

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

GENOME COMPARISON BASED ON RAW READ
AND PAIRED READ DATA.

DIPLOMA THESIS

2016
VLADIMÍR MACKO

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

GENOME COMPARISON BASED ON RAW READ
AND PAIRED READ DATA.

DIPLOMA THESIS

Študijný program: Informatika
Študijný odbor: 2508 Informatika
Školiace pracovisko: Katedra informatiky
Školiteľ: Mgr. Tomáš Vinař, PhD

Bratislava, 2016
Vladimír Macko

Pod'akovanie:

Abstrakt

Klíčové slová:

Abstract

Keywords:

Obsah

Introduction	1
1 Genome differences	2
1.1 Definitions 2-3p	2
1.2 Comparative genomic 1-2p	2
1.3 Probabilistic approach 2-4	2
1.4 Genome alignments 3-4	2
1.5 Evaluation 2-3	2
2 Read statistics	3
2.1 Reads 1-2	3
2.2 Paired reads 1-2	3
2.3 Benchmarks 3-4	3
3 Catchy name for our tool that will be implemented.	4
3.1 Probabilistic model for differences 8	4
3.2 Implementation problems 8	4
3.3 Used Data 3	4
3.4 Results 12	4
3.5 Future improvements 5	4
4 Discussion 3	5
Appendix A	7

Zoznam obrázkov

Zoznam tabuliek

Introduction

Kapitola 1

Genome differences

1.1 Definitions 2-3p

Here we define theory necessary to for this thesis.

1.2 Comparative genomic 1-2p

Here we show some of current methods used in comparative genomic, which studies the differences between genomes.

1.3 Probabilistic approach 2-4

1.4 Genome alignments 3-4

Brief mentions of current genome alignments approaches.

1.5 Evaluation 2-3

Here we discuss what metrics we want to improve and what is the measure of success for our solution.

Kapitola 2

Read statistics

Here we evaluate some statistics about commonly used data. Here we evaluate current used methods and set benchmarks for our solution.

2.1 Reads 1-2

Here we evaluate data for reads and we set our expectations about these data.

2.2 Paired reads 1-2

Here we evaluate data for paired reads and we set our expectations about these data.

2.3 Benchmarks 3-4

Here we benchmark some used tools for our problem.

Kapitola 3

Catchy name for our tool that will be implemented.

3.1 Probabilistic model for differences 8

3.2 Implementation problems 8

3.3 Used Data 3

3.4 Results 12

3.5 Future improvements 5

Kapitola 4

Discussion 3

Here we will discuss.

Literatúra

[1] Sample paper.

Appendix A

TODO