

Método alternativo para análise de agrupamento

Edmar Soares de Vasconcelos⁽¹⁾, Cosme Damião Cruz⁽¹⁾, Leonardo Lopes Bhering⁽¹⁾
e Márcio Fernando Ribeiro Resende Júnior⁽¹⁾

⁽¹⁾Universidade Federal de Viçosa, Dep. de Biologia Geral, Bioagro, Av. P.H. Rolfs, s/nº, CEP 36571-000 Viçosa, MG. E-mail: vasconceloses@yahoo.com.br, cdcruz@ufv.br, leobhering@yahoo.com.br, marciopac@hotmail.com

Resumo – O objetivo deste trabalho foi propor uma alteração no procedimento de agrupamento realizado pelo método de Tocher, passando-o de método simultâneo (original) para seqüencial. Para ilustrar e comparar os métodos, foram realizadas simulações de coleções de acessos com diferentes características, tanto para avaliação individual quanto para avaliação de experimentos com repetições. As simulações e as análises estatísticas foram realizadas com auxílio do programa GENES. O número de grupos formados com o método de Tocher seqüencial foi menor que o número de grupos formados pelo método de Tocher original. No método de Tocher seqüencial, não foi verificada influência dos genótipos já agrupados, no agrupamento dos genótipos mais distantes. O limite de acréscimo na média da distância intragrupo, estimado após a formação de um novo grupo, constitui uma estimativa da dissimilaridade existente entre os acessos dos grupos. O agrupamento dos genótipos com maior dissimilaridade é realizado com maior eficácia pelo método de Tocher seqüencial do que pelo método de Tocher original.

Termos para indexação: método de Tocher original, método de Tocher seqüencial, formação de grupos, genótipos.

Alternative methodology for the cluster analysis

Abstract – The objective of this work was to propose an alteration in the clustering procedure accomplished by the Tocher method, changing it from a simultaneous to a sequential method. Accession collections were generated through simulation to illustrate and compare the original and the sequential methodologies. In each collection, different characteristics for individual data as well as experiments with repetition were evaluated. The simulations and the statistical analyses were accomplished using the GENES applicative. The number of groups formed with the sequential method was smaller than the number formed with the original Tocher method. The genotypes clustering were not influenced by the most distant genotype when the sequential method was used. The increment limit in the mean of the intra-group distance, estimated after the formation of a new group, is an estimative of dissimilarity between the group's accessions. The sequential Tocher method grouped genotypes comprising higher dissimilarity with more effectiveness than the original Tocher method.

Index terms: original Tocher method, sequential Tocher method, clustering, genotypes.

Introdução

Estudos de dissimilaridade atendem a determinados objetivos dos melhoristas por propiciarem informações acerca do grau de semelhança ou de diferença entre dois ou mais genótipos. Entretanto, o número de estimativas de dissimilaridade obtido é relativamente elevado quando se tem grande número de genótipos, o que torna, às vezes, impraticável o reconhecimento de grupos homogêneos por um simples exame visual. Portanto, o uso de métodos que agrupem os genótipos pode ser uma das melhores alternativas para análise e interpretação dos dados (Cruz & Carneiro, 2006).

Há inúmeros métodos de agrupamento, que se distinguem pelo tipo de resultado a ser fornecido e pelas

diferentes formas de definir a proximidade entre um indivíduo e um grupo já formado, ou entre dois grupos quaisquer. Entre esses métodos, o de Tocher apresenta uma inconveniência no agrupamento dos genótipos com maior dissimilaridade: na maioria dos casos, cada genótipo forma um grupo específico (um grupo de apenas um genótipo) em virtude de este agrupamento ser influenciado pela distância dos genótipos já agrupados.

O método de agrupamento por otimização ou método de Tocher, apresentado em Cruz & Carneiro (2006), constitui um método de agrupamento simultâneo, o qual realiza a separação dos genótipos em grupos de uma só vez. Esse método utiliza um único critério de agrupamento e possui a particularidade de apresentar

a distância média dentro dos grupos sempre menor que a distância média entre os grupos.

O método de Tocher vem sendo amplamente empregado no estudo da divergência genética entre acessos. A exemplo do trabalho de Souza et al. (2005), que avaliaram a divergência genética entre 31 genótipos de melancia, por meio de técnicas de agrupamento e de variáveis canônicas. No trabalho de Oliveira et al. (2004), também utilizou-se o método de Tocher na avaliação de divergência genética entre genótipos de alface. Silva et al. (2005) utilizaram esse método a fim de avaliar a diversidade genética de genótipos de cana-de-açúcar. Marchioro et al. (2003) utilizaram o método de agrupamento de Tocher para avaliação da dissimilaridade genética entre genótipos de aveia. Em tomateiro, Karasawa et al. (2005) usaram esse método para avaliar a divergência genética entre acessos de bancos de germoplasma, permitindo indicação de genitores potenciais para a utilização em programas de melhoramento.

No estudo de diversidade genética, pode-se, ainda, utilizar o método de agrupamento de Tocher com critério de aglomeração inverso (Vasconcelos et al., 2007). Por meio desse processo procura-se realizar o agrupamento de indivíduos com maior dissimilaridade. O método, com tais características, foi proposto, principalmente, para avaliação de acessos de bancos de germoplasma visando a estabelecer uma coleção nuclear.

O objetivo deste trabalho foi propor alteração no procedimento de agrupamento realizado pelo método de Tocher, passando-o de método simultâneo para seqüencial. Esta proposta muda a forma de agrupamento dos indivíduos e estabelece grupos que possam propiciar maior interesse em certas áreas da pesquisa, por melhorar o agrupamento entre os indivíduos mais divergentes.

Material e Métodos

Para ilustrar a comparação entre os métodos de Tocher original e seqüencial, foi utilizado um exemplo em que foram simulados valores de sete variáveis para

30 tratamentos (acessos), cujas médias e variâncias são apresentadas na Tabela 1.

No método de otimização de Tocher, realiza-se a partição do conjunto de acessos em subgrupos não-vazios e mutuamente exclusivos, por meio da maximização ou minimização de alguma medida preestabelecida. Nesse método, adota-se o critério de que a média das medidas de dissimilaridade, dentro de cada grupo, deve ser menor que as distâncias médias entre quaisquer grupos.

A partir de uma matriz de dissimilaridade, identificou-se o par de acessos mais similar. Esses acessos formaram o grupo inicial. Em seguida, se avaliou a possibilidade de inclusão de novos acessos, de forma que a distância média dos acessos dentro do grupo fosse menor que as distâncias médias entre quaisquer grupos.

Nesse método, adotou-se o valor máximo da medida de dissimilaridade encontrado no conjunto das menores distâncias envolvendo cada progenitor (designado por α , para os dados avaliados, correspondeu a $\alpha = 1,076$, acessos 27 e 29) (Tabela 2) como o limite de acréscimo na média da distância intragrupo, ou seja, para a formação ou inclusão de acesso a um grupo.

Na formação do grupo 1, foram selecionados acessos com distâncias menores, no caso, os acessos 17 e 23 com distância de 0,124, menor que o valor de α estabelecido. Uma vez formado este grupo, obtiveram-se as medidas de dissimilaridade entre os acessos e o grupo recém-formado, por meio de $d_{(ij)k} = d_{ik} + d_{jk}$, em que d é qualquer medida de dissimilaridade.

Com base nas distâncias entre acessos-grupo, constatou-se, no exemplo, que o acesso 20 foi o mais similar (menor valor de $d_{(ij)k}$) ao grupo inicial. O acréscimo médio no valor da distância dentro do grupo foi dado por:

$$\frac{d_{(17,23,20)} - d_{(17,23)}}{C_{3,2} - C_{2,2}} = \frac{d_{(17,23)20}}{2} = \frac{(0,425+0,368) - 0,124}{2} = 0,3345,$$

em que $C_{3,2}$ e $C_{2,2}$ são combinações de três, dois a dois, e de dois, dois a dois, respectivamente.

Como o valor de $d_{(17,23)20} = 0,3345$ é inferior a α , a inclusão do acesso 20 no grupo inicial foi permitida.

Tabela 1. Valores de média e variância empregados para a simulação de sete características em 30 acessos, utilizados como exemplo para a distinção entre os métodos de Tocher original e seqüencial⁽¹⁾.

Características	Var 1	Var 2	Var 3	Var 4	Var 5	Var 6	Var 7
Média	469,7	1.302,7	823,3	194,0	3.283,3	7.355,3	391,0
Variância	24.864,8	79.671,9	30.408,7	9.547,8	146.447,6	2.492.776,7	46,0

⁽¹⁾Var: variável simulada.

Novamente, foram obtidas as distâncias dos acessos ainda não agrupados, em relação ao grupo em formação e, da mesma forma, se realizou novo teste, a fim de

incluir um outro acesso ao grupo em formação. Quando, na inclusão de um novo acesso ao grupo inicial, obtinha-se um valor no acréscimo da distância, dentro do grupo,

Tabela 2. Medida de dissimilaridade obtida a partir da distância euclidiana entre os 30 tratamentos simulados a partir dos dados da Tabela 1.

Genótipos	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0	1,189	0,524	1,055	1,131	2,083	1,044	1,915	1,160	1,524	1,167	1,740	0,988	1,446	0,495
2	1,189	0	0,869	0,443	0,751	1,412	0,356	1,200	0,510	1,333	1,107	1,399	0,951	1,087	1,155
3	0,524	0,869	0	0,777	0,930	1,810	0,689	1,735	0,857	1,298	1,039	1,570	0,790	1,320	0,533
4	1,055	0,443	0,777	0	0,407	1,648	0,426	1,523	0,356	1,202	1,178	1,541	1,082	0,993	1,157
5	1,131	0,751	0,930	0,407	0	1,843	0,674	1,707	0,410	1,103	1,273	1,457	1,265	1,208	1,357
6	2,083	1,412	1,810	1,648	1,843	0	1,505	0,894	1,603	1,506	1,363	1,989	1,361	1,401	1,922
7	1,044	0,356	0,689	0,426	0,674	1,505	0	1,372	0,340	1,153	0,983	1,347	0,792	1,138	1,006
8	1,915	1,200	1,735	1,523	1,707	0,894	1,372	0	1,500	1,816	1,150	1,844	1,222	1,612	1,814
9	1,160	0,510	0,857	0,356	0,410	1,603	0,340	1,500	0	1,010	1,120	1,303	1,031	1,135	1,238
10	1,524	1,333	1,298	1,202	1,103	1,506	1,153	1,816	1,010	0	1,311	1,455	1,307	1,281	1,598
11	1,167	1,107	1,039	1,178	1,273	1,363	0,983	1,150	1,120	1,311	0	1,748	0,452	1,521	1,148
12	1,740	1,399	1,570	1,541	1,457	1,989	1,347	1,844	1,303	1,455	1,748	0	1,601	1,998	1,758
13	0,988	0,951	0,790	1,082	1,265	1,361	0,792	1,222	1,031	1,307	0,452	1,601	0	1,413	0,809
14	1,446	1,087	1,320	0,993	1,208	1,401	1,138	1,612	1,135	1,281	1,521	1,998	1,413	0	1,440
15	0,495	1,155	0,533	1,157	1,357	1,922	1,006	1,814	1,238	1,598	1,148	1,758	0,809	1,440	0
16	1,429	0,869	1,152	0,757	1,095	1,746	0,901	1,694	0,996	1,682	1,436	2,207	1,325	1,063	1,387
17	1,539	0,916	1,184	0,952	1,158	1,164	0,872	1,543	0,900	0,931	1,412	1,589	1,217	0,815	1,415
18	1,897	1,698	1,930	1,715	1,676	1,823	1,588	1,580	1,539	1,564	1,326	1,877	1,485	1,781	1,964
19	1,125	0,825	0,965	1,116	1,325	1,543	0,903	1,293	1,120	1,601	1,267	1,185	0,965	1,431	0,951
20	1,765	1,201	1,409	1,307	1,520	1,224	1,127	1,695	1,211	1,134	1,608	1,658	1,350	1,115	1,545
21	1,001	0,903	0,791	1,064	1,291	1,365	0,753	1,229	1,028	1,371	0,602	1,618	0,180	1,363	0,757
22	2,367	2,302	2,470	2,321	2,197	2,457	2,187	2,221	2,098	1,985	2,041	1,892	2,133	2,399	2,471
23	1,539	1,005	1,194	1,034	1,232	1,182	0,942	1,602	0,981	0,917	1,437	1,612	1,228	0,839	1,401
24	0,888	1,008	0,726	1,169	1,365	1,715	0,853	1,576	1,119	1,501	1,088	1,362	0,674	1,527	0,563
25	1,578	0,848	1,319	1,102	1,400	1,427	0,826	1,332	1,034	1,591	1,360	1,636	1,087	1,251	1,338
26	1,437	0,930	1,146	0,892	1,049	1,016	0,902	1,371	0,890	0,804	1,148	1,724	1,095	0,698	1,407
27	1,312	1,699	1,492	1,744	1,695	2,168	1,745	1,837	1,784	1,933	1,572	1,658	1,587	2,037	1,532
28	1,455	1,271	1,450	1,147	1,197	1,590	1,219	1,698	1,133	1,097	1,485	1,783	1,436	0,685	1,532
29	1,287	0,988	1,186	1,099	1,048	1,831	1,073	1,539	1,044	1,470	1,473	0,844	1,375	1,615	1,419
Genótipos	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	1,429	1,539	1,897	1,125	1,765	1,001	2,367	1,539	0,888	1,578	1,437	1,312	1,455	1,287	1,571
2	0,869	0,916	1,698	0,825	1,201	0,903	2,302	1,005	1,008	0,848	0,930	1,699	1,271	0,988	1,091
3	1,152	1,184	1,930	0,965	1,409	0,791	2,470	1,194	0,726	1,319	1,146	1,492	1,450	1,186	1,268
4	0,757	0,952	1,715	1,116	1,307	1,064	2,321	1,034	1,169	1,102	0,892	1,744	1,147	1,099	1,132
5	1,095	1,158	1,676	1,325	1,520	1,291	2,197	1,232	1,365	1,400	1,049	1,695	1,197	1,048	1,259
6	1,746	1,164	1,823	1,543	1,224	1,365	2,457	1,182	1,715	1,427	1,016	2,168	1,590	1,831	0,884
7	0,901	0,872	1,588	0,903	1,127	0,753	2,187	0,942	0,853	0,826	0,902	1,745	1,219	1,073	1,009
8	1,694	1,543	1,580	1,293	1,695	1,229	2,221	1,602	1,576	1,332	1,371	1,837	1,698	1,539	1,354
9	0,996	0,900	1,539	1,120	1,211	1,028	2,098	0,981	1,119	1,034	0,890	1,784	1,133	1,044	1,026
10	1,682	0,931	1,564	1,601	1,134	1,371	1,985	0,917	1,501	1,591	0,804	1,933	1,097	1,470	0,712
11	1,436	1,412	1,326	1,267	1,608	0,602	2,041	1,437	1,088	1,360	1,148	1,572	1,485	1,473	1,123
12	2,207	1,589	1,877	1,185	1,658	1,618	1,892	1,612	1,362	1,636	1,724	1,658	1,783	0,844	1,637
13	1,325	1,217	1,485	0,965	1,350	0,180	2,133	1,228	0,674	1,087	1,095	1,587	1,436	1,375	1,053
14	1,063	0,815	1,781	1,431	1,115	1,363	2,399	0,839	1,527	1,251	0,698	2,037	0,685	1,615	0,975
15	1,387	1,415	1,964	0,951	1,545	0,757	2,471	1,401	0,563	1,338	1,407	1,532	1,532	1,419	1,480
16	0	1,159	2,051	1,529	1,449	1,256	2,808	1,237	1,498	1,153	1,112	2,296	1,460	1,781	1,376
17	1,159	0	1,837	1,255	0,425	1,170	2,400	0,124	1,279	1,000	0,492	2,119	1,062	1,474	0,539
18	2,051	1,837	0	1,879	1,975	1,523	0,933	1,869	1,790	1,603	1,652	2,159	1,285	1,876	1,572
19	1,529	1,255	1,879	0	1,365	0,901	2,281	1,277	0,645	1,071	1,341	1,339	1,520	0,836	1,373
20	1,449	0,425	1,975	1,365	0	1,279	2,479	0,368	1,325	1,014	0,844	2,351	1,296	1,707	0,720
21	1,256	1,170	1,523	0,901	1,279	0	2,173	1,181	0,590	0,948	1,101	1,669	1,415	1,400	1,069
22	2,808	2,400	0,933	2,281	2,479	2,173	0	2,415	2,233	2,200	2,304	2,377	1,783	2,120	2,184
23	1,237	0,124	1,869	1,277	0,368	1,181	2,415	0	1,272	1,054	0,526	2,128	1,074	1,518	0,532
24	1,498	1,279	1,790	0,645	1,325	0,590	2,233	1,272	0	1,039	1,370	1,576	1,539	1,225	1,341
25	1,153	1,000	1,603	1,071	1,014	0,948	2,200	1,054	1,039	0	1,201	2,243	1,328	1,567	1,193
26	1,112	0,492	1,652	1,341	0,844	1,101	2,304	0,526	1,370	1,201	0	1,932	0,938	1,465	0,375
27	2,296	2,119	2,159	1,339	2,351	1,669	2,377	2,128	1,576	2,243	1,932	0	1,972	1,076	2,009
28	1,460	1,062	1,285	1,520	1,296	1,415	1,783	1,074	1,539	1,328	0,938	1,972	0	1,556	1,058
29	1,781	1,474	1,876	0,836	1,707	1,400	2,120	1,518	1,225	1,567	1,465	1,076	1,556	0	1,545

superior ao valor de α , o acesso não entrava no grupo e fechava, assim, o grupo 1.

A formação do grupo 2 ocorreu de forma semelhante à do grupo 1, porém com uso da matriz de dissimilaridade apenas dos genótipos ainda não agrupados, contudo, com base no mesmo valor de α .

Se o valor do acréscimo da distância intragrupo fosse menor ou igual ao valor de α , o acesso testado entraria no grupo e a inclusão de novos acessos era testada. Contudo, se o valor de acréscimo médio na distância dentro do grupo superasse o valor de α , então, o grupo era fechado e o acesso avaliado, excluído. A formação dos demais grupos foi realizada de maneira similar à descrita para o grupo 2. O agrupamento de Tocher original, para os dados analisados, está apresentado na Figura 1.

No método de Tocher seqüencial, a obtenção de α e a formação do grupo 1 foi realizada da mesma forma descrita no item anterior, porém, para a formação do grupo 2, foi obtido um novo valor de α , voltando assim ao início da divisão em grupos (antes da formação do grupo 1). A matriz, da qual se extraiu o novo valor de α , era composta apenas pelos acessos ainda não agrupados. Assim, esse valor de α foi obtido do valor máximo da medida de dissimilaridade, encontrado no

conjunto das menores distâncias, envolvendo cada progenitor ainda não-agrupado. Isso propicia um novo valor do critério de agrupamento após a formação de cada novo grupo, e não um critério único, conforme o método de Tocher original. Dessa forma, o método de Tocher passou a ter característica de um método seqüencial (Figura 2).

Com a finalidade de exemplificar a comparação dos métodos e avaliar a eficácia dos agrupamentos, foi realizado um estudo do agrupamento de pontos estrategicamente conhecidos e em um espaço bidimensional (para apenas duas variáveis X e Y) (Figura 3). Na Figura 3 A, os pontos encontram-se distribuídos de maneira a se obter uma estrutura quadrada. Na Figura 3 B, além da estrutura quadrada, foram acrescentados pontos distantes dessa estrutura a fim de verificar se possuíam influência no agrupamento inicial. Na Figura 3 C, os pontos encontram-se distribuídos de maneira a se obter uma estrutura circular, e acrescentaram-se novos pontos distantes dos iniciais para verificar a influência no agrupamento (Figura 3 D).

Na obtenção de maior destaque entre os agrupamentos obtidos pelos métodos de Tocher original

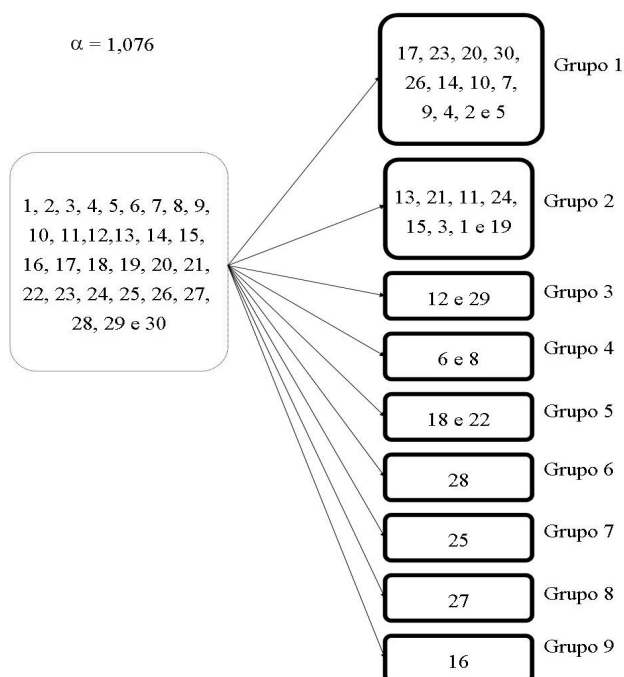


Figura 1. Esquema de agrupamento dos dados simulados, originado pelo método de agrupamento de Tocher, em que cada número representa um acesso.

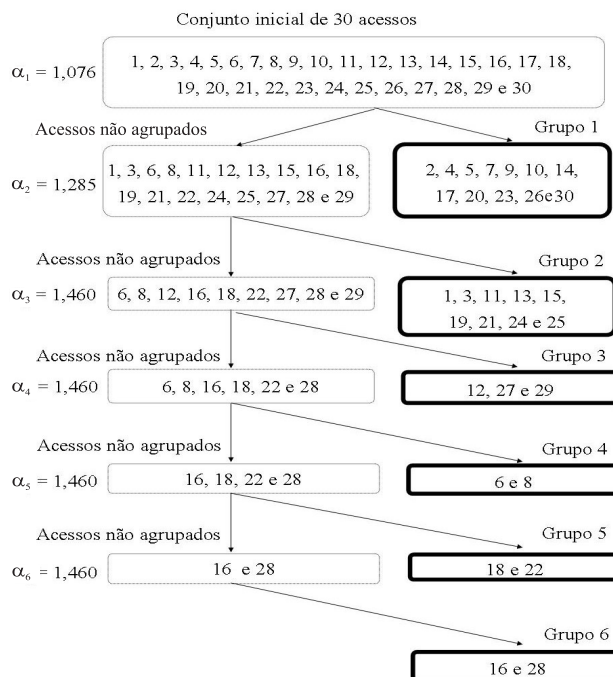


Figura 2. Esquema de agrupamento dos dados simulados, originado do método de Tocher seqüencial, em que cada acesso está representado por um número.

e seqüencial, realizou-se a avaliação de outros dois grupos de genótipos, também em um espaço bidimensional (Figura 4).

Com a finalidade de estabelecer outro exemplo do emprego dos métodos discutidos neste trabalho, foram simuladas cinco variáveis para uma coleção de 100 acessos, em que as médias e as respectivas variâncias se encontram apresentadas na Tabela 3. Realizou-se, ainda, a simulação de um experimento em delineamento em blocos ao acaso, constituído de 50 tratamentos, com quatro repetições. As características deste experimento, com genótipos de alface (Tabela 4), foram retiradas do trabalho de Oliveira et al. (2004). As simulações e as análises estatísticas foram realizadas com auxílio do programa GENES (Cruz, 2006).

Resultados e Discussão

Na Figura 3, são apresentados os genótipos simulados para duas variáveis (X e Y), em que é possível verificar que não existe diferença no padrão de agrupamentos entre os métodos de Tocher original e o seqüencial, tanto para o esquema na forma quadrática quanto na forma circular. O agrupamento manteve o padrão apresentado inicialmente, mesmo quando foram incluídos pontos distantes dos agrupados inicialmente, isto para ambos os métodos de agrupamento empregados (comparação da Figura 3 A com Figura 3 B e da Figura 3 C com Figura 3 D). Estes resultados evidenciam que o método de agrupamento de Tocher seqüencial pode proporcionar resultados semelhantes ao do método de Tocher original, dependendo dos materiais avaliados.

Na Figura 4 A, os genótipos agrupados pelo método original geraram sete grupos. Na Figura 4 B os mesmos genótipos, submetidos ao método seqüencial, geraram um agrupamento diferente do anterior. O método de agrupamento de Tocher seqüencial além de manter a característica de menor distância dentro do grupo do que entre grupos ainda propiciou um agrupamento dos genótipos com maior proximidade com a inclusão, por exemplo, do genótipo 9 no grupo 3, além da formação de novo grupo com os genótipos 15 e 16.

Com relação à Figura 4 C e D, pode ser verificado que nos grupos 1 e 2 não houve diferença para os diferentes métodos de agrupamento. Contudo, o método de Tocher original agrupou apenas os genótipos 7, 9 e 10 no grupo 3, ao passo que o método seqüencial, além

dos genótipos 7, 9 e 10, também agrupou o 8 e o 6 nesse grupo, o que seria esperado, já que os genótipos são

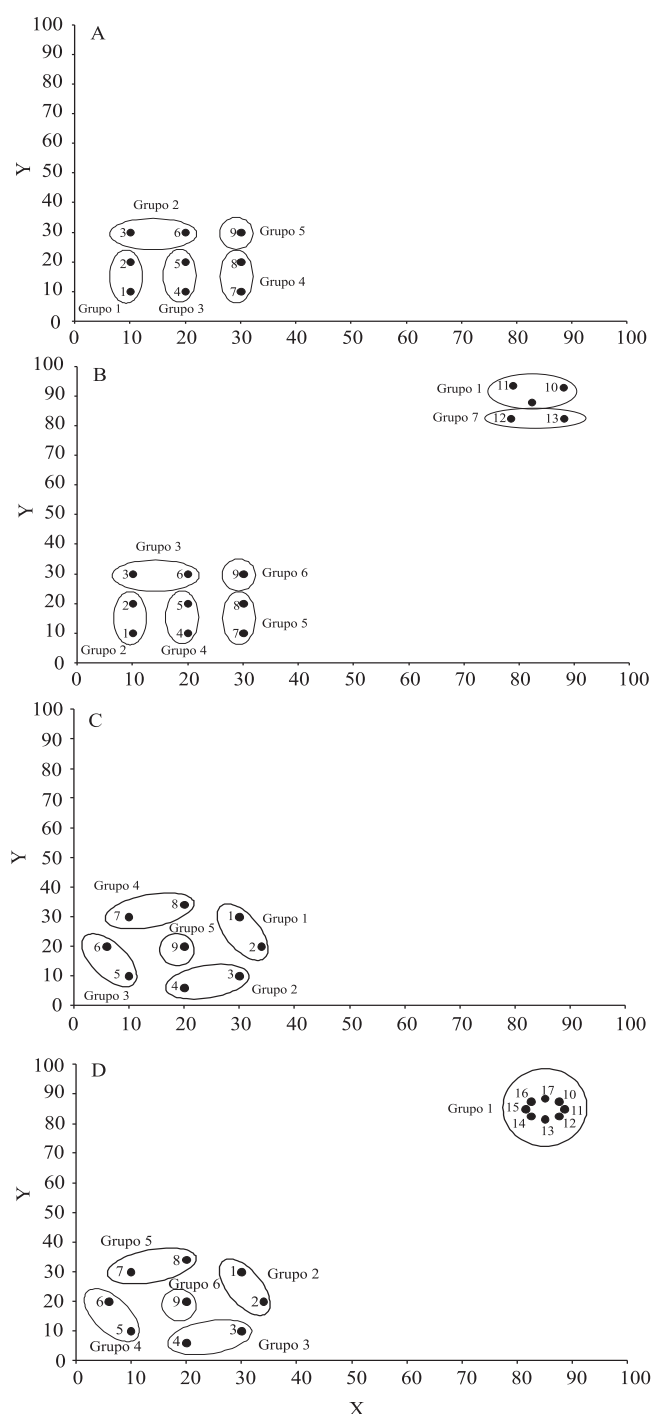


Figura 3. Agrupamento de pontos bidimensionais com dispersão conhecida, ou seja, quadrática com agrupamento pelos métodos de Tocher original (A) e seqüencial (B), e dispersão circular com agrupamento pelos métodos de Tocher original (C) e seqüencial (D).

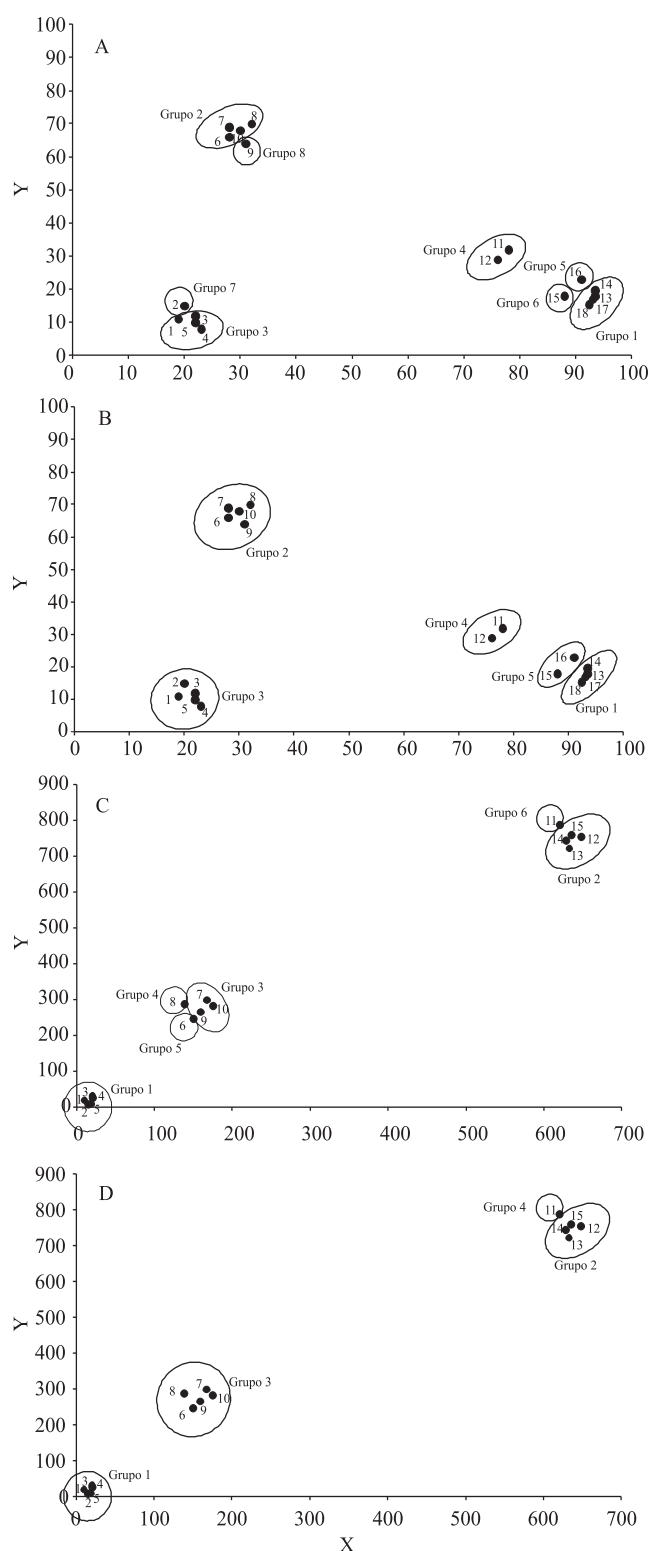


Figura 4. Agrupamento de dois exemplos de pontos bidimensionais com dispersão conhecidas. Exemplo 1: agrupamento pelos métodos de Tocher original (A) e seqüencial (B). Exemplo 2: agrupamento pelos métodos de Tocher original (C) e seqüencial (D).

apresentados próximos no gráfico. Esses resultados evidenciam que o método de agrupamento de Tocher seqüencial, neste caso, foi melhor que o método de Tocher original, já que dispôs genótipos de maior proximidade em grupos semelhantes.

Na Figura 4, era esperado que os genótipos de maior proximidade estivessem em um mesmo grupo, o que foi observado com o uso do método Tocher seqüencial, enquanto o método original separou esses genótipos em maior número de grupos. O método Tocher seqüencial, portanto, apresentou um melhor desempenho, neste caso, do que o método simultâneo.

O agrupamento dos 150 acessos simulados é apresentado na Tabela 5. Estes resultados possibilitam afirmar que o agrupamento pelo método de Tocher seqüencial reduziu o número de grupos formados, quando comparado ao método original (o que era esperado, já que os genótipos já agrupados não afetam o agrupamento dos demais), além de ser obtido aumento no valor máximo da medida de dissimilaridade, presente no conjunto das menores distâncias envolvendo cada progenitor (valor de α) na formação de novos grupos. Assim, à medida que se realiza o agrupamento dos acessos, reduz-se a exigência no acréscimo médio da distância, proporcionado pela inclusão de um novo acesso ao grupo em formação.

O critério de seleção, neste trabalho, possibilitou agrupar genótipos que apresentaram comportamentos próximos, mas que, por apresentarem distância maior que a distância entre grupos próximos, seriam agrupados separadamente pelo método de Tocher original. O valor máximo da medida

Tabela 3. Valores de médias e variâncias utilizados na simulação de cinco variáveis para 100 acessos de uma dada coleção, utilizadas na comparação entre os métodos de Tocher original e alternativo (Tocher seqüencial).

Características	Variável 1	Variável 2	Variável 3	Variável 4	Variável 5
Média	1.410	130	5.820	19	23.200
Variância	865	70	3.040	7	14.644

Tabela 4. Características de alface, utilizadas para simular um experimento de comparação entre os métodos de Tocher original e alternativo (Tocher seqüencial)⁽¹⁾.

Características	RNFO	PFF	PFC	PFR	PFPL
Média	6,3	233,9	26,7	51,7	325,8
Herdabilidade	98,9	92,5	94,5	66,4	81,4
Coefficiente de variação (%)	7,1	7,5	14,0	15,9	7,3

⁽¹⁾RNFO: raiz quadrada do número de folhas; PFF: peso da matéria fresca de folha; PFC: peso da matéria fresca de caule; PFR: peso da matéria fresca de raiz; PFPL: peso da matéria fresca da planta.

de dissimilaridade encontrado no conjunto das menores distâncias, envolvendo cada progenitor ainda não-agrupado, é restabelecido após a formação de um novo grupo. Dessa forma, não existe influência dos genótipos já agrupados no agrupamento dos demais, pelo método de Tocher seqüencial.

O valor de α , estimado após a formação de um novo grupo, no método seqüencial, teve relação apenas com os acessos ainda não-agrupados, não dependendo do grupo 1, o qual possui a menor distância média entre os

acessos, tendo menor valor de α . O último grupo (no nosso exemplo o grupo 10) é o que teve maior distância média intragrupo, que é dependente de α .

Os resultados obtidos pelo emprego dos métodos de Tocher original e seqüencial, sobre os dados simulados com as características de variáveis de alface avaliadas por Oliveira et al (2004), encontram-se apresentados na Tabela 6. O número de grupos formados pelo método de Tocher seqüencial também foi menor que o número de grupos formados pelo método de Tocher original.

Tabela 5. Resultado do agrupamento de 150 acessos, em que avaliaram-se cinco variáveis simuladas para cada acesso.

Grupos	$\alpha^{(1)}$	Tocher seqüencial	Por Tocher ($\alpha = 2,7389$)
		Acessos	Acessos
1	2,7389	2, 3, 4, 5, 7, 8, 9, 11, 13, 16, 17, 19, 25, 30, 31, 32, 33, 35, 36, 38, 40, 43, 45, 48, 50, 51, 53, 54, 56, 57, 58, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 75, 77, 78, 80, 84, 86, 87, 90, 91, 94, 95, 96, 99, 100, 101, 103, 104, 106, 107, 108, 110, 112, 114, 116, 117, 118, 119, 120, 123, 125, 127, 128, 129, 130, 131, 135, 137, 138, 139, 140, 142, 144, 145, 146, 148, 149 e 150	2, 3, 4, 5, 7, 8, 9, 11, 13, 16, 17, 19, 25, 30, 31, 32, 33, 35, 36, 38, 40, 43, 45, 48, 50, 51, 53, 54, 56, 57, 58, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 75, 77, 78, 80, 84, 86, 87, 90, 91, 94, 95, 96, 99, 100, 101, 103, 104, 106, 107, 108, 110, 112, 114, 116, 117, 118, 119, 120, 123, 125, 127, 128, 129, 130, 131, 135, 137, 138, 139, 140, 142, 144, 145, 146, 148, 149 e 150
2	2,9276	6, 18, 26, 29, 34, 41, 46, 59, 60, 79, 81, 88, 102, 109, 111, 126, 133 e 136	6, 18, 26, 29, 46, 59, 60, 79, 81, 88, 111, 126, 133 e 136
3	2,9618	1, 21, 37, 44, 93, 115, 122, 132 e 141	21, 37, 44, 93, 115, 122, 132 e 141
4	2,9618	24, 28, 52, 82, 98, 105 e 143	24, 28, 52, 82, 102, 105 e 143
5	2,9618	10, 22, 42, 55 e 92	14, 41, 76 e 109
6	3,0455	27, 83 e 121	10, 22, 42, 55 e 92
7	3,7986	12, 14, 15, 49, 74, 89, 124, 134 e 147	27 e 83
8	3,7986	20, 39, 85 e 97	89 e 134
9	4,8338	23 e 76	12, 15, 34, 49 e 124
10	4,8338	47 e 113	39 e 74
11	-	-	20 e 97
12	-	-	113
13	-	-	85
14	-	-	98
15	-	-	1
16	-	-	147
17	-	-	121
18	-	-	47
19	-	-	23

⁽¹⁾Valor máximo da medida de dissimilaridade encontrado no conjunto das menores distâncias envolvendo cada progenitor.

Tabela 6. Resultado do agrupamento de 50 tratamentos, oriundos da simulação de um ensaio experimental, em delineamento de blocos ao acaso, em quatro repetições.

Grupos	$\alpha^{(1)}$	Tocher seqüencial	Por Tocher ($\alpha = 22,480$)
		Acessos	Acessos
1	22,480	2, 4, 9, 11, 14, 15, 16, 19, 20, 21, 22, 24, 26, 28, 30, 32, 35, 38, 39, 46 e 47	2, 4, 9, 11, 14, 15, 16, 19, 20, 21, 22, 24, 26, 28, 30, 32, 35, 38, 39, 46 e 47
2	59,329	1, 3, 5, 7, 13, 17, 18, 23, 25, 33, 34, 36, 37, 42, 44 e 45	7, 13, 17, 18 e 45
3	59,329	6, 8, 10, 12, 29, 31, 40, 41, 48 e 50	8 e 29
4	129,654	43 e 49	1, 3, 5, 23, 25, 33, 36 e 44
5	-	27	6, 12, 31 e 50
6	-	-	41 e 48
7	-	-	34 e 42
8	-	-	10 e 40
9	-	-	27
10	-	-	49
11	-	-	43
12	-	-	37

⁽¹⁾Valor máximo da medida de dissimilaridade encontrado no conjunto das menores distâncias envolvendo cada progenitor.

O valor de α aumentou com a formação dos grupos, ou seja, o grupo 1 foi formado utilizando-se um valor de α menor que o utilizado para a formação do grupo 4, uma vez que a formação do grupo 4 não foi influenciada pelos genótipos integrantes do grupo 1.

Conclusões

1. Na utilização do método de Tocher seqüencial não existe influência dos genótipos já agrupados.
2. O valor de α , estimado após a formação de um novo grupo, possui proporcionalidade quanto à dissimilaridade existente entre os acessos dos grupos.
3. O agrupamento dos genótipos com maior dissimilaridade é realizado com maior eficácia pelo método de Tocher seqüencial.

Referências

CRUZ, C.D. **Programa GENES**: análise multivariada e simulação. Viçosa: UFV, 2006. 175p.

CRUZ, C.D.; CARNEIRO, P.C.S. **Modelos biométricos aplicados ao melhoramento genético**. 2.ed. Viçosa: UFV, 2006. 585p.

KARASAWA, M.; RODRIGUES, R.; SUDRÉ, C.P.; SILVA, M.P.; RIVA, E.M.; AMARAL JÚNIOR, A.T. Aplicação de métodos de agrupamento na quantificação da divergência genética entre acessos de tomateiro. **Horticultura Brasileira**, v.23, p.1000-1005, 2005.

MARCHIORO, V.S.; CARVALHO, F.I.F. de; OLIVEIRA, A.C. de; CRUZ, P.J.; LORENCETTI, C.; BENIN, G.; SILVA, J.A.G. da; SCHMIDT, D.A.M. Dissimilaridade genética entre genótipos de aveia. **Ciência e Agrotecnologia**, v.27, p.285-294, 2003.

OLIVEIRA, A.C.B. de; SEDIYAMA, M.A.N.; PEDROSA, M.W.; GARCIA, N.C.P.; GARCIA, S.L.R. Divergência genética e descarte de variáveis em alface cultivada sob sistema hidropônico. **Acta Scientiarum. Agronomy**, v.26, p.211-217, 2004.

SILVA, C.M.; GONÇALVES-VIDIGAL, M.C.; VIDIGAL FILHO, P.S.; SCAPIM, C.A.; DAROS, E.; SILVÉRIO, L. Genetic diversity among sugarcane clones (*Saccharum* spp.). **Acta Scientiarum. Agronomy**, v.27, p.315-319, 2005.

SOUZA, F.F.; QUEIRÓZ, M.A.; DIAS, R.S.C. Divergência genética em linhagens de melancia. **Horticultura Brasileira**, v.23, p.179-183, 2005.

VASCONCELOS, E.S. de; CRUZ, C.D.; BHERING, L.L. FERREIRA, A. Estratégias de amostragem e estabelecimento de coleções nucleares. **Pesquisa Agropecuária Brasileira**, v.42, p.507-514, 2007.

Recebido em 10 de julho de 2007 e aprovado em 31 de agosto de 2007