Ministry of Science and Higher Education of the Russian Federation

ITMO University

GRADUATION THESIS

Chromosome scale genome assembly from long noisy reads using Hi-C data.

Author: Anton Andreevich Zamyatin _____
(full name) (signature)

Subject area 01.04.02 Applied Mathematics and Informatics

Degree level Master

Thesis supervisor: Alexeev N.V., PhD, Lead Researcher

_____
(signature)

St. Petersburg, 2020

Student Anton Andreevich Zamyatin
(full name)
Group M42352 Faculty of Information Technologies and Programming

Subject area, program Bioinformatics and Systems Biology

Consultant(s):
Avdeyev P.V., George Washington University_____
(surname, initials, academic title, degree) (signature)

Thesis received "____"_____2020__

Originality of thesis: _____%

Thesis completed with the grade: _____

Date of defense "____"_____20 ____

Secretary of State Exam Commission_____ _____
(full name) (signature)

Number of pages _____

Number of supplementary materials/Blueprints _____

Ministry of Science and Higher Education of the Russian Federation

ITMO University

APPROVED

Head of educational program

_____ _____
(Surname, initials) (signature)

«____» «_____» 20____

OBJECTIVES
FOR A GRADUATION THESIS

Student Anton Andreevich Zamyatin
(full name)

Group M42352 **Faculty** of Information Technologies and Programming

Degree level Master's

Subject area 01.04.02 Applied Mathematics and Informatics

Major Bioinformatics and Systems Biology

Specialization _____

1 Thesis topic Chromosome scale genome assembly from long noisy reads using Hi-C data.

Thesis supervisor Alexeev Nikita Vladimirovich, PhD, Lead Researcher, ITMO University
(full name, place of employment, position, academic degree, academic title)

2 Deadline for submission of complete thesis «_____» «_____» 20_____

3 Requirements and premise for the thesis

The theoretical analysis of the literature on the topic. Performing the best strategy of genome assembly from long nanopore reads and draft assembly polishing using short Illumina reads for two mosquito species. Performing assembly chromosome-level scaffolding using Hi-C data. Genomes assembly assessment and validation. Performing genome assembly for two barnacle species from long pacbio reads. Polishing of assemblies using short Illumina reads. Genomes assembly assessment and validation.

4 Content of the thesis (list of key issues)

a) The terminology used in the thesis and description of main concepts and technologies. b) Mosquitos project. Project introduction, materials, and methods, project results c) Barnacles project. Project introduction, materials, and methods. Project results. d) Conclusion

5 List of graphic materials (with a list of required material)

Graphic materials representing obtained results are provided along within the thesis text. Additional materials for mosquitos project and barnacles project are in appendix A and B respectively.

6 Source materials and publications *reference materials must not be older than 10 years*

7 Objectives issued on «____**» «**_____**» 20**____

Thesis supervisor _____
(signature)

Objectives assumed by _____ «____» «_____» 20____
(signature)

Ministry of Science and Higher Education of the Russian Federation

ITMO University

SUMMARY
OF A GRADUATION THESIS

Student Anton Andreevich Zamyatin
(full name)
Title of the thesis Chromosome scale genome assembly from long noisy reads using Hi-C data.
Name of organization ITMO University

DESCRIPTION OF THE GRADUATION THESIS

1 Research objective Produce and validate two chromosome-scale assemblies for mosquito species. Produce and validate two chromosome-scale assemblies for barnacle species.

2 Research tasks Performing the best strategy of genome assembly from long nanopore reads and draft assembly polishing using short Illumina reads for two mosquito species. Performing assembly chromosome-level scaffolding using Hi-C data. Genomes assembly assessment and validation. Performing genome assembly for two barnacle species from long pacbio reads. Polishing of assemblies using short Illumina reads. Genomes assembly assessment and validation.

3 Number of sources listed in the review section
4 Total number of sources used in the thesis
5 Sources by years:

Russian			**Foreign**		
In the last 5 years	5 to 10 years	More than 10 years	In the last 5 years	5 to 10 years	More than 10 years
-	-	-			

6 Use of online (internet) resources No
7 Use of modern computer software suites and technologies (List which ones were used and for which section of the thesis)

Software suites and technologies	**Thesis section**

8 Short summary of results/conclusions
Different strategies for genome assembly and polishing were performed, results are assessed. Draft mosquito genomes were scaffolded into chromosome-level assemblies. Assemblies were validated. Two barnacle species were assembled with different assemblers, results are assessed. Genomes were polished and validated.

9 Grants received while working on the thesis No

10 Have you produced any publications or conference reports on the topic of the thesis No

Student Anton Andreevich Zamyatin_____
(Full name) (signature)

Thesis supervisor ____Alexeev N.V _____
(Full name) (signature)

“_____”_____20___