

СРАВНИТЕЛЬНЫЙ АНАЛИЗ НУКЛЕОТИДНОЙ ПОСЛЕДОВАТЕЛЬНОСТИ ГЕНОВ И АМИНОКИСЛОТНОЙ ПОСЛЕДОВАТЕЛЬНОСТИ БЕЛКОВ ОСНОВНЫХ ФЕРМЕНТОВ СИСТЕМЫ ДЕТОКСИКАЦИИ НАСЕКОМЫХ

Н.В. Воронова, С.А. Белая, М.М. Воробьева

Белорусский государственный университет, Минск, Республика Беларусь

e-mail: nvoronova@gmail.ru; merlana.1995@mail.ru; masch.89@mail.ru

Введение

Насекомые – наиболее успешный таксон животных, среди всех обитающих на планете. Это крупный инфракласс, включающий более 1,5 млн. видов, освоивших все возможные среды обитания и источники пищи. Многие насекомые, особенно синантропы и переносчики возбудителей заболеваний человека, в течение десятилетий находятся под постоянным давлением все более совершенных инсектицидов. Однако, как показывают многочисленные исследования, насекомые демонстрируют чрезвычайно высокую способность формировать устойчивость к действию токсинов [1, 2]. Природа такой устойчивости до сих пор окончательно не установлена, однако одним из ключевых звеньев ее реализации являются ферменты системы детоксикации и, в частности, цитохромы р450 [3].

Особое место среди насекомых занимают растительноядные виды. Фитофаги связаны с растениями, на которых они питаются, миллионами лет сопряженной эволюции. Растения за это время выработали комплекс приспособлений, направленных против их вредителей. Большинство видов растений продуцируют вещества, формирующие многокомпонентную биохимическую систему защиты от фитофагов, это как токсины прямого действия, так и вещества модифицирующие развитие насекомых или репелленты [4]. В свою очередь, насекомые ответили на растущий с течением эволюционного времени пресс фитотоксинов совершенствованием системы детоксикации – в первую очередь, комплекса ферментов цитохромов р450.

У насекомых за трансформацию ксенобиотиков отвечают цитохромы 4-го и 6-го семейств (СYP4 и СYP6 соответственно). Это большая группа белков, кодируемая, по последним оценкам, десятками генов (в некоторых случаях более сотни), локализованными в ядерной ДНК. Как известно, в геноме насекомых разных видов, даже в случае их принадлежности к одному семейству, обнаруживают разное количество генов СYP [5, 6]. Кроме того, существует мнение, что эволюция системы СYP 450 у насекомых может протекать именно по пути увеличения числа копий отдельных генов [7], что не только увеличивает их суммарную экспрессию, но и предоставляет материал в виде генов-паралогов для их независимой эволюции и получения продуктов с новыми свойствами. Как результат, в течение нескольких поколений отбора устойчивость насекомых и, в частности, тлей к инсектицидам может возрасти в десятки, а иногда в сотни тысяч раз [8].

Белки, равно как и гены СYP 450 представляют собой логичную мишень в борьбе с устойчивостью насекомых. Обычно предлагается создание и использование совместно с инсектицидом селективных ингибиторов монооксигеназ [9, 10], либо разработка препаратов, вероятно, на основе микро-РНК, блокирующих экспрессию целевых цитохромов [11, 12]. Однако и тот и другой подход в настоящее время остаются лишь делом будущего. Во многом это связано с тем, что разнообразие белков и генов СYP4 и СYP6 у насекомых очень велико. Изучение молекулярного разнообразия последовательностей генов СYP 450 до сих пор, в основном, проводится лишь на модельных организмах и ограниченном числе лабораторных линий, что значительно снижает возможность экстраполяции данных даже на родственные таксоны. В этой работе мы поставили целью изучить варибельность известных последовательностей СYP4 и СYP6 в разных таксонах насекомых, используя методы биоинформационного анализа.

Методы исследования

В работе были использованы последовательности, депонированные в GenBankNCBI (<http://www.ncbi.nlm.nih.gov/genbank/>). Всего было использовано 790 последовательностей генов 15 видов насекомых из 4 отрядов, а именно: Перепончатокрылые (пчелы, шмели и муравьи), Двукрылые (комары, настоящие мухи и плодовые мушки), Полужесткокрылые (тли) и Чешуекрылые (серпокрылые моли).

Последовательности были выровнены в программе MEGA6 поочередным использованием алгоритмов Muscle и Clustal при установленном разрешении на включение пробелов в выровненные последовательности [13]. Аминокислотные последовательности получили прямой трансляцией (таблица кода трансляции № 1).

Уровень нуклеотидного и аминокислотного сходства были рассчитаны для выравненных последовательностей, сгруппированных в соответствии с семейством в первом случае и в соответствии с таксоном – во втором случае. Расчеты провели в программе SIAS и выразили в долях от единицы (<http://imed.med.ucm.es/Tools/sias.html>). Оценка энтропии нуклеотидной композиции была проведена в программе BioEdit v.7.2.5. [14]. Высокой считали энтропию, значение которой превышало единицу.

Генные деревья и деревья аминокислотных последовательностей построили в программе MEGA6 методом минимума эволюции (ME) с частичным удалением пробелов в выравненных последовательностях. Статистическую значимость достоверности топологии ветвей оценили с использованием бутстрэп-индекса для 500 реплик.

Результаты и обсуждение

Для анализа нуклеотидных и аминокислотных последовательностей были выбраны наиболее часто используемые подходы, позволяющие оценить как уровень подобия между последовательностями в выборках (оценка среднего уровня идентичности между последовательностями в группе и средней уровень сходства между последовательностями в выборке), так и предсказуемость появления конкретного нуклеотида в каждом сайте на протяжении гена (так называемое, значение «энтропии-на-сайт»). Результаты выравнивания последовательностей и оценка энтропии на всем протяжении генов показали, что энтропия последовательностей обоих семейств CYP450 чрезвычайно высока. В нуклеотидных последовательностях генов CYP4 и CYP6 насекомых отсутствуют сколько-нибудь протяженные консервативные области (рисунок 1). Падение уровня энтропии после 1600-го нуклеотида в обоих семействах генов связано не с повышением определенности нуклеотидного состава, а с уменьшением числа последовательностей в анализе, что связано с преобладанием в выборке более коротких последовательностей.

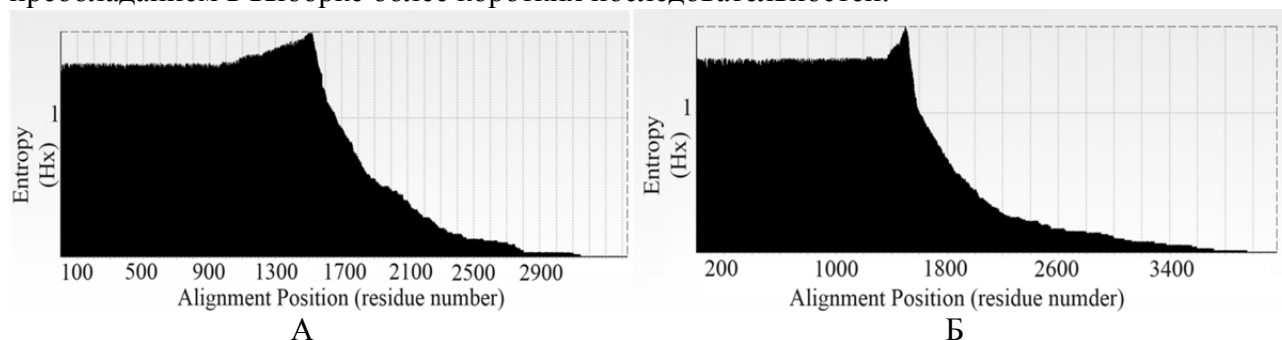


Рисунок 1 – Энтропия нуклеотидной последовательности CYP4 (А) и CYP6 (Б) для общей выборки насекомых разных отрядов на всем протяжении гена

Средние значения индекса идентичности как нуклеотидных, так и аминокислотных последовательностей были чрезвычайно низки (не превышали 60%) вне зависимости от типа формирования выборок (таблицы 1 и 2). Средние значения индекса сходства последовательностей были заметно выше и, в некоторых случаях, когда расчет производился для группы, объединяющей последовательности, относящиеся к одному подсемейству белков, достигали 90%.

Таблица 1 – Анализ нуклеотидных и аминокислотных последовательностей генов СУР4 насекомых разных отрядов

Род насекомых	Идентичность	Сходство	Подсемейство белков	Идентичность	Сходство
Нуклеотидные последовательности					
<i>Acromyrmex</i>	0,40±0,07	0,52±0,18	A	0,42±0,08	0,78±0,05
<i>Aedes</i>	0,48±0,04	0,85±0,01	C	0,42±0,13	0,40±0,08
<i>Acyrtosiphon</i>	0,45±0,11	0,75±0,07	D	0,49±0,14	0,83±0,06
<i>Apis</i>	0,43±0,03	0,77±0,06	E	0,55±0,04	0,84±0,02
<i>Bombus</i>	0,42±0,09	0,74±0,10	G	0,49±0,07	0,78±0,04
<i>Culex</i>	0,48±0,06	0,84±0,03	H	0,52±0,02	0,87±0,01
<i>Drosophila</i>	0,44±0,07	0,79±0,04	J	0,54±0,01	0,87±0,002
<i>Harpegnathos</i>	0,41±0,09	0,71±0,07	P	0,51±0,05	0,51±0,05
<i>Musca</i>	0,45±0,07	0,80±0,04	S	0,60±0,08	0,79±0,06
<i>Solenopsis</i>	0,43±0,12	0,72±0,09	V	0,42±0,08	0,80±0,04
Аминокислотные последовательности					
<i>Acromyrmex</i>	0,35±0,14	0,78±0,04	A	0,39±0,17	0,79±0,05
<i>Acyrtosiphon</i>	0,35±0,16	0,77±0,07	C	0,31±0,13	0,77±0,06
<i>Apis</i>	0,31±0,03	0,80±0,01	D	0,39±0,12	0,82±0,05
<i>Bombus</i>	0,33±0,14	0,80±0,04	E	0,57±0,10	0,87±0,04
<i>Culex</i>	0,38±0,11	0,83±0,04	G	0,42±0,16	0,81±0,06
<i>Drosophila</i>	0,35±0,10	0,81±0,04	P	0,48±0,13	0,85±0,04
<i>Harpegnathos</i>	0,31±0,16	0,73±0,07	S	0,67±0,02	0,90±0,01
<i>Musca</i>	0,35±0,12	0,79±0,05	V	0,29±0,10	0,79±0,04
<i>Plutella</i>	0,31±0,13	0,79±0,04	–	–	–
<i>Solenopsis</i>	0,31±0,20	0,79±0,06	–	–	–

Таблица 2 – Анализ нуклеотидных и аминокислотных последовательностей генов СУР6 насекомых разных отрядов

Род насекомых	Идентичность	Сходство	Подсемейство белков	Идентичность	Сходство
Нуклеотидные последовательности					
<i>Acyrtosiphon</i>	0,51±0,13	0,86±0,05	A	0,45±0,06	0,83±0,07
<i>Apis</i>	0,47±0,02	0,87±0,001	B	0,45±0,08	0,85±0,04
<i>Drosophila</i>	0,37±0,12	0,87±0,03	D	0,51±0,05	0,89±0,02
<i>Harpegnathos</i>	0,48±0,07	0,83±0,04	G	0,48±0,09	0,86±0,03
<i>Musca</i>	0,33±0,08	0,84±0,06	K	0,44±0,07	0,84±0,03
<i>Plutella</i>	0,46±0,09	0,85±0,04	M	0,56±0,12	0,91±0,03
<i>Solenopsis</i>	0,47±0,12	0,85±0,03	N	0,58±0,06	0,92±0,01
–	–	–	T	0,44±0,04	0,87±0,01
–	–	–	Z	0,61±0,03	0,93±0,01
Аминокислотные последовательности					
<i>Acyrtosiphon</i>	0,47±0,14	0,87±0,04	A	0,36±0,08	0,83±0,05
<i>Apis</i>	0,44±0,10	0,85±0,06	B	0,34±0,05	0,83±0,03
<i>Bombus</i>	0,39±0,15	0,85±0,04	D	0,46±0,09	0,87±0,03
<i>Culex</i>	0,37±0,12	0,86±0,03	G	0,46±0,10	0,87±0,02
<i>Drosophila</i>	0,37±0,11	0,85±0,03	K	0,34±0,15	0,82±0,04
<i>Musca</i>	0,35±0,08	0,82±0,07	T	0,41±0,08	0,86±0,02
<i>Solenopsis</i>	0,38±0,18	0,82±0,05	–	–	–

Мы сравнили значения индексов сходства и идентичности, сформировав для каждого семейства генов по 2 дочерние серии выборок для нуклеотидных и аминокислотных последовательностей соответственно: (1) – выборки, включающие последовательности всех

подсемейств генов одного рода насекомых; и (2) – выборки, включающие последовательности генов одного подсемейства всех насекомых, независимо от их таксономической принадлежности. Оказалось, что и для СУР4, и для СУР6 нет значимых различий в средних значениях индексов как при сравнении последовательностей внутри таксономической группы насекомых, так и при сравнении последовательностей внутри подсемейства кодируемых генами белков. Говоря иначе, сходство последовательностей в обоих семействах генов было в равной степени низким, если сравнивать их в соответствии с филогенетической близостью насекомых, для чьих геномов они были расшифрованы, и, так же, если игнорировать филогенетическое родство насекомых и оценивать последовательности только в соответствии с классификацией кодируемого ими продукта (подсемейства белка).

Чтобы определить тенденции формирования сходства/различий нуклеотидных и аминокислотных последовательностей мы дополнительно построили генные деревья на основе расчетных генетических дистанций между анализируемыми последовательностями. Нами намеренно были использованы метод построения филограм, основанный на принципе минимума эволюции (ME), т.е. объединяющий последовательности в соответствии с минимальным количеством необходимых эволюционных событий для превращения одной последовательности в другую (ближайшую). Использование дистанционного метода, как мы полагали, позволит избежать большого числа ошибок, связанных с крайне высокой вариабельностью последовательностей на всем протяжении, в филогенетических построениях. Мы обнаружили, что кластеризация ветвей на деревьях происходит с высоким уровнем достоверности топологии, по многим узлам достигающим 100% (рисунки 2 и 3). Тем не менее, невозможно выявить закономерности в формировании кластеров: на одном и том же дереве обнаруживаются кластеры, объединяющие последовательности насекомых одного отряда, и кластеры, включающие последовательности одного конкретного подсемейства генов насекомых разных таксонов. Индекс бутстэпа, в то же время, как для кластеров первого, так и второго типа оставался неизменно высоким. Поэтапный анализ деревьев позволил нам заключить, что кластеры, формирующиеся в соответствии с таксоном насекомом более типичны для последовательностей генов, а кластеры, образованные в соответствии с подсемейством белка – продукта анализируемого гена – для последовательностей белков. Описанное явление, однако, как было сказано выше, не является правилом, а носит характер нечетко выраженной тенденции.

В начале статьи мы писали о важности изучения генов СУР4 и СУР6 насекомых, прежде всего тех, кто относится к числу фитофагов – вредителей сельскохозяйственных растений и других возделываемых культур, поскольку продукты именно этих генов отвечают у насекомых за способность не только преодолевать неспецифическую защитную реакцию растения, но и выдерживать инсектицидный пресс.

Полученные нами результаты, свидетельствующие о крайне низком уровне сходства между последовательностями генов, кодирующих одни и те же белки или белки, относящиеся к одному подсемейству, у разных насекомых доказывают, что экстраполирование данных, полученных для разных модельных объектов, может быть крайне затруднено, если вообще возможно. Даже насекомые, принадлежащие к близким таксонам, например, одному биологическому роду, как это было показано для *Drosophila*, могут демонстрировать крайне низкий уровень идентичности между генами одного подсемейства.

Выводы

Результаты проведенного анализа показали, что последовательности генов СУР4 и СУР6 насекомых обладают высоким уровнем нуклеотидной энтропии на всем протяжении и крайне низким уровнем генетического сходства и идентичности. Последовательности аминокислот, транслированные с этих генов, также не демонстрируют высоких значений аминокислотной идентичности и аминокислотного сходства.

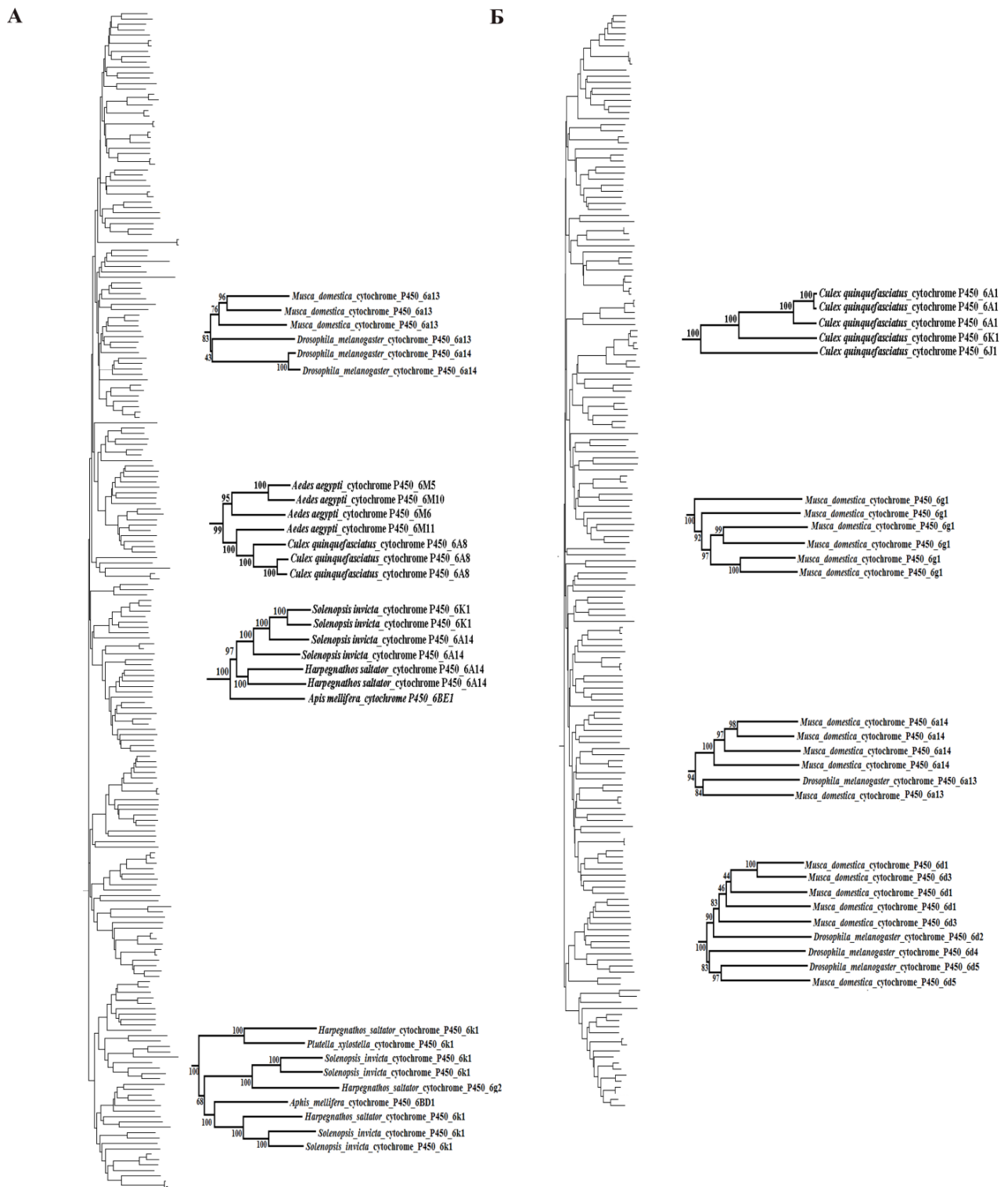


Рисунок 2 – Филогенетическое дерево нуклеотидных (А) и аминокислотных (Б) последовательностей СYP4 насекомых разных отрядов

Анализ филограмм, построенных методом МЕ, показал, что кластеры, формирующиеся в соответствии с таксоном насекомого более типичны для последовательностей генов, а кластеры, образованные в соответствии с подсемейством белка – продукта анализируемого гена – для последовательностей белков. Описанное явление не является правилом, а носит характер тенденции.

2. Markussen, M. Cytochrome P450 monooxygenase-mediated neonicotinoid resistance in the house fly *Musca domestica* L. / M. Markussen, M. Kristensen // *Pesticide Biochemistry and Physiology*. – 2010. – Vol. 98. – P. 50–58.
3. Feyereisen, R. Evolution of insect P450 / R. Feyereisen // *Biochemical Society Transactions*. – 2006. – Vol. 34 (6). – P. 1252–1255.
4. Neal, J.J. Inhibition of insect cytochromes P450 by furanocoumarins / J.J. Neal, D. Wu // *Pesticide Biochemistry and Physiology*. – 1994. – Vol. 50. – P. 43–50.
5. Lawrence, I.G. *Insect Molecular Biology and Biochemistry* / I.G. Lawrence. – Academic Press, 2011. – 574 с.
6. Feyereisen, R. Insect P450 enzymes / R. Feyereisen // *Annual Reviews Entomol.* – 1999. – Vol. 44. – P. 507–503.
7. Harrop, T. Evolutionary changes in gene expression, coding sequence and copy-number at the *cyp6g1* locus contribute to resistance to multiple insecticides in *Drosophila* / T. Harrop [et al.] // *PLOS ONE*. – 2014. – Vol. 9 (3). – P. 1–8.
8. Koo, H.N. Regional susceptibilities to 12 insecticides of melon and cotton aphid, *Aphis gossypii* (Hemiptera: Aphididae) and a point mutation associated with imidacloprid resistance / H.N. Koo [et al.] // *CropProt.* – 2014. – Vol. 55. – P. 91–97.
9. Steele, L. Selective sweep analysis in the genomes of the 91-R and 91-C *Drosophila melanogaster* strains reveals few of the ‘Usual Suspects’ in dichlorodiphenyltrichloroethane (DDT) resistance / L. Steele [et al.] // *PLOS ONE*. – 2015. – Vol. 10 (3). – P. 1–8.
10. Berge, J. Cytochrome P450 monooxygenases and insecticide resistance in insects / Bergé J., Feyereisen R, Amichot M. // *The Royal Society*. – 1998. – Vol. 353. – P. 1701–1705.
11. Оберемок, В.В. Современные инсектициды: их преимущества, недостатки и предпосылки к созданию ДНК-инсектицидов / В.В. Оберемок, А.С. Зайцев // *Ученые записки Таврического национального университета им. В.И. Вернадского*. – 2014. – Т. 27 (66). – С. 112–126.
12. Knipple, Gu. L. Recent advances in RNA interference research in insects: Implications for future insect pest management strategies / Gu L. Knipple // *Crop Protection*. – 2013. – Vol. 45. – P. 36–42.
13. Kumar, S. A biologist-centric software for evolutionary analysis of DNA and protein sequences / S. Kumar [et al.] // *Briefings in Bioinformatics*. – 2008. – Vol. 9. – P. 299–306.
14. Hall, T.A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT / T.A. Hall // *Nucl. Acids. Symp. Ser.* – 1999. – Vol. 41. – P. 95–98.

**COMPARATIVE ANALYSIS OF THE NUCLEOTIDE SEQUENCES OF GENES
AND THE AMINO ACID SEQUENCES OF PROTEINS OF THE MAJOR ENZYMES OF
THE INSECT DETOXIFICATION SYSTEM**

N.V. Voronova, S.A. Belaya, M.M. Varabyova

Belarusian State University, Minsk, Belarus

e-mail: nvoronova@gmail.ru

The results of studying the level of genetic variability of known CYP4 and CYP6 sequences of different taxa of insects are represented. 790 sequences from 15 insect species, which belong to 4 orders, were used in the work. It was established that sequences of CYP4 and CYP6 genes of insects have the high level of nucleotide Entropy (He) and a low level of nucleotide identity and similarity as well as amino acid ones. The analysis of phylograms, which were constructed separately for nucleotide and amino acid sequences, indicates that the clusters formed in accordance with the insect taxa were more typical for the nucleotide sequences, but clusters formed in accordance with subfamily of proteins were more common in the amino acid sequences. A low level of similarity and identity of the nucleotide sequences of the same subfamily of CYP450 that were obtained from the different taxa of insects does not allow extrapolating the data for different model objects.