



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

**DEPARTAMENTO DE CIENCIAS DE LA VIDA
Y DE LA AGRICULTURA**

CARRERA DE INGENIERÍA EN BIOTECNOLOGÍA

**TESIS PREVIO A LA OBTENCIÓN DEL TÍTULO DE
INGENIERO EN BIOTECNOLOGÍA**

AUTOR: MORENO AMORES, JOSÉ ESTEBAN

TEMA: ESTUDIO DE LA DIVERSIDAD GENÉTICA DE CEBADA
Hordeum vulgare L. EN LA COLECCIÓN DEL INIAP USANDO
MARCADORES MOLECULARES SSR.

DIRECTOR: ING. PEÑA PONTÓN, CRISTIAN

CODIRECTOR: ING.-MAT. ROMERO SAKER, PEDRO

SANGOLQUÍ, JULIO 2014

CERTIFICACIÓN

Certifico que el presente trabajo fue realizado en su totalidad por el Sr. JOSÉ ESTEBAN MORENO AMORES como requerimiento parcial a la obtención del título de INGENIERO EN BIOTECNOLOGÍA.

Sangolquí, junio de 2014

Ing. Cristian Peña Pontón
DIRECTOR

Ing.–Mat. Pedro Romero Saker
CODIRECTOR

REVISADO POR:

Dra. María Augusta Chávez
DIRECTORA DE LA CARRERA

DECLARACIÓN DE RESPONSABILIDAD

Yo, José Esteban Moreno Amores

Declaro que:

El proyecto de grado denominado “ESTUDIO DE LA DIVERSIDAD GENÉTICA DE CEBADA *Hordeum vulgare* L. EN LA COLECCIÓN DEL INIAP USANDO MARCADORES MOLECULARES SSR” ha sido desarrollado con base a una investigación exhaustiva, respetando derechos intelectuales de terceros, conforme las citas que constan al pie de las página correspondientes, cuyas fuentes se incorporan en la bibliografía. Consecuentemente este trabajo es de mi autoría.

En virtud de esta declaración me responsabilizo de su contenido, veracidad y alcance científico.

Sangolquí, julio de 2014

José Esteban Moreno Amores

AUTORIZACIÓN

Yo, José Estaban Moreno Amores

Autorizo a la Escuela Politécnica del Ejército la publicación, en la biblioteca virtual de la Institución el trabajo “ESTUDIO DE LA DIVERSIDAD GENÉTICA DE CEBADA *Hordeum vulgare* L. EN LA COLECCIÓN DEL INIAP USANDO MARCADORES MOLECULARES SSR” cuyo contenido, ideas y criterios son de mi exclusiva responsabilidad y autoría.

Sangolquí, julio de 2014

José Estaban Moreno Amores

DEDICATORIA

“El hombre selecciona sólo para su propio bien; lo naturaleza lo hace sólo para el bien del ser que tiene a su cuidado.”

Charles Darwin

A Leonor, Héctor, Mauricio y Sofía... mi familia

José

AGRADECIMIENTO

¡Gracias a la vida! Definitivamente, por permitirme realizar esta investigación en el cultivo que fue la base de uno de los primeros alimentos que consumí con apenas unos meses de edad. He aquí, mi contribución.

Gracias a mi familia. A mis padres, Leonor y Héctor debido a quienes debo gran parte de lo soy. A mi hermano Mauricio por todo su apoyo y ejemplo brindado y a mi hermana y mejor amiga Sofi. Gracias a mis tíos Carlos y Cruz por su confianza.

Gracias al Instituto Nacional Autónomo de Investigaciones Agropecuarias INIAP y a SENESCYT, al Departamento Nacional de Biotecnología DNB y su líder Eduardo Morillo. Al Departamento de Recursos Fitogenéticos, al Programa de Cereales y en este a Luis Ponce y Esteban Falconí.

En mención aparte debo decir gracias a las personas que conocí en el DNB, gracias Katerine Orbe por tus sugerencias, apoyo, confianza y amistad. Gracias a Johanna Buitrón por brindarme su tiempo, paciencia, observaciones y consejos durante todo el trabajo de laboratorio. Gracias Sory, Seño Mary, Seño Geito; a todos mis compañeros y amigos tesisistas: Naty, Liz, Dianita, Migue (el aprendizaje también fue mío), Kari, Santy y a todo el departamento por haber hecho del tiempo que compartí en INIAP una de las mejores experiencias en mi vida.

Gracias a la colaboración de Carlos Jiménez, Denisse Peña y Jazmín Salazar en la prospección de semillas en el austro ecuatoriano. De igual manera a los investigadores Ariel Castro, José Vicente-Vallardón, Agnese Kolodinska y Jhonny Demey por su inmensa ayuda técnico-científica.

Gracias a la Universidad de las Fuerzas Armadas – ESPE, mi *alma máter*; al Ing. Pedro Romero y al Ing. Cristian Peña por su contribución en la dirección y codirección de este trabajo.

Gracias a mis entrañables amigos que me han acompañado desde la secundaria, a todos mis compañeros y amigos de la universidad y a todas las personas que han formado parte de este viaje. ¡Gracias!

LISTADO DE CONTENIDOS

CERTIFICACIÓN	i
DECLARACIÓN DE RESPONSABILIDAD.....	ii
AUTORIZACIÓN	iii
DEDICATORIA	iv
AGRADECIMIENTO	v
LISTADO DE CONTENIDOS.....	vi
LISTADO DE TABLAS.....	x
LISTADO DE FIGURAS	xiii
LISTADO DE ANEXOS	xvii
RESUMEN.....	xx
ABSTRACT.....	xxi
CAPÍTULO 1: INTRODUCCIÓN	1
1.1 Formulación del problema.....	1
1.2 Justificación	2
1.3 Objetivos.....	3
1.3.1 Objetivo general.....	3
1.3.2 Objetivos específicos	3
1.4 Marco teórico.....	4
1.4.1 Cebada (<i>Hordeum vulgare</i> L.).....	4
1.4.1.1 Origen y domesticación.....	4
1.4.1.2 Clasificación taxonómica.	4
1.4.1.3 Descripción botánica.	5
1.4.1.4 Tipos de cebada.	6
1.4.1.5 Genoma de la cebada.....	7
1.4.1.6 Evolución del cultivo en el Ecuador.....	8
1.4.1.7 Zonificación del cultivo de cebada en el Ecuador.....	9
1.4.1.8 Usos y valor nutricional.	10
1.4.1.9 Importancia económica y social.	12
1.4.2 Diversidad genética.....	13

1.4.2.1	Bancos de germoplasma.....	14
1.4.2.2	Germoplasma de cebada.....	15
1.4.3	Marcadores moleculares.....	16
1.4.3.1	Marcadores moleculares microsatélites.....	17
1.4.3.2	Marcadores moleculares disponibles para cebada.....	18
1.4.3.3	Reacción en cadena de la polimerasa.....	19
1.4.3.4	Plataforma de detección de marcadores microsatélites.....	20
1.5	Sistema de hipótesis.....	21
CAPÍTULO 2: MATERIALES Y MÉTODOS		22
2.1	Participantes.....	22
2.1.1	Instituciones.....	22
2.1.2	Responsable del proyecto	22
2.1.3	Colaboradores científicos.....	22
2.2	Zona de estudio.....	23
2.2.1	Fase de invernadero.....	23
2.2.2	Fase de campo.....	23
2.2.3	Fase de laboratorio.....	23
2.3	Periodo de tiempo de investigación.....	23
2.4	Procedimiento	24
2.4.1	Germoplasma en estudio.....	24
2.4.1.1	Material acriollado.....	24
2.4.1.2	Variedades mejoradas.....	25
2.4.1.3	Líneas promisorias.....	25
2.4.2	Recolección de tejido vegetal.....	25
2.4.3	Extracción de ADN genómico de cebada.....	26
2.4.4	Cuantificación de ADN.....	26
2.4.5	Validación de marcadores y de muestras de ADN.....	27
2.4.6	Estandarización de combinaciones de marcadores para amplificación.....	29
2.4.7	Determinación del número de individuos.....	30
2.4.8	Genotipaje de marcadores seleccionados SSR en LI-COR 4300S.....	30
2.4.9	Análisis de datos.....	31
2.4.9.1	Análisis de diversidad genética.....	31
2.4.9.2	Análisis multivariantes.....	33

2.4.9.3	Análisis de remuestreos.....	34
2.4.9.4	Biplot logístico externo	35
2.4.9.5	Análisis de estructura genética	35
2.4.9.6	Identificación de genotipos duplicados.....	37
2.4.9.7	Determinación de una colección núcleo de cebada.....	37
CAPÍTULO 3: RESULTADOS		38
3.1	Germoplasma analizado.....	38
3.2	Extracción, cuantificación y validación de ADN genómico de cebada.....	39
3.3	Validación y estandarización de marcadores y combinaciones para genotipaje en plataforma LI-COR® 4300S	41
3.4	Determinación del número de individuos.....	45
3.5	Genotipaje de marcadores SSR seleccionados en la plataforma LI-COR® 4300S	46
3.6	Análisis de datos	48
3.6.1	Diversidad genética.....	48
3.6.1.1	Diversidad genética de la colección completa.	48
3.6.1.2	Diversidad genética en el germoplasma de cebada acriollado	49
3.6.1.3	Diversidad genética en el germoplasma de cebada mejorado.....	51
3.6.1.4	Diversidad genética comparativa entre el germoplasma acriollado y de mejoramiento.....	53
3.6.2	Análisis de genotipos duplicados.....	54
3.6.3	Análisis de coordenadas principales	56
3.6.4	Análisis de conglomerados.	65
3.6.4.1	Construcción de grupos con el algoritmo UPGMA sobre la matriz de similaridad inicial.....	65
3.6.4.2	Construcción de grupos con el algoritmo UPGMA sobre las coordenadas principales.	69
3.6.4.3	Comparación de los grupos obtenidos en base a la matriz de similaridad inicial y las coordenadas principales.....	71
3.6.4.4	Diversidad genética inherente a los conglomerados ordinales.....	72
3.6.5	Análisis de remuestreos.	73
3.6.6	Análisis moleculares de varianzas complementarios	75
3.6.7	Estadísticos F	76
3.6.8	Biplot logístico externo.....	78

3.6.9	Determinación de un set núcleo de genotipos de cebada.....	83
CAPÍTULO 4: DISCUSIÓN.....		87
4.1.	Protocolo de extracción, rendimiento y calidad del ADN genómico.	87
4.2.	Marcadores microsatélites seleccionados.	88
4.3.	Determinación del número de individuos	90
4.4.	Diversidad genética.....	91
4.4.1.	Diversidad genética de cebada.....	92
4.4.2.	Diversidad genética comparativa.....	93
4.5.	Duplicados	96
4.6.	Análisis multivariados.	97
4.6.1.	Análisis de coordenadas principales	98
4.6.2.	Análisis de conglomerados	98
4.7.	Análisis de remuestreos	99
4.8.	Biplot logístico externo.....	100
4.9.	Set core.....	100
CAPÍTULO 5: CONCLUSIONES		102
CAPÍTULO 6: RECOMENDACIONES		104
BIBLIOGRAFÍA		105
ANEXOS		123

LISTADO DE TABLAS

Tabla 1:	Clasificación taxonómica de la cebada <i>Hordeum vulgare</i>	5
Tabla 2:	Contenido de macro y micronutrientes del grano de cebada (Yufera, 1987).	12
Tabla 3:	Descripción del germoplasma en estudio de acuerdo a su tipo, proveniencia.	24
Tabla 4:	Esquema del análisis molecular de varianza AMOVA.	36
Tabla 5:	Descripción del germoplasma en estudio de acuerdo a su tipo, proveniencia y número de accesiones o variedades.	39
Tabla 6:	Detalle del número de individuos de cada grupo de germoplasma de cebada de los que se extrajeron muestras de ADN genómico.	40
Tabla 7:	Descripción del número y proporción de alelos nulos para los marcadores Bmag0135, Bmag0853 y EBmac0701.	43
Tabla 8:	Detalle de las condiciones de temperatura y marcaje M13 para las combinaciones y marcadores en estudio.	44
Tabla 9:	Resumen del número de individuos y genotipos total y por cada grupo de germoplasma analizados.	46
Tabla 10:	Resumen del número de individuos y genotipos total y por cada grupo de germoplasma analizados.	46
Tabla 11:	Resumen de los parámetros de diversidad genética de 219 genotipos de cebada de la colección completa analizados con 16 loci SSR.	49
Tabla 12:	Resumen de los parámetros de diversidad genética de 156 genotipos de cebada acriollados analizados con 16 loci SSR.	50
Tabla 13:	Promedios de los parámetros de diversidad genética de los genotipos de tres grupos de germoplasma de cebada acriollados analizados con 16 loci SSR.	51
Tabla 14:	Resumen de los parámetros de diversidad genética de 63 genotipos de cebada de mejoramiento analizados con 16 loci SSR.	52
Tabla 15:	Promedios de los parámetros de diversidad genética de los genotipos de dos grupos de germoplasma de cebada mejorado analizados con 16 loci SSR.	53

Tabla 16: Resumen de la identificación de genotipos duplicados.....	55
Tabla 17: Valores Eigen, porcentaje absoluto y acumulativo de la varianza por coordenada principal del ACoP de 219 genotipos de cebada con 16 loci SSR.....	57
Tabla 18: Valores eigen, porcentaje absoluto y acumulativo de la varianza por coordenada principal del análisis ACoP de 219 genotipos de cebada con 16 loci SSR.....	58
Tabla 19: Análisis molecular de varianza (AMOVA) para 16 loci de la configuración con 6 clusters iniciales.	66
Tabla 20: Descripción de los clusters obtenidos mediante algoritmo UPGMA de 219 genotipos de cebada con 16 loci SSR.....	67
Tabla 21: Análisis molecular de varianza (AMOVA) para 16 loci de los 10 clusters ordinales.	69
Tabla 22: Descripción de los clusters obtenidos mediante algoritmo UPGMA en base a los resultados del ACoP de 219 genotipos de cebada con 16 loci SSR.....	70
Tabla 23: Análisis molecular de varianza para 16 loci de los grupos, establecidos <i>a priori</i> , de cebada acriollada y de mejoramiento.	75
Tabla 24: Análisis molecular de varianza para 16 loci de los clusters ordinales 1, 3, 9 y 10 obtenidos en el ACoP.....	76
Tabla 25: Análisis molecular de varianza para 16 loci de los clusters ordinales 2, 4, 5, 6, 7 y 8 obtenidos en el ACoP.....	76
Tabla 26: Estimación de Weir & Cockerham (1984) de los índices F_{IT} , F_{ST} , y F_{IS} para cada locus de las grupos establecidos <i>a priori</i> de cebada acriollada y de mejoramiento.	77
Tabla 27: Estimación de Weir & Cockerham (1984) de los promedios de los índices F_{IT} , F_{ST} , y F_{IS} de los clusters ordinales en distintos casos.....	77
Tabla 28: Alelos significativos ($R^2 \geq 0.5$, $p < 0.05$) resultantes del biplot logístico externo.	79
Tabla 29: Porcentajes de las predicciones de probabilidad de presencia los alelos más representativos ($R^2 \geq 0.5$, $p < 0.05$) en cada cluster ordinal.....	81

Tabla 30: Lista de 51 genotipos del set núcleo de la colección completa de 219 genotipos de cebada analizados con 16 loci SSR, el cluster ordinal al que pertenecen, tipo de material y el software con que fueron encontrados.	83
Tabla 31: Cuadro comparativo de los de diversidad genética para los 16 loci SSR en la colección completa y el set core de cebada.	85

LISTADO DE FIGURAS

Figura 1:	Espiguillas de cebada.....	6
Figura 2:	Espigas de cebada: izquierda: cebada dos carreras; derecha: cebada dos carreras	7
Figura 3:	Superficie cosechada, producción y rendimiento de cebada en Ecuador para el periodo 2002-2008 (FAO, 2014).....	9
Figura 4:	Zonificación del cultivo de cebada en el Ecuador (INEC, 2011).	10
Figura 5:	Participación del PIB de cebada en el PIB nacional y PIB agrícola del Ecuador en el periodo 2002-2009 (INEC, 2011).....	13
Figura 6:	Cartografía de los 21 marcadores SSR de cebada inicialmente considerados.....	28
Figura 7:	Localidades donde fueron colectadas las accesiones de cebada acriollada del banco de germoplasma y de las colectas realizadas en este estudio.....	38
Figura 8:	Cuantificación en gel de agarosa 1% (p/v) de muestras de ADN genómico de cebada (Carril 1 al 12: muestras, carril M: marcador de peso molecular).....	40
Figura 9:	Amplificación de 8 muestras de ADN de genómico de cebada con el marcador Bmac0209. (Carriles 1-8: muestras de ADN, carril M: marcador de peso molecular, carril (+): control positivo, carril (-): control negativo).....	41
Figura 10:	Amplificación de 8 muestras de ADN de genómico de cebada para tres marcadores microsatélites. a) Carriles 1-8: Bmac0213, carriles 9-16: EBmac0501. b) carriles 1-8: Bmac0399. Carriles M: marcador de peso molecular.	42
Figura 11:	Amplificación de muestras de ADN de genómico de cebada para dos marcadores microsatélites. a): Carriles 1-27: Bmag0692. b) carriles 1-10: Bmag0135; 11-20: EBmac0501. Carriles M: marcador de peso molecular.	43
Figura 12:	Perfiles de amplificación de primers SSR detectados por la plataforma LI-COR 4300S. a) con el marcador Bmac0209. b) Combinación <i>dúplex</i> de los marcadores Bmag0007 y Bmac0303.	44

- Figura 13:** Perfiles de amplificación para accesiones de cebada acriollada del Banco de Germoplasma. a) Con el marcador EBmac0501. b) Con el marcador Bmac0040. 45
- Figura 14:** Gel digital de una placa de 88 muestras de ADN de cebada amplificado con la combinación dúplex de los marcadores Bmag0007 y Bmac0303..... 47
- Figura 15:** Gel digital de una placa de 88 muestras de ADN de cebada amplificado con el marcador Bmac0314. 47
- Figura 16:** Comparación de los índices de diversidad genética entre el germoplasma acriollado y el mejorado. 54
- Figura 17:** Análisis de coordenadas principales a partir de la matriz de similitud calculada con el coeficiente SM (Sokal & Michener, 1958), de 219 genotipos de cebada con 16 marcadores SSR. Los genotipos están representados de acuerdo a su tipo de germoplasma. 59
- Figura 18:** Análisis de coordenadas principales ACoP a partir de la matriz de similitud calculada con el coeficiente SM (Sokal & Michener, 1958) de 219 genotipos de cebada con 16 marcadores SSR. a) Individuos correspondientes a materiales acriollados de cebada (grupos A.1, A2 y A.3) b) individuos de las variedades mejoradas de cebada (grupos A.1 y A.2) (véase anexos I.1 e I.2 para identificar cada individuo por su código). 60
- Figura 19:** Análisis de coordenadas principales a partir de la matriz de similitud calculada con el coeficiente SM (Sokal & Michener, 1958) de 219 genotipos de cebada con 16 marcadores SSR. a) Germoplasma tipo A.1, b) germoplasma tipo A.2, c) germoplasma tipo A.3, d) germoplasma tipo M.1 y e) germoplasma tipo M.2. 61
- Figura 20:** Análisis de coordenadas principales a partir de la matriz de similitud calculada con el coeficiente SM (Sokal & Michener, 1958), de 219 genotipos de cebada con 16 marcadores SSR. Se representan los genotipos del germoplasma tipo A.1 de acuerdo a la provincia de su proveniencia. 62

- Figura 21:** Análisis de Coordenadas Principales a partir de la matriz de similitud calculada con el coeficiente SM (Sokal & Michener, 1958), de 219 genotipos de cebada con 16 marcadores SSR. Se representan los individuos de los germoplasmas tipo A.1 que fueron recolectados en las provincias de Azuay, Cañar, Loja e Imbabura y los individuos del germoplasma tipo A.2..... 63
- Figura 22:** Análisis de Coordenadas Principales a partir de la matriz de similitud calculada con el coeficiente SM (Sokal y Michener, 1958) de 219 genotipos de cebada con 16 marcadores SSR, obtenido con el software NTSYS-pc 2.02. 64
- Figura 23:** Porcentaje de variación debido a la comparación entre configuraciones de clusters del dendrograma UPGMA de 219 genotipos de cebada con 16 marcadores SSR..... 65
- Figura 24:** Análisis de conglomerados de 219 genotipos de cebada con 16 loci SSR. a) Dendrograma obtenido con el algoritmo UPGMA con el programa NTSYS-pc 2.02. b) dendrograma obtenido con el algoritmo UPGMA en el programa MULTBILOT c) Representación en el ACoP de los clusters obtenidos con el algoritmo UPGMA (véase anexo I.3 para mejor resolución de los datos)..... 68
- Figura 25:** Porcentaje de variación debido a la comparación entre configuraciones de clusters del dendrograma UPGMA en base a al resultado del ACoP de 219 genotipos de cebada con 16 marcadores SSR..... 69
- Figura 26:** Análisis de conglomerados de 219 genotipos de cebada con 16 loci SSR en base a la matriz de coordenadas principales. a) Dendrograma obtenido con el algoritmo UPGMA en el programa MULTBILOT y b) Representación en el ACoP de los clusters obtenidos con el algoritmo UPGMA. 71
- Figura 27:** Análisis comparativo de pertenencia entre los clusters iniciales y ordinales obtenidos mediante el algoritmo UPGMA..... 72

- Figura 28:** Análisis comparativo de los índices de diversidad genética para los conglomerados o clusters ordinales. 73
- Figura 29:** Análisis de remuestreo sobre residuales tipo procrustes representado el diagrama del ACoP de 219 genotipos de cebada con 16 loci SSR..... 74
- Figura 30:** Biplot logístico externo del ACoP de 219 genotipos de cebada con 16 loci SSR, de un total de 157 alelos reportados se ilustran los 26 más representativos ($R^2 \geq 0.5$, $p < 0.05$). 80
- Figura 31:** Proyecciones binarias de la probabilidad de presencia de individuos sobre cuatro alelos. a) Para el alelo Ebmac0501(135bp), b) Bmac0127(124pb), c) Bmag0125 (109bp) y d) EBmac0679(148bp)..... 82
- Figura 32:** Proyecciones de la probabilidad de presencia de individuos sobre cuatro alelos SSR. a) Para el alelo Ebmac0501(135bp) b) Bmac0127(124pb) c) Bmag0125 (109bp) y d) EBmac0679(148bp). 83
- Figura 33:** Análisis de coordenadas principales, que muestra la distribución de los 51 genotipos del set core, calculado a partir de la matriz de similitud con el coeficiente SM (Sokal y Michener, 1958) de 219 genotipos de cebada con 16 marcadores SSR (véase anexo I.4 para identificar cada individuo por su código). 86

LISTADO DE ANEXOS

ANEXO A:	Colecciones de germoplasma de cebada en el mundo.....	123
ANEXO B:	Descripción de las 160 variedades de cebada analizadas.....	125
Anexo B.1:	Identificación de las 160 variedades de cebada de acuerdo a su tipo de germoplasma	125
Anexo B.2:	Datos pasaporte de las accesiones de cebada acriollada de los tipos A.1 y A.2	127
ANEXO C:	Lista de <i>primers</i> o cebadores de 21 marcadores microsatélite de cebada disponibles en el Departamento Nacional de Biotecnología – INIAP.....	129
ANEXO D:	Teoría estadística.....	131
Anexo D.1:	Análisis de coordenadas principales	132
Anexo D.2:	Análisis de conglomerados	133
Anexo D.3:	Análisis de remuestreos.....	135
Anexo D.4:	Biplot logístico externo	137
ANEXO E:	Concentración de ADN extraído y razón OD_{260nm}/OD_{280nm} de los 305 genotipos de cebada analizados a nivel de 16 marcadores SSR.....	139
ANEXO F:	Alineamiento de secuencias de los iniciadores o <i>primers</i> de los marcadores Bmac0399, Bmac0213, Bmac0134, Bmac0316 y Bmag0007 a sus respectivos <i>targets</i>	143
ANEXO G:	Matriz genotípica de pesos alélicos de 219 genotipos de cebada para 16 marcadores SSR.....	148
ANEXO H:	Índices de diversidad genética.....	159
Anexo H.1:	Frecuencias alélicas de los 16 loci SSR analizados en la colección completa de 219 genotipos de cebada.....	159

Anexo H.2:	Índices de diversidad genética de los individuos de los germoplasmas A.1, A.2 y A.3.....	167
Anexo H.3:	Índices de diversidad genética de los individuos de los germoplasmas M.1 y M.2.....	169
Anexo H.4:	Frecuencias alélicas de los genotipos de la colección completa, de acuerdo a cada cluster ordinal	171
Anexo H.5:	Valores promedio de los índices de diversidad genética de los genotipos de cebada de la colección completa de acuerdo a cada cluster ordinal.....	187
ANEXO I:	Representación de los genotipos de cebada acriollado y mejorado en el diagramas del ACoP a partir de la matriz de similitud calculada con el coeficiente SM (Sokal & Michener, 1958) de 219 genotipos de cebada con 16 marcadores SSR.....	186
Anexo I.1:	Genotipos de cebada acriollada de los tipos A.1, A.2 y A.3.	188
Anexo I.2:	Genotipos de cebada mejorada de los tipos M.1 y M.2.	189
Anexo I.3:	Análisis de conglomerados de 219 genotipos de cebada con 16 loci SSR. a) Dendrograma obtenido con el algoritmo UPGMA con el programa NTSYS-pc 2.02. b) dendrograma obtenido con el algoritmo UPGMA en el programa MULTBILOT c) Representación en el ACoP de los clusters obtenidos con el algoritmo UPGMA.....	190
Anexo I.4:	Análisis de coordenadas principales, que muestra la distribución de los 51 genotipos del set core, calculado a partir de la matriz de similitud con el coeficiente SM (Sokal y Michener, 1958) de 219 genotipos de cebada con 16 marcadores SSR	191
ANEXO J:	Conformación de los clusters iniciales y ordinales.....	192
Anexo J.1:	Conformación de los clusters iniciales.....	192
Anexo J.2:	Conformación de los clusters ordinales.	194

ANEXO K:	Estadísticos F para los casos en que fueron asumidas como subpoblaciones las configuraciones de cluster ordinales.....	196
-----------------	--	-----

RESUMEN

La cebada, *Hordeum vulgare* L., es el cuarto cereal de mayor producción en el mundo y en el Ecuador ha sido y es uno de los principales rubros en la alimentación de los habitantes de la región andina. El presente trabajo tuvo como objetivo generar información molecular que permita analizar la diversidad genética concerniente al germoplasma de cebada del INIAP. Se analizaron 219 genotipos de cebada correspondientes a 160 materiales entre variedades cultivadas, acriolladas y líneas promisorias. Se obtuvieron los perfiles alélicos con 16 loci SSR altamente polimórficos y distribuidos en los siete cromosomas de cebada. Se detectaron 157 alelos en un rango desde 104 hasta 232bp. El número de alelos por locus varió desde 3 hasta 14 con un promedio de 9.81. Excepto para el marcador Bmag0217 el índice de diversidad genética mostró altos valores, que variaron desde 0.60 hasta 0.85 y resultaron en un promedio de 0.75. Los análisis de conglomerados y el de coordenadas principales demostraron que no existe una diferenciación entre materiales acriollados y mejorados. Esto fue corroborado por análisis molecular de varianza y estadísticos F ($F_{ST}=0.014$) que mostraron además que una configuración con 10 clusters evidencia la mayor diferenciación ($F_{ST}=0.422$). El análisis mediante biplot logístico externo reveló 27 alelos significativos ($R^2 \geq 0.5$, $p < 0.05$) y sus predicciones de presencia de individuos pertenecientes a sus respectivos clusters. La gran diversidad genética de cebada reportada en este estudio debe ser técnicamente considerada en programas de fitomejoramiento y conservación de este germoplasma.

Palabras Clave: DIVERSIDAD GENÉTICA, CEBADA, SSR, ACRIOLLADAS, BILOT, GERMOPLASMA.

ABSTRACT

Barley, *Hordeum vulgare* L., is the fourth most produced cereal crop around the world and in Ecuador it has been one of the most important Andean people's foods. The principal aim of this study was to generate molecular data that allow the assessment of genetic diversity regarding to barley germplasm of INIAP. 219 barley genotypes comprising 160 varieties; among landraces, cultivars and breeding lines were analyzed with 16 loci SSR highly polymorphic and distributed among the seven barley's chromosomes. A total of 157 alleles were found with lengths ranging from 104 to 232bp. The number of alleles per locus was ranged from 3 to 14 with a mean of 9.81. The mean value of Nei's genetic diversity was 0.75, and excluding the marker Bmag0217, the values for this parameter were from 0.60 to 0.85. The average of polymorphic index content PIC was 0.71. Clustering Analysis and Principal Coordinate Analysis demonstrated that there was not a genetic differentiation between landraces and improved materials and this was corroborated by AMOVA analysis and F statistics ($F_{ST}=0.014$). The latest parameters also showed the highest differentiation ($F_{ST}=0.422$) and it was obtained when a configuration with 10 ordinal clusters was performed. External Logistic Biplot determined 27 representative alleles ($R^2 \geq 0.5$, $p < 0.05$) including 14 loci and their respective predictions of presence of individuals forming clusters. The high genetic diversity of barley reported in this study must be technically considered in breeding and conservation programs.

Keywords: GENETIC DIVERSITY, BARLEY, SSR, LANDRACES, BILOT, GERMPLASM.

CAPÍTULO 1: INTRODUCCIÓN

1.1 Formulación del problema

La cebada en la actualidad ocupa el cuarto lugar entre los cereales de mayor producción a nivel mundial con 132 millones de toneladas después del maíz, arroz y trigo; y un área cosechada de 49 millones de hectáreas en el año 2012 (FAO, 2014). La Unión Europea se destaca por ser la mayor zona exportadora de cebada y malta mientras que Arabia Saudí, Japón y China son los mayores importadores.

Según cifras de la FAO para el periodo comprendido entre los años 2000 y 2009 la producción y el rendimiento promedio de cebada por año en el Ecuador fueron de 23.585 t. y 0,60 t/ha respectivamente, colocando al país en el puesto número 82 en producción y 101 en rendimiento del ranking mundial de países productores de este cereal (FAO, 2014). Para el año 2012, en el Ecuador la superficie de cebada cosechada alcanzó 20.017 ha, distribuidas en todas las provincias de la Sierra, siendo Chimborazo, Imbabura y Cotopaxi las de mayor área con el cultivo (INEC, 2011).

Por su parte, el Instituto Nacional Autónomo de Investigaciones Agropecuarias INIAP a través del Programa de Cereales de la Estación Experimental Santa Catalina, ha trabajado por más de 30 años generando y transfiriendo nuevas tecnologías a los productores, enfocándose en la generación de germoplasma de cebada con características superiores de rendimiento, resistencia a enfermedades y calidad nutricional.

La comunidad científica ha manifestado el problema de la falta de caracterización y evaluación y la importancia de que las colecciones de germoplasma cuenten con suficientes datos fenológicos, morfológicos, agronómicos y moleculares (Frankel & Brown, 1984). La recolección y conservación de los recursos fitogenéticos sin su complemento de información sobre sus características, convierte a las colecciones en simples y poco útiles depósitos de materiales (Abadie & Berreta, 2001).

A largo de los años, los marcadores morfológicos han sido esenciales para la gestión de los recursos vegetales, y por ello es de gran importancia continuar con

dicha caracterización. Por otro lado, con el reciente advenimiento de técnicas moleculares, los estudios a este nivel ofrecen una serie de oportunidades, entre ellas la de ser una alternativa para estimar la diversidad genética de las colecciones (Westtamm & Kresovich, 1997), y por tanto ayudar a establecer criterios para mejorar la representatividad de las mismas.

Sin embargo; una caracterización molecular no necesariamente sustituye la realizada para caracteres morfológicos y agronómicos, ya que los dos tipos de información tienen historias evolutivas diferentes y pueden mostrar facetas diferentes de la diversidad (Abadie & Berreta, 2001).

No se han reportado estudios de diversidad genética en germoplasma de cebada, a nivel morfológico o molecular en el país. Por lo tanto es necesario analizar y evaluar el germoplasma que se encuentra en los campos de agricultores, “variedades acriolladas”, las variedades mejoradas generadas por los centros de investigación, así como los materiales promisorios en evaluación, para conocer su composición genética, para una futura orientación de este trabajo a la identificación de genes y características de interés agronómico.

1.2 Justificación

La cebada ha sido y es en la actualidad uno de los principales rubros en la alimentación de los habitantes de las zonas altas de la región andina como fuente de almidón, proteínas, vitaminas y minerales, y tiene gran potencial para que se incremente su consumo en las zonas urbanas. Hoy en día la cebada, a pesar de la reducción en su superficie cultivada, es después del maíz el cereal de más amplia distribución en la región interandina y el que mejor se adapta a zonas de alturas superiores a los 3000 m.s.n.m.

Una de las aplicaciones de los marcadores moleculares en Biotecnología agrícola es el estudio la de diversidad y variabilidad genética en germoplasmas. Dentro de estos marcadores, los microsatélites (SSR) tienen una alta capacidad de detección de polimorfismos además que poseen característica codominante para discriminar individuos homo y heterocigotos en especies diploides como la cebada.

El germoplasma en estudio está compuesto por variedades mejoradas y acriolladas de cebada, del que no se tiene información molecular por lo que se propone para su caracterización a este nivel, el uso de marcadores moleculares SSR. Estos resultados se complementarán con los resultados que se obtengan de la caracterización morfoagronómica de los mencionados genotipos, y brindarán los insumos para que el Programa de Cereales cuente con una “colección de trabajo” que concentre la mayor cantidad de diversidad genética disponible y en cuya base se sustenten los procesos de generación de germoplasma con caracteres deseados que considere emprender dicho programa.

1.3 Objetivos

1.3.1 Objetivo general

- Estudiar la diversidad genética de la colección de cebada *Hordeum vulgare* L. del INIAP utilizando marcadores moleculares microsatélites.

1.3.2 Objetivos específicos

- Caracterizar la diversidad alélica y el nivel de polimorfismo de 160 accesiones de cebada con 16 marcadores microsatélites SSR polimórficos distribuidos en los siete cromosomas del genoma.
- Identificar genotipos duplicados de la diversidad genética en la colección de cebada del INIAP.
- Estimar la base genética disponible en variedades acriolladas en relación a los genotipos de mejoramiento.
- Interpretar los datos de diversidad genética obtenidos y procesados en los paquetes bioinformáticos respectivos.

1.4 Marco teórico

1.4.1 Cebada (*Hordeum vulgare* L.)

1.4.1.1 Origen y domesticación.

La cebada *Hordeum vulgare* L. es uno de los cultivos más antiguos, y ha desempeñado un importante rol en el desarrollo de la agricultura y en el de las ciencias como la agronomía, fisiología y genética (Ullrich, 2010).

Existen restos arqueológicos de granos de cebada encontrados en la zona de la Media Luna Fértil (Oriente Próximo) que indican que este cultivo fue domesticado hace 10.000 años aproximadamente. Sobre la domesticación de la cebada, algunas teorías defienden la presencia de una sola línea evolutiva teniendo como único progenitor a *H. vulgare* ssp. *spontaneum* (Bothmer *et al.*, 2003). Sin embargo, hay evidencia molecular que sostiene que hay un origen polifilético de la cebada (Molina-Cano *et al.*, 2005; Komatsuda *et al.*, 2007; Morrell & Clegg, 2007; Saisho & Purugganan, 2007).

En América, la cebada fue introducida a través de los viajes de Cristóbal Colón y fue sembrada por primera vez por los españoles que se encontraban en la isla Isabela y Puerto Rico. En América del Sur es difícil determinar su primer cultivo pero es probable que haya llegado, en primera instancia, a los territorios de la actual Argentina junto con el trigo y de allí se haya propagado a la región andina (Arias, 1995).

En el país el cultivo de la cebada se remonta a tiempos de la conquista española, cuando por la necesidad de forraje para alimentar a los caballos fue cultivada con anterioridad al trigo (Coronel & Jiménez, 2011).

1.4.1.2 Clasificación taxonómica.

La cebada, así como otras gramíneas de importancia como el trigo o el arroz pertenece a la familia *Poaceae*. El género *Hordeum* tiene 32 especies, entre ellas el

ancestro de la cebada cultivada, *H. vulgare spontaneum* o cebada silvestre. En la Tabla 1 se detalla la clasificación taxonómica de la cebada.

Tabla 1: Clasificación taxonómica de la cebada *Hordeum vulgare*

Reino: Plantae
Sub-reino: Tracheobionta
División: Magnoliophyta
Clase: Liliopsida
Subclase: Commelinidae
Orden: Poales
Familia: Poaceae
Género: <i>Hordeum</i>
Especie: <i>H. vulgare</i> L.

Fuente: (Gramene Triticum, 2010).

1.4.1.3 Descripción botánica.

La cebada se caracteriza fundamentalmente por sus inflorescencias. Las flores de la cebada se agrupan para formar una espiga que tiene un eje central o raquis formado por una sucesión de nudos a partir de cada uno de los cuales se desarrollan tres espiguillas. Dichas espiguillas, a diferencia de especies cercanas como el trigo son unifloras. Cada flor consta de glumas en la parte exterior que son estrechas, alargadas y terminadas en una fina y corta arista (Figura 1) (Carrera & Mateo, 2005).

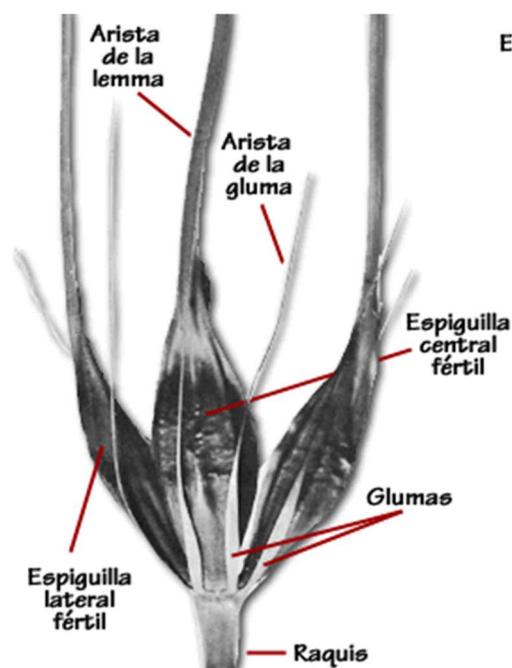


Figura 1: Espiguillas de cebada

La polinización se produce cuando la espiga aún no ha emergido de la cubierta formada por la vaina de la hoja bandera, por lo que la fecundación es autógama. El grano se asemeja al del trigo, excepto que el de la cebada posee cubierta después de la cosecha salvo en las cebadas de grano desnudo (Guerrero, 1999).

Adicionalmente, la cebada se presenta como una planta herbácea con sistema radicular fasciculado que es menos desarrollado que el del trigo y tiene el 60% del peso radicular en los primeros 20-25 cm. de profundidad. El tallo es más flexible que el del trigo y las hojas son más estrechas y de color más claro que las del trigo (Carrera & Mateo, 2005).

1.4.1.4 Tipos de cebada.

El germoplasma de cebada no es homogéneo, según se consideren criterios morfológicos, agronómicos, genéticos o de uso final. Uno de los criterios de clasificación más relevantes es de acuerdo a la morfología de la espiga.

En las cebadas de dos carreras, en la tríada de espiguillas presentes en cada entrenudo del raquis, sólo se desarrolla la espiguilla central, las laterales abortan. Sin embargo en las cebadas de seis carreras se desarrollan las tres espiguillas (Figura 2)

(Briggs, 1978). Los mejoradores se rehúsan a cruzar cebadas de diferentes tipos de carreras, debido a la laboriosa tarea de recuperar en la progenie los tipos estrictos por separado. Por esto, los germoplasmas de estos tipos de cebada constituyen una de las principales divisiones dentro de la diversidad genética del cultivo (Ponce, 2010).



Figura 2: Espigas de cebada: izquierda: cebada dos carreras; derecha: cebada dos carreras

1.4.1.5 Genoma de la cebada.

El género *Hordeum* contiene especies diploides ($2n=2x=14$), tetraploides ($2n=4x=28$), y hexaploides ($2n=6x=42$), aunque la mayoría son de la primera categoría (Briggs, 1978). La especie *H. vulgare* L. (cebada cultivada) es diploide. Comprende un genoma de más de 5.000 Mb que es aproximadamente 12 veces el tamaño del genoma del arroz del cual cerca del 80% de ADN (ácido desoxirribonucleico) es repetitivo (Flavell, Bennett, Smith, & Smith, 1974).

La cebada ha sido un cultivo favorito para la experimentación genética por varias razones: es un cultivo de importancia a nivel mundial; presenta un alto grado de autofecundación; es una especie anula de ciclo corto; es diploide con siete pares de cromosomas; exhibe una amplia diversidad morfológica, fisiológica y genética; se dispone de un amplio reservorio genético y existen mapas genéticos bien definidos (Forster *et al.*, 2000).

La cartografía comparativa entre el genoma de la cebada y otras especies de la tribu de *Triticeae* muestra que los cromosomas 1 a 7 de la cebada son homólogos de los cromosomas 7H, 2H, 3H, 4H, 1H, 6H, 5H, respectivamente, de las otras especies de esa familia (Linde-Laursen, Heslop-Harrison, Shepherd, & Taketa, 1997)

1.4.1.6 Evolución del cultivo en el Ecuador.

En el Ecuador de los años 60 la superficie dedicada a la siembra de cebada superaba las 100.000 ha. Sin embargo, esta se redujo progresiva y considerablemente hasta que en el año 1977 la superficie cultivada fue de 60.000 ha. y en el año 1980 fue de 26.000 ha.; debido a los efectos negativos producidos por la enfermedad conocida como roya amarilla de la cebada que afectó a todos los países de la región andina (Broers, 1994).

Desde 1971 año en que INIAP liberó la variedad de cebada INIAP-Dorada 71 hasta aquellas liberadas en 2003 que suman un total de 11 variedades, han sido varios los esfuerzos que junto con autoridades, se han emprendido para que se retome la producción cebadera en el país. No obstante; dicha recuperación ha sido lenta, debido entre otros motivos, a la insuficiente difusión entre los agricultores de las nuevas variedades mejoradas, falta de mecanismos ágiles de provisión de semillas, alto costo de insumos agrícolas que incluye el precio de semilla certificada y la baja rentabilidad del cultivo en la época de mayor comercialización (Broers, 1994).

En la última década, la producción y superficie cosechada de cebada han disminuido considerablemente respecto del promedio nacional para el mismo periodo, y para el año 2012, estas tan solo superaron las 10 mil toneladas y 20 mil hectáreas, respectivamente. En cuanto al rendimiento del cultivo, y pese a la disminución en la cantidad de cereal producido; no se reportan mayores variaciones excepto por una disminución en el mencionado año 2012 y un considerable aumento registrado en los años 2010 con 0,71 t./ha. y 2011 con y 0,86 t./ha. (Figura 3) (FAO, 2014). A pesar de lo mencionado; las cifras no son alentadoras, el país presenta incluso, los niveles de rendimiento más bajos del continente y la región andina detrás de Colombia, Perú y Bolivia respectivamente.

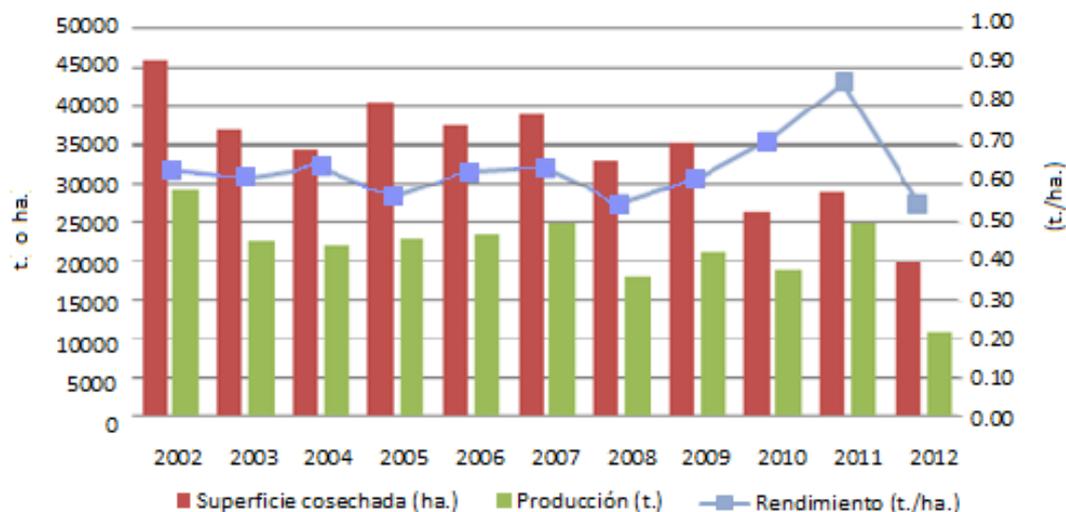


Figura 3: Superficie cosechada, producción y rendimiento de cebada en Ecuador para el periodo 2002-2008 (FAO, 2014).

1.4.1.7 Zonificación del cultivo de cebada en el Ecuador.

La cebada es ampliamente cultivada en el mundo y crece en ambientes muy diversos ecológicamente, incluyendo latitudes y altitudes extremas (Ullrich, 2010). Puede ser cultivada desde el Ecuador hasta Escandinavia, y desde las regiones más húmedas de Europa hasta los oasis del Sahara y Asia (Briggs, 1978). En el país, no existe una zonificación definida para este cultivo y tiene una potencial adaptación sin limitaciones ecológicas que se ubica en un rango altitudinal de 2500 a 3500 m.s.n.m.

El cultivo de cebada está concentrado en la región Sierra, y la provincia de Chimborazo es la más especializada ya que aporta con el 25% de la producción nacional. El Oro es la única provincia fuera de esta región que produce cebada, y aporta con apenas el 0,25% del total nacional (Figura 4) (INEC, 2011).

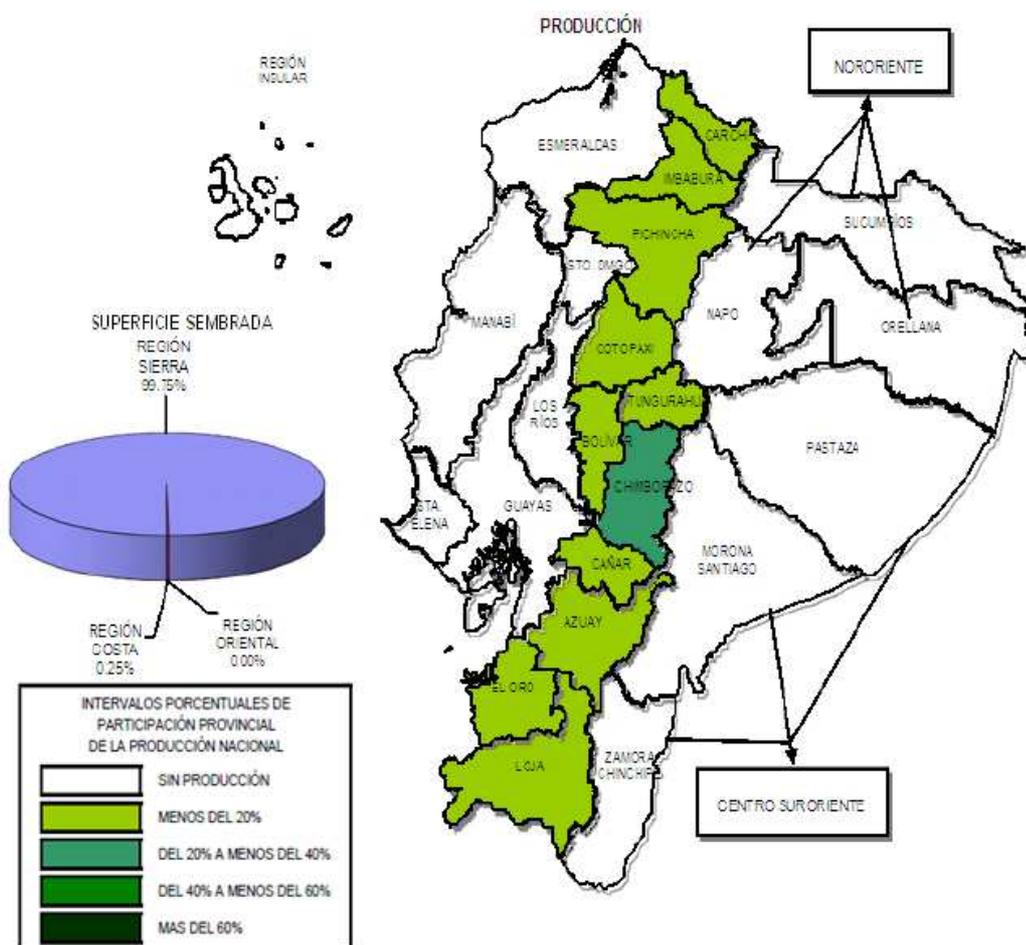


Figura 4: Zonificación del cultivo de cebada en el Ecuador (INEC, 2011).

1.4.1.8 Usos y valor nutricional.

Se estima que a nivel mundial, el 60% de la producción de cebada se destina para alimentar animales en forma de piensos y forraje; 30% para la producción de malta; 5% para alimentación humana y 5% se almacena como semilla (Ullrich, 2010).

La cebada que se destina a fines alimenticios puede ser utilizada de diferentes maneras, sin embargo la harina de cebada tostada o “machica” y el “arroz de cebada” son los productos de mayor demanda representado el 88% del consumo total del grano en el país (INEC, 2011). En el sector rural el consumo de cebada representa el

46% de la producción nacional que en 2009 alcanzó las 6.121 toneladas, lo cual se traduce en una demanda per cápita de 0,46 kg para ese mismo año (FAO, 2014).

De acuerdo a su uso final se puede establecer diferencias en cuanto al tipo de grano de cebada, siendo este cubierto o desnudo. Nutricionalmente, los primeros presentan ventajas sobre los cubiertos ya que poseen un menor contenido de salvado, pierden menos nutrientes durante su procesamiento y presentan mejor digestibilidad (León, 2010). En cuanto a las cebadas cubiertas, suele aplicarse un proceso industrial llamado perlado que consiste en la remoción y eliminación de las capas externas del grano, con la consecuente reducción de nutrientes (Villacrés, 2008).

La cebada es un alimento energético, rico en carbohidratos, especialmente en almidón (Tabla 2). En cuanto a su aporte proteico, es una fuente económica importante de aminoácidos como ácido glutámico, leucina y prolina; aunque deficiente en contenido de lisina y triptófano, déficit que puede ser solventado con el consumo complementario de granos andinos como el amaranto o quinua o de leguminosas como chocho, arveja o lenteja en una misma comida (Villacrés, 2008).

En comparación con otros cereales como el trigo o maíz, la cebada presenta altas cantidades de zinc, oligoelemento que forma parte de la insulina y la anhidrasa carbónica (Tabla 2). El grano perlado presente niveles más bajos de minerales, debido a la pérdida de las cubiertas que contienen el 32% de los minerales del grano (Villacrés, 2008).

Tabla 2: Contenido de macro y micronutrientes del grano de cebada (Yufera, 1987).

	Unidad	Cebada		
		Cubierta	Desnuda	Perlada
Proteína	%	12,2	13,3	12,0
Grasa	%	1,9	2,6	1,5
Carbohidratos totales	%	75,9	80,0	84,3
Carbohidratos fibra	%	6,8	1,9	1,0
Cobre	mg/kg	12	13	12
Hierro	mg/kg	94	72	26
Magnesio	mg/kg	24	19	7
Zinc	mg/kg	49	52	30
Calcio	%	0,05	0,06	0,02
Fosforo	%	0,54	0,47	0,24
Magnesio	%	0,12	0,12	0,07
Potasio	%	0,65	0,45	0,22
Sodio	%	0,1	0,04	0,02

1.4.1.9 Importancia económica y social.

De acuerdo a los datos del III Censo Nacional Agropecuario realizado en 2000, el cultivo de cebada reúne 67.155 UPA's (Unidad de Producción Agropecuaria), mientras que el cultivo de cebada asociado conjuga tan solo 2.278 UPA's. Los pequeños productores son mayoría tanto en el primero como en el segundo sistema de siembra, y es que representan el 78% y el 89%, respectivamente.

La cebada, es uno de los rubros que ha sido incluido dentro de la lista de bienes agrícolas que el Gobierno busca reducir su importación dentro de sus lineamientos de política económica mediante el proceso de Substitución Estratégica de Importaciones; cuya meta para el año 2013 fue la reducción de 40% en la participación de las importaciones de este cereal que en 2009 alcanzaron la cifra de 17 millones de dólares (INEC, 2011).

El comportamiento del PIB (Producto Interno Bruto) de cebada en el período 2002-2009, y su participación tanto en el PIB agrícola y el PIB nacional ecuatoriano, se muestran en la figura 1.5. La tasa de crecimiento del PIB de cebada es de 9,56% en el período mencionado, y en el último año llega a ubicarse en el 0,21% y en el 0,031% del PIB agrícola y del PIB nacional, respectivamente (INEC, 2011).

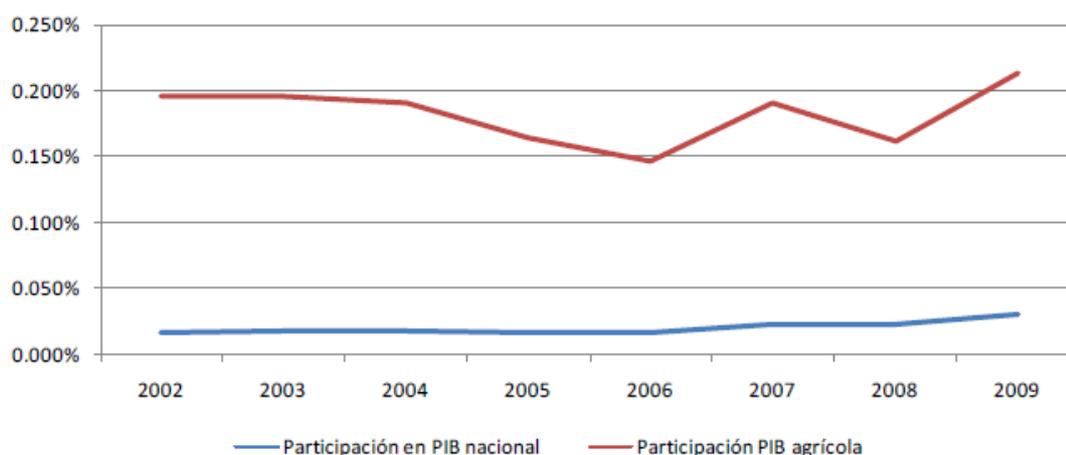


Figura 5: Participación del PIB de cebada en el PIB nacional y PIB agrícola del Ecuador en el periodo 2002-2009 (INEC, 2011)

La connotación social que tiene la cebada radica en que se ha constituido por centurias ahora en un alimento básico del campesino de la región interandina, que es un agricultor de subsistencia, debido a su tolerancia y adaptación a ambientes marginales de producción (INIAP, 2005).

1.4.2 Diversidad genética.

La diversidad genética en sentido amplio es el componente más básico de la biodiversidad y se define como las variaciones heredables que ocurren en cada organismo, entre los individuos de una población y entre las poblaciones dentro de una especie (Piñero, 2008). Se la puede definir como el grado en el cual el material hereditario diferencia internamente una colección de plantas (Demey, 2008).

El Convenio sobre la Diversidad Biológica, en su artículo 2, define a la biodiversidad o diversidad biológica como la variabilidad de organismos vivos de cualquier fuente, incluidos, entre otras cosas, los ecosistemas terrestres y marinos y otros ecosistemas acuáticos y los complejos ecológicos de los que forman parte. Por lo tanto, la biodiversidad se manifiesta a tres niveles: la diversidad a escala de los ecosistemas, la diversidad entre las especies y la diversidad dentro de cada especie (a escala de genes) (Estrella, Manosalvas, Mariaca, & Ribadeneira, 2005).

A partir de los procesos evolutivos que operan sobre la diversidad genética y sus variaciones se deriva la biodiversidad en los dos niveles restantes. De ahí que el

conocimiento y comprensión de la diversidad genética sea de vital importancia tanto para la conservación y el avance de la genética evolutiva, como para la salud pública, la sustentabilidad y la productividad agrícolas, pecuarias, pesqueras y forestales, la domesticación y la biomedicina (Piñero, 2008).

1.4.2.1 Bancos de germoplasma.

Puede entenderse como germoplasma al material genético que constituye la base física de la herencia y que se transmite de una generación a la sucesiva mediante las células germinales (FAO, 2013).

Los bancos de germoplasma poseen colecciones de material vegetal con el objeto de vivir y preservar sus características para el futuro beneficio de la humanidad y del ambiente. Resguardan la fuente de variabilidad requerida por los mejoradores de plantas para el desarrollo de cultivares que permitan al agricultor superar las limitaciones naturales a fin de obtener mayores beneficios de su actividad, así como asegurar la fuente contra la erosión genética (Beeching, Marmey, Hughes, & Charrier, 1994).

Los estudios de diversidad genética se basan en el grado de similitud entre individuos lo que permite la formación de grupos homogéneos que comparten un patrón de diversidad particular. Estos estudios constituyen una de las herramientas que ayudan a efectivizar el control sobre la erosión genética y además, permiten definir los patrones de variación que determinan la incorporación de individuos a programas de mejoramiento genético, ya sea por sus características promisorias o por su resistencia o susceptibilidad a condiciones bióticas o abióticas (Demey, 2008).

A continuación, se presentan algunas definiciones de términos importantes en la presente investigación:

- **Accesión:** Una muestra distinta, singularmente identificable de semillas que representa un cultivar, una línea de cría o una población y que se mantiene almacenada para su conservación y uso. Cada una de las entradas registradas en un banco de germoplasma.

- Variedad: La convención de la Unión Internacional para la Protección de las Obtenciones Vegetales UPOV define una variedad vegetal como:

“un conjunto de plantas de un solo taxón botánico del rango más bajo conocido que, con independencia de si responde o no plenamente a las condiciones para la concesión de un derecho de obtentor, pueda:

 - Definirse por la expresión de los caracteres resultantes de un cierto genotipo o de una cierta combinación de genotipos,
 - Distinguirse de cualquier otro conjunto de plantas por la expresión de por lo menos uno de dichos caracteres
 - Considerarse como una unidad, habida cuenta de su aptitud a propagarse sin alteración”
- Variedad acriollada: Germoplasma que durante siglos ha estado bajo los efectos de la selección natural y de la selección consciente de los agricultores para genotipos deseados. Son genéticamente variables, para caracteres cuantitativos y cualitativos, tienen una buena adaptación a condiciones ambientales específicas y tienen niveles de producción confiables (Harlan, 1992; Ceccarelli, 1996).
- Variedad cultivada o cultivar: incluye los productos finales de los programas de mejoramiento.
- Líneas promisorias: materiales genéticamente puros, que son parte de los procesos de mejoramiento y corresponden a líneas avanzadas de apareamientos.

1.4.2.2 Germoplasma de cebada

Los recursos genéticos de cebada pueden ser divididos en seis grupos principales:

- Cultivares modernos de uso actual
- Cultivares obsoletos

- Variedades criollas (*landraces*)
- Relativos silvestres del género *Hordeum*
- Stocks genéticos y citogenéticos
- Líneas promisorias (*Breeding lines*)

En el anexo A se enlistan las mayores colecciones de cebada en el mundo que albergan al menos 500 accesiones cada una, incluyendo dos colecciones menores de Estonia y Ecuador, que han ratificado el Tratado Internacional sobre los Recursos Fitogenéticos para la Alimentación y la Agricultura (ITPGRFA).

El total de 402.034 accesiones de cebada se aproxima al número estimado por la FAO y por el Inventario Global de Recursos Genéticos de Cebada (GBI). De este total existen 290.820 accesiones con tipo de germoplasma conocido, que a su vez consta de 15% de relativos silvestres, 44% pertenece a *landraces*, 17% líneas promisorias, 9% stocks genéticos y 15% cultivares.

1.4.3 Marcadores moleculares.

Los marcadores moleculares son definidos como “todo y cualquier fenotipo molecular oriundo de la expresión de un gen o estado alélico” (Ferreira & Grattapaglia, 1998).

Estos marcadores tienen varias características favorables en relación a los caracteres morfológicos o bioquímicos: permiten revelar un mayor número de polimorfismos al identificar también las mutaciones silenciosas y sinónimas escondidas en las proteínas; no están influenciados por factores ambientales o por interacciones génicas complejas; y son considerados en general caracteres neutrales, con muy poco o valor adaptativo nulo (Ajmone-Marsan *et al.*, 2001).

La puesta a punto de la Reacción en Cadena de la Polimerasa (PCR por sus siglas en inglés de Polymerase Chain Reaction, véase apartado 1.4.3.3 más adelante) a mediados de los años 80 revolucionó la genética molecular y particularmente la velocidad de desarrollo de nuevos marcadores basados en ADN (Liu, 1998). Los primeros marcadores basados en la PCR se desarrollaron a finales de los 80 y

principios de los noventa: los RAPDs (Random Amplified Polymorphic DNA) y los microsatélites o SSR (Simple Sequence Repeat).

1.4.3.1 Marcadores moleculares microsatélites.

Los marcadores microsatélites, también conocidos como Repeticiones en Tándem Simples (STRs), son repeticiones en tándem de máximo 6 pares de bases de longitud (van Oppen, Rico, Turner, & Hewitt, 2000). Se encuentran repetidos consecutivamente (de 5 a 20 veces) en el genoma con una longitud mínima de repetición de 12 pares de bases (Ellegren, 2004). Además estos marcadores pueden ser clasificados como:

- Perfectos simples: repeticiones en arreglos de una sola secuencia repetida. Ejemplo: $[AGG]_n$
- Imperfectos simples: consisten de una o más unidades de repetición de diferentes longitudes. Ejemplo: $[AAC]_n[ACT][AAC]_{n+1}$
- Perfectos compuestos: están compuestos de dos o más unidades de repetición diferentes de la misma longitud. Ejemplo: $[AGG]_n[AATC]_n$.
- Imperfectos compuestos: son interrumpidos por una o más repeticiones de diferente longitud $[GGAT]_n[ACT][GTAA]_{n+1}$ (Roy, Nazeer, & Saha, 2004).

Los microsatélites más frecuentes en plantas son $(A)_n$, $(AT)_n$, $(GA)_n$ y $(GAA)_n$, los cuales el 10% de las veces se ubican en regiones del genoma no codificantes. Los SSR exhiben altos niveles de polimorfismo debido a que las regiones del genoma con repeticiones en tándem son muy sensibles a sufrir errores durante el apareamiento de las cromátidas en el cross-over de la meiosis I y durante la replicación del ADN. Como consecuencia, las células hijas adquieren un número diferencial de repeticiones que son las que generan polimorfismo en la longitud de las bandas observadas en un análisis molecular basado en el uso de microsatélites. A pesar de su alto grado de polimorfismo, los microsatélites están delimitados por secuencias altamente conservadas, lo que permite el diseño de cebadores para su amplificación por PCR (Chistiakov, Hellemans, & Volckaert, 2006).

La abundancia, distribución relativamente uniforme a lo largo del genoma y el alto grado de polimorfismo, han convertido a los microsatélites en los marcadores moleculares de mayor acogida. Los SSR, por su carácter codominante, permiten diferenciar genotipos homocigotos y heterocigotos, son abundantes, confiables, informativos, están basados en técnicas sencillas de PCR, son de alta reproducibilidad y económicos cuando se cuenta con cebadores previamente diseñados (Kumapatla & Mukhopadhyay, 2005).

Existen muchos marcadores moleculares disponibles, pero para los fines de caracterización genética solamente se deben utilizar marcadores bien establecidos y repetibles como los SSR (Simple Sequence Repeat). En general, la amplificación de microsatélites ha demostrado que son más variables que isoenzimas, RFLP, AFLP y RAPD. Por ejemplo, en *Cucumis* 29 isoenzimas no mostraron polimorfismo entre dos genotipos comerciales. Dentro del mismo género marcadores RAPD detectaron un 38% de polimorfismo, mientras que los microsatélites detectaron un 71%, e incluso detectaron diferencias genéticas entre cultivares altamente emparentados que mediante otras técnicas no habían podido ser distinguidos (Becerra & Paredes, 2000).

1.4.3.2 Marcadores moleculares disponibles para cebada.

En cebada, los primeros mapas genéticos se basaron en marcadores RFLP (Graner *et al.*, 1991) y con el paso del tiempo los marcadores basados en PCR llegaron a ser el tipo de marcadores predominantes. Entre los diversos tipos de marcadores moleculares disponibles en cebada, los marcadores microsatélite han demostrado ser los marcadores de elección para programas de selección asistida por marcadores MAS (Marker-Assisted Selection) y estudios de diversidad genética (Varshney, Mahendar, Aggarwal, & Börner, 2007).

Ciertos mapas genéticos han sido construidos también usando marcadores moleculares SSR que son el tipo de marcadores preferidos para análisis en cereales (Ramsay *et al.*, 2000; Thiel, Michalek, Varshney, & Graner, 2003). A la fecha aproximadamente 1.000 marcadores SSR para cebada han sido publicados. Varshney *et al.* (2007) publicaron un mapa genético de alta densidad integrando 775 loci SSR y Marcel *et al.* (2007) construyeron un mapa con 3.258 marcadores moleculares de los

cuales 503 marcadores fueron marcadores SSR y mostraron ser los de mayor nivel de polimorfismo.

En América Latina han habido esfuerzos por establecer programas de mejoramiento del germoplasma de cebada que comprenden estudios de caracterización genómica del mismo, como es el caso de Castro *et al.* (2010) en Uruguay, sin embargo no hay referencias de este tipo de estudios en este cultivo en los países andinos.

1.4.3.3 Reacción en cadena de la polimerasa.

La reacción en cadena de la polimerasa o PCR (siglas de su nombre en inglés (Polymerase Chain Reaction) es una técnica que permite generar una gran cantidad de copias de un fragmento de ADN. El requisito fundamental para llevar a cabo la reacción es disponer de fragmentos cortos de ADN de cadena sencilla complementarios a los extremos del fragmento a amplificar que servirán de cebadores o *primers* para que una enzima polimerasa incorpore nucleótidos a la cadena molde. Los componentes de la reacción son:

- ADN molde: La molécula de ADN a partir del cual queremos obtener una copia de un fragmento
- Enzima ADN polimerasa: capaz de generar una copia de ADN a partir del ADN molde.
- Cofactores de la enzima: cationes divalentes, generalmente en forma de cloruro de magnesio $MgCl_2$.
- Nucleótidos libres: se añaden en forma de desoxirribonucleósidos trifosfato (dNTPs).
- Cebadores o *primers*: moléculas de ADN simple cadena (entre 10 y 30 pares de bases).

El proceso se lleva a cabo en tres fases: desnaturalización, alineamiento y elongación.

En primer lugar es necesario que el DNA se desnaturalice, es decir, que las dos cadenas de DNA se separen. Esta primera fase se conoce como desnaturalización y se lleva a cabo elevando la temperatura a alrededor de 94°C.

El siguiente paso consiste en un descenso de la temperatura para permitir que los cebadores se unan por complementariedad al DNA molde. Esta segunda fase se conoce como hibridación. Temperaturas habituales en esta fase oscilan entre 35 y 60°C.

Por último, en la fase de elongación o extensión, la enzima polimerasa incorpora nucleótidos complementarios a partir del extremo 3' libre de la región en que han hibridado los cebadores. La temperatura a la que se lleva a cabo esta fase depende de la enzima polimerasa empleada; si se utiliza Taq polimerasa la temperatura de elongación suele ser de 72°C.

Con un ciclo, un solo segmento de molde de DNA de doble cadena se amplifica en dos piezas separadas de ADN de doble cadena. Estas dos piezas son entonces disponible para la amplificación en el siguiente ciclo. A medida que los ciclos se repiten, más y más copias se generan y el número de copias de la plantilla se incrementa exponencialmente.

1.4.3.4 Plataforma de detección de marcadores microsatélites.

Antes de los años 90 y del advenimiento de secuenciadores automatizados, muchos laboratorios usaban geles de agarosa para la detección de los alelos de microsatélites; pero la desventaja principal era la pérdida de resolución entre alelos que difieran en pocas pares de bases, en comparación con el alto poder de resolución de los geles de poliacrilamida.

En la actualidad, varios laboratorios tiene acceso a al menos un tipo de sistema automatizado (capilar o en base a gel). Estos secuenciadores están basados en un fluoróforo que es incorporado en los productos PCR que permite la detección de los alelos microsatélites. Un láser de diodo, que emite una longitud de onda específica de luz, es absorbido por el fluoróforo. Luego, un detector captura la luz emitida por el

fluoróforo y digitaliza la excitación de luz en la forma de una banda en un gel o un cromatograma.

Una de las opciones existentes para la obtención de *primers* o cebadores, es que sean enviados a sintetizar de manera que solo el primer forward sea marcado con un fluoróforo añadido al extremo 5' y el primer *reverse* no sea marcado. Otra opción puede ser escoger la estrategia M13-tailing para reducir significativamente los costos totales (Schuelke, 2000). En esta estrategia se requieren dos *primers* no marcados (forward y reverse) y un segmento o primer M13 marcado fluorescentemente. El primer forward SSR consiste de una cola M13 específica en el extremo 5' de 19 pares de bases (CACGACGTTGTAAAACGAC). Los primers M13 marcado y reverse están en exceso en relación al primer forward. Esto permite que los primers reverse y forward M13 marcado inicien la reacción y cuando el primer limitado sea empobrecido, el primer M13 marcado tome su lugar en los siguientes ciclos PCR.

Una vez que los alelos sean amplificados, deben ser separados y determinados con un software apropiado, por ejemplo, para la plataforma LI-COR 4300S es el software SAGA^{GT}. A pesar de la automatización de las plataformas, los datos deben ser cuidadosamente evaluados por una persona conocedora y corregidos también por ciertas inconsistencias que pueda tener el software.

1.5 Sistema de hipótesis

Existe diversidad genética en la colección de cebada del INIAP a nivel de los 16 marcadores moleculares SSR seleccionados.

CAPÍTULO 2: MATERIALES Y MÉTODOS

2.1 Participantes.

2.1.1 Instituciones.

Las instituciones auspiciantes de esta investigación son la Secretaría de Educación Superior, Ciencia, Tecnología e Innovación SENESCYT y el Instituto Nacional Autónomo de Investigaciones Agropecuarias INIAP como parte del proyecto: “Desarrollo e innovación biotecnológica para la potenciación de rubros agrícolas de importancia en seguridad alimentaria, competitividad exportable y adaptación al cambio climático” (CODIGO SENESCYT PIC-12-INIAP001).

2.1.2 Responsable del proyecto

José Esteban Moreno Amores

2.1.3 Colaboradores científicos.

Eduardo Morillo, Ph.D.	Líder del Departamento Nacional de Biotecnología - INIAP
Ing. Luis Ponce M.Sc.	Programa Nacional de Cereales - INIAP
Ing. Cristian Peña Pontón	Director de tesis
Ing.-Mat. Pedro Romero Saker.	Codirector de tesis

2.2 Zona de estudio.

2.2.1 Fase de invernadero.

Las muestras vegetales de las accesiones de cebada (*Hordeum vulgare* L.) de las colectas de semillas realizadas en Imbabura y en el Austro se recolectaron de plántulas sembradas en condiciones de invernadero del Departamento Nacional de Biotecnología en la Estación Experimental Santa Catalina EESC ubicada en la parroquia Cutuglagua, cantón Mejía en la provincia de Pichincha. Esta estación se encuentra a una altitud de 3050 m.s.n.m, latitud 0°22'15''S y longitud 78°33'14''O, posee una precipitación anual de 1400 mm y una temperatura media de 11.6 °C.

2.2.2 Fase de campo.

El muestreo de material vegetal de las variedades de cebada de la Colección Nacional y de las colectas de semillas del Programa de Cereales del INIAP además de las accesiones de cebada del Banco de Germoplasma del INIAP fue llevado a cabo en los ensayos de campo que el mencionado programa mantiene en la EESC.

2.2.3 Fase de laboratorio.

La fase experimental fue llevada a cabo en el Laboratorio de Biología Molecular del Departamento Nacional de Biotecnología en la EESC.

El mencionado laboratorio cuenta con distintas salas dependiendo de los procedimientos que se lleven a cabo como son las de preparación de medios, extracción de ADN, PCR, sala oscura, electroforesis vertical, genotipaje, autoclavado y sala de gases.

2.3 Periodo de tiempo de investigación.

La presente investigación fue llevada a cabo en 15 meses, desde octubre de 2012 hasta diciembre de 2013.

2.4 Procedimiento

2.4.1 Germoplasma en estudio.

El germoplasma en estudio constó de 160 variedades de cebada de distinto tipo, tanto acriolladas como provenientes de programas de mejoramiento (Anexo B.1). Los materiales acriollados provinieron de accesiones del Banco de Germoplasma y de colectas de semillas, mientras que el resto de los materiales fueron conformados por variedades cultivadas y líneas promisorias. En resumen, se clasificaron las variedades de cebada en cinco grupos de acuerdo al tipo de germoplasma y a su proveniencia (Tabla 3).

Tabla 3: Descripción del germoplasma en estudio de acuerdo a su tipo, proveniencia.

Grupo	Código interno	Tipo de material	Proveniencia	
A.1	CDF	Acriollado	Accesiones	Banco de Germoplasma INIAP
A.2	CCA	Acriollado	Variedades de prospecciones	Colectas Austro e Imbabura
A.3	CCU	Acriollado	Variedades de feria de semillas	Programa de Cereales
M.1	CCN	Mejorado	Variedades Colección Nacional	Programa de Cereales
M.2	CLP	Mejorado	Líneas Promisorias	Programa de Cereales

2.4.1.1 Material acriollado.

Se realizaron dos prospecciones de germoplasma de cebada, con el fin de coleccionar semillas de cebadas acriolladas y complementar las existentes en el Banco de Germoplasma de INIAP. La primera prospección fue en el Austro ecuatoriano en las provincias de Cañar, Azuay y Loja y la segunda en la provincia de Imbabura. Los datos pasaporte de las accesiones de cebada acriollada de los grupos A.1 y A.2 analizadas se enlistan en el anexo B.2.

Adicionalmente, en este estudio se tomaron en cuenta 23 materiales de cebada acriolladas recolectados por el Programa de Cereales en una feria de semillas llevada a cabo en Alausí, en el año 2012.

Las semillas de las accesiones de cebada colectadas (grupos A.2 y A.3) fueron sembradas directamente en vasos de plástico que contenían un sustrato a base de tierra negra y mantenidos en condiciones de invernadero hasta su germinación.

2.4.1.2 Variedades mejoradas.

Esta investigación contempló el estudio de las variedades que el Programa de Cereales de INIAP maneja con su Colección Nacional de 51 variedades mejoradas, entre las que están las 11 variedades comerciales que el programa ha liberado oficialmente desde el año 1971.

2.4.1.3 Líneas promisorias.

Se estudiaron también, 11 líneas promisorias de cebada que el Programa de Cereales mantiene en distintos ensayos; adicionalmente de una línea de cebada colombiana perteneciente a CORPOICA.

2.4.2 Recolección de tejido vegetal.

En el caso de los individuos pertenecientes a los grupos de germoplasma de cebada A.1, M1 y M.2 el tejido vegetal (foliar fresco) para el proceso de extracción de ADN fue colectado directamente de plantas de cebada en periodo de espigazón, sembradas en parcelas a cargo del Programa de Cereales.

Por otro lado, en las variedades de los grupos A.2 y A.3 que fueron sembradas en invernadero, se recogieron muestras de tejido foliar joven de plántulas.

El número de individuos tomados en cuenta por accesión, variedad o línea de cebada difirió tomando en cuenta el criterio de homogeneidad genética dentro de las mismas. Es así, que se recogieron muestras de 10 individuos diferentes de cada una

de las accesiones de los grupos A.1 y A.2; pero, se incluyeron 2 individuos diferentes por cada una de las variedades de los grupos A.3, M1 y M.2.

2.4.3 Extracción de ADN genómico de cebada.

Para la extracción de ADN se utilizó el protocolo descrito por Pallotta *et al.* (2003). En microtubos de 1,7 ml con 0,5 g de muestra (tejido foliar fresco) se añadió 500µl de una solución buffer de extracción (0.1M Tris-HCl pH 7,5; 0,05M EDTA pH 8,0; 1,25% SDS) se maceró y agitó la mezcla. Posteriormente, los tubos fueron incubados a 65 °C en baño maría por 1 hora. Se enfriaron los tubos hasta temperatura ambiente antes de añadir 250 µl de acetato de amonio 6 M, se agitaron los tubos con la nueva mezcla y se dejaron reposando a 4 °C por 15 minutos. Una vez finalizado este tiempo se centrifugaron los tubos a 5.000 rpm por 15 minutos para separar las proteínas precipitadas y el tejido vegetal de la fase acuosa. A 600 µl de este sobrenadante se añadieron 360 µl de isopropanol. Se agitaron los nuevos tubos y se dejó precipitar el ADN a -20 °C por al menos 1 hora. A continuación se centrifugaron los tubos por 15 minutos a 5.000 rpm para la formación del pellet de ADN y se secaron los tubos por inversión sobre papel con cuidado de no perder el pellet de ADN. Se hicieron dos lavados del pellet añadiendo 1.000 µl de etanol 70 % y uno con etanol 95 % y centrifugándolos por 20 minutos a 5.000 rpm. Finalmente se secaron los tubos y se resuspendió el pellet en 100 µl de buffer TE 0,1 M.

2.4.4 Cuantificación de ADN.

La cuantificación de ADN genómico de cebada se realizó utilizando el espectrofotómetro para microplacas EPOCH™ de Biotek®. Mediante el software Gen5™ en el que se configuró el número de muestras para el dispositivo para lectura de microvolúmenes Take3™ y se cargaron 2,0 µl de volumen por muestra al igual que los lugares en donde se hizo la lectura de muestras blanco, que fue la solución en la que se diluyó el ADN. Una vez realizada la lectura de muestras, el programa resume los valores de absorbancias a 260 nm. y 280 nm. y la concentración de ADN en ng/µl.

El valor de la concentración de ADN obtenido fue llevado hasta una de 5 ng/μl que es la concentración inicial del templado para la reacción de amplificación PCR con marcadores microsatélites.

Las muestras de ADN genómico también fueron cuantificadas en geles de agarosa 1% (p/v) mediante electroforesis horizontal. Se tomaron 4 μl de ADN y se mezcló con 1 μl de Blue Juice 1x Invitrogen No. Cat. 10816-015 y se procedió a cargar el gel también con un marcador de peso molecular Low DNA Mass Ladder Invitrogen No. Cat. 10068-013. La solución buffer de corrida fue TAE 1X en condiciones de voltaje constante de 100 V durante 30 minutos y posteriormente se realizó la tinción de los mismos en una solución de bromuro de etidio (15 ppm) por un periodo de tiempo de 15-20 minutos en agitación continua (Morillo & Miño, 2011).

2.4.5 Validación de marcadores y de muestras de ADN.

Se seleccionaron y enviaron a sintetizar 21 primers para los marcadores microsatélites escogidos en base a la información reportada por Ramsay *et al.* (2000), Wang *et al.* (2010), Chen *et al.* (2012) y Kolodinska-Brantestam *et al.* (2012) que corresponden a 3 marcadores por cromosoma seleccionados por su índice de polimorfismo PIC (Polimorfism Index Content) y secuencia motivo (Figura 6) (anexo C).

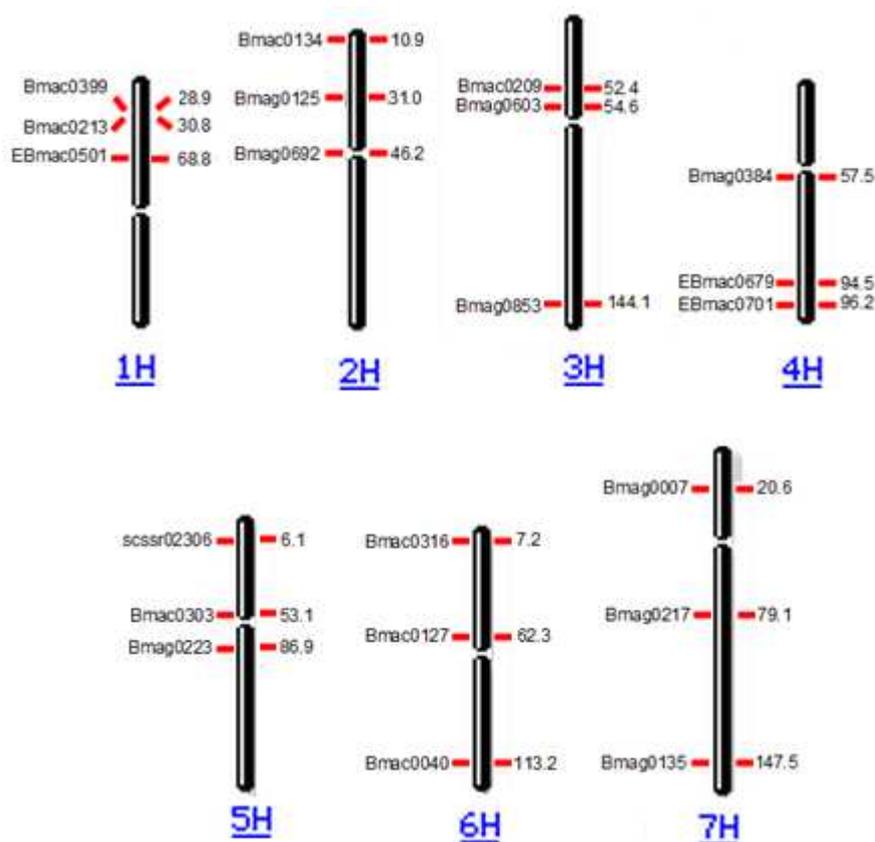


Figura 6: Cartografía de los 21 marcadores SSR de cebada inicialmente considerados.

Se escogieron al azar ocho muestras de ADN para ser amplificadas con cada uno de los marcadores con la metodología descrita por Morillo & Miño (2011) con las siguientes concentraciones finales en un volumen de reacción de 7,5 μ l: ADN (5,0 ng/ μ l): 1,33 ng/ μ l; Buffer GoTaq PROMEGA® (10 X): 1X; MgCl₂ PROMEGA® (25,0 mM): 2,0 mM; dNTPs INVITROGEN®: 0,235 mM; Primer Forward y Reverse (10 μ M): 0,5 μ M; Taq DNA Polimerasa PROMEGA® (5,0 U/ μ l): 0,067 U/ μ l. El programa de amplificación fue el siguiente: Un ciclo de denaturación inicial de 94 °C por 5 minutos; seguido de la repetición de 30 ciclos con 94 °C por 45 segundos, ° TA por 1 minuto y 72 °C por 2 minutos. Un ciclo de extensión final de 72 °C por 7 minutos y uno de 10 °C por 5 minutos.

Los productos PCR se visualizaron mediante electroforesis horizontal en geles de agarosa 2% (p/v) y fueron cargados mezclando 4 μ l de producto con 2 μ l de Blue Juice 1X Marca Invitrogen No. Cat. 10816-015 adicionalmente del marcador de

TrackIt™ 100bp DNA Ladder marca Invitrogen No. Cat. 10488-058. La solución buffer de electroforesis fue TAE 1X a voltaje constante de 100 voltios durante 35 minutos y posteriormente se realizó la tinción de los mismos en una solución de bromuro de etidio (15 ppm) por un periodo de tiempo de 15-20 minutos en agitación continua. Los geles se visualizaron en el fotodocumentador Dolphin View Wealtec.

Para validar las muestras de ADN obtenidas, se las sometieron a reacciones de amplificación PCR usando los marcadores microsatélite: EBmac0501, Bmac0209, Bmag0223, Bmac0040 y Bmag0007 con la metodología antes descrita.

2.4.6 Estandarización de combinaciones de marcadores para amplificación.

Para verificar que amplifiquen y que sus productos sean detectados por la plataforma de LI-COR se utilizó la metodología M13-*tailing* que consiste en añadir a uno de los primers la secuencia M13 de 19 pares de bases para marcar las moléculas por fluorescencia y detectarlas a través de dos canales de detección (en 700 y 800 nm), sistema adaptado al genotipador LI-COR 4300S (LI-COR Biosciences).

Para realizar la amplificación de los marcadores microsatélites de cebada se siguió la metodología descrita por Morillo & Miño (2011) realizándose reacciones de amplificación PCR con las siguientes concentraciones finales en un volumen de reacción de 5,0 µl: ADN (5,0 ng/µl): 2,0 ng/µl; Buffer GoTaq PROMEGA® (10X): 1X; MgCl₂ PROMEGA® (25,0 mM): 2,5 mM; dNTPs INVITROGEN® (5,0 mM): 0,20 mM; Primer Forward (1,0 µM): 0,01 µM; Primer Reverse (10 µM): 0,16 µM; M13 Forward IRDye® (1,0 µM): 0,16 µM; DNA Taq Polymerasa PROMEGA® (5,0 U/µl): 0,05 U/µl. El programa de amplificación fue el siguiente: un ciclo de 94 °C por 2 minutos; un ciclo de 95 °C por 4 minutos; 25 ciclos de 94 °C por 1 minuto, TA por 2 minutos y 72 °C por 2 minutos; un ciclo de 72 °C por 10 minutos y uno de 4 °C por 10 minutos.

Adicionalmente, se realizaron pruebas de amplificación multiplex, específicamente de *duplexaje*, que consistieron en amplificar dos marcadores SSR en una misma reacción PCR. Para seleccionar los primers de cada combinación se tomó en cuenta que éstos tengan una temperatura de alineamiento lo más similar posible,

que exista una diferencia de al menos 30 pb en los productos amplificados y que la marcación de fluorescencia M13 sea la misma.

2.4.7 Determinación del número de individuos.

Debido al incremento del número de individuos, con la consideración de incluir en el estudio 2 ó 10 por cada variedad o accesión, se realizó un *screening* previo de aquellos pertenecientes a las accesiones de cebada acriollada.

Se sometió a reacciones de amplificación PCR a las muestras de ADN de los individuos mencionados anteriormente mediante la técnica *M13-tailing* con los dos marcadores que reportaron tener los índices de polimorfismo PIC más altos que fueron EBmac0501 y Bmac0040. A continuación, se analizaron los perfiles de amplificación para determinar el mínimo número de individuos en el que consten todos los perfiles de cada accesión, y sean estos los escogidos para el genotipaje con la totalidad de marcadores.

2.4.8 Genotipaje de marcadores seleccionados SSR en LI-COR 4300S.

Los productos de PCR se diluyeron en una solución de parada Blue Stop Solution (LI-COR), en proporción 1:1, luego se denaturaron a 95°C por 5 minutos y se pasaron a hielo dejándose reposar por 5 minutos. La preparación de los geles de poliacrilamida al 6.5%, consistió en la adición de 20 ml de Gel Matrix KB Plus al 6.5% (827-05607, LI-COR®), 150 µl Persulfato de Amonio (APS) al 10% y 15 µl de TEMED, esta solución se la colocará entre dos placas de vidrio, previamente lavadas con el detergente e isopropanol, se colocó el peine y se dejó polimerizar por 1 hora. Posteriormente, se ensambló la placa de poliacrilamida en el equipo LI-COR® 4300 S, para la corrida electroforética se utilizará el tampón TBE 1x KB Plus LI-COR®.

Una vez que la placa polimerice y termine la pre-corrida se cargará 0.8 µl de cada muestra en cada uno de los pocillos del peine, en el orden preestablecido, además se cargará 0,5 µl de marcador de peso molecular IRDye® Sizing Standard 50-350 bp. Las corridas electroforéticas se llevaron por 1,5 horas a 1500 voltios. El análisis de la imagen del gel y genotipaje se realizó con el programa SAGA-GT, se

marcó cada banda que represente un alelo con una “X”, al confirmar el genotipaje se generó un reporte que muestra las tallas de los alelos de cada locus SSR en pares de bases y las frecuencias alélicas en porcentaje.

2.4.9 Análisis de datos.

El software SAGA GT-SSR es un asistente de lectura de las imágenes proporcionadas por el LI-COR®. La matriz de datos obtenidos de SAGA es importada a MICROSOFT EXCEL, donde se la depuró de acuerdo a los motifs de cada locus, se disminuyeron en 19 pb las tallas alélicas (debido a la técnica *M13-tailing*) y para los casos de no presentarse amplificación se asignó el valor -1. Esta matriz genotípica, de dimensiones $n \times m$ siendo n el número de genotipos y m el número de loci analizados, fue la base para los análisis en los distintos programas estadísticos.

2.4.9.1 Análisis de diversidad genética.

Para el análisis de diversidad genética se utilizaron los siguientes paquetes informáticos: PowerMarker V3.25 (Liu & Muse, 2005) y GenAlEx V6.5 (Peakall & Smouse, 2012) que es un complemento del programa Microsoft® Excel®, con los que se determinaron los siguientes parámetros:

Suponiendo sean dados n individuos, se usará el símbolo A para cada locus con una serie de alelos A_i . Un alelo A_i tendrá una frecuencia respecto a la población denotada como p_i .

Número de observaciones (Obs): es definido como el número de genotipos no perdidos (para datos diploides) observados en una muestra.

Disponibilidad: está definida como:

$$Disponibilidad = \frac{(1 - Obs)}{n}$$

Frecuencia alélica: la proporción de todos los alelos de ese gen en la población que corresponden específicamente a ese tipo.

$$p_i = \frac{n_i}{2n}$$

Número efectivo de alelos en un locus (Effective Number of Allele ENA) (Kimura & Crow, 1964): hace referencia a los alelos con capacidad de pasar a la siguiente generación.

$$ENA = \frac{1}{\sum_{i=1}^n p_i^2}$$

Heterocigosidad observada: es la proporción de individuos heterocigotos en la población, para un locus dado se estima como:

$$H_o = 1 - \sum_{i=1}^k P_{ii}$$

Heterocigosidad esperada (Nei, 1973): es definida como la probabilidad de que dos alelos escogidos al azar dentro de una población sean iguales.

$$H_e = 1 - \sum_{i=1}^k p_i^2$$

Contenido de información polimórfica (Polimorfism Index Content PIC) (Botstein, White, Skolnik, & Davis, 1980): mide cuan informativo es un marcador, que depende del número de alelos para ese locus y de sus frecuencias relativas.

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

Alelos únicos: Se caracterizan por estar presentes en una población que posee una característica proveniente de un lugar único (Alstad, Hotchkiss, & Corbin, 1991).

2.4.9.2 Análisis multivariantes

Los análisis multivariantes empleados fueron: Análisis de coordenadas principales (ACoP), análisis de conglomerados (AC) y biplot logístico externo (BLE).

Análisis de coordenadas principales

Es el método de ordenación más utilizado en estudios de diversidad genética, desarrollado por Gower (1966) construye una configuración de n puntos (Operative Taxonomic Units OTUs) en el espacio euclídeo, a partir de una matriz de similitud, de manera tal que la distancia entre dos puntos de la configuración aproxime tanto como sea posible la disimilitud entre las OTUs representadas por esos, respetando la estructura de similaridad definida por la matriz de similitud (véase anexo D.1) (Demey, 2008).

Con la ayuda del programa PowerMarker V3.25 (Liu & Muse, 2005) y a partir de la matriz genotípica ($n \times m$) se obtuvo la matriz de frecuencias o matriz X ($n \times q$) donde q es el número de alelos encontrados.

Esta matriz de frecuencias o binaria fueron editadas para depurar los datos de individuos heterocigotos y datos perdidos, y utilizadas en el programa NTSYS 2.02 (Rohlf, 2002) en el que se calcularon las matrices de similitud basada en el coeficiente de emparejamiento simple o Simple Matching Coefficient (Sokal & Michener, 1958):

$$SMC_{ij} = \frac{a + d}{a + b + c + d}$$

Donde, i y j son las OTUs a comparar, a es el número de caracteres presentes comunes, b es el número de caracteres presentes en i pero ausentes en j , c es el número de caracteres presentes en j pero ausentes en i y d número de caracteres ausentes simultáneamente.

A continuación, se calculó la matriz producto del “doble centrado” con la que se obtuvieron los vectores y valores propios o eigen que son graficados en el espacio.

En cuanto al programa MULTBILOT (Vicente-Villardón, 2014), la matriz de frecuencias fue importada y en base a esta se realizó el ACoP, con el coeficiente de emparejamiento simple.

Análisis de conglomerados

El análisis de conglomerados, AC (en inglés, *Cluster Analysis*) es una técnica multivariante que busca agrupar elementos y lograr la máxima homogeneidad en cada grupo y la mayor diferencia entre los grupos. Se pueden emplear dendrograma que son representaciones gráficas que ayudan a interpretar el resultado de los análisis *cluster*.

Los métodos de agrupamiento más usados son: (1) los métodos jerárquicos la clasificación resultante tiene un número creciente de grupos (Dillon & Goldstein, 1984) y (2) los no jerárquicos las clases no son anidadas y los grupos se forman a través de criterios predefinidos (López & Hidalgo, 1994).

Con la matriz de similitud generada y mediante el software NTSYS 2.02 se obtuvieron los arboles con el algoritmo UPGMA “Unweighted Pair Group Method with Arithmetic Mean” (Sneath & Sokal, 1973) que construye un árbol jerárquico utilizando encadenamiento promedio (véase anexo D.2).

Mientras que en el programa MULTBILOT se realizó el AC con el algoritmo UPGMA sobre la matriz de distancias inicial y sobre la resultante del ACoP. Los análisis arrojaron también sus índices de correlación cofenética.

2.4.9.3 Análisis de remuestreos

El método de remuestreo escogido fue el Bootstrap total tipo 3 (procrustes) sobre los residuales. En el programa MULTBILOT se modificó el número de muestras a 1000, y se mantuvieron los valores por defecto de tolerancia 0.0001 y 25 iteraciones. Además se seleccionó la opción para usar las coordenadas consenso (véase anexo D.3).

2.4.9.4 Biplot logístico externo

Los métodos biplot son representaciones gráficas simultáneas de filas (individuos) y columnas (variables) de una matriz de datos dada. El biplot puede ser ajustado para realizar regresiones alternantes e interpolaciones, no obstante, cuando se tratan de datos binarios, resultantes de marcadores moleculares, el análisis Biplot Lineal Clásico no es adecuado ya que su respuesta a través de las dimensiones no es lineal, más bien es logística. Es llamado externo porque las coordenadas de los genotipos son calculadas en un procedimiento externo que es el ACoP. Vicente-Villardón, Galindo, & Blázquez-Zaballos (2006) describieron la geometría del Biplot Logístico, en el que cada individuo está representado como un punto y cada variable como una dirección hacia el origen. La proyección de un individuo sobre la dirección de un carácter predice la probabilidad de presencia de ese carácter (véase anexo D.4).

El análisis de BLE se efectuó en el software MULTBILOT. Se generaron los gráficos de relaciones entre alelos y genotipos, los de proyecciones binarias y los de proyecciones de individuos sobre variables.

2.4.9.5 Análisis de estructura genética

Estadísticos F

Los estadísticos F (Wright, 1965) y revisados posteriormente por (Chakraborty & Danker-Hopfe, 1991) se relacionan de acuerdo a la siguiente ecuación:

$$(1 - F_{IT}) = (1 - F_{IS}) - (1 - F_{ST})$$

El estadístico F_{IS} mide el parecido entre los alelos de un individuo, es una medida del grado de endogamia en una población.

El parámetro F_{IT} mide la desviación de las proporciones esperadas de individuos heterocigotos respecto de las observadas del conjunto de la población.

Weir & Cockerham (1984) construyen formulaciones para los parámetros F. según este método, los parámetros F_{IS} y F_{IT} toman valores positivos cuando hay un déficit de heterocigotos y negativos en caso de exceso de los mismos.

El parámetro F_{ST} mide el parecido entre los individuos de una población y explica el porcentaje de diversidad genética que se debe a la existencia de una estructura en la población.

Para la obtención de estos análisis se empleó el programa GENALEX v6.5 (Peakall & Smouse, 2012).

Análisis molecular de varianza

El Análisis Molecular de Varianza, AMOVA (Excoffier, Smousse, & Quattro, 1992) es un análisis jerárquico que permite analizar la varianza entre grupos y dentro de los mismos, es decir, la medida de estructura genética de una población.

Este tipo de análisis fue una metodología transversal ya que se aplicó también en el análisis de conglomerados. Se utilizó para ello el software ARLEQUIN v3.5.1.3 (Excoffier & Lischer, 2010). Se realizaron análisis moleculares de varianza para verificar configuraciones de número de clusters a formarse en las que se obtenga la mayor variación entre clusters propuestos.

El análisis se basó en el modelo llamado *infinite alleles model* (IAM) considera que el cambio en el tamaño de los alelos puede ocurrir a saltos grandes o pequeños con la misma probabilidad. El esquema del análisis se muestra en la tabla 4.

Tabla 4: Esquema del análisis molecular de varianza AMOVA.

Fuente de variación	Grados de libertad*	Suma de cuadrados	Cuadrado medio esperado**	Porcentaje de varianza
Entre grupos	P - 1	Suma de cuadrados entre grupos	$n\sigma_a^2 + \sigma_b^2$	$\frac{n\sigma_a^2 + \sigma_b^2}{\sigma_T^2}$
Dentro de grupos	n - P	Suma de cuadrados dentro de grupos	σ_b^2	$\frac{\sigma_b^2}{\sigma_T^2}$
Total	n - 1		σ_T^2	

*P: número total de grupos analizados, n: número total de individuos.

$$** n = \frac{N - \sum_p \frac{N_p^2}{N}}{P - 1}$$

Donde los valores de los cuadrados medio esperados corresponden a los componentes de varianza que son utilizados para calcular el índice de fijación de Weir & Cockerham (1984) con la siguiente relación:

$$F_{ST} = \frac{\sigma_a^2}{\sigma_T^2}$$

2.4.9.6 Identificación de genotipos duplicados.

Se realizó la identificación de genotipos 100% similares con la ayuda del complemento de Microsoft® Excel® Microsatellite Toolkit (Park, 2001) con un número de alelos no idénticos tolerados de 2 antes de rechazar un emparejamiento.

2.4.9.7 Determinación de una colección núcleo de cebada.

Una colección núcleo es una submuestra de una gran colección de germoplasma que contiene el mínimo número de individuos que representan la diversidad genética total de la colección original. Existen varias estrategias que pueden adoptarse para extraer una colección núcleo. Los programas usados para este propósito fueron COREFINDER (Policriti & Sgarro, 2014) y POWERCORE (Kim *et al.*, 2007) que utilizan la estrategia M.

La estrategia M (maximización de alelos) (Shoen & Brown 1993; Gouesnard *et al.*, 2001; Escribano, Viruel, & Hormaza, 2008) consiste en detectar el mejor tamaño de muestra que capture el 100% de la diversidad presente dentro de una colección entera de germoplasma. El procedimiento es un algoritmo aleatorio estilo *Las Vegas*: un número de repeticiones es provisto por el usuario, y el algoritmo iniciando con un set aleatorio inicial, usa una estrategia codiciosa para buscar una accesión “A” que brinde una mejor diversidad genética total que alguna accesión “B” perteneciente a la actual colección núcleo. En dicha hipótesis, la accesión “A” es incluida y “B” excluida de la colección núcleo. Este paso es repetido exhaustivamente y cada iteración empieza con un set aleatorio diferente, reduciendo la probabilidad de terminar en un máximo local.

CAPÍTULO 3: RESULTADOS

3.1 Germoplasma analizado.

Como resultado de las prospecciones de semillas de cebada realizadas, en la primera de ellas, realizada en las provincias de Cañar, Azuay y Loja, se obtuvieron 24 materiales; mientras que en la segunda que tuvo lugar en la provincia de Imbabura se recolectaron 4 materiales. Las coordenadas de los lugares de colecta de las accesiones de los grupos A.1 y A.2 se ilustran en la figura 7.

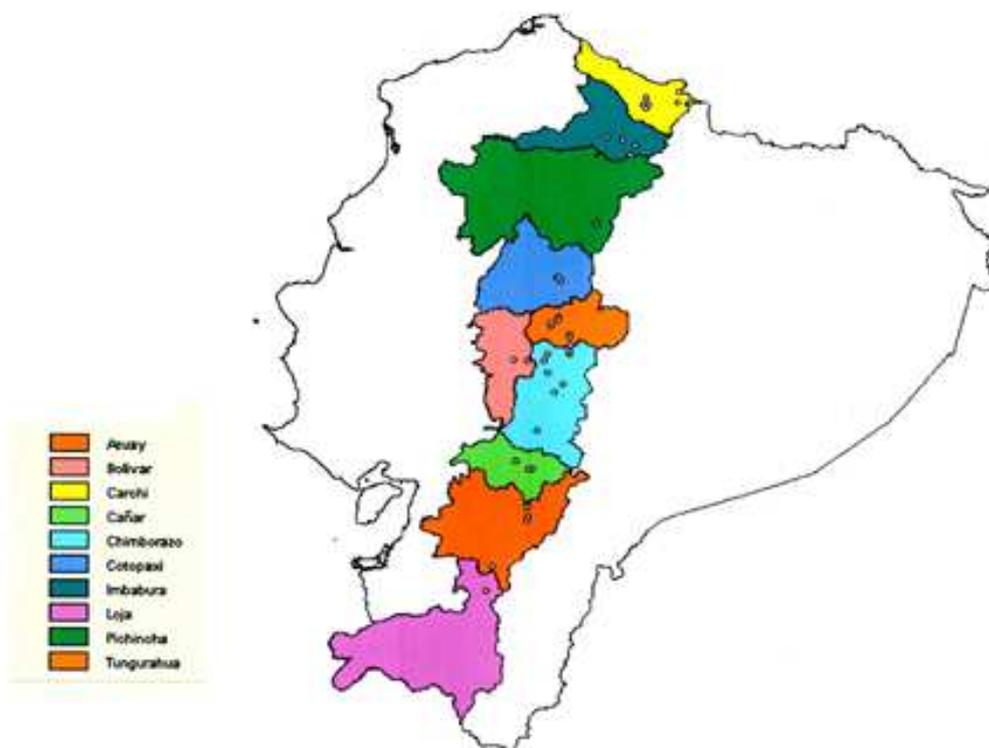


Figura 7: Localidades donde fueron colectadas las accesiones de cebada acriollada del banco de germoplasma y de las colectas realizadas en este estudio.

El germoplasma analizado quedó así compuesto por 160 materiales correspondientes a tres grupos de accesiones acriolladas y dos grupos materiales de mejoramiento, como se describe en la tabla 5.

Tabla 5: Descripción del germoplasma en estudio de acuerdo a su tipo, proveniencia y número de accesiones o variedades.

Grupo	Tipo de material	Proveniencia	No. de accesiones/ variedades
A.1	Acriollado	Banco de Germoplasma INIAP	46
A.2	Acriollado	Colectas Austro e Imbabura	28
A.3	Acriollado	Programa de Cereales	23
M.1	Mejorado	Programa de Cereales	51
M.2	Mejorado	Programa de Cereales	12
			160

3.2 Extracción, cuantificación y validación de ADN genómico de cebada

Se realizaron las extracciones de ADN genómico de un total de 912 individuos del germoplasma en estudio; considerando que, como se mencionó en el apartado 2.4.2., se tomó en cuenta la naturaleza de cada tipo de germoplasma para determinar el número de individuos que iban a considerarse de cada grupo para posteriores análisis (Tabla 6).

El protocolo desarrollado por Pallotta *et al.* (2003) para extraer ADN genómico de trigo y cebada permitió obtener un alto rendimiento que en promedio fue de $54.4 \pm 1.8 \mu\text{g}$ de ADN (Anexo E). Se obtuvieron mayores rendimientos en las muestras que se logró mejor nivel de maceración del tejido foliar junto con la solución buffer de extracción. El pellet de ADN obtenido después de la precipitación con isopropanol presentó coloración marrón, en la mayoría de los casos, pero al final de los lavados con etanol 70% y 95% este cambió a blanquecino.

Tabla 6: Detalle del número de individuos de cada grupo de germoplasma de cebada de los que se extrajeron muestras de ADN genómico.

Grupo	Individuos por accesión/ variedad	No. de accesiones/ variedades	Individuos totales
A.1	10	46	460
A.2	10	28	280
A.3	2	23	46
M.1	2	51	102
M.2	2	12	24
TOTAL		160	912

En cuanto a la calidad del ADN extraído, no se utilizó tratamiento enzimático con RNasa para eliminar contaminación ARN. Los coeficientes de la razón entre los valores de absorbancias OD_{260nm}/OD_{280nm} tuvieron en promedio un valor de 2.10 ± 0.02 (Anexo E), siendo una razón óptima de pureza de ADN aquella que se encuentre entre 1.7 y 1.9; mientras que valores superiores a este rango denotan contaminación por sales y los inferiores interferencia de proteínas, fenol y otros compuesto aromáticos (Touil *et al.*, 2008). También se comprobó la calidad del ADN con electroforesis horizontal en geles de agarosa 1 % (p/v). En la figura 8 se observan muestras de ADN genómico de cebada que presentan bandas bien definidas, con poca degradación y con presencia de contaminación por ARN.

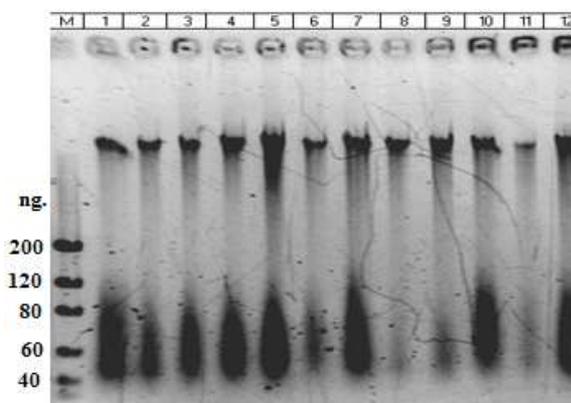


Figura 8: Cuantificación en gel de agarosa 1% (p/v) de muestras de ADN genómico de cebada (Carril 1 al 12: muestras, carril M: marcador de peso molecular).

Se verificaron los productos PCR amplificados de las muestras de ADN que fueron diluidas respectivamente en nuevas alícuotas de concentración 5,0 ng/μl en todas las 912 muestras. No se detectaron problemas en la amplificación de las muestras de ADN, los amplicones presentaron buena intensidad además se confirmaron sus tamaños de acuerdo a lo reportado en referencias, como por ejemplo con el marcador Bmac0209 que se ilustra en la figura 9 con amplicones alrededor de 176 pb.

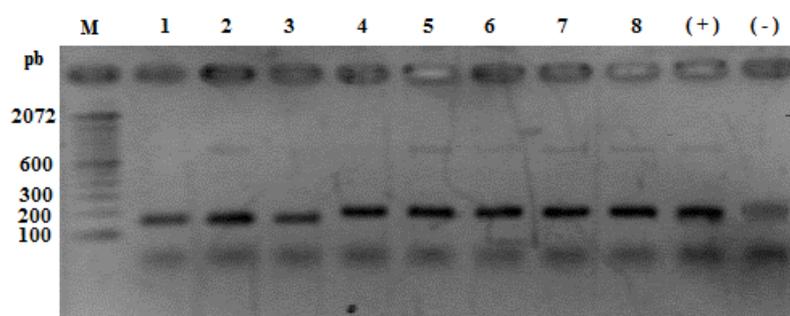


Figura 9: Amplificación de 8 muestras de ADN de genómico de cebada con el marcador Bmac0209. (Carriles 1-8: muestras de ADN, carril M: marcador de peso molecular, carril (+): control positivo, carril (-): control negativo).

3.3 Validación y estandarización de marcadores y combinaciones para genotipaje en plataforma LI-COR® 4300S

En el proceso de validación de los marcadores microsatélites preseleccionados se constataron las condiciones de amplificación y se reportó la modificación de la temperatura de alineamiento para el marcador Bmag0384 y Bmac0127 a 60 y 58 °C respectivamente. Adicionalmente, en el mix PCR para la amplificación de estos marcadores se añadió el coadyuvante de reacción BSA (Bovine Serum Albumina) en concentración final de 0.1%.

Las secuencias genómicas de los productos PCR de 5 marcadores usados en este estudio se detallan en el anexo F. Se verificaron sus secuencias iniciadoras, tamaños aproximados y secuencias motivo.

Las muestras amplificadas con el marcador Bmac0399 no presentaron amplificación cuando se utilizó el programa con la temperatura de alineamiento reportado de 58 °C (Figura 10 a y b), por lo que se realizaron pruebas con gradientes de temperatura sin lograr el resultado esperado, por lo que este marcador fue descartado.

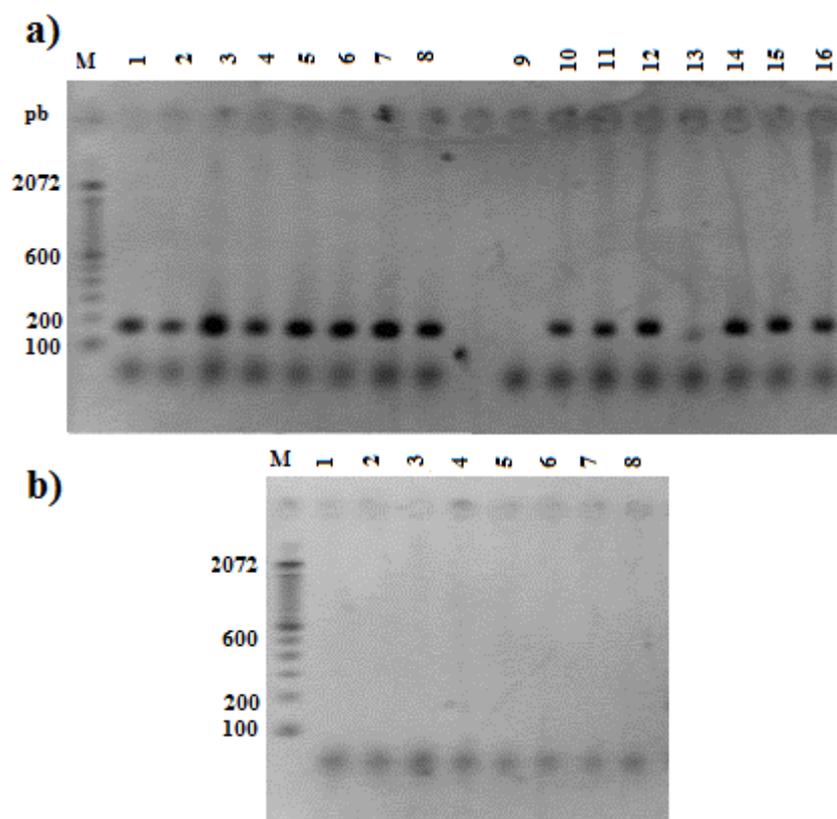


Figura 10: Amplificación de 8 muestras de ADN de genómico de cebada para tres marcadores microsatélites. a) Carriles 1-8: Bmac0213, carriles 9-16: EBmac0501. b) carriles 1-8: Bmac0399. Carriles M: marcador de peso molecular.

El marcador Bmag0692 presentó amplicones de gran tamaño (superiores a 500 pb) que no lo convirtió en un buen candidato para el genotipaje en la plataforma LICOR® 4300 S en la que se pueden visualizar fragmentos de hasta 350 pb (Figura 11.a). Por otro lado, los marcadores Bmag0853, EBmac0701 y Bmag0135 presentaron alelos nulos en sus perfiles de amplificación para un porcentaje considerable (superior al 30%) de muestras en estudio (Tabla 7); por lo que fueron descartados del genotipaje definitivo (Figura 11.b).

Se determinó la presencia de “alelos nulos” al someter a reacciones de amplificación PCR a muestras de ADN, que incluyen además, controles positivos para los marcadores mencionados y comprobar la no amplificación de las muestras.

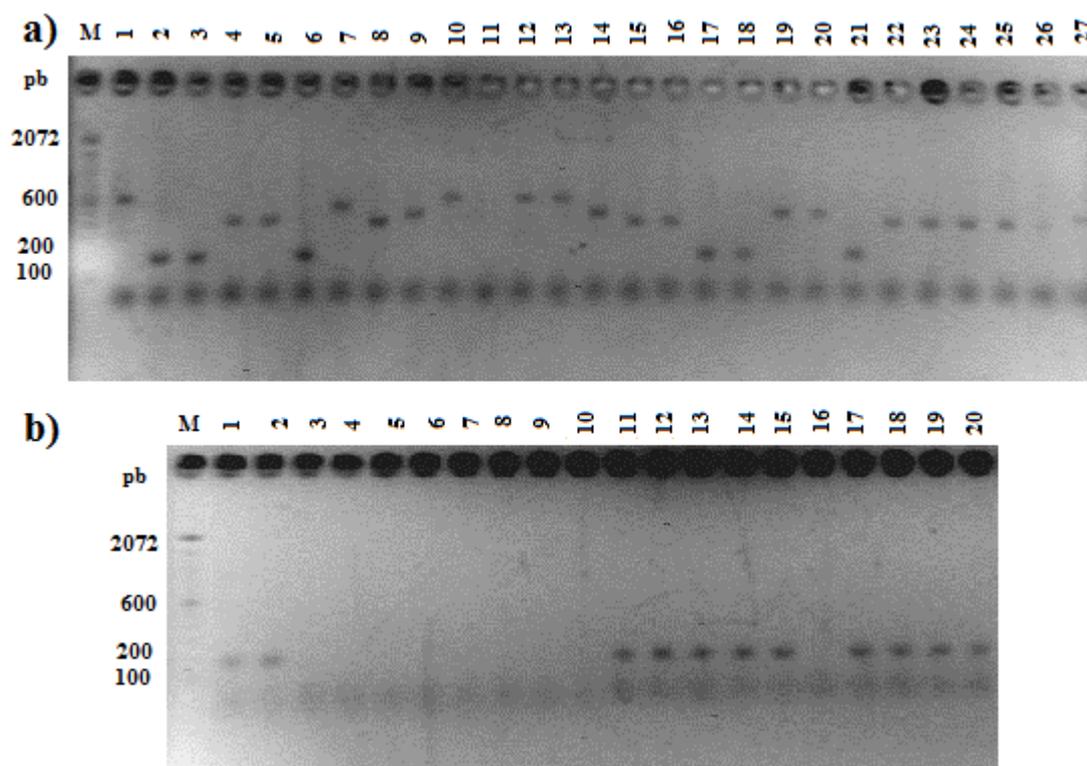


Figura 11: Amplificación de muestras de ADN de genómico de cebada para dos marcadores microsatélites. a): Carriles 1-27: Bmag0692. b) carriles 1-10: Bmag0135; 11-20: EBmac0501. Carriles M: marcador de peso molecular.

Tabla 7: Descripción del número y proporción de alelos nulos para los marcadores Bmag0135, Bmag0853 y EBmac0701.

Material	No. de genotipos	Marcador					
		Bmag0135		Bmag0853		EBmac0701	
		GAN*	%	GAN*	%	GAN*	%
Acriollado	156	68	43.6	70	44.9	78	50.0
Mejorado	63	10	15.9	9	14.3	22	34.9
TOTAL	219	78	35.6	79	36.1	100	45.7

*GAN=Número genotipos que presentan alelos nulos

Los 16 marcadores seleccionados que fueron validados se sometieron a reacciones PCR para ser analizados en el genotipador LI-COR® 4300S, en reacciones simplex en primera instancia para después probarlos combinaciones

multiplex (Figura 12), obteniéndose 2 combinaciones de amplificación en *dúplex* y el resto de marcadores en reacciones *monoplex* como se describe en la tabla 8.

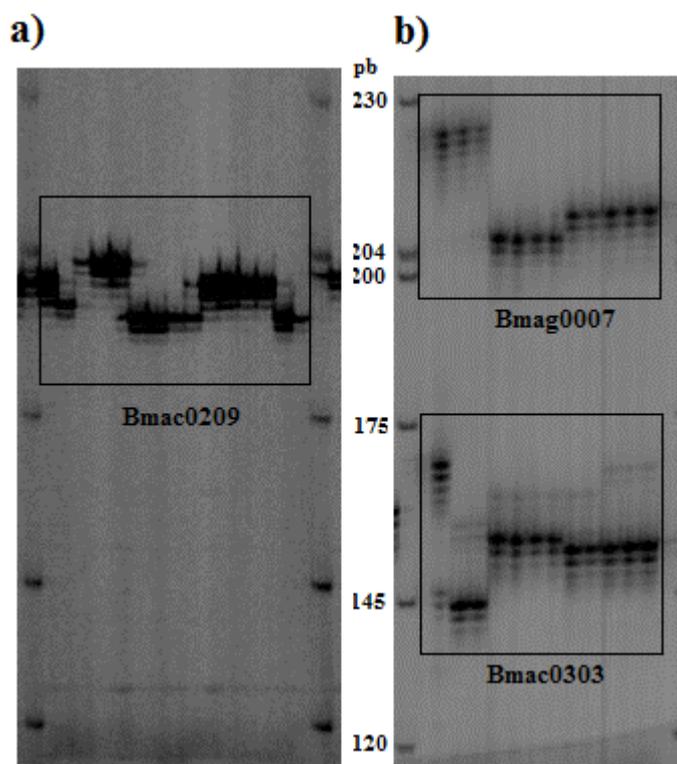


Figura 12: Perfiles de amplificación de primers SSR detectados por la plataforma LI-COR 4300S. a) con el marcador Bmac0209. b) Combinación *dúplex* de los marcadores Bmag0007 y Bmac0303.

Tabla 8: Detalle de las condiciones de temperatura y marcaje M13 para las combinaciones y marcadores en estudio.

Nº	Marcador - combinación	Marcaje	Ta (°C)
1	Bmac0303 – Bmag0007	700	58
2	Bmac0040 – Bmac0127	800	58
3	Bmac0213	800	58
4	EBmac0501	700	58
5	Bmac0209	800	58
6	Bmag0223	700	58
7	Bmag0217	700	58
8	Bmag0384	800	60
9	Bmac0134	700	55
10	Bmag0125	800	55
11	Bmag603	800	55
12	EBmac0679	700	55
13	Bmac0316	700	55
14	scssr02306	700	60

3.4 Determinación del número de individuos

Se confirmó la heterogeneidad genética existente en las accesiones de cebada acriollada (grupos A.1 y A.2), al observar diferentes perfiles de amplificación entre los individuos de la una accesión dada analizados con los marcadores EBmac0501 y Bmac0040 (Figura 13).

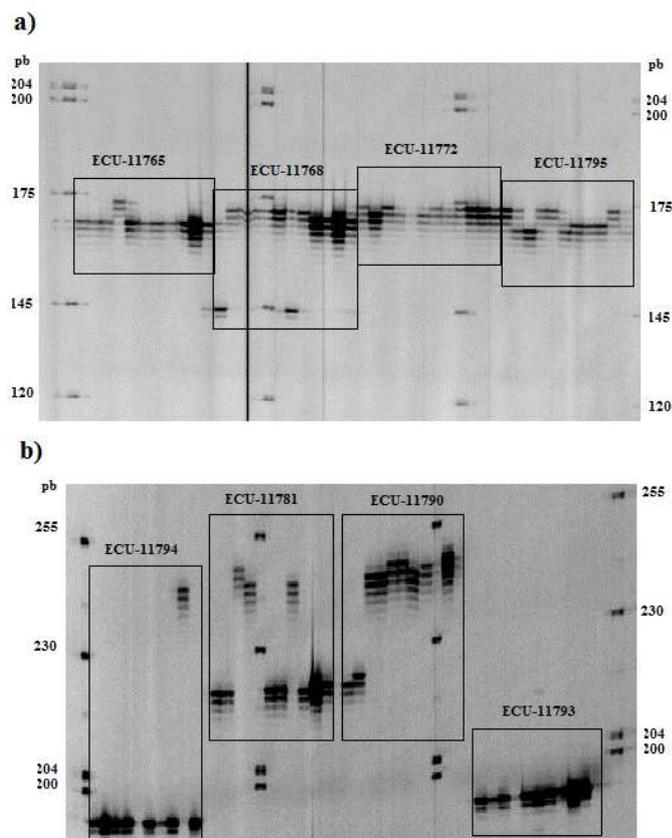


Figura 13: Perfiles de amplificación para accesiones de cebada acriollada del Banco de Germoplasma. a) Con el marcador EBmac0501. b) Con el marcador Bmac0040.

En total, se determinó que existieron 29 accesiones con patrones de amplificación monomórficos a nivel de los dos marcadores mencionados, representado el 39,2% del total de accesiones analizadas. El número de accesiones que presentaron dos genotipos por cada accesión fue de 35, siete accesiones presentaron tres genotipos, dos accesiones cuatro genotipos y tan solo una accesión cinco genotipos (Tabla 9).

Tabla 9: Resumen del número de individuos y genotipos total y por cada grupo de germoplasma analizados.

Nº de individuos seleccionados por accesión	Nº de accesiones representadas	Nº total de individuos seleccionados
1	29	29
2	35	70
3	7	21
4	2	8
5	1	5
	74	133

En definitiva, de las 74 accesiones analizadas, el número de individuos escogidos para el análisis final, asciende a 133; que sumados al resto de materiales de los que se tomó en cuenta dos individuos por material resultan en 305 individuos correspondientes a 219 distintos genotipos (Tabla 10).

Tabla 10: Resumen del número de individuos y genotipos total y por cada grupo de germoplasma analizados.

Grupo	Número de accesiones/ variedades	Individuos seleccionados	Nº Genotipos
A.1	46	92	92
A.2	28	41	41
A.3	23	46	23
M.1	51	102	51
M.2	12	24	12
TOTAL	160	305	219

3.5 Genotipaje de marcadores SSR seleccionados en la plataforma LI-COR® 4300S

Las muestras de ADN de los 305 individuos fueron amplificadas para los 16 marcadores SSR y combinaciones descritos en la tabla 3.3 Se generaron los geles digitales para su edición y genotipaje en el software SAGA™ como se observa en las figuras 14 y 15. Para los marcadores en los que no se alcanzó el porcentaje mínimo de datos perdidos de 10% se repitieron ampliaciones, de la misma manera que

para las muestras hasta que estas alcancen un mínimo de 30% de datos perdidos con todos los marcadores.

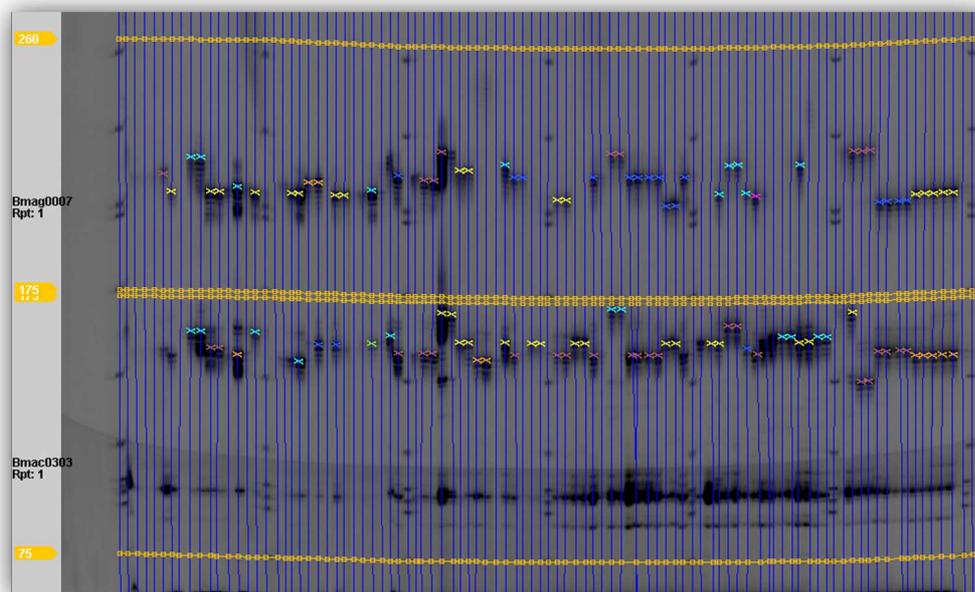


Figura 14: Gel digital de una placa de 88 muestras de ADN de cebada amplificado con la combinación dúplex de los marcadores Bmag0007 y Bmac0303.

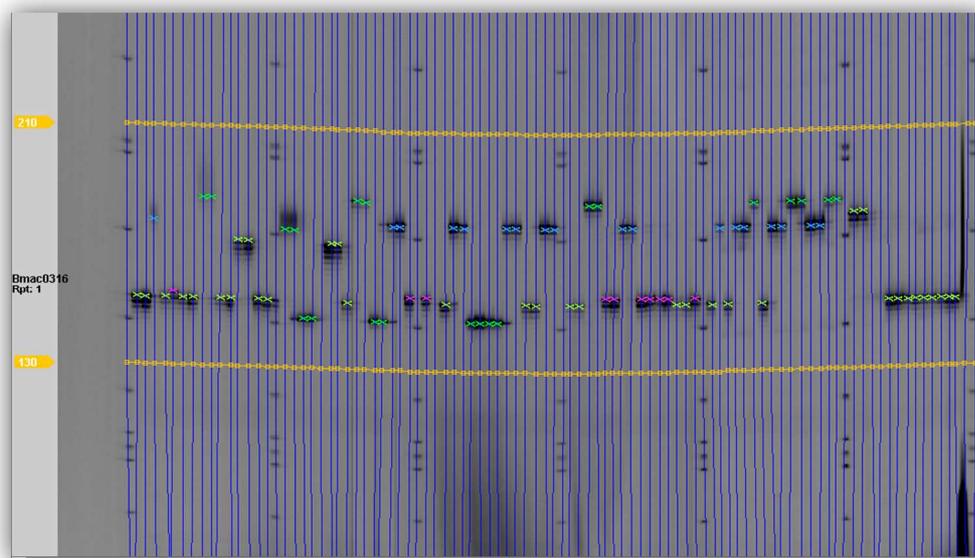


Figura 15: Gel digital de una placa de 88 muestras de ADN de cebada amplificado con el marcador Bmac0314.

3.6 Análisis de datos.

3.6.1 Diversidad genética

Se consideró como colección completa a la integrada por los 219 genotipos de cebada de los distintos tipos de germoplasma descritos en la tabla 10. Se realizaron análisis de diversidad genética en la base de esta colección, así como también del germoplasma y grupos de materiales acriollado (tipos A.1, A.2 y A.3) y mejorado (M.1 y M.2) y análisis comparativos entre estos tipos de germoplasma.

3.6.1.1 Diversidad genética de la colección completa.

La matriz genotípica de la colección completa, con los datos de los 219 genotipos para los 16 marcadores SSR se muestra en el anexo G. Se detectaron 157 alelos con un rango de longitud en pares de bases que fueron desde 104 hasta 232. Las frecuencias alélicas estuvieron en el rango de 0.002 para el alelo de 129bp del locus EBmac0501 hasta 0.68 para el alelo de 186bp del locus Bmag0217. La lista completa de frecuencias alélicas para cada uno de los loci se detallan en el anexo H.1.

El número de alelos por locus varió desde 3 (scssr2306) a 14 (Bmag0007 y Bmac0316) con una media de 9.81 alelos por locus. En cuanto al número de alelos raros, fueron los marcadores Bmac0213 y Bmac0316 los de mayor proporción. La disponibilidad de los datos se promedió en 95% (Tabla 11).

Adicionalmente, el valor promedio para la heterocigosidad observada es 0.027 ya que 10 marcadores no detectaron individuos heterocigotos. El índice de diversidad genética de Nei o heterocigosidad esperada tuvo su valor máximo de 0.85 en el marcador Bmag0007 y el mínimo de 0.51 en Bmag0217 con un promedio de 0.75. El valor del índice de contenido de polimorfismo (PIC) también fue el más alto para el marcador Bmag0007 de 0.83, mientras que el promedio fue de 0.715 (Tabla 11).

Tabla 11: Resumen de los parámetros de diversidad genética de 219 genotipos de cebada de la colección completa analizados con 16 loci SSR.

Marcador	Disp.*	Na*	Nea*	Nar*	Ho*	He*	PIC*
Bmac0213	0.94	13	2.87	9	0.024	0.65	0.63
Ebmac0501	0.98	10	4.05	5	0.005	0.75	0.72
Bmac0209	0.99	10	5.12	5	0.000	0.81	0.78
Bmag0223	0.93	9	4.90	2	0.000	0.80	0.77
Bmac0303	0.94	10	4.77	5	0.000	0.79	0.76
Bmac0040	0.95	13	5.55	8	0.005	0.82	0.80
Bmag0007	0.91	14	6.59	8	0.000	0.85	0.83
Bmag0217	0.94	7	2.03	3	0.117	0.51	0.48
Bmag0384	0.95	5	3.47	1	0.000	0.71	0.66
Bmac0127	0.95	10	5.87	4	0.000	0.83	0.81
Bmac0134	0.94	12	4.37	7	0.000	0.77	0.75
Bmag0125	0.93	7	3.73	3	0.260	0.73	0.69
Bmag0603	0.96	12	3.71	8	0.019	0.73	0.70
EBmac0679	0.96	8	3.61	3	0.000	0.72	0.68
Bmac0316	0.98	14	5.88	9	0.000	0.83	0.81
scssr2306	0.97	3	2.47	0	0.000	0.60	0.53
Promedio	0.95	9.81	2.87	5.06	0.027	0.75	0.71

*Disp= disponibilidad, Na= número de alelos, Nea= número efectivo de alelos, Nar= número de alelos raros ($p_i \leq 0.05$), Ho= heterocigosidad observada, He= heterocigosidad esperada o diversidad genética de Nei (1973), PIC= *Polimorfism Index Content* o Contenido de información polimórfica (véase apartado 2.4.9.1).

3.6.1.2 Diversidad genética en el germoplasma de cebada acriollado

De la colección completa en estudio, 156 genotipos corresponden a accesiones acriolladas de cebada en donde se encontró en promedio 8.75 alelos por locus. Este grupo presenta una heterocigosidad observada de 0.040, que representó un valor más alto que el encontrado en la población total; debido a que solamente en este tipo de germoplasma fue posible detectar genotipos heterocigotos. Mientras que la heterocigosidad esperada y el índice PIC tuvieron un valor medio de 0.73 y 0.69, respectivamente. Al igual que para la colección completa, el marcador con los índices Ho y PIC más alto fue Bmag0007 y el de menor el marcador Bmag0217 (Tabla 12).

Se encontraron en promedio 1.88 alelos únicos por locus, siendo los marcadores Bmac0213 y Bmag0603 en los que se reportaron los mayores valores, como también los valores más altos de alelos raros por locus; mientras que en cinco marcadores (Bmac0303, Bmag0217, Bmag0384, EBmac0679 y scssr2306) no fue posible encontrar alelos exclusivos o únicos (Tabla 12).

Tabla 12: Resumen de los parámetros de diversidad genética de 156 genotipos de cebada acriollados analizados con 16 loci SSR.

Marcador	Disp.*	Na*	Nea*	Nar*	Nau*	Ho*	He*	PIC*
Bmac0213	0.93	12	2.40	9	5	0.03	0.58	0.56
Ebmac0501	0.97	8	3.82	3	2	0.01	0.74	0.70
Bmac0209	0.98	10	5.17	5	3	0.00	0.81	0.78
Bmag0223	0.90	9	4.78	4	2	0.00	0.79	0.77
Bmac0303	0.92	8	4.45	4	0	0.00	0.78	0.75
Bmac0040	0.94	12	4.87	7	3	0.01	0.79	0.77
Bmag0007	0.90	11	6.17	5	2	0.00	0.84	0.82
Bmag0217	0.91	5	1.67	2	0	0.17	0.40	0.38
Bmag0384	0.92	5	3.39	1	0	0.00	0.70	0.65
Bmac0127	0.94	10	5.87	4	2	0.00	0.83	0.81
Bmac0134	0.93	10	4.72	5	2	0.00	0.79	0.77
Bmag0125	0.90	7	3.80	3	2	0.38	0.74	0.69
Bmag0603	0.94	12	3.62	9	5	0.03	0.72	0.69
EBmac0679	0.96	7	3.35	2	0	0.00	0.70	0.66
Bmac0316	0.97	11	5.62	5	2	0.00	0.82	0.80
scssr2306	0.96	3	2.39	0	0	0.00	0.58	0.52
PROMEDIO	0.94	8.75	4.13	4.25	1.88	0.04	0.73	0.69

*Disp= disponibilidad, Na= número de alelos, Nea= número efectivo de alelos, Nar= número de alelos raros ($p_i \leq 0.05$), Nau= número de alelos únicos, Ho= heterocigosidad observada, He= heterocigosidad esperada o diversidad genética de Nei (1973), PIC= *Polimorfism Index Content* o Contenido de información polimórfica (véase apartado 2.4.9.1).

Diversidad genética dentro de los grupos A.1, A.2 y A.3

Los grupos de mayor diversidad, de acuerdo al índice PIC y a la heterocigosidad esperada fueron A.2 y A.1 respectivamente. No obstante el grupo A.1 presenta la mayor cantidad de alelos únicos y alelos raros por locus entre los tres grupos (Tabla 13). En comparación con los índices promedio de la colección completa, en estos tres

grupos se encontraron menos alelos por locus, al igual que los índices PIC y Ho que fueron inferiores en los tres casos. Los índices de diversidad para cada locus SSR de los tres grupos descritos en la tabla 13 se describen en el anexo H.2.

Tabla 13: Promedios de los parámetros de diversidad genética de los genotipos de tres grupos de germoplasma de cebada acriollados analizados con 16 loci SSR.

	Disp.*	Na*	Nea*	Nar*	Nau*	Ho*	He*	PIC*
A.1	0.94	7.63	3.93	3.13	0.56	0.06	0.70	0.66
A.2	0.91	6.88	3.83	1.88	0.44	0.01	0.71	0.67
A.3	0.97	4.94	2.72	1.63	0.06	0.00	0.60	0.55

*Disp= disponibilidad, Na= número de alelos, Nea= número efectivo de alelos, Nar= número de alelos raros ($p_i \leq 0.05$), Nau= número de alelos únicos, Ho= heterocigosidad observada, He= heterocigosidad esperada o diversidad genética de Nei (1973), PIC= *Polimorfism Index Content* o Contenido de información polimórfica (véase apartado 2.4.9.1).

3.6.1.3 Diversidad genética en el germoplasma de cebada mejorado.

Dentro de la colección completa, 63 genotipos corresponden a variedades mejoradas de cebada en donde se encontró en promedio 7.94 alelos por locus. Este grupo presenta una heterocigosidad observada nula. El índice de contenido polimórfico (PIC) promedio y el de diversidad genética de los marcadores en estudio para esta tipo de germoplasma es de 0.72 y 0.76 respectivamente.

En este grupo de germoplasma se encontraron 1.06 alelos únicos y 3.13 alelos raros por locus. Los marcadores Bmac0134 y Bmac0316 fueron capaces de detectar el mayor número de alelos únicos o exclusivos, mientras que siete marcadores no detectaron ninguno (Bmac0209, Bmag0223, Bmag0384, Bmac0217, Bmag0125, Bmag0603 y scssr2306) (Tabla 14).

Tabla 14: Resumen de los parámetros de diversidad genética de 63 genotipos de cebada de mejoramiento analizados con 16 loci SSR.

Marcador	Disp.*	Na*	Nea*	Nar*	Nau*	Ho*	He*	PIC*
Bmac0213	0.95	8	3.98	4	1	0.00	0.75	0.72
Ebmac0501	1.00	8	4.28	3	2	0.00	0.77	0.73
Bmac0209	1.00	7	4.92	2	0	0.00	0.80	0.77
Bmag0223	0.98	7	4.76	1	0	0.00	0.79	0.76
Bmac0303	1.00	10	4.45	5	2	0.00	0.78	0.75
Bmac0040	1.00	10	7.02	2	1	0.00	0.86	0.84
Bmag0007	0.92	12	7.13	5	3	0.00	0.86	0.84
Bmag0217	1.00	7	3.04	4	2	0.00	0.67	0.63
Bmag0384	1.00	5	3.44	2	0	0.00	0.71	0.66
Bmac0127	1.00	8	5.47	2	0	0.00	0.82	0.79
Bmac0134	0.97	10	3.39	6	2	0.00	0.70	0.68
Bmag0125	1.00	5	2.99	1	0	0.00	0.67	0.62
Bmag0603	1.00	7	3.80	3	0	0.00	0.74	0.70
EBmac0679	0.98	8	3.93	4	1	0.00	0.75	0.71
Bmac0316	1.00	12	6.15	6	3	0.00	0.84	0.82
scsr2306	1.00	3	2.63	0	0	0.00	0.62	0.55
PROMEDIO	0.99	7.94	4.46	3.13	1.06	0.00	0.76	0.72

*Disp= disponibilidad, Na= número de alelos, Nea= número efectivo de alelos, Nar= número de alelos raros ($p_i \leq 0.05$), Nau= número de alelos únicos, Ho= heterocigosidad observada, He= heterocigosidad esperada o diversidad genética de Nei (1973), PIC= *Polimorfism Index Content* o Contenido de información polimórfica (véase apartado 2.4.9.1).

Diversidad genética dentro de los grupos M.1 y M.2

El tipo de germoplasma M.1 presenta una considerable mayor diversidad genética que M.2; hecho que puede explicarse, en parte, a que el primero supera en número de individuos por más de cuatro veces al segundo. Todos los índices calculados fueron superiores en M.1, y el más llamativo fue el número de alelos raros por locus que tuvo un valor nulo en M.2 (Tabla 15). Los índices de diversidad para cada locus SSR de los tres grupos se describen en el anexo H.3.

Tabla 15: Promedios de los parámetros de diversidad genética de los genotipos de dos grupos de germoplasma de cebada mejorado analizados con 16 loci SSR.

	Disp.*	Na*	Nea*	Nar*	Nau*	Ho*	He*	PIC*
M.1	0.99	7.50	4.38	2.19	0.56	0.00	0.75	0.72
M.2	0.98	4.19	3.04	0.00	0.31	0.00	0.61	0.56

*Disp= disponibilidad, Na= número de alelos, Nea= número efectivo de alelos, Nar= número de alelos raros ($p \leq 0.05$), Nau= número de alelos únicos, Ho= heterocigosidad observada, He= heterocigosidad esperada o diversidad genética de Nei (1973), PIC= *Polimorfism Index Content* o Contenido de información polimórfica (véase apartado 2.4.9.1).

3.6.1.4 Diversidad genética comparativa entre el germoplasma acriollado y de mejoramiento.

Pese a que los parámetros: número de alelos, alelos raros, alelos únicos y número efectivo de alelos por locus en el germoplasma acriollado son mayores; los índices de diversidad genética y de polimorfismo PIC promedio fueron superiores en el germoplasma mejorado (Figura 16).

De entre los cinco grupos de germoplasma establecidos *a priori*, el más diverso fue el grupo M.1; sin embargo, el grupo A.1 presentó los valores más altos para alelos raros y alelos únicos por locus.

Por otra parte, los valores más altos de frecuencias alélicas en los grupos de germoplasmas en cuestión, se reportaron en el alelo de 186bp del locus Bmag0217 y fueron de 0.76 y 0.50 para los tipos acriollado y mejorado respectivamente. Todos los valores de frecuencias alélicas de cada locus para este germoplasma se detallan en el anexo H.4.

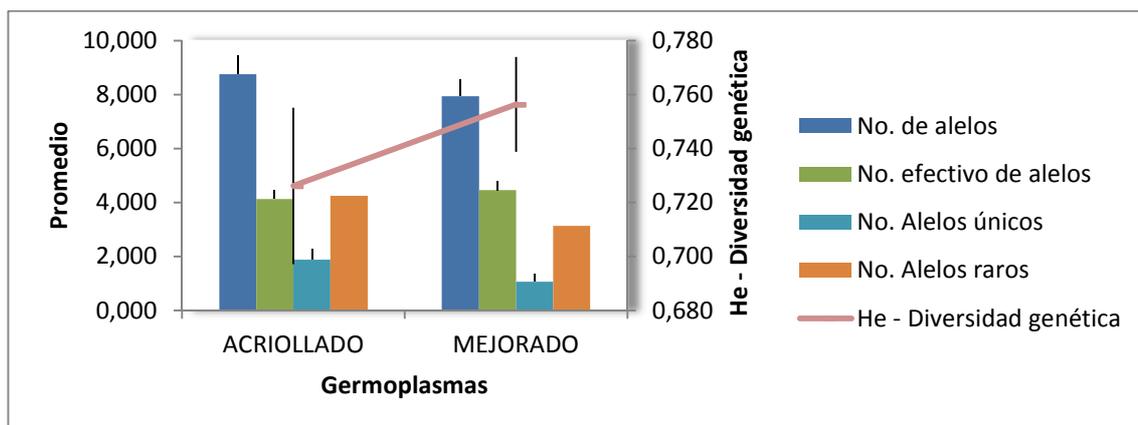


Figura 16: Comparación de los índices de diversidad genética entre el germoplasma acriollado y el mejorado.

3.6.2 Análisis de genotipos duplicados

Un total de 66 genotipos fueron emparejados como duplicados en 18 grupos que contuvieron desde 2 hasta 10 genotipos con 100% de similitud. La lista de estos genotipos se detalla en la Tabla 16.

Diez de los grupos de genotipos duplicados están conformados por individuos pertenecientes a un solo tipo de material de cebada. No se encontraron genotipos duplicados dentro o entre las líneas promisorias del PNC. Para los materiales acriollados se verificó a las accesiones para las que todos sus individuos hayan sido emparejados antes de identificarlas como accesiones duplicadas.

De las accesiones de cebada del Banco de Germoplasma se puede distinguir que todos los individuos de las accesiones CDF49 (ECU-11814) y CDF50 (ECU-11815) fueron identificados como duplicados entre sí; al igual que los de las accesiones CDF13 (ECU-11777), CDF21 (ECU-ECU-11785) y CDF22 (ECU-11786). Los datos pasaporte de las tres últimas accesiones demuestran que fueron recolectadas en la provincia de Chimborazo, en distintas localidades y con distintos nombres comunes.

De las variedades de cebada acriollada de la Feria de Semillas Alausí 2012, los pares de variedades CCU16, CCU23 y CCU1, CCU2 fueron identificadas como genotipos duplicados en grupos (4 y 5) que las contienen exclusivamente. Mientras,

que las variedades CCU4, CCU7 y CCU19, también idénticas entre sí pero además con otras variedades en el grupo 11.

En cuanto a las variedades de cebada acriollada de las colectas realizadas en el Austro e Imbabura, se destacan las CCA12, CCA13, CCA16 y CCA21 en un grupo de genotipos idénticos (grupo 13), en adición, fueron emparejadas las cuatro variedades colectadas en Imbabura (CCI1, 2, 3 y 4) con las CCA1 y CCA2. El par de accesiones CCA23 y CCA24 están emparejadas en un grupo exclusivo.

Las variedades de cebada pertenecientes a la Colección Nacional del PNC que fueron emparejadas como genotipos en dos grupos exclusivos fueron CCN44 y CCN45 (grupo 9) y CCN13, CCN25, CCN26 y CCN46 (grupo 10).

En términos de variedades duplicadas y al observar los grupos 11, 12, 13, 14, 16 y 17, se puede inferir en el origen de las variedades colectadas emparejadas respectivamente. De las variedades comerciales de cebada liberadas por el INIAP, se destaca la variedad CCN10 o INIAP-Cañicapa 2003 que correspondería a la misma que se detectó en las variedades colectadas en Imbabura y las variedades CCA1 y CCA2 del Austro.

Tabla 16: Resumen de la identificación de genotipos duplicados

	GENOTIPOS				No.
	A.1	A.2	A.3	M.1	
1	CDF6.4 CDF34a CDF17.1	--	--	--	3
2	CDF1.5 CDF46.2 CDF48.4	--	--	--	3
3	CDF14.5 CDF17.7	--	--	--	2
4	--	CCU16 CCU23	--	--	2
5	--	CCU1 CCU2	--	--	2
6	--	--	CCA8.2 CCA8.4	--	2
7	--	--	CCA5.1 CCA5.2	--	2
8	--	--	CCA23.3 CCA24.6	--	2

Continúa

GENOTIPOS					No.
9	--	--	--	CCN44 CCN45	2
10	--	--	--	CCN13 CCN25 CCN26 CCN46	4
11	CDF50.3 CDF50.7 CDF49.6	CCU4 CCU7 CCU19	CCA18.1	CCN14	8
12	--	--	CCA11.2	CCN1 CCN41 CCN42	4
13	--	--	CCA12.2 CCA12.8 CCA13.8 CCA163 CCA164 CCA21.3 CCA21.8 CCA1.4 CCA2.1	CCN3 CCN31	9
14	--	--	CC11.1 CC12.6 CC13.8 CC14.2	CCN10	7
15	CDF13.4 CDF21.5 CDF22.7 CDF23.7 CDF24.5 CDF26.3 CDF26.4 CDF30.9	CCU3 CCU6	--	--	10
16	--	CCU9	--	CCN15	2
17	--	--	CCA22.1	CCN12	2
TOTAL					66

3.6.3 Análisis de coordenadas principales

A partir de la matriz binaria X (véase apartado 2.9.4.2) y mediante el coeficiente de emparejamiento simple se realizó este análisis, en primera instancia en el programa MULTBILOT, para ilustrar los ejes de variabilidad de los genotipos y su ordenación en el espacio. Las primeras tres dimensiones del ACoP (análisis de coordenadas principales) explican el 32.63% de la variación total, siendo las coordenadas 1 y 2 las de mayor porcentaje de varianza con el 15.02% y 10.03% respectivamente (Tabla 17).

Tabla 17: Valores Eigen, porcentaje absoluto y acumulativo de la varianza por coordenada principal del ACoP de 219 genotipos de cebada con 16 loci SSR

Coordenada	Valores Eigen	Porcentaje de residuo	Varianza acumulada
1	2.44	15.02	15.02
2	1.63	10.03	25.04
3	1.23	7.58	32.63

En la distribución espacial generada en el ACoP (Figura 17) no se pudieron distinguir grupos definidos por los tipos de material establecidos *a priori*: mejorados y acriollados (Figura 18). Existe una disposición de las OTUs para formar “nubes” o conglomerados distinguibles a través de los ejes de varianza, independientemente del tipo de material que se trate excepto las líneas promisorias que se distribuyen principalmente alrededor del origen de los ejes de varianza y el segundo cuadrante. Los tipos de materiales más numerosos A.1 y M.1 presentan una mejor distribución a lo largo del plano de variación, aunque el segundo tipo con una menor cantidad de aglutinamiento (Figura 19).

La bondad de ajuste de la representación de los individuos del ACoP resultó en un promedio general para los 219 genotipos del $97.74 \pm 4.92\%$, esto es la medida en que el gráfico de la ordenación refleja con precisión la matriz de distancias.

Además, se obtuvo el gráfico que representa los genotipos del grupo A.1 de acuerdo a la provincia en que fueron colectados (Figura 20). En este *plot* tampoco se pudieron distinguir agrupamientos de los individuos de acuerdo al lugar de su procedencia. Se distribuyeron indistintamente a lo largo del plano de variación; excepto los materiales provenientes de Loja y Pichincha, los cuales junto con otros genotipos formaron un mismo grupo, debido a la pequeña cantidad de accesiones que representan estas provincias.

Fue generada también, la ilustración de los individuos de las provincias de Cañar, Azuay, Loja e Imbabura pertenecientes a los grupos A.1 y A.2. En el caso de las provincias de Loja e Imbabura, se distinguió que los nuevos materiales colectados (grupo A.2) tuvieron una disposición que difirió en gran medida de los individuos del grupo A.1. Para los individuos de las provincias de Azuay y Cañar, como tuvieron

una mejor representación en el grupo A.1, los nuevos materiales aportaron en menor medida a incrementar la dispersión de los mismos en el espacio (Figura 21).

El análisis de coordenadas principales realizado en el programa NTSYS-pc 2.02 los valores Eigen y los porcentajes de variación de cada coordenada se describen en la Tabla 18. Las coordenadas que extraen la mayor cantidad de variación son las dos primeras, por lo que en el gráfico en 2D son distinguibles los clusters (Figura 22). Este gráfico es semejante al representado en la figura 17 aunque en este último las dos primeras dimensiones extraen un menor porcentaje de variación total.

Tabla 18: Valores eigen, porcentaje absoluto y acumulativo de la varianza por coordenada principal del análisis ACoP de 219 genotipos de cebada con 16 loci SSR.

Coorde nada	Valores Eigen	Porcentaje de residuo	Varianza acumulada
1	5.373	18.4	18.4
2	3.635	12.4	30.8
3	2.636	9.1	39.9

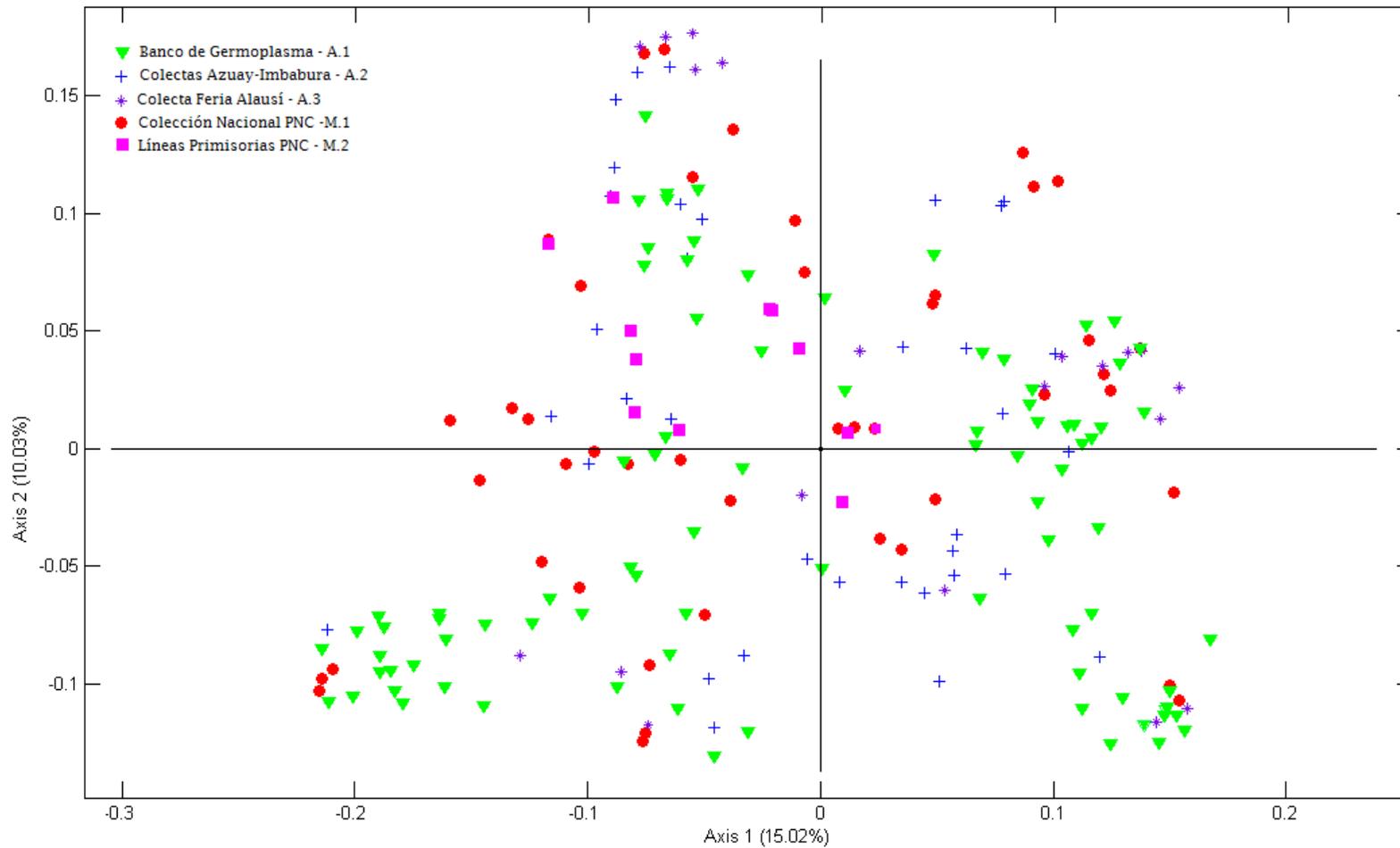


Figura 17: Análisis de coordenadas principales a partir de la matriz de similitud calculada con el coeficiente SM (Sokal & Michener, 1958), de 219 genotipos de cebada con 16 marcadores SSR. Los genotipos están representados de acuerdo a su tipo de germoplasma.

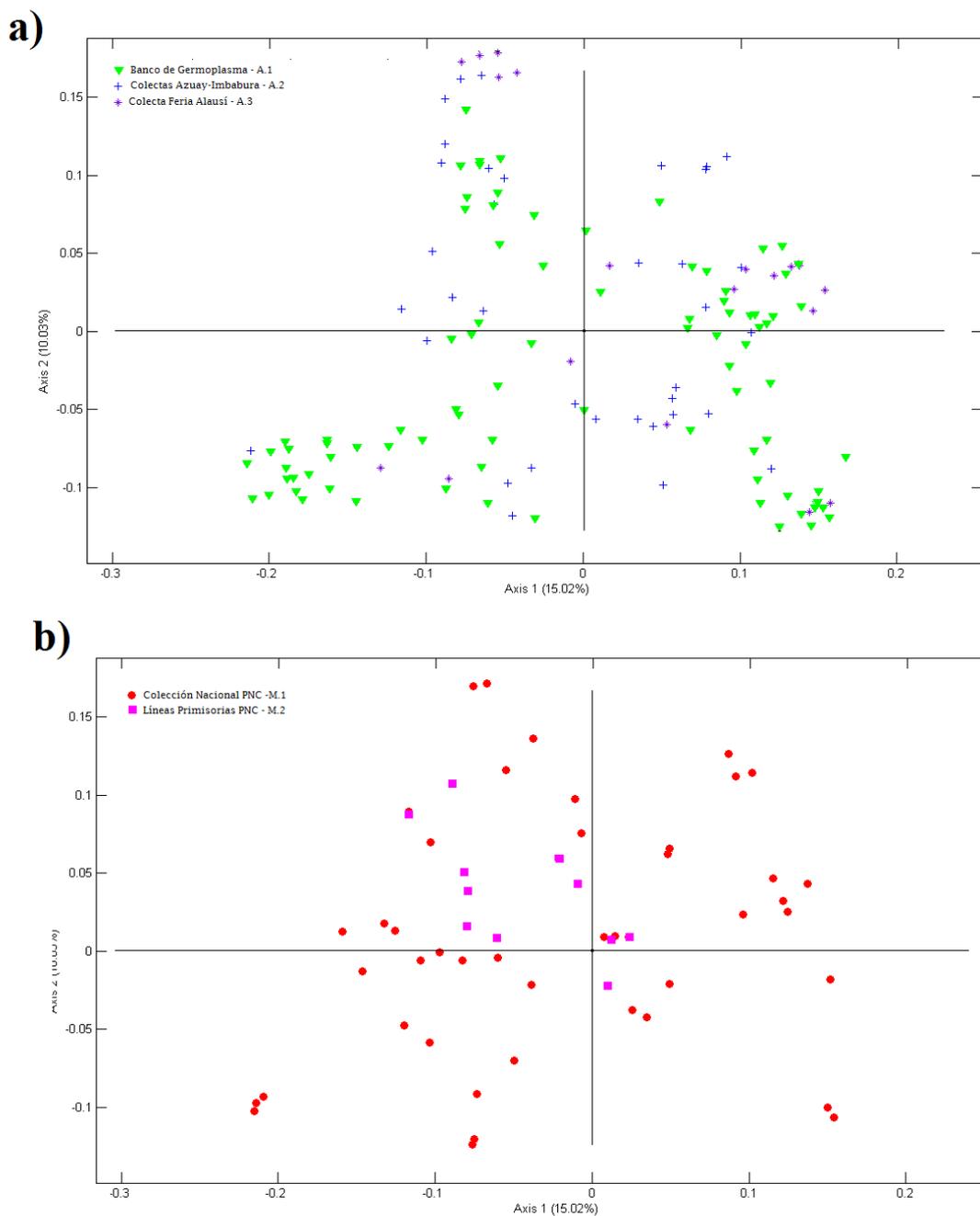


Figura 18: Análisis de coordenadas principales ACoP a partir de la matriz de similitud calculada con el coeficiente SM (Sokal & Michener, 1958) de 219 genotipos de cebada con 16 marcadores SSR. a) Individuos correspondientes a materiales acriollados de cebada (grupos A.1, A.2 y A.3) b) individuos de las variedades mejoradas de cebada (grupos A.1 y A.2) (véase anexos I.1 e I.2 para identificar cada individuo por su código).

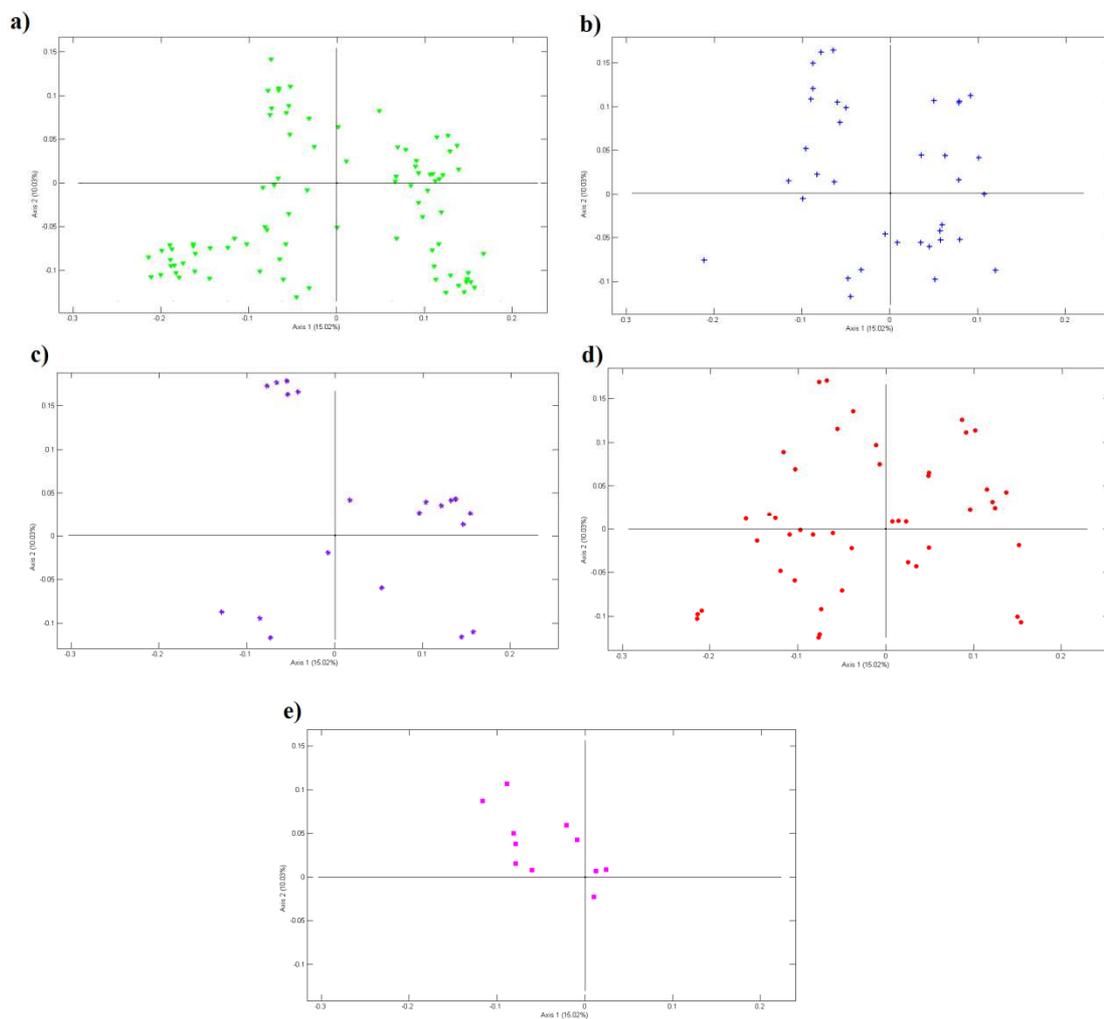


Figura 19: Análisis de coordenadas principales a partir de la matriz de similitud calculada con el coeficiente SM (Sokal & Michener, 1958) de 219 genotipos de cebada con 16 marcadores SSR. a) Germoplasma tipo A.1, b) germoplasma tipo A.2, c) germoplasma tipo A.3, d) germoplasma tipo M.1 y e) germoplasma tipo M.2.

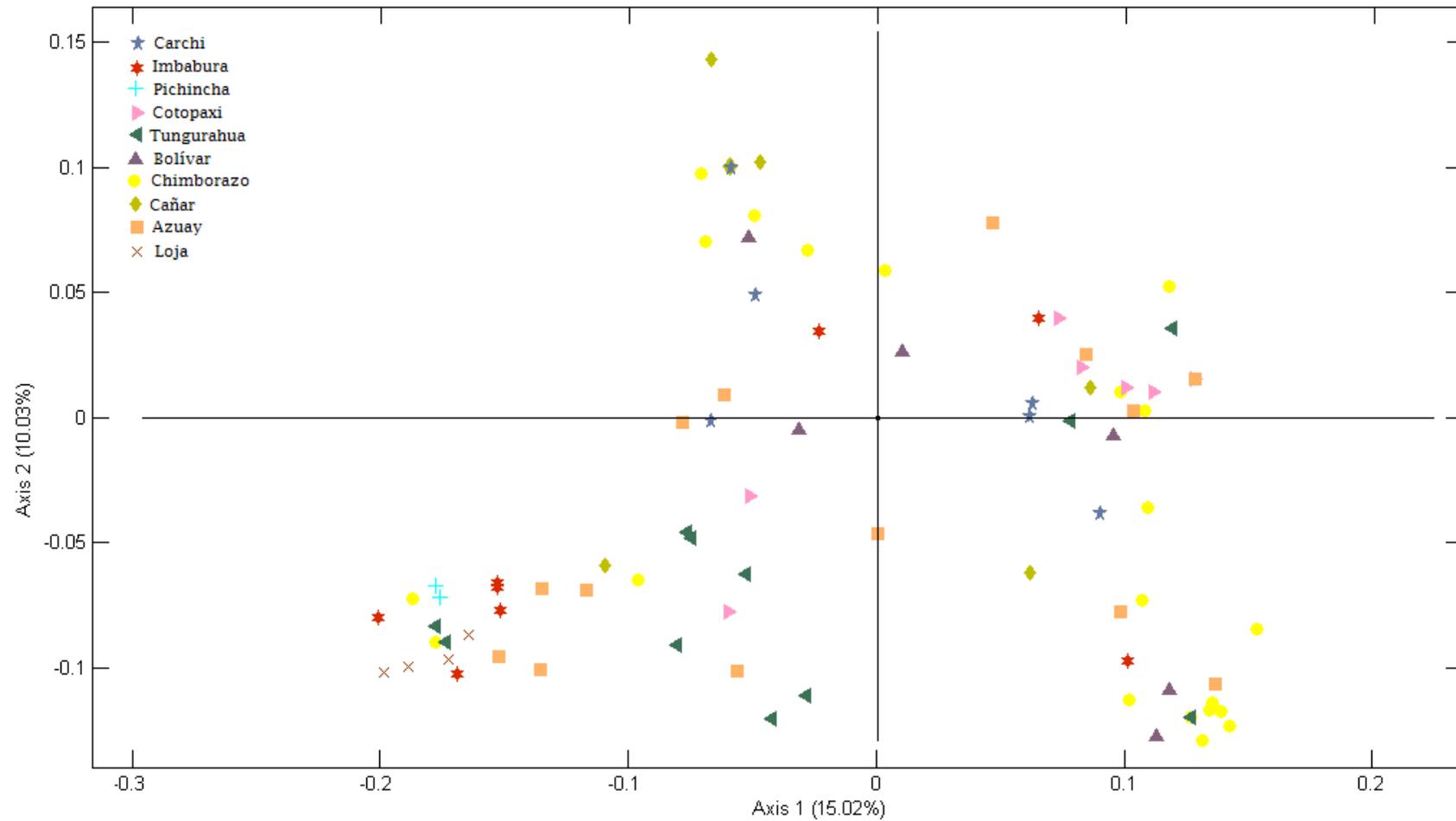


Figura 20: Análisis de coordenadas principales a partir de la matriz de similitud calculada con el coeficiente SM (Sokal & Michener, 1958), de 219 genotipos de cebada con 16 marcadores SSR. Se representan los genotipos del germoplasma tipo A.1 de acuerdo a la provincia de su proveniencia.

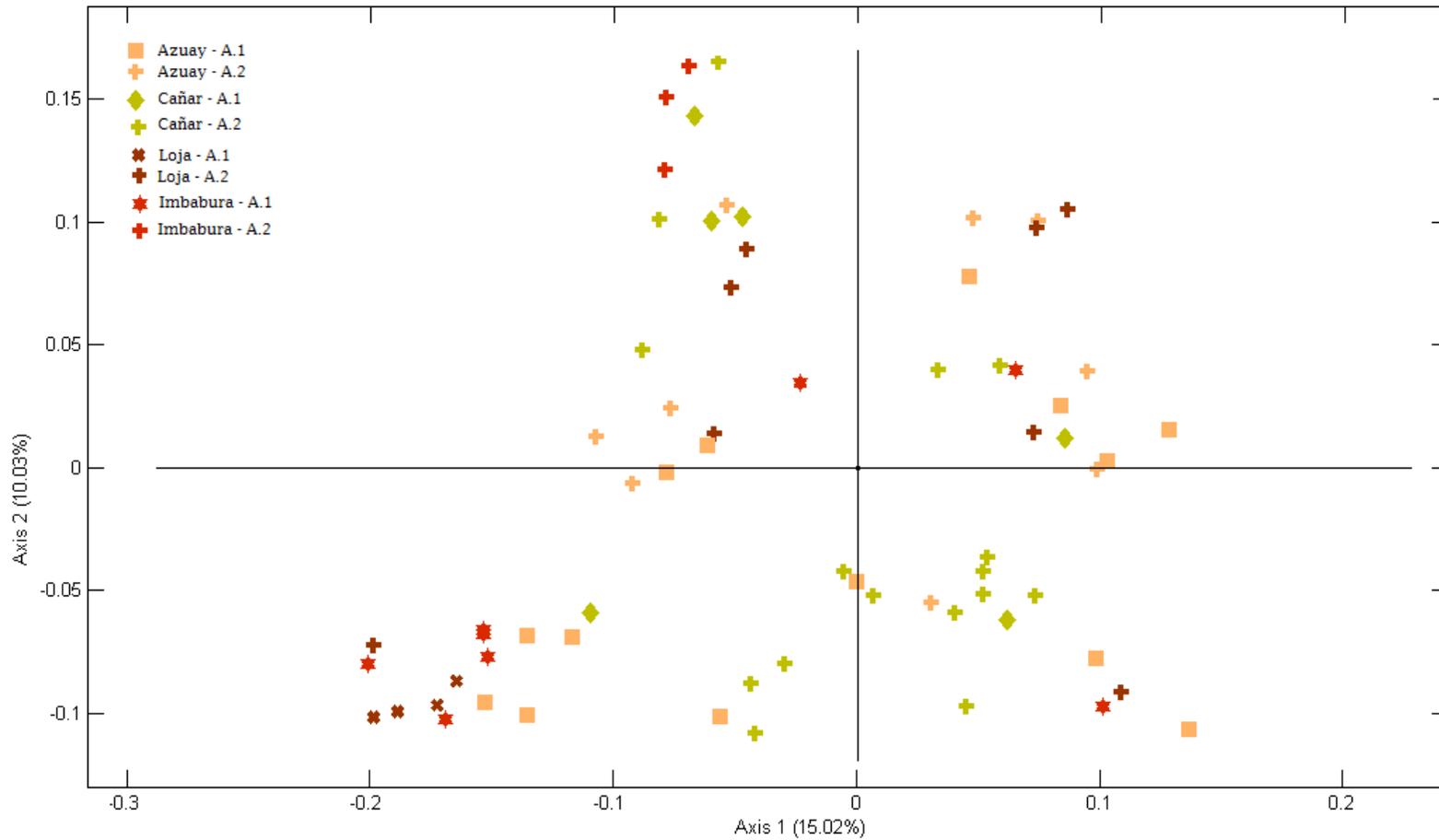


Figura 21: Análisis de Coordenadas Principales a partir de la matriz de similitud calculada con el coeficiente SM (Sokal & Michener, 1958), de 219 genotipos de cebada con 16 marcadores SSR. Se representan los individuos de los germoplasmas tipo A.1 que fueron recolectados en las provincias de Azuay, Cañar, Loja e Imbabura y los individuos del germoplasma tipo A.2.

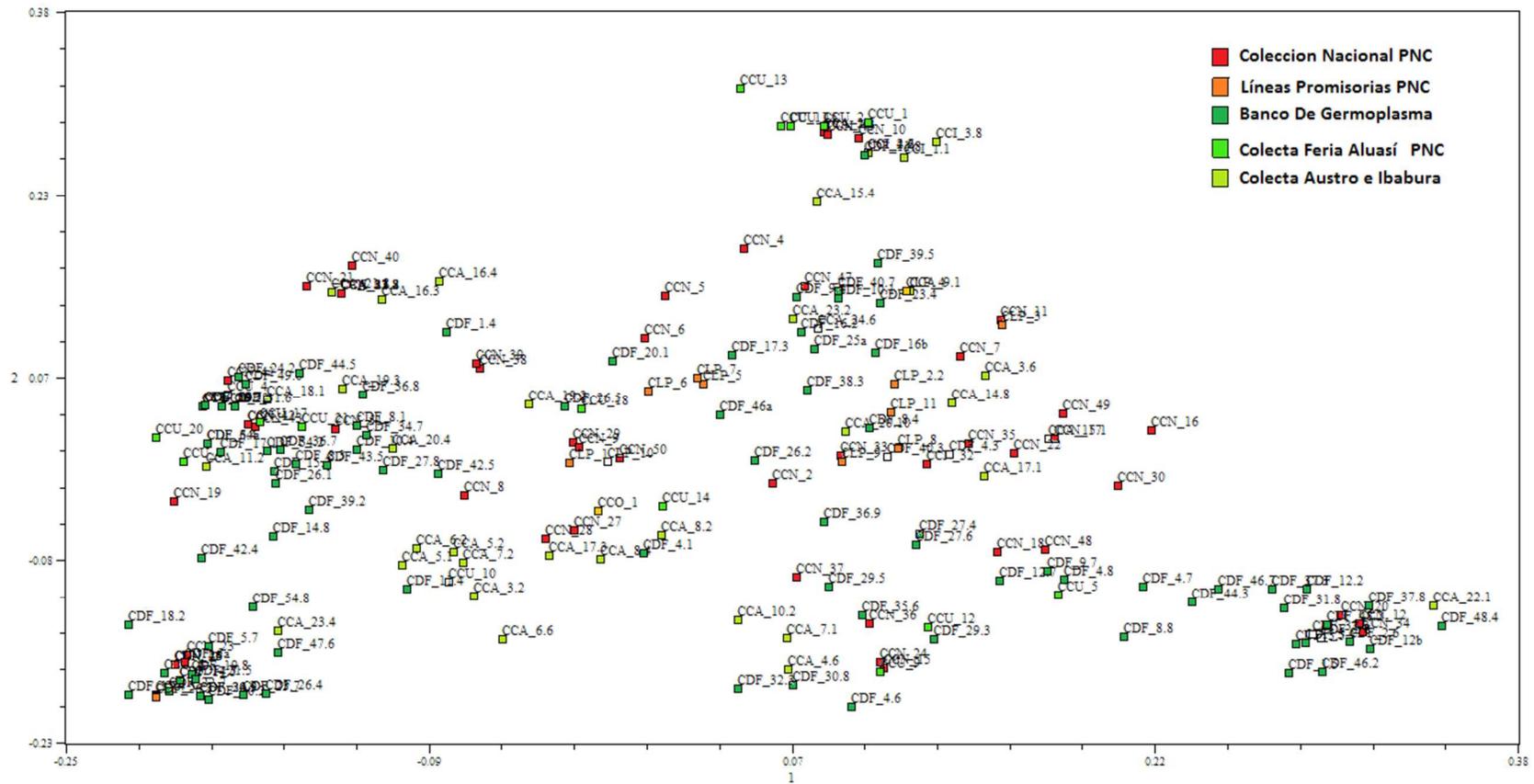


Figura 22: Análisis de Coordenadas Principales a partir de la matriz de similitud calculada con el coeficiente SM (Sokal y Michener, 1958) de 219 genotipos de cebada con 16 marcadores SSR, obtenido con el software NTSYS-pc 2.02.

3.6.4 Análisis de conglomerados.

Se sometieron tanto a la matriz de similitud inicial como a la matriz resultante del ACoP al análisis de conglomerados mediante el algoritmo UPGMA. Se obtuvieron dendogramas que, de acuerdo al programa en el que fueron obtenidos, tuvieron disposiciones diferentes. Para compararlos, se estableció el número de clusters a formarse que reflejaran la mayor variabilidad genética, mediante análisis molecular de varianza AMOVA.

3.6.4.1 Construcción de grupos con el algoritmo UPGMA sobre la matriz de similitud inicial.

Se propuso considerar cada cluster como una subpoblación y someterlos a análisis AMOVA, a las configuraciones formadas desde dos hasta seis clusters, debido a que a partir de la configuración con siete clusters se formaron grupos con un individuo. Los valores de la variación entre clusters para cada configuración se muestran en la figura 23. El valor más alto de variación entre clusters fue para la configuración con seis, la cual fue reportó el 37.53% de variación intergrupala (Tabla 19).



Figura 23: Porcentaje de variación debido a la comparación entre configuraciones de clusters del dendrograma UPGMA de 219 genotipos de cebada con 16 marcadores SSR.

Tabla 19: Análisis molecular de varianza (AMOVA) para 16 loci de la configuración con 6 clusters iniciales.

Fuente de variación	g.l.	Suma de cuadrados	Componentes de varianza	Porcentaje de variación
Entre poblaciones	5	701.88	2.285	37.53
En las poblaciones	213	1638.11	3.809	62.47
Total	218	2339.99	6.098	

En el dendrograma UPGMA obtenido en NTSYS-pc 2.02 se eligió el punto de corte a la distancia de 0.83 y se distinguieron los seis clusters mencionados (Figura 24.a). Esta configuración fue la misma cuando se obtuvo el dendrograma UPGMA a partir de los datos iniciales de la matriz de similitud en el programa MULTBILOT y para esta configuración se obtuvo el valor de coeficiente de correlación cofenética de 0.91 (Figura 24.b). Adicionalmente en este programa se obtuvo el plot de los individuos dispuestos en el plano de variación resultado del ACoP, agrupados en sus respectivos clusters (Figura 24.c).

El detalle de la conformación de los clusters en número de individuos y tipo de material se describe en la tabla 20 y en el anexo J.1 se muestra la conformación de cada cluster con sus respectivos genotipos. El cluster menos numeroso tuvo tres genotipos mientras que el más numeroso tuvo 105. Los clusters formados por genotipos de todos los materiales fueron los dos más numerosos (2 y 5). En cuanto a los tipos de materiales los más numerosos, A.1, A.2 y M.1, tuvieron individuos en todos los clusters, excepto del cluster 1 conformado por 3 individuos del tipo A.1.

Tabla 20: Descripción de los clusters obtenidos mediante algoritmo UPGMA de 219 genotipos de cebada con 16 loci SSR.

Cluster	No. de miembros	Tipo de material				
		A.1	A.2	A.3	M.1	M.2
1	3	3	--	--	--	--
2	105	42	21	15	24	3
3	11	5	2	--	4	--
4	34	23	2	--	9	--
5	50	13	13	5	10	9
6	16	6	3	3	4	--

Los dendogramas de la figura 24 coinciden en la formación de clusters en su mayoría, con la excepción del genotipo CCA20.10 que en la figura 24.a no se empareja en ningún cluster, mientras que en el dendrograma 24.b forma parte del cluster 4. También se comprobó que en el plot del ACoP existieron genotipos que no formaron parte del área señalada para cada uno de los clusters, entre ellos se destacan CCA20.10 y CDF40.3 del cluster 4, CCN40, CLP7 y CCN5 del cluster 2, CCA17.3 del cluster 3, CCN50 y CLP10 del cluster 5.

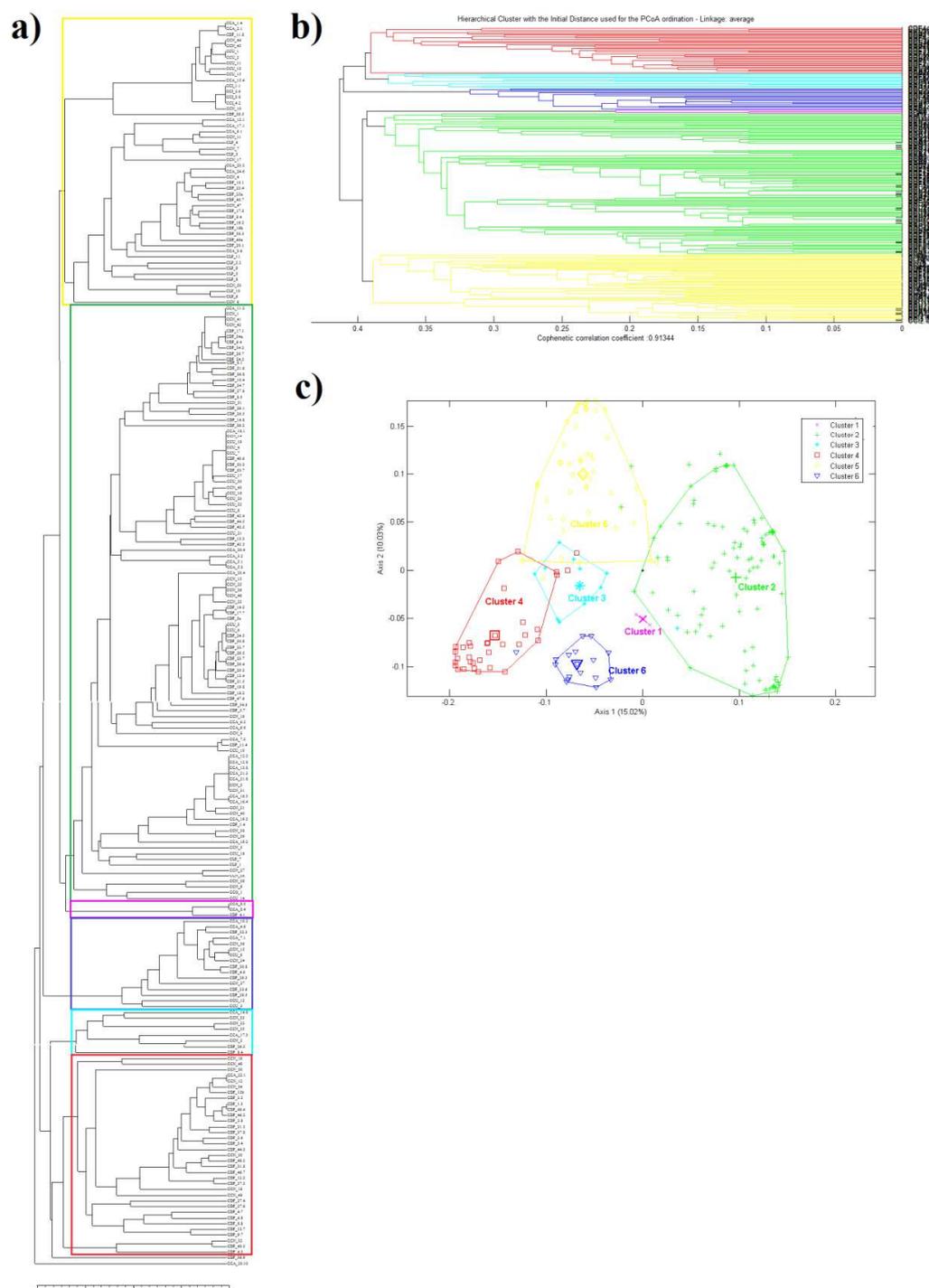


Figura 24: Análisis de conglomerados de 219 genotipos de cebada con 16 loci SSR. a) Dendrograma obtenido con el algoritmo UPGMA con el programa NTSYS-pc 2.02. b) dendrograma obtenido con el algoritmo UPGMA en el programa MULTIBIPLLOT c) Representación en el ACOP de los clusters obtenidos con el algoritmo UPGMA (véase anexo I.3 para mejor resolución de los datos).

3.6.4.2 Construcción de grupos con el algoritmo UPGMA sobre las coordenadas principales.

Para escoger el mejor número de clusters se realizaron análisis de varianza con las configuraciones desde tres hasta 10 clusters que fue el número máximo permitido por el programa. Los valores de la variación entre clusters para cada configuración se muestran en la figura 25. La configuración escogida fue la de 10 clusters en la que se evidenció que el 45.73% de la variación fue explicada por la existente entre los grupos (Tabla 21).

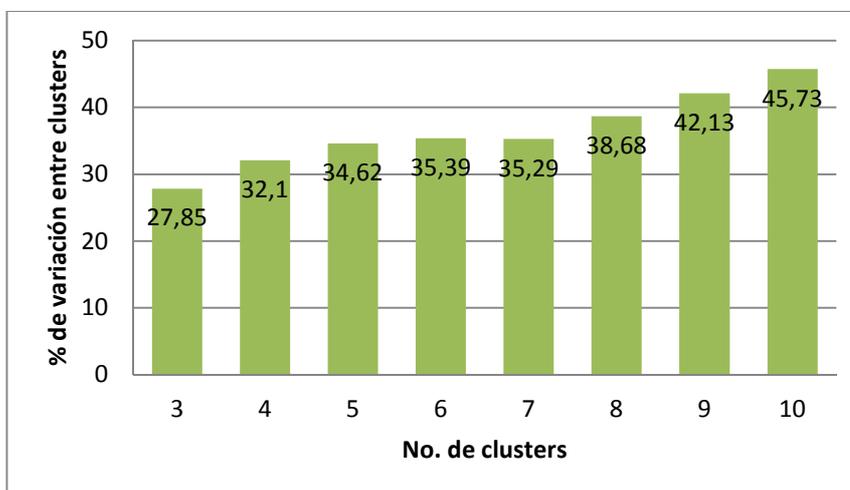


Figura 25: Porcentaje de variación debido a la comparación entre configuraciones de clusters del dendrograma UPGMA en base a al resultado del ACoP de 219 genotipos de cebada con 16 marcadores SSR.

Tabla 21: Análisis molecular de varianza (AMOVA) para 16 loci de los 10 clusters ordinales.

Fuente de variación	g.l.	Suma de cuadrados	Componentes de varianza	Porcentaje de variación
Entre poblaciones	9	1027.57	2.596	45.73
En las poblaciones	209	1312.42	3.081	54.27
Total	218	2339.99	5.676	

El dendrograma obtenido presentó un valor para el coeficiente de correlación cofenética de 0.75 (Figura 26.a). También se generó el plot del ACoP con los 10 clusters UPGMA a partir de los datos resultantes de la ordinación. La lista de los 10 clusters formados y los genotipos pertenecientes a cada uno se detallan en el anexo J.2.

El número de miembros de cada cluster y el tipo de material al que corresponden se describen en la tabla 22. El cluster con más miembros fue el número 4 con 45 y el cluster 3 apenas con 14 fue el menos numeroso. Los tres tipos de materiales más numerosos (A.1, A.2 y M.1) tuvieron presencia en todos los clusters con al menos un miembro, mientras que en los clusters 7 y 8 existieron miembros de cada tipo de material. Es llamativo el hecho de que el germoplasma de tipo A.1 tenga presencia en los clusters 1 y 3 de tan solo un individuo y en el cluster 7 de dos individuos; en comparación con el del tipo M.1 que posee una mejor representación en los clusters con un mínimo de tres individuos en los clusters 2 y 9.

Tabla 22: Descripción de los clusters obtenidos mediante algoritmo UPGMA en base a los resultados del ACoP de 219 genotipos de cebada con 16 loci SSR.

Cluster	No. de miembros	Tipo de material				
		A.1	A.2	A.3	M.1	M.2
1	15	1	5	5	4	--
2	23	9	7	--	3	4
3	14	1	8	--	6	--
4	45	24	4	10	7	--
5	19	5	3	--	9	2
6	23	11	3	3	6	--
7	17	2	9	2	4	1
8	16	4	1	1	5	5
9	22	18	1	--	3	--
10	25	17	1	2	5	--

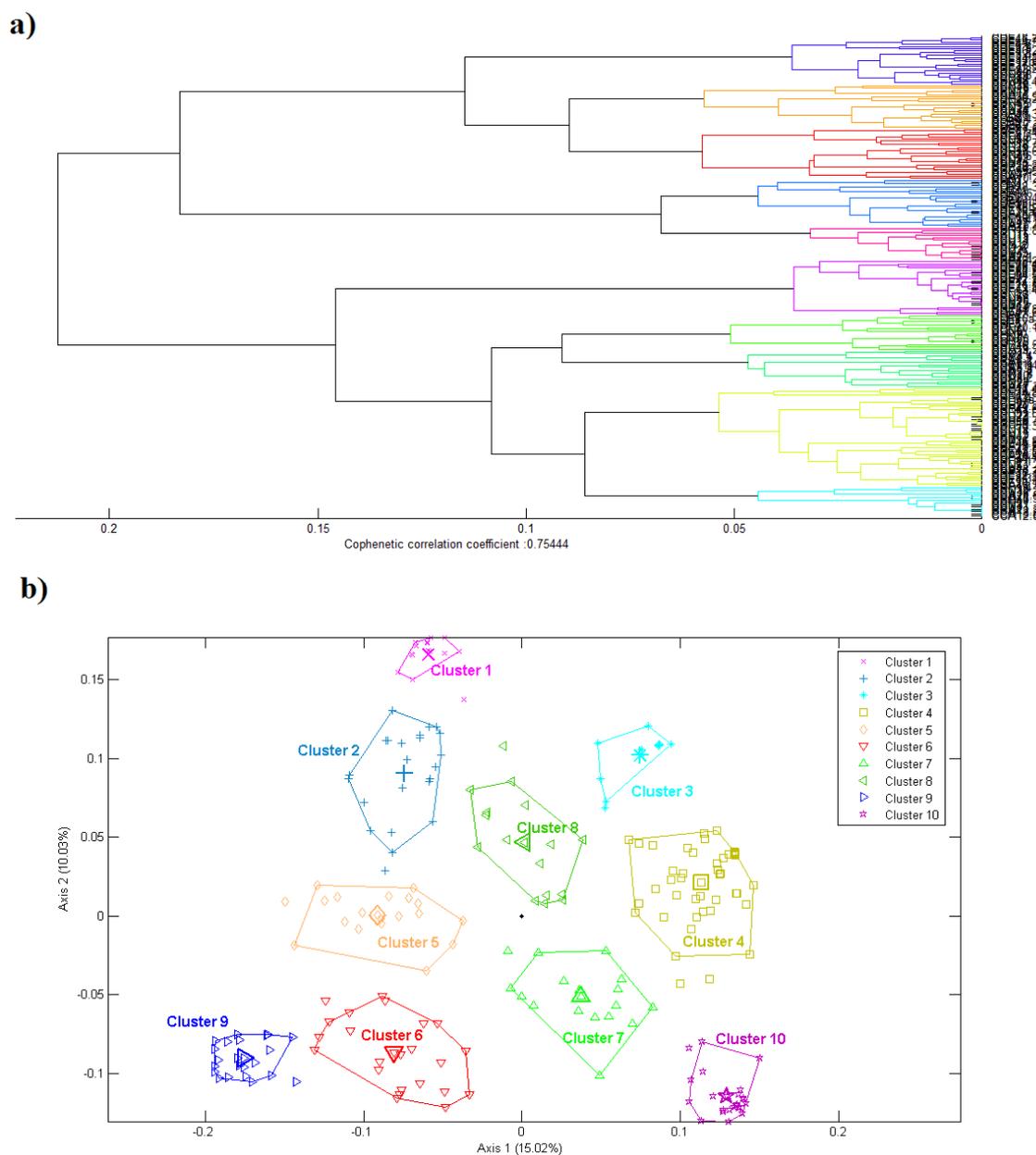


Figura 26: Análisis de conglomerados de 219 genotipos de cebada con 16 loci SSR en base a la matriz de coordenadas principales. a) Dendrograma obtenido con el algoritmo UPGMA en el programa MULTBILOT y b) Representación en el ACoP de los clusters obtenidos con el algoritmo UPGMA.

3.6.4.3 Comparación de los grupos obtenidos en base a la matriz de similitud inicial y las coordenadas principales.

Al hacer una comparación entre los análisis inicial y ordinal de conglomerados, se demostró que el segundo tuvo mejor capacidad para formar un mayor número de

clusters, que por lo tanto permitieron evidenciar un mayor porcentaje de la variación entre clusters formados. En cuanto a la conformación de los clusters obtenidos, se pudo resaltar que todos los miembros de los clusters ordinales 1 y 2 pertenecen también al cluster inicial 5, que además está conformado por proporciones de los clusters ordinales 5 y 8. También todos los miembros de los clusters ordinales 3, 4 y 10 pertenecen al cluster inicial 2 que a su vez incluye a miembros de los clusters iniciales 7 y 8. Los clusters iniciales 5 y 2 fueron los más números, por lo que el análisis de conglomerados que formó un mayor número de clusters los subdividió adecuadamente e incrementó la variación entre los nuevos clusters (Figura 27).

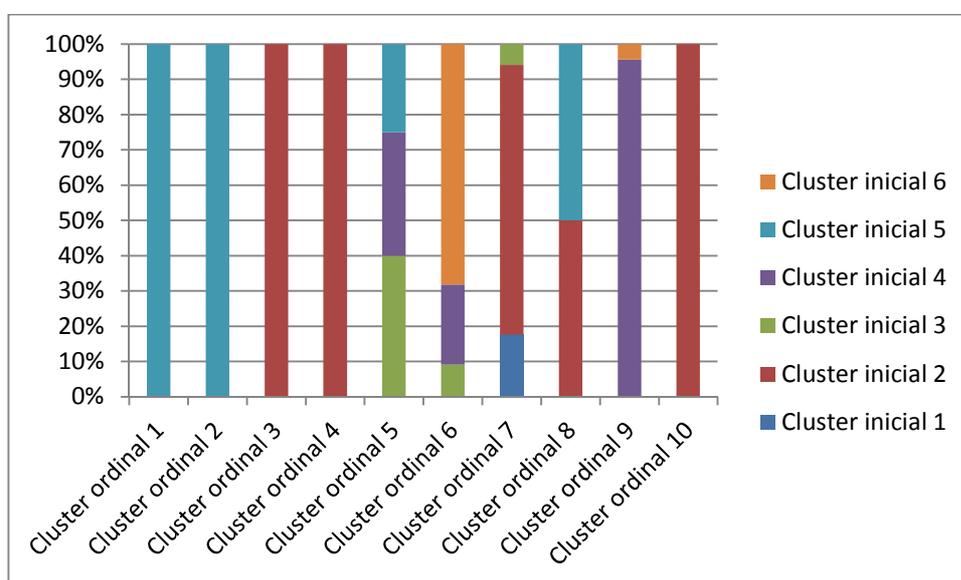


Figura 27: Análisis comparativo de pertenencia entre los clusters iniciales y ordinales obtenidos mediante el algoritmo UPGMA.

3.6.4.4 Diversidad genética inherente a los conglomerados ordinales

Se sometieron a análisis de diversidad genética a los genotipos de cebada de cada uno de los clusters ordinales. En el cluster ordinal cinco, fue detectada la mayor cantidad de alelos por locus al igual que el mayor valor para el índice de diversidad genética de Nei. De acuerdo a este último índice y al número efectivo de alelos y de alelos únicos, los clusters más diversos en orden decreciente fueron: cinco, ocho, siete, seis, dos y cuatro; mientras que en el resto de grupos para este índice de diversidad genética no se superó el valor de 0.30 (Figura 28). Los clusters más diversos fueron los que en el plano de variación se ubicaron más cerca del centro de

coordinadas, y los que explicarían la mayor parte del 54.27% de la variación total debida a la existente dentro de los 10 grupos ordinales (Tabla 21), evidenciado por el respectivo análisis molecular de varianza. En los anexos H.4 y H.5 se muestran las frecuencias alélicas de y los índices de diversidad genética de los clusters ordinales.

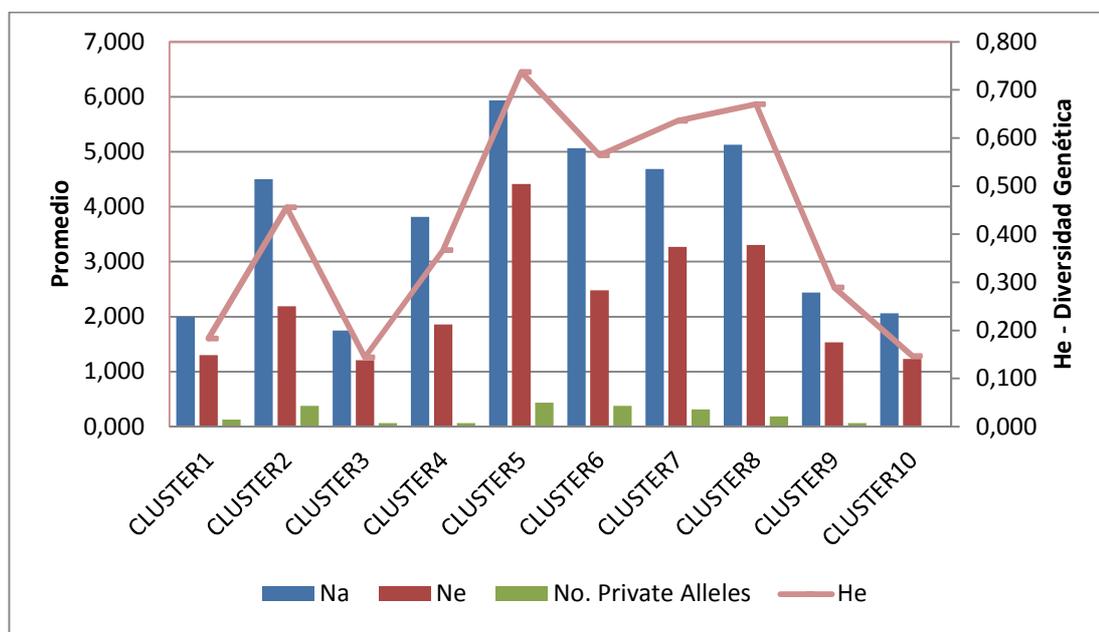


Figura 28: Análisis comparativo de los índices de diversidad genética para los conglomerados o clusters ordinales.

3.6.5 Análisis de remuestreos.

El método de remuestreo escogido fue el Bootstrap total tipo 3 (procrustes) sobre los residuales, el cual presentó los mejores resultados y el mejor ajuste (0.99). Cuando se seleccionó la opción de ilustración de los intervalos de confianza con elipses para los individuos identificados por *clusters* ordinales se detectaron pocas intersecciones entre estas (Figura 29).

El valor promedio de los errores estándar de los intervalos de confianza para los valores eigen fue de 0.014; mientras que para los individuos en el eje 1 y 2 fueron de 0.008 y 0.010 respectivamente.

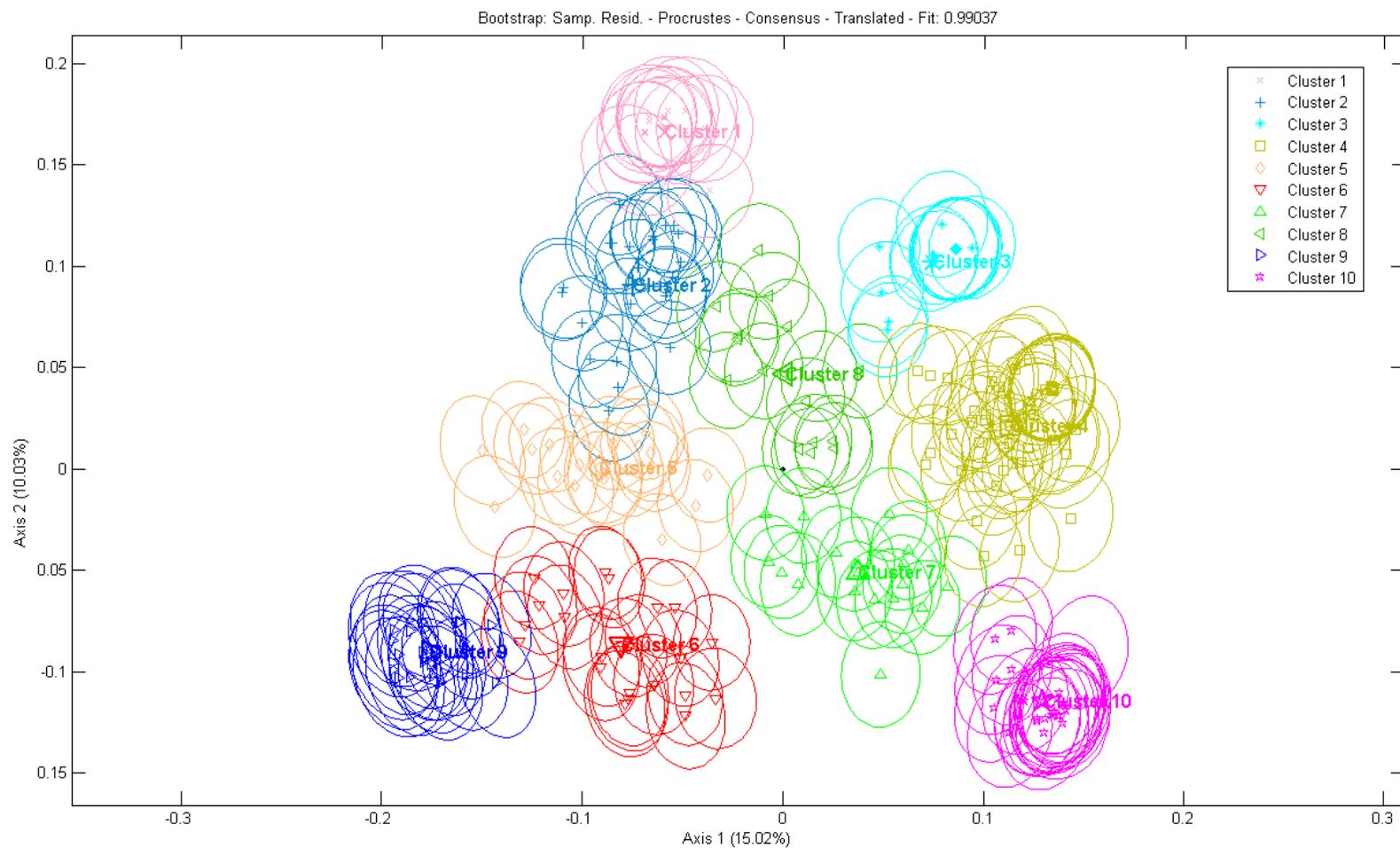


Figura 29: Análisis de remuestreo sobre residuales tipo procrustes representado el diagrama del ACoP de 219 genotipos de cebada con 16 loci SSR.

3.6.6 Análisis moleculares de varianza complementarios

Se realizó el análisis molecular de varianza que confirmó la inexistencia de una significativa relación de agrupación entre los genotipos de cebadas acriolladas y mejoradas. Como se muestra en la tabla 23, la variación molecular explicada entre estos grupos de genotipos, establecidos *a priori*, fue mínima en comparación con la existente dentro de estos grupos.

Tabla 23: Análisis molecular de varianza para 16 loci de los grupos, establecidos *a priori*, de cebada acriollada y de mejoramiento.

Fuente de variación	g.l.	Suma de cuadrados	Componentes de varianza	Porcentaje de variación
Entre poblaciones	1	26.669	0.119	1.98
En las poblaciones	217	2449.248	5.909	98.02
Total	218	2475.917	6.030	

Con el objetivo de robustecer el análisis de conglomerados se efectuaron análisis de varianza tanto entre los clusters ordinales más alejados (1, 3, 9 y 10) y más cercanos (2, 4, 5, 6, 7 y 8) del origen del plano de variación por separado.

Los clusters 1, 3, 9 y 10 evidenciaron menores índices de diversidad genética dentro de los mismos. Se evidenció que la fuente de variación entre estos grupos (77.10%) fue muy superior a la variación dentro estos clusters (Tabla 24) y esto explicó gran parte del 45.73% de variación intergrupala detectado en el análisis molecular de varianza de los diez clusters.

Tabla 24: Análisis molecular de varianza para 16 loci de los clusters ordinales 1, 3, 9 y 10 obtenidos en el ACoP.

Fuente de variación	g.l.	Suma de cuadrados	Componentes de varianza	Porcentaje de variación
Entre poblaciones	3	527.946	4.80781	77.10
En las poblaciones	70	205.655	1.42816	22.90
Total	73	873.762	6.23326	

Mientras que por el contrario; para los clusters 2, 4, 5, 6, 7 y 8 fue mayor la fuente de variación dentro de estos clusters que la existente entre los mismos (Tabla 25) y esto explicó que en estos conglomerados fueran reportados los índices de diversidad genética de mayor valor.

Tabla 25: Análisis molecular de varianza para 16 loci de los clusters ordinales 2, 4, 5, 6, 7 y 8 obtenidos en el ACoP.

Fuente de variación	g.l.	Suma de cuadrados	Componentes de varianza	Porcentaje de variación
Entre poblaciones	5	470.201	2.06227	32.32
En las poblaciones	139	1142.467	4.3191	67.68
Total	144	873.762	6.23326	

3.6.7 Estadísticos F

En la Tabla 26 se muestran los valores de los estadísticos F_{IS} (coeficiente de consanguinidad), F_{ST} (índice de fijación) y F_{IT} (coeficiente de consanguinidad total). En el primer índice que mide el déficit de heterocigotos fue distinto de 1.00 en los locus que detectaron estos individuos, así como el coeficiente F_{IT} ; siendo el locus Bmag0125 el que más se aleja del valor +1 que indica ausencia de heterocigotos. Los valores promedio para los índices F_{IT} y F_{IS} se desvían levemente de +1, indicando que el número de eventos en los que es posible encontrar individuos heterocigotos es muy bajo; mientras que para el índice F_{ST} el valor promedio para las grupos de cebada acriollada y de mejoramiento indica una mínima diferenciación genética.

Tabla 26: Estimación de Weir & Cockerham (1984) de los índices F_{IT} , F_{ST} , y F_{IS} para cada locus de las grupos establecidos *a priori* de cebada acriollada y de mejoramiento.

LOCUS	F_{IT}	F_{ST}	F_{IS}
Bmac0213	0.964	0.056	0.962
Ebmac0501	0.994	0.009	0.994
Bmac0209	1.000	-0.009	1.000
Bmag0223	1.000	0.004	1.000
Bmac0303	1.000	0.032	1.000
Bmac0040	0.994	0.006	0.994
Bmag0007	1.000	-0.001	1.000
Bmag0217	0.782	0.091	0.761
Bmag0384	1.000	0.007	1.000
Bmac0127	1.000	-0.001	1.000
Bmac0134	1.000	0.012	1.000
Bmag0125	0.655	0.044	0.639
Bmag603	0.974	-0.002	0.974
EBmac0679	1.000	0.018	1.000
Bmac0316	1.000	-0.002	1.000
scssr2306	1.000	-0.007	1.000
Promedio	0.964	0.014	0.964

En base a los resultados de los análisis de agrupamiento y multivariados que indicaron la diferenciación significativa de clusters, estos fueron sometidos a análisis para los estadísticos F. El índice de fijación F_{ST} confirmó que la mayor diferenciación genética se obtuvo cuando fueron analizados los clusters ordinales más alejados del origen y se evidenció incluso, una diferenciación menor a la revelada por todos los clusters, cuando se analizaron los grupos más cercanos al origen (Tabla 27). La lista detallada de los estadísticos F para cada locus SSR de los casos en estudio de la tabla 27 se muestra en el anexo K.

Tabla 27: Estimación de Weir & Cockerham (1984) de los promedios de los índices F_{IT} , F_{ST} , y F_{IS} de los clusters ordinales en distintos casos.

Casos en estudio	F_{IT}	F_{ST}	F_{IS}
Clusters del 1 al 10.	0.966	0.452	0.939
Clusters 2, 4, 5, 6, 7 y 8	0.971	0.308	0.958
Clusters 1, 3, 9 y 10	0.962	0.764	0.840

3.6.8 Biplot logístico externo

Se efectuó el análisis de biplot logístico externo, BLE, sobre la base del ACoP que permitió identificar relaciones entre individuos (genotipos de cebada) y variables (alelos de los loci SSR) en el espacio euclídeo. En este biplot, los alelos estuvieron representados por líneas en la dirección de incremento de predicción de probabilidad. El inicio del vector fue el punto que predijo 0.5 y el final fue el punto que predijo 0.75. La longitud de los vectores fue relacionada a la capacidad de predecir la presencia o ausencia de los alelos en cada individuo.

Con el criterio de corrección Bonferroni y un valor para pseudo $R^2 \geq 0.5$ se establecieron 27 alelos de 157 alelos encontrados, pertenecientes a 14 de los 16 loci SSR. Los marcadores Bmag0217 y Bmag0223 son los loci sin alelos representativos. El número de alelos por locus varió entre uno para 7 loci, hasta tres en 6 loci. El mayor valor de R^2 fue para el alelo EBmac0501(125pb) y el menor para los alelos Bmag0125(113pb) y Bmac0213(134) (Tabla 28).

Tabla 28: Alelos significativos ($R^2 \geq 0.5$, $p < 0.05$) resultantes del biplot logístico externo.

No.	Alelo	p -valor	R^2
1	Bmac0040(202pb)	0	0.67
2	Bmac0127(124pb)	0	0.81
3	Bmac0134(162pb)	0	0.57
4	Bmac0209(179pb)	0	0.58
5	Bmac0213(150pb)	0	0.63
6	Bmac0213(160pb)	1.68E-12	0.62
7	Bmac0213(170pb)	0	0.81
8	Bmac0303(136pb)	0	0.56
9	Bmac0316(134pb)	0	0.51
10	Bmac0316(136pb)	0	0.66
11	Bmac0316(160pb)	0	0.69
12	Bmag0007(192pb)	0	0.71
13	Bmag0007(198pb)	0	0.61
14	Bmag0125(107pb)	0	0.61
15	Bmag0125(109pb)	0	0.70
16	Bmag0125(113pb)	0	0.51
17	Bmag0384(114pb)	0	0.71
18	Bmag0384(116pb)	0	0.65
19	Bmag0384(118pb)	0	0.64
20	Bmag0303(110pb)	0	0.53
21	EBmac0501(125pb)	0	0.84
22	EBmac0501(149pb)	1.89E-13	0.56
23	EBmac0501(155pb)	0	0.67
24	EBmac0679(116pb)	0	0.80
25	EBmac0679(144pb)	0	0.70
26	EBmac0679(148pb)	0	0.75
27	scssr02306(151pb)	0	0.79

Adicionalmente se relacionaron los alelos mencionados con la probabilidad de presencia en los individuos de los clusters ordinales (Figura 30). En la tabla 29 se muestra la probabilidad de encontrar cada uno de los alelos descritos en la tabla 28 en los clusters ordinales. Los clusters 1, 3, 4, 9 y 10 presentaron probabilidades iguales o superiores a 90% de que sus individuos contengan al menos dos de los alelos respectivos para cada uno. Sin embargo, para que un alelo sea considerado como representativo, debe presentar una elevada frecuencia alélica debida exclusivamente a dicho cluster, como puede ser comprobado en el anexo H.4.

Por lo tanto, los alelos Bmag0125(109pb), EBmac0679(148pb) y Bmac0127(124pb) presentan una probabilidad superior al 90% de pertenecer casi exclusivamente a los clusters 1, 9 y 10 respectivamente.

Se tomaron como ejemplo demostrativo el alelo con mayor R^2 y los tres alelos mencionados para verificar sus predicciones binarias y de relación con los individuos (genotipos). Los gráficos de predicciones binarias generados presentan los vectores de las variables, los individuos en forma de puntos y un eje perpendicular a la dirección del vector. Los puntos que se encuentran en el área donde se incrementa la probabilidad de presencia del alelo (sombreada roja) y efectivamente lo poseen se pintan como puntos rojos caso contrario como puntos azules, entonces la capacidad de un alelo de discriminar a los individuos se estimó al observar la manera de distribución de los puntos con sus respectivas áreas sombreadas (Figura 31).

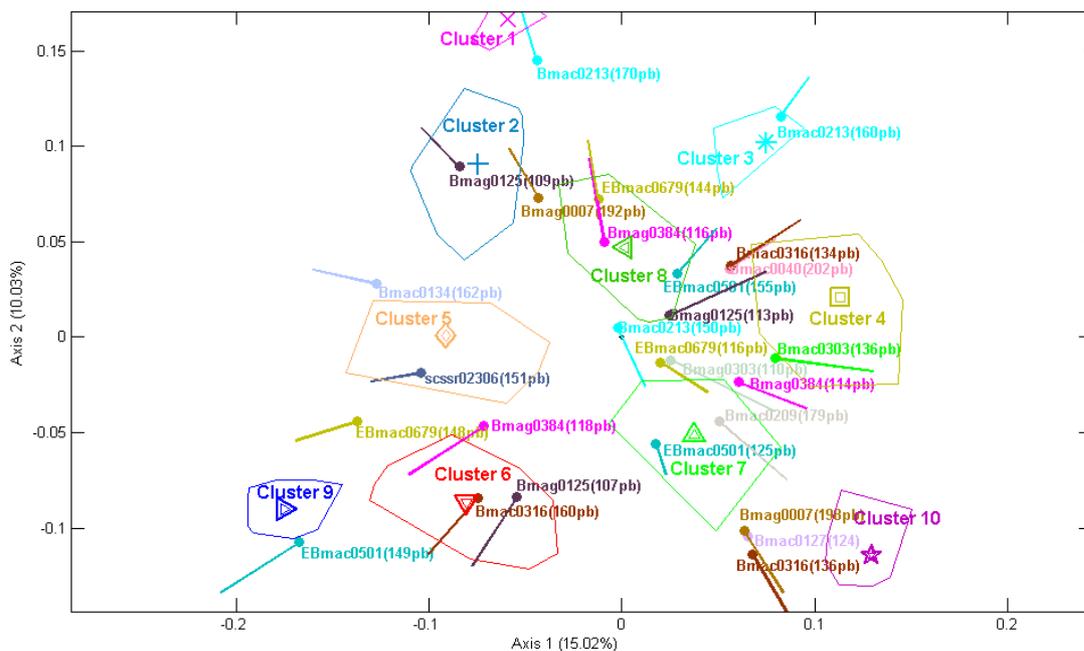


Figura 30: Biplot logístico externo del ACoP de 219 genotipos de cebada con 16 loci SSR, de un total de 157 alelos reportados se ilustran los 26 más representativos ($R^2 \geq 0.5$, $p < 0.05$).

Tabla 29: Porcentajes de las predicciones de probabilidad de presencia los alelos más representativos ($R^2 \geq 0.5$, $p < 0.05$) en cada cluster ordinal.

Alelo No.*	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10
1	0.0	0.0	100.0	86.7	10.0	0.0	29.4	25.0	0.0	8.0
2	0.0	0.0	0.0	2.2	0.0	0.0	5.9	0.0	0.0	92.0
3	46.67	4.55	0.00	0.00	0.00	6.25	15.00	4.55	78.26	0.0
4	0.0	0.0	0.0	48.9	10.0	4.5	64.7	31.3	0.0	92.0
5	0.0	4.5	14.3	80.0	15.0	81.8	64.7	25.0	73.9	100.0
6	0.0	0.0	78.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	86.7	9.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8	0.0	0.0	78.6	46.7	0.0	0.0	58.8	18.8	0.0	80.0
9	0.0	0.0	100.0	86.7	10.0	0.0	29.4	25.0	0.0	8.0
10	0.0	4.5	0.0	0.0	0.0	0.0	5.0	0.0	0.0	88.0
11	0.0	0.0	0.0	0.0	20.0	45.5	11.8	6.3	91.3	0.0
12	100.0	86.4	0.0	0.0	20.0	0.0	11.8	25.0	0.0	0.0
13	0.0	0.0	0.0	4.4	5.0	0.0	23.5	12.5	0.0	84.0
14	0.0	0.0	0.0	0.0	15.0	63.6	23.5	18.8	87.0	0.0
15	93.3	31.8	0.0	0.0	25.0	4.5	0.0	0.0	0.0	0.0
16	6.7	54.5	100.0	75.6	10.0	4.5	29.4	56.3	0.0	80.0
17	0.0	0.0	0.0	80.0	5.0	0.0	35.3	18.8	0.0	96.0
18	100.0	86.4	100.0	6.7	20.0	0.0	17.6	56.3	0.0	0.0
19	0.0	0.0	0.0	2.2	65.0	45.5	23.5	12.5	100.0	0.0
20	0.0	0.0	85.7	60.0	15.0	31.8	76.5	43.8	0.0	92.0
21	0.0	0.0	0.0	2.2	0.0	63.6	70.6	0.0	4.3	100.0
22	0.0	0.0	0.0	0.0	0.0	9.1	0.0	6.3	52.2	0.0
23	93.3	4.5	100.0	91.1	20.0	4.5	17.6	31.3	0.0	0.0
24	0.0	0.0	0.0	95.6	25.0	0.0	76.5	31.3	0.0	100.0
25	100.0	54.5	100.0	0.0	20.0	4.5	5.9	25.0	0.0	0.0
26	0.0	4.5	0.0	0.0	5.0	9.1	0.0	6.3	95.7	0.0
27	0.0	9.1	0.0	0.0	40.0	31.8	0.0	6.3	100.0	0.0

*Alelo No: Referencia de la tabla 3.15

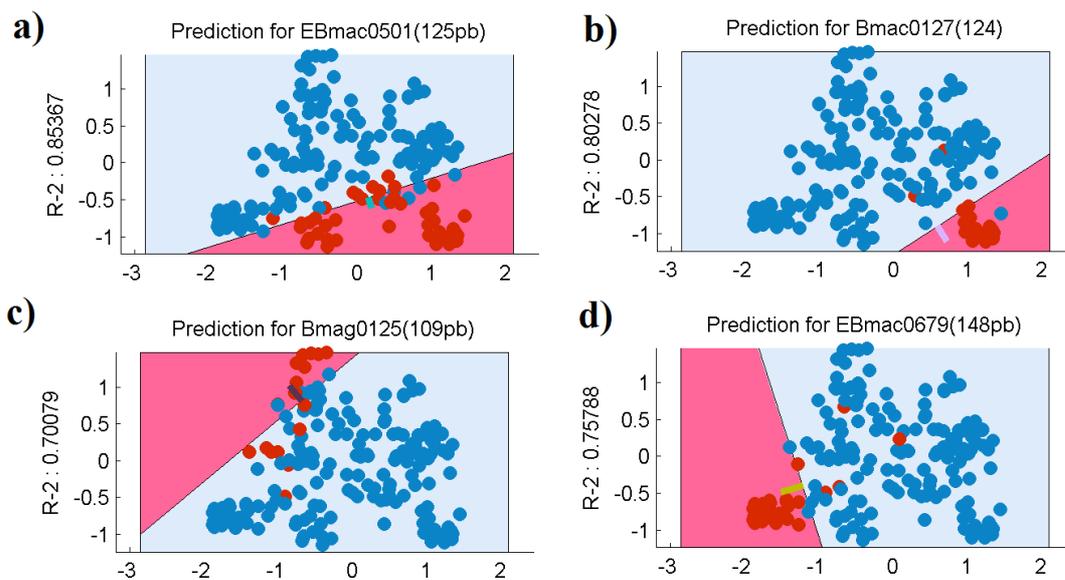


Figura 31: Proyecciones binarias de la probabilidad de presencia de individuos sobre cuatro alelos. a) Para el alelo Ebmac0501(135bp), b) Bmac0127(124pb), c) Bmag0125 (109bp) y d) EBmac0679(148bp).

Las predicciones de las relaciones individuos-variable, se ilustran en la figura 32 para los cuatro alelos descritos en la predicción binaria. En esta representación del biplot se resalta cada alelo seleccionado y la dirección de su vector es prolongada en ambas direcciones y cada individuo es proyectado perpendicularmente sobre esta. Si esta proyección tiene una línea continua el individuo efectivamente presenta el alelo caso contrario la línea es discontinua.

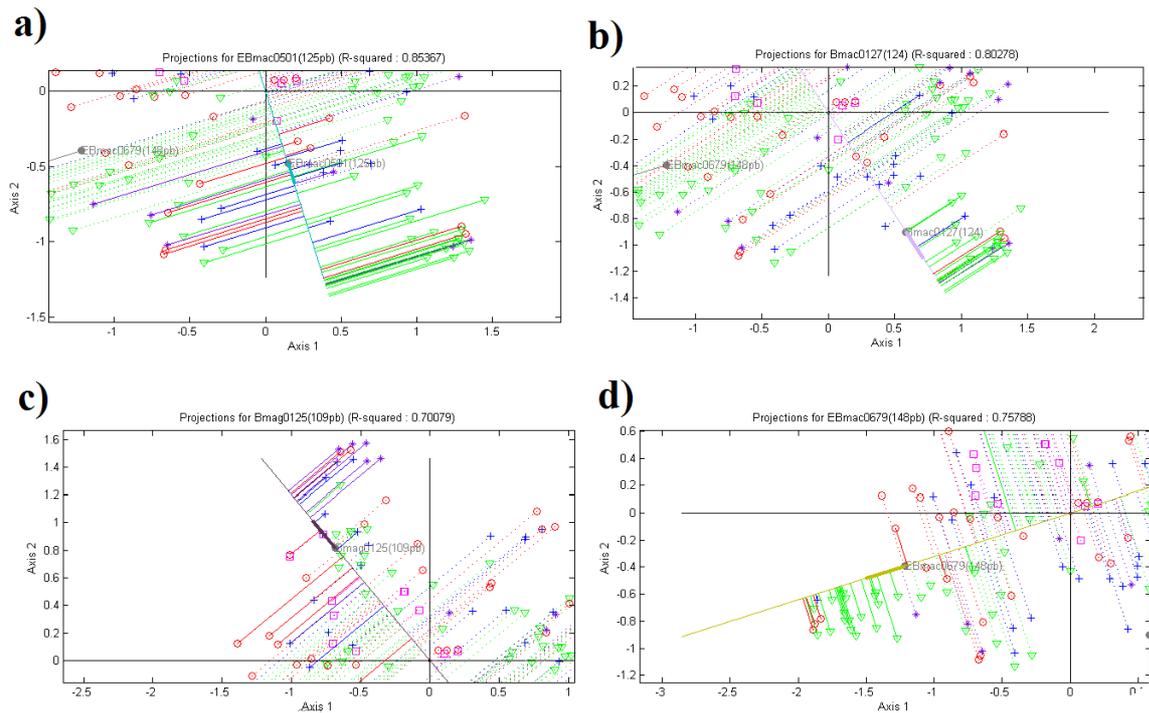


Figura 32: Proyecciones de la probabilidad de presencia de individuos sobre cuatro alelos SSR. a) Para el alelo Ebmac0501(135bp) b) Bmac0127(124pb) c) Bmag0125 (109bp) y d) EBmac0679(148bp).

La bondad de ajuste global del BLE (promedio para los individuos y para las variables) como porcentaje de clasificaciones correctas resultó en $93.22 \pm 7.33\%$; es decir, si usamos el biplot para predecir cada genotipo de cada individuo, más del 93% de ellos serán correctos.

3.6.9 Determinación de un set núcleo de genotipos de cebada.

Se sometió al análisis de colección núcleo a la matriz alélica depurada sin los genotipos duplicados mediante la estrategia *M*. El software POWERCORE arrojó una colección núcleo de 41 genotipos mientras que el programa COREFINDER de 38 genotipos. En la tabla 30 se describen los genotipos mencionados, el tipo de material y el cluster ordinal al que pertenecen.

Tabla 30: Lista de 51 genotipos del set núcleo de la colección completa de 219 genotipos de cebada analizados con 16 loci SSR, el cluster ordinal al

que pertenecen, tipo de material y el software con que fueron encontrados.

	Genotipo	Corefinder	Powercore	Cluster ordinal	Tipo de material
1	CCU_11	X	X	1	A3
2	CLP_11	X	X	2	M2
3	CDF_39.5	X	X	2	A1
4	CCA_15.4		X	2	A2
5	CCA_3.6	X	X	2	A2
6	CLP_2		X	2	M2
7	CLP_4	X	X	2	M2
8	CDF_38.3	X		2	A1
9	CCA_24.6	X		2	A2
10	CDF_40.7		X	2	A1
11	CDF_1.4		X	3	A1
12	CCN_40	X		3	M1
13	CCA_20.4	X	X	4	A2
14	CDF_36.8		X	4	A1
15	CDF_44.5	X		4	A1
16	CDF_8.3	X	X	4	A1
17	CCA_20.10		X	5	A2
18	CCN_49	X	X	5	M1
19	CCN_2	X	X	5	M1
20	CCN_33	X	X	5	M1
21	CDF_36.9	X	X	5	A1
22	CCN_35	X		5	M1
23	CDF_4.3	X	X	5	A1
24	CDF_8.4		X	5	A1
25	CCN_17	X	X	5	M1
26	CLP_8		X	5	M2
27	CLP_9	X		5	M2
28	CCA_14.8	X		5	A2
29	CDF_4.8		X	6	A1
30	CDF_29.5	X	X	6	A1
31	CCU_12	X	X	6	A3
32	CDF_27.4	X	X	6	A1
33	CCN_36		X	6	M1
34	CCN_37	X	X	6	M1
35	CCN_24	X	X	6	M1
36	CDF_29.3	X		6	A1
37	CDF_9.7	X		6	A1
38	CCA_8.4	X	X	7	A2
39	CCA_6.2	X	X	7	A2
40	CCU_10	X	X	7	A3
41	CCA_6.6	X	X	7	A2

Continúa

Genotipo	Corefinder	Powercore	Cluster	Tipo de	Genotipo
-----------------	-------------------	------------------	----------------	----------------	-----------------

			ordinal	material	
42	CCN_5		X	8	M1
43	CCN_6	X	X	8	M1
44	CCA_19.2	X	X	8	A2
45	CLP_5	X		8	M2
46	CLP_7	X	X	8	M2
47	CCU_18		X	8	A3
48	CCU_5	X	X	9	A3
49	CDF_44.3	X	X	9	A1
50	CCN_23		X	10	M1
51	CDF_19.8	X	X	10	A1

De los 51 genotipos, el 55% (28 genotipos) fueron comunes entre los resultados de los dos programas. El cluster ordinal 5 es el que más genotipos de este set presentó, mientras que tan solo se reportó un genotipo perteneciente al cluster 1. En cuanto al tipo de germoplasma, el grupo mayoritario fue el acriollado (63%) y dentro de este el tipo M1 (Banco de Germoplasma). Los materiales mejorados representaron el 37% del set núcleo.

Se realizó un análisis de diversidad genética comparativa entre el set núcleo establecido y el resto de la colección (colección reserva) y se describió en la tabla 31. Todos los índices calculados fueron superiores en el set núcleo, el número de alelos por locus se acerca al valor reportado para la colección completa. El índice de diversidad genética de Nei o heterocigosidad esperada fue superior en el set núcleo, incluso al de la colección completa. El número de alelos por locus fue casi equivalente entre el set núcleo y la colección completa, y al tener un menor número de individuos este presentó un mayor número efectivo de alelos por locus que la colección reserva y completa.

Tabla 31: Cuadro comparativo de los de diversidad genética para los 16 loci SSR en la colección completa y el set core de cebada.

	Set núcleo	Colección reserva	Colección completa
No. de alelos por locus	9.75	8.63	9.81
No. de alelos efectivos por locus	5.67	4.21	2.86
No. de alelos únicos por locus	1.75	0.63	--
Heterocigosidad esperada	0.80	0.74	0.75

Se rescataron en el plot del ACoP la disposición de los genotipos del set núcleo en el espacio. Se pudieron diferenciar que la mayoría de los genotipos se distribuyen cercanos al origen de variación. Uno de los tipos de germoplasma mejor distribuidos en el espacio fueron los de los materiales de la Colección Nacional del PNC (Figura 33).

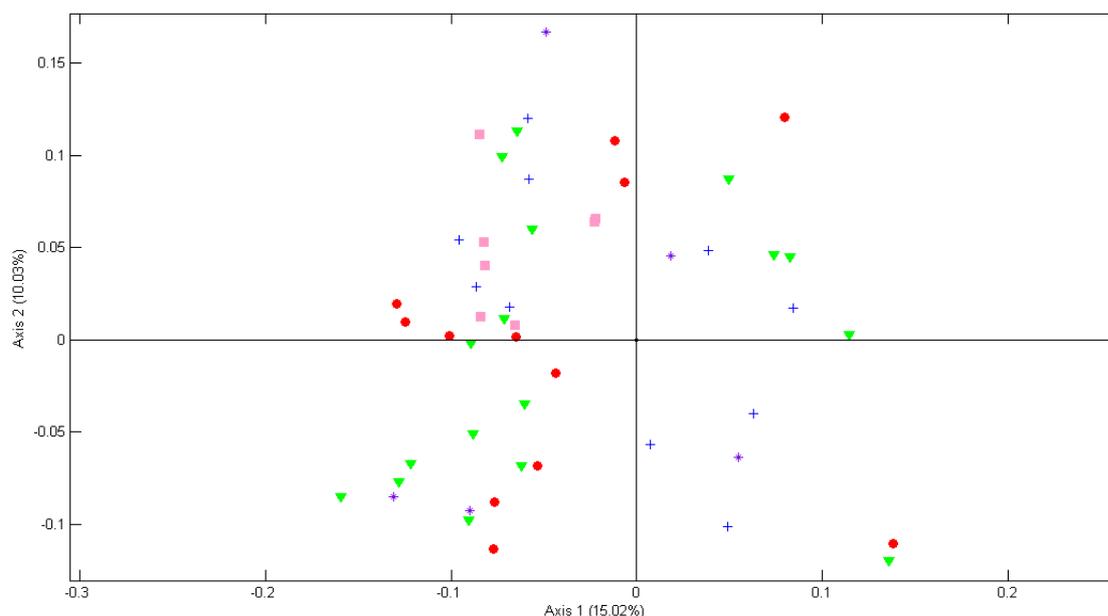


Figura 33: Análisis de coordenadas principales, que muestra la distribución de los 51 genotipos del set core, calculado a partir de la matriz de similitud con el coeficiente SM (Sokal y Michener, 1958) de 219 genotipos de cebada con 16 marcadores SSR (véase anexo I.4 para identificar cada individuo por su código).

CAPÍTULO 4: DISCUSIÓN

4.1. Protocolo de extracción, rendimiento y calidad del ADN genómico.

Existen varios kits comerciales que son capaces de extraer ADN genómico de tejido vegetal con suficiente calidad, pero el rendimiento de ADN producido es a menudo bajo (Xin & Chen, 2006), con la consideración además, que sus costos pueden convertirlos en inalcanzables para pequeños laboratorios.

En cuanto a rendimiento y calidad de los procesos caseros de extracción de ADN, Xin & Chen (2006) desarrollaron un protocolo en el que mejoraron estos parámetros en comparación con un kit comercial, obteniendo 40 veces mayor rendimiento, es decir, 49.8 ± 9.1 μg de ADN genómico total en hojas de sorgo y la calidad en términos de la relación de observancias $\text{OD}_{260\text{nm}}/\text{OD}_{280\text{nm}}$ fue de 1.88 ± 0.02 .

Pallotta *et al.* (2003) proponen un protocolo de extracción de ADN genómico de cebada y trigo, cuyo buffer de extracción se basa en el descrito por Struss & Plieske (1998). La innovación de los primeros autores en su protocolo radica en la el proceso de incubación en frío con acetato de amonio 6M después de la incubación en baño maría de las muestras con el buffer de extracción. Crouse & Amorese (1987) destacan al acetato de amonio como una alternativa al acetato de sodio en protocolos que empleen la precipitación del ADN con alcohol y lo describen como un eficiente agente removedor de proteínas al centrifugar las muestras previo a la adición del agente precipitante. Este aspecto del acetato de amonio puede ser visto como una ventaja, en cuanto al manejo que debe tener el operador cuando se emplean agentes removedores de proteínas como el cloroformo que es corrosivo, tóxico y un contaminante persistente.

Respecto al agente precipitante recomendado y acogido en el protocolo escogido que fue isopropanol, Edwards *et al.* (1991) en su protocolo de extracción de ADN genómico de plantas para análisis con PCR, emplearon una proporción de sobrenadante:isopropanol de 1:1 seguido de una incubación a temperatura ambiente por 5 minutos. En comparación con etanol, isopropanol es ideal en casos en los que se (i) requieran realizar procedimientos rápidos, debido a que necesita menor tiempo

de incubación, (ii) se cuente con gran cantidad de muestra, ya que se requiere de menos volumen. Sin embargo, debido a que las sales son menos solubles en isopropanol, precipitan en mayor cantidad con el pellet, por lo que fueron necesarios varios tratamientos de lavado con etanol 70% y 95% que redujeron la interferencia de las sales de acetato hasta niveles de pureza que permitieron la amplificación del ADN.

4.2. Marcadores microsatélites seleccionados.

Los marcadores SSR son considerados competitivos, en relación a otros marcadores neutrales, debido a que son multialélicos, altamente polimórficos, codominantes, altamente reproducibles y proveen información sobre riqueza genética con buena cobertura del genoma. Adicionalmente, son asequibles económicamente y dóciles para la mayoría de procedimientos de mejoramiento y por ende, aplicables en los programas públicos de mejoramiento que no sean capaces de solventar costosas técnicas para la evaluar diversidad genética (Turyagyenda *et al.*, 2012).

La reducción en el número de marcadores SSR que fueron empleados en comparación con el número inicial, se debió a distintas causas para los 5 marcadores descartados. El marcador Bmac0399 no presentó productos de amplificación, aun cuando fueron modificadas las condiciones PCR en cuanto a temperatura de alineamiento y coadyuvantes de reacción. El mencionado marcador en el estudio de Kolodinska-Brantestam *et al.* (2012) presenta alelos que van desde 123 hasta 147 pb y una diversidad genética de 0.79.

Una cuestión concerniente a los marcadores SSR es la posibilidad de que “alelos nulos” (Callen *et al.*, 1993; Pemberton, Slate, Bancroft, & Barrett, 1995) o no amplificantes sean reportados. Los individuos que son heterocigotos para un alelo nulo parecen ser homocigotos para el alelo visible, mientras que un individuo homocigoto para el alelo nulo corresponde a una reacción fallida (Ellis & Burke, 2007). Cuando están presentes en una población, estos alelos sesgan las frecuencias alélicas, reducen la heterocigosidad observada y por ende incrementan aparentes valores de consanguineidad (DeWoody, Nason, & Hipkins, 2006). En este sentido, los marcadores génicos EST-SSR presentan la ventaja de menor probabilidad de

encontrar alelos nulos ya que son obtenidos de secuencias relativamente mejor conservadas (Rungis *et al.*, 2004).

Fueron reportados y descartados tres marcadores debido a que presentaron una considerable proporción de alelos nulos: Bmag0135, EBmac0701 y Bmag0853. Los alelos nulos para marcadores SSR han sido descritos en trigo y cebada y pueden considerarse relativamente comunes en estas especies (Donini, Stephenson, Bryan, & Koebner, 1998). Stachel *et al.* (2000) atribuyeron este fenómeno al polimorfismo en la región flanqueante del microsatélite y consideraron que un alelo nulo representa uno de una serie de múltiples alelos. Las mencionadas variaciones en secuencia, impiden el alineamiento del primer durante la reacción de amplificación por PCR, resultando en un alelo nulo. Otras posibles causas incluyen la amplificación preferencial de pequeños fragmentos (debido a una inconsistente calidad o cantidad de templado de ADN) o slippage durante la reacción PCR (Gagneux, Boesch, & Woodruff, 1997; Shinde, Lai, Sun, & Arnheim, 2003).

Bolouri-Moghaddam, Safarnejad, & Kazemitabar (2010) detectaron alelos nulos para dos de los 10 loci SSR estudiados en cultivares de cebada iraní. Malysheva-Otto, Ganal, & Roder (2006) reportaron alelos nulos para 5 de los 48 marcadores SSR que estudiaron, mas no fueron excluidos de los análisis; entre los que se destaca el marcador Bmag0135, uno de los que reportaron alelos nulos y fue descartado en este estudio.

Algunas investigaciones han intentado corregir los datos perdidos ocasionados por los alelos nulos al asumir un nuevo alelo y sustituirlo por todos los que se reportan como nulos (Roques, Duchense, & Bernatchez, 1999), sin embargo estudios experimentales sugieren que los alelos nulos corresponden a alelos con diferentes tamaños.

En cuanto a los productos PCR del marcador Bmag0692 que resultaron en alelos de longitudes no reportadas y que superan las 400pb, es muy probable que algún un evento de mutación haya insertado elementos móviles en el microsatélite. Ha sido demostrado que en plantas con grandes genomas, los retrotransposones son la principal clase de ADN repetitivo (Boopathi, 2012). Los retrotransposones son elementos genéticos móviles encontrados en el reino vegetal. (Kumar & Bennetzen, 1999; Fedoroff, 2000). La asociación de los microsatélites genómicos de cebada a

retrotransposones ha sido advertida por Marcel (2007) y Ramsay *et al.* (1999) reportaron homología en las secuencias de marcadores SSR con elementos transponibles conocidos.

Una nueva clase de marcadores muy útiles son los denominados marcadores moleculares génicos (GMMs por sus siglas en inglés de Genic Molecular Markers) que utilizan la información de los crecientes archivos de secuencias de genes, acumulados en los proyectos de secuenciación de cDNA/EST (Expressed Sequence Tag) de un gran número de especies vegetales en los últimos años. Al ser parte de este tipo de secuencias, se espera que representen componentes funcionales del genoma como genes (Varshney *et al.*, 2007).

Se ha demostrado que los marcadores génicos pueden ser superiores a los marcadores genómicos para su uso en procesos de Selección Asistida por Marcadores, mapeo comparativo y la exploración de diversidad genética funcional en germoplasmas adaptados a distintos ambientes (Varshney *et al.*, 2007).

4.3. Determinación del número de individuos

La estrategia de extraer ADN a partir de un *pool* de individuos es significativa cuando se pretende acelerar el genotipaje de variedades criollas, las cuales debido a su más alta variabilidad genética requieren analizar más plantas por accesión para producir una evaluación más aproximada de la diversidad en un germoplasma de este tipo (Brondani *et al.*, 2006).

En cuanto al número y la estrategia idóneos de individuos que debieron ser escogidos por accesión, no existe un consenso, reportándose en estudios de diversidad genética en cebada desde 20 individuos separados por accesión (Bolouri-Moghandam *et al.*, 2011) y en *pool* (Struss & Plieske, 1998), 10 individuos (Hamza, Hamida, Rebai, & Harrabi, 2004), 8 individuos en *pool* (Sardou, Baghizadeh, Tavasoli, & Babaei, 2011) hasta 5-10 plántulas por accesión en *pool* (Malysheva-Otto *et al.*, 2006, Malysheva-Otto *et al.*, 2007).

Un enfoque que concuerda con el que fue escogido en el presente estudio es el reportado por Kolondinska-Brantestam *et al.* (2012), quienes hicieron una diferenciación de acuerdo al tipo de material de cebada; utilizando un mayor número

de individuos para variedades criollas y cultivares liberados hasta 1950 y tan solo un par de individuos en el caso de líneas promisorias y otros materiales de mejoramiento.

De hecho, en su investigación, Brondani *et al.* 2006 luego de analizar los perfiles de amplificación de los *pools* (5 individuos de arroz) de ADN, escogieron las muestras que presentaban heterocigosis en al menos dos marcadores y entonces realizaron una segunda amplificación esta vez para cada individuo.

4.4. Diversidad genética

La agricultura, hoy en día, se caracteriza por una reducción en la diversidad de plantas cultivadas. Tan sólo 30 de las más de 30.000 especie comestibles, son la base de la alimentación mundial. Esta disminución de la diversidad inter específica de cultivos debido a las prácticas agrícolas más la pérdida de la diversidad intra específica de las especies puede derivar en: epidemias de pestes y enfermedades, pérdida de adaptación a creciente tipos de estrés abiótico y pérdida de variación genética para rasgos específicos (Hausmann *et al.*, 2004).

Según la Convención de Biodiversidad de Rio de Janeiro de 1992, los países están comprometidos a conservar los recursos genéticos de sus cultivos de importancia. Existen más de 1300 colecciones de bancos de genes que conservan más de 6 millones de accesiones (FAO, 1996).

La diversidad genética es la base para el mejoramiento genético. El conocimiento de la diversidad que atañe a los germoplasmas tiene un significativo impacto en el mejoramiento de cultivos (Huang, Borner, Roder, & Ganal, 2002). La caracterización molecular es el medio predilecto para cuantificar la variación dentro de grandes números de muestras en germoplasmas. La precisión y robusteza de los patrones revelados por estos estudios radican en una pertinente elección de materiales, que implica un número suficiente de accesiones (Glaszmann, Kilian, Upadhyaya, & Varshney, 2010). Winter & Kahl (1995) recomiendan que para garantizar una mejor resolución en las clasificaciones se debe considerar la mayor cantidad de locus; escoger un bajo número de iniciadores represente un riesgo en la explicación de la variabilidad.

4.4.1. Diversidad genética de cebada

El número de alelos por locus que ha sido reportado en este trabajo fue de 9.81, en estudios similares de diversidad genética en cebada se han encontrado valores para este índice que varían dependiendo del número de materiales y marcadores empleados. Cuando se emplearon un mayor número de marcadores y accesiones se reportaron valores para este índice de que fueron desde 16.7 (Malysheva-Otto *et al.*, 2006), 10.65 (Yahiaoui, 2006) hasta 8.0 (Malysheva-Otto *et al.*, 2007). Mientras que, cuando se emplearon un menor número de marcadores y accesiones Chen *et al.* (2012) reportaron 6.0 alelos por locus. Un estudio que puede ser comparable al presente, en cuanto al número de datos generados, es el de Struss & Plieske (1998) que analizaron 163 líneas de cebada con 15 loci SSR y reportaron 8.6 alelos por locus.

En programas de monitoreo de conservación, algunos autores (Petit, El-Mousadik, & Pons, 1998; Barker, 1999) consideran a la riqueza alélica como el parámetro más relevante, ya que un mayor número de alelos implica una fuente de variación de loci para rasgos de importancia, además en términos de perspectiva largo plazo, el límite de respuesta a la selección está determinado por el número inicial de alelos (Hill & Rasbash, 1986).

Por otra parte, el parámetro más ampliamente usado para medir la diversidad genética dentro de poblaciones es la heterocigosidad esperada o diversidad genética de Nei (1973) (Toro & Caballero, 2005). El índice PIC está estrechamente relacionado con el índice de diversidad genética de Nei y pueden ser considerados equivalentes (Anderson, Churchill, Autrique, SD., & Sorrells, 1993). Los promedios para estos índices en el análisis de la colección completa fueron de 0.75 y 0.71 respectivamente.

En estudios similares, se han reportado índices de diversidad genética similares como en Fu & Horbach (2012) que analizaron 269 accesiones de cebada silvestre de una colección núcleo de germoplasma de Canadá con 25 marcadores SSR y reportan un índice PIC promedio de 0.75; Struss & Plieske (1998) que analizaron 163 líneas de cebada de distintos tipos con 15 loci SSR reportaron el valor de 0.74 de promedio para índice de diversidad genética y Malysheva-Otto *et al.* (2006) analizaron 953

accesiones de cebada, provenientes de la International Barley Core Collection 48 marcadores SSR un promedio de 0.74 en diversidad genética.

Los estudios que obtuvieron índices de diversidad genética inferiores están: Yahiaoui (2006) que en un estudio que consideró 225 individuos de líneas avanzadas de cebada de la colección nuclear española a nivel de 65 marcadores SSR encontró un índice la diversidad genética promedio de 0.66; Kolodinska-Brantestam *et al.* (2012) estudiaron 64 accesiones de cebada de primavera de países Bálticos con 21 marcadores SSR determinaron 0.56 el valor de la diversidad genética en promedio y Malysheva-Otto *et al.* (2007) cuando estudiaron 504 variedades de cebada europeas con 35 marcadores SSR detectaron promedio de valor PIC de 0.58.

El marcador Bmag0007 fue el más informativo de este estudio, de acuerdo a su número de alelos (14), al índice de polimorfismo PIC (0.83) y de diversidad genética de Nei (0.85), mientras que este mismo locus SSR en el estudio de Kolodinska-Brantestam *et al.* (2012) fue reportado el valor de 0.79 para el índice PIC de este marcador. Por otro lado, el menos informativo, Bmag0217 (PIC=0.48) según Wang *et al.* (2010) tuvo un índice PIC de 0.69.

El locus SSR Bmac0040 fue el marcador más polimórfico en estudio de cultivares europeos de Malysheva-Otto *et al.* (2007) y entre landraces y cultivares italianos según Bellucci *et al.* (2013). En el presente estudio este marcador fue el más polimórfico entre los materiales de mejoramiento y ocupó el segundo lugar entre los acriollados.

4.4.2. Diversidad genética comparativa

Las variedades criollas son una gran fuente de diversidad para mejoradores y aún representan una importante herramienta en el descubrimiento de genes y liberación de variedades. Una ventaja de los materiales acriollados puede ser explicada por su capacidad de amortiguamiento, donde sus individuos mejor adaptados compensan la pérdida de líneas menos adaptadas a condiciones severas de estrés biótico o abiótico (Hausmann *et al.*, 2004).

Grando, Ceccarelli, & Tekle (2000) mencionaron la considerable heterogeneidad existente entre variedades criollas de cebada colectadas de diferentes agricultores (y

conocidas con el mismo nombre) y entre los individuos de la misma accesión para varios caracteres.

En cuanto al posible origen de los materiales acriollados del Ecuador y la región, Jilal *et al.* (2011) concluyeron, en un estudio que incluyó variedades criollas de la región andina, que este material introducido pudo haber llegado específicamente desde África del este (Eritrea y Etiopía); es decir detectó un flujo de genes de cebada desde África a Sudamérica. Fueron analizadas 283 variedades criollas de distintas regiones de tres continentes: América del sur, África y Asia a nivel de 20 marcadores SSR (17 genómicos y 3 génicos). Entre los resultados que arrojaron se destacan un promedio para el índice PIC de 0.78 y 20.15 alelos por locus, adicionalmente determinaron que el grupo de variedades criollas sudamericanas fue el menos diverso de entre otras regiones del mundo; aunque fue el que tuvo menos representatividad.

En una comparación entre los materiales de Etiopía, basados en el estudio de (Abebe, 2010) que analizaron 199 accesiones de cebadas criollas con 15 marcadores SSR y los materiales acriollados ecuatorianos del presente estudio; se encontraron 3.86 y 8.75 alelos por locus y 0.43 y 0.69 promedio del índice PIC, respectivamente. El mencionado estudio coincide con el presente en que los marcadores Bmac0040 y Bmag0007 fueron, en este orden, los de mayor número de alelos e índice PIC.

No obstante, existen más semejanzas entre el presente estudio y el de Hadado, Rau, Bitocchi, & Papa (2010), quienes encontraron un promedio para el índice de diversidad genética de Nei de 0.66 y 9.4 alelos por locus, en un análisis de 106 variedades criollas etíopes con 7 loci SSR entre los que son comunes para ambos estudios los loci empleados Bmac0040 y Bmac0134.

En el estudio de Jilal *et al.* (2011) el marcador Bmac0316 fue el que mayor número de alelos detectó con 37. En el presente estudio, para los 156 genotipos correspondientes a 97 accesiones de cebada acriollada, el mismo marcador detectó 11 alelos, de los cuales cinco fueron determinados como alelos raros y dos como alelos únicos.

Varios estudios han demostrado cambios en la diversidad debido a programas de mejoramiento y han evidenciado una cierta disminución de la diversidad genética en los materiales modernos (Russel *et al.*, 2000; Matus & Hayes, 2002; Koebner *et al.*,

2003). Sin embargo, en un meta-estudio de Wouw *et al.* (2010) no se encontró una disminución significativa de la diversidad genética en variedades liberadas en el siglo XX de ocho diferentes cultivos que incluyeron a la cebada.

Struss & Plieske (1998) estudiaron la diversidad genética de varios tipos de germoplasma de cebada entre los que se incluyen variedades silvestres, criollas y cultivares y concluyeron que existió ligera mayor diversidad, de acuerdo al índice de diversidad genética de Nei, en los tipos silvestres y criollos que en los cultivares. Por otro lado, el estudio de Bellucci *et al.* (2013) hizo una comparación de la diversidad entre poblaciones de variedades criollas y cultivares de cebada italianas, y mostraron que la diversidad dentro de las primeras es comparable a la observada dentro de las segundas.

En cuanto a cultivares de cebada, Lamara *et al.* 2013 estudiaron 92 cultivares canadienses usando 50 loci SSR determinaron 7.36 alelos por locus y un promedio PIC de 0.69. Malysheva-Otto *et al.* (2006) en su estudio que incluyó 9 accesiones de cebada provenientes de Ecuador (probablemente cultivares), dentro de un grupo de 145 accesiones de América determinaron una diversidad genética para este grupo de 0.75. Este resultado pudo ser sobrevalorado para el caso de América del Sur, ya que considera un gran número de accesiones canadienses y norteamericanas.

Una desventaja de utilizar la riqueza alélica (número de alelos) como una medida de diversidad genética es que es altamente dependiente del número de la muestra: se espera que grandes muestras contengan más alelos que una pequeña. Un estadístico relacionado, el número de alelos raros, posee una definición que es muy arbitraria, en este estudio se consideró un alelo raro aquel con una frecuencia inferior a 0.05 y es bien conocido que el número de estos alelos también está afectado por el tamaño de la muestra (Chakraborty, Fuerst, & Nei, 1979).

El marcador Bmac0213 fue uno de los que detectó un mayor número de alelos raros en la colección completa, sin embargo, fue uno de los marcadores con menor valor de índice de diversidad genética de Nei (0.65) e índice PIC (0.63). Esto pudo explicarse ya que dado que los índices mencionados están estrechamente relacionados a los valores de las frecuencias alélicas, este marcador presentó una frecuencia para el alelo de 150pb de 0.56, una de las mayores reportadas. Chen *et al.*

(2012) y Guo *et al.* (2012) informaron un índice PIC para este locus de 0.79 y 0.76, respectivamente.

El número de alelos únicos en una población (riqueza alélica privativa) es una medida simple de distintividad genética (Kalinowski, 2004). Los alelos únicos son importantes porque detectan regiones del genoma que son específicas para un tipo de genotipo particular (Senior, Murphy, Goodman, & Stuber, 1998). La presencia de alelos únicos es un indicativo de la diversidad presente en un germoplasma y su potencial como reservorio de nuevos alelos para programas de mejoramiento (Matus & Hayes, 2002). Los resultados evidenciados en cuanto a alelos únicos, reveló los materiales acriollados presentaron mayor cantidad de estos, sin embargo; no existe una marcada diferencia entre los estos genotipos y los de la Colección Nacional respecto de este parámetro.

La alta variabilidad genética detectada, tanto entre materiales acriollados y mejorados, podría hacer que se descarte un enfoque de introgresión de genes desde las variedades criollas hacia los cultivares en futuros programas de mejoramiento; como lo recomienda Ram, Thiruvengadam, & Vinod (2007) en el germoplasma de arroz.

4.5. Duplicados

Los marcadores SSR pueden proveer un sistema óptimo para identificar materiales duplicados en una colección de germoplasma de cebada (Struss & Plieske, 1998) y son muy útiles en la protección y el manejo de dichas colecciones. Los 16 marcadores seleccionados en este estudio fueron capaces de detectar un total de 66 genotipos que compartían 100% de identidad a este nivel de loci analizados.

Según van Hintum & Künpffer (1995) pueden distinguirse varios tipos de duplicación: (i) la duplicación idéntica, cuando accesiones son genéticamente idénticas, (ii) la duplicación común que se refiere a aquellas accesiones derivadas de la misma población de parentales, (iii) duplicación parcial y compuesta donde no todos los alelos están duplicados y (iv) duplicación parental referida a las relaciones entre cruces, progenies y parentales.

Por razones de seguridad de las colecciones base, la duplicación es un procedimiento estándar en muchos bancos de genes, y por razones prácticas existe usualmente un alto grado de duplicación de materiales útiles entre colecciones activas. Por otra parte, la duplicación puede ser considerada inútil e incluso una causa de problemas en programas de investigación y mejoramiento (Hintum & Visser, 1995). Lo expuesto ratifica la utilidad del método en esta investigación, enmarcada en un programa de mejoramiento del germoplasma de cebada; sin embargo, también se destaca el aspecto de la conservación de las accesiones en los bancos de germoplasma.

4.6. Análisis multivariados.

En esta investigación se siguió la propuesta planteada por Vicente-Vallardón *et al.* (2006) y (Demey, 2008) del uso combinado del Análisis de Coordenadas Principales (AcoP), Análisis de Conglomerados (AC) y el Biplot Logístico Externo (BLE) para identificar los alelos responsables de la clasificación de los genotipos de cebada usando la información provista por marcadores moleculares SSR.

En cuanto al coeficiente de similitud escogido, no existe un criterio universal de cuando elegir uno u otro coeficiente, pero se acuerda en que la elección dependerá del peso que se desea dar a las frecuencias a , b , c y d (véase apartado 2.9.4.2), el tipo de datos y la situación experimental (Legèndre & Legèndre, 1979; Gower & Lejendre, 1986). Para el caso de marcadores codominantes, se pueden utilizar los coeficientes de similitud de Jaccard, Emparejamiento Simple, Dice o Roger cuando se codifique sobre alelos. El coeficiente de Emparejamiento Simple ha sido utilizado en estudios de marcadores moleculares SSR en varios organismos diploides (Demey, 2008).

En el contexto de estudios de diversidad genética en bancos de germoplasma se desconocen los modelos evolutivos específicos, por lo que las medidas de distancias genéticas entre poblaciones son de uso más restringido (Demey, 2008). El germoplasma de cebada del Ecuador analizado no constaba de poblaciones definidas o estructuradas, por lo que este tipo de análisis careció de importancia en este estudio.

4.6.1. Análisis de coordenadas principales

La representación de los individuos en dimensión reducida simplifica el análisis de dispersión y permite evidenciar posibles agrupamientos (Vicente-Villardón *et al.*, 2006). No se detectaron conglomerados de individuos clasificados de acuerdo a su tipo de germoplasma, sea este mejorado o acriollado, como tampoco –para el caso de los genotipos provenientes de colectas- de acuerdo a la provincia de su proveniencia.

En los análisis realizados, fueron escogidas las dos primeras dimensiones que recogían el 25.04% de la variación total para representar a todos los individuos en el plano y realizar además los análisis posteriores, ya que (Demey, 2008) corroboró la hipótesis de que dos dimensiones suelen ser suficientes para lograr buenas representaciones de los individuos.

El ACoP presenta además la ventaja metodológica, sobre el Análisis de Componentes Principales (ACP), que cada individuo es identificado únicamente por sus coordenadas principales y favorece así la generación de grupos de individuos a través del Análisis de Conglomerados (AC) (Gower & Harding, 1988). Estos grupos se correspondieron con los clusters formados bajo el algoritmo de conglomeración UPGMA tanto a partir de la matriz de similitud inicial como de la generada del ACoP.

Chae & Warde (2006) demostraron que la capacidad de recuperación de información se mejora usando las coordenadas principales de los individuos. Se esperaba, que para un mismo método de agrupación, la clasificación generada utilizando las coordenadas principales sea similar o superior a la generada usando la matriz de distancia observada. Sin embargo; en este estudio, el coeficiente de correlación cofenética del análisis de conglomerados en base a las coordenadas principales usando el algoritmo UPGMA, evidenció un valor inferior al obtenido sobre las similaridades originales.

4.6.2. Análisis de conglomerados

El agrupamiento de individuos con patrones moleculares similares y su representación en el espacio en un sistema de coordenadas son las dos grandes

técnicas multivariantes de uso general en la visualización de la diversidad genética (Demey, 2008). Para la caracterización usando marcadores moleculares, ya sean dominantes o codominantes, la técnica de clasificación mayormente usada es el dendrograma generado por el algoritmo UPGMA.

Sin embargo, el problema con este análisis es que puede generar o identificar grupos que no existen en forma natural (Everitt, 1979; Everitt, Landau, & Leese, 2001). Es el caso del algoritmo UPGMA que introduce un error por la adecuación de una distancia no ultramétrica a un árbol ultramétrico. Un procedimiento que valida estas clasificaciones, es el coeficiente de correlación cofenética que compara la matriz de distancias original y la derivada del dendrograma (Sokal & Rohlf, 1962). El coeficiente que resultó de la matriz original de distancias fue superior a 0.90 que es el recomendado para buenas interpretaciones.

Los análisis moleculares de varianza llevados a cabo permitieron establecer la configuración de número de clusters (diez) que revela la mayor variación detectable entre y dentro de ellos. En el caso de los clusters ordinales, la primera pudo ser explicada al reanalizar los conglomerados más alejados del origen y mostrar una mayor diferenciación entre ellos; mientras que la segunda pudo ser explicada ya que se obtuvieron los mayores índices de diversidad genética en los clusters más cercanos al origen del plano de variación.

4.7. Análisis de remuestreos

Ha sido demostrada la capacidad de los métodos de remuestreo tales como el Jackknife (Tukey, 1958) y el Bootstrap (Efron & Tibshirani, 1993) para determinar la estabilidad de configuraciones generadas por técnicas multivariantes. De estas técnicas el Bootstrap total es la mejor estrategia metodológica para construir regiones de confianza de los n elementos. Además ha sido revelado que el análisis Procrustes (Bootstrap total tipo 3) muestra los mejores resultados de estabilidad y sensibilidad al efecto configuración de referencia (Demey, 2008) por lo que demostró un muy buen ajuste (0.99) y errores estándar mínimos.

A menudo, los ejes de las coordenadas principales tienen escalas disímiles que no permiten la comparación o representación en el mismo sistema de coordenadas,

sin antes realizar alguna transformación. Esto se debe a que el Análisis de Coordenadas Principales (ACoP) las direcciones de las dimensiones que se generan son arbitrarias y dependen estrechamente del tipo de atributo y de las escalas de medición, entre otros (Demey, 2008). La estrategia metodológica, que permitió corregir los problemas mencionados fue la generación de configuraciones de consenso de las configuraciones generadas por los remuestreos.

4.8. Biplot logístico externo.

Los métodos biplot son enfoques que facilitan la interpretación genética, comparada con las clásicas técnicas de clasificación u ordenación. (Chapman, Schenk, Kazan, & Manners, 2002; Gabriel, 1971; Sharov, Dudekula, & Ko, 2005). En el caso del BLE, este favorece el estudio simultáneo de las relaciones entre individuos, individuos-variables y variables-variables e incrementa la calidad de información presentada en estudios de diversidad genética (Demey, 2008).

Cuando se trata de datos binarios, como los obtenidos del análisis de información molecular, un biplot lineal clásico y un análisis de componentes principales (ACP) no son factibles, mientras que el enfoque BLE descrito por Vicente-Villardón *et al.* (2006) representa una solución alternativa.

Cuando se empleó un valor para la medida de calidad pseudo $R^2 > 0.8$ y se mantuvo el valor de $p \leq 0.05$ para la corrección de Bonferroni en este estudio se determinaron los tres alelos más representativos: Bmac0127(124pb), Bmac0213(170pb), EBmac0501(125pb); no obstante cuando se manejan grandes conjuntos de datos, los p valores son afectados considerablemente por el tamaño de muestra y el número de alelos. En estos casos Demey, Vicente-Villardón, Galindo-Villardón, & Zambrano (2008) recomiendan utilizar un pseudo R^2 con un valor altamente restrictivo porque es menos sensible al tamaño de muestra.

4.9. Set core

A la fecha, la estrategia M es claramente la función más poderosa para seleccionar accesiones con la mayor diversidad alélica y eliminar redundancia inherente a alelos no informativos (Franco, Crossa, Warburton, & Taba, 2006).

Se ha reportado el uso de la estrategia M en el establecimiento de colecciones núcleo. Muzzalupo, Vendramin, & Chiappetta (2014) determinaron una colección núcleo de 23 de 489 accesiones de aceituno (*Olea europaea*) con información molecular de 11 loci SSR. Santos-García *et al.* (2012) seleccionaron 23 y 13 accesiones de entre 134 y 192 accesiones de *S. macrocephala* y *S. capitata* respectivamente al analizarlas con 13 marcadores SSR. Los estudios anteriores priorizan sus resultados al basarse en el programa COREFINDER.

POWERCORE propone sobre la estrategia M una metodología heurística que aplica una función de evaluación para determinar la vía de búsqueda más corta en comparación con búsquedas completamente aleatorias (Kim *et al.*, 2007). Además, este programa difiere del anterior en que analiza los perfiles de los haplotipos de cada individuo.

En cuanto a cebada, se ha reportado el proyecto The International Barley Core Collection Project (Knupffer & Hintum, 1995) que estableció una colección núcleo de alrededor de 2000 accesiones, entre las que se destacan 500 cultivares y 800 landraces, a partir de 373.000 accesiones de cebada. También existe el caso de la colección núcleo de cebadas españolas, establecida por Igartua *et al.* (1998) y comprende un set de 160 entradas a partir de 2000 accesiones de cebadas españolas.

Brondani *et al.* (2006) demostraron que una manera de seleccionar genotipos distintivos de una colección de germoplasma es elegir accesiones representativas de cada cluster formado, y complementar la decisión basándose en los individuos con los mejores rasgos agronómicos de interés. En general, estudios que incorporen descriptores morfológicos y marcadores moleculares proveerán una mejor descripción e interpretación de la diversidad genética (Wilson, Sarich, & Maxson 1974; Hillis & Wiens, 2000; Demey, Zambrano, Fuenmayor, & Segovia, 2003).

CAPÍTULO 5: CONCLUSIONES

- El protocolo de extracción de ADN genómico de cebada desarrollado por Pallotta *et al.* (2003) que fue empleado en este estudio evidenció aceptables rendimientos, un producto extraído cualitativamente adecuado para su amplificación con marcadores moleculares SSR y un desempeño amigable con el ambiente.
- Se comprobó, al nivel molecular del estudio, que el 60.8% de las accesiones de cebada acriollada del Banco de Germoplasma y de la colecta Austro-Imbabura fueron genéticamente heterogéneas.
- La metodología empleada para la selección de los individuos de las accesiones heterogéneas permitió evidenciar una mayor diversidad genética demostrada en los respectivos índices.
- Los análisis de diversidad genética mostraron que al estudiar 219 distintos genotipos de 160 variedades de cebada a nivel de 16 marcadores moleculares SSR fue posible detectar una riqueza alélica de 9.81 alelos por locus y un valor promedio de 0.715 para el índice de polimorfismo PIC de los marcadores.
- Los materiales de mejoramiento fueron el tipo de germoplasma en el que se reportó mayor diversidad genética, específicamente en las variedades de la colección nacional del PNC; sin embargo, en los materiales acriollados se comprobó la existencia de un mayor número de alelos únicos.
- El 30.14% de la colección completa comprendió materiales identificados como duplicados genéticos, agrupados en 17 clases distintas que contuvieron desde 2 hasta 10 genotipos por clase.
- Los análisis multivariados mostraron que no existió una relación de agrupación de los genotipos de cebada analizados de acuerdo al tipo de germoplasma o, de ser el caso, a la provincia en el que fueron recolectados.
- El Análisis de conglomerados bajo el algoritmo UPGMA sobre la matriz de coordenadas principales reflejó una mejor conformación de diez grupos o

clusters entre los que se distinguió una alta diferenciación genética que fue comprobada con los análisis AMOVA y estadísticos F.

- La diferenciación genética fue incluso superior cuando se analizaron por separado los clusters más alejados del origen del plano de variación, mientras que la gran variación genética encontrada dentro de los clusters más próximos al origen fue explicada por los altos índices de diversidad genética que fueron reportados para estos grupos.
- Mediante el análisis del biplot logístico en base al ACoP se determinaron 27 alelos representativos ($R^2 \geq 0.5$, $p < 0.05$) pertenecientes a 14 de los 16 loci SSR analizados.
- Se estableció un set núcleo de 51 genotipos de cebada, en el que se comprobó una riqueza alélica y diversidad genética equivalentes a las de la colección completa.
- La teoría estadística concerniente al tratamiento de datos moleculares es la base para el desarrollo de programas informáticos, y el conocimiento y manejo de los anteriores constituyen las herramientas que un investigador debe priorizar en la interpretación de resultados de un estudio.
- A la luz de los resultados expuestos, se puede afirmar que existe una gran diversidad genética inherente a la colección de cebada del INIAP; y este estudio puede ser considerado un insumo en actividades de fitomejoramiento y de conservación de accesiones en el Banco de Germoplasma.

CAPÍTULO 6: RECOMENDACIONES

- Realizar la estandarización de las condiciones de genotipaje y la validación de los marcadores SSR con una mayor cantidad de muestras de ADN de cebada (más de 30) para descartar posibles problemas de amplificación inherentes al marcador o a los productos de amplificación.
- Tomar en cuenta el tipo de germoplasma que se pretenda estudiar genéticamente, y adoptar una metodología que no descarte del análisis posibles mezclas de genotipos en variedades criollas especialmente.
- Incorporar en próximos estudios de diversidad genética controles positivos, es decir, muestras de ADN de variedades de un cultivo dado de las que se tengan reportes de sus perfiles genéticos a nivel de los marcadores a analizarse.
- Realizar análisis de las secuencias de los productos amplificados de los marcadores Bmag0692, Bmag0853, Bmag0135 y EBmac0701; en especial del primero ya que este mostró amplicones de longitudes no reportadas.
- Correlacionar estadísticamente los datos moleculares de este estudio con los de su caracterización morfoagronómica y establecer los materiales representativos tanto para programas de mejoramiento como para estrategias de conservación en bancos de genes.

BIBLIOGRAFÍA

- Abadie, T., & Berreta, A. (2001). FAGRO. *Caracterización y evaluación de Recursos Fitogenéticos*. Obtenido de http://www.fagro.edu.uy/~fitotecnia/docencia/materiales%20apoyo/Caracterizacion_y_Evaluacion_de_Recursos_Fitogeneticos.pdf
- Abebe, T. (2010). Genetic diversity and population differentiation analysis of Ethiopian barley (*Hordeum vulgare* L.) landraces using morphological traits and SSR markers.
- Ajmone-Marsan, P., Negrini, R., Crepaldi, P., Milanesi, E., Gorni, C., Valentini, A., & Cicogna, M. (2001). Assessing genetic diversity in italian goat populations using AFLP markers. *Animal Genetics*, 32, 281-288.
- Alstad, D., Hotchkiss, S., & Corbin, K. (1991). Gene flow estimates implicate selection as a cause of scale insect population structure. *Evol. Ecol.*, 5, 88-92.
- Anderson, J., Churchill, G., Autrique, J., SD., T., & Sorrells, M. (1993). Optimizing parental selection for genetic linkage maps. . *Genome*, 36:181-186.
- Arias, G. (1995). Estructura, rendimiento y calidad de las cebada brasileñas.
- Bailo, A., & A., G. (2008). *100 Problemas Resueltos de Estadística Multivariante*. Madrid: Delta.
- Barker, J. (1999). Conservation of livestock breed diversity. *AGRI*, 25,33-43.
- Becerra, V., & Paredes, M. (2000). Uso de Marcadores Bioquímicos y Moleculares en Estudios de Diversidad Genética. *Agricultura Técnica*, 60(3), 270-281. Obtenido de <http://www.inia.cl/at/espanol/v60n3/html/art07.htm>.
- Beeching, J., Marmey, P., Hughes, M., & Charrier, A. (1994). Evaluation of molecular approaches for determining genetic diversity in Cassava germplasm. *Proc. 2nd Internat. Scient Meet*, (págs. 22-26). Bogor-Indonesia.
- Bellucci, E., Bitocchi, E., Rau, D., Nanni, L., & Ferradini, N. (2013). Population Structure of Barley Landrace Populations and Gene-Flow with Modern Varieties. *PIOS ONE*, 8(12).

- Bolouri-Moghaddam, M., Safarnejad, A., & Kazemitabar, S. (2011). Genetic diversity assessment in several barley (*Hordeum vulgare* L.) cultivars using microsatellite markers. *Notulae Scientia Biologicae*, 3(2):140-144.
- Boopathi, N. (2012). *Mapping and Marker Assisted Selection*. Springer-Verlag Gmdh Dez.
- Bothmer von, R., Hintum van, T., Kunpffer, H., & Sato, K. (2003). *Diversity in barley (Hordeum vulgare)*. Amsterdam, The Netherlands: Elsevier B.V.
- Botstein, D., White, R., Skolnik, M., & Davis, R. (1980). Construction of a genetic linkage map using restriction fragment length polymorphisms. *Am J Hum Genet*, 32:314-331.
- Briggs, D. (1978). *Barley*. London: Chapman Hall.
- Broers, L. (1994). Resistencia duradera en cultivos alto andinos. *Memorias del primer taller sobre resistencia duradera en cultivos alto andino de Bolivia, Perú, Colombia y Ecuador* (págs. 44-45). Quito-Ecuador: INIAP-WAU-DGIS Editores.
- Brondani, C., Oliveira, T., Nakano, P., Pereira, R., & Brondani, V. (2006). Determination of genetic variability of traditional varieties of Brazilian rice using microsatellite markers. *Genetics and Molecular Biology*, 29(4):676-684.
- Callen, D., Thompson, A., Shen, Y., Phillips, H., Richards, R., Mulley, J., & Sutherland, G. (1993). Incidence and origins of “null” alleles in (AC)n microsatellite markers. *Am J Hum Genet*, 52:922-927.
- Carrera, M., & Mateo, J. (2005). *Prontuario de agricultura*. Madrid: Mundi-Prensa Libros.
- Castro, A., Viega, L., Pritsch, C., Hoffman, E., Hayes, P., & Gomez, B. (2010). Caracterización genómica del germoplasma de cebada, por variables de calidad maltera, agronómicas y sanitarias. *Universidad de la República de Paraguay*.

- Ceccarelli, S. (1996). Adaptation to low/high input cultivation. *Euphytica*, 92: 203-214.
- Chae, S., & Warde, W. (2006). Effect of using principal coordinates and principal. *Computational Statistics and Data Analysis*, 50:1407-1417.
- Chakaraborty, R., Fuerst, P., & Nei, M. (1979). Statistical studies on protein polymorphism in natural populations. III. Distribution of allele frequencies and the number of alleles per locus. *Centerfor Demographic and Population Genetics*.
- Chakraborty, R., & Danker-Hopfe, H. (1991). Analysis of population structure: A comparative analysis of different estimators of Wright's fixation indices. En C. Rao, & R. Chakraborty, *Handbook of Statistics: Statistical Methods in Biological and Medical Sciences* (Vol. 8, págs. 203-254). Elsevier.
- Chapman, S., Schenk, P., Kazan, K., & Manners, J. (2002). Using biplots to interpret gene expression patterns in plants. *Bioinformatics*, 18,202-204.
- Chen, Z., Lu, R., Zou, L., Du, Z., Gao, R., He, T., & Huang, J. (2012). Genetic diversity analysis of barley landraces and cultivars in the Shanghai region of China. *Genet. Mol. Res.*, 11(1):644-650.
- Chistiakov, D., Hellemans, B., & Volckaert, F. (2006). Microsatellites and their genomic distribution, evolution, function and applications: a review with special reference to fish genetics. *Aquaculture*, 255, 1-29.
- Coronel, J., & Jiménez, C. (2011). Guía Práctica para los productores de cebada de la Sierra Sur. En E. E. Austro, *Boletín divulgativo N°404*. Cuenca-Ecuador.
- Crouse, J., & Amorese, D. (1987). Ethanol Precipitation: Ammonium Acetate as an Alternative to Sodium Acetate. *Focus* 9, 3-5.
- Demey, J. (2008). Diversidad genética en bancos de germoplasma: un enfoque biplot. *Tesis Doctoral. Universidad de Salamanca*.

- Demey, J., Vicente-Villardón, J., Galindo-Villardón, M., & Zambrano, A. (2008). Identifying molecular markers associated with classification of genotypes by External Logistic Biplots. *Bioinformatics*, 1-7.
- Demey, J., Zambrano, A., Fuenmayor, F., & Segovia, V. (2003). Relación entre caracterizaciones molecular y morfológica en una colección de Yuca. *Interciencia*, 28(12):1-7.
- DeWoody, J., Nason, J., & Hipkins, V. (2006). Mitigating scoring errors in microsatellite data from wild populations. *Mol Ecol Notes*, 6:951-957.
- Dillon, W., & Goldstein, M. (1984). *Multivariate analysis. Methods and applications*. New York: John Wiley and Sons.
- Donini, P., Stephenson, P., Bryan, G., & Koebner, R. (1998). The potential of microsatellites for high throughput diversity assessment in wheat and barley. *Genet Res Crop Evol*, 45:415-421.
- Edwards, K., Johnstone, C., & Thompson, C. (1991). A simple and rapid method for the preparation of genomic plant DNA for PCR analysis. *Nucleid Acid Research*, 19: 1349.
- Efron, B., & Tibshirani, R. (1993). *An introduction to the bootstrap*. New York: Capman and Hall.
- Ellegren, H. (2004). Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics*, 5, 435-445.
- Ellis, J., & Burke, J. (2007). EST-SSRs as a resource for population genetic analyses. *Heredity*, 99,125-132.
- Escribano, P., Viruel, M., & Hormaza, J. (2008). Comparison of different methods to construct a core germplasm collection in woody perennial species with simple sequence repeat markers. A case study in cherimoya (*Annona cherimola*, Annonaceae), an underutilised subtropical fruit tree species. *Ann Appl Biol.*, 153:25-32.

- Estrella, J., Manosalvas, R., Mariaca, J., & Ribadeneira, M. (2005). Biodiversidad y Recursos Genéticos: Una guía para su uso y acceso en el Ecuador. (Ecociencia, INIAP, MAE, & A. Yala, Edits.)
- Everitt, B. (1979). Unresolved problems in cluster analysis. *Biometrics*, 35:169-181.
- Everitt, B., Landau, S., & Leese, M. (2001). *Cluster analysis* (4th edition ed.). Oxford, UK: Hodder Arnold.
- Excoffier, L., & Lischer, H. (2010). Arlequin suite version 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour*, 10:564-567.
- Excoffier, L., Smousse, P., & Quattro, J. (1992). Analysis of molecular variance inferred from metric distance among DNA haplotypes Applications to human mitochondrial DNA restriction data. *Genetics*, 131,479-491.
- FAO. (1 de Marzo de 2014). *FAOSTAT*. Obtenido de <http://faostat.fao.org>
- Fedoroff, N. (2000). Transposons and genome evolution in plants. *Proc. Natl. Acad. Sci.*, 97:7002-7007.
- Ferreira, M., & Grattapaglia, D. (1998). *Introducción al uso de marcadores Moleculares en el Análisis Genético*. Brasilia-Brasil: EMBRAPA-CENARGEN.
- Flavell, R., Bennett, M., Smith, J., & Smith, D. (1974). Genome size and the proportion of repeated nucleotide sequence DNA in plants. *Biochem Genet*, 12:257-269.
- Forster, B., Ellis, R., Thomas, W., Newton, A., Tuberosa, R., This, D., . . . Ben Salem, M. (2000). The development and application of molecular markers for abiotic stress tolerance in barley. *Journal of Experimental Botany*, 51:000-000.
- Franco, J., Crossa, J., Warburton, M., & Taba, S. (2006). Sampling strategies for conserving maize diversity when forming core subsets using genetic markers. *Crop Sci.*, 46,854-864.

- Frankel, O., & Brown, A. (1984). Plant genetics resources today: a critical appraisal. En J. Holden, & J. Williams, *Crop Genetic Resources: conservation & evaluation* (págs. 249-257). Londres: George Allen & Unwin .
- Fu, Y., & Horbach, C. (2012). Genetic Diversity in A Core Subset of Wild Barley Germplasm. *Diversity*, 4,239-257.
- Gabriel, K. (1971). The biplot - graphic display of matrices with applications to. *Biometrika*, 58,453-467.
- Gagneux, P., Boesch, C., & Woodruff, D. (1997). Microsatellite scoring errors associated with noninvasive genotyping based on nuclear DNA amplified from shed hairs. *Mol. Ecol.*, 6: 861-868.
- Glaszmann, J., Kilian, B., Upadhyaya, H., & Varshney, R. (2010). Accessing genetic diversity for crop omprovement. *Current Opinion in Plant Biology*, 13:1-7.
- Gouesnard, B., Bataillon, T., Decoux, G., Rozale, C., Schoen, D., & David, J. (2001). MSTRAT: An algorithm for building germplasm core collections by maximizing allelic or phenotypic richness. *The Journal of Heredity*, 92,93-94.
- Gower, & Harding, S. (1988). Nonlinear biplots. *Biometrika*, 75(3):445-455.
- Gower, J. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, 53:325-338.
- Gower, J., & Dijksterhuis, G. (2004). *Procrustes Problems*. Londres: Chapman and Hall.
- Gower, J., & Lejendre, P. (1986). Metric and Euclidean Properties of Dissimilarity. *Journal of Classification*, 3:5-48.
- Grando, S., Ceccarelli, S., & Tekle, B. (2000). Diversity in Barley landraces from the Near East. *8th International Barley Germplasm Genetics Symposium*, (págs. 13-15). Adelaide, South Australia.

- Graner, A., Jahoor, A., Schondelmaier, J., Siedler, H., Pillen, K., Fischbeck, G., . . . Herrmann, R. (1991). Construction of an RFLP map of barley. *Theor. Appl. Genet.*, 83, 250-256.
- Guerrero, A. (1999). *Cultivos herbáceos extensivos* (Sexta ed.). Mundi-Prensa Libros.
- Guo, Y., Li, Y., Huang, Y., Jarvis, D., Sato, K., Kato, K., . . . Long, C. (2012). Genetic diversity analysis of hullless barley from Shangri-la region revealed by SSR and AFLP markers. *Genet Resour Crop Evol*, 59:1543-1552.
- Hadado, T., Rau, D., Bitocchi, E., & Papa, R. (2010). Adaptation and diversity along an altitudinal gradient in Ethiopian barley (*Hordeum vulgare* L.) landraces revealed by molecular analysis. *BMC Plant Biology*, 10:121.
- Hamrit, S., Heckenberger, M., Huan, X., Van-Kaauwen, M., Kochieva, E., Koebner, R., . . . Vosman, B. (2004). Changes over time in the genetic diversity of four major European crops: a report from the Gediflux . *Proceedings of the XVIIth EUCARPIA General Congress*, (págs. 3-7). Tulln Australia.
- Hamza, S., Hamida, W., Rebai, A., & Harrabi, M. (2004). SSR-based genetic diversity assessment among Tunisian winter barley and relationship with morphological traits. . *Euphytica*, 135:107-118.
- Harlan, J. (1992). Crops and man. *American Society of Agronomy*.
- Hausmann, B., Parzies, H., Presterl, T., Susic, Z., & Miedaner, T. (2004). Plant genetic resources in crop improvment. *Plant Genetic Resources*.
- Hill, W., & Rasbash, J. (1986). Models of long term artificial selection in finite populations. *Genet. Res.*, 48,41-50.
- Hillis, D., & Wiens, J. (2000). Molecules versus morphology in systematics: Conflicts, artifacts, and misconceptions. En J. Wiens, *Phylogenetic analysis of morphological data* (págs. 1-19). Washington D.C.: Smithsonian Intitution Press.

- Hintum, T. v., & Knupffer, H. (1995). Duplication within and between germplasm collections I: Identifying duplication on the basis of passport data. *Genetic Resources and Crop Evolution*, 24(2):127-133.
- Hintum, T. v., & Visser, D. (1995). Duplication within and between germplasm collections. II. Duplication in four European Barley collections. *Genet. Resources and Crop Evol.*, 42:135-145.
- Huang, X., Borner, A., Roder, M., & Ganal, M. (2002). Assessing genetic diversity of wheat (*Triticum aestivum* L.) germplasm using microsatellite markers. *Theor Appl Genet*, 105:699-707.
- Igartua, E., Gracia, M., Lasa, J., Medina, B., Molina-Cano, J., Montoya, J., & Romagosa, I. (1998). The Spanish Barley Core Collection . *Genet. Res. Crop. Evol.*, 129:45-52.
- INEC, I. N. (2011). Encuesta de Superficie y Producción Agropecuaria.
- INIAP, I. N. (2005). Inventario Tecnológico del Programa de Cereales.
- Jilal, A., Grando, S., Henry, R., Lee, S., Rice, N., Hill, H., . . . Ceccarelli, S. (2011). Genetic diversity of ICARDA's worldwide barley landrace collection. *Genet Resour Crop Evol.*
- Kalinowski, S. (2004). Counting alleles with rarefaction: Private alleles and hierarchical sampling designs. *Conser Genet*, 5,539-543.
- Kim, K., Chung, H., Cho, G., Ma, K., Chandrabalan, D., Gwag, J., . . . Park, Y. (2007). PowerCore: A program applying the advanced M strategy with a heuristic search for establishing core sets. *Bioinformatics*, 23:2155-2162.
- Kimura, M., & Crow, J. (1964). The number of alleles that can be maintained in a finite population. *Genetics*, 49:725-738.
- Koebner, R., Donini, P., Reeves, J., Cooke, R., & Law, J. (2003). Temporal flux in the morphological and molecular diversity of UK barley. . *Theoretical and Applied Genetics*, 106:550-558.

- Kolodinska-Brantestam, A., Rashal, I., Tuveesson, S., Weibull, J., & von Bothmer, R. (2012). Genetic profiles and diversity of baltic spring barley material. *Proceedings of the Latvian Academy of Sciences*, 66, 10-20.
- Komatsuda, T., Pourkheirandish, M., He, C., Azhaguvel, P., Kanamori, H., Perovic, D., . . . Yano, M. (2007). Six-rowed barley originated from mutation in a homeodomain-leucine zipper I-class homeobox gen. *Proc Natl Acad Sci*, 104:1424-1429.
- Krzanowski, W. (2000). *Pinciples of Multivariate Analysis: A User's Perspective* . Oxford: Oxford University Press.
- Kumar, A., & Bennetzen, J. (1999). Plant retrotransposons. *Annu. Rev. Genet.*, 33:479-532.
- Kumpatla, L., & Mukhopadhyay, S. (2005). Mining and survey of simple sequences repeats in expressed sequence tags of dicotyledonous species. *Genome*, 48, 985-998.
- Kunpffer, H., & van Hintum, T. (1995). The Barley Core Collection - an international effort. En T. Hodking, A. Brown, T. van Hintum, & E. Morales, *Core Collections of Plant Genetic Resources* (págs. 171-178). John Wiley and Sons.
- Lamara, M., Zhang, L., Marchand, S., Tinker, N., & Belzile, F. (2013). *Comparative analysis of genetic diversity in Canadian barley assessed by SSR, DarT, and pedigree data*.
- Lebart, L. (2007). Which Bootstrap For Principal Axes Methods . En P. Brito, G. Cucumel, P. Bertrand, & F. de Carvalho, *Selected Contributions in Data Data Analysis and Classification* (págs. 581-588). Berlin : Springer.
- Legèndre, L., & Legèndre, P. (1979). *Ecologie Numérique*. Masson.
- León, D. (2010). Evaluación del rendimiento de dos variedades mejoradas y una tradicional de cebada (*Hordeum vulgare* L.) en Tunshi, parroquia Licto,

cantón Riobamba, provincia de Chimborazo. *Tesis de Ingeniería. Facultad de Recursos Naturales. Escuela de Ingeniería Agronómica. ESPOCH.*

- Linde-Laursen, I., Heslop-Harrison, J., Shepherd, K., & Taketa, S. (1997). The barley genome and its relationship with the wheat genomes. A survey with an internationally agreed recommendation for barley chromosome nomenclature. *hereditas*, 126:1-16.
- Liu, B. (1998). *Statistical Genomics: Linkage, mapping and QTL analysis. CRC Press.*
- Liu, K., & Muse, S. (2005). Power Marker: Integrated analysis environment for genetic marker data. *Bioinformatics*, 21(9), 2128-2129.
- Long, J. (1997). *Regression Models for Categorical and Limited Dependent Variables.* Londres: Sage Publications.
- López, J., & Hidalgo, M. (1994). Análisis de Conglomerados. En M. Alto, & L. JJ., *Fundamentos de estadística en Systat* (págs. 505-532). Addison Wesley Iberoamericana.
- Malysheva-Otto, L., Ganal, M., & Roder, M. (2006). Analysis of molecular diversity, population structure and linkage disequilibrium in a worldwide survey of cultivated barley germplasm (*Hordeum vulgare* L.). *BMC Genet*, 7, 6.
- Malysheva-Otto, L., Ganal, M., Law, J., Reeves, J., & Röder, M. (2007). Temporal trends of genetic diversity in European barley cultivars (*Hordeum vulgare* L.). *Mol. Breed.*, 20:309-322.
- Marcel, T. (2007). Genetic architecture of basal resistance of barley to *Puccinia hordei*. *PhD. Thesis. Wageningen University. The Netherlands.*
- Marcel, T., Varshney, R., Barbieri, M., Jafary, H., de Kock, M., Graner, A., & Niks, R. (2007). A high-density consensus map of barley to compare the distribution of QTLs for partial resistance to *Puccinia hordei* and of defence gene homologues. *Theoretical and Applied Genetics*, 487-500.
- Matus, I., & Hayes, P. (2002). Genetic diversity in three groups of barley germplasm assessed by simple sequence repeats. *Genome*, 45(6):1095-1106.

- Milan, L., & Whittaker, J. (1995). Application of the Parametric Bootstrap to Modelsthat Incorporate a Singular Value Decomposition. *Applied Statistics*, 44(I):31-49.
- Molina-Cano, J., Russell, J., Moralejo, M., Escacena, J., Arias, G., & Powell, W. (2005). Chloroplast DNA microsatellite anlysis supports a polyphyletic origin for barley. *Theoretical Applied Genetics*, 110: 613-619.
- Morillo, E., & Miño, G. (2011). Marcadores Moleculares en Biotecnología Agrícola: Manual de procedimientos y técnicas en INIAP. *Manual No. 91 INIAP Estación Experimental Santa Catalina*.
- Morrel, P., & Clegg, M. (2007). Genetic evidence for a second domestication of barley (*Hordeum vulgare*) east of the Fertile Crescent. *Proc Natl Acad Sci*, 104:3289-3294.
- Muzzalupo, I., Vendramin, G., & Chiappetta, A. (2014). Genetic Biodiversity of Italian Olives (*Olea europaea*) Germplasm Analyzed by SSR Markers. *The Scientific World Journal*.
- Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences*, 70: 3312-3323.
- Pallotta, M., Warner, P., Fox, R., Kuchel, H., Jefferies, S., & Langridge, P. (2003). Marker assisted wheat breeding in the southern region of Australia. *Proceedings of the Tenth International Wheat Genetics Symposium*, 789-791.
- Park, S. (2001). Trypanotolerance in West African cattle and the population genetic effect of selection. *PhD Thesis. University of Dublin*.
- Peakall, R., & Smouse, P. (2012). Gen AIEx 6.5: Genetic Analysis in Excel. *The Australian National University*.
- Pemberton, J., Slate, J., Bancroft, D., & Barrett, J. (1995). Nonamplifying alleles at microsatellite loci: a caution for parentage and population studies. *Mol Ecol*, 4:249:252.

- Petit, R., El-Mousadik, A., & Pons, O. (1998). Identifying populations for conservation on the basis of genetic markers. *Conserv. Biol.*, 12,844-855.
- Piñero, D. (2008). La diversidad genética como instrumento para la conservación y el aprovechamiento de la biodiversidad: estudios en especies mexicanas. En *Capital natural de México, vol I: Conocimiento actual de la biodiversidad*. (págs. 437-494). México: CONABIO.
- Policriti, A., & Sgarro, A. (2014). *Istituto di Genomica Applicata (IGA)*. Obtenido de <http://services.appliedgenomics.org/software/corecollections/>
- Ponce, L. (2010). Effect of vernalization and photoperiod response genes on flowering time in a large barley population. *M.Sc. Thesis. International Centre for Advanced Mediterranean Agronomic Studies and University of Lleida*.
- Ram, S., Thiruvengadam, V., & Vinod, K. (2007). Genetic diversity among cultivars, landraces and wild relatives of rice as revealed by microsatellite markers. *J. Appl. Genet.*, 48:337-345.
- Ramsay, L., Macaulay, M., Cardle, L., Morgante, M., degli-Ivanissevich, S., maestri, E., . . . Waugh, R. (1999). Intimate association of microsatellite repeats with retrotransposons and other dispersed repetitive elements in barley. *The Plant Journal*, 17(4):415-425.
- Ramsay, L., Macaulay, M., Ivanissevich, D., MacLean, K., Cardle, L., Fuller, J., . . . Waugh, R. (2000). A simple sequence repeat-based linkage map of barley. *Genetics*, 156:1997-2005.
- Rohlf, J. (2002). Numerical Taxonomy and Multivariate Analysis System Version 2.02. *Department of Ecology and Evolution. State University of New York*.
- Roques, S., Duchense, P., & Bernatchez, L. (1999). Potential of microsatellites for individual assignment: the North Atlantic redbfish (genus *Sebastes*) species complex as a case study. *Mol. Ecol.*, 8: 1703-1717.

- Roy, C., Nazeer, M., & Saha, T. (2004). Identification of simple sequence repeats in rubber (*Hevea brasiliensis*). *Current Science*, 87, 807-811.
- Rungis, D., Berube, Y., Zhang, J., Ralph, S., Ritland, C., & Ellis, B. (2004). Robust simple sequence repeat markers for spruce (*Picea* spp.) from expressed sequence tags. *Theor Appl Genet*, 109:1283-1294.
- Russel, J., Ellis, R., Thomas, W., Waugh, R., J., P., & Booth, A. (2000). A retrospective analysis of spring barley germ plasm development from 'foundation genotypes' to current successful cultivars. *Molecular breeding*, 6:553-568.
- Saisho, D., & Purugganan, M. (2007). Molecular phylogeography of domesticated barley traces expansion of agriculture in the old world. *Genetics*, 177: 1765-1776.
- Santos-Garcia, M., Toledo-Silva, G., Possidonio-Sasaki, R., Ferreira, T., Simeão-Resende, R., Chiari, L., . . . Pereira, A. (2012). Using genetic diversity information to establish core collections of *Stylosanthes capitata* and *Stylosanthes macrocephala*. *Genetics and Molecular Biology*, 35, 4, 847-861.
- Sardou, M., Baghizadeh, A., Tavasoli, A., & Babaei, S. (2011). The use of microsatellite markers for genetic diversity assessment of genus *Hordeum* L. in Kerman province (Iran). *Afr. J. Biotechnol.*, 10:1516-1521.
- Schoen, D., & Brown, A. (1993). Conservation of allelic richness in wild crop relatives is aided by assessment of genetic markers. *Proceedings National Academy of Sciences*, 22:10623-10627.
- Schuelke, M. (2000). An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology*, 18, 233-234.
- Senior, M., Murphy, J., Goodman, M., & Stuber, C. (1998). Utility of SSRs for determining genetic similarities and relationships in maize using an agarose gel system. *Crop. Sci.*, 40,226-232.

- Sharov, A., Dudekula, D., & Ko, M. (2005). A web-based tool for principal component and significance analysis of microarray data. *Bioinformatics*, 21,2548-2549.
- Shinde, D., Lai, Y., Sun, F., & Arnheim, N. (2003). Taq DNA polymerase slippage mutation rates measured by PCR and quasi-likelihood analysis: (CA/GT)_n and (A/T)_n microsatellites. *Nucleic Acid Research*, 31:974-980.
- Sneath, P., & Sokal, R. (1973). *Numerical taxonomy: The principles and practice*. San Francisco: Freeman W.H. and Co.
- Sokal, R., & Michener, C. (1958). A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin*, 38:1409-1438.
- Sokal, R., & Rohlf, F. (1962). The comparison of dendograms by objective methods. *Taxon*, 11:33-40.
- Stachel, M., Lelley, T., Grausgruber, H., & Vollmann, J. (2000). Application of microsatellites in wheat (*Triticum aestivum* L.) for studying genetic differentiation caused by selection for adaptation and use. *Theoretical and Applied Genetics*, 100:242-248.
- Struss, D., & Plieske, J. (1998). The use of microsatellite markers for detection of genetic diversity in barley populations. *Theor. Appl. Genet.*, 97,308-315.
- Thiel, T., Michalek, W., Varshney, R., & Graner, A. (2003). Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *theoretical and Applied Genetics*, 411-422.
- Torcida, S., & Perez, S. (2012). Análisis de Procrustes y el estudio de la variación morfológica. *Revista Argentina de Antropología Biológica*, 14: 131-141.
- Toro, M., & Caballero, A. (2005). Characterization and conservation of genetic diversity in subdivided populations. *Phil. Trans. R. Soc. B.*, 360,1367-1378.

- Touil, L., Guesmi, F., Fares, K., Zgrouba, C., & Ferchichi, A. (2008). Genetic diversity of some Mediterranean population of the cultivated alfalfa (*Medicago sativa* L.) using SSR markers. *Pak. J. Biol. Sci.*, 11, 1923-1928.
- Tukey, J. (1958). Bias and confidence in not quite large samples. *Annals of Mathematical Statistics*, 29:614.
- Turyagyenda, L., Kizito, E., Ferguson, M., Baguma, Y., Harvey, J., Gibson, P., . . . Osiru, D. (2012). Genetic diversity among farmed-preferred Cassava landraces in Uganda. *African Crop Science Journal*, 15-30.
- Ullrich, S. (2010). *Barley: Production, Improvement and Uses* (Vol. 12). John Wiley & Sons.
- van Oppen, M., Rico, C., Turner, G., & Hewitt, G. (2000). Extensive homoplasy, nonstepwise mutations, and shared ancestral polymorphism at a complex microsatellite locus in Lake Malawi cichlids. *Molecular and Evolution*, 17, 489-498.
- Varshney, R., Mahendar, T., Aggarwal, R., & Börner, A. (2007). Genic Molecular Markers in Plants: Development and Applications. En R. Varshney, & R. Tuberosa, *Genomics-Assisted Crop Improvement*. Springer.
- Varshney, R., Marcel, T., Ramsay, L., Russell, J., Roder, M., Stein, N., . . . Graner, A. (2007). A high density barley microsatellite consensus map with 775 SSR loci. *Theoretical and Applied Genetics*, 1091-1103.
- Vicente-Villardón, J. (2014). MULTBILOT: A package for Multivariate Analysis using Biplots. *Departamento de Estadística. Universidad de Salamanca*.
- Vicente-Villardón, J., Galindo, M., & Blázquez-Zaballos, A. (2006). Logistic Biplots. En M. Greenacre, & J. Blasius, *Multiple correspondence analysis and related methods*. (pág. 608). Cham-man and Hall.
- Villacrés, E. (2008). La cebada un cereal nutritivo (50 recetas para preparar). *INIAP*, 1-5,81-83.

- Wang, J., Yang, J., Zhu, J., Jia, Q., & Tao, Y. (2010). Assessment of genetic diversity by simple sequence repeat markers among forty elite varieties in the germplasm for malting barley. . *Journal of Zhejiang University-SCIENCE B*.
- Weir, B., & Cockerham, C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, 1358-1370.
- Westtamm, A., & Kresovich, S. (1997). Use of molecular techniques of plant genetic variation. En J. Callow, B. Ford-Lloyd, & H. Newbury, *Biotechnology and Plant Resources, conservation and use*. New York: CAB International.
- Wilson, A., Sarich, V., & Maxson, L. (1974). The importance of gene rearrangement in evolution: evidence from studies of rates of chromosomal, protein and anatomical evolution. *Proc. Natl. Acad. Sci.*, 71:3028-3030.
- Winter, P., & Kahl, G. (1995). Molecular marker technologies for plant improvement. *World Journal Microbiology & Biotechnology*, 11:438-448.
- Wouw, M. v., Hintum, T. v., Kik, C., Treuren, R. v., & Visser, B. (2010). Genetic diversity trends in twentieth century crop cultivars: A meta analysis. *TAG Teorretical and Applied Genetics*, 120(6),1241-1252.
- Wright, S. (1965). The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution*, 19:395-420.
- Xin, Z., & Chen, J. (2006). DNA Sequencing II: Optimizing Preparation and Cleanup. En K. Sudbury, *Extraction of Genomic DNA from Plant Tissue* (págs. 47-59). Jones and Bartlett Publishers.
- Yahiaoui, S. (2006). La colección nuclear española de cebada: diversidad genética y potencial agronómico. . *Tesis Doctoral. Universidad de Lleida*.
- Yuferá, E. (1987). *Química Agrícola III*. Madrid España.