# **Caracterización fisicoquímica y funcional para identificar clones élite de cacao (*Theobroma cacao l*.) en tres zonas productoras de Colombia**

## Análisis de resultados

Objetivo 1

Los cambios en la composición del cotiledón pre y post fermentación de los granos de cacao, se evaluaron en función de las siguientes variables: contenido de aceite, ácidos grasos, fenoles totales y contenido de catequina y epicatequina.

Para este fin se evaluó el supuesto de normalidad en cada una de las variables en su estado pre y posterior a la fermentación a través del test de Shapiro-Wilks, dando como resultado el rechazo de la hipótesis de normalidad en todos los casos; por consiguiente, no es apropiado usar test paramétricos (t-student, ANOVA,…) para la comparación de medias y verificación de diferencias significativas.

En este sentido fue necesario hacer las comparaciones con test no paramétricos como Wilcoxon y Kruskal-Wallis los cuales están exentos del supuesto de normalidad en los datos. Teniendo en cuenta la falta de independencia en las observaciones, por efecto del método de siembra y la ubicación espacial de los cultivos, se desarrolló el siguiente esquema de análisis para garantizar una pseudo-independencia de las observaciones:

1. Selección aleatoria de una sub-muestra (aplicando muestreo sin reemplazo) del 80% de la muestra original.
2. Aplicación del test no paramétrico de interés (Wilcoxon y/o Kruskal-Wallis) sobre la sub-muestra, comparando los datos registrados pre y post fermentación para cada variable.
3. Extracción del p-valor para el test calculado.
4. Repetir los 3 pasos anteriores 10.000 veces.
5. Con los 10.000 p-valores obtenidos evaluar los percentiles 5% y 95% de la distribución generada para determinar con precisión el p-valor real con el fin de contrastar la hipótesis nula de interés a un nivel de significancia de α = 0.05.

Para comparar el cambio en la composición del cotiledón en función de las variables medidas, se planteó la hipótesis nula en la cual no existen diferencias entre los registros medidos antes de la fermentación y después de la misma. Como hipótesis alternativa se evaluó si existen diferencias significativas, bien sean estas en términos de pérdida o ganancia.

Para esto se ejecutaron test de Wilcoxon sobre todas las observaciones sin tener en consideración el tipo de clon y la zona de acuerdo la metodología propuesta. Posteriormente, se evaluaron diferencias significativas inducidas por la zona con el test de Kruskal-Wallis y finalmente se concluyó sí existían diferencias significativas por clon y zona con un análisis descriptivo.

A continuación los resultados obtenidos. Para la variable contenido de aceite en la Figura 1, se muestran los datos obtenidos para los clones/repeticiones diferenciados por sus mediciones previas y posteriores al proceso de fermentación. Como se observa en todos los casos, el contenido de aceite es superior previo al proceso de fermentación que posterior al mismo.

|  |  |
| --- | --- |
| *Figura 1. Contenido de aceite pre y post fermentación* | *Figura 2. Distribución de p-valores para el test de Wilcoxon para evaluar diferencias significativas en el contenido de aceite* |

Ahora, para evaluar si las diferencias entre todos los clones para granos frescos y fermentados a nivel general son significativas, se realizaron test de Wilcoxon de acuerdo al esquema propuesto, dando como resultado la distribución del valor-p para el test (Figura 2), donde se evidencia la existencia de diferencias significativas entre los dos momentos de tiempo evaluados (p-valores ≈ 0).

Acto seguido, se exploraron las diferencias en el contenido de aceite entre zonas sin tener en cuenta el tipo de clon, inicialmente de forma descriptiva tal como se muestra en la Figura 3. Como se puede observar existe un comportamiento diferencial por zonas, destacándose las muestras tomadas en Arauca las cuales presentan las diferencias en contenido de aceite más altas respecto a los departamentos de Huila y Santander. No obstante, existen 3 registros en el departamento de Santander que son los que presentan perdidas más altas, incluso por encima de los 20 [unidades] en el contenido de aceite.

|  |  |
| --- | --- |
| Figura 3. Diferencia en el contenido de aceite por zonas | Figura 4. Distribución de p-valores para el test de Kruskal-Wallis para evaluar diferencias significativas en el Contenido de aceite por zona |

Luego para determinar si dichas diferencias observadas son estadísticamente significativas se realizó el test de Kruskal-Wallis dando como resultado p-valores próximos a cero, lo que indica que existen diferencias significativas entre zonas (Figura 4).

Finalmente se exploraron las diferencias entre zonas y clones para determinar si existe un efecto de interacción. Como se observa en la Figura 5, algunos clones presentan pérdidas o ganancias diferenciadas por zona. Un caso a destacar es el clon FLE 3, que para la región de Arauca exhibe una pérdida de 10 [unidades], mientras para la región del Huila se observan diferencias próximas a cero y finalmente para Santander se presentan pérdidas por encima de las 20 [unidades]. Clones como SCC 80, FLE 3 y FEAR 5 en el departamento del Huila no muestran cambios significativos en el contenido de aceite (diferencias cercanas a cero o por debajo). Los demás clones muestran comportamientos diferenciados en cuanto a diferencias en el contenido de aceite.

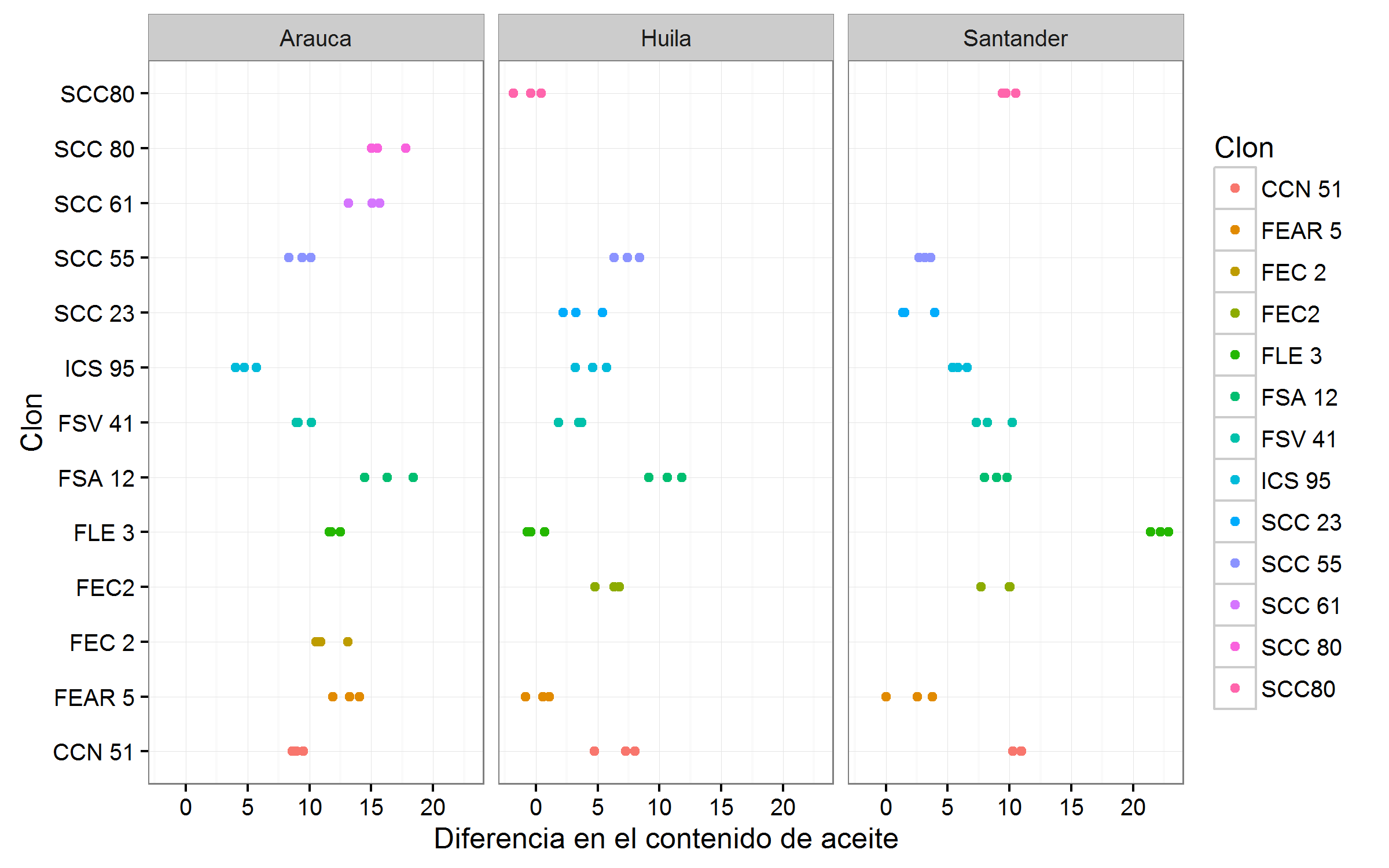


Figura 5. Diferencia en el contenido de aceite por zona y clon

Prosiguiendo con el análisis, se exploraron los cambios en la composición del cotiledón en función del contenido de fenoles totales. Como se observa en la Figura 6, al igual que el contenido de aceite existe un patrón claro de reducción del contenido de fenoles totales pre y post fermentación, para la mayoría de clones evaluados. Para determinar si existen diferencias significativas se procedió a la ejecución del test de Wilcoxon.

|  |  |
| --- | --- |
| Figura 6. Fenoles totales pre y post fermentación | Figura 7. Distribución de p-valores para el test de Wilcoxon para evaluar diferencias significativas en el contenido de fenoles totales |

En la Figura 7 se observa la distribución de p-valores, los cuales oscilan en un rango próximo a cero, lo que indica que para un nivel de significancia del 5%, las diferencias observadas son estadísticamente significativas entre el conteo de fenoles pre y post fermentación (Intervalo del 90% de confianza para p-valores: [0, 0]). Acto seguido se evaluaron las diferencias por zona.

Como se observa en la Figura 8, parece presentarse un comportamiento diferencial producido por la zona, aquí se observa que las perdidas más grandes en el contenido de fenoles totales ocurren en el departamento de Santander mientras las diferencias más pequeñas e incluso ganancias ocurren el departamento de Arauca.

|  |  |
| --- | --- |
| Figura 8. Diferencia en el contenido de fenoles totales por zonas | Figura 9. Distribución de p-valores para el test de Kruskal-Wallis para evaluar diferencias significativas en el contenido de fenoles totales por zona |

Para evaluar las diferencias entre zonas se aplicó el test de Kruskal-Wallis encontrando que existen diferencias estadísticamente significativas entre las zonas (Intervalo del 90% de confianza para p-valores: [0, 0.00002]) (ver Figura 9).

Analizando la Figura 10, clones como FEC 2, FLE 3, FSV 41, ICS 95 y SCC 55 presentan diferencias próximas a cero en el departamento de Arauca, lo que indica un cambio no significativo en términos pre y post fermentación. Por el contrario el clon FSA 12 presenta una de las pérdidas más altas por encima de 40 [unidades] en el departamento de Santander y una pérdida de 20 unidades en el Huila, mientras en el departamento de Arauca se presenta una leve ganancia del orden de 5 [unidades]. En términos generales, las pérdidas más altas ocurren en el departamento de Santander seguidas de las pérdidas ocurridas en el departamento del Huila.

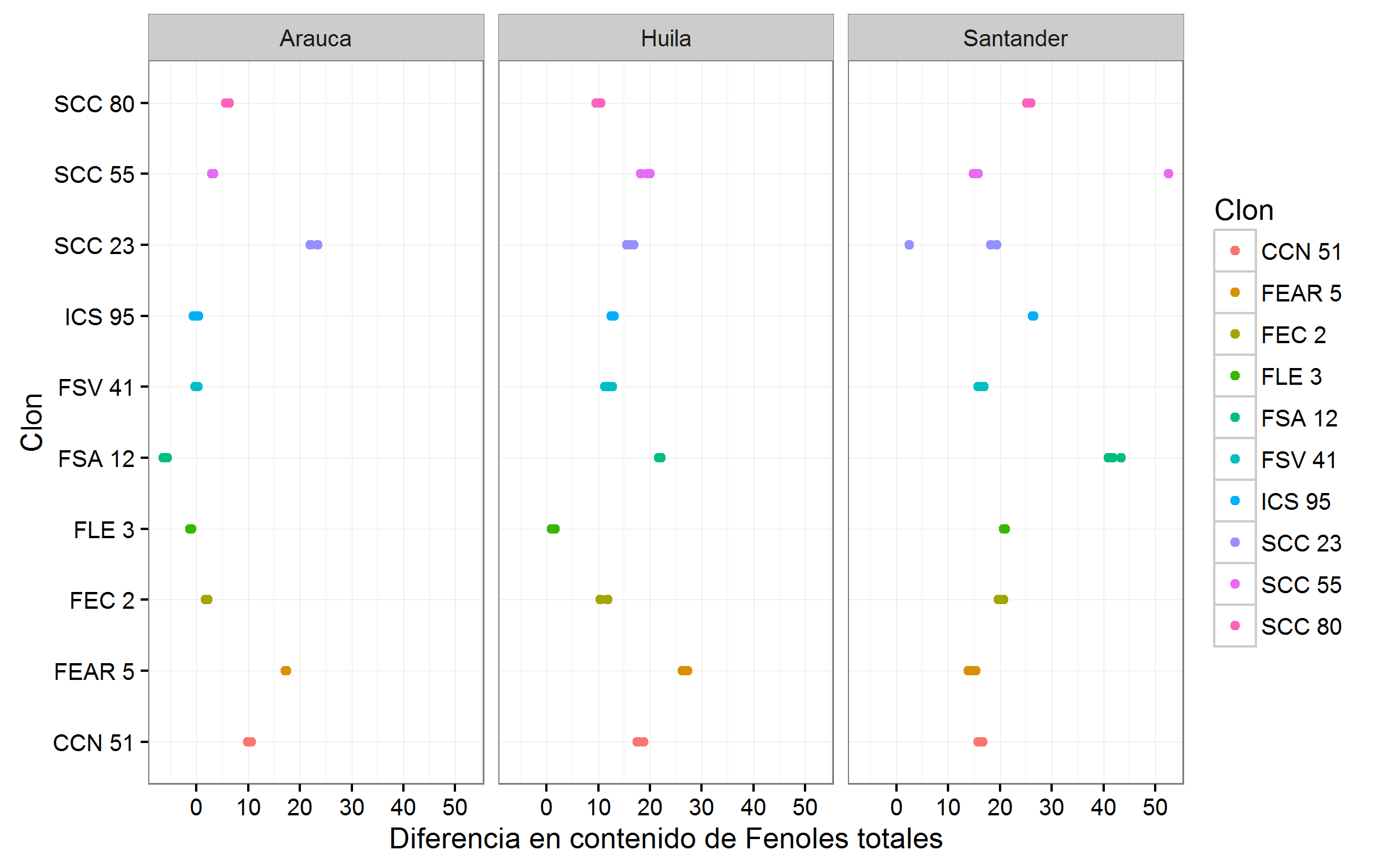


Figura 10. Diferencia en el contenido de fenoles totales por zona y clon

Prosiguiendo con el análisis, se evaluaron los cambios en términos de ácidos grasos siguiendo la misma metodología. De este modo, iniciando con el ácido palmítico se procedió a verificar si existían diferencias significativas entre las mediciones pre y posteriores a fermentación sin considerar clon y/o zona, llegando a la conclusión que no existen diferencias significativas para esta variable en consideración (Intervalo del 90% de confianza para p-valores [0.41, 0.96]) (ver Figuras 11 y 12).

|  |  |
| --- | --- |
| Figura 11. Ácido palmítico pre y post fermentación | Figura 12. Distribución de p-valores para el test de Wilcoxon para evaluar diferencias significativas en el ácido palmítico |

Explorando ahora si existen diferencias significativas entre las zonas donde se evaluaron los diferentes clones, si se observa la Figura 13 se puede notar unas leves diferencias entre las zonas evaluadas (aunque alrededor de 0), no obstante al realizar el test de Kruskal-Wallis para la verificación de la hipótesis de diferencias entre zonas se obtuvo un rango de p-valores [0.06, 0.71] con una alta tendencia a presentar un p-valor alrededor de 0.2 (Figura 14), con el cual no se podría verificar la existencia de diferencias significativas.

|  |  |
| --- | --- |
| Figura 13. Diferencia en ácido palmítico por zona | Figura 14. Distribución de p-valores para el test de Kruskal-Wallis para evaluar diferencias significativas en el ácido palmítico por zona |

Finalmente analizando la Figura 15, se observa que para casi todas las combinaciones clon/ambiente, los ligeros cambios oscilan alrededor de cero, lo que no indica ningún cambio posterior al proceso de fermentación para el ácido palmítico.

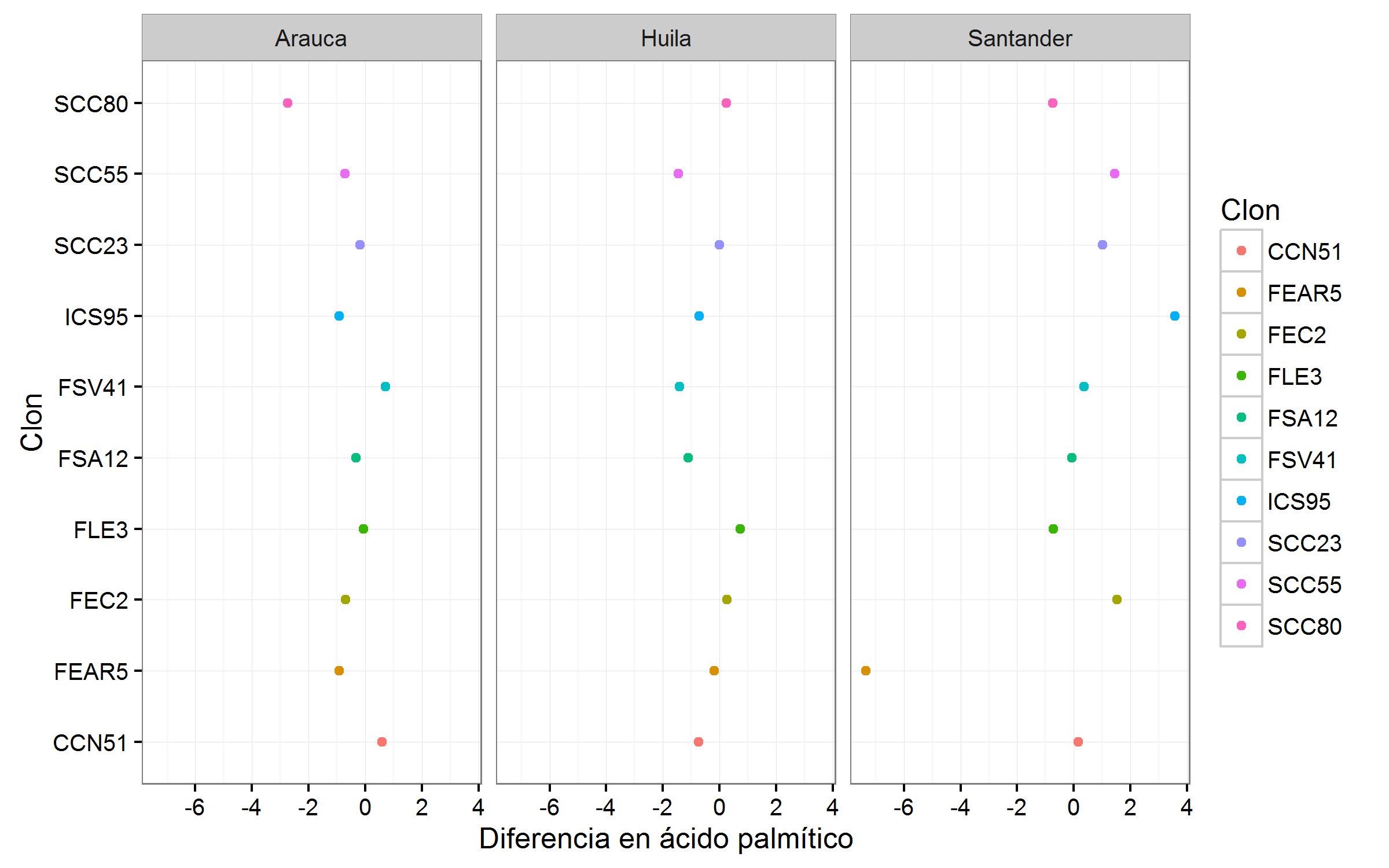


Figura 15. Diferencia en ácido palmítico por zona y clon

En términos del ácido esteárico, aunque se observa una mayor variabilidad en los datos (ver Figura 16), al realizar el test de Wilcoxon para verificar si existen diferencias entre la etapa previa y posterior a la fermentación se logró verificar que no existen diferencias significativas (p-valores, intervalo del 90% [0.33, 0.94]).

|  |  |
| --- | --- |
| Figura 16. Ácido esteárico pre y post fermentación | Figura 17. Distribución de p-valores para el test de Wilcoxon para evaluar diferencias significativas en el ácido esteárico |

Acto seguido, evaluando si existen diferencias entre zonas (Figura18), el test de Kruskal-Wallis indica que la distribución de p-valores oscila en un intervalo de [0.44, 0.99] para un 90% de las veces, por ende se descartan diferencias entre zonas.

|  |  |
| --- | --- |
| Figura 18. Diferencia en ácido esteárico por zona | Figura 19. Distribución de p-valores para el test de Kruskal-Wallis para evaluar diferencias significativas en el ácido esteárico por zona |

Ahora, explorando las diferencias por clon/zona, en la Figura 20 se puede notar que la mayoría de clones presentan variaciones alrededor de 0 lo que índica cambios nulos, sin embargo, clones como SCC 80, ICS 95 y FEAR 5 muestran fuertes variaciones dependiendo de la zona.

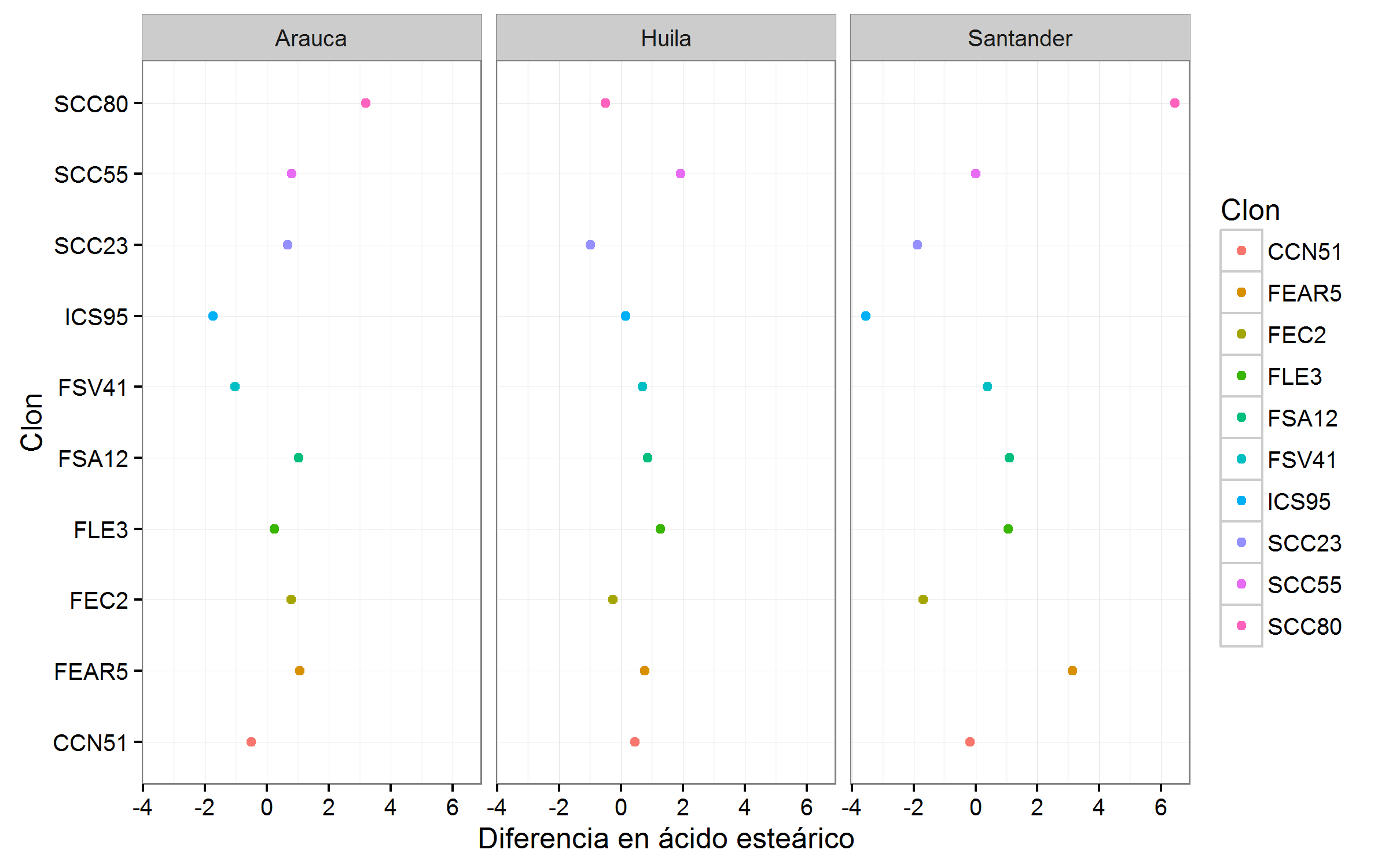


Figura 20. Diferencia en ácido esteárico por zona y clon

Continuando con el proceso, a continuación se examinaron los cambios relacionados con el ácido oleico, el cual presenta baja variación y el test de Wilcoxon, permite concluir que no existen diferencias significativas entre las mediciones pre y post fermentación (p-valores, intervalo 90% [0.52, 0.99]).

|  |  |
| --- | --- |
| Figura 21. Ácido oleico pre y post fermentación | Figura 22. Distribución de p-valores para el test de Wilcoxon para evaluar diferencias significativas en el ácido oleico |

Por otro lado, al explorar la relación que podría tener la zona sobre el contenido de ácido oleico (ver Figuras 23 y 24) se encontró que bien podría existir una diferenciación por zona pero muy débil (distribución de p-valores: [0.03, 0.50]).

|  |  |
| --- | --- |
| Figura 23. Diferencia en ácido oleico por zona | Figura 24. Distribución de p-valores para el test de Kruskal-Wallis para evaluar diferencias significativas en el ácido oleico por zona |

Luego contrastando las diferencias a nivel clon/zona, se tienen que aunque la mayoría de diferencias oscila alrededor de 0, algunos clones muestran un comportamiento diferencial por zona, tales como ICS 95, FL 3 y FEAR 5.



Figura 25. Diferencia en ácido oleico por zona y clon

Finalmente evaluando las diferencias en el contenido de ácido linoleico de manera general, se tiene que las diferencias observadas no son estadísticamente significativas, por tanto no se puede concluir que haya diferenciación alguna entre el contenido de ácido linoleico pre y post fermentación (ver Figuras 26 y 27. Para la Figura 27, el intervalo del 90% de confianza muestra que los p-valores del test de Wilcoxon oscilan entre [0.47, 0.96], por tanto no se puede rechazar la hipótesis nula de igual en las mediciones para los dos tiempos evaluados).

|  |  |
| --- | --- |
| Figura 26. Ácido linoleico pre y post fermentación | Figura 27. Distribución de p-valores para el test de Wilcoxon para evaluar diferencias significativas en el ácido linoleico |

Nuevamente evaluando si existen diferencias entre las zonas (Figura 28), las simulaciones del test Kruskal-Wallis arrojaron como resultado que no existen diferencias significativas por zona (Intervalo de p-valores 90% de confianza: [0.17, 0.87]).

|  |  |
| --- | --- |
| Figura 28. Diferencia en ácido linoleico por zona | Figura 29. Distribución de p-valores para el test de Kruskal-Wallis para evaluar diferencias significativas en el ácido linoleico por zona |

Y posteriormente, analizando el comportamiento desagregado, se identifican ciertos clones que presentan un comportamiento diferencial por zona, estos son: SCC 80, SCC 23 y FEAR 5.

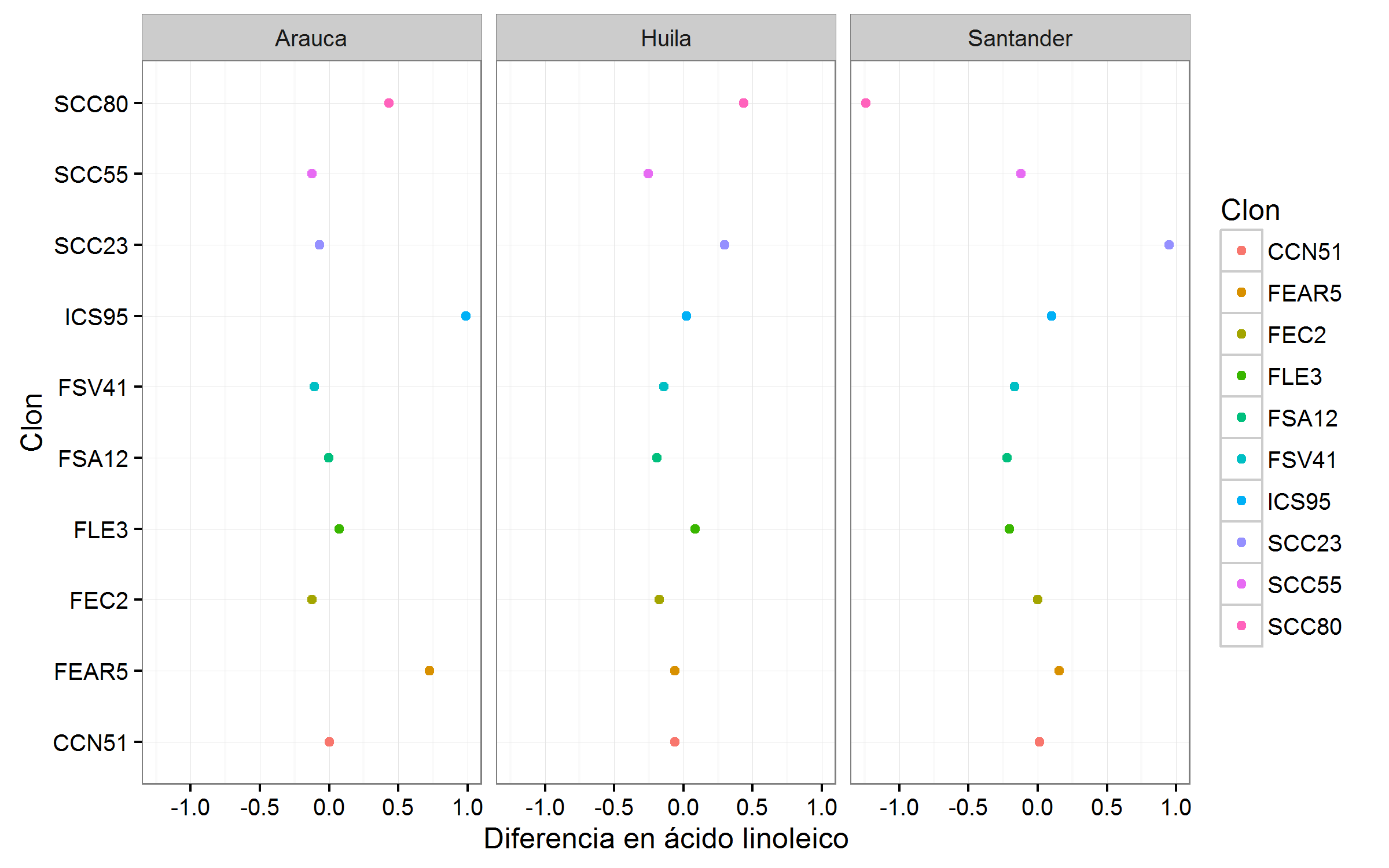


Figura 30. Diferencia en ácido linoleico por zona y clon

Prosiguiendo con el análisis de la composición del cotiledón, a continuación se presentan los resultados relacionados con las variables funcionales contenido de catequina y epicatequina. Para el contenido de catequina en las etapas pre y post fermentación se puede observar en la Figura 31 que en términos generales para la mayoría de clones/repeticiones evaluadas el contenido de catequina tiende a reducir posterior al proceso de fermentación. Para verificar sí dichas diferencias son estadísticamente significativas se realizó un test de Wilcoxon a partir del cual se logró concluir a partir de un nivel de significancia del 5% que las diferencias observadas son estadísticamente significativas y diferentes de cero (distribución de p-valores [0.00000, 0.00001], ver Figura 32).

|  |  |
| --- | --- |
| Figura 31. Catequina pre y post fermentación | Figura 32. Distribución de p-valores para el test de Wilcoxon para evaluar diferencias significativas en el contenido de catequina |

Luego para verificar si dichas diferencias están inducidas por la zona, se realizó un test de Kruskal-Wallis a partir del cual se pudo verificar que no existen diferencias significativas por zona (Intervalo del 90% de confianza para la distribución de p-valores [0.18124, 0.94892], Figuras 33 y 34).

|  |  |
| --- | --- |
| Figura 33. Diferencia en el contenido de catequina por zona | Figura 34. Distribución de p-valores para el test de Kruskal-Wallis para evaluar diferencias significativas en el contenido de catequina por zona |

Finalmente realizando la comparación a través de las zonas y clones se identifica que los clones con reducciones más abruptas en el contenido de catequina posterior al proceso de fermentación corresponden a ICS 95 y SCC 23 que presentan reducciones próximas a los 2000 [unidades] en el departamento de Santander. Por otro lado, clones como SCC 55, FSV 41, FSA 12 y CCN 51 son los que presentan leves cambios posterior a la fermentación indistintamente para los tres departamentos evaluados.

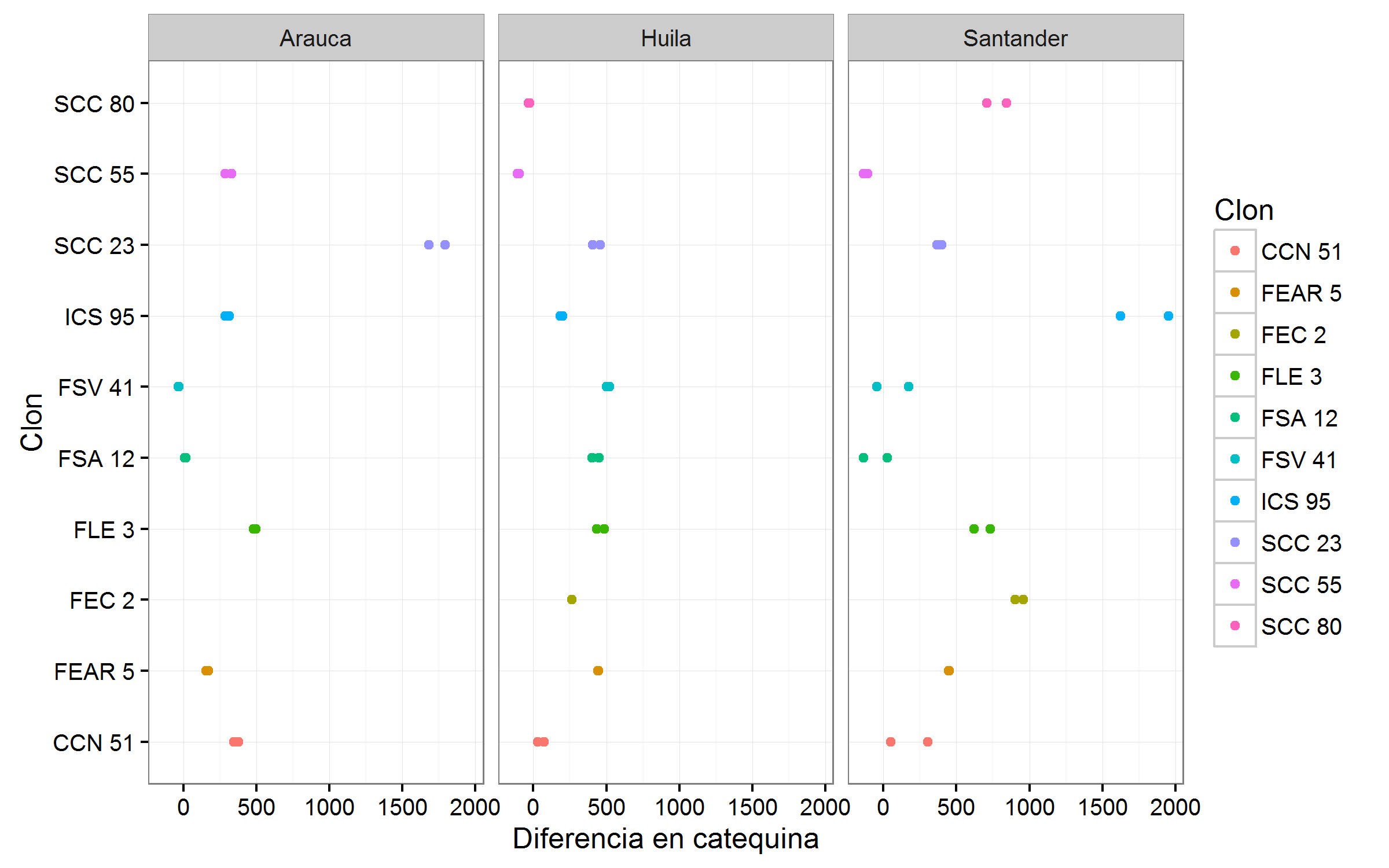


Figura 35. Diferencia en el contenido de catequina por zona y clon

Evaluando el contenido de epicatequina pre y post fermentación, como se observa en la Figura 36, en términos generales este tiende a disminuir después del proceso de fermentación, no obstante este cambio se presenta de forma abrupta solamente para algunos clones. Luego para comprobar si dichas diferencias que se observan de forma descriptiva son estadísticamente significativas, se procedió a realizar un test de Wilcoxon dando como resultado que las diferencias observadas son significativas (distribución de p-valores, Figura 37; acumulándose todos los p-valores ligeramente por encima de 0).

|  |  |
| --- | --- |
| Figura 36. Epicatequina pre y post fermentación | Figura 37. Distribución de p-valores para el test de Wilcoxon para evaluar diferencias significativas en el contenido de epicatequina |

A continuación para evaluar si existe un efecto inducido por la zona, en primera medida en la Figura 38 se observa que todas las diferencias en contenido de epicatequina pre y post fermentación están por encima de cero excepto para un clon. Adicional a esto el departamento del Huila es el que presenta las reducciones más altas en contenido de epicatequina seguido de los departamentos de Arauca y Santander respectivamente. Para evaluar si dichas diferencias son estadísticamente significativas se realizó el test de Kruskal-Wallis encontrándose que con un intervalo de confianza del 90%, el p-valor se encuentra alrededor de [0.00004, 0.00266], lo que indica que dichas diferencias son significativas.

|  |  |
| --- | --- |
| Figura 38. Diferencia en el contenido de epicatequina por zona | Figura 39. Distribución de p-valores para el test de Kruskal-Wallis para evaluar diferencias significativas en el contenido de epicatequina por zona |

Posteriormente se exploraron las diferencias por zona y clon, encontrando que el único clon que presenta un incremento en el contenido de epicatequina posterior a la fermentación es el clon SCC 23 para el departamento de Arauca. Mientras por el contrario, los clones con la mayor reducción se encuentran en el departamento del Huila y corresponden a los clones CCN 51, FEAR y FSA 12 cuyo nivel de reducción sobrepasa los 4000 [unidades].

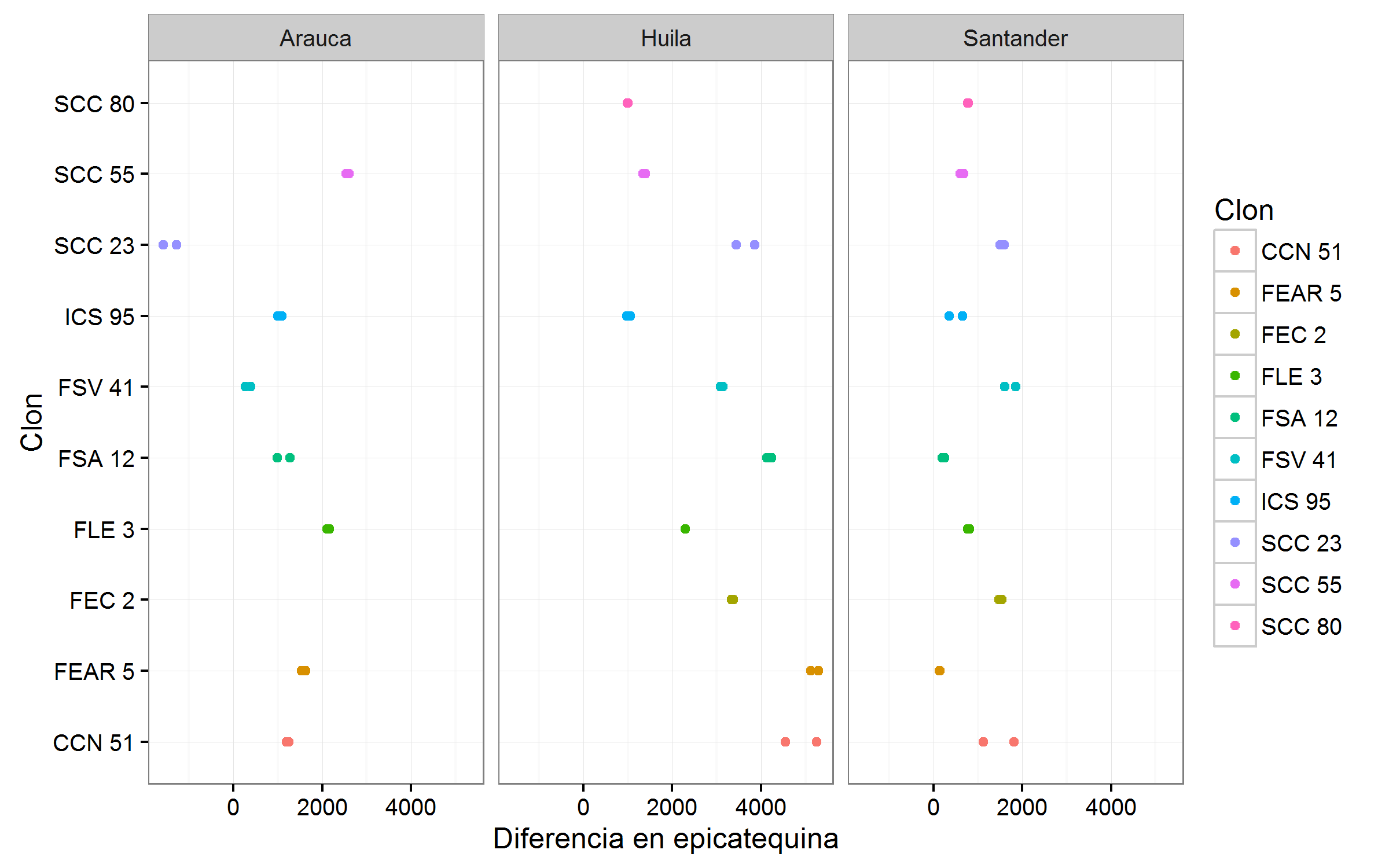


Figura 40. Diferencia en el contenido de epicatequina por zona y clon

Ahora, continuando el análisis con la variable actividad antioxidante, se puede observar en la Figura 41 que en términos generales esta se ve incrementada posterior a la fermentación. Luego, para verificar que dichas diferencias observadas son significativas, se procedió a la realización de un test de Wilcoxon, encontrándose que la distribución de p-valores oscila alrededor de 0, lo que es muestra de la significancia del test y en otras palabras que las diferencias observadas son estadísticamente significativas.

|  |  |
| --- | --- |
| Figura 41. Actividad antioxidante pre y post fermentación | Figura 42. Distribución de p-valores para el test de Wilcoxon para evaluar diferencias significativas en la actividad antioxidante |

Acto seguido para evaluar la influencia del sitio donde se realizó el estudio se procedió a la realización de un test de Kruskal-Wallis. Si se observa en la Figura 43, todas las diferencias indican ganancia en términos de actividad antioxidante para las tres zonas, no obstante, parece no haber un patrón diferenciador por zona. Prueba de esto es el intervalo de confianza del 90% para el p-valor el cual oscila alrededor de 0.06148 y 0.71881, como se observa en la Figura 44.

|  |  |
| --- | --- |
| Figura 43. Diferencia en actividad antioxidante por zona | Figura 44. Distribución de p-valores para el test de Kruskal-Wallis para evaluar diferencias significativas en la actividad antioxidante por zona |

Como paso posterior se exploraron las diferencias por clon a través de los tres ambientes, dando como resultado el hecho que el mismo clon puede tener un comportamiento diferente en cada ambiente evaluado, prueba de esto constituyen los clones: FSA 12, SCC 23 y FEC 2.

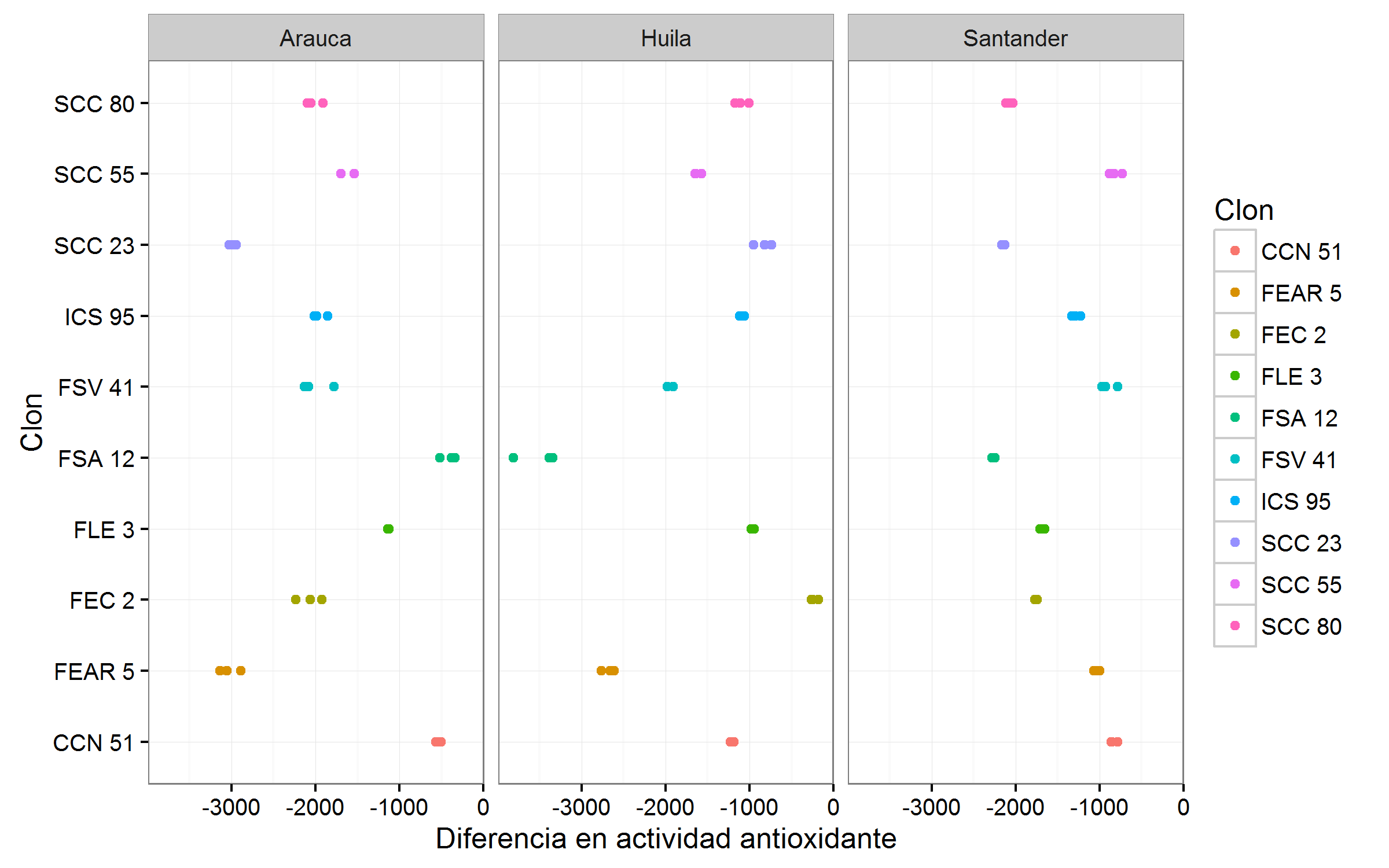


Figura 45. Diferencia en actividad antioxidante por zona y clon

Como hallazgos generales de este objetivo se podría concluir que las variables que ejercen un efecto diferenciador claro en la composición del cotiledón entre zonas y clones son el contenido de aceite [reducción], catequina y epicatequina [reducción], fenoles totales [reducción] y la actividad antioxidante [aumento], por el contrario todos los ácidos grasos parecen no alterarse significativamente posterior al proceso de fermentación de granos.

# Objetivo 2

Para identificar los mejores clones según variables fisicoquímicas y funcionales se realizó un Análisis de Componentes Principales (ACP) utilizando las siguientes variables:

***Fisicoquímicas***: ácidos grasos (esteárico, oleico y palmítico) y la relación Tauromina/Cafeína

***Funcionales***: contenido de fenoles totales, actividad antioxidante, contenido de catequina y epicatequina.

Adicional a esto se incluyeron en el análisis los índices de mazorca y grano como variables de cosecha para determinar su influencia en la composición fisicoquímica y funcional de los clones evaluados. Como primera medida se exploró la relación entre todos los posibles pares de variables a través de un análisis de correlación, calculando el coeficiente de correlación no paramétrico de Spearman, este indicador permite identificar relaciones de tipo creciente o decreciente sin necesidad que estas sean de tipo lineal.

En la Figura 46 se muestra la matriz de correlaciones de Spearman, identificando grupos de variables que se relacionan de manera positiva como el contenido de fenoles totales, catequina porcentaje de aceite y el índice de grano, por otro lado, se encuentran variables como la actividad antioxidante, la epicatequina, el porcentaje de cenizas y el índice de mazorca que aunque se asocian de manera positiva entre ellas, se relacionan de manera negativa con el primer grupo de variables.

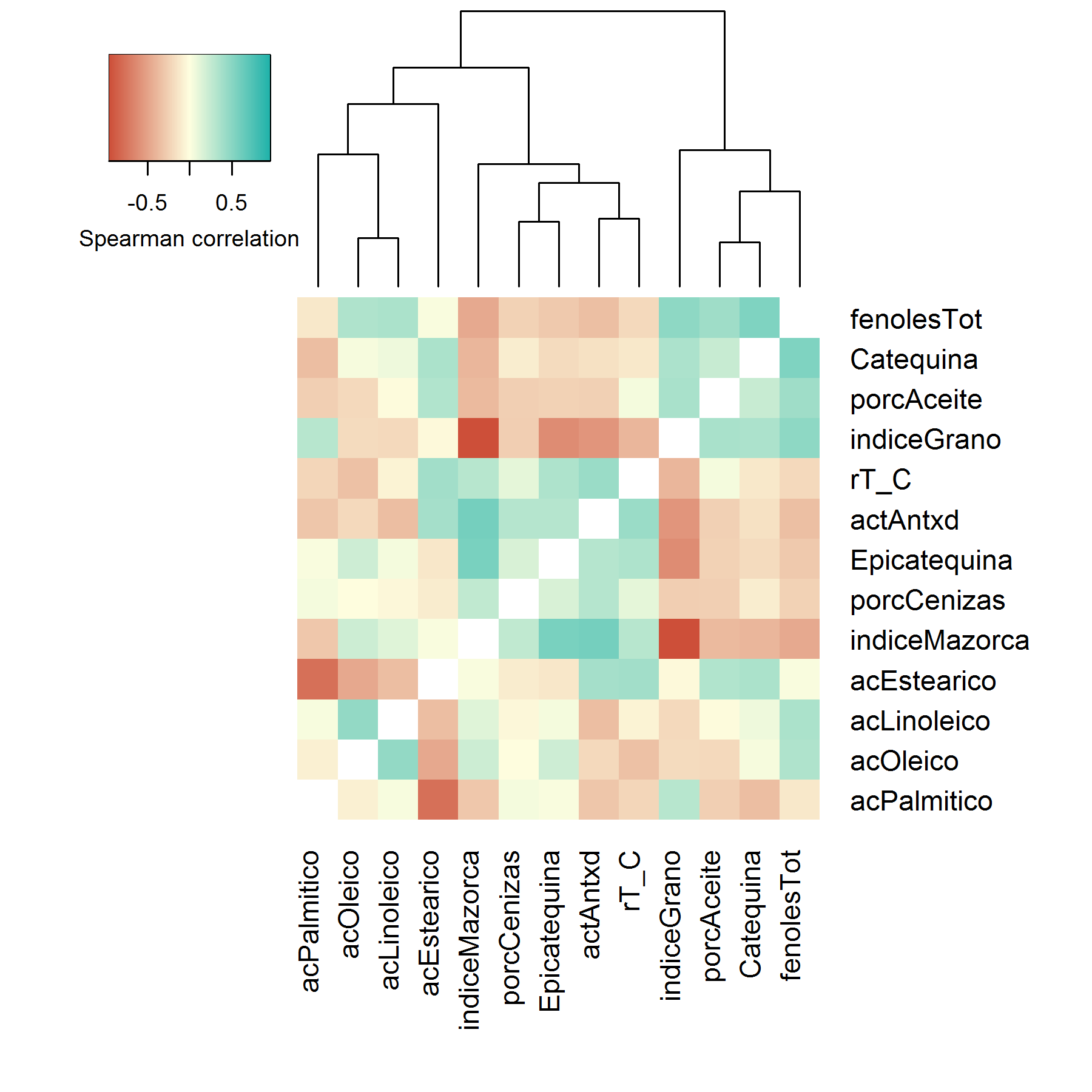


Figura 46. Matriz de correlaciones entre todos los pares de variables.

Por otro lado, los ácidos grasos parecen tener correlaciones casi nulas con el resto de variables y únicamente se destacan algunos casos como la relación entre el ácido palmítico y el ácido esteárico donde la relación negativa es del orden de -0.78. La presencia de altas correlaciones, tanto positivas como negativas, da vía libre al Análisis de Componentes Principales (ACP) el cual busca explicar en un menor número de variables las principales relaciones existentes entre los datos.

De este modo, realizando el ACP se logró descomponer la variabilidad de los datos en un total de 3 componentes principales explicando un 63.6% de la variación total en los datos, específicamente la matriz de correlaciones. La selección del número de componentes principales a utilizar para la interpretación se realizó a partir de la identificación del “codo” en la Figura 47, a partir del cual se escoge el número de componentes previo al codo del gráfico. Para este caso, el codo se forma en la cuarta componente principal por lo tanto se seleccionan las tres primeras para fines de interpretación. La justificación para realizar dicha elección viene dada por el hecho que adicionar más componentes para la interpretación aportaría un bajo porcentaje de explicación frente al que ya se cuenta.

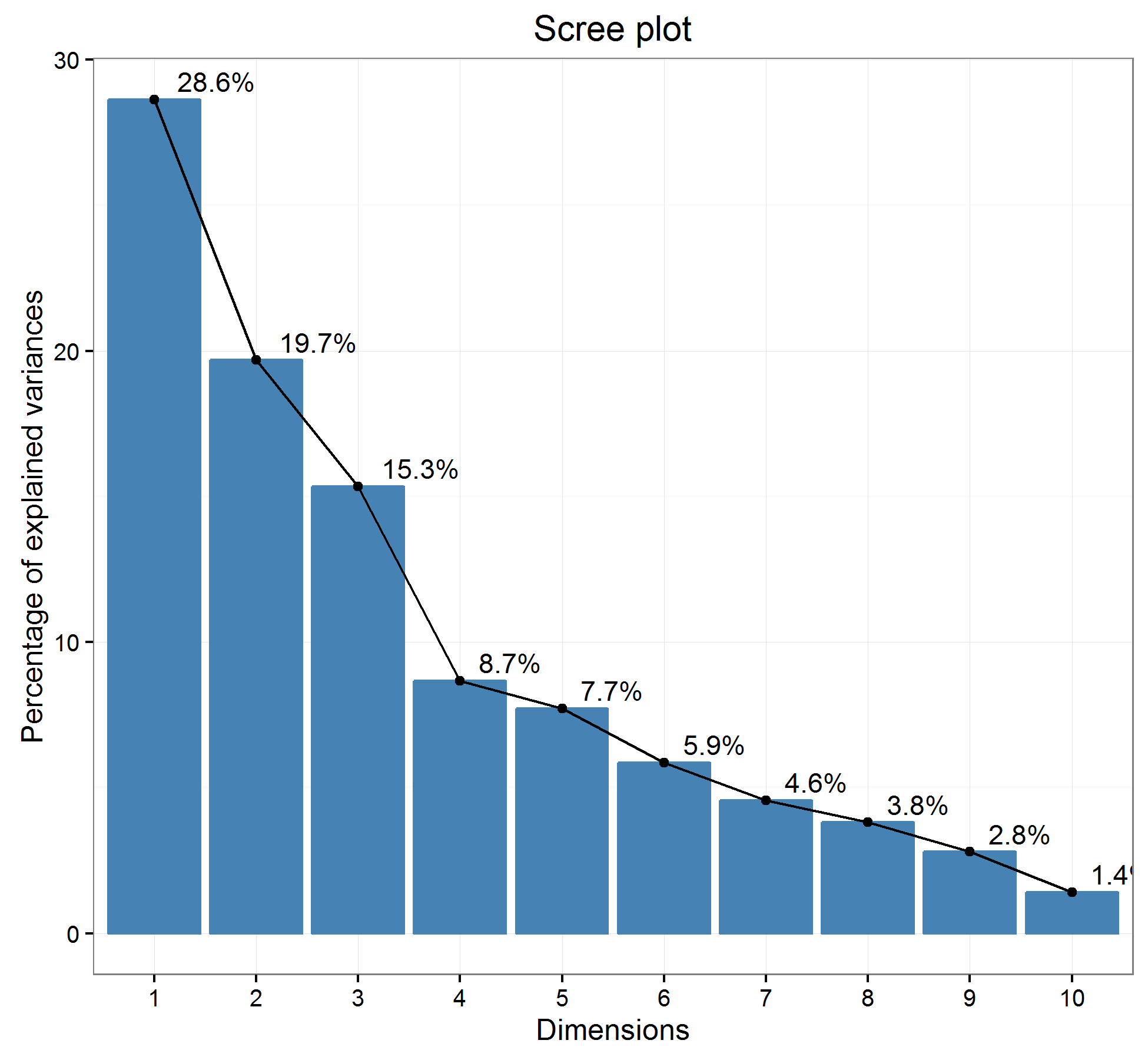


Figura 47. Porcentaje de variabilidad explicado por cada componente principal

Debido a que los componentes principales constituyen una combinación lineal de las variables originales se determinó el aporte de cada variable por componente principal. Para esto se calcularon tres indicadores que miden la importancia, contribución y correlación de las variables originales con cada una de las componentes principales seleccionadas, estos índices se muestran en la Figura 48.

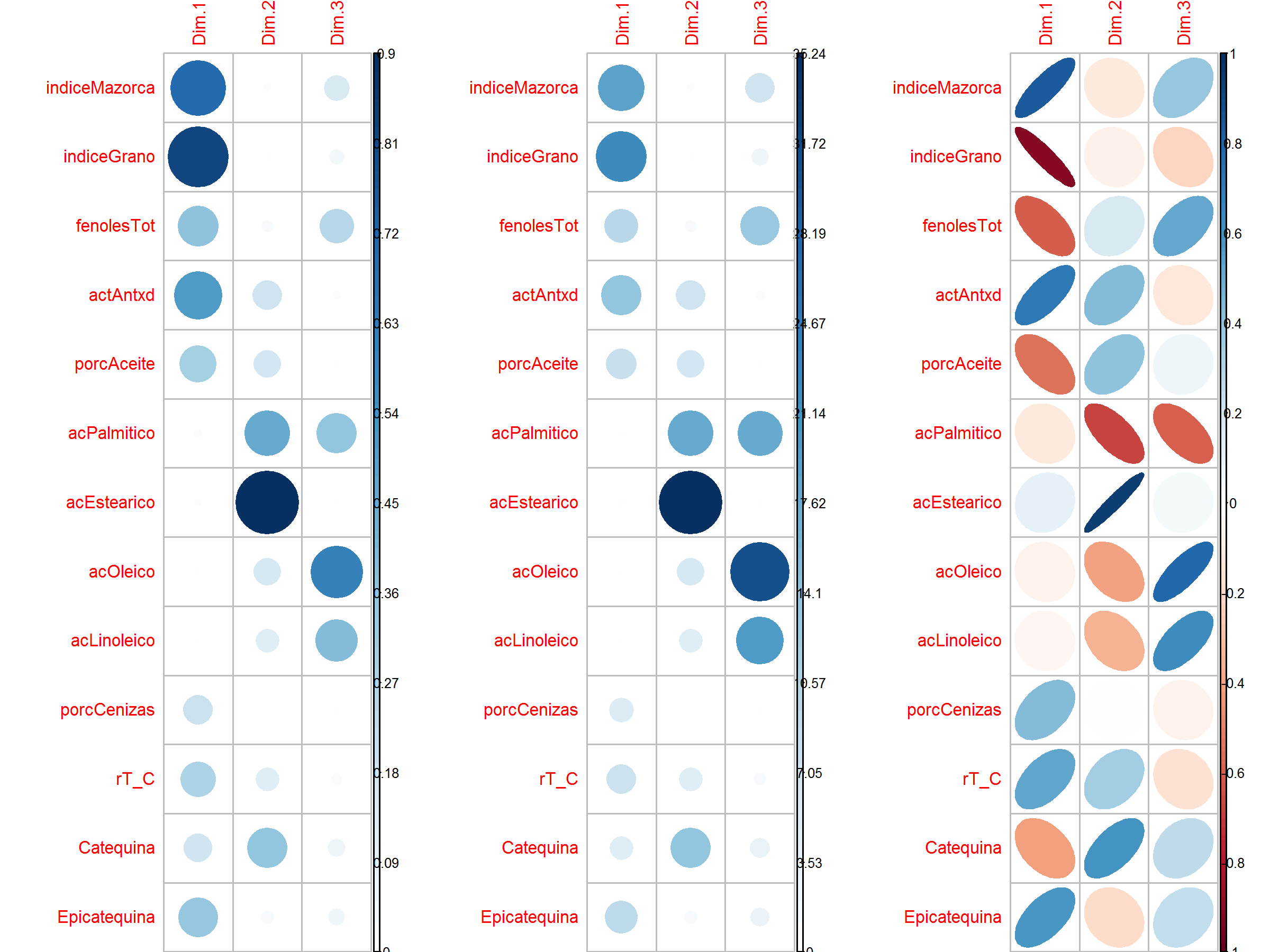


Figura 48. Índices de calidad de representación. Importancias, contribuciones y correlaciones.

Los primeros dos índices de la Figura 48 (de izquierda a derecha) miden el grado aporte que ejerce cada variable en la construcción de cada componente principal, aquí cada columna corresponde a las componentes principales seleccionadas y cada fila a las variables utilizadas para hacer el ACP. Por otro lado, el tercer indicador mide la correlación existente entre cada variable y cada componente principal con esto se busca cuantificar si el aporte de la variable a la componente se realiza en sentido positivo o negativo.

De este modo se tiene para la primera componente principal que su principal poder discriminador entre los clones se ve influenciado por: los índices de cosecha [Mazorca: positivo, Grano: negativo], el contenido de fenoles totales [negativo], la actividad antioxidante [positivo], el porcentaje de aceite [negativo] y el contenido de epicatequina [positivo].

Por su parte la segunda componentes discrimina los clones principalmente por: los ácidos grasos palmítico [negativo], esteárico [positivo] y el contenido de catequina [positivo]. Y finalmente la tercera componente principal discrimina específicamente por: los ácidos oleico [positivo] y linoleico [positivo].

Por otro lado, variables como el porcentaje de cenizas y la relación Tauromina/Cafeína ejercen un efecto muy bajo en la clasificación de los clones y por tanto no forman parte de una contribución significativa a las componentes principales escogidas. A continuación se muestran los círculos de correlaciones entre los compontes principales y las variables originales (Figuras 49, 50 y 51), estos indican las direcciones de máxima relación o variación y las variables que influyen en ellas, adicional a esto la escala de colores indica la importancia que tiene cada variable en la construcción de la componente, complementando el análisis de la Figura 48.

|  |  |
| --- | --- |
| Figura 49. Circulo de correlaciones para variables – Componentes principales 1 y 2 | Figura 50. Circulo de correlaciones para variables – Componentes principales 1 y 3 |
| Figura 51. Circulo de correlaciones para variables – Componentes principales 2 y 3 | |

Las direcciones de máxima variación son de utilidad para determinar cuáles son los individuos u observaciones que más se destacan en función de las variables que más aportan en la construcción de las componentes principales. Como se puede observar en las Figuras 52, 53 y 54 se tiene una vista conjunta de clones y variables discriminados por el ambiente en el cual se recopilaron los datos. Aquí los clones próximos al centro de coordenadas corresponden a los individuos cercanos al promedio en la mayoría de variables, mientras los individuos más alejados del centro y próximos a los ejes principales son los clones que más se destacan bien sea por mediciones extremas (mínimos o máximos) de las variables originales.

|  |  |
| --- | --- |
| Figura . Biplot dimensiones 1 y 2 | Figura . Biplot dimensiones 1 y 3 |
| Figura . Biplot dimensiones 2 y 3 | |

En términos generales a partir de las Figuras 52, 53 y 54 se observa una clara distinción entre los ambientes, esto nos da como resultado un comportamiento diferencial de los clones en los diferentes ambientes tal y como se observó en el análisis de composición del cotiledón. Ahora para identificar los clones que más se destacan en cada componente principal e interpretarlos según las variables originales, se presenta en la Figura 55 el gráfico de contribuciones de cada clon sobre los ejes principales, esto teniendo en cuenta las direcciones máxima variación y las correlaciones de las variables originales.

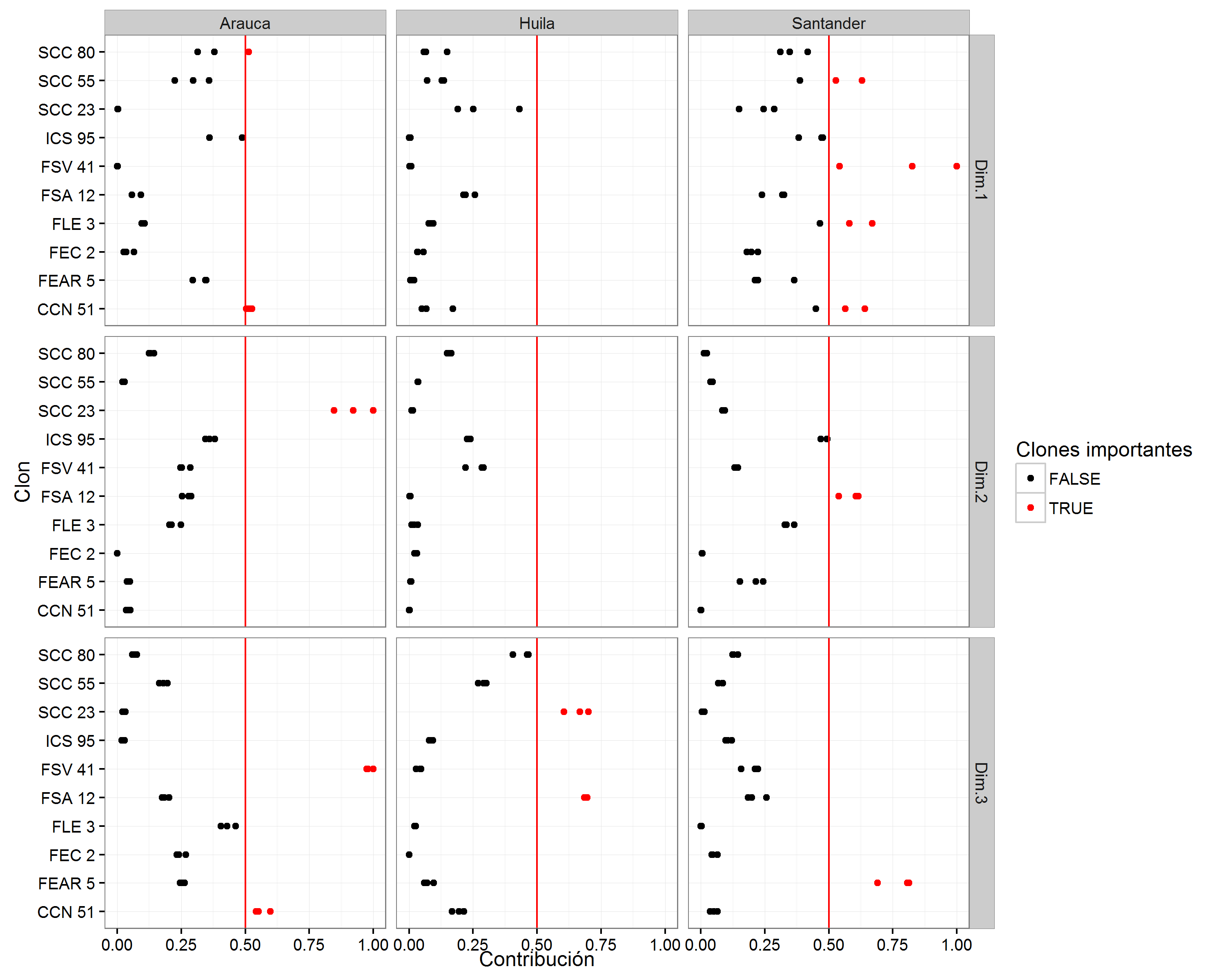


Figura 55. Contribución de clones por componente principal y ambiente

Como se observa en la Figura 55, existen ciertos clones que tienden a destacarse más en un determinado ambiente en función de las variables que definen las componentes principales. Casos destacados son los clones de cacao en el ambiente de Santander donde se destacaron los clones CCN 51, FLE 3, FSV 41 y SCC 55 en la primera componente principal.

En este sentido, el clon CCN 51 del departamento de Santander se destaca por tener unos de los más bajos registros en contenido de epicatequina así como una baja actividad antioxidante. Mientras, el clon FLE 3 se destaca por tener el porcentaje más alto de aceite. Prosiguiendo el clon FSV 41 muestra el menor índice de mazorca y por consiguiente el mayor índice de grano agregándole a esto la más baja relación Tauromina/Cafeína. Y finalmente, el clon SCC 55 se destaca por tener la más baja actividad antioxidante y un bajo porcentaje de ceniza.

Por otro lado, el departamento de Arauca cuenta únicamente con un clon el cual sus tres replicas pasan el umbral de importancia del 0.5, el clon CCN 51 destaca por presentar una de las más altas actividades antioxidantes, el más bajo porcentaje de aceite, el más bajo contenido de ácido oleico y los más altos valores en porcentaje de cenizas y relación Tauromina/Cafeína.

Continuando con la lectura de resultados, se tiene que para la segunda componente principal relacionada con los ácidos grasos palmítico, esteárico y el contenido de catequina, se destacan dos clones uno en el departamento de Arauca que es el clon SCC 23 el cual se destaca por la más alta actividad antioxidante y el más alto contenido del ácido esteárico. Mientras el clon FSA 12 en el departamento de Santander muestra los conteos más altos de fenoles totales y catequina.

En términos de la tercera componente principal relacionada principalmente con los ácidos oleico y linoleico, se tiene que los clones más destacados por departamento son:

1. Santander: clon FEAR 5 que presenta el contenido más bajo de ácido palmítico y el más alto de ácido oleico.
2. Huila: clon FSA 12 con uno de los más altos contenidos de ácido oleico así como un alto índice de mazorca y más bajos porcentajes de aceite. Del mismo modo se destaca el clon SCC 23 con el índice de mazorca más alto, más bajo de índice de grano y uno de los más altos contenidos de ácido oleico.
3. Arauca: clon CCN 51 con alta actividad antioxidante, el más bajo porcentaje de aceite al igual que el más bajo contenido de ácido oleico y los valores más altos del porcentaje de cenizas y de la relación Tauromina/Cafeína. Por otro lado, también se destaca el clon FSV 41 con uno de los conteos más bajos de fenoles totales, el contenido más alto de ácido palmítico, el contenido más bajo de ácido linoleico y catequina.

# Objetivo 3

Con el fin de identificar los principales compuestos que caracterizan los olores y sabores del cacao para los clones evaluados se realizó un Análisis de Correspondencias Múltiple (MCA, por sus siglas en inglés), utilizando como variables de entrada los diferentes compuestos medidos a través del análisis organoléptico donde se midieron 49 tipos de compuestos volátiles distintos.

En primera medida se buscó medir algún tipo de asociación entre los distintos compuestos volátiles, para esto se desarrollaron Test Chi-Cuadrado entre pares de compuestos con el evaluar la independencia o asociación entre dichas variables. Como se observa en la Figura 56 existe un reducido número de asociaciones entre los compuestos volátiles evaluados (p-valores significativos o próximos a 0 de color rojo), no obstante para explorar más a fondo las asociaciones descubiertas se presentan a continuación los resultados del MCA.

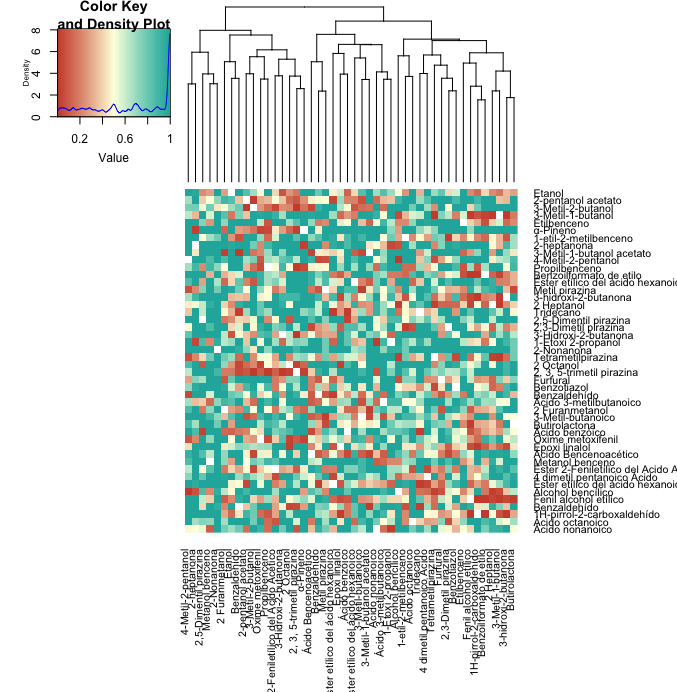


Figura 56. Búsqueda de asociaciones significativas entre pares de variables

Previo al análisis se identificaron aquellos compuestos volátiles que se encontraban presentes en todos los clones y por ende no aportaban rasgos de variación útiles, estos fueron:

* Ácido acético
* 2,3 Butaneidol
* Acetofenona

Posteriormente se realizó el MCA, encontrando que para 7 ejes principales se logra alcanzar un porcentaje de varianza explicada del 60.1% con respecto al total. Este es un indicador de reducidas relaciones entre los compuestos volátiles, ya que la variación en los datos se distribuye por un elevado número de ejes principales, no obstante, inferior al número de compuestos volátiles medidos en total (Figura 57).

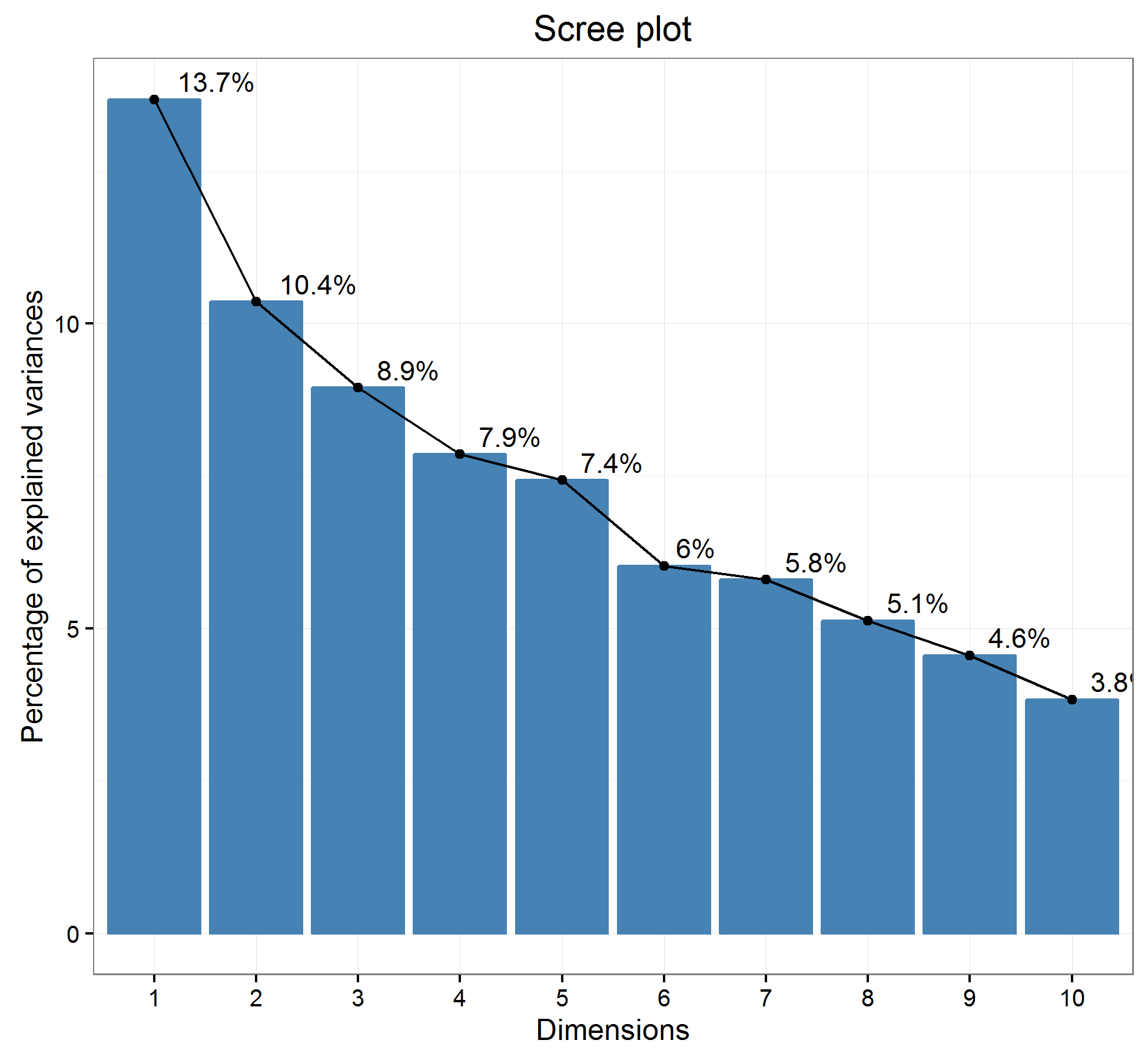


Figura 57. Porcentaje de variabilidad explicado por los ejes principales en el MCA

Dado el elevado número de compuestos volátiles en estudio, en la Tabla 1 se muestran los compuestos que tienen la mayor representatividad y por ende la mejor calidad de representación en cada uno de los ejes principales.

Tabla 1. Compuestos volátiles determinantes por eje principal

|  |  |
| --- | --- |
| **Eje principal** | **Compuesto volátil** |
| 1 | Fenil alcohol etílico |
| 1H-pirrol-2-carboxaldehído |
| Benzoilformato de etilo |
| 2 Heptanol |
| 3-Metil-1-butanol |
| 3-hidroxi-2-butanona |
| 2 | Ácido bencenoacético |
| 3-Metil-2-butanol |
| 2-pentanol acetato |
| Benzaldehído |
| 3-Metil-butanoico |
| 3 | 3-Hidroxi-2-butanona |
| Etanol |
| 2, 3, 5-trimetil pirazina |
| Éster etílico del ácido hexanoico |
| Éster 2-Feniletílico del ácido acético |
| 4 | Benzaldehído |
| 5 | Éster etílico del ácido hexanoico |
| 2 Furanmetanol |
| 6 | Ácido nonanoico |
| 7 | Ácido octanoico |

Al representar la pertenencia o no de cada uno de los compuestos volátiles en los clones estudiados (ejemplo, **compuesto\_volátilXX\_1**: el clon presenta el compuesto o, **compuesto\_volátilXX\_0**: el clon no lo presenta), en un plano factorial de dos dimensiones, se puede observar la gran variación de condiciones representadas, ver Figura 58. Un caso a destacar consiste en el compuesto fenil alcohol etílico en los extremos izquierdo y derecho, aquí representando a cada uno de los clones en el mismo plano se tendría que los clones que se ubiquen a la izquierda del plano y cuente con una alta contribución presentaran dicho compuesto, mientras los que se encuentren a la derecha carecerán de él (Figura 59).

|  |  |
| --- | --- |
| Figura 58. Correlaciones para variables – Ejes principales 1 y 2 | Figura 59. Posicionamiento de los clones en el primer plano factorial |

Los clones más destacados por eje principal son:

Eje principal 1: Santander-FLE3, Arauca-FSA12, Arauca-ICS95

Eje principal 2: Santander-FSA12, Arauca-SCC55, Arauca-SCC80

Eje principal 3: Arauca-FLE3, Santander-FEAR5, Arauca-CCN51

Eje principal 4: Arauca-FEC2

Eje principal 5: ningún clon tiene el límite de importancia mínimo

Eje principal 6: Huila-FSV41, Santander-FSV41

Eje principal 7: ningún clon tiene el límite de importancia mínimo

Acto seguido se realizó un análisis clúster para identificar clones con características similares, encontrando los siguientes resultados:

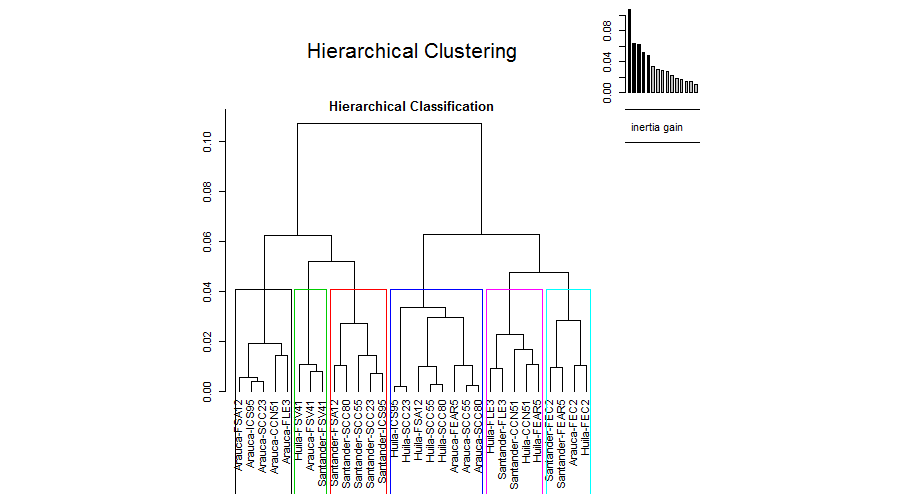


Figura 60. Análisis clúster para individuos

De igual manera, se realizó un análisis clúster para los compuestos volátiles encontrando los siguientes resultados:

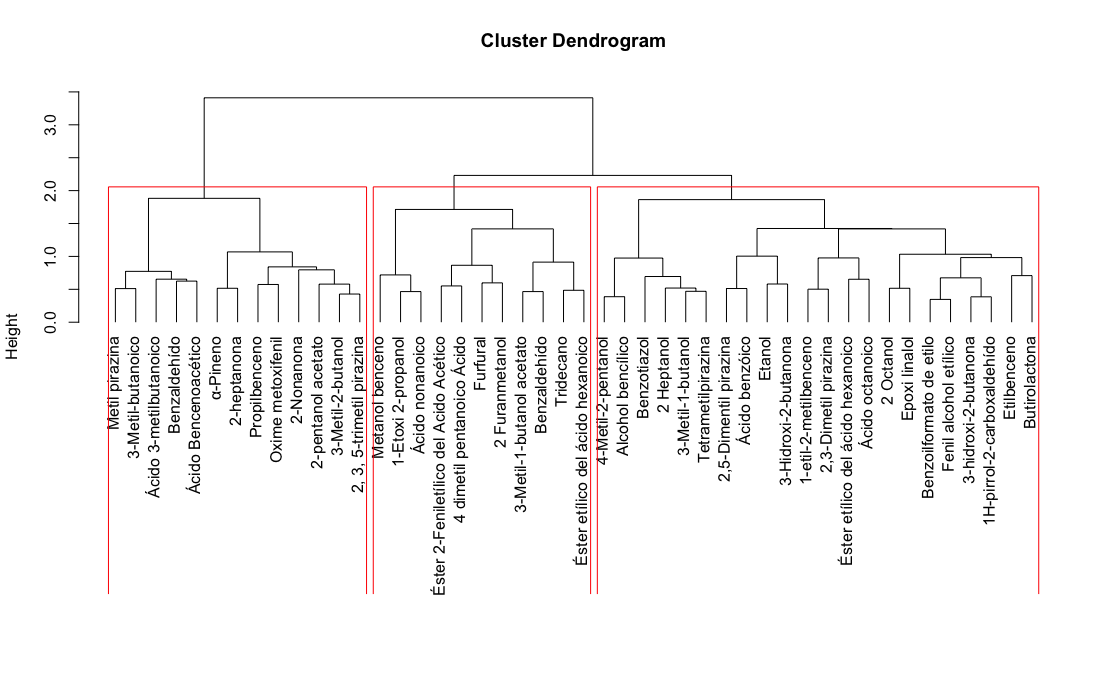


Figura 61. Análisis clúster para variables

Aquí la elección del número de grupos depende de la interpretación que se les dé a las asociaciones entre compuestos volátiles. Se seleccionaron tres clústers, pero depende del criterio del investigador.

# Objetivo 4

Para determinar la influencia genotipo-ambiente en las características fisicoquímicas y funcionales de los clones de cacao se calcularon los coeficientes no paramétricos de Spearman en función del cruce de: condiciones ambientales (clima) y características fisicoquímicas y funcionales. El heatmap o mapa de calor que se muestra en la Figura 62 ilustra la fuerza de la relación entre cada par de variables contrastadas en función del color, de este modo, colores próximos al amarillo indicarían una relación positiva muy alta (no necesariamente lineal), mientras colores próximos al azul oscuro indicarían relaciones negativas altas.

Por otro lado, colores de tonalidad verde indican nula o escasa relación entre variables. Mientras el color gris indica que no se pudo calcular el coeficiente de correlación debido al escaso número de datos. De modo general se puede observar que las variables de clima inciden sobre las diferentes características fisicoquímicas y funcionales de los clones de cacao. Vale la pena resaltar la incidencia del incremento de la lluvia en la reducción del contenido de epicatequina (Correlación de Spearman: -0.65), mientras esta tiene un efecto positivo en el porcentaje de aceite (0.44) y en el contenido de fenoles totales (0.37).

En cuanto a la humedad relativa se puede observar que a medida que esta aumenta se reduce el contenido de los ácidos oleico (-0.64) y linoleico (-0.69). Finalmente evaluando la incidencia de la temperatura máxima se destaca que esta afecta de manera negativa el porcentaje de aceite (-0.60) y el contenido de fenoles totales (-0.62) a diferencia de la lluvia, la cual tiene el efecto contrario. Por otro lado dos características que se ven beneficiadas a medida que aumenta la temperatura máxima son la epicatequina (0.70) y la actividad antioxidante (0.59).

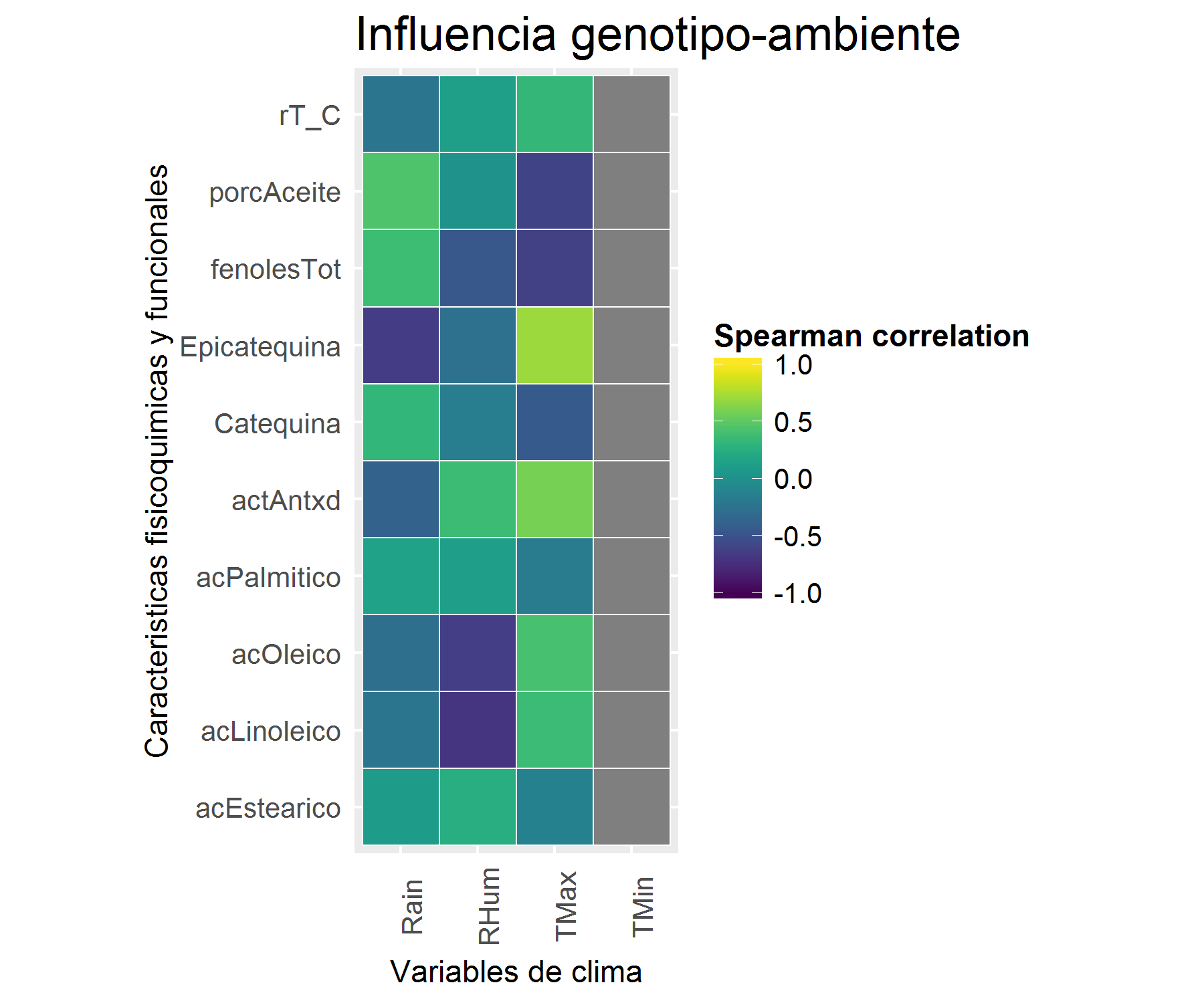


Figura 62. Búsqueda de relaciones significativas entre características fisicoquímicas y funcionales con clima

Para determinar la influencia genotipo-ambiente en las características fisicoquímicas y funcionales de los clones de cacao se realizó el ajuste de regresiones lineales a partir de un muestreo bootstrap, entre cada variable fisicoquímica o funcional con cada indicador climático. El proceso consistió en generar 500 muestras aleatorias a partir de la muestra de datos original y para cada una de ellas se ajustó un modelo de regresión lineal simple utilizando como variable de respuesta cada una de las variables fisicoquímicas o funcionales y como variable predictor cada indicador de clima.

De este modo, el resultado final consiste en el ajuste de 500 modelos lineales por cada par de variables, permitiendo estimar con mayor exactitud la relación existente entre las variables en consideración. Aquí la asignación de estaciones climáticas a los sitios donde se desarrollaron los experimentos, se realizó a partir de los siguientes lineamientos:

* Cercanía de la estación al sitio del experimento
* Menor cantidad de datos faltantes en el periodo de interés (año 2013)

A pesar de definir los lineamientos anteriores, se presentaron casos donde algunas variables de clima medidas en una estación contenían errores como: gran cantidad de datos faltantes o valores repetidos. Estas variables se omitieron del cálculo de los indicadores climáticos con los cuales se ajustaron los modelos de regresión. El periodo de interés consistió en los 6 meses previos a la fecha recolección de los granos y los indicadores de clima definidos fueron:

* ***Temperatura máxima promedio***\*: para su cálculo se usaron los datos diarios de temperatura máxima registrados en la estación, calculando su valor promedio para el periodo de interés.
* ***Temperatura mínima promedio***\*: para su cálculo se usaron los datos diarios de temperatura mínima registrados en la estación, calculando su valor promedio para el periodo de interés.
* ***Temperatura promedio***\*: para su cálculo se usaron los datos diarios de temperatura máxima y mínima registrados en la estación, calculando su valor medio y promediando para el periodo de interés.
* ***Amplitud térmica promedio***\*: para su cálculo se usaron los datos diarios de temperatura máxima y mínima registrados en la estación, calculando su diferencia y promediando para el periodo de interés.
* ***Precipitación acumulada***: para su cálculo se usaron los datos diarios de lluvia registrados en la estación, calculando su suma para el periodo de interés.
* ***Humedad relativa promedio***: para su cálculo se usaron los datos diarios de humedad relativa registrados en la estación, calculando su valor promedio para el periodo de interés.

(\*) Estos indicadores climáticos únicamente fueron calculados para el departamento del Huila para los granos recolectados en la fecha 18 de septiembre de 2013, por ende corresponden a un dato único que no puede ingresar a un modelo de regresión.

A continuación los resultados obtenidos para las variables fisicoquímicas: las variables porcentaje de aceite, contenido de ácido esteárico y contenido de ácido palmítico se ven influenciadas de manera positiva a medida que aumentan las variables de clima. Por otro lado el contenido de ácido oleico y el contenido de ácido linoleico mostraron una tendencia negativa, es decir, a medida que se incrementa la precipitación y la humedad relativa el contenido de estas dos variables se reduce. Mientras la relación tauromina/cafeína permaneció constante sin mostrar ningún cambio respecto la precipitación acumulada y la humedad relativa.

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

Figura 63. Modelos de regresión lineal para variables fisicoquímicas en función del clima

Finalmente, las variables funcionales contenido de catequina y actividad antioxidante muestran una tendencia positiva asociada a los factores climáticos, mientras los fenoles totales y el contenido de epicatequina presentan una relación negativa con respecto a la precipitación total y la humedad relativa.

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |

Figura 64. Modelos de regresión lineal para variables funcionales en función del clima