



UNIVERSIDAD NACIONAL DE COLOMBIA

# **Análisis comparativo de las redes metabólicas de metatranscriptomas edáficos de la rizósfera de *Solanum phureja* en tres diferentes estados fenológicos.**

**José Leonardo Bocanegra Silva**

Universidad Nacional de Colombia

Facultad de Ciencias

Bogotá, Colombia

2015

# **Análisis comparativo de las redes metabólicas de metatranscriptomas edáficos de la rizósfera de *Solanum phureja* en tres diferentes estados fenológicos.**

**José Leonardo Bocanegra Silva**

Tesis de investigación presentada como requisito parcial para optar al título de:  
**Magister en Ciencias Microbiología**

Director:

Ph. D., Q.F. Fabio Ancízar Aristizábal Gutiérrez

Codirector:

Ph.D., Ciencias Biológicas. Andrés Mauricio Pinzón Velasco

Línea de Investigación:

Microbiología Ambiental.

Grupo de Investigación:

Metagenómica de Suelos

Universidad Nacional de Colombia

Facultad de Ciencias

Bogotá, Colombia

2015

*(Dedicatoria o lema)*

*A mi hija, esposa y familia que me han apoyado y dado todo su amor para poder cumplir con las metas en este largo recorrido. Como el más valioso de los regalos que los seres queridos nos pueden ofrecer que es la educación.*

*La educación ayuda a las personas a aprender a ser lo que es capaz de ser.*

*Hesíodo*

## **Agradecimientos**

Quisiera agradecer especialmente al Dr. Fabio Ancízar Aristizábal, por haberme permitido desarrollar mi Tesis de Maestría junto a su equipo de trabajo y por haber contribuido con mi formación profesional y científica; y a mi codirector el Dr. Andrés Mauricio Pinzón, por sus valiosos aportes a esta investigación y especialmente por la amistad brindada, el optimismo entregado y los momentos compartidos durante el desarrollo de esta Tesis.

Quisiera agradecer además, a los compañeros Laura Emilia Cerón, Ibonne García, Guillermo Torres y Sandra González Sayer, con los cuales compartí durante el desarrollo de esta investigación y viví momentos muy especiales.

Un agradecimiento muy especial a Dios por la fuerza y guía que me brindo para el cumplimiento de una meta más en el largo caminar de la vida.

## Resumen

Una de las metas principales de la biología es descubrir los principios fundamentales que determinan la estructura y función de las células y los microorganismos. En este sentido, los proyectos de secuenciación a gran escala no solo han proveído la información de genomas completos de un gran número de organismos, sino, que han permitido el desarrollo de bases de datos que integran rutas metabólicas asociadas a su información genética, las cuales proveen la conectividad específica de mapas metabólicos, así como otro tipo de redes celulares. En el presente trabajo empleó la teoría de grafos para analizar la reconstrucción de redes metabólicas de tres metatranscriptomas obtenidos de muestras de suelo rizosférico de *Solanum phureja*, bajo condiciones de manejo orgánico en tres diferentes estados fenológicos (emergencia, aporque y floración).

A las redes reconstruidas se les realizó un análisis topológico para describir sus características estructurales, encontrándose un comportamiento de escala libre. En cuanto a la composición taxonómica, se encontró diferenciación entre los estados fenológicos a nivel de género especialmente en emergencia y, siendo menos o en algunos casos tendiente a desaparecer entre aporque y floración. A nivel funcional el metabolismo de Carbohidratos, Aminoácidos, Lípidos y Biosíntesis de Vitaminas fueron los que presentaron el mayor porcentaje de funciones enzimáticas asociadas, no obstante, los patrones observados a nivel funcional fueron muy globales y homogéneos entre los tres metatranscriptomas.

**Palabras clave:** Metatranscriptómica, redes metabólicas, teoría de grafos.

## Abstract

One of the most important goal in biology is to uncover the fundamental design principles that provide the common underlying structure and function in all cells and microorganisms. In this sense, large-scale sequencing projects have not only provided complete sequence information for a number of genomes, but also allowed the development of integrated pathway genome databases that provide organism specific connectivity maps of metabolic and, to a lesser extent, other cellular networks. Here, we use graph theory to analyze metabolic network reconstruction of three metatranscriptomics from rhizosphere of *Solanum phureja*, under organic management at three time points corresponding to distinct stages of plant development (seedling, hilling and flowering)

The reconstructed networks were analyzed topologically to describe their structural characteristics, being found a scale free behaviour. Regarding to taxonomic composition, it was observed that the rhizosphere microbial communities varied with plant development and that microbial communities in early plant development were more distinct to the hilling and flowering and that this difference decreased with plant age. The metabolism of carbohydrates, amino acids, lipids and biosynthesis of vitamins were those with the highest percentage of enzymatic functions associated, however, the observed patterns at functional level were global and homogeneous among the three metatranscriptomics.

**Keywords:** Metatranscriptomics, metabolic networks, graph theory.

# Contenido

	Pág.
<b>Resumen</b> .....	<b>V</b>
<b>Lista de figuras</b> .....	<b>IX</b>
<b>Lista de tablas</b> .....	<b>XI</b>
<b>Lista de símbolos y abreviaturas</b> .....	<b>XII</b>
<b>Introducción</b> .....	<b>1</b>
<b>1. Marco teórico</b> .....	<b>4</b>
1.1 El cultivo de la papa.....	4
1.2 El ecosistema edáfico.....	6
1.3 El ADN y ARN en el estudio de la microbiota edáfica .....	9
1.4 Biología de sistemas.....	11
1.5 Teoría de grafos y redes en biología.....	13
1.5.1 El crecimiento y conexionado de una red de escala libre .....	15
1.6 Aplicaciones de las reconstrucciones metabólicas.....	18
1.7 Métodos empleados en la reconstrucción del metabolismo .....	19
1.7.1 Métodos computacionales en la reconstrucción metabólica .....	20
1.8 Análisis de Redes .....	21
1.8.1 Clasificación de redes .....	21
1.8.2 Propiedades de las Redes .....	22
<b>Pregunta de investigación</b> .....	<b>24</b>
<b>Objetivo general</b> .....	<b>24</b>
<b>Objetivos específicos</b> .....	<b>24</b>
<b>2. Metodología</b> .....	<b>25</b>
2.1 Extracción del ARN.....	26
2.2 Procesamiento de datos .....	27
2.2.1 Procesamiento de los metatranscriptomas.....	28
2.2.2 Reducción de redundancia y obtención del Genotipo Metabólico.....	29
2.2.3 Identificación de las rutas metabólicas .....	30
2.2.4 Reconstrucción de la red del metabolismo .....	31
2.2.5 Análisis de Topología .....	31
<b>3. Resultados y discusión</b> .....	<b>32</b>

3.1	Preprocesamiento de los Datos e Identificación del Genotipo metabólico .....	32
3.2	Identificación de las rutas metabólicas .....	35
3.3	Reconstrucción de las redes y su caracterización topológica .....	38
<b>4.</b>	<b>Conclusiones .....</b>	<b>46</b>
<b>5.</b>	<b>Recomendaciones .....</b>	<b>48</b>
<b>A.</b>	<b>Anexo: Calidad de secuencias y anotación por función enzimática.....</b>	<b>49</b>
<b>B.</b>	<b>Anexo: Gráficas obtenidas de los análisis topológicos para los tres metatranscriptomas.....</b>	<b>52</b>
	<b>Bibliografía .....</b>	<b>55</b>

## Lista de figuras

	Pág.
Figura 1-1: Enfoques empleados en el estudio de las redes complejas: (a) el estudio de sus propiedades estructurales (topológicas) y (b) el estudio de sus propiedades dinámicas.....	13
Figura 1-2: Diferentes tipos de redes metabólicas. (a) Los nodos son sustratos o productos (metabolito-metabolito). (b) Los nodos son las enzimas, que se conectan cuando algún producto de una enzima es el sustrato de otra. (c) Bipartitas incluyen tanto sustratos y productos, como enzimas y son una combinación de los dos tipos de redes anteriores. ....	14
Figura 1-3: (a) Tres tipos de redes que se pueden obtener aumentando la reconexión de sus nodos aleatoriamente. (b) Valores de la distancia entre los nodos L y su agrupamiento C en una red en función del incremento de la aleatoriedad de las reconexiones. Modificado de Watts y Strogatz 1998. ....	15
Figura 1-4: Atributos de las estructuras de redes genéricas. (a) Estructura de una red aleatoria. (b) Los grados de conectividad entre los nodos siguen una distribución de Poisson, lo que indica que la mayoría de los nodos tienen aproximadamente el mismo número de enlaces. (c) Estructura red libre de escala unos pocos nodos altamente conectados, llamados hubs (gris). (d) El grado de conectividad en las redes de escala libre sigue una ley de potencias. Modificado de Barabási y Oltvai, 2004. ....	16
Figura 1-5: Clasificación de las redes de acuerdo a la direccionalidad de sus conexiones. ....	22
Figura 2-1: Ubicación del lugar de muestreo en el municipio de Tenjo, Cundinamarca. .	25
Figura 2-2: Desarrollo esquemático de la obtención de muestras de suelo rizosférico de papa criollo ( <i>Solanum phureja</i> ) hasta la secuenciación. ....	26
Figura 2-3: Aproximación esquemática de los diferentes enfoques metodológicos que se están empleando para el modelamiento y estudio de las redes de metabolismos, en donde se involucran diferentes niveles de detalle y complejidad, resaltando enfoques que son más cualitativos y otros más cuantitativos. Modificado de Steuer, 2007. ....	28
Figura 2-4: Esquema metodológico para la reconstrucción de un red metabólica. a) Pasos realizados desde las secuencias hasta la realización de la matriz estequiométrica. b) Procedimiento para la realización del análisis de la red de metabolismo a nivel topológico.....	30
Figura 3-1: Las 15 rutas KEGG más abundantes para cada uno de los metatranscriptomas para los diferentes estados fenológicos.....	36

Figura 3-2: Modelo escala libre para a) Emergencia, b) Aporque y c) Floración, en cada uno de los modelos las zonas rojas indican los nodos más conectados y las zonas verdes son las menos conectadas.....	39
Figura 3-3: Principales phyla encontrados en los diferentes metatranscriptomas para cada uno de los estados fenológicos de acuerdo al software MEGAN 5.0 en relación a los transcritos o lecturas asociadas. ....	41
Figura 3-4: Principales generos encontrados en los diferentes metatranscriptomas, para cada uno de los estados fenológicos de acuerdo al software MEGAN 5.0 en relación a los transcritos o lecturas asociadas. ....	42
Figura 3-5: Gráfica nube de palabras de la composición taxonómica a nivel de genero a. emergencia, b. aporque y c. floración. Obtenida a través del software MEGAN 5.0 en relación a los transcritos o lecturas asociadas. ....	43
Figura 3-6: Subredes obtenidas para los tres metatranscriptomas a) Emergencia, b) Aporque y c) Floración, en cada uno de los modelos las zonas rojas indican los nodos más conectados y las zonas verdes son las menos conectadas.....	44
Figura A-1: Histogramas de Calidad de los datos para cada uno de los diferentes estados fenológicos seleccionados.....	49
Figura B-2: Resultados gráficos de análisis topológico obtenido del software Cytoscape, para el metatranscriptoma en el estado de emergencia. A) Histograma de la distribución del Path más corto, B) Distribución del coeficiente de clusterización, C) Coeficientes topológicos y D) Distribución de la conectividad entre vecinos.....	52
Figura B-2: Resultados gráficos de análisis topológico obtenido del software Cytoscape, para el metatranscriptoma en el estado de aporque. A) Histograma de la distribución del Path más corto, B) Distribución del coeficiente de clusterización, C) Coeficientes topológicos y D) Distribución de la conectividad entre vecinos.....	53
Figura B-3: Resultados gráficos de análisis topológico obtenido del software Cytoscape, para el metatranscriptoma en el estado de floración. A) Histograma de la distribución del Path más corto, B) Distribución del coeficiente de clusterización, C) Coeficientes topológicos y D) Distribución de la conectividad entre vecinos.....	54

## Lista de tablas

	<b>Pág.</b>
Tabla 3-1: Número de lecturas por librerías tras cada paso de preprocesamiento. ....	33
Tabla 3-2: Número de secuencias anotadas y funciones enzimáticas encontradas para cada uno de los estados fenológicos, las cuales representan el genotipo metabólico de cada uno de los metatranscriptomas. ....	35
Tabla 3-3: Representación de cada una de las categorías funcionales del KEGG para los diferentes estados fenológicos. ....	37
Tabla 3-4: Resultados obtenidos para los parámetros de análisis de topología de redes realizado a cada uno de los metatranscriptomas para cada uno de los estados fenológicos en Cytoscape. ....	40
Tabla A-1: Rutas metabólicas encontradas en los metatranscriptomas para cada uno de los estados fenológicos. ....	50

# Lista de símbolos y abreviaturas

## Abreviaturas

<b>Abreviatura</b>	<b>Término</b>
ADN	Ácido Desoxirribonucleico
ADNc	Ácido Desoxirribonucleico Complementario
ARN	Ácido Ribonucleico
ARNm	Ácido Ribonucleico Mensajero
RNasa	Ribonucleasa
KEGG	Enciclopedia de genes y genomas de Kyioto

# Introducción

Los microorganismos constituyen la principal fuente de biodiversidad del planeta, y en especial en el suelo cumplen un papel relevante, ya que llevan a cabo funciones esenciales en el mantenimiento de los procesos reguladores de la biósfera (Streit y Daniel, 2010). Con base en lo anterior, para los ecosistemas agrícolas los microorganismos se convierten en un factor decisivo para el mantenimiento de la producción primaria de biomasa (comida, fibra, energía). No obstante, persiste un gran desconocimiento de la diversidad real presente en el ecosistema edáfico, a pesar de la importancia que los microorganismos representan para las funciones existentes en el suelo. Uno de los factores que ha mantenido esta tendencia ha sido que la investigación se ha enfocado al estudio de organismos cultivables (National Research Council (US) Committee on Metagenomics: Challenges and Functional Applications, 2007; Streit y Daniel 2010).

Con el desarrollo de nuevos enfoques metodológicos basados en tecnologías de ácidos nucleicos (ADN y ARN), se ha puesto de manifiesto que los métodos tradicionales de cultivo no revelan la totalidad de la diversidad microbiana presente en una muestra y tan solo el 1% de ésta es recuperada por cultivo tradicional (Amann *et al.* 1995; Torsvik *et al.* 1990), esto implica que a través de estas metodologías más del 99% de los microorganismos del suelo aún permanecen sin identificar (Borneman *et al.* 1996).

En virtud de lo anterior, es necesario avanzar más allá del escrutinio taxonómico de las poblaciones presentes en los ecosistemas edáficos, con el fin de abordar las incertidumbres existentes en relación de cómo los grupos funcionales microbianos responden a diferentes factores en las condiciones ambientales, la cual depende de las características funcionales (metabólicas) de los organismos involucrados más que de su identidad taxonómica (He *et al.* 2007). Por lo tanto, la riqueza de especies como una medida de diversidad biótica, y su capacidad funcional del ecosistema edáfico, proporcionarán suficiente información para la descripción de la composición y el rol que

cumplen las comunidades en un ecosistema determinado; en este caso el ecosistema edáfico.

Uno de los enfoques que hoy día permite abordar el estudio de los microorganismos de una forma más holística, que facilite el entendimiento de un sistema complejo como son los ecosistemas edáficos, han sido las aproximaciones que emplean metodologías de alto rendimiento, y son en especial los enfoques de la metagenómica y la metatranscriptómica, los que han permitido explorar los genomas y transcritos que son expresados en un momento dado por las comunidades microbianas (Zwolinski, 2007). Adicionalmente, ésta aproximación tiene la ventaja de no necesitar un conocimiento previo de la comunidad microbiana, ni el cultivo de los organismos, abriendo paso a la exploración de la diversidad taxonómica y funcional de las comunidades microbianas, en condiciones *in situ* de ambientes no explorados (Mouillot *et al.* 2004).

Las metodologías de alto rendimiento han puesto a disposición una gran cantidad de datos sobre microorganismos individuales, así como diferentes ambientes completos a nivel molecular (Urich *et al.* 2008). No obstante, el manejo de esta información ha hecho necesario abordar su análisis desde diversos métodos computacionales. Un método que hoy día ha ganado gran fuerza para el estudio de estos datos a escalas genómicas, es el de modelado, el cual emplea matemáticas exactas que permiten explicar la mecánica del sistema a detalle y pueden modelar incluso moléculas, pero al aumentar la complejidad por el volumen de datos, el modelado simple se vuelve ineficiente para abordar sistemas de gran tamaño.

Dado lo anterior, el modelamiento de datos biológicos se enfrenta al reto de tratar de explicar aquellos sistemas que exhiben una estructura o un comportamiento que no puede entenderse ni como un patrón ordenado, ni como un fenómeno altamente aleatorio, que da lugar a explicaciones estadísticamente simples. A dichos sistemas se los suele denominar complejos, y al menos en la mayoría de los casos se los puede pensar como un conjunto de muchos componentes de cuya interacción emerge la complejidad. Las interacciones en un sistema de muchos componentes puede caracterizarse por el concepto de red y es de esperar que las redes que representan a los sistemas complejos puedan heredar tal propiedad. De hecho la mayoría de las redes en la naturaleza exhiben una topología compleja.

El enfoque de redes para el modelado de datos biológicos ha resultado eficaz y su uso se ha difundido en años recientes (Balazsi *et al.* 2005), aplicándose en la reconstrucción metabólica, redes de regulación, entre otras, para explicar los diferentes fenómenos en los sistemas biológicos que exhiben un comportamiento complejo (Kerlebach y Shair, 2008). Este método es elegido para poder analizar sistemas de gran tamaño con una alta cantidad de datos, con el fin de tener una escala a la cual se pueda manejar y hacer predicciones sobre el comportamiento del sistema a analizar en su conjunto (Kerlebach y Shair, 2008).

En este sentido, el presente trabajo propone la aplicación de metodologías de reconstrucción de redes metabólicas, para caracterizar el ecosistema edáfico de la rizósfera de *Solanum phureja*, a partir del análisis topológico de las redes de metabolismo reconstruidas de tres metatranscriptomas, obtenidos en tres diferentes estados fenológicos (emergencia, aporque y floración), en un ciclo de cultivo bajo manejo orgánico, buscando la descripción del potencial funcional de las comunidades microbianas edáficas.

# 1. Marco teórico

## 1.1 El cultivo de la papa

La papa se encuentra dentro de los cuatro cultivos más importantes en la economía global, la agricultura y la seguridad alimentaria del mundo, después del trigo, el arroz y el maíz (Rodríguez 2009). Los cultivares de papa criolla, han sido agrupados dentro de *Solanum tuberosum* Grupo Phureja según la clasificación de Huamán y Spooner (2002). En Colombia la papa criolla representa un 10% de total de la producción de papa, y representa un potencial de exportación a Europa y Japón (Fajardo y Díaz, 2002), ya que posee cualidades como un 25% más de proteína que la papa “de año” (*Solanum tuberosum*), precursores de vitamina A y existen accesiones con resistencia al tizón tardío (*Phytophthora infestans*) entre otras características (Ochoa, 2001).

La papa criolla se encuentra cultivada, preferentemente, entre los 2000 – 3400 msnm, sus tubérculos no tienen periodo de reposo y es posible establecer ciclos de siembra - cosecha tres o cuatro veces al año (Ochoa, 2001). Adicionalmente, la calidad de los tubérculos de papa criolla no solo depende del manejo cultural del cultivo, sino de condiciones ambientales como suelo, microclima, altura, radiación solar y luminosidad de los ecosistemas donde se produce (Rivera *et al.* 2003). El desarrollo del cultivo de papa puede dividirse en: crecimiento vegetativo, desarrollo de las partes vegetativas cosechables, desarrollo reproductivo y senescencia (Valbuena, 2000).

La etapa de emergencia se evidencia 15 días después de la siembra, pero sólo hasta el día 21 ha emergido más de 50% de la población de plantas, posteriormente en el desarrollo de la planta se realizan otras prácticas culturales como es el aporque, periodo durante el cual se presenta la etapa de desarrollo de raíces y estolones, razón por la cual se ha dado un mal manejo en decir que el aporque es un estado fenológico, ya que esta práctica tiene como fin cubrir con suelo el mayor número de nudos para estimular la formación de raíces y estolones en cuyos tejidos apicales se desarrollan los tubérculos;

de ahí su relevancia e importancia, ya que los estolones descubiertos no forman tubérculos y crecen como tallos aéreos, los cuales son improductivos. El atraso de esta práctica cultural aumenta el daño en las raíces y perjudica el número de tubérculos (Lujan, 1994). El aporque se realiza en el primer mes después de que emergen las plantas, con el fin de evitar la afectación del sistema de raíces y estolones (Rivera, 2005). En papa criolla los primeros botones florales se observan, para más de 50% de la población después de los 50 días (Meier, 2001).

Español (2002), ha reportado que durante la emergencia y en la floración existe una fuerte actividad de insectos que perjudican el cultivo, de ahí que estas estepas sean críticas en el desarrollo de la planta; adicionalmente se ha encontrado que durante la floración existe un período máximo de fotosíntesis que presenta una relación lineal con el llenado de los tubérculos (Monroy, 2001). Por otro lado, el cultivo de papa hoy día se enfrenta a grandes retos dentro de los cuales encontramos el continuo deterioro de los suelos, conllevando a una afectación directa sobre la microbiota que puede llegar a estar presente en la matriz suelo y que como consecuencia puede generar alteraciones en la relación planta – microorganismo. Razón por la cual el estudio de las relaciones planta - microorganismo en el cultivo de papa cobra una gran relevancia con el fin de poder establecer las mejores relaciones entre los tipos de suelo, variedad vegetal y microbiota asociada.

Dado lo anterior, se han desarrollado investigaciones donde se ha mirado el efecto de diferentes fuentes de materia orgánica como complemento a la fertilización en papa criolla, encontrándose que el abono orgánico mejora en las variables densidad aparente, humedad volumétrica del suelo y conductividad, reflejando una recuperación de las propiedades del suelo (Neira *et al.* 1996; Rodríguez y Saa, 2004), así mismo las prácticas de manejo orgánico en el cultivo de papa criolla han mostrado diferencias significativas en cuanto a rendimientos y costos en relación con las prácticas de abono químico (Muñoz y Lucero, 2008), del mismo modo Faccini *et al.* (2007), demostró que la aplicación de inóculos de solubilizadores de fosfato mejora los rendimientos del cultivo y en los costos del mismo, resaltándose la importancia de la comunidad edáfica como un factor crucial para el cultivo de la papa criolla.

## 1.2 El ecosistema edáfico

Las funciones y estructuras de las comunidades microbianas del ecosistema edáfico, están fuertemente influenciadas por las características físicas del suelo (Feeney *et al.* 2006), así como, de un rango de parámetros ambientales, incluidos: pH (Dennis *et al.* 2010), temperatura (Ward *et al.* 1998), contenido de humedad (Zhou *et al.* 2002), disponibilidad y complejidad de sustratos (Dennis *et al.* 2010), así como la exposición a las raíces de diferentes especies de plantas (Kuske *et al.* 2002) entre otras.

Las raíces de las plantas desempeñan un papel activo en modificar el medio ambiente del suelo a través de una serie de procesos físicos, químicos y biológicos (Jaillard *et al.* 2001), que son conocidos como rizodeposición y la zona en la cual se desarrolla como rizósfera (Bowen y Rovira, 1999). La rizósfera puede definirse como el volumen de suelo que interactúa directamente con las raíces de las plantas (Bowen y Rovira, 1999; Mengel *et al.* 2001). Nanómetros o centímetros de distancia radial a partir de la superficie de la raíz, actúan como interfaz entre ésta y el mundo mineral, sirviendo como foco de retroalimentación ecológica y comunicación entre organismos (Richter, 2007).

La extensión de la rizósfera no está definida con precisión, pero algunos investigadores la han estimado entre 2 a 5 mm (Youssef *et al.* 1989; Yeates y Darrah, 1991; Badalucco *et al.* 1996). Sin embargo, Pepper y Bezdicsek (1990) informaron que la rizósfera podría extenderse hasta 20 mm, debido a una serie de gradientes como sustratos orgánicos, microorganismos, pH, oxígeno (O<sub>2</sub>), dióxido de carbono (CO<sub>2</sub>) y agua (H<sub>2</sub>O). Además, los microorganismos varían en su sensibilidad frente a las sustancias solubles y volátiles procedentes desde la raíz. Asimismo, la extensión de la rizósfera dependerá de la difusión de los numerosos compuestos liberados desde la raíz, de las características de difusión, del estado hídrico y las características físicas del suelo (Bowen y Rovira, 1999).

A pesar de no existir un consenso respecto a la extensión de la rizósfera, lo que sí es claro, es que esta representa una región de intensa actividad microbiológica, la cual es mantenida a través de la liberación de una serie de compuestos orgánicos como exudados (moléculas pequeñas, tales como ácidos orgánicos, aminoácidos y azúcares), secreciones (como las enzimas), lisados de las células muertas y mucílago (Araújo *et al.* 2003; Garbeva *et al.* 2004; van Overbeek y van Elsas, 2008). A nivel cuantitativo y cualitativo estos compuestos pueden variar no solamente entre las diferentes especies

vegetales, sino también por los diferentes genotipos de los cultivares, generando efectos sobre la estructura y funcionamiento de las comunidades microbianas asociadas con las plantas (Gu y Mazzola, 2003; Bouffaud, 2012; Inceoglu, 2012); evidencia de esto se ha descrito en el maíz (Schilling *et al.* 1998), especies de gramíneas (Morse *et al.* 2000), guisantes y tomate (Whipps, 1987). En papa (*Solanum tuberosum*), la caracterización de la comunidad microbiana en la rizósfera de tres cultivares creciendo en dos diferentes sitios de cultivo revelaron que, dependiendo del tipo de suelo, del 4-9% de las unidades taxonómicas operacionales detectadas por el PhyloChip<sup>1</sup> fueron dependientes del cultivar (Weinert *et al.* 2011).

Ha sido observado también en cultivos de papa que el estado de desarrollo tiene gran influencia en la composición de la comunidad de microorganismos presentes en la rizósfera, observándose que en los estadios tempranos de desarrollo existe alto grado de variabilidad en la comunidad bacteriana, mientras que en la floración la variación en comparación con otros estados más adultos (Inceoglu, 2011). En estados jóvenes de desarrollo se ha observado que en la rizósfera de papa existe una abundancia relativa de Actinobacteria de aproximadamente (49%) and comparado con otros estadios de desarrollo, este efecto desaparece en estados como floración y senescencia donde las abundancias relativos son similares (Inceoglu, 2011). Por otro lado Inceoglu (2011) encontró que las Alphaproteobacteria mostraron un efecto rizosférico positivo al cultivar en todos los estados de desarrollo.

Las prácticas agrícolas y estado de desarrollo de la planta han sido encontradas como elementos fundamentales en la relación planta - microorganismos (Smalla *et al.* 2001; Gu y Mazzola, 2003). En este sentido, se ha descrito que la aplicación de enmiendas con compost, altera la estructura de la comunidad microbiana del suelo transitoriamente (Ros *et al.* 2006; Saison *et al.* 2006). Sin embargo, la evidencia adicional apunta a la persistencia de microorganismos derivados de la enmienda, sobre todo en asociación con plantas en donde las semillas son germinadas en suelos con compostaje, las cuales son colonizadas por microorganismos derivados de la enmienda y esta comunidad

---

<sup>1</sup> Es un microarreglo de ADN, que emplea sondas de genes de 16S ARNr para identificar múltiples organismos.

cambia durante la transición de la espermosfera a rizósfera (Green *et al.* 2006). La comunidad microbiana es diferente en el suelo con compost en comparación de suelo sin compost (Tiquia *et al.* 2002; Benítez *et al.* 2007), y pueden llegar a tener una constitución de microbiota derivada de suelo sin compost y con compost (Inbar *et al.* 2005).

En relación al desarrollo de la planta se ha visto cambios en los patrones de exudación; generalmente se ha encontrado un alto patrón de exudados durante la fase inicial de crecimiento, posiblemente debido al almacenamiento en la semilla de diferentes tipos de compuestos (Hamlen *et al.* 1972). Existe evidencia donde la estructura de la comunidad microbiana de la rizósfera sufre cambios, como fue observado en la rizósfera *Medicago truncatula* en diferentes estados vegetativos y reproductivos (Mougel *et al.* 2006); tales sucesiones de la comunidad microbiana inician en los primeros estadios de desarrollo de la planta, a partir de la liberación de carbono de las semillas durante la germinación (Nelson, 2004). Importante resaltar que los microorganismos no se distribuyen de forma homogénea en la rizósfera, sino, que estos se distribuyen dependiendo del tipo de raíces (primarias o secundarias) y zonas de las raíces (cofia radicular, meristemo lateral y células del borde), a su vez, el crecimiento de la raíz va influenciar la distribución y composición de las comunidades microbiana a través del suelo (Laurent *et al.* 2013).

Hoy día se reconoce que la rizósfera es la zona del ecosistema edáfico que representa la fuente de mayor diversidad microbiológica encontrada en el planeta tierra (Nannipieri, 2003; Elsas *et al.* 2008, Kakirde *et al.* 2010, Mocali y Benedetti, 2010). Diferentes autores estiman que las comunidades microbianas allí presentes contienen millones de bacterias, arqueas, virus y microorganismos eucariotas (Kakirde *et al.* 2010). Algunos reportes señalan que la diversidad encontrada en tan solo 1 gramo de suelo rizosférico oscila entre  $10^3$  y  $10^7$  especies (Fierer *et al.* 2007). En consecuencia, el estudio de las comunidades microbianas edáficas, reviste una gran complejidad que requiere un análisis más allá del escrutinio taxonómico de las poblaciones, sino que además se debe tener en cuenta la dinámica de los procesos del ecosistema, la cual depende de las características funcionales (metabólicas) de los organismos involucrados más que de su identidad taxonómica, como indican algunos autores (He *et al.* 2007).

La diversidad del suelo describe complejidad y variabilidad en diferentes niveles de organización biológica. Cubre la variabilidad genética dentro de taxón (especies), el

número (riqueza), abundancia relativa (uniformidad) de taxones y grupos funcionales (dinámicas metabólicas) en las comunidades (Torsvik y Øvreås, 2002). Los microorganismos constituyen el componente primario de la biodiversidad presente en el ecosistema edáfico, y a los cuales, se les atribuye un papel esencial en el mantenimiento de los procesos reguladores de la biósfera (Streit y Daniel, 2010), a partir de funciones ecosistémicas críticas como el ciclado de compuestos orgánicos e inorgánicos (ciclos biogeoquímicos) (Quijano *et al.* 2013; Holtgrieve *et al.* 2009)

### **1.3 El ADN y ARN en el estudio de la microbiota edáfica**

El conocimiento de la diversidad activa dominante se considera una pieza clave para entender el funcionamiento de un ecosistema (Mouillot *et al.* 2004), y cómo sus comunidades (biomasa y actividad) son responsables de mantener dicha dinámica funcional a través de las variaciones del tiempo y el espacio. En este sentido, los ecosistemas edáficos han presentado limitaciones en su estudio debido a la imposibilidad de poder hacer crecer la mayoría de los microorganismos en medios de cultivo (Vartoukian *et al.* 2010). No obstante, el desarrollo de metodologías independientes de cultivo ha permitido comprender nuevas perspectivas acerca de diversidad, distribución, función y adaptación potencial de las comunidades microbianas (Urich *et al.* 2008; Frias *et al.* 2008), a través de la extracción de ADN, ARN, proteínas y metabolitos directamente de muestras en ambientes naturales y, analizándolos vía metagenómica, metatranscriptómica, metaproteómica y metabolómica, respectivamente (Carvalhais *et al.* 2012).

Las metodologías independientes de cultivo han sufrido un cambio significativo en los últimos años gracias al desarrollo de las tecnologías de secuenciación de alto rendimiento (Edwards *et al.* 2006). Estas tecnologías son insumo básico en las aproximaciones metagenómicas y metatranscriptómicas, las cuales han permitido obtener de forma masiva y rápida conjuntos de datos que van acorde a la complejidad de las comunidades microbianas edáficas (Carvalhais *et al.* 2012). Estas aproximaciones rápidas y eficientes de secuenciación y ensamblaje de fragmentos de ácidos nucleicos, permiten explorar los genomas y transcritos que son expresados en un momento dado por las comunidades microbianas, a partir del análisis de la mezcla total de ADN o ARN de la comunidad de organismos, en los que se incluyen, bacterias, arqueas, eucariotas y

virus (Zwolinski, 2007), con la ventaja de no necesitar un conocimiento previo de la comunidad microbiana, ni el cultivo de los organismos, por lo que se abre paso a un estudio holístico sobre diversidad taxonómica y funcional de las comunidades microbianas, conocimiento clave para caracterizar los ecosistemas edáficos (Mouillot *et al.* 2004; Edwards *et al.* 2006).

Sin embargo, se ha encontrado que los estudios basados en ADN pueden presentar patrones globales similares (Urich *et al.* 2008); lo que hace que esta aproximación no sea la más adecuada para describir roles funcionales de genes u organismos en un ambiente determinado, ya que no permite analizar el estado de expresión de los mismos (Urich *et al.* 2008). Lo anterior ha puesto de manifiesto la necesidad de no solo responder la pregunta de ¿quiénes se encuentran? sino ¿qué están haciendo? dichas comunidades. En este sentido, el uso del ARN ha sido la aproximación más adecuada para analizar el potencial genético activo de los microorganismos, a diferentes condiciones ambientales.

No obstante, la extracción de ARN ha presentado limitaciones técnicas asociadas a la poca estabilidad de estas moléculas, que a su vez, la hace una molécula con un promedio de vida en el rango de segundos a pocos minutos (Deutscher, 2006); así mismo, se ha descrito que estas características pueden variar entre las diferentes especies de microorganismos (Bernstein *et al.* 2002; Selinger *et al.* 2003), y estar influenciadas por el estatus nutricional de las células a nivel individual (Redon *et al.* 2005). En el caso de muestras ambientales provenientes de suelo, se suman otros elementos como la presencia de RNasa, presencia de moléculas orgánicas complejas como ácidos húmicos y fúlvicos, que típicamente se precipitan a la par con los ácidos nucleicos en los procesos de extracción y pueden llegar a inhibir posteriores procesos moleculares (Arbeli y Fuentes, 2007; Opel *et al.* 2010).

Frente a las dificultades para acceder al ARN, en los últimos años se han desarrollado métodos de extracción más eficientes, que logran abordar la mayoría de las limitantes identificadas en su extracción (DeCoste *et al.* 2011). Superadas las limitaciones, el análisis del ARNm de muestras complejas permite el desarrollo de la aproximación metatranscriptómica (Gilbert *et al.* 2008), que en el caso de estudios de comunidades edáficas, logra convertirse en una poderosa herramienta para dilucidar las complejas interacciones allí presentes (Carvalhais *et al.* 2012). La metatranscriptómica ha permitido

obtener secuencias de genes funcionales que se están transcribiendo y cuya expresión puede ser clave en los procesos de los ecosistemas (Botero *et al.* 2005); esto último, aislando directamente de muestras ambientales el ARNm eucariótico (Grant *et al.* 2006) y ARNm procariótico (Poretsky *et al.* 2005; Frias *et al.* 2008).

A pesar de que la aproximación metatranscriptómica no siempre es equiparable con la expresión fenotípica, ya que pueden existir modificaciones postranscripcionales de las proteínas, los procesos postranscripcionales y postraduccionales regulan la expresión génica para la síntesis de proteínas, siendo el control de nivel transcripcional de la expresión génica el que permite a los organismos como las bacterias, adaptarse rápidamente a las condiciones ambientales cambiantes (Moran, 2009), razón por la cual, las respuestas inmediatas a los cambios ambientales pueden ser mejor reflejadas por el metatranscriptoma que por el metaproteoma (Moran, 2009).

## 1.4 Biología de sistemas

El estudio de la biología molecular con sus remarcables avances ha permitido la identificación de un gran número de secuencias de genes y las funciones de sus productos transcripcionales, dando como resultado un entendimiento del comportamiento y atributo de biomoléculas a nivel individual (Bader *et al.* 2008). No obstante, el estudio de las biomoléculas una por una, no permite predecir las consecuencias de determinados mecanismos biológicos para el organismo en su conjunto (Hess, 1970; Barabási y Oltvay, 2004); mecanismos sujetos a propiedades de funciones supramoleculares tales como el ciclo celular, estados metabólicos entre otros; los cuales son producto de interacciones complejas entre los diferentes y numerosos constituyentes de la célula (Hartwell *et al.* 1999).

La complejidad de los organismos ha hecho que los investigadores exploren nuevos enfoques para lograr un mayor entendimiento de los procesos celulares; encontrando en la integración de las llamadas ciencias “ómicas”, la aproximación más adecuada para abordar tal problemática. La integración de las ciencias ómicas permite asociar análisis de conjuntos de datos de múltiples escalas que a través de procedimientos tradicionales no es posible (Zhang *et al.* 2010). Este esfuerzo interdisciplinario que combina trabajos experimentales con modelamiento matemático y análisis computacionales, es lo que hoy

día se conoce como Biología de Sistemas (Guebel *et al.* 2012). Esta última, logra un nuevo enfoque hacia el entendimiento del metabolismo celular como un “sistema integrado”, gracias al carácter holístico de las metodologías de alto rendimiento (Metzker, 2010), las cuales proporcionan datos a gran escala, posibilitando enfocar la investigación de moléculas individuales hacia el análisis de redes metabólicas (Barabási y Oltvay, 2004).

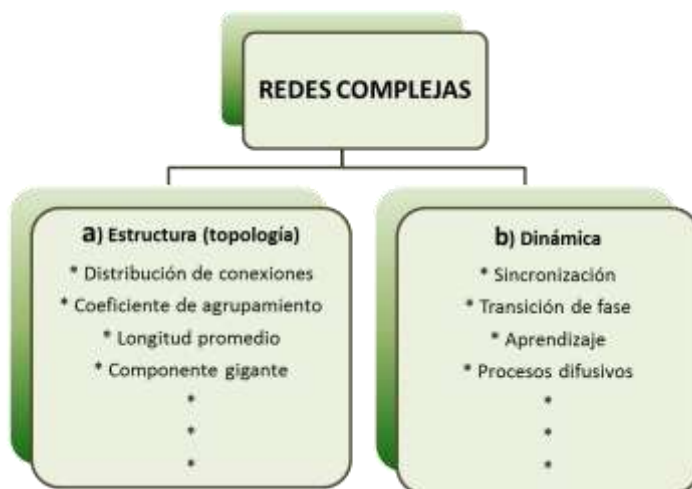
La reconstrucción de redes surge como una aproximación que busca determinar las actividades metabólicas y fisiológicas de las células en diferentes especies (Ma y Zeng, 2006), a través de modelos computacionales que logran simplificar el fenómeno biológico e integrar información heterogénea y de altos niveles de complejidad, para generar modelos que representen todas las relaciones entre metabolitos dentro de la célula por medio de vías metabólicas (Pinzón *et al.* 2010). A partir de esta aproximación, la biología de sistemas ha permitido el estudio integrado y el desciframiento de redes metabólicas y regulatorias, presente en los procesos celulares en diferentes organismos y bajo diferentes condiciones ambientales (Ma y Zeng, 2006).

Las reconstrucciones de metabolismo ha tenido un enfoque a organismos individuales tanto eucariotas como procariotas (Feist *et al.* 2009), analizando dos aspectos principales de las redes biológicas (Barabási y Oltvay, 2004) (**Figura 1-1**). El primero, es describir sus propiedades topológicas, tanto globales, como la libertad de escala y la existencia de módulos funcionales como locales, dándose especial relevancia al agrupamiento y la formación de motivos. El segundo aspecto son los estudios dinámicos, en los cuales se trata de simular analíticamente o de dar seguimiento a cierto fenómeno, como el metabolismo, la transcripción génica, la señalización o la división celular.

Típicamente las reconstrucciones metabólicas se realizan a escala genómica, es decir, teniendo en cuenta el conjunto de genes codificados en un genoma completo (Becker y Palsson, 2005; Feist y Palsson, 2008), aunque estas pueden ser también realizadas a nivel transcriptómico (es decir teniendo en cuenta el conjunto de genes expresados) también conocidas como reconstrucciones dirigidas (Pinzón *et al.* 2010). Sin embargo, esta aproximación no ha sido aplicada a nivel de reconstrucciones de metatranscriptomas, lo cual convierte a este en un proceso novedoso que permite

entender las interacciones de reacciones y metabolitos, integrando la información de varios microorganismos en un ambiente común.

**Figura 1-1:** Enfoques empleados en el estudio de las redes complejas: (a) el estudio de sus propiedades estructurales (topológicas) y (b) el estudio de sus propiedades dinámicas.



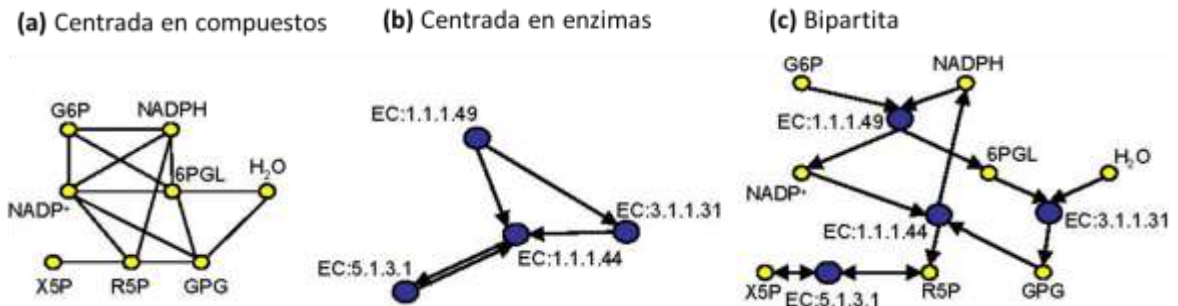
## 1.5 Teoría de grafos y redes en biología

Las reacciones químicas que ocurren dentro de las células, las cuales están ligadas unas a otras formando una intrincada y compleja red, donde los metabolitos son transformados de acuerdo a la ley de la química y termodinámica, tradicionalmente han sido organizadas en términos de rutas, que se encuentran interconectadas formando un mapa interconectado de las reacciones, esta representación es un grafo, el cual describe un conjunto de elementos entrantes y otro conjunto de elementos salientes, y todo en conjunto es lo que se denomina red de metabolismo.

Dado lo anterior, la aplicación de la teoría de grafos en modelos biológicos especialmente a lo que se refiere a metabolismo (**Figura 1-2**), ha tenido como fin el de ofrecer una abstracción matemática para la descripción de las relaciones existentes entre las diferentes moléculas intervinientes a través de una representación gráfica, esto ha hecho que se emplee el término grafo y red indistintamente (Huber *et al.* 2007). Un grafo consiste en un conjunto de nodos y un conjunto de vértices que conectan los nodos; los nodos son las entidades de interés (proteínas, productos químicos, genes u otras

entidades biológicas) y los vértices (conexiones) representan las relaciones entre las entidades (Alison y Lydia, 2009).

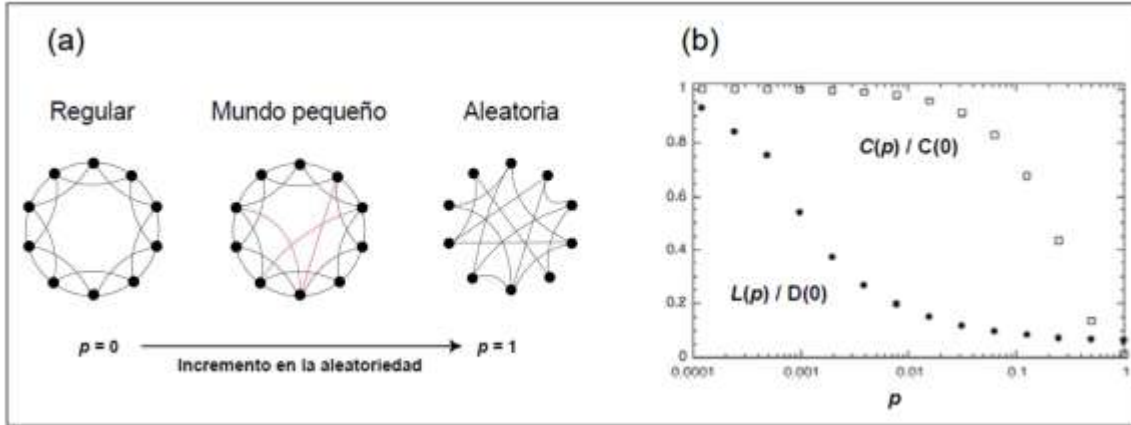
**Figura 1-2:** Diferentes tipos de redes metabólicas. (a) Los nodos son sustratos o productos (metabolito-metabolito). (b) Los nodos son las enzimas, que se conectan cuando algún producto de una enzima es el sustrato de otra. (c) Bipartitas incluyen tanto sustratos y productos, como enzimas y son una combinación de los dos tipos de redes anteriores.



A pesar de la utilidad que representa la teoría de grafos en el estudio de sistemas complejos como los biológicos, es tan solo hasta cuando se define que las redes reales exhiben una serie de características previamente descritas a nivel matemático, como ser redes de mundo pequeño y una organización de escala libre, cuando se despertó el interés de estas aproximaciones de análisis para el entendimiento del comportamiento celular (Huber *et al.* 2007).

La probabilidad de que los nodos de una red o grafo que no son vecinos entre sí, puedan ser alcanzados a través de una distancia corta (número de pasos), a pesar de que las redes sean grandes y dispersas, es lo que se conoce como redes de mundo pequeño (Watts y Strogatz, 1998; Barabási, 2002). Además, en estas últimas existe un alto grado de agrupamiento; supongamos que existe un nodo A que se conecta con otros tres nodos (B, C y D), el agrupamiento implica que además de las conexiones de A con sus vecinos; B, C y D también tienen conexiones entre ellos (**Figura 1-3**). O sea, los “amigos de mis amigos, son amigos entre sí”. Estos trayectos cortos de comunicación entre los nodos como el alto grado de agrupamiento, tienen un gran significado a nivel biológico, ya que tienen un impacto inmediato en la funcionalidad de la red, en el sentido que va a determinar la velocidad y la forma con la que se puede generar una respuesta a una situación dada (Barabási y Oltvai, 2004).

**Figura 1-3:** (a) Tres tipos de redes que se pueden obtener aumentando la reconexión de sus nodos aleatoriamente. (b) Valores de la distancia entre los nodos  $L$  y su agrupamiento  $C$  en una red en función del incremento de la aleatoriedad de las reconexiones. Modificado de Watts y Strogatz 1998.



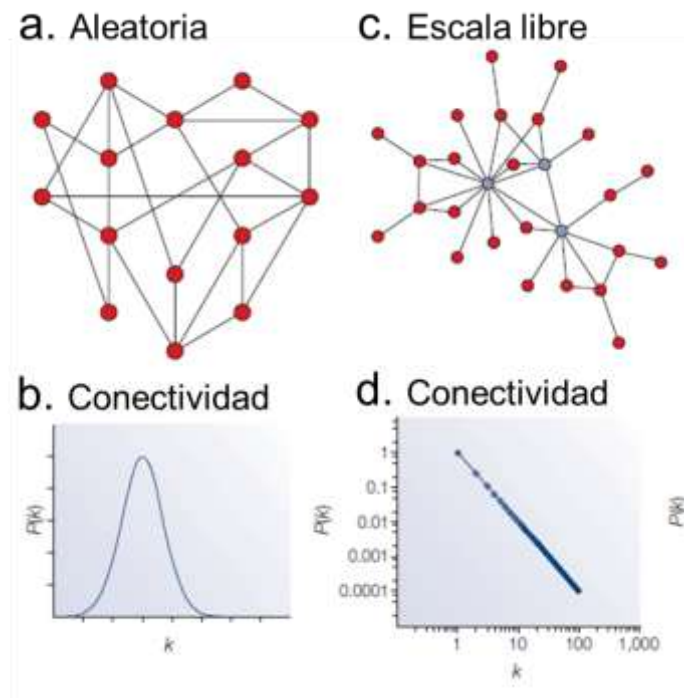
Por otro lado, la existencia de una pequeña fracción de nodos altamente conectados, los cuales son llamados conectores o "hubs", mientras que la mayoría de los nodos están pobremente conectados (**Figura 1-4**), es a lo que se le denomina redes de escala libre que son un tipo de red de mundo pequeño. Estas redes tienen la característica de una particular distribución (grados de conectividad) tipo leyes de potencia (o de cola larga), donde la probabilidad de que un nodo  $p(k)$  esté conectado con  $k$  nodos es proporcional a  $k^{-\gamma}$ , por lo que  $(K)$  no representaría propiamente a la red; este hallazgo sugiere que las redes complejas en la vida real no se conectan de forma aleatoria (Barabási y Albert 1999). En este sentido, Price (1976) fue más allá de lo empírico, al proponer un modelo por el cual se pueden generar *hubs* y redes libres de escala, basándose en el hecho de que los artículos muy citados tienden a ser más y más citados con el tiempo, en un proceso que llamó "ventaja acumulativa". La idea de este proceso, es hoy día más conocida como "acoplamiento preferencial con los nodos más conectados".

### 1.5.1 El crecimiento y conexionado de una red de escala libre

Con el fin de entender como los nodos se distribuyen en una dada posición espacial, y se conectan a nodos existente a través de un mecanismo de probabilidad que favorece cortas distancias y altos grados de conectividad, Barabási y su grupo proponen el modelo de acoplamiento preferencial, el cual es conocido también como el modelo "BA" (Barabási y Albert 1999). Modelo ampliamente aceptado como una de las explicaciones

de la existencia de las redes libres de escala; ya que por mucho tiempo se planteó que existía una población fija de  $N$  nodos y un número fijo  $M$  conexiones añadidas aleatoriamente. Sin embargo, en la realidad esto no ocurre, las redes no están fijas. Por el contrario, las redes complejas evolucionan y crecen en el tiempo a través de la adición simultánea tanto de conexiones como de nodos, por lo que el número de nodos crece durante toda la existencia de la red. En segundo lugar, los patrones aleatorios presuponen que la probabilidad de que dos nodos se conecten es aleatoria y uniforme. Sin embargo, la mayor parte de las redes reales muestran preferencias a la hora de conectarse (Barabási y Albert, 1999; Guimera y Nunes, 2005).

**Figura 1-4:** Atributos de las estructuras de redes genéricas. (a) Estructura de una red aleatoria. (b) Los grados de conectividad entre los nodos siguen una distribución de Poisson, lo que indica que la mayoría de los nodos tienen aproximadamente el mismo número de enlaces. (c) Estructura red libre de escala unos pocos nodos altamente conectados, llamados hubs (gris). (d) El grado de conectividad en las redes de escala libre sigue una ley de potencias. Modificado de Barabási y Oltvai, 2004.



Para explicar el comportamiento antes mencionado, supongamos que el proceso inicia en el tiempo  $t=1$  con 2 nodos unidos por  $m$  conexiones; en cada tiempo  $t \geq 2$ , nosotros adicionamos un nuevo nodo con  $m$  conexiones que unen el nuevo nodo a los nodos que

ya están presentes en la red. Se asume que la probabilidad  $\pi_i$  de que el nuevo nodo será conectado al nodo  $i$  depende en el grado  $deg(i)$  de  $i$  de manera que:

$$\pi_i = \frac{deg(i)}{\sum_j deg(j)}$$

En pocas palabras, cuando se añade un nuevo nodo se añaden conexiones una a la vez, con la segunda y subsiguientes conexiones que hacen conexión preferencial con el grado de los mismos. De forma más simple, se espera que una nueva publicación tenga más probabilidades de citar a publicaciones previamente muy citadas que a publicaciones desconocidas, lo que significa que las publicaciones mas citadas a su vez tienen más probabilidades de ser nuevamente citadas.

Un segundo modelo generado por el grupo de Barabási (Ravasz *et al.* 2002; Ravasz y Barabási, 2003), trata de explicar el origen de la modularidad<sup>2</sup> en las redes libres de escala, basándose en los componentes del modelo BA, pero añade un grado más de complejidad al formar grupos de nodos altamente conectados de manera sistemática emulando un crecimiento fractal. Lo anterior plantea, un modelo como un triángulo cuyos vértices son nodos que se conectan a un cuarto nodo central, por definición los cuatro nodos están bien agrupados. Ahora tripliquemos ese grupo y conectemos los nuevos nodos con los preexistentes, recordemos que el nodo está ganando cada vez más conexiones por acoplamiento preferencial. Siguiendo con el proceso reiteradamente se generara una red con *hubs* como el nodo central, altamente agrupada, jerárquica y modular.

El modelo de conexionado preferencial es uno de los principales mecanismos que explica la formación de la topología en redes empíricas, ya que ha sido fácil la obtención de registros de la evolución de las mismas. Dentro de estas redes tenemos las redes sociales o tecnológicas. No obstante, en el caso de las redes biológicas la obtención de registros de su evolución suele ser más difícil, lo que ha hecho que no sea clara del todo esta cuestión en las redes biológicas. Aunque, no está claro que el modelo de

---

<sup>2</sup> En algunas redes libres de escala se pueden identificar grupos de nodos altamente conectados entre ellos, y no así con el resto de la red, a esos grupos se les llama **módulos**.

conexionado preferencial sea el mecanismo principal de topologías libres de escala en redes biológicas y en particular las ecológicas (Jordan y Scheuring, 2002; Bascompte y Stouffer, 2009). Lo anterior plantea que puede haber otros mecanismos a demás del conexionado preferencial en la formación de *hubs* en redes biológicas.

Kunin y Pereira (2004), reportaron para *Saccharomyces cerevisiae* que en la red de interacción de proteínas, la funcionalidad de una clase de proteínas es un factor tan o más importante que la antigüedad en la determinación de la conectividad de las proteínas. Adicionalmente, en la red de interacción de proteínas de la *S. cerevisiae* un mecanismo de reconexionado ocurre a una tasa significativamente mayor que el mecanismo de duplicación de nodos, sugiriendo que el mecanismo de reconexionado constituye un factor tan o más relevante que el conexionado preferencial en la formación de la topología de la red (Berg *et al.* 2004).

## 1.6 Aplicaciones de las reconstrucciones metabólicas

Se describen cinco potenciales áreas de aplicación (Oberhardt *et al.* 2009): (1) La contextualización de datos a gran escala, que permite desarrollar nuevas herramientas estadísticas para modelar, organizar y comprender las redes biológicas a diferentes niveles que permitan inferir relaciones funcionales; (2) La orientación de la ingeniería metabólica, direcciona la creación de nuevas tecnologías analíticas para entender el metabolismo celular (ej. componentes transcripcionales, proteínas y enzimas metabólicas implicadas en la regulación celular). (3) El redireccionamiento de la hipótesis dado el descubrimiento, genera modelos informales para resolver los problemas metodológicos, que surgen al analizar grandes datos de información, bajo la validación de la hipótesis inicial de investigación. Por otra parte, (4) El cuestionamiento de las múltiples relaciones entre especies, crea nuevas estrategias *in silico* que permite realizar estudios de microorganismos de difícil obtención y cultivo, para reconocer y estudiar sus relaciones celulares individuales y entre especies. Por último, (5) la detección de las propiedades de las redes, permite generar nuevas técnicas de análisis entre las relaciones de los componentes (metabolitos y reacciones) que participan en los complejos procesos celulares, y de ahí, derivar la hipótesis acerca de su estructura y funcionamiento biológico, un ejemplo de esto se puede encontrar cuando se necesita saber si las proteínas que interactúan también son transcripcionalmente co-regulados. Esto podría

ser contestado mirando las respectivas redes, el gráfico de la interacción proteína-proteína y el gráfico de coexpresión, y las pruebas de sí y cómo la superposición de estos gráficos es más de lo que se esperaría por casualidad (Ge *et al.* 2001).

## 1.7 Métodos empleados en la reconstrucción del metabolismo

Con el fin de realizar una reconstrucción metabólica y representarla en un grafo, es necesario iniciar con la identificación del genotipo metabólico, el cual es entendido como todas las reacciones químicas que un organismo puede catalizar por las enzimas codificadas en su genoma (Oberhardt *et al.* 2009), en consecuencia, se debe tomar el genoma o conjunto de secuencias del organismo objetivo e identificar los genes codificados, para posteriormente, realizar la identificación del subconjunto de genes implicados en los procesos metabólicos.

Teniendo el conjunto de genes, el siguiente paso de la reconstrucción es determinar las reacciones bioquímicas que estas enzimas llevan a cabo; este procedimiento puede ser manual o automatizado, empleando bases de datos metabólicas tales como: KEGG (Kanehisa *et al.* 2006), Brenda (Schomburg *et al.* 2004), MetaCyc (Krieger *et al.* 2004), SEED (DeJongh *et al.* 2007) y Transport DB (Ren *et al.* 2007), las cuales contienen colecciones de información de reacciones metabólicas y de transportadores presentes en un rango de diferentes organismos, y que utilizan como nomenclatura la base de datos de Enzyme EC (Enzyme commission) (Moss, 2011) o en algunos casos la TC (Transport commission) (NCBI 2011).

El siguiente paso consiste en la curación manual de la reconstrucción, dado que al emplear métodos automatizados se puede generar un modelo incompleto por la ausencia o la pérdida de reacciones bioquímicas, conocidas como “gaps”. La estrategia consiste en realizar una curación manual, para completar los “gaps” en las redes por inferencia a través de evidencia directa de la literatura científica. Esta laboriosa tarea de la curación manual crea una base de conocimiento estructurado de información (Oberhardt *et al.* 2009). Posteriormente se transforma toda la información en un modelo computacional, con base en una matriz de estequiometría “S”, en la cual se expresan las relaciones entre

metabolitos, producidos y consumidos, en cada una de las reacciones que hacen parte de la reconstrucción.

### **1.7.1 Métodos computacionales en la reconstrucción metabólica**

Durante una reconstrucción metabólica existen una serie de pasos expuestos en el apartado anterior; pasos que en la mayoría de los casos son realizados a través de métodos computacionales que emplean modelos matemáticos, para demostrar gráficamente la presencia de metabolito, enzima, las relaciones de producto-sustrato entre metabolitos y entre las mismas reacciones (Pitkanen *et al.* 2010). En este sentido, se ha descrito que los métodos computacionales para la predicción de vías metabólicas pueden ser divididos en dos grupos (Pitkanen *et al.* 2010). El primero corresponde al método comparativo que predice si una ruta metabólica de referencia o una variación de esta, se encuentra presente en el organismo objetivo; y el segundo es el método exploratorio, que se encarga de detectar las rutas metabólicas posibles a partir de una colección de rutas de referencia en bases de datos, donde ya se han reportado en otros organismos. Cabe señalar que ambos métodos tienen los mismos enfoques, dado que se apoyan en la incorporación de datos experimentales que permiten guiar la búsqueda de una mejor predicción del modelo metabólico.

#### **1.7.1.1 Método Comparativo**

El método comparativo tiene como base la inferencia de la presencia de reacciones metabólicas en modelo predicho, con respecto a una anotación funcional ya existente, donde se emplean las anotaciones de la base de datos de la EC (Moss, 2011) comparados contra las reacciones entre especies de otras bases de datos (Pitkanen *et al.* 2010; Aziz *et al.* 2008). Sin embargo, al realizar estas comparaciones, también se deben emplear métodos para encontrar y reparar reacciones de pérdidas o “gaps” que pueden generar modelos de reconstrucciones metabólicas inconsistentes (Karp *et al.* 2002; Kumar y Dasika, 2007). Por esta razón se utilizan herramientas que computan la probabilidad de que una reacción metabólica, tenga o no, una enzima dentro de un grupo de enzimas posibles, incorporando esta información al comparar con bases de datos de secuencias homologas (Pitkanen *et al.* 2010).

### **1.7.1.2 Método de Predicción Exploratorio**

Este método tiene como fundamento la búsqueda de bases de datos de referencia del metabolismo; estos métodos se dividen en dos clases: 1) gráfico-teórico y 2) métodos basados en restricciones (Pitkanen *et al.* 2010). Para reconstruir un metabolismo de un modelo biológico, el primer método busca identificar las rutas bioquímicas en las bases de datos de metabolismo, empleando solo la información que ha sido validada experimentalmente, creando una matriz estequiométrica de metabolitos y reacciones (Blum y Kohlbacher, 2008; Faust *et al.* 2009). Por otra parte, el segundo método, combina variables internas (ej: números de metabolitos disponibles para la reacciones bioquímicas) o externas (ej: introducción nutrientes) del sistema, para crear restricciones bioquímicas en el modelo reconstruido, cuya finalidad es la optimización de los resultados, en comparación a la realidad del comportamiento de los sistemas biológicos bajo condiciones específicas de cultivo en el laboratorio o en el ambiente (Planes y Beasley, 2009).

## **1.8 Análisis de Redes**

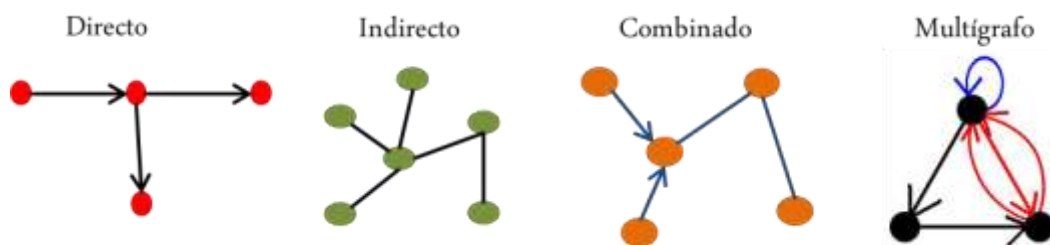
La teoría de grafos un campo de la matemáticas, está ligado habitualmente a la topología que se encarga del análisis estructural de un grafo. En este sentido para entender la información que se obtiene a través de un proceso de reconstrucción metabólica, se emplea el análisis topológico, con el fin de poder entender la estructura y dinámica en una red de relaciones, este proceso consiste en una descripción cuantificable de las redes que caracterizan varios sistemas biológicos (Barabási y Oltvai, 2004).

### **1.8.1 Clasificación de redes**

Las redes pueden ser clasificadas en Indirectas, Directas, Combinadas y Multígrafos (*multigraphs*, en inglés) (**Figura 1-5**). En las redes indirectas las conexiones entre los vértices no tienen dirección y biológicamente están representadas en interacciones de proteínas, redes filogenéticas y correlación de redes; por otra parte, las redes directas, presentan una sola dirección, por lo cual se visualizan como una línea con cabeza de flecha hacia la dirección del producto; los ejemplos en las redes biológicas están relacionados en redes metabólicas y redes de regulación de genes (Junker y Schreiber, 2008). Una gráfica mixta o combinada se caracteriza por tener vínculos directos e indirectos; esta grafica es relevante en biología; como ejemplo, se observa en las redes

de interacción de proteínas-proteínas donde algunas interacciones son indirectas (obtenidas por 2 experimentos híbridos) y otras son directas, representando activaciones como fosforilaciones y otras interacciones directas. Por otra parte, los multígrafos, son grafos que contiene múltiples aristas, que es 2 o más aristas que inciden a los mismos 2 vértices, este grafo se presenta biológicamente, cuando es útil representar el modelo de una vía metabólica donde las mismas sustancias pueden ser transformadas por diferentes reacciones (Junker y Schreiber, 2008).

**Figura 1-5:** Clasificación de las redes de acuerdo a la direccionalidad de sus conexiones.



## 1.8.2 Propiedades de las Redes

Al obtener la información de un proceso biológico y representarlo mediante una red, se pueden emplear los análisis de la teoría de redes o grafos, para la caracterización sistemática de sus propiedades estructurales (Ejemplo: diámetro, conectividad, centralidad entre otras). Esta aproximación nos permite estudiar las propiedades estructurales que subyacen a los principios de organización de los sistemas celulares, lo que refleja propiedades esenciales del sistema, tales como la robustez (Barabási y Oltvai, 2004).

### 1.8.2.1 Grados y conectividad

Una de las características más elementales de un nodo es el grado (degree en inglés) o conectividad,  $k$ , el cual cuantifica el número de conexiones de un nodo con otros nodos en la red. En redes directas, en las cuales se presentan una dirección de comunicación de los nodos dentro de la red, se puede mostrar los grados de entrada,  $k_{in}$ , el cual denota el número de aristas que se dirigen a un nodo, y un grado de salida,  $k_{out}$ , el cual denota el número de vértices que salen del nodo. En la redes indirectas, con  $N$  nodos y  $L$  aristas, es caracterizado por un grado promedio  $\langle k \rangle = 2L/N$  (donde  $\langle \rangle$  denota el promedio).

### **1.8.2.2 Distribución de Grados**

La distribución de grados,  $P(k)$ , da la probabilidad de que un nodo seleccionado tenga exactamente un número de aristas  $k$ . El  $P(k)$  es obtenido por conteo del número de nodos  $N(k)$  con  $k = 1, 2, \dots$  aristas y dividido por el número total de nodos  $N$ . Este parámetro de distribución permite distinguir entre diferentes clases de redes.

### **1.8.2.3 Path más corto**

La distancia entre dos nodos escogidos aleatoriamente en una red es una medida del número de aristas necesarias para conectar dichos nodos. De esta forma, dadas varias vías para conectar dos nodos, la más corta de estas vías es denominada el "path más corto" (*Shortest path*, en inglés), que es el *path* con el número más pequeño de aristas entre los nodos seleccionados. En las redes directas, la distancia de DAB del nodo A al nodo B es frecuentemente diferente de la distancia DBA del nodo B al A. Uno de los parámetros relacionados con los pasos más cortos, es el diámetro  $d_m = \max(d_{ij})$  de una red, que es definida como la máxima distancia de un par de nodos (Barabási, 2002).

### **1.8.2.4 Coeficiente de Clusterización**

Si un nodo A es conectado a B, y B es conectado a C, es altamente probable que A también tenga una arista o conexión con C. Este fenómeno puede ser cuantificado con el coeficiente de clustering,  $Cl = 2n_l / (k(k-1))$ , donde  $n_l$  es el número de aristas que conectan  $k$  vecinos de un nodo  $l$  a otro nodo. Valores de 1, indica que la conectividad es la más alta y valores de 0 la más baja (Barabási, 2002).

## Pregunta de investigación

¿Es posible identificar diferencias funcionales en tres metatranscriptomas edáficos, bajo diferentes estados fenológicos, mediante la reconstrucción y caracterización computacional de sus redes metabólicas?

## Objetivo general

Caracterizar las redes de metabolismo, a nivel topológico, de metatranscriptomas edáficos en tres diferentes estados fenológicos (emergencia, aporque y floración) para un ciclo de cultivo de papa criolla (*Solanum phureja*) bajo manejo orgánico, con el fin de establecer posibles diferencias funcionales.

## Objetivos específicos

- Identificar el conjunto de genes implicados en procesos metabólicos (genotipo metabólico) correspondiente a cada uno de los estados fenológicos (emergencia, aporque y floración).
- Establecer el conjunto de rutas metabólicas y reacciones asociadas al genotipo metabólico identificado en cada uno de los estados fenológicos.
- Reconstruir la red del metabolismo a nivel metatranscriptómico y generar un modelo computacional de esta reconstrucción.
- Caracterizar topológicamente cada una de las redes de metabolismo reconstruidas e identificar posibles características particulares en cada una de ellas, que lleven a la inferencia de propiedades biológicas particulares para cada uno de los estados fenológicos.

## 2. Metodología

Para el desarrollo del presente estudio se obtuvo suelo rizosférico de cultivos de papa criolla (*Solanum phureja*) bajo manejo orgánico de la finca Gabeno (4° 50' 36.82" Norte, 74° 08' 07.99" Oeste) localizada en el departamento de Cundinamarca, en el municipio de Tenjo (**Figura 2-1**). La finca se caracteriza por tener prácticas de fertilización y control de plagas, de acuerdo con métodos de agricultura biodinámica por más de veinte años. Las muestras de suelo rizosférico se colectaron para cada uno de los estados fenológicos propuestos (emergencia, aporque y floración), con espátulas estériles y empacado en tubos estériles, los cuales fueron congelados en N<sub>2</sub> líquido inmediatamente después de la toma, buscando preservar las muestras para la obtención de ARN de la comunidad de microorganismos de la muestra; los procedimientos de extracción de ARN se realizaron durante las siguientes 36 horas después de colectadas las muestras.

**Figura 2-1:** Ubicación del lugar de muestreo en el municipio de Tenjo, Cundinamarca.



## 2.1 Extracción del ARN

El ARN fue extraído de 0,5 g (peso seco) de suelo para cada uno de los estados fenológicos (**Figura 2-2**) mediante el kit RNA POWER SOIL® (Epicentre), al que se adecuaron fases de lisis y purificación de acuerdo a Chen *et al* (2008), seguido de una extracción orgánica con cloroformo alcohol isoamílico y la precipitación con isopropanol. El ARN total extraído se sometió a una metodología de enriquecimiento de ARNm, en la cual el ARN total aislado y purificado, se somete a una hidrólisis de los extremos 5' monofosfato de las secuencias de los ARNr, utilizando una exonucleasa comercial (mRNA ONLY™ Prokaryotic mRNA Isolation Epicentre). Posteriormente y considerando que los ARNm representan una pequeña fracción del ARN total, se requirió de un método de amplificación, para lo cual se empleó un método que se ha desarrollado para amplificar ARNm eucariota a través de la T7 RNA polimerasa.

**Figura 2-2:** Desarrollo esquemático de la obtención de muestras de suelo rizosférico de papa criollo (*Solanum phureja*) hasta la secuenciación.



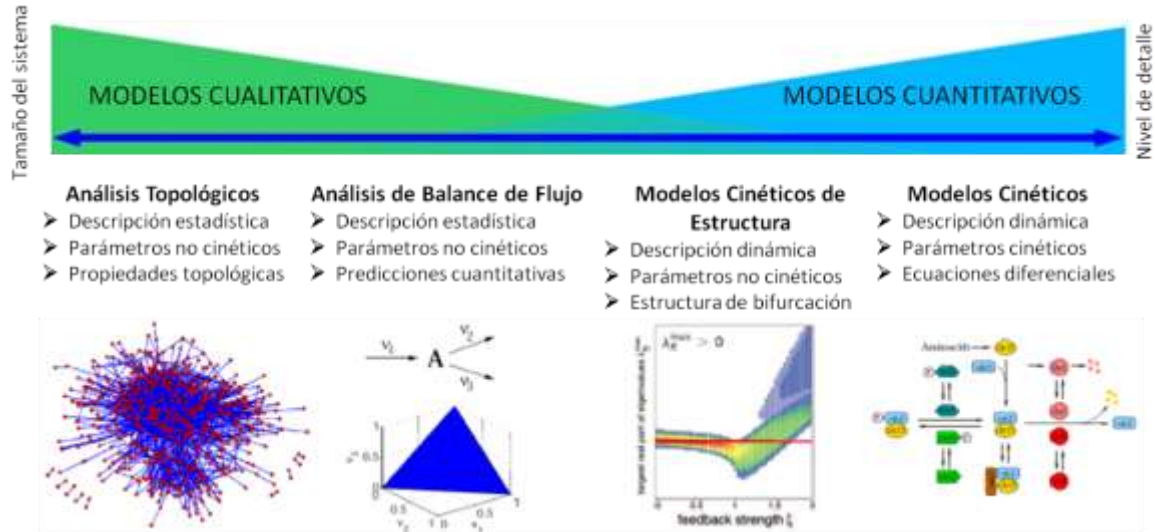
Se tomaron los transcriptos bacterianos y se poliadenilaron utilizando la poli(A) polimerasa de *Escherichia coli*, disponible comercialmente (Poly(A) Polymerase Tailing Kit Epicentre), de acuerdo con lo reportado para el análisis del metatranscriptoma aislado de océanos (Frias *et al.* 2008). Se ha probado por tecnología de microarreglos que la amplificación de ARN comunitario por esta metodología presenta la ventaja de mantener la abundancia relativa original de los transcritos (Frias *et al.* 2008; Gao *et al.* 2007). Después del procedimiento anterior se continuó con la síntesis de la primera cadena de ADNc a partir de los transcritos poliadelinados con PowerScript™ Reverse Transcriptase (Invitrogen), utilizando como cebadores oligo (dT) y bajo las condiciones de reacción recomendadas por el fabricante; posteriormente de obtenido el ADNc se siguió con la realización de la secuenciación, la cual se realizó por medio de la tecnología Illumina Solexa 1G Genome Analyzer (**Figura 2-2**).

## 2.2 Procesamiento de datos

Dado que los procesos celulares poseen una inherente complejidad, describir estos procesos en un solo modelo matemático no necesariamente sea la mejor estrategia metodológica, más bien, las representaciones matemáticas del metabolismo celular pueden tener muchas facetas, que van desde las descripciones puramente topológicas o estequiométricas a los modelos cinéticos mecanicistas de vías metabólicas (Steuer, 2007). La variedad de diferentes representaciones del metabolismo corresponden a los diferentes niveles de detalle y de información que se disponga, de ahí que una posible forma de representar las diferentes metodologías aplicadas al estudio de redes sea la que se observa en la (**Figura 2-3**).

En este sentido, el análisis topológico de la red tiene un número de ventajas considerables, en comparación con la construcción de modelos cinéticos explícitos, ya que un análisis topológico no presupone ningún conocimiento de los parámetros cinéticos, permitiendo así un análisis de organismos o datos poco caracterizados. Esta aproximación metodológica facilita el estudio de grandes sistemas, que consta de varios miles de nodos, mucho más allá del ámbito de los modelos cinéticos actuales, lo que contribuye a una mejor comprensión de la arquitectura de la red metabólica.

**Figura 2-3:** Aproximación esquemática de los diferentes enfoques metodológicos que se están empleando para el modelamiento y estudio de las redes de metabolismos, en donde se involucran diferentes niveles de detalle y complejidad, resaltando enfoques que son más cualitativos y otros más cuantitativos. Modificado de Steuer, 2007.



En virtud de lo señalado, y considerando que un metatranscriptoma es una muestra compleja poco caracterizada, el presente estudio aborda un enfoque metodológico de reconstrucción metabólica a nivel computacional y análisis topológico, para realizar una descripción global de los metatranscriptomas edáficos obtenidos para los tres estados fenológicos estudiados, procediendo de la forma como se describe a continuación para el tratamiento datos.

### 2.2.1 Procesamiento de los metatranscriptomas

Las secuencias obtenidas para los metatranscriptomas de cada uno de los estados fenológicos (emergencia, aporque y floración), se sometieron a un procedimiento de limpieza inicial, que consistió en eliminar las secuencias de baja calidad y la presencia de homopolímeros, así como las secuencias correspondientes a los adaptadores empleados para los procedimientos de secuenciación y los empleados en la metodología de obtención de ADNc. Para la realización de estos procedimientos, se emplearon scripts realizados bajo el lenguaje de programación perl.

Adicionalmente, al tratarse de metatranscriptomas, la existencia de ARNr debe ser eliminada para poder contar con el potencial de expresión de los ecosistemas edáficos

en estudio. En este sentido, las secuencias se sometieron a un procesamiento con la herramienta informática de RiboPicker (Schmieder *et al.* 2011), que genera alineamientos globales usando transformaciones de Burrows Wheeler para indexar las secuencias de las bases de datos de referencia. En este caso se usaron los parámetros estándares y la base de datos no redundantes de ARNr (<ftp://edwards.sdsu.edu:7009/ribopicker/db/>) que contiene secuencias de ARNr 16S, 18S, 23S, 28S, 5S y 5.8S de: Silva, GreenGenes, RDP, Rfam y secuencias ribosomales de NCBI, compiladas en un sólo aplicativo, obteniéndose para cada uno de los estados fenológicos un conjunto de secuencias de ARN sin secuencias ribosomales (ARNm) y otro con las secuencias ribosomales (ARNr).

### **2.2.2 Reducción de redundancia y obtención del Genotipo Metabólico**

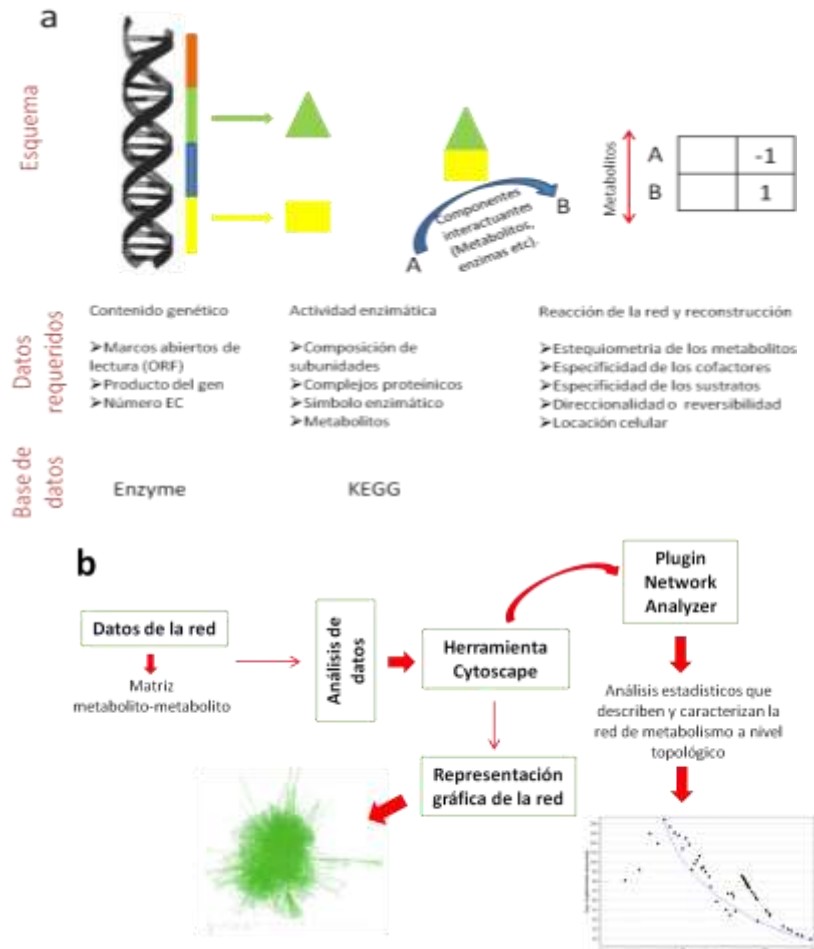
Dado que un mismo gen puede estar representado varias veces a nivel de secuencia en cada uno de los metatranscriptomas y con el fin de eliminar la redundancia presente, se tomó el conjunto de secuencias que correspondían solo a los ARNm para los diferentes estados fenológicos, y se sometieron a un proceso de alineamiento en grupos (clústeres), con un nivel de similaridad mínimo del 97% usando el programa CD-HIT (Li y Godzik, 2006). Realizado el procedimiento anterior, se tomó un conjunto de secuencias representantes (semilla) para cada uno de los estados fenológicos, los cuales fueron empleados para los análisis subsiguientes.

Con el conjunto de secuencias representantes, se procedió a identificar el genotipo metabólico, esto se realizó a través de la búsqueda de similitud de estas secuencias contra la base de datos Enzyme (Moss, 2011), la cual fue procesada mediante un protocolo similar al utilizado por el sistema PRIAM (Claudel-Renard *et al.* 2003), y almacenada como una base de datos local de perfiles. Los metatranscriptomas completos fueron comparados contra esta base de datos usando RPS-BLAST ([file:///C:/Users/Jose/Downloads/summer00%20\(2\).pdf](file:///C:/Users/Jose/Downloads/summer00%20(2).pdf)) (siglas en inglés Reverse Position Specific BLAST), el genotipo metabólico se asignó con base en el mejor hit y cobertura, evidenciado por el valor E.

### 2.2.3 Identificación de las rutas metabólicas

Identificado el genotipo metabólico, cada miembro del genotipo fue mapeado en rutas metabólicas usando como referencia la base de datos KEGG (sigla en inglés para Kyoto Encyclopedia of Genes and Genomes) (Kanehisa *et al.* 2006). Esta base de datos posee una colección de reacciones metabólicas que han sido demostradas existen en un gran número de diferentes organismos; información que en la mayoría de los casos está ligada a un número EC (siglas en inglés enzyme commission) o un número TC (siglas en inglés transport commission). Lo anterior permitió determinar que reacciones bioquímicas están siendo llevadas a cabo por cada una de las enzimas asociada a la anotación de los diferentes genes en los metatranscriptomas. Las reacciones mapeadas para cada ruta metabólica, se descargaron y almacenaron en una base de datos local (**Figura 2-4 a**).

**Figura 2-4:** Esquema metodológico para la reconstrucción de un red metabólica. a) Pasos realizados desde las secuencias hasta la realización de la matriz estequiométrica. b) Procedimiento para la realización del análisis de la red de metabolismo a nivel topológico.



### 2.2.4 Reconstrucción de la red del metabolismo

Las reacciones almacenadas se curaron manualmente asegurando que no existiera redundancia o fallos en su estructura. Después de la curación manual, se hizo necesario convertir esta información en un modelo matemático, que permitiera delinear las interacciones de los diferentes componentes celulares y sus reacciones bioquímicas, dado esto, las reacciones fueron expresadas mediante una matriz estequiométrica, obteniendo así un formato de lectura computacional para análisis **(Figura 2-4 a)**.

La matriz estequiométrica se cargó en el paquete de análisis MATLAB ([www.mathworks.com/products/matlab](http://www.mathworks.com/products/matlab)), y mediante el paquete COBRA ToolBox (Becker *et al.* 2007), se exportó como un grafo bipartito del tipo metabolito-reacción. No obstante, debido a la dificultad que representa el análisis de grafos bipartitos, dicho grafo fue procesado mediante scripts locales y expresado como una red metabolito-metabolito, para entender mejor el fenómeno presente en cada uno de los metatranscriptomas.

### 2.2.5 Análisis de Topología

Se tomó la matriz metabolito-metabolito y se cargó en la herramienta Cytoscape (Shannon *et al.* 2003), con el fin de realizar la representación gráfica de la red y poder realizar los respectivos análisis. A través de un conjunto de algoritmos implementados en la herramienta Network Analyzer se realizaron los análisis de las redes biológicas y calcularon los parámetros de la topología de la red, incluyendo el diámetro de la red, número de vecinos, números de pares de nodos conectados, grados de los nodos, coeficientes de clustering, y la trayectoria más corta **(Figura 2-4 b)**.

## 3.Resultados y discusión

### 3.1 Preprocesamiento de los Datos e Identificación del Genotipo metabólico

El ARN total fue extraído de las muestras de suelo rizosférico para cada uno de los estados fenológicos seleccionados (emergencia, aporque y floración). Las tres muestras fueron secuenciadas en una línea indexada de la plataforma Illumina HiSeq 2000 empleando el sistema de secuenciación de ambos extremos 3' y 5' (paired-end sequencing). La calidad fue evaluada a partir del Q-score (Phred Quality score) (Ewing, *et al.* 1998) de las bases para cada lectura, donde se identificó globalmente una menor calidad en la secuencia del extremo 5' del fragmento, con relación a la secuencia del extremo 3'. Los diagramas de caja de los puntajes Phred de calidad generados por FastQC se pueden apreciar en la **Figura A-1** y el **AnexoA**. En virtud de lo anterior, se decidió trabajar con las librerías correspondientes al extremo 3'.

Con el fin de eliminar la alta variabilidad en la composición de nucleótidos al inicio de las lecturas, las librerías del extremo 3' se procesaron, eliminando las secuencias de baja calidad, la presencia de homopolímeros, las secuencias correspondientes a los adaptadores empleados para los procedimientos de secuenciación y los empleados en la metodología de obtención de ADNc, así como una selección de las lecturas más largas de 70 bases. Este procesamiento permite eliminar algunos patrones que se han observado en diferentes experimentos transcriptómicos (Mortazavi *et al.* 2008; Bullard *et al.* 2010), en donde el uso de hexámeros aleatorios en el proceso de amplificación de la muestra produce y a demás distorsiona el proceso de mapeo sobre esta zona (Hansen *et al.* 2010). Después de procesadas se obtuvieron 43.866.300, 98.613.897 y 91.221.872 secuencias para emergencia, aporque y floración respectivamente. Los conteos de lecturas tras cada paso del preprocesamiento se muestran en la **Tabla 3-1**.

Durante el proceso de remoción de las secuencias ribosomales con el software RiboPicker, se observó un incremento en la cantidad de ARNm recuperado para los tres diferentes metatranscriptomas, ya que en promedio en este tipo de muestras se espera una concentración de ARNr mayor al 95% del ARN total (Peano *et al.* 2013). No obstante, los resultados obtenidos en el presente trabajo muestran que el procedimiento de enriquecimiento no fue tan eficiente, ya que tan solo se logró recuperar un 22%, 32%, y 39% de ARNm del ARN total, para emergencia, aporque y floración respectivamente (**Tabla 3-1**), en comparación con estudios de la misma naturaleza donde han reportado una contaminación de ARNr del 53% (Frias-Lopez *et al.* 2008) y 37% (Poretsky *et al.* 2009), después de haber realizado un procedimiento de enriquecimiento de ARNm igual al empleado en este estudio.

**Tabla 3-1:** Número de lecturas por librerías tras cada paso de preprocesamiento.

Muestra	Lecturas Iniciales	Lecturas Editadas	Lecturas sin Ribosomales	Lecturas Ribosomales	% Ribosomales	Lecturas Clusterizadas	% de secuencias para los análisis
Emergencia	78.018.750	43.866.300	9.650.586	34.215.714	78	3.120.750	4
Aporque	119.285.714	98.613.897	28.687.679	69.926.218	71	8.350.000	7
Floración	124.444.444	91.221.872	35.576.530	55.645.342	61	11.200.000	9
# Total de Lecturas	321.748.908	233.702.069	73.914.795	159.787.274		22.670.750	

Después de la remoción de ribosomales, el conjunto de secuencias de ARNm obtenido para cada uno de los estados fenológicos, se alinearon en grupos (clústeres) con un nivel de similitud mínimo del 97%, lográndose obtener la mayor cantidad de información pero un mínimo volumen de datos para analizar. En este set de secuencias el porcentaje de las lecturas que permanecieron tras la clusterización para ser los datos de entrada en la identificación del genotipo metabólico fue del 32%, 29% y 31% para emergencia, aporque y floración respectivamente (**Tabla 3-1**), sugiriendo un promedio del 70% de redundancia alcanzado con la plataforma Illumina.

Los resultados anteriores comparados con el conjunto de datos iniciales arrojados por el proceso de secuenciación, ponen de manifiesto que del conjunto de secuencias obtenidas, tan solo el 4%, 7% y 9% para emergencia, aporque, y floración respectivamente (**Tabla 3-1**), fueron los datos disponibles para el análisis. Una de las razones que puede explicar esto, se debe a los bajos niveles de ARNm que se

obtuvieron del proceso de enriquecimiento, que hizo que un alto porcentaje de las secuencias presentes en las librerías de los metatranscriptomas correspondiera a ARNr el cual se descartó para los análisis aquí realizados.

Los metatranscriptomas tienen la características de que las secuencias presentes sean marcos abiertos de lecturas (ORF siglas en inglés), ya que hacen parte de ARNm convertido en ADNc, razón por la cual, las secuencias procesadas y seleccionadas, son secuencias que pertenecen al potencial de expresión presente en la comunidad de microorganismos del ecosistema edáfico bajo estudio. Con esta premisa, se procedió a obtener el genotipo metabólico, el cual corresponde a todas las reacciones químicas que un organismo puede catalizar a través de enzimas codificadas en su genoma (Oberhardt *et al.* 2009). Las enzimas en su mayoría son proteínas, las cuales con frecuencia contienen muchos dominios, cada uno con diferentes funciones. Tales dominios han evolucionado como módulos que son combinados en diferentes arreglos para producir proteínas de función específica. Estos dominios son útiles para realizar la anotación de las secuencias a partir de la identificación de los motivos proteicos, que permiten encontrar homologías con interés biológico funcional o estructural.

Una de las formas a través de las cuales hoy día se encuentran las homologías, es la utilización de base de datos precalculadas de matrices de puntuaciones por posiciones (PSSM siglas en inglés) (Altschul *et al.* 1997), y con el programa RPS-BLAST (Reverse Position-Specific BLAST siglas en inglés) identificar dominios conservados en una secuencia de proteína. En este sentido, se tomaron las secuencias seleccionadas de cada uno de los metatranscriptomas para los diferentes estados fenológicos y se anotaron, empleando el programa RPS-BLAST y comparando las secuencias contra una base de datos local, de dominios conservados construida a partir de los dominios presentes en la base de datos Enzyme. La anotación de los metatranscriptomas se realizó asignando a los genes la categoría enzimática de acuerdo al número EC.

A partir del listado de los números EC, se obtuvo un conjunto de identificadores únicos que permitió indicar como los productos de los genes interactúan catalizando reacciones metabólicas. No obstante, dada la naturaleza del conjunto de datos de un metatranscriptoma, el cual representa a gran escala los diferentes potenciales de expresión de los microorganismos presentes en el ecosistema edáfico, se esperaba

redundancia en los resultados obtenidos para las reacciones enzimáticas; por lo cual, se realizó una curación basada en la literatura, con información experimental que permitió obtener la anotación de 2.529.207, 2.752.334 y 2.680.937 secuencias, representadas en 1105, 1253 y 1209 funciones enzimáticas para emergencia, aporque y floración respectivamente (**Tabla 3-2**).

**Tabla 3-2:** Número de secuencias anotadas y funciones enzimáticas encontradas para cada uno de los estados fenológicos, las cuales representan el genotipo metabólico de cada uno de los metatranscriptomas.

Clase enzimática principal	Nombre	Emergencia		Aporque		Floración	
		# de secuencias anotadas	# de funciones enzimática por muestra	# de secuencias anotadas	# de funciones enzimática por muestra	# de secuencias anotadas	# de funciones enzimática por muestra
EC 1	Oxidoreductasa	740445	312	845675	386	796327	352
EC 2	Transferasa	865646	342	858273	340	849504	344
EC 3	Hidrolasas	338995	199	369104	217	343564	211
EC 4	Liasas	288976	114	363470	156	369357	153
EC 5	Isomerasas	132474	60	147111	72	160274	72
EC 6	Ligasas	162671	78	168701	82	161911	77
<b>Total de secuencias anotadas a una función enzimáticas</b>		2529207	1105	2752334	1253	2680937	1209

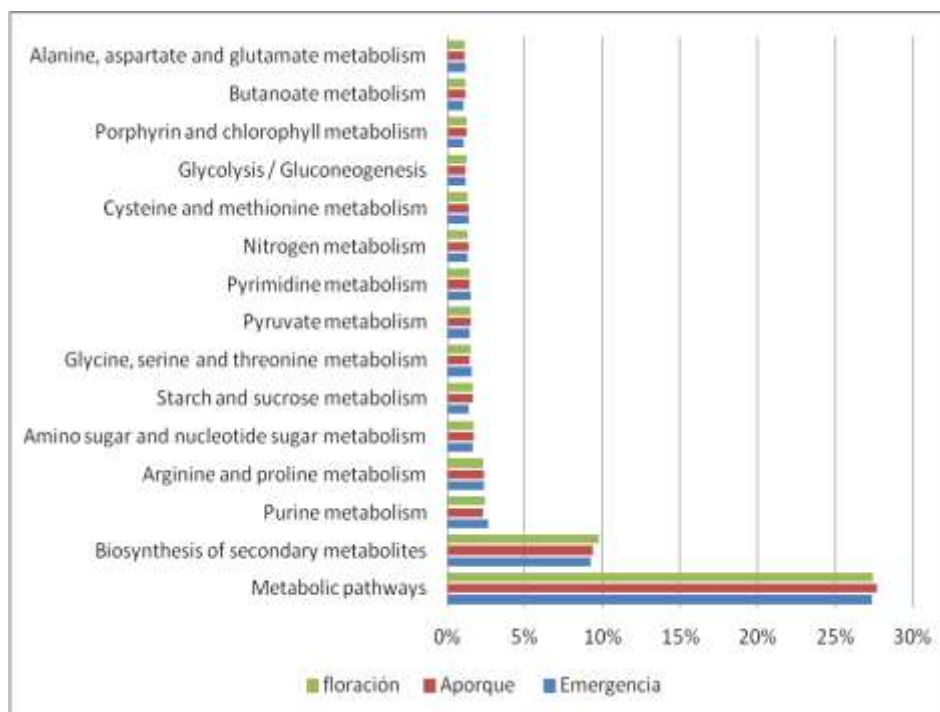
### 3.2 Identificación de las rutas metabólicas

Para indicar como se encontraba la red de reacciones químicas catalizadas por las enzimas, y conectadas vía sus sustratos y productos, para cada uno de los estados fenológicos, se tomaron los perfiles metabólicos y se mapearon a través de una base de datos de rutas metabólicas, con el fin de identificar como éstas enzimas se agrupaban en procesos metabólicos descritos de forma tradicional en rutas metabólicas en las cuales se encuentran rutas bioquímicas y sus componentes (enzimas, sustratos, productos) y las correspondientes interacciones y reacciones químicas, como la glicólisis, ruta de las pentosas fosfato entre otras.

A partir de las 1105, 1253 y 1209 funciones enzimáticas obtenidas para emergencia aporque y floración respectivamente, se logró determinar la asociación de éstas reacciones con las rutas metabólicas en la base de datos KEGG (Kanehisa *et al.* 2006), identificando la existencia de 155, 156 y 158 rutas metabólicas para emergencia, aporque y floración respectivamente (**Tabla A-1 AnexoA**). La mayoría de los transcriptos

anotados se relacionaron con genes del metabolismo central (**Figura 3-1**), sugiriendo para los tres estados fenológicos un conjunto de funciones particularmente similar; esto suele ser común en este tipo de aproximaciones por sus características holísticas, donde se observan patrones globales por la naturaleza de la metodología empleada (Urich *et al.* 2008)

**Figura 3-1:** Las 15 rutas KEGG más abundantes para cada uno de los metatranscriptomas para los diferentes estados fenológicos.



No obstante, al analizar los resultados de las rutas metabólicas encontradas para cada uno de los estados fenológicos, en relación de las 11 categorías funcionales (**Tabla 3-3**), de la base de datos KEGG (Kanehisa *et al.* 2006), el metabolismo de Carbohidratos represento 24%, 25% y 26% para emergencia aporque y floración, Aminoácidos 22% en las tres muestras, Lípidos 10%, 9% y 9% para emergencia aporque y floración y Biosíntesis de Vitaminas 9% en las tres muestras, siendo éstas las funciones que presentaron un mayor porcentaje de funciones enzimáticas asociadas. En la categoría del metabolismo de carbohidratos se identificaron más reacciones en procesos relacionados con el metabolismo de almidón y sacarosa (21,31%), y el metabolismo de

Galactosa (21, 84%). En el metabolismo de energía se identificó en la fosforilación oxidativa (51,06%) y en el metabolismo de metano (23,04%).

**Tabla 3-3:** Representación de cada una de las categorías funcionales del KEGG para los diferentes estados fenológicos.

CATEGORÍA METABÓLICA KEGG	EMERGENCIA	APORQUE	FLORACIÓN
1.1 Metabolismo de carbohidratos	24%	25%	26%
1.2 Metabolismo energético	6%	6%	6%
1.3 Metabolismo de lípidos	10%	9%	9%
1.4 Metabolismo de nucleótidos	7%	6%	6%
1.5 Metabolismo de aminoácidos	22%	22%	22%
1.6 Metabolismo de otros amino ácidos	5%	5%	5%
1.7 Metabolismo y biosíntesis de glicanos	5%	3%	3%
1.8 Metabolismo de cofactores y vitaminas	9%	9%	9%
1.9 Metabolismo de terpenoides y policétidos	3%	3%	3%
1.10 Biosíntesis de otros metabolitos secundarios	4%	4%	4%
1.11 Metabolismo de biodegradación de xenobióticos	5%	7%	7%

En el metabolismo de Lípidos se encontró la mayoría de reacciones implicadas en el metabolismo de glicerofosfolípidos(17,64%) y glicerolípidos (17,24%). En la categoría del metabolismo de nucleótidos se observó en el metabolismo de purinas (48,70%) y pirimidinas (51,29%). En el metabolismo de aminoácidos, se observó una distribución relevante en el metabolismo de arginina y prolina (13,78%), metabolismo de glicina, serina, treonina (12,72%) y degradación de valina, leucina e isoleucina (10,60%).

En la categoría funcional donde se relacionan otros aminoácidos, se observó el metabolismo glutatiónina (20,83%), seguido del metabolismo de ácido cianoamino (16,66%) y beta-alanina (16,66%). Por otro lado, en el grupo de biosíntesis de glicanos, se observó con una mayor distribución el metabolismo de Glicoesfingolípidos biosíntesis (27,655), biosíntesis N-Glicano (21,27%) y de degradación de glicosaminoglicano (19,14).

Algunas reacciones se identificaron en el grupo de cofactores y vitaminas, como el metabolismo porfirina y clorofila (27,24%) y folato (15,53%). Por otro lado, se observó el metabolismo de terpenoides (35,18%) dentro de la misma categoría, seguido de la degradación de geraniol (14,81%). En el metabolismo de metabolitos secundarios, presentó una mayor distribución, el metabolismo de biosíntesis de fenilpropanoides

(28,76%), seguido de la biosíntesis de estreptomicina (15,06%). En la biodegradación de xenobióticos, se identificó el metabolismo de xenobióticos por citocromo P450 (30,26%), seguido de la degradación de benzoato por ligación de CoA (15,78%).

### 3.3 Reconstrucción de las redes y su caracterización topológica

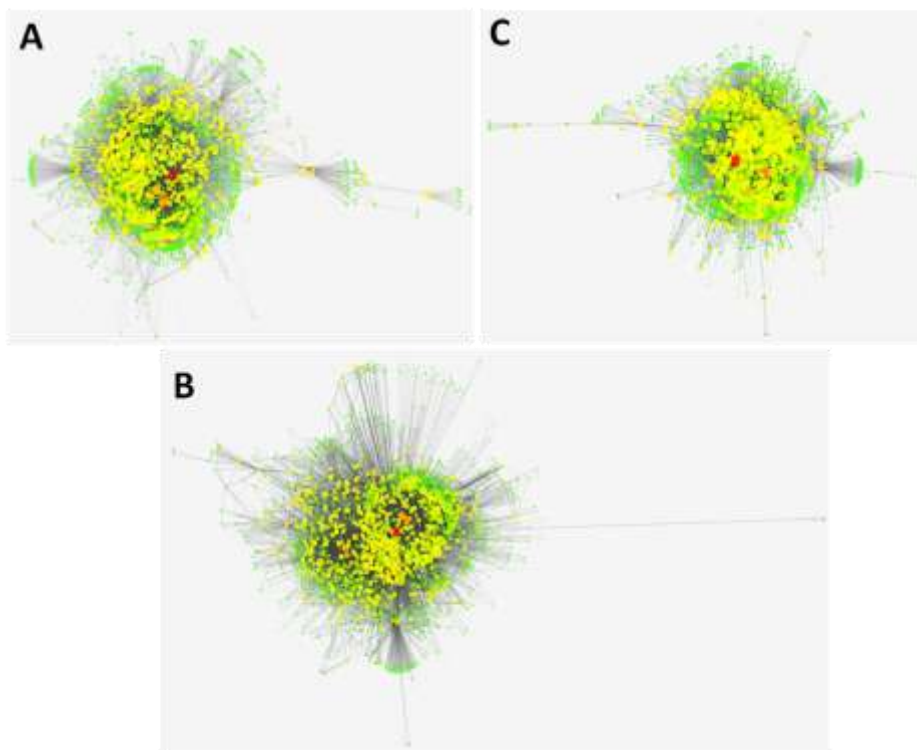
Descrita la estructura funcional de cada uno de los metatranscriptomas en los diferentes estados fenológicos, se procedió a representar el conjunto de reacciones en una matriz llamada matriz de estequiometría. Esta matriz facilita tener un modelo del metabolismo, donde los componentes son los coeficientes estequiométricos de las reacciones, permitiendo lograr abordar el balance de metabolitos (Edwards y Palsson, 2001). Los modelos se basan en balances de masa de intermediarios metabólicos suponiendo un estado estacionario. A partir de la información obtenida desde la estequiometría, se logra un modelo de precursores o bloques de construcción hacia la producción de biomasa. Aunque la composición de biomasa cambia bajo distintas condiciones fisiológicas, puede suponerse constante, puesto que se ha observado *in-silico* que cambios en la composición de biomasa tienen impactos muy leves en los resultados de simulaciones (Varma *et al.* 1993).

A partir de la de la matriz de estequiometría se obtuvo un archivo XML en el cual se encontraban las indicaciones para la representación del metabolismo como un grafo bipartito del tipo metabolito-reacción. No obstante, debido a la dificultad que representa el análisis de grafos bipartitos, dicho grafo fue procesado mediante scripts locales y expresado como una red metabolito-metabolito. EL archivo metabolito-metabolito se cargó en el software Cytoscape (Shannon *et al.* 2003), y se obtuvo una representación gráfica para cada uno de los metatranscriptomas (**Figura 3-2**). Realizada la representación gráfica, se empleó el plugin Network Analyzer, con el que se obtuvo los análisis topológicos.

Dentro de las redes biológicas las características topológicas más ampliamente estudiadas ha sido su Grado de Distribución, Característica de la distancia del Path y las Propiedades de Clusterización (Goh, 2005). Las redes de escala libre se caracterizan por presentar un alto grado de clusterización y una baja distancia del Path, que en el caso de

redes metabólicas, Jeong *et al.* (2000) calculó que el promedio de la distancia del path para 43 organismos estudiados fue de 3,2. Lo anterior significa que los metabolitos pueden ser convertidos a otros en aproximadamente 3 pasos, ya que es la distancia promedio más corta, para que un par de metabolitos se conecten.

**Figura 3-2:** Modelo escala libre para a) Emergencia, b) Aporque y c) Floración, en cada uno de los modelos las zonas rojas indican los nodos más conectados y las zonas verdes son las menos conectadas.



Los análisis topológicos obtenidos para los tres metatranscriptomas en los diferentes estados fenológicos (**Tabla 3-4**), evidenciaron (**Figura B-1,2.3 y AnexoB**) que los modelos tienen un comportamiento de red Libre de Escala (**Figura 3-2**), modelo Barabási–Albert. Lo anterior, en la base de que los resultados muestran que existe un alto grado de clusterización, considerando que el rango de distribución de los valores del coeficiente de clusterización se encuentra entre 0 y 1, y en el presente trabajo se encontró valores entre 0,8 o por encima para los tres metatranscriptomas indicando una alta agrupación de los nodos.

Por otro lado, el promedio de distancia del path y el diámetro de la red, han sido asociadas a un mismo orden de magnitud, propiedad que ha sido previamente descrita para una variedad de redes biológicas, tecnológicas y sociales (Jeong *et al.* 2000; Albert y Barabasi, 2002), y es con frecuencia referido como propiedad de redes de mundo pequeño (Watts y Strogatz, 1998).

**Tabla 3-4:** Resultados obtenidos para los parámetros de análisis de topología de redes realizado a cada uno de los metatranscriptomas para cada uno de los estados fenológicos en Cytoscape.

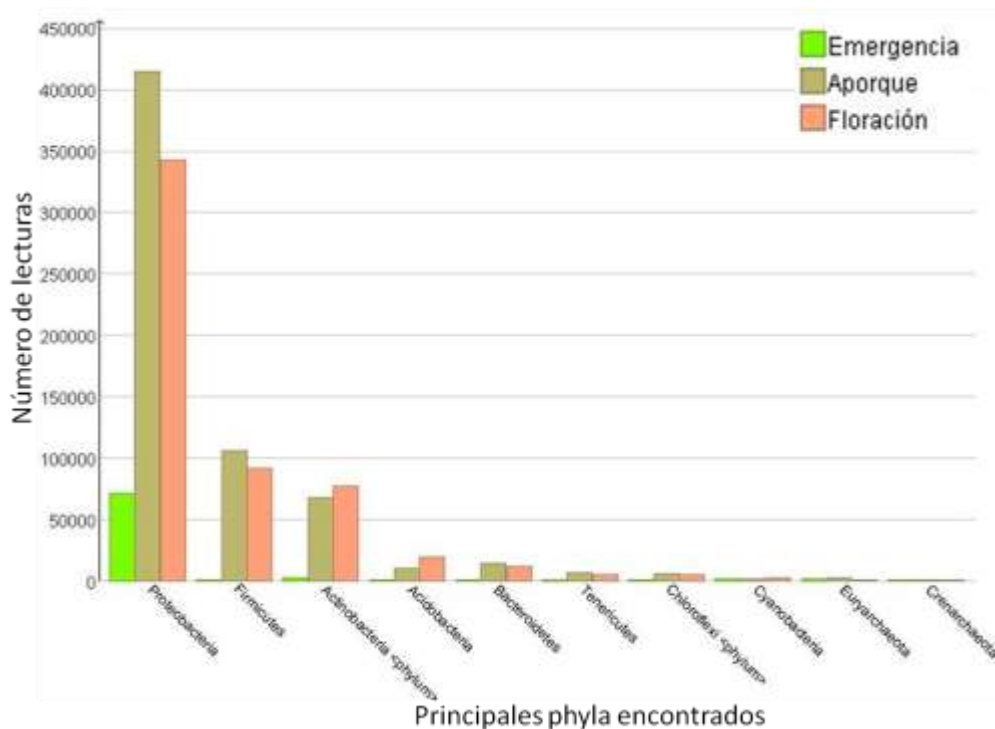
Parámetros de Análisis	Emergencia	Aporque	Floración
Coeficiente de Clusterización	0,8	0,83	0,8
Diámetro de la Red	8	7	8
Centralización de la Red	0.537	0,54	0,53
Path más corto	6172796(97%)	6741936(96%)	624082(94%)
Promedio de # Vecindad	9.803	10	9,861
Densidad de la Red	0.0039	0,004	0,004
Heterogeneidad	4.27	4,478	4,380
Número de nodos	2519	2648	2604
Número de vértices	29099	31556	30418

No obstante, el diámetro encontrado en este trabajo para las redes reconstruidas de los metatranscriptomas estuvo por encima (8) del rango reportado para el promedio de distancia del path (Jeong *et al.* 2000); este fenómeno puede ser explicado en el hecho de que se está trabajando con una muestra compuesta, en la cual se encuentran una gran cantidad de organismos de diferentes grupos taxonómicos, incrementando el número de enzimas y sustratos que participan en la red biológica con diferentes niveles de especialización, lo cual, contribuiría a aumentar el diámetro de la red; sin embargo, cabe señalar que el incremento de la red se ha discutido como una estrategia o una habilidad, que puede llegar a tener un sistema biológico para responder eficientemente a los cambios externos o errores internos normales en un proceso de modulación evolutiva en un ecosistema (Jeon *et al.* 2000). En este sentido, se ha observado que existen claras diferencias en la estructura de las redes de los tres dominios de organismos, resaltándose que Eucariotas y Archeas tienen promedios de distancia del path más altos que en las Bacterias (Hongwu y Zeng, 2003).

Por otro lado, se realizó una descripción de la composición taxonómica con el fin de conocer la dinámica de los microorganismos en el ecosistema edáfico, para lo cual, se empleó el Software MEGAN 5.0 (Huson 2007). Se tomaron las secuencias de cada uno

de los metatranscriptomas para los diferentes estados fenológicos y se anotaron a través de un BLAST contra la base de datos no redundante de microorganismos del NCBI, resaltándose que los grupos más representados en los tres metatranscriptomas fue el phylum de las Proteobacterias (**Figura 3-3**), así mismo para aporque y floración existió una alta representación de Firmicutes y Actinobacterias (**Figura 3-3**).

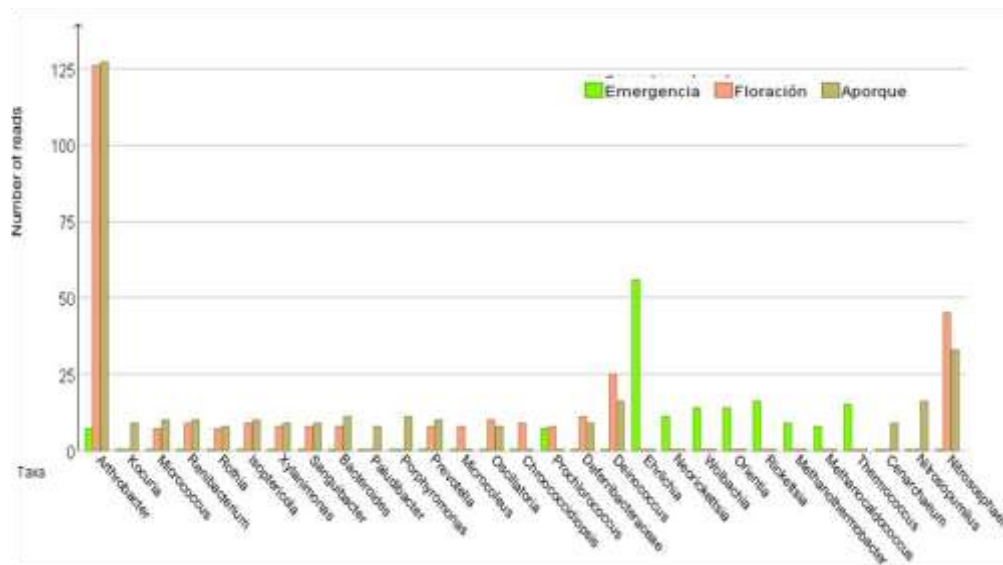
**Figura 3-3:** Principales phyla encontrados en los diferentes metatranscriptomas para cada uno de los estados fenológicos de acuerdo al software MEGAN 5.0 en relación a los transcritos o lecturas asociadas.



La alta representatividad de Proteobacteria y Actinobacteria a través del crecimiento vegetal ha sido previamente observada en la composición de la comunidad microbiana de la rizósfera de plantas (Bulgarelli *et al.*, 2012; Lundberg *et al.*, 2012). Dentro de las Proteobacterias las Alphaproteobactrias fueron las que mayor número de secuencias tuvieron asociadas. Estos resultados son consistentes con reportes en los que han estudiado el efecto del desarrollo de las plantas sobre la comunidad microbiana, y han reportado que el efecto no se observa sobre la estructura global de la comunidad, sino que los cambios se observan sobre microorganismos específicos a través del desarrollo vegetal (Chaparro *et al.* 2014). En contraste, se observa que emergencia en relación a aporque y floración, presentan diferencias en el número de secuencias asociadas a las

grupos identificados, existiendo una tendencia más consistente entre aporque y floración, respecto a emergencia, resultados similares se han observado en la rizósfera de plantas de papa, en donde por medio de técnicas dependientes de cultivo, la rizósfera de plantas jóvenes muestran el mismo patrón pero este desaparece a medida que la planta se torna más adulta. (Inceoglu *et al.* 2011).

**Figura 3-4:** Principales generos encontrados en los diferentes metatranscriptomas, para cada uno de los estados fenológicos de acuerdo al software MEGAN 5.0 en relación a los transcritos o lecturas asociadas.

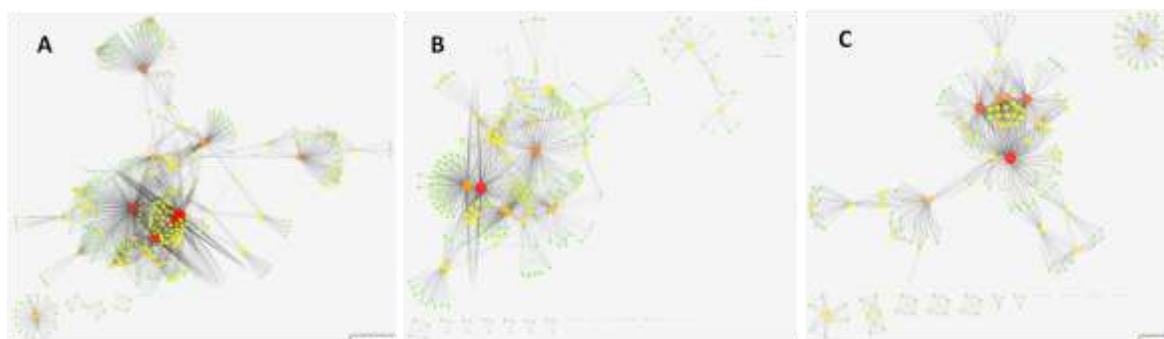


Sin embargo, cuando se analiza la composición de microorganismos a nivel de géneros, se observa que existen diferencias entre los metatranscriptomas, para los diferentes estados fenológicos en relación a los géneros existente en cada uno de los estados fenológicos (**Figura 3-4**). Se observa que emergencia tiende a tener una composición de microorganismos diferentes a aporque y floración; en emergencia dentro de los géneros más representados se resalta *Rhizobium*, *Starkeya*, *Ehrlichia*, *Propionibacteria* y *Pseudomonas*; En el caso de aporque y floración compartieron los géneros *Rhodospirillum*, *Clostridium*, *Heliobacterium*, *Burkholderia*, *Acinetobacter*, y *Pseudomonas* (**Figura 3-5**), estos últimos tienden a presentar patrones mas similares en relación a la composición de microorganismos al nivel de género, lo anterior se relaciona con los resultados obtenidos por Micallef *et al.* (2009), donde reporta que las comunidades microbianas de *Arabidopsis* varían en composición de géneros respecto al estado de desarrollo de la planta, pero esta diferencia decrece con el aumento de la edad



prácticas orgánicas, y la degradación de la lisina se asocia a la descomposición de materia orgánica, el hecho de que se haya fuertemente representado en esta fase puede estar relacionado a las prácticas de cultivo, en donde suelen emplear abonos con alto contenido de materia orgánica en estas etapas. Adicionalmente, Ciertos aminoácidos (glutamato, treonina, serina, cisteína y arginina), a elevadas concentraciones ( $10^{-2}$  a  $10^{-3}$  M) atraen a ciertas especies de género *Pseudomonas*, como por ejemplo, *P. lachrymans* (Chet *et al.* 1973) y *P. aeruginosa* (Nikata *et al.* 1992).

**Figura 3-6:** Subredes obtenidas para los tres metatranscriptomas A) Emergencia, B) Aporque y C) Floración, en cada uno de los modelos las zonas rojas indican los nodos más conectados y las zonas verdes son las menos conectadas.



Se ha demostrado por Vílchez *et al.* (2000), que los aminoácidos constituyen una fracción importante del total de los exudados de raíz, por ejemplo en *Brassica napus* pueden alcanzar el 60- 85% del total (Svenningsson *et al.* 1990). Prueba de ello son los experimentos llevados a cabo que reflejan cómo mutantes afectados en el catabolismo de aminoácidos están alterados en su capacidad para colonizar la raíz frente al parental (Bayliss *et al.* 1997; Espinosa-Urgel & Ramos 2001). En este sentido, una alta funcionalidad de en la degradación de lisina puede estar asociado al hecho que en estadio de aporque (desarrollo de raíces y estolones) y floración se intensifica la fotosíntesis que presenta una relación lineal con el llenado de los tubérculos (Monroy, 2001), ya que la lisina ha sido asociada al aumento de clorofila de las hojas y retraso en el envejecimiento, con lo que se intensifica el rendimiento de la fotosíntesis. En este sentido, puede existir una gran producción de exudados de lisina de la papa que estén activando mecanismos de degradación de lisina en la microbiota de la rizósfera.

Existe una elevada variabilidad de rutas para el catabolismo de lisina en la naturaleza, conociéndose al menos nueve rutas distintas en el conjunto de los seres vivos. En distintas bacterias, entre ellas *Streptomyces sp.* y *Pseudomonas*, degradan L-lisina hasta CO<sub>2</sub> y H<sub>2</sub>O vía cadaverina (Rius & Demain 1997) y considerando que *Pseudomonas* fue uno de los géneros más representados en aporque y floración puede existir una relación en la composición taxonómica observada y la funcionalidad. No obstante hay que resaltar que el análisis de subredes consiste en una resta de las redes donde se están representando lo que es diferente en cada una de las redes metabólicas, esto nos indica que la redundancia funcional juega un papel relevante en cuanto que un organismo puede estar dotado con más de un mecanismo para realizar una función o varios organismos pueden realizar la misma función, como se ha observado en especies del género *Clostridium* que utilizan la lisina como la principal fuente de energía degradándola hacia acetato, butirato y amonio (Baker *et al.* 1973) y Las diferentes especies de *Pseudomonas* que son capaces de utilizar el aminoácido L-lisina como fuente de carbono y/o nitrógeno (Espinosa-Urgel & Ramos 2001). Lo anterior podría explicar cómo los géneros de *Clostridium* y *Pseudomonas* se vieron fuertemente representados y que pueden estar jugando un papel fundamental en la alta representatividad de la función de degradación de lisina para aporque y floración.

## 4. Conclusiones

EL presente trabajo empleó una metodología novedosa para la reconstrucción del metabolismo a nivel metatranscriptómico de muestras de rizósfera de *Solanum phureja*, ya que al no tener un conjunto de genes codificados en un genoma completo, como es el caso de las reconstrucciones a escala genómica (Pinzon *et al.* 2010; Becker y Palsson, 2005), o teniendo en cuenta los genes expresados, como las reconstrucciones dirigidas (Pinzon *et al.* 2010 ), la aproximación reflejó un comportamiento topológico parecido en relación a aquellos observados en sistemas biológicos individuales y reales de un sistema biológico (Jeong *et al.* 2000).

Se observó que el estado fenológico tiene efectos en la composición de la comunidad de microorganismos presentes en los diferentes estados fenológicos, no obstante, estas diferencias se observan a niveles taxonómicos muy específicos, ya que entre más alto sea el nivel jerárquico en el que se analice la composición, más global va ser el comportamiento, sugiriendo una distribución homogénea.

Los géneros más representados en el estado de emergencia fueron *Rhizobium*, *Starkeya*, *Ehrlichia*, *Propionibacteria* y *Pseudomonas*; mientras que en el caso de aporque y floración fueron los mismos en ambos estados fenológicos *Rhodospirillum*, *Clostridium*, *Heliobacterium*, *Burkholderia*, *Acinetobacter*, y *Pseudomonas*.

Se observó que la composición y estructura de la comunidad bacteriana tiene una relación con la funcionalidad del ecosistema edáfico, y que la influencia del estado de la planta es positivo para definir estas relaciones.

Si bien se observaron patrones funcionales similares para los tres estados fenológicos al realizar procesos de comparación entre ellos a través de las restas de redes se pudo observar que existen diferencias y adicionalmente se evidencia la redundancia funcional

que existe en el ecosistema edáfico. En este sentido la existencia de patrones funcionales similares no quiere decir que los tres estados tenga influencia similares sino que resalta el hecho de que puede existir un microorganismo haciendo una función de diferentes formas o diferentes microorganismos realizando la misma función.

La aplicación de reconstrucciones metabólicas puede llegar a ser una buena estrategia para estudiar esta clase de muestras complejas, ya que como se observo se pueden hacer sub redes con el fin de poder generar modelos específicos de interés, que permitan definir elementos o objetos de investigación más dirigidos a preguntas puntuales.

## 5. Recomendaciones

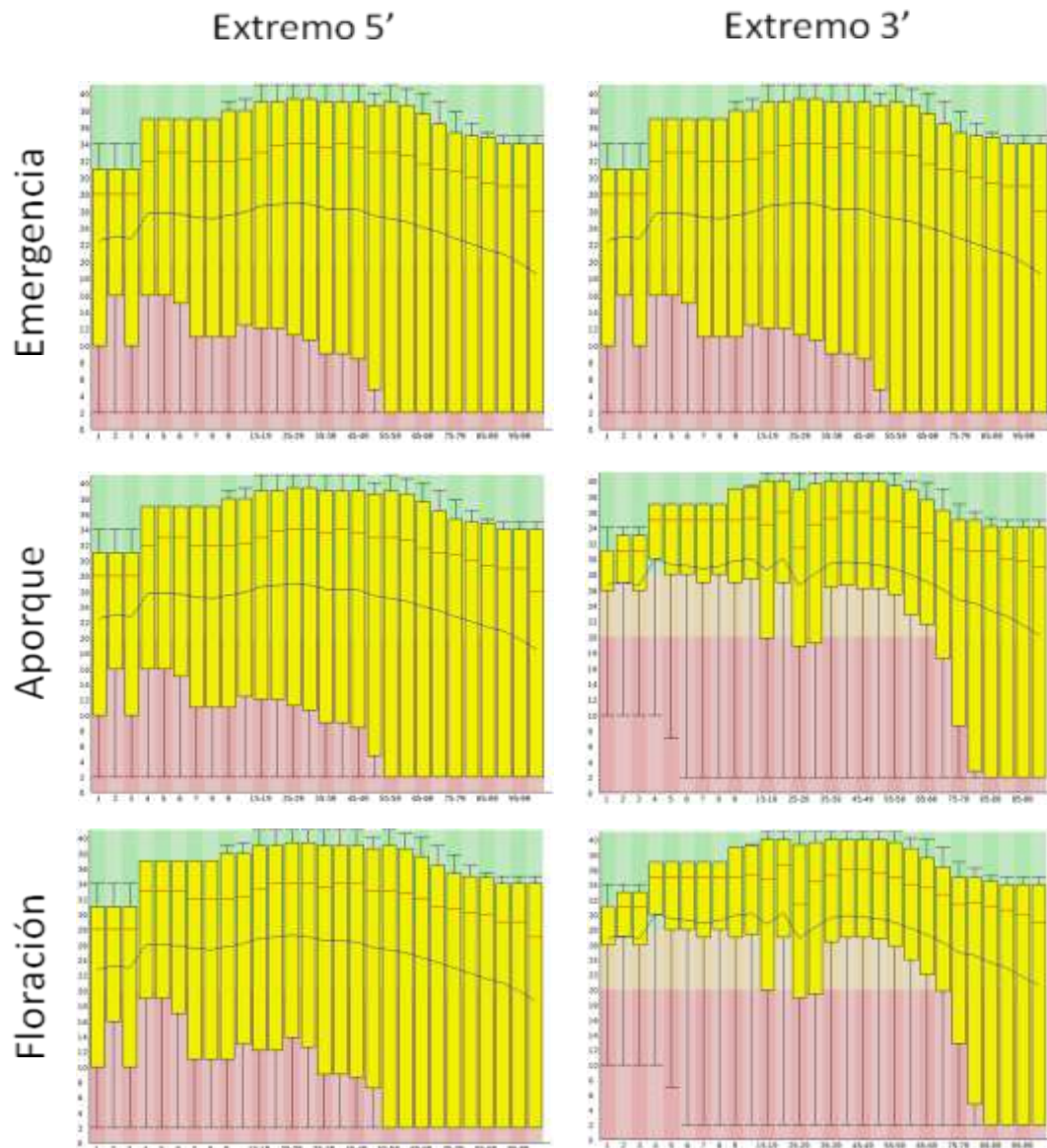
Es necesario que en procesos posteriores de análisis se amplíe el nivel en que se analiza la estructura taxonómica, con el fin que se pueda mapear no sólo a género, sino, especie, donde posiblemente sea más factible realizar observaciones más profundas de las posibles diferencias que pueden existir en el efecto que el desarrollo de la planta puede ejercer sobre la composición y estructura de las comunidades edáficas.

Si bien las reconstrucciones exhibieron un modelo de escala libre, es necesario mejorar los procesos de análisis que permita abordar algunas limitantes en relación a la compartimentalización, ya que por ser una muestra constituida por la información genética de muchos organismos, puede generar que algunos patrones presentes variaciones a aquellos observados en modelos individuales

El empleo de bases no redundantes son una buena estrategia para reducir la demanda computacional a la hora de los análisis, pero puede ser una limitante en cuanto a la profundidad del análisis que se puede realizar, razón por la cual sería de gran importancia el poder emplear las bases con funciones putativas donde se podría recabar una mayor información los sistemas a estudiar.

## A. Anexo: Calidad de secuencias y anotación por función enzimática

Figura A-1: Histogramas de Calidad de los datos para cada uno de los diferentes estados fenológicos seleccionados.



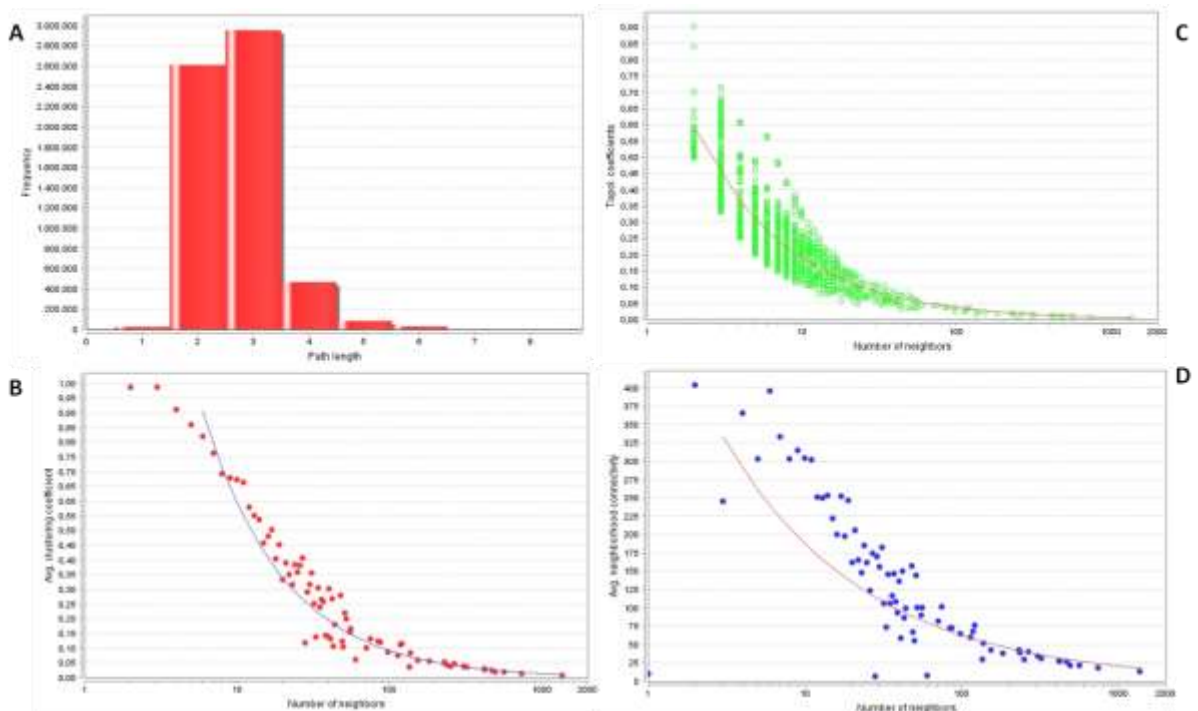
**Tabla A-1:** Rutas metabólicas encontradas en los metatranscriptomas para cada uno de los estados fenológicos.

Nombre de la Ruta	Emer	Apor	Flora	Nombre de la Ruta	Emer	Apor	Flora	Nombre de la Ruta	Emer	Apor	Flora
Glycolysis / Gluconeogenesis	x	x	x	Glutathione metabolism	x	x	x	Thiamine metabolism	x	x	x
Citrate cycle (TCA cycle)	x	x	x	Starch and sucrose metabolism	x	x	x	Riboflavin metabolism	x	x	x
Pentose phosphate pathway	x	x	x	N-Glycan biosynthesis	x	x	x	Vitamin B6 metabolism	x	x	x
Pentose and glucuronate interconversions	x	x	x	Other glycan degradation	x	x	x	Nicotinate and nicotinamide metabolism	x	x	x
Fructose and mannose metabolism	x	x	x	O-Glycan biosynthesis	x	x	x	Pantothenate and CoA biosynthesis	x	x	x
Galactose metabolism	x	x	x	High-mannose type N-glycan biosynthesis	x	x	x	Biotin metabolism	x	x	x
Ascorbate and aldarate metabolism	x	x	x	O-Mannosyl glycan biosynthesis	x	x		Lipoic acid metabolism	x	x	x
Fatty acid biosynthesis	x	x	x	Amino sugar and nucleotide sugar metabolism	x	x	x	Folate biosynthesis	x	x	x
Fatty acid elongation in mitochondria	x	x	x	Streptomycin biosynthesis	x	x	x	Atrazine degradation	x	x	x
Fatty acid metabolism	x	x	x	Biosynthesis of 12-, 14- and 16-membered macrolides	x	x	x	Retinol metabolism	x	x	x
Synthesis and degradation of ketone bodies	x	x	x	Polyketide sugar unit biosynthesis	x	x	x	Porphyrin and chlorophyll metabolism	x	x	x
Steroid biosynthesis	x	x	x	Butirosin and neomycin biosynthesis	x	x	x	Terpenoid backbone biosynthesis	x	x	x
Primary bile acid biosynthesis	x	x	x	Glycosaminoglycan degradation	x	x	x	Indole alkaloid biosynthesis	x	x	x
Secondary bile acid biosynthesis	x	x	x	Glycosaminoglycan biosynthesis - chondroitin sulfate	x	x	x	Monoterpenoid biosynthesis	x	x	x
Ubiquinone and other terpenoid-quinone biosynthesis	x	x	x	Glycosaminoglycan biosynthesis - keratan sulfate	x		x	Limonene and pinene degradation	x	x	x
Steroid hormone biosynthesis	x	x	x	Glycosaminoglycan biosynthesis - heparan sulfate	x	x	x	Diterpenoid biosynthesis	x	x	x
Oxidative phosphorylation	x	x	x	Lipopolysaccharide biosynthesis	x	x	x	Carotenoid biosynthesis	x	x	x
Photosynthesis	x	x	x	Peptidoglycan biosynthesis	x	x	x	Zeatin biosynthesis	x	x	x
Purine metabolism	x	x	x	Glycerolipid metabolism	x	x	x	Sesquiterpenoid biosynthesis		x	x
Puromycin biosynthesis			x	Inositol phosphate metabolism	x	x	x	Nitrogen metabolism	x	x	x
Caffeine metabolism	x	x	x	Glycosylphosphatidylinositol(GPI)-anchor biosynthesis	x	x	x	Sulfur metabolism	x	x	x
Pyrimidine metabolism	x	x	x	Glycerophospholipid metabolism	x	x	x	Caprolactam degradation	x	x	x
Alanine, aspartate and glutamate metabolism	x	x	x	Ether lipid metabolism	x	x	x	Phenylpropanoid biosynthesis	x	x	x
Tetracycline biosynthesis	x	x	x	Arachidonic acid metabolism	x	x	x	Flavonoid biosynthesis	x	x	x
Glycine, serine and threonine metabolism	x	x	x	Linoleic acid metabolism	x	x	x	Anthocyanin biosynthesis	x		
Cysteine and methionine metabolism	x	x	x	alpha-Linolenic acid metabolism	x	x	x	Isoflavonoid biosynthesis	x		
Valine, leucine and isoleucine degradation	x	x	x	Sphingolipid metabolism	x	x	x	Flavone and flavonol biosynthesis	x	x	x

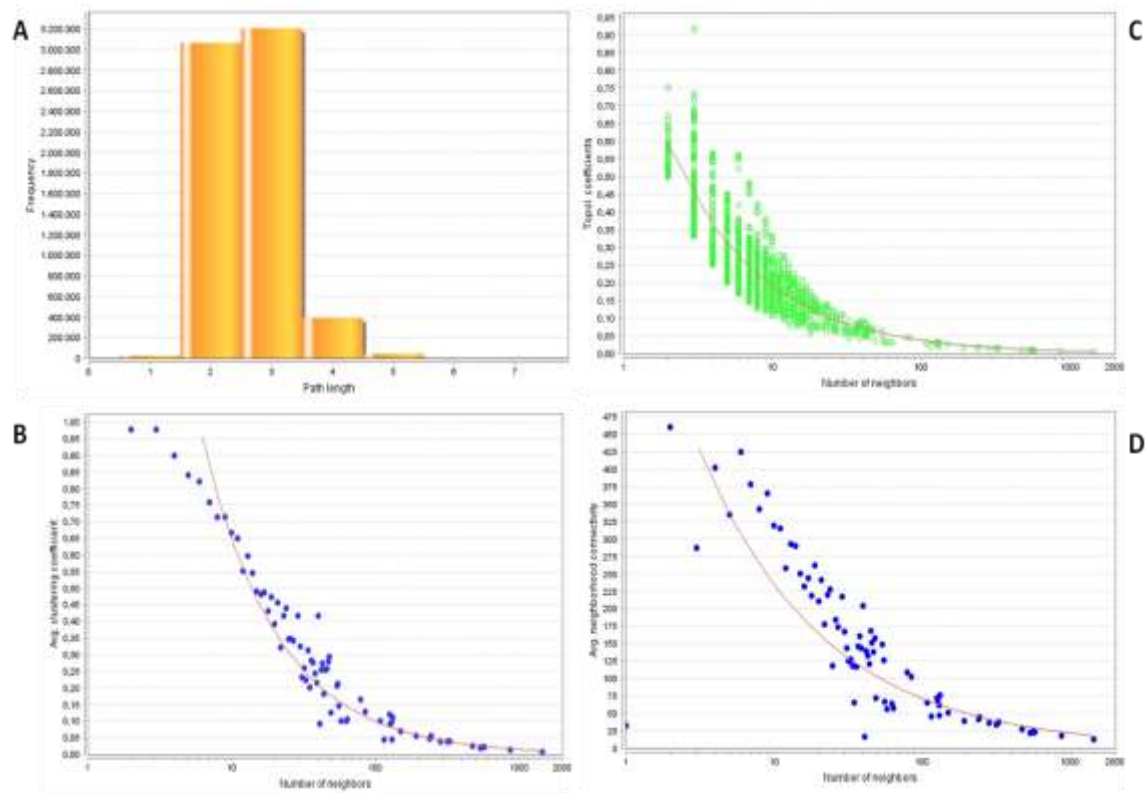
Geraniol degradation	x	x	x	Glycosphingolipid biosynthesis - lacto and neolacto series	x	x	Stilbenoid, diarylheptanoid and gingerol biosynthesis	x	x
Valine, leucine and isoleucine biosynthesis	x	x	x	Glycosphingolipid biosynthesis - globo series	x	x	Isoquinoline alkaloid biosynthesis	x	x
Lysine biosynthesis	x	x	x	Glycosphingolipid biosynthesis - ganglio series	x	x	Tropane, piperidine and pyridine alkaloid biosynthesis	x	x
Lysine degradation	x	x	x	Pyruvate metabolism	x	x	Betalain biosynthesis	x	x
Penicillin and cephalosporin biosynthesis	x	x	x	Biphenyl degradation	x	x	Glucosinolate biosynthesis	x	x
beta-Lactam resistance	x	x	x	Toluene and xylene degradation	x	x	Aminoacyl-tRNA biosynthesis	x	x
Arginine and proline metabolism	x	x	x	2,4-Dichlorobenzoate degradation	x	x	Metabolism of xenobiotics by cytochrome P450	x	x
Clavulanic acid biosynthesis		x	x	1- and 2-Methylnaphthalene degradation	x	x	Insect hormone biosynthesis		x
Histidine metabolism	x	x	x	Tetrachloroethene degradation		x	Drug metabolism - cytochrome P450	x	x
Tyrosine metabolism	x	x	x	Naphthalene and anthracene degradation	x	x	Drug metabolism - other enzymes	x	x
Phenylalanine metabolism	x	x	x	1,4-Dichlorobenzene degradation	x	x	Biosynthesis of unsaturated fatty acids	x	x
gamma-Hexachlorocyclohexane degradation	x	x	x	Fluorene degradation	x	x	Biosynthesis of ansamycins	x	x
Benzoate degradation via hydroxylation	x	x	x	Carbazole degradation	x	x	Biosynthesis of siderophore group nonribosomal peptides	x	x
Bisphenol A degradation	x	x	x	Glyoxylate and dicarboxylate metabolism	x	x	Biosynthesis of vancomycin group antibiotics	x	x
Fluorobenzoate degradation	x	x	x	1,2-Dichloroethane degradation	x	x	Biosynthesis of phenylpropanoids	x	x
Tryptophan metabolism	x	x	x	Benzoate degradation via CoA ligation	x	x	Biosynthesis of terpenoids and steroids	x	x
Phenylalanine, tyrosine and tryptophan biosynthesis	x	x	x	Trinitrotoluene degradation		x	Biosynthesis of alkaloids derived from shikimate pathway	x	x
Novobiocin biosynthesis	x	x	x	Propanoate metabolism	x	x	Biosynthesis of alkaloids derived from ornithine, lysine and nicotinic acid	x	x
Benzoxazinoid biosynthesis		x	x	3-Chloroacrylic acid degradation	x	x	Biosynthesis of alkaloids derived from histidine and purine	x	x
beta-Alanine metabolism	x	x	x	Ethylbenzene degradation	x	x	Biosynthesis of alkaloids derived from terpenoid and polyketide	x	x
Taurine and hypotaurine metabolism	x	x	x	Styrene degradation	x	x	Biosynthesis of plant hormones	x	x
Phosphonate and phosphinate metabolism	x	x	x	Butanoate metabolism	x	x	Metabolic pathways	x	x
Selenoamino acid metabolism	x	x	x	C5-Branched dibasic acid metabolism	x	x	Biosynthesis of secondary metabolites	x	x
Cyanoamino acid metabolism	x	x	x	One carbon pool by folate	x	x	Phosphatidylinositol signaling system	x	x
D-Glutamine and D-glutamate metabolism	x	x	x	Methane metabolism	x	x	mTOR signaling pathway	x	x
D-Arginine and D-ornithine metabolism	x	x	x	Carbon fixation in photosynthetic organisms	x	x	T cell receptor signaling pathway	x	x
D-Alanine metabolism	x	x	x	Reductive carboxylate cycle (CO2 fixation)	x	x	Melanogenesis	x	x
				<b>Emergencia</b>			<b>Aporque</b>		<b>Floración</b>
<b>Total</b>				<b>155</b>			<b>156</b>		<b>158</b>

## B. Anexo: Gráficas obtenidas de los análisis topológicos para los tres metatranscriptomas

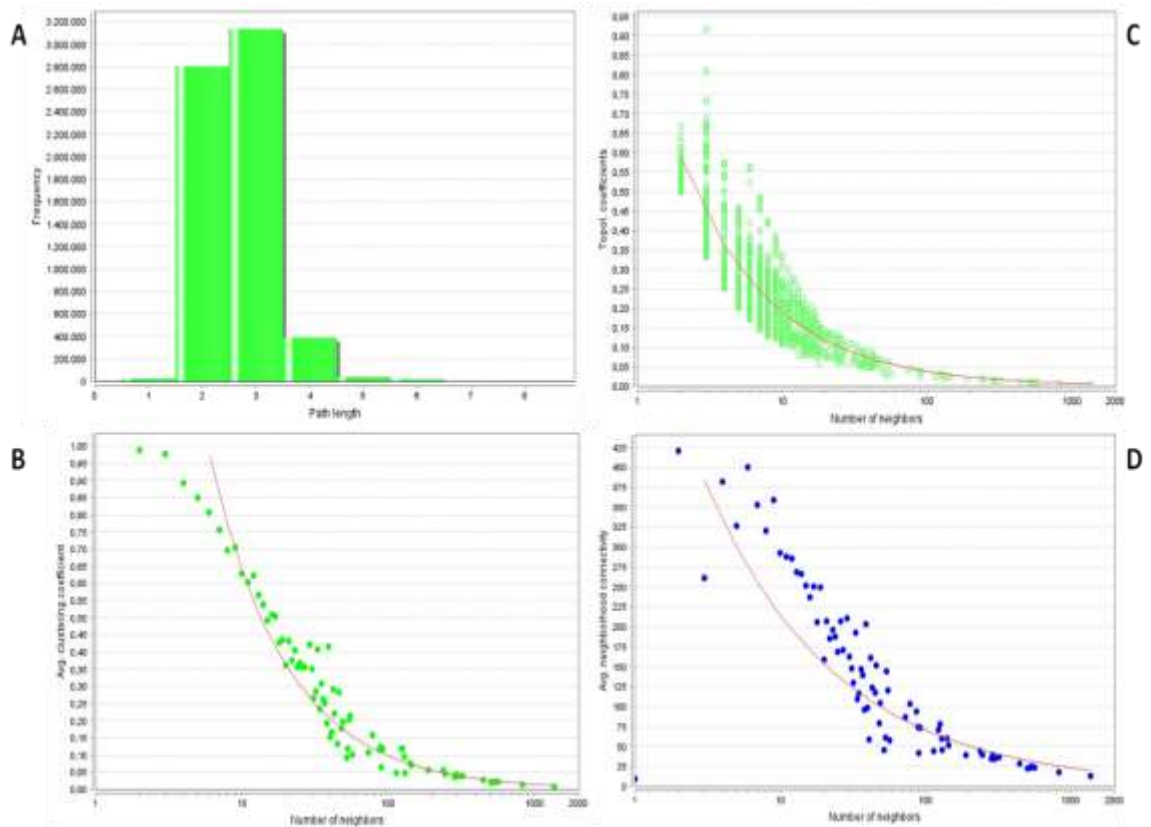
**Figura B-2:** Resultados gráficos de análisis topológico obtenido del software Cytoscape, para el metatranscriptoma en el estado de emergencia. A) Histograma de la distribución del Path más corto, B) Distribución del coeficiente de clusterización, C) Coeficientes topológicos y D) Distribución de la conectividad entre vecinos.



**Figura B-2:** Resultados gráficos de análisis topológico obtenido del software Cytoscape, para el metatranscriptoma en el estado de aporque. A) Histograma de la distribución del Path más corto, B) Distribución del coeficiente de clusterización, C) Coeficientes topológicos y D) Distribución de la conectividad entre vecinos.



**Figura B-3:** Resultados gráficos de análisis topológico obtenido del software Cytoscape, para el metatranscriptoma en el estado de floración. A) Histograma de la distribución del Path más corto, B) Distribución del coeficiente de clusterización, C) Coeficientes topológicos y D) Distribución de la conectividad entre vecinos.



## Bibliografía

Allison P. Heath, Lydia E. Kavraki. (2009). Computational challenges in systems biology. *Computer Science Review*. 1-17.

Albert RK, Barabasi AL. (2002). Statistical mechanics of complex networks. *Rev Mod Phys*. 74: 47.

Altschul SF, *et al.* (1997) "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs," *Nucleic Acids Res*. 25:3389-3402.

Amann RI, Ludwig W, Schleifer KH. (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol. Rev*. 59:143–169.

Araújo da Silva KR, Salles JF, Seldin L, van Elsas JD. (2003). Application of a novel *Paenibacillus*-specific PCR-DGGE method and sequence analysis to assess the diversity of *paenibacillus* spp. in the maize rhizosphere. *J of Micr Meth*. 54(2): 213-231.

Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M. *et al.* (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics*. 9(75).

Badalucco, L. Kuikman, P. J. and Nannipieri, P. (1996). Protease and deaminase activities in wheat rhizosphere and their relation to bacterial and protozoan populations. *Biology and Fertility of Soils* 23: 99–104.

Bader S, Kuhner S, Gavin AC. (2008). Interaction networks for systems biology. *FEBS Letters*. 582: 1220–1224.

Baker JJ, van der Drift C, Stadtman TC. (1973) Purification and properties of lysine mutase a pyridoxal phosphate and B12 coenzyme dependent enzyme. *Biochemistry* 12: 1054-1063

Balazsi G, Barabasi AL, Oltvai ZN. (2005). Topological units of environmental signal processing in the transcriptional regulatory network of *Escherichia coli*. *Proc Natl Acad Sci USA* 2005. 102:7841-7846.

Barabasi AL, Albert R. (1999). Emergence of scaling in random networks. *Science*. 286: 509-512.

Barabasi AL, Oltvai ZN. (2004). Network biology: understanding the cell's functional organization. *Nat Rev Genet*. 5: 101-13.

Bascompte J, Stouffer DB. (2009). The assembly and disassembly of ecological networks. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 1781-1787.

Bayliss C, Bent E, Culham DE, MacLellan S, Clarke AJ, Brown GL, Wood JM. (1997) Bacterial genetic loci implicated in the *Pseudomonas putida* GR12-2R3 canola mutualism: identification of an exudate-inducible sugar transporter. *Can J Microbiol* 43: 809-818

Becker SA. *et al.* (2007). Quantitative prediction of cellular metabolism with constraint-based models: the COBRA toolbox. *Nature Protoc.* 2: 727–738.

Becker SA, Palsson BO. (2005). Genome-scale reconstruction of the metabolic network in *Staphylococcus aureus* N315: an initial draft to the two dimensional annotation. *BMC Microbiol.* 5-8.

Benitez MS, Tustas FB, Rotenberg D, Kleinhenz MD, Cardina J, Stinner D, Miller SA, McSpadden BB. (2007). Multiple statistical approaches of community fingerprint data reveal bacterial populations associated with general disease suppression arising from the application of different organic field management strategies. *Soil Biol. Biochem.* 39: 2289-2301.

Berg J, Lassig M, Wagner A. (2004). Structure and evolution of protein interaction networks: a statistical model for link dynamics and gene duplications, *BMC Evolutionary Biology*, 4: 1-51.

Blum T, Kohlbacher O. (2008). MetaRoute: fast search for relevant metabolic routes for interactive network navigation and visualization. *Bioinformatics.* 24:18.

Botero LM, D'Imperio S, Burr M, McDermott TR, Young M, Hassett DJ. (2005). Poly(A) Polymerase Modification and Reverse Transcriptase PCR Amplification of Environmental RNA. *Applied and Environmental Microbiology.* 71(3): 1267–1275.

Bouffaud ML. *et al.* (2012). Is diversification history of maize influencing selection of soil bacteria by roots? *Mol Ecol.* 21: 195-206.

Bullard JH, Purdom E, Hansen KD, Dudoit S. (2010). Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics* 11:94.

Borneman J, Skroch PW, O'Sullivan KM, Palus JA, Rumjanek NG, Jansen JL, *et al.* (1996). Molecular microbial diversity of an agricultural soil in Wisconsin. *Appl. Environ. Microbiol.* 62:1935–1943.

Bowen, G. D. and Rovira, A. D. (1999). The rhizosphere and its management to improve plant growth. *Advances in Agronomy* 66: 1–102.

Bulgarelli D, Rott M, Schlaeppi K, Ver Loren van Themaat E, Ahmadinejad N, Assenza F *et al.* (2012). Revealing structure and assembly cues for *Arabidopsis* root-inhabiting bacterial microbiota. *Nature* 488: 91–95.

Carvalhais LC, Dennis PG, Tyson GW, Schenk PM. (2012). Application of metatranscriptomics to soil environments. *J of Micr Meth.* 91(2): 246-251.

Claudel-Renard C, Chevalet C, Faraut T, Kahn D. (2003). Enzyme-specific profiles for genome annotation: PRIAM. *Nucleic Acids Research.* 31: 22. 6633-6639

Chaparro JM, Badri DV and Vivanco JM. (2014). Rhizosphere microbiome assemblage is affected by plant development. *The ISME Journal* 8, 790 – 803

Chen CR, Condrón LM, Xu ZH. (2008). Forest Impacts of grassland afforestation with coniferous trees on soil phosphorus dynamics and associated microbial processes: A review. *Ecology and Management* 255: 396-409.

Chet I, Zilberstein Y, Henis Y. (1973) Chemotaxis of *Pseudomonas lachrymans* to plant extracts and to water droplets collected from the leaf surfaces of resistant and susceptible plants. *Physiol. Plant Pathol* 3: 473-479.

DeCoste, N.J., Gadkar, V.J., Filion, M., 2011. Relative and absolute quantitative Real-time PCR-based quantifications of *hcnC* and *phlD* gene transcripts in natural soil spiked with *Pseudomonas* sp. strain LBUM300. *Appl. Environ. Microbiol.* 77, 41–47.

Deutsher MP. (2006) Degradation of RNA in Bacteria: Comparison of mRNA and Stable RNA. *Nucleic Acids Res.* 34:659-666.

DeJongh M. *et al.* (2007). Toward the automated generation of genome-scale metabolic networks in the SEED. *BMC Bioinformatics* 8: 139.

Dennis PG, Miller AJ, Hirsch PR. (2010). Are root exudates more important than other sources of rhizodeposits in structuring rhizosphere bacterial communities? *FEMS Microbiol Ecol* 72: 313–327.

Edwards JS, Ibarra RU, and Palsson BØ. (2001) In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nat Biotechnol*, 19(2):125–130.

Edwards RA, Rodríguez-Brito B, Wegley L, Haynes M, Breitbart M, Peterson DM, *et al.* (2006). Using pyrosequencing to shed light on deep mine microbial ecology. *BMC Genomics* 7:57.

Elsas JD, Speksnijder AJ, van Overbeek LS (2008). A procedure for the metagenomics exploration of disease-suppressive soils. *Journal of Microbiological Methods*, 75: 515-522.

Español J. (2002). Desarrollo de técnicas para la captura, muestreo y seguimiento de la Polilla Guatemalteca *Tecia solanivora* (Povolny) y el Gusano Blanco de la Papa *Premnotrypes vorax* (Hustache) en condiciones de campo. Trabajo de Grado Maestría en Ciencias Agrarias. Escuela de Posgrados. Facultad de Agronomía. Bogotá D.C. Universidad Nacional de Colombia. 140 p.

Espinosa-Urgel M, Ramos JL. (2001). Expression of a *Pseudomonas putida* aminotransferase involved in lysine catabolism is induced in the rhizosphere. *Appl. Environ. Microbiol.* 67: 5219-5224.

Ewing B, Hillier L, Wendl MC, Green P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* Mar;8(3):175-85.

Faccini G, Garzón S, Marínez M, Varela A. (2007). Evaluation of effect of a dual inoculum of phosphate – solubilizing bacteria and *Azotobacter chroococcum*, in crops of creole potato (papa “criolla), “yema de huevo” variety (*Solanum phureja*). In: Velásquez E. and C. Rodríguez-Barrueco (eds.). *First International Meeting on Microbial Phosphate Solubilization*. Springer. 301-308.

Fajardo L y Diego D. (2002). Estudio de factibilidad para la exportación de papa criolla precocida y congelada. Trabajo de Grado Especialista en Gerencia de Negocios Internacionales. Bogotá D.C. Fundación Universidad Jorge Tadeo Lozano. 112.

Faust K, Croes D, van Helden J. (2009). Metabolic Path finding using RPAIR annotation. *J Mol Biol.* 388: 390-414.

Frank JB, Hans VW. (2006). The nature of systems biology. *TRENDS in Microbiology* Vol.15 No.1.

Feist AM, Palsson BO. (2008). The growing scope of applications of genome scale metabolic reconstructions using *Escherichia coli*. *Nat Biotechnol.* 26: 659–667.

Feist AM, Herrgard MJ, Thiele I, Reed JL, Palsson BO. (2009). Reconstruction of biochemical networks in microorganisms. *Nature Reviews Microbiology.* 7: 129.

Feeney DS, Crawford JW, Daniell T, Hallett PD, Nunan N, Ritz K, Rivers M, Young IM. (2006). Three-dimensional microorganization of the soil-rootmicrobe system. *Microbial Ecology*, 52: 151–158.

Fierer N, Breitbart M, Nulton J, Salamon P, Lozupone C, Jones R. (2007). Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of bacteria, archaea, fungi, and viruses in soil. *Applied and Environmental Microbiology*, 73: 7059-7066.

Frias J, Shi Y, Tyson GW, Coleman ML, Schuster SC, Chisholm SW, DeLong EF. (2008). Microbial community gene expression in ocean surface waters. *PNAS.* 105(10): 3805–3810.

Gao H, Yang ZK, Gentry TJ, Wu L, Schadt CW, Zhou J. (2007). Microarray-Based Analysis of Microbial Community RNAs by Whole-Community RNA Amplification. *Applied and Environmental Microbiology.* 73(2): 563-571

Garbeva P, van Veen JA, van Elsas JD. (2004). Microbial diversity in soil: Selection of microbial population by plant and soil type and implications for disease suppressiveness. *Ann Rev Phy.* 42: 243.

Ge H, Liu Z, Church G, *et al.* (2001). Correlation between transcriptome and interactome mapping data from *Saccharomyces cerevisiae*. *Nature Genetics*. 29: 482-486.

Gilbert JA, Field D, Huang Y, Edwards R, Li W, Gilna P, Joint I. (2008). Detection of Large Numbers of Novel Sequences in the Metatranscriptomes of Complex Marine Microbial Communities. *PLoS ONE* 3(8): e3042. doi:10.1371/journal.pone.0003042

Green SJ, Inbar E, Michel FC, Hadar Y, Minz D. (2006). Succession of bacterial communities during early plant development: transition from seed to root and effect of compost amendment. *Appl. Environ. Microbiol.* 72: 3975-3983.

Goh, K., Kahng, B., and Kim, D. (2005) Graph theoretic analysis of protein interaction networks of eukaryotes. *Physica A*. 357:501–512.

Gu YH & Mazzola M (2003) Modification of fluorescent pseudomonad community and control of apple replant disease induced in a wheat cultivar-specific manner. *Appl Soil Ecol* 24:57–72.

Guebel DV, Schmitz U, Wolkenhauer O, Vera J. (2012). Analysis of cell adhesion during early stages of colon cancer based on an extended multi-valued logic approach. *Mol.BioSyst.* 8: 1230-1242.

Guimera R & Nunes LA. (2005). Functional cartography of complex metabolic networks. *Nature* 433, 895-900.

Hansen KD, Brenner SE, Dudoit S. (2010). Biases in Illumina transcriptome sequencing caused by random hexamer priming. *Nucleic Acids Res.* 38:e131.

Hartwell LH, Hopfield JJ, Leibler S, Murray AW. (1999). From molecular to modular cell biology. *Nature*. 402: C47-C52.

He Z, Gentry TJ, Schadt CW, Wu L, Liebich J, Chong SC, *et al.* (2007). GeoChip: a comprehensive microarray for investigating biogeochemical, ecological and environmental processes. *ISME J* 1:67–77.

Hess EL. (1970) Origins of molecular biology. *Science* 168: 664-669

Hongwu M and Zeng A. (2003). Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. *Bioinformatics*. Vol. 19 no. 2:270–277

Huamán Z, Spooner D. (2002). Reclassification of landrace populations of cultivated potatoes (*Solanum* sect. *Petota*). *American Journal of Botany*. 89(6): 947-965.

Huber W, Vincent JC, Long L, Falcon S, and Gentleman R. (2007). Graphs in molecular biology. *BMC Bioinformatics*. Volume 8. Supplement 6.

Holtgrieve GW, Schindler DE, and Jewett PK. (2009). Large predators and biogeochemical hotspots: brown bear (*Ursus arctos*) predation on salmon alters nitrogen cycling in riparian soils. *Ecological Research* 24, 1125–1135.

Huson DH, Auch AF, Qi J, Schuster SC. (2007). MEGAN analysis of metagenomic data. *Genome Res.* 17:377–386.

Inbar E, Green SJ, Hadar Y, Minz D. (2005). Competing factors of compost concentration and proximity to root affect the distribution of streptomycetes. *Microb Ecol.* 50: 73-81.

Inceoglu O, Salles JF, Van Overbeek L, Van Elsas JD (2010) Effect of plant genotype and growth stage on the  $\beta$ -proteobacterial community associated with different potato cultivars in two fields. *Appl Environ Microbiol* 76(11): 3675–3684.

Inceoglu O, Al-Soud WA, Salles JF, Semenov AV, van Elsas JD. (2011) Comparative Analysis of Bacterial Communities in a Potato Field as Determined by Pyrosequencing. *PLoS ONE* 6(8): e23321. doi:10.1371/journal.pone.0023321

Inceoglu O, Salles JF, van Elsas JD. (2012). Soil and cultivar type shape the bacterial community in the potato rhizosphere. *Microb Ecol.* 63: 460-470.

Jaillard, B., Plassard, C. and Hinsinger, P. (2001). Measurements of H<sub>2</sub> fluxes and concentrations in the rhizosphere. p. 231–266. In Z. Rengel (ed.) *Soil acidity handbook*. Marcel Dekker Inc., New York.

Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N. and Barabasi, A.L. (2000) The large-scale organization of metabolic networks. *Nature*, 407, 651–654.

Jordan F, Scheuring I. (2002). Searching for keystones in ecological networks, *Oikos*, 99(3). 607-612.

Junker BH, Schreiber F. (2008). *Analysis of Biological Networks*. John Wiley & Sons, Inc, Hoboken, New Jersey, USA.

Kakirde KS, Parsley LC, Liles MR. (2010). Size does matter: Application-driven approaches for soil metagenomics. *Soil biology & biochemistry*, 42: 1911-1923.

Karlebach G, Shair R. (2008). Modelling and analysis of gene regulatory networks. *Nat Rev Mol Cell Biol.* 9:770-780

Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M. (2006). From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.* 34: D354–D357

Karp PD, Paley S, Romero P. (2002). The pathway tools software. *Bioinformatics.* 18(Suppl 1): 1-S8.

Krieger CJ. *et al.* (2004). MetaCyc: a multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Res.* 32: D438–D442.

- Kumar VS, Dasika MS. (2007). Maranas CD: Optimization based automated curation of metabolic reconstructions. *BMC Bioinformatics* 8: (212).
- Kunin JB, Pereira CA. (2004). Ouzounis, Functional Evolution of the Yeast Protein Interaction Network, *Molecular Biology and Evolution*, 21, 7, p\_ags. 1171 -1176.
- Kuske CR, Ticknor LO, Miller ME, Dunbar JM, Davis JA, *et al.* (2002) Comparison of soil bacterial communities in rhizospheres of three plant species and the interspaces in an arid grassland. *Applied Environmental Microbiology* 68: 1854–1863.
- Laurent P, Raaijmakers JM, Philippe L, van der Putten WH. (2013). Going back to the roots: the microbial ecology of the rhizosphere. *Nature Reviews Microbiology*. 11: 789-799
- Lestrangle KK, Bender GL, Djordjevic MA, Rolfe BG, Redmond JW. (1990). The Rhizobium strain Ngr234 Nodd1 gene-product responds to activation by the simple phenolic-compounds vanillin and isovanillin present in wheat seedling extracts. *Mol Plant-Microbe Interac* 3: 214-220
- Li W and Godzik A. (2006). "CD-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequence". *Bioinformatics*. 22:1658-9.
- Lujan L. (1994). Ecología de la papa. *Revista Papa* N° 12. Federación Colombiana de Productores de Papa. Bogotá. pp. 4 – 28
- Lundberg DS, Lebeis SL, Paredes SH, Yourstone S, Gehring J, Malfatti S *et al.* (2012). Defining the core *Arabidopsis thaliana* root microbiome. *Nature* 488: 86–90.
- Ma HU, Zeng AP. (2006). Reconstruction of Metabolic Networks from Genome Information and Its Structural and Functional Analysis. *Computational Systems Biology*, 169-189.
- Metzker M. (2010) «Sequencing technologies the next generation» *Nature Reviews Genetics* 11: 31-46.
- Mazzola M, Funnell DL, Raaijmakers JM. (2004). Wheat cultivar-specific selection of 2,4-diacetylphloroglucinol-producing fluorescent *Pseudomonas* species from resident soil populations. *Microb. Ecol.* 48: 338-348.
- Meier, U. 2001. Estadios de las plantas mono- y dicotiledóneas. BBCH Monografía. Centro Federal de Investigaciones Biológicas para Agricultura y Silvicultura. 149.
- Mengel, K., Kirkby, E. A., Kosegarten, H. and Appel, T. (2001). Principles of plant nutrition. 5<sup>th</sup> Ed. Editorial Kluwer Academic Publisher: 84-85.
- Micallef SA, Channer S, Shiaris MP, Colon-Carmona A. (2009). Plant age and genotype impact the progression of bacterial community succession in the *Arabidopsis* rhizosphere. *Plant Signal Behav* 4:777–780.

Mocali S, Benedetti A. (2010). Exploring research frontiers in microbiology: The challenge of metagenomics in soil microbiology. *Research in Microbiology*, 161: 497-505.

Monroy LP. (2001). Evaluación de variables agronómicas y fisiológicas en cinco clones avanzados del programa de mejoramiento de papas diploides en dos localidades de Cundinamarca. Trabajo de Grado Ingeniero Agrónomo. Facultad de Ingeniería Agronómica. Universidad Nacional de Colombia. Bogotá. 45 p.

Moran AM, (2009). Metatranscriptomics: eavesdropping on complex microbial communities. *Microbe* 4, 329–335.

Morse CC, Yevdokimov IV, DeLuca, TH. (2000). In situ extraction of rhizosphere organic compounds from contrasting plant communities. *Commun. Soil Sci. Plant Anal.* 31, 725–742.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5:621–628.

Moss GP. (2011). International Union of Biochemistry and Molecular Biology. [en línea] School of Biological and Chemical Sciences, Queen Mary, University of London, UK. <http://www.chem.qmul.ac.uk/iubmb/>.

Mouillot D, Mason WHN, Dumay O, Wilson JB. (2004). Functional regularity: a neglected aspect of functional diversity. *Oecologia* 142: 353-359.

Mougel C. *et al.* (2006). Dynamic of the genetic structure of bacterial and fungal communities at different developmental stages of *Medicago truncatula* Gaertn. cv. Jemalong line J5. *New Phytol.* 170: 165-175.

Muñoz LA, Lucero A. (2008). Efecto de la fertilización orgánica en el cultivo de papa criolla *Solanum phureja*. *Agronomía Colombiana*. Bogotá. 26(2). 340-346.

Nannipieri P, Ascher J, Ceccherini MT, Landi L, Pietramellara G and Renella G (2003). Microbial diversity and soil functions. *European Journal of Soil Science*, 54: 655-670.

National Research Council (US) Committee on Metagenomics: Challenges and Functional Applications. (2007). *The New Science of Metagenomics: Revealing the Secrets of Our Microbial Planet*. National Academies Press (US): Washington (DC).

NCBI (National Center for Biotechnology Information). (2011) [en línea] U.S. National Library of Medicine. 8600 Rockville Pike, Bethesda MD, 20894 USA. <http://www.ncbi.nlm.nih.gov>.

Neira I, Jose P, Gabriel A. (1996). Aplicación de tres fuentes de materia orgánica como complemento a la fertilización con 13-26-6 en papa criolla “Yema de Huevo” *Solanum phureja* en Toca Boyacá. En: *Memorias II Simposio Nacional de Papa Criolla “Yema de Huevo” Solanum phureja*. Revista Papa. Federación Colombiana de Productores de Papa. Bogotá. 21-23

Nelson E. (2004). Microbial dynamics and interactions in the spermosphere. *Annu. Rev. Phytopathol.* 42: 271-309.

Nikata T, Sumida K, Kato J, Ohtake H. (1992) Rapid Method for Analyzing Bacterial Behavioral Responses to Chemical Stimuli. *Appl Environ Microbiol.* 58: 2250-2254

Ochoa CM. (2001). *Las Papas de Sudamérica: Bolivia*. CIP-COSUDE-CID-IFEA. Bolivia. 536 p.

Olaf W, Charles A, Robert J, Gustav S, Olaf D. (2013). The road from systems biology to systems medicine. *Pediatric Research.* 73: 502-507.

Oberhardt MA, Palsson BO, Papin JA. (2009). Applications of genome-scale metabolic reconstructions. *Molecular Systems Biology* 5:320

Palsson BO. (2006). *Systems Biology: Properties of Reconstructed Networks*. Cambridge University Press. New York. 336p.

Peano C, Pietrelli A, Consolandi C, Rossi E, Petiti L, Tagliabue L, De Bellis G and Landini P. (2013). An efficient rRNA removal method for RNA sequencing in GC-rich bacteria. *Microbial Informatics and Experimentation* 3:1.

Pepper IL, and Bezdicek DF. (1990). Root microbial interactions and rhizosphere nutrient dynamics. In *Crops as enhancer of nutrient use*, eds. V. C. Baligar and R. R. Duncan, 375–410. San Diego, CA: Academic Press.

Pinzon A, Rodriguez LM, Gonzalez A, Bernal A, Restrepo S. (2010). Novel approach for the characterization of plant- pathogen interactions. *Briefings in Bioinformatics*: 1-12.

Pitkanen E, Rousu J and Ukkonen E. (2010). Computational methods for metabolic reconstruction. *Current Opinion in Biotechnology.* 21:70–77.

Planes FJ, Beasley JE. (2009). An optimization model for metabolic pathways. *Bioinformatics.* 25(20): 2723-2729.

Price DJ. (1976). A general theory of bibliometric and other cumulative disadvantage processes. *J Am Soc Inform Sci.* 27: 292-306.

Poretsky RS, Bano N, Buchan A, LeCleir G, Kleikemper J, Pickering M, Pate WM, Moran MA, Hollibaugh JT. (2005). Analysis of Microbial Gene Transcripts in Environmental Samples. *Applied and Environmental Microbiology.* 71(7): 4121-4126

Poretsky RS, Hewson I, Sun S, Allen AE, Zehr JP, Moran MA. (2009). Comparative day/night metatranscriptomic analysis of microbial communities in the North Pacific subtropical gyre. *Environ Microbiol* 11: 1359–1375.

Quijano, J.C., Kumar, P., Drewry, D.T., 2013. Passive regulation of soil biogeochemical cycling by root water transport. *Water Resources Research* 49, 3729-3746.

Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabasi AL. (2002). Hierarchical organization of modularity in metabolic networks. *Science* 297: 1551-5.

Ravasz E, Barabasi AL. (2003). Hierarchical organization in complex networks. *Phys Rev E*. 67: 2611-2.

Ren Q, Chen K, Paulsen IT. (2007). TransportDB: a comprehensive database resource for cytoplasmic membrane transport systems and outer membrane channels. *Nucleic Acids Res*. 35: D274-D279.

Richter, D.D. 2007. Humanity's transformation of Earth's soil: Pedology's new frontier. *Soil Sci*. 172: 957-967. doi:10.1097/ss.0b013e3181586bb7.

Rivera Díaz I. (2005). Aplicación de un modelo matemático para evaluar la pérdida de suelo (tn/ha/año) en una zona de ladera bajo cultivo de papa (*Solanum phureja*) en la Sabana de Bogotá. Trabajo de Grado Magíster en Ciencias RECOPIACIÓN DE LA INVESTIGACIÓN DEL SISTEMA PRODUCTIVO PAPA CRIOLLA | 145 Agrarias. Facultad de Agronomía. Universidad Nacional de Colombia. Bogotá. 81p.

Rodríguez J, Saa C. (2004). Evaluación de algunas propiedades físicas y rendimientos de papa criolla amarilla (*Solanum phureja* Juz *et* Buk) en dos lotes sometidos a recuperación mediante la utilización de zanjas de fertilidad, en el corregimiento de Mapachico, municipio de Pasto. Trabajo de Grado Ingeniero Agrónomo. Facultad de Ciencias Agrícolas, Universidad de Nariño. San Juan de Pasto. 89 p.

Ros M, Klammer S, Knapp B, Aichberger K, Insam H. (2006). Long-term effects of compost amendment of soil on functional and structural diversity and microbial activity. *Soil Use Manage*. 22: 209-218.

Rius N, Demain AL. (1997). Lysine epsilon-aminotransferase, the initial enzyme of cephalosporin biosynthesis in actinomycetes. *J Microbiol Biotechnol* 7: 95-100

Saison C, Degrange V, Oliver R, Millard P, Commeaux C, Montange D, Le Roux X. (2006). Alteration and resilience of the soil microbial community following compost amendment: effects of compost level and compost-borne microbial community. *Environ. Microbiol*. 8: 247-257.

Schomburg I. *et al.* 2004. BRENDA, the enzyme database: updates and major new developments. *Nucleic Acids Res*. 32: D431–D433.

Schmieder R, Lim YW, and Edwards R. (2011). Identification and removal of ribosomal RNA sequences from metatranscriptomes. *Oxford Journals Bioinformatics*. 1-3.

Shannon P, Markiel A, Ozier O. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res*.13: 2498-2504.

Schilling G, Gransee A, Deubel A, Lezovic G, Ruppel S. (1998). Phosphorus availability, root exudates, and microbial activity in the rhizosphere. *Z. Pflanzenernähr. Bodenk.* 161, 465–478.

Smalla K, Wieland G, Buchner A, Zock A, Parzy J, Kaiser S, Roskot N, Heuer H & Berg G (2001) Bulk and rhizosphere soil bacterial communities studied by denaturing gradient gel electrophoresis: plant-dependent enrichment and seasonal shifts revealed. *Appl Environ Microbiol* 67: 4742–4751.

Steuer R. (2007). Computational approaches to the topology, stability and dynamics of metabolic networks. *Phytochemistry.* 68(1618): 2139-2151.

Streit WR, Daniel R. (2010). *Metagenomics.* Humana Pr Inc.

Svenningsson H, Sundin P, Liljenberg C (1990) Lipids, carbohydrates and amino-acids exuded from the axenic roots of rape seedlings exposed to water-deficit stress. *Plant Cell Environ* 13: 155-162

Tiquia SM, Lloyd J, Herms DA, Hoitink HAJ, Michel FC. (2002). Effects of mulching and fertilization on soil nutrients, microbial activity and rhizosphere bacterial community structure determined by analysis of TRFLPs of PCR-amplified 16S rRNA genes. *Appl Soil Ecol.* 21: 31-48.

Torsvik V, Goksøyr J, Daae FL. (1990). High diversity in DNA of soil bacteria. *Appl. Environ. Microbiol.* 56:782–787.

Torsvik, V, Øvreås, L. (2002). Microbial diversity and function in soil: from genes to ecosystems. *Current Opinion in Microbiology*, 5(3), 240-245.

Urich T, Lanzen A, Qi J, Huson DH, Schleper C, *et al.* (2008) Simultaneous Assessment of Soil Microbial Community Structure and Function through Analysis of the Meta-Transcriptome. *PLoS ONE* 3(6): e2527. doi:10.1371/journal.pone.0002527.

Valbuena, B.I. (2000). Aspectos básicos sobre el crecimiento y desarrollo en el cultivo de la papa. Manejo integrado del cultivo de la papa en Colombia. Corporación Colombiana de Investigación Agropecuaria (Corpoica), Bogotá.

van Overbeek L, van Elsas JD (2008) Effects of plant genotype and growth stage on the structure of bacterial communities associated with potato (*Solanum tuberosum* L.). *FEMS Microbiol Ecol* 64(2): 283–296.

Varma A, Boesch BW, and Palsson BØ (1993). Stoichiometric interpretation of *Escherichia coli* glucose catabolism under various oxygenation rates. *Appl Environ Microbiol*, 59(8):2465–2473,

Vartoukian SR, Palmer RM & Wade WG. (2010). Cultivation of a *Synergistetes* strain representing a previously uncultivated lineage. *Environ Microbiol* 12: 916–928.

Vilchez S, Molina L, Ramos C, Ramos JL. (2000) Proline catabolism by *Pseudomonas putida*: cloning, characterization, and expression of the put genes in the presence of root exudates. *J Bacteriol* 182: 91-99

Ward DM, Ferris MJ, Nold SC, Bateson MM. (1998). A natural view of microbial biodiversity within hot spring cyanobacterial mat communities. *Microbiology and Molecular Biology Review* 62: 1353–1370.

Watts DJ, Strogatz SH. (1998). Collective dynamics of ‘small-world’ networks. *Nature* 393: 440-442.

Weinert N. *et al.* (2011). PhyloChip hybridization uncovered an enormous bacterial diversity in the rhizosphere of different potato cultivars: many common and few cultivar-dependent taxa. *FEMS Microbiol Ecol.* 75: 497–506.

Whipps, J.M., 1987. Carbon loss from the roots of tomato and pea seedlings grown in soil. *Plant and Soil* 103, 95–100.

Yeates, G., and Darrah P.R. (1991). Microbial changes in a model rhizosphere. *Soil Biology and Biochemistry* 23: 963–971.

Youssef, R.A., and Chino, M. (1989). Root-induced changes in the rhizosphere of plants. I. pH changes in relation to the bulk soil. *Soil Science and Plant Nutrition* 35: 461–468.

Zhang W, Li F, Nie L. (2010). Integrating multiple ‘omics’ analysis for microbial biology: application and methodologies. *Microbiology.* 156: 287-301.

Zhou J, Xia B, Treves D, Wu L, Marsh T, O’Neill R *et al.* (2002). Spatial and resource factors influencing high microbial diversity in soil. *Appl Environ Microbiol* 68: 326–334.

Zwolinski MD. (2007). DNA Sequencing: Strategies for Soil Microbiology. *Soil Science Society of America Journal* 71:592.