MASTER'S THESIS

UNIVERSITY OF RENNES 1

BIOINFORMATICS AND GENOMICS MASTER'S DEGREE

(2014 - 2015)

VISUAL REPRESENTATION AND COMPARATIVE EVALUATION OF SCAFFOLDING TOOLS FOR CIRCULAR GENOMES

INSTITUTE FOR RESEARCH IN IT AND RANDOM SYSTEMS, GENSCALE

263 AVENUE GENERAL LECLERC, 35000 RENNES, FRANCE

Author:

Alexandrina BODRUG

Supervisors:

Pr. R. ANDONOV

Pr. D. LAVENIER

April 30, 2015

Thanks

Abbreviations

Contents

1 Introduction **4**

1.1 Context 4

1.2 Scaffolding 4

1.3 Workflow 5

2 Challenges of the scaffolding problem **6**

2.1 Definition 6

2.2 Proposed models 6

2.3 Distinctive features of circular genomes 6

3 Methods for benchmarking **7**

3.1 Available scaffolding tools to benchmark against 7

3.2 Visualization 7

3.3 Comparisons 7

3.3.1 QUAST 7

3.3.2 Comparison function 7

4 Results **8**

4.1 Data sets compared 8

4.2 Visualization 8

4.3 Comparisons 8

4.3.1 QUAST 8

4.3.2 Comparison function 8

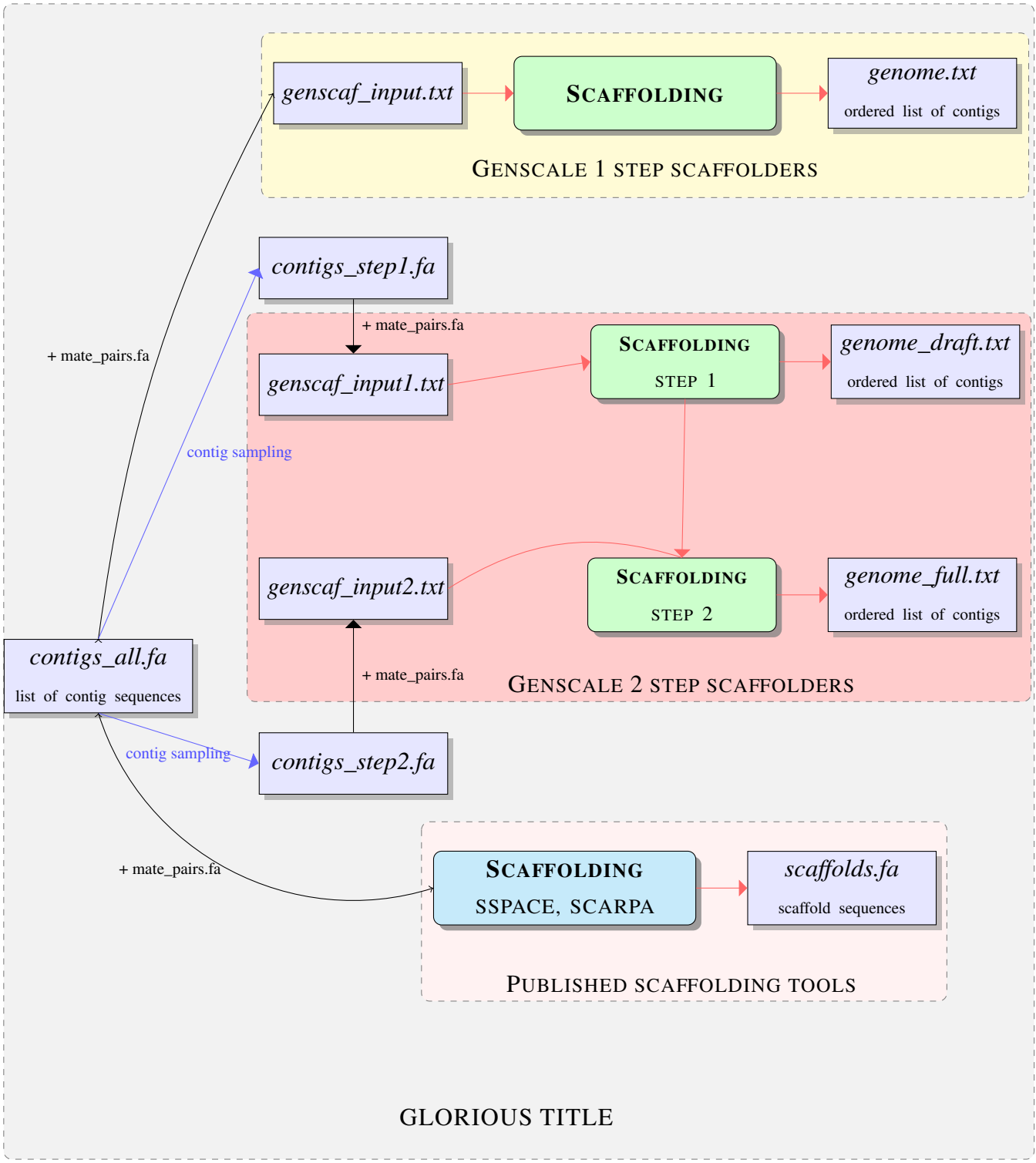
5 Conclusion **9**

1 Introduction

1.1 Context

1.2 Scaffolding

1.3 Workflow


genscaf_input.txt
SCAFFOLDING
genome.txt
ordered list of contigs
GENSCALE 1 STEP SCAFFOLDERS
contigs_step1.fa
+ mate_pairs.fa
+ mate_pairs.fa
genscaf_input1.txt
SCAFFOLDING
STEP 1
genome_draft.txt
ordered list of contigs
contig sampling
genscaf_input2.txt
SCAFFOLDING
STEP 2
genome_full.txt
ordered list of contigs
contigs_all.fa
list of contig sequences
+ mate_pairs.fa
GENSCALE 2 STEP SCAFFOLDERS
contig sampling
contigs_step2.fa
+ mate_pairs.fa
SCAFFOLDING
SSPACE, SCARPA
scaffolds.fa
scaffold sequences
PUBLISHED SCAFFOLDING TOOLS
GLORIOUS TITLE

2 Challenges of the scaffolding problem

2.1 Definition

2.2 Proposed models

2.3 Distinctive features of circular genomes

3 Methods for benchmarking

3.1 Available scaffolding tools to benchmark against

3.2 Visualization

3.3 Comparisons

3.3.1 QUAST

3.3.2 Comparison function

4 Results

4.1 Data sets compared

4.2 Visualization

4.3 Comparisons

4.3.1 QUAST

4.3.2 Comparison function

5 Conclusion