

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL**

**ANDREA GOMES TAVANTI**

**Reconstrução da história filogenética e distribuição do retrotransposon CR1\_E  
em genomas disponíveis em banco de dados.**

**Porto Alegre  
2018**

**ANDREA GOMES TAVANTI**

**Reconstrução da história filogenética e distribuição do retrotransposon CR1\_E  
em genomas disponíveis em banco de dados.**

Trabalho de Conclusão de Curso  
apresentado ao Curso de (Ciências  
Biológicas) da Universidade Federal do  
Rio Grande do Sul, como requisito parcial  
para obtenção do Título de Bacharel em  
(Ciências Biológicas).

Orientadora: Prof.<sup>a</sup> Dr.<sup>a</sup> Maríndia Deprá

Coorientadora: Msc. Natasha Ávila  
Bertocchi

**Porto Alegre  
2018**

## AGRADECIMENTO

Não está em uma ordem específica porque todas as pessoas que quero agradecer contribuíram de formas diferentes e não menos importantes.

Começo com a minha família, que em todos os dias da minha vida me deu todas as condições necessárias para ter uma educação de qualidade. Obrigada a minha mãe, Niara, pelo amor e por sempre ter me escutado e me aconselhado nas horas difíceis. Obrigada ao meu pai, Ricardo, pelo amor e por ser uma fonte inesgotável de sabedoria, persistência, carinho e exemplo. Obrigada aos meus irmãos Umberto, Márcio e Marcelo por compartilharem suas vivências e por serem amigos incondicionais. Obrigada aos meus tios Maristela, Nestor e Renato que são acima de tudo grandes amigos que acalmam o meu coração sempre que eu preciso. Eu amo cada um de vocês e sem vocês essa conquista seria quase impossível.

Obrigada Lucas, que durante esses 13 anos de amor e carinho sempre me apoiou. Nas piores horas, tua energia me impulsionou para caminhos melhores e vou te agradecer eternamente por isso. Eu te amo!

A minha amiga Tuti, que me escuta, que me aconselha estando disposta a ajudar sempre, não importa o tamanho do problema. Uma irmã de coração!

As minhas outras famílias, Fialho e Zawacki que são mais uma fonte de amor e apoio familiar na minha vida. Obrigada por TUDO!

Obrigada Professora Maríndia, que me acolheu como orientada, com muito carinho e atenção e me introduziu ao mundo dos transposons. Muito obrigada pela oportunidade de trabalhar contigo, torço para que possamos repetir isso!

Obrigada Natasha, por toda a ajuda com o TCC, pelo incentivo e pelo ouvido amigo com as indecisões do futuro. Espero que possamos trabalhar mais juntas!

E finalmente, obrigada a todas as pessoas que fazem parte do Laboratório de Drosophila. Trabalhar com vocês em um ambiente cooperativo, feliz e amigável fez toda a diferença na minha vida.

Obrigada! Vocês são TUDO!

## RESUMO

Elementos transponíveis (TEs) são sequências que possuem a particularidade de se mobilizar nos genomas, estando presentes ubiquamente nos organismos. Os retrotransposons *Chicken repeat 1* (CR1) constituem uma superfamília de TEs com estudos incipientes e restritos aos genomas aviários. O objetivo desse trabalho foi reconstruir a história filogenética e distribuição de elementos CR1\_E-like através da busca por sequências homólogas disponíveis em bancos de dados. Para isso, utilizamos as sequências CR1\_E\_pass, CR1\_E\_PPU e CR1\_E\_VSP como sonda para buscas, através do BLASTn, nos bancos de dados *Genbank* e *Flybase*. Somente os resultados de BLAST menores que  $e^{-10}$  foram analisados. As sequências foram alinhadas utilizando os parâmetros *default* da plataforma MAFFT v. 7 e os resultados visualizados e manualmente editados através do software *Aliview*. Foram construídas matrizes de identidade através do software UGENE para identificação da conservação das sequências. Por fim, foram estabelecidas as relações filogenéticas das três sondas com suas sequências homólogas, e das mesmas com o acervo de TEs CR1 disponíveis no banco de dados *Repbse*. Utilizamos o método *Maximum Likelihood* (ML) através do software MEGA. Observamos sequências principalmente nas maiores linhagens de amniotas (mamíferos, répteis e aves), além de grupos distintos como peixes e bactéria (*Escherichia coli*). Portanto a distribuição de sequências CR1\_E-like não foram de acordo com a filogenia dos genomas hospedeiros, sugerindo possíveis eventos de transferência horizontal. Além disso, a presença destes TEs nos genomas de amniotas corrobora a hipótese de terem sido inseridos no ancestral em comum desse grupo, porém evoluíram distintamente em cada linhagem.

Palavras-chave: Elementos transponíveis; *Long Interspersed Nuclear Element* (LINE); Aves

## ABSTRACT

Transposable Elements (TEs) are sequences that have the ability to move in the genome and are present ubiquitously throughout organisms. The Chicken repeat 1 retrotransposons (CR1) constitute a superfamily of TEs with incipient and restricted studies within avian genomes. This work aims to reconstruct the phylogeny history and distribution of the CR1\_E-like elements by searching for homologous sequences available in on-line data banks. CR1E\_pass, CR1\_E\_PPU and CR1\_E\_VSP were used as reference sequences when performing BLASTn on Genbank and on Flybase and only BLAST results with E-value lower than  $e^{-10}$  were analyzed. Sequences were aligned using MAFFT v.7 with default parameters and results were visualized and manually edited through Aliview software. Distance matrices were built using UGENE software in order to identify the conservation of the sequences. Finally, phylogenetic relationships were established among reference sequences and their respective homologous sequences as well as with CR1 TE sequences available on Rebase. Maximum Likelihood method was applied using MEGA software to construct phylogenetic trees. Our BLAST results showed mainly major amniotes groups (mammals, reptiles and birds) and distinct groups (fishes and bacteria - *Escherichia coli*) as hits to the reference sequences. Thus, distribution of CR1\_E like sequences was not only detected according to the phylogeny of host genomes which could be explained by horizontal transfer events. Furthermore, the presence of CR1\_E-like TEs on amniote genomes corroborates the hypothesis that they were present in the common ancestor of this group, although they evolved distinctly in each lineage.

Keywords: Transposable elements, Long Interspersed Nuclear Element (LINE), Birds

## LISTA DE TABELAS

Tabela 1. Espécies com sequências homólogas ao elemento CR1_E_Pass.....	24
Tabela 2. Espécies com sequências homólogas ao elemento CR1_E_PPU.....	25
Tabela 3. Espécies com sequências homólogas ao elemento CR1_E_VSP.....	26
Tabela 4. Ausência/presença de sequências de espécies homólogas com relação às sondas.....	27

## LISTA DE FIGURAS

Figura 1. Esquema gráfico do retroelemento CR1. ....	13
Figura 2. Esquema geral de um elemento CR1 .....	16
Figura 3. Conservação das sequências nucleotídicas.....	28
Figura 4. Filogenia de CR1_E_Pass e suas sequências homólogas. ....	29
Figura 5. Filogenia de CR1_E_PPU e suas sequências homólogas.....	30
Figura 6. Filogenia de CR1_E_VSP e suas sequências homólogas.....	31
Figura 7. Árvore guia de alinhamento entre os elementos CR1.....	32

## SUMÁRIO

<b>1 INTRODUÇÃO .....</b>	<b>9</b>
<b>1.1. Histórico.....</b>	<b>9</b>
<b>1.2. Caracterização e classificação dos elementos de transposição .....</b>	<b>11</b>
<b>1.3 Distribuição, classificação e caracterização de TEs da superfamília CR1... </b>	<b>13</b>
<b>2 JUSTIFICATIVA E OBJETIVOS .....</b>	<b>15</b>
<b>3 METODOLOGIA .....</b>	<b>15</b>
<b>3.1 Sequências de referência .....</b>	<b>15</b>
<b>3.2 Buscas <i>in silico</i> em diferentes bancos de dados e alinhamentos .....</b>	<b>16</b>
<b>3.3 Filogenias e matrizes de distância .....</b>	<b>17</b>
<b>4 RESULTADOS .....</b>	<b>18</b>
<b>4.1 Sequências homólogas .....</b>	<b>18</b>
<b>4.2 Conservação nucleotídica das sequências homólogas .....</b>	<b>18</b>
<b>4.3 Análise comparativa das sequências homólogas .....</b>	<b>19</b>
<b>4.4 Relações filogenéticas entre elementos CR1 .....</b>	<b>20</b>
<b>5 DISCUSSÃO .....</b>	<b>21</b>
<b>6 CONSIDERAÇÕES FINAIS .....</b>	<b>23</b>
<b>REFERÊNCIAS.....</b>	<b>33</b>

# 1 INTRODUÇÃO

## 1.1 Histórico

Os séculos XIX e XX foram muito relevantes com relação aos avanços no âmbito das Ciências Biológicas. Durante esse período os estudos de base em áreas como embriologia, hereditariedade e evolução foram realizados em nível de células, de tecidos ou de populações e foram gradualmente conectados com os avanços dentro da biologia molecular pós segunda guerra mundial (ALLEN, 1978)

Como exemplos de avanços importantes desse período temos a descoberta do núcleo das células, a formulação da teoria de que uma origem celular é comum a todos os seres vivos, a descrição do processo de mitose e a descoberta e descrição dos cromossomos (SEDGWICK; TYLER, 1939). Pode-se destacar também, a formulação das “Leis de Mendel” – pois tratavam de como as características seriam transmitidas pelas gerações – e a explicação inicial de alguns mecanismos e processos que ocasionam a diversidade da vida na Terra com a publicação da obra seminal “A origem das espécies”.

Após a metade do século XX marcos importantes aconteceram, como a elucidação da estrutura do ácido desoxirribonucleico – DNA, do inglês *Deoxyribonucleic acid* – e a descoberta dos elementos de transposição – TEs, do inglês *Transposable Elements*, – (MCCLINTOCK, 1956; WATSON; CRICK, 1953). Essas descobertas revolucionaram os campos da biologia molecular, hereditariedade e evolução, abrindo caminho para o estudo dos genomas de diferentes organismos.

O trabalho realizado por Barbara McClintock, quebrou um dos principais paradigmas com relação ao material genético, a ideia de que ele era estático e a informação contida nele também. McClintock mostrou que a visão sobre a estaticidade do genoma estava equivocada através de sua pesquisa inovadora na qual descrevia os TEs como “elementos controladores do gene” (MATIOLI; FERNANDES, 2012). Ela os descobriu através da observação de rearranjos cromossômicos que promoviam quebras no cromossomo 9 do milho (*Zea mays*). Esse movimento foi descrito como aleatório e ocasionava uma coloração em mosaico nos pericarpos do milho (CARARETO; MONTEIRO-VITORELLO; SLUYS,

2015; JONES, 2005; MCCLINTOCK, 1951). Na época de seus achados McClintock sofreu grande rejeição pelos acadêmicos da área e seu trabalho foi classificado como “obscuro, misterioso e herético” (JONES, 2005). Porém, com os avanços das técnicas de biologia molecular, a partir dos anos 70, a pesquisa realizada pela cientista foi reconhecida como revolucionária garantindo à pesquisadora Barbara McClintock, na década de 80, o prêmio Nobel de fisiologia ou medicina pela descoberta (CARARETO; MONTEIRO-VITORELLO; SLUYS, 2015).

Nos últimos 40 anos, com o desenvolvimento da técnica de sequenciamento de DNA por Sanger, em 1977, iniciaram-se grandes esforços com o objetivo de sequenciar o genoma de diversos organismos (MATIOLI; FERNANDES, 2012). Além disso, o desenvolvimento da técnica de Reação em Cadeia da Polimerase (PCR), em 1983, possibilitou grandes avanços na manipulação do DNA (MATIOLI; FERNANDES, 2012). Com isso, progredimos no entendimento de como o genoma é organizado e estamos descobrindo novos elementos que o compõem.

Atualmente, com o Sequenciamento de Nova Geração (NGS, do inglês *Next Generation Sequencing*) tem-se um contínuo aumento de genomas sequenciados, como no projeto G10k que visa sequenciar dez mil genomas (KOEPLI; PATEN; O'BRIEN, 2015). Como resultado desses esforços, uma quantidade extraordinária de dados moleculares é depositada em diferentes bancos de dados (MATIOLI; FERNANDES, 2012). Alguns exemplos de bancos de dados são o *Genbank*, que armazena sequências genéticas anotadas, o *Repbase*, que armazena elementos do DNA repetitivo e o *Flybase*, que armazena informações genéticas e moleculares específicas de insetos (BAO; KOJIMA; KOHANY, 2015; CLARK et al., 2016; GRAMATES et al., 2017). Com a quantidade massiva de dados amplamente disponíveis pesquisas relacionadas à estruturação e à organização do genoma dos seres vivos foram facilitadas, porém esses dados ainda são pouco explorados. A principal maneira de avançar nos estudos utilizando esses dados moleculares são através da chamada pesquisa *in silico*. Os estudos *in silico* possibilitaram muitas descobertas, com ênfase no estudo de elementos de transposição. Alguns destaques são: a descoberta do TE *Helitron*, e o aumento significativo de eventos de transferência horizontal de elemento transponível (HTT, do inglês *Horizontal Transfer of Transposable Element*). Atualmente estão registrados 2.248 eventos de HTT que

ocorreram ao longo de 10 milhões de anos em 195 genomas de insetos (PECCOUD et al., 2017).

## **1.2. Caracterização e classificação dos elementos de transposição**

Os TEs são sequências de DNA encontradas repetidamente nos genomas dos organismos, que possuem a capacidade de se mobilizar dentro e entre genomas conferindo uma maior flexibilidade e dinâmica ao genoma hospedeiro (CARARETO; MONTEIRO-VITORELLO; SLUYS, 2015). Encontram-se presentes em todos os ramos estudados da árvore da vida, desde organismos como bactérias e fungos, como também em vegetais, invertebrados e todas as classes de vertebrados (PRITHAM, 2009).

Esses elementos do DNA foram chamados de “DNA lixo”, “genes saltadores”, “parasitas genômicos” e “DNA egoísta” por conseguirem se inserir e copiar dentro de um genoma hospedeiro (CARARETO; MONTEIRO-VITORELLO; SLUYS, 2015). Pela capacidade desses elementos de se movimentar pelo genoma eles são tanto promotores de novidades evolutivas, quanto potencialmente deletérios. Eles podem tanto se inserir dentro da sequência codificante de um gene quanto, mais comumente, em regiões não-codificantes do genoma (BIÉMONT; VIEIRA, 2006; MCCULLERS; STEINIGER, 2017).

Devido à diversidade de TEs se fez necessária uma classificação. De forma geral ela dependerá da natureza do intermediário de transposição, das características estruturais do elemento e da homologia a elementos conhecidos (PRITHAM, 2009). A primeira foi de Finnegan (1989) e posteriormente teríamos uma complementação dessa classificação por Wicker et al. (2007). Entretanto, Kapitonov e Jurka (2008) argumentam terem proposto uma classificação complementar a de Finnegan (1989) antes da divulgada por Wicker et al. (2007) para os elementos de transposição, porém atualmente a classificação mais difundida na literatura é a de Wicker (2007).

Segundo Wicker et al. (2007) a classificação segue uma lógica hierárquica onde os TEs são divididos em Classe, Ordem (para os de Classe I) ou Subclasse (para os de Classe II), Superfamílias, Famílias e Subfamílias. Tanto a classificação de Finnegan (1989) quanto a de Wicker et al. (2007) baseiam-se em uma primeira

grande divisão dos TEs que se refere a natureza do intermediário de transposição. Os TEs de Classe I também chamados de retrotransposons ou retroelementos dependem de uma molécula intermediária de RNA para a transposição e os TEs de Classe II mobilizam-se por molécula intermediária de DNA. Cada Classe tem suas peculiaridades com relação as suas próximas divisões no que diz respeito aos aspectos estruturais dos elementos e as homologias.

Wicker et al. (2007) dividem os elementos de Classe II em duas Subclasses de acordo com o mecanismo de transposição, sendo a primeira transposons “corta-e-cola” e a segunda “copia-e-cola”. Os TEs de Subclasse I possuem nove Superfamílias que são diferenciadas baseando-se nas suas repetições terminais invertidas (TIRs - do inglês *Terminal Inverted Repeats*) que no tamanho da duplicação do sítio alvo (TSD - do inglês *Target Site Duplications*) (WICKER et al., 2007). A Subclasse II é dividida em duas Superfamílias, a *Helintron* que se replica por Replicação por círculo rolante sem gerar TSDs e a *Maverick* que tem como principais características um grande tamanho (variando de 10 a 20 mil pares de bases de DNA) e suas longas TIRs (WICKER et al., 2007).

Ainda segundo os autores os elementos Classe I podem ser divididos em cinco Ordens dependendo de como se arranjam os seus domínios internos, das relações filogenéticas existentes entre as sequências das transcriptases reversa e dos mecanismos de transposição. Em Superfamília os elementos que estão dentro de uma mesma Ordem compartilham estratégia de replicação, porém, diferem-se com relação a estrutura proteica, domínios não codificantes e a presença e ao tamanho de TSDs (WICKER et al., 2007). As Famílias são definidas pela conservação da sequência de DNA e as Subfamílias pelas informações de filogenia (WICKER et al., 2007).

Outra maneira de organizar os TEs é quanto a sua capacidade de se mobilizar autonomamente. Os elementos autônomos possuem toda a maquinaria enzimática para promover sua transposição independentemente, enquanto os não-autônomos não possuem as enzimas necessárias para mobilização, mas são reconhecidos pelos TEs autônomos (CARARETO; MONTEIRO-VITORELLO; SLUYS, 2015; WICKER et al., 2007).

### 1.3 Distribuição, classificação e caracterização de TEs da superfamília CR1

O elemento *Chicken reaper 1* (CR1) foi identificado e descrito inicialmente em *Gallus gallus* (galo doméstico), por E. Stumph et al. (1981). Atualmente é classificado como uma superfamília de retrotransposons autônomos, fazendo parte da Ordem LINE - do inglês *Long interspersed nuclear element* - pertencente a Classe I (RUGGIERO; BOURGEOIS; BOISSINOT, 2017). Atualmente foram identificados nos genomas de espécies de urocordados, cefalocordados, vermes achatados, insetos, anfíbios, peixes de nadadeiras raiadas e peixes pulmonados, répteis, mamíferos e outras aves (CHALOPIN et al., 2014; DREW; BRINDLEY, 1997; KAPUSTA; SUH, 2017; SUH et al., 2014).

Os elementos CR1 quando estão completos são compostos por duas fases abertas de leitura (ORF, do inglês *Open Reading Frames*) com seu tamanho total podendo chegar a 4.5Kb – Figura 1 – (BERTOCCHI, 2017; BURCH; DAVIS; HAAS, 1993). A primeira ORF e a porção 5'UTR desses elementos normalmente são truncadas (bem deterioradas) com a ORF codificando uma proteína *gag-like* com motivos *zincfingers-like* (HAAS et al., 1997, 2001; KAPITONOV; JURKA, 2003). A segunda ORF é, normalmente, conservada codificando as enzimas essenciais para a sua mobilização que são a Transcriptase Reversa – RT, do inglês *Reverse Transcriptase* – e uma Endonuclease apurinica (BURCH; DAVIS; HAAS, 1993; HAAS et al., 1997, 2001; KAPITONOV; JURKA, 2003). Possuem uma porção 3'UTR que supõe-se que sirva como região de reconhecimento para a RT codificada por elementos CR1 (SUH et al., 2014). Nessa região também contém motivos *hairpin* e microssatélite octâmero, os quais podem influenciar a estabilidade e a regulação da transcrição de regiões genômicas adjacentes (KAPUSTA; SUH, 2017).



Figura 1: Esquema gráfico do retroelemento CR1. Figura retirado de Kapusta; Suh (2017).

Os elementos CR1 tem sua distribuição estudada, analisada e caracterizada principalmente na Classe Aves. São os TEs mais abundante nessas espécies, chegando a representar aproximadamente 90% dos elementos nesses genomas, equivalente à 7-10% do genoma total (KAPUSTA; SUH, 2017; SUH et al., 2014; ZHANG et al., 2014). Estão presentes em toda a filogenia das Aves e encontram-se bastante degenerados, isto é, sofreram muitas mutações, deleções, ocasionado por estarem presentes nesses genomas há muito tempo (SUH et al., 2014). Na classe Aves são descritas pelo menos 14 subfamílias, todas elas diversificaram de um ancestral em comum de CR1 (KAPUSTA; SUH, 2017). Enquanto, outras seis linhagens antigas de CR1, presentes no ancestral pássaro/crocodylo, foram extintas (KAPUSTA; SUH, 2017). No genoma de répteis e mamíferos também foram identificadas diferentes subfamílias de CR1, contudo de diferentes linhagens (SUH et al., 2014). Em outros grupos de organismos, como peixes e insetos, as linhagens de CR1 presentes possuem as suas sequências a nível de nucleotídeo tão distintas que tornam difícil a análise comparativa com os amniotas - aves, répteis e mamíferos – (BERTOCCHI et al. *in prep.*, 2018; SUH et al., 2014). É uma superfamília de extrema importância nos estudos de TEs visto a sua distribuição e influência na história evolutiva dos genomas de diferentes grupos de organismos.

A subfamília denominada CR1\_E, foi inicialmente identificada no genoma de *Taeniopygia guttata* (Zebra Finch, ordem Passeriformes) (SUH et al., 2011). Posteriormente, também foi encontrada nos genomas de Piciformes, aproximadamente dez espécies de pica-paus, capitães e tucanos através de análises *in silico* (BERTOCCHI et al. *in prep.*, 2018; MANTHEY et al., 2018). Atualmente, a subfamília CR1\_E foi identificada *in situ* em quatro espécies de pica-paus brasileiros e também em *Gallus gallus* (Galo doméstico, Ordem Galliformes) (BERTOCCHI, 2017; BERTOCCHI et al. *in prep.*, 2018). Os estudos abordam principalmente espécies de aves Piciformes, pois esses genomas possuem a característica distinta das outras espécies de aves, ostentando aproximadamente 22% do genoma constituído por TEs CR1 (MANTHEY et al., 2018; ZHANG et al., 2014).

A subfamília CR1\_E caracteriza-se por possuir a região da RT e porção 3'UTR do elemento conservada (BERTOCCHI et al. *in prep.*, 2018). Contudo, a parte inicial

do elemento e a ORF1 são degeneradas ou inexistentes, portanto são elementos não-autônomos (BERTOCCHI et al. *in prep*, 2018; SUH et al., 2014). Embora, tenha aumentado gradativamente a identificação de TEs CR1\_E-like nos genomas de Piciformes, ainda não houve estudos buscando a distribuição desses elementos em outros grupos de aves e de diferentes organismos com genomas disponíveis.

## **2 JUSTIFICATIVA E OBJETIVOS**

Com o aumento no número de sequências disponíveis as pesquisas relacionadas à investigação dos genomas são continuamente necessárias e essenciais. Com a disponibilidade de ferramentas de bioinformática e computadores com grande capacidade de processamento, o estudo da composição dos genomas está cada vez mais avançado. Dada a influência dos TEs nos genomas hospedeiros, a distribuição ubíqua nos organismos e o aumento contínuo de genomas sequenciados disponíveis, os estudos sobre elementos transponíveis são de grande importância para compreender melhor a história evolutiva dos seres vivos.

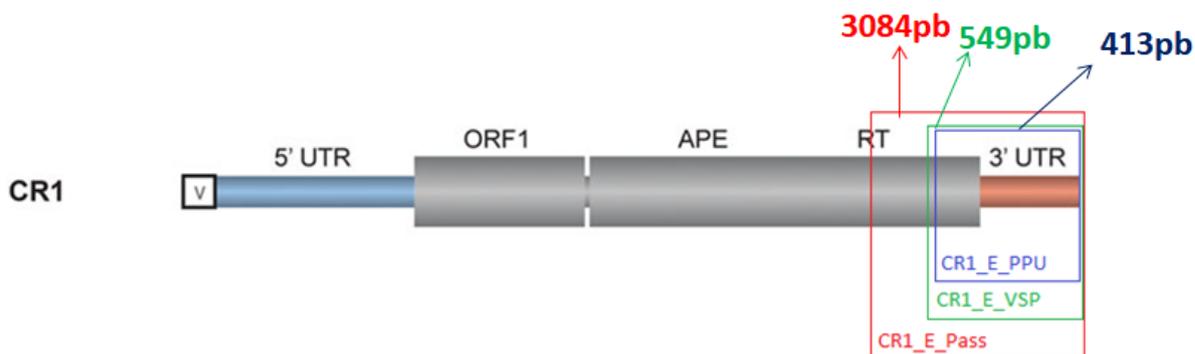
O foco deste trabalho foi aprofundar o conhecimento sobre os retroelementos de transposição CR1\_E. Para isso analisou-se a sua distribuição nos genomas sequenciados disponíveis e reconstrução da história filogenética desses TEs. As análises para elucidar as relações filogenéticas e busca por sequências homólogas foram realizadas *in silico* através de *softwares* e plataformas *online*. Esse trabalho, visou a produção de conhecimento científico para a área de pesquisa com elementos de transposição, da superfamília CR1, no âmbito da bioinformática e pode servir como base para outros projetos que se relacionam ao assunto.

## **3 METODOLOGIA**

### **3.1 Sequências de referência**

Três sequências foram utilizadas como referência para a realização de buscas de retrotransposons CR1\_E em diferentes organismos, todas identificadas inicialmente nos genomas de Aves. A sequência CR1\_E\_Pass de Passeriformes foi

retirada do banco de dados *Repbase* (BAO; KOJIMA; KOHANY, 2015), e as sequências CR1\_E\_PPU consenso de *Picoides pubescens* e a CR1\_E\_VSP de *Veniliornis spilogaster* foram retiradas de Bertocchi (2017). Na figura 2 pode-se verificar a região de abrangência dessas sondas em um típico elemento CR1 bem como seus tamanhos.



**Figura 2. Esquema geral dos elementos CR1.** Em destaque, as regiões das sequências utilizadas como sonda nesse trabalho e seus respectivos tamanhos. Em vermelho CR1\_E\_Pass com 3084pb, em verde CR1\_E\_VSP com 549pb e em azul CR1\_E\_PPU com 413pb. Figura retirada e adaptada de KAPUSTA & SUH (2017).

### 3.2 Buscas *in silico* em diferentes bancos de dados e alinhamentos

As buscas por sequências homólogas foram realizadas nos bancos de dados *Genbank* e *Flybase* com as três sequências sondas (CLARK et al., 2016; GRAMATES et al., 2017).

Os parâmetros utilizados para a seleção das sequências homólogas foram o *Expected value (E-value)* e o tamanho aproximado da sonda. Para esse trabalho o valor de corte estabelecido de *E-value* foi menor que  $e^{-10}$ . A exclusão manual foi realizada quando as sequências homólogas representavam tamanho inferior a aproximadamente 10% da sequência sonda.

As buscas no *Genbank* foram realizadas através do BLASTn usando as configurações padrões, em outubro de 2017 (ALTSCHUL et al., 1990; CLARK et al., 2016). Inicialmente o BLASTn foi feito sem especificação de procura em grupos de organismos, porém os resultados mostrados se limitaram aos melhores escores - que nesse caso eram os genomas de Aves - omitindo possíveis resultados em

outros grupos. Portanto, com o objetivo de se ter abrangência entre diferentes grupos de organismos, para a realização do BLASTn, utilizou-se os filtros disponíveis para selecionar individualmente grandes grupos de organismos - 14 grupos distintos, sendo eles: *Bacteria*, *Archae*, *Reptilia*, *Mammalia*, *Amphibia*, *Aves*, *Bony fishes*, *Cartilaginous fish*, *Flatworms*, *Roundworms*, *Arthropoda*, *Fungi*, *Plants* e *Viruses*.

A busca no banco de dados *Flybase* através de BLASTn foi realizada em Março de 2018 e contra todas as espécies de insetos presentes nesse banco de dados (Ordens Diptera, Lepidoptera, Coleoptera, Hymenoptera, Hemiptera, Phthiraptera e Ixodida) (ALTSCHUL et al., 1990; GRAMATES et al., 2017).

As sequências disponíveis no banco de dados *Repbase* classificadas como CR1-like foram acessadas em março de 2018 para serem alinhadas também às sequências de referência (BAO; KOJIMA; KOHANY, 2015).

Os alinhamentos foram efetuados pelo programa *online* MAFFT v.7 (do inglês *Multiple alignment for amino acid or nucleotide sequences*) com os parâmetros *default* (KATOH; STANDLEY, 2013). Contudo, com as sequências oriundas do banco de dados *Genbank* utilizou-se o parâmetro “*Adjust direction according to the first sequence*” (CLARK et al., 2016). Os alinhamentos foram visualizados, avaliados e refinados manualmente através do *software* Aliview (LARSSON, 2014).

### **3.3 Filogenias e matrizes de distância**

Todas as árvores filogenéticas foram construídas usando o *software* MEGA versão 7.0.26 através do teste *Maximum likelihood*, utilizando as configurações padrão (KUMAR; STECHER; TAMURA, 2016).

Analisou-se, a conservação nucleotídica das sequências homólogas em relação a cada elemento sonda, através da análise da identidade das mesmas. Para isso, utilizou-se a matriz de distância com o *software* Unipro UGENE versão 1.29.0 (OKONECHNIKOV et al., 2012).

## 4 RESULTADOS

### 4.1 Sequências homólogas

Identificamos sequências homólogas a sonda CR1\_E\_Pass em 31 espécies, nos grupos de aves, répteis, bactéria, mamíferos e peixes cartilagosos. Contudo, após refinamento manual, devido à baixa cobertura em relação a sequência sonda, foram excluídas as espécies representantes dos grupos de mamíferos, peixes cartilagosos e duas espécies de répteis. Após a edição, o número final para essa sonda foi de 27 espécies (Tabela 1).

As buscas com a sonda CR1\_E\_PPU encontrou sequências homólogas nos grupos aves, répteis e bactéria, totalizando 27 espécies (Tabela 2). E com a sonda CR1\_E\_VSP observou-se sequências homólogas nos grupos de aves, répteis e bactéria, abrangendo 16 espécies (Tabela 3).

Ao total, foram encontradas 35 espécies diferentes com as três sondas, compreendendo aves, répteis e bactéria (*Escherichia coli*). Desse total, 12 espécies são homólogas para as 3 sondas (Tabela 4). Além disso, observamos espécies exclusivas para cada sonda, 6 para CR1\_E\_Pass, 5 para CR1\_E\_PPU e 1 para CR1\_E\_VSP, sendo elas espécies de Aves para as duas primeiras e 1 espécie de réptil para a última (Tabela 4).

Encontrou-se sequências homólogas a CR1\_E em 28 espécies e em 13 ordens de aves, que não haviam sido registrados anteriormente, aumentando então o registro desse retrotransposon dentro da Classe Aves. Além disso, também encontramos a presença de sequências homólogas em 4 espécies de lagartos (Ordem Squamata) e também em uma espécie de procaríoto.

Não foram encontrados resultados significativos ( $E\text{-value} < e^{-10}$ ) no banco de dados *Flybase*.

### 4.2 Conservação nucleotídica das sequências homólogas

Foram geradas matrizes de identidade para aos três elementos usados como sonda e suas respectivas sequências homólogas, com o intuito de analisar a conservação das sequências dos TEs nos diferentes genomas.

Encontramos para o elemento CR1\_E\_Pass, no grupo das aves, maior conservação nucleotídica na espécie *Aptenodytes patagonicus* (Ordem Sphenisciformes), com 87% de identidade e também, menor conservação em *Lepidothrix coronata* (Ordem Passeriformes), com apenas 7% de identidade (Figura 3A). As outras 25 espécies abrangeram diferentes níveis de conservação de suas sequências entre 17% e 85% de identidade (Figura 3A).

Com o elemento CR1\_E\_PPU, observamos também no grupo das aves a maior conservação da sequência, contudo em duas espécies diferentes, a *Lophotibis cristata* (Ordem Pelecaniformes) e *Aptenodytes patagonicus* (Ordem Sphenisciformes) com 79% de identidade (Figura 3B). Entretanto, a menor conservação foi no grupo dos répteis com a espécie *Iberolacerta cyreni* (Ordem Squamata), com 18% de identidade (Figura 3B). As outras 24 espécies abrangeram diferentes níveis de conservação entre 23% e 78% identidades (Figura 3B).

Identificamos nas sequências homólogas de CR1\_E\_VSP maior conservação da sequência nas espécies de aves *Ciconia boyciana* (Ordem Pelecaniformes) e *Gallus gallus* (Ordem Galliformes), com 79% de identidade (Figura 3C). E a sequência menos conservada, com 21% de identidade, foi da espécie de réptil *I. cyreni* (Ordem Squamata) (Figura 3C). As sequências do TE presentes nas outras 13 espécies abrangeram diferentes níveis identidades entre 29% e 78% (Figura 3C).

### 4.3 Análise comparativa das sequências homólogas

Na figura 4, observamos a filogenia do elemento CR1\_E\_Pass e suas sequências homólogas, onde há a formação de cinco grupos distintos: três deles abrangendo apenas aves (quadros azul, rosa e branco); um com espécies de aves e *E. coli* (quadro amarelo); e um com as três espécies de répteis (quadro verde). O elemento, CR1\_E\_Pass, agrupou-se próximo as sequências das espécies *Corvus brachyrhynchos* e *Serinus canaria*, ambos da Ordem Passeriformes, formando um único grupo (Figura 4 - quadro rosa).

A árvore filogenética do elemento CR1\_E\_PPU e suas sequências homólogas (Figura 5), dividiram-se em três grupos distintos, sendo um grande grupo com as sequências de espécies de aves e *E. coli* (quadro rosa), outro com apenas uma espécie de ave (quadro branco) e um com as três espécies de répteis (quadro

verde). O elemento CR1\_E\_PPU agrupou-se próximo as sequências das espécies *Apaloderma vittatum* e *P. pubescens*, das Ordens Trogoniformes e Piciformes, respectivamente (Figura 5).

Na filogenia com o elemento CR1\_E\_VSP e suas sequências homólogas (Figura 6), observou-se que três grupos distintos formaram-se, sendo um grande grupo com as sequências de espécies de aves e *E. coli* (quadro rosa), um grupo com apenas uma espécie de ave (quadro branco) e um com as quatro espécies de répteis (quadro verde). O TE CR1\_E\_VSP agrupou-se próximo a sequência da espécie *P. pubescens* da Ordem Piciformes (Figura 6- quadro rosa).

As espécies de aves da Ordem Psittaciformes, *Ara ararauna*, *Nestor notabilis* e *Chalcopsitta cardinalis*, agruparam-se com a bactéria *E. coli*, em todas as filogenias.

#### 4.4 Relações filogenéticas entre elementos CR1

Encontra-se disponível no *Repbase*, 839 sequências de retrotransposons CR1-*like*, destas selecionamos, após refinamento manual, 166 sequências (BAO; KOJIMA; KOHANY, 2015). Essas sequências eram de espécies de variados grupos de organismos como peixes, anfíbios, répteis, aves e mamíferos. Para estabelecer as relações entre os elementos CR1 uma árvore filogenética foi construída abrangendo a parte final da RT do elemento e a porção 3'UTR a nível de nucleotídeo, em razão desta região ser altamente conservada em TEs CR1.

Na árvore filogenética gerada podemos observar que as sequências de CR1 estão divididas em três grandes grupos (A, B e C), e os grupos B e C estão subdivididos em dois braços cada (Figura 7). No grupo A, estão as espécies de répteis das ordens Crocodylia (crocodilos e jacaré) e Testudines (tartaruga); Anura (rãs); Coelacanthiformes (peixes ósseos) e Chimaeriformes (peixes cartilagosos); e Monotremata (ornitorrinco) Diprotodontia (canguru) (Figura 7). O ramo B envolve somente espécies de répteis da ordem Crocodylia (crocodilos, jacarés e gaviais) (Figura 7). E por último, o grupo C, possui espécies de répteis (tartarugas e crocodilos); mamíferos (humano, canguru e ornitorrinco); e aves (Figura 7).

O grupo C está subdividido em dois braços, organizado de maneira que o maior braço contém todas as espécies de aves, incluindo as nossas sequências

(Figura 7 – quadro azul). Os elementos utilizados como sonda, CR1\_E\_Pass, CR1\_E\_PPU e CR1\_E\_VSP, agruparam-se com os elementos CR1\_AVI e CR11\_MUn, das espécies de *Gallus gallus* (ordem Galliformes) e *Melopsittacus undulatus* (ordem Psittaciformes), respectivamente.

## 5 DISCUSSÃO

Como demonstrado em nossos resultados, a investigação sobre a distribuição de TEs da subfamília CR1\_E vai além das espécies e ordens de aves onde as três sequências sondas foram identificadas inicialmente. Nas aves identificamos elementos CR1\_E desde espécies basais, Ordem Apterygiformes, até a espécies mais derivadas (Ordem Passeriformes) (Tabela 4). Com isso, podemos inferir a presença desses elementos ao longo de toda a filogenia da Classe Aves, ocasionado possivelmente pela inserção dos TEs CR1\_E no último ancestral em comum da Classe (JARVIS et al., 2014; SUH et al., 2011; ZHANG et al., 2014).

Nos répteis, a presença de elementos CR1\_E foi restrita aos lagartos (Ordem Squamata) e inédita para as quatro espécies *Anolis carolinensis*, *Cyclura pinguis*, *Iberolacerta cyreni* e *Sceloporus undulatus* (Tabela 4). Contudo, somente no genoma de *A. carolinensis* foram registrados outros retrotransposons CR1 (RUGGIERO; BOURGEOIS; BOISSINOT, 2017; SOTERO-CAIO et al., 2017). A distribuição restrita de CR1 nos genomas de lagartos e a baixa conservação dessas sequências (Figura 3), pode estar associada a seleção negativa que esses retrotransposons estão sofrendo (RUGGIERO; BOURGEOIS; BOISSINOT, 2017; SOTERO-CAIO et al., 2017).

As sequências de CR1\_E nos genomas de Squamatas formaram um grupo isolado das outras sequências de Aves e *E. coli* (Figuras 4, 5 e 6). Além disso, não foi observado TE CR1\_E, nas outras ordens de répteis como, crocodilos e tartarugas, porém estão presentes em toda a Classe Aves (JARVIS et al., 2014; PINCHEIRA-DONOSO et al., 2013). Com isso, podemos inferir que os elementos CR1\_E foram inseridos nos genomas anteriormente a diversificação dos répteis e das Aves, porém evoluiu distintamente nos diferentes grupos de répteis,

possivelmente tendo sido perdido ou estar tão degenerado que torna difícil a identificação nos genomas de alguns grupos.

Inesperadamente, identificamos sequências homólogas aos elementos CR1\_E na espécie de bactéria *E. coli*, com a conservação das sequências chegando a, aproximadamente, 80% de identidade (Figura 3). Embora os genomas bacterianos possuam TEs, são as chamadas sequências de inserção (SI) (CARARETO; MONTEIRO-VITORELLO; SLUYS, 2015). As SI são estruturalmente similares aos outros TEs de DNA, da classe II, com TIRS e somente a enzima transposase para mobilização, ou seja, elementos distintos dos CR1 (CARARETO; MONTEIRO-VITORELLO; SLUYS, 2015). Por isso, foi inesperado identificar um retrotransposon em um genoma bacteriano.

E quando observada as relações filogenéticas das sequências CR1\_E de *E. coli* com as das outras espécies, evidenciou-se padrão de agrupamento da bactéria com as espécies da Ordem Psittaciformes (Figuras 4, 5 e 6). A alta conservação das sequências e relações próximas com os CR1 de Psittaciformes podem sugerir um possível evento de HTT, entre aves dessa ordem e o genoma procarioto. Na Classe Aves encontra-se descrito um evento de HTT entre duas espécies de aves com TE de DNA (Classe II) (BERTOCCHI et al., 2017). Com retrotransposons, existe descrito cinco eventos de HTT envolvendo o elemento AviRTE, todos eles entre vermes parasitários de humanos e diferentes linhagens de aves (SUH et al., 2016). Com elementos CR1 existe descrito na literatura eventos de HTT apenas entre espécies de insetos (DOTTO et al., 2017) portanto, ainda não foram descritos eventos de HTT entre retrotransposon CR1 de ave e um organismo procarioto.

A reconstrução da história filogenética dos elementos CR1 mostrou a distribuição desses retrotransposons em variados grupos de organismos, desde peixes a mamíferos (Figura 7). Porém, a proliferação desses retrotransposons ocorreu nos genomas de amniotas. Visto que o grupo dos répteis foram representados por espécies das quatro Ordens que o compõe e a Classe Aves e dos mamíferos tiveram representantes desde espécies basais (ratitas e ornitorrinco) às mais derivadas (passarinhos e humano) (Figura 7) (GREEN et al., 2014; ZHANG et al., 2014). Considerando as relações filogenéticas entre os amniotas (BRADLEY SHAFFER et al., 2013; GREEN et al., 2014; SHEDLOCK et al., 2007) e assumindo a

herança do TE CR1 através de transmissão vertical podemos assumir que as linhagens de CR1 estão presentes no ancestral em comum dos amniotas, como sugerido por Suh et al. (2014).

## **6 CONSIDERAÇÕES FINAIS**

Foi possível com esse estudo aprofundar o conhecimento acerca dos retroelementos de transposição CR1\_E. Isso se evidencia pelos achados inéditos da distribuição dos três elementos em outras 14 Ordens de aves, em 1 Ordem de répteis e em bactéria. A análise da distribuição dos três elementos também mostrou uma distribuição contínua do TE ao longo de toda a filogenia das aves e uma descontinuidade no grupo dos répteis. Além disso, foi detectado para esses retroelementos um possível evento de HTT em *E. coli*, sendo que não há registro na literatura desse tipo de transferência envolvendo bactérias e aves. Contudo, lembrando que os genomas bacterianos são altamente dinâmicos e tendem a manter somente as sequências necessárias para sobrevivência, fica a incógnita de porquê essa espécie procariótica, *E. coli*, introduziu no seu genoma um retrotranspon. Ainda, com a grande representação de amniotos na filogenia para CR1, foi possível observar uma proliferação desses retrotransposons para esse grupo, corroborando a hipótese de que as linhagens desses elementos se inseriram no ancestral comum desses organismos (SUH et al., 2014).

Como perspectivas futuras com relação aos principais achados desse trabalho, são analisar o número de cópias, características funcionais e estruturais desses três elementos nas espécies onde foram detectados. Além disso, a conservação das sequências desses retrotransposons também pode ser verificada. E também, análises mais aprofundadas quanto a real distribuição desses elementos nos genomas de répteis. Com isso pode-se realizar uma análise mais detalhada com relação a sua distribuição, proliferação e evolução.

Vale ressaltar que o número de genomas armazenados em banco de dados aumenta continuamente. Possibilitando ampliar o conhecimento sobre a distribuição e relações filogenéticas desses elementos nos genomas hospedeiros.

**Tabela 1.** Espécies com sequências homólogas ao elemento CR1\_E\_Pass.

<b>Grupo</b>	<b>Espécie</b>	<b>E-value</b>	<b>Código NCBI</b>	<b>Tamanho do genoma no NCBI (em Mb)</b>
Aves	<i>Anas platyrhynchos</i>	0	XR_002399042.1	1.105,05
Aves	<i>Anser cygnoides</i>	0	XM_013193510.1	1.119,15
Aves	<i>Aptenodytes forsteri</i>	0	XM_019473346.1	1.254,35
Aves	<i>Aptenodytes patagonicus</i>	0	LC072333.1	Sem dados
Aves	<i>Ara ararauna</i>	0	JN583888.1	Sem dados
Aves	<i>Chalcopsitta cardinalis</i>	0	EU738434.1	Sem dados
Aves	<i>Chlamydotis macqueenii</i>	0	XM_010127510.1	1.086,57
Aves	<i>Ciconia boyciana</i>	0	LC072369.1	1.364,95
Aves	<i>Columba livia</i>	0	XR_002421707.1	1.107,99
Aves	<i>Corvus brachyrhynchos</i>	0	XR_001862582.1	1.091,31
Aves	<i>Egretta garzetta</i>	0	LC072327.1	1.206,50
Aves	<i>Gallus gallus</i>	0	AC243550.3	1.230,26
Aves	<i>Gymnogyps californianus</i>	0	AC183449.3	Sem dados
Aves	<i>Haliaeetus leucocephalus</i>	0	XM_010574287.1	1.178,41
Aves	<i>Lepidothrix coronata</i>	0	XR_001878440.1	1.079,58
Aves	<i>Lophotibis cristata</i>	0	LC072485.1	Sem dados
Aves	<i>Morus capensis</i>	0	LC072486.1	Sem dados
Aves	<i>Nestor notabilis</i>	0	BK009169.1	1.053,56
Aves	<i>Nipponia nippon</i>	0	KP182409.1	1.223,86
Aves	<i>Opisthocomus hoazin</i>	0	XM_009934347.1	1.203,71
Aves	<i>Pelecanus occidentalis</i>	0	LC072483.1	Sem dados
Aves	<i>Procellaria aequinoctialis</i>	0	LC072332.1	Sem dados
Aves	<i>Serinus canaria</i>	0	XM_009093841.2	1.152,10
Bactéria	<i>Escherichia coli</i>	0	LN877770.1	5.3582
Répteis	<i>Anolis carolinensis</i>	4e <sup>-63</sup>	FJ158987.1	1.799,14
Répteis	<i>Iberolacerta cyreni</i>	1e <sup>-24</sup>	DQ435780.1	Sem dados
Répteis	<i>Sceloporus undulatus</i>	2e <sup>-15</sup>	DQ784617.1	Sem dados

**Tabela 2.** Espécies com sequências homólogas ao elemento CR1\_E\_PPU.

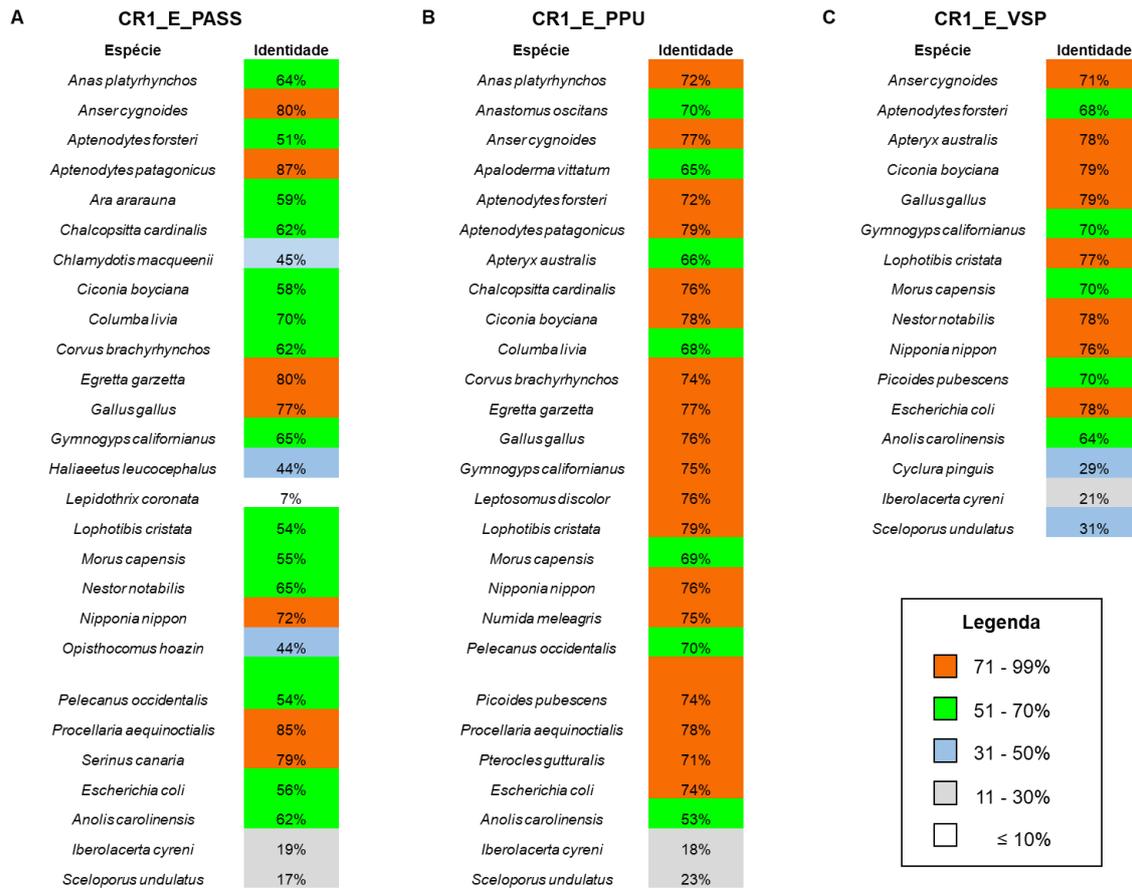
<b>Grupo</b>	<b>Espécie</b>	<b>E-value</b>	<b>Código NCBI</b>	<b>Tamanho do genoma no NCBI (em Mb)</b>
Aves	<i>Anas platyrhynchos</i>	1e <sup>-67</sup>	XM_021273399.1	1.105,05
Aves	<i>Anastomus oscitans</i>	6e <sup>-93</sup>	LC072487.1	Sem dados
Aves	<i>Anser cygnoides</i>	6e <sup>-71</sup>	XR_001212670.1	1.119,15
Aves	<i>Apaloderma vittatum</i>	6e <sup>-80</sup>	XM_009876757.1	1.070,84
Aves	<i>Aptenodytes forsteri</i>	2e <sup>-91</sup>	XM_019473346.1	1.254,35
Aves	<i>Aptenodytes patagonicus</i>	6e <sup>-99</sup>	LC072333.1	Sem dados
Aves	<i>Apteryx australis</i>	2e <sup>-70</sup>	LK064922.1	1523,97
Aves	<i>Chalcopsitta cardinalis</i>	2e <sup>-79</sup>	EU738434.1	Sem dados
Aves	<i>Ciconia boyciana</i>	4e <sup>-93</sup>	LC072369.1	1.364,95
Aves	<i>Columba livia</i>	5e <sup>-72</sup>	XM_021286190.1	1.107,99
Aves	<i>Corvus brachyrhynchos</i>	1e <sup>-68</sup>	XR_001863814.1	1.091,31
Aves	<i>Egretta garzetta</i>	7e <sup>-92</sup>	LC072548.1	1.206,50
Aves	<i>Gallus gallus</i>	2e <sup>-84</sup>	AC186867.3	1.230,26
Aves	<i>Gymnogyps californianus</i>	3e <sup>-89</sup>	AC183449.3	Sem dados
Aves	<i>Leptosomus discolor</i>	4e <sup>-82</sup>	XM_009959315.1	1.136,24
Aves	<i>Lophotibis cristata</i>	7e <sup>-105</sup>	LC072366.1	Sem dados
Aves	<i>Morus capensis</i>	2e <sup>-80</sup>	LC072426.1	Sem dados
Aves	<i>Nipponia nippon</i>	9e <sup>-93</sup>	KP182409.1	1.223,86
Aves	<i>Numida meleagris</i>	8e <sup>-79</sup>	XM_021390231.1	1.043,26
Aves	<i>Pelecanus occidentalis</i>	3e <sup>-83</sup>	LC072351.1	Sem dados
Aves	<i>Picoides pubescens</i>	5e <sup>-102</sup>	XM_009899795.1	1.167,32
Aves	<i>Procellaria aequinoctialis</i>	2e <sup>-91</sup>	LC072332.1	Sem dados
Aves	<i>Pterocles gutturalis</i>	4e <sup>-82</sup>	XM_010084581.1	1.069,32
Bacteria	<i>Escherichia coli</i>	1e <sup>-59</sup>	LN877770.1	5.3582
Répteis	<i>Anolis carolinensis</i>	1e <sup>-13</sup>	FJ158987.1	1.799,14
Répteis	<i>Iberolacerta cyreni</i>	9e <sup>-10</sup>	DQ435780.1	Sem dados
Répteis	<i>Sceloporus undulatus</i>	3e <sup>-10</sup>	DQ784632.1	Sem dados

**Tabela 3.** Espécies com sequências homólogas ao elemento CR1\_E\_VSP.

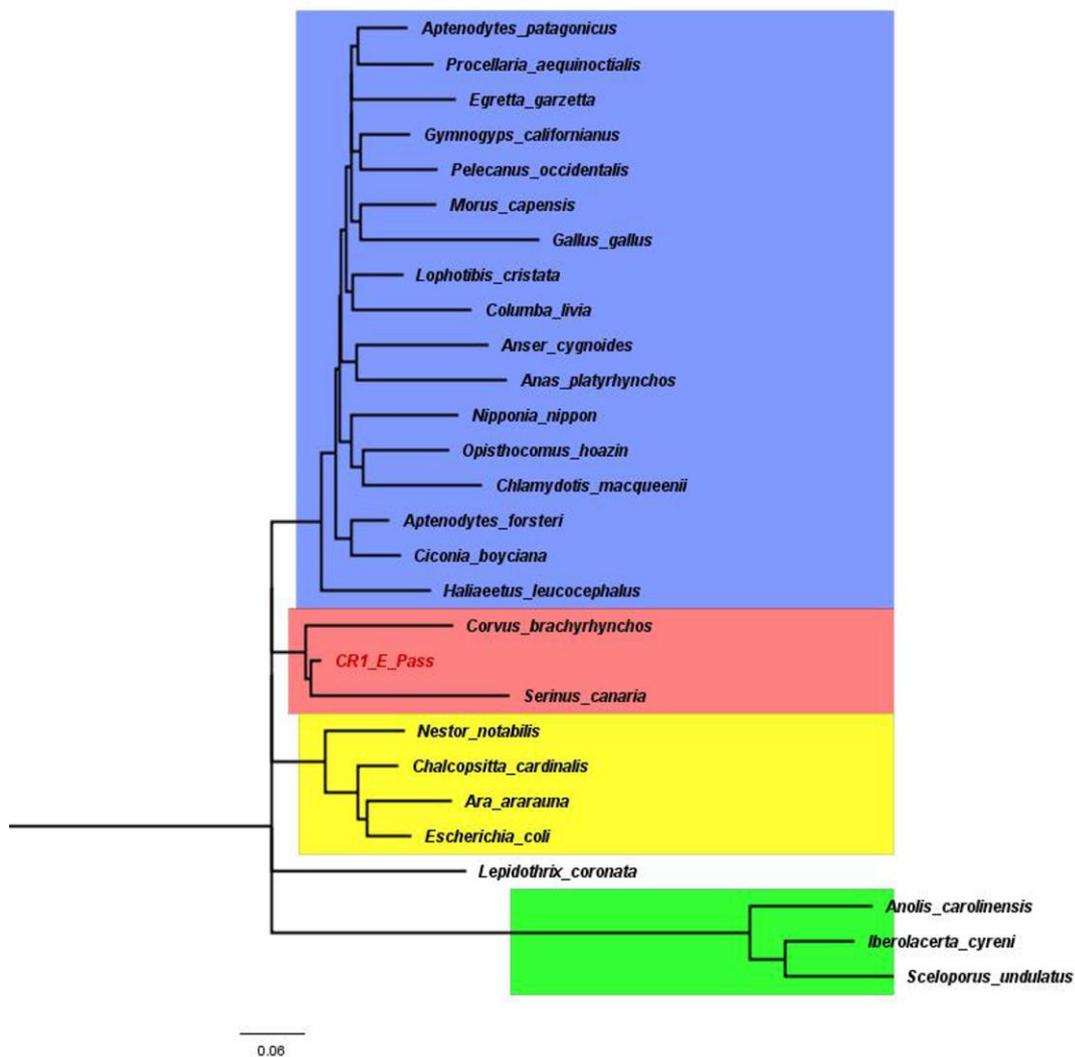
Grupo	Espécie	E-value	Código NCBI	Tamanho do genoma no NCBI (em Mb)
Aves	<i>Anser cygnoides</i>	5e <sup>-88</sup>	XR_001212670.1	1.119,15
Aves	<i>Aptenodytes forsteri</i>	6e <sup>-107</sup>	XM_019473346.1	1.254,35
Aves	<i>Apteryx australis</i>	6e <sup>-87</sup>	LK064922.1	1.523,97
Aves	<i>Ciconia boyciana</i>	2e <sup>-106</sup>	LC072464.1	1.364,95
Aves	<i>Gallus gallus</i>	3e <sup>-94</sup>	AC270333.1	1.230,26
Aves	<i>Gymnogyps californianus</i>	5e <sup>-108</sup>	AC183449.3	Sem dados
Aves	<i>Lophotibis cristata</i>	8e <sup>-116</sup>	LC072561.1	Sem dados
Aves	<i>Morus capensis</i>	1e <sup>-88</sup>	LC072486.1	Sem dados
Aves	<i>Nestor notabilis</i>	2e <sup>-86</sup>	BK009169.1	1.053,56
Aves	<i>Nipponia nippon</i>	1e <sup>-95</sup>	KP182409.1	1.223,86
Aves	<i>Picooides pubescens</i>	3e <sup>-124</sup>	XM_009899795.1	1.167,32
Bactéria	<i>Escherichia coli</i>	5e <sup>-80</sup>	LN877770.1	5.3582
Répteis	<i>Anolis carolinensis</i>	1e <sup>-27</sup>	FJ158987.1	1.799,14
Répteis	<i>Cyclura pinguis</i>	1e <sup>-21</sup>	FJ555497.1	Sem dados
Répteis	<i>Iberolacerta cyreni</i>	9e <sup>-18</sup>	DQ435780.1	Sem dados
Répteis	<i>Sceloporus undulatus</i>	9e <sup>-18</sup>	DQ784617.1	Sem dados

**Tabela 4.** Ausência/presença de sequências de espécies homólogas com relação às sondas.

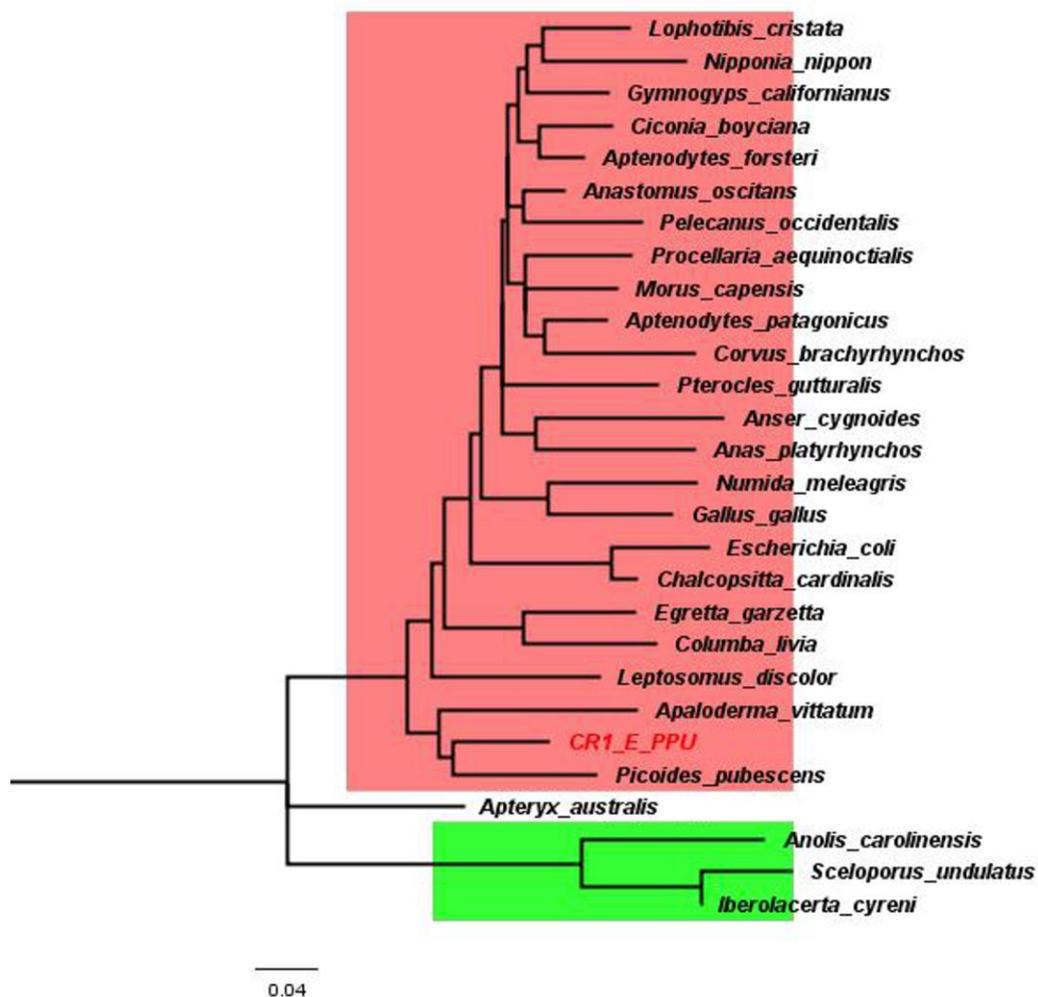
<b>Grupo</b>	<b>Ordem</b>	<b>Espécies</b>	<b>CR1_ E_ Pass</b>	<b>CR1_ E_ PPU</b>	<b>CR1_ E_ VSP</b>
Aves	Accipitriformes	<i>Gymnogyps californianus</i>	+	+	+
Aves	Accipitriformes	<i>Haliaeetus leucocephalus</i>	+	-	-
Aves	Anseriformes	<i>Anas platyrhynchos</i>	+	+	-
Aves	Anseriformes	<i>Anser cygnoides</i>	+	+	+
Aves	Apterygiformes	<i>Apteryx australis</i>	-	+	+
Aves	Columbiformes	<i>Columba livia</i>	+	+	-
Aves	Galliformes	<i>Gallus gallus</i>	+	+	+
Aves	Galliformes	<i>Numida meleagris</i>	-	+	-
Aves	Leptosomiformes	<i>Leptosomus discolor</i>	-	+	-
Aves	Opisthocomiformes	<i>Opisthocomus hoazin</i>	+	-	-
Aves	Otidiformes	<i>Chlamydotis macqueenii</i>	+	-	-
Aves	Passeriformes	<i>Corvus brachyrhynchos</i>	+	+	-
Aves	Passeriformes	<i>Lepidothrix coronata</i>	+	-	-
Aves	Passeriformes	<i>Serinus canaria</i>	+	-	-
Aves	Pelecaniformes	<i>Anastomus oscitans</i>	-	+	-
Aves	Pelecaniformes	<i>Ciconia boyciana</i>	+	+	+
Aves	Pelecaniformes	<i>Egretta garzetta</i>	+	+	-
Aves	Pelecaniformes	<i>Lophotibis cristata</i>	+	+	+
Aves	Pelecaniformes	<i>Morus capensis</i>	+	+	+
Aves	Pelecaniformes	<i>Nipponia nippon</i>	+	+	+
Aves	Pelecaniformes	<i>Pelecanus occidentalis</i>	+	+	-
Aves	Piciformes	<i>Picoides pubescens</i>	-	+	+
Aves	Procellariiformes	<i>Procellaria aequinoctialis</i>	+	+	-
Aves	Psittaciformes	<i>Ara ararauna</i>	+	-	-
Aves	Psittaciformes	<i>Chalcopsitta cardinalis</i>	+	+	-
Aves	Psittaciformes	<i>Nestor notabilis</i>	+	-	+
Aves	Pteroclidiformes	<i>Pterocles gutturalis</i>	-	+	-
Aves	Sphenisciformes	<i>Aptenodytes forsteri</i>	+	+	+
Aves	Sphenisciformes	<i>Aptenodytes patagonicus</i>	+	+	-
Aves	Trogoniformes	<i>Apaloderma vittatum</i>	-	+	-
Bactéria	Enterobacteriales	<i>Escherichia coli</i>	+	+	+
Répteis	Squamata	<i>Anolis carolinensis</i>	+	+	+
Répteis	Squamata	<i>Cyclura pinguis</i>	-	-	+
Répteis	Squamata	<i>Iberolacerta cyreni</i>	+	+	+
Répteis	Squamata	<i>Sceloporus undulatus</i>	+	+	+



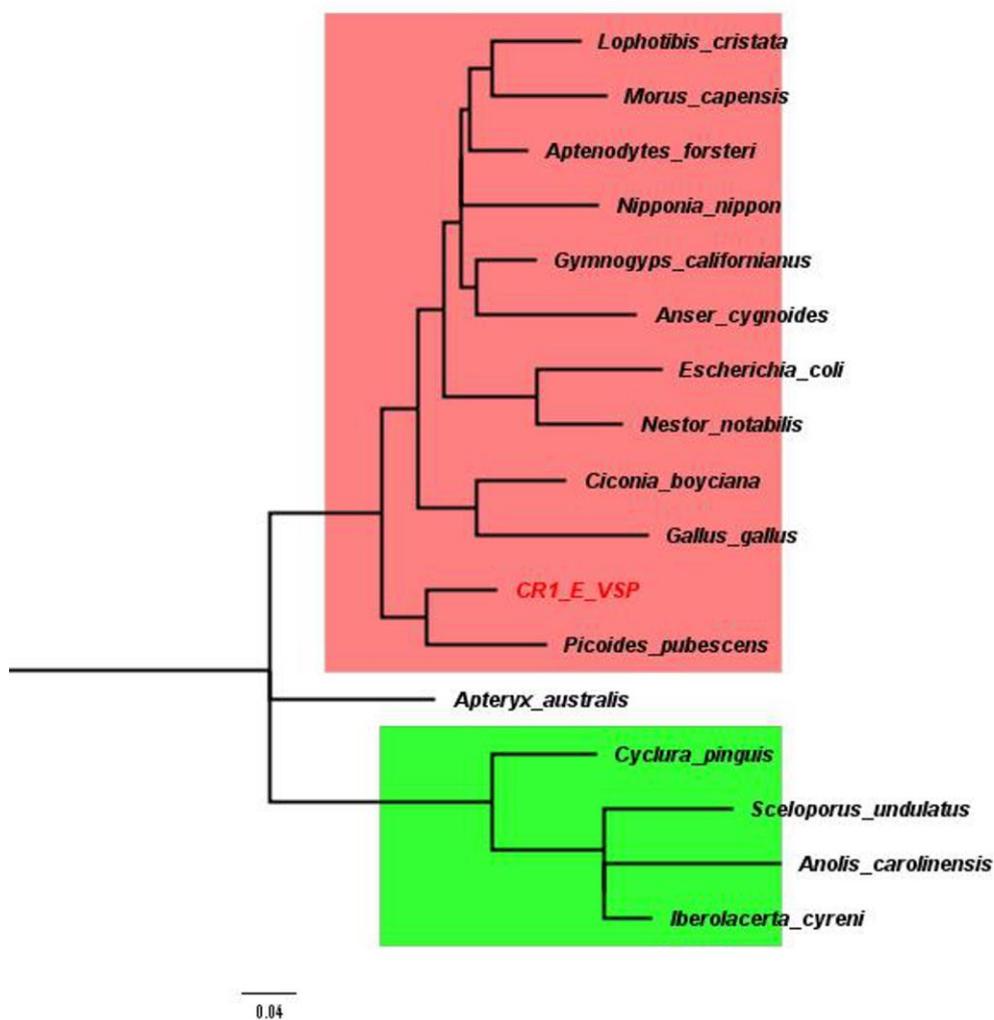
**Figura 3. Conservação das seqüências nucleotídicas.** Matriz de identidade para as sondas CR1\_E\_Pass (A), CR1\_E\_PPU (B) e CR1\_E\_VSP (C) em relação as suas seqüências homólogas de diferentes espécies. A matriz de distância foi calculada por um algoritmo de distância de similaridade através do software Unipro UGENE versão 1.29.0 (OKONECHNIKOV et al., 2012).



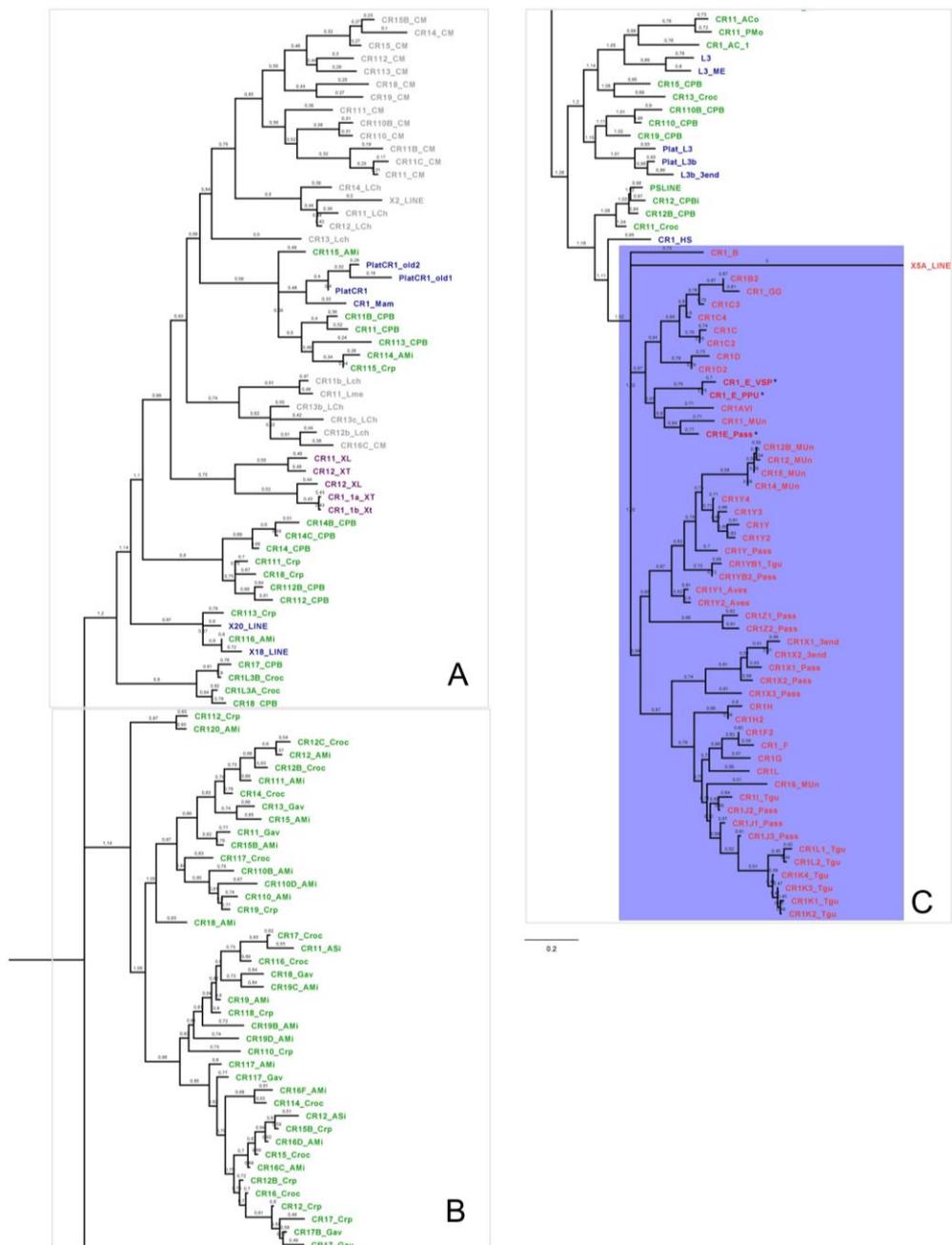
**Figura 4. Filogenia de CR1\_E\_Pass e suas seqüências homólogas.** A filogenia foi construída usando o software MEGA versão 7.0.26 através do teste *Maximum likelihood*, utilizando as configurações padrão (KUMAR; STECHER; TAMURA, 2016). A filogenia foi construída com a região da RT e porção 3'UTR das seqüências. Os quadros azul, rosa, amarelo e verde não marcados são compostos por de aves e *E. coli*. O quadro verde é composto somente por seqüências de répteis. O elemento CR1\_E\_Pass agrupou-se com as espécies *Serinus canaria* e *Corvus brachyrhynchos* (quadro rosa). O elemento CR1\_E\_Pass está identificado em vermelho.



**Figura 5. Filogenia de CR1\_E\_PPU e suas seqüências homólogas.** A filogenia foi construída usando o software MEGA versão 7.0.26 através do teste *Maximum likelihood*, utilizando as configurações padrão (KUMAR; STECHER; TAMURA, 2016). A filogenia foi construída com a região da RT e porção 3'UTR das seqüências. O quadro rosa e a seqüência não marcada são constituídos de aves e *E. coli*. O quadro verde possui somente seqüências de répteis. Dentro do quadro rosa a sonda se agrupou próxima de *Apaloderma vittatum* e *Picoides pubescens*. A sonda CR1\_E\_PPU está identificada em vermelho.



**Figura 6. Filogenia de CR1\_E\_VSP e suas sequências homólogas.** A filogenia foi construída usando o software MEGA versão 7.0.26 através do teste *Maximum likelihood*, utilizando as configurações padrão (KUMAR; STECHER; TAMURA, 2016). A filogenia foi construída com a região da RT e porção 3'UTR das sequências. O quadro rosa e a sequência não marcada são constituídos de aves e *E. coli*. O quadro verde possui somente sequências de répteis. Dentro do quadro rosa a sonda se agrupou principalmente com *Picoides pubescens*. A sonda CR1\_E\_VSP está identificada em vermelho.



**Figura 7. Árvore guia de alinhamento entre os elementos CR1.** A filogenia foi construída usando o software MEGA versão 7.0.26 através do teste *Maximum likelihood*, utilizando as configurações padrão (KUMAR; STECHER; TAMURA, 2016). A filogenia foi construída com a região da RT e porção 3'UTR das sequências. Os "\*" indicam os TEs CR1\_E sonda utilizados nesse trabalho. Todas as outras sequências CR1 presentes foram retiradas do Repbase e estão representadas pelos seus identificadores do banco de dados. As sequências estão indicadas em cores diferentes dependendo de se sua origem são peixes (cinza), anfíbios (roxo), verde (répteis), mamíferos (azul) e aves (vermelho). Os elementos CR1 estão divididos em 3 grupos principais como indicados pelas letras "A", "B" e "C" sendo que cada grupo se subdivide em vários ramos.

## REFERÊNCIAS

ALLEN, G. The Origin and Development of Molecular Biology. In: Life Science in the Twentieth Century. [s.l: s.n.]. p. 187–228. 1978.

ALTSCHUL, S. F. et al. Basic local alignment search tool. Journal of molecular biology, [s. l.], v. 215, n. 3, p. 403–10, 1990. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0022283605803602>>

BAO, W.; KOJIMA, K. K.; KOHANY, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mobile DNA, [s. l.], v. 6, n. 1, p. 4–9, 2015. Disponível em: <<http://dx.doi.org/10.1186/s13100-015-0041-9>>

BERTOCCHI, N. A. Estudo de elementos Galluhop e CR1-like nos genomas de aves. 2017. UNIVERSIDADE FEDERAL DO PAMPA, [s. l.], 2017.

BERTOCCHI, N. A. et al. Evolutionary history of the mariner element galluhop in avian genomes. Mobile DNA, [s. l.], v. 8, n. 1, p. 11, 2017. Disponível em: <<http://mobilednajournal.biomedcentral.com/articles/10.1186/s13100-017-0094-z>>

BERTOCCHI, N. A. et al. Chromosomal distribution of CR1-like transposable element in woodpeckers (Aves Piciformes). 2018 [s.l: s.n.].

BIÉMONT, C.; VIEIRA, C. Genetics: Junk DNA as an evolutionary force. Nature, [s. l.], v. 443, n. 7111, p. 521–524, 2006.

BRADLEY S., H. et al. The western painted turtle genome, a model for the evolution of extreme physiological adaptations in a slowly evolving lineage. Genome Biology, [s. l.], v. 14, n. 3, 2013.

BURCH, J. B.; DAVIS, D. L.; HAAS, N. B. Chicken repeat 1 elements contain a pol-like open reading frame and belong to the non-long terminal repeat class of retrotransposons. Proceedings of the National Academy of Sciences of the United States of America, [s. l.], v. 90, n. 17, p. 8199–8203, 1993.

CARARETO, C. M. A.; MONTEIRO-VITORELLO, C. B.; SLUYS, M.-A. VAN. Elementos de transposição: diversidade, evolução, aplicações e impacto nos genomas dos seres vivos. 1. ed. Rio de Janeiro: FIOCRUZ, 2015.

CHALOPIN, D. et al. Evolutionary active transposable elements in the genome of the coelacanth. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, [s. l.], v. 322, n. 6, p. 322–333, 2014.

CHALOPIN, D. et al. Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates. *Genome Biology and Evolution*, [s. l.], v. 7, n. 2, p. 567–580, 2015.

CLARK, K. et al. GenBank. *Nucleic Acids Research*, [s. l.], v. 44, n. D1, p. D67–D72, 2016.

DREW, A. C.; BRINDLEY, P. J. A retrotransposon of the non-long terminal repeat class from the human blood fluke *Schistosoma mansoni*. Similarities to the chicken-repeat-1-like elements of vertebrates. *Molecular Biology and Evolution*, [s. l.], v. 14, n. 6, p. 602–610, 1997.

DOTTO, Bruno Reis et al. HTT-DB: new features and updates. **Database**, [s. l.], v. 2018, p. 1–5, 2017.

E.STUMPH, W. et al. A chicken middle-repetitive DNA sequence which shares homology with mammalian ubiquitous repeats. *Nucleic Acids Research*, [s. l.], v. 9, p. 5383–5397, 1981.

FINNEGAN, D.J. EUKARYOTIC TRANSPOSABLE ELEMENTS and GENOME EVOLUTION DAVID. *Tig*, [s. l.], v. 5, n. 4, p. 103–107, 1989.

GRAMATES, L. S. et al. FlyBase at 25: Looking to the future. *Nucleic Acids Research*, [s. l.], v. 45, n. D1, p. D663–D671, 2017.

GREEN, R. E. et al. Three crocodylian genomes reveal ancestral patterns of evolution among archosaurs. [s. l.], v. 2014, n. 0, 2014.

HAAS, N. B. et al. Chicken repeat 1 (CR1) elements, which define an ancient family of vertebrate non-LTR retrotransposons, contain two closely spaced open reading frames. *Gene*, [s. l.], v. 197, n. 1–2, p. 305–309, 1997.

HAAS, N. B. et al. Subfamilies of CR1 non-LTR retrotransposons have different 5'UTR sequences but are otherwise conserved. *Gene*, [s. l.], v. 265, n. 1–2, p. 175–183, 2001.

JARVIS, E. D. et al. Whole-genome analyses resolve early branches in the tree of life of modern birds. [s. l.], v. 346, n. 6215, p. 1126–1138, 2014.

JONES, R. N. McClintock's controlling elements: The full story. *Cytogenetic and Genome Research*, [s. l.], v. 109, n. 1–3, p. 90–103, 2005.

KAPITONOV, V. V.; JURKA, J. The esterase and PHD domains in CR1-like non-LTR retrotransposons. *Molecular Biology and Evolution*, [s. l.], v. 20, n. 1, p. 38–46, 2003.

KAPUSTA, A.; SUH, A. Evolution of bird genomes—a transposon's-eye view. *Annals of the New York Academy of Sciences*, [s. l.], v. 1389, n. 1, p. 164–185, 2017.

KATOH, K.; STANDLEY, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, [s. l.], v. 30, n. 4, p. 772–780, 2013.

KOEPFLI, K.-P.; PATEN, B.; O'BRIEN, S. J. The Genome 10K Project: A Way Forward. *Annual Review of Animal Biosciences*, [s. l.], v. 3, n. 1, p. 57–111, 2015. Disponível em: <<http://www.annualreviews.org/doi/10.1146/annurev-animal-090414-014900>>

KUMAR, S.; STECHER, G.; TAMURA, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular biology and evolution*, [s. l.], v. 33, n. 7, p. 1870–1874, 2016.

LARSSON, A. AliView: A fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*, [s. l.], v. 30, n. 22, p. 3276–3278, 2014.

MANTHEY, J. D. et al. Multiple and independent phases of transposable element amplification in the genomes of Piciformes (woodpeckers and allies). [s. l.], n. June, p. 1–35, 2018. Disponível em: <https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evy105/5020728>.

MATIOLI, S. R.; FERNANDES, F. M. C. *Biologia molecular e evolução*. 2 ed. Ribeirão preto. Sociedade Brasileira de Genética. 2012.

MCCLINTOCK, B. Chromosome organization and genic expression. *Cold Spring Harbor symposia on quantitative biology*, [s. l.], v. 16, p. 13–47, 1951.

MCCLINTOCK, B. Intranuclear systems controlling gene action and mutation. Brookhaven symposia in biology, [s. l.], n. 8, p. 58, 1956.

MCCULLERS, T. J.; STEINIGER, M. Transposable elements in *Drosophila*. Mobile Genetic Elements, [s. l.], v. 7, n. 3, p. 1–18, 2017. Disponível em: <<https://www.tandfonline.com/doi/full/10.1080/2159256X.2017.1318201>>

OKONECHNIKOV, K. et al. Unipro UGENE: A unified bioinformatics toolkit. Bioinformatics, [s. l.], v. 28, n. 8, p. 1166–1167, 2012.

PECCOUD, J. et al. Massive horizontal transfer of transposable elements in insects. Proceedings of the National Academy of Sciences, [s. l.], v. 114, n. 18, p. 4721–4726, 2017. Disponível em: <http://www.pnas.org/lookup/doi/10.1073/pnas.1621178114>

PINCHEIRA-DONOSO, D. et al. Global Taxonomic Diversity of Living Reptiles. PLoS ONE, [s. l.], v. 8, n. 3, p. 1–10, 2013.

PRITHAM, E. J. Transposable elements and factors influencing their success in eukaryotes. Journal of Heredity, [s. l.], v. 100, n. 5, p. 648–655, 2009.

RUGGIERO, R. P.; BOURGEOIS, Y.; BOISSINOT, S. LINE insertion polymorphisms are abundant but at low frequencies across populations of *Anolis carolinensis*. Frontiers in Genetics, [s. l.], v. 8, n. APR, p. 1–14, 2017.

SEDGWICK, W. T.; TYLER, H. W. A Short History of Science. 1939. [s.l: s.n.].

SHEDLOCK, A. M. et al. Phylogenomics of nonavian reptiles and the structure of the ancestral amniote genome. Proceedings of the National Academy of Sciences, [s. l.], v. 104, n. 8, p. 2767–2772, 2007. Disponível em: <<http://www.pnas.org/cgi/doi/10.1073/pnas.0606204104>>

SOTERO-CAIO, C. G. et al. Evolution and diversity of transposable elements in vertebrate genomes. Genome Biology and Evolution, [s. l.], v. 9, n. 1, p. 161–177, 2017.

SUH, A. et al. Mesozoic retrotransposons reveal parrots as the closest living relatives of passerine birds. Nature Communications, [s. l.], v. 2, n. 1, 2011.

SUH, A. et al. Multiple lineages of ancient CR1 retroposons shaped the early genome evolution of amniotes. *Genome Biology and Evolution*, [s. l.], v. 7, n. 1, p. 205–217, 2014.

SUH, A. et al. Ancient horizontal transfers of retrotransposons between birds and ancestors of human pathogenic nematodes. *NATURE COMMUNICATIONS*, [s. l.], 2016.

WATSON, J. D.; CRICK, F. H. D. A Structure for Deoxyribose Nucleic Acid. *Nature*, [s. l.], v. 171, n. 4356, p. 737–738, 1953.

WICKER, T. et al. A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics*, [s. l.], v. 8, n. 12, p. 973–982, 2007. Disponível em: <<http://www.nature.com/doi/10.1038/nrg2165>>

ALLEN, Garland. The Origin and Development of Molecular Biology. In: **Life Science in the Twentieth Century**. [s.l: s.n.]. p. 187–228.

BERTOCCHI, Natasha Avila. **ESTUDO DE ELEMENTOS GALLUHOP E CR1-LIKE NOS GENOMAS DE AVES**. 2017. UNIVERSIDADE FEDERAL DO PAMPA, [s. l.], 2017.

BERTOCCHI, Natasha Avila et al. Evolutionary history of the mariner element galluhop in avian genomes. **Mobile DNA**, [s. l.], v. 8, n. 1, p. 11, 2017. Disponível em: <<http://mobilednajournal.biomedcentral.com/articles/10.1186/s13100-017-0094-z>>

BRADLEY SHAFFER, H. et al. The western painted turtle genome, a model for the evolution of extreme physiological adaptations in a slowly evolving lineage. **Genome Biology**, [s. l.], v. 14, n. 3, 2013.

CARARETO, C. M. A.; MONTEIRO-VITORELLO, C. B.; SLUYS, M. A. V. **Elementos de transposição: diversidade, evolução, aplicações e impacto nos genomas dos seres vivos**. [s.l: s.n.].

DOTTO, Bruno Reis et al. HTT-DB: new features and updates. **Database**, [s. l.], v. 2018, p. 1–5, 2017.

E.STUMPH, William et al. A chicken middle-repetitive DNA sequence which shares homology with mammalian ubiquitous repeats. **Nucleic Acids Research**, [s. l.], v. 9,

p. 5383–5397, 1981.

FINNEGAN, David J. EUKARYOTIC TRANSPOSABLE ELEMENTS and GENOME EVOLUTION DAVID. **Tig**, [s. l.], v. 5, n. 4, p. 103–107, 1989.

GREEN, Richard E. et al. Three crocodylian genomes reveal ancestral patterns of evolution among archosaurs. [s. l.], v. 2014, n. 0, 2014.

JARVIS, Erich D. et al. Whole-genome analyses resolve early branches in the tree of life of modern birds. [s. l.], v. 346, n. 6215, p. 1126–1138, 2014.

JONES, R. N. McClintock's controlling elements: The full story. **Cytogenetic and Genome Research**, [s. l.], v. 109, n. 1–3, p. 90–103, 2005.

KOEPFLI, Klaus-Peter; PATEN, Benedict; O'BRIEN, Stephen J. The Genome 10K Project: A Way Forward. **Annual Review of Animal Biosciences**, [s. l.], v. 3, n. 1, p. 57–111, 2015. Disponível em: <<http://www.annualreviews.org/doi/10.1146/annurev-animal-090414-014900>>

MATIOLI, Sergio Russo; FERNANDES, Flora M. C. **Biologia molecular e evolução**. [s.l: s.n.].

MCCLINTOCK, B. Chromosome organization and genic expression. **Cold Spring Harbor symposia on quantitative biology**, [s. l.], v. 16, p. 13–47, 1951.

MCCLINTOCK, B. Intranuclear systems controlling gene action and mutation. **Brookhaven symposia in biology**, [s. l.], n. 8, p. 58, 1956.

PINCHEIRA-DONOSO, Daniel et al. Global Taxonomic Diversity of Living Reptiles. **PLoS ONE**, [s. l.], v. 8, n. 3, p. 1–10, 2013.

SHEDLOCK, A. M. et al. Phylogenomics of nonavian reptiles and the structure of the ancestral amniote genome. **Proceedings of the National Academy of Sciences**, [s. l.], v. 104, n. 8, p. 2767–2772, 2007. Disponível em: <<http://www.pnas.org/cgi/doi/10.1073/pnas.0606204104>>

SUH, Alexander et al. Mesozoic retroposons reveal parrots as the closest living relatives of passerine birds. **Nature Communications**, [s. l.], v. 2, n. 1, 2011.

SUH, Alexander et al. Ancient horizontal transfers of retrotransposons between birds

and ancestors of human pathogenic nematodes. **NATURE COMMUNICATIONS**, [s. l.], 2016.

WATSON, James D.; CRICK, Francis H. D. A Structure for Deoxyribose Nucleic Acid. **Nature**, [s. l.], v. 171, n. 4356, p. 737–738, 1953.

ZHANG, Guojie et al. Comparative genomic data of the Avian Phylogenomics Project. **GigaScience**, [s. l.], v. 3, n. 1, p. 1–8, 2014.