



UNIVERSIDAD NACIONAL DE SAN ANTONIO ABAD DEL CUSCO
ESCUELA DE POSGRADO
MAESTRÍA EN ESTADÍSTICA

TESIS

**ANÁLISIS DE LA DIVERSIDAD GENÉTICA DE QUINUA
(*Chenopodium Quinoa* Willd.) DEL BANCO DE GERMOPLASMA DE
PUNO, MEDIANTE APLICACIÓN DE TÉCNICAS MULTIVARIADAS,
2021**

**PARA OPTAR AL GRADO ACADÉMICO DE MAESTRO EN
ESTADÍSTICA**

AUTOR

Br. JOSE LUIS CAHUANA JORGE

ASESOR

Dr. CLETO DE LA TORRE DUEÑAS
CÓDIGO ORCID: 0000-0003-0921-7217

CO ASESOR

Mtro. ARTURO ZUÑIGA BLANCO
CÓDIGO ORCID: 0000-0002-8576-3415

CUSCO-PERÚ

2024

INFORME DE ORIGINALIDAD

(Aprobado por Resolución Nro.CU-303-2020-UNSAAC)

El que suscribe, **Asesor** del trabajo de investigación/tesis titulada “ANÁLISIS DE LA DIVERSIDAD GENÉTICA DE QUINUA (Chenopodium quinoa Willd.) DEL BANCO DE GERMOPLASMA DE PUNO, MEDIANTE APLICACIÓN DE TÉCNICAS MULTIVARIADAS, 2021”, presentado por: JOSE LUIS CAHUANA JORGE. con Nro. de DNI: ...01342917, para optar el grado académico de **Maestro en Estadística**. Informo que el trabajo de investigación ha sido sometido a revisión por 02 (dos) veces, mediante el Software Antiplagio, conforme al Art. 6° del *Reglamento para Uso de Sistema Antiplagio de la UNSAAC* y de la evaluación de originalidad se tiene un porcentaje de **9% (Nueve por ciento)**.

Evaluación y acciones del reporte de coincidencia para trabajos de investigación conducentes a grado académico o título profesional, tesis

Porcentaje	Evaluación y Acciones	Marque con una (X)
Del 1 al 10%	No se considera plagio.	X
Del 11 al 30 %	Devolver al usuario para las correcciones.	----
Mayor a 31%	El responsable de la revisión del documento emite un informe al inmediato jerárquico, quien a su vez eleva el informe a la autoridad académica para que tome las acciones correspondientes. Sin perjuicio de las sanciones administrativas que correspondan de acuerdo a Ley.	----

Por tanto, en mi condición de asesor, firmo el presente informe en señal de conformidad y **adjunto** la primera hoja del reporte del Sistema Antiplagio.

Cusco,...04.... de.....diciembre..... de 2023.....



Firma

Post firma...: CLETO DE LA TORRE DUEÑAS

Nro. de DNI: 23988416

ORCID del Asesor <https://orcid.org/0000-0003-0921-7217>

Se adjunta:

1. Reporte generado por el Sistema Antiplagio.
2. Enlace del Reporte Generado por el Sistema Antiplagio:
<https://unsaac.turnitin.com/viewer/submissions/oid:27259:292249274>

NOMBRE DEL TRABAJO

TESIS_joseluisCahuana.docx

AUTOR

JOSE LUIS CAHUANA

RECUENTO DE PALABRAS

11579 Words

RECUENTO DE CARACTERES

63726 Characters

RECUENTO DE PÁGINAS

61 Pages

TAMAÑO DEL ARCHIVO

3.0MB

FECHA DE ENTREGA

Dec 4, 2023 2:51 PM GMT-5

FECHA DEL INFORME

Dec 4, 2023 2:53 PM GMT-5**● 9% de similitud general**

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para cada base c

- 9% Base de datos de Internet
- Base de datos de Crossref
- 2% Base de datos de publicaciones
- Base de datos de contenido publicado de Crossr

● Excluir del Reporte de Similitud

- Base de datos de trabajos entregados
- Material citado
- Coincidencia baja (menos de 8 palabras)
- Material bibliográfico
- Material citado
- Bloques de texto excluidos manualmente

DEDICATORIA

*De manera muy especial a mis padres
Cristobal Cahuana y Maria Jorge, por su gran
esfuerzo y perseverancia en la vida,
que han hecho posible mi formación
profesional, así como su invaluable apoyo que
me ha permitido conseguir
mis objetivos.*

*A mis dos hermanitas Yudith y Mery,
por estar ahí en todo momento,
dándome el aliento para seguir adelante
y brindándome siempre
su apoyo incondicional.*

*Reiteradamente a ustedes papá, mamá y
hermanitas, que son mi alegría,
mi fortaleza y mi motivo
de superación
en la vida.*

AGRADECIMIENTO

En principio dar gracias a Dios, por permitirme avanzar en este camino de la vida guiándome y protegiéndome siempre.

Así mismo, agradecer profundamente a mis asesores: Dr. Cleto de la Torre Dueñas y Maestro Arturo Zúñiga Blanco, por permitirme absorber sus conocimientos, experiencia, consejos y paciencia para concluir con el desarrollo del presente trabajo de investigación.

Mi reconocimiento a la Dra. Karla Aparicio Arenas, Mgt. Carla Zúñiga Vilca, Mtra. Soriana Alida Jimenez Coa y Mtra. Yariani Barreto Serrano, por cada uno de sus aportes y sugerencias en el dictamen y que han permitido mejorar el presente trabajo de investigación.

Mi agradecimiento al Dr. Edwin Molina Porcel director de la Escuela de Posgrado y todos los docentes de la Maestría en Estadística de la Universidad Nacional San Antonio Abad del Cusco, que me apoyaron y contribuyeron a mi formación, con sus conocimientos y experiencia de su trayectoria profesional.

De la misma manera, mi agradecimiento al Dr. Rafael Velásquez Huallpa (UNAP), Rigoberto Laura (UNSAAC), Tulio Medina (MINAM), Wilfredo Rojas (PROINPA), David Soraide (UATF-Bolivia), Policarpo Catacora (INIA), Fernando Casanoves (CATIE-Costa Rica), George Argota (AMTAWI), quienes me han permitido continuar por la senda de investigación de la quinua; de la misma manera, a mis docentes de pregrado de la Universidad Nacional del Altiplano de Puno quienes en su momento apoyaron en la consolidación de la información fuente de este trabajo.

Finalmente, mi agradecimiento especial a Leonila Marcelo, Quinia Ariza, Karem Larrauri, Waldir Arbildo, Luis Escurra, Luis Enrique Rodriguez, Alex Arana, Jorge Ingunza, Sergio Valcarcel, Josue Tunqui, Ina Lopez, Magaly Mera y tod@s mis amistades con los cuales tuve la oportunidad de conversar y compartir conocimientos en las que tuve siempre su apoyo incondicional.

ÍNDICE GENERAL

I. Planteamiento del problema.....	5
1.1. Situación Problemática.....	5
1.2. Formulación del problema	6
a. Problema general	6
b. Problemas específicos	6
1.3. Justificación de la investigación.....	6
1.4. Objetivos de la investigación	7
a. Objetivo General	7
b. Objetivos específicos	7
II. Marco teórico conceptual	8
2.1. Bases teóricas	8
2.1.1. La quinua (Chenopodium quinoa Willd.).....	8
2.1.2. Origen, domesticación y distribución geográfica.....	8
2.1.3. Diversidad genética del cultivo de quinua.....	9
2.1.4. Importancia de la quinua.....	10
2.1.5. Fases fenológicas de la quinua.....	10
2.1.6. Evaluación y caracterización de la quinua.	13
2.1.7. Análisis de componentes principales	14
2.1.8. Análisis de agrupamiento o clustering.....	16
2.1.9. Análisis Discriminante.	21
2.2. Marco conceptual (Palabra clave).....	27
2.2.1. Chenopodium quinoa Willd.....	27
2.2.2. Fenología.	27
2.2.3. Diversidad genética.	27
2.2.4. Técnica multivariada.....	28
2.2.5. Análisis Clúster.....	28
2.2.6. Análisis discriminante.	29
2.3. Antecedentes de la investigación.....	29
2.3.1. Antecedentes internacionales	29
2.3.2. Antecedentes nacionales	31
III. Hipótesis y variables	33
3.1. Hipótesis	33

a.	Hipótesis general.....	33
b.	Hipótesis específicas.....	33
3.2.	Identificación de variables e indicadores.....	33
3.3.	Operacionalización de variables.....	34
IV.	Metodología.....	35
4.1.	Ámbito de estudio: Localización política y geográfica.....	35
4.2.	Tipo y nivel de investigación.....	36
4.3.	Unidad de Análisis.....	36
4.4.	Población de estudio.....	36
4.5.	Tamaño de muestra.....	36
4.6.	Técnicas de selección de la muestra.....	36
4.7.	Técnicas de recolección de información.....	36
4.8.	Técnicas de análisis e interpretación de la información.....	36
V.	Resultados y discusión.....	38
5.1.	Preprocesamiento.....	38
5.2.	Identificación de datos outliers.....	39
5.3.	Estandarización de variables.....	42
5.4.	Correlación de las variables.....	42
5.5.	Análisis de componentes principales.....	44
5.6.	Asociación de las colecciones de quinua.....	48
5.7.	Las variables que discriminan a los grupos de colecciones de quinua.....	51
5.8.	Índice de diversidad.....	53
	Conclusiones.....	55
	Recomendaciones.....	56
	Bibliografía.....	57
	Anexos.....	59
a.	Matriz de consistencia.....	59
b.	Instrumentos de recolección de datos a nivel de campo.....	60
c.	Medidas resumen de las variables cuantitativas general.....	61
d.	Medidas resumen de las variables cuantitativas (Perú).....	62
e.	Medidas resumen de las variables cuantitativas (Bolivia).....	63
f.	Código R utilizado para el procesamiento de datos.....	64
g.	Documento de verificación de datos (fuente de datos).....	68
h.	Equipo de profesionales en el campo de germoplasma de quinua, Camacani Puno...	68
i.	Presentación de resultados en el 8vo Congreso Mundial de la quinua 2023.....	69

ÍNDICE DE TABLAS

Tabla 1 Tabla de contingencia 2x2 de variable binaria.....	19
Tabla 2 Operacionalización de variables.....	34
Tabla 3 Resultados de los coeficientes.....	52
Tabla 4 Índice de diversidad de las variables.....	53

ÍNDICE DE FIGURAS

Figura 1 Mapa de ubicación geográfica de la investigación.	35
Figura 2 Proceso de análisis de datos.....	37
Figura 3 Distribución de datos faltantes.....	39
Figura 4 Distribución de las mediciones en los descriptores de caracterización de fenología.	40
Figura 5 Distribución de las mediciones en los descriptores de caracterización de planta.....	40
Figura 6 Distribución de las mediciones en los descriptores de caracterización de panoja.....	41
Figura 7 Distribución de las mediciones en los descriptores de caracterización de grano.	41
Figura 8 Correlación de las variables de caracterización.....	42
Figura 9 Biplot de componentes principales (Variables).....	44
Figura 10 Biplot de componentes principales (Colecciones de quinua).....	45
Figura 11 Gráfico de componentes principales de las variables de caracterización. ..	46
Figura 12 Gráfico de sedimentación para el número de componentes principales.....	47
Figura 13 Asociación de las colecciones de quinua mediante técnica de Clúster.	48
Figura 14 Matriz gráfica sobre la cantidad óptima de grupos en las colecciones de quinua.....	49
Figura 15 Dendograma de las colecciones de quinua.	51

RESUMEN

El propósito de esta investigación consistió en utilizar técnicas de análisis multivariado para describir la diversidad genética de la quinua (*Chenopodium quinoa* Willd.) presente en el Banco de Germoplasma de Puno. Se realizó desde junio del 2021 hasta abril de 2022, la fuente de información fue tomada del Catálogo de Banco de Germoplasma de quinua de Puno, los cuales fueron analizados, recogiendo información de 734 colecciones con registros completos, se aplicó las técnicas de clúster, discriminante y análisis de componentes principales (PCA). En el análisis clúster con distancia euclídea, se encontró dos grupos claramente diferenciados; para lo cual, mediante técnica discriminante se ha determinado que la variable “país de procedencia” de las colecciones es la que ha influenciado para la formación de los grupos, los mismos que se verificaron con los datos de pasaporte de las colecciones (Perú y Bolivia); así mismo, en los resultados de PCA, se ha reportado los 3 primeros componentes explicaron el 53.1% de la variación total de los datos, con un KMO-MSA de 0.83 que indica la pertinencia del análisis. En referencia a la técnica discriminante, las variables: “altura de planta” (altuPlan_cm), “días de panojamiento” (Panoj_dias), “rendimiento de grano por planta” (rendim_gpla) y “días de emergencia” (Emerg) fueron las que mejor discriminan a cada uno de los grupos encontrados, así mismo, el índice de diversidad encontrado promedio respecto a todas las variables es de 0.82 que se traduce en una diversidad moderadamente alta. Estos resultados encontrados tienen relación con lo indicado por otros autores, validando la genética y morfología de la colección del Banco de Germoplasma de quinua de Puno.

Palabras clave: Análisis multivariados, clúster, discriminante, quinua, recursos genéticos, variabilidad genética,

I. Planteamiento del problema

1.1. Situación Problemática

La diversidad genética, en un enfoque general es el componente básico de la biodiversidad, y ello está referido a todas las variaciones heredables que se suelen presentar en una especie y por consiguiente, entre las distintas poblaciones dentro de la misma especie, siempre y cuando se considere las condiciones naturales de su forma estable (Rimieri, 2017).

Entre las plantas superiores, se tiene cerca de 400,000 genes, cuyas variantes se van acumulando dentro de la especie y la suma de todas las características de los genes y sus variantes generan la variabilidad.

En ese contexto, uno de los cultivos alto andinos más importante es la quinua, que empezó a Bolivia y el Perú, pero posteriormente en otros países como: Colombia, Ecuador, los valles interandinos y planicies de Chile y otros (Lescano, 1994).

Específicamente, el altiplano de la región Puno y Bolivia, es conocida como el origen de la mayor variabilidad y diversidad del cultivo de la quinua, ya que en esas regiones se cultivan la mayor cantidad de líneas, cultivares o variedades; lo que confirma que la quinua es una especie autóctona de Sudamérica, especialmente de los andes peruanos y bolivianos (Cardenas, 1994).

Sin embargo, los estudios que se han realizados solo corresponden a pequeños grupos de colecciones de manera descriptiva; por lo tanto, aun no se evidencia estudios de variabilidad genética de todo el banco de germoplasma de las colecciones de quinua mediante la aplicación de las técnicas multivariadas.

1.2. Formulación del problema

a. Problema general

¿Cómo las técnicas multivariantes describen la diversidad genética de las colecciones de quinua en el Banco de Germoplasma de Puno?

b. Problemas específicos

- ¿Cómo se concentra la mayor variabilidad, en la colección de quinua del Banco de Germoplasma de Puno?
- ¿Cómo es la asociación de las colecciones de quinua en el Banco de Germoplasma de Puno?
- ¿Cuál o cuáles son las variables que mejor discriminan los grupos de las colecciones de quinua en el Banco de Germoplasma de Puno?
- ¿Cuánto es la diversidad de la colección de quinua del Banco de Germoplasma de Puno?

1.3. Justificación de la investigación

La presente permitió describir, la diversidad genética de las colecciones de Banco de Germoplasma de quinua en la región Puno; lo cual, posibilita las implicaciones prácticas, la mejora genética, así como desde la relevancia social, el uso agrícola de las variedades genéticas. De igual manera, se reconoce desde la utilidad metodológica de los métodos multivariados mediante análisis clúster, análisis de componentes principales y análisis discriminante para conocer la asociación de las colecciones e importancia de las variables de caracterización de las colecciones.

1.4. Objetivos de la investigación

a. Objetivo General

Describir mediante técnicas multivariantes, la diversidad genética de la colección de quinua correspondiente al Banco de Germoplasma de Puno.

b. Objetivos específicos

- Determinar la concentración de la mayor variabilidad, en la colección de quinua del Banco de Germoplasma de Puno.
- Determinar la asociación de la colección de quinua correspondiente al Banco de Germoplasma de Puno.
- Determinar las variables que mejor discriminan las colecciones de quinua del Banco de Germoplasma de Puno.
- Cuantificar la diversidad de la colección de quinua del Banco de Germoplasma de Puno.

II. Marco teórico conceptual

2.1. Bases teóricas

2.1.1. La quinua (*Chenopodium quinoa* Willd.)

Según la FAO et al. (2013), la quinua tiene diferentes denominaciones que varían de acuerdo al idioma y del país, por ejemplo además de conocerlo como quinua, tiene los nombres tales como: quinua, jiura (Perú), quiuna, quingua (Chile), luba (Colombia) y otros. Además de ello, también el nombre varía en función al idioma; por ejemplo, en otros países le denominan: quingua, kinoa, quinoa, trigo inca, arroz del Perú, sweet quinoa, peruvian rice, inca rice, etc.

Esta especie del género *Chenopodium*, cuenta con al menos 120 especies (Fuentes et al., 2009), donde *Chenopodium quinoa* Willd. es la especie económicamente más importante, la primera fue descrita botánicamente por Carl Ludwig Willdenow en el año 1778 de ahí que el nombre científico lleva la abreviación Willd. y taxonómicamente fue ubicada en la familia Chenopodiaceae; sin embargo, según el reciente sistema de clasificación AGP IV (Chase et al., 2016) la quinua se ubica en la familia Amaranthaceae y subfamilia Chenopodioideae.

2.1.2. Origen, domesticación y distribución geográfica

La quinua es alotetraploide, lo que significa que se ha originado mediante un binomio de especies diploides (N° de cromosomas=18 cada una) que aún se están identificando. Mediante el análisis de ADN, se ha descubierto que uno de los subgenomas tiene origen americano, mientras que el otro subgenoma tiene origen euro-asiático (Jellen et al., 2011).

Así mismo, es un cultivo oligocéntrico, lo cual significa que su centro de origen, diversidad y distribución genética se ubica en la región del anillo circunlacustre del lago Titicaca (Gandarillas, 1979); por lo tanto, esta región se reconoce como el lugar de origen

de esta especie, con una antigüedad de aproximadamente 7000 años (Jacobsen y Mujica, 2001) y su desarrollo en las culturas de Tiwanaku e incaica (Bonifacio, 2004) los mismos que desempeñaron un papel importante en su domesticación y conservación; además, esta domesticación ha permitido mejoras en varios aspectos tales como: planta, inflorescencia (panoja), semillas, pigmentación y otros aspectos; sin embargo, los granos de quinua contienen un alcaloide que es la “saponina” en diferentes concentraciones, los mismos que se tienen que eliminar antes de ser consumida (Mujica et al., 2004).

En referencia a su distribución geográfica, la quinua se encuentra distribuida desde sur de Chile hasta el norte de Colombia, abarcando países de Argentina, Bolivia, Perú, Colombia y Ecuador (Lescano, 1994; Rojas et al., 2010 y Mujica y Jacobsen, 2006). Así mismo, es importante destacar que esta planta se encuentra presente en altitudes que van hasta los 4000 m.s.n.m. del altiplano peruano-boliviano y que actualmente, este cultivo se ha expandido a varias regiones del mundo tales como: Canadá, América del Norte, Europa, África y Asia. En cuanto a su adaptación y producción, ha mostrado resultados muy prometedores (Jacobsen y Mujica, 2001).

2.1.3. Diversidad genética del cultivo de quinua

La zona de mayor diversidad genética de la quinua y sus parientes silvestres es el altiplano peruano boliviano y de manera especial las zonas denominadas “aynoqas”, son lugares de sistemas ancestrales que permiten la organización, producción y conservación in situ (Mujica y Jacobsen, 2006); no obstante, últimamente su uso ha disminuido, por varios motivos como: la incorporación de nuevas tecnologías, las divulgación de nuevas variedades por parte de los investigadores y fitomejoradores, los mismos que presentan mejor rendimiento (Mujica y Jacobsen, 2001).

2.1.4. Importancia de la quinua

Es una planta de periodo anual que pertenece al grupo de las dicotiledóneas, es un cultivo se considera altamente prometedor contra las adversidades del del cambio climático, por su adaptabilidad a las condiciones adversas del clima, suelo, temperatura y humedad. Es capaz de tolerar suelos salinos, tolerancia a déficit de agua y temperaturas bajas que van entre -4° hasta 38°C (Christensen et al., 2007); por otro lado, un aspecto destacado de la quinua es su calidad nutricional de las semillas maduras, las hojas verdes, y las panojas en estado tierno. Todas estas características hacen que la quinua sea reconocida como el único alimento de origen vegetal que posee una proporción adecuada de todos los aminoácidos esenciales necesarios para una dieta equilibrada y saludable. Así mismo, estos aminoácidos son superiores a los encontrados en cultivos como la soja, la cebada y el trigo, los mismos que se comparan solamente con las proteínas presentes en la leche. Además, la quinua presenta una alta concentración de vitaminas, minerales y ácidos grasos y su característica más destacada es su ausencia de gluten; debido a esto, la quinua es una opción alimentaria recomendada para individuos que padecen trastornos abdominales (Zurita-Silva et al., 2014).

2.1.5. Fases fenológicas de la quinua.

a. Emergencia

Ocurre en el momento de que la plántula empieza a brotar del suelo y se producen las primeras hojas del cotiledón. Generalmente, esto sucede entre 7 y 10 días después de la siembra. Es durante este momento que la planta es más susceptible al ataque de aves, ya que al ser una planta dicotiledónea, emergen dos hojas protegidas por el episperma, lo que hace que la semilla sea más visible en el tallo (Jacobsen y Mujica, 2001). Es importante indicar que las semillas pueden germinar rápidamente bajo condiciones óptimas de temperatura, humedad y oxígeno. Entre estos factores, el agua y la temperatura del suelo

son elementos especialmente importantes para promover la germinación y el metabolismo adecuado de la planta (Gomez y Aguilar, 2016).

b. Dos hojas verdaderas

Son las hojas que crecen de 15 a 20 días aproximadamente después de haber sembrado la semilla, tienen forma romboidal que crecen en forma extendida; las hojas verdaderas de la quinua, que tienen una forma alargada en lanceolada, emergen después de las hojas cotiledonales. Estas hojas experimentan un crecimiento rápido y suelen ser propensas al ataque de insectos cortadores como la *Copitarsia turbata*, ello debido a que las plantas aún están tiernas.

c. Cuatro hojas verdaderas

Entre los 25 y 30 días después de la siembra, se produce una fase en la cual las plántulas de quinua exhiben una buena tolerancia a la sequía y a bajas temperaturas. Sin embargo, también son susceptibles al ataque de plagas masticadoras como *Epitrix subcrinita*. Es destacable que durante esta etapa aún se pueden observar las hojas cotiledonales de color verde por su estado tierno.

d. Seis hojas verdaderas

Ocurre en un rango de 35 hasta 45 días posteriores a la siembra, en esta fase se puede apreciar de forma extendida tres pares de hojas, además se puede observar que las hojas adultas protegen al ápice vegetativo especialmente cuando la planta está expuesta al descenso de temperatura o deficiencia de agua. Así mismo es importante resaltar que en esta fase las hojas cotiledonales ya se ponen amarillentas.

e. Ramificación

Ocurre entre los 45 hasta los 50 días posteriores a la siembra, se puede visibilizar ocho hojas verdaderas que se encuentran extendidas. En esta fase ya se puede observar la inflorescencia (panoja) que se encuentran protegidas por un conjunto de hojas, además de

ello se observa que las hojas de la parte axilar se extienden llegando al tercer nudo del tallo, y las cotiledonales se caen, dejando marcas en el tallo. Es importante destacar en esta fase, la planta puede ser afectada por las bajas temperaturas, por lo que los agricultores suelen complementar con fertilizantes y realizar el aporque.

f. Inicio de panojamiento

Se presenta entre los 55 hasta los 60 días posteriores a la siembra, donde la inflorescencia de la planta emerge desde el ápice y a su alrededor se observa un conjunto de hojas de tamaño pequeño los cuales cubren casi la tercera parte de la panoja; por otro lado, también resaltar que ya se evidencia el amarillamiento de las primeras hojas verdaderas, así como un acelerado engrosamiento y crecimiento del tallo. Por otro lado, cabe señalar que suele presentarse el ataque de la plaga conocida con el nombre de qhona qhona (*Eurissacca quinoae*) que suele minar y enrollar las hojas formando nidos.

g. Panojamiento

Se presenta entre los 65 hasta los 70 días posteriores a la siembra, donde parte que destaca con claridad es la inflorescencia sobre las hojas, además se observa los glomérulos que lo constituyen.

h. Inicio de floración

Se presenta entre los 75 hasta los 80 días después de haber realizado, se observa que las flores hermafroditas ubicadas en la apical de la panoja se van abriendo hasta exponer los estambres de la flor de manera separada y muy visible los glomérulos. Cabe señalar que en este periodo también la planta es muy sensible a las heladas y el estrés hídrico.

i. Floración

Denominada también como antesis y se presenta entre los 90 hasta los 100 días posteriores a la siembra, ocurre cuando la mitad de todas las flores que conforma la inflorescencia, están abiertas, la misma se puede percibir al promediar el medio día;

asimismo, es importante señalar que en esta fase las flores están susceptibles a ser afectadas por temperaturas negativas y temperaturas superiores a los 38°C las que pueden producir un aborto en las flores.

j. Grano lechoso

Se presenta entre los 100 hasta los 130 días posteriores a la siembra, es cuando el grano de los glomérulos de la panoja al ser aplastado revienta exponiendo un líquido blanquecino muy similar al color de la leche.

k. Grano pastoso

Esta etapa se produce entre los 130 y 160 días después de la siembra. Se caracteriza por la consistencia pastosa y color blanquecino de los granos cuando se los presiona. También puede haber presencia de plagas como la qhona qhona (*Eurisacca quinoae*) que podría causar daños en el cultivo.

l. Madurez fisiológica

Se presenta entre los 160 hasta 180 días posteriores a la siembra, en esta fase se puede observar que al ser presionado el grano con las uñas demuestra una dureza y su humedad varía entre 14 y 16%. Es importante señalar también que se presenta el amarillamiento y defoliación completa de las hojas.

2.1.6. Evaluación y caracterización de la quinua.

a. Caracterización fenológica

Este procedimiento se realiza respecto a las fases fenológicas de la planta, las cuales van relacionadas con la parte vegetativa y está comprendida desde el momento de la siembra, desarrollo y su maduración. Es importante señalar la parte fenológica de la quinua está representada por: emergencia, ramificación, panojamiento, floración, formación de grano y su madurez fisiológica (Rojas et al., 2014).

b. Caracterización de planta

Se realiza considerando su crecimiento, su altura suele ser variable desde 30 a 300 cm, sujeto a varios factores, tales como: tipo de crecimiento, genotipo, condiciones ambientales, fertilidad del suelo, inclusive su diferenciación con respecto a su origen, que pueden ser quinuas de valle o del altiplano.

c. Caracterización de panoja

Se realiza observando la inflorescencia de la planta, la misma que se encuentra en el ápice de la planta, para lo cual se toma mediciones de su tamaño respecto a su longitud, espesor y está relacionado al tipo de panoja que presenta la planta.

d. Caracterización del grano

Se realiza sobre los frutos de la quinua, que es un fruto en forma de aquenio que está cubierto por el perigonio, los cuales se desprenden fácilmente cuando se frotran cuando están secas. Hay tres formas de granos: cónica, cilíndrica u ovalada (Tapia et al., 2014) y que de acuerdo a la normatividad, se pueden considerar tres tamaños: grande (2.4 ± 2 mm), mediana (1.95 ± 0.15 mm) y pequeña inferior a 1.8 mm (PromPerú et al., 2009).

2.1.7. Análisis de componentes principales

Vertel et al. (2016) indica que el análisis de componentes principales (ACP) es muy utilizado para el procesado de datos de tipo cuantitativo, el objetivo de este análisis es comparar los valores de cada uno de los individuos que representan a las unidades estadísticas que están representados en cada una de las variables numéricas, las mismas que se tienen que reducir en dimensionalidad; así mismo, una alta correlación entre las variables ocasiona que un número menor de variables concentren lo más importante de todas las variables originales. Además de ello según lo descrito por Blanco (2015) acerca de ACP, indica que es una técnica estadística cuyo objetivo principal es la transformación de un conjunto de p variables que están correlacionadas a otro nuevo conjunto con un número

menor que p variables no correlacionadas, su cálculo considerando una matriz de tamaño $n \times p$ con sus datos, se define de la siguiente forma:

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix},$$

Esta notación, es considerando la población con la matriz de covarianza para las p variables y_p y están referidas a las variables y que son elementos de un vector aleatorio

$\bar{x} = (x_1, x_2, \dots, x_p)$, con el ACP se construye las siguientes combinaciones lineales:

$$\begin{aligned} y_1 &= \bar{a}'_1 \bar{x} = a_{11}x_1 + a_{12}x_2 + \dots + a_{1p}x_p, \\ y_2 &= \bar{a}'_2 \bar{x} = a_{21}x_1 + a_{22}x_2 + \dots + a_{2p}x_p, \\ &\vdots \\ y_p &= \bar{a}'_p \bar{x} = a_{p1}x_1 + a_{p2}x_2 + \dots + a_{pp}x_p \end{aligned}$$

En estas combinaciones se busca que la primera componente y_1 explique la mayor varianza posible; sin embargo, está sujeto a las restricciones de los elementos \bar{a}_1 además resulta de la combinación lineal de las primeras variables en \bar{x} . Para el cálculo de las componentes principales, se asume que $\|\bar{a}_1\| = 1$ cuando se haya encontrado la solución y el vector óptimo de pesos $\bar{a}_1' \bar{x}$, la misma que es llamada como la primera componente principal, luego de ello se va encontrando las demás componentes hasta y_p .

Para la extracción de los factores, se inicia desde la primera componente, de tal manera que y_1 posea la varianza máxima posible que está sujeto a la restricción $\|\bar{a}_1\| = 1$, para ello corresponde el desarrollo del siguiente cálculo:

$$\text{Var}(y_1) = \text{Var}(\bar{a}_1' \bar{x}) = \bar{a}_1' \sum \bar{a}_1$$

Luego de ello, siguiendo un proceso de maximización de cada una de las expresiones se logra la determinación de lambda λ que sería la varianza de y_1 .

2.1.8. Análisis de agrupamiento o clustering

Es conocido también como “análisis conglomerado” y busca agrupar los datos con el objetivo de lograr una mayor similitud de elementos dentro de cada grupo y la mayor diferencia entre cada grupo, es decir su propósito es identificar los grupos y dividir un conjunto de observaciones en diferentes grupos o conglomerados, de manera que las observaciones dentro de cada grupo muestren similitud, mientras que los datos en diferentes grupos no son similares. Los datos se representan como $n \times p$, una matriz

$y = [Y_{ij}]$ donde:

$$\mathbf{Y}_{n \times p} = \begin{bmatrix} Y_1' \\ Y_i' \\ \vdots \\ Y_n' \end{bmatrix}$$

El objetivo de este análisis por conglomerados, es desarrollar un esquema de clasificación que dividirá las filas de \mathbf{Y} en k grupos distintos (clusters). Las filas de la matriz suelen representar elementos u objetos. Para descubrir las agrupaciones en los datos, es necesario definir la medida. Es importante indicar que las dos medidas naturales de cercanía son la distancia o "disimilitud" entre grupos y el grado de asociación o "similitud". La elección de la escala aproximada depende del objetivo, la escala de medición puede ser ordinal, nominal, de intervalo o de proporción y el tipo de variable que se analiza puede ser continua o categórica. Muchas aplicaciones del análisis de conglomerados comienzan con una matriz de vecindario en lugar de una matriz; por ello, dada la matriz de proximidad de orden $(n \times n)$, las entradas pueden representar diferencias $[d_{rs}]$ o similitudes $[S_{rs}]$ entre los objetos r^{th} y s^{th} .

Medidas de proximidad

Se utilizan para representar la proximidad de dos objetos. Si una medida de proximidad representa similitud, el valor de la medida aumenta a medida que dos objetos se vuelven más similar. Alternativamente, si la medida de proximidad representa disimilitud, el valor de la medida disminuye de valor a medida que dos objetos se vuelven más parecidos. Dejando que y_r e y_s representen dos objetos en un espacio p-variable, un ejemplo de una medida de disimilitud es el euclidiano: distancia entre años y años. Como medida de similitud, se puede utilizar la proporción de los elementos de los dos vectores que coinciden. Formalmente, es necesario establecer un conjunto de axiomas matemáticos para crear las medidas de similitud.

Además de ello, para representar la proximidad de dos objetos, si una medida de proximidad representa similitud entonces el valor de la medida aumenta a medida que dos objetos se vuelven más similares. Alternativamente, si la medida de proximidad representa disimilitud, el valor de la medida de valor a medida que los objetos disminuyen se vuelven más parecidos.

Por ejemplo, dados dos objetos y_r e y_s en un espacio p-dimensional, una medida de disimilitud satisface las siguientes condiciones.

- 1) $d_{rs} \geq 0$ para todos los objetos y_r e y_s
- 2) $d_{rs} = 0$ si y solo si $y_r = y_s$
- 3) $d_{rs} = d_{sr}$

La condición 3) implica que la medida es simétrica de modo que la medida de disimilitud que compara y_r (objeto r) con y_s (objeto s), es lo mismo que la comparación para el objeto “s” versus objeto “r” la condición 2) requiere que la medida sea cero siempre que el objeto r sea igual al objeto s, los objetos son idénticos solo si $d_{rs} = 0$ y en ninguna

otra situación. Finalmente, 1) implica que la medida nunca es negativa. Una medida de disimilitud que satisface las condiciones 1) para 3) se dice que es una medida semi métrica.

Para las variables continuas (intervalo, escala de razón), la medida de disimilitud más común es la distancia euclidiana entre dos objetos y el cuadrado de la distancia euclidiana entre dos filas y_r e y_s está definido como:

$$d_{rs}^2 = (y_r - y_s)'(y_r - y_s) = \|y_r - y_s\|^2$$

La matriz de datos ($n \times n$) $D = [d_{rs}]$ se denomina matriz de distancia euclidiana. El cambio en las unidades de medida puede hacer que una variable domine la clasificación de distancias, la matriz de distancia euclidiana es más eficaz para las variables medibles; cuando las variables no lo son, se pueden ponderar las diferencias al cuadrado por:

$$s_j^2 = \sum_i^n 1(y_{ij} - \bar{y}_{.j})^2 / (n-1), j=1,2,\dots,p \text{ donde } s_j^2 \text{ y } \bar{y}_{.j} \text{ representan las}$$

estimaciones del promedio y varianza referida a la variable j .

$$d_{rs}^2 = (y_r - y_s)'(\text{diag } S)^{-1} (y_r - y_s)$$

Este proceso elimina la dependencia del análisis de las unidades de medida. Sin embargo, a menudo hace que las distancias dentro del clúster sean más grandes que las distancias entre clústeres, las diferencias tienden así a enmascarar las agrupaciones (Hartigan, 1975). Este problema también ocurre cuando se usa el cuadrado de las distancias de Mahalanobis como una medida de proximidad en la métrica de la matriz de covarianza S definida como:

$$d_{rs}^2 = (y_r - y_s)' S^{-1} (y_r - y_s)$$

Las distancias de Mahalanobis, independientemente de que las variables sean acordes o no, tienden a enmascarar los conglomerados más aún, dado que las correlaciones tienden a reducirse aún más (Hartigan, 1975).

No existe una solución favorable para un problema de la unidad de medida inconsistente, debido a que la distancia euclidiana es un caso especial de la métrica de Minkowski, las medidas de disimilitud pueden representarse como:

$$d_{rs} = \left(\sum_{j=1}^p |y_{rj} - y_{sj}|^\lambda \right)^{1/\lambda}$$

La variación de λ en (a) cambia el peso asignado a distancias mayores y menores. Para $\lambda = 1$ y para $\lambda = 2$, tenemos distancias euclidianas. Al definir las medidas de disimilitud, no incluimos la condición de desigualdad triangular

que $d_{rs} \leq d_{rq} + d_{qs}$ para todos los puntos r, s y q . Con esta condición, la disimilitud las medidas forman una métrica. Si bien todas las distancias de Minkowski forman una métrica, el requisito es una condición suficiente pero no una condición necesaria para una medida de proximidad. Por ejemplo, podemos agrupar utilizando la matriz de distancias euclidianas al cuadrado que no forman una métrica o use la norma $d_{rs} = \|y_r - y_s\|$, que forma una métrica con la condición de que $d_{rs} \leq \max(d_{rq}, d_{qs})$ para todos los puntos r, s y q .

Tabla 1

Tabla de contingencia 2x2 de variable binaria

		Fila s		Total
		1	0	
Fila r	1	a	b	a+b
	0	c	d	c+d
Total		a+c	b+d	p=a+b+c+d

El sentido semi métrico se convierte en ultra métrico ya que esta condición es más fuerte que el requisito de desigualdad triangular para una métrica.

Para reducir el tamaño de las distancias al cuadrado (o distancias), se dividen por el número de variables “p”, debido a que todas las diferencias se dividen por el mismo número

“p”, este no afecta los resultados de la agrupación. Además, las distancias euclidianas al cuadrado son a menudo media centrada restando de cada variable su media correspondiente. Entonces $Y_d = [y_{ij} - \bar{y}_{.j}]$ es analizado. La centrada variable no afecta las distancias euclidianas desde que se eliminó las medias de las “p” variables no cambian la distancia euclidiana entre dos objetos.

Otras dos medidas de disimilitud que se han propuesto cuando todas las y_{ij} son positivas son el seguimiento.

$$\text{Camberra Metric } d_{rs} = \sum_{j=1}^p \left\{ \frac{|y_{rj} - y_{sj}|}{(y_{rj} + y_{sj})} \right\}$$

$$\text{Czekanowski Coefficiente } d_{rs} = \frac{\sum_{j=1}^p |y_{rj} - y_{sj}|}{\sum_{j=1}^p (y_{rj} + y_{sj})}$$

Ambas medidas son normas L1 modificadas y se utilizan cuando los datos están sesgados y / o contener valores atípicos. Una norma L1 escalada que se puede utilizar para datos que incluyen tanto positivos y los valores negativos es la métrica de Gower.

$$\text{Gower Metric } d_{rs} = \sum_{j=1}^p |y_{rj} - y_{sj}| / R_j$$

donde R_j representa el rango de la variable j , se asume que el desarrollo de algunas medidas de disimilitud asumiendo que los datos son continuos con niveles de medición, al menos son de intervalo. Para datos categóricos que tienen valores nominales o escalas ordinales de medida, la situación se vuelve más compleja, (Anderberg, 1973). Para un caso simple, suponga que cada fila y_i' de Y contiene solo datos binarios. Entonces las distancias euclidianas al cuadrado solo proporcionan un recuento de desajustes, (1 - 0) o (0 - 1) y las coincidencias, (1 - 1) y (0 - 0) por igual. Cuando las variables se codifican 0 o 1 a indicar la ausencia o presencia de una característica, una tabla (2 × 2) como se muestra en la Tabla 1. pueden crearse para evaluar medidas de similitud y disimilitud.

En la Tabla 1, las frecuencias b y c representan desajustes mientras que las frecuencias a y c representan coincidencias. Así, la distancia euclidiana al cuadrado dividida por p se convierte en:

$$\sum_{j=1}^p (y_{rj} - y_{sj})^2 / p = (b + c) / p = d_{rs}^2 / p$$

2.1.9. Análisis Discriminante.

Se emplea en el ámbito de la estadística y el aprendizaje automático para clasificar un conjunto de observaciones o elementos en distintos grupos o categorías, utilizando un conjunto de variables predictoras. El objetivo es establecer una regla que permita discriminar los elementos y asignarlos a una de las varias categorías posibles. Para lograr esto, el investigador debe obtener una muestra aleatoria de los elementos experimentales para cada posible grupo de clasificación. (Vertti, 2019).

Definición de la función discriminante (Fisher).

Se obtiene mediante una función lineal proveniente de un número de k variables denominadas explicativas denotadas por X, tal como se muestra a continuación:

$$Z = K_1X_1 + K_2X_2 + K_3X_3 + \dots + K_nX_n$$

Donde:

Z: Corresponde al puntaje discriminante o puntaje discriminatorio, el cual es un valor numérico utilizado para llevar a cabo la clasificación.

K: Indica el factor de ponderación o importancia que se asigna a cada una de las variables para llevar a cabo la discriminación.

X: Son las variables independientes.

Esta función permitirá obtener variables significativas para que se puedan diferenciar los grupos; lo que va permitir encontrar pesos de cada una de las variables.

En este análisis discriminante, el objetivo es la separación de las medias entre los grupos, mientras se busca aumentar la razón de dispersión de los datos.

Tipos de análisis discriminante

Según la cantidad de grupos referidos a la variable que es dependiente, el modelo discriminante puede clasificarse en múltiple o simple. Cuando la variable dependiente tiene dos grupos, se utiliza una técnica de discriminación simple. Por otro lado, si la variable dependiente tiene más de dos grupos, se emplea un método discriminante múltiple.

Modelo discriminante simple

La obtención de una ecuación para el discriminante simple con dos variables independientes se representa mediante el siguiente término:

$$Z = K_1X_1 + K_2X_2$$

La derivación de este modelo intenta averiguar los valores del peso correspondiente a todas las variables a discriminar; es decir, de los valores que tiene K. Cabe recordar que el criterio es aumentar la distancia a la media de cada grupo, tomando en cuenta que los datos están dispersos. Para continuar con la derivación de este modelo discriminante simple y calcular los valores que tiene K, procedemos a resolver las siguientes ecuaciones:

$$\begin{aligned} \sum X_1'^2 K_1 + \sum X_1' X_2'^2 K_2 &= \bar{X}_1 | \text{Gpo.1} - \bar{X}_1 | \text{Gpo.2} \\ \sum X_1' X_2' K_1 + \sum X_2'^2 K_2 &= \bar{X}_2 | \text{Gpo.1} - \bar{X}_2 | \text{Gpo.2} \end{aligned}$$

Donde:

$$\begin{aligned} \sum X_1'^2 &= \sum (\sum X_1^2 - n\bar{X}_1^2) \\ \sum X_2'^2 &= \sum (\sum X_2^2 - n\bar{X}_2^2) \\ \sum X_1' X_2' &= \sum (\sum X_1 X_2 - n\bar{X}_1 \bar{X}_2) \end{aligned}$$

En donde $\sum X_i'^2$ hace referencia a la varianza que tienen los datos correspondientes a la variable dentro del conglomerado, por otro lado: $\sum X_i' X_j'$ constituye la covarianza de

los datos referidos a las variables. Es importante indicar que la varianza es la dispersión que tiene los datos en relación a su promedio y la covarianza refleja la dispersión que tiene los datos respecto a las variables que se tratan de manera agrupada.

La derivación de la ecuación referida a la discriminante, teniendo en cuenta las tres variables independientes se representa tal como sigue:

$$Z = K_1X_1 + K_2X_2 + K_3X_3$$

Para efectuar la derivada de este modelo, se tiene que determinar los valores referidos a los pesos de cada una de las variables que permiten realizar la discriminación, vale decir aquellos valores que tienen las K. Es importante señalar que el procedimiento consiste en incrementar las distancias de las medias de cada uno de los grupos, tomando en cuenta que los datos presentan dispersión. Así mismo, para efectuar la derivada de este modelo y determinar el valor de K se tiene que efectuar las siguientes ecuaciones:

$$\sum X_1'^2 K_1 + \sum X_1' X_2' K_2 + \sum X_1' X_3' K_3 = \bar{X}_1 | \text{Grupo1} - \bar{X}_1 | \text{Grupo2}$$

$$\sum X_1' X_2' K_1 + \sum X_2'^2 K_2 + \sum X_2' X_3' K_3 = \bar{X}_2 | \text{Grupo1} - \bar{X}_2 | \text{Grupo2}$$

$$\sum X_1' X_2' K_1 + \sum X_2' X_3' K_2 + \sum X_3'^2 K_3 = \bar{X}_3 | \text{Grupo1} - \bar{X}_3 | \text{Grupo2}$$

Donde:

$$\sum X_1'^2 = \sum \left(\sum X_1^2 - n \bar{X}_1^2 \right)$$

$$\sum X_2'^2 = \sum \left(\sum X_2^2 - n \bar{X}_2^2 \right)$$

$$\sum X_3'^2 = \sum \left(\sum X_3^2 - n \bar{X}_3^2 \right)$$

$$\sum X_1' X_2' = \sum \left(\sum X_1 X_2 - n \bar{X}_1 \bar{X}_2 \right)$$

$$\sum X_1' X_3' = \sum \left(\sum X_1 X_3 - n \bar{X}_1 \bar{X}_3 \right)$$

$$\sum X_2' X_3' = \sum \left(\sum X_2 X_3 - n \bar{X}_2 \bar{X}_3 \right)$$

En Donde $\sum X_i'^2$ hace referencia a la varianza de todos los datos de la variable analizada considerado en un conjunto de datos que forma el grupo y la $\sum X_i'X_j'$ indica la covarianza que tienen los datos en referencia a las variables en cada uno de los grupos.

El modelo en referencia podría generalizarse en casos de que en el modelo se analicen una mayor cantidad de variables independientes, tal como se muestran en la siguiente expresión:

$$Y_1 = K_1X_1 + K_2X_2 + K_3X_3 + \dots + K_nX_n$$

Al desarrollar este sistema de ecuaciones, se consigue como resultado la ecuación discriminante. Luego de ello, se prosigue con la validación del modelo para su prueba.

Prueba de hipótesis para el modelo

Para realizar la prueba se utiliza el test de Mahalanobis, para encontrar el valor de F-calculada y luego compararlo con la F-crítica de las tablas. El planteamiento de las hipótesis sería las siguientes:

H_0 : El Modelo discriminante no es adecuado para clasificar las observaciones

H_1 : El Modelo discriminante es adecuado para clasificar las observaciones.

Además, D^2 (distancia) se obtiene mediante la siguiente fórmula :

$$D^2 = (n_1 + n_2 - 2) \left[\sum K_i (\bar{X}_i|_1 - \bar{X}_i|_2) \right]$$

Donde D^2 cuyo resultado es un número positivo, debido a que es resultado de las medias. Luego D^2 se contrasta con la F-crítica a fin de determinar la conclusión de rechazar o no rechazar la H_0 , es por ello que se tiene que transformar D^2 en F-calculada, desarrollando la siguiente fórmula:

$$F_{\text{calculada}} = \frac{(n_1 + n_2 - p - 1)(n_1 * n_2)D^2}{[(n_1 + n_2 - 2)p] * (n_1 + n_2)}$$

Posteriormente F-calculada se contrasta con la F-crítica de la tabla considerando los siguientes grados de libertad:

$$F_{\text{crítica}} = [\alpha, p, (n_1 + n_2 - p - 1)]$$

En donde:

n_1 : Cantidad de datos operados en el grupo 1

n_2 : Cantidad de datos operados en el grupo 2

p : Cantidad de grupos operados con la variable dependiente.

Luego de haber terminado con la determinación del modelo, es importante realizar las clasificaciones para su validación.

Etapas de validación del modelo discriminante

En la distribución normal, es común que los resultados referidos a los grupos se superpongan, por consiguiente, se requiere establecer un límite para separar estos resultados. Este límite se conoce como $Z_{\text{crítica}}$ y representa el punto de separación en la línea divisoria. El objetivo es minimizar la superposición de los últimos valores de cada grupo, logrando así una mayor separación entre los promedios de los grupos, los pasos para validar son:

1. Se calcula Z promedio por cada grupo, lo cual se consigue reemplazando las medias de las variables que son de tipo independiente y que corresponden a cada uno de los grupos en la ecuación obtenida.
2. Así mismo, al calcular $Z_{\text{crítica}}$ por medio de la siguiente fórmula, cuando $n_1 = n_2$ se tiene:

$$Z_{\text{crítica}} = (Z_1 + Z_2) / (2)$$

Cuando $n_1 \neq n_2$, se tiene que calcular el promedio ponderado y para $Z_{\text{crítica}}$ la fórmula sería la siguiente:

$$Z_{\text{crítica}} = (n_2 Z_1 + n_1 Z_2) / (n_1 + n_2)$$

3. Luego procedemos a calcular el valor de Z para cada una de las observaciones que están acertadas y se realiza una predicción sobre qué grupo pertenece cada una de ellas, basándose en el valor crítico de $Z_{crítica}$.
4. Se comprueba si esta clasificación es precisa o incorrecta; es decir, se realiza una evaluación de la exactitud de la clasificación.
5. Basado en lo mencionado anteriormente, se determina las clasificaciones que son correctas (porcentaje), la misma que podría representarse una medida de confiabilidad o certeza al realizar pronósticos.
6. A continuación, se prosigue con el cálculo del porcentaje mínimo de los grupos que fueron correctos "C" mediante la fórmula siguiente:

$$C = \sum p_i^2 \text{ o también } p_i^2 + (1 - p_i)^2$$

En donde:

p_i : Es el porcentaje de los datos que puede tener cualquier grupo.

7. Para concluir, se hace una comparación respecto al porcentaje de clasificaciones correctas y el umbral mínimo aceptable de agrupaciones correctas.

Supuestos del análisis discriminante:

Los datos deben seguir una distribución normal, lo que implica que la mayor parte de ellos se encuentren alrededor de su promedio.

En cuanto a la dispersión que tienen los datos, es necesario que las varianzas y las covarianzas sean similares entre los grupos. Esto implica que el experimento se realice en condiciones consistentes y que no haya fuentes de variación que provoquen diferencias significativas entre los grupos.

Al llevar a cabo una clasificación, se parte del supuesto de que el resultado obtenido es confiable, tal como se refleja en el porcentaje de las agrupaciones correctas que son obtenidas durante la validación del modelo planteado.

2.2. Marco conceptual (Palabra clave)

2.2.1. *Chenopodium quinoa* Willd.

El nombre común es “quinua” y es también denominada pseudo cereal, parte de los granos andinos, ha sido objeto de varios estudios en cuanto a su genética utilizando marcadores moleculares. Estos estudios sugieren que la quinua ha experimentado procesos de domesticación en dos ocasiones: la primera en zona del altiplano entre Perú y Bolivia y la segunda en la costa del país de Chile. (Christensen et al., 2007)

2.2.2. Fenología.

La fenología es un campo de estudio que se centra en el momento en que ocurren eventos periódicos en el proceso de desarrollo de algún organismo, como la germinación crecimiento, floración, caída de las hojas en plantas y termino de ciclo de vida, todo ello está relacionado con factores ambientales.

La definición original de fenología vegetal se le atribuye a Carlos Linneo en el año 1751. En ese año, Linneo publicó una metodología para crear calendarios botánicos anuales que puedan servir como guía para observar las etapas biológicas de las plantas en referencia a las condiciones climáticas a las que estaban expuestas. Su objetivo inicial era desarrollar un calendario del ciclo de vida que complementara el calendario astronómico (Heuveldop et al., 1986).

2.2.3. Diversidad genética.

Es la cantidad completa de las características morfogénicas presentes en la conformación de individuos dentro de una especie, esta disminuye cuando ocurre un evento conocido como "cuello de botella", en el cual la población disminuye significativamente quedando una mínima cantidad de individuos. Es importante conocer que, cuanto mayor sea la diversidad genética, más probabilidades tendrá una especie de sobrevivir ante las adversidades ambientales. De la misma forma, también se define como la heterogeneidad de

los tipos de genes existentes en una especie o un conjunto de especies, constituyéndose como la biodiversidad (Jiménez et al., 1999). Así mismo, según lo presentado por Moreno (2001), la diversidad también se le como biodiversidad y se define como “la variabilidad presente entre los seres ya sean del medio terrestre, marino o acuático” los mismos que forman comunidades en diferentes hábitats del mundo.

2.2.4. Técnica multivariada.

Permite examinar las diferentes características provenientes de algún individuo y, debido a su interrelación, no tiene sentido analizar su efecto de forma aislada. Estas técnicas se utilizan con mayor frecuencia en estudios observacionales sobre las causas, la predicción de enfermedades, entre otros, porque posibilita explicar, pronosticar y revisar aquellas variables que pueden distorsionar o sesgar los resultados, lo que a su vez brinda una mejor seguridad en la inferencia estadística (Sagaró y Zamora, 2020).

Asimismo, se define como un conjunto de técnicas estadísticas diseñadas para analizar conjuntos de datos multivariados de manera simultánea. Dado que se registran múltiples variables para cada caso, estas técnicas permiten una comprensión más completa del fenómeno bajo investigación y brindan información que no está disponible mediante métodos univariados o bivariados. (Collet et al., 2004).

2.2.5. Análisis Clúster.

Es también conocido como el análisis para agrupar, esta técnica multivariada que pretende conglomerar elementos o variables con el objeto de conseguir la alta similitud dentro del conglomerado y la máxima disimilitud respecto a cada uno de los conglomerados (De la Fuente, 2011).

De manera similar, esta técnica pretende agrupar a un conjunto de observaciones en un número determinado de conglomerados o grupos. Este proceso se basa en la noción de

distancia o similitud entre las observaciones, y su objetivo principal es lograr una agrupación significativa y coherente de los datos (Marín y Rodríguez-Bernal, 2012).

2.2.6. Análisis discriminante.

Esta técnica está orientada a clasificar a diferentes personas o individuos en distintos grupos o poblaciones según los valores de múltiples variables asociadas a las personas que están siendo clasificadas. Al final del proceso, cada persona se asigna únicamente a un grupo (De la Fuente, 2011).

Además, se define como una técnica de clasificación cuyo propósito es desarrollar una función que pueda asignar nuevos individuos a grupos conocidos en función de los valores de variables específicas. Haciendo un contraste con el análisis de conglomerados, los diferentes grupos obtenidos en el análisis discriminante ya deben ser conocidos previamente y se requiere tener información sobre a qué grupo pertenece cada individuo en particular, así como los valores de las variables discriminantes (Visauta, 1998).

2.3. Antecedentes de la investigación.

2.3.1. Antecedentes internacionales

Debido a su gran diversidad de especies, respecto a los 17 países que cuentan con mayor biodiversidad en el mundo, el Instituto de Investigaciones Agropecuarias - INIAP del Ecuador hizo la investigación denominada “Caracterización morfológica de quinua en tres provincias de la sierra ecuatoriana” con el objetivo de conocer la diversidad genética y agrobiodiversidad de la quinua en un ámbito de 03 provincias, para ello se ha utilizado 40 descriptores de Bioversity International; se recogió datos durante el ciclo de desarrollo de la planta, evaluándose cada una de las parcelas, para realizar el análisis estadístico se ha utilizado el software estadístico SPSS, aplicando a los datos de caracterización con descriptores cuantitativos y cualitativos. Los resultados revelaron una amplia diversidad entre las accesiones estudiadas, lo que permitió reconocer y catalogar un total de 211

colecciones. Estos hallazgos confirman la trascendencia de conservar estos recursos fitogenéticos en Ecuador (MINAGRI et al, 2017).

En Bolivia, Cruz (2016) hizo una investigación denominada Caracterización de variables continuas y discretas del grano de quinua (*Chenopodium quinoa* Willd.) del Banco de Germoplasma de la Estación Experimental Patacamaya, con el objetivo de caracterizar morfológicamente las accesiones de quinua conservadas en el Banco de Germoplasma de granos altoandinos, provenientes del altiplano Centro (La Paz y Oruro) para ello, en el recojo de información ha utilizado 40 descriptores internacionales, los cuales fueron analizados utilizando estadísticas descriptivas, análisis de correlación simple, componentes principales y un análisis de frecuencias.

En Colombia en una investigación realizada por Morillo et al.(2022) denominada “Phenotypic diversity of agromorphological characteristics of quinoa (*Chenopodium quinoa* willd.) germplasm in Colombia” se evaluaron 50 genotipos del cultivo de quinua procedentes de las colecciones de germoplasma de Boyacá, al tomar datos de sus características agronómicas y morfológicas, para lo cual se aplicó un diseño DCA, análisis multivariados seguidos de un enfoque de agrupamiento utilizando descriptores agromorfológicos, de los cuales 16 eran descriptores cualitativos (color del epispermo, forma de hojas, forma de la panícula, etc.) y cinco eran descriptores cuantitativos (N° de panículas, altura de la planta, etc.). Respecto a las conclusiones, se determinó que las características asociadas al rendimiento presentaban una mayor variabilidad. El PCA de las variables de tipo cuantitativo reveló que el primero y segundo componente explicaban hasta un 88% de la variación total, donde las variables: cantidad, diámetro y longitud de panícula, así como altura de planta, fueron los rasgos con mayor variabilidad. Respecto a los grupos de características cuantitativas incluyeron: “altura de planta”, “diámetro”, “longitud de panícula” y “peso de mil semillas”; mientras que los grupos de características cualitativas

incluyeron “forma del tallo”, “hábito de ramificación”, “forma de la panícula” y “color de las axilas”. El análisis factorial de datos mixtos permitió discriminar los materiales con destacadas características morfoagronómicas, esto reveló una amplia variabilidad que debe conservarse y utilizarse en programas de mejoramiento genético del cultivo de la quinua.

2.3.2. Antecedentes nacionales

En la Región Puno, Cárdenas (2017), hizo un trabajo de investigación denominada “Caracterización molecular de 129 accesiones de quinua (*Chenopodium quinoa* Willd.) de la Región Puno mediante marcadores microsatélites” con una cantidad de 129 colecciones de quinua, realizándose la caracterización de diversidad a nivel molecular considerando una cantidad de 15 marcadores microsatélites. Se han logrado identificar 179 alelos entre las colecciones, se observaron características moleculares que formaron 3 grupos genéticamente diferentes; sin embargo, el grado de los niveles de distinción genética respecto a los grupos se tornaron bajos, lo que indica que no hubo claridad en formación de los grupos según los análisis PCA. Así mismo, en las 129 colecciones no se registraron duplicados, lo que expresa una alta diversidad genética carente de una definición de una estructura poblacional. Este resultado forma parte de los primeros trabajos que permitirán la implementación de métodos orientado a un manejo apropiado de banco de germoplasma

De la misma manera, Dominguez (2014), en su estudio sobre el análisis de las características agronómicas y morfológicas, además de conocer como es la variabilidad de todo el grupo de generaciones autofecundadas (S2) correspondiente a las cruzas genéticamente cercanas y distantes del cultivo de la quinua. Se evaluaron un total de 46 caracteres agromorfológicos, de los cuales 21 eran cuantitativos y 25 eran cualitativos, según los descriptores de la quinua y sus parientes silvestres de Bioversity International. Sin embargo, solo 32 caracteres (21 cuantitativos y 11 cualitativos) se incluyeron en la base de datos excluyéndose 14 registros. También se realizaron análisis estadísticos simples y

análisis multivariados y se ha encontrado una alta variabilidad respecto a sus características tales como: diámetro de panoja con 31,33%, rendimiento por planta con 38,45% y N° de ramas primarias con 44,26%. En el ACP se ha calculado un total de 21 componentes; sin embargo, los cuatro componentes iniciales explican el 72% de la variabilidad de las colecciones analizadas, siendo el orden de las proporciones las siguientes: 40, 15, 9 y 6%. Estos componentes proporcionaron información sobre la discriminación de las variables, siendo las variables fenológicas las más distintivas. Finalmente, por medio del análisis clúster se ha generado un dendograma que visualizó claramente las diferencias entre las progenies, se utilizó distancia euclidiana como medida, lo que permitió visualizar la diversidad entre las progenies de manera clara.

Así mismo, Laura et al.(2023), en su tesis titulada “La gestión educacional y producción de quinua orgánica (*Chenopodium Quinoa* Wild) del banco de germoplasma de la universidad nacional del Altiplano según Bioversity Internacional”, hizo un estudio de con la finalidad de determinar la variabilidad de las colecciones del banco de germoplasma de quinua con 41 descriptores, encontrando 10 grupos distintos que agrupan a todas las colecciones analizadas.

III. Hipótesis y variables

3.1. Hipótesis

a. Hipótesis general.

La colección de quinua del Banco de Germoplasma de Puno analizada mediante técnicas multivariadas presenta alta diversidad genética.

b. Hipótesis específicas

- Las variables de la colección de quinua del Banco de Germoplasma de Puno, concentran su variabilidad con mayor proporción en los primeros componentes, mediante aplicación de técnicas multivariadas.
- La colección de quinua del Banco de Germoplasma de Puno, se diferencia claramente al formar grupos con la aplicación de la técnica de clúster.
- En los grupos de la colección de quinua en el Banco de Germoplasma de Puno, las variables referidas a producción de semilla y longitud de planta, son las que mejor discriminan mediante aplicación de la técnica discriminante.
- La cuantificación de la diversidad en la colección de quinua del Banco de Germoplasma de Puno es alta.

3.2. Identificación de variables e indicadores

Variable : Diversidad genética.

Dimensiones : Fenotipo planta, panoja y grano.

3.3. Operacionalización de variables

Tabla 2

Operacionalización de variables

VARIABLES	Definición Conceptual	Definición Operacional	Dimensiones	Indicadores
Diversidad genética	Es el conjunto de genes con funciones particulares acumulados, dentro de una misma especie; que en el tiempo ha desarrollado variantes por su evolución y efectos el medio ambiente. (Hidalgo, 2017).	La diversidad genética está en función a los aspectos de expresión fenotípica de una especie	Fenología	<ul style="list-style-type: none"> - Emergencia (Nº días) - Ramificación (Nº días) - Inicio de panojamiento(Nºdías) - Inicio de floración (Nºdías) - G. estado lechoso (Nºdías) - G. estado pastoso (Nºdías) - Madurez fisiológica(Nºdías)
			Planta	<ul style="list-style-type: none"> - Altura planta (cm) - Altura inicio de panoja (cm) - Diámetro de tallo (cm) - Color del tallo (*) - Número total hojas - Ancho de la hoja (cm) - Longitud de hoja (cm) - Hábito de crecimiento (*) - Número de dientes en hoja - Color de hojas (*)
			Panoja	<ul style="list-style-type: none"> - Longitud de panoja (cm) - Nº total de glomérulos - Color de panoja (*) - Diámetro central (cm) - Diámetro de la base (cm) - Forma de la panoja (*)
			Grano	<ul style="list-style-type: none"> - Tamaño de grano (*) - Forma de grano (*) - Color del perigonio del fruto(*) - Color del pericarpio (*) - Color del epispermo (*) - Contenido de saponina (cm) - Rendim. por planta (g/planta) - Peso de biomasa - Índice de cosecha por planta

Fuente: Elaborado propia, según el soporte teórico

(*) Variables categorizadas

IV. Metodología

4.1. **Ámbito de estudio: Localización política y geográfica**

Este estudio de investigación se ha llevado a cabo en el siguiente lugar:

- Lugar : CIP Camacani Puno.
- Distrito : Puno.
- Provincia : Puno.
- Región : Puno.
- Altitud : 3824 m.s.n.m.
- Latitud : 15 56' 34"
- Longitud : 69 51' 27"

Figura 1

Mapa de ubicación geográfica de la investigación.



4.2. Tipo y nivel de investigación.

Tipo : Básico.

Nivel de investigación : Descriptivo.

Diseño de investigación : No experimental, transversal descriptivo.

4.3. Unidad de Análisis

Colección correspondiente al Banco de Germoplasma de quinua de Puno.

4.4. Población de estudio

753 colectas de quinua del Banco de Germoplasma de Puno.

4.5. Tamaño de muestra

Se ha desarrollado con todas las colecciones que considera el Banco de Germoplasma de quinua de Puno.

4.6. Técnicas de selección de la muestra

Se ha considerado todos los datos de caracterización y evaluación registrados en el catálogo del Banco de Germoplasma de quinua de Puno (no probabilístico).

4.7. Técnicas de recolección de información

Técnica : Documental

Instrumento : Ficha de registro documental

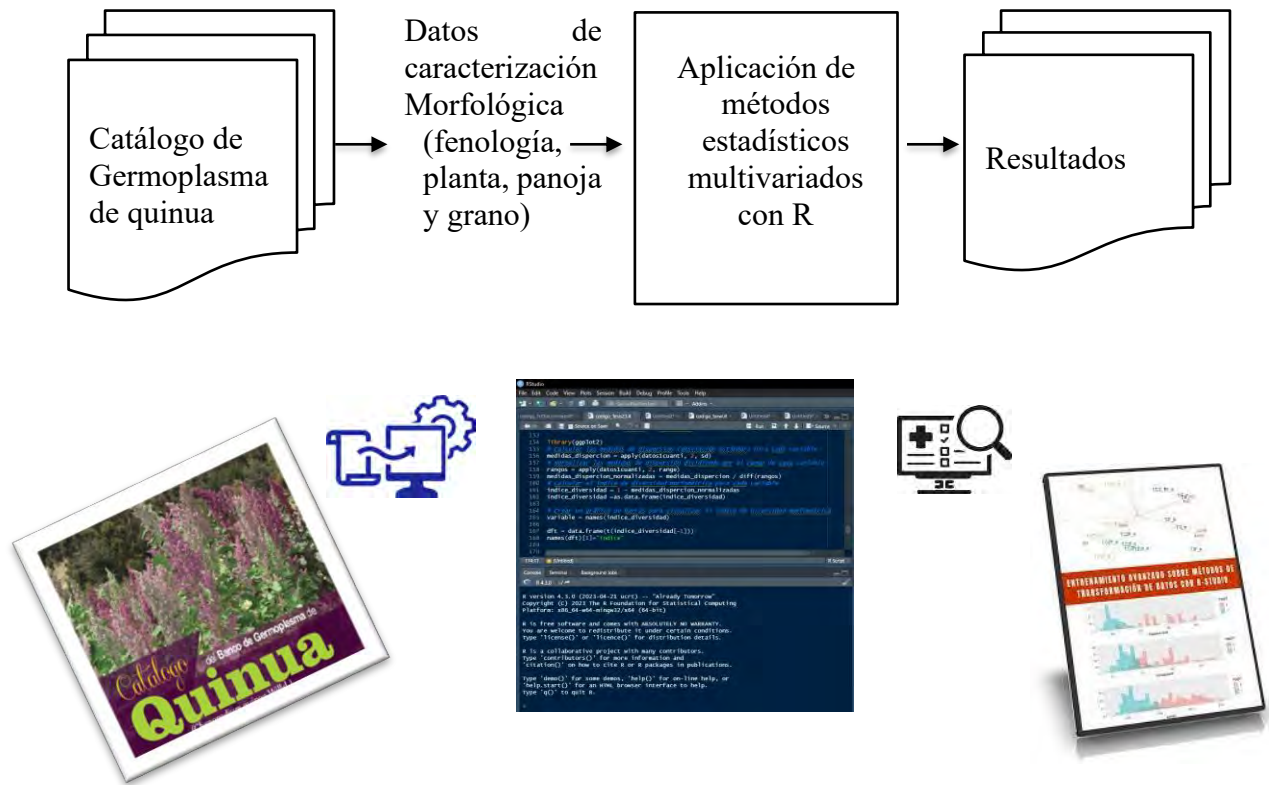
4.8. Técnicas de análisis e interpretación de la información.

Se ha tomado la información del documento denominado: “catálogo del Banco de Germoplasma de quinua de Puno”, las mismas que se han sistematizado para generar una sola base de datos, que se ha utilizado posteriormente para el

procesamiento estadístico utilizando el software estadístico R aplicando algoritmos para el análisis multivariado de PCA, clúster y análisis discriminante.

Figura 2

Proceso de análisis de datos.



V. Resultados y discusión

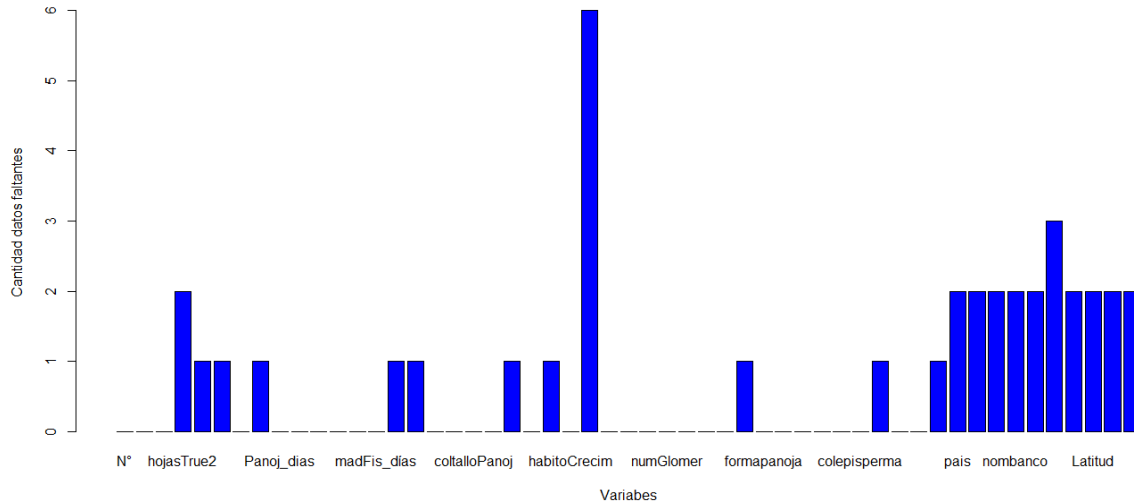
Toda la información de la colección del Banco de Germoplasma de quinua ha tenido que ser sistematizada en una base de datos para su análisis mediante técnicas multivariadas, para ello, se ha realizado los procedimientos que se detallan a continuación.

5.1. Preprocesamiento

Los datos de la caracterización y evaluación de la colección del Banco de Germoplasma de quinua, consta de los siguientes grupos de información:

- Características fenológicas
- Características de la planta
- Características de la panoja
- Características del grano

Inicialmente se tenían 753 registros contenidas en 51 variables tanto cuantitativos y cualitativos, para lo cual se hizo una revisión de todo el conjunto de datos, encontrándose un total de 39 datos faltantes que representan el 5.04% de los datos totales, los mismos que se distribuyen de la siguiente manera:

Figura 3*Distribución de datos faltantes.*

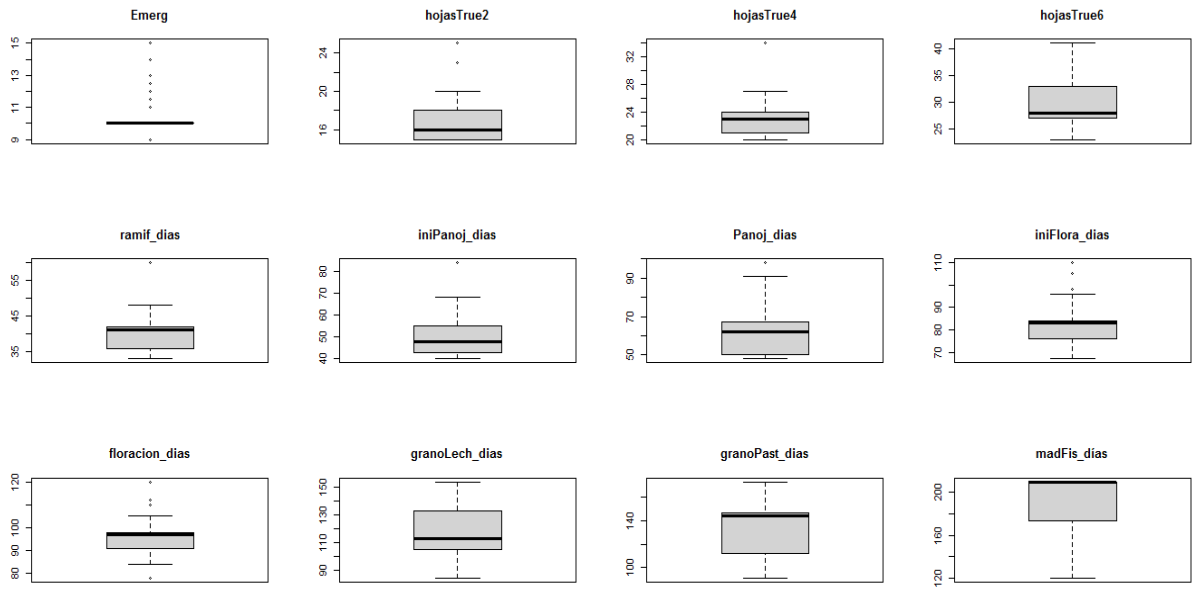
Donde la cantidad de datos faltantes, son menores al 10%, motivo por el cual se ha optado por eliminar aquellos registros que contiene a los datos faltantes, teniendo como resultado una matriz de 734 registros, los cuales cuentan con 31 variables cuantitativas y 20 son de tipo cualitativo.

5.2. Identificación de datos outliers.

Son conocidos también como datos extremos dentro del conjunto de datos que corresponden a cada uno de los descriptores que las denominamos “variables”. Es por tal motivo que, en el análisis realizado a cada una de las variables se observan varios valores dispersos, ello se debe a que corresponden colecciones diferentes, además indicar que cada colección tiene sus características propias distintas a las demás; motivo por el cual, las calificaciones suelen presentarse de manera heterogénea; es por ello que, cada una de las colecciones representa una riqueza de diversidad biológica dentro de la especie. Para verificar como es la distribución de las mediciones dentro de cada una de las variables, se tiene los siguientes gráficos:

Figura 4

Distribución de las mediciones en los descriptores de caracterización de fenología.

**Figura 5**

Distribución de las mediciones en los descriptores de caracterización de planta.

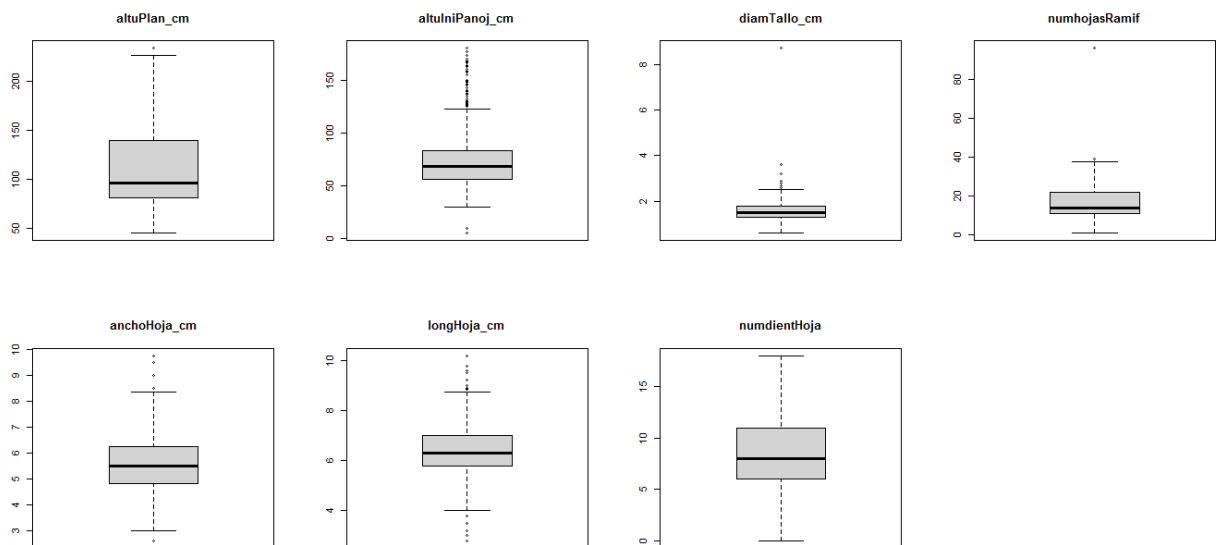
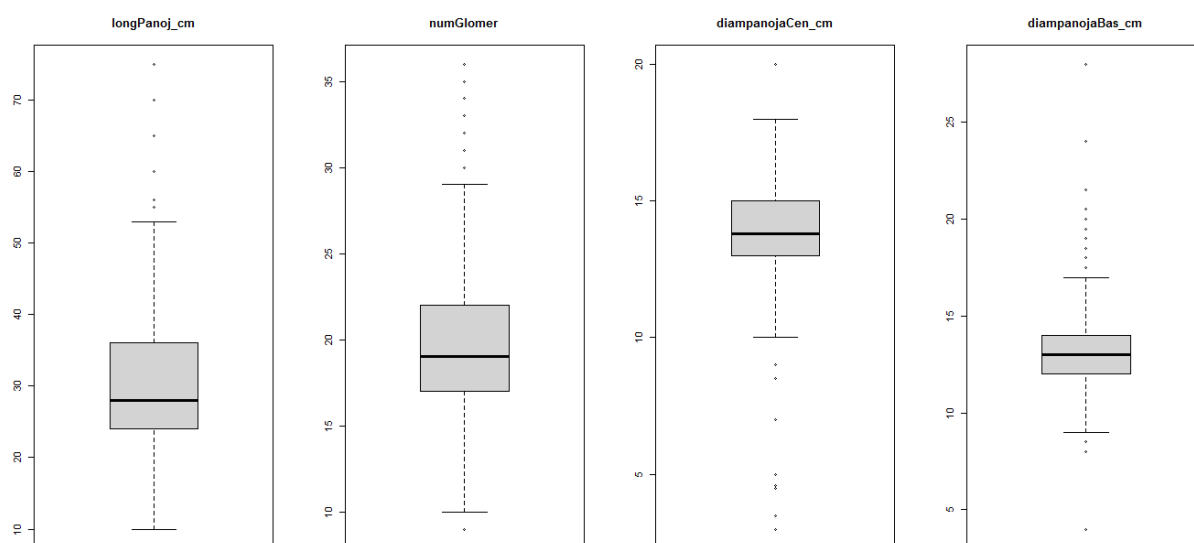
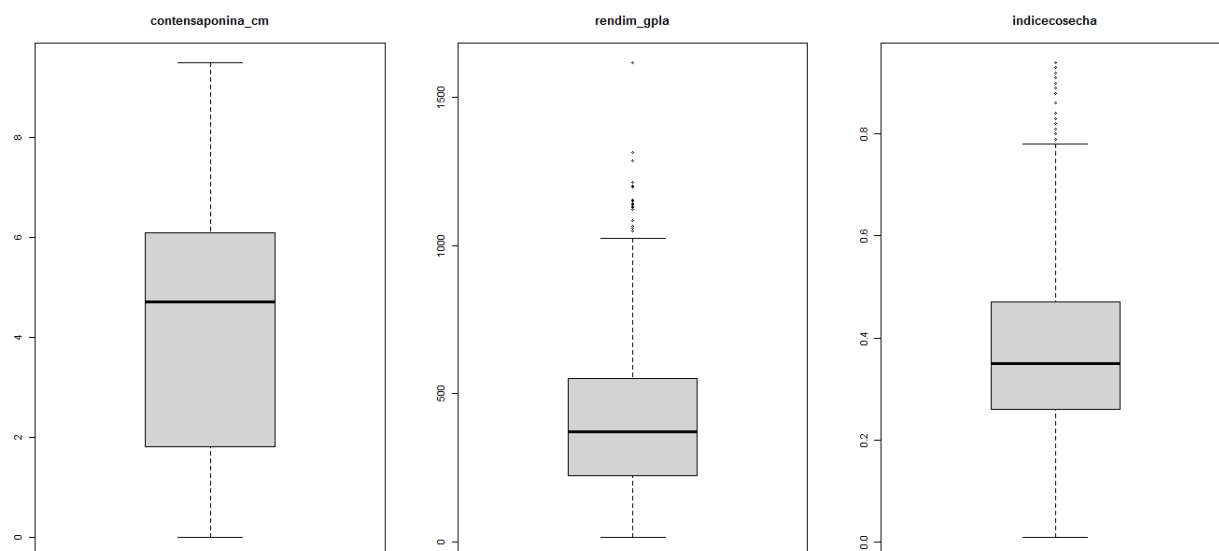


Figura 6

Distribución de las mediciones en los descriptores de caracterización de panoja.

**Figura 7**

Distribución de las mediciones en los descriptores de caracterización de grano.



En las figuras 4,5,6 y 7 se puede visualizar que en algunas variables analizadas se presenta valores outliers, que fueron revisados y contrastados con los registros haciéndose las correcciones correspondientes a fin de tener datos correctos y luego proceder con los análisis multivariados.

5.3. Estandarización de variables

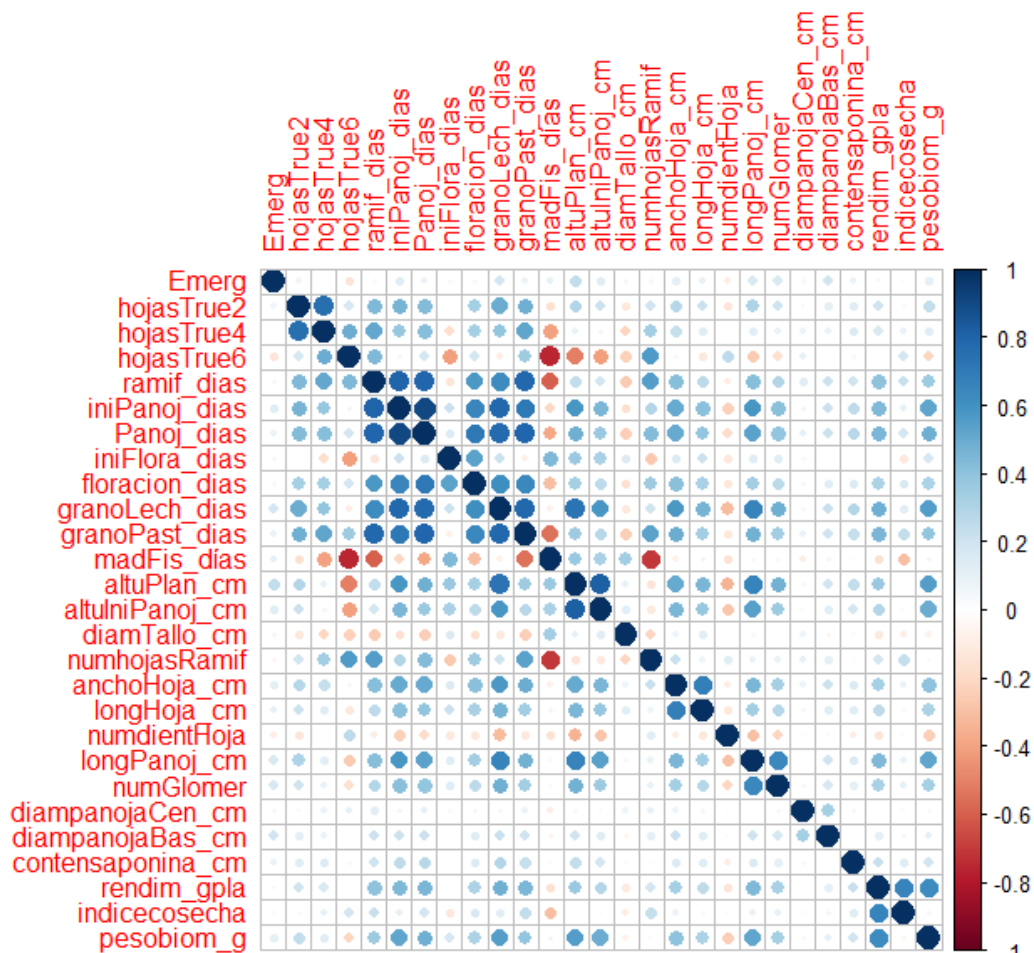
La estandarización de variables se realizó con la finalidad de homogenizar las diferentes unidades de medida de los descriptores o variables y así proceder con la aplicación del análisis estadístico. Es importante indicar que las variables tienen diferentes de medida tales como: cantidad de días, centímetros y gramos.

5.4. Correlación de las variables

Se ha procedido a verificar la correlación que existe entre cada una de las variables de caracterización, para ello se ha encontrado el siguiente gráfico que visualiza la correlación:

Figura 8

Correlación de las variables de caracterización.

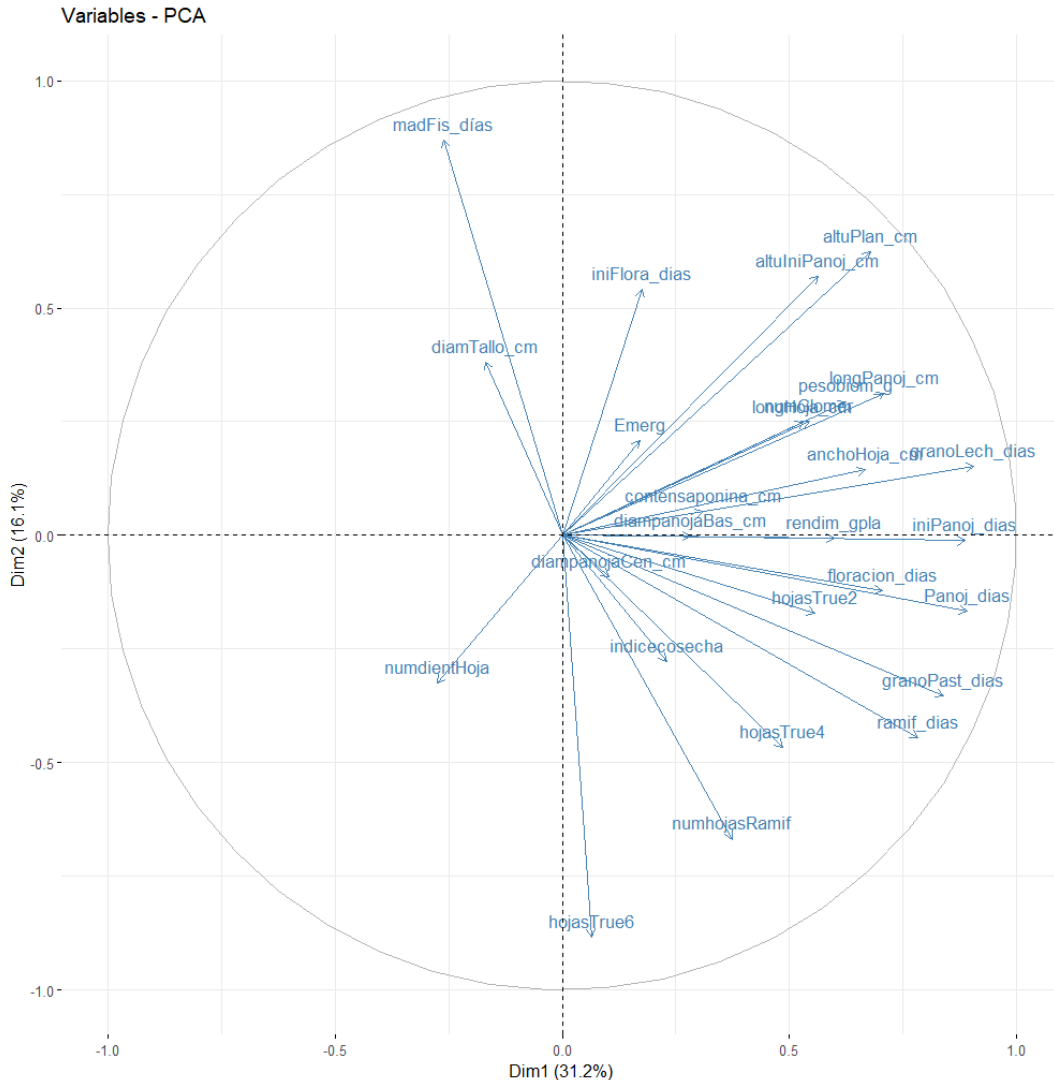


En la figura anterior se observan las correlaciones existentes entre las variables cuantitativas correspondientes al Banco de Germoplasma de quinua, en ello se observa una correlación positiva alta entre los descriptores (círculos azules de color azul más oscuro): “número de días en las que aparece 2 hojas verdaderas” (hojasTrue2) y “número de días en las que aparece 4 hojas verdaderas” (hojasTrue4), “días de inicio de panojamiento” (iniPanoj_días) y “número de días de ramificación” (ramif_días), “número de días de panojamiento” (Panoj_días) y “número de días de ramificación” (ramif_días), “número de días de formación de grano pastoso” (granoPast_días) y “número de días de formación de grano de leche” (granoLech_días), “número de días de formación de grano pastoso” (granoPast_días) y “número de días de panojamiento” (Panoj_días), “altura de planta” (altuPlan_cm) y “número de días de formación de grano de leche” (grano_lech_días), y “número de días de formación de grano de leche” (granoLech_días) y “días de inicio de panojamiento” (iniPanoj_días), “altura de inicio de panojamiento” (altuiniPanoj_cm) y “altura de planta” (altuPlan_cm), “longitud promedio de hoja” (longHoja_cm) y “encho de hoja” (anchoHoja_cm), “longitud de panoja” (longPanoj_cm) y “altura de planta” (altuPlan_cm), “numero de glomérulos” (numGlomer) y “longitud de panoja” (longPanoj_cm) por otro lado también existe correlación negativa alta entre los descriptores: “número de días de madurez fisiológica” (madFis_días) y “número de días de aparición de 6 hojas verdaderas” (hojasTrue6), “número de hojas ramificadas” (numhojasRamif) y “número de días de madurez fisiológica” (madFis_días).

5.5. Análisis de componentes principales

Figura 9

Biplot de componentes principales (Variables)

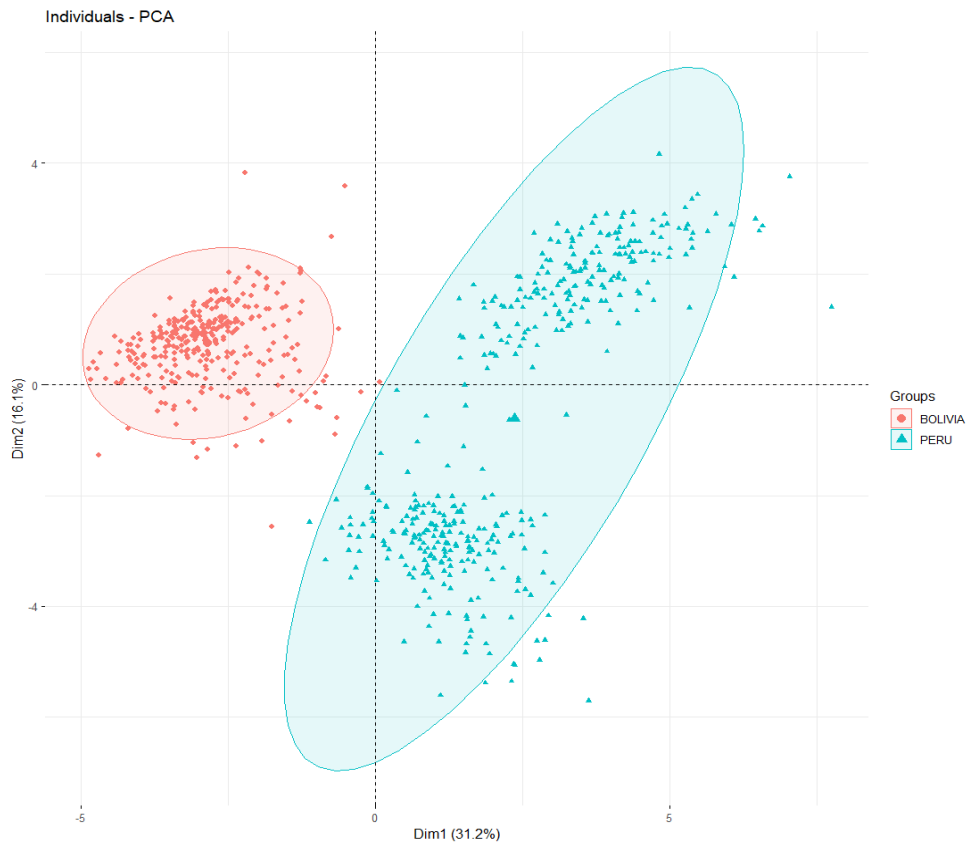


En la figura 9, se muestra un biplot con la representación de todas las variables cuantitativas, en las cuales se observa la presencia de una correlación directa o positiva entre algunas de ellas tales como: “Longitud de panoja” (longPanoj_cm) y “peso de biomasa” (pesobiom_g), “días de floración” (floración_dias) y “días de panojamiento” (Panoj_dias); así mismo, existen otras que tienen correlación inversa o negativa, por ejemplo: “número de hojas en ramificación” (numhojasRamif) y “diámetro de tallo” (diamTallo), “numero de dientes en la hoja” (numdienHoja) y “días de emergencia” (Emerg). Así mismo, se puede apreciar señalar respecto al tamaño de los vectores que representa a las variables, por ejemplo las variables “número de días de la madures

fisiológica” (madFis_días) es la que tiene mayor contribución con la Dim2, por otro lado la variable “número de días de inicio de panojamiento” (iniPanoj_dias) tiene mayor contribución con la Dim1.

Figura 10

Biplot de componentes principales (Colecciones de quinua)



En la figura anterior se puede observar la representación de cada una de las colecciones de quinua, en la que, por la disposición misma en cada una ellas se ha encontrado que pertenecen a 2 grupos (Bolivia y Perú), ello se puede explicar con la representación conjunta de las variables con las que están relacionadas cada uno de estos grupos, para ello, en la siguiente figura 11, se hace una representación del biplot con todas las colecciones de quinua y las variables cuantitativas analizadas.

Figura 11

Gráfico de componentes principales de las variables de caracterización.



Kaiser-Meyer-Olkin factor adequacy

Call: KMO (r = datos1cuanti)

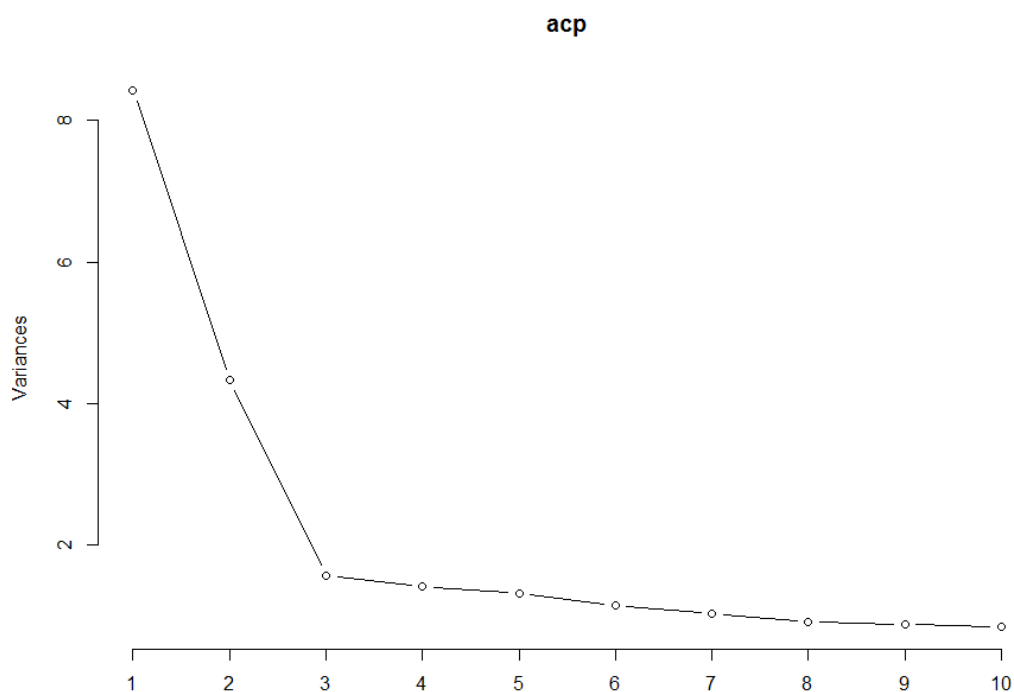
Overall MSA = 0.83

Según lo mostrado en la Figura 11, se muestra un biplot con cada una de las colecciones analizadas en color de cada una de las colecciones y las variables en forma de vectores, en ello se observa que los primeros cinco componentes se acumula un 63.2%, siendo las proporciones las siguientes: 31.2, 47.3, 53.1, 58.4 y 63.3%, donde los primeros tres componentes explican el 53.1% de la variación en los datos, ello considerando los componentes 1, 2 y 3; además de ello, en el resultado de KMO (medida de adecuación muestral) se tiene un valor alto, ello indica su perfecta pertinencia de análisis. Este

resultado se aproxima a los resultados encontrado por Dominguez (2014) que a nivel marcadores genéticos llegó a un 72% de acumulación de la variabilidad con cuatro componentes. Por otro lado según los resultados de Morillo et al. (2022) el PCA de las variables cuantitativas los dos primeros componentes explicaron hasta el 88% de la variación total, ello se explica por la mayor cantidad de descriptores cuantitativos utilizadas. Así mismo, respecto a los resultado de Cárdenas (2017), realizado con 129 colecciones de quinua del mismo banco, señala haber encontrado tres grupos diferentes; sin embargo, no hay una especificación de país de origen de las colecciones, pero si es importante señalar la existencia de una alta diversidad y ello coincide con los resultados encontrados en esta investigación.

Figura 12

Gráfico de sedimentación para el número de componentes principales.



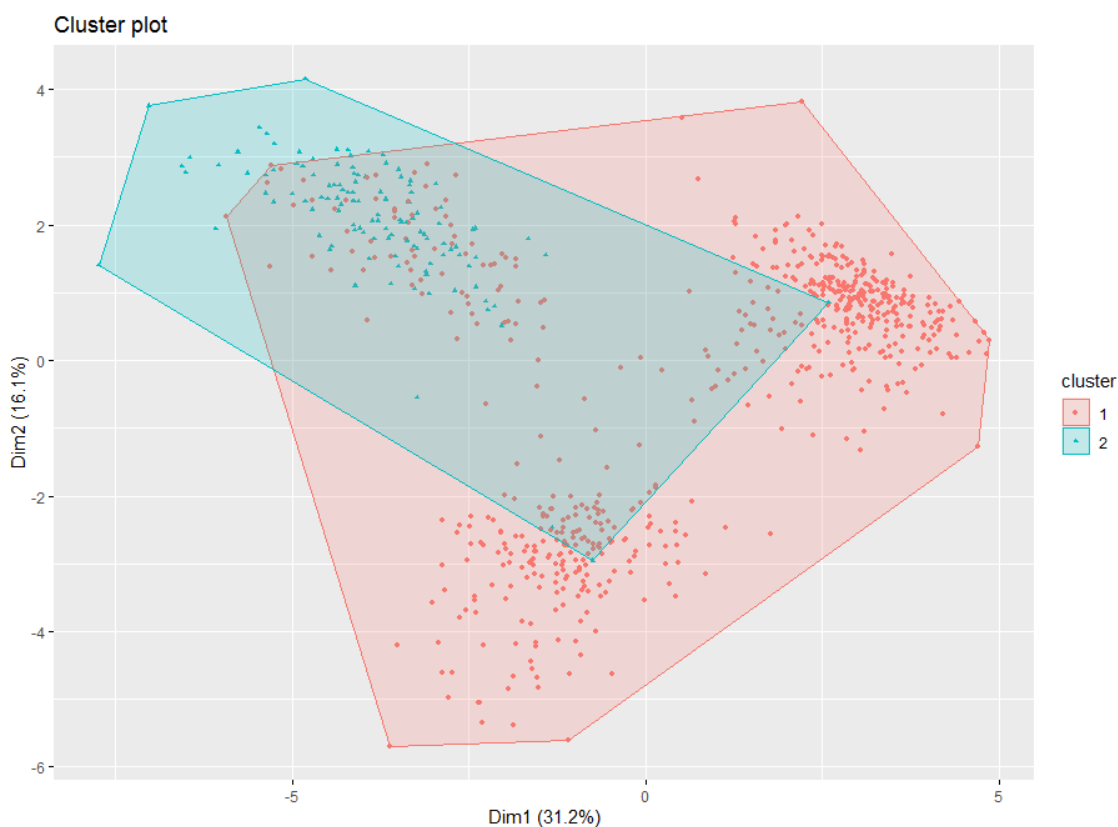
En la figura anterior, se observa el gráfico de sedimentación para el número de componentes principales, en el cual sugiere considerar hasta 3 componentes principales que concentran la mayor variabilidad de los datos.

5.6. Asociación de las colecciones de quinua

Para determinar las asociaciones de las colecciones de quinua, se ha utilizado la técnica de clúster, con el método Ward y distancia Euclídea. Siendo el resultado tal como se muestra en el gráfico siguiente:

Figura 13

Asociación de las colecciones de quinua mediante técnica de Clúster.

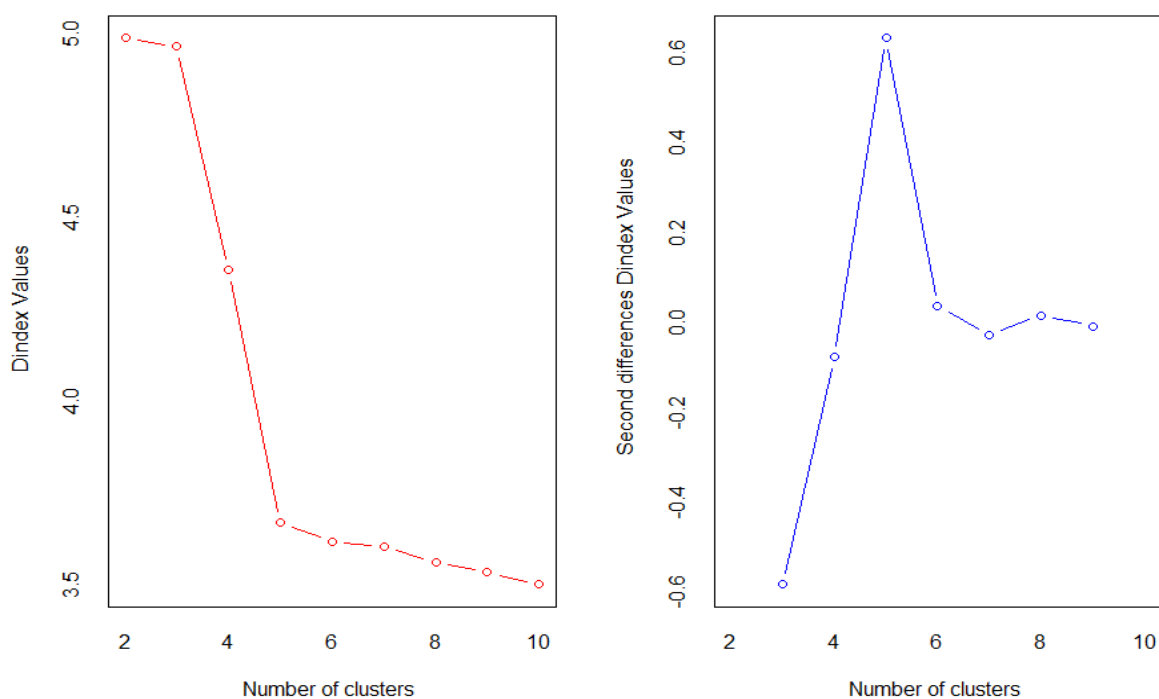


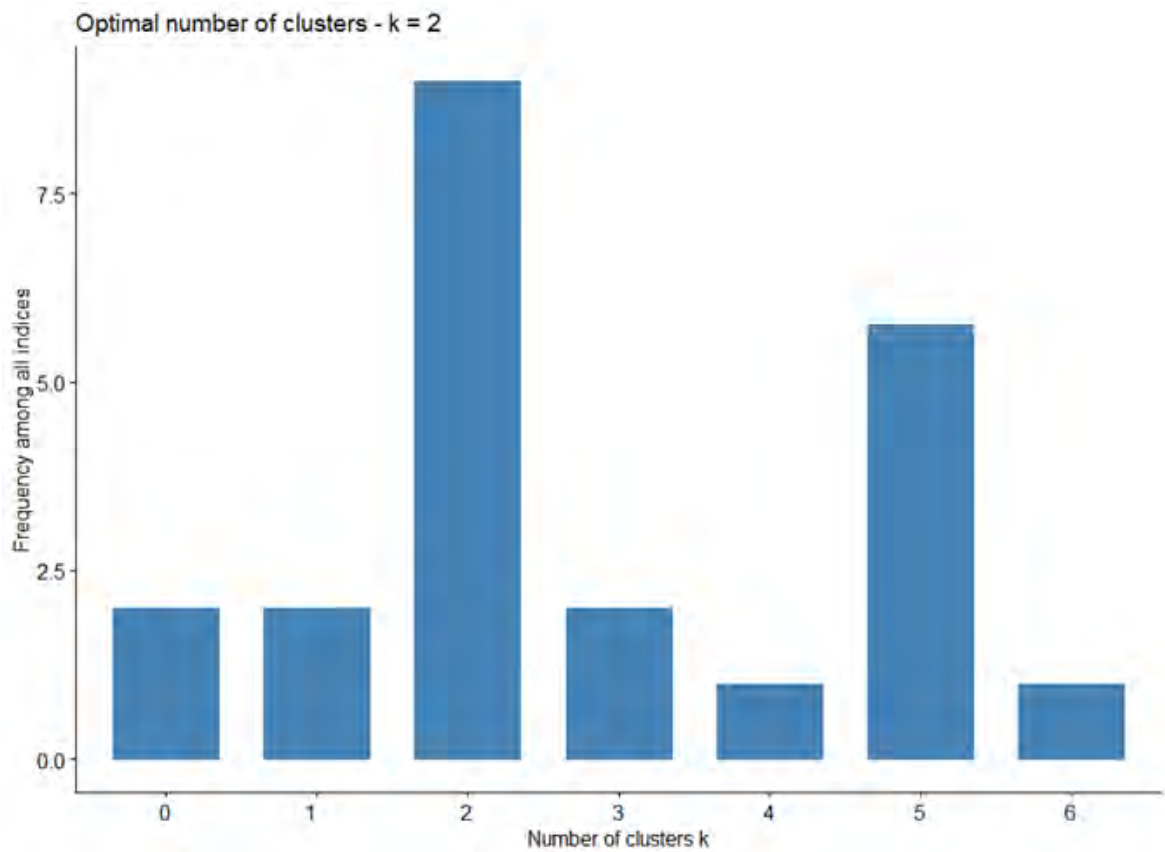
Para conocer esta cantidad adecuada de grupos, se sometió los datos a un procesamiento del algoritmo de NbClust, la misma que por decisión mayoritaria (CONSENSO), como mejor número de agrupaciones la cantidad de grupos sugeridos es 2

grupos, este resultado contrasta con lo indicado por Cárdenas (2017), quien señala la existencia de una conformación de grupos diferentes entre las colecciones estudiadas, ello básicamente se diferencia por la presencia de pigmentación de la planta, inflorescencia, grano y otros, debido a las amplias adaptaciones de las colecciones a diferentes condiciones agroecológicas que ocurre entre las colecciones de Perú y Bolivia, las que muestran resultados diferentes. Así mismo, la cantidad adecuada de grupos se muestra a continuación en los siguientes gráficos:

Figura 14

Matriz gráfica sobre la cantidad óptima de grupos en las colecciones de quinua.

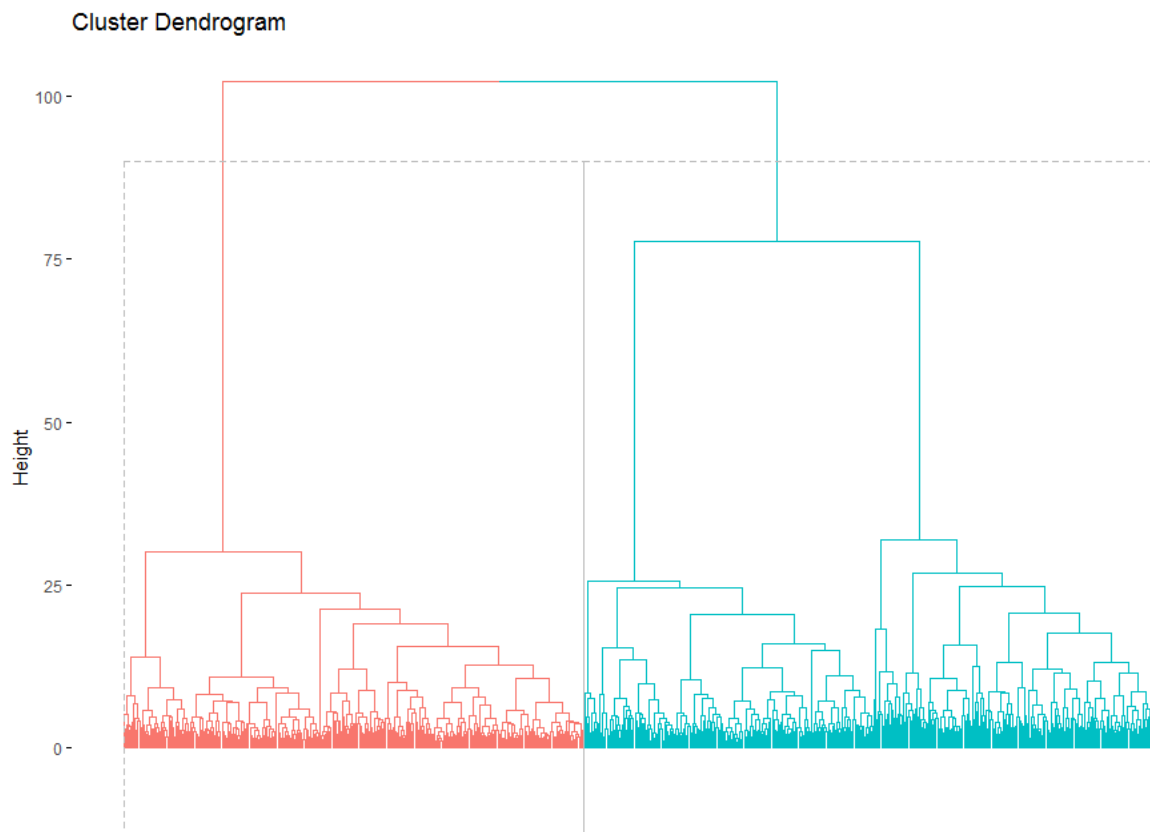




Con los gráficos de la Figura 14, se muestra los gráficos que confirman que se tiene 2 grupos claramente definidos, aplicando la técnica multivariada de Clúster, se procedió a representar en un dendograma, resultando dos grupos definidos que se separan claramente por el país de origen, este resultado tiene relación con lo manifestado por Dominguez (2014), muestra la representación de un dendograma que agrupa a sus colecciones de quinua diferenciadas por su progenie y ello se muestra a continuación:

Figura 15

Dendrograma de las colecciones de quinua.



En la Figura anterior, se observa dos grupos grandes diferenciados, sin embargo si se adopta un corte a menor distancia, se podrá tener varios subgrupos, estos resultados contrastados con la investigación realizada por Sucasaca et al. (2023) se puede confirmar la existencia de resultados similares ya que obtuvo 10 grupos en todo el conjuntos de colecciones analizadas.

5.7. Las variables que discriminan a los grupos de colecciones de quinua.

Para conocer las variables que discriminan, se ha analizado mediante la función LDA. Esta función permite realizar la clasificación mediante la aproximación de Fisher.

$$\text{modelo_LDA} = \text{LDA}(\text{pais} \sim \text{Emerg} + \text{hojasTrue2} + \dots + \text{pesobiom_g})$$

Obteniéndose los siguientes resultados:

BOLIVIA

0.4523161

PERÚ

0.5476839

Tabla 3

Resultados de los coeficientes

Variable	Coefficiente
Emerg	0.00862
hojasTrue2	0.07462
hojasTrue4	-0.04701
hojasTrue6	-0.01754
ramif_dias	0.10592
iniPanoj_dias	0.05578
Panoj_dias	0.08806
iniFlora_dias	-0.15978
floracion_dias	0.13648
granoLech_dias	0.06306
granoPast_dias	0.01883
madFis_días	-0.06729
altuPlan_cm	0.00938
altuIniPanoj_cm	0.00060
diamTallo_cm	-0.17408
numhojasRamif	0.00627
anchoHoja_cm	0.16535
longHoja_cm	-0.05015
numdientHoja	-0.04838
longPanoj_cm	0.03133
numGlomer	0.00454
diampanojaCen_cm	-0.04330
diampanojaBas_cm	0.02752
contensaponina_cm	-0.00940
rendim_gpla	0.00087
Indicecosecha	-0.30729
pesobiom_g	-0.00001

En la tabla mostrada, se tiene los resultados de los coeficientes LD1 y de ello las variables: “ancho de hoja” (anchoHoja_cm), “días de floración” (floracion_dias), “días de ramificación” (ramif_dias), “días de panojamiento” (Panoj_dias), “aparición de 2 hojas

verdaderas” (Panoj_dias), “días formación grano lechoso” (granoLech_dias) y “días de inicio de panojamiento” (iniPanoj_dias), son las que mejor discriminan en el modelo para la formación de los grupos, la mayor parte de ellos son fenológicas, estos resultados son similares a lo encontrado por Cárdenas (2017) que señala que las variables discriminantes están conformadas por las variables fenológicas.

5.8. Índice de diversidad.

Se realizó la estimación de los índices de diversidad de cada una de las variables cuantitativas en estudio, para ello mediante el cálculo de diversidad morfométrica, se ha obtenido los siguientes resultados:

Tabla 4

Índice de diversidad de las variables

VARIABLE	INDICE DE DIVERSIDAD
Emerg	0.924623433
hojasTrue2	0.820450891
hojasTrue4	0.830460235
hojasTrue6	0.804912997
ramif_dias	0.852594423
iniPanoj_dias	0.852112117
Panoj_dias	0.813289944
iniFlora_dias	0.865519916
floracion_dias	0.848640976
granoLech_dias	0.784777636
granoPast_dias	0.780365895
madFis_días	0.799464024
altuPlan_cm	0.781451286
altuIniPanoj_cm	0.838325171
diamTallo_cm	0.941074257
numhojasRamif	0.919536608
anchoHoja_cm	0.848723288
longHoja_cm	0.855077453
numdientHoja	0.822355029
longPanoj_cm	0.833790252
numGlomer	0.837070466
diampanojaCen_cm	0.890461224
diampanojaBas_cm	0.918882145
contensaponina_cm	0.741098739
rendim_gpla	0.839953654
indicecosecha	0.818344505
pesobiom_g	0.862745641

En la tabla mostrada, se tiene calculado el índice de diversidad de cada uno de los descriptores cuantitativos que se han considerado para la investigación, los mismos que muestran valores por encima del 0.74 y que en promedio todas las variables tienen un índice de 0.8204509, lo cual indica que las variables en consideración presentan una diversidad moderadamente alta. Esto significa que existe una variabilidad significativa en las mediciones morfométricas que esta tiene, estos resultados contrastan lo mencionado en los análisis antes expuestos e inclusive lo señalado por los estudios realizados por Cárdenas, (2017), que fue aplicado a 129 colecciones de quinua del mismo banco de germoplasma.

Conclusiones

1. Respecto al análisis de componentes principales (PCA), se ha encontrado que los tres primeros componentes explican el 53.1% de la variación total de los datos, con un KMO-MSA de 0.83 que indica la pertinencia del análisis.
2. Aplicando la técnica multivariada de clúster, se ha encontrado dos grupos definidos, las mismas que corresponden al grupo de colecciones de Perú y otro grupo a las colecciones de las quinuas de Bolivia, ello de acuerdo a los descriptores de fenología, planta, panoja y grano.
3. Las variables que mejor discriminan a los grupos que conforman las colecciones de quinua se tienen principalmente las variables: “ancho de hoja” (anchoHoja_cm), “días de floración” (floracion_dias), “días de ramificación” (ramif_dias), “días de panojamiento” (Panoj_dias), “aparición de 2 hojas verdaderas” (Panoj_dias), “días formación grano lechoso” (granoLech_dias) y “días de inicio de panojamiento” (iniPanoj_dias).
4. Finalmente, respecto al índice de diversidad se ha calculado un índice de diversidad promedio respecto a todas las variables estudiadas es de 0.82 el cual se traduce como una diversidad moderadamente alta.

Recomendaciones

1. Los datos de caracterización fenológica, de planta, panoja y grano constituyen una primera fuente de datos analizados acerca de las colecciones de quinua del Banco de Germoplasma de Puno; sin embargo, se recomienda hacer una caracterización molecular de cada una de estas colecciones y someterlos a otro análisis similar, la misma que permitiría contrastar los resultados de la presente investigación.
2. Existen otras variables que se encuentran en el documento internacional denominado “Descriptor internacional de granos y sus parientes silvestres” tales como: datos de pasaporte (georreferenciación), las mismas que pueden ser recogidos para hacer un estudio de geoespacial de toda la colección del Banco de Germoplasma de Puno, así mismo la información de suelo, temperatura y otras características serían importantes también para explicar el comportamiento de cada una de las colecciones de quinua, los mismos que pueden permitir mejorar las explicaciones acerca de las colecciones del Banco de Germoplasma de Puno.
3. Los resultados obtenidos de estos análisis deberían ser compartidos con los agricultores, técnicos e investigadores con la finalidad de desarrollar planes de mejoramiento genético en el cultivo y la toma de decisiones adecuadas sobre las colecciones del Banco de Germoplasma de quinua en a región Puno.

Bibliografía

- Anderberg, M. (1973). *Análisis de conglomerados para aplicaciones*. (N. Y. Academic Press (ed.)).
- Bioersivity International, FAO, PROINPA, INIAF, & FIDA. (2013). *Descriptorios para quinua (Chenopodium quinoa Willd.) y sus parientes silvestres*. FAO Publications Catalogue 2022. Retrieved. <http://www.fao.org/3/aq658s/aq658s.pdf>
- Blanco, F. (2015). *Introducción al análisis de componentes principales*.
- Bonifacio, A. (2004). *Genetic variation in cultivated and wild Chenopodium species for quinoa breeding*.
- Cardenas, M. (1994). Descripción preliminar de las variedades de Chenopodium quinoa de Bolivia. *Revista de Agricultura*, 2, 13–26.
- Cárdenas, R. G. (2017a). *Caracterización molecular de 129 accesiones de quinua (Chenopodium quinoa Willd.) de la región puno mediante marcadores microsatélites* (Issue January).
- Cárdenas, R. G. (2017b). *Caracterización molecular de 129 accesiones de quinua (Chenopodium quinoa Willd.) de la región puno mediante marcadores microsatélites Para optar el Título Profesional de Biólogo Genetista*. January, 147.
- Chase, M. W., Christenhusz, M. J. M., Fay, M. F., Byng, J. W., Judd, W. S., Soltis, D. E., Mabberley, D. J., Sennikov, A. N., Soltis, P. S., Stevens, P. F., Briggs, B., Brockington, S., Chautems, A., Clark, J. C., Conran, J., Haston, E., Möller, M., Moore, M., Olmstead, R., ... Weber, A. (2016). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society*, 181(1), 1–20. <https://doi.org/10.1111/boj.12385>
- Christensen, S., Pratt, D., Pratt, C., & Al, E. (2007). *Assessment of genetic diversity in the USDA and CIP-FAO international nursery collections of quinua (Chenopodium quinoa Willd.) using microsatellite markers*. Plant Genet Resources: Charact.
- Collet, K., Shepher, S., & Miller, L. (2004). Atlantic International Chapter. *Annual Meeting Info*, 1–51. http://www.fisheriessociety.org/aic/pdf/2004_07_Newsletter.pdf
- Cruz, M. (2016). *Caracterización agronómica de la colección de germoplasma de quinua (Chenopodium quinoa Willd.) del banco nacional de granos altoandinos, del altiplano centro de la estación experimental de patacamaya*. 84.
- De la Fuente, S. (2011). *Análisis de Conglomerados* (U. (Universidad A. de Madrid) (ed.); 2011th ed.). http://www.estadistica.net/Master-Econometria/Analisis_Cluster.pdf
- Dominguez M., R. (2014). *Caracterización agromorfológica de progenies autofecundadas S2, de cruza dobles genéticamente distantes y cercanas en quinua (Chenopodium quinoa willd.) bajo condiciones de invernadero*. Universidad Nacional San Agustín.
- Fuentes, F., Maughan, J., & Jellen, E. (2009). Diversidad genética y recursos genéticos para el mejoramiento de la quinua (Chenopodium quinoa Willd.). *Revista Geográfica de Valparaiso*, 42/2009, 20–23.
- Gandarillas, H. (1979). *Quinua y Kañiwa, cultivos andinos*. CIID, Oficina Regional para América Latina.
- Gomez, L., & Aguilar, E. (2016). Guía del cultivo de la quinua. In *Statewide Agricultural Land Use Baseline 2015* (Vol. 1). <https://doi.org/10.1017/CBO9781107415324.004>
- Hartigan, J. A. (1975). *Clustering Algorithms* (I. John Wiley & Sons (ed.); 99th ed.).
- Heuveldop, J., Pardo, J., Quirós, S., & Espinoza, L. (1986). *Agroclimatología Tropical* (U. E. a Distancia (ed.)).
- Jacobsen, S., & Mujica, A. (2001). *El potencial de la quinua en la alimentación global*.

- Jellen, E., Kolano, B., Sederberg, M., Bonifacio, A., & Maughan, P. (2011). In Wild Crop Relatives: Genomic and Breeding Resources. *Chenopodium*, 35–61.
- Jiménez, Q., Estrada, A., Rodríguez, A., & Arroyo, P. (1999). *Manual Dendrológico de Costa Rica*.
- Laura, R., Alave Chata, D. E., Quispe Ramos, E., Colque Ramos, J., & Cora Machaca, C. G. (2023). La gestión educacional y producción de quinua orgánica (*Chenopodium Quinoa Wild*) del banco de germoplasma de la universidad nacional del Altiplano según Bioversity Internacional. *Ciencia Latina Revista Científica Multidisciplinar*, 7(2), 8260–8278. https://doi.org/10.37811/cl_rcm.v7i2.5954
- Lescano, J. L. (1994). *Genética y mejoramiento de cultivos altoandinos: quinua, ka ihua, tarwi, kiwicha, papa amarga, olluco, mashua y oca*.
- Marín, J., & Rodríguez-Bernal, M. T. (2012). Multiple hypothesis testing and clustering with mixtures of non-central t-distributions applied in microarray data analysis. *Computational Statistics and Data Analysis*.
- Moreno, C. (2001). *Métodos para medir la biodiversidad* (M&T–Manuales y Tesis (ed.); SEA, Vol. 1).
- Morillo, A. C., Manjarres, E. E., & Morillo, Y. M. (2022). Phenotypic diversity of agromorphological characteristics of quinoa (*Chenopodium quinoa willd.*) germplasm in colombia. *Scientia Agricola*, 79(4). <https://doi.org/10.1590/1678-992X-2021-0017>
- Mujica, A., Cahahua, A., & Saravia, R. (2004). Agronomía de la quinua. *Quinoa: Ancestral Cultivo Andino, Alimento Del Presente y Futuro.*, 26–59.
- Mujica, A., & Jacobsen, E. (2001). *Biodiversidad un desafío en la región centro oeste de Sudamérica en agricultura andina*.
- Mujica, A., & Jacobsen, S. (2006). *La quinua (Chenopodium quinoa Willd.) y sus parientes silvestres*. Universidad Mayor de San Andrés.
- PromPerú, Indecopi, & Quinoa, M. de T. de. (2009). *Norma técnica peruana 205.062:2009. Quinoa (Chenopodium quinoa Willd.)*. 15.
- Rimieri, P. (2017). La diversidad genética y la variabilidad genética. *Journal of Basic and Applied Genetics*, 1(2), 17–130. https://sag.org.ar/jbag/wp-content/uploads/2019/11/A1_7-13-2.pdf
- Rojas, W., Pinto, M., Alanoca, C., Gómez, L., Alercia, A., Leon-lobos, P., Diulgheroff, S., Padulosi, S., & Bazile, D. (2014). *Estado de la conservación ex situ de los recursos genéticos de quinua*.
- Rojas, W., Pinto, M., & Soto, J. (2010). *Distribución geográfica y variabilidad genética de los granos andinos*. Bioversity International, Roma, Italia.
- Sagaró del Campo, N. M., & Zamora Matamoros, L. (2020). Técnicas estadísticas multivariadas para el estudio de la causalidad en Medicina. *Revista de Ciencias Médicas de Pinar Del Río*, 24(2), 287–300.
- Tapia, M., Canahua, A., & Ignacio, S. (2014). *Razas de quinua del Perú*.
- UNALM, MINAGRI, F. (2017). *Resúmenes de exposiciones del VI Congreso Mundial de la Quinoa y III Simposio Internacional de Granos Andinos, Perú 2017*.
- Vertel, M., Botero, L., & Cépeda, J. (2016). *Análisis multivariado de datos* (E. Mejoras (ed.); Universida).
- Vertti, J. B. (2019). *Javier Bech Vertti*.
- Visauta, B. (1998). *Análisis estadístico con SPSS para Windows* (Estadística & multivariante. M. M. Graw-Hill (eds.); Volumen II).
- Zurita-Silva, A., Fuentes, F., Zamora, P., Jacobsen, S.-E., & Schwember, A. R. (2014). Breeding quinoa (*Chenopodium quinoa Willd.*): potential and perspectives. *Molecular Breeding*, 34(1), 13–30. <https://doi.org/10.1007/s11032-014-0023-5>

Anexos

a. Matriz de consistencia

TÍTULO: Análisis de la diversidad genética de quinua (*Chenopodium quinoa* Willd.) del banco de germoplasma de Puno, mediante aplicación de técnicas multivariadas, 2021.

Problema	Objetivo	Hipótesis	Variables
<p>General</p> <p>¿Cómo es la diversidad genética de la colección del banco de germoplasma de quinua en Puno, mediante aplicación de técnicas multivariadas, 2021?</p>	<p>General</p> <p>Determinar la diversidad genética de la colección del banco de germoplasma de quinua en Puno, mediante aplicación de técnicas multivariadas, 2021.</p>	<p>General</p> <p>La diversidad genética de quinua (<i>Chenopodium quinoa</i> Willd.) del banco de germoplasma de Puno, mediante aplicación de técnicas multivariadas, 2021, es alta.</p>	<p>La diversidad genética</p>
<p>Específicos</p> <p>a) ¿Cómo es la concentración de la mayor variabilidad, en la colección de quinua del banco de germoplasma de Puno?</p> <p>b) ¿Cómo es el nivel de asociación de la colección de quinua en el banco de germoplasma de Puno, mediante aplicación de métodos multivariados?</p> <p>c) ¿Cuál es el grado de discriminación o importancia de las variables de la colección de quinua del banco de germoplasma de Puno, mediante aplicación de métodos multivariados?</p> <p>d) ¿Cuánto es la diversidad de la colección de quinua del banco de germoplasma de Puno?</p>	<p>Específicos</p> <p>a) Determinar la concentración de la mayor variabilidad, en la colección de quinua del banco de germoplasma de Puno</p> <p>b) Establecer la asociación de la colección de quinua correspondiente al banco de germoplasma de Puno, mediante aplicación de métodos multivariados.</p> <p>c) Determinar las variables que mejor discriminan a los grupos de colecciones de quinua de Puno, mediante aplicación de métodos multivariados.</p> <p>d) Cuantificar la diversidad de la colección de quinua del banco de germoplasma de Puno</p>	<p>Específicos</p> <p>a) Las variables de la colección de quinua del banco de germoplasma de Puno, se concentran con mayor proporción en los primeros componentes, mediante aplicación de técnicas multivariadas.</p> <p>b) La colección de quinua del banco de germoplasma de Puno, se diferencia claramente al formar grupos con la aplicación de la técnica de cluster.</p> <p>c) En los grupos de la colección de quinua en el banco de germoplasma de Puno, las variables referidas a producción de semilla y longitud de planta y son las que mejor discriminan mediante aplicación de métodos multivariados.</p> <p>d) La cuantificación de la diversidad en la colección de quinua del banco de germoplasma de Puno es alta.</p>	

b. Instrumentos de recolección de datos a nivel de campo

CARACTERIZACIÓN DE FENOLOGIA

Código de Accesoión :.....

Emergencia (N° de días) :		Ramificación (N° de días) :	
Inicio de panojamiento (N° de días) :		Inicio de floración (N° de días) :	
Inicio de Formación de grano lechoso (N° días).....		Inicio de Formación de grano Pastoso (N° días).....	
Inicio de Madurez fisiológica (N° días)			

CARACTERIZACIÓN DE PLANTA

Código de Accesoión :.....

Altura de la planta (cm)		Altura de inicio de panoja	
Diámetro del tallo (cm)		Color del tallo	
Cantidad total de hojas		Hábito de crecimiento	
Ancho de hoja (cm)		Longitud de hoja	
Numero de dientes de la hoja		Color de las hojas	

CARACTERIZACIÓN DE PANOJA

Código de Accesoión :.....

Longitud de panoja (cm)		N° total de glomérulos	
Color de panoja		Diámetro central	
Diámetro de la base		Forma de la panoja	

CARACTERIZACIÓN DE GRANO

Código de Accesoión :.....

Tamaño de grano (mm)		Forma del grano	
Color de perigonio del fruto		Color del pericarpio	
Color de epispermo		Contenido de saponina(cm)	
Rendimiento por planta		Peso de biomasa	
Indica de cosecha por planta			

c. Medidas resumen de las variables cuantitativas general

Descripción	Variable	Media	D.E.	CV	Mín	Máx
- Emergencia(N° días)	Emerg	10.16	2.23	21.94	9	70
- Aparición de 2 hojas verdaderas (N° días)	2hojasTrue	17	1.79	10.74	15	25
- Aparición de 4 hojas verdaderas (N° días)	4hojasTrue	23	2.39	10.26	20	34
- Aparición de 6 hojas verdaderas (N° días)	6hojasTrue	30	3.54	11.83	23	41
- días de Ramificación(N° días)	ramif_dias	40	4	10.08	26	60
- Inicio de panojamiento(N° días)	iniPanoj_dias	49	6.91	14.13	40	84
- Dias de panojamiento(N° días)	Panoj_dias	59	9.58	16.19	37	98
- Inicio de floración (N° días)	iniFlora_dias	82	6.1	7.44	46	110
- días de floración (N° días)	floracion_dias	95	7.66	8.06	78	210
- Grano en estado lechoso (N° días)	granoLech_dias	124	146.05	117.67	84	220
- Grano en estado pastoso (N° días)	granoPast_dias	131	17.88	13.61	91	167
- Madurez fisiológica(N° días)	madFis_dias	198	17.93	9.07	120	210
- Altura planta (cm)	altuPlan_cm	110.89	41.28	37.22	45	233.5
- Altura inicio de panoja (cm)	altuIniPanoj_cm	76.86	38.41	49.98	4	698
- Diámetro de tallo (cm)	diamTallo_cm	1.57	0.57	36.19	0.6	10
- Número total hojas ramificación	numhojasRamif	17.56	8.33	47.46	1	96
- Ancho de la hoja (cm)	anchoHoja_cm	5.6	1.17	20.97	2.6	18
- Longitud de hoja (cm)	longHoja_cm	7.47	26.64	356.76	2.8	735
- Número de dientes en hoja	numdientHoja	8.05	3.19	39.62	0	18
- Longitud de panoja (cm)	longPanoj_cm	31.25	10.76	34.44	10	75
- N° total de glómerulos	numGlomer	19.58	4.55	23.25	9	52
- Diámetro panoja central (cm)	diampanojaCen_cm	13.89	3.02	21.75	3	78
- Diámetro de la base (cm)	diampanojaBas_cm	13.82	9.38	67.87	4	16
- Contenido de saponina (cm)	contensaponina_cm	4.15	2.46	59.38	-1.7	9.5
- Rendim. por planta (g/planta)	rendim_gpla	425.3	257.28	60.49	17	1616.6
- Índice de cosecha por planta	indicecosecha	0.47	1.86	395.85	0.01	45
- Peso de biomasa	pesobiom_g	1144.29	638.39	55.79	150	9000

d. Medidas resumen de las variables cuantitativas (Perú)

Descripción	pais	Variable	Media	D.E.	CV	Mín	Máx
- Emergencia(N° días)	PERU	Emerg	10.12	0.49	4.89	9	14
- Aparición de 2 hojas verdaderas (N° días)	PERU	2hojasTrue	17.48	1.48	8.48	15	25
- Aparición de 4 hojas verdaderas (N° días)	PERU	4hojasTrue	24.4	2.1	8.62	20	34
- Aparición de 6 hojas verdaderas (N° días)	PERU	6hojasTrue	30.94	4.22	13.65	23	41
- días de Ramificación(N° días)	PERU	ramif_dias	42.81	2.63	6.15	38	60
- Inicio de panojamiento(N° días)	PERU	iniPanoj_dias	53.77	5.88	10.94	48	84
- Dias de panojamiento(N° días)	PERU	Panoj_dias	66.69	6.34	9.5	37	98
- Inicio de floración (N° días)	PERU	iniFlora_dias	81.56	6.58	8.07	46	110
- días de floración (N° días)	PERU	floracion_dias	98.84	7.12	7.2	84	210
- Grano en estado lechoso (N° días)	PERU	granoLech_dias	129.64	11.56	8.92	111	154
- Grano en estado pastoso (N° días)	PERU	granoPast_dias	147.13	3.69	2.51	130	167
- Madurez fisiológica(N° días)	PERU	madFis_dias	188.63	19.39	10.28	158	210
- Altura planta (cm)	PERU	altuPlan_cm	126.39	48.3	38.22	50	233.5
- Altura inicio de panoja (cm)	PERU	altuIniPanoj_cm	86.63	48.19	55.63	4	698
- Diámetro de tallo (cm)	PERU	diamTallo_cm	1.44	0.38	26.14	0.6	3
- Número total hojas ramificación	PERU	numhojasRamif	25.26	5.55	21.96	9	39
- Ancho de la hoja (cm)	PERU	anchoHoja_cm	6.12	1.06	17.35	3.5	18
- Longitud de hoja (cm)	PERU	longHoja_cm	6.9	2.46	35.6	3	51.5
- Número de dientes en hoja	PERU	numdientHoja	7.5	3.44	45.87	0	18
- Longitud de panoja (cm)	PERU	longPanoj_cm	36.85	10.89	29.55	13	70
- N° total de glomérulos	PERU	numGlomer	21.18	4.74	22.36	12	52
- Diámetro panoja central (cm)	PERU	diampanojaCen_cm	14.09	3.97	28.15	3	78
- Diámetro de la base (cm)	PERU	diampanojaBas_cm	13.95	8.12	58.19	4	169.5
- Contenido de saponina (cm)	PERU	contensaponina_cm	4.7	2.43	51.7	0	9.5
- Rendim. por planta (g/planta)	PERU	rendim_gpla	544.44	275.42	50.59	22	1616.6
- Índice de cosecha por planta	PERU	indicecosecha	0.42	0.19	45.69	0.01	0.94
- Peso de biomasa	PERU	pesobiom_g	1381.98	752.47	54.45	150	9000

e. Medidas resumen de las variables cuantitativas (Bolivia)

Descripción	pais	Variable	Media	D.E.	CV	Mín	Máx
- Emergencia(N° días)	BOLIVIA	Emerg	10.21	3.29	32.16	10	70
- Aparición de 2 hojas verdaderas (N° días)	BOLIVIA	2hojasTrue	15.72	1.66	10.54	15	23
- Aparición de 4 hojas verdaderas (N° días)	BOLIVIA	4hojasTrue	21.99	2.03	9.25	20	28
- Aparición de 6 hojas verdaderas (N° días)	BOLIVIA	6hojasTrue	28.76	1.88	6.55	23	34
- días de Ramificación(N° días)	BOLIVIA	ramif_dias	35.88	0.85	2.36	26	36
- Inicio de panojamiento(N° días)	BOLIVIA	iniPanoj_dias	42.99	0.16	0.38	40	43
- Dias de panojamiento(N° días)	BOLIVIA	Panoj_dias	50.09	1.86	3.71	48	84
- Inicio de floración (N° días)	BOLIVIA	iniFlora_dias	82.56	5.42	6.57	72	98
- días de floración (N° días)	BOLIVIA	floracion_dias	90.31	5.33	5.9	78	113
- Grano en estado lechoso (N° días)	BOLIVIA	granoLech_dias	117.37	217.61	185.4	84	220
- Grano en estado pastoso (N° días)	BOLIVIA	granoPast_dias	112.17	4.51	4.02	91	126
- Madurez fisiológica(N° días)	BOLIVIA	madFis_dias	208.87	5.58	2.67	120	210
- Altura planta (cm)	BOLIVIA	altuPlan_cm	91.95	16.93	18.41	45	180
- Altura inicio de panoja (cm)	BOLIVIA	altuIniPanoj_cm	64.8	13.51	20.85	5	100
- Diámetro de tallo (cm)	BOLIVIA	diamTallo_cm	1.74	0.71	40.76	1	10
- Número total hojas ramificación	BOLIVIA	numhojasRamif	12.47	5.48	43.95	1	96
- Ancho de la hoja (cm)	BOLIVIA	anchoHoja_cm	4.96	0.97	19.63	2.6	8.5
- Longitud de hoja (cm)	BOLIVIA	longHoja_cm	8.16	39.73	486.7	2.8	735
- Número de dientes en hoja	BOLIVIA	numdientHoja	8.7	2.71	31.13	4	17
- Longitud de panoja (cm)	BOLIVIA	longPanoj_cm	24.49	5.45	22.26	10	75
- N° total de glomérulos	BOLIVIA	numGlomer	17.62	3.44	19.52	9	30
- Diámetro panoja central (cm)	BOLIVIA	diampanojaCen_cm	13.65	1.02	7.49	12	17.6
- Diámetro de la base (cm)	BOLIVIA	diampanojaBas_cm	13.67	10.76	78.77	10.65	130
- Contenido de saponina (cm)	BOLIVIA	contensaponina_cm	3.47	2.34	67.3	-1.7	8.5
- Rendim. por planta (g/planta)	BOLIVIA	rendim_gpla	277.69	122.05	43.95	17	685.3
- Índice de cosecha por planta	BOLIVIA	indicecosecha	0.53	2.77	520.93	0.09	45
- Peso de biomasa	BOLIVIA	pesobiom_g	849.21	229.38	27.01	188.9	1517.8

f. Código R utilizado para el procesamiento de datos

```

library(readxl)
datos=read_xlsx("datos_carac_okSENDTESIS1_todo1.xlsx",sheet = 1)
#----- Verificar Cant. NA por columna
colSums(is.na(datos))
barplot(colSums(is.na(datos)), xlab = "Variabes", ylab="Cantidad datos faltantes",col = "blue")
sum(is.na(datos))
# Se tiene 38 columnas (5.04%)
38/753*100
# Omitir las FILAS con observaciones NA
datos1 = na.omit(datos)
colSums(is.na(datos1))
sum(is.na(datos1))

#----- comparando dimensiones de la matriz
dim(datos)
dim(datos1)
#----- corrigiendo tipos de datos
colSums(is.na(datos1))
sum(is.na(datos1))
str(datos1)
#----- Graficando boxplot para cada variable por grupos
par(mfrow=c(3,4))
for (i in 3:14) {
  boxplot(datos1[,i],main=names(datos1)[i])
}

par(mfrow=c(2,4))
for (i in c(15:17,21:23,25)) {
  boxplot(datos1[,i],main=names(datos1)[i])
}

par(mfrow=c(1,4))
for (i in c(28,29,32,33)) {
  boxplot(datos1[,i],main=names(datos1)[i])
}

par(mfrow=c(1,3))
for (i in c(40:42)) {
  boxplot(datos1[,i],main=names(datos1)[i])
}

par(mfrow=c(1,1))

#----- Correlacion de las variables

datos1cuanti=datos1[,c(3:17,21:23,25,28,29,32,33,40:43)]
library(corrplot)
library(grDevices)

```

```

library(Hmisc)
dato_R=rcorr(as.matrix(datos1cuanti))
corrplot(dato_R$r)

#----- Analisis multivariado
library(factoextra)
library(FactoMineR)
library(ggplot2)

datos1cuanti=datos1[,c(3:17,21:23,25,28,29,32,33,40:43)]
View(datos1cuanti)

resul_pca = PCA(datos1cuanti, scale.unit=TRUE, ncp=6, graph=TRUE)
plot(resul_pca, main="PCA colecciones de quinua")
resul_pca$var

acp=prcomp(datos1cuanti,center = T,scale=T)

acp$sdev
summary(acp)
plot(acp,type="l")

library(psych)
KMO(datos1cuanti)
scree(datos1cuanti,main = "Grafico_de_Sedimentacion")

fa.parallel(datos1cuanti,fa="pc")

fviz_pca_var(resul_pca, col.var = "steelblue")
datos1$pais=as.factor(datos1$pais)

#----- graficando las accesiones en 2 dimensiones y pintados por pais.
fviz_pca_ind(resul_pca, label="none", habillage=datos1$pais)

fviz_pca_ind(resul_pca, label="none", habillage=datos1$pais, addEllipses=TRUE, ellipse.level=0.95)

#agregamos biplot
fviz_pca_biplot(resul_pca, label = "var", habillage=datos1$pais, addEllipses=TRUE,
ellipse.level=0.95, ggtheme = theme_minimal())

summary(resul_pca)
resul_pca$ind

#analisis cluster
library("cluster")
clarax = clara(datos1cuanti, 2)
# Cluster plot
fviz_cluster(clarax, stand = T, geom = "point", pointsize = 1)

```

```

#----- numero optimo de grupos
my_data = scale(datos1cuanti)
fviz_nbclust(my_data, kmeans, method = "gap_stat")

library("NbClust")
set.seed(123)
res_nbclust = NbClust(my_data, distance = "euclidean",
                     min.nc = 2, max.nc = 10,
                     method = "complete", index = "all")

factoextra::fviz_nbclust(res_nbclust)

#----- Enhanced hierarchical clustering, cut in 2 groups
res_hc = eclust(my_data, "hclust", k = 2, graph = FALSE)
# Visualizando
fviz_dend(res_hc, rect = TRUE, show_labels = FALSE)

# Graficando
fviz_screplot(resul_pca, addlabels = TRUE, ylim = c(0, 20))

#----- funcion discriminante
datos1cuanti$pais=datos1$pais

library(MASS)
modelo_lda = lda(pais ~ ., data = datos1cuanti)
modelo_lda

plot(modelo_lda, col = c("red", "blue", "green"))

modelo_lda$prior
modelo_lda$counts
modelo_lda$means
modelo_lda$scaling #para ver mejor la importancia de las variables

#-----indice de diversidad -----:

library(ggplot2)
# Calcular las medidas de dispersión (desviación estándar) para cada variable
medidas_dispercion = apply(datos1cuanti, 2, sd)
# Normalizar las medidas de dispersión dividiendo por el rango de cada variable
rangos = apply(datos1cuanti, 2, range)
medidas_dispercion_normalizadas = medidas_dispercion / diff(rangos)
# Calcular el índice de diversidad morfométrica para cada variable
indice_diversidad = 1 - medidas_dispercion_normalizadas
indice_diversidad =as.data.frame(indice_diversidad)

# Crear un gráfico de barras para visualizar el índice de diversidad morfométrica
variable = names(indice_diversidad)
dft = data.frame(t(indice_diversidad[-1]))
names(dft)[1]="indice"

```

```
barplot(dft$indice)
library(openxlsx)
write.xlsx(indice_diversidad,"indice_diver.xlsx")
mean(dft$indice)
min(dft$indice)

df_resultados = data.frame(variable, indice_diversidad*100)
grafico = ggplot(dft$indice, aes(x = variable, y = dft$indice)) +
  geom_bar(stat = "identity", fill = "skyblue") +
  labs(title = "Índice de Diversidad Morfométrica", x = "Variable", y = "Índice de Diversidad") +
  theme_minimal()

# Imprimiendo el gráfico
print(grafico)

# Interpretar los resultados
print(indice_diversidad)

# ----- Fin de código -----
```


g. Documento de verificación de datos (fuente de datos)

CATÁLOGO DEL BANCO DE GERMOPLASMA DE QUINUA



h. Equipo de profesionales en el campo de germoplasma de quinua, Camacani Puno.



i. Presentación de resultados en el 8vo congreso mundial de la quinua 2023.



APLICACIÓN DE MÉTODOS MULTIVARIADOS EN EL ANÁLISIS DE DATOS DE CARACTERIZACIÓN Y EVALUACIÓN DE ACCESIONES DE QUINUA (*Chenopodium quinoa Willd.*), UNAP. PERÚ.

José Luis Cahuana Jorge

Potosí, marzo 2023



VIII Congreso Mundial de la Quinoa - Bolivia 2023	
Libro de Resúmenes	
Recursos genéticos para el descubrimiento de genes y el mejoramiento agronómico de la quinua <i>Genetic resources for gene discovery and agronomic improvement of quinoa</i> David Jarvis Estados Unidos	111
Método 4 Celdas: una herramienta para identificar, evaluar y recuperar la biodiversidad de cultivares de quinua en proceso de erosión <i>Method 4 Cells: a tool for identifying, assessing and restoring biodiversity of eroding quinoa cultivars</i> José Luis Soto Mendizabal Perú	112
Rendimiento de grano de 14 líneas de quinua (<i>Chenopodium quinoa Willd.</i>) en el Centro Agronómico Kayra Cusco Perú <i>Grain yield of 14 quinoa (<i>Chenopodium quinoa Willd.</i>) lines at Centro Agronómico Kayra Cusco Perú</i> Elisabet Céspedes Florez Perú	113
Estudio en invernadero de bacterias promotoras del crecimiento para mejorar la tolerancia de la quinua a la sequía <i>Greenhouse study of growth-promoting bacteria to improve drought tolerance in quinoa</i> Virginia Rigoberta Gonzales Mamani Suecia	114
Estructura, ultraestructura y acumulación de cationes en vejigas de sal de quinua <i>Structure, ultrastructure, and cation accumulation in quinoa salt bladders</i> Hernán Pablo Burrieza Argentina	116
Genes candidatos involucrados en principales procesos biológicos en respuesta al estrés salino, diferencian a los genotipos de quinua (<i>Chenopodium quinoa Willd.</i>) como tolerante o susceptible <i>Candidate genes involved in major biological processes in response to salt stress differentiate quinoa (<i>Chenopodium quinoa Willd.</i>) genotypes as tolerant or susceptible</i> José David Apaza Calcina México	117
Aplicación de Métodos Multivariados en el Análisis de Datos de Caracterización y Evaluación de Accesiones de Quinoa (<i>Chenopodium quinoa Willd.</i>), UNAP. Perú Jose Luis Cahuana Jorge Bolivia	118
Evaluación agronómica de 10 variedades de quinua tropical bajo condiciones agroclimáticas de la comunidad de Algarrobal, municipio de Yacuiba <i>Agronomic evaluation of 10 varieties of tropical quinoa under agro-climatic conditions in the community of Algarrobal, municipality of Yacuiba</i> José Gonzalo Herbas Meneses Bolivia	119



ANÁLISIS DE LA DIVERSIDAD GENÉTICA DE QUINUA (*Chenopodium Quinoa* Willd.) DEL BANCO DE GERMOPLASMA DE PUNO, MEDIANTE APLICACIÓN DE TÉCNICAS MULTIVARIADAS, 2021

Cahuana Jorge, José Luis¹
De la Torre Cleto, Arturo Zuñiga²

El propósito de esta investigación consistió en utilizar técnicas de análisis multivariado para describir la diversidad genética de la quinua (*Chenopodium quinoa* Willd.) presente en el Banco de Germoplasma de Puno. Se realizó desde junio del 2021 hasta abril de 2022, la fuente de información fue tomada del Catálogo de Banco de Germoplasma de quinua de Puno, los cuales fueron analizados, recogiendo información de 734 colecciones con registros completos, se aplicó las técnicas de clúster, discriminante y análisis de componentes principales (PCA). En el análisis clúster con distancia euclídea, se encontró dos grupos claramente diferenciados; para lo cual, mediante técnica discriminante se ha determinado que la variable “país de procedencia” de las colecciones es la que ha influenciado para la formación de los grupos, los mismos que se verificaron con los datos de pasaporte de las colecciones (Perú y Bolivia); en PCA, se ha reportado que en los 3 primeros componentes explican el 53.1% de la variación total de los datos, con un KMO-MSA de 0.83 que indica la pertinencia del análisis. En referencia a la técnica discriminante, las variables: “altura de planta” (altuPlan_cm), “días de panojamiento” (Panoj_dias), “rendimiento de grano por planta” (rendim_gpla) y “días de emergencia” (Emerg) fueron las que mejor discriminan a cada uno de los grupos encontrados, así mismo, el índice de diversidad encontrado promedio respecto a todas las variables es de 0.82 que se traduce en una diversidad moderadamente alta. Estos resultados encontrados tienen relación con lo indicado por otros autores, validando la genética y morfología de la colección del Banco de Germoplasma de quinua de Puno.

ABSTRACT

The purpose of this research was to use multivariate analysis techniques to describe the genetic diversity of quinoa (*Chenopodium quinoa* Willd.) present in the Puno Germplasm Bank. It was carried out from June 2021 to April 2022, the source of information was taken from the Puno Quinoa Germplasm Bank Catalog, which were analyzed, collecting information from 734 collections with complete records, clustering and discriminant techniques were applied. and principal component analysis (PCA). In the cluster analysis with Euclidean distance, two clearly differentiated groups were found; For which, using a discriminant technique, it has been determined that the variable “country of origin” of the collections is the one that has influenced the formation of the groups, the same ones that were verified with the passport data of the collections (Peru and Bolivia).); In PCA, it has been reported that the first 3 components explain 53.1% of the total variation in the data, with a KMO-MSA of 0.83 indicating the relevance of the analysis. In reference to the discriminant technique, the variables: “plant height” (altuPlan_cm), “days of tasseling” (Panoj_dias), “grain yield per plant” (rendim_gpla) and “emergence days” (Emerg) were the ones that They better discriminate each of the groups found, likewise, the average diversity index found with respect to all variables is 0.82, which translates into moderately high diversity. These results found are related to what was indicated by

¹ Ingeniero agrónomo, Universidad Nacional del Altiplano Puno, jluis.inia@hotmail.com

² Docentes de la Facultad de Ciencias Físicas, Químicas y Matemáticas de la Universidad San Antonio Abad del Cusco.

other authors, validating the genetics and morphology of the Puno quinoa Germplasm Bank collection.

Palabras clave: Análisis multivariado, clúster, discriminante, quinua, recursos genéticos, variabilidad genética.

INTRODUCCIÓN

La quinua (*Chenopodium quinoa* Willd.) es un cultivo andino que ha sido cultivado por más de 5000 años en la región andina de América del Sur, (Bazile et al, 2016), es considerado un alimento básico en la dieta de las poblaciones andinas debido a su alto valor nutricional y su capacidad para crecer en condiciones climáticas extremas (Carrasco et al, 2009). Además, la quinua es una fuente importante de ingresos para los agricultores de la región, ya que su demanda ha aumentado en los últimos años debido a su creciente popularidad en todo el mundo (Bhargava et al, 2006). El Banco de Germoplasma de Puno es una institución que se dedica a la conservación y el estudio de la diversidad genética de la quinua y otros cultivos andinos (INIA, 2021). La diversidad genética es importante para la adaptación de los cultivos a diferentes condiciones ambientales y para la mejora de las características agronómicas y nutricionales de las plantas (Frankel, 1984). Por lo tanto, el estudio de la diversidad genética de la quinua en el Banco de Germoplasma de Puno es fundamental para la conservación y el mejoramiento de este cultivo. El análisis de la diversidad genética de la quinua es crucial para su conservación y uso sostenible, especialmente en el contexto del cambio climático y la creciente demanda de alimentos nutritivos y resistentes a condiciones adversas (Jarvis et al, 2011). Además, el estudio de la diversidad genética de la quinua en el Banco de Germoplasma de Puno proporcionará información valiosa para el desarrollo de estrategias de conservación y mejoramiento genético de este cultivo en la región andina.

En este contexto, el objetivo de esta tesis es analizar la diversidad genética de la quinua del Banco de Germoplasma de Puno mediante la aplicación de técnicas multivariadas. Las técnicas multivariadas son herramientas estadísticas que permiten analizar múltiples variables al mismo tiempo y encontrar patrones y relaciones entre ellas (Jolliffe et al, 2002). El uso de estas técnicas en el análisis de la diversidad genética de la quinua permitirá identificar grupos de plantas con características similares y seleccionar las mejores plantas para su conservación y mejora.

MATERIALES Y MÉTODOS

En este estudio, fue no experimental, transversal y descriptivo, se ha llevado a cabo en el CIP Camacani a una altitud de 3824 msnm, se utilizó los datos de caracterización de las colecciones del Banco de Germoplasma de Puno, que consistió en una cantidad de 753 colecciones de quinua, de las cuales solo 734 colecciones tenían los datos completos con los cuales se hizo estudio aplicando las técnicas multivariadas del análisis de componentes principales (ACP) y el análisis de conglomerados. El ACP permitió reducir la dimensionalidad de los datos y encontrar patrones de variación entre las accesiones de quinua (Jackson, 1993). Por otro lado, el análisis de conglomerados agrupó las accesiones en función de su similitud genética, lo que permitió identificar grupos de plantas con características morfológicas y fisiológicas similares (Jain & Dubes, 1988). Estas técnicas multivariadas son ampliamente utilizadas en estudios de diversidad genética de cultivos y han demostrado ser eficaces para identificar variabilidad genética y patrones de agrupamiento en poblaciones de plantas (Peakall & Smouse, 2012).

Además, se calculó el índice de diversidad genética que permitió cuantificar la diversidad genética de todo el banco de germoplasma, todos estos análisis se hicieron utilizando el lenguaje de programación R aplicando comandos y paquetes para cada uno de los análisis.

RESULTADOS

Figura 1. Matriz de correlación de las variables cuantitativas.

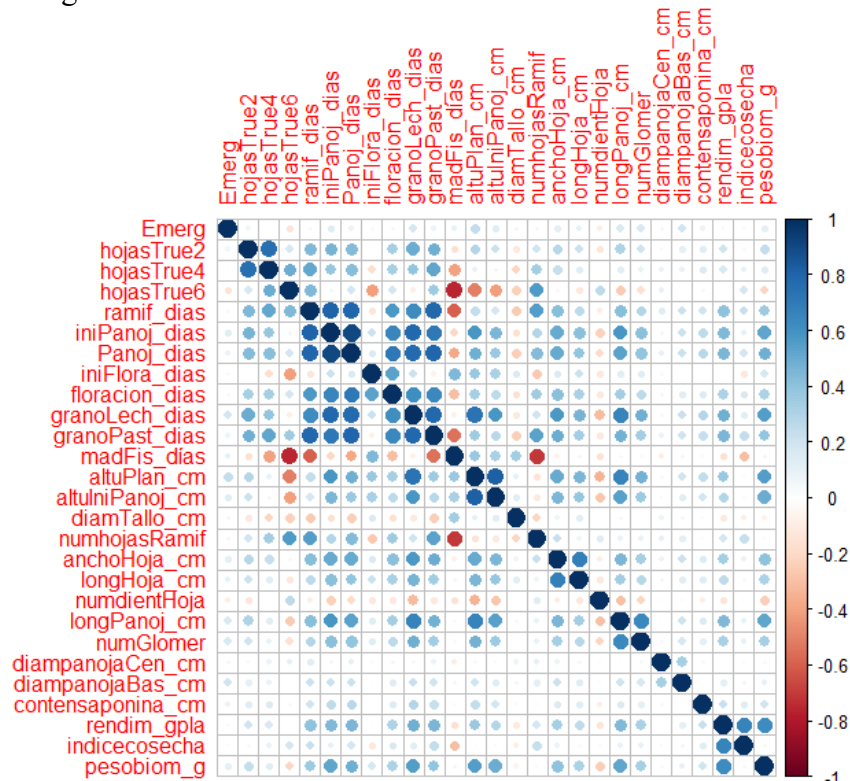


Figura 2. Biplot de las variables cuantitativas.



Figura 3. Matriz gráfica sobre la cantidad óptima de grupos en las colecciones de quinua.

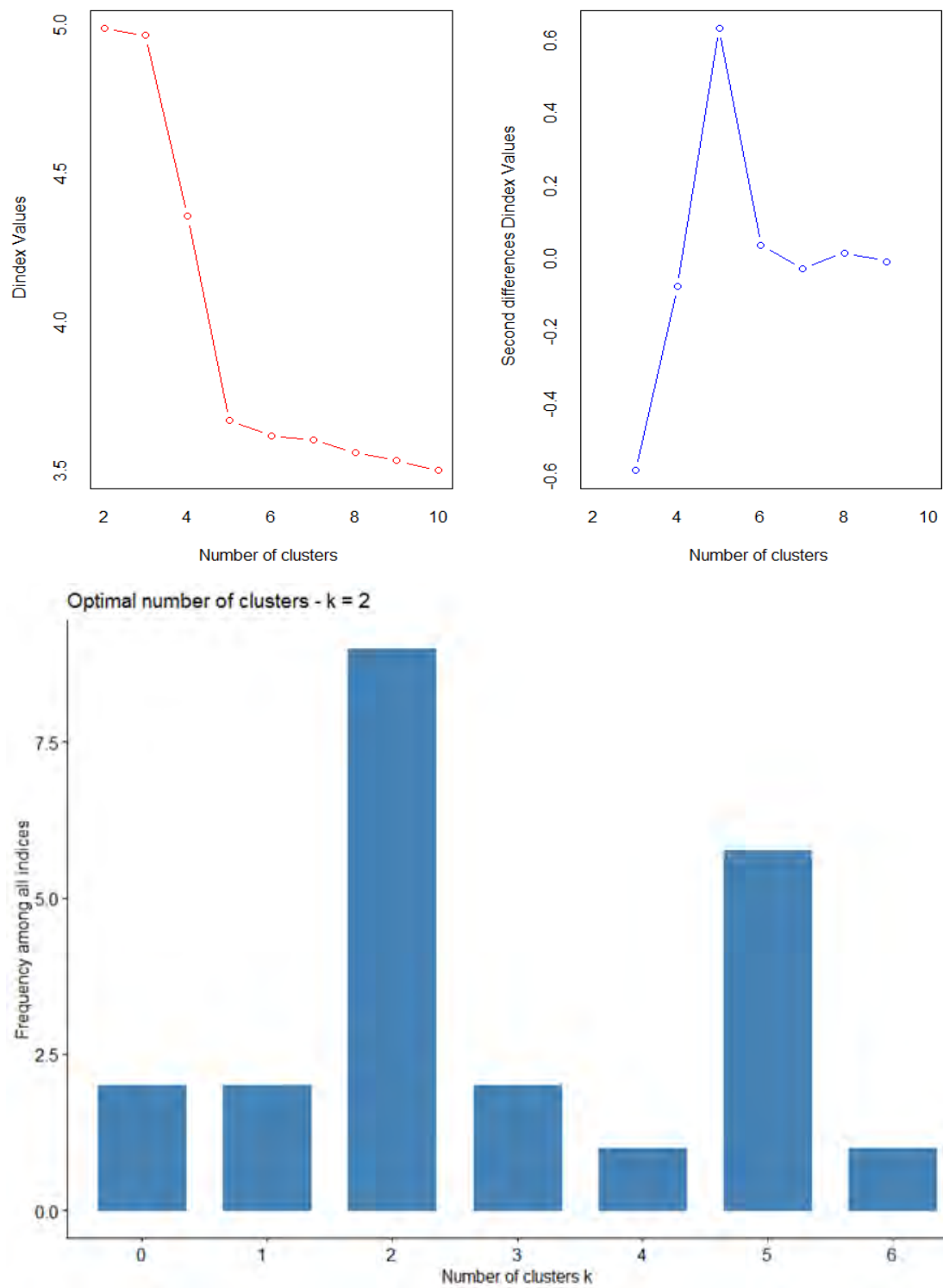


Figura 4. Dendrograma de las colecciones del banco de germoplasma de quinua

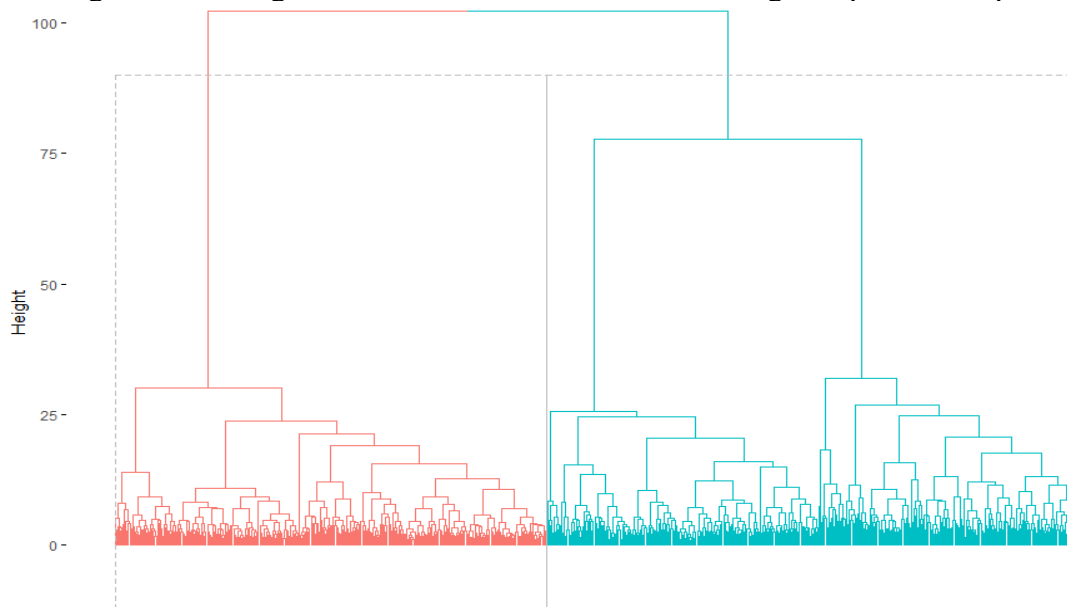


Tabla 1. Resultados de los coeficientes del análisis discriminante

Variable	Coefficiente
Emerg	0.00862
hojasTrue2	0.07462
hojasTrue4	-0.04701
hojasTrue6	-0.01754
ramif_dias	0.10592
iniPanoj_dias	0.05578
Panoj_dias	0.08806
iniFlora_dias	-0.15978
floracion_dias	0.13648
granoLech_dias	0.06306
granoPast_dias	0.01883
madFis_dias	-0.06729
altuPlan_cm	0.00938
altuIniPanoj_cm	0.00060
diamTallo_cm	-0.17408
numhojasRamif	0.00627
anchoHoja_cm	0.16535
longHoja_cm	-0.05015
numdientHoja	-0.04838
longPanoj_cm	0.03133
numGlomer	0.00454
diampanojaCen_cm	-0.04330
diampanojaBas_cm	0.02752
contensaponina_cm	-0.00940
rendim_gpla	0.00087
Indicecosecha	-0.30729
pesobiom_g	-0.00001

Tabla 2. Índice de diversidad de las variables

VARIABLE	INDICE DE DIVERSIDAD
Emerg	0.924623433
hojasTrue2	0.820450891
hojasTrue4	0.830460235
hojasTrue6	0.804912997
ramif_dias	0.852594423
iniPanoj_dias	0.852112117
Panoj_dias	0.813289944
iniFlora_dias	0.865519916
floracion_dias	0.848640976
granoLech_dias	0.784777636
granoPast_dias	0.780365895
madFis_días	0.799464024
altuPlan_cm	0.781451286
altuIniPanoj_cm	0.838325171
diamTallo_cm	0.941074257
numhojasRamif	0.919536608
anchoHoja_cm	0.848723288
longHoja_cm	0.855077453
numdientHoja	0.822355029
longPanoj_cm	0.833790252
numGlomer	0.837070466
diampanojaCen_cm	0.890461224
diampanojaBas_cm	0.918882145
contensaponina_cm	0.741098739
rendim_gpla	0.839953654
indicecosecha	0.818344505
pesobiom_g	0.862745641

DISCUSIÓN

Respecto a los componentes principales, en la Figura 1 se muestra un biplot con la representación de cada una de las colecciones analizadas así como las variables en forma de vectores, en ello se observa que los primeros cinco componentes se acumula un 63.2%, siendo las proporciones las siguientes: 31.2, 47.3, 53.1, 58.4 y 63.2, donde los primeros tres componentes explican el 53.1% de la variación en los datos, ello considerando los componentes 1, 2 y 3; además de ello, en el resultado de KMO (medida de adecuación muestral) se tiene un valor alto, ello indica su perfecta pertinencia de análisis. Este resultado se aproxima a los resultados encontrado por Dominguez (2014) que a nivel marcadores genéticos llegó a un 72% de acumulación de la variabilidad con cuatro componentes. Por otro lado, según los resultados de Morillo et al. (2022) el PCA de las variables cuantitativas los dos primeros componentes explicaron hasta el 88% de la variación total, ello se explica por la mayor cantidad de descriptores cuantitativos utilizadas. Así mismo, respecto a los resultados de Cárdenas (2017), realizado con 129 colecciones de quinua del mismo banco, señala haber encontrado tres grupos diferentes; sin embargo, no hay una especificación de país de origen de las colecciones, pero si es importante señalar la existencia de una alta diversidad y ello coincide con los resultados

encontrados en esta investigación. Para conocer la cantidad adecuada de grupos, se sometió los datos a un procesamiento del algoritmo de NbClust, la misma que por decisión mayoritaria (CONSENSO), se tiene como mejor número de agrupaciones la cantidad de 3 grupos sugeridos, este resultado contrasta con lo indicado por Cárdenas (2017), quien señala la existencia de una conformación de grupos diferentes entre las colecciones estudiadas, ello básicamente se diferencia por la presencia de pigmentación de la planta, inflorescencia, grano y otros, debido a las amplias adaptaciones de las colecciones a diferentes condiciones agroecológicas que ocurre entre las colecciones de Perú y Bolivia, las que muestran resultados diferentes. Respecto a los grupos en el dendograma de la Figura 4, se observa dos grupos grandes diferenciados; sin embargo, si se adopta un corte a menor distancia, se podrá tener varios subgrupos, estos resultados contrastados con la investigación realizada por Laura et al. (2023), se puede confirmar la existencia de resultados similares ya que obtuvo 10 grupos en todo el conjunto de colecciones analizadas. Por otro lado, respecto a las variables que discriminan los grupos los resultados de los coeficientes LD1 en la Tabla 1, se observa que las variables: “ancho de hoja” (anchoHoja_cm), “días de floración” (floracion_dias), “días de ramificación” (ramif_dias), “días de panojamiento” (Panoj_dias), “aparición de 2 hojas verdaderas” (Panoj_dias), “días formación grano lechoso” (granoLech_dias) y “días de inicio de panojamiento” (iniPanoj_dias), son las que mejor discriminan en el modelo para la formación de los grupos, la mayor parte de ellos son fenológicas, estos resultados son similares a lo encontrado por Cárdenas (2017) que señala que las variables discriminantes están conformadas por las variables fenológicas. Finalmente, en el cálculo del índice de diversidad de cada uno de los descriptores cuantitativos que se han considerado para la investigación, en la Tabla 2 se muestran valores por encima del 0.74 y que en promedio se tiene un índice de 0.8204509, lo cual indica que las variables en consideración presentan una diversidad moderadamente alta, esto significa que existe una variabilidad significativa en las mediciones morfométricas que esta tiene, estos resultados contrastan lo mencionado en los análisis antes expuestos e inclusive lo señalado por los estudios realizados por Cárdenas, (2017).

REFERENCIAS BIBLIOGRÁFICAS

- Bazile, D., Jacobsen, S. E., & Verniau, A. (2016). The global expansion of quinoa: trends and limits. *Frontiers in plant science*, 7, 622
- Cárdenas, R. G. (2017). Caracterización molecular de 129 accesiones de quinoa (*Chenopodium quinoa* Willd.) de la región Puno, mediante marcadores microsatélites Para optar el Título Profesional de Biólogo Genetista. January, 147.
- Carrasco Valencia, R., Acevedo de la Cruz, A., & Icochea-Álvarez, J. C. (2009). Quinoa (*Chenopodium quinoa* Willd.): composición, propiedades nutricionales y potencial uso en productos panificados funcionales. *Archivos Latinoamericanos de Nutrición*, 59(2), 135-142.
- Dominguez M., R. (2014). Caracterización agromorfológica de progenies autofecundadas S2, de cruza dobles genéticamente distantes y cercanas en quinoa (*Chenopodium quinoa* willd.) bajo condiciones de invernadero. Universidad Nacional San Agustín.
- Bhargava, A., Shukla, S., & Ohri, D. (2006). *Chenopodium quinoa*—an Indian perspective. *Industrial crops and products*, 23(1), 73-87.
- Instituto Nacional de Innovación Agraria. (2021). Banco de Germoplasma de Puno. Recuperado de <https://www.inia.gob.pe/banco-de-germoplasma-de-puno/>

- Frankel, O. H., & Brown, A. H. D. (1984). Plant genetic resources today: a critical appraisal. In *Crop genetic resources: conservation and evaluation* (pp. 249-257). Allen & Unwin.
- Jolliffe, I. T. (2002). *Principal component analysis*. Wiley Online Library.
- Jackson, D. A. (1993). Stopping rules in principal components analysis: a comparison of heuristical and statistical approaches. *Ecology*, 74(8), 2204-2214.
- Jain, A. K., & Dubes, R. C. (1988). *Algorithms for clustering data*. Prentice-Hall, Inc.
- Jarvis, D. I., Hodgkin, T., Sthapit, B. R., Fadda, C., & Lopez-Noriega, I. (2011). An heuristic framework for identifying multiple ways of supporting the conservation and use of traditional crop varieties within the agricultural production system. *Critical Reviews in Plant Sciences*, 30(1-2), 125-176.
- Laura, R., Alave Chata, D. E., Quispe Ramos, E., Colque Ramos, J., & Cora Machaca, C. G. (2023). La gestión educacional y producción de quinua orgánica (*Chenopodium Quinoa Wild*) del banco de germoplasma de la universidad nacional del Altiplano según Bioversity Internacional. *Ciencia Latina Revista Científica Multidisciplinar*, 7(2), 8260–8278. https://doi.org/10.37811/cl_rcm.v7i2.5954
- Morillo, A. C., Manjarres, E. E., & Morillo, Y. M. (2022). Phenotypic diversity of agromorphological characteristics of quinoa (*Chenopodium quinoa willd.*) germplasm in colombia. *Scientia Agricola*, 79(4). <https://doi.org/10.1590/1678-992X-2021-0017>