



UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA
DEPARTAMENTO DE ESTATÍSTICA



Avaliação numérica de um modelo de segregação alélica em espécies tetraploides: um estudo de simulação

Autor: Felipe Grillo Pinheiro
Orientador: Professor Dr. Cleber Bisognin

Porto Alegre, 20 de Dezembro de 2013.

Universidade Federal do Rio Grande do Sul
Instituto de Matemática
Departamento de Estatística

Avaliação numérica de um modelo de
segregação alélica em espécies tetraploides:
um estudo de simulação

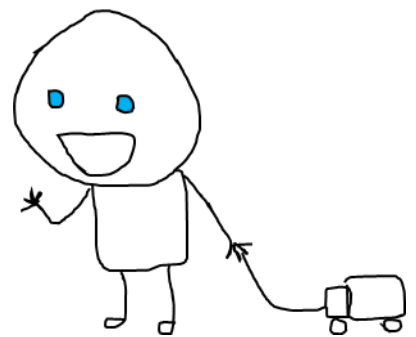
Autor: Felipe Grillo Pinheiro

Monografia apresentada para obtenção
do grau de Bacharel em Estatística.

Banca Examinadora:
Professor Dr. Cleber Bisognin (orientador)
Professora Suzi Alves Camey

Porto Alegre, 20 de Dezembro de 2013.

Do Davico



“

I'm a great believer in luck, and I find the harder I
work the more I have of it.

Thomas Jefferson

AGRADECIMENTOS

Ao meu orientador, Dr. Cleber Bisognin, pela grande ajuda no desenvolvimento das rotinas utilizadas neste trabalho, conselhos e ideias. Por ter sido um grande exemplo de profissional e aceitado o desafio do desenvolvimento em conjunto deste trabalho. Pelas longas conversas em manhãs de plenas sextas-feiras.

À minha orientadora na genética, Dra. Fernanda Bered, pela orientação, confiança, compreensão e por apostar que aquele guri poderia encarar o desafio dos poliploides. (Na verdade por me jogar na arena com esses monstros poliploides)

À professora Suzi Camey por ter gentilmente cedido mais tempo para a elaboração deste manuscrito.

Aos membros do Núcleo Genética e Conservação de Plantas, principalmente as minhas orientadoras práticas e co-autoras Camila Zanella e Márcia Goetze por toda a ajuda na etapa anterior deste trabalho, enquanto estudava uma espécie de bromélia tetraploide, no laboratório e no campo, pela convivência, exemplos, compreensão e críticas indispensáveis.

Ao Bruno e à Cris por me ensinarem outro conceito de família e sempre proporcionarem momentos dos quais serei eternamente grato. Amo vocês.

Aos meus pais, Maristela e Vilmar pelo estímulo, compreensão e apoio. À minha irmã Débora que sempre faz o meu sumário ficar lindo! À minha irmã Márcia e meu cunhado Wagner por todo o apoio, preocupação e por terem me dado o delicioso prazer de desfrutar de um sobrinho-afilhado lindo ao qual dedico este trabalho.

Aos meus amigos mais importantes do mundo estatístico: Criatiano Lima Hackmann, Vinícius Roglio, Lucas Schimidt, Lucas Serafim, João Lucas, João Dewes e Thaís Neves. Por aguentarem conviver comigo durante estes quatro anos de graduação. Pelas risadas,

*companherismo, conselhos, xingões e tantos trabalhos realizados com empenho e esforço,
mas sempre de última hora. Orgulho pela nossa amizade!*

À minha monstrenga favorita Nicole Vetto por ser simplesmente uma amiga fantástica.

*Aos meus amigos e colegas de trabalho na UFCSPA, Milena, Laíse, Letícia, Mário. Pela
disponibilidade, conselhos e ombro amigo. Obrigado por compartilharem sua amizade
comigo.*

A todos que de alguma forma contribuíram para a realização deste trabalho

MUITO OBRIGADO!!!

RESUMO

Poliploidia, definida como a existência de mais de dois conjuntos de cromossomos (genomas) no mesmo núcleo, tem sido um importante mecanismo na evolução de eucariotos. Autopoliploidia e aloploidia são as duas principais categorias em que são classificados os poliploides de acordo com a origem de seu genoma. A classificação de um tetraploide como auto ou aloploide pode ser inferida a partir do tipo de segregação alélica que a espécie apresenta. Caracterizar os mecanismos e padrões de herança genética em poliploides é crucial para estudar a evolução, reprodução e melhoramento genético das populações desses organismos. Stiff *et al* (2008), desenvolveu um procedimento geral, baseado na log verossimilhança da distribuição multinomial, para avaliar qual modelo de herança (se dissômica, tetrassômica ou intermediária), explica melhor a segregação alélica de marcadores genéticos em tetraploides.

O objetivo do trabalho foi investigar as propriedades do modelo proposto por Stiff através de simulações da segregação alélica em parentais com diferentes genótipos e sob diversos cenários de combinações alélicas, avaliando o EQM das estimativas em cada simulação. A análise das propriedades do modelo indicou o EQM das estimativas da maior parte dos parâmetros diminuem conforme o tamanho da amostra de gametas aumenta nos genótipos parentais ABCD e ABCA. Por outro lado as estimativas do EQM de τ podem aumentar se a quantidade de alelos repetidos no genótipo parental é grande, como no caso ABAA. Dessa forma, sugere-se que somente parentais com genótipos ABCD e ABCA sejam utilizados para avaliar a segregação alélica em tetraploides, uma vez que as estimativas produzidas nos cenários de parentais ABAA têm um alto EQM.

Palavras-chave: *Poliploidia, segregação alélica, herança tetrassômica, distribuição multinomial, simulação*

ABSTRACT

Polyploidy, defined by existence of more than two chromosome sets (genomes) in the same nucleus, has been a major mechanism in eukaryotic evolution. Autopoliploidy and allopoliploidy are the two main categories which polyploids are classified accordingly to their genome origin. The classification of a tetraploid as auto or allopolyploid can be inferred based on allelic segregation presented by the species. Characterize the mechanisms and patterns of genetic inheritance in polyploids is crucial to study evolution, reproduction, breeding and populations of these organisms. Stiff et al (2008) developed a general procedure based on the log-likelihood of the multinomial distribution to assess which model of inheritance (if disomic, tetrasomic or intermediate), best explains the allelic segregation of genetic markers in tetraploids. The aim of this study was to investigate the statistical properties of the model proposed by Stiff through simulations of parental allelic segregation considering different parental genotypes and different allelic combinations scenarios, evaluating parameter MSE estimations in each simulation. The model analysis indicated that MSE of most parameters decreases as the sample size increases in parental gamete genotypes ABCD and ABCA. On the other hand, the MSE of τ can increase if the number of repeated alleles in parental genotype is large, as in ABAA. Therefore, we have suggested that only parental genotypes ABCA and ABCD are used to evaluate the allelic segregation in tetraploids, since the estimations produced by scenarios in parental ABAA show high MSE.

Key words: *Poliploidy, allelic segregation models, tetrasomic inheritance, multinomial distribution, simulation*

SUMÁRIO

RESUMO	6
ABSTRACT	7
CAPÍTULO I: Introdução Geral	9
1.Poliploidia	10
2.Origens de poliploidia	10
3.Padrões de herança em poliploides	11
4.Consequências da poliploidia	13
5.Modelos de segregação alélica em tetraploides	14
6.Objetivo	16
CAPÍTULO II: Numeric evaluation of allelic segregation models in tetraploids: a simulation study	17
CAPÍTULO III: Considerações Finais	37
CAPÍTULO IV: Referências Bibliográficas	40

CAPÍTULO I

Introdução Geral

INTRODUÇÃO GERAL

1. Poliploidia

A Poliploidia pode ser definida como a existência de mais de dois conjuntos de cromossomos (genomas) no mesmo núcleo. Ela tem sido um importante fator na evolução de eucariotos (OTTO, 2007; PARISOD et al., 2009). Estima-se que 30% a 70% das angiospermas (plantas superiores) tenham ancestralidade poliploide (GRANT, 1971; STEBBINS, 1971). Estudos recentes demonstraram que o genoma de poliploides pode sofrer extensa e rápida reorganização, através de alterações funcionais e estruturais (como a diploidização), o que reforça 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).

A duplicação do genoma de um organismo é um evento importante na especiação simpátrica (surgimento de uma nova espécie sem que haja separação física entre indivíduos da nova espécie e da espécie que a originou), pois pode levar ao surgimento de uma nova espécie em apenas uma única geração, caso o fluxo gênico entre o novo indivíduo poliploide e a espécie progenitora seja interrompido devido à incompatibilidade no número de cromossomos (PERRIE et al., 2010). Para que tal especiação tenha sucesso, outros indivíduos poliploides são necessários para a reprodução e manutenção da nova linhagem, o que pode ser solucionado através do desenvolvimento da autofecundação ou reprodução assexuada (OTTO; WHITTON 2000). Entretanto, em plantas, indivíduos triploides podem servir de ponte para o fluxo gênico entre espécies com níveis de ploidia distintos (PARISOD et al., 2009).

2. Origens de poliploidia

Autopoliploidia e aloploidia são as duas principais categorias em que são classificados os poliploides (RAMSEY; SCHEMSKE, 1998). Tradicionalmente, é considerado que autopoliploides surjam através da fusão de gametas não reduzidos de parentais diploides, resultando na duplicação de genomas estruturalmente similares e homólogos ($AA + AA =$

AAAA, um autotetraploide). Alopóliploides surgem via hibridização interespecífica acompanhada da duplicação de genomas não-homólogos (homeólogos), ou seja, $(AA + BB = AB \rightarrow \text{duplicação} \rightarrow AABB)$, no caso de alotetraploides; há dois grupos homeólogos contendo, cada um, dois conjuntos de cromossomos homólogos entre si (OLSON, 1997; PARISOD et al., 2009; SOLTIS; SOLTIS, 2000; STIFT et al., 2008). Essa duplicação pode restaurar a viabilidade e fertilidade do híbrido ao permitir a regularidade no pareamento meiótico (STEBBINS, 1971).

Em autotetraploides cada cromossomo pode parear aleatoriamente com qualquer um dos seus homólogos durante a divisão meiótica, formando geralmente tetravalentes (os quatro cromossomos juntos) ou ainda bivalentes (dois pares de cromossomos). Isto leva à herança tetrassômica, pois as combinações cromossômicas são irrestritas e os gametas produzidos em iguais frequências (Figura 1a). Em alotetraploides, se um cromossomo pareia exclusivamente com seu homólogo, ou seja, há apenas formação de bivalentes, isto leva a uma herança dissômica, pois há restrições no pareamento cromossômico e, conseqüentemente, na formação dos gametas (Figura 1c), apresentando a frequência zero para algumas combinações genotípicas (JACKSON, 1982; STIFT et al., 2008; WU et al., 2001). Adicionalmente, Stebbins (1971) classificou táxons alopóliploides que formam multivalentes durante a meiose como “alopóliploides 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 alopóliploides é que os “subgenomas” A e A' são semelhantes o suficiente para serem considerados homólogos em alguns dos conjuntos cromossômicos (Figura 1 b). Por isso, alopóliploides segmentares podem apresentar tanto bivalentes quanto tetravalentes na meiose, e portanto, padrão de segregação intermediário entre dissômico e tetrassômico, considerando o padrão de segregação de vários marcadores moleculares como um todo, não o resultado de um marcador específico.

3. Padrões de herança em poliploides

Um padrão de herança em poliploides não é uma característica estática. O comportamento dos cromossomos na meiose tende a sofrer mudanças temporais. Por exemplo, a diploidização é um evento comum nos poliploides, isto é, a tendência de poliploides

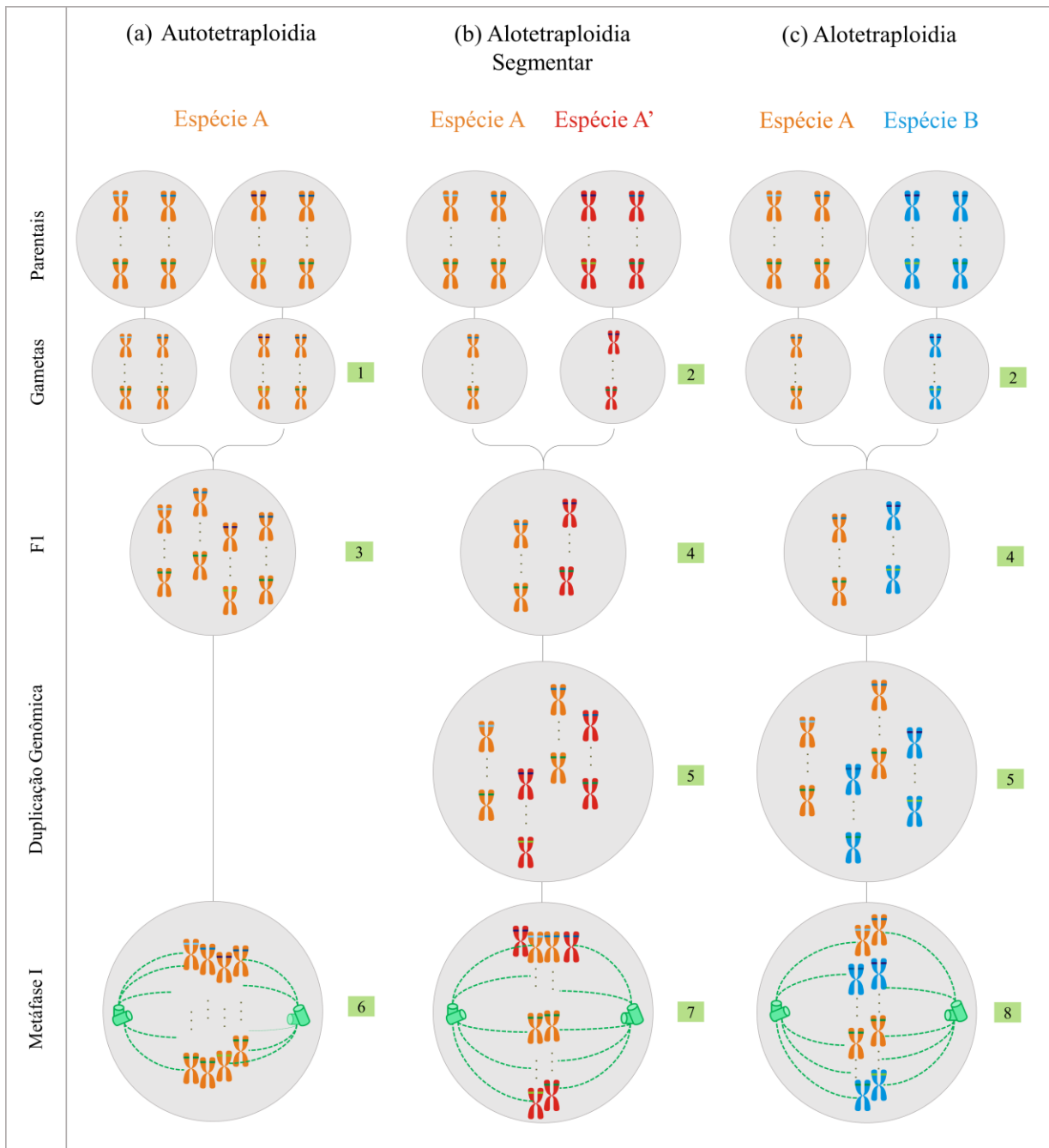


Figura 1. As diferentes origens de um tetraploide: (a) Autotetraploidia: os conjuntos de cromossomos provêm da mesma fonte (espécie parental A), por isso são inteiramente homólogos. A tetravalentes ou bivalentes aleatórios são esperados no pareamento, (b) Alotetraploidia segmentar: os conjuntos cromossômicos são oriundos de espécies distintas (A e A'), mas há certo grau de homologia intergenômico, por isso esperam-se pareamentos tetravalentes em alguns cromossomos e bivalentes preferenciais em outros e (c) Alotetraploidia: os conjuntos cromossômicos provêm cada qual de espécies parentais diferentes (A e B) resultando na existência de dois subgenomas. Ocorrem bivalentes dentro de cada subgenoma no pareamento cromossômico (pareamento intrasubgenômico, comportando-se como um diploide). (1) Gametas não reduzidos, (2) gametas normais, (3) F1 tetraploide, (4) F1 diploide híbrido, (5) duplicação genômica, (6) pareamentos tetravalentes, (7) pareamentos tetravalentes e bivalentes e (8) pareamentos bivalentes.

comportarem-se como diploides, ao longo do tempo (RAMSEY; SCHEMSKE, 2002). Schifino-Wittmann (2003) salienta que há muitas evidências demonstrando rápidas transformações genômicas em poliploides recentes. Apesar disso, os fatores envolvidos nessas transformações ainda não são completamente entendidos.

A consequência prática é que a herança alélica pode alterar-se de tetrassômica para dissômica (e vice-versa). Em autotetraploides (herança tetrassômica) os quatro cromossomos homólogos podem iniciar e manter um pareamento preferencial (dois a dois) que resulta na diploidização citogenética (RAMSEY; SCHEMSKE, 2002; STIFT et al., 2008; WOLFE, 2001). Em aloploiploides (herança dissômica), nem sempre o pareamento é preferencial, podendo ocorrer *crossing-over* entre cromossomos homeólogos, homogeneizando o genoma ao longo das gerações e permitindo a multivalência (SYBENGA, 1996). Já a aloploidia segmentar é observada em sistemas onde a hibridização é comum e os indivíduos são caracterizados por um padrão de herança intermediário entre dissômico e tetrassômico. Neste caso, os parentais de híbridos interespecíficos ainda mantêm certo grau de homologia cromossômica, mas divergiram suficientemente para atingir seus *status* de espécie (RAMSEY; SCHEMSKE, 2002; STIFT et al., 2008).

4. Consequências da poliploidia

Qualquer que seja sua origem, uma espécie poliploide tem potencialmente mais variabilidade genética do que seus progenitores diploides. A presença de dois genomas diferentes permite ao aloploiploide obter todas as enzimas monoméricas e diméricas encontradas em seus parentais (MAHY et al., 2000). Além disso, a aloploidia pode levar à heterozigosidade fixada, se cada “subgenoma” contiver alelos distintos (STIFT et al., 2008). Teoricamente, populações autoploiploides devem exibir maior heterozigosidade do que aquelas dos seus progenitores diploides devido à herança polissômica. Esse aumento na variação genética pode ser um elemento chave na adaptação e tolerância ecológica dos poliploides para habitar novos nichos. (BEVER; FELBER, 1993; MAHY et al., 2000; PARISOD et al., 2009).

Segundo Riesenber (2001), os genomas de aloploiploides estão sujeitos a uma série de mudanças em taxas muito mais elevadas do que em autoploiploides. Tais mudanças

evitariam incompatibilidades entre seus genomas divergentes. Em contrapartida, autopoliploides sofreriam com a redução da fertilidade devido às irregularidades meióticas causadas pela formação de multivalentes (PARISOD et al., 2009).

Caracterizar os mecanismos e padrões de herança genética em poliploides é crucial para estudar a evolução, reprodução, melhoramento e das populações desses organismos. Diferentes modelos devem ser aplicados para calcular coeficientes de variabilidade genética e estrutura populacional, para estimar níveis de endogamia e fluxo gênico e inferir as mudanças resultantes de interações intra ou intergenômicas dependendo da quantidade de genomas envolvidos na formação do poliploide (OLSON, 1997; RONFORT et al., 1998).

5. Modelos de segregação alélica em tetraploides

A investigação dos padrões de segregação de *loci* duplicados e a origem de espécies poliploides normalmente envolvem cruzamentos experimentais controlados e testes de suas progênes (ARNAU et al., 2009; BOUSALEM et al., 2006; GRANT, 1971; SOLTIS; SOLTIS, 1993; STEBBINS, 1971; WENDEL, 2000). Tais estudos têm sido aplicados tanto em espécies cultivadas quanto em populações naturais.

Stift *et al* (2008), desenvolveu um procedimento genérico, baseado na log verossimilhança da distribuição multinomial, para avaliar qual modelo de herança (se dissômica, tetrassômica ou intermediária), explica melhor a segregação alélica de marcadores genéticos em tetraploides. O modelo explica as proporções (ou probabilidades) de todos os possíveis gametas produzidos por um indivíduo com genótipo ABCD conforme o seguinte sistema de equações na forma matricial:

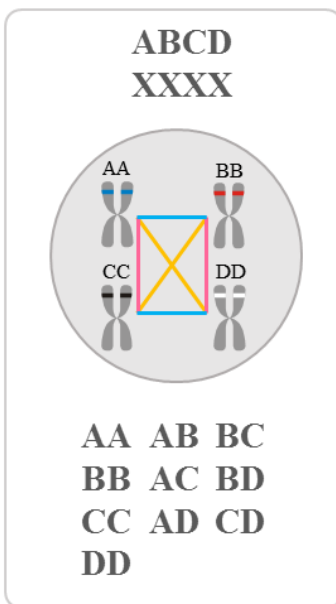
$$\begin{bmatrix} \Pr(AA) \\ \Pr(BB) \\ \Pr(CC) \\ \Pr(DD) \\ \Pr(AB) \\ \Pr(AC) \\ \Pr(AD) \\ \Pr(BC) \\ \Pr(BD) \\ \Pr(CD) \end{bmatrix} = \begin{bmatrix} 0 & 1/4 & 0 & 0 & 0 \\ 0 & 1/4 & 0 & 0 & 0 \\ 0 & 1/4 & 0 & 0 & 0 \\ 0 & 1/4 & 0 & 0 & 0 \\ 1/6 & -1/6 & 0 & 1/4 & 1/4 \\ 1/6 & -1/6 & 1/4 & 0 &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

O modelo é definido por um sistema de equações não-lineares com cinco parâmetros, sujeito às seguintes restrições:

$$\begin{cases} 0 < \tau, \delta_1, \delta_2, \delta_3 < 1 \\ 0 < \beta < 1/6 \\ P(AA) + P(BB) + \dots + P(CD) = 1 \\ \delta_1 + \delta_2 + \delta_3 = 1 \\ \delta_1 \times \delta_2 \times \delta_3 = 0 \end{cases}$$

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.

Segregação Tetrassômica



Segregação Dissômica

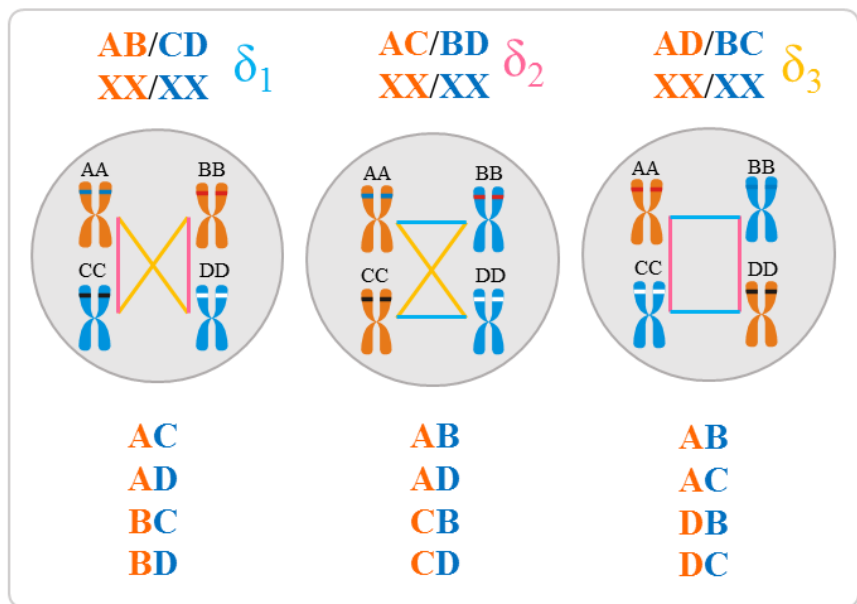


Figura 2. Os genótipos dos gametas são produzidos com probabilidades diferentes, dependendo do tipo de segregação alélica do marcador em questão, além do tipo de associação alelo-genoma se a segregação for dissômica. Por exemplo, na segregação dissômica onde os alelos A e B pertencem ao subgenoma de cor laranja e os alelos C e D estão associados com o genoma de cor azul, gametas com genótipo AB e CD nunca serão produzidos porque, neste caso, haverá a separação dos cromossomos na anáfase I e após, a separação das cromátides irmãs, ou seja, cada cromátide de um mesmo subgenoma acabará migrando sozinha na formação do gameta. A associação dos alelos aos subgenomas na forma AB/CD é representada pelo parâmetro $\delta_1=1$. Nestas condições, os parâmetros δ_2 e $\delta_3=0$.

Essas associações alelo-genoma representam as três diferentes formas em que os quatro alelos de um parental podem se associar a um subgenoma, definindo assim os diferentes pareamentos preferenciais AB/CD, AC/BD e AD/BC (Figura 2). O parâmetro β representa a frequência do fenômeno redução-dupla relativo à frequência de formação de quadrivalentes na meiose. A função log verossimilhança da distribuição multinomial é a função objetivo a ser maximizada. Portanto, o interesse é encontrar os parâmetros do modelo que resultem na maior log verossimilhança entre as probabilidades esperadas pelo modelo para cada genótipo dos gametas e sua frequência observada. Entretanto, não se sabe o tamanho de amostra dos gametas adequado para avaliar a segregação em um marcador de um determinado genótipo parental.

6. Objetivo

O objetivo deste trabalho foi investigar as propriedades do modelo teórico proposto por Stift (2008) através de simulações verificando segregação alélica a partir de parentais com diferentes genótipos e avaliando a média e o EQM das estimativas $\bar{\theta}$ para os valores $\bar{\theta} = \langle \tau, \beta, \delta_1, \delta_2, \delta_3 \rangle$ do modelo fornecidos para a geração da amostra. Além disso buscou-se observar a influência do tamanho da amostra nas estimativas do EQM.

CAPÍTULO II

Numeric evaluation of allelic segregation models in tetraploids: a simulation study

Artigo a ser submetido ao periódico *Genetics*

1 Sample size and numeric evaluation of allelic segregation models in tetraploids: A simulation
2 study

3 ^{1*}§Pinheiro FG, *Bisognin C, §Zanella CM, §Goetze M², §Buttow MV and §Bered F

4

5 **Departamento de Estatística, Instituto de Matemática, Universidade Federal do Rio Grande*
6 *do Sul, Avenida Bento Gonçalves 9500, P.O. Box 15053, 91501-970, Porto Alegre, RS, Brazil.*

7 *§Núcleo de Genética e Conservação de Plantas, Departamento de Genética, Instituto de*
8 *Biociências, Universidade Federal do Rio Grande do Sul, Avenida Bento Gonçalves 9500,*
9 *P.O. Box 15053, 91501-970, Porto Alegre, RS, Brazil.*

10

11 Short running title: Evaluation of segregation models in tetraploids

12

13 Key words: polyploidy, segregation models, tetrasomic inheritance,

14

15 ¹ Corresponding author: Felipe Grillo Pinheiro, Universidade Federal do Rio Grande do Sul,
16 Instituto de Biociências, P.O. Box 15053, 91501-970, Porto Alegre, RS, Brazil.

17 E-mail: pinheirofg@gmail.com

18

19 **ABSTRACT**

20 Polyploidy has been a challenging study area among geneticists especially in studies
21 focused on exploring polyploids inheritance patterns Autopolyploidy and allopolyploidy can
22 be considered extreme categories in which polyploids have been classified. Auto and
23 allotetraploids are expected to provide different gamete genotype frequencies after meiosis,
24 i.e. different frequencies among all possible allelic combinations from a given marker.
25 However, diploidization and homogenization process can affect the inheritance patterns of
26 allelic *loci* over time and generations, inducing shift from tetrasomic into disomic, in strict
27 autotetraploids, and from disomic into tetrasomic, in strict allopolyploids, respectively. These
28 explain the existence of tetraploids with intermediate inheritance between disomic and
29 tetrasomic. Stift *et al.* (2008) first developed segregation models to discriminate among
30 disomic, tetrasomic and several intermediate inheritance scenarios in tetraploids. Despite this
31 method having survived validation through tests in species with well-known segregation
32 patterns, no statistical property has been investigated. In this paper, we have build R-scripts to
33 simulate observed offspring data according to different parental genotypes, parameter
34 scenarios and sample sizes. In parental genotype ABCD, MSEs of all parameters reduced as
35 the gamete sample size increases. In ABCA and ABAA parental genotypes, two parameters
36 have MSE around 0.4 irrespective of parameter scenarios, and in the last case, the main
37 parameter model were strongly affected by parameter scenarios. For segregation analysis we
38 recommended researchers always to use fully-distinguished ABCD parental genotypes and, if
39 not possible, ABCA and its analogous ABCB or ABCC. We have examined the statistical
40 properties of this model and the usage of parental genotypes other than ABCD case, in order
41 to extend method applicability to less variable genetic markers or available individuals for

42 crossing. On the other hand, researchers should avoid ABAA and other non-fully
43 distinguished parental genotypes to estimate segregation parameters because they may not
44 lead to appropriated results, using this approach, even under large sample sizes.

45

46 Polyploidy has been a challenging study area among geneticists especially in studies
47 focused on exploring polyploids inheritance patterns (Olson 1997, Luo *et al.* 2004, 2006, Wu
48 *et al.* 2001, Stift *et al.* 2008). The inference of a tetraploid species inheritance pattern are of
49 great interest from both evolutionary perspective, and breeding purposes (Stift *et al.* 2008).
50 Each particular inheritance pattern may strongly affect segregation of genetic variation,
51 fitness, heterozigosity, production of novel phenotypes, ecological tolerances and fertility of
52 tetraploid offsprings, although in a different way (Kamiri *et al.* 2011, Stift *et al.* 2008, Wu *et*
53 *al.* 2001).

54 Autopolyploidy and allopolyploidy can be considered extreme categories in which
55 polyploids have been classified (Stebbins, 1971, Parisod *et al.* 2009). These two classes
56 essentially differ from each other in the origins of their genomes (set of chromosomes) and,
57 therefore, the consequences each genome origin implies to poliploids (Stift *et al.* 2008, Wu *et*
58 *al.* 2001). Autopolyploids may arise through fusion of unreduced gametes from genetically
59 similar or identical genomes, i.e, from one single species (e.g. parental unreduced gametes:
60 AA + AA → AAAA autotetraploid). During meiosis, all four homologue chromosomes may
61 pair as quadrivalents or random bivalents. Both bivalent and quadrivalent pairings lead to
62 tetrasomic segregation, i.e. all homologous chromosomes have equal opportunities to pair at
63 meiosis (Mather 1934, Stift *et al.* 2008, Wu *et al.* 2001). Moreover, when quadrivalent pairing
64 and crossing-over are present at meiosis, double reduction may occur. Allopolyploids usually
65 derived from hybridization events, which involves chromosome combination of distinct
66 genomes (i.e. different species), with subsequent chromosome doubling (e.g. parental normal
67 gametes from different genomes: A + B → AB hybrids → chromosome doubling → AABB
68 allotetraploid). Allotetraploids have two homeologous sets, i.e. A and B genomes, each one
69 consisting of two homologous chromosome sets (Stift *et al.* 2008, 2010). Chromosome pairing
70 occurs exclusively between homologous sets leading to bivalent pairing and disomic
71 inheritance, i.e. pairing between homeologous chromosome sets is not allowed and its specific
72 allelic associations will not be found on gametes.

73 However, diploidization and homogenization process can affect the inheritance
74 patterns of allelic *loci* over time and generations, inducing shift from tetrasomic into disomic,
75 in strict autotetraploids, and from disomic into tetrasomic, in strict allopolyploids, respectively
76 (Sybenga 1994, 1996, Soltis & Soltis 2000, Wolfe *et al.* 2001, Ramsey & Schemske 2002,
77 Stift *et al.* 2008). A tetraploid at some point in one of these processes may show inheritance
78 patterns, which are intermediate between disomic and tetrasomic. These individuals are
79 similar to individuals termed as segmental allopolyploids except by, in the first case, the
80 segregation of each chromosome set will be classified in itself, as intermediate, instead each
81 chromosome set has its segregation classified as disomic or tetrasomic and then, the overall
82 results from segregation markers characterize individuals as auto, allo or segmental
83 allopolyploids. Segmental allotetraploids also result from hybridization, however parental
84 species are close enough to maintain high degrees of homology, even between homeologous
85 chromosome sets, that justify polyvalent formation.

86 As an implication from disomic, tetrasomic or intermediate segregation, tetraploids are
87 expected to provide different gamete genotype frequencies after meiosis, i.e. different
88 frequencies among all possible allelic combinations from a given marker. Since diverse
89 segregation patterns are linked to distinguished types of polyploids, these models may help to
90 associate offspring genotype data to polyploidy origin in tetraploid species. Some segregation
91 models have been developed in order to discriminate among these several inheritance patterns
92 by analyzing gametes or offspring genotypes obtained through controlled crosses between
93 individuals of species under study (Olson 1997, Wu *et al.* 2001, Luo *et al.* 2004, Stift *et al.*
94 2008). Stift *et al.* (2008) first developed segregation models to discriminate among disomic,
95 tetrasomic and several intermediate inheritance scenarios in tetraploids. Authors designed a
96 five-parameter-model in which one random and three preferential pairing gamete proportions
97 and double reduction rate were estimated, based on maximum log-likelihood estimation
98 (MLE) of a multinomial distribution.

99 Stift's approach cannot be directly applied in cases where alleles are not fully
100 distinguished regarding parental genotype (e.g. AABC or AABB). It is not possible to
101 distinguish which alleles have multiple dosage (e.g. ambiguity as one single parental
102 phenotype can lead to different genotypes: ABC_= AABC, ABBC, ABCC or ABC0) by using
103 simple genotyping techniques, such as electrophoresis. (De Silva, *et al* 2005). Even if

104 assessment of complete parental genotype were possible, the observed gamete genotypes from
105 a parent, in which segregation is being investigated, are manually derived from the observed
106 offspring genotypes. For this reason, parental genotypes from crossing individuals have to be
107 carefully chosen in order to avoid ambiguity in these reconstructions (e.g. A0 and AA gamete
108 genotypes are not distinguishable, in the case of ABC0 parental genotype, since both have A
109 phenotype and 0 represents a null allele).

110 Despite this method having survived validation through tests in species with well-
111 known segregation patterns, no statistical property has been investigated in order to verify
112 whether model parameter estimations are biased, highly variable or fail to converge, according
113 to different possible parameter scenarios (i.e. parameter combinations) and parental genotype.
114 Since segregation studies may involve crossing experiments of wild populations *in loco*,
115 occasionally, some genetic markers or individuals available for crossing did not exhibit fully
116 distinguished alleles at that specific *locus* and the sampled genome amplitude might then be
117 compromised. Several segregation models are tested against a full tetrasomic null model for
118 every gamete genotype sample. Stift *et al.* 2010 calculated the sample sizes to distinguish the
119 best alternative model (i.e. largest log-likelihood) from null hypothesis (tetrasomic inheritance
120 model), for a 80%-power, in a four-allele-distinguished parental genotype (e.g ABCD).
121 However, suitable offspring sample size that guarantees sufficient numerical properties -if
122 they exist- to estimate model parameters also needs to be calculated for evaluation of different
123 parental genotypes segregation. These sample sizes have not been clearly suggested yet.

124 In this paper, we have developed functions in R-language that automatically generate
125 expected probabilities for all gamete genotype outcomes obtained from parental genotypes
126 under different allelic dosage including null alleles. These functions were also adapted to
127 support parental genotype data in cases where genotypic classes of gametes are ambiguous
128 (e.g. AABC, AAAB). We used these R-scripts to simulate observed offspring data and answer
129 the following questions: **I**) In the most informative case (i.e. ABCD parental genotype), do the
130 model meet minimum requirements such as precise and unbiased parameter estimations? **II**) Is
131 it viable to use non-fully distinguished parental genotypes to accurately access model
132 segregation parameters? How do these estimations behave in different parameter combination
133 scenarios and sample size increases? We will discuss the limitations of estimations of these

134 different parental genotype conditions under several parameter combination scenarios
135 regarding variable sample sizes.

136

137

SIMULATION STUDY

138

139 **Stift's model**

140

141 We based our approach on the gamete formation model developed in Stift *et al.* (2008).
142 Authors considered a tetraploid where each chromosome is marked by a different allele (e.g.
143 ABCD). Each possible allelic combination (exactly ten combinations: AA, ..., BC) was taken
144 as a class from a multinomial distribution. The exact probabilities for all these possible allelic
145 combinations were calculated in terms of the unknown parameter values that, after having
146 been estimated under different model situations, were chosen by the likelihood ratio test and
147 support evidence for any of the tested inheritance patterns (disomic, tetrasomic or
148 intermediate). Unknown parameters are estimated based on observed frequency of each
149 gamete genotype and they represent values that are most likely to have produced the observed
150 distribution (i.e. observed frequency of gamete genotypes) (see Stift *et al.* 2008).

151

152 **Gamete genotype and phenotype outcomes and their expected probabilities**

153

154 We started by building a general R-function (R Core Team, 2013) to list all possible
155 gamete genotype outcomes from a given parental genotype. It can be applied to any chosen
156 parental genotype (i.e. ABCD, AABC, ABC0, AABB, ..., A000, 0000). In all parental
157 genotype, except for ABCD, gamete genotypes represented by the same phenotype, were
158 joined in a single category represented by their gamete phenotype (e.g. in ABC0 parental
159 genotype AA and A0 gamete genotypes are represented by A phenotype). This approach was
160 developed in order to generalize the original model and include non-fully parental genotypes
161 that we want to evaluate. We have also modified this function to reduce the expected
162 probabilities of gamete genotype outcomes into gamete phenotype outcomes, if parameter
163 values $\bar{\theta}$ are informed in addition to parental genotype. These R-functions and some examples
164 are available in supplemental material.

165

166 **Model adaptations**

167

168 For R-package availability, routine management and simulation logistic reasons, we
169 have implemented Stiff's model in R-language. Parameters were estimated using `gosolnp`, a
170 nonlinear optimization function based on augmented Lagrange method, from `Rsolnp` R-
171 package (Alexios Ghalanos & Stefan Theussl, 2012) rather than a constrained nonlinear
172 regression (CNLR) used in the original model. Because `gosolnp` minimizes an objective
173 function (OF) instead of maximize, we have adapted the original OF as $-\log$ -likelihood of
174 multinomial distribution:

$$175 \quad -\ell(\vec{\theta} | obs_1, obs_2, \dots, obs_k) = -\left(\ln(n!) - \sum_{i=1}^k \ln(obs_i!) + \sum_{i=1}^k obs_i \ln(esp_i) \right). \quad (1)$$

176 where k is the number of possible gamete phenotype (or genotype in ABCD case) outcomes,
177 obs_i is the observed counts of i th gamete phenotype, $n = obs_1 + obs_2 + \dots + obs_k$ is the gamete
178 sample size, esp_i is the expected probability for the i th gamete phenotype and
179 $\vec{\theta} = \langle \tau, \beta, \delta_1, \delta_2, \delta_3 \rangle$ represents the parameter vector used to calculate expected probabilities of
180 each gamete phenotype \overline{esp} . The constrained functions $\delta_1 + \delta_2 + \delta_3 = 1$ and $\delta_1 \times \delta_2 \times \delta_3 = 0$
181 were implemented as linear and nonlinear restrictions, respectively.

182

183 **Simulations**

184

185 Simulation experiments were performed to demonstrate the statistical properties of
186 MLE under different model conditions. The experiment were designed to consider the effects
187 of different parental genotypes, different model situations and different sample sizes.

188 In this paper, we have evaluated three parental genotypes ABCD, ABCA and ABAA.
189 We have not computed simulations for parental genotypes containing null alleles (e.g. ABC0,
190 AAB0, A000). In the original model, in order to support evidence for a particular inheritance
191 pattern, parameters are estimated under seven model situations, according to the arrangement
192 of fixed parameters and the number of parameters to be estimated (Table 1). For
193 convenience/simplicity, we have chosen only three of these model situations (i.e. m_0 , m_1 and

194 m_6). Other model types and arrangements will have similar behavior by analogy. For a given
195 model situation, estimable parameters were set through several combination values to better
196 explore the parameters range, within their respective boundaries. Fixed parameter values were
197 set as in their model parameter arrangement. Therefore, simulation scenarios were defined by
198 combination of different values of $\tau=(0.1, 0.5, 0.9, 1)$, $\beta=(0, 1/6)$, $\Delta=\delta_1-\delta_2=(0, 0.5, 1)$ and $\delta_3=0$,
199 resulting in two, eight and 24 scenarios, according to parameter configuration m_0 , m_1 and m_6 ,
200 respectively.

201 We also have considered four gamete genotype sample sizes $n=(25,50,100,150)$. The
202 largest sample size was 150 because, in general, it is difficult to obtain offspring samples over
203 this value. We have used 200 replications ($re=200$) of each simulated experiment. This
204 number was decided based on the evaluation of some experiments where the mean, from the
205 first to n th estimation, stabilized. Mean squared error (MSE) and bias were used to assess the
206 estimation properties. We have considered a $MSE < 0.20$ as a sufficient error measure for the
207 estimatives, since the estimators empirical distributions seemed not to follow a normal
208 distribution.

209 Since objective function, equality constrained functions, parent genotypes, gamete
210 phenotype (or genotype) outcomes, parameter scenarios and sample size were defined,
211 simulations were performed as follows:

- 212 1) For a given sample size, parameter scenario (i.e. $\vec{\theta}$) and parental genotype,
213 produce gamete phenotypic classes and their respective probabilities using R-
214 function in supplemental material.
- 215 2) Generate re gamete phenotype sample replicates of size n using `rmultinom`
216 R-function and estimate their parameters $\vec{\theta}$ through `gosolnp`.
- 217 3) Calculate mean, bias, variance and mean squared error (MSE) of these
218 estimations $\vec{\theta}$ based on true values of $\vec{\theta}$ and save data.
- 219 4) Repeat steps 1-3 for each other sample size n , parameter scenario and parental
220 genotype.

221

222 Plots were separated in panels according to the combinations of true parameters $\tau=(0.1,$
223 $0.5, 0.9, 1)$, $\beta=(0, 1/6)$, $\Delta=\delta_1-\delta_2=(0, 0.5, 1)$ and number of fixed parameters $F=(1, 3, 4)$ (i.e.

224 according to the scenarios). Fixed parameters were not plotted. Moreover, since δ_3 is fixed,
225 any wrong estimation on δ_1 is associated to a bias and variance of same magnitude on
226 estimations of δ_2 . Consequently, δ_1 and δ_2 MSE will be exactly the same. For this reason we
227 have plotted only δ_1 . We also transform the MSEs of parameter β in order to adjust the scale
228 and allow parameter comparison because the limit of this parameter is 1/6 instead of 1.

229 The MSEs from different scenarios, for number of fixed parameters $F=1$ and parental
230 genotype ABCD, were plotted in Figure 1. One can note that (Figure 1), when true value of τ
231 increases the MSEs of $\hat{\tau}$ also increase in scenarios where $\beta=0$, but decrease if $\beta=1/6$. On the
232 other hand, as β parameter value increases from zero to one, $\hat{\beta}$ MSE increases especially
233 when τ becomes next to zero. Regarding different β s, as Δ increases, $\hat{\delta}_1$ MSEs decrease when
234 τ less than 0.5, but increases when τ approaches to one, especially in small sample sizes,
235 however as the sample size increases, $\hat{\delta}_1$ MSE convergence seems to enhance faster.

236 Scenarios where the three δ s were set fixed (i.e. $F=3$), in comparison to same scenarios
237 with only δ_3 fixed (i.e. $F=1$), show better estimations of τ and β since MSEs are lower (Figure
238 4), as well in $F=4$ scenery (Figure 5). Disregarding $\tau=1$, in $F=1$ scenarios (Figure 1), parental
239 genotype ABCD show adequate overall estimations $\bar{\theta}$ for the true parameters $\bar{\theta}$ considering
240 that, in the worst scenery (i.e. when $\tau=0.9$, $\beta= 1/6$ and $\Delta= 1$), MSEs of 0.27 or 0.18 are still
241 fairly reasonable for a sample size of 100 and 150, respectively.

242 Similar MSEs behavior was found in ABCA parental genotype concerning to $\hat{\tau}$ and $\hat{\beta}$
243 MSE (Figure 2). We have found high $\hat{\tau}$ and $\hat{\beta}$ MSEs in scenarios where β and τ , respectively,
244 are low. However, $\hat{\delta}_1$ MSEs increase as Δ increases, independently of τ . Moreover,
245 convergence of $\hat{\delta}_1$ MSEs seemed not be affected by different Δ as in ABCD parental
246 genotype. In $F=3$ simulation scenarios (Figure 4), $\hat{\tau}$ and $\hat{\beta}$ MSEs obtained were similar to
247 ones of $F=1$ scenarios. Overall MSE of τ and β behaved also well in this parental genotype.
248 Nevertheless, in scenarios where Δ is high, $\hat{\delta}_1$ MSEs will possibly have unconfident
249 estimations for δ s.

250 Parental genotype ABAA scenery simulations resulted in same consequences of
251 parental ABCA concerning $\hat{\beta}$ and $\hat{\delta}_1$ MSEs and its relations to different τ and Δ , respectively,
252 although $\hat{\tau}$ MSEs showed different results. In ABAA parental genotype, $\hat{\tau}$ MSEs started to

253 increase strongly to its maximum limit, as τ approached to one (Figure 3). $\hat{\tau}$ MSEs were more
254 than 0.25 in small sample sizes, even when $\beta=1/6$, but a reasonable increasing in that sample
255 can minimize this problem. Figure 5 compares the estimation of β between different parental
256 genotypes. $\hat{\beta}$ MSEs of ABCD and ABCA are lower than $\hat{\beta}$ MSEs from ABAA parental
257 genotype are.

258

259

DISCUSSION

260

261 In this paper, we have access the MSE of parameter estimations from the segregation
262 model proposed by Stift *et al* (2008). This model is a likelihood-based approach to estimate
263 parameters of a general tetraploid model that best fits observed segregation data. We have
264 examined the statistical properties of this model and the usage of parental genotypes other than
265 ABCD case, in order to extend method applicability to less variable genetic markers or
266 available individuals for crossing.

267 In parental genotype ABCD, MSEs of all parameters reduced as the gamete sample
268 size increases. Because τ and β parameter are positively associated, scenarios where one of
269 them is high and the other is low, can lead to high values of MSE from the higher parameter.
270 However, despite most parameter estimations shows reasonable MSEs, MSE from δ_1
271 estimations were higher in scenarios where τ are close to one, but lower as τ decreases. This
272 can be happening because when τ is low, it means that most part of the gamete are formed by
273 preferential pairing of some allele-subgenome association, in this case represented by δ_1 , and
274 the estimations of δ_1 can be quite confident. When τ becomes high, most part of gametes are
275 formed by random association between homologue chromosomes and, therefore, the
276 proportion of gamete genotypes explained by a preferential pairing is so small that could be
277 explained by any allele-subgenome association (i.e. δ_1 or δ_2). This might lead to high values of
278 δ_1 MSE. In addition, when the difference between $\delta_1-\delta_2$ is high (i.e. high values of Δ), which
279 means that the association AB/CD is the most likely to have produced the gametes explained
280 by preferential pairing, an increasing in sample size induce reduction in δ_1 MSEs. This may
281 happened because gamete genotypes that depends on δ_1 will show up more frequently than if
282 preferential pairing was not majority influenced by δ_1 . With these results in mind, we are able
283 to answer the first question: In the most informative case (i.e. ABCD parental genotype), do

284 the model meet minimum requirements such as precise and unbiased parameter estimations?
285 Yes, but it seems that it is necessary to increase the sample size in order to guarantee
286 reasonable estimations of δ_s , in case of the true τ are close to its maximum value. We have
287 noted that when $\tau=1$, in $F=1$ scenarios, even $\bar{\theta}$ with wrong δ_1 estimations, produce similar
288 log-likelihood to $F=3$ scenarios. For this reason and because these scenarios are too close to
289 tetrasomic null model we do not need to be so rigid on that wrong δ_1 estimations, since this
290 wrong estimations will not explain the segregation better than the null model.

291 The answer for the second question: Is it viable to use non-fully distinguished parental
292 genotypes to accurately access model segregation parameters? How do these estimations
293 behave in different parameter combination scenarios and sample size increases? arise from
294 results of ABCA and ABAA parental genotypes. High $\hat{\delta}_1$ MSEs were more evident as we
295 analyze non-fully distinguished parental genotypes. For ABCA and ABAA parental genotype,
296 allele A in duplex or triplex dosages will reduce the number of gamete genotypic classes, and
297 all possible gamete genotypes will have more similar probabilities irrespective of their
298 scenarios and, consequently, $\hat{\delta}_1$ MSEs will be higher. However, as we suggest in the case of
299 ABCD parental genotype, we do not need to exclude the analysis of ABCA parental genotype
300 because of the high $\hat{\delta}_1$ MSEs its scenarios are subjected, but we have to keep in mind the
301 possibility of error in δ_s estimations. Furthermore, for parental genotype ABAA, only BB
302 gamete genotype left to reflect the real double reduction, because we are not able to
303 distinguish AA from double reduction from AA by normal combination between alleles from
304 two different chromosomes. When there is no double reduction (i.e. $\beta=0$), BB gamete
305 genotype will be absent and the $\hat{\tau}$ MSEs will be strongly high because possible gamete
306 genotypes show similar probabilities in each scenario. This problem did not affect scenarios
307 when double reduction rate was 1/6, but only if sample has reasonable size, at least 100,
308 otherwise BB gamete genotype will not be present and we might face error in estimations of τ .

309 In summary, for segregation analysis we recommended researchers always to use fully-
310 distinguished ABCD parental genotypes and, if not possible, ABCA and its analogous ABCB
311 or ABCC. Despite the problems a small sample size can imply to parameter estimations,
312 especially to δ_s , we have observed that, in ABCD, ABCA parental genotype, gamete genotype
313 samples of size 50 already produced τ and β MSEs of 0.15 or less. This means that even

314 subjected to the possibility high δ_1 MSE values, we are able to measure appropriately the real
315 proportion of gametes formed by a random chromosome pairing, although we might wrongly
316 estimate allele-subgenome associations. On the other hand, researchers should avoid ABAA,
317 ABBB, ACCC and other non-fully distinguished parental genotypes to estimate segregation
318 parameters because they may not lead to appropriated results, using this approach, even under
319 large sample sizes

320

321 REFERENCES CITED

322

323 Alexios Ghalanos and Stefan Theussl (2012). Rsolnp: General Non-linear Optimization Using
324 Augmented Lagrange Multiplier Method. R package version 1.14.

325 De Silva HN: Estimation of allele frequencies in polyploids under certain patterns of
326 inheritance. *Heredity* 2005, 95:327

327 Doyle, J.J. & Doyle, J.L. (1987) A rapid DNA isolation procedure for small quantities of fresh
328 leaf tissue. *Phytochemical Bulletin*, 19, 11-15.

329 Jackson, R.C. & Jackson, J.W. (1996) Gene segregation in autotetraploids: Prediction from
330 meiotic configurations. *American Journal of Botany*, 83, 673-678.

331 Kamiri M, Stift M, Srairi I, Costantino G, Moussadik A, Hmyene A. *et al.* (2011). Evidence
332 for non-disomic inheritance in a Citrus interspecific tetraploid somatic
333 between *C. reticulata* and *C. lemon* hybrid using SSR markers and cytogenetic
334 analysis. *Plant Cell Rep* (doi:10.1007/s00299-011-1050-x).

335 Luo, Z.W., Zhang, R.M. & Kearsey, M.J. (2004) Theoretical basis for genetic linkage analysis
336 in autotetraploid species. *Proceedings of the National Academy of Sciences of the United*
337 *States of America*, 101, 7040-7045.

338 Luo, Z.W., Zhang, Z., Zhang, R.M., Pandey, M., Gailing, O., Hattemer, H.H. & Finkeldey, R.
339 (2006) Modeling population genetic data in autotetraploid species. *Genetics*, 172, 639-
340 646.

341 Mather, K. (1934) Reductional and equational separation of the chromosomes in bivalents and
342 multivalents. *Journal of Genetics*, 30, 53-78.

343 Otto, S.P. & Whitton, J. (2000) Polyploid incidence and evolution. *Annual Review of*
344 *Genetics*, 34, 401-437.

345 R Core Team (2013). R: A language and environment for statistical computing. R Foundation
346 for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

347 Ramsey, J. & Schemske, D.W. (2002) Neopolyploidy in flowering plants. Annual Review of
348 Ecology and Systematics, 33, 589-639.

349 Soltis, P.S. & Soltis, D.E. (2000) The role of genetic and genomic attributes in the success of
350 polyploids. Proceedings of the National Academy of Sciences of the United States of
351 America, 97, 7051-7057.

352 Stift M, Berenos C, Kuperus P, Van Tienderen P (2008) Segregation models for disomic,
353 tetrasomic and intermediate inheritance in tetraploids: a general procedure applied
354 to *Rorippa* (Yellow Cress) microsatellite data. Genetics 179:2113–2123

355 Stift M, Reeve R, van Tienderen PH (2010) Inheritance in tetraploid yeast revisited:
356 segregation patterns and statistical power under different inheritance models. J Evol Biol
357 23:1570–1578

358 Stebbins, G.L. (1971) Chromosomal evolution in higher plants. Columbia University Press,
359 New York & London.

360 Sybenga, J. (1994) Preferential pairing estimates from multivalent frequencies in tetraploids.
361 Genome, 37, 1045-1055.

362 Sybenga, J. (1996) Chromosome pairing affinity and quadrivalent formation in polyploids:
363 Do segmental allopolyploids exist? Genome, 39, 1176-1184.

364 Wolfe, K.H. (2001) Yesterday's polyploids and the mystery of diploidization. Nature Reviews
365 Genetics, 2, 333-341.

366 Wu, R.L., Gallo-Meagher, M., Littell, R.C. & Zeng, Z.B. (2001) A general polyploid model
367 for analyzing gene segregation in outcrossing tetraploid species. Genetics, 159, 869-882
368
369

370 **TABLES**

371 **Table 1** Model types, parameter configurations, number of fixed parameters and number of
 372 simulated scenarios according to parameter

Model Type	Parameter Arrangement	τ	β	δ_1	δ_2	δ_3	N° of Fixed Parameters (F)	N° of Simulated Scenarios
Tetrasomic Null Model	m_0	1	•	x	x	x	4	2*
Constrained Intermediate Models	m_1	•	•	1	0	0	3	8*
	m_2	•	•	0	1	0	3	-
	m_3	•	•	0	0	1	3	-
Unconstrained Intermediate Models	m_4	•	•	0	•	•	1	-
	m_5	•	•	•	0	•	1	-
	m_6	•	•	•	•	0	1	24*

• Parameters to be estimated by the model. Scenarios were formed by their combinations.

x Any value because when $\tau=1$ δ_1, δ_2 and δ_3 cancel out in the model

* Scenarios of model types evaluated in simulations

373

374 **LEGENDS OF FIGURES**

375 **Figure 1** Parameter MSE of simulations in ABCD parental genotype under 24 different
 376 scenarios with one fixed parameter ($\delta_3=0$).

377

378 **Figure 2** Parameter MSE of simulations in ABCA parental genotype under 24 different
 379 scenarios with one fixed parameter ($\delta_3=0$).

380

381 **Figure 3** Parameter MSE of simulations in ABAA parental genotype under 24 different
 382 scenarios with one fixed parameter ($\delta_3=0$).

383

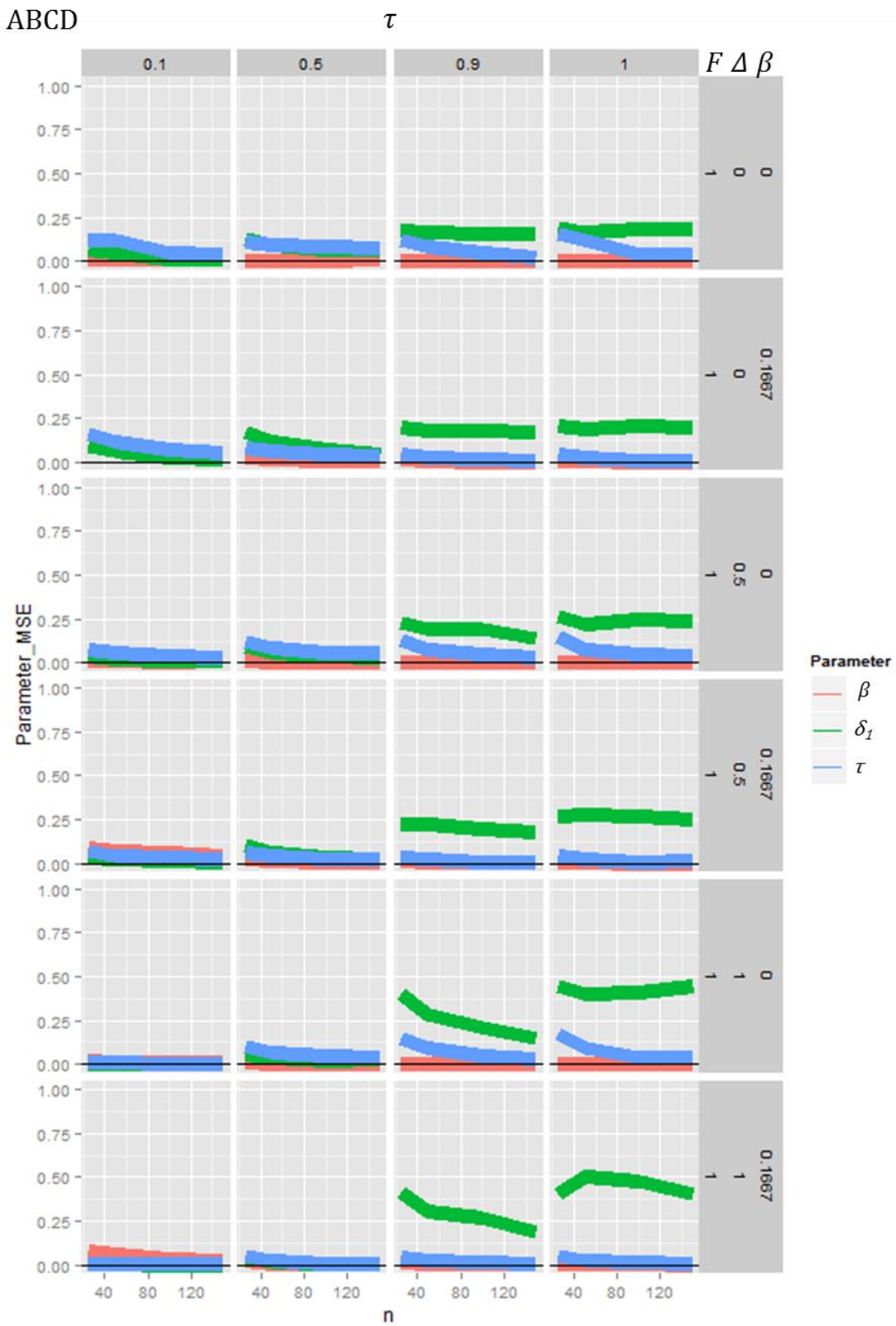
384 **Figure 4** Parameter MSE of simulations in ABCD, ABCA and ABAA parental genotypes
 385 under eight different scenarios with three fixed parameter ($\delta_2=0, \delta_3=0$ and $\delta_1=1$).

386

387 **Figure 5** Parameter MSE of simulations in ABCD, ABCA and ABAA parental genotypes
 388 under two different scenarios with four fixed parameter ($\delta_2=0, \delta_3=0, \delta_1=1$ and $\tau=1$).

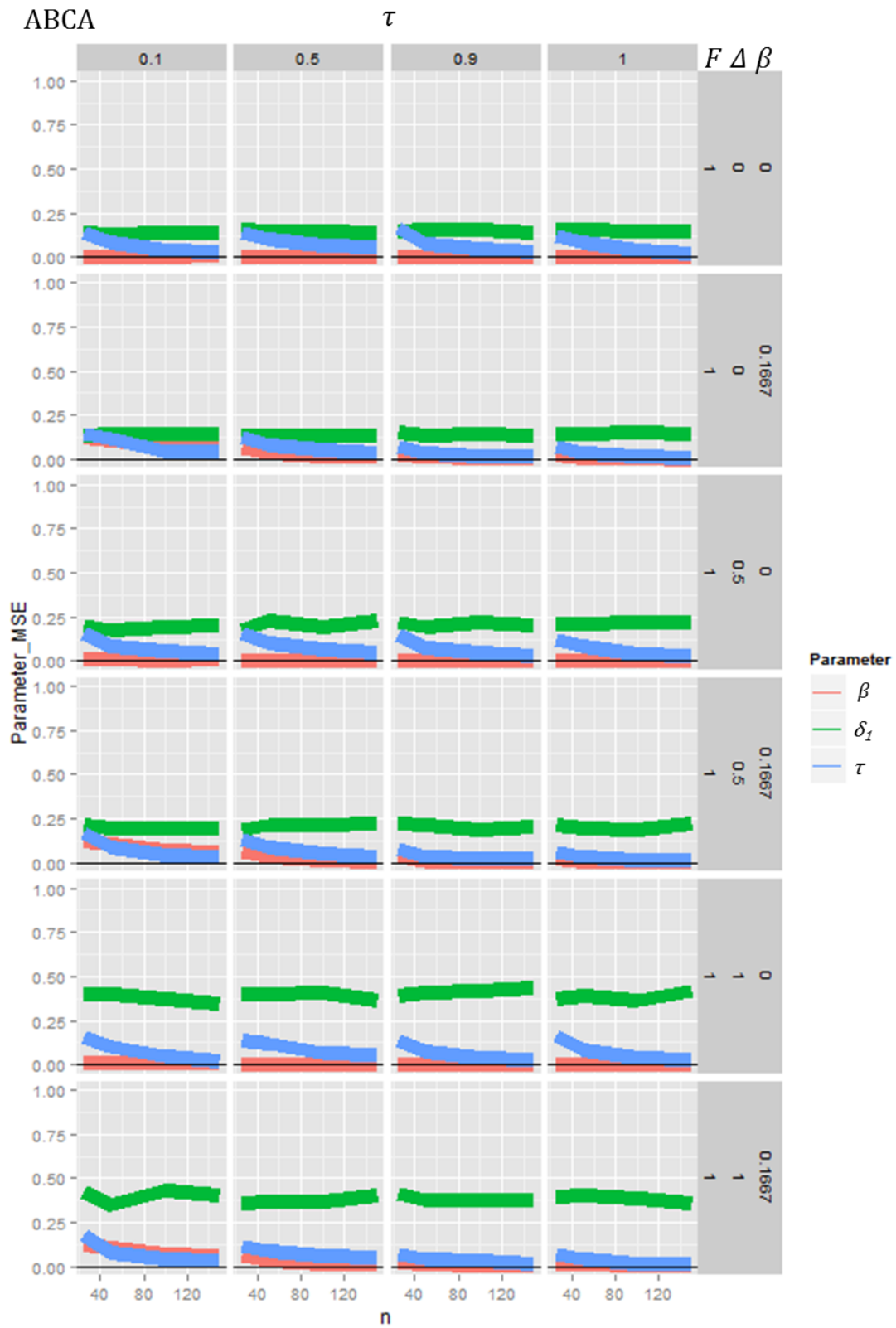
389

390 Figure 1

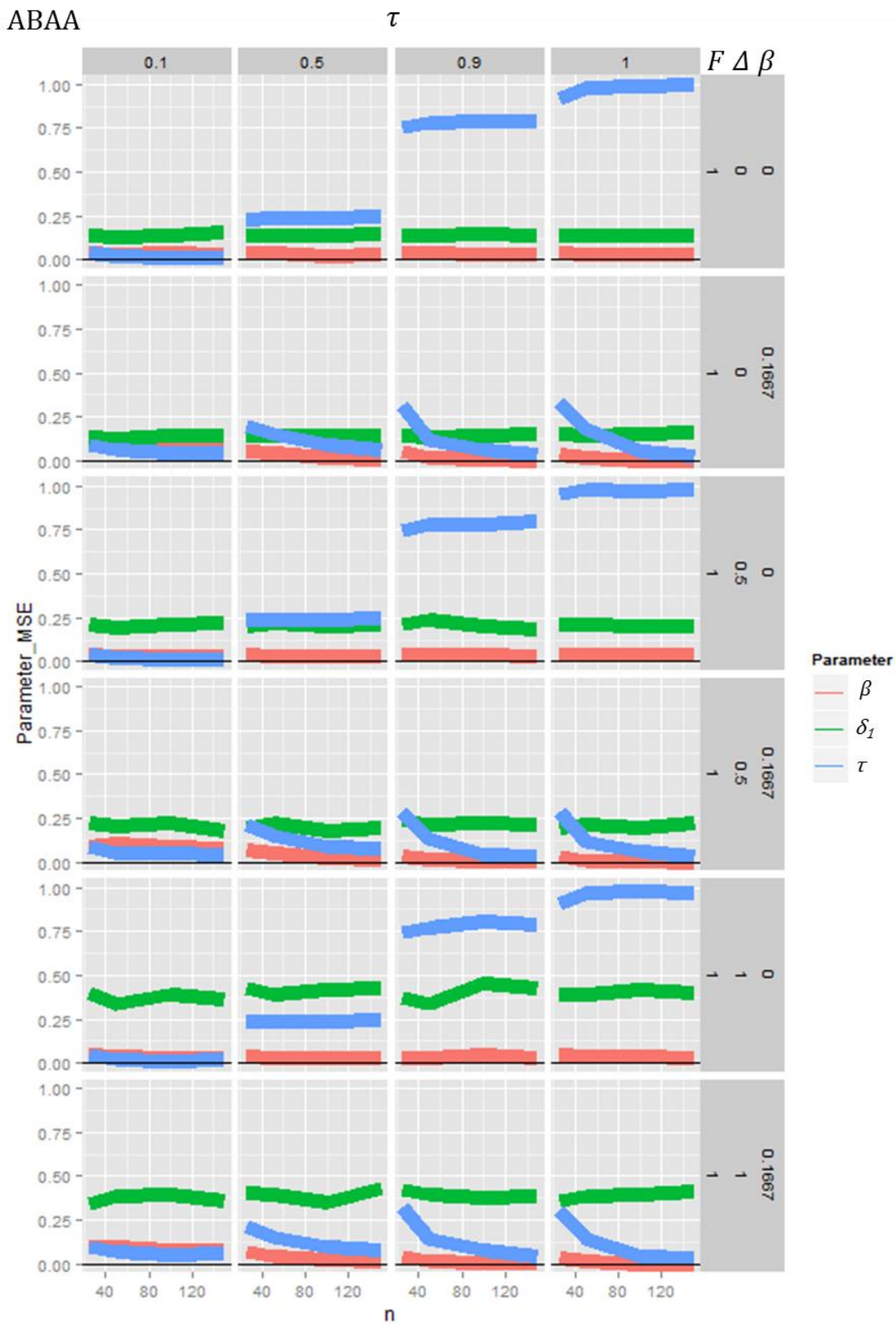


391

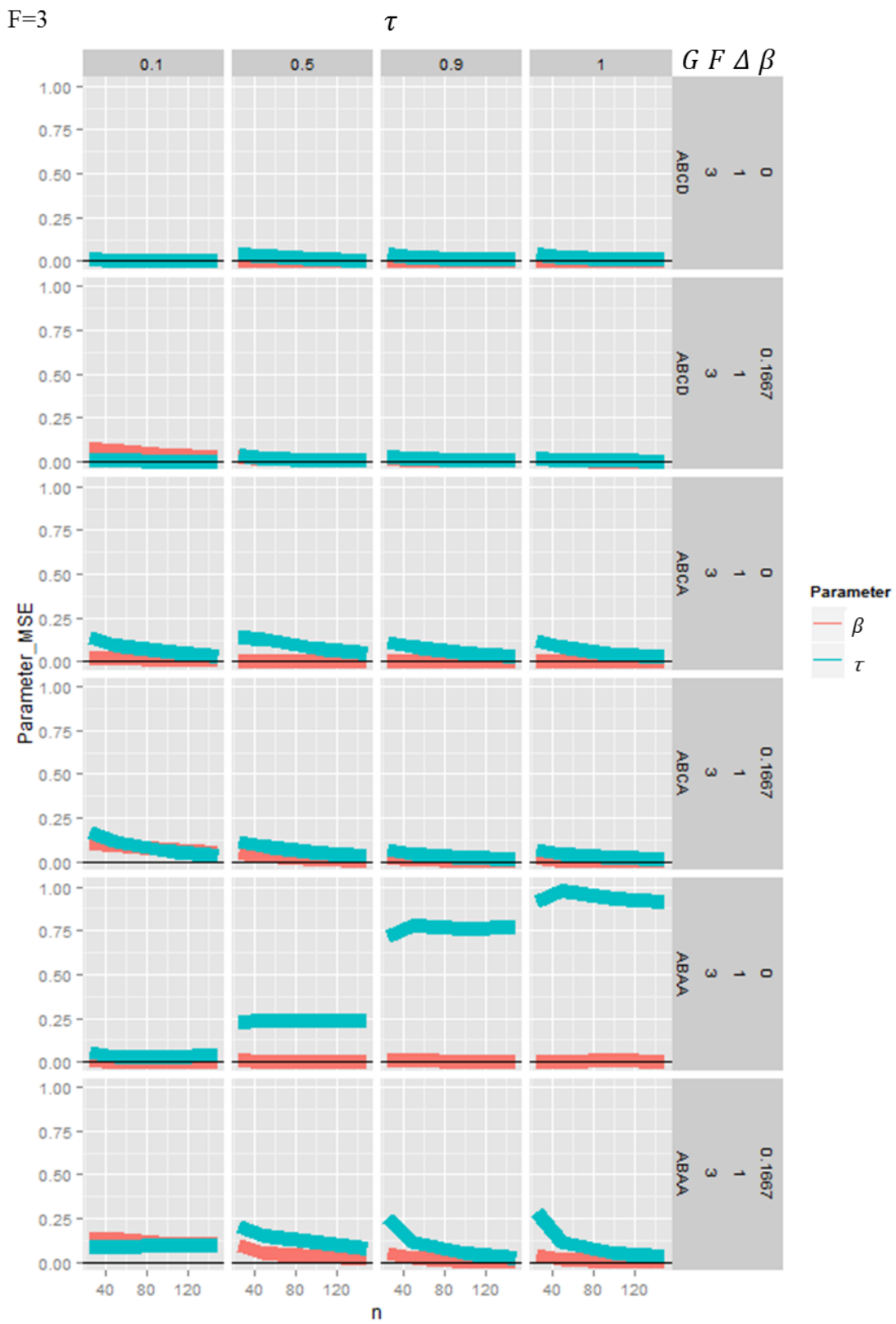
392 Figure 2



393

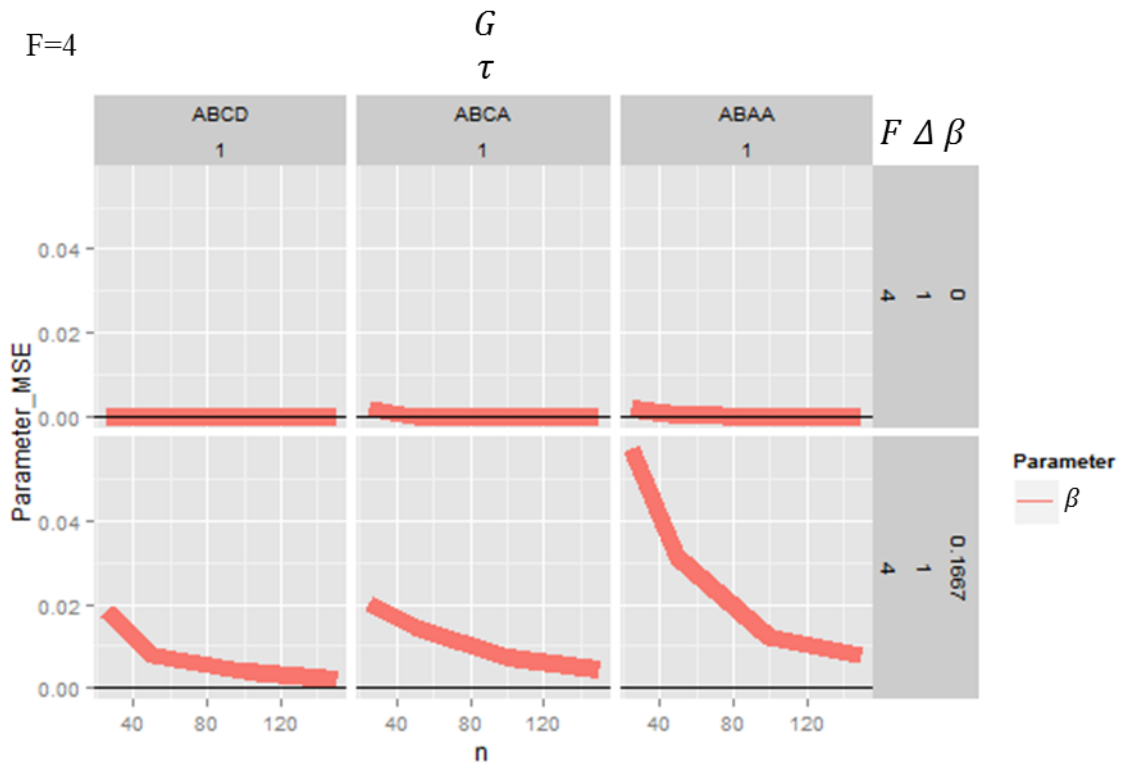


396 Figure 4



397

398 Figure 5



399
 400
 401
 402
 403
 404

CAPÍTULO III

Considerações Finais

CONSIDERAÇÕES FINAIS

Como mencionado na primeira frase do artigo, a poliploidia tem sido uma área de estudo muito desafiadora especialmente no que diz respeito aos trabalhos que buscam traduzir a segregação alélica em tetraploides através de modelos matemáticos. Parte desse desafio se deve ao conhecimento e interesse multidisciplinares, por parte do pesquisador, necessário para a realização destes estudos. Tanto o conhecimento teórico da meiose, de biologia molecular e da genética de poliploides quanto o entendimento de distribuições de probabilidade, métodos de estimação de parâmetros e de otimização de funções, entre outros pormenores, são indispensáveis.

Através das simulações dos genótipos parentais ABCD, ABCA e ABAA, em diversos cenários, representados pelas combinações de valores dos parâmetros a seres estimados, com diferentes números de parâmetros a serem estimados e diferentes tamanho amostrais verificou-se as propriedades estatísticas, mais especificamente o EQM, de um dos últimos modelos de segregação alélica proposto na literatura. As simulações revelaram que o EQM dos parâmetros diminuem com o aumento do tamanho da amostra para o caso ABCD e na maioria dos parâmetros no genótipo ABCA. Entretanto para o genótipo parental ABAA, determinados cenários podem levar a grandes EQMs do parâmetro τ . Apoiando-se nesses resultados, sugere-se que os pesquisadores utilizem o genótipo parental ABCD sempre que possível, sendo o genótipo parental ABCA também passível de análise. Além disso, a utilização do genótipo ABAA deve ser evitada.

Estes genótipos parentais representam apenas uma porção dos possíveis genótipos existentes em tetraploides. Nos próximos passos serão analisados os demais casos, incluindo genótipos parentais com alelos nulos. Além disso, pretende-se expandir a análise para uma avaliação conjunta do genótipo da progênie originada pelo cruzamento entre dois parentais ao invés de se trabalhar com a reconstrução da frequência dos gametas de um único parental. Caso os EQMs e demais estatísticas como a variância e vício continuem se mostrando grandes, buscar-se-á outras alternativas de estimação dos parâmetros, como a máxima verossimilhança penalizada, a alteração da estrutura do modelo na tentativa de reduzir o número de parâmetros estimados ou ainda alternativas que reduzam as estimativas de erro. Além disso, uma maior quantidade de tamanhos amostrais será avaliada com o fim de definir

os tamanhos amostrais adequados para a avaliação da segregação alélica em progênies derivadas do cruzamento de parentais com determinado genótipo. Ressalta-se que a versão do artigo apresentada neste trabalho é, na realidade, seu primeiro esboço, portanto necessita de correções.

CAPÍTULO IV

Referências Bibliográficas

REFERÊNCIAS BIBLIOGRÁFICAS

- ARNAU, G. et al. Revision of ploidy status of *Dioscorea alata* L. (Dioscoreaceae) by cytogenetic and microsatellite segregation analysis. *Theoretical and Applied Genetics* [S.I.], v. 118, n. 7, p. 1239-1249. 2009
- BEVER, J. D.; FELBER, F. The theoretical population genetics of autopolyploidy. *Oxford Surveys in Evolutionary Biology* [S.I.], v. 8, p. 185-217, 1992.
- BOUSALEM, M. et al. Microsatellite segregation analysis and cytogenetic evidence for tetrasomic inheritance in the American yam *Dioscorea trifida* and a new basic chromosome number in the Dioscoreae. *Theoretical and Applied Genetics* [S.I.], v. 113, p. 439-451, 2006.
- DE SILVA, H. N. et al. Estimation of allele frequencies in polyploids under certain patterns of inheritance. *Heredity* [S.I.], v. 95, p. 327-334, 2005.
- 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.
- GRANT, V. *Plant speciation*. New York, NY, USA: Columbia University Press, p.435, 1971.
- JACKSON, R. C. Polyploidy and diploidy: new perspectives on chromosome pairing and its evolutionary implications. *American Journal of Botany* [S.I.], v. 69, p. 1512-1523, 1982.
- MAHY, G. et al. Allozyme evidence for genetic autopolyploidy and high genetic diversity in tetraploid cranberry, *Vaccinium oxycoccos* (Ericaceae). *American Journal of Botany* [S.I.], v. 87, n. 12, p. 1882-1889, 2000.
- OLSON, M. S. Bayesian procedures for discriminating among hypotheses with discrete distributions: inheritance in the tetraploid *Astilbe biternata*. *Genetics* [S.I.], v. 147, p. 1933-1942, 1997.
- OTTO, S. P. The evolutionary consequences of polyploidy. *Cell* [S.I.], v. 131 p. 452-462, 2007.
- OTTO, S. P.; WHITTON, J. Polyploid incidence and evolution. *Annual Review of Genetics* [S.I.], v. 34, p. 401-4037, 2000.
- PARISOD, C.; HOLDEREGGER, R.; BROCHMANN, C. Evolutionary consequences of autopolyploidy. *New Phytologist* [S.I.], v. 186, n. 1, p. 5-17, 2009.
- RAMSEY, J.; SCHEMSKE, D.W. Neopolyploidy in flowering plants. *Annual Review of Ecology and Systematics*, Palo Alto, v.33, p. 589-639. 2002.

- 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.
- RIESEBERG, L.H. Polyploid evolution: keeping the peace at genomic reunions. *Current Biology*, London, v. 11, p. 925-928. 2001.
- RONFORT J. et al. Analysis of population structure in autotetraploid species. *Genetics* [S.I.], v. 150, p. 921-930, 1998.
- SCHIFINO-WITTMANN, M. T. Poliploidia e seu impacto na origem e evolução das plantas silvestres e cultivadas. *R. Bras. Agrociência* [S.I.], v. 10, n. 2, p. 151-157, 2004.
- SOLTIS D. E.; SOLTIS, P. S. Molecular-data and the dynamic nature of polyploidy. *Critical Reviews in Plant Sciences* [S.I.], v. 12, p. 243-273, 1993.
- 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.
- STEBBINS, G.L. *Chromosomal evolution in higher plants*. London, UK. p. 216, 1971.
- 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.
- SYBENGA, J. Chromosome pairing affinity and quadrivalent formation in polyploids: Do segmental allopolyploids exist? *Genome* [S.I.], v. 39, p. 1176-1184, 1996.
- WOLFE, K.H. Yesterday's polyploids and the mystery of diploidization. *Nature Reviews Genetics*, New York, v.2, p. 333-341, 2001.
- WU, R. L.; MA, C.-X.; CASELLA G. A mixed polyploid model for linkage analysis in tetraploids. *Journal of Computational Biology* [S.I.], v. 11, p. 562-580, 2004.
- WU, S. S. et al. A multivalent pairing model of linkage analysis in autotetraploids. *Genetics* [S.I.], v. 159, p.1339-1350, 2001.