

**ВОРОНОВА Н.В.**✉, **ВОРОБЬЁВА М.М.**, **БОНДАРЕНКО Ю.В.**

Белорусский государственный университет,

Беларусь, 220030, г. Минск, пр. Независимости, 4,

✉ [nvoronova@bsu.by](mailto:nvoronova@bsu.by), [ninavvoronova@gmail.com](mailto:ninavvoronova@gmail.com), (375) 29-620-20-00

### ВАРИАБЕЛЬНОСТЬ ГЕНОВ CYP4 И CYP6 У ТЛЕЙ *APHIS FABAE MORDVILKOI* BÖRNER & JANISCH, 1922

**Цель.** Оценить вариабельность генов CYP450 4-го и 6-го семейств, извлеченных их данных полногеномного секвенирования тли *Aphis fabae mordvilkoï*, развивающейся в Беларуси на чубушнике (*Philadelphus coronaries* L.). **Методы.** Секвенирование генома тлей провели в Центре секвенирования ДНК Университета штата Юта (г. Солт-Лейк-Сити, США). Последовательности генов CYP4 и CYP6 были извлечены из пула прочтений генома путем последовательной фильтрации выравниваний прочтений по референсным последовательностям генов CYP450 4-го и 6-го семейства, извлеченных из 3 находящихся в открытом доступе геномов тлей (*Acyrtosiphon pisum* Harris, 1776, *Muzus persicae* (Sulzer, 1776) и *Diuraphis noxia* (Mordvilkoëx-Kurdjumov, 1913)). Все обнаруженные уникальные варианты сборки оценивали как индивидуальный ген. **Результаты.** В геноме *A. fabae mordvilkoï* обнаружен 31 ген CYP4, из которых 24 представляют собой варианты гена CYP4C1. В семействе CYP6 обнаружено 19 генов, 8 из которых представляют собой варианты CYP6A13. Вариабельность нуклеотидных и аминокислотных последовательностей CYP4 и CYP6 у *A. fabae mordvilkoï* оказалась в равной степени высокой. **Заключение.** В геноме тли *A. fabae mordvilkoï* среди генов CYP450 семейства CYP4 преобладают гены CYP4C1, а гены CYP6 представлены преимущественно генами CYP6A13. Остальные гены указанных семейств представлены в геноме *A. fabae mordvilkoï* единичными копиями.

**Ключевые слова:** тли, цитохромы р450, *Aphis fabae*, трофическая специализация, копияность генов.

Тли (Aphidoidea) это некрупный таксон растительноядных насекомых, обладающих рядом особенностей биологии, в частности таких, как облигатный партеногенез, сложный биологический цикл со сменой растений-хозяев и полифенизм, что делает их интересным модель-

ным объектом для изучения процессов экологической диверсификации и связанного с этим видообразования [1]. Большинство видов тлей тесно связаны с конкретными видами кормовых растений, которые представляют собой не только источник питания и влаги, но и фактически являются их средой обитания. Особый микроклимат, формируемый кормовым растением, с его влажностью, освещенностью и привлекательностью для энтомофагов и, что более важно, состав вторичных метаболитов кормового растения определяет перечень видов тлей, способных на нем питаться и размножаться [2]. Этот перечень жестко детерминирован, что позволяет использовать способность к питанию и репродукции на конкретном растении как важный видовой признак в таксономии тлей. Около 1 % рецентных видов тлей, однако, являются истинно многоядными, и круг их кормовых растений включает тысячи видов из десятков ботанических семейств. Из 35 видов тлей, которых Blackman и Eastop относят к числу полифагов [3], 32 относятся к наиболее молодому и процветающему семейству Aphididae и только 3 – к семейству Eriosomatidae.

Возникновение полифагии у тлей, как считается, является эволюционно прогрессивным приобретением. По общепринятому мнению [4], полифагия требует от тлей развития чрезвычайно гибких биохимических механизмов, обеспечивающих нейтрализацию вторичных метаболитов растений, многие из которых имеют направленную инсектицидную активность. Вероятно, по этой причине эволюция тлей в основной массе шла по пути адаптации к питанию на ограниченном числе кормовых растений [5], несмотря на то, что полифагия предоставляет неоспоримые преимущества. До сих пор не известно, в чем причина того, что полифагия среди тлей остается чрезвычайно редким явлением: связано это с высокой «метаболической стоимостью» нейтрализации колоссально многообразия вторичных метаболитов расте-

© **ВОРОНОВА Н.В.**, **ВОРОБЬЁВА М.М.**, **БОНДАРЕНКО Ю.В.**

ний, или способность к полифагии является эволюционным новшеством, которое, благодаря выраженным преимуществам такого рода трофической стратегии, будет распространяться по таксону с течением эволюционного времени.

Молекулярные механизмы полифагии полностью не ясны. Существует мнение, что полифагия обеспечивается гиперинтенсификацией работы системы детоксикации, включающей группу белков: эстераз, глутатион-S-трансфераз и цитохромов р450, а в частности CYP4504-го и 6-го семейств, которые непосредственно ассоциированы с нейтрализацией ксенобиотиков [6]. При этом эволюция системы детоксикации идет не только по пути усиления экспрессии генов, кодирующих входящие в нее ферменты, но и по пути физического увеличения числа копий генов, кодирующих белки определенных семейств. До сих пор, однако, изучение особенностей геномов тлей, отличающихся способностью к полифагии, не было осуществлено на уровне, позволяющем делать заключения о том, является ли увеличение копияности генов общим признаком многоядных видов тлей.

В Международных базах данных нуклеотидных последовательностей NCBI GenBank и AphidBase по состоянию на 2017 г. были представлены 6 полных геномов тлей, среди которых только один вид – *Myzus persicae* (Šulzer, 1776) – относится к полифагам. Мы провели полногеномное секвенирование для другого представителя семейства Aphididae, также относящегося к полифагам, – *Aphis fabae* subsp. *mordvilkoii* Börner&Janisch, 1922, который в Беларуси является фоновым и входит в перечень вредителей культивируемых растений. На этом этапе исследования мы оценили число и вариабельность генов 4-го и 6-го семейств цитохромов р450 (CYP4 и CYP6 соответственно) в сравнении с этими же показателями у *M. persicae* и других видов тлей, чьи геномы в настоящее время доступны.

### Материалы и методы

Тли *A. fabae mordvilkoii* были коллектированы с чубушника венечного (*Philadelphus coronaries* L.) в г. Минск. Видовая принадлежность образцов была дополнительно верифицирована по последовательности гена COI. Полногеномное секвенирование провели в Центре ДНК-секвенирования Университета штата Юта (г. Солт-Лейк-Сити, США). Данные секвенирования были получены из препаратов тотальной

ДНК тлей на основании двух отдельных библиотек секвенирования с использованием секвенатора IonTorrent Proton и представляли собой «single-end» прочтения. Исходные данные содержали 82 480 385 прочтений общей длиной 11 553 721 036 нуклеотидов, длиной от 25 до 390 и средней длиной 140 нуклеотидов.

Поскольку данные оценки размера генома исследуемого вида путем проточной цитофлуориметрии отсутствуют, при оценке глубины покрытия генома мы ориентировались на средний размер генома, рассчитанный для четырех других видов рода *Aphis*: *A. craccivora*, *A. glycines*, *A. gossypi* и *A. nerii*. Средний размер генома, таким образом, составил 356 875 000 пар нуклеотидов, что позволило нам рассчитать предположительную глубину покрытия генома *A. fabae mordvilkoii*, которая составила 32,375 раза.

Оценка качественного и количественного состава данных была произведена с использованием программы FastQC версии 0.11.7. Обрезку и фильтрацию прочтений производили в программе Trimmomatic версии 0.36. Выбранные параметры обработки: CROP=200, MINLEN=130. После удаления коротких прочтений и обрезки некачественных концов пула данных составил 61 724 382 прочтения общей длиной 9 852 001 469 и средней длиной 160 нуклеотидов.

После очистки исходных данных провели приблизительную оценку ширины покрытия генома *A. fabae mordvilkoii*. Поскольку секвенирование генома для данного вида производилось впервые, для оценки ширины покрытия использовали геном тли *A. glycines* в качестве референсного. Ширину покрытия оценивали по результатам маппинга прочтений генома *A. fabae mordvilkoii* на референсный геном. Маппинг проводили по алгоритму Bowtie2, а результаты маппинга оценивали с использованием утилиты «bedtoolsgenomecov». Ширина покрытия референсного генома составила 76,31 % со средней глубиной – 23,5 раза.

Последовательности генов CYP4 (CYP4C1, CYP4C3 CYP4G15, CYP4V2) и CYP6 (CYP6A2, CYP6A13, CYP6A14, CYP6B1, CYP6J1, CYP6K1) были извлечены из пула прочтений генома путем последовательной фильтрации выравниваний прочтений по белок-кодирующим областям генов CYP450 4-го и 6-го семейства, которые, в свою очередь извлекли из 3, находящихся в открытом доступе, геномов тлей

(*A. pisum*, *M. persicae* и *D. noxia*), ориентируясь на сопровождающие геномы аннотации. Еще 3 доступных генома тлей использованы в этой работе не были, т. к. они не сопровождаются достаточно подробной аннотацией генов CYP450.

Алгоритм сборки генов CYP4 и CYP6 прочтений генома *A. fabae mordvilko* состоял из следующих шагов:

1) выравнивание прочтений генома *A. fabae mordvilko* по полному референсному геному;

2) извлечение «островов», собранных на участках референсного генома, которые, согласно аннотации, представляют собой белок-кодирующие области генов CYP4 и CYP6 (фильтрация аннотации генома в формате GFF3 по колонке описания с удалением всех записей, не соответствующих генам CYP450 4 и 6 семейств, и последующая фильтрация результатов выравнивания с использованием утилиты «bedtoolsintersect»);

3) построение из выровненных прочтений консенсусной последовательности (с использованием утилиты «vcf-consensus»);

4) использование оставшихся после выравнивания прочтений (фильтрация результатов выравнивания утилитой «samtoolsview» с последующей конвертацией результатов утилитой «bedtoolsbamtofastq») для выравнивания по следующему референсному геному (повторение шагов 1–4).

Последовательным перебором референсов было установлено, что порядок использования референсных геномов на полученные результаты влияния не имел.

Для верификации определения полученных консенсусных последовательностей как CYP4 и CYP6, а также для выявления в этих последовательностях белок-кодирующих областей (CDS) использовали систему BLAST(NCBI).

Таблица 1. Количество генов CYP450 4-го семейства у *Aphis fabae mordvilko* в сравнении с референсными геномами

Гены	Количество копий генов, обнаруженных в геноме соответствующих видов тлей			
	<i>Aphis fabae mordvilko</i>	<i>Myzus persicae</i>	<i>Acyrtosiphon pisum</i>	<i>Diuraphis noxia</i>
CYP4C1	24	18	21	13
CYP4C3	1	–	2	–
CYP4G15	2	1	4	3
CYP4V2	4	–	2	3
Всего генов	31	19	29	19

Примечание. «-» – гены в геноме не обнаружены.

Для CDS каждого гена в отдельности была проверена рамка считывания, после чего нуклеотидные последовательности белок-кодирующих областей генов были транслированы в аминокислотные последовательности.

Выравнивание аминокислотных последовательностей осуществили в программе MEGA7 с использованием алгоритмов Muscle и ClustalW (алгоритмы использовались поочередно). Парные дистанции между аминокислотными последовательностями также рассчитали в программе MEGA7 [7, 8].

Генные деревья построили с использованием Байесовского анализа методом Монте-Карло для Марковских цепей (MCMC) в программе BEAST2 версии 2.4.7. Аннотацию деревьев провели в программе FigTree версии 1.4.3.

### Результаты и обсуждение

В процессе дополнительной верификации в программе BLAST все аминокислотные последовательности были распознаны как соответствующие последовательности CYP4 и CYP6 с высоким процентом идентичности (91–100 %) и перекрытия (99–100 %).

В результате в геноме *A. fabae mordvilko* была обнаружена 31 последовательность, соответствующая генам CYP4 референсных геномов (табл. 1).

Что касается генов CYP6 еще одного семейства цитохромов р450, отвечающего за детоксикацию у насекомых, то в геноме *A. fabae mordvilko* обнаруживается больше генов, чем это было показано для CYP4, но каждый ген представлен в геноме меньшим числом копий (табл. 2).

На основе аминокислотных последовательностей также построили дендрограммы и рассчитали апостериорную вероятность топологии ветвей (рис.).

Таблица 2. Количество генов CYP450 6-го семейства у *Aphis fabae mordvilkoii* в сравнении с референсными геномами

Гены	Количество копий генов, обнаруженных в геноме соответствующих видов тлей			
	<i>Aphis fabae mordvilkoii</i>	<i>Myzus persicae</i>	<i>Acyrtosiphon pisum</i>	<i>Diuraphis noxia</i>
CYP6A2	1	1	1	1
CYP6A13	8	7	13	8
CYP6A14	4	4	3	5
CYP6B1	1	–	2	3
CYP6J1	1	1	1	1
CYP6K1	4	5	4	3
Всего генов	19	18	24	21

Примечание. «-» – гены в геноме не обнаружены.

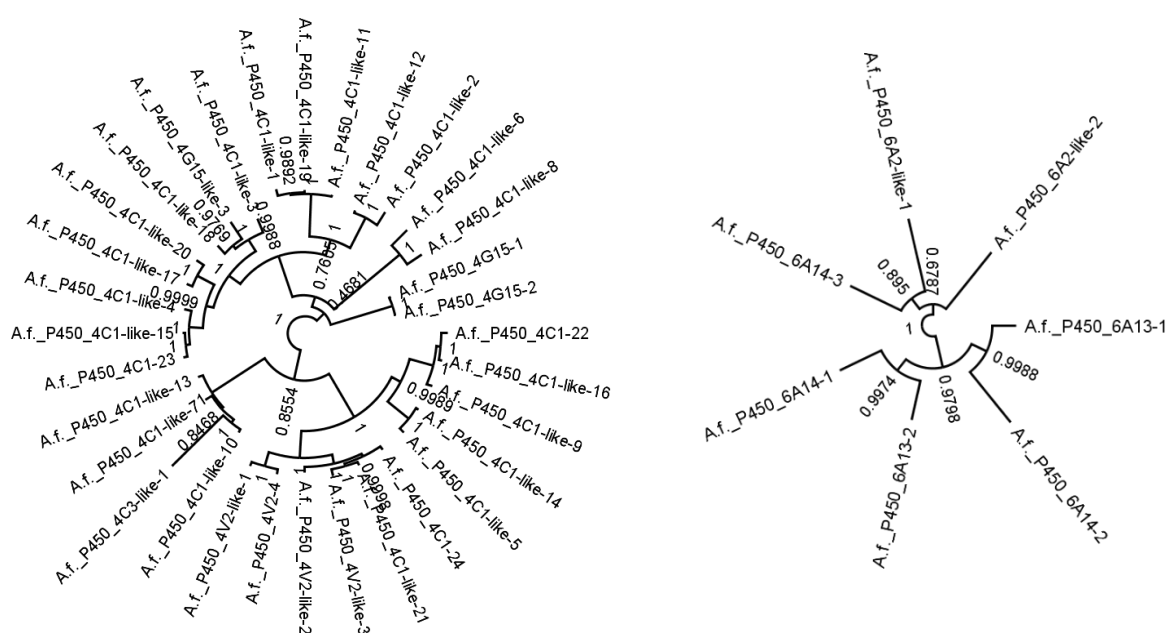


Рис. Дендрограммы сходства аминокислотных последовательностей белок-кодирующей области генов CYP4 (слева) и CYP6 (справа) *Aphis fabae mordvilkoii*.

Оказалось, что аминокислотные последовательности CYP4C1 образуют два крупных кластера, в один из которых входят также последовательности CYP4G15, а в другой – последовательности CYP4V2, причем последовательности CYP4G15 отделяются от своего кластера сразу же после разделения CYP4C1 на две группы. На дендрограмме, построенной для аминокислотных последовательностей CYP6, видно, что последовательности CYP6A14 не образуют единый кластер, объединяясь с остальными генами семейства CYP6. Апостериорная вероятность, поддерживающая отдельные ветви на дендрограмме, тем не менее, была высока. Высокое значение апостериорной вероятности топологии ветвей подтверждает предположение о

происхождении генов семейств CYP450 в результате дупликации исходного гена.

### Выводы

В результате анализа данных полногеномного секвенирования тлей *A. fabae mordvilkoii* оказалось, что в геноме этих тлей обнаруживается 50 последовательностей генов суперсемейства CYP450, относящихся к семействам, ассоциированным с трансформацией ксенобиотиков CYP4 и CYP6). В семействе CYP4 обнаружена 31 копия гена, из которых 24 представляют собой ген CYP4C1. В семействе CYP6 обнаружено 19 копий генов, 8 из которых идентифицируются как CYP6A13.

## Література

1. Brisson J.A., Stern D.L. The pea aphid, *Acyrtosiphon pisum*: an emerging genomic model system for ecological, developmental and evolutionary studies. *Bio Essays*. 2006. Vol. 28, Iss. 7. P. 747–755. doi: 10.1002/bies.20436.
2. Hawthorne D.J., Via S. Genetic linkage of ecological specialization and reproductive isolation in pea aphids. *Nature*. 2001. P. 904–907. doi: 10.1038/35091062.
3. Blackman R.L., Eastop V.F. Aphids on the world's herbaceous plants and shrubs. Vol. 2, The aphids. John Wiley & Sons, 2008. 1460 p.
4. Dowd P.F., Smith C.M., Sparks T.C. Detoxification of plant toxins by insects. *Insect Biochemistry*. 1983. Vol. 13, No. 5. P. 453–468. doi: 10.1016/0020-1790(83)90002-1.
5. Powell G., Tosh C.R., Hardie J. Host-plant selection by aphids: Behavioral, evolutionary, and applied perspectives. *Annual Review of Entomology*. 2006. Vol. 51. P. 309–330. doi: 10.1146/annurev.ento.51.110104.151107.
6. Ferry N., Edwards M.G., Gatehouse J.A. Plant-insect interactions: molecular approaches to insect resistance. *Current Opinion in Biotechnology*. 2004. Vol. 15, Iss. 2. P. 155–161. doi: 10.1016/j.copbio.2004.01.008.
7. Zuckerkandl E., Pauling L. Evolutionary divergence and convergence in proteins. Edited in *Evolving Genes and Proteins*. New York: Academic Press, 1965. P. 97–166.
8. Kumar S., Stecher G., Tamura K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*. 2016. Vol. 33. P. 1870–1874.

**VORONOVA N.V., VARABYOVA M.M., BANDARENKA Y.V.**

*Belarusian State University,*

*Belarus, Minsk, e-mail: nvoronova@bsu.by, ninavvoronova@gmail.com*

## **CYP4 AND CYP6 GENE VARIABILITY IN GENOME OF APHIS FABAE MORDVILKOI BÖRNER & JANISCH, 1922**

**Aim.** To estimate the variability of genes of 4th and 6th families of CYP450, which were extracted from the whole genome data of *Aphis fabae mordvilkoii* collected from *Philadelphus coronaries* L. in Belarus. **Methods.** The whole genome sequencing was carried out in the University of Utah DNA Sequencing and Genomic Core Facilities (USA). CYP4 and CYP6 gene sequences were extracted from the whole genome data by sequential mapping the whole genome reads to CYP4 and CYP6 CDSs of three reference genomes (*Acyrtosiphon pisum* Harris, 1776, *Myzus persicae* (Sulzer, 1776) и *Diuraphis noxia* (MordvilkoexKurdjumov, 1913)). All found unique version of assembling were taken as a single gene. **Results.** In *A. fabae mordvilkoii* genome we found out 31 CYP4 genes and 24 from them were copies of CYP4C1s. We also found out 19 CYP6 gene sand 8 from them were identified as CYP6A13s. Variability of nucleotide an damino acid sequences of CYP4 and CYP6 CDSs were high. **Conclusions.** In *A. fabae mordvilkoii* genome most CYP4 genes were identified as CYP4C1 and most CYP6 genes were CYP6A13s. Other CYP4 and CYP6 were mostly presented as single copies of different genes.

**Keywords:** aphids, cytochrome p450, *Aphis fabae*, trophic specialization, gene copies.

**Воронова Н.В., Вороб'єва М.М., Бондаренко Ю.В.**

*Білоруський Державний Університет,*

*Білорусь, м. Мінськ, e-mail: nvoronova@bsu.by, ninavvoronova@gmail.com*

## **ВАРІАБЕЛЬНІСТЬ ГЕНІВ СYP4 І СYP6 У ПОПЕЛИЦІ APHIS FABAE MORDVILKOI BÖRNER & JANISCH, 1922**

**Мета.** Оцінити варіабельність генів СYP450 4-го і 6-го сімейств, витягнутих їхніх даних про повногеномне секвенування попелиці *Aphis fabae mordvilkoii*, що розвивається в Білорусі, на бузок (*Philadelphus coronaries* L.). **Методи.** Секвенування генома попелиць провели в Центрі секвенування ДНК Університету штату Юта (Солт-Лейк-Сіті, США). Послідовності генів СYP4 і СYP6 були витягнуті з пулу прочиток генома шляхом послідовної фільтрації вирівнювань прочиток по референсним послідовностям генів СYP450 4-го і 6-го сімейства, витягнутих з 3 знаходяться у відкритому доступі геномів попелиць (*Acyrtosiphon pisum* Harris, 1776, *Myzus persicae* (Sulzer, 1776) і *Diuraphis noxia* (MordvilkoexKurdjumov, 1913)). Всі виявлені унікальні варіанти збірки оцінювали як індивідуальний ген. **Результати.** У геномі *A. fabae mordvilkoii* виявлений 31 ген СYP4, з яких 24 є варіанти гена СYP4C1. У сімействі СYP6 виявлено 19 генів, 8 з яких є варіанти СYP6A13. Варіабельність нуклеотидних і амінокислотних послідовностей СYP4 і СYP6 у *A. fabae mordvilkoii* виявилася, рівною мірою, високою. **Висновки.** У геномі попелиці *A. fabae mordvilkoii* серед генів СYP450 сімейства СYP4 переважають гени СYP4C1, а гени СYP6 представлені переважно генами СYP6A13. Решта гени зазначених родин представлені в геномі *A. fabae mordvilkoii* одиничними копіями.

**Ключові слова:** попелиці, цитохроми p450, *Aphis fabae*, трофічна спеціалізація, копійність генів.