

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

Juracy Mendes MOREIRA¹
Paulo César LIMA²
Daniel Furtado FERREIRA²
Renato Ribeiro de LIMA²
André Almeida LIMA³
Antonio CHALFUN-JÚNIOR⁴

- 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. 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, a população é constituída por 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 e da relação de custos.
- PALAVRAS CHAVE: qPCR; dois estágios; amostra biológica; expressão gênica.

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

¹ Universidade Federal de Lavras – UFLA, Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, CEP: 37200-000, Lavras, MG, Brasil. E-mail: juracimendesmoreira@yahoo.com.br

² Universidade Federal de Lavras – UFLA, Departamento de Ciências Exatas – DEX, CEP: 37200-000, Lavras, MG, Brasil. E-mail: pbola@dex.ufla.br / danieljff@dex.ufla.br / rrlima@dex.ufla.br

³ Universidade Federal de Lavras – UFLA, Programa de Pós-Graduação em Agronomia – Fisiologia Vegetal, CEP: 37200-000, Lavras, MG, Brasil. E-mail: andrelima4@yahoo.com.br

⁴ Universidade Federal de Lavras – UFLA, Departamento de Biologia, CEP: 37200-000, Lavras, MG, Brasil. E-mail: chalfunjunior@dbi.ufla.br

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 (qRT-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 qRT-PCR é muito passível de variações principalmente durante a fase da transcrição reversa.

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

O objetivo proposto nesse trabalho foi determinar o tamanho ótimo de amostra biológica para o estudo da quantificação da expressão gênica utilizando a PCR 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 produtos transgênicos e no estudo de expressão gênica (Padilha *et al.*, 2011).

2.2 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.

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 tem sido feitos diversos estudos com base na síntese de etileno.

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 qRT-PCR a partir de RNA total, livre de DNA, extraído de diferentes tecidos e de três estádios distintos de maturação dos frutos.

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 qRT-PCR.

2.3 Função de custos para amostragem em dois estágios

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_0 para uma amostragem em duas fases é dada por:

$$c_0 = c_1 n_1 + c_2 n_1 n_2.$$

3 Material e métodos

Para esse estudo foram utilizados os dados de expressão gênica determinados para três genes diferentes em quatro experimentos realizados na 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 a 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 Catucaí, possui porte baixo a médio, arquitetura compacta, bom vigor, resistência à ferrugem e ao *Meloidogyne exiguae* estágio de maturação precoce, com boa produtividade.

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. Cada amostra biológica para a utilização na qRT-PCR foi

constituída por 3 a 5 frutos coletados de 6 plantas distintas. A extração do RNA realizada pelo protocolo descrito por Chang *et al.* (1993).

Para a análise da expressão gênica quantitativa por qRT-PCR foi utilizado o modelo ABI PRISM7500 Real-Time PCR (Applied Biosystems), utilizando o sistema de detecção SYBR Green e o cDNA obtido a partir de RNA extraído dos frutos. As condições térmicas da reação foram 2 minutos a 50°C, 10 minutos a 95°C, seguidos por 40 ciclos de 15 segundos a 95°C e 1 minuto a 60°C, e finalizando com 15 segundos a 95°C.

3.2 Métodos

Para esse estudo da amostragem ótima em pesquisas com expressão gênica as plantas do cafezal 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 Modelo aleatório para 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 , onde:

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

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 de plantas e frutos 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_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 – modelo aleatório

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 1.

Tabela 1 -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 (σ_p^2 e σ_f^2) podem ser obtidas por:

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

$$QM_f = S_f^2;$$

ou:

$$S_f^2 = QM_f;$$

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

A estimativa da variância da média amostral é:

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

3.2.4 Função de custos para a amostragem 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_0 é dado por:

$$c_0 = c_p n_p + c_f n_p n_f.$$

3.2.5 Tamanho ótimo de amostra

Nesse estudo procurou-se estudar o tamanho ótimo de amostra fixando a variância da média amostral (v_0) e minimizando o custo de amostragem (c_0).

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:

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

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 conhece as variâncias σ_p^2 e σ_f^2 , utiliza-se estimativas dessas variâncias obtidas de pesquisas já realizadas. Os dados dos experimentos utilizados não possibilitaram a obtenção dessas estimativas. A Tabela 2 apresenta os componentes de variância para os experimentos utilizados.

Tabela2 - 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 folha pode ser determinado por:

$$n'_2 = \sqrt{\frac{P_c}{P_v}};$$

com:

$$P_c = \frac{c_p}{c_f} \text{ e } P_v = \frac{\sigma_p^2}{\sigma_f^2}.$$

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 (P_v) como:

$$n'_p = \frac{1}{v_0} \left[P_v \sigma_f^2 + \frac{\sigma_f^2}{n'_f} \right];$$

substituindo n'_f pela expressão em função das proporções de custos e de variância, a expressão para o número ótimo de plantas será:

$$n'_p = \frac{\sigma_f^2}{v_0} \left[P_v + \sqrt{\frac{P_v}{P_c}} \right].$$

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 às 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, 2011-07-08.

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 media 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) ou o custo total da amostra (c_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 (P_c) e a relação de variâncias (P_v), o número ótimo de plantas n'_1 pode ser determinado por:

$$n'_1 = \frac{\sigma_f^2}{v_0} \left[P_v + \sqrt{\frac{P_v}{P_c}} \right].$$

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 de frutos por planta é dado por:

$$n'_2 = \sqrt{\frac{c_p \sigma_f^2}{c_f \sigma_p^2}};$$

onde: n'_2 é o número ótimo de frutos por planta; c_p é o custo de amostragem de uma planta; c_f é o custo de amostragem de um fruto; σ_p^2 é a variância entre plantas e σ_f^2 é a variância entre frutos dentro de plantas.

Designando $P_c = c_p/c_f$ (proporção entre o custo de seleção e acesso a uma planta pelo custo de colheita de um fruto) e $P_v = \sigma_p^2/\sigma_f^2$ (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:

$$n'_2 = \sqrt{\frac{P_c}{P_v}}. \quad (1)$$

A expressão (1) demonstra 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.

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):

$$n'_1 = \frac{1}{v_0} \left[\sigma_p^2 + \frac{\sigma_f^2}{n'_2} \right];$$

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

$$n'_1 = \frac{\sigma_f^2}{v_0} \left[P_v + \sqrt{\frac{P_v}{P_c}} \right]. \quad (2)$$

Apresentam-se na Tabela 3 as estimativas das médias, das variâncias residuais e dos coeficientes de variação obtidos.

Verifica-se nesta Tabela 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.

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 (P_v), foram obtidos os valores para as variâncias entre plantas (σ_p^2).

Tabela 3 - 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.

Apresentam-se na Tabela 4 as proporções de variâncias e desvio padrão amostral fixado em 10% da média.

Tabela 4 - 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 ($\mu = 3,87; \sigma_f^2 = 8,65$), 2008

Proporção	1:1			10:1		
	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	31,6	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 4 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.

Na Tabela 5 os números ótimos de plantas e frutos por planta foram determinados para as mesmas condições da Tabela 4 com mudanças apenas nos valores da média e da variância entre frutos.

Tabela 5 - Valores de n'_1 e n'_2 com diferentes proporções de variâncias e desvio padrão amostral fixado em 10% da média. Ref.: gene ETR, cultivar Acauã ($\mu = 5,30$; $\sigma_f^2 = 3,81$), 2009

Proporção	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 na Tabela 5 que amostras com um total de 20 frutos ou menos são conseguidas para diferentes proporções de variâncias entre plantas e entre frutos. Para P_c igual a 1, proporções menores que 1/10 e para P_c igual a 10, menores que 1/100.

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 6 e 7 apresentam os tamanhos ótimos de amostra para o gene ERF e cultivar Catucaí nos anos agrícolas de 2008 e 2009, respectivamente.

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 4 a 7).

Tabela 6 - 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 ($\mu = 1,40$; $\sigma_f^2 = 0,180$), 2008

Proporção	1:1			10:1		
	P_v	Plantas	Frutos por planta	Frutos na amostra	Plantas	Frutos por planta
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

Tabela 7- 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 ($\mu = 1,95; \sigma_f^2 = 0,811$), 2009

Proporção	1:1			10:1		
	P_v	Plantas	Frutos por planta	Frutos na amostra	Plantas	Frutos por planta
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

Para o gene ERF (Tabela 6) amostras com um máximo de 20 frutos são conseguidas para proporções entre as variâncias entre plantas e entre frutos menores que 1/1 quando $P_c = 1$ e menores que 1/10 para uma $P_c = 10$ em 2008 enquanto que no ano 2009, para

qualquer relação de custos, não se observou amostras com menos de 20 frutos no total, para todas as proporções de variâncias estudadas (Tabela 7).

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:

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

a equação (2) pode ser reescrita como:

$$n'_1 = \frac{p^2}{d^2} \left[P_v + \sqrt{\frac{P_v}{P_c}} \right] ;$$

e o número ótimo de frutos na amostra ($n'_1 n'_2$) será dado por:

$$n'_1 n'_2 = \frac{p^2}{d^2} (1 + \sqrt{P_c P_v}). \quad (3)$$

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 8). 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 $P_v = 1/100$.

A Tabela 9 apresenta os valores obtidos de amostras ótimas para diferentes proporções de variâncias e simulando uma relação entre média e variância de 43% nesse experimento.

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

Resultados semelhantes foram obtidos 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 folhas teve que ser reduzida para de 1,63 para 0,3271 (CV reduzido de 96 para 43%).

Tabela 8 - 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	1:1			10:1		
	P_v	Plantas	Frutos por planta	Frutos na amostra	Plantas	Frutos por planta
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

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

Proporção	1:1			10:1		
	P_v	Plantas	Frutos por planta	Frutos na amostra	Plantas	Frutos por planta
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

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 folhas na amostra ótima foi menor ou igual a 20 (Tabela 8). 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.

Tabela 10 - 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 ($\mu = 1,67e\sigma_f^2 = 0,137$), 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

No caso desse experimento foram utilizados aproximadamente 5 frutos para formar a amostra para a análise de PCR. Como pode ser observado na Tabela 10, 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 (s_0) igual a 10% da média amostral.

Na Tabela 11 estão apresentados os resultados para amostras ótimas admitindo um valor para a variância entre frutos dentro de plantas igual a 0,0274, correspondendo a 20% de 0,137 (que foi o valor obtido para a estimativa da variância residual).

Tabela 11 - 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 12).

Tabela 12 - 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 ($\mu = 7,58\sigma_f^2 = 11,32$), 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

Tabela 13 - 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

Na Tabela 13 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.

Nessa simulação, uma amostra ótima com 5 frutos (e apenas uma planta) iria exigir uma $P_v \geq 1/100$ ou seja, a variância entre plantas seria cem vezes menor que a variância entre frutos dentro de plantas.

Conclusão

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 e da relação de custos.

MOREIRA, J. M.; LIMA, P. C.; FERREIRA, D. F.; LIMA, R. R.; LIMA, A. A. Optimum biological sample size for real-time PCR study. *Rev. Bras. Biom.*, São Paulo, v.30, n.2. 258-277, 2012.

- **ABSTRACT:** *The polymerase chain reaction (PCR) is a scientific technique used in molecular biology for amplifying fragments of a sample of DNA or RNA, 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 in two stages or sub-sampling, called hierarchical sampling, 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 faithfully express the variability in coffee plants and fruits of plants to allow the conditions may vary according to the genes studied.*
- **KEYWORDS:** *qPCR; two stages; biological samples; gene expression.*

Referências

CASTRO, R. D.; ESTANISLAU, W. T.; MESQUITA, P. R.; HILHORST, H. W. M. A semente do café: desenvolvimento e perspectiva genômica. In: SIMPÓSIO DE PESQUISA DOS CAFÉS DO BRASIL, 2, 2001, Vitória. *Resumos...* Brasília: Embrapa Café, 2001, p.27.

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

GALLINA, A. P.; ALMEIDA, R. F.; CAIXETTA, E. T.; MACIEL-ZAMBOLIM, E.; FREITAS, F. S.; ZAMBOLIM, L.; SAKIYAMA, N. S. Identificação de genes diferencialmente expressos durante a interação incompatível cafeeiro-*Hemileia vastatrix*. In: SIMPÓSIO DE PESQUISAS DOS CAFÉS DO BRASIL, 6, 2009, Vitória. *Anais...* Vitória: Consórcio Pesquisa Café, 2009.

GALVÃO, R. M.; KOBAYASHI, A. K.; RIBAS, A. F.; BESPALHOK FILHO, J. C.; PEREIRA, L. F. P.; VIEIRA, L. G. E. Identificação e caracterização de genes de ACC oxidase de café. In: SIMPÓSIO DE PESQUISAS DOS CAFÉS DO BRASIL, 6., 2009, Vitória. Anais... p.32-33.

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. 124f. Dissertação (Mestrado em Fisiologia Vegetal) – Universidade Federal de Lavras, Lavras, 2011.

PADILHA, C. E.; MONTEIRO, F.; FREITAS, H. A.; LENOCH, R. *Uso da reação em cadeia pela polimerase (pcr) como ferramenta de diagnostico de doenças na aquicultura*. Araquari: Instituto Federal de Educação, Ciência e Tecnologia Catarinense, 2011.7p.

R DEVELOPMENT CORE TEAM. *R: a language and environment for statistical computing*. Versão 2.13.1. Vienna, Austria, R Foundation for Statistical Computing, 2011. Disponível em: <<http://www.R-project.org>>.

Recebido em 26.06.2012

Aprovado após revisão 12.10.2012