

# TRDscreen v.2.03: software para el análisis de distorsión de la segregación con efectos epistáticos

Joaquim Casellas<sup>1\*</sup>

<sup>1</sup> Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona

\* Corresponding author: joaquim.casellas@uab.cat

## Resumen

La distorsión de la segregación es un fenómeno genético caracterizado por la desviación de las frecuencias mendelianas esperadas en la descendencia, donde los alelos de un progenitor heterocigoto no se transmiten con igual probabilidad. Estas distorsiones pueden originarse en múltiples fases reproductivas (gametogénesis, fecundación, implantación, desarrollo embrionario, o incluso, vida neonatal temprana) y deberse a un amplio abanico de mecanismos biológicos. Dentro de este contexto, la distorsión de la segregación debería considerarse como un fenómeno relevante en múltiples campos de investigación por su incidencia, magnitud y origen biológico. Aunque el abordaje estadístico de este fenómeno se ha venido centrado históricamente en metodologías generalistas (p.ej., test de  $\chi^2$  sobre poblaciones F2 y retrocruces), disponemos de modelos específicos desarrollados durante los últimos 20 años que permiten una caracterización más detallada de los mecanismos genéticos implicados, sugiriendo incluso la etapa reproductiva en la que se origina la distorsión. Dentro de este contexto, el programa TRDscreen v.1.07 representó un hito relevante en este campo al facilitar el análisis de modelos de distorsión de la segregación tanto a nivel alélico (etapas haploides) como genotípico (etapas diploides), permitiendo la comparación entre ambos dentro de un contexto de máxima verosimilitud. No obstante, la versión actual (TRDscreen v.2.03) representa un avance sustancial al integrar también la predicción de efectos epistáticos entre dos *loci*, manteniendo las dos parametrizaciones originales, pero expandiéndolas a un marco conceptual de mayor complejidad genética. Asumiendo dos *loci* polimórficos, existen un total de 32 cruces informativos con solo un genotipo heterocigoto, 24 cruces con dos genotipos heterocigotos, ocho cruces con tres genotipos heterocigotos, y un solo cruce implicando dobles heterocigotos (65 cruces informativos en total). Para cualquiera de estos cruces, el programa obtiene la información relevante a partir del genotipo de los progenitores y el recuento de hijos para cada genotipo, y lo integra en un procedimiento de máxima verosimilitud mediante el método de *scoring* de Fisher (solo se puede obtener una fórmula cerrada de los estimadores de máxima verosimilitud en el caso del cruce entre dobles heterocigotos). Ambos modelos implican cuatro efectos genéticos directos (dos para cada *loci*) más dos (modelo alélico) o cuatro (modelo genotípico) efectos epistáticos, para los que se calcula el correspondiente *p*-valor mediante un test de Wald basado en la aproximación normal. Resulta importante destacar que el número de combinaciones dos a dos para *n* marcadores polimórficos equivale a el número triangular  $T_{n-1} = (n - 1)(n - 2) / 2$ , alcanzando casi los 1.250 millones para un panel de 50K. Es por ello que la eficiencia y velocidad de computación resultan esenciales para este tipo de análisis. Asumiendo un número moderado de familias por combinación ( $\leq 25$ ), el TRDscreen v.2.03 realiza más de 4.000 análisis por segundo en un procesador Intel® Core™ Ultra 7 165U, abordando el chip de 50K en 3,5 días de cálculo (asumiendo que todos los marcadores son polimórficos y se ha generado el fichero de datos con anterioridad). El programa está escrito en Fortran77, se ha testado con el compilador GNU Fortran (<https://gcc.gnu.org/fortran/>) y se puede descargar de forma gratuita en <https://www.casellas.info/software/TRDscreen.zip>.

*Keywords: distorsión de la segregación, Fortran, máxima verosimilitud, Mendel, SNP*