

IDENTIFICAÇÃO E CARACTERIZAÇÃO DE MICROSSATÉLITES PROVENIENTES DE SEQUÊNCIAS EXPRESSAS DO PROJETO BRASILEIRO DO GENOMA CAFÉ

Guilherme Ferreira Pena², Eveline Teixeira Caixeta³, Juliano Lino Ferreira⁴, Eunize Maciel Zambolim⁵, Laércio Zambolim⁶, Ney Sussumu Sakiyama⁷

¹ Trabalho financiado pelo Consórcio Brasileiro de Pesquisa e Desenvolvimento do Café – CBP&D/Café e FINEP

² Bolsista, Bs, Embrapa Café/Universidade Federal de Viçosa, Viçosa-MG, penabio@yahoo.com.br

³ Pesquisadora, D. Sc., Embrapa Café, Brasília-DF, eveline.caixeta@embrapa.br, autora para correspondência

⁴ Bolsista, D. Sc., Embrapa Café/Universidade Federal de Viçosa, Viçosa-MG, julianolf@yahoo.com.br

⁵ Pesquisadora, D. Sc., Universidade Federal de Viçosa, Viçosa-MG, eunize@ufv.br

⁶ Professor, D. Sc., Universidade Federal de Viçosa, Viçosa-MG, zambolim@ufv.br

⁷ Professor, D. Sc., Universidade Federal de Viçosa, Viçosa-MG, sakiyama@ufv.br

RESUMO: Dentre os marcadores genéticos disponíveis, os microssatélites ou SSRs (*Simple Sequence Repeats*) têm sido os escolhidos para diferentes estudos genéticos por apresentarem características que agregam simplicidade técnica, grande poder de resolução e alto nível de polimorfismo. Com a disponibilidade das seqüências ESTs (*Expressed Sequence Tags*), geradas pelo Projeto Brasileiro do Genoma Café, surgiu a oportunidade de desenvolver esses marcadores, de forma direta e eficiente, por meio de análise eletrônica. Visando o desenvolvimento futuro de marcadores SSRs para café, nesse trabalho foram identificados e caracterizados microssatélites nas seqüências expressas do genoma do cafeeiro. A mineração dos dados foi realizada utilizando-se todas as combinações de di-, tri- e tetranucleotídeos formadores dos microssatélites, perfazendo um total de 46 combinações (quatro de dinucleotídeos, 10 de trinucleotídeos e 32 de tetranucleotídeos). Foram considerados os microssatélites perfeitos e com tamanho mínimo de 12 pares de bases. Do total de 130.792 ESTs proveniente de *C. arabica* foram identificadas 37.826 contendo microssatélites, que após a clusterização resultou em 24.031 EST-SSR. Dentre estas, 45,11% apresentaram tetranucleotídeos, 36,67% trinucleotídeos e 18,23% dinucleotídeos. As unidades repetitivas mais abundantes dentro de cada categoria de EST-SSR foram (AG)_n encontrada em 57,08% das seqüências contendo dinucleotídeos, (AGG)_n com 30,79% dos trinucleotídeos e (AGGG)_n com 33,24% dos tetranucleotídeos. A grande quantidade de SSRs identificada nas seqüências transcritas do genoma do cafeeiro demonstra que essas são fontes valiosas para o desenvolvimento dos marcadores moleculares EST-SSRs.

Palavras-chave: SSR, EST-SSR, marcador molecular, *Coffea*.

IDENTIFICATION AND CHARACTERIZATION OF MICROSATELLITE FROM EXPRESSED SEQUENCES OF THE BRAZILIAN COFFEE GENOME PROJECT

ABSTRACT: Among the available genetic markers, the microsatellites or SSRs (Simple Sequence repeats) have been chosen for genetic studies because they present different characteristics that add technical simplicity, high resolution and high level of polymorphism. With the availability of Expressed Sequence Tags (EST), generated by the Brazilian Coffee Genome Project, it come the opportunity to develop these markers, in a directly and efficiently manner via electronic analysis. Aiming at the future development of SSRs markers for coffee, in this work, microsatellites were identified and characterized in expressed sequences of the coffee genome. The data mining was performed using all the combinations of di-, tri- and tetranucleotide microsatellite motifs, being a total of 46 combinations (four dinucleotide, 10 trinucleotide and 32 tetranucleotides). Of the total of 130,792 ESTs from *C. arabica*, 37,826 containing microsatellites were identified, which after clustering resulted in 24,031 EST-SSR. Among these, 45.11% had tetranucleotide, 36.67% trinucleotide and 18.23% dinucleotide motifs. The most abundant repetitive units within each category of EST-SSR were (AG)_n found in 57.08% of sequences containing dinucleotide, (AGG)_n in 30.79% of trinucleotides, and (AGGG)_n in 33.24% of tetranucleotides. A large number of SSRs identified in the transcribed sequences of the genome of coffee shows that these are valuable sources for the development of molecular markers EST-SSRs.

Key words: SSR, EST-SSR, molecular marker, *Coffea*.

INTRODUÇÃO

Dentre os marcadores de DNA, destaca-se o baseado em microssatélites ou SSR (*Simple Sequence Repeats*).

Este marcador tem sido largamente utilizado em diferentes culturas por apresentar alta reprodutibilidade, simplicidade técnica (PCR), custo relativamente baixo, grande poder de resolução, co-dominância e alto nível de polimorfismo (Yu et al., 2000). Apesar destas vantagens, poucos destes eficientes marcadores foram desenvolvidos para o café, principalmente para *Coffea arabica* (Poncet et al., 2006; Aggarwal et al., 2007).

Marcadores microssatélites derivado de DNA genômico têm sido extensivamente usado ao longo da última década, por causa do alto interesse nas suas propriedades (Ellegren, 2004). Entretanto, eles correspondem a sequências não codantes e assim não facilitam a identificação de genes de interesse. Com o aumento de ESTs (*Expressed Sequence Tags*) disponíveis, surge uma fonte complementar para identificação de microssatélites. As ESTs são fragmentos sequenciados do DNA complementar ao RNA mensageiro e representam parte de uma região transcrita do genoma em determinadas condições fisiológicas. Dessa forma, marcadores SSRs derivados de ESTs têm alta probabilidade de estarem relacionados a porções funcionais do genoma, apesar de a natureza conservada das sequências codantes poder limitar o seu polimorfismo (Varshney et al., 2005).

Em novembro de 2002, iniciou o Projeto Brasileiro do Genoma Café, desenvolvido através de um acordo entre o Consórcio Brasileiro de Pesquisa e Desenvolvimento do Café (CBP&D-Café), a Empresa Brasileira de Pesquisa Agropecuária (Embrapa), a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) e o Fórum Permanente para Relações entre Companhias Universitárias (UNIEMP) (Vieira et al., 2006). Esse Projeto resultou no sequenciamento do genoma das espécies *C. arabica*, *C. canephora*, e *C. racemosa*, a partir de EST, ou seja, de sequências transcritas. Os dados disponibilizados pelo Projeto Genoma fornecem a oportunidade para o desenvolvimento dos marcadores SSR de forma relativamente rápida e eficiente, por meio da análise das sequências do banco de dados, utilizando-se recursos computacionais.

Visando o desenvolvimento futuro de marcadores SSRs para café, *Coffea arabica* L., nesse trabalho foram identificados e caracterizados SSR a partir de sequências ESTs disponibilizadas pelo Projeto Brasileiro do Genoma Café.

MATERIAL E MÉTODOS

Este trabalho foi realizado no Laboratório de Biotecnologia do Cafeeiro, localizado no Instituto de Biotecnologia Aplicado à Agropecuária (BIOAGRO) da Universidade Federal de Viçosa, Minas Gerais.

As sequências ESTs utilizadas foram disponibilizadas pelo Projeto Brasileiro do Genoma Café. Estas sequências foram geradas a partir de um sequenciamento em larga escala de cDNAs que envolveu a participação de 25 laboratórios pertencentes a Universidades Públicas e Instituições de Pesquisa. No Projeto Brasileiro do Genoma Café foram sequenciados 214.964 clones provenientes de três espécies *C. arabica*, *C. canephora* e *C. racemosa*, resultando em 130.792, 12.381 e 10.566 sequências de cada espécie, respectivamente, após processo de trimagem (Vieira et al., 2006). No presente estudo, só foram analisadas as ESTs de *C. arabica*.

Para a identificação dos SSRs nas ESTs do Projeto Brasileiro do Genoma Café, foram consideradas todas as combinações de repetições com dois, três e quatro nucleotídeos, totalizando 46 classes (quatro dinucleotídeos, 10 e trinucleotídeos e 32 tetranucleotídeos). Para obtenção dessas classes foram eliminadas as sobreposições e complementaridade de bases. Foi realizada uma mineração, utilizando como critério microssatélites que possuíssem um mínimo de 12 pares de bases (pb), independente do tamanho da unidade repetitiva. Nessa mineração, foram considerados apenas os SSRs perfeitos, ou seja, aqueles que não apresentam nenhuma interrupção.

A mineração dos SSRs foi realizada a partir das informações disponíveis na plataforma de bioinformática do Genoma Café, desenvolvido pelo laboratório de Genoma e Expressão (LGE), Unicamp (www.lge.ibi.unicamp.br/cafe). Usando o aplicativo *Gene Project* foram criados projetos correspondentes a cada uma das classes de SSRs. Os projetos correspondem a arquivos virtuais contendo as ESTs de interesse. Uma vez designado o projeto, utilizou-se a opção *Pattern Search* como ferramenta que permitiu a identificação de todas as EST que continham a classe de SSRs correspondente.

Para cada projeto, as sequências mineradas que continham SSRs foram clusterizadas. Este processo as agrupou em *contigs* com a finalidade de aumentar o tamanho da sequência a ser analisada e reduzir a redundância, ou seja, impedir que futuramente sejam desenhados *primers* que amplifiquem a mesma região do genoma.

As sequências contendo os SSRs, previamente identificados, foram analisadas utilizando-se o programa *Gramene* (www.gramene.org/db/searches/ssrtool). Esta análise permitiu confirmar e localizar os SSRs nas sequências, bem como verificar a presença de mais de um SSR na mesma sequência. Após a identificação, os SSRs foram caracterizados. Neste estudo, somente foram consideradas as repetições perfeitas.

RESULTADOS E DISCUSSÃO

Foram mineradas do banco de dados do Projeto Brasileiro do Genoma Café 37.826 ESTs de *C. arabica* contendo microssatélites, que após a clusterização e, consequentemente redução da redundância, resultaram em 24.031 EST-SSRs. Esses resultados demonstraram que 18,37% das ESTs apresentaram microssatélites.

Os dados da mineração estão apresentados na Tabela 1. Para cada classe de EST-SSR foi mostrado o número total de ESTs contendo SSR, o número de *contigs* e *singlets* formados após o agrupamento das ESTs e a frequência (%) da classe considerando todas as unidades repetitivas. Além disso, foram obtidas as frequências de cada classe dentro dos di-, tri- e tetranucleotídeos.

Tabela 1. Número de clones minerados, a partir de ESTs do Projeto Brasileiro do Genoma Café, contendo repetições de di-, tri- e tetranucleotídeos, com tamanho mínimo de 12 pares de bases.

| Dinucleotídeos (6x) | | | | Classes | ESTs | Contigs | Singlets | Total ¹ | Freq. 1(%) ² | Freq. 2(%) ³ |
|------------------------|-----------|---------|-----------|---------|-------|---------|----------|--------------------|-------------------------|-------------------------|
| AC/TG | CA/GT | | | (AC)n | 976 | 108 | 603 | 711 | 16,23 | 2,96 |
| AG/TC | GA/CT | | | (AG)n | 3376 | 307 | 2193 | 2500 | 57,08 | 10,4 |
| AT | TA | | | (AT)n | 1419 | 177 | 617 | 794 | 18,13 | 3,3 |
| CG | GC | | | (CG)n | 388 | 9 | 366 | 375 | 8,56 | 1,56 |
| | | | | Total | 6159 | 601 | 3779 | 4380 | 100 | 18,23 |
| Trinucleotídeos (4x) | | | | Classes | ESTs | Contigs | Singlets | Total | Freq. 1(%) | Freq. 2(%) |
| AAC/TTG | ACA/TGT | CAA/GTT | | (AAC)n | 553 | 69 | 318 | 387 | 4,39 | 1,61 |
| AAG/TTC | AGA/TCT | GAA/CTT | | (AAG)n | 3203 | 411 | 1480 | 1891 | 21,46 | 7,87 |
| AAT/TTA | ATA/TAT | TAA/ATT | | (AAT)n | 863 | 108 | 443 | 551 | 6,25 | 2,29 |
| ACC/TGG | CAC/GTG | CCA/GGT | | (ACC)n | 2664 | 256 | 571 | 827 | 9,39 | 3,44 |
| ACG/TGC | GAC/CTG | CGA/GCT | | (ACG)n | 1201 | 202 | 212 | 414 | 4,7 | 1,72 |
| ACT/TGA | CTA/GAT | TAC/ATG | | (ACT)n | 1600 | 206 | 326 | 532 | 6,04 | 2,21 |
| AGC/TCG | CAG/GTC | GCA/CGT | | (AGC)n | 1730 | 228 | 262 | 490 | 5,56 | 2,04 |
| AGG/TCC | GAG/CTC | GGA/CCT | | (AGG)n | 3443 | 265 | 2448 | 2713 | 30,79 | 11,29 |
| AGT/TCA | GTA/CAT | TAG/ATC | | (AGT)n | 1040 | 152 | 255 | 407 | 4,62 | 1,69 |
| CCG/GGC | CGC/GCG | GCC/CGG | | (CCG)n | 1221 | 137 | 462 | 599 | 6,8 | 2,49 |
| | | | | Total | 17518 | 2034 | 6777 | 8811 | 100 | 36,67 |
| Tetranucleotídeos (3x) | | | | Classes | ESTs | Contigs | Singlets | Total | Freq. 1(%) | Freq. 2(%) |
| AAAC/TTTG | AACA/TTGT | ACAA/TG | CAAA/GTTT | (AAAC)n | 710 | 54 | 423 | 477 | 4,4 | 1,98 |
| AAAG/TTTC | AAGA/TTCA | AGAA/TC | GAAA/CTTT | (AAAG)n | 2440 | 173 | 1673 | 1846 | 17,03 | 7,68 |
| AAAT/TTTA | AATA/TTAT | ATAA/TA | TAAA/ATTT | (AAAT)n | 1520 | 101 | 586 | 687 | 6,34 | 2,86 |
| AACC/TTGG | ACCA/TGGT | CCAA/GG | CAAC/GTTG | (AACC)n | 254 | 26 | 144 | 170 | 1,57 | 0,71 |
| AACG/TTGC | AGCA/TCGT | CGAA/GC | GAAC/CTTG | (AACG)n | 260 | 23 | 56 | 79 | 0,73 | 0,33 |
| AACT/TTGC | ACTA/TGAT | CTAA/GA | TAAC/ATTG | (AACT)n | 253 | 27 | 65 | 92 | 0,85 | 0,38 |
| AAGC/TTCG | AGCA/TCGT | GCAA/CG | CAAG/GTTC | (AAGC)n | 439 | 24 | 375 | 399 | 3,68 | 1,66 |
| AAGG/TTCC | AGGA/TCCT | GGAA/CC | GAAG/CTTC | (AAGG)n | 1111 | 58 | 851 | 909 | 8,39 | 3,78 |
| AAGT/TTCA | AGTA/TCAT | GTAA/CA | TAAG/ATTC | (AAGT)n | 91 | 16 | 48 | 64 | 0,59 | 0,27 |
| AATC/TTCC | ATCA/TAGT | TCAA/AG | CAAT/GTTA | (AATC)n | 93 | 15 | 53 | 68 | 0,63 | 0,28 |
| AATG/TTAC | ATGA/TACT | TGAA/AC | GAAT/CTTA | (AATG)n | 97 | 10 | 62 | 72 | 0,66 | 0,3 |
| AATT/TTAA | ATTA/TAAT | | | (AATT)n | 160 | 24 | 79 | 103 | 0,95 | 0,43 |
| ACAG/TGTC | CAGA/GTCT | AGAC/TC | GACA/CTGT | (ACAG)n | 112 | 10 | 67 | 77 | 0,71 | 0,32 |
| ACAT/TGTA | CATA/GTAT | ATAC/TA | TACA/ATGT | (ACAT)n | 250 | 25 | 143 | 168 | 1,55 | 0,7 |
| ACCC/TGGG | CACC/GTGG | CCAC/GG | CCCA/GGGT | (ACCC)n | 542 | 6 | 509 | 515 | 4,75 | 2,14 |
| ACCG/TGGC | CCGA/GGCT | CGAC/GC | GACC/CTGG | (ACCG)n | 68 | 7 | 29 | 36 | 0,33 | 0,15 |
| ACCT/TGGA | CCTA/GGAT | CTAC/GA | TACC/ATGG | (ACCT)n | 180 | 22 | 65 | 87 | 0,8 | 0,36 |
| ACGC/TGCG | CGCA/GCGT | GCAC/CG | CACG/GTGC | (ACGC)n | 94 | 3 | 85 | 88 | 0,81 | 0,37 |
| ACGG/TGCC | CGGA/GCCT | GGAC/CC | GACG/CTGC | (ACGG)n | 47 | 4 | 34 | 38 | 0,35 | 0,16 |
| ACGT/TGCA | CGAT/GCAT | GTAC/CA | TACG/ATGC | (ACGT)n | 130 | 25 | 49 | 74 | 0,68 | 0,31 |
| ACTC/TGAC | CTCA/GAGT | TCAC/AG | CACT/GTGA | (ACTC)n | 195 | 12 | 116 | 128 | 1,18 | 0,53 |
| ACTG/TGAC | CTGA/GACT | | | (ACTG)n | 26 | 3 | 20 | 23 | 0,21 | 0,1 |
| AGAT/TCTA | GATA/CTAT | ATAG/TA | TAGA/ATCT | (AGAT)n | 205 | 26 | 97 | 123 | 1,13 | 0,51 |
| AGCC/TCGG | GCCA/CGGT | CCAG/GG | CAGC/GTCG | (AGCC)n | 57 | 9 | 26 | 35 | 0,32 | 0,15 |
| AGCG/TCGC | GCGA/CGCT | CGAG/GC | GAGC/CTCG | (AGCG)n | 106 | 6 | 80 | 86 | 0,79 | 0,36 |
| AGCT/TCGA | GCTA/CGAT | CTAG/GA | TAGC/ATCC | (AGCT)n | 114 | 21 | 23 | 44 | 0,41 | 0,18 |
| AGGC/TCCG | GGCA/CCGT | GCAG/CG | CAGG/GTCC | (AGGC)n | 68 | 9 | 27 | 36 | 0,33 | 0,15 |
| AGGG/TCCC | GGGA/CCCT | GGAG/CC | GAGG/CTCC | (AGGG)n | 3696 | 38 | 3565 | 3603 | 33,24 | 14,99 |
| AGGT/TCCA | GGTA/CCAT | GTCG/CA | TACC/ATCC | (AGGT)n | 131 | 19 | 55 | 74 | 0,68 | 0,31 |
| AGTC/TCAG | GTCA/CAGT | | | (AGTC)n | 29 | 6 | 10 | 16 | 0,15 | 0,07 |
| CCCG/GGGC | CCGC/GGCG | CGCC/GC | GCCC/CGGG | (CCCG)n | 553 | 11 | 516 | 527 | 4,86 | 2,19 |
| CCGG/GGCC | CGGC/GCCG | | | (CCGG)n | 118 | 12 | 84 | 96 | 0,89 | 0,4 |
| | | | | Total | 14149 | 825 | 10015 | 10840 | 100 | 45,11 |
| TOTAL | | | | | 37826 | | | 24031 | | |

¹ Corresponde a soma dos *contigs* e *singlets*

² Freq. 1(%) - porcentagem de cada classe de EST-SSR, dentro de cada unidade repetitiva (di-, tri- ou tetranucleotídeos)

³ Freq. 2(%) - porcentagem de cada classe de EST-SSR, em relação a todas as unidades repetitivas (di-, tri- e tetranucleotídeos)

Considerando todas as 46 classes microssatélites analisadas, (AGGG)_n foi a mais abundante, sendo encontrada em 14,99% das sequências mineradas, seguida pelas classes (AGG)_n (11,29%) e (AG)_n (10,40%). As demais ficaram com frequência abaixo de 10,00%.

Dentre os dinucleotídeos, foi obtido um total de 4.380 EST-SSRs (Tabela 1) e a classe (AG)_n foi a mais abundante, correspondendo a 57,08% (307 contigs e 2.193 singlets) dos dinucleotídeos identificados. Em estudos com outras espécies de plantas, essa classe também foi identificada como a mais frequente (Scott et al., 2000; Kantety et al., 2002; Morgante et al., 2002; Gao et al., 2003; Thiel et al., 2003; Saha et al., 2004; Jung et al., 2005). As classes (AT)_n e (AC)_n foram encontradas, respectivamente, em 18,13% e 16,23% das sequências contendo dinucleotídeos. A classe menos abundante foi (CG)_n que apresentou frequência de 8,56%. Poncet et al. (2006), também analisaram os microssatélites em café, no entanto na espécie *C. canephora*, e encontraram resultados semelhantes. Observaram que a classe (AG)_n também foi a mais abundante entre os dinucleotídeos, ocorrendo em 62,00% das sequências; e a classe (CG)_n não foi encontrada, evidenciando a menor abundância dessa classe. Em outro trabalho com café, a classe (AG)_n apresentou frequência de 58,20%, seguido por (AT)_n com 24,8%, (AC)_n com 21,6% e (CG)_n com 1,1% (Aggarwal et al., 2007). Comparando os resultados do presente trabalho com estes dois estudos de caracterização de EST em café, observa-se uma concordância dos resultados, sendo sempre a classe (AG)_n a mais abundante e (CG)_n a menos frequente. Estudos realizados com soja (Gao et al. 2003), trigo (Nicot et al. 2004), *Arabidopsis thaliana* (Jung et al. 2005) e seringueira (Feng et al. 2009) também demonstram que o dinucleotídeo (CG)_n é raro em plantas.

Considerando os trinucleotídeos, foram identificados 8.811 EST-SSRs. Destes, a classe mais abundante foi a (AGG)_n, sendo identificada em 30,79% das sequências mineradas, seguida da (AAG)_n com 21,46%. A frequência das demais classes foi menor que 10,00% (Tabela 1). Esses resultados concordam com os obtidos por Poncet et al. (2006) para EST de *C. canephora*, cujos dados demonstraram que as classes mais abundantes dentro dos trinucleotídeos foram (AGG)_n com 23,00% e (AAG)_n com 20,30%. Para outras espécies, como em *Medicago truncatula* e videira, Eujayl et al. (2004) e Scott et al. (2000) obtiveram resultados que também concordam com as frequências obtidas no presente trabalho. Por outro lado, Aggarwal et al. (2007), analisando 2.553 ESTs de café, encontraram frequência de (AAG)_n de 28,80%, seguido de (ACT)_n com 12,60%, (ACC)_n com 11,70% e (AAT)_n com 10,80%.

Dentre os tetranucleotídeos, foram mineradas um total de 10.840 EST-SSRs, dos quais as classes mais frequentes foram (AGGG)_n com 33,24% e (AAAG)_n com 17,03%. Na Tabela 1 nota-se que das 32 classes de tetranucleotídeos, 20 apresentaram frequência inferior a 1,00%.

Na Figura 1 está apresentada a distribuição dos microssatélites, classificados de acordo com a unidade repetitiva, considerando o critério mínimo de 12 pares de bases. Observou-se uma predominância dos tetranucleotídeos (45,11%), seguido pelos trinucleotídeos (36,67%) e os dinucleotídeos (18,23%). Nos outros dois trabalhos de identificação de microssatélites em café foram encontradas frequências diferentes para cada categoria, sendo que no estudo de Poncet et al. (2006) os trinucleotídeos foram os mais abundantes e no de Aggarwal et al. (2007), os dinucleotídeos. Essa discrepância de dados pode ser explicada pela diferença de critérios usados para a mineração dos dados. Segundo Varshney et al. (2005), a frequência, a distribuição e a abundância de ESTs-SSR podem ser altamente variáveis, dependendo do critério de busca dessas sequências repetitivas, do tamanho do conjunto de dados e da ferramenta de busca utilizada na mineração da base de dados. Aggarwal et al. (2007) sugerem uma padronização universal do critério de busca de EST-SSR no genoma, com o intuito de facilitar comparações com os dados oriundos de outros estudos.

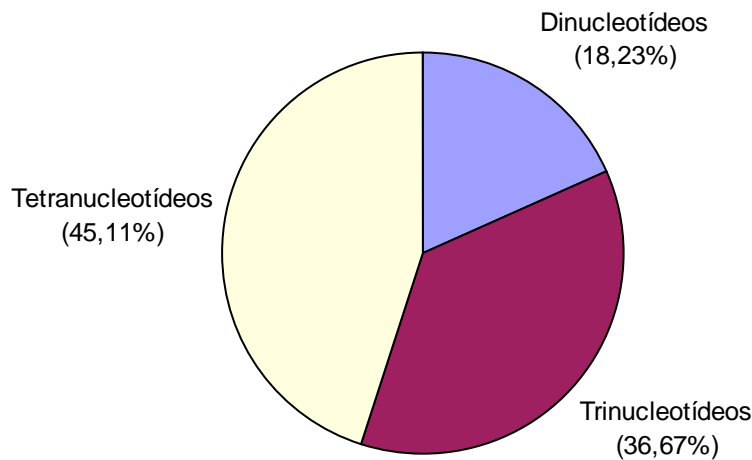


Figura 1. Comparação entre a ocorrência de EST-SSR de acordo com a unidade repetitiva.

CONCLUSÕES

- O genoma do café apresenta grande quantidade de SSRs em suas sequências transcritas, demonstrando que ESTs de café constituem em valioso recurso para o desenvolvimento dos marcadores moleculares EST-SSRs.

Considerando como critério de mineração os microssatélites perfeitos com di-, tri- e tetranucleotídeos de tamanho mínimo de 12pb, observou-se que os tetranucleotídeos foram os mais frequentes e a classe mais abundante foi (AGGG)_n. Dentro dos dinucleotídeos, a mais abundante foi (AG)_n e dos trinucleotídeos (AGG)_n.

REFERÊNCIAS BIBLIOGRÁFICAS

AGGARWAL, R.K.; HENDRE, P.S.; VARSHNEY, R.K.; BHAT, P.R.; KRISHNAKUMAR, V.; SINGH, L. Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analyses of coffee and related species. **Theoretical and Applied Genetics**, Berlin, v.114, n.5, p.359-372, 2007.

ELLEGREN, H. Microsatellites: simple sequences with complex evolution. **Nature Review Genetics**, London, v.5, n.6, p.435-445, 2004.

EUJAYL, I.; SLEDGE, M.K.; WANG, L.; MAY, G.D.; CHEKHOVSKIY, K.; ZWONITZER, J.C.; MIAN, M.A. *Medicago truncatula* EST-SSRs reveal cross-species genetic markers for *Medicago* spp. **Theoretical and Applied Genetics**, Berlin, v.108, n.3, p.414-422, 2004.

FENG, S.P.; LI, W.G.; HUANG, H.C.; WANG, J.Y.; WU, Y.T. Development, characterization and cross-species/genera transferability of EST-SSR markers for rubber tree (*Hevea brasiliensis*). **Molecular Breeding**, Lleida, v.23, n.1, p.85-97, 2009.

GAO, L.F.; TANG, J.F.; LI, H.W.; JIA, J.Z. Analysis of microsatellites in major crops assessed by computational and experimental approaches. **Molecular Breeding**, Lleida, v.12, n.3, p.245-261, 2003.

JUNG, S.; ALBERT, A.; JESUDURAI, C.; TOMKINS, J.; MAIN, D. Frequency, type, distribution and annotation of simple sequence repeats in Rosaceae ESTs. **Functional and Integrative Genomics**, Perth, v.5, n.3, p.136-143, 2005.

KANTETY, R.V.; LA ROTA, M.; MATTHEWS, D.E.; SORRELLS, M.E. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. **Plant Molecular Biology**, Zurich, v.48, p.501-510, 2002.

MORGANTE, M.; HANAFEY, M.; POWELL, W. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. **Nature Genetics**, London, v.30, p.194-200, 2002.

NICOT, N.; CHIQUET, V.; GANDON, B.; AMILHAT, L.; LEGEAI, F.; LEROY, P. et al. Study of simple sequence repeat (SSR) markers from wheat expressed sequence tags (ESTs). **Theoretical and Applied Genetics**, Berlin, v.109, n.4, p.800-805, 2004.

PONCET V., RONDEAU M., TRANCHANT C., CAYREL A., HAMON S., KOCHKO A., HAMON P. SSR mining in coffee tree EST databases: potential use of EST-SSRs as markers for the *Coffea* genus. **Molecular Genetics and Genomics**, Berlin, v.276, n.5, p.436-449, 2006.

SCOTT, K.D.; EGGLER, P.; SEATON, G.; ROSSETTO, M.; ABLETT, E.M.; LEE, L.S.; HENRY, R.J. Analysis of SSRs derived from grape ESTs. **Theoretical and Applied Genetics**, Berlin, v.100, n.5, p.723-726, 2000.

SAHA, M.C.; MIAN, M.A.; EUJAYL, I.; ZWONITZER, J.C.; WANG, L.; MAY, G.D. Tall fescue EST-SSR markers with transferability across several grass species. **Theoretical and Applied Genetics**, Berlin, v.109, n.4, p.783-791, 2004.

THIEL, T.; MICHALEK, W.; VARSHNEY, R.K.; GRANER, A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). **Theoretical and Applied Genetics**, Berlin, v.106, n.3, p.411-422, 2003.

VARSHNEY, R.K., SIGMUND, R., BORNER, A., KORZUN, V., STEIN, N., SORRELLS, M.E., LANGRIDGE, P., GRANER, A. Interspecific transferability and comparative mapping of barley EST-SSR markers in wheat, rye and rice. **Plant Science**, Limerick, v.168, n.1, p.195-202, 2005.

VIEIRA, L.G.E.; ANDRADE, A.C.; COLOMBO, C. A. et al. Brazilian coffee genome project: An EST-based genomic resource. **Brazilian Journal of Plant Physiology**, Londrina, v.18, n.1, p.95-108, 2006.

YU K, PARK SJ, POYSA V, GEPTS P. Integration of simple sequence repeat (SSR) markers into a molecular linkage map of common bean (*Phaseolus vulgaris* L.). **Journal of Heredity**, Saint Louis, v. 91, n.6, p.429-434, 2000.