



# Tutorial para la visualización cromosómica de material con introgresiones utilizando TASSEL y GGT2

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



## Índice general

1.	Resumen de las ideas clave.....	2
2.	Objetivos.....	2
3.	Introducción.....	2
4.	Desarrollo.....	3
4.1.	Preparación del software y uso de TASSEL .....	3
4.2.	Adaptación de datos entre TASSEL y GGT2.....	4
4.3.	Visualización cromosómica con introgresiones usando GGT2.....	5
5.	Cierre .....	6
6.	Bibliografía .....	6



## 1 Resumen de las ideas clave

El desarrollo de poblaciones experimentales, como las líneas de introgresión (IL), es fundamental para estudiar los caracteres de interés en mejora. Para seleccionar e identificar eficientemente los caracteres deseables en estas poblaciones, es esencial una correcta visualización de los datos de genotipado de plantas. Sin embargo, obtener genotipos gráficos de cromosomas con diferentes introgresiones puede ser complejo debido a la gran cantidad de datos y la necesidad de adaptar estos datos a los formatos requeridos por el software de análisis. Para resolver este problema, presentamos una guía paso a paso para generar genotipos gráficos de cromosomas con introgresiones usando los programas Análisis de Caracteres por Asociación, Evolución y Ligamiento (TASSEL) y Genotipos Gráficos (GGT2). Usaremos como caso de estudio el genotipado de un conjunto de retrocruzamientos avanzados (ABs) con introgresiones de un pariente silvestre de berenjena en el fondo genético de la berenjena cultivada (*Solanum melongena*), visualizando los 12 cromosomas del genoma de la berenjena. En el tutorial, proporcionamos imágenes de los programas para facilitar su uso y explicamos cómo seleccionar los parámetros adecuados para cada análisis y visualización. Primero, utilizamos TASSEL para filtrar los datos de genotipado e identificar las introgresiones de interés. Luego, empleamos GGT2 para generar representaciones gráficas de los cromosomas con las introgresiones destacadas. El tutorial también incluye información clave sobre la adaptación intermedia necesaria entre los dos programas, ejemplos prácticos y consejos para superar las dificultades comunes en el proceso. Esta guía está diseñada para estudiantes con conocimientos básicos de genotipado, análisis de datos y herramientas bioinformáticas. La metodología presentada puede mejorar la precisión y eficiencia en la generación de genotipos gráficos de cromosomas para estudios de fitomejoramiento, evolución del genoma y diversidad genética

## 2 Objetivos

Una vez que el alumnado se lea con detenimiento este documento, será capaz de:

- Analizar datos genotípicos utilizando el programa informático TASSEL.
- Filtrar e imputar datos de marcadores SNPs.
- Identificar agrupaciones mediante análisis de componentes principales (PCA).
- Realizar un estudio de asociación de genoma completo (GWAS) para detectar asociaciones genómicas.
- Generar visualizaciones cromosómicas con introgresiones de otras especies usando GGT2.

## 3 Introducción

La mejora genética vegetal es crucial para aumentar el rendimiento de los cultivos, garantizar la seguridad alimentaria mundial y enfrentar desafíos ambientales. Las líneas de introgresión (IL), como poblaciones experimentales, son recursos valiosos para identificar genes y loci de caracteres cuantitativos (QTLs) asociados con características deseables. Las



IL se desarrollan para incorporar segmentos genómicos de un progenitor donante, a menudo un pariente silvestre, en el fondo genético de un progenitor recurrente cultivado. Esto introduce nuevos alelos y diversidad genética, permitiendo explorar la base genética de caracteres complejos y mejorar los cultivos. Para llevar a cabo esta tarea, deberemos ser capaces de analizar datos de genotipado de poblaciones experimentales, lo que puede ser desafiante por la complejidad de los datos y la necesidad de una visualización eficaz. Los genotipos gráficos de cromosomas con diferentes introgresiones permiten analizar e interpretar la composición genómica de las ILs. Estas visualizaciones ayudan a identificar regiones genómicas responsables de caracteres específicos, monitorear la introgresión y diseñar estrategias óptimas de cruzamiento entre plantas.

Generar estos genotipos cromosómicos puede ser complejo y llevar tiempo, requiriendo softwares y formatos de datos adecuados. TASSEL (Análisis de Caracteres por Asociación, Evolución y Vinculación) y GGT2 (Genotipos Gráficos) son herramientas útiles para esto. TASSEL se usa para evaluar asociaciones de caracteres, patrones evolutivos y desequilibrio de ligamiento, mientras que GGT2 se centra en generar representaciones gráficas de cromosomas con introgresiones. Por todo ello, este tutorial presenta una guía paso a paso para generar genotipos gráficos cromosómicos con diferentes introgresiones usando TASSEL y GGT2.

## 4 Desarrollo

En este artículo docente se proporciona un método accesible y práctico para visualizar la composición genética de las ILs y otras poblaciones experimentales, facilitando la comprensión de las bases genéticas de caracteres complejos y promoviendo estrategias eficientes de mejora vegetal.

### 4.1 Preparación del software y uso de TASSEL

El software TASSEL es una herramienta versátil para el análisis de datos genotípicos. Para comenzar, es necesario cargar los datos en formato VCF (Variant Call Format). Una vez cargados los datos, se procede al filtrado de SNPs para asegurar la calidad del análisis.

Para realizar el proceso de filtrado, vaya a Filtrar > sitios > Alinear filtros en TASSEL. Aunque se pueden aplicar diferentes criterios de filtrado en función de la población específica y los objetivos de investigación, un enfoque estándar sugerido incluye establecer un valor mínimo de recuento del 97 %, una frecuencia alélica mínima del 5 % (0,05) y una proporción máxima de heterocigosidad del 70 % (0,7). Estos parámetros permiten el refinamiento inicial de los marcadores moleculares para obtener un conjunto de datos más manejable para su posterior análisis (Imagen 1). Además, se puede realizar un filtrado más estricto estableciendo una distancia mínima en pares de bases (pb) entre sitios adyacentes (Data > Thin Sites by Position). Especifique el archivo de salida que se generará y, como sugerencia, establezca la distancia mínima entre sitios adyacentes en 2.000 pb. Este paso de filtrado adicional refina aún más los marcadores moleculares, lo que garantiza que los marcadores poco espaciados no influyan de manera desproporcionada en el análisis y la visualización de los datos de genotipado.

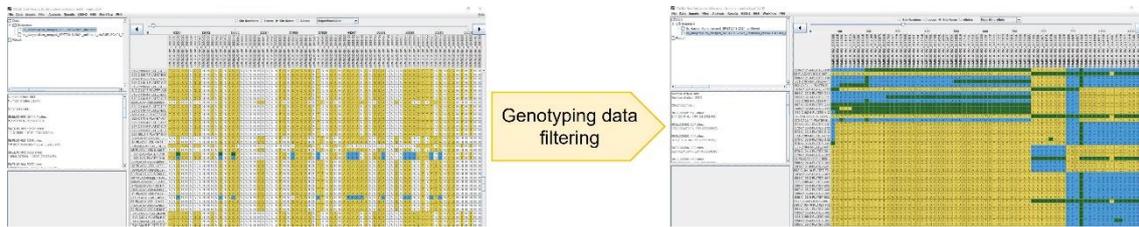


Imagen 1. Comparación de las posiciones de los SNPs antes (izquierda) y después (derecha) del filtrado.

## 4.2 Adaptación de datos entre TASSEL y GGT2

Una vez finalizado el filtrado de marcadores moleculares, debemos obtener los datos de genotipado en formato ABH (Data > ABH Genotype). Proporcionaremos el nombre del archivo de salida y especificaremos los nombres de ambas líneas principales en el formato de texto (.txt). Este proceso generará un archivo de texto que contiene los datos de genotipado de cada individuo, utilizando "A" para representar una línea parental, "B" para la otra línea parental y "H" para los híbridos. Para preparar los datos para su uso en GGT2, primero abra una hoja de cálculo de Excel y pegue el contenido del archivo de texto ABH en ella. Debemos separar los marcadores en columnas a través de la ruta Datos > Texto a columnas > Delimitado > Coma.

Los datos de los individuos deben estar en columnas y los marcadores en filas. Organice los datos como se muestra en la Imagen 2, en la primera columna los nombres de los marcadores, en la segunda el número de cromosomas ("Chrom"), y en la tercera se indica la posición en Mb ("pos") y las columnas posteriores contienen los nombres deseados para los individuos tal como aparecerán en la imagen, junto con sus datos de genotipado correspondientes ("alelos").

Siguiendo estos pasos, los datos de genotipado se reformatearán para cumplir con los requisitos de GGT2, lo que permitirá una integración y visualización adecuada de los datos utilizando el software GGT2.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	
1		Chrom	pos	alleles	ELE 24-1	ELE 24-5	ELE 89-4	ELE 96-4	ELE 96-6	ELE 58-2	ELE 58-8	ELE 35-2	ELE 35-5	ELE 35-6	ELE 21-1	ELE 59-3	ELE 59-5	ELE 90-4	ELE 90-6	ELE 90-3	ELE 90-5	ELE 42-1
2	aliases																					
3	alternative layout=true																					
4																						
5	name=alternative																					
6	SSMEL3CH01_510967	1	0.510967	B	B	H	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
7	SSMEL3CH01_663715	1	0.663715	B	B	H	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
8	SSMEL3CH01_683014	1	0.683014	B	B	H	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
9	SSMEL3CH01_753880	1	0.753880	B	B	H	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
10	SSMEL3CH01_1053533	1	1.053533	B	B	H	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
11	SSMEL3CH01_1419780	1	1.419780	B	B	H	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
12	SSMEL3CH01_1801626	1	1.801626	B	B	H	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
13	SSMEL3CH01_1865382	1	1.865382	B	B	H	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
14	SSMEL3CH01_1925859	1	1.925859	B	B	H	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
15	SSMEL3CH01_1977665	1	1.977665	B	B	H	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
16	SSMEL3CH01_2193786	1	2.193786	B	B	H	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
17	SSMEL3CH01_2208918	1	2.208918	B	B	H	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
18	SSMEL3CH01_2407422	1	2.407422	B	B	H	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
19	SSMEL3CH01_2542770	1	2.542770	B	B	H	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
20	SSMEL3CH01_2584608	1	2.584608	B	B	H	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
21	SSMEL3CH01_2605013	1	2.605013	B	B	H	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
22	SSMEL3CH01_2613197	1	2.613197	B	B	H	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
23	SSMEL3CH01_2665851	1	2.665851	B	B	H	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
24	SSMEL3CH01_2698833	1	2.698833	B	B	H	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
25	SSMEL3CH01_3046146	1	3.046146	B	B	H	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
26	SSMEL3CH01_3059932	1	3.059932	B	B	H	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
27	SSMEL3CH01_3087869	1	3.087869	B	B	H	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
28	SSMEL3CH01_3106753	1	3.106753	B	B	H	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
29	SSMEL3CH01_3312254	1	3.312254	B	B	H	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
30	SSMEL3CH01_3714804	1	3.714804	B	B	H	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
31	SSMEL3CH01_3721980	1	3.721980	B	B	H	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
32	SSMEL3CH01_3751772	1	3.751772	B	B	H	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
33	SSMEL3CH01_3786137	1	3.786137	B	B	H	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
34	SSMEL3CH01_3797453	1	3.797453	B	B	H	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
35	SSMEL3CH01_3865409	1	3.865409	B	B	H	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B

Imagen 2. Formateo en Excel de datos de genotipado para visualización de GGT2.





Imagen 4. Ejemplo de representación gráfica del genoma de retrocruzamientos avanzados de berenjena obtenidos mediante el software GGT2.

Los genotipos gráficos generados proporcionan información valiosa sobre la distribución de las introgresiones, ayudando a los usuarios en el análisis de la diversidad genética, los esfuerzos de mejora y los estudios de selección del genoma.

## 5 Cierre

A lo largo de este objeto de aprendizaje se ha proporcionado un método accesible y eficiente para generar genotipos gráficos cromosómicos con diferentes introgresiones utilizando los softwares que tienen numerosas funcionalidades, por lo que sería posible utilizarlos para otros fines o para el análisis de muestras de distintos orígenes. Además, el uso combinado de TASSEL y GGT2 permite una gestión eficiente de los datos genotípicos, facilitando el análisis y la visualización de introgresiones cromosómicas. El enfoque paso a paso permite a los estudiantes visualizar de manera efectiva datos de genotipado complejos, lo que facilita la identificación y selección de caracteres deseables en poblaciones experimentales. En última instancia, esta metodología contribuye al avance de la mejora vegetal y al estudio de la evolución del genoma y la diversidad genética y proporciona un enfoque práctico y accesible.

## 6 Bibliografía

Prohens, J., et al. (2017). "Introgressomics: a new approach for using crop wild relatives in breeding for adaptation to climate change". *Euphytica*, 213, 158.

Broman, K.W., et al. (2003). "R/qtl: QTL mapping in experimental crosses". *Bioinformatics*, 19, 889-890.

Van Berloo, R. (2008). "GGT 2.0: Versatile software for visualization and analysis of genetic data". *Journal of Heredity*, 99, 232-236.

Reuscher, S., et al. (2015). "First annual TASSEL hackathon". (<https://tassel.bitbucket.io/>)