

Prueba de proceso de GWAS con varios enfoques y múltiples modelos

15 de octubre de 2019

1. Enfoques y Modelos

Para las pruebas se adaptó el pipeline propuesto para GWAS ahora con elementos del software GWASpoly para polyploides. Las pruebas se realizaron para cuatro enfoques de GWAS y cuatro modelos de acción del gen.

1.1. Enfoques

Los enfoques corresponden a la manera de realizar el proceso GWAS, es decir, si se realiza con corrección de estructura poblacional o sin ella. Los siguientes son los 4 modelos ejecutados:

- **Naive (N):** GWAS por defecto, sin adicionar parentesco ni corrección de estructura poblacional. La matriz de parentesco que se construyó para este enfoque es una matriz identidad, donde cada fila, columna corresponde a un marcador y cada marcador tiene parentesco solo consigo mismo (valor 1 en la diagonal).

$$\begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

- **Kinship (K):** GWAS sólo con parentesco, sin corrección de estructura poblacional. La matriz de parentesco se construyó utilizando las funciones del software GWASpoly (*función set.K*).
- **Structure (Q):** GWAS sólo con corrección de estructura poblacional. Para esto se utilizaron dos tipos de datos: la estructura poblacional calculada por Berdugo2016; y los componentes principales (PCs) calculados por la misma herramienta GWASpoly, con número de PCs=20).
- **Componentes Principales (P):** GWAS con corrección de estructura poblacional usando los componentes principales calculados por el software GWASpoly. Se calculan 20 componentes
- **K+Q, K+P, Q+P, K+Q+P:** GWAS con combinaciones de los dos anteriores enfoques.

1.2. Modelos de acción de los marcadores

Estos modelos se refieren a la forma como los alelos de un marcador contribuyen a la producción de fenotipos cualitativos. El software GWASpoly tienen la característica de permitir modelar diferentes tipos de acciones de gen poliploide y de acuerdo a sus autores [XXX], esta característica permite elevar su poder estadístico para detectar el modelo verdadero implícito en el genotipo, es decir se aumenta el poder estadístico cuando la acción de gen asumida en el modelo GWAS coincide con la acción del gen en los lugares de caracteres cuantitativos (QTL) no observados [XXX].

De acuerdo al manual del software GWASpoly, los modelos que se pueden realizar son:

- **General:** Efecto de cada clase de genotipo (AAAA, ..., BBBB) se considera arbitraria y estadísticamente equivalente.
- **Additive:** Efecto del SNP es proporcional a la dosis del alelo menor.
- **2-Dom-X:** Dominante doble donde el estado duplex (AABB) tiene el mismo efecto que los otros estados (AAAB, AAAA, AB BB, BBBB). X se refiere al alelo por referencia (como aparece en el genoma) o alterno.
- **1-Dom-X:** Dominante simple donde todos los heterocigotos (AAAB, AABB, AB BB) son equivalentes a uno de los homocigotos (AAAA o BBBB).

2. Datos y Métodos

2.1. Genotipo, fenotipo, y estructura poblacional

Se utilizó el genotipo, fenotipo, y estructura poblacional de papa obtenidos en Agrosavia. Los datos de genotipo y fenotipo se preprocesaron para unificar tanto nombres de marcadores como nombres de muestras. Igualmente, las nombres de las muestras del fenotipo se unificaron con los nombres de los datos de estructura poblacional.

2.2. Pipeline para GWAS de papas tetraploides

Todas las pruebas se realizaron usando el pipeline propio para GWAS de papas poliploides, el cual está implementado en R y que está usando principalmente para implementar sus distintas etapas las funciones de la librería GWASpoly.

3. Resultados

Los resultados se muestran de acuerdo a los tres tipos de corrección para evaluación de hipótesis múltiples disponibles en el software GWASpoly. Para cada tipo de corrección se muestran a través de una tabla y dos figuras que relaciona los modelos de GWAS evaluados (Naive, Kinship, Structure, Kinship+Structure) y los modelos de acción de gen disponibles en el software GWASpoly (general, aditivo, 1-dominante, 2-dominante).

3.1. Corrección por Bonferroni

3.1.1. Bondad de los diferentes modelos

La figura muestra a través de gráficos Q-Q la bondad de los diferentes modelos GWAS para el rasgo "gota" de los valores $-\log_{10}(p)$ observados vs. los esperados.

3.1.2. SNPs significantes

Y la segunda figura muestra a través de gráficos Manhattan el umbral de significancia para cada locus en su ubicación genómica. Este umbral (línea horizontal punteada) está basado sobre la tasa de falsos positivos en todo el genoma ($\alpha=0.05$) establecida mediante el método de la corrección Bonferroni implementado en el software GWASpoly. Los marcadores (SNPs) asociados que cruzan este umbral corresponden a los QTLs significativos descritos en la tabla anterior.

3.1.3. QTLs significativos

La tabla resume los QTLs (Quantitative Trait Locus) más significativos encontrados para los diferentes modelos de GWAS.

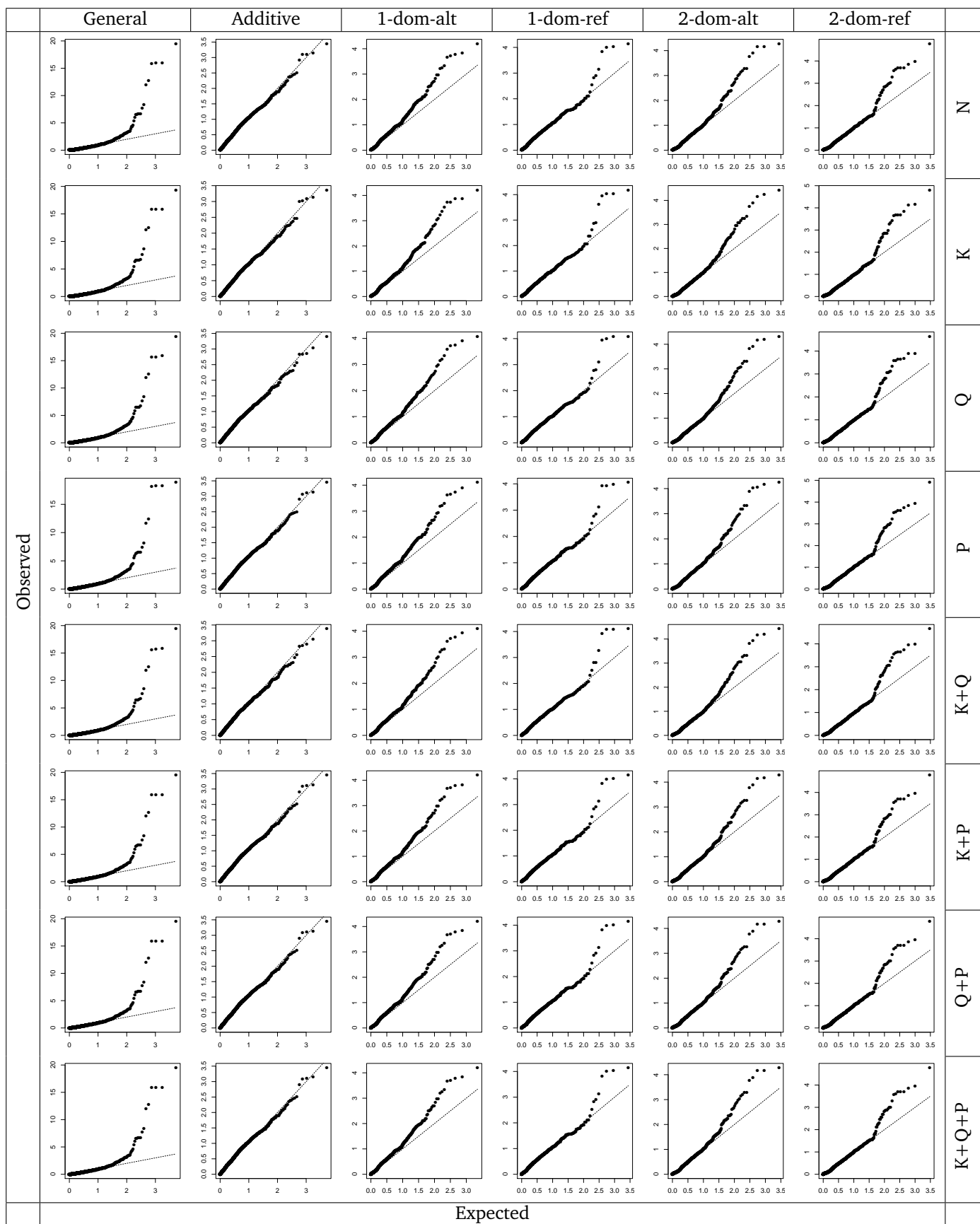


Figura 1: Comparación de gráficos Quartile-Quartile (QQ plots) para los 6 modelos de GWAS (Naive, Kinship, Structure, Structure+Kinship) y los 4 modelos de acción de gen disponibles en el software GWASpoly (general, additive, 1-dom, 2-dom).

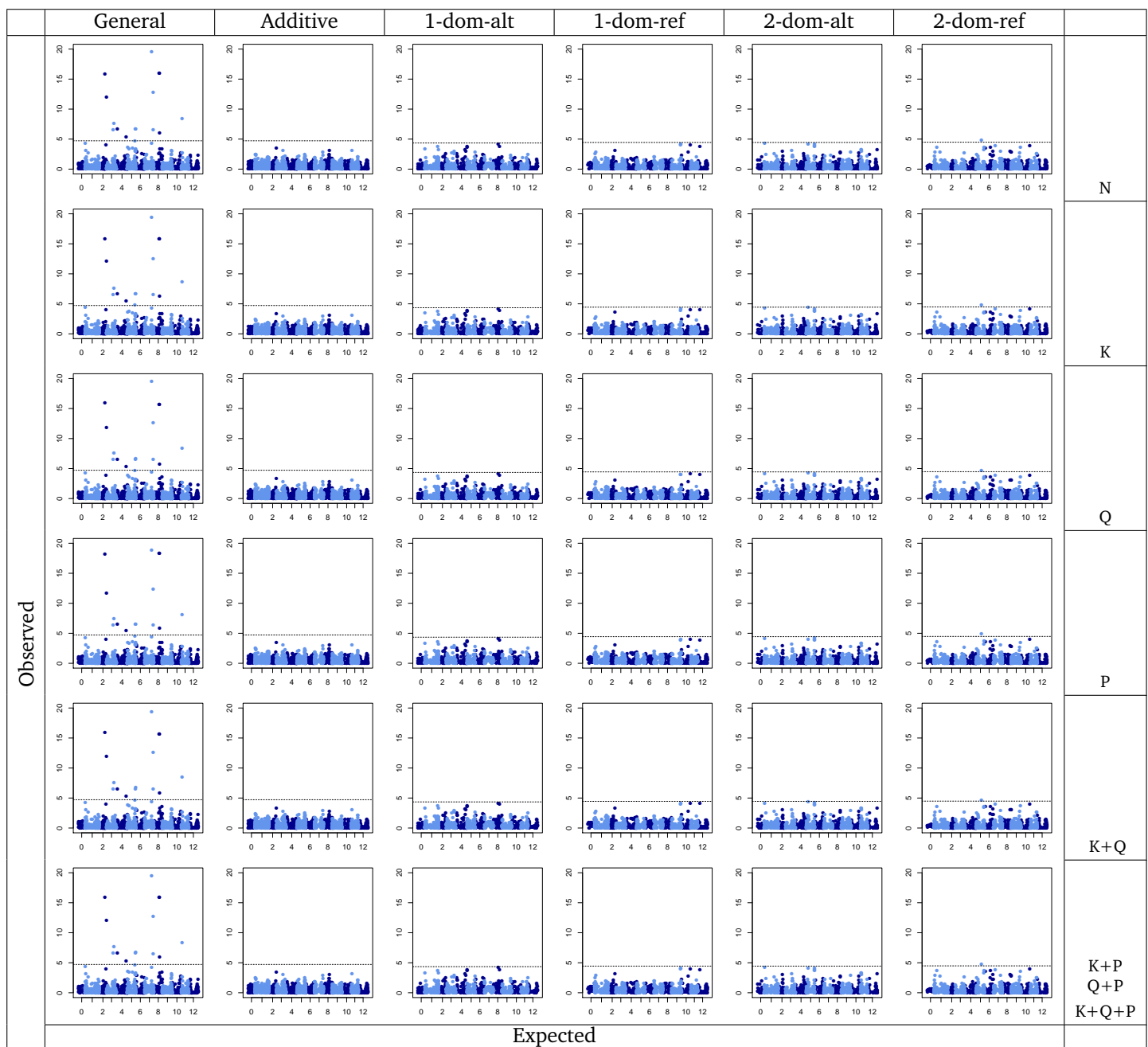


Figura 2: Comparación de gráficos Manhattan para los 4 modelos de GWAS (Naive, Kinship, Structure, Structure+Kinship) y los 4 modelos de acción de gen disponibles en el software GWASpoly (general, additive, 1-dom, 2-dom).

Cuadro 1: QTLs significantes del model Naive vs. modelo de acción de gen.

Rasgo	Modelo	Umbral	Marcador	Crom	Posición	Ref	Alt	Score	Effect
gota	general	4.72	solcap_snp_c1_12386	2	37411847	0	1	6.95	NA
gota	general	4.72	solcap_snp_c1_5895	2	48564909	0	1	7.54	NA
gota	general	4.72	solcap_snp_c2_32854	5	10418984	0	1	5.85	NA
gota	general	4.72	solcap_snp_c2_15929	7	43192596	0	1	8.87	NA
gota	general	4.72	solcap_snp_c2_12404	7	52544646	0	1	4.74	NA
gota	general	4.72	solcap_snp_c2_12601	7	53106007	0	1	4.9	NA
gota	general	4.72	solcap_snp_c2_49245	8	35651378	0	1	7.7	NA
gota	general	4.72	solcap_snp_c2_41463	8	35922019	0	1	7.71	NA
gota	general	4.72	solcap_snp_c1_821	8	37494529	0	1	8.18	NA
gota	general	4.72	solcap_snp_c2_2998	9	58689196	0	1	9.67	NA
gota	general	4.72	solcap_snp_c1_11246	11	4745919	0	1	12.76	NA
gota	1-dom-alt	4.35	solcap_snp_c2_36664	1	535454	0	1	5.15	-0.71
gota	1-dom-alt	4.35	solcap_snp_c2_4875	1	83177518	0	1	5.18	-0.76
gota	1-dom-alt	4.35	solcap_snp_c1_3484	4	71945107	0	1	5.08	-0.78
gota	1-dom-ref	4.44	solcap_snp_c1_7325	2	42764457	0	1	4.52	0.73
gota	1-dom-ref	4.44	solcap_snp_c2_40748	12	12654160	0	1	4.86	0.66
gota	2-dom-alt	4.44	solcap_snp_c2_32854	5	10418984	0	1	6.65	0.88
gota	2-dom-alt	4.44	solcap_snp_c1_6992	6	53677174	0	1	4.66	-0.55
gota	2-dom-alt	4.44	solcap_snp_c1_15787	11	6943511	0	1	5.53	-0.75
gota	2-dom-ref	4.48	solcap_snp_c1_10855	7	9998099	0	1	4.48	0.63
gota	2-dom-ref	4.48	solcap_snp_c1_11907	9	51386555	0	1	5.25	0.5
gota	2-dom-ref	4.48	solcap_snp_c2_2998	9	58689196	0	1	11.19	-1

Cuadro 2: QTLs significantes del model GWAS Kinship vs. modelo de acción de gen.

Rasgo	Modelo	Umbral	Marcador	Crom	Posición	Ref	Alt	Score	Effect
gota	general	4.72	solcap_snp_c1_12386	2	37411847	0	1	6.95	NA
gota	general	4.72	solcap_snp_c1_5895	2	48564909	0	1	7.54	NA
gota	general	4.72	solcap_snp_c2_32854	5	10418984	0	1	5.85	NA
gota	general	4.72	solcap_snp_c2_15929	7	43192596	0	1	8.87	NA
gota	general	4.72	solcap_snp_c2_12404	7	52544646	0	1	4.74	NA
gota	general	4.72	solcap_snp_c2_12601	7	53106007	0	1	4.9	NA
gota	general	4.72	solcap_snp_c2_49245	8	35651378	0	1	7.7	NA
gota	general	4.72	solcap_snp_c2_41463	8	35922019	0	1	7.71	NA
gota	general	4.72	solcap_snp_c1_821	8	37494529	0	1	8.18	NA
gota	general	4.72	solcap_snp_c2_2998	9	58689196	0	1	9.67	NA
gota	general	4.72	solcap_snp_c1_11246	11	4745919	0	1	12.76	NA
gota	1-dom-alt	4.35	solcap_snp_c2_36664	1	535454	0	1	5.15	-0.71
gota	1-dom-alt	4.35	solcap_snp_c2_4875	1	83177518	0	1	5.18	-0.76
gota	1-dom-alt	4.35	solcap_snp_c1_3484	4	71945107	0	1	5.08	-0.78
gota	1-dom-ref	4.44	solcap_snp_c1_7325	2	42764457	0	1	4.52	0.73
gota	1-dom-ref	4.44	solcap_snp_c2_40748	12	12654160	0	1	4.86	0.66
gota	2-dom-alt	4.44	solcap_snp_c2_32854	5	10418984	0	1	6.65	0.88
gota	2-dom-alt	4.44	solcap_snp_c1_6992	6	53677174	0	1	4.66	-0.55
gota	2-dom-alt	4.44	solcap_snp_c1_15787	11	6943511	0	1	5.53	-0.75
gota	2-dom-ref	4.48	solcap_snp_c1_10855	7	9998099	0	1	4.48	0.63
gota	2-dom-ref	4.48	solcap_snp_c1_11907	9	51386555	0	1	5.25	0.5
gota	2-dom-ref	4.48	solcap_snp_c2_2998	9	58689196	0	1	11.19	-1

3.2. Corrección por False Discovery Rate (FDR)

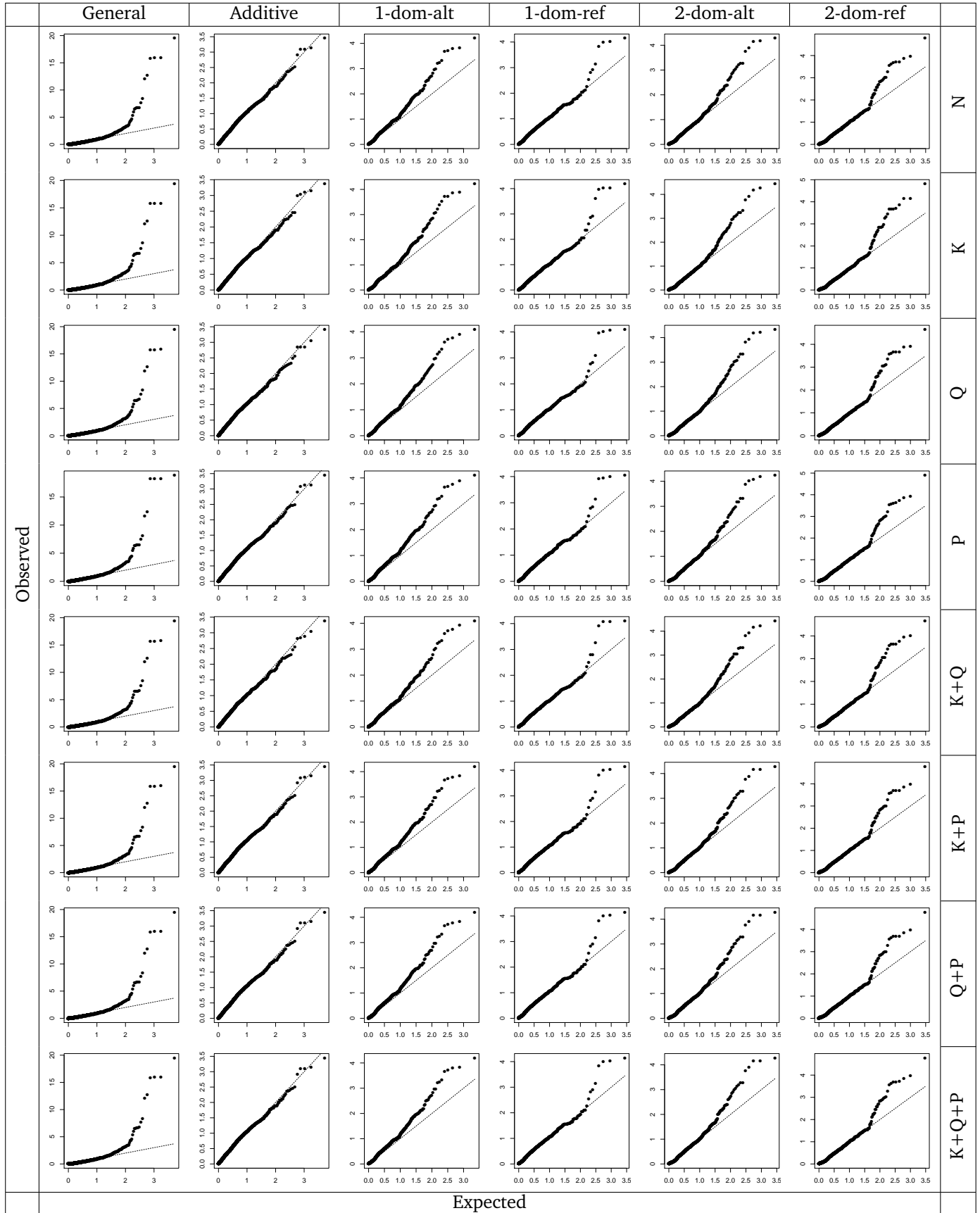


Figura 3: Comparación de gráficos Quartile-Quantile (QQ plots) para los 6 modelos de GWAS (Naive, Kinship, Structure, Structure+Kinship) y los 4 modelos de acción de gen disponibles en el software GWASpoly (general, additive, 1-dom, 2-dom).

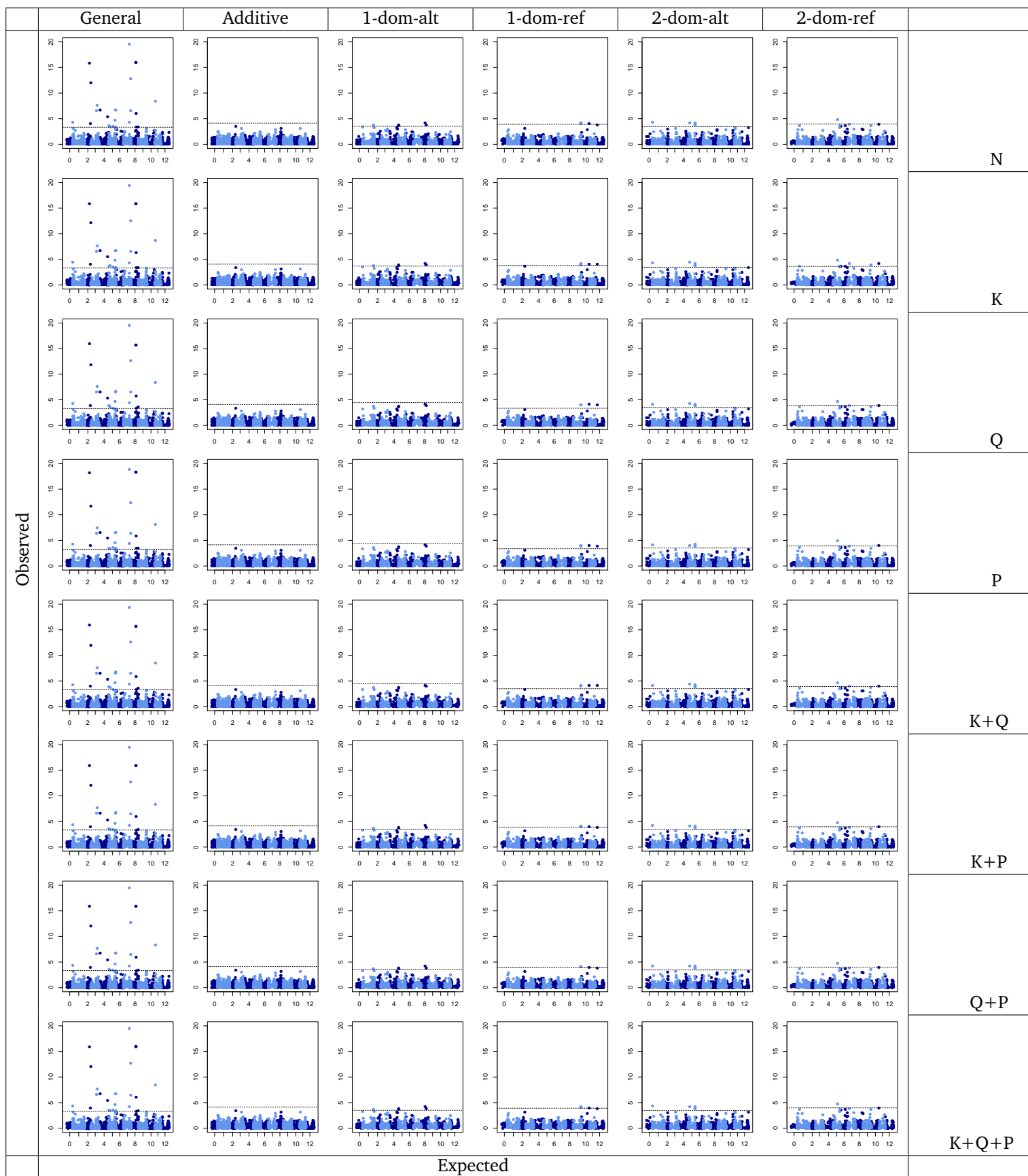


Figura 4: Comparación de gráficos Manhattan para los 4 modelos de GWAS (Naive, Kinship, Structure, Structure+Kinship) y los 4 modelos de acción de gen disponibles en el software GWASpoly (general, additive, 1-dom, 2-dom).

