SÃO GABRIEL

CIÊNCIAS BIOLÓGICAS

CARACTERIZAÇÃO DE SNPs PUTATIVOS ENTRE *Eugenia uniflora L.* E *Eucalyptus grandis* HILL EX MAIDEN.

NATÁLIA MENEZES CERQUEIRA

São Gabriel

2016

NATÁLIA MENEZES CERQUEIRA

CARACTERIZAÇÃO DE SNPs PUTATIVOS ENTRE *Eugenia uniflora* E *Eucalyptus grandis* HILL EX MAIDEN.

Monografia apresentada como exigência para obtenção do grau de Bacharelado em CIÊNCIAS BIOLÓGICAS da UNIVERSIDADE FEDERAL DO PAMPA.

Orientador: Valdir Marcos Stefenon

Co-orientadora: Deise Schröder Sarzi

São Gabriel

2016

CARACTERIZAÇÃO DE SNPs PUTATIVOS ENTRE *Eugenia uniflora* E *Eucalyptus grandis* HILL EX MAIDEN.

NATÁLIA MENEZES CERQUEIRA

Orientador: Valdir Marcos Stefenon

Co-orientadora: Deise Schröder Sarzi

Monografia submetida à Comissão de Avaliação do Trabalho de Conclusão do Curso de Ciências Biológicas, como parte dos requisitos necessários à obtenção do grau de Bacharel em Ciências Biológicas**.**

Aprovada por:

_____

Presidente, Prof. Dr. Valdir Marcos Stefenon

_____

Prof. Dr. Juliano Tomazzoni Boldo

_____

Biólogo Ms. Rafael Plá Matielo Lemos

AGRADECIMENTO

Primeiramente gostaria de agradecer a minha mãe, Giles por ter me apoiado durante todo curso, e que mesmo com as dificuldades que passamos durante esse tempo, sempre me incentivou a seguir em frente.

Ao meu pai, Francisco por tornar as minhas tropeçadas mais divertidas.

A minha irmã Gabriela por me ajudar todas as vezes que eu precisei.

A minhas avós Maria Helena e Marta que sempre estavam dispostas a me aconselhar, mesmo que muitas vezes eu não escutasse.

A minha bisavó Darcy que infelizmente não está mais entre nós, mas que tenho certeza que de onde ela estiver, está feliz por esta conquista.

A toda a minha família.

A todos os meus amigos por me ampararem e incentivarem sempre.

Ao meu orientador Valdir por estar ao meu lado, ajudando sempre que foi preciso.

A minha co-orientadora Deise, por me ensinar, ajudar e principalmente por ter paciência e me acalmar quando o desespero batia.

A todos os colegas do grupo NuGEM pelo apoio.

A UNIPAMPA pelo suporte.

O meu muito obrigada a todos!

“Para todos que tiveram um momento de fraqueza. Não vai doer para sempre, então não deixe isso afetar o que há de melhor em você.”

J. A. REDMERSKI

RESUMO

Eugenia uniflora (pitangueira) é uma espécie da família Myrtaceae bastante explorada economicamente devido a suas propriedades farmacológicas e, por este motivo, são importantes os estudos a nível molecular sobre esta espécie. Os *SNPs* tem sido uma ferramenta amplamente utilizada na área de genômica, assim como os *ESTs* que também apresentam diversas vantagens devido sua presença em diversos bancos de dados. O objetivo deste estudo foi buscar *SNPs* em regiões gênicas de *E. uniflora* em comparação com o banco de dados *ESTs* de *Eucalyptus grandis.* Para a realização deste trabalho foi feito o sequenciamento do DNA total de uma amostra de pitangueira, que gerou uma cobertura de aproximadamente 25% do genoma desta espécie*,* montados em torno de 2.600 *contigs*. Destes, selecionamos 150 para fazer a busca de regiões gênicas hipotéticas por meio da comparação do banco de dados de *ESTs* de *E. grandis.* Após, foram selecionados somente os *contigs* que continham cobertura mínima de 80pb. Dos 150 *contigs* analisados, 35 obtiveram similaridade de *ESTs* entre as duas espécies. Após a retirada dos íntrons, alguns *contigs* mostraram alinhamentos fragmentados e outros mostraram uma cobertura inferior a 80pb. Do total analisado, 8 *contigs* atenderam todas as qualidades exigidas. O *contig* que mostrou uma menor frequência (*SNPs*/pb), apresentou semelhança com a sequência que expressa a proteína Eto1, o *contig* que apresentou um valor intermediário em relação ao máximo e mínimo da frequência obteve semelhança na proteína *RPK*, já o *contig* que apresentou maior frequência mostrou semelhança com a proteína *AGP*. O número de transições foi maior que o de tranversões. Estes dados foram classificados como baixos, indicando que estas regiões gênicas são bem conservadas. A técnica utilizada é de fácil aplicação e muito útil mesmo sendo pouco utilizada visando à identificação de SNPs putativos entre espécies diferentes.

Palavras-chave: Pitanga; Eucalipto; BLAST; *Single Nucleotide Polymorphism.*

ABSTRACT

Eugenia uniflora (Surinam cherry), is a species of the Myrtaceae family, quite economically exploited due to their pharmacological properties. Therefore, studies at the molecular level on this species are important. SNPs have been a widely used tool in the area of genomics, as well as ESTs also offer several advantages due to its presence in various data banks. The aim of this study was to search for SNPs in gene regions of *E. uniflora* compared to the ESTs database of *E. grandis*. For this study, total DNA of *E. uniflora* was sequenced, generating a coveragem of about 25% of the species genome, in about 2,600 contigs. Out of these, 150 were selected to search for hypothetical gene regions, by comparing to the EST database of *Eucalyptus grandis*. Only contigs containing minimum coverage 80pb were selected. Out of the 150 contigs analyzed, 35 ESTs were obtained with similarity between the two species. After the removal of introns, some contigs showed fragmented alignments and others showed a lower coverage 80pb. Of the total analyzed, 8 contigs met all the required qualities. The contig that showed a lower frequency (SNPs / pb) showed similarity to the sequence that expresses the Eto1 protein. The contig that presented a intermediary value regarding the maximum and minimum frequency revealed similarity with the RPK protein, while the contig that presented higher SNPs freqeuncy showed similarity with the AGP protein. The number of transitions was higher than the tranversions. These data were classified as low, and indicate that these gene regions are well conserved. The technique is easy to use and very useful even if it is little used.

Keywords: Pitanga; Eucalyptus; BLAST; Single Nucleotide Polymorphism

LISTA DE TABELAS

SUMÁRIO

Conteúdo

1 INTRODUÇÃO

A família Myrtaceae compreende aproximadamente 100 gêneros e 3.500 espécies distribuídas principalmente em regiões tropicais e subtropicais do mundo, com centros de diversidade na América tropical e Austrália e poucas espécies ocorrendo nas regiões temperadas (BARROSO, *et. a.l,* 1984).

Esta é uma das famílias mais importantes do Brasil (LANDRUM E KAWASAKI 1997) destacando-se, com mais de uma centena de espécies, os gêneros *Eugenia, Myrcia* e *Calyptranthes*, enquanto o restante dos gêneros possui menos de 60 espécies brasileiras (LANDRUM E KAWASAKI 1997).

O perfil químico da família Myrtaceae é bem conhecido e caracteriza-se pela presença de taninos, flavonoides, mono e sesquiterpenos, triterpenos e caracteristicamente derivados do floroglucinol (CRUZ E KAPLAN, 2005).

Eugenia uniflora, popularmente conhecida no Brasil como pitangueira, é uma espécie arbórea da família Myrtaceae, nativa da mata atlântica e com capacidade de adaptação a diversas condições de solo e clima (ALMEIDA *et al*., 2012). Devido a esta adaptabilidade, a pitangueira foi disseminada e é atualmente cultivada nas mais variadas regiões do globo (BEZERRA *et al.*, 2000).

Ela também é considerada uma espécie pioneira, pois é uma das espécies utilizadas em reflorestamentos heterogêneos destinados à recomposição e recuperação de áreas degradadas (BAGETTI, *et. al.,* 2009).

Devido às propriedades benéficas à saúde atribuídas às folhas, pesquisadores começaram a estudar mais profundamente esta espécie, pois existe uma enorme utilização na medicina popular para inúmeras desordens (ADEBAJO *et al.*, 1989). Dentre estes benefícios podemos ressaltar a atividade anti-inflamatória, diurética, hipotensora, inibidora do aumento da glicose e de triglicerídeos séricos (MATSUMURA *et. al,* 2000).

Nas últimas décadas, esta espécie tem sido amplamente explorada pelas indústrias alimentícia, farmacêutica e cosmética, devido ao seu grande potencial econômico e a tendência na atualidade é de um crescimento no cultivo e exploração desta espécie (ALMEIDA *et al*., 2012). De modo geral, assim como do ponto de vista econômico, a pitangueira é muito apreciada pela população devido à produção de saborosos frutos carnosos do tipo baga.

Com perspectivas futuras de uma exploração comercial em grande escala, é importante um maior estudo e um maior conhecimento dos dados disponíveis até o momento que acercam a espécie (ALMEIDA *et al*., 2012).

Nos últimos anos, novos marcadores tem surgido como uma importante ferramenta em genômica, denominados *Single Nucleotide Polymorphism* (*SNP*) (TABASSUM e SUMAN, 2006), que são a mudança de um único nucleotídeo numa mesma posição de molécula de DNA entre indivíduos, diferindo de mudanças em múltiplas bases em posições aleatórias (VIDAL *et al*, 2007). Conceitualmente*, SNPs* correspondem a posições onde ocorrem duas bases alternativas em locais do genoma que são altamente conservados (SANTORO, 2010).

Os *SNPs* são cada vez mais utilizados como marcadores moleculares para diversas aplicações (TABASSUM E SUMAN, 2006), pois são distribuídos por todo o genoma. Além do mais, um SNP localizado na região codificadora pode ter impacto relevante na formação e na atividade de uma proteína. Um *SNP intrônico* pode influenciar o *splicing* do mRNA (KRAWEZAK *et al.,* 1992), assim como, um *SNP* na região promotora pode influenciar a expressão gênica (DRAZEN *et al.,* 1999).

Os *SNPs* são menos mutáveis em comparação com outros marcadores, como os microssatélites. A baixa taxa de mutação recorrente os torna evolutivamente estáveis, por este motivo eles são excelentes marcadores para o estudo de características genéticas complexas e para a melhor clareza da evolução genômica (TABASSUM E SUMAN, 2006).

Estes *SNPs* podem ser responsáveis por importantes variações e modificações nas características fenotípicas entre indivíduos de uma mesma espécie (EMAHAZION *et al.,* 2001) ou na modificação em locais altamente conservados na estrutura dos genes em espécies semelhantes, como é o caso da pitangueira e do *Eucalyptus grandis*, que tem uma grande proximidade evolutiva, sendo que o último possui seu genoma totalmente sequenciado e é facilmente encontrado em bancos de dados.

É importante salientar que a frequência de *SNPs* é heterogênea ao longo do genoma, diferindo entre regiões codificadoras e regiões não codificadoras. Em geral, os *SNPs* são menos frequentes em regiões gênicas que codificam alguma proteína do que em outras regiões (FLADUNG E BUSCHBOM, 2009).

Os *SNPs* podem ser classificados em dois tipos: as não sinônimas, que são aquelas que resultam na modificação de aminoácidos de uma proteína; ou sinônimas que não causam a mudança de aminoácido (EMARA E KIM 2003). No entanto, um *SNP* sinônimo pode modificar a estrutura e a estabilidade do RNA mensageiro e, consequentemente, afetar a quantidade de proteína produzida (GRIFFITHS *et al*., 2001).

A grande quantidade de *SNPs* encontradas no genoma vem estimulando a sua utilização em pesquisas e é aplicada em diferentes espécies, tanto animal como vegetal. *SNPs* têm sido utilizados com sucesso, por exemplo, na construção de mapas genéticos de alta densidade, caracterizados por uma ocorrência no genoma muito além da capacidade dos marcadores tradicionais (HYTEN *et al.,* 2010)

Alguns métodos de detecção de *SNPs* buscam por diferenças individuais em uma sequência através de análises computacional (BUETOW *et al.,* 1999; PICOULT-NEWBERG *et al.,* 1999).

Expressed Sequence Tags (*ESTs*) oferecem algumas vantagens, como a detecção de variação na porção expressa do genoma. Assim, a utilização de *ESTs* disponíveis em bancos de dados implica na redução dos custos de geração de novos dados (KANTETY *et al.,* 2002), como é o caso de sua utilização para busca de regiões gênicas putativas.

Nas áreas de filogenética, a importância do uso de novas sequências nucleares complementares tem sido proposta como uma alternativa para melhorar a resolução das relações filogenéticas (Qiu et al, 1999; Soltis et al 1999; Small et al., 2004). Além do que, os atributos como a elevada taxa de evolução sequencial, a existência de múltiplos locus independentes e a heranças de genes nucleares entre parentes fazem desta uma alternativa muito interessante para estabelecer espécies arbóreas (Small et al., 2004). Visando isto, uma investigação filogenética com base em genes nucleares é necessário começar por selecionar genes para um estudo preliminar (Small et al., 2004).

Visando identificar regiões *ESTs* com potencial filogenético dentro da família Myrtaceae, o objetivo deste trabalho foi mapear possíveis *SNPs* que ocorrem em regiões gênicas putativas da pitangueira (*E. uniflora*) em comparação com o banco de dados de *ESTs* de *Eucalyptus grandis*, testando a hipótese de que estas mutações tenham correlação com eventos evolutivos entre estas duas espécies.

2 METODOLOGIA

2.1 Obtenção das sequências analisadas

As sequências analisadas neste estudo foram gentilmente cedidas pela mestranda de PPGCB Deise Schröder Sarzi antes de sua submissão ao GeneBank (NCBI). A obtenção das sequências é resumidamente descrita a seguir:

Primeiramente, foi coletada uma amostra de *Eugenia uniflora* de uma população natural, localizada no campus São Gabriel da Universidade Federal do Pampa. O local foi devidamente marcado no GPS e o voucher depositado no herbário Bruno Edgar Irgang da mesma universidade, sob o número HBEI1150.

O DNA total do indivíduo amostrado foi extraído utilizando-se o kit DNAeasy plant mini kit da Quiagen. Esta amostra então foi sequenciada através do Sequenciador de nova Geração (NGS) *Ion Torrent*™ Personal Genome Machine® (*PGM*).

2.2 Montagem de contigs

Com o programa Velvet v. 1.2.10 (ZERBINO E BIRNEY, 2008), realizamos a montagem dos *reads* (sequências curtas de aproximadamente 300 pares de base) resultantes do sequenciamento, transformando-os em *contigs*, que são clones de sobreposição que formam um mapa físico do genoma que é usado para guiar a montagem do DNA. No total foram formados 2.600 *contigs* com cerca de 1.000 a 3.000 pares de bases cada um.

2.3 Comparação com o banco de dados EST

Neste trabalho, foram analisados 150 *contigs* com o propósito de busca de regiões gênicas putativas. Estas buscas foram feitas através da comparação das sequências com o banco de dados de *ESTs* (*Expressed Sequence Tags*) de *Eucalyptus grandis* por meio da ferramenta de alinhamento local BLAST (*Basic Local Alignment Search Tool*) (ALTSHUL, 1997), no site da *National Center for Biotechnology Information* (*NCBI*).

Após a comparação com o banco de dados, foram selecionados somente os *contigs* que apresentavam cobertura mínima de 80 bases idênticas. O número mínimo de 80 pares de base é utilizado como padrão mínimo para regiões gênicas, pois os Eucariotos em geral apresentam, normalmente, regiões gênicas de cerca de 90 a 120 pares de bases, regiões maiores e menores podem também ocorrer, porém com menor frequência. (ZAHA, 2003)

2.4 Retirada de *íntrons* e busca de *SNPs*

Os *contigs* com as especificações já citadas foram analisadas no *site* ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html), onde pudemos fazer a retirada dos *íntrons* destas sequências (regiões não codificantes do DNA), deixando somente os *éxons* (regiões codificantes do DNA e que sintetizam alguma proteína).

Tendo agora estas sequências diminuído de tamanho por possuírem somente a porção codificante das mesmas, novos alinhamentos foram realizados com a ferramenta BLAST, através do *software* MEGA7 (KUMAR E TAMURA, 2016), contra o banco de dados de todas as sequências de *E. grandis*, avaliados somente os *contigs* com cobertura mínima de 80 pares de bases (pb).

Posteriormente estas sequências foram alinhadas através da ferramenta Muscle que também se encontra no *software* MEGA7. Esta, então alinhou a sequência das duas espécies e evidenciou os *SNPs* presentes. Os parâmetros do alinhamento foram modificados para que se igualassem aos do BLAST. Os parâmetros modificados foram: *Gap Open* -400 e *Gap Extend* -100, assim em comparação com *E. grandis* pode se observar *SNPs* entre as duas espécies fazendo a contagem e verificação das bases modificadas.

3 RESULTADOS E DISCUSSÃO

3.1. Extração de DNA e Sequenciamento

O sequenciamento gerou aproximadamente 7,0 milhões de bases, que depois da montagem e eliminação de regiões redundantes, corresponde a cerca de 3,15 milhões de bases, resultando em uma cobertura de cerca de 25% do genoma de *E. uniflora.*

3.2. Busca por regiões gênicas putativas

Durante a busca de *ESTs* dos 150 *contigs* analisados, obteve-se somente 35 que mostraram similaridade com a espécie *Eucalyptus grandis, p*ois grande parte do DNA é composto por sequências repetitivas, e embora seja tentador se referir a estas sequências como DNA “lixo”, a manutenção estável dessassequências durante milhares de gerações sugere que o DNA intergênico atribui algum valor positivo (vantagem seletiva) ao organismo (WATSON, 2015)

3.3. Análise dos *SNPs* em regiões codificantes putativas

Dos 35 contigs analisados 18 mostraram baixa cobertura (abaixo de 80 pb) e identidade menor que 70% quando comparadas com o banco de dados de *E. grandis* e outras nove apresentaram alinhamento muito fragmentado, resultando em um total de oito amostras que continham um mínimo de 80 pares de bases, assim estas foram enumeradas de 1 a 8.

Dos 8 *contigs* analisados, a maior frequência de *SNPs* putativos se encontra no *contig* correspondente ao número 7 onde a frequência foi de um a cada 7,53 pares de bases, a mais alta frequência dentre os contigs analisados. Já o *contig* 1 demonstrou a frequência mais baixa em relação aos demais. O *contig* 5 apresentou uma frequência intermediaria entre os *contigs* analisados (TABELA 1).

Tabela 1. Frequências de *SNPs* por pares de bases na comparação de ESTs de *Eugenia uniflora* e *Eucalyptus grandis*.

N° do *contig*	**N° total de pb**	**N° total de SNPs**	**Frequência dos *SNPs***	**Proteínas putativas**
1	313	13	1 a cada 24,08 pb	ETO 1
2	204	19	1 a cada 10,78 pb	Remorin-*like*
3	93	9	1 a cada 10,33 pb	Ubiquitina
4	365	44	1 a cada 8,29 pb	IQ Domain
5	277	22	1 a cada 12,60 pb	LRR-RPK
6	123	14	1 a cada 8,78 pb	Inositol
7	433	56	1 a cada 7,53 pb	AGP
8	49	5	1 a cada 9,80 pb	Adenosina quinase

O contig 8 mostrou uma cobertura abaixo de 80pb, porém demonstrou uma identidade de 90%, sendo assim foi incluído nas análises.

Realizamos uma busca mais aprofundada em relação as possíveis proteínas codificadas para cada *contig*.

O *contig* 1 apresentou semelhança na sequência que pode expressar a proteína putativa ETO1 *like*. Esta é uma proteína que regula negativamente as atividades do gene ACS e a produção de etileno. Esta proteína interage diretamente com a inibição e a atividade da enzima ACS5, da enzima, resultando numa acumulação significativa de proteína ACS5 e etileno (WANG *et al.,* 2004). A super expressão de ETO1 inibe a indução da produção de acetato no regulador de

citicinina e assim no crescimento da planta, e promove a degradação ACS5. Assim ela demostra um mecanismo duplo, inibindo a atividade da enzima ACS encaminhando para a degradação de proteínas. Isto permite a rápida modulação da concentração do etileno.

O *contig* 2 apresentou a proteínas Remorin, elas não contêm domínios transmembranares, até agora têm sido detectados quase que exclusivamente em frações insolúveis em detergente membrana (comumente chamadas balsas lipídicas) (LEFEBVREA *et al*., 2009). Plantas exigem circuitos regulatórios altamente eficientes para modular cascatas de transdução de sinal celular. Planta com proteínas específicas remorin, podem atuar como suportes moleculares regulando a transdução de sinal, foram descritos como sendo altamente fosforilada e para associar com compartimentos de sinalização na membrana plasmática. Estas proteínas evoluíram em todas as plantas (MARÍN *et al.,* 2012).

O *Contig* 3 foi encontrada a proteína Ubiquitina que em células eucarióticas é utilizada na degradação da maioria das proteínas indesejadas (por exemplo proteínas mal-dobradas) utilizando o sistema ubiquitina proteassoma (GOLDBERG, 2003). Além da função proteolítica, a ubiquitinação tem sido associada a vários processos celulares como endocitose, transdução de sinal, controle da transcrição gênica, reparo do DNA e replicação do DNA (HAGLUND e DIKIC, 2005). Nas plantas, há relatos de transcritos de genes pertencentes à via ubiquitina acumulados durante a senescência foliar e durante o ataque de patógenos (BUCHANAN-WOLLASTON, 1997; DEVOTO et al., 2003).

O contig 4 encontrou o Domínio IQ, as funções deste domínio em proteínas vegetais não foram investigados em detalhe, apesar da sua presença conter, por exemplo, em miosinas. O domínio de QI (IQD) possui ativadores de transcrição de ligação a came (CAMTAs) e canais de nucleótidos cíclicos fechados (CNGCs) (Bahler et al., 2002, Abel et al., 2005).

O *contig* 5 mostrou um *SNP* a cada 12,60, ou seja seria um meio termo em relação ao máximo e mínimo da frequência de *SNPs* por pb. Este demostrou a proteína receptor *LRR* serina / treonina-proteína-quinase. *Receptor protein kinases (RPK)*, ou seja, Proteína Receptora Quinases, ativam um complexo conjunto de vias

de sinalização intracelular em resposta ao ambiente extracelular (VAN DER GEER *ET AI.,* 1994;. PADGETT, 1999). *RPK* são proteínas transmembranares de passagem única, que contêm uma sequência de sinal amino terminal, os domínios únicos extracelulares para cada receptor, e um domínio citoplasmático da quinase.

Os *LRRs* formam um solvente exposto que cria uma superfície que faz a mediação de interações proteína-proteína em outros sistemas (KOBE E DEISENHOFER, 1995). Em plantas *LRR-RLKs* estão envolvidos em vários processos, incluindo regulação do desenvolvimento, resistência a doenças, e sinalização de hormônios esteroides (TORII *et al.,* 1996). Mutações em qualquer um dos *LRRs* ou os domínios de quinase leva à perda de função, confirmando a importância destes domínios para a função (TORII *et al.,* 1996). Os SNPs encontrados nessa região representam mutações mantidas, portanto, com valor filogenético/adaptativo.

O *contig* 6 demonstrou a proteína Inositol que é um importante metabólito celular, necessário para o crescimento e desenvolvimento de vegetais (Valluru e Ende, 2011). Em plantas, ele desempenha um papel de destaque no metabolismo do inositol, fornecendo inositol e inositídeos envolvidos em processos metabólicos e em estruturas vegetativas (Goya et al., 2011).

O *contig* que mostrou maior frequência foi o 7 que possui um SNP a cada 7,53 pb, neste foi encontrada a proteína arabinogalactânicas (AGPs), altamente glicosiladas e encontradas em abundância à superfície das células vegetais, localizando-se na membrana celular, espaço periplásmico, parede celular e secreções (SHOWALTER *et al.,* 2010; ELLIS *et al.,* 2010). E mais precisamente a subclasse desta proteína fasciclina que é composta por aproximadamente de 110 a 150 aminoácidos de comprimento e têm baixa similaridade de sequência. Esta semelhança de sequências baixas pode explicar a falta de uma única sequência de consenso para domínios de fasciclina. Contudo, todos os domínios fasciclina contêm duas regiões altamente conservadas (H1 e H2) de aproximadamente 10 aminoácidos cada (KAWAMOTO *et al.,* 1998).

O *contig* 8 mostrou a proteína Adenosina quinase que possui atividade de diversas proteínas em um organismo é regulada por fosforilação e desfosforilação,

através de quinases e fosfatases, respectivamente. Por isso, as quinases são consideradas proteínas regulatórias e utilizam a fosforilação proteica no controle das vias de sinalização celular, podendo propagar ou regular um sinal através da adição de um grupo fosfato ao substrato (BARTELS e SUNKAR, 2005). A enzima adenosina quinase é tipicamente constitutiva e catalisa, em eucariotos, a fosforilação da adenosina em adenosina monofosfato (AMP) (MOFFATT *et al.,* 2000; PARK & GUPTA, 2008).

A estimativa do número de *SNPs* observados entre espécies dependem, naturalmente, das amostras utilizadas na análise. Se as amostras apresentam grande diversidade genética, a tendência é se observar maior número de *SNPs* por par de bases (pb) analisado.

Regiões codificantes são importantes para a manutenção das características essenciais à vida dos organismos. Portanto, mutações ocorrem em uma taxa relativamente baixa e esta taxa varia de região gênica para região gênica. A posição de um *SNP* dentro da região codificadora pode ser importante, pois pode ocorrer em um local altamente conservado ou no centro ativo de uma enzima. Assim, mesmo que a frequência *SNP*/pb seja apenas uma medida conceitual, ela serve como referência ao número de mutações que ocorre entre espécies diferentes.

Além dos estudos de diversidade de DNA, a grande ocorrência de *SNPs* no genoma, teoricamente aumentaria a possibilidade de que estas mutações fossem próximas a genes de interesse agronômico. Assim, estes seriam utilizados para identificação de variantes alélicas em regiões codificadoras levando a associação entre o genótipo observado e o fenótipo, o que facilitaria a seleção assistida para características complexas.

A frequência de *SNPs* encontrados neste trabalho foi classificada como baixa, porém aceitável quando comparada ao que pode ser visto em algumas espécies de angiospermas, onde a frequência pode chegar a até 60 *SNP*/Kb, como é o exemplo de *Populus tremula* (SEBASTIANI *et al.,* 2004). Além da baixa frequência de *SNPs* no genoma das plantas e principalmente nas regiões expressas pode indicar uma elevada conservação dos genes. Por outro lado, pode-se levantar a hipótese de que os genomas das duas espécies aqui avaliadas resistem a possíveis modificações,

consequentemente a baixa frequência de mutações pode ser reflexo desta estabilidade e a relação das frequências de *SNP*/kb podem atuar como um indicador para esta situação.

Como pode ser observado na Tabela 2, as frequências de transições são relativamente maiores do que as transversões em todos os *contigs*, com exceção do *contig* 8, o que é corroborado pelo trabalho de BROOKES (1999), que citou que as mutações mais recorrentes são as do tipo transição, em que há troca de purina por outra purina (A <-> G) ou de uma pirimidina por outra pirimidina (C <-> T). O número de transições é geralmente maior que o de transversões, pois a substituição de bases nitrogenadas de mesma estrutura molecular é um evento mutacional muito mais provável de ocorrer do que uma substituição de uma base púrica por uma pirimídica e vice-versa (KLABUNDE 2016).

Tabela 2: Distribuição dos *SNPs* putativos por tipo de mutação.

	Contig 1		*Contig 2*		*Contig 3*		*Contig 4*		*Contig 5*		*Contig 6*		*Contig 7*		*Contig 8*	
Transições	N°	%	N°	%	N°	%	N°	%	N°	%	N°	%	N°	%	N°	%
A↔G	5	38,46	4	21,05	4	44,44	17	41,46	5	22,72	4	30,76	17	30,35	1	20
C↔T	4	30,76	6	31,57	0	0	12	29,26	10	45,45	4	30,76	20	35,71	1	20
Transversões																
A↔T	2	15,38	2	10,52	2	22,22	7	17,07	1	4,54	0	0	1	1,78	0	0
G↔T	1	7,7	0	0	0	O	1	2,43	0	0	1	7,7	3	5,36	1	20
C↔G	1	7,7	6	31,57	1	11,11	2	4,87	4	18,18	2	15,39	10	17,86	0	0
A↔C	0	0	1	5,26	2	22,22	2	4,87	2	9,09	2	15,39	4	7,14	2	40
Total	13	100	19	100	9	100	41	100	22	100	13	100	55	100	5	100

4 CONCLUSÕES

As frequências dos *SNPs* encontrados foram consideradas baixas em comparação a outras espécies e por isso enfatizam a hipótese de que estas seriam regiões com um alto grau de conservação entre as duas espécies.

A busca de regiões gênicas putativas através do banco de dados *ESTs* é uma técnica pouco explorada, mas se mostrou muito eficiente e prática para a realização de trabalhos nesta área.

5 PERSPECTIVAS

Os dados gerados neste trabalho abrem uma porta para novos estudos com aplicações destas ferramentas que demonstraram praticidade e expressaram um grande número de informações. A partir disso é possível realizar maiores estudos nas áreas de filogenia e diversidade genética da espécie *Eugenia uniflora.*

Outra perspectiva, é explorar a viabilidade do uso da proteína LRR como ferramenta de análise evolutiva e filogenética da família mirtácea.

REFERÊNCIAS

ABEL, S., SAVCHENKO, T. AND LEVY, M. Genome-wide comparative analysis of the IQD gene families in Arabidopsis thaliana and Oryza sativa. **BMC Evol. Biol**. 5: 72, 2005.

ADEBAJO, A. C.; OLOKI, K. J; ALADESANMI, A. Antimicrobial activity of the leaf extract of *Eugenia uniflora*. **Journal Phytotherapy Resource,** v. 3, n.3, p. 258-259, 1989.

ALMEIDA D. J.; FARIA M. V.; DA SILVA P. R.; Biologia experimental em Pitangueira: uma revisão de cinco décadas de publicações científicas. **Ambiência Guarapuava**. 2012

BAGETTI, M. Caracterização físico-química e capacidade antioxidante de pitanga (Eugenia, uniflora L.). 2009. 84 f. Dissertação (Mestrado em Ciência e Tecnologia de Alimentos) - universidade Federal de Santa Maria, Santa Maria, 2009.

BAHLER, M. AND RHOADS, A. Calmodulin signaling via the IQ motif**. FEBS Lett**. 513: 107–113, 2002.

BARTELS D, SUNKAR R Drought and Salt Tolerance in Plants. **Critical Reviews in Plant Sciences** 24:23–58. doi: 10.1080/07352680590910410, 2005.

BARROSO, G. M.; PEIXOTO, A. L.; COSTA, C. G.; ICHASO, C. L. & LIMA, H. C. 1984. Sistemática das Angiospermas do Brasil. Myrtaceae. v.2. Viçosa, Ed. Univ. Fed. Viçosa, 377pp.

BEZERRA, J. E. F.; SILVA Jr., J. F.; LEDERMAN, I.E. **Pitanga (*Eugenia uniflora* L.)**Jaboticabal: Funep, 2000. 30p. (Série Frutas Nativas, 1).

BROOKES, A.J. The essence of SNPs. **Gene,** v.234, n.2, p.177-186, 1999.

BUCHANAN-WOLLASTON, V. The molecular biology of leaf senescence. **Journal of Experimental Botany**, v. 48, n. 2, 181-199, 1997.

BUETOW, K.H., EDMONSON, M.N., and CASSIDY, A.B., Reliable identification of large numbers of candidate SNPs from public EST data, **Nat. Genet.**, 21:323–325, 1999.

CRUZ, A. V. M.; KAPLAN, M. A. C. A importância das famílias *Myrtaceae* e *Melastomataceae* na etnomedicina Brasileira. **Revista Cubana de Plantas Medicinales**, v.10, n. 5, 2005.

DRAZEN, J. M.; YANDAVA, C. N.; DUBE, L.; SZCZERBACK, N.; HIPPENSTEEL, R.; PILLARI, A.; ISRAEL. E.; SCHORK, N.; SILVERMAN, E. S.; KATZ, D. A.; DRAJESK, J. Pharmacogenetic association between ALOX5 promoter genotype and the response to anti-asthma treatment. Nature Genetics, v. 22, p. 168-170, 1999

DEVOTO, A.; MUSKETT, P. R.; SHIRASU, K. Role of ubiquitination in the regulation of plant defence against pathogens. **Curr Opin Plant Biol**, v. 6, n. 4, 307-11, 2003.

ELLIS M, EGELUND J, SCHULTZ C, BACIC A. Arabinogalactan-proteins (AGPs): **Key regulators at the cell surface? Plant Physiology** 153, 403-419, 2010.

EMAHAZION T., FEUK L., JOBS M., SAWYER SL., FREDMAN D., ST CLAIR D,; PRINCE JA., BROOKES AJ. SNP association studies in Alzheimer's disease highlight problems for complex disease analysis. **Trends in Genética**. 17:407-413, 2001

EMARA MG, KIM H. Genetic Markers and their Application in Poultry Breeding. **Poultry Science**. 2003; 82: 952–957.

Fladung M.; Buschbom J. Identification of single nucleotide polymorphisms in diferente Populus species; **Springer-Verlag** 2009.

GRIFFITHS, P. E. & NEUMANN-HELD, E. The many faces of the gene. BioScience, 49, 8, p. 656-62, 1999. Guimarães, R. C. & Moreira, C. H. C. O conceito sistêmico de gene – uma década depois. In: D'Ottavia

GOYOAGA C, BURBANO C, CUADRADO C, ROMERO C, GUILLAMÓN E, VARELA A, PEDROSA MM, MUZQUIZ M. Content and distribution of protein, sugars and inositol phosphates during the germination and seedling growth of two cultivars of Vicia faba. **Journal of Food Composition and Analysis** 24:391–397, 2011.

GOLDBERG, A. L. Protein degradation and protection against misfolded or damaged proteins. **Nature**, v. 426, n. 6968, 895-9, 2003.

HAGLUND, K.; DIKIC, I. Ubiquitylation and cell signaling. **EMBO J**, v. 24, n. 19, 3353-9, 2005.

HYTEN, D. L.; CHOI, IK-YOUNG; SONG, QIJIAN; SPECHT, JAMES E.; CARTER, FTOMAS E. JR.; SHOEMAKER, RANDY C.; HWANG, EUN-YOUNG; MATUKUMALLI, LAKSHMI K.; AND CREGAN, P. B., "A High Density Integrated Genetic Linkage Map of Soybean and the Development of a 1536 Universal Soy Linkage Panel for Quantitative Trait Locus Mapping" (2010). **Agronomy & Horticulture -- Faculty Publications**.

KANTETY, R.V., M.L. ROTA, D.E. MATHEWS, AND M.E. SORRELLS. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. **Plant. Mol. Biol. Rep. v. 48, p.501-510,** 2002.

KAWAMOTO T, NOSHIRO M, SHEN M, NAKAMASU K, HASHIMOTO K, KAWASHIMA-OHYA Y, GOTOH O, KATO Y (1998) Structural and phylogenetic analyses of RGD-CAP/ ig-h3, a fasciclin-like adhesion protein expressed in chick chondrocytes. **Biochim Biophys Acta** 1395: 288–292

KLABUNDE G. H. F. Análise transcriptômica nas coníferas brasileiras *araucaria angustifolia* (bert.) o. kuntze (araucariaceae) e *podocarpus lambertti* klotzsch ex eichler (podocarpaceae): anotação funcional e mineração de marcadores moleculares ssrs E SNPs. **Tese de Doutorado, Florianópolis**, 2016.

KRAWEZAK, M.; REISS, J.; COOPER, D.N.; The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences**. Human Genetics**, v. 90, p. 41-54, 1992.

KOBE B.; DEISENHOFER J. Proteins with leucine-rich repeats. **Current Opinion in Structure Biology**, v.8, n.5, p.409-416. 1995.

LANDRUM L. R.; KAWASAKI M. L. The genera of Myrtaceae in Brazil: an illustrated synoptic treatment and identification Keys. Brittonia. **The New York Botanical Garden**, 1997.

LEFEBVREA B., TIMMERSA T., MBENGUEA M., MOREAUA S., HERVÉA C., TÓTHB K., BITTENCOURT- SILVESTREB J., KLAUSA D., DESLANDESA L., GODIARDA L., J. MURRAYC D., UDVARDIC M. K., RAFFAELED S., MONGRANDD, CULLIMOREA S. J., GAMASA P., NIEBELA A., E OT T. A remorin protein interacts with symbiotic receptors and regulates bacterial infection, 2009.

MARÍN M., THALLMAIR V. , and OTT T. The Intrinsically Disordere dN-terminal Region of AtREM1.3 Remorin Protein Mediates Protein-Protein Interactions. **The journal of biological chemistry** vol. 287, NO. 47, November 16, 2012

MATSUMURA, T.; KASAI, M.; HAYASHI, T.; ARISAWA, M.; MOMOSE, Y.; ARAI, I., AMAGAYA S., KOMATSU Y., a-Glucosidase inhibitors from Paraguayan natural medicine, Nangapiry, the leaves of *Eugenia uniflora*. **Pharmaceutical Biology**, v.38, p.302–307, 2000.

MOFFATT B A, WANG L, ALLEN MS, *et al*. Adenosine kinase of Arabidopsis. Kinetic properties and gene expression. Plant physiology 124:1775–1785, 2000.

PADGETT RW Intracellular signaling: fleshing out the TGF b pathway. Curr Biol 9: R408–R411 Sambrook J, Fritsch EF, Maniatis T (1989) **Expression of cloned genes in cultured mammalian cell**, 1999.

PARK J, GUPTA RS Adenosine kinase and ribokinase - the RK family of proteins. Cell Mol Life Sci 65:2875–2896. doi: 10.1007/s00018-008-8123-, 2008.

PICOULT-NEWBERG, L. ET AL., Mining SNPs from EST databases, Genome Res., 9(2):167–174, 1999.

QIU Y. L., LEE J., QUADRONI F. B., SOLTIS D. E., SOLTIS P. S., ZANIS M., ZIMMER E. A., CHEN Z., SAVOLAINENK V. AND CHASEK M. W. The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. **Nature** 402, 404-407, 1999.

SANTORO A. Identificação de Single Nucleotide Polymorphisms (SNPs) no gene Nove-cis-epoxicarotenóide dioxigenase (NCED) em *Eucalyptus.* Dissertação (mestrado) – **Instituto de Biociências de Botucatu, Universidade Estadual Paulista,** 2010.

SEBASTIANI, F., S. CARNEVALE, AND G. G. VENDRAMIN. 2004. A new set of mono- and dinucleotide chloroplast microsatellites in Fagaceae. Molecular Ecology Notes 4: 259–261.

SHOWALTER AM, KEPPLER B, LICHTENBERG J, GU D, WELCH LR.. A Bioinformatics Approach to the Identification, Classification, and Analysis of Hydroxyproline-Rich Glycoproteins. **Plant Physiology** 153, 485-513. 2010.

SMALL R. L., CRONN R. C. AND WENDEL J. F. Use of nuclear genes for phylogeny reconstruction in plants. Aust. Syst. Bot. 17, 145-170, 2004.

SOLTIS P. S., SOLTIS D. E. AND CHASE M. W. Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. **Nature** 402, 402-404, 1999.

TABASSUM J.; SUMAN L. Single nucleotide polymorphism (SNP)–Methods and applications in plant genetics**: Department of Botany, Delhi University, Delhi** 110 007, India, 2006.

TORII KU, MITSUKAWA N, OOSUMI T, MATSUURA Y, YOKOYAMA R, WHITTIER RF, KOMEDA Y (1996) The Arabi- dopsis erecta gene encodes putative receptor protein kinase with extracellular leucine-rich repeats. **Plant Cell 8: 735–746**

VALLURU R, ENDE WV. Myo-inositol and beyond- emerging networks under stress. **Plant Science** 18:387–400, 2011.

VAN DER GEER P, HUNTER T, LINDBERG RA Receptor protein-tyrosine kinases and their signal transduction pathways. Annu Rev Cell Biol 10: 251–337, 1994.

VIDAL R. O.; CARAZZOLLE M. F.;. SAMPAIO C. L. M; COSTA G. G. L.; FORMIGHIERI E. F.;. POT D; MONDEGO J. M. C.; PEREIRA G. A. G. Identificação de polimorfismos de nucleotídeos únicos em montagem de ests de três espécies de café. Laboratório de Genômica e Expressão, Instituto de Biologia, UNICAMP, Campinas, SP, Brazil; 2Cirad, UMR PIA, Avenue d'Agropolis, Montpellier Cedex 5, France.

WANG, K.L., YOSHIDA, H., LURIN, C. AND ECKER, J.R.Regulation of ethylene gas biosynthesis by the Arabidopsis ETO1 protein. **Nature, 428, 945–950,** 2004.

WATSON J. D. Biologia Molecular do Gene – 7. Ed. – Porto Alegre : Artmed, 2015.

ZAHA. A.; FERREIRA H. B.; PASSAGLIA L. M. P. Biologia Molecular Básica – 3. Ed. Porto Alegre: Mercado Aberto 2003.

ANEXOS

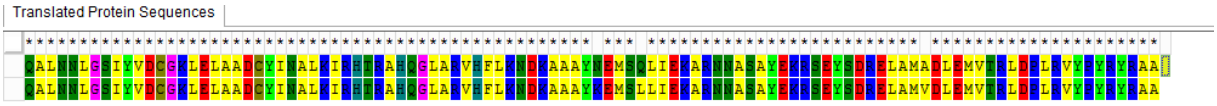
Contig 1

Range 1: 2473 to 2785 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
507 bits(274)	3e-142	300/313(96%)	0/313(0%)	Plus/Plus

```
Query  31    CAGGCGCTGAACAATCTTGGAAGTATCTATGTTGACTGTGGGAAGCTAGAATTGGCAGCT  90
             ||||| |||||||||||| |||||||||||| || ||||||||||||||| ||||||||||
Sbjct  2473  CAGGCACTGAACAATCTCGGAAGTATCTACGTCGACTGTGGGAAGCTGGAATTGGCAGCT  2532

Query  91    GATTGCTATATTAACGCTCTCAAAATCAGGCACACAAGAGCCCATCAAGGGCTTGCTCGA  150
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2533  GATTGCTATATTAACGCTCTCAAAATCAGGCACACAAGAGCCCATCAAGGGCTTGCTCGA  2592

Query  151   GTACATTTTCTGAAAAATGACAAAGCTGCAGCATACAATGAGATGTCACAACTGATAGAG  210
             |||||||||| |||||||||||||||||||||||||| |||||||||| ||||||||||
Sbjct  2593  GTACATTTTCTCAAAAATGACAAAGCTGCAGCATACAAGGAGATGTCACTACTGATAGAG  2652

Query  211   AAGGCGAGGAATAATGCGTCTGCATATGAGAAGAGGTCAGAATATTCTGATCGTGAACTC  270
             ||||| |||||||||||| ||||| |||||||||||||| ||||||||||||||||||||
Sbjct  2653  AAGGCAAGGAATAATGCATCTGCGTATGAGAAGAGGTCTGAATATTCTGATCGTGAACTC  2712

Query  271   GCTATGGCCGACCTCGAGATGGTTACCAGATTAGACCCACTTCGTGTCTATCCATACAGA  330
             |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2713  GCTATGGTCGACCTCGAGATGGTTACCAGATTAGACCCACTTCGTGTCTATCCATACAGA  2772

Query  331   TACCGGGCTGCAG  343
             |||||||||||||
Sbjct  2773  TACCGGGCTGCAG  2785
```

Translated Protein Sequences

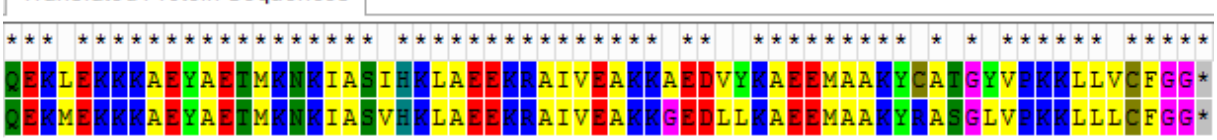


Contig 2

Score	Expect	Identities	Gaps	Strand
272 bits(147)	6e-72	185/204(91%)	0/204(0%)	Plus/Plus

```
Query  1    CAGGAAAAACTGGaaaaaaaGAAAGCGGAATATGCAGAGACAATGAAGAACAAGATTGCT  60
            ||||||||| |||||||| ||||||||| ||||||||||||||||||||||||||||||
Sbjct  544  CAGGAAAAAATGGAAAAGAAGAAAGCAGAATATGCAGAGACAATGAAGAACAAGATTGCT  603

Query  61   TCGATCCACAAGCTCGCTGAAGAAAAGAGGGCGATTGTGGAAGCCAAGAAAGCGGAAGAT  120
            ||| |||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Sbjct  604  TCGGTCCACAAGCTCGCTGAAGAAAAGAGGGCGATTGTGGAAGCCAAGAAAGGGGAAGAT  663

Query  121  GTCTACAAGGCAGAGGAGATGGCTGCTAAGTACTGCGCAACTGGATATGTTCCAAAGAAG  180
            ||   |||||||||||||||||||||||||||||   |||||||  |||  |||||| ||||||
Sbjct  664  CTCCTGAAGGCAGAGGAGATGGCTGCTAAGTATCGCGCAAGCGGACTTGTTCCGAAGAAG  723

Query  181  CTCCTTGTCTGCTTTGGAGGCTGA  204
            || ||  ||||||||||||||||||
Sbjct  724  CTGCTCCTCTGCTTTGGAGGCTGA  747
```

Translated Protein Sequences



Contig 3

Score	Expect	Identities	Gaps	Strand
122 bits(66)	3e-27	84/93(90%)	0/93(0%)	Plus/Plus

```
Query  6    AAAATCGAAGAACAGGACACCATGCATGATATTGCTTGATTCACTGCACATGGCTAACCC  65
            |||||| |||  ||| ||||||||||||||||||||||||||||||||||||| |||||
Sbjct  567  AAAATCAAAGGCCAGAACACCATGCATGATATTGCTTGATTCACTGCACATGGCAAACCC  626

Query  66   TAAGAGACTGGAACCTGACATAAGAAAGTATGT  98
            ||| ||||||||||||  |||||||||| |||
Sbjct  627  TAAAAGACTGGAACCTGCGATAAGAAAGTTTGT  659
```

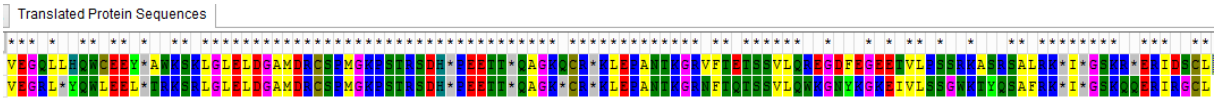
Translated Protein Sequences

Contig 4

Score	Expect	Identities	Gaps	Strand
427 bits(231)	2e-118	321/365(88%)	3/365(0%)	Plus/Plus

```
Query  5     TAATCCA-CAGGTGGAGGGCCAGCTCTTACACCAATGGTGTGAAGAATATTGAGCTTGGA  63
             ||| ||| || ||||||||||   |||| | |||||||| ||||||||  ||| || |||
Sbjct  829   TAA-CCATCA-GTGGAGGGCCGACTCTAATACCAATGGCTTGAAGAATTATGAACTCGGA  886

Query  64    AAAGCAAGTTGGGGTTGGAGCTGGATGGAGCGATGGATCGCTGCTCGCCCATGGGAAAGC  123
             |||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  887   AAAGCCGGTTGGGGTTGGAGCTGGATGGAGCGATGGATCGCTGCTCGCCCATGGGAAAGC  946

Query  124   CGAGTACCCGTTCAGACCATTAGCCCGAAGAAACCACTTAACAAGCAGGGAAGCAATGTC  183
             |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Sbjct  947   CGAGTACCCGTTCAGACCATTAGCCCGAAGAAACCACTTAACAAGCAGGGAAGTAATGTC  1006

Query  184   GCTAAAAGCTTGAACCAGCAAACACCAAAGGCCGTGTCTTCACCGAAACCTCCTCTGTCC  243
             |||||||||||||||||||||||||||||||||||  ||||||| |||||||||| ||||
Sbjct  1007  GCTAAAAGCTTGAACCAGCAAACACCAAAGGCCGTAACTTCACCCAAACCTCCTCCGTCC  1066

Query  244   TCCAACGGGAAGGCGACTTCGAAGGCGAGGAGACTGTCTTACCCAGCAGCCGAAAAGCCA  303
             ||||| || |||||  ||| | ||||| |||||| |||||| ||||| || | ||| ||
Sbjct  1067  TCCAATGGAAAGGCAACTACAAAGGCAAGGAGATTGTCTTATCCAGCGGCTGGAAAACCT  1126

Query  304   GCCGGTCAGCTCTCAGAAAGTAAATCTGAGGAAGTAAACGCTAAGAAAGAATCGACAGTT  363
              || ||| ||| |||||||||||||||||||| ||||   |||||||||||  | |||
Sbjct  1127  ACCAGTCTGCTTTCAGAAAGTAAATCTGAGGAAGCAAACAACAAGAAAGAATCCGCGGTT  1186

Query  364   GCTTA  368
             |||||
Sbjct  1187  GCTTA  1191
```

Translated Protein Sequences



Contig 5

Score	Expect	Identities	Gaps	Strand
390 bits(211)	3e-107	255/277(92%)	0/277(0%)	Plus/Plus

```
Query  1     GCGACCGACAATTTCAGCTCCAAGAACATACTCGGCACCGGAGGCTTCGGGAACGTGTAC  60
             ||||| |||||||||||| ||||||||||||| || || ||||| ||||| || || |||
Sbjct  1186  GCGACAGACAATTTCAGCACCAAGAACATACTTGGTACTGGAGGTTTCGGCAATGTATAC  1245

Query  61    AAGGGGAAGCTTGGGGACGGGACTATGGTCGCGGTGAAACGGCTAAAAGATGTAGTCGGG  120
             ||||||||||||||||| ||||||||||||||||||||||| || ||||||||||| |||
Sbjct  1246  AAGGGGAAGCTTGGGGATGGGACTATGGTCGCGGTGAAACGCCTGAAAGATGTAGTTGGG  1305

Query  121   GCAACCGGAGAATCCCAGTTCCGGACTGAACTGGAGATGATCAGCTTGGCGGTCCATCGC  180
              |||| ||||| |  ||||||||||||||||||||||||||||||||||| |||||||||
Sbjct  1306  ACAACTGGAGAGTTGCAGTTCCGGACTGAACTGGAGATGATCAGCTTGGCAGTCCATCGC  1365

Query  181   AACTTGCTCCGATTGATCGGATACTGCGCCACTTCGAACGAAAGGCTTCTGGTCTACCCT  240
             |||||||| |||||||| ||||||||||||||||||||||||||||||||||||||||||
Sbjct  1366  AACTTGCTGCGATTGATTGGATACTGCGCCACTTCGAACGAAAGGCTTCTGGTCTACCCT  1425

Query  241   CACATGTCAAATGGTAGTGTTGCTTCAAGGCTCAAAG  277
             |||||||| ||||||||||||||||||||||||||||
Sbjct  1426  CACATGTCCAATGGTAGTGTTGCTTCAAGGCTCAAAG  1462
```

Translated Protein Sequences

```
****** ********************************* *** ***********************************************
ATDNFSSKNILGTGGFGNVYKGKLGDGTMVAVKRLKDVVGATGESQFRTELEMISLAVHRNLLRLIGYCATSNERLLVYPHMSNGSVASRLK
ATDNFSTKNILGTGGFGNVYKGKLGDGTMVAVKRLKDVVGTTGELQFRTELEMISLAVHRNLLRLIGYCATSNERLLVYPHMSNGSVASRLK
```

Contig 6

Score	Expect	Identities	Gaps	Strand
159 bits(176)	1e-38	109/123(89%)	0/123(0%)	Plus/Plus

```
Query  12    AGTTGCCAAACCCCTAGTGGTGGATGGGAGCCCCAAGTCCCATGAATTGTTTCTTGCTTA  71
             ||||||||| || |||||||||||||| | | |||||||||||||| |||||||||||||
Sbjct  363   AGTTGCCAAGCCACTAGTGGTGGATGGAACCTCCAAGTCCCATGAACTGTTTCTTGCTTA  422

Query  72    TGATCGATACTCCCTTGTAGAGCTTGAGCGCCCCCTGGTTTTGCAAGAGTTTGTTAATCA  131
             |||||  | ||||||||||||||||||||  ||||| ||||||||||||||||| |||||
Sbjct  423   TGATCAGTCCTCCCTTGTAGAGCTTGAGCCTCCCCTTGTTTTGCAAGAGTTTGTCAATCA  482

Query  132   TGG  134
              ||
Sbjct  483   CGG  485
```

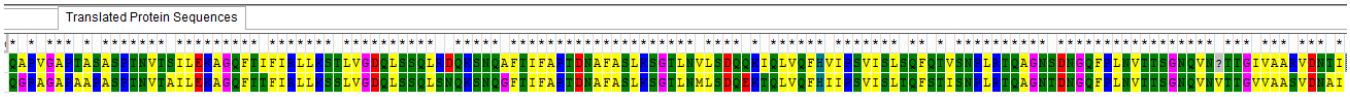
Translated Protein Sequences

```
***  ****  **** ****** ******* * ***** * 
SCQTPSGGWEPQVP*IVSCL*SILPCRA*APPGFARVC*SW
SCQATSGGWNLQVP*TVSCL*SVLPCRA*ASPCFARVCQSR
```

Contig 7

Score	Expect	Identities	Gaps	Strand
526 bits(582)	2e-148	377/433(87%)	1/433(0%)	Plus/Plus

```
Query  1    CAAGCTCCGGTAGGGGCTCCGACGGCTTCCGCCAGCCCCACCAACGTTACCTCAATCCTC  60
            |||| |||||  |||||||||| | ||  |||||||||| ||||| || ||| ||||||||
Sbjct  232  CAAGGTCCGGCCGGGGCTCCGGCTGCCCCCGCCAGCCCGACCAATGTCACCGCAATCCTC  291

Query  61   GAGAAGGCGGGTCAGTTCACTATCTTCATTAGGCTGCTCAAGAGCACTTTGGTGGGCGAC  120
            ||||| |||||||||||||| | ||||||||||||||||||||||| | ||||||| |||
Sbjct  292  GAGAAAGCGGGTCAGTTCACCACCTTCATTAGGCTGCTCAAGAGCAGTCTGGTGGGAGAC  351

Query  121  CAACTCAGCTCTCAGTTGAGGGACCAGAAGTCGAACCAGGCCTTCACTATATTCGCGCCG  180
            || |||||||||||||||   | |||||||||||||||| |||||| |||||  |||||||
Sbjct  352  CAGCTCAGCTCTCAGTTGAGCAATCAGAAGTCGAACCAGGGCTTCACCATATTTGCGCCG  411

Query  181  ACGGACAATGCCTTCGCTAGCCTCAAGTCGGGCACACTGAACGTGCTCTCCGATCAGCAG  240
            || ||||||||  ||||||||  |||||  |||||||||  |||||||  ||||  ||||||||   ||
Sbjct  412  ACTGACAATGCGTTCGCTAGTCTCAAATCGGGCACGCTGAACATGCTTTCCGATCAAGAG  471

Query  241  AAGATCCAGCTGGTGCAGTTCCACGTCATCCCTTCGGTCATCTCGCTCTCGCAGTTCCAG  300
            |||| ||||||  |||||||||||| ||||||||||  |||||||||||| |||||||||  |
Sbjct  472  AAGACCCAGCTAGTGCAGTTCCACATCATCCCTTCCGTCATCTCGCTCACGCAGTTCTCG  531

Query  301  ACCGTCAGCAACCCACTACGGACCCAAGCTGGCAACAGCGACAATGGCCAGTTTCCGCTC  360
            ||| ||||||||||||| |||||||||||  |||||||  |||||  |||||||||||||||
Sbjct  532  ACCATCAGCAACCCACTGCGGACCCAAGCCGGCAACACTGACAACGGCCAGTTTCCGCTC  591

Query  361  AACGTCACCACCTCGGGGAATCAAGTGAA-TATACAACCGGGATTGTGGCCGCAAGAGTC  419
            ||||| ||||| ||||||||||||||||    ||| || ||| ||||||||||||| |||
Sbjct  592  AACGTGACCACATCGGGGAATCAAGTGAACGTTACGACTGGGGTTGTGGCCGCAAGCGTC  651

Query  420  GACAACACAATAA  432
            |||||| ||||||
Sbjct  652  GACAACGCAATAA  664
```



Contig 8

Score	Expect	Identities	Gaps	Strand
66.2 bits(72)	2e-10	44/49(90%)	0/49(0%)	Plus/Plus

```
Query  83   GTATGATGAAATGGCTAGCAACTACAGTGTCGATTACATTGCGGGGGGT  131
            |||||| |||||||||  |||||||||||  |||||||||||  |||||||
Sbjct  226  GTATGACGAAATGGCTAAAAACTACAGTGTAGATTACATTGCTGGGGGT  274
```