

UNIVERSIDAD DE LA REPÚBLICA - FACULTAD DE CIENCIAS
PROGRAMA DE POSGRADO EN BIOTECNOLOGÍA

UTILIZACIÓN DE SECUENCIAS BARCODE PARA CLASIFICAR MUESTRAS DE ESPECIES VEGETALES DE INTERÉS INDUSTRIAL

Tesis para optar por el título de Magister en
Biotecnología

Lic. Bioq. Fabiana Rey Bentos

Montevideo, 2017



UNIVERSIDAD DE LA REPÚBLICA - FACULTAD DE CIENCIAS

PROGRAMA DE POSGRADO EN BIOTECNOLOGÍA

TESIS DE MAestrÍA EN BIOTECNOLOGÍA

**"UTILIZACIÓN DE SECUENCIAS *BARCODE* PARA CLASIFICAR
MUESTRAS DE ESPECIES VEGETALES DE INTERÉS INDUSTRIAL"**

Lic.Bioq. Fabiana Rey Bentos

ORIENTADORES:

Dr. Fabián Capdevielle

Dra. Mary Lopretti

TRIBUNAL:

Dr. Claudio Martínez

Dra. Alexandra Castro

Dr. Fernando Alvarez

Montevideo, 2017

AGRADECIMIENTOS

Al Laboratorio Tecnológico del Uruguay (LATU) por permitirme llevar a cabo esta tesis y financiarla.

A los miembros del Tribunal por acceder a corregirla, por sus opiniones, aportes y enseñanzas.

A mis tutores por la paciencia y comprensión, y por no abandonar hasta el final.

A mi familia por el apoyo y el sostén en los momentos mas difíciles, por el tiempo y los consejos.

A mis amigos que tantas veces me oyeron hablar de "la tesis" y me dieron ánimo para seguir adelante.

A mis compañeros de trabajo por su apoyo y no dejarme flaquear hasta el final de este largo camino.

A todos muchas gracias por estar de una forma u otra!

TABLA DE CONTENIDO

AGRADECIMIENTOS	3
TABLA DE CONTENIDO	4
GLOSARIO Y ABREVIATURAS	6
LISTA DE FIGURAS Y TABLAS	8
RESUMEN	11
1. INTRODUCCIÓN	12
1.1. TRAZABILIDAD Y AUTENTICIDAD DE ALIMENTOS	12
1.2. PLANTAS AROMÁTICAS Y MEDICINALES	14
1.2.1. Definiciones.....	14
1.2.2. Comercio internacional y antecedentes nacionales	15
1.2.3. Características de las especies vegetales evaluadas.....	17
1.3. METODOLOGÍAS PARA EL RECONOCIMIENTO DE ESPECIES VEGETALES DE INTERÉS INDUSTRIAL	20
1.4. DNA BARCODING	23
1.4.1. Generalidades	23
1.4.2. <i>DNA barcoding</i> en plantas	26
1.4.3. <i>Barcode of Life</i>	29
1.4.4. Aplicaciones del <i>DNA barcoding</i>	31
1.5. PROCEDIMIENTOS PARA CLASIFICACIÓN DE MUESTRAS	33
2. OBJETIVOS	35
2.1. OBJETIVO GENERAL	35
2.2. OBJETIVOS ESPECÍFICOS	35
3. METODOLOGÍA	36
3.1. MATERIAL VEGETAL	36
3.2. EXTRACCIÓN DE ADN	37
3.3. OBTENCIÓN DE SECUENCIAS <i>BARCODE</i>	39
3.3.1. Amplificación del locus <i>rbcl</i>	39
3.3.2. Locus <i>matK</i>	40

3.3.3. Secuenciación.....	41
3.4. ANÁLISIS DE DATOS MEDIANTE BARCODE OF LIFE DATA SYSTEMS (BOLD)	42
3.4.1. Ingreso de datos	42
3.4.2. Clasificación.....	43
3.4.3. Clasificación usando base de datos GenBank.....	44
3.5. ANÁLISIS DE DATOS MEDIANTE UN PROCEDIMIENTO ESTADÍSTICO DE CLASIFICACIÓN (IMPLEMENTADO MEDIANTE PLATAFORMA WEKA)	45
4. RESULTADOS Y DISCUSIÓN	46
4.1 EXTRACCIÓN DE ADN	46
4.2. OBTENCIÓN DE SECUENCIAS <i>BARCODE</i>	50
4.2.1. Locus <i>rbcl</i>	50
4.2.2. Locus <i>matK</i>	57
4.3. ANÁLISIS DE DATOS MEDIANTE BARCODE OF LIFE DATA SYSTEMS (BOLD)	60
4.3.1. Clasificación.....	60
4.3.2. Clasificación usando base de datos Genbank.....	69
4.4. ANÁLISIS DE DATOS MEDIANTE UN PROCEDIMIENTO ESTADÍSTICO DE CLASIFICACIÓN (IMPLEMENTADO MEDIANTE PLATAFORMA WEKA)	76
5. CONCLUSIONES Y PERSPECTIVAS	90
6. REFERENCIAS BIBLIOGRÁFICAS	92
7. ANEXOS	101

GLOSARIO Y ABREVIATURAS

ADN: ácido desoxirribonucleico

ARFF: Attribute-Relation File Format

Barcode: código de barras

BID: Banco Interamericano de Desarrollo

BOLD: Barcode of Life Data Systems

BSA: seroalbúmina bovina

CBI: Centre for the Promotion of Imports from Developing Countries

CBOL: Consortium for the Barcode of Life

CTAB: bromuro de cetiltrimetilamonio

DBGW: Database Working Group

DDBJ: DNA Database of Japan

DMSO: dimetilsulfóxido

DNA barcoding: metodología basada en el uso de una secuencia *barcode* para la identificación de la especie a la que pertenece un espécimen

dNTPs: desoxinucleótidos trifosfato

EDTA: ácido etilendiaminotetraacético

EMBL: European Molecular Biology Laboratory

UE: Union Europea

FAO: Organización de las Naciones Unidas para la Alimentación y la Agricultura

FIDA: Fondo Internacional para el Desarrollo Agrícola

FPTA: Fondo de Promoción de Tecnología Agropecuaria

FUNDASOL: Fundación Uruguaya de Cooperación y Desarrollo Solidarios

IAEA: Agencia Internacional de Energía Atómica

INIA: Instituto Nacional de Investigaciones Agropecuarias

LATU: Laboratorio Tecnológico del Uruguay

k-NN: algoritmo del vecino mas cercano (Nearest Neighbour algorithm)

Locus: posición de un gen o secuencia en un cromosoma

NCBI: National Center for Biotechnology Information

MERCOSUR: Mercado Común del Sur

OMS: Organización Mundial de la Salud

PACPYMES: Programa de Apoyo a la Competitividad y Promoción de Exportaciones

PCR: reacción en cadena de la polimerasa

PVP: polivinilpirrolidona

Secuencia *barcode*: secuencia de ADN proveniente de una región definida del genoma, utilizada para identificar especies

TAE: solución tampón Tris - Acetato - EDTA

USDA: United States Department of Agriculture

USP: United States Pharmacopeia

Voucher: muestra representativa de un organismo identificado taxonómicamente por un experto que se encuentra depositada en un herbario.

LISTA DE FIGURAS Y TABLAS

Figura 1. Componentes principales del proyecto <i>Barcode of Life</i> (BOL) y su contribución a la taxonomía, reconstrucción de filogenias moleculares e investigaciones en genética de poblaciones.	25
Figura 2. Poder de discriminación entre especies de los 7 loci por separado, y en combinaciones de 2 y 3.	28
Figura 3. Diagrama de las etapas y componentes del proceso de <i>DNA barcoding</i>	30
Tabla 1. Descripción de los materiales vegetales de referencia utilizados.	36
Tabla 2. Descripción de las muestras industrializadas comerciales analizadas.....	37
Tabla 3. Secuencias de los cebadores utilizados.	40
Figura 4. Electroforesis en gel de agarosa al 1% de ADN extraído de muestras comerciales.	47
Figura 5. Electroforesis en gel de agarosa al 1% de ADN extraído de muestras comerciales.	47
Tabla 4. Concentración y pureza de los ADN obtenidos de la extracción a partir de tejido fresco de muestras de referencia utilizando el protocolo de FAO/IAEA.....	48
Tabla 5. Concentración y pureza de los ADN obtenidos de la extracción a partir de muestras comerciales industrializadas utilizando el protocolo de FAO/IAEA.	49
Tabla 6. Concentración y pureza de los ADN obtenidos de la extracción a partir de muestras comerciales industrializadas utilizando el kit comercial.	50
Figura 6. Mapa de la ubicación en el gen <i>rbcl</i> de <i>Arabidopsis thaliana</i> de los cebadores utilizados.	51
Figura 7 . Amplificación del locus <i>rbcl</i> en muestras de referencia.	51
Figura 8. Amplificación del locus <i>rbcl</i> en muestras de referencia.	52
Figura 9. Reamplificación del locus <i>rbcl</i> en muestras en las que no se obtuvo amplicón en la amplificación inicial.....	52
Tabla 7. Resultados de amplificación y secuenciación de los loci <i>rbcl</i> y <i>matK</i> en muestras de referencia	53

Tabla 7. Resultados de amplificación y secuenciación de los loci <i>rbcl</i> y <i>matK</i> en muestras de referencia	54
Tabla 7. Resultados de amplificación y secuenciación de los loci <i>rbcl</i> y <i>matK</i> en muestras de referencia	55
Figura10. Mapa de la ubicación en el gen <i>matK</i> de <i>Arabidopsis thaliana</i> de los cebadores utilizados	57
Figura 11. Amplificación del locus <i>matK</i>	58
Figura 12. Amplificación del locus <i>matK</i>	58
Figura 13. Amplificación del locus <i>matK</i>	59
Tabla 8. Clasificación de muestras de referencia mediante BOLD usando secuencias del locus <i>rbcl</i>	61
Tabla 8. Clasificación de muestras de referencia mediante BOLD usando secuencias del locus <i>rbcl</i>	62
Tabla 9. Clasificación de muestras de referencia mediante BOLD usando secuencias del locus <i>matK</i>	62
Tabla 9. Clasificación de muestras de referencia mediante BOLD usando secuencias del locus <i>matK</i>	64
Figura 14. Árbol NJ construido a partir de las secuencias <i>rbcl</i> de las muestras de referencia	66
Figura 15. Árbol NJ construido a partir de las secuencias <i>matK</i> de las muestras de referencia	¡Error! Marcador no definido.
Tabla 10. Clasificación de muestras comerciales mediante BOLD usando secuencias del locus <i>rbcl</i>	70
Tabla 11. Clasificación de muestras comerciales mediante BOLD usando secuencias del locus <i>matK</i>	70
Tabla 12. Identificación de muestras de referencia en GenBank usando secuencias del locus <i>rbcl</i>	71
Tabla 13. Identificación de muestras de referencia en GenBank usando secuencias del locus <i>matK</i>	74
Figura 16. Clasificación de secuencias barcode del locus <i>rbcl</i> de muestras de referencia.	78

Figura 17. Clasificación de secuencias barcode del locus <i>matK</i> de muestras de referencia	¡Error! Marcador no definido.
Figura 18. Clasificación de secuencias barcode del locus <i>rbcl</i> de muestras de referencia y comerciales.....	789
Figura 19. Calsificación de secuencias barcode del locus <i>matK</i> en muestras de referencia y comerciales.....	80
Figura 20. Clasificación de muestras de referencia del locus <i>rbcl</i> correspondientes a grupo de especies aromáticas medicinales.....	81
Figura 21. Clasificación de muestras de referenciadel locus <i>matK</i> correspondientes a grupo de especies aromáticas medicinales.....	81
Figura 22. Clasificación de secuencias barcode <i>matK</i> de muestras de referencia y muestra comercial	82
Figura 23. Clasificación de secuencias barcode <i>rbcl</i> de muestras de referencia y muestra comercial	82
Figura 24. Clasificación de secuencias barcode <i>rbcl</i> de secuencias de referencia y comercial pertenecientes al género <i>Origanum</i>	83
Figura 25. Clasificación de secuencias barcode <i>matK</i> de secuencias de referencia y comercial pertenecientes al género <i>Origanum</i>	83
Figura 26. Clasificación de muestra comercial de ajo utilizando secuencias barcode <i>rbcl</i> de muestras de referencia de <i>Allium sativum</i>	84
Figura 27. Clasificación de muestra comercial de ajo utilizando secuencias barcode <i>matK</i> de muestras de referencia de <i>Allium sativum</i>	84
Figura 28. Clasificación de muestra comercial de ajo utilizando secuencias barcode <i>rbcl</i> de referencia	845
Figura 29. Clasificción de muestra comercial de ajo utilizando secuencias barcode <i>matK</i> de referencia	85
Figura 30. Clasificación de muestras de referencia utilizando secuencias barcode <i>rbcl</i> y <i>matK</i> unificadas por muestra	877
Figura 31. Clasificación de muestras comerciales con respecto a muestras de referencia utilizando secuencias barcode <i>rbcl</i> y <i>matK</i> unificadas para cada muestra.....	88

RESUMEN

En los últimos años la trazabilidad ha sido reconocida como una herramienta esencial para garantizar la seguridad y la calidad de los alimentos. Calidad, autenticidad y trazabilidad se han convertido en temas de sumo interés no solamente para los consumidores, sino también para la industria, productores, comerciantes, y agencias reguladoras.

El mercado de hierbas medicinales y especias se encuentra en aumento a nivel mundial así como sus precios, debido a la preferencia de los consumidores por productos naturales. Esto hace que estos productos sean blanco de adulteración, la cual es más común y difícil de detectar en el caso de sus derivados industrializados. Por este motivo es crítico contar con procedimientos confiables que permitan su autenticación.

En el presente trabajo se aplicaron herramientas moleculares basadas en ADN para la generación de información genómica aplicada con el objetivo de clasificar muestras de hierbas aromáticas y plantas medicinales frescas e industrializadas. El trabajo se basó en la estrategia del *DNA barcoding* que utiliza secuencias denominadas *barcodes*, correspondientes a fragmentos específicos de genes presentes en el genoma plasmídico de todas las plantas, los cuales presentan mayor variación interespecífica que intraespecífica. Se utilizó la plataforma *Barcode of Life Data Systems* (BOLD) para el manejo, análisis de datos y clasificación por asignación de un individuo desconocido a una especie mediante sus *barcodes*. Asimismo se evaluó la aplicabilidad de un procedimiento basado en minería de datos para la clasificación de las muestras, implementado en base al *Waikato Environment for Knowledge Analysis* (WEKA).

Los resultados obtenidos indican que ambos marcadores son amplificables con los cebadores utilizados obteniéndose secuencias bidireccionales de alta calidad en el caso de tejido fresco en el 83% de los casos para *rbcl* y en el 63% de los casos para *matK*. En las muestras industrializadas los valores fueron de 50 y 35% respectivamente. Ambas herramientas permitieron la clasificación de las muestras de tejido fresco. En el caso de las muestras industrializadas, el procedimiento algorítmico del vecino más cercano implementado por WEKA alcanzó un alto porcentaje de clasificación correcta dentro de los grupos de referencia.

Podemos concluir que la metodología *DNA barcoding* es una herramienta útil y accesible que permite la clasificación en grupos, si bien su desempeño se ve acotado por la representación de las especies en las bases de datos. El enfoque basado en minería de datos utilizando secuencias *barcode* mostró un buen desempeño para la clasificación de muestras vegetales frescas e industrializadas, por lo que se propone como una metodología factible de ser utilizada a nivel industrial con distintas matrices.

1. INTRODUCCIÓN

1.1. TRAZABILIDAD Y AUTENTICIDAD DE ALIMENTOS

En los últimos años los aspectos relacionados con la trazabilidad han sido reconocidos como una herramienta esencial para garantizar la seguridad y la calidad de los alimentos.

Calidad, autenticidad y trazabilidad se han convertido en temas de sumo interés no solamente para los consumidores, sino también para la industria, productores, comerciantes, y agencias reguladoras.

Actualmente los consumidores exigen alimentos frescos, sabrosos, nutritivos y seguros. Más aún, un número cada vez más grande se inclina por el consumo de alimentos funcionales que ofrezcan beneficios específicos para la salud. Esto ha originado un cambio en la agricultura, que de enfocarse en la cantidad producida pasó a poner el énfasis en la calidad, seguridad, funcionalidad y sustentabilidad; con la consiguiente necesidad de desarrollar y adoptar sistemas de trazabilidad en las cadenas de suministro (Opara, 2003).

En el contexto de los alimentos, el Parlamento Europeo definió la trazabilidad en la Regulación EC N°178/2002 como "la habilidad de trazar y seguir el rastro, a través de todas las etapas de producción, transformación y distribución, de un alimento, una ración, un animal destinado a la producción de alimentos o una sustancia destinados a ser incorporados en alimentos o raciones o con probabilidad de serlo" (European Parliament and of the Council, 2002). En 2006 el Codex Alimentarius propuso la siguiente definición: "la capacidad de seguir el movimiento de un alimento a través de etapas especificadas de la producción, procesamiento y distribución" (Codex Alimentarius Commission, 2006).

La trazabilidad contribuye a agregar transparencia a la cadena de distribución mediante el etiquetado y la existencia de registros verificables. De esta forma agrega valor al sistema de gestión de la calidad permitiendo la identificación, verificación y aislamiento de las fuentes de no conformidades de modo de cumplir con las regulaciones y satisfacer las expectativas de los consumidores.

Desde el punto de vista del consumidor la trazabilidad genera confianza en la calidad y seguridad de los alimentos que consume. Para los agricultores, la industria, los distribuidores y los reguladores, es una parte importante de un sistema que colabora en la mejora continua y en la disminución del impacto en la salud de la población a través de la rápida detección y aislamiento de las fuentes de riesgo (Agencia Española

de Seguridad Alimentaria y Nutrición (AESAN), 2009). Ante una alerta alimentaria, el poder conocer el origen y el movimiento de los productos mediante un sistema de trazabilidad permite definir el campo de acción y potencial del problema, facilitar la recuperación y retiro del mercado de los lotes afectados, establecer claramente las responsabilidades y reducir el impacto económico sobre los integrantes de la cadena de suministro (López, 2003).

En relación a lo mencionado anteriormente, uno de los riesgos que está ganando la atención de todos los actores de la cadena alimentaria es el fraude con fines económicos. El panel de expertos en adulterantes intencionales de los ingredientes alimentarios de la United States Pharmacopeia (USP) definió la adulteración intencional o por intereses económicos de los ingredientes alimentarios como " la adición fraudulenta de sustancias no auténticas, o la remoción, o sustitución de sustancias auténticas sin el conocimiento del comprador para la ganancia del vendedor" (Moore J.C., 2012).

Esta no es solo una cuestión de importancia comercial, sino que involucra la seguridad alimentaria dado que cualquier adulteración resulta en un cambio en la identidad o la pureza del ingrediente original mediante su sustitución, dilución, o modificación por medios físicos o químicos. En general, la práctica más común es el agregado intencional de un adulterante para aumentar el valor del alimento mediante el engaño. A menudo el adulterante es seguro para consumo humano a pesar que no esté expresamente permitido su agregado en los alimentos, pero en algunos casos se han adulterado alimentos con sustancias tóxicas (Manning & Soon, 2014).

Por todo lo expuesto, es de suma importancia el desarrollo de metodología analítica de detección que permita garantizar la autenticidad de los alimentos y de esta forma prevenir el fraude. El objetivo principal es la identificación de marcadores únicos o grupos de marcadores que permitan caracterizar la autenticidad del alimento o sus potenciales adulterantes y/o contaminantes, y utilizarlos para la resolución de casos de autenticación.

En este trabajo nos centramos en la identificación de especie en plantas aromáticas y medicinales. De acuerdo a un estudio publicado por el Centro para la Promoción de Importaciones desde países en desarrollo a Europa (CBI), el fraude en el mercado de hierbas y especias está aumentando (CBI Market Intelligence, 2015). El aumento de precios y de mercados para este tipo de productos es causa de la adulteración con fines de ganancia económica. En el caso de hierbas y especias industrializadas (molidas y/o procesadas) la adulteración es más común y difícil de detectar. Por este motivo Europa y Estados Unidos, principales mercados para la exportación de estos

productos, tienen requerimientos de calidad y autenticidad que brinden garantías para la seguridad alimentaria de sus consumidores.

1.2. PLANTAS AROMÁTICAS Y MEDICINALES

1.2.1. Definiciones

Se define como plantas aromáticas aquellas que contienen una cantidad sobresaliente de aceites esenciales y están representadas por una gran variedad de especies ampliamente distribuidas en el reino vegetal principalmente en algunas familias como las Compuestas, Coníferas, Labiadas, Mirtáceas, Rutáceas y Umbelíferas. Son valoradas por sus aromas y sabores característicos así como por sus propiedades medicinales.

Los aceites esenciales son sustancias líquidas, aromáticas y volátiles, conformados por un grupo heterogéneo de sustancias orgánicas situadas en distintas partes de estos vegetales (hojas, flores, rizoma, corteza, frutos, raíz, semillas).

Las plantas medicinales son aquellas que en uno o más de sus órganos contienen sustancias que pueden ser utilizadas con fines terapéuticos o que son precursores para la semisíntesis químico-farmacéutica. La mayoría de las plantas aromáticas tienen propiedades medicinales, pero de éstas últimas no todas pertenecen al grupo de las aromáticas (Davies, Serie FPTA-INIA, 2004 a).

Tienen diferentes usos entre ellos culinario, medicinal, en perfumería y cosmética, aromaterapia, suplementos alimenticios y extractos para la elaboración de nuevas drogas. Para ello se utilizan distintas partes de la planta: semillas, frutos, raíces, tallos, corteza, rizoma de las flores.

A menudo se habla de especias haciendo referencia a las partes vegetales desprovistas de clorofila como clavo de olor, canela o nuez moscada, y se denomina hierbas a aquellas que contienen clorofila (Southern African Development Community, 2016).

Las plantas de uso medicinal permiten obtener un amplio espectro de productos de acuerdo a la forma de presentación al consumidor o el uso de distintos procedimientos extractivos (tisanas, cápsulas, extractos y tinturas, nuevas drogas, intermediarios para la síntesis de nuevas drogas, etc).

1.2.2. Comercio internacional y antecedentes nacionales

El comercio de estas especies vegetales mantiene un crecimiento sostenido a nivel mundial en las últimas décadas debido a varios factores. Un segmento cada vez más grande de la población se ha tornado más consciente de la necesidad de establecer hábitos de vida más saludables, lo que ha tenido un efecto positivo en la demanda de hierbas y especias. La necesidad de disminuir el consumo de sal y aditivos sintéticos en los alimentos, el gusto por la cocina internacional, y la tendencia al consumo de platos preparados, así como el retorno al uso de productos de origen natural con fines terapéuticos han abierto oportunidades para la producción y exportación de plantas aromáticas y medicinales (Douglas, Heyes, & Smallfield, 2005).

En el mercado mundial el 50% de las hierbas son usadas en la dieta humana, 25% en cosmética, 20% en la industria farmacéutica y 5% en otros rubros (Dellacassa, Moyna, & Nieto, 2005).

En 2004, la producción mundial total de hierbas y especias rondó los 6.5 millones de toneladas y su valor fue de aproximadamente 2.8 miles de millones de dólares. El 96,4% de la producción mundial se da en países en vías de desarrollo de Asia, África y América, siendo responsables del 84% del mercado mundial.

El principal importador de estos productos a nivel mundial es la Unión Europea (UE), siendo Alemania el país líder. Le siguen en importancia Estados Unidos y Japón (Southern African Development Community, 2016).

En 2014, la importación de hierbas y especias por parte de la UE alcanzó las 533 mil toneladas con un valor de 1.9 miles de millones de euros. El volumen de importaciones creció un promedio de 3.8% por año entre 2010 y 2014 y el valor de dichas importaciones aumentó 10% por año (CBI Market Intelligence, 2015).

En el mismo año la importación de especias por parte de Estados Unidos fue de 337 mil toneladas representando 1.3 miles de millones de dólares (United States Department of Agriculture, 2015).

En América Latina la situación de las cadenas productivas de hierbas y especias es bastante dispar. En varios países sobresale la producción de materias primas y extractos para el mercado de exportación regional e internacional, sin embargo en otros la producción depende de recolecciones silvestres o de importaciones (Dellacassa, Moyna, & Nieto, 2005).

En las últimas décadas, en nuestro país se llevaron a cabo acciones tendientes a consolidar y potenciar la producción y comercialización de productos de origen vegetal con fines terapéuticos bajo criterios de sostenibilidad ecológica, social y económica.

En este sentido, en la década de los 90, el Instituto Nacional de Investigaciones Agropecuarias (INIA) junto a una empresa privada comenzaron a estudiar el cultivo de especies vegetales nativas de uso medicinal y a validar técnicas de producción de especies introducidas. A fines de los 90 en un proyecto interinstitucional cofinanciado por la UE se enfocaron en la obtención de aceites esenciales a partir de especies vegetales nativas para ofrecer en el mercado local e internacional. Para este fin también era necesario hacer énfasis en el proceso de domesticación de estos cultivos. Este proyecto continuó a partir del año 2000 financiado por el Fondo de Promoción de Tecnología Agropecuaria (FPTA) de INIA (Davies, Serie FPTA-INIA, 2004 a). En 2005 comenzó a ejecutarse un proyecto entre Fundación Uruguaya de Cooperación y Desarrollo Solidarios (FUNDASOL) - Banco Interamericano de Desarrollo (BID), con el objetivo de favorecer el desarrollo y organización del sector de la producción e industrialización de plantas aromáticas y medicinales, así como el acceso al mercado de dichos productos. Paralelamente se inició un programa regional Fondo Internacional para el Desarrollo Agrícola (FIDA) - Mercado Común del Sur (MERCOSUR), de apoyo a la Red de Fitoterapicos en el MERCOSUR, con el objetivo de reducir la pobreza rural a través de la diversificación de su producción con cultivos de plantas medicinales y su inserción en cadenas de producción de fitoterapicos.

Ese mismo año se formó la Cámara Uruguaya de Fitomedicamentos, Productos Naturales y Afines integrada por 25 empresas. Un año más tarde, con el apoyo del Programa de Apoyo a la Competitividad y Promoción de Exportaciones (PACPYMES) de la UE se creó Phyto Uruguay, un consorcio de diez empresas uruguayas que comenzó a exportar y comercializar productos y medicamentos naturales con el apoyo del Estado (Fitoterapia, 2016).

El control de calidad es fundamental en todo tipo de iniciativa de exportación de estas hierbas o productos derivados de ellas, para cumplir con las legislaciones de los países compradores. Asimismo es un requisito del Reglamento Bromatológico Nacional, con el cual debe cumplir todo producto de consumo interno de origen nacional o importado (Reglamento Bromatológico Nacional, 2012). Los parámetros de calidad para este tipo de plantas incluyen: identidad, pureza, contenido y composición de los principios activos, entre otras. Es importante el control de los mismos para evitar la presencia de adulterantes o sustitutos que en algunos casos podrían tener efectos tóxicos.

Los criterios de identidad y pureza juegan un rol preponderante en el aseguramiento de la calidad y reproducibilidad de los productos derivados de estas materias primas, lo cual contribuye a su seguridad y eficacia (Joshi, Chavan, Warude, & Patwardhan, 2004).

1.2.3. Características de las especies vegetales evaluadas.

La selección de las especies vegetales evaluadas en esta tesis fue realizada tomando en cuenta varios factores. El principal fue la necesidad expresada tanto por productores como por organismos certificadores de identificar determinadas especies y poder diferenciarlas de otras muy similares o utilizadas como adulterantes. Además se consideraron las propiedades medicinales y los niveles de consumo e industrialización. Se incluyeron especies nativas que han sido estudiadas por sus propiedades y que están siendo utilizadas en cosmética y fitoterapia. A continuación se describen brevemente los géneros y especies estudiadas.

Allium es un género muy grande que incluye aproximadamente 750 especies de las cuales al menos 20 son comestibles. Contiene varias plantas de importancia económica, entre ellas ajo (*Allium sativum* L.), cebolla (*Allium cepa* L.), ciboulette (*A. schoenoprasum* L.), cebollin chino (*A. tuberosum* Rottler ex Sprengel syn. *A. odorum*) y puerro (*A. ampeloprasum* L. syn. *A. porrum* G. Don). La taxonomía de un género tan grande es dificultosa ya que, entre otras cosas, no hay un rasgo morfológico único que los agrupe. La división del género en subgéneros, secciones y variedades es artificial y está basada en relativamente pocos caracteres morfológicos (Havey, 1991) (Ipek, Ipek, & Simon, 2014). La especie evaluada en este trabajo fue el ajo (*Allium sativum*). Se encuentra entre las plantas cultivadas más antiguas y ha sido usado como especia, alimento y agente medicinal por miles de años por sus múltiples efectos beneficiosos, entre ellos antimicrobiano, antitrombótico, hipolipídico, antiartrítico, hipoglicémico y antitumoral (Thomson & Ali, 2003).

Asteraceae es una de las familias más grandes dentro de las plantas con flor, con alrededor de 25000 especies distribuidas en aproximadamente 1500 géneros. Tiene una considerable importancia ecológica y económica. Miembros de la familia se distribuyen desde las regiones polares hasta los trópicos, extendiéndose por todos los habitats desde el desierto hasta los pantanos, y desde la selva lluviosa hasta los picos de las montañas (Royal Botanic Gardens, 2016). Esta familia encabeza la lista de las plantas con mayor ocurrencia de flavonoides en su forma aglicona, por lo que se consideran prometedoras para ser utilizadas como protectores de las células contra el daño oxidativo. Dentro de ella seleccionamos cuatro especies para evaluar: *Artemisia dracunculus* (estragón), *Achyrocline flaccida* (marcela amarilla), *Achyrocline satureioides* (marcela blanca) y *Baccharis trimmera* (carqueja).

El estragón (*Artemisia dracunculus*) es utilizado en la cocina como especia y conservante en la preparación de frutos, infusiones, vinagres, mostazas, bebidas refrescantes y alcohólicas. También se utiliza con fines medicinales por sus propiedades antiinflamatorias, antipiréticas, antiespasmódicas, entre otras (Aglarova, 2008).

La marcela blanca (*Achyrocline satureioides*) y la marcela amarilla (*Achyrocline flaccida*) son especies nativas del sudeste de América del Sur. Se distribuyen en el Sur de Brasil, Argentina, Paraguay y Uruguay. Ambas se utilizan indistintamente como medicamentos naturales. En estudios realizados en el marco del proyecto FPTA 137 "Desarrollo del sector de las plantas medicinales y aromáticas en Uruguay" se evaluó la concentración de flavonoides, la capacidad antioxidante y citoprotectora de infusiones de las especies de marcela. Se pudo demostrar que disminuyen en un 56% las lesiones debidas a accidentes cerebrovasculares en rata, así como sus propiedades antioxidantes de gran interés para su uso en cosmética. La marcela es una planta muy utilizada popularmente, sin embargo las distintas especies son usadas de forma indistinta atribuyéndoles a todas ellas las mismas propiedades. Los autores concluyeron que no todas presentan la misma composición y que su actividad biológica es claramente diferente (Morquio, y otros, 2004). Estudios farmacológicos con animales realizados por otros grupos de investigación indicaron la presencia de actividades antiinflamatoria, analgésica, espasmolítica y acción sedativa central (Davies, Serie FPTA-INIA, 2004 a).

El género *Baccharis* comprende más de 400 especies nativas de América tropical y subtropical. La carqueja (*Baccharis trimera*) se encuentra distribuida a lo largo de la selva amazónica en Brasil, Perú y Colombia así como en Argentina, Paraguay y Uruguay. Contiene más de 20% de flavonoides siendo sus principales principios activos. Tiene una larga historia de uso con fines medicinales. Se han reportado propiedades como antiácido, antiúlceras, estimulante de la digestión, hepatotónico y detoxificante (Taylor, 1996).

Lamiaceae es una familia extensa, que comprende cerca de 7000 especies pertenecientes a 264 géneros, cuya taxonomía ha mostrado ser dificultosa. Las relaciones filogenéticas entre tribus, subtribus y géneros son poco claras y ambiguas (Kaufman & Wink, 1994). Las especies que pertenecen a esta familia se caracterizan por ser generalmente hierbas o arbustos de varios tamaños. Son cosmopolitas, teniendo dos centros de biodiversidad principales, la cuenca del Mediterráneo y Asia central. Debido a que producen aceites esenciales y metabolitos secundarios son muy usadas desde la antigüedad como especias y como plantas medicinales (De Mattia, y otros, 2011) (Theodoridis, y otros, 2012). Dentro de esta familia seleccionamos las siguientes especies: *Ocimum basilicum* (albahaca), *Origanum vulgare* (orégano), *Origanum majorana* (mejorana) y *Thymus vulgaris* (tomillo).

El género *Origanum* comprende 43 especies y 18 híbridos. Muchas de estas especies se distribuyen en la región mediterránea donde crecen en áreas montañosas en suelos secos, rocosos y a menudo calcáreos. Los pelos glandulares que cubren las partes aéreas de las plantas pertenecientes a este género secretan aceites esenciales con aromas característicos que han hecho que desde la antigüedad hayan sido utilizadas para aromatizar platos tradicionales así como para usos medicinales. En la actualidad tanto la

mejorana (*O. majorana* L.) como el orégano (generalmente *O. vulgare* u *O. onites* L.) son ampliamente utilizados y comercializados a nivel mundial como especias populares. Además de sus propiedades sensoriales estas dos especies poseen actividades antifúngicas, antibacteriales, antioxidantes, analgésicas y anticancerígenas.

Los límites del género *Origanum* han sido objeto de intensas discusiones taxonómicas. Actualmente se acepta la revisión de Letswaart (citado en Lukas, 2010), que toma al género como un todo que comprende tres grupos, diez secciones y 43 especies. Sin embargo la taxonomía no está exenta de incertidumbres ya que todo el género muestra una variación morfológica considerable. Algunas especies poseen características intermedias entre dos secciones. Otras poseen características que no se encuentran dentro del género, pero si en otros estrechamente relacionados como *Thymus* o *Satureja*. Esto podría ser el resultado de la especiación por hibridación, un fenómeno frecuente no sólo dentro del género sino también con especies relacionadas pertenecientes a los géneros *Thymus* y *Satureja* (Lukas, 2010).

El género *Thymus* comprende unas 150 especies distribuidas a través de las regiones áridas, templadas y frías del viejo mundo al norte del Ecuador y en las costas de Groenlandia. La región oeste del Mediterráneo parece ser el centro de origen de este género. La hibridación es un fenómeno muy común en este género, cuando dos o más especies comparten el mismo territorio. En general no se observa incompatibilidad genética entre especies lo que hace que los estudios taxonómicos dentro de este género sean bastante complejos. Con respecto a su uso, las especies del género *Thymus* son muy apreciadas como plantas aromáticas, por sus propiedades medicinales y por su uso en cosmética y perfumería (Morales, 1996).

El género *Ocimum* comprende hierbas anuales y perennes y arbustos nativos de las regiones tropicales y subtropicales de Asia, África y América del Sur. El estudio de la taxonomía de este género es difícil debido a la interferencia del hombre con el cultivo, selección e hibridación. Se han descrito unas 150 especies pertenecientes a este género basado en caracteres morfológicos como color y morfología de las hojas. Recientemente algunos autores han propuesto que de esas solo 65 son especies diferenciables, el resto corresponde a sinónimos o asignaciones falsas.

La albahaca (*Ocimum basilicum* L.) tiene una larga historia de uso culinario como hierba aromática. Tradicionalmente ha sido usada como hierba medicinal para el tratamiento de dolores de cabeza, tos, diarrea, parásitos y disfunciones renales. Se le considera una fuente de compuestos aromáticos y aceites esenciales que tienen actividad repelente de insectos, nematocida y antibacterial (Labra, y otros, 2004).

El género *Lippia* o *Aloysia*, perteneciente a la familia de las Verbenáceas, incluye aproximadamente 200 especies de hierbas, arbustos y árboles pequeños. Las especies se

distribuyen principalmente en América del Sur y Central, y en las regiones tropicales de África. Muchas de ellas tienen historia de uso como medicinas para aliviar problemas gastrointestinales y respiratorios. Se ha demostrado que algunas especies poseen actividad citostática y antimalaria. El género presenta, en general, un perfil consistente en cuanto a composición química, actividades farmacológicas y usos en medicina casera (Pascual, Slowing, Carretero, Sánchez Mata, & Villar, 2001).

1.3. METODOLOGÍAS PARA EL RECONOCIMIENTO DE ESPECIES VEGETALES DE INTERÉS INDUSTRIAL

Como ya fue comentado en las secciones precedentes, tanto las hierbas y las especias como los productos derivados de ellas mediante su industrialización, deben ser seguros para el consumo humano y cumplir con las reglamentaciones de etiquetado, asegurando identidad y pureza. Principalmente los productos industrializados, deshidratados y en polvo, son más propicios para las adulteraciones ya que el adulterante es más difícil de distinguir visualmente. Para estos fines es crítico contar con procedimientos confiables que permitan la autenticación. La Organización Mundial de la Salud (OMS) hace hincapié en la importancia de las metodologías cualitativas y cuantitativas que permitan estandarizar los materiales vegetales, mediante su caracterización botánica, composición química y actividad biológica (Nikam, Kareparamban, Jadhav, & Kadam, 2012).

Las técnicas analíticas utilizadas para estos fines se basan en métodos físicos, métodos químicos/ bioquímicos, inmunoensayos y más recientemente se han incorporado los métodos moleculares basados en ADN (Dhanya & Sasikumar, 2010).

La primera aproximación la constituyen los métodos macroscópicos, que permiten la identificación mediante la inspección de características morfológicas u organolépticas relevantes de ciertas partes de la planta. Tienen varias ventajas entre ellas que simultáneamente permiten hacer una evaluación de calidad. Pero poseen algunas desventajas como la necesidad de personal altamente entrenado y ejemplares de herbario para utilizar como referencia. Por otra parte estos métodos pueden no llegar a identificar a nivel de especie en algunas plantas, y no tienen la sensibilidad para poder llegar a reconocer diferencias sutiles entre plantas como para diferenciar ecotipos o quimiotipos (Techen, S.L. Crockett, Khan, & Scheffler, 2004).

Los métodos microscópicos permiten la identificación del material vegetal entero, fragmentado, e incluso en polvo. El uso de microscopios electrónicos de barrido ha permitido mejorar la exactitud del método. Una de las desventajas de estos métodos es el costo y el mantenimiento del equipamiento. Al igual que los métodos macroscópicos, los

métodos microscópicos tienen la desventaja de requerir personal altamente calificado y material de referencia (Nikam, Kareparamban, Jadhav, & Kadam, 2012) (Techen, S.L. Crockett, Khan, & Scheffler, 2004).

La evaluación estructural ya sea por métodos macro o microscópicos requiere experiencia e idoneidad para el análisis visual de estructuras y características de las plantas, fundamentalmente en aquellas que han sido deshidratadas y molidas, en mezclas de especies o muestras que se han degradado durante el almacenamiento (Dhanya & Sasikumar, 2010).

Otras metodologías químicas y bioquímicas también son utilizadas para la identificación, así como para la detección de adulterantes. Entre ellas podemos mencionar Cromatografía Líquida de Alta Performance (HPLC), Cromatografía en Capa Fina (TLC), Cromatografía Gaseosa (GC), Cromatografía Gaseosa acoplada a Espectrometría de Masas (GC-MS), Resonancia Magnética Nuclear (MNR), Cromatografía Líquida acoplada a Espectrometría de Masas (LC - MS), Electroforesis Capilar (CE), Electroforesis en Geles de Poliacrilamida (PAGE), y *Enzyme Linked Immunosorbent Assay* (ELISA) (Dhanya & Sasikumar, 2010). Estos métodos permiten identificar, y en muchos casos cuantificar, los ingredientes o principios activos que se conoce tienen actividad medicinal o de interés en las hierbas o especias en estudio. En los casos en lo que no se conoce el principio activo se usan sustancias marcadoras que se definen como constituyentes químicos presentes en la planta que son importantes para la calidad del producto final (Kunle, Egharevba, & Ahmadu, 2012). La desventaja de éstos métodos es que los perfiles fitoquímicos pueden variar dependiendo de la parte de la planta que sea procesada, las condiciones ambientales en las que la planta creció, los métodos de recolección y las condiciones de almacenamiento, entre otros. En la mayoría de los casos con el análisis de los compuestos químicos no es posible diferenciar entre especies inequívocamente. En algunos productos, las sustancias marcadoras pueden solaparse con las presentes en especies relacionadas pero que no se desea su presencia en el producto en cuestión. Asimismo en algunos casos los estándares son demasiado caros o difíciles de conseguir, o no han sido identificadas sustancias marcadoras (Dhanya & Sasikumar, 2010) (Khanuja, Shasany, Darokar, & Kumar, 1999).

Muchos de estos productos agrícolas son sometidos a procesos severos durante su industrialización. Éstos alteran la estructura de la planta por lo cual los métodos macroscópicos y microscópicos no permiten la identificación en el producto final. Por este motivo es que actualmente los métodos basados en ADN y/o proteínas son los más utilizados para la trazabilidad de estos productos.

Los marcadores de ADN presentan ventajas sobre otro tipo de marcadores fenotípicos como los químicos o las proteínas. Brindan información más confiable sobre polimorfismos ya que la composición genética es única para cada individuo y no se ve

afectada por la edad, las condiciones fisiológicas ni los factores ambientales (Joshi, Chavan, Warude, & Patwardhan, 2004). Como los marcadores no son tejido específicos pueden detectarse en cualquier estadio del desarrollo del vegetal (Heubl, 2010). El ADN es una macromolécula extremadamente estable que puede extraerse de tejido vegetal fresco, seco y/o procesado. Es resistente a los procesos industriales como desgarro, cocción a presión, hervor y tratamiento con agentes químicos. Esta característica permite realizar la identificación aún contando solo con trazas de material genético en el producto final industrializado. Por otra parte, hay kits comerciales para la extracción que cuentan con tecnología avanzada lo que permite obtener un rendimiento aceptable de ADN aun partiendo de material vegetal degradado (Galimberti, Labra, Sandionigi, & Bruno, 2014).

Existen distintos tipos de técnicas moleculares basadas en ADN que pueden utilizarse para evaluar sustituciones con la finalidad de autenticar taxones de plantas. Entre ellos se encuentran los métodos basados en la reacción en cadena de la polimerasa (PCR), métodos basados en hibridación y metodologías basadas en secuenciación (Sucher & Carles, 2008).

Los métodos basados en hibridación consisten en hibridar sondas marcadas con filtros que contienen el ADN de interés digerido con enzimas de restricción. Los polimorfismos se detectan por la presencia o ausencia de bandas luego de la hibridación. Este tipo de metodología es costosa en tiempo y mano de obra y requiere cantidades de ADN de buena calidad y que no esté degradado (Heubl, 2010) (Sudha Revathy, Rathinamala, & Murugesan, 2012).

Los marcadores basados en PCR implican la amplificación de un determinado locus mediante una reacción en la que participan cebadores específicos o arbitrarios y una ADN polimerasa, enzima termoestable que se encarga de la polimerización. Las ventajas de las metodologías basadas en PCR es que requieren pequeñas cantidades de ADN, no es indispensable el conocimiento previo de la secuencia completa de la región a amplificar (si parte de la misma para el diseño de los cebadores) y los resultados se obtienen en tiempos cortos. Dependiendo del tipo de cebador usado estas técnicas se clasifican en:

- arbitrarias o semi arbitrarias: son aquellas que no requieren conocimiento previo de la secuencia. Por ejemplo: AP-PCR (Arbitrary Polymerase Chain Reaction), RAPD (Random Amplified Polymorphic DNA), DAF (DNA Amplification Fingerprinting), AFLP (Amplified Fragment Length Polymorphism) e ISSR (Inter Simple Sequence Repeat).

- sitio específicas: desarrolladas a partir de secuencias de ADN conocidas. Por ejemplo: CAPS (Cleaved Amplified Polymorphic Sequences), SSR (Simple Sequence Repeats), SCAR (Sequence Characterized Amplified Regions) , STS (Sequence Tagged Sites) (Heubl, 2010).

El desarrollo de estrategias de detección cuantitativa como la PCR cuantitativa competitiva (QC - PCR) y la PCR en tiempo real permitieron la confirmación y cuantificación de los adulterantes presentes en una muestra (Dhanya & Sasikumar, 2010).

Las metodologías basadas en la secuenciación de ADN permiten caracterizar especies, analizar relaciones filogenéticas, genética de poblaciones y procesos evolutivos. Consiste en determinar las sustituciones en el ADN mediante el alineamiento de la secuencia nucleotídica de una determinada región del genoma de una muestra en estudio con una secuencia homóloga perteneciente a organismos relacionados. Se definen regiones apropiadas del genoma nuclear, mitocondrial o cloroplástico. Para obtener una identificación de especies confiable basado en datos de secuenciación es necesario disponer de especímenes de herbario o una base de datos confiable que asegure que los especímenes de referencia hayan sido identificados por taxónomos expertos (Heubl, 2010).

Una aproximación basada en la amplificación y secuenciación de una región del genoma estandarizada y ubicada en una posición específica, denominada *DNA barcoding*, fue desarrollada por científicos canadienses (Hebert, Ratnasingham, & de Waard, 2003). Esta estrategia permite la asignación de una muestra desconocida a una especie conocida mediante la comparación de su secuencia *barcode* contra una base de datos de secuencias correspondientes a especímenes *voucher* previamente identificados.

1.4. DNA BARCODING

1.4.1. Generalidades

La idea de utilizar una pequeña región del genoma para identificar especies data de varias décadas. Ya en 1993, Amot y colaboradores publicaron un trabajo describiendo la posibilidad de utilizar un gen para discriminar aislamientos de *Plasmodium falciparum* (Pečnikar & Buzan, 2013).

En 2003, científicos de la Universidad de Guelph (Canadá) desarrollaron el uso de secuencias *barcode* de ADN como parte de un sistema capaz de ayudar en el reconocimiento e identificación taxonómica de especies. El principio de esta metodología fue descrito por Herbert y colaboradores (2003) y se basa en el uso de una pequeña región del genoma ubicada en una posición específica, que se encuentra presente en forma universal en todos los linajes y que funciona como identificador único de cada especie (Hebert, Cywinska, Ball, & deWaard, 2003). El objetivo era generar una

herramienta para identificar especies que fuera rápida, de bajo costo, útil para los taxónomos y accesible a aquellos que no tuvieran tal experticia.

Sigue el mismo principio que la taxonomía básica, el asociar un nombre a una referencia específica de una colección en asociación con un entendimiento funcional del concepto de especies (Kress, Wurdack, Zimmer, Weigt, & Janzen, 2005). Esto en una primera instancia generó controversia, ya que fue visto por ciertos grupos como un intento de sustituir la taxonomía tradicional por " taxonomía de ADN " (Ebach & Holdrege, 2005). Varios autores refutaron estas opiniones aclarando el alcance y objetivos de esta nueva metodología. Uno de los propósitos del *DNA barcoding* es convertirse en una herramienta para clasificar nuevos organismos en unidades de acuerdo a sus secuencias *barcode*, pero no para delinear especies. En esos casos se requiere el trabajo de taxónomos expertos para asignar estos grupos a especies. Por otro lado, esta metodología puede ayudar a la taxonomía a asignar especímenes en estudio a especies conocidas cuando los caracteres morfológicos ya no están presentes (por ejemplo en especímenes inmaduros o dañados) o en especies que tienen ciclos de vida polimórficos (Cowan, Chase, Kress, & Savolainen, 2006) (Gregory, 2005) (Hajibabaei, Singer, & Hickey, 2007) (Schindel & Miller, 2005) (Newmaster, Fazekas, & Ragupathy, 2006). En la figura 1 se presenta un esquema de la relación de complementación entre el *DNA barcoding*, la taxonomía tradicional, la genética de poblaciones y la filogenética molecular.

En la práctica, a partir de una pequeña muestra de tejido del organismo en estudio se genera la secuencia de ADN de la región estándar o *barcode*. Esta es comparada con una librería de secuencias de referencia originada a partir de especies conocidas. La coincidencia entre la secuencia del organismo en estudio con una de las de la librería de referencia permite una identificación rápida y reproducible. Idealmente una secuencia *barcode* debe tener suficiente variación entre especies pero poca variación intraespecífica, como para permitir una identificación de especie inequívoca. Por lo tanto la selección de un locus *barcode* no es sencilla, ya que debe encontrarse un equilibrio entre la posibilidad de aplicación universal con una tasa máxima de divergencia entre especies (Kress & Erickson, 2007).

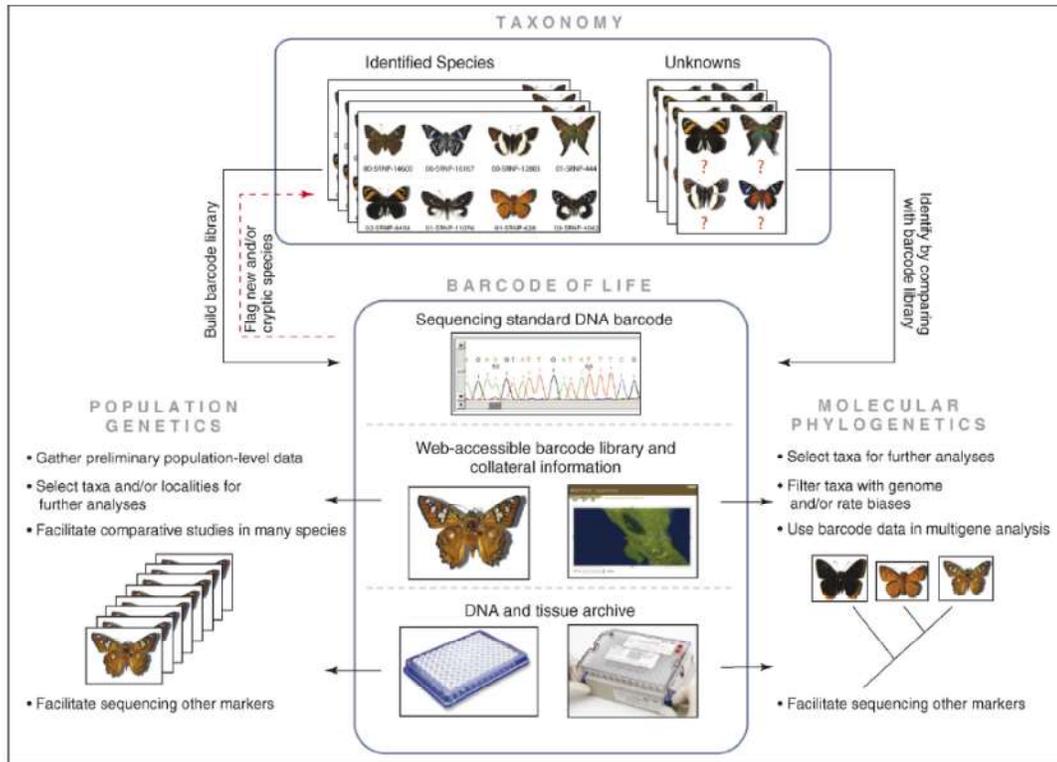


Figura 1. Componentes principales del proyecto *Barcode of Life* (BOL) y su contribución a la taxonomía, reconstrucción de filogenias moleculares e investigaciones en genética de poblaciones. Extraído de Hajibabaei, 2007.

Las características necesarias y/o deseables que requiere una región del genoma para ser candidata a convertirse en *barcode* se describen a continuación:

- debe tener suficiente variación entre especies para permitir la diferenciación pero suficiente similitud entre individuos pertenecientes a la misma especie para permitir agruparlos.
- debe poder estandarizarse de forma que pueda ser amplificada en todas las especies. Para ello el locus debe ser accesible y las regiones que lo flanquean deben estar lo suficientemente conservadas para permitir el diseño de cebadores universales.
- el largo de la secuencia debe ser tal que permita su secuenciación en una única reacción y no debe contener elementos que la dificulten, como por ejemplo repetidos largos. Por otra parte debe poder recuperarse a partir de todo tipo de muestras, incluidas las que estén degradadas por tratarse de productos industrializados, tejidos que datan de muchos años, o que estén presentes en muy pequeña cantidad en la muestra por lo que también es importante que no sean regiones de genes presentes en copia única en el genoma.

- el alineamiento entre secuencias pertenecientes a diferentes especies no debe ser dificultoso, para lo cual se requiere que no presente muchas inserciones ni deleciones (Howard, 2010) (Valentini, Pompanon, & Taberlet, 2008) (Pečnikar & Buzan, 2013).

En animales, el sistema se basa en la amplificación de la región 5' del gen de la subunidad 1 de la citocromo oxidasa c (*cox1*) ubicado en el genoma mitocondrial. Esta región cumple con la mayoría de los criterios descritos anteriormente para un *barcode*. Es haploide, de herencia uniparental y tiene un alto poder de discriminación. Es una región codificante, presente en alto número de copias por célula y en animales no sufre grandes variaciones de tamaño, no presenta estructuras secundarias fuertes, microinversiones o repeticiones frecuentes de mononucleótidos (Hebert, Ratnasingham, & de Waard, 2003) (Hollingsworth, Graham, & Little, 2011).

1.4.2. *DNA barcoding* en plantas

En plantas, la búsqueda de un locus *barcode* ha resultado bastante más dificultosa que en animales. Los genes mitocondriales presentan muy baja tasa de variación de secuencias entre familias, por lo que la región *COI* no es aplicable para la diferenciación de especies. Más aún, las plantas cambian rápidamente su genoma mitocondrial lo que excluye la posibilidad de utilizar los espaciadores intergénicos como identificadores variables que permitan identificar a nivel de especie (Kress, Wurdack, Zimmer, Weigt, & Janzen, 2005). Por otra parte, en plantas es común la ocurrencia de procesos evolutivos complejos como hibridación, poliploidía y apomixis, lo que hace aún más difícil la definición de los límites entre especies (Fazekas, y otros, 2009).

En estudios filogenéticos en plantas es muy utilizada la región ITS (*Internal Transcribed Spacer*), correspondiente al cistrón ribosomal nuclear (18S-5.8S-26S), por lo que en una primera instancia fue postulada como secuencia *barcode* debido a su capacidad de discriminar a nivel de especies y a la cantidad de secuencias existentes en bases de datos. Sin embargo presenta algunas limitaciones. Kress y colaboradores (2005) realizaron un estudio en el que compararon la performance de la región ITS y algunas regiones del genoma del cloroplasto más utilizadas en estudios filogenéticos. Este presenta algunas ventajas con respecto al genoma nuclear como su estabilidad estructural, su herencia uniparental, no sufre recombinación y que se han diseñado cebadores universales para distintas regiones del mismo (Haider, 2011). Los autores concluyeron que para lograr la discriminación a nivel de especie en plantas era necesario el uso de más de un locus simultáneamente, y recomendaron la región nuclear ITS y el espaciador intergénico *trnH-psbA* ubicado en el cloroplasto como punto de partida (Kress, Wurdack, Zimmer, Weigt, & Janzen, 2005).

A partir de entonces la búsqueda de secuencias *barcode* en plantas se enfocó en el genoma del cloroplasto por las ventajas que ya fueron descritas, con el objetivo de identificar regiones de rápida evolución para lograr suficiente variación de secuencias entre especies, pero no demasiado como para asegurar que estén presentes en todas las plantas y que permitan el diseño de cebadores universales robustos. Varios autores realizaron estudios utilizando distintos loci codificantes y no codificantes del genoma plastídico, evaluando su eficiencia en identificar numerosas especies representativas del reino planta (Kress & Erickson, 2007) (Fazekas, y otros, 2008) (Newmaster S. , Fazekas, Steeves, & Janovec, 2008) (Lahaye, y otros, 2008) (Taberlet, y otros, 2007).

De estos estudios surgieron siete *loci* candidatos, de los cuales cuatro son porciones de genes codificantes (*matK*, *rbcl*, *rpoB* y *rpoCl*) y tres son espaciadores no codificantes (*atpF-atpH*, *trnH-psbA* y *psbK-psbI*). Los distintos grupos de investigación probaron distintas combinaciones de estos loci pero no se logró llegar a un consenso.

En 2009 el grupo de trabajo en plantas de CBOL hizo público un estudio en el que compararon la performance de estos siete loci por separado y en combinaciones de dos y tres de ellos, siguiendo los criterios de CBOL en cuanto a calidad de datos y selección de *loci* candidatos (universalidad, calidad de secuencias y cobertura, y poder discriminatorio). Para ello evaluaron 907 secuencias pertenecientes a distintas especies que incluían angiospermas, gimnospermas y criptógamas, obtenidas por distintos grupos de investigación. En base a los resultados obtenidos concluyeron que lo más adecuado era el uso de una combinación de dos *loci* y recomendaron *rbcl* + *matK* como *barcode* estandar para plantas terrestres (CBOL Plant Working Group, 2009). Con esta combinación y este grupo de muestras los autores lograron 72% de identificaciones correctas a nivel de especie. En la figura 2 se observa en forma gráfica el poder de discriminación de los *loci* por separado, y en combinaciones de dos y tres.

El *barcode rbcl* comprende una región de 599 pares de bases ubicada en la región 5' del gen, localizada entre las bases 1 - 599 en la secuencia completa del genoma plastídico de *Arabidopsis thaliana*. Codifica la subunidad mayor de la ribulosa bifosfato carboxilasa (RuBisCO). Esta enzima, una de las más abundantes en la tierra, cataliza la asimilación del CO₂ a carbono orgánico a través del ciclo de Calvin. Se encuentra en la mayoría de las proteobacterias, cianobacterias, algas y plantas superiores (Mondal, Shit, & Kundu, 2013) (Vijayan & Tsou, 2010).

La región *barcode matK* comprende 841 bases en el centro del gen, localizada entre las bases 205 y 1046 en la secuencia completa del genoma plastídico de *A. thaliana*. Codifica la enzima maturasas K, cuya función aún no está establecida. Es el único gen para una proteína de unión al ARN no ribosomal codificada por el cloroplasto en plantas superiores. Está relacionado con maturasas de intron y posiblemente esté involucrado con la eliminación de intrones del grupo II en cloroplastos (Research Group Molecular

Genetics - Humboldt Universitat Zu Berlin, 2016). Su tasa de sustitución a nivel nucleotídico es tres veces más alta y seis veces más alta a nivel de aminoácidos que *rbcl*, lo que muestra que es un gen de rápida evolución (Barthet & Hilu, 2007) (Vijayan & Tsou, 2010).

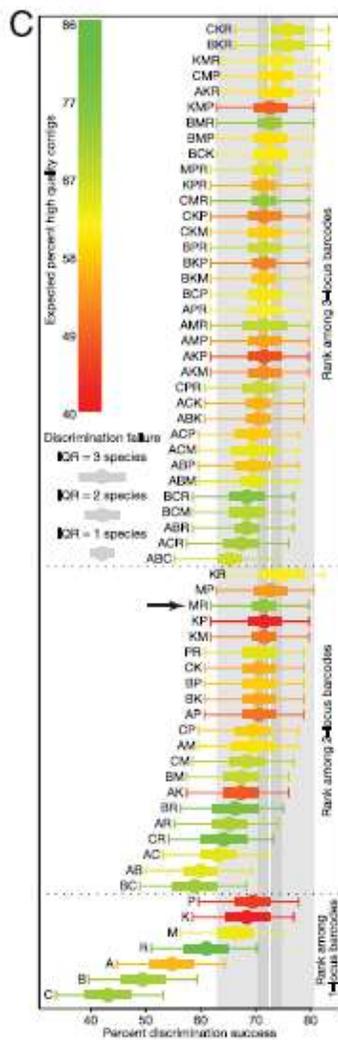


Figura 2. Poder de discriminación entre especies de los 7 loci por separado, y en combinaciones de 2 y 3. Las barras laterales finas indican el intervalo de confianza de 95%, las barras laterales gruesas indican la magnitud relativa de falla en la discriminación. La línea blanca representa el porcentaje de éxito en la discriminación de los siete loci, la barra gris claro el intervalo de confianza de 95%. Las barras gris oscuro indican la magnitud relativa de la falla en la discriminación.

Tomado y modificado de CBOL Plant Working Group, 2009

La elección de esta combinación de *loci* por parte de CBOL se basó en la facilidad de recuperación de la región *rbcl* y el poder de discriminación que ofrece *matK*. El *locus rbcl* es sencillo de amplificar, secuenciar y alinear en la mayoría de las plantas. A pesar de su bajo poder de discriminación oficia de ancla ya que se encuentra presente en todas las especies de plantas. Por el contrario, *matK* resulta más difícil de amplificar particularmente en las plantas que no son angiospermas, pero por su rápida evolución es la región más parecida a *COI* en plantas por su alto poder de discriminación.

1.4.3. *Barcode of Life*

Todo proyecto de *ADN barcoding* tiene cuatro componentes principales:

- Especímenes: es de fundamental importancia la existencia de especímenes identificados por taxónomos, que oficien de referencia y que se encuentren depositados en repositorios de material biológico.
- Análisis en el laboratorio: integrantes del CBOL trabajaron en la elaboración de protocolos de análisis para laboratorios que trabajan con grandes volúmenes de muestras. Los mismos contienen sugerencias que son relevantes para disminuir los costos de análisis y mejorar los tiempos de ejecución (Ivanova, deWaard, Hajibababei, & Hebert, 2017)
- Base de datos: uno de los componentes más importantes del *DNA barcoding* es la construcción de una base de datos de referencia pública que contenga identificadores de especies, que permita la asignación de especímenes desconocidos a especies conocidas. Actualmente existen dos bases de datos que cumplen con estas características, *Barcode of Life Data System* (BOLD) y la *International Nucleotide Sequence Database Collaboration* (INSDC).
- Análisis de datos: la identificación se realiza por comparación de la secuencia problema con las secuencias depositadas en la base de datos. Se busca la coincidencia más cercana y se asigna el espécimen a esa especie.

En la figura 3 se observa un esquema del proceso de *DNA Barcoding* y los cuatro componentes que fueron descritos previamente.

Uno de los objetivos de la iniciativa de *DNA barcoding* es la construcción de una base de datos de secuencias que contenga todas las especies, disponible online y de acceso público. A través del *International Barcode of Life Project* (iBOL), cientos de científicos especialistas en biodiversidad y genómica pertenecientes a 25 países trabajan juntos para construir una librería de referencia parametrizada que constituya el cimiento para el sistema de identificación (Consortium for the Barcode of Life, 2010 - 2017).

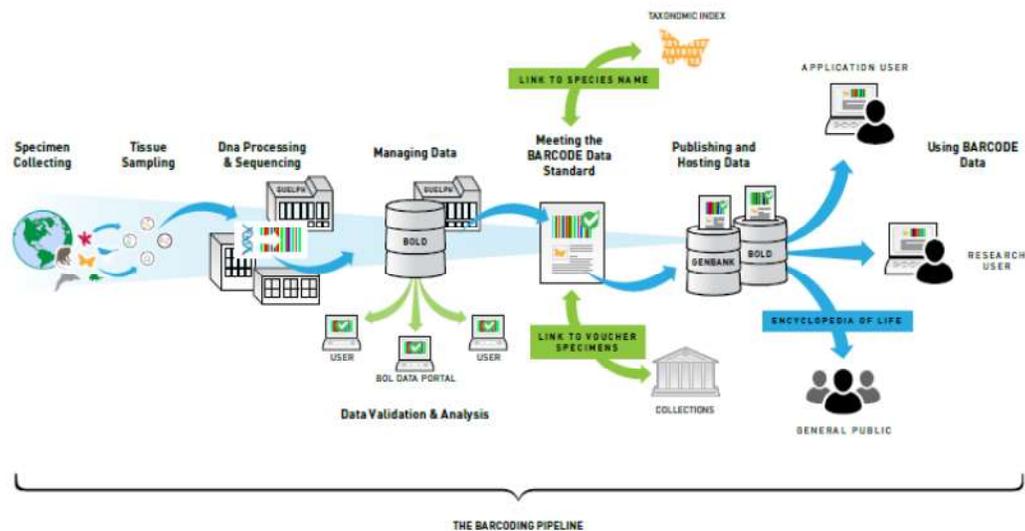


Figura 3. Diagrama de las etapas y componentes del proceso de DNA barcoding (Extraído de Barcode of Life 2017 <http://www.barcodeoflife.org/content/about/what-dna-barcoding>)

En el año 2005 se estableció el *Barcode of Life Data Systems* (BOLD) (<http://www.boldsystems.org>), una plataforma bioinformática web de acceso público desarrollada por el *Consortium for the Barcode of Life* (CBOL), que da soporte a todas las fases de la secuencia analítica desde la colecta del espécimen hasta la librería de secuencias *barcode* validada. Incorpora un repositorio de registros de estas secuencias y una infraestructura que permite el manejo de datos de proyectos de gran escala. Incluye herramientas analíticas que apuntan al aseguramiento de la calidad de los datos que se depositan en ella, así como a la identificación de especímenes. Permite la adquisición, almacenamiento, análisis y publicación de datos, así como el ensamblaje de datos moleculares, morfológicos y de distribución geográfica. Actualmente cuenta con 5443073 secuencias *barcode*, la gran mayoría pertenecientes a animales (178250 especies), siguiéndole en número las plantas (65763 especies), los hongos (20900 especies) y finalmente otros eucariotas unicelulares (Ratnasingham & Hebert, 2007) (International Barcode of Life, 2014).

Cualquier usuario registrado en BOLD puede generar un proyecto y protegerlo con una contraseña, lo que le permite ingresar sus datos y analizarlos en forma personal o con aquellos colaboradores a los que les otorgue acceso. Una vez que los datos se hacen públicos, una copia de las secuencias y los datos de los especímenes migran a GenBank, la base de datos de secuencias de *National Center for Biotechnology Information* (NCBI) y a los repositorios genómicos *DNA Data Bank of Japan* (DDBJ) y *European Molecular Biology Laboratory* (EMBL) ya que los tres forman parte de la *International Nucleotide Sequence*

Database Collaboration (INSDC) (BOLD Systems, Biodiversity Institute of Ontario, 2014-2017).

El grupo de trabajo de bases de datos (DBGW) del CBOL estableció ciertas recomendaciones para el tratamiento de las secuencias que son depositadas como *barcode* en la base de datos BOLD con el objetivo de asegurar la exactitud y confiabilidad de los datos. Previo a la creación del consenso de las secuencias bidireccionales los extremos de ambas deben recortarse para minimizar la cantidad de asignaciones de bases de baja calidad en cada extremo, así como las secuencias de los cebadores. En general las bases con QV >20 se consideran de alta calidad y QV>30 de muy alta calidad. Al editar las bases con un QV<20 deben registrarse como N. Al generar la secuencia consenso, las bases con alta y muy alta calidad de una de las secuencias deben mantenerse por sobre otras de menor calidad de la secuencia en la dirección contraria. Al momento de decidir cual registro será repetible y confiable para la identificación de especies, los usuarios deben seleccionar solo aquellos para los cuales el consenso se haya basado en una cobertura bidireccional con por lo menos 40% de bases cuya asignación sea certera (no Ns) (Hanner, 2009).

1.4.4. Aplicaciones del *DNA barcoding*

Uno de los beneficios de esta metodología es hacer que los resultados de la biología sistemática sean de fácil acceso a muchos tipo de usuarios finales a través de herramientas estandarizadas y de alta tecnología. Esto hace que sus aplicaciones sean muy variadas incluyendo por ejemplo la biomedicina (a través de la identificación de patógenos, parásitos y vectores), bioseguridad, agricultura (identificación de pestes que afectan los cultivos), estudios ambientales y ecológicos, conservación de especies en peligro de extinción, evaluación de la biodiversidad, etc (Ali, y otros, 2014) (Pečnikar & Buzan, 2013) (Stoeckle, Chenery, & Schindel, 2011).

Varios autores han propuesto el *DNA barcoding* como una herramienta universal para la trazabilidad de alimentos. Es un método sensible, rápido, económico y confiable que permite certificar el origen y la calidad de materias primas frescas, así como identificar adulteraciones que ocurren en la cadena de producción (por ejemplo la mezcla de especies) aún en productos con alto nivel de procesamiento. Pero su performance está claramente influenciada por la variabilidad molecular del y/o los organismos en estudio y por la disponibilidad de bases de datos de secuencias de referencia que contengan datos confiables. Es por este motivo que el grupo de trabajo de bases de datos de CBOL estableció criterios de calidad para los datos que son enviados a las bases de datos BOLD y GenBank como se comentó en la sección anterior (Galimberti, y otros, 2013).

Durante los procesos de industrialización las materias primas son sometidas a procesos físicos (calentamiento, ebullición, radiación UV) y/o químicos (agregado de conservantes, estabilizantes, edulcorantes, cambios de pH y reacciones químicas) que alteran la estructura del ADN mediante reacciones físicas y químicas, degradación y fragmentación. La ventaja que tiene el enfoque de *DNA barcoding* frente a otras metodologías que utilizan ADN es que se basa en la amplificación de fragmentos ubicados en el genoma de mitocondria y cloroplastos, los cuales son más resistentes al procesamiento.

Se ha visto que tiene gran efectividad en la trazabilidad en alimentos derivados del mar (peces, moluscos, crustáceos y equinodermos) en su forma fresca o como parte de alimentos procesados. Tanto es así que la *U.S. Food and Drug Administration* (FDA) lo ha propuesto como metodología para la autenticación de productos comerciales basados en pescado (Galimberti, y otros, 2013).

Asimismo tiene potencial para la autenticación y trazabilidad de carnes, leche y productos lácteos, y cultivos a través de la cadena de suministro (Galimberti, Labra, Sandionigi, & Bruno, 2014).

También ha sido utilizada con éxito para separar especies de plantas tóxicas de no tóxicas (Bruni, y otros, 2010), para identificar muestras de té comerciales (Stoeckle, Gamble, Kirpekar, Young, & Ahmed, 2011), especies de frutas en yogures, residuos de frutas en jugos, purés, chocolates y galletitas (Galimberti, y otros, 2013).

Galimberti y colaboradores (2014) propusieron el uso de esta metodología como una herramienta adecuada y universal para caracterizar y trazar a través de la cadena de suministro tanto las especias como las plantas medicinales. En un estudio realizado por su grupo de investigación utilizaron distintas partes de plantas aromáticas y sus productos industrializados (aceites y extractos) conservados en distintas condiciones (deshidratados, congelados, etc). Tomaron muestras en las distintas etapas del proceso de industrialización y también testearon el desempeño de la metodología en extractos de dichas plantas. Concluyeron que la combinación de marcadores propuesta por el CBOL permitía la trazabilidad de este tipo de hierbas aun partiendo de productos industrializados (De Mattia, y otros, 2011).

El mismo tipo de enfoque fue utilizado por diversos autores para la identificación y autenticación de diversas especias, plantas medicinales y productos derivados de éstas (De Mattia, y otros, 2011) (Federici, y otros, 2013) (Gismondi, Fanali, Martínez Labarga, Grilli Caiola, & Canini, 2013) (Theodoridis, y otros, 2012) (Wang, y otros, 2013) (Stoeckle, Gamble, Kirpekar, Young, & Ahmed, 2011) (Chen, y otros, 2010) (Gao, y otros, 2010 a) (Gao, y otros, 2010 b) (Parvathy, y otros, 2014) (Federici, y otros, 2014) (Li, y otros, 2012) (Newmaster, Grguric, Shanmughanandhan, & Ramalingam, 2013) (Wallace, y otros, 2012) (Mahadani & Ghosh, 2013) (Kool, y otros, 2012).

1.5. PROCEDIMIENTOS PARA CLASIFICACIÓN DE MUESTRAS

La clasificación de especies mediante *DNA barcoding* consiste en asignar un espécimen desconocido a una especie conocida mediante el análisis de sus secuencias *barcode*. Para lograr esto se usan tres tipos de métodos *ad hoc*: basados en árboles, basados en similitud y basados en caracteres.

Weitschek y colaboradores proponen que el problema de la clasificación de secuencias *barcode* también puede abordarse como un problema de *machine learning* supervisado. Con este tipo de metodología el usuario provee como entrada una librería de referencia compuesta por especímenes ya clasificados. Basado en este conjunto de referencia el *software* computa el modelo de clasificación y a continuación éste puede aplicarse al conjunto de datos problema que contiene especímenes que requieren clasificación. El conjunto de datos puede contener individuos que no se conoce a que especie pertenecen o estar ya asignados a una especie, y en este caso serviría como una validación del modelo de clasificación. Para obtener datos confiables es preciso que el set de referencia contenga por lo menos cuatro secuencias de cada especie, y que estas contengan todos los polimorfismos posibles (Weitschek, Fiscon, & Felici, 2014).

La plataforma WEKA (*Waikato Environment for Knowledge Analysis*) es un paquete Java de uso libre y local que agrupa los algoritmos más populares para resolver problemas de clasificación, predicción numérica o de agrupamientos. Actualmente es reconocido como un sistema de referencia en minería de datos y *machine learning* que ha tenido amplia aceptación tanto en ámbitos académicos como de negocios (Hall, y otros, 2009).

En un estudio realizado por Weitschek y colaboradores se comparó el desempeño de algoritmos de clasificación supervisada usando la plataforma WEKA con las metodologías de clasificación *ad Hoc* usadas por *DNA barcoding*, usando grupos de datos empíricos y sintéticos. Pudieron concluir que los métodos basados en *machine learning* y minería de datos son muy prometedores para resolver el problema de clasificación en especies, obteniendo un desempeño excelente en ambos tipos de datos (Weitschek, Fiscon, & Felici, 2014).

Otros autores se plantearon hipótesis semejantes con respecto a los métodos utilizados para la clasificación de secuencias *barcode* (van Velzen, Weitschek, Felici, & Bakker, 2012) (Bertolazzi, Felici, & Weitschek, 2009). Austerlitz y colaboradores, realizaron un estudio comparando el desempeño de métodos filogenéticos y métodos estadísticos para la asignación de secuencias *barcode* de muestras problema a especies conocidas. Su conclusión fue que no hay un método universal cuyo desempeño sea más adecuado en

todos los casos, si bien la metodología del vecino más cercano resultó ser la más confiable respecto a cambios en el set de datos. La aplicabilidad de los distintos métodos está fuertemente influenciada por la historia de los especímenes en estudio (eventos de especiación, coalescencia y mutaciones) y de la cantidad de datos disponibles (tamaño de la muestra, largo de las secuencias y cantidad de loci) (Austerlitz, y otros, 2009).

2. OBJETIVOS

2.1. OBJETIVO GENERAL

Demostrar la aplicabilidad de herramientas moleculares basadas en la estrategia del *DNA barcoding* para la clasificación de muestras de especies vegetales - frescas y procesadas - de interés industrial.

2.2. OBJETIVOS ESPECÍFICOS

1. Evaluar el desempeño de la metodología de *DNA barcoding* en la clasificación de muestras de especias y hierbas medicinales, frescas e industrializadas, en grupos de referencia.
2. Evaluar la aplicación de un procedimiento basado en minería de datos, independiente de comparación con bases de datos externas, para clasificar muestras de especias y hierbas medicinales, frescas e industrializadas, utilizando información de secuencias *barcode*.

3. METODOLOGÍA

3.1. MATERIAL VEGETAL

La selección de las especies incluidas en el alcance de esta tesis se realizó en base a necesidades planteadas por productores y organismos certificadores, por sus propiedades medicinales y/o por su uso industrial. La descripción taxonómica de las mismas se observa en la Tabla 1.

Se utilizaron tres ejemplares de cada especie excepto *Origanum majorana* y *Petroselinum crispum* de las cuales se obtuvo una sola muestra. Los mismos fueron proporcionados por INIA de su colección de germoplasma, vinculada al proyecto FPTA N° 137" Desarrollo del sector de las plantas medicinales y aromáticas en Uruguay" (INIA, 2004). Todas ellas fueron usadas como muestras de referencia.

Muestras industrializadas de algunas de las especies en estudio fueron proporcionadas por el Departamento de Análisis de Productos Agropecuarios del LATU y por una empresa del ramo (Tabla 2).

Tabla 1. Descripción taxonómica de los materiales vegetales de referencia utilizados.

Nombre vulgar	Especie	Género	Familia	Orden	Clase
Ajo	<i>Allium sativum</i>	Allium	Amaryllidaceae	Asparagales	Liliopsida
Albahaca	<i>Ocimum</i> sp	Ocimum	Lamiaceae	Lamiales	Magnoliopsida
Estragón	<i>Artemisia dracunculus</i>	Artemisia	Asteraceae	Asterales	Magnoliopsida
Orégano	<i>Origanum vulgare</i>	Origanum	Lamiaceae	Lamiales	Magnoliopsida
Mejorana	<i>Origanum majorana</i>	Origanum	Lamiaceae	Lamiales	Magnoliopsida
Perejil	<i>Petroselinum crispum</i>	<i>Petroselinum</i>	Apiaceae	Apiales	Magnoliopsida
Marcela blanca	<i>Achyrocline satureioides</i>	<i>Achyrocline</i>	Asteraceae	Asterales	Magnoliopsida
Marcela amarilla	<i>Achyrocline flaccida</i>	<i>Achyrocline</i>	Asteraceae	Asterales	Magnoliopsida
Carqueja	<i>Baccharis trimera</i>	<i>Baccharis</i>	Asteraceae	Asterales	Magnoliopsida
Cedrón	<i>Aloysia triphylla</i> (<i>citriodora</i>)	<i>Aloysia</i>	Verbenaceae	Lamiales	Magnoliopsida
Tomillo	<i>Thymus vulgaris</i>	<i>Thymus</i>	Lamiaceae	Lamiales	Magnoliopsida

Tabla 2. Descripción de las muestras industrializadas comerciales analizadas

Muestra	Presentación
Albahaca	Tejido vegetal deshidratado y molido
Orégano	Tejido vegetal deshidratado y molido
Estragón	Tejido vegetal deshidratado
Tomillo	Tejido vegetal deshidratado
Perejil	Tejido vegetal deshidratado y molido
Ajo	Bulbos deshidratados en escamas y molidos
Broccoli (<i>Brassica oleracea</i>)	Tejido vegetal deshidratado y molido en cápsulas de 500 mg (equivalentes a 50 mg de Broccoli fresco)
Bilberry (<i>Vaccinium myrtillus</i>)	Extracto de hoja de arándano negro en cápsulas de 500 mg

3.2. EXTRACCIÓN DE ADN

Las plantas aromáticas y medicinales producen aceites esenciales y una gran diversidad de metabolitos secundarios, entre ellos: alcaloides, flavonoides, fenoles, polisacáridos gomosos, terpenos y quinonas. Se ha visto que este tipo de compuestos pueden dificultar la extracción de ADN e inhibir la reacción de PCR. Los problemas más frecuentes que han sido reportados para la extracción y purificación de ADN de alto peso molecular en este tipo de plantas son su degradación debido a la presencia de endonucleasas, y la extracción simultánea de polisacáridos de alta viscosidad, polifenoles y metabolitos secundarios que inhiben la reacción enzimática. La presencia de polifenoles en muchas de estas especies disminuye el rendimiento y la pureza del ADN extraído (Khanuja, Shasany, Darokar, & Kumar, 1999) (Schori & Showalter, 2011).

Para la obtención de ADN de tejido fresco de muestras de referencia así como de muestras comerciales en las que el tejido vegetal se encontraba deshidratado y molido, se utilizó el método propuesto por FAO/IAEA para la extracción de ADN de alto peso molecular a partir de tejidos vegetales con algunas modificaciones (Joint FAO/IAEA Division of Nuclear Techniques in Food and Agriculture, 2002). El mismo se basa en la ruptura de las paredes celulares mediante la molienda en un mortero del tejido congelado con nitrógeno líquido. El ADN es liberado del núcleo celular por acción del detergente bromuro de cetiltrimetilamonio (CTAB) contenido en el buffer de lisis y protegido de la acción de las endonucleasas por el ácido etilendiamino tetraacético

(EDTA). Una extracción con la mezcla cloroformo-álcohol isoamílico separa el ADN de las proteínas y finalmente es recuperado por medio de una precipitación con isopropanol.

El tejido vegetal fue congelado con nitrógeno líquido y molido en mortero. Se tomó entre 0,5 y 1 g, se colocó en un tubo y se agregaron 600 μ l de buffer de extracción (CTAB 2%, NaCl 1.4 M, EDTA 20 mM, Tris-Cl pH 8 100 mM, PVP 40 2%) precalentado a 60°C y 1,2 μ l de β -mercaptoetanol. Las muestras se incubaron a 60°C por 30 min, mezclando de vez en cuando. Posteriormente se le agregaron 600 μ l de mezcla cloroformo-álcohol isoamílico (24:1) mezclando cuidadosamente, seguido por una centrifugación a 10000 rpm por 20 min a temperatura ambiente. El sobrenadante fue transferido a otro tubo donde se precipitó el ADN con 500 μ l de isopropanol frío, mezclando suavemente hasta lograr la precipitación de los ácidos nucleicos. Las muestras fueron centrifugadas a 10000 rpm por 20 min. El sobrenadante fue descartado y se dejó escurrir el tubo boca abajo sobre un papel absorbente para eliminar los restos de isopropanol. El pellet fue lavado con 500 μ l de etanol 70% seguido de una centrifugación a 10000 rpm por 5 min. Este paso se repitió dos veces. Se dejó secar a temperatura ambiente hasta que todo el etanol se hubiera evaporado. Una vez secos los pellets fueron resuspendidos en 100 μ l de agua milliQ estéril y conservados en freezer a -20°C.

En el caso de algunas muestras industrializadas de origen comercial en las que se obtuvo muy bajo rendimiento de ADN utilizando el protocolo descrito anteriormente se utilizó el kit comercial Nucleospin Food (Macherey - Nagel GmbH & Co. KG, Düren, Germany), según las instrucciones del fabricante. Este kit fue diseñado para la extracción de ADN de alimentos procesados y matrices complejas. Se basa en el uso de un *buffer* de lisis que contiene sales caotrópicas, agentes desnaturalizantes y detergentes, y la adsorción del ADN a una membrana de sílica, lo que permite la remoción de inhibidores de la PCR mediante lavados con *buffer*.

Se tomaron precauciones para evitar contaminaciones entre muestras que pudieran afectar los resultados. Todos los materiales y reactivos utilizados fueron autoclavados a 121°C por 15 minutos previo a su uso. El mortero y las espátulas fueron lavados entre muestra y muestra con hipoclorito de sodio al 10% durante 5 minutos, luego con etanol al 70% durante 5 minutos y enjuagados con agua destilada estéril. El mismo tratamiento se llevó a cabo con las mesadas de trabajo.

La evaluación de la calidad del ADN obtenido fue realizada mediante electroforesis en gel de agarosa al 1% (w/v) utilizando bromuro de etidio (6,25 μ g en 10 μ l de solución) para visualizar las bandas. Se cargaron en promedio 4 μ g de ADN por carril. Se utilizó Mass Ruler DNA Ladder Mix (103 ng/ μ l) (Fermentas, Waltham, MA, USA) como marcador de peso molecular. Las condiciones de corrida fueron 100 V, 200 mA durante 50 min. Los geles fueron visualizados en un documentador de geles Omnidoc (Clever Scientific Ltd., Warwickshire, UK).

La evaluación del rendimiento de la extracción se realizó mediante la medida de absorbancia a 260 nm y a 280 nm utilizando un espectrofotómetro Nanodrop 1000 (Thermo Fisher Scientific, Wilmington, DE, USA). El cociente A_{260}/A_{280} se utilizó como indicador de pureza del ADN extraído.

3.3. OBTENCIÓN DE SECUENCIAS *BARCODE*

Uno de los cuellos de botella para la aplicación de la metodología de *DNA barcoding* en plantas terrestres ha sido la definición de la o las regiones más adecuadas para lograr una asignación de especie confiable. Los genes mitocondriales presentan muy baja tasa de sustitución en plantas, a diferencia de lo que ocurre en los animales, por lo cual la investigación se ha centrado en la búsqueda de regiones alternativas.

En 2009, el grupo de trabajo de plantas del CBOL, recomendó la combinación de dos locus *rbcl* y *matK* como el *barcode* estandar para la identificación de plantas terrestres (CBOL Plant Working Group, 2009).

3.3.1. Amplificación del locus *rbcl*

Para amplificar esta región se utilizaron los cebadores *rbclA_F* y *rbclA_R* propuestos por Erickson en el grupo de plantas de CBOL (2009) (Tabla 3) y las condiciones de reacción por él descritas fueron tomadas como punto de partida para la puesta a punto (CBOL Plant Working Group, 2009).

Se utilizaron Taq polimerasas de distintas procedencias; se modificó la temperatura de unión de los cebadores y el ciclo; se modificaron las concentraciones de reactivos y ADN molde; y se agregó seroalbúmina bovina (BSA) y dimetilsulfóxido (DMSO) como potenciadores de la reacción (Frackman, Kobs, Simpson, & Storts, 1998) (Roux, 1995).

Para evitar contaminaciones cruzadas los materiales fueron autoclavados a 121°C durante 15 minutos. Las mesadas de trabajo fueron tratadas con hipoclorito de sodio al 10% durante 5 minutos, luego etanol 10% durante 5 minutos. Las mezclas de reacción se prepararon en una cabina de PCR y para las pipetas se usaron puntas con filtro.

La reacción se llevó a cabo en un termociclador Palm-Cycler CG1-96 (Corbett Research, Mortlake, Australia), utilizando 0.5 µl de ADN molde (200 ng en promedio) en 20 µl de mezcla de reacción. La misma contenía *buffer* 1X (10 mM Tris-HCl pH 8,3, 50 mM KCl), 1.0 mM MgCl₂, 0.2 mM de cada dNTP, 0.5 µM de cada cebador, 2% DMSO, 0.1 µg/ µl BSA y

1.0 U de Standard Taq Polimerasa (New England Biolabs Inc., Ipswich, MA, USA). Se realizaron controles negativos incluyendo un tubo con reactivos y sin ADN molde en cada reacción.

Las condiciones de reacción fueron las siguientes: un ciclo de desnaturalización a 95°C por 3 min; 33 ciclos de desnaturalización a 94°C por 30 seg, unión de los cebadores a 55°C por 30 seg y extensión a 72°C por 1 min; y una extensión final a 72°C por 10 min.

Los productos de PCR fueron sometidos a electroforesis en gel de agarosa al 1.5% en buffer TAE y visualizados por tinción con bromuro de etidio (6,25 µg en 10 µl de solución). Las condiciones de corrida fueron 80V, 200 mA durante 30 min. Los geles fueron visualizados en un documentador de geles Omnidoc (Clever Scientific Ltd., Warwickshire, UK).

Tabla 3. Secuencias de los cebadores utilizados.

Locus	Primer	Secuencia (5' - 3')	Tamaño esperado del amplicón (pb)
<i>rbcl</i>	rbclLa _F	ATGTCACCACAAACAGAGACTAAAGC	580
<i>rbcl</i>	rbclLa _R	GTAAAATCAAGTCCACCRCG	
<i>matK</i>	3F_KIM f	CGTACAGTACTTTTGTGTTTACGAG	850
<i>matK</i>	1R_KIM r	ACCCAGTCCATCTGGAAATCTTGGTTC	

3.3.2. Locus *matK*

Para amplificar esta región se utilizaron los cebadores 3F_KIM f y 1R_KIM r diseñados por Ki-Joong Kim y sugeridos por el grupo de trabajo de plantas del CBOL (2009) (Tabla 3) y las condiciones de reacción por el descritas fueron tomadas como punto de partida para la puesta a punto (CBOL Plant Working Group, 2009). Se modificaron las concentraciones de reactivos y de ADN molde en la mezcla de reacción hasta lograr poner a punto la amplificación, y se utilizó BSA y DMSO como potenciadores de la reacción (Frackman, Kobs, Simpson, & Storts, 1998) (Roux, 1995).

Para evitar contaminaciones cruzadas los materiales utilizados fueron autoclavados a 121°C durante 15 minutos. Las mesadas de trabajo fueron tratadas con hipoclorito de sodio al 10% durante 5 minutos, luego etanol 10% durante 5 minutos. Las mezclas de reacción se prepararon en una cabina de PCR y para las pipetas se usaron puntas con filtro.

La reacción se llevó a cabo en un termociclador Palm-Cycler CG1-96 (Corbett Research, Mortlake, Australia), utilizando 0.5 µl de ADN molde (200 ng en promedio) en 20 µl de mezcla de reacción. La misma contenía buffer 1X (10 mM Tris-HCl pH 8,3, 50 mM KCl), 1.0 mM MgCl₂, 0.2 mM de cada dNTP, 1.0 µM de cada primer, 4% DMSO, 0.4 µg/ µl BSA y 2.0 U de Standard Taq Polimerasa (New England Biolabs Inc., Ipswich, MA, USA). Se realizaron controles negativos incluyendo un tubo con reactivos y sin ADN molde en cada reacción.

Las condiciones de reacción fueron las siguientes: un ciclo de desnaturalización a 94°C por 1 min; 35 ciclos de desnaturalización a 94°C por 30 seg, unión de los cebadores a 52°C por 20 seg y extensión a 72°C por 50 seg; y una extensión final a 72°C por 5 min.

Los productos de PCR fueron sometidos a electroforesis en gel de agarosa al 1.5% en *buffer* TAE y visualizados por tinción con bromuro de etidio (6,25 µg en 10 µl de solución). Las condiciones de corrida fueron 80V, 200 mA durante 30 min. Los geles fueron visualizados en un documentador de geles Omnidoc (Clever Scientific Ltd., Warwickshire, UK).

3.3.3. Secuenciación

Los productos de amplificación obtenidos fueron enviados a MacroGen Inc. (Seoul, Korea) para su purificación y secuenciación utilizando los mismos cebadores que para la amplificación (Tabla 3). La misma se llevó a cabo en un secuenciador automático Applied Biosystems 3730x/ DNA Analyzer (Life Technologies, Carlsbad, USA). Este equipo utiliza el sistema de secuenciación en ciclos con fluorescencia basado en el método de Sanger. Al igual que la secuenciación por Sanger, este método requiere un ADN molde, un cebador de secuenciación, una ADN polimerasa térmicamente estable, nucleótidos (dNTPs), dideoxinucleótidos (ddNTPs) y *buffer*. Pero a diferencia del método de Sanger que usa material radiactivo, la secuenciación en ciclos emplea colorantes fluorescentes para marcar los productos de extensión y los componentes se combinan en una mezcla de reacción que es sometida a ciclos de hibridación, extensión y desnaturalización en un termociclador. Los fragmentos marcados con fluorescencia son separados por tamaño de acuerdo a su carga total. Poco antes de alcanzar el electrodo positivo pasan a través de la trayectoria de un haz de laser que hace que los colorantes emitan fluorescencia, la que es detectada por un detector óptico. El *software* del equipo convierte la señal de fluorescencia en datos digitales. Como cada colorante, cuando es excitado por el laser, emite luz a distintas longitudes de onda los cuatro colores correspondientes a las cuatro bases pueden detectarse y ser distinguidos en una sola inyección del capilar (Applied Biosystems, 2009). Los resultados de la muestra analizada se presentan en forma de electroferograma en el que cada base asignada tiene asociado un valor de calidad (QV). Se calcula para cada base según: $QV = -10 \log (Pe)$, siendo Pe la probabilidad de error. Por

lo que QV 20 corresponde a un 99% de exactitud y QV 16 a un 97.5 (Ewing & Green, 1998).

En aquellos casos en los que no se obtuvieron secuencias o la calidad de las mismas fue baja en toda su extensión, se procedió a amplificar nuevamente y secuenciar.

En este trabajo se utilizó el *software* Geneious 7.1 (<http://geneious.com>) (Kearse, y otros, 2012) para el análisis y edición de las secuencias. Se verificó la asignación de bases y la calidad de las mismas. En base a esto se recortaron los extremos, eliminando las posiciones en que no fue posible asignar una base (N) y aquellas cuya calidad fue muy baja. Las secuencias directa y reversa editadas correspondientes a cada ejemplar fueron ensambladas para dar lugar a una secuencia consenso.

No se siguieron las recomendaciones de DBGW, descritas en el punto 1.4.3, en forma estricta dado que en muchos casos la calidad de las secuencias obtenidas no lo permitió. Por otra parte el objetivo de este trabajo era comprobar la robustez del sistema para su uso en productos industrializados cuyas secuencias en general son de inferior calidad a las sugeridas por el DBGW.

3.4. ANÁLISIS DE DATOS MEDIANTE BARCODE OF LIFE DATA SYSTEMS (BOLD)

3.4.1. Ingreso de datos

BOLD es una base de datos de libre acceso, que permite el trabajo en línea. Se accede a ella a través de la siguiente dirección: <http://www.boldsystems.org>.

Para ingresar datos en la base es necesario el registro del investigador y la institución en la que se llevó a cabo la investigación. Este hecho permite el acceso a otros proyectos que aun no se han hecho públicos solicitando el permiso al responsable del mismo, así como colaborar en el curado y anotación de la información publicada.

Todos los datos que se ingresan en la base de datos se organizan en proyectos, por lo que el primer paso fue la creación de los proyectos a los que se vincularían los datos. Luego se registraron los especímenes. Para cada registro BOLD asigna un código de identificación que permite rastrear la muestra a lo largo de todo el proceso. A cada espécimen se le asociaron datos de taxonomía, colecta, secuencias obtenidas para los loci locus *rbcl* y *matK*, y los cebadores utilizados para obtenerlas.

Para analizar las secuencias y poder extraer el máximo de información de ellas se trabajó con tres proyectos. Uno de ellos, denominado FRB, contenía las secuencias consenso obtenidas del alineamiento y edición de las secuencias directa y reversa. El segundo, denominado FWD, contenía solo las secuencias directas editadas manualmente; y el tercero, identificado como REV, las secuencias reversas editadas manualmente.

3.4.2. Clasificación

Con la secuencia consenso obtenida para cada muestra se realizó la clasificación de muestras empleando la plataforma BOLD que utiliza toda la librería de secuencias de proyectos públicos y privados depositadas en la base de datos.

En el caso de plantas utiliza los marcadores *rbcl* y *matK* y el algoritmo BLAST que proporciona una lista de las noventa y nueve coincidencias más cercanas pero no la probabilidad de clasificación en un taxón. A cada coincidencia BLAST le asigna un porcentaje de identidad, un *score* y un valor E. El *score* indica el grado de similitud entre las secuencias que están siendo comparadas y tiene relación con el largo de las mismas. El valor E indica la probabilidad de que una coincidencia se haya generado por azar y está relacionado con el tamaño de la base de datos. La clasificación mediante BOLD no puede realizarse con ambos marcadores al mismo tiempo, es decir que se obtienen tantas asignaciones como secuencias *barcode* se utilicen.

Se definió como coincidencia más cercana el blanco con mayor porcentaje de identidad, tomando como límite de aceptación arbitrario un 99,5%. En la mayoría de los casos este correspondió a la secuencia con mayor *score*. En las búsquedas en las que se obtuvieron múltiples coincidencias con igual porcentaje de identidad, se seleccionó el blanco más cercano a la especie de referencia cuando fue posible.

Las muestras de referencia fueron utilizadas para validar la identificación taxonómica. La secuencia se consideró correctamente asignada cuando la especie a la que pertenece obtuvo el mayor *score* entre todos los candidatos. Se consideró una falla en la asignación cuando la especie correcta obtuvo un *score* menor o igual que otras especies.

Para el análisis de los datos de secuencias y especímenes, y como forma de validar la identificación se utilizó la herramienta Taxon ID *Tree* disponible en BOLD. Permite la construcción de dendogramas usando el algoritmo *Neighbor Joining* (Saitou & Nei, 1987). Las secuencias fueron alineadas con MUSCLE y para calcular las distancias se utilizó el modelo *Kimura 2 Parameter* (Milton, Pierossi, & Ratnasingham, 2013) (Edgar, 2004) .

3.4.3. Clasificación usando base de datos GenBank

Adicionalmente a la identificación utilizando la base de datos BOLD se utilizó la base de datos GenBank.

GenBank es la base de datos de secuencias genéticas del *National Institute of Health* (NIH), una colección anotada de todas las secuencias de ADN públicas disponibles. Es parte de la *International Nucleotide Sequence Database Collaboration*; la cual comprende la Base de Datos de ADN de Japón (DDBJ), el Laboratorio Europeo de Biología Molecular (EMBL) y GenBank del Centro Nacional de Información en Biotecnología (NCBI). Estas tres organizaciones intercambian datos diariamente con el objetivo de asegurar el acceso de la comunidad científica a la información completa y actualizada.

GenBank contiene secuencias nucleotídicas de diferentes tipos de organismos tanto eucariotas como procariontas. Actualmente el número de secuencias depositadas supera los 180 millones. Parte de ellas corresponden a *barcodes*. BOLD envía una copia de las secuencias así como los datos de los especímenes a los que corresponden tan pronto como los datos toman carácter público.

La búsqueda en GenBank fue llevada a cabo usando el algoritmo Megablast. Se seleccionaron los parámetros del algoritmo establecidos por defecto excepto el número de secuencias blanco que fue fijado en 500. Al igual que se hizo con la identificación en BOLD, se utilizaron las secuencias consenso generadas por el alineamiento de las secuencias bidireccionales de cada ejemplar.

Se definió la coincidencia más cercana como el blanco con mayor porcentaje de identidad y mayor porcentaje de cobertura de la secuencia problema, usando como límite arbitrario un 90%. En la mayoría de los casos esto correspondió a la secuencia con mayor *score*. En los casos en que la búsqueda arrojó múltiples coincidencias idénticas se priorizó el blanco perteneciente a la especie más cercana a la muestra.

Las muestras de referencia fueron utilizadas para validar la identificación taxonómica. La secuencia se consideró correctamente asignada cuando la especie a la que pertenece obtuvo el mayor *score* entre todos los candidatos. Se consideró una falla en la asignación cuando la especie correcta obtuvo un *score* menor o igual que otras especies.

3.5. ANÁLISIS DE DATOS MEDIANTE UN PROCEDIMIENTO ESTADÍSTICO DE CLASIFICACIÓN (IMPLEMENTADO MEDIANTE PLATAFORMA WEKA)

El objetivo fue utilizar las secuencias *barcode* generadas para asignar un espécimen desconocido a una especie conocida, considerandolo un problema de *machine learning* supervisado, de modo de desarrollar una metodología para la clasificación independiente de bases de datos.

El primer paso fue obtener las secuencias barcode para las muestras en estudio. En este caso disponíamos de las secuencias que generamos de acuerdo a la descripción realizada en los apartados anteriores. Las mismas se encontraban en formato FASTA y debían ser transformadas al formato ARFF para trabajar en Weka. Un archivo en formato ARFF (Attribute-Relation Format File) es un archivo de texto ASCII que describe una serie de instancias que comparten un grupo de atributos. Este tipo de archivos fue desarrollado por el proyecto de *Machine Learning* en la Universidad de Waikato. Para realizar la conversión se utilizó el software Fasta2Weka Converter (dmb.iasi.cnr.it/supbarcodes.php) (Weitschek, Fiscon, & Felici, 2014).

La plataforma WEKA contiene varios métodos para realizar la clasificación y cualquiera de ellos puede aplicarse a la clasificación de *barcodes*. En este trabajo se utilizó un algoritmo de tipo “vecino más cercano” (IBk) perteneciente a una familia de algoritmos “*Lazy*” basados en la estimación de distancias entre casos cuyas clases se encuentran pre-definidas (referencias ó *training set*) y casos que se analizan como problemas de clasificación a resolver. Para evaluar el grado de precisión ó acierto del algoritmo aplicado, se utilizó la opción de validación cruzada (*leave-one-out cross validation*) implementada en la plataforma WEKA (versión 3.6.12. Waikato Environment for Knowledge Analysis, University of Waikato, Hamilton, NZ, <http://www.cs.waikato.ac.nz/ml/weka/>) (Hall, y otros, 2009).

Para medir la capacidad de acierto de la clasificación se utilizó el coeficiente estadístico Kappa (Cohen, 1960).

$$\text{Kappa} = \frac{\text{Pr}(a) - \text{Pr}(e)}{1 - \text{Pr}(e)}$$

donde: Pr(a) es la proporción de aciertos observados y Pr(e) es la probabilidad esperada de aciertos por azar. Kappa permite estimar si la clasificación obtenida puede ser producto del azar ($k \leq 0$) así como el grado de robustez del clasificador utilizado (mayor cuanto más cercano a 1).

4. RESULTADOS Y DISCUSIÓN

4.1 EXTRACCIÓN DE ADN

En la Tabla 4 se presentan los datos de concentración y pureza de las muestras de ADN extraído de tejido fresco. Como se puede observar en todos los casos el rendimiento fue considerable. En cuanto a la pureza, en la mayor parte de las muestras se obtuvo un cociente A_{260}/A_{280} superior a 1,5, lo que indica que no hay presencia de proteínas y contaminantes orgánicos.

En la Tabla 5 se presentan los datos de concentración y pureza del ADN extraído de las muestras comerciales. Durante la industrialización de estas muestras el tejido vegetal fue sometido a algún tipo de procesamiento (deshidratación, molienda, extracción), lo cual pudo haber afectado la integridad del ADN (Bauer, Weller, Hammes, & Hertel, 2003). Esto se corroboró con el análisis electroforético, en el que se observó el ADN parcialmente degradado y en algunos casos bajo rendimiento de la extracción (figuras 4 y 5).

Se repitió la extracción de ADN de las muestras comerciales de ajo y arándano utilizando el kit comercial Nucleospin Food (Macherey - Nagel GmbH & Co. KG, Düren, Germany) en un intento de mejorar la pureza. El objetivo fue alcanzado en las muestras de ajo ya que, como se observa en la tabla 6, aumentó el rendimiento y la pureza. En el caso del arándano, por tratarse de un tejido con alta presencia de compuestos fenólicos, la obtención de ADN de alta pureza no pudo lograrse.

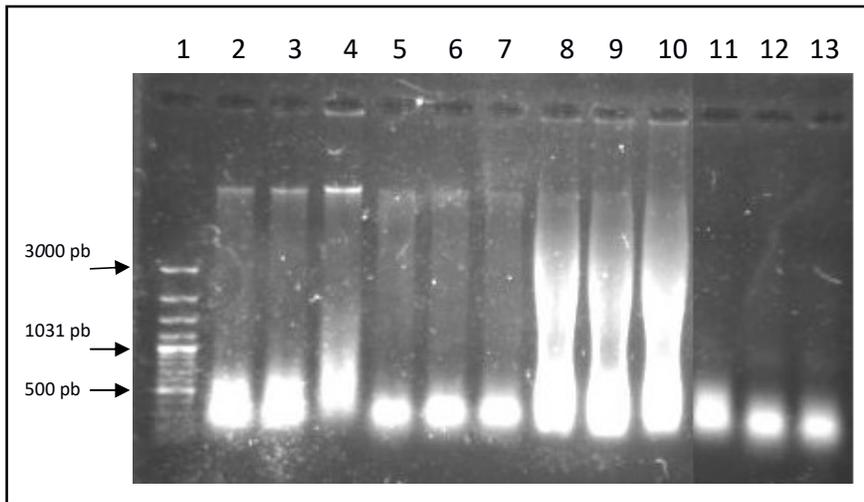


Figura 4. Electroforesis en gel de agarosa al 1% de ADN extraído de muestras comerciales.
Carril 1: Marcador de peso molecular (Mass Ruler DNA Ladder Mix, Fermentas), carril 2: Tomillo 1, carril 3: Tomillo 2, carril 4: Tomillo 3, carril 5: Albahaca 1, carril 6: Albahaca 2, carril 7: Albahaca 3, carril 8: Estragón 1, carril 9: Estragón 2, carril 10: Estragón 3, carril 11: Orégano 1, carril 12: Orégano 2, carril 13: Orégano 3

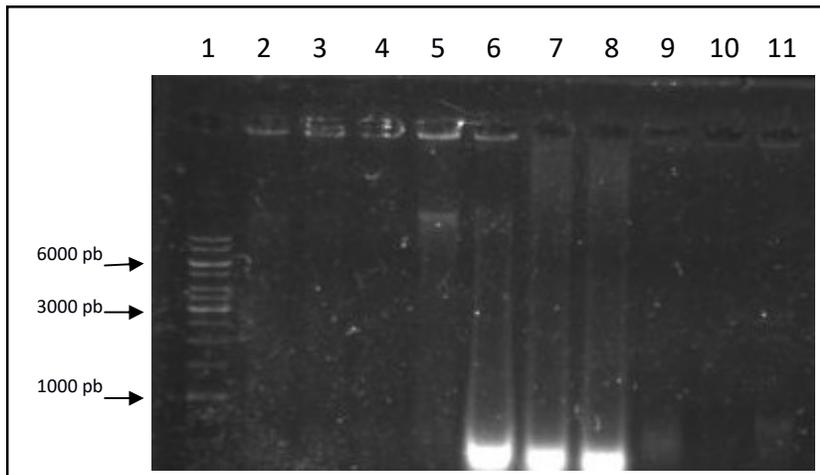


Figura 5. Electroforesis en gel de agarosa al 1% de ADN extraído de muestras comerciales.
Carril 1: Marcador de peso molecular (GeneRuler 1 Kb DNA Ladder, Fermentas), carril 2: Ajo en escamas 1, carril 3: Ajo en escamas 2, carril 4: Ajo molido 1, carril 5: Ajo molido 2, carril 6: Brócoli 1, carril 7: Brócoli 2, carril 8: Brócoli 3, carril 9: Arándano 1, carril 10: Arándano 2, carril 11: Arándano 3

Tabla 4. Concentración y pureza de los ADN obtenidos de tejido fresco de muestras de referencia utilizando el protocolo puesto a punto para la extracción de ADN de alto peso molecular a partir de tejidos vegetales.

Muestra	Concentración	Absorbancia	Absorbancia	A ₂₆₀ /A ₂₈₀
	ng/μl	260 nm	280 nm	
<i>Achyrocline flaccida</i> 1	1832	36,6	17,7	2,1
<i>Achyrocline flaccida</i> 2	684	13,7	6,3	2,2
<i>Achyrocline flaccida</i> 3	653,1	13,1	6,9	1,9
<i>Achyrocline satuireioides</i> 1	661,5	13,2	6,8	1,9
<i>Achyrocline satuireioides</i> 2'	386,4	7,7	3,7	2,1
<i>Achyrocline satuireioides</i> 3	1402,3	28	13,6	2,1
<i>Origanum majorana</i> 2	213,5	4,3	2,9	1,5
<i>Origanum vulgare</i> 1	191,4	3,8	2,8	1,4
<i>Origanum vulgare</i> 2	234,2	4,7	3	1,6
<i>Origanum vulgare</i> 3	262,3	5,2	3,5	1,5
<i>Thymus vulgaris</i> 1	309,3	6,2	3,1	2
<i>Thymus vulgaris</i> 2	251	5	2,4	2,1
<i>Thymus vulgaris</i> 3	463	9,3	4,6	2
<i>Allium sativum</i> 1	235,4	4,7	2,3	2
<i>Allium sativum</i> 2	516,4	10,3	5,6	1,8
<i>Allium sativum</i> 3	818,5	16,4	8	2
<i>Baccharis trimera</i> 1	642,9	12,9	6,6	1,9
<i>Baccharis trimera</i> 2	496,1	9,9	5	2
<i>Baccharis trimera</i> 3'	515,9	10,3	5,2	2
<i>Petroselinum crispum</i> 1	2807,5	56,2	26,9	2,1
<i>Ocimum spp</i> 1	270,4	5,4	2,6	2,1
<i>Ocimum spp</i> 2	20,3	0,4	0,3	1,4
<i>Ocimum spp</i> 3	50,6	1	0,6	1,6
<i>Artemisia dracunculus</i> 1	219,5	4,4	2,9	1,5
<i>Artemisia dracunculus</i> 2	186,1	3,7	2,5	1,5
<i>Artemisia dracunculus</i> 3	497,4	10	5,9	1,7
<i>Aloysia triphylla</i> 1	1467	29,3	13,6	2,2
<i>Aloysia triphylla</i> 3	401,7	8	2,5	3,2

Tabla 5. Concentración y pureza de los ADN obtenidos de muestras comerciales industrializadas utilizando el protocolo puesto a punto para la extracción de ADN de alto peso molecular a partir de tejidos vegetales.

Muestra	Concentración	Absorbancia	Absorbancia	A_{260}/A_{280}
	ng/ μ l	260 nm	280 nm	
Tomillo 1	266,1	5,3	2,8	1,9
Tomillo 2	254,5	5,1	2,6	1,9
Tomillo 3	213,7	4,3	2,4	1,8
Albahaca 1	264,4	5,3	3,3	1,6
Albahaca 2	230,5	4,6	2,7	1,7
Albahaca 3	224,4	4,5	2,6	1,7
Estragón 1	929,11	18,6	10,8	1,7
Estragón 2	575,4	11,5	6,4	1,8
Estragón 3	749,5	15,0	8,1	1,8
Orégano 1	217,1	4,3	2,1	2,1
Orégano 2	198,9	4,0	1,7	2,3
Orégano 3	205,8	4,1	1,8	2,2
Ajo en escamas 1	113,4	2,3	2,1	1,1
Ajo en escamas 2	87,7	1,8	1,3	1,4
Ajo molido 1	39,6	0,8	0,6	1,4
Ajo molido 2	109,0	2,2	1,3	1,7
Brócoli en cápsulas 1	3112,1	62,2	30,8	2,0
Brócoli en cápsulas 2	2300,9	46,0	36,5	1,3
Brócoli en cápsulas 3	62,1	1,2	1,2	1,1
Arándano en cápsulas 1	72,6	1,4	1,0	1,4
Arándano en cápsulas 2	19,3	0,4	0,5	0,8
Arándano en cápsulas 3	35,9	0,7	1,0	0,7

Tabla 6. Concentración y pureza de los ADN obtenidos de la extracción a partir de muestras comerciales industrializadas utilizando el kit comercial.

Muestra	Concentración ng/μl	Absorbancia 260 nm	Absorbancia 280 nm	A₂₆₀/A₂₈₀
Ajo en escamas	559,9	11,2	5,7	2,0
Ajo molido	408,1	8,2	4,2	2,0
Arándano	38,5	0,8	0,8	1,0

4.2. OBTENCIÓN DE SECUENCIAS *BARCODE*

4.2.1. Locus *rbcl*

Se definió el uso de una región en el extremo 5' del gen correspondiente a la ribulosa bifosfato carboxilasa (*rbcl*). En la secuencia de *Arabidopsis thaliana* esta región se encuentra comprendida entre las bases 1 y 599 como se observa en la figura 6. Se seleccionaron los primers *rbcl*_{a_f} y *rbcl*_{a_r}, debido a que en varios estudios quedó demostrada su universalidad (Kress & Erickson, 2007) (CBOL Plant Working Group, 2009).

Con el protocolo de amplificación puesto a punto se analizaron un total de 29 muestras de referencia pertenecientes a las familias Amaryllidaceae, Apiaceae, Astereaceae, Lamiaceae, Verbenaceae. En todos los casos se obtuvieron productos de amplificación del tamaño esperado como se puede ver en la tabla 7 y en las figuras 7 y 8.

Fueron analizadas 9 muestras industrializadas de origen comercial de perejil, tomillo, albahaca, estragón, orégano, ajo en escamas, ajo molido, brócoli y arándano en las presentaciones que se detallan en la tabla 2. En todos los casos se obtuvieron productos de amplificación excepto la muestra de arándano que no amplificó en la primera instancia. Este hecho era de esperar ya que en la extracción se obtuvo poco ADN de baja pureza. Se intentó resolver esta situación haciendo una segunda amplificación en la cual el ADN molde fuera el producto de la primera amplificación. En este caso tampoco se obtuvo producto para la muestra de arándano en cápsulas (Figura 9).

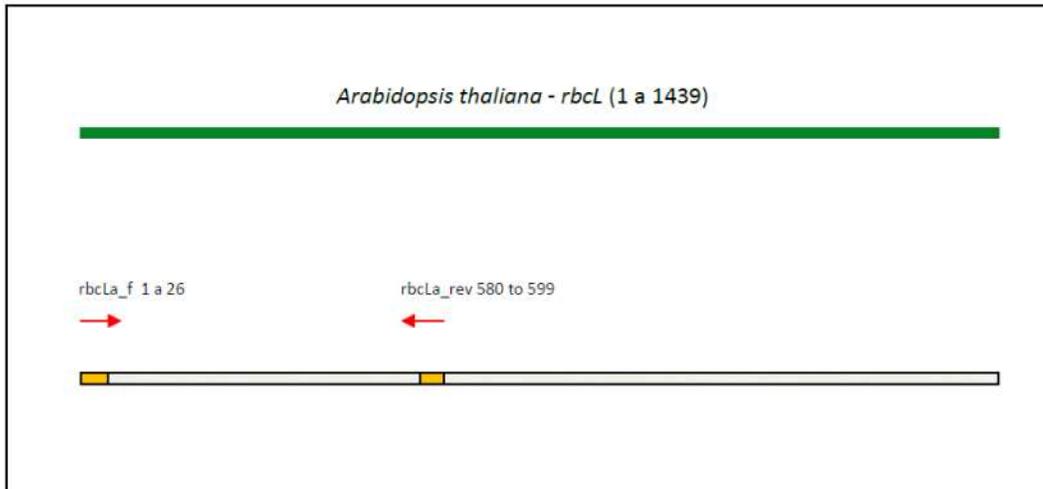


Figura 6. Mapa de la ubicación en el gen *rbcL* de *Arabidopsis thaliana* de los cebadores utilizados. Extraído y modificado de <http://www.kew.org/barcoding>.

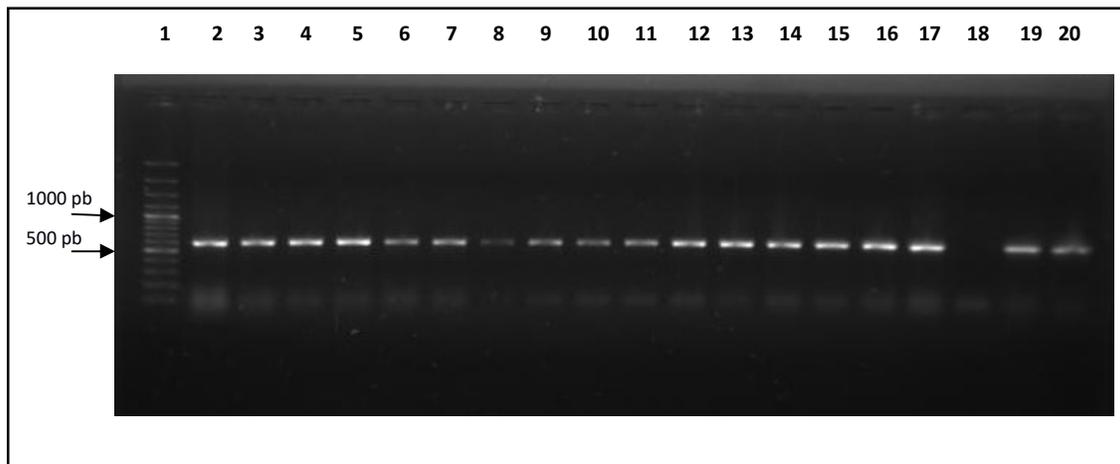


Figura 7. Amplificación del locus *rbcL* en muestras de referencia. Carril 1: Gene Ruler 100 bp DNA Ladder Plus (Fermentas), carril 2: *A.flaccida* (1), carril 3: *A.flaccida* (2), carril 4: *A.flaccida* (3), carril 5: *A. satureioides* (1), carril 6: *A. satureioides* (2'), carril 7: *A. satureioides* (3), carril 8: *O. majorana* (2), carril 9: *O. vulgare* (1), carril 10: *O. vulgare* (2), carril 11: *O. vulgare* (3), carril 12: *T. vulgaris* (1), carril 13: *T. vulgaris* (2), carril 14: *T. vulgaris* (3), carril 15: *A. sativum* (1), carril 16: *A. sativum* (2), carril 17: *A. sativum* (3), carril 18: control negativo, carril 19: *B. trimmera* (1), carril 20: *B. trimmera* (2).

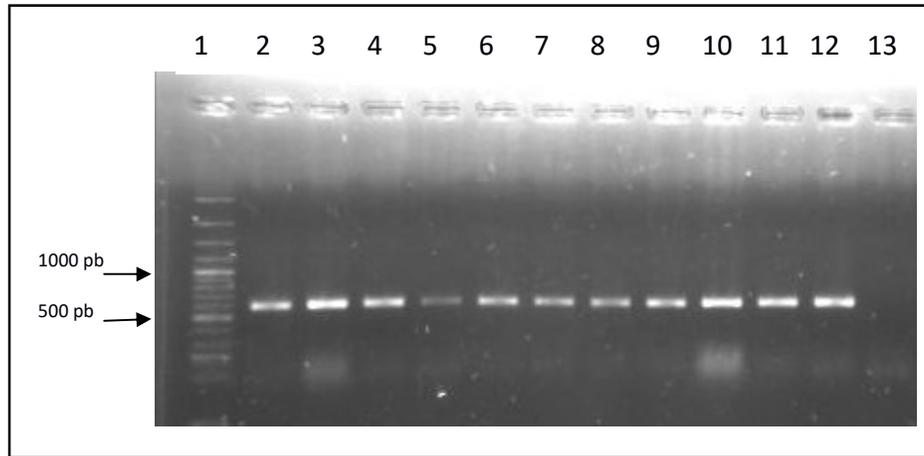


Figura 8. Amplificación del locus *rbcL* en muestras de referencia. Carril 1: Gene Ruler 100 bp DNA Ladder Plus (Fermentas), carril 2: *B. trimmera* (3'), carril 3: *P. crispum* (1), carril 4: *Ocimum* (1), carril 5: *Ocimum* (2), carril 6: *Ocimum* (3), carril 7: *A. dracunculus* (1), carril 8: *A. dracunculus* (2), carril 9: *A. dracunculus* (3), carril 10: *A. triphylla* (1), carril 11: *A. triphylla* (2), carril 12: *A. triphylla* (3), carril 13: control negativo.

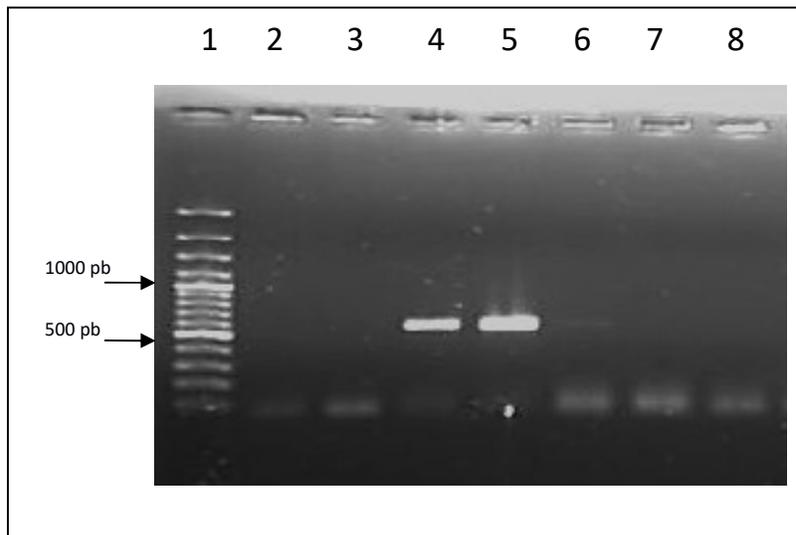


Figura 9. Reamplificación del locus *rbcL* en muestras en las que no se obtuvo amplicón en la amplificación inicial. Carril 1: Gene Ruler 100 bp DNA Ladder Plus (Fermentas), carril 2: control negativo, carril 3: control negativo, carril 4: *O. vulgare* (2), carril 5: *Ocimum* (3), carril 6: mezcla estragón- albahaca (1:3), carril 7: Bilberry (arándano en cápsulas), carril 8: control negativo reamplificación.

Tabla 7. Resultados de amplificación y características de secuencias obtenidas de los loci *rbcl* y *matK* en muestras de referencia

Especie	Muestras	Amplificación		Secuencias <i>rbcl</i>					Secuencias <i>matK</i>				
		<i>rbcl</i>	<i>matK</i>	Directa (bases)	% QV20 directa (bases)	Reversa (bases)	% QV20 reversa (bases)	Secuencia consenso (bases)	Directa (bases)	% QV20 directa (bases)	Reversa (bases)	% QV20 reversa (bases)	Secuencia consenso (bases)
<i>Allium sativum</i>	1	si	si	857	546	812	537	511	838	837	838	378	588
	2	si	si	804	532	815	541	510	853	842	908	837	645
	3	si	si	926	68	876	539		845	844	840	828	640
<i>Ocimum sp.</i>	1	si	si	912	534	856	550	501	811	811	810	128	432
	2	si	no	652	217	839	539	484	*	*	*	*	*
	3	si	si	610	537	846	536	426	807	806	809	180	600
<i>Baccharis trimmera</i>	1	si	si	822	529	862	552	516	824	592	921	371	737
	2	si	si	838	536	812	551	520	822	821	924	824	713
	3	si	si	825	537	856	542	515	820	819	921	494	713
<i>Aloysia triphylla</i>	1	si	si	848	341	846	551	501	825	824	807	417	594
	2	si	si	919	534	840	567	499	829	828	928	830	566
	3	si	si	826	342	838	543	494	826	825	922	829	580

En la tabla se detalla para cada muestra y locus si se obtuvieron productos de amplificación, si se obtuvieron secuencias directa y reversa, y en caso afirmativo el largo en pb y la cantidad de bases de calidad \geq QV20%. (*) No hubo amplificación de ese locus para esa muestra por lo cual no se generó una secuencia.

Tabla 7. Resultados de amplificación y características de secuencias obtenidas de los loci *rbcl* y *matK* en muestras de referencia

Especies	Muestras	Amplificación		Secuencias <i>rbcl</i>					Secuencias <i>matK</i>				
		<i>rbcl</i>	<i>matK</i>	Directa (bases)	% QV20 directa (bases)	Reversa (bases)	% QV20 reversa (bases)	Secuencia consenso (bases)	Directa (bases)	% QV20 directa (bases)	Reversa (bases)	% QV20 reversa (bases)	Secuencia consenso (bases)
<i>Artemisia dracunculus</i> L.	1	si	si	654	526	655	569	504	826	825	828	827	595
	2	si	si	663	527	845	541	492	827	826	832	687	633
	3	si	si	841	283	850	551	499	830	829	922	830	592
<i>Achyrocline satureioides</i>	1	si	si	616	529	729	546	507	843	829	922	785	384
	2	si	si	917	543	854	564	435	840	825	909	790	496
	3	si	si	672	288	670	553	439	839	822	892	787	467
<i>Achyrocline flaccida</i>	1	si	si	831	534	837	546	500	828	819	927	789	550
	2	si	si	807	545	849	547	501	822	821	916	835	551
	3	si	si	804	528	845	542	501	840	824	908	782	500

En la tabla se detalla para cada muestra y locus si se obtuvieron productos de amplificación, si se obtuvieron secuencias directa y reversa, y en caso afirmativo el largo en pb y la cantidad de bases de calidad \geq QV20%. (*) No hubo amplificación de ese locus para esa muestra por lo cual no se generó una secuencia.

Tabla 7. Resultados de amplificación y características de secuencias obtenidas de los loci *rbcl* y *matK* en muestras de referencia

Especies	Muestras	Amplificación		Secuencias <i>rbcl</i>					Secuencias <i>matK</i>				
		<i>rbcl</i>	<i>matK</i>	Directa (bases)	% QV20 directa (bases)	Reversa (bases)	% QV20 reversa (bases)	Secuencia consenso (bases)	Directa (bases)	% QV20 directa (bases)	Reversa (bases)	% QV20 reversa (bases)	Secuencia consenso (bases)
<i>Origanum majorana</i>	1	si	si	846	527	852	532	519	924	824	846	115	451
<i>Origanum vulgare</i>	1	si	no	625	304	597	539	498	*	*	*	*	*
	2	si	si	667	529	792	539	498	802	801	799	179	507
	3	si	si	650	529	649	531	509	826	457	794	158	525
<i>Petroselinum crispum</i>	1	si	si	787	345	677	548	505	918	819	924	379	550
<i>Thymus vulgaris</i>	1	si	si	859	534	814	343	474	811	810	746	210	545
	2	si	si	806	506	631	538	521	814	813	827	303	532
	3	si	si	808	529	829	533	505	790	24	907	313	**

En la tabla se detalla para cada muestra y locus si se obtuvieron productos de amplificación, si se obtuvieron secuencias directa y reversa, y en caso afirmativo el largo en pb y la cantidad de bases de calidad \geq QV20%. (*) No hubo amplificación de ese locus para esa muestra por lo cual no se generó una secuencia. (*) No hubo amplificación de ese locus para esa muestra por lo cual no se generó una secuencia. (**) La secuencia directa no fue de buena calidad por lo que no se pudo generar una secuencia consenso con la reversa. Se realizaron los análisis posteriores tomando únicamente la secuencia reversa.

La puesta a punto de la reacción de PCR para la amplificación de la región *rbcl* implicó varias modificaciones de las condiciones descritas por Erickson (CBOL Plant Working Group, 2009). Entre ellas, se probaron Taq polimerasas de distintas procedencias; se modificó la temperatura de hibridación de los cebadores y el ciclo; se modificaron las concentraciones de reactivos y ADN molde; y se agregó BSA y DMSO como potenciadores de la reacción (Frackman, Kobs, Simpson, & Storts, 1998).

De acuerdo a los resultados obtenidos confirmamos la universalidad de los cebadores *rbclLa_f* y *rbclLa_rev* en nuestras condiciones de trabajo.

Se obtuvieron secuencias bidireccionales de buena calidad en el 83% de las muestras de referencia como se detalla en la Tabla 7. En los Anexos 1 y 2 se presentan los electroferogramas obtenidos para las muestras *T. vulgaris* 2 (locus *rbcl*, secuencia reversa) y *A. triphylla* 3 (locus *matK*, secuencia directa).

Algunos problemas surgieron con una muestra de *Origanum vulgare*, dos muestras de *Ocimum*, una de *Thymus vulgaris* y *Origanum majorana*. Se pudieron resolver volviendo a amplificar y secuenciar estas muestras. El largo promedio de las secuencias fue 575 bases, siendo muy homogéneo en todas las especies.

En el caso de las muestras comerciales industrializadas el 50% de las secuencias bidireccionales obtenidas fueron de calidad aceptable. Las muestras de albahaca y estragón debieron ser amplificadas y secuenciadas en una segunda instancia con lo que se logró mejorar la calidad. En el anexo 5 se puede observar la superposición de picos en la primera secuencia directa del locus *rbcl* correspondiente a la muestra comercial de albahaca. No se obtuvieron resultados para la muestra de arándano.

Para cada muestra se ensamblaron la secuencia directa y la reversa para generar una secuencia consenso, utilizada posteriormente para la identificación.

Los resultados obtenidos para la región *rbcl* avalan las conclusiones del estudio realizado por el grupo de plantas del CBOL en las especies objeto de estudio de este trabajo. Este locus es fácilmente amplificable con los cebadores disponibles, lo que demuestra su universalidad, y las secuencias generadas son de buena calidad en general requiriendo por tanto poca edición manual (CBOL Plant Working Group, 2009).

4.2.2. Locus *matK*

La región *matK* propuesta como *barcode* se encuentra comprendida entre las bases 205 y 1046 del gen de la maturasa K en la secuencia de *Arabidopsis thaliana*. Se utilizaron los cebadores 3F_KIM f y 1R_KIM r ya que son los sugeridos por el grupo de plantas del CBOL para amplificar la región *matK* en angiospermas (CBOL Plant Working Group, 2009). La ubicación de los mismos se señala en la figura 10.

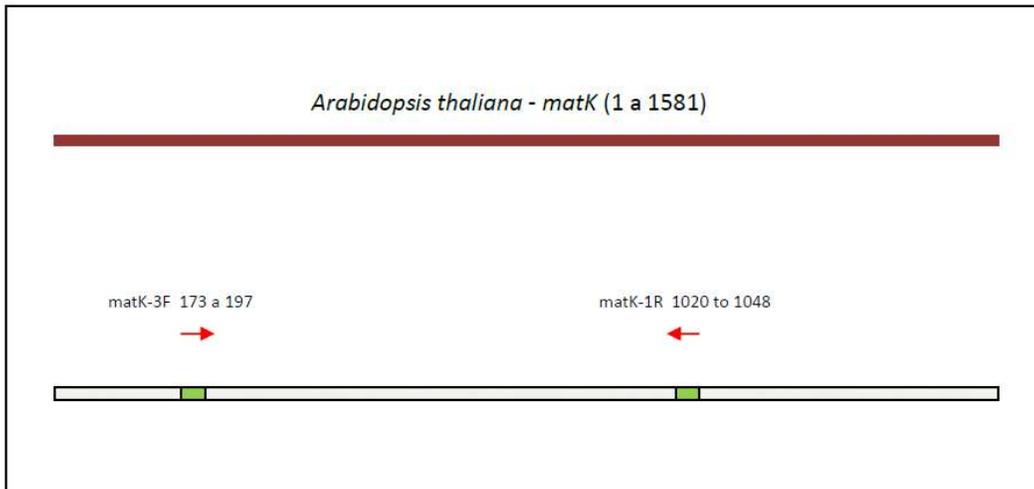


Figura 10. Mapa de la ubicación en el gen *matK* de *Arabidopsis thaliana* de los cebadores utilizados. Extraído y modificado de <http://www.kew.org/barcoding>.

Con el protocolo puesto a punto se analizaron un total de 29 muestras de referencia pertenecientes a las familias Amaryllidaceae, Apiaceae, Astereaceae, Lamiaceae, Verbenaceae. En el 86% de los casos se obtuvieron productos de amplificación del tamaño esperado como se observa en las figuras 11, 12, 13 y en la tabla 7. La amplificación falló en las especies *Origanum majorana*, *Origanum vulgare* y *Ocimum* pertenecientes a la familia de las Lamiaceae. En estos casos se repitió la amplificación partiendo de la mitad de la concentración de ADN (100 ng), obteniéndose resultados exitosos.

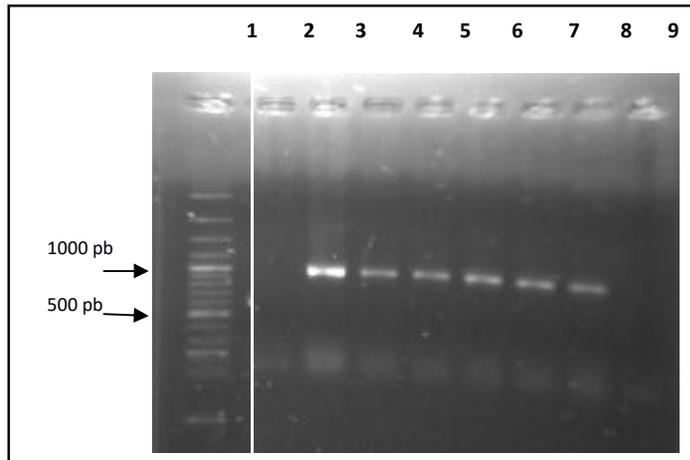


Figura 11. Amplificación del locus *matK*. Carril 1: Gene Ruler 100 bp DNA Ladder Plus (Fermentas), carril 2: control negativo, carril 3: *A. flaccida* (1), carril 4: *A. flaccida* (2), carril 5: *A. flaccida* (3), carril 6: *A. satureioides* (1), carril 7: *A. satureioides* (2), carril 8: *A. satureioides* (3), carril 9: *O. majorana*

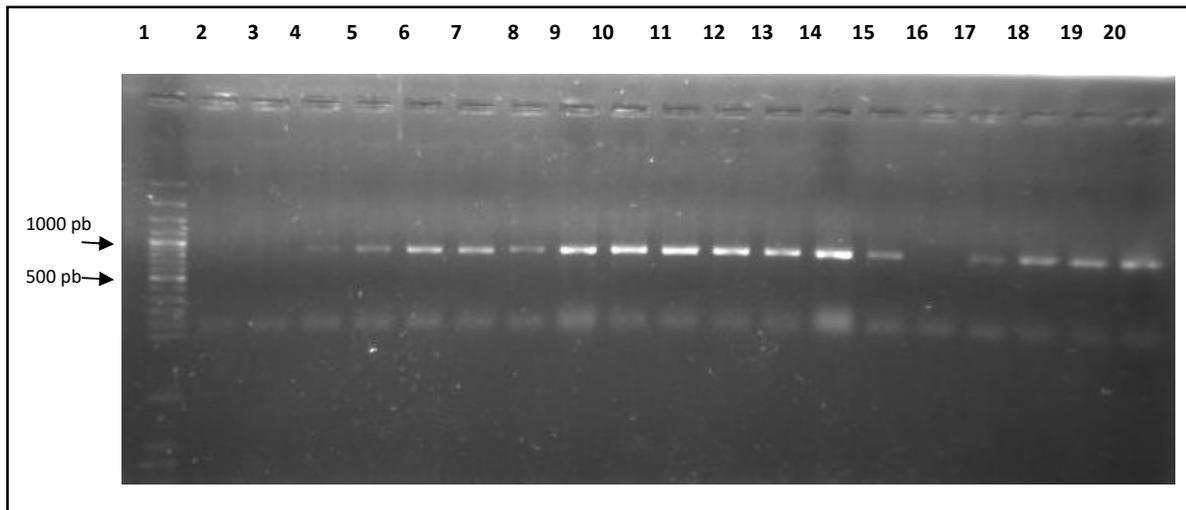


Figura 12. Amplificación del locus *matK*. Carril 1: Gene Ruler 100 bp DNA Ladder Plus (Fermentas), carril 2: *O. vulgare* (1), carril 3: *O. vulgare* (2), carril 4: *O. vulgare* (3), carril 5: *T. vulgaris* (1), carril 6: *T. vulgaris* (2), carril 7: *T. vulgaris* (3), carril 8: *A. sativum* (1), carril 9: *A. sativum* (2), carril 10: *A. sativum* (3), carril 11: *B. trimmera* (1), carril 12: *B. trimmera* (2), carril 13: *B. trimmera* (3), carril 14: *P. crispum* (1), carril 15: *Ocimum* (1), carril 16: *Ocimum* (2), carril 17: *Ocimum* (3), carril 18: *A. dracunculus* (1), carril 19: *A. dracunculus* (2), carril 20: *A. dracunculus* (3)

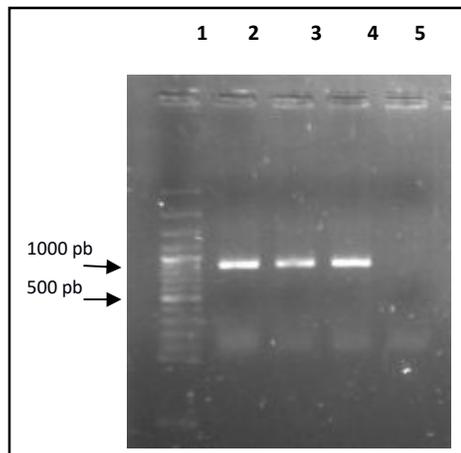


Figura 13. Amplificación del locus *matK*. Carril 1: Gene Ruler 100 bp DNA Ladder Plus (Fermentas), carril 2: *A. triphylla* (1), carril 3: *A. triphylla* (2), carril 4: *A. triphylla* (3), carril 5: control negativo

Fueron analizadas 9 muestras comerciales industrializadas de perejil, tomillo, albahaca, estragón, orégano, ajo en escamas, ajo molido, brócoli y arándano en las presentaciones que se detallan en la tabla 2. No se obtuvieron productos de amplificación para la muestra de arándano. En el caso de albahaca y orégano fue necesaria una segunda amplificación cambiando las condiciones de reacción.

Las secuencias obtenidas para las muestras de referencia fueron de mediana y baja calidad en el 37 % de los casos. En todas las secuencias se observó una pérdida de resolución alrededor de las 500 bases, como puede verse en la secuencia correspondiente a una muestra de *A. triphylla* en el Anexo 2. El largo promedio fue de 793 bases, observándose mucha variación entre las muestras. No se obtuvieron resultados en una muestra de *Origanum vulgare* y en una de *Ocimum*. Las dos muestras de *Ocimum* restantes debieron ser amplificadas y secuenciadas nuevamente con lo que se mejoró la intensidad de señal y la calidad. Lo mismo ocurrió con una muestra de *Thymus vulgaris* y *Petroselinum crispum*.

Las secuencias bidireccionales obtenidas fueron de calidad aceptable en el 35 % de los casos de las muestras comerciales. La muestra de albahaca fue secuenciada nuevamente pero no se logró mejorar la calidad del resultado. En los Anexos 4 y 6 se presentan las secuencias obtenidas para las muestras de orégano y albahaca.

Varios autores han reportado dificultades en la amplificación y la obtención de secuencias bidireccionales de buena calidad de la región *matK* en distintas especies (Fazekas, y otros, 2008) (Kress & Erickson, 2007) (Hollingsworth, Graham, & Little, 2011) (Theodoridis, y otros, 2012). Este hecho fue corroborado en este estudio observándose dificultades para la amplificación y secuenciación de esta región tanto en las muestras de tejido fresco como en muestras procesadas. A pesar de ello, *matK*

sigue siendo considerado como locus de referencia junto a *rbcl* por el CBOL Plant Group. Esto es debido a que es una de las regiones codificantes del genoma plastídico que evoluciona mas rápidamente lo que le da un gran poder de discriminación entre especies (CBOL Plant Working Group, 2009) (Hollingsworth, Graham, & Little, 2011).

4.3. ANÁLISIS DE DATOS MEDIANTE BARCODE OF LIFE DATA SYSTEMS (BOLD)

4.3.1. Clasificación

Con las secuencias bidireccionales obtenidas y editadas manualmente se generó una secuencia consenso para cada muestra y cada *locus*. Con ellas se llevó a cabo la identificación en la base de datos de *Barcode of Life* (BOLD) de cada *locus* por separado, ya que el motor de BOLD no está diseñado para utilizar ambos marcadores en la misma instancia, pese a que se recomienda el uso de dos *loci* para plantas.

Se registró la coincidencia más cercana, el vecino más cercano perteneciente al mismo género (VMCIG) y el vecino más cercano perteneciente a un género distinto (VMCDG). En las tablas 8 y 9 pueden observarse los resultados correspondientes al *locus rbcl* y al *locus matK* respectivamente. En todos los casos se reportó el *score* y el porcentaje de identidad entre la secuencia problema y la secuencia blanco. En aquellos casos en que la identificación arrojó varios taxones con el mismo *score* y porcentaje de identidad, se reportaron los más cercanos a la especie de referencia cuando fue posible.

Tabla 8. Clasificación de muestras de referencia mediante BOLD usando secuencias del locus *rbcl*.

Especie	Muestra	Mejor coincidencia	Score	% Identidad	Vecino más cercano igual género	Score	% Identidad	Vecino más cercano distinto género	Score	% Identidad	Especie/ Género asignado
<i>Allium sativum</i>	1	<i>Allium sativum</i>	511	100	<i>Allium ursinum</i>	507	99.61	*	*	*	<i>A.sativum</i>
	2	<i>Allium sativum</i>	510	100	<i>Allium ursinum</i>	506	99.61	*	*	*	<i>A.sativum</i>
	3	<i>Allium sativum</i>	521	100	<i>Allium ursinum</i>	517	99.62	*	*	*	<i>A.sativum</i>
<i>Ocimum sp.</i>	1	<i>Ocimum tenuiflorum/O. americanum/O. sp.</i>	501	100	<i>Ocimum sanctum</i>	499	100	<i>Plectranthus barbatus/Coleus blumei/Solenostemon scutellarioides/Anisochilus carnosus/Ortosiphon stamineus</i>	495	99.6	Ocimum
	2	<i>Ocimum tenuiflorum/O. americanum/O. sp.</i>	484	100	<i>Ocimum sanctum/tenuiflorum</i>	482	100	<i>Clerodendrathus spicatus/Coleus blumei/Solenostemon scutellarioides/Anisochilus carnosus/Plectranthus mollis</i>	478	99.59	Ocimum
	3	<i>Ocimum basilicum/O. kilimandscharicum/O. americanum/etc</i>	424	99.53	<i>Ocimum tenuiflorum/ etc</i>	424	99.3	<i>Origanum majorana/ Mentha sp./ Rivea hypocrateriformis/Garcinia forbesi</i>	424	99.06	**
<i>Baccharis trimera</i>	1	<i>Baccharis trimera</i>	515	100	<i>Baccharis trinervis</i>	504	99.8	<i>Symphiotrichum novae-angliae</i>	512	99.61	<i>B. trimera</i>
	2	<i>Baccharis trimera</i>	518	100				<i>Solidago nemoralis</i>	513	99.23	<i>B. trimera</i>
	3	<i>Baccharis trimera</i>	514	100	<i>Baccharis trinervis</i>	504	99.8	<i>Symphiotricum novae -angliae</i>	512	99.61	<i>B. trimera</i>
<i>Aloysia triphylla</i>	1	<i>Aloysia citrodora</i>	501	100	<i>Lippia stoechadifolia</i>	477	99.39	<i>Acantholippia seriphioides</i>	487	99.39	<i>A.citrodora</i>
	2	<i>Aloysia citrodora</i>	499	100	<i>Lippia myriocephala</i>	477	97.8	<i>Acantholippia seriphioides</i>	487	99.39	<i>A.citrodora</i>
	3	<i>Aloysia citrodora</i>	494	100	<i>Lippia stochadifolia</i>	477	98.38	<i>Acantholippia seriphioides</i>	488	99.39	<i>A.citrodora</i>
<i>Artemisia dracunculus</i> L.	1	<i>Artemisia dracunculus</i>	502	100	<i>Artemisia frigida/A. scoparia</i>	502	99.6	<i>Chrysanthemum indicum</i>	500	99.4	Artemisia
	2	<i>Artemisia dracunculus</i>	491	99.8	<i>Artemisia annua/A. frigida/A.lavandulifolia/A. gmelinii/A.scoparia</i>	491	99.8	<i>Chrysanthemum indicum</i>	489	99.59	Artemisia
	3	<i>Artemisia dracunculus</i>	499	100	<i>Artemisia annua/A. frigida/A.lavandulifolia/A. gmelinii/A.scoparia</i>	499	100	<i>Chrysanthemum indicum/C. morifolium</i>	497	99.8	Artemisia

Especie	Muestra	Mejor coincidencia	Score	% Identidad	Vecino más cercano igual género	Score	% Identidad	Vecino más cercano distinto género	Score	% Identidad	Especie/ Género asignado
<i>Achyrocline satureioides</i>	1	<i>Gnaphalium</i>	420	99.76	<i>Gnaphalium sylvaticum/G.norvegicum/G.su pinum</i>	418	99.52	<i>Helichrysum cymosum</i>	420	99.76	**
	2	<i>Gnaphalium</i>	434	99.77	<i>Gnaphalium sylvaticum/G.supinum</i>	432	99.54	<i>Helichrysum cymosum/Helychrisum populifolium/H.kraussii</i>	434	99.77	**
	3	<i>Gnaphalium</i>	438	99.77	<i>Gnaphalium sylvaticum/G.norvegicum/G.su pinum</i>	436	99.54	<i>Helichrysum cymosum/Metalasia densa</i>	438	99.77	**
<i>Achyrocline flaccida</i>	1	<i>Gnaphalium</i>	499	99.8	<i>Gnaphalium uliginosum</i>	493	99.2	<i>Annaphilis margaritacea/Antennaria neglecta</i>	495	99.4	<i>Gnaphalium</i>
	2	<i>Gnaphalium</i>	499	99.6	<i>Gnaphalium uliginosum</i>	493	99.0	<i>Annaphilis margaritacea/Antennaria neglecta</i>	495	99.2	<i>Gnaphalium</i>
	3	<i>Gnaphalium</i>	501	100	<i>Gnaphalium uliginosum</i>	495	99.4	<i>Annaphilis margaritacea/Antennaria neglecta</i>	497	99.6	<i>Gnaphalium</i>
<i>Origanum majorana</i>		<i>Origanum majorana</i>	519	100	<i>Origanum vulgare</i>	517	100	<i>Ocimum sp.</i>	519	100	**
<i>Origanum vulgare</i>	1	<i>Origanum vulgare</i>	498	100	<i>Origanum. majorana</i>	498	100	<i>Ocimum sp.</i>	498	100	**
	2	<i>Origanum vulgare</i>	498	100	<i>Origanum. majorana</i>	498	100	<i>Ocimum sp./Satureja vulgaris/Pycnanthemum verticillatum</i>	498	100	**
	3	<i>Origanum vulgare</i>	509	100	<i>Origanum majorana</i>	509	100	<i>Ocimum sp.</i>	509	100	**
<i>Petroselinum crispum</i>		<i>Petroselinum crispum</i>	502	100	<i>Petroselinum segetum</i>	485	99.0	<i>Sium suave/Anethum graveolens</i>	494	98.8	<i>P. crispum</i>
<i>Thymus vulgaris</i>	1	<i>Thymus vulgaris</i>	474	100	<i>Thymus praecox/T. Polytrichus/ T. pulegioides</i>	474	100	<i>Mentha suaveolens/M. spicata/M.pulegium/M.suaveolens/M. arvensis/C. asinos</i>	472	99.79	<i>Thymus</i>
	2	<i>Thymus vulgaris</i>	521	100	<i>Thymus quinquecostatus</i>	517	99.6	<i>Bystropogon origanifolius</i>	518	99.81	<i>T. vulgaris</i>
	3	<i>Thymus vulgaris</i>	505	100	<i>Thymus quinquecostatus</i>	501	99.6	<i>B. origanifolius/Mentha spicata</i>	503	99.8	<i>T. vulgaris</i>

En la tabla se presentan los datos obtenidos de la comparación de las secuencias correspondientes a las muestras de referencia con la base de datos BOLD. Se describe la coincidencia mas cercana, el/los vecino/s mas cercano de igual género y el/los vecino/s mas cercanos pertenecientes a otros género clasificados según su score y % de identidad con la secuencia problema. Por último se informa la especie/género asignado de acuerdo a los datos obtenidos. (*) No se encontraron vecinos mas cercanos de otro género en la base de datos. (**) De acuerdo a los datos no se puede asignar especie ni género. (***) No se obtuvieron secuencias para ese locus

Tabla 9. Clasificación de muestras de referencia mediante BOLD usando secuencias del locus *matK*.

Especie	Muestra	Mejor coincidencia	Score	% Identidad	Vecino más cercano igual género	Score	% Identidad	Vecino más cercano distinto género	Score	% Identidad	Especie/ Género asignado
<i>Allium sativum</i>	1	<i>Allium sativum</i>	587	99.83	<i>A. scorodoprasum</i>	577	98.98	*			<i>A. sativum</i>
	2	<i>Allium sativum</i>	645	100	<i>A. vineale/ A. scorodoprasum/ A. ampeloprasum</i>	633	99.07	<i>Mesua ferrea</i>	575	96.15	<i>A. sativum</i>
	3	<i>Allium sativum</i>	640	100	<i>A. vineale/A. scorodoprasum/A. ampeloprasum</i>	628	99.06	*			<i>A. sativum</i>
<i>Ocimum sp.</i>	1	<i>Ocimum basilicum</i>	429	99.31	<i>O.sp./O.kilimandscharicum/O. citriodorum/O.americanum</i>	429	99.31	<i>Mentha spicata</i>	429	99.31	**
	2	***			***			***			***
	3	<i>Ocimum sp.</i>	598	100	<i>O.citrodorum/O. kilimandscharicum/O. basilicum</i>	598	99.67	<i>Mentha spicata</i>	598	99.67	**
<i>Baccharis trimera</i>	1	<i>Baccharis trimera</i>	737	100	<i>Baccharis trinervis</i>	731	99.59	<i>Oclemena nemoralis/ Aster umbellatus/ A. ontarionis</i>	731	99.59	<i>B. trimera</i>
	2	<i>Baccharis trimera</i>	713	100	<i>Baccharis trinervis</i>	707	99.58	<i>Oclemena nemoralis/ O. acuminata/ Aster umbellatus/ A. ontarionis/ /Doellingeria umbelatta</i>	707	99.58	<i>B. trimera</i>
	3	<i>Baccharis trimera</i>	712	100	<i>Baccharis trinervis</i>	704	99.3	<i>Doellingeria umbellata/Oclemena nemoralis/O. acuminata/Aster ontarionis/A. umbellatus</i>	704	99.3	<i>B. trimera</i>
<i>Aloysia triphylla</i>	1	<i>Aloysia citrodora</i>	594	100	<i>Lippia stoechadipholia</i>	558	97.13	<i>Lantana camara</i>	562	97.47	<i>A. citrodora</i>
	2	<i>Aloysia citrodora</i>	566	100	<i>Lippia stoechadipholia</i>	533	97.17	<i>Lantana camara</i>	537	97.52	<i>A. citrodora</i>
	3	<i>Aloysia citrodora</i>	580	100	<i>Lippia stoechadipholia</i>	547	97.24	<i>Lantana camara</i>	551	97.58	<i>A.citrodora</i>
<i>Artemisia dracunculus L.</i>	1	<i>Artemisia dracunculus</i>	595	100	<i>A. frigida/A. alaskana/A.rupestris/A.norvegica/A.borealis/A.capillaris/A.glomerata/A.hyperborea/A.campestris/A.biennis</i>	587	99.33	<i>Chrysanthemum indicum/C. lavandulifolium</i>	585	99.16	<i>A. dracunculus</i>
	2	<i>Artemisia dracunculus</i>	633	100	<i>A. frigida/ A.alaskana/A.norvegica/A.borealis/A.campestris/A. biennis</i>	625	99.37	<i>Chrysanthemum indicum/C. lavandulifolium/Ajania gracilis</i>	623	99.21	<i>A. dracunculus</i>
	3	<i>Artemisia dracunculus</i>	592	100	<i>A.frigida/A.alaskana/A.norvegica/A.borealis/A.campestris/A.biennis/A.rupestris /A.hyperborea</i>	582	99.16	<i>Chrysanthemum indicum/C.lavandulifolium/Ajania gracilis</i>	580	98.99	<i>A. dracunculus</i>

En la tabla se presentan los datos obtenidos de la comparación de las secuencias correspondientes a las muestras de referencia con la base de datos BOLD. Se describe la coincidencia mas cercana, el/los vecino/s mas cercano de igual género y el/los vecino/s mas cercanos pertenecientes a otros género clasificados según su score y % de identidad con la secuencia problema. Por último se informa la especie/género asignado de acuerdo a los datos obtenidos. (*) No se encontraron vecinos mas cercanos de otro género en la base de datos. (**) De acuerdo a los datos no se puede asignar especie ni género. (***) No se obtuvieron secuencias para ese locus

Tabla 9. Clasificación de muestras de referencia mediante BOLD usando secuencias del locus *matK*

Especie	Muestra	Mejor coincidencia	Score	% Identidad	Vecino más cercano igual género	Score	% Identidad	Vecino más cercano distinto género	Score	% Identidad	Especie/ Género asignado
<i>Achyrocline satureioides</i>	1	<i>Gnaphalium</i>	384	100	**			<i>Helichrysum / Metalasia densa</i>	384	100	**
	2	<i>Gnaphalium</i>	496	100	**			<i>Helichrysum/Metalasia densa/Senecio</i>	496	100	**
	3	<i>Gnaphalium</i>	467	100	**			<i>Helichrysum/Metalasia/Senecio</i>	467	100	**
<i>Achyrocline flaccida</i>	1	<i>Gnaphalium</i>	550	100	<i>Gnaphalium luteo-album</i>	542	99.27	<i>Helichrysum petiolare/ Metalasia densa/Senecio</i>	550	100	**
	2	<i>Gnaphalium</i>	551	100	<i>Gnaphalium luteo-album</i>	543	99.27	<i>Helichrysum petiolare/Metalasia densa/ Senecio</i>	551	100	**
	3	<i>Gnaphalium</i>	500	100	<i>Gnaphalium luteo-album</i>	492	99.2	<i>Helichrysum petiolare/H. scabrum/Metalasia densa</i>	500	100	**
<i>Origanum majorana</i>		<i>Origanum majorana</i>	451	100	<i>Origanum vulgare</i>	451	100	<i>Thymus vulgaris/T. caespitus</i>	447	99.56	Origanum
<i>Origanum vulgare</i>	1	*			*			*			
	2	<i>Origanum vulgare</i>	506	99.8	<i>Origanum sp./ O. majorana/ O. microphyllum/ O. dayi/O. elongatum/O. rotundifolium/O. dictamnus</i>	506	99.8	<i>Thymus vulgaris/T.caespitus</i>	504	99.61	Origanum
	3	<i>Origanum vulgare</i>	518	100	<i>Origanum sp./ O. majorana/ O. microphyllum/ O. dayi/O. elongatum/O. rotundifolium/O. dictamnus/O. onites</i>	518	98.67	<i>Thymus vulgaris/T. caespitus</i>	516	98.48	Origanum
<i>Petroselinum crispum</i>		<i>Petroselinum crispum</i>	550	100	<i>Petroselinum segetum</i>	550	100	<i>Apium graveolens</i>	544	99.45	Petroselinum
<i>Thymus vulgaris</i>	1	<i>Thymus vulgaris</i>	544	100	<i>T. striatus/T. spinulosus/T. polytrichus</i>	542	99.82	<i>Origanum sp./O. vulgare/O. dictamnus/O. quinquecostatu/O. onites</i>	534	99.26	<i>T. vulgaris</i>
	2	<i>Thymus vulgaris</i>	531	100	<i>T. praecox/T.oenipuntanus/T. pulegioides/T. serpyllum</i>	529	99.62	<i>Origanum sp./O. vulgare/O. onites</i>	523	99.06	<i>T. vulgaris</i>
	3	<i>Thymus vulgaris</i>	560	98.95	<i>T. longicaulis/T. serpyllum/T. striatus/T. spinulosus/T. sipyleus/T. pulegioides</i>	552	98.25	<i>Origanum sp.</i>	543	97.72	Thymus

En la tabla se presentan los datos obtenidos de la comparación de las secuencias correspondientes a las muestras de referencia con la base de datos BOLD. Se describe la coincidencia mas cercana, el/los vecino/s mas cercano de igual género y el/los vecino/s mas cercanos pertenecientes a otros género clasificados según su score y % de identidad con la secuencia problema. Por último se informa la especie/género asignado de acuerdo a los datos obtenidos. (*) No se encontraron vecinos mas cercanos de otro género en la base de datos. (**) De acuerdo a los datos no se puede asignar especie ni género. (***) No se obtuvieron secuencias para ese locus

De los 29 haplotipos *rbcl* obtenidos para las muestras de referencia, 12 pudieron asignarse correctamente a nivel de especie, 9 a nivel de género y 12 a nivel de familia. Esto representa un 41 % de asignación correcta a nivel de especie.

De los 27 haplotipos *matK* obtenidos para las muestras de referencia, 14 pudieron asignarse correctamente a nivel de especie, 5 a nivel de género y 8 a nivel de familia. Esto representa un 52 % de asignación correcta a nivel de especie.

El criterio utilizado para considerar que la secuencia fue correctamente asignada a la especie a la que pertenece, fue que la misma obtuviera el mayor score entre todos los candidatos. Se consideró una falla en la asignación cuando la especie correcta obtuvo un score menor o igual al de las otras especies. En los casos en que la diferencia entre los scores no era significativa se tomó en cuenta el porcentaje de identidad para reforzar el criterio.

Por ejemplo, en el caso de la especie *Aloysia triphylla*, las tres muestras mostraron diferencias de score y de porcentaje de identidad que permiten discriminarlas de otros taxones cercanos pertenecientes al mismo género. Este comportamiento se dio en el caso de los dos marcadores lo cual permite confirmar la clasificación.

Por el contrario, con los resultados obtenidos para las muestras de *Achyrocline satureioides* no es posible discriminar especie ni género. Para ambos marcadores se repite el mismo patrón

Esto concuerda con lo expuesto por varios autores con respecto al poder discriminatorio de *rbcl* con respecto a *matK*. *rbcl* es un marcador muy utilizado en estudios filogenéticos de plantas por su robustez para ubicar un espécimen no identificado en la familia, género y en algunos casos especie a la que pertenece. Su poder discriminatorio es menor especialmente en aquellos géneros ricos en especies como ocurre en angiospermas (Kress & Erickson, 2007) (CBOL Plant Working Group, 2009).

Las secuencias obtenidas para las muestras de referencia fueron utilizadas para construir un árbol filogenético basado en el algoritmo *Neighbor Joining* y el modelo de cálculo de distancias *Kimura 2 Parameter*, para cada marcador. Esta es una forma de validar el agrupamiento de las muestras en conjuntos a partir de sus secuencias *barcode*. En las figuras 14 y 15 se pueden observar los árboles obtenidos.

Las secuencias *rbcl* se agrupan correctamente por género (Figura 14). En aquellos géneros donde se analizó más de una especie, como *Achyrocline* y *Origanum*, las distintas especies son agrupadas en el mismo conjunto. Dos de las tres muestras de *Thymus vulgaris* son agrupadas junto con las muestras de *Origanum vulgare* y *Origanum majorana*.

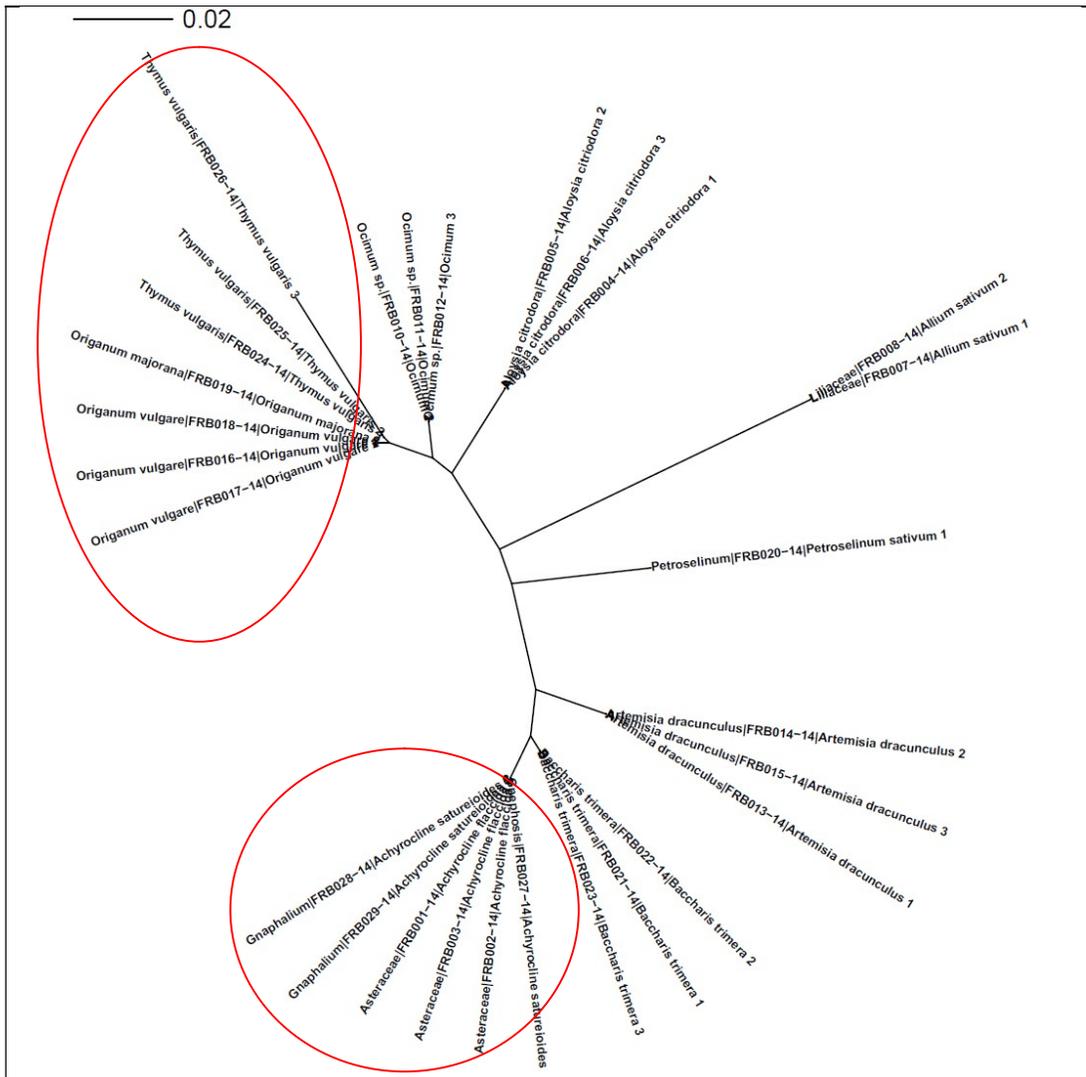


Figura 14. Árbol filogenético basado en el algoritmo *Neighbor Joining* y el modelo de cálculo de distancias *Kimura 2 Parameter* construido a partir de las secuencias *rbcl* de las muestras de referencia. Los círculos rojos señalan el agrupamiento de especies muy cercanas en el mismo *cluster*.

Las secuencias *matK* de las muestras de referencia se agrupan de la misma forma que fue descrita para las secuencias *rbcl*. El agrupamiento se realiza por género y en el caso de *Origanum* y *Thymus* los géneros se encuentran en el mismo *cluster* (Figura 15).

Estos resultados son concordantes con los presentados en las tablas 8 y 9, es decir que la asignación a la especie correcta en las muestras analizadas es relativamente baja y que la discriminación de muestras pertenecientes a especies muy cercanas no fue lograda.

Fazekas y colaboradores (2009) reportaron que la habilidad de discriminar entre especies de los *loci barcodes* es menor en plantas que en animales, lo cual estaría asociado al hecho que los límites entre especies son mas difusos en las primeras. Esto es debido a fenómenos que ocurren en plantas como la reproducción asexual, poliploidía e hibridación. En estudios realizados encontraron que las especies monofilicas bien soportadas son menos comunes en plantas y que la brecha entre distancias genéticas intra e interespecíficas es menos pronunciada que en animales. Como resultado de estos hechos, discriminar entre especies de plantas usando *barcodes* simples o *multilocus* a partir del genoma plastídico es un desafío (Fazekas, y otros, 2009).

Según Hollingsworth, para que el *DNA barcoding* funcione correctamente en plantas requiere que haya transcurrido suficiente tiempo desde la especiación para que las mutaciones o la deriva conduzcan a un grupo de caracteres que agrupen a individuos pertenecientes a la misma especie. En clados en los que la especiación ha sido muy reciente o las tasas de mutaciones son muy bajas, es muy probable que las secuencias *barcode* se compartan entre especies relacionadas (Hollingsworth, Graham, & Little, 2011).

Un ejemplo de esta situación se da en el género *Origanum*. En nuestro estudio lo incluimos por la importancia de sus especies de uso culinario, cosmético y medicinal, y con el objetivo de diferenciar entre dos especies muy cercanas: *Origanum vulgare* y *Origanum majorana*. Los resultados que obtuvimos están de acuerdo con las conclusiones de un trabajo realizado por De Mattia y colaboradores (2011). Este es un género que muestra una gran promiscuidad genética. En general, se dan eventos de hibridación entre distintas especies por lo que la diversidad genética intraespecífica es mayor que la interespecífica, lo que hace que el enfoque del *barcoding* no sea aplicable, ya que no permite discriminar entre ambas especies (De Mattia, y otros, 2011).

Thymus es otro género en el que las diferencias entre taxones muy relacionados se limitan a pocos caracteres, lo que hace que sean difíciles de asignar a una especie (Galimberti, Labra, Sandionigi, & Bruno, 2014) (Federici, y otros, 2013).

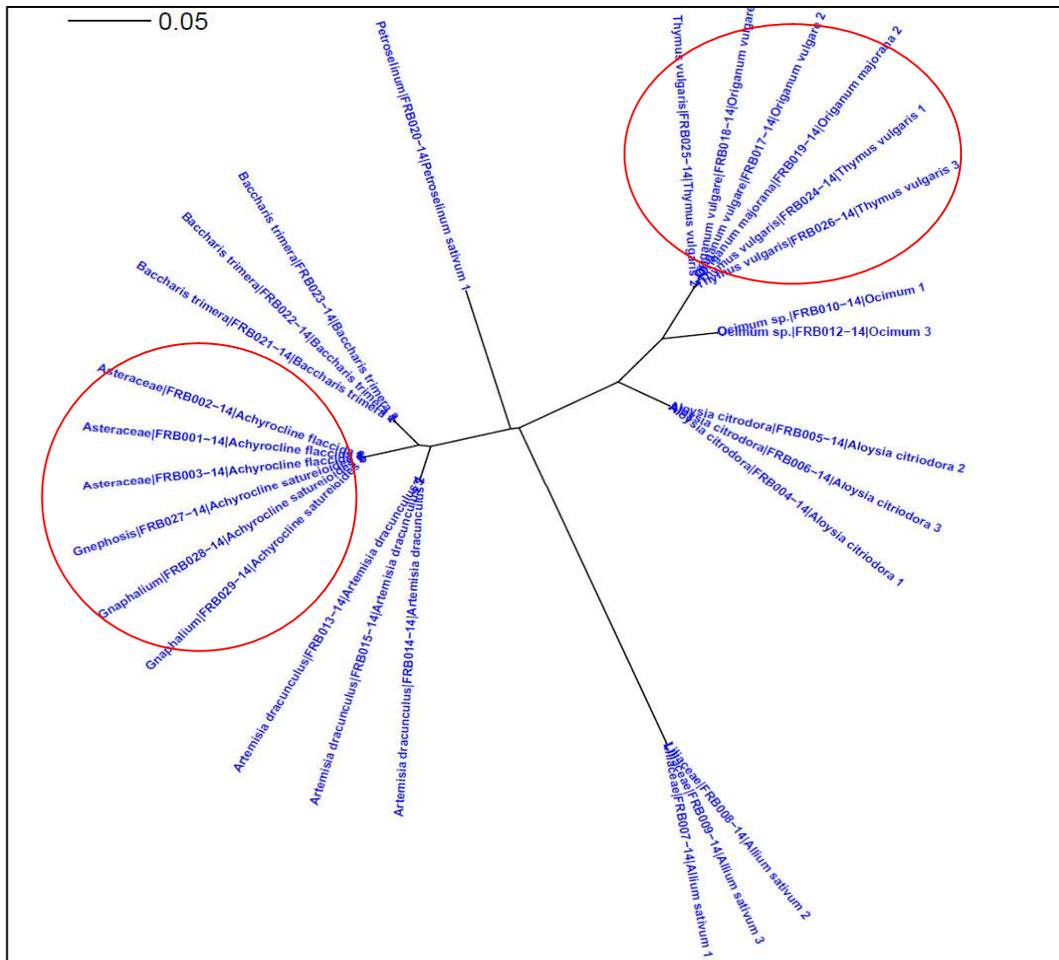


Figura 15. Árbol filogenético basado en el algoritmo *Neighbor Joining* y el modelo de cálculo de distancias *Kimura 2 Parameter* construido a partir de las secuencias *matK* de las muestras de referencia. Los círculos rojos señalan el agrupamiento de especies muy cercanas en el mismo *cluster*.

El mismo procedimiento fue realizado con las muestras industrializadas. Los resultados obtenidos se presentan en las tablas 10 y 11. Como puede observarse la muestra de ajo molido pudo asignarse correctamente a nivel de género con ambos marcadores. El ajo en escamas pudo asignarse a nivel de especie con el marcador *matK*. Estragón no pudo asignarse correctamente usando el marcador *rbcl* pero si con *matK*. En la muestra de perejil sucedió la situación a la inversa, *rbcl* fue mas eficiente para permitir la asignación que *matK*. El brocoli en cápsulas pudo asignarse correctamente a nivel de género pero no de especie. Tanto la albahaca como el orégano no pudieron diferenciarse de otras especies pertenecientes a la familia de las Lamiaceas mediante el uso de estos marcadores. En los Anexos 3 a 6 se pueden observar las secuencias obtenidas para estas muestras. El hecho de no poder diferenciarse en principio no lo atribuimos a una contaminación entre muestras ya que se tomaron todas las precauciones necesarias para que no ocurriera. Tampoco pensamos que se deba a una

adulteración ya que en las secuencias no se aprecian indicios de la presencia de una mezcla como si se vió en las correspondientes a mezclas realizadas artificialmente (datos no mostrados).

4.3.2. Clasificación usando base de datos Genbank

Además de la identificación utilizando BOLD se decidió realizar el mismo procedimiento usando la base de datos GenBank con el objetivo de observar si se lograba mejorar el porcentaje de asignaciones correctas. Como fue mencionado anteriormente, GenBank es el mayor repositorio de secuencias genéticas públicas e incluye las secuencias depositadas en BOLD. Por este motivo era de esperar que se lograra disminuir el efecto que podía tener el que no hubieran suficientes secuencias de las especies en estudio contra las cuales comparar en la base de datos.

Al igual que en BOLD, se utilizaron las secuencias consenso generadas para cada muestra y cada *locus*. Se realizaron las búsquedas de los loci por separado ya que la base de datos no permite la búsqueda empleando una combinación de ambos.

Se registró la coincidencia más cercana, el vecino más cercano perteneciente al mismo género (VMCIG) y el vecino más cercano perteneciente a un género distinto (VMCDG). En las tablas 12 y 13 pueden observarse los resultados correspondientes a los loci *rbcl* y *matK* respectivamente. En todos los casos se reportó el *score* y el porcentaje de identidad entre la secuencia problema y la secuencia blanco. En los casos en los que no está aclarado, el porcentaje de cobertura de la secuencia problema es del 100%.

En aquellos casos en que la clasificación arrojó varios taxones con el mismo *score* y porcentaje de identidad, se reportaron los más cercanos a la especie de referencia cuando fue posible.

El criterio utilizado para considerar que la secuencia fue correctamente asignada a la especie a la que pertenece, fue que la misma obtuviera el mayor *score* entre todos los candidatos. Se consideró una falla en la asignación cuando la especie correcta obtuvo un *score* menor o igual al de las otras especies. En los casos en que la diferencia entre los *scores* no era significativa se tomó en cuenta el porcentaje de identidad para reforzar el criterio.

Tabla 10. Clasificación de muestras comerciales mediante BOLD usando secuencias del locus *rbcl*

Descripción	Mejor coincidencia	Score	% Identidad	Vecino más cercano igual género	Score	% Identidad	Vecino más cercano distinto género	Score	% Identidad	Especie/Género asignado
Ajo molido	<i>Allium sativum</i>	403	98.53	<i>Allium scorodoprasum</i>	403	98.53	<i>Ocimum basilicum</i>	396	97.8	Allium
Albahaca	<i>Ocimum basilicum</i>	411	99.52	<i>Ocimum kilimadcharicum/americanum/tenuiflorum</i>	411	99.52	<i>Origanum majorana/Mentha sp.</i>	411	99.52	**
Estragon	<i>Artemisia dracunculus</i>	411	99.76	<i>Artemisia vulgaris/absinthiu/annua/campestris</i>	411	99.76	<i>Tanacetum atkinsonii/Jacobaea gibbosa/Hulteniella integrifolia</i>	411	99.76	**
Orégano	<i>Origanum vulgare</i>	381	98.96	<i>Origanum majorana/onites/sipyleum/dictamnus</i>	381	98.96	<i>Satureja vulgaris/Pycnanthemum verticillatum</i>	381	98.96	**
Perejil	<i>Petroselinum crispum</i>	417	99.63	<i>Petroselinum segetum</i>	409	98.81	<i>Amni seubertianum/trifoliatum/majus</i>	415	99.52	<i>P. crispum</i>

Tabla 11. Clasificación de muestras comerciales mediante BOLD usando secuencias del locus *matK*

Descripción	Mejor coincidencia	Score	% Identidad	Vecino más cercano igual género	Score	% Identidad	Vecino más cercano distinto género	Score	% Identidad	Especie/Género asignado
Ajo en escamas	<i>Allium sativum</i>	345	99.43	<i>Allium vineale/A.scorodoprasum/A.ampeloprasum</i>	341	98.85	<i>Mesua ferrea</i>	317	95.66	A.sativum
Ajo molido	<i>Allium sativum</i>	280	99.64	<i>Allium scorodoprasum</i>	278	99.29	<i>Mesua ferrea</i>	260	96.09	A.sativum
Albahaca							<i>Amaranthus</i>	243	98.79	**
Broccoli	<i>Brassica oleracea</i>	298	100	<i>Brassica napus/B.rapa</i>	298	100	<i>Raphanus raphanistrum</i>	294	99.33	Brassica
Estragon	<i>Artemisia dracunculus</i>	329	99.4	<i>Artemisia annua/A.kruhsiana</i>	323	98.49	<i>Hulteniella integrifolia</i>	325	98.8	A.dracunculus
Orégano	<i>Origanum vulgare</i>	244	98.39	<i>O.dictamnus/O.onites/O.dayi/O.elongatum/O.rotundifolium</i>	244	98.39	<i>Thymus vulgaris</i>	244	98.39	**
Perejil	<i>Petroselinum crispum</i>	266	99.63	<i>Petroselinum segetum</i>	266	99.63	<i>Apium graveolens</i>	264	99.25	Petroselinum

En las tablas se presenta la especie a la que pertenece la muestra y el resultado arrojado por la base de datos BOLD para la clasificación con cada marcador. Se describen la especie con mejor nivel de coincidencia, el vecino más cercano perteneciente al mismo género y el vecino más cercano correspondiente a otro género. De acuerdo al score y % de identidad se define la especie asignada. En algunos casos la clasificación se realiza a nivel de género (mismo score y % de identidad) y en otros no se puede clasificar. (**) No se puede asignar especie ni género .

Tabla 12. Identificación de muestras de referencia en GenBank usando secuencias del locus rbcL.

Especie	Muestra	Mejor coincidencia	Score	% Identidad	Vecino más cercano igual género	Score	% Identidad	Vecino más cercano distinto género	Score	% Identidad	Especie/ Género asignado
<i>Allium sativum</i>	1	<i>Allium sativum</i>	939	99	<i>Allium cepa</i>	944	100	<i>Nothoscordum inutile</i>	900	98	Allium
	2	<i>Allium sativum</i>	937	99	<i>Allium cepa</i>	942	100	<i>Nothoscordum inutile</i>	898	98	Allium
	3	<i>Allium sativum</i>	931	99	<i>Allium cepa</i>	937	100				Allium
<i>Ocimum sp.</i>	1	<i>Ocimum basilicum</i>	926	100	<i>Ocimum tenuiflorum</i>	926	100	<i>Origanum majorana</i>	926	100	**
	2	<i>Ocimum basilicum</i>	894	100	<i>Ocimum tenuiflorum</i>	894	100	<i>Origanum majorana</i>	894	100	**
	3	<i>Ocimum basilicum</i>	780	99	<i>O.tenuiflorum/ O. americanum/ O. kilimandscharicum/O.sp</i>	780	99	<i>Origanum majorana/ Mentha sp.</i>	780	99	**
<i>Baccharis trimera</i>	1	<i>Baccharis genistelloides</i>	950	99	<i>Baccharis tricuneata</i>	944	99	<i>Diplostephiumserratifolium/D.spinolosu m/D.azureum/Exostigma notobellidiastrum</i>	950	99	**
	2	<i>Baccharis genistelloides</i>	953	99	<i>Baccharis tricuneata</i>	948	99	<i>Diplostephiumserratifolium/D.spinolosu m/D.azureum/Exostigma notobellidiastrum /Archibaccharis asperifolia</i>	953	99	**
	3	<i>Baccharis genistelloides</i>	948	99	<i>Baccharis tricuneata</i>	942	99	<i>Diplostephiumserratifolium/D.spinolosu m/D.azureum/Exostigma notobellidiastrum /Archibaccharis asperifolia</i>	948	99	**
<i>Aloysia triphylla</i>	1	<i>Aloysia citrodora</i>	926	100				<i>Populus fremontii/ Jacaranda puberula</i>	881	98	<i>A. citrodora</i>
	2	<i>Aloysia citrodora</i>	922	100				<i>Populus fremontii/ Jacaranda puberula</i>	878	98	<i>A. citrodora</i>
	3	<i>Aloysia citrodora</i>	913	100	<i>Lippia oxyphyllaria</i>	850 cobertura 99%	98	<i>Populus fremontii/ Jacaranda puberula</i>	869	98	<i>A. citrodora</i>
<i>Artemisia dracunculus L.</i>	1	<i>Artemisia dracunculus</i>	902 cobertura 97%	99	<i>A.annua/A.capillaris/A.gmelinii/A.vulgaris/A.frigida/A.roxburghiana</i>	924	99	<i>Ajania gracilis /Chrysanthemum indicum</i>	918	99	Artemisia
	2	<i>Artemisia dracunculus</i>	894 cobertura 98%	99	<i>A.annua/A.capillaris/A.gmelinii/A.vulgaris/A.frigida/A.roxburghiana/A.sieversana/A.scoparia</i>	905	99	<i>Ajania gracilis /Chrysanthemum indicum/C. lavandulifolium</i>	900	99	Artemisia
	3	<i>Artemisia dracunculus</i>	909 cobertura 98%	100	<i>A. annua/ A. roxburghiana/ A. annua/ A. frigida/ A. sieversiana/ A. capillaris/ A. scoparia/A. gmelinii/A.vulgaris/A.frigida</i>	922	100	<i>Ajania gracilis/Chrysanthemum indicum/C. lavandulifolium</i>	917	99	Artemisia

Especie	Muestra	Mejor coincidencia	Score	% Identidad	Vecino más cercano igual género	Score	% Identidad	Vecino más cercano distinto género	Score	% Identidad	Especie/ Género asignado
<i>Achyrocline satureioides</i>	1	<i>Gnaphalium sp.</i>	769	99	<i>Gnaphalium supinum/G. sylvaticum</i>	769	99	<i>Craspedia incana/Metalasia densa/Helychrisum cymosum /H.glumaceum</i>	774	99	Fallo en la asignación
	2	<i>Gnaphalium sp.</i>	795	99	<i>Gnaphalium supinum/G. sylvaticum</i>	795	99	<i>Metalasia densa/Helychrisum cymosum/H.glumaceum</i>	800	99	Fallo en la asignación
	3	<i>Gnaphalium sp.</i>	802	99	<i>Gnaphalium supinum/G. sylvaticum</i>	802	99	<i>Metalasia densa/Helychrisum cymosum/H.glumaceum</i>	808	99	Fallo en la asignación
<i>Achyrocline flaccida</i>	1				<i>Gnaphalium affine</i>	909	99	<i>Antennaria neglecta/A. howellii/A.parvifolium/Anaphilis margaritacea/A. sinica /Lentopodium leiolepis /Helychrisum stoechas /Gamochaeta pensylvanica</i>	909	99	**
	2				<i>Gnaphalium affine</i>	905 cobertura 99%	99	<i>Antennaria neglecta/A. howellii/A.parvifolium/Anaphilis margaritacea/A. sinica /Lentopodium leiolepis /Helychrisum stoechas /Gamochaeta pensylvanica</i>	907	99	Fallo en la asignación
	3				<i>Gnaphalium affine</i>	913 cobertura 99%	99	<i>Antennaria neglecta/A. howellii/A.parvifolium/Anaphilis margaritacea/A. sinica /Lentopodium leiolepis /Helychrisum stoechas /Gamochaeta pensylvanica</i>	915	99	Fallo en la asignación
<i>Origanum majorana</i>				<i>Origanum vulgare</i>	959	100	<i>Mentha longifolia /Bystropogon origanifolius</i>	953	99	Origanum	
<i>Origanum vulgare</i>	1	<i>Origanum vulgare</i>	920	100	<i>Origanum saccatum/O.minutiflorum</i>	920	100	<i>Mentha longifolia /M.suaveolens/M. spicata/M. canadensis/Bystropogon origanifolius/Clinopodium chinense</i>	915	99	Origanum
	2	<i>Origanum vulgare</i>	920	100	<i>Origanum saccatum/O.minutiflorum</i>	920	100	<i>Mentha longifolia/M.spicata/ M. suaveolens/ M. canadensis /M. pulegium/ Clinopodium vulgare/ Clinopodium chinense /Bystropogon origanifolius</i>	915	99	Origanum
	3	<i>Origanum vulgare</i>	941	100	<i>Origanum saccatum/O.minutiflorum</i>	933 cobertura 99%	100	<i>Mentha longifolia/ M. spicata/ Bystropogon origanifolius</i>	935	99	<i>O. vulgare</i>
<i>Petroselinum crispum</i>		<i>Petroselinum crispum</i>	924	99	<i>Petroselinum segetum</i>	887 cobertura 97%	99	<i>Ammi majus</i>	913	99	<i>P. crispum</i>

Especie	Muestra	Mejor coincidencia	Score	% Identidad	Vecino más cercano igual género	Score	% Identidad	Vecino más cercano distinto género	Score	% Identidad	Especie/ Género asignado
<i>Thymus vulgaris</i>	1	<i>Thymus vulgaris</i>	876	100	<i>Thymus sibthorpii/T. praecox/T.pulegioides/T. alsinoides</i>	876	100	<i>Mentha longifolia/M.spicata/ M. suaveolens/ M. canadensis/ M. pulegium/ M. aquatica /Clinopodium vulgare/C. acinos/ C. chinense/ Bystropogon origanifolius</i>	870	99	Thymus
	2	<i>Thymus vulgaris</i>	963	100	<i>Thymus alsinoides</i>	963	100	<i>Mentha longifolia</i>	957	99	Thymus
	3	<i>Thymus vulgaris</i>	933	100	<i>Thymus alsinoides</i>	933	100	<i>Bystropogon origanifolius /Mentha longifolia</i>	928	99	Thymus

En la tabla se presentan los datos obtenidos de la comparación de las secuencias correspondientes a las muestras de referencia con la base de datos BOLD. Se describe la coincidencia mas cercana, el/los vecino/s mas cercano de igual género y el/los vecino/s mas cercanos pertenecientes a otros género clasificados según su score y % de identidad con la secuencia problema. Por último se informa la especie/género asignado de acuerdo a los datos obtenidos. (*) No se encontraron vecinos mas cercanos de otro género en la base de datos. (**) De acuerdo a los datos no se puede asignar especie ni género. (***) No se obtuvieron secuencias para ese locus

Tabla 13. Identificación de muestras de referencia en GenBank usando secuencias del locus *matK*

Especie	Muestra	Mejor coincidencia	Score	% Identidad	Vecino más cercano igual género	Score	% Identidad	Vecino más cercano distinto género	Score	% Identidad	Especie/ Género asignado
<i>Allium sativum</i>	1	<i>Allium sativum</i>	1083	99	<i>A. scorodoprasum/A. vineale/A. ampelo prasum</i>	1050	99				<i>A. sativum</i>
	2	<i>Allium sativum</i>	1192	100	<i>A. scorodoprasum/A. vineale/A. ampelo prasum</i>	1158	99				<i>A. sativum</i>
	3	<i>Allium sativum</i>	1182	100	<i>A. scorodoprasum/A. vineale/A. ampelo prasum</i>	1149	99				<i>A. sativum</i>
<i>Ocimum sp.</i>	1	<i>Ocimum basilicum</i>	780	99	<i>O. tenuiflorum/O. americanum/O. kilimandscharicum/O.sp</i>	780	99	<i>Mentha sp./Origanum majorana</i>	780	99	**
	2	*	*	*	*	*	*	*	*	*	*
	3	<i>Ocimum basilicum</i>	1101	99	<i>Ocimum x africanum /O.americanum</i>	1101	99	<i>Basilicum polystachium</i>	1064	99	Ocimum
<i>Baccharis trimera</i>	1	<i>Baccharis genistelloides</i>	1356	99	<i>Baccharis tricuneata</i>	1351	99	<i>Diplostephium coriaceum/D.frontinense/D.serratifoliu/D.romeroi/D.colombianum/Laennecia sopherifolia/L.coultieri</i>	1345	99	<i>B. genistelloides</i>
	2	<i>Baccharis genistelloides</i>	1312	99	<i>Baccharis tricuneata</i>	1306	99	<i>Diplostephium coriaceum/D.inesianum/D.frontinense/D.serratifolium/D.romeroi/D.colombianum/Laennecia sopherifolia/L. coultieri/Doellingeria umbellata/Oclemena acuminata</i>	1301	99	<i>B. genistelloides</i>
	3	<i>Baccharis tricuneata</i>	1297	99	<i>Baccharis genistelloides</i>	1291	99	<i>Diplostephium coriaceum /D.inesianum /D.frontinense/ D.serratifolium /D.romeroi/D.colombianum/Laennecia sopherifolia/L. coultieri/ Doellingeria umbellata/Oclemena acuminata</i>	1291	99	<i>B. tricuneata</i>
<i>Aloysia triphylla</i>	1	<i>Aloysia citrodora</i>	1098	100	<i>Lippia javanica /L. oxyphyllaria</i>	994 cobertura 99%	97	<i>Lantana camara</i>	1011 cobertura 99%	97	<i>A. citrodora</i>
	2	<i>Aloysia citrodora</i>	1046	100	<i>Lippia javanica /L. oxyphyllaria</i>	950 cobertura 99%	97	<i>Lantana camara</i>	966 cobertura 99%	98	<i>A. citrodora</i>
	3	<i>Aloysia citrodora</i>	1072	100	<i>Lippia javanica /L. oxyphyllaria</i>	976 cobertura 99%	97	<i>Lantana camara</i>	992 cobertura 99%	98	<i>A. citrodora</i>
<i>Artemisia dracunculus L.</i>	1	<i>Artemisia dracunculus</i>	1099	100	<i>A. borealis/A. hyperborea/A. afra/A. siebersiana/A. campestris/A. japonica</i>	1077	99	<i>Ajania gracilis</i>	1072	99	<i>A. dracunculus</i>
	2	<i>Artemisia dracunculus</i>	1170	100	<i>Artemisia borealis/A. frigida/A. afra/A. siebersiana /A. campestris/A. japonica</i>	1147	99	<i>Ajania gracilis/Chrysanthemum indicum</i>	1142	99	<i>A. dracunculus</i>

Especie	Muestra	Mejor coincidencia	Score	% Identidad	Vecino más cercano igual género	Score	% Identidad	Vecino más cercano distinto género	Score	% Identidad	Especie/ Género asignado
	3	<i>Artemisia dracunculus</i>	1088	99	<i>Artemisia borealis/A. hyperborea/A.frigida/A.afra/A.sieversiana/A.campestris/A.japonica</i>	1066	99	<i>Ajania gracilis/Chrysanthemum indicum</i>	1061	99	A. <i>dracunculus</i>
<i>Achyrocline satureioides</i>	1							<i>Metalsia densa/Helychrisum petiolare/H.scabrum/H.pandurifolium</i>	710	100	Fallo en la asignación
	2							<i>Metalsia densa/Helychrisum petiolare/H.scabrum</i>	917	100	Fallo en la asignación
	3							<i>Metalsia densa/Helychrisum petiolare</i>	863	100	Fallo en la asignación
<i>Achyrocline flaccida</i>	1				<i>Gnaphalium supinum</i>	966	98	<i>Metalsia densa/Helychrisum petiolare/H.scabrum</i>	1016	100	Fallo en la asignación
	2				<i>Gnaphalium supinum</i>	968	98	<i>Metalsia densa/Helychrisum petiolare/H.scabrum</i>	1018	100	Fallo en la asignación
	3							<i>Metalsia densa/Helychrisum petiolare/H.scabrum</i>	924	100	Fallo en la asignación
<i>Origanum majorana</i>		<i>Origanum majorana</i>	824	99	<i>Origanum vulgare</i>	828	99	<i>Thymus caespititius</i>	824	99	Origanum
<i>Origanum vulgare</i>	1	*	*	*	*		*	*	*	*	*
	2	<i>Origanum vulgare</i>	933	99	<i>Origanum onites/O.microphyllum/O. dayi/O. elongatum/O. dictamnus/O. rotundifolium</i>	933	99	<i>Thymus caespititius</i>	929	99	Origanum
	3	<i>Origanum vulgare</i>	944	99	<i>Origanum onites/O.microphyllum/O. dayi/O. elongatum/O. dictamnus/O. rotundifolium</i>	944	99	<i>Thymus caespititius</i>	941	98	Origanum
<i>Petroselinum crispum</i>		<i>Petroselinum crispum</i>	1016	100	<i>Petroselinum segetum</i>	983 cobertura 96%	100	<i>Apium graveolens</i>	1000	99	<i>P. crispum</i>
<i>Thymus vulgaris</i>	1	<i>Thymus vulgaris</i>	990	99	<i>Thymus striatus/T. praecox</i>	1002	99	<i>Origanum vulgare</i>	981 cobertura 99%	99	Thymus
	2	<i>Thymus vulgaris</i>	974	99	<i>Thymus pulegioides</i>	974	99	<i>Origanum vulgare</i>	957	99	Thymus
	3	<i>Thymus vulgaris</i>	817	99	<i>Thymus serpyllum/T. pulegioides/T.praecox/T.sibthorpii</i>	822	99	<i>Origanum vulgare</i>	811	99	Thymus

En la tabla se presentan los datos obtenidos de la comparación de las secuencias correspondientes a las muestras de referencia con la base de datos BOLD. Se describe la coincidencia mas cercana, el/los vecino/s mas cercano de igual género y el/los vecino/s mas cercanos pertenecientes a otros género clasificados según su score y % de identidad con la secuencia problema. Por último se informa la especie/género asignado de acuerdo a los datos obtenidos. (*) No se encontraron vecinos mas cercanos de otro género en la base de datos. (**) De acuerdo a los datos no se puede asignar especie ni género. (***) No se obtuvieron secuencias para ese locus

A diferencia de lo que ocurrió con la clasificación mediante BOLD, con el uso de la base de datos GenBank el porcentaje de asignaciones correctas a nivel de especie fue significativamente más bajo.

De los 29 haplotipos *rbcl* obtenidos solamente 5 pudieron asignarse correctamente a nivel de especie, 12 a nivel de género y 10 a nivel de familia.

Contrariamente a lo que observamos en la base de datos BOLD, *matK* no nos permitió profundizar en el nivel de identificación.

De los 27 haplotipos *matk* obtenidos, 10 pudieron asignarse correctamente a nivel de especie, 7 a nivel de género y 10 a nivel de familia.

A pesar que GenBank es una base de datos que contiene mayor volumen de secuencias, solo contiene los registros de barcodes que se han hecho públicos. De hecho para alguna de las especies en estudio, existen muy pocas o ninguna secuencia correspondientes a *rbcl* o *matK* que las representen en GenBank. Este es el caso de *Baccharis trimera*, *Aloysia triphylla*, *Achyrocline satureioides*, *Achyrocline flaccida*.

Como varios autores lo han resaltado, la existencia de bases de datos de secuencias *barcode* de referencia confiables y accesibles son un elemento crítico para el éxito de la metodología *DNA barcoding* en la asignación correcta de una muestra desconocida a la especie correcta. En el caso de plantas las bases de datos contienen menor representación de especies que para animales (aproximadamente 65000 especies de plantas frente a 178000 de animales). Este hecho lo hemos podido corroborar en este trabajo. Las especies en estudio tenían muy baja representación en las bases de datos públicas, lo que coincide con lo expuesto por Galimberti y colaboradores (2014) que sostiene que especies de plantas aromáticas y medicinales, denominadas cultivos pequeños, así como especies locales presentan pocas o ninguna secuencia *barcode* en las bases mencionadas (Galimberti, Labra, Sandionigi, & Bruno, 2014).

4.4. ANÁLISIS DE DATOS MEDIANTE UN PROCEDIMIENTO ESTADÍSTICO DE CLASIFICACIÓN (IMPLEMENTADO MEDIANTE PLATAFORMA WEKA)

Algunos autores han propuesto el uso métodos puramente estadísticos para resolver las dificultades que se presentan en la aplicación de la estrategia de *DNA barcoding* en algunos casos problemáticos como en especies que han divergido recientemente (van Velzen, Weitschek, Felici, & Bakker, 2012) (Weitschek, Fiscon, & Felici, 2014) (Austerlitz, y otros, 2009).

En este trabajo evaluamos el desempeño de un método estadístico de clasificación para clasificar nuestras muestras vegetales en grupos de referencia, y valorar la posibilidad de utilizarlo como alternativa a los métodos filogenéticos utilizados en el enfoque propuesto por la metodología *DNA barcoding* a través de su plataforma BOLD.

Utilizamos el método de clasificación estadística no paramétrico k- Nearest Neighbour (k-NN) implementado en la plataforma WEKA. El mismo permite clasificar una instancia problema en la clase correspondiente a la más cercana, es decir que tiene menor distancia estimada en el espacio definido por las variables utilizadas, en este caso posiciones en la secuencia de ADN en las que se comparan ambas instancias. El valor de k es seleccionado por el usuario y depende de los datos, generalmente valores altos reducen el efecto del ruido en la clasificación pero crean límites entre clases parecidas. En este caso seleccionamos un valor de k igual a 1 ya que se ha reportado para este tipo de estudios que brinda los mejores resultados (Austerlitz, y otros, 2009).

Con las secuencias *barcode* obtenidas para las muestras de referencia y las industrializadas comerciales según se describió en apartados anteriores, se hicieron análisis utilizando distintos agrupamientos de las mismas en virtud de la información que queríamos extraer del análisis.

En primera instancia se corrieron las muestras de referencia separadas por marcador *rbcl* y *matK*.

Con las secuencias *barcode* correspondientes a *rbcl* se obtuvo un 100% de instancias correctamente clasificadas con un Kappa igual a 1. Kappa es un coeficiente estadístico que permite estimar si la clasificación obtenida puede ser producto del azar ($k \leq 0$) así como el grado de robustez del clasificador utilizado (mayor cuanto más cercano a 1) (Cohen, 1960). En el caso de *matK* se obtuvo un 92,6% de instancias correctamente clasificadas con un Kappa de 0,92. Un resumen de los datos obtenidos y la matriz de confusión para ambos casos se presenta en las figuras 16 y 17. De esta forma validamos el modelo utilizado, ya que corroboramos que clasifica correctamente las muestras de referencia.

Luego probamos clasificar las muestras comerciales industrializadas en especies conocidas utilizando las muestras de referencia. Para eso agrupamos las secuencias *barcode* correspondientes a ambos tipos de muestras por marcador en un único archivo y procedimos a realizar la clasificación en la plataforma WEKA utilizando el modelo previamente descrito. Para *rbcl* se obtuvo un 89,2% de acierto en la clasificación con un Kappa de 0,88. Como se puede observar en la matriz de confusión (figura 18), las muestras que fueron clasificadas incorrectamente fueron las de origen comercial. Utilizando las secuencias *barcode* correspondientes a *matK* las instancias correctamente asignadas fueron un 80,6% con un Kappa igual a 0,79. Al igual que con

las secuencias *rbcl*, las muestras comerciales no fueron correctamente clasificadas (figura 19).

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances   58      100 %
Incorrectly Classified Instances  0        0 %
Kappa statistic                  1

=== Confusion Matrix ===

 a b c d e f g h i j k <-- classified as
6 0 0 0 0 0 0 0 0 0 0 | a = Achyrocline_flaccida
0 6 0 0 0 0 0 0 0 0 0 | b = Achyrocline_satureioides
0 0 6 0 0 0 0 0 0 0 0 | c = Allium_sativum
0 0 0 6 0 0 0 0 0 0 0 | d = Aloysia_citriodora
0 0 0 0 6 0 0 0 0 0 0 | e = Artemisia_dracunculus
0 0 0 0 0 6 0 0 0 0 0 | f = Baccharis_trimera
0 0 0 0 0 0 6 0 0 0 0 | g = Ocimum
0 0 0 0 0 0 0 2 0 0 0 | h = Origanum_majorana
0 0 0 0 0 0 0 0 6 0 0 | i = Origanum_vulgare
0 0 0 0 0 0 0 0 0 2 0 | j = Petroselinum_sativum
0 0 0 0 0 0 0 0 0 0 6 | k = Thymus_vulgaris

```

Figura 16. Resultados obtenidos en WEKA de clasificación de muestras de referencia usando secuencias barcode del locus *rbcl*. Se reportan el porcentaje de instancias correctamente clasificadas, el de clasificadas incorrectamente y el coeficiente Kappa. Los números en la matriz de confusión corresponden al número de muestras.

```

=== Stratified cross-validation ===
Correctly Classified Instances   50      92.5926 %
Incorrectly Classified Instances  4       7.4074 %
Kappa statistic                  0.9177

=== Confusion Matrix ===

 a b c d e f g h i j k <-- classified as
4 0 0 0 2 0 0 0 0 0 0 | a = Achyrocline_flaccida
0 6 0 0 0 0 0 0 0 0 0 | b = Achyrocline_satureioides
0 0 6 0 0 0 0 0 0 0 0 | c = Allium_sativum
0 0 0 6 0 0 0 0 0 0 0 | d = Aloysia_citriodora
2 0 0 0 4 0 0 0 0 0 0 | e = Artemisia_dracunculus
0 0 0 0 0 6 0 0 0 0 0 | f = Baccharis_trimera
0 0 0 0 0 0 4 0 0 0 0 | g = Ocimum
0 0 0 0 0 0 0 2 0 0 0 | h = Origanum_majorana
0 0 0 0 0 0 0 0 4 0 0 | i = Origanum_vulgare
0 0 0 0 0 0 0 0 0 2 0 | j = Petroselinum_sativum
0 0 0 0 0 0 0 0 0 0 6 | k = Thymus_vulgaris

```

Figura 17. Resultados obtenidos en WEKA de clasificación de muestras de referencia usando secuencias barcode del locus *matK*. Se reportan el porcentaje de instancias correctamente clasificadas, el de clasificadas incorrectamente y el coeficiente Kappa. Los números en la matriz de confusión corresponden al número de muestras. Los números en verde indican las instancias mal clasificadas.

```

=== Stratified cross-validation ===
Correctly Classified Instances   58      89.2308 %
Incorrectly Classified Instances  7      10.7692 %
Kappa statistic                  0.8823
=== Confusion Matrix ===

 a b c d e f g h i j k l m n o p q <-- classified as
6 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | a = Achyrocline_flaccida
0 6 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | b = Achyrocline_satureioides
0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 | c = Ajo_molido
0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 | d = Albahaca
0 0 0 0 0 6 0 0 0 0 0 0 0 0 0 0 0 | e = Allium_sativum
0 0 0 0 0 0 6 0 0 0 0 0 0 0 0 0 0 | f = Aloysia_citriodora
0 0 0 0 0 0 0 6 0 0 0 0 0 0 0 0 0 | g = Artemisia_dracunculus
0 0 0 0 0 0 0 0 6 0 0 0 0 0 0 0 0 | h = Baccharis_trimera
0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 | i = Estragon
0 0 0 0 0 0 0 0 0 6 0 0 0 0 0 0 0 | j = Ocimum
0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 | k = Oregano
0 0 0 0 0 0 0 0 0 0 2 0 0 0 0 0 0 | l = Origanum_majorana
0 0 0 0 0 0 0 0 0 0 0 6 0 0 0 0 0 | m = Origanum_vulgare
0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 | n = Perejil
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 2 0 0 | o = Petroselinum_sativum
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 6 0 | p = Thymus_vulgaris
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 | q = Tomillo

```

Figura 18. Resultados obtenidos en WEKA de clasificación de muestras de referencia y comerciales usando secuencias barcode del locus *rbcl*. Se reporta el porcentaje de instancias correctamente clasificadas, el de clasificadas incorrectamente y el coeficiente Kappa. Los números en la matriz de confusión corresponden al número de muestras. Los números en verde indican las instancias mal clasificadas.

```

=== Stratified cross-validation ===
Correctly Classified Instances   50      80.6452 %
Incorrectly Classified Instances  12     19.3548 %
Kappa statistic                  0.7893
=== Confusion Matrix ===

 a b c d e f g h i j k l m n o p q r <-- classified as
4 0 0 0 0 0 0 2 0 0 0 0 0 0 0 0 0 | a = Achyrocline_flaccida
0 6 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | b = Achyrocline_satureioides
0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 | c = Ajo_escamas
0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 | d = Ajo_molido
0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 | e = Albahaca
0 0 0 0 0 6 0 0 0 0 0 0 0 0 0 0 0 | f = Allium_sativum
0 0 0 0 0 0 6 0 0 0 0 0 0 0 0 0 0 | g = Aloysia_citriodora
2 0 0 0 0 0 0 4 0 0 0 0 0 0 0 0 0 | h = Artemisia_dracunculus
0 0 0 0 0 0 0 0 6 0 0 0 0 0 0 0 0 | i = Baccharis_trimera
0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 | j = Broccoli
0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 | k = Estragon
0 0 0 0 0 0 0 0 0 0 0 4 0 0 0 0 0 | l = Ocimum
0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 | m = Oregano
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 2 0 0 | n = Origanum_majorana

```

Figura 19. Resultados obtenidos en WEKA de clasificación de muestras de referencia y comerciales usando secuencias barcode del locus *matK*. Se reporta el porcentaje de instancias correctamente clasificadas, el de clasificadas incorrectamente y el coeficiente Kappa. Los números en la matriz de confusión corresponden al número de muestras. Los números en verde indican las instancias mal clasificadas.

Luego comenzamos a realizar pruebas agrupando muestras de acuerdo a criterios que nos parecían interesantes de analizar. Por ejemplo agrupamos las muestras de referencia que fueran aromáticas y con propiedades medicinales (*Achyrocline flaccida*, *Achyrocline satureioides*, *Aloysia citriodora*, *Artemisia dracunculus*, *Baccharis trimmera*, *Ocimum* sp. y *Thymus vulgaris*). La clasificación utilizando las secuencias barcode *rbcl* dió como resultado un 100% de acierto en la clasificación con un Kappa de 1. Con las secuencias barcode *matK*, se dió una situación similar a lo observado usando el total de las muestras de referencia. Se obtuvo un 90% de instancias correctamente clasificadas con un Kappa igual a 0,88). En ambos casos se observa un error en la clasificación entre *Achyrocline flaccida* y *Artemisia dracunculus* (Figuras 20 y 21).

Si agrupamos las muestras de referencia aromáticas y con propiedades medicinales e incluimos la muestra comercial de estragón, analizando las secuencias barcode *matK* lo que se observa es que es clasificada como *Achyrocline satureioides*. Sin embargo dentro de las muestras de referencia dos muestras de *Artemisia dracunculus* con clasificadas como *Achyrocline flaccida*. Utilizando *rbcl* las instancias correctamente clasificadas son del 100% con un Kappa de 1 (Figuras 22 y 23).

Posteriormente hicimos una prueba con *Origanum vulgare*, *Origanum majorana* y la muestra de orégano comercial. El objetivo fue ver que sucede con la clasificación si tenemos una muestra problema que suponemos a qué especie pertenece y la comparamos con muestras de referencia correspondientes a esa especie. En este caso comparamos contra dos especies que son muy similares y que muchas veces se cofunden cuando la muestra está industrializada y pierde sus características macroscópicas que permiten identificarla en forma visual.

Utilizando las secuencias barcode *rbcl* se obtuvo un 88,9% de aciertos en la clasificación con un Kappa igual a 0,74. Como puede observarse en la figura 24, la clasificación es correcta ya que la muestra comercial la asigna a la etiqueta *Origanum vulgare*. Pero como no la asigna a su propia etiqueta el porcentaje de asignación no es del 100. Utilizando las secuencias *matK* los resultados son coincidentes (figura 25). Con estos resultados podemos decir que la muestra comercial se clasifica correctamente, lo que no pudo lograrse usando las herramientas de la plataforma BOLD.

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances   42    100 %
Incorrectly Classified Instances  0     0 %
Kappa statistic                  1

=== Confusion Matrix ===

 a b c d e f g <-- classified as
6 0 0 0 0 0 0 | a = Achyrocline_flaccida
0 6 0 0 0 0 0 | b = Aloysia_citriodora
0 0 6 0 0 0 0 | c = Ocimum
0 0 0 6 0 0 0 | d = Artemisia_dracunculus
0 0 0 0 6 0 0 | e = Baccharis_trimera
0 0 0 0 0 6 0 | f = Thymus_vulgaris
0 0 0 0 0 0 6 | g = Achyrocline_satureioides

```

Figura 20. Resultados obtenidos en WEKA de clasificación de muestras de referencia agrupadas como aromáticas medicinales usando secuencias *barcode* del locus *rbCL*. Se reporta el porcentaje de instancias correctamente clasificadas, el de clasificadas incorrectamente y el coeficiente Kappa. Los números en la matriz de confusión corresponden al número de muestras. El círculo rojo muestra las muestras asignadas correctamente

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances   36    90 %
Incorrectly Classified Instances  4    10 %
Kappa statistic                  0.883

=== Confusion Matrix ===

 a b c d e f g <-- classified as
4 0 0 2 0 0 0 | a = Achyrocline_flaccida
0 6 0 0 0 0 0 | b = Achyrocline_satureioides
0 0 6 0 0 0 0 | c = Aloysia_citriodora
2 0 0 4 0 0 0 | d = Artemisia_dracunculus
0 0 0 0 6 0 0 | e = Baccharis_trimera
0 0 0 0 0 4 0 | f = Ocimum
0 0 0 0 0 0 6 | g = Thymus_vulgaris

```

Figura 21. Resultados obtenidos en WEKA de clasificación de muestras de referencia agrupadas como aromáticas medicinales usando secuencias *barcode* del locus *matK*. Se reporta el porcentaje de instancias correctamente clasificadas, el de clasificadas incorrectamente y el coeficiente Kappa. Los números en la matriz de confusión corresponden al número de muestras. El círculo rojo muestra las muestras asignadas correctamente. Los números en verde indican las instancias mal clasificadas.

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances   37      88.0952 %
Incorrectly Classified Instances  5      11.9048 %
Kappa statistic                 0.8618

=== Confusion Matrix ===

 a b c d e f g h <-- classified as
5 0 0 1 0 0 0 0 | a = Achyrocline_flaccida
0 6 0 0 0 0 0 0 | b = Achyrocline_satureioides
0 0 6 0 0 0 0 0 | c = Aloysia_citriodora
2 0 0 4 0 0 0 0 | d = Artemisia_dracunculus
0 0 0 0 6 0 0 0 | e = Baccharis_trimera
0 2 0 0 0 0 0 0 | f = Estragon
0 0 0 0 0 4 0 0 | g = Ocimum
0 0 0 0 0 0 6 0 | h = Thymus_vulgaris

```

Figura 22. Resultados obtenidos en WEKA de clasificación de muestras de referencia y comerciales agrupadas como aromáticas medicinales usando secuencias *barcode* del locus *matK*. Se reporta el porcentaje de instancias correctamente clasificadas, el de clasificadas incorrectamente y el coeficiente Kappa. Los números en la matriz de confusión corresponden al número de muestras. El círculo rojo muestra las muestras asignadas correctamente. Los números en verde indican las instancias mal clasificadas.

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances   44      100 %
Incorrectly Classified Instances  0       0 %
Kappa statistic                 1

=== Confusion Matrix ===

 a b c d e f g h <-- classified as
6 0 0 0 0 0 0 0 | a = Achyrocline_flaccida
0 6 0 0 0 0 0 0 | b = Aloysia_citriodora
0 0 6 0 0 0 0 0 | c = Ocimum
0 0 0 6 0 0 0 0 | d = Artemisia_dracunculus
0 0 0 0 6 0 0 0 | e = Baccharis_trimera
0 0 0 0 0 6 0 0 | f = Thymus_vulgaris
0 0 0 0 0 0 6 0 | g = Achyrocline_satureioides
0 0 0 0 0 0 0 2 | h = Estragon

```

Figura 23. Resultados obtenidos en WEKA de clasificación de muestras de referencia y comerciales agrupadas como aromáticas medicinales usando secuencias *barcode* del locus *rbcl*. Se reporta el porcentaje de instancias correctamente clasificadas, el de clasificadas incorrectamente y el coeficiente Kappa. Los números en la matriz de confusión corresponden al número de muestras. El círculo rojo muestra las muestras asignadas correctamente.

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      8      88.8889 %
Incorrectly Classified Instances    1      11.1111 %
Kappa statistic                    0.7429

=== Confusion Matrix ===

a b c <-- classified as
6 0 0 | a = Origanum_vulgare
0 2 0 | b = Origanum_majorana
1 0 0 | c = Oregano

```

Figura 24. Resultados obtenidos en WEKA de clasificación de muestras de referencia y comerciales pertenecientes al género *Origanum* usando secuencias *barcode* del locus *rbcL*. Se reporta el porcentaje de instancias correctamente clasificadas, el de clasificadas incorrectamente y el coeficiente Kappa. Los números en la matriz de confusión corresponden al número de muestras.

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      6      85.7143 %
Incorrectly Classified Instances    1      14.2857 %
Kappa statistic                    0.72

=== Confusion Matrix ===

a b c <-- classified as
0 0 1 | a = Oregano
0 2 0 | b = Origanum_majorana
0 0 4 | c = Origanum_vulgare

```

Figura 25. Resultados obtenidos en WEKA de clasificación de muestras de referencia y comerciales pertenecientes al género *Origanum* usando secuencias *barcode* del locus *matK*. Se reporta el porcentaje de instancias correctamente clasificadas, el de clasificadas incorrectamente y el coeficiente Kappa. Los números en la matriz de confusión corresponden al número de muestras.

Un análisis similar al realizado con *Origanum* fue hecho con *Allium*. Se agruparon las muestras de referencia de *Allium sativum* y se clasificaron las dos muestras de ajo comerciales comparando con las de referencia. Utilizando las secuencias *barcode rbcL* se obtuvo un Kappa de 0 con lo cual los resultados de la clasificación no difieren significativamente de lo esperable por azar para el número de casos considerados. En este caso al igual que en orégano la muestra comercial es clasificada como perteneciente a la muestra de referencia, por lo que la clasificación es correcta aunque con ese valor de Kappa no podemos asegurar que no sea debido al azar (figura 26). Con las secuencias *barcode matK* las muestras comerciales son asignadas a la clase *Allium sativum* con un Kappa de 0,24 (Figura 27).

Si aumentamos el número de muestras y especies en el grupo a ser comparado la asignación es correcta en el 94% de los casos con un Kappa de 0,92 utilizando las secuencias *barcode rbcl*. Con *matK* el 87,5% de las instancias fueron correctamente asignadas con un Kappa de 0,83 (Figuras 28 y 29). Cabe destacar que las secuencias que incorporamos al grupo de datos en análisis son una parte reducida del total de muestras de referencia y que podrían ser especies que encontráramos juntas en una mezcla comercial formulada como condimento para platos específicos. En este caso fueron ajo, perejil y orégano por ejemplo.

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      6      85.7143 %
Incorrectly Classified Instances    1      14.2857 %
Kappa statistic                    0

=== Confusion Matrix ===

a b <-- classified as
6 0 | a = Allium_sativum
1 0 | b = Ajo_molido

```

Figura 26. Resultados obtenidos en WEKA de clasificación de muestras de referencia y comerciales pertenecientes al género *Allium* usando secuencias *barcode* del locus *rbcl*. Se reporta el porcentaje de instancias correctamente clasificadas, el de clasificadas incorrectamente y el coeficiente Kappa. Los números en la matriz de confusión corresponden al número de muestras.

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      6      75 %
Incorrectly Classified Instances    2      25 %
Kappa statistic                    0.2381

=== Confusion Matrix ===

a b c <-- classified as
0 1 0 | a = Ajo_escamas
0 0 1 | b = Ajo_molido
0 0 6 | c = Allium_sativum

```

Figura 27. Resultados obtenidos en WEKA de clasificación de muestras de referencia y comerciales pertenecientes al género *Allium* usando secuencias *barcode* del locus *matK*. Se reporta el porcentaje de instancias correctamente clasificadas, el de clasificadas incorrectamente y el coeficiente Kappa. Los números en la matriz de confusión corresponden al número de muestras.

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances   16       94.1176 %
Incorrectly Classified Instances  1       5.8824 %
Kappa statistic                 0.9163

=== Confusion Matrix ===

 a b c d e <-- classified as
6 0 0 0 0 | a = Allium_sativum
0 6 0 0 0 | b = Origanum_vulgare
0 0 2 0 0 | c = Origanum_majorana
0 0 0 2 0 | d = Petroselinum_sativum
1 0 0 0 0 | e = Ajo_molido

```

Figura 28. Resultados obtenidos en WEKA de clasificación de una muestras comercial de ajo respecto a un grupo de muestras de referencia usando secuencias *barcode* del locus *rbcl*. Se reporta el porcentaje de instancias correctamente clasificadas, el de clasificadas incorrectamente y el coeficiente Kappa. Los números en la matriz de confusión corresponden al número de muestras.

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances   14       87.5 %
Incorrectly Classified Instances  2       12.5 %
Kappa statistic                 0.8307

=== Confusion Matrix ===

 a b c d e f <-- classified as
0 1 0 0 0 0 | a = Ajo_escamas
0 0 1 0 0 0 | b = Ajo_molido
0 0 6 0 0 0 | c = Allium_sativum
0 0 0 2 0 0 | d = Origanum_majorana
0 0 0 0 4 0 | e = Origanum_vulgare
0 0 0 0 2 2 | f = Petroselinum_sativum

```

Figura 29. Resultados obtenidos en WEKA de clasificación de una muestras comercial de ajo respecto a un grupo de muestras de referencia usando secuencias *barcode* del locus *matK*. Se reporta el porcentaje de instancias correctamente clasificadas, el de clasificadas incorrectamente y el coeficiente Kappa. Los números en la matriz de confusión corresponden al número de muestras.

La estrategia de *DNA barcoding* para plantas implica el uso de por lo menos dos *loci* para llevar a cabo la asignación de especies. Pero la plataforma BOLD no permite por el momento trabajar con las secuencias de ambos *loci* simultáneamente como sucede en el caso de animales u hongos. En este caso se ingresa la secuencia barcode correspondiente a *COI* y el sistema devuelve la identificación a nivel de especie. En el caso de plantas la asignación se hace usando cada *locus* por separado y la base de datos devuelve los 100 taxones mas cercanos a la secuencia problema.

Para ver si lográbamos resolver este impedimento realizamos una prueba unificando las secuencias *barcode rbcl* y *matK* para cada una de las muestras de referencia con las que trabajamos. Luego utilizando la plataforma WEKA con el clasificador Lazy (IBK) para realizar la clasificación, obteniendo un 100% de asignaciones correctas con un Kappa de 1 como se observa en la figura 30. Utilizando las secuencias de ambos *loci* unidas pasamos de evaluar 348 atributos a 806. De acuerdo con algunos autores al aumentar el número de datos mejora el desempeño del método (Weitschek, Ficon, & Felici, 2014) (Austerlitz, y otros, 2009). Austerlitz (2009) propone como una alternativa, entre otras, el aumentar el largo de la secuencia con lo que aumenta el número de caracteres en el análisis.

Con esta modificación no logramos identificar todas las muestras comerciales en un solo análisis, tal como ocurrió cuando realizamos la clasificación con las secuencias *barcode rbcl* y *matK* por separado (Figura 31).

Aplicando la misma lógica que cuando trabajamos con las secuencias *barcode rbcl* y *matk* por separado, analizamos las muestras industriales comerciales de a una con respecto al total de muestras de referencia. Lo que observamos es que tanto la muestra de ajo como las de estragón las clasifica en forma correcta, en el grupo de *Allium sativum* y *Artemisia dracunculus* respectivamente, en ambos casos con un 96,8% de asignaciones correctas y un Kappa de 0,96.

Si disminuimos la cantidad de especies en el grupo de muestras de referencia y mantenemos solo las más relacionadas a la muestra problema vemos que cuanto menos especies hay, más acertada es la clasificación, pero el valor de Kappa va disminuyendo. Este hecho también fue observado cuando trabajamos con *rbcl* y *matK* por separado. Por ejemplo al tratar de clasificar la muestra de orégano comercial utilizando como muestras de referencia *Artemisia dracunculus*, *Origanum majorana*, *Origanum vulgare* y *Thymus vulgaris* vemos que la clasifica como *Artemisia dracunculus* con 94,7% de aciertos y un Kappa de 0,93. Si quitamos del grupo de datos las muestras de *Artemisia dracunculus* entonces lo clasifica como *Origanum majorana* con un 92,3% de aciertos y un Kappa de 0,88.

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances   54      100 %
Incorrectly Classified Instances  0        0 %
Kappa statistic                  1

=== Confusion Matrix ===

 a b c d e f g h i j k <-- classified as
6 0 0 0 0 0 0 0 0 0 0 | a = Achyrocline_flaccida
0 6 0 0 0 0 0 0 0 0 0 | b = Achyrocline_satureioides
0 0 6 0 0 0 0 0 0 0 0 | c = Allium_sativum
0 0 0 6 0 0 0 0 0 0 0 | d = Aloysia_citriodora
0 0 0 0 6 0 0 0 0 0 0 | e = Artemisia_dracunculus
0 0 0 0 0 6 0 0 0 0 0 | f = Baccharis_trimera
0 0 0 0 0 0 4 0 0 0 0 | g = Ocimum
0 0 0 0 0 0 0 2 0 0 0 | h = Origanum_majorana
0 0 0 0 0 0 0 0 4 0 0 | i = Origanum_vulgare
0 0 0 0 0 0 0 0 0 2 0 | j = Petroselinum_sativum
0 0 0 0 0 0 0 0 0 0 6 | k = Thymus_vulgaris

```

Figura 30. Resultados obtenidos en WEKA de clasificación de muestras de referencia usando secuencias *barcode* de los *loci matK* y *rbcL* de cada muestra unificadas. Se reporta el porcentaje de instancias correctamente clasificadas, el de clasificadas incorrectamente y el coeficiente Kappa. Los números en la matriz de confusión corresponden al número de muestras. El círculo rojo señala las instancias correctamente clasificadas.

Con los resultados obtenidos podemos decir que el procedimiento estadístico de clasificación utilizado funciona adecuadamente para fines taxonómicos, es decir clasifica correctamente muestras de especies conocidas llamadas "de referencia" en esta tesis.

Con respecto a la aplicación a muestras industrializadas, pudimos observar que en algunos casos funciona y permite clasificar las muestras problema en forma correcta. Lo que se puede destacar es que la significación estadística de los resultados mejora al aumentar el número de secuencias de referencia incluídas en el análisis y comparar una muestra comercial a la vez, como pudimos observar en el caso del ajo. El desempeño del método sin embargo no fue bueno al incluír todas las muestras de referencia y todas las comerciales. En ese caso falló en la asignación de las comerciales.

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances   54      91.5254 %
Incorrectly Classified Instances  5       8.4746 %
Kappa statistic                  0.9069
=== Confusion Matrix ===

 a b c d e f g h i j k l m n o p <-- classified as
6 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | a = Achyrocline_flaccida
0 6 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | b = Achyrocline_satureioides
0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 | c = Ajo_molido
0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 | d = Albahaca
0 0 0 0 6 0 0 0 0 0 0 0 0 0 0 0 | e = Allium_sativum
0 0 0 0 0 6 0 0 0 0 0 0 0 0 0 0 | f = Aloysia_citriodora
0 0 0 0 0 0 6 0 0 0 0 0 0 0 0 0 | g = Artemisia_dracunculus
0 0 0 0 0 0 0 6 0 0 0 0 0 0 0 0 | h = Baccharis_trimera
0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 | i = Estragon
0 0 0 0 0 0 0 0 4 0 0 0 0 0 0 0 | j = Ocimum
0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 | k = Oregano
0 0 0 0 0 0 0 0 0 0 2 0 0 0 0 0 | l = Origanum_majorana
0 0 0 0 0 0 0 0 0 0 0 4 0 0 0 0 | m = Origanum_vulgare
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 | n = Perejil
0 0 0 0 0 0 0 0 0 0 0 0 0 0 2 0 | o = Petroselinum_sativum
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 6 | p = Thymus_vulgaris

```

Figura 31. Resultados obtenidos en WEKA de clasificación de muestras comerciales con respecto a muestras de referencia usando secuencias *barcode* de los *loci matK* y *rbcL* de cada muestra unificadas. Se reporta el porcentaje de instancias correctamente clasificadas, el de clasificadas incorrectamente y el coeficiente Kappa. Los números en la matriz de confusión corresponden al número de muestras.

Austerlitz y colaboradores (2009) llevaron a cabo un estudio en el que compararon distintos métodos de clasificación para el tratamiento de secuencias *barcode*. De acuerdo al principio del método los agruparon en dos categorías: filogenéticos y de clasificación estadística supervisada. Entre los métodos estadísticos trabajaron con k-Nearest Neighbour (1-NN) y observaron que era el que tenía mejor desempeño y que resultó el más confiable en todas las situaciones que analizaron. Este método es menos proclive a ser confundido por la presencia de individuos de otra especie que compartan mutaciones con la especie problema. Este hecho también explicaría porque este método que utiliza un solo vecino más cercano ($k=1$) se desempeña mejor que aquellos que también consideran otros vecinos mas lejanos (2-NN, 3-NN, etc). De acuerdo a sus resultados concluyeron que al aumentar el número de datos mejora el desempeño de todos los métodos en general. Hay varias formas de hacerlo, ya sea aumentando el tamaño de la muestra, obtener mas información de las secuencias aumentando el largo, o secuenciar mas *loci* independientes incluyendo regiones nucleares (Austerlitz, y otros, 2009). Esta última alternativa contradice el espíritu de universalidad, simplicidad y estandarización del enfoque propuesto para la metodología *DNA barcoding*, si bien algunos autores lo han propuesto (Fazekas, y otros, 2009) (Wang, y otros, 2013).

5. CONCLUSIONES Y PERSPECTIVAS

El presente trabajo buscó demostrar la factibilidad de aplicación de técnicas moleculares basadas en ADN conformes con la estrategia de *DNA barcoding* para la clasificación de muestras de especias y hierbas medicinales, frescas e industrializadas.

De acuerdo a lo que se había planteado, se puso a punto la extracción de ADN de este tipo de muestras así como la amplificación y secuenciación de regiones *barcode* propuestas por el grupo de trabajo de plantas del *Consortium for the Barcode of Life*, *rbcl* y *matK*.

Se utilizó la plataforma *Barcode of Life Data Systems* para validar la clasificación de muestras de referencia de especie conocida y ubicar muestras problema en grupos de referencia. Se observó que la clasificación de muestras desconocidas es más difícil de lo esperado, especialmente si las muestras de referencia corresponden a poblaciones distintas que las muestras problema. Una de las principales limitantes es la cobertura a nivel taxonómico de las bases de datos de referencia.

Para que esta metodología sea factible de aplicar a nivel industrial es fundamental la existencia de una base de datos global que contenga una gran cantidad de secuencias *barcode* de referencia pertenecientes a distintas poblaciones de las mismas especies. Esto aseguraría una representación de la variabilidad genética dentro de cada especie, haciendo posible la diferenciación de muestras pertenecientes a especies cercanas.

Las dificultades con las que nos encontramos utilizando el enfoque de *DNA barcoding* nos indican la necesidad de encontrar métodos alternativos para asignar secuencias *barcode* de muestras desconocidas a especies conocidas.

Con este objetivo, como parte de esta tesis evaluamos la aplicabilidad de un procedimiento estadístico de clasificación basado en minería de datos, independiente de comparación con bases de datos externas, para la clasificación de secuencias *barcode* de muestras de plantas aromáticas y medicinales.

Nuestros resultados indican que este tipo de métodos de clasificación son promisorios candidatos para llevar a cabo la clasificación de especímenes desconocido en especies conocidas mediante el uso de secuencias *barcode*, obteniendo buenos resultados.

Sin embargo, algunas consideraciones deben ser tomadas en cuenta a la hora de aplicar estos métodos. Es necesario construir una base de datos local de muestras de referencia que incluya las posibles especies a la que pertenecen las muestras problema

para ser comparadas. El diseño del grupo de datos es importante e influye en la obtención de resultados. En nuestra experiencia el clasificar de a una muestra problema por vez da mejores resultados que incluirlas todas en el mismo grupo de análisis. Por otro lado, en casos donde se requiera diferenciar una especie contaminante o adulterante, observamos que comparar una muestra problema con un grupo de muestras de referencia acotado que incluya las especies en cuestión influye positivamente en la correcta clasificación.

Asimismo, realizamos algunas evaluaciones uniendo las secuencias *barcode* de ambos *loci* para cada muestra y observamos que permite clasificar correctamente las muestras de referencia.

Como perspectivas de este trabajo consideramos que sería interesante profundizar en el estudio del uso de procedimientos estadísticos de clasificación, utilizando mayor cantidad de muestras por especie tanto de referencia como comerciales de forma de ampliar la diversidad genética. Esta estrategia permitirá mejorar los porcentajes de asignaciones correctas. Otro punto interesante es el uso de ambas secuencias *barcode* en el mismo análisis. Esto permitiría agregar datos con lo cual mejora el desempeño de este tipo de métodos.

6. REFERENCIAS BIBLIOGRÁFICAS

- Agencia Española de Seguridad Alimentaria y Nutrición (AESAN). (2009). Guía para la aplicación de la trazabilidad en la empresa agroalimentaria.
- Aglarova, A. Z. (2008). Biological characteristics and useful properties of Tarragon (*Artemisia dracunculus* L.) (Review). *Pharmaceutical Chemistry Journal*, 42(2).
- Ali, M., G., G., Hidvégi, N., Kerti, B., Al Hemaïd, F., Pandey, A., & Lee, J. (2014). The changing epitome of species identification - DNA barcoding. *Saudi Journal of Biological Sciences*(21), 204 - 231.
- Applied Biosystems. (2009). DNA Sequencing by Capillary Electrophoresis. *Applied Biosystems Chemistry Guide*(Second Edition).
- Austerlitz, F., David, O., Schaeffer, B., Bleakley, K., Olteanu, M., Leblois, R., Laredo, C. (2009). DNA barcode analysis: a comparison of phylogenetic and statistical classification methods. *BMC Bioinformatics*, 10(Suppl 14):S10, doi: 10.1186/1471-2105-10-S14-S10.
- Barthet, M., & Hilu, K. (2007). Expression of matK: Functional and evolutionary implications. *American Journal of Botany*, 94(8), 1402 - 1412.
- Bauer, T., Weller, P., Hammes, W. P., & Hertel, C. (2003). The effect of processing parameters on DNA degradation in food. *European Food Research Technology*(217), 338–343.
- Bertolazzi, P., Felici, G., & Weitschek, E. (2009). Learning to classify species with barcodes. *BMC Bioinformatics* (10), S7 doi: 10.1186/1471-2105-10-S14-S7.
- Biomatters. (2014). <http://www.geneious.com>.
- BOLD Systems, Biodiversity Institute of Ontario. (2014-2017). *BOLD Systems*. Recuperado el 16 de Mayo de 2017, de <http://v4.boldsystems.org/index.php>
- Bruni, I., De Mattia, F., Galimberti, A., Galasso, G., Banfi, E., Casiraghi, M., & Labra, M. (2010). Identification of poisonous plants by DNA barcoding approach. *International Journal of Legal Medicine*(124), 595-603.

- CBI Market Intelligence. (Junio de 2015). *CBI Centre for the Promotion of Imports from Developing Countries*. Recuperado el Febrero de 2016, de www.cbi.eu/market-information
- CBOL Plant Working Group. (2009). A DNA barcode for land plants. *Proceedings of the National Academy of Sciences USA (PNAS)*, 106(31), 12794 - 12797.
- Chen, S., Yao, H., Han, J., Liu, C., Song, J., Shi, L., . . . Leon, C. (2010). Validation of the ITS2 Region as a Novel DNA Barcode for Identifying Medicinal Plant Species. *PloS ONE*, 5(1).
- Codex Alimentarius Commission. (2006). *FAO's Animal Production and Health Division*. Recuperado el 25 de 01 de 2016, de http://www.fao.org/ag/againfo/themes/en/meat/quality_trace.html
- Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(37), <http://epm.sagepub.com/content/20/1/37>.
- Consortium for the Barcode of Life. (2010 - 2017). *Barcode of Life. Identifying species with DNA barcoding* . Recuperado el 16 de Mayo de 2017, de <http://www.barcodeoflife.org/content/about/what-dna-barcoding>
- Cowan, R., Chase, M., Kress, W., & Savolainen, V. (2006). 300,000 species to identify: problems, progress, and prospects in DNA barcoding of land plants. *Taxon*, 55(3), 611-616.
- Davies, P. (Noviembre de 2004 a). Plantas aromáticas y medicinales. *Estudios en domesticación y cultivo de especies medicinales y aromáticas nativas*(11). (I. N. (INIA), Ed.) Montevideo, Uruguay.
- Davies, P. (Noviembre de 2004 b). La marcela: revisión de bibliografía. *Estudios en domesticación y cultivo de especies medicinales y aromáticas nativas, Serie FPTA-INIA 11*, 193 - 199. (INIA, Ed.) Montevideo.
- De Mattia, F., Bruni, I., Galimberti, A., Cattaneo, F., Casiraghi, M., & Labra, M. (2011). A comparative study of different DNA barcoding markers for the identification of some members of Lamiaceae. *Food Research International*, 693 - 702.
- Dellacassa, E., Moyna, P., & Nieto, A. (2005). *Aportes para el desarrollo del sector de plantas medicinales y aromáticas en Uruguay* . Montevideo: Fundaquim, Red Propymes, URU.TEC, Fundasol, Cooperación Técnica Alemana (GTZ).
- Dhanya, K., & Sasikumar, B. (2010). Molecular Marker Based Adulteration Detection in Traded Food and Agricultural Commodities of Plant Origin with Special

- Reference to Spices. *Current Trends in Biotechnology and Pharmacy*, 4(1), 454 - 489.
- Douglas, M., Heyes, J., & Smallfield, B. (2005). Recuperado el Febrero de 2016, de <http://www.fao.org/3/a-ad420e.pdf>
- Ebach, M., & Holdrege, C. (2005). DNA barcoding is no substitute for taxonomy. *Nature*(434), 697.
- Edgar, R. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5), <https://doi.org/10.1093/nar/gkh340>.
- European Parliament and of the Council. (28 de January de 2002). Regulation (EC) No 178/2002. *Official Journal of the European Communities*. Brussels.
- Ewing, B., & Green, P. (1998). Base-calling of automated sequencer traces using Phred. II. Error probabilities. *Genome Research*, 8, 186- 194. Obtenido de genome.cshlp.org.
- Fazekas, A. J., Burgess, K., Kesanurti, P. R., Graham, S., Newmaster, S., Husband, B., Barrett, S. (2008). Multiple Multilocus DNA Barcodes from the Plastid Genome Discriminate Plant Species Equally Well. *PLoS ONE*, 3(7), e2802.
- Fazekas, A., Kesanakurti, P., Burgess, K. S., Percy, D., Graham, S. W., Barrett, S. C., Husband, B. C. (2009). Are plant species inherently harder to discriminate than animal species using DNA barcoding markers? *Molecular Ecology Resources*, 9 (Supl 1), 130-139.
- Federici, S., Fontana, D., Galimberti, A., Bruni, I., de Mattia, F., Cortis, P., Labra, M. (2014). A rapid diagnostic approach to identify poisonous plants using DNA barcoding data. *Plant Biosystems*.
- Federici, S., Galimberti, A., Bartolucci, F., Bruni, I., De Mattia, F., Cortis, P., & Labra, M. (2013). DNA barcoding to analyse taxonomically complex groups in plants: the case of Thymus (Lamiaceae) groups in plants: the case of Thymus (Lamiaceae). *Botanical Journal of the Linnean Society*(171), 687-689.
- Fitoterapia*. (2016). Recuperado el Febrero de 2016, de <http://dec.fq.edu.uy/wwds2008/fitoterapia/index.html>
- Frackman, S., Kobs, G., Simpson, D., & Storts, D. (1998). Betaine and DMSO: Enhancing Agents for PCR. *Promega Notes*(25), 27 - 30.

- Galimberti, A., De Mattia, F., Losa, A., Bruni, I., Federici, S., Casiraghi, M., Labra, M. (2013). DNA barcoding as a new tool for food traceability. *Food Research International*(50), 55 - 63.
- Galimberti, A., Labra, M., Sandionigi, A., & Bruno, A. (2014). DNA Barcoding for Minor Crops and Food Traceability. *Advances in Agriculture*, 1 - 8.
- Gao, T., Yao, H., Song, J., Zhu, Y., Liu, C., & Chen, S. (2010 a). Evaluating the feasibility of using candidate DNA barcodes in discriminating species of the large Asteraceae family. *BMC Evolutionary Biology*, 10, 324.
- Gao, T., Yao, H., Song, J., Liu, C., Zhu, Y., Ma, X., Chen, S. (2010 b). Identification of medicinal plants in the family Fabaceae using a potential DNA barcode ITS2. *Journal of Ethnopharmacology*(130), 116-121.
- Gismondi, A., Fanali, F., Martínez Labarga, M. L., Grilli Caiola, M., & Canini, A. (2013). *Crocus sativus* L. genomics and different DNA barcode applications. *Plant Systematics Evolution*.
- Gregory, T. (2005). DNA barcoding does not compete with taxonomy. *Nature*(434), 1067.
- Haider, N. (2011). Chloroplast-specific universal primers and their uses in plant studies. *Biologia Plantarum*, 55(2), 225-236.
- Hajibabaei, M., Singer, G. H., & Hickey, D. (2007). DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *TRENDS in Genetics*, 23(4), 167-172.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. (2009). The WEKA Data Mining Software: An Update. *SIGKDD Explorations*, 11(1), 10 - 18.
- Hanner, R. (26 de Marzo de 2009). Data Standards for BARCODE Records in INSDC (BRIs).
- Havey, M. (1991). Phylogenetic relationships among cultivated *Allium* species from restriction enzyme analysis of the chloroplast genome. *Theoretical and Applied Genetics*, 81, 752-757.
- Hebert, P., Cywinska, A., Ball, S., & deWaard, J. (2003). Biological identifications through DNA barcodes. *Proc. R. Soc. Lond. B* (270), 313 - 321.
- Hebert, P., Ratnasingham, S., & de Waard, J. (2003). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc. R. Soc. Lond. B (Suppl)*(270), S96-S99.

- Heubl, G. (2010). New Aspects of DNA-based Authentication of Chinese Medicinal Plants by Molecular Biological Techniques. *Planta Medica*, 76, 1963 - 1974.
- Hollingsworth, P. M., Graham, S. W., & Little, D. P. (2011). Choosing and Using a Plant DNA Barcode. *PLoS ONE*, 6(5), e19254. doi:10.1371/journal.pone.0019254.
- Howard, C. (Junio de 2010). The development of Deoxyribonucleic Acid (DNA) based methods for the identification and authentication of medicinal plant material.
- INIA (Ed.). (Noviembre de 2004). Estudios en domesticación y cultivo de especies medicinales y aromáticas nativas. *Serie FPTA-INIA(11)*. Uruguay.
- International Barcode of Life. (2014). *Bold Systems*. Recuperado el 23 de Mayo de 2017, de <http://www.boldsystems.org/>
- Ipek, M., Ipek, A., & Simon, P. (2014). Testing the utility of matK and ITS DNA regions for discrimination of *Allium* species. *Turkish Journal of Botany*, 38, 203 - 212.
- Ivanova, N., deWaard, J., Hajibababei, M., & Hebert, P. (2017). *Protocols for High Volume DNA Analysis*. Recuperado el 15 de Mayo de 2017, de Barcode of Life: <http://www.barcodeoflife.org/content/about/what-dna-barcoding>
- Joint FAO/IAEA Division of Nuclear Techniques in Food and Agriculture. (2002). Mutant Germplasm Characterization using Molecular Markers. A Manual. *Technical Courses Series 19*. Vienna, Austria: International Atomic Energy Agency.
- Joshi, K., Chavan, P., Warude, D., & Patwardhan, B. (2004). Molecular markers in herbal drug technology. *Current Science*, 87, 159 - 165.
- Kaufman, M., & Wink, M. (1994). Molecular Systematics of the Nepetoideae (Family Labiateae): Phylogenetic Implications from rbcL Gene Sequences. *Z. Naturforsch*, 635 - 645.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12), 1647-1649.
- Khanuja, S. P., Shasany, A. K., Darokar, M. P., & Kumar, S. (1999). Rapid Isolation of DNA from Dry and Fresh Samples of Plants Producing Large Amounts of Secondary Metabolites and Essential Oils. *Plant Molecular Biology Reporter*, 17, 1 - 7.
- Kool, A., de Boer, H., Krüger, A., Rydberg, A., Abbad, A., Björk, L., & Martin, G. (2012). Molecular Identification of Commercialized Medicinal Plants in Southern Morocco. *PLoS ONE*, 7(6), e39459. doi:10.1371/journal.pone.0039459.

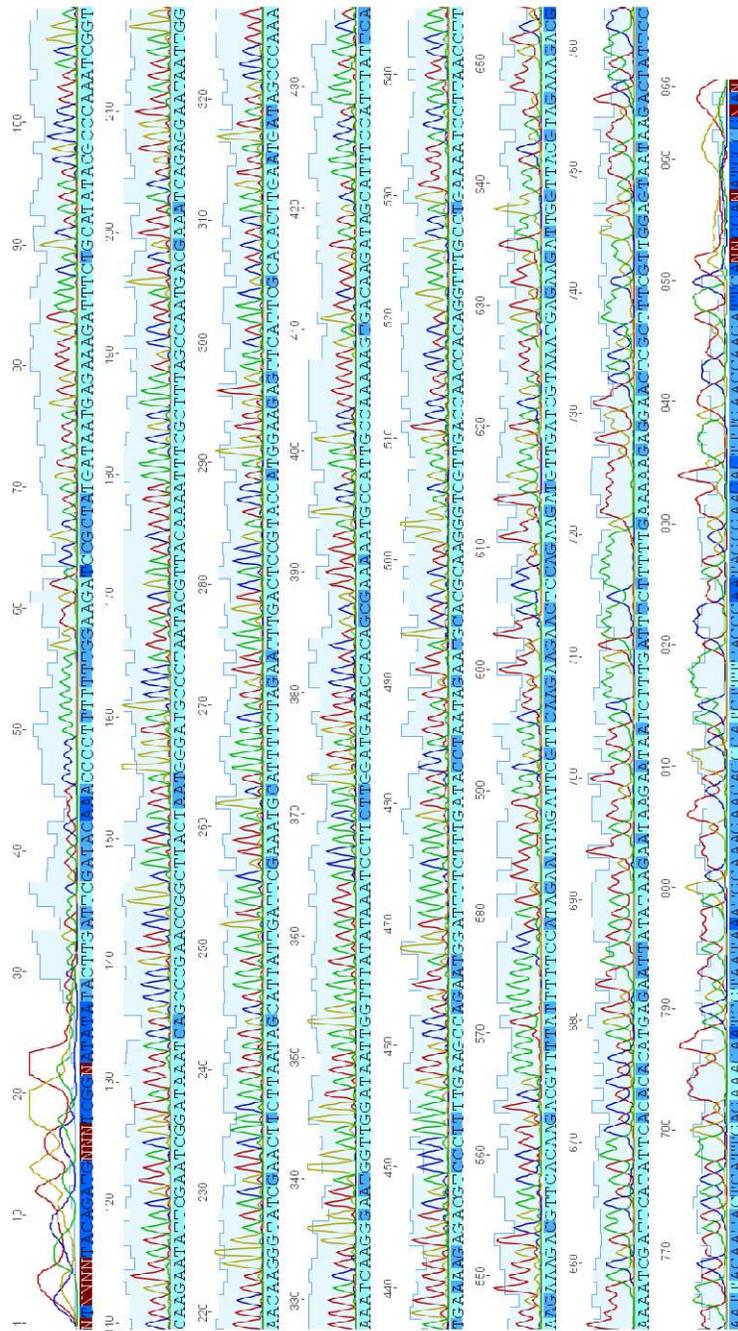
- Kress, W., & Erickson, D. (2007). A Two-Locus Global DNA Barcode for Land Plants: The Coding *rbcl* Gene Complements the Non-Coding *trnHpsbA*. *PLoS ONE*, 2(6), e508. doi:10.1371/journal.pone.0000508.
- Kress, W., Wurdack, K., Zimmer, E., Weigt, L., & Janzen, D. (2005). Use of DNA barcodes to identify flowering plants. *PNAS*, 102(23), 8369-8374.
- Kunle, O. F., Egharevba, H. O., & Ahmadu, P. O. (2012). Standardization of herbal medicines - A review. *International Journal of Biodiversity and Conservation*, 4(3), 101 - 112.
- Labra, M., Miele, M., Ledda, B., Grassi, F., Mazzei, M., & Sala, F. (2004). Morphological characterization, essential oil composition and DNA genotyping of *Ocimum basilicum* L. cultivars. *Plant Science*, 167, 725 - 731.
- Lahaye, R., van der Bank, M., Bogarin, D., Warner, J., Pupulin, F., Gigot, G., Savolainen, V. (February de 2008). DNA barcoding the floras of biodiversity hotspots. *PNAS*, 105(8), 2923 - 2928.
- Li, M., Au, K.-Y., Lam, H., Cheng, L., Jiang, R.-W., Pui-Hay But, P., & Shaw, P.-C. (2012). Identification of Baiying (*Herba Solani Lyrati*) commodity and its toxic substitute Xungufeng (*Herba Aristolochiae Mollissimae*) using DNA barcoding and chemical profiling techniques. *Food Chemistry*(135), 1653-1658.
- López, M. M. (2003). *Tecnologías Moleculares de Trazabilidad Alimentaria, Informe de Vigilancia Tecnológica*. Genoma España/CIBT-FGUAM.
- Lukas, B. (Junio de 2010). Molecular and phytochemical analyses of the genus *Origanum* L. (Lamiaceae). *Disertación*. Viena: Universidad de Viena.
- Mahadani, P., & Ghosh, S. (2013). DNA Barcoding: A tool for species identification from herbal juices. *DNA*, 35-38.
- Manning, L., & Soon, J. (2014). Developing systems to control food adulteration. *Food Policy*, 49, 23 - 32.
- Milton, M., Pierossi, P., & Ratnasingham, S. (2013). *Barcode of Life Data Systems*. Recuperado el Diciembre de 2014, de <http://www.boldsystems.org>
- Mondal, S., Shit, S., & Kundu, S. (January de 2013). A comparative computational study of the *rbcl* gene in plants and in the three prokariotic families Archea, Cyanobacteria and Proteobacteria. *Indian Journal of Biotechnology*, 12, 58-66.
- Moore J.C., S. J. (2012). Development and Application of a Database of Food Ingredient Fraud and Economically Motivated Adulteration from 1980 to 2010. *Journal of Food Science*, 77(4), 118 - 126.

- Morales, R. (Febrero de 1996). Studies on the genus *Thymus* L. *Lamiales Newsletter*(ISSN 1358-2305). Royal Botanic Gardens Kew.
- Morquio, A., Ferreira, M., Arredondo, F., Blasina, F., Echeverry, C., Abin, J., & Dajas, F. (2004). Estudio comparativo de compuestos polifenólicos, actividad citoprotectora y antioxidante en especies autóctonas del género *Achyrocline*. En P. Davies, *Estudios en Domesticación y Cultivo de Especies Medicinales y Aromáticas Nativas* (págs. 169 - 178). Montevideo: INIA.
- Newmaster, S., Fazekas, A., & Ragupathy, S. (2006). DNA barcoding in land plants: evaluation of rbcL in a multigene tiered approach. *Can.J.Bot*(84), 335-341.
- Newmaster, S., Fazekas, A., Steeves, R., & Janovec, J. (2008). Testing candidate plant barcode regions in the Myristicaceae. *Molecular Ecology Resources*(8), 480 - 490.
- Newmaster, S., Grguric, M., Shanmughanandhan, D., & Ramalingam, S. (2013). DNA barcoding detects contamination and substitution in North American herbal products. *BMC Medicine*(11), 222.
- Nikam, P., Kareparamban, J., Jadhav, A., & Kadam, V. (2012). Future Trends in Standardization of Herbal Drugs. *Journal of Applied Pharmaceutical Science 02 (06)*; 2012: 38-44, 2(6), 38-44.
- Opara, L. (2003). Traceability in agriculture and food supply chain: A review of basic concepts, technological implications, and future prospects. *Food, Agriculture & Environment*, 1(1), 101 - 106.
- Parvathy, V., Swetha, V., Sheeja, T., Leela, N., Chempakam, B., & Sasikumar, B. (2014). DNA barcoding to detect chilli adulteration in traded black pepper powder. *Food Biotechnology*, 28(1), 2-40.
- Pascual, M., Slowing, K., Carretero, E., Sánchez Mata, D., & Villar, A. (2001). *Lippia*: traditional uses, chemistry and pharmacology: a review. *Journal of Ethnopharmacology*, 76, 201-214.
- Pečnikar, Z., & Buzan, E. (2013). 20 years since the introduction of DNA barcoding: from theory to application. *Journal of Applied Genetics*.
- Ratnasingham, S., & Hebert, P. (2007). Bold: The Barcode of Life Data System (www.barcodinglife.org). *Molecular Ecology Notes*(7), 355 - 364.
- Reglamento Bromatológico Nacional. (Febrero de 2012). *5ta Edición*. IMPO (Dirección Nacional de Impresiones y Publicaciones Oficiales).

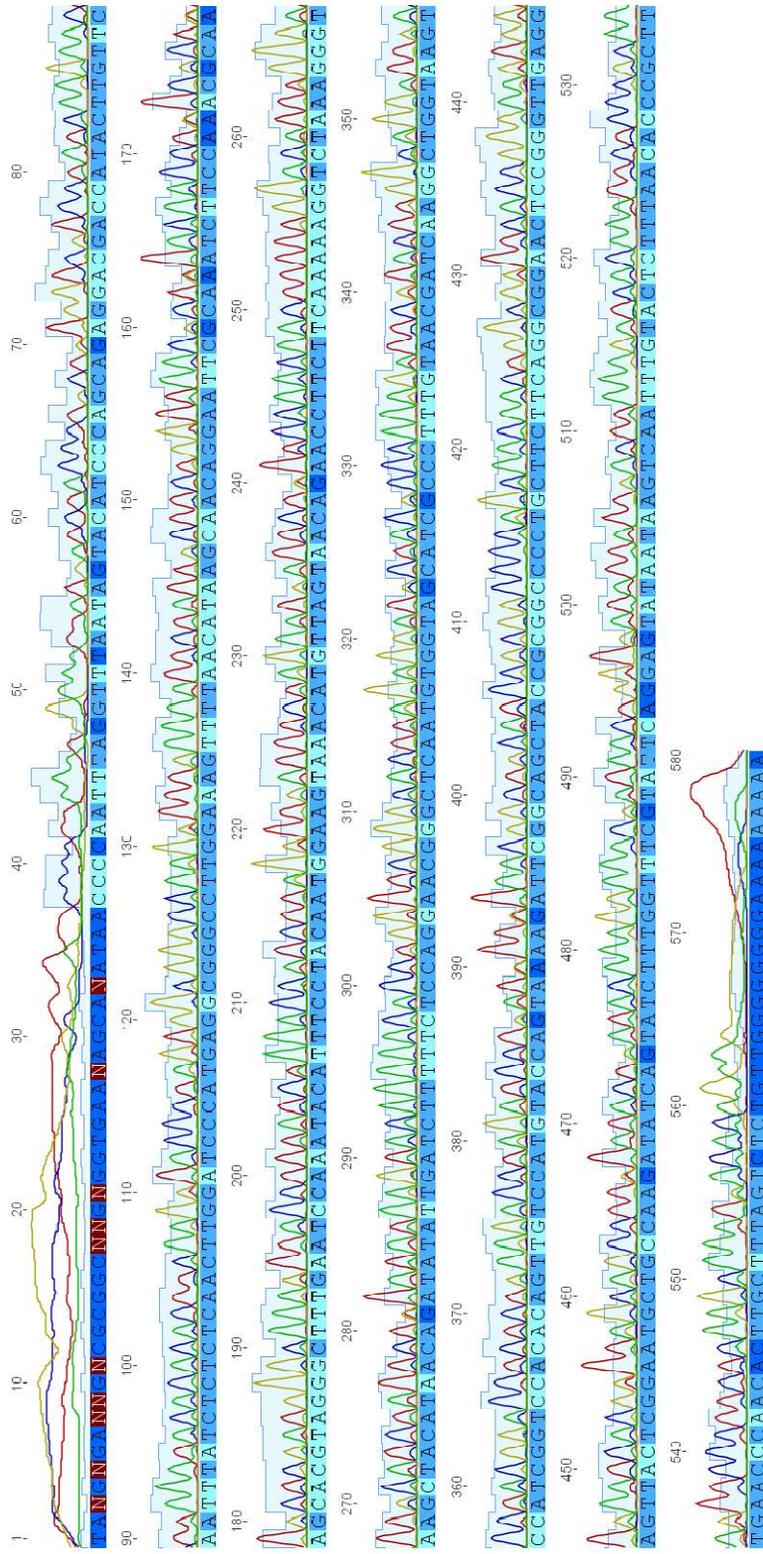
- Research Group Molecular Genetics - Humboldt Universitat Zu Berlin. (2016). *Molecular Genetics - Humboldt Universitat Zu Berlin*. Recuperado el 10 de Mayo de 2017, de <https://www.molgen.hu-berlin.de/projects-matk.php>
- Roux, K. H. (1995). Optimization and troubleshooting in PCR. *Genome Research*, 4, S185 - S194.
- Royal Botanic Gardens. (29 de Marzo de 2016). *Royal Botanic Gardens, Kew*. Recuperado el 29 de Marzo de 2016, de <http://www.kew.org/science/compositae.html>
- Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4(4), 406-425.
- Schindel, D., & Miller, S. (2005). DNA barcoding a useful tool for taxonomists. *Nature*(435), 17.
- Schori, M., & Showalter, A. (2011). DNA Barcoding as a means for identifying medicinal plants of Pakistan. *Pak. J. Bot.*(43), 1-4.
- Southern African Development Community. (2016). *SADC Trade*. Recuperado el Febrero de 2016, de www.sadctrade.org/TIB/spices
- Stoeckle, M., Chenery, J., & Schindel, D. (2011). *Barcode of Life*. Recuperado el 17 de Mayo de 2017, de <http://www.barcodeoflife.org/content/resources/outreach-materials>
- Stoeckle, M., Gamble, C., Kirpekar, R., Young, G., & Ahmed, S. (2011). Commercial Teas Highlight Plant DNA Barcode Identification Successes and Obstacles. *Nature Scientific Reports*, 1-7.
- Sucher, N., & Carles, M. (2008). Genome-Based Approaches to the Authentication of Medicinal Plants. *Planta Medica*(74), 603-623.
- Sudha Revathy, S., Rathinamala, R., & Murugesan, M. (2012). Authentication methods for drugs used in Ayurveda, Siddha and Unani systems of medicine: an overview. *International Journal of Pharmaceutical Sciences and Research (IJPSR)*, 3(8), 2352 - 2361.
- Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A., Brochmann, C. (2007). Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. *Nucleic Acids Research*, 35(3), e14.
- Taylor, L. (1996). *Wealth of the Rainforest, Pharmacy to the World*. Recuperado el Mayo de 2016, de <http://www.rain-tree.com>

- Techen, N., S.L. Crockett, S., Khan, I., & Scheffler, B. (2004). Authentication of Medicinal Plants Using Molecular Biology Techniques to Compliment Conventional Methods. *Current Medicinal Chemistry*, 1391 - 1401.
- Theodoridis, S., Stefanaki, A., Tezcan, M., Aki, C., Kokkini, S., & Vlachonasios, K. (2012). DNA barcoding in native plants of the Labiatae (Lamiaceae) family from Chios Island (Greece) and the adjacent Cesme-Karaburun Peninsula (Turkey). *Molecular Ecology Resources*, 620 - 633.
- Thomson, M., & Ali, M. (2003). Garlic [*Allium sativum*]: A Review of its Potential Use as an Anti-Cancer Agent. *Current Cancer Drug Targets*, 3, 67 - 81.
- USDA. (Abril de 2015). *USDA ERS - U.S. Food imports*. Recuperado el Febrero de 2016, de <http://www.ers.usda.gov/data-products/us-food-imports.aspx>
- Valentini, A., Pompanon, F., & Taberlet, P. (2008). DNA barcoding for ecologists. *Trends in Ecology and Evolution*, 24(2), 110-117.
- van Velzen, R., Weitschek, E., Felici, G., & Bakker, F. (2012). DNA Barcoding of Recently Diverged Species: Relative Performance of Matching Methods. *PLoS ONE*, 7(1), e30490. doi:10.1371/journal.pone.0030490.
- Vijayan, K., & Tsou, C. H. (2010). DNA barcoding in plants: taxonomy in a new perspective. *Current Science*, 99(11), 1530 - 1541.
- Wallace, L., Boilard, S., Eagle, S., Spall, J., Shokralla, S., & Hajibabaei, M. (2012). DNA barcodes for everyday life: Routine authentication of Natural Health Products. *Food Research International*(49), 446-452.
- Wang, M., Zhao, H.-X., Wang, L., Wang, T., Yang, R.-W., Wang, X.-L., Zhang, L. (2013). Potential use of DNA barcoding for the identification of *Salvia* based on cpDNA and nrDNA sequences. *Gene*, 528, 206-215.
- Weitschek, E., Fison, G., & Felici, G. (2014). Supervised DNA Barcodes species classification: analysis, comparisons and results. *BioData Mining* (7), 4.

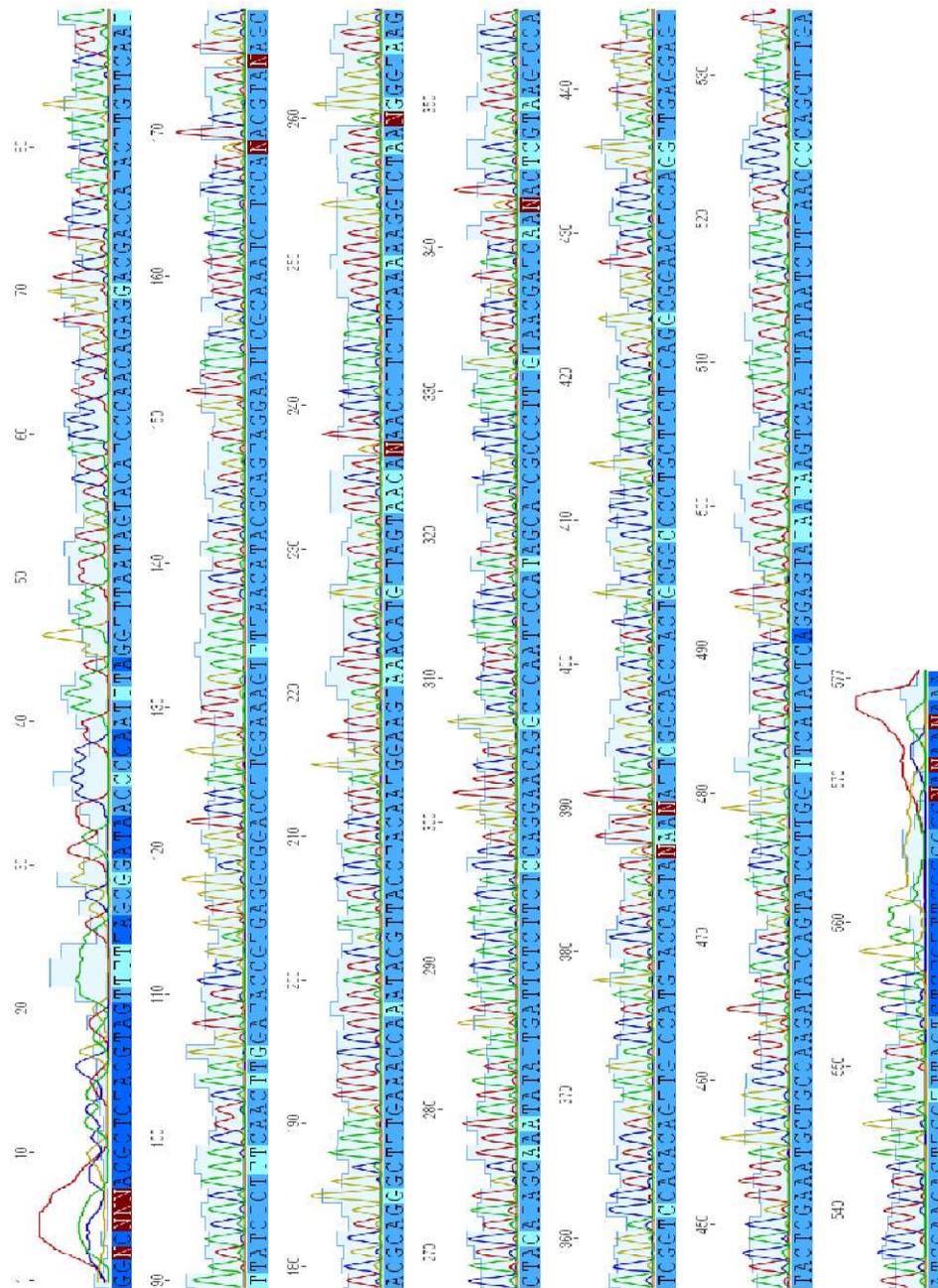
Anexo 2. Electroferograma correspondiente a la secuencia forward del locus *matK* de la muestra *Aloysia triphylla* 3.



Anexo 3. Electroferograma correspondiente a la secuencia reversa del locus *rbcl* de la muestra comercial de orégano.



Anexo 5. Electroferograma correspondiente a la secuencia directa del locus *rbcl* de la muestra comercial albahaca.



Anexo 6. Electroferograma correspondiente a la secuencia directa del locus *matK* de la muestra comercial albahaca.

