

Explorando um modelo de segregação alélica em tetraplóides

Felipe Grillo Pinheiro¹

Cleber Bisognin²

Resumo: Poliploidia, existência de mais de dois conjuntos de cromossomos (genomas) no mesmo núcleo, tem sido um importante mecanismo na evolução de eucariotos. Autopoliploidia e alopoliploidia são as duas principais categorias em que são classificados os poliploides e sua classificação pode ser efetuada a partir do tipo de segregação alélica que o tetraploide apresenta. O objetivo do trabalho foi em investigar as propriedades de um modelo da literatura através de simulações verificando o tamanho da amostra de gametas necessário para analisar a segregação alélica a partir de parentais com diferentes genótipos e avaliando a média e o EQM das estimativas. A análise preliminar das propriedades do modelo indicou o EQM das estimativas de $\hat{\theta}$ aumentam conforme aumenta-se a quantidade de alelos repetidos no genótipo parental, considerando os parâmetros testados. Além disso a média das mesmas estimativas convergem mais rapidamente quando se tem os quatro alelos distintos no parental. O comportamento das estimativas variam de acordo com o valor de τ .

Palavras-chave: Poliploidia, segregação alélica, distribuição multinomial, tamanho amostral.

1 Introdução

A Poliploidia, existência de mais de dois conjuntos de cromossomos (genomas) no mesmo núcleo, tem sido um importante mecanismo na evolução de eucariotos (OTTO, 2007; PARISOD et al., 2009). Estudos recentes demonstraram que o genoma de poliploides pode sofrer extensa e rápida reorganização, através de alterações funcionais e estruturais, reforçando o interesse em entender a multiplicação genômica como um processo evolutivo, bem como suas consequências (genéticas e ecológicas) sobre o estabelecimento e manutenção de poliploides na natureza (PARISOD et al., 2009; DOYLE et al., 2008).

Autopoliploidia e alopoliploidia são as duas principais categorias em que são classificados os poliploides (RAMSEY; SCHEMSKE, 1998), dependendo da origem dos genomas envolvidos na formação do tetraploide. Autopoliploides surgem a partir de parentais diploides da mesma espécie, como resultado da duplicação de genomas estruturalmente similares e homólogos ($AA + AA = AAAA$).

¹ UFRGS - Universidade Federal do Rio Grande do Sul. Email: grillopf@gmail.com

² UFRGS - Universidade Federal do Rio Grande do Sul. Email: cleberbisognin@gmail.com

Dessa forma, cada cromossomo pode parear aleatoriamente com qualquer um dos seus homólogos, formando tetravalentes e levando à herança tetrassômica, pois as combinações cromossômicas são aleatórias e todos os genótipos possíveis para os gametas, a partir do genótipo parental são produzidos em frequências iguais. Aloploidos surgem através da duplicação de genomas não-homólogos (homeólogos), ou seja, $(AA + BB = AB \rightarrow \text{duplicação} \rightarrow AABB)$. Como há o envolvimento de duas espécies diferentes existem dois genomas homeólogos entre si, cada um composto por dois conjuntos de cromossomos homólogos entre si (OLSON, 1997; PARISOD et al., 2009; SOLTIS; SOLTIS, 2000; STIFT et al., 2008). Assim, um cromossomo parecia exclusivamente com seu homólogo, ou seja, há apenas formação de bivalentes, levando a uma herança dissômica onde os genótipos possíveis dos gametas também são produzidos na mesma frequência, mas alguns genótipos jamais são produzidos.

Adicionalmente, Stebbins (1971) classificou táxons aloploidos que formam multivalentes durante a meiose como “aloploidos segmentares”, com o objetivo de indicar a duplicação de genomas parcialmente diferenciados $(AA + A'A' = AA' \rightarrow \text{duplicação} \rightarrow AAA'A')$. A diferença em relação aos aloploidos é que os subgenomas A e A' são semelhantes o suficiente para serem considerados homólogos em alguns conjuntos cromossômicos. Nesse caso a segregação alélica é intermediária.

Stift (2008), desenvolveu um procedimento, baseado na verossimilhança, para avaliar qual modelo de herança, se dissômica, tetrassômica ou intermediária, explica melhor a segregação alélica de marcadores genéticos em tetraploides. Entretanto, não se sabe o tamanho de amostra dos gametas adequado para avaliar a segregação em um marcador de um determinado genótipo parental. O objetivo deste trabalho consiste em investigar as propriedades do modelo teórico proposto por Stift (2008) através de simulações verificando o tamanho da amostra de gametas necessário para analisar a segregação alélica a partir de parentais com diferentes genótipos e avaliando a média e o EQM das estimativas $\hat{\theta}$ para os valores $\vec{\theta} = \langle \tau, \beta, \delta_1, \delta_2, \delta_3 \rangle$ do modelo fornecidos para a geração da amostra.

2 Metodologia

O modelo proposto na literatura explica as proporções (probabilidades) de todos os possíveis gametas produzidos por um indivíduo com genótipo ABCD conforme o seguinte sistema na forma matricial:

$$\begin{array}{l}
 \Pr(AA) \\
 \Pr(BB) \\
 \Pr(CC) \\
 \Pr(DD) \\
 \Pr(AB) \\
 \Pr(AC) \\
 \Pr(AD) \\
 \Pr(BC) \\
 \Pr(BD) \\
 \Pr(CD)
 \end{array}
 =
 \begin{bmatrix}
 0 \\
 0 \\
 0 \\
 0 \\
 1/6 \\
 1/6 \\
 1/6 \\
 1/6 \\
 1/6 \\
 1/6
 \end{bmatrix}
 \tau +
 \begin{bmatrix}
 1/4 \\
 1/4 \\
 1/4 \\
 1/4 \\
 -1/6 \\
 -1/6 \\
 -1/6 \\
 -1/6 \\
 -1/6 \\
 -1/6
 \end{bmatrix}
 \beta\tau +
 \begin{bmatrix}
 0 & 0 & 0 \\
 0 & 0 & 0 \\
 0 & 0 & 0 \\
 0 & 0 & 0 \\
 0 & 1/4 & 1/4 \\
 1/4 & 0 & 1/4 \\
 1/4 & 1/4 & 0 \\
 1/4 & 1/4 & 0 \\
 1/4 & 0 & 1/4 \\
 0 & 1/4 & 1/4
 \end{bmatrix}
 (1 - \tau) \langle \delta_1, \delta_2, \delta_3 \rangle \quad (1)$$

Estas equações definem um conjunto de equações não-lineares com cinco parâmetros. O parâmetro τ indica a proporção de gametas formados por associações aleatórias entre os cromossomos durante a meiose, portanto é o parâmetro que está ligado à segregação tetrassômica e seu valor pode variar de zero (segregação dissômica) à um (segregação tetrassômica). Se $\tau < 1$, as proporções dos gametas esperados depende de um conjunto de três parâmetros “dissômicos” (δ_1 , δ_2 e δ_3) que indicam, respectivamente, o grau de pareamento preferencial em cada tipo de associação alelo-genoma. O parâmetro β representa a frequência do fenômeno redução-dupla relativo à frequência de formação de quadrivalentes na meiose.

Estimação de parâmetros: A função logverossimilhança da distribuição multinomial é a função objetivo a ser maximizada. Portanto, o interesse é encontrar os parâmetros do modelo que resultem na maior logverossimilhança.

Simulação: Para avaliar o modelo e suas propriedades, uma rotina foi escrita (no R) para mimetizar a segregação dos gametas de um marcador genético, a partir das probabilidades esperadas, considerando um dado genótipo parental. Pelo caráter inicial do estudo, foram avaliados alguns valores para os parâmetros $\vec{\theta}$. O algoritmo da rotina seguiu os seguintes passos:

- 1) Escolha dos valores para os parâmetros $\vec{\theta} = \langle \tau, \beta, \delta_1, \delta_2, \delta_3 \rangle$;
- 2) Produzir re amostras de tamanho n
- 3) Obter estimativa do vetor $\hat{\vec{\theta}}$ para cada uma das re amostras de tamanho n criadas.
- 4) Avaliar a média e o EQM destas estimativas em relação ao vetor de parâmetros $\vec{\theta}$ conhecido;
- 5) Repetir os passos 2) e 3) para cada tamanho de amostra escolhido para avaliação.

Com o propósito de demonstração foram utilizados nas simulações um número de replicações $re=100$, tamanhos de amostra $n=10, 25, 50, 75, 100$ e 150 e dois conjuntos de valores iniciais para $\vec{\theta}$ com $\tau=0.9$, e 0.67 , e o restante dos parâmetros fixos com os valores $\beta=0.004$, $\delta_1=0.22$, $\delta_2=0$ e $\delta_3=0.78$.

3 Resultados e discussão

A tabela 1 apresenta a média e o EQM das 100 estimativas para os dois conjuntos de parâmetros escolhidos, com os tamanhos de amostra estudados, em dois genótipos parentais. Para o genótipo parental ABCD, que contém as quatro informações alélicas distintas, a média das estimativas dos parâmetros tendem a convergir para o valor real dos parâmetros, quando o valor $\tau = 0.67$. A mesma analogia pode ser feita para as estimativas de β , δ_1 , δ_2 e δ_3 . Entretanto, quando os valores de τ são aumentados para 0.9, as estimativas de todos os parâmetros não convergem rapidamente para os verdadeiros valores dos parâmetros, mesmo com tamanhos de amostras relativamente altos.

Para o caso em que o genótipo parental é ABCA, isto é, contém alelos repetidos, o mesmo comportamento se mantém da convergência se mantém como no caso ABCD, entretanto a convergência é mais lenta, sendo o maior valor de tamanho de amostra ($n=150$) insuficiente para a obtenção da convergência adequada. Para valores de $\tau = 0.67$ as estimativas dos parâmetros possuem menor EQM se compara no caso em que $\tau = 0.9$ (Tabela1).

Tabela 1. Estimativas da média e do EQM para os parâmetros em 100 replicações nos diferentes tamanhos amostrais dos gametas a partir dos vetores de parâmetros conhecidos e o genótipos parentais.

Genótipo	n	Média					EQM					
		τ	β	δ_1	δ_2	δ_3	τ	β	δ_1	δ_2	δ_3	
ABCD	$\vec{\theta}$	0.67	0.04	0.22	0.00	0.78						
	10	0.46	0.05	0.35	0.22	0.43	0.16	0.01	0.17	0.18	0.29	
	25	0.51	0.06	0.21	0.23	0.56	0.14	0.00	0.08	0.18	0.18	
	50	0.60	0.06	0.26	0.13	0.61	0.08	0.00	0.11	0.09	0.15	
	75	0.58	0.07	0.25	0.10	0.65	0.07	0.00	0.09	0.06	0.11	
	100	0.63	0.05	0.25	0.09	0.66	0.05	0.00	0.08	0.06	0.11	
	150	0.68	0.05	0.19	0.14	0.67	0.04	0.00	0.06	0.11	0.11	
	$\vec{\theta}$	0.90	0.04	0.22	0.00	0.78						
	10	0.48	0.07	0.32	0.28	0.4	0.3	0.01	0.15	0.23	0.31	
	25	0.58	0.09	0.31	0.24	0.44	0.17	0.01	0.15	0.18	0.26	
	50	0.71	0.07	0.26	0.29	0.45	0.08	0.00	0.12	0.22	0.24	
	75	0.73	0.06	0.28	0.26	0.46	0.07	0.00	0.13	0.2	0.26	
	100	0.76	0.05	0.28	0.24	0.47	0.05	0.00	0.11	0.18	0.21	
	150	0.81	0.05	0.26	0.19	0.55	0.02	0.00	0.11	0.13	0.18	
	ABCA	$\vec{\theta}$	0.67	0.04	0.22	0.00	0.78					
		10	0.40	0.06	0.19	0.22	0.60	0.20	0.01	0.09	0.14	0.17
25		0.44	0.07	0.19	0.22	0.59	0.17	0.01	0.08	0.16	0.16	
50		0.48	0.07	0.23	0.16	0.61	0.14	0.01	0.08	0.09	0.13	
75		0.46	0.08	0.16	0.17	0.67	0.15	0.01	0.06	0.10	0.09	
100		0.45	0.08	0.16	0.18	0.65	0.16	0.01	0.06	0.09	0.08	
150		0.46	0.09	0.17	0.15	0.67	0.13	0.01	0.06	0.07	0.07	
$\vec{\theta}$		0.90	0.04	0.22	0.00	0.78						

10	0.36	0.07	0.29	0.24	0.47	0.42	0.01	0.15	0.18	0.24
25	0.60	0.07	0.24	0.30	0.46	0.21	0.01	0.10	0.23	0.26
50	0.59	0.07	0.27	0.26	0.48	0.19	0.00	0.12	0.18	0.24
75	0.67	0.07	0.28	0.22	0.50	0.14	0.00	0.13	0.14	0.22
100	0.61	0.08	0.26	0.31	0.43	0.18	0.00	0.12	0.21	0.25
150	0.64	0.08	0.29	0.23	0.47	0.17	0.00	0.12	0.16	0.21

O gráfico1, ilustra o comportamento da média e do EQM das 100 estimativas para diferentes valores de τ e tamanhos de amostra e cenários de genótipos parentais. A convergência para os verdadeiros parâmetros é mais lenta conforme aumenta a quantidade de alelos repetidos.

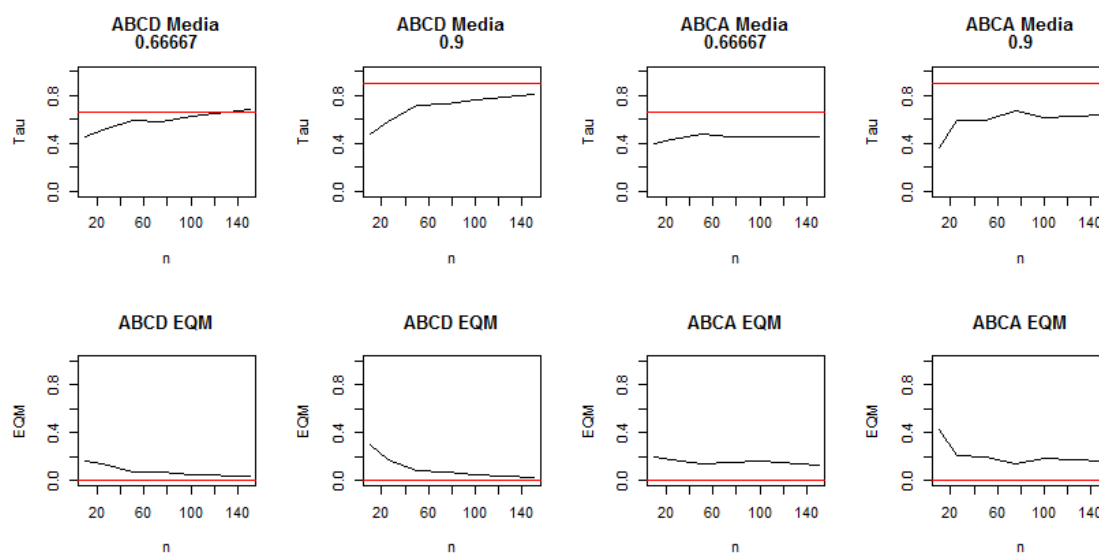


Figura 1. Comportamento da média e do EQM das 100 estimativas para diferentes valores de τ e tamanhos de amostra.

Como no caso ABCD existem várias categorias de genótipos possíveis para os gametas, conforme se aumenta o tamanho da amostra, mais refinada se tornam as probabilidades oriundas do modelo em relação àquela esperada pela amostra e a verossimilhança é maior. No caso de um parental ABCA, o alelo A está repetido, portanto, não são esperados nenhum dos genótipos envolvendo o alelo D que são encontrados nos gametas de ABCD. Como existem menos categorias de genótipos dos gametas a capacidade de discriminar as probabilidades entre diversas possibilidades de parâmetros, a convergência torna-se mais lenta ou até impossível na prática, pois a partir de $n=100$ a obtenção de mais progênie é comprometida.

3 Conclusão

A análise preliminar das propriedades do modelo indicou que a capacidade de discriminação entre as probabilidades de cada categoria de genótipo dos gametas de diferentes valores dos parâmetros estimados $\hat{\theta}$ diminui conforme aumenta-se a quantidade de alelos repetidos no genótipo parental, considerando os parâmetros testados. Isso significa que deve-se tomar cuidado ao utilizar este método quando não temos completa distinção dos alelos parentais. Além disso, as estimativas do parâmetro τ podem ser subestimadas quando o tamanho da amostra de gametas é pequeno, mesmo quando o genótipo parental não tem alelos repetidos. Uma limitação do estudo consiste no fato de que foi utilizado um número muito baixo de replicações, especialmente por se tratar de um problema envolvendo proporções. Entretanto, estas análises são preliminares e serão estendidas para outros genótipos, vetor de parâmetros, tamanhos de amostra e número de replicações.

Referências

- [1] DOYLE, J. J. *et al.* Evolutionary genetics of genome merger and doubling in plants. *Annual Review of Genetics* [S.I.], v. 42, p.443-461, 2008.
- [2] OTTO, S. P. The evolutionary consequences of polyploidy. *Cell* [S.I.], v. 131 p. 452-462, 2007.
- [3] PARISOD, C.; HOLDEREGGER, R.; BROCHMANN, C. Evolutionary consequences of autopolyploidy. *New Phytologist* [S.I.], v. 186, n. 1, p. 5–17, 2009.
- [4] RAMSEY, J.; SCHEMSKE, D.W. Pathways, mechanisms and rates of polyploid formation in flowering plants. *Annual Review of Ecology and Systematics*, Palo Alto, v. 29, p. 467-501, 1998.
- [5] SOLTIS, P. S.; SOLTIS D. E. The role of genetic and genomic attributes in the success of polyploids. *Proceedings of the National Academy of Sciences, USA*, v. 97, p. 7051-7057, 2000.
- [6] STEBBINS, G.L. Chromosomal evolution in higher plants. London, UK. p. 216, 1971.
- [7] STIFT, M. *et al.* Segregation models for disomic, tetrasomic and intermediate inheritance in tetraploids: a general procedure applied to *Rorippa* (yellow cress) microsatellite data. *Genetics* [S.I.], v. 179, p. 2113-2123, 2008.