

Aplicaciones del Algebra de Vectores a la Estimación del Desequilibrio de Ligamiento

Luis Gomez-Raya^{1*}, Ramon Peiró-Pastor¹, and Wendy M. Rauw¹

¹Departamento de Mejora Genética Animal; INIA-CSIC; Ctra de La Coruña km 7,5; Madrid

*Corresponding author: gomez-luis@inia.csic.es

Resumen

El desequilibrio de ligamiento (LD) es la asociación no aleatoria de alelos en loci diferentes. Los vectores en geometría son segmentos orientados que poseen dirección, sentido y módulo (longitud). En este trabajo resumimos nuevas aplicaciones del álgebra vectorial a la Genómica, particularmente a la interpretación geométrica del desequilibrio de ligamiento y a las consecuencias prácticas derivadas de dicha interpretación. Un vector en el espacio euclidiano real de dimensión (R^n) tiene componentes según el genotipo de cada individuo. Así los componentes de los vectores \vec{u}_A y \vec{u}_B correspondientes a SNPs con alelos A/a and B/b son:

$$\begin{aligned} u_{AA} &= (1 - (f_{AA} - f_{\alpha\alpha}))\alpha_A; & u_{BB} &= (1 - (f_{BB} - f_{bb}))\alpha_B \\ u_{Aa} &= -(f_{AA} - f_{\alpha\alpha})\alpha_A; & u_{Bb} &= -(f_{BB} - f_{bb})\alpha_B \\ u_{aA} &= (-1 - (f_{AA} - f_{\alpha\alpha}))\alpha_A; & u_{bB} &= (-1 - (f_{BB} - f_{bb}))\alpha_B, \end{aligned}$$

donde f_i es la frecuencia del genotipo i and α_i es el efecto aditivo del SNP i . El producto escalar de vectores se define como

$$\vec{u}_A \cdot \vec{u}_B = \|\vec{u}_A\| \|\vec{u}_B\| \cos \varphi_{AB}$$

Donde los módulos $\|\vec{u}_A\|$ and $\|\vec{u}_B\|$ son funciones de la raíz cuadrada de la varianza aditiva y del tamaño de la muestra. Resolviendo esta ecuación para φ_{AB} y poniéndolo en el sistema sexagesimal obtenemos

$$\varphi_{AB} = \frac{180}{\pi} \arccos \left(\frac{\vec{u}_A \cdot \vec{u}_B}{\|\vec{u}_A\| \|\vec{u}_B\|} \right)$$

Definimos este ángulo, φ_{AB} , como la interpretación geométrica del desequilibrio de ligamiento (Desequilibrio Geométrico de Ligamiento).

La utilización del Algebra de Vectores aplicada a la Genómica tiene diversas aplicaciones que enumeramos a continuación: 1) *Normalización Vectorial del desequilibrio de ligamiento*. La normalización del desequilibrio de ligamiento se realiza para obtener una medida entre 0 y 1 que describa la asociación entre alelos en una escala fácilmente interpretable. Proponemos una nueva medida normalizada del desequilibrio de ligamiento basada en la norma del vector suma de los dos vectores. Dicha norma presenta un valor mínimo cuando los vectores son ortogonales (equilibrio) y un valor máximo cuando el desequilibrio es completo. Estos valores permiten construir una medida entre 0 y 1 que depende en menor medida de las frecuencias alélicas. El desequilibrio normalizado puede interpretarse biológicamente como un incremento de la variabilidad genética de la suma de los dos vectores debido al desequilibrio, ya que la norma del vector es función de la raíz cuadrada de la varianza aditiva, 2) *Generalización del Desequilibrio Geométrico de Ligamiento a SNPs con caracteres cuantitativos (aditivos, dominantes, interacciones) y enfermedades hereditarias*. Hemos desarrollado ecuaciones de predicción y explorado las propiedades de los distintos sistemas, 3) *Estimación algebraica de LD entre pares de SNPs*. Hemos desarrollado fórmulas para estimar LD de una manera directa sin necesidad de métodos iterativos como el método standard y 4) *Computación masiva de LD*. La utilización de vectores en matrices que incluyen SNPs (aditivo, dominante, interacción etc.) y/o caracteres cuantitativos permite un cálculo masivo y rápido del

desequilibrio. Esto es debido a que la multiplicación de matrices es simple y esta incorporada en los lenguajes de programación.

Keywords: Linkage Disequilibrium, Geometry, Vector Space, SNP

Agradecimientos. Este trabajo está financiado por proyectos PID2022-142738OB-I00 y PIE-202540E044. Los autores agradecen al Centro de Supercomputación de Galicia (CESGA) por el uso del superordenador Finis Terrae III.