

**UNIVERSIDADE FEDERAL DO PAMPA**

**NATASHA AVILA BERTOCCHI**

**ESTUDO DE ELEMENTOS GALLUHOP E CR1-LIKE NOS GENOMAS DE AVES**

**São Gabriel  
2017**

**NATASHA AVILA BERTOCCHI**

**ESTUDO DE ELEMENTOS GALLUHOP E CR1-LIKE NOS GENOMAS DE AVES**

Dissertação apresentada ao Programa de Pós-graduação em Ciências Biológicas da Universidade Federal do Pampa como requisito parcial para obtenção do Título de Mestre em Ciências Biológicas.

Orientadora: Prof. Dr<sup>a</sup>. Analía Del Valle  
Garnero

Coorientador: Prof. Dr. Fabiano Pimentel  
Torres

**São Gabriel  
2017**

Ficha catalográfica elaborada automaticamente com os dados fornecidos  
pelo(a) autor(a) através do Módulo de Biblioteca do  
Sistema GURI (Gestão Unificada de Recursos Institucionais) .

B545e Bertocchi, Natasha Avila  
ESTUDO DE ELEMENTOS GALLUHOP E CR1-LIKE NOS GENOMAS DE AVES  
/ Natasha Avila Bertocchi.  
82 p.  
  
Dissertação(Mestrado)-- Universidade Federal do Pampa,  
MESTRADO EM CIÊNCIAS BIOLÓGICAS, 2017.  
"Orientação: Analía del Valle Garnero".  
  
1. Elementos transponíveis . 2. Aves. I. Título.

**NATASHA AVILA BERTOCCHI**

**ESTUDO DE ELEMENTOS GALLUHOP E CR1-LIKE NOS GENOMAS DE AVES**

Dissertação apresentada ao programa de Pós-graduação *stricto sensu* em Ciências Biológicas pela Universidade Federal do Pampa como requisito parcial para obtenção do Título de Mestre em Ciências Biológicas.

Dissertação de mestrado defendida e aprovada em 24 de março de 2017

Banca examinadora:

---

Prof. Dr<sup>a</sup>. Analía Del Valle Garnero  
Orientador  
UNIPAMPA

---

Prof. Dr. Daniel Ângelo Sganzerla Graichen  
UFSM

---

Prof. Dr. Ricardo José Gunski  
UNIPAMPA

Dedico esta dissertação às mulheres extremamente fortes e corajosas que me criaram, amaram e são os meus exemplos de persistência, luta e amor... Em especial à minha mãe, Juliana que me mostrou o Mundo e a Ciência.

## AGRADECIMENTO

Ao longo dessa trajetória muitas pessoas auxiliaram das mais variadas formas e eu só tenho a agradecer imensamente.

Ao Prof. Fabiano, por ser ter me orientado em toda essa trajetória desde a graduação e ter me apresentado ao mundo dos elementos transponíveis. Sou imensamente grata por todo o apoio para que essa dissertação fosse possível.

À minha orientadora, Prof. Analía, pela disponibilidade em me orientar confiando na minha capacidade e no meu trabalho. Também pelo grande empurrão para que fosse para São Paulo no primeiro ano do mestrado, a lembrança contínua de como dois anos passam rápido e a companhia prazerosa sempre.

Ao Dr. Gabriel Wallau pela disponibilidade impar em me ajudar e por todo aprendizado sobre elementos transponíveis e bioinformática nesses anos.

Ao Prof. Ricardo, por me inspirar com sua dedicação no trabalho com as aves e acreditar que os transposons renderiam frutos.

Ao Prof, Cesar Martins, UNESP- Botucatu, por abrir as portas de seu laboratório e possibilitar o segundo capítulo dessa dissertação.

Aos meus amigos, colegas de laboratório e dos laboratórios vizinhos pela parceria e companhia toooooodos os dias. Levo ótimas lembranças e deixo um grande obrigada por esses momentos que me foram muito importantes pessoal e profissionalmente. Os de sempre Thays, Rafa, Vanusa, Jéssica, Vivi um grande abraço!

Às minhas amigas pelo apoio emocional, mesmo de longe, com conversas intermináveis na madrugada em que me ouviam pacientemente mesmo sem fazer idéia de como os termos elemento, transponível, genomas e aves poderiam estar conectados.

À colega de trabalho e amiga Thays por ter sonhado, acredito e realizado junto comigo a execução dos experimentos. Sem tua companhia, ajuda e incentivo não seria possível realizar os experimentos de FISH, assim como todo o resto do trabalho. E principalmente meu imenso obrigada pelo apoio e amizade em todas as

horas. Os meus dias se tornaram muito menos árdios e mais divertidos em tua companhia.

Ao Tarta por ser e estar comigo continuamente em todo esse processo maluco que é e foi graduação e mestrado. Tenho certeza que nem Alzheimer poderá apagar ou diminuir a importância que tens.

Aos meus pais, que mesmo não fazendo ideia do que eu faço, me apóiam. Ao meu pai, Leandro, por ter aceitado essa empreitada comigo e assim ter se tornado o principal órgão de fomento neste estágio da minha formação. E principalmente por me auxiliar a ter foco no meu objetivo e incentivar a fazer o que eu sei fazer, como ele mesmo diz. A minha mãe Juliana, não existem palavras que exemplifiquem a tua importância nesse momento. Muito obrigada por me ensinar à importância das escolhas e a pensar. Ah! Levei a sério quando me disseste que estudaria a vida toda (eu devia ter uns 5 anos), hahah!

E por fim, com muito amor e saudade à minha pequena família 'inventada' (minha mãe, meu padrasto Paulo e meu irmão Pedro) por estarem comigo sempre, e serem onde estiverem o meu lar.

“Somos um planeta vivo, Sofia! Somos um grande barco navegando ao redor de um sol incandescente no universo. Mas cada um de nós é um barco em si mesmo, um barco carregado de genes navegando pela vida. Se conseguirmos levar essa carga ao porto mais próximo, nossa vida não terá sido em vão”

Jostein Gaarder



## RESUMO

Elementos transponíveis (TEs, do inglês *Transposable Elements*) são sequências que possuem a particularidade de se mobilizar dentro e entre genomas, estando presentes ubiquamente nos organismos e distribuídos por todos os ramos da árvore da vida. Os TEs influenciam os genomas hospedeiros de diferentes maneiras tendo um papel fundamental na evolução dos mesmos. São classificados em duas grandes classes, com base no seu intermediário de transposição: Classe I (retrotransposons) se transpõe via RNA, Classe II (transposons) se movem via DNA. Em aves, o conhecimento sobre elementos transponíveis ainda é muito insipiente, restringindo-se a poucos genomas sequenciados, principalmente o de *Gallus gallus*. Neste trabalho foram estudados dois tipos distintos de TEs: *galluhop* - elemento de Classe II, pertencente à superfamília *mariner/Tc1*, e CR1 - elemento de Classe I, ordem conhecida como LINE (do inglês *Long Interspersed Nuclear Element*), pertencente à superfamília CR1. Os objetivos desse trabalho foram caracterizar a história evolutiva das seqüências de homólogas de *galluhop* encontradas em genomas disponíveis de aves e, também, caracterizar a distribuição cromossômica de CR1-like em genomas de aves Piciformes (pica-paus). No primeiro capítulo, através de análises *in silico*, foi possível estimar a diversidade de cópias, as características estruturais e funcionais, a distribuição descontínua na linhagem aviária, a descrição de uma nova subfamília para família *mariner* e sugerir um evento de transferência horizontal do elemento *galluhop*. Destacamos nesta abordagem o primeiro registro de uma possível transferência horizontal de elemento transponível entre aves, e a nova subfamília *Gallus* restrita, até o momento, a elementos *galluhop* encontrados em aves. No segundo capítulo, mostramos por meio da técnica de FISH, a distribuição cromossômica de um mesmo elemento transponível, CR1-E-like, em diferentes genomas de aves, evidenciando padrão de distribuição muito diferente em cada genoma, mesmo entre espécies do mesmo gênero.

Palavras chave: Elementos transponíveis, CR1, *galluhop*, ordem Piciformes, ordem Galliformes

## ABSTRACT

Transposable Elements (TEs) are sequences that have the particularity of mobilizing within and between genomes, being ubiquitously present in organisms and distributed throughout the branches of the tree of life. TEs influence host genomes in different ways and play a key role in their evolution. They are classified into two major classes, based on their transposition intermediary: Class I (retrotransposons) transposes via RNA, Class II (transposons) move through DNA. In birds, the knowledge about transposable elements is still very insipient, being restricted to a few genomes sequenced, mainly the one of *Gallus gallus*. In this work two distinct types of TEs were studied: *galluhop* - element of Class II, belonging to the superfamily *mariner* / Tc1, and CR1 - element of Class I, order known as LINE (of the English Long Interspersed Nuclear Element), belonging to the superfamily CR1. The objectives of this work were to characterize the evolutionary history of *galluhop* homolog sequences found in avian available genomes and also to characterize the chromosomal distribution of CR1-like in genomes of woodpeckers. In the first chapter, through *in silico* analysis, it was possible to estimate the diversity of copies, the structural and functional characteristics, the discontinuous distribution in the avian line, the description of a new subfamily for the *mariner* family and to suggest a horizontal transfer event of the *galluhop* element. We highlight in this approach the first record of a possible horizontal transfer of transposable element between birds, and the new *Gallus* subfamily restricted, until now, to *galluhop* elements found in birds. In the second chapter, we show through the FISH technique the chromosomal distribution of the same transposable element, CR1-E-like, in different bird genomes, showing a very different distribution pattern in each genome, even among species of the same genus.

Keywords: Transposable elements, CR1, *galluhop*, Piciformes order, Galliformes order

## LISTA DE FIGURAS

### INTRODUÇÃO

FIGURA 1- Classificação dos elementos transponíveis.....	19
FIGURA 2- Filogenia das aves e características genômicas das ordens. ....	22
FIGURA 3- Principais elementos transponíveis identificados em amniotas. ....	23

### CAPÍTULO 1

Figure 1- Schematic representation of the reconstructed of <i>galluhop</i> copies compared to the <i>galluhop</i> consensus.....	47
Figure 2- Kimura 2 Parameters distance distribution of each copy to its corresponding consensus sequence .....	48
Figure 3- Phylogeny of mariner-like transposases .....	49
Figure 4- Horizontal Transfer hypothesis of <i>galluhop</i> elements.....	50
Figure S1- Experimental design procedure showing steps of the analysis.....	53
Figure S2- Graphical representation of the conservation of terminal inverted repeats (TIRs) 5' e 3' <i>galluhop</i> element in the six genomes .....	54
Figure S3- Graphical representation of the conservation of target site duplication (TSDs) 5' e 3' <i>galluhop</i> element in the six genomes.....	55

### CAPÍTULO 2

FIGURA 1- Fluxograma do trabalho .....	61
FIGURA 2- Amplificação do TE CR1-E-like em duas espécies de pica-paus.....	66
FIGURA 3- Cariótipo parcial de <i>V. spilogaster</i> (VSP).. ..	67
FIGURA 4- Bandeamento C em cromossomos metafásicos de um macho de <i>V. spilogaster</i> .....	67
FIGURA 5- Experimentos de FISH com sondas de CR1-like em cromossomos metafásicos de quatro espécies pica-paus e <i>G. gallus</i> . ....	69
FIGURA 6- Representação esquemática da distribuição de CR1-E-like no primeiro par e cromossomo sexual Z.....	70

## LISTA DE TABELAS

### CAPÍTULO 1

Table 1- Avian genomes with <i>galluhop</i> and characteristics of copies.....	51
Table 2- Identity of the <i>galluhop</i> consensus sequences in the genomes of birds.....	52
Table S1- GenBank access numbers birds genomes .....	56

### CAPÍTULO 2

TABELA 1- Espécies amostradas, quantidade de espécimes e local das coletas. ...	63
---	----

## LISTA DE ABREVIATURAS E SIGLAS

µl - microlitro

CCA - *Colaptes campestris*

CEUA - Comitê de Ética no Uso de Animais

CME - *Colaptes melanochloros*

CR1 - *Chicken repeat 1*

DAPI - *4',6-diamidino-2-phenylindole*

DNA - Ácido desoxirribonucleico - do inglês: DeoxyriboNucleic Acid

FISH - Hibridização *in situ* fluorescente - do inglês *Fluorescence in situ Hybridization*

Gb - 1 giga base equivalente a 1.000.000.000 pares de bases

GGA - *Gallus gallus*

HTT- Transferência horizontal de elementos de transposição - do inglês *Horizontal Transposon Transfer*

Kb- 1 kilo base equivalente a 1.000 pares de bases

LINE- Elementos nucleares dispersos longos - do inglês Long Interspersed Nuclear Elements

LTR- Repetição terminal longa - do inglês *Long Terminal Repeat*

MCA - *Melanerpes candidus*

MgCl<sub>2</sub> - Cloreto de Magnésio

MITE - Elemento de transposição miniatura - do inglês *Miniature Transposable Element*

mM - milimolar

mtDNA - DNA mitocondrial

Mya - Milhões de anos atrás - do inglês *Millions of Years Ago*

ORF - Fase aberta de leitura - do inglês *Open Reading Frame*

pb - pares de bases

PCR - Reação em cadeia da polimerase - do inglês *Polymerase Chain Reaction*

pMol - pico Moles

PRINS - *Primed in-situ labelling*

RNA - Ácido ribonucleico - do inglês: RiboNucleic Acid

RNase - Ribonuclease

SINE - Elementos nucleares dispersos curtos - do inglês *Short Interspersed Element*

SISBIO - Sistema de Autorização e Informação em Biodiversidade

SSC - Solução salina concentrada - do inglês *Saline-Sodium Citrate*

TEs - Elemento transponível – do inglês *Transposable Elements*

TIR - Repetição terminal invertida – do inglês *Terminal Inverted Repeat*

TSD - Duplicação de sítio - do inglês *Target Site Duplication*

UTR - Regiões não codificantes - do inglês: *Untranslated Region*

UV - ultravioleta

VSP - *Veniliornes spilogaster*

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	16
1.1 Elementos transponíveis .....	16
1.2 Classificação dos TEs .....	18
1.3 Singularidade dos genomas e TEs das Aves .....	20
1.4 Elementos transponíveis em aves.....	23
<b>2 OBJETIVOS</b> .....	25
2.1 Objetivo geral .....	25
2.2 Objetivos específicos .....	25
<b>3 CAPÍTULO 1</b> .....	26
<b>Evolutionary history of the <i>mariner galluhop</i> element in the avian genomes</b>	26
3.1 Abstract.....	27
3.2 Introduction .....	28
3.3 Material and Methods.....	30
3.4 Results .....	32
3.5 Discussion.....	34
3.6 Acknowledgments .....	39
3.7 References.....	39
3.8 Internet Resources.....	43
<b>4 CAPÍTULO 2</b> .....	59
<b>Distribuição cromossômica do element transponível CR1-like em pica-paus (Aves Piciformes)</b> .....	59
4.1 Introdução .....	59
4.2 Material e Métodos.....	61
4.3 Resultados .....	66
4.4 Discussão .....	71
4.5 Referências .....	74
<b>5 CONSIDERAÇÕES FINAIS</b> .....	78
<b>REFERÊNCIAS</b> .....	79

# 1 INTRODUÇÃO

## 1.1 Elementos transponíveis

Os elementos transponíveis (TEs do inglês *Transposable Elements*) são sequências de DNA, encontradas repetidamente nos genomas, que possuem como particularidade a capacidade de mudarem de posição tanto dentro como entre genomas. Essa característica, a mobilização, revolucionou o conhecimento sobre o que, atualmente, denomina-se genoma e o seu funcionamento dentro das células.

Até, aproximadamente, metade do século XX acreditava-se, devido à análise de microscopia óptica dos cromossomos, que os genes eram “invioláveis” e estavam fixos dentro dos cromossomos (VARANI et al., 2015). Contudo, Barbara McClintock observou um intrincado mecanismo de dinâmica cromossômica, onde uma quebra no braço curto do cromossomo 9 de milho (*Zea mays*) estaria associada a modificações na coloração dos grãos de milho (JONES, 2005; MCCLINTOCK, 1950). McClintock percebeu que os locos (Ds e Ac) poderiam mudar de posição dentro do genoma de forma aleatória, ocasionando a mudança de coloração nos grãos. No entanto, foi fortemente desacreditada pela comunidade científica devido à crença na estaticidade do material genético.

Com o desenvolvimento de técnicas de biologia molecular e assim o aprimoramento do conhecimento sobre as características estruturais do genoma, na década de 80, os TEs foram redescobertos mobilizando-se no genoma de *Escherichia coli*, associados a mutações em *Drosophila melanogaster* e em milho (VARANI et al., 2015). Desde então, os TEs foram encontrados em todos os ramos da árvore da vida estudados, desde organismos mais simples como bactérias e fungos, à plantas, invertebrados, todas as classes de vertebrados, e podem ser considerados virtualmente presente em todas as espécies eucarióticas analisadas (PRITHAM, 2009; WICKER et al., 2007). Entretanto, existem algumas exceções, restrita a espécies unicelulares estudadas como no genoma da alga vermelha *Cyanidioschyzon merolae*, 6 espécies de Apicomplexa (*Babesia bovis*, *Cryptosporidium hominis*, *Cryptosporidium parvum*, *Plasmodium falciparum* e *Theileria parva*), além de uma espécie de Unikont, *Encephalitozoon cuniculi* (PRITHAM, 2009).



As características, de mobilização e parasitária, tornaram os TEs as sequências mais abundantes e ubíquos na natureza (AZIZ; BREITBART; EDWARDS, 2010), influenciando a trajetória evolutiva dos genomas hospedeiros de diferentes formas. Os TEs são geralmente encontrados em regiões não codificadoras, contudo também podem ser observados junto à éxons ou mesmo íntrons. Os TEs podem, por exemplo: fornecer promotores alternativos para expressão de genes; quando inseridos em genes inativá-los; viabilizar variantes de proteínas através do *splicing*; ter ação epigenética quando modificam a expressão de um gene sem modificar sua sequência; e ocasionar rearranjos cromossômicos (CARARETO; MONTEIRO-VITORELLO; SLUYS, 2015). Nas regiões não codificantes os TEs podem, por exemplo, assumir um papel estrutural estando associados ao centrômero e/ou regiões terminais dos cromossomos (OLIVEIRA et al., 2013)

Os TEs tiveram várias denominações conforme a evolução do conhecimento sobre a estrutura e função do DNA, como o termo “DNA lixo” devido aos TEs fazerem parte de uma grande porção do genoma que não codificam para polipeptídeos e “genes saltadores” devido à capacidade de se mobilizarem de um sítio a outro do DNA (VARANI et al., 2015). Já os termos “parasitas genômicos” e “DNA egoísta” advêm dos TEs se mobilizarem indiferente das necessidades estruturais e funcionais do genoma hospedeiro e por também ser uma grande parcela desses genomas, em humanos representa 50%, aproximadamente 40% do genoma de roedores e podem chegar a mais de 60% do genoma, como em milho (BIÉMONT; VIEIRA, 2006; GIBBS et al., 2004; LANDER et al., 2001).

Com isso, podemos observar uma rede de interação complexa entre os TEs e os genomas. Com o aumento dos projetos de sequenciamento de genomas de organismos em todos os ramos da árvore da vida, ocorreu uma “explosão” de informações sobre TEs, sendo descobertos continuamente novos elementos, incluindo elementos muitas vezes com características diferentes das conhecidas e com funções diferentes nos genomas. Por isso, o estudo contínuo dos TEs, principalmente em genomas pouco estudados até o momento, como das aves, é extremamente importante para compreender melhor a interação e evolução destas sequências com os genomas.

## 1.2 Classificação dos TEs

Conforme o conhecimento sobre TEs foi se ampliando surgiram diferentes propostas de classificação (FINNEGAN, 1989; KAPITONOV; JURKA, 2008; WICKER et al., 2007). As propostas possuem abordagens diferentes, porém concordam com o padrão similar com a classificação sistemática utilizado para espécies e em uma primeira divisão enzimática, baseada na molécula intermediária de transposição. Por exemplo, os elementos que se transpõem via uma molécula de RNA, utilizando uma transcriptase reversa homóloga às usadas por vírus, pertencem a Classe I e também são chamados de retrotransposons; e a Classe II, os transposons, é constituída dos elementos que se movem via DNA utilizando geralmente uma transposase. Os TEs de Classe I são transcritos em um RNA, reversamente transcritos em DNA e inseridos em outra região do genoma; e os TEs de Classe II são excisados pela transposase e reinseridos em outro local do genoma (FINNEGAN, 1989).

Nas divisões subsequentes as propostas divergem, a amplamente aceita é de Wicker e colaboradores (2007) (Figura 1). A proposta divide hierarquicamente os elementos em:

Subclasse - subclasse I e II para os elementos de classe II conforme o mecanismo de excisão dos elementos;

Ordem – conforme a organização geral e enzimologia dos TEs, dentro de uma mesma ordem os elementos compartilham as mesmas estratégias de replicação;

Superfamília – distinguem-se umas das outras por características como estrutura das proteínas, domínios não codificantes, presença e tamanho dos sítios de duplicação;

Família – são definidas pela conservação das sequências de aminoácidos;

Subfamília – são definidas por análises filogenéticas, podem distinguir grupos de elementos autônomos de elementos não autônomos.

Classification		Structure	TSD	Code	Occurrence
Order	Superfamily				
<b>Class I (retrotransposons)</b>					
LTR	Copia	→ GAG AP INT RT RH →	4-6	RLC	P, M, F, O
	Cypsy	→ GAG AP RT RH INT →	4-6	RLG	P, M, F, O
	Bel-Pao	→ GAG AP RT RH INT →	4-6	RLB	M
	Retrovirus	→ GAG AP RT RH INT ENV →	4-6	RLR	M
	ERV	→ GAG AP RT RH INT ENV →	4-6	RLE	M
DIRS	DMS	→ GAG AP RT RH YR ←	0	RYD	P, M, F, O
	Ngano	→ GAG AP RT RH YR →	0	RYN	M, F
	VIPER	→ GAG AP RT RH YR →	0	RYY	O
PLE	Penelope	← RT EN →	Variable	RPP	P, M, F, O
LINE	R2	← RT EN →	Variable	RIR	M
	RTE	← APE RT →	Variable	RIT	M
	Jockey	← ORF1 APE RT →	Variable	RIJ	M
	LI	← ORF1 APE RT →	Variable	RIL	P, M, F, O
	I	← ORF1 APE RT RH →	Variable	RII	P, M, F
SINE	IRNA	← →	Variable	RST	P, M, F
	YSL	← →	Variable	RSL	P, M, F
	5S	← →	Variable	RSS	M, O
<b>Class II (DNA transposons) - Subclass 1</b>					
TIR	Tc1-Mariner	→ Tase* ←	TA	DTT	P, M, F, O
	hAT	→ Tase* ←	8	DIA	P, M, F, O
	Mutator	→ Tase* ←	9-11	DTM	P, M, F, O
	Merlin	→ Tase* ←	8-9	DTE	M, O
	Transib	→ Tase* ←	5	DTR	M, F
	P	→ Tase ←	8	DTP	P, M
	PiggyBac	→ Tase ←	TTAA	DTB	M, O
	FB - Harbinger	→ Tase* CR2 ←	3	DTH	P, M, F, O
	CACTA	→ Tase CR2 ←	2-3	DTC	P, M, F
Crypton	Crypton	← YR →	0	DYC	F
<b>Class II (DNA transposons) - Subclass 2</b>					
Helitron	Helitron	← RPA Y2 HEL →	0	DWH	P, M, F
Maverick	Maverick	← C-INT ATP CYP POL B →	6	DMM	M, F, O

Structural features				
→	Long terminal repeats	←	Terminal inverted repeats	→
→	Diagnostic feature in non-coding region	→	Region that can contain one or more additional CRFs	→
→	Coding region	→	Non-coding region	→

Protein coding domains				
AP, Aspartic proteinase	APE, Apurinic endonuclease	ATP, Packaging ATPase	C-INT, C-integrase	CYP, Cysteine protease
ENV, Envelope protein	GAG, Capsid proteins	HEL, Helicase	INT, Integrase	EN, Endonuclease
POL B, DNA polymerase B	RH, RNase H	RPA, Replication protein A (found only in plants)	ORF, Open reading frame of unknown function	RT, Reverse transcriptase
Tase, Transposase (* with DDE motif)		YR, Tyrosine recombinase		Y2, YR with YY motif

Species groups			
P, Plants	M, Metazoans	F, Fungi	O, Others

FIGURA 1- Classificação dos elementos transponíveis. Proposta por Wicker et al. (2007), imagem adaptada

Outra forma de ordenar os TEs é quanto à autonomia para se mobilizar. Os elementos que possuem intacta toda a codificação para produzir a maquinaria enzimática para transposição são chamados autônomos, e os TEs sem ou com grandes deleções na região de codificação são chamados não-autônomos. Na Classe II, os elementos não-autônomos, possuem deleções na região da enzima de mobilização transposase e são conhecidos como MITEs (do inglês *Miniature Inverted-repeat Transposable Elements*). Os MITEs são pequenos elementos (aproximadamente 50-500 pb), com deleções na região da ORF (do inglês *Open Reading Frame*), sequências terminais conservadas, duplicação do sítio alvo de

inserção e geralmente alto número de cópias (FATTASH et al., 2013). São considerados parasitas das cópias autônomas, pois utilizam a transposase desses elementos para se mobilizarem, podendo inclusive sobrepor o número de cópias autônomas (LE ROUZIC; CAPY, 2006). Devido a essa característica, os MITEs, são considerados superparasitas, pois parasitam os parasitas genômicos (WALLAU et al., 2014; WALLAU; ORTIZ; LORETO, 2012).

Os elementos estudados e apresentados nessa dissertação pertencem às duas classes descritas anteriormente: o primeiro é um elemento de Classe II, pertencente à superfamília *mariner/Tc1*, família *mariner* que é constituída de várias subfamílias, o segundo é um elemento de Classe I, ordem conhecida como LINE (do inglês *Long Interspersed Nuclear Element*), pertencente à superfamília CR1.

### 1.3 Singularidade dos genomas e TEs das Aves

A classe Aves é um grupo monofilético, com aproximadamente dez mil espécies espalhadas por todo o globo terrestre que compartilham a capacidade de voar, não restritiva ao grupo, e que condicionou adaptações inclusive nos seus genomas. As necessidades metabólicas exigidas pelo voo condicionaram adaptações aos seus genomas, podendo assim ter favorecido o desenvolvimento de genomas menores (WRIGHT; GREGOR; WITT, 2014). A variação de tamanho de genomas das aves vai de 0,89 Gb a 2,11 Gb enquanto que em mamíferos variam de 1,59 Gb a 8,22 Gb (Figura 3) (KAPUSTA; SUH, 2016).

Em contrapartida, é o grupo dentro dos amniotas, com maior número cromossômico,  $2n = 136-142$  em *Corythaixoides concolor* e uma grande variação sendo o menor  $2n = 40$  em *Falco columbarius* (CHRISTIDIS, 1990; KAPUSTA; SUH, 2016; NISHIDA et al., 2008). Esse perfil estrutural do genoma se dá pela pequena quantidade de macrocromossomos (geralmente de 10-14 pares) e o restante é formado por microcromossomos que, provavelmente, se originaram de inúmeras fissões dos macrocromossomos ao longo da trajetória evolutiva da classe, e é onde também se concentram as regiões gênicas (BURT, 2002; HILLIER et al., 2004).

A diminuição do tamanho dos genomas de aves está associada à diminuição da fração repetitiva do genoma, que é formada principalmente pelos TEs (ZHANG et al.,

2014). Nas aves os TEs representam entre 8-10% do genoma total e em outros grupos de animais, como os mamíferos, muitas vezes chegam a representar aproximadamente 50% do genoma. Contudo, na ordem Piciformes observou-se maior densidade de TEs chegando a 22% do genoma de *Picoides pubescens* (GIBBS et al., 2004; KAPUSTA; SUH, 2016; ZHANG et al., 2014).

Os dados sobre TEs aumentaram drasticamente com a disponibilidade de novos genomas de aves sequenciados, até recentemente os genomas disponíveis eram de *Gallus gallus* (galinha doméstica), *Meleagris gallopavo* (peru) e *Taeniopygia guttata* (mandarim), os dois primeiros devido à importância econômica e o terceiro como organismo modelo para estudos (CHICKEN; SEQUENCING, 2004; DALLOUL et al., 2010). Com o consórcio de sequenciamento de genomas G10k Project foram sequenciados 50 genomas com espécies representativas de todas as ordens da classe Aves, fornecendo assim dados expressivos sobre os TEs (Figura. 2)(JARVIS et al., 2014; KOEPFLI et al., 2015). Atualmente são mais de 70 genomas de aves seqüenciados disponíveis.

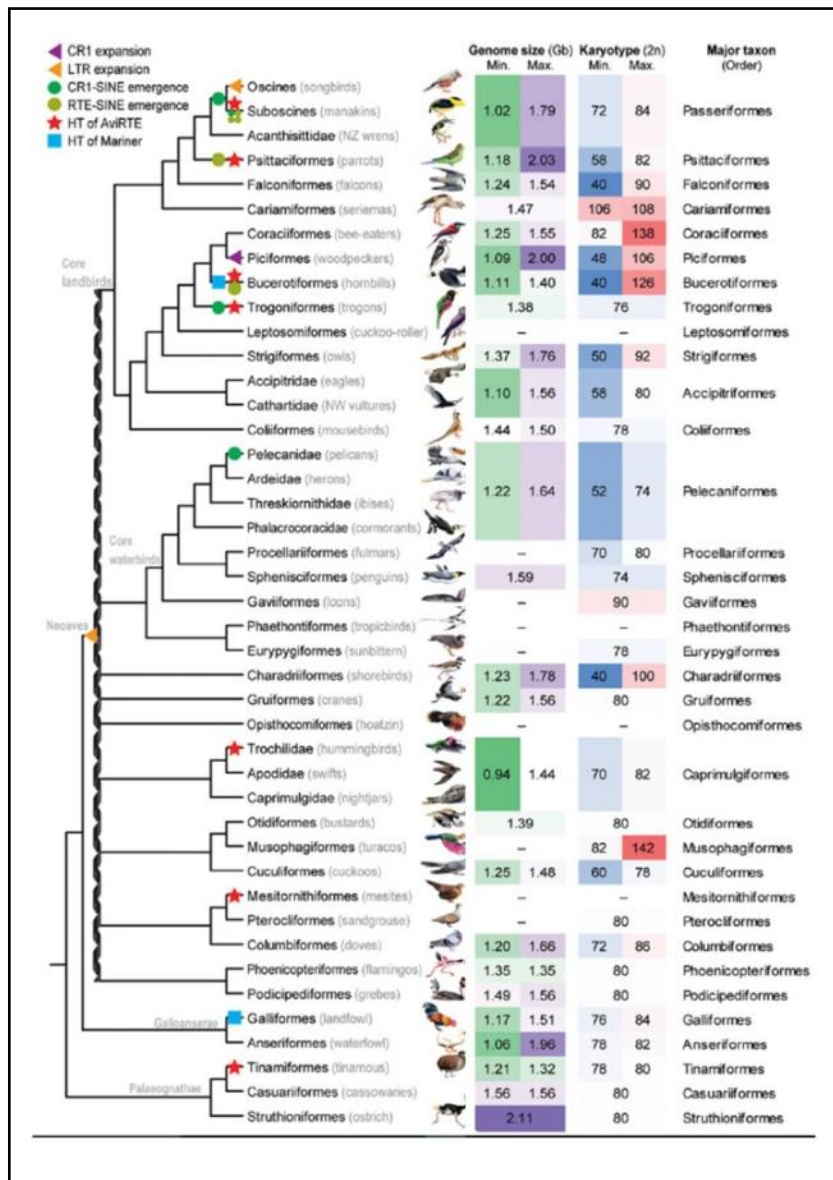


FIGURA 2- Filogenia das aves e características genômicas das ordens. Adaptado de Kapusta e Suh (2016)

## 1.4 Elementos transponíveis em aves

As principais superfamílias de TEs encontradas nas Aves e as suas características estruturais podem ser observadas na figura 3.

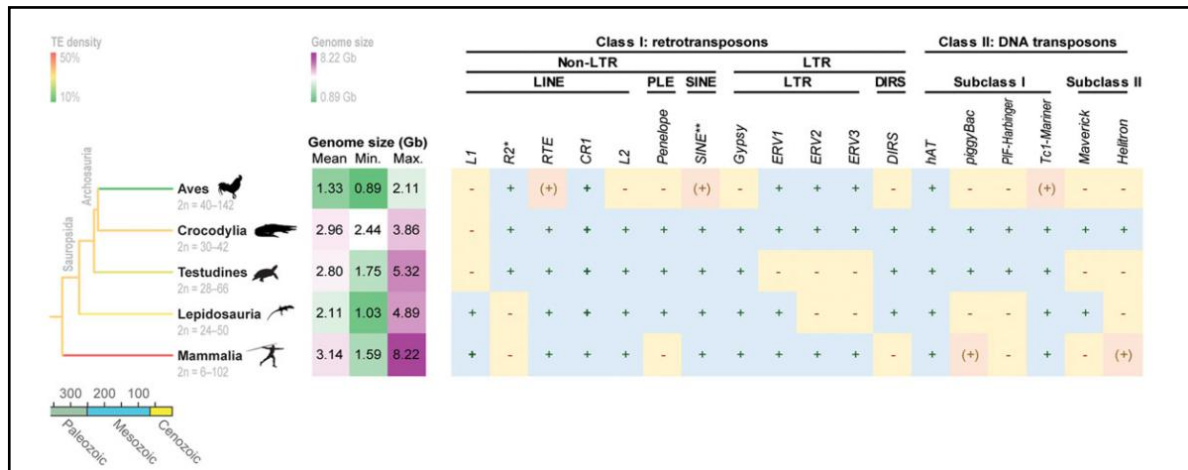


FIGURA 3- Principais elementos transponíveis identificados em amniotas. Retirado de Kapusta e Suh (2016).

Em aves, foram encontrados apenas dois transposons até o momento, o *charlie* (superfamília hAT) e o *galluhop* (família *mariner*), ambos identificados inicialmente em *Gallus gallus* (WICKER et al., 2005) (Figura 3). O elemento *galluhop* possui as características comuns à família *mariner* que são: aproximadamente 1300 pb de tamanho, com uma única ORF que codifica a transposase do elemento, repetições terminais invertidas conservadas de aproximadamente 28 pb e duplicação da sequência TA no sítio de inserção (ROUAULT et al., 2009; WALLAU et al., 2011). O primeiro elemento *mariner* descrito foi em *Drosophila mauritiana* e desde então são continuamente descobertos em outras espécies, incluindo outros invertebrados e vertebrados como mamíferos (HARTL; LOHE; LOZOVSKAYA, 1997; JACOBSON; MEDHORA; HARTL, 1986; OLIVEIRA et al., 2013).

Os elementos de Classe I possuem mais representantes nos genomas aviários, porém menos do que podemos observar em outros grupos de amniotas (Figura 3). Destaca-se em Aves principalmente a superfamília de retrotransposons CR1 que é considerada ubíqua nos amniotas, e a mais abundante nos genomas dos saurapsidas (KORDIS, 2010; SUH et al., 2014).

A superfamília CR1 (Chicken Repeat 1) foi identificada primeiramente no genoma de *G. gallus*, esta inserida na ordem dos LINEs (do inglês *Long Interspersed Nuclear Elements*) (STUMPH et al., 1981; WICKER et al., 2005, 2007). A superfamília possui as características comuns esta ordem de TEs como: duas ORFs, a ORF 2 responsável pela codificação das enzimas necessárias para mobilização incluindo a transcriptase reversa; a codificação para a enzima transcriptase reversa conservada dentro da superfamília e; altamente também são elementos truncados na porção 5' UTR e conservados na porção 3' UTR (BURCH; DAVIS; HAAS, 1993; HAAS et al., 1997, 2001; KAJIKAWA; OHSHIMA; OKADA, 1997; KAPITONOV; JURKA, 2003). Os elementos CR1 possuem diferentes subfamílias descritas com base principalmente no genoma de *G. gallus* e répteis, as caracterizadas em *G. gallus* são nomeadas em ordem alfabética de CR1- A até CR1- F, CR1- X e CR1-Y, suas sequências são a maioria pequenos fragmentos do elemento de aproximadamente 500 pb (WICKER et al., 2005).



## 2 OBJETIVOS

### 2.1 Objetivo geral:

- Contribuir para o conhecimento sobre a presença, dinâmica e evolução de elementos transponíveis nos genomas de aves.

### 2.2 Objetivos específicos:

- Caracterizar a história evolutiva das seqüências de homólogas de *galluhop* encontradas em genomas disponíveis de aves.
- Caracterizar a distribuição cromossômica de CR1-like em genomas de pica-paus.

### 3 CAPÍTULO 1

Artigo a ser submetido à Revista Genetics and Molecular Biology (GMB)

#### **Evolutionary history of the *mariner galluhop* element in the avian genomes**

Natasha Avila Bertocchi<sup>1,2</sup>, Gabriel da Luz Wallau<sup>3</sup>, Fabiano Pimentel Torres<sup>2</sup>,  
Ricardo José Gunski<sup>1,2</sup>, Analía del Valle Garnero<sup>\*1,2</sup>

<sup>1</sup>Programa de Pós-graduação em Ciências Biológicas, Universidade Federal do Pampa (Unipampa), São Gabriel, RS, Brazil

<sup>2</sup>Laboratório de Diversidade Genética Animal, Universidade Federal do Pampa (Unipampa), São Gabriel, RS, Brazil

<sup>3</sup>Departamento de Entomologia, Centro de Pesquisas Aggeu Magalhães – FIOCRUZ-CPqAM, Recife, PE, Brazil

**Short title:** *mariner* transposon in avian genomes

**Keywords:** *galluhop*, *mariner*, avian genome, horizontal transfer

**Corresponding author:** Analía del Valle Garnero, Aluizio Barros Macedo street, BR290 Km 423, [analiagarnero@yahoo.com.br](mailto:analiagarnero@yahoo.com.br)

### 3.1 Abstract

Taking advantage of rich genomic data available for birds we characterized the evolutionary history of the *galluhop* element, originally described in *Gallus gallus*, through the use of several bioinformatic analysis. *galluhop* homologous sequences were found in six from 72 genomes analysed: *Gallus gallus*, *Buceros rhinoceros*, *Coturnix japonica*, *Colinus virginianus*, *Lyrurus tetrrix* and *Meleagris gallopavo* and its copy number varied from 5 copies in *C. japonica* to 10158 in *G. gallus*. All 6 species presented short elements suggesting the presence of Miniature Inverted repeats Transposable Elements (MITEs), but only four species showed potential full length partners. Although, we found no potential coding copies we do detected an ancient massive MITE amplification in *G. gallus* and *M. gallopavo* genomes. The other four species also presented MITEs suggesting that *galluhop*-like elements generates MITEs frequently. Phylogenetic analysis of reconstructed coding sequences showed that *galluhop* homolog sequences form a new *mariner* subfamily which we named *Gallus*. Inter species distance analysis showed high identity between the consensus of *B. rhinoceros* and the other five related species (from 88 to 92%), suggesting that probably a horizontal transfer took place between the ancestral of *L. tetrrix* and *B. rhinoceros*.

### 3.2 Introduction

Transposable elements (TEs) are mobile genetic genes widely distributed and abundant component of many eukaryotic genomes. TEs can be classified in two main classes based on its transposition mechanism: Class I (moves through an RNA intermediate) and Class II (through DNA intermediate) (Finnegan 1989; Wicker et al. 2007; Kapitonov and Jurka 2008). Successful proliferation of TEs in genomes is linked with its replicative and mobile capacity inside of the host genome and also between genomes (Carareto et al. 2015). On the other hand, such mobility capacity is most of the time neutral or deleterious to the host organism. New TE insertions in genes coding regions or even at upstream/downstream positions can have a huge impact on flanking genes (Feschotte 2008). In addition, such highly similar and repetitive sequences across the genome also generate several recombination spots that can lead to chromosomal inversions and deletions (Cáceres et al. 2001; McVean 2010). However, an increasing body of evidence is showing that new TEs insertions introduce variability and can eventually be adaptive to the host genome (Volff 2006; Casola et al. 2008).

TEs are an integral part of host genomes and hence are vertically inherited to descendants through the male and female germ line DNA and from ancestral to extant species in a evolutionary scale (Schaack et al. 2010). However, nowadays, compelling evidences, in a wide variety of taxa, are showing that, Horizontal Transposon Transfer (HTT), the exchange of genetic material between isolated sexual species, is an effective way in which TEs invade new genomes and colonize other species (Schaack et al. 2010; Wallau et al. 2012). Currently, more than 577 HTTs events were reported (Dotto et al. 2015). and the *mariner* family are the Class

II DNA transposons with the highest number of HTT cases reported (52) so far (Dotto et al. 2015). Such events were reported in a wide variety of taxa as insects and mammals (Wallau et al. 2011; Oliveira et al. 2012). However, only two HTT events involving birds species were reported so far: two retrotransposons (AviRTE) which took place between the ancestral of birds and human pathogenic nematodes (Suh et al. 2016).

Non-autonomous elements can emerge at any step of the TE “life cycle” through deletion or internal region degeneration eventually keeping its transposition capacity in the presence of autonomous or codifying copies (Carareto et al. 2015). Non-autonomous elements originated from Class II transposons are known as Miniature Inverted-repeat Transposable Elements (MITEs) (Wicker et al. 2007). Such elements presents deletions or degenerated coding region and conserved Terminal Inverted Repeats (TIRs) which can be recognized by functional transposases (González and Petrov 2009; Fattash et al. 2013). MITEs have already been associated with several Class II superfamilies such as hAT, *P* element and Tc1/mariner (Osborne et al. 2006; Yang et al. 2009; Deprá et al. 2012). Usually MITEs reach higher copy number than its autonomous counterparts, which can determines a dead end for this TE family in long term (Le Rouzic and Capy 2006).

Although TEs are nowadays recognized as major players in genome evolution the knowledge about TEs is scarce in some group of species as birds (Kordis 2010; Kapusta and Suh 2016). One of the reasons, was the scarcity of available genomes, but since 2014, more than 70 avian genomes are available (Koepli et al. 2015). Among few studies focusing on TEs in birds genomes, a reduction in the repetitive DNA were detected in Sauropsids probably due to the purifying selection pressure

acting on metabolism optimization (Organ et al. 2007; Zhang and Edwards 2012; Kapusta e Suh 2016). In particular, are the Class II TEs which are abundant in other eukaryotic species but presented a limited diversity in the few avian genomes studied so far: *Gallus gallus* and *Meleagris gallopavo* (Chicken Sequencing 2004; Dalloul et al. 2010; Kordis 2010).

Elements from the *mariner* family generally are 1.3 kb long, contain TIRs of approximately 28 bp and a unique ORF (Open Reading Frame) which codes for a transposase (Wallau et al. 2011). Due to the great diversity of the *mariner* family such elements were classified in subfamilies based on phylogenetic analysis. The classification proposed by Rouault et al. (2009) includes 12 subfamilies (*mauritiana*, *cecropia*, *rosa*, *mellifera*, *lineata*, *capitata*, *irritans*, *briggsae*, *elegans*, *Atlantis* and *CR1*). Among Class II TEs found in avian genomes a *mariner*-like element named *galluhop* was characterized (Chicken Sequencing 2004; Wicker et al. 2005), but up to now no other work focused in understand its evolution in other avian species.

Based on the reasoning presented above we aimed to characterize the evolutionary history of *galluhop* homolog sequences found in avian available genomes. Our results showed that *galluhop*-like sequences compose a new *mariner* subfamily which was probably involved in a horizontal transfer event between avian species and that the emergence of MITEs is a common process in this new subfamily.

### **3.3 Material and Methods**

#### **3.3.1 Bioinformatic workflow**

##### **3.3.1.1 Genome search for *galluhop* homologs**

Nucleotide sequence from the *galluhop* consensus described by Wicker et al. (2005) was obtained from the Repbase database (Jurka et al. 2005). 72 avian genomes were available up to October of 2016 (Table S1). BLASTn searches were performed using *galluhop* consensus sequence from Repbase using default parameters. Only blast results which presented E-value lower than  $e^{-10}$  were further analyzed. *In house* perl scripts were used to retrieve all sequences and 200 base pairs of flanking sequences from each copy.

Sequence alignment of all copies plus flanking sequences from each species were performed with MAFFT v.7 (Kato and Standley 2013) (Figure S1).

### 3.3.1.2 Functional characterization

Resulting alignments were manually inspected and corrected in order to correctly identify TIRs and duplication sites. TIRs conservation were visually determined using Weblogo (Crooks et al. 2004). After identification and definition of element copies boundaries all copies were characterized by the presence or not of ORFs using the OrfFinder script implemented in UGENE (Okonechnikov et al. 2012) and the script getorf implemented in Emboss (Emboss getorf) with the following parameters: -minsize 900 -find 1 -methionine Y. Copies were classified as i) presenting a predicted coding protein => than 300 aa and conserved TIRs as potential autonomous copies; ii) presenting potential coding protein <= than 300 and conserved TIRs as potential non-autonomous copies; iii) copies with a missing TIR but with ORFs => than 300 aa as potential coding copies and iv) elements with a missing TIR and ORFs <= 300 aa as defective elements (Figure S1).

### 3.3.2 Nucleotide Distance and Phylogenetic Analysis

In order to estimate the interspecies distance of TEs we reconstructed the a majority consensus ancestral copy of each TE in each genome (Okonechnikov et al. 2012). Kimura 2 parameters (K2P) distance between all copies and its corresponding consensus sequence was estimated with distmat script from the Emboss package (Emboss distmat) and histogram distribution plotted with ggplot2 (Wickham 2009) in the R environment (R Development Core Team 2016). Dating between *galluhop* consensus elements was performed according to the equation  $T = k/2r$  (Graur and Li 2000). T represent the divergence time between TEs, k is the divergence value between TE consensus and r is the average evolutionary rate from birds genomes ( $\sim 1.9 \times 10^{-3}$  substitutions per site per million years Zhang et al.(2014).

Alignment of reconstructed *galluhop* coding region (almost complete ORF and partial for those who only presented MITEs) from all 6 species that presented *galluhop* homologs were performed using a previously published transposase alignment covering most of the *mariner* subfamilies (Wallau et al. 2014) using MAFFT v.7 (Kato and Standley 2013).

Phylogenetic reconstruction was performed maximum likelihood using PHYML (Guindon and Gascuel 2003) and branch support was evaluated by SH-like support (Anisimova and Gascuel 2006).

## 3.4 Results

### 3.4.1 *galluhop* homologs in avian genomes



Six out of 72 avian genomes analyzed presented *galluhop*-like sequences (Table 1). Five of those are related species from the Galliformes order that diverged at least 46 Mya from each other (CI: 37 – 55 Mya - (Hedges et al. 2015)): *Colinus virginianus*, *Coturnix japonica*, *Lyrurus tetrrix*, *Gallus gallus* e *Meleagris gallopavo*. We also identified *galluhop* homolog sequences in *Buceros rhinoceros* species from the Bucerotiformes order which diverged from Galiformes 98 Mya (CI: 92.1-104.0 Mya - (Hedges et al. 2015)).

Those elements presented a high copy number in both *G. gallus* and *M. gallopavo* genomes, 10158 and 8317 respectively. The remaining four species presented a lower copy number varying from 5 to 96 copies (Table 1). Among all copies analyzed no potential autonomous or coding copies were found (Table 1). Four out of 5 Galliformes species (*G. gallus*, *M. gallopavo*, *C. virginianus* and *L. tetrrix*) presented elements with similar size to the reference *galluhop* element deposited in Repbase (around 1300 bp - Table 1) although most of them presented two 12 bp insertions that hindered any transposase to be codified (Figure 1). The remaining Galliformes species, *C. japonica*, presented only 5 short elements of 550 bp and conserved TIRs resembling MITEs. Lastly, *B. rhinoceros* presented 14 copies of 575 bp but with conserved TIRs and subterminal regions of the elements (Figure 1 and S2). Most of all *galluhop*-like sequences found presented both imperfect TIRs (Figure S2) and target site duplication (TSD) TA characteristic of *mariner* elements (Figure S3).

#### **3.4.2 *galluhop* intra and interspecies dynamics**

Intragenomic divergence between each copy and its corresponding ancestral consensus sequence was performed in order to evaluate the amplification dynamics in each genome bearing *galluhop*-like sequences except for one Galliformes species *C. japonica* which presented only 4 copies. As can be seen in Figure 2A and B all Galliformes species presented a wide distribution of elements ages suggesting that *galluhop* elements are ancient parasites of those genomes (from 0.10 to 0.65 K2P). However, the only *Buceriformes* species, *B. rhinoceros*, presented a much younger elements distribution ranging from 0.10 to 0.16 (Figure 2B).

Consensus elements from each genome were built by majority rule and its Kimura 2 parameters distance was estimated. Among all Galliformes consensus sequences the distance at the nucleotide level varied from 0.0382 to 0.1654 (Table 2). The *B. rhinoceros* consensus presented a K2P distance of 0.0571 to 0.1843 being the lowest distance comparison with *L. tetricus* consensus (Table 2). Moreover, comparing the entire mitochondrial genome of both species an K2P distance of 0.2659 was found.

### 3.4.3 *Mariner* family Phylogenetic Reconstruction

In order to establish the phylogenetic positioning of *galluhop*-like sequences we performed phylogenetic reconstruction using a maximum likelihood approach (Okonechnikov et al. 2012) using the most complete copy recovered from each genome in which we were able to reconstruct transposase aminoacid sequences. Then we aligned all sequences with a dataset comprising all major *mariner* subfamilies available from (Wallau et al. 2014). Our results shows that *galluhop*-like

sequences form a strongly supported new clade in the *mariner* family tree (Figure 3) being an basal branch of *mauritiana*, *mellifera* and other small *mariner* subfamilies.

### 3.5 Discussion

Transposable elements evolution usually presents complex patterns as patchy distribution inside taxa associated with high similarity of TEs in host species that diverged long time ago. The presence of such patterns can be explained by an exchanged of TEs by those species, which characterizes a phenomenon known as HTT. HTT events have been reported all over the eukaryote tree of life in the last years (around 577) and several of those were reported among vertebrate species (Dotto et al. 2015). For instance, the SPIN transposon was found in 5 distantly related mammalian species as well as in a African frog and a lizard presenting high similarity and patch distribution (Gilbert et al. 2009). Although those recent findings in vertebrates, only two HTT event were reported involving Avian clade and parasitic Nematodes (Suh et al. 2016).

Our results are showing that *galluhop* homolog presents a patchy distribution in the avian tree, being found in only 5 Galliformes species (*C. virginianus*, *C. japonica*, *L. tetrrix*, *G. gallus* and *M. gallopavo*) and *B. rhinoceros*, a species from the Neoaves – Buceriformes order. Those two avian orders diverged around 85-98 Mya in the end of the Cretaceous period (Jarvis et al. 2014; Hedges et al. 2015). In addition, we found that the consensus sequences built from copies of each genome presented a lower nucleotide divergence than the vertically transmitted mitochondrial genome. A similar or higher TE-mtDNA distance is expected if TEs were evolving by vertical transfer. On the other hand, a lower TE distance compared with host genes

distance is expected if a horizontal transfer took place. Patchy distribution, high similarity of *galluhop* sequences in distantly related species associated with lower TEs distance compared with host genes suggest that a HTT event took place between some species of the Galliformes order (probably *L. tetrrix*) and *Buceros rhinoceros* ancestor. Another supporting evidence is the intragenomic dynamics of those elements showing that *B. rhinoceros* copies are younger than copies found in *Galliformes* species suggesting that the most probable transfer direction was from *Galliformes* to *B. rhinoceros*.

Those two species currently presents a distinct distribution area: while *L. tetrrix* is restricted to Northern Eurasia as Swiss-Italian-French Alps until Scandinavia, Estonia and Russia, *B. rhinoceros* presents a Southeast distribution range in countries like Borneo, Singapore, Malaysia and Thailand (Collar et al. 2001; Storch 2007) suggesting such HTT event occurred in the ancestral species that may shared the same environment (Jarvis et al. 2014).

With the pass of evolutionary time TEs evolving at a neutral rate are expected to degenerate. Therefore, if an ancient invasion of *B. rhinoceros* genome occurred we can expect that some degree of degeneration would appears in current TE copies. It is in agreement with the absence of potential coding *galluhop* elements in those two genomes. In addition, when we used the (Zhang et al. 2014) background evolutionary rate of  $\sim 1.9 \times 10^{-3}$  substitutions per site per million years between *L. tetrrix* and *B. rhinoceros* consensus elements we could estimate the HTT date probably took place at the end of the Cenozoic era around 15.02 Mya (Figure. 4). Altogether our data supports an ancient HTT event which probably took place between the ancestral of Galliformes species and the *B. rhinoceros* ancestor (Figure. 4).

Phylogenetic analysis using all *galluhop*-like consensus sequences and several sequences from characterized *mariner* subfamilies indicates that *galluhop*-like elements compose a new *mariner* subfamily which we called *Gallus* (Figure 3). TEs from the *Gallus* family emerged in the ancestral of Galliformes order (around 55-65 Mya (Jarvis et al. 2014) increasing its copy number particularly in *G. gallus* and *M. gallopavo* genomes. Only non-autonomous copies of the *Gallus* subfamily were found presenting several mutations, multiple stop codons and changes in the element reading frame (Figure 1). However, we also found a large number of short non-autonomous elements (around 500-600 bp) which presented conserved 5' and 3' region of the element, including TIRs (Figure 1 and S2), but presented a large deletion compared with the full length consensus element (Figure 1). Such shorter elements presented all MITEs characteristics (Fattash et al. 2013) and amplified successfully in *G. gallus* and *M. gallopavo* composing the large majority of *galluhop* copies found in those genomes (97.7% in *G. gallus* and 98.4% in *M. gallopavo* genomes). *C. virginianus* and *L. tetrrix* also presented amplification of MITEs in a smaller scale and *C. japonica* and *B. rhinocerus* only presented MITEs elements and no trace of its possible autonomous counterparts (Figure 1). All these findings suggests that MITEs originated frequently in this new *mariner* subfamily which probably defined the fate of those elements leading to the elements lineages dead end in all avian genomes studied. Such view is in agreement that the emergence of superparasites as MITEs can lead TE subfamilies decay and disappear through time (Naito et al. 2006; Fattash et al. 2013)

Our results bring new light on the evolution of the *mariner* family of class II transposons describing a new subfamily called *Gallus*. The distribution pattern and similarity analysis suggests the first HTT event between avian genomes and

highlights successful amplification of MITEs of this subfamily in some avian genomes. However, further analysis on other Bucerotiformes related species are needed in order to evaluate if the *galluhop*-like sequences distribution in this taxa support or not this HTT event

### 3.6 Acknowledgments

The authors are grateful to Dr. Elgion Loreto for its inspiring comments on the manuscript and Thays de Oliveira assistance with the figures.

### 3.7 References

- Anisimova M and Gascuel O (2006) Approximate Likelihood-Ratio Test for Branches : A Fast , Accurate , and Powerful Alternative. *Syst Biol* 55:539–552. doi: 10.1080/10635150600755453
- Cáceres M, Puig M and Ruiz A (2001) Molecular characterization of two natural hotspots in the *Drosophila buzzatii* genome induced by transposon insertions. *Genome Res* 11:1353–1364. doi: 10.1101/gr.174001
- Carareto CMA, Monteiro-Vitorello CB and Sluys M-A Van (2015) Elementos de transposição: diversidade, evolução, aplicações e impacto nos genomas dos seres vivos, 1st ed. FIOCRUZ, Rio de Janeiro
- Casola C, Hucks D and Feschotte C (2008) Convergent domestication of pogo-like transposases into centromere-binding proteins in fission yeast and mammals. *Mol Biol Evol* 25:29–41. doi: 10.1038/jid.2014.371
- Chicken I and Sequencing G (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432:695–716. doi: 10.1038/nature03394
- Collar NJ, Andreev A, Chan S, Crosby MJ, Subramanya S and Tobias JA (2001) Threatened Birds of Asia: The BirdLife International Red Data Book. Natural History Book\_Service Ltd, Cambridg
- Crooks G, Hon G, Chandonia J and Brenner S (2004) NCBI GenBank FTP Site\nWebLogo: a sequence logo generator. *Genome Res* 14:1188–1190. doi: 10.1101/gr.849004.1
- Dalloul R a., Long J a., Zimin A V., Aslam L, Beal K, Blomberg LA, Bouffard P, Burt DW, Crasta O, Crooijmans RPM a et al. (2010) Multi-platform next-generation sequencing of the domestic Turkey (*Meleagris gallopavo*): Genome assembly and analysis. *PLoS Biol.* doi: 10.1371/journal.pbio.1000475
- Deprá M, Ludwig A, Valente VL and Loreto EL (2012) Mar, a MITE family of hAT transposons in *Drosophila*. *Mob DNA* 3:13. doi: 10.1186/1759-8753-3-13
- Dotto BR, Carvalho EL, Silva AF, Duarte Silva LF, Pinto PM, Ortiz MF and Wallau GL (2015) HTT-DB: Horizontally transferred transposable elements database. *Bioinformatics* 31:2915–2917. doi: 10.1093/bioinformatics/btv281

Fattash I, Rooke R, Wong A, Hui C, Luu T, Bhardwaj P and Yang G (2013) Miniature inverted-repeat transposable elements: discovery, distribution, and activity 1. *486:475–486*.

Feschotte C (2008) Transposable elements and the evolution of regulatory networks. *Nat Genet 9:397–405*. doi: 10.1038/nrg2337

Finnegan DJ (1989) Eukaryotic transposable elements and genome evolution. *Trends Genet 5:103–107*. doi: 10.1016/0168-9525(89)90039-5

Gilbert C, Pace JK and Feschotte C (2009) Horizontal SPINning of transposons. *Commun Integr Biol 2:117–119*. doi: 10.1073/pnas.0806548105. www.landesbioscience.com

González J and Petrov D (2009) MITEs — The Ultimate Parasites. *Science (80- ) 325:1352–1353*. doi: 10.1126/science.1179556

Graur D and Li W (2000) *Fundamentals of molecular evolution*, 2nd ed. Sinauer Associates, Sunderland

Guindon S and Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol 52:696–704*. doi: 10.1080/10635150390235520

Hedges SB, Marin J, Suleski M, Paymer M and Kumar S (2015) Tree of life reveals clock-like speciation and diversification. *Mol Biol Evol 32:835–845*. doi: 10.1093/molbev/msv037

Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, Ho SYW, Faircloth BC, Nabholz B, Howard JT et al. (2014) Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science (80- ) 346:1126–1138*. doi: 10.1126/science.1251385

Jurka J, Kapitonov V V., Pavlicek A, Klonowski P, Kohany O and Walichiewicz J (2005) Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res 110:462–467*. doi: 10.1159/000084979

Kapitonov V V and Jurka J (2008) A universal classification of eukaryotic transposable elements implemented in Repbase. *Nat Rev Genet 9:411–412*; author reply 414. doi: 10.1038/nrg2165-c2

Kapusta A and Suh A (2016) Evolution of bird genomes—a transposon’s-eye view. *Ann N Y Acad Sci 1–22*. doi: 10.1111/nyas.13295

Katoh K and Standley DM (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol 30:772–780*. doi: 10.1093/molbev/mst010



- Koepfli K, Paten B, Scientists G 10K C of and O'Brien SJ (2015) The Genome 10K Project: a way forward. *Annu Rev Anim Biosci* 3:57–111. doi: 10.1146/annurev-animal-090414-014900
- Kordis D (2010) Transposable elements in reptilian and avian (Sauropsida) genomes. *Cytogenet Genome Res* 127:94–111. doi: 10.1159/000294999
- Le Rouzic A and Capy P (2006) Population genetics models of competition between transposable element subfamilies. *Genetics* 174:785–793. doi: 10.1534/genetics.105.052241
- McVean G (2010) What drives recombination hotspots to repeat DNA in humans? *Philos Trans R Soc* 365:1213–8. doi: 10.1098/rstb.2009.0299
- Naito K, Cho E, Yang G, Campbell M a, Yano K, Okumoto Y, Tanisaka T and Wessler SR (2006) Dramatic amplification of a rice transposable element during recent domestication. *Proc Natl Acad Sci U S A* 103:17620–17625. doi: 10.1073/pnas.0605421103
- Okonechnikov K, Golosova O, Fursov M, Varlamov A, Vaskin Y, Efremov I, German Grehov OG, Kandrov D, Rasputin K, Syabro M et al. (2012) Unipro UGENE: A unified bioinformatics toolkit. *Bioinformatics* 28:1166–1167. doi: 10.1093/bioinformatics/bts091
- Oliveira SG, Bao W, Martins C and Jurka J (2012) Horizontal transfers of Mariner transposons between mammals and insects. *Mob DNA* 3:14. doi: 10.1186/1759-8753-3-14
- Organ CL, Shedlock AM, Meade A, Pagel M and Edwards S V. (2007) Origin of avian genome size and structure in non-avian dinosaurs. *Nature* 446:180–184. doi: 10.1038/nature05621
- Osborne PW, Luke GN, Holland PWH and Ferrier DEK (2006) Identification and characterisation of five novel Miniature Inverted-repeat Transposable Elements (MITEs) in amphioxus (*Branchiostoma floridae*). *Int J Biol Sci* 2:54–60.
- R Development Core Team (2016) R: A language and environment for statistical computing.
- Rouault JD, Casse N, Chénais B, Hua-Van A, Filée J and Capy P (2009) Automatic classification within families of transposable elements: Application to the mariner Family. *Gene* 448:227–232. doi: 10.1016/j.gene.2009.08.009
- Schaack S, Gilbert C and Feschotte C (2010) Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. *Trends Ecol Evol* 25:537–546. doi: 10.1016/j.tree.2010.06.001.Promiscuous

Storch I (2007) Status Survey and Conservation Action Plan 2006-2010. IUCN-SSC-Birdlife-WPA Grouse Specialist Group, Switzerland and Fordingbridge

Suh A, C. Witt C, Menger J, R.Sadanandan K, Podsiadlowski L, Gerth M, Weigert A, McGuire JA, Mudge J, Edwards S V. et al. (2016) Ancient horizontal transfers of retrotransposons between birds and ancestors of human pathogenic nematodes. *Nat Commun* 1–9. doi: 10.1038/pj.2016.37

Volff JN (2006) Turning junk into gold: Domestication of transposable elements and the creation of new genes in eukaryotes. *BioEssays* 28:913–922. doi: 10.1002/bies.20452

Wallau G, Capy P, Loreto E and Hua-Van A (2014) Genomic landscape and evolutionary dynamics of mariner transposable elements within the *Drosophila* genus 5. 1–19. doi: 10.1186/1471-2164-15-727

Wallau GL, Hua-Van A, Capy P and Loreto ELS (2011) The evolutionary history of mariner-like elements in Neotropical drosophilids. *Genetica* 139:327–338. doi: 10.1007/s10709-011-9552-6

Wallau GL, Ortiz MF and Loreto ELS (2012) Horizontal transposon transfer in eukarya: Detection, bias, and perspectives. *Genome Biol Evol* 4:689–699. doi: 10.1093/gbe/evs055

Wicker T, Robertson JS, Schulze SR, Feltus FA, Magrini V, Morrison JA, Mardis ER, Wilson RK, Peterson DG, Paterson AH et al. (2005) The repetitive landscape of the chicken genome. *Genome Res* 15:126–136. doi: 10.1101/gr.2438004

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O et al. (2007) A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* 8:973–982. doi: 10.1038/nrg2165-c4

Wickham H (2009) *ggplot2*, Elegant graphics for data analysis.

Yang G, Nagel DH, Feschotte C, Hancock CN and Wessler SR (2009) Tuned for transposition: molecular determinants underlying the hyperactivity of a Stowaway MITE. *Science* 325:1391–1394. doi: 10.1126/science.1175688

Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, Storz JF, Antunes A, Greenwold MJ, Meredith RW et al. (2014) Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* (80- ) 346:1311–1321. doi: 10.1126/science.1251385

Zhang Q and Edwards S V. (2012) The evolution of intron size in amniotes: A role for powered flight? *Genome Biol Evol* 4:1033–1043. doi: 10.1093/gbe/evs070

### 3.8 Internet Resources

Emboss distmat,  
<http://emboss.sourceforge.net/apps/release/6.6/emboss/apps/distmat.html> (February, 2017)

Emboss *getorf*, <http://emboss.sourceforge.net/apps/cvs/emboss/apps/getorf.html>, (February, 2017)

NCBI birds genomes, <https://www.ncbi.nlm.nih.gov/genome/browse/> (October 2016)

Weblogo, <http://weblogo.berkeley.edu/logo.cgi> (November 2016)

Rebase *galluhop* sequence, <http://www.girinst.org/rebase/update/browse.php> (October 2016).

Figure 1: Schematic representation of the reconstructed of *galluhop* copies compared to the *galluhop* consensus. Regions terminal inverted repeats indicated in red, transposase coding region shown light gray and Insertion region in dark gray. Galliformes order: four genomes (*G. gallus*, *M. gallopavo*, *C. virginianus* and *L. tetrrix*) showed potential complete partners although there is no potential coding copies, and *C. japonica* presented short elements. Bucerotiformes order: *B. rhinoceros* presented only short elements.

Figure 2: Kimura 2 Parameters distance distribution of each copy to its corresponding consensus sequence. A) Relative intragenomic dating of copies found in *G. gallus* and *M. gallopavo*. B) Relative intragenomic dating of copies found in *B. rhinoceros*, *C. virginianus* and *L. tetrrix*.

Figure 3: Phylogeny of *mariner*-like transposases. Phylogeny of *mariner*-like transposases, by maximum likelihood using PHYML (Guindon and Gascuel 2003). Clade colors denote the different subfamilies of *mariner* family, indicated to the left of the tree. In gray the new subfamily: Gallus.

Figure 4: Horizontal Transfer hypothesis of *galluhop* elements. Chronogram TENT avian tree from Jarvis et al 2014 with the addition of probable *L. tetrrix* positioning and split data following TimeTree data (Hedges et al. 2015) Red branches denotes the evolutionary hypothesis of *Gallus* subfamily vertical evolution in the *Galliformes* and horizontal transfer from *L. tetrrix* to *B. rhinoceros* ancestrals. X bar bellow the tree denotes the time in million of years.

Figure S1: Experimental design procedure showing steps of the analysis. *galluhop* homologous sequences were found in six from 72 genomes analyze. We analyzed the functional, structural characteristics and phylogenetic reconstruction of the putative transposases.

Figure S2: Graphical representation of the conservation of terminal inverted repeats (TIRs) 5' e 3' *galluhop* element in the six genomes, using WebLogo (Crooks et al. 2004). Galliformes order: *G. gallus* (A – A'), *M. gallopavo* (B – B'), *C. virginianus* (C – C'), *L. tetrrix* (D – D') and *C. japonica* (E – E'). Bucerotiformes order: *B. rhinoceros* (F – F').

Figure S3: Graphical representation of the conservation of target site duplication (TSDs) 5' e 3' *galluhop* element in the six genomes, using WebLogo (Crooks et al. 2004). Galliformes order: *G. gallus* (A – A'), *M. gallopavo* (B – B'), *C. virginianus* (C – C'), *L. tetrrix* (D – D') and *C. japonica* (E – E'). Bucerotiformes order: *B. rhinoceros* (F – F').

Supplementary material - the following online material is available for this article:

Table S1 – GenBank access numbers birds genomes

Figure S1 - Experimental design procedure showing steps of the analysis

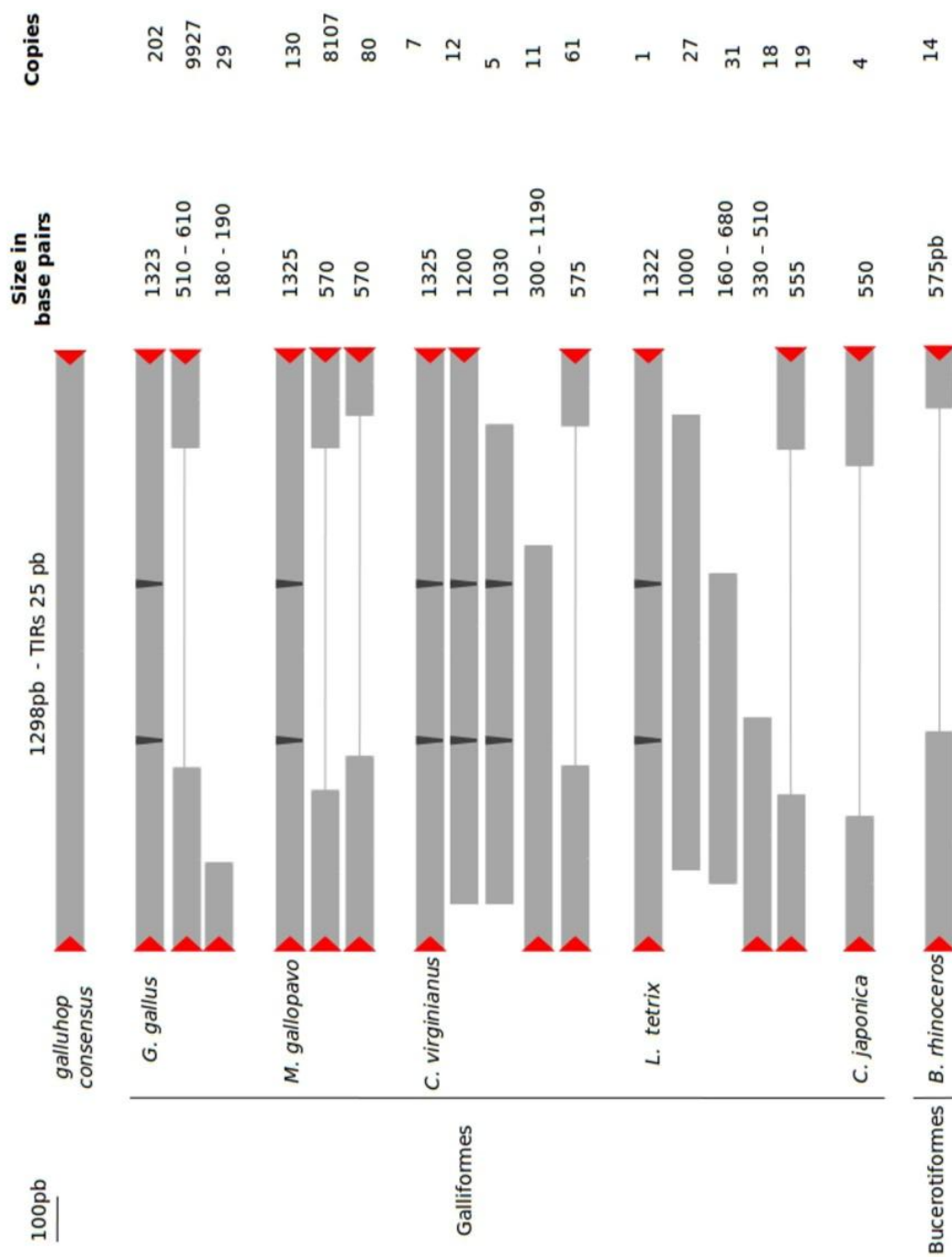
Figure S2 – Graphical representation of the TIRs conservation

Figure S3 – Graphical representation of the TSDs conservation

This material is available as part of the online article from

<http://www.scielo.br/gmb>

**Figure 1- Schematic representation of the reconstructed of galluhop copies compared to the *galluhop* consensus**



**Figure 2- Kimura 2 Parameters distance distribution of each copy to its corresponding consensus sequence.**

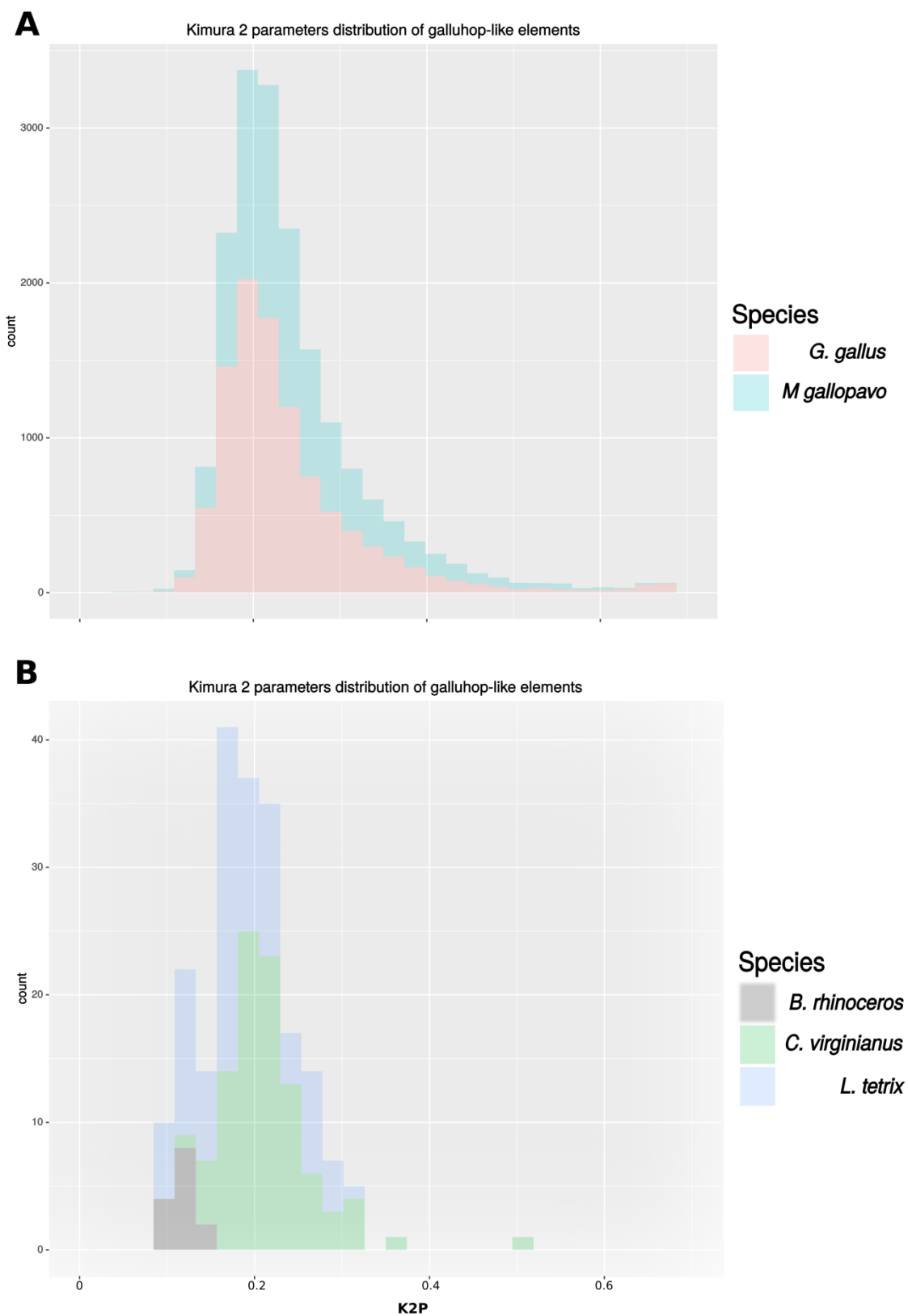




Figure 3- Phylogeny of mariner-like transposases

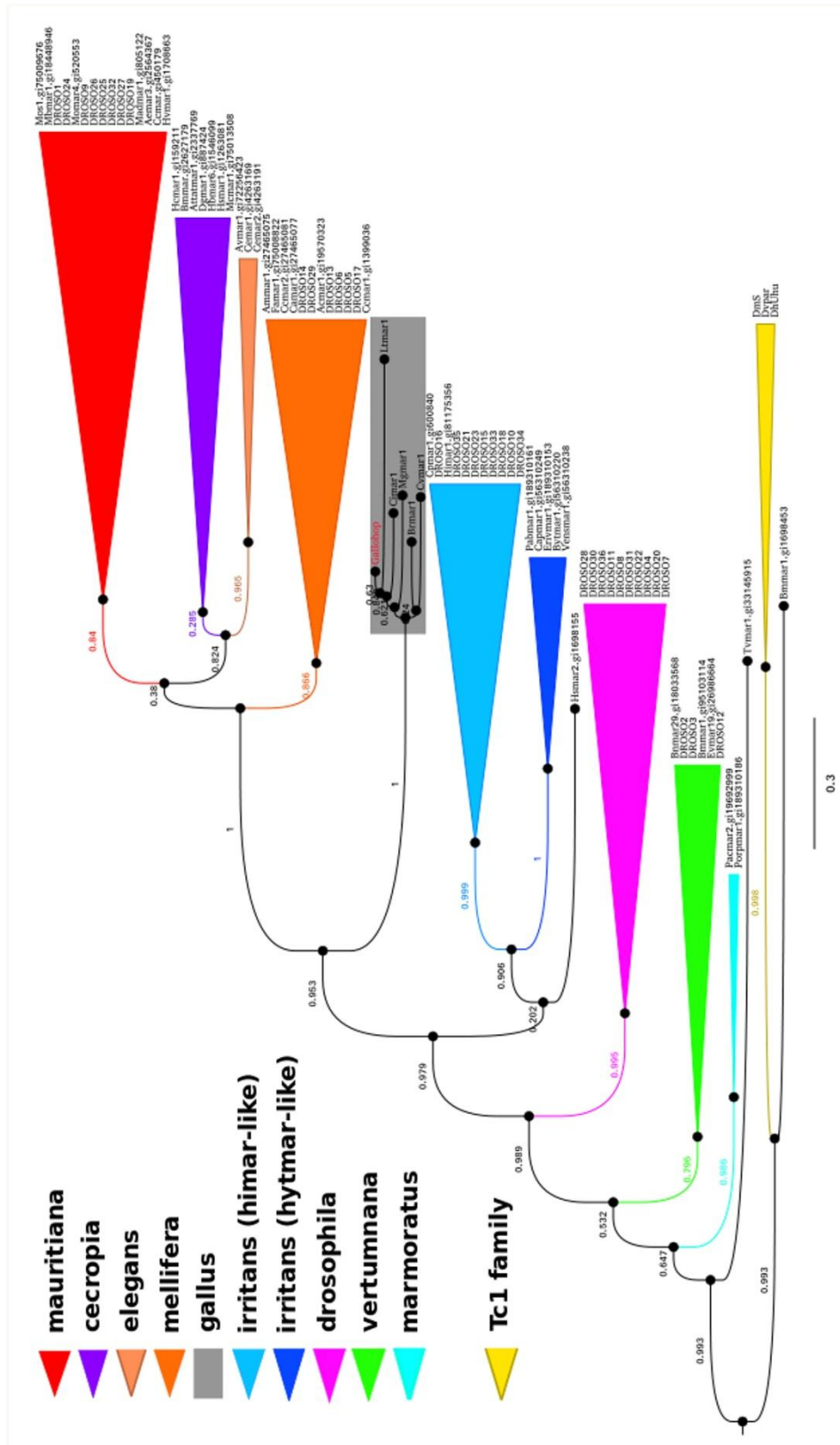
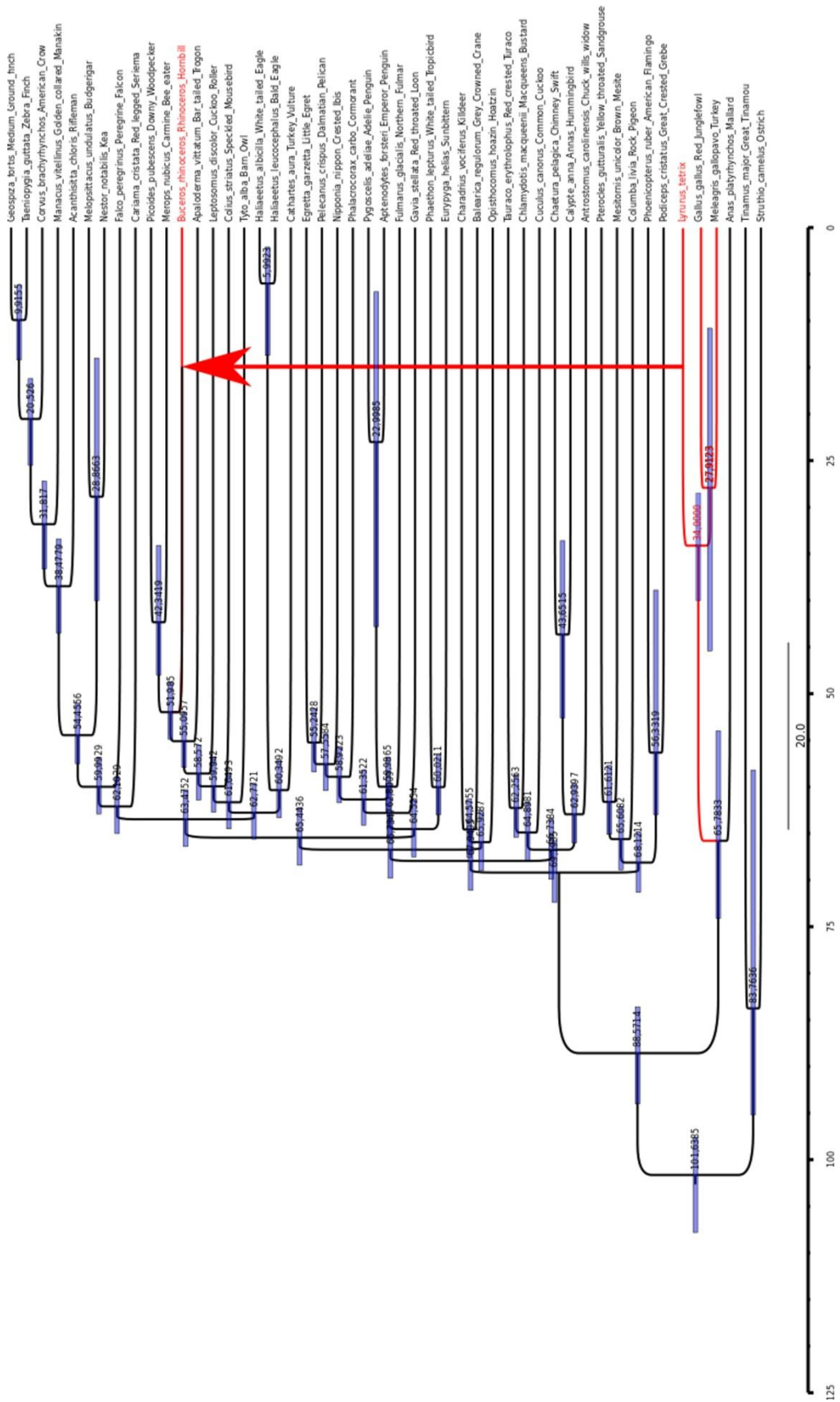


Figure 4- Horizontal Transfer hypothesis of *galluhop* elements



**Table 1- Avian genomes with *galluhop* and characteristics of copies**

Order	Species	Genome Size (Mb)	Nº de copies	*ORFs	Full length elements (around 1200 – 1300 pb)	Non-autonomous elements (around 500-610 pb)	Partial Elements (around 160-1200)
Galliformes	<i>G. gallus</i>	1046.93	10158	N	202	9927	29
Galliformes	<i>M. gallopavo</i>	1061.82	8317	N	130	8187	0
Galliformes	<i>C. virginianus</i>	1171.86	96	N	7	61	28
Galliformes	<i>L. tetricus</i>	657.025	96	N	1	19	76
Galliformes	<i>C. japonica</i>	531.96	4	N	0	4	0
Bucerotiformes	<i>B. rhinoceros</i>	1065.78	14	N	0	14	0

\* No ORFs were found

Table 2- Identity of the *galluhop* consensus sequences in the genomes of birds

Species	<i>B.</i> <i>rhinoceros</i>	<i>C.</i> <i>virginianus</i>	<i>C.</i> <i>japonica</i>	<i>G.</i> <i>gallus</i>	<i>L.</i> <i>tetrix</i>	<i>M.</i> <i>gallopavo</i>
<i>B. rhinoceros</i>		24.23	48.5	19.86	15.02	20.97
<i>C. virginianus</i>	0.0921		43.52	13.65	13.65	14.07
<i>C. japonica</i>	0.1843	0.1654		44.5	37.10	43.68
<i>G. gallus</i>	0.0755	0.0519	0.1691		10.94	5.92
<i>L. tetrix</i>	0.0571	0.0519	0.141	0.0416		10.05
<i>M. gallopavo</i>	0.0797	0.0535	0.166	0.0225	0.0382	

Figure S1- Experimental design procedure showing steps of the analysis

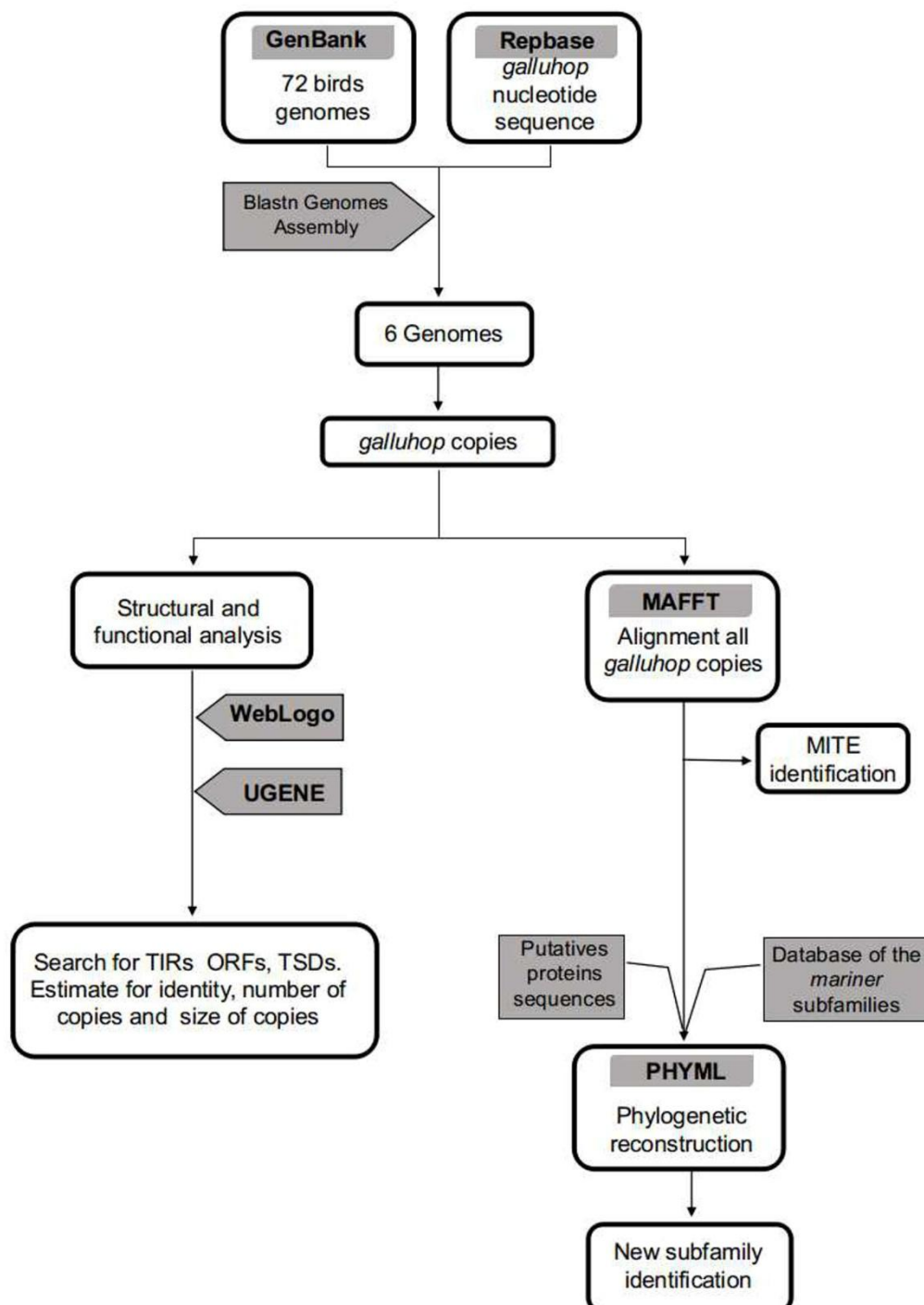


Figure S2- Graphical representation of the conservation of terminal inverted repeats (TIRs) 5' e 3' *galluhop* element in the six genomes

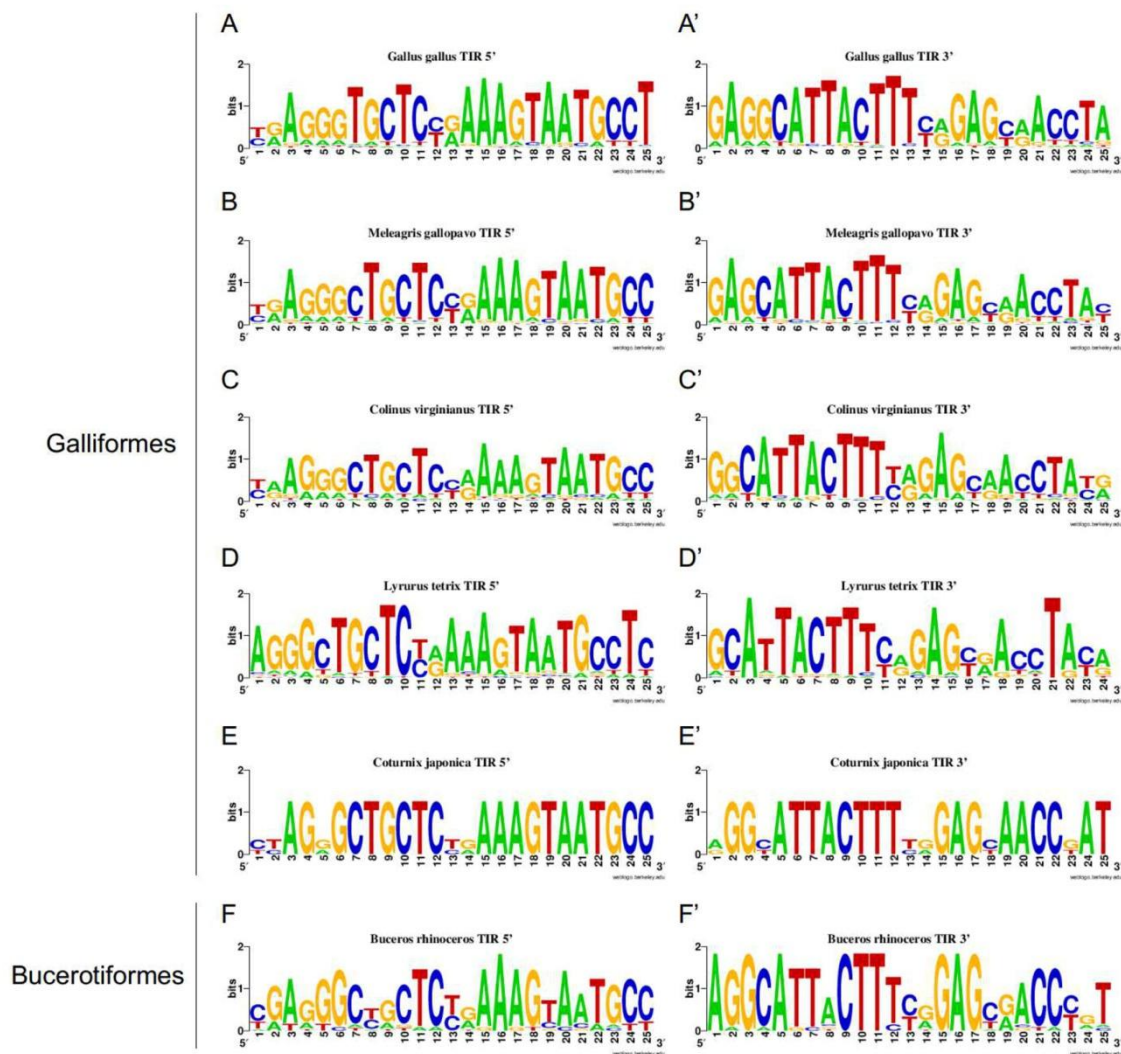
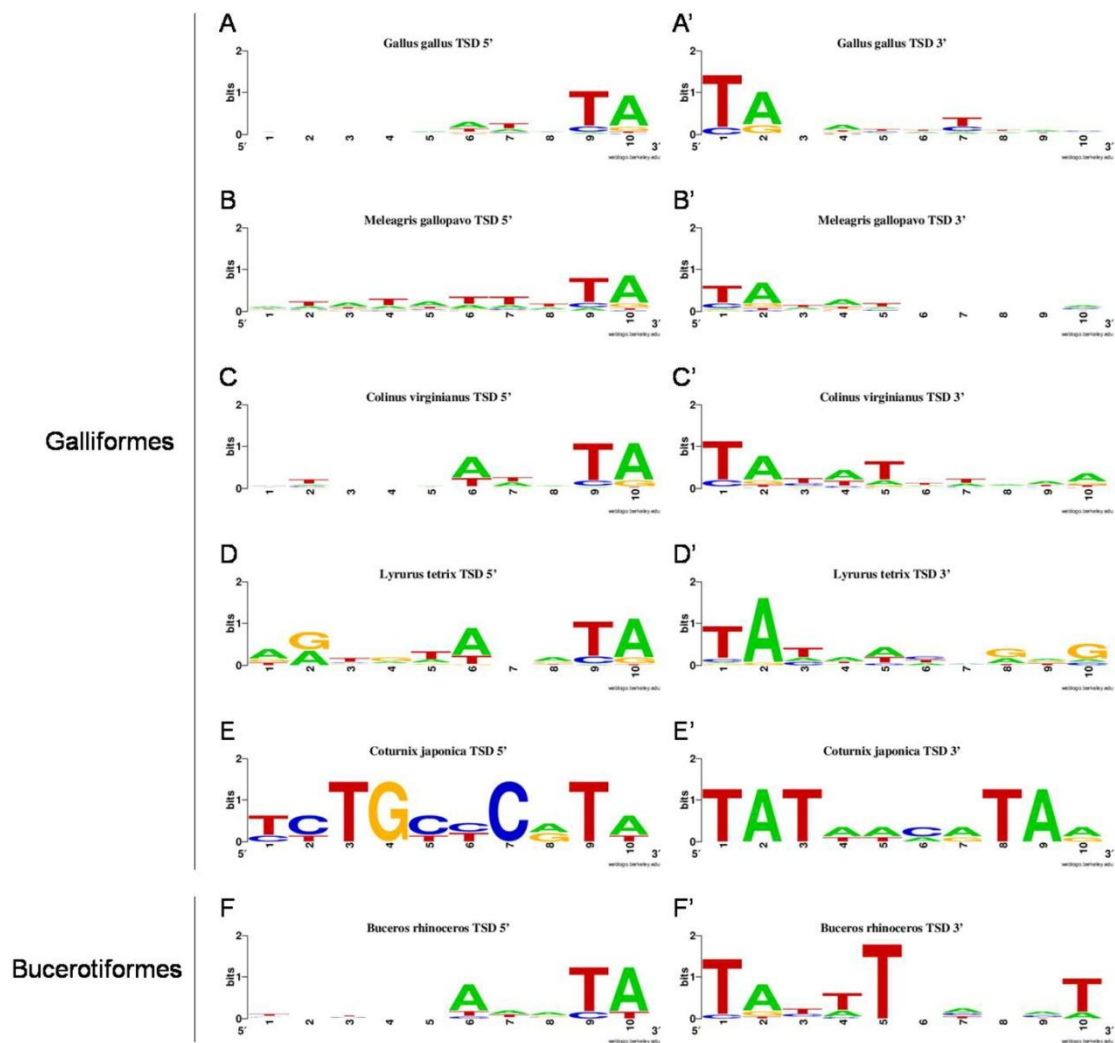


Figure S3- Graphical representation of the conservation of target site duplication (TSDs) 5' e 3' *galluhop* element in the six genomes



**Table S1: GenBank access numbers birds genomes**

<b>Ordem/ espécie</b>	<b>ID</b>	<b>GenBank assembly accession</b>	<b>RefSeq assembly accession</b>
<b>Anseriformes</b>			
<i>Anas platyrhynchos</i>	2793	GCA_000355885.1 (latest)	GCF_000355885.1 (latest)
<i>Anser cygnoides</i>	31397	GCA_000971095.1 (latest)	GCF_000971095.1 (latest)
<b>Apodiformes</b>			
<i>Chaetura pelagica</i>	33278	GCA_000747805.1 (latest)	GCF_000747805.1 (latest)
<i>Pterocles gutturalis</i>	32063	GCA_000699245.1 (latest)	GCF_000699245.1 (latest)
<b>Apterygiformes</b>			
<i>Apteryx australix</i>	10907	GCA_001039765.1 (replaced)	GCF_001039765.1 (latest)
<b>Bucerotiformes</b>			
<i>Buceros rhinoceros</i>	32403	GCA_000710305.1 (latest)	GCF_000710305.1 (latest)
<b>Caprimulgiformes</b>			
<i>Caprimulgus carolinensis</i>	32067	GCA_000700745.1 (latest)	GCF_000700745.1 (latest)
<b>Charadriiformes</b>			
<i>Calidris pugnax</i>	41082	GCA_001431845.1 (latest)	GCF_001431845.1 (latest)
<i>Charadrius vociferus</i>	32124	GCA_000708025.2 (latest)	GCF_000708025.1 (latest)
<b>Ciconiiformes</b>			
<i>Cathartes aura</i>	7839	GCA_000699945.1 (latest)	n/a
<b>Coliiformes</b>			
<i>Colius striatus</i>	31969	GCA_000690715.1 (latest)	GCF_000690715.1 (latest)
<b>Columbiformes</b>			
<i>Columba livia</i>	10719	GCA_000337935.1 (latest)	GCF_000337935.1 (latest)
<b>Coraciiformes</b>			
<i>Leptosomus discolor</i>	32000	GCA_000691785.1 (latest)	GCF_000691785.1 (latest)
<i>Merops nubicus</i>	31978	GCA_000691845.1 (latest)	GCF_000691845.1 (latest)
<b>Cuculiformes</b>			
<i>Cuculus canorus</i>	32170	GCA_000709325.1 (latest)	GCF_000709325.1 (latest)
<b>Falconiformes</b>			
<i>Aquila chrysaetos</i>	32031	GCA_000766835.1 (latest)	GCF_000766835.1 (latest)
<i>Falco cherrug</i>	14103	GCA_000337975.1 (latest)	GCF_000337975.1 (latest)
<i>Falco peregrinus</i>	132	GCA_000337955.1 (latest)	GCF_000337955.1 (latest)
<i>Haliaeetus albicilla</i>	31975	GCA_000691405.1 (latest)	GCF_000691405.1 (latest)
<i>Haliaeetus leucocephalus</i>	32665	GCA_000737465.1 (latest)	GCF_000737465.1 (latest)
<b>Galliformes</b>			
<i>Colinus virginianus</i>	12817	GCA_000599465.1 (latest)	n/a
<i>Coturnix japonica</i>	113	GCA_000511605.1 (latest)	n/a
<i>Gallus gallus</i>	111	GCA_000002315.3 (latest)	GCF_000002315.4 (latest)
<i>Lyrurus tetrix</i>	31278	GCA_000586395.1 (latest)	n/a
<i>Meleagris gallopavo</i>	112	GCA_000146605.3 (latest)	GCF_000146605.2 (latest)
<b>Gaviiformes</b>			
<i>Gavia stellata</i>	7634	GCA_000690875.1 (latest)	GCF_000690875.1 (latest)
<b>Gruiformes</b>			
<i>Balearica regulorum</i>	17144	GCA_000709895.1 (latest)	GCF_000709895.1 (latest)
<i>Cariama cristata</i>	31967	GCA_000690535.1 (latest)	GCF_000690535.1 (latest)



<b><i>Chlamydotis macqueenii</i></b>	34357	GCA_000695195.1 (latest)	GCF_000695195.1 (latest)
<i>Eurypyga helias</i>	31970	GCA_000690775.1 (latest)	GCF_000690775.1 (latest)
<i>Mesitornis unicolor</i>	32034	GCA_000695765.1 (latest)	GCF_000695765.1 (latest)
<b>Musophagiformes</b>			
<i>Tauraco erythrophus</i>	32247	GCA_000709365.1 (latest)	GCF_000709365.1 (latest)
<b>Opisthocomiformes</b>			
<i>Opisthocomus hoazin</i>	31992	GCA_000692075.1 (latest)	GCF_000692075.1 (latest)
<b>Passeriformes</b>			
<i>Acanthisitta chloris</i>	32002	GCA_000695815.1 (latest)	GCF_000695815.1 (latest)
<i>Corvus brachyrhynchos</i>	32035	GCA_000691975.1 (latest)	GCF_000691975.1 (latest)
<i>Corvus cornix</i>	18230	GCA_000738735.1 (latest)	GCF_000738735.1 (latest)
<i>Ficedula albicollis</i>	11872	GCA_000247815.2 (latest)	GCF_000247815.1 (latest)
<i>Geospiza fortis</i>	13302	GCA_000277835.1 (latest)	GCF_000277835.1 (latest)
<i>Lepidothrix coronata</i>	44392	GCA_001604755.1 (latest)	GCF_001604755.1 (latest)
<i>Manacus vitellinus</i>	31979	GCA_000692015.2 (latest)	GCF_000692015.1 (latest)
<i>Parus major</i>	12863	GCA_001522545.1 (replaced)	GCF_001522545.1 (latest)
<i>Passer domesticus</i>	17653	GCA_001700915.1 (latest)	n/a
<i>Phylloscopus plumbeitarsus</i>	45124	GCA_001655115.1 (latest)	n/a
<i>Phylloscopus trochiloides</i>	45123	GCA_001655095.1 (latest)	n/a
<i>Pseudopodoces humilis</i>	15114	GCA_000331425.1 (latest)	GCF_000331425.1 (latest)
<i>Serinus canaria</i>	11251	GCA_000534875.1 (latest)	GCF_000534875.1 (latest)
<i>Setophaga coronata</i>	46404	GCA_001746935.1 (latest)	n/a
<i>Sturnus vulgaris</i>	41647	GCA_001447265.1 (latest)	GCF_001447265.1 (latest)
<i>Taeniopygia guttata</i>	367	GCA_000151805.2 (latest)	GCF_000151805.1 (latest)
<i>Zonotrichia albicollis</i>	17510	GCA_000385455.1 (latest)	GCF_000385455.1 (latest)
<i>Zosterops lateralis</i>	40104	GCA_001281735.1 (latest)	n/a
<b>Pelecaniformes</b>			
<i>Egretta garzetta</i>	31706	GCA_000687185.1 (latest)	GCF_000687185.1 (latest)
<i>Nipponia nippon</i>	7967	GCA_000708225.1 (latest)	GCF_000708225.1 (latest)
<i>Pelecanus crispus</i>	31933	GCA_000687375.1 (latest)	GCF_000687375.1 (latest)
<i>Phaethon lepturus</i>	31929	GCA_000687285.1 (latest)	GCF_000687285.1 (latest)
<i>Phalacrocorax carbo</i>	32171	GCA_000708925.1 (latest)	GCF_000708925.1 (latest)
<b>Phoenicopteriformes</b>			
<i>Phoenicopus ruber</i>	31928	GCA_000687265.1 (latest)	n/a
<b>Piciformes</b>			
<i>Picoides pubescens</i>	32059	GCA_000699005.1 (latest)	GCF_000699005.1 (latest)
<b>Podicipediformes</b>			
<i>Podiceps cristatus</i>	7975	GCA_000699545.1 (latest)	n/a
<b>Procellariiformes</b>			
<i>Fulmarus glacialis</i>	31971	GCA_000690835.1 (latest)	GCF_000690835.1 (latest)
<b>Psittaciformes</b>			
<i>Amazona aestiva</i>	40915	GCA_001420675.1 (latest)	n/a
<i>Amazona vittata</i>	15170	GCA_000332375.1 (latest)	n/a
<i>Ara macao</i>	17770	GCA_000400695.1 (latest)	n/a
<i>Melopsittacus undulatus</i>	10765	GCA_000238935.1 (latest)	GCF_000238935.1 (latest)
<i>Nestor notabilis</i>	33272	GCA_000696875.1 (latest)	GCF_000696875.1 (latest)

---

<b>Sphenisciformes</b>			
<i>Aptenodytes forsteri</i>	32061	GCA_000699145.1 (latest)	GCF_000699145.1 (latest)
<i>Pygoscelis adeliae</i>	17559	GCA_000699105.1 (latest)	GCF_000699105.1 (latest)
<b>Strigiformes</b>			
<i>Tyto alba</i>	31927	GCA_000687205.1 (latest)	GCF_000687205.1 (latest)
<b>Struthioniformes</b>			
<i>Struthio camelus</i>	122	GCA_000698965.1 (latest)	GCF_000698965.1 (latest)
<b>Tinamiformes</b>			
<i>Tinamus guttatus</i>	32250	GCA_000705375.2 (latest)	GCF_000705375.1 (latest)
<b>Trochiliformes</b>			
<i>Calypte anna</i>	32060	GCA_000699085.1 (latest)	GCF_000699085.1 (latest)
<b>Trogoniformes</b>			
<i>Apaloderma vittatum</i>	32169	GCA_000703405.1 (latest)	GCF_000703405.1 (latest)

---

## 4 CAPÍTULO 2

### Distribuição cromossômica do elemento transponível CR1-like em pica-paus (Aves Piciformes)

#### 4.1 Introdução

O retrotransposon Chicken repeat 1 (CR1) foi identificado inicialmente no genoma de *Gallus gallus*, atualmente formam uma superfamília de elementos transponíveis (TEs) ubíquos em todos os amniotas (STUMPH et al., 1981; SUH, 2015). Os TEs CR1-like são elementos nucleares interdispersos longos (LINEs), com aproximadamente 4,5 kb quando completos (BURCH; DAVIS; HAAS, 1993). São constituídos por duas ORFs, sendo a ORF1 codificante para uma proteína *gag-like* com motivos *zincfinger-like* e a ORF2 responsável pela codificação da transcriptase reversa e outras enzimas envolvidas na mobilização do TE (BURCH; DAVIS; HAAS, 1993; HAAS et al., 1997, 2001; KAJIKAWA; OHSHIMA; OKADA, 1997; KAPITONOV; JURKA, 2003). Como ocorre em outros LINEs, os TEs CR1-like são altamente truncados na porção 5' UTR e conservados na porção 3' UTR, e também possuem uma sequência microssatélite conservada nessa porção terminal com provável função de reconhecimento para mobilização (SUH et al., 2014; WICKER et al., 2005).

A caracterização das subfamílias de elementos CR1, em aves, é basicamente restrita ao genoma de *G. gallus* (WICKER et al., 2005). Entretanto, devido à “explosão” de genomas aviários recentemente sequenciados, vem ocorrendo também uma massiva identificação de elementos CR1 em outras linhagens de aves, porém, carecendo ainda de análise mais detalhadas quanto as suas caracterizações estruturais e funcionais (KAPUSTA; SUH, 2016; SUH et al., 2011). Nos outros grupos de Sauropsidas, os répteis, são observadas algumas subfamílias diferentes das encontradas em aves, nos mostrando a diversidade desse elemento (SUH et al., 2014).

Os elementos CR1 estão presentes nas maiores linhagens de amniotas sugerindo que foram inseridos no ancestral em comum desse grupo, contudo

evoluíram distintamente em cada linhagem, em mamíferos foram substituídos devido uma massiva expansão de retrotransposon LINE-1 e LINE-2 (KORDIS; LOVSIN; GUBENSEK, 2006). Nas linhagens de lepidossauros (lagartos) e aves estão associados à diminuição do tamanho dos genomas, devido à perda de linhagens antigas de CR1 e assim diminuindo o tamanho do genoma (SUH et al., 2014). No caso das aves, essa contração do genoma esta associada à necessidade de otimizar o metabolismo para o voo (SMITH; GREGORY, 2009).

CR1 é o único TE distribuído por toda a linhagem da classe Aves, possui poucas cópias viáveis para mobilização, em *G. gallus*, 98% são fragmentos curtos com aproximadamente 500 pb da região 3' UTR e a enzima transcriptase reversa (WICKER et al., 2005). Também representam praticamente a totalidade da fração repetitiva dos genomas aviários, aproximadamente 10%, uma porcentagem inferior do que é visto em outros grupos de vertebrados, por exemplo, os mamíferos com aproximadamente 50% do genoma formado por TEs (GIBBS et al., 2004).

Há uma exceção nas aves, na família Picidae (ordem Piciformes), a espécie *Picoides pubescens* (pica-pau) possui uma fração bem mais expressiva de elementos CR1-like, com aproximadamente 22% do seu genoma (JARVIS et al., 2014; ZHANG et al., 2014). Além disso, em outras espécies de pica-paus foram relacionado o tamanho grande dos primeiros pares de macrocromossomos, incluindo o cromossomo sexual Z, com o maior acúmulo de sequências repetitivas (DE OLIVEIRA et al., 2017). Porém, o estudo da fração repetitiva nos genomas das aves, principalmente Piciformes ainda é bastante insipiente. Por essas razões supõe-se que elementos CR1 estão relacionados com as características estruturais dos cromossomos de pica-paus.

O objetivo deste estudo foi conhecer o padrão de distribuição de um elemento CR1-like nos cromossomos de pica-paus, com ênfase nos maiores pares do complemento cariotípico (par 1 e cromossomo sexual Z). E, a partir desses padrões, fazer uma análise comparativa entre as espécies e das mesmas com *Gallus gallus*, espécie com características basais.

## 4.2 Material e Métodos

### 4.2.1 Fluxograma do trabalho

As principais etapas realizadas neste trabalho são apresentadas na figura 1, cujos procedimentos experimentais encontram-se detalhados ao longo do texto.

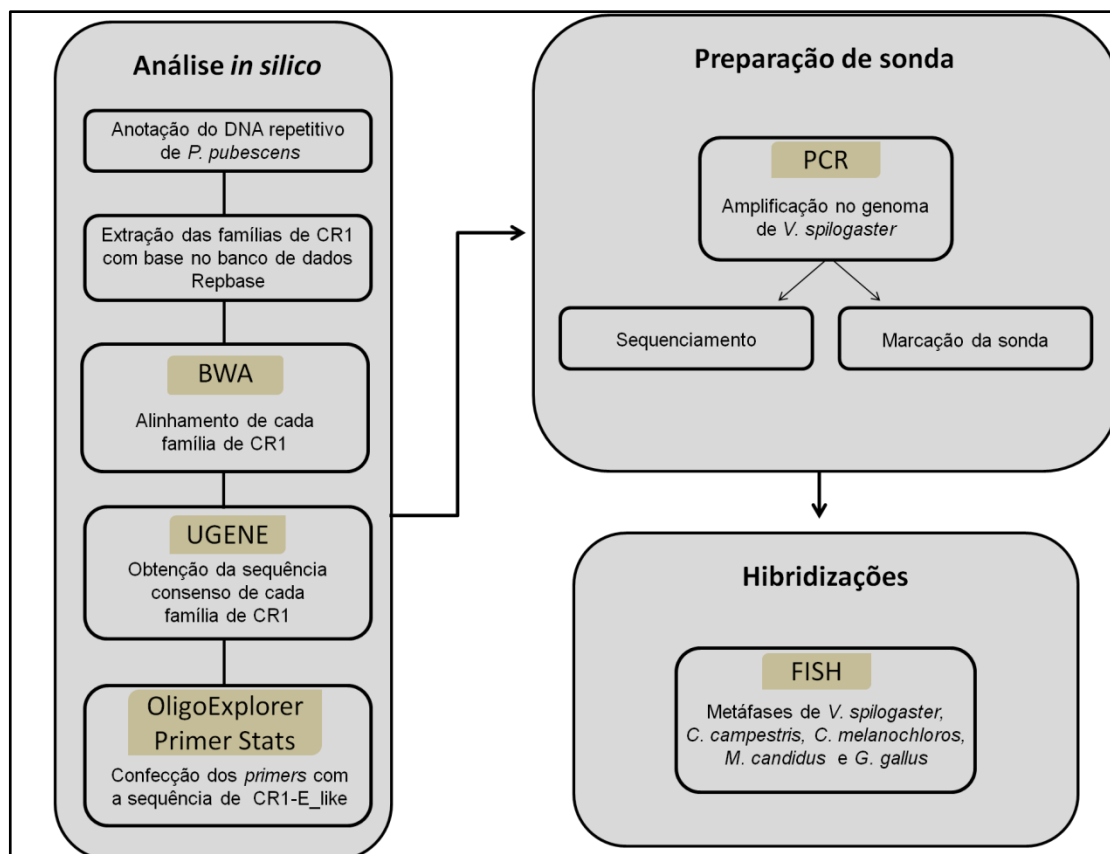


FIGURA 1- Fluxograma do trabalho

### 4.2.2 Screening de CR1-like em *Picoides pubescens* e prospecção dos primers

Com base na anotação do genoma de *P. pubescens* foi possível acessar os elementos repetitivos disponibilizado em <http://gigadb.org/dataset/101012>, as sequências foram extraídas segundo o script de Bailly-bechet; Haudry; Lerat (2014), e então subdivididas através de um script em python (*in house*), separando cada família de elemento em um arquivo. Para referência dos elementos CR1-like, utilizamos o banco de dados do RepBase Update (BAO; KOJIMA; KOHANY, 2015)

no dia 06/10/2015 para extrair, via website, todos os elementos CR1-like disponíveis no RepBase até a data, as sequências ficaram em um único arquivo de referência.

Os elementos de *P. pubescens* foram alinhados com base no arquivo de referência, através do BWA (LI; DURBIN, 2009) utilizando os parâmetros padrão. Os resultados dos alinhamentos foram visualizados no *software* Ugene (OKONECHNIKOV et al., 2012), e após o refinamento manual foram obtidas sequências consenso para a prospecção de *primers* específicos para cada CR1-like.

Foram selecionadas sequências consenso dos elementos com maior número de cópias no genoma, visando os experimentos de FISH (do inglês *Fluorescence in situ Hybridization*). Os *primers* foram construídos através do *software* OligoExplorer 1.2 (<http://www.genelink.com/tools/gl-oe.asp>), e da plataforma *online* Primer Stats (STOTHARD, 2000). Posteriormente foram testados no primer-BLAST (YE et al., 2012) para possíveis ampliações além do genoma de *P. pubescens*.

Os *primers* foram construídos para amplificar aproximadamente 500 pb do elemento CR1-E\_pass do genoma de *P. pubescens*. O fragmento amplificado contém parte da ORF2 e da porção terminal 3' do elemento. As sequências dos *primers* são:

- direto CR1\_E500S: 5'- ATGGACAGTGGCATTGAGTG -3'

- reverso CR1\_E500A: 5'- GCCCTCTGATCATCTTTGTG -3'

#### 4.2.3 Coleta de amostras

As coletas dos espécimes ocorreram no Estado do Rio Grande do Sul, Brasil conforme a tabela 1. A captura dos espécimes foi realizada com redes ornitológicas tipo *mist net* conforme licença do SISBIO, registro 33860-1 e 44173-1 e aprovação pelo Comitê de Ética no Uso de Animais em pesquisa (CEUA) da Universidade Federal do Pampa, número 026/2012.

TABELA 1- Espécies amostradas, quantidade de espécimes e local das coletas.

	<b>Espécies</b>	<b>Número de indivíduos e sexo</b>	<b>Localização</b>
	<b><i>Gallus gallus</i> (GGA)</b>	1 ♀	São Gabriel
	<b><i>Veniliornis spilogaster</i> (VSP)</b>	1 ♂	Porto Vera Cruz
	<b><i>Colaptes campestris</i> (CCA)</b>	3 ♂ 1 ♀	Dom Pedrito, São Gabriel
	<b><i>Colaptes melanochloros</i> (CME)</b>	5 ♀	Dom Pedrito, São Gabriel
	<b><i>Melanerpes candidus</i> (MCA)</b>	1 ♂	Porto Vera Cruz

#### 4.2.4 Obtenção da sonda

As extrações de DNA genômico foram realizadas a partir de amostras de sangue do espécime *V. spilogaster*, com o kit PureLink® Genomic DNA Mini Kit (Invitrogen) seguindo as recomendações do fornecedor.

As reações de PCR foram realizadas em volume final de 50 µl, da seguinte forma: 50ng de amostra de DNA, 10 pmol de cada *primer*, 2.5 mM MgCl<sub>2</sub> e 1U de Taq DNA polimerase. As condições de amplificações foram: primeira desnaturação a 95°C por 2 minutos, 35 ciclos de desnaturação a 95°C por 40 segundos, anelamento a 55°C por 30 segundos e extensão de 72°C por 1 minuto, seguido de uma extensão a 72°C por 5 minutos.

O produto amplificado foi submetido à eletroforese em gel de agarose 1%, corado com GelRed, visualizado e fotografado em fotodocumentador L-Pix Ex (Loccus Biotecnologia), sobre luz UV.

Posteriormente, o produto amplificado foi purificado a partir de gel de agarose 1% com o kit Quick Gel Extraction & PCR (Invitrogen). Com o DNA purificado foi

novamente realizada a PCR, verificada a amplificação em gel de agarose 1% e novamente purificada com o mesmo *kit*. A partir de 1 µg de DNA do fragmento isolado foi produzida a sonda para os experimentos de FISH através do *kit* DIG-nick translation mix (ROCHE) de acordo com o fabricante.

O produto da PCR purificado foi sequenciado seguindo as recomendações do serviço de sequenciamento da Macrogen Inc. (Korea). As sequências foram submetidas ao BLASTn para confirmação da identidade do elemento CR1, e depositadas no banco de dados Repbase (aguardando código da sequência a ser liberada pelo banco de dados).

#### **4.2.5 Preparação cromossômica**

As preparações cromossômicas foram obtidas através da técnica de cultura de medula óssea de curta duração segundo Garner and Gunski (2000) e também foi realizado o cultivo de fibroblastos de acordo com Sasaki et al. (1968). Independentemente da metodologia utilizada houve o tratamento com colchicina e a amostra foi lavada e fixada com metanol e ácido acético (3:1). Foram coletadas amostras de sangue para extração de DNA genômico, através do kit Purelink genomic DNA mini kit (Invitrogen) e quantificado em espectrofotômetro NanoVue (GE Healthcare Life Sciences).

#### **4.2.6 Determinação do número diploide e cariótipo parcial de *V. spilogaster***

Para determinação do número diploide e morfologia dos cromossomos, foram analisadas 30 metáfases com coloração convencional (GIEMSA 5% em 0,07M tampão fosfato com pH 6,8). E, para detecção das regiões de heterocromatina constitutiva, utilizou-se o protocolo de banda C segundo SUMNER, (1972)

#### **4.2.7 Experimentos de FISH e detecção de sinal**



Os experimentos de FISH foram realizados segundo Pinkel et al. 1986 com modificações. As lâminas com os cromossomos fixados foram incubadas em ácido clorídrico 10 mM e pepsina 1% por 5 minutos em temperatura ambiente, após lavadas com 2xSSC por 5 minutos duas vezes e então desidratadas em série de etanol (70%, 90% e 100%) a 4°C por 2 minutos cada. Em seguida foram incubadas a 37°C em câmara úmida com RNase A (10mg/ml) por 30 minutos. E então as lâminas foram desidratadas em formamida 70% a 70 °C por 1 minuto e 30 segundos, e imediatamente lavadas em série de etanol (70%, 90% e 100%) por 5 minutos cada a -20°C. O mix de hibridização (sonda e tampão de hibridização) foi desnaturado a 95°C por 10 minutos e transferido imediatamente para o gelo e posteriormente aplicado nas lâminas. A hibridização foi realizada em câmara úmida *overnight*. Após a hibridização, as lâminas foram desidratadas 3 vezes em formamida 50 % a 45 °C por 3 minutos e imediatamente lavadas em 2xSSC por 2 minutos, 5 vezes a 45 °C. As lâminas foram incubadas, por 1 hora a 37 °C em câmara úmida, na solução de detecção (marcador fluorescente e 4xSSC/Tween), a detecção foi realizada com Anti-digoxigenina-Rhodamine (Roche). Subseqüentemente as lâminas foram lavadas em 4xSSC/tween por 3 vezes, 5 minutos a 37 °C com agitação leve. Depois, sem secar, as lâminas foram contrastadas com Antifade+DAPI.

#### **4.2.8 Análises no microscópio**

Aproximadamente 30 metáfases foram analisadas para confirmar número cromossômico, estrutura do cariótipo e resultados de FISH. As imagens foram capturadas no microscópio epifluorescente Olympus BX53 (Olympus Corporation, Ishikawa, Japan).

## 4.3 Resultados

### 4.3.1 Obtenção do TE CR1-like e prospecção de *primers*

Foram identificados diferentes elementos CR1 na anotação da fração repetitiva do genoma de *P. pubescens*, aproximadamente 380 elementos CR1-like distintos, que variaram de uma a mais de 100 mil cópias. Foram prospectados *primers* para os elementos CR1-like com os maiores números de cópias identificados no genoma de *P. pubescens*, CR1-E com 83.818 cópias e CR1-J3 com 118.492 cópias, visando obtenção de sonda para os experimentos de FISH. Por questões técnicas, foram utilizados apenas os primers para CR1-E para a realização deste trabalho. O elemento CR1-E-like apresentou amplificação, através da PCR, de fragmentos de aproximadamente 500 pb no genoma de VSP como esperado (Figura 2).

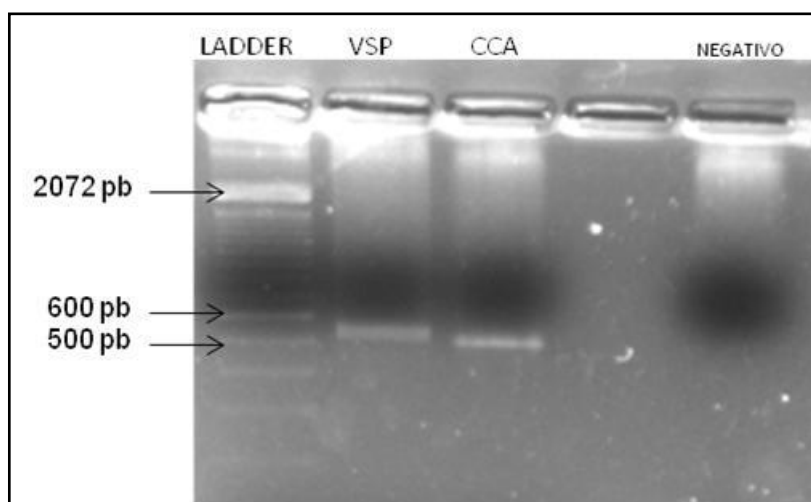


FIGURA 2- Amplificação do TE CR1-E-like em espécies duas espécies de pica-paus. A amplificação foi feita utilizando *primers* específicos prospectados para o elemento e os fragmentos foram separados por eletroforese em gel de agarose (1%). À esquerda marcador de peso molecular 100pb DNA Ladder; *V. spilogaster* (VSP); *C. campestris* (CCA); controle negativo.

### 4.3.2 Cariótipo de *V. spilogaster*

Na espécie (VSP) foi encontrado  $2n=80$  cromossomos. O cariótipo é formado por 14 pares de macrocromossomos, incluindo o cromossomo sexual Z, e 26 pares de microcromossomos, apresentando as seguintes morfologias: par 1 - submetacêntrica, dos pares 2 ao 14 - acrocêntrica, cromossomos Z - metacêntrica e

este cromossomo é o maior do complemento cariotípico, dos pares 15 ao 26 - telocêntrica (Figura 4). As regiões de heterocromatina são observadas nas regiões centroméricas dos cromossomos (Figura 4). O número diploide e as regiões de heterocromatina das outras espécies de pica-paus analisada por FISH estão disponíveis em De Oliveira et al. (2017), e de *G. gallus* em Stock; Arrighi; Stefos (1974).

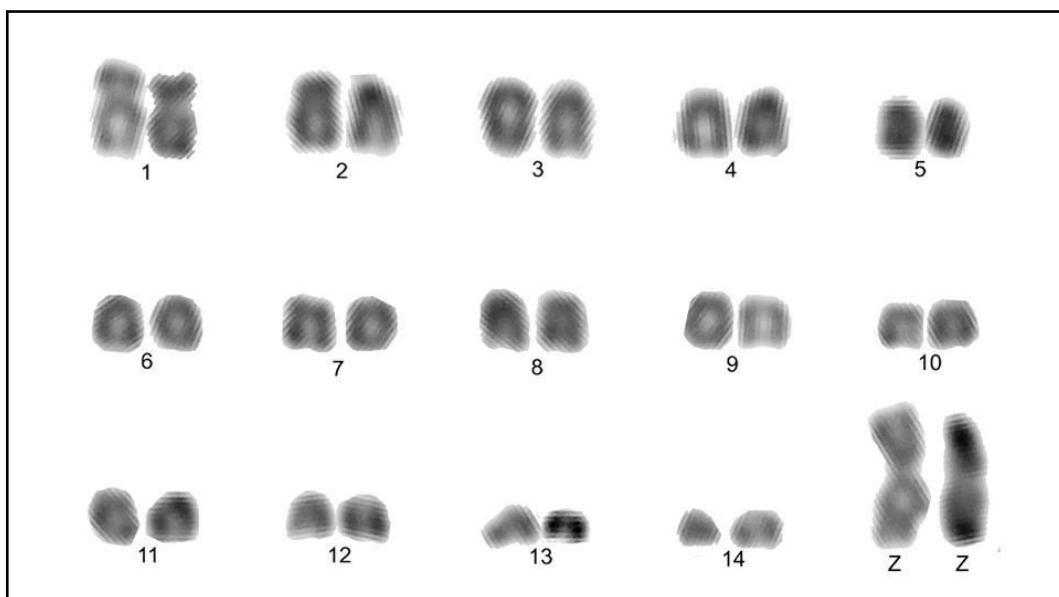


FIGURA 3- Cariótipo parcial de *V. spilogaster* (VSP). Os primeiros 14 pares e os cromossomos sexuais ZZ do espécime macho de VSP.

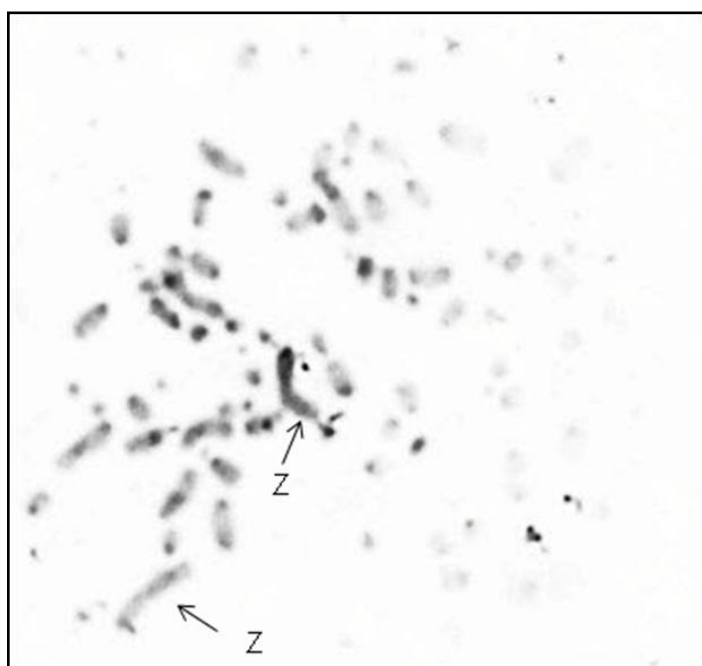


FIGURA 4- Bandejamento C em cromossomos metafásicos de um macho de *V. spilogaster*. As setas indicam os cromossomos sexuais.

### **4.3.3 Análises de FISH nas espécies de pica-paus e *G. gallus***

Observamos hibridização do fragmento de CR1-like nas cinco espécies estudadas, com padrões distintos por espécie. Houveram marcações tanto em macrocromossomos como em microcromossomos em todas as espécies. Destacou-se, principalmente, o primeiro par de macrocromossomos e o cromossomo sexual Z nas cinco espécies (Figura 5 e 6).

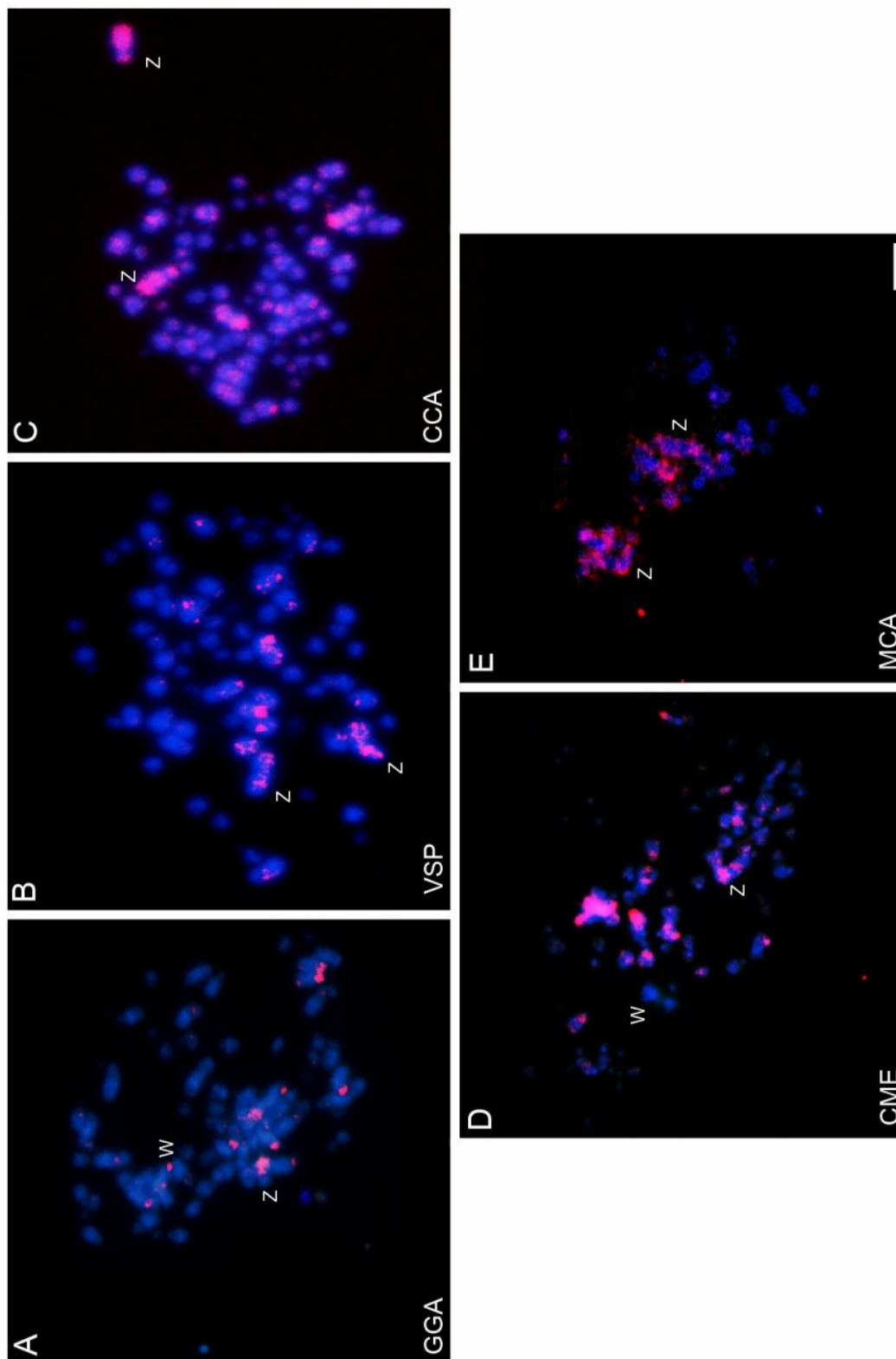


FIGURA 5- Experimentos de FISH com sondas de CR1-like em cromossomos metafásicos de quatro espécies pica-paus e *G. gallus*. (A) *Gallus gallus* (GGA), (B) *Veniliornis spilogaster* (VSP), (C) *Colaptes campestris* (CCA), (D) *Colaptes melanochloros* (CME) e (E) *Melanerpes candidus* (MCA). Os cromossomos foram contracolorados com DAPI (azul) e as sondas de CR1-like detectadas com Rodamina (vermelho). No canto inferior esquerdo estão indicadas as espécies. Os cromossomos sexuais estão indicados em cada metáfase. Barra de escala = 5 $\mu$ m.

No primeiro par de macrocromossomos obtivemos o seguinte padrão de marcações: em GGA na região centromérica e terminal do braço curto; em VSP ao longo de todo o braço longo, regiões centroméricas e terminais; em CCA ao longo de todo o cromossomo; em CME ao longo do braço longo e região terminal do braço curto; e em MCA nas regiões terminais de ambos os braços (Figura 6A).

No cromossomo sexual Z obtivemos o seguinte padrão: em GGA marcações pericentroméricas no braço curto; em VSP banda intersticial do braço longo; em CCA marcações na região terminal e ao longo de todo o braço longo; em CME nas regiões terminais e uma banda intersticial no braço longo; e em MCA região terminal e pericentromérica e intersticial do braço longo (Figura 6B).

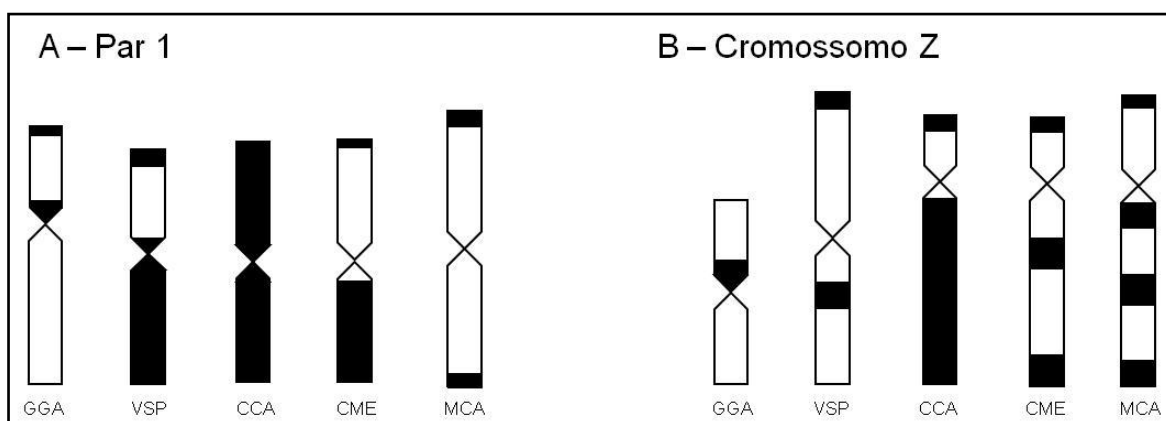


FIGURA 6- Representação esquemática da distribuição de CR1-E-like no primeiro par e cromossomo sexual Z. As regiões em preto correspondem às hibridizações de CR1-like tanto em (A) par 1 como em (B) cromossomo Z. As espécies estão indicadas abaixo de cada cromossomo, *Gallus gallus* (GGA), *Veniliornis spilogaster* (VSP), *Colaptes campestris* (CCA), *Colaptes melanochloros* (CME) e *Melanerpes candidus* (MCA).

#### 4.4 Discussão

Identificamos elementos CR1-like no genoma de *P. pubescens* de subfamílias descritas de CR1, como as subfamílias CR1-A a CR1-F, CR1- X e CR1-Y (HAAS et al., 2001; WICKER et al., 2005), contudo muitos dos elementos encontrados ainda não estão completamente caracterizados. A escolha pela subfamília CR1-E ocorreu devido ao elevado número de cópias dessa subfamília no respectivo genoma.

O TE CR1-E<sub>pass</sub> que identificamos no genoma de *P. pubescens* foi anteriormente identificado no genoma de aves Passeriformes por Suh et al. (2011) e classificado como pertencente a subfamília CR1-E. As hibridizações positivas em GGA (clado Galloanserae) e nas quatro espécies de pica-paus (clado Neoaves) corrobora com a teoria da subfamília CR1-E estar presente em toda a filogenia das aves Neognathas (SUH et al., 2011).

Com a FISH, foi possível observar a distribuição de CR1-E-like nos cromossomos de GGA, principalmente nos macrocromossomos e cromossomo sexual Z (Figura 5A), corroborando com o observado em Coullin et al. (2005).

Na ordem Piciformes estudos sobre a fração repetitiva dos genomas e sua distribuição nos cromossomos são insipientes, resumem-se à análise de sequências microssatélites e satélites (telômeros), com três das espécies (CCA, CME, MCA) estudadas neste trabalho (DE OLIVEIRA et al., 2017). Os estudos, na espécie VSP estavam restritos a análises de DNA mitocondrial para esclarecer incongruências filogenéticas dos gêneros *Veniliornes* e *Picoides*, visto que são filogeneticamente muito próximos (WEBB; MOORE, 2005).

A espécie VSP apresentou número cromossômico típico para a classe Aves,  $2n=80$  cromossomos (Figura 2), dentro da variação encontrada para os pica-paus que é de 64 a 108 cromossomos e um grande cromossomo Z (DE OLIVEIRA et al., 2017). Quanto à distribuição de heterocromatina constitutiva foram observadas marcações nas regiões centroméricas, os mesmos padrões encontrados para os outros Picidae e comum entre as aves (Figura 3) (DE OLIVEIRA et al., 2017; SHIBUSAWA et al., 2004)

As marcações, em MCA no cromossomo Z, demonstraram estar colocalizadas com as sequências microssatélites encontradas em De Oliveira (2017), entretanto apenas algumas hibridizações de CR1-E-like estão colocalizadas com as sequências de microssatélites em CCA e CME (Figuras 5C-E e 6B). E em VSP apenas a marcação na região terminal do cromossomo coincide com as hibridizações de sequências microssatélites, entretanto a espécie VSP não está entre as analisadas na literatura (Figuras 5B e 6B) (DE OLIVEIRA et al., 2017). Esses acúmulos coincidentes entre CR1-E-like e sequências microssatélites, podem estar associados à sequência de reconhecimento utilizadas pelas enzimas de mobilização para reconhecer os TEs CR1 (KAPUSTA; SUH, 2016). Contudo, nas subfamílias de CR1, que foram identificadas sequências microssatélites, estas possuíam oito pares de base (ATTCTRTG)<sub>n</sub>, e não três como as utilizadas por De Oliveira (2017). Mesmo assim, é possível que o elemento analisado possua uma sequência de reconhecimento similar às sequências microssatélites, mas distinta das anteriormente analisadas. Em GGA não foram observadas hibridizações de sequências microssatélites no cromossomo Z, somente em W que não coincidem com os nossos resultados (MATSUBARA et al., 2016).

Observou-se também, nos cromossomos sexuais Z de pica-paus, bandas de hibridização maiores do que é observado em GGA (Figura 6), indicando um acúmulo de TEs nesse cromossomo. Com isso, é possível sugerir que o aumento deste cromossomo, o maior do complemento cariotípico (Figura 2), pode estar relacionado com o acúmulo de diferentes sequências repetitivas (DE OLIVEIRA et al., 2017; ZHANG et al., 2014).

As marcações, nas cinco espécies, não estão confinadas às regiões de heterocromatina dos cromossomos (centrômeros), porém no par 1 observa-se hibridizações nas regiões centroméricas em GGA, VSP e CCA. Somente GGA apresenta essas marcações no par 1 e cromossomo Z (Figura 5). Esse padrão não restritivo a determinadas regiões dos cromossomos e distribuição do TE diferente no mesmo cromossomo em espécies filogeneticamente próximas é visto também para outros retrotransposons, como Rex 1, 3 e 6, nos genomas de peixes (SUPIWONG et al., 2013)



As espécies de pica-paus apresentam padrão de marcações bastante distinto em relação à GGA. Em GGA observa-se uma quantidade menor de marcações, o que apoia as hipóteses, de que houve uma grande dispersão de CR1-like em Piciformes e de que a subfamília CR1-E está presente em menor quantidade neste genoma (KAPUSTA; SUH, 2016; ZHANG et al., 2014).

Nossos resultados demonstram que o elemento CR1-like está presente em diferentes linhagens das aves Neognathas, visto que houve hibridizações do TE em GGA (ave basal do clado) e nas espécies de pica-paus estudadas. Evidenciou-se também, que a distribuição do elemento, não está restrita às regiões de heterocromatina e sua localização no cromossomo varia entre as espécies, porém a maior concentração de hibridizações ocorre nos maiores pares do complemento (cromossomo Z e par 1).

Em Piciformes, uma característica marcante é o cromossomo Z ser o maior do complemento cariotípico, diferentemente de outras linhagens de aves. Vem sendo atribuído a isso o acúmulo de sequências repetitivas nesse cromossomo para explicar o seu maior tamanho. Nesse sentido, o maior número de marcações nos cromossomos Z das espécies de pica-paus em relação à espécie basal GGA, mostradas nesse trabalho corrobora com essa hipótese na perspectiva dos elementos de transposição.

#### 4.5 Referências

BAILLY-BECHET, M.; HAUDRY, A.; LERAT, E. “One code to find them all”: a perl tool to conveniently parse RepeatMasker output files. **Mobile DNA**, v. 5, n. 1, p. 13, 2014.

BAO, W.; KOJIMA, K. K.; KOHANY, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. **Mobile DNA**, v. 6, n. 1, p. 11, 2015.

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**, v. 90, n. 17, p. 8199–8203, 1993.

COULLIN, P. et al. Cytogenetic repartition of chicken CR1 sequences evidenced by PRINS in Galliformes and some other birds. **Chromosome Research**, v. 13, n. 7, p. 665–673, 2005.

DE OLIVEIRA, T. D. et al. Genomic Organization of Repetitive DNA in Woodpeckers (Aves, Piciformes): Implications for Karyotype and ZW Sex Chromosome Differentiation. **Plos One**, v. 12, n. 1, p. e0169987, 2017.

GARNERO, A. D. V.; GUNSKI, R. J. Comparative analysis of the karyotypes of *Nothura maculosa* and *Rynchotus rufescens* (Aves, Tinamidae). A case of chromosomal polymorphism. **The Nucleus**, v. 43, p. 64–70, 2000.

GIBBS, R. A et al. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature*, v. 428, n. 6982, p. 493–521, 2004.

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**, 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**, 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. **Science**, v. 346, n. 6215, p. 1126–1138, 2014.

KAJIKAWA, M.; OHSHIMA, K.; OKADA, N. Determination of the entire sequence of turtle CR1: the first open reading frame of the turtle CR1 element encodes a protein with a novel zinc finger motif. **Molecular biology and Evolution**, v. 14, n. 12, p. 1206–1217, 1997.

KAPITONOV, V. V.; JURKA, J. The esterase and PHD domains in CR1-like non-LTR retrotransposons. **Molecular Biology and Evolution**, 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**, p. 1–22, 2016.

KORDIS, D.; LOVSIN, N.; GUBENSEK, F. Phylogenomic analysis of the L1 retrotransposons in Deuterostomia. **Systematic biology**, v. 55, n. 6, p. 886–901, 2006.

LI, H.; DURBIN, R. Fast and accurate short read alignment with Burrows-Wheeler transform. **Bioinformatics**, v. 25, n. 14, p. 1754–1760, 2009.

MATSUBARA, K. et al. Amplification of microsatellite repeat motifs is associated with the evolutionary differentiation and heterochromatinization of sex chromosomes in Sauropsida. **Chromosoma**, v. 125, n. 1, p. 111–123, 2016.

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

PINKEL, D.; STRAUME, T.; GRAY, J. W. Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. **Proceedings of the National Academy of Sciences of the United States of America**, v. 83, n. 9, p. 2934–8, 1986.

SASAKI, M.; IKEUCHI, T.; MAKINO, S. A feather pulp culture technique for avian chromosomes, with notes on the chromosomes of the peafowl and the ostrich. **Experientia**, v. 24, n. 12, p. 1292–1293, 1968.

SHIBUSAWA, M. et al. Karyotypic evolution in the Galliformes: an examination of the process of karyotypic evolution by comparison of the molecular cytogenetic findings

with the molecular phylogeny. **Cytogenetic and Genome Research**, v. 106, n. 1, p. 111–9, 2004.

SMITH, J. D. L.; GREGORY, T. R. The genome sizes of megabats (Chiroptera: Pteropodidae) are remarkably constrained. *Biology letters*, v. 5, n. 3, p. 347–51, 2009.

STOCK, A. D.; ARRIGHI, F. E.; STEFOS, K. Chromosome homology in birds: banding patterns of the chromosomes of the domestic chicken, ring-necked dove, and domestic pigeon. **Cytogenet Cell Genet**, v. 13, p. 410–418, 1974.

STOTHARD, P. The Sequence Manipulation Suite: javascript programs for analysing and formatting protein and DNA sequences. **Biotechniques**, v. 28, p. 1102–1104, 2000.

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

SUH, A. et al. Mesozoic retroposons reveal parrots as the closest living relatives of passerine birds. **Nature Communications**, v. 2, n. 1, p. 443, 2011.

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

SUH, A. The Specific Requirements for CR1 Retrotransposition Explain the Scarcity of Retrogenes in Birds. **Journal of Molecular Evolution**, v. 81, n. 1, p. 18–20, 2015.

SUMNER, A. T. A simple technique for demonstrating centromeric heterochromatin. *Exptl Cell Res*, v. 75, p. 304–306, 1972.

SUPIWONG, W. et al. Karyotype and cytogenetic mapping of 9 classes of repetitive DNAs in the genome of the naked catfish *Mystus bocourti* (Siluriformes, Bagridae). **Molecular Cytogenetics**, v. 6, p. 51, 2013.

WEBB, D. M.; MOORE, W. S. A phylogenetic analysis of woodpeckers and their allies using 12S, Cyt b, and COI nucleotide sequences (class Aves; order Piciformes). **Molecular Phylogenetics and Evolution**, v. 36, n. 2, p. 233–248, 2005.

WICKER, T. et al. The repetitive landscape of the chicken genome. **Genome Research**, v. 15, n. 1, p. 126–136, 2005.

YE, J. et al. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. **BMC Bioinformatics**, v. 13, n. 1, p. 134, 2012.

ZHANG, G. et al. Comparative genomics reveals insights into avian genome evolution and adaptation. **Science**, v. 346, n. 6215, p. 1311–1321, 2014.

## 5 CONSIDERAÇÕES FINAIS

Com este trabalho acreditamos ter contribuído para o conhecimento de elementos transponíveis em aves, utilizando duas abordagens diferentes e eficientes.

No primeiro capítulo, através de análises *in silico*, foi possível estimar a diversidade de cópias, as características estruturais e funcionais, a distribuição descontínua na linhagem aviária, a descrição de uma nova subfamília para família *mariner* e sugerir um evento de transferência horizontal do elemento *galluhop*. Destacamos nesta abordagem o primeiro registro de uma possível transferência horizontal de elemento transponível entre aves, e a nova subfamília *Gallus* restrita, até o momento, a elementos *galluhop* encontrados em aves.

No segundo capítulo, conseguimos visualizar a distribuição cromossômica de um mesmo elemento transponível, CR1-E-like, em diferentes genomas de aves, evidenciando um padrão de distribuição muito diferente em cada genoma, mesmo entre espécies do mesmo gênero. Destacamos a utilização e confecção precursora de sonda específica um elemento de ave na hibridização *in situ* fluorescente .

A disponibilidade crescente de novos genomas de aves sequenciados possibilitará avanços contínuos no estudo sobre elementos transponíveis em aves, visto que é uma área cheia de possibilidades como demonstramos nesse trabalho.

## REFERÊNCIAS

- AZIZ, R. K.; BREITBART, M.; EDWARDS, R. A. Transposases are the most abundant, most ubiquitous genes in nature. **Nucleic Acids Research**, v. 38, n. 13, p. 4207–4217, 2010.
- BIÉMONT, C.; VIEIRA, C. Genetics: junk DNA as an evolutionary force. **Nature genetics**, v. 443, n. 5, p. 521–524, 2006.
- 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**, v. 90, n. 17, p. 8199–8203, 1993.
- BURT, D. W. Origin and evolution of avian microchromosomes. **Cytogenetics and Genome Research**, v. 96, n. 6232, p. 97–112, 2002.
- 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.
- CHICKEN, I.; SEQUENCING, G. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. **Nature**, v. 432, n. 7018, p. 695–716, 2004.
- CHRISTIDIS, L. Aves. In: **Animal Cytogenetics**. Berlin: [s.n.]. v. Chordata 3.
- DALLOUL, R. A. et al. Multi-platform next-generation sequencing of the domestic Turkey (*Meleagris gallopavo*): Genome assembly and analysis. **PLoS Biology**, v. 8, n. 9, 2010.
- FATTASH, I. et al. Miniature inverted-repeat transposable elements: discovery, distribution, and activity. **Genome**, v. 56, n. 9, p. 475–486, 2013.
- FINNEGAN, D. J. Eukaryotic transposable elements and genome evolution. Trends in **Genetics**, v. 5, n. C, p. 103–107, 1989.
- GIBBS, R. A et al. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. **Nature**, v. 428, n. 6982, p. 493–521, 2004.

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**, 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**, v. 265, n. 1–2, p. 175–183, 2001.

HARTL, D. L.; LOHE, A. R.; LOZOVSKAYA, E. R. Modern thoughts on an ancient mariner: function, evolution, regulation. **Annual review of genetics**, v. 31, p. 337–58, 1997.

HILLIER, L. W. et al. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. **Nature**, v. 432, n. 7018, p. 695–716, 2004.

JACOBSON, J. W.; MEDHORA, M. M.; HARTL, D. L. Molecular structure of a somatically unstable transposable element in *Drosophila*. **Proceedings of the National Academy of Sciences of the United States of America**, v. 83, n. 22, p. 8684–8688, 1986.

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

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

KAJIKAWA, M.; OHSHIMA, K.; OKADA, N. Determination of the entire sequence of turtle CR1: the first open reading frame of the turtle CR1 element encodes a protein with a novel zinc finger motif. **Molecular Biology and Evolution**, v. 14, n. 12, p. 1206–1217, 1997.

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

KAPITONOV, V. V.; JURKA, J. A universal classification of eukaryotic transposable elements implemented in Repbase. **Nature reviews. Genetics**, v. 9, n. 5, p. 411–412; author reply 414, 2008.



KAPUSTA, A.; SUH, A. Evolution of bird genomes-a transposon's-eye view. **Annals of the New York Academy of Sciences**, p. 1–22, 2016.

KOEPFLI, K. et al. The Genome 10K Project: a way forward. **Annual review of animal biosciences**, v. 3, p. 57–111, 2015.

KORDIS, D. Transposable elements in reptilian and avian (Sauropsida) genomes. **Cytogenetic and Genome Research**, v. 127, n. 2–4, p. 94–111, 2010.

LANDER, E. S. et al. Initial sequencing and analysis of the human genome. **Nature**, v. 409, n. 6822, p. 860–921, 2001.

LE ROUZIC, A.; CAPY, P. Population genetics models of competition between transposable element subfamilies. **Genetics**, v. 174, n. 2, p. 785–793, 2006.

MCCLINTOCK, B. The Origin and Behavior of Mutable Loci in Maize. **Genetics**, v. 36, n. 6, p. 344–355, 1950.

NISHIDA, C. et al. Characterization of chromosome structures of Falconinae (Falconidae, Falconiformes, Aves) by chromosome painting and delineation of chromosome rearrangements during their differentiation. **Chromosome Research**, v. 16, n. 1, p. 171–181, 2008.

OLIVEIRA, S. G. et al. Chromosomal organization and evolutionary history of Mariner transposable elements in Scarabaeinae coleopterans. **Molecular Cytogenetics**, v. 6, n. 1, p. 54, 2013.

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

ROUAULT, J. D. et al. Automatic classification within families of transposable elements: Application to the mariner Family. **Gene**, v. 448, n. 2, p. 227–232, 2009.

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

- SUH, A. et al. Multiple lineages of ancient CR1 retroposons shaped the early genome evolution of amniotes. **Genome Biology and Evolution**, v. 7, n. 1, p. 205–217, 2014.
- VARANI, A. DE M. et al. Elementos de transposição: Classificação e Mecanismos de Mobilização. In: **Elementos de transposição: diversidade, evolução, aplicações e impacto nos genomas dos seres vivos**. 1. ed. Rio de Janeiro: FIOCRUZ, 2015. p. 11–42.
- WALLAU, G. et al. Genomic landscape and evolutionary dynamics of mariner transposable elements within the *Drosophila* genus. **BMC Genomics**, v.15, p. 1–19, 2014.
- WALLAU, G. L. et al. The evolutionary history of mariner-like elements in Neotropical drosophilids. **Genetica**, v. 139, n. 3, p. 327–338, 2011.
- WALLAU, G. L.; ORTIZ, M. F.; LORETO, E. L. S. Horizontal transposon transfer in eukarya: Detection, bias, and perspectives. **Genome Biology and Evolution**, v. 4, n. 8, p. 689–699, 2012.
- WICKER, T. et al. The repetitive landscape of the chicken genome. **Genome Research**, v. 15, n. 1, p. 126–136, 2005.
- WICKER, T. et al. A unified classification system for eukaryotic transposable elements. **Nature reviews. Genetics**, v. 8, n. 12, p. 973–982, 2007.
- WRIGHT, N. A.; GREGORY, T. R.; WITT, C. C. Metabolic 'engines' of flight drive genome size reduction in birds. **Proceedings of the Royal Society B**, v. 281, n. January, 2014.
- ZHANG, G. et al. Comparative genomics reveals insights into avian genome evolution and adaptation. **Science**, v. 346, n. 6215, p. 1311–1321, 2014.