

ANÁLISIS DE EXPRESIÓN DIFERENCIAL EN PELOS RADICULARES DEL ARROZ: COMPARACIÓN ENTRE DESEQ2 Y EXPRESIÓN ATLAS DEL EXPERIMENTO E-CURD-28.

Andrea Arredondo Restrepo.

Los pelos radiculares son extensiones tubulares de las células epidérmicas que se encuentran en las raíces de las plantas y aumentan su área superficial, lo cual facilita la absorción de agua y nutrientes. Se desarrollan en tres etapas básicas las cuales son, especificación del destino de las células ciliadas de la raíz, inicio de un crecimiento del pelo y el alargamiento de este a través del crecimiento en la punta. Se encuentran en casi todas las plantas vasculares, incluidas las angiospermas, gimnospermas y licófitas, las cuales muestran características celulares similares, lo que indica un origen evolutivo en común.

Por su parte, el programa genético molecular para el desarrollo de los pelos radiculares se ha estudiado ampliamente en la planta *Arabidopsis thaliana*, identificando mecanismos genéticos bien establecidos y caracterizados; sin embargo, el conocimiento de como funcionan estos genes en plantas monocotiledóneas como el arroz (*Oryza sativa*) es escaso, por lo que, los autores buscaron en el artículo identificar, comparar y caracterizar los genes encargados del desarrollo de los pelos radiculares en diferentes especies vegetales.

Para ello, él estudió E-CURD-28 realizado por Huang et al. (2017), tuvo como objetivo caracterizar el transcriptoma de las células ciliares del arroz, a través del análisis de expresión génica diferencial por RNA-Seq. utilizando la línea transgénica OsEXPA30::GFP, y comparando su perfil genético con el de *Arabidopsis thaliana*. Identificando genes ortólogos implicados en el desarrollo de los pelos radiculares, y determinando que tan conservados se encuentran. Además, pudieron detectar genes específicos en el arroz que podrían desempeñar funciones únicas en la morfogénesis de los pelos radiculares.

Metodología.

Para determinar los genes del desarrollo de pelos radiculares en *Arabidopsis thaliana* y en arroz (*Oryza sativa*), se realizó la siguiente metodología.

Material vegetal y aislamiento del ARN.

Las semillas de las líneas transgénicas para *Arabidopsis thaliana* y *Oryza sativa*, son respectivamente COBL9::GFP y OsEXPA30::GFP, las cuales se cultivaron en medio sólido y bajo luz continúa. Luego, las puntas del crecimiento radicular se recolectaron y se prepararon protoplastos (es decir, se eliminó la pared celular de las células vegetales). Después, las células señalizadas por GFP se aislaron mediante citometría de flujo, y el ARN total se extrajo de las muestras.

En las muestras de control (sin marcador de GFP), los pelos radiculares se aislaron mediante inmersión y agitación en nitrógeno líquido, con un filtraron posteriormente. Las bibliotecas del ADNc se prepararon con el kit de Illumina TruSeq y se realizó la secuenciación con Illumina HiSeq 2000.

Procesamiento de las lecturas y cuantificación de la expresión génica.

Se recordaron las primeras 15 pb de las lecturas con 50pb, y se alinearon mediante un genoma de referencia usando TopHat (los genomas de referencia y las anotaciones para Arabidopsis y arroz se descargaron de la base de datos Ensembl Plant) usando como parámetro una longitud del fragmento de 17. Luego, se calculó la expresión génica mediante Cufflinks2 con corrección multilectura (-u -G).

Análisis de expresión diferencial.

El número de los recuentos crudos mapeados a cada gen se cuantificó mediante HTSeq y se con edgeR se hizo el análisis diferencial. Para ello, primero se filtraron los genes con expresión inferior a 1 en más de tres de seis muestras (límite inferior), luego los datos crudos se normalizaron con el método Trimmed Mean Of M-values, y la dispersión se estimó con el modelo tagwise de edgeR.

Los valores P se calcularon y se consideraron genes diferencialmente expresados aquellos con un cambio de pliegue (fold change) ≥ 2 ($|\log_2FC| \geq 1$) y un valor q de FDR ≤ 0.001 .

Análisis filogenético y comparativo.

Se realizó un árbol filogenético obteniendo las familias génicas de la base de datos GreenPhyl. Las secuencias de las proteínas se alinearon con MAFT y el árbol se construyó con FastTree.

Diseño experimental.

Tabla 1. *Resumen del diseño experimental.*

Variable biológica.	Niveles experimentales.	Tipo de muestra.	Número de réplicas.	Tipo de datos generado.	Contraste analizado.
Genotipo de la planta.	1. <i>Wildtype</i> (control). 2. <i>OsEXPA30::GFP</i> (línea transgénica).	Células ciliadas de los pelos radiculares.	3 por cada nivel.	RNA-Seq (lecturas Illumina HiSeq 2000).	<i>OsEXPA30::GFP</i> vs <i>Wildtype</i> .

Resultados.

Figura 1. Gráfica del volcano plot: *EXPA30::GFP* vs *Wildtype*.

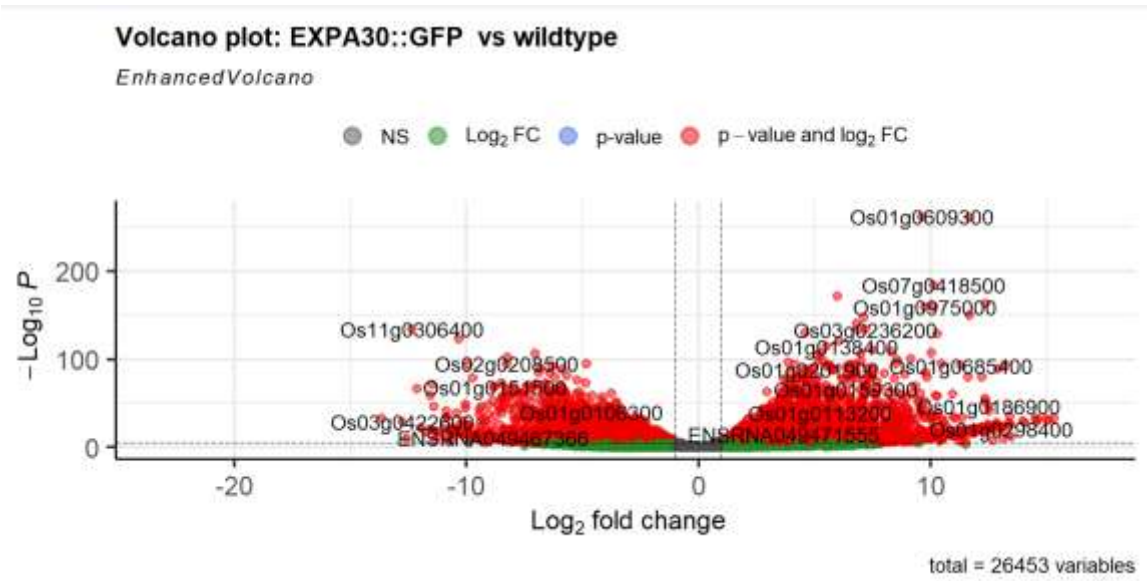


Tabla 2. Resumen de genes.

	Gene.ID	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj	Gene.Name
1764	Os01g0609300	92294.553	9.624148	0.2560708	37.58393	0.000000e+00	0.000000e+00	OsPDR9
20164	Os08g0482600	16300.409	11.687947	0.3382664	34.55249	1.307816e-261	1.512358e-257	OsUCL29
17672	Os07g0418500	7591.619	10.127065	0.3490218	29.01557	4.186303e-185	3.227361e-181	OsCYP709C9
6667	Os02g0806600	9601.034	6.002445	0.2143513	28.00284	1.500426e-172	8.675461e-169	
19101	Os08g0137900	20111.031	12.368433	0.4530037	27.30316	3.890702e-164	1.799683e-160	OsUCL25
11443	Os04g0497000	4923.917	10.114289	0.3744484	27.01117	1.092541e-160	4.211380e-157	OsNADPH1
3802	Os01g0975000	3342.553	9.762550	0.3623207	26.94450	6.617352e-160	2.186373e-156	OsDSR3
5551	Os02g0605900	9501.985	11.669635	0.4468278	26.11663	2.360403e-150	6.823925e-147	Cht6
24003	Os11g0213700	2271.280	7.114997	0.2747603	25.89529	7.525873e-148	1.933982e-144	
23423	Os10g0568900	3501.806	6.795410	0.2682255	25.33469	1.325545e-141	3.065721e-138	

Tabla 3.

Total de genes	Genes significativos (padj<0.05)	Genes altamente significativos (padj<1e-5)	Genes sobreexpresados en EXPA30GFP	Genes sobreexpresados en Wildtype
26453,00	14404,00	7077,00	7540,00	6864,00

Análisis de resultados.

El análisis de expresión diferencial que se realizó con DESeq2, comparando la línea OsEXPA30::GFP y el wildtype de la planta de arroz, permitió identificar 26453 genes analizados en total como se puede observar en la Tabla 3. De estos, 14404 son significativos en su expresión (padj<0.05), entre ellos 7540 genes tuvieron una sobreexpresión en la línea transgénica (presentaron un $\log_2\text{FoldChange} > 0$) y 6864 tuvieron un $\log_2\text{FoldChange} < 0$ lo que indica sobreexpresión en el wildtype.

La sobreexpresión de un gen en análisis diferenciales hace referencia a un aumento significativo en la cantidad de ARNm de un gen en comparación con su control, lo cual indica que, en este caso, el gen *EXPA30* afecta las redes de regulación, transporte y metabolismo en las células de los pelos radiculares. Lo cual es un efecto esperado ya que este gen es una expansina (NIH, 2024) que le confiere a la pared celular características diferentes a las normales, como aumentar la tasa de crecimiento a través del aflojamiento de la pared celular, lo cual modifica su transcriptoma completo, alterando la estructura celular (Jiménez, 2023). Por lo que, es coherente que más de la mitad del transcriptoma analizado presente cambios significativos hacia la línea modificada.

Por su parte, el volcano plot representa cada gen con dos coordenadas, en el eje x se muestra que tanto cambia la expresión entre las dos muestras y el y que tan significativa es esa diferencia. La Figura 1 evidencia genes cerca del centro, lo que significa que no cambian entre ambas líneas analizadas, y dos grupos grandes hacia los extremos, que indican la sobreexpresión de los genes hacia ambos lados. Los que se encuentran a la derecha muestran gran sobreexpresión en EXP30::GFP y hacia la izquierda son genes reprimidos en esta línea y sobre expresados en el control. A su vez, los genes más apartados o muy alejados del resto como el *Os01g0609300* tienen valores de $\log_2\text{FD}$ muy altos y valores p extremadamente bajos (como se ve en la Tabla 2); indicando que cambian mucho entre las condiciones y representan los efectos más fuertes del gen *EXPA30* sobre las células ciliadas.

De esta manera, la Tabla 2 representa los 10 genes más significativos, en los cuales el $\log_2\text{FD}$ tiene un valor muy elevado y el padj es muy pequeño, lo que indica que son genes involucrados activamente en los pelos radiculares y pertenecen a la familia de la fitocianina en el arroz, los cuales están relacionados con procesos de crecimiento, transporte y

señalización celular (Ma et al., 2011). Lo que confirma que el cambio en la línea transgénica representa cambios profundos y la activación de genes específicos que permiten modificar la pared celular e incrementan la capacidad de elongación de los pelos radiculares.

Por último, al hacer una comparación de los resultados obtenidos con el análisis de DESeq2 realizado, y lo mencionado en el artículo y los resultados publicados en Atlas, se evidencia que en este utilizaron edgeR para la expresión diferencial y trabajaron con parámetros de $\log_2FD \geq 2$ y $unFDR \leq 0.001$ que corresponde a una corrección estadística que controla el porcentaje esperado de falsos positivos entre los genes significativos (Columbia university Irving medical center, s.f.). Sin embargo, en ambos casos se utilizan las mismas muestras, por lo que posible realizar un análisis comparativo entre ambos métodos.

Por su parte, el artículo confirma que el transcriptoma de los pelos radiculares contiene grandes grupos de genes diferencialmente expresados, con un perfil transcriptómico altamente especializado entre especies, lo que coincide con los resultados y su análisis en el método DESeq2. A su vez, se sugiere que las expansinas podrían haber evolucionado para su uso en el desarrollo de pelos radiculares, y generan un cambio en las rutas, activando genes transportadores como los arrojados por DESeq2 (*OsPDR9*, *OsUCL29*, etc.), concordando los resultados de la Tabla 2, que son genes que se esperan que estén involucrados, con las funciones mencionadas por los autores del estudio.

También, al comparar los resultados obtenidos del análisis DESeq2 y los de Expression Atlas, se confirma que todos los genes, en específico los 10 más significativos aparecen también en los datos del artículo, lo que significa el análisis fue exitoso y que ambos métodos con parámetros diferentes arrojan los genes con cambios muy marcados, lo que biológicamente significa que el gen *EXPA30* reprograma masivamente el transcriptoma de los pelos radiculares, y demuestra que los efectos observados no dependen del software de análisis diferencial, sino que reflejan una señal biológica real y robusta.

Conclusión.

En conclusión, el análisis diferencial realizado mediante DESeq2 permitió identificar una modificación del transcriptoma en la línea *OsEXPA30::GFP*, con 14000 mil genes diferencialmente expresados y 10 genes altamente significativos, los cuales tienen funciones relacionadas al crecimiento, transporte y metabolismo de la pared celular. Los genes hallados coinciden con los reportados por el artículo y Expression Atlas, confirmando que las células de los pelos radiculares presentan un programa transcriptómico especializado y altamente diferenciado entre las especies de plantas. Y que a pesar de la elección de parámetros diferentes entre DESeq2 y edgeR, los resultados demuestran que el efecto observado biológicamente es real.

Referencias.

Columbia university Irving medical center. (s.f.). *Tasa de falsos descubrimientos*. Columbia university Irving medical center. <https://www.publichealth.columbia.edu/research/population-health-methods/false-discovery-rate>

Expression Atlas. (2017). Diversification of Root Hair Development Genes in Vascular Plants - rice dataset. Expression Atlas. <https://www.ebi.ac.uk/gxa/experiments/E-CURD-28/Downloads>

Huang, L., Shi, X., Wang, W., Hui, K. & Schiefelbein, J. (2017). Diversification of Root Hair Development Genes in Vascular Plants. *Plant Physiology*, 174(3), 1697-1712. <https://doi.org/10.1104/pp.17.00374>

Jiménez, E. *Resumen sobre las Expansinas y su función*. Studocu. <https://www.studocu.com/es/document/universidad-de-cordoba-espana/biologia-molecular-y-celular-de-las-plantas/resumen-expansinas/90762359>

Ma, H., Zhao, H., Liu, Z. & Zhao, J. (2011). The Phytocyanin Gene Family in Rice (*Oryza sativa* L.): Genome-Wide Identification, Classification and Transcriptional Analysis. *Plos one*, 6(10). <https://doi.org/10.1371/journal.pone.0025184>

NIH. (s.f.). *LOC107277540 presunta expansina-A30 [Grupo Oryza sativa Japonica (arroz japonés)]*. NIH. <https://www.ncbi.nlm.nih.gov/gene/?term=EXPA30>