

cell data represented cell cycle. Finally, we estimated the optimal number of components for extracting individual cell types in pancreas scRNA-seq data.

REFERENCES

1. Sompairac N., Nazarov P. V., Czerwinska U. et al. Independent Component Analysis for Unraveling the Complexity of Cancer Omics Datasets // *Int. J. Mol. Sci.* 2019. Vol. 20, № 18. P. 4414–4441.
2. Nazarov P. V., Wienecke-Baldacchino A. K., Zinovyev A. et al. Deconvolution of transcriptomes and miRNomes by independent component analysis provides insights into biological processes and clinical outcomes of melanoma patients // *BMC Med. Genomics*. 2019. Vol. 12, №1, P. 132–149.
3. McLendon R., Friedman A., Bigner D. et al. Comprehensive Genomic Characterization Defines Human Glioblastoma Genes and Core Pathways // *Nature*. 2008. Vol. 455, № 7216. P. 1061–1068.
4. Bailey P., Chang D. K., Nones K. et al. Genomic Analyses Identify Molecular Subtypes of Pancreatic Cancer // *Nature*. 2016. Vol. 531, № 7592. P. 47–52.
5. Dirkse A., Golebiewska A., Buder T. et al. Stem cell-associated heterogeneity in Glioblastoma results from intrinsic tumor plasticity shaped by the microenvironment // *Nat Commun*. 2019. Vol. 10, № 1787.
6. Enge M., Arda H. E., Mignardi M. et al. Single-Cell Analysis of Human Pancreas Reveals Transcriptional Signatures of Aging and Somatic Mutation Patterns // *Cell*. 2017. Vol. 171, №2. P. 321–330.

СОСТАВ И ВАРИАБЕЛЬНОСТЬ ГЕНОВ GST И UDT В ГЕНОМЕ APHIS CRACCIVORA

Р. С. Шулинский, Н. В. Воронова

*Белорусский государственный университет,
биологический факультет, Минск, Беларусь
E-mail: bio.shulinsk@bsu.by, nvoronova@bsu.by*

В работе представлены результаты аннотации генов GST и UDT в геноме *Aphis craccivora*, также проведен сравнительный анализ генного состава с другими геномами тлей. Показано, что генный состав семейств UDT и GST в геноме *A. craccivora* является типичным для тлей рода *Aphis*, однако отличаясь от такового у тлей других родов. Среди особенностей этого рода можно выделить уникальные гены 2A2 и 2B33, а также отсутствие генов 1_1.

Ключевые слова: геномика; тли; системы детоксикации.

Введение. В детоксикации ксенобиотиков участвуют три группы ферментов, которые поочередно преобразуют субстрат. При изолированном изучении генов, продукты которых участвуют в детоксикации на ее первой стадии (CYP450 и/или эстеразы и карбоксил-эстеразы) невозможно получить полную картину работы системы детоксикации. В связи

с этим изучение генов, продукты которых «включаются» на последующем этапе – GST и UDT является фундаментальной задачей. Тли, относящиеся к группе полифагов, такие как *Aphis craccivora*, имеют ряд особенностей в числе и вариабельности генов системы детоксикации, в связи с чем они способны метаболизировать большое число вторичных метаболитов растений, в т.ч. обладающих инсектицидным действием. Учитывая потенциальную связь между полифагией и формированием устойчивости к инсектицидам, понимание особенностей организации системы детоксикации у полифагов позволит не просто получить новые знания, но сформировать целостное представление об эволюции, организации и работе системы детоксикации у тлей. Получение новых знаний в этой области чрезвычайно важно, поскольку без создания целостной картины о генетической основе полифагии как ключевого приспособления, непосредственно связанного с устойчивостью к инсектицидам, невозможно предложить новые, действенные подходы к осуществлению контроля над процессом формирования устойчивости к инсектицидам у насекомых-вредителей, таких как тли.

Материалы и методы. Полногеномное секвенирование *A. craccivora* было проведено на платформе MiSeq с использованием библиотеки TruSeq 350. При процессинге прочтений были проведены оценка качества прочтений и тримминг некачественных оснований в FastQC и Trimmomatic, соответственно. Фильтрация эукариотических прочтений проводилась путем маппинга по всем бактериальным геномам в программе Bowtie2 (sensitive local alignment) с последующим отбором невыравненных прочтений.

Вышеуказанные прочтения использовались в дальнейшей сборке в ассемблере AbySS2.0. Аннотация проводилась в пайплайне Maker. Мы использовали итеративную стратегию обучения скрытых марковских моделей. Первым этапом являлось обучение скрытых в программе snar на моделях генов, полученных в результате выравнивания EST. Далее предсказанные модели snar использовались для переобучения snar, а также обучения AUGUSTUS. В качестве самообучающегося HMM предиктора был выбран genemark. Финальные модели генов синтезировались как консенсусные с HMM предсказаний, а также EST и белковой гомологии в программе Evidence Modeller.

Идентификация генов GST и UDT проводилась при помощи BLAST, а также с использованием поиска IPR и Pfam-доменов в InterProScan. Курирование аннотации целевых моделей генов проводили в геномном браузере Apollo.

Результаты. Количество, а также сравнительная характеристика состава генов данной группы с их составом в геноме других тлей представлена в таблицах 1 и 2.

Таблица 1

Количество генов UDT в геномах различных видов тлей

UDP ген	<i>Acyrtosiphon pisum</i>	<i>Aphis gossypii</i>	<i>Diuraphis noxia</i>	<i>Sipha flava</i>	<i>Aphis craccivora</i>
1_3	1	0	0	0	0
2C1	15	12	18	8	8
2B15	10	4	5	6	9
2B13	5	5	1	5	3
2A1	1	0	0	0	1
2B17	3	2	10	2	3
2B7	2	1	0	2	1
2B1	6	4	3	4	3
2B10	1	3	0	1	3
2B2	14	17	4	6	16
2A3	1	1	7	3	7
2B37	4	1	0	0	4
2B18	1	0	0	4	0
1-7C	4	1	1	0	1
2B14	1	0	0	1	0
2B20	8	2	1	0	1
1_1	3	0	1	3	0
1_7	6	1	4	0	1
1_5	0	1	0	2	1
2B16	2	0	0	0	0
2B30	0	0	0	1	0
2B9	0	0	0	5	0
2B19	0	4	0	1	1
2A2	0	2	0	0	1
2B33	0	5	0	0	1
2b23	0	0	1	0	0
1_9	0	0	1	0	0
Total	88	66	57	54	65

Таблица 2

Количество генов GST в геномах различных видов тлей

GST ген	<i>Acyrtosiphon pisum</i>	<i>Aphis gossypii</i>	<i>Diuraphis noxia</i>	<i>Sipha flava</i>	<i>Aphis craccivora</i>
N	23	10	10	14	6

Таким образом можно заключить что генный состав семейств UDT и GST в геноме *A. craccivora* является типичным для тлей рода *Aphis*, од-

нако отличаясь от такового у тлей других родов. Среди особенностей этого рода можно выделить уникальные гены 2A2 и 2B33, а также отсутствие генов 1_1. Относительно генов GST можно отметить, что геном *A. craccivora* имеет наименьшее количество этих генов среди вышеуказанных геномов.

Следует отметить, что UDT относятся к числу наименее изученных белков системы детоксикации насекомых и тлей, в частности.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Armstrong J., Fiddes I. T., Diekhans M., Paten B. Whole-Genome Alignment and Comparative Annotation // *Annu. Rev. Anim. Biosci.* 2019. Vol. 7. P. 41–64.
2. Dewey, C. N. Whole-Genome Alignment // *Methods Mol. Biol.* 2012. Vol. 855. P. 237–257.
3. Narzisi G., Mishra B. Comparing De Novo Genome Assembly: The Long and Short of It // *PLoS One*. 2011. Vol. 6. № 4.
4. Wenger J. A., Cassone B. J., Legeai F. et al. Whole genome sequence of the soybean aphid, *Aphis glycines* // *Insect Biochem. Mol. Biol.* 2017.
5. Gouin A., Legeai F., Nouhaud P. et al. Whole-genome re-sequencing of non-model organisms: lessons from unmapped reads. *Heredity*. 2015. Vol. 114. P. 494–501.
6. Langmead B., Salzberg S. Fast gapped-read alignment with Bowtie 2 // *Nature Methods*. 2012. Vol. 9. P. 357–359.
7. Simpson J. T., Wong K., Jackman S. D. et al. ABySS: A parallel assembler for short read sequence data // *Genome Res.* 2009. Vol. 19. № 6. P. 1117–1123.
8. Cantarel B. L., Korf I., Robb S. M. C. et al. MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes // *Genome Res.* 2008. Vol. 18. №1. P. 188–196.
9. Tempel S. Using and Understanding RepeatMasker. *Mobile Genetic Elements*. New York: Humana Press. 2012. P. 29–51.