

ANGELO GABRIEL MENDES CORDEIRO

PADRÃO DE USO DE CÓDONS EM GENES DUPLICADOS DE *Corynespora cassicola*

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-graduação em Genética e Melhoramento, para a obtenção do título de *Magister Scientiae*.

Orientador: Luiz Orlando de Oliveira

Coorientadora: Thaís Carolina da Silva Dal'sasso

VIÇOSA - MINAS GERAIS

2024

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade
Federal de Viçosa - Campus Viçosa**

T

C794p
2024

Cordeiro, Angelo Gabriel Mendes, 1998-

Padrão de uso de códons em genes duplicados de
Corynespora cassicola / Angelo Gabriel Mendes Cordeiro. –
Viçosa, MG, 2024.

1 dissertação eletrônica (53 f.): il. (algumas color.).

Orientador: Luiz Orlando de Oliveira.

Dissertação (mestrado) - Universidade Federal de Viçosa,
Departamento de Bioquímica e Biologia Molecular, 2024.

Referências bibliográficas: f. 48-53.

DOI: <https://doi.org/10.47328/ufvbbt.2024.290>

Modo de acesso: World Wide Web.

1. Fungos fitopatogênicos - Genética. 2. Genética
molecular. 3. Genes. 4. Mutação (Biologia). 5. Código genético.
I. Oliveira, Luiz Orlando de, 1961-. II. Universidade Federal de
Viçosa. Departamento de Bioquímica e Biologia Molecular.
Programa de Pós-Graduação em Genética e Melhoramento.
III. Título.

CDD 22. ed. 579.56


ANGELO GABRIEL MENDES CORDEIRO

PADRÃO DE USO DE CÓDONS EM GENES DUPLICADOS DE *Corynespora cassiicola*

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-graduação em Genética e Melhoramento, para a obtenção do título de *Magister Scientiae*.


APROVADA: 26 de fevereiro de 2024.

Assentimento:

Documento assinado digitalmente
 **ANGELO GABRIEL MENDES CORDEIRO**
Data: 23/07/2024 12:04:46-0300
Verifique em <https://validar.iti.gov.br>

Angelo Gabriel Mendes Cordeiro

Autor

Documento assinado digitalmente
 **LUIZ ORLANDO DE OLIVEIRA**
Data: 22/07/2024 12:24:05-0300
Verifique em <https://validar.iti.gov.br>

Luiz Orlando de Oliveira

Orientador

AGREDECIMENTOS

Primeiramente a Deus, por me conceder a vida, coragem, sabedoria e resiliência, dons indispensáveis para que pudesse estar aqui hoje.

Aos meus pais (Tânia Pauli Mendes Cordeiro e Edson Cordeiro), meus irmãos e demais familiares que contribuíram com minha formação pessoal.

A minha esposa (Fernanda dos Santos de Almeida Cordeiro) pela companhia, carinho e paciência ao longo desses 2 anos.

A UNEMAT e aos colegas do GENBIOMOL (a Larissa em reparação histórica kkk) pela minha graduação, essencial para o ingresso ao mestrado.

Ao Laboratório de Biologia Molecular e Filogeografia, ao Vinicius, a Thaís e em especial ao professor Luiz Orlando de Oliveira, pela disposição, paciência e principalmente pela orientação e aprendizado ao longo deste mestrado.

Ao Instituto de Biotecnologia Aplicada à Agropecuária (BIOAGRO) pela disponibilidade de espaço para realização deste trabalho de pesquisa.

Ao Programa de Pós-Graduação em Genética e Melhoramento (PPGGM) pela oportunidade de realização o mestrado.

A Universidade Federal de Viçosa (UFV) por seus professores e funcionários, que contribuíram para fornecer ensino e condições de trabalho de excelente qualidade.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) - Código de Financiamento 001.

RESUMO

CORDEIRO, Angelo Gabriel Mendes, M.Sc., Universidade Federal de Viçosa, fevereiro de 2024. **Padrão de uso de códons em genes duplicados de *Corynespora cassiicola***. Orientador: Luiz Orlando de Oliveira. Coorientadora: Thaís Carolina da Silva Dal'sasso.

Os genes duplicados podem seguir diferentes destinos evolutivos. Após eventos de duplicação, o acúmulo distinto de mutações não-sinônimas pode fazer com que os parálogos se tornem capazes de codificar produtos gênicos divergentes. Com o passar do tempo evolutivo, estes produtos podem adquirir funções complementares ou até mesmo funções distintas. Mutações sinônimas, ou silenciosas também podem alterar a estrutura ou a função de parálogos em famílias multigênicas. O acúmulo de mutações sinônimas altera o padrão de uso de códons e assim, potencialmente, modula diferencialmente a tradução do transcrito gerado por cada parálogo. Investigar o padrão de uso de códons em famílias multigênicas pode fornecer evidências importantes para o estudo evolutivo destas famílias e como este fenômeno pode contribuir para modular o repertório de produtos gênicos a disposição do organismo. Em patógenos, a investigação poderia ajudar a aprofundar a compreensão de eventos que acontece durante as interações patógeno-hospedeiro. *Corynespora cassiicola* é bastante conhecido por ser um fungo patogênico cosmopolita de alto índice de polifagia. Em seu genoma, este fungo acumula uma grande quantidade de genes com o potencial de serem efetores e, portanto, atuantes durante a sua interação com os distintos hospedeiros. As proteínas indutoras de necrose e que estimulam a produção de etileno (NLP), as enzimas do tipo pectinas metilesterases (PME) e as enzimas deuterolisina metaloproteases (M35) são exemplos de famílias multigênicas que estão presentes no genoma de *Corynespora cassiicola*. Membros destas famílias multigênicas possuem o potencial de atuarem como efetores durante as primeiras horas de infecção. Para o presente estudo, foram obtidos conjuntos de dados moleculares compostos por sequências de DNA das famílias multigênicas NLP, PME e M35. Foram obtidos também dados de genes ortólogos de cópia única (single copy orthologs, SCO), que foram utilizados como conjunto de referência durante algumas análises. Por fim, foram conduzidas uma série de análises que mensuraram o viés de uso de códons e estimam vários parâmetros a ele associados. Os resultados mostraram que a grande maioria dos parálogos dentro de cada família exibiram preferência por códons ricos em bases nitrogenadas GC. Essa preferência por GC

ocorreu principalmente em detrimento de A. Mesmo sendo constatado que genes diferentes compartilham códons preferenciais em comum, ficou evidenciado que existe mais de uma força moldando o padrão de uso de códons. No geral a maior influência é de pressão de mutação e principalmente seleção natural. Foi possível notar também que membros de uma mesma família genica podem ser afetados por diferentes forças evolutivas. Tal fato tornou difícil a tarefa de encontrar relações entre os padrões de uso de códons em membros mais próximos filogeneticamente. Outra descoberta interessante é que os parálogos que apresentaram os padrões de uso de códons mais distinto do conjunto de dados de referência foram os que apresentaram menores taxas de expressão. Esta descoberta indica uma possível relação entre os códons preferenciais de um organismo e a taxa de tradução e conseqüentemente expressão apresentada pelo mesmo.

Palavras-chave: Evolução molecular; Fungos fitopatogênicos; Genes efetores; Mutação silenciosa; Viés de códons.

ABSTRACT

CORDEIRO, Angelo Gabriel Mendes, M.Sc., Universidade Federal de Viçosa, February, 2024. **Pattern of codon usage in duplicated genes of *Corynespora cassiicola***. Advisor: Luiz Orlando de Oliveira. Co-advisor: Thaís Carolina da Silva Dal'sasso.

The duplicated genes can follow different evolutionary fates. After duplication events, the distinct accumulation of non-synonymous mutations can lead paralogs to encode divergent gene products. Over evolutionary time, these products may acquire complementary or even distinct functions. Synonymous mutations, also known as silent mutations, can also alter the structure or function of paralogs in multigene families. The accumulation of synonymous mutations changes the codon usage pattern and thus, potentially, differentially modulates the translation of the transcript generated by each paralog. Investigating the codon usage pattern in multigene families can provide important evidence for the evolutionary study of these families and contribute to modulating the repertoire of gene products available to the organism. In pathogens, such investigation could help deepen the understanding of events occurring during pathogen-host interactions. *Corynespora cassiicola* is well-known for being a cosmopolitan pathogenic fungus with a high degree of polyphagy. In its genome, this fungus accumulates a large number of genes with the potential to be effectors and therefore act during its interaction with different hosts. Necrosis-inducing and ethylene-stimulating proteins (NLP), pectin methylesterase-like enzymes (PME), and deuterolysin metalloprotease enzymes (M35) are examples of multigene families present in the genome of *Corynespora cassiicola*. Members of these multigene families have the potential to act as effectors during the early hours of infection. For this study, molecular data sets were obtained, consisting of DNA sequences from the NLP, PME, and M35 multigene families. A data set of single-copy orthologous genes (SCO) was also obtained, which was used as a reference set during some analyses. Finally, a series of analyses measured the bias in codon usage and estimated various associated parameters. The results showed that the vast majority of paralogs within each family exhibited a preference for codons rich in GC nitrogenous bases. This preference for GC occurred mainly at the expense of A. Even though different genes shared common preferred codons, it was evident that more than one force shaped the codon usage pattern. Overall, the greatest influence is from mutation pressure and, primarily, natural selection. It was also noted that members of the same gene family

can be affected by different evolutionary forces. This fact made it difficult to establish relationships between codon usage patterns in closely related phylogenetic members. Another interesting discovery is that paralogs that exhibited the most distinct codon usage patterns from the reference data set were those with lower expression rates. This finding suggests a possible relationship between an organism's preferred codons and the translation rate and, consequently, the expression exhibited by it.

Keywords: Codon bias; Effector genes; Molecular evolution; Phytopathogenic fungi; Silent mutation.

SUMÁRIO

1. INTRODUÇÃO	09
2. OBJETIVOS	14
2.1 Objetivos gerais	14
2.2 Objetivos específicos	14
3. MATERIAIS E MÉTODOS	15
3.1 Construção dos bancos de dados	15
3.2 Índices utilizados na estimativa do viés no uso de códons	16
3.2.1 Índice de Adaptação de Códons – CAI	16
3.2.2 Uso Relativo de Códons Sinônimos – RSCU	17
3.2.3 Número Efetivo de Códons – ENC	17
3.2.4 Índice Intrínseco de Desvio de Códon – ICDI	18
3.2.5 Índice de Viés de Códon – CBI	18
3.2.6 Índice de Ordem do Uso de Códon Sinônimo – SCUO	19
3.2.7 Frequência Ótima de Códons – FOP	19
3.2.8 Índice de Similaridade de Uso de Códons – COUSIN	20
3.3 Gráficos de viés (PR2) e de neutralidade	21
3.4 Softwares utilizados	22
4. RESULTADOS	22
4.1 Conjunto de dados de referência – 7962 genes SCO	22
4.2 Família de genes das proteínas Indutoras de Necrose e Etileno – NLP	26
4.3 Família de genes das enzimas Pectinas Metilesterases – PME	31
4.4 Família de genes das enzimas Deuterolisinas Metaloproteases – M35	36
5. DISCUSSÃO	42
5.1 Padrões no uso de códons e suas implicações	42
5.2 As diferentes forças que moldam o padrão de uso de códons e suas possíveis relações com a filogenia	44
5.3 As preferencias no Uso Relativo de Códons Sinônimos	45
6. CONCLUSÃO	47
7. REFERÊNCIAS BIBLIOGRÁFICAS	48

1. INTRODUÇÃO

Segundo Ohno (1970), os genes podem seguir três destinos evolutivos distintos após terem sido submetidos a eventos de duplicação. Um destes destinos é a pseudogenização, que ocorre quando uma das cópias do gene duplicado sofre mutações que a torna não funcional. O segundo é a sub-funcionalização, um destino que ocorre quando as duas cópias são mantidas em um estado ainda funcional, mas suas funções se tornam complementares, passando a dividir as funções que antes era exercida apenas pelo único gene ancestral. Por fim, temos a neo-funcionalização, destino no qual uma das cópias duplicadas evolui para executar uma nova função biológica que não estava presente no gene ancestral, enquanto a outra cópia é submetida a fortes pressões seletivas e mantém a função original do gene ancestral (Ohno, 1970).

Dentre as modificações que podem alterar as funções de genes duplicados, as mais conhecidas e compreendidas são as mutações ditas não-silenciosas. Estas mutações ocorrem quando um evento de substituição, adição ou deleção de uma base nitrogenada no DNA resulta na mudança de um aminoácido no produto gênico, a proteína. Entretanto, também existem as mutações ditas silenciosas, também denominadas de mutações sinônimas. Quando estas mutações ocorrem, a base nitrogenada do DNA é alterada; mas, devido a degenerância do códon genético, a alteração da base não resulta na mudança do aminoácido na proteína. Por muito tempo, as mutações sinônimas foram ignoradas sob alegação de que, como produto final, não apresentavam consequência alguma para o organismo, visto que elas não alteravam a estrutura primária da proteína (Lehninger; Nelson; Cox, 1995).

Inicialmente, o que se tinha de conhecimento claro e estabelecido para a ciência era que o código genético consistia em um conjunto de 64 códons. Quando excluimos os 3 códons de parada (que não codifica nenhum aminoácido) e os códons AUG e UGG (que codificam apenas para os aminoácidos Metionina e Triptofano, respectivamente), os códons restantes (um total de 59) codificam 18 aminoácidos, ou seja, diferentes códons codificam um mesmo aminoácido. Por essa característica, foi dado ao código genético a designação de degenerado, enquanto que aqueles códons que codificam para um mesmo aminoácido, ficaram conhecidos como códons sinônimos (Watson et al., 2006).

Já existia a informação de que o uso de códons sinônimos não era algo uniforme, entretanto como mencionado, acreditava-se que, quando por ventura, ocorresse uma mutação que alterasse a primeira ou a última base de um códon, porém este continuava a codificar o mesmo aminoácido, diziam que esta era uma mutação sinônima ou silenciosa. Além disso, a partir do momento que constataram que as mutações sinônimas acontecem em uma velocidade bem maior em relação às mutações não sinônimas, foi fortemente assumida a neutralidade das mutações sinônimas (Kimura, 1977).

Com o passar dos anos e os grandes avanços tecnológicos nas áreas de genética e biotecnologia, que hoje permitem sequenciar genes e genomas com maior facilidade que antes, os estudos nessas áreas também avançaram. Deste modo, as mutações sinônimas, antes ignoradas, agora passam a despertar o interesse de pesquisadores, justamente por não serem tão silenciosas como se imaginavam. Hoje já se sabe que, apesar de sinônimas, tais mutações influenciam vários fatores de aspecto funcional e estrutural como alterações nos níveis de expressão de genes individuais, eficiência global da tradução (velocidade e precisão), aptidão celular e até no enrolamento de mRNA (Kudla et al., 2009).

O viés de uso de códons caracterizado pelo uso preferencial de alguns tipos de códons em detrimento de outros que codificam o mesmo aminoácido durante a síntese proteica, é agora um fenômeno empiricamente bem estabelecido (Deng et al., 2020). No entanto, ainda não está claro porque o padrão de uso de códons evoluiu e quais as forças evolutivas que impulsionam a sua evolução. Em termos gerais, pode-se pensar em forças mutacionais ou resultantes da seleção moldando o padrão de uso de códons. As forças de seleção ainda podem ser subdivididas entre aquelas que afetam diretamente a frequência relativa dos códons em todo o genoma, de uma maneira que afete todos os genes uniformemente ou de tal modo que a seleção opere em genes individuais (Deng et al., 2020). Neste último caso, algumas combinações de escolhas de códons podem ser preferidas para alguns genes, enquanto outras combinações sejam preferidas por outros genes.

Mutações, especialmente as ditas não-sinônimas, são eventos evolutivamente importantes pois podem contribuir para o surgimento de novas raças de patógenos, capazes de quebrar a resistência altamente específica adquirida pelo hospedeiro

(Castanho, 2017). Genes caracterizados por codificar proteínas com capacidade de promover virulência, são conhecidos como genes efetores. Mais especificamente, as proteínas produto dos genes efetores são capazes de alterar as funções e a estrutura das células do hospedeiro, facilitando a infecção e colonização (Lo Presti et al., 2015). Essas proteínas desempenham um papel efetivo no estabelecimento de condições que permitem a sobrevivência, o desenvolvimento e reprodução dos agentes patogênicos (Kamoun, 2009).

Definir com certeza que um gene é efetor não é uma tarefa fácil e exige uma série de análises e critérios científicos. Por isso, algumas características compartilhadas entre fungos patogênicos têm sido amplamente utilizadas na determinação de genes candidatos a efetores. Dentre essas características, podemos citar o sinal de secreção existente na região N-terminal em proteínas codificadas por genes efetores, na maioria dos fungos patogênicos; serem enriquecidos em resíduos de cisteína; demonstrarem ausências de similaridade nas bases de dados e possuírem tamanhos relativamente pequenos. (Lowe; Howlett, 2012; Duplesis; Bakkeren; Hamelin, 2014; De Carvalho et al., 2017).

O *Corynespora cassiicola* pertence à família Corynesporascaceae (filo Ascomycota, classe Dothideomycetes, ordem Pleosporales) popularmente conhecido por ser o causador da “mancha alvo” em plantas (Voglmayr; Jaklitsch, 2017; Déon et al., 2012). Esta espécie de fungo vem causando grande preocupação devido aos danos causados a um amplo número de culturas economicamente importantes, como algodão (*Gossypium hirsutum* L.), soja (*Glycine max* L.), café (*Coffea canephora* cv. *Conilon* L.), feijão (*Phaseolus vulgaris* L.), mandioca (*Manihot* sp.), dentre outras (Qi et al., 2011).

Além de seu considerado índice de polifagia mencionado, outras características do *Corynespora cassiicola* que chamam bastante atenção são sua capacidade de atingir desde a parte aérea até sistema radicular e seu poder de infectar tanto plantas mais lenhosas e de maiores portes, como seringueira (*Hevea brasiliensis* L.) e eucalipto (*Eucalyptus grandis* W. Hill ex Maiden), quanto plantas menores, como tomate (*Solanum lycopersicum* L.) e pepino (*Cucumis sativus* L.) (Qi et al., 2011; Reis et al., 2014). Em revisão bibliográfica recente (Farr; Rossman; Castlebury, 2021), a abrangência deste cenário é demonstrada em cerca de 200 trabalhos publicados

entre 1930 e 2021, nos quais verificou-se existência de *Corynespora cassiicola* em cerca de 400 espécies de diferentes hospedeiros. O estudo constatou também que, apesar da preferência por regiões tropicais e subtropicais, esta espécie de fungo encontra-se distribuído por mais de 70 países de todos os continentes do globo terrestre (Farr; Rossman; Castlebury, 2021).

Acredita-se que parte dessa “alta capacidade infecciosa” do *Corynespora cassiicola*, relatada anteriormente, pode estar relacionada ao grande repertório de genes supostamente efetores compondo seu genoma (Lopez et al., 2018). Alguns exemplos de genes em *Corynespora cassiicola* que se enquadram na designação de possíveis efetores, são os genes que codificam para as proteínas indutoras de necrose e que estimulam a produção de etileno (NLP), enzimas do tipo pectinas metilesterases (PME) e as enzimas deuterolisina metaloproteases (M35).

Esses três exemplos de famílias de genes possivelmente efetores citadas acima, foram amplamente estudadas em trabalhos anteriores de nosso grupo de pesquisa (Dal’Sasso et al., 2022; Rocha et al., 2023a). A partir do sequenciamento de genomas de isolados do *Corynespora cassiicola* e diversas análises desses mesmo, foram obtidas informações úteis quanto o número de cópias que existem de cada gene, além de suas possíveis relações filogenéticas (Dal’Sasso et al., 2022; Rocha et al., 2023a).

As NLP (Necrosis-and Ethylene-inducing peptide 1-like proteins), como o próprio nome já diz, são proteínas caracterizadas por estimular respostas que induzem necrose e a produção de etileno em seus hospedeiros. Estas proteínas compõem uma superclasse de proteínas não exclusiva de *Corynespora cassiicola*, podendo ser identificadas em bactérias, oomicetos e outros fungos (Gijzen; Nurnberger, 2006). As NLP são geralmente consideradas efetores devido à sua citotoxicidade para plantas eudicotiledôneas que induz respostas imunes de plantas e morte celular por necrose (Irieda et al., 2019; Ono et al., 2020). As NLP constituem três famílias (NLP1, NLP2 e NLP3), sendo a família NLP1 dividida em duas subfamílias (NLP1.1, NLP1.2). Recentemente foi descoberto que no genoma de *Corynespora cassiicola* podem existir quatro cópias de NLP: NLP1.1, NLP1.2A, NLP1.2B e NLP1.3 (Dal’Sasso et al., 2022).

As PME (Pectinmethylesterases) são enzimas que possuem a capacidade de remodelar a pectina na parede celular para desempenhar vários papéis no crescimento e desenvolvimento (Pelloux et al., 2007). Do ponto de vista fúngico, esta capacidade das PME desempenham um papel crucial na invasão de tecidos vegetais durante a infecção, pois, ao alterar o padrão de metilesterificação da pectina ocorre o rompimento da matriz coesa da pectina que envolve a rede celulose-hemicelulose consequentemente reduzindo a rigidez da parede celular da planta (Sella et al., 2016).

Por sua vez, as M35 (Deuterolysin metalloproteases) são enzimas proteolíticas dependentes de zinco, encontrados em fungos e bactérias (Li; Zhang, 2014; Arnadottir et al., 2009; Rawlings et al., 2018). Embora o papel funcional da deuterolisina ainda não seja amplamente conhecido, acredita-se que as M35 desempenhem papéis funcionais na patogenicidade tanto em bactérias quanto em fungos (Rocha, 2023b). Uma série de evidências relatam o envolvimento das M35 em funções relacionadas a supressão de sistema imunológico de hospedeiros, degradação de paredes celulares, indução de apoptose, dentre outras (Pan et al., 2020; Zhang et al., 2021).

Além de serem possivelmente efetoras, outra característica que essas proteínas mencionadas acima têm em comum é serem codificadas por parálogos que formam famílias gênicas. Ou seja, essas proteínas são produtos transcritos de mais de um único gene. A existência de parálogos resulta de eventos de duplicação de regiões do genoma ao longo do curso da evolução (Ohno, 1970).

Tendo em vista tudo que foi exposto e com o intuito de entender as relações estabelecidas entre genes duplicados e nas interações entre o *Corynespora cassiicola* e seus hospedeiros, a análise do viés no uso de códons surge como uma alternativa que apesar de nova, já é bastante utilizada neste contexto. Tais análises podem auxiliar também na compreensão do comportamento entre o padrão de uso de códons e as relações filogenéticas existente entre os genes duplicados.

Levando em consideração que a diferenciação e o funcionamento das cópias oriundas de eventos de duplicação ainda não estão bem definidos na literatura, visto que, esta é uma área de estudo em constante desenvolvimento. Além disso, quando pesquisamos por uso de códons em genes duplicados e efetores em fungos a

escassez de trabalhos publicados a este respeito é ainda maior, fica evidente a importância desse tipo de estudos.

A identificação dos padrões no uso de códons adotados por genes duplicados e potencialmente efetores podem fornecer informações importantes, que possibilitam compreender as interações entre patógenos e hospedeiros a níveis mais profundos. Inclusive, a realização e os avanços de estudos nessas áreas vêm permitindo o desenvolvimento de estratégias de substituição de códons através de genes sintéticos, baseados em parâmetros de otimização de códons de um patógeno, visando diminuir a virulência do patógeno ou aumentar a resistência em plantas (Bansal et al., 2022). Por fim, o conhecimento adquirido nesse processo pode ser de extrema importância tanto para a biologia geral, quanto na hora de traçar estratégias que visam o controle ou o desenvolvimento da resistência sustentável, úteis para o melhoramento convencional e biotecnológico (Castanho, 2017).

2. OBJETIVOS

2.1 Objetivos gerais

Analisar o padrão de uso de códons em três famílias de genes duplicados e possivelmente efetores em *Corynespora cassiicola*, bem como compreender as forças que moldam esse padrão e suas implicações.

2.2 Objetivos específicos

Verificar a existência de padrões e vieses no uso de códons em genes duplicados e possivelmente efetores, comparando-os com um conjunto de dados de referência constituído de genes ortólogos de cópia única.

Conhecer quais as forças que moldam os padrões de uso de códons nas três famílias de genes duplicados e possivelmente efetores de *Corynespora cassiicola*.

Identificar as possíveis relações entre a filogenia e as preferências de uso de códons em genes oriundos de processo de duplicação.

Indicar quais são os códons preferidos em comum pelos membros das diferentes famílias gênicas, bem como suas principais peculiaridades.

3. MATERIAIS E MÉTODOS

3.1 Construção dos bancos de dados

Foram construídos quatro bancos de dados. O banco de dados 1, utilizado na elaboração do conjunto de dados de referência, consistiu de 7962 genes ortólogos de cópia única (SCO de Single Copy Orthologs) de um isolado de *Corynespora cassiicola* (CCP) que havia sido utilizado anteriormente em um trabalho de filogenia molecular (Dal'Sasso et al., 2022). O conjunto final de genes que constituíram o banco de dados 1 foi obtido com auxílio de scripts em Python, a partir de um conjunto inicial de 9.018 genes. Este conjunto inicial foi submetido a processos de filtragem e eventualmente excluídos os genes que: 1) não iniciavam em ATG; 2) possuíam mais de um códon de parada (TAA, TAG ou TGA); 3) possuíam o número total de bases que não era múltiplo de três; e, por fim 4) possuíam um tamanho inferior a 549 pares de bases, pois, na melhor das hipóteses, estes tamanho não permitiriam cada um dos 61 códons codificantes de aminoácidos (61 códons x3 repetição mínima x3 bases por códons = 549 pb) estarem representados 3 vezes, no mínimo. Estes critérios também foram utilizados na construção dos demais bancos de dados descritos abaixo. Finalmente, o banco de dados 1 restou com um total de 7962 genes SCO.

O banco de dados 2 foi constituído de 135 CDSs (coding DNA sequences) provenientes dos 4 membros conhecidos da família de genes NLP (Necrosis and Ethylene inducing proteins) obtidos de 44 isolados de *Corynespora cassiicola* (Dal'Sasso et al., 2022). O quarto membro dessa família multigênica foi encontrado no genoma de apenas 3 dos 44 isolados disponíveis, logo: 44 CDSs da NLP 1.1 + 44 CDSs da NLP 1.2A + 44 CDSs da NLP 1.2B + 3 CDSs da NLP 1.3 = 135 CDSs (Dal'Sasso et al., 2022).

O banco de dados 3 e o banco de dados 4 foram compostos por dados genômicos de um estudo anterior de evolução molecular (Rocha, 2023b). Mais especificamente, o banco de dados 3 continha 300 CDSs dos 5 parálogos do gene PME (Pectina Metilesterase): PME 1.1, PME 1.2, PME 2.1, PME 2.2 e PME 2.3 oriundos de 60 isolados de *Corynespora cassiicola* (5 parálogos x 60 isolados = 300 CDSs) (Rocha et al., 2023a).

O conteúdo genômico dos genes que compõem o banco de dados 4 refere-se a quatro cópias dos genes M35 (Deuterolisina metaloprotease) (Rocha, 2023b). Para esta análise foram utilizados 61 isolados de *Corynespora cassiicola*. Entretanto, dos 60 isolados que possuíam a sequência gênica que codifica para a M35 2.1, 24 deles apresentaram um códon de parada prematuro, fazendo com que esses genes fossem eliminados da nossa análise, seguindo os critérios estabelecidos anteriormente (descritos acima). Dos 36 isolados restantes, 22 apresentaram esses genes classificados como possíveis não efetores (Rocha, 2023b). Sendo assim, o conjunto de genes de M35 2.1 foi dividido em M35 2.1e (possíveis efetores) e M35 2.1n (possíveis não-efetores). Os genes do clado M35 2.2 foram encontrados em apenas 5 indivíduos. (61 CDSs da M35 1.1 + 38 CDSs da M35 1.2 + 22 CDSs da M35 2.1n + 14 CDSs da M35 2.1e + 5 CDSs da M35 2.2 = 140 CDSs totais).

3.2 Índices utilizados na estimativa do viés no uso de códons

3.2.1 Índice de Adaptação de Códons – CAI

O CAI (Codon Adaptation Index) é uma medida que quantifica a adaptabilidade relativa dos códons, usada entre genes altamente expressos, em outras palavras, ela estima a extensão da tendência para códons que são conhecidos por serem favorecidos em genes altamente expressos (Sharp; Li, 1987). Este índice fornece uma evidência primária sobre a eficiência relativa da tradução (Gustafsson et al., 2012).

Um valor de “adaptação relativa”, w_i , para o códon i é calculado a partir de sua frequência relativa de uso em um conjunto de referência de genes altamente expressos que é próprio para cada espécie; sendo assim:

$$w_i = \frac{RSCU_i}{RSCU_{max}} = \frac{X_i}{X_{max}}$$

Onde $RSCU_{max}$ e X_{max} são os valores RSCU e X para o códon usado com mais frequência para um aminoácido. O CAI para um gene é então definido como a média geométrica dos valores w para códons nesse gene:

$$\text{CAI} = \left(\prod_{i=1}^L w_i \right)^{1/L} \quad \text{or} \quad \exp \left(\frac{1}{L} \sum_{i=1}^L \ln(w_i) \right)$$

Onde L é o número de códons no gene, excluindo metionina, triptofano e códons de parada. O valor CAI pode variar de 0 até 1, com valores mais próximos a 1 representando uma maior proporção dos códons mais abundantes (Sharp; Li, 1987).

3.2.2 Uso Relativo de Códons Sinônimos – RSCU

O RSCU (Relative synonymous codon usage) é calculado obtendo a razão entre a frequência observada de códons em relação a frequência esperada que leva em consideração um uso uniforme entre todos os códons sinônimos (Sharp; Li, 1986).

$$\text{RSCU}_i = \frac{X_i}{\frac{1}{n} \sum_{i=1}^n X_i}$$

Onde X_i é o número de ocorrências do códon i ; e n é a quantidade de códons que codifica o aminoácido (1, 2, 3, 4 ou 6). O códon que obtiver um valor RSCU acima de 1,6 pode ser considerado como super-representado. Por sua vez, o códon que apresentar valor RSCU abaixo de 0,6 pode ser considerado sub-representados (Deb; Uddin; Chakraborty, 2020).

3.2.3 Número Efetivo de Códons – ENC

O ENC (Effective Number of Codons) é uma medida geral do viés do uso igual de códons em gene individuais, logo, o conhecimento dos códons ideais ou do “conjunto de referência” de genes altamente expressos não é necessário (Wright, 1990). Este índice pode ser estimado da seguinte maneira:

$$N_c = 2 + \frac{9}{\bar{F}_2} + \frac{1}{\bar{F}_3} + \frac{5}{\bar{F}_4} + \frac{3}{\bar{F}_6}$$

Onde \bar{F}_k ($k = 2, 3, 4$ ou 6) é a média dos valores de F_k para aminoácidos k vezes degenerados. F_k para cada aminoácido k vezes degenerado pode ser estimado por:

$$F_k = \frac{nS - 1}{n - 1}$$

Onde n é o número total de códons para esse aminoácido, e S é:

$$S = \sum_{i=1}^k \left(\frac{n_i}{n}\right)^2$$

Onde n_i é o número de ocorrências do i -ésimo códon para este aminoácido. Os valores ENC variam entre 20 (quando o uso é não-aleatório) até 61 (quando o uso é aleatório, ou seja, os 61 códons sinônimos são utilizados igualmente). O autor deste índice sugere que valores de ENC abaixo de 35 como sendo de viés significativo (Wright, 1990).

3.2.4 Índice Intrínseco de Desvio de Códon – ICDI

O ICDI (Intrinsic Codon Deviation Index) fornece uma estimativa independente do valor do qui-quadrado com base no RSCU e na degeneração de aminoácidos da sequência. O ICDI é mais útil para estimar o viés de códons em espécies onde os códons ideais são desconhecidos (Freire-Picos et al., 1994). As estimativas do ICDI foram calculadas conforme a seguinte equação:

$$ICDI = \sum_{\alpha \in A} F_{\alpha} S_{\alpha}$$

Onde F_{α} é uma frequência relativa do aminoácido α ; e S_{α} é:

$$S_{\alpha} = \frac{1}{k_{\alpha}(k_{\alpha}-1)} \sum_{c \in C_{\alpha}} (r_{\alpha c} - 1)^2$$

Onde, $r_{\alpha c}$ é o RSCU e k_{α} é a degeneração do aminoácido α . As estimativas do ICDI variam entre 0 (uso igual de todos os códons) até 1 (de um códon para cada aminoácido).

3.2.5 Índice de Viés de Códon – CBI

O CBI (Codon Bias Index) estima a tendência do padrão de uso de códons da sequência de codificação com base no grau de códons preferidos (Bennetzen; Hall, 1982). Este índice pode ser calculado pela seguinte fórmula:

$$CBI = \frac{N_o - N_r}{N_t - N_r}$$

Onde N_o é a ocorrência total de códons superiores na sequência de codificação; N_r é o total de códons superiores quando todos os códons sinônimos são aleatórios e N_t refere-se à frequência de aminoácidos correspondentes aos códons superiores nas sequências de codificação (Bennetzen; Hall, 1982).

Os valores CBI também variam de 0 até 1, sendo que valores CBI igual a 0 refere-se a uma escolha aleatória de códons, enquanto CBI igual 1 indica que a sequência usa principalmente códons preferenciais (Bennetzen; Hall, 1982).

3.2.6 Índice de Ordem do Uso de Códons Sinônimo – SCUO

O (SCUO) (Synonymous Codon Usage Order) quantifica a excentricidade da distribuição uniforme como uma diferença normalizada entre a entropia máxima e a observada (Wan; Xu; Zhou, 2003). O índice SCUO médio para sequências de codificação é calculado pela seguinte fórmula:

$$SCUO = \sum_{i=1}^{n_i} \left(\frac{\sum_{j=1}^{n_i} x_{ij}}{\sum_{i=1}^{18} \sum_{j=1}^{n_i} x_{ij}} \right) SCUO_i$$

Onde j é o códon do aminoácido e $SCUO_i$ é:

$$SCUO_i = \frac{H_i^{max} - H_i}{H_i^{max}}$$

Aqui, $SCUO_i$ é o SCUO do aminoácido em cada sequência e H_i e H_i^{max} são a entropia e o máximo para um i -ésimo aminoácido em uma sequência. Este índice, assim como outros, também varia de 0 a 1, sendo SCUO igual a 0 para nenhum viés e SCUO igual a 1 para alto viés no uso de códons (Wan; Xu; Zhou, 2003).

3.2.7 Frequência Ótima de Códons – FOP

O FOP (Frequency of Optimal Codons) pode ser compreendido como a razão entre o número de códons ótimos e o número total de códons (ótimos ou não) (Ikemura, 1981). Os “códons ótimos” podem ser definidos como aqueles que ocorrem em genes de alta expressão e por consequência, são significativamente mais frequentes no genoma do que códons que ocorrem em genes de baixa expressão

(Stenico; Lloyd; Sharp, 1994). Em outras palavras, os “códon ótimos” são definidos como aqueles cuja frequência aumenta com o aumento do nível de expressão dos genes que os contêm (Duret; Mouchiroud, 1999).

Esta medida pode ser considerada como a medida mais simples do viés de uso de códon próprios da espécie, entretanto, é essencial compreender que o índice FOP depende do contexto ou da espécie, uma vez que os seus valores dependem do código genético da espécie em particular.

$$F_{op} = \frac{X_{op}}{X_{op} + X_{non}}$$

Onde X_{op} é o número de códon “ótimos” e X_{non} é o número de códon “não ótimos” em um gene. Como na maioria dos índices citados, códon de parada e códon para metionina, triptofano e outros aminoácidos com códon ideais indeterminados, são excluídos do cálculo. O FOP também varia de 0 (quando o viés é nulo) até 1 (viés é máximo) (Ikemura, 1981).

3.2.8 Índice de Similaridade de Uso de Códon – COUSIN

O COUSIN (Codon Usage Similarity Index) compara as preferências de uso de códon de uma sequência de consulta, com a referência e normaliza a saída sobre a suposição da hipótese nula de uso aleatório de códon. O COUSIN pode ser calculado como COUSIN₁₈ permitindo a contribuição igual de cada uma das 18 famílias de códon sinônimos para o índice global, ou COUSIN₅₉ que permite a contribuição proporcional de cada família à frequência do aminoácido correspondente nas sequências de consulta (Bourret; Alizon; Bravo, 2019). Para este trabalho, foram calculados ambos os índices.

O COUSIN₁₈ é a razão entre a soma dos pesos de todos os códon sinônimos para este aminoácido no conjunto de dados de consulta sobre a soma correspondente dos pesos no conjunto de dados de referência, logo:

$$\text{COUSIN}_{18}^a = \frac{1}{\mathcal{N}} \times \frac{\sum_{c \in k_a} W_{c,a}^{\text{que}}}{\sum_{c \in k_a} W_{c,a}^{\text{ref}}}$$

Onde n é o número de aminoácidos presentes tanto na consulta quanto na referência e k_a é o conjunto de códons sinônimos que codificam o aminoácido 'a'.

O COUSIN₅₉ pode ser calculado através de uma formula bem semelhante:

$$\text{COUSIN}_{59}^a = f_a^{\text{que}} \times \frac{\sum_{c \in k_a} W_{c,a}^{\text{que}}}{\sum_{c \in k_a} W_{c,a}^{\text{ref}}}$$

Onde f_a^{que} é a frequência do aminoácido 'a' na consulta. A pontuação final do COUSIN é obtida somando as pontuações individuais do COUSIN de todos os aminoácidos.

Os autores desse índice sugerem a seguinte interpretação dos resultados: uma pontuação COUSIN igual 1 indica que o uso preferencial de códons na consulta são semelhantes aos do conjunto de dados de referência; COUSIN igual a 0 indica que o uso preferencial de códons na consulta são semelhantes aos da hipótese nula (ou seja, uso igual de códons sinônimos); valores COUSIN acima de 1 indica uso preferencial de códons na consulta são semelhantes aos da referência, mas de maior magnitude; enquanto entre 0 e 1, o uso preferencial de códons na consulta são semelhantes aos da referência, mas de menor magnitude; por fim, COUSIN abaixo de 0, os uso preferencial de códons na consulta são opostos aos da referência (Bourret; Alizon; Bravo, 2019).

3.3 Gráficos de viés (PR2) e de neutralidade

O gráfico de viés PR2 (Parity rule 2) fornece informações sobre o efeito de mutação ou seleção no viés do uso de códons. Este gráfico foi construído calculando o viés de AT através da seguinte formula $A3/(A3+T3)$ no eixo y e o viés de GC $G3/(G3+C3)$ no eixo x (Sueoka, 1995). Seus resultados fornecem informações sobre a preferência pelo conteúdo de purina (A e G) ou pirimidina (C e T/U) na terceira posição do códon. Os pontos alocados no centro do gráfico, onde o valor de ambas as coordenadas coincide e resulta em 0,5, refletem a ausência de mutação ou viés de seleção. Portanto, qualquer desvio do centro do gráfico PR2 permite estimar se o viés de uso de códons está sendo afetado por mutação, seleção ou ambas.

O gráfico de neutralidade é calculado através dos valores médios de GC1 e GC2 (GC12) em relação a GC3 (Sueoka, 1988). Nos gráficos de neutralidade, a regressão

com inclinação 0 sugere a ausência de pressão de mutação direcional ou restrições seletivas completas. Por outro lado, uma inclinação de 1 indica o mesmo módulo de mutação entre GC12 e GC3 e que a neutralidade completa foi o principal elemento no processo evolutivo (Wu; Zhao; Tao, 2015).

3.4 Softwares utilizados

Para todas as sequencias em estudo, os resultados referentes aos dados de composição de bases nitrogenadas, como a porcentagem de GC total ou então a porcentagem de uma (ou mais) determinada base em uma posição específica na trinca, como por exemplo a %A(3), foram calculados utilizando a versão online do Software COUSIN, disponível em <https://cousin.ird.fr/index.php> (Bourret; Alizon; Bravo, 2019).

Os índices ICDI, CBI, SCUO, FOP, COUSIN₁₈ e COUSIN₅₉ também foram calculados pela versão online do Software COUSIN (Bourret; Alizon; Bravo, 2019). Os demais índices (RSCU, CAI e ENC) foram obtidos através da versão online do Software CAIcal, disponível em <http://genomes.urv.es/CAIcal/> (Puigbò; Bravo; Garcia-Vallve, 2008).

Os gráficos de dispersão (neutralidade e correlações) foram construídos com a versão 4.13 do software Past (Hammer; Harper; Ryan, 2001). Este mesmo Software também foi utilizado nas análises de componentes principais (PCA) e nos agrupamentos. O método de agrupamento hierárquico utilizado foi o UPGMA (Unweighted Pair Group Method using Arithmetic averages) calculado com base nos valores RSCU e utilizando o índice de dissimilaridade de Dice (1945).

Os gráficos de linha também utilizaram os valores do uso relativo de códons sinônimos como base; porém estes gráficos foram construídos no software EXCEL 2016. Este último software (EXCEL 2016) também foi utilizado na construção dos gráficos de viés PR2.

4. RESULTADOS

4.1 Conjunto de dados de referência – 7962 genes SCO

A tabela a seguir contém informações sobre a composição média de bases do conjunto de dados de referência, composto por 7962 SCO obtidos do genoma de *Corynespora cassiicola* (Tabela 1).

Tabela 1. Composições de bases nitrogenadas em 7962 genes ortólogos de cópia única (SCO) de *Corynespora cassiicola*.

	length	%A(3)	%T(3)	%G(3)	%C(3)	%GC(all)	%GC(1)	%GC(2)	%GC(3)
Mínimo	549	0	2,29	9,94	16,27	42,35	36,78	28,52	40,65
Máximo	16041	33,94	44,58	55,14	74,66	70,08	88,83	78,92	94,80
Média	1653	15,47	19,86	28,14	36,53	55,42	57,23	44,36	64,67

**O número entre parênteses refere-se a posição da(s) determinada(s) base no códon.

A porcentagem média de adenina na terceira posição do códon (%A(3)), foi extremamente baixa. Alguns genes apresentaram esse valor inferior a 1%, quando o esperado seria de 25%, tendo em vista que adenina é uma das quatro bases que podem ocupar uma determinada posição ao longo do gene. O gráfico apresentado na figura 2 mostra que essa baixa %A(3) está significativamente correlacionada ao viés de códons calculado pelo ENC (Numero Efetivo de Códons).

Contrastando com os valores obtidos para %A(3), a porcentagem média de citosina na terceira posição do códon, ou seja, %C(3), foi 36,53%. Portanto, citosina foi a base mais abundante, seguido de guanina (com 28,14%). Mesmo com o valor médio da %GC(2) ficando abaixo de 45%, a porcentagem média de GC total (%GC(all)) foi de 55,42, evidenciando novamente preferência por citosina e guanina.

O gráfico de neutralidade é uma ferramenta comumente utilizado para determinar os principais fatores que estão influenciando no viés do uso de códons (figura 1). A inclinação da reta de regressão do gráfico de neutralidade construído com base na porcentagem de GC dos 7962 SCO obtidos do genoma de *Corynespora cassiicola*, foi de 0,076. Isso significa dizer que o viés de códon existente nesses genes é pouco afetado pela evolução neutra, visto que, o efeito da pressão de mutação é de apenas 7,6%, enquanto a influência de outros fatores, como por exemplo a seleção natural, é de 92,4%.

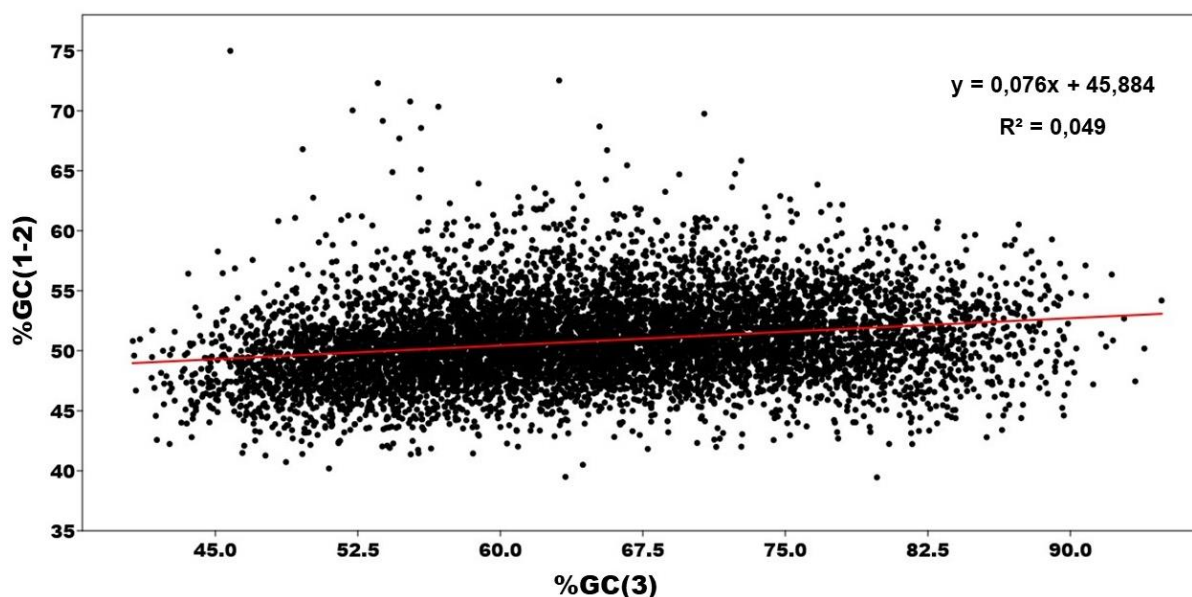


Figura 1. Gráfico de neutralidade relacionando valores de %GC3 ou seja, porcentagem de Guanina ou Citosina na terceira base do códon (ao longo do eixo X) com os valores de %GC12 ou seja, porcentagem média de Guanina ou Citosina na primeira ou segunda base dos códons (ao longo do eixo Y) dos 7962 SCO de *Corynespora cassiicola*.

Os índices apresentados na tabela 2 corroboraram entre si ao inferir informações sobre a preferência no uso de códons de 7962 genes SCO do conjunto de referência. O Número Efetivo de Códons (ENC) entre todos os genes do conjunto de dados de referências variou de 26,4 a 61,0 com valor médio de 51,1 resultado que pode ser compreendido como uso de códons razoavelmente aleatório.

Tabela 2. Principais índices utilizados para quantificar o viés no uso de códons, valores referentes a 7962 SCO de *Corynespora cassiicola*.

	ENC	SCUO	FOP	CBI	ICDI
Valor mínimo	26,4	0,02	0,23	-0,13	0,01
Valor máximo	61,0	0,47	0,84	0,74	0,73
Valor médio	51,1	0,12	0,47	0,21	0,17

ENC = Número Efetivo de Códons; SCUO = Índice de Ordem do Uso de Códons Sinônimo; FOP = Frequência Ideal de Códons; CBI = Índice de Viés de Códons; ICDI = Índice de Desvio de Códons Intrínseco.

Apesar de serem calculados levando em consideração aspectos diferentes, os índices SCUO, CBI e ICDI apresentaram valores médios bem próximos, todos abaixo de 0,25 indicando novamente baixo viés no uso de códons dos 7962 SCO de *Corynespora cassiicola*.

Houve uma correlação positiva entre o valor ENC e a %A(3), sendo que, quanto maior o valor ENC (ou seja, menor viés), maior a porcentagem de adenina na terceira base (figura 2).

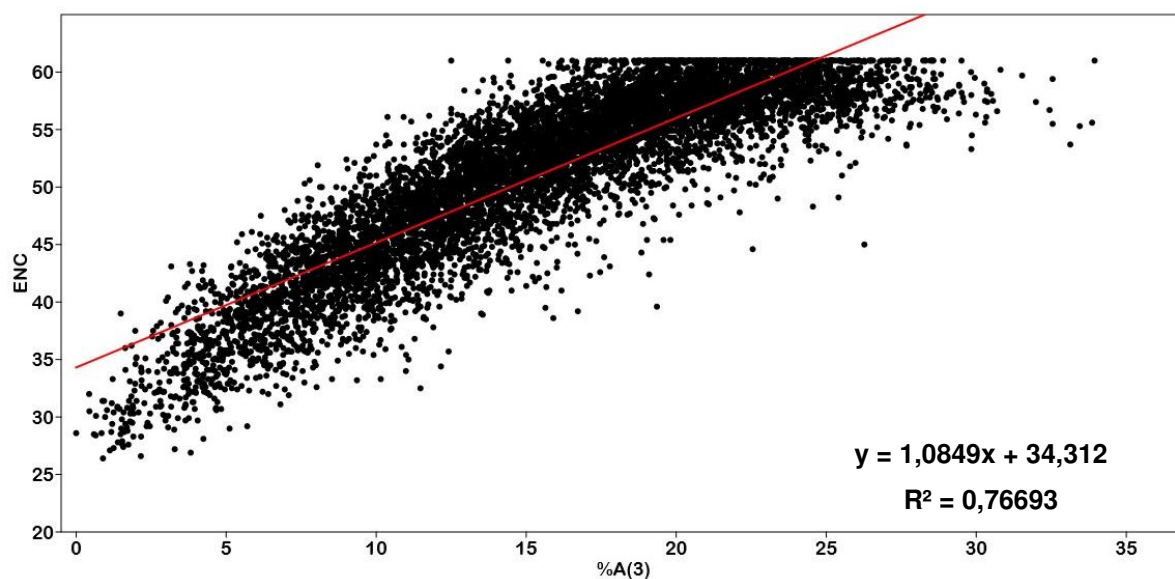


Figura 2. Gráfico de dispersão relacionando valores de %A(3), ou seja, porcentagem de adenina na terceira base dos códons (ao longo do eixo X) com valores de ENC, ou seja, número efetivo de códons (ao longo do eixo Y) dos 7962 SCO de *Corynespora cassiicola*.

A correlação negativa entre os valores de ENC e a %GC(all) mostra que o viés no uso de códons dos 7962 SCO de *Corynespora cassiicola* se deve também ao aumento da %CG(all) (figura 3).

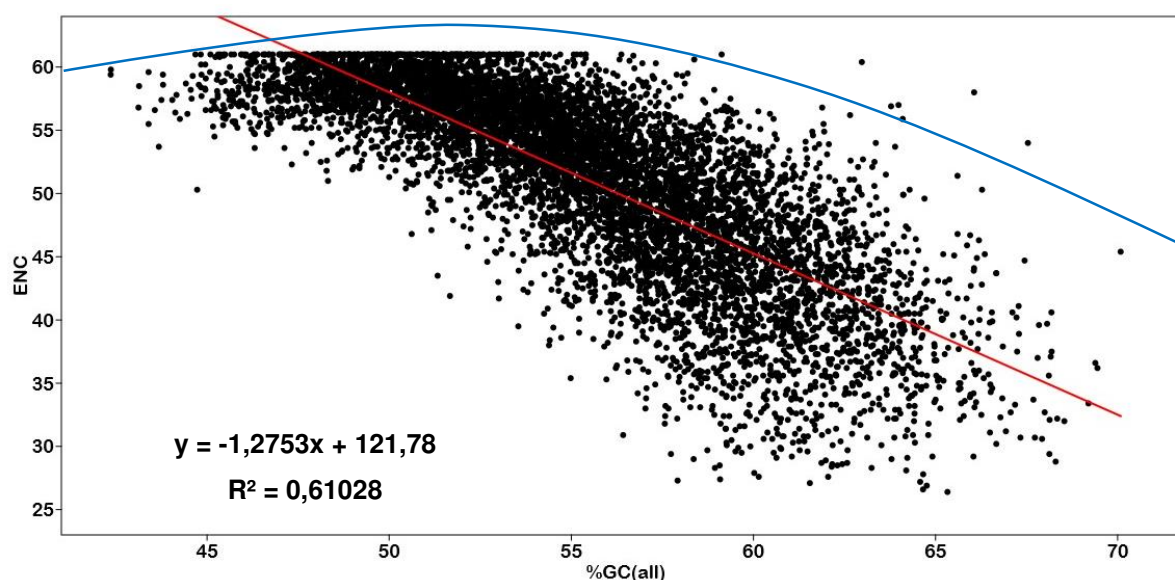


Figura 3. Gráfico de dispersão relacionando valores de %GC(all), ou seja, porcentagem de Guanina e Citosina total (ao longo do eixo X) com valores de ENC, ou seja, número efetivo de códons (ao longo do eixo Y) dos 7962 SCO de *Corynespora cassiicola*.

Em sua imensa maioria, os genes ficaram distribuídos abaixo da curva padrão (linha azul). Tal resultado pode ser interpretado como um viés no uso de códons sendo influenciado/afetado pela seleção natural ou restrições composicionais, entre outros fatores.

O Uso Relativo de Códon Sinônimos (RSCU) médio entre os 7962 genes de *Corynespora cassiicola* analisados variou de 0 a 6 entre todos os códon. Segundo Sharp e Li (1987), códon com valores RSCU entre 0,6 e 1,6 são considerados aleatórios e chamados de códon imparciais. Houve sete códon com os RSCU inferior a 0,6 (TTA, ATA, GTA, CTA, TGT, AGT e AAA). Estes sete códon estão, conseqüentemente, sub-representados no conjunto de sequências destes genes. Já outros cinco códon (CGC, CTC, GGC, GTC e ATC) obtiveram valores RSCU acima de 1,6 e por isso estariam super-representados.

Outro detalhe que chamou a atenção nessa análise foi o fato de que todos os códon super-representados são terminados em C e a maioria ricos em GC. Enquanto que todos os códon sub-representados (RSCU inferior a 0,6) apresentam, no mínimo, duas bases nitrogenadas A ou T em sua composição e sempre são terminados por uma dessas duas bases.

4.2 Família de genes das proteínas Indutoras de Necrose e Etileno – NLP

A composição de bases nitrogenadas das quatro cópias conhecidas dos genes que codificam para as proteínas NLP, provenientes de 44 isolados fúngicos pertencentes a *Corynespora cassiicola*, estão apresentados na tabela 3.

Tabela 3. Composição de bases nitrogenadas dos genes das NLP de 44 isolados de *Corynespora cassiicola*.

NLP	Header	length	%A(3)	%T(3)	%G(3)	%C(3)	%GC(all)	%GC(1)	%GC(2)	%GC(3)
NLP1.1	Min.	714	14,71	26,05	17,65	36,98	53,36	51,68	52,52	55,46
	Max.	714	16,39	28,57	21,43	38,24	54,48	52,10	52,94	58,82
	Média	714	15,86	26,90	19,51	37,73	53,96	52,07	52,56	57,24
NLP1.2A	Min.	717	5,86	17,16	20,08	51,46	58,16	58,16	43,93	71,97
	Max.	717	7,95	20,08	20,92	56,07	59,69	59,00	44,77	76,15
	Média	717	6,44	19,25	20,70	53,61	59,11	58,81	44,20	74,32
NLP1.2B	Min.	732	9,02	30,33	15,16	41,80	55,74	56,97	52,05	56,97
	Max.	732	10,66	33,20	16,80	43,85	56,97	58,61	52,87	60,66
	Média	732	9,94	31,98	15,51	42,58	56,15	57,81	52,55	58,08
NLP1.3	Min.	762	15,35	29,26	22,05	31,48	50,99	52,36	45,19	54,44
	Max.	810	16,30	29,53	22,96	33,07	51,31	53,70	46,46	55,12
	Média	794	15,98	29,35	22,66	32,01	51,14	53,13	45,61	54,67

**O número entre parênteses refere-se a posição da(s) determinada(s) base no códon.

De modo similar da forma como ocorreu com os 7962 SCO, os genes das NLP também exibiram percentual médio de GC(all) acima dos 50%, bem como, baixa percentagem de A(3). Apesar das similaridades de valores médios apresentado pelos diferentes membros da família de genes que codificam para as NLP, é possível notar que o parálogo NLP 1.2A foi a que mais se diferenciou das demais, apresentando uma média de %A(3) baixíssima (6,44), bem como, a maior média de %GC(all) (59,11) e uma percentagem média de GC(3) superior a 74%.

Embora a família dos genes que codificam para NLP seja constituída de cópias de um mesmo gene e obtidas de isolados de uma mesma espécie, diferentes forças evolutivas podem estar atuando e influenciando nas mudanças de padrões no uso de códons entre os membros (figura 4).

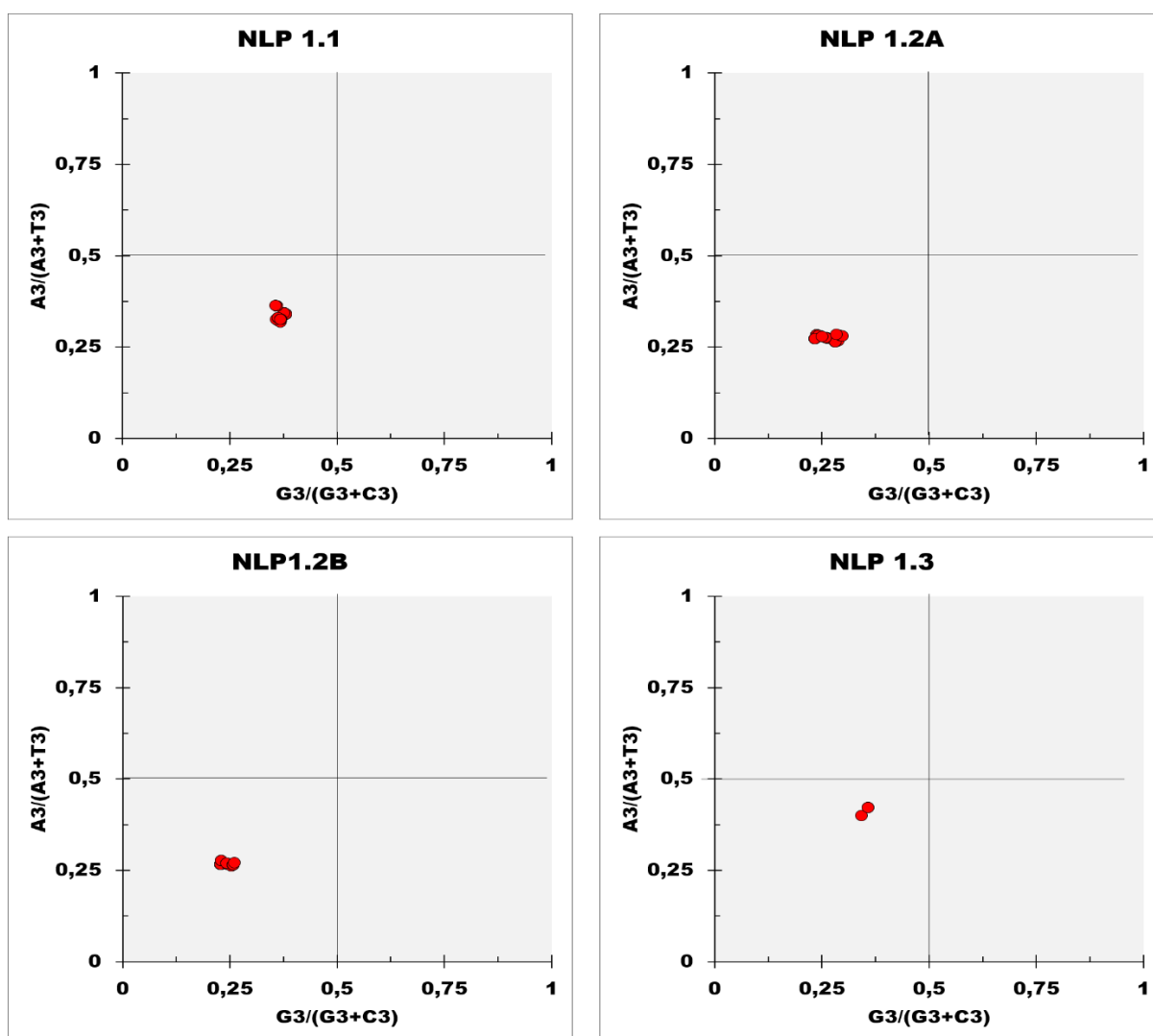


Figura 4. Gráfico de viés PR2 relacionando valores de $G3/(G3+C3)$ (ao longo do eixo X) com valores de $A3/(A3+T3)$ (ao longo do eixo Y) dos genes das NLP de 44 isolados de *Corynespora cassiicola*.

As dispersões gráficas apresentaram-se bastante similares. Elas mostram os genes das NLP no quadrante inferior esquerdo. Isso significa que T e C foram as bases preferidas por todos os membros dessa família. Devido ao fato de que os pontos não ocupam a posição central no gráfico, mas estão desviados do centro, fica evidente que o uso preferencial de códons observado nas sequências não é apenas função da pressão de mutação, mas também da pressão de seleção.

A maioria dos índices apresentados na tabela 4 também corroboraram entre si ao inferir informações sobre a preferência no uso de códons dos parálogos dos genes que codificam as NLP provenientes de 44 isolados *Corynespora cassiicola*. Essa corroboração entre os índices infere maior confiabilidade nas análises.

Tabela 4. Principais índices utilizados para quantificar o viés no uso de códons, valores referentes aos genes das NLP de 44 isolados de *Corynespora cassiicola*.

NLP		COUSIN ₁₈	COUSIN ₅₉	CAI	ENC	SCUO	FOP	CBI	ICDI
NLP1.1	Min.	0,92	0,82	0,75	47,20	0,09	0,43	0,18	0,18
	Max.	1,12	0,94	0,76	50,00	0,11	0,44	0,19	0,24
	Média	1,07	0,91	0,75	49,08	0,10	0,44	0,18	0,22
NLP1.2A	Min.	2,16	2,15	0,83	38,60	0,18	0,63	0,44	0,36
	Max.	2,40	2,46	0,85	43,40	0,24	0,67	0,50	0,43
	Média	2,33	2,32	0,84	41,71	0,19	0,65	0,46	0,40
NLP1.2B	Min.	0,86	1,08	0,75	46,00	0,10	0,46	0,21	0,20
	Max.	1,03	1,27	0,76	48,40	0,17	0,49	0,26	0,23
	Média	0,92	1,16	0,75	47,40	0,13	0,47	0,23	0,21
NLP1.3	Min.	0,65	0,56	0,72	54,50	0,09	0,39	0,10	0,16
	Max.	0,70	0,60	0,73	57,10	0,10	0,41	0,12	0,19
	Média	0,67	0,58	0,72	56,23	0,09	0,40	0,11	0,17

COUSIN = Índice de Similaridade de Uso de Códons; CAI = Índice de adaptação de códons; ENC = Número Efetivo de Códons; SCUO = Índice de Ordem do Uso de Códons Sinônimo; FOP = Frequência Ideal de Códons; CBI = Índice de Viés de Códons; ICDI = Índice de Desvio de Códons Intrínseco.

Novamente, o parálogo NLP 1.2A foi a que mais se distanciou das demais em todas os índices calculados, sendo a que apresentou o maior viés no uso de códons (menor ENC e maior SCUO, FOP, CBI e ICDI). Assim como conjunto de dados de referência deste estudo, os índices ICDI, CBI e SCUO apresentaram valores médios bem próximos, a maioria abaixo de 0,25, com algumas exceções para o membro NLP 1.2A.

Quanto ao Índice de Similaridade de Uso de Códons (COUSIN), tanto o indicador que permite a contribuição igual de cada uma das 18 famílias de códons sinônimos para o índice global (COUSIN₁₈), quanto o que permite a contribuição proporcional de cada família à frequência do aminoácido (COUSIN₅₉) apresentaram valores bem próximos.

Segundo Bourret; Alizon e Bravo, (2019) uma pontuação COUSIN de 1 indica que o uso preferencial de códons dos dados consultados é semelhante aos do conjunto de referência. Isso se aplica para os parálogos NLP 1.1 e NLP1.2B. Valores COUSIN maiores (NLP 1.2A) e menores (NLP 1.3) que 1, são considerados semelhantes ao conjunto de dados de referência, porém em maior e menor magnitude, respectivamente. Os valores COUSIN apresentado pelo membro NLP 1.3 são compreendidos entre 0 e 1 e podem ser considerados semelhantes aos da referência, porém, de menor magnitude. Assim como o clado NLP 1.2A que apresentou COUSIN acima de 1, considerado semelhantes aos da referência, mas de maior magnitude.

No geral, os valores médios do uso relativo de códons sinônimos (RSCU) apresentaram um padrão de distribuição semelhante entre os 59 códons (figura 5).

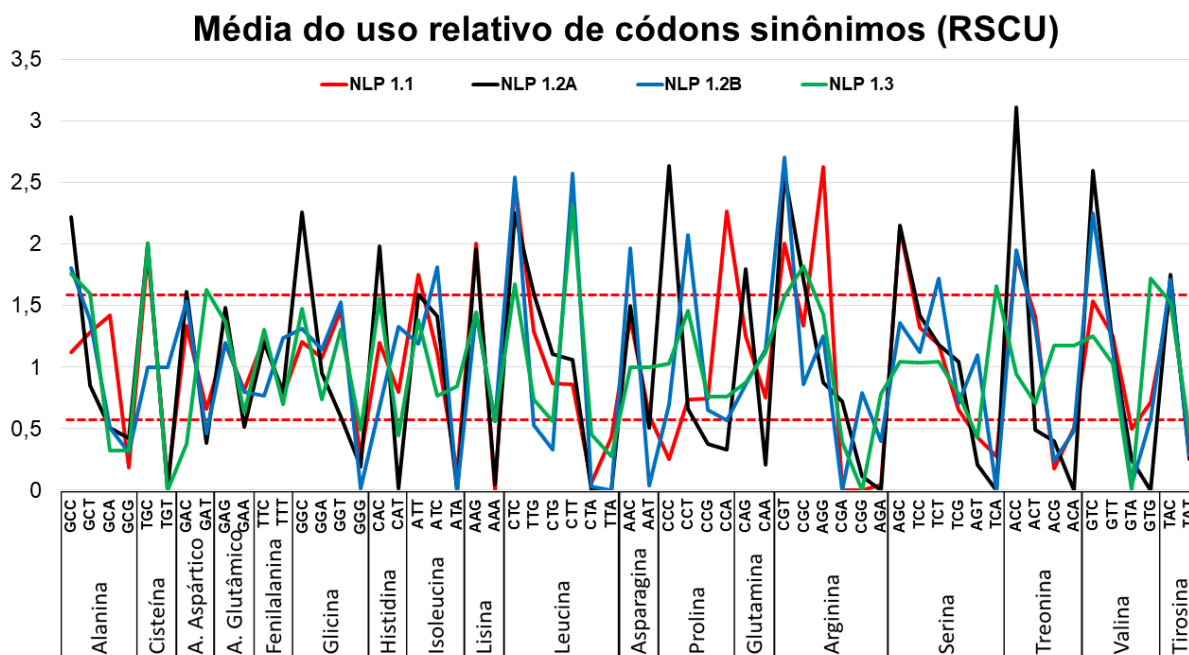


Figura 5. Comparação do Uso Relativo de Códons Sinônimos (RSCU) de 59 aminoácidos dos genes das NLP de 44 isolados de *Corynespora cassiicola*.

O parálogo NLP 1.2A exibiu a maior variação dos valores RSCU entre todos os cladados (variando entre 0 e 3,11). Esta também, foi o membro que apresentou maior número, tanto de códons sub-representados (um total de 26), quanto de códons super-representados (15) entre todas as cópias do gene das NLP. Enquanto isso, o membro NLP 1.3 apresentou a menor quantidade de códons sub-representados (15) e super-representados (8), seguido do clado NLP 1.1 com 17 e 9, respectivamente.

Nota-se que em 100% dos membros da família de genes das NLP, os códons considerados como sub-representados possuem predominância de aminoácidos A/T (28 A/T e 23 C/G em NLP 1.1; 44 A/T e 34 C/G em NLP 1.2A; 35 A/T e 25 C/G em NLP 1.2B; 23 A/T e 22 C/G em NLP 1.3). Do mesmo modo, há predominância de C/G nos códons considerados super-representados, novamente em todos os cladados (12 A/T e 15 C/G em NLP 1.1; 13 A/T e 32 C/G em NLP 1.2A; 15 A/T e 18 C/G em NLP 1.2B; 9 A/T e 15 C/G em NLP 1.3).

O códon CCC (Prolina) apresentou uma variação considerável, sendo classificado como sub-representados na NLP 1.1 e super-representados na NLP 1.2A. Já os códons TCA (Serina) e GTG (Valina) foram classificados como super-representados na NLP 1.3 e como sub-representados em outros membros. Tais variações se mostraram ainda mais evidentes entre todos os códons, após a realização da análise de componentes principais (figura 6).

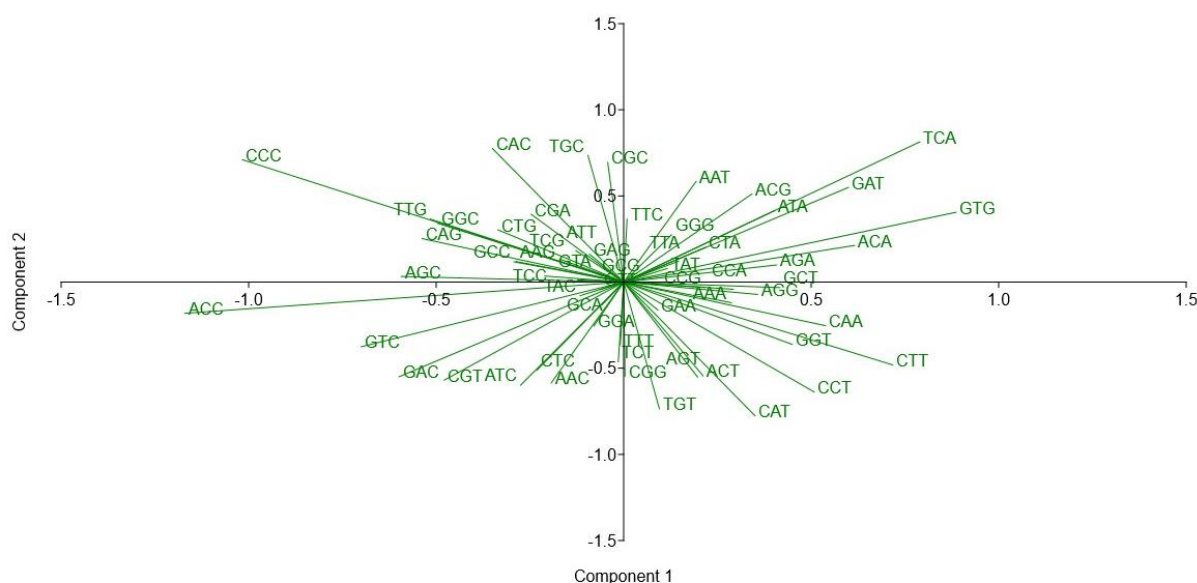


Figura 6. Análise de componentes principais (PCA) com base nos valores médios do RSCU (uso relativo dos códons sinônimos) calculados a partir das cópias do gene das NLP em 44 isolados de *Corynespora cassiicola*.

O agrupamento UPGMA construído com base nos valores médios do RSCU (uso relativo dos códons sinônimos) calculados a partir das cópias do gene das NLP em 44 isolados de *Corynespora cassiicola* esta representados na figura 7.

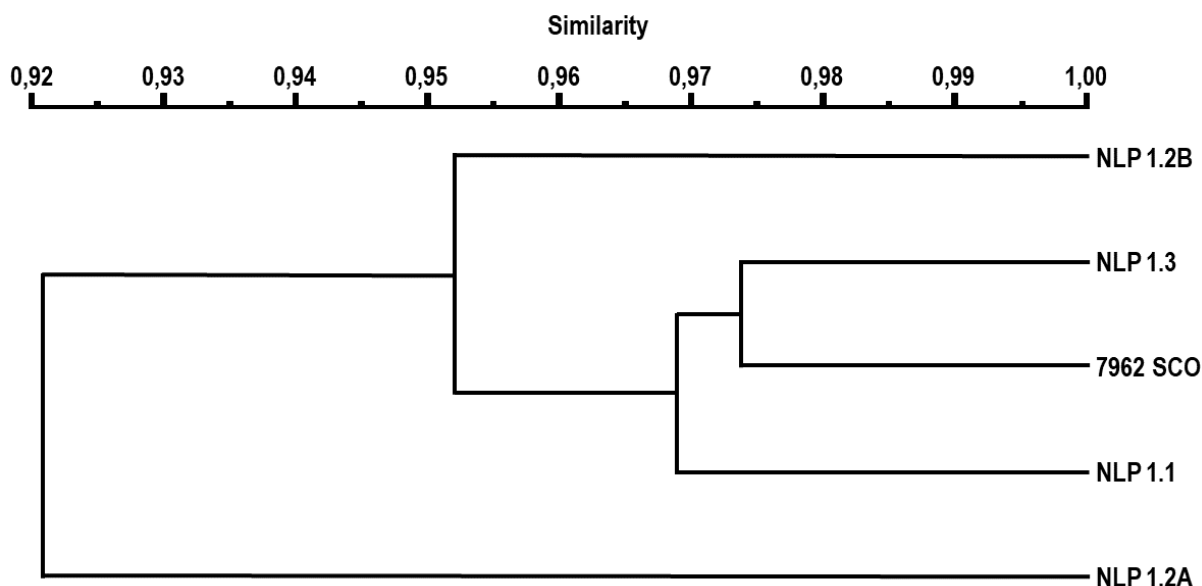


Figura 7. Agrupamento UPGMA com base nos valores médios do RSCU dos genes das NLP de 44 isolados de *Corynespora cassiicola*. Índice de similaridade de Dice (1945).

Corroborando com as demais análises, o agrupamento também mostrou que os parálogos NLP1.3 e 1.1 apresentam padrões de uso de códons mais próximos dos 7962 SCO do conjunto de dados de referência. Do mesmo modo, novamente o membro NLP 1.2A foi quem apresentou maior índice de dissimilaridade.

4.3 Família de genes das enzimas Pectinas Metilesterases – PME

Durante as análises realizadas na família de genes das enzimas Pectinas Metilesterases (PME) foi possível constatar que a porcentagem média de GC total (%GC(all)) variou de 46,03 no clado PME1.2 até 65,02 no clado PME 1.1. Até aqui, tanto os 7962 SCO, que compõe o conjunto de dados de referência, quanto as NLP, todas as médias tinham exibido o percentual médio de GC(all) acima dos 50%. Outro destaque foi percentual de A(3), que nos genes analisados anteriormente (7962 SCO e família de genes das NLP) sempre apresentou médias, no máximo, até a casa dos 15%, entretanto, na tabela 5, o parálogo PME 1.2 apresentou esse índice próximo ao valor esperado em normalidade (Observado: 22,10 / Esperado: 25). Contudo, na

maioria dos parálogos dos genes das PME, o %A(3) ainda pode ser considerado baixo.

As principais informações acerca da composição de bases nitrogenadas dos cinco parálogos que são membros da família de genes que codificam para PME provenientes de 60 isolados de *Corynespora cassiicola*, estão contidas na tabela 5.

Tabela 5. Composição de bases da família de genes das PME de 60 isolados de *Corynespora cassiicola*.

PME	Header	length	%A(3)	%T(3)	%G(3)	%C(3)	%GC(all)	%GC(1)	%GC(2)	%GC(3)
PME 1.1	Min.	1215	6,42	6,17	28,06	53,48	62,92	54,82	49,44	82,18
	Max.	1347	9,80	8,58	31,85	57,28	65,51	57,31	53,83	86,91
	Média	1228,8	7,11	6,83	30,22	55,84	65,02	55,74	53,26	86,07
PME 1.2	Min.	1080	21,39	31,76	13,91	28,89	45,23	45,23	46,11	43,57
	Max.	1227	23,12	33,92	16,39	31,43	46,67	46,98	48,05	46,39
	Média	1162,4	22,10	32,58	15,05	30,27	46,29	46,03	47,51	45,32
PME 2.1	Min.	1062	17,27	24,79	18,01	35,25	49,12	48,59	40,95	53,45
	Max.	1704	19,54	27,01	20,07	38,16	51,88	55,63	44,06	57,94
	Média	1424,9	18,64	25,63	19,12	36,62	50,42	52,42	43,09	55,74
PME 2.2	Min.	1002	10,78	17,96	19,76	48,20	55,69	48,80	49,40	68,86
	Max.	1002	12,58	20,36	21,26	50,00	56,39	49,40	49,70	70,36
	Média	1002	11,62	18,39	20,44	49,56	56,17	49,07	49,45	69,99
PME 2.3	Min.	750	14,04	32,24	23,26	25,38	47,60	43,20	48,29	48,64
	Max.	1050	17,83	35,20	24,32	27,16	48,94	46,29	51,06	51,03
	Média	966,6	16,56	33,47	23,98	25,99	48,39	45,52	49,70	49,96

**O número entre parênteses refere-se a posição da(s) determinada(s) base no códon.

No geral, o membro PME 2.1 foi quem exibiu os índices médios mais próximos aos apresentados pelos 7962 SCO. Enquanto que os membros PME 2.2 e principalmente a PME 1.1 apresentaram os primeiros indícios de possuir os maiores vieses entre todos os membros da família de genes das PME, visto que apresentaram percentuais contrastantes em vários índices.

Os gráficos da figura 8 mostram que com exceção da PME 1.1, todos os outros clados apresentaram seus pontos distribuídos no quadrante inferior esquerdo. Isso significa que T e C foram as bases preferidos por esses membros. Além disso, devido ao padrão de distribuição dos pontos, podemos concluir que as forças que estão moldando o padrão de uso de códons nesses parálogos não se tratam apenas de pressão de mutação, mas também da pressão de seleção.

O clado PME 1.1 foi o único dentre todos os membros das três famílias gênicas estudadas que apresentou a maior parte de seus pontos distribuídos fora do quadrante inferior esquerdo.

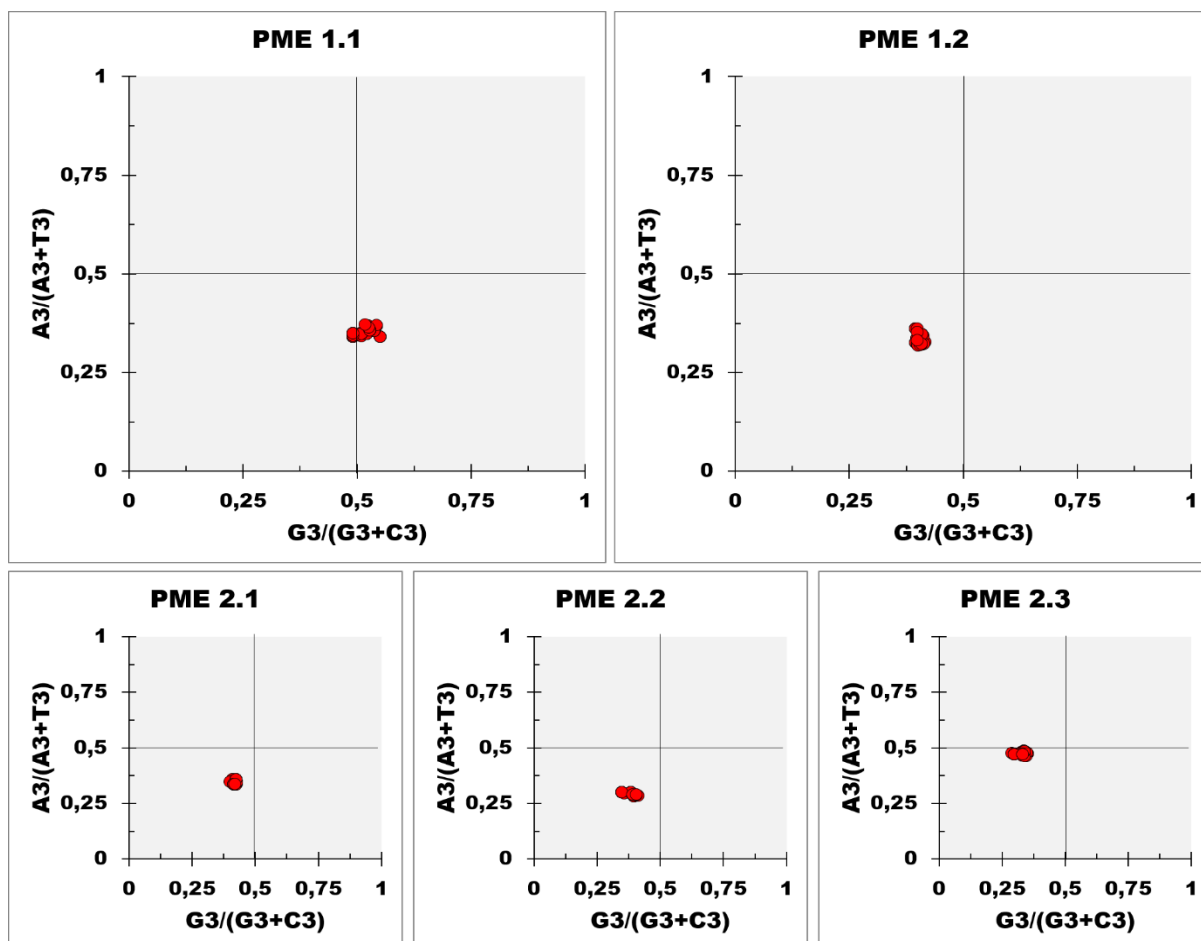


Figura 8. Gráfico de viés PR2 relacionando valores de $G3/(G3+C3)$ (ao longo do eixo X) com valores de $A3/(A3+T3)$ (ao longo do eixo Y) dos genes das PME de 60 isolados de *Corynespora cassiicola*.

Todos os índices corroboraram entre si ao apontar a existência de um viés relativamente baixo entre os parálogos da família de genes que codifica para as Pectinas Metilesterases. O parálogo PME 1.1 foi o membro que apresentou o maior viés em relação ao uso de códons. O membro PME 2.2 apresentou o segundo maior viés (tabela 6).

Novamente o clado PME 2.1 apresentou todos os índices mais parecidos com os 7962 SCO, entretanto esse não foi o membro de menor viés na família. Este posto ficou para os parálogos PME 1.2 e 2.3 que além de apresentarem os valores de todos os índices bem próximos estatisticamente, também exibiram os menores vieses dessa família (tabela 6).

Tabela 6. Principais índices utilizados para quantificar o viés no uso de códons, valores referentes a família de genes das PME em 60 isolados de *Corynespora cassiicola*.

PME		COUSIN ₁₈	COUSIN ₅₉	CAI	ENC	SCUO	FOP	CBI	ICDI
PME 1.1	Min.	1,858	2,056	0,815	36,0	0,220	0,574	0,382	0,302
	Max.	2,255	2,384	0,839	38,3	0,260	0,613	0,443	0,381
	Média	2,119	2,271	0,831	36,9	0,241	0,599	0,422	0,348
PME 1.2	Min.	-0,381	-0,112	0,672	53,8	0,056	0,312	-0,028	0,075
	Max.	-0,013	0,126	0,684	60,6	0,095	0,346	0,016	0,125
	Média	-0,136	0,045	0,679	58,6	0,067	0,334	0,001	0,087
PME 2.1	Min.	0,487	0,475	0,716	53,7	0,066	0,430	0,135	0,078
	Max.	0,785	0,824	0,732	60,5	0,088	0,473	0,199	0,149
	Média	0,591	0,619	0,725	55,7	0,073	0,449	0,165	0,103
PME 2.2	Min.	1,665	1,717	0,8	44,2	0,169	0,549	0,336	0,256
	Max.	1,815	1,802	0,809	46,4	0,220	0,564	0,357	0,333
	Média	1,713	1,769	0,806	44,8	0,206	0,559	0,350	0,306
PME 2.3	Min.	-0,240	-0,120	0,677	52,7	0,049	0,303	-0,019	0,053
	Max.	-0,099	-0,026	0,697	60,7	0,081	0,335	0,011	0,117
	Média	-0,170	-0,069	0,685	57,4	0,061	0,320	-0,001	0,080

COUSIN = Índice de Similaridade de Uso de Códons; CAI = Índice de adaptação de códons; ENC = Número Efetivo de Códons; SCUO = Índice de Ordem do Uso de Códons Sinônimo; FOP = Frequência Ideal de Códons; CBI = Índice de Viés de Códons; ICDI = Índice de Desvio de Códons Intrínseco.

Novamente, os valores médios do uso relativo de códons sinônimos (RSCU) apresentaram um padrão de distribuição bem similar entre os 59 códons (figura 9).

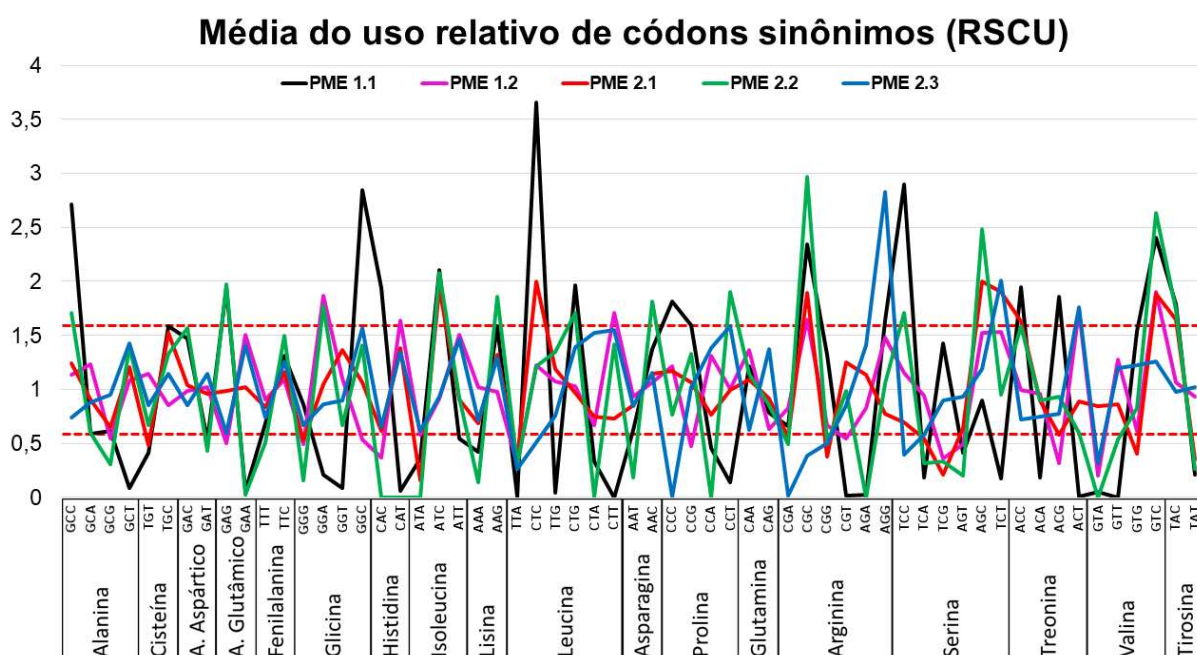


Figura 9. Comparação do Uso Relativo de Códons Sinônimos (RSCU) de 59 códons dos genes das PME em 60 isolados de *Corynespora cassiicola*.

Discorrendo a respeito das maiores variações, no gráfico anterior é possível notar que, enquanto praticamente todos os genes das PME tem um pico considerável ao passar pela trinca CGC (arginina), o clado PME 2.3 (linha azul, figura 9) fica abaixo de 0,5. Outras variações destacáveis podem ser observadas nas trincas CTC (leucina) e GGC (glicina), estas ocasionadas pelos picos altos e discrepantes do parálogo PME 1.1 já mencionados.

O agrupamento UPGMA calculados a partir nos valores médios do RSCU (uso relativo dos códons sinônimos) com base nas cópias do gene das PME em 60 isolados de *Corynespora cassiicola* dividiu as sequencias em três grupos (figura 11).

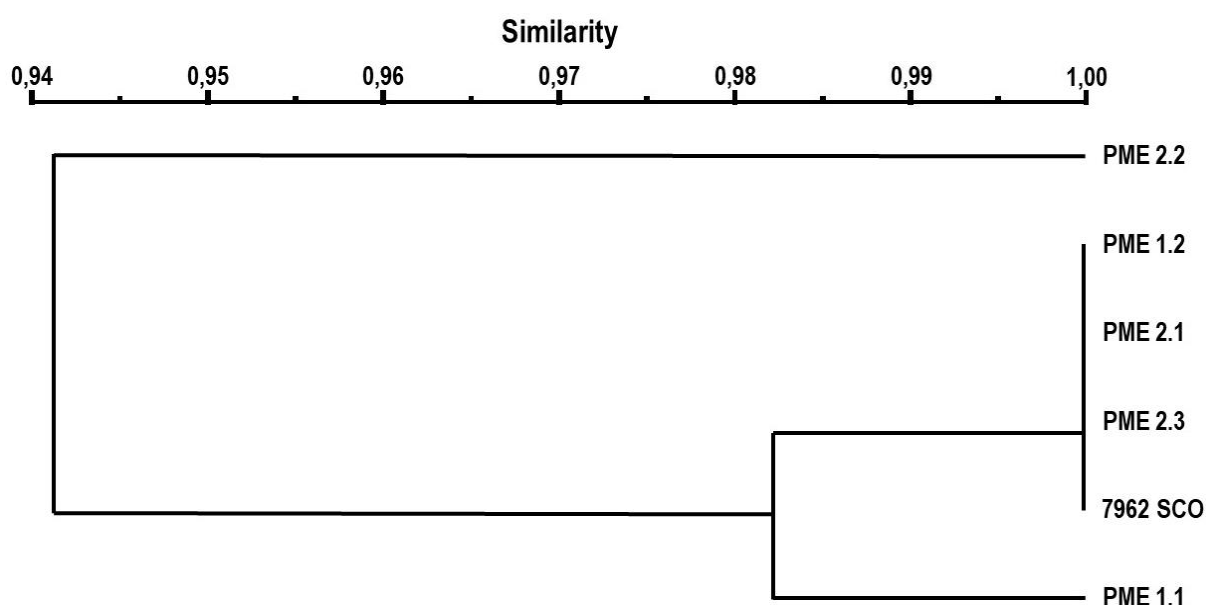


Figura 11. Agrupamento UPGMA com base nos valores médios do RSCU dos membros da família de genes das PME de 60 isolados de *Corynespora cassiicola*. Índice de similaridade de Dice (1945).

Corroborando com as demais análises, o agrupamento também mostrou que os parálogos PME 2.2 e principalmente a PME 1.1 são as mais distintas. As demais PME foram agrupadas no mesmo grupo dos genes do conjunto de dados de referência.

4.4 Família de genes das enzimas Deuterolisinas Metaloproteases – M35

Diferente das outras famílias gênicas deste estudo, nas M35, a falta de um padrão claro dificultou na hora de eleger o clado mais parecido com o conjunto de dados de referência apenas com informações acerca da composição de bases nitrogenadas (tabela 7).

Tabela 7. Composição de bases nitrogenadas dos genes das M35 de 61 isolados de *Corynespora cassiicola*.

M35	Header	length	%A(3)	%T(3)	%G(3)	%C(3)	%GC(all)	%GC(1)	%GC(2)	%GC(3)
M35 1.1	Min.	1059	7,93	25,21	19,26	45,04	56,09	52,98	50,14	64,87
	Max.	1059	8,78	26,91	20,40	46,74	57,22	53,54	51,56	66,57
	Média	1059	8,37	26,14	19,83	45,66	56,42	53,24	50,53	65,49
M35 1.2	Min.	963	19,50	32,20	16,51	23,08	43,00	48,43	39,25	40,74
	Max.	1464	24,79	35,04	21,71	27,57	47,06	53,89	40,98	47,85
	Média	1263	21,59	33,63	19,54	25,24	45,60	51,75	40,27	44,78
M35 2.1n	Min.	1341	6,10	23,75	17,94	43,20	55,90	54,47	49,44	61,29
	Max.	2007	12,26	26,62	19,91	50,55	58,39	55,71	52,02	70,15
	Média	1502	7,91	25,77	19,35	46,97	57,25	55,16	50,26	66,32
M35 2.1e	Min.	1050	2,00	25,43	16,29	47,22	57,33	55,12	49,43	65,71
	Max.	1671	8,80	26,00	18,49	56,00	59,71	56,86	51,17	72,57
	Média	1183	3,52	25,77	16,96	53,75	59,06	56,46	50,01	70,71
M35 2.2	Min.	1050	16,29	37,71	17,43	28,57	46,00	46,29	45,71	46,00
	Max.	1050	16,29	37,71	17,43	28,57	46,00	46,29	45,71	46,00
	Média	1050	16,29	37,71	17,43	28,57	46,00	46,29	45,71	46,00

**O número entre parênteses refere-se a posição da(s) determinada(s) base no códon.

O parálogo M35 2.2 que havia sido encontrada em apenas cinco indivíduos, apresentou valores bem próximos estatisticamente aos do clado M35 1.2. Esses dois membros foram exceções a baixa percentagem de A(3) e o percentual médio de GC(all) acima dos 50%, aparentemente característicos da espécie. Tal fato, fez com que essas M35 (2.2 e principalmente a 1.2) fossem as que apresentaram os percentuais que mais se diferenciam das demais.

A baixíssima percentagem média de A(3) (3,52%) e o percentual médio de GC(3) acima dos 70% fazem do membro M35 2.1e um candidato promissor ao posto de membro de maior viés no uso de códons. Essa teoria ganha mais embasamento ao analisar os índices da tabela 8.

No gráfico de viés (PR2) Todos os parálogos dos genes das M35 ficaram dispostos no quadrante inferior esquerdo, sendo o membro M35 1.2 o que mais se aproximou do centro (figura 12).

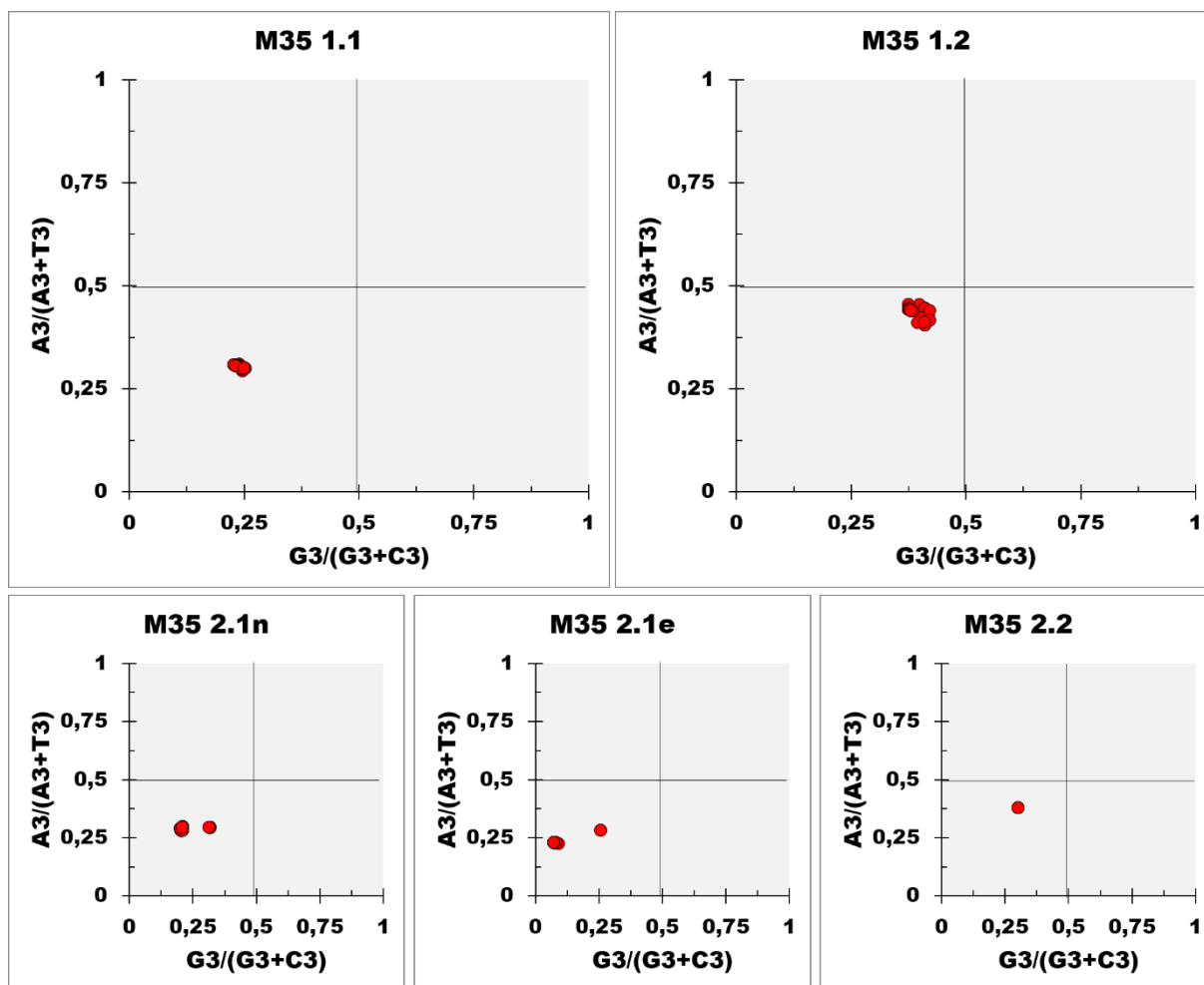


Figura 12. Gráfico de viés PR2 relacionando valores de $G3/(G3+C3)$ (ao longo do eixo X) com valores de $A3/(A3+T3)$ (ao longo do eixo Y) dos genes das M35 de 61 isolados de *Corynespora cassiicola*.

Novamente, as bases T e C foram as preferidas por todos os membros da família gênica M35. Pelo padrão de distribuição dos pontos, podemos inferir que as forças que estão moldando o padrão de uso de códons nesses parálogos são resultantes de pressão de mutação e também da pressão de seleção.

A tabela 8 contém informações acerca do viés existente no uso de códons dos membros da família de genes das M35. Os principais índices utilizados para quantificar o viés no uso de códons concordaram entre si. Os parálogos M35 1.2 e 2.2 novamente apresentaram valores bem parecidos entre si, e foram os membros que exibiram os menores vieses. Em contrapartida, a M35 2.1e foi o clado que apresentou maior viés na grande maioria dos índices utilizados, seguidos desse, destacam-se também os membros M35 1.1 e M35 2.1n.

Tabela 8. Principais índices utilizados para quantificar o viés no uso de códons, valores referentes aos genes das M35 de 61 isolados de *Corynespora cassiicola*.

M35		COUSIN ₁₈	COUSIN ₅₉	CAI	ENC	SCUO	FOP	CBI	ICDI
M35 1.1	Min.	1,888	1,669	0,800	42,900	0,163	0,526	0,311	0,300
	Max.	2,113	1,796	0,807	46,300	0,188	0,540	0,333	0,372
	Média	1,991	1,732	0,803	43,880	0,178	0,534	0,321	0,349
M35 1.2	Min.	-0,561	-0,494	0,650	50,700	0,065	0,267	-0,117	0,070
	Max.	0,114	0,085	0,691	56,400	0,104	0,336	-0,006	0,141
	Média	-0,149	-0,128	0,675	52,882	0,080	0,308	-0,048	0,100
M35 2.1n	Min.	1,469	1,441	0,761	41,000	0,167	0,495	0,267	0,192
	Max.	2,205	2,096	0,817	47,900	0,252	0,578	0,388	0,402
	Média	1,927	1,824	0,796	43,436	0,217	0,544	0,339	0,321
M35 2.1e	Min.	1,840	1,780	0,785	31,700	0,190	0,538	0,328	0,265
	Max.	2,730	2,507	0,849	43,600	0,222	0,646	0,481	0,561
	Média	2,514	2,316	0,834	34,657	0,210	0,619	0,443	0,483
M35 2.2	Min.	-0,002	0,096	0,688	54,500	0,081	0,352	0,043	0,090
	Max.	-0,002	0,096	0,688	54,500	0,081	0,352	0,043	0,090
	Média	-0,002	0,096	0,688	54,500	0,081	0,352	0,043	0,090

COUSIN = Índice de Similaridade de Uso de Códons; CAI = Índice de adaptação de códons; ENC = Número Efetivo de Códons; SCUO = Índice de Ordem do Uso de Códons Sinônimo; FOP = Frequência Ideal de Códons; CBI = Índice de Viés de Códons; ICDI = Índice de Desvio de Códons Intrínseco.

Os resultados apresentados pelos diferentes quantificadores de viés expostos na tabela acima fornecem luz a dúvida que tinha ficado acerca do clado mais similar aos 7962 SCO. Comparando todos os membros, foi possível notar que o parálogo M35 1.2 possui os índices mais parecidos com os genes do conjunto de dados de referência.

Os valores médios do uso relativo de códons sinônimos (RSCU) da família de genes que codificam para as enzimas deuterolisina metaloprotease em *Corynespora cassiicola* também apresentaram no geral, um padrão de distribuição semelhante entre os 59 códons (figura 13). As famílias analisadas anteriormente (NLP e PME) também apresentaram um padrão de distribuição semelhante entre os 59 códons na maioria de seus respectivos membros.

outros (figura 14). No gráfico da figura 13 é possível notar que o maior responsável pelas altas variações envolvendo a aminoácido arginina é o membro M35 2.2.

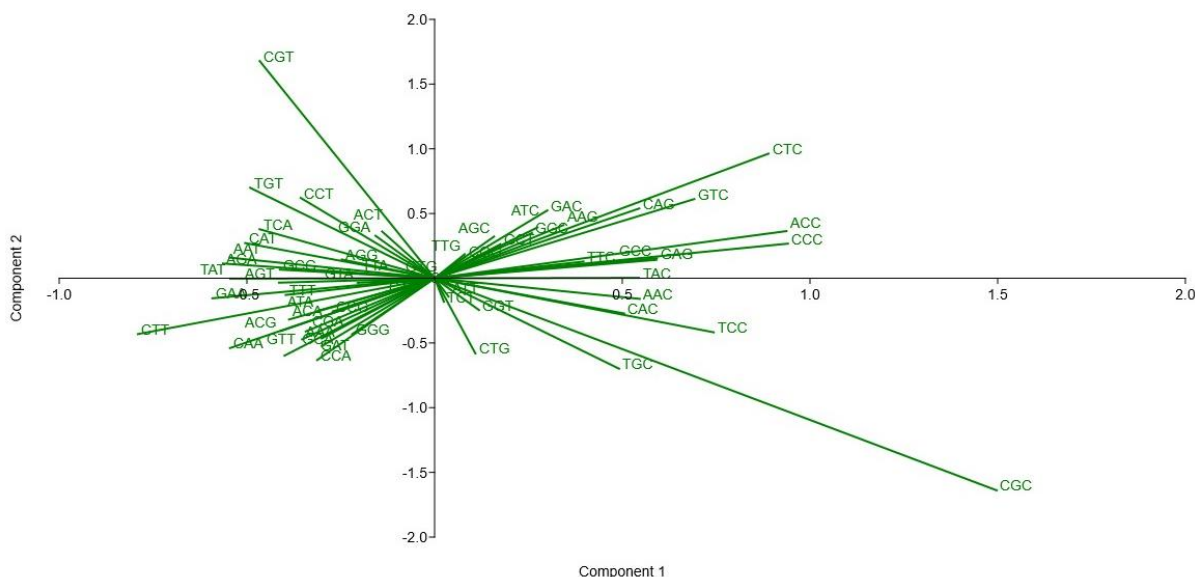


Figura 14. Análise de componentes principais (PCA) com base nos valores médios do RSCU (uso relativo dos códons sinônimos) calculados a partir dos membros da família de genes das M35 provenientes de 61 isolados de *Corynespora cassiicola*.

Os valores médios do RSCU (uso relativo dos códons sinônimos) também foram utilizados na construção de um agrupamento UPGMA estabelecido com base no índice de dissimilaridade de Dice utilizando as sequências dos genes que codificam para as enzimas deuterolisina metaloprotease em 61 isolados de *Corynespora cassiicola* (figura 15).

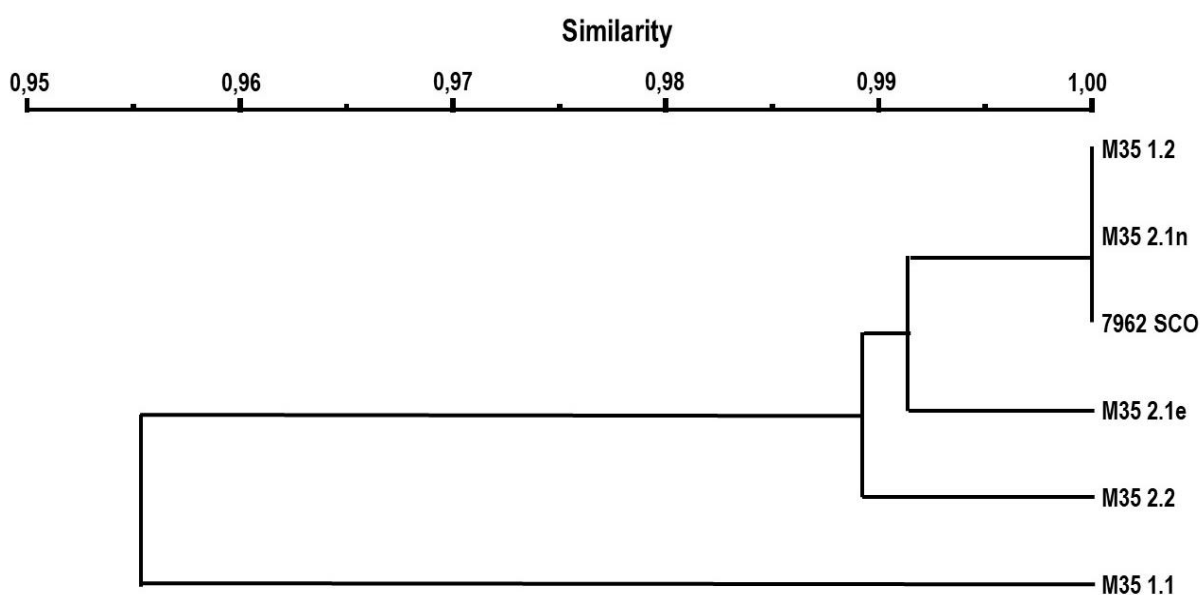


Figura 15. Agrupamento UPGMA com base nos valores médios do RSCU dos genes das M35 de 61 isolados de *Corynespora cassiicola*. Índice de similaridade de Dice (1945).

Pelo agrupamento UPGMA podemos concluir que os membros M35 1.2 e 2.1n são as que possuem um uso relativo de códons sinônimos mais similar aos 7962 SCO do conjunto de dados de referência. O clado M35 1.1 por sua vez, foi quem apresentou o uso relativo de códons sinônimos mais distinto dos demais.

5. DISCUSSÃO

5.1 Padrões no uso de códons e suas implicações

No geral, a grande maioria dos índices e análises concordaram em mostrar que, apesar de baixo, existe sim restrições composicionais e viés no uso de códons em *Corynespora cassiicola*. A otimização de códons é um fenômeno comum na evolução de parasitas que os ajuda a se adaptar e colonizar em seus respectivos hospedeiros (Badet et al., 2017).

O Número Efetivo de Códons (ENC) é uma medida simples, porém bastante eficaz em estimar o viés no uso de códons. Este índice, assim como os outros (SCUO, FOP, CBI e ICDI), também mostraram eficiência em distinguir os membros com relação ao viés. O valor médio de ENC também foi alto (53,27 a 54,45) nos mais de 90 mil genes presentes nos genomas de cinco espécies de fungos do gênero *Puccinia*, composto exclusivamente por fungos patógenos vegetais (Roy; Van Staden, 2019).

Correlacionar a %A(3) e %GC(all) com o ENC nos mostrou que grande parte do viés ocorre pela preferência dos genes por G ou C em detrimento a porcentagem de A, principalmente na última posição da trinca. Conhecer e estudar o nucleotídeo presente na terceira posição da trinca é importante, pois este pode ser determinante na definição do aminoácido.

A preferência por códons terminados em C ou G, bem como a %GC total acima dos 50% evidenciado na grande maioria dos genes de *Corynespora cassiicola* do presente estudo, parecem ser características comum em várias espécies do filo Ascomycota, pois, também foram encontradas em *Alternaria alternata*, *Bipolaris maydis*, *Bipolaris oryzae*, *Colletotrichum graminicola*, *Fusarium graminearum*, *Gaeumannomyces tritici*, *Pyricularia oryzae* e *Verticillium dahliae* (Bansal et al., 2022); *Aspergillus flavus*, *Aspergillus niger*, *Fusarium oxysporum* e *Colletotrichum*

gloeosporioides (Gupta; Singh, 2021); *Claviceps purpúrea* e *Zymoseptoria tritici* (Roy; Van Staden, 2019); *Beauveria bassiana* (Muthabathula; Suneetha; Grace, 2018).

A alta %C(3) observada na maioria dos genes em estudo, pode ser justificada pelo fato de que, segundo Rocha (2004), a maioria dos anticódons de aminoácidos degenerados duplos, geralmente preferem a guanina na primeira posição e a disponibilidade de citosina na terceira posição do códon seria combiná-los adequadamente e, subsequentemente, aumentar a velocidade e a precisão da tradução. Outro motivo que poderiam explicar essa preferência, é que códons ricos em GC codificam aminoácidos simples e biossinteticamente mais baratos falando em termos do gasto de energia associado a biossíntese (Smith; Chapman, 2010). Além disso, a otimização da base nitrogenada final da trinca para C ou G pode também, favorecer a alta expressão de genes envolvidos na invasão de patógenos no hospedeiro e, assim, auxiliar no estabelecimento do parasitismo (Gupta; Singh, 2021).

Economia e principalmente eficiência parecem ser características desejáveis quando lembramos que os genes deste estudo são genes de cópia única (conjunto de dados de referência) ou então são genes possivelmente efetores (3 famílias). Vale enfatizar que os genes efetores estão associados a processos biológicos cruciais e que necessitam ser expressos em abundância, pois se tratam de moléculas secretadas por patógenos que muitas vezes atuam diretamente na planta, modificando a fisiologia de suas células e facilitando o processo de infecção (Lo Presti et al., 2015). A família de genes das NLP por exemplo, apresentou expressão logo nas primeiras horas após a infecção (Dal'Sasso et al., 2022), o que potencialmente caracteriza esta família como sendo de genes efetores.

Algumas evidências que podem reforçar essas hipóteses são mostradas pelos clados que exibiram preferência por AT. Os parálogos PME 1.2 e M35 1.2 que apresentaram valores %GC(3) e %GC(all) próximos a 45%, também foram aqueles que apresentaram pouca ou nenhuma expressão nos testes realizados por Rocha (2023)b. A preferência por GC apresentada pelos códons do conjunto de dados de referência (7962 genes SCO) pode sugerir quais são as bases e os códons mais abundantes no organismo dos isolados. Deste modo, a existência de genes com preferência por AT, como os que codificam para a PME 1.2 e M35 1.2 em um

organismo rico em GC pode ser uma das justificativas para a ausência ou atraso na expressão dos genes.

O parálogo NLP 1.1 que apresentou os padrões no uso de códons mais semelhante com os 7962 genes SCO, também foi a NLP que apresentou expressão logo nas primeiras horas após a infecção/inoculação, atingindo o nível mais alto de expressão antes dos outros três membros (Dal'Sasso et al., 2022). Vários autores já relataram que o nível de expressão gênica é o determinante mais crucial que influencia o comportamento de uso de códons, ditando o modo de uso de códons em vários organismos (Roy; Van Staden, 2019; Song et al., 2017; Duret, 2000).

O padrão de uso de códons mais conservado e semelhante ao conjunto de dados de referência apresentado pelo gene que codifica para a NLP 1.1 pode estar relacionado com os resultados de testes de neutralidade (Tajima's D, Fu e Li's D* e F*) conduzidos por Dal'Sasso e colaboradores (2022). Em estudo recente sobre as NLP, estes autores ponderaram que o clado NLP 1.1 estaria sobre a ação intensa de forças de seleção negativa que estaria resultando na eliminação de polimorfismos deletérios e conseqüentemente mantendo conservada a sequência dos genes (Dal'Sasso et al., 2022). Essa descoberta pode ser um indicativo de que estas forças seletivas também conservariam o padrão de uso de códons ao longo da evolução neste clado da família das NLP.

Do mesmo modo, a NLP 1.2A ter sido o clado que apresentou os valores de composição nucleotídica mais contrastantes e o maior viés no uso de códons, vai de encontro com o que outros autores já haviam estabelecido, que esta NLP estaria evoluindo sob restrição seletiva mais relaxada em comparação aos outros membros (Dal'Sasso et al., 2022). Outro fato interessante é que além de ser o membro que mais se distanciou dos demais membros em todos os índices calculados, a NLP 1.2A também foi que menos apresentou expressão (Dal'Sasso et al., 2022). Tal fato reforça as ideias anteriores, de que por esse clado possuir maior viés, significa que ela precisa de códons mais específicos, o que conseqüentemente poderia estar atrasando sua tradução e expressão.

5.2 As diferentes forças que moldam o padrão de uso de códons e suas possíveis relações com a filogenia

A desproporcionalidade no uso de GC/AT por si só representa a influência da pressão seletiva no viés no uso de códons (Wang et al., 2018). Entretanto, os gráficos de neutralidade e de viés PR2 são mais eficazes em mostrar a existência e quais são as forças que estão moldando o padrão de uso de códons em *Corynespora cassiicola*.

No geral, o padrão de utilização de códons foi impulsionado por forças mutacionais e principalmente pela seleção natural, visto que nenhum dos genes das três famílias se encontraram distribuídos exatamente no centro de seus respectivos gráficos PR2 (figuras 4, 8 e 12). Alguns autores já sugeriram anteriormente que o viés no uso de códons resulta de uma parceria mútua entre seleção e mutação, que é equilibrada por várias forças desconhecidas em diferentes níveis de organização (Hershberg; Petrov, 2008; Labella et al., 2019).

A rara ocorrência do parálogo M35 2.2, tendo ele sido encontrado em apenas cinco dos 61 isolados de *Corynespora cassiicola*, sugere que este membro da família das M35 possa ter surgido muito recentemente em um ancestral comum destes isolados. O alto grau de semelhança da sua composição nucleotíca e padrão de uso de codons com os parálogo M35 1.2 poderiam indicar que eles compartilham um ancestral comum mais recente; entretanto, os parálogos de M35 1.2 e M35 2.2 ocupam posição filogeneticamente distintas na árvore das M35 de *Corynespora cassiicola* (Rocha, 2023).

Do mesmo modo, os clados PME 2.2 e 2.3 que são filogeneticamente próximos, visto que compartilham de um evento de duplicação em comum e relativamente recente, apresentam padrões bem distintos na preferência do uso de códons, inclusive ficando bastante separadas no agrupamento (figura 11).

Tais fatos evidenciam que apesar de se tratar de famílias gênicas, na qual seus membros foram originados de duplicação, diferentes forças evolutivas podem atuar sobre os parálogos de um determinado gene, independente do grau de relacionamento filogenético com seus parálogos. Fazendo-os apresentarem diferentes padrões na utilização de códons. Outros estudos também já relataram a ação de diferentes forças moldando o padrão de uso de códons (Dean et al., 2012; Jiang et al., 2017).

5.3 As preferencias no Uso Relativo de Códons Sinônimos

Nossos resultados mostraram um compartilhamento de vários códons preferidos entre genes das famílias NLP, PME e M35 e o conjunto de dados que foi utilizado como referência (7962 genes SCO). Entretanto, o único códon 100% em comum foi o TTA, classificado como sub-representado em todos os genes de todas as famílias estudadas e 7962 SCO. O uso do códon TTA também foi menor do que o esperado em outros organismos com genomas ricos em GC (Li et al., 2007), sendo, inclusive, considerado um códon extremamente raro em alguns (Silov et al., 2021).

Apesar do compartilhamento de códons preferidos em comum relatado, vários cladogramas apresentaram particularidades, alguns até contrastando com membros de sua própria família, conforme demonstrado pela análise de componentes principais. As altas variações encontradas nos aminoácidos arginina e leucina pelas famílias PME e M35, bem como as preferências por códons ricos em GC exibidas pelas três famílias, também foram relatados em estudos anteriores com outras espécies de ascomicetos (Bansal et al., 2022).

As diferentes preferências por códons que codificam para prolina, apresentados pelos membros da família de genes das NLP, podem se tratar de estratégias de manutenção e uso, visto que esse aminoácido desempenha um papel crucial na interação hospedeiro-patógeno, protegendo o patógeno sob condições estressantes (Christgen; Becker, 2019).

Já foi comprovado também que essas diferenças de códons preferidos são estratégias de patógenos e hospedeiros que empregam aminoácidos biossinteticamente econômicos, para que a energia economizada possa ser canalizada para conferir mais virulência ou resistência ao sistema imunológico da planta (Bansal et al., 2022). Um exemplo disso é a Glicina, um aminoácido codificado majoritariamente por guanina, que também teve sua preferência de uso compartilhada entre diferentes genes desse estudo. Pesquisas anteriores mostram que a glicina pode ter um custo de sintetização até 6 vezes mais baixo quando comparado a outros aminoácidos, em relação ao gasto de ATP durante sua síntese (Smith; Chapman, 2010).

O fato é que o compartilhamento de códons preferidos entre genes das famílias NLP, PME e M35 e o conjunto de dados que foi utilizado como referência (7962 genes

SCO) é uma evidência direta da conservação desses códons durante o curso da evolução e da importância desses códons para a expressão de seus respectivos genes.

6. CONCLUSÃO

A grande maioria dos genes deste estudo exibiram preferência por códons ricos e GC. Essa preferência por GC ocorre principalmente em detrimento de A e pode ser justificada por estratégias de economia e eficiência durante o processo de tradução.

Mesmo genes diferentes compartilhando códons preferenciais em comum, ficou evidenciado que existe mais de uma força moldando o padrão de uso de códons. No geral a maior influência é de pressão de mutação e principalmente seleção natural. Foi possível notar também que, membros de uma mesma família genica, podem ser afetados por diferentes forças evolutivas. Tal fato tornou difícil a tarefa de encontrar relações entre os padrões de uso de códons em membros mais próximos filogeneticamente.

Clados que apresentaram os padrões de uso de códons mais diferente do conjunto de dados de referência são os que apresentaram menores taxa expressão e vice-versa. Esta constatação pode indicar uma relação entre os códons preferenciais de um organismo e a taxa de tradução e conseqüentemente expressão apresentada pelo mesmo.

7. REFERÊNCIAS BIBLIOGRÁFICAS

ARNADOTTIR, H.; HVANNDAL, I.; ANDRESDOTTIR, V.; BURR, S. E.; FREY, J.; GUDMUNDSDOTTIR, B. K. The AsaP1 peptidase of *Aeromonas salmonicida* subsp. *achromogenes* is a highly conserved deuterolysin metalloprotease (family M35) and a major virulence factor. **Journal of bacteriology**. v. 191, n. 1, p. 403-410, 2009.

BADET, T.; PEYRAUD, R.; MBENGUE, M.; NAVAUD, O.; DERBYSHIRE, M.; OLIVER, R. P.; BARBACCI, A.; RAFFAELE, S. Codon optimization underpins generalist parasitism in fungi. **Elife**. v. 6, p. e22472, 2017.

BANSAL, S.; MALLIKARJUNA, M. G.; BALAMURUGAN, A.; NAYAKA, S. C.; PRAKASH, G. Composition and Codon Usage Pattern Results in Divergence of the Zinc Binuclear Cluster (Zn(II)₂Cys₆) Sequences among Ascomycetes Plant Pathogenic Fungi. **Journal of Fungi**. v. 8, n. 11, p. 1134, 2022.

BENNETZEN, J. L.; HALL, B. D. Codon selection in yeast. **Journal of Biological Chemistry**. v. 257, n. 6, p. 3026-3031, 1982.

BOURRET, J.; ALIZON, S.; BRAVO, I. G. COUSIN (COdon Usage Similarity INdex): a normalized measure of codon usage preferences. **Genome Biology and Evolution**. v. 11, n. 12, p. 3523-3528, 2019.

CASTANHO, F. M. **Caracterização estrutural e polimorfismo de genes candidatos à efetores de *phakopsora pachyrhizi***. 2017. Dissertação (Mestrado em Agronomia) - Universidade Estadual Do Norte Do Paraná, Bandeirantes, 2017.

CHRISTGEN, S. L.; BECKER, D. F. Role of proline in pathogen and host interactions. **Antioxidants & redox signaling**. v. 30, n. 4, p. 683-709, 2019.

DAL'SASSO, T. C. S. ROCHA, V. D.; RODY, H. V. S.; COSTA, M. D-B. L.; OLIVEIRA, L. O. The necrosis-and ethylene-inducing peptide 1-like protein (NPL) gene Family of the plant pathogen *Corynespora cassiicola*. **Current Genetics**. v. 68, n. 5-6, p. 645-659, 2022.

DEAN, R.; VAN KAN, J. A. L.; PRETORIUS, Z. A.; HAMMOND-KOSACK, K. E.; DI PIETRO, A.; SPANU, P. D.; RUDD, J. J.; DICKMAN, M.; KAHMANN, R.; ELLIS, J.; FOSTER, G. D. The Top 10 fungal pathogens in molecular plant pathology. **Molecular plant pathology**. v. 13, n. 4, p. 414-430, 2012.

DEB, B.; UDDIN, A.; CHAKRABORTY, S. Codon usage pattern and its influencing factors in different genomes of hepadnaviruses. **Archives of Virology**. v. 165, p. 557-570, 2020.

DE CARVALHO, M. C. C. G.; NASCIMENTO, L. C.; DARBEN, L. M.; POLIZEL-PADANOSQUI, A. M.; LOPES-CAITAR, V. S.; QI, M.; ROCHA, C. S.; CARAZZOLLE, M. F.; KUWAHARA, M. K.; PEREIRA, G. A. G.; ABDELNOOR, R. V.; WHITHAM, S. A.; MARCELINO-GUIMARÃES, F.C. Prediction of the in planta *Phakopsora pachyrhizi* secretome and potential effector families. **Molecular plant pathology**. v. 18, n. 3, p. 363-377, 2017.

DENG, Y.; HEDAYIOGLU, F. L.; KALFON, J.; CHU, D.; VON DER HAAR, T. Hidden patterns of codon usage bias across kingdoms. **Journal of The Royal Society Interface**. v. 17, n. 163, p. 20190819, 2020.

DÉON, M.; BOURRÉ, Y.; GIMENEZ, S.; BERGER, A.; BIEYSSE, D.; LAMOTTE, F.; PONCET, J.; ROUSSEL, V.; BONNOT, F.; OLIVER, G.; FRANCHÉL, J.; SEGUIN, M.; LEROY, T.; ROECKEL-DREVET, P.; PUJADE-RENAUD, V. Characterization of a cassiicolin-encoding gene from *Corynespora cassiicola*, pathogen of rubber tree (*Hevea brasiliensis*). **Plant Science**. v. 185/186 p. 227-237, 2012.

DICE, L. R. Measures of the amount of ecological association between species. **Ecology**. v. 26, n. 3, p. 297-302. 1945.

DUPLESSIS, S.; BAKKEREN, G.; HAMELIN, R. Chapter Six -Advancing knowledge on biology of rust fungi through genomics. **Advances in botanical research**. v.70, p. 173-209, 2014.

DURET, L.; MOUCHIROUD, D. Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. **Proceedings of the National Academy of Sciences**. v. 96, n. 8, p. 4482-4487, 1999.

DURET, L. tRNA gene number and codon usage in the *C. elegans* genome are co-adapted for optimal translation of highly expressed genes. **Trends in Genetics**. v. 16, n. 7, p. 287-289, 2000.

FARR, D. F.; ROSSMAN, A. Y.; CASTLEBURY, L. A. (2021). United States National Fungus Collections Fungus-Host Dataset. Ag Data Commons. **USDA**. Disponível em: <https://doi.org/10.15482/USDA.ADC/1524414>. Acesso em: 4 jan. 2024.

FREIRE-PICOS, M. A.; GONZALEZ-SISO, M. I.; RODRÍGUEZ-BELMONTE, E.; RODRÍGUEZ-TORRES, A. M.; RAMIL, E.; CERDAN, M. E. Codon usage in *kluveromyces lactis* and in yeast cytochrome c-encoding genes. **Gene**. v. 139, n. 1, p. 43-49, 1994.

GIJZEN, M.; NÜRNBERGER, T. Nep1-like proteins from plant pathogens: recruitment and diversification of the NPP1 domain across taxa. **Phytochemistry**. v. 67, n. 16, p. 1800-1807, 2006.

GUSTAFSSON, C.; MINSHULL, J.; GOVINDARAJAN, S.; NESS, J.; VILLALOBOS, A.; WELCH, M. Engineering genes for predictable protein expression. **Protein Expression and Purification**. v. 83, n. 1, p. 37-46, 2012.

GUPTA, S.; SINGH, R. Comparative study of codon usage profiles of *Zingiber officinale* and its associated fungal pathogens. **Molecular Genetics and Genomics**. v. 296, n. 6, p. 1121-1135, 2021.

HAMMER, Ø.; HARPER, D. A. T.; RYAN, P. D. Past: paleontological statistics software package for education and data analysis. **Palaeontologia electronica**. v. 4, n. 1, p. 1, 2001.

HERSHBERG, R.; PETROV, D. A. Selection on codon bias. **Annual review of genetics**. v. 42, p. 287-299, 2008.

IRIEDA, H.; INOUE, Y.; MORI, M.; YAMADA, K.; OSHIKAWA, Y.; SAITOH, H.; UEMURA, A.; TERAUCHI, R.; KITAKURA, S.; KOSAKA, A.; SINGKARAVANIT-OGAWA, S.; TAKANO, Y. Conserved fungal effector suppresses PAMP-triggered immunity by targeting plant immune kinases. **Proceedings of the National Academy of Sciences**. v. 116, n. 2, p. 496-505, 2019.

IKEMURA, T. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. **Journal of molecular biology**. v. 146, n. 1, p. 389-409, 1981.

JIANG, W.; LV, B.; WU, X.; WANG, J.; WU, G.; SHI, C.; TANG, X. Analysis of synonymous codon usage patterns in the edible fungus *Volvariella volvacea*. **Biotechnology and Applied Biochemistry**. v. 64, n. 2, p. 218-224, 2017.

KAMOUN, S. The secretome of plant-associated fungi and oomycetes. **The mycota**. In: Deising HB **plant relationships**, 2. ed. v. 5, p. 173-180, 2009.

KIMURA, M. Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. **Nature**. v. 267, p. 275-276, 1977.

KUDLA, G.; MURRAY, A. W.; TOLLERVEY, D.; PLOTKIN, J. B. Coding-sequence determinants of gene expression in *Escherichia coli*. **Science**. v. 324, n. 5924, p. 255-258, 2009.

LABELLA, A. L.; OPULENTE, D. A.; STEENWYK, J. L.; HITTINGER, C. T.; ROKAS, A. Variation and selection on codon usage bias across an entire subphylum. **PLoS genetics**. v. 15, n. 7, p. e1008304, 2019.

LEHNINGER, A. L.; NELSON, D. L.; COX, M. M. **Princípios de bioquímica**. 2. ed. São Paulo: Sarvier, 1995.

LI, J.; ZHANG, K-Q. Independent expansion of zinc metalloproteinases in Onygenales fungi may be associated with their pathogenicity. **PLoS One**. v. 9, n. 2, p. e90225, 2014.

LI, W.; WU, J.; TAO, W.; ZHAO, C.; WANG, Y.; HE, X.; CHANDRA, G.; ZHOE, X.; DENG, Z.; CHATER, K. F.; TAO, M. A genetic and bioinformatic analysis of *Streptomyces coelicolor* genes containing TTA codons, possible targets for regulation by a developmentally significant tRNA. **FEMS Microbiology Letters**. v. 266, n. 1, p. 20-28, 2007.

LO PRESTI, L.; LANVER, D.; SCHWEIZER, G.; TANAKA, S.; LIANG, L.; TOLLLOT, M.; ZUCCARO, A.; REISSMANN, S.; KAHMANN, R. Fungal effectors and plant susceptibility. **Annual review of plant biology**. v. 66, p. 513-545, 2015.

LOPEZ, D.; RIBEIRO, S.; LABEL, P.; FUMANAL, B.; VENISSE, J-S.; KOHLER, A.; OLIVEIRA, R. R.; LABUTTI, K.; LIPZEN, A.; LAIL, K.; BAUER, D.; OHM, R. A.; BARRY, K. W.; SPATAFORA, J.; GRIGORIEV, I. V.; MARTIN, F. M.; PUJADE-RENAUD, V. Genome-wide analysis of *Corynespora cassiicola* leaf fall disease putative effectors. **Frontiers in microbiology**. v. 9, p. 276, 2018.

LOWE, R. G. T; HOWLETT, B. J. Indifferent, affectionate, or deceitful: lifestyles and secretomes of fungi. **PLoS pathogens**. v. 8, n. 3, p. e1002515, 2012.

MUTHABATHULA, P.; SUNEETHA, S.; GRACE, R. Genome-wide codon usage bias analysis in *Beauveria bassiana*. **Bioinformatics**. v. 14, n. 9, p. 580-586, 2018.

OHNO, S. **Evolution by gene duplication**. 1^a ed. New York: Springer-Verlag, 1970.

ONO, E.; MISE, K.; TAKANO, Y. RLP23 is required for *Arabidopsis* immunity against the grey mould pathogen *Botrytis cinerea*. **Scientific Reports**. v. 10, n. 1, p. 13798, 2020.

PAN, L.; WEN, S.; YU, J.; LU, L.; ZHU, X.; ZHANG, Z. Genome-wide identification of M35 family Metalloproteases in *Rhizoctonia cerealis* and functional analysis of RcMEP2 as a virulence factor during the fungal infection to Wheat. **International Journal of Molecular Sciences**. v. 21, n. 8, p. 2984, 2020.

PELLOUX, J.; RUSTÉRUCCHI, C.; MELLEROWICZ, E. J. New insights into pectin methylesterase structure and function. **Trends in plant Science**. v. 12, n. 6, p. 267-277, 2007.

PUIGBÒ, P.; BRAVO, I. G.; GARCIA-VALLVE, S. CAIcal: a combined set of tools to assess codon usage adaptation. **Biology direct**, v. 3, n. 38, p. 1-8, 2008.

QI, Y-X.; ZHANG, X.; PU, J-J.; LIU, X-M.; LU, Y.; ZHANG, H.; ZHANG, H-Q.; LV, Y-C.; XIE, Y-X. Morphological and molecular analysis of genetic variability within isolates of *Corynespora cassiicola* from different hosts. **European Journal of Plant Pathology**. v. 130, p. 83-95, 2011.

RAWLINGS, N. D.; BARRETT, A. J.; THOMAS, P. D.; HUANG, X.; BATEMAN, A.; FINN, R. D. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. **Nucleic acids research**. v. 46, n. D1, p. D624-D632, 2018.

REIS, B. P.; LANNA FILHO, R.; ALFENAS, R. F.; ALFENAS, A. C. First report of *Corynespora cassiicola* causing severe leaf blight on *Eucalyptus* in Brazil. **New Disease Reports**. v. 29, n. 7, p. 2044-0588, 2014.

ROCHA, E. P. Codon usage bias from tRNA's point of view: Redundancy, specialization, and efficient decoding for translation optimization. **Genome Res**. v.14, p. 2279-2286, 2004.

ROCHA, V. D.; DAL'SASSO, T. C. S.; DAL-BIANCO, M.; OLIVEIRA, L. O. Genome-wide survey and evolutionary history of the pectin methylesterase (PME) gene family in the Dothideomycetes class of fungi. **Fungal Genetics and Biology**. v.169, p.103841, 2023.

ROCHA, V. D. **Molecular evolution of two pathogenicity-associated gene families across the dothideomycetes class of fungi**. 2023. Tese (Doutorado em Genética e Melhoramento) - Universidade Federal de Viçosa, Minas Gerais, 2023.

ROY, A.; VAN STADEN, J. Insights into the riddles of codon usage patterns and codon context signatures in fungal genus *Puccinia*, a persistent threat to global agriculture. **Journal of Cellular Biochemistry**. v. 120, n. 12, p. 19555-19566, 2019.

SELLA, L.; CASTIGLIONI, C.; PACCANARO, M. C.; JANNI, M.; SCHÄFER, W.; D'OVIDIO, R.; FARARON, F. Involvement of fungal pectin methylesterase activity in the interaction between *Fusarium graminearum* and wheat. **Molecular Plant-Microbe Interactions**. v. 29, n. 4, p. 258-267, 2016.

SILOV, S.; ZABURANNYI, N.; ANISIMOVA, M.; OSTASH, B. The Use of the Rare TTA Codon in *Streptomyces* Genes: Significance of the Codon Context? **Indian Journal of Microbiology**. v. 61, n. 1, p. 24-30, 2021.

SONG, H.; LIU, J.; SONG, Q.; ZHANG, Q.; TIAN, P.; NAN, Z. Comprehensive Analysis of Codon Usage Bias in Seven *Epichloë* Species and Their Peramine-Coding Genes. **Front. Microbiol.** v. 8, p.1419, 2017.

SUEOKA, N. Directional mutation pressure and neutral molecular evolution. **Proceedings of the National Academy of Sciences**. v. 85, n. 8, p. 2653-2657, 1988.

SUEOKA, N. Intrastrand parity rules of DNA base composition and usage biases of synonymous codons. **Journal of molecular evolution**. v. 40, p. 318-325, 1995.

SHARP, P. M.; LI, W. H. An evolutionary perspective on synonymous codon usage in unicellular organisms. **Journal of molecular evolution**, v. 24, p. 28-38, 1986.

SHARP, P. M.; LI, W. H. The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. **Nucleic acids research**. v. 15, n. 3, p. 1281-1295, 1987.

SMITH, D. R.; CHAPMAN, M.R. Economical evolution: microbes reduce the synthetic cost of extracellular proteins. **MBio**. v. 1, n. 3, p. 10.1128, 2010.

STENICO, M.; LLOYD, A. T.; SHARP, P. M. Codon usage in *Caenorhabditis elegans*: delineation of translational selection and mutational biases. **Nucleic acids research**, v. 22, n. 13, p. 2437-2446, 1994.

VOGLMAYR, H.; JAKLITSCH, W. M. *Corynespora*, *Exosporium* and *Helminthosporium* revisited-new species and generic reclassification. **Studies in Mycology**. v. 87, n. 1, p. 43-76, 2017.

WAN, X.; XU, D.; ZHOU, J. A new informatics method for measuring synonymous codon usage bias. **Intelligent engineering systems through artificial neural networks**. v. 13, p. 1-8, 2003.

WANG, L.; XING, H.; YUAN, Y. C; WANG, X. L; SAEED, M.; TAO, J. C; FENG, W.; ZHANG, G.; SONG, X.; ZHEN, S. Genome-wide analysis of codon usage bias in four sequenced cotton species. **PloS one**. v. 13, n. 3, p. e0194372, 2018.

WATSON, J. D.; BAKER, T. A.; BELL, S. P.; GANN, A.; LEVINE, M.; LOSICK, R. **Biologia Molecular do Gene**. 5. ed. Porto Alegre: Artmed, 2006.

WU, Y.; ZHAO, D.; TAO, J. Analysis of codon usage patterns in herbaceous peony (*Paeonia lactiflora* Pall.) based on transcriptome data. **Genes**. v. 6, n. 4, p. 1125-1139, 2015.

WRIGHT, F. The 'effective number of codons' used in a gene. **Gene**, v. 87, n. 1, p. 23-29, 1990.

ZHANG, X.; HUANG, H.; WU, B.; XIE, J.; VILJOEN, A.; WANG, W.; MOSTERT, D.; XIE, Y.; FU, G.; XIANG, D.; LYU, S.; LIU, S.; LI, C. The M35 metalloprotease effector FocM35_1 is required for full virulence of *Fusarium oxysporum* f. sp. *ubense* tropical race 4. **Pathogens**. v. 10, n. 6, p. 670, 2021.