

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
БИОЛОГИЧЕСКИЙ ФАКУЛЬТЕТ
Кафедра зоологии

ШУЛИНСКИЙ
Роман Сергеевич

СБОРКА И АННОТАЦИЯ ГЕНОМА *APHIS CRASSIVORA*

Аннотация к дипломной работе

Научный руководитель:
кандидат биологических наук,
доцент Н.В. Воронова

Допущена к защите
«__» _____ 2020 г.
Зав. кафедрой зоологии

доктор биологических наук, профессор
_____ С.В. Буга

Минск, 2020

РЕФЕРАТ

Дипломная работа 36 страниц, 4 рисунка, 4 таблицы, 130 источников.

КЛЮЧЕВЫЕ СЛОВА: ТЛИ, ГЕНОМИКА ТЛЕЙ.

Объект исследования: Геномная ассамблея *Aphis craccivora*.

Целью дипломной работы является сборка и аннотация генома опасного фитофага *A. craccivora* для дальнейшего изучения молекулярных основ устойчивости к инсектицидам и полифагии.

Материалы и методы исследования: Полногеномные прочтения генома *A. craccivora*, собственноручно собранные, с последующей структурной и функциональной аннотацией. **Методы:** компьютерные, биоинформатические.

Размер генома *A. craccivora* составил 336,445 Мб, что соответствует среднему размеру генома тли и имеет следующие статистические показатели: N_{50} – 47498 п.н., размер длиннейшей контиги – 632732 п.н., количество контиг – 17896.

Консенсусный набор генов включает в себя 15957 генов со средней длиной CDS в размере 1541 п.н., средней длиной экзона в размере 215 п.н., средним количеством экзонов в гене в размере 6 и средней длиной гена в размере 5430.

Оценка сборки с использованием SEGMA показала, что 93,47 % потенциально консервативных генов были полностью восстановлены, а 98,31 % из них были частично восстановлены. Кроме того, подход BUSCO показал, что 95,0 % универсальных однокопийных ортологов были аннотированы. Эти особенности предполагают, что сборка генома *A. craccivora* представляет относительно полное содержание генов, которое сопоставимо или превосходит по качеству ранее опубликованные сборки генома тлей.

Функциональная аннотация по международным базам данных нуклеотидных и аминокислотных последовательностей, а также по базам данных белковым доменам показала следующие результаты: UniRef – подтверждено 91,29 % от общего числа генов, RefSeq – подтверждено 88,57 % от общего числа генов, KO – подтверждено 47,56 % от общего числа генов, InterPro – подтверждено 60,24 % от общего числа генов, PFAM – подтверждено 56,38 % от общего числа генов, GO – подтверждено 44,58 % от общего числа генов. Почти 34% генов были поддержаны транскриптомными последовательностями.

РЭФЕРАТ

Дыпломная работа, 36 старонак, 4 малюнкi, 4 табліцы, 130 крыніц.

КЛЮЧАВЫЯ СЛОВЫ: тлі, геноміка тлей.

Аб'ект даследавання: геномная асамблея *Aphis craccivora*.

Мэтай дыпломнай работы з'яўляецца зборка і анатацыя генома небяспечнага фітафага *A. craccivora* для наступнага вывучэння малекулярных асноў устойлівасці да інсектыцыдаў і паліфагіі.

Матэрыялы і метады даследавання: поўнагеномныя чытанні геному *A. craccivora*, ўласнаручна сабраныя, з наступнай структурнай і функцыянальнай анатацыяй. Метады: камп'ютарныя, біяінфарматычныя.

Памер геному *A. craccivora* склаў 336,445 Мб, што адпавядае сярэдняму памеру геному тлі і мае наступныя статыстычныя паказчыкі: N50 – 47498 п.н., памер даўжэйшай паслядоўнасці – 632732 п.н., колькасць паслядоўнасцей – 17896.

Кансэнсусны набор генаў ўключае ў сябе 15957 генаў з сярэдняй даўжынёй CDS, раўнай 1541 п.н., сярэдняй даўжынёй экзонаў – 215 п.н., сярэдняй колькасцю экзонаў ў гене – 6 і сярэдняй даўжынёй гена, раўнай 5430.

Ацэнка зборкі з выкарыстаннем SEGMA паказала, што 93,47 % патэнцыйна кансерватыўных генаў былі цалкам адноўлены, а 98,31 % з іх былі часткова адноўлены. Акрамя таго, падыход BUSCO паказаў, што 95,0 % універсальных аднакапійных арталагаў былі анатаваны. Гэтыя асаблівасці мяркуюць, што зборка геному *A. craccivora* уяўляе адносна поўнае ўтрыманне генаў, якое супастаўна або пераважае па якасці раней апублікаваныя зборкі геному тлей.

Функцыянальная анатацыя па міжнародных базах дадзеных нуклеатыдных і амінакіслотных паслядоўнасцей, а таксама па базах дадзеных бялковых даменаў паказала наступныя вынікі: UniRef – падцверджана 91,29% ад агульнай колькасці генаў, RefSeq – падцверджана 88,57% ад агульнай колькасці генаў, KO – падцверджана 47,56% ад агульнай колькасці генаў, InterPro – падцверджана 60,24% ад агульнай колькасці генаў, PFAM – падцверджана 56,38% ад агульнай колькасці генаў, GO – падцверджана 44,58% ад агульнай колькасці генаў. Амаль 34% генаў былі падтрыманы транскрыптомнымі паслядоўнасцямі.

ABSTRACT

The graduation project includes: pages 36, figures 4, tables 4, sources 130.

Key words: Aphids, Aphids genomics.

Object of study: *Aphis craccivora* genome assembly.

The aim of the thesis is the assembly and annotation of the dangerous phytophage *A. craccivora* to further study the molecular basis of insecticide resistance and polyphagy.

Materials and research methods: Full-genomic readings of the genome of *A. craccivora*, personally assembled, with subsequent structural and functional annotation. Methods: computer, bioinformatics. The genome size of *A. craccivora* was 336.445 Mb, which corresponds to the average size of the aphid genome and has the following statistical indicators: N50 –47498 bp, the length of the longest contig — 632732 bp, the number of contigs — 17896.

The consensus set of genes includes 15,957 genes with an average long CDS of 1541 bp, an average long exon of 215 bp, an average number of exons in a gene of 6, and an average long gene of 5430.

Assembly quality assesment using CEGMA showed that 93.47% of potentially conserved genes were completely restored, and 98.31% of them were partially restored. In addition, the BUSCO approach showed that 95.0% of universal one-copy orthologs of orthologs were annotated. These features suggest that the *A. craccivora* genome assembly represents a relatively complete gene content that is comparable or superior in quality to previously published aphid genome assemblies.

The functional annotation on international databases of nucleotide and amino acid sequences, as well as on databases of protein domains showed the following results: UniRef - confirmed 91.29% of the total number of genes, RefSeq - confirmed 88.57% of the total number of genes, KO - confirmed 47 , 56% of the total number of genes, InterPro– confirmed 60.24% of the total number of genes, PFAM– confirmed 56.38% of the total number of genes, GO– confirmed 44.58% of the total number of genes. Almost 34% of the genes were supported by transcriptome sequences.