



UNIVERSIDAD NACIONAL DE LA AMAZONÍA PERUANA

Centro de Investigaciones de Recursos Naturales de la Amazonía
(CIRNA)

Unidad Especializada de Biotecnología (UEB)

PROYECTO

ANÁLISIS ESTRUCTURAL Y FUNCIONAL DEL GENOMA DE *Myrciaria dubia* “CAMU-CAMU” COMO BASE PARA SU MEJORAMIENTO GENÉTICO

INVESTIGADORES

Dr. Juan Carlos Castro Gómez (CIRNA-UNAP) [Investigador Principal]

Dr. Aureliano Bombarely Gomez (DH-VPISU)

Dr. Miguel Angel Botella Mesa (DBMB-UM)

Dr. Jorge Luis Marapara del Águila (CIRNA-UNAP)

Dra. Marianela Cobos Ruiz (LBB-UCP)

M.Sc. Sixto Alfredo Imán Correa (ACRF-INIA)

TESISTAS

Pre-grado: 04 tesis

Post- Grado: 02 tesis

PRACTICANTES: 08 practicantes

DURACIÓN: 02 años

MONTO: S/. 70,000.00

IQUITOS – PERÚ

2015

I. TITULO

Análisis estructural y funcional del genoma de *Myrciaria dubia* “camu-camu” como base para su mejoramiento genético.

1.1. ÁREA DE INVESTIGACIÓN EN EL QUE SE INSERTA EL PROYECTO

Biotechnología

1.2. DURACIÓN DEL PROYECTO

02 años

Inicio: Jun/2015

Término: Jun/2017

1.3. COSTO TOTAL DEL PROYECTO

S/. 70,000.00

1.4. NOMBRES Y APELLIDOS DE LOS INVESTIGADORES

Dr. Juan Carlos Castro Gómez

Investigador Principal, Unidad Especializada de Biotechnología (UEB), Centro de Investigaciones de Recursos Naturales de la Amazonía (CIRNA), Universidad Nacional de la Amazonía Peruana (UNAP). Psje. Los Paujiles S/N, San Juan Bautista, Teléfono: 965-017215, correo electrónico: juanccgomez@yahoo.es.

Dr. Aureliano Bombarely Gomez

Investigador, Department of Horticulture (DH), Virginia Polytechnic Institute and State University (VPISU), Latham Hall 216, Blacksburg, Virginia, USA, Teléfono: (504) 2312426, correo electrónico: aurebg@vt.edu.

Dr. Miguel Angel Botella Mesa

Investigador, Departamento de Biología Molecular y Bioquímica (DBMB), Facultad de Ciencias, Universidad de Málaga (UM), España, Teléfono: (34) 952134268, correo electrónico: mabotella@uma.es.

Dr. Jorge Luis Marapara Del Águila

Investigador, Unidad Especializada de Biotechnología (UEB). Centro de Investigaciones de Recursos Naturales de la Amazonía (CIRNA), Universidad Nacional de la Amazonía

Peruana (UNAP). Psje. Los Paujiles S/N, San Juan Bautista, Teléfono: 065236121, correo electrónico: jmandela@hotmail.com.

Dra. Marianela Cobos Ruiz

Investigadora, Laboratorio de Biotecnología y Bioenergética (LBB), Universidad Científica del Perú (UCP), Av. Abelardo Quiñones Km 2,5, San Juan Bautista, Iquitos, Teléfono: 065-263569, correo electrónico: marianela_cobosr@yahoo.es.

MSc. Sixto Alfredo Imán Correa

Investigador, Área de Conservación de Recursos Fitogenéticos (ACRF), Instituto Nacional de Innovación Agraria (INIA), Calle San Roque 209, San Juan Bautista, Teléfono: 065261132, correo electrónico: siman@inia.gob.pe

1.5. INSTITUCIONES COMPROMETIDAS

Universidad Nacional de la Amazonía Peruana (UNAP)

Es la Institución ejecutora principal, cuenta con el CIRNA implementado con equipos para realizar varios de los análisis propuestos en el Proyecto.

Virginia Polytechnic Institute and State University (VPISU)

Es una Institución participante que cuenta con el Departamento de Horticultura implementado con programas y equipos de cómputo para realizar los análisis bioinformáticos del Proyecto.

Universidad de Málaga (UM)

Es una Institución participante que cuenta con el Departamento de Bioquímica y Biología Molecular implementado con equipos para realizar análisis moleculares considerados en el Proyecto.

Universidad Científica del Perú (UCP)

Institución participante que cuenta con el LBB proporcionará el soporte en la formación de recursos humanos y análisis moleculares en el presente proyecto.

Instituto Nacional de Innovación Agraria (INIA)

Institución participante, que cuenta con una colección de germoplasma de *Myrciaria dubia* y proporcionará las muestras biológicas para realizar los análisis requeridos en el Proyecto.

II. ÍNDICE

I. Título	2
II. Índice	3
III. Resumen	5
IV. Planteamiento del Problema	7
V. Antecedentes de la Investigación	9
VI. Hipótesis	13
VII. Objetivos de la Investigación	14
VIII. Metodología del Proyecto	15
IX. Metas por componentes	18
X. Aspectos éticos	19
XI. Resultados esperados	20
XII. Estrategias a utilizar para la transferencia y comunicación de los resultados	21
XIII. Impactos esperados	22
XIV. Infraestructura y equipos a utilizarse	25
XV. Cronograma de actividades por componentes mensualizado y responsables	26
XVI. Resumen del presupuesto por rubros, partidas y componentes en soles	27
XVII. Monitoreo y evaluación: Matriz de Marco Lógico	29
XVIII. Referencias bibliográficas	31
XIX. Anexos	39

III. RESUMEN DEL PROYECTO

La amazonía peruana sobresale por su mega biodiversidad de especies vegetales. Esta biodiversidad viene siendo aprovechada por pobladores de comunidades indígenas y pueblos ribereños como fuente de alimentos, fitofármacos y materiales para construcción, entre otros diversos usos. Pero es necesario explorar nuevas alternativas para aprovechar sosteniblemente la biodiversidad amazónica. Es por tanto esperable que muchos de los genes únicos que poseen estas especies sean potencialmente útiles para realizar mejora genética y desarrollar procesos biotecnológicos para la producción de compuestos bioactivos de alta cotización en el mercado mundial. Esta mina de genes con valor incalculable debe ser estudiada y explotada para posicionarnos favorablemente a nivel regional y mundial. Para lograr esta meta debemos realizar análisis estructurales y funcionales de los genomas de especies promisorias como *M. dubia*. Si hace algunos años esto era utópico, actualmente, estos análisis son posibles con el desarrollo de plataformas de secuenciamiento masivo, con las nuevas herramientas bioinformáticas y los potentes sistemas de cómputo. Por tanto, el objetivo principal de este proyecto es realizar el análisis estructural y funcional del genoma de *Myrciaria dubia* “camu-camu”. Las muestras botánicas serán obtenidas de la colección única de germoplasma de *M. dubia* del INIA y los análisis se realizarán en tres fases. En la primera fase se estimará el tamaño del genoma usando citometría de flujo a partir de núcleos celulares de plántulas de *M. dubia*. Adicionalmente, se estimará la heterocigosidad de los miembros de la colección secuenciando de 10 a 20 regiones variables. Ambos resultados permitirán seleccionar el cultivar o acceso de la colección más adecuado para la secuenciación del genoma. En la segunda fase se procederá a la secuenciación masiva y ensamblaje del genoma de *M. dubia* usando una combinación de lecturas cortas producidas por el secuenciador de Illumina (con una cobertura de 100X) y lecturas largas producidas por el secuenciador de Pacific Biosystems (con una cobertura de 20X). En una tercera fase se procederá a la anotación estructural y funcional del genoma para lo cual se necesitan datos transcriptómicos que capturen la mayor parte los genes de esta especie. Para ello se procederá a la purificación de ARN y secuenciación de flor, hoja, tallo, raíz, plántula y tres estadios de maduración del fruto. Las lecturas producidas se usarán para la anotación del genoma. Una vez se haya producido un borrador del genoma de *M. dubia*, se realizará diferentes análisis bioinformáticos tales como análisis de familias génicas y comparación con genomas secuenciados de la familia Mirtaceae como *Eucalyptus grandis*. El proyecto planteado concluirá con un equipo de investigadores consolidado y experimentado para analizar genomas y transcriptomas de otras especies de la biodiversidad amazónica. Además, se contará con un atlas de genes

potencialmente útiles para realizar mejoramiento genético y desarrollar procesos biotecnológicos para la generación de productos con valor agregado.

Palabras clave: biotecnología, frutal amazónico, genes y genomas, secuenciamiento masivo, transcriptoma.

IV. PLANTEAMIENTO DEL PROBLEMA

La amazonía peruana cuenta con una excepcional biodiversidad, razón por la que nuestro país es considerado uno de los más megabiodiversos del planeta (1–4). Esta megabiodiversidad es el soporte fundamental para los habitantes de las múltiples comunidades y ciudades asentadas en la amazonía. Porque, además de brindarles diversos servicios ecosistémicos, les suministran alimentos, medicinas, materiales para la construcción de viviendas, entre otros múltiples beneficios (5–7). Además, la extracción y comercialización de diversos recursos renovables (madera, peces, fauna, etc) y no renovables (petróleo, gas natural, minerales, etc) les proporcionan importantes ingresos económicos (6) incluido a los gobiernos locales y regionales a través del canon y sobrecanon.

Del conjunto de frutales nativos de la amazonía, el camu-camu (*Myrciaria dubia*) es considerado una especie bandera regional y ha sido priorizada para su investigación científica, tanto básica como aplicada. *M. dubia* se caracteriza por crecer naturalmente en diversas cuencas amazónicas (8) y el gobierno regional ha impulsado su cultivo, mediante créditos agrarios, en varias comunidades por los ingresos económicos que proporciona mediante la comercialización de sus frutos. La creciente demanda de estos frutos, se atribuye al alto contenido de ácido L-ascórbico (AsA, vitamina C) que puede superar los 2 g por cada 100 g de pulpa (9,10). Además, varias investigaciones han demostrado que los frutos, hojas, raíces y semillas producen múltiples compuestos fitoquímicos bioactivos de interés para las industrias alimentarias, farmacológicas, cosméticas, entre otras (11–14).

Sin embargo, el gran potencial de este frutal nativo aún no puede ser aprovechado en plenitud. Esto se debe a varios problemas que necesitan ser resueltos a corto, mediano y largo plazo. Primero, varios reportes muestran que existe una gran variación en el contenido de AsA en los frutos de las plantaciones naturales y cultivadas (9,10,15,16), lo cual impide que los productores ofrezcan frutos con alto y homogéneo contenido de AsA. Segundo, el ataque de plagas a diferentes tejidos de la planta y particularmente a los frutos (p.ej. causadas por *Conotrachelus dubiae*), provoca grandes pérdidas en la producción (17,18) y aún no se disponen de procedimientos para el control de estas plagas. Finalmente, es limitado el uso de tecnologías apropiadas para dar valor agregado a los productos de esta especie, porque comúnmente a nivel local y regional se comercializan los frutos sin ningún procesamiento cuando a nivel nacional e internacional se comercializa como pulpa, néctares, etc.(19).

Varios de los problemas mencionados, particularmente los dos primeros, pueden ser solucionados a mediano plazo con el desarrollo de variedades genéticamente mejoradas de *M. dubia* que muestren alta y homogénea producción de AsA e independiente de factores medioambientales y con resistencia a factores bióticos y abióticos adversos. La mejora de cultivos mediante técnicas tradicionales de cruce y selección puede necesitar decenas de generaciones antes de alcanzar resultados satisfactorios. No obstante la aplicación de técnicas de mejora basadas en metodologías genómicas puede acelerar notablemente este proceso. La disposición de esta información nos permitirá desarrollar marcadores moleculares y genéticos imprescindibles para el mejoramiento genético asistido de *M. dubia*. Además, con los genes identificados podremos implementar procesos biotecnológicos para la biosíntesis de diversos productos altamente cotizados. Cabe resaltar que las ciencias de la vida y las biotecnologías (biología celular y molecular, ingeniería genética, genómica) y la bioinformática, tienen una importancia estratégica para el desarrollo regional y nacional, pues se relacionan con cinco de los siete sectores productivos prioritarios seleccionados y tres de los cuatro sectores sociales (20). Por tanto, en esta investigación planteamos la siguiente pregunta ¿Es factible realizar el análisis estructural y funcional del genoma de *Myrciaria dubia* “camu-camu”? Consecuentemente, los resultados de esta investigación nos permitirán responder esta pregunta y contribuirá significativamente en el desarrollo de una especie como *M. dubia* con un enorme potencial económico.

V. ANTECEDENTES DE LA INVESTIGACIÓN

Tecnologías para el análisis estructural y funcional del genoma

Para realizar los análisis genómicos estructurales y funcionales de las plantas y de otros organismos se emplean tecnologías de secuenciamiento que han sido innovadas recientemente. Por ejemplo, el método de secuenciamiento de terminación de cadena desarrollado a finales de los años 70 (21), y posterior automatización (22), es considerada de “primera generación” y ha predominado por más de 30 años, permitiendo obtener la secuencia de genomas del humano, animales, plantas, bacterias, virus, entre otros (23,24). La limitante de esta tecnología es su elevado costo, mantenimiento y consumibles, de tal manera que emprender proyectos de secuenciamiento de genomas era una exclusividad de consorcios internacionales.

Sin embargo, a mediados del 2000 se desarrollaron nuevos métodos, referidos como secuenciadores de segunda generación. Estas nuevas tecnologías (FLX 454 de Life Sciences/Roche, SOLiD de Applied Biosystems, Illumina Genome Analyzer, etc), aplican varias estrategias combinadas para la preparación de la muestra, secuenciamiento e imagen, alineamiento del genoma y métodos de ensamblado. La mayor ventaja ofrecida por estas tecnologías es la capacidad de secuenciar un gran volumen de información, en algunos casos mayores a 1 billón (0,45 a 90 Gb) de lecturas cortas (25 a 250 bases) por corrida del instrumento, con gran exactitud y a un bajo costo (25–29). Es importante remarcar que, estas nuevas tecnologías se caracterizan por ser multipropósitos, los cuales nos permiten estudiar no sólo el genoma, sino además, el transcriptoma, epigenoma, código de barras molecular y determinar eventos de unión proteínas-ADN en el genoma de cualquier especie (30–32).

Recientemente, algunas compañías han desarrollado secuenciadores de tercera generación, que tienen mayor escala de producción de secuencias y con costos más reducidos. La primera de estas tecnologías, comercializada por Helicos Biosciences (Cambridge, MA USA), emplea una estrategia de secuenciamiento de moléculas simples y se caracteriza por no requerir amplificación del ADN (33) . En el 2011, Pacific Biosciences lanzó la plataforma PacBio RS, que emplea una poderosa tecnología de molécula simple en tiempo real ó tecnología SMRT™ con capacidad de realizar lecturas de fragmentos grandes (≥ 1 kpb) y producir varias gigabases de secuencias en menos de 24 horas (34). Otras compañías como Complete Genomics (35), NABsys (36), ZS Genetics (37) y Oxford

Nanopore Technologies (38) están desarrollando sistemas de secuenciamiento de tercera generación.

Avances en el análisis estructural del genoma de plantas

Los genomas de las plantas son frecuentemente grandes, complejos y poliploides, haciendo que su secuenciamiento sea un enorme desafío (39). El gran tamaño de los genomas de plantas se atribuyen a su duplicación parcial o total y a la amplificación de elementos repetitivos, que representan del 25 al 73% del genoma (40,41). En consecuencia, los dos grandes retos en el secuenciamiento y ensamblaje del genoma de plantas es resolver los elementos repetitivos y discriminar las regiones duplicadas del genoma (28).

Pero estos retos han sido superados y actualmente se ha completado el análisis genómico estructural de varias especies de plantas. Inicialmente los genomas fueron secuenciados con la tecnología estándar desarrollada por Frederick Sanger (21,22), siendo el genoma de la planta modelo *Arabidopsis thaliana* (tamaño de genoma = 125Mb) el primero en ser secuenciado y ensamblado (42). Luego fueron secuenciados los genomas de dos subespecies del arroz *Oryza sativa L. ssp. Japonica* (43) y *Oryza sativa L. ssp. Indica* (44). *Populus trichocarpa*, es la cuarta especie secuenciada y presenta más de 45,000 genes que codifican proteínas (45). Otros genomas de plantas secuenciados con tecnologías de primera generación y publicados desde el 2006 son los de *Medicago truncatula* (24), *Vitis vinífera* (46), *Carica papaya* (47), *Lotus japonicus* (48), *Physcomitrella patens* (49), *Sorghum bicolor* (50), *Zea mays* (51), *Glycine max* (52), *Ricinus communis* (53), *Brachypodium distachyon* (54), *Arabidopsis lyrata* (55) y *Eucalyptus grandis* (56). Esta última especie, al igual que *M. dubia* pertenece a la familia *Myrtaceae*, consecuentemente pueden tener una alta homología.

También, actualmente se dispone de la estructura de genomas de plantas analizadas por la combinación de tecnologías de secuenciadores de primera y segunda generación. Por ejemplo, Huang *et al.* (57) combinando tecnologías de secuenciamiento tradicional con la plataforma de Illumina GA ensamblaron el genoma de *Cucumis sativus*, en base a una cobertura del genoma de 72.2 X. Asimismo, otros investigadores emplearon las plataformas de secuenciamiento tradicional y del GS FLX Titanium de Roche para secuenciar el genoma de *Malus domestica* (58). Otras secuencias de genomas completos generados con estrategias similares se han publicado como el de *Jatropha curcas* (59), *Theobroma cacao* (60), *Fragaria vesca* (61), *Cajanus cajan* (62), *Brassica rapa* (63), *Solanum tuberosum* (64), *Phoenix dactylifera* (65), *Musa acuminata* (66), *Prunus mume* (67), *Triticum aestivum* (68),

Amborella trichopoda (69), *Citrus sinensis* (70), *Elaeis guineensis* (71), *Pyrus bretschneideri* (72), *Triticum urartu* (73), *Camelina sativa* (74), *Beta vulgaris* (75), *Solanum pennellii* (76), *Oryza glaberrima* (77) y *Coffea canephora* (78). En conclusión, con las tecnologías disponibles actualmente es posible realizar el análisis genómico estructural de *M. dubia*, en un tiempo relativamente corto y con un limitado presupuesto,

Avances en el análisis funcional del genoma de plantas

El análisis funcional del genoma o también conocido como análisis del transcriptoma, se refiere al análisis de la expresión del conjunto completo y la cantidad de todos los transcritos en una célula o tejido, para una etapa de desarrollo específico o condición fisiológica. Uno de los objetivos de este tipo de análisis es elaborar un atlas de todos los tipos de transcritos, particularmente de los ARN mensajeros y cuantificar sus niveles de expresión (79–83). Con estas tecnologías de análisis, las moléculas de ARN (total o fraccionada) son convertidas en una librería de fragmentos de ADNc con adaptadores unidos a uno o ambos extremos. Cada molécula, con o sin amplificación, es secuenciada para obtener lecturas cortas (30-400pb). En principio, cualquier plataforma de segunda generación nos permite hacer este proceso y los sistemas de Illumina IG, SOLiD y GS FLX Titanium han sido empleados para este propósito. Después de completar el secuenciamiento, las lecturas son alineadas a un genoma de referencia o transcritos de referencia, o ensamblados *de novo* sin la secuencia genómica, para producir un mapa de transcripción a escala genómica que consta de la estructura transcripcional y/o el nivel de expresión de cada gen (84,85).

Aunque el secuenciamiento del ARN es una nueva tecnología y está desarrollándose activamente, nos proporciona varias ventajas con respecto a las otras tecnologías existentes. Primero, que no está limitado a detectar transcritos que corresponden a una secuencia genómica existente, consecuentemente es particularmente útil para estudiar *M. dubia* u otras especies de la amazonía que no disponen de un genoma de referencia. Además, permite identificar polimorfismos como los SNPs, indels y microsatélites en las regiones transcritas. Adicionalmente, presenta un amplio rango dinámico de niveles de expresión genética, que le permite ser más exacto para cuantificar los niveles de expresión. Finalmente, los resultados del análisis genómico funcional muestran altos niveles de reproducibilidad, tanto para las réplicas técnicas como las biológicas (27,82,84).

Debido a las múltiples ventajas mencionadas de estas nuevas tecnologías, recientemente se ha incrementado notablemente las publicaciones científicas sobre los análisis genómicos funcionales de varias especies de plantas. Por ejemplo, nuestro equipo

ha realizado el análisis del transcriptoma de frutos de *Myrciaria dubia* (86). También, estudios similares se han realizado con otras especies de plantas tales como *Ananas comosus* (87), *Capsicum annuum* (88), *Citrus sinensis* (70), *Litchi chinensis* (89), *Mangifera indica* (90), *Myrica rubra* (91), *Momordica cochinchinensis* (92), *Pyrus bretschneideri* (93), *Vaccinium* spp. (94) y *Ziziphus jujuba* (95). Adicionalmente, aproximaciones similares fueron efectivamente empleadas para realizar el análisis genómico funcional de otros tipos de tejidos y órganos en varias especies de plantas (96–111). Todas estas investigaciones muestran que con las tecnologías de secuenciamiento masivo fácilmente podemos abordar análisis genómicos funcionales relativamente complejos y responder importantes preguntas de investigación como la que planteamos en esta propuesta.

VI. HIPÓTESIS

El análisis estructural y funcional del genoma de *Myrciaria dubia* nos permitirá identificar genes promisorios para el mejoramiento genético y con potencial aplicación para el desarrollo de procesos biotecnológicos de importancia para las industrias alimentarias, farmacéuticas, agroindustrias entre otras.

VII. OBJETIVOS DE LA INVESTIGACIÓN

Objetivo general

Realizar el análisis estructural y funcional del genoma de *Myrciaria dubia* “camu-camu”

Objetivos específicos

Determinar el tamaño del genoma de *M. dubia*

Secuenciar y ensamblar *de novo* el genoma de *M. dubia*

Desarrollar un atlas de expresión génica de *M. dubia*.

VIII. METODOLOGÍA DEL PROYECTO

8.1. Obtención del material botánico

Las hojas, flores y frutos (verdes, pintones y maduros) serán obtenidas de plantas de *M. dubia* de la Colección Nacional de Germoplasma, Campo Experimental El Dorado, Estación Experimental San Roque-Loreto, Instituto Nacional de Innovación Agraria (INIA), ubicado en el Km 25½ de la carretera Iquitos-Nauta (03°57'17" LS, 73°24'55" LO). Esta colección está conformada por 43 accesiones obtenidas de ocho cuencas hidrográficas de la Región Loreto (Nanay, Itaya, Napo, Ucayali, Putumayo, Curaray, Tigre y Amazonas), establecidas hace aproximadamente 20 años.

8.2. Determinación del tamaño del genoma

Para realizar este análisis, primero se separarán las semillas de los frutos maduros. Estas serán lavadas meticulosamente con abundante agua potable y agua destilada. Posteriormente, las semillas serán escarificadas específicamente en el sector donde se ubica el embrión. Luego, las semillas serán puestas entre cuatro capas de gasa embebidas con agua destilada, depositadas en recipientes de plástico e incubadas bajo condiciones de oscuridad a temperatura ambiente por 10 a 14 días. Posteriormente, las plántulas obtenidas serán cultivadas en una solución hidropónica bajo exposición de luz natural y a temperatura ambiente por 3 a 4 semanas. De estas plántulas se obtendrán las hojas y se procederá a estimar el tamaño del genoma de acuerdo a Arumuganathan y Earle (39). Con este protocolo se obtienen núcleos celulares intactos a partir de las hojas de plántulas. Estos núcleos serán teñidos con yoduro de propidio y tratados con ARNasa. Para determinar el contenido de ADN genómico de 5000 a 10000 núcleos teñidos, se empleará un citómetro de flujo FACS scalibur a 532 nm. El valor C monoploide (1C) será calculado y expresado en Mb empleando el factor de conversión de 1 pg de ADN = 978 Mb.

8.3. Secuenciamiento y ensamblado *de novo* del genoma

El ADN genómico será purificado a partir de las hojas de plántulas de *M. dubia* de acuerdo a Castro *et al.* (112). Para la secuenciación y ensamblaje del genoma se usará una estrategia híbrida combinando dos tipos de tecnologías de secuenciamiento. Por un lado se realizará un secuenciamiento usando la tecnología de Illumina (HiSeq™ 2500) de lecturas cortas pareadas con insertos de 300 pb y longitud de lectura de 150 pb (2x150) con una cobertura de 100 veces el tamaño estimado del genoma (100X) para crear una base de secuencias consenso. Dichas lecturas se suplementarán con lecturas pareadas de 100 pb con insertos de 3 y 5 Kb para extender el tamaño del ensamblaje (20X para cada una).

Estas secuencias se ensamblarán usando dos programas, SOAPdeNovo (113) y ABySS (114). Por otro lado, a fin de resolver los elementos repetitivos, se realizará un secuenciamiento de lecturas largas (>5 Kb) usando la tecnología desarrollada por Pacific Bioscience (34). Para integrar estas lecturas, se usarán dos programas, SSPACE-Long (115) para extender las secuencias consenso y PBJelly (116) para completar los gaps entre estas secuencias consenso.

La anotación estructural de los modelos génicos se realizará usando la combinación de programas de predicción *de novo* tales como SNAP (117) o Augustus (118) que se basan en el entrenamiento de programa de análisis de secuencias para predecir ORF con programas de alineamiento de transcritos tales como Blast (119) y Exonerate (120). Ambos resultados serán integrados con Maker-P (121) para generar una predicción consenso. Adicionalmente, las repeticiones serán anotadas con RepeatModeler (122) y RepeatScout (123). La anotación funcional de los modelos génicos se realizará comparando la secuencia de dichos modelos con las secuencias conocidas de bases de datos como GenBank, SwissProt y TAIR usando Blast. Los dominios de proteína serán anotados usando InterProScan (124). Finalmente, todos los datos serán puestos a disposición de los integrantes del grupo a través de Blast (para buscar secuencias homólogas), JBrowse (para explorar los modelos génicos) y Pathway Tools (para explorar las anotaciones metabólicas).

Para conocer la historia evolutiva de *M. dubia* se realizará un análisis comparativo de su genoma con otros genomas vegetales que comprenderá: 1) Análisis de familias génicas con otras *Myrtaceas* secuenciadas (*Eucalyptus grandis*), otras rosidas de otras familias como *Brassicaceas* (*Arabidopsis thaliana*), *Malvaceas* (*Theobroma cacao*), *Leguminosas* (*Glycine max*, *Medicago truncatula*), *astéridas* (*Solanum lycopersicum*, *Solanum tuberosum*) y distintos linajes de monocotiledóneas (*Oryza sativa*, *Zea Mays* y *Musa acuminata*). Este análisis nos permitirá conocer si *M. dubia* posee eventos de poliploidía ancestrales, o si bien existen familias bajo fenómenos de expansión o contracción propios del género *Myrciaria* o de la familia *Myrtaceae*. 2) Análisis de sintenia con especies de la misma familia como *Eucalyptus grandis*, para conocer hasta qué punto se mantiene el orden posicional de genes en el genoma de ambas especies. 3) Análisis de los genes involucrados en el metabolismo de AsA en un entorno comparativo con otras especies, para conocer si la diferencia en el contenido de AsA se debe a una variación estructural (por ejemplo polimorfismos específicos de secuencia o variación en número de copias de un gen) o de dosis génica (expresión diferencial de genes implicados en la biosíntesis de AsA).

8.4. Desarrollar un atlas de expresión génica

El ARN total será purificado de hojas, flores, frutos (verdes, pintones y maduros) y plántulas (raíces, tallos y plúmulas) empleando el método CTAB, extracciones con solventes orgánicos y tratamiento con ADNasa como describe Castro *et al.* (128). La calidad y cantidad del ARN purificado será determinado mediante análisis espectrofotométrico y electroforético (129). Posteriormente, se purificará el ARNm y sintetizará el ADN complementario (ADNc) usando el SuperScript double-stranded cDNA synthesis kit (Invitrogen, CA). El ADNc será usado para construir las librerías y proceder a su secuenciamiento masivo con la plataforma de Illumina HiSeq™ 2000 (Illumina Inc., San Diego, CA, USA) por la empresa Macrogen Inc.

Previo al ensamblado las lecturas de las secuencias “crudas” serán filtradas usando Trimmomatic v0.32 (130) y luego se procederá a ensamblar las secuencias siguiendo la aproximación de múltiples k-mer de Melicher *et al.* (131). Para dilucidar las funciones potenciales de los genes expresados se utilizará el programa en línea FastAnnotator (132). Para reconstruir las vías metabólicas las secuencias asignadas con un código de enzimas serán mapeadas en la base de datos de vías metabólicas del Kyoto Encyclopedia of Genes and Genomes (133). Finalmente, para enriquecer la anotación funcional las secuencias génicas serán enviadas al KEGG Automatic Annotation Server – KAAS (134). En base a todos los análisis bioinformáticos indicados se podrá elaborar el atlas de genes expresados de *M. dubia*.

8.5. Control de calidad y bioseguridad

La calidad de las diferentes actividades en este proyecto de investigación será garantizada con la aplicación de protocolos estandarizados, equipos calibrados, materiales apropiados e insumos químicos grado biología molecular. Asimismo, se seguirán estrictas medidas de bioseguridad en el trabajo de campo y del Laboratorio a fin de minimizar los riesgos que atenten en contra de la salud del equipo de investigadores, practicantes, tesisistas y otras personas de la comunidad directa e indirectamente involucradas. Todas las personas participantes del proyecto recibirán charlas sobre bioseguridad y todas las cartillas de información de seguridad de las sustancias que serán manipuladas se estudiarán minuciosamente. Las sustancias químicas empleadas serán inactivadas antes de su eliminación en base a protocolos que se implementarán.

IX. METAS POR COMPONENTES

Componentes	Metas	
	Primer año	Segundo año
Determinación del tamaño del genoma	Se ha determinado el tamaño del genoma de <i>M. dubia</i> al 100%	
Secuenciamiento y ensamblado de novo del genoma	Se ha secuenciado y ensamblado de novo el genoma de <i>M. dubia</i> en un 75%	Se ha secuenciado y ensamblado de novo el genoma de <i>M. dubia</i> en un 100%
Desarrollo de un atlas de expresión génica		Se ha desarrollado el atlas de genes expresados de <i>M. dubia</i> en un 100%
Formación de recursos humanos	Se ha contribuido con la formación de 4 voluntarios, 4 practicantes, 2 tesis de pre-grado y 1 tesis de Maestría. También, se ha realizado 2 cursos teóricos-prácticos	Se ha contribuido con la formación de 4 voluntarios, 4 practicantes, 2 tesis de pre-grado y 1 tesis de Doctorado. Además, se ha realizado 2 cursos teóricos-prácticos
Difusión de resultados	Se ha expuesto los avances de las investigaciones en 2 encuentros científicos internacionales, se ha enviado 2 artículos para su publicación en revista nacional o internacional indexada	Se ha expuesto los avances de las investigaciones en 1 encuentro científico internacional, se ha enviado 2 artículos para su publicación en revista nacional o internacional indexada

X. ASPECTOS ÉTICOS

Todos los miembros del equipo están comprometidos en salvaguardar la información generada en esta investigación, manteniendo la confidencialidad de la misma. Asimismo, garantizaremos que la información generada sea fidedigna.

XI. RESULTADOS ESPERADOS

Al término del Proyecto se ha completado el análisis genómico estructural y funcional de *M. dubia*. Por tanto, con el desarrollo de una base de datos genómica con acceso público de los resultados y la entrega de datos crudos a repositorios públicos como GenBank tal como se solicita en las publicaciones científicas internacionales. Con todos estos aportes estaremos contribuyendo a sentar las bases para el desarrollo de variedades mejoradas asistida con marcadores genéticos y para la implementación de procesos biotecnológicos. Además, de contribuir con la formación del recurso humano, se ha consolidado una red científica constituida por un equipo multidisciplinario e interinstitucional de investigadores potencializada en el análisis de genomas y transcriptomas de especies vegetales de la biodiversidad amazónica.

XII. ESTRATEGIAS A UTILIZAR PARA LA TRANSFERENCIA Y COMUNICACIÓN DE LOS RESULTADOS

La transferencia de los conocimientos científicos generados como resultado de la investigación se hará a través de los artículos científicos que se publiquen en las revistas especializadas indexadas nacionales e internacionales. También, presentando los resultados de la investigación en eventos científicos como: congresos, simposios y mesas redondas. Además, los resultados y publicaciones estarán disponibles en el portal web de la UNAP y de las Instituciones participantes. Finalmente, el conocimiento generado se impartirá a los estudiantes de las diferentes facultades en las disciplinas de Bioquímica, Biología Celular y Molecular, Procesos Biotecnológicos y Genética.

XIII. IMPACTOS ESPERADOS

a) Mejora de las capacidades técnicas

El desarrollo de este Proyecto contribuirá de manera significativa en la mejora de nuestras capacidades técnicas y científicas. Porque las diversas actividades previstas incluyen el uso de tecnologías de última generación y manejo de nuevas herramientas bioinformáticas. Consecuentemente, el capital humano será potencializado y fortalecido con estas nuevas tecnologías y herramientas que nos permitirán mejorar el nivel de nuestras investigaciones. Asimismo, nuestra red de colaboración científica nacional e internacional con el Instituto Nacional de Innovación Agraria, Universidad Científica del Perú, Virginia Polytechnic Institute and State University (VPISU) y la Universidad de Málaga será robustecida y propiciará el desarrollo de nuevas investigaciones con *M. dubia* y otras especies de nuestra biodiversidad amazónica.

b) Formación de investigadores jóvenes

Nosotros tenemos claro que nuestra región y el país requieren de más investigadores que contribuyan a dar solución a los múltiples problemas que nos agobian. De acuerdo a ello, nuestro equipo ha asumido la responsabilidad de contribuir con la formación de jóvenes investigadores. Por eso, tenemos la política de captar a jóvenes desde los primeros niveles de estudios universitarios de las Facultades de Ciencias Biológicas, Ciencias Forestales y áreas afines a la biología. Varios de estos jóvenes inician su carrera científica en la Unidad Especializada de Biotecnología como voluntarios, posteriormente realizan sus prácticas pre-profesionales (I y II) y finalmente sus tesis. Este proceso, que toma varios años garantiza una sólida formación de nuestros futuros científicos.

c) Tesis de pregrado y/o postgrado

Una de nuestras misiones como docentes e investigadores de la UNAP es realizar el asesoramiento de tesis de pre y post-grado, actividad que constantemente venimos realizando en la Unidad Especializada de Biotecnología del CIRNA. En concordancia con esta misión, en este Proyecto brindaremos el apoyo para la realización de 4 tesis de pre-grado y 2 tesis de post-grado (1 de Maestría y 1 de Doctorado). Este apoyo es permanente y va desde la elaboración y presentación del anteproyecto de tesis, asesoramiento en el desarrollo de experimentos y trabajos de Laboratorio, redacción del informe final de tesis y sustentación. Con este acompañamiento permanente a los tesisistas, nosotros hemos logrado un 100% de éxito y debido a ello contamos con más de 20 tesisistas que fueron entrenados en nuestro laboratorio y varios de ellos actualmente están realizando estudios de post-grado en el extranjero.

d) N° de personas capacitadas

Se estima que con el desarrollo de los 4 cursos teóricos-prácticos previstos en el Proyecto se capacitará a un aproximado de 200 personas. Además, teniendo en cuenta que los resultados de nuestras investigaciones fortalecerán las actividades académicas de dictado de cursos de bioquímica, procesos biotecnológicos y genética, entonces estaremos contribuyendo con la capacitación de unos 400 estudiantes.

e) Acceso para servicios especializados

Para el desarrollo de tres componentes del Proyecto (Determinación del tamaño del genoma, Secuenciamiento y ensamblado *de novo* del genoma y Elaboración de un atlas de genes expresados) se requerirá el acceso a servicios especializados como el análisis por citometría de flujo, el secuenciamiento masivo con la plataforma de Illumina HiSeq™ 2000 y el análisis con nuevas herramientas bioinformáticas del Department of Horticulture (DH), Virginia Polytechnic Institute and State University (VPISU).

f) Publicaciones, eventos científicos

Nuestro equipo de investigadores constantemente publica los resultados de sus investigaciones científicas en revistas nacionales e internacionales indexadas. Asimismo, frecuentemente ha participado en encuentros científicos internacionales exponiendo nuestros trabajos de investigación. De acuerdo con esta tradición, durante la ejecución de este proyecto se tiene planificado publicar en revistas nacionales e internacionales por lo menos 4 artículos científicos y participar en por lo menos 3 encuentros científicos internacionales que se desarrollan en la ciudad de Lima.

g) Generación de conocimientos o producción de nuevas tecnologías

Al conocer la estructura y funcionamiento del genoma de *M. dubia*, será posible identificar, aislar y caracterizar todos los genes promisorios para el mejoramiento genético y desarrollo de procesos biotecnológicos. Asimismo, se podrá implementar estudios básicos complementarios relacionados a la genómica funcional, tales como el estudio del proteoma y metaboloma. Además, permitirá el desarrollo de millones de marcadores moleculares (microsatélites, SNPs, etc) que serán fundamentales para realizar estudios de diversidad genética, estudios de asociación de marcadores moleculares con características fenotípicas y permitirá establecer nuevas estrategias de mejoramiento genético.

En el ámbito socioeconómico, nuestros resultados darán el soporte para el desarrollo de investigaciones aplicadas para solucionar problemas puntuales que afectan a los agricultores y que dependen de la producción de *M. dubia*. Por ejemplo, al desarrollar programas de mejoramiento genético para crear variedades de plantas con mayor producción, mejor calidad de frutos y resistentes a factores bióticos (p.ej. resistentes al ataque de la plaga *Conotrachelus dubiae*) y/o abióticos adversos (p.ej. sequías, acidez del pH, etc). En consecuencia, de manera indirecta, esta investigación contribuirá a mejorar la calidad de vida de los pobladores amazónicos que dependen de este importante recurso.

XIV. INFRAESTRUCTURA Y EQUIPOS A UTILIZARSE

La **Unidad Especializada de Biotecnología del CIRNA** tiene equipos para realizar técnicas bioquímicas y de biología molecular consideradas en el Proyecto, tales como PCR en Tiempo Real, Microcentrífugas refrigeradas, Fluorómetro de microplacas, Citómetro de flujo (para analizar células y determinar el tamaño de genomas), Equipo de HPLC, Analizador Genético 3130XL (para el secuenciamiento del ADN), equipos de electroforesis, y un Sistema de Fotoregistro BioDocAnalyze.

El **Department of Horticulture (DH), Virginia Polytechnic Institute and State University (VPISU)** cuenta con equipos necesarios para la preparación de muestras para secuenciación masiva tales como un fluorímetro Qubit para la medida exacta de la concentración de ácidos nucleicos, un equipo de electroforesis BioAnalyzer para la medida exacta de los tamaños de librerías para secuenciación, así como cámaras de electroforesis BioRad, centrifugas, termocicladores y otros equipos necesarios para la purificación y manipulación de ácidos nucleicos. Asimismo, dispone del equipo necesario para el análisis bioinformático consistente en dos servidores con 64 núcleos y 256 y 128 Gb de memoria RAM y 7 y 3 Tb de disco duro para almacenamiento respectivamente.

El **Departamento de Biología Molecular y Bioquímica (DBMB), Universidad de Málaga** cuenta con equipos para realizar análisis bioquímicos y de biología molecular, como espectrofotómetros UV/Visible, microcentrífugas, sistemas de electroforesis, equipos para el aislamiento y la clonación molecular de genes, entre otros equipos menores.

El **Laboratorio de Biotecnología y Bioenergética de la UCP** cuenta con equipos para purificación y análisis de ácidos nucleicos tales como Centrífuga Universal Refrigerada, Espectrofotómetro NanoDrop 2000, cabina de bioseguridad, entre otros equipos menores.

La **Estación Experimental Agraria San Roque del INIA**, cuenta con la Colección Nacional de Germoplasma de *M. dubia*, constituido por 43 accesiones (muestras representativas de la variabilidad genética de esta especie) obtenidos de 43 lugares diferentes de la amazonía peruana pertenecientes a 8 principales cuencas hidrográficas de la Región Loreto (Nanay, Itaya, Napo, Ucayali, Putumayo, Curaray, Tigre y Amazonas).

XV. CRONOGRAMA DE ACTIVIDADES POR COMPONENTES MENSUALIZADO Y RESPONSABLES

Componentes/Actividades/Responsables	Meses																								
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	
Determinación del tamaño del genoma																									
Colecta de material botánico/SI	■	■																							
Germinación de semillas y obtención de plántulas/SI		■	■																						
Análisis por citometría de flujo de los núcleos celulares/MC, JM			■	■	■	■																			
Secuenciación y ensamblado de novo del genoma																									
Purificación del ADN genómico a partir de plántulas/JM, MC						■	■																		
Secuenciar el genoma/AB, MB, JC							■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Ensamblar <i>de novo</i> y anotar el genoma/AB, MB, JC								■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Desarrollo de un atlas de expresión génica																									
Purificación de ARN total y ARNm de diferentes tejidos/JM, MC													■	■	■	■	■	■	■	■	■	■	■	■	■
Síntesis del ADNc y secuenciación/AB, MB, JC																									
Ensamblado y anotación del transcriptoma/AB, MB, JC																									
Formación de recursos humanos																									
Adiestramiento de voluntarios y practicantes/JM, MC, JC		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Realización de tesis de pre- y post-grado/JC, JM, MC	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Realización de cursos teóricos-prácticos/JC, JM, MC, SI	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Difusión de resultados																									
Participación como ponentes en encuentros científicos/JC, JM								■								■					■				
Publicación de artículos científicos en revistas indexadas/Todos												■													■

Leyenda: AB: Aureliano Bombarely, JC: Juan Castro, JM: Jorge Marapara, MB: Miguel Botella, MC: Marianela Cobos, SI: Sixto Imán

XVI. RESUMEN DEL PRESUPUESTO POR RUBROS, PARTIDAS Y COMPONENTES EN SOLES

Presupuesto por rubros

RUBRO	MONTO (S/.)	%
a) Desarrollo de los componentes y actividades	42,000.00	60.00
b) Movilidad local para el equipo de investigadores	6,000.00	8.57
c) Pasajes y viáticos para capacitación	3,000.00	4.29
d) Adquisición de accesorios para equipos	7,500.00	10.70
e) Mantenimiento preventivo y reparación de equipos	3,000	4.29
f) Servicios de terceros vinculados a las actividades del Proyecto	1,000	1.43
g) Pago para adquirir bibliografía especializada	1,500	2.14
h) Organización de seminarios, talleres y similares	3,000.00	4.29
i) Gastos operativos del Proyecto	3,000.00	4.29
TOTAL S/.	70,000.00	100.00

Presupuesto por partidas genéricas y componentes

PARTIDA	DESCRIPCIÓN	COMPONENTES					TOTAL (S/.)
		Determinación del tamaño del genoma	Secuenciamiento y ensamblado de novo del genoma	Desarrollo de un atlas de expresión génica	Formación de recursos humanos	Difusión de resultados	
23.12.11	Vestuario, accesorios y prendas diversas	100.00	100.00	100.00	0.00	0.00	300.00
23.15.12	De oficina	100.00	100.00	100.00	200.00	300.00	800.00
23.15.31	Aseo, limpieza y tocador	100.00	500.00	900.00	0.00	0.00	1,500.00
23.15.41	Electricidad, Iluminación y electrónica	100.00	100.00	100.00	0.00	0.00	300.00
23.16.199	Otros accesorios y repuestos para maquinari:	2,500.00	2,500.00	2,500.00	0.00	0.00	7,500.00
23.18.12	Medicamentos	0.00	0.00	0.00	100.00	0.00	100.00
23.18.21	Material, insumos, instrumental y acc. Medic	2,000.00	5,000.00	5,000.00	0.00	0.00	12,000.00
23.19.11	Libros, textos y otros materiales impresos	0.00	500.00	500.00	500.00	0.00	1,500.00
23.21.21	Pasajes y gastos de transporte	0.00	0.00	0.00	0.00	1,500.00	1,500.00
23.21.22	Viaticos y asignaciones por comisión de servi	0.00	0.00	0.00	0.00	1,500.00	1,500.00
23.21.299	Movilidad Local	2,000.00	2,000.00	2,000.00	0.00	0.00	6,000.00
23.22.44	servicio de impresiones, encuadernación y er	0.00	0.00	0.00	500.00	500.00	1,000.00
23.24.15	Servicios de mantenimiento, acondicionamier	0.00	1,500.00	1,500.00	0.00	0.00	3,000.00
23.27.101	Seminarios, talleres y similares	0.00	0.00	0.00	3,000.00	0.00	3,000.00
23.27.1199	Servicios diversos	0.00	10,000.00	20,000.00	0.00	0.00	30,000.00
TOTAL (S/.)		6,900.00	22,300.00	32,700.00	4,300.00	3,800.00	70,000.00

Presupuesto por componentes

COMPONENTE	UNIDAD DE MEDIDA	CANTIDAD	MONTO (S/.)
Determinación del tamaño del genoma	Informe	1	6,900.00
Secuenciamiento y ensamblado de novo del genoma	Base de datos	1	22,300.00
Desarrollo de un atlas de expresión génica	Base de datos	1	32,700.00
Formación de recursos humanos	Informes	8 voluntarios	4,300
	Informes	8 practicantes	
	Tesis	4 tesis de pre-grado	
	Tesis	1 tesis de Maestría	
	Tesis	1 tesis de Doctorado	
	Afiches	4 cursos teóricos-prácticos	
	Difusión de resultados	Certificados	
	Artículos	4 artículos científicos	
TOTAL S/.			70,000.00

VII. MONITOREO Y EVALUACIÓN: MATRIZ DE MARCO LÓGICO

OBJETIVOS	INDICADORES VERIFICABLES	MEDIOS DE VERIFICACIÓN	SUPUESTOS DE IMPORTANCIA
<p>FIN</p> <p>Desarrollar variedades mejoradas de <i>M. dubia</i>, caracterizadas por su alto y homogéneo contenido de vitamina C en sus frutos y por ser resistentes a factores bióticos y abióticos adversos</p>	Existencia de cultivos con variedades mejoradas de <i>M. dubia</i> e incremento en la demanda por los mercados de sus productos generados en condiciones naturales y mediante procesos biotecnológicos.	Registros estadísticos de SUNAT, ADEX, INEI, Artículos científicos publicados en revistas nacionales e internacionales indexadas, libros, tesis de pre y post-grado, informes	Apoyo decidido y priorizado de las autoridades competentes para la realización de investigaciones científicas básicas y aplicadas. <i>M. dubia</i> es una de las especies banderas de la región y el país.
<p>PROPÓSITO</p> <p>Realizar el análisis estructural y funcional del genoma de <i>M. dubia</i> “camu-camu”</p>	Base de datos con información del análisis estructural y funcional del genoma. Se dispone de un atlas de genes e identificados aquellos con potencial uso como marcadores para el mejoramiento genético asistidos y desarrollo de procesos biotecnológicos.	Bases de datos, artículos publicados en revistas nacionales e internacionales indexadas, cuadernos de laboratorio, informes de prácticas pre-profesionales, tesis de pre y post grado publicadas. Resúmenes de eventos científicos, posters científicos y memorias anuales del Vicerrectorado de Investigación	Priorización de la investigación científica en la UNAP. Desembolso oportuno de los fondos, importación de productos en corto tiempo, disponibilidad de materiales y reactivos, existe agilidad en los trámites administrativos. Condiciones eléctricas adecuadas.
<p>RESULTADOS/PRODUCTOS</p> <p>Determinación del tamaño del genoma</p>	Se ha determinado el tamaño del genoma de <i>M. dubia</i> en un 100% al finalizar el primer año del Proyecto	Cuadernos de laboratorio, informes de prácticas pre-profesionales, registros de datos del citómetro de flujo, artículos publicados	Se dispone de cantidades suficientes de materiales y reactivos, no hay recorte presupuestal
Secuenciamiento y ensamblado de novo del genoma	Se ha secuenciado y ensamblado de novo el genoma de <i>M. dubia</i> en un 100%, por tanto, al segundo año de ejecución del Proyecto conocemos cómo es la estructura del genoma de esta especie	Bases de datos, artículos publicados en revistas nacionales e internacionales indexadas, cuadernos de laboratorio, posters científicos y memorias anuales del Vicerrectorado de Investigación	Importación adecuada y rápida de reactivos y materiales, condiciones eléctricas adecuadas, equipos están en buen estado
Desarrollo de un atlas de expresión génica	Al final el segundo año de ejecución del Proyecto se ha desarrollado un atlas de expresión génica de <i>M. dubia</i> en un 100%.	Bases de datos, artículos publicados en revistas nacionales e internacionales indexadas, cuadernos de laboratorio, informes de prácticas pre-profesionales, tesis de pre y post grado publicadas.	No hay recorte presupuestal, las autoridades priorizan el apoyo de la investigación, la importación de reactivos y materiales es rápida, condiciones eléctricas adecuadas, equipos están en buen estado
Formación de recursos humanos	Al finalizar el proyecto se habrá adiestrado a 8 voluntarios y 8 practicantes. Además, 4 tesis de pre y 2 de post-grado han sustentado sus tesis. También, se habrán realizado 4 cursos teóricos-prácticos	Tesis de pre y post-grado sustentadas y publicadas, informes de prácticas pre-profesionales, cuadernos de laboratorio, artículos publicados en revistas nacionales e internacionales indexadas	Las autoridades priorizan el apoyo de la investigación, reactivos y materiales en buen estado, condiciones eléctricas adecuadas, equipos están en buen estado
Difusión de resultados	Al cabo del segundo año del Proyecto se ha participado como ponentes en 3 encuentros científicos internacionales y se tendrá 4 artículos en proceso de publicación en revistas nacionales e internacionales indexadas. Además, 6 tesis (4 de pre-grado y 2 de post-grado) han sido sustentadas y publicadas	Certificados como ponentes de 3 encuentros científicos, 4 artículos publicados o en proceso de publicación, Tesis de pre y post-grado sustentadas y publicadas, informes de prácticas pre-profesionales, cuadernos de laboratorio	Apoyo de las autoridades para la difusión de los resultados, desembolso oportuno de los fondos

OBJETIVOS	INDICADORES VERIFICABLES	MEDIOS DE VERIFICACIÓN	SUPUESTOS DE IMPORTANCIA
ACTIVIDADES			
Determinación del tamaño del genoma Colecta de material botánico Germinación de semillas y obtención de plántulas Análisis por citometría de flujo de los núcleos celulares	El tamaño del genoma de <i>M. dubia</i> ha sido determinado al 100% en el primer año de investigación. El presupuesto requerido es de S/. 6,900.00	Cuadernos de laboratorio, informes de prácticas pre-profesionales, registros de datos del citómetro de flujo, artículos publicados	Se dispone de cantidades suficientes de materiales y reactivos, no hay recorte presupuestal
Secuenciamiento y ensamblado de novo del genoma Purificación del ADN genómico a partir de plántulas Secuenciar el genoma Ensamblar de novo y anotar el genoma	Al segundo año de ejecución del Proyecto se conoce la estructura del genoma de <i>M. dubia</i> en un 100%. El presupuesto requerido es de S/. 22,300.00	Bases de datos, artículos publicados en revistas nacionales e internacionales indexadas, cuadernos de laboratorio, posters científicos y memorias anuales del Vicerrectorado de Investigación	Importación adecuada y rápida de reactivos y materiales, condiciones eléctricas adecuadas, equipos están en buen estado
Desarrollo de un atlas de expresión génica Purificación de ARN total y ARNm de diferentes tejidos Síntesis del ADNc y secuenciamiento Ensamblado y anotación del transcriptoma	Al término del Proyecto se ha desarrollado al 100% un atlas de expresión génica de <i>M. dubia</i> . El presupuesto requerido es de S/. 32,700.00	Bases de datos, artículos publicados en revistas nacionales e internacionales indexadas, cuadernos de laboratorio, informes de prácticas pre-profesionales, tesis de pre y post grado publicadas.	Las autoridades priorizan el apoyo de la investigación, la importación de reactivos y materiales es rápida, condiciones eléctricas adecuadas, equipos están en buen estado
Formación de recursos humanos Adiestramiento de voluntarios y practicantes Realización de tesis de pre- y post-grad Realización de cursos teóricos-prácticos	Al finalizar el proyecto se ha adiestrado a 8 voluntarios y 8 practicantes. Además, 4 tesis de pre y 2 de post-grad han sustentado sus tesis. También, se habrán realizado 4 cursos teóricos-prácticos. El presupuesto requerido es de S/. 4,300.00	Tesis de pre y post-grad sustentadas y publicadas, informes de prácticas pre-profesionales, cuadernos de laboratorio, artículos publicados en revistas nacionales e internacionales indexadas	Las autoridades priorizan el apoyo de la investigación, reactivos y materiales en buen estado, condiciones eléctricas adecuadas, equipos están en buen estado
Difusión de resultados Participación como ponentes en encuentros científicos Publicación de artículos científicos en revistas indexadas	Al finalizar el Proyecto hemos participado como ponentes en 3 encuentros científicos internacionales y se tendrá 4 artículos en proceso de publicación en revistas nacionales e internacionales indexadas. Además, 6 tesis (4 de pre-grad y 2 de post-grad) han sido sustentadas y publicadas. El presupuesto requerido es de S/. 3,800.00	Certificados como ponentes de 3 encuentros científicos, 4 artículos publicados o en proceso de publicación, Tesis de pre y post-grad sustentadas y publicadas, informes de prácticas pre-profesionales, cuadernos de laboratorio	Apoyo de las autoridades para la difusión de los resultados, desembolso oportuno de los fondos

XVIII. REFERENCIAS BIBLIOGRÁFICAS

1. Brooks TM, Mittermeier RA, Fonseca GAB da, Gerlach J, Hoffmann M, Lamoreux JF, et al. Global Biodiversity Conservation Priorities. *Science*. 2006;313(5783):58-61.
2. Rodríguez LO, Young KR. Biological Diversity of Peru: Determining Priority Areas for Conservation. *AMBIO J Hum Environ*. 2000;29(6):329-37.
3. Hoorn C, Wesselingh FP, Steege H ter, Bermudez MA, Mora A, Sevink J, et al. Amazonia Through Time: Andean Uplift, Climate Change, Landscape Evolution, and Biodiversity. *Science*. 2010;330(6006):927-31.
4. Wittmann F, Schöngart J, Montero JC, Motzer T, Junk WJ, Piedade MTF, et al. Tree species composition and diversity gradients in white-water forests across the Amazon Basin. *J Biogeogr*. 2006;33(8):1334-47.
5. Stagegaard J, Sørensen M, Kvist LP. Estimations of the importance of plant resources extracted by inhabitants of the Peruvian Amazon flood plains. *Perspect Plant Ecol Evol Syst*. 2002;5(2):103-22.
6. Gavin MC, Anderson GJ. Socioeconomic predictors of forest use values in the Peruvian Amazon: A potential tool for biodiversity conservation. *Ecol Econ*. 2007;60(4):752-62.
7. Balick MJ, Elisabetsky E, Laird SA. Medicinal Resources of the Tropical Forest: Biodiversity and Its Importance to Human Health. Columbia University Press; 1996. 460 p.
8. Imán, SA, Pinedo, S, Melchor, M. Caracterización morfológica y evaluación de la colección nacional de germoplasma de camu camu *Myrciaria dubia* (H.B.K) Mc Vaugh, del INIA Loreto-Perú. *Sci Agropecu*. 2011;2:189-201.
9. Alves R, Filgueiras H, Moura C, Araujo N, Almeida A. Camu camu (*Myrciaria dubia* Mc Vaugh): a Rich Natural Source of Vitamin C. *Proc Interamer Soc Trop Hort*. 2002;46:11-3.
10. Imán S, Bravo L, Sotero V, Oliva C. Contenido de vitamina C en frutos de camu camu *Myrciaria dubia* (H.B.K) Mc Vaugh, en cuatro estados de maduración, procedentes de la Colección de Germoplasma del INIA Loreto, Perú. *Sci Agropecu*. 2011;2(3):123-30.
11. Akachi T, Shiina Y, Kawaguchi T, Kawagishi H, Morita T, Sugiyama K. 1-methylmalate from camu-camu (*Myrciaria dubia*) suppressed D-galactosamine-induced liver injury in rats. *Biosci Biotechnol Biochem*. 2010;74(3):573-8.
12. Akter MS, Oh S, Eun J-B, Ahmed M. Nutritional compositions and health promoting phytochemicals of camu-camu (*Myrciaria dubia*) fruit: A review. *Food Res Int*. 2011;44(7):1728-32.
13. Da Silva FC, Arruda A, Ledel A, Dauth C, Romão NF, Viana RN, et al. Antigenotoxic effect of acute, subacute and chronic treatments with Amazonian camu-camu (*Myrciaria dubia*) juice on mice blood cells. *Food Chem Toxicol Int J Publ Br Ind Biol Res Assoc*. 2012;50(7):2275-81.
14. Fracassetti D, Costa C, Moulay L, Tomás-Barberán FA. Ellagic acid derivatives, ellagitannins, proanthocyanidins and other phenolics, vitamin C and antioxidant capacity of two powder products from camu-camu fruit (*Myrciaria dubia*). *Food Chem*. 2013;139(1-4):578-88.
15. Castro JC, Gutiérrez F, Acuña C, Cerdeira LA, Tapullima A, Marianela C, et al. Variación del contenido de vitamina C y antocianinas en *Myrciaria dubia* «camu-camu». *Rev Soc Quím Perú*. 2013;79(4):319-30.
16. Justi KC, Visentainer JV, Evelázio de Souza N, Matsushita M. Nutritional composition and vitamin C stability in stored camu-camu (*Myrciaria dubia*) pulp. *Arch Latinoam Nutr*. 2000;50(4):405-8.
17. Delgado C, Couturier G, Fine PVA. Survival of Seasonal Flooding in the Amazon by the Terrestrial Insect *Conotrachelus dubiae* O'Brien & Couturier (Coleoptera: Curculionidae), a Pest of the Camu-Camu Plant, *Myrciaria dubia* (Myrtaceae). *Neotrop Entomol*. 2014;43(4):380-4.

18. Farro S, Pinedo M. Possible factors which produce fruit drop of *Myrciaria dubia* (HBK) Mc Vaugh, «camu camu» during the reproductive phenology in the collection «cinco cuencas» from the San Miguel experimental center-IIAP, Loreto, Peru. *Sci Agropecu.* 2010;1(2):117-23.
19. SIICEX. Reportes de Productos de Biocomercio [Internet]. [citado 19-05-15]. Recuperado a partir de: <http://www.siicex.gob.pe/siicex/apb/ReporteProducto.aspx?psector=1025&preporte=prodmercvolu&pvalor=1920>
20. Consejo Nacional de Ciencia, Tecnología e Innovación Tecnológica CONCYTEC. Plan Nacional Estratégico de Ciencia, Tecnología e Innovación para la Competitividad y el Desarrollo Humano 2006 – 2021. CONCYTEC; 2006.
21. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A.* 1977;74(12):5463-7.
22. Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR, et al. Fluorescence detection in automated DNA sequence analysis. *Nature.* 1986;321(6071):674-9.
23. Home - Genome - NCBI [Internet]. [citado 20 de mayo de 2015]. Recuperado a partir de: <http://www.ncbi.nlm.nih.gov/genome/>
24. All Eukaryotic Genomes Annotated at NCBI [Internet]. [citado 20 de mayo de 2015]. Recuperado a partir de: http://www.ncbi.nlm.nih.gov/genome/annotation_euk/all/
25. Metzker ML. Sequencing technologies - the next generation. *Nat Rev Genet.* 2010;11(1):31-46.
26. Varshney RK, Nayak SN, May GD, Jackson SA. Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends Biotechnol.* 2009;27(9):522-30.
27. Lister R, Gregory BD, Ecker JR. Next is now: new technologies for sequencing of genomes, transcriptomes, and beyond. *Curr Opin Plant Biol.* 2009;12(2):107-18.
28. Imelfort M, Edwards D. De novo sequencing of plant genomes using second-generation technologies. *Brief Bioinform.* 2009;10(6):609-18.
29. Berkman PJ, Lai K, Lorenc MT, Edwards D. Next-generation sequencing applications for wheat crop improvement. *Am J Bot.* 2012;99(2):365-71.
30. SOLiD Next Generation Sequencing | Life Technologies [Internet]. [citado 20 de mayo de 2015]. Recuperado a partir de: <http://www.lifetechnologies.com/pe/en/home/life-science/sequencing/next-generation-sequencing/solid-next-generation-sequencing.html>
31. Products - GS FLX+ System : 454 Life Sciences, a Roche Company [Internet]. [citado 20 de mayo de 2015]. Recuperado a partir de: <http://454.com/products/gs-flx-system/>
32. HiSeq 2500 Ultra-High-Throughput Sequencing System [Internet]. [citado 20 de mayo de 2015]. Recuperado a partir de: http://www.illumina.com/systems/hiseq_2500_1500.html
33. Thompson JF, Steinmann KE. Single Molecule Sequencing with a HeliScope Genetic Analysis System. *Current Protocols in Molecular Biology* [Internet]. John Wiley & Sons, Inc.; 2001 [citado 21-05-15]. Recuperado a partir de: <http://onlinelibrary.wiley.com/doi/10.1002/0471142727.mb0710s92/abstract>
34. Pacific Biosciences: Home www.pacificbiosciences.com [Internet]. [citado 20 de mayo de 2015]. Recuperado a partir de: <http://www.pacificbiosciences.com/>
35. Complete Genomics [Internet]. [citado 20 de mayo de 2015]. Recuperado a partir de: <http://www.completegenomics.com/>
36. Nabsys - Nabsys | Introducing Positional Sequencing for DNA Analysis [Internet]. [citado 20 de mayo de 2015]. Recuperado a partir de: <http://www.nabsys.com/>

37. ZS Genetics | Working At The Scale of Life [Internet]. [citado 20 de mayo de 2015]. Recuperado a partir de: <http://www.zsgenetics.com/>
38. Oxford Nanopore Technologies [Internet]. [citado 20 de mayo de 2015]. Recuperado a partir de: <https://www.nanoporetech.com/>
39. Arumuganathan K, Earle ED. Nuclear DNA content of some important plant species. *Plant Mol Biol Report*. 1991;9(3):208-18.
40. Zuccolo A, Sebastian A, Talag J, Yu Y, Kim H, Collura K, et al. Transposable element distribution, abundance and role in genome size variation in the genus *Oryza*. *BMC Evol Biol*. 2007;7:152.
41. Meyers BC, Tingey SV, Morgante M. Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. *Genome Res*. 2001;11(10):1660-76.
42. Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*. 2000;408(6814):796-815.
43. Goff SA, Ricke D, Lan T-H, Presting G, Wang R, Dunn M, et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science*. 2002;296(5565):92-100.
44. Yu J, Hu S, Wang J, Wong GK-S, Li S, Liu B, et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science*. 2002;296(5565):79-92.
45. Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, et al. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Lawrence Berkeley Natl Lab [Internet]. 2006 [citado 16 de mayo de 2014]; Recuperado a partir de: <http://escholarship.org/uc/item/3101x2rn#page-2>
46. Jaillon O, Aury J-M, Noel B, Policriti A, Clepet C, Casagrande A, et al. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*. 2007;449(7161):463-7.
47. Ming R, Hou S, Feng Y, Yu Q, Dionne-Laporte A, Saw JH, et al. The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). *Nature*. 2008;452(7190):991-6.
48. Sato S, Nakamura Y, Kaneko T, Asamizu E, Kato T, Nakao M, et al. Genome structure of the legume, *Lotus japonicus*. *DNA Res Int J Rapid Publ Rep Genes Genomes*. 2008;15(4):227-39.
49. Rensing SA, Lang D, Zimmer AD, Terry A, Salamov A, Shapiro H, et al. The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. *Science*. 2008;319(5859):64-9.
50. Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, et al. The Sorghum bicolor genome and the diversification of grasses. *Nature*. 2009;457(7229):551-6.
51. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, et al. The B73 maize genome: complexity, diversity, and dynamics. *Science*. 2009;326(5956):1112-5.
52. Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, et al. Genome sequence of the palaeopolyploid soybean. *Nature*. 2010;463(7278):178-83.
53. Chan AP, Crabtree J, Zhao Q, Lorenzi H, Orvis J, Puiu D, et al. Draft genome sequence of the oilseed species *Ricinus communis*. *Nat Biotechnol*. 2010;28(9):951-6.
54. International Brachypodium Initiative. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature*. 2010;463(7282):763-8.
55. Hu TT, Pattyn P, Bakker EG, Cao J, Cheng J-F, Clark RM, et al. The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nat Genet*. 2011;43(5):476-81.

56. Myburg AA, Grattapaglia D, Tuskan GA, Hellsten U, Hayes RD, Grimwood J, et al. The genome of *Eucalyptus grandis*. *Nature*. 2014;510(7505):356-62.
57. Huang S, Li R, Zhang Z, Li L, Gu X, Fan W, et al. The genome of the cucumber, *Cucumis sativus* L. *Nat Genet*. 2009;41(12):1275-81.
58. Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, et al. The genome of the domesticated apple (*Malus x domestica* Borkh.). *Nat Genet*. 2010;42(10):833-9.
59. Sato S, Hirakawa H, Isobe S, Fukai E, Watanabe A, Kato M, et al. Sequence analysis of the genome of an oil-bearing tree, *Jatropha curcas* L. *DNA Res Int J Rapid Publ Rep Genes Genomes*. 2011;18(1):65-76.
60. Argout X, Salse J, Aury J-M, Guiltinan MJ, Droc G, Gouzy J, et al. The genome of *Theobroma cacao*. *Nat Genet*. 2011;43(2):101-8.
61. Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, Delcher AL, et al. The genome of woodland strawberry (*Fragaria vesca*). *Nat Genet*. 2011;43(2):109-16.
62. Varshney RK, Chen W, Li Y, Bharti AK, Saxena RK, Schlueter JA, et al. Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. *Nat Biotechnol*. 2012;30(1):83-9.
63. The Brassica rapa Genome Sequencing Project Consortium, Wang X, Wang H, Wang J, Sun R, Wu J, et al. The genome of the mesopolyploid crop species *Brassica rapa*. *Nat Genet*. 2011;43(10):1035-9.
64. Potato Genome Sequencing Consortium, Xu X, Pan S, Cheng S, Zhang B, Mu D, et al. Genome sequence and analysis of the tuber crop potato. *Nature*. 2011;475(7355):189-95.
65. Al-Dous EK, George B, Al-Mahmoud ME, Al-Jaber MY, Wang H, Salameh YM, et al. De novo genome sequencing and comparative genomics of date palm (*Phoenix dactylifera*). *Nat Biotechnol*. 2011;29(6):521-7.
66. D'Hont A, Denoeud F, Aury J-M, Baurens F-C, Carreel F, Garsmeur O, et al. The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature*. 2012;488(7410):213-7.
67. Zhang Q, Chen W, Sun L, Zhao F, Huang B, Yang W, et al. The genome of *Prunus mume*. *Nat Commun*. 2012;3:1318.
68. Brenchley R, Spannagl M, Pfeifer M, Barker GLA, D'Amore R, Allen AM, et al. Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature*. 2012;491(7426):705-10.
69. Amborella Genome Project. The Amborella genome and the evolution of flowering plants. *Science*. 2013;342(6165):1241089.
70. Xu Q, Chen L-L, Ruan X, Chen D, Zhu A, Chen C, et al. The draft genome of sweet orange (*Citrus sinensis*). *Nat Genet*. 2013;45(1):59-66.
71. Singh R, Ong-Abdullah M, Low E-TL, Manaf MAA, Rosli R, Nookiah R, et al. Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. *Nature*. 2013;500(7462):335-9.
72. Wu J, Wang Z, Shi Z, Zhang S, Ming R, Zhu S, et al. The genome of the pear (*Pyrus bretschneideri* Rehd.). *Genome Res*. 2013;23(2):396-408.
73. Ling H-Q, Zhao S, Liu D, Wang J, Sun H, Zhang C, et al. Draft genome of the wheat A-genome progenitor *Triticum urartu*. *Nature*. 2013;496(7443):87-90.
74. Kagale S, Koh C, Nixon J, Bollina V, Clarke WE, Tuteja R, et al. The emerging biofuel crop *Camelina sativa* retains a highly undifferentiated hexaploid genome structure. *Nat Commun*. 2014;5:3706.

75. Dohm JC, Minoche AE, Holtgräwe D, Capella-Gutiérrez S, Zakrzewski F, Tafer H, et al. The genome of the recently domesticated crop plant sugar beet (*Beta vulgaris*). *Nature*. 2014;505(7484):546-9.
76. Bolger A, Scossa F, Bolger ME, Lanz C, Maumus F, Tohge T, et al. The genome of the stress-tolerant wild tomato species *Solanum pennellii*. *Nat Genet*. 2014;46(9):1034-8.
77. Wang M, Yu Y, Haberer G, Marri PR, Fan C, Goicoechea JL, et al. The genome sequence of African rice (*Oryza glaberrima*) and evidence for independent domestication. *Nat Genet*. 2014;46(9):982-8.
78. Denoeud F, Carretero-Paulet L, Dereeper A, Droc G, Guyot R, Pietrella M, et al. The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *Science*. 2014;345(6201):1181-4.
79. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. 2009;10(1):57-63.
80. Wilhelm BT, Landry J-R. RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. *Methods San Diego Calif*. 2009;48(3):249-57.
81. Costa V, Angelini C, De Feis I, Ciccodicola A. Uncovering the complexity of transcriptomes with RNA-Seq. *J Biomed Biotechnol*. 2010;2010:853916.
82. Nagalakshmi U, Waern K, Snyder M. RNA-Seq: a method for comprehensive transcriptome analysis. *Curr Protoc Mol Biol Ed Frederick M Ausubel Al*. 2010;Chapter 4:Unit 4.11.1-13.
83. Garg R, Jain M. RNA-Seq for transcriptome analysis in non-model plants. *Methods Mol Biol Clifton NJ*. 2013;1069:43-58.
84. Tang F, Barbacioru C, Nordman E, Li B, Xu N, Bashkirov VI, et al. RNA-Seq analysis to capture the transcriptome landscape of a single cell. *Nat Protoc*. 2010;5(3):516-35.
85. Ozsolak F, Milos PM. RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet*. 2011;12(2):87-98.
86. Castro J, Maddox D, Cobos M, Requena D, Zimic M, Imán S, et al. De novo assembly and functional annotation of *Myrciaria dubia* fruit transcriptome reveals multiple metabolic pathways for L-ascorbic acid biosynthesis. *BMC Genomics*. 2015;In peer review.
87. Ong WD, Voo L-YC, Kumar VS. De novo assembly, characterization and functional annotation of pineapple fruit transcriptome through massively parallel sequencing. *PloS One*. 2012;7(10):e46937.
88. Martínez-López LA, Ochoa-Alejo N, Martínez O. Dynamics of the chili pepper transcriptome during fruit development. *BMC Genomics*. 2014;15:143.
89. Li C, Wang Y, Huang X, Li J, Wang H, Li J. De novo assembly and characterization of fruit transcriptome in Litchi chinensis Sonn and analysis of differentially regulated genes in fruit in response to shading. *BMC Genomics*. 2013;14:552.
90. Wu H, Jia H, Ma X, Wang S, Yao Q, Xu W, et al. Transcriptome and proteomic analysis of mango (*Mangifera indica* Linn) fruits. *J Proteomics*. 2014;105:19-30.
91. Feng C, Chen M, Xu C, Bai L, Yin X, Li X, et al. Transcriptomic analysis of Chinese bayberry (*Myrica rubra*) fruit development and ripening using RNA-Seq. *BMC Genomics*. 2012;13:19.
92. Hyun TK, Rim Y, Jang H-J, Kim CH, Park J, Kumar R, et al. De novo transcriptome sequencing of *Momordica cochinchinensis* to identify genes involved in the carotenoid biosynthesis. *Plant Mol Biol*. 2012;79(4-5):413-27.
93. Xie M, Huang Y, Zhang Y, Wang X, Yang H, Yu O, et al. Transcriptome profiling of fruit development and maturation in Chinese white pear (*Pyrus bretschneideri* Rehd). *BMC Genomics*. 2013;14(1):823.

94. Li X, Sun H, Pei J, Dong Y, Wang F, Chen H, et al. De novo sequencing and comparative analysis of the blueberry transcriptome to discover putative genes related to antioxidants. *Gene*. 2012;511(1):54-61.
95. Li Y, Xu C, Lin X, Cui B, Wu R, Pang X. De Novo Assembly and Characterization of the Fruit Transcriptome of Chinese Jujube (*Ziziphus jujuba* Mill.) Using 454 Pyrosequencing and the Development of Novel Tri-Nucleotide SSR Markers. *PLoS ONE* [Internet]. 3 de septiembre de 2014 [citado 20 de marzo de 2015];9(9). Recuperado a partir de: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4153635/>
96. Que Y, Su Y, Guo J, Wu Q, Xu L. A Global View of Transcriptome Dynamics during *Sporisorium scitamineum* Challenge in Sugarcane by RNA-seq. *PloS One*. 2014;9(8):e106476.
97. Wang Z, Hu H, Goertzen LR, McElroy JS, Dane F. Analysis of the *Citrullus colocynthis* Transcriptome during Water Deficit Stress. *PloS One*. 2014;9(8):e104657.
98. Lai Z, Lin Y. Analysis of the global transcriptome of longan (*Dimocarpus longan* Lour.) embryogenic callus using Illumina paired-end sequencing. *BMC Genomics*. 2013;14:561.
99. Liu J, Mei D, Li Y, Huang S, Hu Q. Deep RNA-Seq to Unlock the Gene Bank of Floral Development in *Sinapis arvensis*. *PloS One*. 2014;9(9):e105775.
100. Ge X, Chen H, Wang H, Shi A, Liu K. De novo assembly and annotation of *Salvia splendens* transcriptome using the Illumina platform. *PloS One*. 2014;9(3):e87693.
101. Fan Z, Li J, Li X, Wu B, Wang J, Liu Z, et al. Genome-wide transcriptome profiling provides insights into floral bud development of summer-flowering *Camellia azalea*. *Sci Rep*. 2015;5:9729.
102. Padovan A, Patel HR, Chuah A, Huttley GA, Krause ST, Degenhardt J, et al. Transcriptome sequencing of two phenotypic mosaic eucalyptus trees reveals large scale transcriptome re-modelling. *PloS One*. 2015;10(5):e0123226.
103. Zhao P, Zhang L, Zhao L. Dissection of the style's response to pollination using transcriptome profiling in self-compatible (*Solanum pimpinellifolium*) and self-incompatible (*Solanum chilense*) tomato species. *BMC Plant Biol*. 2015;15(1):119.
104. Shemesh-Mayer E, Ben-Michael T, Rotem N, Rabinowitch HD, Doron-Faigenboim A, Kosmala A, et al. Garlic (*Allium sativum* L.) fertility: transcriptome and proteome analyses provide insight into flower and pollen development. *Front Plant Sci*. 2015;6:271.
105. Huang J-Z, Lin C-P, Cheng T-C, Chang BC-H, Cheng S-Y, Chen Y-W, et al. A de novo floral transcriptome reveals clues into phalaenopsis orchid flower development. *PloS One*. 2015;10(5):e0123474.
106. Sharma R, Mishra M, Gupta B, Parsania C, Singla-Pareek SL, Pareek A. De Novo Assembly and Characterization of Stress Transcriptome in a Salinity-Tolerant Variety CS52 of *Brassica juncea*. *PloS One*. 2015;10(5):e0126783.
107. Han L, Li JL, Jin M, Su YH. Transcriptome analysis of Arabidopsis seedlings responses to high concentrations of glucose. *Genet Mol Res GMR*. 2015;14(2):4784-801.
108. Choudhary M, Jayanand null, Padaria JC. Transcriptional profiling in pearl millet (*Pennisetum glaucum* L.R. Br.) for identification of differentially expressed drought responsive genes. *Physiol Mol Biol Plants Int J Funct Plant Biol*. 2015;21(2):187-96.
109. Jin J, Kim MJ, Dhandapani S, Tjhang JG, Yin J-L, Wong L, et al. The floral transcriptome of ylang ylang (*Cananga odorata* var. *fruticosa*) uncovers biosynthetic pathways for volatile organic compounds and a multifunctional and novel sesquiterpene synthase. *J Exp Bot*. 2015;

110. Yamamoto N, Takano T, Tanaka K, Ishige T, Terashima S, Endo C, et al. Comprehensive analysis of transcriptome response to salinity stress in the halophytic turf grass *Sporobolus virginicus*. *Front Plant Sci.* 2015;6:241.
111. Zhao J-L, Wang Y-L, Yao D-Q, Zhu W-Y, Chen L, He H-L, et al. Transcriptome profiling of trichomeless reveals genes associated with multicellular trichome development in *Cucumis sativus*. *Mol Genet Genomics MGG.* 2015;
112. Castro JC, Cobos M, Ramírez R, Imán S. Aislamiento de ADN genómico de *Myrciaria dubia* (HBK) «camu-camu» apropiado para análisis moleculares. *Cienc Amaz.* 2012;2(1):7-15.
113. Li R, et al. De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res.* 2010;20:265-72.
114. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Birol I. ABySS: a parallel assembler for short read sequence data. *Genome Res.* junio de 2009;19(6):1117-23.
115. Boetzer M, Pirovano W. SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinformatics.* 2014;15:211.
116. English AC, Salerno WJ, Reid JG. PBHoney: identifying genomic variants via long-read discordance and interrupted mapping. *BMC Bioinformatics.* 2014;15(1):180.
117. Korf I. Gene finding in novel genomes. *BMC Bioinformatics.* 2004;5(1):59.
118. Stanke M, Morgenstern B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* 2005;33(suppl 2):W465-7.
119. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics.* 2009;10:421.
120. Slater GS, Birney E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics.* 2005;6(1):31.
121. Campbell MS, Law M, Holt C, Stein JC, Moghe GD, Hufnagel DE, et al. MAKER-P: a tool kit for the rapid creation, management, and quality control of plant genome annotations. *Plant Physiol.* 2014;164(2):513-24.
122. RepeatModeler Download Page [Internet]. [citado 22 de mayo de 2015]. Recuperado a partir de: <http://www.repeatmasker.org/RepeatModeler.html>
123. RepeatScout [Internet]. [citado 22 de mayo de 2015]. Recuperado a partir de: <http://bix.ucsd.edu/repeatscout/>
124. Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, et al. InterProScan: protein domains identifier. *Nucleic Acids Res.* 2005;33(suppl 2):W116-20.
125. Haas B, et al. Automated eukaryotic gene structure annotation using EVIDENCEModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* 2008;9:R7.
126. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinforma Oxf Engl.* 2005;21(18):3674-6.
127. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997;25(17):3389-402.
128. Gómez JCC, Reátegui ADCE, Flores JT, Saavedra RR, Ruiz MC, Correa SAI. Isolation of high-quality total RNA from leaves of *Myrciaria dubia* «camu camu». *Prep Biochem Biotechnol.* 2013;43(6):527-38.

129. Sambrook, J, Fritsch, EF, Maniatis, T. *Molecular Cloning: A Laboratory Manual*. Second. Cold Spring Harbor Laboratory Press; 1989. 2344 p.
130. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinforma Oxf Engl*. 2014;30(15):2114-20.
131. Melicher D, Torson AS, Dworkin I, Bowsher JH. A pipeline for the de novo assembly of the *Themira biloba* (Sepsidae: Diptera) transcriptome using a multiple k-mer length approach. *BMC Genomics*. 2014;15:188.
132. Chen T-W, Gan R-CR, Wu TH, Huang P-J, Lee C-Y, Chen Y-YM, et al. FastAnnotator--an efficient transcript annotation web tool. *BMC Genomics*. 2012;13 Suppl 7:S9.
133. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 1 de enero de 2000;28(1):27-30.
134. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res*. 2007;35(Web Server issue):W182-5.

XIX. ANEXOS