

Анализ генетических отличий некоторых патоваров *Pseudomonas syringae*

С.И. ПРИХОДЬКО¹, А.Б. ЯРЕМКО²,
Е.В. СТАРИКОВА³, К.П. КОРНЕВ⁴

ФГБУ «Всероссийский центр карантина растений»
(ФГБУ «ВНИИКР»), р. п. Быково, г. Раменское,
Московская обл., Россия

¹ ORCID 0000-0002-1281-4410,
e-mail: svetlana.prik@yandex.ru

² ORCID 0000-0003-3295-8080, e-mail: an_ya94@mail.ru

³ ORCID 0000-0001-6582-210X,
e-mail: hed.robin@gmail.com

⁴ ORCID 0000-0002-3490-1857,
e-mail: konstantin.kornev@gmail.com

АННОТАЦИЯ

Pseudomonas syringae pv. *maculicola* является фитопатогеном, поражающим крестоцветные растения во всем мире. *P. s.* pv. *maculicola* регулируется фитосанитарными требованиями следующих стран: Китай, Израиль, Мексика, Египет, Судан и Индонезия. Таким образом, возникла необходимость в установлении соответствия мест производства и экспортной продукции требованиям стран – торговых партнеров России методами лабораторной диагностики. В связи с высоким генетическим сходством патоваров вида *P. syringae*, при проведении лабораторной диагностики особый интерес представляет необходимость отличить патовар *maculicola* от патовара *tomato*. Данные бактерии могут одновременно присутствовать на растениях семейства Крестоцветные (Капустные) и вызывать бактериоз. Также трудности в установлении возбудителя бактериоза на крестоцветных растениях представляет бактерия *P. cannabina* pv. *alisalensis*. Этот вид имеет широкий круг растений-хозяев, включая крестоцветные. Кроме того, ареалы хозяев и симптомы проявления заболевания *P. s.* pv. *maculicola* и *P. c.* pv. *alisalensis* совпадают, что также может привести к путанице в определении возбудителя. С целью анализа генетических отличий некоторых патоваров *Pseudomonas syringae*, а также поиска специфических генетических маркеров, которые могут быть использованы при проведении лабораторной диагностики *P. s.* pv. *maculicola*, нами был проведен поиск предположительной мишени путем изучения белков, соответствующих 293 общедоступным геномным сборкам некоторых патоваров вида *P. syringae*, а также 43 геномным сборкам близкородственных видов *P. cannabina* и *P. savastanoi*. Анализ показал высокое сходство большинства анализируемых последовательностей с последовательностями патовара *P. c.* pv. *alisalensis*.

Ключевые слова. Генетический тест, геномные сборки, белок, филогенетическое древо.

Analysis of genetic differences of some *Pseudomonas syringae* pathovars

S.I. PRIKHODKO¹, A.B. YAREMKO², E.V. STARIKOVA³,
K.P. KORNEV⁴

FGBU "All-Russian Plant Quarantine Center"
(FGBU "VNIICR"), Bykovo, Ramenskoye,
Moscow Oblast, Russia

¹ ORCID 0000-0002-1281-4410,
e-mail: svetlana.prik@yandex.ru

² ORCID 0000-0003-3295-8080, e-mail: an_ya94@mail.ru

³ ORCID 0000-0001-6582-210X,
e-mail: hed.robin@gmail.com

⁴ ORCID 0000-0002-3490-1857,
e-mail: konstantin.kornev@gmail.com

ABSTRACT

Pseudomonas syringae pv. *maculicola* is a phytopathogen affecting cruciferous plants worldwide. *P. s.* pv. *maculicola* is regulated by the phytosanitary requirements of the following countries: China, Israel, Mexico, Egypt, Sudan and Indonesia. Thus, there was a need to establish the compliance of the places of production and export products with the requirements of the countries, Russia's trading partners, using laboratory diagnostics methods. Due to the high genetic similarity of *P. syringae* pathovars, when conducting laboratory diagnostics, of particular interest is the need to distinguish *maculicola* pathovar from *tomato* pathovar. These bacteria can simultaneously be present on plants of the Cruciferous (Cabbage) family and cause bacteriosis. It is also difficult to identify the causative agent of bacteriosis on cruciferous plants, the bacterium *P. cannabina* pv. *alisalensis*. This species has a wide range of host plants, including cruciferous plants. In addition, host plants areas and symptoms of *P. s.* pv. *maculicola* and *P. c.* pv. *alisalensis* coincide, which can also lead to confusion in the identification. In order to analyze the genetic differences of some *Pseudomonas syringae* pathovars, as well as the search for specific genetic markers that can be used in laboratory diagnosis of *P. s.* pv. *maculicola*, we searched for a potential target by studying proteins corresponding to 293 available genomic assemblies of some *P. syringae* pathovars, as well as 43 genomic assemblies of closely related species *P. cannabina* and *P. savastanoi*. The analysis showed a high similarity of most of the analyzed sequences with the sequences of *P. c.* pv. *alisalensis* pathovar.

Key words. Genetic test, genomic assemblies, protein, phylogenetic tree.

ВВЕДЕНИЕ



Бактерии рода *Pseudomonas* относятся к классу Gammaproteobacteria, семейству Pseudomonadaceae и были впервые описаны Migula (1894 г.). Впоследствии развитие молекулярных методов исследований привело к значительному пересмотру классификации псевдомонад. Группы псевдомонад были сформированы на основе исследований последовательностей гена 16S рРНК (Anzai et al., 2000). На сегодняшний день род *Pseudomonas* считается наиболее сложным в плане диагностики. Некоторые виды псевдомонад являются возбудителями заболеваний сельскохозяйственных и декоративных растений. Наиболее распространенный представитель фитопатогенов данного рода – вид *P. syringae* (<https://www.cabi.org>, 2021). Бактерии этого вида способны колонизировать филосферу широкого круга как культурных, так и дикорастущих растений (Morriset al., 2007). Симптомы заболеваний, вызываемых *P. syringae*, разнообразны и проявляются в форме гнилей, пятнистостей, галловых образований (Sarkar, Guttman, 2004).

Вид *P. syringae* подразделяется на патовары в зависимости от растения-хозяина. Тем не менее неоднократно были установлены случаи поражения одним бактериальным изолятом патовара разных растений-хозяев, что привело исследователей к необходимости пересмотра классификации вида. Сложность в дифференциации патоваров также состоит в их близком филогенетическом родстве (Hwang et al., 2005; Gironde, Manceau, 2012). На сегодняшний день на основе гомологии ДНК выделяют 9 геномовидов *P. syringae* (Gardan et al., 1999; Ilić et al., 2016).

Pseudomonas syringae pv. *maculicola* поражает растения семейства Капустные и является причиной спорадических эпифитотий бактериальной пятнистости в посевах различных крестоцветных культур во всем мире, в том числе и на территории России. Необходимость в диагностике *P. s.* pv. *maculicola* возникла в связи увеличением объемов экспорта масличного рапса. Места производства экспортной продукции должны быть свободны от вредных организмов, регулируемых страной-импортером. Для установления соответствия таким требованиям проводятся мониторинги распространения вредных организмов, а также исследования продукции, предназначенной на экспорт. *P. s.* pv. *maculicola* регулируется фитосанитарными требованиями следующих стран: Китай, Мексика, Египет, Израиль, Судан и Индонезия (Приходько и др., 2021). При проведении диагностики в лаборатории особый интерес представляет необходимость отличить патовар *P. s.* pv. *maculicola* от *tomato*, поскольку эти бактерии могут одновременно присутствовать на растениях и вызывать бактериоз (Cuppels, Ainsworth, 1995; Gironde, Manceau, 2012). Также близкий вид *P. cannabina* pv. *alisalensis* имеет широкий круг растений-хозяев, включая крестоцветные. Ареалы хозяев *P. s.* pv. *maculicola* и *P. cannabina* pv. *alisalensis* совпадают. *P. cannabina* pv. *alisalensis* вызывает бактериальную пятнистость и ожог растений семейств Brassicaceae и Poaceae. Сходный круг поражаемых растений и симптоматика проявления заболевания

INTRODUCTION

Bacteria of the genus *Pseudomonas* belong to the class Gammaproteobacteria, family Pseudomonadaceae and were first described by Migula (1894). Subsequently, the development of molecular research methods led to a significant revision of the classification of pseudomonads. Pseudomonas groups were formed on the basis of studies of the 16S rRNA gene sequences (Anzai et al., 2000). Today, the genus *Pseudomonas* is considered the most difficult to diagnose. Some species of pseudomonas are causative agents of diseases of agricultural and ornamental plants. The most common representative of phytopathogens of this genus is the species *P. syringae* (<https://www.cabi.org>, 2021). Bacteria of this species are able to colonize the phyllosphere of a wide range of both cultivated and wild plants (Morriset al., 2007). Symptoms of diseases caused by *P. syringae* are varied and manifest in the form of rot, spots, gall formations (Sarkar, Guttman, 2004).

P. syringae is subdivided into pathovars depending on a host plant. Nevertheless, cases of damage by one bacterial isolate of pathovar to different host plants have been repeatedly established, which led researchers to the need to revise the classification of the species. The difficulty in differentiating pathovars also lies in their close phylogenetic relationship (Hwang et al., 2005; Gironde, Manceau, 2012). To date, based on DNA homology, 9 genomospecies of *P. syringae* have been isolated (Gardan et al., 1999; Ilić et al., 2016).

Pseudomonas syringae pv. *maculicola* infects plants of the Cabbage family and is the cause of sporadic epiphytotic of bacterial spotting in crops of various cruciferous crops throughout the world, including in Russia. The need for diagnosis of *P. s.* pv. *maculicola* arose due to the increase in exports of oilseed rape. Places of production of export products must be free from pests regulated by the importing country. To establish compliance with such requirements, monitoring of the pests spread is carried out, as well as studies of products intended for export. *P. s.* pv. *maculicola* is regulated by the phytosanitary requirements of the following countries: China, Mexico, Egypt, Israel, Sudan and Indonesia (Prihodko et al., 2021). When conducting diagnosis in the laboratory, it is necessary to distinguish the pathovar *P. s.* pv. *maculicola* from *tomato*, since these bacteria can simultaneously be present on plants and cause bacteriosis (Cuppels, Ainsworth, 1995; Gironde, Manceau, 2012). Also, the closely related species *P. cannabina* pv. *alisalensis* has a wide range of host plants, including cruciferous. Host plants areas of *P. s.* pv. *maculicola* and *P. cannabina* pv. *alisalensis* coincide. *P. cannabina* pv. *alisalensis* causes bacterial blotch and blight on plants of the families Brassicaceae and Poaceae. A similar range of affected plants and symptoms of the manifestation of the disease often lead to confusion in identifying the pathogen only by classical microbiological and biochemical methods (Sakata et al., 2021). The aim of our research was to analyze the genetic differences of some *Pseudomonas syringae* pathovars.

часто приводят к путанице в определении возбудителя только классическими микробиологическими и биохимическими методами (Sakata et al., 2021). Целью наших исследований являлся анализ генетических отличий некоторых патоваров *Pseudomonas syringae*.

МАТЕРИАЛЫ И МЕТОДЫ

Подбор геномных сборок бактерий рода *Pseudomonas*. Анализ генетических отличий некоторых патоваров *Pseudomonas syringae* осуществляли путем исследования аннотированных белков, соответствующих 293 общедоступным геномным сборкам из базы NCBI Reference Sequence Database (NCBI RefSeq), а также 43 геномным сборкам близкородственных видов *P. cannabina* и *P. savastanoi* (табл. 1).

Кластеризация белковых последовательностей патоваров. Белковые последовательности, соответствующие геномным сборкам, помечали в соответствии с видом и патоваром, затем производили кластеризацию последовательностей при помощи биоинформатического инструмента CD-HIT (v. 4.1). Данный инструмент производит группировку последовательностей в кластеры по заданному порогу сходства на основе соответствия распределений коротких подпоследовательностей-пептидов («слов»).

Использовались следующие пороги сходства для каждой из последовательностей, попадающей в кластер: не менее 70% идентичности другим последовательностям кластера и не менее 70% от длины наиболее длинной последовательности кластера. Это условный порог, позволяющий не отбрасывать последовательности, слегка различающиеся между собой по длине, что является допустимым, и в то же время избежать попадания в кластер коротких последовательностей с частичным сходством с анализируемыми – такие последовательности затрудняют анализ. Такой порог длины позволяет анализировать примерно «равные» последовательности и минимизировать количество «шума».

Длина используемого «слова» составила 5 аминокислотных остатков. Такая длина является желательной длиной при данном пороге сходства согласно рекомендациям авторов алгоритма CD-HIT. Такая идентичность является минимальным порогом сходства, на котором работает «белковая» версия алгоритма CD-HIT. Такой порог позволил проанализировать максимальное количество кластеров. Повышение порога приводит к потере данных, а снижение его невозможно.

Выравнивание белковых и нуклеотидных последовательностей избранных кластеров и построение деревьев. Для каждого из 12 избранных белковых кластеров было проведено множественное выравнивание последовательностей с использованием инструмента MUSCLE (v. 3.8.31) (https://drive5.com/muscle/downloads_v3.htm). На основе полученных выравниваний были построены филогенетические деревья с использованием методов максимального правдоподобия с помощью пакета PhyML (v. 3.0) (Guindon et al., 2010). В качестве модели замен использовалась модель JTT. Данная модель была выбрана автоматически инструментом PhyML как наиболее подходящая для данного

MATERIALS AND METHODS

Selection of genomic assemblies of bacteria of the genus *Pseudomonas*. The analysis of genetic differences of some *Pseudomonas syringae* pathovars was performed by examining annotated proteins corresponding to 293 available genomic assemblies from the NCBI Reference Sequence Database (NCBI RefSeq), as well as 43 genomic assemblies of the closely related species *P. cannabina* and *P. savastanoi* (Table 1).

Таблица 1

Использованные штаммы и изоляты *Pseudomonas syringae* pv. *maculicola*
(<https://www.ncbi.nlm.nih.gov/refseq/>)

Table 1
Used strains and isolates of *Pseudomonas syringae* pv. *maculicola*
(<https://www.ncbi.nlm.nih.gov/refseq/>)

№ п/п	Геномная сборка <i>P. s. pv. maculicola</i> Genomic assembly <i>P. s. pv. maculicola</i>	Штамм/ изолят Strain/ Isolate	Происхождение/ растение Origin/Plant
1	GCF_000145845.2	ES4326	США USA/ <i>Raphanus sativus</i> L.
2	GCF_000935725.1	CFBP 1657	Новая Зеландия New Zealand/ <i>Brassica oleracea</i> var. <i>botrytis</i> L.
3	GCF_001293575.1	YM7930	Япония Japan/ <i>Raphanus sativus</i> L.
4	GCF_001293855.1	90_32	США USA/ <i>Brassica oleracea</i> var. <i>botrytis</i> L.
5	GCF_001293925.1	H7608	Япония Japan/ <i>Brassica rapa</i> subsp. <i>pekinensis</i> (Lour.) Hanelt
6	GCF_001294165.1	M6	Великобритания UK/ <i>Brassica oleracea</i> var. <i>botrytis</i> L.
7	GCF_001294185.1	KN91	–/ <i>Raphanus sativus</i> L.
8	GCF_001294305.1	M4a	США USA/ <i>Raphanus sativus</i> L.
9	GCF_001400855.1	ICMP 3935	Новая Зеландия New Zealand/ <i>Brassica oleracea</i> var. <i>botrytis</i> L.
10	GCF_003699635.1	4981	Зимбабве Zimbabwe/ <i>Brassica oleracea</i> var. <i>botrytis</i> L.
11	GCF_003700765.1	ICMP 2744	Великобритания UK/ <i>Brassica nigra</i> (L.) W.D.J. Koch
12	GCF_003702755.1	ICMP 11281	Китай China/ <i>Brassica rapa</i> L.

набора последовательностей. Модель использовалась для построения всех деревьев.

Визуализация полученных деревьев проводилась с использованием инструмента TreeDyn (v. 198.3) (http://phylogeny.lirmm.fr/phylo.cgi/one_task.cgi?task_type=treedyn).

Выравнивание репрезентативных белковых последовательностей каждого из отобранных в ходе анализа 8 кластеров проводили при помощи алгоритма BLASTP (NCBI).

Выравнивание нуклеотидных последовательностей проводилось при помощи инструмента Muscle (v. 3.8). Визуализация результатов выравнивания производилась в MView (v. 1.20.1) (<https://www.ebi.ac.uk/Tools/msa/mview/>).

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

В результате кластеризации 1 760 372 белковых последовательностей с помощью инструмента CD-HIT было получено 38 498 кластеров, содержащих от 1 до 920 аминокислотных последовательностей различной длины. Полученные кластеры анализировали с помощью скрипта на языке Python, позволяющего находить кластеры с заданными параметрами в большом массиве текстовых данных.

Анализ полученных кластеров показал отсутствие белковых последовательностей, специфичных только для патовара *P. s. pv. maculicola*. Часто белковые последовательности патовара *P. syringae* *pv. maculicola* кластеризовались вместе с патоваром *P. s. pv. tomato*. Ввиду близости данных патоваров, а также сложности их дифференцирования, что могло привести к ошибочному аннотированию патоваров в NCBI RefSeq, мы производили поиск кластеров, содержащих максимальное количество белков патоваров *P. s. pv. maculicola* и *P. s. pv. tomato* и минимальное количество белков других патоваров. В результате такого анализа нам удалось обнаружить 2 кластера, содержащие белковые последовательности только *P. s. pv. maculicola* и *P. s. pv. tomato* и не содержащие белковых последовательностей других патоваров. В кластерах, содержащих последовательности *P. s. pv. maculicola* и *P. s. pv. tomato*, также часто встречались последовательности патоваров *P. s. pv. lachrymans* и *P. cannabina* *pv. alisalensis*. Данные анализа избранных белковых кластеров приведены в таблице 2.

Как видно из данных, представленных в таблице 2, избранные белковые кластеры разделились на II типа. Кластеры I типа присутствовали в 10 из 12 проанализированных штаммов *P. s. pv. maculicola*, в 5 из 20 проанализированных штаммов *P. s. pv. tomato*, во всех 4 проанализированных штаммах *P. s. pv. alisalensis* и в 2 из 11 проанализированных штаммов *P. s. pv. lachrymans*. Кластеры II типа присутствовали в 7 из 12 проанализированных штаммов *P. s. pv. maculicola*, в 9 из 20 проанализированных штаммов *P. s. pv. tomato* и отсутствовали в проанализированных штаммах *P. s. pv. alisalensis* и *P. s. pv. lachrymans*. Подобные результаты ставят под вопрос корректность аннотаций использованных геномных сборок, депонированных в NCBI RefSeq.

В качестве примера ниже представлены 4 из 12 филогенетических деревьев, построенных для кластеров обоих типов (рис. 1).

По результатам анализа филогенетических деревьев нами были исключены кластеры,

Clustering of pathovars protein sequences. Protein sequences corresponding to genomic assemblies were labeled according to species and pathovar, then the sequences were clustered using the bioinformatics tool CD-HIT (v. 4.1). This tool groups sequences into clusters according to a given threshold of similarity based on the correspondence of distributions of short peptide subsequences ("words").

The following similarity thresholds were used for each of the sequences in the cluster: at least 70% identity to other cluster sequences and at least 70% of the length of the longest cluster sequence. This is a conditional threshold that makes it possible not to discard sequences that slightly differ in length from each other, which is acceptable, and at the same time to avoid getting into the cluster of short sequences with partial similarity to the analyzed ones – such sequences make analysis difficult. Such a length threshold allows us to analyze approximately "equal" sequences and minimize the amount of "noise".

The length of the used "word" was 5 amino acid residues. This length is the desired length for a given similarity threshold, as recommended by the authors of the CD-HIT algorithm. This identity is the minimum similarity threshold on which the "protein" version of the CD-HIT algorithm works. This threshold made it possible to analyze the maximum number of clusters. Raising the threshold results in data loss, but lowering it is not possible.

Alignment of protein and nucleotide sequences of selected clusters and construction of trees. For each of the 12 selected protein clusters, multiple sequence alignments were performed using the MUSCLE tool (v. 3.8.31) (https://drive5.com/muscle/downloads_v3.htm). Based on the resulting alignments, phylogenetic trees were constructed using maximum likelihood methods using the PhyML (v. 3.0) package (Guindon et al., 2010). The JTT model was used as a substitution model. This model was automatically selected by the PhyML tool as the most suitable for this set of sequences. The model was used to build all trees.

The resulting trees were visualized using the TreeDyn tool (v. 198.3) (http://phylogeny.lirmm.fr/phylo.cgi/one_task.cgi?task_type=treedyn).

Alignment of representative protein sequences of each of the 8 clusters selected during the analysis was performed using the BLASTP algorithm (NCBI).

Alignment of nucleotide sequences was carried out using the Muscle tool (v. 3.8). Alignment results were visualized in MView (v. 1.20.1) (<https://www.ebi.ac.uk/Tools/msa/mview/>).

RESULTS AND DISCUSSION

As a result of clustering 1,760,372 protein sequences using the CD-HIT tool, 38,498 clusters were obtained containing from 1 to 920 amino acid sequences of various lengths. The resulting clusters were analyzed using a Python script that allows finding clusters with given parameters in a large array of text data.

Analysis of the resulting clusters showed the absence of protein sequences specific only for the pathovar *P. s. pv. maculicola*. Often the protein sequences of

Таблица 2
Избранные кластеры с последовательностями патовара
P. syringae* pv. *maculicola

Table 2
Selected clusters with *P. syringae* pv. *maculicola* pathovar sequences

Кластер Cluster	Аннотация Annotation	Длина Length	N. mac	N. tom	N. ali	N. lach
1415	TonB-dependent receptor	785	10/12	5/20	4/4	2/11
9438	DJ-1/Pfpl family protein	333	10/12	5/20	4/4	2/11
11492	MBL fold metallo-hydrolase	296	10/12	5/20	4/4	2/11
12691	DJ-1/Pfpl family protein	276	10/12	5/20	4/4	2/11
17663	putative natural product biosynthesis protein	208	10/12	5/20	4/4	2/11
18924	putative natural product biosynthesis protein	193	10/12	5/20	4/4	2/11
19719	OsmC family protein	184	7/12	9/20	0/4	0/11
21897	copper chaperone PCu(A)C	162	10/12	5/20	4/4	2/11
22842	DUF2946 domain-containing protein	154	10/12	5/20	4/4	2/11
26276	DUF2946 domain-containing protein	127	10/12	5/20	4/4	2/11
28504	carboxymuconolactone decarboxylase family protein	110	7/12	9/20	0/4	2/11
30723	cytochrome c5 family protein	95	10/12	5/20	0/4	2/11

«Кластер» – номер кластера. «Аннотация» – аннотация репрезентативной последовательности в NCBI RefSeq. «Длина» – максимальная длина последовательности в кластере в аминокислотных остатках. «N. mac (tom/ali/lach)» – количество штаммов патовара *maculicola* (*tomato/alisalensis/lachrymans*), в которых присутствует данная белок-кодирующая последовательность, от общего количества проанализированных штаммов данного патовара.

“Cluster” – cluster number. “Annotation” is an annotation of a representative sequence in NCBI RefSeq. “Length” – the maximum length of the sequence in the cluster in amino acid residues. “N. mac (tom/ali/lach)” – number of pathovar *maculicola* (*tomato/alisalensis/lachrymans*) strains, containing this protein-coding sequence, from the total number of analyzed strains of this pathovar.

не позволяющие отличать последовательности патоваров *maculicola* и *tomato* от патовара *alisalensis*. Возможные пересечения маркерных последовательностей с последовательностями данного патовара нежелательны ввиду схожести спектра поражаемых культур и симптомов поражения. Таким образом, было отобрано 6 кластеров: 9438, 12691, 11492, 19719, 28504 и 30723.

При анализе результатов выравнивания последовательностей этих кластеров было установлено высокое сходство большинства анализируемых последовательностей с последовательностями патовара *P. s. pv. alisalensis*, а также возник вопрос о корректности аннотации патоваров использованных геномных сборок, депонированных в NCBI RefSeq.

С целью поиска стратегии решения данных проблем нами была проведена перепроверка использовавшихся в анализе геномных сборок *P. s. pv. maculicola* и *P. s. pv. tomato*. В результате из анализа были исключены 3 сборки *P. s. pv. maculicola*, имеющие статус “Taxonomy check: Inconclusive”. Такой статус означает, что средняя нуклеотидная идентичность (Average Nucleotide Identity, ANI) для

the pathovar *P. syringae* pv. *maculicola* were clustered together with the pathovar *P. s. pv. tomato*. Due to the proximity of these pathovars, as well as the complexity of their differentiation, which could lead to erroneous annotation of pathovars in NCBI RefSeq, we searched for clusters containing the maximum number of proteins of the pathovars *P. s. pv. maculicola* and *P. s. pv. tomato*, and the minimum number of proteins of other pathovars. As a result of this analysis, we were able to detect 2 clusters containing protein sequences only of *P. s. pv. maculicola* and *P. s. pv. tomato* and not containing protein sequences of other pathovars. In clusters containing sequences of *P. s. pv. maculicola* and *P. s. pv. tomato*, there were often sequences of the pathovars *P. s. pv. lachrymans* and *P. cannabina* pv. *alisalensis*. Analysis data for selected protein clusters are shown in Table 2.

As can be seen from the data presented in Table 2, the selected protein clusters were divided into two types. Type I clusters were present in 10 of the 12 analyzed strains of *P. s. pv. maculicola*, in 5 of the 20 analyzed strains of *P. s. pv. tomato*, in all 4 analyzed strains of *P. c. pv. alisalensis* and in 2 of the 11 analyzed strains of *P. s. pv. lachrymans*. Type II clusters were present in 7 of the 12 analyzed strains of *P. s. pv. maculicola*, in 9 of the 20 analyzed strains of *P. s. pv. tomato* and absent in the analyzed strains of *P. c. pv. alisalensis* and *P. s. pv. lachrymans*. Such results call into question the correct-

ness of the annotations of the used genomic assemblies deposited in NCBI RefSeq.

As an example, 4 out of 12 phylogenetic trees constructed for clusters of both types are presented below (Fig. 1).

Based on the results of the analysis of phylogenetic trees, we excluded clusters that did not allow us to distinguish the sequences of the pathovars *maculicola* and *tomato* from the pathovar *alisalensis*. Possible intersections of marker sequences with sequences of this pathovar are undesirable due to the similarity of the spectrum of affected crops and the symptoms of the disease. Thus, 6 clusters were selected: 9438, 12691, 11492, 19719, 28504 и 30723.

When analyzing the results of alignment of the sequences of these clusters, a high similarity of most of the analyzed sequences with the sequences of the pathovar *P. c. pv. alisalensis* was found, and also there was a question about the correctness of the annotation of

данной сборки не соответствует ожидаемой. Таким образом, из анализа были исключены геномные сборки GCF_001293575.1, GCF_001294185.1 и GCF_000145845.2 (табл. 1). Белки именно этих геномных сборок проявляли наибольшее сходство с последовательностями *P. c. pv. alisalensis*.

Нами были перестроены филогенетические деревья для 12 ранее отобранных кластеров с обновленными последовательностями. На рисунке 2 представлены 4 из 12 филогенетических деревьев с обновленными последовательностями.

Сравнивая приведенные на рисунках 1 и 2 деревья, можно сделать вывод, что фильтрация геномных сборок с неподтвержденной аннотацией позволила исключить последовательности *P. c. pv. alisalensis* из клад, в которых находятся последовательности *P. s. pv. maculicola* и *P. s. pv. tomato*, лишь для части кластеров. Таким образом, из анализа были исключены последовательности кластеров 22842, 21897, 18924, 17663.

Для оставшихся кластеров наблюдалась вариативность последовательностей *P. s. pv. maculicola*. Несмотря на то, что нами была проведена

pathovars of the used genomic assemblies deposited in NCBI RefSeq.

In order to find a strategy for solving these problems, we rechecked the genomic assemblies of *P. s. pv. maculicola* and *P. s. pv. tomato* used in the analysis. As a result, 3 assemblies of *P. s. pv. maculicola* were excluded from the analysis with the status "Taxonomy check: Inconclusive". This status means that the average nucleotide identity (ANI) for this assembly does not match the expected. Thus, genomic assemblies GCF_001293575.1, GCF_001294185.1 and GCF_000145845.2 were excluded from the analysis (Table 1). The proteins of these genomic assemblies showed the greatest similarity to the sequences of *P. c. pv. alisalensis*.

We rebuilt phylogenetic trees for 12 previously selected clusters with updated sequences. Figure 2 shows 4 out of 12 phylogenetic trees with updated sequences.

Comparing the trees shown in Figures 1 and 2, we can conclude that the filtering of genomic assemblies with an unconfirmed annotation made it possible to exclude sequences of *P. c. pv. alisalensis* from the clades containing the sequences of *P. s. pv. maculicola* and *P. s. pv. tomato*, only for some clusters. Thus, sequences

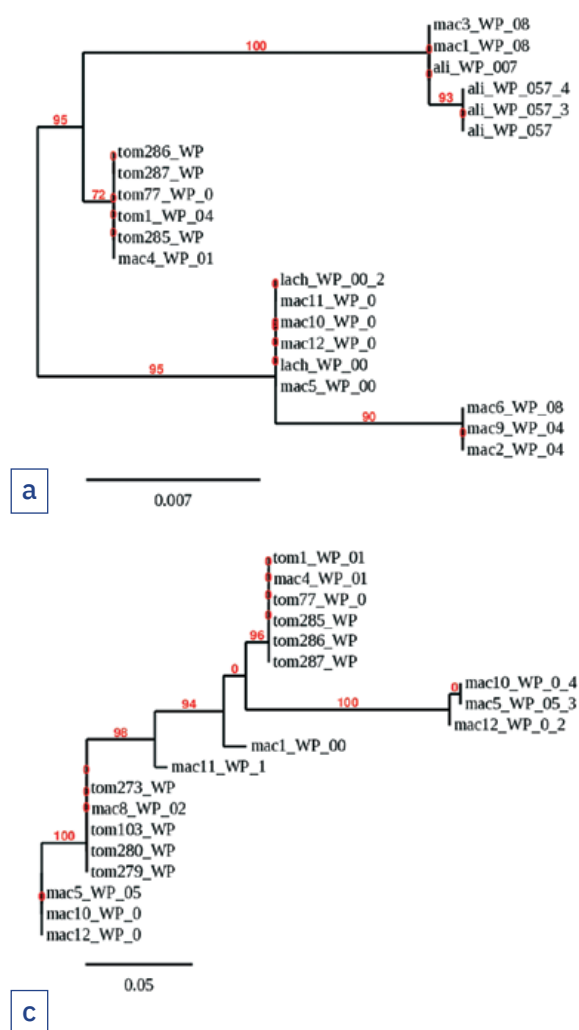


Рис. 1. Филогенетические деревья белковых последовательностей для кластеров 1415 (a), 18924 (b), 19719 (c), 28504 (d)

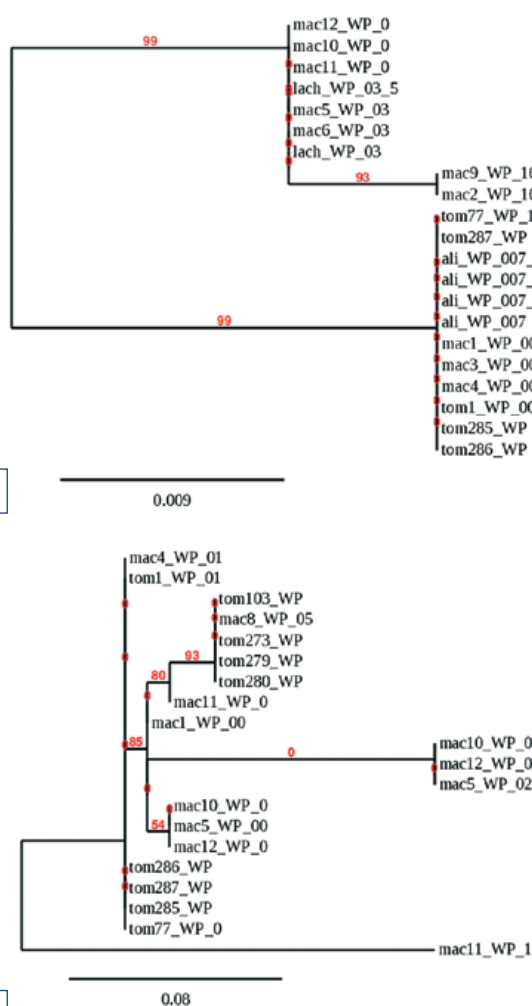


Fig. 1. Phylogenetic trees of protein sequences for clusters 1415 (a), 18924 (b), 19719 (c), 28504 (d)

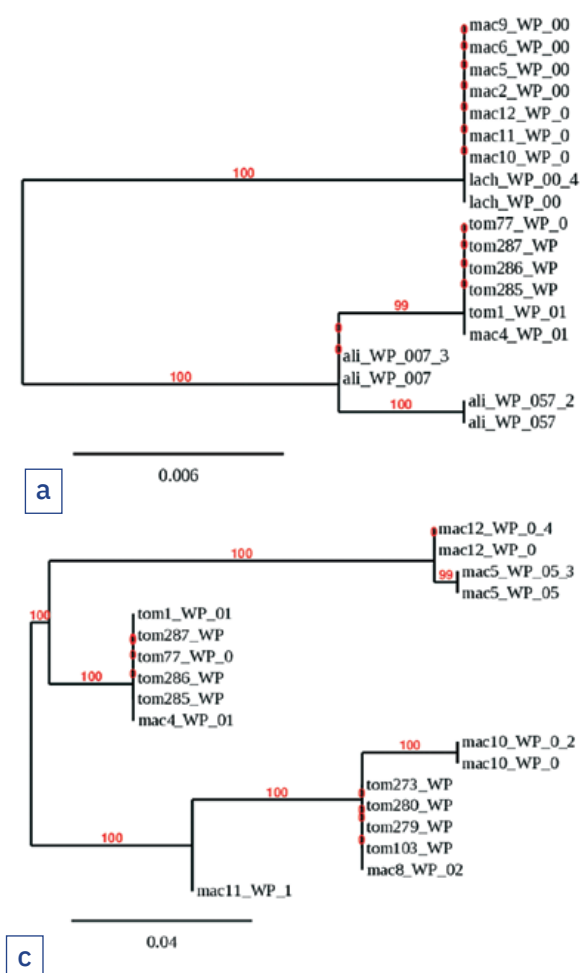


Рис. 2. Филогенетические деревья белковых последовательностей для отобранных кластеров 1415 (а), 18924 (б), 19719 (с), 28504 (д) с учетом обновленного набора последовательностей

Fig. 2. Phylogenetic trees of protein sequences for selected clusters 1415 (a), 18924 (b), 19719 (c), 28504 (d) taking into account the updated set of sequences

фильтрация использовавшихся геномных сборок, последовательности данного патовара выделялись в различные клады, что поставило под вопрос возможность создания универсальной тест-системы, обладающей чувствительностью ко всем штаммам патовара. По-видимому, для данного патовара характерна высокая вариабельность последовательностей.

Результаты выравнивания восьми отобранных кластеров приведены в таблице 3. Данные выравнивания были проведены для установления их сходства с последовательностями других фитопатогенных видов, не использовавшихся на предыдущем этапе работы. При этом не оценивалась точность аннотации геномных сборок других видов рода *Pseudomonas*. Из всех видов, сходство с последовательностями которых обнаруживалось по результатам выравнивания анализируемых последовательностей, приведены только значимые фитопатогенные и широко распространенные виды.

Для дальнейшего анализа мы выбрали только те кластеры, в которых белковая идентичность последовательностям *Pseudomonas fluorescens*, *P. viridiflava* и *P. cannabina* pv. *alisalensis* не превышает 95%. Таким условиям соответствовали 2 кластера: 11492 и 9438 (выделены зеленым). Для каждой из белковых последовательностей, входящих в кластеры, были загружены соответствующие

of clusters 22842, 21897, 18924, 17663 were excluded from the analysis.

For the remaining clusters, variability was observed for the sequences of *P. s. pv. maculicola*. Despite the fact that we filtered the used genomic assemblies, the sequences of this pathovar were allocated to different clades, which questioned the possibility of creating a universal test system that is sensitive to all pathovar strains. Apparently, this pathovar is characterized by high sequence variability.

The results of the alignment of 8 selected clusters are shown in Table 3. These alignments were carried out to establish their similarity with the sequences of other phytopathogenic species that were not used at the previous stage of work. At the same time, the accuracy of the annotation of genomic assemblies of other species of the genus *Pseudomonas* was not evaluated. Of all the species, the similarity with the sequences of which was found by the results of the alignment of the analyzed sequences, only significant phytopathogenic and widespread species are shown.

For further analysis, we selected only those clusters in which the protein identity for the sequences of *Pseudomonas fluorescens*, *P. viridiflava* and *P. cannabina*

Таблица 3
Результаты белкового выравнивания анализируемых кластеров с последовательностями других фитопатогенных псевдомонад

Table 3
Results of protein alignment of the analyzed clusters with sequences of other phytopathogenic pseudomonads

Кластер Cluster	Идентификатор белка Protein ID	Максимальная идентич- ность с <i>P. fluorescens</i> Maximum identity with <i>P. fluorescens</i>	Максимальная идентич- ность с <i>P. viridiflava</i> Maximum identity with <i>P. viridiflava</i>	Максимальная идентичность с <i>P. cannabina</i> (в т. ч. <i>pv. alisalensis</i>) Maximum identity with <i>P. cannabina</i> (including <i>pv. alisalensis</i>)
11492	WP_044392274.1	94,59%	94,93%	–
12691	WP_005771418.1	91,67%	91,30%	97,46%
1415	WP_080898718.1	88,46%	97,77%	96,56%
19719	WP_029290780.1	99,46%	97,28%	–
26276	WP_080898717.1	79,53%	89,76%	98,43%
28504	WP_053481123.1	100%	–	–
30723	WP_005764646.1	56,84%	95,89%	96,84%
9438	WP_005770349.1	77,99%	92,55%	92,55%

последовательности кодирующих их генов. Для каждого из белков была оставлена только 1 соответствующая ему репрезентативная нуклеотидная последовательность.

В ходе дальнейшего анализа корректности аннотации геномных сборок не были обнаружены последовательности генов, соответствующих белкам *P. fluorescens* и *P. viridiflava*, что может быть объяснено примененным нами фильтром корректности. Таким образом, нуклеотидные последовательности генов *P. fluorescens* и *P. viridiflava* в анализ включены не были.

На рисунках 3 и 4 представлены результаты выравнивания нуклеотидных последовательностей, соответствующих кластерам 11492 и 9438.

На рисунке 3 видно, что в различных штаммах патовара *P. s. pv. maculicola* полиморфизмы

pv. alisalensis does not exceed 95%. Two clusters corresponded to these conditions: 11492 and 9438 (highlighted in green). For each of the protein sequences included in the clusters, the corresponding sequences of the genes encoding them were loaded. For each of the proteins, only 1 corresponding representative nucleotide sequence was left.

In the course of further analysis of the correctness of the annotation of genomic assemblies, no sequences of genes corresponding to proteins *P. fluorescens* and *P. viridiflava* were found, which can be explained by the correctness filter applied by us. Thus, the nucleotide sequences of the genes of *P. fluorescens* and *P. viridiflava* were not included in the analysis.

Figures 3 and 4 show the results of the alignment of the nucleotide sequences corresponding to clusters 11492 and 9438.

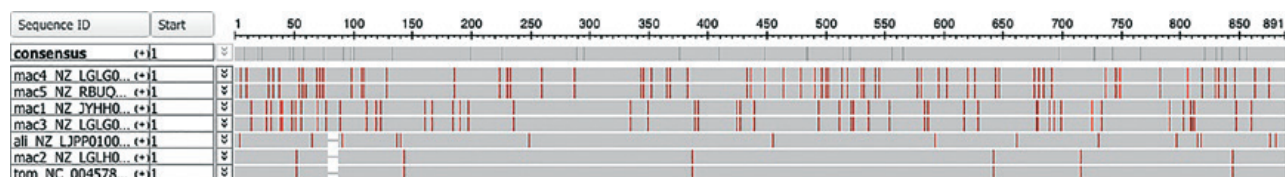


Рис. 3. Выравнивание нуклеотидных последовательностей, соответствующих белкам из кластера 11492. Красным обозначены полиморфные участки, белым – «гэпы».

Fig. 3. Alignment of nucleotide sequences corresponding to proteins from cluster 11492. Polymorphic areas are marked in red; "gaps" are marked in white.

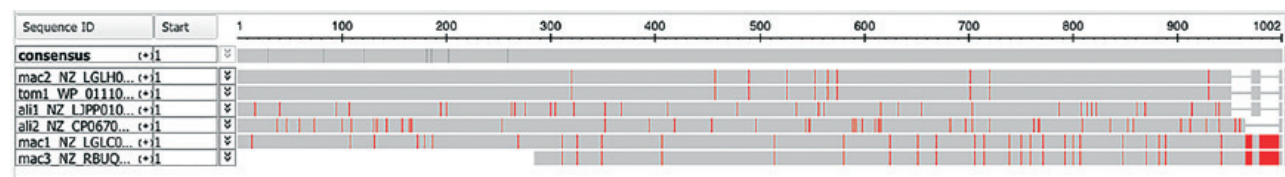


Рис. 4. Выравнивание нуклеотидных последовательностей, соответствующих белкам из кластера 9438. Красным обозначены полиморфные участки, белым – «гэпы».

Fig. 4. Alignment of nucleotide sequences corresponding to proteins from cluster 9438. Polymorphic areas are marked in red; "gaps" are marked in white.

генов кластера 11492, хоть и отличаются от соответствующей последовательности *P. c. pv. alisalensis*, вариабельны и не имеют консенсуса. Характер распределения полиморфизмов разделяет последовательности патовара на 3 различных типа, отличающихся как от *P. c. pv. alisalensis*, так и друг от друга. Это не позволяет подобрать участок для разработки универсальных праймеров на патовар с использованием данных последовательностей.

Аналогичное распределение было получено для последовательностей кластера 9438 (рис. 4): последовательности вариабельны внутри патовара и разделяются на 2 типа, что не позволяет подобрать на данные последовательности праймеры для всех изученных сегодня групп штаммов *P. s. pv. maculicola*.

ЗАКЛЮЧЕНИЕ

Анализ генетических отличий некоторых патоваров *Pseudomonas syringae*, проведенный путем исследования аннотированных белков, соответствующих 293 общедоступным геномным сборкам некоторых патоваров вида *P. syringae* (включая 12 штаммов патовара *P. s. pv. maculicola*), а также 43 геномным сборкам близкородственных видов *P. cannabina* и *P. savastanoi*, показал высокое сходство большинства анализируемых последовательностей с последовательностями патовара *P. c. pv. alisalensis*.

СПИСОК ЛИТЕРАТУРЫ

1. Приходько С., Яремко А., Корнев К., 2021. Испытание различных методов идентификации возбудителя бактериальной пятнистости листьев цветной капусты *Pseudomonas syringae* pv. *maculicola* (McCulloch) Young et al. – Таврический вестник аграрной науки, 1 (25): 174–186. URL: <https://doi.org/10.33952/2542-0720-2021-1-25-174-186>.
2. Anzai Y., Kim H., Park J., Wakabayashi H., Oyaizu H., 2000. Phylogenetic affiliation of the pseudomonads based on 16S rRNA sequence. – International journal of systematic and evolutionary microbiology, 50 (4): 1563–1589. URL: <https://doi.org/10.1099/00207713-50-4-1563>.
3. Cuppels D., Ainsworth T., 1995. Molecular and Physiological Characterization of *Pseudomonas syringae* pv. *tomato* and *Pseudomonas syringae* pv. *maculicola* Strains That Produce the Phytotoxin Coronatine. – Applied and Environmental Microbiology, 61 (10): 3530–3536. URL: <https://doi.org/10.1128/aem.61.10.3530-3536.1995>.
4. Gardan L., Shafik H., Belouin S., Broch R., Grimont F., Grimont P.A.D., 1999. DNA relatedness among the pathovars of *Pseudomonas syringae* and description of *Pseudomonas tremiae* sp. nov. and *Pseudomonas cannabina* sp. nov. (ex Sutic and Dowson 1959). – Int. J. System. Bact., № 49: 469–478. URL: <https://doi.org/10.1099/00207713-49-2-469>.
5. Gironde S., Manceau C., 2012. Housekeeping Gene Sequencing and Multilocus Variable-Number Tandem-Repeat Analysis to Identify Subpopulations within *Pseudomonas syringae* pv. *maculicola* and *Pseudomonas syringae* pv. *tomato* That Correlate with Host Specificity. – Applied and Environmental Microbiology, 78 (9): 3266–3279. URL: <https://doi.org/10.1128/AEM.06655-11>.
6. Guindon S., Dufayard J.-F., Lefort V., Anisimova M., Hordijk W., Gascuel O., 2010. New Algorithms

Figure 3 shows that in various strains of the pathovar *P. s. pv. maculicola*, the polymorphisms of genes of cluster 11492, although they differ from the corresponding sequence of *P. c. pv. alisalensis*, are variable and have no consensus. The pattern of distribution of polymorphisms separates pathovar sequences into 3 different types, different both from *P. c. pv. alisalensis* and each other. This makes it impossible to select a site for the development of universal primers for the pathovar using these sequences.

A similar distribution was obtained for sequences of cluster 9438 (Fig. 4): sequences are variable within pathovar and are divided into 2 types, which does not allow selecting primers for these sequences for all strain groups of *P. s. pv. maculicola* studied today.

CONCLUSION

Analysis of genetic differences of some *Pseudomonas syringae* pathovars conducted by examining annotated proteins corresponding to 293 available genomic assemblies of some *P. syringae* pathovars (including 12 strains of the pathovar *P. s. pv. maculicola*), as well as 43 genomic assemblies of closely related species *P. cannabina* and *P. savastanoi*, showed a high similarity of most of the analyzed sequences with the sequences of the pathovar *P. c. pv. alisalensis*.

REFERENCES

1. Prikhodko S.I., Iaremko A.B., Kornev K.P. Testing of different methods for identification of bacterial leaf spot (*Syringae* pv. *maculicola* (McCulloch) Young et al.) plant pathogen in cauliflower leaves. *Taurida Bulletin of Agrarian Science*, 2021; 1 (25): 174–186. URL: <https://doi.org/10.33952/2542-0720-2021-1-25-174-186> (in Russian).
2. Anzai Y., Kim H., Park J., Wakabayashi H., Oyaizu H. Phylogenetic affiliation of the pseudomonads based on 16S rRNA sequence. *International journal of systematic and evolutionary microbiology*, 2000; 50 (4): 1563–1589. URL: <https://doi.org/10.1099/00207713-50-4-1563>.
3. Cuppels D., Ainsworth T. Molecular and Physiological Characterization of *Pseudomonas syringae* pv. *tomato* and *Pseudomonas syringae* pv. *maculicola* Strains That Produce the Phytotoxin Coronatine. *Applied and Environmental Microbiology*, 1995; 61 (10): 3530–3536. URL: <https://doi.org/10.1128/aem.61.10.3530-3536.1995>.
4. Gardan L., Shafik H., Belouin S., Broch R., Grimont F., Grimont P.A.D. DNA relatedness among the pathovars of *Pseudomonas syringae* and description of *Pseudomonas tremiae* sp. nov. and *Pseudomonas cannabina* sp. nov. (ex Sutic and Dowson 1959). *Int. J. System. Bact.*, 1999; No. 49: 469–478. URL: <https://doi.org/10.1099/00207713-49-2-469>.
5. Gironde S., Manceau C. Housekeeping Gene Sequencing and Multilocus Variable-Number Tandem-Repeat Analysis to Identify Subpopulations within *Pseudomonas syringae* pv. *maculicola* and *Pseudomonas syringae* pv. *tomato* That Correlate with Host Specificity. *Applied and Environmental Microbiology*, 2012; 78 (9): 3266–3279. URL: <https://doi.org/10.1128/AEM.06655-11>.

and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. – *Systematic Biology*, 59 (3): 307–321.

7. Hwang M., Morgan R., Sarkar S., Wang P., Guttman D., 2005. Phylogenetic characterization of virulence and resistance phenotypes of *Pseudomonas syringae*. – *Applied and Environmental Microbiology*, 71 (9): 5182–5191. URL: <https://doi.org/10.1128/AEM.71.9.5182-5191.2005>.

8. Ilić R., Balaž J., Stojšin V., Bagi F., Pivić R., Stanojković-Sebić A., Jošić D., 2016. Molecular characterization of *Pseudomonas syringae* pvs. from different host plants by repetitive sequence-based PCR and multiplex-PCR. – *Zemdirbyste-Agriculture*, 103 (2): 199–206. URL: <https://doi.org/10.13080/z-a.2016.103.026>.

9. Morris C., Kinkel L., Xiao K., Prior P., Sands D., 2007. Surprising niche for the plant pathogen *Pseudomonas syringae*. – *Infection, Genetics and Evolution*, № 7: 84–92.

10. Sakata N., Ishiga T., Ishiga Y., 2021. *Pseudomonas cannabina* pv. *alisalensis* TrpA Is Required for Virulence in Multiple Host Plants. – *Frontiers in microbiology*, № 12: 1–11. URL: <https://doi.org/10.3389/fmicb.2021.659734>.

11. Sarkar S., Guttman D., 2004. Evolution of the core genome of *Pseudomonas syringae*, a highly clonal, endemic plant pathogen. – *Applied and Environmental Microbiology*, 70 (4): 1999–2012. URL: <https://doi.org/10.1128/AEM.02553-07>.

12. CABI Crop Protection Compendium, 2021. – URL: <https://www.cabi.org>.

13. MUSCLE (v. 3.8.31). – URL: https://drive5.com/muscle/downloads_v3.htm.

14. A multiple alignment viewer (MView). – URL: <https://www.ebi.ac.uk/Tools/msa/mview/>.

15. RefSeq: NCBI Reference Sequence Database, 2021. – URL: <https://www.ncbi.nlm.nih.gov/refseq/>.

16. TreeDyn (v. 198.3). – URL: http://phylogeny.lirmm.fr/phylo.cgi/one_task.cgi?task_type=treedyn.

ИНФОРМАЦИЯ ОБ АВТОРАХ

Приходько Светлана Игоревна, научный сотрудник, заведующая лабораторией бактериологии и анализа ГМО Испытательного лабораторного центра ФГБУ «ВНИИКР», р. п. Быково, г. Раменское, Московская обл., Россия; ORCID 0000-0002-1281-4410, e-mail: svetlana.prik@yandex.ru.

Яремко Анастасия Богдановна, младший научный сотрудник научного отдела молекулярно-генетических методов диагностики ФГБУ «ВНИИКР», р. п. Быково, г. Раменское, Московская обл., Россия; ORCID 0000-0003-3295-8080, e-mail: an_ya94@mail.ru.

Старикова Елизавета Валентиновна, агроном лаборатории вирусологии Испытательного лабораторного центра ФГБУ «ВНИИКР», р. п. Быково, г. Раменское, Московская обл., Россия; ORCID 0000-0001-6582-210X, e-mail: hed.robin@gmail.com.

Корнев Константин Павлович, кандидат биологических наук, заместитель директора ФГБУ «ВНИИКР», р. п. Быково, г. Раменское, Московская обл., Россия; ORCID 0000-0002-3490-1857, e-mail: konstantin.kornev@gmail.com.

6. Guindon S., Dufayard J.-F., Lefort V., Anisimova M., Hordijk W., Gascuel O. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology*, 2010; 59 (3): 307–321.

7. Hwang M., Morgan R., Sarkar S., Wang P., Guttman D. Phylogenetic characterization of virulence and resistance phenotypes of *Pseudomonas syringae*. *Applied and Environmental Microbiology*, 2005; 71 (9): 5182–5191. URL: <https://doi.org/10.1128/AEM.71.9.5182-5191.2005>.

8. Ilić R., Balaž J., Stojšin V., Bagi F., Pivić R., Stanojković-Sebić A., Jošić D. Molecular characterization of *Pseudomonas syringae* pvs. from different host plants by repetitive sequence-based PCR and multiplex-PCR. *Zemdirbyste-Agriculture*, 2016; 103 (2): 199–206. URL: <https://doi.org/10.13080/z-a.2016.103.026>.

9. Morris C., Kinkel L., Xiao K., Prior P., Sands D. Surprising niche for the plant pathogen *Pseudomonas syringae*. *Infection, Genetics and Evolution*, 2007; No. 7: 84–92.

10. Sakata N., Ishiga T., Ishiga Y. *Pseudomonas cannabina* pv. *alisalensis* TrpA Is Required for Virulence in Multiple Host Plants. *Frontiers in microbiology*, 2021; No. 12: 1–11. URL: <https://doi.org/10.3389/fmicb.2021.659734>.

11. Sarkar S., Guttman D. Evolution of the core genome of *Pseudomonas syringae*, a highly clonal, endemic plant pathogen. *Applied and Environmental Microbiology*, 2004; 70 (4): 1999–2012. URL: <https://doi.org/10.1128/AEM.02553-07>.

12. CABI Crop Protection Compendium, 2021. URL: <https://www.cabi.org>.

13. MUSCLE (v. 3.8.31). URL: https://drive5.com/muscle/downloads_v3.htm.

14. A multiple alignment viewer (MView). URL: <https://www.ebi.ac.uk/Tools/msa/mview/>.

15. RefSeq: NCBI Reference Sequence Database, 2021. URL: <https://www.ncbi.nlm.nih.gov/refseq/>.

16. TreeDyn (v. 198.3). URL: http://phylogeny.lirmm.fr/phylo.cgi/one_task.cgi?task_type=treedyn.

INFORMATION ABOUT THE AUTHORS

Svetlana Prikhodko, Researcher, Head of Bacteriology and GMO Analysis Laboratory, Laboratory Testing Center, FGBU “VNIICR”, Bykovo, Ramenskoye, Moscow Oblast, Russia; ORCID 0000-0002-1281-4410, e-mail: svetlana.prik@yandex.ru.

Anastasia Yaremko, Junior Researcher, Molecular and Genetic Diagnostic Methods Department, FGBU “VNIICR”, Bykovo, Ramenskoye, Moscow Oblast, Russia; ORCID 0000-0003-3295-8080, e-mail: an_ya94@mail.ru.

Elizaveta Starikova, Agronomist, Virology Laboratory, Laboratory Testing Center, FGBU “VNIICR”, Bykovo, Ramenskoye, Moscow Oblast, Russia; ORCID 0000-0001-6582-210X, e-mail: hed.robin@gmail.com.

Konstantin Kornev, PhD in Biology, Deputy Director, FGBU “VNIICR”, Bykovo, Ramenskoye, Moscow Oblast, Russia; ORCID 0000-0002-3490-1857, e-mail: konstantin.kornev@gmail.com.