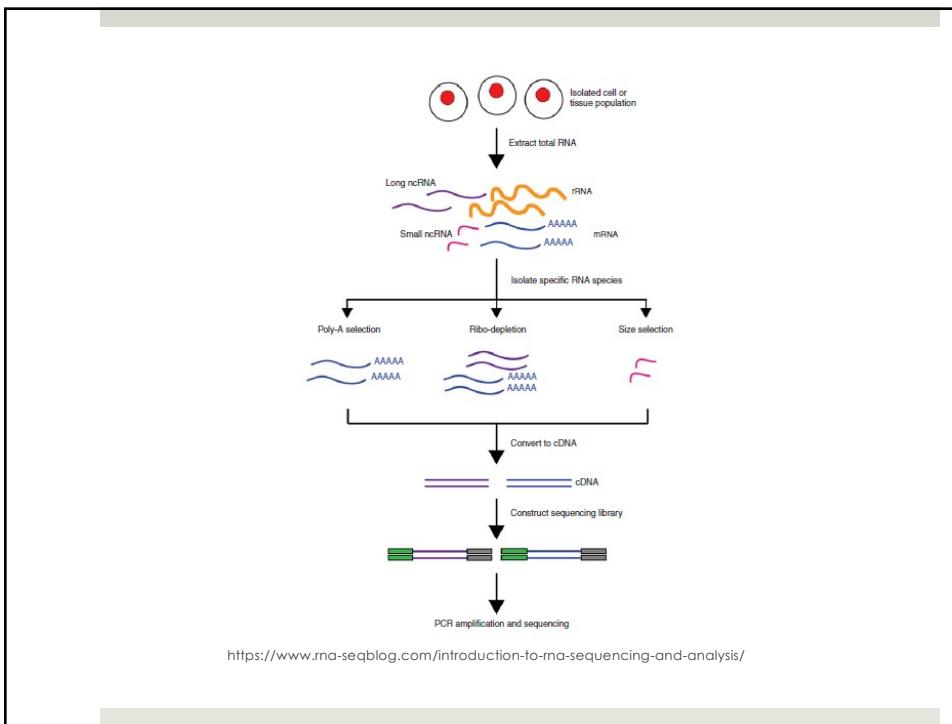




1



2

¿Qué utilidad tiene?

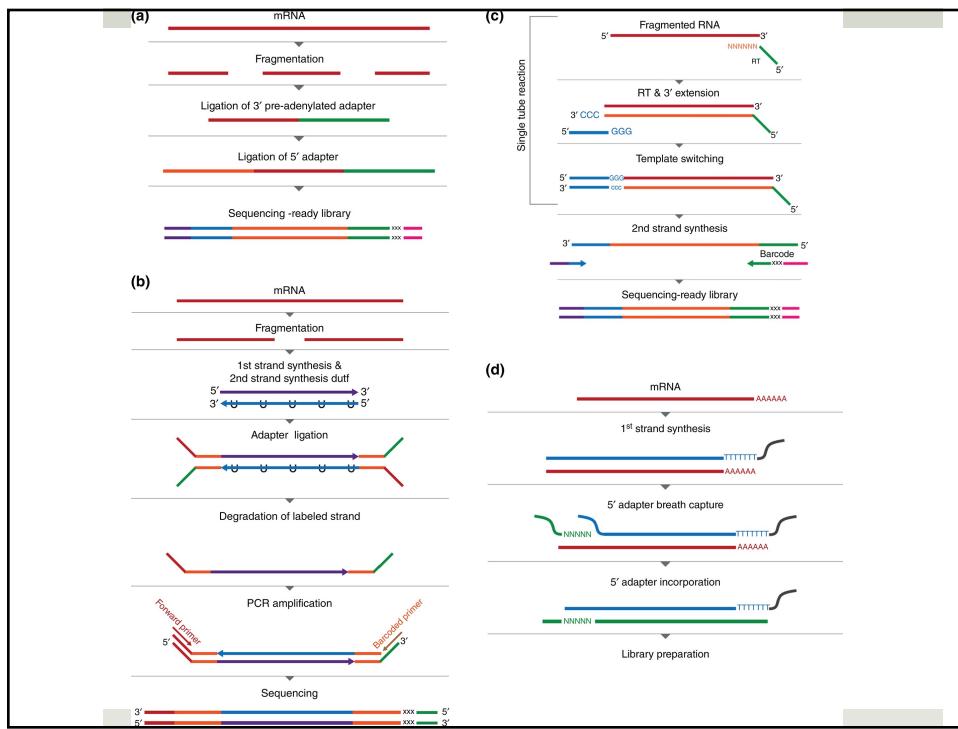
- Medir cuantitativamente expresión génica.
- Construir un mapa transcriptómico.
- Descubrir nuevos genes.
- Identificar sitios de *splicing*.

3

Consideraciones generales

- Generalmente se secuencia cDNA
- La preparación de librerías dependen del tipo de RNA, y varían en cuanto a la forma de:
 - Capturar las moléculas de RNA interés
 - Convertir RNA a cDNA
 - Agregar los adaptadores
- Detalles en WIREs RNA 2017, 8:e1364. doi: 10.1002/wrna.1364

4



5

Metodos de secuenciación

- DGE-seq (Tag-Seq):
 - Secuenciación profunda
 - Digestión con enzima de restricción
 - Solo una región del transcripto (~21pb)
 - Económico, nivel expresión del gen
- Secuenciación de extremo 3'
 - Usan oligo(dT) como primer de RT
 - Económico, nivel expresión del gen

- Alta variabilidad
- Sesgo
- Efectos de batch

6

Métodos para discriminación de transcritos

- Detección de alternativas de splicing (isoformas)
 - Uniones exón-exón (ej. RASL-Seq)
 - Comparación estadística de lecturas dentro de exones
- Detección de fusiones de genes (ej. En cáncer)
 - Enriquecimiento para genes de interés
 - Captura de exones
 - Single-molecule real-time sequencing (SMRT) on the PacBio
 - SLR-RNA-Seq en Illumina MOLECULO system

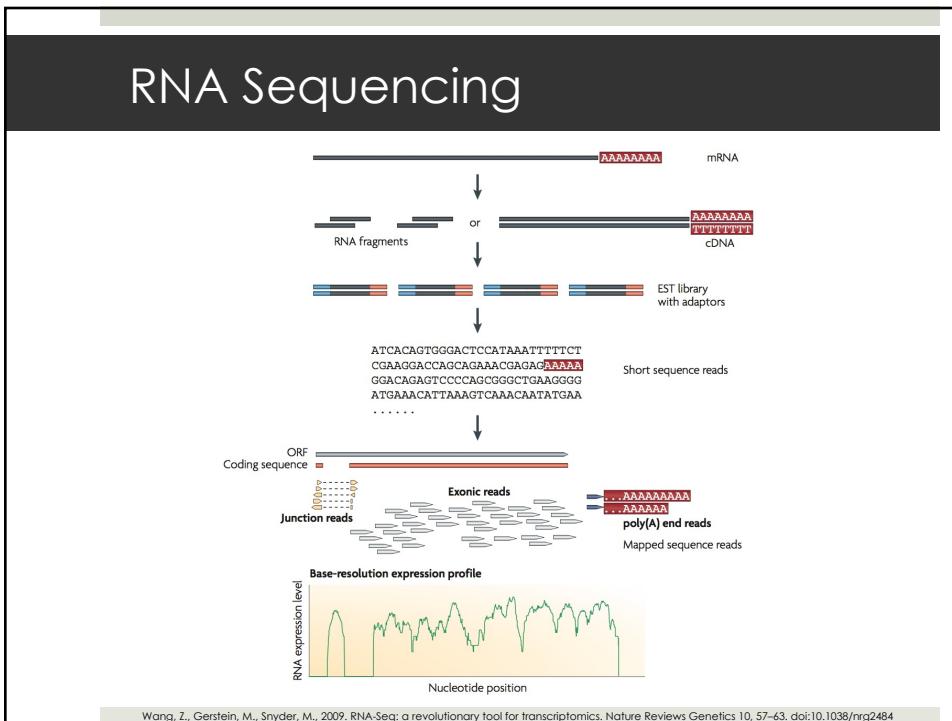
7

Distintos métodos para interrogar cada aspecto del ciclo de vida del RNA

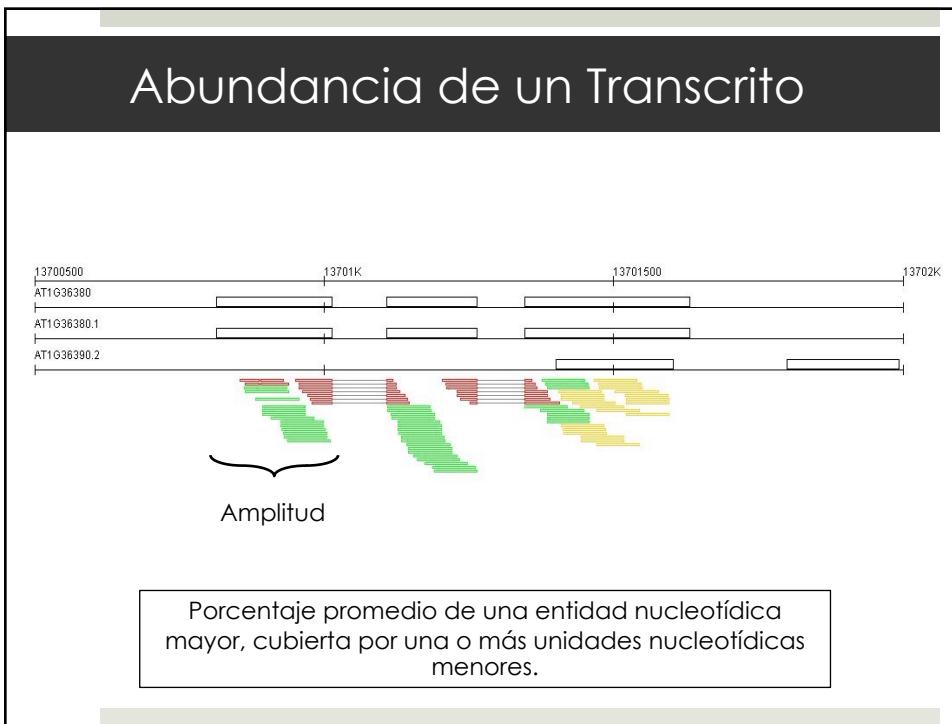
Purpose	Methods	RNA Species Sequenced
Transcribing RNA polymerase	GRO-Seq ¹⁰² NET-Seq ¹⁰³⁻¹⁰⁵	Nascent RNA transcribed <i>in vitro</i> Nascent RNA associated with RNA polymerase <i>in vivo</i>
RNA synthesis, processing and degradation	4sU-Seq ¹⁰⁶⁻¹⁰⁸	Newly synthesized RNA labeled with 4-thiouridine (4sU) <i>in vivo</i>
Translational status	Ribosome profiling (Ribo-seq) ⁹⁹	Ribosome-bound mRNA fragments protected from RNase digestion
RNA–protein interactions	RIP-Seq ¹⁰⁹⁻¹¹¹ HIT-CLIP/CLIP-seq, ¹¹²⁻¹¹⁴ iCLIP ¹¹⁵ PAR-CLIP ¹¹⁶	RNA in immunoprecipitated ribonucleoprotein complex RNA fragments crosslinked to interacting proteins by UV (254 nm) RNA labeled with 4sU and crosslinked to interacting proteins by UV (365 nm)
RNA structure probing	PARS, ¹¹⁷ FragSeq ¹¹⁸	RNA fragments generated by RNases digesting double-stranded regions (V1) or single-stranded regions (S1 or P1)
	DMS-Seq ¹¹⁹ , SHAPE-S ¹²⁰ icSHAPE ¹²¹	Unstructured RNA regions labeled with reactive chemicals
RNA–RNA interactions	RAP-RNA ¹²²	RNA isolated by antisense probe

Hrdlickova, R., Toloue, M., Tian, B., 2017. RNA-Seq methods for transcriptome analysis: RNA-Seq. Wiley Interdisciplinary Reviews: RNA. doi:10.1002/wrna.1364

8

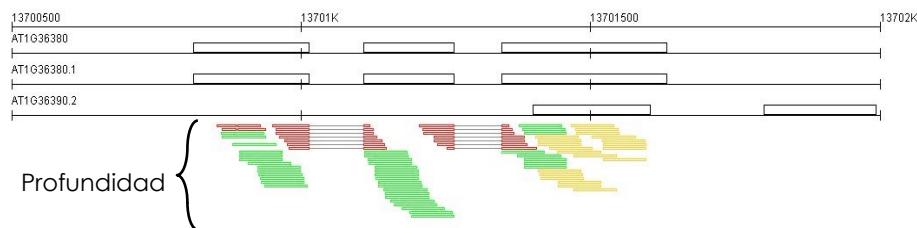


9



10

Abundancia de un Transcrito

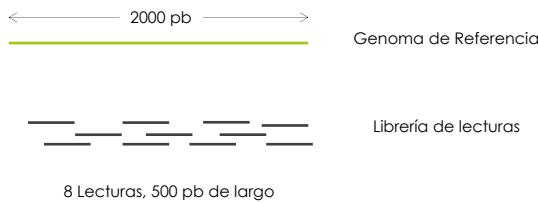


Número de veces que una entidad nucleotídica determinada, ha sido leída mediante un proceso de secuenciación.

11

Profundidad de Cobertura

■ Ejemplo



Profundidad = ?

12

Profundidad de Cobertura

- G = Largo completo del genoma.
- N = Número de lecturas.
- L = Largo de lecturas.

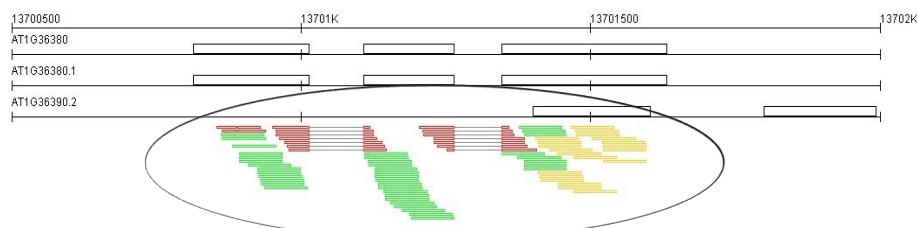
$$D = \frac{N * L}{G}$$



$$D = \frac{8 * 500}{2000} = 2X$$

13

Abundancia de un Transcrito



Número de lecturas dependen del tamaño

14

Normalización de Abundancia

RPKM

Reads per Kilobase of exon model per Million of reads mapped.

$$RPKM = \frac{10^9 \cdot C}{N \cdot L}$$

■ Donde:

- C = Número de lecturas mapeadas por gen.
- N = Total de lecturas.
- L = Largo del exón.

15

Expresión Diferencial

- Secuenciar un organismo bajo 2 o más condiciones experimentales diferentes.
 - Estimulado vs No-estimulado.
 - Wild-type vs Mutante.
- Comparar niveles de expresión en genes de interés.
- Pruebas estadísticas:
 - T-test.
 - F-test.
 - **Exact Test.**

16

Exact Test

- Sea Y_{ij} la abundancia observada para el gen 'j' en la librería 'i'

$$Y_{ij} \sim NB(\mu_i, \phi)$$

- Distribución **Binomial Negativa** modela sucesos ocurridos en un intervalo de tiempo ó espacio.

$$E(Y_{ij}) = \mu_{ij} = m_i \lambda_{ij}$$

$$\text{var}(Y_{ij}) = \mu_{ij} + \phi_g \mu_{ij}^2$$

17

Exact Test

$$H_0 : \mu_{1j} - \mu_{2j} = 0$$

- Definir si diferencias de conteos entre librerías son significativas.

- Dada una diferencia entre promedios observada:

¿Cuál es la probabilidad de encontramos una diferencia igual o mayor asumiendo H_0 ?



Valor P

18

Exact Test

- ❑ Dado un valor de α previamente definido
 - ❑ 0.05
 - ❑ 0.1
 - ❑ 0.2
- ❑ Rechazar H_0 si:
 - ❑ P-valor $< \alpha$

!Gen J está diferencialmente expresado!

19

Práctico

- ❑ *Sulfolobus acidocaldarius*
 - ❑ Arqueobacteria termo-acidófila.
 - ❑ Habita en medios por sobre 80° C y con PH ~ 2.
- ❑ Lrs14-like: Factor de transcripción con rol clave en la formación de biopelícula.
 - ❑ Medio preferencial de crecimiento de este organismo.
- ❑ ¿Qué otros genes se ven afectados por la expresión de Lrs14-like?
- ❑ Si se cambia el medio de crecimiento ¿Cómo afecta en la expresión génica?

20

Hipótesis

1. *Lrs14-like*, regula la expresión de otros genes en el proceso de formación de biopelículas.
2. Realizar un cultivo de esta bacteria en un medio diferente a biopelícula, afecta la expresión génica.

21

Experimento

■ Secuenciación de 4 muestras

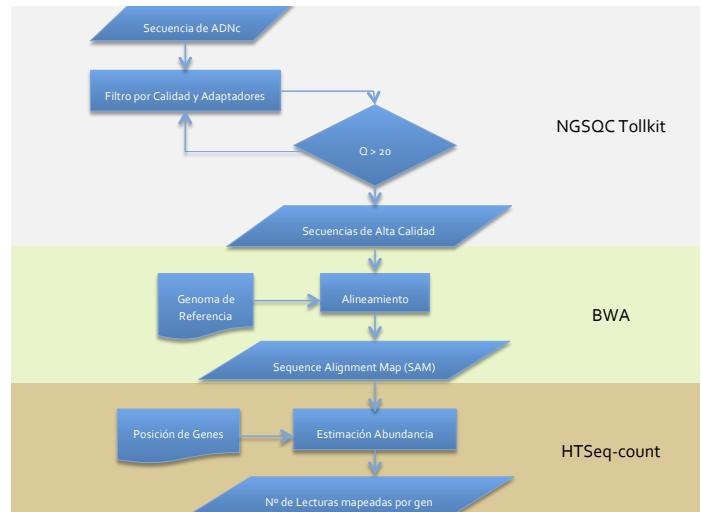
Wild Type, Plantónico	Wild Type, Biopelícula
Mutante, Plantónico	Mutante, Biopelícula

■ Plataforma Illumina HiSeq 2500

Muestra	Tamaño	Largo Lecturas
Wild Type, Plantónico	~ 22.4 M	94 pb
Wild Type, Biopelícula	~ 24.9 M	
Mutante, Plantónico	~ 22.4 M	
Mutante, Biopelícula	~ 25.3 M	

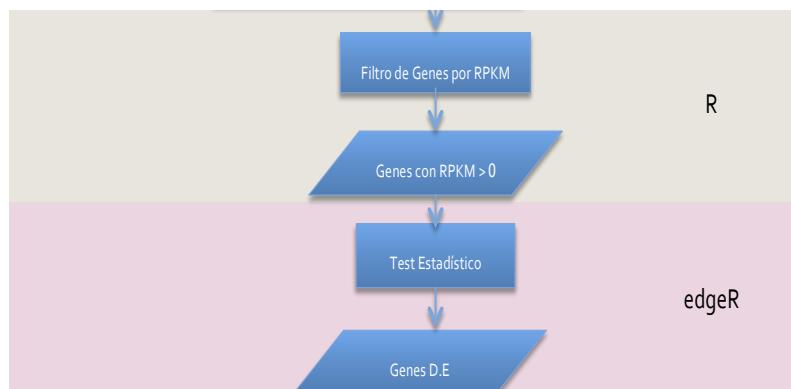
22

Flujo de Trabajo



23

Flujo de Trabajo



24

Test para Medio de cultivo

$$y_{ij} = \mu + \text{Medio}_i + e_{ij}$$

donde,

- | | |
|------------------|--|
| y_{ij} | : Expresión de un gen |
| μ | : expresión promedio en medio planctónico |
| Medio_i | : Efecto (diferencia) de medio biopelícula |
| e_{ij} | : efecto residual |

25

Test para Genotipo de Lrs14-like

$$y_{ij} = \mu + \text{Genotipo}_j + e_{ij}$$

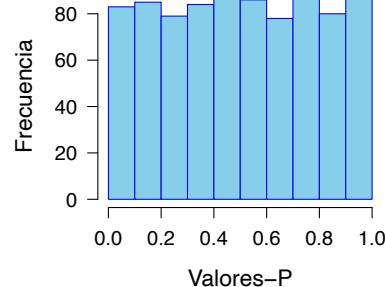
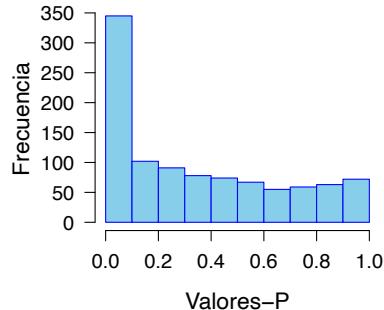
donde,

- | | |
|---------------------|--|
| y_{ij} | : Expresión de un gen |
| μ | : expresión promedio en genotipo wildtype |
| Genotipo_j | : Efecto (diferencia) del genotipo mutante
(Knock Down) para Lrs14-like |
| e_{ij} | : efecto residual |

26

Resultados

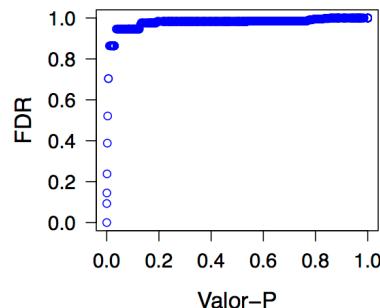
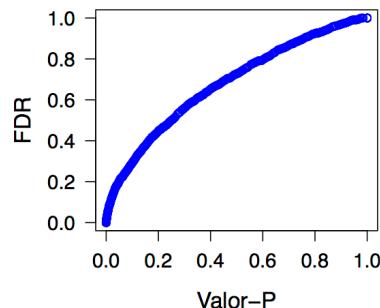
■ Histograma de Valores P



27

Resultados

■ Valor P vs FDR

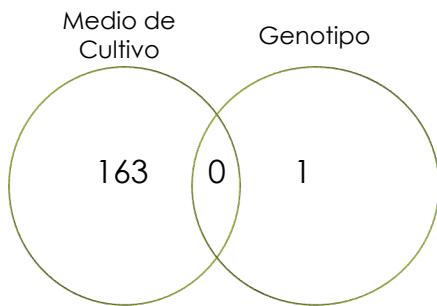


FDR: *False discovery rate* (tasa de falsos positivos). Control del error cuando se realizan muchos tests (uno por gen). Se seleccionan genes con $FDR < 0.1$

28

Resultados

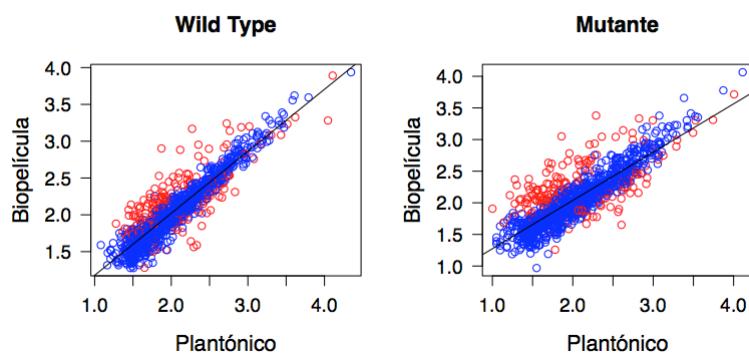
■ Número de genes diferencialmente expresados (FDR < 0.1)



29

Resultados

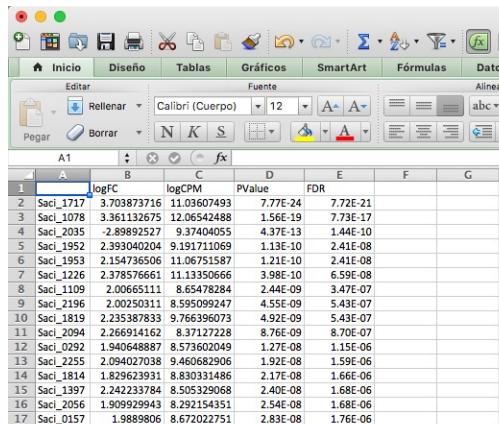
■ Scatterplot de Valores de Expresión



30

Resultados

- Tabla de genes diferencialmente expresados.



A	B	C	D	E	F	G
1	logFC	logCPM	PValue	FDR		
2	Saci_1717	3.703873716	11.03607493	7.77E-24	7.72E-21	
3	Saci_1078	3.361132675	12.065424488	1.56E-19	7.73E-17	
4	Saci_2035	-2.898925277	9.37404055	4.37E-13	1.44E-10	
5	Saci_1952	2.393040204	9.191711069	1.13E-10	2.41E-08	
6	Saci_1953	2.154736506	11.06751587	1.21E-10	2.41E-08	
7	Saci_1226	2.378576661	11.13350666	3.98E-10	6.59E-08	
8	Saci_1109	2.00665111	8.65478284	2.44E-09	3.47E-07	
9	Saci_219	2.00250311	8.595099247	4.55E-09	5.43E-07	
10	Saci_1819	2.235387833	9.766396073	4.92E-09	5.43E-07	
11	Saci_2059	2.266914162	8.37127228	8.76E-09	8.70E-07	
12	Saci_0292	1.940648887	8.573602049	1.27E-08	1.15E-06	
13	Saci_2255	2.094027038	9.460682906	1.92E-08	1.59E-06	
14	Saci_1814	1.829623931	8.830331486	2.17E-08	1.66E-06	
15	Saci_1397	2.242233784	8.505329068	2.40E-08	1.68E-06	
16	Saci_2056	1.909929943	8.292154351	2.54E-08	1.68E-06	
17	Saci_0157	1.9889806	8.672022751	2.83E-08	1.76E-06	

31

Conclusiones

- Se observa expresión diferencial entre 163 genes de *S. acidocaldarius* cuando la archeobacteria crece en diferentes medios de cultivo.
- Se observa solo un gen (Saci_2195) diferencialmente expresado entre los genotipos silvestres y mutantes para el gen Lrs14-like. Este gen está anotado como "Hypothetical protein".
- No hay suficiente evidencia para determinar si la delección de Lrs14-like genera efectos en la regulación génica de otros genes.
- Se sugiere realizar un nuevo experimento con réplicas experimentales.

32