

Ejercicio 4

a) Podría en función de esta información determinar un gradiente de similaridad entre los tres individuos?

Para poder establecer un gradiente de similaridad entre los individuos es necesario primero especificar un criterio que defina “similaridad”. Si consideramos la primer notacion, donde se representan los loci de cada individuos con letras, podriamos establecer que los individuos son similares si coinciden en letras y subindices, sumando un punto por cada letra, con su respectivo subindice, coincidente. Luego, de esta manera podemos registrar el siguiente puntaje para cada uno de los pares de comparaciones.

- Individuo 1 vs Individuo 2: $S_{12} = 5$
- Individuo 1 vs Individuo 3: $S_{13} = 6$
- Individuo 2 vs Individuo 3: $S_{23} = 5$

Por consiguiente, podriamos decir que los individuos 1 y 3 son los mas semejantes ya que registraron el mayor puntaje.

Por otro lado, tambien es posible establecer un gradiente de similaridad considerando el patron de bandas de cada individuo. Bajo esta representacion resulta util redefinir el concepto de similaridad, sumando un punto si la coincidencia en bandas es exacta y asignando medio punto si al menos coinciden una de sus bandas en cada locus. De esta forma, los puntajes que se registran son los siguientes:

- Individuo 1 vs Individuo 2: $S_{12} = 2,5$
- Individuo 1 vs Individuo 3: $S_{13} = 3$
- Individuo 2 vs Individuo 3: $S_{23} = 2,5$

Al igual que antes podemos observar que los individuos 1 y 3 son los mas semejantes.

Luego, si quisieramos establecer un gradiente de similaridad podriamos expresarlo de la siguiente manera:

$$S_{12} = S_{23} < S_{13}$$

b) Con el fin de utilizar un coeficiente de similaridad para cuantificar la semejanza entre individuos como codificaría el resultado obtenido del patrón de bandas ?

	A1	A2	B1	B2	C1	C2	D1	D2
Individuo 1	0	1	0	1	0	1	1	1
Individuo 2	1	1	0	1	1	1	0	1
Individuo 3	0	1	1	1	1	1	1	1

En la tabla anterior podemos observar una nueva codificacion del patron de bandas, donde se indica con ceros y unos la ausencia y presencia, respectivamente, de cada alternativa alelica.

c) Utilizando de los coeficientes de similaridad SM, Jaccard, Dice y Russel-Rao cuantifique las similaridades entre los individuos 1-2 y 2-3.

	Individuo 1 vs Individuo 2	Individuo 2 vs Individuo 3
SM	5/8	5/8
Jaccard	4/7	5/8
Dice	8/11	10/13
Russel-Rao	4/8	5/8

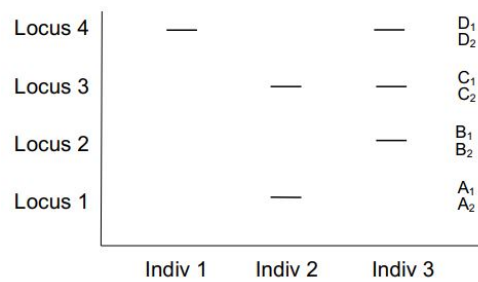
En la tabla anterior podemos observar las similitudes obtenidas bajo los distintos coeficientes entre los individuos 1-2 y 2-3.

d) Concuerda lo hallado en el inciso (c) respecto a lo esperado en (a) ? Comente al respecto.

Cabe destacar que solamente el coeficiente Simple Matching coincide con la similitud subjetiva establecida en el inciso anterior, donde observamos que la Similitud entre el Individuo 1 y 2 era la misma que para los individuos 2 y 3. Esto se corresponde, a que el coeficiente SM considera coincidencias entre “cosas” (ya sea en presencia o ausencia), tal como el criterio adoptado por nosotros en el inciso anterior para poder establecer la Similitud subjetiva.

Por otra parte podemos observar que la similitud mas alta se obtiene utilizando el coeficiente Dice, ya que el mismo le da mayor importancia a las coincidencias en presencia e ignora si existen coincidencias en ausencias.

e) Como sería el patrón de bandas si hubiese utilizado un marcador molecular dominante ? (tomemos a la alternativa alélica 1 como dominante sobre la 2)



f) Como codificaría esta información para poder utilizar como medida de semejanza un coeficiente de similitud ?

	Locus 1	Locus 2	Locus 3	Locus 4
Individuo 1	0	0	0	1
Individuo 2	1	0	1	0
Individuo 3	0	1	1	1

Bajo esta nueva codificación pueden calcularse los coeficientes de similitud SM, Jaccard, Dice y Russel-Rao de manera tal que sea posible cuantificar las similitudes entre los individuos 1-2 y 2-3.

	Individuo 1 vs Individuo 2	Individuo 2 vs Individuo 3
SM	1/4	1/4
Jaccard	0/3	1/4
Dice	0/3	2/5
Russel-Rao	0/4	1/4

g) Verificar resultados con R

Para verificar los resultados obtenidos al calcular los distintos coeficientes de similitud propuestos, se utiliza la función *dist.binary* del paquete *ade4*

Coeficientes de similitud para la codificación que considera alelos codominantes:

	Individuo 1 vs Individuo 2	Individuo 2 vs Individuo 3
SM	0.62	0.62
Jaccard	0.57	0.62
Dice	0.73	0.77
Russel-Rao	0.50	0.62

Coeficientes de similaridad para la codificación que considera al alelo 1 como dominante:

	Individuo 1 vs Individuo 2	Individuo 2 vs Individuo 3
SM	0.25	0.25
Jaccard	0.00	0.25
Dice	0.00	0.40
Russel-Rao	0.00	0.25