



**Caracterización molecular de variedades de arroz adaptadas a la provincia de Loja utilizando
marcadores microsatélites (SSRs)**

Reyes Salinas, Fabian Andres

Departamento de Ciencias de la Vida y de la Agricultura

Carrera de Ingeniería en Biotecnología

Trabajo de titulación, previo a la obtención del título de Ingeniero en Biotecnología

Ochoa Tufiño, Andrea Valeria Ph.D.

17 de septiembre de 2021

Informe de originalidad

NOMBRE DEL CURSO

Revisión de tesis

NOMBRE DEL ALUMNO

FABIAN ANDRÉS REYES SALINAS

NOMBRE DEL ARCHIVO

FABIAN ANDRÉS REYES SALINAS - Documento sin título

SE HA CREADO EL INFORME

14 sept 2021

Resumen

Fragmentos marcados	2	0,5 %
Fragmentos citados o entrecorridos	0	0 %

Coincidencias de la Web

tec.ac.cr	1	0,4 %
espe.edu.ec	1	0,2 %

1 de 2 fragmentos

Fragmento del alumno [MARCADO](#)

Oryza sativa es una gramínea anual, de tallos redondos y huecos compuestos por nudos y entrenudos, hojas de lámina plana, lineal y larga, unidas al tallo por la vaina de forma alternada y su ...

[Mejor coincidencia en la Web](#)

La planta de arroz (**Oryza sativa**), es una gramínea anual, de tallos redondos y huecos compuestos de nudos y entrenudos, con hojas de lámina plana unidas al tallo por una vaina y con inflorescencia...

Oryza sativa - Repositorio
TEC

<https://repositorio.tec.ac.cr/bitstream/handle/2238/5875/Determinaci%C3%B3n%20preliminar%20de%20g%C3%A9neros%20y%20densidades%20poblacion%20sequence=1&isAllowed=y>

2 de 2 fragmentos

Fragmento del alumno [MARCADO](#)

La asignación de genotipos a poblaciones predeterminadas se realizó con el software STRUCTURE vs. 2. 3.4 desarrollado por Pritchard y colaboradores (2000) con...

[Mejor coincidencia en la Web](#)

Asignación genética **La asignación de los genotipos a poblaciones predeterminadas, se realizó con el programa Structure vs. 2.3.4.**

ARTICULO CIENTIFICO CARACTERIZACION MOLECULAR <http://repositorio.espe.edu.ec/bitstream/21000/8016/1/AC-B-ESPE-047648.pdf>

<https://classroom.google.com/g/tg/MzlwODY2NDMzNjI1/MzlwOTUzMDk2NDIz#u=NzY2NTAxOTY2MjJa&t=f>

1/1



Firmado electrónicamente por:
**ANDREA
VALERIA OCHOA
TUFINO**



Departamento de Ciencias de la Vida y de la Agricultura

Carrera de Ingeniería en Biotecnología

Certificación

Certifico que el trabajo de titulación, “**Caracterización molecular de variedades de arroz adaptadas a la provincia de Loja utilizando marcadores microsatélites (SSRs)**” fue realizado por el señor **Reyes Salinas Fabian Andres** el cual ha sido revisado y analizado en su totalidad por la herramienta de verificación de similitud de contenido; por lo tanto cumple con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, razón por la cual me permito acreditar y autorizar para que lo sustente públicamente.

Sangolquí, 13 de septiembre de 2021



Firmado electrónicamente por:

**ANDREA
VALERIA OCHOA**

TUFINO

Ochoa Tufiño, Andrea Valeria PhD

CI. 1719478669



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

Departamento de Ciencias de la Vida y de la Agricultura

Carrera de Ingeniería en Biotecnología

Responsabilidad de Autoría

Yo, **Fabian Andres Reyes Salinas**, con cédula de ciudadanía N° 1804086344, declaro que el contenido, ideas y criterios del trabajo de titulación, "**Caracterización molecular de variedades de arroz adaptadas a la provincia de Loja utilizando marcadores microsatélites (SSRs)**" es de mi autoría y responsabilidad. Por lo tanto, cumple con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE; respetando los derechos intelectuales de terceros y referenciando las citas bibliográficas.

Sangolquí, 17 de septiembre de 2021

Reyes Salinas, Fabian Andres

CI. 1804086344



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

Departamento de Ciencias de la Vida y de la Agricultura

Carrera de Ingeniería en Biotecnología

Autorización de Publicación

Yo, **Reyes Salinas Fabian Andres**, con cédula de ciudadanía N° 1804086344, autorizo a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de titulación, "**Caracterización molecular de variedades de arroz adaptadas a la provincia de Loja utilizando marcadores microsatélites (SSRs)**" en el Repositorio Institucional, cuyo contenido, ideas y criterios son de mi responsabilidad.

Sangolquí, 17 de septiembre de 2021

Reyes Salinas, Fabian Andres

CI. 1804086344

Dedicatoria

Dedico este trabajo de investigación a todos aquellos que han contribuido a mi crecimiento y han sido parte de este camino.

A mis padres, Fabián y Fanny quienes han sido un pilar importante en mi vida y siempre me han apoyado en cualquier sueño y objetivo que me planteo por más difícil que parezca.

A mis hermanos Steven y Carolina por ser amigos y compañeros en todas las etapas de mi vida, sin duda su presencia ha llenado de colores cada historia.

A mis abuelos Teresa, Juan y Piedad por todas sus oraciones, sus consejos y su amor. Son mi fuente inagotable de energía.

Al resto de mi familia por ser inspiración, apoyo y refugio.

Finalmente, a todas aquellas personas que se han convertido en motivación, tinta y muelle.

Agradecimiento

Quiero agradecer a Dios por todas las bendiciones y oportunidades que se han presentado y se presentarán a lo largo de mi vida.

A mi directora de tesis Valeria Ochoa, PhD por su tiempo, consejos, conocimiento y guía durante el proceso de realización de este proyecto, sin duda alguna aspiro llegar a ser un científico como ella.

Al Instituto Nacional de Investigaciones Agropecuarias INIAP y al doctor Eduardo Morillo por haberme dado la oportunidad de desarrollarme y obtener nuevos conocimientos con el proyecto al que me encontré vinculado, de manera muy especial a la doctora Elisa Quiala por acogerme como familia durante mi estancia en la Estación Litoral Sur y por toda su guía personal y profesional.

A las MsC Johana Buitron y Lizeth Ojeda por ser apoyo en el laboratorio, todos sus consejos han sido valiosos para mi crecimiento personal.

A mis amigos Karla, Gabriela, Diego, Dani, Brillo, Massiel, Kerly, Lorena y David por ser la familia que me ha acompañado durante mi vida universitaria.

De manera especial a aquellas personas que han sido motor para salir adelante y me han ayudado a crecer. Gracias por darme confianza en mi mismo y razones para seguir adelante.

Índice de Contenido

Listado de Abreviaturas	13
Resumen	14
Abstract	15
Capítulo 1: Introducción	16
Planteamiento del problema	16
Justificación del problema	17
Objetivos	17
Objetivo general.....	17
Objetivos específicos	17
Marco Teórico	18
Arroz (<i>Oryza sativa</i> L.)	18
Diversidad y Variabilidad Genética.	22
Marcadores Moleculares.....	23
Parámetros para la medición de variabilidad genética	27
Hipótesis	28
Capítulo 2: Materiales y Métodos	28
Participantes	28
Zona de Estudio	29
Duración de la Investigación.....	29
Metodología	29
Material Vegetal.....	29

Siembra y Recolección del Material Vegetal	30
Extracción y Cuantificación de ADN.....	30
Dilución y Amplificación de Muestras	31
Pruebas de Polimorfismo, Multiplexaje y Selección de Cebadores	32
Genotipaje de las Muestras con Tecnología M13-Tailing	34
Análisis de Diversidad Genética.....	35
Análisis de Estructura Genética	35
Análisis Molecular de Varianza y Estadística F	36
Identificación de Genotipos Duplicados	36
Capítulo 3: Resultados.....	37
Extracción y Cuantificación de ADN	37
Pruebas de Polimorfismo, Multiplexaje y Selección de Cebadores.....	37
Genotipaje de las Muestras con Tecnología M13-Tailing	39
Diversidad Genética	40
Análisis de Estructura Genética.....	44
Análisis Molecular de Varianza y Estadística F.....	54
Identificación de Genotipos Duplicados.....	57
Capítulo 4: Discusión	62
Análisis de Diversidad Genética	62
Análisis de Estructura Génica.....	63
Análisis Molecular de Varianza y Estadística F.....	66
Identificación de Genotipos Duplicados.....	66

Capítulo 5: Conclusiones	69
Capítulo 6: Recomendaciones	70
Capítulo 7: Bibliografía.....	71
Capítulo 8: Anexos.....	83

Índice de Tablas

Tabla 1. <i>Clasificación de marcadores moleculares</i>	24
Tabla 2. <i>Combinaciones múltiplex</i>	34
Tabla 3. <i>Alelos encontrados en cada locus</i>	41
Tabla 4. <i>Parámetros de diversidad genética para 81 variedades de arroz en 17 loci.</i>	43
Tabla 5. <i>Análisis Molecular de Varianza AMOVA para muestras de arroz.</i>	54
Tabla 6. <i>Estadísticos F de Wright estimados para los 17 loci.</i>	56
Tabla 7. <i>Identificación de Genotipos duplicados.</i>	57
Tabla 8. <i>Agrupamientos formados por genotipos duplicados.</i>	58
Tabla 9. <i>Muestras de campo sin identificación.</i>	59
Tabla 10. <i>Porcentaje de identidad entre variedades provenientes del programa de arroz.</i>	61

Índice de Figuras

Figura 1. <i>Morfología de Oryza sativa</i>	18
Figura 2. <i>Línea de tiempo de desarrollo del arroz</i>	19
Figura 3. <i>Prueba de polimorfismo para 32 cebadores</i>	38
Figura 4. <i>Combinaciones tríplex y dúplex obtenidas</i>	39
Figura 5. <i>Gel obtenido mediante el programa SAGA-GT</i>	40
Figura 6. <i>Valores de k vs Δk resultado del análisis de 17 loci para las muestras</i>	45
Figura 7. <i>Asignación de las muestras a las dos poblaciones determinadas</i>	46
Figura 8. <i>Dendograma UPGMA generado mediante la matriz de disimilaridad</i>	48
Figura 9. <i>Dendograma NJ para las muestras en estudio</i>	51
Figura 10. <i>Análisis de Coordenadas Principales (PCoA)</i>	53
Figura 11. <i>Porcentaje de varianza molecular</i>	55
Figura 12. <i>Análisis de redes para los genotipos duplicados de las muestras en estudio</i>	60

Listado de Abreviaturas

ADN	Ácido Desoxirribonucleico
AMOVA	Análisis Molecular de Varianza
PCR	Reacción en Cadena de la Polimerasa (siglas en inglés)
PIC	Contenido de Información Polimórfica
UPGMA	Unweighted pair group method with arithmetic mean
PCoA	Principal Coordinates Analysis
mL	Mililitros
µL	Microlitros
NJ	Neighbor Joining

Resumen

La provincia de Loja lidera el ranking nacional de rendimiento con un valor promedio de 8.96 t/ha por lo que es de importancia identificar las variedades cultivadas en este sector. Este estudio tiene como objetivo la caracterización molecular de variedades de arroz cultivadas en los cantones de Zapotillo y Macará de la provincia de Loja, utilizando marcadores microsatélites (SSRs) para lo cual se utilizó 20 variedades provenientes del programa de arroz de INIAP como referenciales y 61 materiales recolectados en la zona de estudio.

Se estableció un panel de 17 marcadores con un total de 75 alelos y un promedio de 4 alelos/locus que permiten identificar las variedades de arroz provenientes del programa de arroz y de campo. Se realizó un análisis de diversidad genética en el que se encontró que existe variabilidad entre las variedades estudiadas pese a la naturaleza endogámica del arroz de acuerdo a los valores PIC promedio, el análisis molecular de varianza y los valores obtenidos de los estadísticos F de Wright FIS, FIT y FST; el análisis de estructura genética dividió a los individuos en estudio en dos poblaciones, se obtuvieron 127 genotipos multilocus de los 226 totales.

Por último, se identificaron genotipos duplicados en donde el 18.90% corresponden a la variedad FED-60 y es probable que el 0.79% correspondan a la variedad SFL-09 debido a que comparten 26 de los 34 alelos analizados, se encontraron también 8 agrupaciones y no se logró identificar 53 de los 127 genotipos multilocus.

Palabras Clave

- **ARROZ**
- **DIVERSIDAD GENÉTICA**
- **GENOTIPAJE**

Abstract

The province of Loja leads the national yield ranking with an average value of 8.96 t / ha, so it is important to identify the varieties grown in this sector. The objective of this study is the molecular characterization of rice varieties grown in the cantons of Zapotillo and Macará in Loja's province, using microsatellite markers (SSRs) for which 20 varieties from the INIAP rice program were used as references and 61 materials collected in the study area.

A panel of 17 markers was established with a total of 75 alleles and an average of 4 alleles / locus that allow identifying the rice varieties from field and the rice program. A genetic diversity analysis was carried out in which it was found that there is variability between the studied varieties despite the inbred nature of the rice according to the average PIC values, the molecular analysis of variance and the values obtained from the Wright FIS F statistics. , FIT and FST; The genetic structure analysis divided the study individuals into two populations, 127 multilocus genotypes were obtained out of the 226 total.

Finally, duplicate genotypes were identified where 18.90% correspond to the FED-60 variety and it is probable that 0.79% correspond to the SFL-09 variety because they share 26 of the 34 analyzed alleles, 8 clusters were also found and 53 of the 127 multilocus genotypes could not be identified.

Key Words

- **RICE**
- **GENETIC DIVERSITY**
- **GENOTYPING**

Capítulo 1: Introducción

Planteamiento del problema

En el Ecuador el sector agrícola aporta con el 8% a la producción total anual del país (Producto Interno Bruto) (Ministerio de Agricultura y Ganadería, 2019), este sector es una fuente importante de trabajo además de ser un pilar fundamental en la economía. A nivel nacional, 5.11 millones de hectáreas de tierra se encuentran bajo labor agropecuaria, de las cuales 849355 hectáreas fueron destinadas a cultivos de ciclo corto o transitorios, principalmente: maíz duro seco, arroz en cáscara y papa. (Instituto Nacional de Estadísticas y Censos , 2019).

El arroz es considerado uno de los principales rubros dentro del sector, al contribuir con el 4% del Valor Agregado Bruto (VAB) Agropecuario (Banco Central del Ecuador, 2019), con una superficie sembrada total de 261770 hectáreas y una producción total anual de 1668523 toneladas.

En el país, existen 18 variedades de arroz liberadas por el Instituto Nacional de Investigaciones Agropecuarias (INIAP), y aproximadamente 11 variedades liberadas por la empresa privada, de ellas, las principales variedades que se sembraron en el 2019 fueron: SFL-11 (59%), Feron (12%) e INIAP 11 (4%) (García, 2019). La diversidad genética entre las variedades de arroz, provee la materia prima para la adaptación de las plantas a los cambios y la obtención de nuevas variedades, mientras que la pérdida de diversidad genética en las plantas de cultivo, generalmente debido a la selección antropogénica de unos pocos rasgos de interés, reduce la capacidad de las poblaciones de adaptarse a lo largo del tiempo (Rogers & McGuire, 2015). Por lo tanto, es necesario desarrollar información sobre la variabilidad genética de las variedades del país para asegurar su correcto manejo.

Justificación del problema

La producción nacional de arroz en el año 2019, presentó una disminución del 3% en relación al año 2018, como consecuencia de una reducción del 7% de la superficie cosechada a nivel nacional (García, 2019). A pesar de esto, el rendimiento nacional del cultivo de arroz en cáscara, durante el 2019 fue de 5.78 t/ha, lo que significa un aumento de 0.21 t/ha con respecto al año anterior (Ministerio de Agricultura y Ganadería, 2019).

La provincia de Loja lidera el ranking de rendimiento a nivel nacional con 11.4 t/ha a pesar de participar únicamente con el 2.10% de la producción total del país (Ministerio de Agricultura y Ganadería, 2019). La identificación de las variedades sembradas en la zona de estudio a través de marcadores moleculares SSRs, permitirá la selección de nuevos parentales con características de alto rendimiento para el programa de mejoramiento de arroz, además de identificar la variabilidad y relación genética entre los diferentes individuos.

Objetivos

Objetivo general

Caracterizar molecularmente variedades de arroz cultivadas en los cantones de Zapotillo y Macará de la provincia de Loja, utilizando marcadores microsatélites (SSRs).

Objetivos específicos

- Caracterizar variedades de arroz con tecnología de genotipaje M13-Tailing a través de marcadores microsatélites SSRs.
- Realizar un análisis de diversidad y estructura génica del material en estudio.
- Determinar la relación genética de cada variedad y su parentesco mediante un árbol filogenético.

Marco Teórico

Arroz (*Oryza sativa* L.)

Taxonomía, Morfología y Desarrollo. El arroz es una planta perteneciente al subtipo Angiospermas, clase Monocotiledóneas, género *Oryza* y a la tribu Oryzaceae de la familia Gramineae (Poaceae) (Degiovanni, Berrío, & Charry, 2010). El género *Oryza* contiene 25 especies reconocidas, de las cuales 23 son especies silvestres y dos, *O. sativa* y *O. glaberrima* son cultivadas (Sanchez, Wing, & Brar, 2013).

Oryza sativa es una gramínea anual, de tallos redondos y huecos compuestos por nudos y entrenudos, hojas de lámina plana, lineal y larga, unidas al tallo por la vaina de forma alternada y su inflorescencia es en panícula como se muestra en la **Figura 1**. El tamaño de la planta varía entre 0.4 m (enana) hasta más de 7 m (flotantes), la altura de la planta de arroz es una función de la longitud y número de los entrenudos. Estas dos características, son caracteres varietales definidos (Centro Internacional de Agricultura Tropical CIAT, 2005).

Figura 1.

Morfología de Oryza sativa.

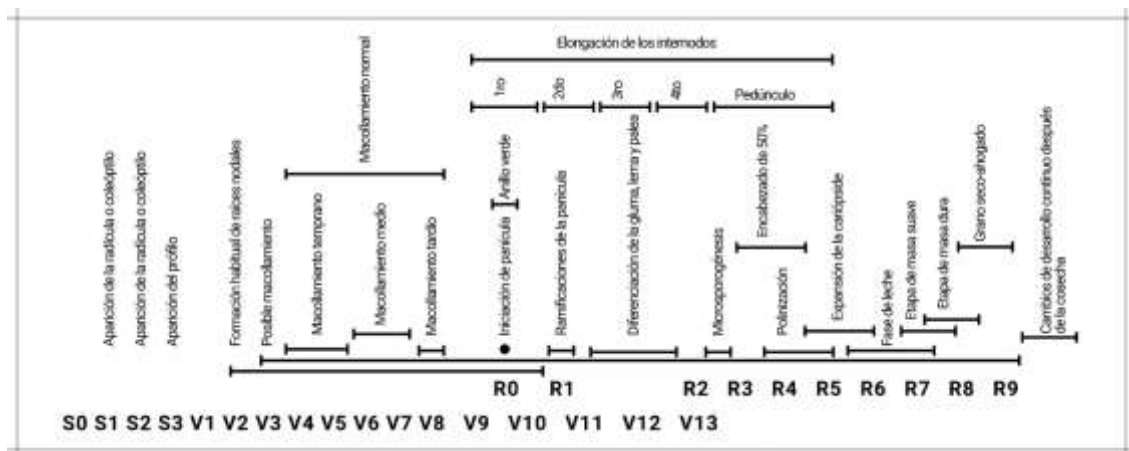


Nota. Adaptado de Köhler's Medizinal-Pflanzen, de Hermann Adolph Koehler, (1914).

Counce, Keisling y Mitchell (2000), propusieron un sistema para expresar el desarrollo morfológico de las plantas de arroz que consta de 3 fases. La **Figura 2**, muestra la línea de tiempo del desarrollo de las plantas de arroz propuesta por los autores.

Figura 2.

Línea de tiempo de desarrollo del arroz.



Nota. Fases de desarrollo morfológico de plantas de arroz: plántula (**S**, por su nombre en inglés), vegetativa (**V**) y reproductiva (**R**). Adaptado de *A Uniform, Objective, and Adaptive System for Expressing Rice Development*, de Counce, Keisling y Mitchell, 2000.

La fase S normalmente ocurre en 4 etapas de crecimiento (S0 a S3) que comprenden la aparición y desarrollo de radícula, coleóptilo y la primera hoja rudimentaria denominada prófido. La fase V consiste en las etapas V1 a VN, donde N es el número final de hojas con collares en el tallo principal. Esta fase termina con el desarrollo de la última hoja u hoja bandera.

La fase R consta de 10 etapas de crecimiento, las etapas reproductivas tempranas (R0 a R2) ocurren simultáneamente con las etapas vegetativas tardías. Las etapas entre R0 a R3 se refieren al desarrollo de la panícula, entre R4 a R8 al florecimiento y llenado de grano individual de cada panícula y la etapa R9 se refiere a la maduración y secado de los granos.

Importancia. El arroz es un alimento básico para más de la mitad de la población mundial y es el segundo cereal más cultivado después del maíz (Awika, 2011), con un consumo alrededor de 486.62 millones de toneladas métricas en todo el mundo durante el año de cosecha 2018/2019 (Shahbandeh, 2021). Se estima que la producción mundial alcanzará las 582 millones de toneladas y que el consumo aumente en 29 millones de toneladas en el 2029 (Organisation for Economic Co-operation Development (OECD) and Food and Agriculture Organization (FAO), 2020). India y China concentran más de la mitad de la producción mundial de arroz con 43.79 millones y 30.18 millones de hectáreas destinadas al cultivo respectivamente (Shahbandeh, 2021).

En América Latina y el Caribe el arroz es el cuarto alimento más consumido contribuyendo con el 11% de la ingesta calórica per cápita de la región y alrededor de 5.3 millones de hectáreas cultivadas, en su mayoría por pequeños productores (FONTAGRO, 2020).

A nivel nacional, el último informe de rendimiento proporcionado por el Ministerio de Agricultura y Ganadería, menciona que para arroz, en el 2019, se destinó una superficie sembrada total de 261770 hectáreas y se obtuvo una producción total anual de 1668523 toneladas (García, 2019).

Variedades Genéticas. Se cree que el proceso de domesticación del arroz se remonta aproximadamente a entre 8000-9000 años atrás, a partir de dos progenitores salvajes *O. nivara* y *O. rufipogon* (Chen, Huang, Tian, Wing, & Han, 2019), (Sang & Ge, 2007). Este proceso implicó la adaptación de los parentales a condiciones geográficas y ecológicas distintas, dando origen a dos subespecies, japónica e indica (Chen, Huang, Tian, Wing, & Han, 2019) que son fenotípica y ecológicamente divergentes.

Estudios moleculares de repeticiones de secuencia simple (SSRs) y polimorfismos de un solo nucleótido (SNPs) han permitido la clasificación de estas dos subespecies en 5 grupos, *índica* en *índica* y *aus* y *japónica* en *aromática*, *japónica tropical* y *japónica templada* (Muto, y otros, 2019), (Kishor, Seo, Chin, & Koh, 2020).

El proyecto 3000 Genomas de Arroz (3K-RG, por sus siglas en inglés), desarrollado por la Academia China de Ciencias Agrícolas, BGI-Shenzhen y el Instituto Internacional de Investigación del Arroz, permitió el análisis de 3000 genomas de varias accesiones de arroz de la Colección del Banco Internacional de Genética de Arroz (IRGC) en el Instituto Internacional de Investigación del Arroz (IRRI) y del Banco Nacional de Genética de Cultivos de China (CNCGB). Las 3000 accesiones muestreadas procedían de 89 países/regiones diferentes, de las cuales el 77.1% pertenecían a los centros de diversidad genética de arroz: Asia sudoriental (33.9%), Asia meridional (25.6%) y China (17.6%) (The 3,000 rice genomes project, 2014). Este proyecto permitió corroborar la clasificación en los 5 grupos previamente descritos, además de sugerir varias subpoblaciones no reportadas que se correlacionan con la ubicación geográfica (Wang, Mauleon, & Leung, 2018).

La gran importancia económica y alimentaria de esta especie, a nivel mundial, ha provocado una amplia diversidad de variedades genéticas específicas para diferentes localidades, por lo que no se puede medir con exactitud su diversidad.

Diversidad y Variabilidad Genética.

La diversidad genética es la variabilidad existente en una especie o entre especies (Nevo, 2001), originada por mutaciones espontáneas por errores durante la replicación del ADN (Ellegren & Galtier, 2016), y el sustrato para la evolución considerando las condiciones ambientales bióticas y abióticas cambiantes (Çalışkan, 2012), (FAO, 2008).

Desde los inicios de la agricultura, la variabilidad genética ha permitido desarrollar cultivares nuevos y mejorados con características deseables (Bhandari, y otros, 2017). Estas características incluyen aquellas deseadas por los productores, como por ejemplo un mayor rendimiento y aquellas deseadas por los fitomejoradores como la resistencia a plagas o enfermedades (Govindaraj, Vetriventhan, & Srinivasan, 2015), sin embargo, el desarrollo de cultivos que cubren con todas las demandas actuales ha conllevado a una progresiva erosión genética.

La erosión genética es la pérdida de la diversidad genética en las plantas de cultivo, generalmente acelerada por actividad antropogénica (Rogers & McGuire, 2015), debido a tres factores: prácticas de reproducción asimétricas que se centran en la mejora de solo unos pocos rasgos, el uso frecuente de pocos genotipos seleccionados como progenitores en el programa de desarrollo varietal y la introducción de pocas líneas destacadas en muchos países, lo que conduce a una mayor similitud genética entre cultivares modernos (Bhandari, y otros, 2017).

La pérdida de la diversidad genética reduce la habilidad de las poblaciones a adaptarse a lo largo del tiempo, reduce el potencial evolutivo, la aptitud reproductiva y la capacidad de generar nuevas variedades mejoradas (Rogers & McGuire, 2015). El reto actual para la ciencia, es el desarrollo de procesos de fitomejoramiento que identifiquen e incorporen eficientemente la diversidad genética de distintas fuentes fitogenéticas incluidas las variedades actualmente cultivadas, las variedades recientemente desarrollados, las variedades locales, los parientes silvestres y cercanos de las variedades cultivados y las colecciones de germoplasma con plantas de élite y / o mutantes (Swarup, y otros, 2020) para la conservación de la diversidad genética en las plantas de cultivo.

Existen dos tipos principales de variabilidad genética, aquella que tiene una influencia directa en el éxito reproductivo y el fenotipo del individuo, en donde la selección natural tiene efecto, denominada variabilidad adaptativa o selectiva (Holderegger, Kamm, & Gugerli, 2006) y aquella que es producto de sistemas de mutación, migración, deriva génica y apareamiento en lugar de la selección natural, denominada variabilidad genética neutra (Neale & Wheeler, 2019). Es necesario añadir que la mayoría de las técnicas moleculares desarrolladas en la actualidad se utilizan para medir variabilidad genética neutra (Bhandari, y otros, 2017).

Marcadores Moleculares

En las ciencias vegetales, los marcadores moleculares son utilizados para estudios de diversidad genética, evolución y filogenia, identificación de ploidias, identificación y retrocruzamiento de genes de interés, mapeo genético, mejoramiento, entre otras aplicaciones (Muhammad Azhar, y otros, 2017).

Los marcadores moleculares son secciones del genoma o fragmentos que contienen variaciones (Adhikari, y otros, 2017). Estas variaciones se utilizan para reconocer el polimorfismo entre diferentes genotipos o alelos de un gen en una población o grupo de genes (Karlik & Tombulođlu, 2016). Los marcadores de ADN pueden asociarse con fragmentos de ADN particulares dentro del gen de interés o estar vinculados a un rasgo.

Según Adhikari y colaboradores (2017), un marcador molecular ideal se caracteriza por: encontrarse en todo el genoma, altos niveles de polimorfismo, codominancia, alelos con características diferentes y fáciles de reconocer y efectos pleiotrópicos o espistáticos reducidos o nulos.

Tipos de Marcadores. Los marcadores moleculares se pueden clasificar por su dominancia (dominante o codominante), método de detección (hibridación o PCR) y su modo de transmisión (biparental, organular materno o paterno, etc.) (Semagn, Bjørnstad, & Ndjiondjop, 2006).

La **tabla 1** muestra los distintos tipos de marcadores moleculares propuesto por (Karlik & Tombulođlu, 2016):

Tabla 1.

Clasificación de marcadores moleculares

	Marcador	Acrónimo
Marcadores Clásicos	Fragmentos de restricción de longitud polimórfica	RFLP
	Fragmentos polimórficos amplificados al azar	RAPD
	Polimorfismos en la longitud de fragmentos amplificados	AFLP

	Marcador	Acrónimo
	Microsatélites	SSRs/ STRs/ SSLPs
	Polimorfismos de nucleótido único	SNPs
Marcadores de alto rendimiento	Tecnología de matrices de diversidad	DArT
	Marcadores de matriz etiquetada	TAMs
	Marcadores ADN asociados a sitios de restricción	RAD
	PCR específico alelo competitivo	KASpar
	Variaciones en el número de copia	CNV
	Variaciones de presencia- ausencia	PAVs

Nota. Adaptado de *Molecular Markers and Their Applications*, de Karlik y colaboradores, 2016.

Marcadores Microsatélites SSRs. Los marcadores microsatélites, también conocidos como repeticiones de secuencia simple (SSRs), repeticiones cortas en tándem (STRs) o polimorfismos de longitud de secuencia simple (SSLPs), son motivos repetidos en tándem que constan de 1 a 6 nucleótidos ampliamente distribuidos en el genoma (Miah, y otros, 2013). El polimorfismo de los SSRs está determinado por la longitud de la secuencia o el número de unidades repetidas en los alelos del genoma de individuos de una misma especie (Karlik & Tombuloğlu, 2016). Los microsatélites son marcadores basados en PCR, en donde las secuencias que flanquean los microsatélites se utilizan como cebadores ya que usualmente son fragmentos conservados (Henry, 2013).

La abundancia y el tipo de marcador microsatélite varía entre especies, en plantas, la unidad de repetición más abundante para microsatélites reportada es (AT)_n (Miah, y otros, 2013). Los SSRs, al ser marcadores que cumplen con las características ideales descritas anteriormente, ha sido ampliamente utilizados en

estudios de genotipaje, genética poblacional, conservación, bioseguridad y mejoramiento vegetal (Henry, 2013).

El genotipaje, es la técnica por la cual se detectan variaciones en el ADN entre individuos de una misma especie (Getachew, Bille, Bell, & Gebreselassie, 2019). El genotipaje determina diferencias en el material genético comparando una secuencia de ADN con la de otra muestra o secuencia de referencia. Identifica pequeñas variaciones en la secuencia genética dentro de las poblaciones como los microsatélites SSRs (Mason, 2015).

Tecnología M13-Tailing. La tecnología M13-Tailing es una técnica de genotipaje por fluorescencia que tiene como principio el uso de 3 cebadores para la amplificación PCR: 1) un cebador forward específico para la secuencia de interés que contiene una cola M13 en su extremo 5' (5'-CAC GAC GTT GTA AAA CGA C-3'); 2) un cebador reverse específico para la secuencia de interés; 3) y un cebador universal M13 de secuencia complementaria a la cola M13 del cebador forward marcado con fluorescencia (Schuelke, 2000) derivado de un vector bacteriano (Rodriguez Oblessucl, y otros, 2009). Durante la PCR, las condiciones de amplificación son escogidas de modo que, durante los primeros ciclos, el cebador forward y reverse específicos se incorporan en los productos PCR. Cuando el cebador forward se agota por una concentración inicial baja (Blacket, Robi, Good, Lee, & Miller, 2012), la temperatura de alineamiento se reduce para facilitar la hibridación del cebador universal M13 y marcar los productos PCR.

El genotipaje por fluorescencia permite una recolección más amplia y rápida de información que los métodos tempranos basados en isotopos radioactivos y teñido por plata (Blacket, Robi, Good, Lee, & Miller, 2012). La utilización de un único cebador universal marcado con fluorescencia reduce costos al evitar el diseño de varios cebadores específicos marcados.

Parámetros para la medición de variabilidad genética

Frecuencia alélica. La frecuencia alélica es el término utilizado para describir la fracción de copias de un gen que pertenecen a un alelo particular en una población definida (Silver, 2001). La frecuencia alélica se puede calcular de forma empírica con la ecuación de Hardy-Weingber:

$$\text{Frecuencia del alelo } A = \frac{\text{número de copias del alelo } A \text{ en la población}}{\text{Número total de copias del gen en la población}}$$

$$p^2 + 2pq + q^2 = 1$$

Donde p y q son la frecuencia de los alelos, p^2 y q^2 representan a los individuos homocigotos y $2pq$ a los individuos heterocigotos.

Heterocigosidad observada y esperada. La heterocigosidad es la presencia de alelos diferentes para un determinado gen. La heterocigosidad es un parámetro calculado usualmente para estimar el grado de variación genética de una población (Sbordoni, Allegrucci, & Cesaroni, 2012).

El nivel observado de heterocigosidad (H_o) es la proporción de loci que son genéticamente heterocigotos y mide la combinación de alelos en genotipos diploides. El nivel esperado de heterocigosidad (H_e) es el nivel de heterocigosidad que se espera del apareamiento aleatorio dado un conjunto de frecuencias alélicas (p_i). Esta heterocigosidad esperada es una medida de la diversidad genética conocida como el índice Nei de diversidad genética (Jarvis, y otros, 2016).

Contenido de información Polimórfica (PIC). El contenido de información polimórfica de un marcador, corresponde a su habilidad para detectar el polimorfismo entre los individuos de una población (Serrote, Reiniger, & Rabaiolli, 2019). El PIC para n alelos, puede ser calculado con la fórmula:

$$PIC = 1 - \sum_{i=1}^n p_i^2 - \sum_{i=1}^{n-1} \sum_{j=i+1}^n 2p_i^2 p_j^2$$

Donde n es el número de alelos, p_i y p_j son la frecuencia alélica de las poblaciones i y j (Guo & Elston, 1998).

Hipótesis

Ho: No existen diferencias genéticas entre las variedades de arroz sembradas en los cantones de Zapotillo y Macará de la provincia de Loja.

Hi: Existen diferencias genéticas entre las variedades de arroz sembradas en los cantones de Zapotillo y Macará de la provincia de Loja

Capítulo 2: Materiales y Métodos

Participantes

La presente investigación ha sido elaborada por Fabian Andres Reyes Salinas egresado de la Carrera de Ingeniería en Biotecnología, bajo la tutoría de la Dra. Valeria Ochoa PhD. Docente-Investigador de la Universidad de las Fuerzas Armadas ESPE y la asesoría del Dr. Eduardo Morillo PhD. Investigador Principal y responsable del área de Biotecnología del Instituto Nacional de Investigaciones Agropecuarias (INIAP) y la Dra. Elisa Quiala PhD. Investigadora del INIAP. El financiamiento para la investigación estuvo a cargo del Departamento Nacional de Biotecnología del INIAP.

Zona de Estudio

El trabajo de investigación se realizó en el laboratorio de Biología Molecular del Departamento Nacional de Biotecnología de la Estación Experimental Litoral Sur ubicado en el Km. 26 vía Durán – Tambo, Cantón Yaguachi, Provincia del Guayas, latitud 2°15'29"S, longitud 79°38'39"O y en el laboratorio de Biología Molecular del Departamento Nacional de Biotecnología de la Estación Experimental Santa Catalina ubicado en la Panamericana Sur, km 1 – Vía a Tambillo, Cantón Mejía, Provincia de Pichincha, latitud 2°15'24"S, longitud 79°38'40"O.

Duración de la Investigación

El tiempo de duración de la investigación fue de aproximadamente 9 meses. Se inicio en noviembre de 2020 y culminó en agosto de 2021.

Metodología

Material Vegetal

Se utilizaron un total de 246 muestras en la investigación, de las cuales 186 (61 variedades, 3 repeticiones por variedad) provienen de los cantones Zapotillo y Macará, provincia de Loja y 60 muestras (20 variedades genéticas puras, 3 repeticiones por variedad) provienen del programa de arroz de la estación Litoral Sur del Instituto Nacional de Investigaciones Agropecuarias (INIAP).

Las semillas de arroz fueron proporcionadas por el Ministerio de Agricultura y Ganadería (MAG) recolectadas de los productores en la zona de estudio y las variedades mejoradas: INIAP IMPACTO, INIAP FL-ELITE, FL-Arenillas, INIAP FL-1480, SFL-11, INIAP415, INIAP10, SFL12, PUNGSAM, INIAP12, INIAP14, INIAP15, INIAP16, INIAP17, INIAP18, SFL09, F-50, Feron, Tinajones y TAEABUEG fueron proporcionadas por el programa de arroz del INIAP. El listado total de las muestras para la investigación se encuentra en el **Anexo 1**.

Siembra y Recolección del Material Vegetal

Para la pre-germinación del material se realizaron 3 repeticiones, utilizando 105 semillas en total para cada una de las variedades. Se colocaron 35 semillas en cada caja Petri con un disco de papel filtro como base. A continuación, se humedeció el papel filtro con agua destilada y se mantuvo a condiciones ambientales durante 5 días.

El sustrato utilizado fue tierra de río previamente pasteurizada a 60-70 °C que se colocó en macetas, se humedeció el sustrato con agua de riego hasta llegar a nivel de fangoso y se sembró 10 semillas en cada maceta por cada caja Petri con semillas pre-germinadas (un total de 3 macetas con 10 semillas por variedad). Se cultivaron las semillas a condiciones ambientales.

La toma de muestras del material fue realizada cuando el arroz alcanzó la quinta hoja (V5), alrededor de los 20 días después de la siembra.

Extracción y Cuantificación de ADN

Se realizó la extracción de ADN a partir de tejido foliar de una planta por cada maceta, utilizando el protocolo descrito por Ferreira y Grattapaglia (1998).

En un tubo Eppendorf de 1.5 mL se agregó 0.05 g de tejido fresco del material y 700 μ L de tampón de extracción 2X CTAB constituido de 2g de CTAB, 10 mL de Tris HCl 1M pH 8, 8.12g de NaCl, 4 mL de EDTA 0.5M, 1g de PVP, 200 μ L de β -mercaptoetanol en un volumen final de 100 mL y se trituró la muestra con pistilos previamente esterilizados. A continuación, se incubó las muestras a 65°C por 1 hora agitando cada 30 minutos para luego centrifugar a 13200 r.p.m. por 15 minutos.

Se tomó el sobrenadante y se añadió 600 μ L de una solución cloroformo-alcohol isoamílico en proporción 24:1 denominada CIA, se homogeneizó las muestras y se centrifugó a 13000 r.p.m. por 5 minutos, se tomó el sobrenadante y se

repitió el proceso con la solución CIA. Luego, se añadió 400 μL de isopropanol, se incubó a -20°C durante 30 minutos y se centrifugó por 4 minutos a 13000 r.p.m.

Se retiró el isopropanol y se realizó dos lavados con 1 mL de etanol al 70% al ADN precipitado, después se secó en microestufa por 30 minutos a 37°C para luego resuspender el ADN en 100 μL de tampón TE. Por último, se añadió 3 μL de RNAsas y se incubó por 30 minutos a 37°C . Las muestras de ADN se mantuvieron a -20°C hasta su uso.

Para la cuantificación del ADN se utilizó el espectrofotómetro de microplacas Epoch de BioTek Instruments© para medir la absorbancia a longitudes de onda de 260 y 280 nm y cuantificar la concentración de ADN en $\text{ng}/\mu\text{L}$. Se obtuvo la relación 260/280 para determinar la pureza del ADN extraído y se corrió las muestras en gel de agarosa al 1% y 100 voltios por 30 min para observar la integridad del ADN.

Dilución y Amplificación de Muestras

Una vez cuantificada la concentración de ADN de las muestras, se procedió a realizar una dilución a una concentración de 10 $\text{ng}/\mu\text{L}$ en agua destilada ultrapura, a continuación, se determinó si las muestras amplificaban mediante PCR. El master mix para la PCR contiene 2.18 μL de agua ultra pura, 1.50 μL de buffer PCR 1X, 0.60 μL de MgCl_2 2 mM, 0.38 μL de dNTP's 0.253 mM, 0.38 μL de cada cebador a concentración 0.5 μM y 0.10 μL de taq polimerasa a concentración 0.067 U/mL, por reacción se dispensa un volumen final de 5.5 μL de master mix y 2 μL de muestra de ADN.

El programa del termociclador consta de una desnaturalización inicial a 94°C por 5 min, seguido de 30 ciclos de desnaturalización a 94°C por 45 segundos, alineamiento de los cebadores por 1 min a 55°C y una elongación a 72°C por 2 min, seguido de una elongación final por 7 min a 72°C . Los amplicones se

visualizaron en gel de agarosa al 2% mediante electroforesis a 100 voltios por 30 min.

Pruebas de Polimorfismo, Multiplexaje y Selección de Cebadores

Un set de 32 cebadores proporcionados por la Estación Experimental Santa Catalina del INIAP, basados en la publicación de Jiang, Zhang, Wang, Chen, & Xu (2010), se amplificaron mediante reacción en cadena de la polimerasa (PCR) para determinar si los cebadores son polimórficos o monomórficos en 4 muestras de ADN previamente amplificadas en gel de agarosa, con el propósito de contar con diversidad genética dentro de los ensayos de amplificación de los cebadores.

Se empleó el protocolo descrito por Morillo y Miño (2011) para la amplificación con marcadores universales M13-Tailing en el cual se añade el fragmento M13 (5'-CAC GAC GTT GTA AAA CGA C-3') a los cebadores forward. Durante la PCR, el marcador IRDye M13 se une a la cola M13 añadida a los cebadores seleccionados de los marcadores microsatélites. El master mix para la amplificación M13 consta de 0.32 μ L de agua ultrapura, 1 μ L de buffer de PCR 1x, 0.50 μ L de MgCl₂ 2.5 mM, 0.20 μ L de dNTP's 0.2 mM, 0.80 μ L del marcador M13 700/800 0.16 mM, 0.05 μ L de cebador F-M13 0.01 μ M, 0.08 μ L de cebador reverse 0.16 μ M, 0.05 μ L de taq polimerasa 0.05 U/ μ L y 2 μ L de muestra de ADN.

El programa para el termociclador consta de una desnaturalización inicial de 2 min a 94 °C, seguida de una desnaturalización secundaria a 95 °C por 4 min, 25 ciclos que constan de desnaturalización por 1 min a 95 °C, alineamiento por 2 min a la temperatura específica del cebador seleccionado y una elongación a 72 °C por 2 min y una elongación final a 72 °C por 10 min. La fluorescencia del marcador será detectada en gel de poliacrilamida casera al 3.5% por el analizador ADN-LICOR 4300S de LI-COR Biosciences®, a longitudes de onda de 700 y 800 nm. Los

cebadores utilizados fueron: RM499, RM151, RM9, RM246, PSM332, RM529, RM327, PSM122, RM208, PSM378, RM247, RM487, PSM132, PSM326, PSM363, RM430, RM50, RM418, RM70, RM44, RM223, PSM338, PSM406, PSM167, RM286, RM20B, RM229, PSM415, RM21, RM144, RM415, RM453. Con una temperatura de alineamiento de 55 °C para todos los cebadores.

Los cebadores RM208, RM247, PSM132, PSM332, PSM122, RM9, PSM326, RM50, RM286, RM20B, RM21, PSM167, RM229, PSM415, RM415, PSM363 y RM144 fueron escogidos para la conformación de las combinaciones múltiplex donde se tomaron en cuenta que la temperatura de alineamiento sea similar y el rango de amplificación sea diferente, las combinaciones realizadas se amplificaron con el protocolo de Morillo y Miño (2011).

El master mix para las pruebas de duplexaje M13 es el mismo descrito anteriormente para la amplificación M13-Tailing con la diferencia que el volumen de agua ultrapura es de 0.19 µL, mientras que para las pruebas de triplexaje M13, el master mix consta de 0.88 µL de agua ultrapura, 1.40 µL de buffer de PCR 1x, 0.70 µL de MgCl₂ 2.5 mM, 0.28 µL de dNTP's 0.2 mM, 1.12 µL del marcador M13 700/800 0.16 mM, 0.07 µL de cada cebador F-M13 0.01 µM, 0.11 µL de cada cebador reverse 0.16 µM, 0.05 µL de taq polimerasa 0.05 U/µL y 2 µL de muestra de ADN. El programa de amplificación para el termociclador es el mismo descrito anteriormente para ambos casos y la fluorescencia del marcador será detectada en gel de poliacrilamida casera al 3.5% por el analizador ADN-LICOR 4300S de LI-COR Biosciences®, a longitudes de onda de 700 y 800 nm.

Se obtuvo un total de 9 combinaciones, 1 triplex, 6 duplex y 2 monoplex que se muestran en la **Tabla 2**. Los pesos esperados, la secuencia de los cebadores y del producto de amplificación se encuentran en el **Anexo 2**.

Tabla 2.*Combinaciones múltiplex*

Combinación	Nombre Cebador
TRIPLEX 1	RM208
	RM247
	PSM132
DUPLEX 2	PSM332
	PSM122
DUPLEX 3	RM9
	PSM326
DUPLEX 4	RM50
	RM286
DUPLEX 5	RM20B
	RM21
DUPLEX 6	PSM167
	RM229
DUPLEX 7	PSM415
	RM415
MONOPLEX 8	PSM363
MONOPLEX 9	RM144

Nota. La temperatura de alineamiento para todos los cebadores fue de 55°C.

Genotipaje de las Muestras con Tecnología M13-Tailing

Para el genotipaje se realizó una amplificación de las muestras en placas para PCR con las combinaciones, master mix y programa del termociclador descritos anteriormente, a los productos de amplificación se añadió solución Blue Stop constituida de EDTA, formamida y azul de bromofenol, en una proporción 1:5, a continuación, se desnaturalizó las muestras a 95°C por 3 minutos y luego se incubó en hielo evitando el contacto con la luz. En el gel de poliacrilamida casera al 3.5% se dispensó 0.8 µL de cada muestra por pocillo y 0.5 µL del marcador de peso

molecular IRDye 700 y 800. Se realizó la corrida electroforética por 1 hora y media a 1500 voltios.

El análisis de la corrida se realizó en el software SAGA-GT para el análisis de microsatélites de LI-COR Biosciences® en el cual se ingresó los datos sobre las tallas del marcador de peso molecular, rango del tamaño esperado de a amplificación, marcaje, combinaciones de los loci SSR, nombres de las muestras y posición en las que fueron cargadas. Como resultado, se obtiene un gel digital en el que se marcan las bandas que representan a los alelos amplificados en cada muestra.

Por último, se genera una matriz de datos de dimensiones $n \times m$ siendo n el número de muestras y m el número de loci analizados, la información obtenida se sintetiza en una matriz general (**Anexo 3**) para su posterior análisis bioinformático.

Análisis de Diversidad Genética

Para el análisis de diversidad genética se utilizaron los paquetes informáticos Power Marker 3.25 (Liu & Muse, 2005) y GenAlex V6.5 (Peakall & Smouse, 2012). Se determinaron los siguientes parámetros de variabilidad: número de observaciones, frecuencias alélicas, número efectivo de alelos, heterocigosidad observada, heterocigosidad esperada (Nei, 1973) y contenido de información polimórfica (PIC).

Análisis de Estructura Genética

La asignación de genotipos a poblaciones predeterminadas se realizó con el software STRUCTURE vs. 2.3.4 desarrollado por Pritchard y colaboradores (2000) con una duración del período de quemado de 1000, un número de repeticiones de MCMC después del quemado de 10000, 10 iteraciones y el modelo *Admixture*.

Se calculó una matriz de disimilaridad a través del método de emparejamiento simple SM (por sus siglas en inglés) y 1000 iteraciones como

bootstrapping, a continuación, se realizó un dendograma con el método UPGMA y un dendograma con el método del vecino más cercano (Neighbor Joining) para visualizar las poblaciones obtenidas por el método bayesiano con el programa DARWIN 6.0.21 (Perrier & Jacquemoud-Collet, 2006).

Con los datos obtenidos se realizó un análisis de coordenadas principales (ACoP) y análisis de conglomerados (AC) en el programa GenAlex V6.5 (Peakall & Smouse, 2012).

Análisis Molecular de Varianza y Estadística F

Con la información obtenida en el análisis de estructura genética se procedió a realizar un análisis molecular de varianza (AMOVA) (Excoffier, Smouse, & Quattro, 1992), para determinar la diferenciación genética de los grupos formados y las estadísticas F, para este propósito se utilizó GenAlex V6.5 (Peakall & Smouse, 2012).

Identificación de Genotipos Duplicados

Se realizó la identificación de genotipos duplicados con ayuda del complemento de Microsoft Excel Microsatellites Toolkit (Park, 2000) con un número de alelos no idénticos tolerados de 1 antes de rechazar un emparejamiento.

Capítulo 3: Resultados

Extracción y Cuantificación de ADN

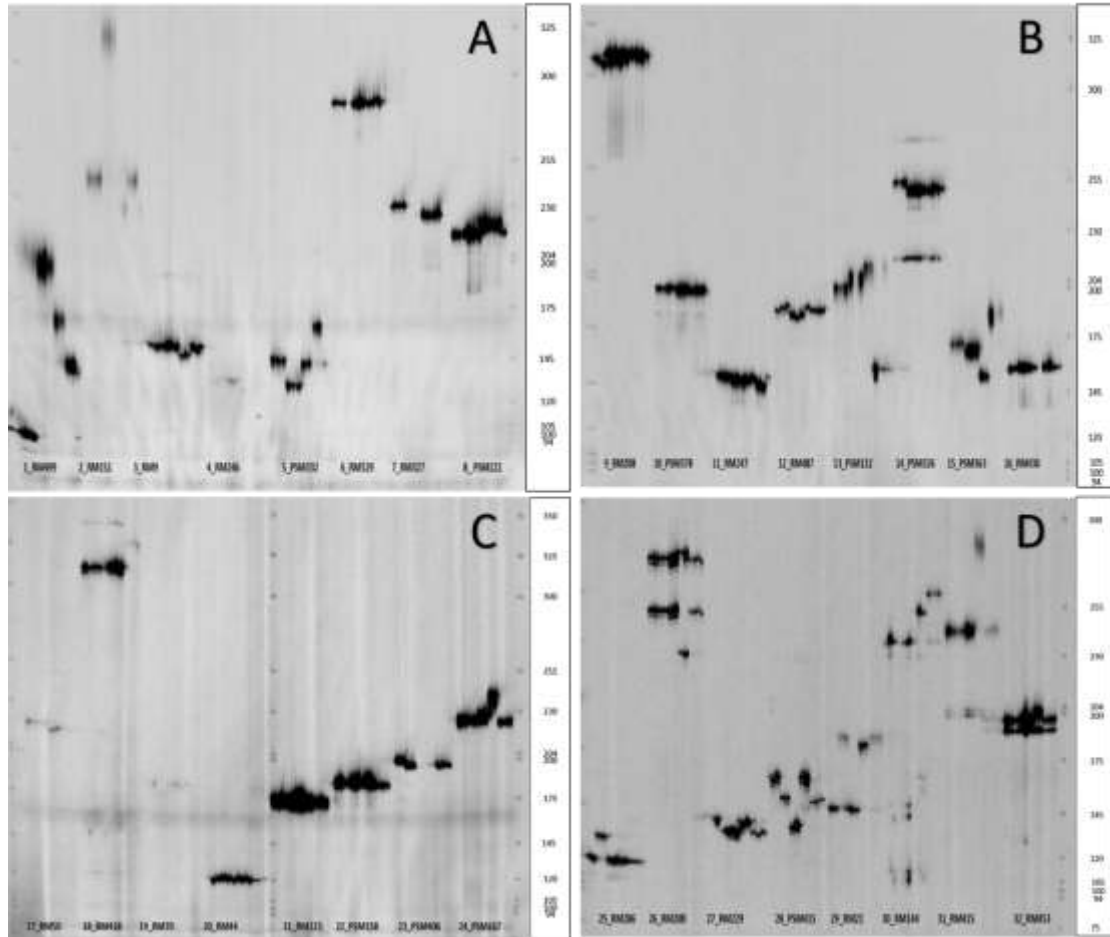
Se obtuvieron concentraciones de ADN en un rango de 30 a 2000 ng/ μ L con un ratio 260/280 que oscilaba entre 1.7 y 2.

Pruebas de Polimorfismo, Multiplexaje y Selección de Cebadores

En la **Figura 3** se observa los productos de la amplificación para las pruebas de polimorfismo de los 32 cebadores proporcionados. Los cebadores RM208, RM247, PSM132, PSM332, PSM122, RM9, PSM326, RM50, RM286, RM20B, RM21, PSM167, RM229, PSM415, RM415, PSM363 y RM144 fueron escogidos para la conformación de las combinaciones múltiplex, debido a su temperatura de alineamiento similar y su rango de amplificación diferente.

Figura 3.

Prueba de polimorfismo para 32 cebadores.

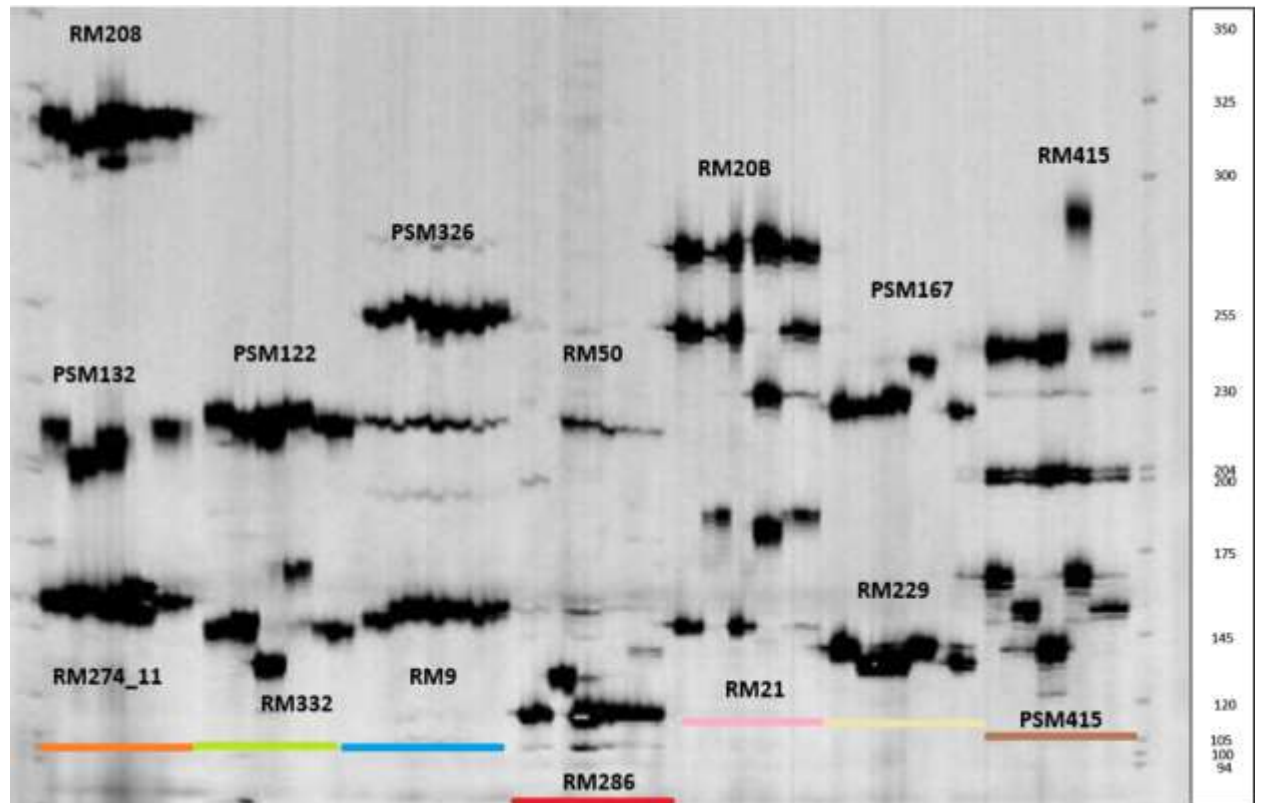


Nota. En las imágenes A, B, C Y D se pueden observar los alelos amplificados de 32 cebadores para 4 variedades de arroz en gel de poliacrilamida casera al 3.5% y marcador molecular IRDye de 350 pb. Se seleccionaron los marcadores que mostraban dos o más bandas de diferente tamaño como por ejemplo el cebador PSM415 (Figura 3D, 4 cebador) para los posteriores análisis.

En la prueba de multiplexaje se obtuvo un total de 9 combinaciones, 1 tríplex, 6 dúplex y 2 monoplex que se muestran en la **Figura 4**.

Figura 4.

Combinaciones tríplex y dúplex obtenidas.



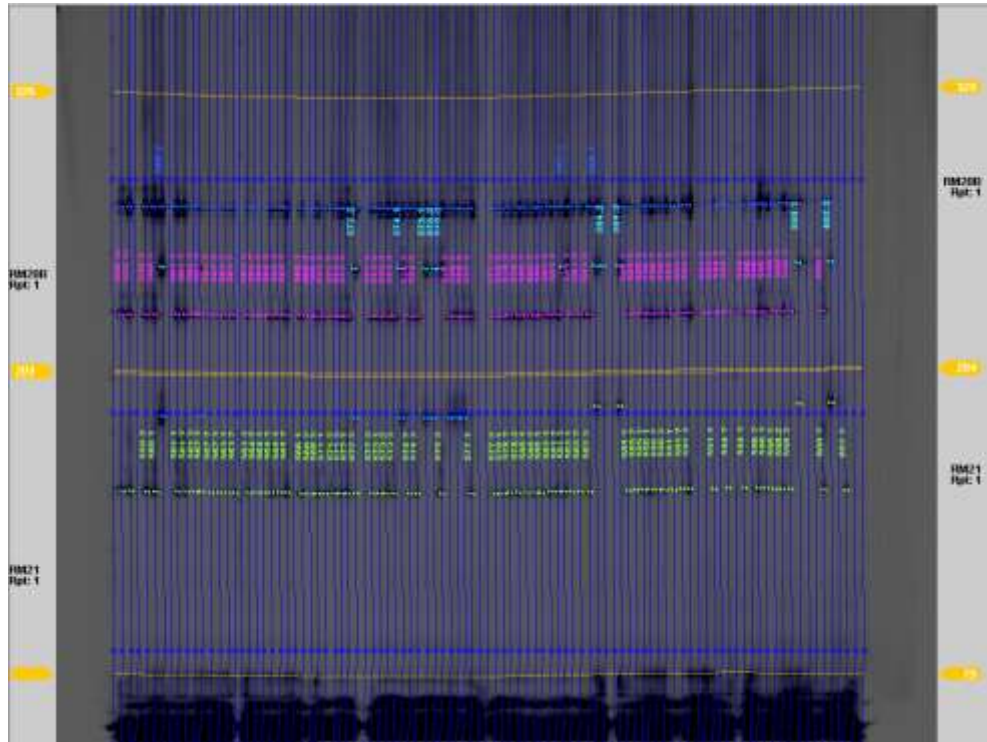
Nota. En la imagen se puede observar los productos de amplificación múltiple para 7 combinaciones. De izquierda a derecha: 1 tríplex y 6 dúplex, para 11 muestras de arroz en gel de poliacrilamida casera al 3.5% y marcador molecular IRDye de 350 pb, todos los marcadores poseen una temperatura de alineamiento de 55 °C y un rango de amplificación diferente, condiciones necesarias para las combinaciones multiplex.

Genotipaje de las Muestras con Tecnología M13-Tailing

Un total de 27 placas de 96 pocillos para PCR fueron amplificadas con la tecnología M13-Tailing, se obtuvo el mismo número de geles digitales que placas amplificadas en el software SAGA-GT para el análisis de microsatélites de LI-COR Biosciences®, en los que se marcaron las bandas que representan a los alelos amplificados en cada muestra. La **Figura 5** muestra un ejemplo de los geles digitales obtenidos.

Figura 5.

Gel obtenido mediante el programa SAGA-GT



Nota. En la imagen se puede observar un ejemplo de los gels digitales obtenido mediante el programa SAGA-GT, este es el gel para la combinación dúplex de los cebadores RM20B (amplificación superior) y RM21 (amplificación inferior). Las líneas verticales de color azul marcan el número de pocillos del gel y líneas horizontales en amarillo determinan el posible rango de amplificación de los marcadores. Cruces de color azul, morado muestran los alelos amplificados para el cebador RM20B y las cruces en verde muestran los productos de amplificación para el cebador RM21 de 88 muestras.

Diversidad Genética

20 variedades proporcionadas por el programa de arroz del INIAP y 61 muestras provenientes de los cantones de Zapotillo y Macará fueron evaluadas por triplicado para 17 loci microsatélites. La **Tabla 3** presenta un total de 75 alelos

encontrados con su respectiva frecuencia en cada locus. Los 17 loci presentaron un promedio de 4 alelos/locus, el cebador PSM363 es el más polimórfico con 8 alelos y el cebador RM208 el menos polimórfico con 2 alelos, el valor más alto para el contenido de información polimórfica PIC fue del cebador PSM363 con un valor de 0.6729 y el valor más bajo obtenido fue de 0.1296 para el cebador RM274 con un promedio de 0.3507 para los 17 loci. Los cebadores PSM132, RM208, RM415 y RM144 presentaron una heterocigosidad observada (H_o) de 0, es decir que las variedades estudiadas son homocigotas para estos loci, el cebador como mayor heterocigosidad observada fue RM20B con un valor de 0.9865 y 4 alelos y la media para los 17 loci fue $H_o=0.0793$.

La **Tabla 4** presenta un resumen de los parámetros de diversidad genética para los 17 loci en estudio.

Tabla 3.

Alelos encontrados en cada locus

Marcador	Alelo (pb)	No. de repeticiones	Frecuencia
PSM132	169	2	0.0046
	185	32	0.0731
	191	28	0.0639
	195	370	0.8447
	199	6	0.0137
RM208	290	346	0.7723
	294	102	0.2277
RM274	132	9	0.0200
	136	419	0.9311
	138	6	0.0133
	140	10	0.0222
	154	6	0.0133
PSM122	198	344	0.7644
	202	98	0.2178
	204	8	0.0178
PSM332	113	23	0.0511
	127	399	0.8867

Marcador	Alelo (pb)	No. de repeticiones	Frecuencia
	131	14	0.0311
	147	14	0.0311
PSM326	233	120	0.2667
	235	6	0.0133
	237	324	0.7200
RM9	130	309	0.6928
	134	8	0.0179
	136	90	0.2018
	180	39	0.0874
RM286	112	95	0.2121
	116	2	0.0045
	122	278	0.6205
	98	73	0.1629
RM50	181	13	0.0290
	197	359	0.8013
	201	70	0.1563
	205	6	0.0134
RM20B	210	143	0.3221
	231	80	0.1802
	258	209	0.4707
	261	12	0.0270
RM21	129	306	0.6861
	149	10	0.0224
	159	18	0.0404
	165	30	0.0673
	169	82	0.1839
PSM167	203	8	0.0178
	205	102	0.2267
	207	297	0.6600
	219	43	0.0956
RM229	114	6	0.0133
	116	105	0.2323
	122	341	0.7544
PSM415	115	12	0.0267
	118	6	0.0133
	133	67	0.1489
	145	339	0.7533
	148	12	0.0267
	163	14	0.0311
RM415	145	6	0.0135
	223	410	0.9234
	257	14	0.0315

Marcador	Alelo (pb)	No. de repeticiones	Frecuencia
	263	12	0.0270
	267	2	0.0045
PSM363	137	12	0.0268
	141	1	0.0022
	149	26	0.0580
	151	114	0.2545
	155	203	0.4531
	159	32	0.0714
	169	42	0.0938
	175	18	0.0402
RM144	219	384	0.8649
	234	12	0.0270
	237	8	0.0180
	240	6	0.0135
	243	32	0.0721
	252	2	0.0045

Nota. La tabla muestra los alelos encontrados para los 17 loci analizados. En la segunda columna se encuentra el tamaño del alelo en pares de bases, la tercera columna muestra el número de veces que dicho alelo se encontró en el total de las muestras y la última columna muestra la frecuencia de cada alelo.

Tabla 4.

Parámetros de diversidad genética para 81 variedades de arroz en 17 loci.

Marcador	Mayor Frecuencia Alélica	No. de genotipos	Tamaño de la Muestra	No. de observaciones	No. de alelos	He	Ho	PIC
PSM132	0.8447	5.0000	226.0000	219.0000	5.0000	0.2768	0.0000	0.2630
RM208	0.7723	2.0000	226.0000	224.0000	2.0000	0.3517	0.0000	0.2898
RM274	0.9311	6.0000	226.0000	225.0000	5.0000	0.1318	0.0444	0.1296
PSM122	0.7644	4.0000	226.0000	225.0000	3.0000	0.3679	0.0444	0.3121
PSM332	0.8867	5.0000	226.0000	225.0000	4.0000	0.2093	0.0044	0.2021
PSM326	0.7200	4.0000	226.0000	225.0000	3.0000	0.4103	0.0356	0.3364
RM9	0.6928	5.0000	226.0000	223.0000	4.0000	0.4713	0.0314	0.4239
RM286	0.6205	7.0000	226.0000	224.0000	4.0000	0.5434	0.0804	0.4859
RM50	0.8013	5.0000	226.0000	224.0000	4.0000	0.3324	0.0045	0.2997
RM20B	0.4707	6.0000	226.0000	222.0000	4.0000	0.6415	0.9865	0.5739
RM21	0.6861	7.0000	226.0000	223.0000	5.0000	0.4888	0.0179	0.4502
PSM167	0.6600	5.0000	226.0000	225.0000	4.0000	0.5036	0.0756	0.4496
RM229	0.7544	4.0000	226.0000	226.0000	3.0000	0.3767	0.0044	0.3151
PSM415	0.7533	7.0000	226.0000	225.0000	6.0000	0.4078	0.0133	0.3796

Marcador	Mayor Frecuencia Alélica	No. de genotipos	Tamaño de la Muestra	No. de observaciones	No. de alelos	He	Ho	PIC
RM415	0.9234	5.0000	226.0000	222.0000	5.0000	0.1454	0.0000	0.1421
PSM363	0.4531	8.0000	226.0000	224.0000	8.0000	0.7103	0.0045	0.6729
RM144	0.8649	6.0000	226.0000	222.0000	6.0000	0.2456	0.0000	0.2359
Media	0.7412	5.3529	226.0000	223.7059	4.4118	0.3891	0.0793	0.3507

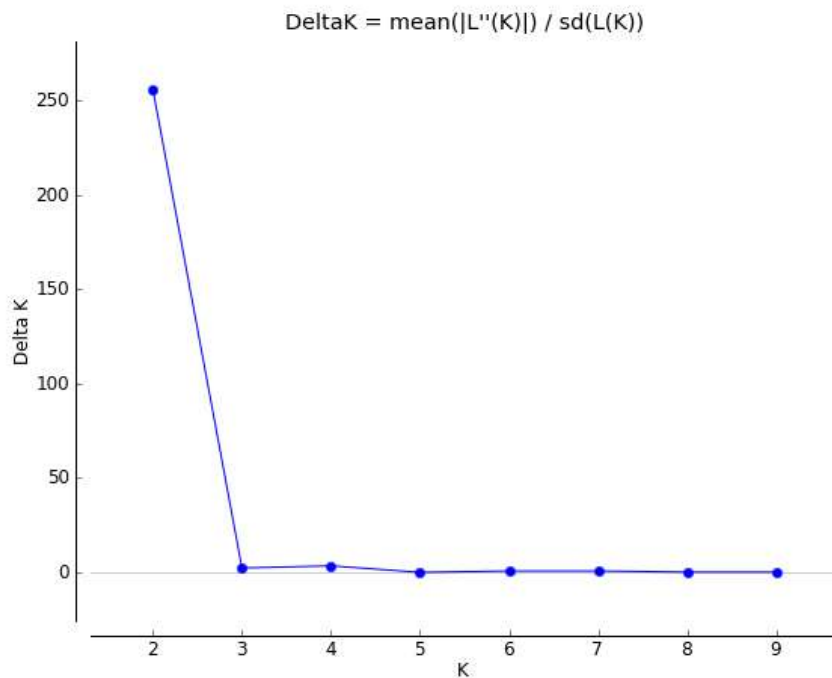
Nota. En la tabla se muestran los parámetros de diversidad genética analizados en el presente estudio para los 17 loci analizados. En la primera columna se muestra la mayor frecuencia de los alelos; en la segunda columna el número de combinaciones existente de los alelos; en la tercera, el tamaño de la muestra se refiere al número de individuos totales y el número de observaciones descrito en la cuarta columna se refiere al número de individuos que mostraban una banda; la quinta columna muestra el número de alelos encontrados; la sexta, He= Heterocigosidad esperada o número de heterocigotos esperados si hubiera un apareamiento aleatorio en la población; Ho= Heterocigosidad observada, es decir el número de heterocigotos encontrados y PIC= Contenido de Información Polimórfica o el grado de información que nos puede brindar un marcador.

Análisis de Estructura Genética

Se realizó la asignación genética de los individuos a poblaciones predeterminadas y se encontró que el valor de $k=2$ genera el mayor Δk con un valor de 255.70 como se muestra en la **Figura 6**. En el **Anexo 4** se encuentra la información complementaria para tomar $k=2$ como el número de poblaciones para las muestras del estudio.

Figura 6.

Valores de k vs Δk resultado del análisis de 17 loci para las muestras.

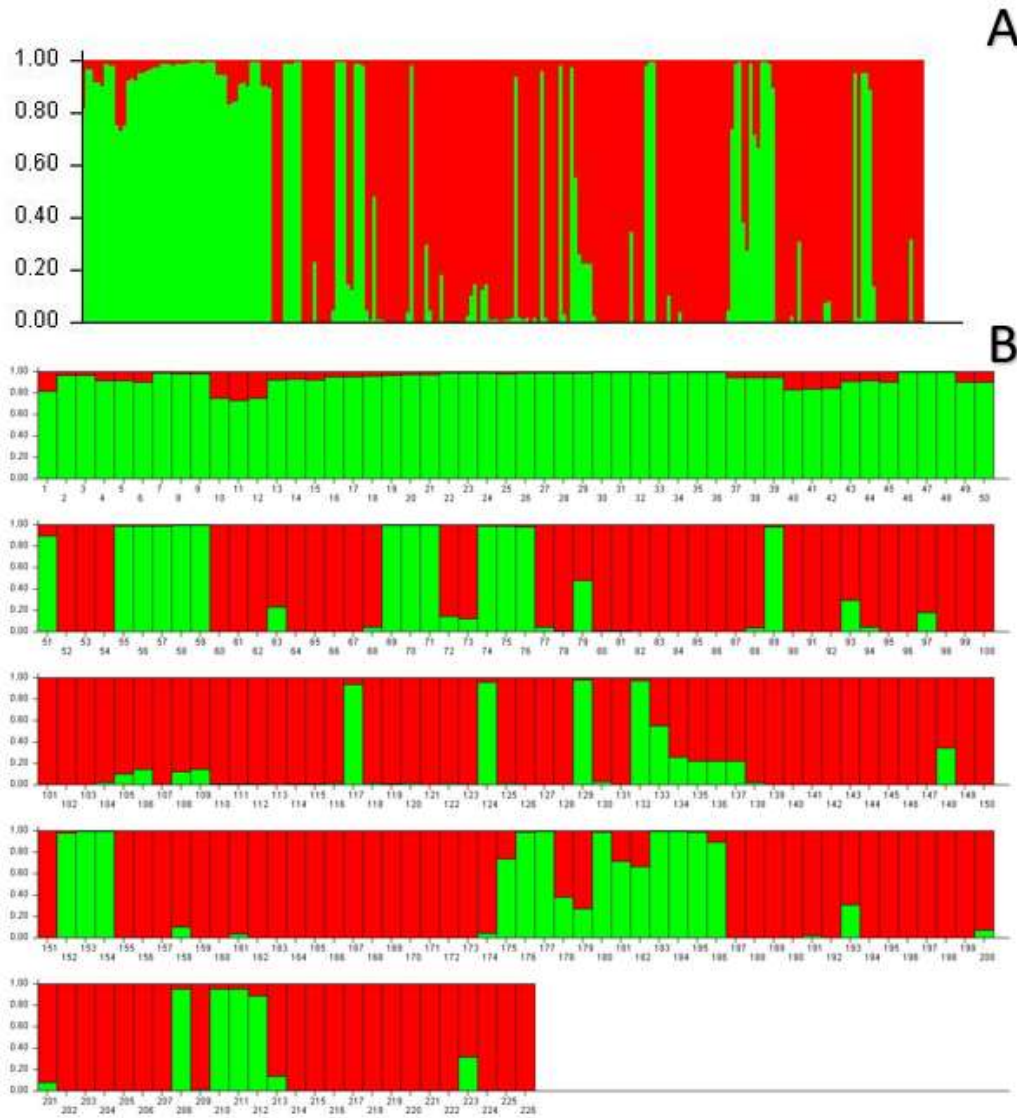


Nota. Gráfica de k vs Δk , el valor más alto de Δk determinado por el método de Evanno permite determinar el verdadero número de poblaciones para las muestras en estudio. En el eje x se muestran los valores de k o el número de poblaciones posibles en las que se puede distribuir a los individuos en estudio y en el eje y se muestran los valores de Δk .

La asignación de las muestras a las dos poblaciones determinadas se encuentra en la **Figura 7**, la población 1 (POP1) cuenta con 85 individuos mientras que la población 2 (POP2) cuenta con 141 individuos

Figura 7.

Asignación de las muestras a las dos poblaciones determinadas.



Nota. A) Vista general de la asignación de los individuos a las dos poblaciones, B) vista por individuo. En las gráficas el eje x representa a las muestras y el eje y muestra la probabilidad de pertenencia a las dos poblaciones determinadas, en rojo probabilidad de pertenencia a la población 1 (POP1) y en verde probabilidad de pertenencia a la población 2 (POP2). La población a la que cada muestra fue asignada se determino por el valor mayor entre las dos poblaciones.

Tras la asignación genética de individuos por el método bayesiano, se calculó una matriz de disimilaridad para estimar la similitud entre las distintas muestras a través del coeficiente de emparejamiento simple SM (simple matching) y 1000 iteraciones, obteniendo un valor medio de disimilaridad de 0.37099, un valor mínimo de 0, es decir muestras que eran completamente similares, y un valor máximo de 0.90909 entre la variedad INIAP 12 (testigo 10) y la muestra 593 proveniente de campo.

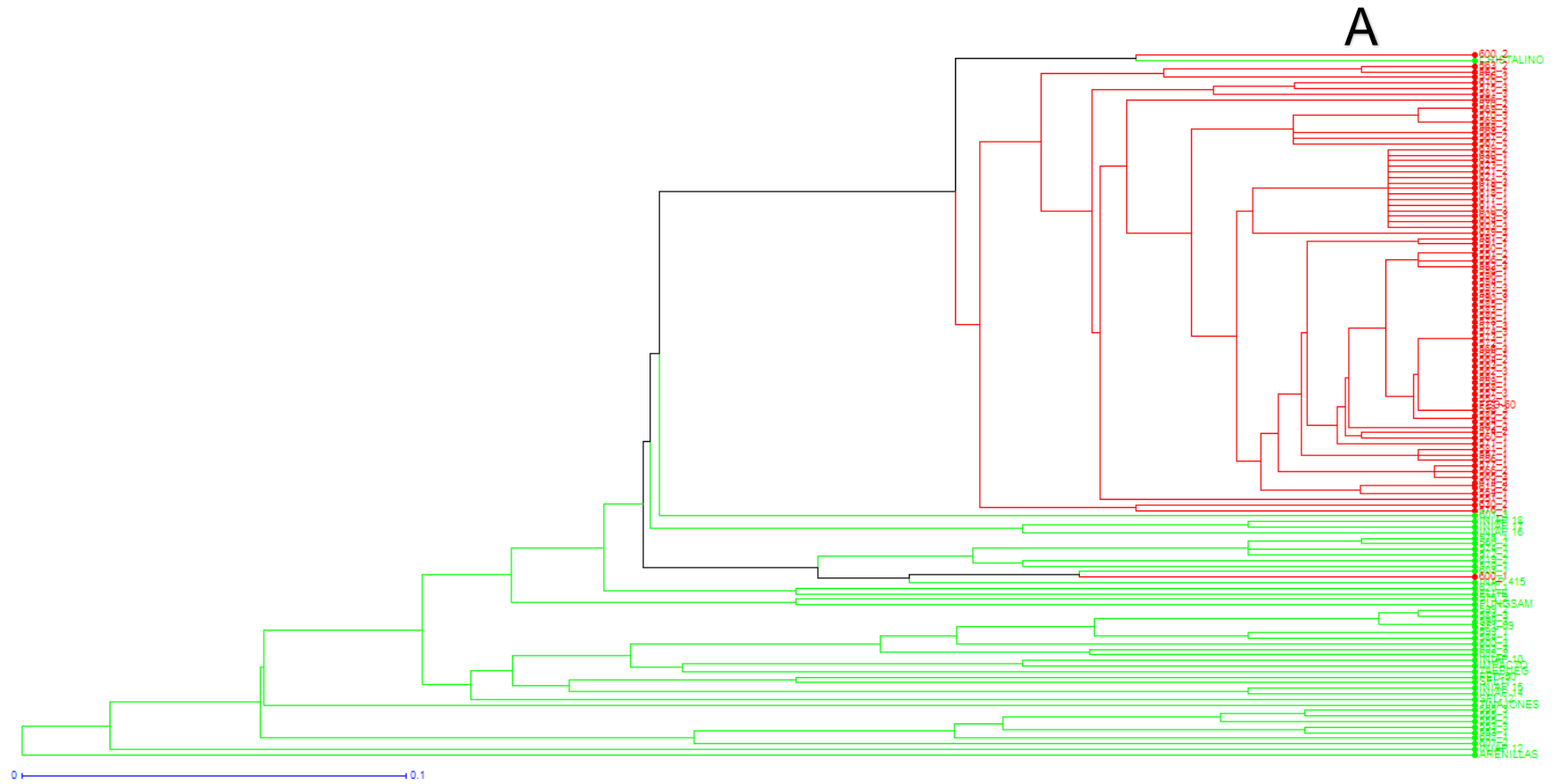
Con los valores obtenidos de disimilaridad se realizaron un dendograma con el método UPGMA (**Figura 8**) y un dendograma mediante el método del vecino más cercano (NJ) (**Figura 9**) para determinar cual de los dos métodos distribuía mejor las muestras según las poblaciones obtenidas mediante el método bayesiano.

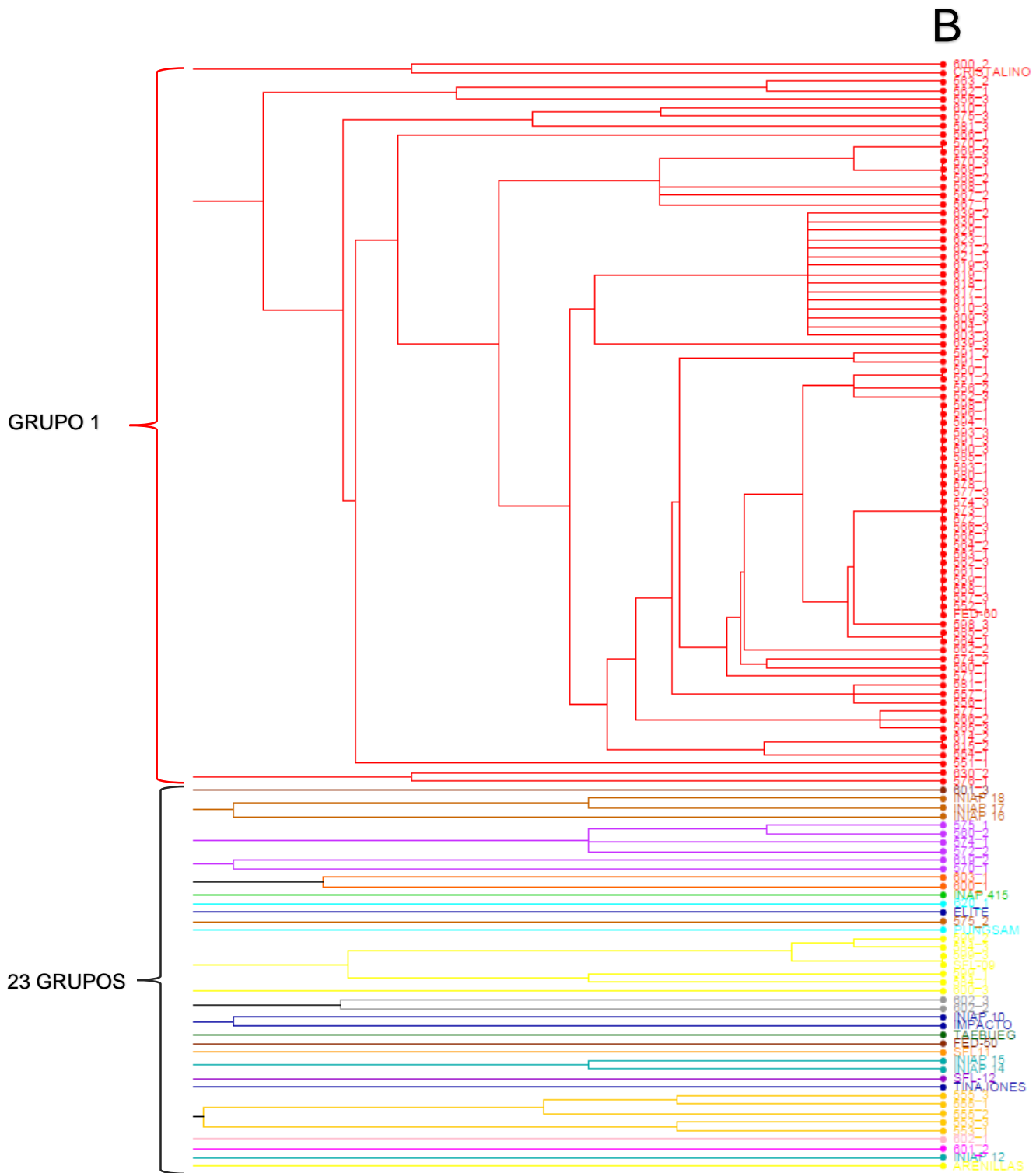
El dendograma UPGMA mostró que las variedades Arenillas (testigo 3) e INIAP 12 (testigo 10) son las más distantes dentro de la población de estudio y la variedad FED-60 (testigo 18) es la más emparentada con las muestras obtenidas en campo.

Se utilizó la asignación obtenida en el método bayesiano para determinar la posición de los individuos de ambas poblaciones dentro del dendograma como se muestra en la **Figura 8A**, se encontró que la mayoría de las muestras se encontraban agrupadas conforme a las dos poblaciones a excepción de la variedad Cristalino y la muestra 600.1 que se encontraban en la población contraria a la que fueron asignadas. Luego se realizaron los agrupamientos de acuerdo a la distancia de las ramas y se encontró que las muestras se agrupaban en 24 grupos, en donde el primer grupo contenía a la población 1 del método bayesiano pero la población 2 se distribuía en los 23 grupos restantes como se muestra en la **Figura 8B**.

Figura 8.

Dendograma UPGMA generado mediante la matriz de disimilaridad.





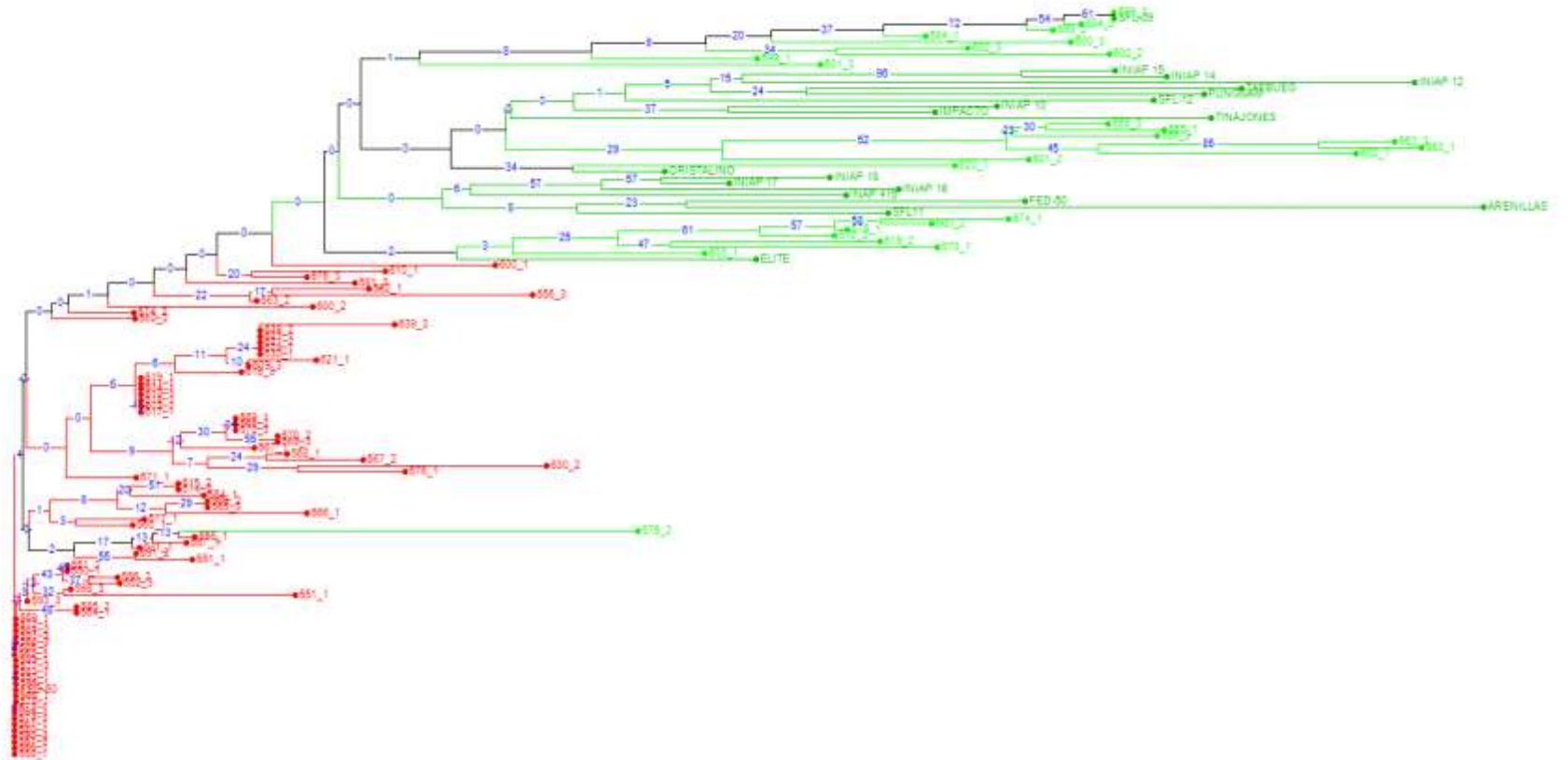
Nota. A) Vista general del dendrograma UPGMA para las variedades de estudio distribuidas de acuerdo al método bayesiano, B) Clusters en los que se agruparon a las variedades de arroz, en rojo grupo 1 que corresponde a la población 2 del método bayesiano y en distintos colores los 23 grupos en los que se dividió a la población 1 del método bayesiano de acuerdo a la longitud de las ramas.

El dendograma realizado con el método de Neighbor Joining (NJ), confirmó que las muestras en estudio se agrupan en las dos poblaciones obtenidas por el método Bayesiano, la muestra 575.2 se agrupó con la población 2 a pesar de corresponder a la población 1, como se muestra en la **Figura 9**.

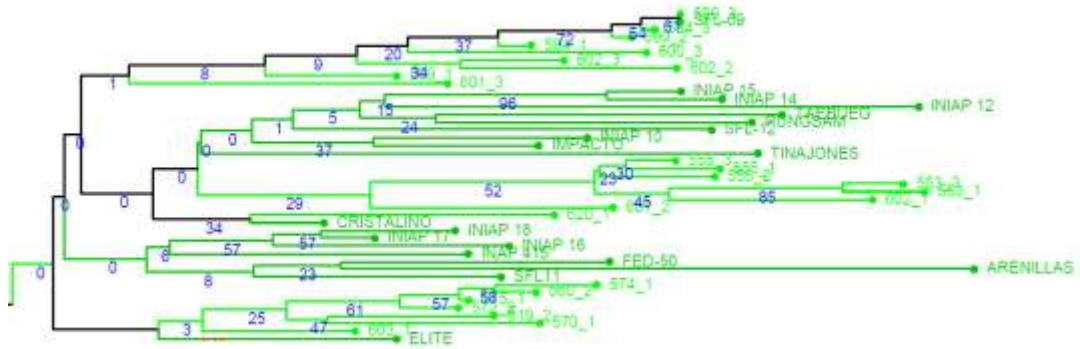
Tras la asignación de poblaciones, se realizó un Análisis de Coordenadas Principales (PCoA, por sus siglas en inglés) para analizar la distribución de los datos (**Figura 10**), el PCoA mostró que las muestras forman dos grupos confirmando la asignación descrita anteriormente.

Figura 9.

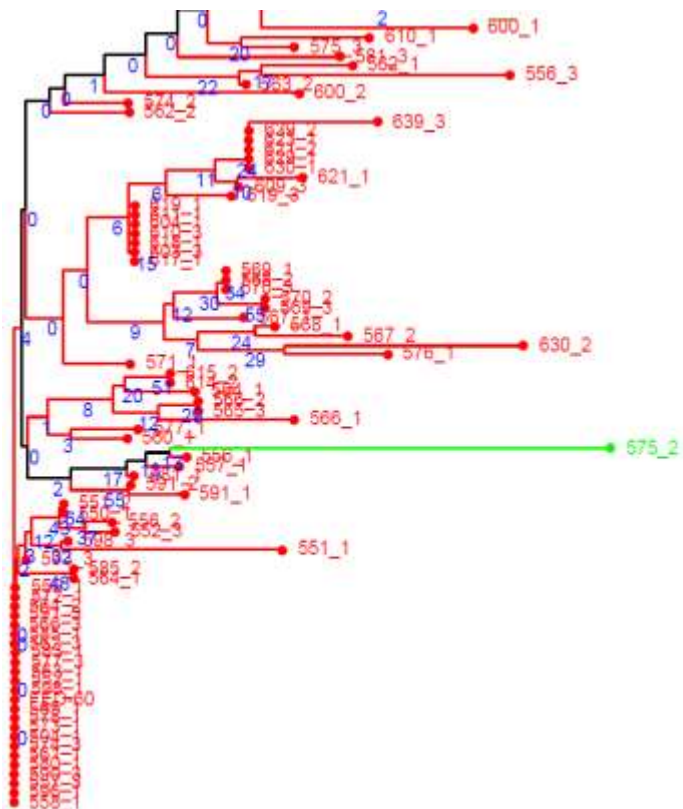
Dendrograma NJ para las muestras en estudio

A

B



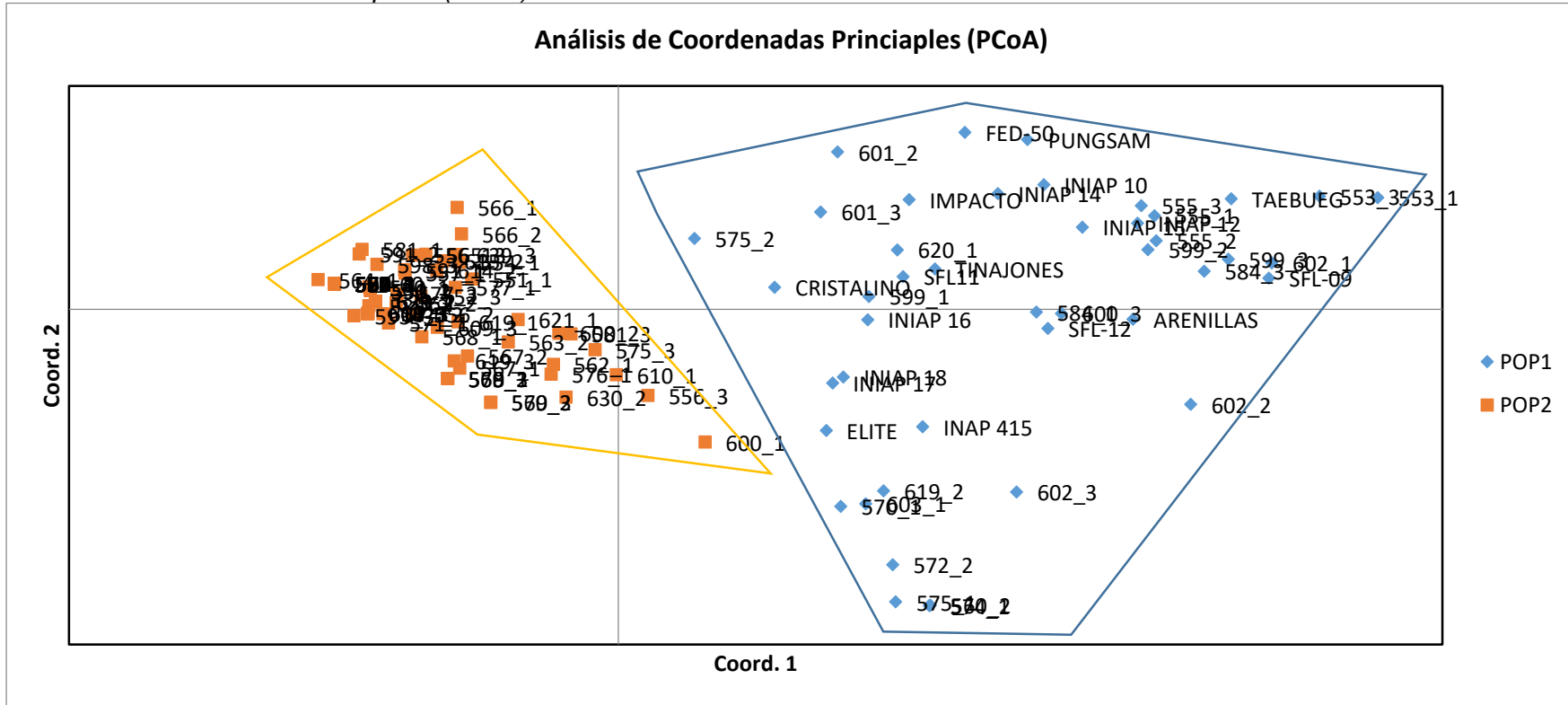
C



Nota. A) Vista general del dendrograma realizado con el método NJ para las muestras de estudio; B) Población 1 (POP1) determinada mediante el método bayesiano vista en el dendrograma; C) Población 2 (POP2) determinada mediante el método bayesiano vista en el dendrograma. La muestra 575.2 pertenece a POP1 a pesar de ser agrupada en POP2 en el dendrograma. La longitud de las ramas representa cantidad de cambio genético.

Figura 10.

Análisis de Coordenadas Principales (PCoA)



Nota. El PCoA permite la visualización de la relación de similitud o disimilitud entre los individuos. En la imagen se muestra el PCoA realizado para los individuos en estudio, las muestras se agruparon de acuerdo a las poblaciones asignadas por el método bayesiano.

Las tres primeras coordenadas acumulan el 44.85% de la variación total de datos, donde la primera coordenada principal explica el 30.31% de la variación, la segunda coordenada el 7.48 % y la tercera explica el 7.06% de la varianza, con valores propios de 56.26, 13.88 y 13.10 respectivamente.

Análisis Molecular de Varianza y Estadística F

El análisis molecular de Varianza (AMOVA) se realizó a partir de las distancias calculadas con un valor p de 0.001. El AMOVA mostró que el mayor porcentaje de varianza se da dentro de la población con un 82% mientras que la varianza entre poblaciones es únicamente del 18% como se muestra en la **Tabla 5** y la **Figura 11**.

Tabla 5.

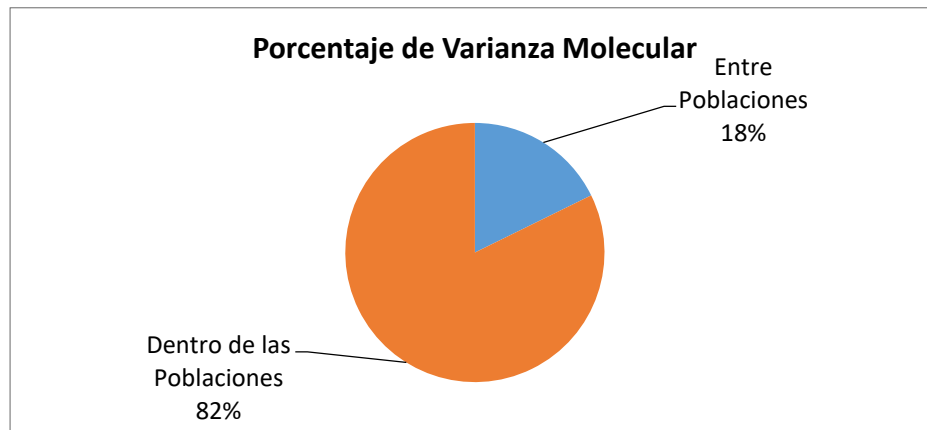
Análisis Molecular de Varianza AMOVA para muestras de arroz.

Fuente	g.l	Suma de Cuadrados	Promedio de Cuadrados	Variación Estándar	Porcentaje de Varianza
Entre Poblaciones	1	145.052	145.052	0.669	18%
Dentro de las Poblaciones	450	1397.098	3.105	3.105	82%
Total	451	1542.150		3.774	100%

Nota. La tabla muestra el porcentaje de varianza de las dos poblaciones determinadas por el análisis del programa STRUCTURE para determinar el porcentaje de variabilidad entre poblaciones y dentro de las poblaciones.

Figura 11.

Porcentaje de varianza molecular.



Nota. El diagrama de pastel muestra el porcentaje de varianza de las dos poblaciones descritas por el método bayesiano teniendo un valor mayor para la varianza dentro de las poblaciones.

Los estadísticos F de Wright calculados para los 17 loci se presentan en la **Tabla 6**. Las estimaciones para el coeficiente de endogamia F_{IS} tuvieron un valor negativo en el locus RM20B que significa un exceso de heterocigotos con respecto al equilibrio de Hardy-Weinberg, en el resto de locus se obtuvo un valor positivo y un valor de 1 en los loci PSM132, RM208, RM415, RM144, la media fue de 0.845 que indica que existe endogamia dentro de la población de estudio.

El índice de fijación F_{IT} tuvo un comportamiento similar al coeficiente de endogamia en los distintos loci, el valor medio fue de 0.855 que significa que la diferenciación de la población en estudio con respecto a la población total se debe a la endogamia, por último, el coeficiente de fijación F_{ST} tiene un valor de 0.090 que muestra que no existe una subdivisión en la población, es decir que existe una diferenciación genética baja dentro de la población de estudio.

Tabla 6.

Estadísticos F de Wright estimados para los 17 loci.

Locus	F_{IS}	F_{IT}	F_{ST}
PSM132	1.000	1.000	0.053
RM208	1.000	1.000	0.229
RM274	0.721	0.729	0.030
PSM122	0.885	0.894	0.084
PSM332	0.975	0.976	0.052
PSM326	0.906	0.922	0.171
RM9	0.925	0.938	0.179
RM286	0.828	0.854	0.151
RM50	0.990	0.990	0.018
RM20B	-0.593	-0.510	0.053
RM21	0.960	0.964	0.099
PSM167	0.805	0.812	0.038
RM229	0.990	0.992	0.156
PSM415	0.974	0.977	0.131
RM415	1.000	1.000	0.023
PSM363	0.995	0.995	0.035
RM144	1.000	1.000	0.026
Media	0.845	0.855	0.090

Nota. En la primera columna se muestra el estadístico F_{IS} (coeficiente de endogamia) que puede tener valores de -1 a 1 siendo los valores negativos un exceso de heterocigotos y los valores positivos una deficiencia de los mismo, este coeficiente determina la probabilidad de que dos alelos de un individuo sean iguales por descendencia con respecto a las dos poblaciones determinadas por el método bayesiano. La segunda columna muestra el estadístico F_{IT} (índice de fijación) que también puede tomar valores entre -1 y 1, este índice se define como la probabilidad de que dos alelos sean idénticos en un mismo individuo por descendencia con respecto a la población total de estudio. Por último el coeficiente F_{ST} (coeficiente de fijación) puede tomar valores entre 0 y 1 y se define como el grado de divergencia entre las dos poblaciones determinadas por el método bayesiano; el valor de 0 indica panmixia y el valor de 1 una subdivisión extrema de las dos poblaciones. La tabla muestra los valores calculados para cada locus y la media para los 17 loci.

Identificación de Genotipos Duplicados

Se realizó un análisis de genotipos duplicados para determinar si las variedades obtenidas en campo correspondían a una de las variedades procedentes del programa de arroz, tras el análisis se obtuvo que, de 127 genotipos obtenidos, el 18.90% corresponden a la variedad FED-60 y es probable que el 0.79% correspondan a la variedad SFL-09 como se muestra en la **Tabla 7**.

Además, se observó que varias muestras eran similares entre sí, pero no correspondían a ninguna variedad procedente del programa de arroz por lo que se procedió a agruparlas obteniendo 8 grupos (**Tabla 8**), 53 muestras era diferentes entre sí y con las variedades procedentes del programa de arroz como se observa en la **Tabla 9**. La muestra 609.3 comparte relación con el grupo 6 y 8 debido a que de los 34 alelos analizados únicamente se obtuvieron 30, en la **Figura 12** se observa el análisis de redes de la relación de los genotipos duplicados.

Tabla 7.

Identificación de Genotipos duplicados.

Variedad	Genotipos Duplicados	Porcentaje de Similitud	Número de Alelos Idénticos	Número total de alelos analizados
FED-60	552.1	100%	34	34
	557.3	100%	34	34
	558.1	100%	34	34
	559.1	100%	34	34
	561.1	100%	34	34
	562.3	100%	34	34
	563.1	100%	34	34
	564.2	100%	34	34
	565.1	100%	34	34
	566.3	100%	34	34
	572.1	100%	34	34
	573.1	100%	34	34
	574.3	100%	34	34
	577.3	100%	34	34

Variedad	Genotipos Duplicados	Porcentaje de Similitud	Número de Alelos Idénticos	Número total de alelos analizados
	578.1	100%	34	34
	580.1	100%	34	34
	583.1	100%	34	34
	585.1	100%	34	34
	590.3	100%	34	34
	591.3	100%	34	34
	593.3	100%	34	34
	594.1	100%	34	34
	596.1	100%	34	34
	598.1	100%	34	34
SFL-09	599.3	100%	26	34

Nota. En la tabla se observan las muestras de campo (columna 2) que tuvieron un 100% de similitud con una variedad proveniente del programa de arroz (columna 1).

Tabla 8.

Agrupamientos formados por genotipos duplicados.

Grupo	Genotipos Duplicados	Porcentaje de Similitud	Número de Alelos Idénticos	Número total de alelos analizados
1	550.1	100%	34	34
	551.2	100%	34	34
2	564.1	100%	32	34
	585.2	100%	32	34
3	565.3	100%	32	34
	566.2	100%	32	34
	577.1	100%	30	34
4	568.2	100%	34	34
	569.1	100%	34	34
	570.3	100%	34	34
5	569.3	100%	34	34
	570.2	100%	34	34
6	603.3	100%	34	34
	604.1	100%	34	34
	609.3	100%	30	34
	610.3	100%	34	34
	611.1	100%	34	34
	617.1	100%	34	34
	618.1	100%	34	34
	619.1	100%	30	34

Grupo	Genotipos Duplicados	Porcentaje de Similitud	Número de Alelos Idénticos	Número total de alelos analizados
7	619.3	100%	30	34
	614.2	100%	32	34
	615.2	100%	32	34
8	609.3	100%	30	34
	621.2	100%	34	34
	623.1	100%	34	34
	629.1	100%	34	34
	630.1	100%	34	34
	639.2	100%	34	34

Nota. La tabla muestra las muestras provenientes de campo que eran idénticas entre sí pero no correspondían a ninguna variedad proveniente del programa de arroz.

Tabla 9.

Muestras de campo sin identificación.

Muestras sin identificar				
552.3	562.1	575.2	600.2	639.3
553.1	562.2	575.3	600.3	
553.3	563.2	576.1	601.2	
554.1	566.1	581.1	601.3	
555.1	567.1	581.3	602.1	
555.2	567.2	584.1	602.2	
555.3	568.1	584.3	602.3	
556.1	570.1	591.1	603.1	
556.2	571.1	591.2	610.1	
556.3	572.2	598.3	619.2	
557.1	574.1	599.1	620.1	
560.1	574.2	599.2	621.1	
560.2	575.1	600.1	630.2	

Nota. La tabla indica las muestras provenientes de campo que no tuvieron ninguna similitud.

Por último, se analizó el porcentaje de identidad de las variedades provenientes del programa de arroz para los 17 loci, se encontró que las variedades INIAP 14 e INIAP 15 comparten el 88.24% de identidad al igual que las variedades INIAP 17 e INIAP18, es decir que, de 34 alelos analizados, las variedades comparten entre sí 30 alelos; las variedades Impacto e INIAP10, así como las variedades INIAP16, INIAP17 e INIAP18 comparten el 76.47% de similitud; por último las variedades INIAP415 y SFL-09 comparten el 73.33% como se muestra en la **Tabla 10**.

Tabla 10.

Porcentaje de identidad entre variedades provenientes del programa de arroz.

Variedad 1	Variedad 2	Porcentaje de Similitud	Número de Alelos idénticos	Número total de alelos analizados
INIAP 14	INIAP 15	88.24%	30	34
INIAP 17	INIAP 18	88.24%	30	34
IMPACTO	INIAP 10	76.47%	26	34
INIAP 16	INIAP 17	76.47%	26	34
INIAP 16	INIAP 18	76.47%	26	34
INIAP 415	SFL-09	73.33%	22	34

Nota. La tabla muestra el porcentaje de similitud de las variedades provenientes del programa de arroz con un valor mayor al 70%.

Capítulo 4: Discusión

Análisis de Diversidad Genética

El contenido de información polimórfica de un marcador corresponde a su capacidad para detectar el polimorfismo entre individuos de una población (Serrote Caetano, Reiniger, Silva, Rabaiolli, & Stefanel, 2019). Según Botstein y colaboradores (1980), un marcador codominante con valores PIC mayores que 0.5 se considera altamente informativo, valores entre 0.5 y 0.25 son razonablemente informativos y valores menores a 0.25 son poco o nada informativos.

Para los 17 loci analizados se determinó un valor PIC máximo de 0.6729 para el cebador PSM363 siendo este altamente informativo, lo que concuerda con el estudio realizado por Jiang y colaboradores (2010) quienes reportan un valor PIC de 0.5679 (>0.5) para el mismo marcador, por otro lado, el cebador RM274 es considerado poco o nada informativo por su bajo valor PIC (0.1296), Seetharam y colaboradores (2009) al analizar 30 genotipos de arroz que obtuvieron un valor PIC de 0.064 para este marcador, determinando su baja capacidad informativa para ciertos genotipos.

El valor PIC promedio fue de 0.3507, este valor es inferior a los reportados por Terang y colaboradores (2020) quienes al analizar la diversidad en cultivares de arroz de Asia Central y Occidental obtuvieron un valor PIC de 0.71, también al valor de 0.56 obtenido para 49 marcadores SSR por Pathaichindachote y colaboradores (2019) y al valor de 0.624 obtenido para variedades de Ecuador con 30 marcadores microsatélites (Pérez-Almeida et al., 2019). La diferencia en el contenido de información polimórfica del presente estudio con respecto a otras investigaciones puede deberse al grado de divergencia genética de las muestras estudiadas

(Ngangkham et al., 2019), indicando que los genotipos estudiados son moderadamente diversos (Kumar et al., 2012).

La heterocigosidad observada promedio fue de 0.0793, esto concuerda con estudios realizados por Becerra y colaboradores (2017), Park y colaboradores (2019) y Yelome y colaboradores (2018), esta baja heterocigosidad se debe a que el arroz es un cultivo altamente endogámico (Alvarez et al., 2007) en donde el cultivo intensivo promueve eventos de apareamiento dentro de una misma población (Gao & Gao, 2016).

Las variedades provenientes del programa de arroz son líneas puras liberadas a los productores para su cultivo, las líneas puras tras la progenie F4 en cultivos autógamos como el arroz, deben poseer aproximadamente el 94% de alelos heterocigotos (Singh et al., 2021), el cebador como mayor heterocigosidad observada, RM20B con un valor de 0.9865, fue capaz de detectar heterocigosidad residual en las variedades provenientes del programa de arroz.

Análisis de Estructura Génica

La asignación de los individuos a poblaciones determinadas realizada en el programa STRUCTURE agrupó a las muestras en dos poblaciones $k=2$ debido a que genera el mayor valor de Δk (255.70) y también debido al resto de información obtenida del programa, Lawson y colaboradores (2018), mencionan que a pesar de que generalmente se utilice el valor de Δk como factor determinante, es necesario analizar el resto de factores que arroja el programa para no sobre interpretar la información obtenida.

Todas las variedades procedentes del programa de arroz del INIAP fueron agrupadas en la población uno (POP1) que cuenta con 85 individuos con excepción de la variedad Ferón que fue asignada a la población dos (POP2) que posee 141

individuos. La agrupación de todas las variedades procedentes del programa de arroz puede ser explicada debido a que la mayoría de las variedades liberadas en América Latina y el Caribe tienen como origen los mismos progenitores, un estudio realizado por Berrio y colaboradores (2016) mostró que las variedades liberadas por el Fondo Latinoamericano para Arroz de Riego (FLAR) hasta el 2014, en 13 países miembros entre ellos Ecuador, tienen 3 ancestrales que han aportado con el 35.6% de sus genes a las variedades actuales. La similitud en la constitución genética puede provocar la vulnerabilidad al estrés biótico y abiótico, además de la dificultad para superar el techo de productividad existente en las variedades comerciales (Acevedo et al., 2006).

El valor promedio de disimilitud obtenido mediante emparejamiento simple para datos de información alélica fue de 0.37099. Se utilizó el coeficiente de emparejamiento simple ya que es un método generalmente utilizado para determinar la similitud o disimilitud entre individuos de una población que no posee información *a priori* (Kosman & Leonard, 2005) y recomendado para poblaciones con individuos muy cercanos (Mete Çilingirtürk & Ergüt, 2014) en donde todos los datos proveen la misma información.

El dendograma construido por el método UPGMA agrupó a los individuos en 24 clusters, las variedades INIAP 14 e INIAP 15 comparten un mismo nodo dentro del árbol, al igual que las variedades INIAP 16, 17 y 18, estos resultados concuerdan con los obtenidos por Pérez y colaboradores (2019). De acuerdo a los autores las variedades INIAP 14 y 15 comparten rasgos agronómicos como precocidad, alta labranza y buen rendimiento, además se menciona que las variedades 16, 17 y 18 se encuentran emparentadas con un tipo de planta similar y buena calidad de grano.

La variedad SFL-11 y Fedearroz50 (FED-50) también comparten un mismo nodo, esto puede deberse a que ambas variedades comparten características de tolerancia a enfermedades como *Pyricularia* y la enfermedad de la hoja blanca (Caicedo, 2008), además de que la variedad FED-50 ha sido ampliamente utilizada como progenitora en varios programas de cruzamiento en Latino América (Pérez-Almeida et al., 2019). El resultado obtenido con el método bayesiano concuerda con los resultados obtenidos en el dendograma realizado en este estudio, que determinó que FED-60 está más emparentada con las muestras procedentes de campo. El dendograma realizado mediante el método de NJ agrupó a las muestras en dos clusters de acuerdo a las dos poblaciones determinadas por el método bayesiano a excepción de la muestra 575.2 que a pesar de pertenecer a la población 1 fue agrupada en POP2, esto se puede deber a que los valores de probabilidad de pertenencia de la muestra en el método bayesiano fueron de 0.49 para la población 2 y 0.51 para la población 1, por lo que tenía probabilidad casi similar de pertenecer a cualquiera de las dos poblaciones.

Se encontró una diferencia en el número de poblaciones o clusters en los que agrupó el programa STRUCTURE a los individuos con respecto al método de agrupamiento UPGMA, esta diferencia se debe a que el programa STRUCTURE utiliza un algoritmo iterativo bayesiano mediante la colocación de muestras en grupos cuyos miembros comparten patrones de variación similares (Porrás-Hurtado et al., 2013) que a diferencia de los métodos tradicionales de agrupamiento brinda mayor información y mayor fiabilidad para asignar a los individuos a poblaciones (Breton et al., 2008). El método de NJ (Saitou & Nei, 1987) agrupa a las muestras de mejor manera que el dendograma generado por el método de UPGMA, esto se debe a que a diferencia del método UPGMA, NJ asume una tasa de evolución distinta para cada rama acercándose más a la realidad.

El PCoA realizado corroboró los datos obtenidos mediante el método de agrupamiento bayesiano al mostrar dos grupos definidos, el análisis de coordenadas principales es un método robusto y conveniente para mostrar relaciones complejas multivariadas (Gower, 2005).

Análisis Molecular de Varianza y Estadística F

El AMOVA mostró que el mayor porcentaje de varianza se da dentro de la población con un 82% mientras que la varianza entre poblaciones es únicamente del 18%, estos resultados son similares a los obtenidos por Nachimuthu y colaboradores (2015), Suvi y colaboradores (2020) e Islam y colaboradores (2018). La alta variabilidad intrapoblacional y la homogeneidad genética entre las poblaciones podrían haber surgido debido a altos niveles de flujo genético. Alternativamente, puede ser que estas poblaciones simplemente no se hayan separado el tiempo suficiente para acumular diferencias genéticas detectables (SHENG et al., 2005).

Los estadísticos F: F_{IS} , F_{IT} y F_{ST} tuvieron valores medios de 0.845, 0.855 y 0.090 respectivamente, el cálculo de estos valores reveló la desviación de la ley de Hardy-Weinberg para la variación molecular dentro de la población, el resultado de F_{ST} indica una baja divergencia existente en la población mientras que el valor F_{IT} alto indica una falta de heterocigosidad en la población, esto se debe probablemente a la naturaleza endogámica del arroz (Nachimuthu et al., 2015). Los datos obtenidos de los estadísticos F concuerdan con los obtenidos en el análisis molecular de varianza.

Identificación de Genotipos Duplicados

Se establecieron 8 grupos de muestras que eran similares entre sí, pero no correspondían a ninguna variedad procedente del programa de arroz, todos los agrupamientos se encuentran dentro de la población 2 como se observa en el

dendograma NJ, también se determinó que 53 muestras no tenían ninguna identidad y estas se encontraban distribuidas en ambas poblaciones determinadas por el método bayesiano. Esto puede deberse a que en el presente estudio no se utilizaron todas las variedades comerciales como referenciales para la identificación de genotipos duplicados, por lo que tanto las agrupaciones como las muestras sin identificación pueden corresponder a otras variedades no genotipadas.

Otra posible explicación puede ser que tanto las agrupaciones como los materiales no identificados probablemente son variedades acriolladas. Loja es un *hotspot* de diversidad debido a su ubicación geográfica, diversidad de pisos altitudinales, la depresión de Huancabamba, geomorfología y la presencia de cuatro cuencas hidrográficas que generan múltiples hábitats y microhábitats (Aguirre Mendoza et al., 2017) obligando a las variedades a adaptarse a las condiciones agro-ecológicas cambiantes.

A veces estos cambios son lo suficientemente intensos como para causar estrés en las plantas que no es detectable a nivel de fenotipo, pero tiene una respuesta molecular para adaptarse (Gallego & Benavides, 2019) provocando la aparición de nuevos genotipos. En el caso de los agrupamientos, estos pueden ser segregantes de la variedad FED-60 ya que se encuentran en la misma población determinada por el método bayesiano, como en el dendograma realizado por el método NJ. Por otro lado, las muestras no identificadas que se encuentran distribuidas entre ambas poblaciones pueden ser híbridos de las variedades comerciales ya que, a pesar de que el arroz es una planta con una alta tasa de autopolinización, en las condiciones meteorológicas adecuadas o en la presencia de polinizadores puede haber eventos de hibridación natural (MATSUI & KAGATA, 2003; "Rice Pollen Goes the Distance," 2014). La presencia de variedades acriolladas puede explicar también el alto rendimiento de producción de la provincia,

múltiples estudios han descrito el fenómeno denominado heterosis o vigor híbrido en el que los híbridos tienen mayor rendimientos que los progenitores heterocigotos (Huang et al., 2015).

Por último, se analizó el porcentaje de identidad de las variedades provenientes del programa de arroz para los 17 loci, se encontró que las variedades INIAP 14 e INIAP 15 comparten el 88.24% de identidad al igual que las variedades INIAP 17 e INIAP18. Esto reafirma los datos obtenidos en el dendograma en los que las variedades que presentan una alta identidad se encuentran en el mismo nodo.

Capítulo 5: Conclusiones

- Se determinaron 75 alelos en los 17 loci analizados con un promedio de 4 alelos/locus que permitieron identificar tanto las variedades provenientes del programa de arroz como las muestras de campo.
- El marcador RM20B, con un valor $H_o = 0.0793$, pudo identificar heterocigosidad residual en las variedades provenientes del programa de arroz pese a que estas son líneas puras.
- A pesar de la naturaleza endogámica del arroz existe variabilidad entre las variedades estudiadas que pudo ser detectada con los 17 marcadores analizados de acuerdo al valor PIC promedio de 0.3507, el análisis molecular de varianza y los valores obtenidos de los estadísticos F de Wright F_{IS} , F_{IT} y F_{ST} (0.845, 0.855 y 0.090 respectivamente).
- Mediante el análisis de agrupamiento por el método NJ y el método bayesiano se concluyó que las variedades provenientes del programa de arroz se encuentran más relacionadas entre sí que con las variedades provenientes de campo a excepción de la variedad FED-60.
- De 127 genotipos obtenidos, el 18.90% corresponden a la variedad FED-60 y es probable que el 0.79% correspondan a la variedad SFL-09 debido a que comparten 26 de los 34 alelos analizados, debido a la información perdida en la variedad SFL-09 no podemos realizar una afirmación concreta.

- Los 8 grupos obtenidos y las 53 muestras no identificadas pueden tratarse de variedades acriolladas que surgieron debido a la presión de selección por las condiciones agro-climáticas del lugar de estudio y podrían explicar el alto rendimiento de producción de la provincia de Loja.

Capítulo 6: Recomendaciones

- Utilizar un mayor número de marcadores para poder determinar de una manera más precisa el grado de diversidad presente en las variedades de arroz.
- Realizar la caracterización molecular de las variedades provenientes del programa de arroz del INIAP que no fueron genotipadas en el estudio para poder tener un mayor número de referenciales para la identificación de genotipos en la provincia de Loja.
- Realizar estudios más profundos de los genotipos que no se lograron identificar en el presente estudio pues sus características podrían ser de interés para el programa nacional de arroz.

Capítulo 7: Bibliografía

- Adhikari, S., Saha, S., Biswas, A., Rana, T. S., Kumar Bandyopadhyay, T., & Ghosh, P. (2017). Application of molecular markers in plant genome analysis: a review. *The Nucleus*.
- Awika, J. (2011). Major Cereal Grains Production and Use around the World. En J. Awika, V. Piironen, & S. Bean, *Advances in Cereal Science: Implications to Food Processing and Health Promotion*. American Chemical Society.
- Bhandari, H., Bhanu, A. N., Srivastava, K., Singh, M., Shreya, & Hemantaranjan, A. (2017). Assessment of genetic diversity in crop plants - an overview. *Advances in PLant and Agriculture Research*.
- Blacket, M. J., Robi, C., Good, R. T., Lee, S. F., & Miller, A. D. (2012). Universal primers for fluorescent labelling of PCR fragments—an efficient and cost-effective approach to genotyping by fluorescence. *Molecular Ecology Resources*.
- Botstein, D., White, R. L., Skolnick, M., & Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American journal of human genetics*.
- Çalışkan, M. (2012). *Genetic diversity in plants*. IntechOpen.
- Centro Internacional de Agricultura Tropical CIAT. (2005). *Guía de Estudio: Morfología de la Planta de Arroz*. Cali: CIAT.
- Chen, E., Huang, X., Tian, Z., Wing, R. A., & Han, B. (2019). The Genomics of *Oryza* Species Provides Insights into Rice Domestication and Heterosis. *Annual Review of Plant Biology*.

- Counce, P., Keisling, T., & Mitchell, A. (2000). Uniform, Objective, and Adaptive System for Expressing Rice Development. *Crop Science*.
- Degiovanni, V., Berrío, L., & Charry, R. (2010). Origen, taxonomía, anatomía y morfología de la planta de arroz. En V. Degiovanni, C. Martínez, & F. Motta, *Producción Eco-Eficiente del Arroz en América Latina* (pág. 38). Cali: Centro Internacional de Agricultura Tropical CIAT.
- Ellegren, H., & Galtier, N. (2016). Determinants of genetic diversity. *Nature Reviews Genetics*.
- Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of Molecular Variance Inferred From Metric Distances Among DNA Haplotypes: Application to Human Mitochondrial DNA Restriction Data. *Genetics*.
- FAO. (2008). *FAO Newsroom*. Obtenido de Erosion of plant genetic diversity: http://www.fao.org/newsroom/en/focus/2004/51102/article_51107en.html
- Ferreira, M., & Grattapaglia, D. (1998). Introdução ao uso de marcadores moleculares em análise genética. *EMBRAPA*.
- FONTAGRO. (7 de Octubre de 2020). *FONTAGRO anuncia una nueva iniciativa: Arroz más productivo y sustentable para Latinoamérica*. Obtenido de FONTAGRO anuncia una nueva iniciativa: Arroz más productivo y sustentable para Latinoamérica: <https://www.fontagro.org/es/publicaciones/prensa/comunicados-de-prensa/fontagro-anuncia-una-nueva-iniciativa-arroz-mas-productivo-y-sustentable-para-latinoamerica/>
- García, D. (2019). Informe de Rendimientos Objetivos de Arroz en Cáscara 2019. *Coordinación General de Información Nacional Agropecuaria*.

- Getachew, S. E., Bille, N. H., Bell, J. M., & Gebreselassie, W. (2019). Genotyping by Sequencing for Plant Breeding- A Review. *Advances in Microbiology and Biotechnology*.
- Govindaraj, M., Vetriventhan, M., & Srinivasan, M. (2015). Importance of Genetic Diversity Assessment in Crop Plants and Its Recent Advances: An Overview of Its Analytical Perspectives. *Genetic Research International*.
- Guo, X., & Elston, R. C. (1998). Linkage Information Content of Polymorphic Genetic Markers. *Human Heredity*.
- Henry, R. J. (2013). Evolution of DNA Marker Technology in Plants. En R. J. Henry, *Molecular Markers in Plants*. Wiley-Blackwell.
- Holderegger, R., Kamm, U., & Gugerli, F. (2006). Adaptive vs. neutral genetic diversity: implications for landscape genetics. *Landscape Ecology*.
- Jarvis, D. I., Hodgkin, T., Brown, A. H., Tuxill, J., López Noriega, I., Smale, M., & Sthapit, B. (2016). *Crop Genetic Diversity on the Field and on the Farm*. Bioversity International.
- Jiang, S.-k. H., Zhang, X.-j., Wang, J.-y., Chen, W.-f., & Xu, Z.-j. (2010). Development of a Highly Informative Microsatellite (SSR) Marker Framework for Rice (*Oryza sativa* L.) Genotyping. *Agricultural Science in China*.
- Karlik, E., & Tombuloğlu, H. (2016). Molecular Markers and Their Applications. En K. Rehman Hakeem, H. Tombuloğlu, & G. Tombuloğlu, *Plant Omics: Trends and Applications* (págs. 137-157). Springer Nature.
- Kishor, D., Seo, J., Chin, J. H., & Koh, H.-J. (2020). Evaluation of Whole-Genome Sequence, Genetic Diversity, and Agronomic Traits of Basmati Rice (*Oryza sativa* L.). *Frontiers in Genetics*.

- Koehler, H. A. (1914). *Köhler's Medizinal-Pflanzen*. Franz Eugen Köhler.
- Liu, K., & Muse, S. V. (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics*.
- Mahalingam, A., Saraswathi, R., & Ramalingam, J. (2013). Simple sequence repeat (SSR) markers for assessing genetic diversity among the parental lines of hybrid rice (*Oryza sativa* L.). *African Journal of Biotechnology*.
- Mason, A. S. (2015). SSR Genotyping. En J. Batley, *Plant Genotyping: Methods and Protocols*. New York: Springer.
- Miah, G., Rafii, M. Y., Ismail, M. R., Puteh, A. B., Rahim, H. A., Islam, K. N., & Latif, M. A. (2013). A Review of Microsatellite Markers and Their Applications in Rice Breeding Programs to Improve Blast Disease Resistance. *International Journal of Molecular Sciences*.
- Morillo, E., & Miño, G. (2011). Marcadores Moleculares en Biotecnología Agrícola: Manual de Técnicas y Procedimientos en INIAP. *INIAP*.
- Muhammad Azhar, N., Muhammad Amjad, N., Muhammad Qasim, S., Yıldız, D., Gonul, C., Mehtap, Y., . . . Faheem Shehzad, B. (2017). DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. *Agriculture and Environmental Biotechnology*.
- Muto, C., Ebana, K., Kawano, K., Bounphanousay, V., Bounphanousay, C., Kanyavong, K., . . . Fukuta, Y. (2019). Genetic variation in rice (*Oryza sativa* L.) germplasm from northern Laos. *Breeding Science*.

- Neale, D. B., & Wheeler, N. C. (2019). Neutral Genetic Variation. En D. B. Neale, & N. C. Wheeler, *The Conifers: Genomes, Variation and Evolution*. Cham: Springer.
- Nei, M. (1973). Analysis of Gene Diversity in Subdivided Populations. *Proceedings of the National Academy of Sciences of the United States of America*.
- Nevo, E. (2001). Genetic Diversity. En S. Levin, *Encyclopedia of Biodiversity* (pág. 662). Haifa: Elsevier.
- Organisation for Economic Co-operation Development (OECD) and Food and Agriculture Organization (FAO). (2020). *OECD-FAO Agricultural Outlook 2020-2029*. París: Rome/OECD Publishing.
- Park, S. (2000). *User's manual, Microsatellite ToolKit for MS Excel 97 or 2000 (PC)*. Dublin, Ireland: Molecular Population Genetics Lab, Smurfit Institute of Genetics, Trinity College.
- Peakall, R., & Smouse, P. E. (2012). GenAEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics*.
- Perrier, X., & Jacquemoud-Collet, J.-P. (2006). *DARwin software*. Obtenido de DARwin software: <https://darwin.cirad.fr/index.php>
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics*.
- Rodrigues Oblessucl, P., de CamposII, T., Morini Kupper Cardosol, J., Sforçall, D. A., Moro Baronil, R., Pereira de SouzaII, A., & Lasry Benchimol, L. (2009). Adaptation of fluorescent technique for genotyping with new microsatellite markers in common bean. *Pesquisa Agropecuária Brasileira*.

- Rogers, D., & McGuire, P. (2015). Genetic Erosion: Context Is Key. En M. Ahuja, & S. Mohan Jain, *Genetic Diversity and Erosion in Plants Indicators and Prevention*. Springer.
- Rohlf, J. (2002). Numerical taxonomy and Multivariate analysis system version 2.1. *Department of ecology and evolution state University of New York*.
- Sanchez, P., Wing, R., & Brar, D. (2013). The Wild Relative of Rice: Genomes and Genomics. *Springer Science*.
- Sang, T., & Ge, S. (2007). Genetics and phylogenetics of rice domestication. *Science Direct*.
- Sbordoni, V., Allegrucci, G., & Cesaroni, D. (2012). POPULATION STRUCTURE. En W. B. White, & D. C. Culver, *Encyclopedia of Caves*. Elsevier.
- Schuelke, M. (2000). An economic method for the fluorescent labeling of PCR fragments. *Nature*.
- Semagn, K., Bjørnstad, Å., & Ndjondjop, M. (2006). An overview of molecular marker methods for plants. *African Journal of Biotechnology*.
- Serrote Caetano, M., Reiniger, L., Silva, K., Rabaiolli, S., & Stefanel, C. (2019). Determining the Polymorphism Information Content of a Molecular Marker. *Gene*.
- Serrote, C. M., Reiniger, L. R., & Rabaiolli, S. M. (2019). Determining the Polymorphism Information Content of a Molecular Marker. *Gene*.
- Shahbandeh, M. (20 de Enero de 2021). *Total global rice consumption 2008-2020*. Obtenido de Statista: <https://www.statista.com/statistics/255977/total-global-rice->

- Alvarez, A., Fuentes, J. L., Puldón, V., Gómez, P. J., Mora, L., Duque, M. C., Gallego, G., & Tohme, J. M. (2007). Genetic diversity analysis of Cuban traditional rice (*Oryza sativa* L.) varieties based on microsatellite markers. *Genetics and Molecular Biology*, 30(4), 1109–1117. <https://doi.org/10.1590/S1415-47572007000600014>
- Becerra, V., Paredes, M., Ferreira, M. E., Gutiérrez, E., Díaz, L. M., Becerra, V., Paredes, M., Ferreira, M. E., Gutiérrez, E., & Díaz, L. M. (2017). Assessment of the genetic diversity and population structure in temperate japonica rice germplasm used in breeding in Chile, with SSR markers. *Chilean Journal of Agricultural Research*, 77(1), 15–26. <https://doi.org/10.4067/S0718-58392017000100002>
- Berrio-Orozco, L. E., Torres-Toro, É. A., Barona-Valencia, J., & Cuásquer-Sedano, J. B. (2016). Diversidad genética de las variedades de arroz FLAR liberadas entre 2003-2014. *Agronomía Mesoamericana*, 27(2), 217. <https://doi.org/10.15517/AM.V27I2.20695>
- Breton, C., Pinatel, C., Médail, F., Bonhomme, F., & Bervillé, A. (2008). Comparison between classical and Bayesian methods to investigate the history of olive cultivars using SSR-polymorphisms. *Plant Science*, 175(4), 524–532. <https://doi.org/10.1016/j.plantsci.2008.05.025>
- Caicedo, Y. (2008). *EVALUACIÓN DE CARACTERÍSTICAS AGRONÓMICAS DE CUATRO LÍNEAS INTERESPECÍFICAS DE ARROZ (*Oryza sativa*/*Oryza latifolia*) COMPARADAS CON DOS VARIEDADES COMERCIALES Y UNA NATIVA EN EL UNIVERSIDAD DEL PACÍFICO FACULTAD DE CIENCIAS NATURALES Y EXACTAS PROGRAMA DE. 48.*
- Gallego, S. M., & Benavides, M. P. (2019). Cadmium-Induced Oxidative and

- Nitrosative Stress in Plants. *Cadmium Toxicity and Tolerance in Plants: From Physiology to Remediation*, 233–274. <https://doi.org/10.1016/B978-0-12-814864-8.00010-3>
- Gao, L.-Z., & Gao, C.-W. (2016). Lowered Diversity and Increased Inbreeding Depression within Peripheral Populations of Wild Rice *Oryza rufipogon*. *PLoS ONE*, *11*(3). <https://doi.org/10.1371/JOURNAL.PONE.0150468>
- Gower, J. C. (2005). Principal Coordinates Analysis. *Encyclopedia of Biostatistics*. <https://doi.org/10.1002/0470011815.b2a13070>
- Huang, X., Yang, S., Gong, J., Zhao, Y., Feng, Q., Gong, H., Li, W., Zhan, Q., Cheng, B., Xia, J., Chen, N., Hao, Z., Liu, K., Zhu, C., Huang, T., Zhao, Q., Zhang, L., Fan, D., Zhou, C., ... Han, B. (2015). Genomic analysis of hybrid rice varieties reveals numerous superior alleles that contribute to heterosis. *Nature Communications 2015 6:1*, *6*(1), 1–9. <https://doi.org/10.1038/ncomms7258>
- Islam, M. Z., Khalequzzaman, M., Bashar, M. K., Ivy, N. A., Mian, M. A. K., Pittendrigh, B. R., Haque, M. M., & Ali, M. P. (2018). Variability Assessment of Aromatic Rice Germplasm by Pheno-Genomic traits and Population Structure Analysis. *Scientific Reports 2018 8:1*, *8*(1), 1–14. <https://doi.org/10.1038/s41598-018-28001-z>
- Kosman, E., & Leonard, K. J. (2005). Similarity coefficients for molecular markers in studies of genetic relationships between individuals for haploid, diploid, and polyploid species. *Molecular Ecology*, *14*(2), 415–424. <https://doi.org/10.1111/j.1365-294X.2005.02416.x>
- Kumar, R., Singh, A. K., & Kumar, A. (2012). Evaluation of genetic diversity in rice using simple sequence repeats (SSR) markers. *African Journal of Biotechnology*, *11*(84), 14988–14995. <https://doi.org/10.4314/ajb.v11i84>

- Lawson, D. J., van Dorp, L., & Falush, D. (2018). A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nature Communications* 2018 9:1, 9(1), 1–11. <https://doi.org/10.1038/s41467-018-05257-7>
- MATSUI, T., & KAGATA, H. (2003). Characteristics of Floral Organs Related to Reliable Self-pollination in Rice (*Oryza sativa* L.). *Annals of Botany*, 91(4), 473. <https://doi.org/10.1093/AOB/MCG045>
- Mete Çilingirtürk, A., & Ergüt, Ö. (2014). Hierarchical clustering with simple matching and joint entropy dissimilarity measure. *Journal of Modern Applied Statistical Methods*, 13(1), 329–338. <https://doi.org/10.22237/jmasm/1398918000>
- Nachimuthu, V. V., Muthurajan, R., Duraijalaguraja, S., Sivakami, R., Pandian, B. A., Ponniah, G., Gunasekaran, K., Swaminathan, M., K, S. K., & Sabariappan, R. (2015). Analysis of Population Structure and Genetic Diversity in Rice Germplasm Using SSR Markers: An Initiative Towards Association Mapping of Agronomic Traits in *Oryza Sativa*. *Rice*, 8(1), 30. <https://doi.org/10.1186/S12284-015-0062-5>
- Ngangkham, U., Dash, S., Parida, M., Samantaray, S., Nongthombam, D., Yadav, M. K., Kumar, A., Chidambaranathan, P., Katara, J. L., Patra, B. C., & Bose, L. K. (2019). The potentiality of rice microsatellite markers in assessment of cross-species transferability and genetic diversity of rice and its wild relatives. 3 *Biotech*, 9(6), 1–19. <https://doi.org/10.1007/s13205-019-1757-x>
- Park, J. R., Yang, W. T., Kwon, Y. S., Kim, H. N., Kim, K. M., & Kim, D. H. (2019). Assessment of the genetic diversity of rice germplasms characterized by black-purple and red pericarp color using simple sequence repeat markers. *Plants*, 8(11), 1–12. <https://doi.org/10.3390/plants8110471>

- Pathaichindachote, W., Panyawut, N., Sikaewtung, K., Patarapuwadol, S., & Muangprom, A. (2019). Genetic Diversity and Allelic Frequency of Selected Thai and Exotic Rice Germplasm Using SSR Markers. *Rice Science*, 26(6), 393–403. <https://doi.org/10.1016/J.RSCI.2018.11.002>
- Pérez-Almeida, I., Celi-Herán, R., Sánchez-Mora, F., Paz-Carrasco, L., & Ramos-Viteri, B. (2019). Assessment of molecular genetic diversity of ecuadorian rice cultivars using simple sequence repeat markers. *Bioagro*, 31(1), 3–12.
- Porras-Hurtado, L., Ruiz, Y., Santos, C., Phillips, C., Carracedo, Á., & Lareu, M. V. (2013). An overview of STRUCTURE: applications, parameter settings, and supporting software. *Frontiers in Genetics*, 4(MAY). <https://doi.org/10.3389/FGENE.2013.00098>
- Rice pollen goes the distance. (2014). *Nature* 2014 511:7507, 511(7507), 8–8. <https://doi.org/10.1038/511008d>
- Saitou, N., & Nei, M. (1987). The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees. *Mol. Biol. Evol*, 4(4), 406–425.
- Seetharam, K., Thirumeni, S., & Paramasivam, K. (2009). Estimation of genetic diversity in rice (*Oryza sativa* L.) genotypes using SSR markers and morphological characters. *African Journal of Biotechnology*, 8(10), 2050–2059. <https://doi.org/10.4314/ajb.v8i10.60480>
- SHENG, Y., ZHENG, W., PEI, K., & MA, K. (2005). Genetic Variation Within and Among Populations of a Dominant Desert Tree *Haloxylon ammodendron* (Amaranthaceae) in China. *Annals of Botany*, 96(2), 245. <https://doi.org/10.1093/AOB/MCI171>
- Singh, D. P., Singh, A. K., & Singh, A. (2021). Recurrent selection in self-pollinated

crops. *Plant Breeding and Cultivar Development*, 331–343.

<https://doi.org/10.1016/B978-0-12-817563-7.00007-6>

Suvi, W. T., Shimelis, H., Laing, M., Mathew, I., & Shayanowako, A. I. T. (2020).

Assessment of the genetic diversity and population structure of rice genotypes using SSR markers. *Acta Agriculturae Scandinavica Section B: Soil and Plant Science*, 70(1), 76–86. <https://doi.org/10.1080/09064710.2019.1670859>

Yelome, O. I., Audenaert, K., Landschoot, S., Dansi, A., Vanhove, W., Silue, D., Van Damme, P., & Haesaert, G. (2018). Analysis of population structure and genetic diversity reveals gene flow and geographic patterns in cultivated rice (*O. sativa* and *O. glaberrima*) in West Africa. *Euphytica*, 214(11).

<https://doi.org/10.1007/S10681-018-2285-1>

Capítulo 8: Anexos