



Tutorial práctico para mapeo de QTLs

Apellidos, nombre	Villanueva Párraga, Gloria Vilanova Navarro, Santiago Gramazio, Pietro Plazas Ávila, Mariola
Departamento	Instituto de Conservación y Mejora de la Agrodiversidad Valenciana Biotecnología
Centro	Universitat Politècnica de València



Índice general

1	Resumen de las ideas clave.....	2
2	Objetivos.....	2
3	Introducción.....	2
4	Desarrollo.....	3
4.1	Introducción a <i>R/qtl</i>	3
4.2	Preparación de los datos de entrada a los programas.....	4
4.3	Análisis de QTLs en <i>R/qtl</i>	4
4.4	Resultados.....	6
5	Cierre.....	7
6	Bibliografía.....	7



1 Resumen de las ideas clave

El mapeo de loci de caracteres cuantitativos (QTLs) es una técnica eficaz para identificar regiones genéticas asociadas con caracteres complejos, potenciada por la disponibilidad de datos de genotipado y fenotipado de alto rendimiento. Sin embargo, su análisis puede ser complicado, especialmente con poblaciones experimentales no convencionales. R/qtl es una herramienta interactiva útil para este propósito, aunque su interfaz puede resultar complicada para principiantes. Por ello, se propone un tutorial práctico que cubre la preparación de datos, el uso de funciones y parámetros de R/qtl, y la selección de modelos adecuados. El tutorial incluye un ejemplo práctico con retrocruces avanzados de berenjena, que han sido fenotipados y genotipados. R/qtl proporciona resultados útiles como puntuaciones LOD y gráficos de efectos para QTLs identificados, facilitando la comprensión de la genética de caracteres complejos y apoyando programas de mejora y selección vegetal. El enfoque práctico del tutorial permitirá a los usuarios con poca experiencia en programación analizar sus datos y entender mejor los métodos de mapeo de QTL.

2 Objetivos

Este documento está pensado para alumnado con conocimientos previos en genética y pretende:

- Proporcionar una guía práctica como introducción al paquete R/qtl, enfocada a usuarios con experiencia limitada en programación.
- Dilucidar los aspectos básicos de la preparación de datos, incluyendo el formato de archivo de entrada y la información requerida en el programa.
- Instruir al alumnado en cómo utilizar varias de las funciones y parámetros de R/qtl, incluyendo la selección del modelo más apropiado para cada conjunto de datos y métodos para estimar mapas genéticos.
- Presentar un ejemplo práctico con retrocruces avanzados de berenjena con introgresiones de una especie silvestre relacionada, previamente fenotipada y genotipada.
- Mostrar cómo R/qtl proporciona diversos resultados útiles, incluyendo puntuaciones LOD de marcadores moleculares significativos y gráficos de efectos para QTLs identificados, facilitando la interpretación y el análisis de los resultados.

3 Introducción

Los avances en genotipado y fenotipado han incrementado la disponibilidad de marcadores moleculares y datos de alto rendimiento, mejorando la comprensión de la genética de caracteres complejos en plantas. Identificar un locus de caracteres cuantitativos (QTL) es crucial para mejorar la eficiencia de los programas de mejora de cultivos. Sin embargo, los análisis de QTLs son complejos y requieren de herramientas bioinformáticas y modelos estadísticos especializados. El paquete R/qtl es un recurso valioso para realizar análisis de QTLs con precisión y eficiencia.

4 Desarrollo

Este artículo docente ofrece una guía, explicada paso a paso, para realizar análisis de QTLs usando R/qtl, dirigida a principiantes y personas con poca experiencia en programación. La guía cubre la instalación de R/qtl, el uso básico y la preparación de datos de entrada, utilizando un ejemplo práctico con retrocruces avanzados (ABs) de berenjena. El objetivo es proporcionar conocimientos fundamentales para realizar análisis de QTLs y capacitar a los usuarios para aplicar estas técnicas en sus propias investigaciones.

4.1 Introducción a R/qtl

Antes de comenzar el análisis de QTL, debemos instalar R y RStudio:

- R (<https://posit.co/download/rstudio-desktop/>)
- RStudio (<https://posit.co/download/rstudio-desktop/>)

Después del proceso de instalación, los usuarios pueden familiarizarse con la interfaz de R como se muestra en la Imagen 1. La interfaz consiste en un área superior izquierda donde se puede abrir un Script de R navegando a "File > New File > R Script". Aquí es donde se escribirá el código para el análisis de QTL. El área inferior derecha muestra la consola donde se ejecutará el código escrito y se presentarán los resultados. En el área superior derecha, el entorno proporciona una vista general de los archivos añadidos y creados durante el análisis. Finalmente, el área inferior derecha permite a los usuarios instalar los paquetes necesarios para el análisis de QTL y ver los gráficos generados.

Para comenzar a usar el paquete R/qtl, el primer paso es instalarlo. En el área inferior derecha, vaya a la pestaña "Packages" y haga clic en "Install". En la ventana emergente, busque y seleccione "qtl" y haga clic en "Install" para completar el proceso. Una vez instalado el paquete, escriba `library(qtl)` en el Script de R para cargar el paquete.

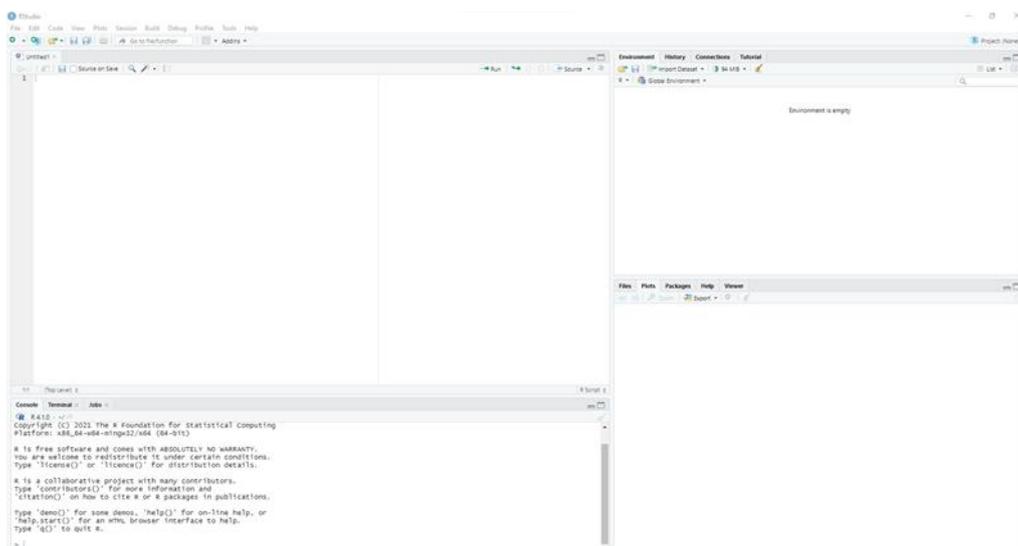


Imagen 1. Información general sobre el espacio de trabajo de RStudio.

4.2. Preparación de los datos de entrada a los programas

El archivo de entrada para realizar el análisis con R/qtl debe estar en formato CSV y seguir una estructura específica. Las primeras columnas del archivo de entrada deben contener los datos fenotípicos, con cada carácter representado en una columna separada. La fila del encabezado debe contener el nombre del carácter seguido por los valores medidos para cada accesión. Se pueden incluir múltiples caracteres, cada uno ocupando su respectiva columna. A continuación, deben añadirse los datos genotípicos en sucesivas columnas. Para cada marcador, la primera fila debe contener el nombre del marcador, la segunda fila debe indicar el cromosoma al que pertenece el marcador y la tercera fila debe especificar la posición del marcador. Los datos genotípicos para cada accesión deben proporcionarse en las filas siguientes. Para hacer el análisis más eficiente, los datos genotípicos pueden prefiltrarse usando otros programas, como el software TASSEL (Trait Analysis by aSSociation, Evolution and Linkage), versión 5.0 (ver objeto de aprendizaje <http://hdl.handle.net/10251/193564>). Este paso de preprocesamiento puede ayudar a optimizar los datos de entrada para el análisis con R/qtl. Un ejemplo del formato del archivo de entrada puede verse en la Imagen 2, que ilustra cómo deben organizarse los datos para su procesamiento eficiente por el paquete R/qtl.

Fenotipos		Genotipos					
A	B	C	D	E	F	G	H
Prod	F-Weight	SSMEL3CH01	SSMEL3CH01	SSMEL3CH01	SSMEL3CH01	SSMEL3CH01	SSMEL3CH01
		1	1	1	1	1	1
		10192176	11614810	11614831	21355012	21355034	22429440
4	793 46.65	B	H	H	A	B	B
5	2063 71.14	B	H	H	A	B	B
6	2524 58.7	B	H	H	A	B	B
7	2343 41.84	B	H	H	A	B	B
8	1625 34.57	B	H	H	A	B	B
9	2135 38.82	B	H	H	A	B	B
10	2257 66.38	H	H	H	A	B	B
11	2981 76.44	B	H	H	A	B	B
12	1261 63.05	B	H	H	A	B	B
13	1324 50.92	B	H	H	A	B	B
14	2115 34.67	B	H	H	A	B	B
15	2709 48.38	B	H	H	A	B	B
16	3139 58.13	B	H	B	A	B	B
17	3869 82.32	B	H	B	A	B	B
18	2019 61.18	B	H	B	A	B	B
19	2327 75.06	B	H	B	A	B	B
20	1791 57.77	B	H	B	A	B	B
21	1354 54.16	B	H	B	A	B	B
22	1477 54.7	B	H	B	A	B	B
23	1059 42.36	B	H	B	A	B	B
24	1616 50.5	B	H	B	A	B	B
25	1682 50.97	B	H	B	A	B	B
26	884 44.2	B	H	B	A	B	B
27	1817 60.57	B	B	B	A	B	B
28	2356 60.41	B	B	B	A	B	B

Imagen 2. Muestra del archivo de entrada CSV con datos fenotípicos y genotípicos en R/qtl.

4.3. Análisis de QTLs en R/qtl

En este artículo docente se va a explicar el uso del paquete R/qtl para el análisis de QTLs con ejemplos sencillos para cada paso. Podemos encontrar un manual con más información del paquete R/qtl en el siguiente link (<https://rqt.org/manual/qtl-manual.pdf>).

Importación de datos

Para importar los datos para el análisis con R/qtl, se utiliza la función 'read.cross'. Al usar esta función, es necesario especificar el formato del archivo de entrada (por ejemplo, format = "csv"), el nombre del archivo (por ejemplo, file = "RQTL.csv"), los genotipos presentes (por ejemplo, genotypes = c("A", "H", "B")) y los alelos (por ejemplo, c("A", "B")) para indicar los genotipos de las líneas parentales y el híbrido. Además, deben indicarse las



generaciones de retrocruzamiento (por ejemplo, BC.gen = 4) y autofecundación (por ejemplo, F.gen = 1) en las que se encuentran los accesos incluidos en el análisis.

Cálculos intermedios

Antes de realizar los análisis de QTLs, se deben realizar algunos cálculos intermedios. Estos cálculos implican la estimación de probabilidades genotípicas condicionales dados los datos de marcadores de múltiples posiciones. Para ello, se utiliza la función 'calc.genoprob'. El archivo de entrada importado previamente debe proporcionarse como una variable a la función 'calc.genoprob'. Además, se deben especificar varios parámetros, incluyendo el 'step', que es la distancia máxima en cM entre posiciones en las que se calculan las probabilidades genotípicas. En el ejemplo, el 'step' se establece en 0, lo que significa que las probabilidades se calculan solo en las ubicaciones de los marcadores. El parámetro 'off.end' se refiere a la distancia en cM más allá de los marcadores terminales en cada cromosoma hasta la cual se llevarán a cabo los cálculos de probabilidad de genotipo (por ejemplo, off.end=0.0). El parámetro 'error.prob' representa la tasa de error de genotipado asumida utilizada en el cálculo de la penetrancia, que es la probabilidad de observar un genotipo dado el genotipo verdadero (por ejemplo, error.prob=1.0e-4). También se debe especificar la función de mapa elegida para convertir distancias genéticas en fracciones de recombinación, con opciones que incluyen Haldane, Kosambi o Carter-Falconer. Finalmente, el parámetro 'stepwidth' indica si los puntos intermedios deben tener tamaños de paso fijos o variables (por ejemplo, stepwidth="fixed").

Cálculo de probabilidades genotípicas

El siguiente paso implica calcular las probabilidades genotípicas utilizando la función 'scanone'. Esta función realiza un escaneo a lo largo del genoma con un modelo de QTL único con posible consideración de covariables, utilizando cualquiera de varios modelos posibles para el fenotipo y varios métodos numéricos posibles. El parámetro para esta función es la variable en la que se ha guardado el objeto cross. Es necesario indicar el número de la columna donde se encuentra el fenotipo en el archivo de entrada (por ejemplo, pheno.col=1). El modelo de fenotipo se puede especificar como el modelo normal usual, un modelo para rasgos binarios, un modelo de dos partes o un análisis no paramétrico (por ejemplo, model="normal"). El método a utilizar se puede especificar como si se debe usar el algoritmo EM, imputación, regresión Haley-Knott, el método extendido Haley-Knott o regresión de marcadores (por ejemplo, method="em").

Además, la función 'scanone' se utiliza nuevamente, pero añadiendo el parámetro 'n.perm', que especifica el número de réplicas por permutación a realizar (por ejemplo, n.perm=100). Los resultados de la prueba de permutación proporcionan valores LOD y probabilidades para cada QTL, lo que permite la identificación de QTLs significativos. La función 'summary' se puede usar para mostrar los resultados del escaneo del genoma y la prueba de permutación, con el argumento alpha especificando el(los) umbral(es) de significancia a utilizar (por ejemplo, alpha=c(0.6, 0.10, 0.05)). Un ejemplo de script puede observarse en la Imagen 3.

Visualización de los resultados del análisis de QTL

Los resultados del análisis de QTL se pueden visualizar utilizando la función 'plot', que muestra los puntajes LOD obtenidos para diferentes marcadores en diferentes cromosomas o en un solo cromosoma si se especifica (por ejemplo, chr=3). Los gráficos se pueden ver en la pestaña "Plots" en la esquina inferior derecha del espacio de trabajo. Además, se puede obtener un resumen de los marcadores significativos obtenidos en el análisis utilizando la

función 'summary', especificando la variable que contiene los valores de permutación y el valor alpha (por ejemplo, alpha=0.05). En el caso de un QTL significativo, el efecto del marcador significativo sobre el fenotipo se puede visualizar utilizando un gráfico de efectos. Para identificar el marcador asociado con el QTL significativo, se utiliza la función 'find.marker' con el archivo de datos importado, el número de cromosoma y la posición del marcador como entradas (por ejemplo, RQTL, chr=1, pos=130615804). El gráfico de efectos se genera utilizando la función 'effectplot', especificando el marcador y el número de columna del fenotipo. Este gráfico puede proporcionar información valiosa sobre el efecto del QTL sobre el fenotipo, lo que puede ser útil para un análisis e interpretación adicionales de los resultados.

```
#2.3 QTL analysis in R/qtl
#2.3.1 Importing data
RQTL <- read.cross("csv", sep=";", file= "RQTL.csv", genotypes = c("A","H", "B"),
                 alleles = c("A","B"), BC.gen = 4, F.gen = 1)

#2.3.2 Intermediate calculations
RQTL.cal<- calc.genoprob(RQTL, step=0, off.end = 0.0, error.prob = 1.0e-4,
                       map.function = "kosambi", stepwidth = "fixed")

#2.3.3 Calculating genotype probabilities
yield.scan1 <- scanone(RQTL.cal, pheno.col = 1, method="em")
yield.scan1.perm <- scanone(RQTL.cal, pheno.col = 1, method="em", n.perm = 100)
summary(yield.scan1.perm, alpha=c(0.6, 0.10, 0.05))

#2.3.4 Visualizing QTL analysis results
plot(yield.scan1, ylim=c(0,4), main="Yield")
plot(yield.scan1, ylim=c(0,3), chr=2)
summary(yield.scan1, perm=yield.scan1.perm, lodcolumn = 1, alpha=0.5)

mar <- find.marker(RQTL, chr=2, pos= 80288728)
plotyield <- plotPXG(RQTL, marker=mar, pheno.col = 1)
plotyieldeffect<-effectplot(RQTL, mname1=mar, pheno.col = 1)
```

Imagen 3. Código de ejemplo en formato R Script para el análisis de QTL mediante el paquete R/qtl.

4.4 Resultados

El análisis de QTL mediante R/qtl se realizó en un conjunto ABs de berenjena (*S. melongena*) con introgresiones de un pariente silvestre, previamente fenotipados y genotipados. Durante el análisis se obtuvo un resumen de los valores de permutación (Tabla 1) y un resumen del análisis de marcadores significativos del QTL (Tabla 1). Se identificó un marcador molecular significativo asociado a un carácter complejo, como se refleja en las puntuaciones LOD obtenidas del análisis.

Tabla 1. Ejemplo resumen de los valores de permutación (izquierda) y del marcador significativo del análisis QTL (derecha).

LOD puntuación (100 permutaciones)					
%	LOD	Marker	chr	pos	LOD
60%	1.71	SSMEL3CH02_80288728	2	80288728	3.23
10%	2.54				
5%	2.82				

Se generó un gráfico de las puntuaciones LOD para diferentes marcadores en diferentes cromosomas, proporcionando una representación visual de los resultados (Imagen 4). Además, el efecto del QTL sobre el fenotipo se visualizó mediante un gráfico de efectos (Imagen 4), que proporcionó información valiosa para interpretar los resultados y diseccionar las bases genéticas de caracteres complejos.

Estos resultados demuestran el potencial del mapeo de QTL utilizando R/qtl para dilucidar la base genética de caracteres complejos en poblaciones experimentales, que se pueden utilizar para informar los programas de mejora y selección.

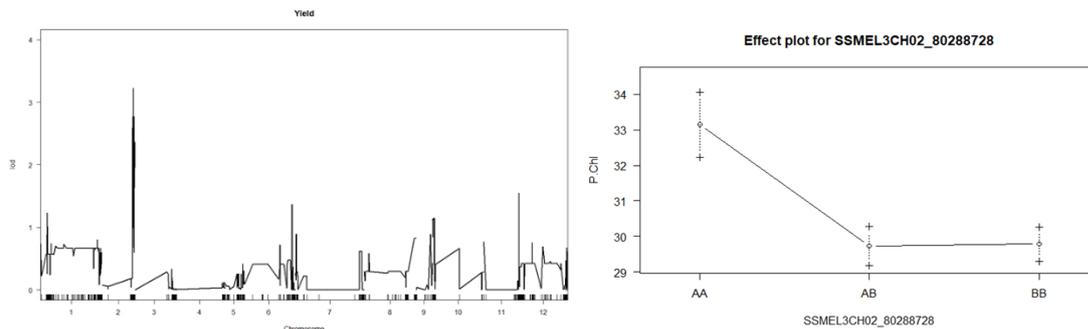


Imagen 4. Ejemplo de diagrama cromosómico de puntuaciones LOD (izquierda) y gráfico del efecto del marcador significativo (derecha)

5. Cierre

El uso de R/qtl para el mapeo de QTL en poblaciones experimentales puede ayudar en gran medida en la identificación de regiones genéticas asociadas con caracteres complejos, proporcionando información sobre la base genética de estos caracteres que se puede utilizar para informar los programas de mejora y selección. El tutorial práctico proporcionado en este artículo docente tiene como objetivo equipar a los principiantes con los conocimientos fundamentales necesarios para realizar análisis QTL utilizando R/qtl, permitiéndoles explorar y aplicar más a fondo estas técnicas en sus propios estudios de investigación.

6. Bibliografía

Würschum, T. 2012. "Mapping QTL for agronomic traits in breeding populations", *Theoretical Applied Genetics*, vol. 125, pp. 201-210.

Cobb, J. N., Biswas, P. S., Platten, J. D. 2019. "Back to the future: revisiting MAS as a tool for modern plant breeding", *Theoretical Applied Genetics*, vol. 132, pp. 647-667.

Broman, K. W., Wu, H., Sen, S., Churchll, G. A. 2003. "R/qtl: QTL mapping in experimental crosses", *Bioinformatics*, vol. 19, pp. 889-890.

R Core Team R: a language and environment for statistical computing, 2021. Retrieved from: <https://www.r-project.org/>

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., Buckler, E. S. 2007. "TASSEL: Software for association mapping of complex traits in diverse samples", *Bioinformatics*, vol. 23, pp- 2633-2635.