

Полногеномное секвенирование энтомопатогенной микроспоридии *Tubulinoosema loxostegi* 2020

Румянцева А.^{1*}, Ефейкин Б.², Токарев Ю.¹

¹ Всероссийский институт защиты растений, Санкт-Петербург, Россия

² Институт проблем экологии и эволюции им. А.Н. Северцова РАН, Москва, Россия

* rumiantseva.arina@yandex.ru

Ключевые слова: микроспоридии; секвенирование; сборка генома

Мотивация и цель: *Tubulinoosema loxostegi* 2020 принадлежит к микроспоридиям, группе облигатных внутриклеточных эукариотических паразитов, родственных грибам. Геномные данные играют важную роль в оценке эволюции вирулентности микроспоридий и поиске взаимосвязи между генетическими различиями разных видов и их гостальной специфичностью [1]. Филогенетический анализ, основанный на гене 16S рРНК, уже недостаточен для идентификации близкородственных видов микроспоридий (сходство данного локуса у представителей рода *Tubulinoosema* превышает 99 %) и разрешения более глубоких взаимоотношений таксона для оценки его реального видового разнообразия. В настоящее время доступна сборка геномов примерно 50 видов микроспоридий, однако эта выборка должна быть намного больше, чтобы считаться надежной и полезной, учитывая известное количество адекватно описанных видов, родов и таксонов более высокого ранга [2]. Целью данной работы было секвенирование генома *T. loxostegi* 2020 для сравнения его основных характеристик с единственным представленным на сегодняшний день референсным геномом родственного вида *T. ratisbonensis*.

Методы и алгоритмы: Микроспоридия *Tubulinoosema loxostegi* 2020 была выделена из лугового мотылька, отловленного в Новосибирской области в 2020 г., и размножена в гусеницах сибирского шелкопряда (*Dendrolimus sibiricus*), содержащихся в лабораторных условиях. Для экстракции геномной ДНК клеточную стенку спор разрушали при помощи жидкого азота. На платформе NovaSeq 4000 Illumina для нашего генома было получено порядка 40 млн прочтений, контроль качества которых осуществлялся при помощи программы FastQC. Выровненные и очищенные от адаптеров прочтения не короче 55 нуклеотидов были использованы для геномной сборки на ассемблере SPAdes. Расчет основных метрик сборки и поиск генов проводили при помощи QUAST и Augustus.

Результаты: Размер полученного генома составил 7 542 987 п.о., распределенных по 1162 контигам с общим содержанием GC 23,12 %, показатель N50 составил ~15 Kb. Размер генома (7.5 Mb), общее количество генов (3086), количество генов, кодирующих белки (3015), рРНК (8) и тРНК (58) практически совпали с таковым для *T. ratisbonensis* (табл. 1), что вполне ожидаемо для двух близкородственных видов.

Сравнительный анализ видов был проведен на основании 6 белок-кодирующих генов, перспективных с точки зрения молекулярных маркеров для дифференциации близкородственных видов, а именно гексокиназы, актина, Hsp70, ДНК-

хеликазы, α - и β -тубулина. Сходство нуклеотидных и аминокислотных последовательностей большинства из них колебалось в пределах 98–99 и 98–100 % соответственно. Для гексокиназы и ДНК-хеликазы эти показатели были несколько ниже, что указывает на более высокий полиморфизм данных локусов в пределах рода *Tubulinosema* (табл. 2).

Таблица 1. Сравнение сборок *T. loxostegi* 2020 (SPAdes) и *T. ratisbonensis* (IDBA-UD)

Характеристика	<i>T. loxostegi</i> 2020	<i>T. ratisbonensis</i>
Номер доступа в Генбанке	–	GCA_004000155.1
Число контигов	1162	1011
Самый длинный контиг Mb	1.0	–
Размер генома Mb	7.5	7.6
Количество генов	3086	3082
Количество белков	3015	3013
Количество рРНК	8	8
Количество тРНК	58	55
N50, kb	14.981	11.9
N70, kb	6.887	–
GC, %	23.12	23
Покрытие генома	180.0x	200.0x
Количество неопределенных пар оснований (N's)	11	–
Доля неопределенных пар оснований в 100 Kb сборки	0.67	–
Ссылка	Данная работа	Polonais et al., 2019 [3]

Таблица 2. Сравнение некоторых генов «домашнего хозяйства» *T. loxostegi* 2020 и *T. ratisbonensis*

Локус	Сходство нуклеотидной / аминокислотной последовательности гена, %
Гексокиназа (1190 п.о.)	96.64 / 98.48
Актин (1128 п.о.)	98.85 / 98.94
Hsp70 (2002 п.о.)	98.50 / 99.70
α -тубулин (1341 п.о.)	98.36 / 99.11
β -тубулин (1320 п.о.)	98.48 / 100
ДНК-хеликаза (1748 п.о.)	95.88 / 93.48

Выводы: Впервые проведены полногеномное секвенирование и предварительная сборка генома для энтомопатогенной микроспоридии из лугового мотылька. О полноте собранного генома свидетельствует сопоставимость основных характеристики сборки *T. loxostegi* 2020 и близкородственного вида *T. ratisbonensis*. Сравнительный анализ шести генов «домашнего хозяйства» позволил выявить несколько перспективных молекулярных маркеров для дальнейшей дифференциации близкородственных видов рода *Tubulinosema*, в частности гены гексокиназы и ДНК-хеликазы, показавшие более высокие различия в нуклеотидных последовательностях, чем ген мсрРНК.

Финансирование: Исследование поддержано грантом РФ (№ 23-16-00262).

Whole genome sequencing of *Tubulinosema loxostegi* 2020

Rumiantseva A.^{1*}, Efeikin B.², Tokarev Y.¹

¹ All-Russian Institute of Plant Protection, St. Petersburg, Russia

² A.N. Severtsov Institute of Ecology and Evolution, RAS, Moscow, Russia

* rumiantseva.arina@yandex.ru

Key words: microsporidia; sequencing; genome assembly

Motivation and Aim: *Tubulinosema loxostegi* 2020 belongs to the Microsporidia, a group of obligate intracellular eukaryotic parasites related to fungi. Genomic data play an important role in assessing the evolution of microsporidia virulence and searching for the relationship between genetic differences among different species and their host specificity [1]. Phylogenetic analysis based on the 16S rRNA gene is no longer sufficient to identify closely related microsporidia species (the similarity of this locus in representatives of the genus *Tubulinosema* exceeds 99 %) and to resolve deeper relationships of the taxon to assess its real species diversity. Genome assemblies of approximately 50 microsporidia species are currently available, but this sample would need to be much larger to be considered reliable and useful given the known number of species, genera, and higher-ranking taxa adequately described to date. [2]. The purpose of this work was to assemble of the *T. loxostegi* 2020 genome to compare its main characteristics with the only reference genome of the related species *T. ratisbonensis* presented to date.

Methods and Algorithms: The microsporidium *Tubulinosema loxostegi* 2020 was isolated from the beet webworm larvae (*Loxostegi sticticalis*) captured in the Novosibirsk region in 2020. Spores were propagated in the Siberian silkworm (*Dendrolimus sibiricus*) larvae under laboratory conditions. To extract genomic DNA, the cell wall of the spores was destroyed using liquid nitrogen. Approximately 40 million reads were obtained for our genome on the NovaSeq 4000 Illumina platform, and their quality was assessed using the FastQC program. Reads that were aligned and free of adapters, with a minimum length of 55 nucleotides, were utilized for genome assembly using the SPAdes assembler. The main assembly metrics and gene research were calculated using QUAST and Augustus.

Results: The resulting genome size was 7,542,987 bp, distributed over 1,162 contigs with a total GC content of 23.12 %. The N50 indicator was approximately 15 Kb. The genome size (7.5 Mb), total number of genes (3086), number of genes encoding proteins (3015), rRNA (8), and tRNA (58) closely matched those of *T. ratisbonensis* (Table 1), as expected for two closely related species.

A comparative analysis of species was conducted based on six protein-coding genes, which show promise as molecular markers for differentiating closely related species. These genes include hexokinase, actin, Hsp70, DNA helicase, α -tubulin, and β -tubulin. The similarity of the nucleotide and amino acid sequences for most of these genes ranged from 98 to 99 % and 98 to 100 %, respectively. However, for hexokinase and DNA helicase, these indicators were slightly lower, suggesting a higher level of polymorphism within the genus *Tubulinosema* at these locus (Table 2).

Table 1. Comparison of *T. loxostegi* 2020 (SPAdes) and *T. ratisbonensis* (IDBA-UD) assemblies

Metrics	<i>T. loxostegi</i> 2020	<i>T. ratisbonensis</i>
GenBank accession number	–	GCA_004000155.1
Number of contigs	1162	1011
Largest contig, Mb	1.0	–
Total assembly length, Mb	7.5	7.6
Number of genes	3086	3082
Number of proteins	3015	3013
Number of rRNA	8	8
Number of tRNA	58	55
N50, kb	14.981	11.9
N70, kb	6.887	–
GC, %	23.12	23
Genome coverage	180.0x	200.0x
N's	11	–
N's per 100 Kbp	0.67	–
Link	–	Polonais et al., 2019 [3]

Table 2. Comparison of some housekeeping genes of *T. loxostegi* 2020 and *T. ratisbonensis*

Locus	Gene nucleotide/amino acid sequence similarity, %
Hexokinase (1190 bp)	96.64 / 98.48
Actin (1128 bp)	98.85 / 98.94
Hsp70 (2002 bp)	98.50 / 99.70
α - tubulin (1341 bp)	98.36 / 99.11
β - tubulin (1320 bp)	98.48 / 100
DNA helicase (1748 bp)	95.88 / 93.48

Conclusion: For the first time, whole-genome sequencing and preliminary genome assembly were carried out for the entomopathogenic microsporidium from the beet webworm larvae. The completeness of the assembled genome is evidenced by the comparability of the main characteristics of the assembly with those of the closely related species *T. ratisbonensis*. A comparative analysis of six housekeeping genes made it possible to identify several promising molecular markers for further differentiation of closely related species within the genus *Tubulinosema*, particularly the hexokinase and DNA helicase genes, which exhibited higher differences in nucleotide sequences than the 16s rRNA gene.

Funding: The study is supported by RSF (# 23-16-00262).

Список литературы/References

1. Willis A.R., Reinke A.W. Factors That Determine Microsporidia Infection and Host Specificity. In: Weiss L.M., Reinke A.W. (Eds.). *Microsporidia. Experientia Supplementum*. Vol. 114. Springer, 2022;91-114
2. Bojko J. et al. Microsporidia: a new taxonomic, evolutionary, and ecological synthesis. *Trends Parasitol.* 2022;38(8):642-659
3. Polonais V. et al. Draft genome sequence of *tubulinosema ratisbonensis*, a microsporidian species infecting the model organism *drosophila melanogaster*. *Microbiol Resour Announc.* 2019;8(31):e00077-19. doi 10.1128/mra.00077-19