

COMPARAÇÃO DE ABORDAGENS BAYESIANAS DE ESTRUTURA GENÉTICA SOB POPULAÇÕES EM DIFERENTES TAXAS DE VARIABILIDADE E DIVERSIDADE

**LUIS WILLIAN PACHECO ARGE¹; DAIANE DE PINHO BENEMANN²;
NEWTON ALEX MAYER³; JOSÉ ANTONIO PETERS⁴; VALMOR JOÃO
BIANCHI⁵**

¹UFPEL- Centro de Genômica e Fitomelhoramento/ FAEM – l.willianpacheco@yahoo.com.br

²UFPEL- Lab. Cultura de Tecidos de Plantas/ Depto de Botânica – daiane_bio@yahoo.com.br

³EMBRAPA Clima Temperado – alex.mayer@embrapa.br

⁴UFPEL- Lab. Cultura de Tecidos de Plantas/ Depto de Botânica – japeters1@hotmail.com

⁵UFPEL- Lab. Cultura de Tecidos de Plantas/ Depto de Botânica – valmorjb@yahoo.com

1. INTRODUÇÃO

Métodos estatísticos Bayesianos tem recentemente promovido grandes avanços em várias áreas da ciência, que agora estendidos para a genética quantitativa e molecular, como por exemplo na determinação de taxas de associação entre variantes genéticas com doenças, fenotipagem (STEPHENS & BALDING, 2009), estrutura genética (PRITCHARD et al., 2000; FALUSH et al., 2003; GAO et al., 2007), entre outras aplicações.

Em genética de populações os modelos Bayesianos tem grande vantagem em relação aos métodos frequentistas por serem caracterizados como métodos probabilísticos ($Pr(y|x)$) e de aprendizagem, ou seja, são modelos estocásticos que buscam padrões através de algoritmos que empregam o modelo de cadeias Markovianas escondidas e a integração pelo método Monte Carlo. Já os métodos frequentistas, são caracterizados por terem suas inferências sobre o desconhecido, baseadas nas medidas de desempenho amostral, ou seja, $P(x|y)$.

De maneira mais simples e aplicada, os modelos Bayesianos capturam a estrutura genética da população pela variância molecular de cada subpopulação usando uma distribuição de probabilidade conjunta e separada pelo conjunto de loci observados. Assim, uma população alvo (natural ou melhorada) tem potencial de ser geneticamente estruturada, de tal forma que as fronteiras que limitam o fluxo gênico existem, ou existiram. A aplicação de abordagens Bayesianas podem ter a capacidade de inferir sobre a estrutura genética da população com maior eficiência que vários dos métodos frequentistas (STEPHENS & BALDING, 2009).

Entretanto, vários são os métodos Bayesianos disponíveis atualmente, dentre estes as abordagens dos softwares STRUCTURE (PRITCHARD et al., 2000; FALUSH et al., 2003) e INSTRUCT (GAO et al., 2007) têm sido extensivamente aplicadas e indicadas para análise da estrutura genética de populações alógamas e autógamas, respectivamente.

Frente ao contexto, o objetivo do presente trabalho foi comparar qual das abordagens Bayesianas é a mais indicada para inferir sobre a estrutura genética molecular de populações em diferentes taxas de equilíbrio de Hardy Weinberg sobre uma população constituída por acessos de portaenxertos de *Prunus* melhorados e de outras espécies.

2. MATERIAL E MÉTODOS

O material vegetal avaliado no presente estudo foi proveniente da "Coleção Porta-enxerto de *Prunus*", da EMBRAPA Clima Temperado, sendo constituída por 34 acessos sul-brasileiros; 27 acessos de portaenxertos introduzidos; e 14 acessos representados por genótipos de diferentes espécies de *Prunus*.

O DNA genômico foi isolado pelo procedimento descrito DOYLE & DOYLE (1987), quantificado e padronizado para 20 ng μl^{-1} .

A PCR foi realizada com 15 pares de oligonucleotídeos iniciadores para loci SSR da série BPPCT (DIRLEWANGER et al., 2002) e UDP (CIPRIANI et al., 1999; TESTOLIN et al., 2000).

A eletroforese foi conduzida em gel de poliacrilamida 6% em voltagem de 9 V cm^{-1} por 2 horas e 30 min., e a revelação foi realizada com nitrato de prata conforme o protocolo descrito por BASSAM et al. (1991).

As análises das abordagens Bayesianas do software STRUCTURE Ver. 2.3.1 e INSTRUCT Ver. 3.2.09 foram conduzidas a partir do conjunto de dados obtido pela avaliação de 15 pares de oligonucleotídeos iniciadores. Em ambas as abordagens realizou-se a análise dos K s variando de 1 a 10, com dez repetições cada, período de aquecimento de 20^4 e de interações de 20^5 . O método de frequência alélica utilizada foi a correlacionada, e ancestralidade comum (Admixture model). Os resultados do logaritmo neperiano (\ln) de probabilidade dos dados ($\ln \text{Pr}(X|K)$) foram utilizados para a detecção do melhor valor de K (EVANNO et al., 2005) e para a identificação de qual abordagem é a mais adequada. Após a detecção do melhor K estatístico realizou-se a sumarização das réplicas de cada K pelo software CLUMPP Ver. 1.1.2 (JAKOBSSON & ROSENBERG, 2007) e a representação gráfica realizada pelo software DISTRUCT Ver. 1.1 (ROSENBERG, 2004).

3. RESULTADOS E DISCUSSÕES

Os resultados do \ln de probabilidade dos dados [$\ln \text{Pr}(X|K)$] (Figura 1A) permitiram inferir que a abordagem Bayesiana do software STRUCTURE foi mais eficiente do que a abordagem do INSTRUCT para inferir sobre a probabilidade de um acesso ao acaso (randômico) estar inserido dentro de uma determinada população (K). Para cada informação a priori (K), a probabilidade a posteriori ($\text{Pr}(X|K)$) é calculada somando o logaritmo da probabilidade a priori com os logaritmos das probabilidades condicionais de cada indivíduo da população.

A identificação dos melhores valores de K estatístico [Figura 1B (EVANNO, et al., 2005)] não correspondeu com os resultados biológicos de agrupamento, pois a melhor informação biológica está contida no $K3$. Os resultados da abordagem hierárquica de agrupamento (dendrograma) pelo coeficiente de Nei & Li (1979) foram utilizados para a ordenação dos acessos de forma hierárquica. A informação a priori de $K3$ juntamente com a ordenação hierárquica (Figura 2) possibilitou a categorização dos acessos referente a um determinado grupo em no mínimo 85% em cada pool alélico, ou seja, o pool alélico 'a' abrange 91,17% dos acessos sul-brasileiros, 'b' 88,88% dos portaenxertos introduzidos; e, 'c' 85,71% dos acessos de outras espécies de *Prunus*. A figura 2 mostra também os diferentes índices de estruturação de cada grupo em relação a toda a coleção. A abordagem Bayesiana do software STRUCTURE leva em consideração a maximização do equilíbrio de Hardy-Weinberg em um determinado grupo para inferir sobre a estrutura genética de populações alógamas e categorizar o acesso em subpopulações (PRITCHARD et al., 2000; FALUSH et al., 2003). Diferentemente,

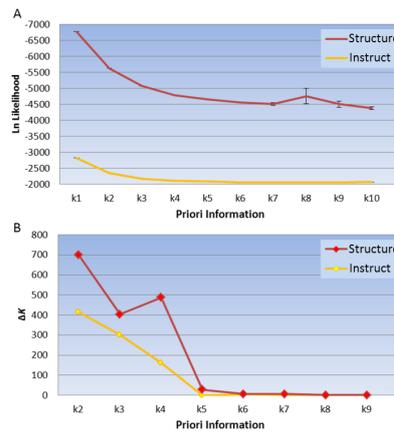


Figura 1: Sumário de informação das abordagens Bayesianas de estrutura genética para cada informação a priori. (A) Resultados dos dados de $\ln \Pr(X|K)$. (B) Resultados da estatística de detecção dos K s com maior conteúdo de informação sobre a "Coleção Porta-enxerto de *Prunus*".

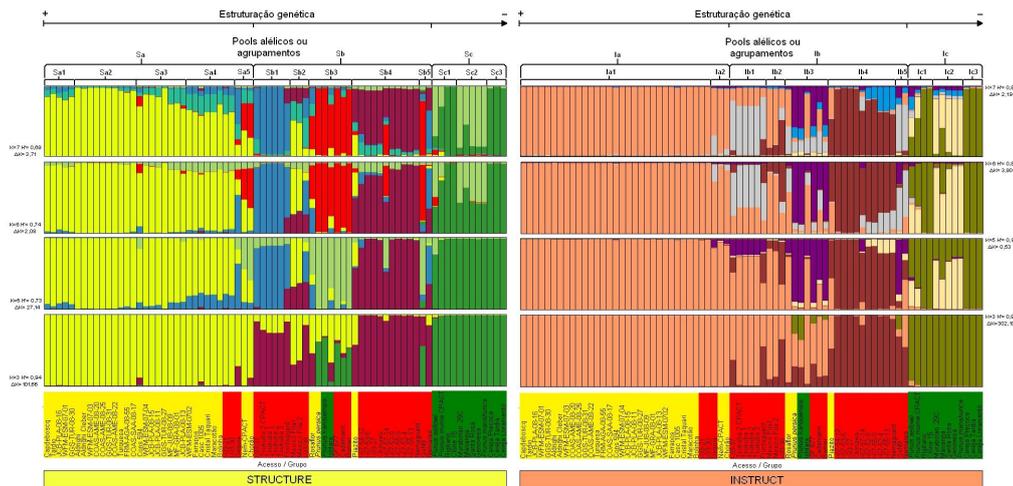


Figura 2: Estrutura genética da "Coleção Portaenxerto de *Prunus*" com base nos resultados da abordagem hierárquica de agrupamento pelo coeficiente de Nei & Li (1979) e Bayesianas do software STRUCTURE e INSTRUMENT. A identificação dos acessos em amarelo são provenientes do grupo sul-brasileiro, vermelho- introduzidos e verde- outras espécies.

a abordagem do software INSTRUMENT não leva em consideração o equilíbrio de Hardy-Weinberg, mas sim a taxa de autofecundação ou endogamia para inferir sobre a estrutura genética e a categorização dos acessos em subpopulações (GAO et al., 2007). Como os acessos deste estudo são caracterizados como de espécies diferentes e de acessos melhorados, há desta maneira uma grande taxa de alelos de baixa frequência, principalmente no grupo de acessos introduzidos e no de outras espécies, o que na maioria das vezes é uma característica (de frequência alélica) encontrada em populações naturais alógamas, mais do que em populações autógamas. Embora o maior grupo analisado seja formado por acessos

de *Prunus persica*, ou seja, por plantas autógamas, era de se esperar a presença de grande número de alelos de alta frequência, o qual não foi observado. Sendo assim, a composição alélica da população pode justificar o fato da abordagem do software STRUCTURE apresentar melhores resultados do que a abordagem do software INSTRUCT, sobre a inferência de estrutura genética na população analisada e a categorização dos acessos em determinados agrupamentos.

4. CONCLUSÕES

O presente estudo pode concluir que a aplicação da abordagem Bayesiana que leva em consideração o equilíbrio de Hardy Weinberg é a mais eficaz para a inferência de estrutura genética em populações autógamas com presença de diferentes acessos melhorados, e com grande diversidade genética.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- BASSAM, B.J.; et al. Fast and sensitive silver staining of DNA in polyacrylamide gels. **Anal Biochemistry**, v.196, p. 80-83, 1991.
- CIPRIANI, G.; et al. AC/GT and AG/CT microsatellite repeats in peach [*Prunus persica* (L.) Batsch]: isolation, characterization and cross-species application in *Prunus*. **Theoretical and Applied Genetics**, n. 99, p. 65-72, 1999.
- DIRLEWANGER, E.; et al. Development of microsatellite markers in peach [*Prunus persica* (L.) Batsch] and their use in genetic diversity analysis in peach and sweet cherry (*Prunus avium* L.). **Theoretical and Applied Genetics**, v. 105, p. 127-138, 2002.
- DOYLE, J.J.; DOYLE, J.L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. **Phytochemistry Bulletin**, v. 19, p. 11-15, 1987.
- EVANNO, G.; et al. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. **Molecular Ecology**, v. 14: p. 2611-2620, 2005.
- FALUSH, D.; et al. Inference of pop. structure using multilocus genotype data: linked loci and correlated allele frequency. **Genetics**, v. 164: 1567-1587, 2003.
- GAO, H.; et al. An MCMC Approach for Joint Inference of Population Structure and Inbreeding Rates from Multi-Locus Genotype Data. **Genetics**, (online), 2007.
- JAKOBSSON, M.; ROSENBERG, N.A. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. **Bioinformatics**, v. 23: p. 1801-1806, 2007.
- NEI, M.; LI, W.H. Mathematical-model for studying genetic-variation in terms of restriction endonucleases. **PNAS**, v. 76: p. 5269-5273, 1979.
- PRITCHARD, J.K.; et al. Inference of population structure using multilocus genotype data. **Genetics**, v. 155: p. 945-959, 2000.
- ROSENBERG, N.A. DISTRUCT: a program for the graphical display of population structure. **Molecular Ecology Note**, v. 4: p. 137-138, 2004.
- STEPHENS, M; BALDING, DJ. Bayesian statistical methods for genetic association studies. **NATURE REVIEWS Genetics**, v. 10, p. 681-690, 2009.
- TESTOLIN R, et al. Microsatellite DNA in peach (*Prunus persica* L. Batsch) and its use in fingerprinting and testing the genetic origin of cultivars. **Genome**, v. 43, n 3: p. 512-520, 2000.