

<https://helda.helsinki.fi>

First draft genome assembly of Coleochaete orbicularis

Sablok, Gaurav

2019-09-29

Sablok , G , He , X , Miranto , M , Peltomaa , E , Sleith , R , Karol , K , Delwiche , C , Bell , N , Paulin , L , Poczai , P & Hyvönen , J 2019 , ' First draft genome assembly of Coleochaete orbicularis ' , 5th Conference on Plant Genome Evolution , Spain , 29/09/2019 - 01/10/2019 .

<http://hdl.handle.net/10138/305923>

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

First draft genome assembly of *Coleochaete orbicularis*

Gaurav Sablok^{1,2}, Xiaolan He¹, Mari Miranto¹, Elina Peltomaa³, Robin Sleith⁴, Kenneth Karol⁴, Charles Delwiche⁵, Neil Bell⁶, Lars Paulin⁷, Peter Poczai¹, Jaakko Hyvonen^{1,2*}

¹Finnish Museum of Natural History, University of Helsinki, Finland; ²OEB, Faculty of Biology and Environmental Sciences, University of Helsinki, Finland; ³ECOENV, Faculty of Biology and Environmental Sciences, University of Helsinki, Finland; ⁴New York Botanical Garden, USA; ⁵University of Maryland USA; ⁶Royal Botanic Garden, Edinburgh, UK; ⁷Institute of Biotechnology, University of Helsinki, Finland.

*Email: jaakko.hyvonen@helsinki.fi

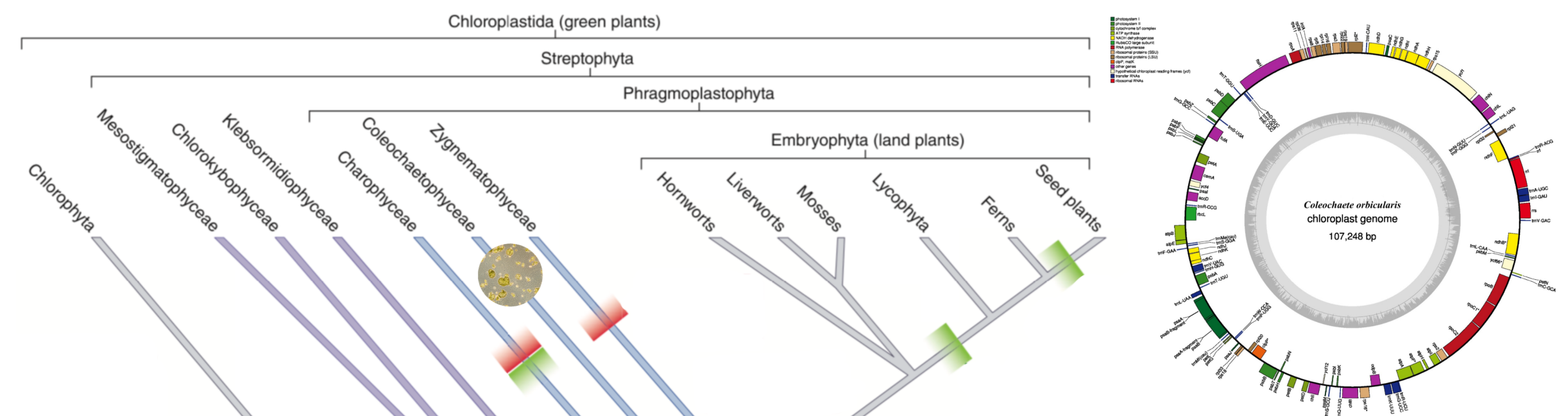
Introduction:

Comparative genomics of the embryophytes (land plants) provide basis to understand their divergence and adaptation to the terrestrial environment. The focus of the genome sequencing has been on the understanding of the patterns of genetic development and adaptation, particularly of the crop and some model species. Rapidly increased transcriptome sequencing effort and availability of the ample genetic marker data have allowed comparisons of early embryophytes and tracheophytes (vascular plants). The key to understand the deepest nodes requires the availability of the genomes and high-quality gene models from the early lineages of the streptophytes to understand the genomic synteny, gene duplication and losses as well as the diverged genes across the streptophytes, and in particular embryophytes. We present the first genome assembly of *Coleochaete orbicularis* and compare it also with the recently released genome of *Chara braunii*, another freshwater green alga.

Material and Methods:

In-vitro grown cultures of *Coleochaete orbicularis* (UTEX LB 2651) on MS medium were used for genomic DNA extraction with the Plant-EZ DNA extraction kit and were sequenced using PacBio-SMRT at the Institute of Biotechnology, University of Helsinki, Finland. Additionally, Illumina sequencing was performed using Nova-seq (150PE) with an average insert size of ~450-600bp. Prior to genome assembly, reads were mapped to the contamination database for removal of the contaminating reads using bowtie2 (Langmead, 2012). Filtered genomic reads were assembled using SGA (String Graph Assembler, Simpson, 2014). Additionally, a hybrid assembly was also assembled incorporating the Nova-seq and the PacBio reads using the MaSuRCA genome assembler (Zimin et al. 2013). Assembly contiguity was evaluated using the LASTz alignment program (Harris, 2007) against *Chara braunii* genome (Nishiyama et al. 2018) and coverage plots were visualized with Circos (Krzywinski et al. 2009). Reads were mapped to the assembled genome and genome coverage was visualized using the Integrated Genomics Viewer (2011).

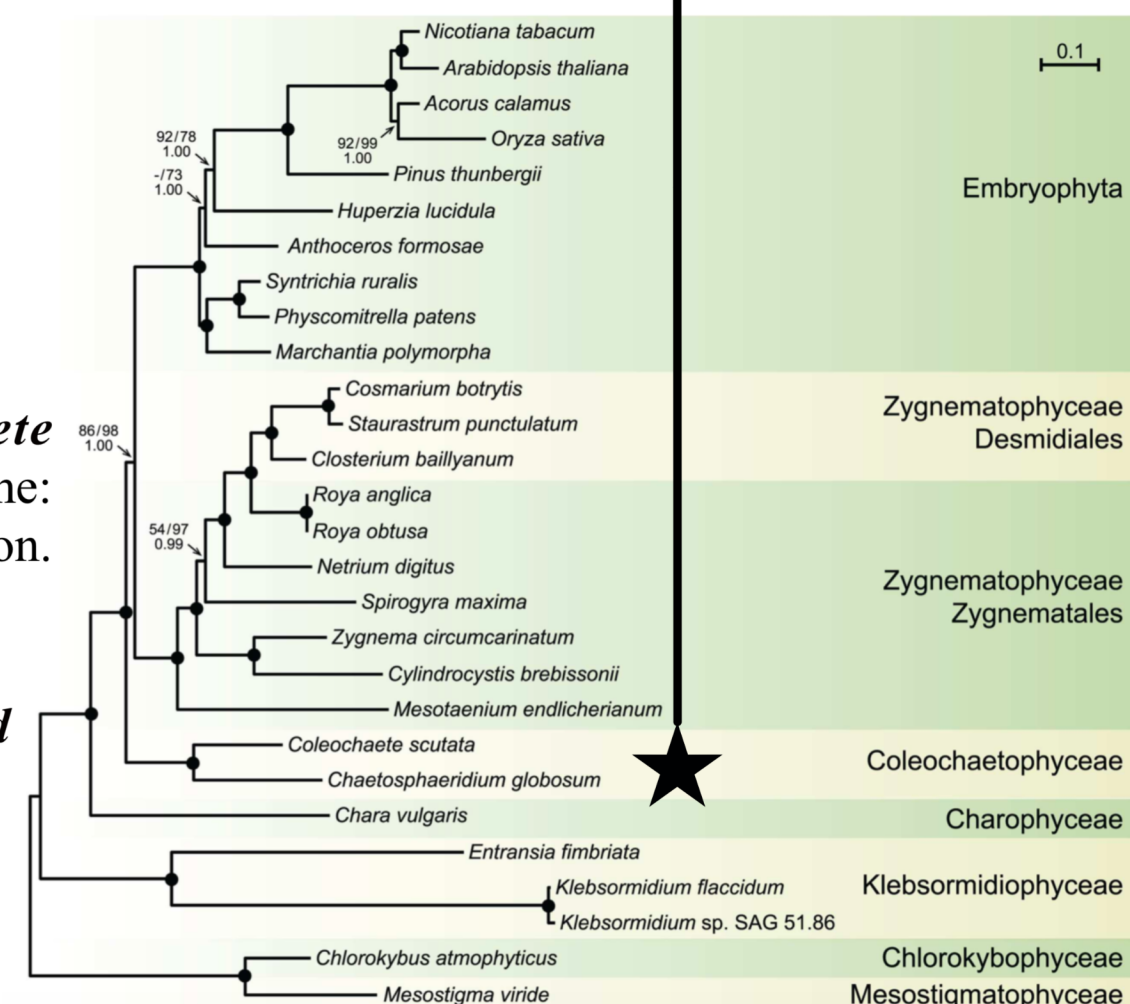
Position of *Coleochaete orbicularis* as based on recent phylogenetic analyses



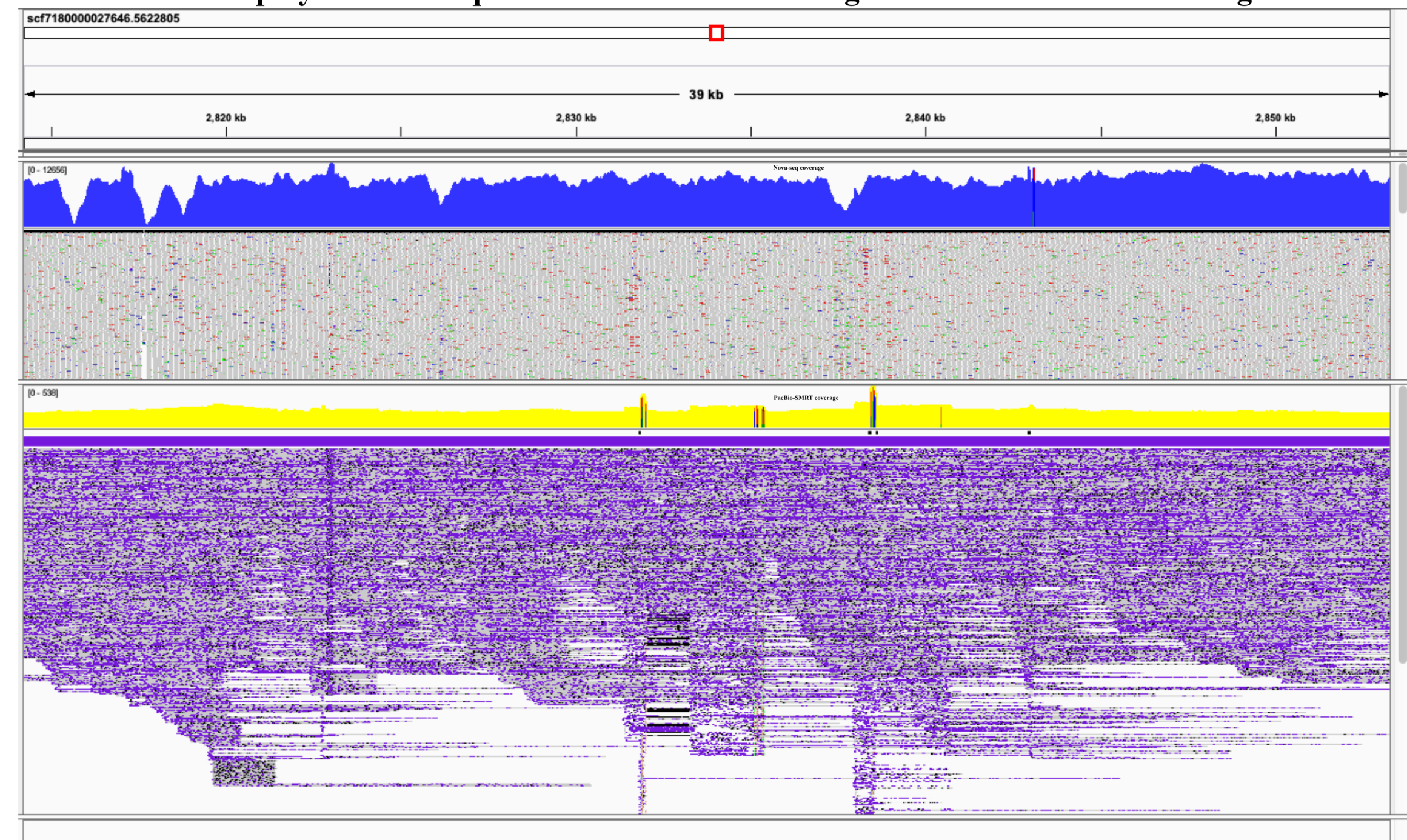
Sablok et al. (2018) Plastomics of charophycean *Coleochaete orbicularis* and embryophyte *Blasia pusilla*. European Conference on Computational Biology, poster 541.

Systematics and phylogenetic placement of *Coleochaete orbicularis* based on Nishiyama T et al. (2018) The Chara Genome: Secondary Complexity and Implications for Plant Terrestrialization. Cell.174:448-464.e24.

Systematics and phylogenetic placement of *Coleochaete orbicularis* based on Lemieux C et al. (2016) Front Plant Sci. 2016;7:697



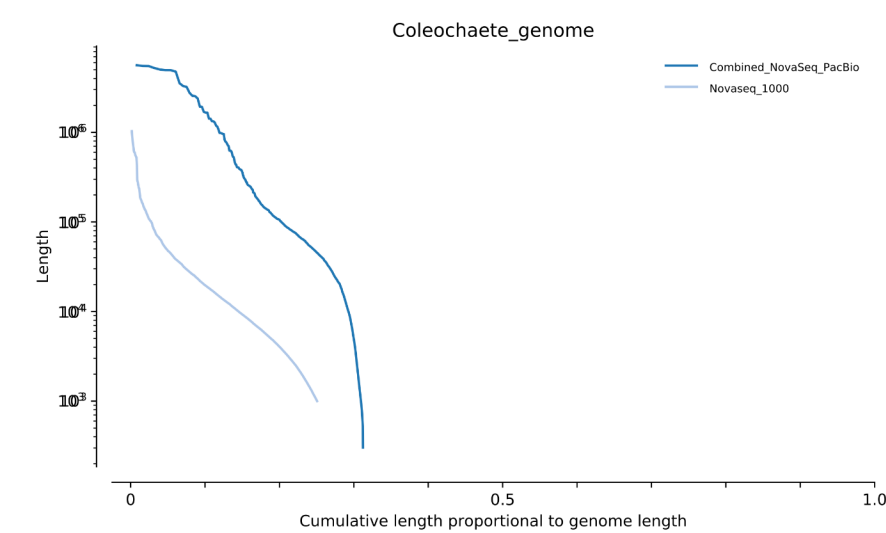
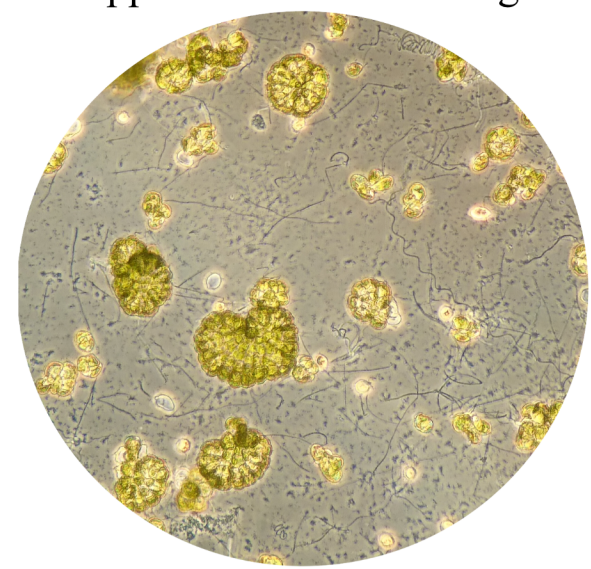
IGV display of Nova-seq and PacBio-SMRT coverage of *Coleochaete orbicularis* genome



References:

Robinson JT et al (2011) Nat Biotechnol. 29:24–26; Simpson JT (2014) 30:1228–1235; Zimin et al. (2013) Bioinformatics 29:2669–2677; Langmead et al. (2012) Nature Methods 9:357–359; Harris (2007) Penn State University; Krzywinski et al. (2009) Genome Research 19:1639-1645; Lemieux C et al. (2016) Front Plant Sci. 7:697; Nishiyama T et al. (2018) The Chara Genome: Secondary Complexity and Implications for Plant Terrestrialization. Cell.174:448-464.e24; Sablok et al. (2018) European Conference on Computational Biology, poster 541.

Acknowledgements: Academy of Finland for their grant 295595 to JH & XH, and to CSC - IT Centre for Science for providing computing environment.



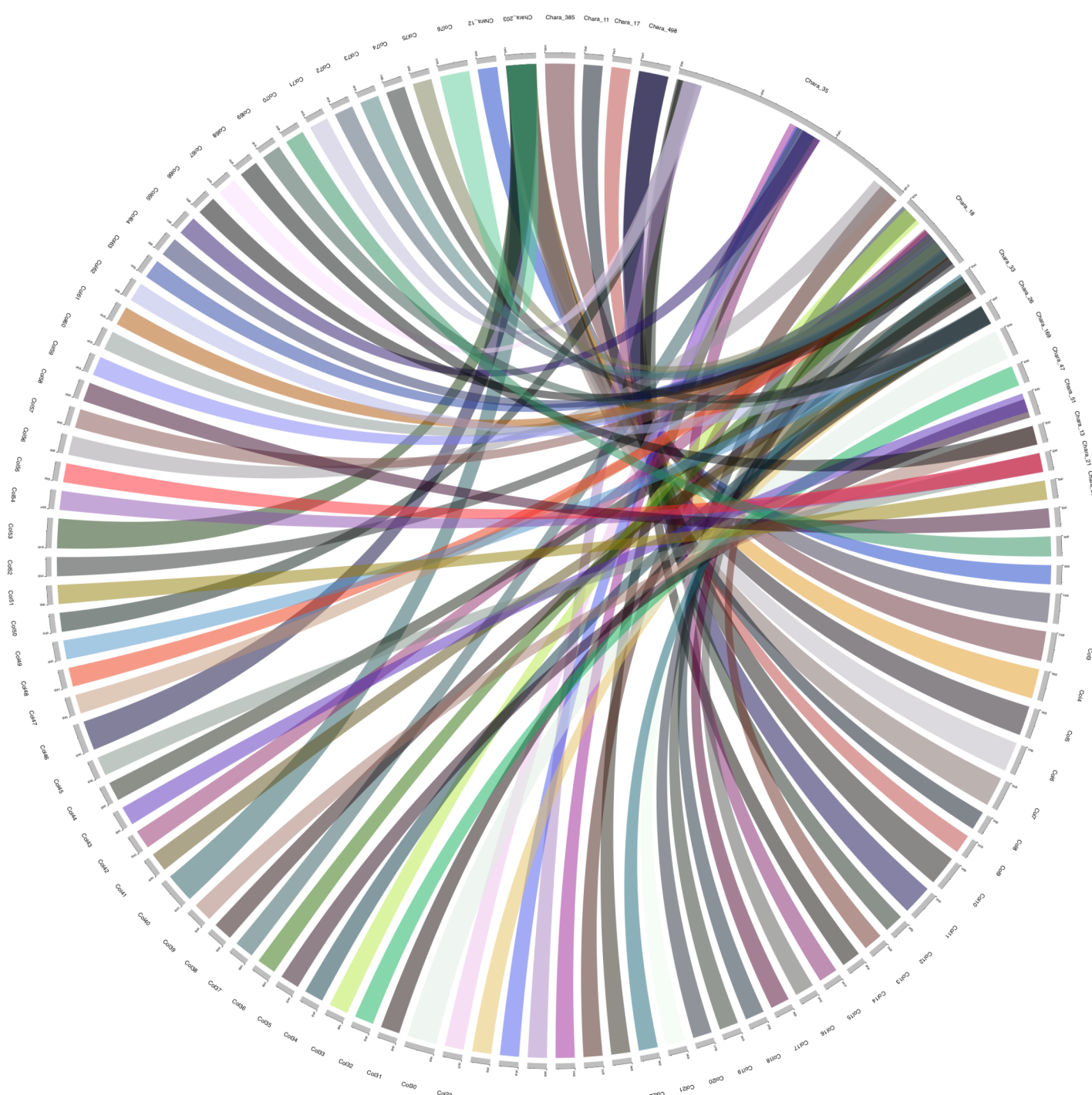
Coleochaete orbicularis

Genome assembly stats of *Coleochaete orbicularis*

<i>Coleochaete orbicularis</i> Nova-Seq	4.082728 G
<i>Coleochaete orbicularis</i> Filtered Nova-Seq	3.804648 G
<i>Coleochaete orbicularis</i> PacBio SMRT	1182891
<i>Coleochaete orbicularis</i> Filtered PacBio SMRT	1151746
<i>Coleochaete orbicularis</i> Genome Assembly	215226293bp
<i>Coleochaete orbicularis</i> N50	263465bp
<i>Coleochaete orbicularis</i> Number of Scaffolds	8758
<i>Coleochaete orbicularis</i> Number of Scaffolds (> 2000bp)	4054
Number of <i>Coleochaete orbicularis</i> scaffolds mapped to <i>Chara braunii</i> genome*	979/2132

*LASTz identity=50, coverage=50

Comparative visualization of the *Coleochaete orbicularis* genome assembly against *Chara braunii*



***Coleochaete orbicularis* scaffolds > 12KB are mapped to *Chara braunii* genome**