

De novo Assembly and Comparative Genomics

on Eukaryotic Species Mixtures

Bastian Greshake[†], Andreas Blaumeiser[†], Simonida Zehr[†],

Francesco Dal Grande^{*}, Anjuli Meiser[§], Imke Schmitt^{*§}, Ingo Ebersberger[†]

[†] Department for Applied Bioinformatics, Institute for Cell Biology and Neuroscience, Goethe University, Frankfurt am Main, Germany

^{*} Biodiversity and Climate Research Centre, Senckenberg Gesellschaft für Naturforschung, Frankfurt am Main, Germany

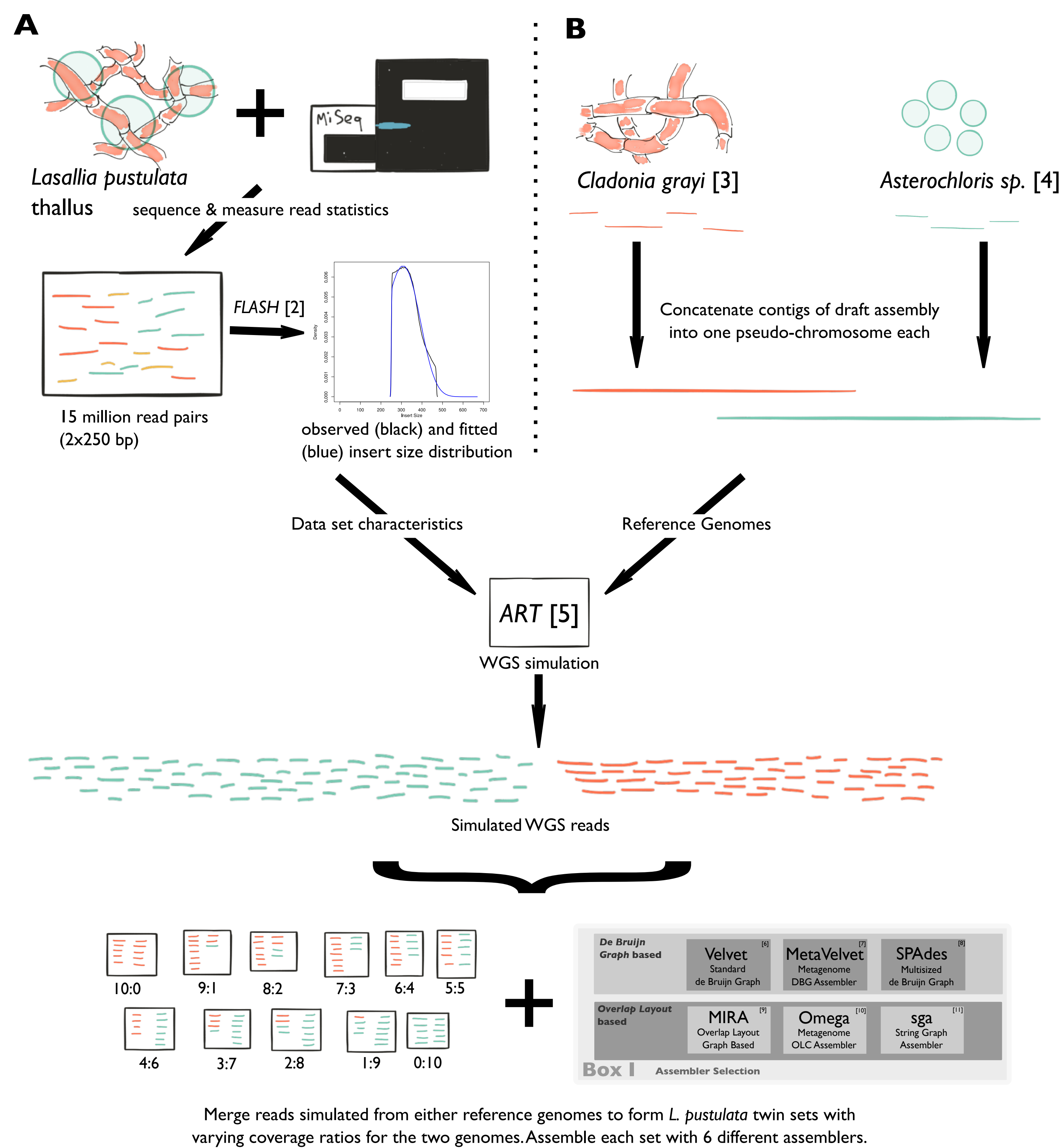
[§] Institute of Ecology, Evolution and Diversity, Goethe University, Frankfurt am Main, Germany



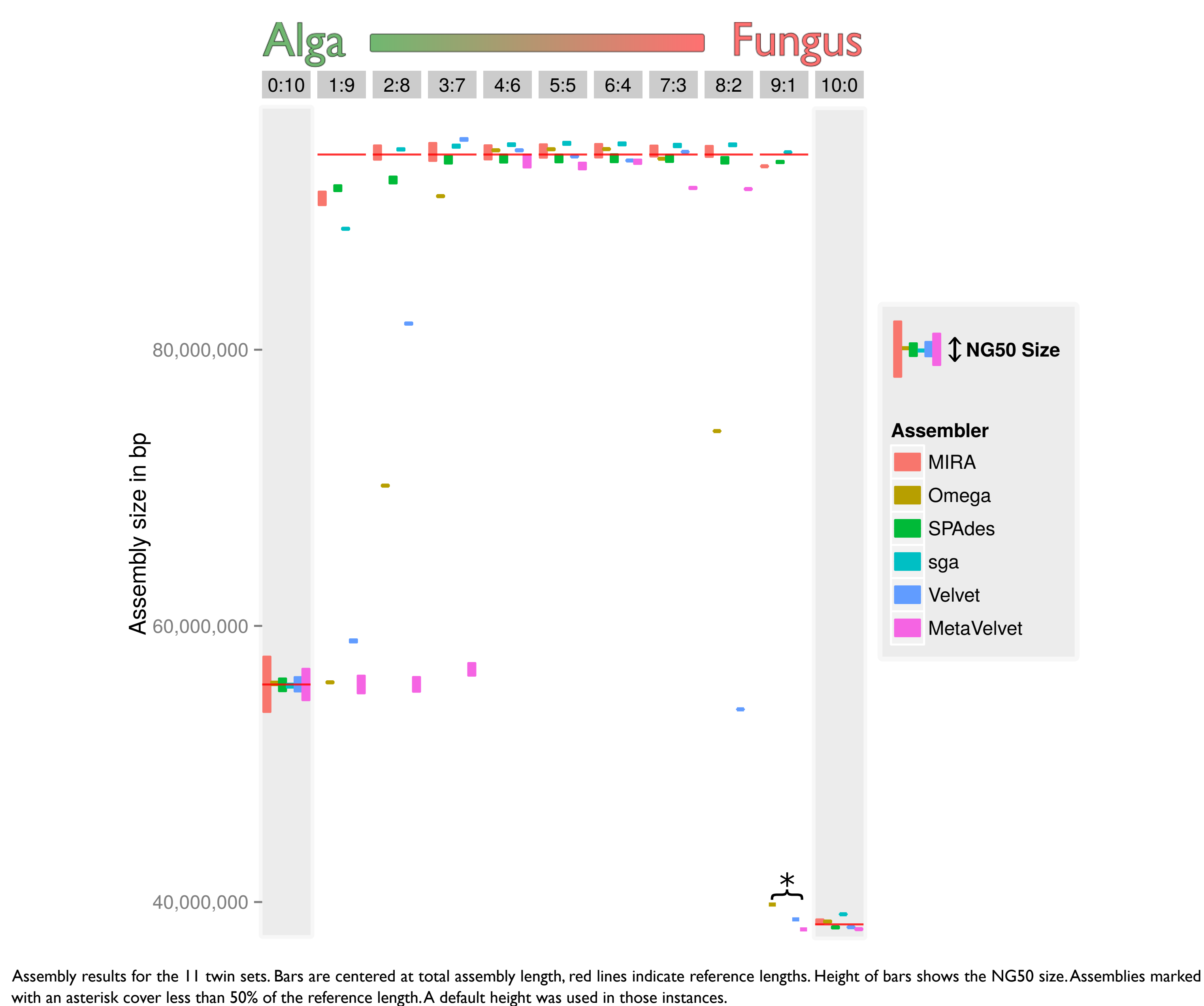
Summary

Mutualistic symbiotic relationships are found across organisms of all complexity. In extreme instances, as in some lichens, the interaction appears so close that the participating organisms grow only poorly – or even not at all – when cultivated in isolation. This renders mutualistic symbionts valuable objects to study the genomic basis of adaptation and co-evolution. The close interdependence in such communities, however, confounds genomic studies. In many cases separate sequencing of the participating organisms is not feasible, leaving metagenomics approaches as the method of choice. Here we address how and to what extent eukaryotic genomes can be reconstructed from such data.

I. Assembler Evaluation with Simulated Twin Sets [1]

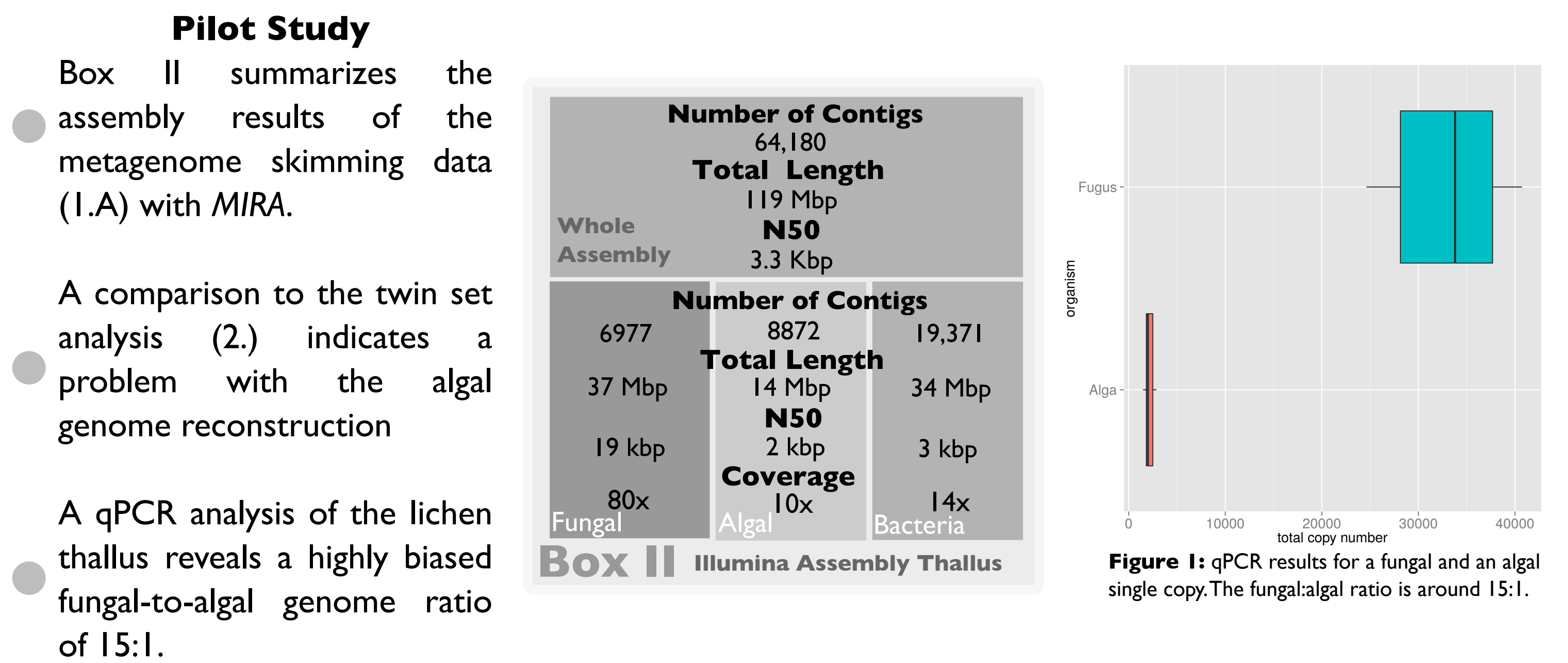


2. Assembly Results of the Twin Sets

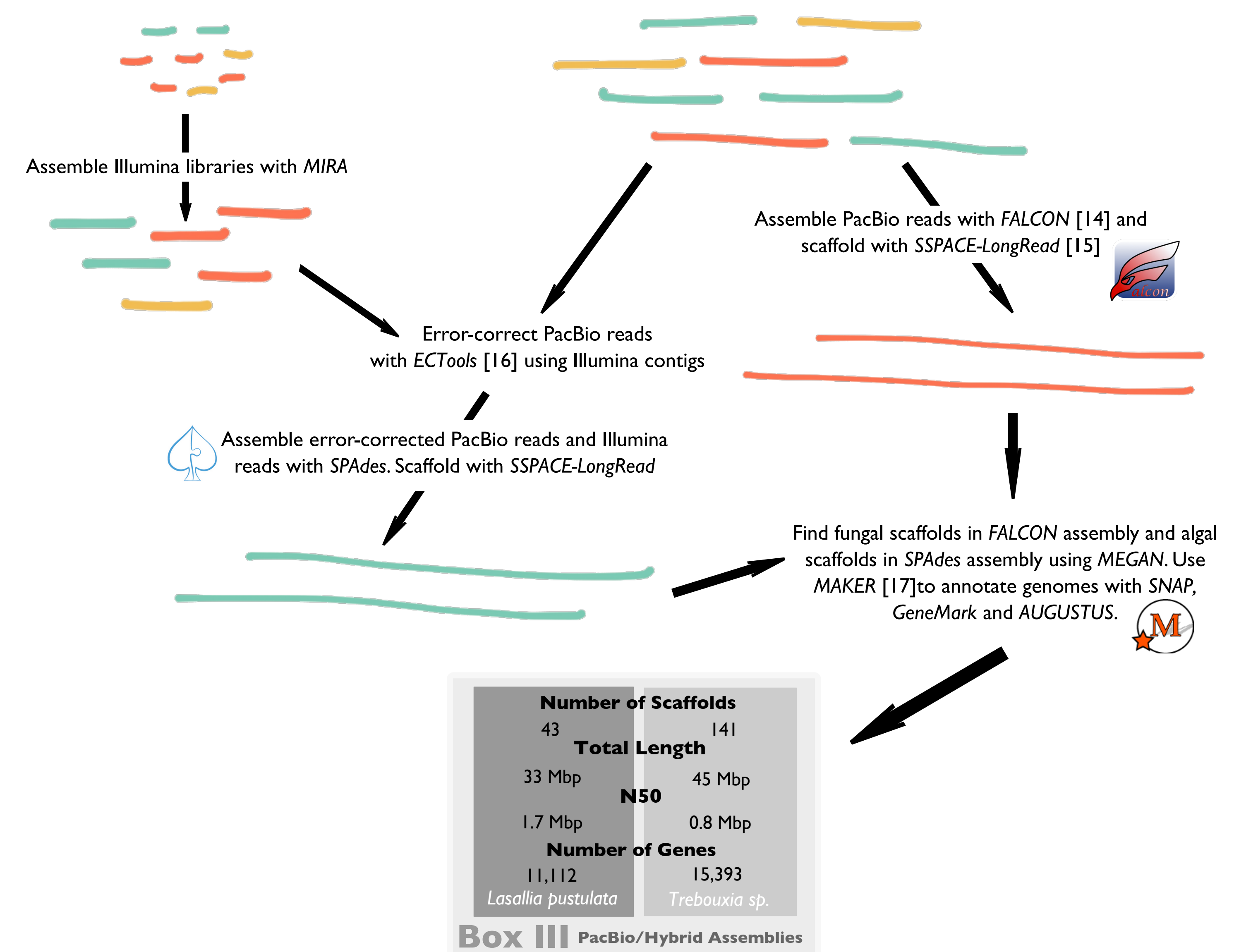


We use in silico-generated data sets to sound out the performance of different assembly paradigms on Whole Genome Shotgun data from eukaryotic species mixtures. On this basis we have begun reconstructing the metagenome of *Lasallia pustulata*. Using a hybrid sequencing approach, combining Illumina short read and PacBio long read data, we have assembled the genome of the mycobiont and a major fraction of the algal photobiont. We integrate this data with genome sequences of closely related non-lichenized fungi as a first step towards analyzing how lichenization affects genome evolution.

3. Sequencing the *L. pustulata* metagenome



Hybrid Assembly: Short-Read meets Long-Read PacBio sequencing was done alongside further Illumina sequencing: Using PacBio 2,705,256 polymerase reads with a read N50 of 15kb were sequenced. A 250 bp mate pair library (5kb inserts) of 15 million reads was sequenced using Illumina MiSeq. To cope with the coverage differences we performed two different assemblies, targeting the fungal and the algal genome respectively. For high coverage data PacBio-only assemblies are state of the art, low-coverage data require hybrid assemblies using Illumina and PacBio data [13].

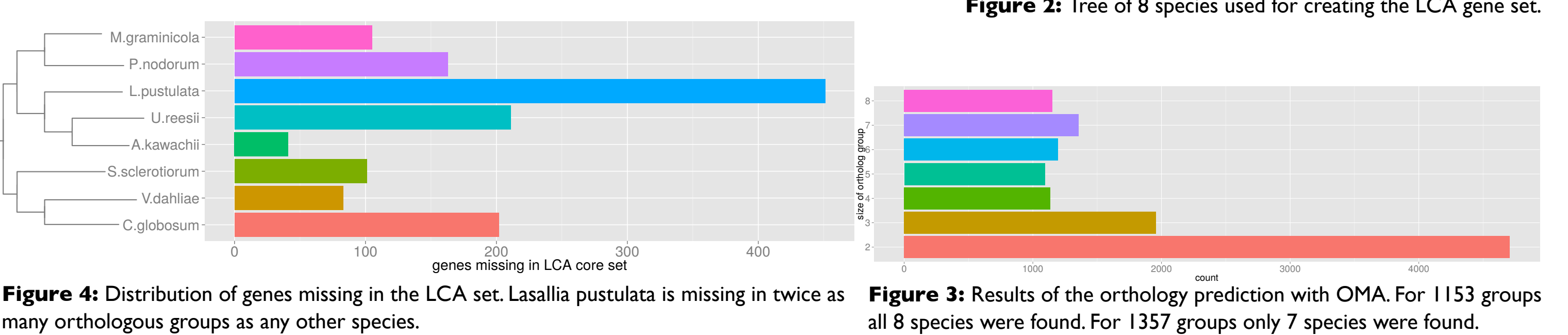


Outlining the assembly workflow for the genomes of the mycobiont and the photobiont. The mycobiont can be assembled directly from the PacBio data. The photobiont is assembled using a hybrid approach using PacBio and Illumina data.

4. Does Lichenization Facilitate Gene Loss?

Ancestral Gene Set To investigate lineage specific gene loss, the Last Common Ancestor (LCA) gene set of the Pezizomycotina was reconstructed using OMA [18] (Figure 2). In total 12,595 orthologous groups were formed (Figure 3).

Absence of LCA Genes For 1357 further groups genes were only found in 7 species. For these groups it is *L. pustulata* which is absent most often, hinting that these genes are lost over time in lichenization (Figure 4).



[1] Greshake B, Zehr S, Dal Grande F et al. Mol Ecol Res (2015) epub ahead of print
[2] Magoc T and Salzberg S. Bioinformatics (2011) 27 (21):2957-63
[3] <http://genome.jgi.doe.gov/Claag2/>
[4] <http://genome.jgi.doe.gov/Aspho2/>
[5] Huang W, Li L, Myers JR, Marth GT. Bioinformatics (2012) 28 (4):593-594
[6] Zerbinio DR and Birney E. Genome Research (2008) 18:821-829.
[7] Namiki T, Hachiya T, Tanaka H, Sakakibara Y. Nucleic Acids Res. (2012) 40(20), e155
[8] Bankevich A, Nurk S, Antipov D et al. Journal of Computational Biology (2012) 19(5):455-477
[9] <http://sourceforge.net/projects/mira-assembler/>
[10] Haider B, Ahn T, Bushnell B et al. Bioinformatics (2014) 29:395
[11] Simpson JT and Durbin R. Bioinformatics (2010) 26 (12):1367-1373
[12] Huson DH, Mitra S, Ruscheweyh H et al. Genome Research (2011) 21: 1552-1560
[13] Mike Schatz, PAG 2014 (<http://schatzlab.cshl.edu/presentations/2014-01-14.PAG.Single%20Molecule%20Assembly.pdf>)
[14] <https://github.com/PacificBiosciences/FALCON>
[15] Boetzer M and Pirovano V. BMC Bioinformatics (2014) 15:211
[16] <https://github.com/jgurtowski/ectools>
[17] Campbell MS, Holt C, Moore B, Yandell M. Curr Protoc Bioinformatics (2014) 48:4.11.1-4.11.39
[18] <http://omabrowser.org/standalone/>



Contact

Bastian Greshake
bgreshake@gmail.com
Goethe University, Frankfurt am Main, Germany
Max-von-Laue-Straße 13, 60438 Frankfurt am Main

References

