IGOR RODRIGUES DA COSTA

Locos anônimos

Rio de Janeiro

2015

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

INSTITUTO DE BIOQUÍMICA MÉDICA LEOPOLDO DE MEIS

PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA BIOLÓGICA

IGOR RODRIGUES DA COSTA

LOCOS ANÔNIMOS

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Química Biológica, Instituto de Bioquímica Médica Leopoldo de Meis, Universidade Federal do Rio de Janeiro, como requisito parcial à obtenção do título de Mestre.

Orientador: Prof. Francisco Prosdocimi, Ph.D

Co-orientador: William Bryan Jennings, Ph.D

Rio de Janeiro

2015

|  |
| --- |
| Costa, Igor Rodrigues da  Locos Anônimos/ Igor Rodrigues da Costa. – Rio de Janeiro: UFRJ/IBqM, 2015.  XXII, 132 p.: il.; 29,7 cm.  Orientador: Francisco Prosdocimi  Dissertação (mestrado) – UFRJ/IBqM/Programa de Pós-graduação em Química Biológica, 2015.  1. Locos Anônimos. 2. Bioinformática. 3. Genômica Comparativa. II. Universidade Federal do Rio de Janeiro, Instituto de Bioquímica Médica Leopoldo de Meis, Programa de Pós-graduação em Química Biológica. III. Título. |

IGOR RODRIGUES DA COSTA

LOCOS ANÔNIMOS

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Química Biológica, Instituto de Bioquímica Médica Leopoldo de Meis, Universidade Federal do Rio de Janeiro, como requisito parcial à obtenção do título de Mestre.

Aprovada em / /

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Prof. Francisco Prosdocimi, Ph.D, UFRJ

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Bryan Jennings, Ph.D, UFRJ

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Prof. <Nome do membro da banca>, <título ex.: Ph.D, DSc>, <filiação ex.: UFRJ>

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Prof. <Nome do membro da banca>, <título ex.: Ph.D, DSc>, <filiação ex.: UFRJ>

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Prof. <Nome do membro da banca>, <título ex.: Ph.D, DSc>, <filiação ex.: UFRJ>

<Dedicatória>

Agradecimentos

<Epígrafe>

Resumo

COSTA, Igor Rodrigues da. **Locos Anônimos**. 2015. <número de folhas>. Tese (Mestrado em Química Biológica) – Instituto de Bioquímica Médica Leopoldo de Meis, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2015.

O resumo deve ser redigido em só parágrafo, de preferência, na 3ª pessoa do singular e o verbo na voz ativa com, no máximo, 500 palavras e no mínimo 150 palavras. Sugere-se que o resumo venha antecedido por uma referência bibliográfica do trabalho.

Palavras-chave: <lista de palavras-chave>

Abstract

COSTA, Igor Rodrigues da. **Locos Anônimos**. 2015. <número de folhas>. Tese (Mestrado em Química Biológica) – Instituto de Bioquímica Médica Leopoldo de Meis, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2015.

O resumo deve ser redigido em só parágrafo, de preferência, na 3ª pessoa do singular e o verbo na voz ativa com, no máximo, 500 palavras e no mínimo 150 palavras.Sugere-se que o resumo venha antecedido por uma referência bibliográfica do trabalho.

Keywords: <lista de palavras-chave>

Lista de Figuras

[Figura 1: <descrição da figura> 20](#_Toc405213978)

[Figura 2: <descrição da figura> 20](#_Toc405213979)

[Figura 1: Frequência das topologias observadas 22](#_Toc405213980)

Lista de Tabelas

[Quadro 1: <descrição da tabela> 3](#_Toc338858942)

[Quadro 2: <descrição da tabela> 3](#_Toc338858943)

Lista de Siglas

|  |  |
| --- | --- |
| AL | Loco Anônimo |
| <sigla 2> | <palavras e expressões correspondentes grafadas por extenso> |
| <sigla 3> | <palavras e expressões correspondentes grafadas por extenso> |
| <sigla 4> | <palavras e expressões correspondentes grafadas por extenso> |
| <sigla 5> | <palavras e expressões correspondentes grafadas por extenso> |

Sumário

[Introdução 16](#_Toc405213423)

[1.1 Marcadores moleculares 16](#_Toc405213424)

[1.1.1 Utilidade dos marcadores moleculares 16](#_Toc405213425)

[1.1.2 Microssatélites 16](#_Toc405213426)

[1.1.3 SNPs 16](#_Toc405213427)

[1.1.4 Locos Anônimos 16](#_Toc405213428)

[1.2 Genômica 17](#_Toc405213429)

[1.2.1 Sequenciamento de nova geração 17](#_Toc405213430)

[1.2.2 Genomas completos e parciais 17](#_Toc405213431)

[1.3 Genômica Computacional 17](#_Toc405213432)

[1.3.1 Bancos de dados 17](#_Toc405213433)

[1.3.2 Softwares para análise de dados 17](#_Toc405213434)

[1.4 Genética de População 17](#_Toc405213435)

[1.4.1 Parâmetros de interesse para estudo 17](#_Toc405213436)

[1.4.2 O modelo dos *Hominidae* na genética de população 17](#_Toc405213437)

[2 Objetivo 18](#_Toc405213438)

[3 Desenvolvimento 19](#_Toc405213439)

[3.1 Pipeline para busca de Locos Anônimos em genomas completos 19](#_Toc405213440)

[3.1.1 Busca por regiões anônimas 19](#_Toc405213441)

[3.1.2 Filtragem por conservação e unicidade 19](#_Toc405213442)

[3.1.3 Alinhamento dos Locos Anônimos 19](#_Toc405213443)

[3.2 Pipeline para busca de Locos Anônimos em genomas parciais 19](#_Toc405213444)

[3.2.1 Filtro de qualidade das sequencias 19](#_Toc405213445)

[3.2.2 Concatenamento das sequencias *pair-ended* 19](#_Toc405213446)

[3.2.3 Predição de primers 19](#_Toc405213447)

[3.2.4 Verificação de primers por PCR *in silico* 19](#_Toc405213448)

[3.3 Figuras 20](#_Toc405213449)

[3.4 Tabelas 20](#_Toc405213450)

[3.5 Notas de rodapé 20](#_Toc405213451)

[4 Resultados 21](#_Toc405213452)

[4.1 Predição das regiões anônimas do genoma humano 21](#_Toc405213453)

[4.2 Homologia e unicidade dos Locos Anônimos hipotéticos humanos em quatro genomas de *Hominidae* 21](#_Toc405213454)

[4.3 Análise dos Locos Anônimos de *Hominidae* 21](#_Toc405213455)

[4.3.1 Mapeamento dos Locos Anônimos por cromossomo 21](#_Toc405213456)

[4.3.2 Modelos de substituição 21](#_Toc405213457)

[4.3.3 Filogenia 21](#_Toc405213458)

[4.3.4 Estimativa da população efetiva e tempo de divergência 21](#_Toc405213459)

[4.3.5 CpG e taxa de mutação 21](#_Toc405213460)

[4.3.6 Teste de neutralidade 21](#_Toc405213461)

[Conclusão 22](#_Toc405213462)

[Referências 23](#_Toc405213463)

[Glossário 24](#_Toc405213464)

[Anexos 25](#_Toc405213465)

[ANEXO A - PÁGINA DE ROSTO 25](#_Toc405213466)

[Apêndices 26](#_Toc405213467)

[APÊNDICE A – Código Fonte Dos programas utilizados 26](#_Toc405213468)

[Índice 27](#_Toc405213469)

Introdução

Neste trabalho é descrito um novo programa criado para predição *in* silico de marcadores moleculares ideais para genética de população. Além disso, são apresentadas análises de marcadores criados a partir da aplicação deste programa no genoma dos *Hominidae*. Por fim, os parâmetros estimados são comparados favoravelmente com o estado da arte.

## Marcadores genéticos

Marcadores genéticos são sequências de DNA que, por apresentarem variações, podem ser usadas para diferenciar organismos ou espécies. O primeiro marcador genético a ser desenvolvido, 30 anos atrás, se aproveitava de variações nos sítios alvo de enzimas de restrição para criar *fingerprints* genéticos e são usados até hoje em testes forensicos e de paternidade [[1](#_ENREF_1" \o "Jeffreys, 1985 #200)]. Avanços nas técnicas de sequênciamento nos últimos 30 anos permitiram o desenvolvimento de diversos novos marcadores genéticos, como microssatélites, polimorfismos de nucleotídeo único (SNP, *single nucleotide polymorphism*) e os locos anônimos (AL, *anonymous loci*).

### Utilidade dos marcadores genéticos

Os marcadores genéticos são ferramentas versáteis, sendo úteis em campos como a ciência forense, testes de paternidade, medicina personalizada e diagnóstica, genética de população, filogenia e agropecuária, entre outros.

### Microssatélites

Um dos marcadores genéticos mais usados são os microssatélites

### SNPs

Outra classe de marcadores genéticos, os polimorfismos de nucleotídeos únicos (SNPs) são variações de uma única base.

### Locos Anônimos

Locos anônimos (ALs), os marcadores de interesse neste trabalho, são regiões não codificantes com caracteristicas ideais para estudos de genética de população e filogenia. Essas caracteristicas são:

1. Cópia única no genôma;
2. Sob seleção neutra;
3. Segregação independente.

#### Vantages e desvantagens dos Locos Anônimos sobre os outros marcadores

Locos anônimos, por apresentarem estas caracteristicas ideais,

#### Estado da arte na descoberta e descrição de Locos Anônimos

Pegar da apresentação

## Genômica

### Sequenciamento de nova geração

### Genomas completos e parciais

## Genômica Computacional

### Bancos de dados

### Softwares para análise de dados

## Genética de População

### Parâmetros de interesse para estudo

#### Estimativa de tempo de divergência

#### Estimativas de população efetiva

### O modelo dos *Hominidae* na genética de população

# Objetivo

Desenvolver uma metodologia para descoberta de Locos Anônimos tanto em genomas completos quanto para dados brutos de sequenciamento e aplicá-la no modelo dos *Hominidae* e dos Beija-Flores.

# Desenvolvimento

# Resultados

Após o desenvolvimento do programa ALFie foi realizada a busca de Locos Anônimos nos genomas de Humano, Chimpanzé, Gorila e Orangutango. Os LAs encontrados foram analisados com diversas ferramentas para obtenção de dados populacionais destas espécies.

## Pipeline para busca de Locos Anônimos em genomas completos

The current version of the human genome was downloaded as FASTA sequences and cut into regions that set apart from genes, generating two datasets of **anonymous regions** that were distant at least (a) 10Kb and (b) 200Kb from genes. Genes were considered as any feature present in the human annotated file from ENSEMBL and included also pseudogenes, microRNA genes and other transcribed annotated regions. We used a filter of 10Kb for the distance between the regions and the telomeric portions of the chromosomes. These anonymous regions were then cut into smaller 1Kb pieces that were BLASTed against the genomes of human, chimpanzee, gorilla and orang-utan. These 1Kb candidate loci were then filtered based on being single-copy in each genome and having at least 90% of nucleotidic similarity and coverage in all genomes analyzed.

All the loci that passed through the describe filters were accessed based on the distances between them in the human genome. Once more, 10Kb and 200Kb filters were used to provide final informative loci for population genetics. At least, the four ortholog loci were aligned and a minimum alignment size of 500 bp filter was applied excluding indels.

### Busca por regiões anônimas

#### Separação em Locos Anônimos hipotéticos

1. 1;
2. 2;
3. 3;

### Filtragem por conservação e unicidade

### Alinhamento dos Locos Anônimos

1. A;
2. B;
3. C.

## Pipeline para busca de Locos Anônimos em genomas parciais

### Filtro de qualidade das sequencias

### Concatenamento das sequencias *pair-ended*

### Predição de primers

### Verificação de primers por PCR *in silico*

“citação citaçãocitaçãocitaçãocitaçãocitaçãocitaçãocitaçãocitaçãocitaçãocitaçãocitaçãocitaçãocitaçãocitaçãocitaçãocitaçãocitaçãocitaçãocitaçãocitaçãocitaçãocitaçãocitaçãocitaçãocitaçãocitaçãocitaçãocitaçãocitaçãocitaçãocitaçãocitaçãocitaçãocitaçãocitaçãocitaçãocitaçãocitaçãocitação”

## Figuras



Figura : <descrição da figura>



Figura : <descrição da figura>

## Tabelas

Quadro : <descrição da tabela>

|  |  |  |
| --- | --- | --- |
| Coluna 1 | Coluna 2 | Coluna 3 |
| A | B | C |
| B | C | A |

Quadro : <descrição da tabela>

|  |  |
| --- | --- |
| Coluna A | Coluna B |
| 23 | 55 |
| 34 | 12 |

## Notas de rodapé

Nota[[1]](#footnote-1)

## Predição das regiões anônimas do genoma humano

Normal normal normal

## Homologia e unicidade dos Locos Anônimos hipotéticos humanos em quatro genomas de *Hominidae*

## Análise dos Locos Anônimos de *Hominidae*

### Mapeamento dos Locos Anônimos por cromossomo

### Modelos de substituição

### Filogenia

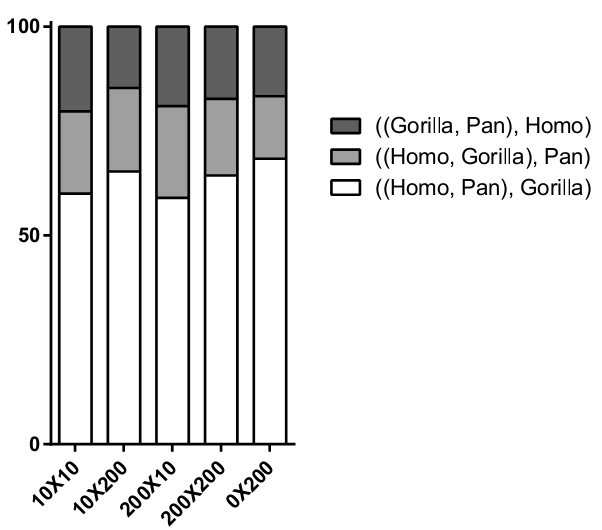


Figura : Frequência das topologias observadas

### Estimativa da população efetiva e tempo de divergência

### CpG e taxa de mutação

CpG

Close: 0,490479667 +- 0,094620764

### Teste de neutralidade

Normal normal normal

## Análise dos Locos Anônimos do genôma parcial de *Amazilia versicolor*

### Amplificação em *Calypte anna*

### Sequenciamento

Conclusão

Normal normal normal

Referências

Glossário

Relação de palavras ou expressões técnicas de uso restrito ou de sentido obscuro, utilizadas no texto, acompanhadas das respectivas definições. É um elemento opcional, elaborado em ordem alfabética.

Anexos

Elemento opcional, que consiste em um texto ou documento não elaborado pelo autor, que serve de fundamentação, comprovação e ilustração.

ANEXO A - PÁGINA DE ROSTO

Apêndices

Elemento opcional que consiste em um texto ou documento elaborado pelo autor, a fim de complementar sua argumentação,sem prejuízo da unidade nuclear do trabalho.

APÊNDICE A – Código fonte dos programas utilizados

Normal normal normal

Índice

É a lista de palavras ou frases, ordenadas segundo um determinado critério, que localiza e remete para as informações contidas no texto. O índice aparece no final da publicação.

1. Jeffreys, A.J.; Wilson, V.; Thein, S.L. Hypervariable 'minisatellite' regions in human DNA. *Nature* **1985**, *314*, 67-73.

1. Descrição da nota de rodapé [↑](#footnote-ref-1)