



JURACY MENDES MOREIRA

**TAMANHO ÓTIMO DE AMOSTRA BIOLÓGICA
PARA ESTUDO NA PCR EM TEMPO REAL**

**LAVRAS – MG
2012**

JURACY MENDES MOREIRA

**TAMANHO ÓTIMO DE AMOSTRA BIOLÓGICA PARA ESTUDO NA
PCR EM TEMPO REAL**

Dissertação apresentada à Universidade Federal de Lavras como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, para a obtenção do título de Mestre.

Orientador

Dr. Paulo César Lima

Co-orientador

Dr. Daniel Furtado Ferreira

LAVRAS - MG

2012

**Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da
Biblioteca da UFLA**

Moreira, Juracy Mendes.

Tamanho ótimo de amostra biológica para estudo na PCR em tempo real / Juracy Mendes Moreira. – Lavras : UFLA, 2012.
59 p. : il.

Dissertação (mestrado) – Universidade Federal de Lavras, 2012.

Orientador: Paulo César Lima.

Bibliografia.

1. qPCR. 2. Dois estágios. 3. Amostragem hierárquica. 4.
Expressão gênica. I. Universidade Federal de Lavras. II. Título.

CDD – 519.5

JURACY MENDES MOREIRA

**TAMANHO ÓTIMO DE AMOSTRA BIOLÓGICA PARA ESTUDO NA
PCR EM TEMPO REAL**

Dissertação apresentada à Universidade Federal de Lavras como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, para a obtenção do título de Mestre.

APROVADA em 05 de junho de 2012.

Dr. Renato Ribeiro de Lima UFLA

PhD. Antônio Chalfun Júnior UFLA

Dr. Paulo César Lima
Orientador

LAVRAS - MG

2012

Com muito amor, à minha esposa Sara,
e aos meus filhos Guilherme, Leonardo e Pedro Vitor.

DEDICO

AGRADECIMENTOS

A Deus, pela presença constante em minha vida e por ter me permitido mais essa conquista. A minha adorada esposa e aos meus filhos, meus mais preciosos bens, obrigado pelo carinho e incentivo. Aos meus pais, pelo apoio em todos os momentos, fazendo-se sempre presentes ao longo desta caminhada, e por serem exemplos de luta e determinação para mim, aos meus irmãos, sogro, sogra e cunhados.

Ao meu Professor Orientador e, acima de tudo, meu amigo prof. Dr. Paulo César Lima, pela sábia e dedicada orientação, pelo apoio e compreensão em todos os momentos, requisitos que contribuíram para a conclusão deste trabalho.

Ao meu co-orientador, prof. Dr. Daniel Furtado Ferreira, pela paciência com que me coorientou, pela disponibilidade em auxiliar-me a qualquer momento, pelas críticas e sugestões.

A todos do corpo docente do Departamento de Ciências Exatas desta universidade, em especial ao prof. Dr. Renato Ribeiro de Lima, pela colaboração, amizade e dedicação. Seus ensinamentos foram de grande relevância para a realização deste trabalho e meu crescimento profissional.

Aos funcionários do Departamento de Ciências Exatas, pela simpatia e boa vontade no atendimento.

À Universidade Federal de Lavras e ao Departamento de Ciências Exatas, pela oportunidade concedida para a realização do mestrado. E à cidade de Lavras pela acolhedora hospitalidade durante a realização do curso.

A Capes (Coordenação de aperfeiçoamento de Nível superior), pela concessão da bolsa de estudos.

Ao amigo Juliano Bortolini, pela amizade, companheirismo nos estudos e pela paciência no convívio diário e a todos os demais colegas de turma.

Meu mais sincero agradecimento ao professor Paulo Eustáquio Rezende do Nascimento (Ex. Reitor da FESURV - Universidade de Rio Verde) pela concessão da bolsa de estudos durante a graduação, sem a qual não teria sido possível a realização desse mestrado.

Aos professores Nagib Yassin, Marcelo Judce e Mozaniel Batista da Silva da FESURV - Universidade de Rio Verde e José Waldemar da Silva da UFU - Universidade Federal de Uberlândia.

Ao doutorando em Agronomia/Fisiologia Vegetal desta universidade André Almeida Lima, pela concessão dos dados. Ao prof. PhD. Antônio Chalfun Júnior, pela valorosa contribuição na minha banca de qualificação e defesa.

SARA

Dizem que a mulher é o sexo frágil, mas que mentira absurda! Eu que faço parte da rotina de uma delas sei que a força está com elas...

Vejam como é forte a que eu conheço, sua sapiência não tem preço, satisfaz meu ego se fingindo submissa, mas no fundo me enfeitiça...

Quando eu chego em casa à noitinha quero uma mulher só minha, mas pra quem deu luz não tem mais jeito porque o filho quer seu peito...

O outro já reclama a sua mão e o outro quer o amor que ela tiver quatro homens dependentes e carentes da força da mulher...

Mulher! Mulher! Do barro de que você foi gerada me veio inspiração pra decantar você nessa canção...

Mulher! Mulher! Na escola em que você foi ensinada jamais tirei um 10 sou forte, mas não chego aos seus pés...

Erasmoo Carlos

RESUMO

A reação em cadeia da polimerase (PCR) é uma técnica científica usada na biologia molecular para amplificar fragmentos de uma amostra de DNA ou RNA. Uma inovação na PCR denominada de PCR em tempo real (qPCR), trouxe mais rapidez e precisão nos diagnósticos clínicos e laboratoriais. A reação em cadeia da polimerase da transcrição reversa em tempo real (RTq-PCR) é outro procedimento semelhante a qPCR com a vantagem de ser realizada em apenas uma etapa. As pesquisas genéticas na cultura do café, têm tido grande expansão e, na maioria dos casos, as amostras são constituídas da coleta de folhas ou de frutos em diferentes plantas constituindo amostragem em mais de um estágio. Na amostragem em dois estágios ou sub-amostragem, designada amostragem hierárquica, a população é constituída por N_1 unidades primárias e cada unidade primária por N_2 indivíduos. São selecionadas n_1 unidades primárias e, de cada uma delas, selecionados n_2 indivíduos. Para a determinação do tamanho ótimo da amostra biológica é necessário que se tenha dados obtidos de experimentos bem conduzidos e que expressem fielmente a variabilidade entre plantas de café e entre frutos, nas plantas, para condições que possam variar de acordo com os genes pesquisados. Em geral, o tamanho da amostra biológica utilizado pode estar sendo subestimado em função principalmente da relação entre as variâncias.

Palavras-chave: qPCR. Dois estágios. Amostra biológica. Expressão gênica.

ABSTRACT

The polymerase chain reaction (PCR) is a scientific technique used in molecular biology for amplifying fragments of a sample of DNA or RNA. An innovation in the PCR called real time PCR (qPCR) brought more quickly and accurately for clinical and laboratory diagnostic. Thereverse polymerase chain reaction transcription in real time (RTq-PCR) is another procedure similar to qPCR with the advantage of being performed in one step. Genetic research in coffee culture have been greatly expanded and in most cases the samples consist of collecting leaves or fruits of different plants representing more than one sampling stage. In the sampling int wost age sorsub-sampling, called hierarchical sampling, the population consists of N_1 primary units, and each primary uni consist of N_2 individuals. n_1 primary units are selected, and n_2 individuals are selected from each of them. To determine the optimum size of the biological sample is necessary to have data obtained from experiments conducted and which faith fully express the variability in coffee plants and fruits of plant stoallow the conditions may vary according to the genes studied. In general the size of biological sample used can be underestimated due mainly the relationship between the variances.

Keywords: qPCR. Two stages. Biological samples. Gene expression.

SUMÁRIO

1	INTRODUÇÃO.....	11
2	REVISÃO DE LITERATURA.....	13
2.1	Reação em cadeia da polimerase (PCR).....	13
2.1.1	Reação em cadeia da polimerase em tempo real (qPCR)....	13
2.1.2	Amostragem na cultura do café para estudos com PCR	15
2.2	Fundamentos Teóricos.....	17
2.2.1	Modelo para amostragem hierárquica em dois estágios.....	18
2.2.2	Variância da média amostral - modelo aleatório.....	19
2.2.3	Componentes de variância - modelo aleatório.....	19
2.2.4	Função de custos para amostragem hierárquica.....	21
2.2.5	Tamanho ótimo de amostra.....	21
3	MATERIAL E MÉTODOS.....	24
3.1	Material.....	24
3.2	Métodos.....	25
3.2.1	Amostragem hierárquica em dois estágios.....	25
3.2.2	Variância da média amostral em dois estágios.....	26
3.2.3	Componentes de variância em dois estágios.....	27
3.2.4	Função de custos em dois estágios.....	28
3.2.5	Tamanho ótimo de amostra em dois estágios.....	28
3.2.6	Proporções entre custos de amostragem e entre variâncias	30
3.2.7	Simulação de dados.....	32
4	RESULTADOS E DISCUSSÃO.....	34
4.1	Número ótimo de frutos por planta.....	34
4.2	Número ótimo de plantas na amostra.....	39
5	CONCLUSÃO.....	55
	REFERÊNCIAS.....	56

1 INTRODUÇÃO

A reação em cadeia da polimerase (PCR) é uma técnica científica usada na biologia molecular para amplificar fragmentos de uma amostra que pode ser de DNA ou RNA, e replicá-los várias vezes permitindo a detecção de um sinal de fluorescência que é emitido, quando uma molécula repórter hibridiza com a sequência alvo.

A técnica da PCR é importante para tornar válidos experimentos que envolvem a expressão de um gene, formando cadeias de DNA e amplificando-as de modo a permitir sua quantificação, para, em seguida, possibilitar a comparação com outro gene. A PCR é também muito importante no estudo da genética molecular e na determinação rápida da paternidade e no diagnóstico de doenças infecciosas.

Uma inovação na PCR, denominada PCR em tempo real, (qPCR) trouxe mais rapidez e precisão nos diagnósticos clínicos e laboratoriais e o uso dessa técnica abriu grandes perspectivas na análise da expressão gênica.

A reação em cadeia da polimerase da transcrição reversa, em tempo real, (RTq-PCR) é outro procedimento semelhante a qPCR com a vantagem de ser realizada em apenas uma etapa ao invés de duas como na qPCR. Apesar de já ser bastante comum, a RTq-PCR é muito passível de variações principalmente durante a fase da transcrição reversa.

Um dos fundamentos experimentais, ainda pouco estudado, refere-se ao tipo de amostra biológica necessário para a obtenção de resultados estatisticamente confiáveis. A maioria dos trabalhos experimentais com PCR tem utilizado poucas amostras biológicas, sem considerar o processo de seleção dos indivíduos e tampouco o tamanho da amostra para diferentes genes e em diferentes condições de amostragem.

Em muitas pesquisas, utilizando-se a técnica da PCR, as amostras biológicas são obtidas em mais de uma fase ou mais de um estágio, como por exemplo, o estudo sobre o controle do bicudo-do-algodoeiro, em que Artico et al.(2009) utilizaram amostras de tecidos de diferentes órgãos florais, em diferentes estágios de desenvolvimento e de diferentes plantas de algodão. Mbella et al.(2011), em estudos com culturas geneticamente modificadas, utilizaram folhas coletadas de diferentes plantas para a detecção de genes endógenos de referência através da qPCR caracterizando, neste caso, uma amostragem biológica em duas fases: plantas e folhas das plantas selecionadas.

As pesquisas genéticas na cultura do café têm tido grande expansão e, na maioria dos casos, as amostras são constituídas de folhas, frutos ou outro tecido qualquer, coletado sem diferentes plantas constituindo amostragem em mais de um estágio.

Devido ao emprego de recursos computacionais e da robótica nos laboratórios de análises genéticas, a variação entre amostras é minimizada. Porém, a representatividade das amostras biológicas ainda é pouco estudada.

O objetivo proposto neste trabalho foi de estudar o tamanho ótimo de amostra biológica para a quantificação da expressão gênica utilizando a PCR em tempo real em amostragem hierárquica.

2 REVISÃO DE LITERATURA

2.1 Reação em cadeia da polimerase (PCR)

A reação em cadeia da polimerase (PCR) é utilizada para a amplificação de um fragmento de DNA sem o uso de um organismo vivo. É uma técnica muito comum em laboratórios de pesquisas médicas e biológicas para diversas finalidades, como o sequenciamento de genes, diagnósticos de doenças hereditárias e infecciosas, testes de paternidade, na criação de produtos transgênicos e no estudo de expressão gênica (PADILHA et al., 2011).

2.1.1 Reação em cadeia da polimerase em tempo real (qPCR)

Na PCR em tempo real, as aplicações mais comuns são a análise de expressão gênica e a detecção e quantificação de micro RNA's (agentes reguladores da expressão).

A Figura 1 representa a curva de amplificação da PCR em tempo real. São observadas três fases distintas: a primeira fase (linha de base) corresponde à região onde não há produtos da PCR suficientes para que a fluorescência seja detectada. Na segunda fase, (fase exponencial) a quantidade de produtos da PCR dobra a cada ciclo e na fase do platô, não há mais aumento no número de produtos.

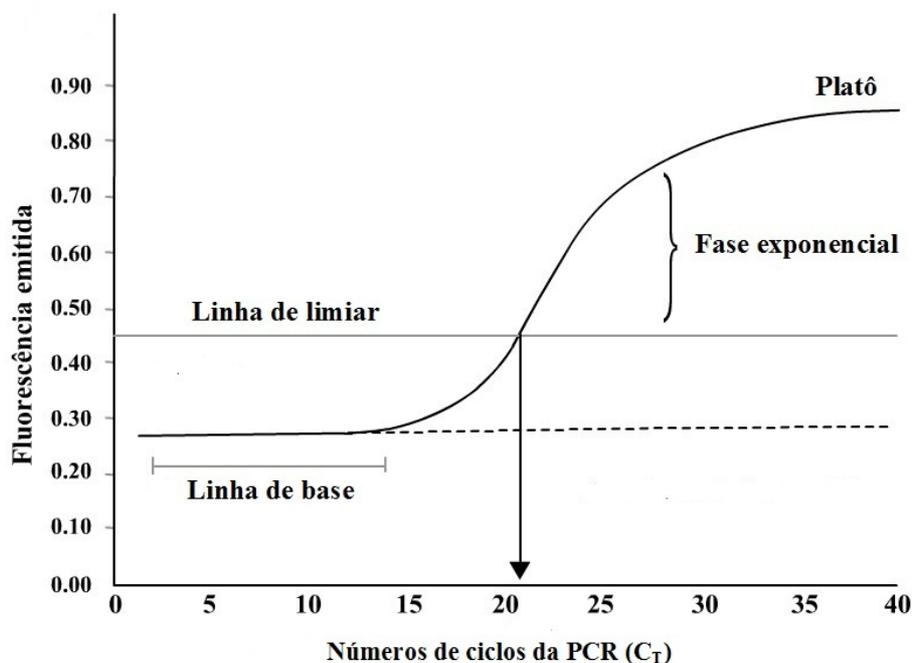


Figura 1 Curva de amplificação da PCR em tempo real

A validação do procedimento é realizada com uma amostra sem DNA, mas com todos os reagentes, representada na Figura 1 pela linha tracejada. A linha de limiar ou linha de C_T é traçada de acordo com o objetivo do estudo. A intersecção com a curva de produtos representa uma leitura (C_T) na qual todas as amostras atingiram uma função exponencial (NOVAIS; ALVES, 2004).

A reação em cadeia da polimerase da transcrição reversa em tempo real (RTq-PCR) é um dos métodos mais precisos no estudo da expressão gênica. É um importante desenvolvimento na técnica de PCR permitindo a detecção precisa e a quantificação do RNA em apenas uma fase. Os dados obtidos, com a técnica, são altamente informativos e, por isso, muito usado no diagnóstico de

doenças virais (BUSTIN; MUELLER, 2005). Apesar de sua popularidade, a RTq-PCR é muito passível de variações (LADEIRA; ISAAC; FERREIRA, 2011).

As principais fontes de variação, que afetam os resultados da PCR em tempo real, são as diferenças dos níveis de expressão gênica entre os organismos, as diferenças devidas à variabilidade na técnica de laboratório (pipetagem, calibração, extração de RNA) e às variações nas amostras biológicas devidas à escolha de tecidos e amostras de células, tipo e local de coleta das amostras, constituição das amostras biológicas e outros (LADEIRA; ISAAC; FERREIRA, 2011).

Os pesquisadores têm seguido as recomendações do guia publicado por Bustin et al. (2009) que contém as informações consideradas mínimas para os experimentos com PCR. Embora os autores tenham dado detalhes da técnica de PCR, do número de repetições a ser utilizado nos laboratórios e de análises estatísticas, não discutem os procedimentos para as amostragens biológicas.

2.1.2 Amostragem na cultura do café para estudos com PCR

O melhoramento em café tem como objetivo principal o desenvolvimento de cultivares mais resistentes a doenças, mais produtivas, mais adaptadas às diversas condições climáticas e com produção de maior qualidade. Como exemplo recente, Gallina et al. (2009) realizaram estudos para identificar genes envolvidos na resistência e defesa do cafeeiro à ferrugem.

Dentre os principais estudos têm se destacado aqueles que tratam do controle da floração do café, uma vez que a maior uniformidade na maturação dos frutos depende de uma floração homogênea.

A uniformidade de maturação dos frutos do cafeeiro está diretamente relacionada com a qualidade final do produto. Diferentes floradas significam estádios de maturação diferentes que podem resultar em um maior gasto com a colheita e queda na qualidade do grão (BUDZINSKI et al., 2005).

O cafeeiro apresenta uma grande heterogeneidade na maturação dos frutos devido ao florescimento sequencial. Geralmente o estágio de maturação é mais uniforme nas cultivares precoces, porém pouco se sabe sobre os fatores que regulam essa característica, Castro et al. (2001). Na tentativa de padronizar a maturação dos frutos de café têm sido feitos diversos estudos com base na síntese de etileno.

Pezzopane et al. (2005) coletaram frutos nas diferentes fases fenológicas do café (chumbinho, expansão, verde, verde cana e cereja) para estudarem o relacionamento de alguns genes na maturação dos frutos de café.

Galvão et al.(2009), com o objetivo de isolar e caracterizar genes de ACC oxidase de café *SPP*, em projetos de transformação genética visando o controle da maturação dos frutos de café, realizaram ensaios de RTq-PCR a partir de RNA total, livre de DNA, extraído de diferentes tecidos e de três estádios distintos de maturação dos frutos. Segundo os autores, a expressão do gene de ACC oxidase em frutos de café fortalece a hipótese de que a inibição desse gene poderá produzir plantas com maturação mais uniforme.

Em um estudo para análise da expressão gênica via RTq-PCR, Budzinski et al.(2005) estudaram o estágio de maturação dos frutos de café coletando frutos em diferentes estádios e outros tecidos.

Para analisar a expressão de genes relacionados às rotas da biossíntese e sinalização do etileno, durante a maturação de frutos de cafeeiro, Lima (2011) realizou estudo com frutos coletados de cultivares precoce e tardia e utilizou a análise da expressão gênica via RTq-PCR.

Observa-se, em quase todos os trabalhos revisados, que o tamanho e a constituição das amostras biológicas variam muito e não se identificou nenhum critério ou referência sobre o assunto. Gallina et al. (2009) selecionaram duas folhas por planta do cafeeiro em um trabalho para identificar genes envolvidos na resistência e defesa do cafeeiro a ferrugem, mas não citam a quantidade de plantas amostradas.

Em um estudo para isolar e caracterizar genes de ACC oxidase de café, visando o controle da maturação dos frutos, Galvão et al. (2009) coletaram diferentes tecidos e de diferentes estádios de maturação via qRT-PCR para a realização do estudo sem citar maiores detalhes do tipo de amostra.

Estudando a expressão de genes relacionados às rotas de biossíntese e sinalização do etileno, durante a maturação de frutos de café, Lima (2011) selecionou seis plantas e 20 frutos, em cada planta, para determinar o percentual de maturação. Desses vinte frutos, tomou de três a cinco frutos que foram macerados e tomados como amostra para a realização da RTq-PCR.

2.2 Fundamentos teóricos

A amostragem em mais de um estágio é muito comum nas pesquisas, podendo variar segundo a população objetivo ou com a finalidade dos trabalhos como, por exemplo, amostragem hierárquica, amostragem hierárquica com frequências pré-fixadas, amostragem estratificada e outras.

Na amostragem em dois estágios ou sub amostragem, designada amostragem hierárquica (GANGULI, 1941), a população é constituída por N_1 unidades primárias e cada unidade primária por N_2 indivíduos. São selecionadas n_1 unidades primárias e, de cada uma delas, selecionados n_2 indivíduos.

Para a amostragem em apenas um estágio, existem várias propostas para a determinação do tamanho ótimo de amostra, como o método empírico de Smith (1938), o método da inspeção visual da curvatura máxima do coeficiente de variação citado por Federer (1955), o método da curvatura máxima do coeficiente de variação de Lessman e Atkins (1963) e recentemente foram propostos novos métodos como o método da função de curvatura do coeficiente de variação e a utilização do ajuste de modelos segmentados como o modelo linear com platô citado por Paranaíba, Ferreira e Morais (2009).

Quando as características da população objetivo exigem que a amostragem seja realizada em estágios, o procedimento, utilizado para a determinação do tamanho ótimo de amostra, depende do conhecimento das variâncias para os diferentes estágios e da fixação da precisão ou do custo total da amostra.

2.2.1 Modelo para amostragem hierárquica em dois estágios

Admitindo-se que os efeitos nos diferentes estágios sejam aditivos, pode-se definir uma observação obtida por amostragem hierárquica em dois estágios pelo modelo:

$$Y_{ij} = \mu + \alpha_i + \beta_{ij}$$

em que Y_{ij} representa a observação realizada no indivíduo j do segundo estágio, na unidade i do primeiro estágio; μ é uma constante inerente a cada observação;

α_i representa o efeito da unidade i do primeiro estágio; β_{ij} representa o efeito do indivíduo j do segundo estágio dentro da unidade i com $i = 1, 2, \dots, n_1$ e $j = 1, 2, \dots, n_2$ para cada i .

2.2.2 Variância da média amostral - modelo aleatório

A média amostral (\bar{Y}) para o modelo de amostragem em dois estágios é dada por:

$$\bar{Y} = \mu + \frac{1}{n_1} \sum_{i=1}^{n_1} \alpha_i + \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \beta_{ij}.$$

Quando a determinação das unidades do primeiro estágio e a seleção dos indivíduos, em cada unidade selecionada, é aleatória, os parâmetros do modelo, exceto μ , têm efeitos aleatórios. Admitindo-se que os parâmetros do modelo tenham médias e covariâncias nulas e variâncias iguais à σ_α^2 e σ_β^2 , a variância da média amostral para uma amostragem em dois estágios é dada por:

$$VAR(\bar{Y}) = \frac{1}{n_1} \sigma_\alpha^2 + \frac{1}{n_1 n_2} \sigma_\beta^2.$$

2.2.3 Componentes de variância - modelo aleatório

O modelo de análise de variância para a amostragem em dois estágios com as esperanças matemáticas dos quadrados médios, considerando o modelo aleatório, é apresentado na Tabela 1.

Tabela 1 Modelo de análise de variância para amostragem hierárquica em dois estágios – modelo aleatório.

Fontes de Variação	GL	QM	E(QM)
Entre unidades	$n_1 - 1$	QM_1	$\sigma_\beta^2 + n_2 \sigma_\alpha^2$
Entre indivíduos dentro de unidades	$n_1(n_2 - 1)$	QM_2	σ_β^2

As estimativas dos componentes de variância (S_α^2 e S_β^2) podem ser obtidas igualando os quadrados médios às suas respectivas E(QM), ou seja,

$$QM_1 = S_\beta^2 + n_2 S_\alpha^2,$$

$$QM_2 = S_\beta^2, \quad e$$

$$S_\beta^2 = QM_2,$$

$$S_\alpha^2 = \frac{QM_1 - QM_2}{n_2}.$$

O estimador da variância da média amostral é:

$$\widehat{VAR}(\bar{Y}) = \frac{1}{n_1} S_\alpha^2 + \frac{1}{n_1 n_2} S_\beta^2.$$

2.2.4 Função de custos para amostragem hierárquica

Os projetos de pesquisa envolvem custos, em maior ou menor escala associado também com o processo de amostragem. Em se tratando de amostragem em duas fases, esses custos podem ser ainda mais elevados, principalmente considerando as pesquisas com expressão gênica onde o processo de escolha do material biológico é mais crítico.

Assumindo que o custo é proporcional ao número de unidades amostradas, nas duas fases, e que o custo por unidade amostrada é conhecido, ou seja, custo de deslocamento, material utilizado e outros, o custo total da amostra pode ser considerado como uma função linear do número de unidades amostradas.

Designando c_1 e c_2 como o custo de cada uma das n_1 unidade e cada um dos n_2 indivíduos amostrados respectivamente, o custo total da amostra c_t para uma amostragem em duas fases é dada por:

$$c_t = c_1 n_1 + c_2 n_1 n_2$$

2.2.5 Tamanho ótimo de amostra

O procedimento estatístico para o dimensionamento ótimo da amostra para a amostragem em dois estágios consiste na solução do sistema de equações constituído pela variância da média amostral e do custo total de amostragem como funções de n_1 e n_2 :

$$VAR(\bar{Y}) = \frac{1}{n_1} \sigma_a^2 + \frac{1}{n_1 n_2} \sigma_b^2,$$

$$c_2 = c_1 n_1 + c_2 n_1 n_2.$$

A amostra ótima é constituída por $n_1 n_2$ indivíduos sendo n_1 unidades primárias, com n_2 indivíduos selecionados em cada uma das unidades primárias. As soluções ótimas são obtidas fixando-se a variância da média amostral ou o custo total da amostra.

Marcuse (1949) apresenta a alocação ótima para n estágios em amostragem hierárquica, amostragem hierárquica com frequências pré-fixadas e amostragem estratificada para populações finitas.

As soluções ótimas para a amostragem hierárquica em dois estágios são apresentadas a seguir. Nas expressões, a variância entre unidades primárias (σ_1^2) é designada por σ_1^2 e a variância entre indivíduos dentro das unidades, (σ_2^2) representada por σ_2^2 .

O número de unidades primárias, fixada a variância da média (v_0) e minimizando o custo total da amostra é:

$$n_1 = \sqrt{\frac{\sigma_1^2}{c_1} \left(\frac{\sum_{i=1}^n \sqrt{c_i \sigma_i^2}}{v_0} \right)}.$$

Se o interesse for fixar o custo total da amostra (c_0) e minimizar a variância da média amostral, o número ótimo de unidades primárias é:

$$n_1 = \sqrt{\frac{\sigma_1^2}{c_1} \left(\frac{c_0}{\sum_{i=1}^n \sqrt{c_i \sigma_i^2}} \right)}.$$

O número ótimo de indivíduos em cada unidade primária não depende de que seja fixada a variância da média ou o custo total da amostra e é dado por:

$$n_2^r = \sqrt{\frac{c_1 \sigma_2^2}{c_2 \sigma_1^2}}.$$

As expressões para o número ótimo de unidades primárias, em função de n_2^r , fixada a variância da média amostral ou o custo total da amostra são, respectivamente:

$$n_1^r = \frac{1}{V_0} \left(\sigma_1^2 + \frac{\sigma_2^2}{n_2^r} \right),$$

e

$$n_1^r = \frac{c_0}{c_1 + c_2 n_2^r}.$$

3 MATERIAL E MÉTODOS

Para esse estudo, foram utilizados os dados de expressão gênica determinados para três genes diferentes em quatro experimentos realizados com a cultura do café (*Coffea arabica*) por Lima (2011).

Foram determinadas as estimativas dos componentes de variância e médias nesses experimentos para serem utilizadas como parâmetros na simulação de dados e para o estudo do tamanho ótimo de amostra.

3.1 Material

Os experimentos de Lima (2011) foram conduzidos na Fazenda do Ministério da Agricultura em Varginha – MG, nos anos de 2008 e 2009. Dois experimentos foram realizados com a cultivar Acauã (*Coffea arabica* L. cv. *Acauã*) e os outros dois com a cultivar Catucaí (*Coffea arabica* L. cv. *Catucaí*).

A cultivar Acauã, proveniente do cruzamento entre Sarchimor e Mundo Novo, apresenta porte baixo, arquitetura compacta, alto vigor, resistência à ferrugem, tolerância ao fitonematoide *Meloidogyne exigua* e à seca. O estágio de maturação é tardio, com frutos de cor vermelho escura e boa produtividade.

A cultivar Catucaí, proveniente do cruzamento entre Icatu e Catuaí, possui porte baixo a médio, arquitetura compacta, bom vigor, resistência à ferrugem e ao *Meloidogyne exigua* e estágio de maturação precoce, com boa produtividade.

O plantio das cultivares ocorreu no ano de 2000, com espaçamento entre linhas de 3,5 m e entre plantas de 1,0 m. A florada principal, no ano de 2007, ocorreu no final do mês de Outubro, sendo que no ano de 2008, não foi possível determinar a ocorrência de uma florada principal.

Em cada experimento foram coletados frutos em quatro períodos diferentes do ano, iniciando-se no mês de março e prosseguindo com as coletas a cada trinta dias. Foram utilizadas três repetições no delineamento inteiramente casualizado.

A coleta dos frutos deu-se no terço médio das plantas, no período da manhã, nos lados ensolarados das plantas. Os frutos coletados foram imediatamente colocados em tubos Falcon de 50 ml e inseridos em nitrogênio líquido, sendo armazenados em freezer a -80°C .

Cada amostra biológica, para a utilização na RTq-PCR, foi constituída por 3 a 5 frutos coletados de 6 plantas distintas.

A extração do RNA foi realizada pelo protocolo descrito por Chang, Puryear e Cairney(1993).

3.2 Métodos

Para esse estudo sobre a amostragem ótima, em pesquisas com expressão gênica, as plantas do café foram consideradas unidades primárias e os frutos das plantas considerados unidades secundárias. Estudou-se a amostragem hierárquica a dois estágios consistindo da seleção aleatória de n_p plantas de café e, em cada planta, seleção aleatória de n_f frutos.

3.2.1 Amostragem hierárquica em dois estágios

Admitindo-se que os efeitos das plantas e dos frutos na expressão gênica são aditivos, pode-se definir uma observação obtida por amostragem hierárquica em dois estágios pelo modelo:

$$Y_{ij} = \mu + \alpha_i + \beta_{ij}$$

com $i = 1, 2, \dots, n_p$ e $j = 1, 2, \dots, n_f$ para cada i em que:

Y_{ij} representa a expressão gênica observada no fruto j na planta i ; μ é uma constante inerente a cada observação; α_i representa o efeito da planta i ; β_{ij} representa o efeito do fruto j da planta i .

3.2.2 Variância da média amostral em dois estágios

A média amostral da expressão gênica (\bar{Y}) para o modelo de amostragem de plantas e frutos é:

$$\bar{Y} = \mu + \frac{1}{n_p} \sum_{i=1}^{n_p} \alpha_i + \frac{1}{n_p n_f} \sum_{i=1}^{n_p} \sum_{j=1}^{n_f} \beta_{ij}$$

Quando a seleção das plantas e a escolha dos frutos, em cada planta, são aleatórias, os parâmetros do modelo, exceto μ , têm efeitos aleatórios. Admitindo-se que os parâmetros do modelo tenham médias e covariâncias nulas e variâncias iguais à σ_p^2 e σ_f^2 , a variância da média amostral para a amostragem de plantas e frutos é dada por:

$$VAR(\bar{Y}) = \frac{1}{n_p} \sigma_p^2 + \frac{1}{n_p n_f} \sigma_f^2$$

3.2.3 Componentes de variância em dois estágios

O modelo de análise de variância para a amostragem de plantas e frutos com as esperanças matemáticas dos quadrados médios, considerando o modelo aleatório, é apresentado na Tabela 2.

Tabela 2 Modelo de análise de variância para amostragem de plantas e frutos – modelo aleatório.

Fontes de Variação	GL	QM	E(QM)
Entre plantas	$n_p - 1$	QM_p	$\sigma_f^2 + n_f \sigma_p^2$
Entre frutos dentro de plantas	$n_p(n_f - 1)$	QM_f	σ_f^2

As estimativas dos componentes de variância (S_p^2 e S_f^2) podem ser obtidas por:

$$QM_p = S_f^2 + n_f S_p^2,$$

$$QM_f = S_f^2,$$

ou

$$S_f^2 = QM_f,$$

$$S_p^2 = \frac{QM_p - QM_f}{n_f}.$$

O estimador da variância da média amostral é:

$$\widehat{VAR}(Y) = \frac{1}{n_p} S_p^2 + \frac{1}{n_p n_f} S_f^2.$$

3.2.4 Função de custos em dois estágios

O custo total de amostragem foi assumido como proporcional ao número de plantas e de frutos. Designando c_p , o custo de chegada e de seleção de cada planta e c_f , o custo de escolha e colheita de cada fruto, o custo total da amostra c_c é dado por:

$$c_c = c_p n_p + c_f n_p n_f.$$

3.2.5 Tamanho ótimo de amostra em dois estágios

Neste estudo, procurou-se estudar o tamanho ótimo de amostra, fixando a variância da média amostral (v_0) e minimizando o custo de amostragem (c_c).

A amostra ótima é constituída por $n_p' n_f'$ frutos sendo n_p' plantas e n_f' frutos selecionados em cada planta. Os números ótimos de plantas e de frutos, por planta, são dados por:

$$n_p' = \frac{1}{v_0} \left(\sigma_p^2 + \frac{\sigma_f^2}{n_f'} \right),$$

e

$$n_f' = \sqrt{\frac{c_p \sigma_f^2}{c_f \sigma_p^2}}.$$

O critério adotado para fixar a precisão da média amostral foi tomar o desvio padrão amostral como um percentual (d) da média amostral (\bar{Y}). Assim:

$$v_a = \left(\frac{d\bar{Y}}{100} \right)^2.$$

em que d é uma percentagem ($d > 0$) da média amostral (\bar{Y}). Dessa forma, o coeficiente de variação amostral é dado por $d\%$.

Na prática, como não se conhecem as variâncias σ_x^2 e σ_f^2 , utilizam-se estimativas dessas variâncias obtidas de pesquisas já realizadas. Os dados dos experimentos utilizados não possibilitaram a obtenção dessas estimativas. A Tabela 3 apresenta os componentes de variância para os experimentos utilizados.

Tabela 3 Modelo de análise de variância para os experimentos de Lima, (2011).

Fontes de Variação	GL	QM	E(QM)
Entre Períodos do ano	3	$QM_{\text{Períodos}}$	$\sigma^2 + 3 \sum_{i=1}^4 (t_i - \bar{t})^2$
Resíduo	8	$QM_{\text{Resíduo}}$	σ^2

t_i = efeito do período i \bar{t} = média dos efeitos dos períodos

A variância residual, nesses experimentos, foi devida à variabilidade entre as plantas, entre os diferentes estádios de maturação, entre os frutos e outras variáveis. Assim sendo, foram tomadas proporções entre a variância entre plantas e a variância entre frutos dentro de plantas limitadas pela estimativa da variância residual de cada experimento, considerada como o valor máximo para a variância entre frutos dentro de plantas.

3.2.6 Proporções entre custos de amostragem e entre variâncias

Para a amostragem em duas fases, em que na primeira fase, selecionam-se plantas e, na segunda fase, selecionam-se frutos dentro de plantas, pode ser verificado que o número ótimo de frutos depende apenas das relações de custos e de variâncias independentemente da precisão requerida para a média amostral:

$$n'_f = \sqrt{\frac{c_p \sigma_f^2}{c_f \sigma_p^2}}$$

Designando P_c , à razão entre os custos unitários de plantas (c_p) e de frutos (c_f), e P_v à razão entre a variância entre plantas (σ_p^2), e entre frutos dentro de plantas (σ_f^2), o número ótimo de frutos por planta pode ser determinado por:

$$n'_f = \sqrt{\frac{P_c}{P_v}}$$

com

$$R_c = \frac{c_p}{c_f},$$

e

$$R_v = \frac{\sigma_p^2}{\sigma_f^2}.$$

A expressão para o número ótimo de plantas na amostra

$$n_p^* = \frac{1}{v_0} \left[\sigma_p^2 + \frac{\sigma_f^2}{n_f} \right],$$

pode ser reescrita em função da proporção de variâncias (R_v) como:

$$n_p^* = \frac{1}{v_0} \left[R_v \sigma_f^2 + \frac{\sigma_f^2}{n_f} \right].$$

Substituindo n_f^* pela expressão em função das proporções de custos e de variância, a expressão para o número ótimo de plantas será:

$$n_p^* = \frac{\sigma_f^2}{v_0} \left[R_v + \sqrt{\frac{R_v}{R_c}} \right].$$

Essa expressão demonstra que o número ótimo de plantas na amostra (n_1') depende da variância fixada para a média amostral (v_0) e do valor da variância entre frutos, além das proporções de custos e de variâncias.

3.2.7 Simulação de dados

As variâncias residuais e as médias para a expressão gênica dos quatro experimentos, para cada um dos três genes, foram utilizadas para as simulações de dados de expressão gênica de frutos e plantas de café, segundo o modelo estatístico:

$$Y_{ij} = \mu + \alpha_i + \beta_{ij}$$

em que Y_{ij} é valor da expressão gênica referente ao fruto j na planta i ; μ é uma constante inerente a toda observação; α_i representa o efeito aleatório da planta i ; β_{ij} representa efeito aleatório do fruto j na planta i .

Admitiu-se que os parâmetros do modelo têm médias e covariâncias nulas e variâncias iguais à σ_p^2 e σ_f^2 e a técnica utilizada foi a de Monte Carlo.

Para cada experimento, foram realizadas 5000 simulações e obtidas as estimativas das variâncias entre plantas e entre frutos dentro de plantas para a determinação do tamanho ótimo de amostra para uma determinada precisão da média amostral. Os dados foram simulados com auxílio do programa R versão 2.13.1.

Para o cálculo do número ótimo de plantas (n_1'), além da proporção de custos (P_c) e da proporção de variâncias (P_v), necessitou-se também da

variância entre frutos dentro de plantas (σ_f^2) e da fixação da variância para a média amostral (v_0) ou o custo total da amostra desejado (C_0).

Para este estudo, fixou-se a variância da média amostral (v_0) como uma função da média da amostra, dada por:

$$v_0 = \left(\frac{d\bar{Y}}{100} \right)^2.$$

O número ótimo de plantas de café na amostra fixando a variância da média (v_0) depende dos valores das variâncias entre plantas e entre frutos dentro de plantas, como apresentado:

$$n_1^* = \frac{1}{v_0} \left[\sigma_p^2 + \frac{\sigma_f^2}{n_2} \right].$$

Usando a relação de custos (R_c) e a relação de variâncias (R_v), o número ótimo de plantas n_1^* pode ser determinado por:

$$n_1^* = \frac{\sigma_f^2}{v_0} \left[R_v + \sqrt{\frac{R_v}{R_c}} \right].$$

4 RESULTADOS E DISCUSSÃO

Na apresentação dos resultados optou-se primeiramente pela discussão do número ótimo de frutos por planta, porque esses resultados não dependem de que seja fixada a precisão requerida para a média amostral ou o custo total da amostra. Em seguida, são apresentados os resultados para o número ótimo de plantas com base nos resultados obtidos por Lima (2011) e nas simulações de dados.

4.1 Número ótimo de frutos por planta

O número ótimo de frutos por planta independe de que seja fixada a variância da média amostral ou o custo total da amostra. O número ótimo de frutos por planta é dado por:

$$n_p^f = \sqrt{\frac{c_p \sigma_f^2}{c_f \sigma_p^2}}$$

onde:

n_p^f é o número ótimo de frutos por planta, fixada a variância da média amostral ou fixado o custo total da amostra; c_p é o custo de amostragem de uma planta; c_f é o custo de amostragem de um fruto; σ_p^2 a variância entre plantas; σ_f^2 é a a variância entre frutos dentro de plantas.

Designando $R_c = \frac{c_p}{c_f}$ (proporção entre o custo de seleção e acesso a uma planta pelo custo de colheita de um fruto), e $R_v = \frac{\sigma_p^2}{\sigma_f^2}$ (proporção entre a variância entre plantas e a variância entre frutos dentro de plantas), o número ótimo de frutos de cafeeiro na amostra pode ser expresso por:

$$n_s^* = \sqrt{\frac{R_c}{R_v}} \quad (1)$$

Pela expressão (1), verifica-se que o número ótimo de frutos na amostra diminui com o aumento da proporção de variância, isto é, o número de frutos por planta diminui, quando a variância entre plantas aumenta em relação à variância entre frutos. Verifica-se também que o número de frutos aumenta com o aumento da proporção de custos, ou seja, quanto maior o custo de amostragem de planta, em relação ao custo de amostragem de frutos, maior o número ótimo de frutos na amostra.

A Tabela 4 apresenta os números ótimos de frutos de cafeeiro para diferentes proporções da variância entre plantas e variância entre frutos dentro de plantas, considerando a proporção de custos igual a 1:1. Sem considerar o número de plantas e tomando $R_c = 1$, para obter um máximo de 20 frutos é necessário ter-se uma $R_v \geq \frac{1}{400}$. Isso significa que a variância entre frutos dentro de plantas poderá ser no máximo 400 vezes maior que a variância entre plantas.

Tabela 4 Números ótimos de frutos por planta para diferentes proporções de variância entre plantas e variância entre frutos ($R_v = \frac{\sigma_p^2}{\sigma_f^2}$). Proporção de custos ($\frac{C_p}{C_f}$) de 1:1.

P_v	Frutos	P_v	Frutos	P_v	Frutos	P_v	Frutos	P_v	Frutos
1/1	1,00	2/1	0,71	5/1	0,45	10/1	0,32	100/1	0,10
1/2	1,41	2/2	1,00	5/2	0,63	10/2	0,45	100/2	0,14
1/3	1,73	2/3	1,22	5/3	0,77	10/3	0,55	100/3	0,17
1/4	2,00	2/4	1,41	5/4	0,89	10/4	0,63	100/4	0,20
1/5	2,24	2/5	1,58	5/5	1,00	10/5	0,71	100/5	0,22
1/10	3,16	2/10	2,24	5/10	1,41	10/10	1,00	100/10	0,32
1/20	4,47	2/20	3,16	5/20	2,00	10/20	1,41	100/20	0,45
1/40	6,32	2/40	4,47	5/40	2,83	10/40	2,00	100/40	0,63
1/50	7,07	2/50	5,00	5/50	3,16	10/50	2,24	100/50	0,71
1/100	10,00	2/100	7,07	5/100	4,47	10/100	3,16	100/100	1,00
1/200	14,14	2/200	10,00	5/200	6,32	10/200	4,47	100/200	1,41
1/400	20,00	2/400	14,14	5/400	8,94	10/400	6,32	100/400	2,00
1/500	22,36	2/500	15,81	5/500	10,00	10/500	7,07	100/500	2,24
1/800	28,28	2/800	20,00	5/800	12,65	10/800	8,94	100/800	2,83
1/1000	31,62	2/1000	22,36	5/1000	14,14	10/1000	10,00	100/1000	3,16

Na maioria dos trabalhos revisados, foram utilizados 5 frutos por amostra exigindo uma $R_v \geq \frac{1}{20}$, indicando que a variância entre frutos dentro de plantas poderá ser no máximo 20 vezes maior que a variância entre plantas.

Na prática, geralmente o custo de seleção de uma planta é maior que o custo da coleta de um fruto. As Tabelas 5 e 6 apresentam os números ótimos de frutos por planta para relações de custos mais condizentes com o que ocorre na prática.

Tabela 5 Números ótimos de frutos por planta para diferentes proporções de variância entre plantas e variância entre frutos ($R_v = \frac{\sigma_p^2}{\sigma_f^2}$). Proporção de custos ($\frac{C_p}{C_f}$) de 10:1.

P_n	Frutos	P_n	Frutos	P_n	Frutos	P_n	Frutos	P_n	Frutos
1/1	3,16	2/1	2,24	5/1	1,41	10/1	1,00	100/1	0,10
1/2	4,47	2/2	3,16	5/2	2,00	10/2	1,41	100/2	0,14
1/3	5,48	2/3	3,87	5/3	2,45	10/3	1,73	100/3	0,17
1/4	6,32	2/4	4,47	5/4	2,83	10/4	2,00	100/4	0,20
1/5	7,07	2/5	5,00	5/5	3,16	10/5	2,24	100/5	0,22
1/10	10,00	2/10	7,07	5/10	4,47	10/10	3,16	100/10	0,32
1/20	14,14	2/20	10,00	5/20	6,32	10/20	4,47	100/20	0,45
1/40	20,00	2/40	14,14	5/40	8,94	10/40	6,32	100/40	0,63
1/50	22,36	2/50	15,81	5/50	10,00	10/50	7,07	100/50	0,71
1/100	31,62	2/100	22,36	5/100	14,14	10/100	10,00	100/100	1,00
1/200	44,72	2/200	31,62	5/200	20,00	10/200	14,14	100/200	1,41
1/400	63,25	2/400	44,72	5/400	28,28	10/400	20,00	100/400	2,00
1/500	70,71	2/500	50,00	5/500	31,62	10/500	22,36	100/500	2,24
1/800	89,44	2/800	63,25	5/800	40,00	10/800	28,28	100/800	2,83
1/1000	100,00	2/1000	70,71	5/1000	44,72	10/1000	31,62	100/1000	3,16

Assumindo P_c igual a 10, para obter um máximo de 20 frutos na amostra é necessário ter-se uma $R_v \geq \frac{1}{40}$ e, para um máximo de 5 frutos é necessário, ter-se uma $R_v \geq \frac{1}{2}$ significando que a variância entre frutos dentro de plantas deverá ser no máximo 40 vezes maior que a variância entre plantas, no primeiro caso, e 2 vezes maior, no segundo caso (Tabela 5).

Tabela 6 Números ótimos de frutos por planta para diferentes proporções de variância entre plantas e variância entre frutos ($R_v = \frac{\sigma_p^2}{\sigma_f^2}$). Proporção de custos ($\frac{C_p}{C_f}$) de 100:1.

P_v	Frutos	P_v	Frutos	P_v	Frutos	P_v	Frutos	P_v	Frutos
1/1	10,00	2/1	7,07	5/1	4,47	10/1	3,16	100/1	1,00
1/2	14,14	2/2	10,00	5/2	6,32	10/2	4,47	100/2	1,41
1/3	17,32	2/3	12,25	5/3	7,75	10/3	5,48	100/3	1,73
1/4	20,00	2/4	14,14	5/4	8,94	10/4	6,32	100/4	2,00
1/5	22,36	2/5	15,81	5/5	10,00	10/5	7,07	100/5	2,24
1/10	31,62	2/10	22,36	5/10	14,14	10/10	10,00	100/10	3,16
1/20	44,72	2/20	31,62	5/20	20,00	10/20	14,14	100/20	4,47
1/40	63,25	2/40	44,72	5/40	28,28	10/40	20,00	100/40	6,32
1/50	70,71	2/50	50,00	5/50	31,62	10/50	22,36	100/50	7,07
1/100	100,00	2/100	70,71	5/100	44,72	10/100	31,62	100/100	10,00
1/200	141,42	2/200	100,00	5/200	63,25	10/200	44,72	100/200	14,14
1/400	200,00	2/400	141,42	5/400	89,44	10/400	63,25	100/400	20,00

1/500	223,61	2/500	158,11	5/500	100,00	10/500	70,71	100/500	22,36
1/800	282,84	2/800	200,00	5/800	126,49	10/800	89,44	100/800	28,28
1/1000	316,23	2/1000	223,61	5/1000	141,42	10/1000	100,00	100/1000	31,62

Como pode ser observado, na Tabela 6, quando a proporção de custos é igual a 100, para um máximo de 20 frutos na amostra é necessário ter-se uma

$R_c \approx \frac{1}{4}$, isso significa que a variância entre frutos poderá ser, no máximo, 4 vezes a variância entre plantas.

Para um máximo de 5 frutos na amostra, é necessário ter-se uma

$R_c \approx \frac{5}{1}$, isso significa que a variância entre plantas poderá ser, no máximo, 5 vezes a variância entre frutos.

Verifica-se que quando o custo da amostragem de uma planta aumenta em relação ao custo de amostragem de um fruto, para uma mesma proporção de variâncias, a amostra deverá conter mais frutos.

4.2 Número ótimo de plantas na amostra

Nesse estudo, optou-se por fixar a variância da média amostral (v_0) como uma função da média da amostra, dada por:

$$v_0 = \left(\frac{d\bar{Y}}{100} \right)^2 .$$

O número ótimo de plantas de café na amostra fixada a variância da média amostral (v_0):

$$n_1^i = \frac{1}{v_0} \left[\sigma_p^2 + \frac{\sigma_f^2}{n_2} \right]$$

pode ser determinado usando a relação de custos (P_c) e a relação de variâncias (P_v) como:

$$n_1^i = \frac{\sigma_f^2}{v_0} \left[R_v + \sqrt{\frac{R_v}{R_c}} \right] \quad (2)$$

De acordo com a expressão (2), para o cálculo do número ótimo de plantas (n_1^i), além da proporção de custos (P_c), da proporção de variâncias (P_v) e da variância que se deseja fixar para a média amostral (v_0), necessita-se também do conhecimento da variância entre frutos dentro de plantas (σ_f^2).

Para o estudo do número ótimo de plantas e de frutos, foram utilizados os componentes de variância obtidos nas análises estatísticas para a expressão dos genes ERF, ETR e ACO em frutos de cafeeiro (Lima, 2011). A Tabela 7 apresenta as estimativas dos componentes de variância obtidas.

Verifica-se uma grande variação entre os resultados das médias e das variâncias para diferentes épocas de amostragem mesmo considerando um único gene, uma mesma cultivar, mesmos procedimentos de coleta e de técnicas de laboratório.

Tabela 7 Estimativas de médias e variâncias residuais para a expressão gênica de três genes em duas cultivares de café e dois anos agrícolas.

Cultivar	Gene	Ano	Média	Variância
Acauã	ERF	2008	1,35	0,215
Acauã	ERF	2009	1,67	0,137
Acauã	ETR	2008	3,87	8,650
Acauã	ETR	2009	5,03	3,810
Acauã	ACO	2008	23,75	123,880
Acauã	ACO	2009	7,58	11,320
Catucai	ERF	2008	1,40	0,180
Catucai	ERF	2009	1,95	0,811
Catucai	ETR	2008	2,20	17,410
Catucai	ETR	2009	13,56	51,880
Catucai	ACO	2008	1,33	1,630
Catucai	ACO	2009	10,95	8,150

Considerando que os dados de Lima (2011) foram obtidos de amostras de frutos e plantas de cafeeiro para cada tratamento, a variância residual engloba a variabilidade entre frutos dentro da mesma planta e a variabilidade entre diferentes plantas, além de outras fontes de variação. Assim, na impossibilidade de se estimar a variância entre plantas e a variância entre frutos dentro de plantas separadamente, os quadrados médios residuais das análises de variância foram utilizados como valores máximos para as variâncias entre frutos dentro de plantas (σ_f^2) na obtenção dos valores ótimos do número de plantas e de frutos nas amostras.

Através de diferentes proporções simuladas para variâncias entre plantas e entre frutos dentro de plantas (R_w), foram obtidos os valores para as variâncias entre plantas (σ_p^2).

Os números ótimos de plantas e de frutos por planta obtidos considerando os resultados experimentais do gene ETR, na cultivar Acauã, nos anos agrícolas de 2008 e 2009, são apresentados nas Tabelas 8 e 9, respectivamente.

Tabela 8 Valores de n_1^f e n_2^f com diferentes proporções de variâncias e desvio padrão amostral fixado em 10% da média. Referência: gene ETR, cultivar Acauã ($\mu = 3,87$, $\sigma_f^2 = 8,65$), 2008.

Proporção de Custos	1:1			10:1			
	P_v	Plantas	Frutos por planta	Frutos na amostra	Plantas	Frutos por planta	Frutos na amostra
1/100000		0,18	316,2	58	0,06	1000,0	58
1/10000		0,58	100,0	58	0,19	316,2	60
1/1000		1,88	31,62	60	0,64	100,0	64
1/400		3,03	20,0	61	1,06	63,2	67
1/200		4,37	14,1	62	1,58	44,7	71
1/100		6,35	10,0	64	2,40	31,6	76
1/50		9,32	7,0	66	3,74	22,3	84
1/40		10,58	6,3	67	4,33	20,0	87
1/25		13,86	5,0	70	5,96	15,8	94
1/20		15,80	4,4	71	6,97	14,1	99
1/10		24,04	3,1	76	11,55	10,0	116

1/5	37,38	2,2	84	19,72	7,0	139
1/4	43,32	2,0	86	23,57	6,3	149
1/3	52,24	1,7	91	29,55	5,5	163
1/2	69,72	1,4	99	41,79	4,4	187
1/1	115,51	1,0	116	76,02	3,1	240
2/1	197,19	0,7	139	141,34	2,2	316
4/1	346,53	0,5	173	267,55	1,5	423
5/1	417,92	0,4	187	329,62	1,4	466
10/1	760,20	0,3	240	635,31	1,0	635

Verifica-se na Tabela 8 que, mesmo para proporções extremamente grandes, entre as variâncias como um para cem mil, o número total de frutos na amostra é sempre maior que maior que 50, mesmo para apenas uma planta na amostra ótima. Como era de se esperar, para a proporção de custos maior, a amostra ótima deverá conter mais frutos.

Na Tabela 9, os números ótimos de plantas e frutos por planta foram determinados para as mesmas condições da Tabela 8, com mudanças apenas nos valores da média e da variância entre frutos.

Tabela 9 Valores de n_1^* e n_2^* com diferentes proporções de variâncias e desvio padrão amostral fixado em 10% da média. Referência: gene ETR, cultivar Acauã ($\mu = 5,30$, $\sigma^2 = 3,81$), 2009.

Proporção de Custos	1:1			10:1		
	Plantas	Frutos por planta	Frutos na amostra	Plantas	Frutos por planta	Frutos na amostra
1/100000	0,0	316,2	15	0,0	1000,0	15
1/10000	0,2	100,0	15	0,0	316,2	16
1/1000	0,5	31,6	16	0,2	100,0	17
1/400	0,8	20,0	16	0,3	63,2	17
1/200	1,1	14,1	16	0,4	44,7	18
1/100	1,7	10,0	17	0,6	31,6	20
1/50	2,4	7,1	17	1,0	22,4	22
1/40	2,8	6,3	17	1,1	20,0	23
1/25	3,6	5,0	18	1,6	15,8	25
1/20	4,1	4,5	18	1,8	14,1	26
1/10	6,3	3,2	20	3,0	10,0	30

1/5	9,7	2,2	22	5,1	7,1	36
1/4	11,3	2,0	23	6,1	6,3	39
1/3	13,6	1,7	24	7,7	5,5	42
1/2	18,2	1,4	26	10,9	4,5	49
1/1	30,1	1,0	30	19,8	3,2	63
2/1	51,4	0,7	36	36,9	2,2	82
4/1	90,4	0,5	45	69,8	1,6	110
5/1	109,0	0,4	49	85,9	1,4	122
10/1	198,2	0,3	63	165,6	1,0	166

Observa-se que amostras com um total de 20 frutos ou menos são conseguidas para diferentes proporções de variâncias entre plantas e entre frutos.

Para F_c igual a 1, proporções menores que $\frac{1}{10}$ e para F_c igual a 10, menores que $\frac{1}{100}$.

Tabela 10 Valores de n'_1 e n'_2 com diferentes proporções de variâncias e desvio padrão amostral fixado em 10% da média. Referência: gene ERF, cultivar Catucaí ($\mu = 1,40$, $\sigma^2 = 0,180$), 2008.

Proporção de Custos	1:1			10:1		
	Plantas	Frutos por planta	Frutos na amostra	Plantas	Frutos por planta	Frutos na amostra
F_v						
1/1000000	0,0	1000,0	9	0,0	3162,3	9
1/100000	0,0	316,2	9	0,0	1000,0	9
1/10000	0,1	100,0	9	0,0	316,2	9
1/1000	0,3	31,6	9	0,1	100,0	10
1/400	0,5	20,0	10	0,2	63,2	11
1/200	0,7	14,1	10	0,3	44,7	11
1/100	1,0	10,0	10	0,4	31,6	12
1/50	1,5	7,1	10	0,6	22,4	13

1/40	1,7	6,3	11	0,7	20,0	14
1/25	2,2	5,0	11	0,9	15,8	15
1/20	2,5	4,5	11	1,1	14,1	16
1/10	3,8	3,2	12	1,8	10,0	18
1/5	5,9	2,2	13	3,1	7,1	22
1/4	6,9	2,0	14	3,7	6,3	24
1/3	8,3	1,7	14	4,7	5,5	26
1/2	11,1	1,4	16	6,6	4,5	30
1/1	18,4	1,0	18	12,1	3,2	38
2/1	31,4	0,7	22	22,5	2,2	50
4/1	55,1	0,5	28	42,5	1,6	67
5/1	66,5	0,4	30	52,4	1,4	74
10/1	120,9	0,3	38	101,0	1,0	101

Resultados semelhantes foram observados para outros valores de média e da variância entre frutos obtidos dos experimentos com os outros genes (ERF e ACO) e com a outra cultivar. As Tabelas 10 e 11 apresentam os tamanhos ótimos de amostra para o gene ERF e cultivar Catucaí, nos anos agrícolas de 2008 e 2009, respectivamente.

Tabela 11 Valores de n_1' e n_2' com diferentes proporções de variâncias e desvio padrão amostral fixado em 10% da média. Referência: gene ERF, cultivar Catucaí ($\mu = 1,95$, $\sigma^2 = 0,811$), 2009.

Proporção de Custos	1:1			10:1		
	Plantas	Frutos por planta	Frutos na amostra	Plantas	Frutos por planta	Frutos na amostra
$P_{3\%}$						
1/1000000	0,0	1000,0	21	0,0	3162,3	21
1/100000	0,1	316,2	21	0,0	1000,0	22
1/10000	0,2	100,0	22	0,1	316,2	22
1/1000	0,7	31,6	22	0,2	100,0	23
1/400	1,1	20,0	22	0,4	63,2	25
1/200	1,6	14,1	23	0,6	44,7	26
1/100	2,3	10,0	23	0,9	31,6	28
1/50	3,4	7,1	24	1,4	22,4	31
1/40	3,9	6,3	25	1,6	20,0	32

1/25	5,1	5,0	26	2,2	15,8	35
1/20	5,8	4,5	26	2,6	14,1	36
1/10	8,9	3,2	28	4,3	10,0	43
1/5	13,8	2,2	31	7,3	7,1	51
1/4	16,0	2,0	32	8,7	6,3	55
1/3	19,3	1,7	34	10,9	5,5	60
1/2	25,7	1,4	36	15,4	4,5	69
1/1	42,7	1,0	43	28,1	3,2	89
2/1	72,8	0,7	51	52,2	2,2	117
4/1	128,0	0,5	64	98,8	1,6	156
5/1	154,3	0,4	69	121,7	1,4	172
10/1	280,7	0,3	89	234,6	1,0	235

Ressalta-se que, com o aumento da relação de custos, o número ótimo de plantas diminui e o número ótimo de frutos por planta aumenta, mas, mesmo assim, o total de frutos na amostra aumenta (Tabelas 8 a 11).

Para o gene ERF (Tabela 10), amostras com um máximo de 20 frutos são conseguidas para proporções entre as variâncias entre plantas e entre frutos menores que $\frac{1}{1}$ quando $P_G = 1$ e menores que $\frac{1}{10}$ para uma $P_G = 10$, em 2008, enquanto que no ano 2009, para qualquer relação de custos, não se observaram amostras com menos de 20 frutos no total, para todas as proporções de variâncias estudadas (Tabela 11).

Tais resultados eram esperados devidos principalmente às variações nas amostras biológicas já que a expressão gênica é muito sensível a todo tipo de estímulo externo. A Tabela 12 apresenta os valores obtidos para os coeficientes de variação obtidos nos experimentos de Lima (2011).

Observa-se que as relações entre médias e variâncias dos experimentos variaram de 22 a 190%. Dos doze experimentos realizados, a maioria apresentou coeficiente de variação entre 30 e 50% e, mesmo considerando o mesmo gene, os coeficientes de variação diferiram bastante entre si.

Tabela 12 Estimativas das médias, das variâncias residuais e dos coeficientes de variação para a expressão dos genes ERF, ETF e ACO em frutos de duas cultivares de café e dois anos agrícolas.

Cultivar	Gene	Ano	Variância	Média	CV (%)
Acauã	ACO	2008	123,880	23,750	46,86
Acauã	ACO	2009	11,320	7,580	44,39
Catucaí	ACO	2008	1,630	1,330	95,99
Catucaí	ACO	2009	8,150	10,950	26,07
Acauã	ERF	2008	0,215	1,350	34,35
Acauã	ERF	2009	0,137	1,670	22,16
Catucaí	ERF	2008	0,180	1,400	30,30
Catucaí	ERF	2009	0,811	1,950	46,18
Acauã	ETR	2008	8,650	3,870	76,00
Acauã	ETR	2009	3,810	5,030	38,81
Catucaí	ETR	2008	17,410	2,200	189,66
Catucaí	ETR	2009	51,880	13,560	53,12

CV = coeficiente de variação

Como o tamanho de amostra biológica para estudos com PCR em café tem sido de, no máximo 20 frutos, independentemente dos genes envolvidos e de outras condições experimentais, procurou-se verificar as condições necessárias para uma amostra ótima com 20 frutos, independentemente do número de plantas amostradas e utilizando os dados de Lima (2011) como referência.

Tomando-se o valor da variância entre frutos dentro de plantas como um percentual $p(p > 0)$ da média amostral de forma semelhante ao que foi proposto para a variância fixada para a média amostral, de tal forma que:

$$\sigma_f^2 = \left(\frac{p\bar{Y}}{100} \right)^2$$

sendo σ_f^2 uma nova variância denominada de variância considerada a equação (2) pode ser reescrita como:

$$n_1 = \frac{p^2}{d^2} \left[R_v + \sqrt{\frac{R_v}{R_c}} \right]$$

assim, o número ótimo de frutos na amostra ($n_1 n_2$) será dado por:

$$n_1 n_2 = \frac{p^2}{d^2} (1 + \sqrt{R_c R_v}) \quad (3)$$

O coeficiente de variação obtido para o gene ETR, cultivar Acauã, no experimento de 2008, foi igual a 76% com estimativa da variância residual igual a 8,65 (Tabela 12). Para esses valores, não se conseguiu dimensionar uma amostra ótima com 20 frutos ou menos (Tabela 8).

Usando a expressão (3), obteve-se o valor igual a 43% de p necessário para a obtenção de uma amostra ótima com 20 frutos, considerando uma relação de custos igual a um, fixando o desvio padrão amostral em 10% da média e usando uma proporção de variâncias $R_v = \frac{1}{100}$.

A Tabela 13 apresenta os valores obtidos de amostras ótimas para diferentes proporções de variâncias e simulando uma relação entre média e variância de 43% nesse experimento.

Tabela 13 Valores de n_1 e n_2 com diferentes proporções de variâncias e desvio padrão amostral fixado em 10% da média. Referência: gene ETR, cultivar Acauã ($\mu = 3,87$), 2008. Variância entre frutos considerada igual a 2,77.

Proporção de Custos	1:1			10:1			
	P_v	Plantas	Frutos por planta	Frutos na amostra	Plantas	Frutos por planta	Frutos na amostra
1/100000		0,1	316,2	19	0,0	1000,0	19
1/10000		0,2	100,0	19	0,1	316,2	19
1/1000		0,6	31,6	19	0,2	100,0	20
1/400		1,0	20,0	19	0,3	63,2	21
1/200		1,4	14,1	20	0,5	44,7	23
1/100		2,0	10,0	20	0,8	31,6	24
1/50		3,0	7,1	21	1,2	22,4	27
1/40		3,4	6,3	21	1,4	20,0	28
1/25		4,4	5,0	22	1,9	15,8	30
1/20		5,1	4,5	23	2,2	14,1	32
1/10		7,7	3,2	24	3,7	10,0	37
1/5		12,0	2,2	27	6,3	7,1	45
1/4		13,9	2,0	28	7,5	6,3	48
1/3		16,7	1,7	29	9,5	5,5	52
1/2		22,3	1,4	32	13,4	4,5	60
1/1		37,0	1,0	37	24,3	3,2	77
2/1		63,1	0,7	45	45,2	2,2	101
4/1		110,9	0,5	55	85,7	1,6	135
5/1		133,8	0,4	60	105,5	1,4	149
10/1		243,4	0,3	77	203,4	1,0	203

Nessas condições, para $P_v = \frac{1}{100}$, a variância entre plantas deveria ser igual a 0,0277 e a variância entre frutos igual a 2,77 bem menor que o valor encontrado no experimento (8,65).

Resultados semelhantes foram simulados, utilizando a variância e a média do experimento de 2008 para o gene ACO com a cultivar Catucaí e mantidas as mesmas pressuposições anteriores. Nesse caso, a variância entre frutos foi reduzida para de 1,63 para 0,3271 (CV reduzido de 96 para 43%).

Tabela 14 Valores de n'_1 e n'_2 com diferentes proporções de variâncias e desvio padrão amostral fixado em 10% da média. Referência: gene ACO,

cultivar Catucaí($\mu = 1,33$), 2008. Variância entre frutos considerada igual a 0,3271.

Proporção de Custos	1:1			10:1		
	Plantas	Frutos por planta	Frutos na amostra	Plantas	Frutos por planta	Frutos na amostra
1/100000	0,1	316,2	19	0,0	1000,0	19
1/10000	0,2	100,0	19	0,1	316,2	19
1/1000	0,6	31,6	19	0,2	100,0	20
1/400	1,0	20,0	19	0,3	63,2	21
1/200	1,4	14,1	20	0,5	44,7	23
1/100	2,0	10,0	20	0,8	31,6	24
1/50	3,0	7,1	21	1,2	22,4	27
1/40	3,4	6,3	21	1,4	20,0	28
1/25	4,4	5,0	22	1,9	15,8	30
1/20	5,1	4,5	23	2,2	14,1	32
1/10	7,7	3,2	24	3,7	10,0	37
1/5	12,0	2,2	27	6,3	7,1	45
1/4	13,9	2,0	28	7,5	6,3	48
1/3	16,7	1,7	29	9,5	5,5	52
1/2	22,3	1,4	32	13,4	4,5	60
1/1	37,0	1,0	37	24,3	3,2	77
2/1	63,1	0,7	45	45,2	2,2	101
4/1	110,9	0,5	55	85,7	1,6	135
5/1	133,8	0,4	60	105,5	1,4	149
10/1	243,4	0,3	77	203,4	1,0	203

Esses resultados mostram que a precisão do experimento, tomado como referência para o dimensionamento ótimo de amostra, deve ser a maior possível para permitir boas estimativas dos componentes de variância e médias amostrais e, conseqüentemente, garantindo amostras representativas.

Como pôde ser verificado, o tamanho da amostra utilizado pode estar subestimado em função principalmente da relação entre a variabilidade entre frutos dentro de plantas e a média. Isso irá afetar a precisão do experimento, diminuindo a confiabilidade nos resultados obtidos.

Como provavelmente a variância entre frutos, nesses experimentos, tenha sido menor que a variância residual tomada como referência, foram feitos estudos através da simulação de dados, utilizando as mesmas condições desses

experimentos, mas supondo uma variância entre frutos menor, correspondendo a uma parte da estimativa da variância residual correspondente.

Tomando como exemplo o experimento com o gene ERF, cultivar Acauã, no ano de 2009 que apresentou boa precisão (CV igual a 22,16%), para proporções entre variâncias de 1/1000 a 1/1, o número de frutos na amostra ótima foi menor ou igual a 20 (Tabela 15). Como provavelmente a variância entre frutos foi menor que a variância residual obtida nesse experimento, o número de frutos para uma amostragem ótima seria menor que 20, se fossem utilizadas as condições desse experimento como referência para a alocação ótima de amostra.

Como no caso do trabalho de Lima (2011) em que foram utilizados aproximadamente 5 frutos para formar a amostra para a análise de PCR, provavelmente a amostra utilizada tenha sido insuficiente mesmo considerando as várias possibilidades de P_v (1/1000 a 1/1), uma relação de custo de 10:1 e desvio padrão igual a 10% da média amostral.

Tabela 15 Tamanho ótimo de amostra para diferentes proporções de variâncias, desvio padrão amostral fixado em 10% da média, relação de custos de 10:1. Referência: gene ERF, cultivar Acauã ($\mu = 1,67$ e $\sigma_f^2 = 0,137$), 2009.

P_v	Número de plantas	Número de frutos	Frutos na amostra	CV(%)
1/1000	0,1	96,6	5	22,1
1/400	0,1	63,0	6	22,1
1/200	0,1	45,3	6	22,2
1/100	0,2	31,1	7	22,2
1/50	0,3	22,4	7	22,2
1/40	0,4	20,2	7	22,2
1/25	0,5	15,7	8	22,1
1/20	0,6	14,0	8	22,2

1/10	1,0	10,0	10	22,2
1/5	1,6	7,1	12	22,1
1/4	2,0	6,3	13	22,1
1/3	2,6	5,4	14	22,1
1/2	3,6	4,4	16	22,1
1/1	6,5	3,1	20	22,1
2/1	11,8	2,2	27	22,1
4/1	23,2	1,6	37	22,4
5/1	27,7	1,4	39	22,1
10/1	53,6	1,0	53	22,0

Na Tabela 16, estão apresentados os resultados para amostras ótimas simulando um valor para a variância entre frutos dentro de plantas igual a 0,0274, correspondendo a 20% de 0,137.

Tabela 16 Tamanho ótimo de amostra para diferentes proporções de variâncias, desvio padrão amostral fixado em 10% da média, relação de custos de 10:1. Referência: gene ERF, cultivar Acauã ($\mu = 1,67$), 2009. Variância entre frutos considerada igual a 0,0274.

P_v	Número de plantas	Número de frutos	Frutos na amostra	CV(%)
1/1000	0,0	100,0	1	9,9
1/400	0,0	65,2	1	9,9
1/200	0,0	44,8	1	9,9
1/100	0,0	31,7	1	9,9
1/50	0,1	22,2	1	9,9
1/40	0,1	20,1	1	9,9
1/25	0,1	15,8	2	9,9
1/20	0,1	14,1	2	9,9

1/10	0,2	10,0	2	9,9
1/5	0,3	7,0	2	9,9
1/4	0,4	6,3	3	9,9
1/3	0,5	5,5	3	9,9
1/2	0,7	4,5	3	9,9
1/1	1,3	3,2	4	9,9
2/1	2,4	2,2	5	9,9
4/1	4,6	1,6	7	9,9
5/1	5,6	1,4	8	9,9
10/1	10,7	1,0	11	9,9

Em outro experimento onde foi estudado o gene ACO na cultivar Acauã em 2009, o coeficiente de variação determinado foi de 44,39%, para todas as proporções de variâncias e uma proporção de custos de 10:1 com um desvio padrão fixado de 10% da média amostral, seriam necessários mais de 20 frutos na amostra ótima (Tabela 17).

Tabela 17 Tamanho ótimo de amostra para diferentes proporções de variâncias, desvio padrão amostral fixado em 10% da média, relação de custos de 10:1, Referência: gene ACO, cultivar Acauã ($\mu = 7,58$ e $\sigma_f^2 = 11,32$), 2009.

P_v	Número de plantas	Número de frutos	Frutos na amostra	CV(%)
1/1000	0,2	110,7	21	44,3
1/400	0,4	60,5	23	44,3
1/200	0,6	43,7	24	44,5
1/100	0,8	31,1	26	44,4
1/50	1,3	22,0	29	44,3
1/40	1,5	20,2	29	44,4
1/25	2,0	15,7	32	44,4
1/20	2,4	14,3	34	44,5
1/10	4,1	9,8	40	44,6

1/5	6,8	7,0	48	44,3
1/4	7,9	6,4	51	44,4
1/3	9,9	5,5	55	44,2
1/2	14,4	4,5	65	44,8
1/1	25,2	3,2	80	43,9
2/1	48,7	2,2	109	44,7
4/1	93,4	1,6	149	45,3
5/1	108,6	1,4	153	43,3
10/1	222,4	1,0	221	44,8

Na Tabela 18, são apresentados os resultados de amostras ótimas obtidas para as mesmas condições do experimento anterior, mas considerando uma variância entre frutos de 2,2983 correspondendo a 20% da variância residual estimada no experimento,

Tabela 18 Tamanho ótimo de amostra para diferentes proporções de variâncias, desvio padrão amostral fixado em 10% da média, relação de custos de 10:1, Referência: gene ACO, cultivar Acauã ($\mu = 7,58$), 2009, Variância entre frutos considerada igual a 2,2983.

P_v	Número de plantas	Número de frutos	Frutos na amostra	CV(%)
1/1000	0,0	100,7	4	20,0
1/400	0,1	59,4	5	20,0
1/200	0,1	45,4	5	20,0
1/100	0,2	32,3	5	20,0
1/50	0,3	22,7	6	20,0
1/40	0,3	19,7	6	20,0
1/25	0,4	15,7	7	20,0
1/20	0,5	14,1	7	20,0

1/10	0,8	10,0	8	20,0
1/5	1,4	7,1	10	20,0
1/4	1,6	6,3	10	20,0
1/3	2,1	5,5	11	19,9
1/2	2,8	4,5	13	20,0
1/1	5,3	3,2	17	20,0
2/1	9,6	2,3	22	20,0
4/1	19,1	1,6	30	20,1
5/1	22,2	1,4	32	19,9
10/1	43,8	1,0	44	20,0

Nessa simulação, uma amostra ótima com 5 frutos (e apenas uma planta)

iria exigir uma $R_v \geq \frac{1}{100}$ ou seja, a variância entre plantas seria cem vezes menor que a variância entre frutos dentro de plantas,

5 CONCLUSÃO

Para a determinação do tamanho ótimo da amostra biológica é necessário que se tenham dados obtidos de experimentos bem conduzidos e que expressem fielmente a variabilidade entre plantas de café e entre frutos nas plantas para condições que possam variar de acordo com os genes pesquisados.

Destaca-se que, na literatura não foram encontrados tamanhos de amostra (PCR em tempo real) determinados com base estatística.

Em geral, o tamanho da amostra biológica utilizado pode estar sendo subestimado em função principalmente da relação entre as variâncias e da relação de custos,

Para uma relação de custo alta, o tamanho ótimo de amostra se torna ainda maior. Com isso, a obtenção de amostras muito pequena poderá afetar a precisão do experimento diminuindo a confiabilidade,

O número de frutos na amostra ótima aumenta à medida que o custo de amostrar uma planta aumenta em relação ao custo de amostrar um fruto, para uma mesma proporção de variâncias.

REFERÊNCIAS

ARTICO, S. et al. Desenvolvimento de novas ferramentas biotecnológicas para o controle mais efetivo do bicudo-do-algodoeiro: identificação de gene expressos predominantemente em tecidos florais de algodão e clonagem de promotores específicos de flor. In: CONGRESSO BRASILEIRO DO ALGODÃO, 7., 2009, Foz do Iguaçu. Sustentabilidade da cotonicultura brasileira e expansão dos mercados: **Anais**. Campina Grande: Embrapa Algodão, 2009. 1 CD-ROM

BUDZINSKI, F. G. I. et al. Análise de genes expressos durante estádios finais da maturação de frutos de café. In: Simpósio de Pesquisa dos Cafés do Brasil, 4., 2005, Londrina. **Anais...** Brasília, DF: Embrapa Café, 2005.

BUSTIN, S.; MUELLER, R. Real-time reverse transcription PCR (qRT-PCR) and its potential use in clinical diagnosis. **Clinical science**, v. 109, p. 365-379, 2005.

BUSTIN, S. A. et al. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. **Clinical chemistry**, v. 55, n. 4, p. 611-622, 2009.

CASTRO, R.D.et al. A semente de café: desenvolvimento e perspectivas genômicas. In: SIMPÓSIO DE PESQUISA DOS CAFÉS DO BRASIL, 2, Vitória, 2001. **Resumo**. Brasília: Embrapa Café, 2001. p.27.

CHANG, S.; PURYEAR, J.; CAIRNEY, J.A simple and efficient method for isolating RNA from pine trees.**Plant Molecular BiologyReporter**, Amsterdam, v.11, p.113-116, 1993.

FEDERER, W.T. **Experimental design**: theory and application. New York: MacMillan Publishing Company, 1955. 544p.

GALLINA, A. P. et al. Identificação de genes diferencialmente expressos durante a interação incompatível cafeeiro-Hemileiavastatrix. In: Simpósio de Pesquisa dos Cafés do Brasil, 6., 2009, Vitória. Inovação científica, competitividade e mudanças climáticas: **Anais...** Vitória: Consórcio Pesquisa Café, 2009.

GALVÃO, R. M. et al. Identificação e caracterização de genes de ACC oxidase de café. In: II Simpósio de Pesquisa dos Cafés do Brasil, Vitória, ES, **Anais...** Brasil, p.32-33, 2001.

GANGULI, M. A note on nested sampling. **Sankya**. v. 5. p.449-452, 1941.

LADEIRA, P. R. S.; ISAAC, C.; FERREIRA, M. C. Reação em cadeia da polimerase da transcrição reversa em tempo real. Real-time reverse transcription polymerase chain reaction. **Rev Med**, São Paulo, v. 90, n.1, p. 47-51, jan./mar. 2011.

LESSMAN, K. J.; ATKINS, R. E. Optimum plot size and relative efficiency of lattice designs for grain sorghum yield test. **Crop Science**, Madison, v. 3, p. 477-481, 1963.

LIMA, A. A. **Caracterização *in silico* e análise da expressão de genes das rotas de biossíntese e sinalização do etileno em frutos de diferentes cultivares de café (*Coffea arabica*)**. 2011. 124 p. Dissertação (Mestrado em Fisiologia Vegetal)- Universidade Federal de Lavras, Lavras, 2011.

MBELLA, E. G. M. et al. Green qPCR methods for detection of endogenous reference genes in commodity crops: a step ahead in combinatory screening of genetically modified crops in food and feed products. **European Food Research and Technology**, v. 232, n. 3, p. 485-496, 2011. ISSN 1438-2377. Disponível em: <<http://dx.doi.org/10.1007/s00217-010-1408-2>>. Acesso em: 26 maio 2012

MARCUSE, S. Optimum allocation and variance components in nested sampling with an application to chemical analysis. **Biometrics**, v. 5 p. 189-206, 1949.

NOVAIS, C, M; ALVES, M, P. PCR em tempo real: Uma inovação tecnológica da reação em cadeia da polimerase (PCR). **Revista Biotecnologia Ciência e Desenvolvimento**, ed. 33, 10p, jul./dez. 2004. Disponível em: <<http://www.biotecnologia.com.br/revista/bio33/pcr.pdf>> Acesso em: 26 maio 2012

PADILHA, C. E. et al. **Uso da reação em cadeia pela polimerase (pcr) como ferramenta de diagnostico de doenças na aquicultura**. Instituto Federal de Educação, Ciência e Tecnologia Catarinense, Campus Araquari, 2011.

PARANAÍBA, P. F.; FERREIRA, D. F.; MORAIS, A. R. Tamanho ótimo de parcelas experimentais, proposição de métodos de estimação. **Revista Brasileira de Biometria**, São Paulo, v. 27, n. 2, p. 255-268, 2009.

PEZZOPANE, C. G. et al. Expressão gênica diferencial em frutos de Coffea Arábica L, em diferentes estádios de desenvolvimento e maturação. In: SIMPÓSIO DE PESQUISA DOS CAFÉS DO BRASIL, 4., 2005, Londrina. **Anais...** Brasília, DF: Embrapa Café, 2005.

R DEVELOPMENT CORE TEAM. **R**: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, Disponível em: <<http://www.r-project.org/>> Acesso em 26 de maio de 2012

SMITH, H.F. An empirical law describing heterogeneity in the yields of agricultural crops. **Journal Agricultural Science**, Camberra, v. 28, p. 1-23, 1938.