

High-quality genome assembly of *Capsella bursa-pastoris* reveals asymmetry of regulatory elements at early stages of polyploid genome evolution

Kasianov A., Klepikova A., Kulakovskiy I., Gerasimov E., Fedotova A., Besedina E., Kondrashov A., Logacheva M., Penin A.

Kazan Federal University, 420008, Kremlevskaya 18, Kazan, Russia

Abstract

© 2017 The Authors The Plant Journal © 2017 John Wiley & Sons Ltd Polyploidization and subsequent sub- and neofunctionalization of duplicated genes represent a major mechanism of plant genome evolution. *Capsella bursa-pastoris*, a widespread ruderal plant, is a recent allotetraploid and, thus, is an ideal model organism for studying early changes following polyploidization. We constructed a high-quality assembly of *C. bursa-pastoris* genome and a transcriptome atlas covering a broad sample of organs and developmental stages (available online at <http://travadb.org/browse/Species=Cbp>). We demonstrate that expression of homeologs is mostly symmetric between subgenomes, and identify a set of homeolog pairs with discordant expression. Comparison of promoters within such pairs revealed emerging asymmetry of regulatory elements. Among them there are multiple binding sites for transcription factors controlling the regulation of photosynthesis and plant development by light (PIF3, HY5) and cold stress response (CBF). These results suggest that polyploidization in *C. bursa-pastoris* enhanced its plasticity of response to light and temperature, and allowed substantial expansion of its distribution range.

<http://dx.doi.org/10.1111/tpj.13563>

Keywords

Capsella bursa-pastoris, genome assembly, homeologs, polyploidy, regulatory elements, RNA-seq

References

- [1] Agren, J.Å., Wang, W., Koenig, D., Neuffer, B., Weigel, D. and Wright, S.I. (2014) Mating system shifts and transposable element evolution in the plant genus *Capsella*. *BMC Genom.* 15, 602.
- [2] Anders, S. and Huber, W. (2010) Differential expression analysis for sequence count data. *Genome Biol.* 11, R106.
- [3] Aoki, K., Yano, K., Suzuki, A. et al. (2010) Large-scale analysis of full-length cDNAs from the tomato (*Solanum lycopersicum*) cultivar Micro-Tom, a reference system for the Solanaceae genomics. *BMC Genom.* 11, 210.
- [4] Benjamin, A.M., Nichols, M., Burke, T.W., Ginsburg, G.S. and Lucas, J.E. (2014) Comparing reference-based RNA-Seq mapping methods for non-human primate data. *BMC Genom.* 15, 570.
- [5] Borrill, P., Ramirez-Gonzalez, R. and Uauy, C. (2016) expVIP: a customizable RNA-seq data analysis and visualization platform. *Plant Physiol.* 170, 2172–2186.

- [6] Buggs, R.J.A., Wendel, J.F., Doyle, J.J., Soltis, D.E., Soltis, P.S. and Coate, J.E. (2014) The legacy of diploid progenitors in allopolyploid gene expression patterns. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369, 20130354.
- [7] Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T.L. (2009) BLAST+: architecture and applications. *BMC Bioinform.* 10, 421.
- [8] Chen, D., Xu, G., Tang, W., Jing, Y., Ji, Q., Fei, Z. and Lin, R. (2013) Antagonistic basic helix-loop-helix/bZIP transcription factors form transcriptional modules that integrate light and reactive oxygen species signaling in *Arabidopsis*. *Plant Cell*, 25, 1657-1673.
- [9] Clark, S.C., Egan, R., Frazier, P.I. and Wang, Z. (2013) ALE: a generic assembly likelihood evaluation framework for assessing the accuracy of genome and metagenome assemblies. *Bioinformatics*, 29, 435-443.
- [10] Cock, P.J., Antao, T., Chang, J.T. et al. (2009) Biopython: freely available python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 25, 1422-1423.
- [11] Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29, 15-21.
- [12] Douglas, G.M., Gos, G., Steige, K.A. et al. (2015) Hybrid origins and the earliest stages of diploidization in the highly successful recent polyploid *Capsella bursa-pastoris*. *Proc. Natl Acad. Sci.* 112, 2806-2811.
- [13] Doyle, J. and Doyle, J. (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 11-15.
- [14] Du, X., Servin, B., Womack, J.E., Cao, J., Yu, M., Dong, Y., Wang, W. and Zhao, S. (2014) An update of the goat genome assembly using dense radiation hybrid maps allows detailed analysis of evolutionary rearrangements in Bovidae. *BMC Genom.* 15, 625.
- [15] Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucl. Acids Res.* 32, 1792-1797.
- [16] Edwards, K.D., Bombarely, A., Story, G.W., Allen, F., Mueller, L.A., Coates, S.A. and Jones, L. (2010) TobEA: an atlas of tobacco gene expression from seed to senescence. *BMC Genom.* 11, 142.
- [17] Forrest, A.R.R., Kawaji, H., Rehli, M. et al. (2014) A promoter-level mammalian expression atlas. *Nature*, 507, 462-470.
- [18] Goodstein, D.M., Shu, S., Howson, R. et al. (2012) Phytozome: a comparative platform for green plant genomics. *Nucl. Acids Res.* 40, D1178-D1186.
- [19] Han, T.-S., Wu, Q., Hou, X.-H., Li, Z.-W., Zou, Y.-P., Ge, S. and Guo, Y.-L. (2015) Frequent introgressions from diploid species contribute to the adaptation of the tetraploid Shepherd's Purse (*Capsella bursa-pastoris*). *Mol. Plant.* 8, 427-438.
- [20] Hegarty, M.J., Barker, G.L., Wilson, I.D., Abbott, R.J., Edwards, K.J. and Hiscock, S.J. (2006) Transcriptome shock after interspecific hybridization in senecio is ameliorated by genome duplication. *Curr. Biol.* 16, 1652-1659.
- [21] Hehl, R. and Bülow, L. (2014) AthaMap web tools for the analysis of transcriptional and posttranscriptional regulation of gene expression in *Arabidopsis thaliana*. *Meth. Mol. Biol.* Clifton NJ, 1158, 139-156.
- [22] Hehl, R., Norval, L., Romanov, A. and Bülow, L. (2016) Boosting AthaMap database content with data from protein binding microarrays. *Plant Cell Physiol.* 57, e4.
- [23] Hintz, M., Bartholmes, C., Nutt, P., Ziermann, J., Hameister, S., Neuffer, B. and Theissen, G. (2006) Catching a "hopeful monster": shepherd's purse (*Capsella bursa-pastoris*) as a model system to study the evolution of flower development. *J. Exp. Bot.* 57, 3531-3542.
- [24] Huang, D.W., Sherman, B.T. and Lempicki, R.A. (2009a) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucl. Acids Res.* 37, 1-13.
- [25] Huang, D.W., Sherman, B.T. and Lempicki, R.A. (2009b) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44-57.
- [26] Hurka, H., Freundner, S., Brown, A.H. and Plantholt, U. (1989) Aspartate aminotransferase isozymes in the genus *Capsella* (Brassicaceae): subcellular location, gene duplication, and polymorphism. *Biochem. Genet.* 27, 77-90.
- [27] Ibarra-Laclette, E., Lyons, E., Hernández-Guzmán, G. et al. (2013) Architecture and evolution of a minute plant genome. *Nature*, 498, 94-98.
- [28] Iorizzo, M., Ellison, S., Senalik, D. et al. (2016) A high-quality carrot genome assembly provides new insights into carotenoid accumulation and asterid genome evolution. *Nat. Genet.* 48, 657-666.
- [29] Jaillon, O., Aury, J.-M., Noel, B. et al. (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*, 449, 463-467.
- [30] Johnston, J.S. (2005) Evolution of genome size in Brassicaceae. *Ann. Bot.* 95, 229-235.
- [31] Kajitani, R., Toshimoto, K., Noguchi, H. et al. (2014) Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Res.* 24, 1384-1395.
- [32] Klepikova, A.V., Kasianov, A.S., Gerasimov, E.S., Logacheva, M.D. and Penin, A.A. (2016) A high resolution map of the *Arabidopsis thaliana* developmental transcriptome based on RNA-seq profiling. *Plant J.* 88, 1058-1070.

- [33] Kryvokhyzha, D., Holm, K., Chen, J., Cornille, A., Glémin, S., Wright, S.I., Lagercrantz, U. and Lascoux, M. (2016) The influence of population structure on gene expression and flowering time variation in the ubiquitous weed *Capsella bursa-pastoris* (Brassicaceae). *Mol. Ecol.* 25, 1106-1121.
- [34] Kulakovskiy, I.V., Boeva, V.A., Favorov, A.V. and Makeev, V.J. (2010) Deep and wide digging for binding motifs in ChIP-Seq data. *Bioinforma. Oxf. Engl.* 26, 2622-2623.
- [35] Kulakovskiy, I.V., Vorontsov, I.E., Yevshin, I.S. et al. (2016) HOCOMOCO: expansion and enhancement of the collection of transcription factor binding sites models. *Nucl. Acids Res.* 44, D116-D125.
- [36] Libault, M., Farmer, A., Joshi, T. et al. (2010) An integrated transcriptome atlas of the crop model *Glycine max*, and its use in comparative analyses in plants: soybean transcriptome atlas. *Plant J.* 63, 86-99.
- [37] Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550.
- [38] Magoc, T. and Salzberg, S.L. (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, 27, 2957-2963.
- [39] Mathelier, A., Fornes, O., Arenillas, D.J. et al. (2016) JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucl. Acids Res.* 44, D110-D115.
- [40] Mézard, C. (2006) Meiotic recombination hotspots in plants: figure 1. *Biochem. Soc. Trans.* 34, 531-534.
- [41] Mironova, V.V., Omelyanchuk, N.A., Wiebe, D.S. and Levitsky, V.G. (2014) Computational analysis of auxin responsive elements in the *Arabidopsis thaliana* L. genome. *BMC Genom.* 15, S4.
- [42] Ni, M., Tepperman, J.M. and Quail, P.H. (1998) PIF3, a phytochrome-interacting factor necessary for normal photoinduced signal transduction, is a novel basic helix-loop-helix protein. *Cell*, 95, 657-667.
- [43] Nielsen, R., Paul, J.S., Albrechtsen, A. and Song, Y.S. (2011) Genotype and SNP calling from next-generation sequencing data. *Nat. Rev. Genet.* 12, 443-451.
- [44] Nutt, P., Ziermann, J., Hintz, M., Neuffer, B. and Theißen, G. (2006) *Capsella* as a model system to study the evolutionary relevance of floral homeotic mutants. *Plant Syst. Evol.* 259, 217-235.
- [45] O'Malley, R.C., Huang, S.-S.C., Song, L., Lewsey, M.G., Bartlett, A., Nery, J.R., Galli, M., Gallavotti, A. and Ecker, J.R. (2016) Cistrome and episcistrome features shape the regulatory DNA landscape. *Cell*, 165, 1280-1292.
- [46] Ortiz-Ramírez, C., Hernandez-Coronado, M., Thamm, A., Catarino, B., Wang, M., Dolan, L., Feijó, J.A. and Becker, J.D. (2016) A transcriptome atlas of *Physcomitrella patens* provides insights into the evolution and development of land plants. *Mol. Plant.* 9, 205-220.
- [47] Park, S., Lee, C.-M., Doherty, C.J., Gilmour, S.J., Kim, Y. and Thomashow, M.F. (2015) Regulation of the *Arabidopsis* CBF regulon by a complex low-temperature regulatory network. *Plant J.* 82, 193-207.
- [48] Sato, Y., Takehisa, H., Kamatsuki, K. et al. (2013) RiceXPro Version 3.0: expanding the informatics resource for rice transcriptome. *Nucl. Acids Res.* 41, D1206-D1213.
- [49] Schmid, M., Davison, T.S., Henz, S.R., Pape, U.J., Demar, M., Vingron, M., Schölkopf, B., Weigel, D. and Lohmann, J.U. (2005) A gene expression map of *Arabidopsis thaliana* development. *Nat. Genet.* 37, 501-506.
- [50] Severin, A.J., Woody, J.L., Bolon, Y.-T. et al. (2010) RNA-Seq Atlas of *Glycine max*: a guide to the soybean transcriptome. *BMC Plant Biol.* 10, 160.
- [51] Slotte, T., Ceplitis, A., Neuffer, B., Hurka, H. and Lascoux, M. (2006) Intrageneric phylogeny of *Capsella* (Brassicaceae) and the origin of the tetraploid *C. bursa-pastoris* based on chloroplast and nuclear DNA sequences. *Am. J. Bot.* 93, 1714-1724.
- [52] Slotte, T., Holm, K., McIntyre, L.M., Lagercrantz, U. and Lascoux, M. (2007) Differential expression of genes important for adaptation in *Capsella bursa-pastoris* (Brassicaceae). *Plant Physiol.* 145, 160-173.
- [53] Slotte, T., Huang, H.-R., Holm, K., Ceplitis, A., Onge, K.S., Chen, J., Lagercrantz, U. and Lascoux, M. (2009) Splicing variation at a FLOWERING LOCUS C homeolog is associated with flowering time variation in the tetraploid *Capsella bursa-pastoris*. *Genetics*, 183, 337-345.
- [54] Slotte, T., Hazzouri, K.M., Ågren, J.A. et al. (2013) The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. *Nat. Genet.* 45, 831-835.
- [55] Soy, J., Leivar, P., González-Schain, N., Sentandreu, M., Prat, S., Quail, P.H. and Monte, E. (2012) Phytochrome-imposed oscillations in PIF3 protein abundance regulate hypocotyl growth under diurnal light/dark conditions in *Arabidopsis*: PIF3 promotes growth under short days. *Plant J.* 71, 390-401.
- [56] St Onge, K.R., Foxe, J.P., Li, J., Li, H., Holm, K., Corcoran, P., Slotte, T., Lascoux, M. and Wright, S.I. (2012) Coalescent-based analysis distinguishes between allo- and autopolyploid origin in Shepherd's Purse (*Capsella bursa-pastoris*). *Mol. Biol. Evol.* 29, 1721-1733.
- [57] Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30, 1312-1313.
- [58] Stanke, M., Schöffmann, O., Morgenstern, B. and Waack, S. (2006) Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinform.* 7, 62.

- [59] Stracke, R., Favory, J.-J., Gruber, H., Bartelniewoehner, L., Bartels, S., Binkert, M., Funk, M., Weisshaar, B. and Ulm, R. (2010) The Arabidopsis bZIP transcription factor HY5 regulates expression of the PFG1/MYB12 gene in response to light and ultraviolet-B radiation. *Plant, Cell Environ.* Available at: <http://doi.wiley.com/10.1111/j.1365-3040.2009.02061.x> (accessed 8 October 2016).
- [60] Tomato Genome Consortium (2012) The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*, 485, 635–641.
- [61] Vorontsov, I.E., Kulakovskiy, I.V. and Makeev, V.J. (2013) Jaccard index based similarity measure to compare transcription factor binding site models. *Algorithms Mol. Biol. AMB*, 8, 23.
- [62] Wesse, C., Pietzka, D., Tietmeyer, J. and Neuffer, B. (2016) Latitudinal genome size variation in *Capsella bursa-pastoris*. Available at: <https://doi.org/10.13140/rg.2.2.36595.32805> (accessed 19 February 2017).
- [63] Yang, J., Liu, D., Wang, X. et al. (2016) The genome sequence of allopolyploid *Brassica juncea* and analysis of differential homoeolog gene expression influencing selection. *Nat. Genet.* Available at: <http://www.nature.com/doi/10.1038/ng.3657> (accessed 8 September 2016).
- [64] Yilmaz, A., Mejia-Guerra, M.K., Kurz, K., Liang, X., Welch, L. and Grotewold, E. (2011) AGRIS: the Arabidopsis gene regulatory information server, an update. *Nucl. Acids Res.* 39, D1118–D1122.
- [65] Yu, J., Wang, J., Lin, W. et al. (2005) The genomes of *Oryza sativa*: a history of duplications. *PLoS Biol.* 3, e38.
- [66] Zhang, B., Wang, L., Zeng, L., Zhang, C. and Ma, H. (2015) Arabidopsis TOE proteins convey a photoperiodic signal to antagonize CONSTANS and regulate flowering time. *Genes Dev.* 29, 975–987.
- [67] Zhang, H., Gou, X., Zhang, A., Wang, X., Zhao, N., Dong, Y., Li, L. and Liu, B. (2016a) Transcriptome shock invokes disruption of parental expression-conserved genes in tetraploid wheat. *Sci. Rep.* 6, 26363.
- [68] Zhang, Y., Liu, Z., Khan, A.A. et al. (2016b) Expression partitioning of homeologs and tandem duplications contribute to salt tolerance in wheat (*Triticum aestivum* L.). *Sci. Rep.* 6, 21476.