

## ПРИЛОЖЕНИЯ

к статье А.Ю. Пронозина, М.К. Брагиной, Е.А. Салиной  
«Пангеномы сельскохозяйственных растений»

### Приложение 1

#### Основные характеристики пангеномов растений

Название вида	Кол-во образцов	Стратегия сборки	Общий размер пангенома (кол-во генов)	Универсальные гены, %	Вариабельные гены, %	Литературный источник
Резуховидка Таля ( <i>Arabidopsis thaliana</i> )	19	Сборка-выравнивание	37.789	69.7	30.3	Contreras-Moreira et al., 2017
Трахиния двуколосковая ( <i>Brachypodium distachyon</i> )	54		37.886	54	46	Gordon et al., 2017
Капуста ( <i>Brassica oleracea</i> )	10 <sup>a</sup>	Выравнивание-сборка	61.379	81.3	18.7	Golicz et al., 2016
Капсикум ( <i>Capsicum</i> )	355		51.757	55.7	44.3	Ou et al., 2018
Люцерна ( <i>Medicago</i> )	15	Сборка-выравнивание	74.700	41.9	58.1	Zhou et al., 2017
Рис ( <i>Oryza sativa</i> ), Asian rice	66 <sup>a</sup>		42.580	61.9	38.1	Zhao et al., 2018
Рис ( <i>Oryza sativa</i> ), Asian rice	453/3.000		46.115 <sup>b</sup> /47.288 <sup>b</sup>	52.9/61.3	47.1/28.7	Sun et al., 2016
Рис ( <i>Oryza sativa</i> ), African rice	120 культурных/ 74 диких		35.198/36.252	86.5/98.6	13.5/1.4	Monat et al., 2018
Соя ( <i>Glycine max</i> )	7 <sup>a</sup>		59.080	80.1	19.9	Li et al., 2010
Пшеница мягкая ( <i>Bread wheat</i> )	18	Выравнивание-сборка	128.656	64.3	33.7	Montenegro et al., 2017
Рапс ( <i>Brassica napus</i> )	53		94.013	62	38	Hurgobin et al., 2017
Томат ( <i>Solanum lycopersicum</i> )	725		40.369	74.2	–	Gao et al., 2019
Репа ( <i>Brassica rapa</i> )	3	Сравнительная реконструкция	36.882	84.56	15.44	Lin et al., 2014
Подсолнечник однолетний ( <i>Helianthus annuus</i> )	493	Сборка-выравнивание	45.302	72.7	5.4	Hubner et al., 2019
Кукуруза ( <i>Zea mays</i> )	503		40.500	63	37	Hirsch et al., 2014

Примечание. <sup>a</sup> Дикие и культивируемые; <sup>b</sup> подтип *indica*; <sup>c</sup> подтип *japonica*.

## Приложение 2

### Методы биоинформатического анализа пангенома растений

Название вида	Метод анализа	Литературный источник
Кукуруза ( <i>Zea mays</i> )	Транскриптомные библиотеки подготовлены для каждого исследуемого генотипа для семи растений. Для каждой библиотеки осуществлен контроль качества прочтений с помощью FastQC. Для расчета уровня экспрессии и SNP последовательности прочтений для каждой библиотеки выровнены относительно референсного генома кукурузы с помощью Bowtie версии 0.12.7 и TopHat версии 1.4.1	Hirsch et al., 2014
Соя ( <i>Glycine max</i> , <i>Glycine soja</i> )	Сборка каждого генома выполнена с помощью SOAPdenovo. Гены, кодирующие белки, идентифицированы для каждого генома на основе гомологии, оценки уровня экспрессии мРНК и предсказания структуры гена с помощью программы Augustus. Верификация предсказанных генов проведена путем сравнения с белковыми последовательностями 18 видов растений с применением BLASTP. Для идентификации ортологов использована БД OrthoMCL. Белковые последовательности 670 семейств выровнены при помощи MUSCLE. Филогенетические деревья построены с применением метода Neighbor-joining в пакете MEGA. Набор основных и изменчивых генов оценивали, основываясь на кластеризации семейств генов OrthoMCL, найдены семейства генов, присутствующих в образцах <i>G. soja</i> , которые можно считать основными генными семействами. Гомозиготные SNP и небольшие делеции/вставки (меньше 100 п. н.) найдены из результатов выравнивания контигов программой MUMmer. При помощи SAMtools на основе выравнивания коротких прочтений определены гетерозиготные SNP	Li et al., 2010
Рапс ( <i>Brassica napus</i> )	Собранные контиги сравниваются с базой данных NCBI с использованием BLAST+. Анализ генов PAV осуществлен с применением пакета SGSGeneLoss. Прочтения выровнены относительно пангенома программой Bowtie2; выровненные прочтения анализировали с помощью SAMtools и ERANET-ASSYST. Функциональную аннотацию проводили с помощью Blast2GO, кластеризацию генов – OrthoMCL v2.09. Гены исследуемого пангенома кластеризовали вместе с генами <i>A. thaliana</i> (TAIR 10). Поиск генов устойчивости к патогенам (R гены) в пангеноме проводили с помощью InterProScan, основываясь на доменах из баз данных Pfam и Coils. SNPs найдены при помощи пакетов VCFtools и vcflib	Hurgobin et al., 2017
Рис ( <i>Oryza sativa</i> )	Для выравнивания геномов друг с другом применяли программу Nucmer, для поиска геном-специфических и общих областей генома - BEDTools. Программа jellyfish/AMOS использована для анализа <i>k</i> -меров с целью классификации уникальных и повторяющихся последовательностей. BEDTools использовали для определения пересечения геном-специфических/общих областей относительно соответствующих аннотаций с целью определить общие для всех геномов экзонные участки	Schatz et al., 2014
Пшеница мягкая ( <i>Triticum aestivum</i> )	Контроль качества и удаление адаптеров осуществлены с помощью TRIMMOMATIC v.0.33, последовательности короче 73 п. н. удалены. Программу VELVET v.1.2.10 применяли для сборки пангенома с размером <i>k</i> -меров 71. Для картирования прочтений ДНК использовали программу BOWTIE 2 v.2.2.5. Пакет SAMTOOLS применяли для расчета покрытия аннотированных генов относительно наличия/отсутствия для каждого гена. Прочтения RNA-seq выровнены относительно референсного генома с применением TOPHAT2 v.2.1.0.1. REPEATMASKER используется для маркировки повторяющихся областей с помощью БД RepBase. Программу AUGUSTUS v.2.1.0 использовали для предсказания генов с применением транскриптомных данных. Для оценки полноты сборки генома использована программа CEGMA. PANGP – для подсчета общего количества генов и основных генов для 19 генотипов	Montenegro et al., 2017
Капуста ( <i>Brassica oleracea</i> )	Кластеризация генов проведена программой OrthoMCL 58 v2.09 (со стандартными параметрами). Гены пангенома <i>B. oleracea</i> кластеризированы с генами <i>A. thaliana</i> (TAIR 10). Характеристика наличия/отсутствия генов проведена с использованием пакета SGSGeneLoss, на основе этого анализа проведено разделение семейств генов на универсальные и переменные.	Golicz et al., 2016

## Окончание приложения 2

Название вида	Метод анализа	Литературный источник
Люцерна усеченная ( <i>Medicago truncatula</i> )	15 геномов люцерны собраны с применением ALLPATHS-LG, аннотированы при помощи пакета AUGUSTUS. На основе сравнения каждой <i>de novo</i> сборки с референсом получен набор последовательностей новых генов. Последовательности низкого качества просканированы и удалены с помощью программ Dustmasker и Tandem Repeat Finder. Последовательности с потенциальной контаминацией отфильтрованы программой BLAST с использованием базы данных NCBI. Новые последовательности из 12 образцов собраны и выровнены с использованием Para-Mugsy. Построенный пангеном включает референсную последовательность как основу и новые переменные сегменты, определенные в других генотипах	Zhou et al., 2017
Томат ( <i>Solanum lycopersicum</i> )	Геномы из каждой выборки выровнены относительно пангенома с применением BWA-MEM 60 со стандартными параметрами. Наличие/отсутствие для каждого гена определено с использованием SGGeneLoss	Gao et al., 2019
Подсолнечник однолетний ( <i>Helianthus annuus</i> )	Проводили выравнивание прочтений из каждого образца на референсный геном. Дополнительно осуществляли сборку геномов <i>de novo</i> . Контиги из всех образцов объединяли в одну базу данных, содержащую все последовательности, не найденные в референсе. Набор объединенных данных кластеризовали при помощи CD-HIT. С целью определения полной совокупности генов для каждого образца сырые риды из каждого образца выравнивали относительно собранного пангенома. Плохо аннотированные гены удаляли, перекрывающиеся аннотации объединяли.	Hubner et al., 2019

### Приложение 3

#### Программы для обработки данных, используемые при анализе пангенома растений

Название программы	Функционал	Литературный источник	Ссылка в Интернете
trimmomatic	Фильтрация библиотек коротких прочтений	Bolger et al., 2014	<a href="http://www.usadellab.org/cms/index.php?page=trimmomatic">http://www.usadellab.org/cms/index.php?page=trimmomatic</a>
FastQC	Анализ библиотек коротких прочтений	Wingett et al., 2018	<a href="http://www.bioinformatics.babraham.ac.uk/projects/fastqc/">http://www.bioinformatics.babraham.ac.uk/projects/fastqc/</a>
bwa	Картирование коротких прочтений на референс без учета сплайсинга	Li, Durbin, 2009	<a href="http://bio-bwa.sourceforge.net/">http://bio-bwa.sourceforge.net/</a>
bowtie2	Картирование коротких прочтений на референс без учета сплайсинга	Langmead et al., 2012	<a href="http://bowtie-bio.sourceforge.net/bowtie2/index.shtml">http://bowtie-bio.sourceforge.net/bowtie2/index.shtml</a>
SOAPdenovo	Сборка <i>de novo</i> контигов из коротких прочтений секвенирования второго поколения	Luo et al., 2012	<a href="https://github.com/aquaskyline/SOAPdenovo2">https://github.com/aquaskyline/SOAPdenovo2</a>
ALLPATHS-LG	Сборка <i>de novo</i> контигов из библиотек коротких прочтений	Gnerre et al., 2011	<a href="http://software.broadinstitute.org/allpaths-lg/blog/">http://software.broadinstitute.org/allpaths-lg/blog/</a>
MaSuRCA	Сборщик геномов <i>de novo</i> из библиотек коротких ридов	Zimin et al., 2013	<a href="https://github.com/alekseyzimin/masurca">https://github.com/alekseyzimin/masurca</a>
Samtools	Обработка и анализ картирования последовательностей	Li et al., 2009	<a href="http://www.htslib.org/">http://www.htslib.org/</a>
Freebayes	Поиск однонуклеотидных делеций, вставок и замен	Garrison, Marth, 2012	<a href="https://github.com/ekg/freebayes">https://github.com/ekg/freebayes</a>
Tersect	Анализ однонуклеотидных делеций, инсерций и замен в пангеномных данных	Kurowski, Mohareb, 2020	<a href="https://github.com/tomkurowski/tersect">https://github.com/tomkurowski/tersect</a>
MUMmer	Выравнивание последовательностей в масштабах геномов	Kurtz et al., 2004	<a href="http://mummer.sourceforge.net/">http://mummer.sourceforge.net/</a>
lastz	Выравнивание нуклеотидных последовательностей	Harris, 2007	<a href="http://www.bx.psu.edu/~rsharris/lastz/">http://www.bx.psu.edu/~rsharris/lastz/</a>
ppsPCP	Поиск вариаций в представленности генов и аннотация пангенома	Tahir Ul Qamar et al., 2019	<a href="https://github.com/Zhuxitong/ppsPCP">https://github.com/Zhuxitong/ppsPCP</a>
EUPAN	Сборка индивидуальных геномов, поиск вариаций и аннотация пангенома	Hu et al., 2017	<a href="http://cgm.sjtu.edu.cn/eupan/">http://cgm.sjtu.edu.cn/eupan/</a>
InterProScan	Предсказание доменной структуры в представленных последовательностях	Jones et al., 2014	<a href="ftp://ftp.ebi.ac.uk/pub/software/unix/iprscan/5/5.39-77.0/">ftp://ftp.ebi.ac.uk/pub/software/unix/iprscan/5/5.39-77.0/</a>
blast	Выравнивание нуклеотидных или аминокислотных последовательностей	Camacho et al., 2009	<a href="https://blast.ncbi.nlm.nih.gov">https://blast.ncbi.nlm.nih.gov</a>
CEGMA	Оценка полноты представленного набора последовательностей путем поиска характерных фрагментов	Parra et al., 2007	<a href="http://korflab.ucdavis.edu/Datasets/">http://korflab.ucdavis.edu/Datasets/</a>
BUSCO	Поиск характерных последовательностей в представленной выборке для оценки полноты этой выборки	Simão et al., 2015	<a href="https://busco.ezlab.org/">https://busco.ezlab.org/</a>
PGAP	Осуществляет масштабный поиск генов, проводит функциональную аннотацию, обогащение кластеров ортологичных генов терминами онтологии, анализ эволюции видов, производит структурный анализ пангенома, идентификацию универсальной и варибельной части пангенома	Zhao et al., 2012	<a href="https://github.com/ncbi/pgap">https://github.com/ncbi/pgap</a>
BPGA	Проводит кластеризацию генов на основе сходства последовательностей, анализ наличия/отсутствия ортологов, построение графика зависимости размеров пангенома, реконструкцию филогенетического дерева, анализ метаболических путей	Cahudhari et al., 2019	<a href="https://github.com/jbadavis/bpga">https://github.com/jbadavis/bpga</a>
panX	Идентификация кластеров ортологичных генов	Ding et al., 2018	<a href="http://pangenome.tuebingen.mpg.de/">http://pangenome.tuebingen.mpg.de/</a>
Pan4Draft	Улучшение аннотации пангеномов за счет добавления к ней информации о последовательностях незавершенных геномов (unfinished genomes)	Veras et al., 2018	<a href="http://www.computationalbiology.ufpa.br/pan4drafts">http://www.computationalbiology.ufpa.br/pan4drafts</a>

## Список литературы / References

- Bolger A.M., Lohse M., Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114-2120.
- Cahudhari N.M., Gupta V.K., Dutta C. BPGA-an ultra-fast pan-genome analysis pipeline. *Sci. Rep.* 2019;6(1):1-10.
- Camacho C., Coulouris G., Avagyan V., Ma N., Papadopoulos J., Bealer K., Madden T.L. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009;10(1):421.
- Contreras-Moreira B., Cantalapiedra C.P., García-Pereira M.J., Gordon S.P., Vogel J.P., Igartua E., Casas A.M., Vinuesa P. Analysis of plant pan-genomes and transcriptomes with GET\_HOMOLOGUES-EST, a clustering solution for sequences of the same species. *Front. Plant Sci.* 2017;8:184.
- Ding W., Baumdicker F., Neher R.A. panX: pan-genome analysis and exploration. *Nucleic Acids Res.* 2018;46(1):e5-e5.
- Gao L., Gonda I., Sun H., Ma Q., Bao K., Tieman D.M., Burzynski-Chang E.A., Fish T.L., Stromberg K.A., Sacks G.L., Thannhauser T.W., Foolad M.R., Diez M.J., Blanca J., Canizares J., Xu Y., van der Knaap E., Huang S., Klee H. J., ..., Fei Z. The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. *Nat. Genet.* 2019;51(6):1044-1051. <https://doi.org/10.1038/s41588-019-0410-2>.
- Garrison E., Marth G. Haplotype-based variant detection from short-read sequencing. *arXiv preprint arXiv*. 2012:1207.3907.
- Gnerre S., MacCallum I., Przybylski D., Ribeiro F.J., Burton J.N., Walker B.J., Sharpe T., Hall G., Shea T.P., Sykes S., Berlin A.M., Aird D., Costello M., Daza R., Williams L., Nicol R., Gnirke A., Nusbaum C., Lander E.S., Jaffe D.B. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc. Natl. Acad. Sci. USA*. 2011;108(4):1513-1518.
- Golicz A.A., Bayer P.E., Barker G.C., Edger P.P., Kim H., Martinez P.A., Chan C.K., Severn-Ellis A., McCombie W.R., Parkin I.A., Paterson A.H., Pires J.C., Sharpe A.G., Tang H., Teakle G.R., Town C.D., Batley J., Edwards D. The pangenome of an agronomically important crop plant *Brassica oleracea*. *Nat. Commun.* 2016;7(1):1-8.
- Gordon S.P., Contreras-Moreira B., Woods D.P., Des Marais D.L., Burgess D., Shu S., Stritt C., Roulin A.C., Schackwitz W., Tyler L., Martin J., Lipzen A., Dochy N., Phillips J., Barry K., Geuten K., Budak H., Juenger T. E., Amasino R., Caicedo A.L., Goodstein D., Davidson P., Mur L.A.J., Figueroa M., Freeling M., Catalan P., Vogel J.P. Extensive gene content variation in the *Brachypodium distachyon* pan-genome correlates with population structure. *Nat. Commun.* 2017;8(1):1-13.
- Harris R.S. Improved Pairwise Alignment of Genomic DNA: PhD thesis, The Pennsylvania State University, 2007.
- Hirsch C.N., Foerster J.M., Johnson J.M., Sekhon R.S., Muttoni G., Vaillancourt B., Penagaricano F., Lindquist E., Pedraza M.A., Barry K., de Leon N., Kaeppeler S.M., Buell C.R. Insights into the maize pan-genome and pan-transcriptome. *Plant Cell*. 2014;26(1):121-135.
- Hu Z., Sun C., Lu K.C., Chu X., Zhao Y., Lu J., Shi J., Wei C. EUPAN enables pan-genome studies of a large number of eukaryotic genomes. *Bioinformatics*. 2017;33(15):2408-2409.
- Hubner S., Bercovich N., Todesco M., Mandel J.R., Odenheimer J., Ziegler E., Lee J.S., Baute G.J., Owens G.L., Grassa C.J., Ebert D.P., Ostevik K.L., Moyers B.T., Yakimowski S., Masalia R.R., Gao L., Čalić I., Bowers J.E., Kane N.C., Swanvellder D.Z.H., Kubach T., Muñoz S., Langlade N.B., Burke J.M., Rieseberg L.H. Sunflower pan-genome analysis shows that hybridization altered gene content and disease resistance. *Nat. Plants*. 2019;5(1):54-59.
- Hurgobin B., Edwards D. SNP discovery using a pangenome: has the single reference approach become obsolete. *Biology*. 2017;6(1):21.
- Jones P., Binns D., Chang H.Y., Fraser M., Li W., McAnulla C., McWilliam H., Maslen J., Mitchell A., Nuka G., Pesseat S., Quinn A.F., Sangrador-Vegas A., Scheremetjew M., Yong S.Y., Lopez R., Hunter S. InterProScan 5: genome-scale protein function classification. *Bioinformatics*. 2014;30(9):1236-1240.
- Kurowski T.J., Mohareb F. Tersect: a set theoretical utility for exploring sequence variant data. *Bioinformatics*. 2020;36(3):934-935.
- Kurtz S., Phillippy A., Delcher A.L., Smoot M., Shumway M., Antonescu C., Salzberg S.L. Versatile and open software for comparing large genomes. *Genome Biol.* 2004;5(2):R12.
- Langmead B., Salzberg S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods*. 2012;9(4):357.
- Li H., Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. 2009;25(14):1754-1760.
- Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25(16):2078-2079.
- Li Y.H., Li W., Zhang C., Yang L., Chang R.Z., Gaut B.S., Qiu L.J. Genetic diversity in domesticated soybean (*Glycine max*) and its wild progenitor (*Glycine soja*) for simple sequence repeat and single-nucleotide polymorphism loci. *New Phytol.* 2010;188(1):242-253.
- Lin K., Zhang N., Severing E.I., Nijveen H., Cheng F., Visser R.G.F., Wang X., de Ridder D., Bonnema G. Beyond genomic variation-comparison and functional annotation of three *Brassica rapa* genomes: a turnip, a rapid cycling and a Chinese cabbage. *BMC Genomics*. 2014;15(1):250.

- Luo R., Liu B., Xie Y., Li Z., Huang W., Yuan J., He G., Chen Y., Pan Q., Liu Y., Tang J., Wu G., Zhang H., Shi Y., Liu Y., Yu C., Wang B., Lu Y., Han C., Cheung D.W., Yiu S.M., Peng S., Xiaoqian Z., Liu G., Liao X., Li Y., Yang H., Wang J., Lam T.W., Wang J. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience*. 2012;1(1):18.
- Monat C., Tranchant-Dubreuil C., Engelen S., Labadie K., Paradis E., Tando N., Sabot F. Comparison of two African rice species through a new pan-genomic approach on massive data. *bioRxiv*. 2018:245431.
- Montenegro J.D., Golicz A.A., Bayer P.E., Hurgobin B., Lee H., Chan C.K., Visendi P., Lai K., Doležel J., Batley J., Edwards D. The pangenome of hexaploid bread wheat. *Plant J*. 2017;90(5):1007-1013.
- Ou L., Li D., Lv J., Chen W., Zhang Z., Li X., Yang B., Zhou S., Yang S., Li W., Gao H., Zeng Q., Yu H., Wang B., Dai X., Ma Y., Zou X. Pan-genome of cultivated pepper (*Capsicum*) and its use in gene presence-absence variation analyses. *New Phytol*. 2018;220(2):360-363.
- Parra G., Bradnam K., Korf I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics*. 2007;23(9):1061-1067.
- Schatz M.C., Maron L.G., Stein J.C., Wences A.H., Gurtowski J., Biggers E., Lee H., Kramer M., Antoniou E., Ghiban E., Wright M.H., Ming C.J., Ware D., McCouch S.R., McCombie W.R. Whole genome *de novo* assemblies of three divergent strains of rice, *Oryza sativa*, document novel gene space of *aus* and *indica*. *Genome Biol*. 2014;15(11):506.
- Simão F.A., Waterhouse R.M., Ioannidis P., Kriventseva E.V., Zdobnov E.M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015;31(19):3210-3212.
- Sun C., Hu Z., Zheng T., Lu K., Zhao Y., Wang W., Shi J., Wang C., Lu J., Zhang D., Li Z., Wei C. RPAN: rice pan-genome browser for ~3000 rice genomes. *Nucleic Acids Res*. 2016;45(2):597-605.
- Tahir Ul Qamar M., Zhu X., Xing F., Chen L.L. ppsPCP: a plant presence/absence variants scanner and pan-genome construction pipeline. *Bioinformatics*. 2019;35(20):4156-4158.
- Veras A., Araujo F., Pinheiro K., Guimarães L., Azevedo V., Soares S., Costa da Silva A., Ramos R. Pan4Draft: a computational tool to improve the accuracy of pan-genomic analysis using draft genomes. *Sci. Rep*. 2018;8(1):1-8.
- Wingett S.W., Andrews S. FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Res*. 2018;7:1338.
- Zhao Q., Feng Q., Lu H., Li Y., Wang A., Tian Q., Zhan Q., Lu Y., Zhang L., Huang T., Wang Y., Fan D., Zhao Y., Wang Z., Zhou C., Chen J., Zhu C., Li W., Weng Q., Xu Q., Wang Z.X., Wei X., Han B., Huang X. Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. *Nat. Genet*. 2018;50(2):278-284.
- Zhao Y., Wu J., Yang J., Sun S., Xiao J., Yu J. PGAP: pan-genomes analysis pipeline. *Bioinformatics*. 2012;28(3):416-418.
- Zhou P., Silverstein K.A., Ramaraj T., Guhlin J., Denny R., Liu J., Farmer A.D., Steele K.P., Stupar R.M., Miller J.R., Tiffin P., Mudge J., Young N.D. Exploring structural variation and gene family architecture with *De Novo* assemblies of 15 *Medicago* genomes. *BMC Genomics*. 2017;18(1):1-14.
- Zimin A.V., Marçais G., Puiu D., Roberts M., Salzberg S.L., Yorke J.A. The MaSuRCA genome assembler. *Bioinformatics*. 2013;29(21):2669-2677.