

Gustavo Bortolanza de Andrade

**Análise de plasticidade das vias de senescência celular,
apoptose e estabilidade genômica**

Orientador: Prof Dr José Luiz Rybarczyk Filho

Botucatu, SP

2022

Gustavo Bortolanza de Andrade

Análise de plasticidade das vias de senescência celular, apoptose e estabilidade genômica

Trabalho de Conclusão de Curso
apresentado ao Instituto de Biociências
da Universidade Estadual Paulista
"Júlio de Mesquita Filho", Câmpus de
Botucatu, para obtenção do título de
Bacharel em Ciências Biológicas.

Orientador: Prof Dr José Luiz Rybarczyk Filho

Departamento de Biofísica e Farmacologia - Laboratório de
Biocomplexidade, Instituto de Biociências de Botucatu - UNESP

Botucatu, SP

2022

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP

BIBLIOTECÁRIA RESPONSÁVEL: ROSEMEIRE APARECIDA VICENTE-CRB 8/5651

Andrade, Gustavo Bortolanza de.

Análise de plasticidade das vias de senescência celular,
apoptose e estabilidade genômica / Gustavo Bortolanza de
Andrade. - Botucatu, 2022

Trabalho de conclusão de curso (bacharelado - Ciências
Biológicas) - Universidade Estadual Paulista "Júlio de
Mesquita Filho", Instituto de Biociências de Botucatu

Orientador: José Luiz Rybarczyk Filho

Capes: 20206003

1. Plasticidade celular. 2. Genômica. 3. Apoptose.
4. Células - Envelhecimento.

Palavras-chave: Estabilidade genômica; Manutenção genômica;
Plasticidade evolutiva; Supressão tumoral.

AGRADECIMENTOS

Agradeço primeiramente a minha falecida mãe, Dulcéia Bortolanza de Andrade, a quem dedico este trabalho.

À meu pai, Mauro Roberto de Andrade, bem como a minha madrastra Cornelia Wendel, minhas tias Dilma e Dirce Bortolanza, por me darem todo o suporte que precisei durante minha formação assim como a motivação para seguir adiante.

À minha namorada, Karolina Correia Penetra, que mais esteve presente tanto para me levantar em minhas derrotas quanto para comemorar minhas conquistas em toda essa etapa da minha jornada.

Aos meus irmãos Mateus e Pedro Rochavetz de Lara Andrade. Por me ajudarem com sua sabedoria em meus momentos de decisão.

Aos meus estimados amigos da "Geo Aliança" Arthur Ramalho, Bruno Alves, Giovanna Bergamasco, Marcio Rossatto, Maria Palone e Matheus Lopes por toda a ajuda prática que me deram durante esse período e pelos bons momentos que tivemos durante a faculdade. Especialmente a Arthur Ramalho por todo auxílio que me deu em programação e a Maria Palone por me ajudar a estruturar minhas ideias.

E ao meu orientador José Luiz Rybarczyk Filho, e demais colegas de laboratório, pela oportunidade de trabalhar ao lado de pesquisadores tão competentes.

Se alcanço minha graduação, é porque vocês tornaram isto possível.

RESUMO

Os mecanismos de manutenção genômica são de suma importância para organismos eucariontes pois atuam frente a inconformidades no DNA, reparando seus danos e corrigindo possíveis erros ou até inativando ou destruindo a célula em casos extremos. Desta forma, mecanismos de manutenção genômica promovem longevidade, pois impedem que erros no DNA se acumulem e se propaguem levando a célula a se tornar um tumor. A abordagem evolutiva para estudo destas vias gênicas que agem na estabilidade genômica e na supressão tumoral ajuda a compreender suas origens e suas relações à medida que estas foram se desenvolvendo ao longo da evolução. Com o objetivo de analisar a plasticidade das vias de genes envolvidos nos mecanismos de manutenção genômica, foi utilizado o Geneplast do Bioconductor, uma ferramenta de análise, que se baseia na distribuição dos ortólogos ao longo de determinada árvore evolutiva, e infere seus valores de abundância, diversidade e plasticidade. Os dados foram fornecidos pelos bancos de dados String, e o KEGG (Kyoto Encyclopedia of Genes and Genomes). Verificou-se então os valores de abundância, diversidade e plasticidade dos genes das vias de apoptose, senescência celular e estabilidade genômica, o que revelou que o grupo de estabilidade genômica possui a maior diversidade entre os três, sendo a via com maior presença entre os organismos da árvore. Ao mesmo tempo, esse grupo possui o menor valor de abundância, o que indica que há poucos parálogos desses genes nos organismos. Por conseguinte, seus índices de plasticidade evolutiva possuem os menores valores, o que indica um grupo de genes com baixa aceitação às mutações e alterações na forma como este mecanismo funciona. Este conjunto de ortólogos é essencial tanto em organismos unicelulares, como em organismos mais complexos, o que pode explicar sua aparição antiga na árvore de eucariotos e sua presença constante e quase inalterada ao longo de toda a história evolutiva deste grupo. Por outro ângulo em relação aos ortólogos das vias de apoptose e senescência celular, notou-se que eles são estatisticamente idênticos, tanto em relação aos valores de diversidade, quanto do índice de plasticidade evolutiva, o que pode indicar que eles tiveram aparições em um mesmo ancestral comum, recentemente em comparação a via de estabilidade genômica. Porém, seus valores de abundância são distintos e maiores nos ortólogos que atuam na via de apoptose, o que indicaria que este grupo tem uma aceitação maior a alterações. Embora esta análise seja promissora, ainda há muito a se descobrir em relação à história evolutiva dos genes que atuam mantendo a estabilidade genômica, e mais análises são necessárias para inferir a origem exata destes genes assim como a influência que exercem uns sobre os outros

SUMÁRIO

INTRODUÇÃO.....	7
OBJETIVOS.....	17
MATERIAIS E MÉTODOS.....	17
RESULTADOS.....	22
DISCUSSÃO.....	24
CONCLUSÃO.....	26
REFERÊNCIAS BIBLIOGRÁFICAS.....	27

1. INTRODUÇÃO

1.1 Câncer

O Câncer é uma das principais doenças que afligem a humanidade nos dias atuais. Segundo a Sociedade Brasileira de Cancerologia, anualmente morrem 8,2 milhões de pessoas vítimas de câncer e, apesar do crescente investimento financeiro no combate à doença, a tendência é que o número de novos casos continue subindo, segundo a Organização Mundial da Saúde (OMS).

Um tumor caracteriza-se por um amontoado de células que surge através da replicação descontrolada de uma única célula que acumulou mutações em seu DNA ao longo de muitos anos, corrompendo sua função original criando um comportamento celular inadequado. Essas mutações podem ocorrer tanto por erros internos nos momentos de mitose, como por danos externos, como radiação (GRIFFITHS *et al.*, 2013).

As mutações podem ocorrer em qualquer célula do corpo e, portanto, existem centenas de tipos diferentes de neoplasias com características e tratamentos distintos. Algumas são consideradas benignas pois, apesar de poderem crescer em tamanho, tendo potencial para causar quadros graves, não são capazes de se espalhar pelo corpo e usualmente podem ser removidas cirurgicamente. Diferentemente, os tumores malignos, ou câncer, tem a capacidade de se espalhar para os tecidos adjacentes ou então adentrar os vasos condutores para se alojarem em órgãos distantes, em um processo denominado metástase. Dessa forma, esta enfermidade pode resistir a processos cirúrgicos e seu tratamento se torna muito mais complicado, sendo uma doença bem mais perigosa. (COOPER, 2000)

1.2 Estabilidade Genômica

Os organismos vivos possuem mecanismos naturais de manutenção do genoma para impedir o surgimento de neoplasias. Estas funções celulares agem para impedir que as mutações se acumulem e levem a célula a se tornar um tumor, reparando danos externos, evitando o acúmulo de erros no DNA celular e promovendo assim o bom funcionamento da célula, bem como o fitness celular e a manutenção da sua homeostase. Esses mecanismos incluem, mas não se limitam, às vias de recombinação homóloga, união de extremidades não homóloga, reparo por excisão de bases, reparo por excisão

de nucleotídeos, reparo de incompatibilidade de DNA além de outras mais. O funcionamento destas vias é de extrema importância uma vez que defeitos nesses mecanismos levam à instabilidade genômica, caracterizada por diversas desordens e doenças humanas, incluindo o câncer e demais neoplasias (RODIER *et al.*, 2007).

1.2.1 Reparo por Excisão de Bases:

O reparo por excisão de bases (*Base Excision Repair* - BER) é o principal mecanismo de correção de bases danificadas em pequenas seções de DNA. Esses danos podem ocorrer devido à lesões nas bases nitrogenadas causadas por estresse oxidativo, alquilação ou desaminação. Quando não corrigidos, eles podem levar a um emparelhamento de bases errôneo e se consolidar como uma mutação no DNA, gerando instabilidade genômica e contribuindo para o surgimento de doenças neoplásicas.

O início da via de reparo por excisão de base se dá quando uma enzima DNA glicosilase reconhece a base danificada na fita de DNA e catalisa a sua remoção, deixando apenas um sítio desprovido de base nitrogenada ou sítio AP; na sequência uma enzima AP Endonuclease ou DNA AP liase recorta a cadeia principal do DNA, deixando o espaço de um único nucleotídeo aberto. Em seguida, uma enzima DNA polimerase adiciona uma nova base sem danos ao local do sítio AP, usando a base da fita complementar como molde. Para finalizar o processo, uma enzima DNA ligase conecta o novo nucleotídeo à fita de DNA, como representado na figura 1.

Adicionalmente, existe também uma forma de reparo por excisão de bases que renova múltiplas bases em um único processo, denominado “Long-patch”, em que, após a clivagem da fita de DNA pela AP Endonuclease, DNAs polimerase beta (POLB) e delta (POLD) associadas a PCNA promovem o afastamento de um trecho da fita de DNA e substituem pelas novas bases. Logo mais a enzima Flap Endonuclease (FEN) remove a fita de DNA que foi afastada, separando todo um trecho de nucleotídeos da cadeia principal de DNA, e finalmente uma DNA ligase liga a nova seção do DNA a cadeia principal (ROBERTSON *et al.*, 2009)

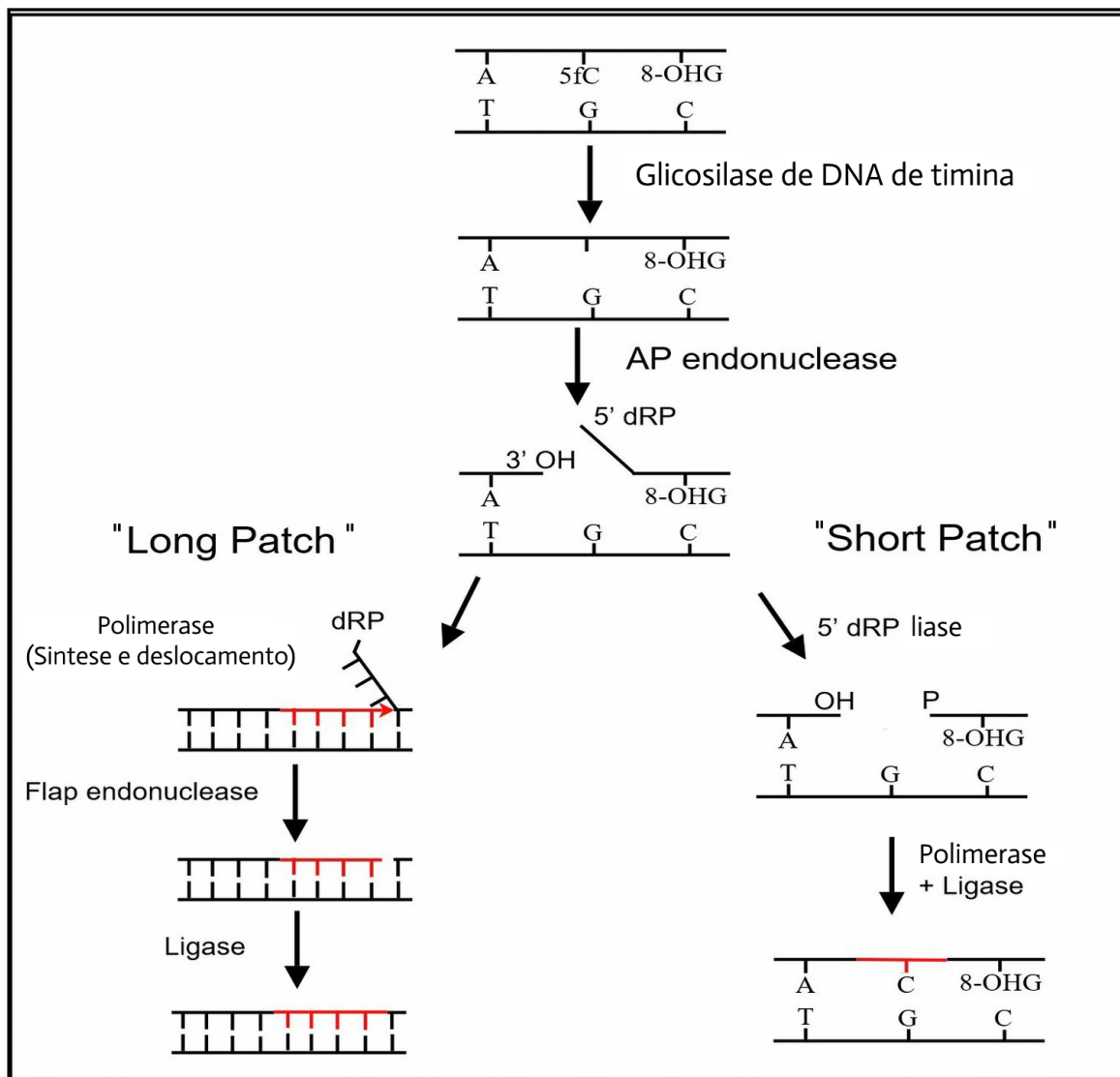


Figura 1: Esquema de reparo por excisão de bases, representações das etapas de reconhecimento da base lesionada, sua remoção e posterior abertura da cadeia principal, finalizando com a retirada de uma seção de DNA e sua substituição por novas bases (Adaptado de https://commons.wikimedia.org/wiki/File:BER_of_5fC_no_caption.jpg).

1.2.2 Reparo por Excisão de Nucleotídeos

Esse processo repara principalmente lesões ambientais do DNA causadas por agentes externos, como luz Ultravioleta ou componentes carcinogênicos, e se divide em duas sub-vias distintas, "transcriptional coupled repair" (TCR) e "global genome repair" (GGR). Esses nomes se referem à etapa inicial do processo de excisão de nucleotídeos, e ambos os processos são demonstrados na figura 2.

Em ambas sub-vias, a primeira etapa é o reconhecimento da anormalidade no DNA, a diferença sendo que o processo TCR ocorre quando RNA polimerase encontra danos na fita molde de DNA que ela iria transcrever em RNA mensageiro. Ela então age como um sinalizador para o complexo proteico TFIIH dar sequência ao processo de reparo; Enquanto que o GGR ocorre nas demais partes do DNA e é reconhecido pelas proteínas do grupo de complementação do xeroderma pigmentoso (XP) que, ao detectar danos, também sinaliza para os mesmos complexos proteicos.

A próxima etapa consiste no desenrolamento da dupla hélice do DNA mediado pelo complexo TFIIH juntamente com XPB e XPD, além de XPA em associação com RPA. Este grande complexo proteico abre o DNA, separando as ligações entre as bases nitrogenadas formando uma 'bolha' de aproximadamente 30 pares de base em torno da região lesionada. Posterior a isso, duas endonucleases recortam essa tira de DNA, e o espaço que foi retirado é preenchido por DNA polimerases e é conectado à cadeia principal por DNA ligases, corrigindo assim a área afetada (SALDIVAR et al., 2007).

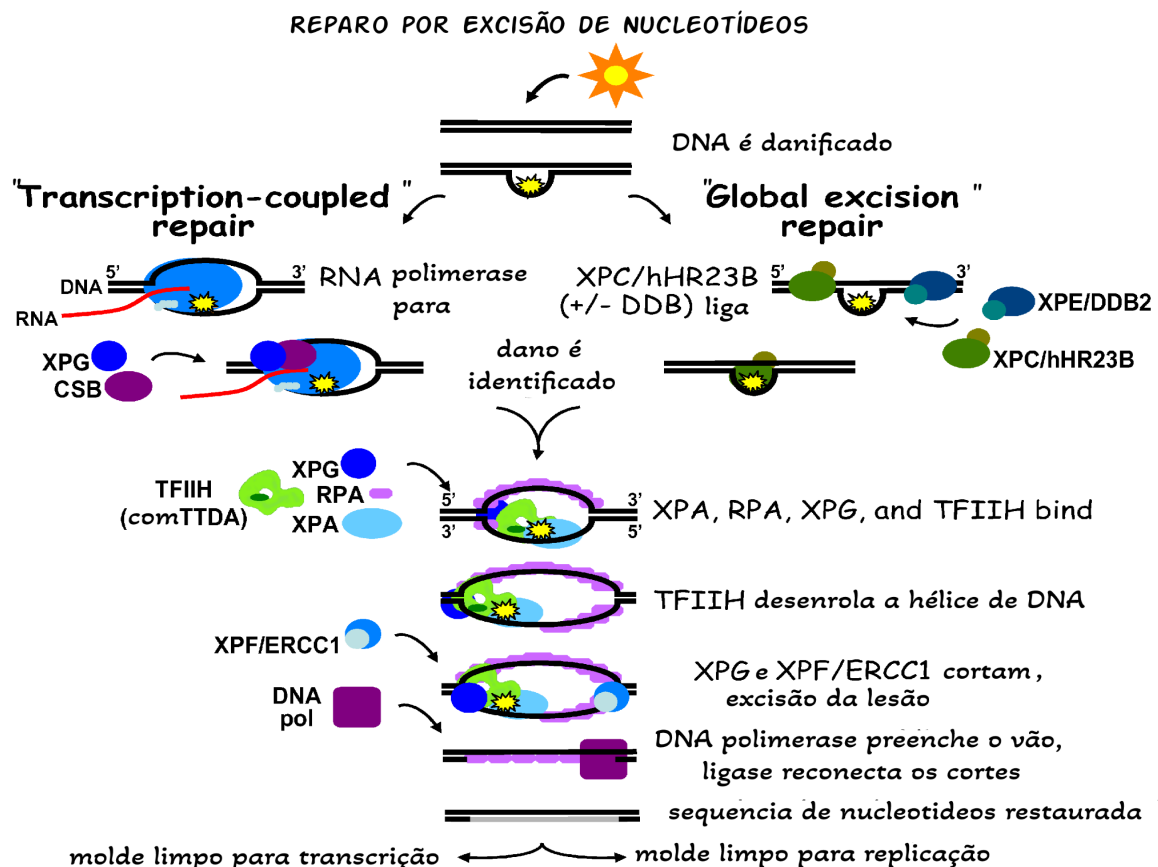


Figura 2: Esquema de reparo por excisão de nucleotídeos, demonstrando as etapas de reconhecimento em abas subvis (TCR a esquerda e GGR a direita), seguido das etapas de desenrolamento do DNA, formação da bolha, recorte da área afetada e repleenchimento.

(Adaptado de: https://commons.wikimedia.org/wiki/File:Nucleotide_Excision_Repair-journal.pbio.0040203.g001.png)

1.2.3 Reparo de Incompatibilidade de DNA

O reparo de incompatibilidades age para reparar eventuais erros que ocorrem nos processos celulares de replicação, recombinação ou até mesmo outros reparos do DNA, onde pode ser que uma base incompatível seja pareada erroneamente, pois por mais precisos que sejam esses processos eles não são infalíveis e ocorrem demasiadas vezes no tempo de vida de um organismo. A falha nessa via está associada a mutações, instabilidade genômica e tumores (LI, 2007)

REPARO DE INCOMPATIBILIDADE DE DNA

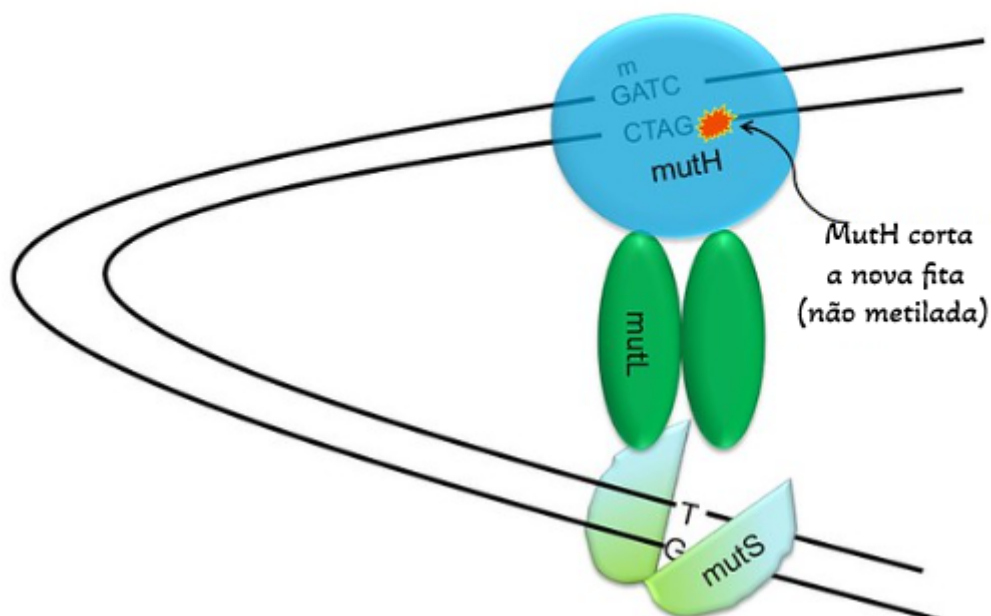


Figura 3: Representação da junção de MutS, MutL e MutH no processo de reparo de incompatibilidade.

(Adaptado de: <https://www.flickr.com/photos/agathman/4977090731>)

A proteína de reconhecimento de incompatibilidade (MutS) corre pela fita de DNA, monitorando se há algum pareamento incorreto. Ao encontrar ela sinaliza para a proteína MutL, que liga MutS a uma terceira proteína MutH. A ação conjunta dessas três proteínas irá localizar uma sequência de bases específicas no DNA (GATC), como exemplificado na figura 3, e iniciar um processo de excisão a partir dali, em que a helicase irá desenrolar o DNA em direção à incompatibilidade e exonucleases irão remover essa porção da fita de DNA até alcançar o erro, deixando um espaço vazio que por sua vez será preenchido corretamente por DNA polimerases e reconectado a fita original por uma DNA ligase, corrigindo assim o pareamento de bases incompatíveis (MARTI et al., 2002).

1.2.4 União de extremidade não-homóloga

A via de união de extremidades não homólogas age frente a danos na dupla fita de DNA que efetivamente separam um segmento do outro. Isso pode ocorrer devido à ação de enzimas ou a diferentes tipos de radiação e podem deixar rompimentos tanto simétricos quanto assimétricos na dupla fita bem como danificar as pontas das fitas de DNA, o que dificulta o processo de reconectá-las umas às outras. Este tipo de dano é um dos mais agressivos que podem ser infligidos em uma molécula de DNA, pois a possibilidade de perda de informações é muito grande o que pode comprometer gravemente a estabilidade genômica dependendo da área afetada.

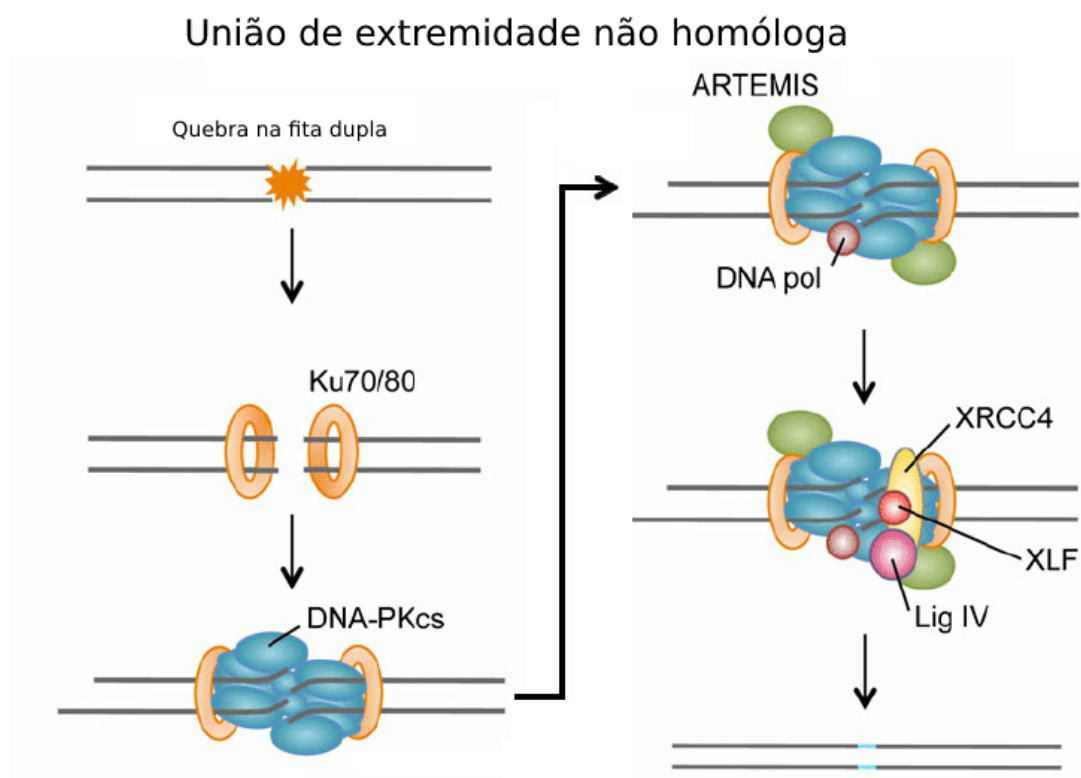


Figura 4: Processo de união de extremidade não homóloga. representando a união das proteínas ku70 e ku80 ao fragmento de DNA, e a ação das DNA Proteína Kinases em aproximar as duas extremidades, além de endonucleases corrigindo o comprimento dos fragmentos e por último a ação de ligases restabelecendo as conexões (Adaptado de https://www.researchgate.net/figure/Double-strand-breaks-repair-pathways-During-homologous-recombination-the-repair-of-DSB_fig2_257135774).

Desta maneira, unir as duas pontas de DNA danificados se torna extremamente importante. Quando possível, proteínas Ku70 e Ku80 se ligam às extremidades soltas do DNA seccionado e recrutam DNA proteína kinases

que aproximam as duas extremidades. Por último, DNA ligases conectam as duas pontas de DNA e, caso haja uma diferença pronunciada no comprimento das duas fitas, endonucleases (como artemis) costumam trimar as bases de uma das fitas até que o tamanho esteja adequado para o procedimento. Este processo não reconstrói nenhuma informação perdida de DNA e, portanto, não é um reparo perfeito, mas ainda é preferível a deixar desconectado os fragmentos da dupla fita (WYMAN et al., 2006).

1.2.5 Recombinação homóloga

A via de reparo por recombinação homóloga, assim como na via de união de extremidade não-homóloga, age frente a quebras na dupla fita de DNA. No entanto, neste caso a informação perdida vai ser recuperada o mais fielmente possível da molécula de cromossomo homóloga, que embora não seja idêntica, possui os mesmos alelos nas mesmas posições o que mantém um alto nível de fidelidade (OUYANG et al., 2008).

Nesta via, um complexo de proteínas formado por Rad50, Mre11 e Nbs1 vão se ligar aos fragmentos 3' das fitas de DNA danificadas, e proteínas da família RPA se ligam as extremidades 5' prevenindo que elas se liguem a outras moléculas. Esse complexo então quebra a dupla fita de DNA danificada em caudas de recombinação de fita única e, posteriormente, a proteína Rad51 irá buscar no cromossomo homólogo por sequências idênticas de pares de base e irá "invadir" essa sequência com uma de suas fitas de DNA. Essa fita liga-se à fita complementar do cromossomo homólogo deslocando a fita que estava posicionada ali, o que cria um "laço de deslocamento" (D-loop). A partir disso, DNA polimerases utilizam a fita complementar para criar um novo filamento de DNA, que se liga à fita invasora com quando o D-loop é desfeito. Esse novo filamento então liga-se ao fragmento danificado e será utilizado como molde para complementar a sequência perdida na lesão (HEYER et al., 2006).

1.3 Supressão Tumoral

Quando os mecanismos de estabilidade genômica falham em reparar as lesões e corrigir os erros no DNA, existem algumas alternativas emergenciais para evitar o crescimento celular desregulado. Essas alternativas são conhecidas como mecanismos de supressão tumoral e podem ser consideradas um tanto drásticas, pois alteram severamente o metabolismo

celular. Dois dos principais mecanismos são as vias de apoptose e senescência celular (KACZANOWSKI, 2016).

1.3.1 Apoptose

A apoptose é um importante mecanismo de supressão tumoral regulado pela proteína P53, que se caracteriza pela morte celular programada frente ao acúmulo de danos e erros de replicação no DNA celular. Esse fenômeno consiste na eliminação de células danosas ao organismo, para evitar que as mutações e os danos de uma célula se propaguem descontroladamente, a célula se autodestrói (DANIAL *et al.*, 2004)

O processo apoptótico tem como função a morte programada da célula e é majoritariamente coordenado pelo processo conhecido como cascata de caspases. Caspases são enzimas de ação hidrolítica que quebram as ligações peptídicas de proteínas alterando-as ou inativando-as e dessa forma interrompem os processos fisiológicos celulares desligando suas funções e preparando a célula para o processo de morte celular programada. Para evitar que causem danos inadequados, necessitam de ativação através de processos de dimerização, ou quebra de ligações, o que pode ser efetuado por outras caspases em uma cascata de reações exponenciais.

As caspases são classificadas em duas categorias referentes a sua função, caspases de ativação (como caspases 8 e 9) e caspases de execução (como caspases 3,6 e 7). as caspases de ativação são ativadas pelo processo de dimerização na etapa de indução, em contrapartida as caspases de execução são ativadas pelas caspases de ativação (ou pelas próprias caspases de execução) através da quebra da suas ligações (MCLLWAIN *et al.*, 2013).

A etapa de indução acontece de forma extrínseca, ou intrínseca à célula, como pode ser observado na figura 5, quando ocorre de forma extrínseca moléculas pró apoptóticas conhecidas como Death Ligands (como proteínas da superfamília Tumor Necrosis Factor - TNF) sinalizam para proteínas de membrana específicas (TNF receptors) essas proteínas por sua vez interagem com proteínas adaptadoras no interior da célula, criando um complexo proteico que sinaliza a indução da morte celular (Death-Inducing signaling complex-DISC) que por sua vez recruta e ativa a cascata de caspases, através da caspase-8, dando início ao processo de apoptose celular (GREEN *et al.*, 2015).

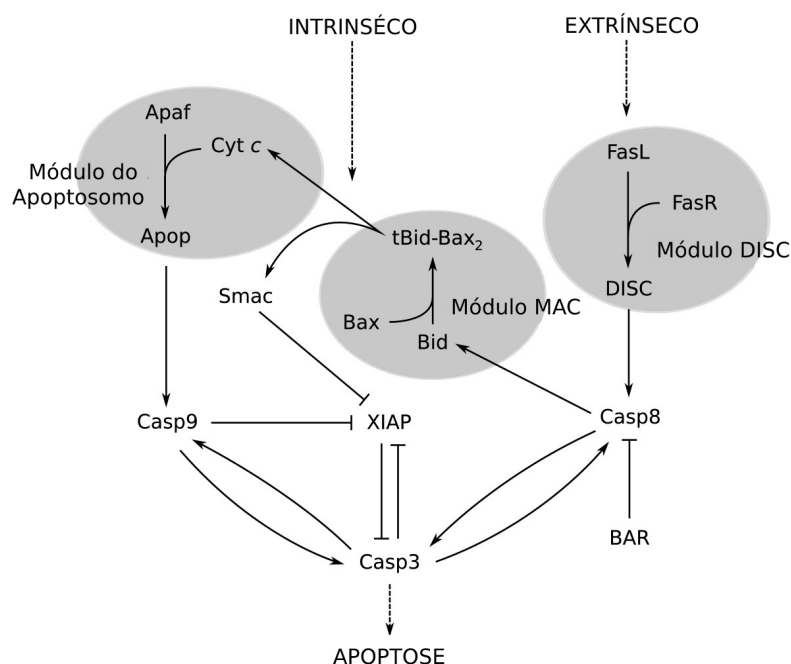


Figura 5: Via de apoptose, intrínseca: utilizando citocromo C para ativar a cascata de caspases através da caspase 9. E extrínseca a célula, formando o complexo proteico indutor da morte celular para ativar a cascata de caspases a partir de caspase 8 (Adaptado de: https://commons.wikimedia.org/wiki/File:Extrinsic_and_intrinsic_pathways_to_caspase-3_activation.jpg).

Quando a etapa de indução ocorre de forma intrínseca, a célula detecta danos ou erros acumulados de forma irreparável pelas vias de estabilidade genômica, ela então ativa proteínas indutoras de apoptose da família Bcl-2, essas proteínas agem na membrana mitocondrial tornando ela permeável e assim permitindo a liberação de citocromo c e outros fatores apoptóticos na matriz citoplasmática, esses fatores vão por sua vez ativar a cascata de caspases a partir de caspase-9 (SAVITSKAYA et al., 2015)

1.3.2 Senescência celular

Senescência celular é uma resposta alternativa a apoptose, que também é regulada pela proteína P53 e age frente a danos no DNA, mantendo a célula em um ciclo não reprodutivo e alterando seu metabolismo. Desta forma, a senescência promove longevidade em organismos multicelulares, pois impede que erros no DNA se acumulem e se propaguem e, desta maneira, suprime o surgimento de tumores no organismo (KOWALD et al., 2020).

O mecanismo da senescência celular é complexo e caracterizado pelo aprisionamento da célula em ciclo não reprodutivo, evitando assim o crescimento celular desregulado e demais instabilidades genômicas. Pode ser provocado como resposta a diferentes problemas celulares, como redução nos telômeros, danos no DNA, estresse oxidativo, entre outros. Esses problemas

ativam vias de sinalização que levam à proteínas responsáveis por inibir o ciclo reprodutivo celular (como p53, p27, p21, p16 e p15), e cada uma dessas vias tem um conjunto de proteínas atuando para dar a resposta adequada, que pode variar de acordo com o tipo celular ou a gravidade do problema (MUÑOSZ-ESPÍN et al., 2014).

Em um dos processos de resposta a danos no DNA, as proteínas ATM, ATR, CHK1 e CHK2 agem fosforilando a proteína p53 e por ação da p53 ativando p21. A proteína p21 é uma quinase inibidora dependente de ciclina que se liga a complexos de quinases dependentes de ciclina interrompendo a sua atividade e impedindo a síntese de DNA inabilitando o ciclo celular reprodutivo (FRIDLYANSKAYA et al., 2015).

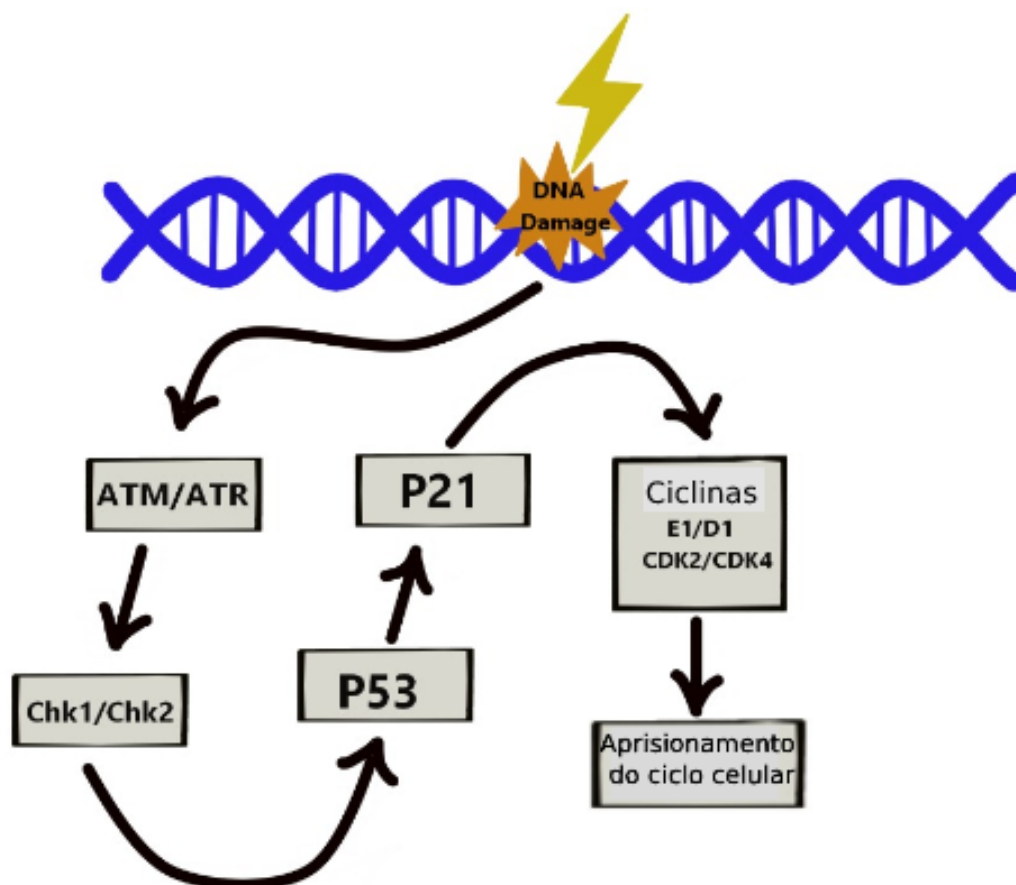


Figura 6: Via de ativação da senescência celular em resposta a danos no DNA, demonstrando a ação das proteínas ATM/ATR e Chk1/Chk2 em ativar P51 através da fosforilação de P53 e assim impedindo os complexos de kinases dependentes de ciclina, o que leva a célula a um ciclo não reprodutivo, figura do autor.

2. OBJETIVO

Tanto os mecanismos de estabilidade genômica quanto os de supressão tumoral são de suma importância para o bom funcionamento de organismos multicelulares, e apesar de fortemente relacionados ainda há muito o que se estudar a respeito da sua origem e evolução. Entender o panorama evolutivo dos mecanismos relacionados ao câncer permite uma visão mais abrangente de seu funcionamento, como seu surgimento e gradual aumento de complexidade de suas vias e de suas interações, culminando nos mecanismos que existem atualmente. (CASTRO *et al.*, 2008)

Com isto em mente, o presente trabalho é dedicado a entender a história evolutiva das vias que atuam na manutenção do genoma através de análises comparativas dos índices de plasticidade, diversidade e abundância dos componentes presentes nas vias dos mecanismos de senescência celular bem como, de apoptose e do conjunto de vias de estabilidade genômica.

3. MATERIAIS E MÉTODOS

3.1 Banco de Dados

Primeiramente, foram selecionadas 404 proteínas distintas presentes em humanos e que participavam de ao menos uma das seguintes vias de interesse: replicação de DNA, reparo por excisão de bases, reparo por excisão de nucleotídeos, reparo de incompatibilidade, união de extremidade não homóloga, recombinação homóloga, apoptose e senescência celular. Estas proteínas foram selecionadas a partir da Kyoto Encyclopedia of Genes and Genomes (KEGG), um banco de dados que contém informações a respeito de diferentes vias de funções celulares, como quais moléculas participam destas vias e como elas interagem umas com as outras formando mecanismos complexos de respostas celulares.

Em seguida, definiu-se uma árvore filogenética contendo 477 espécies de eucariotos, selecionadas com base na versão 11 do banco de dados STRING, que possui amplas informações a respeito de proteínas, bem como suas relações e interações. Uma simplificação do cladograma pode ser observado na figura 7 e, na árvore original, existem 37 ramos indicando o ancestral comum dos diversos grupos de eucariotos.



Figura 7: Simplificação do cladograma de eucariotos considerando humanos como o topo da árvore, figura do autor.

3.2 Inferência de Índice de Plasticidade evolutiva

Como objeto da análise, foram definidos grupos de proteínas eucarióticas distintas, mas que possuem uma origem comum entre si e que conservam funções similares (Grupos Ortólogos Eucaróticos ou KOGs). Tendo como base informações dos conjuntos de grupos ortólogos presentes na versão 11 do banco de dados STRING, e filtrando-o apenas nas 404 proteínas selecionadas e também apenas nas 477 espécies contidas na árvore filogenética, obteve-se 85 grupos ortólogos (ALTENHOFF *et al.*, 2012).

Baseando-se, então, na abundância e diversidade desses grupos ortólogos ao longo da árvore, foi calculado seu índice de plasticidade usando o software R, versão 4.1.2. que pode ser obtido no repositório “The Comprehensive R Archive Network” (CRAN) que é uma rede de servidores que visam manter o versionamento do R devidamente documentado, e as funções do pacote Geneplast versão 1.18.0 (Dalmolin *et al.*, 2015) do Bioconductor versão 3.15.

Uma das funções dessa ferramenta é observar a diversidade e a abundância dos grupos ortólogos dentro da árvore estabelecida, sendo a diversidade o quão espalhado este grupo de ortólogos está entre as espécies da árvore, e é calculado pela equação normalizada de Shannon, aqui representada pela expressão:

$$H\alpha = - \frac{1}{\ln M} \sum p(i, \alpha) \ln p(i, \alpha)$$

Onde “H α ” representa a diversidade de um dado KOG α e “M” representa o número total de organismos analisados, assim para cada dado KOG α a equação vai calcular a frequência que este KOG ocorre nesta árvore, utilizando a somatória da probabilidade dele estar presente em cada uma das espécies selecionadas, com base na quantidade de genes ortólogos e parálogos em comum deste dado KOG com as espécies (DALMOLIN *et al.*, 2011).

E a abundância representa o quanto este grupo de ortólogos possui parálogos na árvore, ou seja, quantas “cópias” destes ortólogos há em um mesmo organismo, sendo importante lembrar que no caso dos parálogos após o evento de duplicação as cópias resultantes tendem a variar sua função ao longo das gerações.

A partir disto, a ferramenta calcula o índice de plasticidade pela dada expressão:

$$EPI = 1 - \frac{DIVERSIDADE}{\sqrt{ABUNDÂNCIA}}$$

Assim sendo, quanto maior a diversidade de um grupo ortólogo menor sua plasticidade e quanto maior sua abundância, maior sua plasticidade (DALMOLIN *et al.*, 2011). A plasticidade é um indicador da capacidade de um genótipo de resistir ou tolerar processos de alterações e se adaptar a eles, muitas vezes adquirindo novas funções (KELLY *et al.*, 2012).

O valor de diversidade é uma escala que varia entre 0 e 1, sendo que um KOG com valor de diversidade 1 estaria presente em todos os organismos da árvore evolutiva estudada, e de forma inversa um KOG com valor de diversidade 0 não estaria presente em nenhuma das espécies do cladograma em questão. Por outro lado o valor de abundância começa em 1 mas não tem um valor limitante pois se refere a uma medida quantitativa de eventos de duplicação, onde um KOG que possua valor de abundância 1 teria permanecido inalterado após seu surgimento ao passo que um KOG que possua valor de abundância elevado, como algumas centenas, teria também centenas de parálogos.

Portanto um KOG com valor de diversidade e abundância baixas indicaria que ele ocorre em poucas espécies e que ele possui poucos parálogos o que aponta para uma origem recente na árvore evolutiva, isso condiz com uma plasticidade alta o suficiente para aceitar surgimento de novos genes.

Para KOGs com alto valor de abundância e baixo valor de diversidade entende-se que ele está presente em poucas espécies e que passou por diversos processos de duplicação, assim sendo sua plasticidade seria muita alta pois aceita com facilidade alterações em seus genes.

Quanto a KOGs com baixo valor de abundância mas com alto valor de diversidade, observa-se um índice de plasticidade extremamente baixo, o que indica uma origem bem antiga na árvore evolutiva, pois estes genes estão presentes em quase todos os organismos do cladograma e mantiveram-se praticamente inalterados ao longo do processo evolutivo.

Por último, entende-se que KOGs com ambos os valores de abundância e diversidade altos, estão presentes em vários organismos da árvore evolutiva em questão e possuem vários parálogos, o que também é um indicativo de alta plasticidade, pois estes genes são capazes de aceitar modificações.

3.3 Análise Estatística

Para avaliar se há ou não diferenças significativas entre os valores obtidos através da análise de plasticidade realizada pelo Geneplast, é necessário uma análise estatística que compara valores de variância de mais de dois grupos considerando um único fator (no caso vias de função celular). Esta análise vai confirmar se os valores obtidos são diferentes por conta de participarem de vias distintas, ou se a diferença observada se deve ao acaso.

Em razão de definir qual teste é o ideal para este caso, é necessário optar por um teste paramétrico ou não paramétrico. Um teste paramétrico parte do pressuposto de que as amostras possuem distribuição normal e variâncias homocedásticas.

Com intenção de verificar estes pressupostos, realizou-se o teste de Shapiro-Wilk, que parte da hipótese nula de que os dados de determinado grupo amostral possuem uma distribuição normal. Se o valor P do teste for superior a um alfa (considerado 5%, para os propósitos deste trabalho), confirma-se a hipótese nula, caso o contrário (o valor de P for inferior a 5%), rejeita-se a hipótese nula e, portanto, a amostra possui distribuição não normal. Este teste foi realizado utilizando o software R sem nenhum pacote adicional.

De forma similar, realiza-se um teste de Levene para verificar a homogeneidade das variâncias (homocedasticidade) de dois grupos ou mais, partindo da hipótese nula de que há homocedasticidade entre as variâncias destes grupos. Entende-se que com um valor de P superior a 5%, confirma-se essa hipótese, caso o contrário, com um valor de P inferior a 5% rejeita-se a hipótese nula e entende-se que não há homocedasticidade entre as variâncias dos grupos. Este teste foi realizado no software R utilizando o pacote “car: Companion to Applied Regression” (Fox et al., 2019).

Posteriormente a verificação dos pressupostos decidiu-se por um teste não paramétrico, conhecido como teste de Kruskal-Wallis. Este teste está presente no pacote básico do R e rejeita a hipótese de que os grupos não possuem diferenças significativas quando o valor de P for inferior a 5%. Após estes testes acusarem diferenças significativas entre os grupos, realizou-se o teste de Dunn para observar quais grupos são diferentes dos demais. O teste de Dunn pode ser realizado no software R utilizando-se do pacote “dunn.test” (Dinno, 2017). Ele compara dois grupos por vez, até cada grupo ter sido comparado com os demais, e acusa diferença entre dois grupos quando o valor de P for inferior a 5%.

4. RESULTADOS

Tendo obtido os valores de abundância, diversidade e também o índice de plasticidade evolutiva para cada um dos grupos de genes ortólogos (KOG), separou-se os KOGs em três grupos de acordo com suas vias de atuação, KOGs que atuavam em estabilidade genômica, apoptose e senescência celular, como demonstrado na figura 8, e então foi verificado se haviam diferenças estatísticas entre os valores de abundância, diversidade e plasticidade de cada um destes grupos em relação aos demais.

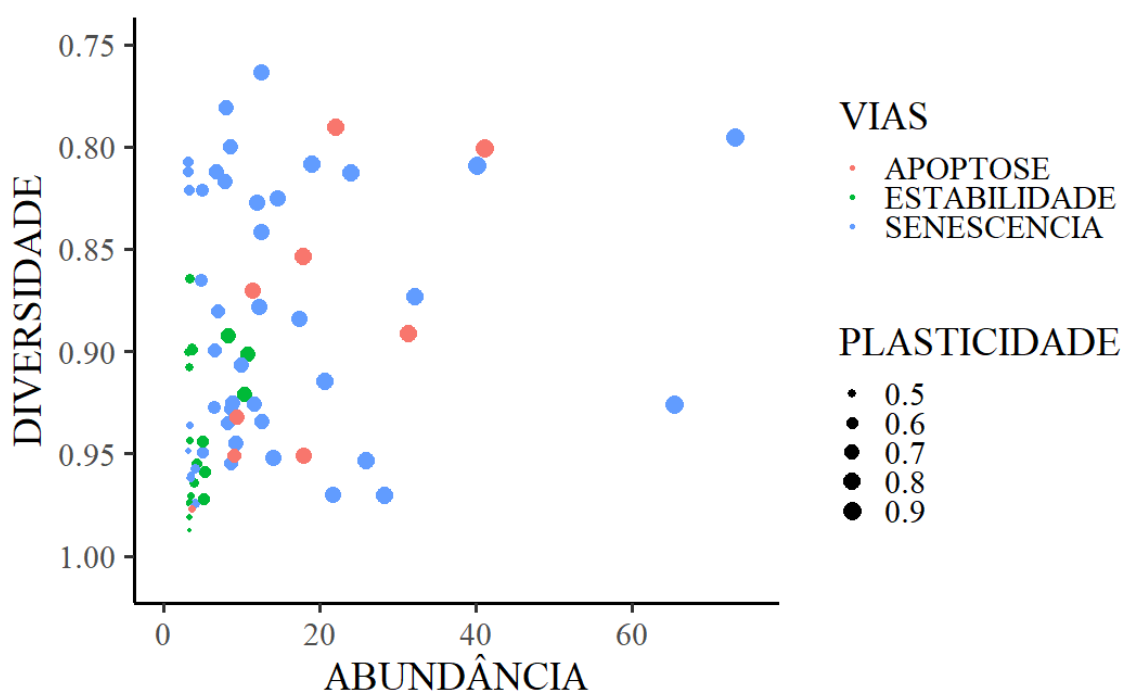


Figura 8: Gráfico de pontos indicando o valor de diversidade e abundância de cada um dos KOGs, onde cada ponto representa um KOG e sua coloração indica a via em que ele participa e o tamanho do ponto indica seu índice de plasticidade evolutiva, figura do autor.

Para isso, primeiramente é necessário obter informações a respeito da normalidade da distribuição destes valores, bem como teste da homocedasticidade de suas variâncias, o que foi realizado através do teste de Shapiro Wilk e teste de Levene. O resultado mostrou que apesar de haver normalidade em algumas distribuições não havia robustez suficiente em nenhum dos grupos para justificar um teste paramétrico, portanto prosseguiu-se para análise não paramétrica de um único fator, teste conhecido como Kruskal-Wallis, que revela se há ou não diferenças estatisticamente relevantes entre os três grupos.

O teste de Kruskal-Wallis mostrou que tanto para os valores de diversidade quanto para os de abundância e ainda para os de plasticidade haviam diferenças significativas. Portanto prosseguiu-se com o teste a

posteriori de Dunn, que indica quais grupos são estatisticamente distintos dos demais.

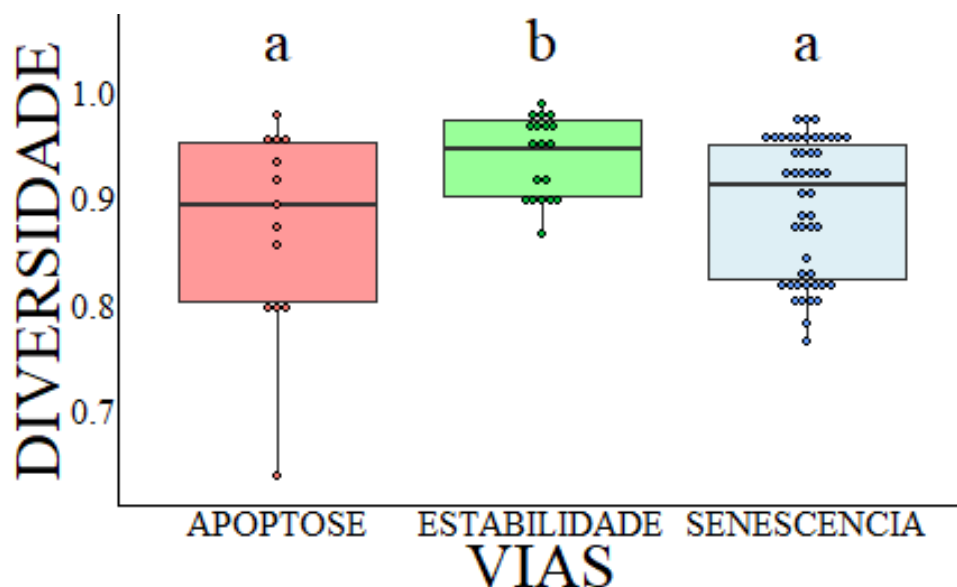


Figura 9: Box Plot comparando os valores de diversidade dos três grupos de KOGs. Obtido com a realização de um teste de Kruskal-Wallis seguido por um teste a posteriori de Dunn, figura do autor.

Os resultados demonstraram que em questão dos valores de diversidade os grupos de apoptose e senescência celular são estatisticamente idênticos entre si e ambos são inferiores ao valor de diversidade do grupo de estabilidade genômica, como evidenciado na figura 9.

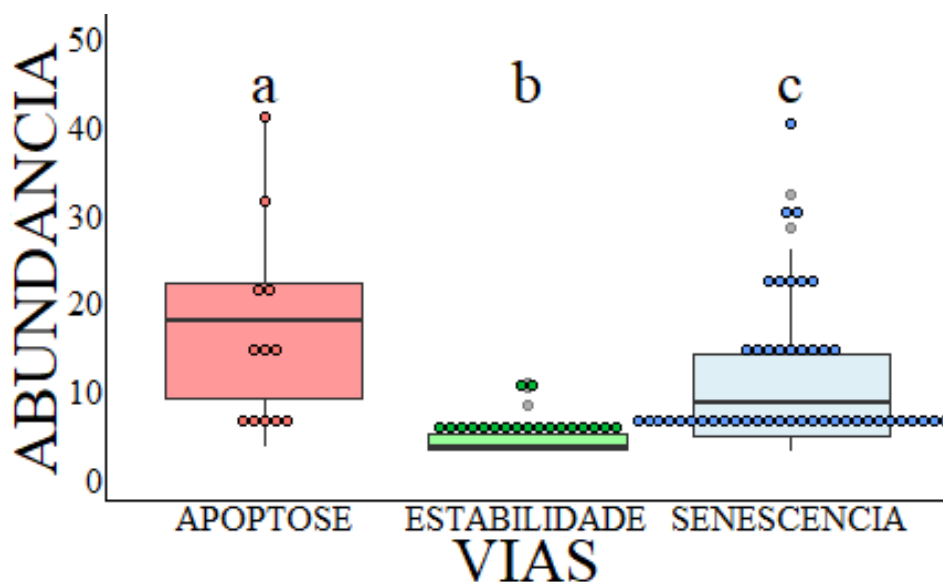


Figura 10: Box Plot comparando os valores de diversidade dos três grupos de KOGs. Obtido com a realização de um teste de Kruskal-Wallis seguido por um teste a posteriori de Dunn, figura do autor.

Por outro lado, em relação aos valores de abundância, a figura 10 mostra o grupo de apoptose com o maior valor, seguido por senescência celular, enquanto o grupo de estabilidade genômica tem os menores valores.

E finalmente pode ser observado na figura 11, que os valores de plasticidade novamente mostram apoptose e senescência celular como sendo grupos estatisticamente indistinguíveis e ambos os valores são superiores ao do grupo de estabilidade genômica.

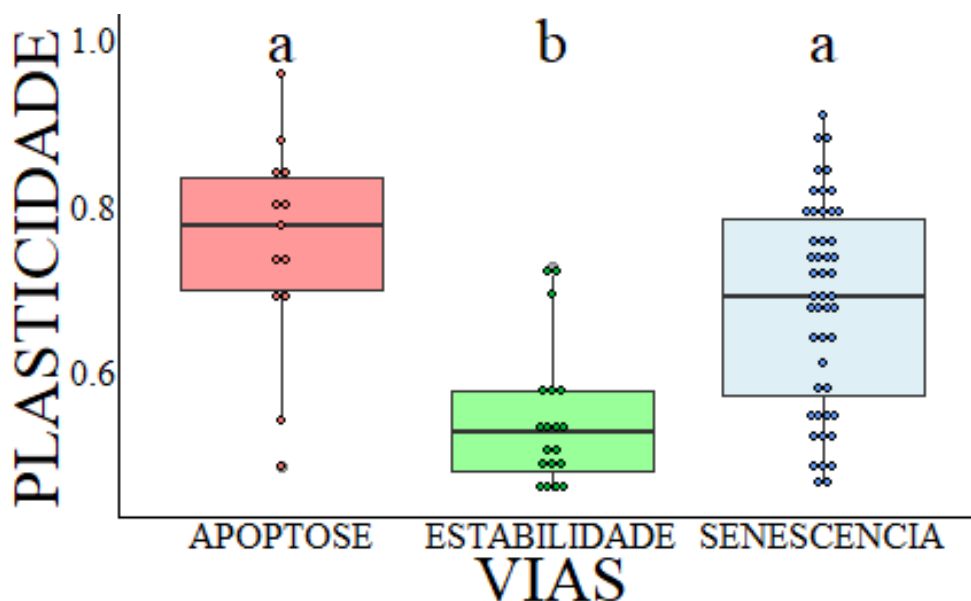


Figura 11: Box Plot comparando os valores de plasticidade dos três grupos de KOGs. Obtido com a realização de um teste de Kruskal-Wallis seguido por um teste a posteriori de Dunn, figura do autor.

5. DISCUSSÃO

Os resultados mostraram que o grupo de estabilidade genômica possui a maior diversidade dentre os três, o que indica que é a via que possui maior presença entre os organismos da árvore e, portanto, a que estava presente no ancestral comum do maior número de espécies. Ao mesmo tempo, esse grupo possui o menor valor de abundância, o que indica que há poucas cópias desses genes nos organismos. Por conseguinte, seus índices de plasticidade evolutiva possuem os menores valores, o que indica um grupo de genes com baixa aceitação às mutações e alterações na forma como estes mecanismos funcionam. Evidentemente, este conjunto de ortólogos é essencial tanto em organismos unicelulares, como em organismos mais complexos, pois aumenta o fitness e ajuda na manutenção dos tecidos, portanto qualquer alteração em seus genes poderia ser catastrófica e dificilmente seria selecionada, o que

explica sua aparição antiga na árvore de eucariotos e sua presença constante e quase inalterada ao longo de toda a história evolutiva deste grupo.

Quanto aos ortólogos das vias de apoptose e senescência celular nota-se que eles são estatisticamente idênticos em relação aos valores de diversidade, o que é um forte indicador que eles tiveram aparições em um mesmo ancestral comum, ou muito próximo, e como seus valores são menores que os de ortólogos da via de estabilidade genômica pode-se entender que sua origem é mais recente. Porém, seus valores de abundância são distintos e maiores nos ortólogos que atuam na via de apoptose ou seja passaram por mais eventos de duplicação e portanto possuem mais parálogos em cada espécie, o que também indicaria que este grupo ortólogo tem uma aceitação maior a alterações em seus componentes, apesar do que mostra seus índices de plasticidade evolutiva.

No artigo “Evolutionary origins of human apoptosis and genome-stability gene networks” de Mauro Castro (2008), o autor observou um crescimento repentino no número de grupos ortólogos de apoptose próximo a origem evolutiva dos metazoários e hipotetizou que uma possível explicação para essa ocorrência seria o aparecimento de outra via de supressão tumoral como a senescência celular, que teria aumentado a plasticidade desta via ao relaxar sua pressão seletiva, aumentando sua tolerância a episódios de mutação.

Ao analisar comparativamente os valores de diversidade e abundância dos grupos ortólogos de estabilidade genômica, apoptose e senescência celular, o objetivo era verificar em qual momento da história evolutiva a via de senescência celular teria surgido em relação às outras, e também como este surgimento teria impactado as demais vias. Porém, verificou-se que através destes valores o momento de surgimento de ambas as vias de supressão tumoral aparenta ter ocorrido simultaneamente, o que impossibilita estabelecer uma relação de causa e efeito entre o aparecimento de uma e a pressão seletiva exercida sobre a outra.

Ainda assim este resultado não descarta a possibilidade da presença da via de senescência celular ter sido a responsável pelo crescimento observado em apoptose, pois apesar de terem um ponto de origem evolutiva muito próximo, este não é o único indicativo que poderia confirmar esta relação. Se fosse analisado os exatos pontos de surgimento de cada KOG presente nessas vias poderia-se obter novos vislumbres a respeito de suas influências uma na outra.

6. CONCLUSÃO

Embora os resultados não tenham sido conclusivos, os valores de abundância de um grupo ortólogo na árvore filogenética de eucariotos é apenas uma das possíveis abordagens que podem ser utilizadas para se analisar a história evolutiva molecular dos mecanismos celulares. Portanto mais estudos se fazem necessários para averiguar esta teoria, especificamente visando estabelecer o surgimento e crescimento de ambas as vias de supressão ao longo da árvore evolutiva de eucariotos e compará-los com os índices de plasticidade evolutiva da outra via, buscando por sinais de correlação.

REFERÊNCIAS BIBLIOGRÁFICAS

ALGUNS NÚMEROS DO CÂNCER NO BRASIL E NO MUNDO. **Sociedade Brasileira de Cancerologia**, 2016. Disponível em: <<http://www.sbcancer.org.br/alguns-numeros-do-cancer-no-brasil-e-no-mundo/>>. Acesso em 06/04/2022

ALTENHOFF, A.M., *et al.* Resolving the ortholog conjecture: orthologs tend to be weakly, but significantly, more similar in function than paralogs. **PLoS Comput Biol.** 2012

COOPER, G. M. **The Cell: A Molecular Approach**. 2º edição. Sunderland (MA): Sinauer Associates; 2000. The Development and Causes of Cancer. Disponível em: <<https://www.ncbi.nlm.nih.gov/books/NBK9963/>>. Acesso em 05/05/2022

DALMOLIN R.J.S., CASTRO M.A.A. Geneplast: evolutionary rooting and plasticity inference. **R package**, 2015.

DALMOLIN, R.J.S. *et al.* Evolutionary plasticity determination by orthologous groups distribution **Biology Direct**. 2011

DANIAL, N. N., *et al.* Cell death: critical control points. **Cell**. v. 23, 2004

DINNO A. dunn.test: Dunn's Test of Multiple Comparisons Using Rank Sums. **R package version 1.3.5.**, 2017 Disponível em <<https://CRAN.R-project.org/package=dunn.test>> Acesso em 26/05/2022

FOX J., WEISBERG S. Car: Companion to Applied Regression, Third Edition. **Thousand Oaks CA:** Sage., 2019. Disponível em <<https://socialsciences.mcmaster.ca/jfox/Books/Companion/>> Acesso em 26/05/2022

FRIDLYANSKAYA, I. *et al.* Senescence as a general cellular response to stress: A mini-review. **Experimental Gerontology**, 2015

GREEN, D. R.; LLAMBI, F. Cell Death Signaling. **Cold Spring Harb Perspect Biol.** v. 7(12), 2015

GRIFFITHS, A. J. F., *et al.* **Introdução à Genética**, 10ª Edição, Guanabara Koogan Ltda., 2013, p. 503

HEYER, W.-D., *et al.* Rad54: the Swiss Army knife of homologous recombination? **Nucleic Acids Research**, v. 34(15), p. 4115–4125 2006

KACZANOWSKI, S. Apoptosis: its origin, history, maintenance and the medical implications for cancer and aging. **Phys Biol**. v. 11, 2016

KEGG: Kyoto Encyclopedia of Genes and Genomes. Kyoto University, Disponível em: <<https://www.genome.jp/kegg/pathway.html>>. Acesso em 12/05/2021

KELLY SA, Panhuis TM, Stoeckl AM. Phenotypic plasticity: molecular mechanisms and adaptive significance. **Compr Physiol**. 2012 Apr;2(2):1417-39

KOWALD, A. *et al.* On the evolution of cellular senescence. **Aging Cell** v. 19, p. 12, 2020

LI, G.-M. Mechanisms and functions of DNA mismatch repair. **Cell Research**, v. 18(1), p. 85–98, 2007

MARTI, T. M., *et al.* DNA mismatch repair and mutation avoidance pathways. **Journal of Cellular Physiology**, v. 191(1), p. 28–41, 2002

MAURO, A. A. *et al.* Evolutionary origins of human apoptosis and genome-stability gene networks, **Nucleic Acids Research**, v. 36, p. 6269–6283, 2008

MCELLWAIN, D. R., *et al.* Caspase Functions in Cell Death and Disease. **Cold Spring Harbor Perspectives in Biology**, 2013

MUÑOZ-ESPIN, D.; SERRANO, M. Cellular senescence: from physiology to pathology. **Nature Reviews Molecular Cell Biology**, 2014

OUYANG, K. J., *et al.* Homologous recombination and maintenance of genome integrity: Cancer and aging through the prism of human RecQ helicases. **Mechanisms of Ageing and Development**, v. 129(7-8), p. 425–440, 2008

ROBERTSON, A. B. *et al.* DNA repair in mammalian cells: Base excision repair: the long and short of it. **Cell Mol Life Sci**. 2009

R. versão 4.1.2: R Core Team, 1993. Disponível em: <<https://cran.r-project.org/bin/windows/base/old/>>. Acesso em: 12 janeiro de 2022

RODIER, F., *et al.* Two faces of aging and tumor suppression. **Nucleic Acids Res.**, v. 35, p. 53, 2007

SALDIVAR, J. S. *et al.* Nucleotide excision repair pathway review I: Implications in ovarian cancer and platinum sensitivity. **Gynecologic Oncology**, v.107(1), p. S56–S71, 2007

SAVITSKAYA, M. A.; ONISHCHENKO, G. E. Mechanisms of apoptosis. **Biochemistry (Moscow)**, v. 80(11), p. 1393–1405, 2015

STRING: functional protein association networks. versão 11.0: Academic Consortium. Disponível em:
<https://version-11-0.string-db.org/cgi/input.pl?sessionId=g0O2gLREHJNW&input_page_show_search=on>

WYMAN, C.; KANAAR, R. DNA Double-Strand Break Repair: All's Well that Ends Well. **Annual Review of Genetics**, V. 40(1), P. 363–383, 2006