

ANÁLISIS BIOINFORMÁTICO DEL TRANSCRIPTOMA DEL MAQUI (*Aristotelia chilensis*): IDENTIFICACIÓN DE GENES INVOLUCRADOS EN LA BIOSÍNTESIS DE ANTOCIANINAS

**MAURICIO ALFONSO RUIZ MORAGA
INGENIERO EN BIOINFORMÁTICA**

RESUMEN

Aristotelia chilensis o comúnmente llamado maqui, es un pequeño árbol endémico chileno que produce un fruto de tipo baya que ha comenzado a tomar relevancia biotecnológica y médica, por su alta actividad antioxidante en comparación con otros frutos de tipo bayas, los cuales están dados principalmente por su alta concentración de antocianinas específicas. Las antocianinas del maqui se destacan por dar pigmentación de tonalidades púrpuras y azules a lo largo de la etapa de maduración del fruto. En el maqui las antocianinas con mayor concentración corresponden a delfinidina y cianidina, mostrando niveles de concentración por encima de la media general en éste tipo de frutos. Actualmente los genes implicados en la regulación y biosíntesis de antocianinas en el maqui permanecen sin ser investigado. Por tanto el objetivo principal del presente proyecto es estudiar el transcriptoma del maqui para identificar y caracterizar los genes implicados en la formación de antocianinas en el fruto del maqui y así poder justificar su alto contenido. Para esto se realizaron secuenciaciones en MiSeq de Illumina, para tres estados diferentes de maduración; fruto verde, pintón y maduro, obteniendo 29.603.685 lecturas de extremos pareados de 150 pb de largo. El total de unigenes ensamblados fue de 64.924, de los cuales se anotaron 32.453 de manera automática utilizando la base de datos no redundante de proteínas del NCBI. De los unigenes no anotados, 28.906 corresponden a secuencias no codificantes, identificadas usando CPC (Coding Protein Calculator). Utilizando una base de datos local de los genes asociados a la ruta metabólica de la antocianina y la herramienta de alineamientos locales BLAST (tblastn), se lograron identificar los unigenes de *Aristotelia chilensis* asociados a ésta ruta. Se observó que genes claves de la ruta de biosíntesis de antocianinas, tales como F3'H, F3'5'H, DFR y UFGT, se encuentran altamente expresos cuando el fruto se encuentra en estado maduro. El análisis de expresión diferencial entre los unigenes de fruto verde y maduro, realizados con EdgeR, resultó con un total de 9.452 unigenes diferencialmente expresos entre los dos estados de maduración.

La buena calidad de secuenciación se demostró en la gran cantidad de unigenes obtenidos

del transcriptoma reconstruido en forma de novo, mostrando también que existen datos interesantes que pueden ser objetivo de estudios posteriores, como los unigenes asociados a ARN no codificante. El análisis de expresión diferencial arrojó resultados que concuerdan con los contenidos de antocianinas totales y perfiles específicos de delphinidina, mostrando también la sobre expresión que existe en el fruto maduro de algunos genes estructurales de la ruta metabólica de la antocianina.

ABSTRACT

Aristotelia chilensis or commonly called maqui, is a Chilean endemic tree that produces a berry type fruit of recently biotechnological and medical importance, well-known for its high antioxidant activity compared to other berry fruits, which are given mainly for its high concentration of specific anthocyanins.

Anthocyanins in maqui stand to give purple pigmentation and blue tones along the ripening stage. In maqui, anthocyanins with the most concentration are delphinidin and cyanidin, showing levels above the average concentration for this type of fruit. Currently, genes involved in regulation and biosynthesis of anthocyanins in maqui remain to be studied. Therefore, the main objective of this project is to study the transcriptome of maqui to identify and characterize genes involved in the formation of anthocyanins in maqui berry and justify their high content. For this, three runs on a MiSeq Illumina sequencer were performed, for three different stages of ripening; for the green fruit, the half ripened fruit and the ripe fruit, obtaining 19.603.685 paired end reads of 151 bp in length. The total number of assembled unigenes were 64.924, of which 32.453 were automatically annotated using the non redundant database from NCBI. From non annotated unigenes, 28.906 correspond to non coding sequences, identified using CPC (Coding Protein Calculator).

Differential expression analysis of unigenes between green and ripe fruits, using EdgeR, resulted in a total of 9.452 differentially expressed unigenes, from which 7.222 were up regulated and 2.230 were down regulated in green fruits compared to ripe fruits.

Using a local database of anthocyanin metabolic pathway genes and BLAST (tblast), achieved identify the unigenes of *Aristotelia chilensis* associated to this pathway. It was observed that key genes of the biosynthesis of anthocyanins, such as F3'5'H, DFR, UFGT and regulator genes (Myb family transcription factors) are highly expressed in ripe fruit.

The good sequencing quality was demonstrated by the large number of unigenes obtained from the de novo transcriptomic analysis, also providing interesting data that can be target of future studies, as unigenes that were transcribed to non coding RNAs. Differential expression analysis resulted in data consistent with the total anthocyanins content and specific profiles of delphinidin, also showing the up regulation of some structural genes of the anthocyanin metabolic pathway on the ripe fruit.