

**UNIVERSIDAD NACIONAL AGRARIA
LA MOLINA**

**ESCUELA DE POSGRADO
MAESTRÍA EN MEJORAMIENTO GENÉTICO
DE PLANTAS**



**“CARACTERIZACIÓN MORFOLÓGICA Y MOLECULAR DE
GERMOPLASMA DE TRES RAZAS DE MAÍZ AMILÁCEO
(*Zea mays* L.) DEL PERÚ”**

**Presentada por:
ISAMAR CELEXE BAZO SOTO**

**TESIS PARA OPTAR EL GRADO DE MAESTRO MAGISTER
SCIENTIAE EN MEJORAMIENTO GENÉTICO DE PLANTAS**

Lima - Perú

2024

ManuscritoFinal_tesis_MgSc_IBAZO.pdf

INFORME DE ORIGINALIDAD

6%

INDICE DE SIMILITUD

6%

FUENTES DE INTERNET

3%

PUBLICACIONES

4%

TRABAJOS DEL ESTUDIANTE

FUENTES PRIMARIAS

1	www.tara.tcd.ie Fuente de Internet	3%
2	hdl.handle.net Fuente de Internet	2%
3	repositorio.lamolina.edu.pe Fuente de Internet	1%

Excluir citas Activo

Excluir bibliografía Activo

Excluir coincidencias < 1%

**UNIVERSIDAD NACIONAL AGRARIA
LA MOLINA**

**ESCUELA DE POSGRADO
MAESTRÍA EN MEJORAMIENTO GENÉTICO DE
PLANTAS**

**“CARACTERIZACIÓN MORFOLÓGICA Y MOLECULAR DE
GERMOPLASMA DE TRES RAZAS DE MAÍZ AMILÁCEO
(*Zea mays* L.) DEL PERÚ”**

**TESIS PARA OPTAR EL GRADO DE MAESTRO
MAGISTER SCIENTIAE**

**Presentada por:
ISAMAR CELEXE BAZO SOTO**

Sustentada y aprobada ante el siguiente jurado:

Mg.Sc. Julián Chura Chuquiya
PRESIDENTE

Dr. Raúl Blas Sevillano
ASESOR

Ph.D. Maria de Lourdes Tapia y Figueroa
MIEMBRO

Ph.D. César López Bonilla
MIEMBRO

DEDICATORIA

“Ahora bien, la fe es la garantía
de los bienes que se esperan,
la plena certeza de las realidades
que no se ven”. Hebreos 11:1.

A Dios, en quién residen todas mis fuerzas, voluntad y amor hacia los demás.

Gracias a ti Padre eterno, que siempre me guías hacia los caminos correctos, desde el hogar y la familia que me concediste hasta las amistades tan apreciadas con las que cuento en este momento, honra y gloria a ti por siempre Señor.

AGRADECIMIENTOS

A mis padres, que siempre me han brindado su apoyo incondicional en todos los aspectos, para cumplir mis metas personales y académicas. Asimismo, agradecer a mis hermanos y familiares cercanos por su apoyo moral constante.

Al Dr. Raúl Blas, mi asesor de tesis, por su participación desde la concepción de la investigación hasta la realización y revisión de los análisis, así como la redacción de la tesis.

A mis profesores y amigos, por su apoyo constante para afrontar las dificultades encontradas durante la realización de la investigación, así como las charlas y ánimos que siempre me brindan.

Finalmente, al proyecto “Mejoramiento genético del maíz amiláceo explotando el germoplasma peruano: uso de secuenciamiento de ADN de última generación, y fenotipificación en campos experimentales” financiado por el STC-CGIAR, proyecto el cual financio la presente investigación.

ÍNDICE GENERAL

I.	INTRODUCCIÓN	1
II.	REVISIÓN DE LITERATURA	3
2.1	RAZAS DE MAÍCES PERUANOS	3
2.2	COLECCIONES EN BANCOS DE GERMOPLASMA.....	5
2.2.1	Bancos de germoplasma	5
2.2.2	Recursos genéticos.....	6
2.2.3	Colectas de germoplasma	6
2.2.4	Accesión o entrada.....	7
2.2.5	Caracterización de germoplasma	7
2.2.6	Principal colección nacional de maíz	7
2.3	DIVERSIDAD Y VARIABILIDAD GENÉTICA DEL MAÍZ EN EL PERÚ .	8
2.3.1	Diversidad y variabilidad genética	8
2.3.2	Caracterización morfológica.....	8
2.3.2.1	Caracteres morfológicos	9
2.3.2.2	Caracterización morfológica en maíz amiláceo.....	9
2.3.3	Caracterización molecular	10
2.3.3.1	Marcadores moleculares	11
2.3.3.2	Genotipado por secuenciación (GBS) y SNPs	12
2.3.3.3	Caracterización molecular en maíz.....	13
III.	MATERIALES Y MÉTODOS.....	14
3.1	MATERIAL VEGETAL	14
3.2	MATERIALES USADOS EN LABORATORIO	14
3.3	ÁREA EXPERIMENTAL	17
3.4	CULTIVO DE MAÍZ EN EL ÁREA EXPERIMENTAL.....	17

3.5 CARACTERIZACIÓN MORFOLÓGICA	18
3.5.1 Caracteres cuantitativos	20
3.5.2 Caracteres cualitativos	21
3.6 CARACTERIZACIÓN MOLECULAR	25
3.6.1 Extracción de ADN	25
3.6.2 Determinación de la calidad del ADN.....	26
3.6.3 Determinación de la concentración del ADN.....	27
3.6.4. Desarrollo e identificación de SNPs mediante GBS	27
3.7 ANÁLISIS DE DATOS.....	30
3.7.1 Análisis de datos para la caracterización morfológica	30
3.7.1.1 Análisis de la estadística descriptiva	30
3.7.1.1 Análisis de la estadística multivariada	31
3.7.2 Análisis de datos para la caracterización molecular y estructura genética	35
IV. RESULTADOS Y DISCUSIONES	40
4.1 ANÁLISIS DE LA CARACTERIZACIÓN MORFOLÓGICA.....	40
4.1.1 Análisis de la estadística descriptiva de caracteres cualitativos.....	40
4.1.2 Análisis de la estadística descriptiva de caracteres cuantitativos.....	46
5.1.3 Análisis de correlación de los caracteres cuantitativos	56
5.1.4 Análisis multivariado de agrupamiento jerárquico	59
5.2 ANÁLISIS DE LA CARACTERIZACIÓN MOLECULAR Y ESTRUCTURA GENÉTICA.....	96
V. CONCLUSIONES	122
VI. RECOMENDACIONES	123
VII. REFERENCIAS BIBLIOGRÁFICAS	124
VIII. ANEXOS	144

ÍNDICE DE TABLAS

Tabla 1: Descriptores morfológicos empleados en la caracterización morfológica de accesiones de maíz amiláceo de las razas Piscorunto, Paro y Chullpi.	19
Tabla 2: Tabla de la moda de los caracteres cualitativos evaluados en las razas Paro, Chullpi y Piscorunto.	40
Tabla 3: Prueba de Wilcoxon – Mann – Whitney entre las razas.....	47
Tabla 4: Tabla de coeficientes de correlación cofenético para el agrupamiento jerárquico basado en distancias a partir de caracteres morfológicos.	60
Tabla 5: Tabla del número de accesiones por raza en cada grupo formado mediante el agrupamiento UPGMA y la distancia de Gower.	62
Tabla 6: Tabla del número de accesiones por raza en cada grupo formado mediante el agrupamiento Complete y la distancia de Gower.	64
Tabla 7: Tabla del número de accesiones por raza en cada grupo formado mediante el agrupamiento Single y la distancia de Gower	66
Tabla 8: Tabla del número de accesiones por raza en cada grupo formado mediante el agrupamiento Ward y la distancia de Gower.....	68
Tabla 9: Tabla del número de accesiones por raza en cada grupo formado mediante el HCPC con las componentes del PCA.....	87
Tabla 10 : Tabla del número de accesiones por raza en cada grupo formado mediante el HCPC con las componentes del PCA.....	89
Tabla 11: Tabla del número de accesiones por raza en cada grupo formado mediante el HCPC con las componentes del FAMD.	91
Tabla 12: Tabla del número de accesiones por raza en cada grupo formado mediante el HCPC con las componentes del MFA.....	92
Tabla 13: Tabla de valores de la Diversidad genética de Nei, Heterocigosidad observada, Índice del contenido polimórfico y Frecuencia del alelo menos frecuente de cada raza y de la población.....	101
Tabla 14: Tabla de coeficientes de correlación cofenético de agrupamiento jerárquico basado en distancias a partir de los marcadores SNPs	107
Tabla 15: Matriz de valores Fst por par de razas de maíz amiláceo.....	118
Tabla 16: Matriz de valores Fst por par de grupos determinados con el programa STRUCTURE.....	118

Tabla 17: Matriz de valores F_{st} por par de grupos determinados por el origen de las accesiones.	119
Tabla 18: Tabla de resultados del AMOVA de las accesiones agrupadas de acuerdo con las razas de maíz amiláceo.....	120
Tabla 19: Tabla de resultados del AMOVA de las accesiones agrupadas con el programa STR.....	120
Tabla 20: Tabla de resultados del AMOVA de las accesiones agrupadas de acuerdo al origen.	121

ÍNDICE DE FIGURAS

Figura 1: Clasificación de las razas de maíces del Perú.	4
Figura 2: Distribución de las accesiones de las tres razas de maíz amiláceo.	15
Figura 3: Colores de tallo empleados para evaluar el color de tallo en las razas Paro, Chullpi y Piscorunto.	22
Figura 4: Tipo de disposición de hileras de granos empleado para evaluar este carácter en las razas Paro, Chullpi y Piscorunto.	22
Figura 5: Formas de mazorca empleadas para evaluar este carácter en las razas Paro, Chullpi y Piscorunto.	23
Figura 6: Colores de grano empleados para evaluar el color de grano en las razas Paro, Chullpi y Piscorunto.	23
Figura 7: Tipos de granos empleados para evaluar el tipo de grano en las razas Paro, Chullpi y Piscorunto.	24
Figura 8: Formas de superficies de granos empleados para evaluar la forma de superficie de grano en las razas Paro, Chullpi y Piscorunto.	25
Figura 9: Pasos principales de la técnica de GBS para el análisis de diversidad genética.	28
Figura 10: Gráfico de barras para los caracteres CT, DHG, FM y CM.	42
Figura 11: Gráfico de barras del tipo de grano.	43
Figura 12: Gráfico de barras de la pubescencia de tallo.	43
Figura 13: Gráfico de barras de la forma de la superficie de grano	44
Figura 14: Gráfico de barras del color de grano.	46
Figura 15: Gráfico de cajas para los caracteres DFM y DFF.	48
Figura 16: Gráfico de cajas para los caracteres AM y AP.	49
Figura 17: Gráfico de cajas para los caracteres NHT y NHEM.	50
Figura 18: Gráfico de cajas para los caracteres LH y AH.	51
Figura 19: Gráfico de cajas para los caracteres de mazorca como DAM, DCM, DBM, DM, DR y DME.	53
Figura 20: Gráfico de cajas para los caracteres NH y NGH.	54
Figura 21: Gráfico de cajas para el carácter LM.	55
Figura 22: Gráfico de cajas para los caracteres LG, GG, AG y PG.	56
Figura 23: Coeficientes de Pearson (r-valor) de los caracteres cuantitativos calculados en las razas Chullpi, Paro y Piscorunto.	56

Figura 24: Histograma de la distancia genética de Gower calculada en pares de accesiones en todas las razas.	61
Figura 25: Dendrograma de accesiones de maíz amiláceo empleando el agrupamiento UPGMA y la distancia de Gower.	63
Figura 26: Dendrograma de accesiones de maíz amiláceo empleando el agrupamiento Complete y la distancia de Gower.	65
Figura 27: Dendrograma de accesiones de maíz amiláceo empleando el agrupamiento Single y la distancia de Gower.	67
Figura 28: Dendrograma de accesiones de maíz amiláceo empleando el agrupamiento Ward y la distancia de Gower.	69
Figura 29: Gráfico de barras de la varianza porcentual de las primeras 10 componentes principales obtenidas del PCA.	70
Figura 30: A. Gráfico “Biplot” de correlación entre la calidad de la representación de los caracteres cuantitativos en las dos primeras componentes principales del PCA. B. Correlograma entre la calidad de la representación de los caracteres cuantitativos desde el primer al cuarto componente principal del PCA.	71
Figura 31: A. Gráfico de barras de la contribución de los caracteres cuantitativos en las cuatro primeras componentes principales del PCA. B. Correlograma entre la contribución de los caracteres cuantitativos desde el primer al cuarto componente principa.	73
Figura 32: Gráfico “Biplot” de las accesiones en las dos primeras componentes principales del PCA.	74
Figura 33: Gráfico “Biplot” de las accesiones y caracteres cuantitativos en las dos primeras componentes principales del PCA.	75
Figura 34: Gráfico de barras de la varianza porcentual de las primeras 10 componentes principales obtenidas del MCA.	76
Figura 35: Gráfico “Biplot” de los caracteres cualitativos en las dos primeras componentes principales del MCA.	77
Figura 36: A. Gráfico “Biplot” de correlación entre la calidad de la representación de los caracteres cualitativos en las dos primeras componentes principales del MCA. B. Correlograma entre la calidad de la representación de los caracteres cualitativos desde el primer al décimo tercer componente principal del PCA.	78
Figura 37: A. Gráfico de barras de la contribución de los caracteres cualitativos en las cuatro primeras componentes principales del MCA. B. Correlograma entre la	

contribución de los caracteres cualitativos desde el primer al décimo tercer componente principal del MCA.....	79
Figura 38: Gráfico “Biplot” de las accesiones y caracteres cualitativos en las dos primeras componentes principales del PCA.	80
Figura 39: A. Gráfico “Biplot” de las accesiones y los caracteres cualitativos CE y TG. B. “Biplot” de las accesiones y los caracteres cualitativos CG y FSG.....	81
Figura 40: Gráfico de barras de la varianza porcentual de las primeras 10 componentes principales obtenidas del FAMD.	81
Figura 41: Gráfico “Biplot” de los caracteres cuantitativos y cualitativos en las dos primeras componentes principales del FAMD.	82
Figura 42: A. Gráfico de barras de la contribución de los caracteres cuantitativos y cualitativos en las dos primeras componentes principales del FAMD. B. Correlograma entre la contribución de los caracteres cuantitativos y cualitativos desde el primer al décimo cuarto componente principal del FAMD.	83
Figura 43: Gráfico “Biplot” de las accesiones en las dos primeras componentes principales del FAMD.....	84
Figura 44: Gráfico de barras de la varianza porcentual de las primeras 10 componentes principales obtenidas del MFA.	85
Figura 45: A. Gráfico de barras de la contribución de los caracteres agrupados en las dos primeras componentes principales del MFA. B. “Biplot” de los caracteres agrupados en las dos primeras componentes principales del MFA.	86
Figura 46: Gráfico “Biplot” de las accesiones en las dos primeras componentes principales del MFA.	86
Figura 47: Gráfico de HCPC de accesiones de maíz amiláceo empleando las primeras cuatro componentes principales del PCA.	88
Figura 48: Gráfico de HCPC de accesiones de maíz amiláceo empleando las primeras 13 componentes principales del MCA.....	90
Figura 49: Gráfico de HCPC de accesiones de maíz amiláceo empleando las primeras 14 componentes principales del FAMD.	93
Figura 50: Gráfico de HCPC de accesiones de maíz amiláceo empleando las primeras 13 componentes principales del MFA.	95
Figura 51: Densidad de los marcadores SNP en cromosomas de maíz.	97
Figura 52: Proporción de marcadores SNPs con datos faltantes	98
Figura 53: Proporción de marcadores SNPs con datos faltantes en las accesiones.....	98

Figura 54: Proporción del MAF en los marcadores SNP.	99
Figura 55: Gráficos del método de Evanno para K=1 hasta K=5.....	102
Figura 56: Diagramas de barras multilínea a partir de probabilidades de membresía a cada grupo (cluster) con etiquetado por accesión de k=1 a k=5.....	103
Figura 57: Diagrama de barras de una sola línea (k=3) a partir de probabilidades de membresía a cada grupo (cluster) con etiquetado por accesión y grupos determinados por STRUCTURE a un $PM \geq 0.5$	104
Figura 58: Diagrama de barras de una sola línea (k=3) a partir de probabilidades de membresía a cada grupo (cluster) con etiquetado por accesión y grupos determinados por la raza de maíz.....	105
Figura 59: Diagrama de barras de una sola línea (k=3) a partir de probabilidades de membresía a cada grupo (cluster) con etiquetado por accesión y grupos determinados por el origen de la accesión	106
Figura 60: Dendrograma con distancia euclidiana y agrupamiento UPGMA de accesiones etiquetadas por colores en concordancia con la raza de maíz amiláceo Paro, Chullpi y Piscorunto.....	108
Figura 61: Dendrograma distancia euclidiana y agrupamiento UPGMA de los grupos determinados por la raza de maíz amiláceo.....	109
Figura 62: Dendrograma con distancia euclidiana y agrupamiento UPGMA de accesiones etiquetadas por colores en concordancia con los grupos determinados por STRUCTURE.....	110
Figura 63: Dendrograma distancia euclidiana y agrupamiento UPGMA de los grupos determinados por el programa STRUCTURE.....	111
Figura 64: Dendrograma con distancia euclidiana y agrupamiento UPGMA de accesiones etiquetadas por colores en concordancia con los grupos por origen de las accesiones.	112
Figura 65: Dendrograma distancia euclidiana y agrupamiento UPGMA de los grupos determinados por el origen de las accesiones.....	113
Figura 66: Biplot del número de PC retenidas y la varianza acumulada producto del DAPC.....	114
Figura 67: Biplot del número de clusters y el valor del BIC producto del DAPC.	114
Figura 68: Biplot de PCA de los grupos determinados por las razas de maíz amiláceo.	115
Figura 69: Biplot de PCA de los grupos determinados por STRUCTURE.....	116

Figura 70: Biplot de PCA de los grupos determinados por el origen de las accesiones.

..... 117

ÍNDICE DE ANEXOS

Anexo 1: Datos pasaporte del material vegetal	144
Anexo 2: Evidencia fotográfica de las semillas que corresponden al material vegetal empleado para extracción de ADN.....	149
Anexo 3: Evidencia fotográfica de la cosecha proveniente de los experimentos en campo.....	152
Anexo 4: Geles de calidad de ADN posterior a la extracción.	176
Anexo 5: Datos de cuantificación de ADN del material vegetal.....	179
Anexo 6: Estadística descriptiva para las razas Chullpi, Paro y Piscorunto.....	182
Anexo 7: Datos completos del análisis de varianza de los caracteres cuantitativos de las razas Chullpi, Paro y Piscorunto.....	185
Anexo 8: Coeficientes de correlación de Pearson y prueba de significancia de la correlación de las razas Chullpi, Paro y Piscorunto.	193
Anexo 9: Análisis multivariado de agrupamiento jerárquico empleando distancias genéticas.	196
Anexo 10: Análisis multivariado de agrupamiento jerárquico empleando componentes principales.....	198
Anexo 11: Información del Δk para el análisis de estructura genética con el programa STRUCTURE desde $k=1$ hasta $k=6$	208
Anexo 12: Diagramas de barras multilínea de probabilidades de membresía a cada grupo (cluster) con etiquetado por accesión de $k=1$ a $k=6$	209
Anexo 13: Tabla de probabilidades de membresía (valor q) a los grupos (clusters) generados con el programa STRUCTURE.....	210
Anexo 14: Valores propios, porcentaje acumulado de varianza y coeficientes de correlación de las accesiones a las primeras cinco CP del PCA.	213
Anexo 15: Matrices de p-valor del F_{st} por pares y tablas del bootstrap y los intervalos de confianza para cada F_{st}	219
Anexo 16: Tabla del Análisis Molecular de Variancia para las razas, grupos de STRUCTURE y origen de accesiones.....	222

RESUMEN

Las 120 accesiones de las razas Paro, Chullpi y Piscorunto que se evaluaron en este estudio, presentan una amplia variabilidad morfológica, ya que, desde un enfoque descriptivo básico, se pudieron identificar varios estados en los caracteres cualitativos que anteriormente no habían sido reportados en la literatura. Asimismo, a partir de las comparaciones de la prueba de Wilcoxon – Mann – Whitney obtenido de las medianas de los caracteres cuantitativos, se observó que existen diferencias significativas entre las medianas de las tres razas, lo cual implica que morfológicamente ha sido posible caracterizar a cada una de ellas. A partir de un enfoque multivariado de tipo jerárquico se halló que la distancia de Gower combinado con el agrupamiento de WARD, con un coeficiente de correlación cofenético de 0.833, permitió la agrupación de las accesiones de acuerdo con la raza, sin presencia de duplicados, pero con cinco accesiones ubicadas en una raza diferente. Así como también, empleando un enfoque jerárquico basado en componentes principales se evidenció la presencia de duplicados al realizar el HCPC, al tomar de base el FAMD, siendo este el método que mejor agrupo a las accesiones por raza. Al aplicar diferentes formas del análisis de CP se dedujo que emplear tanto caracteres cuantitativos como los cualitativos se permite una agrupación más adecuada de las accesiones. Adicionalmente, del análisis de CP se obtuvo al TG, CE, CG y FSG como los caracteres cualitativos más discriminantes, y al NH, DME, DR, DAM, DM, NHT, PG, AG y GG como los cuantitativos más discriminantes. Tomando en cuenta que todos los descriptores evaluados son importantes para la correcta agrupación de las accesiones de acuerdo con las razas. Con respecto al análisis de la variabilidad y estructura genética, se identificaron 30873 SNP, los cuales después del filtrado con datos faltantes al 5% en los SNPs, al 20% en los genotipos y con MAF al 10%, se retuvieron 15486 SNPs, los que se emplearon para la caracterización molecular. Índices de diversidad como GD, Ho, PIC y MAF, resultaron con medias de 0.38, 0.31, 0.3 y 0.28 respectivamente, evidenciando la presencia de variabilidad genética molecular en las accesiones. En cuanto a la estructura genética, los tres métodos de agrupación STR, PCA, distancia genética euclidiana/UPGMA que fueron utilizados mostraron cierto nivel de concordancia en el reconocimiento de los grupos por origen y raza. El AMOVA y el índice de Fst, evidenció la baja diferenciación entre los grupos de STR, por origen y raza, así como la presencia de un gran número de genotipos heterocigotos y la alta variabilidad que reside dentro y entre las accesiones.

Palabras claves: razas de maíz amiláceo peruano, Paro – Chullpi, Piscorunto, análisis de diversidad genética, caracterización morfológica y molecular, estructura genética y marcadores moleculares SNP.

ABSTRACT

The 120 accessions of the Paro, Chullpi, and Piscorunto races evaluated in this study exhibit a wide range of morphological variability. In a basic descriptive analysis, several states in the qualitative characteristics were identified that had not been previously reported in the literature. Based on the comparison of the medians of quantitative traits using the Wilcoxon-Mann-Whitney test, it was observed that there are significant differences between the medians of the three races. This implies that it has been possible to morphologically characterize each of the races. From a hierarchical multivariate approach, it was found that using the Gower distance combined with the WARD grouping, which resulted in a cophenetic correlation coefficient of 0.833, allowed for the grouping of accessions based on race. There were no duplicates present but five accessions were found to be in a different racial group. Using a hierarchical approach based on principal components, duplicates were evident when performing the HCPC with FAMD as a basis. This method effectively grouped the accessions by race. By applying different forms of principal components (CP) analysis, it was deduced that utilizing both quantitative and qualitative traits enables a more accurate grouping of accessions. Additionally, from the CP analysis, TG, CE, CG, and FSG were identified as the most discriminating qualitative characters, while NH, DME, DR, DAM, DM, NHT, PG, AG, and GG were identified as the most discriminating quantitative characters. Considering that all the descriptors evaluated are crucial for accurately grouping accessions by race. Regarding the analysis of variability and genetic structure, 30,873 SNPs were identified. After filtering out data with missing values at 5% in the SNPs, 20% in the genotypes, and with a Minor Allele Frequency (MAF) of 10%, 15,486 SNPs were retained for molecular characterization. Diversity indices such as GD, Ho, PIC, and MAF resulted in means of 0.38, 0.31, 0.3, and 0.28, respectively, evidencing the presence of molecular genetic variability in the accessions. Regarding the genetic structure, the three grouping methods - STR, PCA, and Euclidean genetic distance/UPGMA - that were used showed a certain level of agreement in identifying groups based on origin and race. The analysis of molecular variance (AMOVA) and the Fst index indicated low differentiation among the STR groups based on origin and race. Additionally, a significant number of heterozygous genotypes were observed, highlighting the high variability within and between accessions.

Keywords: Peruvian starchy corn races, Paro – Chullpi – Piscorunto, genetic diversity analysis, morphological and molecular characterization, genetic structure, and SNP molecular markers.

I. INTRODUCCIÓN

La investigación sobre recursos genéticos y fitomejoramiento es una de las actividades de innovación más relevantes para la sostenibilidad de la agricultura (Silva *et al.* 2013); debido a que, un conocimiento sólido de la diversidad genética en el germoplasma es vital para la recolección estratégica, el mantenimiento, la conservación y la utilización de esta en la mejora de los cultivos (Rai *et al.* 2013).

El análisis de la variación genética en las colecciones de germoplasma genera un valor agregado para los bancos de germoplasma, haciendo de esta investigación una buena inversión, ya que la información obtenida es fundamental para iniciar un programa de mejoramiento para cualquier tipo de cultivo.

Un análisis bien documentado de la cantidad y los tipos de polimorfismos útiles permite a los curadores de bancos de germoplasma ofrecer accesiones específicas con las características deseadas a los mejoradores de plantas, quienes pueden seleccionar materiales adaptados a sus objetivos y necesidades (Ortiz y Engels 2004), haciendo más fácil y colaborativa la fase temprana del desarrollo de una nueva variedad o cultivar.

La caracterización morfológica viene siendo utilizada desde hace muchos años como la principal herramienta para la generación del conocimiento del estado de la diversidad genética de las colecciones en los bancos de germoplasma, pero en la actualidad ha sido necesario complementar esta información de carácter morfológico con información derivada de estudios moleculares.

Los avances en las tecnologías secuenciación, como de próxima generación (*Next generation*) o “secuenciamientos de alto rendimiento”, han llevado los costos del Secuenciación del ADN al punto que el genotipado por Secuenciación (GBS) es alcanzable para especies de alta diversidad y con grandes genomas (Elshire *et al.* 2011). Esto permite la creciente disponibilidad de información de secuencia para varias especies de plantas y el análisis directo de las mutaciones puntuales que dan lugar a polimorfismos de nucleótido único (SNP).

Las mutaciones puntuales son el tipo más frecuente de variación del ADN intraespecífico (polimorfismo) y se pueden detectar en una cantidad más o menos ilimitada. Con el estado actual del conocimiento, los marcadores de SNP parecen ser mejores para cumplir con los requisitos para el manejo asistido por marcadores del germoplasma que se tiene almacenado (Garner *et al.* 2004).

De manera específica y sobre la diversidad genética de las razas de maíz en nuestro país, Salhuana (2004) menciona que ha sido reconocida de gran importancia para la investigación, y como consecuencia de la información generada, su posterior utilización en mejoramiento de este cultivo; sin embargo, se considera que la falta de información relevante sobre el material vegetal conservado para usuarios potenciales es la causa del bajo uso de las accesiones conservadas en los bancos de germoplasma (Ortiz y Engels 2004).

Tomando en cuenta lo antes mencionado, fue que se analizó la variabilidad genética de tres razas de maíz amiláceo Piscorunto, Paro y Chullpi, que pertenecen a la colección del Programa de Investigación en Maíz de la Universidad Nacional Agraria La Molina, con la finalidad de generar información básica que permita clasificar el germoplasma, promoviendo de esta manera, la conservación y utilización de estos recursos fitogenéticos. Así mismo, estimó y analizó la variabilidad genética empleando la caracterización morfológica, utilizando descriptores vegetativos, de mazorca y grano; y molecular con marcadores tipo SNPs en 120 accesiones de las razas Piscorunto, Paro y Chullpi. Así como, se estimaron los parámetros genéticos y analizó la estructura genética de las razas Piscorunto, Paro y Chullpi.

II. REVISIÓN DE LITERATURA

2.1 RAZAS DE MAÍCES PERUANOS

La raza es un agregado de poblaciones de una especie que tiene en común caracteres morfológicos, fisiológicos y usos específicos. Sin embargo, estas características distintivas no son suficientes para construir una subespecie diferente, por lo tanto, la clasificación en razas es aplicada a especies cultivadas (Sevilla 2006).

Ávila (1996) menciona que, en el maíz peruano, la continua selección practicada por el hombre en las diversas zonas, complementada con la selección natural que permitía la reproducción diferencial de las plantas mejor adaptadas a cada medio, dio como consecuencia, primero la formación de variedades, luego la formación de razas y finalmente complejos raciales altamente diferenciados. Los patrones selectivos muchas veces podían ser coincidentes entre los diferentes pueblos, como por ejemplo el aumento del nivel de productividad, aunque los sistemas para conseguir ese aumento eran diferentes.

Tras estudiar su colección mundial de maíz, Kuleshov (citado por Salhuana 2004) determinó que la mayor diversidad y especialización de maíz del grupo amiláceo (endospermo blando) tenía lugar en el Perú. Más aún, estableció que el grupo más subdividido y rico en rasgos morfológicos y biológicos era el amiláceo (Salhuana 2004).

Salhuana (2004) menciona que para la clasificación racial en Perú se tomaron datos biométricos en las colecciones de maíz en los caracteres de planta, panoja, externos e internos de mazorca, de grano y también características citológicas. La clasificación racial se estableció en base al análisis de esta información. Además, se estableció la distribución geográfica de cada una de ellas y se postuló su origen, así como su relación con razas de otros países. Se hizo estudios de los maíces prehistóricos que ayudaron a establecer postulados de la evolución del maíz en el Perú y por lo tanto fueron de utilidad para establecer la clasificación racial.

La historia evolutiva del maíz en el Perú comprende las siguientes etapas, presentadas aquí en secuencia cronológica: domesticación en las altitudes bajas a medias en los Andes, formación de razas primitivas y extensión del rango original de adaptación de las especies, introgresión de *Tripsacum*; introducción limitada de maíz, que fue domesticado en áreas fuera de la región andina central, hibridación interracial y formación de razas híbridas tempranas, expansión del área de cultivo, mejoramiento de métodos agrícolas, hibridación interracial y formación de razas híbridas, introducción moderna y la formación de razas modernas incipientes (Grobman *et al.* 1961).

De acuerdo con el proceso evolutivo del maíz se ha podido establecer cinco grupos de razas en total:

1. Razas primitivas
2. Razas derivadas de la primitivas
3. Razas de reciente derivación
4. Razas introducidas
5. Razas incipientes

Adicionalmente se tiene a las razas imperfectamente definidas, en la figura 1, se pueden observar la clasificación de las diferentes razas de maíz.



Figura 1: Clasificación de las razas de maíces del Perú

Fuente: Grobman *et al.* (1961).

Sobre la variabilidad del maíz, Grobman *et al.* (1961) indican lo siguiente: el rango general de variabilidad del maíz en el Perú y en las áreas periféricas es mayor que en las otras regiones primarias en el continente americano, tal variabilidad apunta a los Andes centrales como centro de la evolución activa del maíz tanto en el pasado como en el presente; la presencia de formas indígenas antiguas y la gran variabilidad genética del maíz indican que el área peruana es un centro primario de domesticación de esta especie, la domesticación fue seguida de hibridación, introgresión y selección a gran escala, este centro primario es adicional a una región de domesticación primaria independiente en América Central.

Adicionalmente, estos autores mencionan que la mayor parte de la variabilidad del maíz en Perú se puede clasificar en razas, tomando prestado de la definición de Mayr y otros, se puede definir una raza como una población real o potencialmente entrecruzada, una de las varias que pueden formar una especie que se distingue por tener en común ciertos rasgos morfológicos y fisiológicos y por lo tanto también tener en común los genes que determinan estos rasgos.

2.2 COLECCIONES EN BANCOS DE GERMOPLASMA

2.2.1 Bancos de germoplasma

Las colecciones ex situ de recursos fitogenéticos comenzaron con los jardines botánicos de la Edad Media y evolucionaron desde finales del siglo XIX hasta principios del XX en las llamadas "estaciones de introducción". En ese momento, no existían criterios especiales para las metodologías de conservación, dado que la intención era esencialmente usar germoplasma en lugar de conservarlo (De Vicente 2004).

La implementación del mejoramiento genético moderno trajo consigo una mayor necesidad de tener diversidad en los materiales de cruzamiento. Al mismo tiempo, las nuevas variedades, en particular los híbridos, se utilizaron ampliamente debido a sus mayores rendimientos. Esto, a su vez, llevó a la sustitución a gran escala de variedades tradicionales, incluso cultivos, y disminuyó rápidamente las fuentes de recursos genéticos potencialmente valiosos que se encontraban en el campo (De Vicente 2004, Ortiz y Engels 2004).

Esta erosión genética suscitó preocupación por la disponibilidad futura de recursos fitogenéticos. En respuesta, las estaciones de introducción se convirtieron en bancos de genes, lo que dio a los fitomejoradores un acceso más fácil y rápido a un espectro más amplio

de recursos genéticos. Los bancos de genes comenzaron a asumir el papel adicional de ser depósitos de materiales de germoplasma amenazados que podrían ser útiles para los seres humanos (De Vicente 2004, Sevilla y Holle 2004).

Los bancos de genes son reservorios de biodiversidad y fuentes de alelos que pueden recuperarse con relativa facilidad para mejorar los cultivos genéticamente (Ortiz y Engels 2004).

2.2.2 Recursos genéticos

Por “recursos genéticos” se entiende el material genético de valor real o potencial. El acceso a los recursos genéticos estará sometido al consentimiento fundamentado previo de la parte contratante que proporciona los recursos, a menos que esa parte decida otra cosa (ONU 1992).

El rápido desarrollo de la biotecnología moderna en las últimas décadas nos ha permitido utilizar los recursos genéticos de manera que no solo ha alterado de manera fundamental nuestra comprensión del mundo vivo, sino que también ha llevado al desarrollo de nuevos productos y prácticas que contribuyen al bienestar humano. Siendo, desde medicamentos vitales hasta métodos que mejoran la seguridad de nuestros suministros de alimentos. También ha mejorado los métodos de conservación que ayudan a salvaguardar la biodiversidad global (CBD 2010).

2.2.3 Colectas de germoplasma

La colección de germoplasma se diferencia de las exploraciones etnobotánicas, en que las colecciones son de semillas, y se hacen con algún procedimiento de muestreo que permita hacer inferencias sobre la composición de la población. La colección de germoplasma se hace para lograr cualquiera de los siguientes objetivos: la conservación de la diversidad de la especie, obtención de germoplasma para usarlo en el mejoramiento genético y la búsqueda de formas nuevas, o no disponibles en los bancos de germoplasma (Sevilla y Holle 2004).

2.2.4 Accesoión o entrada

Es la unidad de conservación, es una muestra de semillas, o de cualquier órgano reproductivo, que se identifica con un número o código que lo distingue del resto del germoplasma, y que entra o se “accesa” a un banco de germoplasma (Sevilla y Holle 2004).

2.2.5 Caracterización de germoplasma

La caracterización del germoplasma es un proceso que se inicia con la colección o introducción. Debe finalizar con la publicación y la difusión de la información junto con la semilla, para que pueda ser utilizada (Sevilla y Holle 2004). El objetivo principal de la caracterización de recursos fitogenéticos es medir la variabilidad genética de una colección mediante el uso de descriptores definidos, por lo tanto, la caracterización es el primer paso en el mejoramiento de los cultivos y programas de conservación (Hernández 2013).

2.2.6 Principal colección nacional de maíz

La diversidad del maíz en el Perú se viene conservando *ex situ* en el banco de germoplasma de maíz del Programa Cooperativo de Investigación en Maíz (PCIM) desde su creación en 1953, parte de los objetivos de este programa se encuentran relacionados a mejorar y crear variedades amiláceas con alta calidad nutricional para la sierra del país, así como también el coleccionar, mantener y estudiar el germoplasma nacional (PCIM 1997).

El número total de accesiones colectadas e ingresadas al banco es de 3931 procedentes de 55 grupos raciales, todas las accesiones son regeneradas para mantener la semilla viable y algunas en varias ocasiones. De todas las accesiones se conserva la semilla original y se tiene una foto a color de casi todas las accesiones, por ello es posible observar cambios en la frecuencia de genes de color o textura, lo cual se usa como un identificador de la estabilidad genética de las accesiones. La variabilidad de la semilla se revisa constantemente, hasta el 2005 se contaba con un 55.6% de viabilidad de toda la colección, también se tienen seleccionadas 378 accesiones las cuales conforman la colección núcleo (Oscanoa y Sevilla 2004).

2.3 DIVERSIDAD Y VARIABILIDAD GENÉTICA DEL MAÍZ EN EL PERÚ

2.3.1 Diversidad y variabilidad genética

La diversidad de una especie está constituida por todas las variaciones genéticas, producto de la diferencia de las especies. No hay que confundir la diversidad genética con la variabilidad genética que es característica de poblaciones dentro de una especie. La variación entre poblaciones de una especie, más la variación dentro de poblaciones es la diversidad genética total de una especie. Las especies pueden ser más o menos diversas; las características dentro de las poblaciones pueden ser más o menos variables (Sevilla y Holle 2004).

En todos los sectores, la falta de diversidad genética aumenta los riesgos para la seguridad alimentaria y la nutrición, como ya ha ocurrido en el pasado. El uso de la diversidad y de los rasgos deseados seleccionados de esta diversidad, por otro lado, mejora la seguridad alimentaria y la nutrición. Mantener y utilizar una amplia gama de diversidad, tanto la diversidad entre las especies como la diversidad genética dentro de las especies, por lo tanto, significa mantener la capacidad para responder a los desafíos futuros (FAO 2015).

No es posible estimar la diversidad genética de una especie en términos estadísticos o cuantitativos. En general, lo que se hace es clasificar la especie en categorías intraespecíficas como razas y ecotipos; la diversidad genética relativa de una especie en una región se da en términos del número de categorías intraespecíficas (Sevilla y Holle 2004).

2.3.2 Caracterización morfológica

Durante el proceso de caracterización de una especie se estima la variabilidad existente en la población de individuos que la conforman. El objetivo principal es medir la variabilidad genética de una colección mediante el uso de descriptores definidos. Otros objetivos son el establecimiento de la representatividad de la colección, la investigación de la estructura genética a través de la determinación de poblaciones identificables, la identificación de duplicados dentro de una colección y la identificación de genes especiales tanto como alelos particulares (Franco e Hidalgo 2003).

La caracterización morfológica se centra básicamente en los aspectos morfológicos (el uso de marcadores morfológicos) y fenológicos observados en las accesiones y su comparación

con listas de características descriptivas o también llamados descriptores (Ferreira y Grattapaglia 1998).

2.3.2.1 Caracteres morfológicos

Un descriptor de caracterización es una característica morfológica o atributo cuya expresión es fácil de medir, registrar o evaluar y que hace referencia a la forma, estructura o comportamiento de una accesión (Hernández 2003). Cuando todavía no existen los descriptores de caracterización en alguna especie, el primer paso es elegir los caracteres morfológicos y criterios para seleccionarlos, como son el análisis de varianza, coeficiente de variación, coeficiente de correlación y análisis de componentes principales (López *et al.* 2008).

Existen numerosos tipos de caracteres morfológicos, de acuerdo con su relación con la planta tenemos: externos-internos, embriológicos, citológicos, ultraestructurales y según sus características intrínsecas están los continuos-discontinuos. Absolutos-relativos, objetivos-subjetivos, estables-inestables, cuantitativos-cualitativos. Los marcadores morfológicos presentan ciertas limitaciones como son un polimorfismo limitado, alta interacción del genotipo por el ambiente, son con frecuencia caracteres poligénicos, mayor posibilidad de epistasia o pleiotropía y para algunos caracteres es necesario completar el ciclo vital (González 2016).

2.3.2.2 Caracterización morfológica en maíz amiláceo

Las investigaciones alrededor de la caracterización morfológica de recursos fitogenéticos abarcan varios fines, como por ejemplo la definición de nuevas razas, tal como menciona Chávez (2018) quien evaluó 43 accesiones de maíz amiláceo del PCIM de la UNALM y con ayuda de análisis multivariado generó agrupamientos entre las accesiones de acuerdo a los descriptores de mazorca, pero lo cuales no fueron suficientes para postular una nueva raza; otro objetivo para el empleo de caracterización fue realizado por Machaca (2017) en 12 accesiones de maíz tolerantes al frío, este tipo de estudios son cada vez más comunes ya que permite seleccionar las accesiones más sobresalientes en términos de estabilidad y rendimiento con respecto a una característica de importancia agronómica como lo es la tolerancia al frío.

El Instituto de Innovación Agraria (INIA) ha realizado la caracterización morfológica de muchas de las razas de sierra, como el estudio descriptivo que se realizó en las razas Cusco, Paro, San Gerónimo, Huancavelicano Chullpi, Piscorunto, Granada, Morocho y San Gerónimo Huancavelicano, liderado por Oscanoa *et al.* (2006) y en el cual se recopiló información tanto geográfica como morfológica de mazorca para todas las razas antes mencionadas. En otro estudio de la misma índole, realizado por Oscanoa y Sevilla (2008), se caracterizó las 13 razas de las regiones de Junín, Huancavelica y Ayacucho donde además de describir cada una de las razas evaluadas, realizaron un análisis de conglomerados para corroborar los agrupamientos raciales y la formación de compuestos para mantenimiento de semilla.

Existen también otros estudios en los que se han evaluado accesiones de bancos de germoplasma, así se tienen trabajos como los realizados por Catalán (2015) y Yepez (2010), en los cuales se evaluaron y caracterizaron accesiones del banco de germoplasma de la Universidad San Antonio de Abad del Cusco. De la misma manera en conjunto con el PCIM de la UNALM se han realizado números estudios de caracterización de material promisorio para mejoramiento como los ejecutados por Chavarry (2014) para la determinación racial asistida por análisis multivariado en la forma del análisis de componentes principales, Macuri (2016) con la evaluación realizada en accesiones de las razas de sierra baja y media; y Dávalos (2017) en su estudio de 332 accesiones de maíz de razas de selva en la cual describió morfológicamente y agrupó las accesiones en función de diferentes características.

2.3.3 Caracterización molecular

La caracterización molecular tiene los mismos objetivos que la caracterización morfológica, mas no son excluyentes, en la actualidad ambos enfoques de estudio tienen un carácter complementario para el análisis de la diversidad genética. La explotación eficiente y efectiva de la diversidad genética que se encuentra en las colecciones de grandes bancos de genes consume mucho tiempo y recursos. Por lo tanto, las tecnologías moleculares se deben usar para crear subconjuntos representativos de la colección completa o central. Sin embargo, la caracterización filogenética debe continuar para toda la colección, no solo para el núcleo o subconjunto (De Vicente 2004).

A través de estas nuevas tecnologías de caracterización molecular se han obtenido una gran cantidad de marcadores moleculares dispersos a lo largo de todo el genoma, los marcadores

ligados a genes de interés agronómico se pueden utilizar para realizar selección genotípica temprana, en plántulas, y de esta manera se evita manejar cientos o miles de plantas en la selección fenotípica en invernáculo o a campo. La selección asistida por marcadores moleculares es una herramienta muy interesante para los planes de mejoramiento, ya que permite acelerar la selección y disminuir los costos en el manejo de grandes poblaciones de plantas (Masuelli 1999).

2.3.3.1 Marcadores moleculares

Los marcadores moleculares son herramientas importantes para el manejo de bancos de germoplasma, ya que se usan para estimar las relaciones genéticas entre accesiones dentro de una colección de germoplasma. Se pueden identificar y preservar genotipos únicos, o identificar vacíos en la colección con la ayuda de marcadores de ADN, que se pueden utilizar para optimizar el manejo de la diversidad genética (Ortiz y Engels 2004).

Los marcadores basados en el polimorfismo del ADN revelan sitios de variación de la secuencia del ADN. A diferencia de los marcadores morfológicos, las variaciones no se muestran por sí mismas en el fenotipo, porque pueden tener diferencias en un solo nucleótido del gen o en una secuencia repetitiva del ADN. Su polimorfismo sobrepasa varias veces el de otras variantes genéticas incluyendo las isoenzimas (Rodríguez y Arencibia 2002).

Los avances en la genómica de los principales cultivos y especies modelo deben aplicarse para mejorar la comprensión de la variación intraespecífica en las especies y para cubrir nuevas especies. La tecnología basada en marcadores debe implementarse en la gestión de recursos genéticos (De Vicente 2004).

Los marcadores basados en el polimorfismo del ADN o marcadores de ADN se clasifican en dos categorías según la técnica en que se basen (Rodríguez y Arencibia 2003):

Marcadores detectados por hibridación. Aquí se agrupan las técnicas que emplean sondas para la detección de las variantes. Las sondas son fragmentos de ADN o ácido ribonucleico (ARN) que contienen el código complementario para una secuencia específica del genoma. Marcadores tal como RFLP.

Marcadores detectados por amplificación. Los cuales son identificados por amplificación del ADN y están basados en la reacción en cadena de la polimerasa o PCR. Existen diferentes

técnicas que se basan en la PCR, ellas difieren entre sí en la longitud y secuencia de los cebadores empleados, que pueden ser de secuencias arbitraria, semi arbitraria o específica; condiciones de PCR y el método de separación y detección de los fragmentos. Así se tienen marcadores como: RAPD (Random Amplified Polymorphic DNA), SCAR (Sequences Characterized Amplified Regions), AFLP (Amplified Fragment Length Polymorphism), CAPS (Cleaved Amplified Polymorphic Sequence) o también llamado RCR-RFLP, SSR (Simple Sequence Repeat), RAMPO (Random Amplified Microsatellite Polymorphism) y SNP (Single Nucleotide Polymorphism).

2.3.3.2 Genotipado por secuenciación (GBS) y SNPs

Avances en las tecnologías de próxima generación (*Next generation*) han llevado los costos de la Secuenciación del ADN al punto que el genotipado por Secuenciación (GBS) es ahora alcanzable para especies de alta diversidad y con grandes genomas (Elshire *et al.* 2011). Esto permite la creciente disponibilidad de información de secuencia para varias especies de plantas y el análisis directo de las mutaciones puntuales que dan lugar a polimorfismos de nucleótido único (SNP). Las mutaciones puntuales son el tipo más frecuente de variación del ADN intraespecífico (polimorfismo) y se pueden detectar en una cantidad más o menos ilimitada. Con el estado actual del conocimiento, los marcadores de SNP parecen ser mejores para cumplir con los requisitos para el manejo asistido por marcadores de los recursos genéticos (Graner *et al.* 2004).

El enfoque de GBS se basa en la reducción del genoma con enzimas de restricción, para esto no se requiere de un genoma de referencia para el desarrollo de SNPs. GBS es un proceso combinado de descubrimiento de genotipos y genotipado en un solo paso y proporciona una herramienta rápida, de alto rendimiento y rentable para un genoma. Además, es aplicable a una amplia gama de especies no modelo (He *et al.* 2014, Peterson *et al.* 2014).

De acuerdo con Peterson *et al.* (2014), GBS se puede considerar como uno de los varios métodos de secuenciación de representación reducida para el genotipado. Involucra cinco componentes principales (CP): preparación de muestras, ensamblaje de bibliotecas, secuenciación, identificación de SNP y análisis de diversidad. Los pasos principales incluyen la extracción y digestión de ADN, la ligación del adaptador, la amplificación por PCR, la selección del tamaño de los fragmentos, la agrupación de bibliotecas, la secuenciación, el procesamiento de datos, la identificación de SNP y el análisis de diversidad genética. Estos

pasos pueden variar en el uso de endonucleasas de restricción, la plataforma de Secuenciación y en el análisis bioinformático para el genotipado de SNP.

2.3.3.3 Caracterización molecular en maíz

En la actualidad existe una variedad bastante amplia de estudios enfocados en el análisis de la diversidad genética a través de la caracterización molecular con marcadores SNPs. Uno de los estudios más recientes con respecto a germoplasma nativo de maíz, es el Bedoya *et al.* (2017) en el cual se analizaron landraces de diferentes partes de Latinoamérica y del caribe, así como también el realizado por Van Heerwaarden *et al.* (2011).

Dentro de los estudios que abarcan landraces americanos, la mayoría de ellos se han realizado empleando marcadores de tipo SSR, AFLP y RAPD, así se tienen estudios como los de Bracco *et al.* (2014, 2009) que evaluaron maíces de nativos de Argentina, los de Pineda-Hidalgo *et al.* (2013) y Reif *et al.* (2006) con razas de maíces de México y Molin *et al.* (2013) en maíces nativos de Brasil.

Además del germoplasma nativo, también existen estudios que se enfocan en la caracterización de líneas endogámicas de los programas de mejoramiento como los realizados en China por Leng *et al.* (2019) con marcadores SNPs y Wang *et al.* (2008) con marcadores SSR, también está el trabajo realizado por Wu *et al.* (2016) que evaluaron 538 accesiones provenientes de diferentes partes del mundo incluido material del CIMMYT. En Europa también se han caracterizado este tipo de material siendo Frascoli *et al.* (2013) uno de los muchos estudios que se tienen más recientes al respecto. Gouresnard *et al.* (2017) hace una comparación de líneas endogámicas de origen europeo con germoplasma mantenido por el USDA. Dentro de los estados Unidos también se han realizado este tipo de estudios, así Schaefer *et al.* (2013) evaluaron líneas históricas del programa de mejoramiento de la Universidad de Minnesota y las compararon con las líneas del actual programa.

En estudios más recientes, la caracterización de los recursos genéticos vegetales se encuentra dirigida a líneas endogámicas avanzadas para el mejoramiento de factores de estrés abiótico y biótico, como las líneas evaluadas por Boakyewaa *et al.* (2019) mejoradas para diversos factores que influyen la maduración temprana, las estudiadas por Dinesh *et al.* (2016) con tolerancia al calor, Dao *et al.* (2014) y Lanes *et al.* (2014) con líneas adaptadas a maíces tropicales y Eritro *et al.* (2017) con líneas adaptadas a altitudes medias en África.

III. MATERIALES Y MÉTODOS

3.1 MATERIAL VEGETAL

En este estudio se emplearon accesiones provenientes del banco de germoplasma de maíz del Programa de Investigación en Maíz de la UNALM, estas accesiones proceden de varias regiones de la sierra de Perú, en la Figura 2 se puede observar la ubicación de cada una de las accesiones en el mapa del territorio peruano, y se utilizaron 40 accesiones por raza de maíz amiláceo, siendo estas razas: Piscorunto, Paro y Chullpi. Adicionalmente, la información detallada de departamento, provincia, distrito, latitud, longitud y altitud de cada accesión se encuentra en el Anexo 1.

3.2 MATERIALES USADOS EN LABORATORIO

Reactivos químicos:

- Ácido Acético Glacial
- Agua Destilada
- Alcohol Isoamílico
- Bromuro de Hexadeciltrimetil Amonio (CTAB)
- β Mercaptoetanol
- Cloroformo
- Cloruro de Sodio
- EDTA (Ácido Etilen Diamino Tetra-Acético)
- Etanol Absoluto
- Isopropanol
- Nitrógeno Líquido
- RNAase A
- PVP (Polyvinilpyrrolidone)
- Tris (Tris Hidroxymethyl Amino Ethan)

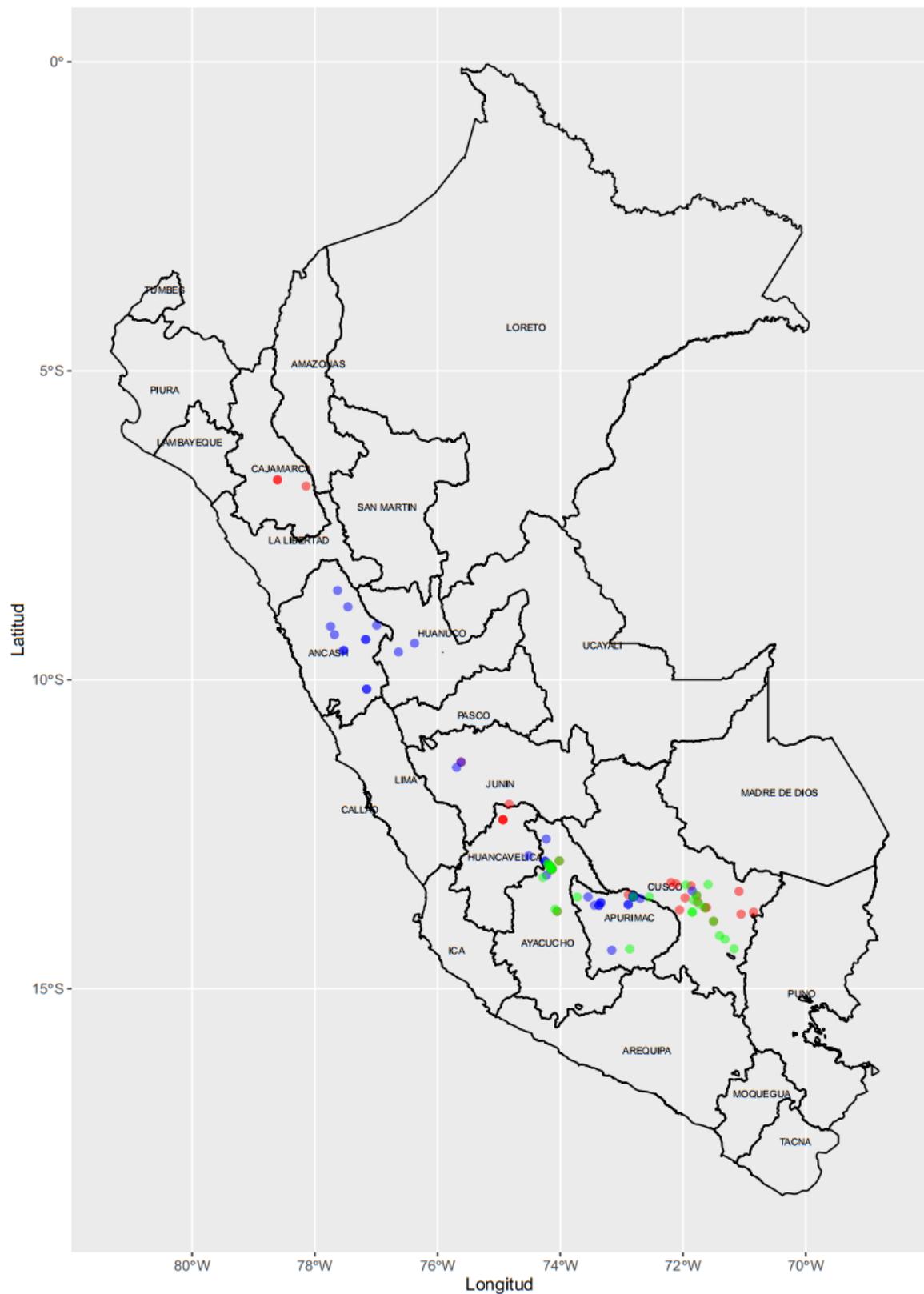


Figura 2: Distribución de las accesiones de las tres razas de maíz amiláceo.

*Raza Chullpi (color rojo), Paro (color azul) y Piscorunto (verde)

Fuente: elaboración propia.

Material de laboratorio

- Agitadores magnéticos
- Espátulas, pinzas y tijeras
- Contenedor de nitrógeno líquido
- Guantes de vinilo
- Magnetos
- Matraces erlenmeyer
- Morteros de porcelana con pistilos
- Pipetores o pipetas
- Puntas plásticas de 10, 200 y 1000 μ l
- Probetas 10, 100, 1000 ml.
- Racks polipropileno
- Tubos para microcentrífuga de 1.5 ml (Eppendorf)
- Vasos de precipitados

Equipos

- Autoclave
- Balanza analítica
- Bomba de vacío
- Cámara de baño maría
- Congelador a -20°C
- Estufa
- Horno microondas
- Potenciómetro
- Refrigeradora
- Vórtex

Electroforesis horizontal

Reactivo - químicos

- Agarosa
- Tris (Tris Hidroxymethyl Amino Ethan)
- Boro
- EDTA (Ácido Etilen Diamino Tetra-Acético)

Materiales de Laboratorio

- Tubos Eppendorf
- Pipetas
- Placas de 96 pocillos, capacidad de 200 μ l
- Racks polypropylne (soporte de tubos Eppendorf)
- Vasos precipitados

Equipos

- Cámara de electroforesis horizontal
- Fuente de poder
- Horno microondas

3.3 ÁREA EXPERIMENTAL

El experimento para llevar a cabo caracterización morfológica se realizó en un campo de cultivo localizado en el distrito de Chiquián, en la provincia de Bolognesi, en la región de Ancash a 3374 msnm, cuya posición geográfica es 10°9'4" Sur y 77°9'22" Oeste.

3.4 CULTIVO DE MAÍZ EN EL ÁREA EXPERIMENTAL

La unidad experimental estuvo constituida por un surco con dimensiones de 10 m de largo y entre 0.8-0.9 de ancho aproximadamente. Se empleo un surco por accesión, en cada raza, empleando un diseño experimental en DBCA, en el cual se emplearon dos bloques, es decir dos repeticiones, cada bloque contenía 40 surcos (un surco por accesión) más 8 surco que correspondían a los testigos.

En cada surco se sembraron 3 semillas por golpe, entre los cuales hubo 0.4 m de separación con un total de 24 golpes por surco aproximadamente. Durante el desahije, se dejó dos plantas por golpe.

El suelo se preparó previamente a la siembra, con el arado, nivelado del suelo y finalmente el surcado, cuyas dimensiones fueron mencionadas previamente. La siembra se realizó en la última semana de diciembre del 2018.

Se realizaron dos abonamientos, a los 15 días y a los 45 días empleando fuentes de NPK, como urea, fosfato di amónico y cloruro de potasio. El control de las malezas se realizó con herbicidas a base de atrazina en dos aplicaciones a los 30 días y a los 60 días aproximadamente. Adicionalmente también se realizó el deshierbo manual y se mantuvo el control de plagas e insectos a lo largo del cultivo.

La cosecha se realizó al llegar a la etapa de madurez fisiológica, a mediados de agosto del 2019. Dentro de cada parcela solo se cosecharon las mazorcas que fueron producto de la polinización manual (cruzadas de planta a planta, es decir, hermanos completos). Luego de cosechadas fueron secadas al sol y a las dos semanas traídas al programa de maíz de la UNALM para la caracterización morfológica de mazorca.

3.5 CARACTERIZACIÓN MORFOLÓGICA

Para la caracterización morfológica, se seleccionaron 30 descriptores del manual de descriptores propuesto por el CIMMYT-IBPGR para el maíz (IBPGR 1991), pero se efectuaron algunas modificaciones durante la toma de datos, siendo detalladas más adelante, del total de descriptores 21 fueron cuantitativos y nueve cualitativos.

Tabla 1: Descriptores morfológicos empleados en la caracterización morfológica de accesiones de maíz amiláceo de las razas Piscorunto, Paro y Chullpi

Caracteres cuantitativos	Abreviatura
Días a la floración masculina	DFM
Días a la floración femenina	DFF
Altura de planta	AP
Altura de mazorca	AM
Numero de hojas por planta	NHT
Número de hojas encima de la mazorca	NHEM
Longitud de hoja	LH
Ancho de hoja	AH
Longitud de mazorca	LM
Diámetro del ápice de la mazorca	DAM
Diámetro del centro de la mazorca	DCM
Diámetro de la base de la mazorca	DBM
Número de hileras	NH
Número de granos por hilera	NGH
Diámetro de marlo	DM
Diámetro de raquis	DR
Diámetro de médula	DME
Longitud de grano	LG
Ancho de grano	AG
Espesor o grosor de grano	GG
Peso de 100 granos	PG
Caracteres cualitativos	
Color de tallo	CT
Pubescencia de tallo	PT
Disposición de hilera de granos	DHG
Color de marlo o elote	CM
Forma de mazorca	FM
Color de grano	CG
Color de endospermo	CE
Tipo de grano	TG
Forma de la superficie del grano	FSG

3.5.1 Caracteres cuantitativos

Descriptores de estado vegetativo, estos se evaluaron en 10 plantas competentes del total de plantas que se encontraban en un surco por accesión

- a. Días a la floración masculina: se tomó el número de días transcurridos desde la siembra hasta que más del 50% de las plantas de una accesión por bloque están emitiendo polen.
- b. Días a la floración femenina: se registró el número de días transcurridos desde la siembra hasta que más del 50% de las plantas de una parcela tengan estigmas como mínimo de 2 centímetros de largo.
- c. Altura de planta: se tomó la medida desde la base de la planta, es decir desde el nivel del terreno hasta la base de la panoja en 10 plantas competentes por accesión en cada bloque.
- d. Altura de mazorca: en las mismas 10 plantas tomadas para medir altura de plantas, se tomó la medida de altura desde la base de la planta hasta el nudo donde nace la mazorca superior.
- e. Numero de hojas por planta: se evaluó después de la floración y esto se realizará en las mismas 10 plantas de (c).
- f. Número de hojas encima de la mazorca: se contaron después de estado lechoso y en 10 plantas a partir de la mazorca más alta en las mismas que en (c).
- g. Longitud de hoja: se midió desde a lígula hasta el ápice de la hoja que sobresale de la mazorca más alta. Después de la floración en 10 plantas por accesión siendo las mismas que en (c).
- h. Ancho de hoja: se midió en las mismas hojas seleccionadas en (g), en el punto medio de su longitud.

Descriptores de mazorca, estos se evaluaron en 10 mazorcas por accesión, estas mazorcas fueron extraídas de la posición más alta de cada planta. En el Anexo 3 se encuentran las fotografías de las mazorcas más representativas de cada accesión que fueron cosechadas para cada raza.

- i. Longitud de mazorca: se desgranó completamente la mazorca y se midió la longitud.
- j. Diámetro del ápice de la mazorca: se desgranó completamente la mazorca y se midió el diámetro.
- k. Diámetro del centro de la mazorca: se desgranó completamente la mazorca y se midió el diámetro.
- l. Diámetro de la base de la mazorca: se desgranó completamente la mazorca y se midió el diámetro.
- m. Número de hileras: se contarán las hileras de granos en la parte central de la mazorca más alta.
- n. Número de granos por hilera: se contaron el número de granos por hilera de 5 hileras por mazorca evaluada.
- o. Diámetro de marlo: se realizó la medición en la posición donde se midió el diámetro de la base de la mazorca.
- p. Diámetro de raquis: se realizó la medición donde se midió el diámetro de la base de la mazorca.
- q. Diámetro de medula: se realizó la medición en la posición donde se midió el diámetro de la base de la mazorca.

Descriptores de grano, estos se evaluaron en una muestra de 10 granos provenientes de la mezcla compuesta por 100 granos tomados de 10 plantas de una misma accesión.

- r. Longitud de grano: este fue el promedio de 10 granos consecutivos de una hilera en el punto medio de la mazorca más alta.
- s. Ancho de grano: se midieron los mismos 10 granos que el (r).
- t. Espesor o grosor de grano: se midieron los mismo 10 granos que el (r).
- u. Peso de 100 granos, esto se midió en granos ajustados a un contenido de humedad entre 10 – 20%

3.5.2 Caracteres cualitativos

- a. Color de tallo: en el momento de la floración, se observó el color de tallo entre las dos mazorcas más altas de 10 plantas. Los colores considerados son verde, morado y café, una referencia de estos se observa en la Figura 3.



Figura 3: Colores de tallo empleados para evaluar el color de tallo en las razas Paro, Chullpi y Piscorunto

- b. Pubescencia de tallo: se evaluó en el momento de la floración, donde los niveles de pubescencia son: escasa, intermedia y densa.

Descriptores de mazorca, estos se evaluaron en 10 mazorcas por accesión, estas mazorcas fueron extraídas de la posición más alta de cada planta.

- c. Disposición de las hileras de granos: se usó la mazorca más alta por planta y se evaluarán 10 mazorcas, de acuerdo con las siguientes categorías nominales: regular, irregular, recta y espiral (ver Figura 4).

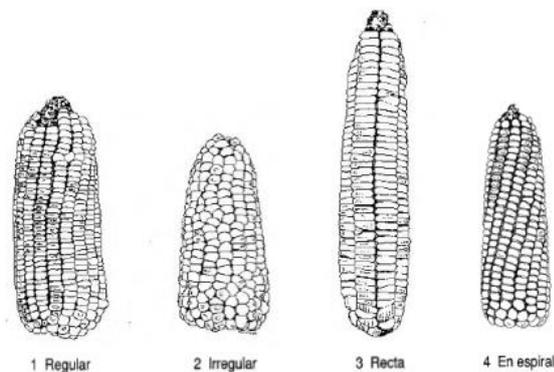


Figura 4: Tipo de disposición de hileras de granos empleado para evaluar este carácter en las razas Paro, Chullpi y Piscorunto

Fuente: Rincón, 2020.

- d. Color de marlo o elote: se indicó como máximo tres colores en orden de frecuencia. Se emplearon los mismos colores que en el carácter de color de grano.
- e. Forma de mazorca fue descrita de acuerdo con las siguientes categorías nominales, cónica, cónica-cilíndrica y cilíndrica, según se ilustra en la Figura 5.

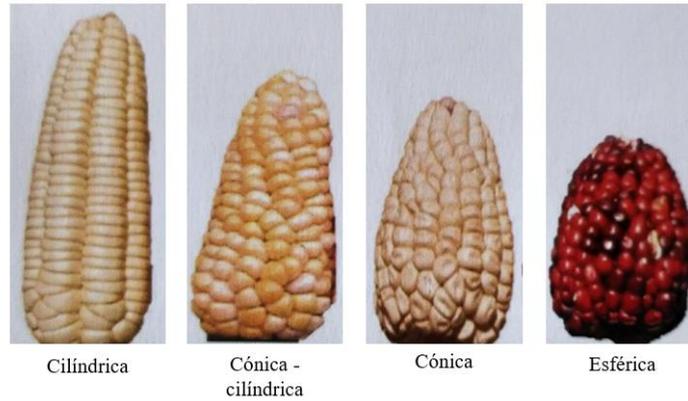


Figura 5: Formas de mazorca empleadas para evaluar este carácter en las razas Paro, Chullpi y Piscorunto

Descriptores de grano, se evaluaron en una muestra de 10 granos provenientes de la mezcla compuesta por 10 granos tomados de 10 plantas de una misma accesión.

- f. Color de grano: se indicó como máximo tres colores en orden de frecuencia. Los colores tomados como referencia son los siguientes: blanco, amarillo, morado, jaspeado, café, anaranjado, moteado, capa blanca y rojo, según se ilustra en la Figura 6.

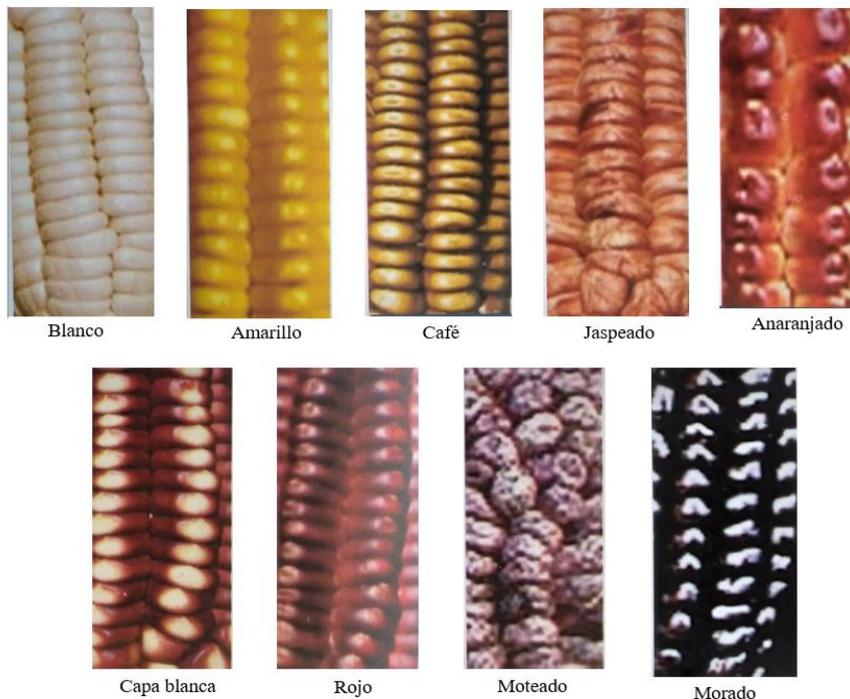


Figura 6: Colores de grano empleados para evaluar el color de grano en las razas Paro, Chullpi y Piscorunto

Fuente: Rincón, 2020.

- g. Color de endospermo: se indicó como máximo tres colores en orden de frecuencia. Se emplearon los mismos colores que en el carácter de color de grano.
- h. Tipo de grano: se indicó como máximo tres tipos de grano en orden de frecuencia. Las categorías nominales empleadas como referencia fueron: harinoso, semi harinoso, semi cristalino, cristalino, reventador, dulce, ceroso. según se ilustra en la Figura 7.

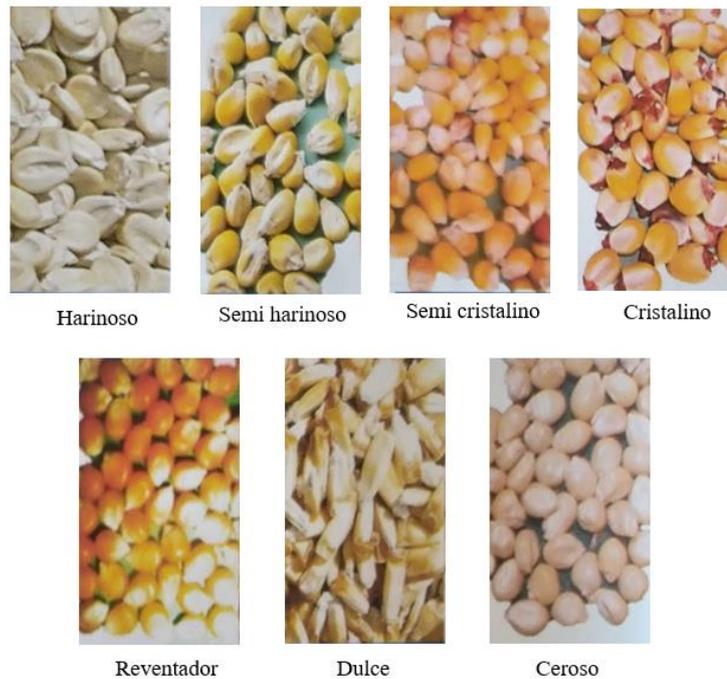


Figura 7: Tipos de granos empleados para evaluar el tipo de grano en las razas Paro, Chullpi y Piscorunto

Fuente: Rincón, 2020.

- i. Forma de la superficie del grano: se indicó como máximo tres categorías nominales en orden de frecuencia. Las categorías tomadas como referencia son: contraído, dentado, plano, redondo, puntiagudo, muy puntiagudo, según se ilustra en la Figura 8.

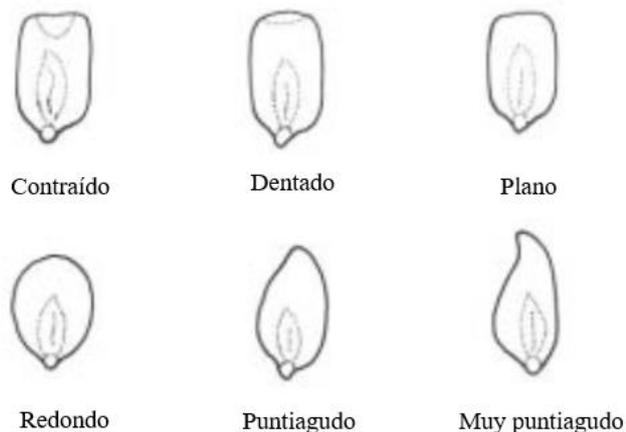


Figura 8: Formas de superficies de granos empleados para evaluar la forma de superficie de grano en las razas Paro, Chullpi y Piscorunto

Fuente: IBPGRI, 1991.

3.6 CARACTERIZACIÓN MOLECULAR

3.6.1 Extracción de ADN

La extracción de ADN de hojas jóvenes de cada accesión se realizó mediante el protocolo con buffer CTAB (Bromuro de hexadecil trimetil amonio) establecido por Doyle y Doyle (1990) con algunas modificaciones, esto se realizó en el laboratorio de Genómica y bioinformática del Instituto de Biotecnología de la UNALM. Las fotografías de las semillas empleadas de cada accesión para la obtención de hojas jóvenes para la extracción de ADN se encuentran en el Anexo 2. Se tomó una hora de la planta en la semana 4 desde la germinación.

La extracción de ADN para cada individuo consistió en los pasos siguientes:

1. En un mortero frío se trituró aproximadamente 100 mg de hoja con 700 μ l de CTAB-buffer de lisis 2X (20 g. CTAB, 81.81 g. NaCl, 40 ml. 0.5 M EDTA a pH 8.0, 100 ml. de 1M Tris:HCl a pH 8.0 y 10 g. de PVP para 1 L.) y 2 μ l de β -mercaptoetanol.
2. El líquido resultante de la trituración fue transferido a un tubo Eppendorf y el contenido del tubo se mezcló por inversión.
3. Posteriormente, el tubo Eppendorf fue incubado a 65 °C durante 45 minutos, agitándose cada 15 minutos.

4. Después de la incubación, se dejó enfriar el tubo Eppendorf para añadir 700 µl de cloroformo:alcohol isoamílico (24:1), mezclándose por inversión. Esta mezcla fue centrifugada a 14000 rpm por 8 minutos.
5. El sobrenadante resultante de la centrifugación fue transferido a un nuevo tubo Eppendorf.
6. Se agregó al nuevo tubo 50 µl de CTAB 10X (10 g. CTAB, 5 g. NaCl, 10 ml. 0.5 M EDTA a pH 8.0 para 100 ml.) y se mezcló por inversión. Luego de obtener una mezcla homogénea se añadió al tubo 700 µl de cloroformo:alcohol isoamílico (24:1) y se mezcló por inversión y se centrifugó a 14000 rpm por 8 minutos.
7. El sobrenadante resultante de la centrifugación fue transferido a un nuevo tubo Eppendorf.
8. Se agregó al nuevo tubo 500 µl de etanol absoluto frío y se mezcló suavemente hasta ver la medusa de ADN.
9. Esta mezcla fue llevada a una temperatura de -20 °C por unos 30 minutos y luego se centrifugó el contenido del tubo a 14000 rpm por 20 minutos.
10. Posteriormente se eliminó el sobrenadante, dejando secar el tubo. El precipitado fue lavado dos veces con un mililitro de etanol de 70 % y al 90% respectivamente a 14000 rpm por 8 minutos.
11. Después se dejó que el precipitado seque a temperatura ambiente durante toda la noche.
12. Finalmente, se suspendió el precipitado, dependiendo de su tamaño, en 50-100 µl de T10E0.1 (100 ml. de 1M Tris: HCl a pH 8.0, 0.5 M EDTA a pH 8.0 para 400 ml.). Al término de este procedimiento se almacenó a -20 °C.

3.6.2 Determinación de la calidad del ADN

La determinación de la calidad del ADN en las muestras extraídas se realizó en geles de agarosa al 1 por ciento (2,5 g. Agarosa en 250 ml. TBE 1X y 2 µl de bromuro de etidio concentrado) en una cámara electroforética horizontal sumergido en buffer TBE 1X (buffer TBE 10X- 108 g. de Tris Base, 55 g de Ácido Bórico, 40ml. de 0.5 M EDTA a pH 8 para 1L.). Las fotos de los geles de calidad de ADN realizados para las tres razas se encuentran en el Anexo 4.

Para ello se realizó previamente una dilución 1:10 de 1 µl de ADN extraído con 9 µl de tampón de carga SALB 1X (SALB 10X- 75 mg. Azul de Bromofenol, 75 mg. Xilene Cianol, 100 mg. Orange G, 2.5 ml. TBE10X y 30 g. Sucrosa en 50 ml.).

Esta dilución se cargó en el gel de agarosa previamente preparado y se corrió con ayuda de una fuente de poder a 100 v durante 40 minutos y posteriormente el gel se visualizó en un transiluminador de UV.

3.6.3 Determinación de la concentración del ADN

Con el fin de eliminar cualquier interferencia en este procedimiento, cada muestra de ADN fue sometida a la acción de 2 ul de la enzima RNAsa (10 mg/ml). Luego, se utilizó 2 ul de cada muestra para medir su concentración en un espectrofotómetro NanoDrop™ (ThermoFisher™). El espectrofotómetro proporcionó la concentración en ng/µl y la relación A260/A280 (tasa de absorbancias a 260 nm sobre 280 nm), considerada como pureza del ADN, la cual debe de estar muy cercana a 2.8 para una óptima calidad (Lee *et al.* 2016). Los datos de cuantificación por espectrofotometría para cada accesión se encuentran en el Anexo 4, en este mismo anexo se encuentran los datos de cuantificación por fluorometría realizado por la entidad que también realizó el servicio de Secuenciación.

3.6.4. Desarrollo e identificación de SNPs mediante GBS

El servicio de Secuenciación fue realizado por el Centro de Genómica de la Universidad de Minnesota, laboratorio el cual siguió el protocolo descrito por Elshire *et al.* (2011) para el desarrollo de SNPs, esto se realizó en el panel de 120 accesiones de las tres razas de maíz amiláceo antes mencionadas.

Existen numerosas metodologías para el análisis de diversidad mediante el desarrollo de SNP por medio del GBS, pero todas cuentan con al menos cinco pasos básicos para el desarrollo de la técnica (Figura 9), los cuales fueron empleados para identificar a los SNP. Estos pasos fueron los siguientes: preparación de muestras, ensamblaje de librerías, secuenciación, estos dos últimos pasos, como se mencionó con anterioridad, fueron realizados como servicio por un laboratorio especializado; posteriormente se realizó la identificación de SNPs, o también denominado *variant calling* (llamado de variantes SNP) y finalmente el análisis de variabilidad Peterson *et al.* (2014).

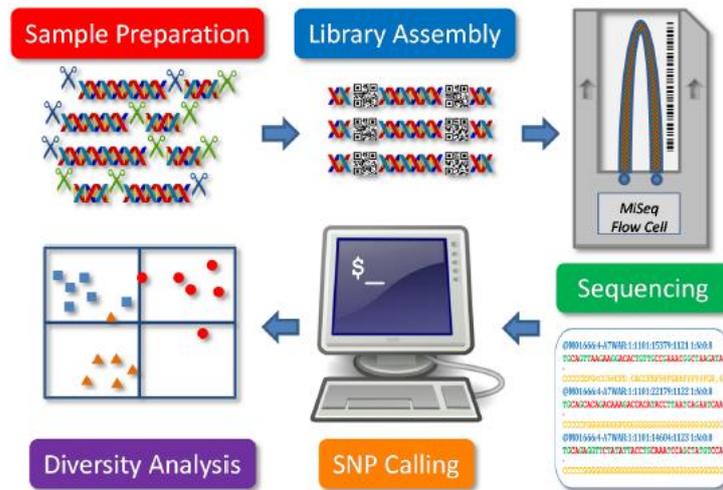


Figura 9: Pasos principales de la técnica de GBS para el análisis de diversidad genética

Fuente: Peterson *et al.* (2014).

Para la creación de las librerías y su posterior ensamblaje, la complejidad del genoma en el maíz se redujo al digerir el ADN genómico de cada muestra con *ApeKI*, la cual es una enzima de restricción sensible a la metilación. Los fragmentos resultantes de todas las muestras se ligaron directamente a un par de adaptadores específicos para los cortes dejados por las enzimas de restricción; luego de esto se adicionó el código de barras, se cuantificó y se seleccionaron los fragmentos con tamaño específico (entre 300-400 pb) para la posterior secuenciación.

La amplificación por PCR se llevó a cabo para generar las librerías genómicas, las cuales se secuenciaron con la tecnología de Illumina HiSeq 2500 (Illumina Inc., EE. UU.). La secuenciación produjo millones de lecturas divididas en varios archivos FASTQ (Elshire *et al.* 2011).

Luego, de acuerdo con Schnable *et al.* (2009), el conjunto completo de *reads* o fragmentos de secuenciación, identificados por un único código de barras, se fusionaron para generar un archivo FASTQ maestro, este fue evaluado y filtrado en función a la calidad de los *reads* y después de ello se procedió al indexarlo con el genoma de referencia de *Zea mays* B73 RefGen_v4 (Jiao *et al.* 2017).

La identificación de los SNPs (*variant calling*) estuvo a cargo del Dr. Carlos Arbizu en el marco del proyecto que ha financiado el presente trabajo de tesis. Para esto se realizó la identificación implementando el programa TASSEL v5.0 (Badbury *et al.* 2007), y con ayuda

de este programa se realizó el alineamiento de los *reads* al genoma de referencia B73, además de la identificación de los SNP en su fase de descubrimiento y de asociar la presencia de estos SNPs a cada uno de los individuos fenotipados en su fase de producción (Bradbury *et al.* 2007).

De aquí en adelante, a los genotipos se les denominara “accesiones” para no perder la denominación que también se le dio en la caracterización morfológica, ya que estos genotipos pertenecen a las accesiones provenientes del Banco de Germoplasma de la UNALM. Una vez empleado el programa TASSEL se obtuvo como producto final un archivo en formato VCF (*variant calling format*), en formato VCF v4.0 (EMBL-EBI 2021), que contenía todos los SNP identificados en este paso y que se utilizaron en el análisis de variabilidad.

A partir de esto, se realizó el análisis de diversidad, el cual inició con el proceso de filtrado e imputación de datos en los SNPs. Para ello, el archivo VCF antes mencionado fue sometido a un primer filtro, el cual utiliza la información de datos faltantes que existe para cada marcador SNP, para esto aquellos marcadores en los que no se logró amplificar la señal que indique la presencia de un nucleótido en específico en las accesiones evaluadas se tomaron como faltantes. Aquellos marcadores que contaban con un número de datos faltantes mayor o igual al 5% se eliminaron.

Asimismo, en el segundo filtro se eliminaron las accesiones en las cuales se halló hasta un 20% de datos faltantes, de manera similar a como se realizó para los marcadores SNPs. Posteriormente, se realizó una imputación de los datos faltantes en cada marcador SNP, de acuerdo a la media que se presentaba para ese marcador.

Después de ello, conforme a la frecuencia del alelo menos común (MAF), se aplicó el tercer filtro, en el cual se eliminaron los marcadores que tenían un porcentaje menor o igual al 10% de alelos menos frecuentes o comunes. Finalmente, se generó un archivo con el programa Excel en formato csv, el cual contenía la información de todos los marcadores (en columnas) y accesiones (filas) restantes del proceso de filtrado, es este archivo el que se usó y del cual se derivaron los diferentes formatos para el análisis de datos para la caracterización molecular y estructura genética.

3.7 ANÁLISIS DE DATOS

3.7.1 Análisis de datos para la caracterización morfológica

3.7.1.1 Análisis de la estadística descriptiva

Como parte del análisis de la estadística descriptiva, se evaluaron las acciones en base a los caracteres cuantitativos y cualitativos. De las 120 accesiones, solo fueron empleadas 116 de ellas para estos análisis morfológicos. Estas fueron 40 accesiones de la raza chullpi, 40 de paro y 36 de piscorunto, excepto por APUC-309, MOQ-016, MOQ-010 y AYA-183, de estas cuatro accesiones no se pudieron obtener las observaciones de los caracteres en grano y mazorca, ya que estas se perdieron durante el proceso de traslado desde el campo experimental, ubicado en Chiquián (Huaraz) hasta la UNALM (Lima), es por ello que no están incluidas en este tipo de análisis.

Entonces, se determinó la moda para los caracteres cualitativos en cada una de las razas en estudio. Así mismo se generaron gráficos de barras comparando las frecuencias de los estados en cada carácter para cada una de las razas.

Continuando con el análisis descriptivo, se determinaron los valores máximos y mínimos, así como el cálculo de la media, desviación estándar y coeficiente de variabilidad para los caracteres cuantitativos, esto en cada una de las razas. Empleando esta información se realizaron gráficos de cajas para cada uno de los caracteres cuantitativos.

Adicionalmente se realizó un análisis comparativo de medianas, aplicando una prueba de Wilcoxon – Mann – Whitney. Las comparaciones se realizaron en pares de razas, así tenemos la comparación de Chullpi vs Paro, Chullpi vs Piscorunto y Paro vs Piscorunto. Las comparaciones de medianas se realizaron a un nivel de significancia del 5% para cada uno de estos caracteres cuantitativos.

Para el análisis de correlación, se calculó el estadístico de Pearson (r-valor) con su respectiva prueba de hipótesis (p-valor) a un nivel de significancia del 1%, esto se realizó para cada una de las razas, así como también en la población completa, la cual está comprendida por las tres razas en conjunto. Con los r-valores se pudo categorizar las correlaciones tanto positivas como negativas en la escala propuesta por Hernández, Fernández y Baptista (2014).

Los análisis de la estadística descriptiva fueron ejecutados haciendo uso del programa estadístico R v4.2.2 (Equipo de R Development Core, 2022) y empleado los paquetes FinCal

v0.6.3 (Fan, 2022), modeest v2.4.0 (Poncert, 2022), lme4 v1.4 (Bates *et al.* 2023), agricolae v1.6 (Mendiburu, 2023), openxlsx v.4.2.5.2 (Schauberges *et al.* 2023), Hmisc v5.1 (Harrell, 2023) y ggplot 2 v3.4.1 (Wickham *et al.* 2023).

3.7.1.1 Análisis de la estadística multivariada

Se analizaron los caracteres cuantitativos y cualitativos mediante diferentes enfoques para el análisis de agrupamiento jerárquico, esto con la finalidad de verificar la correspondencia de las accesiones a las razas que les fueron asignadas, así como también identificar la relevancia de los descriptores cuantitativos y cualitativos en los análisis empleados en este estudio.

El primer enfoque de agrupamiento jerárquico se realizó en base a una matriz de disimilaridad de los caracteres cuantitativos y/o cualitativos, para ello se calcularon las distancias genéticas entre las accesiones empleando los siguientes algoritmos descritos por Mohammadi y Prasanna *et al.* (2003) y Suwanda *et al.* (2020):

La **distancia euclidiana** o de línea recta es la estadística más comúnmente utilizada para estimar la distancia genética entre individuos (genotipos o poblaciones) mediante observaciones morfológicas y cuantitativas. La distancia entre dos individuos i y j , teniendo observaciones sobre los caracteres morfológicos (p) denotados por x_1, x_2, \dots, x_p y y_1, y_2, \dots, y_p para i y j , respectivamente, se pueden calcular mediante la siguiente fórmula:

$$d_{(i,j)} = [(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_p - y_p)^2]^{1/2}$$

La **distancia de Manhattan** es también denominada como “Distancia de la ciudad cuadrada”, esta mide el valor absoluto de la diferencia para un carácter entre dos individuos, esta distancia también emplea datos de caracteres cuantitativos.

$$d = \sum_{i=1}^n |x_i - y_i|$$

La **distancia de Gower** describe un coeficiente general para medir la distancia genética entre dos individuos sobre la base de varios tipos de caracteres, como por ejemplo los cualitativos y cuantitativos. Para los caracteres cualitativos, la distancia entre dos individuos se puntúa como 0 (siempre que haya una coincidencia) y 1 (siempre que haya una discrepancia). Para

caracteres cuantitativos, la distancia entre dos individuos se calcula como la diferencia en los valores del carácter dividida por el rango general del mismo.

Este método convierte la distancia de caracteres cuantitativos a un valor específico en una escala de 0 a 1; esto, a su vez, permite el uso simultáneo de datos cuantitativos y cualitativos para generar una matriz de distancias. Para ello, se suman las distancias de los caracteres individuales para cada par de individuos y luego se dividen entre el número de caracteres puntuados en ambos individuos. La medida de Gower de distancias entre individuos i y j se define como:

$$DG_{ij} = 1/p \sum w_k d_{ijk}$$

Donde p es el número de caracteres, d_{ijk} es la contribución del k -ésimo carácter a la distancia total entre dos individuos; $d_{ijk} = (x_{ik}-x_{jk})$, donde x_{ik} , x_{jk} son los valores del k -ésimo carácter en los individuos i y j , respectivamente, y $w_k = 1/R_k$, donde R_k es el renacimiento del k -ésimo carácter en la muestra.

A partir de la matriz de disimilaridad se procedió a realizar el agrupamiento de las accesiones, y para ello se emplearon cinco diferentes métodos, los cuales serán descritos a continuación (Gallardo 2023):

Método de distancia mínima o amalgamiento simple (**Single**), el cual considera que la distancia o similitud entre dos *clusters* o grupos viene dada por la mínima distancia, es decir la máxima similitud, entre sus componentes.

Método de distancia máxima o amalgamiento completo (**Complete**), el cual considera que la distancia o similitud entre dos *clusters* hay que medirla atendiendo a sus elementos más dispares; o sea la distancia o similitud entre *clusters* viene dada por la máxima distancia, es decir, la mínima similitud entre sus componentes.

Método de distancia aritmética en pares no ponderados (**UPGMA**), este propone utilizar como distancia entre un *cluster* y un individuo a la media de las distancias entre los individuos del de dicho *cluster* y del individuo respectivamente.

Método de Ward (**Ward**) emplea en cada etapa de agrupamiento a la unión de los dos *clusters* para los cuales se tenga el menor incremento en el valor total de la suma de los cuadrados de las diferencias, dentro de cada *cluster*, de cada individuo con respecto al centroide del *cluster*.

Método del Centroide (**Centroid**) utiliza como distancia entre grupos a la distancia entre los centroides de cada grupo. Este método presenta un inconveniente que es dejarse influenciar excesivamente por los grupos de mayor tamaño, esto hace que sea menos utilizado.

Adicionalmente, con la finalidad de seleccionar la distancia y método de agrupamiento más adecuados y en concordancia con las razas conocidas, se calculó el coeficiente cofenético para todas las combinaciones dadas entre las tres distancias (Euclidiana, Manhattan Y Gower) y los cinco agrupamientos realizados (Single, Complete, UPGMA, Ward y Centroid). Asimismo, se generaron dendrogramas para la combinación de distancias-agrupamiento con un coeficiente cofenético (CC) mayor y/o igual a 0.8 (Darkwa *et al.* 2020).

Una vez seleccionados los dendrogramas más adecuados, se realizó la prueba de la silueta (Tsindi *et al.* 2023) con la finalidad de identificar el número de agrupamientos en cada dendrograma con un CC mayor y/o igual a 0.8.

El segundo enfoque de agrupamiento jerárquico se realizó en base a los componentes principales (CP) generados mediante cuatro diferentes métodos, para ello también se usaron los caracteres cuantitativos y/o cualitativos. Para este análisis se calcularon los valores propios, la proporción de la varianza total expresada por cada carácter y los coeficientes por carácter en cada CP.

Para todos los métodos se retuvo el número de CP con los que se alcanzó el 75 % de varianza (Suhr 2020), y estos mismos fueron los que se emplearon para los análisis posteriores de Agrupamiento. Además, se realizaron gráficos de barras para la representatividad, que mide el grado de asociación entre los caracteres y un eje en particular (Kassambara 2017) y/o la contribución de los caracteres en CP seleccionadas para el análisis.

Asimismo, se generaron gráficos tipo “Biplot” con respecto a las accesiones y/o a los caracteres evaluados (a manera de un sistema de ejes formado por los dos primeros CP, esto para apreciar la dispersión y la agrupación de las accesiones de las diferentes razas).

Métodos para la obtención de las CP:

El análisis de componentes principales (**PCA**) es definido como un método para la reducción de la información que es útil para aclarar las relaciones entre dos o más caracteres y para dividir la varianza total de los caracteres originales en un número limitado de variables no correlacionadas (Mohammadi y Prasanna *et al.* 2003).

Los caracteres empleados para este análisis PCA fueron los cuantitativos y para evitar sesgos relacionados con la diferencia de escala entre las variables, los datos se estandarizaron para que su promedio fuera cero y la desviación estándar fuera igual a uno. Por otra parte, se analizó qué caracteres actuaron como principales discriminadores entre las accesiones y, asimismo, estos se agruparon mediante el algoritmo de K-means (Rivas *et al.* 2022).

Análisis de correspondencia múltiple (**MCA**), esta emplea datos de caracteres cualitativos, denominados también como categóricos, y suele emplearse para detectar patrones y las estructuras subyacentes en un conjunto de datos (Girma *et al.* 2014).

Además, se calculó la correlación de Pearson y el coeficiente de determinación entre los caracteres y los cuatro primeras CP para el PCA y MCA respectivamente, así como también una prueba de hipótesis para asegurar la significancia de la correlación mencionada con el CP (Kassambara 2017).

Análisis factorial de datos mixtos (**FAMD**), es un método de componentes principales que explora la información contenida en caracteres continuos y categóricos (Rellstab *et al.* 2016).

Análisis factorial múltiple (**MFA**), este método de análisis resume y permite la visualización de un conjunto de datos mixtos y complejos en donde los individuos son descritos por un conjunto de variables estructuradas en grupos. Además, toma en cuenta la contribución de todos los grupos de variables para definir las distancias entre individuos (Kassambara 2017).

A partir de los CP más significativos de cada método, es decir con los que se alcanzaba el 75% de la varianza, se realizó un análisis de agrupamiento jerárquico (HCPC), el cual permitió determinar patrones de variación y la relación genética existente entre las accesiones de las razas (Zavinnon *et al.* 2019). Este tipo de agrupamiento emplea el método de Ward para minimizar la inercia dentro del grupo y maximizar la inercia entre grupos con la finalidad de obtener los grupos más homogéneos (N'DA *et al.* 2022)

Los análisis y gráficos de la estadística multivariada también fueron ejecutados en R v4.2.2, empleando los siguientes paquetes cluster v2.1.4 (Maechker *et al.* 2023), stats v4.2.2 y dendextend v1.16.0 (Galili *et al.* 2023), ggdendro v0.1.23 (De Vries *et al.* 2022) para el enfoque con matriz de distancias; y los paquetes FactoMineR v1.1.0 (Husson *et al.* 2023), factoextra v1.0.7 (Kassambara *et al.* 2020), dplyr v1.1.0 (Wickham *et al.* 2023^a), ggplot2 v3.4.1 (Wickham *et al.* 2023b), hrbrthemes v0.8.0 (Rudis *et al.* 2020) y corrplot v0.92 (Wei y Simko, 2021) para el enfoque con la matriz de CP y el HPCP.

3.7.2 Análisis de datos para la caracterización molecular y estructura genética

A partir del archivo VCF, que contenía a todos los marcadores SNPs que pudieron ser identificados (en el paso de *Variant calling*), se generó un gráfico de densidad de los marcadores SNPs en los 10 cromosomas de maíz. Por otro lado, el análisis de variabilidad inicio con varios pasos de filtrado e imputación de los datos faltantes para los marcadores SNP en las accesiones, para cada uno de estos pasos se generaron gráficos que ilustraron los SNPs y accesiones que se eliminaron y los que se mantuvieron.

Como producto de los pasos de filtrado e imputación se obtuvo un archivo en formato csv, el cual se empleó como archivo principal y el cual denominaremos de aquí en adelante “data_snp_filt_imp” y del cual derivaron los archivos en los cuales se agregó información como raza, origen, entre otros para posteriores análisis de diversidad.

A continuación, el archivo “data_snp_filt_imp” se empleó para obtener tres archivos diferentes, cada uno de ellos contenía todos los marcadores SNPs para una raza en específico (paro, chullpi y Piscorunto). De esta manera a partir de cada uno de estos archivos y del archivo principal se calculó para cada raza y conjunto de accesiones, lo siguiente y de acuerdo con Granato y Fritsche-Neto (2022):

El Contenido de información polimórfica (PIC)

$$PIC = 1 - (p_j^2 + q_j^2) - 2 p_j^2 q_j^2$$

La frecuencia del alelo menos común (MAF)

$$MAF = \min(p_j, q_j)$$

La heterocigosidad observada (Ho)

$$H_0 = \frac{n H_j}{N}$$

La diversidad genética de Nei o heterocigosidad esperada (GD o He)

$$GD = 1 - p_j^2 - q_j^2$$

Donde para el locus j existen dos alelos A_1 y A_2 , $n H_j$ es el número de genotipos heterocigotos en el locus j , N es el número de accesiones y para estos alelos se tiene las siguientes frecuencias alélicas:

$$f(A_1) = p_j = \frac{n A_1}{2N}$$

$$f(A_2) = q_j = 1 - p_j$$

En el inicio del análisis de la estructura genética se empleó el programa STRUCTURE (STR) (Pritchard *et al.* 2000), el cual emplea un enfoque para el análisis de agrupamiento en base a un modelo bayesiano (Mohammadi y Prasanna 2003). Para esto se convirtió el archivo “data_snp_filt_imp” al formato str, ya con el patron indicado se procedió a ejecutar el programa. Se realizaron pruebas preliminares para evaluar algunos parámetros, sin embargo, otros se mantuvieron constantes como la longitud del Burn-in de 5000, MCMC de 10000, el valor de Lambda (λ) igual a uno y el modelo Admixture de ascendencia, esto debido a que estos parámetros ya fueron evaluados en otros estudios (Zhang *et al.* 2016, Lanes *et al.* 2014, Schaeffer 2013).

Como parte de las pruebas preliminares se realizaron un par corridas de STR variando el modelo de frecuencias alélicas, siendo estas independientes o correlacionadas, para esto no se utilizó información previa sobre el origen o raza de las accesiones para definir los grupos.

Resultando de ello, que no se encontraron diferencias en los resultados producidos por los diferentes modelos (datos no mostrados), de manera similar fue hallado por Lia (2009), es por ello que todos los análisis posteriores se realizaron sobre los resultados obtenidos de la frecuencia de alelos correlacionados, ya que es el modelo más frecuente en ser empleado por múltiples estudios relacionados con SNPs.

Luego, se probó el número de k , siendo este el número de grupos por formar, para ello se realizaron corridas de $k=1$ hasta $k=10$, como resultado se obtuvo que cuando el número de grupos (valor k) se fijó $k \geq 6$, las muestras siempre tendieron a agruparse estrechamente y en un solo grupo (anexo 11), por lo que las siguientes corridas con el programa se realizaron desde $k=1$ hasta $k=5$.

La evaluación del k más adecuado para población, se realizó utilizando un criterio *ad hoc* basado en la tasa de cambio en la probabilidad logarítmica de los datos entre k sucesivos (Braco *et al.* 2012, Evanno *et al.* 2005). Posteriormente, para la selección de la pertenencia de las accesiones a los grupos, se empleó las probabilidades de membresía (PM), que son los valores “ q ” de los resultados de las corridas con el programa (anexo 12). Estas debían ser superiores o iguales a 0,5 para ser asignadas a un grupo correspondiente y aquellas accesiones con una inferior a 0,5 se asignaron a un grupo mixto, de manera similar fue realizado por Leng *et al.* (2019), Dao *et al.* (2014), Semagn *et al.* (2012) y Van Ingheland *et al.* (2010).

Para cada k que se ejecutó con STR, se generaron diagramas de barras a partir de las probabilidades de membresía (matrices de los valores q) a un grupo (*cluster*) y estos se etiquetaron por accesión y por grupos determinados por STR (a un $PM \geq 0.5$), raza y origen.

Por otra parte, continuando con el análisis de la estructura genética de la población, se emplearon métodos basados en distancias. Para ello, se aplicaron procedimientos jerárquicos como las distancias Euclidiana, Simple, Absolute y Manhattan (Gruber *et al.* 2023), combinando estas con dos diferentes formas de agrupamiento más comunes (Mohammadi y Prasanna 2003), UPGMA y Ward. De las combinaciones entre distancias y agrupamientos obtenidos se calculó el coeficiente cofenético para elegir método más adecuado para aglomerar a las accesiones.

A partir de la selección de método y distancias más idóneos para formar los conglomerados, se realizaron dendrogramas para las accesiones y grupos conformados por las razas, los determinados por STR y el origen de las accesiones. Adicionalmente, de acuerdo con la información del origen, proporcionada en los datos pasaporte (anexo 1), se ubicaron las accesiones en los diferentes departamentos del Perú en los que fueron colectados (figura 2), coloreados de acuerdo con la raza consignada en estos mismos datos pasaporte antes mencionados y provistos por el Banco de germoplasma de maíz de la UNALM.

Continuando con los métodos basados en distancia, se efectuó Análisis Discriminante de Componentes Principales (Jombart *et al.* 2010) que es un procedimiento no jerárquico, el cual consiste en ejecutar K-medias sucesivas con un número creciente de *clusters*, después de transformar los datos mediante un análisis de componentes principales. Para cada modelo simulado, se calcula una medida estadística de bondad de ajuste denominada criterio de información bayesiano (BIC), que permite elegir el número *clusters* óptimo (Jombart *et al.* 2023).

De manera similar, se empleó las distancias entre las accesiones en base el coeficiente de correlación de Pearson para realizar un análisis de componentes principales (PCA), el cual proyecta las accesiones (variables) en un nuevo espacio utilizando una nueva matriz que muestra el grado de similitud entre las accesiones. Tradicionalmente, se utiliza un coeficiente de correlación en lugar de la covarianza, ya que el uso de un coeficiente de correlación elimina el efecto de la escala: así, una variable que varía entre 0 y 1 no pesa más en la proyección que una variable que varía entre 0 y 1000 (Luminvero 2023).

En el mismo contexto del análisis de estructura genética, se calculó los valores de F_{st} por pares de razas, grupos determinados mediante STR y grupos de acuerdo con el origen de las accesiones, para estimar la diferenciación genética entre los grupos antes mencionados. Esto se realizó empleando el método propuesto por Wright (1949) y que fue actualizado por Weir y Cockerham (1984). Adicionalmente, se aplicaron tres Bootstrap para cada par de F_{st} , así como el cálculo del intervalo de confianza al 95%.

Al mismo tiempo, tomando en cuenta que la heterocigosidad de las accesiones es una fuente de variación dentro de las mismas y por lo tanto de una posible diferenciación entre las que pertenecen a un mismo grupo, fue que se empleó un Análisis Molecular de Varianza

(AMOVA) (Excoffier *et al.* 2010, 1992) para estimar y descomponer el porcentaje de la varianza molecular que reside entre las accesiones, en las accesiones dentro de cada grupo de raza, de STR y de origen; así como también, entre los grupos.

Los análisis y gráficos de la caracterización molecular y estructura genética fueron ejecutados en R v4.2.2, empleando los siguientes paquetes ggplot2 v3.4.1 (Wickham *et al.* 2023b), vcfR v1.14.0 (Knaus *et al.* 2023), gaston v1.5.9 (Perdry *et al.* 2023), bwardr v0.1 (Ward 2023), SNPReady v0.9.6 (Granato y Fritsche-Neto 2022), pophelper v2.3.1 (Francis 2017), dartR v2.7.2 (Gruber *et al.* 2023), devtools v2.4.5 (Wickham *et al.* 2022), label.switching v1.8 (Papastamoulis 2022), tidyr v1.3.0 (Wickham *et al.* 2023c), poppr v2.9.3 (Kamvar *et al.* 2023), adegenet v2.1.10 (Jombart *et al.* 2023).

IV. RESULTADOS Y DISCUSIONES

4.1 ANÁLISIS DE LA CARACTERIZACIÓN MORFOLÓGICA

4.1.1 Análisis de la estadística descriptiva de caracteres cualitativos

Con respecto a los caracteres cualitativos, se muestra el resumen de las modas para las tres razas evaluadas en la tabla 2. Dentro de estos resultados, se destaca la presencia de la misma moda en los caracteres de como CT, DHG, CM y FM.

Adicionalmente, a partir de la información proporcionada en la tabla 2, se puede observar que el carácter FSG presento una moda muy variada en las tres razas, ya que fue el único carácter cualitativo que presento un estado diferente como moda para cada una de las razas. Al menos dos razas presentaron la misma moda, con respecto a los otros caracteres como PT, CG, CE y TG.

Tabla 2: Tabla de la moda de los caracteres cualitativos evaluados en las razas Paro, Chullpi y Piscorunto

Carácter cualitativo	Raza		
	Chullpi	Paro	Piscorunto
CT	Café	Café	Café
PT	Densa	Intermedia	Intermedia
DHG	Irregular	Irregular	Irregular
CM	Blanco	Blanco	Blanco
FM	Cónica cilíndrica	Cónica cilíndrica	Cónica cilíndrica
CG	Amarillo	Amarillo	Moteado
CE	Blanco	Blanco	Blanco
TG	Dulce	Harinoso	Harinoso
FSG	Contraído	Plano	Dentado

Además de contar con la misma moda, los caracteres CT, DHG, FM y CM, también exhibieron la presencia de otros estados. Así se tiene que para la FM y el CM además de la moda se presentan tres estados adicionales (figura 10).

Para la FM, el estado de cónico-cilíndrico se presenta como la moda de las tres razas, esto fue registrado como característica únicamente de la raza Piscorunto por Salhuana (2004), así como también por Velásquez y Montoro (2011); asimismo, estos autores han encontrado formas cónicas en la raza Paro, lo cual también se encontró en el presente estudio; sin embargo, para la raza Piscorunto también se observaron accesiones con otras formas cónicas.

Adicionalmente, en la raza Chullpi también se encontraron los estados esférico y cilíndrico, lo cual no concuerda con la información recolectada para esta raza con respecto a este carácter, ya que Salhuana (2004), menciona acerca de mazorcas ovoidales, así como también Velásquez y Montero (2011) las describen como esfero-cónicas.

En CM, la moda para las tres razas es el color blanco, lo cual también ha sido resaltado como el color predominante por Salhuana (2004), así como por Velásquez y Montoro (2011), pero solo para las razas Piscorunto y Chullpi, ya que para la raza Paro mencionan que hay una gran variedad de colores, lo cual también se pudo observar en la presente investigación. En la figura 10 se muestra que, para CM, la raza Paro es la única que presenta cuatro diferentes estados en este descriptor, es decir el marlo presentó hasta cuatro diferentes colores.

En el caso del carácter CT, el color predominante para las tres razas es el café, tal como se observa en la figura 10, y siguiendo a este, se tiene al color morado y por último está el color verde, cabe mencionar que este último color solamente está presente en las razas Paro y Piscorunto.

El carácter DHG, además de presentar la misma moda en las tres razas, que en este caso es la dispersión irregular, presenta también dispersión regular en las tres razas, pero en menor frecuencia, esto se puede observar en la Figura 10. Estos hallazgos concuerdan con mencionado por Grobman *et al.* (1961) acerca de la dispersión irregular de los granos en las mazorcas de estas tres razas de maíz amiláceo.

En los caracteres TG y CE, los cuales se encuentran muy relacionados, solo se presentaron dos estados (figura 11), el estado harinoso en el TG y el color blanco de CE estuvo presente en las tres razas, pero con una frecuencia total en Paro y Piscorunto, lo cual concuerda con lo mencionado por (Oscanoa y Sevilla, 2011, Salhuana 2004).

El estado dulce únicamente se presentó en la raza Chullpi, en esta raza se da la presencia de un endospermo arrugado porque es homocigota para el gen *su* que produce germoplasma dulce (*su* es del inglés *sugary* que significa azucarado). El gen *su* no permite que el azúcar se transforme en almidón, por eso cuando el grano se seca, se arruga (Oscanoa y Sevilla, 2011).

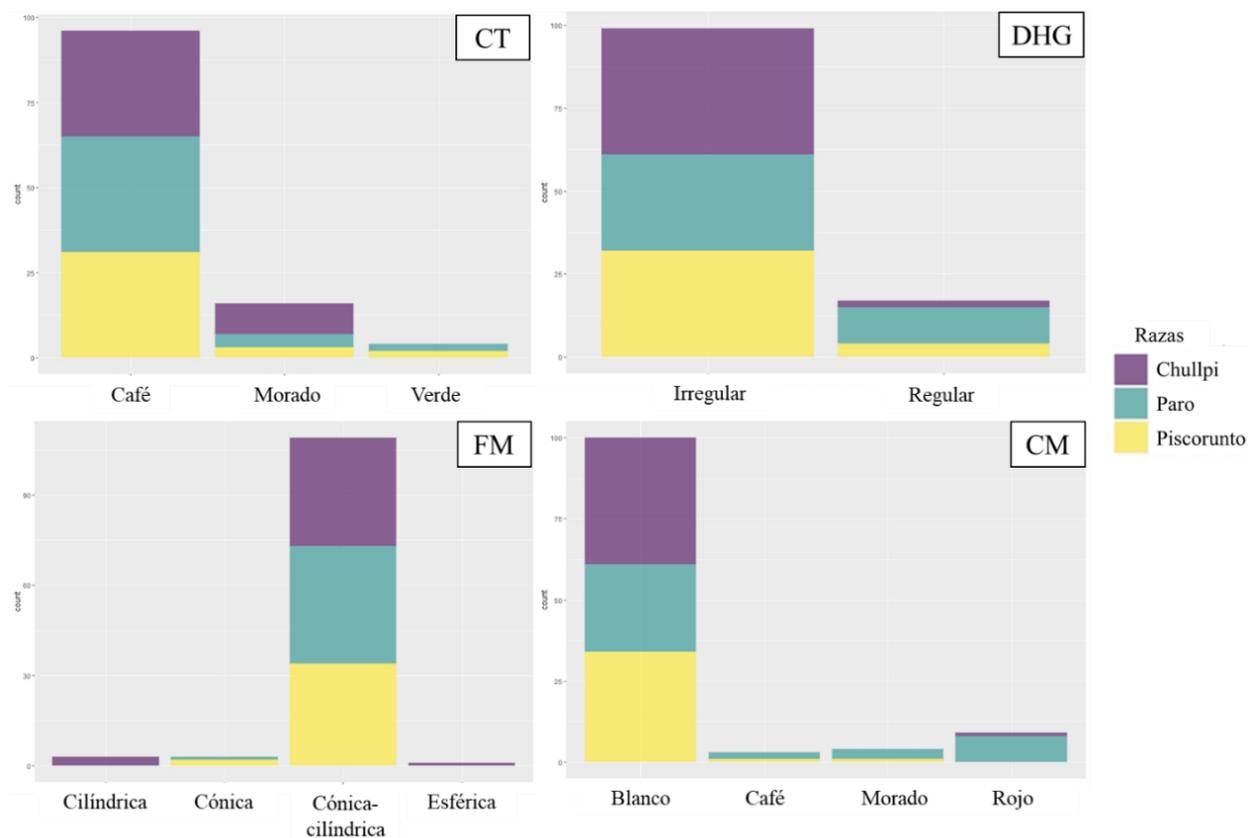


Figura 10: Gráfico de barras para los caracteres CT, DHG, FM y CM

Su es un gen recesivo; si a una planta de la raza Chullpi le cae polen de una raza con granos harinoso, como lo son Paro o Piscorunto, cuyo genotipo es *SuSu*, el grano sería *Sususu*, es decir que tendría una triple dosis alélica y es harinoso normal (Oscanoa y Sevilla, 2011), lo cual ha ocurrido con algunas accesiones de esta raza evaluadas en el presente estudio, en las cuales también se encontraron granos harinosos, pero en menor frecuencia.

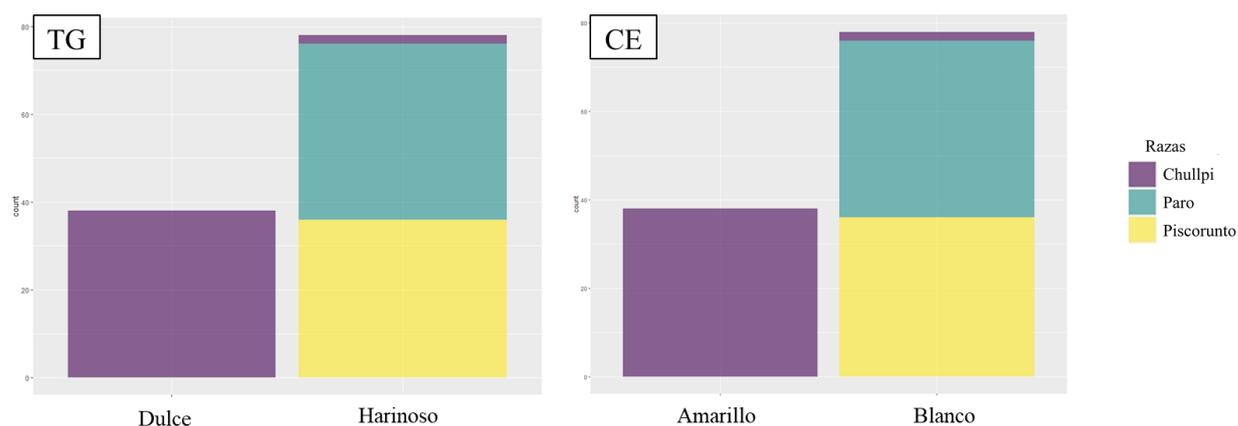


Figura 11: Gráfico de barras del tipo de grano

En el caso de PT, los estados intermedia y densa fueron los más frecuentes para las tres razas (figura 12), siendo la moda, la pubescencia intermedia. Además, la pubescencia escasa solo estuvo presente en las razas Paro y Chullpi, pero en baja frecuencia. Estas observaciones sobre la pubescencia son muy discordantes a las halladas por Grobman *et al.* (1961), ya que ellos mencionan ausencia de pubescencia para las razas Chullpi y Paro, y para la raza Pisorunto una baja pubescencia; cabe mencionar que han pasado 60 años desde esas descripciones para este carácter, en este periodo la selección continua a lo largo de los años puede haber generado dichas diferencias en este carácter para las tres razas evaluadas en el presente estudio.

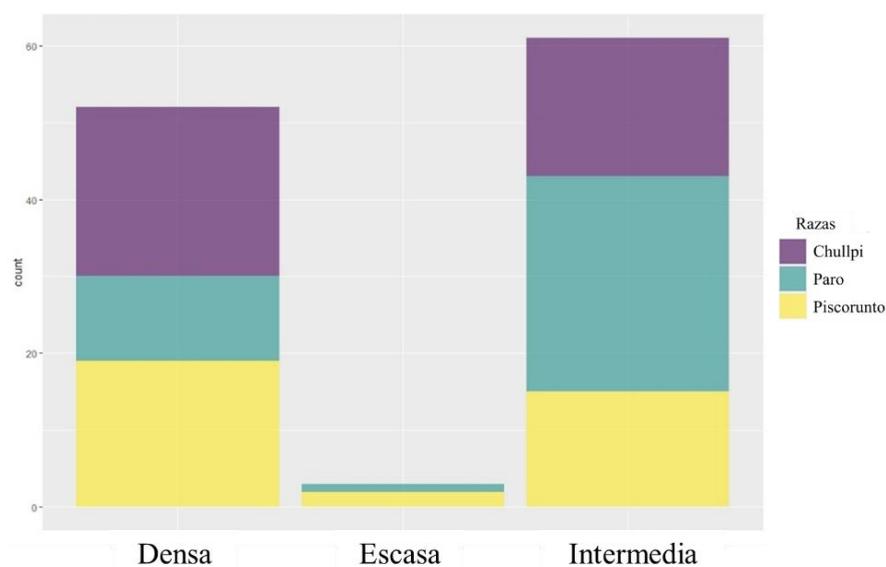


Figura 12: Gráfico de barras de la pubescencia de tallo

El carácter FSG presentó seis estados en este descriptor, así se puede observar en la figura 13, que en la raza Chullpi casi todas las accesiones presentaron forma contraída y algunas dentadas, pero en menor frecuencia. Asimismo, se observó que la mayor frecuencia para la raza Piscorunto fue dentado, con algunas accesiones planas y otras redondas.

Finalmente, se tiene a la raza Paro con la mayor cantidad de estados diferentes, siendo estos el dentado, muy puntiagudo, plano-puntiagudo y redondo. Así como también, se puede observar que la raza Paro existe una variedad de formas y colores más amplia que con respecto a las razas Chullpi y Piscorunto para este carácter FSG en conjunto con CM y CT.

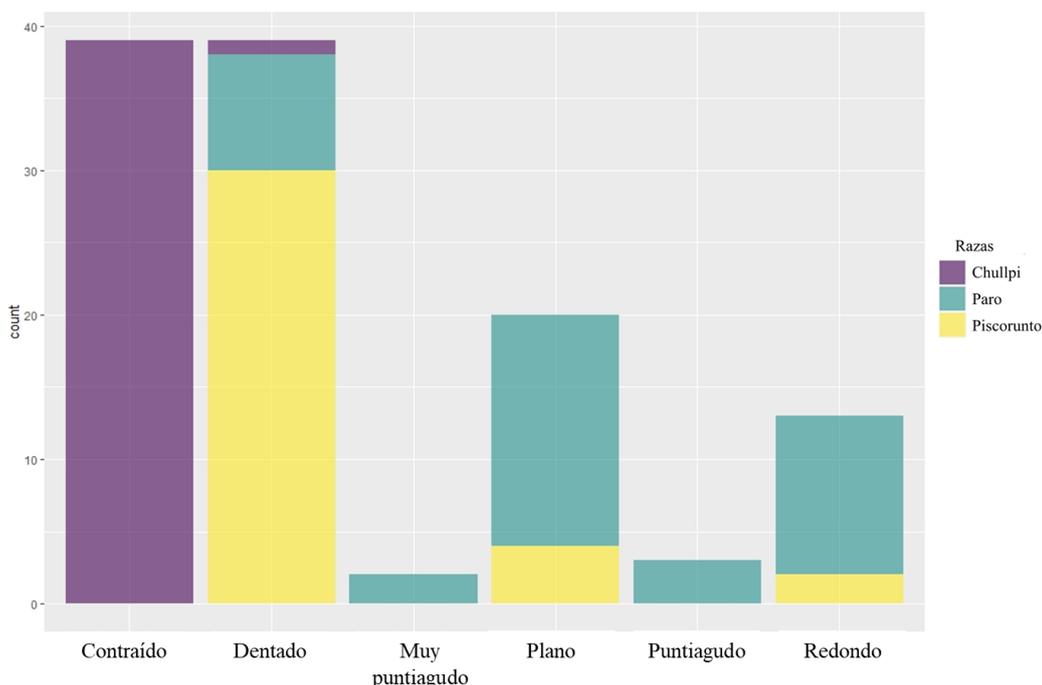


Figura 13: Gráfico de barras de la forma de la superficie de grano

En el CG, la moda de las razas Chullpi y Paro fue el color amarillo, sin embargo, si se observa la figura 14, esta se evidencia lo ambigua que puede ser la interpretación de la moda si es que no se toma en cuenta las demás frecuencias en la muestra completa.

Así se tiene que en la raza Paro, el color amarillo es el color más frecuente, pero este deja de ser un referente característico de esta raza, ya que, seguido a este estado, en términos de frecuencia, se encuentra el color capa blanca, el cual solo tiene algunas accesiones menos que el amarillo; en este caso la moda no es muy informativa, ya que existe varios estados de este descriptor para esta raza. Además, se presentaron accesiones con color anaranjado,

blanco, café, jaspeado y rojo, pero en menor frecuencia que los dos estados antes mencionados.

Lo indicado en el párrafo anterior, permite corroborar la gran variedad de colores que se exhiben en la raza Paro, lo cual no ha sido reportado en investigaciones anteriores, es por ello que cabe resaltar la posibilidad de que esto puede haber ocurrido, en la presente investigación, debido a un mal manejo al realizar los cruzamientos, provocando así una contaminación entre el polen proveniente de las diferentes razas, ya que como se observa en las evidencias fotográficas de la cosecha para esta raza (anexo 3), dentro de las mazorcas es donde más se exhibe la gran variedad de colores en el grano.

En la raza Chullpi, al contrario de lo ocurrido en la raza Paro, tiene como moda el color amarillo y al mismo tiempo tiene una frecuencia total en este estado; por lo cual, se puede decir que este color de grano es predominante en esta raza, lo cual también ha sido corroborado por la evaluación realizada por Cosio *et al.* (2014) en las accesiones de esta misma raza del Banco de Germoplasma de la UNSAAC, donde encontraron que casi todas las accesiones de la raza Chullpi presentaban grano de color amarillo.

Así también, se encontró que para la raza Piscorunto, la moda que es el color moteado tiene una frecuencia predominante, este color tan característico de esta raza también ha sido descrito como tal por Velásquez y Montero (2011), Salhuana (2004) y por Grobman *et al.* (1961), esto puede indicar que este carácter ha sido y es empleado para su selección, ya que se ha mantenido a través de los años.

El moteado típico de esta raza, así como también de algunas otras, es causado por los genes Dt-1 del cromosoma 9, Dt-2 del cromosoma 6 y Dt-3 del cromosoma 7; así como también se sabe que es producido por la presencia de un alelo R y 2r, o sea, el genotipo Rrr es el que se expresa en la aleurona. El color de la aleurona segrega dentro de la mazorca, o sea que dentro de una mazorca se pueden encontrar granos moteados como también granos rojos, púrpuras, etc (Oscanoa y Sevilla, 2011). Esto explica porque varias accesiones de la raza Piscorunto también presentan granos de color morado y café, pero en menor frecuencia.

Tal como se menciona en párrafos anteriores, los resultados obtenidos concuerdan con la literatura citada, sin embargo, se debe mencionar que también se han podido identificar en

varias accesiones de las tres razas evaluadas, en las cuales se halló la presencia de estados en los caracteres que previamente no habían sido documentados para estas. Esto permite proponer algunos motivos por los cuales se puede haber presentado esta situación, así como también la existencia de una amplia variabilidad morfológica presente en las accesiones evaluadas

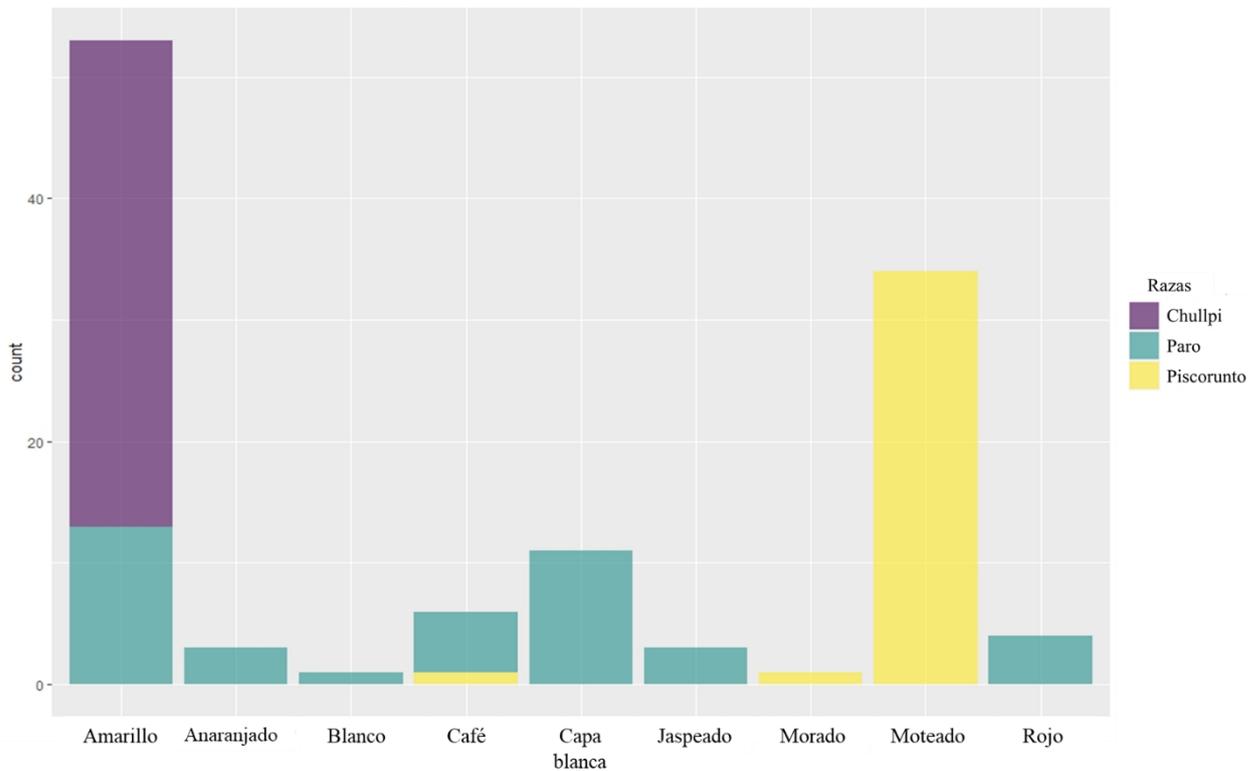


Figura 14: Gráfico de barras del color de grano

4.1.2 Análisis de la estadística descriptiva de caracteres cuantitativos

Se evaluó la media, mediana, máximo, mínimo, desviación estándar y coeficiente de variabilidad en 19 caracteres cuantitativos en las razas Paro, Chullpi y Piscorunto. En el anexo 6, se encuentran tres tablas con todos los resultados obtenidos para las razas Paro, Chullpi y Piscorunto, respectivamente.

Los resultados de las medidas de tendencia central y los de dispersión fueron empleados para realizar un análisis comparativo, para esto se tomaron las medianas de cada raza en cada carácter y se sometieron a una prueba de Wilcoxon – Mann – Whitney. Se empleó esta prueba, ya que previamente se probaron los supuestos de normalidad y homogeneidad de

varianzas, de los cuales los supuestos de normalidad no resultaron significativos (estos resultados se encuentran en el anexo 7).

Tomando en consideración lo anterior, se muestran los p-valores de la prueba de Wilcoxon – Mann – Whitney para cada uno de los caracteres cuantitativos evaluados en la tabla 3. En esta tabla se observa que para todas las medianas comparados de los caracteres entre las razas de Chullpi y Piscorunto, el p-valor es menor que el nivel de significancia de 0.05, por lo tanto, se puede rechazar la hipótesis nula. Esto nos indica que hay una diferencia significativa entre las medianas de las dos muestras, es decir, hay evidencia estadística para afirmar que las medianas son diferentes.

Tabla 3: Prueba de Wilcoxon – Mann – Whitney entre las razas Paro, Chullpi y Piscorunto

Prueba de Wilcoxon - Mann - Whitney			
Carácter cuantitativo	p-valor a un nivel de significancia de 0.05		
	Chullpi vs Paro	Chullpi vs Piscorunto	Paro vs Piscorunto
DFM	< 2.2e-16	< 2.2e-16	< 2.2e-16
DFE	< 2.2e-16	< 2.2e-16	< 2.2e-16
AP (cm)	3.466E-07	4.611E-05	4.313E-02
AM (cm)	< 2.2e-16	< 2.2e-16	2.429E-02
NHT	< 2.2e-16	< 2.2e-16	< 2.2e-16
NHEM	2.712E-03	3.538E-09	3.491E-04
LH (cm)	1.178E-08	2.200E-16	1.314E-14
AH (cm)	8.669E-12	< 2.2e-16	< 2.2e-16
LM (cm)	< 2.2e-16	8.140E-03	< 2.2e-16
DAM (mm)	< 2.2e-16	< 2.2e-16	4.225E-03
DCM (mm)	< 2.2e-16	< 2.2e-16	< 2.2e-16
DBM (mm)	< 2.2e-16	< 2.2e-16	6.699E-04
NH	< 2.2e-16	< 2.2e-16	1.675E-03
NGH	< 2.2e-16	< 2.2e-16	1.011E-06
DM (mm)	< 2.2e-16	< 2.2e-16	3.521E-05
DR (mm)	< 2.2e-16	< 2.2e-16	7.528E-06
DME (mm)	< 2.2e-16	< 2.2e-16	1.657E-03
LG (mm)	0.404	1.350E-06	3.027E-10
AG (mm)	< 2.2e-16	< 2.2e-16	3.137E-02
GG (mm)	< 2.2e-16	< 2.2e-16	3.427E-06
PG (gr)	< 2.2e-16	< 2.2e-16	0.346

Para el caso de las comparaciones realizadas entre las razas Chullpi y Paro, también los p-valores fueron menores que el nivel de significancia establecido, excepto por el carácter de longitud de grano que obtuvo un p-valor de 0.404, lo cual no permite rechazar la hipótesis nula, es decir, no hay suficiente evidencia para concluir que las medianas sean diferentes.

De manera similar a lo mencionado en el párrafo anterior, en las comparaciones para las razas Paro y Piscorunto, todos los p-valores fueron menores que el nivel de significancia, pero en este caso, el carácter en excepción fue el peso de 100 granos, el cual obtuvo 0.218, p-valor que no permite concluir que las medianas son diferentes.

Adicionalmente, se realizaron gráficos de dispersión para todos los caracteres evaluados, la figura 15 muestra el gráfico de cajas para los caracteres de DFM y DFF, en este se puede observar que la raza Paro es la que tiene el conjunto de datos más uniforme. Por el lado de las razas Piscorunto y Chullpi, se puede mencionar que ambas tienen un conjunto de datos más variables que los de la raza Paro.

En las tres razas existe un periodo de solapamiento entre los días de floración femenina y masculina, lo cual permite la polinización dentro cada raza. La raza Paro es la más precoz con el menor número de días a floración y seguida a esta es la raza Piscorunto. La raza Chullpi es la más tardía de todas, ya que se puede observar que para ambos caracteres esta raza es la que tiene la mayor cantidad de días hasta la floración, estos resultados son similares a obtenidos por Salhuana (2004) y Grobman *et al.* (1961), los cuales mencionan que esta raza tiene hasta 144 días a floración en promedio.

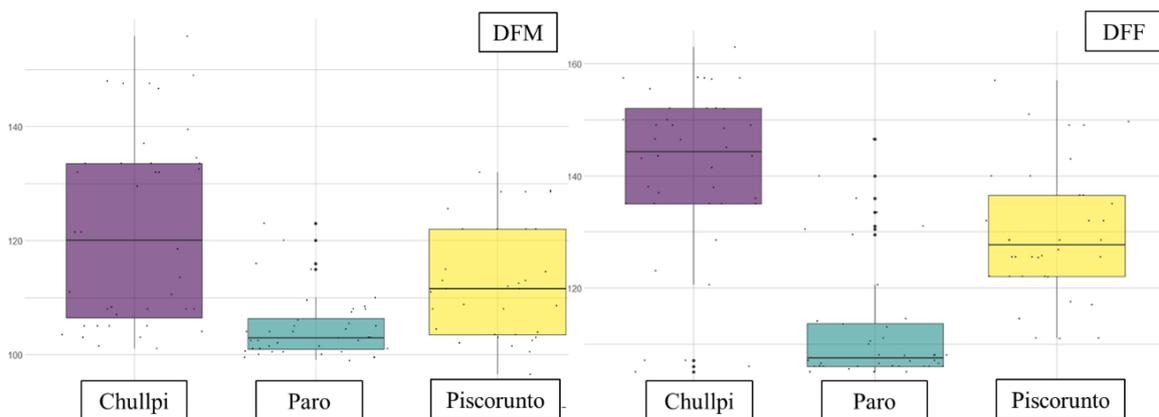


Figura 15: Gráfico de cajas para los caracteres DFM y DFF

Luego de analizar los resultados para el carácter de AP, se puede afirmar que la raza con la mayor AP fue la raza Chullpi con 191.94 cm en promedio, seguida de la raza Paro con 183.98 cm y, por último, a la raza Piscorunto con 174.41 cm, todas estas medias resultaron diferentes grupos de Wilcoxon (tabla 3). De manera similar a la AP, en el carácter de la AM, la raza Chullpi posee la altura de mayor magnitud con 108.78 cm, seguido por la raza Paro con 95.73 cm y, por último, la raza Piscorunto con 86.03 cm.

Los resultados obtenidos en este estudio, con respecto al carácter de AP, concuerdan con los obtenidos para la raza Chullpi, por Velázquez y Montero (2011), Salhuana (2004) y Grobman *et al.* (1961). En estos estudios, mencionan a la raza Chullpi como la de mayor altura entre estas tres razas, seguido por la raza Piscorunto y siendo la de menos altura, la raza Paro.

Con respecto a estas dos últimas razas de menor altura, en el presente estudio lo obtenido se lo contrario, ya que, en promedio la raza Paro es más alta que la raza Piscorunto, cabe resaltar la antigüedad de estas investigaciones, además de los efectos de la domesticación de este cultivo a través de los años.

En la figura 16, se puede observar que las tres razas tienen una variabilidad de respuestas similar. Cabe resaltar, que en esta también se observan muchos valores extremos en las tres razas. Una de las razones para que esto ocurra, puede deberse a la ubicación de las plantas en los campos experimentales; por ejemplo, aquellas plantas que estaban en los extremos de los surcos a los bordes de cada bloque fueron las que presentaron los valores más extremos, ya que los árboles que bordeaban el campo experimental pudieron haber causado algún efecto de sombra y por ende impedir que una mayor cantidad de luz llegué a la planta y que esto de alguna manera haya afectado su crecimiento y desarrollo.

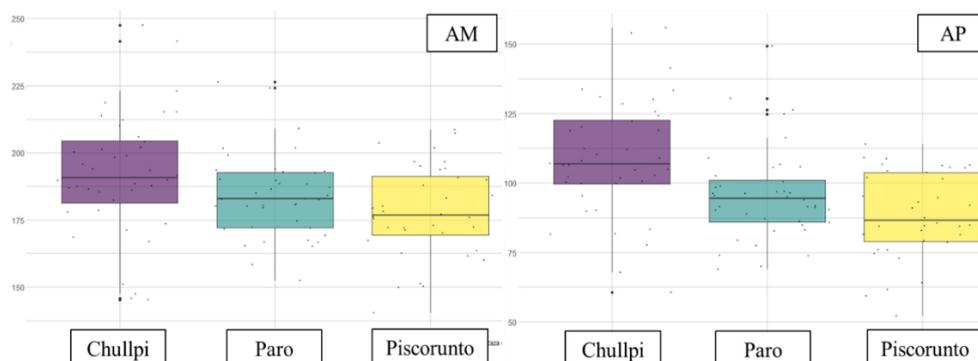


Figura 16: Gráfico de cajas para los caracteres AM y AP

El NHT resultó el mismo en las razas Paro y Piscorunto con 9.32 y 8.39 hojas en promedio respectivamente, razas entre las cuales no se hallaron diferencias significativas y estuvieron consideradas en un mismo grupo de Wilcoxon; mientras que, para la raza Chullpi se halló que su media si era diferente a la de las dos razas ya mencionadas e incluso con un NHT mayor, con 10.6 hojas en promedio por planta.

Con respecto a este carácter, tanto Velásquez y Montero (2011) como Salhuana (2004), mencionan a la raza Piscorunto como la raza con el mayor número de hojas entre las tres razas, seguido de Chullpi y, por último, Paro; sin embargo, lo encontrado en este trabajo no concuerda con esa información. Esto se puede deber a varias razones, entre ellas se tiene, el número de accesiones por raza que han sido muestreadas y la incorrecta denominación de las accesiones como pertenecientes a esas razas sin serlo realmente.

En la figura 17, se muestran los gráficos de cajas para NHT y para el NHEM, la raza con mayor homogeneidad en sus datos para ambos caracteres es Paro, seguida por Piscorunto y la raza con mayor dispersión entre sus observaciones es Chullpi.

En el NHEM, se encontraron tres grupos de Wilcoxon, correspondientes a las tres razas evaluadas, así la raza Chullpi mostró el mayor número de hojas por encima de la mazorca con 4.66 hojas en promedio, seguido por Paro con 4.51 y Piscorunto con 4.34 hojas.

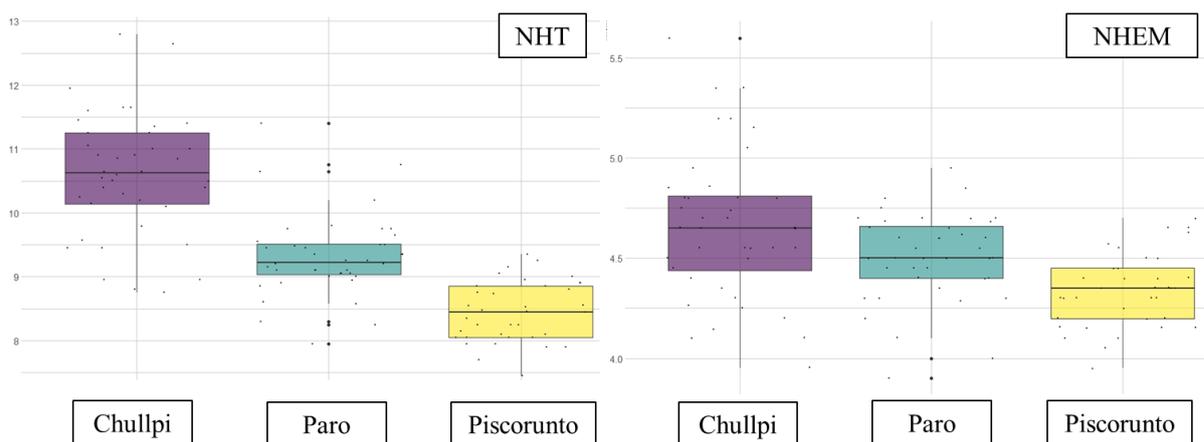


Figura 17: Gráfico de cajas para los caracteres NHT y NHEM

Con respecto a las hojas del maíz, se evaluaron los caracteres de LH y AH, en ambos casos y de acuerdo con la comparación de medianas de Wilcoxon, se hallaron tres diferentes grupos correspondientes a las razas en evaluación.

En el caso de la LH, la raza Chullpi tiene las hojas más largas con 79.87 cm en promedio, luego se tiene a la raza Paro con 76.55 cm y Piscorunto que tiene las hojas de menor longitud con 69.55 cm.

Para el AH, y como caso contrario al observado en LH, la raza Piscorunto es la hoja de mayor ancho con 8.70 cm en promedio, seguido de la raza Paro con 7.96 cm y Chullpi que tiene las hojas más angostas con una medida de 7.41 cm en promedio por planta.

En la figura 18, se puede observar una tendencia contraria en la respuesta de la evaluación de estos dos caracteres, ya que a medida que uno de estos dos caracteres aumenta en magnitud, el otro tiende a disminuir, en este gráfico también se observa que la dispersión en el conjunto de datos es semejante para las tres razas. Adicionalmente, se puede mencionar la existencia de valores extremos únicamente en la raza Paro.

En la mazorca se midieron varios caracteres, entre ellos el diámetro de la mazorca, esta medida se realizó a nivel de la base (DBM), al centro (DCM) y en el ápice de esta (DAM). Para estos tres caracteres mencionados, se halló que pertenecían a tres diferentes grupos de Wilcoxon, correspondientes a las razas en evaluación.

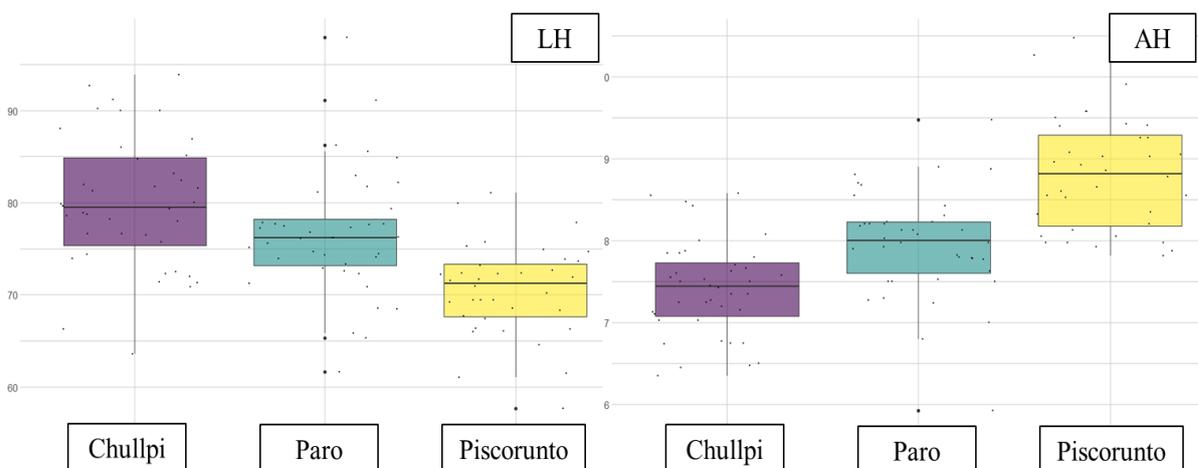


Figura 18: Gráfico de cajas para los caracteres LH y AH

La raza Chullpi para los caracteres de DBM, DCM y DAM, obtuvo las mayores medidas entre las razas, así se tiene 47.85, 54.70 y 43.13 mm respectivamente en promedio. A esta raza, le continuó la raza Paro para los caracteres DBM y DAM, con 43.39 y 35.42 mm respectivamente.

Por último, con las menores longitudes para DBM y DAM, se tiene a la raza Piscorunto con 42.47 y 34.74 mm respectivamente. Entre estas dos últimas razas (Paro y Piscorunto), el orden se invierte para el carácter de DCM, ya que el diámetro de posición intermedia lo tiene la raza Piscorunto con 50.71 mm y el diámetro de menor longitud, lo tiene la raza Paro con 47.87 mm.

En general, estas medidas de DBM, DCM y DAM reflejan y concuerdan con las FM halladas en las evaluaciones de carácter cualitativo, donde en las tres razas predominó la forma cónica-cilíndrica; ya que, para esta forma, el diámetro del centro debe ser de mayor magnitud a la longitud del diámetro de la base y al mismo tiempo este diámetro de la base debe ser de mayor magnitud a la longitud del diámetro del ápice, dando así esta mazorca una forma cónica-cilíndrica.

Con respecto al marlo, que forma la parte central de la mazorca, se realizaron mediciones de diámetro de este (DM), en el raquis (DR) y en la médula (DM). Estas longitudes halladas siguen la misma tendencia que en los caracteres de DBM, DCM y DAM, es decir, se hallaron que pertenecían a tres diferentes grupos de Wilcoxon, correspondientes a las tres razas ya mencionadas.

La raza Chullpi presenta las mayores longitudes en DM, DR y DME, así se tiene 28.56, 19.20 y 10.89 mm respectivamente, luego se encuentra la raza Piscorunto con 23.66, 14.18 y 7.10 mm respectivamente. Paro es la raza con las menores longitudes para estos caracteres con 22.87, 13.32 y 6.80 mm respectivamente en promedio.

Todo lo mencionado en estos párrafos con respecto a la mazorca se puede observar en la figura 19, este gráfico de cajas muestra esta tendencia que se observa entre las razas con respecto a los diferentes caracteres evaluados.

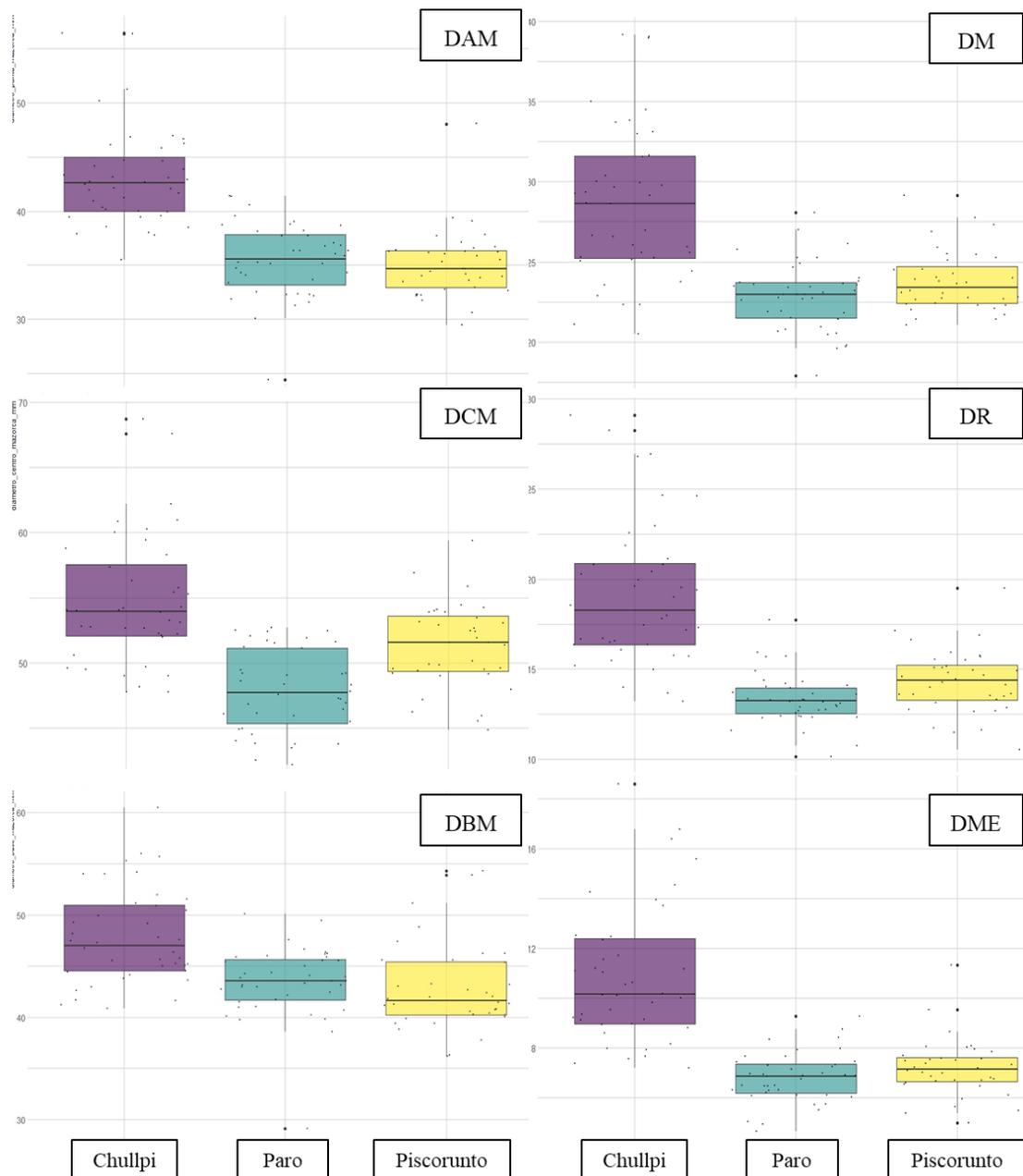


Figura 19: Gráfico de cajas para los caracteres de mazorca como DAM, DCM, DBM, DM, DR y DME

Así mismo, se observa la dispersión de los datos en cada una de ellas y la presencia de valores extremos para todas las razas. Estas observaciones se ven apoyadas en los porcentajes del CV para los caracteres de DM, DR y DME, ya que estos son los más altos entre todos los caracteres evaluadas en este estudio, lo que indica que estos caracteres son los menos confiables entre todas las estudiadas, o por lo menos no en las condiciones del presente estudio.

En el NGH se muestran medias diferentes para las tres razas, tal cual como se puede observar en la tabla 3, para la prueba de Wilcoxon, la raza Chullpi tiene la mayor cantidad de granos por hilera con 18.78 granos en promedio, luego continua la raza Paro con 16.89 granos y con la menor cantidad de granos por hilera se encuentra la raza Piscorunto.

En el caso del NH, los resultados de la prueba de Wilcoxon arrojaron solo dos grupos, de los cuales uno fue formado por la raza Chullpi con 17.78 hilera en promedio por mazorca y en el otro grupo se hallaron a las razas Paro y Piscorunto con 12.34 y 12.75 hileras respectivamente, en estas dos últimas razas no hay diferencias significativas. El NH es un carácter que ha sido estudiado en muchas ocasiones y en el caso del maíz amiláceo, Velázquez y Montero (2011), Salhuana (2004) y Grobman *et al.* (1961) han encontrado que la raza Chullpi cuenta hasta con 18 hilera en promedio, seguido de esta se tiene a la raza Paro con mazorcas que tienen entre 18 a 14 hileras, y por último la raza Piscorunto con 12 hileras en promedio. Estos hallazgos concuerdan con los encontrados en el presente estudio.

En la figura 20, se puede apreciar estos dos caracteres en gráficas de cajas, donde el conjunto de datos del NGH se muestra más homogéneo que el NH, incluso se puede mencionar que el NH tiene en todas las razas por lo menos un valor extremo.

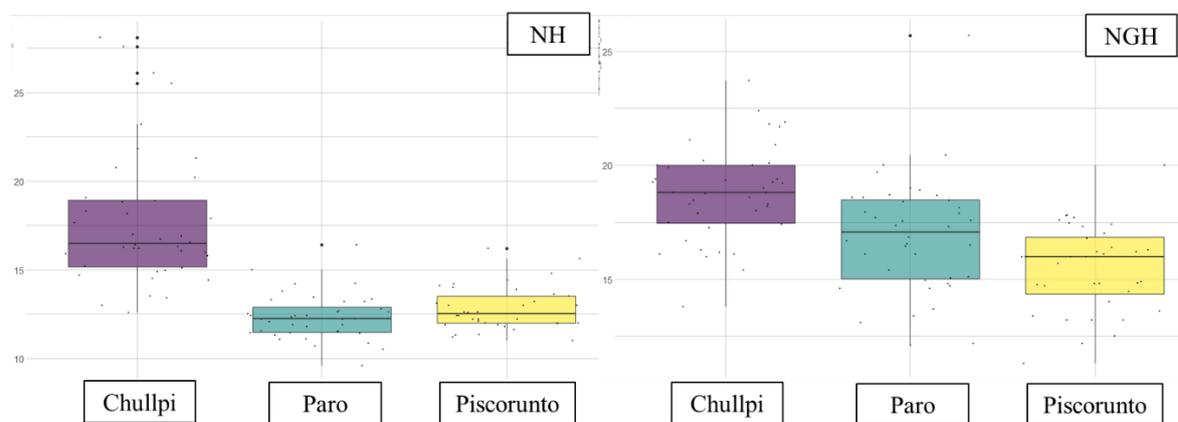


Figura 20: Gráfico de cajas para los caracteres NH y NGH

Continuando con los caracteres evaluados en la mazorca, la longitud de esta fue medida en las tres razas obteniendo como resultado, que las medias de las longitudes en promedio fueron todas diferentes, de acuerdo con la prueba de comparación de Wilcoxon.

Las mazorcas de mayor LM en promedio pertenecen a la raza Paro con 10.47 cm, a esta le continua la raza Chullpi con 9.38 cm y, por último, se tiene a la raza Piscorunto con 9.07cm en promedio.

En la figura 21, se muestra el conjunto de datos para este carácter, en este gráfico de cajas se puede apreciar que el grado de dispersión de los datos es alto; además, en el caso de la raza Paro se observan varios valores extremos, por lo cual, la media para esta raza puede estar siendo influencia por estos valores extremos, haciendo de esta manera que la raza aparente tener una longitud en promedio más larga de lo que en realidad posee la muestra.

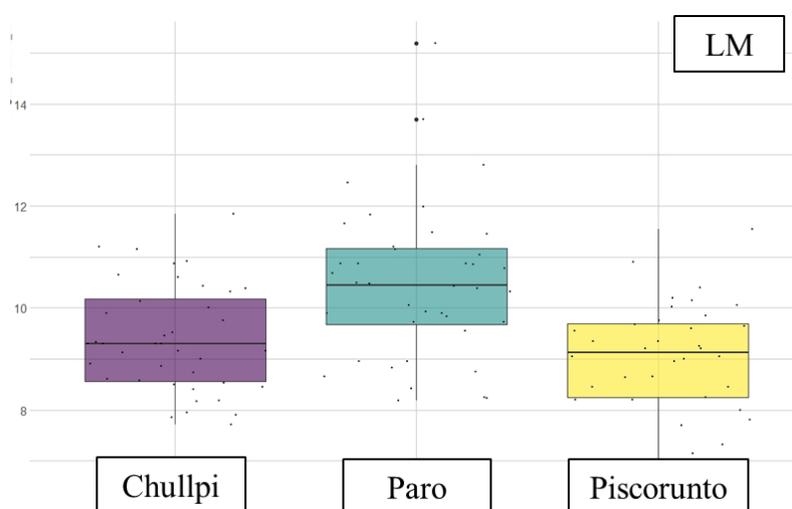


Figura 21: Gráfico de cajas para el carácter LM

Las medidas tomadas para el grano, es decir, el ancho, grosor y peso están relacionadas, ya que la raza Paro es la que tiene las medias de mayor magnitud con 9.75 mm, 5.66 mm y 52 g, seguida por la raza Piscorunto con 9.45 mm, 5.34 mm y 51.45 g, dejando de último a la raza Chullpi con 8.06 mm, 4.66 mm y 37 g, respectivamente. Cabe mencionar que las medias para estos caracteres fueron encontradas como diferentes grupos de Wilcoxon, siendo estos correspondientes a las razas evaluadas.

En el caso del carácter de longitud de grano, solamente presento dos grupos de Wilcoxon, el de mayor longitud fue conformado por la raza Piscorunto con 16.24 mm y en el otro grupo se encontraron las razas Chullpi y Paro con 15.59 y 15.51 mm, respectivamente. En la Figura 22, se muestra el gráfico de cajas de los cuatro caracteres evaluados en el grano.

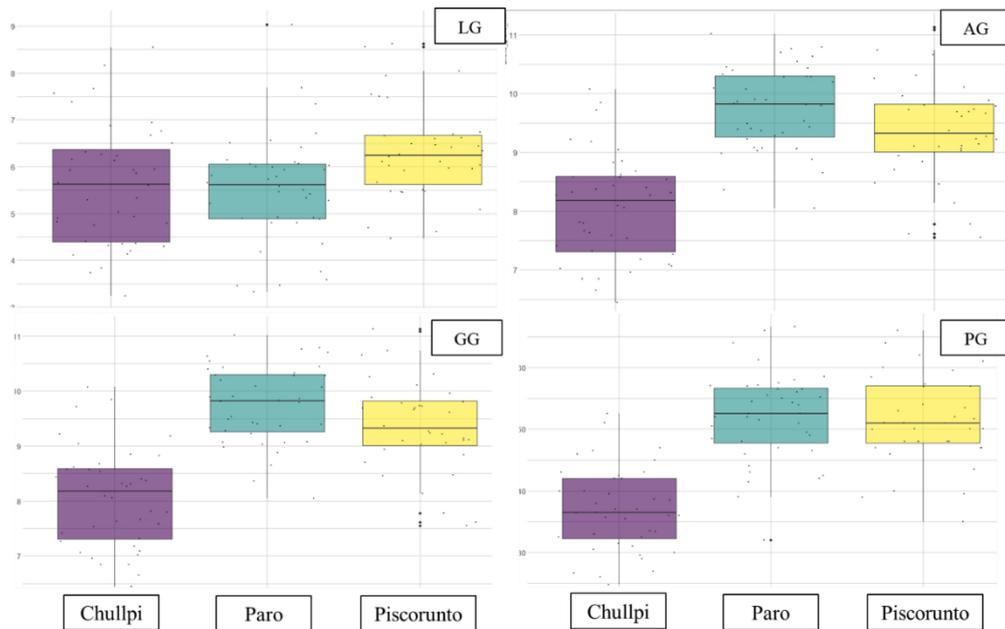


Figura 22: Gráfico de cajas para los caracteres LG, GG, AG y PG

5.1.3 Análisis de correlación de los caracteres cuantitativos

El análisis de correlación evidenció una amplia variedad de correlaciones tanto positivas como negativas y se realizó entre todos los caracteres cuantitativos, los cuales también fueron empleados para el análisis de la estadística descriptiva. Los valores de las correlaciones (r-valor) en la población de estudio completa, que comprenden las accesiones de las tres razas de maíz amiláceo, se muestran en la figura 23, en esta se observa en las celdas de color verde, aquellos r-valores con una categoría de correlación positiva de considerable hasta perfecta; así como también, los de correlación negativa considerable en las celdas de color rojo.

De forma complementaria, también se generaron tres figuras que evidencian los r-valores de las correlaciones obtenidas en cada una de las razas, estas figuras se encuentran en el anexo 8, de estas la figura 23^a muestra las correlaciones en la raza Chullpi, la figura 23b en la raza Paro y la figura 23c en la raza Piscorunto.

	NHEM	NHT	LH	AH	AP	AM	DFM	DFE	LG	AG	GG	LM	DAM	DCM	DBM	NH	NGH	DM	DR	DME	PG
NHEM	1.00	0.73*	0.56*	-0.32*	0.51*	0.45*	0.25*	0.24*	-0.06	-0.29*	-0.30*	0.19*	0.39*	0.28*	0.31*	0.41*	0.45*	0.40*	0.37*	0.35*	-0.33*
NHT	1.00	0.82*	-0.44*	0.71*	0.79*	0.46*	0.44*	-0.19	-0.49*	-0.52*	0.24*	0.61*	0.38*	0.45*	0.67*	0.67*	0.56*	0.61*	0.61*	-0.58*	
LH		1.00	-0.16*	0.80*	0.85*	0.35*	0.32*	-0.13	-0.25*	-0.33*	0.43*	0.46*	0.26*	0.40*	0.49*	0.60*	0.41*	0.42*	0.42*	-0.42*	
AH			1.00	-0.13	-0.15	-0.08	-0.08	0.26*	0.39*	0.14*	0.07	-0.38*	-0.18	-0.26*	-0.44*	-0.25*	-0.30*	-0.35*	-0.38*	0.38*	
AP				1.00	0.92*	0.21*	0.27*	-0.08	-0.16	-0.24*	0.32*	0.34*	0.20*	0.29*	0.36*	0.44*	0.35*	0.36*	0.33*	-0.22*	
AM					1.00	0.36*	0.40*	-0.15	-0.26*	-0.33*	0.28*	0.40*	0.25*	0.32*	0.48*	0.50*	0.41*	0.46*	0.43*	-0.31*	
DFM						1.00	0.68*	-0.07	-0.47*	-0.48*	-0.09	0.28*	0.33*	0.18*	0.51*	0.34*	0.33*	0.40*	0.41*	-0.51*	
DFE							1.00	-0.05	-0.52*	-0.49*	-0.18	0.30*	0.36*	0.17	0.52*	0.28*	0.38*	0.42*	0.40*	-0.52*	
LG								1.00	0.23*	0.05	-0.12	0.15	0.38*	0.31*	-0.02	-0.17	0.04	0.06	0.03	0.16	
AG									1.00	0.63*	0.24*	-0.50*	-0.41*	-0.25*	-0.77*	-0.38*	-0.55*	-0.59*	-0.64*	0.78*	
GG										1.00	0.14	-0.53*	-0.50*	-0.31*	-0.65*	-0.53*	-0.53*	-0.57*	-0.59*	0.68*	
LM											1.00	-0.03	-0.22*	0.13	-0.10	0.54	-0.08	-0.19	-0.18	0.12	
DAM												1.00	0.78*	0.80*	0.73*	0.44*	0.79*	0.79*	0.77*	-0.52*	
DCM													1.00	0.68*	0.71*	0.28*	0.77*	0.78*	0.74*	-0.41*	
DBM														1.00	0.54*	0.35*	0.63*	0.65*	0.63*	-0.28*	
NH															1.00	0.54*	0.78*	0.84*	0.85*	-0.74*	
NGH																1.00	0.43*	0.40*	0.40*	-0.47*	
DM																	1.00	0.89*	0.87*	-0.50*	
DR																		1.00	0.96*	-0.59*	
DME																			1.00	-0.63*	
PG																				1.00	

Figura 23: Coeficientes de Pearson (r-valor) de los caracteres cuantitativos calculados en las razas Chullpi, Paro y Piscorunto

*Las celdas en color verde tienen una correlación positiva desde considerable hasta perfecta y las celdas de color rojo tienen una correlación negativa considerable.

*Correlación es significativa al $p \geq 0.01$.

Las correlaciones empleadas para las discusiones que se presentaran a continuación resultaron significativas a un nivel del 1% de significancia. En el caso de aquellas, que luego de la prueba de hipótesis no fueron significativas, no se han empleado en estas discusiones, ya que carecen de sustento estadístico para asegurar algún tipo de relación lineal entre los caracteres evaluados.

De todas las correlaciones obtenidas la mayoría de ellas fueron positivas, resaltando particularmente, la correlación de AP con AM, la cual resultó con un r-valor de 0.92 (figura 23), siendo esta categorizada como perfecta, además de ser positiva y significativa. Varios grupos de investigación en maíz; por ejemplo, Mahmoog *et al.* (2022), Baduwal *et al.* (2022), Vishnuvardhan *et al.* (2021), Pallabi *et al.* (2018), Chavarry (2014) y Jawaharlal *et al.* (2011) y Martínez *et al.* (2010), han reportado magnitudes de correlaciones similares y muy cercanas a 1 con respecto a estos dos caracteres, así como en el presente estudio.

De modo similar, AP evidenció una correlación categorizada como positiva muy fuerte con LH, cuyo r-valor fue de 0.8. Este tipo de correlación entre estos dos caracteres también fue hallado por Magar *et al.* (2021). A su vez LH mostró un r-valor de 0.85 con AM, y esta correlación se pudo observar no solo en la muestra completa (figura 23), sino también por separado en el análisis de correlación que se hizo para cada una de las razas (anexo 8).

Dentro de esta misma categoría, positiva y muy fuerte, también se tiene a caracteres en grano, como la correlación de 0.78 que se observó entre el AG con el PG. Así como también al marlo (también llamado coronta), el cual tiene correlaciones que van desde 0.78 a 0.85, donde el DM, DR, DME están correlacionados muy fuerte al NH.

Por otro lado, bajando en la categoría de correlación, debido a una menor magnitud en el r-valor, se tiene, por ejemplo, al valor de 0.73 que se dio como correlación positiva y considerable entre NHEM y NHT (esto se dio en la muestra completa, figura 23, y de igual manera dentro de cada raza, anexo 8). Esto se puede deber a que entre algunos caracteres probablemente se estén representando diferentes vías para medir el mismo carácter, aunque también podría deberse a una relación estructural o del desarrollo como, por ejemplo, el NHT con el número de nudos y AP con AM (Martínez *et al.* 2010).

Además, otras correlaciones positivas y considerables fueron halladas entre AM y NGH (r -valor igual a 0.5), lo cual también fue hallado por Jawaharlal *et al.* (2011). Del mismo modo y con magnitudes similares, se tiene al AG con GG (r -valor igual a 0.63) y al GG con el PG (r -valor igual a 0.68), lo cual se puede estar dando debido a que estos caracteres todos han sido medidos en la mazorca.

Del mismo modo, y continuando con los caracteres de mazorca, el NH con DBM, DCM y el DAM exhibieron correlaciones entre 0.54 y 0.73, lo cual también fue hallado por Magar *et al.* (2021), Islam *et al.* (2020) y Jawaharlal *et al.* (2011). Los mismo ocurre entre los caracteres referentes al diámetro de la mazorca, es decir DBM, DCM y el DAM con los diámetros medidos en el marlo, DM, DR y DME, los cuales muestra valores de correlación entre 0.63 y 0.79; valores similares de correlación también fueron reportados por Chávez (2018).

Continuando en la categoría de positivas y considerables, están los DFF y los DFM, los cuales están correlacionados con un r-valor de 0.68. En otros estudios se encontraron correlaciones similares, tales como Mahmooq *et al.* (2022), Vishnuvardhan *et al.* (2021), y Jawaharlal *et al.* (2011).

Según la clasificación propuesta por Hernández, Fernández y Baptista (2014) sobre el valor de r (correlación de Pearson), se deben resaltar también las correlaciones entre los caracteres morfológicos que entreguen magnitudes altas y de asociación negativa, ya que la selección intensiva en el lado negativo de cualquiera de estas características podría mejorar otras características aumentando automáticamente los caracteres relacionados con el rendimiento (Jawaharlal *et al.* 2011).

Entonces, en este estudio se pudo obtener correlaciones negativas y considerables, tales como las relacionadas entre el AG y GG con los diámetros del marlo (DM, DR, DME), cuyos valores de r oscilaron entre -0.53 y -0.64, esto dentro del análisis de la muestra completa (figura 23). Continuando con los diámetros medidos, se tiene que tanto, los ya mencionados DM, DR y DME como los DBM, DCM, DAM se encuentran correlacionados con el PG, con r-valores que van desde -0.50 a -0.63, esto también fue evidenciado por el estudio realizado en Nepal por Baduwal *et al.* (2022).

En relación particular con el PG, este carácter mostro correlaciones negativas no solo con los diámetros en mazorca, como ya se ha mencionado, sino también con NHT y DFF (r-valor igual a -0.58 y -0.52, respectivamente), siendo los DFF una correlación que también fue hallada por Jawaharlal *et al.* (2011).

Habría que decir también que, de todos los caracteres evaluados, especialmente dos, el AH y LG, solamente mostraron algunas correlaciones de categoría media hasta débil, pudiendo ser inherentes a un germoplasma base y producto del proceso evolutivo en este germoplasma en particular, así como también lo podrían ser aquellas correlaciones que a pesar de tener mayor magnitud y de considerables hasta fuertes, no son significativas a nivel estadístico.

5.1.4 Análisis multivariado de agrupamiento jerárquico

La evaluación de la diversidad genética es un pre-requisito para identificar genotipos promisorios y superiores que pueden ser empleados en cualquier programa de hibridación en mejoramiento de plantas. Esta diversidad genética puede ser elucidada mediante el análisis de agrupamiento (Patel *et al.* 2023 y Suman *et al.* 2018).

El análisis de agrupamiento tiene la función de organizar la información y así esta pueda ser entendida y utilizada de forma más eficiente y fácil (Suman *et al.* 2018), como por ejemplo a partir de la visualización de un gráfico de árbol, denominado dendrograma, el cual es generado mediante esta técnica.

Así, por ejemplo, los genotipos pertenecientes a los agrupamientos separados por una gran distancia genética observable en un dendrograma, podrían utilizarse en un programa de hibridación para obtener un amplio espectro de variación entre los individuos segregantes (Patel *et al.* 2023).

Análisis multivariado de agrupamiento jerárquico empleando distancias genéticas

Dado que uno de los objetivos del presente estudio es evaluar y analizar la variabilidad genética, es que se realizaron análisis de agrupamiento jerárquico empleando dos enfoques, en el primero se utilizó una matriz de distancias genéticas y en el segundo se trabajó con Componentes Principales (CP) para la obtención de la matriz de distancias entre accesiones.

En el primer enfoque se emplearon tres diferentes algoritmos para el cálculo de las distancias genéticas (Euclidiana, Manhattan y Gower), así como cinco diferentes métodos de agrupamiento (UPGMA, Single, Complete, Ward y Centroid). Como resultado de estos análisis se obtuvo el Coeficiente Cofenético (CC) para cada una de las combinaciones entre las diferentes distancias y los agrupamientos evaluados (observar la tabla 4).

Tabla 4: Tabla de coeficientes de correlación cofenético para el agrupamiento jerárquico basado en distancias a partir de caracteres morfológicos

	Agrupamiento				
Distancia	UPGMA	SINGLE	COMPLETE	WARD.D2	CENTROID
Euclidiana	0.712	0.671	0.642	0.509	0.694
Manhattan	0.777	0.699	0.720	0.621	0.729
Gower	0.833	0.800	0.812	0.806	0.780

A partir de los CC (tabla 4) se pudo inferir que, para el conjunto de 116 accesiones, la distancia de Gower muestra los coeficientes más altos y los que serían los agrupamientos más adecuados de acuerdo con Darkwa *et al.* (2020), ya que estos son mayores o iguales a 0.8.

Adicionalmente, se debe tomar en cuenta que, de las tres distancias evaluadas, la de Gower es la única que emplea tanto caracteres cuantitativos como los cualitativos esto refleja la importancia del uso de variables cualitativas en el proceso de caracterización y no tan solo de las cuantitativas, ya que éstas de tipo categórico son más estables en el medio ambiente que las cuantitativas, lo cual permite obtener un resultado más adecuado (Chávez 2018).

Los resultados al emplear la distancia genética de Gower indican la ausencia de duplicados, ya que es posible asumir que dos accesiones son idénticas morfológicamente si la distancia entre ellas es igual a cero (Franco e Hidalgo 2003). En el presente estudio no se han observado duplicados al emplear este tipo de distancias, ya que todas las accesiones están en mayor o menor grado distanciadas unas de otras, así se obtuvo una distancia media entre todas las accesiones de 0.279, en un rango de 0.037 hasta 0.640, esto se puede visualizar en el histograma presentado en la figura 24.

Así como en el presente estudio, donde la distancia de Gower indica ser la distancia más adecuada para el agrupamiento jerárquico, Gill (2009) menciona que el agrupamiento del germoplasma de frijol fue mejor definido por la distancia de Gower en base a la raza, al color y origen de la semilla. Para evaluar el número de grupos y cuan definidos estos estaban, se generaron dendrogramas de los agrupamientos obtenidos por medio de los algoritmos de UPGMA, Complete, Ward y Single, siendo estos los agrupamientos que igualaron y superaron el CC de 0.8, tal cual se mencionó en párrafos anteriores.

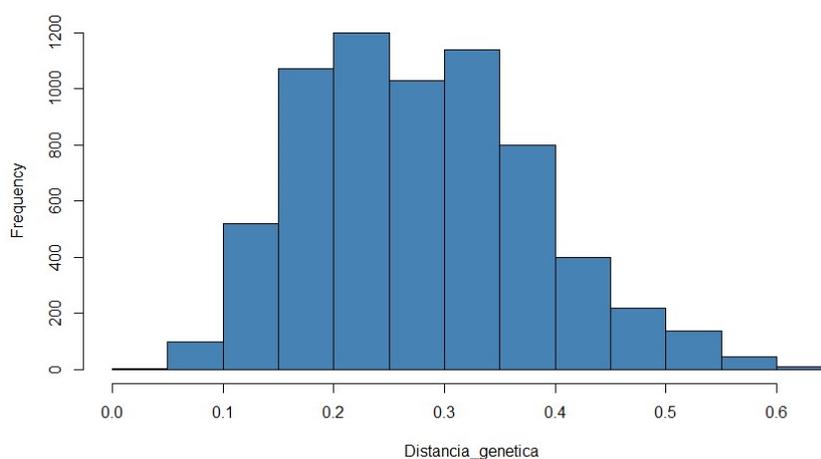


Figura 24: Histograma de la distancia genética de Gower calculada en pares de accesiones en todas las razas

Para estos cuatro tipos de agrupamiento se realizó la prueba de la silueta (Tsindi *et al.* 2023), con la cual para todos ellos se identificaron dos grupos (anexo 9.1). Sin embargo y de acuerdo con la asignación de las razas, brindada por el banco de germoplasma, estos dos grupos formados en cada tipo agrupamiento, no permite agrupar a las accesiones por razas ni por caracteres morfológicos, geográficos ni de origen, lo cual llevó a buscar y emplear otros criterios para la selección del número de grupos más adecuados en los dendrogramas.

Chávez *et al.* (2018) menciona que para la elección del punto de corte en el dendrograma que permita la separación de los grupos, se puede emplear el criterio establecido por Núñez y Escobedo (2011) que sostienen que no existe un criterio definido para establecer dicho punto, por lo que éste puede ser seleccionado en base a los objetivos de la investigación y conocimientos previos del investigador, es por esta razón que en muchos casos este punto de corte es elegido arbitrariamente.

Tabla 5: Tabla del número de accesiones por raza en cada grupo formado mediante el agrupamiento UPGMA y la distancia de Gower

Raza	G1	G2	G3	G4	G5	G6
Chullpi	1	6	32	0	1	0
Paro	0	0	0	1	18	21
Piscorunto	0	0	0	0	36	0
Total	1	6	32	1	55	21

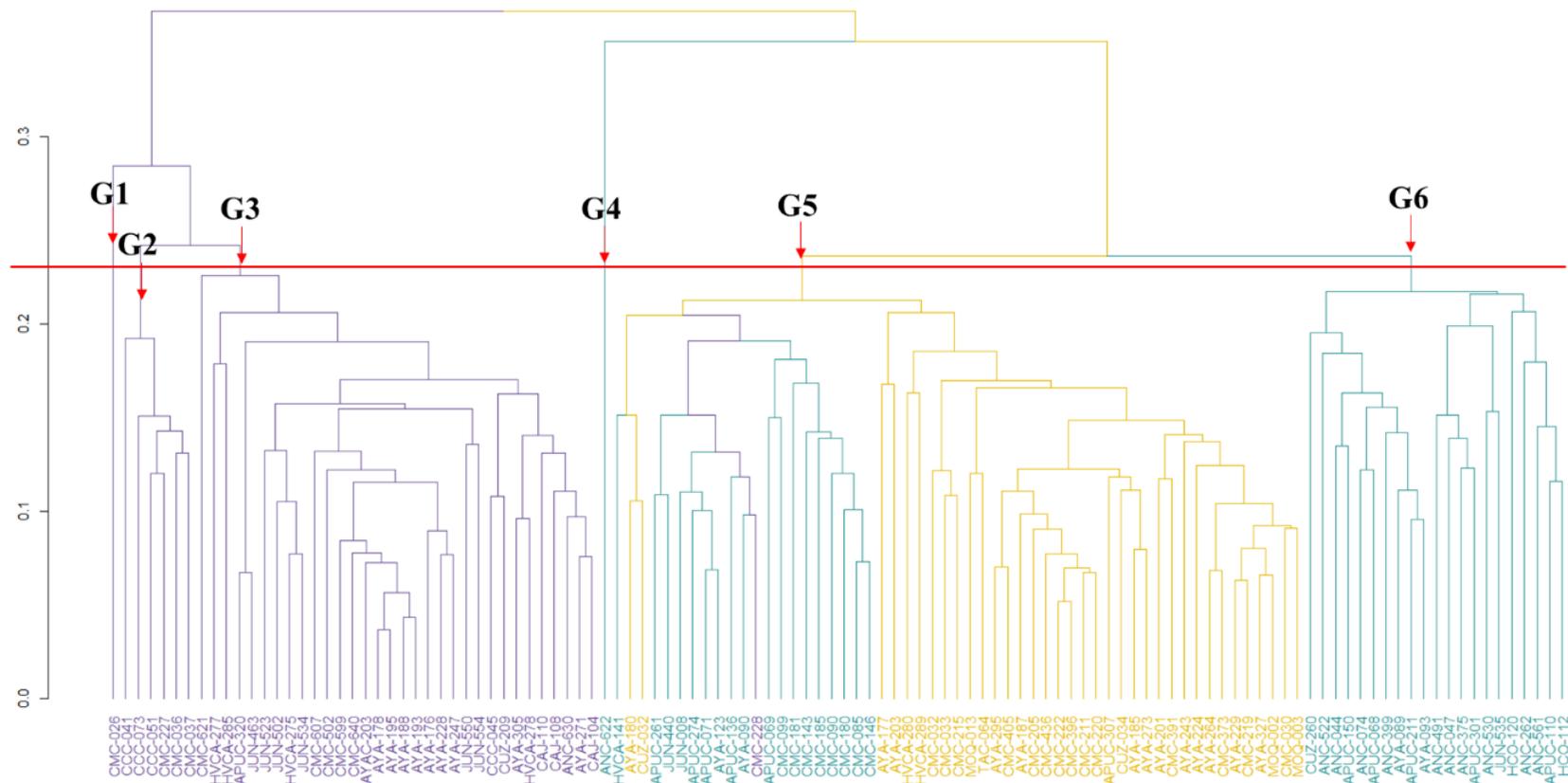


Figura 25: Dendrograma de accesiones de maíz amiláceo empleando el agrupamiento UPGMA y la distancia de Gower

*Accesiones coloreadas de morado pertenecen a la raza Chullpi, de color verde agua a la raza Paro y de color amarillo a la raza Piscorunto. G1: grupo 1, G2: grupo 2, G3: grupo 3, G4: grupo 4, G5: grupo 5 y G6: grupo 6.

Entonces, empleando el criterio establecido por Núñez y Escobedo (2011), el agrupamiento por UPGMA permite establecer el punto de corte aproximadamente a 0.23, con cual se generaron seis grupos (figura 25). Los grupos formados mediante este agrupamiento presenta un número de accesiones dispares en cada uno de ellos. Así en el grupo 1, 2, 3 y 5 se tiene solo una accesión, seis, 32 y una respectivamente de la raza chullpi en estos grupos. Accesiones de la raza paro se encuentra una, 18 y 21 en los grupos 4, 5 y 6 respectivamente. La única raza que tiene a todas sus accesiones en un solo grupo es Piscorunto, con las 36 accesiones en el grupo 5 (tabla 5).

En el caso en que se seleccionara un punto de corte por encima de 0.23, se podría obtener tres grupos, de los cuales uno de ellos tendría las 39 acciones de chullpi, un grupo con tan solo una accesión de paro y un grupo más con 39 accesiones de Paro, 36 de Piscorunto y una de Chullpi, lo cual también nos entregaría grupos con un numero de accesiones muy dispares.

Tabla 6: Tabla del número de accesiones por raza en cada grupo formado mediante el agrupamiento Complete y la distancia de Gower

Raza	G1	G2	G3	G4	G5
Chullpi	7	32	0	0	1
Paro	0	0	1	7	32
Piscorunto	0	0	0	34	2
Total	7	32	1	41	35

Para el agrupamiento tipo Complete, con el punto de corte a 0.35, se obtuvieron cinco grupos (figura 26). Los grupos formados mediante este agrupamiento presenta un número de accesiones diferentes en cada uno de ellos. La raza chullpi presento siete, 32 y una accesión en los grupos 1, 2 y 5 respectivamente; la raza paro tiene una, siete y 32 accesiones en los grupos 3, 4 y 5 respectivamente, y, por último, la raza piscorunto presentó 34 y dos accesiones en los grupos 4 y 5 respectivamente (tabla 6).

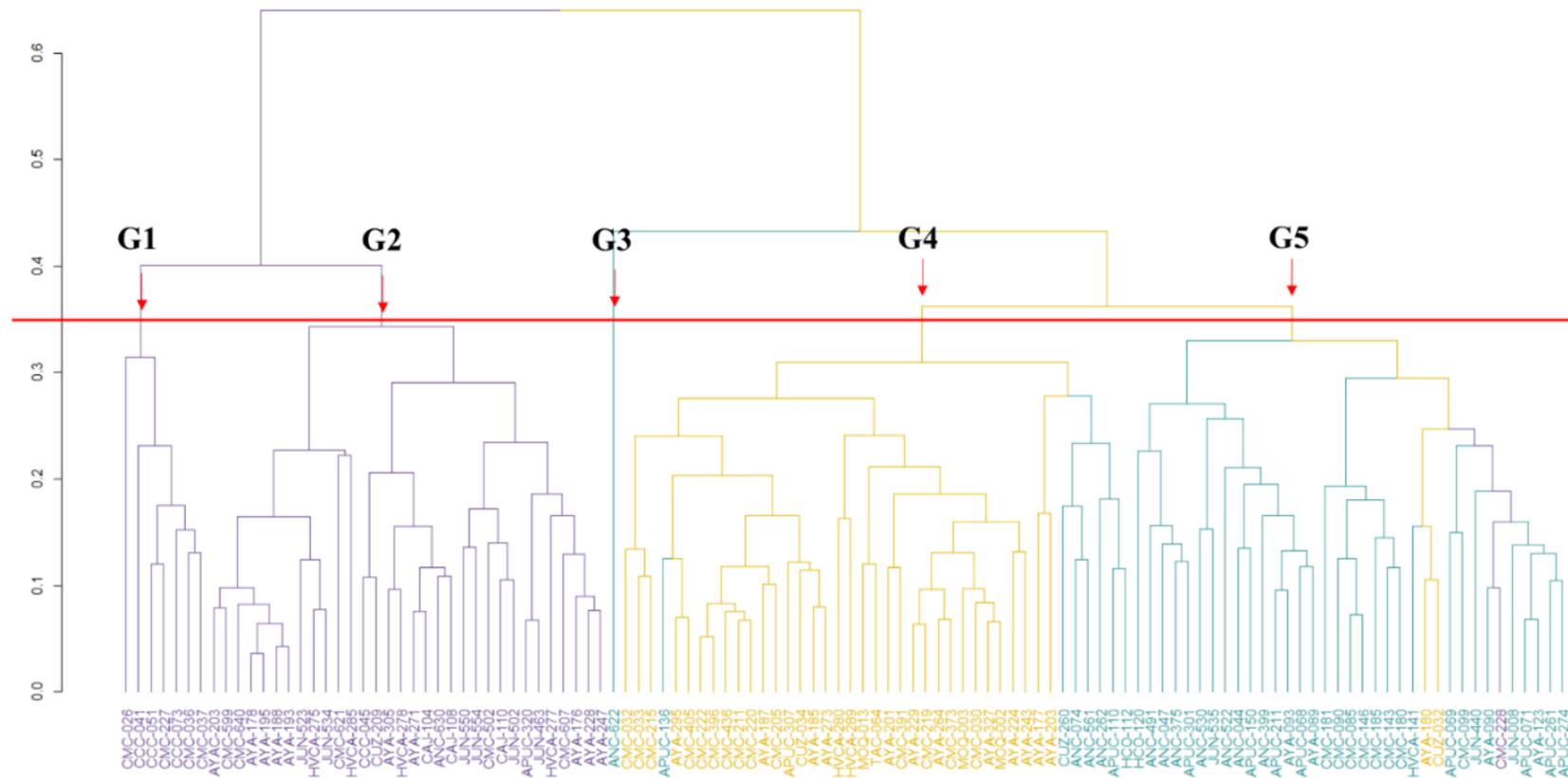


Figura 26: Dendrograma de accesiones de maíz amiláceo empleando el agrupamiento Complete y la distancia de Gower

*Accesiones coloreadas de morado pertenecen a la raza Chullpi, de color verde agua a la raza Paro y de color amarillo a la raza Piscorunto. G1: grupo 1, G2: grupo 2, G3: grupo 3, G4: grupo 4 y G5: grupo 5.

Desde otro punto de corte y superior a 0.35, se formarían dos grupos, uno de ellos compuesto por 39 accesiones de chullpi y el otro grupo con una de chullpi, 36 de piscorunto y 40 de paro. A partir de este agrupamiento Complete y del generado por UPGMA, se puede deducir la cercanía morfológica que existe entre las accesiones de la raza paro y piscorunto, y la distancias que en mayor grado existe de estas dos razas con las accesiones de chullpi.

Tabla 7: Tabla del número de accesiones por raza en cada grupo formado mediante el agrupamiento Single y la distancia de Gower

Raza	G1	G2	G3	G4
Chullpi	0	1	38	1
Paro	1	0	0	39
Piscorunto	0	0	0	36
Total	1	1	38	76

El agrupamiento con el algoritmo de Single muestra cuatro grupos a un punto de corte de 0.16 (figura 27). De la misma manera que con los anteriores agrupamientos, los grupos formados presentan grupos con un número de accesiones muy diferente en cada uno de ellos. El grupo 1 y 2 solamente están conformados por una accesión, cada uno procedente de la raza paro y chullpi respectivamente. El grupo 3 con 38 accesiones de chullpi y el grupo 4, siendo el más numeroso con una accesión de chullpi, 39 de paro y 36 de piscorunto (tabla 7).

Este tipo de agrupamiento es el que ha mostrado tener la estrategia de agrupamiento menos adecuada, ya que si el punto de corte es mayor a 0.16, el número de grupos se reduce a uno con un número muy elevado de accesiones y si es menor a 0.16, el número de grupos aumenta de sobre manera, teniendo en cada grupo de una a dos accesiones.

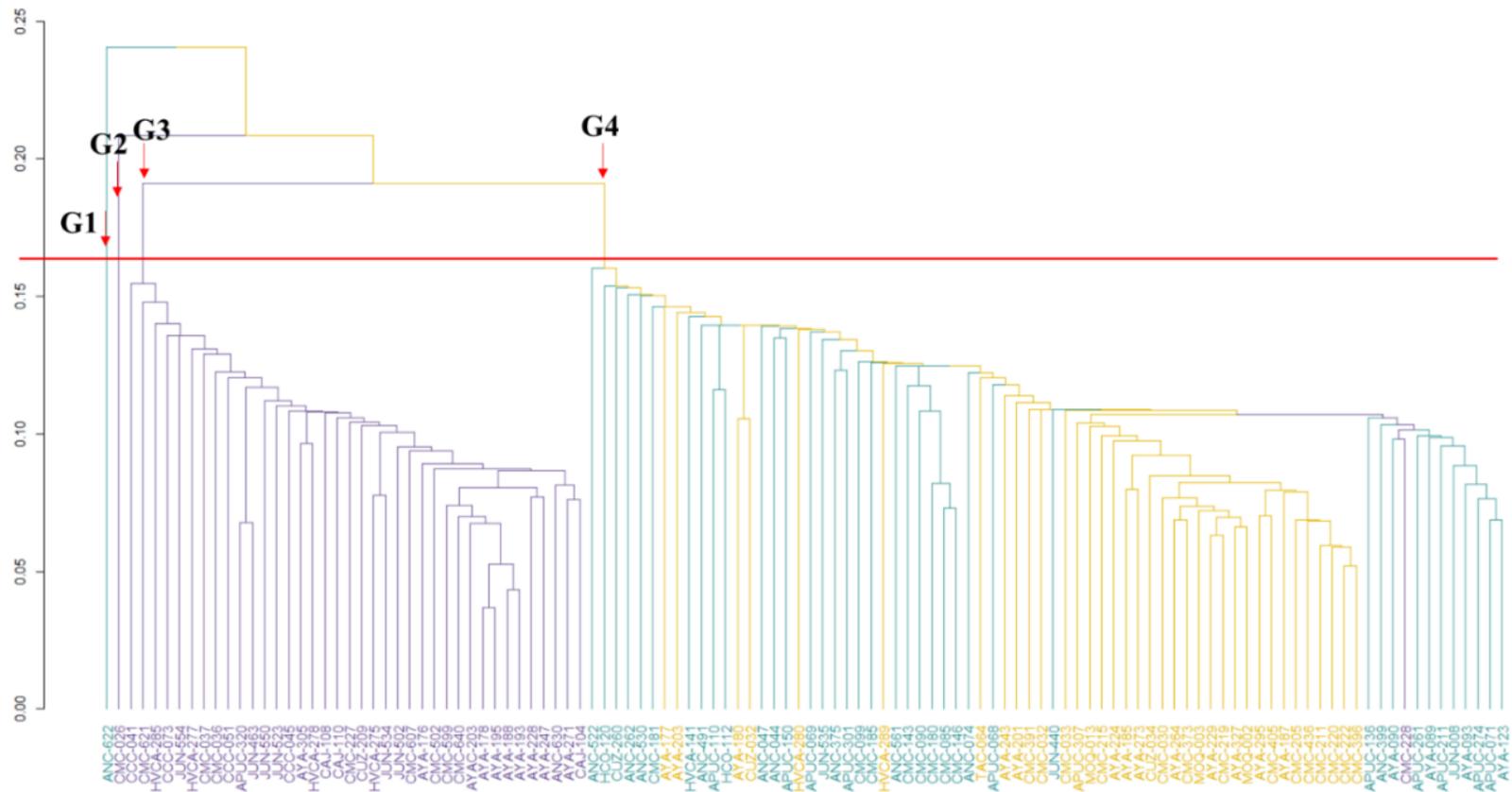


Figura 27: Dendrograma de accesiones de maíz amiláceo empleando el agrupamiento Single y la distancia de Gower

*Accesiones coloreadas de morado pertenecen a la raza Chullpi, de color verde agua a la raza Paro y de color amarillo a la raza Piscorunto. G1: grupo 1, G2: grupo 2, G3: grupo 3 y G4: grupo 4.

Tabla 8: Tabla del número de accesiones por raza en cada grupo formado mediante el agrupamiento Ward y la distancia de Gower

Raza	G1	G2	G3
Chullpi	39	0	1
Paro	0	0	40
Piscorunto	0	34	2
Total	39	34	43

Por último, en el caso del agrupamiento de Ward con el punto de corte a 0.7, se obtuvieron tres grupos (figura 28). Los grupos formados mediante este agrupamiento presentan un número de accesiones diferentes en cada uno de ellos, pero en menor cantidad que con los agrupamientos de UPGMA, Complete y Single. La raza chullpi presentó 39 y una accesión en los grupos 1 y 3 respectivamente; la raza paro tiene todas sus accesiones, es decir 40 en el grupo 3, y, por último, la raza piscorunto presentó 34 y dos accesiones en los grupos 2 y 3 respectivamente (tabla 8).

De manera similar, Franco y Hidalgo (2003) mencionan que para el análisis multivariado utilizando la distancia de Gower y el agrupamiento jerárquico de Ward es posible caracterizar adecuadamente colecciones de germoplasma de interés, lo cual también ha sido comprobado en el presente estudio.

Entre todos los agrupamientos evaluados, el de Ward es el que ha podido agrupar a las accesiones en concordancia con las razas y esto se evidencia en la figura 28, donde claramente se observan tres grupos bien definidos con tan solo tres accesiones que no pertenecen al grupo 3 (observar G3 en la figura 28).

Es necesario resaltar que esto puede haberse dado debido a cierto grado de contaminación ocurrido tanto a nivel del Banco de germoplasma, ya que la recolección de las accesiones procedentes de los agricultores que siembran el maíz en forma extensiva sin ningún tipo de barreras, puede generar polinización cruzada con polen extraño por acción de insectos o del viento, asimismo, esto podría haber ocurrido también en el periodo de cultivo en el campo experimental por la cercanía entre las accesiones dentro de cada bloque, lo cual se puede evidenciar en el presente estudio, en el que para varias accesiones en las razas paro y piscorunto se observan varios estados para los caracteres cualitativos de grano y mazorca (anexo 3).

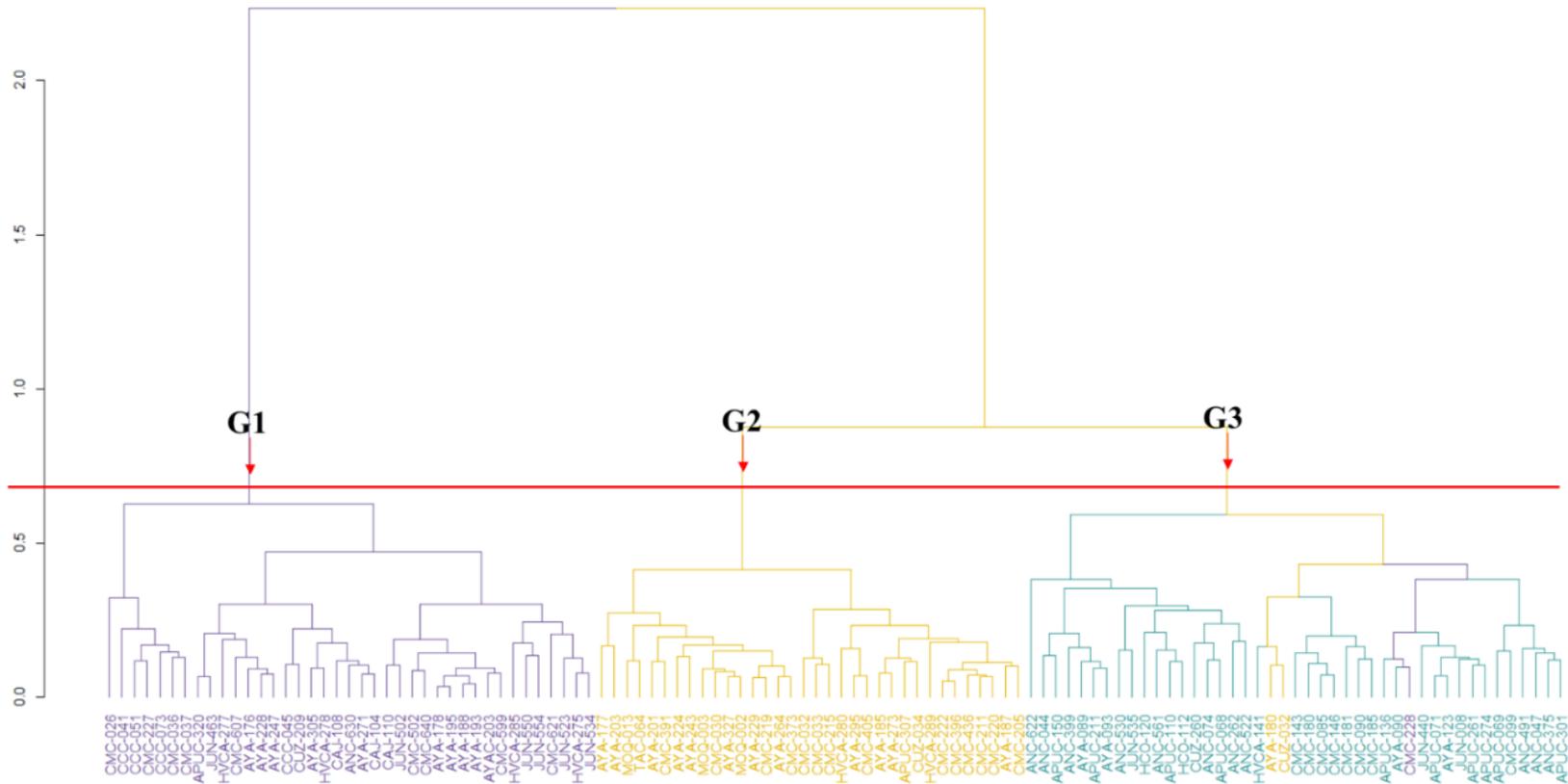


Figura 28: Dendrograma de accesiones de maíz amiláceo empleando el agrupamiento Ward y la distancia de Gower

*Accesiones coloreadas de morado pertenecen a la raza Chullpi, de color verde agua a la raza Paro y de color amarillo a la raza Piscorunto. G1: grupo 1, G2: grupo 2 y G3: grupo 3.

Aunado a esto, se debe tomar en cuenta que las razas se clasifican no solo con criterios morfológicos (cuantitativos y cualitativos), sino también se consideran criterios culturales y ecológicos (Sevilla y Holle, 2004), por lo cual cabe la posibilidad que, durante las etapas de regeneración de la semilla, estas accesiones hayan podido divergir de la raza asignada al ser accesadas al banco de germoplasma.

Análisis multivariado de agrupamiento jerárquico empleando componentes principales

En este segundo enfoque de agrupamiento jerárquico se emplearon cuatro diferentes métodos de CP, y a partir de las CP se realizó el HCPC, el cual emplea el criterio de Ward para la formación de los grupos con las accesiones, el cual fue el más adecuado de emplear para formar los conglomerados a partir de las distancias genéticas para las accesiones del presente estudio.

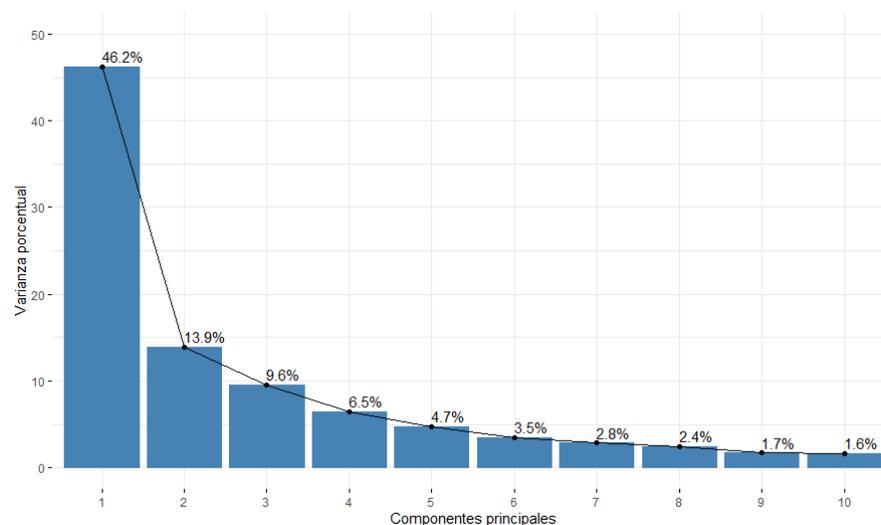


Figura 29: Gráfico de barras de la varianza porcentual de las primeras 10 componentes principales obtenidas del PCA

El PCA fue el primer método de CP ejecutado, el cual permitió reducir el número de caracteres cuantitativos a cuatro componentes adimensionales, los cuales albergaron el 76.16% de la varianza (observar la figura 29), y al mismo tiempo cada una de estas CP tuvo un valor propio mayor o igual a uno (anexos 10.1), que de acuerdo con N'DA *et al.* (2022), es el criterio de Kaiser y que se menciona en esta instancia por ser un buen principio para la selección del número de CP que se deben mantener para analizar la variabilidad, en este caso morfológica cuantitativa, en toda la población.

Empleando estas cuatro primeras CP, se identificó a los caracteres discriminantes, es decir, aquellos que permiten diferenciar de forma más eficiente las muestras caracterizadas (Chávez 2018), para lo cual se evaluó la representatividad y la contribución de cada uno de los caracteres a las primeras cuatro CP.

En la figura 30B, se puede observar el grado de representatividad en el correlograma, donde se observa que, hasta la cuarta CP, todos los caracteres demuestran una alta representatividad por lo menos en una de ellas y por lo tanto en estos términos, todos son igualmente importantes e imprescindibles para representar la variabilidad hallada en esta población de maíz.

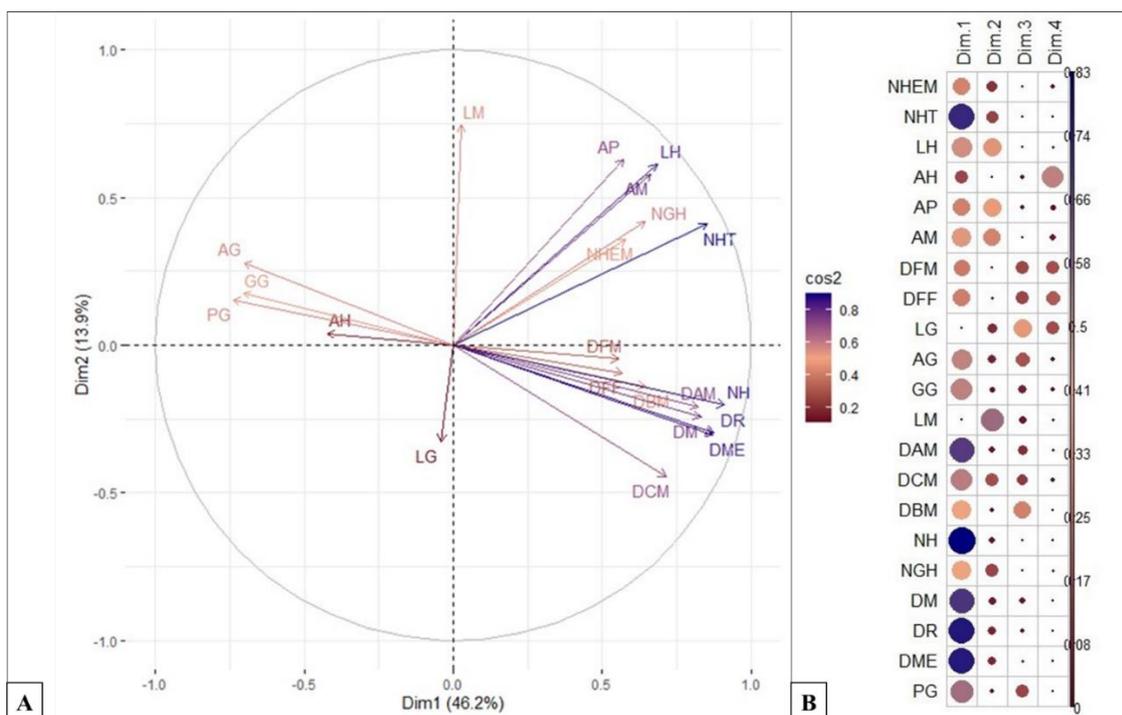


Figura 30: A. Gráfico “Biplot” de correlación entre la calidad de la representación de los caracteres cuantitativos en las dos primeras componentes principales del PCA. B. Correlograma entre la calidad de la representación de los caracteres cuantitativos desde el primer al cuarto componente principal del PCA

De este tipo de análisis también se puede desprender un gráfico tipo “Biplot”, en el cual se puede apreciar algunas relaciones presentes entre los caracteres cuantitativos (figura 30^a). Debido a que estamos utilizando datos estandarizados, las coordenadas de las variables sobre

cada componente principal son iguales a la correlación entre las variables originales y los componentes principales (Macuri 2016).

Así mismo, empleando la metodología de “k-means” se logró hallar tres grupos entre todos los caracteres cuantitativos (anexo 10.3), esto también fue empleado para resaltar las correlaciones entre los caracteres y su comparación para denotar la diversidad genética (Zavinon *et al.* 2019 y Al-Naggar *et al.* 2022), los cuales se encuentran organizados y vinculados de la siguiente manera:

Grupo 1: compuesto por AP, AM, LM, LH, NGH, NHT y NHEM, que están positivamente correlacionados, y todos estrechamente relacionados con la arquitectura de planta y mazorca, estos resultados concuerdan con los obtenidos el análisis de correlaciones desarrollado en el apartado anterior.

Grupo 2: conformado por LG, AG, GG, PG y AH, estos cuatro últimos están negativamente correlacionados con los caracteres del grupo 3. Además, todos los caracteres de este grupo, excepto la AH, están directamente relacionados con las dimensiones y el rendimiento del grano.

Grupo 3: constituido por DFM, DFF, DM, DBM, DAM, DCM, DR, DME y NH, los cuales están correlacionados negativamente con el grupo 2, siendo esto corroborado por lo mencionado en el análisis de correlación. Además, en este grupo se tiene a todos los caracteres relacionados con el diámetro de la mazorca. Cabe mencionar que en la figura 30^a se observa la superposición de variables como DR, DM y DME, lo cual reafirma su alta correlación, siendo esto mencionado también por Chávez *et al.* (2018).

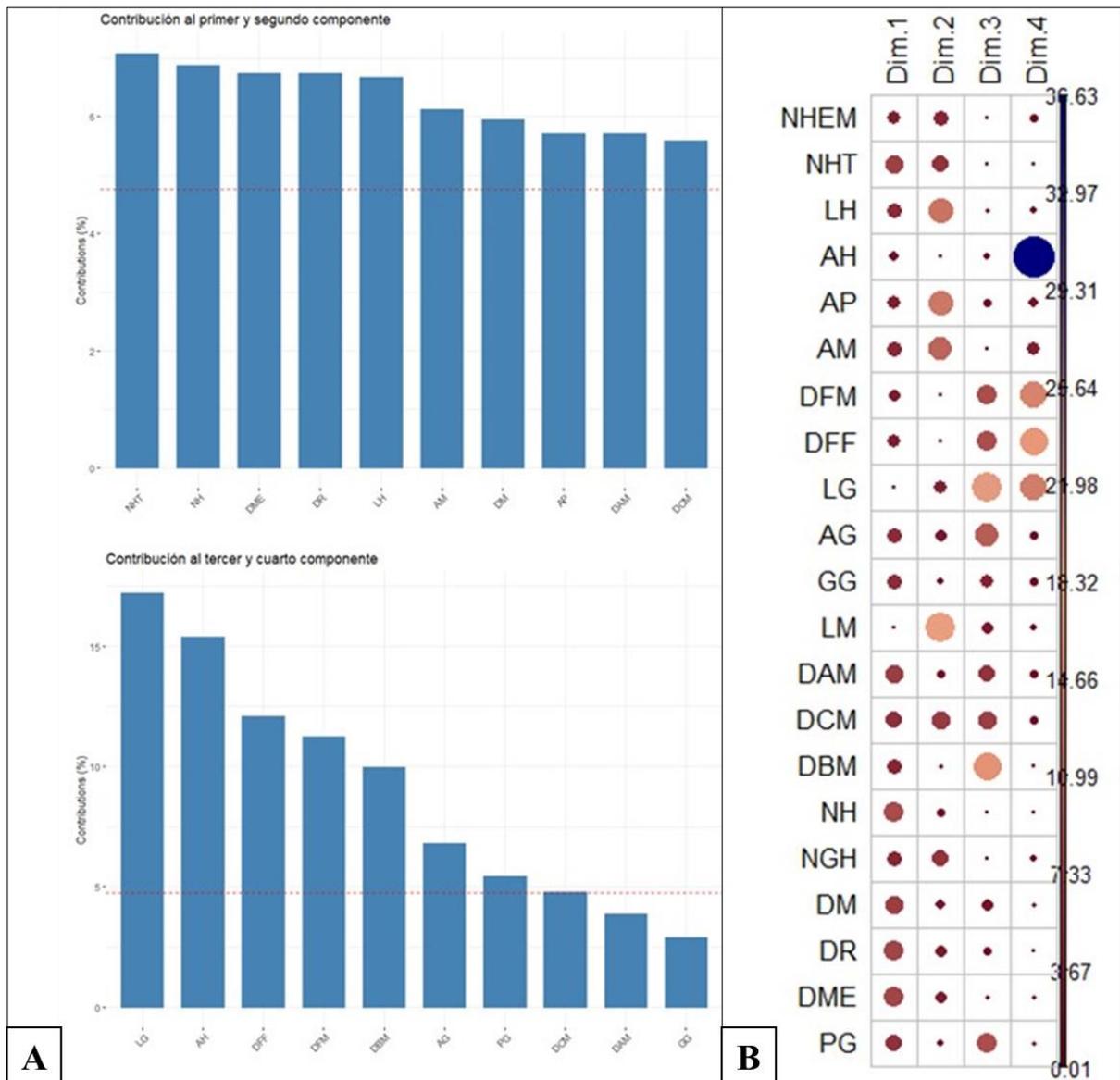


Figura 31: A. Gráfico de barras de la contribución de los caracteres cuantitativos en las cuatro primeras componentes principales del PCA. B. Correlograma entre la contribución de los caracteres cuantitativos desde el primer al cuarto componente principal

La interpretación de las correlaciones entre los caracteres originales y los CP, también se deben focalizar en la contribución de los mismos; mientras más altos sean estos coeficientes de correlación, independientemente del signo, más eficientes serán en la discriminación de las accesiones; por tanto, los caracteres con coeficiente negativo significan que están caracterizando en sentido contrario en relación con los caracteres positivos y viceversa (Macuri 2016).

Tomando en cuenta lo anterior, en cuanto a la contribución todos los caracteres excepto por LG y LM, todos aportan significativamente en el primer componente (anexo 10.2). No obstante, algunos caracteres que no lo hacen en el primer componente como LG, si lo hace en el tercer y/o cuarto CP, así como LM en el segundo (figura 31B), por lo cual, en términos de contribución, tampoco podemos prescindir de ningún carácter, ya que todos están aportando a la variabilidad dentro de las componentes seleccionadas para estos análisis, y esto se puede observar en la figura 31^a.

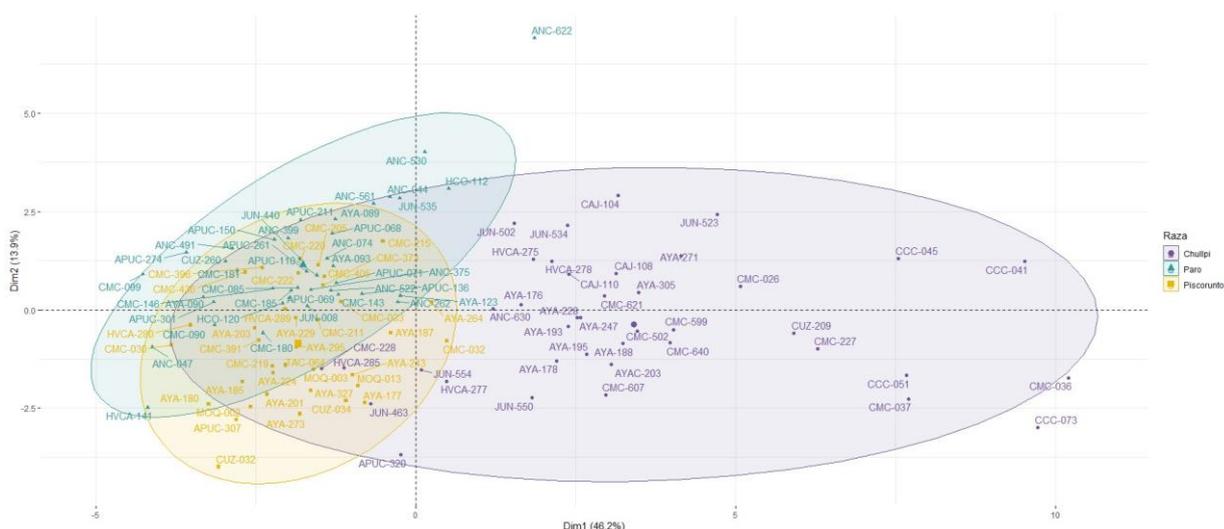


Figura 32: Gráfico “Biplot” de las accesiones en las dos primeras componentes principales del PCA

*Accesiones coloreadas de morado pertenecen a la raza Chullpi, de color verde agua a la raza Paro y de color amarillo a la raza Piscorunto.

Por otro lado, para las accesiones se generó un “Biplot” en los dos primeros CP (figura 32) en cual se pudo observar la dispersión de las mismas y las agrupaciones formadas entre ellas, de las cuales se puede inferir que las accesiones de la raza chullpi son más homogéneas dentro de su misma raza, ya que forman una elipse alejada de las otras dos razas.

Por el contrario, entre las accesiones de la raza paro y piscorunto no hay una diferenciación muy visible, ya que las elipses generadas para cada raza se ven solapadas hasta la mitad de su área. Lo anteriormente expuesto en el enfoque con las distancias genéticas, entrego resultados similares en cuanto a las accesiones y los grupos que se formaron entre ellas.

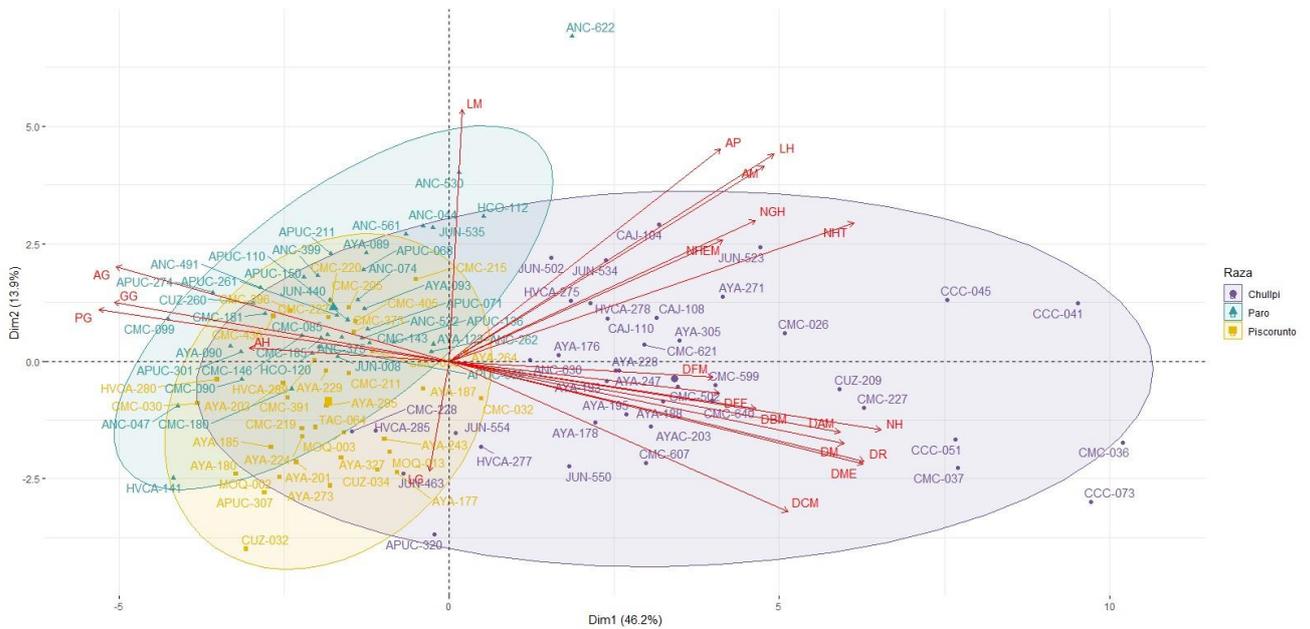


Figura 33: Gráfico “Biplot” de las accesiones y caracteres cuantitativos en las dos primeras componentes principales del PCA

*Accesiones coloreadas de morado pertenecen a la raza Chullpi, de color verde agua a la raza Paro y de color amarillo a la raza Piscorunto.

Recapitulando lo ya mencionado acerca de los caracteres y de las accesiones, también se realizó un “Biplot” adicional, en el cual se puede observar las relaciones entre estas (figura 33). Por ejemplo, en las accesiones que están del mismo lado que los caracteres se pueden encontrar valores altos para esa variable, como ocurre con el NH, que en el caso de las accesiones de la raza chullpi son las que tiene los valores más altos, esta misma información se puede corroborar en el análisis realizado a la estadística descriptiva.

Al mismo tiempo, y continuando con la raza chullpi, se puede mencionar que las accesiones que estén del lado opuesto de ese carácter tienen menores valores, así se tiene para las dimensiones del grano como GG y AG, los valores más bajos se encuentran en las accesiones de chullpi, lo cual caracteriza al grano de esta raza. En consecuencia, y a partir de estas inferencias, se puede deducir que este tipo de gráficos puede ayudar a la caracterización de las accesiones de manera más eficaz.

Adicionalmente, el PCA Biplot reveló las accesiones y los caracteres que se solapan como vectores en la figura 33, esto sugiere que cuando se escoja las accesiones para el programa

de maíz que maneja el Banco de germoplasma, se le debe dar prioridad a los caracteres que se solapan a esas accesiones (Sinana *et al.* 2023).

Anteriormente, se había mencionado la importancia de usar tanto los caracteres cuantitativos como los cualitativos para la caracterización de las accesiones, sin embargo, existía la posibilidad de que tan solo empleando los cualitativos se tuviera la misma capacidad de evaluar la variabilidad, que, si en lugar de ello, se emplearan los dos tipos de variables en conjunto, ya que al evaluar solo las cualitativas permitiría ahorrar esfuerzo, tiempo y presupuesto en la toma de datos y diseño experimental.

En tal sentido, se ejecutó el método de MCA que emplea únicamente a los caracteres cualitativos para formar las CP, con la finalidad de determinar si los caracteres cualitativos permiten evaluar la diversidad.

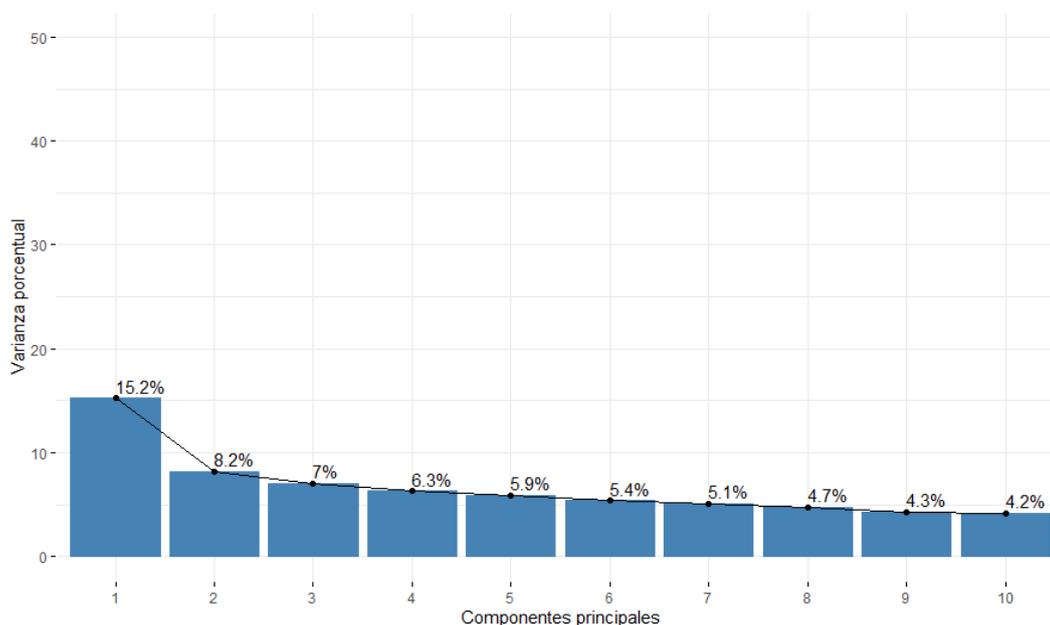


Figura 34: Gráfico de barras de la varianza porcentual de las primeras 10 componentes principales obtenidas del MCA

De la aplicación del MCA, se obtuvieron 26 CP con las cuales se obtuvo el 100% de varianza explicada (anexo 10.1), de estas se retuvieron las 13 primeras, que albergaron el 77.50% de la varianza para los análisis posteriores (figura 34). Con respecto al valor propio de cada CP, se debe mencionar que ninguna de las 13 CP seleccionadas llegó a pasar el valor de uno, lo

cual es un indicador de lo inadecuadas que pueden llegar a ser estas CP para el análisis de la variabilidad.

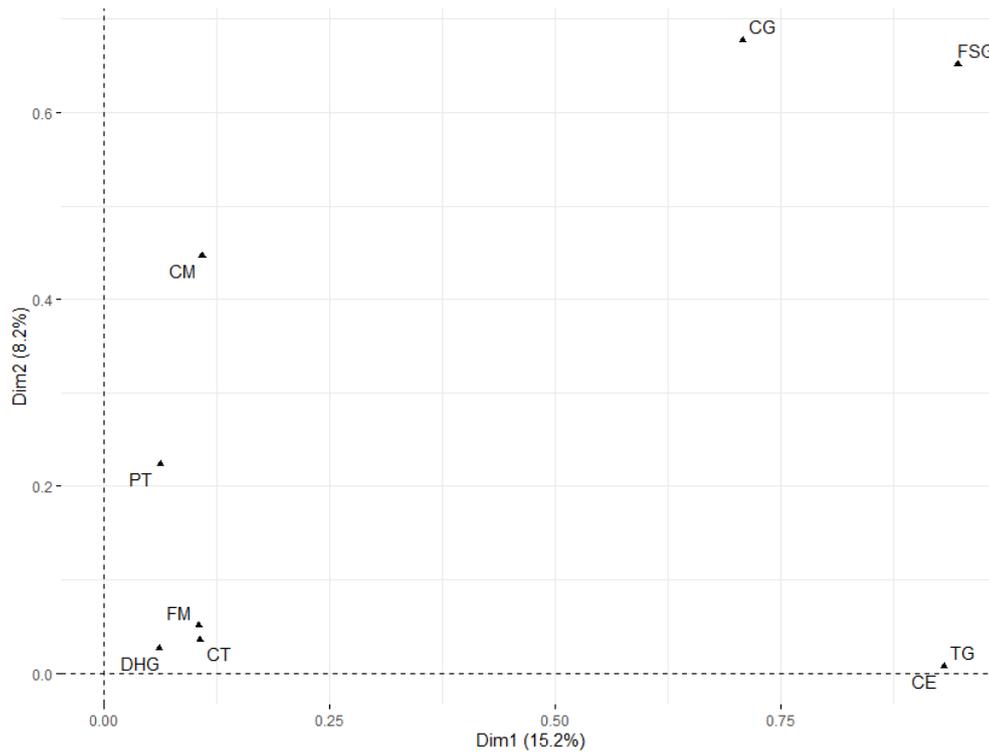


Figura 35: Gráfico “Biplot” de los caracteres cualitativos en las dos primeras componentes principales del MCA

El MCA también permitió dilucidar algunas relaciones existentes entre los caracteres cualitativos. Por ejemplo, tanto el CE, TG, FSG y CG se encuentran altamente correlacionadas con el primer CP, esto se puede observar en la figura 35. Así como el CM y PT, que se encuentra correlacionados con el segundo CP.

Si bien estas correlaciones con las CP pueden ser alta y estadísticamente significativas, observar los valores de los coeficientes de determinación en el anexo 10.2, esto no significa que son capaces de explicar la variabilidad, ya que si sumamos la varianza acumulada en estas dos primeras CP solamente se alcanza 23.4 %.

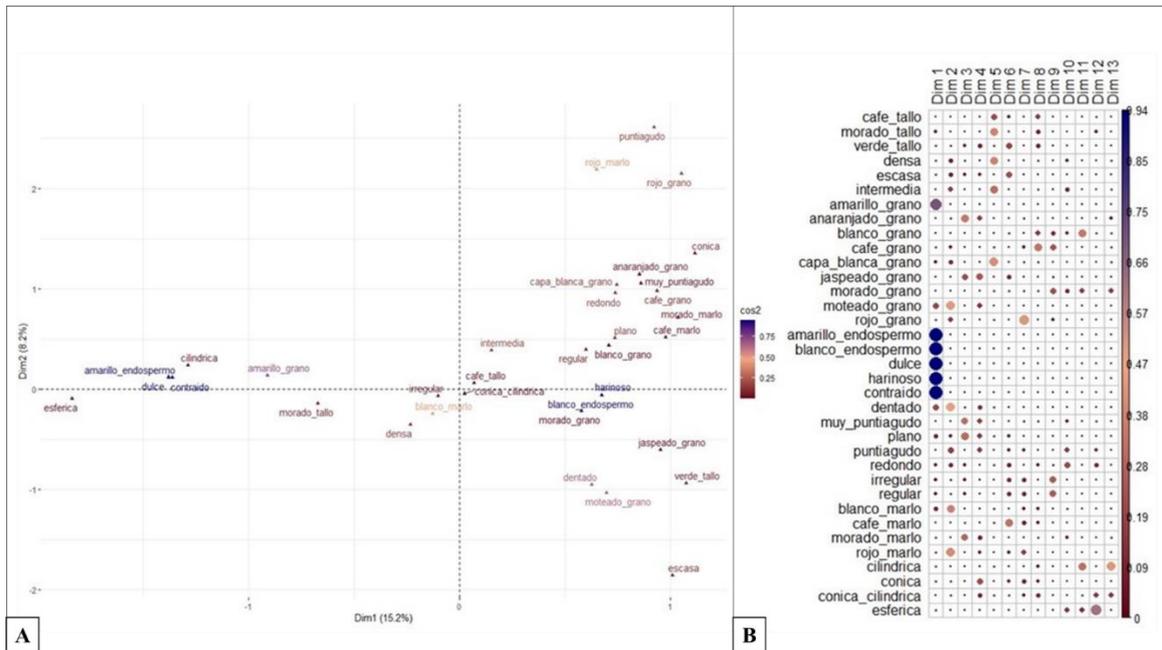


Figura 36: A. Gráfico “Biplot” de correlación entre la calidad de la representación de los caracteres cualitativos en las dos primeras componentes principales del MCA. B. Correlograma entre la calidad de la representación de los caracteres cualitativos desde el primer al décimo tercer componente principal del PCA

En paralelo, otra manera de observar que tan pertinente puede ser utilizar únicamente los caracteres cualitativos es evaluando la representatividad y la contribución de estos a las CP seleccionadas.

En la figura 36^a se puede observar la dispersión de los estados en los caracteres cualitativos y como estos se asocian entre sí, tal cual como en la figura 35. Mientras que en la figura 36B, se observa la calidad de la representación de todos los estados, en los que destacan en el primer CP, los que se encuentran relacionados con el CE y TG ya que poseen una tiene alta calidad.

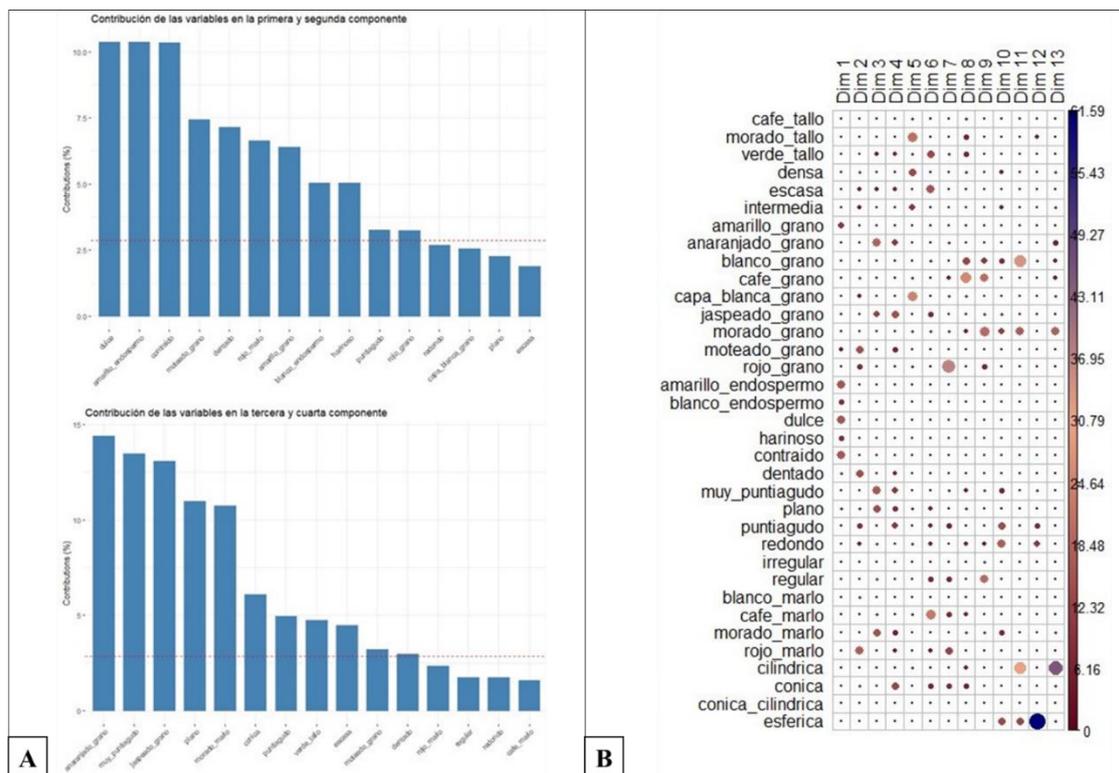


Figura 37: A. Gráfico de barras de la contribución de los caracteres cualitativos en las cuatro primeras componentes principales del MCA. B. Correlograma entre la contribución de los caracteres cualitativos desde el primer al décimo tercer componente principal del MCA

En cuanto a la contribución de los caracteres, los que más contribuyen en la primera CP son el TG (con los estados dulce y harinoso) y CE (con los estados amarillo y blanco), y para los demás caracteres, sus estados contribuyen desde la segunda hasta la treceava CP, esto no solo es observable en la figura 37, sino que estadísticamente los coeficientes de determinación también resultaron significativos (anexo 10.2).

De la evaluación de la representatividad y contribución se desprende que estos dos caracteres, el TG y CE son los que permiten discriminar las razas entre la mayor parte de las accesiones, y se puede inferir que serían de los caracteres cualitativos más importantes por evaluar. Esto también se muestra en el “Biplot” realizado para las accesiones y los caracteres (figura 38 y figura 39^a), en donde se observa una clara separación de las accesiones de la raza chullpi, con las de las razas paro y piscurunto.

Similarmente, esto fue obtenido como resultado en el PCA y en el enfoque por distancias genéticas, lo cual nos permite deducir que chullpi es una raza que mantiene sus características cualitativas y cuantitativas a pesar de algunos procesos de contaminación, ya mencionados en párrafos anteriores.

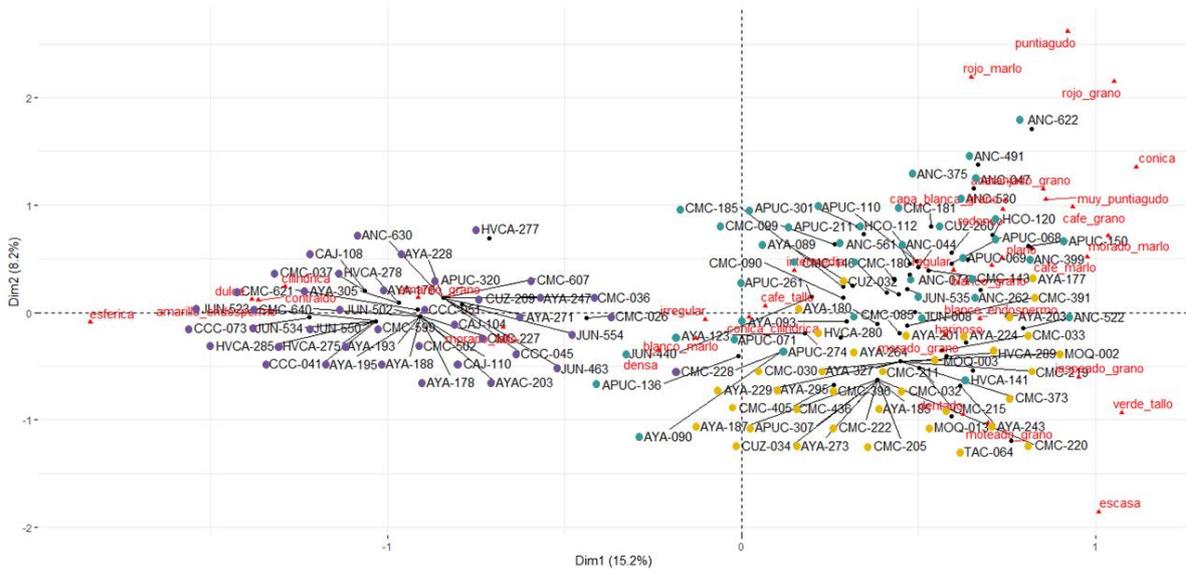


Figura 38: Gráfico “Biplot” de las accesiones y caracteres cualitativos en las dos primeras componentes principales del PCA

*Accesiones con el círculo de color morado pertenecen a la raza Chullpi, de color verde a la raza Paro y de color amarillo a la raza Piscorunto.

Además del CE y TG, en la figura 38 también se observa que los estados del CG y al FSG permiten diferenciar a las accesiones entre razas y no solo de chullpi sino también entre paro y piscorunto, cuyas diferencias son más complejas de identificar, ya que varias accesiones solapan posiciones dentro del “Biplot”. En la figura 39B, se muestran los diferentes estados de estos dos caracteres y como permiten una mejor diferenciación de estas accesiones en estas dos razas.

Sin duda, el MCA ha permitido evaluar la variabilidad presente en las accesiones, más cabe mencionar que no ha logrado disminuir la dimensionalidad de los caracteres empleados, ya que fueron nueve los caracteres originales y se retuvieron 13 CP, esto tomando en cuenta los criterios ya mencionados, que además estadísticamente todos estos caracteres están correlacionados con las CP retenidas para el análisis.

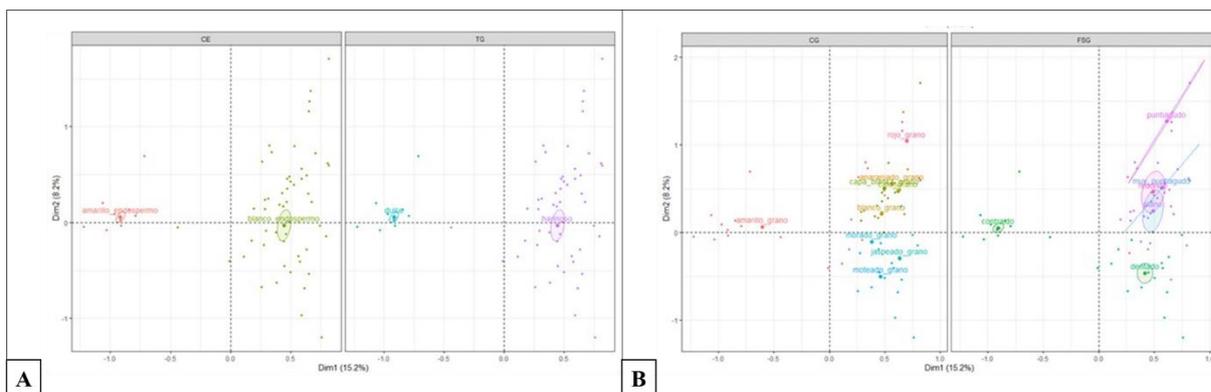


Figura 39: A. Gráfico “Biplot” de las accesiones y los caracteres cualitativos CE y TG. B. “Biplot” de las accesiones y los caracteres cualitativos CG y FSG

Sin embargo, este enfoque de CP si nos ha permitido identificar a los caracteres cualitativos más importantes para la discriminación de las accesiones por raza dentro de los nueve caracteres originales.

Un tercer método de CP fue aplicado al conjunto completo de caracteres, es decir para este análisis se emplearon los cuantitativos y cualitativos. Este método es conocido como FAMD, y mediante este procedimiento se retuvo el 75.52% de la varianza en 14 CP, con valores propios mayores a uno, lo cual cumple con el criterio de Kaiser, que nos permite continuar los análisis posteriores con estas 14 CP, las primeras 10 se muestran en la figura 40.

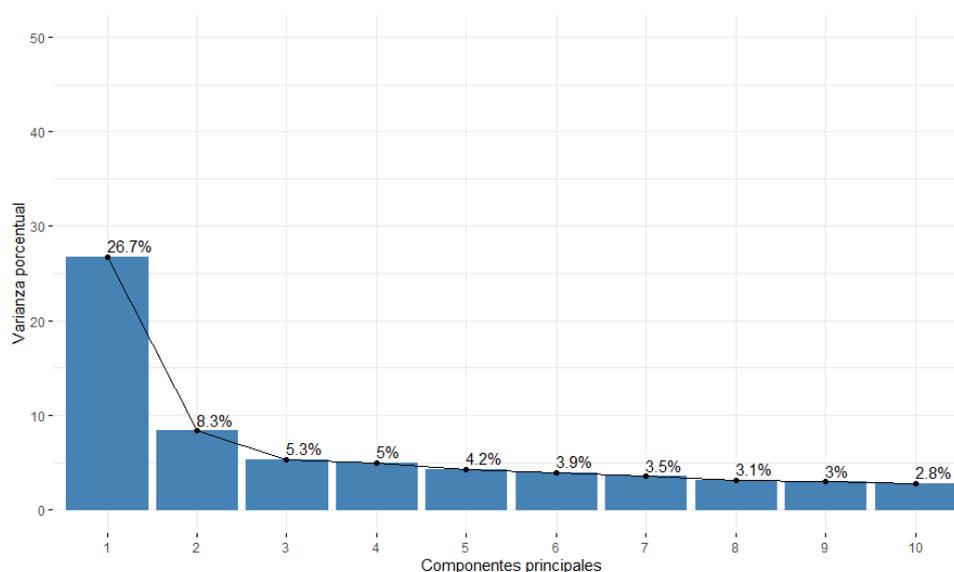


Figura 40: Gráfico de barras de la varianza porcentual de las primeras 10 componentes principales obtenidas del FAMD

Una gran ventaja de FAMD es que permite emplear ambos tipos de caracteres y a partir de este análisis se pueden inferir relaciones entre ellos, a diferencia de los dos anteriores métodos de CP empleados, en donde si bien se observaron relaciones estas fueron entre un solo tipo de carácter.

Para ejemplificar lo mencionado antes, se puede observar en la figura 41, la relación que existe entre el TG y CE, que son caracteres cualitativos, con el NH, DAM, DR, DME que son caracteres cuantitativos. En conjunto estos caracteres son los que permiten la separación de las accesiones de la raza chullpi con respecto a paro y pisorunto.

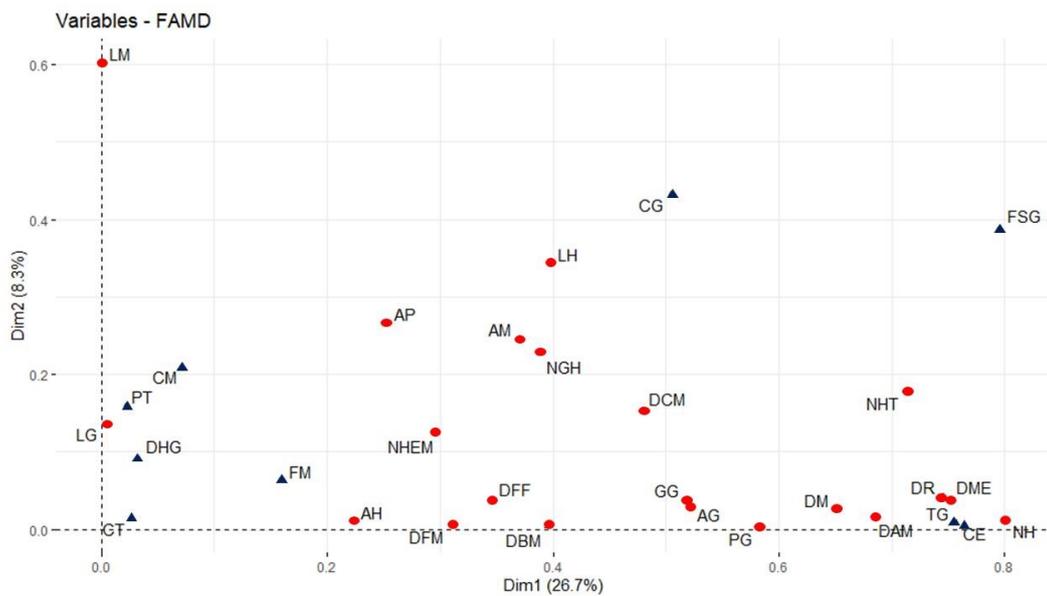


Figura 41: Gráfico “Biplot” de los caracteres cuantitativos y cualitativos en las dos primeras componentes principales del FAMD

Lo anteriormente expuesto para el PCA y MCA, nos da a entender que la contribución de los caracteres a las CP es un buen indicador de la importancia del carácter para discriminar a las accesiones. En cuanto a este principio, en la figura 42B se puede observar como todos los caracteres contribuyen de manera similar en las 14 CP seleccionadas para la evaluación, permitiendo inducir la importancia de todas para la evaluación de la variabilidad de la población.

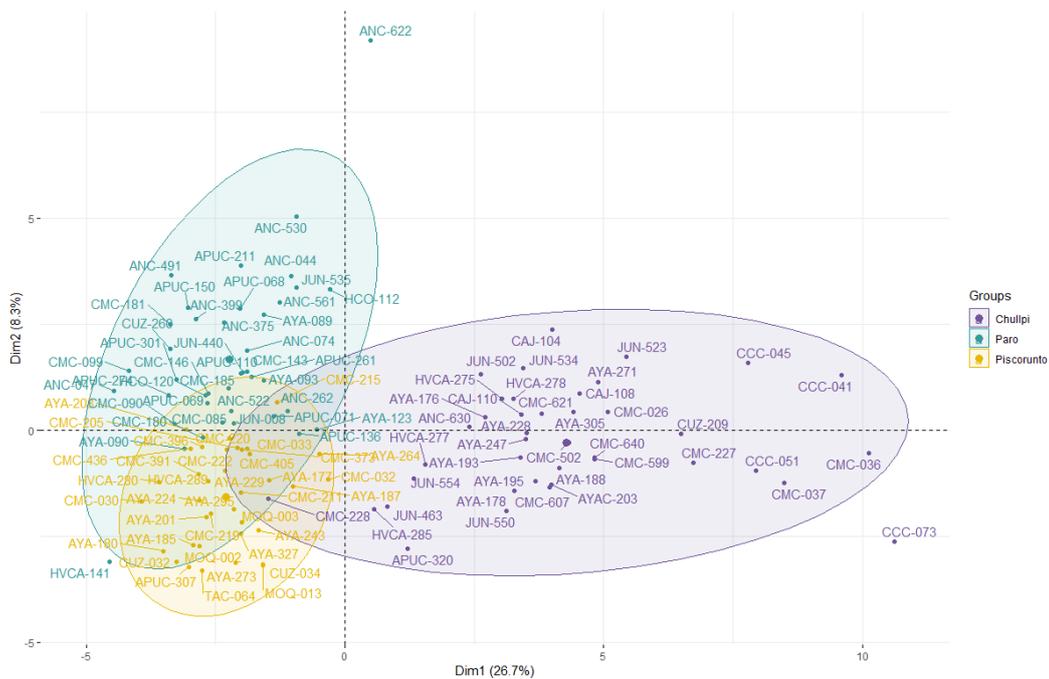


Figura 43: Gráfico “Biplot” de las accesiones en las dos primeras componentes principales del FAMD

*Accesiones coloreadas de morado pertenecen a la raza Chullpi, de color verde agua a la raza Paro y de color amarillo a la raza Piscorunto.

Al mismo tiempo, nos expone la superposición de varias de las accesiones de las razas paro y piscorunto, lo cual fortalece la idea de la posible contaminación al nivel de las accesiones entre estas dos razas, lo que complica la capacidad de diferenciarlas con los métodos de agrupamiento jerárquico.

Por último, se realizó el MFA, que de manera similar al FAMD emplea ambos tipos de caracteres, este último método de CP se enfoca en agrupar los caracteres más allá de solo el tipo de variable que sean (es decir, si es cuantitativa o cualitativa) sino que se pueden agrupar de acuerdo a otros criterios.

Para este análisis se agruparon los caracteres de la siguiente manera:

Grupo de caracteres cualitativos de la planta en estado vegetativo (Cuali_plant_est_veg) conformado por CT y PT.

Grupo de caracteres cuantitativos de la planta en estado vegetativo (Cuanti_plant_est_veg) compuesto por NHT, NHEM, AP, LH, AM y AH.

Grupo de caracteres cualitativos de la mazorca (Cuali_mazorca) conformado por DHG, CM y FM.

Grupo de caracteres cuantitativos de la mazorca (Cuanti_mazorca) compuesto por LM, DAM, DCM, DBM, NH, NGH, DM, DR y DME.

Grupo de caracteres cualitativos del grano (Cuali_grano) conformado por CG, CE, TG y FSG.

Grupo de caracteres cuantitativos del grano (Cuanti_grano) compuesto por LG, AG, GG y PG.

Grupo de caracteres cuantitativos en la floración (Floracion) constituido por DFM y DFF.

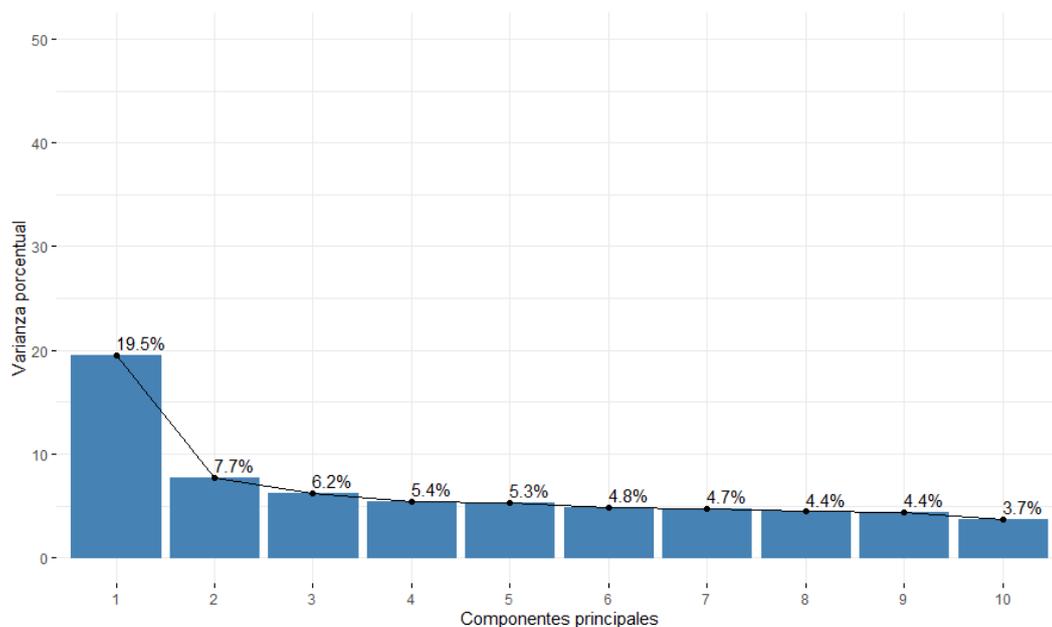


Figura 44: Gráfico de barras de la varianza porcentual de las primeras 10 componentes principales obtenidas del MFA

De este análisis se obtuvieron en total 46 CP, lo cual excede el número de caracteres puestos en evaluación, por ello para reducir el número de dimensiones, se escogieron las primeras 13, que albergaron el 75.42 % de la varianza (figura 44). Aunque de estas, solo las cuatro primeras tienen un valor propio mayor de uno, indicando esto que estas CP deben ser las que se utilicen para evaluar la variabilidad a través de este método.

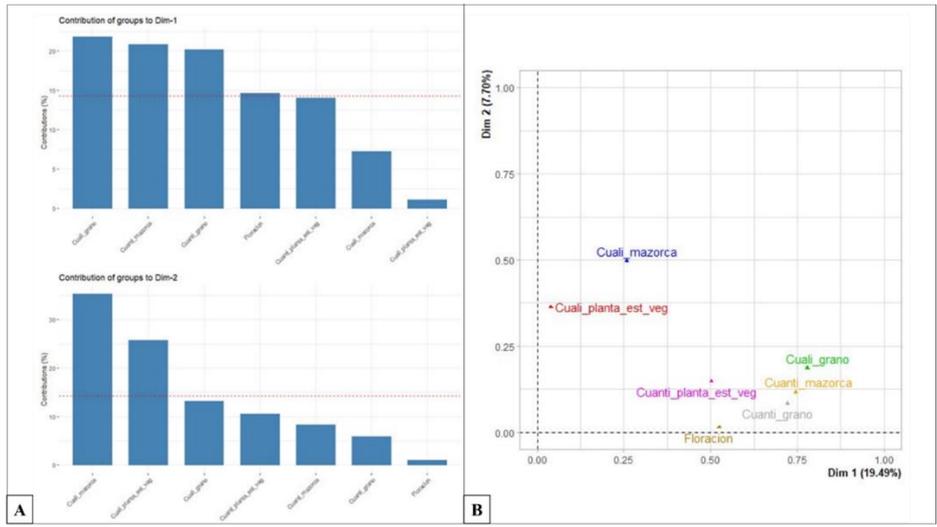


Figura 45: A. Gráfico de barras de la contribución de los caracteres agrupados en las dos primeras componentes principales del MFA. B. “Biplot” de los caracteres agrupados en las dos primeras componentes principales del MFA

Por otra parte, este MFA también ha permitido identificar qué grupos de caracteres son los que más ayudan a discriminar a las accesiones entre las razas. En la figura 45^a, se puede observar que los caracteres de grano son los que más contribuyen a la diferenciación, seguidos de los de mazorca, luego los de la planta en estado vegetativo y por último, los de floración.

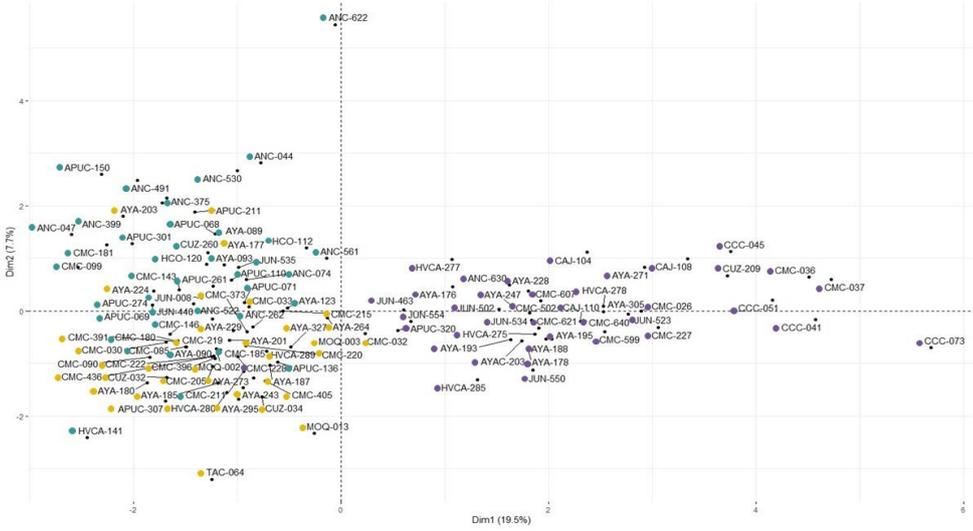


Figura 46: Gráfico “Biplot” de las accesiones en las dos primeras componentes principales del MFA

*Accesiones con el círculo de color morado pertenecen a la raza Chullpi, de color verde a la raza Paro y de color amarillo a la raza Piscorunto.

Igualmente, a lo realizado para los anteriores métodos de CP, se generó un “Biplot” para las accesiones, donde se obtuvieron resultados muy similares a los del PCA, MCA y FAMD. En los que las accesiones de la raza chullpi se encuentran alejados de paro y piscorunto, y que estas dos últimas son más complicadas de diferenciar entre sí. Es interesante observar que esto también se halló en el enfoque empleando distancias genéticas, contemplado en los dendrogramas resultantes de ese análisis.

El análisis de los CP nos permite seleccionar a las más importantes y que retienen la mayor cantidad de la varianza en el mínimo número de CP, de acuerdo a los criterios mencionados con anterioridad. Estos CP se emplearon para efectuar el HCPC, que combina el agrupamiento jerárquico y el análisis de CP para describir las relaciones entre las accesiones y encontrar los grupos más adecuados que permitan diferenciarlas, encontrando dentro de los mismos a las accesiones más homogéneas y entre ellos a las más heterogéneas (Zavinon *et al* 2019).

El HCPC con las cuatro primeras CP del PCA evidenció cuatro grupos a un punto de corte de 1.2 (figura 47). El primer grupo contiene seis accesiones de la raza chullpi, una de paro y 21 de piscorunto. El grupo 2 tiene 35 de paro y 15 de piscorunto, el grupo 3 únicamente tiene nueve de chullpi y el grupo 4 tiene 25 de chullpi y 4 de paro (tabla 9). Si el punto de corte se tomará a un nivel de 2.1, se formarían dos grupos el primero conformado por el grupo 1 y 2 y el segundo formado por el 3 y 4.

Tabla 9: Tabla del número de accesiones por raza en cada grupo formado mediante el HCPC con las componentes del PCA

Raza	G1	G2	G3	G4
Chullpi	6	0	9	25
Paro	1	35	0	4
Piscorunto	21	15	0	0
Total	28	40	9	29

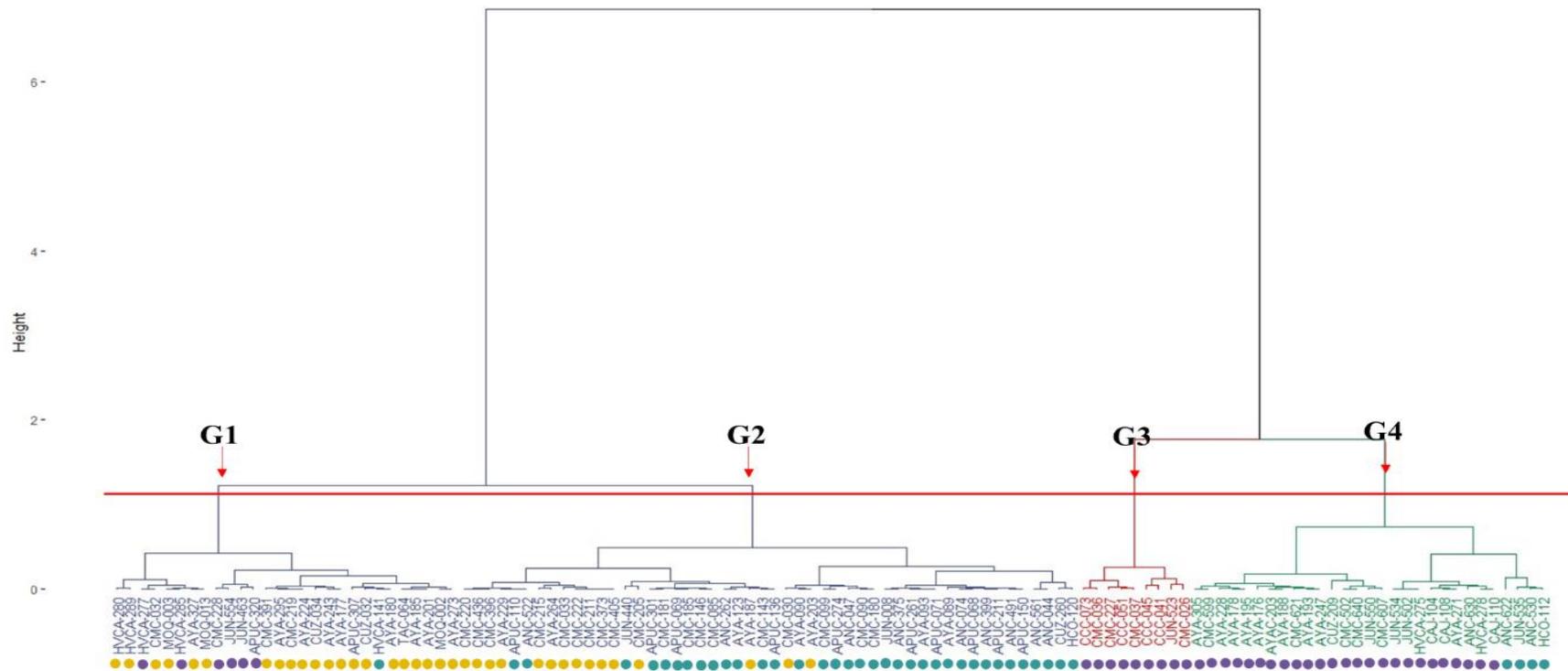


Figura 47: Gráfico de HCPC de accesiones de maíz amiláceo empleando las primeras cuatro componentes principales del PCA

*Accesiones con un círculo de color morado pertenecen a la raza Chullpi, de color verde agua a la raza Paro y de color amarillo a la raza Piscorunto.

G1: grupo 1, G2: grupo 2, G3: grupo 3 y G4: grupo 4.

Estos grupos estarían compuestos similarmente a lo hallado empleando la distancia de Gower y los agrupamientos de UPGMA y Complete. Cabe resaltar, que el HCPC-PCA se realizó utilizando los caracteres cuantitativos, mientras que la distancia de Gower ocupa tanto los cuantitativos como los cualitativos.

En cuanto a presencia de duplicados, tal como se puede apreciar en el anexo 10.4, el HCPC-PCA encontró 33 casos de duplicados, no solo dentro de las accesiones pertenecientes a una misma raza, sino entre razas. Así, por ejemplo, se tienen tres casos entre las razas paro y piscorunto y dos entre chullpi y piscorunto. Además, en el caso particular de la raza piscorunto, en tres casos se mostró triplicados.

Tabla 10 : Tabla del número de accesiones por raza en cada grupo formado mediante el HCPC con las componentes del PCA

Raza	G1	G2	G3
Chullpi	39	0	1
Paro	0	31	9
Piscorunto	0	7	29

Luego está el HCPC generado con las 13 primeras CP del MCA, cuyo dendrograma cuenta con tres grupos a un punto de corte de 0.2 (figura 48), el cual es bastante bajo e indicaría una similitud alta entre todas las accesiones en conjunto. El grupo 1 cuenta con 39 accesiones de la raza chullpi, el grupo 2 con 31 de paro y 7 de piscorunto, y el grupo 3 con una de chullpi, 9 de paro y 29 de piscorunto (tabla 10).

Al respecto de los duplicados, este enfoque HCPC-MCA presentó 16 casos de duplicados, encontrándose para algunos casos hasta 13 accesiones idénticas dentro de una raza, esto ocurrió tanto para chullpi como para piscorunto (anexo 10.4). Lo cual es un indicador que, al emplear solamente los caracteres cualitativos, no se podrían diferenciar las accesiones dentro de una misma raza con precisión. Adicionalmente, también se encontraron duplicados entre razas, esto ocurrió para las razas paro y piscorunto, así como también para chullpi con paro.

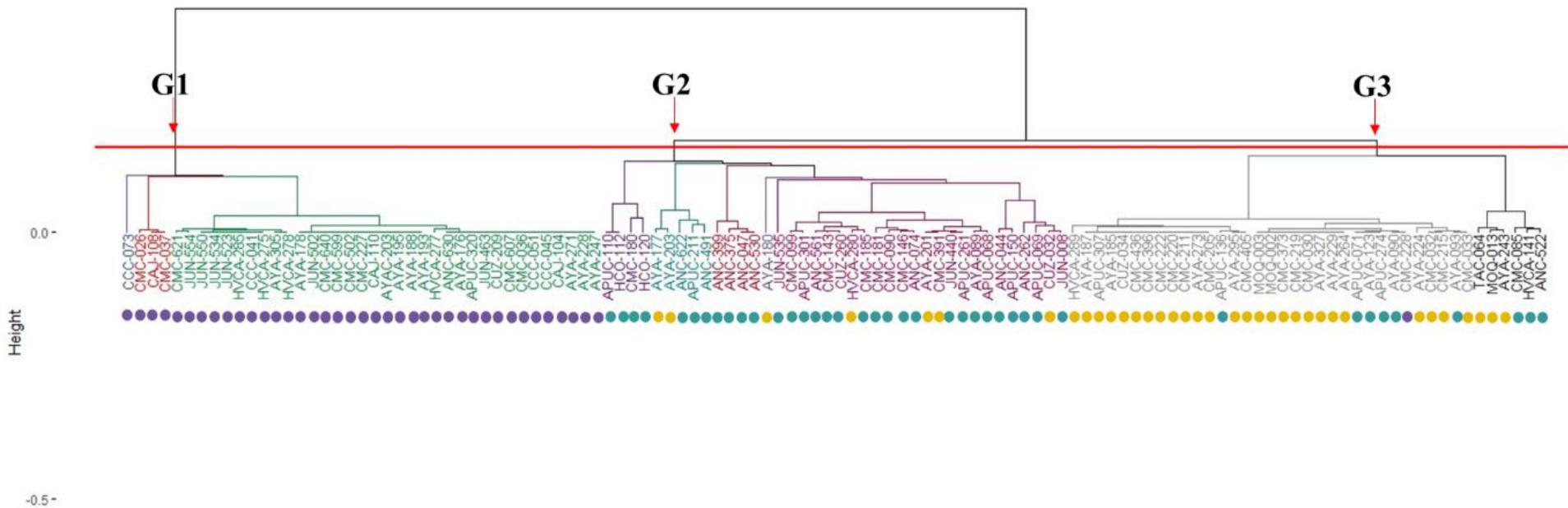


Figura 48: Gráfico de HCPC de accesiones de maíz amiláceo empleando las primeras 13 componentes principales del MCA

*Accesiones con un círculo de color morado pertenecen a la raza Chullpi, de color verde agua a la raza Paro y de color amarillo a la raza Piscorunto.

G1: grupo 1, G2: grupo 2 y G3: grupo 3.

Tabla 11: Tabla del número de accesiones por raza en cada grupo formado mediante el HCPC con las componentes del FAMD

Raza	G1	G2	G3
Chullpi	39	0	1
Paro	0	2	38
Piscorunto	0	34	2

Para el caso del HCPC obtenido con las 14 primeras CP del FAMD, se identificaron tres grupos a un punto de corte de 2.5 (figura 49). El grupo 1 tiene 39 accesiones de chullpi, el grupo 2 con 34 de piscorunto y dos de paro, y el grupo 3 con 38 de paro, una de chullpi y dos de piscorunto (tabla 11).

En la tabla 11, se puede observar para cada uno de los grupos formados es mínimo el número de accesiones (de una a tres) que no corresponden a la raza que alberga en mayoría en ese grupo, lo cual permite deducir que este enfoque de FAMD, sería el más adecuado para agrupar de acuerdo al criterio de las razas en esta población.

En cuanto a los duplicados, se evidenciaron hasta 11 casos, de los cuales en la raza chullpi se pudieron encontrar hasta tres accesiones idénticas (anexo 10.4). Aunado a esto, tanto la raza paro y piscorunto también presentaron casos de duplicación. Sin embargo, a diferencia de lo ocurrido para el HCPC generado con las CP del PCA y MCA, en este tipo de agrupamiento no se encontraron duplicados entre razas.

Tabla 12: Tabla del número de accesiones por raza en cada grupo formado mediante el HCPC con las componentes del MFA

Razas	G1	G2
Chullpi	39	1
Paro	0	40
Piscorunto	0	36

Finalmente, el MFA produjo un HCPC que a un punto de corte de 1.0 presentó dos grupos (figura 50). El primero contenía 39 accesiones de chullpi y el segundo con solo una de chullpi, 40 de paro y 36 de piscorunto (tabla 12).

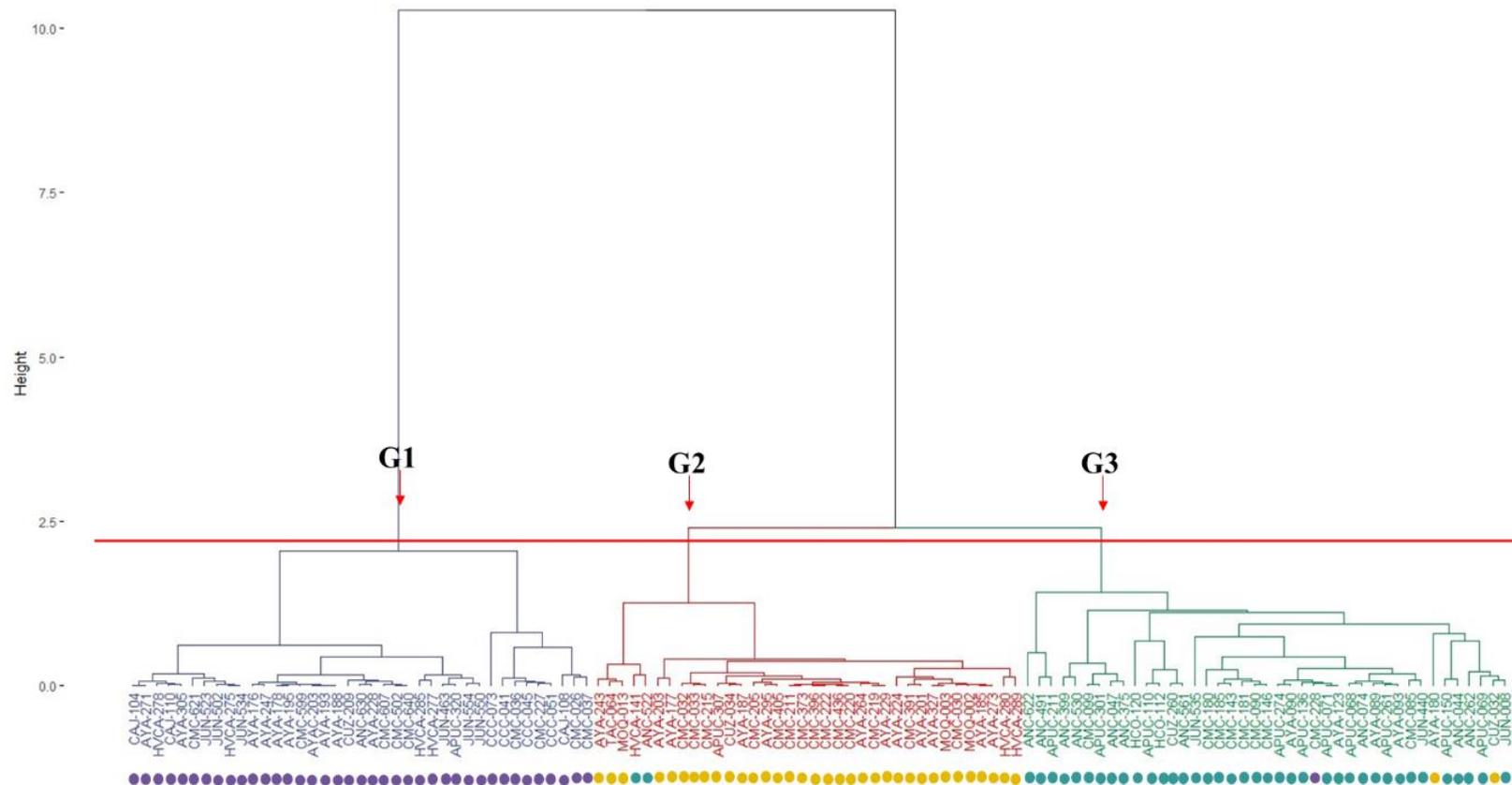


Figura 49: Gráfico de HCPC de accesiones de maíz amiláceo empleando las primeras 14 componentes principales del FAMD

*Accesiones con un círculo de color morado pertenecen a la raza Chullpi, de color verde agua a la raza Paro y de color amarillo a la raza Piscorunto.

G1: grupo 1, G2: grupo 2 y G3: grupo 3.

Dentro de grupo 2, no se pueden diferenciar grupos entre las razas, a pesar de que se consideren grupos más pequeños, ya que los grupos que se forman a partir de este contienen accesiones tanto de paro como de piscorunto. No obstante, de acuerdo con el análisis de duplicados, no se evidenció ningún duplicado en ninguna de las accesiones pertenecientes a una raza, ni tampoco entre razas. Por el contrario, a la suma de casos duplicados que sí se identificaron en los anteriores enfoques del HCPC (anexo 10.4).

En resumen, para este segundo enfoque de agrupamiento jerárquico, se puede decir que tomando en cuenta la reducción del número de caracteres a un número menor de CP; así como, la correlación de estos a cada CP. Además de la capacidad de discriminar a las accesiones de acuerdo con las razas y la posible identificación de duplicados entre las accesiones. Se tiene que, el enfoque HCPC con las CP del FAMD, sería el método más adecuado de emplear para el conjunto de datos del presente estudio.

Con respecto a los dos enfoques empleados para el análisis de agrupamiento jerárquico en general, tanto el uso de las distancias de Gower y el agrupamiento Ward, así como el HCPC con el método de CP del FAMD, hacen uso de ambos tipos de caracteres, por lo que un análisis que utiliza los dos tipos de variables sigue siendo una mejor opción para la evaluación de la variabilidad y la diversidad genética, esto concuerda con lo mencionado por Chávez (2018).

En términos de los agrupamientos formados, ambos enfoques entregan conformaciones similares, esto también es mencionado por Franco e Hidalgo (2003). En el caso de este estudio, ambos enfoques permiten diferenciar a las accesiones de chullpi de las de paro y piscorunto, pero en cuanto a estas dos últimas razas, las similitudes son más visibles que las diferencias, por ello es más complejo discriminarlas con ambos enfoques.

Por otro lado, en el análisis de agrupamiento ya sea usando los caracteres originales o los CP, se observa que hay algunas accesiones pertenecientes a una misma raza que ocupan grupos distintos. Una explicación podría ser que hay razas que son parecidas fenotípicamente, pero tienen origen distinto.

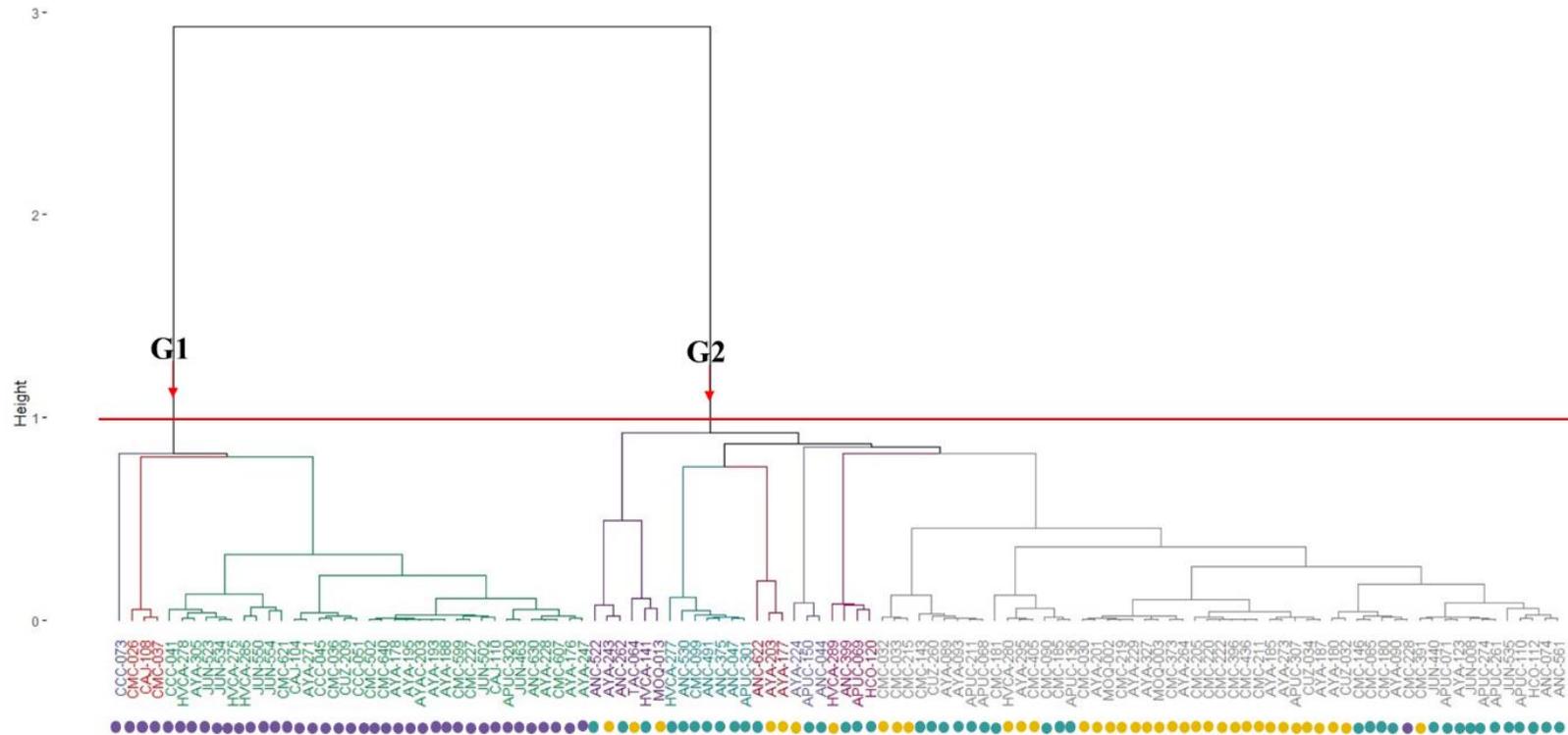


Figura 50: Gráfico de HCPC de accesiones de maíz amiláceo empleando las primeras 13 componentes principales del MFA

*Accesiones con un círculo de color morado pertenecen a la raza Chullpi, de color verde agua a la raza Paro y de color amarillo a la raza Piscorunto.

G1: grupo 1 y G2: grupo 2.

Además, cabe resaltar que la clasificación racial en el Perú además de criterios morfológicos, se emplean también criterios ecológicos y culturales; además de datos de grano y mazorca que fueron decisivos para la clasificación racial. En ese mismo contexto, los análisis de agrupamiento y la construcción de dendrogramas no son decisivos para la clasificación racial, pero pueden ser parte de un método más holístico que considere otros criterios además de los morfológicos (Chavarry 2014).

Asimismo, la falta de cohesión racial de algunas accesiones puede explicarse por la dinámica de producción e intercambio de razas locales entre los agricultores. La mayoría de los productores en América Latina han adoptado prácticas de cultivo que permiten el cruce racial no intencional, ya que cultivan diferentes formas o variedades de maíz en áreas pequeñas cercanas entre sí, lo que facilita el intercambio de polen y flujo de genes entre las razas (Rivas *et al.* 2022).

Adicionalmente, es importante resaltar que la variabilidad morfológica observada en las accesiones estudiadas puede ser atribuida a la variación en la constitución genética de las accesiones, pero también a las reacciones debido a las condiciones ambientales por las que fueron sometidas durante el cultivo en el campo experimental (N'DA *et al.* 2022).

Aunado a esto, la mayor parte del maíz que tienen los agricultores son por lo general cruzamientos, mezclas o similares; por tanto, las muestras de maíz colectadas en los campos de los agricultores y accesadas al banco de germoplasma, van a presentar estas mezclas o cruces generando que algunas razas se agrupen o separen casi al azar cuando se realiza un análisis de conglomerados a estas colectas (Macuri 2016).

5.2 ANÁLISIS DE LA CARACTERIZACIÓN MOLECULAR Y ESTRUCTURA GENÉTICA

Como resultado del *SNP calling* (llamado de variantes) se identificaron 30873 SNPs. Estos marcadores moleculares se encontraron distribuidos en los diez cromosomas de maíz (figura 51), y en términos de porcentaje, aproximadamente el 65% de ellos se encontraron entre el primer y el quinto cromosoma, disminuyendo la densidad a medida que iban distribuyendo hasta llegar al cromosoma diez, resultados similares fueron obtenidos por Leng *et al.* (2019), pero con un menor número de SNP identificados.

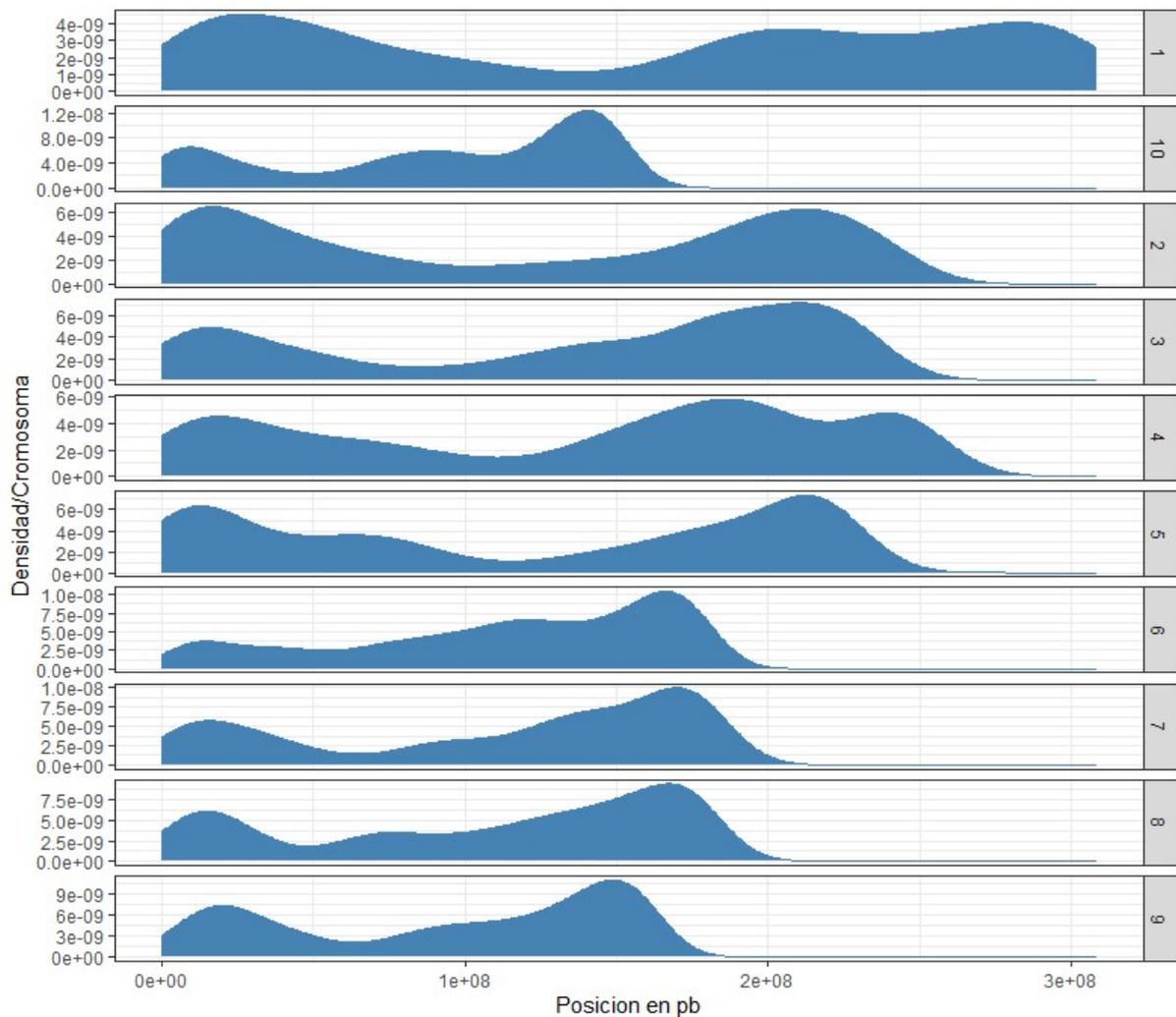


Figura 51: Densidad de los marcadores SNP en cromosomas de maíz

El número total del SNPs, es decir los 30873 marcadores, paso por tres diferentes procesos del filtrado. Se inicio con la eliminación de aquellos marcadores que tenían hasta un 5% de datos faltantes, tal como se muestra en la figura 52, con lo cual se obtuvo como resultado el 53.23% de los marcadores que había en un inicio, es decir se mantuvieron 16437 SNP para el siguiente paso de filtrado.

De manera similar a lo presentado en este estudio, Zhang *et al.* (2016), Wu *et al.* (2014), Ribeiro *et al.* (2013), Yang *et al.* (2011) emplearon un umbral no mayor al 20% de datos faltantes para este primer paso de filtrado. Asimismo, Schaefer *et al.* (2013) empleo los SNPs hasta un 10% de datos faltantes. Los estudios mencionados fueron desarrollados en líneas endogámicas de maíz originarias de China, Brasil y Estados Unidos.

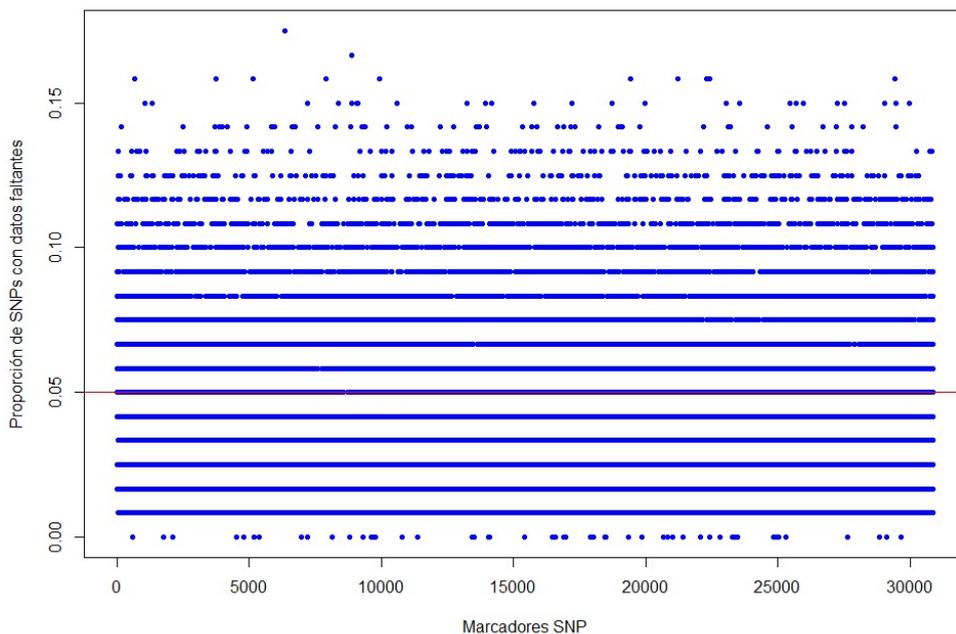


Figura 52: Proporción de marcadores SNPs con datos faltantes

En el segundo paso de filtrado, se eliminaron aquellas accesiones que tenían hasta un 20% de datos faltantes. Esto ocurrió con un genotipo en el cual no amplificó ningún marcador por lo tanto fue eliminado, dejando 119 accesiones por mantener para los posteriores análisis (figura 53). Este mismo criterio de filtrado fue realizado por Van Heerwaarden *et al.* (2010) en variedades locales originarias de México.

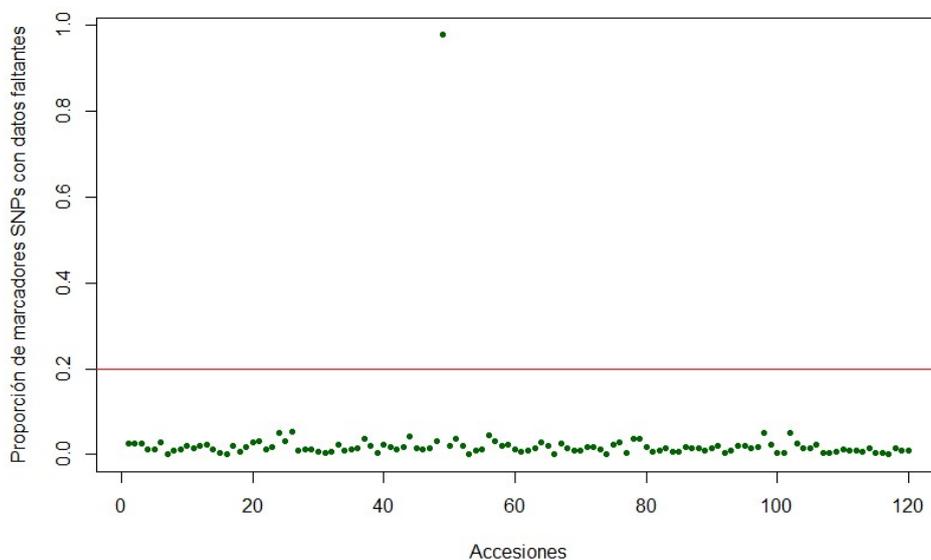


Figura 53: Proporción de marcadores SNPs con datos faltantes en las accesiones

Como tercer paso de filtrado, se realizó una imputación de los datos en cada SNP tomando en cuenta la media de las 119 accesiones remanentes del segundo paso de filtrado. Este paso de imputación se realizó ya que, existe una gran base bibliográfica que sugiere que no imputar los datos faltantes puede tener consecuencias graves para la validez estadística y que afecta la estimabilidad de los parámetros estadísticos que se pretende generalizar a la población de inferencia más amplia (Sun y Kardia 2008).

Finalmente, en el cuarto paso de filtración, los marcadores que tenían alelos con una frecuencia menor al 10% (lo menos comunes) fueron eliminados, con lo cual se obtuvo una reducción en el número de SNPs retenidos, dejando el conteo en 15 486 SNPs en 119 accesiones (figura 54). Este porcentaje que se usó para suprimir los alelos menos comunes también fue empleado por Ribeiro *et al.* (2013) y Yang *et al.* (2011).

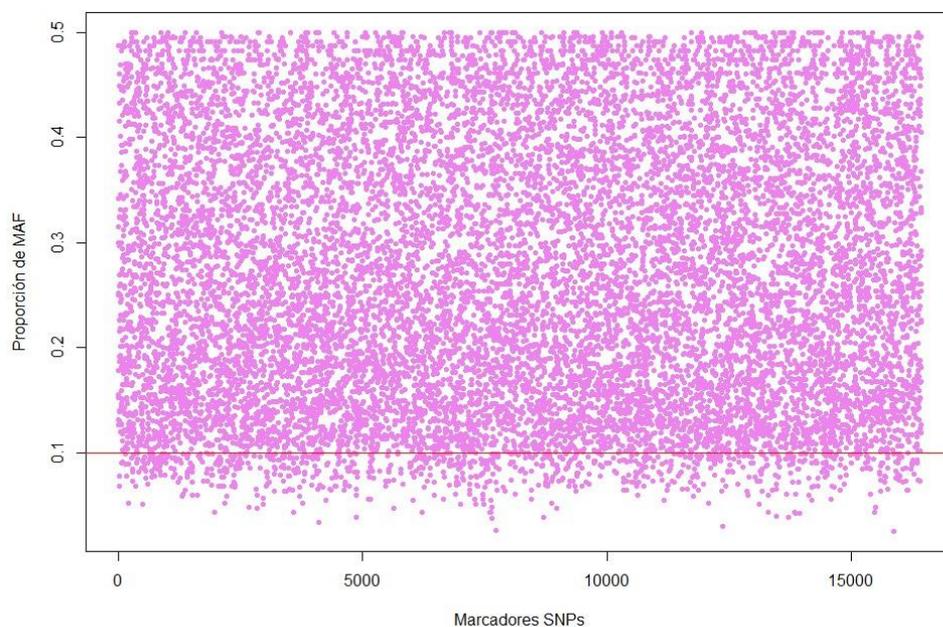


Figura 54: Proporción del MAF en los marcadores SNP

En general el proceso de filtrado permitió obtener un número considerable de SNP. Sin embargo, este es relativo en cuanto a la calidad del análisis en base a estos que se pueda obtener. Acerca de esto último, Nazareno y Knowles (2021) indican que las estimaciones del flujo de genes son sólidas en diferentes rangos de MAF, que fluctúen entre 0,05 –0,35 y con un porcentaje de datos faltantes entre 0 y 20%.

Con respecto a esto, los umbrales empleados en este estudio se encuentran dentro de los rangos denotados por los autores, los cuales también mencionan que no existe una “regla general” para el filtrado bioinformático y que no es aconsejable confiar en la configuración predeterminada del programa estadístico por emplear. En cambio, ellos sugieren que el enfoque a implementar se aplique de forma independiente en cada estudio empírico separado que permita confirmar las configuraciones apropiadas para obtener estimaciones genéticas de poblaciones imparciales, lo cual se realizó en la presente tesis.

En cuanto a la diversidad genética, se calculó la diversidad de Nei (GD – también denominada como heterocigosidad esperada), la heterocigosidad observada (Ho), el índice del contenido de información (PIC) y la frecuencia del alelo menos frecuente para cada raza y en toda la población.

Con respecto a la GD, se puede decir que es la proporción esperada de genotipos heterocigotos bajo el equilibrio de Hardy-Weinberg (Nei 1973), además también se menciona que es la probabilidad de que un par de alelos seleccionados aleatoriamente de una población sea diferentes Luo *et al.* (2019).

Como resultado se obtuvo para las tres razas en conjunto, una GD media de 0.38, con un rango de 0.5 hasta 0.18. Resultados similares fueron obtenidos por Van Inghelandt *et al.* (2010), Schaefer *et al.* (2013), Zhang *et al.* (2016) y Wu *et al.* (2016), con 0.32, 0.393, 0.362 y 0.31 respectivamente, siendo todos hallados en accesiones de maíz que formaban parte de líneas endogámicas.

De manera similar, para el Ho se obtuvo una media de 0.31 con un rango de 0.5 hasta 0.18, esto nos indica la proximidad entre el GD (He) y la Ho. Por otro lado, se tiene que estudios como los de Schaefer *et al.* (2013), Zhang *et al.* (2016) y Wu *et al.* (2016) presentan valores de Ho muy bajos comparados con los de este presente estudio, así se tiene, 0.054, 0.02 y 0.00, respectivamente. Sin embargo, esto se puede explicar debido a que en estos trabajos se evaluaron líneas endogámicas.

Tabla 13: Tabla de valores de la Diversidad genética de Nei, Heterocigosidad observada, Índice del contenido polimórfico y Frecuencia del alelo menos frecuente de cada raza y de la población

Razas	GD			Ho			PIC			MAF		
	Media	Rango inferior	Rango superior									
Chullpi	0.37	0.03	0.5	0.3	0.18	0.5	0.3	0.02	0.38	0.28	0.1	0.5
Paro	0.37	0.05	0.5	0.31	0.26	0.39	0.3	0.05	0.38	0.28	0.1	0.5
Piscorunto	0.37	0	0.5	0.31	0.2	0.41	0.29	0	0.38	0.27	0.1	0.5
Total	0.38	0.18	0.5	0.31	0.18	0.5	0.3	0.16	0.38	0.28	0.1	0.5

*GD es la heterocigosidad esperada, Ho es la heterocigosidad observada, PIC es el índice de contenido polimórfico y el MAF es la frecuencia del alelo menos frecuente.

Luo *et al.* (2019) mencionan que los valores de GD y Ho son medidas de diversidad genética entre genotipos en poblaciones en fase de mejoramiento, por lo cual se suele hallar que el valor general de GD es ligeramente mayor que el valor de PIC lo cual también fue hallado en el presente estudio (tabla 13), en el que el valor medio del PIC fue de 0.30 con un rango de 0.38 hasta 0.16. Relacionado con estos valores de PIC, se tiene a Zawadi *et al.* (2021) con 0.359, Badu *et al.* (2021) con 0.20 y Schaefer *et al.* (2013) con 0.32, valores muy cercanos a los hallados las las tres razas de maíz.

Adicionalmente, se sabe por un estudio previo de Botstein *et al.* (1980), que los marcadores con un valor PIC hasta 0,5 se consideraban altamente informativos, con un valor de 0,25 a 0,5 moderadamente informativos; y con un valor inferior a 0,25 ligeramente informativos. Nuestros resultados mostraron valores de PIC desde ligeramente hasta moderadamente informativos.

Anderson *et al.* (2010) indicaron que no eliminar aquellos alelos de frecuencia menos común da como resultado el uso de información falsa, por dos razones, esto principalmente por dos razones, la primera menciona que las asociaciones observadas en estos SNP son pequeñas porque están dirigidas por los genotipos de pocos individuos; y la segunda, dice que pueden provenir de errores del proceso de genotipado en marcadores que en realidad son monomórficos en la población.

De esto se desprende, que como resultados se obtuvo una media en el valor de MAF de 0.28 con un rango inferior de 0.1 hasta el máximo posible que es 0.5. Estos valores son mayores a los hallado por Badu *et al.* (2021) con una media de 0.173 y un rango de 0.015-0.5; y Boakyewaa Adu *et al.* (2019) con una media de 0.16 y un rango de 0.01-0.5, esto se puede deber a que las muestras empleadas en estas investigaciones son líneas endogámicas y las presentadas en este trabajo son variedades locales y razas.

Para el análisis de la estructura genética, se inició con procedimiento realizado a través del programa STRUCTURE (STR) que emplea un enfoque para el análisis de agrupamiento en base a un modelo bayesiano (Mohammadi y Prasanna 2003). Liu *et al* (2003) reportaron que, para líneas endogámicas de maíz, el enfoque por modelo Bayesiano permitió obtener grupos consistentes con el pedigrí de cada una de las líneas, por lo cual se empleó en primera instancia ese mismo enfoque.

A partir de esto, tomando como criterio la tasa de cambio de la probabilidad logarítmica de los datos entre k sucesivos (Evanno *et al.* 2005) se halló que eran tres los grupos óptimos (k=3) (figura 55) que se formaban a partir de las 119 accesiones en evaluación, para ello se probó el número de grupos óptimos desde K=1 hasta k=5, para ello observar la figura 56.

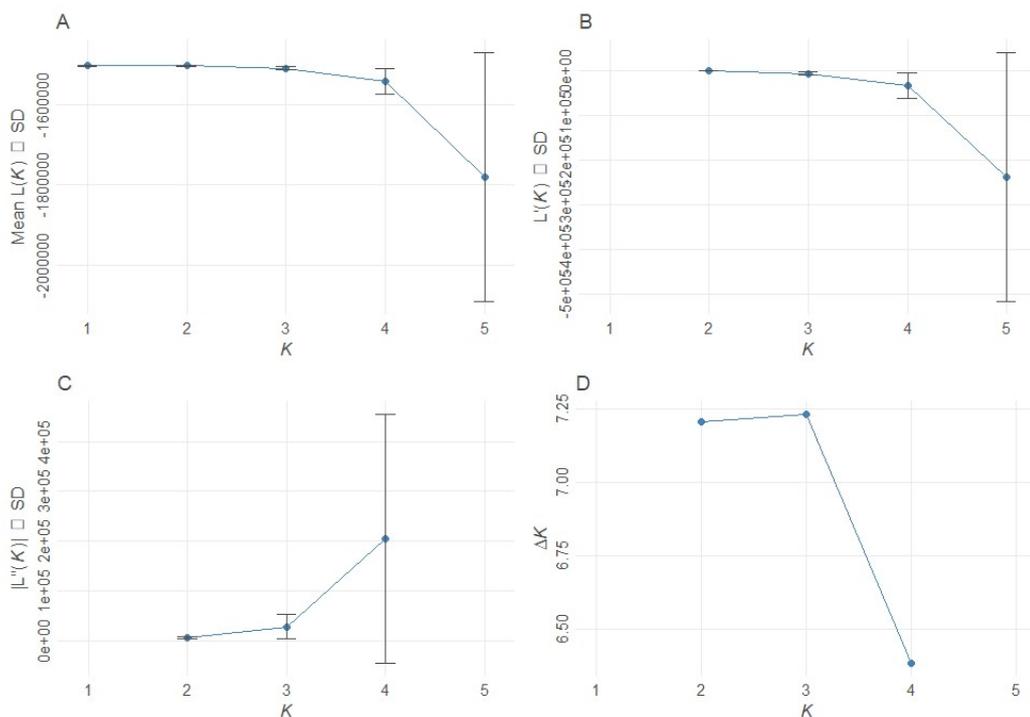


Figura 55: Gráficos del método de Evanno para K=1 hasta K=5

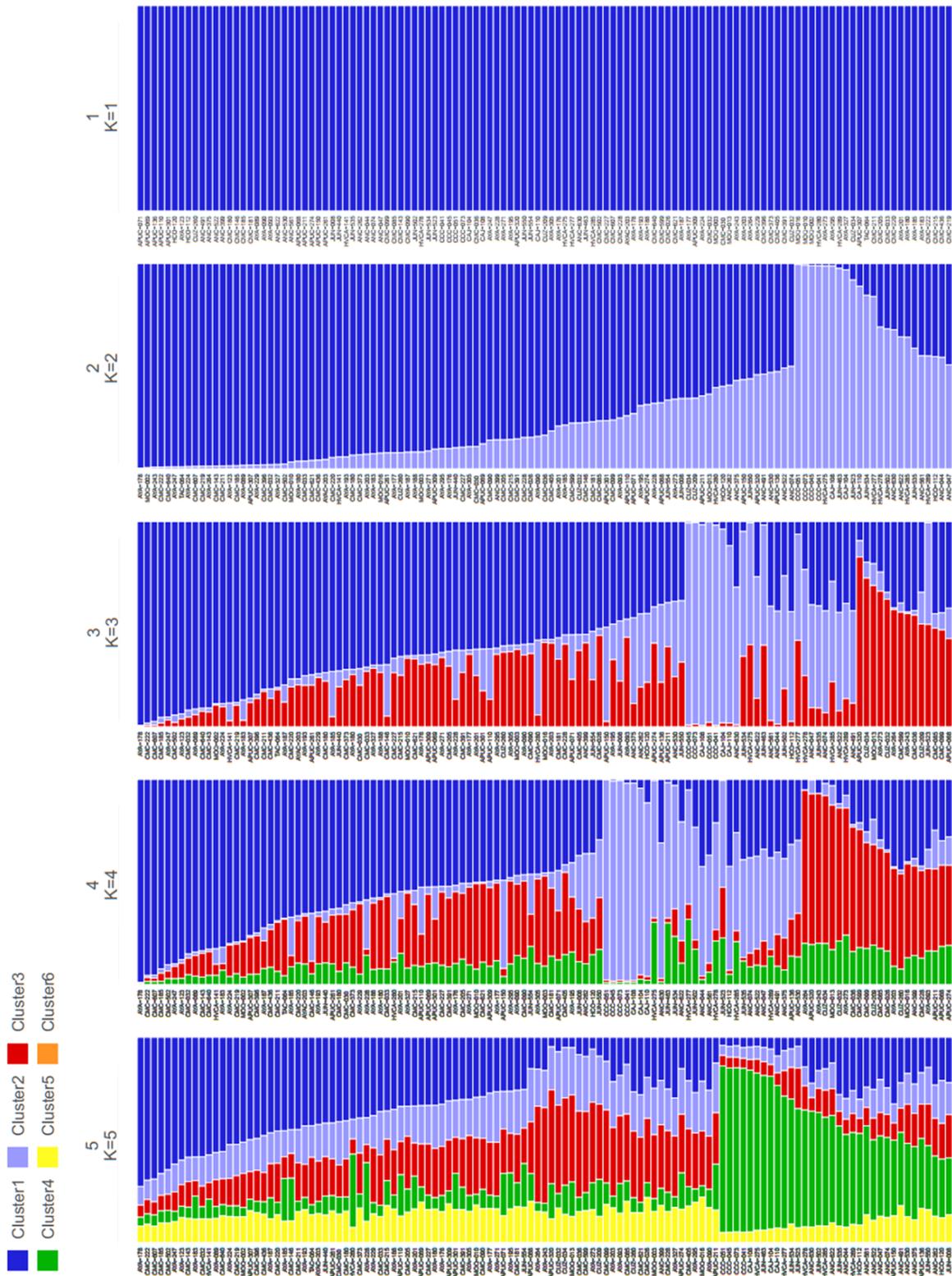


Figura 56: Diagramas de barras multilinea a partir de probabilidades de membresía a cada grupo (cluster) con etiquetado por accesión de k=1 a k=5

En la figura 57, se puede observar los grupos determinados por STR con el número óptimo de grupos de $k=3$, a los cuales se les asignó la etiqueta SG1, para el grupo 1, SG2 para el grupo 2, SG3 para el grupo 3 y M para el grupo que contenía las accesiones que debido a que no llegaban al valor de 0.5 en la probabilidad de membresía (PM) a ningún grupo se les agrupó en un grupo mixto.

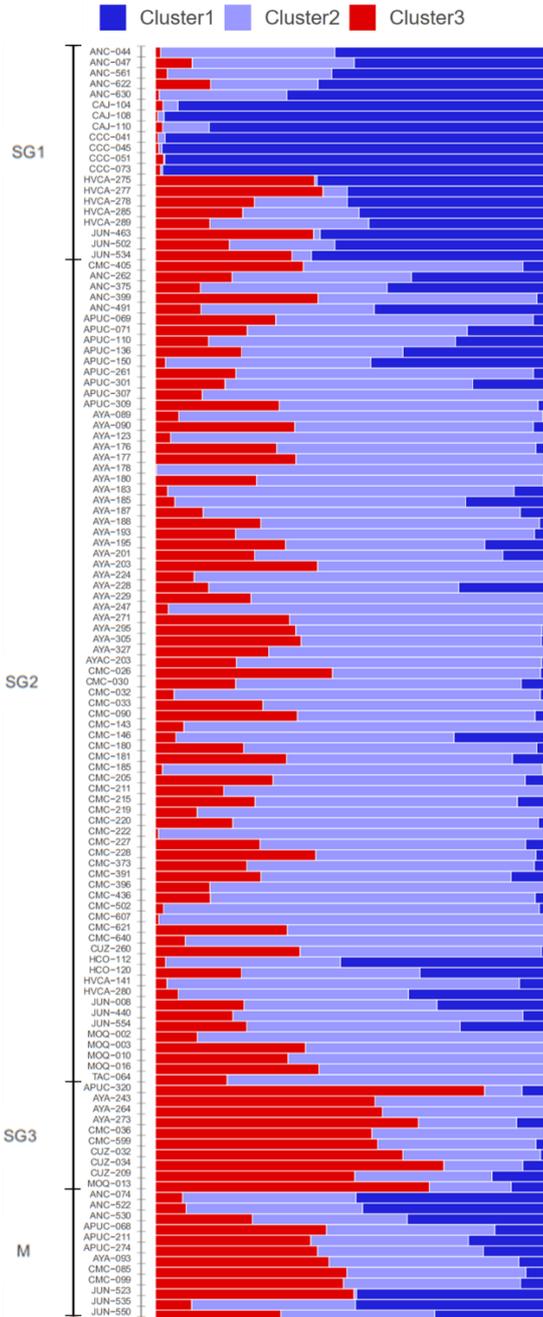


Figura 57: Diagrama de barras de una sola línea ($k=3$) a partir de probabilidades de membresía a cada grupo (cluster) con etiquetado por accesión y grupos determinados por STRUCTURE a un $PM \geq 0.5$

Con la finalidad de corroborar si es que las accesiones que integran los grupos determinados por STR eran las mismas accesiones pertenecientes a las tres razas, se realizaron diagramas de barras multilínea a partir de probabilidades de membresía a cada grupo (*cluster*), las accesiones se organizaron en este diagrama de acuerdo a la raza a la que correspondía. En la figura 58, se puede apreciar que no hay una consistencia entre los grupos determinados por STR y la raza la que corresponde, ya que varias accesiones que pertenecen a un mismo grupo de STR no se encuentran agrupadas en una misma raza.



Figura 58: Diagrama de barras de una sola línea (k=3) a partir de probabilidades de membresía a cada grupo (cluster) con etiquetado por accesión y grupos determinados por la raza de maíz

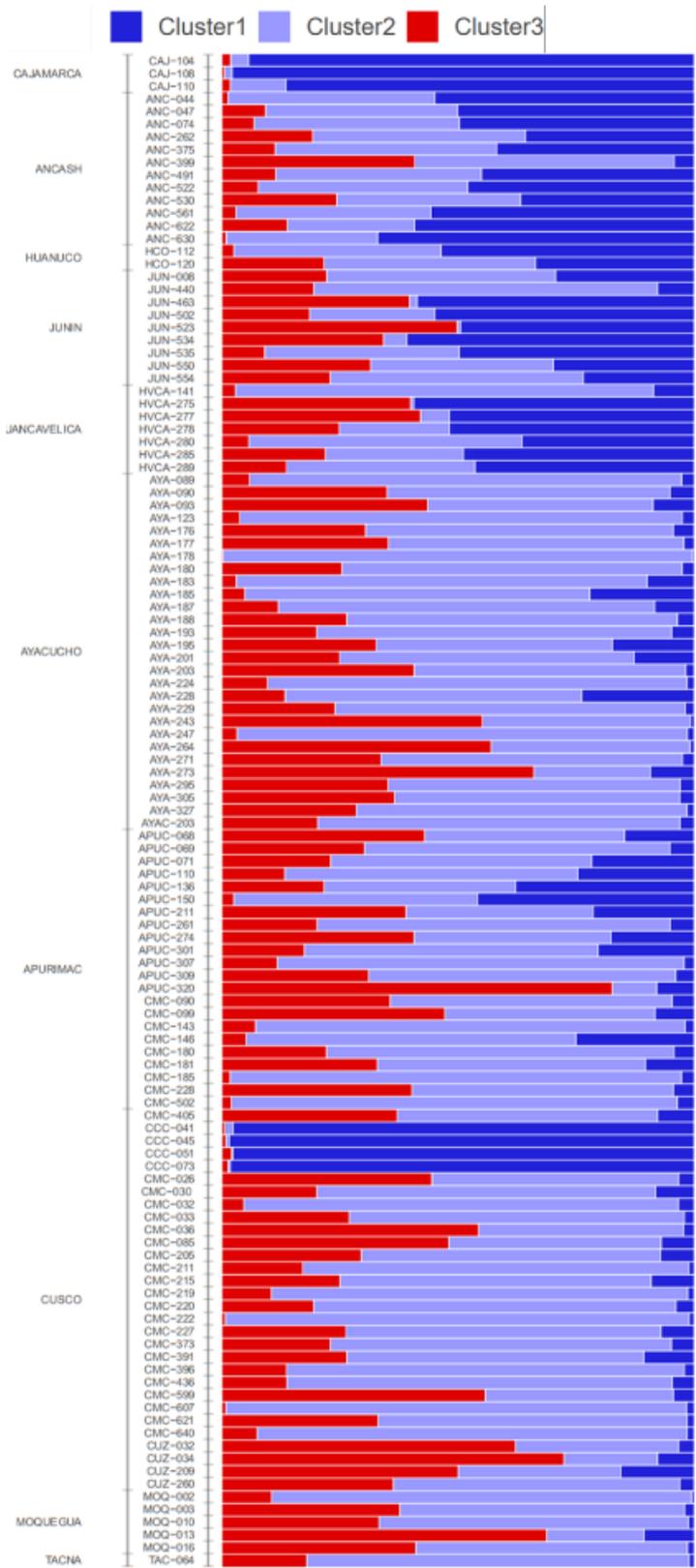


Figura 59: Diagrama de barras de una sola línea (k=3) a partir de probabilidades de membresía a cada grupo (cluster) con etiquetado por accesión y grupos determinados por el origen de la accesión

Adicionalmente, y de acuerdo con lo investigado por Pineda e Hidalgo (2013), quienes sospechaban que la distancia geográfica entre las accesiones podría ser otro factor influyente debido a que las muestras suelen ser recolectadas en diferentes localidades, fue que se realizó un diagrama de barras agrupando a las accesiones de acuerdo con el origen de la misma (es decir del lugar en el que fue colectada para luego ingresarla al Banco de germoplasma).

De este diagrama (figura 59), se puede inferir que tampoco existe una consistencia entre las accesiones que componen los grupos de STR con los grupos determinados por el origen, ya que para determinado origen hay varias accesiones que pertenecen a diferentes grupos de STR. Por el contrario, en un estudio realizado por Boakyewaa Adu *et al.* (2019) mencionan que los marcadores SNP asignaron claramente líneas endogámicas en grupos heteróticos en función de las poblaciones de origen.

Continuando con el análisis de la estructura genética, en esta instancia se emplearon métodos basados en distancias, de los cuales se obtuvo como resultado que la distancia euclidiana con el agrupamiento UPGMA, fueron los que combinados obtuvieron un coeficiente de correlación cofenético de 0.800 (tabla 14), que de acuerdo con Darkwa *et al.* 2020, es el valor adecuado para indicar que aquella distancia con el agrupamiento escogidos son los que agrupan mejor a las accesiones.

Tabla 14: Tabla de coeficientes de correlación cofenético de agrupamiento jerárquico basado en distancias a partir de los marcadores SNPs

Agrupamiento/Distancias	Euclidiana	Simple	Absolute	Manhattan
Ward	0.373	0.150	0.145	0.400
UPGMA	0.800	0.553	0.552	0.774

En base a esta combinación de distancia con agrupamiento, se graficaron dendrogramas coloreando a las accesiones de acuerdo con la raza, grupos determinados por STR y por origen de las accesiones. En la figura 60, se observa que no existen grupos que cuenten únicamente con accesiones pertenecientes a una sola raza, en otras palabras, no hay relación entre las distancias genéticas con la pertenecía a una raza en específico.

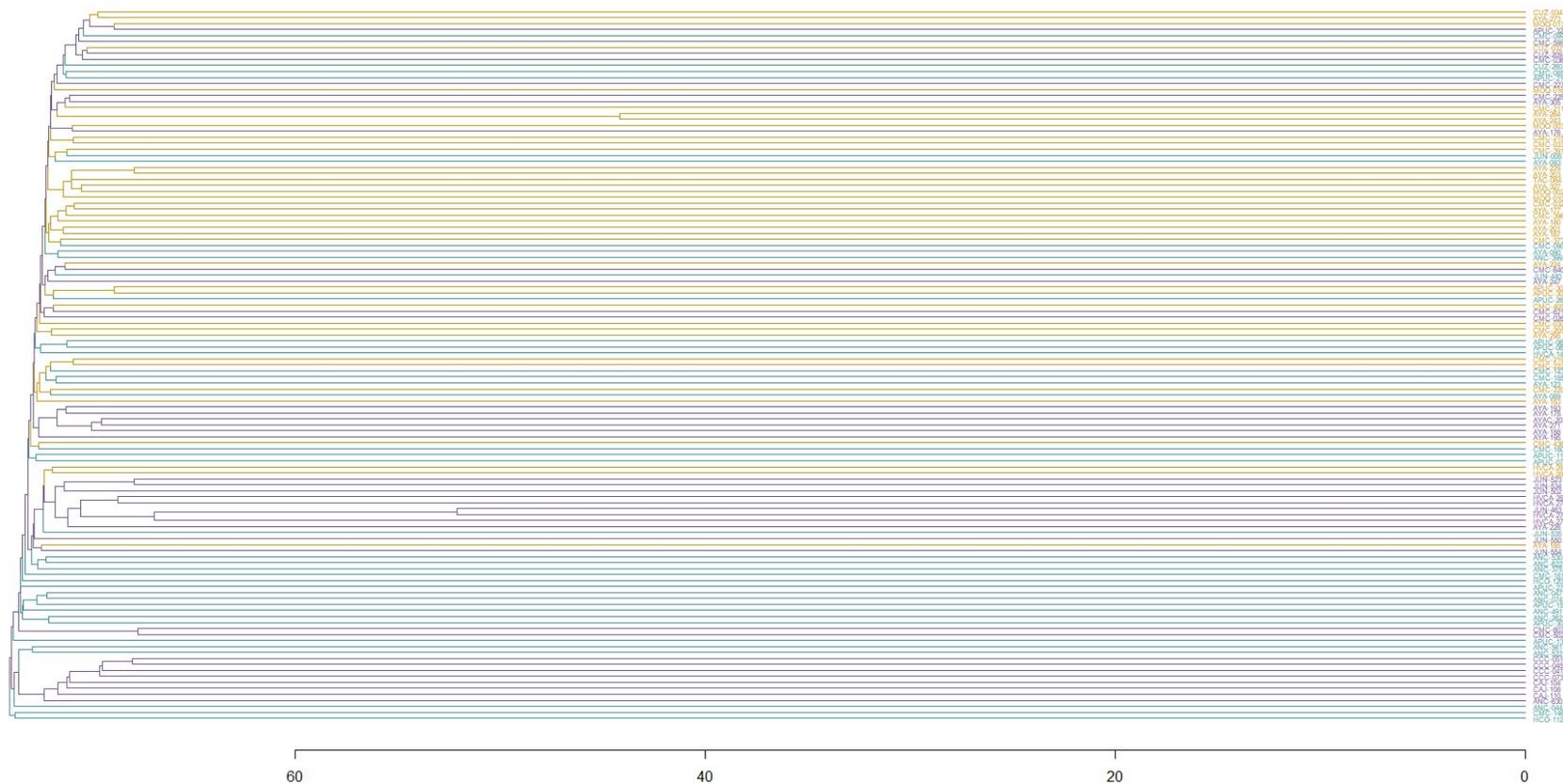


Figura 60: Dendrograma con distancia euclidiana y agrupamiento UPGMA de accesiones etiquetadas por colores en concordancia con la raza de maíz amiláceo ■ Paro, ■ Chullpi y ■ Piscorunto

Vinculado a esto también, se calcularon las distancias de los grupos formados por las accesiones que pertenecían a una misma; así en la figura 61, se muestra que existe una menor distancia genética entre las accesiones de la raza Chullpi con las de Paro, que con respecto a las accesiones de la raza Piscorunto, que sería el grupo más alejado.

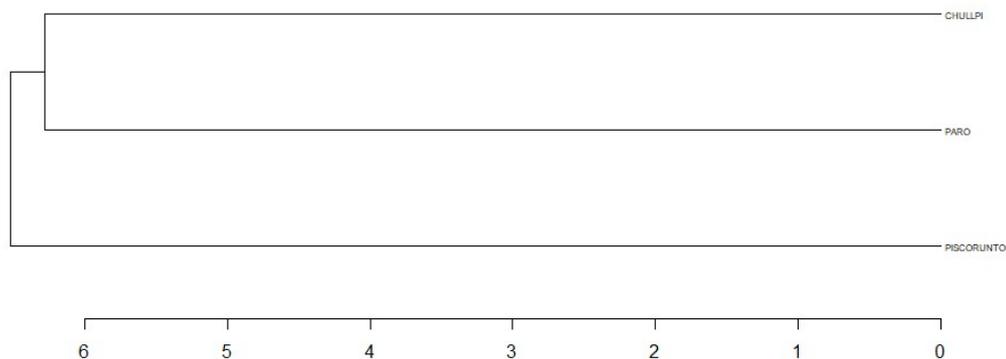


Figura 61: Dendrograma distancia euclidiana y agrupamiento UPGMA de los grupos determinados por la raza de maíz amiláceo

En cuanto a los grupos formados a partir del STR, se realizó un dendrograma coloreando las accesiones de acuerdo con los grupos ya mencionados, que fueron formados mediante el enfoque bayesiano (figura 62). Encontrando que, tampoco existe una consistencia entre las accesiones que pertenecen a los grupos de STR con los grupos que potencialmente se pueden formar a partir del enfoque por distancias.

Resultados similares fueron hallados por Semagn *et al.* (2012), quienes empleando la distancia de Rogers y el agrupamiento de Neighbor-Joining no encontraron concordancia entre el dendrograma de distancias con los resultados obtenidos con STR. Sin embargo, Dao *et al.* (2014) mencionan que diferentes combinaciones de matrices de similitud de distancias genéticas y algoritmos de agrupamiento pueden dar lugar a diferentes grupos. Incluso una única matriz de distancias y un único algoritmo de agrupamiento pueden producir varios grupos alternativos que crean ambigüedad a la hora de seleccionar el mejor.

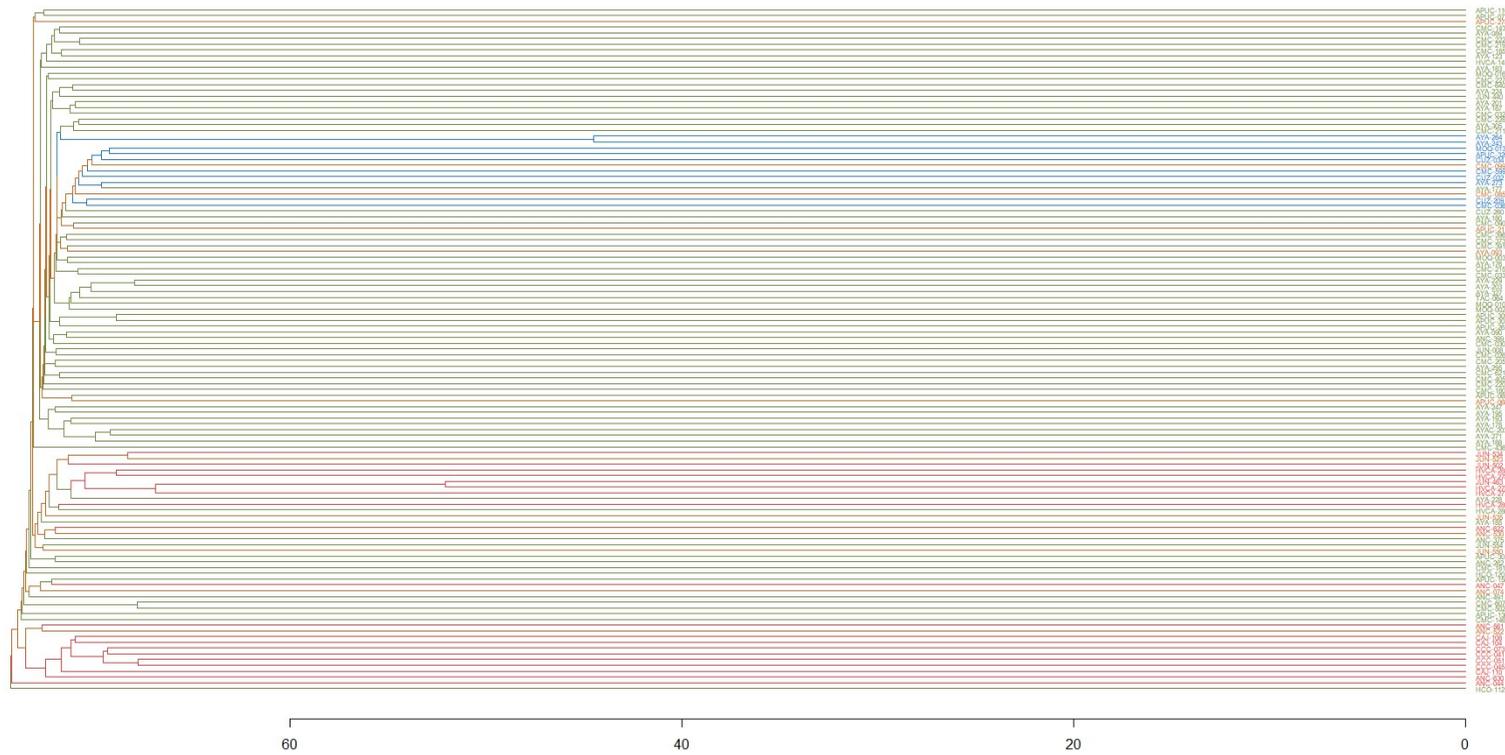


Figura 62: Dendrograma con distancia euclidiana y agrupamiento UPGMA de accesiones etiquetadas por colores en concordancia con los grupos determinados por STRUCTURE

■ SG1, ■ SG2, ■ SG3 y ■ M

De manera semejante a lo realizado para los grupos por razas (figura 63), se calcularon las distancias entre los grupos determinados por STR y se halló que los grupos SG1 y SG2 son más cercanos que el grupo SG3. En cuanto al grupo M, que contenía las accesiones que no fue posible asignarlas a un grupo, se encontró más cercano a los grupos SG1 y SG2.

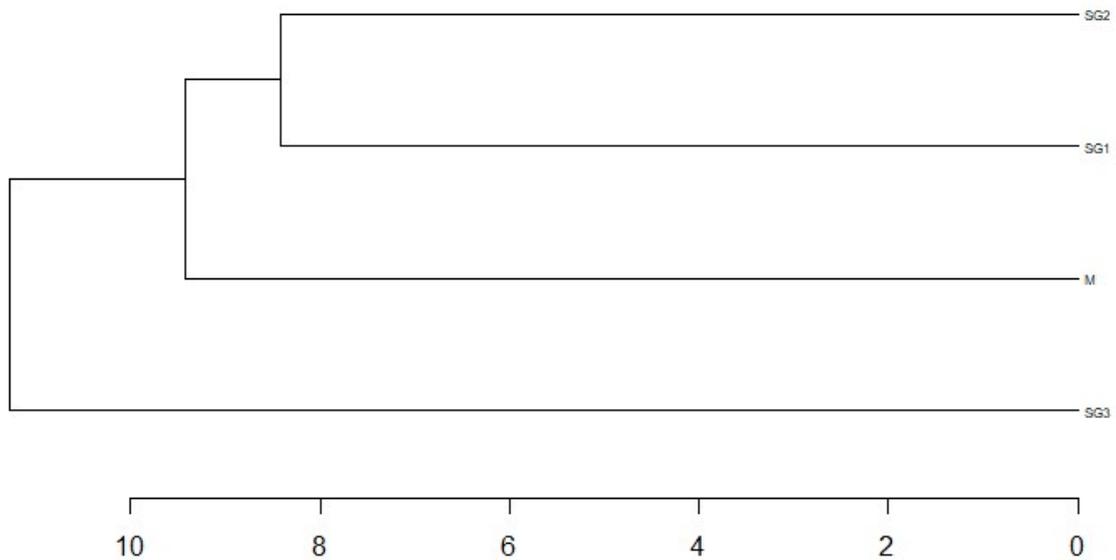


Figura 63: Dendrograma distancia euclidiana y agrupamiento UPGMA de los grupos determinados por el programa STRUCTURE

Por último, se generó un dendrograma coloreando las accesiones de acuerdo con el origen, figura 64. Demostrando también que, de acuerdo con este criterio no se pueden agrupar las accesiones, ya que dentro de los grupos que se muestran en el dendrograma hay accesiones que proceden de diferentes departamentos de la sierra peruana. Pineda e Hidalgo (2013) obtuvieron resultados similares, ya que mencionan que, para sus variedades locales mexicanas, no había ninguna una asociación significativa entre las distancias genéticas por pares y las distancias geográficas por pares, lo que indica que el patrón de diferenciación no se debe al aislamiento por distancia para cada accesión.

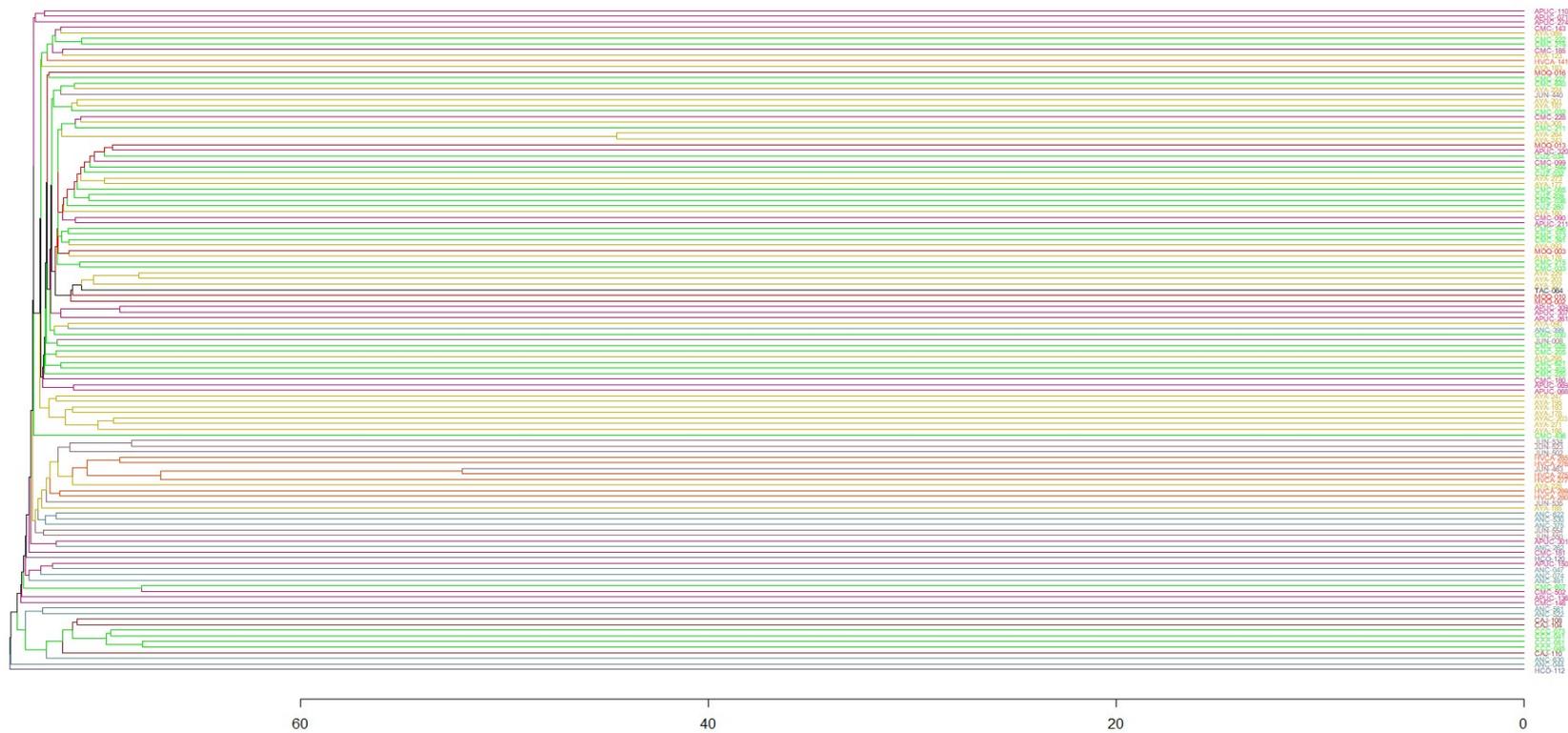


Figura 64: Dendrograma con distancia euclidiana y agrupamiento UPGMA de accesiones etiquetadas por colores en concordancia con los grupos por origen de las accesiones

- Cajamarca, ■ Ancash, ■ Huanúco, ■ Junín, ■ Huancavelica,
- Ayacucho, ■ Apurímac, ■ Cusco, ■ Moquegua, ■ Tacna

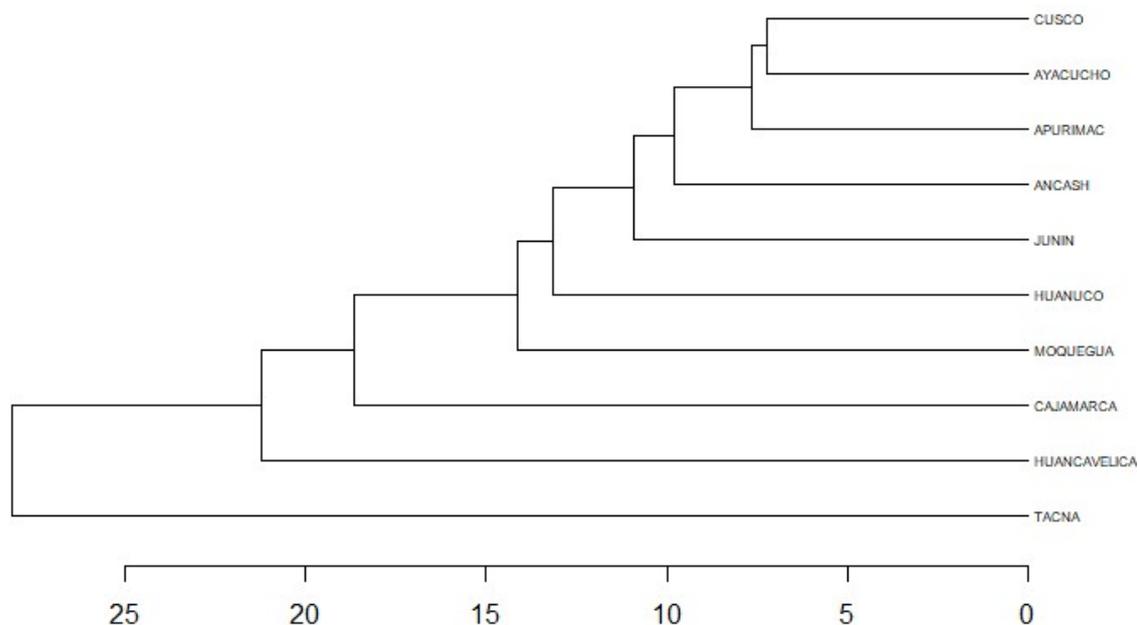


Figura 65: Dendrograma distancia euclidiana y agrupamiento UPGMA de los grupos determinados por el origen de las accesiones

Por otro lado, se agruparon las accesiones por origen para calcular la distancia por grupos. De ello se obtuvo como resultado, que las accesiones provenientes de Cusco, Ayacucho y Apurímac son muy cercanas genéticamente. Coincidentemente, estos departamentos se localizan en la sierra sur del Perú. Luego de ello, se tiene a Ancash, Junín y Huánuco, que se ubican entre el norte y centro de la sierra del Perú. Esto nos da indicios de una asociación entre la distribución geográfica de las muestras y las distancias que existen entre los grupos de accesiones, lo cual es muy diferente a lo hallado para las distancias que se observan entre las accesiones.

Otro enfoque basado en distancias, pero no jerárquico es el Análisis discriminante de componentes principales (DAPC), este método se basa en la transformación de datos de alelos utilizando el análisis de componentes principales como paso previo al análisis discriminante. El DAPC define un modelo en el que la variación genética se divide en un componente entre grupos y dentro del grupo Bracco *et al.* (2021).

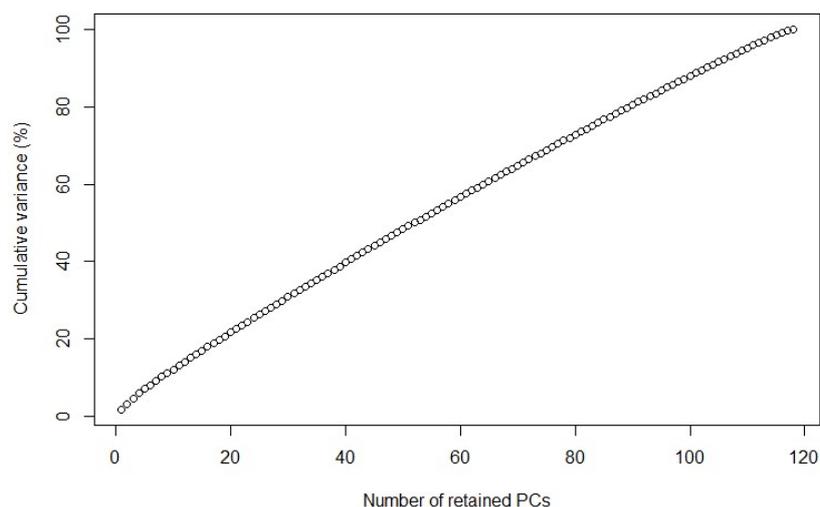


Figura 66: Biplot del número de PC retenidas y la varianza acumulada producto del DAPC

Para la aplicación de este método, como primer paso se hallan las componentes principales, los resultados de ello se observan en la figura 66. Posterior a esto, se procede a elegir el número óptimo de grupos sobre la base del criterio de información bayesiano (BIC) asociado más bajo. Jombart *et al.* (2010) menciona el método del codo para la selección del número óptimo de grupos, sin embargo, tal como se aprecia en la figura 67, para las muestras de este presente estudio, dicho punto de inflexión no se observa en la gráfica, por lo cual este enfoque jerárquico no sería el adecuado para agrupar las accesiones.

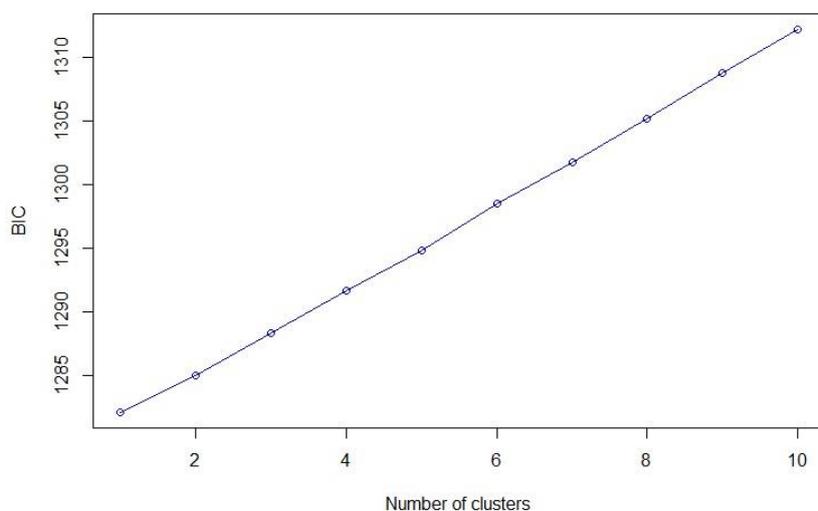


Figura 67: Biplot del número de clusters y el valor del BIC producto del DAPC

Continuando con el enfoque de distancias se realizó un PCA, con el cual se obtuvieron 118 componentes con valores propios mayores a uno, sin embargo, cada componente representa un % de variancia muy pequeño, desde la primera a la CP catorce sumó un 16.74% (anexo 14). Así como también, en las dos primeras CP se acumula 3.4% de la varianza; resultados similares fueron hallados por McLean *et al.* (2021) en variedades locales de maíz procedentes de México.

Es por ello por lo que se debe tomar en cuenta ciertas desventajas que puede presentar el emplear las CP para explicar la distribución de la varianza. Esto mismo es mencionado por Gruber *et al.* 2023, quienes nos indican que, la representación de las entidades en un espacio tan reducido no representará fielmente las distancias en la matriz de distancias de entrada simplemente debido a la pérdida de información en dimensiones informativas más profundas.

A partir de esto se puede inferir, por qué al colorear las accesiones de acuerdo a los grupos formados por las razas, estas no se agrupan en concordancia con este criterio. Esto se puede observar en la figura 68, en la cual las tres razas tienen accesiones muy cercanas unas con otras, sin mostrar una agrupación aparente. Cabe mencionar, que estos resultados concuerdan con los mostrados en el análisis jerárquico.

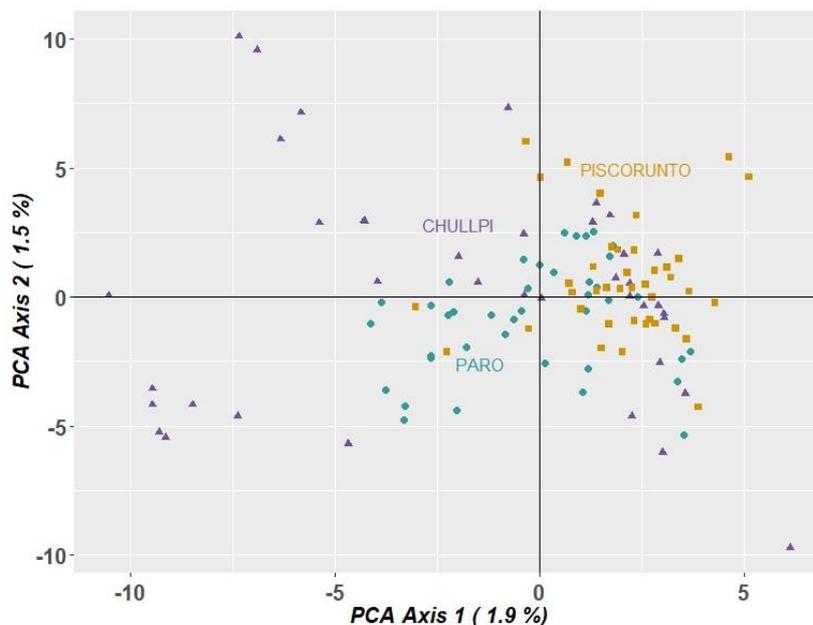


Figura 68: Biplot de PCA de los grupos determinados por las razas de maíz amiláceo

Por el contrario, cuando las accesiones se colorean de acuerdo con los grupos generados por STR, con el PCA se puede observar una agrupación de las accesiones más adecuada, esto se observa en la figura 69. En este gráfico, se puede apreciar la separación en estos tres grupos SG1, SG2 y SG3, e incluso la presencia dispersa de las accesiones que componen el grupo mixto (M).

Existen otros estudios, que ha encontrado esta misma concordancia entre los resultados entregados por el enfoque por modelo bayesiano como lo es el STR y el enfoque por distancias del PCA. Así se tiene tanto a Zhang *et al.* (2016). Dao *et al.* (2014), Semagn *et al.* (2012) y Yang *et al.* (2011) habiendo todos realizado sus investigaciones en líneas endogámicas de maíz procedentes de China y África a través de SNPs.

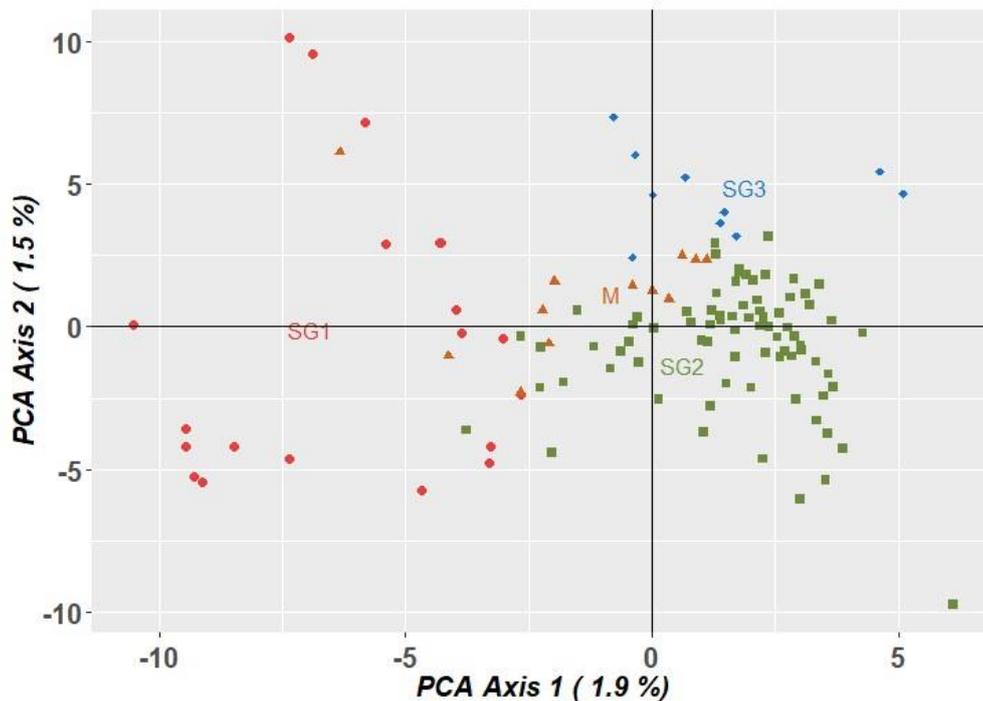


Figura 69: Biplot de PCA de los grupos determinados por STRUCTURE

Finalmente, se colorean las accesiones tomando en cuenta los grupos determinados por el origen de las accesiones. En la figura 70, se observa una aglomeración de los grupos similar a lo hallado con el agrupamiento jerárquico con distancias genéticas. Así los departamentos de la sierra sur del Perú se encuentran muy cercanos (Cusco, Ayacucho, Moquegua, Apurímac y Tacna) y continuando hacia la izquierda se tiene a los departamentos de la sierra central (Junín, Huánuco y Huancavelica) y norte del Perú (Ancash y Cajamarca), esto

refuerza la sospecha de la relación entre la distribución geográfica del origen de las accesiones con su posible agrupación de acuerdo con este criterio.

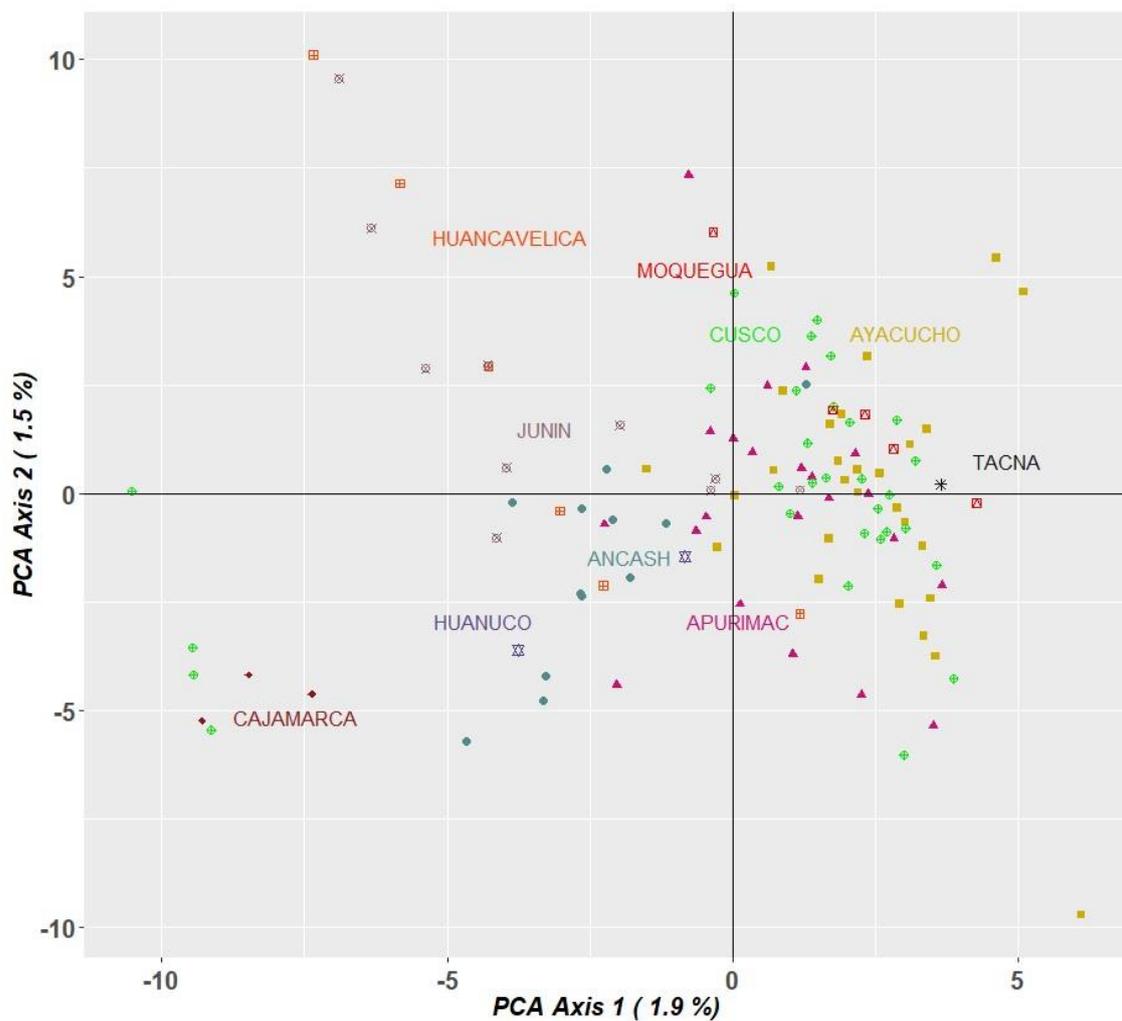


Figura 70: Biplot de PCA de los grupos determinados por el origen de las accesiones

Otro estimador para la estructura genética es el índice de fijación de Wright, el cual es útil como medida general de diferenciación entre poblaciones. Es por esta razón, que se calculó dicho índice para los grupos determinados por las razas de maíz, por el análisis de STR y por el origen de las accesiones. Para cada uno de ellos, se comprobó su validez mediante el uso de Bootstrap e intervalos de confianza para cada índice (anexo 15).

En el caso de los grupos determinados por las razas, los valores de los índices son menores a 0.05, estando en un rango entre 0.0014 y 0.003 (tabla 15), por lo cual se puede categorizar a estos grupos como pequeña diferenciación. Esta clasificación es en base a discutido por Wright (1965), investigación en la que indica que a un valor de $F_{ST}=0$ no hay diferenciación,

FST=1 total diferenciación, FST menor a 0.05 pequeña diferenciación, FST entre 0.05-0.15 diferenciación moderada, FST entre 0.15-0.25 diferenciación fuerte y FST mayor a 0.25 muy fuerte diferenciación.

Tabla 15: Matriz de valores Fst por par de razas de maíz amiláceo

	Paro	Chullpi	Piscorunto
Paro	-	-	-
Chullpi	0.0014	-	-
Piscorunto	0.0017	0.003	-

De manera semejante a lo ocurrido a los grupos por razas, en el caso de los grupos determinados por STR, el valor de los índices permite considerarlos como grupos con pequeña diferenciación, estos se encontraban en un rango de 0.02 a 0.14 (tabla 16). No obstante, estos fueron mayores que los obtenidos con los grupos por raza.

Tabla 16: Matriz de valores Fst por par de grupos determinados con el programa STRUCTURE

	SG1	M	SG2	SG3
SG1	-	-	-	-
M	0.004	-	-	-
SG2	0.008	0.002	-	-
SG3	0.014	0.005	0.006	-

Igualmente, a lo sucedido con los anteriores grupos, en el caso de los determinados por el origen de las accesiones, los valores de los índices Fst varían en un rango de 0 a 0.033, lo que permite inferir que estos grupos presentan una pequeña diferenciación. Cabe resaltar, que los valores más altos se dan entre los grupos que poseen las menores cantidades de muestras procedentes de ese departamento, así como se puede apreciar que ocurre para Huancavelica (seis accesiones), Moquegua (cinco accesiones), Cajamarca (tres accesiones), Huánuco (dos accesiones) y Tacna (una accesión), de esto se puede inferir que este índice se ve afectado por el número de muestras.

Tabla 17: Matriz de valores Fst por par de grupos determinados por el origen de las accesiones

	Ancash	Apurímac	Ayacucho	Cajamarca	Cusco	Huánuco	Huancavelica	Junín	Moquegua	Tacna
Ancash	-	-	-	-	-	-	-	-	-	-
Apurímac	0.003	-	-	-	-	-	-	-	-	-
Ayacucho	0.005	0.002	-	-	-	-	-	-	-	-
Cajamarca	0.009	0.016	0.020	-	-	-	-	-	-	-
Cusco	0.003	0.001	0.002	0.015	-	-	-	-	-	-
Huánuco	0.005	0.008	0.011	0.015	0.008	-	-	-	-	-
Huancavelica	0.008	0.006	0.009	0.021	0.008	0.014	-	-	-	-
Junín	0.004	0.005	0.006	0.016	0.005	0.010	-0.001	-	-	-
Moquegua	0.006	0.002	0.002	0.025	0.002	0.014	0.012	0.008	-	-
Tacna	0.003	-0.001	0.000	0.033	0.001	0.003	0.014	0.008	0.000	-

En general para todos los grupos se obtuvieron valores muy bajos para los índices, todos sin excepción categorizados con pequeña diferenciación. De forma similar, McLean *et al.* (2021) y Zawadi *et al.* (2021) obtuvieron niveles bajos de diferenciación genética, estos estudios fueron desarrollados en variedades locales y líneas endogámicas, respectivamente, y entre ellos los de menor valor fueron los de las líneas endogámicas.

Para complementar la información obtenida con el índice Fst, se realizó un AMOVA (anexo 16), y a partir de método se obtiene los estadísticos ϕ y σ . El estadístico de prueba resultante ϕ es análogo al Fst de Wright y es la relación entre el cuadrado medio entre grupos y el cuadrado medio total. ϕ representa la correlación entre las muestras genéticas aleatorias dentro de un grupo en relación con las muestras aleatorias de la población en general. la sigma representa la varianza, σ , para cada población y el porcentaje de la varianza total explicada por cada fuente de varianza. Además, ϕ proporciona estadísticas de diferenciación de población. Esperaríamos que un valor ϕ más alto representara un mayor nivel de diferenciación (Wright 1951, Cockerham y Weir 1973).

Se realizó el AMOVA para los grupos formados a partir de las razas de maíz amiláceo. Como resultados se obtuvo un 0.2% de la varianza que se localizaba entre las razas Paro, Chullpi y Piscorunto, siendo este ni un uno por ciento permite reforzar la inferencia acerca de que las razas a nivel molecular no permiten establecer grupos distinguibles. En cuanto al resto de la varianza esta se encontraba entre y dentro de las accesiones, mostrado valores negativos entre y dentro las accesiones para el valor de ϕ (tabla 18), esto demuestra la presencia de un

exceso de accesiones de genotipos heterocigóticos (López-Zavala *et al.* 2013), esto concuerda con lo obtenido con los índices de GD y Ho.

Tabla 18: Tabla de resultados del AMOVA de las accesiones agrupadas de acuerdo con las razas de maíz amiláceo

Fuentes de variación	σ	%	ϕ
Varianza entre razas	3.96335	0.2	0.002
Varianza entre las accesiones	-650.885	-32.822	-0.329
Varianza dentro de las accesiones	2630	132.622	-0.326
Varianza total	1983.078	100	

En el caso de los grupos determinados por STR, fueron estos lo que presentaron el mayor porcentaje de diferenciación entre los grupos, esto con 0.617%. Sin embargo, esto continúa siendo un indicativo de muy baja diferenciación entre los grupos. Asimismo, se observó un valor negativo de ϕ entre y dentro las accesiones, lo cual indica un exceso en el número de genotipos heterocigóticos.

Tabla 19: Tabla de resultados del AMOVA de las accesiones agrupadas con el programa STR

Fuentes de variación	σ	%	ϕ
Varianza entre grupos de STR	12.255	0.617	0.006
Varianza entre las accesiones	-654.843	-32.950	-0.332
Varianza dentro de las accesiones	2630.000	132.333	-0.323
Varianza total	1987.412	100.000	

Para los grupos determinados por el origen de las accesiones, ocurrió de manera similar a lo visto con los anteriores criterios de formación de grupos. Se evidencio una baja diferenciación entre los grupos, siendo esta 0.4% de la varianza, mayor a la mostrada por los grupos de razas, pero menor a lo obtenido con los grupos de STR. En cuanto al resto de la varianza se distribuyó entre y dentro de las accesiones, las cuales volvieron a mostrar un exceso en el número de genotipos heterocigotos, ya que se observaron valores de ϕ negativos tanto dentro como entre las accesiones.

Tabla 20: Tabla de resultados del AMOVA de las accesiones agrupadas de acuerdo al origen

Fuentes de variación	σ	%	ϕ
Varianza entre grupos por origen de accesiones	8.875	0.448	0.004
Varianza entre las accesiones	-655.599	-33.056	-0.332
Varianza dentro de las accesiones	2630.000	132.609	-0.326
Varianza total	1983.276	100.000	

Tal como indica Dao (2014) es común utilizar varios métodos multivariados para el análisis de la diversidad genética molecular y estructura genética. Sin embargo, los resultados obtenidos en este estudio no mostraron una separación totalmente precisa por raza, origen ni por los grupos estimados por medio de STR.

Así como lo obtenido en el presente estudio, Pineda – Hidalgo (2013) indican que los tres métodos de agrupación (STR, PCA, euclidiana/UPGMA) utilizados mostraron cierto nivel de concordancia en el reconocimiento de los grupos.

No obstante, entre estos tres criterios de agrupamiento el determinado por STR es el que ha obtenido los agrupamientos más adecuados, así como los índices y estimadores de mayor valor, seguido por los grupos determinados por el origen y, por último, los asignados a las razas de maíz amiláceo, esto también fue hallado por Leng *et al.* (2019), Ertiro *et al.* (2017) y Semagn *et al.* (2012), tomando en cuenta otros criterios como programa de reproducción, instituto de procedencia, resistencia a enfermedades y tolerancia a sequía.

En cuanto a la falta de concordancia con las razas, Bracco *et al.* (2012) menciona que las prácticas de gestión realizadas por los agricultores influyen enormemente en la distribución de la diversidad genética y el flujo de genes, tal vez es por ello que a pesar de que se encuentran asignadas a una misma raza, no hay alta similitud, sin embargo si consideramos el departamento de origen, esto permite agruparlas de forma un poco más precisa, tomando en cuenta que los intercambios y flujo de genes se da entre las accesiones que manejan los agricultores que pertenecen a una misma localidad y por ende departamento.

V. CONCLUSIONES

Se estimó y analizó la variabilidad genética empleando descriptores vegetativos, de mazorca y grano en la caracterización morfológica, con la cual se halló un alto grado de variabilidad en los caracteres evaluados. Además, de la caracterización molecular a través del desarrollo de marcadores tipo SNPs en 120 accesiones de las razas Piscorunto, Paro y Chullpi empleando Genotipado por secuenciamiento.

Se estimaron y analizaron parámetros genéticos, como GD (0.38), Ho (0.31), PIC (0.3) y MAF (0.28); y de la estructura genética de las razas Piscorunto, Paro y Chullpi, así como otros posibles criterios de agrupamiento para todas las accesiones evaluadas, los cuales no evidenciaron una estructura definida empleando marcadores SNP para la muestra de accesiones evaluada.

VI. RECOMENDACIONES

Continuar con los trabajos de investigación en fenotipado y genotipado de otras razas de maíz conservadas en el Banco de Germoplasma del Programa de Investigación en Maíz de la Universidad Nacional Agraria La Molina.

Las evaluaciones para la caracterización morfológica deberían incluir por lo menos dos campañas, ya que esto permitiría emplear esta información con más confiabilidad para la selección sobre los caracteres con menor heredabilidad.

VII. REFERENCIAS BIBLIOGRÁFICAS

- Anderson, C; Pettersson, F; Clarke, G; Cardon, L; Morris, A; Zondervan, K. 2010. Data quality control in genetic case-control association studies. *Nat Protoc* 5: 1564–1573
<https://doi.org/10.1038/nprot.2010.116>
- Al-Naggar, AMM; Soliman, AM; Hussien, MH; Mohamed, AMH. 2022. Genetic diversity of maize inbred lines based on morphological traits and its association with heterosis. *Sabrao Journal of Breeding and Genetics* 54(3):589-597. DOI: <https://doi.org/10.54910/sabrao2022.54.3.11>.
- Badu-Apraku, B; Garcia-Oliveira, AL; Petrolí, CD; Hearne, S; Adewale, SA; Gedil, M. 2021. Genetic diversity and population structure of early and extra-early maturing maize germplasm adapted to sub-Saharan Africa. *BMC Plant Biology* 21(1). DOI: <https://doi.org/10.1186/s12870-021-02829-6>.
- Baduwal, P; Chand, H; Kayastha, P; Lamichhane, P; Pandey, B; Kc, B; Magar, BR; Bhandari, J; Khanal, S. 2022. Correlation analysis of maize (*Zea mays* L.) genotypes: A review. *International Journal of Environment, Agriculture and Biotechnology* 7(6):153-157. DOI: <https://doi.org/10.22161/ijeab.76.17>
- Bates, D; Maechler, M; Bolker, B; Walker, S; Bojesen, R. Singmann, H; Scheipl, F; Dai, B; Grothendieck, G. 2023. Package “lme4”. Linear Mixed-Effects Models using “Eigen” and S4. Consultado 6 dic. 2023. Disponible en <https://cran.r-project.org/web/packages/lme4/lme4.pdf>
- Bedoya, CA; Dreisigacker, S; Hearne, S; Franco, J; Mir, C; Prasanna, BM; Taba, S; Charcosset, A; Warburton, ML. 2017. Genetic diversity and population structure of native maize populations in Latin America and the Caribbean. *PLoS ONE* 12(4):1-21. DOI: <https://doi.org/10.1371/journal.pone.0173488>.

- Boakyewaa Adu, G; Badu-Apraku, B; Akromah, R; Garcia-Oliveira, AL; Awuku, FJ; Gedil, M. 2019. Genetic diversity and population structure of early-maturing tropical maize inbred lines using SNP markers. PLoS ONE 14(4):1-12. DOI: <https://doi.org/10.1371/journal.pone.0214810>.
- Botstein, D; White, RL; Skolnick, M; Davis, RW. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet. (3):314-31. PMID: 6247908; PMCID: PMC1686077.
- Bracco, M; Lia, V; Gottlieb, AM; Cámara, J; Poggio, L. 2009. Genetic diversity in maize landraces from indigenous settlements of Northeastern Argentina. Genetica 135(1):39-49. DOI: <https://doi.org/10.1007/s10709-008-9252-z>.
- Bracco, M; Lia, V; Hernández, JC; Poggio, L; Gottlieb, AM. 2012. Genetic diversity of maize landraces from lowland and highland agro-ecosystems of Southern South America: Implications for the conservation of native resources. Annals of Applied Biology 160(3):308-321. DOI: <https://doi.org/10.1111/j.1744-7348.2012.00544.x>.
- Bradbury, PJ; Zhang, Z; Kroon, DE; Casstevens, TM; Ramdoss, Y; Buckler, ES. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics 23(19):2633-2635. DOI: <https://doi.org/10.1093/bioinformatics/btm308>.
- Catalán, F. 2016. Caracterización agrobotánica de ciento treinta y uno accesiones de maíz (*Zea mays* L.) de la subcuenca del Rio Ollambamba distrito de Mariscal Gamarra provincia de Grau – Apurímac. Título Lic. Cusco, Perú. Universidad Nacional De San Antonio Abad Del Cusco. 124 p.
- CBD (Convention on Biological Diversity). 2010. Uses of genetic resources (en línea). Montreal, Canadá. 5 p. Consultado 14 mar. 2019. Disponible en <https://www.cbd.int/abs/infokit/factsheet-uses-en.pdf>
- Chavarry, B. 2014. Caracterización morfológica de una muestra de accesiones de maíces peruanos del banco de germoplasma de maíz (*Zea mays* L.) de la UNALM. Título Lic. Lima, Perú. Universidad Nacional Agraria La Molina. 116 p.

- Chávez, AP. 2018. Determinación de una nueva raza de maíz (*Zea mays* L.), mediante la caracterización fenotípica de mazorcas recolectadas en Lambayeque. Tesis Lic. Lima, Perú. Universidad Nacional Agraria La Molina. 98 p.
- Cosio, P; Catalan, W; Yepez, E; Meza, Y. 2014. Conservación, caracterización y evaluación del germoplasma de maíz del CICA-DAZ-UNSAAC. ISBN:978-612-46410-7-7.
- Dao, A; Sanou, J; Mitchell, SE; Gracen, V; Danquah, EY. 2014. Genetic diversity among INERA maize inbred lines with single nucleotide polymorphism (SNP) markers and their relationship with CIMMYT, IITA, and temperate lines. *BMC Genetics* 15(1):1-14. DOI: <https://doi.org/10.1186/s12863-014-0127-2>.
- Darkwa, K; Agre, P; Olanmi, B; Iseki, K; Matsumoto, R; Powell, A; Bauchet, G; De Koeyer, D; Muranaka, S; Adebola, P; Asiedu, R; Terauchi, R; Asfaw, A. 2020. Comparative assessment of genetic diversity matrices and clustering methods in white Guinea yam (*Dioscorea rotundata*) based on morphological and molecular markers. *Scientific Reports* 10(1). DOI: <https://doi.org/10.1038/s41598-020-69925-9>.
- Dávalos, A. 2017. Diversidad de maíz (*Zea mays* L.) en la selva peruana. Título Lic. Lima, Perú. Universidad Nacional Agraria La Molina 90 p.
- De Vicente, M. 2004. Introduction. In De Vicente, MC (ed.). *The evolving role of genebanks in the fast-developing field of molecular genetics*. Rome, Italy. *Genetic Resources* 11:8-12. Consultado 25 oct. 2018. Disponible en https://www.researchgate.net/publication/236668218_Genebank_management_and_the_potential_role_of_molecular_genetics_to_improve_the_use_of_conserved_genetic_diversity
- De Vries, A; Ripley, B. 2022. Package “ggdendro”. Create Dendrograms and Tree Diagrams Using “ggplot2”. Consultado 6 dic. 2023. Disponible en <https://cran.r-project.org/web/packages/ggdendro/ggdendro.pdf>

- Dinesh, A; Patil, A; Zaidi, PH; Kuchanur, PH; Vinayan, MT. 2016. Genetic diversity, linkage disequilibrium and population structure among CIMMYT maize inbred lines, selected for heat tolerance study. *Maydica* 61(29):1-7.
- Doyle, JJ; Doyle, JL. 1990. Isolation of plant DNA from fresh tissue. *Focus* 12:13-15.
- Elshire, RJ; Glaubitz, JC; Sun, Q; Poland, JA; Kawamoto, K; Buckler, ES; Mitchell, SE. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6(5):1-10. DOI: <https://doi.org/10.1371/journal.pone.0019379>.
- EMBL-EBI. 2021. Variant Calling Format versión 4.0 (en línea). Consultado el 21 dic. 2023. Disponible en <https://www.internationalgenome.org/wiki/Analysis/vcf4.0/>
- Ertiro, BT; Semagn, K; Das, B; Olsen, M; Labuschagne, M; Worku, M; Wegary, D; Azmach, G; Ogugo, V; Keno, T; Abebe, B; Chibsa, T; Menkir, A. 2017. Genetic variation and population structure of maize inbred lines adapted to the mid-altitude sub-humid maize agro-ecology of Ethiopia using single nucleotide polymorphic (SNP) markers. *BMC Genomics* 18(1):1-11. DOI: <https://doi.org/10.1186/s12864-017-4173-9>.
- Evanno, G; Regnaut, S; Goudet, J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology* 14(8):2611-2620. DOI: <https://doi.org/10.1111/j.1365-294X.2005.02553.x>.
- Excoffier, L; Lischer, HE. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* 10(3):564-7. DOI: 10.1111/j.1755-0998.2010.02847.x.
- Excoffier, L; Smouse, PE; Quattro, JM. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**:479-491.

- Fan, FY. 2022. Package “FinCal”. Time Value of Money, Time Series Analysis and Computational. Consultado 6 dic. 2023. Disponible en <https://cran.r-project.org/web/packages/FinCal/FinCal.pdf>
- FAO (Organización de la Naciones Unidas para la agricultura y la alimentación). 1993. Modalidades del consumo del maíz (en línea). In: El maíz en la nutrición humana. Roma, Italia. Consultado 10 dic. 2019. Disponible en <http://www.fao.org/3/t0395s/T0395S00.htm#Contents>
- FAO (Food and Agriculture Organization of united nations). 2015. Genetic resources for food security and nutrition. Consultado 20 mar. 2019. Disponible en línea: <http://www.fao.org/3/a-i5049e.pdf>
- Ferreira, M.E. & Grattapaglia, D. 1998. Introducción al uso de Marcadores Moleculares en el análisis genético. 1ra Edición. Brasilia: EMBRAPA-CENARGEN, pp 220.
- Francis, RM. 2017. pophelper: an R package and web app to analyse and visualize population structure. *Molecular Ecology Resources* 17:27–32. DOI:10.1111/1755-0998.12509
- Franco, TL; Hidalgo, R. 2003. International Plant Genetic Resources Institute. Análisis estadístico de datos de caracterización morfológica de recursos fitogenéticos. s.l., Instituto Internacional de Recursos Fitogenéticos. 89 p. Disponible en https://cropgenebank.sgrp.cgiar.org/images/file/learning_space/technicalbulletin8.pdf
- Frascaroli, E; Schrag, TA; Melchinger, AE. 2013. Genetic diversity analysis of elite European maize (*Zea mays* L.) inbred lines using AFLP, SSR, and SNP markers reveals ascertainment bias for a subset of SNPs. *Theoretical and Applied Genetics* 126(1):133-141. DOI: <https://doi.org/10.1007/s00122-012-1968-6>.
- Gallardo, JA. 2023. Capítulo 3. Métodos jerárquicos de Análisis Cluster. Universidad de Granada. Granada, España. Consultado 6 dic. 2023. Disponible en <https://www.ugr.es/~gallardo/pdf/cluster-3.pdf>

- Gallardo, M. 2012. Concentrados y subproductos para la alimentación de rumiantes (en línea). In: XXI Curso Internacional de Lechería para Profesionales de América Latina. 10 p. Consultado 10 dic. 2019. Disponible en <http://prodanimal.fagro.edu.uy/cursos/NUTRICION/TEORICOS/Tema%202.%20Material%20de%20lectura.%20Concentrados%20y%20subproductos.pdf>
- Galili, T; Benjamini, Y; Simpson, G; Gallotta, M; Renaudie, J. 2023. Package “dendextend”. Extending “dendrogram” Functionality in R. Consultado 6 dic. 2023. Disponible en <https://cran.r-project.org/web/packages/dendextend/dendextend.pdf>
- Gill Langarica, HI. 2009. Diversidad genética del frijol común y su implicación en el mejoramiento genético asistido en México. Tesis de doctorado. Altamira, Tamaulipas. Mexico. Instituto Politécnico Nacional. 146 p.
- Girma, G; Hyma, KE; Asiedu, R; Mitchell, SE; Gedil, M; Spillane, C. 2014. Next-generation sequencing based genotyping, cytometry and phenotyping for understanding diversity and evolution of guinea yams. Theoretical and Applied Genetics 127(8):1783-1794. DOI: <https://doi.org/10.1007/s00122-014-2339-2>.
- Gonzales, F. 2016. Curso intensivo de posgrado: Caracterización de recursos genéticos (en línea). León, España. 380 p. Consultado 30 jul. 2019. Disponible en http://fernando.gonzalez.unileon.es/web_mex16/LIBRO_ALUMNOS_red.pdf
- 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. Genetics Research International ID 431487. DOI: <https://doi.org/10.1155/2015/431487>.
- Gouesnard, B; Negro, S; Laffray, A; Glaubitz, J; Melchinger, A; Revilla, P; Moreno-Gonzalez, J; Madur, D; Combes, V; Tollon-Cordet, C; Laborde, J; Kermarrec, D; Bauland, C; Moreau, L; Charcosset, A; Nicolas, S. 2017. Genotyping-by-sequencing highlights original diversity patterns within a European collection of

1191 maize flint lines, as compared to the maize USDA genebank. *Theoretical and Applied Genetics* 130(10):2165-2189.

Granato, I. Fritsche-Neto. 2022. Package “snpReady”. Prepariong genotypic datasetd in order to run genomic amalysis (en línea). Consultado el 21 dic. 2023. Disponible en <https://cran.r-project.org/web/packages/snpReady/snpReady.pdf>

Graner, A. Dehmer, K. Thiel, T. Borner, A. 2004. Plant genetic resources: benefits and implications of using molecular markers. In de Vicente, MC (ed.). *The evolving role of genebanks in the fast-developing field of molecular genetics*. Rome, Italy. *Genetic Resources* 11:8-12. Consultado 25 oct. 2018. Disponible en https://www.researchgate.net/publication/236668218_Genebank_management_and_the_potential_role_of_molecular_genetics_to_improve_the_use_of_conserved_genetic_diversity

Grobman, A; Salhuana, W; Sevilla, R; Mangelsdorf, PC. 1961. *Races of maize in Peru* (en línea). Washington, United States. 384 p. Consultado 10 abr. 2019. Disponible en https://www.ars.usda.gov/ARUserFiles/50301000/Races_of_Maize/RoM_Peru_0_Book.pdf.

Gruber, B; George, A; Mijangos, JL; Pacioni, C; Berry, O; Clark, L; Devloo-Delva, F; Archer, E. 2023. Package “dartR”. Importin and análisis SNP and Silicodart data generated by genoma-wide restriction fragment análisis (en línea). Consultado el 21 dic. 2023. Disponible en <https://cran.r-project.org/web/packages/dartR/dartR.pdf>

Harrell Jr, FE; Dupont, C. 2023. Package “Hmisc”. Harrell Miscellaneous. Consultado 6 dic. 2023. Disponible en <https://cran.r-project.org/web/packages/Hmisc/Hmisc.pdf>

He, J. Zhao, X. Laroche, A. Lu, Z. Liu, H. Li, Z. 2014. Genotyping by sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Frontiers in Plant Science* 5(1): 484.

- Hernández, R. Fernández, C. Baptista, P. 2014. Metodología de la investigación. Mexico: McGraw Hill Education.
- Hernández, AE. 2013. Caracterización morfológica de los recursos fitogenéticos. Revista Biológica de Ciencias 2(3): 113-118. ISSN 2007-3380.
- Husson, F; Josse, J; Le, S; Mazet, J. 2023. Package “FactoMineR”. Multivariate Exploratory Data Analysis and Data Mining. Consultado 6 dic. 2023. Disponible en <https://cran.r-project.org/web/packages/FactoMineR/FactoMineR.pdf>
- IBPGR. 1991. Descriptores para Maiz (en línea). Rome, Italy. 100 p. Consultado 30 jun. 2019. DOI: <https://doi.org/10.1017/CBO9781107415324.004>. Disponible en http://archive-cpgr.cgiar.org/fileadmin/bioversity/publications/pdfs/104_Descriptors_for_maize.Descriptores_para_maiz.Descripteurs_pour_le_mais-cache=1415188810.pdf
- Illumina Inc. 2020. Guia del sistema HiSeq 2500 (en línea). Consultado el 21 dic. 2023. Disponible en https://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/translations/hiseq-2500-system-guide-15035786-esp.pdf
- Islam, S; Ferdausi, A; Sweety, AY; Das, A; Ferdoush, A; Haque, MA. 2020. Morphological characterization and genetic diversity analyses of plant traits contributed to grain yield in maize (*Zea mays* L.). Journal of Bioscience and Agriculture Research 25(1):2047-2059. DOI: <https://doi.org/10.18801/jbar.250120.251>.
- Jawaharlal, J; Reddy, G; Kumar, R. 2011. Genetic variability and carácter association studies in maize. Agric. Sci. Digest. 31(3):173-177.
- Jiao, Y; Peluso, P; Shi, J; Liang, T; Stitzer; MC; Wang, B; Campbell, MS; Stein, JC; Wei, X; Chin, CS; Guill, K; Regulski, M; Kumari, S; Olson, A; Gent, J; Schneider, KL; Wolfgruber, TK; May, MR; Springer, NM; Antoniou, E; McCombie, WR; Presting, GG; McMulleu, M; Ross-Ibarra, J; Dawe, RK; Hastie, A; Rank, DR;

- Ware, D. Improved maize reference genome with single-molecule technologies. Nature. 2017. Nature 546(7659):524-527. DOI: 10.1038/nature22971.
- Jombart, T; Kamvar, Z; Collins, C; Lustrik, R; Beugin, MP; Knaus, B; Solynos, P; Mikryukov, V. Schliep, K; Maié, T; Morkovsky, L; Ahmed, I; Cori, A; Calboli, F; Ewing, RJ; Michaud, F; DeCamp, R; Cortiol, A; Clark, L; Dimens, P. 2023. Package “adegenet”. Toolset for the exploration of genetic and genomic data (en línea). Consultado el 21 dic. Disponible en <https://cran.r-project.org/web/packages/adegenet/adegenet.pdf>
- Jombart, T; Devillard, S; Balloux, F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genetics 11, 94. DOI:<https://doi.org/10.1186/1471-2156-11-94>
- Kamvar, Z; Tabima, J; Everhart, S; Brooks, J; Krueger-Hadfield, S; Sotka, E; Knaus, B; Meirmans, P; Chevalier, F; Folarin, D; Grunwald, N. 2023. Package “poppr”. Genetic análisis of populations with mixed reproduction (en línea). Consultado el 21 dic. 2023. Disponible en <https://cran.r-project.org/web/packages/poppr/poppr.pdf>
- Kassambara, A. 2017. Practical Guide to Principal Component Methods in R (en línea). s.l., s.e. Disponible <https://www.facebook.com/groups/stats.ebooksandpapers/>.
- Kassambara, A. 2022. Package “factoextra”. Extract and Visualize the Results of Multivariate Data Analyses. Consultado 6 dic. 2023. Disponible en <https://cran.r-project.org/web/packages/factoextra/factoextra.pdf>
- Knaus, B; Grunwald, N; Anderson, E; Winter, D; Kamvar, Z; Tabina, J. 2023. Package “vcfR”. Manipulate and visualize VCF data (en línea). Consultado el 21 dic. 2023. Disponible en <https://cran.r-project.org/web/packages/vcfR/vcfR.pdf>
- Lanes, ECM; Viana, JMS; Paes, GP; Paula, MFB; Maia, C; Caixeta, ET; Miranda, G V. 2014. Population structure and genetic diversity of maize inbreds derived from tropical hybrids. Genetics and Molecular Research 13(3):7365-7376. DOI: <https://doi.org/10.4238/2014.September.12.2>.

- Lee, HY; Ro, NY; Jeong, HJ; Kwon, JK; Jo, JK; Ha, YS. 2016. Genetic diversity and population structure analysis to construct a core collection from a large Capsicum germplasm. *BMC Genomics* 17:142.
- Leng, Y; Lu, C; Li, L; Xiang, Y; Xia, C; Wei, R; Rong, T; Lan, H. 2019. Heterotic grouping based on genetic variation and population structure of maize inbred lines from current breeding program in Sichuan province, Southwest China using genotyping by sequencing (GBS). *Molecular Breeding* 39(3). DOI: <https://doi.org/10.1007/s11032-019-0946-y>.
- Lia, V V.; Poggio, L; Confalonieri, VA. 2009. Microsatellite variation in maize landraces from Northwestern Argentina: Genetic diversity, population structure and racial affiliations. *Theoretical and Applied Genetics* 119(6):1053-1067. DOI: <https://doi.org/10.1007/s00122-009-1108-0>.
- Liu, K; Goodman, M; Muse, S; Smith, JS; Buckler, E; Doebley, J. 2003. Genetic Structure and Diversity Among Maize Inbred Lines as Inferred From DNA Microsatellites. *Genetics* 165: 2117-2128. Consultado 10 dic, 2023. Disponible en <https://academic.oup.com/genetics/article/165/4/2117/6050529>
- López, S; Nieto, A; Barrietos, A; Rodríguez, P; Colinas, M; Borys, M. 2008. Selección de variables morfológicas para la caracterización del tejocote (*Crataegus* spp). *Revista Chapingo Serie Horticultura* 14(2): 97-111.
- López-Zavala, R; Cano-Camacho, H; Chassin-Noria, O; Oyama, K; Vázquez-Marrufo, G; Zavala-Páramo, G. 2013. Diversidad genética y estructura de poblaciones de pavos domésticos mexicanos. *Revista Mexicana de Ciencias Agropecuarias* 4(4): 417-434. Consultado el 28 dic, 2023. Disponible en línea https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S2007-11242013000400002
- Luminvero. 2023. Analisis de componentes principales (ACP) (en línea). Consultado el 21 dic. Disponible en <https://www.xlstat.com/es/soluciones/funciones/analisis-de-componentes-principales-acp>

- Luo, Z; Brock, J; Dyer, JM; Kutchan, T; Schachtman, D; Augustin, M; Ge, Y; Fahlgren, N; Abdel-Haleem, H. 2019. Genetic diversity and population structure of a *Camelina sativa* spring panel. *Frontiers in Plant Science* 10. DOI: <https://doi.org/10.3389/fpls.2019.00184>.
- Masuelli, R. 1999. Uso de marcadores moleculares en el mejoramiento genético de especies hortícolas. *Avances en Horticultura* 4(1): 54-66.
- Machaca, R. 2017. Caracterización agronómica y morfológica de doce accesiones de maíz altiplánico (*Zea mays* L.) tolerantes al frío en CIP Camacani, Puno. Título Lic. Puno, Perú. Universidad Nacional del Altiplano. 138 p.
- McLean-Rodríguez, FD; Costich, DE; Camacho-Villa, TC; Pè, ME; Dell'Acqua, M. 2021. Genetic diversity and selection signatures in maize landraces compared across 50 years of in situ and ex situ conservation. *Heredity* 126(6):913-928. DOI: <https://doi.org/10.1038/s41437-021-00423-y>.
- Macuri, E. 2016. Estudio de la diversidad fenotípica del maíz (*Zea mays* L.) en la sierra baja y media del Perú. Título de Lic. Lima, Perú. Universidad Nacional Agraria La Molina. 107 p.
- Maechler, M; Rousseeuw, P; Struyf, A; Hubert, M; Hornik, K; Studer, M; Roudier, P; Gonzales, J; Kozłowski, K; Schubert, E; Murphy, K. 2023. Consultado 6 dic. 2023. Disponible en <https://cran.r-project.org/web/packages/cluster/cluster.pdf>
- Magar, BT; Acharya, S; Gyawali, B; Timilsena, K; Upadhayaya, J; Shrestha, J. 2021. Genetic variability and trait association in maize (*Zea mays* L.) varieties for growth and yield traits. *Heliyon* 7(9):e07939. DOI: <https://doi.org/10.1016/j.heliyon.2021.e07939>.
- Mahmood, T; Qasim, M; Ahmad, S; Bin Sajid, H; Abdullah, M; Dilshad, R; Tahir, N. 2022. *Zea mays* L. Germplasm characterization based on various morphological attributes. *The International Journal of Biological Research* 5(1). ISSN Online:2618-14444.

- Martínez, M; Ortiz, R; Ríos, H; Acosta, R. 2010. Análisis de las correlaciones en poblaciones cubanas de maíz. *Cultivos tropicales* 31(2):82-91.
- Mendiburu, F. 2023. Package “agricolae”. *Statistical Procedures for Agricultural Research*. Consultado 6 dic. 2023. Disponible en <https://cran.r-project.org/web/packages/agricolae/agricolae.pdf>
- MINAGRI. 2012. Maíz amiláceo: Principales Aspectos de la Cadena Productiva (en línea). Lima, Perú. 37 p. Consultado 20 nov. 2019. Disponible en <http://agroaldia.minagri.gob.pe/biblioteca/download/pdf/agroeconomia/agroeconomiamaimaizamilaceo.pdf>
- Mohammadi, SA; Prasanna, BM. 2003. REVIEW & INTERPRETATION Analysis of Genetic Diversity in Crop Plants-Salient Statistical Tools and Considerations. *Crop Science* 43:1235-1248.
- Molin, D; Coelho, CJ; Máximo, DS; Ferreira, FS; Gardingo, JR; Matiello, RR. 2013. Genetic diversity in the germplasm of tropical maize landraces determined using molecular markers. *Genetics and Molecular Research* 12(1):99-114. DOI: <https://doi.org/10.4238/2013.January.22.8>.
- Nazareno, AG; Knowles LL. 2021. There Is No 'Rule of Thumb': Genomic Filter Settings for a Small Plant Population to Obtain Unbiased Gene Flow Estimates. *Front Plant Sci*.12:677009. doi: 10.3389/fpls.2021.677009. PMID: 34721447; PMCID: PMC8551369.
- N'DA, HA; Akanvou, L; Akanza, KP; Kouakou, CK. 2022. Morphological characterization of purple-grain maize (*zea mays* L.) diversity in Cte dIvoire. *African Journal of Agricultural Research* 18(10):860-869. DOI <https://doi.org/10.5897/ajar2022.16065>.
- Nei M. 1973. Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci* 70(12):3321-3. doi: 10.1073/pnas.70.12.3321. PMID: 4519626; PMCID: PMC427228.

- Núñez-Colín, C. A., y Escobedo-López, D. (2011). Uso correcto del análisis clúster en la caracterización de germoplasma vegetal. *Agronomía Mesoamericana* 22(2): 415-427. Disponible en <https://www.scielo.sa.cr/pdf/am/v22n2/a18v22n2.pdf>
- ONU (Organización de las Naciones Unidas). 1992. Convenio sobre la diversidad biológica (en línea). Río de Janeiro, Brasil. 30 p. Consultado 30 mar. 2019. Disponible en <https://www.cbd.int/doc/legal/cbd-es.pdf>
- Ortiz, R. Engels, J. 2004. Genebank management and the potential role of molecular genetics to improve the use of conserved genetic diversity. In de Vicente, MC (ed.). *The evolving role of genebanks in the fast-developing field of molecular genetics*. Rome, Italy. *Genetic Resources* 11:8-12. Consultado 25 oct. 2018. Disponible en https://www.researchgate.net/publication/236668218_Genebank_management_and_the_potential_role_of_molecular_genetics_to_improve_the_use_of_conserved_genetic_diversity
- Oscanoa Rodríguez, C; Sevilla Panizo, R. 2011. Mejoramiento conservativo del maíz en la sierra del Perú. *Inia* 3:30-39.
- Oscanoa, C; Sevilla, R. 2008. Razas de maíz de la sierra central del Perú: Junín, Huancavelica y Ayacucho. Junín, Perú. 486 p. Consultado 20 mar. 2020. Disponible en https://repositorio.inia.gob.pe/bitstream/20.500.12955/969/1/Oscanoa-Incremento_de_rendimiento_de_%20maiz_en_la_%20sierra_central%20_del%20_Peru_a_traves_de_conservacion_de_razas.pdf
- Oscanoa, C; Jara, W; Sevilla, R; Narro, T; Altet, A. 2006. Colección de germoplasma de maíz criollo en sierra central del Perú (en línea). Lima, Perú. 83 p. Consultado 20 jun. 2019. Disponible en <https://repositorio.inia.gob.pe/handle/inia/969>
- Papastamoulis, P. 2022. Package “label.switching”. Relabelling MCMC outputs of mixture models (en línea). Consultado el 21 dic. 2023. Disponible en <https://cran.r-project.org/web/packages/label.switching/label.switching.pdf>
- Patel, R; Patel, DA; Memon, J; Das, A; Patil, K. 2023. D 2 clustering of yield and yield accredited attributes for genetic diversity analysis in maize (*Zea mays* L.). *The*

Pharma Innovation Journal 12(3):2210-2213. Disponible en www.thepharmajournal.com.

PCIM (Programa Cooperativo de Investigación en Maíz). 1997. Programa de cereales y leguminosas: área de maíz (en línea). Lima, Perú. Consultado 18 ene. 2019. Disponible en <http://www.lamolina.edu.pe/facultad/agronomia/programa/maiz>

Peterson, G; Dong, Y; Horbach, C; Fu, Y. 2014. Genotyping by sequencing for Plant genetic diversity analysis: a lab guide for SNP genotyping. *Diversity* 6(1):665-680.

Perdry, H; Dandine-Roulland, C; Bandyopadhyay, D; Kettner, L. 2023. Package “gaston”. Genetic data handling & Linear mixed models (en línea). Consultado el 21 dic. 2023. Disponible en <https://cran.r-project.org/web/packages/gaston/gaston.pdf>

Pineda-Hidalgo, KV; Méndez-Marroquín, KP; Alvarez, EV; Chávez-Ontiveros, J; Sánchez-Peña, P; Garzón-Tiznado, JA; Vega-García, MO; López-Valenzuela, JA. 2013. Microsatellite-based genetic diversity among accessions of maize landraces from Sinaloa in México. *Hereditas* 150(4):53-59. DOI: <https://doi.org/10.1111/j.1601-5223.2013.00019.x>.

Poncet, P. 2022. Package “modeest”. Mode Estimation. Consultado 6 dic. 2023. Disponible en <https://cran.r-project.org/web/packages/modeest/modeest.pdf>

Pritchard, JK, Stephens, M, and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.

Rai, VP; Kumar, R; Kumar, S; Rai, A; Kumar, S; Singh, M; Singh, SP; Rai, AB; Paliwal, R. 2013. Genetic diversity in Capsicum germplasm based on microsatellite and random amplified microsatellite polymorphism markers. *Physiology and Molecular Biology of Plants* 19(4):575-586. DOI: <https://doi.org/10.1007/s12298-013-0185-3>.

Reif, JC; Warburton, ML; Xia, XC; Hoisington, DA; Crossa, J; Taba, S; Muminović, J; Bohn, M; Frisch, M; Melchinger, AE. 2006. Grouping of accessions of Mexican

- ances of maize revisited with SSR markers. *Theoretical and Applied Genetics* 113(2):177-185. DOI: <https://doi.org/10.1007/s00122-006-0283-5>.
- Rellstab, C; Bühler, A; Graf, R; Folly, C; Gugerli, F. 2016. Using joint multivariate analyses of leaf morphology and molecular-genetic markers for taxon identification in three hybridizing European white oak species (*Quercus* spp.). *Annals of Forest Science* 73(3):669-679. DOI: <https://doi.org/10.1007/s13595-016-0552-7>
- Ribeiro, CAG; Pastina, MM; Guimaraes, LJM; Guimaraes, PEO; Pacheco, CAP; Malgalhaes, JV; Parentoni, SN; Barros, EG; Guimaraes, CT. 2013. Genetic diversity and population structure assessed by SNP markers in a panel of maize inbred lines. Resumos do 59° Congresso Brasileiro de Genética. Consultado el 10 dic, 2023. Disponible en línea <https://www.alice.cnptia.embrapa.br/alice/bitstream/doc/974205/1/Geneticdiversity1.pdf>
- Rincón Sánchez, F. 2020. Anexo 7: Guía práctica para la Descripción Preliminar de Colectas de Maíz. En Proyecto: Conocimiento de la diversidad y distribución actual del maíz nativo y sus parientes silvestres en México (en línea). México. 6 p. Consultado 26 jun 2020. Disponible en <https://www.biodiversidad.gob.mx/media/1/genes/files/GuiaPracticaMaiz.pdf>
- Rivas, JG; Gutierrez, AV; Defacio, RA; Schimpf, J; Vicario, AL; Hopp, HE; Paniego, NB; Lia, VV. 2022. Morphological and genetic diversity of maize landraces along an altitudinal gradient in the Southern Andes. *PLoS ONE* 17(12 December). DOI: <https://doi.org/10.1371/journal.pone.0271424>.
- Rodríguez, M. & Arencibia, A. 2002. Principales tipos de marcadores del polimorfismo de los ácidos nucleicos. Técnicas analíticas. In Cornide. MT. (ed.). *Marcadores moleculares. Nuevos horizontes en la genética y la selección de las plantas*. 30 p.
- Roy, P; Haque, M; Ferdausi, A; Bari, M. 2018. Field performance based indepth genetic analysis of selected maize (*Zea mays* L.) genotypes. *Fundamental and Applied Agriculture* :1. DOI: <https://doi.org/10.5455/faa.284502>.

- Rudis, B; Kennedy, P; Reiner, P; Wilson, D; Adam, X; Barnnet, J; Leeper, T; Meys, J. 2022. Package “hrbrthemes”. Additional Themes, Theme Components and Utilities for “ggplot2”. Consultado 6 dic. 2023. Disponible en <https://cran.r-project.org/web/packages/hrbrthemes/hrbrthemes.pdf>
- Salhuana W. 2004. Diversidad y descripción de las Razas de Maíz en el Perú (en línea). Lima, Perú. 35 p. Consultado 14 feb. 2020. Disponible en https://www.ars.usda.gov/ARSDocuments/2010/05/0301000/Races_of_Maize/Diversidad%20y%20razas%20de%20maiz%20en%20Peru.pdf
- Schaefer, CM; Bernardo, R. 2013. Population structure and single nucleotide polymorphism diversity of historical Minnesota maize inbreds. *Crop Science* 53(4):1529-1536. DOI: <https://doi.org/10.2135/cropsci2012.11.0632>.
- Sevilla, R. 2006. Definiciones conceptuales básicas. In: Manual para la caracterización In Situ de Cultivos Nativos: conceptos y procedimientos. Estrada, R., Medina, T., Roldan, A (Eds.). Lima, Perú. 167 p.
- Sevilla, R. Holle, M. 2004. Recursos genéticos vegetal. 1era Edición. Edición Luis León Asoc. Lima-Perú, 445 p.
- Semagn, K; Vivek, BS; Makumbi, D; Beyene, Y; Mugo, S; Prasanna, BM; Warburton, ML. (2012). Molecular characterization of diverse CIMMYT maize inbred lines from eastern and southern Africa using single nucleotide polymorphic markers (en línea). s.l., s.e. Disponible en <http://cmpg.unibe.ch/software/arlequin3>.
- Schauberger, P; Walker, A; Braglia, L; Sturm, J; Garbuszus, JM; Barbone, JM. 2023. Package “openxlsx”. Read, Write and Edit xlsx Files. Consultado 6 dic. 2023. Disponible en <https://cran.r-project.org/web/packages/openxlsx/openxlsx.pdf>
- Schnable, PS; Pasternak, S; Liang, C. 2009. The B73 Maize Genome: Complexity, Diversity, and Dynamics. *Science* 326(1):1112-1115.

- Silva, W. Carvalho, S. Duarte, J. 2013. Identification of minimum descriptors for characterization of *Capsicum* spp. germplasm. *Horticultura Brasileira* 31: 190-202. DOI: <https://doi.org/10.1590/s0102-05362013000200004>.
- Sinana, HF; Ravikesavan, R; Iyanar, K; Senthil, A. 2023. Study of genetic variability and diversity analysis in maize (*Zea mays* L.) by agglomerative hierarchical clustering and principal component analysis. *Electronic Journal of Plant Breeding* 14(1):43-51. DOI: <https://doi.org/10.37992/2023.1401.015>.
- Suhr, DD. 2020. Principal Component Analysis vs. Exploratory Factor Analysis. s.l., s.e.
- Suman, SK; Kumar, M; Kumar, A; Kumar, R. 2018. Genetic Diversity Analysis of Maize (*Zea mays* L.) Inbred Lines on the basis of Morphological Traits. *International Journal of Current Microbiology and Applied Sciences* 7(12):3219-3224. DOI: <https://doi.org/10.20546/ijcmas.2018.712.372>.
- Sun, Y V.; Kardia, SLR. 2008. Imputing missing genotypic data of single-nucleotide polymorphisms using neural networks. *European Journal of Human Genetics* 16(4):487-495. DOI: <https://doi.org/10.1038/sj.ejhg.5201988>.
- Suwanda, R; Syahputra, Z; Zamzami, EM. 2020. Analysis of Euclidean Distance and Manhattan Distance in the K-Means Algorithm for Variations Number of Centroid K. In *Journal of Physics: Conference Series*. s.l., Institute of Physics Publishing DOI: <https://doi.org/10.1088/1742-6596/1566/1/012058>.
- Tsindi, A; Eleblu, JSY; Gasura, E; Mushoriwa, H; Tongoona, P; Danquah, EY; Mwadzingeni, L; Zikhali, M; Ziramba, E; Mabuyaye, G; Derera, J. 2023. Analysis of population structure and genetic diversity in a Southern African soybean collection based on single nucleotide polymorphism markers. *CABI Agriculture and Bioscience* 4(1). DOI: <https://doi.org/10.1186/s43170-023-00158-2>.
- Yang, X; Gao, S; Xu, S; Zhang, Z; Prasanna, BM; Li, L; Li, J; Yan, J. 2011. Characterization of a global germplasm collection and its potential utilization for analysis of complex quantitative traits in maize. *Molecular Breeding* 28(4):511-526. DOI: <https://doi.org/10.1007/s11032-010-9500-7>.

- Van Heerwaarden, J; Ross-Ibarra, J; Doebley, J; Glaubitz, JC; De Jesús Sánchez González, J; Gaut, BS; Eguiarte, LE. 2010. Fine scale genetic structure in the wild ancestor of maize (*Zea mays* ssp. *parviglumis*). *Molecular Ecology* 19(6):1162-1173. DOI: <https://doi.org/10.1111/j.1365-294X.2010.04559.x>.
- van Inghelandt, D; Melchinger, AE; Lebreton, C; Stich, B. 2010. Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. *Theoretical and Applied Genetics* 120(7):1289-1299. DOI: <https://doi.org/10.1007/s00122-009-1256-2>.
- Velásquez y Montoro. 2011. Estudio sobre la biodiversidad, Erosión y Contaminación Genética del maíz nativo en Perú. In: *Biodiversidad, Erosión y Contaminación Genética del Maíz Nativo en América Latina del Maíz Nativo en América Latina*.
- Vishnuvardhan, K; Kumar, K; Murthy, K; Devi, M; Kumar, Y. 2021. Genetic variability, heritability and association análisis of morphological and yield components in maize hybrids. *Annals of Plant and Soil Research* 23(4):424-431. DOI: <https://doi.org/10.47815/apsr.2021.10096>
- Wang, R; Yu, Y; Zhao, J; Shi, Y; Song, Y; Wang, T; Li, Y. 2008. Population structure and linkage disequilibrium of a mini core set of maize inbred lines in China. *Theoretical and Applied Genetics* 117(7):1141-1153. DOI: <https://doi.org/10.1007/s00122-008-0852-x>.
- Ward, B. 2023. Package “bwardr”. Brian Ward personal R package (en línea). Consultado el 21 dic. 2023. Disponible en <https://rdrr.io/github/etnite/bwardr/>
- Wickham, H; Francois, R; Henry, L; Muller, K; Vaughan, D. 2023a. Package “dplyr”. A grammar of Data Manipulation. Consultado 6 dic. 2023. Disponible en <https://cran.r-project.org/web/packages/dplyr/dplyr.pdf>
- Wickham, H; Chang, W; Henry, L; Pedersen, TL; Takashi, K; Wilke, C; Woo, K; Yutani, H; Dunnington, D. 2023. Package “ggplot2”. Create Elegant Data Visualisations Using the Grammar of Graphics. Consultado 6 dic. 2023b. Disponible en <https://cran.r-project.org/web/packages/ggplot2/ggplot2.pdf>

- Wickham, H; Vaughan, D; Girlich, M; Ushey, K; PBC, P. 2023c. Package “tidyr”. Tools to helps to create tidy data (en línea). Consultado el 21 dic. 2023. Disponible en <https://cran.r-project.org/web/packages/tidyr/tidyr.pdf>
- Wickham, H; Hester, J; Chang, W; Bryan, J. 2022. Package “devtools”. Collection of package development tools (en línea). Consultado el 21 dic. 2023. Disponible en <https://cran.r-project.org/web/packages/devtools/devtools.pdf>
- Wei, T; Simko, V. 2021. Package “corrplot”. An Intriduction to corrplot package. Consultado 6 dic. 2023. Disponible en <https://cran.r-project.org/web/packages/corrplot/vignettes/corrplot-intro.html>
- Weir, BS; Cockerham, CC. 1984. Estimating F statistics for the análisis of population structure. *Evolution* 38:1358-1370. DOI: 10.2307/2408641
- Wright S. 1965. The Interpretation of Population Structure by F-Statistics with Special Regard to Systems of Mating *Evolution* 19:395–420. DOI: 10.1111/j.1558-5646.1965.tb01731.x.
- Wright, S. 1951. The Genetical Structure of Populations. *Annals of Eugenics*, 15: 323-354. <http://dx.doi.org/10.1111/j.1469-1809.1949.tb02451.x>
- Wright, S. 1949. The genetical structure of populations. *Annals of Human Genetics* 15: 323-354. DOI: 10.1111/j.1469-1809.1949.tb02451.x
- Wu, X; Li, Y; Shi, Y; Song, Y; Wang, T; Huang, Y; Li, Y. 2014. Fine genetic characterization of elite maize germplasm using high-throughput SNP genotyping. *Theoretical and Applied Genetics* 127(3):621-631. DOI: <https://doi.org/10.1007/s00122-013-2246-y>.
- Yepez, E. 2010. Caracterización morfológica y evaluación fenológica de sesenta y cinco entradas de maíz (*Zea mays* L.). Título de Lic. Cusco, Perú. Universidad Nacional De San Antonio Abad Del Cusco. 137 p.

- Zavinon, F; Adoukonou-Sagbadja, H; Bossikponnon, A; Dossa, H; Ahanhanzo, C. 2019. Phenotypic diversity for agro-morphological traits in pigeon pea landraces [(*Cajanus cajan* L.) Millsp.] cultivated in southern Benin. *Open Agriculture* 4(1):487-499. DOI: <https://doi.org/10.1515/opag-2019-0046>
- Zawadi, M; Sibiya, J; Mashingaidze, K; Amelework, AB; Kondwakwenda, A; Musvosvi, C. 2021. Molecular characterization and diversity analysis of selected maize inbred lines using single-nucleotide polymorphism markers. *Canadian Journal of Plant Science* 101(2):240-248. DOI: <https://doi.org/10.1139/cjps-2020-0051>.
- Zhang, X; Zhang, H; Li, L; Lan, H; Ren, Z; Liu, D; Wu, L; Liu, H; Jaqueth, J; Li, B; Pan, G; Gao, S. 2016. Characterizing the population structure and genetic diversity of maize breeding germplasm in Southwest China using genome-wide SNP markers. *BMC Genomics* 17(1). DOI: <https://doi.org/10.1186/s12864-016-3041-3>.

VIII. ANEXOS

Anexo 1: Datos pasaporte del material vegetal

Datos pasaporte de la raza Chullpi

Accesión	Departamento	Provincia	Distrito	Latitud	Longitud	Altitud
JUN-502	Junín	Huancayo	Pariahuanca	-12.02	-74.83	2850
HVCA-278	Huancavelica	Huaribamba	Huaribamba	-12.27	-74.93	3002
JUN-534	Junín					
JUN-523	Junín					
CCC-051	Cusco	Quispicanchis	Quiquija	-13.80	-71.05	
CCC-073	Cusco	Calca	Taray	-13.43	-71.09	
CCC-045	Cusco	Urcos	Urcos	-13.73	-72.06	
CCC-041	Cusco	Urubamba	Yucay	-13.28	-72.20	
CMC-037	Cusco	Cusco	Quispicanchis	-13.69	-71.62	
CAJ-104	Cajamarca	Hualgayoc	Hualgayoc	-6.76	-78.61	3300
CMC-036	Cusco	Urubamba	Urubamba	-13.31	-72.12	
CAJ-108	Cajamarca	Celendin	Lucma	-6.87	-78.14	2750
AYA-247	Ayacucho	La Mar	Tambo	-12.93	-74.02	3240
AYA-228	Ayacucho	Huanta	Iguaín	-12.98	-74.20	2929
AYA-271	Ayacucho	Victor Fajardo	Huancapi	-13.75	-74.05	3131
AYA-195	Ayacucho	Huamanga	Quinua	-13.07	-74.15	2994
APUC-320	Apurímac					
JUN-550	Junín					

JUN-554	Junín					
CAJ-110	Cajamarca	Hualgayoc	Hualgayoc	-6.76	-78.61	2750
CUZ-209	Cusco	Cusco	Cusco	-13.53	-71.97	3220
AYA-305	Ayacucho					
AYA-176	Ayacucho	Huamanga	Quinoa	-13.06	-74.14	2861
HVCA-275	Huancavelica	Huaribamba	Huaribamba	-12.27	-74.93	3002
HVCA-277	Huancavelica	Huaribamba	Huaribamba	-12.27	-74.93	3002
ANC-630	Ancash					
JUN-463	Junín	Tarma	Huaracayo	-11.33	-75.62	3314
HVCA-285	Junín					
CMC-502	Apurímac	Abancay	Huanipaca	-13.48	-72.89	2460
CMC-227	Cusco	Quispicanchis	Oropesa	-13.59	-71.76	
CMC-607	Cusco	Quispicanchis	Cusipata	-13.91	-71.50	3382
CMC-228	Apurímac	Abancay	Cachora	-13.51	-72.81	
AYA-203	Ayacucho	Huamanga	Vinchos	-13.07	-74.15	3362
AYA-178	Ayacucho	Huamanga	Quinoa	-13.07	-74.15	3033
AYA-193	Ayacucho	Huamanga	Quinoa	-13.07	-74.13	3026
AYA-188	Ayacucho	Huanta	Huamanguilla	-13.02	-74.17	3100
CMC-640	Cusco	Quispicanchis	Huaro	-13.69	-71.64	3176
CMC-599	Cusco	Cusco	Calca	-13.49	-71.78	3009
CMC-026	Cusco	Paruro	Paruro	-13.76	-70.85	
CMC-621	Cusco	Cusco	Calca	-13.34	-71.87	3680

Datos pasaporte de la raza Paro

Accesión	Departamento	Provincia	Distrito	Latitud	Longitud	Altitud
APUC-071	Apurímac	Abancay	Cachora	-13.51	-72.81	2750
APUC-069	Apurímac	Abancay	Cachira	-13.51	-72.81	2750
APUC-136	Apurímac	Abancay	Abancay	-13.64	-72.89	2500
APUC-110	Apurímac	Abancay	Cachora	-13.51	-72.81	2377
APUC-301	Apurímac	Andahuaylas	Pacucha	-13.61	-73.34	2820
HCO-120	Huánuco	Dos de mayo	Rondos	-9.41	-76.37	3600
AYA-123	Ayacucho	Huamanga	Socosvinchos	-13.16	-74.22	3100
HCO-112	Huánuco	Huánuco	Quisqui	-9.55	-76.64	3000
CUZ-260	Cusco	Calca	Pisac	-13.42	-71.85	3100
ANC-491	Ancash	Antonio Raymondi	Chingas	-9.12	-76.99	2600
ANC-375	Ancash	Bolognesi	Chiquian	-10.15	-77.16	3350
ANC-522	Ancash	Huari	Huari	-9.35	-77.17	3100
ANC-399	Ancash	Carhuaz	Tinco	-9.27	-77.68	2000
CMC-180	Apurímac	Andahuaylas	Santa Maria Chicmu	-13.65	-73.44	2820
CMC-146	Apurímac	Abancay	Abancay	-13.51	-72.81	2909
CMC-185	Apurímac	Abancay	Cachora	-13.51	-72.81	2909
CMC-181	Apurímac	Abancay	Cachora	-13.51	-72.81	2909
AYA-089	Ayacucho	Huanta	Huanta	-12.94	-74.25	2620
AYA-090	Ayacucho	Huanta	Paquecc	-12.94	-74.25	2570
AYA-093	Ayacucho	Huanta	San Jose	-12.58	-74.23	2400
ANC-622	Ancash	Sihuas	San Juan	-8.55	-77.63	2850
ANC-530	Ancash	Huari	Huari	-9.35	-77.17	3100
ANC-561	Ancash	Pomabamba	Pomabamba	-8.82	-77.46	2900
APUC-068	Apurímac	Abancay	Cachora	-13.51	-72.81	2750
APUC-211	Apurímac	Abancay	Cachora	-13.51	-72.81	2800
APUC-274	Apurímac	Aymaraes	Caraybamba	-14.38	-73.16	3100

APUC-150	Apurímac	Abancay	Abancay	-13.64	-72.89	2500
APUC-261	Apurímac	Abancay	Curahuasi	-13.54	-72.70	3600
JUN-008	Junín	Tarma	Tarma	-11.42	-75.69	3080
JUN-440	Junín	Tarma	Huaragayo	-11.33	-75.62	3030
HVCA-141	Huancavelica	Acobamba	Pomacocha	-12.85	-74.52	3100
JUN-535	Junín					
ANC-262	Ancash	Bolognesi	Chiquian	-10.15	-77.16	3300
ANC-044	Ancash	Huaraz	Restaurac	-9.53	-77.53	3220
ANC-074	Ancash	Yungay	Yungay	-9.14	-77.74	3000
ANC-047	Ancash	Huaraz	Independencia	-9.53	-77.53	3190
CMC-099	Apurímac	Andahuaylas	San Jeronimo	-13.65	-73.37	3500
CMC-085	Cusco	Anta	Mollepata	-13.52	-73.55	2659
CMC-143	Apurímac	Andahuaylas	Pacucha	-13.61	-73.34	3784
CMC-090	Apurímac	Andahuaylas	San Jeronimo	-13.65	-73.37	3500

Datos pasaporte de la raza Piscorunto

Accesión	Departamento	Provincia	Distrito	Latitud	Longitud	Altitud
AYA-187	Ayacucho	Huanta	Huamanguilla	-13.02	-74.17	3100
AYA-177	Ayacucho	Huamanga	Quinua	-13.07	-74.15	3000
APUC-309	Apurímac					
AYA-224	Ayacucho	Huanta	Iguaín	-12.98	-74.20	3086
CMC-032	Cusco	CANCHIS	Tinta	-14.14	-71.41	
MOQ-003	Moquegua					
CMC-030	Cusco	Quispicanchis	Huaro	-13.69	-71.64	
MOQ-013	Moquegua					
AYA-243	Ayacucho	La Mar	Tambo	-12.93	-74.02	3289
AYA-203	Ayacucho	Huamanga	Vinchos	-13.07	-74.15	3362
AYA-264	Ayacucho	Victor Fajardo	Huancaraylla4	-13.72	-74.08	3298

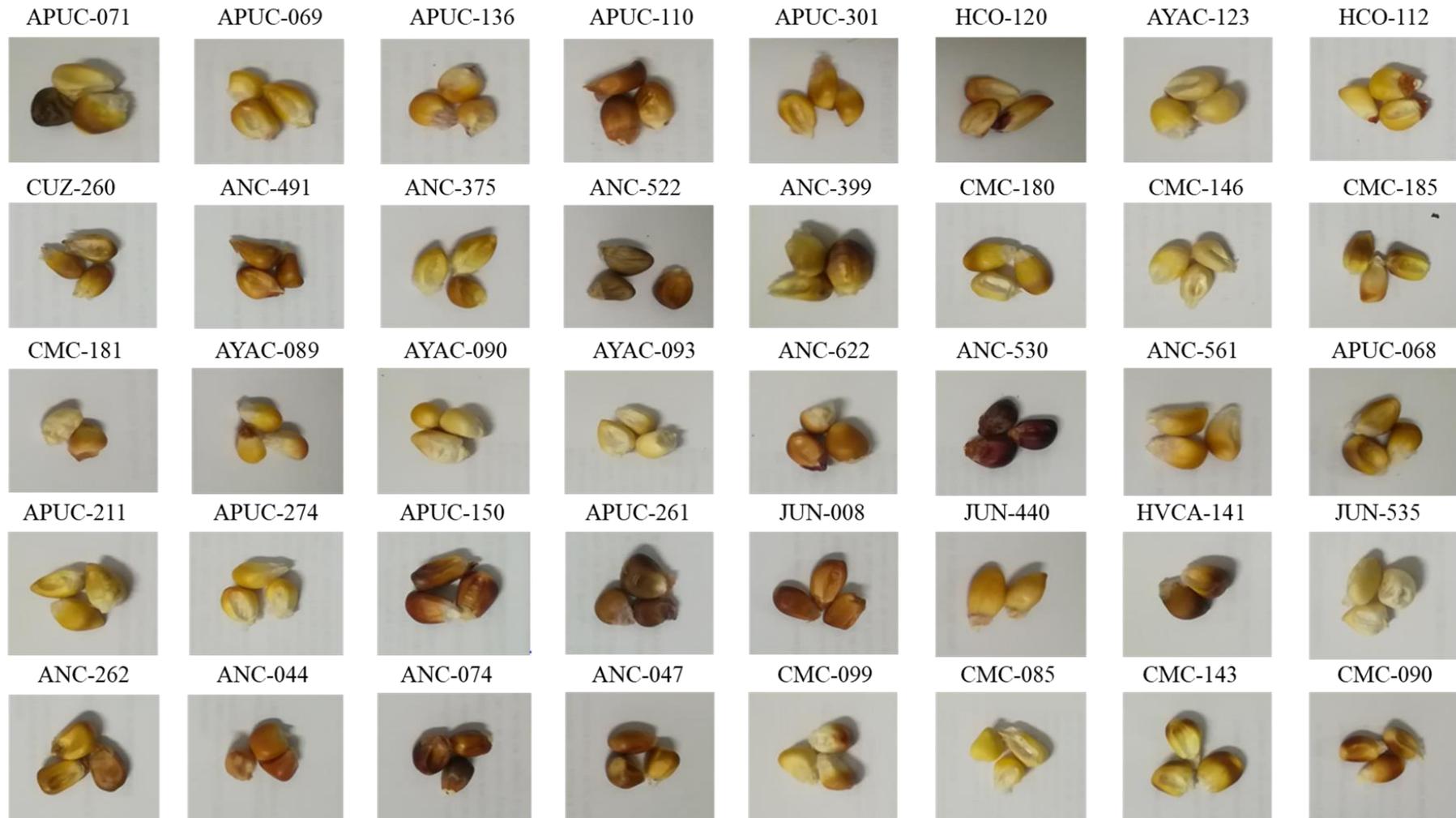
AYA-229	Ayacucho	Huanta	Iguaín	-12.98	-74.20	2927
CMC-396	Cusco	Cusco	Calca	-13.49	-71.78	
CMC-373	Cusco	Paucartambo	Paucartambo	-13.32	-71.59	
CMC-405	Cusco	Canchis	Marangani	-14.36	-71.17	
CMC-391	Cusco	Canchis	Cusipata	-13.91	-71.50	
CUZ-032	Cusco	Quispicanchis	Lucre	-13.63	-71.73	3130
MOQ-016	Moquegua					
MOQ-010	Moquegua					
MOQ-002	Moquegua					
HVCA-280	Huancavelica					
AYA-273	Ayacucho	Victor Fajardo	Huancapi	-13.75	-74.05	3151
AYA-295	Ayacucho	Huamanga	Socos	-13.20	-74.28	3416
HVCA-289	Huancavelica					
AYA-327	Ayacucho					
CUZ-034	Cusco	Calca	Calca	-13.32	-71.95	2850
APUC-307	Apurímac					
TAC-064	Tacna					
CMC-211	Cusco	Paruro	Paruro	-13.76	-71.85	
CMC-205	Cusco	Antabamba	Antabamba	-14.36	-72.87	
CMC-033	Cusco	Paruro	Paruro	-13.76	-71.85	
CMC-220	Cusco	Abancay	Cachora	-13.51	-72.81	
AYA-201	Ayacucho	Huamanga	Vinchos	-13.07	-74.15	3358
AYA-180	Ayacucho	Huamanga	Quinoa	-13.07	-74.13	3029
AYA-185	Ayacucho	Huanta	Huamaguilla	-13.02	-74.17	3090
AYA-183	Ayacucho	Huamanga	Quinoa	-13.05	-74.13	3223
CMC-222	Cusco	Cusco	Saylla	-13.57	-71.83	
CMC-215	Cusco	Chincheros	Chincheros	-13.52	-73.72	
CMC-436	Cusco	Canchis	San Pablo	-14.20	-71.32	
CMC-219	Cusco	Cusco	Anta	-13.52	-72.55	

Anexo 2: Evidencia fotográfica de las semillas que corresponden al material vegetal empleado para extracción de ADN

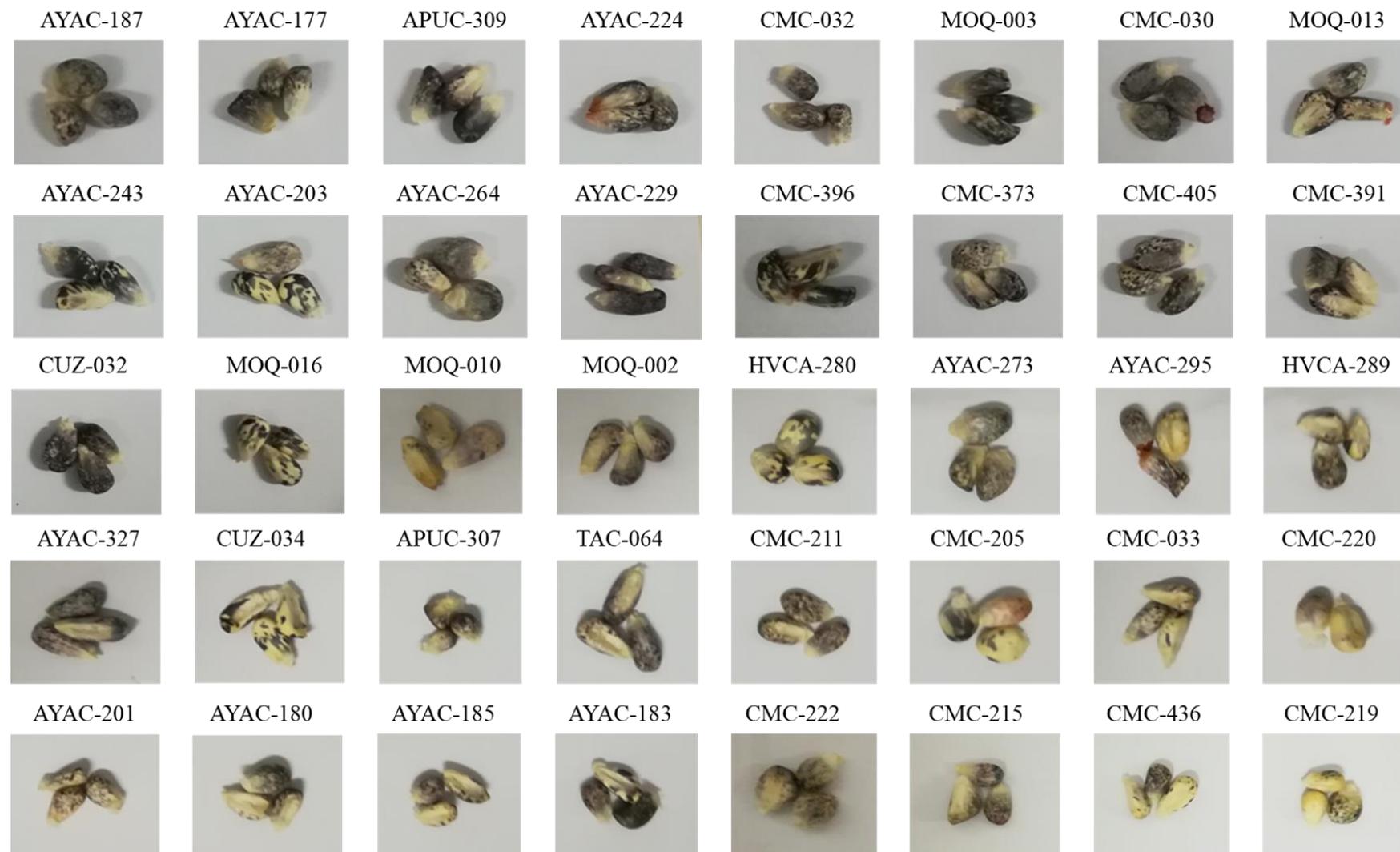
Semillas de la raza Chullpi empleadas para extraer ADN



Semillas de la raza Paro empleadas para extraer ADN



Semillas de la raza Piscorunto empleadas para extraer ADN



Anexo 3: Evidencia fotográfica de la cosecha proveniente de los experimentos en campo

Mazorcas de la cosecha de la raza Chullpi









APUC-320



APUC-320



AYA-271



AYA-271



CCC-041



CCC-041



JUN-523



JUN-523



CAJ-110



CAJ-110





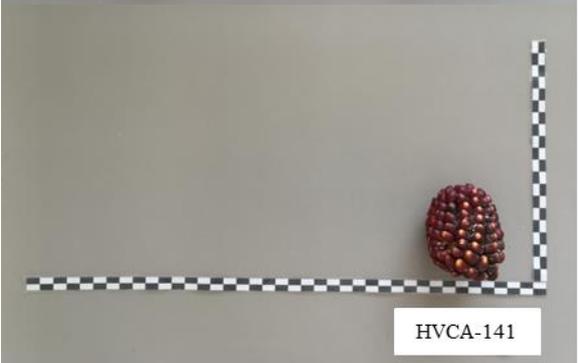


Mazorcas de la cosecha de la raza Paro



















Mazorcas de la cosecha de la raza Piscorunto















MOQ-002



CMC-211



CMC-211



AYA-264



AYA-264



AYA-327





APUC-309



AYA-224



AYA-201



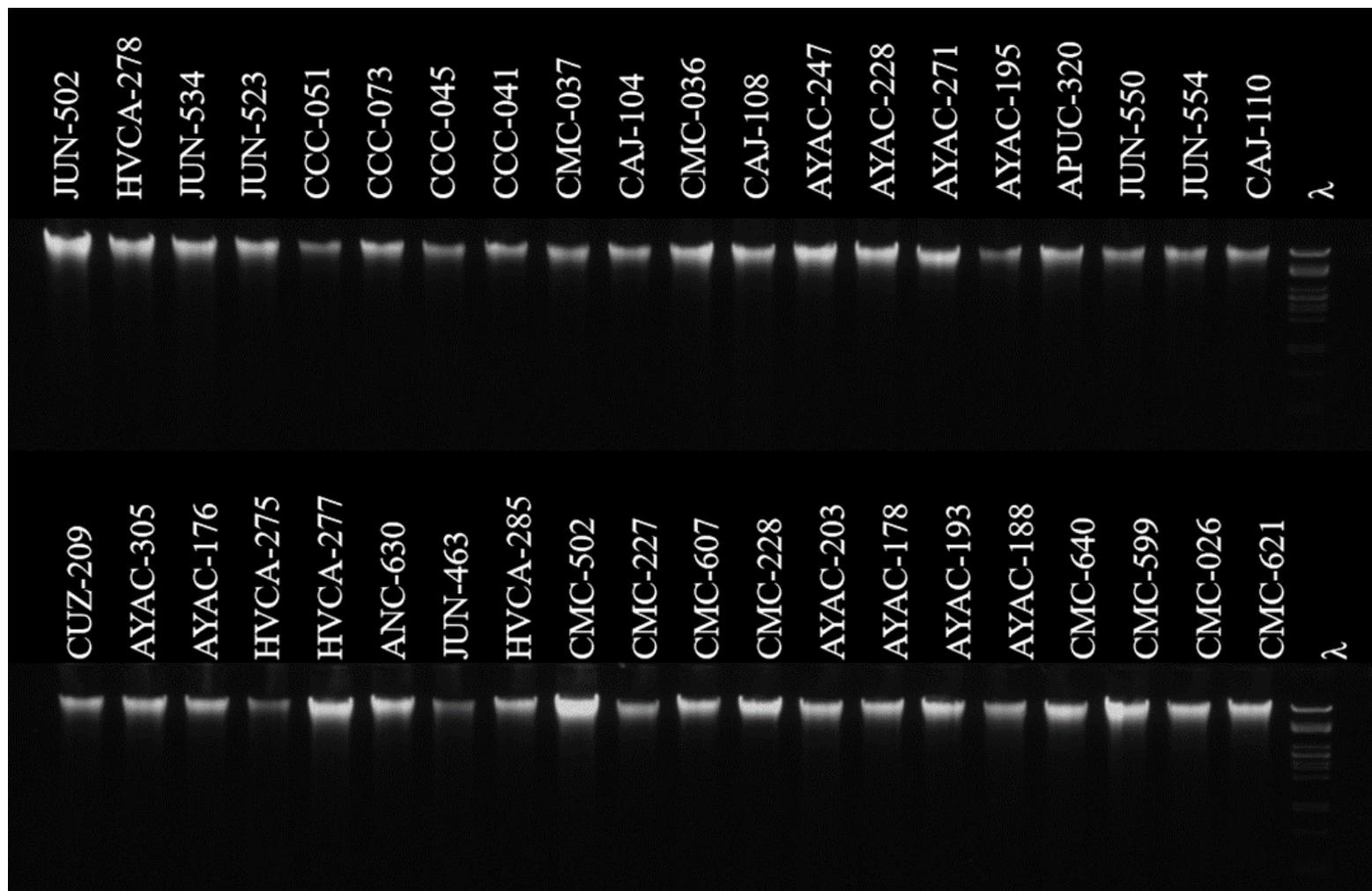
APUC-307



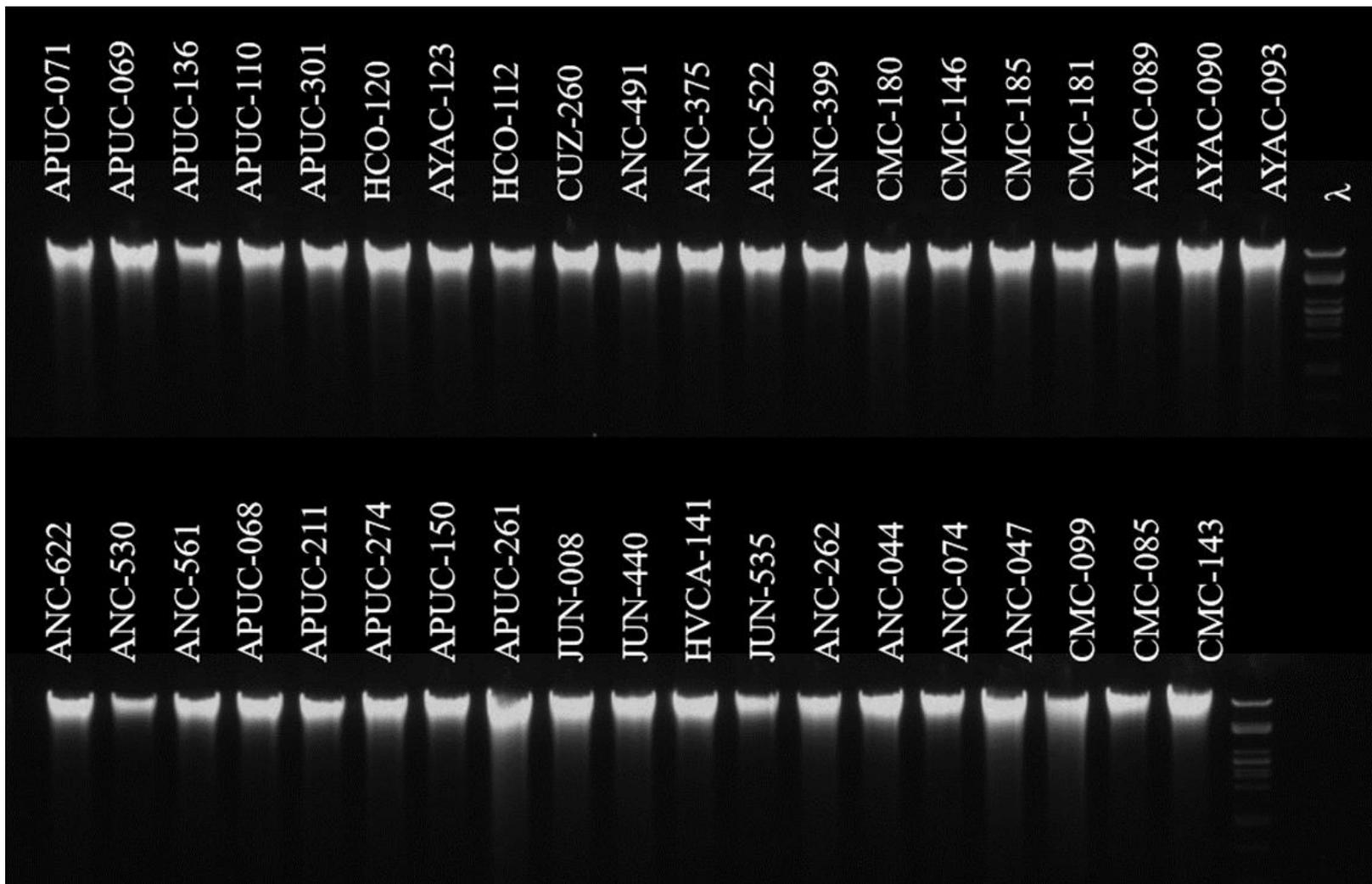
TAC-064

Anexo 4: Geles de calidad de ADN posterior a la extracción

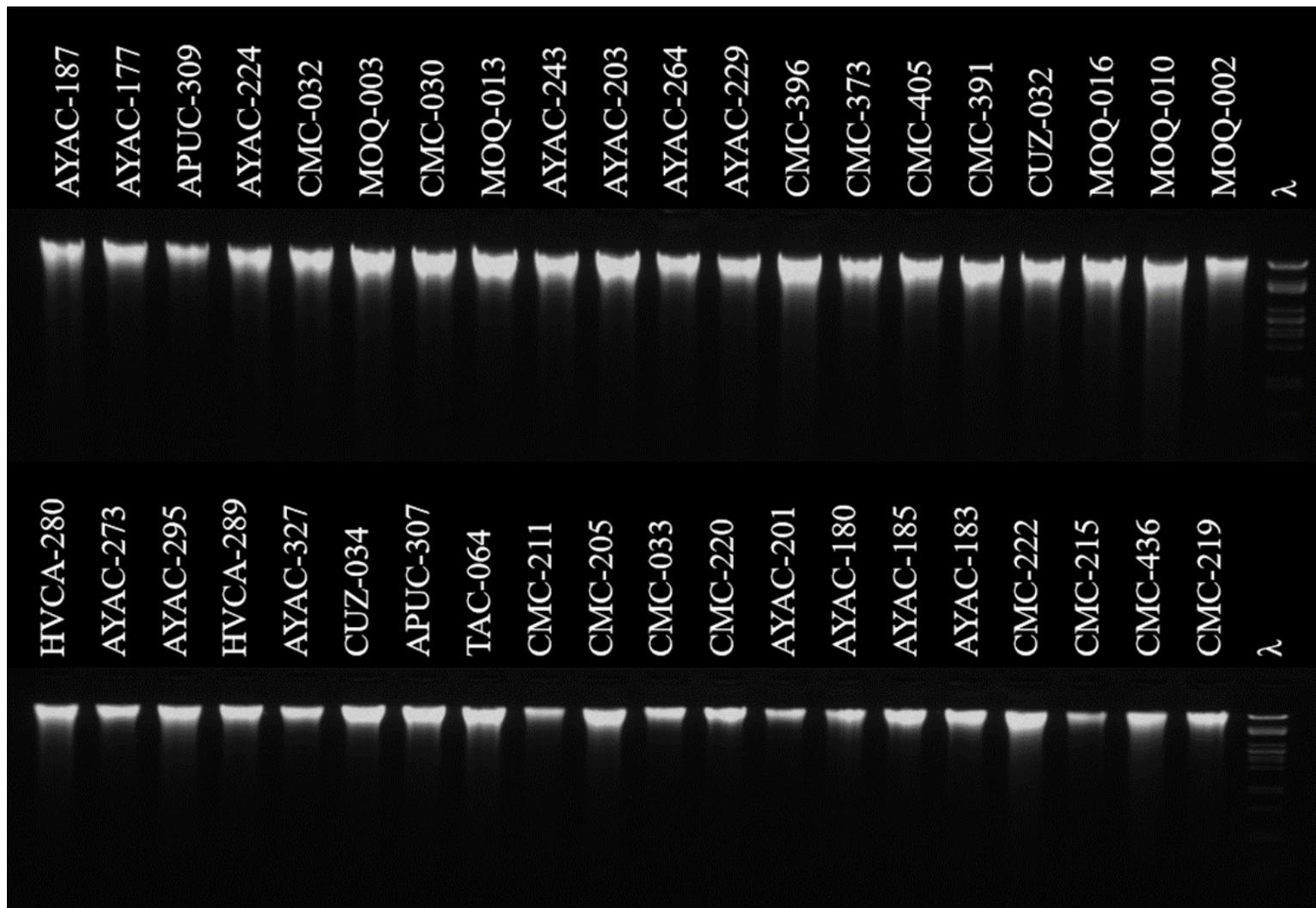
Gel de calidad de ADN de muestras de maíz de la raza Chullpi en agarosa al 1%. Primera muestra de la izquierda: marcador λ (0.28 μ g/ μ l)



Gel de calidad de ADN de muestras de maíz de la raza Paro en agarosa al 1%. Primera muestra de la izquierda: marcador λ (0.28 μ g/ μ l)



Gel de calidad de ADN de muestras de maíz de la raza Piscorunto en agarosa al 1%. Primera muestra de la izquierda: marcador λ (0.28 μ g/ μ l)



Anexo 5: Datos de cuantificación de ADN del material vegetal

Datos de cuantificación de ADN de la raza Chullpi

Accesión	Concentración por espectrometría en ng/ul	Coficiente 260/280	Coficiente 260/230	Concentración por fluorescencia en ng/ul
JUN-502	921.5	1.9	1.81	76.85
HVCA-278	763.2	1.89	1.69	38.81
JUN-534	857.6	1.89	1.75	92.21
JUN-523	759.4	1.88	1.72	77.41
CCC-051	333.9	1.76	1.12	22.81
CCC-073	472	1.83	1.43	67.76
CCC-045	645.5	1.85	1.5	65.29
CCC-041	588.8	1.83	1.38	58.13
CMC-037	517.9	1.85	1.46	66.40
CAJ-104	763.3	1.79	1.38	62.44
CMC-036	1016.7	1.87	1.51	82.79
CAJ-108	1166.2	1.85	1.6	87.89
AYAC-247	967	1.89	1.63	78.57
AYAC-228	813.7	1.89	1.64	81.06
AYAC-271	1022.8	1.85	1.48	68.90
AYA-195	570.7	1.83	1.32	59.12
APU-320	921.3	1.87	1.59	54.94
JUN-550	602.5	1.86	1.48	44.96
JUN-554	668.7	1.87	1.48	50.40
CAJ-110	472.4	1.85	1.46	49.73
CUZ-209	664.2	1.88	1.59	60.16
AYAC-305	813.5	1.89	1.68	79.30
AYAC-176	850.6	1.88	1.63	67.59
HVCA-275	501.3	1.85	1.47	38.14
HVCA-277	1575.2	1.88	1.77	105.61
ANC-630	1136.7	1.89	1.78	85.35
JUN-463	515.7	1.83	1.33	23.12
HVCA-285	1134.9	1.89	1.6	66.06
CMC-502	1257.3	1.91	1.84	114.42
CMC-227	646.6	1.87	1.58	54.81
CMC-607	1066.9	1.86	1.57	73.14
CMC-228	1033.4	1.87	1.64	100.37
AYAC-203	774.8	1.89	1.71	77.34
AYA-178	721.7	1.92	1.71	60.15
AYAC-193	941.5	1.88	1.7	68.21
AYAC-188	868.6	1.9	1.71	86.55
CMC-640	655.3	1.9	1.71	66.82
CMC-599	1393	1.89	1.82	120.24
CMC-026	1367.2	1.85	1.52	82.09

Datos de cuantificación de ADN de la raza Paro

Accesión	Concentración por espectrometría en ng/ul	Coficiente 260/280	Coficiente 260/230	Concentración por fluorescencia en ng/ul
APUC-069	3589	1.95	2.18	132.41
APUC-136	4122.2	1.96	2.19	159.55
APUC-110	3580.1	1.97	2.18	151.90
APUC-301	1048.1	1.96	2.13	34.19
HCO-120	5222.9	1.98	2.23	227.56
AYA-123	3651.6	1.97	2.21	160.42
HCO-112	2524.1	1.97	2.13	124.78
CUZ-260	3313.6	1.95	2.19	182.30
ANC-491	4434.8	1.96	2.19	181.80
ANC-375	3682.3	1.95	2.13	173.43
ANC-522	3833.1	1.96	2.18	171.72
ANC-399	4507.8	1.95	2.15	150.14
CMC-180	6069.3	1.98	2.23	254.53
CMC-146	3065.1	1.98	2.2	147.84
CMC-185	4194.4	1.97	2.2	189.07
CMC-181	2379.5	1.95	2.05	141.99
AYA-089	2724.8	1.97	2.14	172.71
AYA-090	5183.4	1.97	2.24	232.97
AYA-093	4038	1.96	2.23	223.26
ANC-622	3982.5	1.97	2.21	169.98
ANC-530	2014.3	1.96	2.1	98.54
ANC-561	3504	1.95	2.18	185.80
APUC-068	4252.4	1.95	2.21	192.41
APUC-211	4923	1.97	2.21	192.76
APUC-274	3388.8	1.95	2.19	129.55
APUC-150	3930.6	1.96	2.16	108.57
APUC-261	4235.1	1.97	2.21	114.46
JUN-008	5806.2	1.96	2.18	172.84
JUN-440	5606.9	1.97	2.21	99.59
HVCA-141	3997.2	1.94	2.18	94.54
JUN-535	3480.9	1.94	2.13	97.01
ANC-262	2727.1	1.97	2.13	123.82
ANC-044	4641.7	1.97	2.19	89.37
ANC-074	3477.4	1.96	2.17	125.34
ANC-047	5674	1.98	2.25	115.78
CMC-099	4251.1	1.97	2.18	254.53
CMC-085	4633.8	1.97	2.22	149.06
CMC-143	2657.3	1.97	2.13	120.89
CMC-090	4553.2	1.95	2.14	161.59

Datos de cuantificación de ADN de la raza Piscorunto

Accesión	Concentración por espectrometría en ng/ul	Coficiente 260/280	Coficiente 260/230	Concentración por fluorescencia en ng/ul
AYA-187	1048.9	1.93	2.02	99.48
AYA-177	1001.2	1.93	1.89	92.46
APUC-309	623.4	0.189	1.77	88.14
AYA-224	1935	1.92	2.01	175.88
CMC-032	2285.3	1.91	1.91	140.57
MOQ-003	2311.1	1.92	1.99	156.43
CMC-030	2234.5	1.92	1.92	206.22
MOQ-013	2633.8	1.93	2.02	166.20
AYA-243	2211.4	1.93	1.99	123.64
AYA-203	2737.6	1.92	1.96	115.28
AYA-264	1613.8	1.9	1.81	112.36
AYA-229	2392.7	1.93	1.93	133.43
CMC-396	3157.9	1.94	2.02	156.12
CMC-373	2477.3	1.91	1.91	131.34
CMC-405	2232.9	1.92	1.95	190.34
CMC-391	1999.6	1.92	2.01	154.55
CUZ-032	2097.3	1.91	1.93	116.72
MOQ-016	2638.2	1.93	1.93	114.03
MOQ-010	3038.4	1.97	2.13	163.77
MOQ-002	1563.4	1.86	1.5	91.80
HVCA-280	3291.9	1.9	1.96	131.59
AYA-273	1429.2	1.94	2.01	142.82
AYA-295	2464	1.95	2.08	187.87
HVCA-289	3395.7	1.93	1.99	154.59
AYA-327	1895.6	1.94	2.06	138.93
CUZ-034	1027.7	1.93	2.1	210.72
APUC-307	2653.7	1.93	1.96	177.79
TAC-064	2477.8	1.94	2.07	162.67
CMC-211	1172.6	1.91	1.67	83.50
CMC-205	2567.6	1.95	2.07	203.27
CMC-033	1395.1	1.91	1.8	126.70
CMC-220	1717.3	1.89	1.78	118.96
AYA-201	702.1	1.94	1.8	68.95
AYA-180	2075.9	1.93	1.85	134.07
AYA-185	2292.3	1.92	1.91	137.59
AYA-183	1814	1.93	1.89	98.61
CMC-222	3414.3	1.91	1.97	185.32
CMC-215	990.9	1.9	1.71	84.20
CMC-436	1293.3	1.9	1.82	143.77
CMC-219	1368.1	1.91	1.88	136.91

Anexo 6: Estadística descriptiva para las razas Chullpi, Paro y Piscorunto

Datos de la estadística descriptiva para la raza Chullpi

Carácter cuantitativo	Media	Mediana	Mínimo	Máximo	Desviación estándar	Coefficiente de variabilidad (%)
DFM	122.20	111	100	163	19.54	15.98
DFF	141.28	141	103	163	17.76	12.57
AP (cm)	191.94	190	98	295	35.26	18.31
AM (cm)	108.78	110	20	238	30.59	28.12
NHT	10.60	11	7	17	1.59	15.04
NHEM	4.66	5	2	9	0.9	19.36
LH (cm)	79.87	79	30	110	12.38	15.50
AH (cm)	7.41	7	2	13	1.57	21.21
LM (cm)	9.38	9	4.5	15	1.94	20.73
DAM (mm)	43.13	42.13	13.78	66.67	6.95	16.11
DCM (mm)	54.70	54	21.43	78.31	7.97	14.56
DBM (mm)	47.85	47.52	30.17	81.32	7.52	15.71
NH	17.78	16.50	8	35	5.27	29.66
NGH	18.78	18	10	30	4.12	21.93
DM (mm)	28.56	27.32	14.69	53.61	6.41	22.44
DR (mm)	19.20	18.08	8.10	35.29	5.34	27.81
DME (mm)	10.89	9.95	2.40	27.50	3.89	35.73
LG (mm)	15.59	15.63	4.34	24.25	2.57	16.50
AG (mm)	8.06	7.99	3.19	18.32	1.54	19.13
GG (mm)	4.66	4.52	2.18	9.74	0.96	20.56
PG (gr)	37.06	35	20	60	8.44	22.77

Datos de la estadística descriptiva para la raza Paro

Carácter cuantitativo	Media	Mediana	Mínimo	Máximo	Desviación estándar	Coefficiente de variabilidad (%)
DFM	104.61	102.00	99	135	6.63	6.34
DFF	111.62	107	105	161	14.40	12.9
AP (cm)	183.98	183	95	298	26.89	14.62

AM (cm)	95.73	95	15	212	25.87	27.02
NHT	9.32	9	7	15	1.22	13.15
NHEM	4.51	4	3	8	0.71	15.71
LH (cm)	76.55	76.50	45	116	11.06	14.45
AH (cm)	7.96	8	2	13	1.61	20.26
LM (cm)	10.47	10	5	22	2.73	26.04
DAM (mm)	35.42	35.71	15.49	56	6.56	18.53
DCM (mm)	47.87	48.23	20.24	96	6.39	13.36
DBM (mm)	3.39	43.47	22	62.42	6.77	15.60
NH	12.34	12	7	28	2.58	20.87
NGH	16.89	17	7	37	4.77	28.25
DM (mm)	22.87	22.56	12.61	49.88	4.59	20.08
DR (mm)	13.32	13.20	4.47	24.38	2.93	22.03
DME (mm)	6.80	6.61	1.43	17.53	2.12	31.13
LG (mm)	15.51	15.49	5.08	25.51	2.21	14.24
AG (mm)	9.75	9.68	4.49	19.90	1.56	16.07
GG (mm)	5.66	5.48	2.72	10.76	1.22	21.45
PG (gr)	52.00	53	31	73	9.79	18.83

Datos de la estadística descriptiva para la raza Piscorunto

Carácter cuantitativo	Media	Mediana	Mínimo	Máximo	Desviación estándar	Coefficiente de variabilidad (%)
DFM	112.47	105	94	135	12.5	11.12
DFE	130.02	129	100	163	15.07	11.59
AP (cm)	174.41	177	73	280	30.92	17.73
AM (cm)	86.03	84	15	180	28.62	33.27
NHT	8.39	8	6	13	1.03	12
NHEM	4.34	4	3	6	0.63	15.61
LH (cm)	69.55	70	38	105	10.66	15.33
AH (cm)	8.70	9	2	14	1.54	17.73
LM (cm)	9.07	9	4	15	1.88	20.69
DAM (mm)	34.74	34.58	18.15	53.56	4.97	14.32

DCM (mm)	50.71	50.94	24.89	77.74	5.62	11.09
DBM (mm)	42.47	41.77	26.54	64.74	6.44	15.16
NH	12.75	12	8	22	2.36	18.54
NGH	15.61	16	8	26	3.28	21.07
DM (mm)	23.66	23.54	14.87	49.93	3.64	15.40
DR (mm)	14.18	13.94	4.71	29.05	3.18	22.43
DME (mm)	7.10	7.07	2.46	21.21	2.08	29.25
LG (mm)	16.23	16.30	7.22	27.21	2.15	13.25
AG (mm)	9.45	9.53	1.22	15.55	1.57	16.62
GG (mm)	5.34	5.19	2.66	8.95	1.18	22.02
PG (gr)	51.45	51	31	68	8.21	15.95

Anexo 7: Datos completos del análisis de comparación de medianas de los caracteres cuantitativos de las razas Chullpi, Paro y Piscorunto

Prueba de normalidad de los datos a un nivel de significancia de 0.05

Prueba de Kolmogorov-Smirnov						
Carácter cuantitativo	Chllpi		Paro		Piscorunto	
	D	p-valor	D	p-valor	D	p-valor
DFM	0.243	< 2.2e-16	0.358	< 2.2e-16	0.459	< 2.2e-16
DFE	0.206	< 2.2e-16	0.299	< 2.2e-16	0.217	< 2.2e-16
AP (cm)	0.050	0.040	0.048	0.054	0.055	0.015
AM (cm)	0.047	0.060	0.054	0.020	0.029	0.490
NHT	0.171	< 2.2e-16	0.227	< 2.2e-16	0.257	< 2.2e-16
NHEM	0.245	< 2.2e-16	0.285	< 2.2e-16	0.355	< 2.2e-16
LH (cm)	0.039	0.171	0.051	0.029	0.036	0.261
AH (cm)	0.131	0.000	0.100	0.000	0.106	0.000
LM (cm)	0.107	0.000	0.095	0.000	0.093	0.000
DAM (mm)	0.069	0.001	0.034	0.300	0.041	0.361
DCM (mm)	0.047	0.058	0.044	0.086	0.062	0.039
DBM (mm)	0.041	0.138	0.028	0.572	0.075	0.006
NH	0.192	< 2.2e-16	0.184	< 2.2e-16	D = 0.21567	< 2.2e-16
NGH	0.128	0.000	0.099	0.000	0.131	0.000
DM (mm)	0.093	0.000	0.071	0.001	0.050	0.152
DR (mm)	0.091	0.000	0.044	0.091	0.065	0.024
DME (mm)	0.116	0.000	0.066	0.002	0.059	0.053
LG (mm)	0.028	0.544	0.045	0.080	0.547	0.061
AG (mm)	0.041	0.135	0.047	0.061	0.067	0.018
GG (mm)	0.070	0.001	0.077	0.000	0.065	0.024
PG (gr)	0.089	< 2.2e-16	0.072	0.001	0.104	0.000

Prueba de comparación de medianas de Wilcoxon – Mann – Whitney

Carácter: NHT

```
> print(resultado_Chullpi_vs_Paro)

Wilcoxon rank sum test with continuity correction
data: Chullpi and Paro
W = 482786, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0

> print(resultado_Chullpi_vs_Piscorunto)

Wilcoxon rank sum test with continuity correction
data: Chullpi and Piscorunto
W = 358171, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0

> print(resultado_Paro_vs_Piscorunto)

Wilcoxon rank sum test with continuity correction
data: Paro and Piscorunto
W = 271920, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0
```

Carácter: NHEM

```
> print(resultado_Chullpi_vs_Paro)

      Wilcoxon rank sum test with continuity correction
data:  Chullpi and Paro
W = 345457, p-value = 0.002712
alternative hypothesis: true location shift is not equal to 0

> print(resultado_Chullpi_vs_Piscorunto)

      Wilcoxon rank sum test with continuity correction
data:  Chullpi and Piscorunto
W = 244156, p-value = 3.538e-09
alternative hypothesis: true location shift is not equal to 0

> print(resultado_Paro_vs_Piscorunto)

      Wilcoxon rank sum test with continuity correction
data:  Paro and Piscorunto
W = 229311, p-value = 0.0003491
alternative hypothesis: true location shift is not equal to 0
```

Carácter: LH

```
> print(resultado_Chullpi_vs_Paro)

      Wilcoxon rank sum test with continuity correction
data:  Chullpi and Paro
W = 372678, p-value = 1.178e-08
alternative hypothesis: true location shift is not equal to 0

> print(resultado_Chullpi_vs_Piscorunto)

      Wilcoxon rank sum test with continuity correction
data:  Chullpi and Piscorunto
W = 287473, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0

> print(resultado_Paro_vs_Piscorunto)

      Wilcoxon rank sum test with continuity correction
data:  Paro and Piscorunto
W = 259644, p-value = 1.314e-14
alternative hypothesis: true location shift is not equal to 0
```

Carácter: AH

```
> print(resultado_Chullpi_vs_Paro)

      Wilcoxon rank sum test with continuity correction
data:  Chullpi and Paro
W = 257584, p-value = 8.669e-12
alternative hypothesis: true location shift is not equal to 0

> print(resultado_Chullpi_vs_Piscorunto)

      Wilcoxon rank sum test with continuity correction
data:  Chullpi and Piscorunto
W = 92421, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0

> print(resultado_Paro_vs_Piscorunto)

      Wilcoxon rank sum test with continuity correction
data:  Paro and Piscorunto
W = 130846, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0
```

Carácter: AP

```
> print(resultado_Chullpi_vs_Paro)

      Wilcoxon rank sum test with continuity correction
data:  Chullpi and Paro
W = 367068, p-value = 3.466e-07
alternative hypothesis: true location shift is not equal to 0

> print(resultado_Chullpi_vs_Piscorunto)

      Wilcoxon rank sum test with continuity correction
data:  Chullpi and Piscorunto
W = 235129, p-value = 4.611e-05
alternative hypothesis: true location shift is not equal to 0

> print(resultado_Paro_vs_Piscorunto)

      Wilcoxon rank sum test with continuity correction
data:  Paro and Piscorunto
W = 202282, p-value = 0.04313
alternative hypothesis: true location shift is not equal to 0
```

Carácter: AM

```
> print(resultado_Chullpi_vs_Paro)

      Wilcoxon rank sum test with continuity correction
data:  Chullpi and Paro
W = 403324, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0

> print(resultado_Chullpi_vs_Piscorunto)

      Wilcoxon rank sum test with continuity correction
data:  Chullpi and Piscorunto
W = 267681, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0

> print(resultado_Paro_vs_Piscorunto)

      Wilcoxon rank sum test with continuity correction
data:  Paro and Piscorunto
W = 215489, p-value = 0.02429
alternative hypothesis: true location shift is not equal to 0
```

Carácter: DFM

```
> print(resultado_Chullpi_vs_Paro)

      Wilcoxon rank sum test with continuity correction
data:  Chullpi and Paro
W = 506375, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0

> print(resultado_Chullpi_vs_Piscorunto)

      Wilcoxon rank sum test with continuity correction
data:  Chullpi and Piscorunto
W = 279670, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0

> print(resultado_Paro_vs_Piscorunto)

      Wilcoxon rank sum test with continuity correction
data:  Paro and Piscorunto
W = 139270, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0
```

Carácter: DFF

```
> print(resultado_Chullpi_vs_Paro)

      Wilcoxon rank sum test with continuity correction
data:  Chullpi and Paro
W = 477490, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0

> print(resultado_Chullpi_vs_Piscorunto)

      Wilcoxon rank sum test with continuity correction
data:  Chullpi and Piscorunto
W = 291600, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0

> print(resultado_Paro_vs_Piscorunto)

      Wilcoxon rank sum test with continuity correction
data:  Paro and Piscorunto
W = 62270, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0
```

Carácter: LG

```
> print(resultado_Chullpi_vs_Paro)

      Wilcoxon rank sum test with continuity correction
data:  Chullpi and Paro
W = 327709, p-value = 0.4042
alternative hypothesis: true location shift is not equal to 0

> print(resultado_Chullpi_vs_Piscorunto)

      Wilcoxon rank sum test with continuity correction
data:  Chullpi and Piscorunto
W = 174943, p-value = 1.35e-06
alternative hypothesis: true location shift is not equal to 0

> print(resultado_Paro_vs_Piscorunto)

      Wilcoxon rank sum test with continuity correction
data:  Paro and Piscorunto
W = 165041, p-value = 3.027e-10
alternative hypothesis: true location shift is not equal to 0
```

Carácter: AG

```
> print(resultado_Chullpi_vs_Paro)

      Wilcoxon rank sum test with continuity correction
data:  Chullpi and Paro
W = 127461, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0

> print(resultado_Chullpi_vs_Piscorunto)

      Wilcoxon rank sum test with continuity correction
data:  Chullpi and Piscorunto
W = 97070, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0

> print(resultado_Paro_vs_Piscorunto)

      Wilcoxon rank sum test with continuity correction
data:  Paro and Piscorunto
W = 222147, p-value = 0.03137
alternative hypothesis: true location shift is not equal to 0
```

Carácter: GG

```
> print(resultado_Chullpi_vs_Paro)

      Wilcoxon rank sum test with continuity correction
data:  Chullpi and Paro
W = 157052, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0

> print(resultado_Chullpi_vs_Piscorunto)

      Wilcoxon rank sum test with continuity correction
data:  Chullpi and Piscorunto
W = 132879, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0

> print(resultado_Paro_vs_Piscorunto)

      Wilcoxon rank sum test with continuity correction
data:  Paro and Piscorunto
W = 238981, p-value = 3.427e-06
alternative hypothesis: true location shift is not equal to 0
```

Carácter: LM

```
> print(resultado_Chullpi_vs_Paro)

      Wilcoxon rank sum test with continuity correction
data:  Chullpi and Paro
W = 240747, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0

> print(resultado_Chullpi_vs_Piscorunto)

      Wilcoxon rank sum test with continuity correction
data:  Chullpi and Piscorunto
W = 225367, p-value = 0.00814
alternative hypothesis: true location shift is not equal to 0

> print(resultado_Paro_vs_Piscorunto)

      Wilcoxon rank sum test with continuity correction
data:  Paro and Piscorunto
W = 275149, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0
```

Carácter: DAM

```
> print(resultado_Chullpi_vs_Paro)

      Wilcoxon rank sum test with continuity correction
data:  Chullpi and Paro
W = 511816, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0

> print(resultado_Chullpi_vs_Piscorunto)

      Wilcoxon rank sum test with continuity correction
data:  Chullpi and Piscorunto
W = 357381, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0

> print(resultado_Paro_vs_Piscorunto)

      Wilcoxon rank sum test with continuity correction
data:  Paro and Piscorunto
W = 226935, p-value = 0.004225
alternative hypothesis: true location shift is not equal to 0
```

Carácter: DCM

```
> print(resultado_Chullpi_vs_Paro)

      Wilcoxon rank sum test with continuity correction
data:  Chullpi and Paro
W = 486263, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0

> print(resultado_Chullpi_vs_Piscorunto)

      Wilcoxon rank sum test with continuity correction
data:  Chullpi and Piscorunto
W = 275133, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0

> print(resultado_Paro_vs_Piscorunto)

      Wilcoxon rank sum test with continuity correction
data:  Paro and Piscorunto
W = 150635, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0
```

Carácter: DBM

```
> print(resultado_Chullpi_vs_Paro)

      Wilcoxon rank sum test with continuity correction
data:  Chullpi and Paro
W = 423385, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0

> print(resultado_Chullpi_vs_Piscorunto)

      Wilcoxon rank sum test with continuity correction
data:  Chullpi and Piscorunto
W = 295743, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0

> print(resultado_Paro_vs_Piscorunto)

      Wilcoxon rank sum test with continuity correction
data:  Paro and Piscorunto
W = 230589, p-value = 0.0006699
alternative hypothesis: true location shift is not equal to 0
```

Carácter: NH

```
> print(resultado_Chullpi_vs_Paro)

      Wilcoxon rank sum test with continuity correction
data:  Chullpi and Paro
W = 542521, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0

> print(resultado_Chullpi_vs_Piscorunto)

      Wilcoxon rank sum test with continuity correction
data:  Chullpi and Piscorunto
W = 342959, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0

> print(resultado_Paro_vs_Piscorunto)

      Wilcoxon rank sum test with continuity correction
data:  Paro and Piscorunto
W = 186973, p-value = 0.001675
alternative hypothesis: true location shift is not equal to 0
```

Carácter: NGH

```
> print(resultado_Chullpi_vs_Paro)

      Wilcoxon rank sum test with continuity correction
data:  Chullpi and Paro
W = 397456, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0

> print(resultado_Chullpi_vs_Piscorunto)

      Wilcoxon rank sum test with continuity correction
data:  Chullpi and Piscorunto
W = 299319, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0

> print(resultado_Paro_vs_Piscorunto)

      Wilcoxon rank sum test with continuity correction
data:  Paro and Piscorunto
W = 240378, p-value = 1.011e-06
alternative hypothesis: true location shift is not equal to 0
```

Carácter: DM

```
> print(resultado_Chullpi_vs_Paro)

      Wilcoxon rank sum test with continuity correction
data:  Chullpi and Paro
W = 496038, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0

> print(resultado_Chullpi_vs_Piscorunto)

      Wilcoxon rank sum test with continuity correction
data:  Chullpi and Piscorunto
W = 309479, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0

> print(resultado_Paro_vs_Piscorunto)

      Wilcoxon rank sum test with continuity correction
data:  Paro and Piscorunto
W = 179643, p-value = 3.521e-05
alternative hypothesis: true location shift is not equal to 0
```

Carácter: DR

```
> print(resultado_Chullpi_vs_Paro)

      Wilcoxon rank sum test with continuity correction
data:  Chullpi and Paro
W = 542203, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0

> print(resultado_Chullpi_vs_Piscorunto)

      Wilcoxon rank sum test with continuity correction
data:  Chullpi and Piscorunto
W = 332284, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0

> print(resultado_Paro_vs_Piscorunto)

      Wilcoxon rank sum test with continuity correction
data:  Paro and Piscorunto
W = 177336, p-value = 7.528e-06
alternative hypothesis: true location shift is not equal to 0
```

Carácter: DME

```
> print(resultado_Chullpi_vs_Paro)

      Wilcoxon rank sum test with continuity correction
data:  Chullpi and Paro
W = 540508, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0

> print(resultado_Chullpi_vs_Piscorunto)

      Wilcoxon rank sum test with continuity correction
data:  Chullpi and Piscorunto
W = 344376, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0

> print(resultado_Paro_vs_Piscorunto)

      Wilcoxon rank sum test with continuity correction
data:  Paro and Piscorunto
W = 186341, p-value = 0.001657
alternative hypothesis: true location shift is not equal to 0
```

Carácter: PG

```
> print(resultado_Chullpi_vs_Paro)

      Wilcoxon rank sum test with continuity correction
data:  Chullpi and Paro
W = 80659, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0

> print(resultado_Chullpi_vs_Piscorunto)

      Wilcoxon rank sum test with continuity correction
data:  Chullpi and Piscorunto
W = 49210, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0

> print(resultado_Paro_vs_Piscorunto)

      Wilcoxon rank sum test with continuity correction
data:  Paro and Piscorunto
W = 213960, p-value = 0.3463
alternative hypothesis: true location shift is not equal to 0
```

Anexo 8: Coeficientes de correlación de Pearson y prueba de significancia de la correlación de las razas Chullpi, Paro y Piscorunto

	NHEM	NHT	LH	AH	AP	AM	DFM	DFF	LG	AG	GG	LM	DAM	DCM	DBM	NH	NGH	DM	DR	DME	PG
NHEM	1	0.74*	0.47*	-0.24	0.56*	0.47*	0.15	0.2	0.21	0.13	-0.14	0.1	0.34*	0.29	0.32*	0.26	0.19	0.31	0.32*	0.24	0
	NHT	1	0.82*	0.16	0.87*	0.88*	0.35	0.33*	0.14	-0.22	-0.24	0.25	0.43*	0.37*	0.35*	0.51*	0.43*	0.39*	0.5*	0.44*	-0.32*
		LH	1	0	0.84*	0.9*	0.42	0.36*	0.19	0.2	-0.27*	0.27	0.52*	0.48*	0.49*	0.55*	0.43*	0.41*	0.54*	0.51*	-0.35*
			AH	1	-0.24	0.14*	0.25	0.13	0.21	0.37*	0.4	0.26	-0.2	-0.12	0	-0.27	0	-0.17	-0.22*	-0.15	0.33*
				AP	1	0.95*	0.18	0.27	0.17	-0.12	-0.12*	0.24	0.41*	0.33*	0.35*	0.42*	0.34*	0.34*	0.44*	0.4*	-0.22
					AM	1	0.27	0.3	0.05	-0.14	-0.12	0.29	0.42*	0.37*	0.36*	0.5*	0.42*	0.42*	0.49*	0.44*	-0.3
						DFM	1	0.42	0	-0.17	-0.35*	0.19	0.13	0.16	0.14	0.31	0.38*	0.38	0.14	0.21	-0.39*
							DFF	1	0.15	0.2	0.4	0.14	0	0.19	0.55	0.21	0.25	0.25*	0	0	0*
								LG	1	0.6	0.5	-0.35	0.44*	0.52*	-0.14*	0.26	0	0*	0.33*	0.27*	0.13
									AG	1	0.47	0.32*	-0.32*	-0.32*	0.55*	-0.67*	-0.31	-0.31*	-0.55*	-0.59*	0.75*
										GG	1	-0.15	-0.2	-0.21	0	-0.45*	-0.46*	-0.46	-0.32*	-0.36*	0.69*
											LM	1	-0.1	0	0.84	0	0.61*	0	0	0	0
												DAM	1	0.87*	0.84*	0.71*	0.3*	0.79*	0.82*	0.78*	-0.29
													DCM	1	0.81*	0.79*	0.35*	0.83*	0.81*	0.78*	-0.31*
														DBM	0	0.54*	0.22	0.64*	0.68*	0.63*	0
															NH	1	0.52*	0.73*	0.82*	0.82*	-0.7*
																NGH	1	0.38*	0.41*	0.37*	-0.55*
																	DM	1	0.92*	0.9*	-0.37
																		DR	1	0.96*	-0.49*
																			DME	1	-0.52*
																				PG	1

Figura 23a. Coeficientes de Pearson (r-valor) de los caracteres cuantitativos medidas en la raza Chullpi.

*Las celdas en color verde tienen una correlación positiva desde considerable hasta perfecta y las celdas de color rojo tienen una correlación negativa considerable. *Correlación es significativa al $p \geq 0.01$.

	NHEM	NHT	LH	AH	AP	AM	DFM	DFF	LG	AG	GG	LM	DAM	DCM	DBM	NH	NGH	DM	DR	DME	PG
NHEM	1	0.59*	0.43*	0	0.41*	0.23	0.23	0.2	-0.21	-0.1	-0.07	0.37*	0.07	0.14	0	0.33*	0.44*	0.17	0	-0.18	-0.02
	NHT	1	0.85*	0.23	0.77*	0.8*	0.59*	0.51*	0.38*	-0.23	-0.28	0.09*	0.02	0.05	0	0.38*	0.62*	0.28	0.25	0	-0.19
		LH	1	0.34*	0.75*	0.79*	0.48*	0.46*	-0.19	0	-0.18	0.52*	0.12	0.1	0.1	0.22	0.55*	0.26	0.17	-0.12	-0.31
			AH	1	0	0.18	0.24	0.12	0*	0	-0.38*	0.22	0.11	0.12	0	-0.01	0.23	0.23	0.42*	0.26*	-0.3
				AP	1	0.89*	0.33*	0.39*	-0.35*	-0.19	-0.1	0.36*	0.11	0	0	0.23	0.42*	0.3	0.17	0	0
					AM	1	0.54*	0.55*	-0.42*	-0.25	-0.21	0.34*	0	-0.02	0	0.23	0.41*	0.25	0.21	0*	-0.12
						DFM	1	0.75*	-0.43*	-0.22	-0.36*	0.17	-0.38*	-0.23	-0.3	0.3	0.32*	0.01	0.21	0.03	-0.21
							DFF	1	-0.51*	-0.44*	-0.34*	0.24	-0.39*	-0.43*	-0.31	0.47*	0.25	-0.14	-0.01	-0.16	-0.37*
								LG	1	0.71*	0.4*	0	0.34*	0.45*	0.41*	-0.54*	-0.35*	0	0	0.01	0.02
									AG	1	0.2	0	0.27	0.39*	0.28*	-0.63*	0.16	-0.19	0.24	0.21	0.29
										GG	1	0	-0.1	-0.47	0	-0.31	-0.31	-0.24	-0.4*	-0.32*	0.38*
											LM	1	0.2	0*	0.36*	0.21	0.72*	0.16	0	0.1	-0.1
												DAM	1	0.7*	0.67*	0	0.11	0.67*	0.37*	0.19	0.18
													DCM	1	0.53*	-0.11	-0.09	0.55*	0.46*	0.22	0.19
														DBM	1	0	0.17	0.46*	0.33*	0.31*	0.24
															NH	1	0.28	0.13	0	0	-0.17
																NGH	1	0.2	0.15	-0.1	-0.13
																	DM	1	0.72*	0.48*	0.23
																		DR	1	0.78*	0.34
																			DME	1	0.12
																				PG	1

Figura 23b. Coeficientes de Pearson (r-valor) de los caracteres cuantitativos medidas en la raza Paro.

Las celdas en color verde tienen una correlación positiva desde considerable hasta perfecta y las celdas de color rojo tienen una correlación negativa considerable. *Correlación es significativa al $p \geq 0.01$

	NHEM	NHT	LH	AH	AP	AM	DFM	DFF	LG	AG	GG	LM	DAM	DCM	DBM	NH	NGH	DM	DR	DME	PG
NHEM	1	0.64*	0.39	0.3	0.15	0	0.17	0.11	0	-0.09	-0.38*	0.2	-0.11	0	0	0	0.35*	0.25	-0.12	0.09	-0.23
	NHT	1	0.74*	0.59*	0.68*	0.67*	0.24	0.24	0.1	0.12	-0.42*	0.44*	-0.23	-0.12	-0.13	0	0.55*	0	-0.24	-0.12	0
		LH	1	0.58*	0.81*	0.74*	0	0.21	0	0	-0.28	0.63*	0	-0.26	0.01	0	37*	0	-0.24	0.1	0
			AH	1	0.56*	0.59*	0	0	0.23	0.12	-0.3	0.33*	0.18	0	0.05	-0.14	0.22*	0	-0.09	-0.11	0.08
				AP	1	0.88*	0	0	0	0.31	-0.22	0.6*	0	-0.19	0.03	-0.22	0.37*	0	-0.16	-0.06	0.18
					AM	1	0.13	0.25	0	0.37*	-0.13	0.47*	-0.14	-0.24	0	-0.25	0.24	0	-0.15	-0.16	0.36*
						DFM	0	0.73*	-0.1	-0.31	0	-0.09	-0.44*	-0.22	-0.39*	0.15	0	-0.2	-0.22	-0.33*	-0.25
							DFF	1	-0.2	-0.22	0.09	0	-0.34*	-0.29	-0.33*	0	0	-0.36*	-0.15	-0.15	-0.25
								LG	1	0.08	-0.32	0.09	0.24	0.4*	0.36*	0.02	0.38*	-0.25	-0.15	-0.14	0.17
									AG	1	0.29	0.19	0	0	0.21	-0.57*	0*	0	0.18	0	0.75*
										GG	1	0.15	-0.12	-0.21	0	-0.45*	-0.58*	-0.11	0.2	0	0.25
											LM	1	0.22	-0.05	0.37*	0.1	0.6*	0.22	-0.2	-0.15	0.19
												DAM	1	0.44*	0.65*	0.14	0	0.27	0.25	0.18	0
													DCM	1	0.45*	0.43*	0	0.28	0.45*	0.51*	0
														DBM	1	0.02	0	0.28	0.33	0.33*	0.21
															NH	1	0.26*	0.19	0.25	0.34*	-0.53*
																NGH	1	0	-0.36*	-0.21	0
																	DM	1	0.21	0.33*	0.1
																		DR	1	0.75*	0.1
																			DME	1	-0.14
																				PG	1

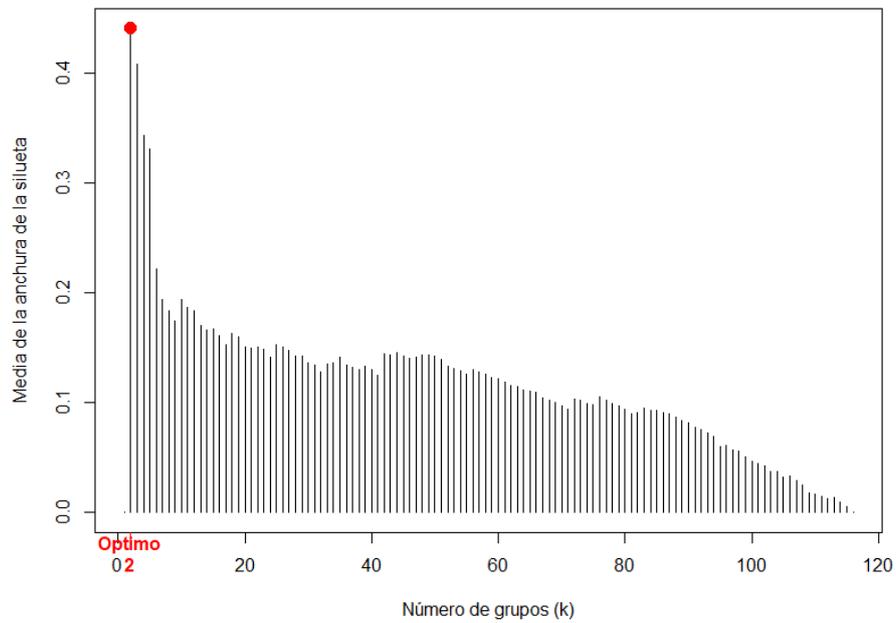
Figura 23b. Coeficientes de Pearson (r-valor) de los caracteres cuantitativos medidas en la raza Piscorunto.

*Las celdas en color verde tienen una correlación positiva desde considerable hasta perfecta y las celdas de color rojo tienen una correlación negativa considerable. *Correlación es significativa al $p \geq 0.01$.

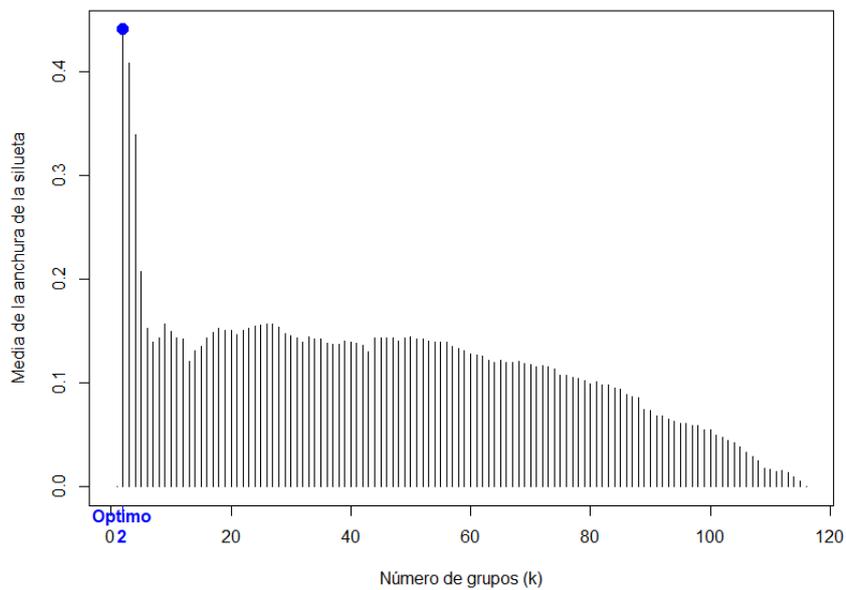
Anexo 9: Análisis multivariado de agrupamiento jerárquico empleando distancias genéticas

Anexo 9.1 Gráficos de las pruebas para la selección del número de grupos óptimos para combinación de distancia Gower con agrupamiento UPGMA, complete, Ward, Single.

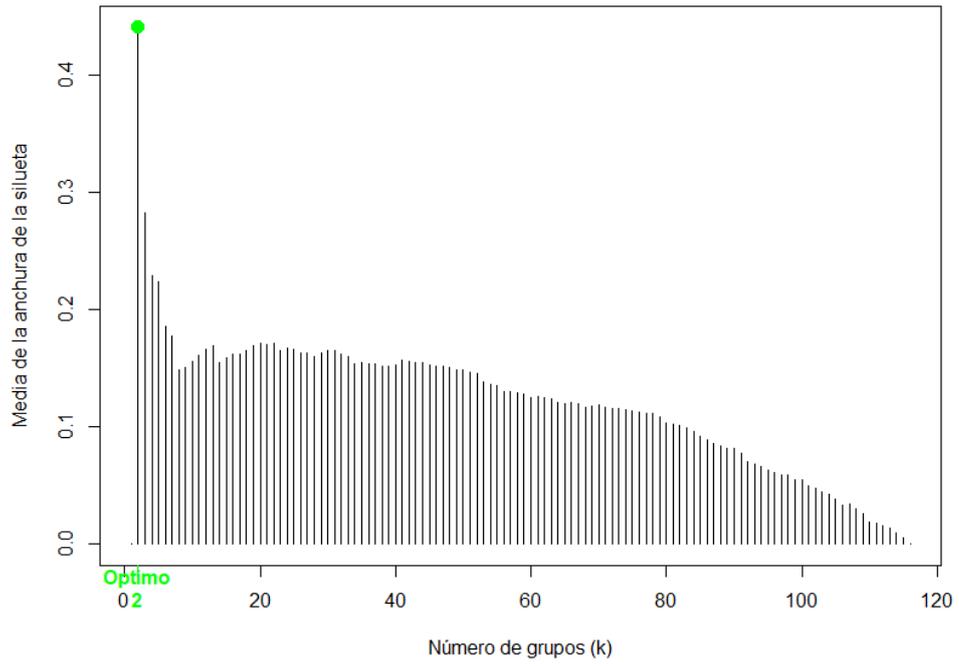
Número de grupos óptimo para dist. Gower y agrupamiento UPGMA



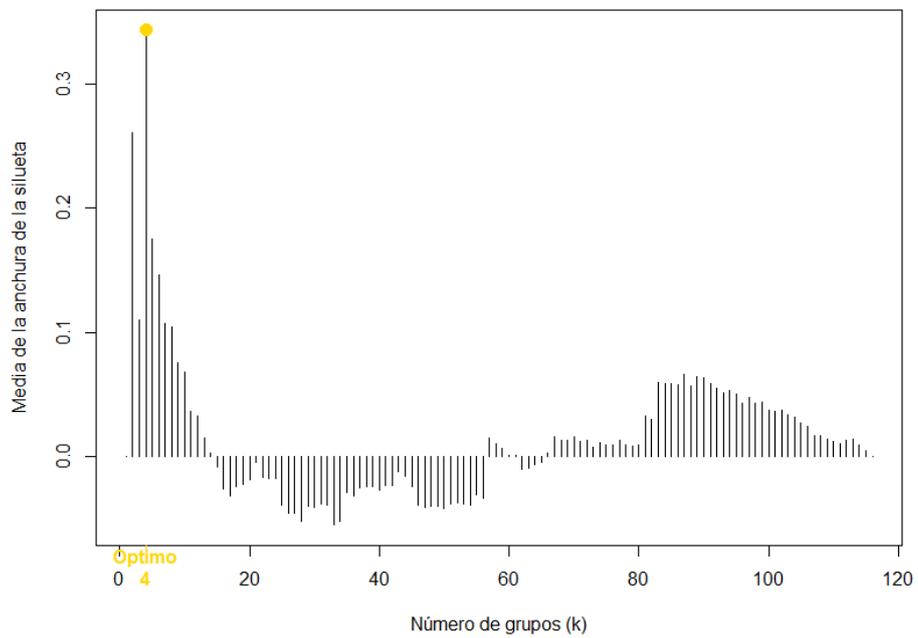
Número de grupos óptimo para dist. Gower y agrupamiento COMPLETE



Número de grupos óptimo para dist. Gower y agrupamiento WARD



Número de grupos óptimo para dist. Gower y agrupamiento WARD



Anexo 10: Análisis multivariado de agrupamiento jerárquico empleando componentes principales

Anexo 10.1 Tablas de los valores propios de las componentes principales para PCA, MCA, FAMD y MFA.

Componentes Principales del PCA	Valor propio	Varianza porcentual	Varianza porcentual acumulada
Dim.1	9.703	46.207	46.207
Dim.2	2.922	13.913	60.120
Dim.3	2.008	9.562	69.682
Dim.4	1.360	6.477	76.159
Dim.5	0.989	4.709	80.868
Dim.6	0.735	3.502	84.370
Dim.7	0.598	2.847	87.217
Dim.8	0.507	2.415	89.632
Dim.9	0.353	1.682	91.314
Dim.10	0.346	1.648	92.962
Dim.11	0.333	1.584	94.546
Dim.12	0.235	1.118	95.664
Dim.13	0.198	0.945	96.608
Dim.14	0.148	0.706	97.314
Dim.15	0.145	0.689	98.003
Dim.16	0.121	0.574	98.577
Dim.17	0.099	0.470	99.048
Dim.18	0.075	0.359	99.406
Dim.19	0.060	0.285	99.691
Dim.20	0.036	0.173	99.865
Dim.21	0.028	0.135	100.000

Componentes principales del MCA	Valor propio	Varianza porcentual	Varianza porcentual acumulada
Dim.1	0.44	15.22	15.225
Dim.2	0.24	8.18	23.405
Dim.3	0.20	6.99	30.395
Dim.4	0.18	6.28	36.680
Dim.5	0.17	5.87	42.551
Dim.6	0.15	5.36	47.916
Dim.7	0.15	5.09	53.004
Dim.8	0.14	4.73	57.730
Dim.9	0.12	4.30	62.026
Dim.10	0.12	4.19	66.221
Dim.11	0.11	3.86	70.084

Dim.12	0.11	3.82	73.900
Dim.13	0.10	3.60	77.501
Dim.14	0.09	3.26	80.764
Dim.15	0.08	2.94	83.705
Dim.16	0.08	2.83	86.537
Dim.17	0.08	2.64	89.180
Dim.18	0.07	2.51	91.695
Dim.19	0.06	2.06	93.758
Dim.20	0.05	1.63	95.384
Dim.21	0.04	1.40	96.783
Dim.22	0.04	1.23	98.011
Dim.23	0.03	1.08	99.093
Dim.24	0.02	0.85	99.939
Dim.25	0.00	0.06	100.000
Dim.26	0.00	0.00	100.000

Componentes principales del FAMD	Valor propio	Varianza porcentual	Varianza porcentual acumulada
Dim.1	12.563	26.730	26.730
Dim.2	3.916	8.333	35.063
Dim.3	2.488	5.295	40.357
Dim.4	2.335	4.968	45.325
Dim.5	1.996	4.247	49.572
Dim.6	1.855	3.946	53.519
Dim.7	1.654	3.520	57.039
Dim.8	1.480	3.148	60.187
Dim.9	1.393	2.963	63.150
Dim.10	1.321	2.810	65.960
Dim.11	1.192	2.535	68.495
Dim.12	1.128	2.401	70.896
Dim.13	1.115	2.373	73.269
Dim.14	1.062	2.259	75.528
Dim.15	0.961	2.045	77.573
Dim.16	0.902	1.920	79.493
Dim.17	0.850	1.808	81.301
Dim.18	0.828	1.762	83.063
Dim.19	0.800	1.703	84.766
Dim.20	0.711	1.513	86.279
Dim.21	0.684	1.455	87.734
Dim.22	0.660	1.403	89.138
Dim.23	0.590	1.254	90.392
Dim.24	0.535	1.139	91.530
Dim.25	0.464	0.988	92.518
Dim.26	0.442	0.940	93.459

Dim.27	0.346	0.737	94.196
Dim.28	0.339	0.720	94.916
Dim.29	0.309	0.658	95.574
Dim.30	0.286	0.608	96.182
Dim.31	0.268	0.570	96.752
Dim.32	0.243	0.518	97.269
Dim.33	0.207	0.441	97.711
Dim.34	0.184	0.391	98.102
Dim.35	0.167	0.356	98.458
Dim.36	0.141	0.301	98.759
Dim.37	0.129	0.276	99.035
Dim.38	0.105	0.224	99.259
Dim.39	0.090	0.191	99.450
Dim.40	0.063	0.133	99.583
Dim.41	0.056	0.118	99.701
Dim.42	0.049	0.103	99.804
Dim.43	0.035	0.075	99.880
Dim.44	0.027	0.057	99.936
Dim.45	0.018	0.039	99.975
Dim.46	0.012	0.025	100.000

Componentes principales del MFA	Valor propio	Varianza porcentual	Varianza porcentual acumulada
Dim.1	3.57	19.49	19.485
Dim.2	1.41	7.70	27.181
Dim.3	1.13	6.17	33.354
Dim.4	1.00	5.44	38.792
Dim.5	0.97	5.30	44.088
Dim.6	0.89	4.85	48.935
Dim.7	0.86	4.70	53.634
Dim.8	0.81	4.44	58.073
Dim.9	0.80	4.36	62.436
Dim.10	0.68	3.73	66.167
Dim.11	0.60	3.28	69.450
Dim.12	0.58	3.16	72.607
Dim.13	0.53	2.91	75.519
Dim.14	0.50	2.72	78.235
Dim.15	0.40	2.17	80.401
Dim.16	0.38	2.07	82.475
Dim.17	0.32	1.76	84.239
Dim.18	0.28	1.51	85.745
Dim.19	0.27	1.49	87.234
Dim.20	0.26	1.42	88.658
Dim.21	0.23	1.27	89.932

Dim.22	0.23	1.26	91.188
Dim.23	0.22	1.20	92.386
Dim.24	0.19	1.04	93.424
Dim.25	0.16	0.85	94.277
Dim.26	0.13	0.73	95.011
Dim.27	0.12	0.63	95.640
Dim.28	0.11	0.60	96.237
Dim.29	0.11	0.58	96.815
Dim.30	0.10	0.55	97.365
Dim.31	0.08	0.44	97.802
Dim.32	0.07	0.38	98.185
Dim.33	0.06	0.34	98.522
Dim.34	0.06	0.31	98.837
Dim.35	0.05	0.26	99.100
Dim.36	0.04	0.24	99.336
Dim.37	0.03	0.14	99.477
Dim.38	0.02	0.12	99.596
Dim.39	0.02	0.10	99.692
Dim.40	0.01	0.08	99.770
Dim.41	0.01	0.06	99.832
Dim.42	0.01	0.05	99.886
Dim.43	0.01	0.05	99.932
Dim.44	0.00	0.03	99.959
Dim.45	0.00	0.02	99.983
Dim.46	0.00	0.02	100.000

Anexo 10.2. Coeficientes de correlación y de determinación (R^2) de cada carácter con las primeras cuatro componentes principales del PCA y MCA.

Los coeficientes son significativos a un $p \geq 0.01$

Primer componente principal del PCA

Caracteres Cuantitativos	Correlación	p-valor
NH	0.901	2.70E-45
DR	0.874	1.56E-37
DME	0.876	4.35E-37
NHT	0.852	7.65E-34
DM	0.831	7.09E-31
DAM	0.823	9.04E-30
DCM	0.713	2.58E-19
LH	0.684	2.72E-17
AM	0.663	5.18E-16
DBM	0.646	4.83E-15
NGH	0.645	5.82E-15
NHEM	0.576	1.31E-11

AP	0.571	2.13E-11
DFF	0.568	2.83E-11
DFM	0.556	9.27E-11
AH	-0.421	2.46E-06
AG	-0.702	1.58E-18
GG	-0.706	8.89E-19
PG	-0.740	2.54E-21

Segundo componente principal del PCA

Caracteres Cuantitativos	Correlación	p-valor
LM	0.745	8.50E-22
AP	0.628	4.29E-14
LH	0.613	2.71E-13
AM	0.578	1.10E-11
NGH	0.417	3.22E-06
NHT	0.410	4.77E-06
NHEM	0.358	7.99E-05
AG	0.280	2.38E-03
NH	-0.203	2.87E-02
DAM	-0.209	2.45E-02
DM	-0.244	8.23E-03
DR	-0.295	1.30E-03
DME	-0.303	9.44E-04
LG	-0.326	3.57E-04
DCM	-0.444	5.78E-07

Tercer componente principal del PCA

Caracteres Cuantitativos	Correlación	p-valor
LG	0.623	8.02E-14
DBM	0.575	1.44E-11
AG	0.460	2.10E-07
PG	0.421	2.46E-06
DCM	0.377	3.03E-05
DAM	0.340	2.07E-04
GG	0.275	2.85E-03
LM	0.241	9.04E-03
DM	0.224	1.55E-02
DFF	-0.421	2.57E-06
DFM	-0.422	2.31E-06

Cuarto componente principal del PCA

Caracteres Cuantitativos	Correlación	p-valor
AH	0.706	8.84E-19
DFE	0.480	5.15E-08
DFM	0.447	5.05E-07
LG	0.438	8.94E-07
AM	0.221	1.69E-02

Primer componente principal del MCA

Caracteres cualitativos	R ²	p-valor
CE	0.931	5.93E-68
TG	0.931	5.93E-68
FSG	0.946	6.60E-68
CG	0.708	2.70E-25
CT	0.106	1.78E-03
CM	0.109	4.80E-03
FM	0.105	6.02E-03
DHG	0.061	7.33E-03
PT	0.063	2.60E-02

Segundo componente principal del MCA

Caracteres cualitativos	R ²	p-valor
FSG	0.652	1.06E-23
CG	0.677	5.13E-23
CM	0.447	2.23E-14
PT	0.224	6.09E-07

Tercer componente principal del MCA

Caracteres cualitativos	R ²	p-valor
FSG	0.646	2.62E-23
CG	0.598	9.39E-18
CM	0.339	4.42E-10
DHG	0.071	3.86E-03
PT	0.090	4.87E-03
CT	0.078	1.00E-02

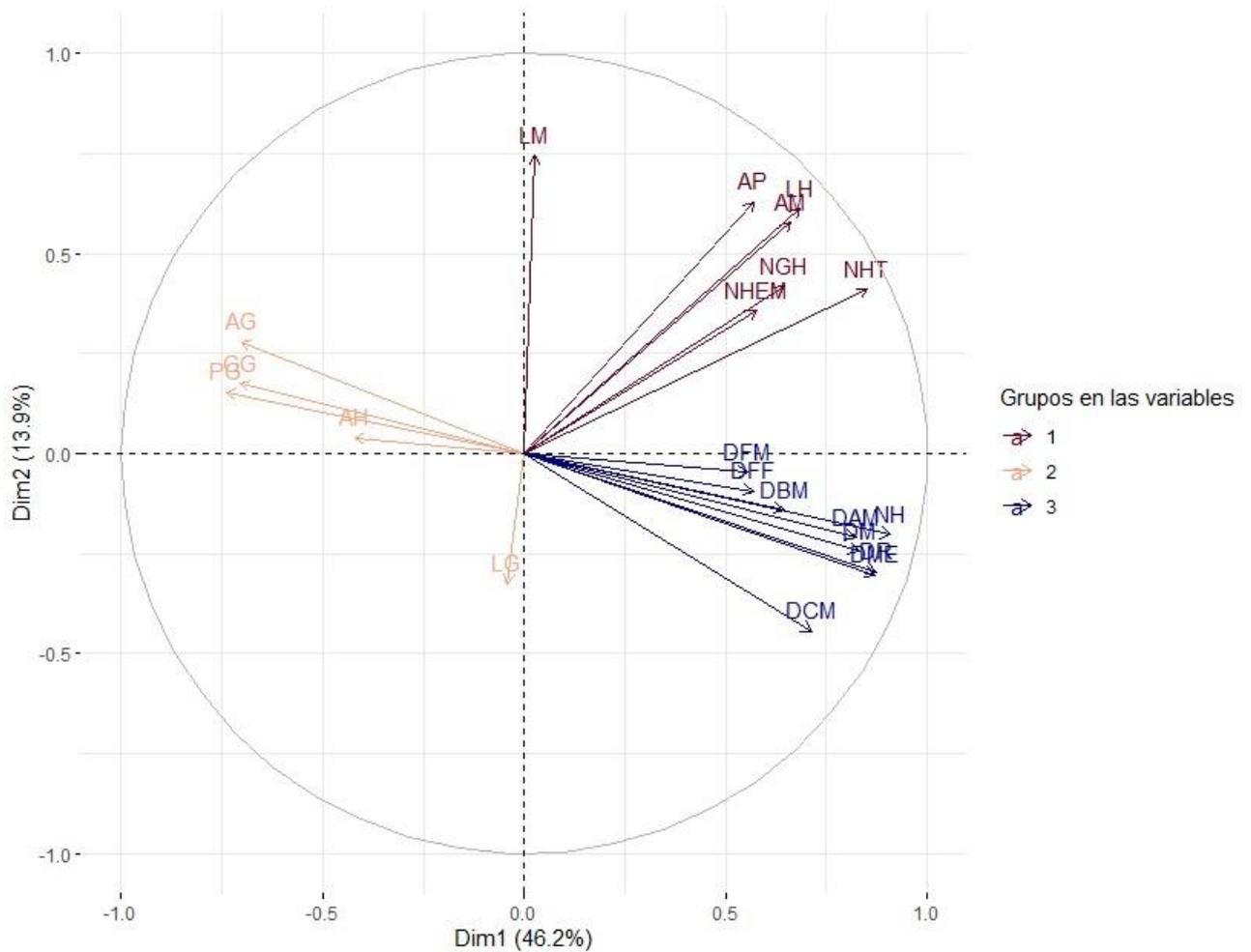
Cuarto componente principal del MCA

Caracteres cualitativos	R ²	p-valor
FSG	0.533	8.20E-17
CG	0.536	6.44E-15
FM	0.215	5.12E-06

CM	0.169	1.12E-04
CT	0.092	4.30E-03
PT	0.085	6.82E-03

Anexo 10.3 Gráfico “Biplot” de los caracteres cuantitativos en las dos primeras componentes principales del PCA

Los caracteres del mismo color indican la pertenencia a un grupo K-means.



Anexo 10.4 Tablas de accesiones duplicadas bajo el análisis HCPC con las CP del PCA, MCA y FAMD

HCPC-PCA

Paro	Piscorunto	Chullpi	Paro - Piscorunto	Piscorunto - Chullpi	Paro - Chullpi
APUC-301/CMC-181	AYA-327/MOQ-013	AYA-305/CMC-599	HVCA-141/AYA-180	HVCA-277/CMC-032	
APUC-069/CMC-185	AYA-295/CMC-219/AYA-224	AYA-178/AYA-195	AYA-299/APUC-110/ANC-522	MOQ-003/HVCA-285	
CMC-146/CMC-085	AYA-243/AYA-177	AYAC-203/AYA-188	AYA-090/AYA-203		
ANC-262/AYA-123	TAC-064/AYA-185-AYA-201	AYA-193/AYA-247			
CMC-143/APUC-136	CMC-220/CMC236/CMC396	CMC-502/CMC-640			
CMC-099/APUC-274	CMC-222/CMC-211/CMC-373/CMC-405	CAJ-108/AYA-271			
CMC-090/CMC-180	MOQ-002/AYA-273	JUN-534/JUN-502			
JUN-008/ANC-373		HVCA278/CAJ-110			
ANC-074/APUC-068					
APUC-261/AYA-093/APUC-071					
ANC-399/APUC-211					
ANC-491/APUC-150					
ANC-561/ANC-044					

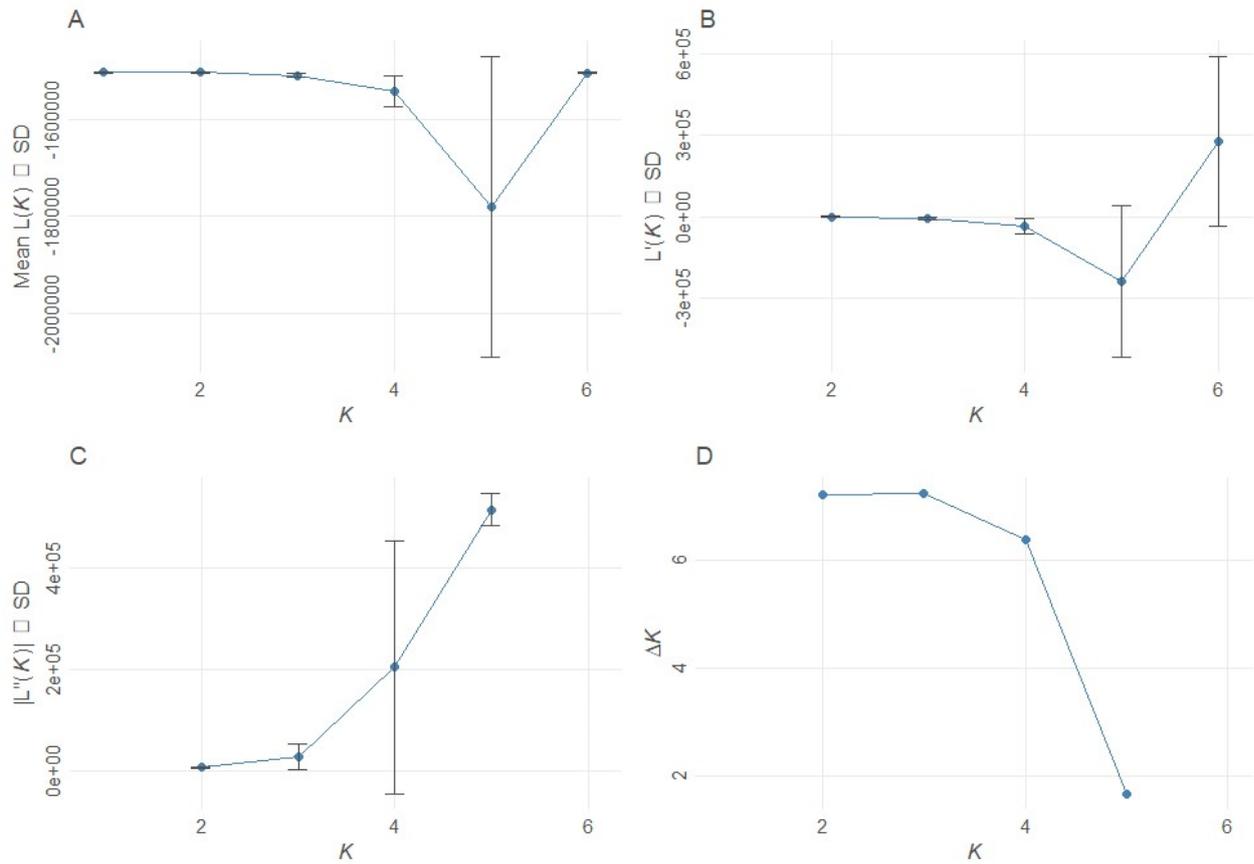
HCPC-MCA

Paro	Piscorunto	Chullpi	Paro - Piscorunto	Piscorunto - Chullpi	Paro - Chullpi
ANC-110/HCO-112	AYA-201/CMC-391	CAJ-108/CMC-037	AYA-093/CMC-033		AYA-090/CMC-228
ANC-047/ANC-503	AYA-187/APUC-307/AYA-185/CUZ-034/CMC-436/CMC-396/CMC-222/CMC-220/CMC-211/AYA-273/CMC-205	JUN-550/JUN-534/JUN-523/HVCA-285/CCC-041/HVCA-275			
JUN-440/APUC-261					
APUC-071/AYA-123/APUC-274		AYA-305/HVCA-278			
	AYA-295/CMC-405	AYA-178/JUN-502/CMC-640/CMC-599/CMC-502/CMC-227/CAJ-110/AYAC-203/AYA-195/AYA-1887AYA-193			
	MOQ-003/MOQ-002/CMC-373/CMC-219/CMC-030/AYA-327/AYA-229/AYA-264				
	CMC-032/CMC-215	ANC-630/AYA-176/APUC-320/JUN-463/CUZ-209/CMC-607/CMC-036/CCC-051/CAL-104/AYA-271/AYA-228/AYA-247			

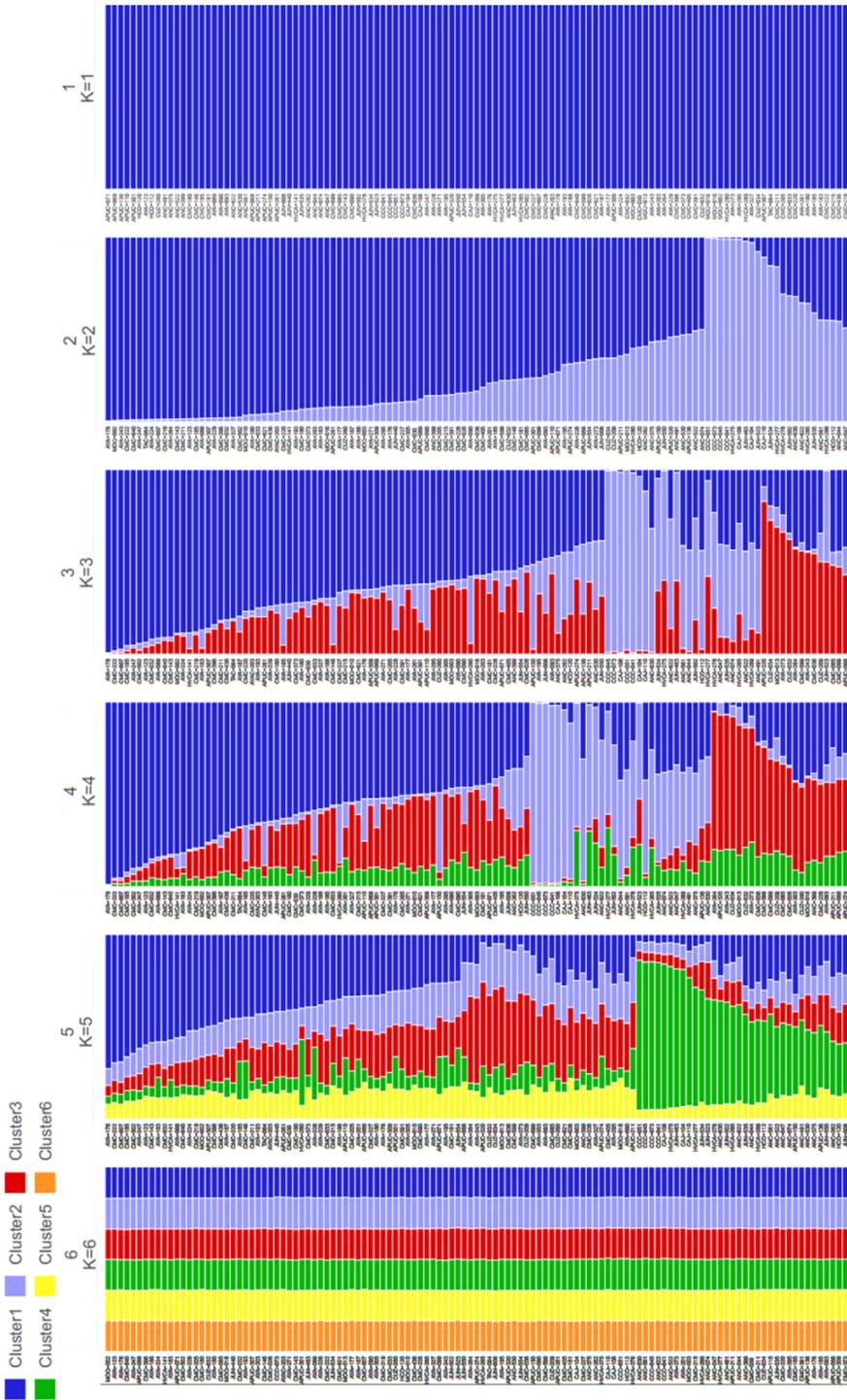
HCPC-FAMD

Paro	Piscorunto	Chullpi	Paro - Piscorunto	Piscorunto - Chullpi	Paro - Chullpi
APUC- 071/AYA-123	CMC-211/CMC- 373	CAJ-104/AYA- 271			
	CMC-396/CMC- 222	AYA- 178/AYA-195			
	CMC-436/CMC- 220	AYAC- 203/AYA- 193/AYA-188			
	CMC-219/AYA- 229	CMC- 502/CMC-640			
	AYA-327/MOQ- 003				
	AYA-185/AYA- 273				

Anexo 11: Información del Δk para el análisis de estructura genética con el programa STRUCTURE desde $k=1$ hasta $k=6$



Anexo 12: Diagramas de barras multilínea de probabilidades de membresía a cada grupo (cluster) con etiquetado por accesión de k=1 a k=6



Anexo 13: Tabla de probabilidades de membresía (valor q) a los grupos (clusters) generados con el programa STRUCTURE

Accesión	Probabilidad de membresía			Grupo con PM ≥ 0.5
	G1	G2	G3	
ANC-044	0.5498	0.4373	0.0128	SG1
ANC-047	0.5011	0.4065	0.0925	SG1
ANC-074	0.4963	0.4357	0.0679	M
ANC-262	0.3575	0.4501	0.1924	SG2
ANC-375	0.4184	0.469	0.1128	SG2
ANC-399	0.0418	0.5497	0.4084	SG2
ANC-491	0.4508	0.4351	0.1142	SG2
ANC-522	0.48	0.4431	0.077	M
ANC-530	0.3678	0.3889	0.2435	M
ANC-561	0.5574	0.4128	0.0297	SG1
ANC-622	0.5918	0.2697	0.1383	SG1
ANC-630	0.6702	0.3204	0.0092	SG1
APUC-068	0.1479	0.4228	0.4294	M
APUC-069	0.0514	0.6466	0.3022	SG2
APUC-071	0.2171	0.5524	0.2306	SG2
APUC-110	0.2468	0.6203	0.1332	SG2
APUC-136	0.3784	0.4054	0.2162	SG2
APUC-150	0.4592	0.5156	0.0252	SG2
APUC-211	0.2142	0.3958	0.3901	M
APUC-261	0.0508	0.7475	0.202	SG2
APUC-274	0.176	0.4167	0.4072	M
APUC-301	0.2035	0.6219	0.1747	SG2
APUC-307	0.0213	0.8612	0.1174	SG2
APUC-309	0.0393	0.6498	0.3109	SG2
APUC-320	0.0797	0.0934	0.8268	SG3
AYA-089	0.0269	0.9146	0.0585	SG2
AYA-090	0.0511	0.5992	0.3497	SG2
AYA-093	0.0876	0.4766	0.4359	M
AYA-123	0.0246	0.9376	0.0378	SG2
AYA-176	0.0446	0.6513	0.3044	SG2
AYA-177	0.022	0.6253	0.3527	SG2
AYA-178	0.0035	0.994	0.0026	SG2
AYA-180	0.0255	0.7205	0.2539	SG2
AYA-183	0.0995	0.8698	0.0305	SG2
AYA-185	0.2207	0.7307	0.0485	SG2
AYA-187	0.084	0.7963	0.1197	SG2

AYA-188	0.0356	0.7001	0.2644	SG2
AYA-193	0.0478	0.7513	0.2011	SG2
AYA-195	0.1732	0.5001	0.3268	SG2
AYA-201	0.1276	0.6234	0.2492	SG2
AYA-203	0.0181	0.5745	0.4073	SG2
AYA-224	0.0154	0.8877	0.097	SG2
AYA-228	0.2382	0.6283	0.1337	SG2
AYA-229	0.0188	0.7415	0.2398	SG2
AYA-243	0.0073	0.4412	0.5514	SG3
AYA-247	0.0145	0.9533	0.0323	SG2
AYA-264	0.0085	0.4215	0.57	SG3
AYA-271	0.0244	0.6388	0.3369	SG2
AYA-273	0.0929	0.2466	0.6604	SG3
AYA-295	0.0296	0.618	0.3523	SG2
AYA-305	0.0303	0.6041	0.3657	SG2
AYA-327	0.0152	0.7004	0.2845	SG2
AYAC-203	0.0298	0.767	0.2033	SG2
CAJ-104	0.9437	0.0376	0.0187	SG1
CAJ-108	0.9788	0.0154	0.0058	SG1
CAJ-110	0.8648	0.1173	0.0177	SG1
CCC-041	0.9758	0.0178	0.0063	SG1
CCC-045	0.9838	0.0079	0.0084	SG1
CCC-051	0.9772	0.0021	0.0208	SG1
CCC-073	0.9826	0.0051	0.0125	SG1
CMC-026	0.033	0.5224	0.4445	SG2
CMC-030	0.0809	0.7178	0.2013	SG2
CMC-032	0.0334	0.9199	0.0464	SG2
CMC-033	0.0204	0.7098	0.2697	SG2
CMC-036	0.0227	0.4338	0.5436	SG3
CMC-085	0.0699	0.4496	0.4807	M
CMC-090	0.0468	0.5967	0.3564	SG2
CMC-099	0.0831	0.445	0.4722	M
CMC-143	0.019	0.9096	0.0714	SG2
CMC-146	0.2505	0.6981	0.0513	SG2
CMC-180	0.0421	0.7362	0.2216	SG2
CMC-181	0.1033	0.5674	0.3292	SG2
CMC-185	0.0274	0.9554	0.0174	SG2
CMC-205	0.0714	0.6332	0.295	SG2
CMC-211	0.0108	0.818	0.1712	SG2
CMC-215	0.0912	0.6587	0.2502	SG2
CMC-219	0.0138	0.8817	0.1047	SG2
CMC-220	0.0382	0.7679	0.194	SG2
CMC-222	0.0105	0.9819	0.0074	SG2

CMC-227	0.0704	0.6671	0.2627	SG2
CMC-228	0.0444	0.553	0.4024	SG2
CMC-373	0.0486	0.7221	0.2296	SG2
CMC-391	0.1069	0.6282	0.2644	SG2
CMC-396	0.0206	0.8425	0.1368	SG2
CMC-405	0.0771	0.5515	0.3713	SG2
CMC-436	0.0466	0.8158	0.1377	SG2
CMC-502	0.0359	0.9438	0.0203	SG2
CMC-599	0.0442	0.3976	0.5581	SG3
CMC-607	0.0163	0.9752	0.0084	SG2
CMC-621	0.0149	0.6538	0.331	SG2
CMC-640	0.0148	0.9101	0.0748	SG2
CUZ-032	0.0333	0.3454	0.6216	SG3
CUZ-034	0.0775	0.1981	0.7241	SG3
CUZ-209	0.1545	0.345	0.5004	SG3
CUZ-260	0.0302	0.6071	0.3631	SG2
HCO-112	0.5356	0.4391	0.0254	SG2
HCO-120	0.3357	0.4483	0.216	SG2
HVCA-141	0.0861	0.8849	0.0289	SG2
HVCA-275	0.5935	0.0068	0.3995	SG1
HVCA-277	0.5179	0.0614	0.4209	SG1
HVCA-278	0.5176	0.2338	0.2485	SG1
HVCA-280	0.3644	0.5786	0.057	SG2
HVCA-285	0.489	0.292	0.2192	SG1
HVCA-289	0.4645	0.3989	0.1367	SG1
JUN-008	0.2935	0.4844	0.2223	SG2
JUN-440	0.0771	0.7288	0.1943	SG2
JUN-463	0.5864	0.0162	0.397	SG1
JUN-502	0.5497	0.2645	0.1857	SG1
JUN-523	0.4951	0.0063	0.4984	M
JUN-534	0.6084	0.0488	0.3427	SG1
JUN-535	0.4977	0.4111	0.0908	M
JUN-550	0.2987	0.3863	0.3149	M
JUN-554	0.2351	0.5357	0.2294	SG2
MOQ-002	0.0061	0.8887	0.1051	SG2
MOQ-003	0.0201	0.6031	0.3766	SG2
MOQ-010	0.0123	0.6547	0.3331	SG2
MOQ-013	0.1065	0.2055	0.6878	SG3
MOQ-016	0.0142	0.5749	0.4107	SG2
TAC-064	0.0106	0.809	0.1801	SG2

Anexo 14: Valores propios, porcentaje acumulado de varianza y coeficientes de correlación de las accesiones a las primeras cinco CP del PCA

Componentes principales	Valor propio	% Varianza
1	12.314	1.862
2	9.838	1.488
3	9.551	1.445
4	8.723	1.319
5	7.871	1.191
6	7.433	1.124
7	7.290	1.103
8	7.021	1.062
9	6.989	1.057
10	6.922	1.047
11	6.747	1.021
12	6.685	1.011
13	6.645	1.005
14	6.615	1.001
15	6.569	0.994
16	6.528	0.987
17	6.423	0.971
18	6.399	0.968
19	6.384	0.966
20	6.371	0.964
21	6.340	0.959
22	6.290	0.951
23	6.277	0.949
24	6.218	0.940
25	6.187	0.936
26	6.151	0.930
27	6.136	0.928
28	6.114	0.925
29	6.091	0.921
30	6.080	0.920
31	6.067	0.918
32	6.040	0.914
33	6.010	0.909
34	6.002	0.908
35	5.950	0.900
36	5.931	0.897
37	5.919	0.895

38	5.900	0.892
39	5.869	0.888
40	5.861	0.886
41	5.816	0.880
42	5.805	0.878
43	5.789	0.876
44	5.767	0.872
45	5.724	0.866
46	5.714	0.864
47	5.686	0.860
48	5.681	0.859
49	5.655	0.855
50	5.620	0.850
51	5.616	0.849
52	5.606	0.848
53	5.583	0.844
54	5.568	0.842
55	5.552	0.840
56	5.509	0.833
57	5.482	0.829
58	5.470	0.827
59	5.456	0.825
60	5.450	0.824
61	5.414	0.819
62	5.398	0.816
63	5.377	0.813
64	5.357	0.810
65	5.353	0.810
66	5.333	0.807
67	5.317	0.804
68	5.289	0.800
69	5.277	0.798
70	5.249	0.794
71	5.239	0.792
72	5.211	0.788
73	5.199	0.786
74	5.189	0.785
75	5.183	0.784
76	5.145	0.778
77	5.139	0.777
78	5.120	0.774
79	5.094	0.770
80	5.061	0.765

81	5.044	0.763
82	5.024	0.760
83	4.995	0.755
84	4.970	0.752
85	4.946	0.748
86	4.940	0.747
87	4.929	0.745
88	4.907	0.742
89	4.896	0.740
90	4.869	0.736
91	4.841	0.732
92	4.836	0.731
93	4.806	0.727
94	4.784	0.724
95	4.755	0.719
96	4.735	0.716
97	4.714	0.713
98	4.691	0.710
99	4.678	0.708
100	4.624	0.699
101	4.611	0.697
102	4.596	0.695
103	4.579	0.693
104	4.566	0.691
105	4.551	0.688
106	4.510	0.682
107	4.478	0.677
108	4.466	0.675
109	4.440	0.672
110	4.402	0.666
111	4.387	0.664
112	4.376	0.662
113	4.343	0.657
114	4.329	0.655
115	4.269	0.646
116	4.222	0.639
117	2.785	0.421
118	2.040	0.309

Accesiones	Coeficientes				
	PC1	PC2	PC3	PC4	PC5
ANC-044	-3.313	-4.771	-0.453	1.402	-2.778
ANC-047	-2.649	-2.371	-1.910	-2.438	-3.527
ANC-074	-2.661	-2.298	-1.651	0.188	-0.601
ANC-262	-1.175	-0.698	-1.252	-0.461	-3.799
ANC-375	-2.652	-0.335	-0.650	0.559	-1.297
ANC-399	1.302	2.520	-0.825	-0.914	0.032
ANC-491	-1.792	-1.938	-1.040	-1.264	-3.120
ANC-522	-2.092	-0.594	-1.030	1.701	1.318
ANC-530	-2.212	0.572	-0.069	-0.747	3.968
ANC-561	-3.277	-4.210	-0.956	-1.129	-1.988
ANC-622	-3.857	-0.204	-1.386	0.570	0.007
ANC-630	-4.667	-5.703	-2.160	0.875	1.646
APUC-068	0.349	0.947	-0.558	-3.655	-4.540
APUC-069	1.686	-0.118	0.766	-0.754	-1.899
APUC-071	-0.460	-0.541	-0.951	-0.292	-2.194
APUC-110	0.138	-2.553	-0.602	-0.795	-3.856
APUC-136	-2.246	-0.712	-0.064	-1.944	-3.593
APUC-150	-2.023	-4.409	-0.158	1.009	-4.968
APUC-211	-0.390	1.435	-0.681	-3.115	-4.228
APUC-261	1.140	-0.522	0.814	-1.281	-2.127
APUC-274	0.011	1.268	-1.105	-0.436	-2.548
APUC-301	-0.634	-0.864	0.332	-1.484	-4.193
APUC-307	2.834	-1.041	1.213	-0.960	-2.944
APUC-309	2.150	0.930	-0.506	-1.306	-2.460
APUC-320	-0.771	7.333	-3.356	-5.073	1.821
AYA-089	3.481	-2.419	2.206	1.239	1.101
AYA-090	1.713	1.589	0.145	-1.348	0.538
AYA-093	0.889	2.355	-1.328	-0.706	0.342
AYA-123	3.352	-3.284	1.951	0.444	-0.067
AYA-176	2.202	0.024	-0.983	-0.999	-0.945
AYA-177	1.919	1.826	-0.260	-2.691	1.287
AYA-178	6.113	-9.713	4.617	2.739	-8.404
AYA-180	2.582	0.468	-1.147	-2.106	0.603
AYA-183	1.514	-1.984	4.334	2.561	3.794
AYA-185	-0.265	-1.239	2.138	1.821	1.902
AYA-187	1.698	-1.043	1.350	0.696	1.812
AYA-188	2.893	-0.333	0.678	-2.246	-7.074
AYA-193	2.931	-2.539	-0.588	-1.424	-8.069
AYA-195	0.043	-0.042	-0.407	-1.326	-2.539

AYA-201	0.724	0.531	1.449	0.697	2.107
AYA-203	2.364	3.161	-0.356	-2.243	3.520
AYA-224	3.332	-1.216	0.545	0.713	3.534
AYA-228	-1.507	0.558	5.482	2.145	-0.065
AYA-229	3.409	1.486	-0.691	-1.198	6.630
AYA-243	5.105	4.647	-12.718	16.879	-2.939
AYA-247	3.561	-3.752	1.934	0.005	0.161
AYA-264	4.631	5.431	-12.911	16.589	-3.090
AYA-271	2.198	0.545	-0.496	-2.386	-2.222
AYA-273	0.679	5.222	-1.535	-2.572	0.379
AYA-295	1.974	0.304	1.017	-1.701	-2.037
AYA-305	1.859	0.741	-1.284	-0.929	-1.208
AYA-327	3.115	1.132	-0.506	-2.279	3.229
AYAC-203	3.028	-0.658	1.014	-1.753	-5.676
CAJ-104	-8.472	-4.175	-3.622	-1.385	-3.824
CAJ-108	-9.291	-5.234	-4.007	0.742	2.857
CAJ-110	-7.369	-4.620	-2.511	1.451	1.166
CCC-041	-9.132	-5.442	-3.864	-0.714	-0.545
CCC-045	-9.454	-4.182	-4.052	-0.594	3.991
CCC-051	-10.524	0.049	-5.845	-2.013	2.844
CCC-073	-9.459	-3.555	-4.867	0.068	3.357
CMC-026	2.056	1.638	-1.940	-1.462	-0.010
CMC-030	1.008	-0.467	1.023	-0.074	0.807
CMC-032	2.020	-2.125	2.050	1.860	3.174
CMC-033	2.752	-0.025	-0.761	-0.110	1.591
CMC-036	1.721	3.169	-2.252	-3.008	-0.015
CMC-085	1.118	2.370	-0.385	-2.422	-1.621
CMC-090	1.215	0.580	0.245	-1.258	-1.727
CMC-099	0.617	2.484	-0.563	-3.210	-1.698
CMC-143	3.679	-2.121	1.369	1.478	1.036
CMC-146	1.056	-3.701	0.629	1.828	3.182
CMC-180	2.384	-0.009	-0.142	1.093	2.039
CMC-181	1.390	0.382	-0.386	-0.890	-0.399
CMC-185	3.522	-5.356	1.307	0.934	0.317
CMC-205	0.805	0.167	-0.389	0.313	1.435
CMC-211	3.209	0.749	0.989	0.512	1.200
CMC-215	1.636	0.363	-1.175	-0.572	1.882
CMC-219	3.584	-1.650	1.763	0.668	1.888
CMC-220	2.313	-0.923	0.931	-0.352	-0.880
CMC-222	3.880	-4.271	3.530	3.916	3.546
CMC-227	2.544	-0.354	-1.243	-1.307	-0.877
CMC-228	1.291	2.909	-0.059	0.480	1.575
CMC-373	2.263	0.345	0.186	-0.180	1.503

CMC-391	1.396	0.237	-0.759	-0.236	-0.734
CMC-396	2.695	-0.878	1.657	-0.961	2.459
CMC-405	1.317	1.167	-0.736	-1.456	-0.637
CMC-436	2.599	-1.050	0.454	1.247	2.023
CMC-502	2.258	-4.631	2.734	4.536	3.676
CMC-599	1.387	3.633	-2.118	-3.182	2.100
CMC-607	3.002	-6.032	3.150	3.200	1.155
CMC-621	2.886	1.685	-0.099	-1.757	-0.725
CMC-640	3.037	-0.815	1.909	0.174	3.764
CUZ-032	1.479	4.007	-2.255	-2.977	1.440
CUZ-034	0.029	4.620	-4.476	-5.343	-1.925
CUZ-209	-0.388	2.425	-1.945	-1.899	4.098
CUZ-260	1.780	1.994	-1.764	-1.442	1.361
HCO-112	-3.761	-3.611	1.714	1.554	2.793
HCO-120	-0.836	-1.463	-1.287	-0.257	1.112
HVCA-141	1.184	-2.768	3.286	-0.568	-0.935
HVCA-275	-7.346	10.101	12.066	6.102	-3.563
HVCA-277	-5.829	7.146	6.088	2.819	-1.493
HVCA-278	-4.270	2.927	3.515	1.683	0.754
HVCA-280	-2.264	-2.126	4.800	2.346	0.309
HVCA-285	-4.293	2.957	4.302	2.951	3.687
HVCA-289	-3.024	-0.404	2.401	2.305	0.804
JUN-008	-0.299	0.338	-0.913	-0.081	-2.460
JUN-440	1.182	0.075	0.226	0.974	3.137
JUN-463	-6.894	9.563	11.763	6.622	-4.380
JUN-502	-3.962	0.598	3.060	1.386	-1.036
JUN-523	-6.330	6.107	-1.152	-1.355	1.180
JUN-534	-5.384	2.878	0.436	0.177	-1.436
JUN-535	-4.133	-1.018	3.260	2.360	2.155
JUN-550	-1.975	1.573	0.362	-0.066	0.538
JUN-554	-0.381	0.066	1.138	-0.496	0.929
MOQ-002	4.285	-0.218	1.256	-0.118	5.916
MOQ-003	1.760	1.921	-1.374	-1.029	1.043
MOQ-010	2.818	1.030	0.466	-2.383	1.933
MOQ-013	-0.336	6.022	-3.852	-2.851	2.157
MOQ-016	2.316	1.816	-0.451	-4.409	-2.679
TAC-064	3.655	0.212	1.763	-0.237	4.951

Anexo 15: Matrices de p-valor del Fst por pares y tablas del bootstrap y los intervalos de confianza para cada Fst

Matriz de p-valor del Fst por pares de razas de maíz amiláceo

	Paro	Chullpi	Piscorunto
Paro	-	-	-
Chullpi	0	-	-
Piscorunto	0	0	-

Tabla del bootstrap y los intervalos de confianza para cada Fst por pares de razas de maíz amiláceo.

Población 1	Población 2	Bootstrap			Intervalo de confianza		p-valor	Fst
		1	2	3	Límite inferior	Límite superior		
Paro	Chullpi	0.001	0.001	0.001	0.001	0.001	0	0.001
Paro	Piscorunto	0.002	0.002	0.002	0.002	0.002	0	0.002
Chullpi	Piscorunto	0.003	0.003	0.003	0.003	0.003	0	0.003

Matriz de p-valor del Fst por pares de grupos determinados por el programa STRUCTURE

	SG1	M	SG2	SG3
SG1	-	-	-	-
M	0	-	-	-
SG2	0	0	-	-
SG3	0	0	0	-

Tabla del bootstrap y los intervalos de confianza para cada Fst por pares de grupos determinados por el origen de la accesión.

Población 1	Población 2	Bootstrap			Intervalo de confianza		p-valor	Fst
		1	2	3	Límite inferior	Límite superior		
SG1	M	0.003	0.003	0.004	0.003	0.003	0	0.004
SG1	SG2	0.008	0.008	0.009	0.008	0.008	0	0.008
SG1	SG3	0.013	0.013	0.014	0.013	0.013	0	0.014
M	SG2	0.002	0.002	0.002	0.002	0.002	0	0.002
M	SG3	0.004	0.004	0.005	0.004	0.004	0	0.005
SG2	SG3	0.006	0.006	0.006	0.006	0.006	0	0.006

Matriz de p-valor del Fst por pares de grupos de origen de las accesiones

	Ancash	Apurímac	Ayacucho	Cajamarca	Cusco	Huánuco	Huancavelica	Junín	Moquegua	Tacna
Ancash	-	-	-	-	-	-	-	-	-	-
Apurímac	0	-	-	-	-	-	-	-	-	-
Ayacucho	0	0	-	-	-	-	-	-	-	-
Cajamarca	0	0	0	-	-	-	-	-	-	-
Cusco	0	0	0	0	-	-	-	-	-	-
Huánuco	0	0	0	0	0	-	-	-	-	-
Huancavelica	0	0	0	0	0	0	-	-	-	-
Junín	0	0	0	0	0	0	1	-	-	-
Moquegua	0	0	0	0	0	0	0	0	-	-
Tacna	0	0.667	0.333	0	0	0	0	0	0.333	-

Tabla del bootstrap y los intervalos de confianza para cada Fst por pares de grupos de origen de las accesiones

Población 1	Población 2	Bootstrap			Intervalo de confianza		p-valor	Fst
		1	2	3	Límite inferior	Límite superior		
Ancash	Apurímac	0.003	0.003	0.003	0.003	0.003	0.000	0.003
Ancash	Ayacucho	0.005	0.006	0.006	0.005	0.006	0.000	0.005
Ancash	Cajamarca	0.009	0.009	0.010	0.009	0.009	0.000	0.009
Ancash	Cusco	0.003	0.003	0.004	0.003	0.003	0.000	0.003
Ancash	Huánuco	0.005	0.005	0.006	0.005	0.005	0.000	0.005
Ancash	Huancavelica	0.008	0.008	0.009	0.008	0.008	0.000	0.008
Ancash	Junín	0.004	0.004	0.005	0.004	0.004	0.000	0.004
Ancash	Moquegua	0.005	0.007	0.007	0.005	0.007	0.000	0.006
Ancash	Tacna	0.002	0.005	0.006	0.002	0.005	0.000	0.003
Apurímac	Ayacucho	0.002	0.002	0.002	0.002	0.002	0.000	0.002
Apurímac	Cajamarca	0.016	0.017	0.018	0.016	0.017	0.000	0.016
Apurímac	Cusco	0.001	0.001	0.001	0.001	0.001	0.000	0.001
Apurímac	Huánuco	0.008	0.008	0.008	0.008	0.008	0.000	0.008
Apurímac	Huancavelica	0.006	0.006	0.007	0.006	0.006	0.000	0.006
Apurímac	Junín	0.004	0.004	0.005	0.004	0.004	0.000	0.005
Apurímac	Moquegua	0.002	0.002	0.002	0.002	0.002	0.000	0.002
Apurímac	Tacna	-0.002	-0.001	0.002	-0.002	-0.001	0.667	-0.001
Ayacucho	Cajamarca	0.020	0.022	0.022	0.020	0.022	0.000	0.020
Ayacucho	Cusco	0.002	0.002	0.002	0.002	0.002	0.000	0.002
Ayacucho	Huánuco	0.009	0.010	0.010	0.009	0.010	0.000	0.011
Ayacucho	Huancavelica	0.009	0.009	0.010	0.009	0.009	0.000	0.009
Ayacucho	Junín	0.006	0.006	0.006	0.006	0.006	0.000	0.006
Ayacucho	Moquegua	0.002	0.002	0.002	0.002	0.002	0.000	0.002
Ayacucho	Tacna	-0.002	0.001	0.003	-0.002	0.001	0.333	0.000
Cajamarca	Cusco	0.014	0.016	0.017	0.014	0.016	0.000	0.015
Cajamarca	Huánuco	0.014	0.015	0.016	0.014	0.015	0.000	0.015
Cajamarca	Huancavelica	0.020	0.021	0.022	0.020	0.021	0.000	0.021

Cajamarca	Junín	0.015	0.016	0.016	0.015	0.016	0.000	0.016
Cajamarca	Moquegua	0.025	0.026	0.026	0.025	0.026	0.000	0.025
Cajamarca	Tacna	0.032	0.034	0.035	0.032	0.034	0.000	0.033
Cusco	Huánuco	0.007	0.007	0.008	0.007	0.007	0.000	0.008
Cusco	Huancavelica	0.007	0.008	0.008	0.007	0.008	0.000	0.008
Cusco	Junín	0.004	0.004	0.005	0.004	0.004	0.000	0.005
Cusco	Moquegua	0.002	0.002	0.002	0.002	0.002	0.000	0.002
Cusco	Tacna	0.001	0.002	0.004	0.001	0.002	0.000	0.001
Huánuco	Huancavelica	0.012	0.014	0.014	0.012	0.014	0.000	0.014
Huánuco	Junín	0.008	0.009	0.009	0.008	0.009	0.000	0.010
Huánuco	Moquegua	0.010	0.013	0.013	0.010	0.013	0.000	0.014
Huánuco	Tacna	0.002	0.004	0.008	0.002	0.004	0.000	0.003
Huancavelica	Junín	-0.002	-0.002	-0.001	-0.002	-0.002	1.000	-0.001
Huancavelica	Moquegua	0.010	0.012	0.013	0.010	0.012	0.000	0.012
Huancavelica	Tacna	0.015	0.015	0.016	0.015	0.015	0.000	0.014
Junín	Moquegua	0.007	0.008	0.008	0.007	0.008	0.000	0.008
Junín	Tacna	0.005	0.008	0.009	0.005	0.008	0.000	0.008
Moquegua	Tacna	-0.001	0.002	0.002	-0.001	0.002	0.333	0.000

Anexo 16: Tabla del Análisis Molecular de Variación para las razas, grupos de STRUCTURE y origen de accesiones

AMOVA de las razas de maíz amiláceo

Fuentes de variación	Df	Suma de cuadrados medios	Media de los cuadrados
Entre razas	2	3285.267	1642.634
Entre las accesiones	116	154074.69	1328.23
Dentro de las accesiones	119	312970	2630
Total	237	470329.96	1984.515

AMOVA de los grupos determinados por el programa STRUCTURE

Fuentes de variación	Df	Suma de cuadrados medios	Media de los cuadrados
Entre grupos de STRUCTURE	3	5523.778	1841.259
Entre las accesiones	115	151836.18	1320.315
Dentro de las accesiones	119	312970	2630
Total	237	470329.96	1984.515

AMOVA de los grupos determinados por el origen de las accesiones

Fuentes de variación	Df	Suma de cuadrados medios	Media de los cuadrados
Entre grupos de origen de accesiones	9	13610.54	1512.282
Entre las accesiones	109	143749.42	1318.802
Dentro de las accesiones	119	312970	2630
Total	237	470329.96	1984.515